

Review Article

Technology to the rescue: how to uncover the role of transposable elements in preimplantation development

Lauryn A. Deaville^{1,2,3} and  Rebecca V. Berrens^{1,2}

¹Institute for Developmental and Regenerative Medicine, Oxford University, IMS-Tetsuya Nakamura Building, Old Road Campus, Roosevelt Dr, Oxford OX3 7TY, U.K.; ²Department of Paediatrics, Oxford University, Level 2, Children's Hospital, John Radcliffe Headington, Oxford OX3 9DU, U.K.; ³MRC Weatherall Institute of Molecular Medicine, Oxford University, John Radcliffe Hospital, Oxford OX3 9DS, U.K.

Correspondence: Rebecca V. Berrens (rebecca.berrens@idrm.ox.ac.uk)



Transposable elements (TEs) are highly expressed in preimplantation development. Preimplantation development is the phase when the cells of the early embryo undergo the first cell fate choice and change from being totipotent to pluripotent. A range of studies have advanced our understanding of TEs in preimplantation, as well as their epigenetic regulation and functional roles. However, many questions remain about the implications of TE expression during early development. Challenges originate first due to the abundance of TEs in the genome, and second because of the limited cell numbers in preimplantation. Here we review the most recent technological advancements promising to shed light onto the role of TEs in preimplantation development. We explore novel avenues to identify genomic TE insertions and improve our understanding of the regulatory mechanisms and roles of TEs and their RNA and protein products during early development.

Introduction

During preimplantation development, the zygote undergoes meticulously regulated cell divisions accompanied by global epigenetic reprogramming and activation of the zygotic genome (reviewed in [1]). Simultaneously, significant changes in gene expression facilitate lineage specification, forming both the inner cell mass and the trophoblast tissue (reviewed in [2]). Lineage specification is essential for implantation and viability of the developing organism. Cells produced during preimplantation contribute to the development of all differentiated cell types in an organism. The expression of transposable elements (TE) plays an important role in preimplantation development, and the intricate relationship between TE expression and this phase of early development has been extensively reviewed in [3–5] (Figure 1). However, repetitive genetics and low cell-numbers present unique problems to researchers studying TEs in preimplantation.

TEs are mobile genetic elements, a quirk which has resulted in hundreds of thousands of copies of TEs with near-identical sequences. Over time, these sequences acquire mutations, diversifying loci. Therefore, with sufficient sequencing depth, many older families of TEs can be studied using canonical short-read sequencing technologies [7]. However, the lack of sequence diversity in younger TEs makes only ~10% of the youngest loci mappable [7,8]. This bias in TE research often underestimates the activity and roles of young, retrotransposition active TEs [8].

Obtaining embryo samples from both human and mouse is difficult, with each approach imposing different challenges on researchers. Human studies are halted both by ethical constraints, and embryo availability and quality. For mouse work, technical challenges impose the main roadblock. Specifically, throughput is limited to tens of embryos per mouse, and isolation, culture, and experimentation steps result in losses to this already minimal material. Due to these limitations many existing techniques are

Received: 14 February 2024
Revised: 23 April 2024
Accepted: 24 April 2024

Version of Record published:
16 May 2024

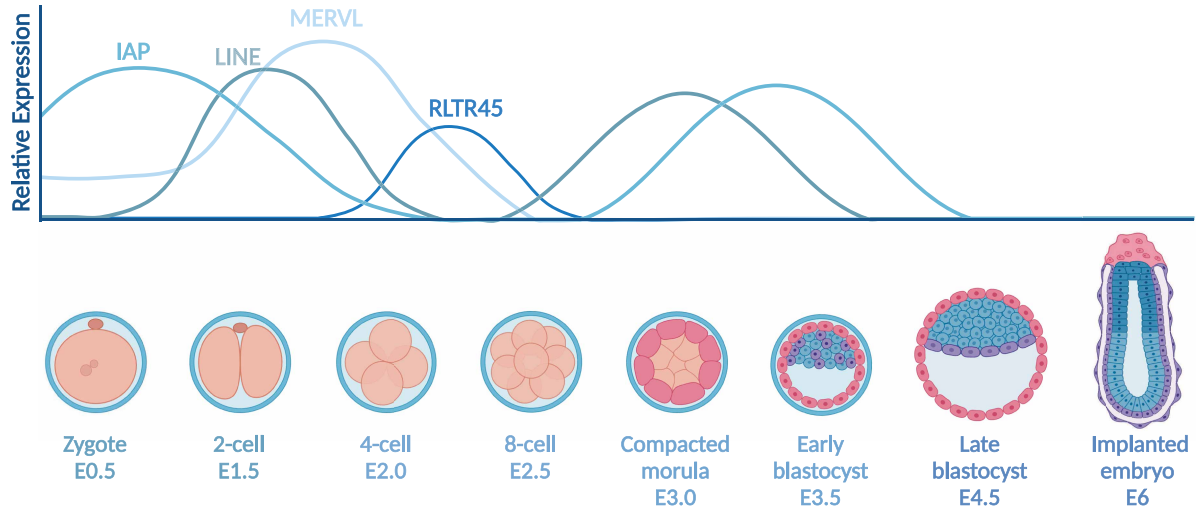


Figure 1. Schematic of TEs expressed at stages of preimplantation development. Adapted from [5,6]. Created with BioRender.com

only available to use in mouse or human pluripotent embryonic stem cells (ESCs), derived from the inner cell mass of the blastocyst. ESCs can be cultured in conditions mimicking various stages of embryo development, and advancements to 3D stem-cell culture and embryo approaches are beginning to provide more diverse embryo models for both human and mouse [9,10]. Using these systems enables the application of existing high-throughput techniques to the study of preimplantation, which have previously been inaccessible due to cell-number constraints. Here, we review existing and novel methodologies with potential to answer open questions surrounding TEs in preimplantation development (Figure 2).

Exploring embryo alternatives

While many existing technologies to study genetics are now able to be applied to the study of TEs, the uniquely low cell numbers in preimplantation development mean many questions remain unanswered. However, novel *in vitro* embryo and organoid culture systems will enable exploration of developmental processes and genetic mechanisms without the limitations of traditional mouse models. By eliminating the uterine barrier, *ex vivo* systems enable mechanistic interrogation of pre- and post-implantation embryo development in mammals [15–17]. For example, *ex utero* manipulation enables injection of CRISPR proteins directly into zygotes to observe effects of genetic knockouts (KOs) or knockdown (KD) [18]. Such culture systems have already been harnessed to tackle questions in symmetry breaking and germ-layer specification in mice [15,16]. Similarly, human embryonic-like structures called blastoids can be derived from human ESCs, enabling dissection of early human development in the absence of maternal tissues [19,20]. Further optimization and characterization of these systems is required to fully validate the use of gastruloids and blastoids as a model system for preimplantation embryos. However, with the help of *in vivo* validation, these *ex vivo* systems provide a valuable opportunity to conduct higher input experiments and functional screens of TEs in preimplantation; which we will explore in this review.

What can we learn from looking closer at TE DNA sequences?

Unlike in somatic cells, novel TE insertions arising during preimplantation have the potential to propagate through the germline and contribute to heritable genetic variation in subsequent generations. Their intrinsic drive for self-preservation is thought to promote TE expression prior to primordial germ cell (PGC) specification [21]. Furthermore, TE activity in oogenesis and spermatogenesis can result in insertions carried through to the zygote [22,23]. High-fidelity genome assemblies are crucial for identifying such transposon insertions. This is especially important with regards to younger TEs. Young TEs exhibit high sequence similarity as they remain retrotransposition active, creating novel insertions in different populations, mouse strains, and cell

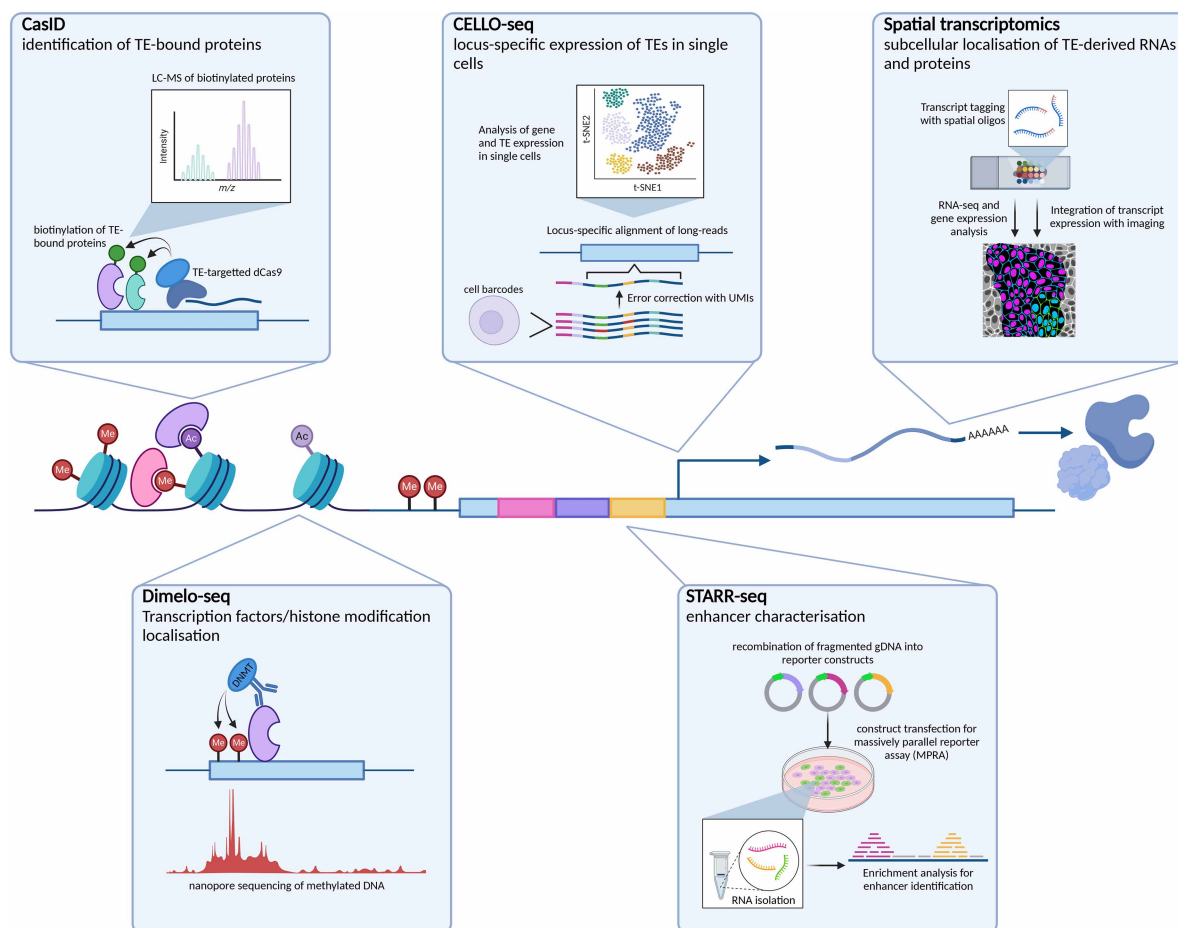


Figure 2. Technology for Transposable elements – Schematic of key technologies for the study various aspects of TE and preimplantation biology.

CasID: Cas9-targeted biotinylation of TE-bound proteins, followed by streptavidin pulldown and Liquid Chromatography-Mass Spectrometry (LC-MS) for protein identification [11]. CELLOseq: addition of cellular barcodes and unique molecular identifiers (UMIs) enables single-cell long-read RNA sequencing for unique mapping of TE-derived transcripts [8]. Spatial Transcriptomics: hybridisation of unique oligos across the sample prior to imaging and RNA isolation enables coupling of imaging with transcriptomic analysis [12]. DiMeLo-seq: Antibody-targeted DNA methyltransferases result in enrichment of m6dA over protein-bound regions which can be resolved with ONT direct DNA sequencing [13]. STARR-seq: genome fragmentation and generation of many reporter constructs is followed by RNA extraction and enrichment analysis for enhancer identification [14]. Created with BioRender.com.

lineages [24–26]. A recent comparison identified strain-specific variations in retrotransposon abundance and activity in mice [27]. For example, long interspersed nuclear element 1 (LINE1) are more abundant in the CB57BL/6 compared with the 129 strain of mice. Despite these variations, many publications continue to use CB57BL/6-derived genome assemblies to study TEs in 129-derived mouse ESCs [28]. This may result in wrongly assigned TEs causing inappropriate assumptions about genomic context, and ultimately roles of TEs. Recent advancements in long-read sequencing technologies, such as PacBio and Oxford Nanopore Technology (ONT) sequencing, have significantly enhanced genome assemblies, particularly in repetitive regions, for example the X chromosome [29,30]. Further advantages for identifying TE insertions might arise through ONTs adaptive sampling. Adaptive sampling allows real-time rejection of undesired sequences by reversing polarity in the nanopore [31,32]. Selection of molecules based on genomic co-ordinates may allow higher fidelity reconstructions of full-length TEs. Although whole genome sequencing technologies can require input of ~100 000 cells, researchers should consider using *in vitro* models to generate their own genome assemblies to

study TEs [27,33]. For example, recent telomere to telomere assemblies may give important insights into sex specific roles of TE insertions in early development [29].

Following identification of TE insertions, exploration of the DNA sequence allows elements to be classified based on their mechanism of transposition, evolutionary age, and sequence similarity (reviewed in [34]). LINE1s for example, can be divided based on the presence of different monomer sequences within the promoter [35,36]. A recent study found that changes in monomer structure of LINE1s can influence nearby reporter gene expression [37]. While we know that LINE1s are expressed in preimplantation development, analysis of monomer structures in expressed LINE1 promoters may decode the mechanism of locus-specific expression. However, expression does not guarantee retrotransposition. Characterizing retrotransposition activity of younger insertions necessitates retrotransposition assays, currently confined to *in vitro* conditions using mouse ESCs [31,32,38]. The exploration of alternative culturing systems or embryos may hold potential for expanding the scope of these assays [39–41].

Many TE families are also predicted to possess enhancer activity; promoting transcription of their own RNA as well as other genes [42–44]. The multifaceted roles of TE-derived enhancers have been reviewed extensively in [45]. However, an open question is whether specific sequences within TEs predispose TE loci to function as cell-type specific enhancers compared with other loci. Existing methods like STARR-seq can be used to test whether a minimal core sequence of a TE loci is sufficient to act as an enhancer [14]. STARR-seq enables a broad-scale examination of enhancers in the genome, identifying regulatory motifs through reporter gene expression. STARR-seq has been previously used to identify bona fide TE enhancers in cancers [46] and human ESCs [47] and, as *in vitro* models improve may facilitate identification of minimal sequence components of TE loci that control developmental gene regulatory networks throughout the course of development, where use of this technique has been limited by low cell numbers.

How do TEs interact with their chromatin environment?

TE enhancers mediate genomic long-range regulatory interactions. Studies illuminated how expression of enhancer RNA and upstream antisense RNA from complementary Alu sequences can modulate enhancer-promoter interactions of oncogenes, underscoring the profound regulatory potential of TEs in chromatin dynamics [48]. Our understanding of TE-mediated chromatin interactions in embryos remains limited, and the influence of TEs on chromatin architecture during preimplantation is still discussed. For example, LINE1 RNA itself seems to be able to influence epigenetic remodeling through recruitment of NUCLEOLIN and KAP1 [49]. Adaptation of high-resolution capture techniques like Micro Capture-C could allow investigation of interactions of individual loci [50]. For example, using probes against regions surrounding loci of interest was employed to profile TE interactions in HeLa cells, identifying TE interactions using proximal unique sequences [51]. Using a similar strategy may enable exploration of individual TE-derived *cis*-regulatory elements and their genic targets during preimplantation. Additionally, techniques such as Pore-C, which link several genomic loci together prior to long read sequencing, may enable dissection of complex interactions of TE loci with multiple genomic locations [52]. This could shed light on the role of TEs in the compaction of chromatin in the early embryo [53].

Specific motifs within TEs enable binding of cell-type specific transcription factors, contributing to their activities as regulatory elements for other genes. For example, binding of OCT4, KLF4 and KLF17 to a long terminal repeat (LTR) of a human endogenous retrovirus (ERV) K is thought to contribute to hominoid-specific patterns of embryonic gene expression [54–56]. More recently, 5' UTRs of young LINE1s bound by elongation factor ELL3 were shown to act as enhancers for *Act3*. ACT3 is necessary for activation of a key kinase pathway, providing an interesting link between TE expression and developmental signaling [57]. As well as *trans*-acting factor binding, TE enhancer activity can also be determined by the presence of specific epigenetic modifications. For example, H4K16ac activates transcription of TEs to enable enhancer activity in human ESCs [47]. Studying the role of TEs as enhancers in early development will enable us to understand whether TEs are involved in whole gene regulatory networks or mostly control individual nearby genes.

Next to contributing to their own regulatory functions, epigenetic modifications and binding of *trans*-acting factors also regulates expression of TEs. For example, deposition of the repressive histone mark H3K9me3 on LTRs leads to stage-specific transcriptional silencing of TEs in eight-cell and blastocyst stage human preimplantation embryos [58]. Further examples have shown increased chromatin accessibility and allele-specific histone modifications suggesting transcriptional activity over TE families during preimplantation development [59–61]. Using updated mapping approaches, analysis of locus-specific transcription of TEs may be possible using these

datasets, however insights into young TE families may be limited by short-reads and low coverage. Alongside epigenetic modifications, association of proteins with TE DNA is also known to contribute to their regulation. For example, LINE1s commonly associate with the transcription factor YY1. Interestingly, while in humans, YY1 binding has been shown to silence LINE1 expression; in mice, YY1 binding stimulates expression from LINE1 [60,61]. Alongside YY1, many TEs are known to experience cell-type specific regulation of gene expression by other zinc-finger nucleases (ZNFs) [55,62,63].

Current state-of-the-art techniques for assessing the enrichment of epigenetic modifications and DNA-bound proteins, such as ChIP-seq, are powerful but encounter limitations in *in vivo* work due to the substantial cell numbers required for high-quality data. This limitation poses a challenge, particularly in the context of preimplantation studies where samples are limited. While low-input alternatives such as CUT&RUN or CUT&TAG are improving in quality, the use of short-read sequencing constrains their applicability to exploring protein interactions with TE families [64,65]. Several advanced analysis approaches have enabled robust analysis of TEs using short read data (reviewed extensively in [7]). Building on this, deep learning holds promise for unique mapping of reads from existing short-read experiments. For example, MATES uses adjacent read alignments surrounding the TE locus to allocate multi-mapping reads to specific loci of TEs in single-cell data [66]. MATES may be particularly useful for re-analysis of existing single-cell preimplantation data from [67–70], reducing the costs associated with data generation. Using such algorithms may extract novel information about locus-specific histone modifications over TEs in single-cells. However, analysis approaches are still limited by sequencing depth and, in very old datasets, by read length and quality, encouraging the use of novel long-read sequencing approaches.

Novel technical approaches, such as ONT long-read sequencing, offer a promising solution to limitations in existing genomics methods. Direct sequencing of native DNA allows identification of base modifications, enabling detection of DNA modifications such as 5mC and m6dA. Co-option of DNA modifying enzymes allows investigation of DNA-protein interactions as well as chromatin features over repetitive regions. DiMeLo-seq uses antibodies to direct DNA adenine methyltransferases to proteins of interest. DiMeLo-seq has provided valuable insights into protein dynamics within mammalian centromeres holding promises for TE research [13]. Similarly, DamID uses DNA adenine methyltransferases, however these are targeted via fusion proteins [71,72]. Another example is STAM-seq, which identifies DNA methylation and open chromatin within repetitive regions [73]. While the requirement for a million cells as input limits the use of these techniques *in vivo*, techniques such as DiMeLo-seq and STAM-seq could be adapted to provide insight into locus-specific changes in chromatin modifications and accessibility during pluripotency exit in *in vitro* cultures such as gastruloids, for example to study the zygote to blastocyst transition. With DiMeLo-seq, it is possible to explore protein binding density on a single chromatin fiber from a single cell, enabling exploration of heterogeneity in transcription factor distribution across individual TEs in single cells in the developing embryo.

Many factors associated with TE function in preimplantation remain elusive. An emergence in dead Cas9-based targeting technologies allows researchers to identify novel proteins binding to TE DNA. Via endonuclease-inactive Cas9, gRNAs target modifier proteins to specific TE families to identify novel protein candidates involved in TE regulation or function. Such technologies have already improved understanding of TE DNA-protein interactions. In human ESCs, CARGO-BioID identified the RNA N^6 -methyladenosine (m6A) reader, YTHDC2, to occupy genomic loci of LTR7/HERV-H through its interaction with m6A-modified HERV-H RNAs. YTHDC2 further recruited the DNA demethylase TET1 thereby preventing epigenetic silencing of TE loci which inhibited neural differentiation of human ESCs [74]. A similar technology, CasID, identified TNRC18 as a chromatin-associated regulator of ERVs. Specifically, TNRC18 binds H3K9me3 and recruits co-repressors like HDAC–SIN3–NCoR complexes to silence ERV1 expression [11]. So far, these techniques have been successful in studying TEs at the family-level. Improvements to genome assemblies and gRNA specificity targeting upstream of the promoter of an individual TE locus may enable a locus-specific TE view of protein binding in the future.

Covalent epigenetic modifications to DNA such as 5mC act as a critical repressor of TE expression [36,75–77]. Direct sequencing of native DNA also provides a unique opportunity to explore changes in DNA methylation of TE loci. The ability to also measure 5mC as well as 5hmC simultaneously eliminates the need for the conventionally used bisulfite conversion [78,79]. Long-read sequencing identified variation in *de novo* methylation of LINE1 loci, which may allow transcription of individual loci during mouse ESC differentiation [36]. Furthermore, ONT sequencing identified *de novo* activity of DNA methyltransferase DNMT1 targeted to IAP retrotransposons, which silences IAP expression in mouse ESCs [80]. ONT sequencing also allows to explore

other base modifications, such as the selective oxidation of 5mC to 5hmC. Oxidation of 5mC to 5hmC occurs in the paternal genome during the earliest stages of preimplantation [81,82]. Several studies have explored TET3 activity on LINE1s, however, locus-specific data on 5hmC distribution during preimplantation from direct DNA sequencing could deepen our understanding of the role of 5hmC in TE regulation during early development [83,84]. This may reveal how methyltransferases are targeted to induce changes in DNA methylation at the TE locus level [84].

What else can we learn from TE expression?

While exploring epigenetic changes around TEs is insightful, often the link between chromatin modifications and TE expression is unclear. Many studies focus on the loss of repressive marks during preimplantation, which alone is insufficient to answer outstanding questions of TE dynamics [76,85–88]. For example, a recent study on locus-specific DNA methylation shows not all demethylated loci are expressed in human ESCs [89]. However, after extensive research, it is known that TEs are expressed during specific stages of preimplantation, prior to silencing in most somatic cells [90] (Figure 1). For example, in the mouse genome, elevated transcription of MERVL and MaLR elements is a known marker of zygotic genome activation (ZGA) [91–94]. Expression of ERVs has also been identified in human preimplantation embryos, starting from ZGA at the eight-cell stage [56,90]. Alongside ERVs, the earliest expressed retrotransposons are LINE1s, which are detected first in the zygote. While expression peaks in the two-cell stage, LINE1s remain transcriptionally active throughout embryonic development [95].

Although family-level expression dynamics of TEs are well characterized, the limitations of short-read transcriptomic data and imaging approaches mean it remains a mystery if loci are expressed preferentially, especially younger elements with more similar sequences. Expression of young TEs is often underestimated using only unique mapping reads [8,96]. Computational methods try to overcome this by assigning reads to specific loci depending on how many reads each locus from a subfamily overlaps [97,98]. Several studies have used short-read single-cell data to study TE expression in single cells, either to assess family level expression of TEs [99] or study TE expression in groups of cells in a range of systems [99–101]. For example, combination of mapping approaches with genomic classification has enabled locus-resolved TE expression in bulk RNA-seq data in zebrafish [100]. Single-cell RNA-seq in mouse and human embryos may also reveal allele-specific variation in locus-specific TE expression [100]. Advanced analysis approaches of short-read data ensure accuracy of TE expression, and begin to provide locus-specific and single-cell resolution of TE dynamics. However, the youngest TEs lacking uniquely mappable sequences from short-read data continue to limit the applicability of such approaches.

Alternatively, current advances in long-read RNA-seq methodologies hold great promise for locus-specific expression analysis. PacBio sequencing combined with target enrichment to investigate LINE1 expression, focusing on 5' UTR primers to discern transcription initiation from TE promoters [102]. However, these methods lack single-cell resolution. CELLO-seq, offers a unique advantage by providing long-reads with error-correction, facilitating precise mapping of TE transcripts at the locus level within single cells [8]. Single-cell long-read RNA-seq proves particularly beneficial for examining preimplantation stages, allowing for the exploration of heterogeneity in TE expression within individual cells of the embryo. Potential improvements, such as multiplexing of samples and automation of the pipeline hold promise for enhancing throughput in TE analysis.

LTRs of ERVs also give rise to TE-derived isoforms that are expressed throughout ZGA and long-read single-cell RNA-seq will enable identification of TE-derived isoforms throughout early development [103]. Exploration of chimeric TE-gene transcripts raises questions about their authenticity, potentially being either artifacts of PCR or true structural variations [104,105]. Overlap with known short-read data, utilization of splice sites, and correlation with chromatin data could enhance the identification of genuine TE transcripts [106]. Additionally, investigating native transcription with techniques like TT-seq involving 4-Thiouridine incorporation may distinguish actively transcribed loci from residual nuclear RNA [107]. However, the inability of reverse transcriptases to work with highly structured RNA and biases introduced during PCR amplification highlight the need for alternative methods without need for amplification.

Direct RNA-seq emerges as a reliable method for qualitatively measuring locus-specific TE expression without the need for PCR amplification. While direct RNA sequencing requires high input and high-quality RNA, which is often not feasible from single embryos, this method may be applicable to *in vitro* samples. Direct RNA-seq also enables exploration of the role of RNA modifications in TE RNA activity. Notably, m6A

was shown to mark retrotransposon-derived RNAs in oocytes and preimplantation embryos, including MTA and MERVL during ZGA at the two-cell stage [108]. This modification extends to young LINE1s in later stages, suggesting a role in RNA stability, localization, and translation of these transcripts during the transition from totipotency to pluripotency [108]. However, the protocol remains limited by 3' biases and difficulties in detecting RNA modifications remain. Limitations aside, new flow-cell chemistry together with more data for improved training of base callers may identify a role for RNA modifications in TE-RNA stability and improve understanding of the function of TE RNAs in preimplantation development. To date, the sequence or RNA modifications of currently active LINE1s in the human genome remain unknown. Further investigation is needed into the impact of TE-RNA modifications on early embryo development.

What is the functional role of TE expression?

Various approaches have been used to understand the regulation of TEs, but most are correlative. As with genes, approaches such as KD or KO systems are essential to study function. Large-scale CRISPR screens involving Cas9-mediated KO of epigenetic regulators were able to identify novel protein factors involved in retroelement silencing alongside SETDB1 in mouse ESC [109]. Similar screens identified factors involved in retrotransposition [110,111]. CRISPR screens could prove useful for the study of TE loci themselves, rapidly providing insight into the contributions of many different TEs into gene regulation in mouse ESC. On a smaller scale, KO of an individual LTR10A-ENG in human trophoblast stem cells identified its role as an enhancer for the growth factor receptor ENG. Inappropriate activation of this enhancer is thought to contribute to pre-eclampsia, another example highlighting the importance of TEs in preimplantation [112]. However, KO approaches can cause disruption by removing large parts of the genome. This is a particular issue when studying younger, intact elements that can be over 6 kb in length.

Alternative methods aim to epigenetically silence genomic locations or remove RNAs to recapitulate TE silencing. Epigenetic silencing targeted using transcription activator-like effector (TALE) proteins or by anti-sense oligos showed precise expression of LINE1 at the two-cell stage is involved in progression to the blastocyst stage [49,113]. However, more recently co-option of the CRISPR system (termed CRISPRi) has improved targeting of sequence-specific silencing. CRISPR inactivation of MERVL highlighted that their transcription is essential for preimplantation [114]. While these approaches might be favorable, concerns arise with KD due to spreading of chromatin modifications causing alterations to surrounding genetics [115]. It will be necessary to constrain KD effects and targeting to ensure KD individual loci instead of whole TE families, because individual loci of TE families might overlap with genic regions, and it remains unclear how much of the KD can be attributed specifically to the TE expression.

What are the roles of TE RNA and protein?

Both RNA and protein products of TE expression are thought to contribute to preimplantation development, and several methodologies can be applied to explore their roles. Imaging techniques have been integral to ZGA studies, with single molecule imaging using the MS2 loop system combined with HALO-tag technologies emerge as novel methods to infer transcription dynamics in early embryos [116], reviewed in [117]. Applying this methodology, researchers found cell cycle dependent subcellular organization of LINE1 RNA in HEK293T cells [118]. Application of imaging technology may provide answers to other outstanding questions surrounding TE RNA biology. Both coding and non-coding nascent RNAs, including those produced by retrotransposons, may play an important role in forming transcriptional condensates in development and disease contexts [119]. Such condensation of LINE1 is also identified as a critical factor for retrotransposition [120]. Studying condensates using single molecule imaging technologies in early embryos could give important insights into the timing of retrotransposition events, as well as the role of TE-derived transcriptional condensates in preimplantation [119].

RNA fluorescence *in situ* hybridization (RNA-FISH) has served as a foundational technique in early embryo investigations. The expression of TEs such as IAPs and LINE1s during ZGA, were initially characterized through RNA-FISH methods [95]. Multiplexing of RNA-FISH probes has enabled the development of spatial transcriptomics, employing techniques like MERFISH, 10× Genomics, and NanoString [12,121,122]. Combining TE-specific probes with epigenetic markers and key ZGA transcripts holds great potential for understanding single-cell regulation within developing embryos. However, application of spatial transcriptomics to TE RNAs remains a challenge due to the requirement for robust and specific probes and multiplex readouts.

Table 1. Techniques and their description with limitations and open questions in preimplantation they could answer

Part 1 of 2

Technique	Description	Limitations	Questions to be answered
Direct DNA-seq (ONT)	Long read direct DNA sequencing by Nanopore by sequencing directly isolated DNA. This allows to identify TE insertions as well as characterize their 5mC and 5hmC levels.	$\times 10^5$ cells <i>in vitro</i>	When do de novo TE insertions happen in the genome? Are these loci differentially methylated?
STARR-seq	Self-transcribing active regulatory region sequencing, a method for identifying and mapping active enhancer elements by capturing nascent RNA transcripts synthesized from active regulatory regions.	$\times 10^6$ cells <i>in vitro</i>	Can we identify minimal core sequences in TE subfamilies that act as enhancers at different stages of preimplantation development?
Micro Capture-C	Cross-linking agents fix chromatin structure ahead of proximity ligation. Probe-mediated pull-down is used to capture chromatin interactions for regions of interest. Captured chromatin fragments are sequenced to identify interacting genomic regions.	$\times 10^6$ cells <i>in vitro</i>	Do TE loci exhibit enhancer function important for the first cell fate choice early in two cell stage embryo changing gene expression patterns at eight cell stage?
Pore-C	A chromatin conformation capture technique that ligates multiple contact sites prior to nanopore sequencing to analyze chromatin interactions, enabling the study of complex chromatin folding and organization at high resolution.	$\times 10^6$ cells <i>in vitro</i>	Do TE loci of the same family interact with each other to build enhancer hubs? Are they involved in droplet formation and play a functional role in pluripotency?
MATES	Computational method that uses adjacent read alignments surrounding the TE locus to allocate multi-mapping reads to specific loci of TEs in single-cell data.	<i>In silico</i>	Is there heterogeneity of histone modifications over individual loci of TEs in single cells of the early embryo?
CUT&RUN/ CUT&TAG	CUT&RUN involves targeted cleavage of chromatin-bound proteins using an antibody and micrococcal nuclease (MNase), before sequencing of DNA fragments. CUT&Tag uses an antibody-targeted Tn5 transposase to directly tag the protein-bound DNA before sequencing.	$\times 10^4$ cells <i>in vitro/in vivo</i>	Do specific transcription factors regulate locus-specific TE expression in early embryos?
DiMeLo-seq	Profiling DNA-protein interactions at single-cell resolution by using antibodies to specifically target DNA adenine methylation to protein-bound sites. Nanopore sequencing identifies DNA methylation patterns in individual cells.	$\times 10^6$ cells <i>in vitro</i>	Do individual TEs show accumulation of specific histone modifications during pluripotency exit?
STAM-seq	Tn5-based ATAC-seq with molecular identifiers, a method for profiling chromatin accessibility at single-cell resolution, allowing the study of regulatory elements in individual cells.	<i>In vitro/in vivo</i>	Is accessibility of individual TE loci heterogenous in preimplantation embryos?
CARGO-BioID	A bait protein fused to a promiscuous biotin ligase (BioID) is expressed in cells, and biotinylates neighboring proteins. Biotinylated proteins are captured by streptavidin beads and identified using mass spectrometry.	$\times 10^6$ cells <i>in vitro</i>	Can we identify new interactors of LINE1 RT to understand whether any host proteins are involved in LINE1 retrotransposition in the early embryo?

Continued

Table 1. Techniques and their description with limitations and open questions in preimplantation they could answer

Part 2 of 2

Technique	Description	Limitations	Questions to be answered
CasID	CRISPR-Cas9-targeted labeling of proteins bound to specific DNA sites, enabling the identification of protein-DNA interactions.	$\times 10^6$ cells <i>in vitro</i>	Do different TE loci show accumulation of distinct protein interactors? Does this impact TE expression?
CELLO-seq	Single-cell long-read RNA sequencing, using cellular barcodes and UMIs to enable single cell and single loci resolution of TE expression.	<i>In vivo</i>	Which TE loci are expressed in single cell blastomeres? Can we detect novel TE-derived isoforms or allele specific TE expression?
TT-seq	Transcriptional run-on followed by high-throughput sequencing, a technique for measuring nascent transcription rates by labeling and sequencing newly synthesized RNA molecules with thiouridine from actively transcribing genes.	$\times 10^7$ cells <i>in vitro</i>	Can we identify actively transcribed TE loci? Does the stability of different TE RNAs vary?
Direct RNA-seq	Direct sequencing of RNA without amplification by Nanopore. This enables transcriptome analysis as well as identification of RNA-modifications, such as m6A.	$\times 10^6$ cells <i>in vitro</i>	Can we detect novel TE-derived isoforms? What is the distribution and impact of RNA modifications on TE-derived transcripts?

Spatial approaches are not limited to RNA, with spatial proteomics emerging as a promising avenue for the study of protein localization in preimplantation [123]. Spatial proteomics could reveal heterogeneity of TE protein translation between the inner and outer cells of eight-cell stage embryos, or differences in subcellular protein localization. However, recent work on LINE1s in cancer underscores the need for careful characterization of TE antibodies to ensure specificity [124]. Such protein tags have also been used in conjunction with retrotransposition assays, potentially facilitating a more precise understanding of protein localization during retrotransposition [38]. Advancements to spatial transcriptomics, proteomics, and advanced imaging may provide a unique perspective on the interaction of TE RNA and protein with other cellular pathways during preimplantation. For example, resolution of ZFPs alongside TE transcripts during preimplantation development could give important insights into the temporal regulation of TE expression.

Perspectives

- The exploration of TE DNA, chromatin, expression, and products in the context of preimplantation development is a multifaceted endeavor. The highly repetitive and mobile qualities make TEs perhaps the most unique genetic elements. Unique genetics alongside the reduced cell numbers available have limited research into TEs in preimplantation. A broad range of recent technological advancements provide the opportunity to explore the roles of TEs.
- TEs are frequently studied as groups of repetitive elements. KD of several TE families in preimplantation embryos have revealed their importance in this phase of development. However, these KDs can affect up to 10% of the genome and, like many other technologies, are limited to investigations of TE families. Since, a technological revolution has enabled a new focus of TE research, at the level of individual loci. Increasingly, such research highlights heterogeneity in the activity, environment, and functions of different TE loci, underlining the importance of a locus-specific view.

- We encourage researchers to move away from the traditional view of TEs as groups of repetitive elements, and to focus on their uniqueness. Discrepancies in TE prevalence among mouse strains emphasizes the need for caution when using genome assemblies. Implementation of long-read technologies will afford a locus-specific view of TE regulation and function and allow a more accurate view of novel TE insertions and their environment. Furthermore, single-cell approaches will teach us about heterogeneity of TEs between cells of the early embryo. In the future, combining these technologies with advancements in *in vitro* methods to study mouse as well as human development will provide unprecedented insights into the role of TEs in preimplantation development.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

Funding

The work is funded a DTP-BBSRC fellowship (L.A.D.) and a IDRMC, Department of Paediatrics Transition Fellowship (R.V.B.).

Open Access

Open access for this article was enabled by the participation of University of Oxford in an all-inclusive *Read & Publish* agreement with Portland Press and the Biochemical Society under a transformative agreement with JISC.

Author Contribution

L.A.D. and R.V.B. wrote the manuscript.

Abbreviations

ERV, endogenous retrovirus; ESC, embryonic stem cell; KD, knockdown; KO, knockout; LINE1, long interspersed nuclear element 1; LTR, long terminal repeat; ONT, Oxford Nanopore Technology; PGC, primordial germ cell; RNA-FISH, RNA fluorescence *in situ* hybridization; TALE, transcription activator-like effector; TE, transposable element; UMI, unique molecular identifier; ZGA, zygotic genome activation; ZNF, zinc-finger nucleases.

References

- Schulz, K.N. and Harrison, M.M. (2019) Mechanisms regulating zygotic genome activation. *Nat. Rev. Genet.* **20**, 221–234 <https://doi.org/10.1038/s41576-018-0087-x>
- Cockburn, K. and Rossant, J. (2010) Making the blastocyst: lessons from the mouse. *J. Clin. Invest.* **120**, 995–1003 <https://doi.org/10.1172/JCI41229>
- Gerdes, P., Richardson, S.R., Mager, D.L. and Faulkner, G.J. (2016) Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol.* **17**, 100 <https://doi.org/10.1186/s13059-016-0965-5>
- Senft, A.D. and Macfarlan, T.S. (2021) Transposable elements shape the evolution of mammalian development. *Nat. Rev. Genet.* **22**, 691–711 <https://doi.org/10.1038/s41576-021-00385-1>
- Guo, Y., Li, T.D., Modzelewski, A.J. and Siomi, H. (2024) Retrotransposon renaissance in early embryos. *Trends Genet.* **40**, 39–51 <https://doi.org/10.1016/j.tig.2023.10.010>
- Low, Y., Tan, D.E.K., Hu, Z., Tan, S.Y.X. and Tee, W.W. (2021) Transposable element dynamics and regulation during zygotic genome activation in mammalian embryos and embryonic stem cell model systems. *Stem Cells Int.* **2021**, 1624669 <https://doi.org/10.1155/2021/1624669>
- O'Neill, K., Brocks, D. and Hammell, M.G. (2020) Mobile genomics: tools and techniques for tackling transposons. *Philos. Trans. R. Soc. B Biol. Sci.* **375**, 20190345 <https://doi.org/10.1098/rstb.2019.0345>
- Berrens, R.V., Yang, A., Laumer, C.E., Lun, A.T.L., Bieberich, F., Law, C.-T. et al. (2022) Locus-specific expression of transposable elements in single cells with CELLO-seq. *Nat. Biotechnol.* **40**, 546–554 <https://doi.org/10.1038/s41587-021-01093-1>
- Ying, Q.-L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J. et al. (2008) The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 <https://doi.org/10.1038/nature06968>
- Williams, R.L., Hilton, D.J., Pease, S., Willson, T.A., Stewart, C.L., Gearing, D.P. et al. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**, 684–687 <https://doi.org/10.1038/336684a0>
- Zhao, S., Lu, J., Pan, B., Fan, H., Byrum, S.D., Xu, C. et al. (2023) TNRC18 engages H3K9me3 to mediate silencing of endogenous retrotransposons. *Nature* **623**, 633–642 <https://doi.org/10.1038/s41586-023-06688-z>

- 12 Moffitt, J.R. and Zhuang, X. (2016) RNA imaging with multiplexed error-robust fluorescence in situ hybridization (MERFISH). *Methods Enzymol.* **572**, 1–49 <https://doi.org/10.1016/bs.mie.2016.03.020>
- 13 Altemose, N., Maslan, A., Smith, O.K., Sundararajan, K., Brown, R.R., Mishra, R. et al. (2022) DiMeLo-seq: a long-read, single-molecule method for mapping protein–DNA interactions genome wide. *Nat. Methods* **19**, 711–723 <https://doi.org/10.1038/s41592-022-01475-6>
- 14 Arnold, C.D., Zabidi, M.A., Pagani, M., Rath, M., Schernhuber, K., Kazmar, T. et al. (2017) Genome-wide assessment of sequence-intrinsic enhancer responsiveness at single-base-pair resolution. *Nat. Biotechnol.* **35**, 136–144 <https://doi.org/10.1038/nbt.3739>
- 15 Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.-C. et al. (2018) Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 <https://doi.org/10.1038/s41586-018-0578-0>
- 16 van den Brink, S.C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotzsch, S., Turner, D.A. et al. (2014) Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 <https://doi.org/10.1242/dev.113001>
- 17 Aguilera-Castrejon, A., Oldak, B., Shani, T., Ghanem, N., Itzkovich, C., Slomovich, S. et al. (2021) Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis. *Nature* **593**, 119–124 <https://doi.org/10.1038/s41586-021-03416-3>
- 18 Yao, H., Sun, N., Shao, H., Wang, T. and Tan, T. (2023) Ex utero embryogenesis of non-human primate embryos and beyond. *Curr. Opin. Genet. Dev.* **82**, 102093 <https://doi.org/10.1016/j.gde.2023.102093>
- 19 Deglincerti, A., Etoc, F., Guerra, M.C., Martyn, I., Metzger, J., Ruvo, A. et al. (2016) Self-organization of human embryonic stem cells on micropatterns. *Nat. Protoc.* **11**, 2223–2232 <https://doi.org/10.1038/nprot.2016.131>
- 20 Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V. et al. (2016) Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* **18**, 700–708 <https://doi.org/10.1038/ncb3347>
- 21 Friedli, M. and Trono, D. (2015) The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* **31**, 429–451 <https://doi.org/10.1146/annurev-cellbio-100814-125514>
- 22 Berteli, T.S., Wang, F., McKerrow, W., Navarro, P.A., Fenyo, D., Boeke, J.D. et al. (2023) Transposon insertion profiling by sequencing (TIPseq) identifies novel LINE-1 insertions in human sperm. *J. Assist. Reprod. Genet.* **40**, 1835–1843 <https://doi.org/10.1007/s10815-023-02852-6>
- 23 Wang, L., Dou, K., Moon, S., Tan, F.J. and Zhang, Z.Z. (2018) Hijacking oogenesis enables massive propagation of LINE and retroviral transposons. *Cell* **174**, 1082–1094.e12 <https://doi.org/10.1016/j.cell.2018.06.040>
- 24 Brouha, B., Schustak, J., Badge, R.M., Lutz-Prigge, S., Farley, A.H., Moran, J.V. 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 <https://doi.org/10.1073/pnas.0831042100>
- 25 Zemojtel, T., Penzkofer, T., Schultz, J., Dandekar, T., Badge, R. and Vingron, M. (2007) Exonization of active mouse L1s: a driver of transcriptome evolution? *BMC Genomics* **8**, 392 <https://doi.org/10.1186/1471-2164-8-392>
- 26 Autio, M.I., Bin Amin, T., Perrin, A., Wong, J.Y., Foo, R.S.Y. and Prabhakar, S. (2021) Transposable elements that have recently been mobile in the human genome. *BMC Genomics* **22**, 789 <https://doi.org/10.1186/s12864-021-08085-0>
- 27 Ferraj, A., Audano, P.A., Balachandran, P., Czechanski, A., Flores, J.I., Radecki, A.A. et al. (2023) Resolution of structural variation in diverse mouse genomes reveals chromatin remodeling due to transposable elements. *Cell Genom.* **3**, 100291 <https://doi.org/10.1016/j.xgen.2023.100291>
- 28 Hooper, M., Hardy, K., Handyside, A., Hunter, S. and Monk, M. (1987) HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292–295 <https://doi.org/10.1038/326292a0>
- 29 Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A. et al. (2020) Telomere-to-telomere assembly of a complete human X chromosome. *Nature* **585**, 79–84 <https://doi.org/10.1038/s41586-020-2547-7>
- 30 Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A. et al. (2022) The complete sequence of a human genome. *Science* **376**, 44–53 <https://doi.org/10.1126/science.abj6987>
- 31 Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B.J. and Loose, M. (2021) Readfish enables targeted nanopore sequencing of gigabase-sized genomes. *Nat. Biotechnol.* **39**, 442–450 <https://doi.org/10.1038/s41587-020-00746-x>
- 32 Kovaka, S., Fan, Y., Ni, B., Timp, W. and Schatz, M.C. (2021) Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. *Nat. Biotechnol.* **39**, 431–441 <https://doi.org/10.1038/s41587-020-0731-9>
- 33 Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E. and Gouil, Q. (2020) Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* **21**, 30 <https://doi.org/10.1186/s13059-020-1935-5>
- 34 Wells, J.N. and Feschotte, C. (2020) A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* **54**, 539–561 <https://doi.org/10.1146/annurev-genet-040620-022145>
- 35 Zhou, M. and Smith, A.D. (2019) Subtype classification and functional annotation of L1Md retrotransposon promoters. *Mob. DNA* **10**, 14 <https://doi.org/10.1186/s13100-019-0156-5>
- 36 Gerdes, P., Chan, D., Lundberg, M., Sanchez-Luque, F.J., Bodea, G.O., Ewing, A.D. et al. (2023) Locus-resolution analysis of L1 regulation and retrotransposition potential in mouse embryonic development. *Genome Res.* **33**, 1465–1481 <https://doi.org/10.1101/gr.278003.123>
- 37 Kong, L., Saha, K., Hu, Y., Tschetter, J.N., Habben, C.E., Whitmore, L.S. et al. (2022) Subfamily-specific differential contribution of individual monomers and the tether sequence to mouse L1 promoter activity. *Mob. DNA* **13**, 13 <https://doi.org/10.1186/s13100-022-00269-z>
- 38 Moran, J.V., Holmes, S.E., Naas, T.P., DeBerardinis, R.J., Boeke, J.D. and Kazazian, H.H. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 [https://doi.org/10.1016/s0092-8674\(00\)81998-4](https://doi.org/10.1016/s0092-8674(00)81998-4)
- 39 Kelley, R.L. and Gardner, D.K. (2017) In vitro culture of individual mouse preimplantation embryos: the role of embryo density, microwells, oxygen, timing and conditioned media. *Reprod. Biomed. Online* **34**, 441–454 <https://doi.org/10.1016/j.rbmo.2017.02.001>
- 40 Oldak, B., Wildschutz, E., Bondarenko, V., Comar, M.-Y., Zhao, C., Aguilera-Castrejon, A. et al. (2023) Complete human day 14 post-implantation embryo models from naive ES cells. *Nature* **622**, 562–573 <https://doi.org/10.1038/s41586-023-06604-5>
- 41 Weatherbee, B.A.T., Gantner, C.W., Iwamoto-Stohl, L.K., Daza, R.M., Hamazaki, N., Shendure, J. et al. (2023) Pluripotent stem cell-derived model of the post-implantation human embryo. *Nature* **622**, 584–593 <https://doi.org/10.1038/s41586-023-06368-y>
- 42 Sundaram, V., Cheng, Y., Ma, Z., Li, D., Xing, X., Edge, P. et al. (2014) Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res.* **24**, 1963–1976 <https://doi.org/10.1101/gr.168872.113>
- 43 Sundaram, V., Choudhary, M.N.K., Pehrsson, E., Xing, X., Fiore, C., Pandey, M. et al. (2017) Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat. Commun.* **8**, 14550 <https://doi.org/10.1038/ncomms14550>

- 44 Sundaram, V. and Wang, T. (2018) Transposable element mediated innovation in gene regulatory landscapes of cells: re-visiting the “gene-battery” model. *BioEssays* **40**, 1700155 <https://doi.org/10.1002/bies.201700155>
- 45 Fueyo, R., Judd, J., Feschotte, C. and Wysocka, J. (2022) Roles of transposable elements in the regulation of mammalian transcription. *Nat. Rev. Mol. Cell Biol.* <https://doi.org/10.1038/s41580-022-00457-y>
- 46 Karttunen, K., Patel, D., Xia, J., Fei, L., Palin, K., Aaltonen, L. et al. (2023) Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. *Nat. Commun.* **14**, 5313 <https://doi.org/10.1038/s41467-023-41081-4>
- 47 Pal, D., Patel, M., Boulet, F., Sundarraj, J., Grant, O.A., Branco, M.R. et al. (2023) H4K16ac activates the transcription of transposable elements and contributes to their cis-regulatory function. *Nat. Struct. Mol. Biol.* **30**, 935–947 <https://doi.org/10.1038/s41594-023-01016-5>
- 48 Liang, L., Cao, C., Ji, L., Cai, Z., Wang, D., Ye, R. et al. (2023) Complementary Alu sequences mediate enhancer–promoter selectivity. *Nature* **619**, 868–875 <https://doi.org/10.1038/s41586-023-06323-x>
- 49 Percharde, M., Lin, C.-J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A. et al. (2018) A LINE1-nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405.e19 <https://doi.org/10.1016/j.cell.2018.05.043>
- 50 Hua, P., Badat, M., Hanssen, L.L.P., Hentges, L.D., Crump, N., Downes, D.J. et al. (2021) Defining genome architecture at base-pair resolution. *Nature* **595**, 125–129 <https://doi.org/10.1038/s41586-021-03639-4>
- 51 Raviram, R., Rocha, P.P., Luo, V.M., Swanzey, E., Miraldi, E.R., Chuong, E.B. et al. (2018) Analysis of 3D genomic interactions identifies candidate host genes that transposable elements potentially regulate. *Genome Biol.* **19**, 216 <https://doi.org/10.1186/s13059-018-1598-7>
- 52 Zhong, J.-Y., Niu, L., Lin, Z.-B., Bai, X., Chen, Y., Luo, F. et al. (2023) High-throughput Pore-C reveals the single-allele topology and cell type-specificity of 3D genome folding. *Nat. Commun.* **14**, 1250 <https://doi.org/10.1038/s41467-023-36899-x>
- 53 Du, Z., Zheng, H., Huang, B., Ma, R., Wu, J., Zhang, X. et al. (2017) Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature* **547**, 232–235 <https://doi.org/10.1038/nature23263>
- 54 Fuentes, D.R., Swigut, T. and Wysocka, J. (2018) Systematic perturbation of retroviral LTRs reveals widespread long-range effects on human gene regulation. *Elife* **7**, e35989 <https://doi.org/10.7554/eLife.35989>
- 55 Pontis, J., Planet, E., Offner, S., Turelli, P., Duc, J., Coudray, A. et al. (2019) Hominoid-specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* **24**, 724–735.e5 <https://doi.org/10.1016/j.stem.2019.03.012>
- 56 Grow, E.J., Flynn, R.A., Chavez, S.L., Bayless, N.L., Wossidlo, M., Wesche, D.J. et al. (2015) Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* **522**, 221–225 <https://doi.org/10.1038/nature14308>
- 57 Meng, S., Liu, X., Zhu, S., Xie, P., Fang, H., Pan, Q. et al. (2023) Young LINE-1 transposon 5′ UTRs marked by elongation factor ELL3 function as enhancers to regulate naive pluripotency in embryonic stem cells. *Nat. Cell Biol.* **25**, 1319–1331 <https://doi.org/10.1038/s41556-023-01211-y>
- 58 Xu, R., Li, S., Wu, Q., Li, C., Jiang, M., Guo, L. et al. (2022) Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos. *Cell Stem Cell* **29**, 1051–1066.e8 <https://doi.org/10.1016/j.stem.2022.06.001>
- 59 Lu, F., Liu, Y., Inoue, A., Suzuki, T., Zhao, K. and Zhang, Y. (2016) Establishing chromatin regulatory landscape during mouse preimplantation development. *Cell* **165**, 1375–1388 <https://doi.org/10.1016/j.cell.2016.05.050>
- 60 Sanchez-Luque, F.J., Kempen, M.-J.H.C., Gerdes, P., Vargas-Landin, D.B., Richardson, S.R., Troskie, R.-L. et al. (2019) LINE-1 evasion of epigenetic repression in humans. *Mol. Cell* **75**, 590–604.e12 <https://doi.org/10.1016/j.molcel.2019.05.024>
- 61 Saha, K., Nielsen, G.I., Nandani, R., Kong, L., Ye, P. and An, W. (2024) YY1 is a transcriptional activator of mouse LINE-1 Tf subfamily. bioRxiv <https://doi.org/10.1101/2024.01.03.573552>
- 62 Imbeault, M., Hellebood, P.-Y. and Trono, D. (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**, 550–554 <https://doi.org/10.1038/nature21683>
- 63 Rowe, H.M., Friedli, M., Offner, S., Verp, S., Mesnard, D., Marquis, J. et al. (2013) De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140**, 519–529 <https://doi.org/10.1242/dev.087585>
- 64 Kaya-Okur, H.S., Wu, S.J., Codomo, C.A., Pledger, E.S., Bryson, T.D., Henikoff, J.G. et al. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 <https://doi.org/10.1038/s41467-019-09982-5>
- 65 Skene, P.J. and Henikoff, S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 <https://doi.org/10.7554/eLife.21856>
- 66 Wang, R., Zheng, Y., Zhang, Z., Zhu, X., Wu, T.P. and Ding, J. (2024) MATES: a deep learning-based model for locus-specific quantification of transposable elements in single cell. bioRxiv <https://doi.org/10.1101/2024.01.09.574909>
- 67 Eckersley-Maslin, M.A., Alda-Catalinas, C. and Reik, W. (2018) Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. *Nat. Rev. Mol. Cell Biol.* **19**, 436–450 <https://doi.org/10.1038/s41580-018-0008-z>
- 68 Zhang, B., Zheng, H., Huang, B., Li, W., Xiang, Y., Peng, X. et al. (2016) Allelic reprogramming of the histone modification H3K4me3 in early mammalian development. *Nature* **537**, 553–557 <https://doi.org/10.1038/nature19361>
- 69 Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y. et al. (2016) The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature* **534**, 652–657 <https://doi.org/10.1038/nature18606>
- 70 Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z. et al. (2018) Chromatin analysis in human early development reveals epigenetic transition during ZGA. *Nature* **557**, 256–260 <https://doi.org/10.1038/s41586-018-0080-8>
- 71 Cheetham, S.W., Jafarani, Y.M.A., Andersen, S.B., Jansz, N., Kindlova, M., Ewing, A.D. et al. (2022) Single-molecule simultaneous profiling of DNA methylation and DNA-protein interactions with Nanopore-DamID. bioRxiv <https://doi.org/10.1101/2021.08.09.455753>
- 72 van den Amele, J., Trauner, M., Hörmanseder, E., Donovan, A.P.A., Battle, O.L., Cheetham, S.W. et al. (2024) Targeted DamID detects cell-type specific histone modifications in vivo. bioRxiv <https://doi.org/10.1101/2024.04.11.589050>
- 73 Mo, W., Shu, Y., Liu, B., Long, Y., Li, T., Cao, X. et al. (2023) Single-molecule targeted accessibility and methylation sequencing of centromeres, telomeres and rDNAs in Arabidopsis. *Nat. Plants* **9**, 1439–1450 <https://doi.org/10.1038/s41477-023-01498-7>
- 74 Sun, T., Xu, Y., Xiang, Y., Ou, J., Soderblom, E.J. and Diao, Y. (2023) Crosstalk between RNA m6A and DNA methylation regulates transposable element chromatin activation and cell fate in human pluripotent stem cells. *Nat. Genet.* **55**, 1324–1335 <https://doi.org/10.1038/s41588-023-01452-5>

- 75 Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A. et al. (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 <https://doi.org/10.1038/nature10960>
- 76 Berrens, R.V., Andrews, S., Spensberger, D., Santos, F., Dean, W., Gould, P. et al. (2017) An endosRNA-based repression mechanism counteracts transposon activation during global DNA demethylation in embryonic stem cells. *Cell Stem Cell* **21**, 694–703.e7 <https://doi.org/10.1016/j.stem.2017.10.004>
- 77 Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W. et al. (2018) Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat. Cell Biol.* **20**, 620–631 <https://doi.org/10.1038/s41556-018-0093-4>
- 78 Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A. et al. (2021) DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 <https://doi.org/10.1186/s13059-021-02510-z>
- 79 Schreiber, J., Wescoe, Z.L., Abu-Shumays, R., Vivian, J.T., Baatar, B., Karplus, K. et al. (2013) Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands. *Proc. Natl Acad. Sci. U.S.A.* **110**, 18910–18915 <https://doi.org/10.1073/pnas.1310615110>
- 80 Haggerty, C., Kretzmer, H., Riemenschneider, C., Kumar, A.S., Mattei, A.L., Bailly, N. et al. (2021) Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* **28**, 594–603 <https://doi.org/10.1038/s41594-021-00603-8>
- 81 Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M. et al. (2011) 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* **2**, 241 <https://doi.org/10.1038/ncomms1240>
- 82 Iqbal, K., Jin, S.-G., Pfeifer, G.P. and Szabó, P.E. (2011) Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. U.S.A.* **108**, 3642–3647 <https://doi.org/10.1073/pnas.1014033108>
- 83 Inoue, A., Matoba, S. and Zhang, Y. (2012) Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. *Cell Res.* **22**, 1640–1649 <https://doi.org/10.1038/cr.2012.160>
- 84 Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W.S., D'Souza, Z. et al. (2016) De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 <https://doi.org/10.1038/ncb3296>
- 85 Rowe, H.M., Kapopoulou, A., Corsinotti, A., Fasching, L., Macfarlan, T.S., Tarabay, Y. et al. (2013) TRIM28 repression of retrotransposon-based enhancers is necessary to preserve transcriptional dynamics in embryonic stem cells. *Genome Res.* **23**, 452–461 <https://doi.org/10.1101/gr.147678.112>
- 86 Bulut-Karslioglu, A., De La Rosa-Velázquez, I.A., Ramirez, F., Barenboim, M., Onishi-Seebacher, M., Arand, J. et al. (2014) Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells. *Mol. Cell* **55**, 277–290 <https://doi.org/10.1016/j.molcel.2014.05.029>
- 87 Walter, M., Teissandier, A., Pérez-Palacios, R. and Bourc'his, D. (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5**, e11418 <https://doi.org/10.7554/eLife.11418>
- 88 Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T. et al. (2010) KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 <https://doi.org/10.1038/nature08674>
- 89 Lanciano, S., Philippe, C., Sarkar, A., Pratella, D., Domrane, C., Doucet, A.J. et al. (2024) Locus-level L1 DNA methylation profiling reveals the epigenetic and transcriptional interplay between L1s and their integration sites. *Cell Genom.* **4**, 100498 <https://doi.org/10.1016/j.xgen.2024.100498>
- 90 Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F. et al. (2015) Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* **16**, 135–141 <https://doi.org/10.1016/j.stem.2015.01.005>
- 91 Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D. et al. (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 <https://doi.org/10.1038/nature11244>
- 92 Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G.J. and Schultz, R.M. (2004) RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 <https://doi.org/10.1016/j.ydbio.2004.01.028>
- 93 Kigami, D., Minami, N., Takayama, H. and Imai, H. (2003) MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 <https://doi.org/10.1095/biolreprod.102.007906>
- 94 Peaston, A.E., Evisikov, A.V., Graber, J.H., de Vries, W.N., Holbrook, A.E., Solter, D. et al. (2004) Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 <https://doi.org/10.1016/j.devcel.2004.09.004>
- 95 Fadloun, A., Le Gras, S., Jost, B., Ziegler-Birling, C., Takahashi, H., Gorab, E. et al. (2013) Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 <https://doi.org/10.1038/nsmb.2495>
- 96 Teissandier, A., Servant, N., Barillot, E. and Bourc'his, D. (2019) Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mob. DNA* **10**, 52 <https://doi.org/10.1186/s13100-019-0192-1>
- 97 Jin, Y., Tam, O.H., Paniagua, E. and Hammell, M. (2015) TET transcripts: a package for including transposable elements in differential expression analysis of RNA-seq datasets. *Bioinformatics* **31**, 3593–3599 <https://doi.org/10.1093/bioinformatics/btv422>
- 98 Yang, W.R., Ardeljan, D., Pacyna, C.N., Payer, L.M. and Burns, K.H. (2019) SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res.* **47**, e27 <https://doi.org/10.1093/nar/gky1301>
- 99 He, J., Babarinde, I.A., Sun, L., Xu, S., Chen, R., Shi, J. et al. (2021) Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat. Commun.* **12**, 1456 <https://doi.org/10.1038/s41467-021-21808-x>
- 100 Chang, N.-C., Rovira, Q., Wells, J., Feschotte, C. and Vaquerizas, J.M. (2022) Zebrafish transposable elements show extensive diversification in age, genomic distribution, and developmental expression. *Genome Res.* **32**, 1408–1423 <https://doi.org/10.1101/gr.275655.121>
- 101 Jönsson, M.E., Garza, R., Johansson, P.A. and Jakobsson, J. (2020) Transposable elements: a common feature of neurodevelopmental and neurodegenerative disorders. *Trends Genet.* **36**, 610–623 <https://doi.org/10.1016/j.tig.2020.05.004>
- 102 Deininger, P., Morales, M.E., White, T.B., Baddoo, M., Hedges, D.J., Servant, G. et al. (2017) A comprehensive approach to expression of L1 loci. *Nucleic Acids Res.* **45**, e31 <https://doi.org/10.1093/nar/gkw1067>
- 103 Franke, V., Ganesh, S., Karlic, R., Malik, R., Pasulka, J., Horvat, F. et al. (2017) Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. *Genome Res.* **27**, 1384–1394 <https://doi.org/10.1101/gr.216150.116>
- 104 Oliveira, D.S., Fablet, M., Larue, A., Vallier, A., Carareto, C.M.A., Rebollo, R. et al. (2023) ChimeraTE: a pipeline to detect chimeric transcripts derived from genes and transposable elements. *Nucleic Acids Res.* **51**, 9764–9784 <https://doi.org/10.1093/nar/gkad671>

- 105 Peng, Z., Yuan, C., Zellmer, L., Liu, S., Xu, N. and Liao, D.J. (2015) Hypothesis: artifacts, including spurious chimeric RNAs with a short homologous sequence, caused by consecutive reverse transcriptions and endogenous random primers. *J. Cancer* **6**, 555–567 <https://doi.org/10.7150/jca.11997>
- 106 Berthelier, J., Furci, L., Asai, S., Sadykova, M., Shimazaki, T., Shirasu, K. et al. (2023) Long-read direct RNA sequencing reveals epigenetic regulation of chimeric gene-transposon transcripts in *Arabidopsis thaliana*. *Nat. Commun.* **14**, 3248 <https://doi.org/10.1038/s41467-023-38954-z>
- 107 Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A. et al. (2016) TT-seq maps the human transient transcriptome. *Science* **352**, 1225–1228 <https://doi.org/10.1126/science.aad9841>
- 108 Wang, Y., Li, Y., Skuland, T., Zhou, C., Li, A., Hashim, A. et al. (2023) The RNA m6A landscape of mouse oocytes and preimplantation embryos. *Nat. Struct. Mol. Biol.* **30**, 703–709 <https://doi.org/10.1038/s41594-023-00969-x>
- 109 Fukuda, K., Okuda, A., Yusa, K. and Shinkai, Y. (2018) A CRISPR knockout screen identifies SETDB1-target retroelement silencing factors in embryonic stem cells. *Genome Res.* **28**, 846–858 <https://doi.org/10.1101/gr.227280.117>
- 110 Ardeljan, D., Steranka, J.P., Liu, C., Li, Z., Taylor, M.S., Payer, L.M. et al. (2020) Cell fitness screens reveal a conflict between LINE-1 retrotransposition and DNA replication. *Nat. Struct. Mol. Biol.* **27**, 168–178 <https://doi.org/10.1038/s41594-020-0372-1>
- 111 Mita, P., Sun, X., Fenyö, D., Kahler, D.J., Li, D., Agmon, N. et al. (2020) BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. *Nat. Struct. Mol. Biol.* **27**, 179–191 <https://doi.org/10.1038/s41594-020-0374-z>
- 112 Frost, J.M., Amante, S.M., Okae, H., Jones, E.M., Ashley, B., Lewis, R.M. et al. (2023) Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat. Struct. Mol. Biol.* **30**, 527–538 <https://doi.org/10.1038/s41594-023-00960-6>
- 113 Jachowicz, J.W., Bing, X., Pontabry, J., Bošković, A., Rando, O.J. and Torres-Padilla, M.E. (2017) LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat. Genet.* **49**, 1502–1510 <https://doi.org/10.1038/ng.3945>
- 114 Sakashita, A., Kitano, T., Ishizu, H., Guo, Y., Masuda, H., Ariura, M. et al. (2023) Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat. Genet.* **55**, 484–495 <https://doi.org/10.1038/s41588-023-01324-y>
- 115 Guthmann, M., Qian, C., Gialdini, I., Nakatani, T., Ettinger, A., Schauer, T. et al. (2023) A change in biophysical properties accompanies heterochromatin formation in mouse embryos. *Genes Dev.* **37**, 336–350 <https://doi.org/10.1101/gad.350353.122>
- 116 Los, G.V., Encell, L.P., McDougall, M.G., Hartzell, D.D., Karassina, N., Zimprich, C. et al. (2008) HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3**, 373–382 <https://doi.org/10.1021/cb800025k>
- 117 Banaz, N., Mäkelä, J. and Uphoff, S. (2018) Choosing the right label for single-molecule tracking in live bacteria: side-by-side comparison of photoactivatable fluorescent protein and Halo tag dyes. *J. Phys. Appl. Phys.* **52**, 064002 <https://doi.org/10.1088/1361-6463/aaf255>
- 118 Mita, P., Wudzinska, A., Sun, X., Andrade, J., Nayak, S., Kahler, D.J. et al. (2018) LINE-1 protein localization and functional dynamics during the cell cycle. *eLife* **7**, e30058 <https://doi.org/10.7554/eLife.30058>
- 119 Asimi, V., Sampath Kumar, A., Niskanen, H., Riemenschneider, C., Hetzel, S., Naderi, J. et al. (2022) Hijacking of transcriptional condensates by endogenous retroviruses. *Nat. Genet.* **54**, 1238–1247 <https://doi.org/10.1038/s41588-022-01132-w>
- 120 Sil, S., Keegan, S., Ettefa, F., Denes, L.T., Boeke, J.D. and Holt, L.J. (2023) Condensation of LINE-1 is critical for retrotransposition. *eLife* **12**, e82991 <https://doi.org/10.7554/eLife.82991>
- 121 Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L. et al. (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* **26**, 317–325 <https://doi.org/10.1038/nbt1385>
- 122 Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R. et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 <https://doi.org/10.1038/ncomms14049>
- 123 Gahurova, L., Tomankova, J., Cerna, P., Bora, P., Kubickova, M., Virnicchi, G. et al. (2023) Spatial positioning of preimplantation mouse embryo cells is regulated by mTORC1 and m7G-cap-dependent translation at the 8- to 16-cell transition. *Open Biol.* **13**, 230081 <https://doi.org/10.1098/rsob.230081>
- 124 Taylor, M.S., Wu, C., Fridy, P.C., Zhang, S.J., Senussi, Y., Wolters, J.C. et al. (2023) Ultrasensitive detection of circulating LINE-1 ORF1p as a specific multicancer biomarker. *Cancer Discov.* **13**, 2532–2547 <https://doi.org/10.1158/2159-8290.CD-23-0313>