

BMB Reports – Manuscript Submission

Manuscript Draft

Manuscript Number: BMB-22-088

Title: Ongoing endeavors to detect mobilization of transposable elements

Article Type: Mini Review

Keywords: Transposable elements; transposition; in vivo retrotransposition assay; Oxford Nanopore Technology; ONT

Corresponding Author: Sungjin Moon

Authors: Yujeong Lee¹, Una Ha¹, Sungjin Moon^{1,*}

Institution: ¹Department of Biological Sciences, Kangwon National University, Chuncheon, 24341, South Korea,

Manuscript Type: Mini Review

Title

Ongoing endeavors to detect mobilization of transposable elements

Author's name

Yujeong Lee*, Una Ha*, Sungjin Moon

* These authors contributed equally to this work

Affiliation

Department of Biological Sciences, Kangwon National University, Chuncheon, 24341, South Korea

Running title

Ongoing endeavors to detect mobilization of transposable elements

Keywords: Transposable elements, transposition, *in vivo* retrotransposition assay, Oxford Nanopore Technology (ONT)

Corresponding Author's Information: +82-33-250-8536, sungjin.moon@kangwon.ac.kr

Abstract

Transposable elements (TEs) are DNA sequences capable of mobilization from one location to another in the genome. Since the discovery of 'Dissociation (*Dc*) locus' by Barbara McClintock in maize (1), mounting evidence in the era of genomics indicates that a significant fraction of most eukaryotic genomes is composed of TE sequences, involving in various aspects of biological processes such as development, physiology, diseases and evolution. Although technical advances in genomics have discovered numerous functional impacts of TE across species, our understanding of TEs is still ongoing process due to challenges resulted from complexity and abundance of TEs in the genome. In this mini-review, we briefly summarize biology of TEs and their impacts on the host genome, emphasizing importance of understanding TE landscape in the genome. Then, we introduce recent endeavors especially *in vivo* retrotransposition assays and long read sequencing technology for identifying *de novo* insertions/TE polymorphism, which will broaden our knowledge of extraordinary relationship between genomic cohabitants and their host.

Simple TE categories with complex characteristics

Broadly, TEs are grouped into class I (retrotransposons) or class II (DNA transposons), depending on a transposition mechanism with or without an RNA intermediate, respectively. Class I TEs further fall into two subclasses: long terminal repeats (LTR) or non-long terminal repeats (non-LTR) retrotransposons (2). In spite of structural difference between LTR and non-LTR retrotransposons, mobilization of class I TEs requires an RNA intermediate, followed by reverse transcription for the integration. Referred as "copy and paste", the original sequence of class I retrotransposons remains intact after integration into the genome. In contrast, class II DNA transposons propagate themselves via "cut and paste" mechanism. Once transcribed and translated from source DNA sequence, transposase machinery from class II recognizes and excises terminal inverted repeats (TIRs) flanking self-DNA sequence from the original location, and finally re-integrates into a new genomic location. TEs from both classes over time accumulate genetic alterations in their internal sequence and often become incompetent for transposition by themselves. TEs with such genetic alterations, called as non-autonomous TEs, still can propagate in the genome by hijacking protein machineries from other TEs competent for autonomous transposition (3).

Categorization of TEs by their mode of transposition looks rather simple. When it comes to details, nevertheless, biology of TEs is not only incredibly complicating but intriguingly puzzling. It is well known that a significant portion of human genome is occupied by repetitive sequences (4) and TEs are one type of such sequences that are interspersed throughout genome in different sizes and shapes. In the case of one heavily studied non-LTR retrotransposon, LINE1s (long interspersed nuclear elements, or L1s in short) take up approximately 17% of the human genome with half million copies (5). Although most of L1 copies exhibit genetic changes such as rearrangements, point mutations, and 5'-truncation unfavorable for mobilization (6), roughly 100 copies of L1s are estimated to be active in the genome. Polymorphism of the particular TEs also varies so that around 300 L1 insertion sites are known to differ between unrelated individuals (7), contributing to complexity and variations within the human population.

TEs that are inactive in one species are not necessarily same cases in other species. Most of LTR type retrotransposons in human (endogenous retrovirus or HERV), if not all of them, are thought to be inactive (8), whereas some LTR-retrotransposons in koala (the koala retrovirus or KoRV), a vertically transmitting endogenous retrovirus, is actively invading the koala population (9). A class II transposon, called as *P-elements*, is active in wild population of *Drosophila melanogaster* whereas DNA transposons in human are unable to mobilize due to mutations accumulated in internal sequences (10). Interestingly, *P-elements* had successfully completed world-wide invasion in the genome of *Drosophila melanogaster* population within a century. Given that activation of *Drosophila P-elements* results in genomic instability and animal sterility (11), survival of wild *Drosophila* population from the invasion infers that there must be an arms race between host and the invading *P-elements* (12-14). As a matter of fact, invasion of *P-elements* in closely related species, *Drosophila simulans*, is still ongoing process (15, 16).

Integration preference from individual TE family seems quite different as well. From a large collection of mutant fly lines with a single transposon insertion (the *Drosophila* Gene Disruption Project), three DNA transposons (*Minos*, *P-element* and *PiggyBac*) used in this project show different preference for their integration (17, 18). Analysis on a collection of fly lines with new insertion points out that preference for *Minos* elements comes close to a random distribution in fly genome, whereas *P-elements* have 200-400 "hotspot loci" accounting for 30-40% of new insertions and tend to be inserted near gene promoters. *PiggyBac* transposons also have hotspots, but are considerably different from those of *P-elements*. In other case, mobilization of TE takes place into a specific locus. R1/R2 non-LTR retrotransposons in *Drosophila* encode sequence-specific endonuclease responsible for mobilization exclusively into 28s rDNA locus, which has been a successful strategy for transposition in the genome at least for 500-800 millions of years (19). In summary, different features of TEs across the eukaryote phylogeny strongly argue that TEs are not simply kept as miscellaneous genetic sequences (20, 21). Instead, TE families from different species indeed deploy their unique strategy for propagation, which shapes a different landscape of host genome.

1 BTCFF97H98.DFCC:

Roles of the 'controlling elements' in genome: an ally or an enemy?

Unprecedented technical advances in genomics since early 2000s, a vast amount of sequencing data sheds light on the composition of the genome from a variety of species. With a few exceptions such as apicomplexan parasites (22), a significant fraction of eukaryotic genome is transposable elements (TEs) that are interspersed within a genome (4). Being once considered as genetic fossils during evolution, however, growing evidence undoubtedly points out that TEs participate in many aspects of biology, which seems to be irrelevant to "selfish elements". When TE mobilization takes place, an insertion can alter the local gene regulatory network or an inserted TE segment might introduce their own intrinsic transcriptional modules such as enhancers, insulators and repressors in new genomic loci. Industrial melanism (23), color/shape change in plant (24), and maize domestication (25) are the widely recognized such instances, illustrating extraordinary relationship between TEs and their hosts during evolution. Notably, abundance in genome and capability to interact with host transcription machinery offer changes in transcriptional regulation, which can interweave the expression of nearby host genes. Numerous studies have shown co-option or exaptation of TEs as a region for gene regulatory innovation (26, 27). For example, analysis from binding sites of orthologous TFs in human and mouse cell lines provided evidence of species-specific TE-derived binding peaks (28). Other cases for enrichment of non-autonomous MER130 family as active enhancers adjacent to neocortical genes for neural development (29) and association of MER20 elements in rewiring the gene network for pregnancy in placental mammals (30) indicate that TEs extensively act as one of driving forces for novel regulatory network during evolution. In addition to modulating the gene expression, hosts take advantage of TE-derived sequences that are repurposed for a new function. One classic example is the recombination activating 1 (RAG1) and recombination activating 2 (RAG2) proteins, which catalyze the V(D)J rearrangement essential for the adaptive immune systems. It was proposed that RAG1 and RAG2 were domesticated from the ancient *Transib* element in that the terminal inverted repeats (TIRs) and its way of arrangement are similar (31). It was found that *ProtoRAG*, an active *Transib* element from

the lancelet encodes RAG1-, RAG2-like genes and it resembles RAG1/2-mediated DNA rearrangement (32), which further solidified the origin of RAG1/2 system in jawed vertebrates.

Telomeres and centromeres, the two vital features of eukaryotic chromosomes, appear to be occupied by transposon sequences in some organisms. The terminal regions of a chromosome to maintain genome stability and integrity, telomeres of *Drosophila* are preserved by three retrotransposons, *Het-A*, *TART* and *TAHRE*, or collectively termed as HTT (33-36). Although *Het-A*- and *TART*-derived sequences are also found in centromeric regions of Y chromosome (37), most copies of HTT exist in telomere regions, pointing out their strong preference for transposition. This observation also brings intriguing view of how to achieve the balance between the rate of retrotransposition in such restricted region of the genome and telomere length regulation. In addition, centromeres essential for proper segregation of sister chromatids during cell division are composed of highly repetitive DNA sequences. Interestingly, recent reports on *Drosophila* (38) and others such as some plants (39), amoeba (40) and kangaroo (41) reveal that centromeres contains TE sequences. Combinatorial sequencing efforts on centromeres from *Drosophila melanogaster* show that all centromeres are enriched by a specific retrotransposon, *G2/Jockey-3*, flanked by satellite repeats (38), opening new questions of functional relationship between particular TEs and centromere biology.

Obviously, all transposition events are not necessarily advantageous to host in terms of physiology, development, and evolution. For examples, a large collection of mutant fly lines with a single transposon insertion (the *Drosophila* Gene Disruption Project) shows largely deleterious effects on nearby genes (17, 18), as well as transposition-induced embryonic lethality from mutagenesis studies using *PiggyBack* (PB)/*Sleeping Beauty* (SB) transposons in mouse (42, 43). Initially found from a patient with hemophilia, retrotransposition of LINE1s is known to cause other disorders as well, including many types of cancer (5, 44). Given occurrence of a new insertion by retrotransposons such as *Alu* elements (1 out of 20 live births), *LINE1* (1 out of 20-200 births) and *SVA* (1 out of 900 births), it is not peculiar to see that 124 cases of disorders are associated with

retrotransposition events (5) and that altered gene expression profile by retrotransposition could transform cells to be oncogenic (44-46). Besides of a role as an internal mutagen, sequence similarity and abundance of TEs in genome as substrates for non-allelic homologous recombination can give rise to intra- or inter-chromosomal crossing over. Although unequal crossing over may result in expansion of gene-rich segmental duplications found in human (47), recombination events mediated by integration sites of TEs such as *Alu* elements have been reported in various human disorders (48, 49).

To prevent deleterious effects from mobilization of TEs, hosts employ evolutionarily conserved molecular strategies to silence TEs transcriptionally and post-transcriptionally (50-52). When host surveillance systems such as DNA methylation 5mC (50) or small noncoding RNAs such as piRNAs in germline (51) are experimentally incapacitated, genome-wide derepression of transposons unsurprisingly conveys genomic catastrophe. It has been shown that unleashed from germline specific piRNA pathway in *Drosophila*, massively derepressed TE mobilome from germ cells takes host machinery to selectively target developing oocytes, the sole channel to the next generation (53). Similarly, genome-wide activation of *P-elements* in *Drosophila* leads to numerous *de novo* insertions in developing germ cells, resulting in genomic instability and female sterility (12, 13). In other species such as human, mouse, and zebrafish, derepression of TEs without host defense systems results in animal sterility as a phenotypic outcome (54-57). Taken together, examples in this section emphasize on extraordinary relationship between TEs and their host. Genome-wide TE activation unquestionably compromises genome integrity, leading to detrimental consequences such as animal sterility. Meanwhile integration events by TEs can rewire transcriptional regulation of genes, which seems to be symbiotic from many cases. Given insertional impacts on host genome, therefore, precise monitoring on TE mobilization is essential for better understanding of the double-edged swords coevolving with host.

Visualizing mobilization of TEs with cellular resolution

A life cycle of retrotransposon is a series of transcription, translation, followed by reverse transcription and finally integration into new locus with their own preference. Given complexity on sequence and abundance of TEs residing in the genome, cataloguing *de novo* insertions of TEs has been a quite challenging task. For last two decades, analysis of transcriptomic profile using RNA-seq is apparently one of choices to estimate the degree of derepression of TEs. In line with this, many studies using *Drosophila* as a model system heavily rely on the transcriptional level of TEs as a proxy for TE activation, which, in fact, is an indirect method to decipher transposon mobilization. For examples, studies using *Drosophila* female germline and a gut showed that actual mobilization events quantified at the genomic DNA level are less correlated to transcriptome or small RNAs (piRNAs) sequencing data, thus suggesting that RNA signatures provide insufficient information on the degree of transposition (53, 58). Even quantification of transposition events at the genomic DNA level lacks cellular resolution because majority of sequencing data represents TE insertions from a total sum of given cells or tissues. In this section, we introduce recent progress to visualize mobilization events spatiotemporally with cellular resolution in mammal and *Drosophila* model systems.

Endeavors to visualize mobilization of non-LTR TEs Estimated about 17% of human genome is LINE1 retrotransposons (L1s), non-long terminal repeat (non-LTR) retroelements. Of these, estimated 100 copies actively transpose (known as hot LINE1s) (59, 60) and trans-mobilize other non-autonomous transposons such as *Alu* elements, SINE, etc (61, 62). As studies on impacts of L1s in physiology and human disease are still ongoing, an engineered reporter system and its derivatives emulating the non-LTR life cycle have been widely used to address several questions on L1 biology. To capture retrotransposition events of L1s, briefly, a retrotransposition cassette contains a reporter such as green fluorescent protein (GFP) in 3' UTR region of LINE1 (Fig. 1A). This reporter, where orientation is in the opposite direction with respect to a normal L1 transcription, is engineered to disrupt its expression by an artificial intron in the middle of the reporter (IVS, intervening sequence

in Fig. 1A). While the engineered L1 cassette is being transcribed, the artificial intron from the reporter is spliced out, followed by the reverse transcription of the entire cassette. After the cassette without the artificial intron successfully is retrotransposed elsewhere in genome, one can observe a reporter signal (GFP) that enables to identify cells with *de novo* insertion events. This conceptual design of retrotransposition assay has been modified, and applied to many studies (Table 1); a retrotransposition cassette with CMV/CAG promoter (63-72); with endogenous human/mouse L1 promoter (73, 74); a same concept applied to *Drosophila I-element* similar with non-LTR retrotransposons containing two ORFs such as mammalian L1s (53, 75). The engineered L1 retrotransposition cassette has served as an excellent tool to appreciate mobilization of LINE1s in human genomes, heterogeneity in mouse brain as well as a selective integration into developing oocytes in *Drosophila melanogaster* (53, 64, 72, 73, 75-77).

Endeavors to visualize mobilization of LTR TEs The *gypsy* elements in *Drosophila melanogaster* are categorized as long terminal repeats (LTR) retrotransposon, similar to the *Ty* elements of *Saccharomyces cerevisiae* and vertebrate provirus of retrovirus (78-80). Interestingly, the *gypsy* elements tend to integrate into seven regions within a 200-bp of the promoter of the *ovo* gene, which is necessary for proper oogenesis in the female germline (81-83). Adopting such behavior of *gypsy* transposon generated a cassette to trace mobilization events, so-called a *gypsy*-TRAP line (Fig. 1B) (83, 84). Without *gypsy* mobilization, α -tubulin promoter in the cassette drives expression of the GAL80 proteins which suppress the activity of GAL4 transcription factors and subsequent downstream UAS-reporter (Fig. 1B, upper panel). When endogenous *gypsy* TEs retrotranspose into the known hotspots (the *ovo* promoter) in the *gypsy*-TRAP cassette, the expression of GAL80 become ceased. In turn, GAL4 transcription factors are then released from the suppression by GAL80, driving the expression of the reporter gene (Fig. 1B, lower panel) (85). This process permits to map cells *in vivo* with new integration events of *gypsy* transposons in a given tissue and developmental stage (Table 1). For examples, the *gypsy*-TRAP line provides evidence to support ideas that *gypsy* are activated in an aged fly brain (84); in aged fat body (equivalent to liver in mammal) (86, 87); in

aged fly intestine (88); in fly model of tauopathy (89); in FTD–ALS causing CHMP2B^{Intron5} mutation (90); in the developing mesodermal tissue with histone 3 lysine 9 (H3K9) substituted by arginine (H3R9) (91).

The *gypsy*-TRAP line, however, has some technical downsides. Firstly, the reporter signal might be a consequence of mobilization by other retrotransposons sharing similar hot spots (81). Secondly, the system needs three transgenes (a reporter under UAS regulatory elements, GAL4 transcription factor, and the engineered cassette with GAL80 suppressor) in one animal, rendering further applications of this system a little difficult. An interesting idea that takes advantages of *gypsy* transposon replication has been proposed to improve the *gypsy*-TRAP line. In the newer version, named as cellular labeling of endogenous retrovirus replication reporter or CLEVR in short (Fig. 1C and Table 1) (92), the conserved features of retrovirus replication is applied to the system. Briefly, 5' end of the 3'-LTR (U3" region in Fig. 1C) in the *gypsy* RNA is used as a template for the synthesis of 5' end of the 5'-LTR during replication. Similarly, 3' end of the 5'-LTR (U5 region in Fig. 1C) is used for the synthesis of 3' end of the 3'-LTR. This system additionally includes a GFP-P2A-mCherry reporter at the 5' end of the 3'-LTR (U3" region in Fig. 1C) and 5X UAS regulatory elements at the 3' end of the 5'-LTR (U5 region in Fig. 1C), where orientation of both the reporter and 5X UAS regulatory elements is opposite with respect to the *gypsy* transcription. Upon completion of *de novo* integration, it generates two hybrid LTRs with 5X UAS regulatory elements in the vicinity of the GFP-P2A-mCherry at both ends. When combined with a tissue-specific driver such as glial-specific GAL4 (*repo*-GAL4), fluorescent reporter signals in the fly brain were detectable in age-dependent manner (92). The signal was all disappeared when *gypsy*-RNAi was introduced to the same animal, verifying its specificity (92). This system was further used to show that 1) *gypsy* transposons like a retrovirus are capable of transmitting intercellularly between *Drosophila* cells grown in the same culture media (93), in consistent with finding that *gypsy* mobilization involves production of *gypsy* viral particles in the follicle cells to infect the oocyte in *Drosophila* (94) and that 2) fly driving glial-specific expression of hTDP43, found to be aggregated in amyotrophic lateral

sclerosis (ALS) and frontotemporal dementia (FTD), experiences *gypsy*-ERV activation, DNA damage and apoptosis in both glia and nearby neurons, hypothesizing the endogenous retrovirus might contribute to TDP-43-mediated neurodegeneration in non-cell-autonomous manner (95). Since CLEVR strategy can be applicable to other LTR type retrotransposons, it will enable to improve our understanding of transposable mobilization further in various contexts of biology.

Detecting mobilization with nucleotide resolution

Detecting TE mobilization events in earlier days relied on traditional methods such as Southern blotting and *in situ* hybridization on polytene chromosomes (96-98), which provide rough information on integration sites in the genome. PCR-based strategies such as L1 display, ATLAS (amplification typing of L1 active subfamilies) and LIDSIP (L1 insertion dimorphisms identification by PCR) (99-101), and transposon display (TD) (102) have been used to locate insertion sites of TEs. In spite of efficiency and versatility, prior annotation for the TE of interest is a prerequisite for successful identification of new insertions. Recent technical advances in DNA sequencing together with computational tools have opened up an unprecedented level of TE landscape from given tissues or contexts with nucleotide resolution, rapidly becoming a standard tool. As numerous computational tools on the basis of mapping events via split-reads or discordant read pairs (or in combination of both) are developed to detect polymorphic TE insertions/deletions (103), tremendous endeavors have identified *de novo* and mosaic TE insertions; and such discoveries engage in dissection of functional impacts on a variety of organisms (13, 44, 53, 104). In terms of mappability, nonetheless, utilization of short reads to study TEs mobilization can be challenging due to repetitive nature of TEs. Mapping of short reads on repetitive regions in genome can create ambiguities and incomplete contiguity, which might mislead data interpretation. For examples, a measurement of *de novo* insertion events by short reads based on an Illumina platform pointed out high false positive rates despite high sensitivity of the mapping strategy (105), suggesting that somatic transposition events might be less prevalent than expected. This study argues a fundamental flaw owing probably to unavoidable chimeric artifacts during library preparation and detection algorithm (105). Another example is that exon trapping by a 2kb SVA transposon insertion in the intron of MFSD8 was initially missed by standard clinical sequencing, suggesting that accurate tracking of transposon is a critical point especially for genomic medicine (106).

Detecting mobilization by Oxford Nanopore Technologies (ONT)

So-called third generation sequencing (TGS) implemented with ultra-long reads have

recently emerged (107), allowing to produce sequencing reads typically more than several kilo-bases long. Since ultra-long sequencing reads can solve the mappability issue from short sequencing reads (108), many platforms have been developed, including Oxford Nanopore Technologies (ONT), Pacific Biosciences (PacBio), etc (109). Although each platform undoubtedly provides its unique potential to decipher the complexity of genome, we briefly review especially Oxford Nanopore Technologies (ONT) related in TE biology due to its feasibility for a small-sized laboratory. Since its finding on the bio-pore that enables to detect ionic current blockage as nucleic acid polymers pass through (110, 111), a pocket-sized device of MinION nanopore technology has been developed to trace changes in voltage corresponding to DNA sequences. Released in 2014 to early-access users (112), it has undergone a series of improvement on signal-to-noise ratio (113, 114) as well as algorithms for long-read data analysis (115). Efforts on improvement of ONT sequencing technology are ongoing process to achieve higher accuracy (116-119), extend read length (120, 121) and increase throughput (122-125). Shown to produce a read size even up to a mega base scale (126), it is now being applied to study unanswered questions such as finishing gaps in reference genome of human, nematode, plant, zebrafish and fruit fly (127-135) or to build non-reference genome by *de novo* assembly (108). ONT sequencing has been in attention as it can complement some of drawbacks from short-read sequencing, including mappability on repetitive regions, technical biases during library preparation and so on. Although read accuracy (87-98 %) of ONT sequencing needs to be improved when compared to short read sequencing (>99.9 %), nanopore technique can detect new insertions and intact target site duplications of TEs from single read (136). ONT sequencing detected 46 new TE insertions/TE losses from *Drosophila* reference stock maintained in a laboratory for more than 350 generations, compared to the reference genome (128), and possibly becomes feasible to study TE dynamics in population scale as well (137). From different types of cancer samples, ONT method further discovered new insertions of TEs that had not been catalogued previously. For instances, ONT method in combination with LDI-PCR uncovered new insertional events of LINE1 in colorectal tumor that were undetectable with 40x sequencing (138). Analysis on

liver cancers, which were previously sequenced for the International Cancer Genome Consortium (ICGC), identified germline and somatic structural variations (SVs) probably caused by non-allelic homolog recombination (NAHR) by SINE transposons, providing evidence that long reads can be a valuable platform for detection of structural variations than short reads based approach (139).

Although ONT technology requires further improvement to accomplish better read accuracy than the current status, there has been efforts in conjunction with other genomic approaches to capture more precise landscape of transposons. In combination with Hi-C scaffolding, ONT sequencing from two wild type strains of *Drosophila melanogaster* identified hundreds of TE insertions missed from short read-based studies before (140). Another study focusing on clonal neoplasia in fly gut tissue combined short and long read sequencing to profile directly somatic TE insertions from genomic DNA samples instead of TE transcriptome (141). In addition, efforts on filling large gaps in a centromere utilized ONT sequencing method in conjunction with ChIP-seq for CENP-A (centromeric histone), and super-resolution chromatin fiber imaging, leading to discovery that centromeres of *Drosophila melanogaster* are occupied by non-long terminal repeats (non-LTR) retroelement of *Jockey* family, *G2/Jockey-3* (38). Although a question of how such retroelement become a major source in all centromeres of *Drosophila melanogaster* remains to be elucidated, it is remarkable that long read sequencing advances our understanding and unveils new questions about centromere of *Drosophila* genus.

It is well known that DNA modification such as 5-methylcytosine transcriptionally suppresses transposons in many organisms (50). 5-methylcytosine is technically detectable by pre-treated chemical reaction to differentiate from cytosine, also known as bisulfite sequencing (142). As ONT platform utilizes native nucleic acids as a substrate, it precludes the pre-chemical reaction during library preparation and is able to directly ask 5-methylcytosine status on repetitive regions of the genome. Studies showed that DNA methyl transferase I (Dnmt1)-dependent DNA methylation is enriched at IAP retrotransposons in mouse embryonic stem cells, and that locus-specific methylation of human TEs from many types of tissues and liver tumor show DNA methylation

dynamics and its effect on the repression of TEs (143, 144).

In addition to reading DNA modification by sequencing native genomic DNA molecules, ONT system is also able to directly sequence long RNA molecules or complementary DNA (cDNA). Sequencing of the 5'-Cap-captured native full length RNAs from *Locust* genome identified widespread TE exonization, which was a computationally challenging task by short-read sequencing (145). Direct reading of long RNA sequences was able to catalogue previously unannotated retrotransposon-related transcripts at the early stage of triticale seed development (146), opening new questions on their role in phenotypic variations. ONT sequencing on cDNA generated from virus-like particles (VLPs) from *dsm1* mutants identified active transposons in *Arabidopsis thaliana* and *Zea mays* without mapping to genomic DNA (147, 148). Despite the requirement of a large quantity of nucleic acids samples, direct sequencing of RNA molecules can expand understanding of TE transcriptome and bypass biases that might be integrated by PCR amplification or reverse transcription. Taken together, ONT platform, since its release, certainly has been a means to scrutinize TE biology with distinct perspective that short read sequencing was unable with and opens up promising opportunities to appreciate further impacts of TEs in physiology, development and evolution.

Summary

Most of living organisms in their genome have experienced TE propagation that takes place regardless of host fitness during evolution. Beneficial *de novo* insertions have established mutualistic relationship between TEs and host that undergo positive selection over time, meanwhile mutagenic transposition cases negatively impact on host fitness. Studies of more than 70 years have been expanded to explore fine mapping of polymorphic TEs between individuals/species; functional impacts of TEs in diverse contexts; the molecular basis on TE silencing in a given tissue/species and so on. Despite exciting discoveries, challenges due to the complexity of TEs in the genome have still placed our understanding of TEs to be ongoing progress. In addition to technical approaches introduced in this article, there are other efforts with distinct perspectives to decipher complexity of TEs. For example, single cell genomics combined with long read sequencing catalogued comprehensive information of TE expression (149). Proteomic approach ('proteomics informed by transcriptomics') characterized active TEs in poorly annotated organisms such as *Aedes aegypti* (150). Together with rapidly developing techniques, multifaceted approaches such as *in vivo* retrotransposition reporter in conjunction with digital droplet PCR (76), PacBio platform (77), short reads (53) show synergetic effects to broaden our knowledge of TEs. In summary, continuous endeavors for technical breakthrough and combination of techniques with different angles will provide better spatiotemporal resolution on TE biology, which will help us to assess further relationship between transposable elements and host in diverse contexts.

Acknowledgements

Authors in this work sincerely apologize for presenting a minuscule fraction of colleagues' work due to page limitations. We deeply thank to Dr. Yun Doo Chung and Dr. Sim Namkoong for comments on the manuscript. SM, YL and UH drafted this manuscript and all agreed the submitted version. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (2020R1C1C1007050) and by 2020 Research Grant from Kangwon National University (520200047) given to SM.

Conflict

The authors declare no conflict of interest.

Figure legend

Figure 1. Strategies for monitoring retrotransposition events with cellular resolution. **A.** An overview of the LINE1 retrotransposition assay (adopted from (64)). The L1-cassette contains reporter (shown is GFP) in the opposite direction to normal L1 transcription. As transcription occurs, the intron that disrupts the reporter is removed by a splicing event. SD, splice donor; SA, splice acceptor; IVS, intervening sequence. **B.** A schematic of *gypsy*-TRAP reporter (adopted from (84)). i) No integration: Expression of GAL80 under α -tubulin promoter with intact hot spots (*ovo* promoter) suppresses GAL4-mediated transcription, thus no GFP signal is detectable. ii) *Gypsy* integration: If integration of *gypsy* into the *ovo* binding site occurs, depicted by triangles, GAL4-driven GFP expression becomes detectable as GAL80 expression is ceased. UAS, Upstream Activating Sequence. **C.** A *gypsy*-CLEVR reporter mimicking the replication cycle of retrovirus (adopted from (92)). During replication of *gypsy* retrotransposon, 5' end of the 3'-LTR region (U3" in black box) and 3' end of the 5' LTR region (U5 in white box) are respectively used for synthesis of 5' end of 5'-LTR region and 3' end of 3'-LTR region. Note that both 5X UAS in 5'-LTR region (U5 in white box) and a reporter (GFP-P2A-mCherry) in 3'-LTR region (U3" in black box) are in opposite direction to *gypsy* transcription. Retrotransposition of *gypsy*-CLEVR leads 5X UAS in vicinity to the reporter (GFP-P2A-mCherry), allowing expression of the reporter by GAL4 activator. PBS, primer binding site; LTR, Long Terminal Repeats.

1. Mc CB (1950) The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences of the United States of America* 36(6):344-355.
2. Kazazian HH, Jr. & Moran JV (2017) Mobile DNA in Health and Disease. *The New England journal of medicine* 377(4):361-370.
3. Richardson SR, *et al.* (2015) The Influence of LINE-1 and SINE Retrotransposons on Mammalian Genomes. *Microbiology spectrum* 3(2):MDNA3-0061-2014.
4. Cordaux R & Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nature reviews. Genetics* 10(10):691-703.
5. Hancks DC & Kazazian HH, Jr. (2016) Roles for retrotransposon insertions in human disease. *Mobile DNA* 7:9.
6. Ostertag EM & Kazazian HH, Jr. (2001) Biology of mammalian L1 retrotransposons. *Annual review of genetics* 35:501-538.
7. Ewing AD & Kazazian HH, Jr. (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome research* 20(9):1262-1270.
8. Mager DL & Stoye JP (2015) Mammalian Endogenous Retroviruses. *Microbiology spectrum* 3(1):MDNA3-0009-2014.
9. Tarlinton RE, Meers J, & Young PR (2006) Retroviral invasion of the koala genome. *Nature* 442(7098):79-81.
10. Lander ES, *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921.
11. Clark JB & Kidwell MG (1997) A phylogenetic perspective on P transposable element evolution in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 94(21):11428-11433.
12. Khurana JS, *et al.* (2011) Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* 147(7):1551-1563.

13. Moon S, *et al.* (2018) A Robust Transposon-Endogenizing Response from Germline Stem Cells. *Developmental cell* 47(5):660-671 e663.
14. Kofler R, Senti KA, Nolte V, Tobler R, & Schlotterer C (2018) Molecular dissection of a natural transposable element invasion. *Genome research* 28(6):824-835.
15. Kofler R, Hill T, Nolte V, Betancourt AJ, & Schlotterer C (2015) The recent invasion of natural *Drosophila simulans* populations by the P-element. *Proceedings of the National Academy of Sciences of the United States of America* 112(21):6659-6663.
16. Hill T, Schlotterer C, & Betancourt AJ (2016) Hybrid Dysgenesis in *Drosophila simulans* Associated with a Rapid Invasion of the P-Element. *PLoS genetics* 12(3):e1005920.
17. Bellen HJ, *et al.* (2011) The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188(3):731-743.
18. Spradling AC, Bellen HJ, & Hoskins RA (2011) *Drosophila* P elements preferentially transpose to replication origins. *Proceedings of the National Academy of Sciences of the United States of America* 108(38):15948-15953.
19. Eickbush TH & Eickbush DG (2015) Integration, Regulation, and Long-Term Stability of R2 Retrotransposons. *Microbiology spectrum* 3(2):MDNA3-0011-2014.
20. Huang CR, Burns KH, & Boeke JD (2012) Active transposition in genomes. *Annual review of genetics* 46:651-675.
21. Sotero-Caio CG, Platt RN, 2nd, Suh A, & Ray DA (2017) Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome biology and evolution* 9(1):161-177.
22. Kissinger JC & DeBarry J (2011) Genome cartography: charting the apicomplexan genome. *Trends in parasitology* 27(8):345-354.
23. Van't Hof AE, *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534(7605):102-105.
24. Lisch D (2013) How important are transposons for plant evolution? *Nature reviews. Genetics* 14(1):49-61.

25. Studer A, Zhao Q, Ross-Ibarra J, & Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature genetics* 43(11):1160-1163.
26. Rebollo R, Farivar S, & Mager DL (2012) C-GATE - catalogue of genes affected by transposable elements. *Mobile DNA* 3(1):9.
27. de Souza FS, Franchini LF, & Rubinstein M (2013) Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Molecular biology and evolution* 30(6):1239-1251.
28. Sundaram V, *et al.* (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome research* 24(12):1963-1976.
29. Notwell JH, Chung T, Heavner W, & Bejerano G (2015) A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nature communications* 6:6644.
30. Lynch VJ, Leclerc RD, May G, & Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature genetics* 43(11):1154-1159.
31. Kapitonov VV & Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS biology* 3(6):e181.
32. Huang S, *et al.* (2016) Discovery of an Active RAG Transposon Illuminates the Origins of V(D)J Recombination. *Cell* 166(1):102-114.
33. Traverse KL & Pardue ML (1988) A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proceedings of the National Academy of Sciences of the United States of America* 85(21):8116-8120.
34. Levis RW, Ganesan R, Houtchens K, Tolar LA, & Sheen FM (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75(6):1083-1093.
35. Abad JP, *et al.* (2004) TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Molecular biology and evolution*

- 21(9):1620-1624.
36. Villasante A, *et al.* (2007) Drosophila telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome research* 17(12):1909-1918.
 37. Agudo M, *et al.* (1999) Centromeres from telomeres? The centromeric region of the Y chromosome of Drosophila melanogaster contains a tandem array of telomeric HeT-A- and TART-related sequences. *Nucleic acids research* 27(16):3318-3324.
 38. Chang CH, *et al.* (2019) Islands of retroelements are major components of Drosophila centromeres. *PLoS biology* 17(5):e3000241.
 39. Neumann P, *et al.* (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* 2(1):4.
 40. Glockner G & Heidel AJ (2009) Centromere sequence and dynamics in Dictyostelium discoideum. *Nucleic acids research* 37(6):1809-1816.
 41. Ferreri GC, *et al.* (2011) Recent amplification of the kangaroo endogenous retrovirus, KERV, limited to the centromere. *Journal of virology* 85(10):4761-4771.
 42. Rad R, *et al.* (2010) PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* 330(6007):1104-1107.
 43. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, & Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 436(7048):221-226.
 44. Lee E, *et al.* (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337(6097):967-971.
 45. Babaian A & Mager DL (2016) Endogenous retroviral promoter exaptation in human cancer. *Mobile DNA* 7:24.
 46. Shukla R, *et al.* (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153(1):101-111.
 47. Bailey JA, Liu G, & Eichler EE (2003) An Alu transposition model for the origin and expansion

- of human segmental duplications. *American journal of human genetics* 73(4):823-834.
48. Montgomery EA, Huang SM, Langley CH, & Judd BH (1991) Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* 129(4):1085-1098.
49. Deininger PL & Batzer MA (1999) Alu repeats and human disease. *Molecular genetics and metabolism* 67(3):183-193.
50. Deniz O, Frost JM, & Branco MR (2019) Regulation of transposable elements by DNA modifications. *Nature reviews. Genetics* 20(7):417-431.
51. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D, & Zamore PD (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nature reviews. Genetics* 20(2):89-108.
52. Yang P, Wang Y, & Macfarlan TS (2017) The Role of KRAB-ZFPs in Transposable Element Repression and Mammalian Evolution. *Trends in genetics : TIG* 33(11):871-881.
53. Wang L, Dou K, Moon S, Tan FJ, & Zhang ZZ (2018) Hijacking Oogenesis Enables Massive Propagation of LINE and Retroviral Transposons. *Cell* 174(5):1082-1094 e1012.
54. Nagirnaja L, *et al.* (2021) Variant PNLDC1, Defective piRNA Processing, and Azoospermia. *The New England journal of medicine.*
55. Houwing S, *et al.* (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129(1):69-82.
56. Carmell MA, *et al.* (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Developmental cell* 12(4):503-514.
57. Bourc'his D & Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431(7004):96-99.
58. Siudeja K, *et al.* (2021) Unraveling the features of somatic transposition in the *Drosophila* intestine. *The EMBO journal* 40(9):e106388.
59. Brouha B, *et al.* (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proceedings of the National Academy of Sciences of the United States of*

- America* 100(9):5280-5285.
60. Beck CR, *et al.* (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141(7):1159-1170.
 61. Hancks DC, Goodier JL, Mandal PK, Cheung LE, & Kazazian HH, Jr. (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. *Human molecular genetics* 20(17):3386-3400.
 62. Raiz J, *et al.* (2012) The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic acids research* 40(4):1666-1683.
 63. Boeke JD, Garfinkel DJ, Styles CA, & Fink GR (1985) Ty elements transpose through an RNA intermediate. *Cell* 40(3):491-500.
 64. Moran JV, *et al.* (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87(5):917-927.
 65. Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV, & Kazazian HH, Jr. (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic acids research* 28(6):1418-1423.
 66. Xie Y, Rosser JM, Thompson TL, Boeke JD, & An W (2011) Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic acids research* 39(3):e16.
 67. Kannan M, *et al.* (2017) Dynamic silencing of somatic L1 retrotransposon insertions reflects the developmental and cellular contexts of their genomic integration. *Mobile DNA* 8:8.
 68. Kopera HC, *et al.* (2016) LINE-1 Cultured Cell Retrotransposition Assay. *Methods in molecular biology* 1400:139-156.
 69. Wei W, Morrish TA, Alisch RS, & Moran JV (2000) A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Analytical biochemistry* 284(2):435-438.
 70. Morrish TA, *et al.* (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature genetics* 31(2):159-165.

71. Feng Q, Moran JV, Kazazian HH, Jr., & Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87(5):905-916.
72. An W, *et al.* (2006) Active retrotransposition by a synthetic L1 element in mice. *Proceedings of the National Academy of Sciences of the United States of America* 103(49):18662-18667.
73. Muotri AR, *et al.* (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435(7044):903-910.
74. Newkirk SJ, *et al.* (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 114(28):E5635-E5644.
75. Jensen S & Heidmann T (1991) An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. *The EMBO journal* 10(7):1927-1937.
76. Sultana T, *et al.* (2019) The Landscape of L1 Retrotransposons in the Human Genome Is Shaped by Pre-insertion Sequence Biases and Post-insertion Selection. *Molecular cell* 74(3):555-570 e557.
77. Flasch DA, *et al.* (2019) Genome-wide de novo L1 Retrotransposition Connects Endonuclease Activity with Replication. *Cell* 177(4):837-851 e828.
78. Marlor RL, Parkhurst SM, & Corces VG (1986) The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Molecular and cellular biology* 6(4):1129-1134.
79. Kim A, *et al.* (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America* 91(4):1285-1289.
80. Song SU, Gerasimova T, Kurkulos M, Boeke JD, & Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes & development* 8(17):2046-2057.

81. Mevel-Ninio M, Mariol MC, & Gans M (1989) Mobilization of the gypsy and copia retrotransposons in *Drosophila melanogaster* induces reversion of the ovo dominant female-sterile mutations: molecular analysis of revertant alleles. *The EMBO journal* 8(5):1549-1558.
82. Dej KJ, Gerasimova T, Corces VG, & Boeke JD (1998) A hotspot for the *Drosophila* gypsy retroelement in the ovo locus. *Nucleic acids research* 26(17):4019-4025.
83. Labrador M, Sha K, Li A, & Corces VG (2008) Insulator and Ovo proteins determine the frequency and specificity of insertion of the gypsy retrotransposon in *Drosophila melanogaster*. *Genetics* 180(3):1367-1378.
84. Li W, *et al.* (2013) Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nature neuroscience* 16(5):529-531.
85. Duffy JB (2002) GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* 34(1-2):1-15.
86. Jones BC, *et al.* (2016) A somatic piRNA pathway in the *Drosophila* fat body ensures metabolic homeostasis and normal lifespan. *Nature communications* 7:13856.
87. Wood JG, *et al.* (2016) Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* 113(40):11277-11282.
88. Sousa-Victor P, *et al.* (2017) Piwi Is Required to Limit Exhaustion of Aging Somatic Stem Cells. *Cell Rep* 20(11):2527-2537.
89. Sun W, Samimi H, Gamez M, Zare H, & Frost B (2018) Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nature neuroscience* 21(8):1038-1048.
90. Fort-Aznar L, Ugbo C, & Sweeney ST (2020) Retrovirus reactivation in CHMP2B Intron5 models of frontotemporal dementia. *Human molecular genetics* 29(16):2637-2646.
91. Penke TJ, McKay DJ, Strahl BD, Matera AG, & Duronio RJ (2016) Direct interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes & development* 30(16):1866-

- 1880.
92. Chang YH, Keegan RM, Prazak L, & Dubnau J (2019) Cellular labeling of endogenous retrovirus replication (CLEVR) reveals de novo insertions of the gypsy retrotransposable element in cell culture and in both neurons and glial cells of aging fruit flies. *PLoS biology* 17(5):e3000278.
 93. Keegan RM, Talbot LR, Chang YH, Metzger MJ, & Dubnau J (2021) Intercellular viral spread and intracellular transposition of *Drosophila gypsy*. *PLoS genetics* 17(4):e1009535.
 94. Song SU, Kurkulos M, Boeke JD, & Corces VG (1997) Infection of the germ line by retroviral particles produced in the follicle cells: a possible mechanism for the mobilization of the gypsy retroelement of *Drosophila*. *Development* 124(14):2789-2798.
 95. Chang YH & Dubnau J (2019) The Gypsy Endogenous Retrovirus Drives Non-Cell-Autonomous Propagation in a *Drosophila* TDP-43 Model of Neurodegeneration. *Current biology : CB* 29(19):3135-3152 e3134.
 96. Wensink PC, Tabata S, & Pacht C (1979) The clustered and scrambled arrangement of moderately repetitive elements in *Drosophila* DNA. *Cell* 18(4):1231-1246.
 97. Potter S, Truett M, Phillips M, & Maher A (1980) Eucaryotic transposable genetic elements with inverted terminal repeats. *Cell* 20(3):639-647.
 98. Biemont C, Ronsseray S, Anxolabehere D, Izaabel H, & Gautier C (1990) Localization of P elements, copy number regulation, and cytotype determination in *Drosophila melanogaster*. *Genetical research* 56(1):3-14.
 99. Badge RM, Alisch RS, & Moran JV (2003) ATLAS: a system to selectively identify human-specific L1 insertions. *American journal of human genetics* 72(4):823-838.
 100. Sheen FM, *et al.* (2000) Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome research* 10(10):1496-1508.
 101. Pornthanakasem W & Mutirangura A (2004) LINE-1 insertion dimorphisms identification by PCR. *BioTechniques* 37(5):750, 752.

102. Kim EY, Fan W, & Cho J (2021) Determination of TE Insertion Positions Using Transposon Display. *Methods in molecular biology* 2250:115-121.
103. Goerner-Potvin P & Bourque G (2018) Computational tools to unmask transposable elements. *Nature reviews. Genetics* 19(11):688-704.
104. Tubio JMC, *et al.* (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345(6196):1251343.
105. Treiber CD & Waddell S (2017) Resolving the prevalence of somatic transposition in *Drosophila*. *eLife* 6.
106. Kim J, *et al.* (2019) Patient-Customized Oligonucleotide Therapy for a Rare Genetic Disease. *The New England journal of medicine* 381(17):1644-1652.
107. van Dijk EL, Jaszczyszyn Y, Naquin D, & Thermes C (2018) The Third Revolution in Sequencing Technology. *Trends in genetics : TIG* 34(9):666-681.
108. Wang Y, Zhao Y, Bollas A, Wang Y, & Au KF (2021) Nanopore sequencing technology, bioinformatics and applications. *Nature biotechnology* 39(11):1348-1365.
109. Logsdon GA, Vollger MR, & Eichler EE (2020) Long-read human genome sequencing and its applications. *Nature reviews. Genetics* 21(10):597-614.
110. Kasianowicz JJ, Brandin E, Branton D, & Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proceedings of the National Academy of Sciences of the United States of America* 93(24):13770-13773.
111. Song L, *et al.* (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274(5294):1859-1866.
112. Jain M, Olsen HE, Paten B, & Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome biology* 17(1):239.
113. Manrao EA, *et al.* (2012) Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nature biotechnology* 30(4):349-353.
114. Chorf GM, *et al.* (2012) Automated forward and reverse ratcheting of DNA in a nanopore at

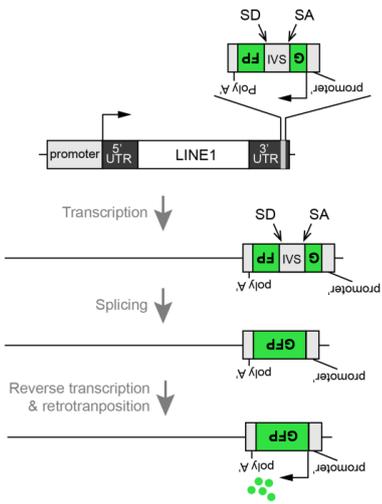
- 5-A precision. *Nature biotechnology* 30(4):344-348.
115. Amarasinghe SL, Ritchie ME, & Gouil Q (2021) long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. *GigaScience* 10(2).
116. Gong L, *et al.* (2018) Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nature methods* 15(6):455-460.
117. Goodwin S, *et al.* (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome research* 25(11):1750-1756.
118. David M, Dursi LJ, Yao D, Boutros PC, & Simpson JT (2017) Nanocall: an open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics* 33(1):49-55.
119. Boza V, Brejova B, & Vinar T (2017) DeepNano: Deep recurrent neural networks for base calling in MinION nanopore reads. *PloS one* 12(6):e0178751.
120. Gong L, Wong CH, Idol J, Ngan CY, & Wei CL (2019) Ultra-long Read Sequencing for Whole Genomic DNA Analysis. *Journal of visualized experiments : JoVE* (145).
121. Branton D & Deamer DW (2018) *Nanopore sequencing : an introduction* (World Scientific, New Jersey) p p.
122. De Coster W, *et al.* (2019) Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome research* 29(7):1178-1187.
123. De Roeck A, *et al.* (2019) NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome biology* 20(1):239.
124. Kim HS, *et al.* (2019) Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information. *GigaScience* 8(12).
125. Nicholls SM, Quick JC, Tang S, & Loman NJ (2019) Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience* 8(5).
126. Payne A, Holmes N, Rakyan V, & Loose M (2019) BulkVis: a graphical viewer for Oxford

- nanopore bulk FAST5 files. *Bioinformatics* 35(13):2193-2198.
127. Jain M, *et al.* (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature biotechnology* 36(4):338-345.
128. Solares EA, *et al.* (2018) Rapid Low-Cost Assembly of the *Drosophila melanogaster* Reference Genome Using Low-Coverage, Long-Read Sequencing. *G3 (Bethesda)* 8(10):3143-3154.
129. Tyson JR, *et al.* (2018) MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome research* 28(2):266-274.
130. Jain M, *et al.* (2018) Linear assembly of a human centromere on the Y chromosome. *Nature biotechnology* 36(4):321-323.
131. Miga KH, *et al.* (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585(7823):79-84.
132. Michael TP, *et al.* (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nature communications* 9(1):541.
133. Chernyavskaya Y, Zhang X, Liu J, & Blackburn J (2022) Long-read sequencing of the zebrafish genome reorganizes genomic architecture. *BMC genomics* 23(1):116.
134. Garg S, *et al.* (2021) Chromosome-scale, haplotype-resolved assembly of human genomes. *Nature biotechnology* 39(3):309-312.
135. Kirov I, *et al.* (2021) Transposons Hidden in *Arabidopsis thaliana* Genome Assembly Gaps and Mobilization of Non-Autonomous LTR Retrotransposons Unravelled by Nanotei Pipeline. *Plants* 10(12).
136. Mohamed M, *et al.* (2020) A Transposon Story: From TE Content to TE Dynamic Invasion of *Drosophila* Genomes Using the Single-Molecule Sequencing Technology from Oxford Nanopore. *Cells* 9(8).
137. Miller DE, Staber C, Zeitlinger J, & Hawley RS (2018) Highly Contiguous Genome Assemblies of 15 *Drosophila* Species Generated Using Nanopore Sequencing. *G3* 8(10):3131-3141.
138. Pradhan B, *et al.* (2017) Detection of subclonal L1 transductions in colorectal cancer by long-

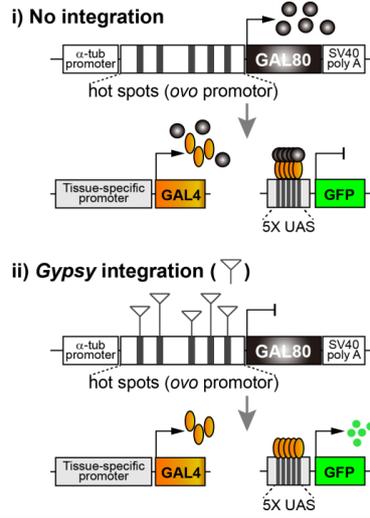
- distance inverse-PCR and Nanopore sequencing. *Scientific reports* 7(1):14521.
139. Fujimoto A, *et al.* (2021) Whole-genome sequencing with long reads reveals complex structure and origin of structural variation in human genetic variations and somatic mutations in cancer. *Genome medicine* 13(1):65.
140. Ellison CE & Cao W (2020) Nanopore sequencing and Hi-C scaffolding provide insight into the evolutionary dynamics of transposable elements and piRNA production in wild strains of *Drosophila melanogaster*. *Nucleic acids research* 48(1):290-303.
141. Siudeja K, *et al.* (2021) Unraveling the features of somatic transposition in the *Drosophila* intestine. *The EMBO journal*:e106388.
142. Frommer M, *et al.* (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proceedings of the National Academy of Sciences of the United States of America* 89(5):1827-1831.
143. Ewing AD, *et al.* (2020) Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling. *Mol Cell* 80(5):915-928 e915.
144. Haggerty C, *et al.* (2021) Dnmt1 has de novo activity targeted to transposable elements. *Nature structural & molecular biology* 28(7):594-603.
145. Jiang F, *et al.* (2019) Long-read direct RNA sequencing by 5'-Cap capturing reveals the impact of Piwi on the widespread exonization of transposable elements in locusts. *RNA biology* 16(7):950-959.
146. Kirov I, *et al.* (2020) Nanopore RNA Sequencing Revealed Long Non-Coding and LTR Retrotransposon-Related RNAs Expressed at Early Stages of Triticale SEED Development. *Plants* 9(12).
147. Lee SC, *et al.* (2020) Arabidopsis retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. *Genome research* 30(4):576-588.
148. Panda K & Slotkin RK (2020) Long-Read cDNA Sequencing Enables a "Gene-Like" Transcript Annotation of Transposable Elements. *The Plant cell* 32(9):2687-2698.

149. Berrens RV, *et al.* (2021) Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nature biotechnology*.
150. Maringer K, *et al.* (2017) Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in *Aedes aegypti*. *BMC genomics* 18(1):101.

A LINE-1 retrotransposition assay



B Gypsy-TRAP



C CLEVR

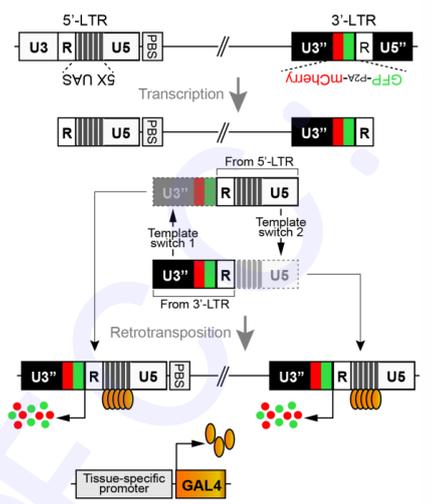


Fig. 1.

Transposition assay	Cell / Tissue	Reporter	Ref
LINE-1 retrotransposition assay	Cell line (Human)	EGFP	59, 65, 73, 77
		NEO ^R	60, 61, 62, 64, 69, 71, 76, 77
		Luciferase	66
		TEM1	67
	Cell line (Mouse)	NEO ^R	64
	Cell line (CHO)	NEO ^R	60, 70
	Tissues (Mouse/Rat)	EGFP	72, 73, 74
Tissues (Drosophila)	EGFP	53	
	NEO ^R	75	
Gypsy-TRAP	Tissues (Drosophila)	EGFP	84, 86, 87, 89, 90
CLEVR	Cell line (Drosophila)	EGFP-mCherry	92, 93
	Tissues (Drosophila)	EGFP-mCherry	92, 95

EGFP, enhanced green fluorescent protein, NEO^R, Neomycin resistant gene; CHO, Chinese hamster ovarian cell; TEM1, beta-lactamase