Promoter methylation, transcription, and retrotransposition of LINE-1 in colorectal adenomas and adenocarcinomas

Milad SHADEMAN
Ferdowsi University of Mashhad
Khadijeh ZARE
Ferdowsi University of Mashhad
Morteza ZAHEDI
Ferdowsi University of Mashhad
Hooman MOSANNEN MOZAFFARI
Mashhad University of Medical Sciences
Hadi BAGHERI HOSSEINI
Mashhad University of Medical Sciences
Kamran GHAFFARZADEGAN
Razavi Hospital
Ladan Goshayeshi
Mashhad University of Medical Sciences
Hesam Dehghani (dehghani@um.ac.ir)
Research Institute of Biotechnology, Ferdowsi University of Mashhad https://orcid.org/0000-0001-6750-0040

Research

Keywords: LINE-1, colorectal adenoma, colorectal adenocarcinoma, methylation, retrotransposition, insertion

Posted Date: May 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-28838/v1

Version of Record: A version of this preprint was published at Cancer Cell International on September 1st, 2020. See the published version at https://doi.org/10.1186/s12935-020-01511-5.

Abstract

Background: LINE-1, Alu, and SVA elements are non-LTR retrotransposons that create approximately one-third of the human genome. The loss of tight control mechanisms on the function of mobile elements has been implicated in many human diseases. The methylation of the CpG islands of the LINE-1 promoter is one of these mechanisms. In this study, we determined the promoter methylation and the expression of LINE-1 in three stages of colorectal non-advanced adenoma, advanced adenoma, and adenocarcinoma. In addition, we analyzed the insertion of LINE-1, Alu, and SVA elements in the genome of colorectal advanced adenomas.

Results: We found that the LINE-1 hypomethylation index in advanced adenoma and adenocarcinoma were significantly higher than that in non-advanced adenomas. The copy number of LINE-1 transcripts in advanced adenoma was significantly higher than that in non-advanced adenomas, and in adenocarcinomas was significantly higher than that in the advanced adenomas. Analysis of the genome of colorectal advanced adenomas revealed that at this stage de novo insertions of LINE-1, Alu, and SVA were approximately 16%, 51%, and 74%, respectively.

Conclusions: Our findings showing a decreased methylation of LINE-1 promoter accompanied by the higher level of LINE-1 transcription, and de novo genomic insertions in advanced (high-grade) adenoma, a precancerous stage before colorectal carcinoma, suggests that the early and advanced polyp stages may host very important pathogenic processes concluding to cancer.

Background

Mobile pieces of DNA called transposable elements occupy more than half of the human genome [1,2]. Insertion of mobile elements into the genome of somatic cells and their movement may lead to cancers. DNA transposons and retrotransposons are the two major classes of mobile elements. Retrotransposons comprise two main groups of human endogenous retroviruses, and poly(A) retrotransposons [2]. In the latter group, LINE-1, Alu, and SVA (SINE-R –VNTR–Alus) make up approximately one-third of the human DNA. One of the most active DNA retrotransposons in mammals is LINE-1, which contains 17% of the human genome and consists of two main open reading frames (ORFs) encoding the necessary proteins for retrotransposition [2,3], one is ORF1 encoding a 40 kDa RNA-binding protein, and the other is ORF2 that encodes a 150 kDa protein with endonuclease and reverse transcriptase activities [4].

CpG dinucleotides which are moderately found at LINE-1 promoter, are methylated in the normal cells [5]. Methylation of the LINE-1 promoter and silencing of its transcription is one of the most well-studied mechanisms of LINE-1 repression [6]. The hypomethylation of the LINE-1 promoter increases the accessibility of the RNA polymerase II and the other regulatory elements to initiate or regulate the transcription of this element [7,8]. It has been shown that in brain cells the young LINE-1 elements that contain truncated or mutated Yin Yang 1 (YY1) binding sites are globally hypomethylated, indicating that the YY1 transcription factor mediates L1 promoter DNA methylation [9]. Binding of Krüppel-associated-box- zinc finger protein (KRAB-ZFP) and recruitment of KAP1 corepressor followed by NURD/HDAC repressor complex, histone methyltransferases, and DNA methyltransferases may as well explain the mechanism of methylation at LINE-1 promoter and several groups of SVA elements [10–12]. The methylation induces silencing of these elements in human ESCs [13]. In addition, it

has been shown that PIWI-interacting RNA-induced silencing complex (piRISC) can also guide the de novo methylation machinery to LINE-1 locus [14,15].

The activity of full-length LINE-1 which can lead to adenoma formation and cancer progression depends on the epigenetic regulation of its promoter [16,17]. The hypomethylation level at the promoter is very important for activation of LINE-1 elements and by this activation and increasing of their expression, the mobility of these elements is increased, tending to copy and paste around the genome and trying to make new insertions all around the genome [18]. This activity makes the genome more unstable, and also provides a setting for cancer progression [19]. LINE-1 promoter hypomethylation has been observed in various cancers [6]. Several studies have analyzed the LINE-1 hypomethylation in gastrointestinal cancers [20]. Hypomethylation of L1 promoter sequences using methylation-sensitive Southern blotting has detected similar hypomethylated status in normal and colon cancer specimens [21]. However, samples taken from colorectal carcinomas with microsatellite instability have shown a significant decrease in LINE-1 methylation in comparison with normal adjacent tissues using pyrosequencing [22]. Hypomethylation of overall LINE-1 sequence in normal colon mucosa has also been associated with poor survival in patients with sporadic colon cancer [23]. The LINE-1 hypomethylation index (LHI) measured by absolute quantitative analysis of methylated alleles (AQAMA) realtime PCR [24] on paraffin-embedded tissue sections treated by in-situ DNA sodium bisulfite modification has shown that LINE-1 is demethylated during the adenomatous and early colorectal cancer stages [25].

In this study, we investigated the relationship between the promoter methylation and the expression of LINE-1 in colorectal non-advanced adenoma, advanced adenoma, and adenocarcinoma samples. Furthermore, to associate the LINE-1 activation with the mobility and insertion of retrotransposons, we analyzed the insertion of the autonomous LINE-1 and nonautonomous Alu and SVA elements in the advanced adenoma stage (not cancerous) after whole-genome sequencing.

Results

LINE-1 promoter is hypomethylated in colorectal non-advanced adenoma, advanced adenoma, and adenocarcinoma

To determine the methylation status of LINE-1 promoter in different stages of early colorectal cancer, the LINE-1 Hypomethylation Index (LHI) was analyzed in colorectal non-advanced (low-grade) adenoma, advanced (high-grade) adenoma, and adenocarcinoma and in their adjacent at-risk (control) tissues. In all three stages, the polyp/tumor tissue samples had significantly lower methylation levels (higher LHI) than their control adjacent counterparts (Figure 1). Also, the methylation levels of LINE-1 promoter in advanced adenoma and adenocarcinoma samples were significantly lower than those in non-advanced adenomas. The finding of LINE-1 promoter hypomethylation in non-advanced adenomas (low-grade polyps) seems to be one of the earliest events for tumor formation (Figure 1).

Increased LINE-1 transcripts in colorectal adenoma, advanced adenoma, and adenocarcinoma

The progressively increased LINE-1 promoter hypomethylation from non-advanced adenoma to advanced adenoma and adenocarcinoma suggested that there could be an association between the progressing levels of promoter hypomethylation and transcriptional activation of LINE-1 during the development of colorectal cancer. RT-qPCR quantification of LINE-1 transcripts revealed that the LINE-1 transcripts in advanced adenoma were

significantly higher than those in non-advanced adenomas, and in the adenocarcinoma, they were significantly higher than those in the advanced adenoma (Figure 2). The LINE-1 transcript levels in adenocarcinomas were significantly higher than those in the adjacent tissues. This significant level of increased transcripts between the polyp and adjacent tissues was not observed in non-advanced and advanced adenomas, while the overall level of LINE-1 transcripts was higher in advanced adenoma and the adjacent tissues than that in non-advanced adenoma and the relevant adjacent tissues (Figure 2).

Somatic insertion of mobile elements in the genome of colorectal advanced adenomas

The advanced adenoma stage is a critical and transient state between the initial polyps (adenomas) and adenocarcinomas. Thus, we decided to analyze the genome of advanced adenomatous tissues to identify the location, size, incidence, and types of mobile elements. For this purpose, MELT analysis was performed on the whole genome sequencing data compared to the reference genome (1000 Genomes Project MEIs). MELT using Burroughs-Wheeler Alignment to align all reads properly to the reference genome and reference insertion list, allowed us to find and compare non-reference insertions. Each new LINE-1 insertion has specific characteristics of retrotransposition including random 5’ inversion, numerous 5’ truncation, a poly(A) tail, and flanking target site duplications (TSDs) [26]. In 6 samples of advanced adenomas, the non-intergenic reference and de novo insertion of LINE-1 elements with different sizes in different chromosomes and genes were identified (Figure 3A). We found approximately 16% de novo insertions of LINE-1 in advanced adenoma samples (Figure 3B; Table 2). One-third of de novo insertions were full length (~6 kb) (Figure 3B). The highest to lowest percentage of LINE-1 insertions were observed in intergenic (64.8%), intronic (30.3%), promoter sequences (3.1%), and transcription terminator regions (1.7%) (Figure 3C).

Alu and SVA elements remain active and are mobile in the genome. We found that similar to LINE-1, the highest and the second-highest incidence of Alu and SVA insertions occurred in intergenic and intronic regions (Figures 4 and 5). In addition, the number of insertions for Alu elements in colorectal advanced adenoma tissues was more than those for LINE-1 and SVA elements. Also, de novo insertions of Alu were markedly higher than those for LINE-1 and SVA elements, and Alu was the only element that showed de novo insertion in the exonic regions (Figure 4B). Three genes that were targeted by Alu in their exons were RUFY1, EZR, and RYR3 in chromosomes 5, 6, and 15, respectively (Figure 4A). While SVA elements had the lowest number of insertions in comparison with LINE-1 and Alu elements, they had the highest percentage of insertions in the transcription termination regions (Figure 5B, C).

Alu insertions within the protein-coding genes

Based on the enrichment analysis of Alu de novo insertions in the affected gene, the GO terms in three categories of molecular function (MF), biological process (BP), and cellular components (CC) were determined (Figure 6). The most important BPs based on the *p*-value for Alu de novo insertions were neuron cellular homeostasis and cell migration. Postsynaptic membrane and cell junction were the most important CCs that were affected by Alu de novo insertions. Transmembrane receptor protein tyrosine kinase activity and calcium channel regulator activity were molecular functions that were significantly affected by Alu de novo insertions (Figure 6).

Discussion

In this study, we investigated the relationship between the promoter methylation and the expression of LINE-1 in three stages of colorectal non-advanced adenoma, advanced adenoma, and adenocarcinoma. In addition, we analyzed the insertion of LINE-1, Alu, and SVA elements in the genome of colorectal advanced adenomas. We found that in all three stages, the polyp/tumor tissue had significantly lower methylated CpG dinucleotides in the LINE-1 promoter than their adjacent control tissues (Figure 1). Also, the LHI in advanced adenoma and adenocarcinoma were significantly higher than those in non-advanced adenomas (Figure 1). The copy number of LINE-1 transcripts in advanced adenoma was significantly higher than that in non-advanced adenomas, and in adenocarcinomas was significantly higher than that in the advanced adenomas (Figure 2). Analysis of the genome of colorectal advanced adenomas revealed that at this stage de novo insertions of LINE-1, Alu, and SVA were approximately 16%, 51%, and 74%, respectively (Figures 3-5).

Numerous cytoplasmic and nuclear mechanisms can restrict or inhibit non-LTR retrotransposon activity and mobilization. These restriction mechanisms are reviewed by Goodier [27]). Regulation of methylation of CpG islands is one of the most important restriction mechanisms that is controlled by several interacting factors to restrict or inhibit LINE-1 activity. In human cells, these factors include DNA methyltransferases (such as DNMT1, DNMT 3a, and DNMT 3b), DNA methyl-binding proteins (such as methyl CpG binding protein 2), Histone methyltransferases (such as SUV39H), DNA repair proteins (such as ATM serine/threonine kinase and ERCC excision repair 4), KRAB-ZFPs and their corepressors (such as PLZF, ZNF91, ZNF93, and KAP1), and transcription factors (such as YY1, P53, RUNX3, SP1, SOX2, RAR, and ETS1) [13,28,37,29–36]. The exact mechanisms to explain the escape of retroelements from the inhibitory mechanisms of cells are not known. For example, the loss of methylation marks at CpG islands in cancer cells could be secondary to a genome-wide hypomethylation event, or a concerted function of effector enzymes, transcription factors, long noncoding RNAs and regulatory proteins working on a specific promoter [38]. The loss or gain of interaction of long noncoding RNAs with DNA methyltransferases could be very important for LINE-1 expresssion. For example, studies have shown that CCAT1 long noncoding RNA which is involved in regulation of proliferation and anchorage-independent growth of cancer cells, and affects other regulatory lncRNAs may function as a scaffold for epigenetic regulators of gene function and determine CpG island methylator phenotype [39–42]. An interesting aspect of LINE-1 function is its relation to pluripotency status of the cell. It has been shown that its transcriptional activation regulates chromatin accessibility in the early mouse embryo, and it is involved in nuclear organization and cellular identity of embryonic stem cells [43,44]. Thus, the activity of LINE-1 may also be secondary to pluripotency and cancer stemness, where transcription factors and methyltransferases cooperate to establish uncontrolled proliferation ability, enhanced potential to self-renew, migration capacity, and differentiation potential into different cell types [45,46].

Alu and SVA are the non-autonomous non-LTR retrotransposons that can hijack LINE-1 retrotransposon machinery, to use the LINE-1-encoded proteins for their mobilization [26,47]. Alu elements are the most successful transposons in the human genome which are transcribed by RNA polymerase III [47]. Studies suggest that the retrotransposition rate of Alu is ten times higher than LINE-1 [48]. We found that the number of total and de novo insertions for Alu elements in colorectal advanced adenoma tissues were more than those for LINE-1 and SVA elements. Alu was the only element that showed de novo insertion in the exonic regions (Figure 4). Three genes that were targeted by Alu in their exons were RUFY1, EZR, and RYR3. RUN And FYVE Domain Containing 1 (RUFY1) gene encodes a protein that contains a RUN domain and a FYVE-type zinc finger domain, playing a role in early endosomal trafficking. A recent study has reported the silencing of RUFY1 expression in

gastric cancer cells [49]. EZR (Ezrin) protein functions as a protein-tyrosine kinase substrate in microvilli and plays a key role in cell surface structure adhesion and migration. Several studies have reported the importance of this molecule in colorectal cancer [50–52]. Ryanodine Receptor 3 (RYR3) is involved in calcium release from intracellular storage. These receptors have also been reported as important players in colorectal cancer [53,54]. In metastatic colorectal cancer, the inadvertent activation of evolutionarily methylation-silenced genes MET, RAB3IP and CHRM3 proto-oncogenes have been identified [55].

Besides the insertional mutagenesis caused by the mobilization of retroelements and the genes that lose their encoding ability, other mechanisms for carcinogenesis of these elements have been postulated. The antisense promoter of human LINE-1 can start the transcription of adjacent genomic sequences generating chimeric RNAs that can perturb transcription of neighboring genes. These chimeric transcripts that have been isolated from breast cancer cell lines and primary tumors and colon cancer cells seem to be unique and of biomarker value for the determination of malignancy [56]. Indeed, the location of insertions being at the promoter, 5’ UTR, exons, introns, 3’ UTR, and transcription termination sequences would be a determining factor. For example, intronic LINE-1 insertion into a regulatory element has increased the expression of the candidate liver oncogene ST18 [57]. Studies have also confirmed that intronic LINE-1 insertions result in decreased expression of the mutated genes [6,58]. This could theoretically contribute to tumorigenesis by decreasing the expression of tumor suppressor genes. In this study, we found that the highest and the second-highest incidence of LINE-1, Alu, and SVA insertions occurred in intergenic and intronic regions (Figures 3-5). SVA had the highest percentage of insertions in the transcription termination regions. All three elements showed 3.1-5.1% insertions in promoters

Conclusions

In summary, we have shown that methylation of CpG islands in the LINE-1 promoter is progressively decreased in each of the three stages of non-advanced adenoma, advanced adenoma, and adenocarcinoma. This stage-to-stage decreased methylation which overlaps onto the course of colorectal cancer was accompanied by a higher level of LINE-1 transcription, suggesting that retrotransposition of LINE-1 and non-autonomous retroelements might be driving cancer. De novo insertions which were found in advanced (high-grade) adenoma, a precancerous stage before colorectal carcinoma, leads us to speculate that the early and advanced polyp stages may host very important pathogenic processes concluding to cancer.

Methods

Polyp and cancer biopsies

An informed consent questionnaire briefly describing the research outline was described for, and filled and signed by each patient. Colorectal biopsy samples and polypectomy specimens were acquired from patients at different hospitals and clinics of Mashhad University of Medical Sciences, Mashhad, Iran. The samples were numbered and in cryovials containing RNA Shield (DENAzist Asia Co., Iran) were transferred to liquid Nitrogen within 30 minutes. Also, the biopsy samples with the same numbers were sent to Mashhad Pathology Laboratory, Mashhad, for histopathological analysis. From each patient, one biopsy sample from the adjacent mucosa was also prepared and processed for deep freezing and histopathological analysis. Portable containers of liquid nitrogen were used to transfer frozen samples to minus eighty freezers at the Research Institute of Biotechnology, Ferdowsi University of Mashhad, for storage and further processing. A total of 5 non-advanced

(low-grade) adenomas, 6 advanced (high-grade) adenomas, and 5 cancer (adenocarcinoma) tissue samples along with their adjacent normal mucosa samples were used in this study.

Genomic DNA isolation and sodium bisulfite modification

Genomic DNA was isolated from adenomas, adenocarcinomas and related adjacent normal tissues using the Animal DNA Isolation Kit (DENAzist Asia Co., Iran). The extracted DNA was subjected to sodium bisulfite modification (SBM). The quality and quantity of extracted DNA and SBM DNA were evaluated using gel electrophoresis and using Epoch 2 nanodrop reader (BioTek Instruments Inc., USA). Bisulfite conversion and subsequent purifications were performed using the EpiTect 96 Bisulfite Kit (Qiagen, GmbH, Hilden, Germany) and according to the manufacturer's protocols.

Evaluation of LINE-1 hypomethylation status

The hypomethylation status of LINE-1 was evaluated by absolute quantitative analysis of methylated alleles (AQAMA) assay [25]. AQAMA requires forward and reverse primers, methylation-specific and unmethylation-specific TaqMan probes (Table 1; Supplementary Figure 1G). A 148-bp fragment at the 5'UTR of the LINE-1 promoter was amplified and subjected to probes to evaluate promoter evaluation. The AQAMA PCR reactions (performed with a Rotor-Gene Q real-time PCR cycler; Qiagen Inc., USA) were performed in triplicates. Each reaction contained 100ng of the SBM DNA template, each primer at 500 nM concentration, and dual-labeled hybridization probes (5'FAM-3'BHQ1-labeled for methylation-specific and 5'HEX-3'BHQ1 for unmethylation-specific probe) at 100 nM concentration in Premix Ex Taq (Probe qPCR) master mix (Takara, Japan). The PCR cycles were 95°C for 5 min, followed by 40 cycles of 94°C for 30 s, 60°C for 40 s, and 72 °C for 30 s. QPCR reactions were repeated to adjust the reaction temperature, the concentration of primers, and to acquire the best amplification curves (Supplementary Figure 1A-D). In the 5' UTR of LINE-1, a cluster of 31 individual CpG sites has been identified in a region stretching across approximately 500 bp [59–62]. Our two AQAMA probes were able to bind the CpG sites 10, 11, and 12.

To make universal unmethylated and methylated controls, normal peripheral blood DNA was used. Fully unmethylated control (UMC) DNA was synthesized by using Phi-29 DNA polymerase (New England Biolabs) and fully methylated control (MC) was synthesized using M.SssI CpG methyltransferase (New England BioLabs Inc., USA) based on the manufacturer's protocol and a previous report [63]. All control DNAs were sodium bisulfite converted and were amplified by PCR using primers for LINE-1 promoter. The size of amplified methylated and unmethylated control fragments was confirmed by gel electrophoresis. Then, these fragments were gel-purified and ligated into the pTZ57R_T cloning vector with the TA cloning strategy. The sequence of constructed plasmids containing methylated and unmethylated control fragments was confirmed by Sanger sequencing. Then, the serial dilutions of both constructs were subjected to AQAMA with methylation and unmethylated specific probes to make standard curves. The absolute copy number of each sample was estimated from two standard curves with known copy numbers of methylated and unmethylated DNA (10^1 to 10^9 copies) (Supplementary Figure 1A and C). The statistical differences for LINE-1 promoter methylation between different groups were tested by the Mann Whitney test.

Quantification of LINE-1 transcripts

From non-advanced adenoma, advanced adenoma, and cancer (adenocarcinoma) tissue samples and their adjacent at-risk tissues total RNA was isolated using the Total RNA Isolation Kit (DENAzist Asia Co., Iran). The quality and quantity of extracted RNA were evaluated using gel electrophoresis and spectrophotometry (Epoch 2, BioTek Instruments Inc., USA). One µg of total RNA was reverse transcribed using random hexamer primers and MMLV reverse transcriptase (Thermo Fisher Scientific, USA). To quantify the level of transcripts for LINE-1, quantitative RT-PCR reactions comprising the RealQ Plus 2x Master Mix Green (containing SYBR Green I; Ampliqon, Denmark), 200 ng cDNA template, and each primer (Table 1) at 500 nM in a 20 µl reaction volume, were carried out in a Rotor-Gene Q real-time PCR cycler (Qiagen, USA). The primers were able to amplify a fragment from cDNA corresponding to 5'UTR of LINE-1. Amplification steps were 95 °C for 15 min, followed by 40 cycles of 94 °C for 30 s, 57 °C for 30 s, and 72 °C for 30 s. The identity of PCR products was confirmed by melt curve analysis, and gel electrophoresis and to check the genomic DNA contamination, the RT-minus reaction for each sample was included.

QPCR reactions were repeated to adjust the reaction temperature, the concentration of primers, and to acquire the best amplification curves (Supplementary Figure 2). Amplified fragments were extracted from agarose gel and after nanodrop spectrophotometry and determination of their serial dilution were used to make quantification standard curves. Each dilution was subjected to triplicate qPCR reactions. The $\log_{10}$ of absolute copy numbers was plotted against the relevant cycle threshold (CT) to draw the standard curve. The efficiency of qPCR was calculated from the slope of standard curves, according to the following equation: $\text{Efficiency}=(10^{-1/\text{slope}}-1)\times 100\%$. The absolute copy numbers for LINE-1 transcripts were quantified based on the standard curve. The statistical differences between differentially expressed groups for the LINE-1 transcript copy number was tested by the Mann Whitney test.

Analysis of LINE-1 insertion

Genomic DNA isolated from 6 advanced adenoma tissue samples and 1 adjacent at-risk tissue was prepared for sequencing on the Illumina HiSeq 2500 platform with the High Throughput Library Preparation Kit (Macrogen Inc., South Korea). DNA was fragmented, and the libraries were prepared using a standard version of the manufacturer's protocol. Libraries were assessed for concentration and fragment size. The libraries were pooled and sequenced on a 100 paired-end Illumina HiSeq 2500 run (Illumina Inc., USA). The samples were sequenced to a depth equivalent to ~30X coverage. The resulting sorted BAM files were used for the mobile element indication.

Mobile Element Insertion (MEI) identification was carried out on WGS Illumina data using the MELT (mobile element locator tool Version 2.1.5) algorithm with default parameters (http://melt.igs.umaryland.edu) to find the insertion of different mobile elements like LINE-1, ALU (a primate SINE) and SVA (SINE-VNTR- Alu). MELT detects Alu, LINE-1, and SVA MEIs by searching for signs of split reads (SRs) and discordant read pairs (DRPs) in Illumina WGS data that are enriched at sites containing new and non- reference mobile element insertions [64]. Different features of insertions including being de novo or reference insertions, their size, chromosomal location, incidence, and target regions (3'UTR, 5'UTR, exonic, intronic, intergenic, promoter, transcription terminator) were compared (15) (Supplementary Figure 3). Insertion sites were manually examined using Golden Helix Genome Browser software (version 7.8.10). The 1000 genomes version of the hg19 human genome reference sequence (available at ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/reference/human_g1k_v37.fasta.gz) was used as a reference sequence to find the new insertions. The circos plots were drawn using shinycircos online

software (shinycircos.ncpgr.cn). De novo insertions are defined as insertions that do not have conformity with the list of reference MEIs (hg19).

Gene Ontology Analysis

Based on Alu de novo (non-reference) insertions in the introns, exons, promoters, terminators, and 3'UTRs of protein-coding genes (PCGs), functional annotation clustering of affected genes with a $-\log_{10}$ *p*-value > 2 was performed using the Gene Ontology category (GO Direct) of the latest released version of DAVID web tool (The Database for Annotation, Visualization, and Integrated Discovery v6.8 Oct. 2016) (http://david.ncifcrf.gov). A *p*-value of < 0.01 was considered statistically significant.

Abbreviations

SVA: SINE-R –VNTR–Alus

ORF: open reading frame

YY1: Yin Yang 1

KRAB-ZFP: Krüppel-associated-box-zinc finger protein

piRISC: PIWI-interacting RNA-induced silencing complex

LHI: LINE-1 Hypomethylation Index

TSDs: target site duplications

MF: molecular function

BP: biological process

CC: cellular component

DNMT: DNA methyltransferases

AQAMA: absolute quantitative analysis of methylated alleles

MEI: Mobile Element Insertion

Declarations

Ethics approval and consent to participate

The ethical approval for this study is issued as the ethical code for grant number 942331 from the National Institute for Medical Research Development of Iran, and by the Committee on Research Ethics of Mashhad University of Medical Sciences (IR.MUMS.REC.1395.228), based on the Ethical Guidelines of Research from The Ministry of Health and Medical Education of Iran, and under the Declaration of Helsinki.

Consent for publication

No identifying patient information is included in this report.

Availability of data and material

The data sets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no conflict of interest.

Funding

This study was financially supported by grant number 942331 from The National Institute for Medical Research Development of Iran (NIMAD), and grant number 41064 from Ferdowsi University of Mashhad, Iran.

Authors' contributions

Conceived and designed the experiments: MS and HD. Performed the experiments: MS, KZ, and MZ. Provided patient biopsy samples: LG, and HMM. Performed histopathological examinations: KG. Analyzed the data: MS and HD. Supervised the experiments: HD. Wrote the manuscript: HD and MS. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Azam Naseri Salanghuch for her great help to categorize the patients' samples and information. We also would like to thank nurses Fahimeh Ahmadi and Mahsa Jalaei for their help for the preparation and preservation of biopsy samples.

References

1. 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 [Internet]. 2017;114:E733–40. Available from: http://www.pnas.org/lookup/doi/10.1073/pnas.1619797114
2. Kazazian HH, Moran J V. Mobile DNA in health and disease. N. Engl. J. Med. 2017.
3. Kemp JR, Longworth MS. Crossing the LINE Toward Genomic Instability: LINE-1 Retrotransposition in Cancer. Front Chem [Internet]. 2015;3:1–9. Available from: http://journal.frontiersin.org/Article/10.3389/fchem.2015.00068/abstract
4. Mathias SL, Scott AF, Kazazian HH, Boeke JD, Gabriel A. Reverse transcriptase encoded by a human transposable element. Science (80-) [Internet]. 1991;254:1808–10. Available from: https://science.sciencemag.org/content/254/5039/1808
5. Hoffmann MJ, Schulz WA. Causes and consequences of DNA hypomethylation in human cancer. Biochem Cell Biol [Internet]. 2005;83:296–321. Available from: http://www.nrcresearchpress.com/doi/10.1139/o05-036
6. Scott EC, Devine SE. The role of somatic L1 retrotransposition in human cancers. Viruses. 2017;9:1–19.

7. Grandi FC, Rosser JM, Newkirk SJ, Yin J, Jiang X, Xing Z, et al. Retrotransposition creates sloping shores: A graded influence of hypomethylated CpG islands on flanking CpG sites. Genome Res. 2015;25:1135–46.
8. Burns KH. Transposable elements in cancer. Nat Rev Cancer [Internet]. Nature Publishing Group; 2017;17:415–24. Available from: http://dx.doi.org/10.1038/nrc.2017.35
9. Sanchez-Luque FJ, Kempen MJHC, Gerdes P, Vargas-Landin DB, Richardson SR, Troskie RL, et al. LINE-1 Evasion of Epigenetic Repression in Humans. Mol Cell. 2019;
10. Wolf G, Greenberg D, Macfarlan TS. Spotting the enemy within: Targeted silencing of foreign DNA in mammalian genomes by the Krüppel-associated box zinc finger protein family. Mob. DNA. 2015.
11. Turelli P, Castro-Diaz N, Marzetta F, Kapopoulou A, Raclot C, Duc J, et al. Interplay of TRIM28 and DNA methylation in controlling human endogenous retroelements. Genome Res. 2014;
12. Jacobs FMJ, Greenberg D, Nguyen N, Haeussler M, Ewing AD, Katzman S, et al. An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. Nature. 2014;
13. Castro-Diaz N, Ecco G, Coluccio A, Kapopoulou A, Yazdanpanah B, Friedli M, et al. Evolutionally dynamic L1 regulation in embryonic stem cells. Genes Dev. 2014;
14. Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science (80-). 2007;
15. Kuramochi-Miyagawa S, Watanabe T, Gotoh K, Totoki Y, Toyoda A, Ikawa M, et al. DNA methylation of retrotransposon genes is regulated by Piwi family members MILI and MIWI2 in murine fetal testes. Genes Dev. 2008;
16. 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;
17. 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.
18. Hancks DC, Kazazian HH. Active human retrotransposons: variation and disease. Curr Opin Genet Dev. 2012;22:191–203.
19. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, et al. Landscape of somatic retrotransposition in human cancers. Science. 2012;337:967–71.
20. Baba Y, Yagi T, Sawayama H, Hiyoshi Y, Ishimoto T, Iwatsuki M, et al. Long Interspersed Element-1 Methylation Level as a Prognostic Biomarker in Gastrointestinal Cancers. Digestion [Internet]. 2018;97:26–30. Available from: https://www.karger.com/Article/FullText/484104
21. Suter CM, Martin DI, Ward RI. Hypomethylation of L1 retrotransposons in colorectal cancer and adjacent normal tissue. Int J Colorectal Dis. 2004;
22. Estécio MRH, Gharibyan V, Shen L, Ibrahim AEK, Doshi K, He R, et al. LINE-1 hypomethylation in cancer is highly variable and inversely correlated with microsatellite instability. PLoS One. 2007;2.
23. Zhuo C, Li Q, Wu Y, Li Y, Nie J, Li D, et al. LINE-1 hypomethylation in normal colon mucosa is associated with poor survival in Chinese patients with sporadic colon cancer. [Internet]. Oncotarget. 2015. p. 23820–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26172297%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4695154

24. de Maat MFG, Umetani N, Sunami E, Turner RR, Hoon DSB. Assessment of methylation events during colorectal tumor progression by absolute quantitative analysis of methylated alleles. Mol Cancer Res [Internet]. 2007;5:461–71. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17510312
25. Sunami E, de Maat M, Vu A, Turner RR, Hoon DSB. LINE-1 hypomethylation during primary colon cancer progression. PLoS One. 2011;6:2–8.
26. Raiz J, Damert A, Chira S, Held U, Klawitter S, Hamdorf M, et al. The non-autonomous retrotransposon SVA is trans-mobilized by the human LINE-1 protein machinery. Nucleic Acids Res. 2012;
27. Goodier JL. Restricting retrotransposons: A review. Mob DNA [Internet]. Mobile DNA; 2016;7. Available from: http://dx.doi.org/10.1186/s13100-016-0070-z
28. Gasior SL, Wakeman TP, Xu B, Deininger PL. The human LINE-1 retrotransposon creates DNA double-strand breaks. J Mol Biol. 2006;
29. Yu F. Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. Nucleic Acids Res. 2001;
30. Gasior SL, Roy-Engel AM, Deininger PL. ERCC1/XPF limits L1 retrotransposition. DNA Repair (Amst). 2008;
31. Becker KG, Swergold G, Ozato K, Thayer RE. Binding of the ubiquitous nuclear transcription factor YY1 to a cis regulatory sequence in the human LINE-1 transposable element. Hum Mol Genet. 1993;
32. Yang Z, Boffelli D, Boonmark N, Schwartz K, Lawn R. Apolipoprotein(a) gene enhancer resides within a LINE element. J Biol Chem. 1998;
33. Tchenio T. Members of the SRY family regulate the human LINE retrotransposons. Nucleic Acids Res. 2000;
34. Yang N, Zhang L, Zhang Y, Kazazian HH. An important role for RUNX3 in human L1 transcription and retrotransposition. Nucleic Acids Res. 2003;
35. Laperriere D, Wang TT, White JH, Mader S. Widespread Alu repeat-driven expansion of consensus DR2 retinoic acid response elements during primate evolution. BMC Genomics. 2007;
36. Martens JHA, O'Sullivan RJ, Braunschweig U, Opravil S, Radolf M, Steinlein P, et al. The profile of repeat-associated histone lysine methylation states in the mouse epigenome. EMBO J. 2005;
37. Puszyk W, Down T, Grimwade D, Chomienne C, Oakey RJ, Solomon E, et al. The epigenetic regulator PLZF represses L1 retrotransposition in germ and progenitor cells. EMBO J. 2013;
38. Zhao Y, Sun H, Wang H. Long noncoding RNAs in DNA methylation: New players stepping into the old game. Cell Biosci. 2016.
39. Liu Z, Chen QJ, Hann SS. The functions and oncogenic roles of CCAT1 in human cancer. Biomed. Pharmacother. 2019.
40. Zare K, Shademan M, Ghahramani Seno MM, Dehghani H. CRISPR/Cas9 Knockout Strategies to Ablate CCAT1 lncRNA Gene in Cancer Cells. Biol Proced Online [Internet]. 2018 [cited 2019 Aug 15];20:21. Available from: https://biologicalproceduresonline.biomedcentral.com/articles/10.1186/s12575-018-0086-5
41. Shademan M, Salanghuch AN, Zare K, Zahedi M, Foroughi MA, Rezayat KA, et al. Expression profile analysis of two antisense lncRNAs to improve prognosis prediction of colorectal adenocarcinoma. Cancer Cell Int. 2019;
42. McCleland ML, Mesh K, Lorenzana E, Chopra VS, Segal E, Watanabe C, et al. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. J Clin Invest. 2016;

43. Percharde M, Lin CJ, Yin Y, Guan J, Peixoto GA, Bulut-Karslioglu A, et al. A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell. 2018;
44. Jachowicz JW, Bing X, Pontabry J, Bošković A, Rando OJ, Torres-Padilla ME. LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. Nat Genet. 2017;
45. Es-Haghi M, Soltanian S, Dehghani H. Perspective: Cooperation of Nanog, NF-κB, and CXCR4 in a regulatory network for directed migration of cancer stem cells [Internet]. Tumor Biol. Feb 29, 2016 p. 1559–65. Available from: http://link.springer.com/10.1007/s13277-015-4690-6
46. Soltanian S, Dehghani H. BORIS: A key regulator of cancer stemness [Internet]. Cancer Cell Int. Dec 5, 2018 p. 154. Available from: https://cancerci.biomedcentral.com/articles/10.1186/s12935-018-0650-8
47. Dewannieux M, Esnault C, Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. Nat Genet. 2003;35:41–8.
48. Ade C, Roy-Engel AM. SINE Retrotransposition: Evaluation of Alu Activity and Recovery of De Novo Inserts. In: Garcia-Pérez JL, editor. Transposons and Retrotransposons [Internet]. New York, NY: Springer New York; 2016. p. 183–201. Available from: http://link.springer.com/10.1007/978-1-4939-3372-3_13
49. Zhi Q, Chen H, Liu F, Han Y, Wan D, Xu Z, et al. Podocalyxin-like protein promotes gastric cancer progression through interacting with RUN and FYVE domain containing 1 protein. Cancer Sci. 2019;
50. Qureshi-Baig K, Kuhn D, Viry E, Pozdeev VI, Schmitz M, Rodriguez F, et al. Hypoxia-induced autophagy drives colorectal cancer initiation and progression by activating the PRKC/PKC-EZR (ezrin) pathway. Autophagy. 2019;
51. Milone MR, Pucci B, Colangelo T, Lombardi R, Iannelli F, Colantuoni V, et al. Proteomic characterization of peroxisome proliferator-activated receptor-γ (PPARγ) overexpressing or silenced colorectal cancer cells unveils a novel protein network associated with an aggressive phenotype. Mol Oncol. 2016;
52. Sethi MK, Thaysen-Andersen M, Kim H, Park CK, Baker MS, Packer NH, et al. Quantitative proteomic analysis of paired colorectal cancer and non-tumorigenic tissues reveals signature proteins and perturbed pathways involved in CRC progression and metastasis. J Proteomics. 2015;
53. Chae YS, Kim JG, Kang BW, Lee SJ, Lee YJ, Park JS, et al. Functional polymorphism in the MicroRNA-367 binding site as a prognostic factor for colonic cancer. Anticancer Res. 2013;
54. Suzuki T, Nishioka T, Ishizuka S, Hara H. A novel mechanism underlying phytate-mediated biological action-phytate hydrolysates induce intracellular calcium signaling by a GαQ protein-coupled receptor and phospholipase C-dependent mechanism in colorectal cancer cells. Mol Nutr Food Res. 2010;
55. Hur K, Cejas P, Feliu J, Moreno-Rubio J, Burgos E, Boland CR, et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of protooncogenes in human colorectal cancer metastasis. Gut. 2014;
56. Cruickshanks HA, Tufarelli C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics. 2009;
57. 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;
58. 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;

59. Nüsgen N, Goering W, Dauksa A, Biswas A, Jamil MA, Dimitriou I, et al. Inter-locus as well as intra-locus heterogeneity in line-1 promoter methylation in common human cancers suggests selective demethylation pressure at specific cpgs. Clin Epigenetics. 2015;
60. El-Maarri O, Walier M, Behne F, van Üüm J, Singer H, Diaz-Lacava A, et al. Methylation at global LINE-1 repeats in human blood are affected by gender but not by age or natural hormone cycles. PLoS One. 2011;
61. Sharma A, Jamil MA, Nuesgen N, Dauksa A, Gulbinas A, Schulz WA, et al. Detailed methylation map of LINE-1 5'-promoter region reveals hypomethylated CpG hotspots associated with tumor tissue specificity. Mol Genet Genomic Med. 2019;
62. El-Maarri O, Becker T, Junen J, Manzoor SS, Diaz-Lacava A, Schwaab R, et al. Gender specific differences in levels of DNA methylation at selected loci from human total blood: A tendency toward higher methylation levels in males. Hum Genet. 2007;
63. Umetani N, De Maat MFG, Mori T, Takeuchi H, Hoon DSB. Synthesis of universal unmethylated control DNA by nested whole genome amplification with φ29 DNA polymerase. Biochem Biophys Res Commun. 2005;329:219–23.
64. Gardner EJ, Lam VK, Harris DN, Chuang NT, Scott EC, Pittard WS, et al. The Mobile Element Locator Tool (MELT): population-scale mobile element discovery and biology. Genome Res. 2017;27:1916–29.

Tables

Table 1. Oligonucleotides used in this study

sv

Gene	Sequence (5’ To 3’)	Product (bp)	Application
LINE-1 Expression Primers (GenBank accession number L19088) (13)	F: TGAGAACGGGCAGACAGACT R: AGGTCTGTTGGAATACCCTGCC	129	RT-qPCR
LINE-1 AQAMA Primers	F: GGGTTTATTTTATTAGGGAGTGTTAGA R: TCACCCCTTTCTTTAACTCAAA	148	qPCR
LINE-1 AQAMA Methylation Specific Probe	FAM-TGCGCGAGTCGAAGT-BHQ1	-	qPCR
LINE-1 AQAMA Unmethylation Specific Probe	HEX-TGTGTGAGTTGAAGTAGGG-BHQ1	-	qPCR

Table 2. De novo (non-reference) LINE-1 insertions in six colorectal advanced adenomas

NCBI-ID	Gene Name	Gene Full Name	Region of Insertion	Insertion Family Type	Size of Insertion (bp)	Sample Count
NR_028067	OR4N3P	olfactory receptor family 4 subfamily N member 3 pseudogene	PROMOTER	L1Ambig	4170	5
NM_001005469	OR5B3	olfactory receptor family 5 subfamily B member 3	PROMOTER	L1Ta	1471	1
NM_001033054	AIPL1	aryl hydrocarbon receptor interacting protein like 1	PROMOTER	L1Ambig	64	1
NM_001270974	HYDIN	HYDIN, Axonemal Central Pair Apparatus Protein	INTRONIC	L1Ta1d	*5984-6019	6
NM_000369	TSHR	Thyroid Stimulating Hormone Receptor	INTRONIC	L1Ta1d	*6017	1
NM_001099771	POTEF	POTE Ankyrin Domain Family Member F	INTRONIC	L1Ta1d	*5986-6010	3
NR_027413	LINC00910	Long Intergenic Non-Protein Coding RNA 910	INTRONIC	L1Ambig	3756	1
NM_145235	FANK1	Fibronectin Type III And Ankyrin Repeat Domains 1	INTRONIC	L1Ambig	2910-2919	7
NM_052900	CSMD3	CUB and Sushi Multiple Domains 3	INTRONIC	L1Ambig	1705	1
NR_002833	DPY19L2P1	DPY19L2 Pseudogene 1	INTRONIC	L1Ambig	83-1683	4
NM_000947	PRIM2	Primase DNA Subunit 2	INTRONIC	L1Ambig	1525	1
NR_135597.1	LOC101928195	-	INTRONIC	L1Ambig	1386	1
NM_001994	F13B	Coagulation Factor XIII B Chain	INTRONIC	L1Ta	604	1
NM_022062	PKNOX2	PBX/Knotted 1 Homeobox 2	INTRONIC	L1Ambig	547	1

NM_001025231	KPRP	Keratinocyte Proline Rich Protein	INTRONIC	L1Ta	218	1
NM_001164315.1	ANKRD36	Ankyrin Repeat Domain 36	INTRONIC	L1Ambig	179	1
NM_001131010	SATB1	SATB Homeobox 1	INTRONIC	L1Ambig	159	1
NM_017812	CHCHD3	Coiled-Coil-Helix-Coiled-Coil-Helix Domain Containing 3	INTRONIC	L1Ambig	147	1
NM_001037175	SUSD4	Sushi Domain Containing 4	INTRONIC	L1Ambig	59	1
NM_033225	CSMD1	CUB And Sushi Multiple Domains 1	INTRONIC	L1Ambig	49	1

* Full-length insertion happened

Figures

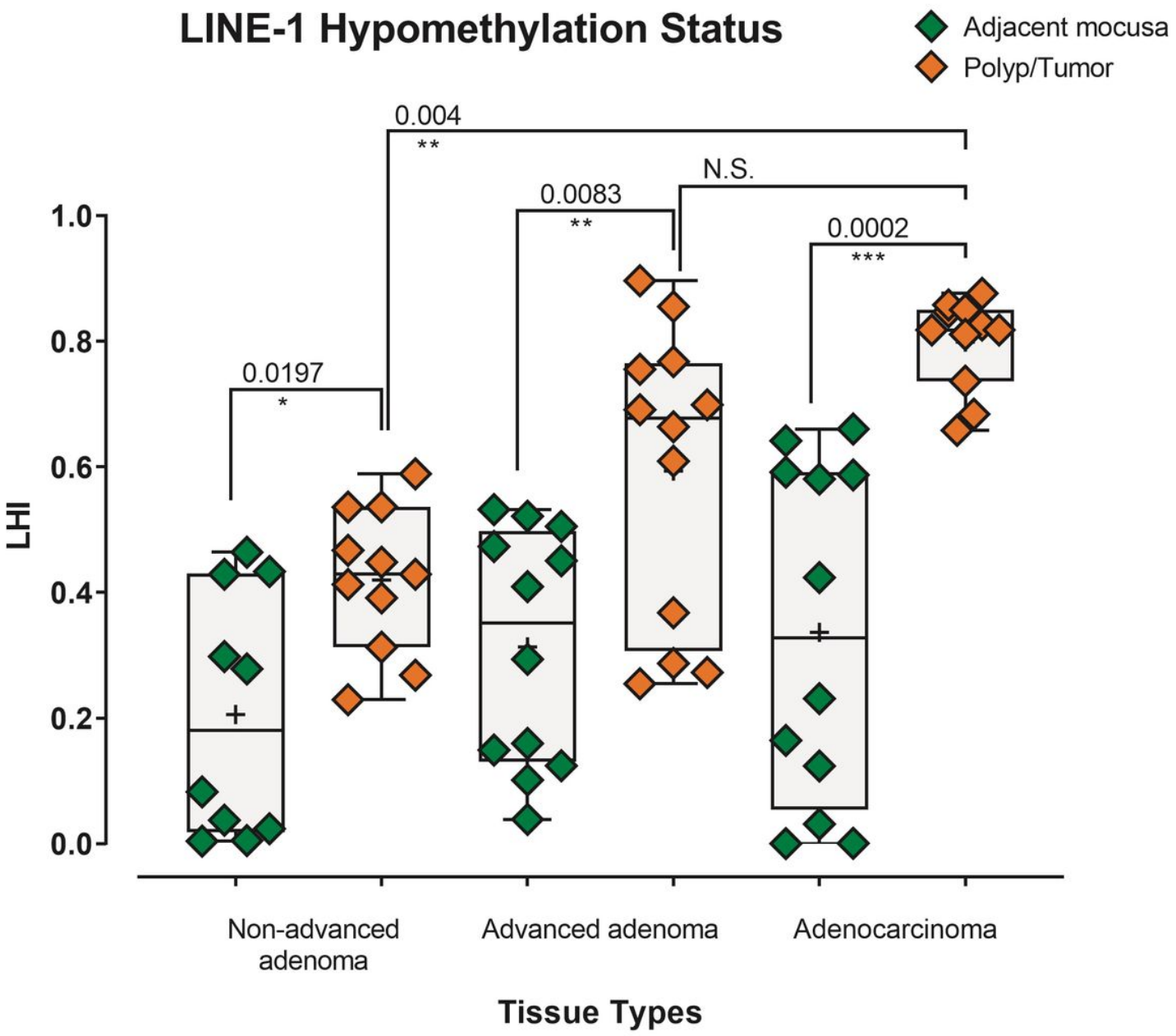

Figure 1

LINE-1 promoter is hypomethylated in colorectal non-advanced adenoma, advanced adenoma, and adenocarcinoma. The univariate scatter/box plot shows the LINE-1 promoter hypomethylation index (LHI) for individual samples in non-advanced adenoma, advanced adenoma, and adenocarcinoma and their adjacent tissues. The Mann Whitney test was used to analyze the significant differences between different groups of samples at the shown p-value. N.S.: not significant.

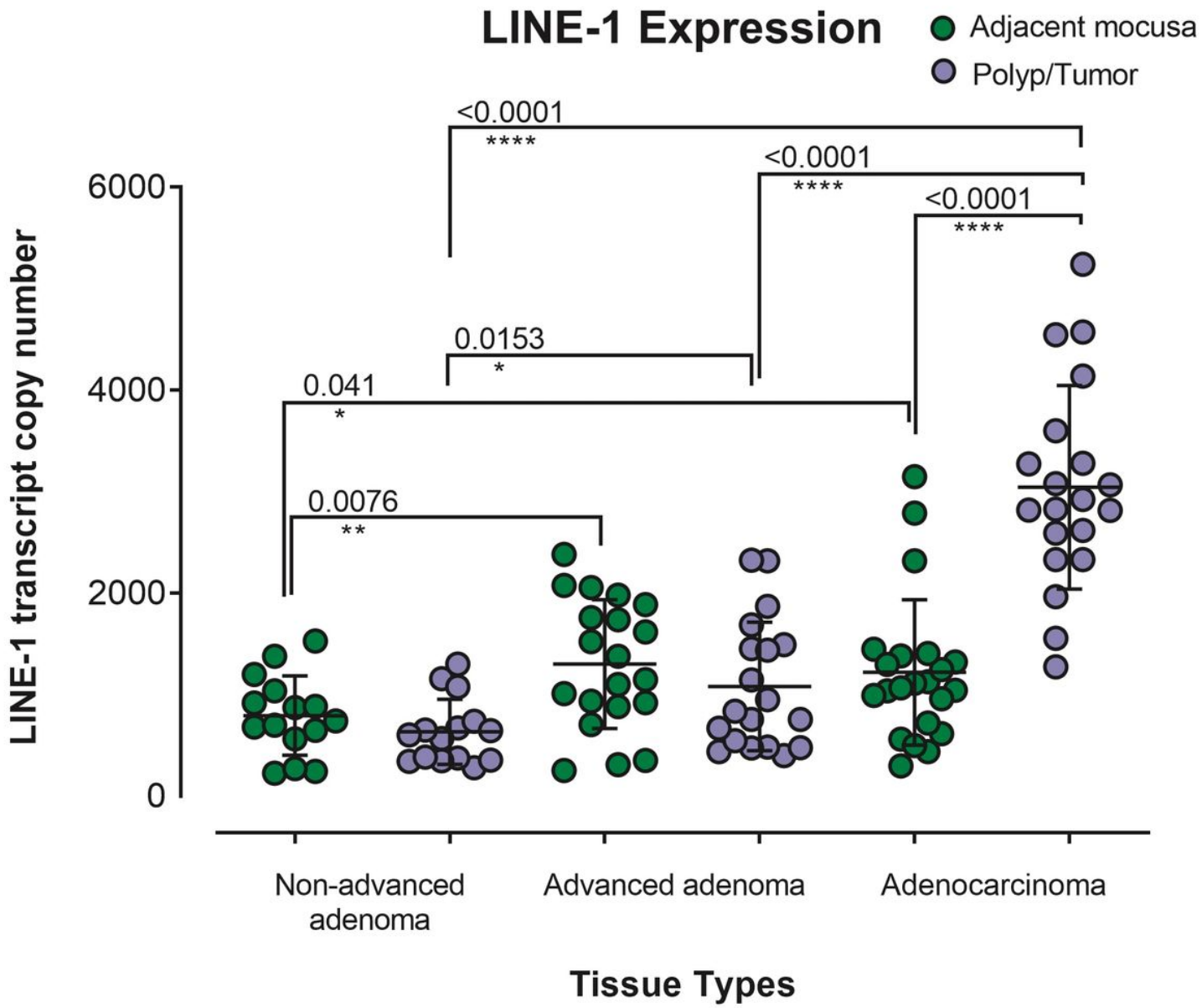

Figure 2

Increased LINE-1 transcripts in colorectal adenoma, advanced adenoma, and adenocarcinoma. The univariate scatter/box plot shows the differential expression of LINE-1 transcript for individual samples in non-advanced adenoma, advanced adenoma, and adenocarcinoma and their adjacent tissues. The Mann Whitney test was used to analyze the significant differences between different groups of samples at the shown p-value.

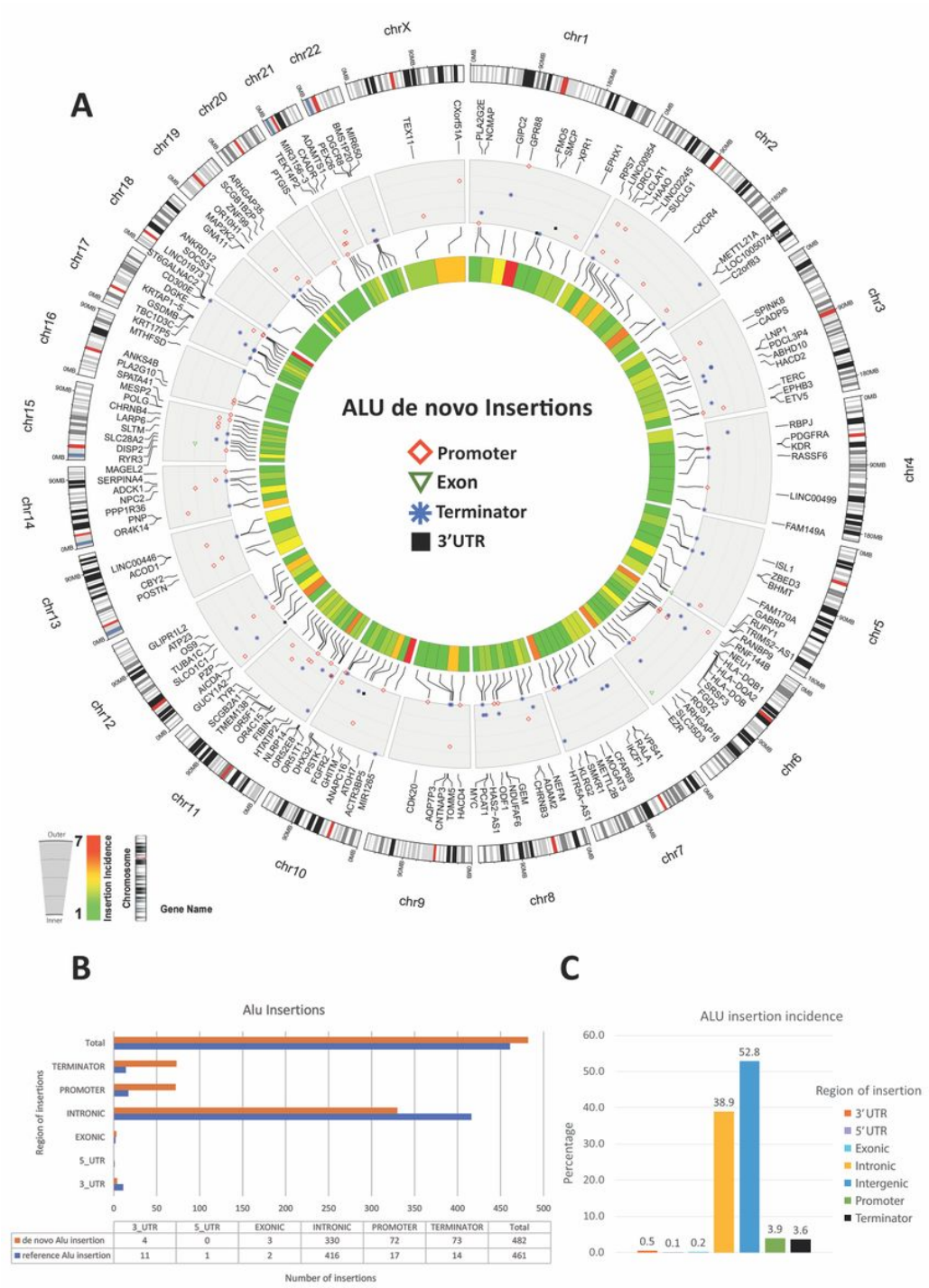

Figure 3

Somatic insertion of LINE-1 elements in the genome of colorectal advanced adenomas. A) The circos plot shows the incidence and location of LINE-1 insertions in six advanced adenoma polyp tissue compared to one normal mucosa. The outer track shows the location of insertions in relation to cytobands of chromosomes. From outside to inside, the second track shows the size and type of insertions (de novo or reference) (the right guide inset). From outside to inside, the third track shows the genes targeted by insertions and target regions (promoters, introns, and transcription termination regions). The innermost heatmap track shows the incidence of insertion for each genomic region (see the left guide inset). B) Length and number of LINE-1 reference, de novo, and de novo full-length insertions. C) Percentage LINE-1 insertions in each of the genomic regions.

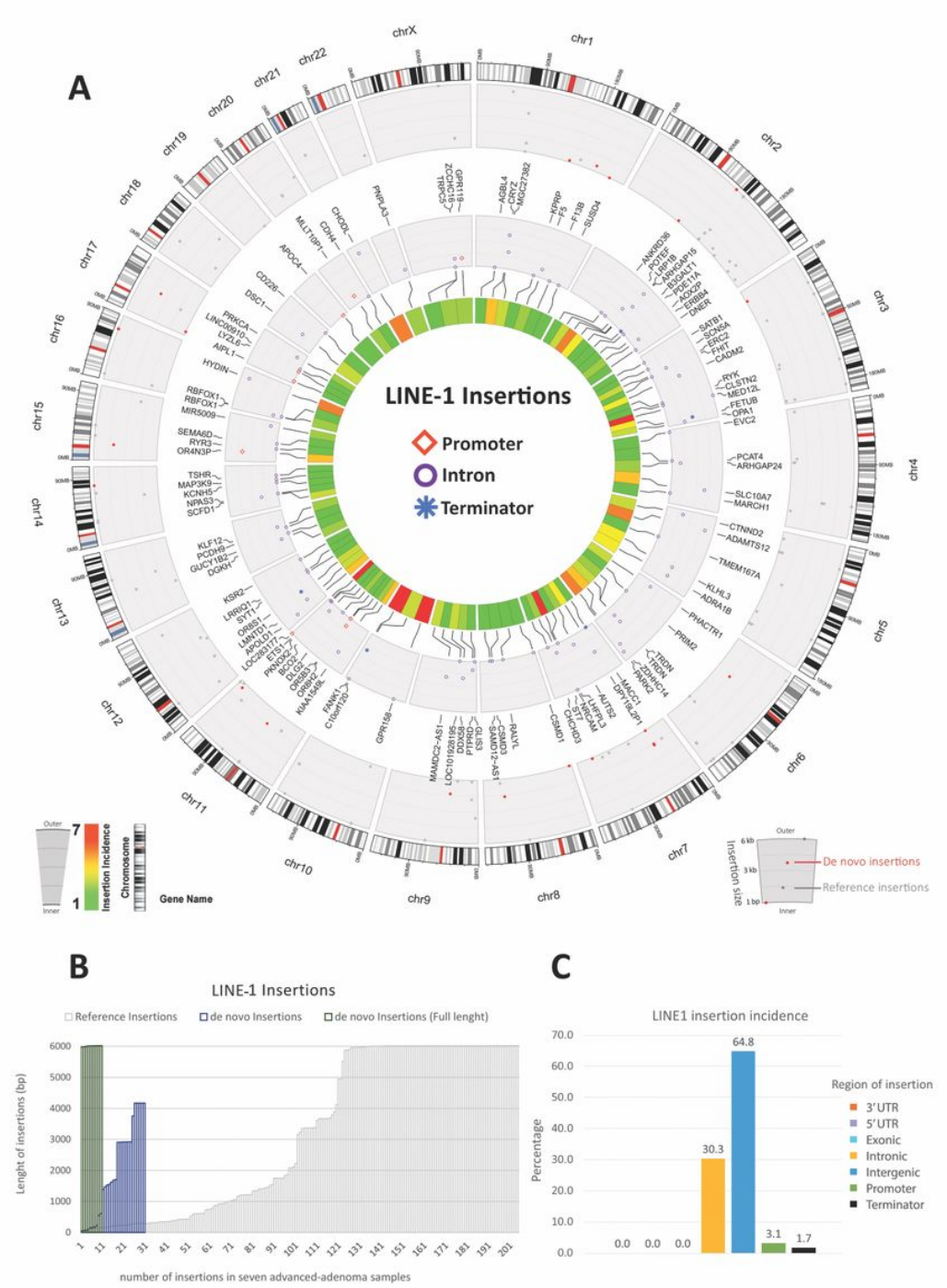

Figure 4

Somatic de novo insertion of Alu elements in the genome of colorectal advanced adenomas. A) The circos plot shows the incidence and location of Alu insertions in six advanced adenoma polyp tissue compared to one normal mucosa. The outer track shows the location of insertions in relation to cytobands of chromosomes. From outside to inside, the second track shows the genes targeted by insertions and target regions (promoters, exons, transcription termination regions, and 3' UTR). The innermost heatmap track shows the incidence of insertion for each genomic region (see the left guide inset). B) The region and number of Alu references and de novo insertions. C) Percentage Alu insertions in each of the genomic regions.

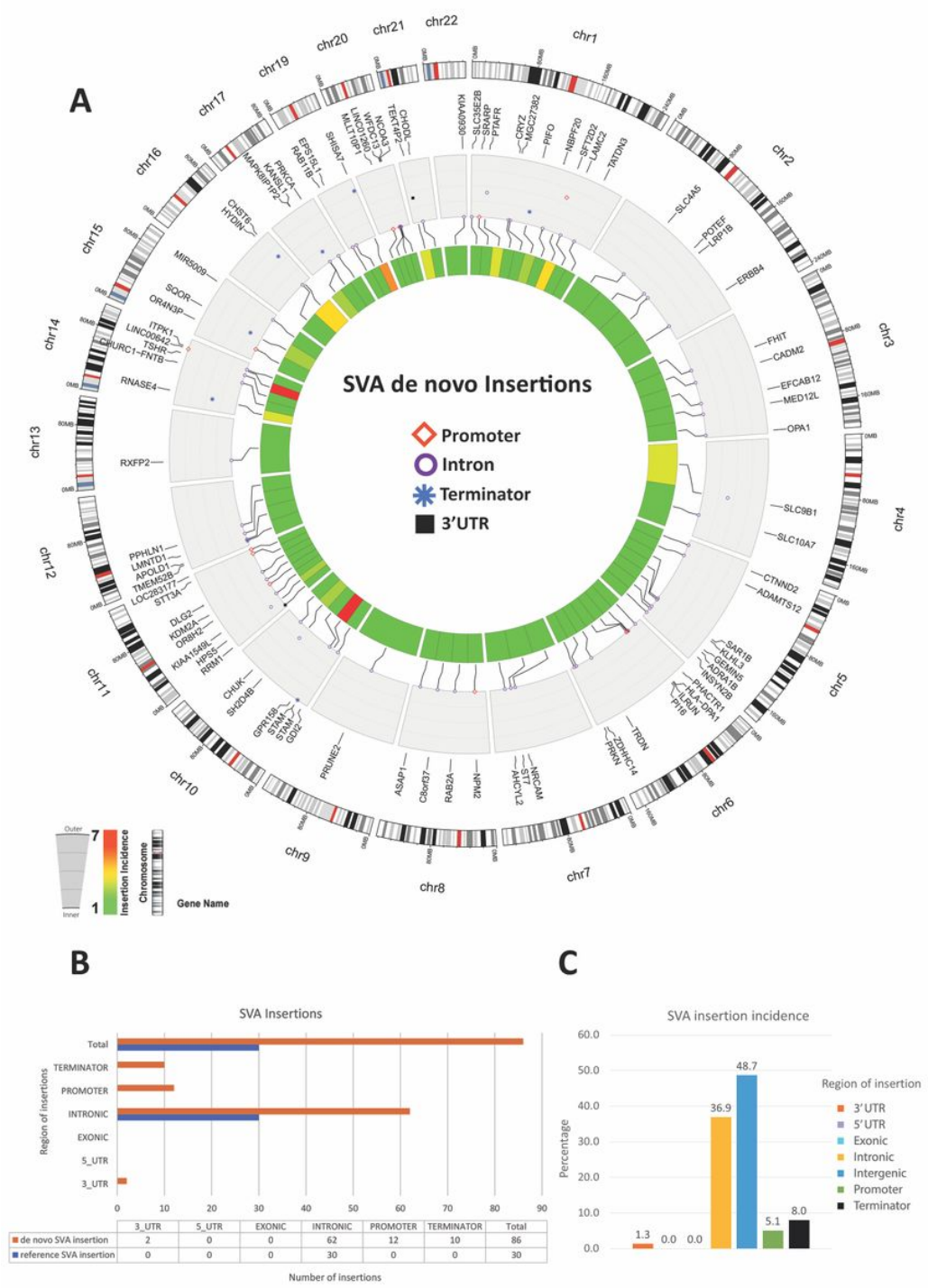

Figure 5

Somatic de novo insertion of SVA elements in the genome of colorectal advanced adenomas. A) The circos plot shows the incidence and location of SVA insertions in six advanced adenoma polyp tissue compared to one normal mucosa. The outer track shows the location of insertions in relation to cytobands of chromosomes. From outside to inside, the second track shows the genes targeted by insertions and target regions (promoters, introns, transcription termination regions, and 3' UTR). The innermost heatmap track shows the incidence of insertion for each genomic region (see the left guide inset). B) The region and number of SVA reference and de novo insertions. C) Percentage SVA insertions in each of the genomic regions.

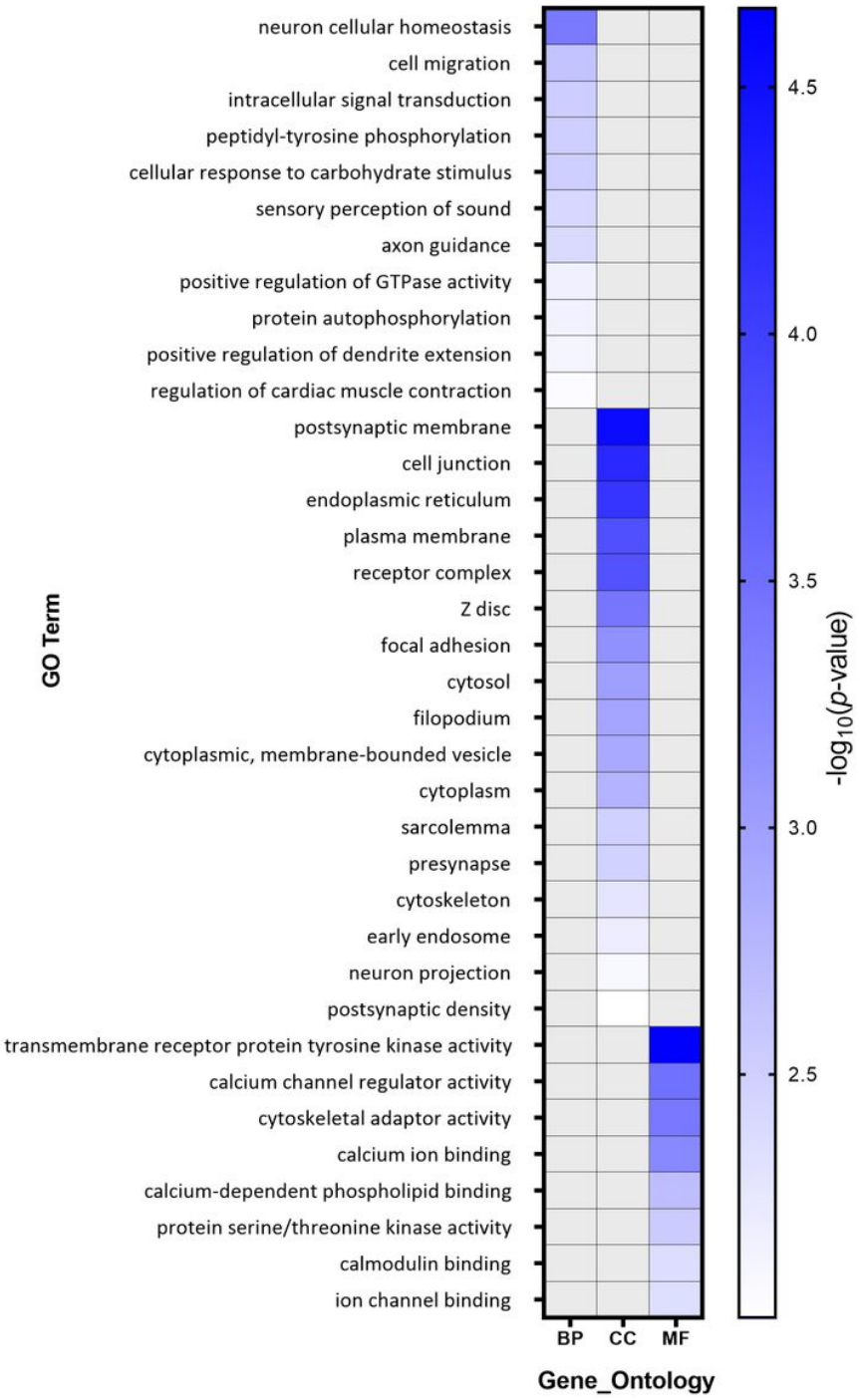

Figure 6

GO terms indicating genes that are affected by Alu retrotransposition in advanced adenoma. GO terms indicating genes that are affected by Alu de novo insertions in three categories of biological processes (BP), cellular component (CC), and molecular function (MF).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplFig3.tif
- SupplFig2.tif
- SupplFig1.tif