**Aberrant gene expression in human lymphomas driven by transcriptional activation of endogenous retroviral promoters**

**Or…**

Artem Babaian, Mark T. Romanish, Liane Gagnier, Stephan Hubold, Jake Lever, Matthew C. Lorincz, Mohammad M. Karimi,Adele Telenius, Christian Steidl, Dixie L. Mager

*(add Steve Jones, others?) (remove or rearrange some people if we don’t do functional analysis of IRF5/RALB)*

**Introduction**

Sequences derived from transposable elements (TEs), including endogenous retroviruses (ERVs), make up a large fraction of the human genome (1). TE insertions are obviously detrimental in certain contexts, but those that have gone to fixation during evolution are assumed to be primarily neutral (ie. junk DNA). While this is likely to be true for the vast majority of the ~4 million TE remnants in the genome, there are clear examples of TE/ERV protein domestication, such as the ERV-derived syncytins involved in placenta development(2), as well as growing reports of exaptation of TE regulatory sequences (3-7). Indeed, TEs have been shown to provide transcriptional regulatory motifs to individual genes (5, 8, 9) and there is increasing evidence that distributed TEs are involved in whole gene regulatory networks (3, 10-13)(**add more?**), a scenario first proposed over 60 years ago (14, 15). One study using CAGE (Cap Analysis of Gene Expression) to map transcriptional start sites (TSSs) in human tissues reported that 10-30% of such sites occur within TEs (16). The long terminal repeats (LTRs) of ERVs, which number ~400,000 in the genome, are particularly well suited to regulatory exaptation since these sequences naturally contain enhancers and promoters and typically remain intact following deletion of ERV internal sequences due to LTR-LTR recombination over time (17).

Although the concept of TE exaptation as a driving force in organismal evolution is becoming increasingly accepted (18) there is also interest in determining the potential role of TEs in human diseases, particularly cancer. Much of the recent focus has been on detection of new somatic insertions of L1 long interspersed elements (LINEs) in human malignancies (19, 20) (21) and on potential carcinogenic roles for human ERV-encoded proteins (22, 23). To date, there have been few investigations on the possibility that TE regulatory sequences could be exploited during oncogenesis. Newly integrated retroviruses have long been known to activate proto-oncogenes via the enhancers or promoters in their LTRs and, indeed, many of the most well studied oncogenes were originally discovered as common sites of retroviral insertion in animal cancer models (24). It is possible that a similar process involving transcriptional activation of normally dormant TEs/LTRs in cancer cells could drive ectopic gene expression or transcription and contribute to somatic evolution of the malignant state – a phenomenon we term “cancer exaptation”. The plausibility of such a scenario is increased in cancer since genome-wide DNA hypomethylation, which is common in tumors (25, 26), is associated with increased transcription of TEs that are generally repressed by DNA methylation in normal somatic cells (27-30).

While the idea of “cancer exaptation” is intriguing, only a few documented cases have been reported to date (31-35), with the example in Hodgkin lymphoma (HL) having the most evidence for a significant effect on disease severity (33). Classical HL is defined by the malignant Hodgkin and Reed-Sternberg (HRS) cells which are known to lose B-cell characteristics and gain expression of non-B lineage genes resembling a macrophage-like signature (36) (37) (38) (39). One such gene, colony stimulating factor 1 receptor (*CSF1R*), is activated inappropriately in HL via the epigenetic derepression of a normally dormant LTR promoter of the ancient THE1B MaLR subfamily (33). This ectopic expression of *CSF1R* is required for survival of HL cell lines and is associated with poorer patient prognosis (33). This study found that other THE1B LTRs are also transcriptionally activated in HL cells but no other significant effects on genes were reported (33). Thus, the genome-wide prevalence of such a phenomenon and the global impact on the cancer transcriptome in HL or any other cancer are largely unknown. In this study, we developed a rigorous computational method, termed “LIONS”, for analyzing whole transcriptome (RNA-seq) data specifically to detect TEs that promote cancer-specific expression of genes or long non-coding RNAs (lncRNAs). We tested and refined this method using ENCODE data (40) and then applied it to HL and diffuse large B cell lymphoma (DLBCL). Our results represent the first comprehensive analysis of TE-promoter activation and resultant transcriptional effects on genes in two human cancer types.

**Results and Discussion**

**Development of a rigorous computational method to identify and quantify TE-promoted chimeric transcripts**

Methods to identify TEs that act as promoters for genes or other transcripts such as lncRNAs rely on the fact that TE-derived sequence will normally form the 5’ end of the resultant “chimeric” transcript, assuming the transcriptional start site (TSS) occurs within the TE itself. Previous “transcriptome-wide” studies to detect TE-derived promoters have therefore analyzed annotated mRNAs (41), ESTs (42), assembled transcripts (6, 43, 44), short CAGE tags (16), PET (paired-end ditag) sequences (45), paired-end RNA-seq data (35, 46) or methods targeted for particular events such as L1-driven transcripts (32). While these methods are useful, they have significant limitations. CAGE provides insufficient information of overall transcript structure and some other methods may not identify the true 5’ end of transcripts or suffer from a high false positive rate due to TE internal exonization events. Moreover, none of the aforementioned studies have attempted to quantify the strength or contribution of the putative TE promoter to overall transcript expression. To quantitatively measure and compare the contribution of TE promoters to normal and cancer transcriptomes we developed a tool that incorporates features of previous methods but significantly builds upon them. Our method, termed “LIONS”, uses paired-end RNA-seq data and a scheme of the bioinformatics workflow is shown in figure 1. This workflow incorporates the following features/steps (*give more details*)….. (further details in Methods and Supplementary Methods).

**Optimization and Evaluation of the method using ENCODE data**

-Encode -ANN?

Discuss Figure 2, differences between K562, H1, GM1278, overrepresentation of LTRs, age of LTRs

**TE-promoter activation in human lymphoma**

Having validated and optimized our computational method, we next applied it to RNA-seq data generated from 12 (*or 7 if we include just the true HL lines which might be better)* Hodgkin lymphoma (HL) cell lines (see Methods or just reference (47)) and to RNA-seq data from a large collection (*how many – 20 or more?)* of primary diffuse large B-cell lymphoma (DLBCL) samples published by Morin et al (48). We have recently examined this DLBCL data using a simpler method (46) to find chimeric transcripts (35) but that method was non-quantitative with a significant false positive rate that did not allow in depth statistical analysis. As a “normal” tissue control for these two cancers of B-cell origin, we used nine RNA-seq libraries generated from germinal center B cells in benign tonsils (48). We first determined the total number of different TE-driven chimeric transcripts passing our thresholds *(need to discuss these)* in each HL, DLBCL and normal B-cell sample. As shown in figure 3A, this number is higher, on average, in the HL and DLBCL samples compared to normal B cells but there is substantial spread among each group. The difference between lymphoma and normal was more striking when just the LTR initiations were considered, with such LTR-driven transcripts being significantly more numerous in the HL and DLBCL samples (figure 3b). When compared to random expectations based on genomic abundance of the different classes of TEs, LTRs are strikingly overrepresented, particularly in the HL and DLBCL samples (figure 3D). (Discuss and present more graphs in supplementary data on different TE/LTR groups, etc.) As shown with the ENCODE data, these results indicate that LTRs are more frequently employed as promoters compared to other classes of TEs. Moreover, this promoter activity is more prominent in HL cell lines and primary DLBCL samples compared to normal B cells. Importantly, these data also demonstrate that this phenomenon is not specific to cancer cell lines but also occurs in primary tumors.

To identify TE promoters that are transcriptionally dormant in normal cells but specifically activated in cancer, we next applied the criteria that a TE-driven chimeric transcript be absent from the normal B-cell libraries. This step eliminates cases where TEs have been exapted during evolution to provide promoter function in normal cells, often as tissue-specific alternative promoters (9). We also applied a “recurrence” threshold and counted only TE promoter activation events that were found in at least 2 of the 12 (**or 7**) HL libraries or at least 10% of the DLBCL libraries and none of the normal libraries. The result was a list of XXX recurrent TE chimeric transcripts in HL and YYY in DLBCL, some of which are sense chimeras with coding genes, some are antisense to genes and many are intergenic transcripts or non-coding RNAs (Tables S1 and S2).

On the top X sense chimeras with coding genes in the HL cell lines (Table S1), we performed RT-PCRs specific for the chimeric transcript on a panel of cell lines including representatives both positive and negative for the chimeric form by RNA-seq (suppl figure). These results were used to judge specificity and sensitivity of the computational method. As shown in figure 3C, for the chimeric transcripts tested, the method has a specificity of 89%, meaning that the large majority of cases detected in the RNA-seq data were detectable by RT-PCR and hence were “true positives”. The sensitivity of the computational method is lower (~55%), meaning that a significant fraction of cell lines deemed negative by RNA-seq, were positive by RT-PCR. This is to be expected since RT-PCR is a more sensitive technique (ref).

**Most active TE subfamilies in HL and DLBCL**

*Stuff to cover: Discussion of THE1Bs in HL and other TE families in HL and DLBCL as appropriate. Discuss similarities and differences in any TE families that are significantly over-represented. Summarize what CSF1R paper found about THEs.*

As mentioned above, Lamprecht et al demonstrated that the *CSF1R* gene is ectopically expressed in HL due to transcriptional activation of an upstream THE1B ERV LTR (33). Moreover, using a RACE approach designed to capture THE1 LTR initiated chimeric transcripts in HL cell lines, they isolated a number of such transcripts initiated by THE1B, THE1A or THE1C LTRs, all members of the THE MaLR superfamily (49). It was also shown that the *CSF1R* THE1B LTR binds NF-kB(33), which is known to be constitutively active in HL (37). In support of a major role of NF-kB in THE1 LTR transcriptional activation, we found that the HL cell line KMH2, which is unusual in having inactivating mutations in at least three NF-kB inhibitors (50), had the highest number of chimeric transcripts initiated by THE1 LTRs (see figure or Table?).

**Genes affected by TE promoter activation**

As mentioned above, several hundred TE-driven chimeric transcripts are recurrently detected in HL and DLBCL and not in the normal B-cell libraries (Tables S1 and S2). YY genes or transcripts, highlighted in the Tables, appear on both lists. *(Say something about the probability of this overlap occurring by chance). Somewhere mention that we found CSF1R as proof of principle. We will need to refer to our DLBCL paper here and discuss a bit. Do some sort of gene ontology analysis since people expect this sort of thing.*

As shown in Table S1, we found several of these genes to be upregulated in microdissected primary HRS cells (39), raising the possibility that their upregulation could be driven by TE activation. *(Not sure this is true for very many but we need to say something. Look at literature for other HL gene expression studies to compare to.)* A notable case is IL1R2, which we and others (39, 51) have reported as one of the highest expressed genes in HL. Figure x shows ....., clearly demonstrating that, indeed, IL1R2 expression in HL is driven from an LTR. *(we could add some of Stephan’s data on IL1R2 in suppl figure here). Should do the same for DLBCL. Need to know if any of the genes we find have been reported to be overexpressed in DLBCL. Could make a column in the tables of genes with this info.*

**Correlations of TE activation patterns with tumor mutations.**

***This is something that is obvious to look at and reviewers may well want to see it. Rita looked at this for our DLBCL paper and didn’t find anything significant but it needs to be done again and more thoroughly with the better pipeline results. This type of section will add meat to the paper if we do not include any functional analysis of genes…***

**Concluding Remarks**

***I think we have decided not to include all the analysis of IRF5 and RALB but below are the “linking” sections we had written. If we do all the above sections, it will still be a very nice paper. Of course, once you get all the results, there will likely be other analyses that we might want to add.***

***Artem’s link section: (If we try to include IRF5 and Ralb but they don’t make Table S1…)***

**The computational analysis provides a specific overview of TEs that initiate potentially functional transcripts [supplementary methods: selection criteria]. To investigate whether there is a biological consequence to chimeric transcription we choose to investigate the candidate genes IRF5 and RALB. Initially identified in our primary chimeric-read clustering screen, these candidates are recurrently chimeric, predicted to encode the complete coding sequence, up-regulated in both RNA-seq and in a primary micro-dissected HL microarray relative to B-cell controls, although IRF5 and RALB do not meet genome-wide significance in the microarray dataset [Steidl, 2012]."**

***Dixie’s attempt at linking (which I still don’t like)***

**Our computational analysis provides a rigorous overview of TEs that initiate potentially functional transcripts [supplementary methods: selection criteria]. As discussed above, we maintained stringent threshold criteria to ensure that the pipeline has a low false positive rate. However, this high stringency results in some cases which we originally detected using a previously published simpler pipeline (ref Mehdi) failing to make our final lists due to complex transcription patterns or failure of ab initio assembly to correctly call the 5’ end of the TE-driven transcript. Two such cases in HL are *IRF5* and *RALB*, which we have chosen to investigate further here due to…. These candidates are recurrently chimeric, predicted to encode the complete proteins and are up-regulated in both the RNA-seq data and in a primary micro-dissected HL microarray relative to B-cell controls, although they did not meet genome-wide significance in the microarray dataset [Steidl, 2012]."**

**Materials and Methods (copied from Artem April 9 version) (much will of course be deleted of we don’t include IRF5 and Ralb)**

*Sequencing, Alignment and Assembly*

*[Steidl protocol?]*

Unaligned RNAseq reads for K562, GM12878 and H1ESC were downloaded from ENCODE (table 2).

RNA sequencing reads were aligned to the human genome (hg19) with tophat2 v.2.0.8 [ref] with the –report-secondary-alignments parameter. The transcriptome was assembled *ab initio* for each library independetly with Cufflinks [ref] with the –u –min-frags-per-transfrag 5 –max-multiread-fraction 0.99 parameters and without a reference gene set. Expression of

*Chimeric discovery pipeline*

*Methodical/Theoretical description*

*ANN*

*ENCODE Cross-validation data*

CAGE transcription start site (TSS) hidden markov model (HMM) clusters of reads were downloaded from UCSC table browser in bed format[ref]. Chromatin State HMM combining 9 ChIPseq signals and input into single chromatin status (i.e. strong promoter, weak/poised enhancer, insulator, heterochromatin...) were downloaded in GTF format from UCSC table browswer. A CAGE cluster or chromatin state was defined as TE-derived if it’s midpoint fell within a TE.

Cell Culture

All Hodgkin’s Lymphoma cell lines were obtained from the DSMZ and cultured by stanard operating procedures. All cultures were not used after 15 passages from starting cultures.

Nucleic Acid Analyses

DNA and RNA was simultaneously extracted from Hodgkin’s cell lines (HDLM-2, KM-H2, L428, L540, L591, L1236, Med-B1 and UH-01) and non-Hodgkin’s cell lines (GM12878, HL60, IM9, Jurkat, K562, NK92, Raji and THP1) using the Allprep kit (Qiagen)and quantified by spectroscopy with a Nanodrop 1000 spectrophotometer (Thermo Scientific). In general, 1 ug of RNA was reverse transcribed using the VILO RT system (Invitrogen) prior to downstream applications.

For 5’ RACE experiments, the FirstChoice RLM kit (Ambion) was employed, however, kit components were replaced for laboratory Superscript III (Invitrogen) at the reverse transcription step. Gene specific primers used in subsequent RT-PCR analysis are listed in Supplementary Table 1. The products were cloned into the pGEMT-vector kit (Promega, Madison, WI, USA). Plasmid preparation and DNA sequencing were performed by Eurofins MWG Operon

RT-PCR

qRT-PCR

Bisulfite conversion was performed on 500ng genomic DNA using the EZ DNA Methylation Kit (Zymo Research) according to the manufacturer's protocol. Converted DNA was used as a template for 35 cycles of 1 round or 2 rounds in a semi-nested PCR reaction with AmpliTaq Gold DNA polymerase (Applied Biosystems) using the primers given in [primer](http://www.jbc.org/cgi/content/full/M111.227637/DC1" \t "pmc_ext) table. Two independent PCRs were performed for each primer pair to eliminate amplification bias of methylated or unmethylated sequences. PCR products were gel-purified (Minelute) and cloned using the pGEM-T Easy Vector kit (Promega). All sequences included in the analyses either displayed unique methylation patterns or unique C to T non-conversion errors (remaining C’s not belonging to a CpG dinucleotide) after bisulfite treatment of the genomic DNA. This avoids considering several PCR-amplified sequences resulting from the same template molecule. All sequences had a conversion rate >96%. Plasmid preparation and DNA sequencing were performed by Eurofins MWG Operon. At least six independent clones were obtained for each region of interest. Data analysis was performed using the QUMA analysis program from RIKEN.

Protein Analyses

For the preperation of protein lysates, 2x106 cells were washed thrice in PBS, resuspended into 100 μl RIPA buffer (0.1% SDS, 1% NP-40, 0.5% sodium deoxycholate in PBS) with Complete protease inhibitor (Roche), homogenized by aspirating through at 21G syringe, incubated on ice for 10 minutes and immediatly stored at -80°C. Upon thawing cell lysates, protein concentration was measured with Bradford reagent (Bio-Rad) and the colometric reaction was measured at 570 nm by Elx808 microplate reader (BioTek). For gel electrophoresis, equal weights of protein were loaded in each lane, ran using the 4-12% Bis-Tris gels and the NuPAGE SDS-PAGE gel system (Life Technologies) and transfered onto PVDF membranes (Millipore). Membranes were blocked with 5% skim milk in TBST for an hour and cut to expected bands. Immunoblotting was performed over-night with agitation in a cold room with monoclonal antibodies mouse anti-RALB (Millipore, cat# 04-037) at 1:1000, mounse anti-IRF5 [AbNova ##] at 1:1500, and rabbit anti-Actin [Sigma ##] at 1:1000 in TBST. Horseradish Peroxidase labelled secondary antibodies anti-rabbit (Santa-Cruz Biotechnology, sc-2005) or anti-mouse [?] at 1:10,000 in TBST were incubated for 1 hour at room temperature with agitation. Protein was visualized with Amersham ECL Western Blotting Analysis System (GE) and developed on BioMax MR Film (Kodak). Protein band intensities were quantified using ImageJ 1.48q software (http://imagej.nih.gov/ij).

Stable Knockdown of Transcripts

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409(6822):860-921.

2. Dupressoir A, Lavialle C, Heidmann T. From ancestral infectious retroviruses to bona fide cellular genes: Role of the captured syncytins in placentation. Placenta. 2012;33(9):663-71.

3. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annual review of genetics. 2012;46:21-42.

4. Wang T, Zeng J, Lowe CB, Sellers RG, Salama SR, Yang M, et al. Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. Proceedings of the National Academy of Sciences. 2007;104(47):18613-8.

5. Rebollo R, Farivar S, Mager DL. C-GATE - catalogue of genes affected by transposable elements. Mobile DNA. 2012;3(1):9.

6. Kapusta A, Kronenberg Z, Lynch VJ, Zhuo X, Ramsay L, Bourque G, et al. Transposable Elements Are Major Contributors to the Origin, Diversification, and Regulation of Vertebrate Long Noncoding RNAs. PLoS genetics. 2013;9(4):e1003470.

7. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D, et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. Nature genetics. 2013;45(7):836-41.

8. Marino-Ramirez L, Lewis KC, Landsman D, Jordan IK. Transposable elements donate lineage-specific regulatory sequences to host genomes. Cytogenet Genome Res. 2005;110(1-4):333-41.

9. Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene. 2009;448(2):105-14.

10. Feschotte C. Transposable elements and the evolution of regulatory networks. Nature reviews Genetics. 2008;9(5):397-405.

11. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nature genetics. 2011;43(11):1154-9.

12. Kunarso G, Chia NY, Jeyakani J, Hwang C, Lu X, Chan YS, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature genetics. 2010;42(7):631-4.

13. Lu X, Sachs F, Ramsay L, Jacques PE, Goke J, Bourque G, et al. The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nature structural & molecular biology. 2014;21(4):423-5.

14. McClintock B. Controlling elements and the gene. Cold Spring Harb Symp Quant Biol. 1956;21:197-216.

15. Britten RJ, Davidson EH. Gene regulation for higher cells: a theory. Science. 1969;165(891):349-57.

16. Faulkner GJ, Kimura Y, Daub CO, Wani S, Plessy C, Irvine KM, et al. The regulated retrotransposon transcriptome of mammalian cells. Nature genetics. 2009.

17. Belshaw R, Watson J, Katzourakis A, Howe A, Woolven-Allen J, Burt A, et al. Rate of Recombinational Deletion among Human Endogenous Retroviruses. The Journal of Virology. 2007;81(17):9437-42.

18. Oliver KR, Greene WK. Mobile DNA and the TE-Thrust hypothesis: supporting evidence from the primates. Mobile DNA. 2011;2(1):8.

19. Solyom S, Ewing AD, Rahrmann EP, Doucet TT, Nelson HH, Burns MB, et al. Extensive somatic L1 retrotransposition in colorectal tumors. Genome research. 2012;22(12):2328-38.

20. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ, et al. Landscape of Somatic Retrotransposition in Human Cancers. Science. 2012;337(6097):967-71.

21. Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. Cell. 2013;153(1):101-11.

22. Hohn O, Hanke K, Bannert N. HERV-K(HML-2), the best preserved family of HERVs: endogenisation, expression and implications in health and disease. Frontiers in Oncology. 2013;3(246):1-12.

23. Chen T, Meng Z, Gan Y, Wang X, Xu F, Gu Y, et al. The viral oncogene Np9 acts as a critical molecular switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting the growth of human leukemia stem/progenitor cells. Leukemia. 2013;27(7):1469-78.

24. Rosenberg N, Jolicoeur P. Retroviral pathogenesis. In: Coffin JM, Hughes SH, Varmus H, editors. Retroviruses Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1997. p. 475-586.

25. De Smet C, Loriot A. DNA hypomethylation in cancer: Epigenetic scars of a neoplastic journey. Epigenetics. 2010;5(3):206-13.

26. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. Nature reviews Cancer. 2011;11(10):726-34.

27. Szpakowski S, Sun X, Lage JM, Dyer A, Rubinstein J, Kowalski D, et al. Loss of epigenetic silencing in tumors preferentially affects primate-specific retroelements. Gene. 2009;448(2):151-67.

28. Ross JP, Rand KN, Molloy PL. Hypomethylation of repeated DNA sequences in cancer. Epigenomics. 2010;2(2):245-69.

29. Romanish MT, Cohen CJ, Mager DL. Potential mechanisms of endogenous retroviral-mediated genomic instability in human cancer. Seminars in cancer biology. 2010;20(4):246-53.

30. Piskareva O, Lackington W, Lemass D, Hendrick C, Doolan P, Barron N. The human L1 element: a potential biomarker in cancer prognosis, current status and future directions. Current Molecular Medicine. 2011;11(4):286-303.

31. Weber B, Kimhi S, Howard G, Eden A, Lyko F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. Oncogene. 2010;29(43):5775-84.

32. Cruickshanks HA, Tufarelli C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. Genomics. 2009;94(6):397-406.

33. Lamprecht B, Walter K, Kreher S, Kumar R, Hummel M, Lenze D, et al. Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nature medicine. 2010;16(5):571-9, 1p following 9.

34. Cruickshanks HA, Vafadar-Isfahani N, Dunican DS, Lee A, Sproul D, Lund JN, et al. Expression of a large LINE-1-driven antisense RNA is linked to epigenetic silencing of the metastasis suppressor gene TFPI-2 in cancer. Nucleic acids research. 2013;41(14):6857-69.

35. Lock FE, Rebollo R, Miceli-Royer K, Gagnier L, Kuah S, Babaian A, et al. Distinct isoform of FABP7 revealed by screening for retroelement-activated genes in diffuse large B-cell lymphoma. Proceedings of the National Academy of Sciences. 2014;111(34):E3534-E43.

36. Schwering I, Brauninger A, Klein U, Jungnickel B, Tinguely M, Diehl V, et al. Loss of the B-lineage-specific gene expression program in Hodgkin and Reed-Sternberg cells of Hodgkin lymphoma. Blood. 2003;101(4):1505-12.

37. Kuppers R, Engert A, Hansmann ML. Hodgkin lymphoma. The Journal of clinical investigation. 2012;122(10):3439-47.

38. Tiacci E, Doring C, Brune V, van Noesel CJ, Klapper W, Mechtersheimer G, et al. Analyzing primary Hodgkin and Reed-Sternberg cells to capture the molecular and cellular pathogenesis of classical Hodgkin lymphoma. Blood. 2012;120(23):4609-20.

39. Steidl C, Diepstra A, Lee T, Chan FC, Farinha P, Tan K, et al. Gene expression profiling of microdissected Hodgkin Reed-Sternberg cells correlates with treatment outcome in classical Hodgkin lymphoma. Blood. 2012;120(17):3530-40.

40. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.

41. van de Lagemaat LN, Landry JR, Mager DL, Medstrand P. Transposable elements in mammals promote regulatory variation and diversification of genes with specialized functions. Trends in genetics : TIG. 2003;19(10):530-6.

42. Nigumann P, Redik K, Matlik K, Speek M. Many human genes are transcribed from the antisense promoter of L1 retrotransposon. Genomics. 2002;79(5):628-34.

43. Huda A, Bushel PR. Widespread Exonization of Transposable Elements in Human Coding Sequences is Associated with Epigenetic Regulation of Transcription. Transcriptomics 2013;1:101.

44. Kelley D, Rinn J. Transposable elements reveal a stem cell specific class of long noncoding RNAs. Genome biology. 2013;13(11):R107.

45. Conley AB, Piriyapongsa J, Jordan IK. Retroviral promoters in the human genome. Bioinformatics. 2008;24(14):1563-7.

46. Karimi MM, Goyal P, Maksakova IA, Bilenky M, Leung D, Tang JX, et al. DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. Cell Stem Cell. 2011;8(6):676-87.

47. Liu Y, Abdul Razak FR, Terpstra M, Chan FC, Saber A, Nijland M, et al. The mutational landscape of Hodgkin lymphoma cell lines determined by whole-exome sequencing. Leukemia. 2014.

48. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476(7360):298-303.

49. Smit AF. Identification of a new, abundant superfamily of mammalian LTR-transposons. Nucleic acids research. 1993;21(8):1863-72.

50. Schmidt A, Schmitz R, Giefing M, Martin-Subero JI, Gesk S, Vater I, et al. Rare occurrence of biallelic CYLD gene mutations in classical Hodgkin lymphoma. Genes, chromosomes & cancer. 2010;49(9):803-9.

51. Ma Y, Visser L, Roelofsen H, de Vries M, Diepstra A, van Imhoff G, et al. Proteomics analysis of Hodgkin lymphoma: identification of new players involved in the cross-talk between HRS cells and infiltrating lymphocytes. Blood. 2008;111(4):2339-46.