Dynamic and Ongoing *De Novo* L1 Retrotransposition Contributes to Genome Plasticity and Intrapatient Heterogeneity in Ovarian Cancer

Barun Pradhan[1,2], Jaana Oikkonen[1], Kaiyang Zhang[3], Juan Manuel Botto[4], Minna R. Eriksson[1], Srividhya Sundaresan[1], Fatih Genç[1], Thomas R. Pisanic II[5,6], Matías Marín Falco[1], Yilin Li[1], Sanna Pikkusaari[1], Kari Lavikka[1], Giulia Micoli[1], Giovanni Marchi[1], Taru A. Muranen[1], Kaisa Huhtinen[7], Anna Vähärautio[1], Richard Badge[8], Kathleen H. Burns[9,10], Sakari Hietanen[11], Johanna Hynninen[11], Geoffrey J. Faulkner[2,4], Sampsa Hautaniemi[1], and Liisa Kauppi[1]

ABSTRACT

Long interspersed element-1 (L1) retrotransposons are the only protein-coding active transposable elements in the human genome. Although typically silenced in normal cells, they are highly expressed in many human epithelial cancers, including high-grade serous ovarian cancer (HGSC), and can integrate into the genome through retrotransposition. *De novo* L1 insertions are known to contribute to genomic instability and cancer evolution in epithelial malignancies, including HGSC, suggesting that they might also play a role in intrapatient tumor heterogeneity. In this study, we quantified *de novo* L1 insertions in clinical HGSC specimens and uncovered high heterogeneity in total L1 insertion events (L1 burden) between patients. HGSC tumors with high L1 burden were highly proliferative, whereas tumors with low or no L1 insertions showed enrichment of immune response and cell death pathways. Although the overall L1 burden was similar across different tumor sites within the same patient, the specific L1 insertions (L1 profiles) diverged significantly more than their single-nucleotide variants profiles. Taken together, these findings demonstrate that L1 activity and retrotransposition are highly dynamic *in vivo* and can contribute substantially to tumor genome plasticity, especially at late stages of cancer progression. The patient-specific propensity of acquiring L1 insertions (L1 burden) could be driven by molecular properties of the progenitor tumor. Retrotransposition-associated DNA damage and/or replication stress could be a potential molecular vulnerability for precision cancer medicine approaches.

Significance: L1 retrotransposition is a dynamic process that continues at late stages of high-grade serous ovarian cancer and can substantially contribute to intrapatient tumor heterogeneity.

Introduction

Long INterspersed Element-1 (LINE-1 or L1) retrotransposons are the only active, protein-coding family of transposable elements in the human genome. They move via an RNA intermediate by a copy-and-paste mechanism called retrotransposition, using two self-encoded proteins called L1-ORF1p and L1-ORF2p. As a consequence, a new copy (L1 repeat sequence) is inserted into a novel location in the genome. Of the approximately 500,000 L1s that occupy the human genome (1), just more than 100 evolutionarily young L1s, with intact sequences encoding functional L1-ORF1p and L1-ORF2p, are retrotransposition-competent (RC; refs. 2, 3). Although transcriptionally silenced in normal adult cells, L1s are highly active in different epithelial cancers (4, 5), and tumor-specific somatic L1 insertions are frequent in cancer genomes (6–12). However, only a small number of *de novo* L1 insertions have been identified as driver events in cancers (10, 13–16), with the majority of insertions thought to be passenger mutations (6, 7, 11).

High-grade serous ovarian carcinoma (HGSC) is the most lethal type of gynecologic malignancy (17). It typically originates in the fallopian tube (18) and spreads to various anatomic regions within the peritoneal cavity, such as the ovaries and omentum. HGSC is characterized by significant L1 dysregulation as indicated by elevated cancer cell–specific L1 transcript expression (19, 20) and a high incidence of L1-ORF1p positivity—94% of HGSCs tested were immunoreactive to an L1-ORF1p antibody (4). Aberrant L1 expression starts early in serous tubal intraepithelial carcinomas, which are precursors to HGSC in the fallopian tube (21, 22). Despite early and persistent L1 expression during HGSC development and

---

[1]Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Helsinki, Finland. [2]Mater Research Institute - University of Queensland, Brisbane, Australia. [3]Department of Pathology, University of Helsinki and HUS Diagnostic Center, Helsinki University Hospital, Helsinki, Finland. [4]Queensland Brain Institute, University of Queensland, Brisbane, Australia. [5]Johns Hopkins Institute for NanoBioTechnology, Johns Hopkins University, Baltimore, Maryland. [6]Department of Oncology, The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins, Baltimore, Maryland. [7]Cancer Research Unit, Institute of Biomedicine and FICAN West Cancer Centre, University of Turku, Turku, Finland. [8]Department of Genetics, Genomics and Cancer Sciences, University of Leicester, Leicester, United Kingdom. [9]Dana-Farber Cancer Institute, Boston, Massachusetts. [10]Harvard Medical School, Boston, Massachusetts. [11]Department of Obstetrics and Gynecology, University of Turku and Turku University Hospital, Turku, Finland.

J. Oikkonen and K. Zhang contributed equally to this article.

Corresponding Author: Liisa Kauppi, Research Program in Systems Oncology, Research Programs Unit, Faculty of Medicine, University of Helsinki, Haartmaninkatu 8 (PO Box 63), Helsinki 00014, Finland. E-mail: liisa.kauppi@helsinki.fi

Cancer Res 2026;86:1073–88

doi: 10.1158/0008-5472.CAN-24-4419

progression, the frequency of *de novo* L1 retrotransposition events per tumor is highly variable in different patients (8, 11, 23), with roughly 30% of cases devoid of any *de novo* L1 insertion in the studies cited. This makes HGSC a pertinent setting for addressing the question of why some tumors are more restrictive to *de novo* L1 insertions than others despite possessing all the obvious components needed for L1 retrotransposition. Moreover, little is known about L1-mediated intrapatient genomic heterogeneity, that is, how tumors at different metastatic sites that originated from a common HGSC progenitor differ in their *de novo* L1 retrotransposition profile.

Here, we quantify and map somatic *de novo* L1 insertion events, both between and within patients, in clinical HGSC tumor specimens collected in the observational DECIDER trial. Furthermore, using multiomics data from the same samples, we assess which molecular factors are associated with tumors with low versus high L1 insertion rates. Our findings indicate that L1 retrotransposition is an ongoing and highly dynamic process during HGSC evolution that contributes to intrapatient tumor heterogeneity.

Materials and Methods

Human tumor specimens

Tumor samples from patients with HGSC (FIGO stage III and above) at the Turku University Hospital, Finland, prior to any cancer treatment were used for this study. Tumors were sampled primarily from two anatomic regions, the tubo-ovarian region, which comprises tumors located at ovaries or the adnexal mass, and the omentum. In total, DNA and RNA were extracted from 50 samples from 28 patients with HGSC (cohort 1). For 21 patients, at least two tumor samples were available from different anatomic regions, whereas for the remaining seven patients, just one (either omental or ovary/adnexus sample) was analyzed. Informed consent was received from all participating patients and the study was approved by the ethics committee of the Hospital District of Southwest Finland (VARHA/28314/13.02.02/2023).

Cell lines

A panel of six HGSC cell lines [OAW28 (RRID: CVCL_1614), COV318 (RRID: CVCL_2419), OVCAR5 (RRID: CVCL_1628), COV362 (RRID: CVCL_2420), KURAMOCHI (RRID: CVCL_1345), and OVSAHO (RRID: CVCL_3114)] and a human teratocarcinoma cell line, PA-1 (RRID: CVCL_0479), were used in this study. OAW28, COV318, and COV362 were acquired from Merck (cat. #85101601, 07071903, and 07071904, respectively); KURAMOCHI and OVSAHO were acquired from Japanese Cancer Research Resources Bank (cat. #JCRB0098 and JCRB1046, respectively); OVCAR5 was acquired from Sigma (SCC259); and PA-1 was obtained from ATCC (CRL-1572). All HGSC cell lines were authenticated by Eurofins Genomics Europe Applied Genomics GmbH using PCR single-locus technology. PA-1's identity was confirmed by deep long-read whole-genome sequencing (WGS) in Faulkner laboratory. *Mycoplasma* testing was done for all cell lines using the MycoAlert Mycoplasma Detection Kit (Lonza, LT07-318) as per the manufacturer's instructions.

OVCAR-5, OVSAHO, and KURAMOCHI were maintained in RPMI-1640 medium (Lonza) + 10% FBS (Gibco); COV318, COV362, and OAW28 were cultured in DMEM (Lonza) + 10% FBS. Additionally, insulin (Sigma) was added as a supplement to OVCAR-3 and OVCAR-8 (10 μg/mL) and OAW28 (20 IU/L). PA-1 cells were maintained in Minimum Essential Media (MEM; Thermo Fisher Scientific, 11095072), 10% heat-inactivated FBS (Life Technologies, 10082147), 1X GlutaMAX I supplement (Gibco, 35050061), 1X nonessential amino acids (Life Technologies, 11140050), and 100 U/mL penicillin–streptomycin solution (Invitrogen, 15140122)—from now on referred to as MEM complete—in 75 cm^2 U-shaped cell culture flask with vent cap (Corning, 430641U) placed in a humidified incubator at 37°C with 5% CO_2 saturation. Cells were split at 80% confluency using trypsin-EDTA 0.25% (Invitrogen, 25200-056).

WGS data

WGS data of the 50 tumor samples from 28 patients with HGSC (cohort 1) and their matching blood DNA, as well as from 57 tumor samples from a second cohort of 38 patients with HGSC (cohort 2), were obtained from the prospective, multiregion, longitudinal, observational DECIDER clinical trial (ClinicalTrials.gov identifier NCT04846933).

DNA and RNA extraction

The Qiagen AllPrep kit (cat. #80204) was used to extract DNA and RNA simultaneously from approximately 1 mg of tissue, which were then stored at −20°C and −80°C, respectively. DNA concentration was measured using a Quantus fluorometer (Promega, RRID: SCR_026279). To check quality and integrity, 100 ng of the extracted genomic DNA was visualized with 1% agarose gel electrophoresis.

Long-distance inverse PCR

Long-distance inverse PCR (LDI-PCR) to detect somatic L1 retrotransposition was performed as previously described (24, 25). Briefly, 100 ng genomic DNA extracted from tumor samples was digested using restriction enzymes *Nsi*I (New England Biolabs, R3127), *Vsp*I (FastDigest, Thermo Fisher Scientific, FD0914), *Sac*I (FastDigest, Thermo Fisher Scientific, FD1133), *Pst*I (New England Biolabs, R0140), *Hind*III (New England Biolabs, R3104), and *Spe*I (New England Biolabs, R3133) according to the manufacturers' instructions. Restriction enzymes were heat-inactivated after the digestion reaction, and digested DNA was self-ligated to form a circular DNA template using T4 DNA ligase (Thermo Fisher Scientific, EL0011). The standard reaction mix was prepared as per the manufacturer's instructions except for *Vsp*I-digested DNA, to which polyethylene glycol was added to a final concentration of 5% w/v. The ligase reaction mix was directly added to the digested DNA and incubated at room temperature for 1 hour, followed by heat inactivation of the enzyme. *Vsp*I-digested DNA was exceptionally incubated in the ligase mix at 4°C overnight and heat inactivated subsequently. Finally, touch-down PCR was performed on thus generated circular DNA at optimized annealing temperature (for cycling conditions, see Supplementary Table S1). Primers used to detect somatic retrotransposition events arising from RC-L1_22q12.1 and RC-L1_Xp22.2 and their optimized annealing temperature are listed in Supplementary Table S1. Circular DNA template libraries generated by *Nsi*I, *Vsp*I, *Sac*I, and *Pst*I and *Nsi*I, *Vsp*I, *Hind*III, and *Spe*I were used for RC-L1_22q12.1– and RC-L1_Xp22.2–mediated insertion detection, respectively.

Nanopore sequencing

In total, 20 different LDI-PCRs were performed per sample (12 reactions for RC-L1_22q12.1 using circular templates generated from four different restriction enzymes and amplifying each with three sets of primer pairs; eight reactions for RC-L1_Xp22.2 using circular templates generated from four different restriction enzymes, amplifying each with two sets of primer pairs), with two technical replicates for each reaction. All LDI-PCR products from each sample were pooled and quantified using Quantus fluorometer (Promega). Using TapeStation, 5 to 20 ng of the pooled PCR product was analyzed (Agilent, RRID: SCR_018435), and the average molecular weight of the pooled amplicon was calculated. Then 0.1 pmol of the pooled PCR product from each sample was barcoded using the ligation-based barcoding kit NBD-104 and NBD-114 (Oxford Nanopore Technologies). Barcoded amplicon libraries were pooled in an equimolar fashion to produce the final sequencing library, which was introduced to a FLO-MIN106D flowcell and run on MinION Mk1B (RRID: SCR_017985). The raw fast5 files were base-called using the high-accuracy base-calling option on Guppy software version 3.4.5 (RRID: SCR_023196).

LDI-PCR/Nanopore sequencing data analysis

In order to identify the target genomic location of somatic L1 retrotransposition, we utilized a structural variation (SV) calling pipeline called pipeline-structural-variation (available at https://github.com/nanoporetech/pipeline-structural-variation). In short, the pipeline employs minimap2 (RRID: SCR_018550; ref. 26) to map the fastq sequencing reads to the human reference genome, calculates suitable parameters based on the read depth (27), and finally calls the SV using Sniffles (RRID: SCR_017619; ref. 28). The minimum read length was set to 200 bp, and the rest of the parameters were the defaults. As Sniffles is designed to detect SV from WGS data, it erroneously detects the restriction enzyme cut-site generated in our LDI-PCR protocol as SV breakpoints. We therefore manually curated the SV calls generated by visualizing them on the Integrative Genomics Viewer (RRID: SCR_011793) and by analyzing the reads associated with each call. To qualify as "true" SV calls, the reads associated with SV detected had to meet the following criteria: (i) be disjointed at one of the restriction enzyme cut-sites and (ii) have supplementary alignments to the unique tag of RC-L1_22q12.1 or RC-L1_Xp22.2. Additionally, we visualized the alignment files for each sample at the breakpoints to observe polyA sequence and target-site modification (duplications or deletions) that are typical sequence signatures of L1 retrotransposition and reported them in Supplementary Table S2.

To infer clonality, we normalized the total read counts supporting each insertion for the number of restriction enzymes and primer pairs that could detect the somatic L1 insertion in question. Normalized read count (NRC) was thus calculated as NRC = (absolute read count supporting an insertion)/(number of restriction enzymes × number of primer pairs detecting it).

Validation PCR for somatic L1 insertion–mediated intrapatient heterogeneity detected by LDI-PCR/Nanopore sequencing

To validate somatic L1 insertions detected by LDI-PCR/Nanopore sequencing (Nanopore-seq), we designed PCR primers across the junction point in the target genomic location as identified by LDI-PCR/Nanopore-seq. Next, we performed two different PCRs. The first one (PCR I) used only the junction-spanning primer pair targeting the insertion site. A genuine somatic insertion was expected to yield two products, one corresponding to the wild-type (WT; empty) allele and another larger product representing the allele containing the insertion. The second one (PCR II) combined the reverse primer from PCR I with one of the primers originally used in LDI-PCR that targets the source RC-L1. This reaction was designed to amplify the unique junction formed by the somatic insertion, resulting in a single, insertion-specific product. The expected size of the insertion-specific amplicon in each PCR was estimated based on LDI-PCR/Nanopore-seq reads supporting the insertion event. A somatic L1 insertion was considered validated if either PCR assay produced an insertion-specific amplicon of the predicted size, consistent with the presence of the insertion. PCR primers and conditions are provided in Supplementary Table S3.

Detection of inactivating variant in RC-L1_22q12.1 sequence

To detect the presence of inactivating alleles in the RC-L1_22q12.1 sequence, we compared the variants present in the RC-L1_22q12.1 sequence in the genomes of RC-L1_22q12.1 (−) samples (samples with no somatic L1 insertions arising from RC-L1_22q12.1) with those of RC-L1_22q12.1 (+) samples (samples that contained somatic *de novo* insertions arising from RC-L1_22q12.1). We applied LONGSHOT (RRID: SCR_025318; ref. 29), a tool to detect single-nucleotide variants (SNV) in single-molecule long-read sequencing data, to identify SNVs in RC-L1_22q12.1 and its flanking sequence. BCFtools (RRID: SCR_005227; ref. 30) was then used to identify RC-L1_22q12.1 variants specific to RC-L1_22q12.1 (−) samples.

Validation of the effect of inactivating variant in RC-L1-22q12.1 by L1 retrotransposition assay

We introduced ORF2p coding region missense variant (rs201455670) into an RC-L1 sequence, LINE retrotransposable element 3 (LRE3; hereafter referred to as LRE3 rs201455670), in an L1 retrotransposition reporter using overlap extension-PCR (31). The L1 reporter construct used for this experiment was a modified version of 99gfpLRE3 (32), in which the cytomegalovirus promoter was replaced with a CBh promoter. Additionally, a coding sequence expressing mCherry and puromycin resistance, linked by a 2A self-cleaving peptide, was inserted downstream of the CBh promoter. As a negative control, we used a retrotransposition-incompetent construct containing a reverse transcriptase–disabling point mutation (D702A) in the ORF2p coding region (33, 34), hereafter referred to as LRE3 RT−.

For retrotransposition assays, PA-1 cells were seeded in six-well plates (Corning, CLS3516-50EA) at 2.5 × 10^5 cells per well. Sixteen hours after seeding, cells were transfected using 1 μg of either LRE3 WT, LRE3 RT−, or LRE3 rs201455670 reporter plasmid. After 24 hours, the transfection medium was replaced with MEM complete, and cells were allowed to recover for another 24 hours. Puromycin was then added to the culture medium at a final concentration of 1 μg/mL to select for transfected cells. Six days after transfection, cells were harvested and analyzed by flow cytometry. Retrotransposition efficiency was quantified using an enhanced GFP (EGFP)–based L1 reporter assay, as previously described (35).

DNA sequencing data analysis

We reanalyzed WGS data (36–38), obtained from the DECIDER project (https://www.deciderproject.eu/), of 219 treatment-naïve tumor samples from 66 patients together with their matched normals. A total of 102 samples belonged to cohort 1 and 117 to cohort 2. Preprocessing, short mutation calling, and copy-number estimation were performed as before (36, 37). Briefly, 150 bp paired-end reads were trimmed using Trimmomatic 0.32 (RRID: SCR_011848; ref. 39) before alignment to GRCh38.d1.vd1 using BWA-MEM 0.7.12-r1039 (RRID: SCR_010910; arXiv 1303.3997). This was followed by duplicate marking with Picard 2.6 (https://github.com/broadinstitute/picard, RRID: SCR_006525) and base quality recalibration with GATK 3.7 (RRID: SCR_001876; ref. 40). We called short somatic variants with GATK 4.1.9.0 Mutect2 (RRID: SCR_026692) with joint calling (JC) and germline variants with allele-specific filtering and joint genotyping, both according to best practices. Germline status in tumor samples was assessed via forced calling with Mutect2. Allele-specific copy numbers were estimated using modified ASCAT (RRID: SCR_016868; ref. 41) with segmentation from GATK.

Single-base substitution (SBS) signatures and ovaHRDscar scores for these tumors/patients were obtained from previous studies (36, 38). SBS signatures were assigned based on the mutation classification of Catalogue of Somatic Mutations in Cancer (COSMIC) reference signatures v3.2 (RRID: SCR_002260; ref. 42). Briefly, SBS signatures were adjusted for GRCh38 trinucleotides, excluding chromosomes Y and M, and SBS signatures were fitted independently on each sample using an R implementation based on SigProfilerAttribution (43). Signatures with at least 20% ovarian cancer occurrence in COSMIC were selected as common signatures. OvaHRDscar scores were estimated using copy-number profiles from GRIDSS (RRID: SCR_027130) and PURPLE (RRID: SCR_022999) from HMFtools (https://github.com/hartwigmedical/pipeline5; ref. 38). Tumors with an ovaHRDscar score of 54 or higher were classified as deficient in homologous recombination (HR; ref. 44), whereas those with scores below 54 were classified as proficient in HR.

To estimate *de novo* L1 retrotransposition events for each tumor sample [obtained from the tubo-ovarian region (ovary/adnexus) or omentum; total 107 tumor samples from 66 patients, cohorts 1 and 2 combined] from WGS data, we used x-Transposable element analyzer (xTea) v0.1 (45) with default parameters. We included only samples that contained at least 20% tumor cells to have sufficient read coverage for L1 insertion detection. Tumor purity was estimated by taking into account the copy-number variation and truncal pathogenic *TP53* mutation exhibited by the tumor (36). The number of filtered high-confident somatic L1 events was used as an estimate for L1 events within the sample. L1 burden was calculated as either the sum of *de novo* L1 insertions detected by xTea and LDI-PCR/Nanopore-seq (for cohort 1) or the total *de novo* L1 insertions detected by xTea only (for cohort 2). Tumors from cohort 1 and 2 were independently classified as L1 high (highest quartile of L1 burden) and L1 low (lowest quartile of L1 burden).

Association of patient-level L1 status with genomic features

Genomic features, including mutational burden, ovaHRDscar (44), and mutational signatures based on SNVs and small insertion and deletion, as well as somatic or germline mutation of key HR-mediated repair (HRR) genes of these tumors, were previously characterized (36, 38). As these genomic features were calculated at the patient level, we categorized patients with HGSC in our cohorts based on their L1 status. Patients with at least one L1-high tumor were considered “L1-high,” whereas those with at least one L1-low tumor were considered L1-low.

Genotyping PCR of RC-L1_22q12.1 and RC-L1_Xp22.2

To detect the presence/absence of RC-L1_22q12.1 and RC-L1_Xp22.2 in their native genomic location, PCR primers were designed across each RC-L1. A total of 1.5 ng of genomic DNA was amplified by conventional PCR using Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific, F549L) and PCR mix as per the manufacturer’s instructions using the Phusion GC buffer. PCR primers used and cycling conditions are provided in Supplementary Table S4. PCR products were then resolved on a 1% agarose gel. Tumors were classified as homozygous for the source loci if only the large L1-containing amplicon was visible upon agarose gel electrophoresis. The presence of both large and small amplicons indicated heterozygosity, whereas only the small amplicon indicated absence of the source RC-L1. In cases of failed or ambiguous PCR results, we referred to LDI-PCR/Nanopore-seq data, which also amplify the source RC-L1 in its native genomic location.

Reverse-transcriptase PCR

RT-PCR was performed to detect the expression of mRNA of RC-L1_22q12.1 and RC-L1_Xp22.2. RT-PCR primers were designed such that the 3′-end of the forward primer specifically recognized RC-L1–specific ACA variant (46) and the reverse primer annealed to the 3′ flanking nonrepetitive region of the RC-L1. Then 1 μg of RNA extracted from tumor samples was used to synthesize complementary DNA using SuperScript IV Reverse Transcriptase (Thermo Fisher Scientific, 18090010) according to the manufacturer’s instructions. Oligo dT primers (Thermo Fisher Scientific, 18418020) were used to ensure that mRNA with polyA tails were captured. The cDNA produced was diluted in nuclease-free water in a 1:10 ratio, and 1 μL of diluted cDNA was used for the RT-PCR reaction to detect RC-L1_22q12.1 and RC-L1_Xp22.2 transcripts. Phusion Hot Start II DNA polymerase (Thermo Fisher Scientific, F549L) was used along with the Phusion HF buffer provided. The cycling conditions for RT-PCR and primers used are listed in Supplementary Table S5. The same cDNA was used for qRT-PCR using the same dilution and primers using HOT FIREPol EvaGreen qPCR Mix Plus (Solis BioDyne, 08-25-00001) according to the manufacturer’s thermal cycler program.

L1-ORF1p IHC

Formalin-fixed, paraffin-embedded sections from HGSC tumors were deparaffinized in xylene, rehydrated through graded ethanol, and treated with BLOXALL (Vector, SP-6000-100) for 30 minutes to quench endogenous peroxidase activity. Antigen retrieval was performed in a Biocare Decloaking Chamber at 110°C for 5 minutes using EnVision FLEX Target Retrieval Solution (Agilent, K8005). Sections were rinsed in tris-buffered saline with 1% Tween and incubated with anti–L1-ORF1p antibody (Sigma-Aldrich, MABC1152; RRID: AB_2941775; 1:500 in Dako REAL Diluent, Agilent S2022) for 1 hour at room temperature. After washing, slides were incubated with EnVision FLEX/HRP (Agilent, SM802) for 30 minutes and developed with DAB substrate (Agilent, SM803) for 5 minutes. Sections were counterstained with Mayer hematoxylin, Lillie modification (Agilent, S3309) for 1 minute and washed in Scott tap water substitute for 2 minutes to develop the hematoxylin blue color. Slides were then dehydrated, cleared, and mounted with Eukitt (Sigma-Aldrich, 03989). Images were captured using a Zeiss Axio Imager (RRID: SCR_018876) microscope.

Differential gene expression and gene set enrichment analysis

We processed, quantified, and deconvolved the RNA sequencing reads as previously described (47). Briefly, read pairs were trimmed using Trimmomatic (version 0.33; ref. 39) and the trimmed reads were aligned to the GRCh38.d1.vd1 reference genome with GENCODE v25 annotation using STAR (version 2.5.2b, RRID: SCR_004463; ref. 48). We quantified the gene-level effective counts using eXpress (version 1.5.1-linux_x86_64, RRID: SCR_006873; ref. 49) and estimated the cancer cell–specific expressions using PRISM (RRID: SCR_027136; ref. 47).

We performed differential gene expression analysis between L1-high ($n = 30$) and L1-low ($n = 29$) samples on cancer cell–specific effective counts using DESeq2 (version 1.36.0; RRID: SCR_015687; ref. 50). Gene set enrichment analysis (GSEA; RRID: SCR_003199; ref. 51) was performed using fgsea (version 1.22.0; RRID: SCR_020938; bioRxiv 060012) by ranking differentially expressed genes (DEG) based on their $-\log_{10}$ (P value) × sign (fold change). We used the hallmark, Kyoto Encyclopedia of Genes and Genomes, PID, and REACTOME gene sets collected from the Molecular Signatures Database (version 7.5.1; RRID: SCR_016863) in GSEA.

Single-cell RNA sequencing data analysis

For nine patients with HGSC (all in cohort 1) for whom single-cell RNA sequencing (scRNA-seq) was performed previously (52), we retrieved preprocessed and annotated scRNA-seq gene count matrices from the Gene Expression Omnibus (RRID: SCR_005012) database. Samples contained on average 446 cancer cells, and samples with fewer than 40 cancer cells were excluded from the analysis. For each sample, we calculated the percentage of cancer cells expressing *LIN28A* and *LIN28B* by determining the proportion of cells with >0 Unique Molecular Identifier (UMI) counts for the respective genes.

Estimates of let-7 miRNA levels

Total RNA, including miRNAs, was extracted from cells and tissues using the Qiagen miRNeasy Tissue/Cells Advanced Mini Kit (cat. #217604) following the manufacturer's protocol. cDNA synthesis was performed using the Qiagen miRCURY LNA RT Kit (cat. #339340) with a synthetic RNA spike-in control, according to the "Pick-&-Mix" protocol. The resulting cDNA was diluted 1:80 and loaded onto custom Qiagen miRCURY LNA miRNA PCR panel plates (cat. #339330), precoated with primers targeting various let-7 and control miRNAs (Supplementary Table S6). qRT-PCR was carried out using the Qiagen miRCURY LNA SYBR Green PCR Kit (cat. #339345) on a Bio-Rad CFX96 thermocycler (RRID: SCR_018064) with the following conditions: 2 minutes at 95°C, followed by 40 cycles of 10 seconds at 95°C and 60 seconds at 56°C. Cq values were calibrated using interplate calibration and normalized by subtracting the mean Cq of 10 reference miRNAs to account for sample-to-sample variation. These normalized values (ΔCq) were used for comparisons across tumor samples.

Functional assay to assess the role of LIN28 in L1 retrotransposition

PA-1 cells were treated with MEM complete containing 2.5 μmol/L LIN28 inhibitor C1632 (cat. #SML3673, Sigma-Aldrich). As a control, PA-1 cells treated with 0.01% of DMSO were used. This treatment was maintained throughout the whole experiment. Twenty-four hours later, cells were seeded in six-well plates (Corning, CLS3516-50EA) at 2.5×10^5 cells per well. Sixteen hours after seeding, cells were transfected using 1 μg of either WT or RT 99CBhgfpLRE3 L1 reporter plasmid. Then 24 hours later, transfection media were replaced with MEM complete containing 2.5 μmol/L LIN28 inhibitor C1632 (or DMSO) to allow 24-hour recovery. Cells were harvested and processed for flow cytometry analysis 6 days after transfection. L1 retrotransposition efficiency was measured via an EGFP L1 reporter system, as described previously (35).

Ki67 immunofluorescence and quantification

Formalin-fixed, paraffin-embedded sections from HGSC tumors were deparaffinized in xylene and rehydrated through a series of decreasing concentrations of ethanol washes. Antigen retrieval was performed using 10 mmol/L citrate buffer, followed by blocking with normal donkey serum. Sections were then incubated overnight at 4°C with primary antibodies against Ki67 (Abcam, ab15580, RRID: AB_443209, 1:1,000 dilution) and pan-cytokeratin (Sigma-Aldrich, C2931, RRID: AB_258824, 1:1,000 dilution). After washing, sections were incubated with fluorescently labeled secondary antibodies at a 1:500 dilution: donkey anti–mouse IgG-Alexa Fluor 488 (Invitrogen, A-21202, RRID: AB_141607) and donkey anti–rabbit IgG-Alexa Fluor 647 (Invitrogen, A-31573, RRID: AB_2536183). Nuclei were counterstained with Hoechst, and slides were scanned at 20× magnification using the Pannoramic 250 FLASH III digital slide scanner (3DHISTECH, RRID: SCR_022184).

Digitized slides were analyzed using QuPath (version 0.5.1, RRID: SCR_018257; ref. 53). For each sample, tissue regions were manually annotated, and a custom pixel classifier for pan-cytokeratin was used to identify tumor-rich areas. In cases with low tumor content or patchy pan-cytokeratin staining, tumor-rich regions were manually selected. Nuclear segmentation was performed using the pre-trained StarDist model dsb2018_heavy_augment (54). Detected nuclei were filtered based on size and circularity to retain only tumor cells. Ki67-positive tumor cells were identified using a custom object classifier. Although the same classifier was applied across all samples, it was adjusted in a few exceptional cases to account for artifacts or low-intensity staining.

WGS-based similarity analysis for intrapatient tumors

Analysis of within-patient tumor heterogeneity of genome-wide L1 insertions

To estimate within-patient tumor heterogeneity, all high-confident insertions detected in any sample were white-listed and kept from the raw xTea calls for that patient. This joint-calling (JC) procedure ensured that the unique L1 insertions were not detected in other samples of a patient. The JC-saved insertions were not used to estimate the L1 insertion count of a single sample or patient to ensure comparability between patients with variable number of samples.

L1 insertions versus SNV comparisons

We assessed the similarity between two tumors from the same L1-high patient by comparing their shared SNVs and L1 retrotransposition events. Only L1 insertions identified using xTea analysis after JC (xTea_JC) were considered for this purpose. To quantify this similarity, we used the Jaccard Index (J.I.), defined as J.I. = (number of events shared by both tumors)/(total number of events in both tumors combined). We calculated the J.I. for all tumor pairs in which each tumor had at least one L1 retrotransposition event. This approach provided a quantitative measure of the overlap between the tumors based on shared genetic alterations.

Data visualization

All the plots presented in this article were created in R (https://www.R-project.org/) using ggplot2 (RRID: SCR_014601; ref. 55), genomicRanges (RRID: SCR_000025; ref. 56), and Rcircos (RRID: SCR_003310; ref. 57) packages.

Results

Global and locus-specific detection of *de novo* L1 retrotransposition

It has been shown across cancer types that some tumors have a high burden of somatic L1 insertions, whereas others have few or no

L1 insertions (6, 7, 11). To estimate the total *de novo* L1 insertion burden in real-world HGSC specimens, we used two independent, complementary techniques (**Fig. 1A**), xTea (45) and LDI-PCR/Nanopore-seq (**Fig. 1B**; refs. 24, 25), on 50 HGSC samples from 28 patients. The first method identifies genome-wide L1 insertions from WGS data, whereas the latter detects *de novo* L1 insertions arising from specific, highly active L1 loci and only picks up insertions that involve 3′ transduction (a process by which L1 mobilizes its 3′ flanking sequence; ref. 58). We designed the LDI-PCR assays to detect L1 insertions from two source L1s located at 22q12.1 and Xp22.2 (L1 AL121825 and L1 AC004554 in ref. 3, respectively) that were shown to be highly active in HGSC (8, 11). These source elements are referred to as RC-L1_22q12.1 and RC-L1_Xp22.2 hereon.

A priori, L1 insertions without any 3′ transduction (dark orange in **Fig. 1C**) cannot be detected by LDI-PCR/Nanopore-seq. In contrast, xTea detects sibling insertion or orphan/3′ transduction genome-wide (light orange) and is not limited to the two hot RC L1 loci that were assessed by LDI-PCR/Nanopore-seq. Given these differences in detection strategy, it is perhaps not surprising that we found only minimal overlap between L1 insertion events detected by the two methods (**Fig. 1C**). These findings underscore that *bone fide* L1 insertions are missed if just one approach is relied upon for L1 insertion detection and that transduction mapping can reveal covert activity of active RC-L1 source loci that contribute significantly to human mutagenesis (59).

The total number of *de novo* somatic L1 insertions, i.e., the L1 burden, for each tumor, was calculated as the sum of retrotransposition events detected by the two methods. Altogether, 938 *de novo* somatic insertions were identified, 712 by genome-wide xTea analysis and 226 by LDI-PCR/Nanopore-seq (**Fig. 1C**, see Supplementary Fig. S1 for examples of agarose gel images). L1 burden per tumor ranged from 0 to 228 insertions, with a median of eight insertions (mean 19). All tumors except two contained at least one *de novo* L1 insertion (Supplementary Table S7; both L1-null samples originated from the same patient, EOC891). At the L1-high end of the spectrum, two independent tumor samples (EOC839_OvaL and EOC839_Ome) of a single patient together accounted for >40% (408 of 938) of all detected L1 insertions. Occasionally, estimates of L1 burden from the genome-wide versus locus-specific approach were discrepant (annotated with asterisks in Supplementary Table S7): four samples showed numerous *de novo* L1 insertions by xTea (13–42 events) but very few (0–3 events) by LDI-PCR/Nanopore-seq; conversely, one sample (EOC124_OvaR) showed just three events by xTea but 29 events from RC-L1_22q12.1 LDI-PCR/Nanopore-seq alone. Although the first scenario might be due to lower tumor-specific 3′ transduction activity from the two source RC-L1s selected for our LDI/Nanopore-seq assay, the latter is likely due to LDI-PCR/Nanopore-seq's ability to detect low-abundance subclonal somatic L1 insertions that are missed by WGS-based approach with 40× coverage (24).

LDI-PCR/Nanopore-seq (**Fig. 1B**) provides complete nucleotide-level resolution of each 3′ transduction-bearing insertion allowing precise identification of key L1 retrotransposition hallmarks, such as 3′ polyA sequence, target-site modification, and occasional 5′ inversion. Analysis of Nanopore-seq reads revealed that the average length of L1 insertions with 3′ transduction was 807 bp (median = 610 bp; range, 119–5601 bp; Supplementary Table S2). None of the somatic L1 copies identified in our HGSC sample cohort were full-length and thus are unable to support further retrotransposition. Orphan transduction events, that is, insertions truncated to the extent that the insertion contained no L1 sequence at all, were remarkably common—59% and 84% of total L1 insertion events stemming from RC-L1_22q12.1 and RC-L1_Xp22.2, respectively, fell into this category (**Fig. 1D**; Supplementary Table S2).

L1 retrotransposition heterogeneity between tumors and patients

We then assessed whether L1 retrotransposition capacity was similar between individual tumors from the same patient with HGSC. To this end, we examined the L1 burden in each tumor (i.e., the total number of *de novo* insertions detected) and categorized tumors into three groups by their L1 status: L1 low, L1 intermediate, or L1 high if they were in the bottom quartile (Q1), middle quartiles (Q2–Q3), or top quartile (Q4) of L1 burden, respectively (**Fig. 2A**). We then compared within-patient L1 insertion burden and L1 status in cases for which two or more tumors were available; typically, one tumor was from the tubo-ovarian region, referred to as ovary or adnexus based on their anatomic location hereafter, and another from the omentum (**Fig. 2B**). Different tumors within each patient were largely concordant, that is, both L1 burden and L1 status (L1 low, L1 intermediate, or L1 high) between the anatomic sites were similar (**Fig. 2B**). These findings were corroborated in a second cohort of 57 tumors from 38 patients with HGSC (referred to as cohort 2) for whom L1 burden was measured solely using xTea (Supplementary Fig. S2). We conclude that the propensity to acquire *de novo* somatic L1 insertions seems to be a patient-specific and not a tumor-specific property in HGSC.

Determinants of somatic L1 retrotransposition activity in HGSC tumors

With the observed wide range of total L1 burden across HGSC tumor samples, we asked whether this variability could be explained by the availability (or lack) of L1 components necessary for retrotransposition. To this end, we performed locus-specific DNA- and RNA-level analyses of two source RC-L1s, RC-L1_22q12.1 and RC-L1_Xp22.2.

The first layer of potential regulation involves source RC-L1s at the DNA level. Evolutionarily young RC-L1s can be polymorphic in the human population—if the source L1 is absent from the patient's (and tumor's) genome, it would naturally be unable to spawn any *de novo* insertions. Indeed, RC-L1_Xp22.2 was missing in five tumors originating from patients EOC44, EOC450, and EOC192, explaining the lack of RC-L1_Xp22.2–derived insertions in these cases (**Fig. 2C**, i and iii; Supplementary Fig. S3).

Transcriptional activity of source RC-L1 loci is one obvious candidate for influencing tumor L1 insertion burden. To test this possibility, we assessed locus-specific RNA expression of RC-L1_22q12.1 and RC-L1_Xp22.2 by RT-PCR (**Fig. 2C**, ii and iii; Supplementary Fig. S4). RC-L1_22q12.1 was consistently expressed in nearly all (98%) tumors, but L1 insertions of RC-L1_22q12.1 origin were only detected in 58% (29 of 50) of them (**Fig. 2C**, iii). This indicates that the expression of RC-L1 mRNA is not sufficient to drive *de novo* insertions.

In contrast to RC-L1_22q12.1, the expression of RC-L1_Xp22.2 was more variable, with transcripts undetectable in 44% (20 of 45) of tumors containing at least one genomic copy of this locus. As expected, most tumors lacking RC-L1_Xp22.2 transcripts also lacked corresponding insertions. However, we identified three tumors (marked with asterisks in **Fig. 2C**, iii) that showed no detectable transcripts of RC-L1_22q12.1 and/or RC-L1_Xp22.2 but still harbored somatic L1 insertions derived from them. The absence of RC-L1_22q12.1 expression in one such tumor was further confirmed by

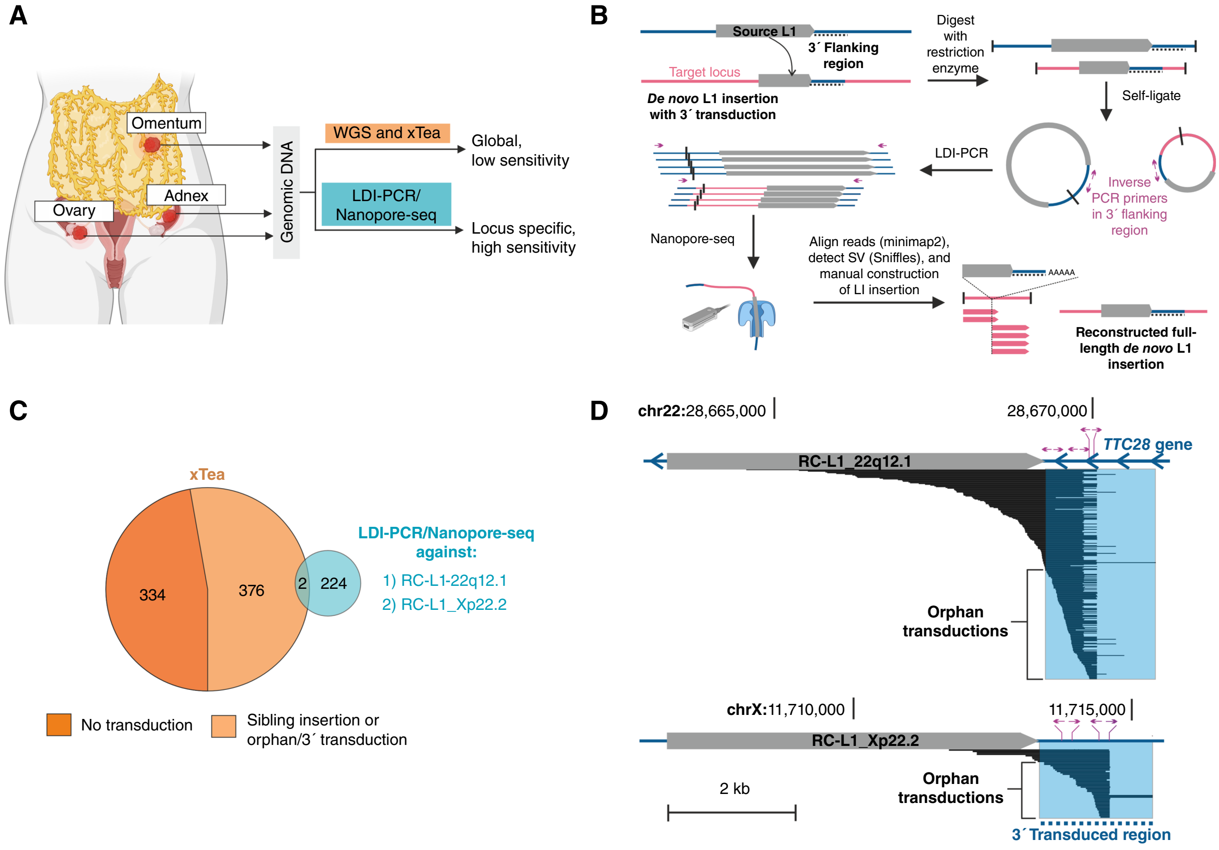


Figure 1.

Detection of *de novo* L1 retrotransposition events in HGSC tumors. **A,** Study design. Clinical HGSC tumor specimens were sampled from one or more anatomic sites from each patient. Somatic L1 retrotransposition events were detected from genomic DNA using two different approaches (WGS/xTea and LDI-PCR/Nanopore-seq). **B,** LDI-PCR/Nanopore-seq detects L1 insertions that involve transduction of the 3′ flanking region (dotted line) from a source L1 locus of choice. Genomic DNA is digested and self-ligated to generate circular DNA templates; black vertical lines indicate restriction enzyme cut and ligation sites. This is followed by LDI-PCR amplification using inverse primers (purple arrows), and LDI-PCR products are sequenced using the Oxford Nanopore Technologies (ONT) platform (referred to as Nanopore-seq). **C,** There was little overlap between *de novo* L1 insertion events detected by xTea and LDI-PCR/Nanopore-seq. **D,** LDI-PCR/Nanopore-seq identified *de novo* somatic insertions originating from RC-L1_22q12.1 and RC-L1_Xq22.2, plotted as black horizontal lines below each source L1 locus. Purple arrows in **B** and **D** depict inverse PCR primers used for LDI-PCR. The transparent blue box highlights 3′ transduced region that was mobilized as part of the insertions. *De novo* insertions both with or without L1 sequence (partnered or orphan 3′ transductions, respectively) were observed. **A,** Created in BioRender. Pradhan, B. (2025) https://BioRender.com/mql6oe2.

qRT-PCR (Supplementary Fig. S4B). These results suggest that current transcriptional activity alone does not explain the variability of L1 insertion burden across tumors. We also tested for the expression of L1-ORF1p protein by IHC in a handful of samples (two L1-high and two L1-low tumors); consistent with earlier reports that L1-ORF1p is near-ubiquitously expressed across HGSC (4), all four tumors were positive for ORF1p (Supplementary Fig. S5). However, ORF1p expression is not exclusive to RC-L1s and may also originate from inactive copies, making it an unreliable indicator of retrotransposition capacity (60).

Another potential contributor to low L1 insertion burden is mutations affecting the protein-coding region of L1 that may render the L1 retrotransposition incompetent. One ORF2p-coding region missense variant (rs201455670) was present heterozygously in the source RC-L1_22q12.1 in both tumors of a patient EOC737 who lacked any insertions from this locus (**Fig. 2C**, iii; Supplementary Fig. S6A). Alteration of the amino acid affected by this variant was reported to reduce retrotransposition efficiency by almost 90% in a scanning mutagenesis study (61). To functionally validate rs201455670, we introduced this variant into an RC LRE3 (62) sequence of a L1-retrotransposition reporter cassette. In L1 retrotransposition assays using PA-1 cells, the construct harboring the rs201455670 mutation completely abolished retrotransposition activity (Supplementary Fig. S6B). Moreover, two commercial HGSC cell lines, KURAMOCHI and OVSAHO, also harbor this variant and by LDI-PCR/Nanopore-seq showed no evidence of insertions from RC-L1_22q12.1, whereas within four other HGSC cell lines, all with the WT ORF2p allele, cells displayed multiple insertions (≥9) from this locus (Supplementary Fig. S6C).

Together, these results demonstrate that the tumor-specific L1 insertion burden in clinical HGSC specimens is shaped by a combination of genomic (whether a functional RC-L1 is present in the tumor genome), transcriptional (whether it is actively transcribed), and posttranscriptional regulatory factors, all of which

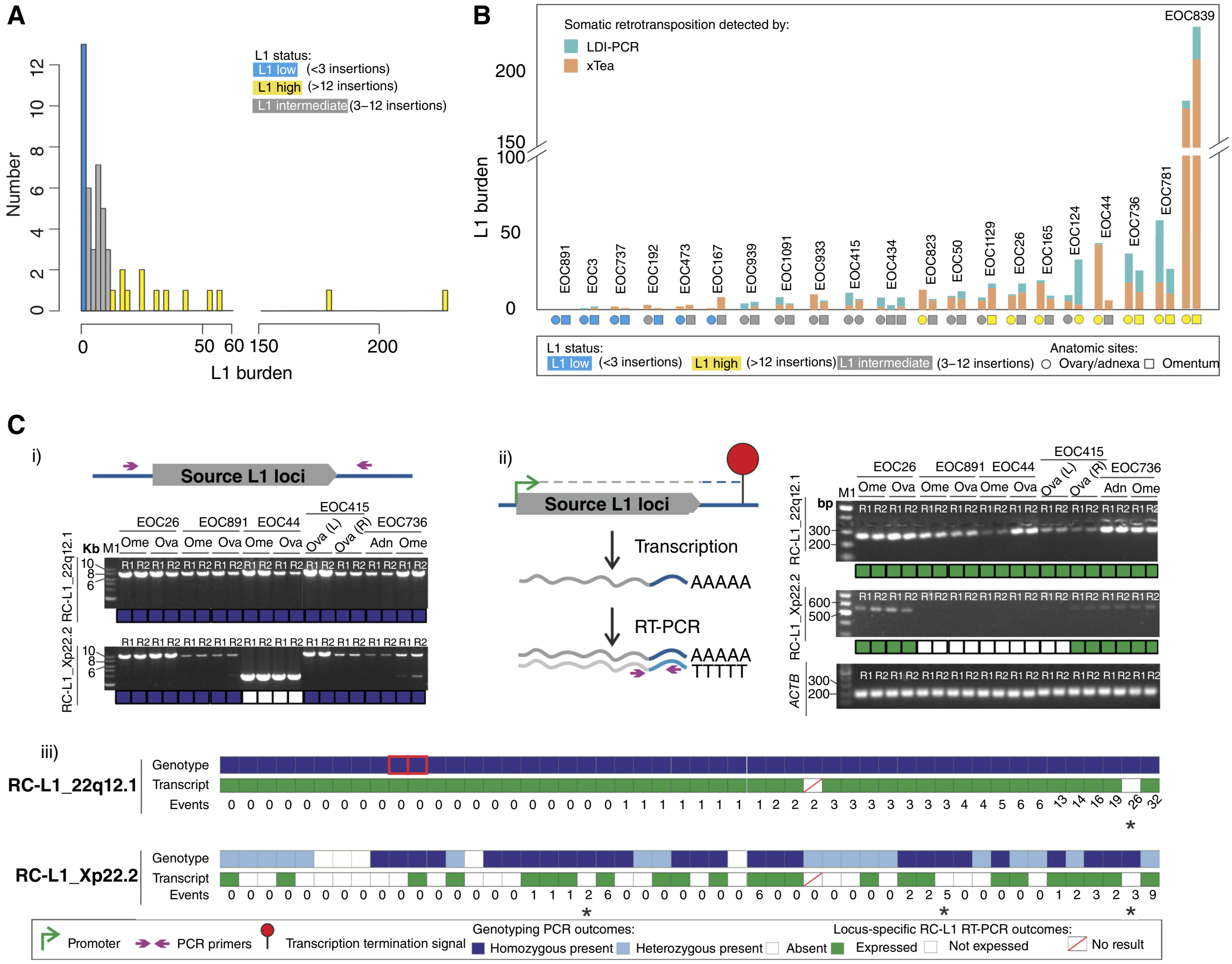


Figure 2.

De novo L1 insertion burden in tumors across and within patients with HGSC. **A,** Distribution of L1 burden across 50 tumors from 28 patients, with each bar representing a bin of two insertions. Tumors in the lowest and highest quartiles of L1 burden were classified as L1 low (blue) and L1 high (yellow), respectively. **B,** L1 burden within each patient for whom more than one tumor sample was available. The L1 status of each tumor is shown using the same color scheme as in **A**. Tumors sampled from the tubo-ovarian region (ovary/adnexus, $n = 23$) and omentum ($n = 20$) are depicted as circles and squares, respectively. **C,** Genotyping and transcriptional analysis of source L1 loci RC-L1_22q12.1 and RC-L1_Xq22.2. (i) Genotyping PCR scheme (PCR primers indicated as purple arrows) and representative results of PCR products, analyzed by agarose gel electrophoresis. R1 and R2 denote replicate reactions. Longer PCR products represent alleles that contain the source L1, whereas shorter PCR products correspond to alleles that lack the source L1. Tumor samples were categorized as homozygous, heterozygous, or absent for the source L1 (dark blue, light blue, and white squares, respectively). Shown are genotyping PCR results from two representative tumors (out of five) that lack RC-L1_Xq22.2. (ii) Transcription of the source L1 in each tumor sample was tested by RT-PCR (primers indicated as purple arrows) targeting 3′ transduction, events that occur when L1 transcription reads through into a noncanonical 3′ flanking termination signal (shown as red lollipop). Green and white squares denote the presence and absence, respectively, of detectable RT-PCR product in at least one of two replicates (R1 and R2). *ACTB* RT-PCR served as a control for RNA integrity and successful cDNA synthesis. (iii) Summary of genotype and expression data of RC-L1_22q12.1 and RC-L1_Xq22.2 in 50 tumor samples, along with the number of *de novo* L1 retrotransposition events detected by LDI-PCR/Nanopore-seq in each sample. Red frames in RC-L1_22q12.1 genotyping data highlight the samples with a heterozygous SNP at the L1-ORF2p coding region, making it incompetent for retrotransposition. Red diagonal lines across white squares indicate that data were not available for the sample. Black asterisks mark tumors lacking transcriptional expression of a source L1 despite having insertions derived from it. Ome, omentum; ova, ovary.

influence whether a given RC-L1 can actively contribute to somatic retrotransposition.

Molecular features associated with L1-high versus L1-low tumors

Given the wide range of L1 burden between tumors, we considered what additional molecular processes, besides those tested above, may contribute to regulating L1 insertional activity. To explore what makes a tumor permissive or restrictive for somatic L1 retrotransposition, we used the matching transcriptomic and genomic data previously produced for these tumor specimens [European Genome-Phenome Archive (EGA) accession numbers EGAS00001004714 and EGAS00001006775, respectively].

Differential gene expression analysis of L1-high versus L1-low HGSC tumors, i.e., tumors with high or low L1 insertion burden, revealed upregulation of several members of the melanoma antigen

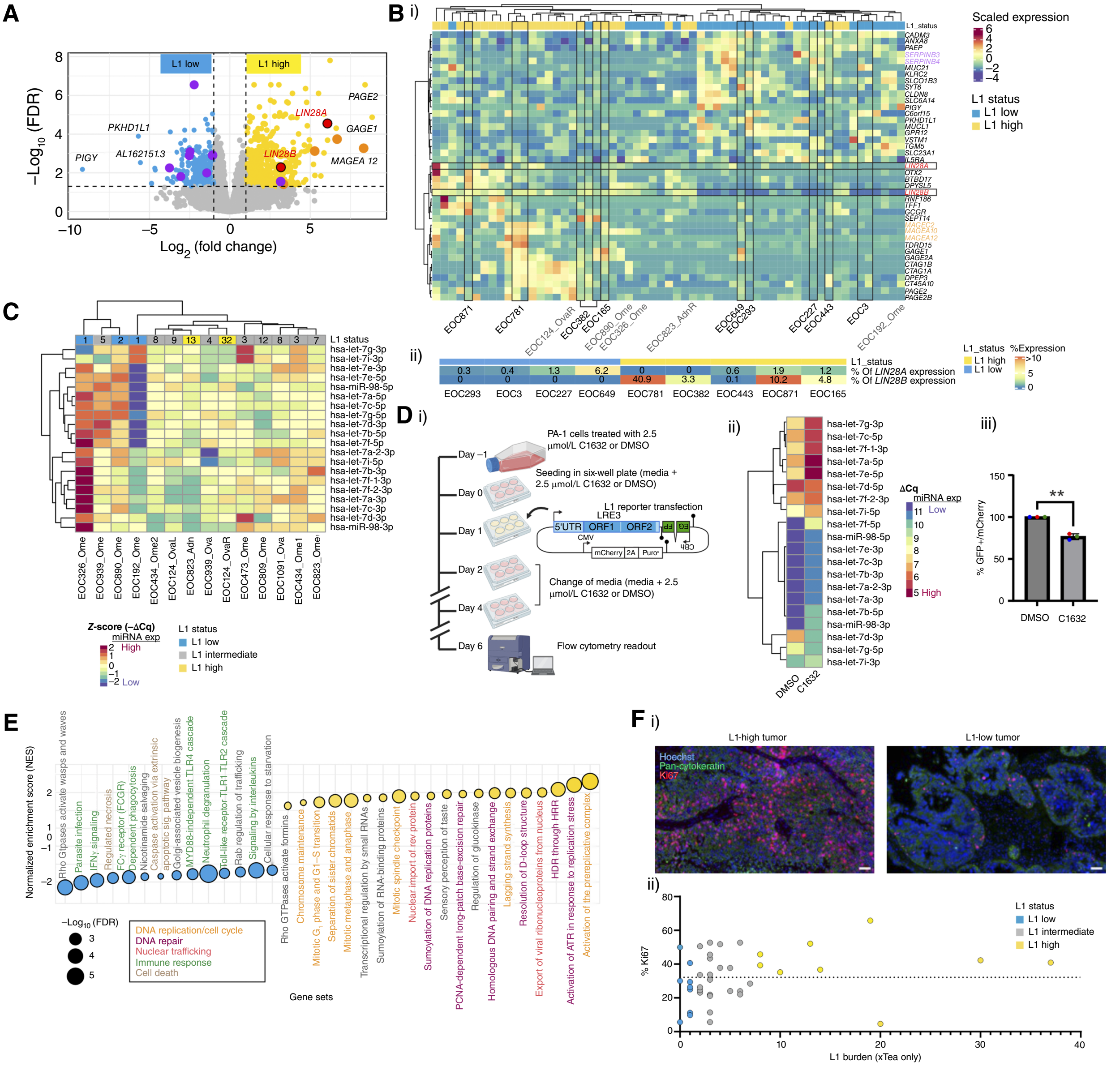
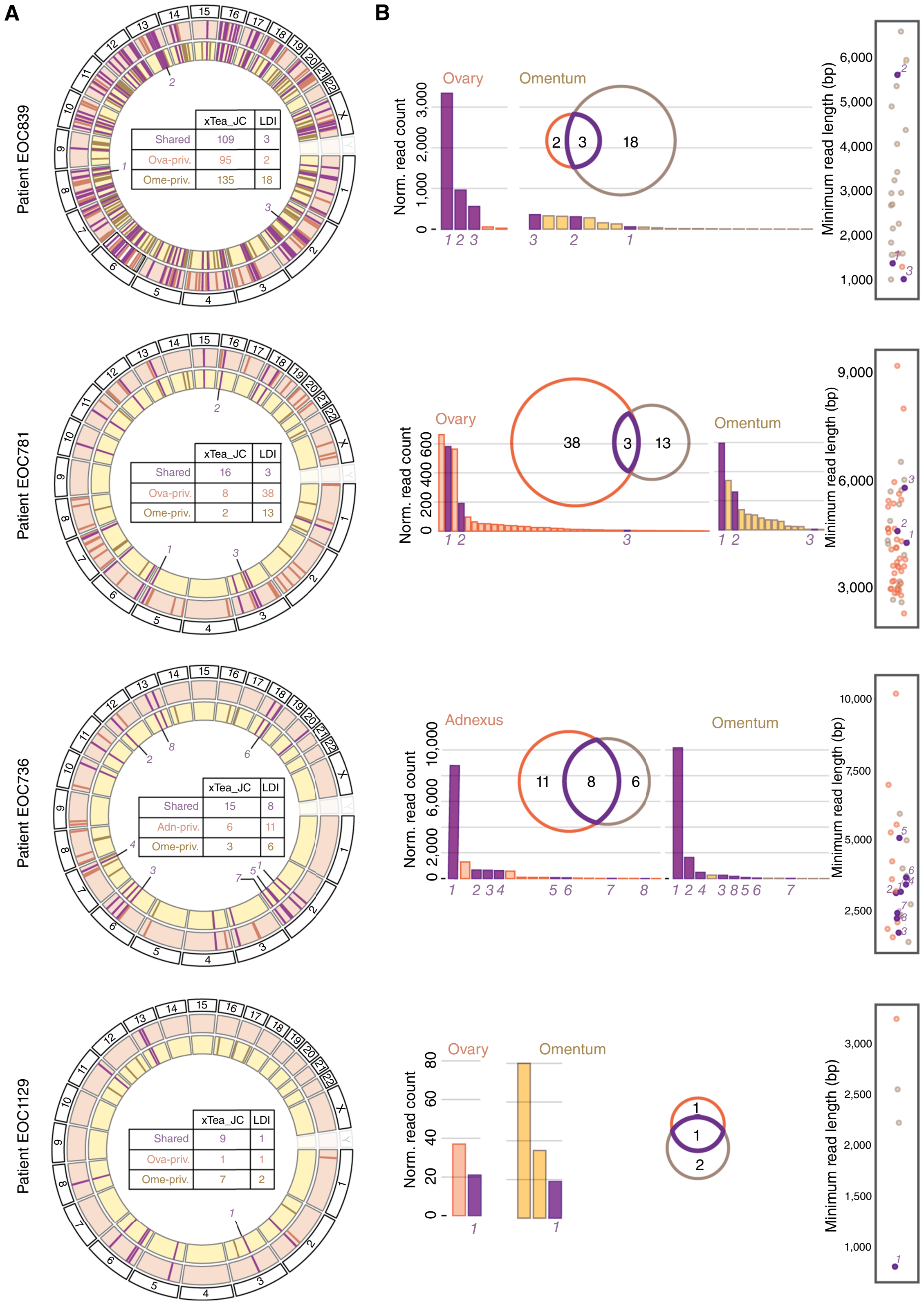


Figure 3.

Transcriptomic features associated with *de novo* L1 retrotransposition frequency. **A,** Volcano plot showing genes that are significantly upregulated (fold change ≥2; FDR < 0.05) in L1-high (yellow) and L1-low (blue) HGSC tumors. The top three genes with the highest fold change in both L1-low and L1-high tumors are labeled with their gene names. Members of the *MAGE* and *SERPIN* family of genes that were significantly upregulated in L1-high and L1-low samples are highlighted in orange and purple circles, respectively. *LIN28A* and *LIN28B* are highlighted in red circles. **B,** Sample-wise expression of top 20 differentially expressed genes, along with *LIN28A* and *LIN28B*. (i) Heatmap displaying gene expression in each L1-low and L1-high tumor sample (*LIN28A* and *LIN28B* highlighted with black horizontal frames). Nine patients for whom scRNA-seq data (32) were available are indicated with vertical black frames. (ii) Sample-wise percentage of tumor cells expressing *LIN28A* and *LIN28B* in scRNA-seq data from nine patient tumors, grouped by L1-high and L1-low status, presented as a heatmap. **C,** Let-7 expression in clinical HGSC tumors. Heatmap shows the relative expression level of 20 let-7 miRNA family members across 14 HGSC tumors. *Z*-scores derived from normalized Cq values are plotted to enhance visualization of expression differences across samples. **D,** LIN28 inhibition using C1632 in the L1 retrotransposition assay. (i) Experimental design. (ii) let-7 miRNA expression in C1632-treated vs. control (DMSO) PA-1 cells. (iii) Barplot showing L1 retrotransposition efficiency in three biological replicates as the percentage of EGFP-positive (EGFP+) events normalized against mCherry-positive (mCherry+) events. *P* values were calculated by a two-tailed *t* test. **, $P \leq 0.01$. **E,** Reactome pathways significantly enriched (FDR < 0.01) in L1-high (yellow dots) and L1-low (blue dots) as determined by GSEA. The pathways or "Gene sets" are ranked according to their normalized enrichment score (NES) and the dot size reflects the FDR. **F,** Quantification of proliferation in clinical tumor samples with variable L1 burden. (i) Ki67 immunofluorescence depicting proliferation activity in representative L1-high and L1-low HGSC tumor sections. Tumor sections were co-stained with Hoechst and pan-cytokeratin to label nuclei and epithelial cancer cells, respectively. Scale bar, 100 μmol/L. (ii) Proliferation levels (Ki67 positivity) plotted against L1 burden (as detected by xTEA) in 40 HGSC tumors. Color of the dots denotes each tumor's L1 status [blue, L1 low (L1 burden ≤1); gray, L1 intermediate (L1 burden = 2–7); and yellow, L1 high (L1 burden ≥8)]. **D,** Created in BioRender. Pradhan, B. (2025) https://BioRender.com/d6rja54.



Patient EOC839

	xTea_JC	LDI
Shared	109	3
Ova-priv.	95	2
Ome-priv.	135	18

Patient EOC781

	xTea_JC	LDI
Shared	16	3
Ova-priv.	8	38
Ome-priv.	2	13

Patient EOC736

	xTea_JC	LDI
Shared	15	8
Adn-priv.	6	11
Ome-priv.	3	6

Patient EOC1129

	xTea_JC	LDI
Shared	9	1
Ova-priv.	1	1
Ome-priv.	7	2

gene, GAGE, and PAGE protein family in L1-high samples (**Fig. 3A** and **B**, i; fold change >2; FDR < 0.05). L1-low samples on the other hand showed higher expression of the *SERPIN* gene family (**Fig. 3A** and **B**, i; fold change >2; FDR <0.05). Among the genes upregulated in L1-high HGSC tumors, *LIN28A* and *LIN28B* (**Fig. 3A**) have a potential mechanistic link to increased somatic L1 insertions: LIN28 has been shown to inhibit the maturation of let-7 miRNA (63, 64), a known repressor of L1 retrotransposition (65).

For more detailed insights on the transcriptomic landscape across L1-high versus L1-low tumor specimens, we examined the sample-wise expression of top 20 DEGs, alongside *LIN28A* and *LIN28B*. *LIN28B* was more consistently expressed across L1-high samples, whereas *LIN28A* was expressed in just a few L1-high samples (**Fig. 3B**, i). As part of an earlier study, tumors from nine of the patients (five L1 high and four L1 low) analyzed herein had been subjected to scRNA-seq (52), allowing for an independent assessment of *LIN28A* and *LIN28B* transcript levels in these tumors. *LIN28B* was expressed in all five L1-high tumors (in 0.1%–40.9% of cancer cells), whereas no expression was detected in the four L1-low tumors (**Fig. 3B**, ii). Thus, both sample-wise bulk RNA-seq and scRNA-seq data implicate *LIN28B*, rather than *LIN28A*, as being robustly associated with retrotransposition in HGSC.

Given that based on earlier reports (64, 65) that *LIN28* is expected to downregulate let-7, we then measured levels of 20 let-7 miRNAs by qRT-PCR in 14 tumors for which fresh-frozen tumor tissue was available (three L1 low, nine L1 intermediate, and two L1 high). L1-low tumor samples showed higher expression of many let-7 miRNAs and clustered together, along with one L1-intermediate tumor (L1 burden = 5). In contrast, L1-high tumors and L1-intermediate tumors with mostly higher L1 burden (≥8) formed a separate cluster and exhibited consistently low let-7 levels (**Fig. 3C**).

To functionally validate the role of LIN28 in regulating L1 retrotransposition, we treated PA-1 cells, a human teratocarcinoma cell line, with the selective LIN28 inhibitor C1632 (**Fig. 3D**, i; ref. 66). Treatment with 2.5 μmol/L C1632 led to upregulation of several let-7 miRNA family members (**Fig. 3D**, ii), consistent with LIN28's known role as a posttranscriptional repressor of let-7 biogenesis. Importantly, this molecular response was accompanied by a ~30% reduction in L1 retrotransposition activity as quantified by an EGFP-based retrotransposition reporter assay (**Fig. 3D**, iii; ref. 67). These findings suggest that LIN28 promotes L1 activity, at least in part, through suppression of let-7 and that pharmacologic inhibition of LIN28 can attenuate L1 mobilization in cancer cells.

We next investigated which biological pathways were enriched in L1-high versus L1-low tumors. GSEA (51) revealed enrichment of several pathways related to DNA replication and cell-cycle progression in L1-high tumors (**Fig. 3E**). In agreement with these findings, the genomes of L1-high tumors also exhibited a significantly higher occurrence of SBS1, a mitotic clock-like mutational signature (Supplementary Fig. S7C; ref. 57). In addition, SBS40, a mutational signature of unknown etiology, and total tumor mutation burden were higher in L1-high tumors (Supplementary Fig. S7D and S7E). We found no significant correlation of patients' age at diagnosis or other systemic factors, such as pregnancy history or hormone replacement therapy use, with their tumor L1 burden (Supplementary Table S8).

To validate whether indeed L1-high tumors were highly proliferative, as suggested by GSEA and enrichment of mutational signature SBS1, we quantified Ki67, a marker of cell proliferation (68), by indirect immunofluorescence in 40 HGSC tumors (nine L1 high, 23 L1 intermediate, and eight L1 low). Tumors with the highest L1 insertion counts tended to also have the highest proportion of Ki67-positive cancer cells (**Fig. 3F**). Overall, there was a statistically significant association between L1 status and Ki67 expression: L1-high tumors were predominantly Ki67 high (i.e., above median Ki67 positivity, which was 32%), whereas L1-low tumors were mostly Ki67 low (below median Ki67 positivity; two-tailed Fisher exact test, $P = 0.0152$; **Fig. 3F**, ii). Together, these results link tumor cell proliferation with high L1 retrotransposition burden, which aligns with the known dependency of L1 insertional activity on DNA replication during S-phase (69).

L1-high tumors were also enriched for DNA replication–related DNA repair pathways such as "Activation of ATR in Response to Replication Stress," suggesting that L1-high tumors experience replication stress caused by high proliferation rate and respond by activating several DNA repair factors. HRR, which is deficient in approximately half of HGSC tumors (70), was also enriched in L1-high tumors (**Fig. 3E**). This finding prompted us to ask whether HRR defects show a negative association with L1 burden. Of 66 patients with HGSC (cohorts 1 and 2 combined), 10 (15%) had biallelic loss-of-function mutations in key HRR genes (*BRCA1*, *BRCA2*, and *RAD51C*), but these mutations were found in both L1-high and L1-low groups (Supplementary Fig. S7B). Likewise, we found no association between L1 burden and HR deficiency (HRD) as assessed by three independent genomics-based HRD indicators, namely SBS3, small insertions and deletions signature ID6, and ovaHRDscar (44). Thus, although GSEA results showed HRR enrichment in L1-high tumors, there was no clear connection between genomic HRD metrics and L1 insertion levels (Supplementary Fig. S7B).

Another key pathway enriched in L1-high HGSC tumors involved the trafficking of viral ribonucleoprotein complexes to and from the nucleus (**Fig. 3E**). This likely reflects the formation of L1 ribonucleoprotein complexes, essential for successful

Figure 4.

Distinct profiles of somatic L1 insertion profiles at different anatomic sites within the same patient. Examples from four representative patients with L1-high HGSC are shown. **A,** Circos plot visualizing *de novo* L1 insertions in two tumor sites [tubo-ovarian (ovary/adnexus) region, outer light orange track and omentum, inner yellow track]. Shared L1 insertions between the two sites are marked as purple lines, whereas those that were unique to tubo-ovarian or omental tumors (private L1 insertions) are marked as dark orange and dark yellow lines, respectively. Shared L1 insertions identified by LDI/Nanopore-seq are numbered based on the number of supporting Nanopore reads in descending order (insertion labeled #1 having the highest read count). The table inside the circos plot summarizes the total shared and private events detected by both xTea (using JC insertions) and LDI-PCR/Nanopore-seq (LDI). **B,** Intrapatient heterogeneity in L1 profiles based on LDI-PCR/Nanopore-seq. Bar plot shows NRCs (normalized for the number of restriction enzymes and primer pairs that can detect the somatic L1 insertion in question) for each *de novo* L1 insertions detected by LDI-PCR/Nanopore-seq. Insertions are ranked from highest to lowest NRC for each tumor sample. L1 insertions unique to ovary/adnexus tumors and omental tumors are in light orange and yellow, respectively, whereas those that are shared by both sites are in purple. Scatter plots on the right show the minimum read lengths by which each L1 insertion was identified (light orange circle, L1 insertion private to ovary/adnexus; yellow circle, L1 insertion private to omentum; purple circle, L1 insertion shared by both sites).

retrotransposition (71), which may utilize the same nuclear trafficking systems as viruses to access the nucleus efficiently. L1-low tumors, on the other hand, were enriched for pathways related to immune response and cell death (both apoptotic and necrotic pathways), along with increased immune signaling activity (**Fig. 3E**). Overall, our findings highlight distinct transcriptomic profiles in L1-high versus L1-low tumors, potentially explaining some of the variability in L1 burden across HGSC tumors.

Distinct subclonal somatic L1 insertions in tumors within the same patient

Given the similarity in total L1 insertion burden between different tumor sites within the same patient (**Fig. 2B**), we asked whether also their *de novo* L1 insertional profiles, that is, the specific target sites in which new L1 insertions occurred, were similar. To this end, *de novo* L1 insertion sites found in the respective tubo-ovarian tumor were compared with those identified in the other tumor site (typically, omentum).

First, we considered genome-wide (i.e., xTea-based) *de novo* L1 insertions. For each patient with at least one L1 insertion detected in both tumors by xTea ($n = 17$ in cohort 1; $n = 14$ in cohort 2; Supplementary Table S9), we annotated insertions as "shared" if they were present at both tumor sites and as "private" if found in only one (**Fig. 4A**; Supplementary Fig. S8). Here, we implemented a JC approach, in which for each within-patient tumor pair, we cross-referenced prefiltered insertion candidates in one tumor against (putative) private insertions identified in the other tumor to ensure that shared insertions were not missed because of stringent default xTea filtering criteria. Most of the L1 insertions so detected were shared between tumors within a patient (55% and 72% shared insertions in cohorts 1 and 2, respectively).

Next, we assessed shared versus private L1 insertions from locus-specific (LDI-PCR/Nanopore-seq) data in the same 17 patients from cohort 1 who had at least one L1 insertion detected in both tumors by xTea (**Fig. 4B**). Here, a very different picture emerged: the vast majority (82%) of L1 insertions identified by LDI-PCR/Nanopore-seq was private (**Fig. 4B**; Supplementary Table S9). In four patients, all with at least one L1 insertion in each tumor, there were no shared insertions at all (Supplementary Table S9). Several of the shared and private insertions were validated using two independent PCR assays (Supplementary Fig. S9).

It seemed likely that the high number of private insertions detected by LDI-PCR/Nanopore-seq would be due to low-abundance insertions present in tumor subclones that may be missed by xTea. To test this, we inferred subclonality by assessing the NRC of Nanopore reads supporting each L1 insertion (**Fig. 4B**; Supplementary Fig. S8). In most (10 of 15) tumors, the shared insertion had the highest NRC, suggesting its presence in the dominant subclone (**Fig. 4B**; Supplementary Fig. S8). Shared insertions also accounted for the majority of total normalized reads (mean = 74%; median = 82%; range = 34%–100%). In the remaining five tumors, private insertions had higher NRCs and made up the majority of total reads (mean = 81%; median = 82%; range = 63%–100%), indicating clonal dominance by lineages that lacked shared insertions. The size of the restriction fragments and/or PCR products, produced in the first step of the LDI-PCR workflow (**Fig. 1C**), could conceivably affect our detection ability as their length can affect the efficiency of self-ligation, LDI-PCR amplification, and Nanopore-seq (Supplementary Fig. S10). No consistent pattern of L1 insertions with the highest NRCs having the shortest read lengths was observed; however (**Fig. 4B**), we found no evidence of any appreciable technical bias affecting these results.

Dynamics of somatic L1 retrotransposition compared with SNVs

Given that tumors at different anatomic sites are thought to arise from a common progenitor, the presence of many private L1 insertions, specifically in L1-high patients (patients with at least one L1-high tumor), raised the possibility of recent and ongoing L1 retrotransposition activity within the individual tumors. To contextualize these findings, we compared joint-called L1 insertion profiles (**Fig. 4A**) of L1-high patients with another class of somatic mutations, namely, SNVs, identified using the same WGS dataset (**Fig. 5A**).

In 14 L1-high patients for whom SNV data existed, intertumor similarity in SNV profiles was significantly higher compared with L1 insertion profiles, reflected in a higher J.I. (Wilcoxon signed-rank test, $P = 0.0354$; **Fig. 5B**). The distribution of J.I. measured from SNVs detected in multiple tumors of L1-intermediate and L1-low patients (combined $n = 13$, who had matching SNV data) was similar to that of L1-high patients. This suggests that most SNVs represent early mutational events and are acquired throughout evolutionary history. In contrast, L1 insertions, specifically in tumors with high L1 burden, seem to reflect a more recent, ongoing mutational process as indicated by the relatively higher number of private L1 insertion events.

Discussion

Here, we comprehensively chart inter- and intra-patient L1 retrotransposition and its potential molecular regulators by multiomic analysis of real-world HGSC specimens. Prior studies have measured a single readout of L1 activity in isolation, demonstrating near-ubiquitous expression of L1 transcripts (19) or L1-ORF1p protein (4), while many HGSC tumors are devoid of *de novo* L1 insertions (8, 11, 23).

High-resolution mapping of *de novo* L1 insertions by LDI-PCR/Nanopore-seq revealed a surprisingly high proportion of orphan transductions, i.e., insertions that did not contain the L1 repetitive sequence at all (**Fig. 1D**). Existing targeted sequencing methods rely on internal (L1 repetitive) sequences of young RC-L1s (8, 10, 72) and would have missed these insertions entirely. For some RC-L1s, orphan transductions may even be the dominant mode of *de novo* L1 insertions as exemplified by insertions arising from RC-L1_Xp22.2. We have now observed prevalent orphan transductions across two different epithelial cancer types, HGSC (herein) and colorectal cancer (24). Thus, the contribution of orphan L1 transductions to genome instability is likely to be sizable in epithelial cancers and may have been substantially underestimated in previous studies because of a lack of suitable, high-sensitivity detection methods.

By concomitantly analyzing the genome and transcriptome of each HGSC tumor, we discovered first a handful of tumors that contained insertions originating from source RC-L1s for which transcripts were undetectable at the time of sampling, indicating that transcriptional status of L1s can switch during tumor evolution. Second, many tumors contained no detectable somatic L1 retrotransposition from "hot" source L1s despite the presence of mature transcripts originating from these elements. We

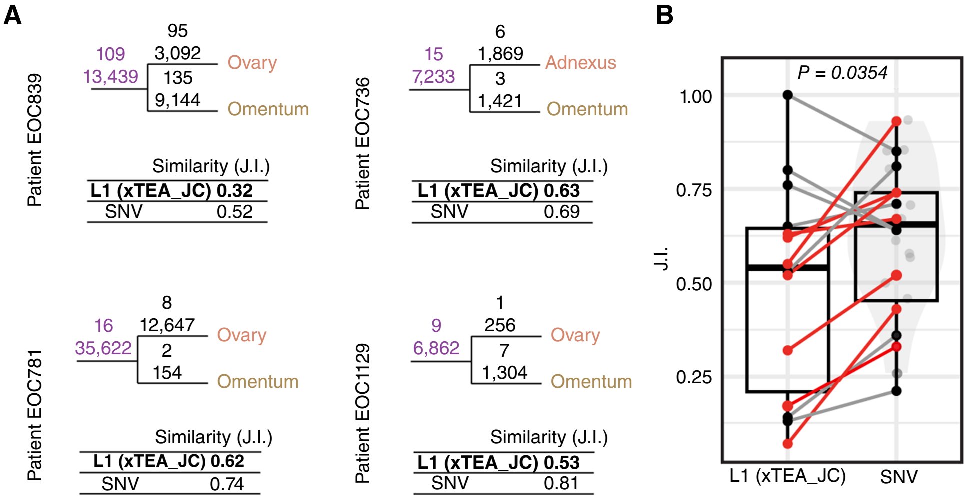


Figure 5.

Comparison of intrapatient heterogeneity based on L1 insertions vs. SNVs. **A,** Shared (premetastatic) vs. private (tumor-specific) L1 insertions identified by xTea with JC correction [L1 (xTea_JC)] and SNVs are shown for the same four representative L1-high patients as in **Fig. 4**. Sample trees display the number of L1 insertions (bold) and somatic SNVs (regular font). Mutations shared between tumor sites are shown in purple; private events are shown in black. The degree of similarity between tumor pairs is quantified using the J.I., for which a higher value indicates greater overlap. **B,** Boxplots show the J.I. calculated for paired tumor samples from the same patient using L1 insertions identified via xTea using JC, L1 (xTea_JC), and SNVs. Each dot represents an L1-high patient, and lines connect paired J.I. values. L1-high patients with total L1 insertions above median are highlighted with red dots and lines. For comparison, J.I. distributions from SNVs in L1-low and L1-intermediate patients are shown as a gray semitransparent violin plot. *P* value was calculated using a two-tailed paired Wilcoxon signed-rank test.

conclude that L1 retrotransposition in HGSC tumors is often regulated posttranscriptionally.

One factor contributing to this posttranscriptional regulation in HGSC could be LIN28 (and LIN28B in particular, in agreement with findings by Piskounova and colleagues, ref. 57), which we found to be lowly expressed in L1-low tumors. RNA-binding protein LIN28 can disrupt let-7 biosynthesis (64, 73, 74), and let-7 microRNAs have been shown to suppress L1 retrotransposition by inhibiting the translation of L1-ORF2p protein (65). Consistent with this, we found that let-7 levels were high in tumors with lower L1 burden and that *in vitro* LIN28 inhibition leads to elevated let-7 expression and reduced L1 retrotransposition efficiency. *In vitro*, let-7 has been shown to restrict L1 retrotransposition without affecting levels of L1 mRNA or L1-ORF1p protein (65). This posttranscriptional mechanism could reconcile also the paradox observed in HGSC: although more than 90% of HGSC tumors are positive for RC-L1 transcripts (19) and for L1-ORF1p by immunostaining (4), only about 70% exhibit somatic L1 retrotransposition events in genomic analyses (8, 11, 23). L1-low HGSC tumors, despite expressing high levels of RC-L1 mRNA and ORF1p, may lack sufficient ORF2p translation because of active let-7, thereby limiting retrotransposition. In L1-high tumors, this let-7–mediated suppression of retrotransposition may be alleviated by *LIN28* overexpression.

Interestingly, LIN28 proteins have also been associated with increased cell proliferation in cancer (75) and here we show that L1-high HGSC tumors were more proliferative, as judged by genomic, transcriptomic, and quantitative Ki67 analyses. Thus, it is tempting to speculate that upregulation of LIN28B may provide HGSC tumors with a milieu that favors somatic retrotransposition both by relieving let-7–mediated repression of L1-ORF2p translation (65) and by promoting cell proliferation, which facilitates greater nuclear access for the L1 machinery during cell division (69).

In gastrointestinal cancers, immune pathways were shown to negatively correlate with somatic L1 burden, and tumors with high immune activity exhibited elevated cytolytic activity (76). Similarly, we show here that L1-low HGSC tumors were enriched for immune signaling, as well as apoptotic and necrotic pathways. It is unclear whether or how this observation links with the innate immune activation that is triggered by the presence of L1 cDNA (77–79). L1 transcripts were expressed broadly across HGSC tumors both in our dataset, and in others analyzed previously (19, 20), and the correlation we report is with low L1 insertion burden, not with L1 RNA expression levels. Whether in these L1-low tumors, a population of “idle” L1 cDNAs exists that could lead to heightened immune activity remains to be investigated.

Similarly, the functional relevance of the DEGs and pathways for L1 retrotransposition remains to be tested. They may report on upstream or downstream processes that are independent of L1 insertional activity *per se* and could, for example, be mediated by let-7 family members. For instance, let-7, which we found to be overexpressed in tumors with low L1 burden, has been shown to downregulate PD-L1, leading to enhanced immune surveillance and reduced cell proliferation (73). It is possible that let-7 acts as a central node regulating L1 suppression, immune activation, and cell proliferation in HGSC.

Our observation that replication stress–associated DNA repair pathways are enriched in L1-high tumors is in contrast with studies that show proteins involved in the same pathway to restrict L1 retrotransposition in *in vitro* assays (80, 81). We reason that although the lack of these DNA repair proteins might initially enable cells to acquire new L1 insertions, eventually the overwhelming number of double-strand breaks induced by the L1 machinery (82) might be detrimental and lead to cell death. Tumors with functional

DNA repair, on the other hand, may allow cells to survive despite L1-mediated DNA damage. This could facilitate the gradual accumulation of *de novo* L1 insertions over time, resulting in a high L1 burden in tumor cells.

In line with previous studies, variability between patients was high in terms of total *de novo* somatic L1 insertions acquired, with tens of insertions in some, whereas 8% of the patients displayed no insertions when assessed using the genome-wide detection method xTea. On the other hand, L1 burden in multiple tumors from the same patient was highly concordant. This indicates that the tendency to accumulate *de novo* somatic L1 insertions in HGSC is more characteristic of the patient rather than individual tumors.

To our knowledge, this is the first study to comprehensively profile intrapatient L1 insertion profiles in HGSC across a large cohort. Despite originating from a common progenitor, and sharing most of the SNVs, different tumor sites within the same patient showed a relatively higher proportion of xTea-detected private L1 insertions as compared with SNVs. The proportion of private L1 insertions in individual tumors of the patient was substantially higher still, when LDI-PCR/Nanopore-seq, a targeted detection method was used. Another study utilizing a targeted approach called Retrotransposon capture sequencing, albeit conducted on just three patients with HGSC, also reported a similarly high proportion of private L1 insertions (8). Targeted methods like LDI-PCR/Nanopore-seq are more sensitive and have the ability to uncover subclonal L1 insertions that are likely to be missed by analyzing WGS data (24). Despite being highly sensitive, it is to be noted that LDI-PCR/Nanopore-seq is locus-specific and can only detect 3′ transduction events that account for roughly 25% of total somatic L1 retrotransposition events across different cancers (7, 11). In contrast, a WGS-based approach like xTea is not limited to preselected RC-L1 or to 3′ transductions and provides a genome-wide estimate of L1 burden. Ideally, a genome-wide strategy that enriches and/or amplifies all RC-L1s, coupled with long-read sequencing, would provide both reliable and sensitive estimates of the somatic L1 retrotransposition landscape in a tumor. Although the magnitude of within-patient heterogeneity in L1 insertions was different when using different detection methods, the picture that emerges is acquisition of *de novo* L1 insertions as an ongoing mutational process in HGSC genomes. This contrasts with findings in endometrioid and clear cell ovarian carcinoma, in which most RC-L1_22q12.1–originated insertions were shared across intrapatient tumor sites, i.e., likely representing clonal events that occurred early during cancer development (83). Our findings in HGSC align with observations in prostate, lung, and pancreatic cancers (7, 46, 84) in which private somatic L1 insertions were found in metastatic lesions, indicating that L1-mediated genome evolution is more prevalent later when the cancer spreads to new anatomic regions. The fact that many *de novo* L1 insertions were unique to each tumor site within one patient also underscores the utility of L1 insertions in tracking intrapatient HGSC evolution and potentially as sensitive markers of recurrence in liquid biopsies.

Taken together, our results indicate that L1 retrotransposition in HGSC genomes is an ongoing and dynamic process *in vivo*. This scenario pinpoints retrotransposition-associated DNA damage and/or replication stress as a molecular vulnerability that could potentially be exploited for precision cancer medicine. L1-high tumors may, for example, be sensitive to drugs that inhibit ATR, a kinase critical for the DNA replication stress repair pathway. Future functional experiments are needed to test the potential for novel, L1-based therapeutic strategies in aggressive epithelial cancers.

Data Availability

WGS data and RNA sequencing data used in this study were obtained from DECIDER clinical trial (ClinicalTrials.gov identifier NCT04846933; refs. 22, 47) and are deposited in EGA with accession numbers EGAS00001006775 and EGAS00001004714, respectively. The scRNA-seq dataset analyzed in this study was retrieved from Gene Expression Omnibus database (accession code: GSE266577). Data used for panel of normals in mutation calling were obtained from The Cancer Genome Atlas's database of Genotypes and Phenotypes under accession number phs000178. LDI-PCR/Nanopore-seq data generated in this study are deposited in EGA with accession number EGAS50000001198. All other raw data are available upon request from the corresponding author.

Authors' Disclosures

B. Pradhan reports grants from Finnish Cultural Foundation during the conduct of the study. J. Oikkonen reports personal fees from European Union during the conduct of the study, as well as personal fees from Orion Corporation outside the submitted work. S. Pikkusaari reports grants from Cancer Foundation Finland outside the submitted work. K. Lavikka reports personal fees from Orion Pharma outside the submitted work. K.H. Burns reports personal fees from Transposon Therapeutics, Regeneron, Scaffold Therapeutics, and Boehringer Ingelheim and grants from Calico outside the submitted work, as well as a patent for 63/423,696 pending. J. Hynninen reports grants from EU HORIZON 2020 during the conduct of the study. S. Hautaniemi reports grants from European Union's Horizon 2020 Research and Innovative Programme [grant agreement 965193 (DECIDER)] and Sigrid Jusélius Foundation and nonfinancial support from CSC-IT Center for Science during the conduct of the study. L. Kauppi reports grants from Research Council of Finland, Sigrid Jusélius Foundation, and Cancer Foundation Finland during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

B. Pradhan: Conceptualization, data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **J. Oikkonen:** Data curation, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **K. Zhang:** Data curation, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **J.M. Botto:** Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–original draft, writing–review and editing. **M.R. Eriksson:** Formal analysis, validation, investigation, visualization, writing–review and editing. **S. Sundaresan:** Data curation, software, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **F. Genç:** Data curation, formal analysis, validation, investigation, visualization, writing–review and editing. **T.R. Pisanic:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **M. Marin Falco:** Data curation, formal analysis, validation, investigation, visualization, writing–review and editing. **Y. Li:** Resources, data curation, software, formal analysis, investigation, methodology, writing–original draft, writing–review and editing. **S. Pikkusaari:** Formal analysis, supervision, investigation, methodology, writing–review and editing. **K. Lavikka:** Data curation, software, formal analysis, investigation, methodology, writing–review and editing. **G. Micoli:** Resources, data curation, formal analysis, writing–review and editing. **G. Marchi:** Resources, data curation, formal analysis, supervision, investigation, writing–review and editing. **T.A. Muranen:** Resources, data curation, formal analysis, supervision, funding acquisition, investigation, methodology, writing–review and editing. **K. Huhtinen:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. **A. Vähärautio:** Formal analysis, validation, investigation, visualization, writing–review and editing. **R. Badge:** Formal analysis, supervision, validation, investigation, visualization, methodology, writing–review and editing. **K.H. Burns:** Formal analysis, validation, investigation, visualization, methodology, writing–review and editing. **S. Hietanen:** Resources, data curation, formal analysis, validation, investigation, visualization, writing–review and editing. **J. Hynninen:** Resources, data curation, formal analysis, supervision, validation, investigation, visualization, methodology, writing–review and editing. **G.J. Faulkner:** Resources, formal analysis, supervision, validation, investigation, visualization, writing–review and editing. **S. Hautaniemi:** Resources, formal analysis, supervision, funding acquisition, validation, investigation, methodology, writing–review and editing. **L. Kauppi:** Conceptualization, resources, data curation, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing.

Acknowledgments

We would like to thank Dr. Manuela Tumiati, Taina Turunen, and Sabrina Gericke for their technical assistance and sample collection and to Dr. Sanna Vuoristo for experimental advice. HGSC cell lines used in this study were kindly provided by Drs. A. Vähärautio and Krister Wennerberg laboratories. This work was supported in part by the European Union's Horizon 2020 Research and Innovation Programme under grant agreement 965193 (DECIDER), the Sigrid Jusélius Foundation, the Cancer Foundation Finland, and the Research Council of Finland (grant number 322178 to L. Kauppi). B. Pradhan was supported by research grants from the Finnish Cultural Foundation. G.J. Faulkner was supported by an Australian National Health and Medical Research Investigator Grant (GNT1173711). We would also like to acknowledge CSC-IT Center for Science, Finland, for computational resources and Biomedicum Functional Genomics Unit (FuGu) for genomics services. The results published here are in part based upon data generated by The Cancer Genome Atlas managed by the NCI and National Human Genome Research Institute. Information about The Cancer Genome Atlas can be found at http://cancergenome.nih.gov. The whole-slide scan images were generated using 3DHISTECH Pannoramic 250 FLASH III digital slide scanner at Finnish Genome Editing Center and analyzed at the Biomedicum Imaging Unit, supported by HiLIFE and the Faculty of Medicine, University of Helsinki, and Biocenter Finland. Generative artificial intelligence assisted in improving clarity and brevity of the author-generated text in parts of the manuscript.

Note

Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Received November 25, 2024; revised July 10, 2025; accepted November 5, 2025; posted first November 12, 2025.

References

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860–921.
2. Beck CR, Collier P, Macfarlane C, Malig M, Kidd JM, Eichler EE, et al. LINE-1 retrotransposition activity in human genomes. Cell 2010;141:1159–70.
3. Brouha B, Schustak J, Badge RM, Lutz-Prigge S, Farley AH, Moran JV, et al. Hot L1s account for the bulk of retrotransposition in the human population. Proc Natl Acad Sci U S A 2003;100:5280–5.
4. Rodić N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS, et al. Long interspersed element-1 protein expression is a hallmark of many human cancers. Am J Pathol 2014;184:1280–6.
5. Ardeljan D, Taylor MS, Ting DT, Burns KH. The human long interspersed element-1 retrotransposon: an emerging biomarker of neoplasia. Clin Chem 2017;63:816–22.
6. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al. Landscape of somatic retrotransposition in human cancers. Science 2012;337:967–71.
7. Tubio JMC, Li Y, Ju YS, Martincorena I, Cooke SL, Tojo M, et al. Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. Science 2014;345:1251343.
8. Nguyen THM, Carreira PE, Sanchez-Luque FJ, Schauer SN, Fagg AC, Richardson SR, et al. L1 retrotransposon heterogeneity in ovarian tumor cell evolution.Cell Rep 2018;23:3730–40.
9. Schauer SN, Carreira PE, Shukla R, Gerhardt DJ, Gerdes P, Sanchez-Luque FJ, et al. L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. Genome Res 2018;28:639–53.
10. Shukla R, Upton KR, Muñoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell 2013;153:101–11.
11. Rodriguez-Martin B, Alvarez EG, Baez-Ortega A, Zamora J, Supek F, Demeulemeester J, et al. Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. Nat Genet 2020;52:306–19.
12. Pitkänen E, Cajuso T, Katainen R, Kaasinen E, Välimäki N, Palin K, et al. Frequent L1 retrotranspositions originating from TTC28 in colorectal cancer. Oncotarget 2014;5:853–9.
13. Miki Y, Nishisho I, Horii A, Miyoshi Y, Utsunomiya J, Kinzler KW, et al. Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. Cancer Res 1992;52:643–5.
14. Scott EC, Gardner EJ, Masood A, Chuang NT, Vertino PM, Devine SE. A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. Genome Res 2016;26:745–55.
15. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. Genome Res 2014;24:1053–63.
16. Cajuso T, Sulo P, Tanskanen T, Katainen R, Taira A, Hänninen UA, et al. Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival. Nat Commun 2019;10:4022.
17. Torre LA, Trabert B, DeSantis CE, Miller KD, Samimi G, Runowicz CD, et al. Ovarian cancer statistics, 2018. CA Cancer J Clin 2018;68:284–96.
18. Labidi-Galy SI, Papp E, Hallberg D, Niknafs N, Adleff V, Noe M, et al. High grade serous ovarian carcinomas originate in the fallopian tube. Nat Commun 2017;8:1093.
19. McKerrow W, Wang X, Mendez-Dorantes C, Mita P, Cao S, Grivainis M, et al. LINE-1 expression in cancer correlates with p53 mutation, copy number alteration, and S phase checkpoint. Proc Natl Acad Sci U S A 2022;119:e2115999119.
20. Perkiö A, Pradhan B, Genc F, Pirttikoski A, Pikkusaari S, Erkan EP, et al. Locus-specific LINE-1 expression in clinical ovarian cancer specimens at the single-cell level. Sci Rep 2024;14:4322.
21. Pisanic TR 2nd, Asaka S, Lin S-F, Yen T-T, Sun H, Bahadirli-Talbott A, et al. Long interspersed nuclear element 1 retrotransposons become deregulated during the development of ovarian cancer precursor lesions. Am J Pathol 2019;189:513–20.
22. Zhouchunyang X, Cochrane DR, Tessier-Coutier B, Leung S, Karnezis AN, Cheng AS, et al. Expression of L1 retrotransposon open reading frame protein 1 in gynecologic cancers. Hum Pathol 2019;92:39–47.
23. Tang Z, Steranka JP, Ma S, Grivainis M, Rodić N, Huang CRL, et al. Human transposon insertion profiling: analysis, visualization and identification of somatic LINE-1 insertions in ovarian cancer. Proc Natl Acad Sci U S A 2017;114:E733–40.
24. Pradhan B, Cajuso T, Katainen R, Sulo P, Tanskanen T, Kilpivaara O, et al. Detection of subclonal L1 transductions in colorectal cancer by long-distance inverse-PCR and nanopore sequencing. Sci Rep 2017;7:14521.
25. Pradhan B, Kauppi L. Detection of retrotransposition activity of hot LINE-1s by long-distance inverse PCR. J Vis Exp 2019;149(e59880). DOI: 10.3791/59880.
26. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 2018;34:3094–100.
27. Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 2018;34:867–8.
28. Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat Methods 2018;15:461–8.
29. Edge P, Bansal V. Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. Nat Commun 2019;10:4660.
30. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011;27:2987–93.
31. Heckman KL, Pease LR. Gene splicing and mutagenesis by PCR-driven overlap extension. Nat Protoc 2007;2:924–32.
32. Macia A, Widmann TJ, Heras SR, Ayllon V, Sanchez L, Benkaddour-Boumzaouad M, et al. Engineered LINE-1 retrotransposition in nondividing human neurons. Genome Res 2017;27:335–48.
33. Moran JV, Holmes SE, Naas TP, DeBerardinis RJ, Boeke JD, Kazazian HH Jr. High frequency retrotransposition in cultured mammalian cells. Cell 1996;87:917–27.

34. Wei W, Morrish TA, Alisch RS, Moran JV. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. Anal Biochem 2000;284:435–8.
35. Sanchez-Luque FJ, Kempen M-JHC, Gerdes P, Vargas-Landin DB, Richardson SR, Troskie R-L, et al. LINE-1 evasion of epigenetic repression in humans. Mol Cell 2019;75:590–604.e12.
36. Lahtinen A, Lavikka K, Virtanen A, Li Y, Jamalzadeh S, Skorda A, et al. Evolutionary states and trajectories characterized by distinct pathways stratify patients with ovarian high grade serous carcinoma. Cancer Cell 2023;41:1103–17.e12.
37. Lavikka K, Oikkonen J, Li Y, Muranen T, Micoli G, Marchi G, et al. Deciphering cancer genomes with GenomeSpy: a grammar-based visualization toolkit, Gigascience 2024;13:giae040.
38. Koskela H, Li Y, Joutsiniemi T, Muranen T, Isoviita V-M, Huhtinen K, et al. HRD related signature 3 predicts clinical outcome in advanced tubo-ovarian high-grade serous carcinoma. Gynecol Oncol 2024;180:91–8.
39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
40. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
41. Van Loo P, Nordgard SH, Lingjærde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A 2010;107:16910–5.
42. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–7.
43. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.
44. Perez-Villatoro F, Oikkonen J, Casado J, Chernenko A, Gulhan DC, Tumiati M, et al. Optimized detection of homologous recombination deficiency improves the prediction of clinical outcomes in cancer. NPJ Precis Oncol 2022;6:96.
45. Chu C, Borges-Monroy R, Viswanadham VV, Lee S, Li H, Lee EA, et al. Comprehensive identification of transposable element insertions using multiple sequencing technologies. Nat Commun 2021;12:3836.
46. Rodić N, Steranka JP, Makohon-Moore A, Moyer A, Shen P, Sharma R, et al. Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. Nat Med 2015;21:1060–4.
47. Häkkinen A, Zhang K, Alkodsi A, Andersson N, Pekcan Erkan E, Dai J, et al. PRISM: recovering cell-type-specific expression profiles from individual composite RNA-Seq samples. Bioinformatics 2021;37:2882–8.
48. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics 2013;29:15–21.
49. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods 2013;10:71–3.
50. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 2014;15:550.
51. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A 2005;102:15545–50.
52. Launonen I-M, Niemiec I, Hincapié-Otero M, Erkan EP, Junquera A, Afenteva D, et al. Chemotherapy induces myeloid-driven spatially confined T cell exhaustion in ovarian cancer. Cancer Cell 2024;42:2045–63.e10.
53. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: open source software for digital pathology image analysis. Sci Rep 2017;7:16878.
54. Schmidt U, Weigert M, Broaddus C, Myers G. Cell detection with star-convex polygons. In: Frangi A, Schnabel J, Davatzikos C, Alberola-López C, Fichtinger G, editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2018. Cham: Springer International Publishing; 2018. p. 265–73.
55. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2018; ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org
56. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. PLoS Comput Biol 2013;9:e1003118.
57. Zhang H, Meltzer P, Davis S. RCircos: an R package for Circos 2D track plots. BMC Bioinformatics 2013;14:244.
58. Goodier JL, Ostertag EM, Kazazian HH Jr. Transduction of 3'-flanking sequences is common in L1 retrotransposition. Hum Mol Genet 2000;9:653–7.
59. Macfarlane CM, Collier P, Rahbari R, Beck CR, Wagstaff JF, Igoe S, et al. Transduction-specific ATLAS reveals a cohort of highly active L1 retrotransposons in human populations. Hum Mutat 2013;34:974–85.
60. Sokolowski M, Chynces M, deHaro D, Christian CM, Belancio VP. Truncated ORF1 proteins can suppress LINE-1 retrotransposition in trans. Nucleic Acids Res 2017;45:5294–308.
61. Adney EM, Ochmann MT, Sil S, Truong DM, Mita P, Wang X, et al. Comprehensive scanning mutagenesis of human retrotransposon LINE-1 identifies motifs essential for function. Genetics 2019;213:1401–14.
62. Brouha B, Meischl C, Ostertag E, de Boer M, Zhang Y, Neijens H, et al. Evidence consistent with human L1 retrotransposition in maternal meiosis I. Am J Hum Genet 2002;71:327–36.
63. Iliopoulos D, Hirsch HA, Struhl K. An epigenetic switch involving NF-kappaB, Lin28, Let-7 MicroRNA, and IL6 links inflammation to cell transformation. Cell 2009;139:693–706.
64. Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, et al. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. Cell 2011;147:1066–79.
65. Tristán-Ramos P, Rubio-Roldan A, Peris G, Sánchez L, Amador-Cubero S, Viollet S, et al. The tumor suppressor microRNA let-7 inhibits human LINE-1 retrotransposition. Nat Commun 2020;11:5712.
66. Roos M, Pradère U, Ngondo RP, Behera A, Allegrini S, Civenni G, et al. A small-molecule inhibitor of Lin28. ACS Chem Biol 2016;11:2773–81.
67. Kopera HC, Larson PA, Moldovan JB, Richardson SR, Liu Y, Moran JV. LINE-1 cultured cell retrotransposition assay. Methods Mol Biol 2016;1400: 139–56.
68. Gerdes J, Schwab U, Lemke H, Stein H. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. Int J Cancer 1983;31:13–20.
69. Mita P, Wudzinska A, Sun X, Andrade J, Nayak S, Kahler DJ, et al. LINE-1 protein localization and functional dynamics during the cell cycle. Elife 2018;7: e30058.
70. Konstantinopoulos PA, Ceccaldi R, Shapiro GI, D'Andrea AD. Homologous recombination deficiency: exploiting the fundamental vulnerability of ovarian cancer. Cancer Discov 2015;5:1137–54.
71. Kulpa DA, Moran JV. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. Nat Struct Mol Biol 2006;13:655–60.
72. Steranka JP, Tang Z, Grivainis M, Huang CRL, Payer LM, Rego FOR, et al. Transposon insertion profiling by sequencing (TIPseq) for mapping LINE-1 insertions in the human genome. Mob DNA 2019;10:8.
73. Chen Y, Xie C, Zheng X, Nie X, Wang Z, Liu H, et al. LIN28/let-7/PD-L1 pathway as a target for cancer immunotherapy. Cancer Immunol Res 2019;7:487–97.
74. Wang L, Rowe RG, Jaimes A, Yu C, Nam Y, Pearson DS, et al. Small-molecule inhibitors disrupt let-7 oligouridylation and release the selective blockade of let-7 processing by LIN28. Cell Rep 2018;23:3091–101.
75. Thornton JE, Gregory RI. How does Lin28 let-7 control development and disease? Trends Cell Biol 2012;22:474–82.
76. Jung H, Choi JK, Lee EA. Immune signatures correlate with L1 retrotransposition in gastrointestinal cancers. Genome Res 2018;28:1136–46.
77. Tunbak H, Enriquez-Gasca R, Tie CHC, Gould PA, Mlcochova P, Gupta RK, et al. The HUSH complex is a gatekeeper of type I interferon through epigenetic regulation of LINE-1s. Nat Commun 2020;11:5387.
78. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, et al. L1 drives IFN in senescent cells and promotes age-associated inflammation. Nature 2019;566:73–8.
79. Baldwin ET, van Eeuwen T, Hoyos D, Zalevsky A, Tchesnokov EP, Sánchez R, et al. Structures, functions and adaptations of the human LINE-1 ORF2 protein. Nature 2024;626:194–206.
80. Mita P, Sun X, Fenyö D, Kahler DJ, Li D, Agmon N, et al. BRCA1 and S phase DNA repair pathways restrict LINE-1 retrotransposition in human cells. Nat Struct Mol Biol 2020;27:179–91.
81. Bona N, Crossan GP. Fanconi anemia DNA crosslink repair factors protect against LINE-1 retrotransposition during mouse development. Nat Struct Mol Biol 2023;30:1434–45.
82. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol 2006;357:1383–93.
83. Xia Z, Cochrane DR, Anglesio MS, Wang YK, Nazeran T, Tessier-Cloutier B, et al. LINE-1 retrotransposon-mediated DNA transductions in endometriosis associated ovarian cancers. Gynecol Oncol 2017;147:642–7.
84. Ketola K, Kaljunen H, Taavitsainen S, Kaarijärvi R, Järvelä E, Rodríguez-Martín B, et al. Subclone eradication analysis identifies targets for enhanced cancer therapy and reveals L1 retrotransposition as a dynamic source of cancer heterogeneity. Cancer Res 2021;81:4901–9.