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

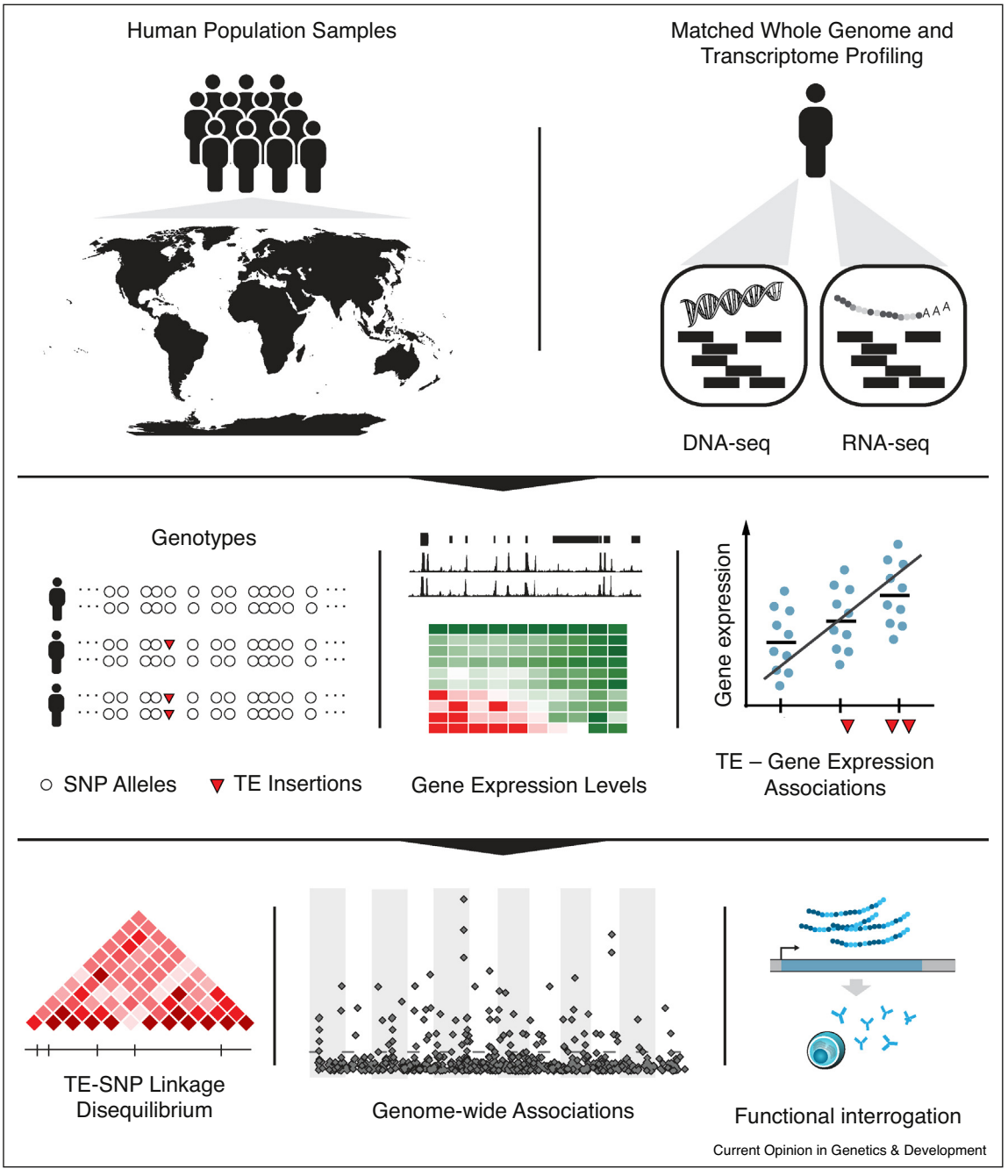
Transposable elements (TEs) are distinguished by their ability to move, i.e. transpose, among genomic locations, often making copies of themselves as they go. TEs can replicate to extremely high copy numbers over time; at least 50% of the human genome sequence is thought to be derived from TE insertions [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.

to differences in gene expression patterns between individuals. Furthermore, human regulatory variation generated by recent TE activity may have important implications for health and disease. This mini-review is focused on recent studies that are beginning to shed light on the ways in which ongoing TE activity can impact human health via changes in genome regulation. These studies are distinguished by their population level approach to the study of TE generated human variation (Figure 1).

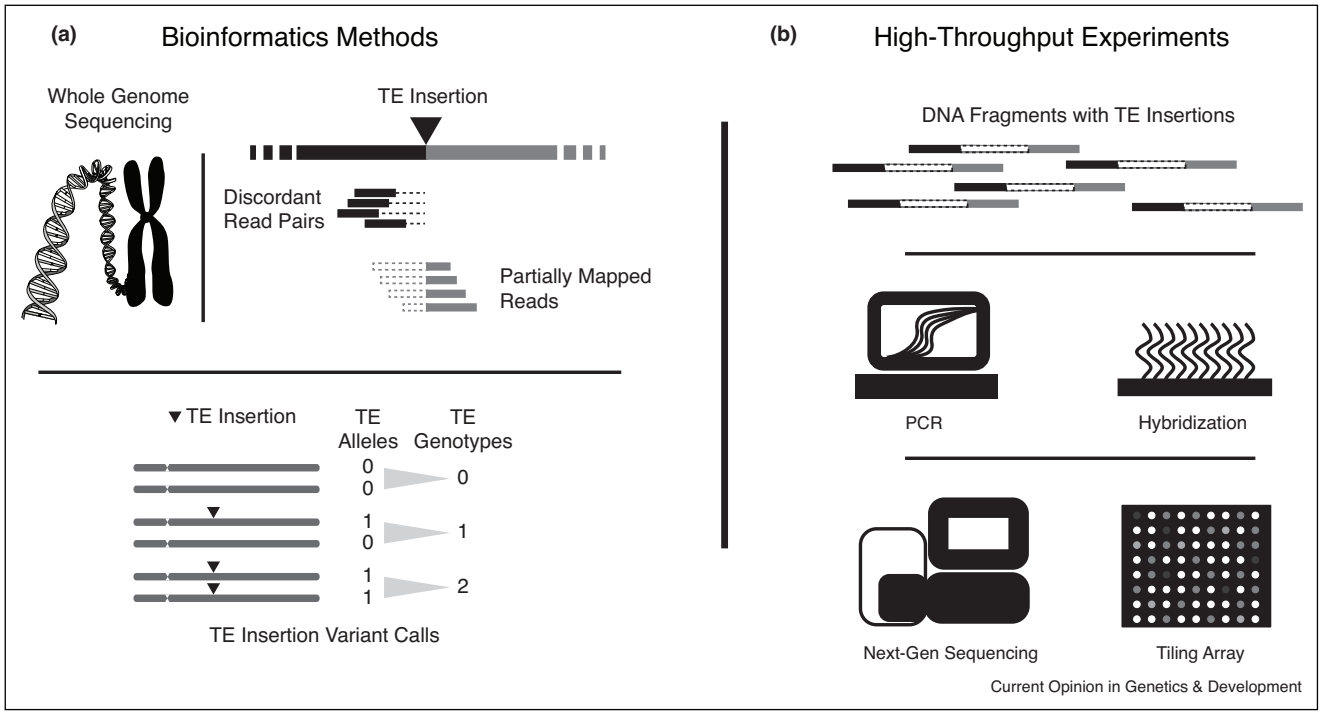
Genome-enabled approaches for characterizing TE insertion variants

Two distinct classes of genome-enabled approaches for the characterization of TE insertion variants have emerged over the last several years [34•]: (1) bioinformatics methods that rely on the analysis of whole genome sequence data to find TE insertions that differ from a reference sequence (Figure 2A), and (2) high-throughput experimental methods that utilize next-generation sequencing to pinpoint the locations of novel TE insertions (Figure 2B).

Computational approaches for the discovery of TE insertion variants rely on one of two methods: (1) discordant read-pair mapping for short read sequencing technology, or (2) split read mapping for long read technology [35•]. Our own group recently performed a benchmarking study on 21 bioinformatics tools designed for detecting human TE insertion variants from whole genome sequence data [36••]. After an initial screen of tools that were found to be unreliable, or no longer maintained, our study focused on seven programs: ITIS [37], MELT [38•], Mobster [39], RetroSeq [40], Tangram [41], TEMP [42], and T-lex2 [43]. We found MELT to have superior performance for human TE variant detection from whole genome sequence data, but also show how a combined approach using two or more methods, including Mobster and RetroSeq, could yield superior performance. Since the publication of our paper, two new computational tools for TE insertion discovery have been published. The program STEAK [44] claims superior performance compared to existing short read methods, whereas LoRTE [45] is designed for PacBio® long read sequence technology.

At this time, given the predominance of Illumina® short read sequencing technology, discordant read-pair mapping approaches are most widely used. It should be noted that some short read methods also employ split, clipped, or insertion junction reads, in addition to discordant read-pair mapping, as part of their TE detection protocols. Nevertheless, these short read methods are still far from perfect and there is substantial room for additional development in the field. As long read sequencing technology becomes more widespread, split read approaches should

Figure 2



Genome-enabled approaches for the discovery and characterization of TE insertion variants. (A) Bioinformatics methods rely on the computational analysis of whole genome sequence read data to characterize genome-wide patterns of TE insertion alleles and genotypes. (B) High-throughput experimental methods use enrichment of genomic fragments that contain known active TE sequences followed by sequence or array based characterization of their genomic locations.

become more popular. Perhaps more importantly, we expect that split read approaches with long reads will be inherently more accurate and reliable than discordant read pair mapping, since long reads that span entire TE insertions should be mapped with much less ambiguity than shorter reads. Long reads should also help to disambiguate complex structural variants resulting from nested TE insertions.

High-throughput experimental techniques for TE variant detection also share several basic features: (1) DNA fragmentation, (2) TE enrichment, and (3) TE calling. The methods are distinguished by the approaches used for each step of the process. DNA fragmentation can be achieved via enzymatic digestion or by mechanical shearing. TE enrichment can be performed using PCR, with active TE-specific primers, or with hybridization to active TE-specific probes. Finally, TE calling is done using next-generation sequencing, for more recent methods, or with tiling arrays for the older methods. The most widely used experimental methods for TE variant detection include ME-Scan [46], L1-Seq [47], RC-seq [48], and Transposon-Seq [49]. One area of ongoing improvement for these methods entails the refinement of algorithms used to map enriched TE fragments to genome reference sequences. For example, the TIPseqHunter algorithm was recently developed to refine and improve human TE variant calls made by the existing TIP-seq experimental method [50].

Genome-scale experimental approaches of this kind have been most widely applied to the study of somatic TE variants that characterize cancer tissues. This is one of the most promising areas of recent human TE research, and it has been extensively reviewed elsewhere [51•]. This mini-review is focused instead on germline mutations that yield inter-individual differences in TE insertion patterns and manifest themselves as human population genetic variations, i.e. TE polymorphisms.

TE polymorphisms and human genome regulation

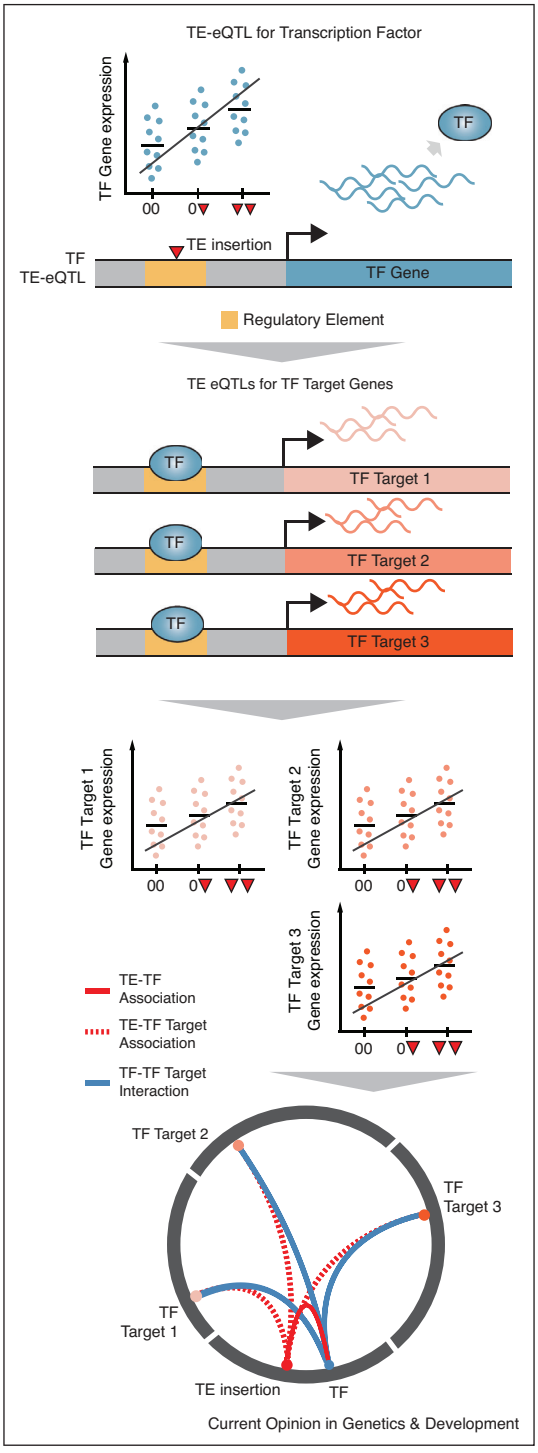
Our own group recently published a population-level view of the regulatory consequences of recent human TE activity [52••]. For this study, we adopted the expression quantitative trait loci (eQTL) analytical paradigm for the analysis of human TE polymorphisms. eQTL are genomic variants associated with changes in gene expression levels [53]. The eQTL approach requires multiple individual samples that have been deeply characterized at both the genomic (DNA-seq) and transcriptomic (RNA-seq) levels. Gene expression levels for individual samples are regressed against locus-specific genotypes for matched individuals to uncover eQTL associations. This approach was developed for single nucleotide polymorphism (SNP) genotypes, whereas in our case, we used locus-specific TE insertion state genotypes. TE insertion genotypes at any locus can be encoded as 0 (homozygous — insertion absent), 1 (heterozygous — one insertion present), or 2 (homozygous — two insertions present). Differences in gene expression levels across these distinct TE insertion states are indicative of TE polymorphism-to-gene expression associations (Figure 1). For the case of either SNPs or TE genotypes, the eQTL approach depends critically on the reliability of individual variant calls. Extensive benchmarking of SNP and TE variant callers has been performed, for example as part of the 1000 Genomes Project (1KGP), as previously described [54,55••]. As is standard for this kind of analysis, we only use variant calls that have been validated, including avoiding low frequency variants, for the purposes of eQTL analysis [52••].

This approach was powered by the 1KGP, phase 3 of which entailed the genome-wide characterization of TE insertion genotypes for 2504 individuals across 26 human populations [54,55••]. B-lymphocyte gene expression data, derived from EBV-transformed lymphoblastoid cell lines or LCLs, for 445 of the same 1KGP individuals, representing one African population and four European populations, were taken from the Genetic European Variation in Health and Disease (GEUVEDIS) RNA-seq project [56]. Merging data from both projects allowed us to directly compare TE insertion site genotypes to gene expression levels from the same individuals. Furthermore, comparison of results for African and European populations allowed us to uncover population-specific regulatory effects of human TE polymorphisms.

Regression of gene expression against TE insertion site genotypes revealed hundreds of eQTL associations, and TE-eQTL were found both within and between the African and European populations. A number of TE polymorphisms were shown to be associated with expression differences between population groups. One advantage of using TE insertion site genotypes for eQTL analysis is that the relatively low number of common TE genotypes across the genome (~16 000) allows for both *cis* and *trans* eQTL analysis. This is because the number of possible eQTL associations is the product of the number of genes and the number of variants being compared; accordingly, the analysis of millions of SNPs times thousands of genes presents a combinatorically daunting bioinformatics analysis challenge. For this reason, most SNP eQTL studies focus exclusively on *cis* SNPs that are found within or in close proximity to individual genes. Since our study was not limited in this way, we were able to discover many *trans* associations of TE polymorphisms with human gene regulation. In fact, we were surprised to find that *trans* regulatory effects for TE polymorphisms were even more common than *cis* effects.

For one particular example, the B cell specific transcription factor PAX5, we uncovered a potential mechanism

Figure 3



The impact of TE polymorphisms on gene regulatory networks. The eQTL approach is used to discover associations between TE insertion variants and tissue-specific gene expression levels (i.e. TE-eQTLs). A TE insertion variant found in *cis* to a transcription factor (TF) can lead to coordinated changes across a gene regulatory network via transitive effects on downstream targets of the TF. An example is shown, similar to what has been observed for the TF gene *PAX5*, where TE associated increase in the expression of a TF leads in turn to increased expression of the TF target genes. This will reveal itself as multiple *trans* TE-eQTL associations for the same TE insertion variant.

that could explain the numerous *trans* TE-eQTL that we observed (Figure 3). This example also underscores how individual TE loci can participate in the rewiring of entire regulatory networks. The *PAX5* gene has a *cis* Alu eQTL that is associated with increased expression in B lymphocytes. This same Alu insertion is associated with increased expression of numerous PAX5 target genes, presumably by virtue of a transitive effect whereby increased PAX5 expression in turn increases the expression of downstream targets in its regulatory network.

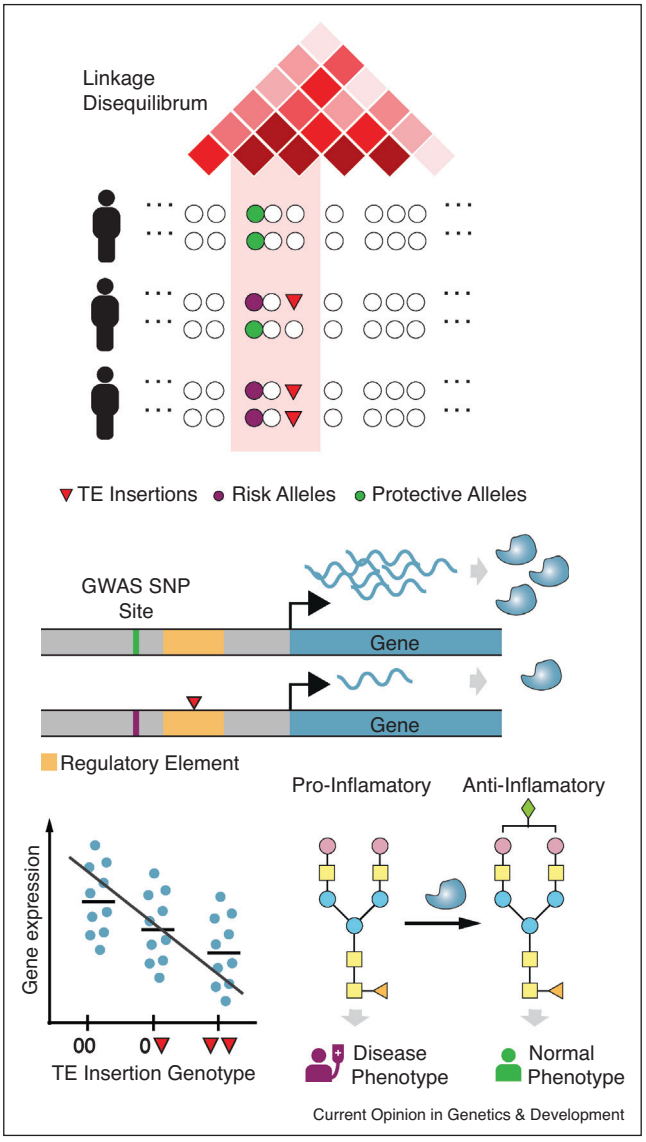
To our knowledge, this is the first and only study of its kind in humans. However, analogous genome-scale approaches have been used to discover TE associations with gene expression in the model organisms Arabidopsis [57] and maize [58]. It is important to point out that the eQTL results summarized here are very much cell-type dependent. Expansion of this kind of eQTL analysis to multiple cell and tissue types is expected to reveal distinct TE-gene regulatory effects. The recently completed Genotype-Tissue Expression (GTEx; https://www.gtexportal.org/) project provides eQTL data for more than 50 cell/tissue types, providing a tremendous opportunity for further work of this kind.

TE polymorphisms and complex common disease

Two studies published in 2017 have taken a similar population-level view of the phenotypic effects of human TE polymorphisms [59••,60••]. For each of these studies, associations between TE insertion site genotypes and complex common diseases were explored. Both studies relied on the analysis of linkage disequilibrium (LD) patterns to discover TE polymorphisms linked to SNPs that were previously associated with health or disease related phenotypes via genome-wide association studies (GWAS). An implicit rationale for genome-scale surveys of this kind is the notion that TE insertions are expected to be more disruptive than SNP variations given the larger scale genomic changes that they entail. Interestingly, both studies report that TE polymorphisms are enriched at GWAS loci, highlighting their potential impact. The first study of this kind, from the group of Kathleen Burns, found 44 Alu insertions in tight LD with previously discovered GWAS trait associated SNPs [59••]. The authors pointed out that this represents a >20-fold increase over the number of polymorphic Alu insertions that were previously known to be associated with human phenotypes, thereby underscoring the power of population genomic approaches for studies on the phenotypic impact of TE polymorphisms. Furthermore, the implicated Alu polymorphisms were found to be associated with a very broad range of health and disease related phenotypes.

Our own study on the impact of TE polymorphisms on complex common disease was designed to explore the connection between TE-mediated genome regulation and disease related phenotypic effects [60••]. To achieve this aim, we used a progressive set of genome-wide bioinformatics screens that searched for polymorphic TE insertions that are: (1) found in LD with known GWAS SNPs, (2) located within tissue-specific enhancers, and (3) associated with tissue-specific gene expression levels. We further narrowed our search for candidate

Figure 4



TE insertion variants impact on human disease via gene regulatory changes. TE insertion variants are found in tight linkage disequilibrium (LD) with previously characterized genome-wide association study (GWAS) SNP risk alleles. The linked TE insertion variant is associated with reduced gene expression, which is in turn associated with elevated disease risk. The scheme shown here corresponds to a TE insertion variant associated with reduced expression of the *B4GALT1* gene, which leads to increased inflammation and related disease pathology.

TE polymorphisms to those associated with genes with blood or immune related functions, consistent with the fact that the gene expression data we analyzed is from B lymphocytes. This progressive and stringent genomic screen uncovered six TE polymorphisms that are likely to be associated with disease phenotypes by virtue of their gene regulatory effects. These included both Alu elements, as previously reported, as well as SVA elements. For example, we discovered an SVA insertion in the cell-type specific enhancer of the *B4GALT1* gene (Figure 4). B4GALT1 acts to convert the Immunoglobulin G (IgG) antibody from a pro-inflammatory to an anti-inflammatory form. The SVA insertion is associated with both down-regulation of the *B4GALT1* gene, thereby potentially leading to increased inflammation, and is linked to a genomic region implicated by GWAS in both inflammatory conditions (Crohn's disease) and autoimmune disease (systemic lupus erythematosus).

Conclusions

One important caveat regarding the surveys of the effects of TE polymorphisms on human gene regulation and disease reviewed here relates to the fact that they rely on association studies. While this class of approaches has great potential to reveal connections between TE generated variation and health related phenotypes, association-based methods do not necessarily uncover causal variants (i.e., correlation $\neq$ causation). In this sense, the TE-phenotype associations uncovered by these studies are perhaps best considered as hypotheses, which will need to be further interrogated by experimental studies in order to provide deeper insight into causality and mechanism. Accordingly, one expectation is that these kinds of large-scale association studies can substantially narrow the experimental search space, with respect to possible TE-phenotype interactions, and thereby serve as a valuable point of departure for subsequent work.

The population genomics view of TEs exemplified by the recent studies reviewed here has the potential to expand our understanding of the phenotypic impact of human TEs. While ongoing human TE activity has widely been considered to be deleterious, the presence of TE insertion variants that segregate as common polymorphisms among human populations indicates that many novel TE insertions must have escaped the action of purifying selection. Accordingly, polymorphic human TE insertion variants comprise an important source of naturally occurring genetic variation with subtle effects on genome regulation and human health. Functionally relevant TE polymorphisms of this kind are likely to provide crucial source material for ongoing human evolution.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**:860-921.
2. de Koning APJ, Gu WJ, Castoe TA, Batzer MA, Pollock DD: **Repetitive elements may comprise over two-thirds of the human genome**. *PLoS Genet* 2011:7.
3. Doolittle WF, Sapienza C: **Selfish genes, the phenotype paradigm and genome evolution**. *Nature* 1980, **284**:601-603.
4. Orgel LE, Crick FH: **Selfish DNA: the ultimate parasite**. *Nature* 1980, **284**:604-607.
5. Gould SJ, Vrba ES: **Exaptation — a missing term in the science of form**. *Paleobiology* 1982, **8**:4-15.
6. Miller WJ, Hagemann S, Reiter E, Pinsker W: **P-element homologous sequences are tandemly repeated in the genome of *Drosophila guanche***. *Proc Natl Acad Sci U S A* 1992, **89**:4018-4022.
7. Chuong EB, Elde NC, Feschotte C: **Regulatory activities of transposable elements: from conflicts to benefits**. *Nat Rev Genet* 2017, **18**:71-86.
8. Feschotte C: **Transposable elements and the evolution of regulatory networks**. *Nat Rev Genet* 2008, **9**:397-405.
9. Rebollo R, Romanish MT, Mager DL: **Transposable elements: an abundant and natural source of regulatory sequences for host genes**. *Annu Rev Genet* 2012, **46**:21-42.
10. Conley AB, Piriyapongsa J, Jordan IK: **Retroviral promoters in the human genome**. *Bioinformatics* 2008, **24**:1563-1567.
11. Jordan IK, Rogozin IB, Glazko GV, Koonin EV: **Origin of a substantial fraction of human regulatory sequences from transposable elements**. *Trends Genet* 2003, **19**:68-72.
12. Marino-Ramirez L, Lewis KC, Landsman D, Jordan IK: **Transposable elements donate lineage-specific regulatory sequences to host genomes**. *Cytogenet Genome Res* 2005, **110**:333-341.
13. Bejerano G, Lowe CB, Ahituv N, King B, Siepel A, Salama SR, Rubin EM, Kent WJ, Haussler D: **A distal enhancer and an ultraconserved exon are derived from a novel retroposon**. *Nature* 2006, **441**:87-90.
14. • Chuong EB, Elde NC, Feschotte C: **Regulatory evolution of innate immunity through co-option of endogenous retroviruses**. *Science* 2016, **351**:1083-1087.
A detailed account of endogenous retrovirus (ERV) sequences that have been exapted to provide interferon inducible enhancers to a diverse group of mammalian genomes. This study is distinguished by its use of the CRISPR-Cas9 system to experimentally confirm the regulatory role of candidate ERV enhancers.
15. Chuong EB, Rumi MA, Soares MJ, Baker JC: **Endogenous retroviruses function as species-specific enhancer elements in the placenta**. *Nat Genet* 2013, **45**:325-329.
16. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, Ng HH, Bourque G: **Transposable elements have rewired the core regulatory network of human embryonic stem cells**. *Nat Genet* 2010, **42**:631-634.
17. Notwell JH, Chung T, Heavner W, Bejerano G: **A family of transposable elements co-opted into developmental enhancers in the mouse neocortex**. *Nat Commun* 2015, **6**:6644.
18. Conley AB, Jordan IK: **Cell type-specific termination of transcription by transposable element sequences**. *Mob DNA* 2012, **3**:15.
19. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo XY, Ramsay L, Bourque G, Yandell M, Feschotte C: **Transposable elements are**

major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet* 2013:9.

20. Piriyapongsa J, Marino-Ramirez L, Jordan IK: **Origin and evolution of human microRNAs from transposable elements**. *Genetics* 2007, **176**:1323-1337.

21. Weber MJ: **Mammalian small nucleolar RNAs are mobile genetic elements**. *PLoS Genet* 2006, **2**:1984-1997.

22. Jacques PE, Jeyakani J, Bourque G: **The majority of primate-specific regulatory sequences are derived from transposable elements**. *PLoS Genet* 2013, **9**:e1003504.

23. Pavlicek A, Jabbari K, Paces J, Paces V, Hejnar JV, Bernardi G: **Similar integration but different stability of Alus and LINEs in the human genome**. *Gene* 2001, **276**:39-45.

24. Schmidt D, Schwalie PC, Wilson MD, Ballester B, Goncalves A, Kutter C, Brown GD, Marshall A, Flicek P, Odom DT: **Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages**. *Cell* 2012, **148**:335-348.

25. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T: **Widespread contribution of transposable elements to the innovation of gene regulatory networks**. *Genome Res* 2014, **24**:1963-1976.

26. Wang J, Vicente-Garcia C, Seruggia D, Molto E, Fernandez-Minan A, Neto A, Lee E, Gomez-Skarmeta JL, Montoliu L, Lunyak VV *et al.*: **MIR retrotransposon sequences provide insulators to the human genome**. *Proc Natl Acad Sci U S A* 2015, **112**:E4428-E4437.

27. Batzer MA, Deininger PL: **A human-specific subfamily of Alu sequences**. *Genomics* 1991, **9**:481-487.

28. Batzer MA, Gudi VA, Mena JC, Foltz DW, Herrera RJ, Deininger PL: **Amplification dynamics of human-specific (HS) Alu family members**. *Nucleic Acids Res* 1991, **19**:3619-3623.

29. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, Kazazian HH Jr: **Hot L1s account for the bulk of retrotransposition in the human population**. *Proc Natl Acad Sci U S A* 2003, **100**:5280-5285.

30. Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE: **Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man**. *Nature* 1988, **332**:164-166.

31. Ostertag EM, Goodier JL, Zhang Y, Kazazian HH Jr: **SVA elements are nonautonomous retrotransposons that cause disease in humans**. *Am J Hum Genet* 2003, **73**:1444-1451.

32. Wang H, Xing J, Grover D, Hedges DJ, Han K, Walker JA, Batzer MA: **SVA elements: a hominid-specific retroposon family**. *J Mol Biol* 2005, **354**:994-1007.

33. Wildschutte JH, Williams ZH, Montesion M, Subramanian RP, Kidd JM, Coffin JM: **Discovery of unfixed endogenous retrovirus insertions in diverse human populations**. *Proc Natl Acad Sci U S A* 2016, **113**:E2326-E2334.

34. • Rishishwar L, Wang L, Clayton EA, Marino-Ramirez L, McDonald JF, Jordan IK: **Population and clinical genetics of human transposable elements in the (post) genomic era**. *Mob Genet Elements* 2017, **7**:1-20.
A recent review that provides a comprehensive overview of the prospects for population genomic studies of human TEs. Both evolutionary and clinical aspects of human TE biology are explored.

35. • Ewing AD: **Transposable element detection from whole genome sequence data**. *Mob DNA* 2015, **6**:24.
A review paper that covers the algorithmic approaches used to discover novel TE insertion variants via the analysis of whole genome sequence data.

36. •• Rishishwar L, Marino-Ramirez L, Jordan IK: **Benchmarking computational tools for polymorphic transposable element detection**. *Brief Bioinform* 2016.
A comprehensive benchmarking and validation study of 21 programs designed to discover and characterize TE insertion variants from whole genome sequence data. The focus of the study is on the discovery of TE insertion variants in human genome sequences.

37. Jiang C, Chen C, Huang Z, Liu R, Verdier J: **ITIS, a bioinformatics tool for accurate identification of transposon insertion sites using next-generation sequencing data**. *BMC Bioinform* 2015, **16**:72.

38. • Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, Mills RE, Genomes Project C, Devine SE: **The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology**. *Genome Res* 2017.
The bioinformatics tool MELT was developed by members of the 1000 Genomes Project structural variation group for the sequence based discovery and characterization of polymorphic human TE insertions. The superior performance of MELT has been independently validated [36].

39. Thung DT, de Ligt J, Vissers LE, Steehouwer M, Kroon M, de Vries P, Slagboom EP, Ye K, Veltman JA, Hehir-Kwa JY: **Mobster: accurate detection of mobile element insertions in next generation sequencing data**. *Genome Biol* 2014, **15**:488.

40. Keane TM, Wong K, Adams DJ: **RetroSeq: transposable element discovery from next-generation sequencing data**. *Bioinformatics* 2013, **29**:389-390.

41. Wu J, Lee WP, Ward A, Walker JA, Konkel MK, Batzer MA, Marth GT: **Tangram: a comprehensive toolbox for mobile element insertion detection**. *BMC Genomics* 2014, **15**:795.

42. Zhuang J, Wang J, Theurkauf W, Weng Z: **TEMP: a computational method for analyzing transposable element polymorphism in populations**. *Nucleic Acids Res* 2014, **42**:6826-6838.

43. Fiston-Lavier AS, Barron MG, Petrov DA, Gonzalez J: **T-lex2: genotyping, frequency estimation and re-annotation of transposable elements using single or pooled next-generation sequencing data**. *Nucleic Acids Res* 2015, **43**:e22.

44. Santander CG, Gambron P, Marchi E, Karamitros T, Katzourakis A, Magiorkinis G: **STEAK: a specific tool for transposable elements and retrovirus detection in high-throughput sequencing data**. *Virus Evol* 2017, **3**:vex023.

45. Disdero E, Filee J: **LoRTE: detecting transposon-induced genomic variants using low coverage PacBio long read sequences**. *Mob DNA* 2017, **8**:5.

46. Witherspoon DJ, Xing J, Zhang Y, Watkins WS, Batzer MA, Jorde LB: **Mobile element scanning (ME-Scan) by targeted high-throughput sequencing**. *BMC Genomics* 2010, **11**:410.

47. Ewing AD, Kazazian HH Jr: **High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes**. *Genome Res* 2010, **20**:1262-1270.

48. Sanchez-Luque FJ, Richardson SR, Faulkner GJ: **Retrotransposon capture sequencing (RC-Seq): a targeted, high-throughput approach to resolve somatic L1 retrotransposition in humans**. *Methods Mol Biol* 2016, **1400**:47-77.

49. Iskow RC, McCabe MT, Mills RE, Torene S, Pittard WS, Neuwald AF, Van Meir EG, Vertino PM, Devine SE: **Natural mutagenesis of human genomes by endogenous retrotransposons**. *Cell* 2010, **141**:1253-1261.

50. Tang Z, Steranka JP, Ma S, Grivainis M, Rodic N, Huang CR, Shih IM, Wang TL, Boeke JD, Fenyo D *et al.*: **Human transposon insertion profiling: analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer**. *Proc Natl Acad Sci U S A* 2017, **114**:E733-E740.

51. • Burns KH: **Transposable elements in cancer**. *Nat Rev Cancer* 2017, **17**:415-424.
A comprehensive and in depth review that covers the very latest developments on the role of TE activity in cancer, which is one of the most promising areas of human TE research.

52. •• Wang L, Rishishwar L, Marino-Ramirez L, Jordan IK: **Human population-specific gene expression and transcriptional network modification with polymorphic transposable elements**. *Nucleic Acids Res* 2017, **45**:2318-2328.
The first study of its kind wherein the regulatory potential of polymorphic human TEs was explored using the expression quantitative trait loci (eQTL) analytical paradigm. Numerous associations between TE polymorphisms and gene expression levels were uncovered including

population-specific gene regulatory effects as well as coordinated changes to gene regulatory networks.

53. Gibson G, Powell JE, Marigorta UM: **Expression quantitative trait locus analysis for translational medicine**. *Genome Med* 2015, **7**:60.

54. Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA *et al.*: **A global reference for human genetic variation**. *Nature* 2015, **526**:68-74.

55. •• Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, Zhang Y, Ye K, Jun G, Fritz MHY *et al.*: **An integrated map of structural variation in 2,504 human genomes**. *Nature* 2015, **526**:75-81.

A comprehensive catalog of human genome structural variants produced by the structural variation working group of the 1000 Genomes Project consortium. This catalog of variants includes a genome-wide collection of polymorphic TE insertions for Alu, L1, and SVA families and serves as an invaluable resource for population genomic studies of human TEs.

56. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, Gonzalez-Porta M, Kurbatova N, Griebel T, Ferreira PG *et al.*: **Transcriptome and genome sequencing uncovers functional variation in humans**. *Nature* 2013, **501**: 506-511.

57. Stuart T, Eichten SR, Cahn J, Karpievitch YV, Borevitz JO, Lister R: **Population scale mapping of transposable element diversity reveals links to gene regulation and epigenomic variation**. *Elife* 2016:5.

58. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, Springer NM: **Transposable elements contribute to activation of maize genes in response to abiotic stress**. *PLoS Genet* 2015, **11**:e1004915.

59. •• Payer LM, Steranka JP, Yang WR, Kryatova M, Medabalimi S, Ardeljan D, Liu C, Boeke JD, Avramopoulos D, Burns KH: **Structural variants caused by Alu insertions are associated with risks for many human diseases**. *Proc Natl Acad Sci U S A* 2017, **114**:E3984-E3992.

The first study linking TE insertion variants with disease risk alleles from genome-wide association studies (GWAS). The findings represent a >20-fold increase in the number of polymorphic Alu insertions associated with human traits, underscoring the power of the population genomic approach for the study of TE phenotypic impacts.

60. •• Wang L, Norris ET, Jordan IK: **Human retrotransposon insertion polymorphisms are associated with health and disease via gene regulatory phenotypes**. *Front Microbiol* 2017, **8**:1418.

Another recent report linking polymorphic human TEs with disease phenotype loci. This study is distinguished by the revealed connections between the regulatory efforts of TE polymorphisms and the molecular mechanisms that underlie particular disease phenotypes.