



Recent Bioinformatic Progress to Identify Epigenetic Changes Associated to Transposable Elements

Emmanuelle Lerat*

Univ Lyon, Univ Lyon 1, CNRS, VetAgro Sup, UMR5558, Laboratoire de Biométrie et Biologie Evolutive, Villeurbanne, France

OPEN ACCESS

Edited by:

Jan Padeken,
Friedrich Miescher Institute for
Biomedical Research (FMI),
Switzerland

Reviewed by:

Yuriy L Orlov,
I.M.Sechenov First Moscow State
Medical University, Russia
Ruslan Kalendar,
University of Helsinki, Finland
Mira Han,
University of Nevada, United States
Hidenori Nishihara,
Tokyo Institute of Technology, Japan

*Correspondence:

Emmanuelle Lerat
emmanuelle.lerat@univ-lyon1.fr

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 07 March 2022

Accepted: 25 April 2022

Published: 13 May 2022

Citation:

Lerat E (2022) Recent Bioinformatic
Progress to Identify Epigenetic
Changes Associated to
Transposable Elements.
Front. Genet. 13:891194.
doi: 10.3389/fgene.2022.891194

Transposable elements (TEs) are recognized for their great impact on the functioning and evolution of their host genomes. They are associated to various deleterious effects, which has led to the evolution of regulatory epigenetic mechanisms to control their activity. Despite these negative effects, TEs are also important actors in the evolution of genomes by promoting genetic diversity and new regulatory elements. Consequently, it is important to study the epigenetic modifications associated to TEs especially at a locus-specific level to determine their individual influence on gene functioning. To this aim, this short review presents the current bioinformatic tools to achieve this task.

Keywords: epigenetics, transposable elements, bioinformatics, epigenomics, NGS data

INTRODUCTION

For many years, the presence of transposable elements (TEs) has been acknowledged in the genomes of all living organisms, not only because of their large prevalence in some of them but also because they have a profound impact on their functioning and evolution (Feschotte, 2008; Biemont, 2010; Almojil et al., 2021). TEs correspond to sequences with a large variety of structural features allowing their grouping in different classes and families (Wicker et al., 2007). They share the common characteristics to be able to move around inside their host genome using their own enzymatic machinery and to duplicate themselves, leading them to be present in a genome in multiple and often very similar copies. The presence and expression of TEs have been associated to various diseases like for example the Haemophilia A, the Alstrom syndrome, the Dent's disease or different cancers [see for a review (Payer and Burns, 2019)]. Their mutational effects are generally summarized into three categories: the disruption or modification of a host protein, the deregulation of gene transcription, and chromosomal rearrangements by ectopic recombination.

The deleterious effects of TEs on their host genomes have led to the evolution of regulatory mechanisms to control their activity. In particular, the different epigenetic mechanisms contribute to their silencing preventing their mobilization (Huda and Jordan, 2009; Lisch, 2009). In mammals and in plants, TEs are mainly silenced by DNA methylation (Deniz et al., 2019). TEs are also targeted by repressive histone modifications leading to heterochromatin formation (Slotkin and Martienssen, 2007). Finally, RNA interference through the action of small RNAs (sRNAs) allows post-transcriptional silencing leading for example to the control of TE activity in the germline of *Drosophila melanogaster* (Sato and Siomi, 2020). sRNAs also allow the targeting of TEs by DNA methylation of homologous sequences and the setting up of histone modification to silence them, which demonstrate the clear interlacing of all epigenetic mechanisms (Hall et al., 2002; Lewsey et al., 2016).

Despite their numerous deleterious effects, TEs are also now largely recognized as important actors in the evolution of genomes by promoting a fair amount of genetic diversity (Schrader and

Schmitz, 2019). In addition, TEs can be implicated in the gene regulation by providing regulatory elements (Chuong et al., 2016). The genetic diversity induced by TEs is particularly valuable for organisms in order for them to adapt to environmental changes. This new response may be very rapid especially when it involves epigenetic changes associated to TEs that may induce an impact on the host gene functioning. For example, a specific TE insertion in *D. melanogaster* was shown to display inactive histone modifications in normal condition but also active modifications under oxidative stress conditions, having an impact on the observed expression pattern of nearby genes (Guio et al., 2018). Similarly in mouse, the methylation level of a TE inserted near the agouti gene leads to variation in its transcription level inducing variation in the fur color (Morgan et al., 1999). In some cases, the TE associated epigenetic modifications may act as facilitator for adaptation and may be conserved through time corresponding to the taming of TEs (Capy, 2021). In addition, some TEs may be organized in genomes as tandem arrays which has been proposed to participate to chromosome rearrangement and structuring in some plants (Kalendar et al., 2020).

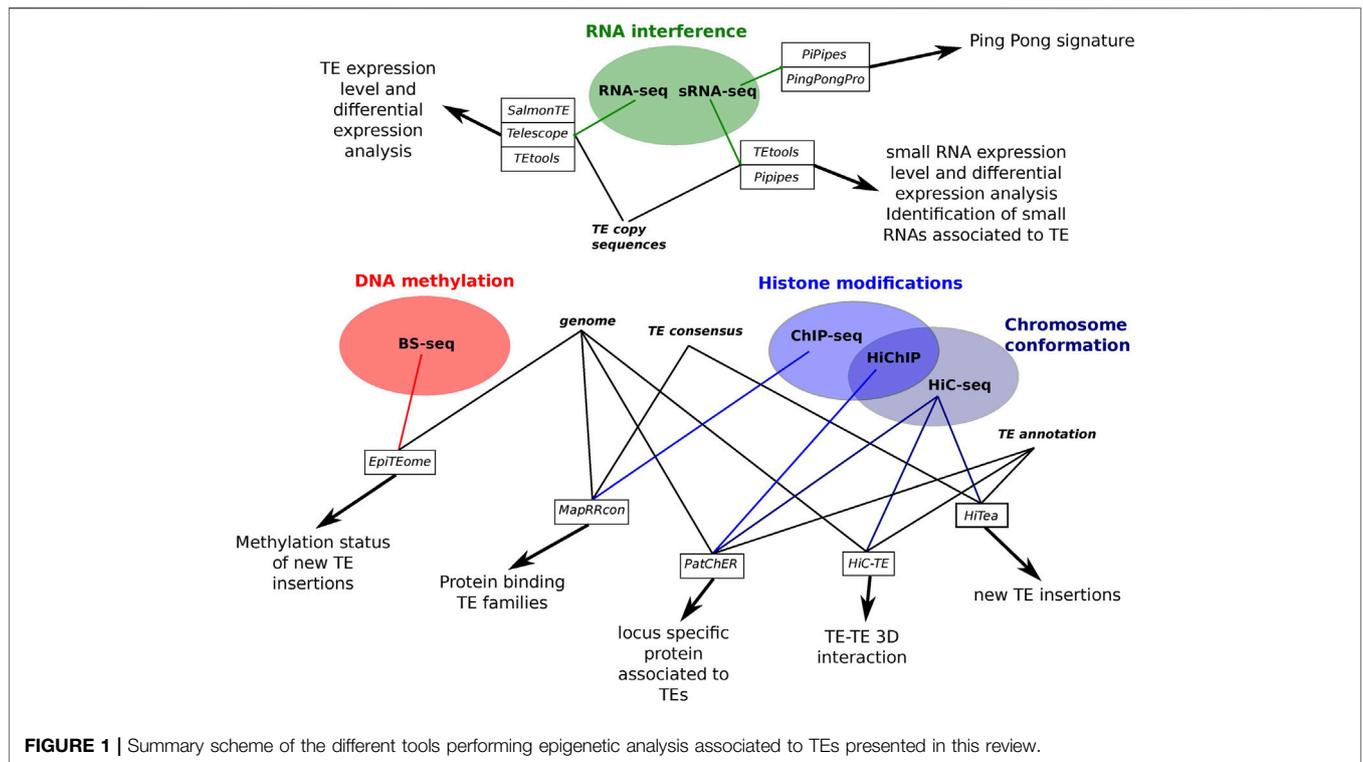
Consequently, it is particularly important to study the epigenetic modifications associated to TEs as already proposed elsewhere (Lerat et al., 2019). More particularly, it is even more important to have information at the copy specific level of TEs, i.e. according to the insertion locus of each element. Indeed, it can be expected that not all copies of a given TE in a genome may have the same impact on genes. Although fixed TE insertions are ancient and likely to be not deleterious in normal condition, they may influence their neighboring genes when a change in the environment occurs (Grégoire et al., 2016). In genomes, there is also a quite large proportion of TE insertions that are polymorphic when comparing different individuals of a same species (Gardner et al., 2017). These insertions can represent a potential source of genetic variability between individuals (Goubert et al., 2020). A very large number of bioinformatic tools have been designed to identify polymorphic insertions using genomic re-sequencing data by comparison to a reference genome. Some of these tools have been developed to identify insertions of interest in specific diseases like cancer, in which TEs have been described for a long time to be reactivated potentially leading to new insertions (Anwar et al., 2017; Jang et al., 2019). Among the twenty existing programs (for a comprehensive list, see https://tehub.org/en/resources/repeat_tools), we can cite for the most recent *TIP_finder* (Orozco-Arias et al., 2020) developed to detect TE insertions in very large genomes and tested on human cancer data, *cloudMELT* (Chuang et al., 2021) which was used to discover thousand of polymorphic TE insertions in numerous human population data, and *xTea* (Chu et al., 2021) which identifies new TE insertions from multiple sequencing technologies including long-reads and 10X linked reads. Although some attempts to perform standardized benchmarks of these methods exist (Rishishwar et al., 2017; Vendrell-Mir et al., 2019), there is nevertheless still room to determine what is the best tool to use considering the investigated biological data.

The question of the locus specificity in the epigenetic analysis of TEs is very difficult to tackle on the bioinformatic point of view

given the sequencing data obtained by the different techniques used to generate them. Contrary to whole genome re-sequencing, it is more difficult to obtain reads long enough to allow unambiguous mapping of reads on the reference genome, especially in the case of histone modification analysis. This comes from the fact that recent TE copies are very similar in sequences and that short reads may not be specific enough to distinguish between very similar TE copies. Thus, the question of the sequenced read mappability is central in the development of bioinformatics tools aimed to study the epigenetic modifications associated to TEs. Indeed, short reads corresponding to TEs are often difficult to map unambiguously because of the repeated nature of TEs and the high intra-family sequence similarity (Li and Freudenberg, 2014; Sexton and Han, 2019; Teissandier et al., 2019). In this mini-review, I will present the most recent bioinformatic tools that have been proposed to decipher epigenetic modifications specifically associated to TEs and trying to take into account the copy specific level of these insertions (Figure 1).

TRANSCRIPTION AND RNA INTERFERENCE ANALYSES

A first indication concerning the activity of TEs and thus the potentiality for them to either promote new insertions or influence neighboring gene expression, is the presence of TE transcripts in the cells. Since few years, specific bioinformatic tools have been developed to measure the transcription level of TEs (Table 1), some also allowing their differential expression analyses among several samples (different tissues or different populations). In their very comprehensive review concerning this topic, Lanciano and Cristofari (Lanciano and Cristofari, 2020) perfectly describe the complexity of the task. Indeed, since TEs are repeats that can be polymorphic, highly similar and with potential overlapping with genes, it drives to the existence of complex transcripts which may be difficult to identify and analyze. The authors identify three challenges to be overcome to deal with TE expression analysis which are the mappability of reads corresponding to TEs which are often of non-unique location, the sequence divergence among copies from a same family and their chimeric or co-transcription with genes blurring the detection of functional TE transcripts (Lanciano and Cristofari, 2020). In an attempt to determine how existing tools can accurately assess the expression of TEs at the copy level, a simulation study was recently performed (Schwarz et al., 2021). In this study, RNA-seq reads from mouse, human and turquoise killifish were generated. The results of this benchmark indicate that three of the tested tools performed very well to detect differentially expressed TEs. The first tool giving the best performance is *SalmonTE* (Jeong et al., 2018). It performs a read mapping against TE sequences and reassigns multi-mapping reads using the expectation-maximization (EM) algorithm before determining the read count of each sequence. The second best tool, *Telescope* (Bendall et al., 2019), was designed to determine TE expression in a copy-specific manner. As the previous tool, it uses EM algorithm to assign ambiguous mapped reads. Finally,



the third position is occupied by *TEtools* (Lerat et al., 2017) which performs read alignment against TE sequences allowing transcription level determination at both family and copy level. Contrary to the other two tools, it randomly assesses multi-mapped reads to a given sequence. According to the

benchmark, the three tools have the potential to precisely assess TE expression at a specific genomic locus with some minor modifications (Schwarz et al., 2021). Other tools continue to be developed with the goal to determine the genomic locus expression like for example *REdiscoverTE*

TABLE 1 | Recent tools for transcription and small RNA analysis.

Tool	Function	Input data	Algorithm for read mapping and selection	References
SalmonTE	Quantification of TE transcript abundance	RNA-seq and TE sequences	EM-algorithm	Jeong et al. (2018)
Telescope	Copy-specific TE expression	Aligned read from RNAseq	EM-algorithm	Bendall et al. (2019)
TEtools	Determine differentially expressed TEs and smallRNAs	SmallRNA-seq, RNA-seq, TE copies	Bowtie2 (mRNA) and Bowtie1 (smallRNA) (random assignment of best match)	Lerat et al. (2017)
REdiscoverTE	Copy-specific TE expression	RNA-seq, TE copy sequences, intron sequences	Salmon (EM-algorithm)	Kong et al. (2019)
ExplorATE	Copy-specific TE expression	Genome sequence, gene annotation, TE annotation RNA-seq	Salmon; target transcript assignment based on the percentage of identity for a class/family of TEs	https://github.com/FemeniasM/ExplorATE_shell_script
TEffectR	Influence of TEs on neighboring gene expression	Gene annotation, TE annotation, aligned RNA-seq reads on genome	Reads with 100% overlap with given TE regions are considered	Karakulah et al. (2019)
LIONS	Identification and analysis of chimeric TE-gene transcripts	RNA-seq, References genome, gene and TE annotation	Tophat2 (random assignment of best match)	Babaian et al. (2019)
PIPipes	Analyze piRNAs and TE-derived RNAs	SmallRNA-seq, RNA-seq, TE copies	STAR (mRNA) and Bowtie1 (smallRNA) (EM-algorithm)	Han et al. (2015)
PingPongPro	Detection of ping-pong cycle activity in piRNA-Seq data	SmallRNA-seq	Weighted read count	Uhrig and Klein, (2019)

(Kong et al., 2019) or *ExplorATE* (https://github.com/FemeniasM/ExplorATE_shell_script). The first tool takes advantage of the weight-mapping method *Salmon* (Patro et al., 2017) using *RepeatMasker* outputs and performs expression analysis down to the TE sub-family level. The quantification at individual locus level is accomplished with regard to the host gene positions i.e. whether the sub-family copies are intergenic, intronic or exonic. The second tool is an R package aims at identifying active TEs among RNA-seq data. It deals with both co-transcription and multi-mapping issues, and performs differential expression analyses. It may also be interesting to directly assess the effect of TEs on gene transcription. This is what is proposed using the R package *TEffectR* (Karakulah et al., 2019). This methods uses a linear regression model to determine the influence of TEs on neighboring gene expression using RNA-seq data, *RepeatMasker* output files containing TE location and Ensembl gene annotations. Similarly, the *LIONS* pipeline (Babaian et al., 2019) aims at identifying and analyzing the expression of chimeric TE-gene transcripts initiated by TEs.

Having the expression level of TEs is however not enough to evaluate whether a family is actually active and able to promote new insertions. To this aim, it is important to study the post-transcriptional regulation of TEs by analyzing the transcription of sRNAs, and more particularly of piwi-RNAs (piRNAs) which are specifically designed for the TE regulation. However, such a type of analyses comes with more technical difficulties. Reads corresponding to piRNAs are even shorter than those corresponding to mRNAs. Indeed, piRNAs have sizes comprised between 21 and 35 bp (Ozata et al., 2019), which usually requires the addition of a mononucleotide tail in order to obtain long enough fragments for high throughput sequencing. These tails need to be removed before the reads can be used, with for example a dedicated trimming tool like *UrQt* (Modolo and Lerat, 2015), designed at first for this task but also able to perform regular unsupervised trimming of any NGS data. To avoid as much as possible the presence of degraded mRNAs, it is usually necessary before sequencing to have a first step of filtering like the anion-exchange chromatography purification method (Grentzinger et al., 2014) which allows the enrichment of genuine piRNAs, resulting in an increase in their sequencing depth. Once the reads have been sequenced, the goal is to be able to correctly assign the piRNAs to their TE family as well as their expression level, to identify the TE families that are actively regulated. In that perspective, very few tools have been proposed to specifically handle this question (Table 1). The *TEtools* pipeline (Lerat et al., 2017) proposes such a possibility, using an alternative mapper than for mRNA to take into account the short size of the reads. One tool of the *piPipes* pipeline (Han et al., 2015) performs transcriptomic analyses of piRNAs based on the mapping of reads on consensus TE sequences from model species. In addition, it can report the “Ping-Pong” signature specific of the piRNAs, assigned to genomic annotations. This signature indicates the presence of the positive feedback loop (called the Ping-Pong cycle) allowing the reinforcement of piRNA production and thus of the TE-silencing effect. The detection of this particular signature is at the core of the *PingPongPro* tool (Uhrig and Klein, 2019) which uses

weighted read count within the region of a TE overlapping by 10 bp and corresponding to the empirical probability that the read is ping-pong-derived.

DNA METHYLATION AND TE INSERTIONS

DNA methylation has been described in numerous organisms as the main silencing mechanism of TEs at the transcriptional level on a long term (Deniz et al., 2019). In majority, the mechanism targets specifically 5-methylcytosine but other modifications exist like the 4-methylcytosine and the 6-methyladenine. Different techniques have been developed to decipher the methylation profile of a genome. The most widely used is the bisulfite sequencing (BS-seq) technique in which genomic DNA is sequenced after bisulfite conversion of unmethylated cytosines into uracils, methylated cytosines being protected from the conversion. The comparison with the sequence of the same genomic DNA without treatment allows to determine which cytosines in the genome are methylated in the studied condition. Currently, only one bioinformatic tool specifically address the question of the methylation of TE sequences. This tool, *EpiTEome* (Daron and Slotkin, 2017), aims at detecting both new TE insertion sites and their corresponding DNA methylation status. To achieve this goal, the first step is to use BS-seq data to identify new TE insertions when compared to a reference genome. Then the methylation status of these insertions is assessed as well as at the region surrounding them. The tool takes as input BS-seq reads that do not map on the reference genome using specific mappers designed to handle such data like *Bismark* (Krueger and Andrews, 2011). These reads are then split and each split-read is mapped on the reference genome to identify breakpoint locations on the read so that part of the read corresponds to a TE and the other part to a genomic sequence into which the new insertion occurred. The procedure allows the identification of all new TE insertion locations. The DNA methylation status is determined using the split-reads that were used to identify the insertion locus. The main drawback of the method is that it only considers insertions not present in the reference genome whereas they are not the only ones expected to be subject to methylation variation. This means that for these other insertions, it is still necessary to rely on classic approaches dealing by BS-seq data, knowing that they have limitations in their mapping abilities when it concerns TEs.

CHROMATIN STRUCTURE AND CHROMOSOME-CHROMOSOME INTERACTIONS

The DNA compaction inside the nucleus is associated to the various biochemical modifications directed on the amino-acid tail of the histone proteins. Some modifications are associated with the opening of the chromatin allowing gene expression whereas other modifications are on the contrary associated with closed chromatin (heterochromatin) which drives to transcription silencing (Lawrence et al., 2016). For example, the regulation

TABLE 2 | Recent tools for the identification of TEs in chromatin structure and chromosome-chromosome interactions.

Tool	Function	Input data	Multimapping handling	References
Crunch	Performs ChIP-seq analysis (mapping, peak calling)	ChIP-seq data and references genome	Multimapping reads are taken into account to avoid a loss of binding peaks in repeated regions; the weight of each of these reads are equally distributed to all mapping locations	Berger et al. (2019)
MapRRcon	Identify proteins binding to TE sequences	ChIP-seq data, references genome and TE consensus sequences	Unique and non-unique aligned reads are extracted and mapped to TE sequences; reads with partial alignment, >3 mismatches and any indels are excluded; Remaining reads are mapped against TE consensus	Sun et al. (2018)
PatChER	Use of chimeric HiC-seq fragments between unique and non-unique reads to identify proteins binding to TE sequences	HiChIP data, HiC-seq data, references genome	Performs random mapping of non-unique reads	Taylor et al. (2021)
HiC-TE	Identification of TEs implicated in 3D conformation	HiC-seq data, references genome	Read mapping performed using Bowtie2	Lexa et al. (2021)
HiTea	Identification of new TE insertions using discarded HiC-seq reads from classical approaches	HiC-seq data, TE consensus, TE annotations in references genome	Identification of close discordant read pairs with one mapping on a unique locus and the other on a TE sequence	Jain et al. (2021)

of retroelements has been associated to the presence of trimethylated histone H3 at lysine 9 (H3K9m3) (Fukuda and Shinkai, 2020). On the contrary, histone acetylation is associated to gene expression (Verdone et al., 2005). Different bioinformatic tools have been developed to analyze the reads produced by Chromatin Immuno-Precipitation sequencing (ChIP-seq) (see for a review Steinhäuser et al., 2016) but the majority of them discard multi-mapping reads preventing the identification of peaks associated to repeat regions. However, some recent tools do consider multi-mapping reads and thus may potentially give some information concerning repeat sequences. In that respect, among the most recent published tools, *Crunch* (Berger et al., 2019) considers multi-mapping reads to avoid a loss of binding peaks in repeated regions and equally distribute the weight of each of these reads to all mapping locations. Although this represents a good starting point, this is however not specific enough to allow the direct analysis of histone modifications associated to TEs.

There are very few tools that propose to assign histone modifications to TEs (Table 2). The first to have been developed estimates an average histone modification enrichment for a set of homologous repeat sequences, like a TE family, but does not give information about the variability among these sequences (Day et al., 2010). Moreover, among the multiple mapping reads on the genome, only those uniquely mapping to sequences belonging to a same repeat family are taken into account. This first tool is interesting but is also limited by the choice of organisms on which the analyses can be made. More recently, the pipeline *MapRRcon* was developed to allow the identification of proteins binding to TE sequences by mapping ChIP-seq reads to TE consensus sequences after whole-genome alignment (Sun et al., 2018). This approach was applied on human data to identify the interaction between transcription factors and a specific family of TE called L1, representing the youngest and most active TE family in human. Associated to transcriptomic analyses, it allowed the authors to suggest that some L1-binding

factors may play a role in the regulation of L1 activity during tumor development. Again in this approach, the read assignment remains global and is only able to identify reads associated to the whole TE family without providing information concerning the genomic location of the individual insertions. More globally, a major problem with ChIP-seq reads in the identification of associated TE sequences is their short sizes which prevents obtaining a sufficient number of reads unambiguously mapped to individual TE locations. Very recently, a solution has been proposed to help bypass this difficulty which takes advantage of random 3D interactions as implemented in the pipeline *PatChER* (Taylor et al., 2021). These random interactions correspond to relatively short distance (few tens of kb) 3D folding of the chromosomes and can be obtained from high throughput technologies like HiC-seq. HiC-seq is a method allowing to decipher 3D architecture of whole genomes by coupling proximity-based ligation with high-throughput sequencing (Lieberman-Aiden et al., 2009). In their approach, Taylor et al. use the random 3D interaction to guide the mapping of short reads obtained from HiChIP, a protein-centric chromatin conformation method, on TE individual insertions. The principle of this method is that since HiC-seq produces chimeric fragments between genomic regions that are close in 3D space, it should thus be possible to pair non-unique reads with unique reads from nearby genomic regions when focusing on random 3D interactions. By applying their method on mouse embryonic stem cell data, the authors were able to show that a particular family of retrotransposon displays a large variation in the enrichment of an active histone modifications according to the genomic location of the TE insertions (Taylor et al., 2021).

Chromosome interactions may also be interesting to study on the TE point of view. Since it has been shown that some TEs may be associated to particular 3D conformation (Cournac et al., 2016; Lu et al., 2021), they may be implicated in long distance gene regulation. Very recently, the pipeline *HiC-TE* (Lexa et al., 2021; <https://gitlab.fi.muni.cz/lexa/hic-te>) was proposed to study

interactions in nuclei of different repeat sequences, and in particular TEs, in long distance or interchromosomal interactions. The tool identifies and quantifies the interactions of repeats in the 3D organization of the genome using as an input HiC-seq data. It can work according two modes, reference-based and reference-free, to enhance the robustness of the results. Using HiC-seq data can also allow the identification of new TE insertions as proposed by the tool *HiTea* (Jain et al., 2021). This approach uses reads that are discarded in classical analyses of HiC data and map them on TE consensus to identify non reference TE insertions. In their application, the authors used split HiC read information and coverage to detect insertions of three major classes of TEs active in human. They were able to detect 1,085 new insertions using Hi-C data from the GM12878 cell line, which represent 62% of the new insertions identified by whole genome re-sequencing using long reads.

CONCLUSION

Advances in sequencing technologies have made it possible to develop new bioinformatic tools for more specific analysis of TEs. The most recent tools are beginning to be able to analyze in a finer

way, i.e. at the level of the individual copy, the epigenetic modifications associated with them. However, there is still a lot of progress to be made in this field, in particular with the consideration of structural variants between the samples studied independently of a reference genome. Cellular heterogeneity within a tissue must also be taken into account, particularly in the case of cancer studies. For this last possibility, a new methodological development was made in this direction that allows the locus-specific expression analysis of TEs in single cells (Berrens et al., 2021). New bioinformatic tools should thus continuously be developed to help analyze these new types of data.

AUTHOR CONTRIBUTIONS

EL conceived and wrote the entire manuscript.

FUNDING

EL has been supported by the CNRS and the LBBE.

REFERENCES

- Almojil, D., Bourgeois, Y., Falis, M., Hariyani, I., Wilcox, J., and Boissinot, S. (2021). The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes. *Genes* 12, 918. doi:10.3390/genes12060918
- Anwar, S., Wulaningsih, W., and Lehmann, U. (2017). Transposable Elements in Human Cancer: Causes and Consequences of Dereglulation. *Ijms* 18, 974. doi:10.3390/ijms18050974
- Babaian, A., Thompson, I. R., Lever, J., Gagnier, L., Karimi, M. M., and Mager, D. L. (2019). LIONS: Analysis Suite for Detecting and Quantifying Transposable Element Initiated Transcription from RNA-Seq. *Bioinforma. Oxf. Engl.* 35, 3839–3841. doi:10.1093/bioinformatics/btz130
- Bendall, M. L., de Mulder, M., Iñiguez, L. P., Lecanda-Sánchez, A., Pérez-Losada, M., Ostrowski, M. A., et al. (2019). Telescope: Characterization of the Retrotranscriptome by Accurate Estimation of Transposable Element Expression. *PLoS Comput. Biol.* 15, e1006453. doi:10.1371/journal.pcbi.1006453
- Berger, S., Pachkov, M., Arnold, P., Omidi, S., Kelley, N., Salatino, S., et al. (2019). Crunch: Integrated Processing and Modeling of ChIP-Seq Data in Terms of Regulatory Motifs. *Genome Res.* 29, 1164–1177. doi:10.1101/gr.239319.118
- Berrens, R. V., Yang, A., Laumer, C. E., Lun, A. T. L., Bieberich, F., Law, C.-T., et al. (2021). Locus-specific Expression of Transposable Elements in Single Cells with CELLO-Seq. *Nat. Biotechnol.* 40, 546–554. doi:10.1038/s41587-021-01093-1
- Biémont, C. (2010). A Brief History of the Status of Transposable Elements: From Junk DNA to Major Players in Evolution. *Genetics* 186, 1085–1093. doi:10.1534/genetics.110.124180
- Capy, P. (2021). Taming, Domestication and Exaptation: Trajectories of Transposable Elements in Genomes. *Cells* 10, 3590. doi:10.3390/cells10123590
- Chu, C., Borges-Monroy, R., Viswanadham, V. V., Lee, S., Li, H., Lee, E. A., et al. (2021). Comprehensive Identification of Transposable Element Insertions Using Multiple Sequencing Technologies. *Nat. Commun.* 12, 3836. doi:10.1038/s41467-021-24041-8
- Chuang, N. T., Gardner, E. J., Terry, D. M., Crabtree, J., Mahurkar, A. A., Rivell, G. L., et al. (2021). Mutagenesis of Human Genomes by Endogenous Mobile Elements on a Population Scale. *Genome Res.* 31, 2225–2235. doi:10.1101/gr.275323.121
- Chuong, E. B., Elde, N. C., and Feschotte, C. (2016). Regulatory Activities of Transposable Elements: from Conflicts to Benefits. *Nat. Rev. Genet.* 18, 71–86. doi:10.1038/nrg.2016.139
- Cournac, A., Koszul, R., and Mozziconacci, J. (2016). The 3D Folding of Metazoan Genomes Correlates with the Association of Similar Repetitive Elements. *Nucleic Acids Res.* 44, 245–255. doi:10.1093/nar/gkv1292
- Daron, J., and Slotkin, R. K. (2017). EpiTEome: Simultaneous Detection of Transposable Element Insertion Sites and Their DNA Methylation Levels. *Genome Biol.* 18, 91. doi:10.1186/s13059-017-1232-0
- Day, D. S., Luquette, L. J., Park, P. J., and Kharchenko, P. V. (2010). Estimating Enrichment of Repetitive Elements from High-Throughput Sequence Data. *Genome Biol.* 11, R69. doi:10.1186/gb-2010-11-6-r69
- Deniz, Ö., Frost, J. M., and Branco, M. R. (2019). Regulation of Transposable Elements by DNA Modifications. *Nat. Rev. Genet.* 20, 417–431. doi:10.1038/s41576-019-0106-6
- Feschotte, C. (2008). Transposable Elements and the Evolution of Regulatory Networks. *Nat. Rev. Genet.* 9, 397–405. doi:10.1038/nrg2337
- Fukuda, K., and Shinkai, Y. (2020). SETDB1-Mediated Silencing of Retroelements. *Viruses* 12, 596. doi:10.3390/v12060596
- Gardner, E. J., Lam, V. K., Harris, D. N., Chuang, N. T., Scott, E. C., Pittard, W. S., et al. (2017). The Mobile Element Locator Tool (MELT): Population-Scale Mobile Element Discovery and Biology. *Genome Res.* 27, 1916–1929. doi:10.1101/gr.218032.116
- Goubert, C., Zevallos, N. A., and Feschotte, C. (2020). Contribution of Unfixed Transposable Element Insertions to Human Regulatory Variation. *Phil. Trans. R. Soc. B* 375, 20190331. doi:10.1098/rstb.2019.0331
- Grégoire, L., Haudry, A., and Lerat, E. (2016). The Transposable Element Environment of Human Genes Is Associated with Histone and Expression Changes in Cancer. *BMC Genomics* 17, 588. doi:10.1186/s12864-016-2970-1
- Grentzinger, T., Armenise, C., Pelisson, A., Brun, C., Mugat, B., and Chambeyron, S. (2014). A User-Friendly Chromatographic Method to Purify Small Regulatory RNAs. *Methods* 67, 91–101. doi:10.1016/j.ymeth.2013.05.011
- Guio, L., Vieira, C., and González, J. (2018). Stress Affects the Epigenetic Marks Added by Natural Transposable Element Insertions in *Drosophila melanogaster*. *Sci. Rep.* 8, 12197. doi:10.1038/s41598-018-30491-w
- Hall, I. M., Shankaranarayana, G. D., Noma, K.-I., Ayoub, N., Cohen, A., and Grewal, S. I. S. (2002). Establishment and Maintenance of a Heterochromatin Domain. *Science* 297, 2232–2237. doi:10.1126/science.1076466
- Han, B. W., Wang, W., Zamore, P. D., and Weng, Z. (2015). piPipes: a Set of Pipelines for piRNA and Transposon Analysis via Small RNA-Seq, RNA-Seq, Degradome- and CAGE-Seq, ChIP-Seq and Genomic DNA

- Sequencing. *Bioinforma. Oxf. Engl.* 31, 593–595. doi:10.1093/bioinformatics/btu647
- Huda, A., and Jordan, I. K. (2009). Epigenetic Regulation of Mammalian Genomes by Transposable Elements. *Ann. N. Y. Acad. Sci.* 1178, 276–284. doi:10.1111/j.1749-6632.2009.05007.x
- Jain, D., Chu, C., Alver, B. H., Lee, S., Lee, E. A., and Park, P. J. (2021). HiTea: a Computational Pipeline to Identify Non-reference Transposable Element Insertions in Hi-C Data. *Bioinforma. Oxf. Engl.* 37, 1045–1051. doi:10.1093/bioinformatics/btaa923
- Jang, H. S., Shah, N. M., Du, A. Y., Dailey, Z. Z., Pehrsson, E. C., Godoy, P. M., et al. (2019). Transposable Elements Drive Widespread Expression of Oncogenes in Human Cancers. *Nat. Genet.* 51, 611–617. doi:10.1038/s41588-019-0373-3
- Jeong, H. H., Yalamanchili, H. K., Guo, C., Shulman, J. M., and Liu, Z. (2018). An Ultrafast and Scalable Quantification Pipeline for Transposable Elements from Next Generation Sequencing Data. *Pac Symp. Biocomput* 23, 168–179. doi:10.1142/9789813235533_0016
- Kalendar, R., Raskina, O., Belyayev, A., and Schulman, A. H. (2020). Long Tandem Arrays of Cassandra Retroelements and Their Role in Genome Dynamics in Plants. *Ijms* 21, 2931. doi:10.3390/ijms21082931
- Karakülah, G., Arslan, N., Yandım, C., and Suner, A. (2019). TEffectR: an R Package for Studying the Potential Effects of Transposable Elements on Gene Expression with Linear Regression Model. *PeerJ* 7, e8192. doi:10.7717/peerj.8192
- Kong, Y., Rose, C. M., Cass, A. A., Williams, A. G., Darwish, M., Lianoglou, S., et al. (2019). Transposable Element Expression in Tumors Is Associated with Immune Infiltration and Increased Antigenicity. *Nat. Commun.* 10, 5228. doi:10.1038/s41467-019-13035-2
- Krueger, F., and Andrews, S. R. (2011). Bismark: a Flexible Aligner and Methylation Caller for Bisulfite-Seq Applications. *Bioinforma. Oxf. Engl.* 27, 1571–1572. doi:10.1093/bioinformatics/btr167
- Lanciano, S., and Cristofari, G. (2020). Measuring and Interpreting Transposable Element Expression. *Nat. Rev. Genet.* 21, 721–736. doi:10.1038/s41576-020-0251-y
- Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. *Trends Genet.* 32, 42–56. doi:10.1016/j.tig.2015.10.007
- Lerat, E., Casacuberta, J., Chaparro, C., and Vieira, C. (2019). On the Importance to Acknowledge Transposable Elements in Epigenomic Analyses. *Genes* 10, 258. doi:10.3390/genes10040258
- Lerat, E., Fablet, M., Modolo, L., Lopez-Maestre, H., and Vieira, C. (2017). TEtools Facilitates Big Data Expression Analysis of Transposable Elements and Reveals an Antagonism between Their Activity and that of piRNA Genes. *Nucleic Acids Res.* 45, gkw953. doi:10.1093/nar/gkw953
- Lewsey, M. G., Hardcastle, T. J., Melnyk, C. W., Molnar, A., Valli, A., Urlich, M. A., et al. (2016). Mobile Small RNAs Regulate Genome-wide DNA Methylation. *Proc. Natl. Acad. Sci. U.S.A.* 113, E801–E810. doi:10.1073/pnas.1515072113
- Lexa, M., Cechova, M., Nguyen, S. H., Jedlicka, P., Tokan, V., Kubat, Z., et al. (2021). HiC-TE: a Computational Pipeline for Hi-C Data Analysis Shows a Possible Role of Repeat Family Interactions in the Genome 3D Organization. *BioRxiv* 1218, 473300. doi:10.1101/2021.12.18.473300
- Li, W., and Freudenberg, J. (2014). Mappability and Read Length. *Front. Genet.* 5, 381. doi:10.3389/fgene.2014.00381
- Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragozcy, T., Telling, A., et al. (2009). Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* 326, 289–293. doi:10.1126/science.1181369
- Lisch, D. (2009). Epigenetic Regulation of Transposable Elements in Plants. *Annu. Rev. Plant Biol.* 60, 43–66. doi:10.1146/annurev.arplant.59.032607.092744
- Lu, J. Y., Chang, L., Li, T., Wang, T., Yin, Y., Zhan, G., et al. (2021). Homotypic Clustering of L1 and B1/Alu Repeats Compartmentalizes the 3D Genome. *Cell Res.* 31, 613–630. doi:10.1038/s41422-020-00466-6
- Modolo, L., and Lerat, E. (2015). UrQT: an Efficient Software for the Unsupervised Quality Trimming of NGS Data. *BMC Bioinforma.* 16, 137. doi:10.1186/s12859-015-0546-8
- Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K., and Whitelaw, E. (1999). Epigenetic Inheritance at the agouti Locus in the Mouse. *Nat. Genet.* 23, 314–318. doi:10.1038/15490
- Orozco-Arias, S., Tobon-Orozco, N., Piña, J. S., Jiménez-Varón, C. F., Tabares-Soto, R., and Guyot, R. (2020). TIP_finder: An HPC Software to Detect Transposable Element Insertion Polymorphisms in Large Genomic Datasets. *Biology* 9, 281. doi:10.3390/biology9090281
- Ozata, D. M., Gainetdinov, I., Zoch, A., O'Carroll, D., and Zamore, P. D. (2019). PIWI-interacting RNAs: Small RNAs with Big Functions. *Nat. Rev. Genet.* 20, 89–108. doi:10.1038/s41576-018-0073-3
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* 14, 417–419. doi:10.1038/nmeth.4197
- Payer, L. M., and Burns, K. H. (2019). Transposable Elements in Human Genetic Disease. *Nat. Rev. Genet.* 20, 760–772. doi:10.1038/s41576-019-0165-8
- Rishishwar, L., Mariño-Ramírez, L., and Jordan, I. K. (2017). Benchmarking Computational Tools for Polymorphic Transposable Element Detection. *Brief. Bioinform.* 18, bbw072–918. doi:10.1093/bib/bbw072
- Sato, K., and Siomi, M. C. (2020). The piRNA Pathway in Drosophila Ovarian Germ and Somatic Cells. *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 96, 32–42. doi:10.2183/pjab.96.003
- Schrader, L., and Schmitz, J. (2019). The Impact of Transposable Elements in Adaptive Evolution. *Mol. Ecol.* 28, 1537–1549. doi:10.1111/mec.14794
- Schwarz, R., Koch, P., Wilbrandt, J., and Hoffmann, S. (2021). Locus-specific Expression Analysis of Transposable Elements. *Brief. Bioinform.* 23, bbab417. doi:10.1093/bib/bbab417
- Sexton, C. E., and Han, M. V. (2019). Paired-end Mappability of Transposable Elements in the Human Genome. *Mob. DNA* 10, 29. doi:10.1186/s13100-019-0172-5
- Slotkin, R. K., and Martienssen, R. (2007). Transposable Elements and the Epigenetic Regulation of the Genome. *Nat. Rev. Genet.* 8, 272–285. doi:10.1038/nrg2072
- Steinhauser, S., Kurzawa, N., Eils, R., and Herrmann, C. (2016). A Comprehensive Comparison of Tools for Differential ChIP-Seq Analysis. *Brief. Bioinform.* 17, bbv110. doi:10.1093/bib/bbv110
- Sun, X., Wang, X., Tang, Z., Grivainis, M., Kahler, D., Yun, C., et al. (2018). Transcription Factor Profiling Reveals Molecular Choreography and Key Regulators of Human Retrotransposon Expression. *Proc. Natl. Acad. Sci. U.S.A.* 115, E5526–E5535. doi:10.1073/pnas.1722565115
- Taylor, D., Lowe, R., Philippe, C., Cheng, K. C. L., Grant, O. A., Zabet, N. R., et al. (2021). Locus-specific Chromatin Profiling of Evolutionarily Young Transposable Elements. *Nucleic Acids Res.* 50, e33. gkab1232. doi:10.1093/nar/gkab1232
- Teissandier, A., Servant, N., Barillot, E., and Bourc'his, D. (2019). Tools and Best Practices for Retrotransposon Analysis Using High-Throughput Sequencing Data. *Mob. DNA* 10, 52. doi:10.1186/s13100-019-0192-1
- Uhrig, S., and Klein, H. (2019). PingPongPro: a Tool for the Detection of piRNA-Mediated Transposon-Silencing in Small RNA-Seq Data. *Bioinforma. Oxf. Engl.* 35, 335–336. doi:10.1093/bioinformatics/bty578
- Vendrell-Mir, P., Barteri, F., Merenciano, M., González, J., Casacuberta, J. M., and Castanera, R. (2019). A Benchmark of Transposon Insertion Detection Tools Using Real Data. *Mob. DNA* 10, 53. doi:10.1186/s13100-019-0197-9
- Verdore, L., Caserta, M., and Mauro, E. D. (2005). Role of Histone Acetylation in the Control of Gene Expression. *Biochem. Cell Biol.* 83, 344–353. doi:10.1139/o05-041
- Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., et al. (2007). A Unified Classification System for Eukaryotic Transposable Elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165-c4

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Lerat. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.