

## Invited Mini Review

## Ongoing endeavors to detect mobilization of transposable elements

Yujeong Lee<sup>#</sup>, Una Ha<sup>#</sup> & Sungjin Moon<sup>\*</sup>

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

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. [BMB Reports 2022; 55(7): 305-315]

## 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

\*Corresponding author. Tel: +82-33-250-8536; Fax: +82-33-259-5665; E-mail: sungjin.moon@kangwon.ac.kr

<sup>#</sup>These authors contributed equally to this work.

<https://doi.org/10.5483/BMBRep.2022.55.7.088>

Received 23 April 2022, Revised 26 May 2022,  
Accepted 14 June 2022

**Keywords:** *In vivo* retrotransposition assay, Oxford nanopore technology (ONT), Transposable elements, Transposition

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.

## 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

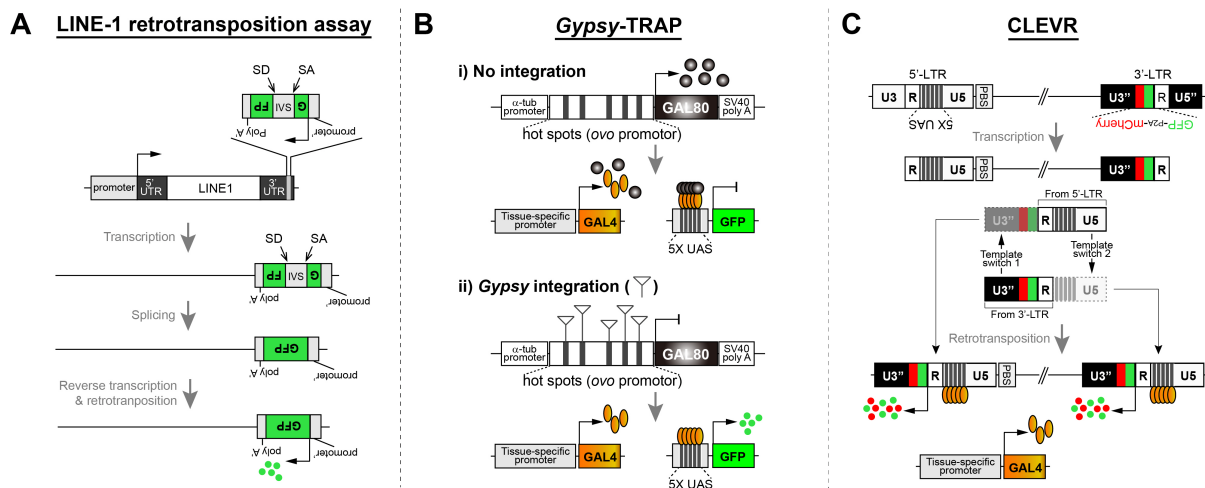
births) and SVA (1 out of 900 births), it is not peculiar to see that 124 cases of disorders are associated with retrotransposition events (5) and that altered gene expression profile by retrotransposition could transform cells to be oncogenic (44-46). Besides of a role as an internal mutagen, sequence 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



**Fig. 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  $\alpha$ -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.

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 l-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,  $\alpha$ -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*

**Table 1.** Retrotransposition assays to monitor transposition events

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

EGFP: enhanced green fluorescent protein; NEO<sup>R</sup>: Neomycin resistant gene; CHO: Chinese hamster ovarian cell; TEM1: beta-lactamase.

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<sup>Intron5</sup> 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 glia-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 *ddm1* 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 spatio-temporal 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 (2020 R1C1C1007050) and by 2020 Research Grant from Kangwon National University (520200047) given to SM.

## CONFLICTS OF INTEREST

The authors have no conflicting interests.

## REFERENCES

1. Mc CB (1950) The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A* 36, 344-355
2. Kazazian HH Jr and Moran JV (2017) Mobile DNA in health and disease. *N Engl J Med* 377, 361-370
3. Richardson SR, Doucet AJ, Kopera HC, Moldovan JB, Garcia-Perez JL and Moran JV (2015) The Influence of LINE-1 and SINE retrotransposons on mammalian genomes. *Microbiol Spectr* 3, MDNA3-0061-2014
4. Cordaux R and Batzer MA (2009) The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 10, 691-703
5. Hancks DC and Kazazian HH Jr (2016) Roles for retrotransposon insertions in human disease. *Mob DNA* 7, 9
6. Ostertag EM and Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35, 501-538
7. Ewing AD and Kazazian HH Jr (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res* 20, 1262-1270
8. Mager DL and Stoye JP (2015) Mammalian endogenous retroviruses. *Microbiol Spectr* 3, MDNA3-0009-2014
9. Tarlinton RE, Meers J and Young PR (2006) Retroviral invasion of the koala genome. *Nature* 442, 79-81
10. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860-921
11. Clark JB and Kidwell MG (1997) A phylogenetic perspective on P transposable element evolution in *Drosophila*. *Proc Natl Acad Sci U S A* 94, 11428-11433
12. Khurana JS, Wang J, Xu J et al (2011) Adaptation to P element transposon invasion in *Drosophila melanogaster*. *Cell* 147, 1551-1563
13. Moon S, Cassani M, Lin YA, Wang L, Dou K and Zhang ZZ (2018) A robust transposon-endogenizing response from germline stem cells. *Dev Cell* 47, 660-671 e663
14. Kofler R, Senti KA, Nolte V, Tobler R and Schlotterer C (2018) Molecular dissection of a natural transposable element invasion. *Genome Res* 28, 824-835
15. Kofler R, Hill T, Nolte V, Betancourt AJ and Schlotterer C (2015) The recent invasion of natural *Drosophila simulans* populations by the P-element. *Proc Natl Acad Sci U S A* 112, 6659-6663
16. Hill T, Schlotterer C and Betancourt AJ (2016) Hybrid dysgenesis in *Drosophila simulans* associated with a rapid invasion of the P-element. *PLoS Genet* 12, e1005920
17. Bellen HJ, Levis RW, He Y et al (2011) The *Drosophila* gene disruption project: progress using transposons with distinctive site specificities. *Genetics* 188, 731-743
18. Spradling AC, Bellen HJ and Hoskins RA (2011) *Drosophila* P elements preferentially transpose to replication origins. *Proc Natl Acad Sci U S A* 108, 15948-15953
19. Eickbush TH and Eickbush DG (2015) Integration, regulation, and long-term stability of R2 retrotransposons. *Microbiol Spectr* 3, MDNA3-0011-2014
20. Huang CR, Burns KH and Boeke JD (2012) Active transposition in genomes. *Annu Rev Genet* 46, 651-675
21. Sotero-Caio CG, Platt RN 2nd, Suh A and Ray DA (2017) Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol Evol* 9, 161-177
22. Kissinger JC and DeBarry J (2011) Genome cartography: charting the apicomplexan genome. *Trends Parasitol* 27, 345-354
23. Van't Hof AE, Campagne P, Rigden DJ et al (2016) The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102-105
24. Lisch D (2013) How important are transposons for plant evolution? *Nat Rev Genet* 14, 49-61
25. Studer A, Zhao Q, Ross-Ibarra J and Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nat Genet* 43, 1160-1163
26. Rebollo R, Farivar S and Mager DL (2012) C-GATE - catalogue of genes affected by transposable elements. *Mob DNA* 3, 9
27. de Souza FS, Franchini LF and Rubinstein M (2013) Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong? *Mol Biol Evol* 30, 1239-1251
28. Sundaram V, Cheng Y, Ma Z et al (2014) Widespread contribution of transposable elements to the innovation

- of gene regulatory networks. *Genome Res* 24, 1963-1976
29. Notwell JH, Chung T, Heavner W and Bejerano G (2015) A family of transposable elements co-opted into developmental enhancers in the mouse neocortex. *Nat Commun* 6, 6644
  30. Lynch VJ, Leclerc RD, May G and Wagner GP (2011) Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 43, 1154-1159
  31. Kapitonov VV and Jurka J (2005) RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. *PLoS Biol* 3, e181
  32. Huang S, Tao X, Yuan S et al (2016) Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell* 166, 102-114
  33. Traverse KL and Pardue ML (1988) A spontaneously opened ring chromosome of *Drosophila melanogaster* has acquired He-T DNA sequences at both new telomeres. *Proc Natl Acad Sci U S A* 85, 8116-8120
  34. Levis RW, Ganesan R, Houtchens K, Tolar LA and Sheen FM (1993) Transposons in place of telomeric repeats at a *Drosophila* telomere. *Cell* 75, 1083-1093
  35. Abad JP, De Pablos B, Osoegawa K, De Jong PJ, Martin-Gallardo A and Villasante A (2004) TAHRE, a novel telomeric retrotransposon from *Drosophila melanogaster*, reveals the origin of *Drosophila* telomeres. *Mol Biol Evol* 21, 1620-1624
  36. Villasante A, Abad JP, Planello R, Mendez-Lago M, Celniker SE and de Pablos B (2007) *Drosophila* telomeric retrotransposons derived from an ancestral element that was recruited to replace telomerase. *Genome Res* 17, 1909-1918
  37. Agudo M, Losada A, Abad JP, Pimpinelli S, Ripoll P and Villasante A (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 Res* 27, 3318-3324
  38. Chang CH, Chavan A, Palladino J et al (2019) Islands of retroelements are major components of *Drosophila* centromeres. *PLoS Biol* 17, e3000241
  39. Neumann P, Navratilova A, Koblikova A et al (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA* 2, 4
  40. Glockner G and Heidel AJ (2009) Centromere sequence and dynamics in *Dictyostelium discoideum*. *Nucleic Acids Res* 37, 1809-1816
  41. Ferreri GC, Brown JD, Oberfell C et al (2011) Recent amplification of the kangaroo endogenous retrovirus, KERV, limited to the centromere. *J Virol* 85, 4761-4771
  42. Rad R, Rad L, Wang W et al (2010) PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science* 330, 1104-1107
  43. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG and Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 436, 221-226
  44. Lee E, Iskow R, Yang L et al (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337, 967-971
  45. Babaian A and Mager DL (2016) Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* 7, 24
  46. Shukla R, Upton KR, Munoz-Lopez M et al (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101-111
  47. Bailey JA, Liu G and Eichler EE (2003) An Alu transposition model for the origin and expansion of human segmental duplications. *Am J Hum Genet* 73, 823-834
  48. Montgomery EA, Huang SM, Langley CH and Judd BH (1991) Chromosome rearrangement by ectopic recombination in *Drosophila melanogaster*: genome structure and evolution. *Genetics* 129, 1085-1098
  49. Deininger PL and Batzer MA (1999) Alu repeats and human disease. *Mol Genet Metab* 67, 183-193
  50. Deniz O, Frost JM and Branco MR (2019) Regulation of transposable elements by DNA modifications. *Nat Rev Genet* 20, 417-431
  51. Ozata DM, Gainetdinov I, Zoch A, O'Carroll D and Zamore PD (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat Rev Genet* 20, 89-108
  52. Yang P, Wang Y and Macfarlan TS (2017) The role of KRAB-ZFPs in transposable element repression and mammalian evolution. *Trends Genet* 33, 871-881
  53. Wang L, Dou K, Moon S, Tan FJ and Zhang ZZ (2018) Hijacking oogenesis enables massive propagation of LINE and retroviral transposons. *Cell* 174, 1082-1094 e1012
  54. Nagimaja L, Morup N, Nielsen JE et al (2021) Variant PNLDC1, defective piRNA processing, and azoospermia. *N Engl J Med* 385, 707-719
  55. Houwing S, Kamminga LM, Berezikov E et al (2007) A role for Piwi and piRNAs in germ cell maintenance and transposon silencing in Zebrafish. *Cell* 129, 69-82
  56. Carmell MA, Girard A, van de Kant HJ et al (2007) MIWI2 is essential for spermatogenesis and repression of transposons in the mouse male germline. *Dev Cell* 12, 503-514
  57. Bourc'his D and Bestor TH (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 431, 96-99
  58. Siudeja K, van den Beek M, Riddiford N et al (2021) Unraveling the features of somatic transposition in the *Drosophila* intestine. *EMBO J* 40, e106388
  59. Brouha B, Schustak J, Badge RM et al (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 100, 5280-5285
  60. Beck CR, Collier P, Macfarlane C et al (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159-1170
  61. Hancks DC, Goodier JL, Mandal PK, Cheung LE and Kazazian HH Jr (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum Mol Genet* 20, 3386-3400
  62. Raiz J, Damert A, Chira S et al (2012) The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. *Nucleic Acids Res* 40, 1666-1683
  63. Boeke JD, Garfinkel DJ, Styles CA and Fink GR (1985) Ty elements transpose through an RNA intermediate. *Cell* 40, 491-500
  64. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD and Kazazian HH Jr (1996) High frequency retrotrans-



- position in cultured mammalian cells. *Cell* 87, 917-927
65. Ostertag EM, Prak ET, DeBerardinis RJ, Moran JV and Kazazian HH Jr (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res* 28, 1418-1423
  66. Xie Y, Rosser JM, Thompson TL, Boeke JD and An W (2011) Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res* 39, e16
  67. Kannan M, Li J, Fritz SE et al (2017) Dynamic silencing of somatic L1 retrotransposon insertions reflects the developmental and cellular contexts of their genomic integration. *Mob DNA* 8, 8
  68. Kopera HC, Larson PA, Moldovan JB, Richardson SR, Liu Y and Moran JV (2016) LINE-1 cultured cell retrotransposition assay. *Methods Mol Biol* 1400, 139-156
  69. Wei W, Morrish TA, Alisch RS and Moran JV (2000) A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal Biochem* 284, 435-438
  70. Morrish TA, Gilbert N, Myers JS et al (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat Genet* 31, 159-165
  71. Feng Q, Moran JV, Kazazian HH Jr and Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905-916
  72. An W, Han JS, Wheelan SJ et al (2006) Active retrotransposition by a synthetic L1 element in mice. *Proc Natl Acad Sci U S A* 103, 18662-18667
  73. Muotri AR, Chu VT, Marchetto MC, Deng W, Moran JV and Gage FH (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903-910
  74. Newkirk SJ, Lee S, Grandi FC et al (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc Natl Acad Sci U S A* 114, E5635-E5644
  75. Jensen S and Heidmann T (1991) An indicator gene for detection of germline retrotransposition in transgenic *Drosophila* demonstrates RNA-mediated transposition of the LINE I element. *EMBO J* 10, 1927-1937
  76. Sultana T, van Essen D, Siol O et al (2019) The landscape of L1 retrotransposons in the human genome is shaped by pre-insertion sequence biases and post-insertion selection. *Mol Cell* 74, 555-570 e557
  77. Flasch DA, Macia A, Sanchez L et al (2019) Genome-wide de novo L1 retrotransposition connects endonuclease activity with replication. *Cell* 177, 837-851 e828
  78. Marlor RL, Parkhurst SM and Corces VG (1986) The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol Cell Biol* 6, 1129-1134
  79. Kim A, Terzian C, Santamaria P, Pelisson A, Purd'homme N and Bucheton A (1994) Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A* 91, 1285-1289
  80. Song SU, Gerasimova T, Kurkulos M, Boeke JD and Corces VG (1994) An env-like protein encoded by a *Drosophila* retroelement: evidence that gypsy is an infectious retrovirus. *Genes Dev* 8, 2046-2057
  81. Mevel-Ninio M, Mariol MC and 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. *EMBO J* 8, 1549-1558
  82. Dej KJ, Gerasimova T, Corces VG and Boeke JD (1998) A hotspot for the *Drosophila* gypsy retroelement in the ovo locus. *Nucleic Acids Res* 26, 4019-4025
  83. Labrador M, Sha K, Li A and Corces VG (2008) Insulator and Ovo proteins determine the frequency and specificity of insertion of the gypsy retrotransposon in *Drosophila melanogaster*. *Genetics* 180, 1367-1378
  84. Li W, Prazak L, Chatterjee N et al (2013) Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat Neurosci* 16, 529-531
  85. Duffy JB (2002) GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* 34, 1-15
  86. Jones BC, Wood JG, Chang C et al (2016) A somatic piRNA pathway in the *Drosophila* fat body ensures metabolic homeostasis and normal lifespan. *Nat Commun* 7, 13856
  87. Wood JG, Jones BC, Jiang N et al (2016) Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in *Drosophila*. *Proc Natl Acad Sci U S A* 113, 11277-11282
  88. Sousa-Victor P, Ayyaz A, Hayashi R et al (2017) Piwi is required to limit exhaustion of aging somatic stem cells. *Cell Rep* 20, 2527-2537
  89. Sun W, Samimi H, Gamez M, Zare H and Frost B (2018) Pathogenic tau-induced piRNA depletion promotes neuronal death through transposable element dysregulation in neurodegenerative tauopathies. *Nat Neurosci* 21, 1038-1048
  90. Fort-Aznar L, Ugboode C and Sweeney ST (2020) Retrovirus reactivation in CHMP2B/Intron5 models of frontotemporal dementia. *Hum Mol Genet* 29, 2637-2646
  91. Penke TJ, McKay DJ, Strahl BD, Matera AG and Duronio RJ (2016) Direct interrogation of the role of H3K9 in metazoan heterochromatin function. *Genes Dev* 30, 1866-1880
  92. Chang YH, Keegan RM, Prazak L and 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 Biol* 17, e3000278
  93. Keegan RM, Talbot LR, Chang YH, Metzger MJ and Dubnau J (2021) Intercellular viral spread and intracellular transposition of *Drosophila* gypsy. *PLoS Genet* 17, e1009535
  94. Song SU, Kurkulos M, Boeke JD and 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, 2789-2798
  95. Chang YH and Dubnau J (2019) The gypsy endogenous retrovirus drives non-cell-autonomous propagation in a *Drosophila* TDP-43 model of neurodegeneration. *Curr Biol* 29, 3135-3152 e3134
  96. Wensink PC, Tabata S and Pacht C (1979) The clustered and scrambled arrangement of moderately repetitive elements in *Drosophila* DNA. *Cell* 18, 1231-1246

97. Potter S, Truett M, Phillips M and Maher A (1980) Eucaryotic transposable genetic elements with inverted terminal repeats. *Cell* 20, 639-647
98. Biemont C, Ronsseray S, Anxolabehere D, Izaabel H and Gautier C (1990) Localization of P elements, copy number regulation, and cytotype determination in *Drosophila melanogaster*. *Genet Res* 56, 3-14
99. Badge RM, Alisch RS and Moran JV (2003) ATLAS: a system to selectively identify human-specific L1 insertions. *Am J Hum Genet* 72, 823-838
100. Sheen FM, Sherry ST, Risch GM et al (2000) Reading between the LINES: human genomic variation induced by LINE-1 retrotransposition. *Genome Res* 10, 1496-1508
101. Pornthanakasem W and Mutirangura A (2004) LINE-1 insertion dimorphisms identification by PCR. *Biotechniques* 37, 750, 752
102. Kim EY, Fan W and Cho J (2021) Determination of TE insertion positions using transposon display. *Methods Mol Biol* 2250, 115-121
103. Goerner-Potvin P and Bourque G (2018) Computational tools to unmask transposable elements. *Nat Rev Genet* 19, 688-704
104. Tubio JMC, Li Y, Ju YS et al (2014) Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343
105. Treiber CD and Waddell S (2017) Resolving the prevalence of somatic transposition in *Drosophila*. *Elife* 6, e28297
106. Kim J, Hu C, Moufawad El Achkar C et al (2019) Patient-customized oligonucleotide therapy for a rare genetic disease. *N Engl J Med* 381, 1644-1652
107. van Dijk EL, Jaszczyszyn Y, Naquin D and Thermes C (2018) The third revolution in sequencing technology. *Trends Genet* 34, 666-681
108. Wang Y, Zhao Y, Bollas A, Wang Y and Au KF (2021) Nanopore sequencing technology, bioinformatics and applications. *Nat Biotechnol* 39, 1348-1365
109. Logsdon GA, Vollger MR and Eichler EE (2020) Long-read human genome sequencing and its applications. *Nat Rev Genet* 21, 597-614
110. Kasianowicz JJ, Brandin E, Branton D and Deamer DW (1996) Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* 93, 13770-13773
111. Song L, Hobaugh MR, Shustak C, Cheley S, Bayley H and Gouaux JE (1996) Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* 274, 1859-1866
112. Jain M, Olsen HE, Paten B and Akeson M (2016) The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17, 239
113. Manrao EA, Derrington IM, Laszlo AH et al (2012) Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. *Nat Biotechnol* 30, 349-353
114. Cherf GM, Lieberman KR, Rashid H, Lam CE, Karplus K and Akeson M (2012) Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. *Nat Biotechnol* 30, 344-348
115. Amarasingham SL, Ritchie ME and Gouil Q (2021) long-read-tools.org: an interactive catalogue of analysis methods for long-read sequencing data. *Gigascience* 10, giab003
116. Gong L, Wong CH, Cheng WC et al (2018) Picky comprehensively detects high-resolution structural variants in nanopore long reads. *Nat Methods* 15, 455-460
117. Goodwin S, Gurtowski J, Ethe-Sayers S, Deshpande P, Schatz MC and McCombie WR (2015) Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome Res* 25, 1750-1756
118. David M, Dursi LJ, Yao D, Boutros PC and Simpson JT (2017) Nanocall: an open source basecaller for Oxford Nanopore sequencing data. *Bioinformatics* 33, 49-55
119. Boza V, Brejova B and Vinar T (2017) DeepNano: deep recurrent neural networks for base calling in MinION nanopore reads. *PLoS One* 12, e0178751
120. Gong L, Wong CH, Idol J, Ngan CY and Wei CL (2019) Ultra-long read sequencing for whole genomic DNA analysis. *J Vis Exp* 145, e58954
121. Branton D and Deamer DW (2018) Nanopore sequencing: an introduction, World Scientific, New Jersey.
122. De Coster W, De Rijk P, De Roeck A et al (2019) Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. *Genome Res* 29, 1178-1187
123. De Roeck A, De Coster W, Bossaerts L et al (2019) NanoSatellite: accurate characterization of expanded tandem repeat length and sequence through whole genome long-read sequencing on PromethION. *Genome Biol* 20, 239
124. Kim HS, Jeon S, Kim C et al (2019) Chromosome-scale assembly comparison of the Korean Reference Genome KOREF from PromethION and PacBio with Hi-C mapping information. *Gigascience* 8, giz125
125. Nicholls SM, Quick JC, Tang S and Loman NJ (2019) Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *Gigascience* 8, giz043
126. Payne A, Holmes N, Rakyen V and Loose M (2019) BulkVis: a graphical viewer for Oxford nanopore bulk FAST5 files. *Bioinformatics* 35, 2193-2198
127. Jain M, Koren S, Miga KH et al (2018) Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* 36, 338-345
128. Solares EA, Chakraborty M, Miller DE et al (2018) Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. *G3 (Bethesda)* 8, 3143-3154
129. Tyson JR, O'Neil NJ, Jain M, Olsen HE, Hieter P and Snutch TP (2018) MinION-based long-read sequencing and assembly extends the *Caenorhabditis elegans* reference genome. *Genome Res* 28, 266-274
130. Jain M, Olsen HE, Turner DJ et al (2018) Linear assembly of a human centromere on the Y chromosome. *Nat Biotechnol* 36, 321-323
131. Miga KH, Koren S, Rhie A et al (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature* 585, 79-84
132. Michael TP, Jupe F, Bemm F et al (2018) High contiguity *Arabidopsis thaliana* genome assembly with a single nanopore flow cell. *Nat Commun* 9, 541

133. Chernyavskaya Y, Zhang X, Liu J and Blackburn J (2022) Long-read sequencing of the zebrafish genome reorganizes genomic architecture. *BMC Genomics* 23, 116
134. Garg S, Functammasan A, Carroll A et al (2021) Chromosome-scale, haplotype-resolved assembly of human genomes. *Nat Biotechnol* 39, 309-312
135. Kirov I, Merkulov P, Dudnikov M et al (2021) Transposons hidden in *Arabidopsis thaliana* genome assembly gaps and mobilization of non-autonomous LTR retrotransposons unravelled by nanotei pipeline. *Plants (Basel)* 10, 2681
136. Mohamed M, Dang NT, Ogyama Y 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, 1776
137. Miller DE, Staber C, Zeitlinger J and Hawley RS (2018) Highly contiguous genome assemblies of 15 *Drosophila* species generated using Nanopore sequencing. *G3 (Bethesda)* 8, 3131-3141
138. Pradhan B, Cajuso T, Katainen R et al (2017) Detection of subclonal L1 transductions in colorectal cancer by long-distance inverse-PCR and Nanopore sequencing. *Sci Rep* 7, 14521
139. Fujimoto A, Wong JH, Yoshii Y 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 Med* 13, 65
140. Ellison CE and 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 Res* 48, 290-303
141. Siudeja K, van den Beek M, Riddiford N et al (2021) Unraveling the features of somatic transposition in the *Drosophila* intestine. *EMBO J* 40, e106388
142. Frommer M, McDonald LE, Millar DS et al (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc Natl Acad Sci U S A* 89, 1827-1831
143. Ewing AD, Smits N, Sanchez-Luque FJ et al (2020) Nanopore sequencing enables comprehensive transposable element epigenomic profiling. *Mol Cell* 80, 915-928 e915
144. Haggerty C, Kretzmer H, Riemenschneider C et al (2021) Dnmt1 has de novo activity targeted to transposable elements. *Nat Struct Mol Biol* 28, 594-603
145. Jiang F, Zhang J, Liu Q 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 Biol* 16, 950-959
146. Kirov I, Dudnikov M, Merkulov P et al (2020) Nanopore RNA sequencing revealed long non-coding and LTR retrotransposon-related RNAs expressed at early stages of triticale SEED development. *Plants (Basel)* 9, 1794
147. Lee SC, Ernst E, Berube B et al (2020) *Arabidopsis* retrotransposon virus-like particles and their regulation by epigenetically activated small RNA. *Genome Res* 30, 576-588
148. Panda K and Slotkin RK (2020) Long-read cDNA sequencing enables a "gene-like" transcript annotation of transposable elements. *Plant Cell* 32, 2687-2698
149. Berrens RV, Yang A, Laumer CE et al (2021) Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat Biotechnol* 40, 546-554
150. Maringer K, Yousuf A, Heesom KJ et al (2017) Proteomics informed by transcriptomics for characterising active transposable elements and genome annotation in *Aedes aegypti*. *BMC Genomics* 18, 101