*Article*

# Human Endogenous Retrovirus K Rec forms a regulatory loop with MITF that opposes the progression of melanoma to an invasive stage

**Manvendra Singh[1,2#], Huiqiang Cai[1#], Mario Bunse[1], Cedric Feschotte[2] and Zsuzsanna Izsvák*[1]**

[1] Max-Delbrück-Center for Molecular Medicine (MDC), in the Helmholtz Society Robert-Rössle-Strasse 10, 13125 Berlin, Germany
[2] Department of Molecular Biology & Genetics, 526 Campus Road, Cornell University, Ithaca, NY 14853
[#] Equal Contributions
* Correspondence:

**Abstract:** In the human genome, HERV-K(HML2) is the most recently endogenized retrovirus (ERV). While HERV-K(HML2) transcription is observed in normal tissues, both retroviral and the envelope (Env)-derived endogenized accessory products (e.g. Np9 and Rec) are upregulated in various cancer. Still, it is not clear whether the different HERV-K-derived genes contribute to a disease or are simple by-products. Here, we focus on the potential role of Rec in melanoma. Our *in vitro* model as well as high throughput data mining, including single cell transcriptome analyses of patent's material reveals that Rec expression marks the *proliferative* (still controllable) stage of melanoma, and is involved in maintaining a delicate balance between cell proliferation and invasion. Thus, similar to melanocyte-inducing transcription factor (MITF), Rec is a sensitive marker of melanoma progression. Our Rec-knockdown *in vitro* system can faithfully model a subpopulation (MITF malignancy) of melanoma cells in human patients. While both Env, Rec modulate an endothelial-mesenchymal transition (EMT)-like process of cancer progression, they affect the phenotype switch inversely. Rec inhibits the transition the to the *invasive* state by affecting the expression level of some key determinants of the EMT-like process, including MITF that directly binds the LTR5 _Hs of HERV-K. The Hominoid-specific HERV-K products might explain certain species-specific features of melanoma progression, and pinpoint to the limitation of using animal models in melanoma studies.

**Keywords:**

## 1. Introduction

Endogenous retroviruses (ERVs) are remnants of retroviruses that once infected the germline and became vertically inherited as part of the host genome. Sequences derived from various ERVs account for 8% of the human genome [1], reflecting multiple waves of retroviral invasion in the human lineage. In the human genome, HERV-K(HML2) is the most recently endogenized HERV, and several copies of this subgroup still expresses retroviral open reading frames (e.g. Gag, Pro-Protease, Pol-Polymerase, Env-Envelope), and are even capable of producing (non-infective) viral particles [3]. The canonical function of HERV-K Env is to enable the retrovirus to traffic through the endosomal membrane of the host cells or directly enter cells by fusing with their membranes. The endogenization process of HERV-K(HML2) (HERV-K thereafter) included the emergence of the accessory proteins, Np9 and Rec, translated from splicing products of the Env mRNA [3, 4]. Env, Np9 and Rec interact with different host-encoded cellular factors and have distinct functional activities. Rec is a small RNA-binding protein considered to be a functional homolog of the HIV Rev accessory protein [5]. Through its interaction with CRM1 nuclear export factor [4, 6], Rec is involved in exporting un-spliced HERV-K RNA from the nucleus [3, 7]. In addition, Rec can directly modulate cell signalling via binding to the promyelocytic leukemia zinc finger (PLZF) protein, a transcriptional repressor of the c-MYC proto-oncogene [8]. Rec also binds the TZFP (testicular zinc-finger protein) and the hSGT

(human small glutamine-rich tetratricopeptide repeat protein), which are transcriptional repressors of the androgen receptor [9, 10]. While it is evident that Rec must have first evolved to facilitate HERV-K replication, there is speculation that the protein may have been coopted for physiological function, such as spermatogenesis [6, 11] and defence against viral infections in the early human embryo [12].

It has been reported that at least 18 HERV-K genomic loci can be transcribed in healthy human tissues and express potentially-coding Rec or Np9 mRNAs (alone or in combination with retroviral genes), though the landscape of transcribed HERV-K loci differs considerably between cell/tissue types [13]. Rec RNA/protein expression has been observed in human embryonic tissues, placenta and retina [14, 15]. Interestingly, HERV-K is expressed in both germinal- and pluripotent stem cells, providing a special link between these cell types [16]. While HERV-K transcription has been observed at low levels in various normal tissues, it is strongly induced upon environmental stress, such as ultraviolet irradiation (e.g. UVB and UVC), starvation or viral infection that may manifest the global DNA hypomethylation [17-21]. Furthermore, many studies have reported robust mRNA upregulation of several genomic loci of HERV-K in various disease states, including certain autoimmune diseases and a number of cancers, especially lung, breast cancers, germ cell tumours and melanoma [18, 22, 23] and reviewed in [24, 25]. Moreover, transcriptional activation and translation of various HERV-K products [22, 26] and even virus-like particles have been observed in certain cancer cells such as teratoma, embryonic carcinoma and melanomas [3, 27-29].

Whether and how the overexpression of HERV-K products may contribute to the etiology or progression of these diseases has been the subject of many studies, debates, and speculations. Several mechanisms have been proposed by which HERV-K might contribute to disease. The activated LTRs could act as alternative promoters and de-regulate tumor suppressor genes or proto-oncogenes (reviewed in [35, 36]). Furthermore, the Env protein of HERV-K, via its fusogenic property, is capable to induce cell-cell fusion and could therefore contribute to tumour invasiveness [18, 37, 38]. Env could also have oncogenic properties through direct interference with cellular signalling pathways (e.g. RSS/MEK/ERK) [39, 40]. Indeed, the overexpression of the HERV-K Env protein has been reported to induce an epithelial to mesenchymal transition (EMT)-like process [39, 41], a crucial event in oncogenesis leading to more malignant phenotype. In addition, Env was suggested to promote tumorigenesis via modulation of the immune response [42-44]. Confusingly, HERV-K may have either positive or negative effect on the immune system [45, 46]. Similar to Env, the expression of the accessory genes, Rec and Np9 were readily increased in pathological conditions [47], including germ cell tumours and melanoma [4, 6, 22, 27, 28, 48-52], but again it was not clear whether and how these factors could contribute to tumorigenesis and other disease states.

To begin answering this question, we investigate the oncogenic properties of HERV-K Rec (Rec thereafter) in melanoma. We used *in vitro* models as well as high throughput data mining, including single cell transcriptome analyses of patient samples, to examine the impact of Rec expression in the spatiotemporal progression of melanoma. We found that Rec marks the *proliferative* state of melanoma, and similarly to Env [39] modulates the EMT-like process of cell transformation. However, surprisingly, in contrast to Env [39], Rec is inhibiting and not activating the transition from proliferative to invasive state, and thus might be in fact a protective factor in melanoma.

## 2. Materials and Methods

### *2.1. Construction of the KD constructs*

All 3 KD targeting the ERVK6(HML2.HOM) locus. KD1 targets 3′UTR, KD2 3′UTR, and KD3 the gene body of Rec.

KD1 construct 1 (3'UTR_I): (as) 5-'ATCCATTCAACTCTGAGTGGA-3'
KD2 construct 2 (5'UTR_I): (as) 5'-TAAGGCTGACTGTAGATGTAC-3'
KD3 construct 4 (Rec): (as) 5'-CAACGGTGCTCGATTGCGATG-3'

### *2.2. Cell Culture*

The melanoma cell line, A375 (ATCC CRL-1619) was cultivated in RPMI1640 (Thermo Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; PAA Laboratories, Pasching, Austria), 80 U/mL penicillin and streptomycin (Lonza, Basel, Switzerland) at 37°C in a humidified atmosphere of 95% air and 5% $CO_2$. The melanoma cell line, SKMel-28 (ATCC HTB-72) was maintained in Dulbecco's modified Eagle's medium (DMEM, Life Technologies, England) containing GlutaMAX™-I supplement (Gibco, catalogue number: 31966-021), with 10% FBS and penicillin and streptomycin in the same condition as above.

### *2.3. Transfection of melanoma cells and cell sorting*

To knock down Rec, melanoma cells were transfected with pT2-MP71-KD1-Rec-EGFP, pT2-MP71-KD2-Rec-EGFP, pT2-MP71-KD3-Rec-EGFP plasmids in the presence of pT2-CAGGS-SB100X, using the Neon transfection system (Life technologies). A piggyBac transposon-based vector, targeting GFP was used as a control. 10µl Kit with the electroporation setting, 1350 V (pulse voltage), 20ms (pulse width), 2 pulses. $2\times10^5$ cells in 11 µL resuspension buffer R were combined with 2 µL of purified plasmid mixture (50 ng transposase and 500 ng transposon). Transfected ells were transferred to 6-well tissue culture plates containing 2 mL culture medium. EGFP+ cells were sorted at week 1 and 3 post transfection, by FACSAria II cell sorter (Becton).

### *2.4. Quantitative Real-Time PCR and semi-quantitative PCR*

Total RNA was extracted using Direct-zol RNA MiniPrep kit (Zymo Research) according to the protocol of the manufacturer. Reverse transcription was performed from 1µg total RNA with the High Capacity RNA-to-cDNA kit (Applied Biosystems). The isolated RNA would be treated with additional DNaseI (Invitrogen), to get rid of potential DNA contamination. Quantitative real-time PCR (qRT-PCR) was carried out on ABI7900 with PowerUp SYBR Green Master Mix kit (Applied Biosystems), according to the recommendations of the manufacturer. For quantification of mRNA expression levels, real-time PCR were run in triplicates for each cDNA sample, using GAPDH and 18s RNA as the internal controls for SKMel-28 and A375 cells, respectively. For primers and annealing temperatures see **Table 1**. Data was analysed using the comparative CT (2-ΔΔCT) method, which describes relative gene expression.

**Table 1.** The primers/oligos used in this study.

| Name | Application | Sequence (5′ – 3′) | Tm(°C) |
|---|---|---|---|
| HML-2-Env | Real-Time qPCR | F: GCTGCCCTGCCAAACCTGAG<br>R: CCTGAGTGACATCCCGCTTACC | 60 |
| HML-2-Gag | Real-Time qPCR | F: AGCAGGTCAGGTGCCTGTAACATT<br>R: TGGTGCCGTAGGATTAAGTCTCCT | 60 |
| HML-2-Np9 | Real-Time qPCR | F: AGATGTCTGCAGGTGTACCCA<br>R: CTCTTGCTTTTCCCCACATTTC | 60 |
| HML-2-Rec | Real-Time qPCR | F: ATCGAGCACCGTTGACTCACAAGA<br>R: GGTACACCTGCAGACACCATTGAT | 60 |

| | | | |
|---|---|---|---|
| MITF-M | Real-Time qPCR | F:ATGCTGGAAATGCTAGAATATAATCACT<br>R: GAATGTGTGTTCATGCCTGG | 60 |
| SPANXB1/B2 | Real-Time qPCR | F: AGGCCAATGAGGCCAACAAGAC<br>R: TCCTCCTGTAGCGAACCACTAG | 60 |
| 18S RNA | Real-Time qPCR | F: GTAACCCGTTGAACCCCATT<br>R: CCATCCAATCGGTAGTAGCG | 60 |
| GAPDH | Real-Time qPCR | F: CAATGACCCCTTCATTGACCTC<br>R: AGCATCGCCCCACTTGATT | 60 |

*2.5. RNA sequencing and data analysis*

Concentration of RNA was measured on NanoDrop Spectrophotometer ND-1000 and the quality of RNA was analysed using Aglient RNA 6000 Nano Kit on Aligent 2100 Bioanalyzer machine. The RNA sequencing library was prepared from 550 ng of RNA, using Illumina TruSeq Stranded mRNA LT Set A kit (Cat. no. RS-122-2101), according to TruSeq Stranded mRNA Sample Prep LS Protocol. Sequencing was performed on an Illumina HiSeq 2000 platform as 100 bp first strand specific paired-end reads.

Sample-specific barcoded sequencing reads were de-multiplexed from multiplexed flow cells and by using CASAVA. 1.8.2 BCL files were converted to FASTQ format files. Quality of raw sequence reads was determined by using the FastQC. Reads with quality score below 30 were removed. We removed the highly variable 2 nt from the ends of the remaining sequencing reads, and mapped them over the reference genome (Human hg19/GRCh37) and transcriptome model (hg19.refseq.gtf). hg19/GRCh37 and hg19.refseq.gtf that were downloaded from USCS tables (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/) by using Bowite 2.0.5.0, Samtools 0.1.17.0 and TopHat v2.0.8 with parameters: "tophat2 -p 8 –r 150--mate-std-dev 140 --library-type fr-firststrand". Transcript assembly for each individual sample was conducted with Cufflinks v2.08, generating FPKM values. In the cross-study or/and cross-platform comparisons we normalized the batch effect using the 'combat' package from R. Differentially expressed genes (DEGs) were calculated as counts per million (CPM) using featureCount and algorithms from "DESeq2". This strategy provided both quantification and statistical inference of systematic changes between conditions (with at least three replicates). We had generated various transcriptomes from A375 cells (e.g. A375-untransfected, A375-scambled transfection control, A375-knockdowns targeting HERV-K(HML2) Rec using three distinct RNAi constructs, Rec-KD1,2,3). We performed the similar studies in SK-MEL28 cell lines. As an additional control, and to increase statistical power for DEGs identified in our transcriptomes upon Rec-KD, we also used publicly available A375 RNA-seq datasets (GSE110948). To analyze single-replicate RNA-seq datasets, we utilized the "GFOLD" algorithm for normalization, which calculates a variance of fold-changes from unreplicated RNA-seq data. Canonical pathways and biological function of the differentially expressed genes (DEGs) were further subjected to KEGG and Gene Set Enrichment Analyses (GSEA).

*2.6. Single cell RNA-seq data processing*

We got the single cell count matrix from GSE72056. We calculated activity of genes in every cell at TPM expression levels. We considered samples expressing over 1000 genes with expression level exceeding the defined threshold (Log2 TPM > 1). We used Seurat_3.1.1, a package from R to normalize

the datasets at logarithmic scale using "scale.factor = 10000". After normalization, we calculated scaled expression (z-scores for each gene) for downstream dimension reduction. We used the original annotations of datasets to classify the malignant, non-malignant and heterogenous cell-types. To define cell population clusters, we employed the FindClusters function of "Seurat" using "PCA" as a reduction method. The specific markers for each cluster identified by "Seurat" were determined by the "FindAllMarkers" function, using "roc" as a test for significance. This provided us two lists of gene sets, 1) Malignant vs Non-Malignant differentially expressed genes and 2) genes differentially expressed between MITF-High and MITF-low tumours, which we applied for comparison with Rec-KD RNA-seq DEGs. Feature plots, violin plots and heatmaps were constructed using default functions, except for the colour scale that was set manually. The annotated cells were re-clustered using the methodologies described above and visualized on the UMAP coordinates.

### *2.7. ATAC-seq and ChIP-seq data analyses*

ATAC-seq and ChIP-seq raw datasets in sra format were downloaded from the listed studies, and converted to fastq format using sratools function fastq-dump –split-3. Fastq reads were mapped against the hg19 reference genome with the bowtie2 parameters: --very-sensitive-local. All unmapped reads, reads with MAPQ < 10 and PCR duplicates were removed using Picard and samtools. All the ATAC-seq peaks were called by MACS2 with the parameters –nomodel -q 0.01 -B. Blacklisted regions were excluded from called peaks (https://www.encodeproject.org/annotations/ENCSR636HFF/). To generate a set of unique peaks, we merged ATAC-seq peaks using the mergeBed function from bedtools, where the distance between peaks was less than 50 base pairs. We then intersected these peak sets with repeat elements from hg19 repeat-masked coordinates using bedtools intersectBed with 50% overlap. To calculate the enrichment over the given repeat elements, we first extended 5KB upstream and 5KB downstream coordinates from the left boundary (TSS) of respective elements in a strand-specific manner. These 10KB windows were further divided into 100 bps bins and tags (bedGraph output from MACS2) were counted in each bin. Tag counts in each bin were normalized by the total number of tags per million in given samples and presented as Counts Per Million (CPM) per 100 bps. We averaged CPM values for each bin between replicates before plotting the figures. To find the TF binding motifs, 25 bps sequences were extended from either side of MITF ChIP-seq peak summits. The extended sequences were analyzed by RSAT tool (http://rsat.sb-roscoff.fr/). The TF binding motifs were calculated from JASPAR libraries of human TF motifs.

### *2.8. Cell invasion assay*

The cell invasion assays were performed using transwell chambers (8 μm pore size; Corning Costar) according to the vendor's instructions. Briefly, the insert of the wells was first coated with 50 μl BD matrigel. A375 cells were resuspended with RPMI1640 medium containing 0.5% FCS to reach a concentration of $5\times10^5$ cells/ml. Then 100 μl cells from each group were seeded into the upper chamber of the insert, with adding DMEM containing 10% FCS to the lower chamber. After incubation at 37 °C in a humidified atmosphere of 5% $CO_2$ for 24 h, non-invaded cells were removed with gentle swabbing while the invaded ones were stained, and then images were taken with a microscopy. Cell numbers were calculated with ImageJ.

### *2.9. Accession of datasets used in this study:*

Encode chromHMM datasets:
ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeBroadHmm
Encode TF ChiP-seq datasets:
ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs
GSE60663: H3K27Ac Primary Melanoma
GSE60663: H3K4Me3 Primary Melanoma
GSE60663: MITF ChIP-seq Primary Melanoma

GSE60664: RNA-seq Primary Melanoma
GSE110948: A375 RNA-seq
GSE46817: RNA-seq of 7 distinct melanoma lines
GSE46805: RNA-seq Melanocyte BRAF overexpression
GSE50681: MITF ChIP-seq Melanoma and Melanocyte post-treatment
GSE82330: ATAC-seq Melanoma A375
GSE72056: Patient's Single cell RNA-seq data

## 3. Results

### *3.1. LTR5_Hs loci form active chromatin in cancer and pluripotent cell lines*

The regulatory region of the youngest subfamily of HERV-K is the human-specific long terminal repeat 5 (LTR5_Hs). To dissect the regulation of HERV-K transcription in different cell types, we adopted a systematic genome-wide approach using publicly available data. For this, we investigated the chromatin state at ~600 full-length LTR5_Hs sequences using the uniquely mappable reads in 9 different cell types from the ENCODE project [50]. Analysis of the chromatin states suggested that a substantial fraction (n = 227) of LTR5_Hs sequences overlap with active or permissive chromatin states (**Figure 1**A). When compared to differentiated somatic cells, these LTR5_Hs genomic loci showed features of active ('open') chromatin state in human embryonic stem cells (H1_ESCs) and in cancer cell lines (**Figure 1**A)**.**

Various transcription factors (TFs) have been reported to bind LTR5_Hs/HERV-K, including the pluripotency factors such as OCT4 [12, 51] or the melanocyte inducing TF (MITF) [52]. In a systematic approach, we mapped the peaks from 142 ChIP-seq TF datasets from 7 cell lines (ENCODE) over LTR5_Hs genomic loci (n ~ 600). Our integrative analysis of various TF occupancy revealed that in H1 embryonic stem cells (H1_ESCs) and leukemia K562 cells, LTR5_Hs recruits significantly more TFs than in HepG2, GM12878 or HUVEC cells (**Figure 1**B), which may indicate elevated levels of HERV-K transcription in embryonic stem cells and cancer cells. While in H1_ESCs, LTR5_Hs was bound prominently by pluripotency specific TFs (e.g. NANOG) (**Figure 1**B-C), the cancer line (K562) displays the binding of multiple TFs that control the cell cycle and proliferation (e.g. c-MYC, MAX, COREST, P300, PU1, GATA2 and JUND) (**Figure 1**B-C and **Table S1**). Intriguingly, NANOG binding over LTR5_Hs is overrepresented in 3iL naïve conditions (**Figure 1**C). The observed HERV-K virion particles in the human blastocyst might be the products of HERVK driven by NANOG [12]. The active chromatin states and TF bindings over LTR5_Hs in cancer lines might be associated with the oncogenic phenotype.

### *3.2. MITF regulated LTR5_Hs/HERV-K expression is a hallmark of the 'proliferative' type of melanoma*

It has been previously reported that in melanoma, the melanocyte/melanoma-specific isoform, MITF-M may bind LTR5_Hs, and promotes the expression of HERV-K transcripts [52]. MITF(-M) is a critical transcription factor associated with melanoma progression [53, 54], and is a sensitive marker of melanoma invasiveness: it is highly expressed in the melanocytes and the *proliferative* state of melanoma (e.g. MITF-high) but lowly expressed in the *invasive* state (e.g MITF-low) [55].

To investigate how HERV-K is controlled in melanoma, we mined ChIP-seq and RNA-seq data from melanocytes, various melanoma cell lines and primary melanoma cultures (**Methods**). Importantly, characterizing primary melanoma cultures according to their MITF levels (**Figure 1**D) revealed a positive correlation between MITF-regulated LTR5_Hs/HERV-K transcription and invasiveness (**Figure 2**A). To see if additional transposable elements (TEs) respond to melanoma proliferative/invasive status, we performed a more systematic analysis, including other TE families. This analysis revealed that the expression of certain ERVs, including LTR5_Hs/HERV-K, LTR2C and LTR13 was correlated with that of a set of genes associated with the *proliferative* state of melanoma (**Figure S1**A). Nevertheless, when compared to the overall expression of other TE families, LTR5_Hs/HERVK was the most abundantly expressed TE family, and highly correlated with MITF level (rho = 0.43, corrected p-value < 0.2e-9) in the melanoma primary cultures (N=10) (**Figure 2**B).

These data suggest that LTR5_Hs/HERV-K expression can be used as a marker that correlates negatively with melanoma invasiveness. Next, we characterized LTR5_Hs loci activated in the *proliferative* type of melanoma cells by layering ChIP-seq occupancy signals of H3K4Me3, H3K27Ac and MITF (**Figure 2**C and **S1**B-C). We found that MITF binding at LTR5_Hs loci coincides with the presence of the canonical E-box motif (CA(C/T)GTG), which is recognized by MITF as well as c-MYC, and MAX TFs [56] (**Figure 2**C).

We also investigated the regulation of HERV-K transcription in the BRAF$^{V600E}$ mutant, *invasive* type of melanoma. PLX4032, a commonly used BRAF$^{V600E}$ inhibitor in the clinic, enhances the expression of MITF by several folds that reverts melanoma from an *invasive* to the *proliferative* state [59]. To examine how PLX4032 treatment affect HERV-K expression, we reanalyzed the publicly available MITF ChIP-seq datasets of primary melanocytes and of a BRAF$^{V600E}$ mutant melanoma cell line (COLO829) with and without PLX4032 treatment (GSE50681). We observed that PLX4032 treatment results in a substantial enrichment of MITF ChIP-seq peaks over LTR5_Hs elements (123 unique loci were bound) (**Figure 2**D). When compared to other TE families, the gain of MITF binding occurred specifically over LTR5_Hs (**Figure 2**E). Importantly, while MITF is expressed in healthy pigmented cells, we found no elevated MITF ChIP-seq signal over LTR5_Hs elements upon PLX4032 treatment in normal, pigmented melanocytes (**Figure 2**D). Thus, MITF binding to LTR5_Hs is specific to the *proliferative* state of melanoma.

Collectively, the MITF binding may lead to LTR5_Hs activation, not typical in healthy melanocytes, but in the *proliferative* type of melanoma, thus can be considered as a sensitive marker of the *proliferative* vs *invasive* phenotype of melanoma.

### *3.3. Depleting HERV-K-Rec induces an EMT-like process in A375 melanoma cells*

The ERVK6 locus (located on the Chr7 p22.1) produces both retroviral ORFs (Gag and Env) and the accessory HERV-K product, Rec [13, 28]. To see how this specific locus responds to MITF in *proliferative* (MITF-High), *invasive* (MITF-Low) primary melanoma lines, we analyzed a published dataset for H3K27ac and MITF ChiP-seq (GSE60663), and also included the proliferative type A375 melanoma cell line in the analysis (GSE82330). This locus showed a higher H3K27Ac signal in both *proliferative* MITF-High and A375 cells relative to MITF-Low cells, and also a significant ATAC-seq signal in A375 (**Figure 3**A), indicating that this locus is preferentially active in *proliferative* type of melanoma. Moreover, ATAC-seq and active histone mark analysis (e.g. H3K27ac and H3K4Me1) supports the accessibility of multiple LTR5_Hs in A375 (**Figure 3**B-C), suggesting that A375 cell line might be an adequate model for this type of melanoma.

To decipher the underlying difference of the invasiveness of various melanoma cells and the potential role of Rec in melanoma, we first used an RNAi approach to deplete Rec expression in two melanoma cell lines (e.g. A375 and SKMel-28) of different invasiveness. A375 and SKMEL-28, both wild type for the tumour suppressor PTEN and carry the BRAF$^{V600E}$ mutation, but A375 has a higher proliferation rate, whereas SKMEL-28 is more invasive [62]. To knock down (KD) Rec, we designed three RNAi constructs. These constructs were delivered using *Sleeping Beauty* as a stable vector [63], and also expressed a fluorescence reporter (GFP). To generate stable KD cell lines for both A375 and SKMel-28, the KD constructs were transfected together with the SB100X transposase [64]. The stable KD-Rec lines were selected in two-rounds by FACS sorting for the GFP-reporter signal. Real time qRT-PCR showed that Rec transcript levels were strongly depleted in each of the three independent KD cell lines isolated for both A375 and SKMel-28 (**Figure 3**D). Notably, as a side effect, knocking down Rec also led to an increased level of alternative retroviral HERV-K transcripts, Gag and Env in SKMel-28 and in A375, respectively (**Figure 3**E).

To detect global transcriptional changes in an unbiased way, we used RNA-seq analysis of the stable KD-Rec cell lines. The analysis of individual HERV-K genomic loci revealed that our KD strategy (using three biological replicates) significantly affected seven Rec expressing loci in KD-Rec_A375 cells (**Figure 3**F). SKMel-28, by contrast, has only three HERV-K loci expressing Rec (not shown) [28].

Having validated our Rec KD lines, we performed a global analysis of differentially expressed genes (DEGs) (|2|-fold and FDR < 0.05, (**Figure S1**E**, Table S1,** see **Methods**) between the KD lines and the parental cell lines. The list of the DEGs was different between the A375 and SKMel-28 knock-downs, which could be in part explained by the distinct patterns of active HERV-K loci in the two melanoma lines. In KD-Rec_SKMel-28, DEGs (**Table S1**) did not include MITF and were not significantly enriched for any Gene Ontology (GO) categories, and formed no obvious pattern, suggesting that knocking-down Rec in this particular (*invasive*) melanoma line was not instructive of biological process alterations. The smaller number of Rec expressing HERV-K loci, and the antagonistic elevation of alternative retroviral product (e.g. Rec) in KD-Rec_SKMel-28 might, at least partially, explain the lack of obvious pattern.

In contrast, GO analysis of 1120 DEGs in KD-Rec_A375 revealed that most of the dysregulated genes were associated with *cell communication*, followed by *cell cycle, cell proliferation* (**Figure 4**A). Narrowing down the list of DEGs by excluding genes of the category of *system development*, the remaining genes were enriched in clusters of *cell differentiation*, *stem cell differentiation, cell death* and *system development*. Genes in the *stem cell differentiation* and *system development* categories could be linked to the *WNT signalling pathway*. Among DEGs in KD-Rec_A375, we identified genes involved in an epithelial-mesenchymal transition (EMT)-like process, has been observed to play a critical role of phenotype switching in the *proliferative* to *invasive* state [33, 66, 67]. Intriguingly, this set of DEGs included matrix metallopeptidase 2 (MMP2), a classical marker of cell invasion and additional 4 genes from the list of the 16 key markers of the canonical EMT process [65] (e.g. N-Cadherin (CDH2), zinc finger protein (SNAI2), forkhead box C2 (FOXC2) and goosecoid Homeobox (GSC) (**Figure 4**B-C). Nevertheless, as the melanocytes and melanoma cells differ from epithelial and mesenchymal cells, respectively, the EMT-like process in melanoma phenotype switching is characteristically different from the canonical EMT process [33]. In melanoma metastasis, the TF network switches to the melanoma-specific SNAI2$^{low}$/ZEB2$^{low}$/ZEB1$^{high}$ phenotype [69-71], that we observed also in our KD-Rec_A375 cells (**Figure 4**B-C), suggesting that Rec modulates the melanoma specific EMT-like phenotype switching. The elevated level of EMT markers in KD-Rec_A375 cells suggest that Rec inhibits the EMT-like process.

Beside canonical EMT markers, we also observed upregulation of further factors in KD-Rec_A375 cells, known to act as inducers of the EMT-like process. For example, non-canonical WNT signalling via WNT5A and GSK3B, which is markedly upregulated in KD-Rec_A375 cells (**Figure 4**D), has been shown to be critical for the switch to an EMT-like invasive phenotype when melanoma gets relapsed [72, 73]. Furthermore, the most extensively studied inducer of EMT, TGF-β [74, 75], reported to control SNAI2 [76], was also upregulated was also upregulated in our KD-Rec_A375 transcriptomes (**Table S1**).

In addition, among the top DEGs, we observe a strong upregulation of several members of the SPANXC (sperm protein associated with the nucleus in the X chromosome) family of genes (**Figure 4**D-E), which are normally expressed in the germline. The upregulation of SPANXB1/2, clinically relevant cancer-testis antigens, has been previously reported in melanoma [75] and was also validated by qRT-PCR in our KD-Rec_A375 samples (**Figure 4**E). Simultaneous upregulation of multiple genes physically located within the SPANX (B/C) chromosomal cluster suggests an epigenetic disturbance affecting the entire locus. Indeed, the epigenetically sensitive Polycomb repressor complex 2 (PRC2) has been implicated in maintaining SPANXB in a transcriptionally silent state in differentiated cells [76]. Our data suggest that Rec may exert directly or indirectly an effect on the epigenome of melanoma, an aspect that we did not follow up in this study.

Importantly, *in our* KD-Rec_A375 cells, we detected robust transcriptional decrease in the level of MITF-M, the governor of the melanocyte differentiation program (**Figure 5**A), confirmed by qPCR. In addition to MITF, we observed a decreased expression of further melanocyte differentiation markers (e.g. TYR, PMEL) [77] (**Figure 5**B).

Collectively, our transcriptome analysis in KD-Rec_A375 cells suggest that Rec might function as a cellular factor inhibiting the melanoma phenotype switching to the *invasive* state.

### *3.4. Rec is a suppressor of cell invasion in melanoma*

Our SNAI2$^{low}$/ZEB2$^{low}$/ZEB1$^{high}$ phenotype, attributed to aggressive melanoma, suggested that knocking down Rec, via induction of an EMT-like process might affect the invasiveness of A375 cells. Further on this line, the decreased level of MITF, and the transcriptome-wide DEGs of *invasive* vs *proliferative* melanoma markers also predicted an enhanced invasiveness of KD-Rec_A375 cells. To test these predictions, we performed a trans-well invasion assay to compare the invasiveness of KD-Rec_A375 cells to their parental line. The results showed that all three KD-Rec_A375 lines exhibited significantly elevated invasiveness (**Figure 5**C-D). The enhanced metastatic potential of the melanoma upon Rec KD supported the idea that Rec functions as a suppressor of the EMT-like transition process, which modulates the invasiveness of melanoma.

### *3.5. The KD-Rec_A375 in vitro model mimics the MITF type malignancy in patients*

Next, we sought to examine whether some of the genes apparently dysregulated upon Rec KD in A375 cells were also differentially regulated in malignant cell types. To assess this, and to monitor the dynamics of Rec-regulated genes in the progression of tumours, we performed a comprehensive single cell (sc)RNA-seq analysis (n=4645) derived from melanoma patients (n=19) [78]. Congruent with the original study [78], our analysis with default parameters from the 'Seurat' algorithm distinguishes clusters of non-malignant and malignant cell types (**Figure 6**A). The five distinct malignant cell types were marked by high expression levels of MAGEC2, MITF, APOE, VIM, and SPP1, respectively (**Figure 6**A-B).

To see how faithfully our KD-Rec_A375 model mimics *in vivo* processes observed in melanoma patients, we first compared the ~1500 DEGs identified between non-malignant and malignant cell types (all five), to those identified in our KD-Rec_A375 experiments (**Figure 7**A). We found that ~200 genes overlapped between the two analyses in opposite direction (**Figure 7**B), a significant enrichment over the random expectation (p-value <0.00001, see Methods).

Importantly, the expression levels of MITF, APOE and SPP1 (but not MAGEC2, and only modestly VIM) were also significantly depleted upon Rec knockdown in A375 cells, suggesting that the KD-Rec_A375 might recapitulate certain aspects of the melanoma process in patients (**Figure 7**C-D). The affected MITF, APOE (Apolipoprotein E) and SPP1 (osteopontin, cytokine) highlight different aspects of melanoma progression, a crosstalk of differentiation, angiogenesis [79] and inflammation [80], respectively.

As MITF directly regulates HERV-K (via binding LTR5_Hs), we next focused on the MITF malignancy sub-cluster of the scRNA-seq dataset [78] (**Figure 6**B). We took the top 100 MITF-correlated genes, and determined their expression in our Rec-KD_A375 data. This approach revealed a highly significant (Wilcoxon-test, p-value < 2e16) reverse differential expression pattern between MITF-type malignancy and Rec-KD_A375 (**Figure 7**E), indicating an anti-correlation between the expression of MITF-target gene expression and Rec depletion. Thus, our KD-Rec_A375 model appears to recapitulate MITF-dependent aspects of melanoma in patients.

## 4. Discussion

HERV-K resembles "complex" retroviruses such as HIV-1 by its ability to encode multiple accessory proteins, including Rec, through alternative splicing of its Env transcript. While Env has the typical structure of a transmembrane protein (e.g. extracellular, transmembrane, cytoplasmic domains), Rec lacks the majority of the domains required for a membrane protein function (**Figure 8**A), but carries nuclear localization/export signals [3, 7]. Although Env and Rec are completely different in sequence and function, our results suggest that Rec, like Env [39], exerts a modulatory effect on the EMT-like process in cancer progression. However, while Env affects signalling via ERK1/2 activation, through the cytoplasmic tail of the transmembrane Env protein [39], this domain is missing from Rec (**Figure 8**A). Rec, the splice variant of Env, inhibits the EMT-like process by affecting the expression levels of key determinants of the EMT-like process, and certain upstream regulators. Thus, the two factors appear to act in the opposite direction. While Env overexpression

(both RNA and protein) may promote tumour progression [39], depletion of Rec appears to result in a comparable phenotype. Curiously, similar to Env, the retroviral gene product, Gag also increases the tumorigenic potential of melanoma cells [38], raising the possibility that retroviral and accessory genes might generally act inversely, deciding on a long-lasting debate on the potential oncogenicity of a retroviral versus accessory genes of HERV-K.

In about 50% of melanomas, the protein kinase BRAF is mutated ($BRAF^{V600E}$) [84]. MITF is considered as a crucial regulator of melanoma invasiveness [85], and plays a critical role in controlling $BRAF^{V600E}$ melanoma [86, 87]. Accordingly, expression of MITF has been widely used as a marker to distinguish between *proliferative* and *invasive* type of melanoma cells [61, 85, 86, 88]. Reduced *MITF* expression generates *invasive* melanomas, with tumour-initiating properties [86]. Importantly, our KD-Rec_A375 *in vitro* system appears to recapitulate a subpopulation of melanoma cells (MITF malignancy) in human patients. Our data suggest that MITF activates LTR5_Hs only in melanoma (not in melanocytes), but Rec expression specifically marks the *proliferative* type of melanoma cells. Importantly, the *proliferative* phase of melanoma is still a better controllable stage of melanoma progression, by contrast to the *invasive* stage that gets frequently relapsed. If Rec expression is involved in maintaining a delicate balance between cell proliferation and invasion, as this study suggests, then Rec activation might be used as a sensitive marker to distinguish between the *proliferative* vs *invasive* phenotypes of melanoma, not expressed in normal melanocytes.

To control cell plasticity during melanomagenesis, the EMT-like network, operates upstream of MITF. In cells harbouring the $SNAI2^{low}/ZEB2^{low}/ZEB1^{high}$ phenotype, *the depleted* SNAI2 and ZEB2, expressed in normal melanocytes and function as tumor-suppressors, would fail to activate the MITF-dependent melanocyte differentiation program [34, 87]. In parallel, the upregulated ZEB1 induces an invasion-associated cascade that operates in concert with downregulating *MITF.*

Our study reveals that depletion of Rec in a $BRAF^{V600E}$ mutant melanoma cells (A375) results in a decline in MITF level. Thus, our KD-Rec_A375 cells suggest a model whereby a MITF-dependent regulatory loop is disrupted through (i) decreased MITF binding to LTR5_Hs, which leads to (ii) reduced Rec levels and (iii) a further decline in MITF level, contributing to (iv) escalated tumour aggressiveness (**Figure 8**B). The observed cross-regulation between Rec and MITF reveals an inhibitory effect of Rec on the invasive transformation of melanoma. Our data suggest that high level of Rec expression marks the *proliferative* (still controllable) state of melanoma cells, and opposes the molecular switch to the *invasive* state. Future research is required to decipher the exact mechanism of Rec control over MITF, and to test whether therapeutically induced Rec transcription, similar to enhanced MITF expression (e.g. PLX4032 treatment), results in a reversion of the *invasive* melanoma phenotype.

Under healthy physiological conditions, only a limited number (n=18, or often less) of HERV-K loci appear to be active [13], whereas in melanoma patients/cell lines, the number of transcribed HERV-K loci is higher (n~24) [28], indicating that certain HERV-K loci are transcribed only under pathologic conditions. Several loci express multiple HERV-K-derived products (both core and accessory replication factors). Although forced expression of MITF-M in non-malignant cells has been shown to enhance the levels of multiple HERV-K derived transcripts (e.g. Gag, *Env* and *Rec) [52], the activated* HERV-K loci might *differ in both intensity of expression and the composition of their* HERV-K-derived transcripts. In line with this, HERV-K loci are variably activated and in various combinations in melanomas [28, 89], suggesting that variability in the amounts of expressed HERV-K products might contribute to melanoma heterogeneity [22, 27], and possibly the progression and outcome of the disease.

It has been widely discussed how precisely melanoma can be recapitulated in animal models. Interestingly, our analysis of single cell transcriptome data from multiple human patients suggest a distinct separation of *proliferative* and *invasive* type of cells within a patient, whereas these two phases occur simultaneously within the same melanoma cell in mice [88] As Rec is hominid-specific, our study has also an evolutionary prospective that might explain certain species-specific features of melanoma progression in human patients, and help us to understand the limitation(s) of using animal models in melanoma studies.

## Figure legends

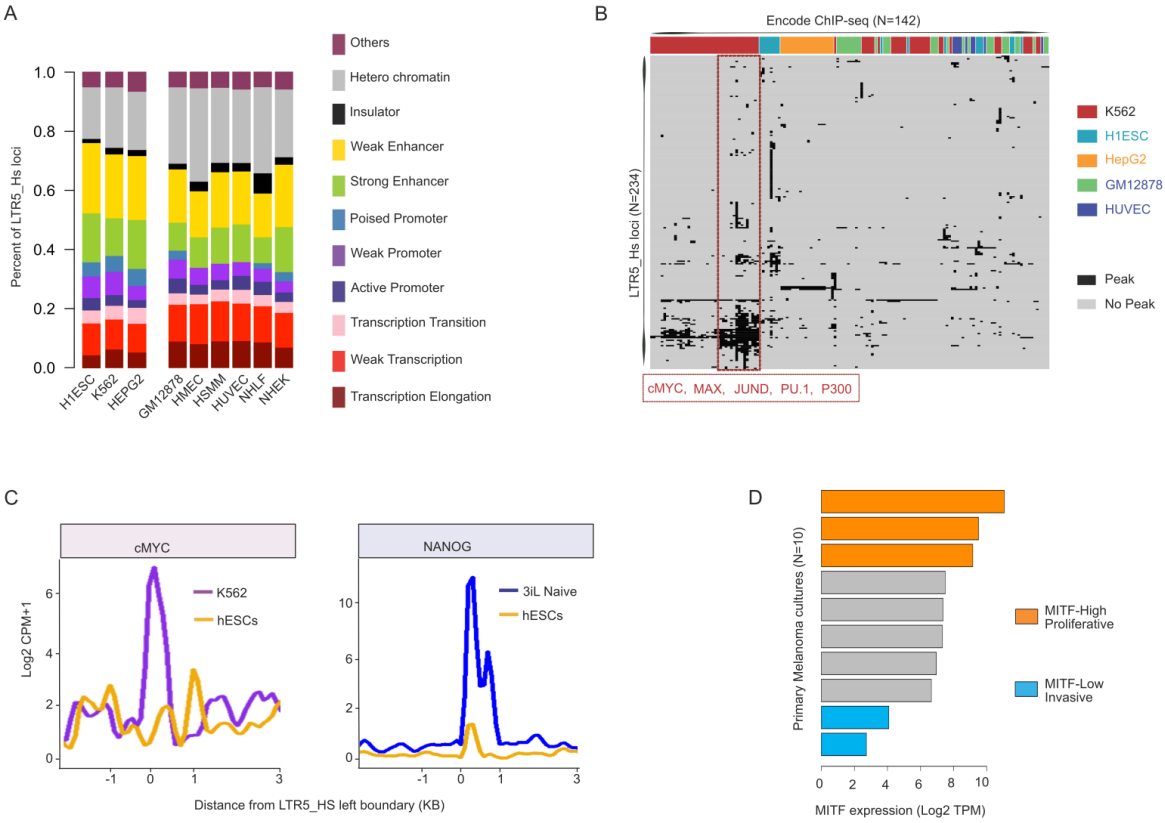


**Figure 1.** (A) Stacked barplots show the chromatin state (chromHMM) of the LTR5_Hs genomic loci (human Refseq genome) across 9 cell types. (B) Heatmap summarizing the transcription factor (TF) occupancy (ChIP-seq peaks, ENCODE) over individual LTR5_Hs loci in five human cell lines. This plot includes 234 out of 615 LTR5_Hs loci annotated in the human Refseq genome, which is occupied by at least one ChIP-seq peak. Each row represents an individual locus of LTR5_Hs. Each column shows TF ChIP-seq peak occupancy of a different cell line. Grey denotes the absence, whereas black is the presence of peaks in a given loci. Various TFs encompassing LTR5_Hs loci is manually annotated in a given cell line. The cluster of TFs binding over a subset of LTR5_Hs loci is shown in the red box. C. Line plots showing the distribution of normalized ChIP-seq tag counts of cMYC and NANOG in K562 and in naive ESCs, respectively. The ChIP-seq tag counts are calculated relative to primed H1_ESCs, in a 3kb genomic window at the left boundaries of LTR5_Hs. (D) Characterizing primary melanoma cultures according to their MITF expression levels. Barplot showing the level of MITF expression (Log2 TPM) in primary melanoma cultures (n=10). According to their MITF expression level, cultures are defined as *proliferative melanoma* (MITF-High) and *invasive melanoma* (MITF-Low) *in vitro* models.

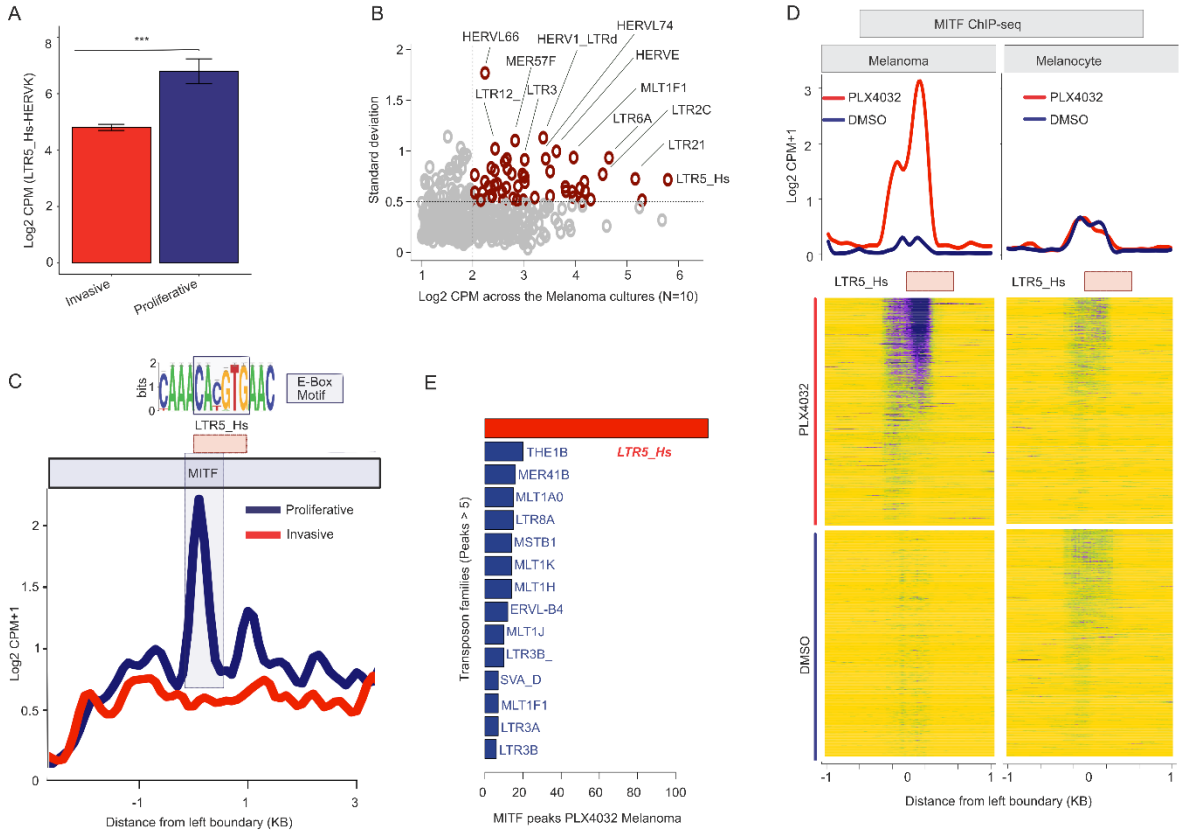


**Figure 2.** (A) Barplot showing the LTR5_Hs/HERV-K-int expression (Log2 CPM) levels in *proliferative melanoma* (MITF-High) and *invasive melanoma* (MITF-Low) cell lines (GSE60664). (B) Scatter plot visualizes the average expression (Log2 CPM) (X-axis) and standard deviation from mean (Y-axis) of transposable element (TE) families in various primary melanoma cultures (n=10). (C) Line plot showing the MITF ChIP-seq signal occupancy in *proliferative melanoma* (MITF-High) and *invasive melanoma* (MITF-Low) cell lines averaged over all of the LTR5_Hs loci. The identified E-Box motif in MITF bound LTR5_Hs in A375 cells is shown. Note that the E-box motif was found using the sequences of LTR5_Hs loci with MITF peaks. (D) The effect of PLX4032 treatment on MITF binding at LTR5_Hs loci. Upper panel: Line plot showing the MITF ChIP-seq signal averaged over LTR5_Hs loci in primary melanocytes and melanoma cells with and without the treatment of PLX4032. Lower panel: Heatmap of MITF ChIP-seq signals. Note: MITF occupies LTR5_Hs genomic loci upon PLX4032 treatment specifically in melanoma, but not in melanocytes (over 100 loci). (E) Barplot showing the number of MITF peaks detected over the genomic loci of various TE families in melanoma cells treated with PLX4032. This plot shows only those families which have at least 5 peaks over the individual locus.

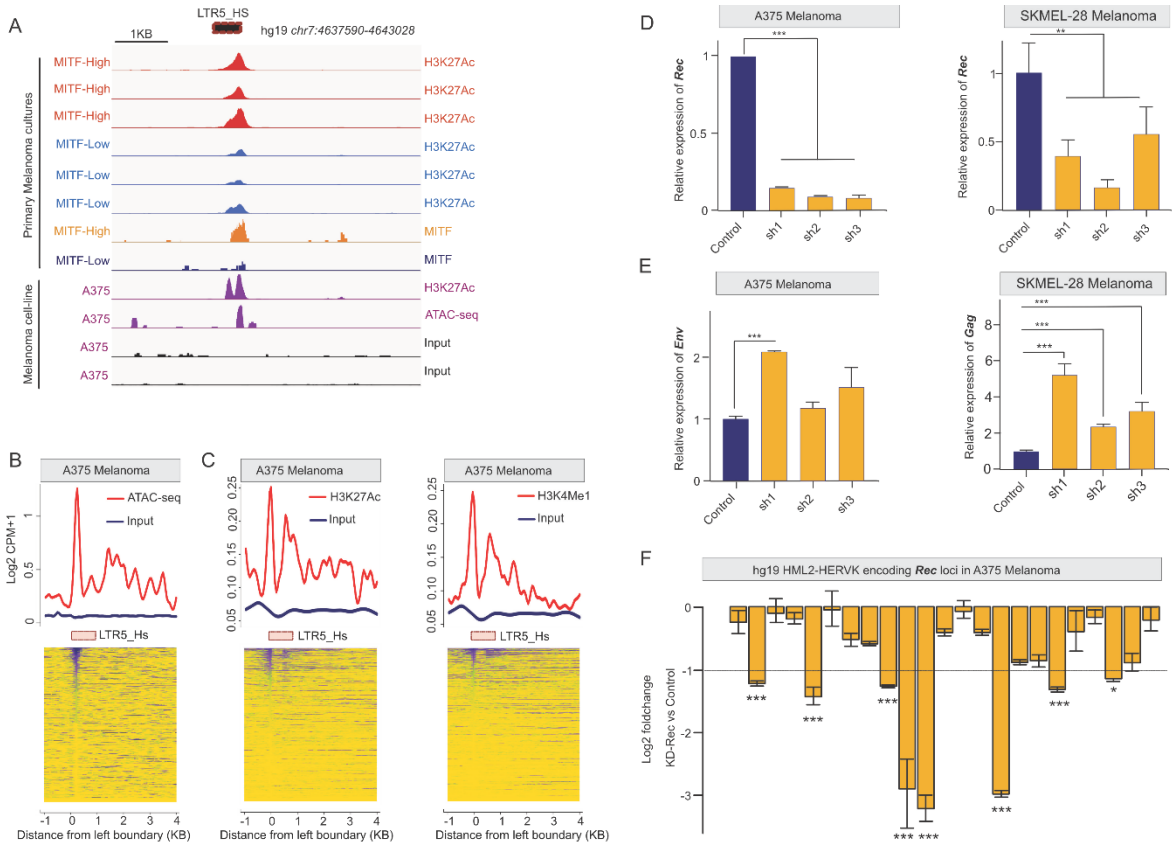


**Figure 3:** (A) Genome browser tracks showing the H3K27Ac, ATAC-seq and MITF signals at a specific LTR5_Hs-HERVK locus (ERVK6, known to produce Rec) in MITF-High primary melanoma cultures and A375 cell line (*proliferative*) and in MITF-low (*invasive*) melanoma primary cultures. (B) Normalized ATAC-seq (nucleosome-associated) signal averaged at LTR5_Hs genomic loci in A375 cells. (C) Normalized ChIP-seq signal of H3K4Me1 and H3K27Ac histone marks (corresponding to active chromatin states) averaged over LTR5_Hs genomic loci in A375 cells.(D) Knocking down (KD) Rec in A375 and SKMEL-28 cells using RNA interference. Stable KD lines were selected in two-rounds by FACS sorting for the GFP-reporter signal. Barplots showing the normalized RT-qPCR quantification of Rec mRNA levels with three KD-Rec constructs. (E) Knocking down Rec affects the transcription of alternative HERV-K products. Barplot is showing the effect of knocking down Rec (KD-Rec) on the expression of alternative HERVK products, determined by real time qPCR. Note: KD-Rec led to an increased level of Gag in SKMel-28 (all the three biological replicates), whereas Env in A375 (KD1 only). (F) Knocking down Rec in A375 affects at least seven genomic HERV-K loci. Combined barplots showing the specificity of the RNAi strategy against the Rec subunit. Relative quantification of HERV-K Rec genomic loci compared with negative and scrambled controls indicates that all three KD constructs significantly affected at least 7 Rec expressing loci in A375 cells.

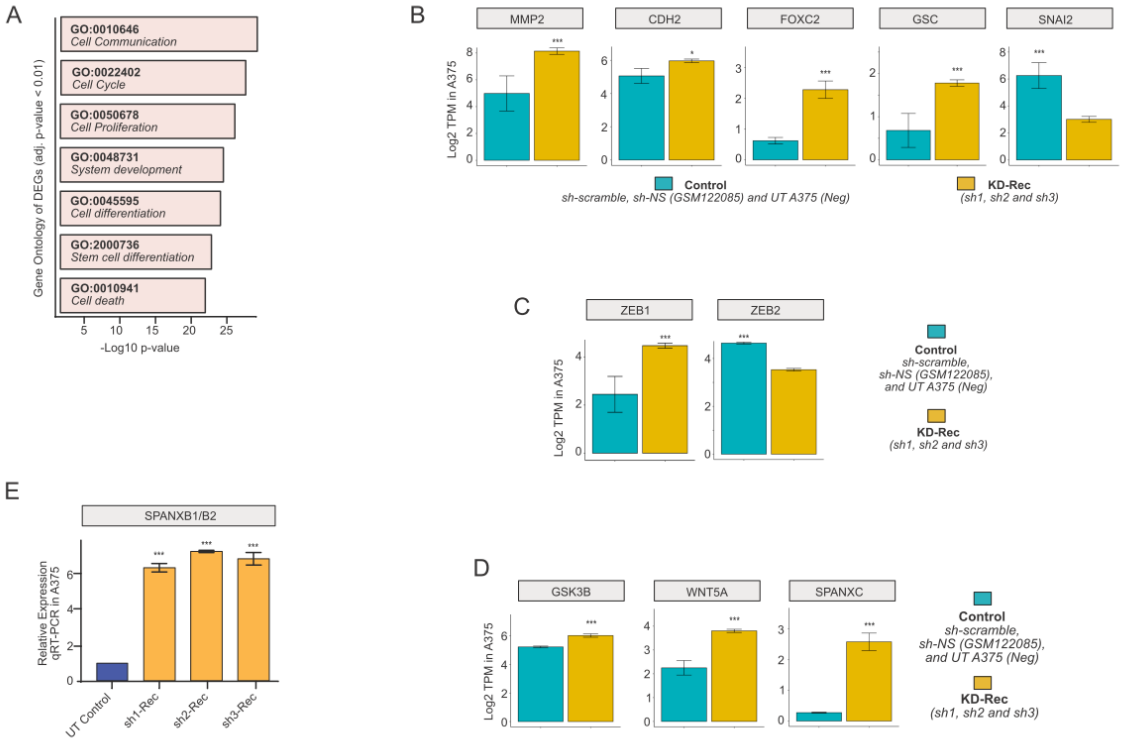


**Figure 4:** (A) Barplots of significant Gene Ontology (GO) Biological Process terms enrichment analysis of DEGs in KD-Rec_A375. Values are negative logarithms of the corrected p-values (Benjamini Hochberg FDR).(B) Barplots representing means ± SEM from RNA-seq expression values (Log2 TPM) for selected DEGs in KD-Rec_A375 (e.g. Matrix metallopeptidase 2 (MMP2), a classical marker of invasion; N-Cadherin (CDH2), Zinc finger protein SNAI2 (SNAI2), Forkhead box C2 (FOXC2) and Goosecoid Homeobox (GSC) are among the key gene regulators of the canonical EMT transition. (C) Barplots representing means ± SEM from RNA-seq expression values (Log2 TPM) for DEGs in KD-Rec_A375 vs control, involved in the EMT-like process in melanoma [34] (e.g. SNAI2, ZEB2 and ZEB1). (D) Barplots representing means ± SEM from RNA-seq expression values (Log2 TPM) for selected DEGs in KD-Rec_A375 vs control, known as inducers of the EMT-like process (e.g. WNT5A and GSK3B) and SPANXC (sperm protein associated with the nucleus in the X chromosome). (E) Barplots are showing the effect of knocking down Rec in A375 vs control on SPANXB1/B2 genes, detected as DEG in RNA-seq experiments, and validated using real time qPCR.

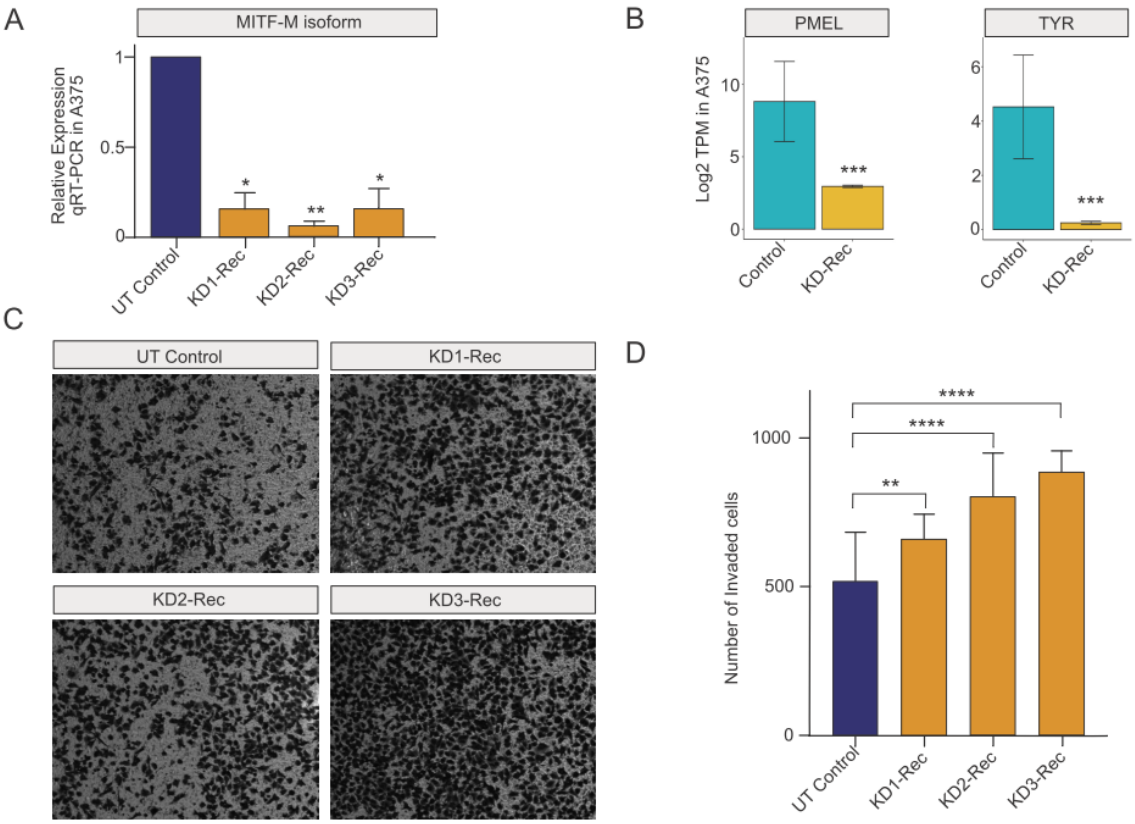


**Figure 5:** (A) Quantifications transcriptional variation of MITF-M by KD-Rec in A375 cells using RNA-seq and by real time qPCR. (B) Barplot representing means ± SEM from RNA-seq expression values (Log2 TPM) for melanocyte markers (e.g. TYR and PMEL) upon Rec-KD in A375. (C) Representative inverted light microscope images of trans-well invasion assay performed in A375 melanoma cells upon knocking-down Rec by three KD constructs (KD1-3). UT, untransfected control. Scale bars, 400 µm. (D) Quantification of cells invading in the trans-well invasion assay (C). Number of invaded cells was counted by imageJ. Columns indicate average number of invaded cells per field from two independent experiments. Error bars represent the means±s.d. of two independent experiments. Mann Whitney test; **$P \leq 0.01$, ****$p \leq 0.0001$ (See methods for details).

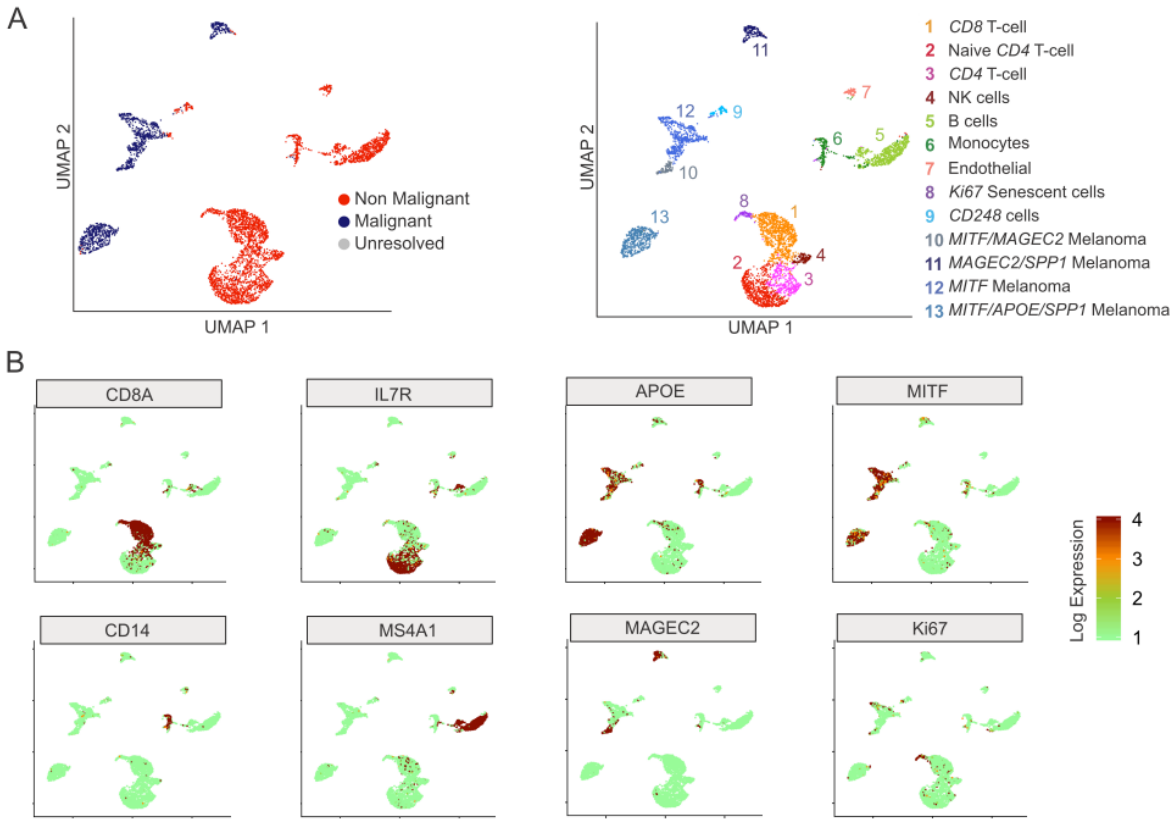


**Figure 6:** (A) Various plots with maximum thirteen distinct clusters are visualized on UMAP coordinates using 10XGenomics scRNA-seq [78] and their original annotations. Left panel shows the Malignant and Non-Malignant cell populations, whereas right panel is displaying all the clusters obtained that were classified using their original annotations. (B) Feature plots based on the UMAP plot shown on (A) visualizing the expression of selected markers used to identify the distinct cell-types in melanoma population. Color intensity gradient indicates the expression of the depicted marker gene. Each dot represents an individual cell. Dots in green denote lower, whereas in dark red, higher level of gene expression in a given single cell. This resulted in five distinct malignant cell types, marked by MAGEC2, MITF, APOE and Ki67 respectively (See Fig. 7A for SPP1 expression).

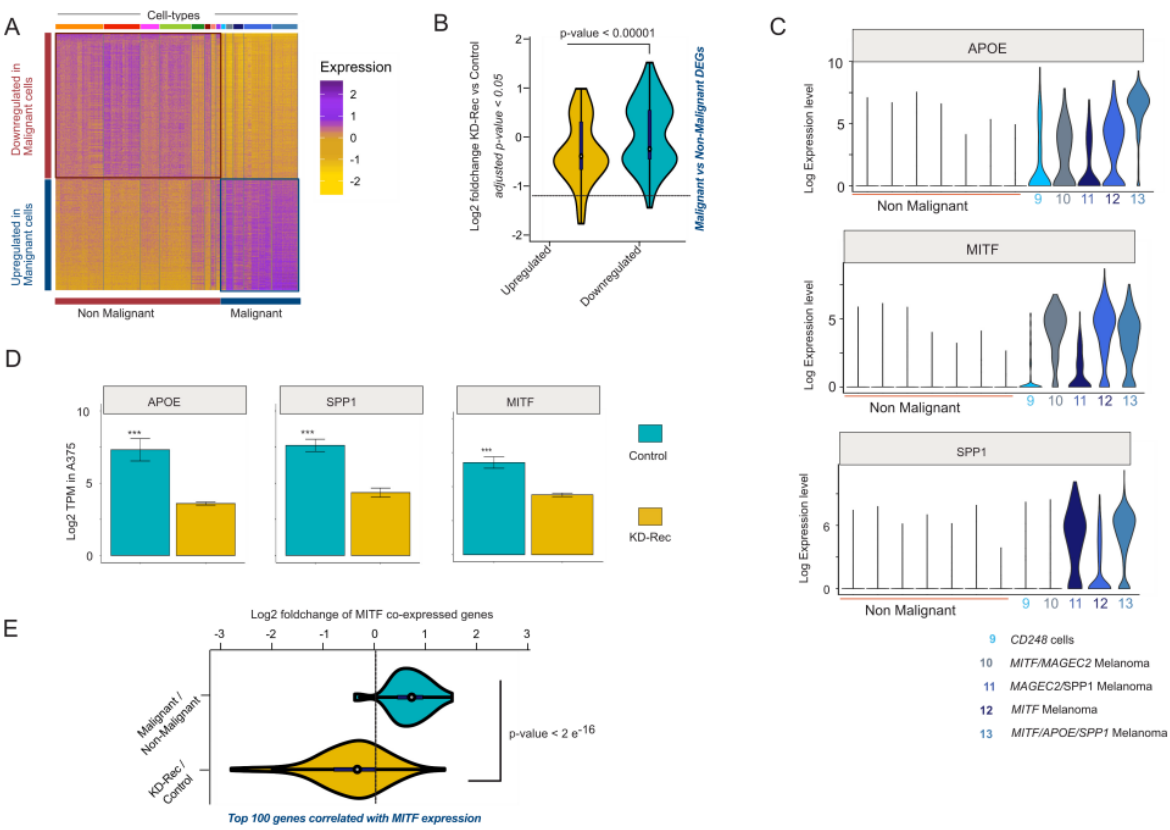


**Figure 7: (A)** Heatmap visualization of scaled expression [log TPM (transcripts per million)] values of distinctive set of ~ 1500 genes (log2 fold change > 1 and > 50% of cells expressing either in healthy or malignant cells), which are differentially expressed in scRNA-seq data. Colour scheme is based on Z-score distribution from -2.5 (gold) to 2.5 (purple). Left margin colour bars highlight representative gene sets specific to the respective clusters. (B) Violin plot visualizes the density and distribution of expression (Log2 TPM values) of ~ 200 DEGs upon Rec-KD in A375 cells which are common in the set of genes shown in figure above (C). (p value calculated by Wilcoxon test). (C) Violin plots visualize the density and distribution of expression of distinct marker genes, for melanoma malignant clusters. (D) Barplots with +/- SEM shows the down-regulation of MITF, SPP1 and APOE upon knock down of Rec in A375 cells. (E) Violin plot showing the comparison of previously identified gene set as "MITF-high", consisting of *MITF* itself and the top 100 genes correlated with MITF expression between malignant and non-malignant cells (blue) and upon Rec-KD in A375 by individual shRNA constructs (gold). (p value calculated by Wilcoxon test).

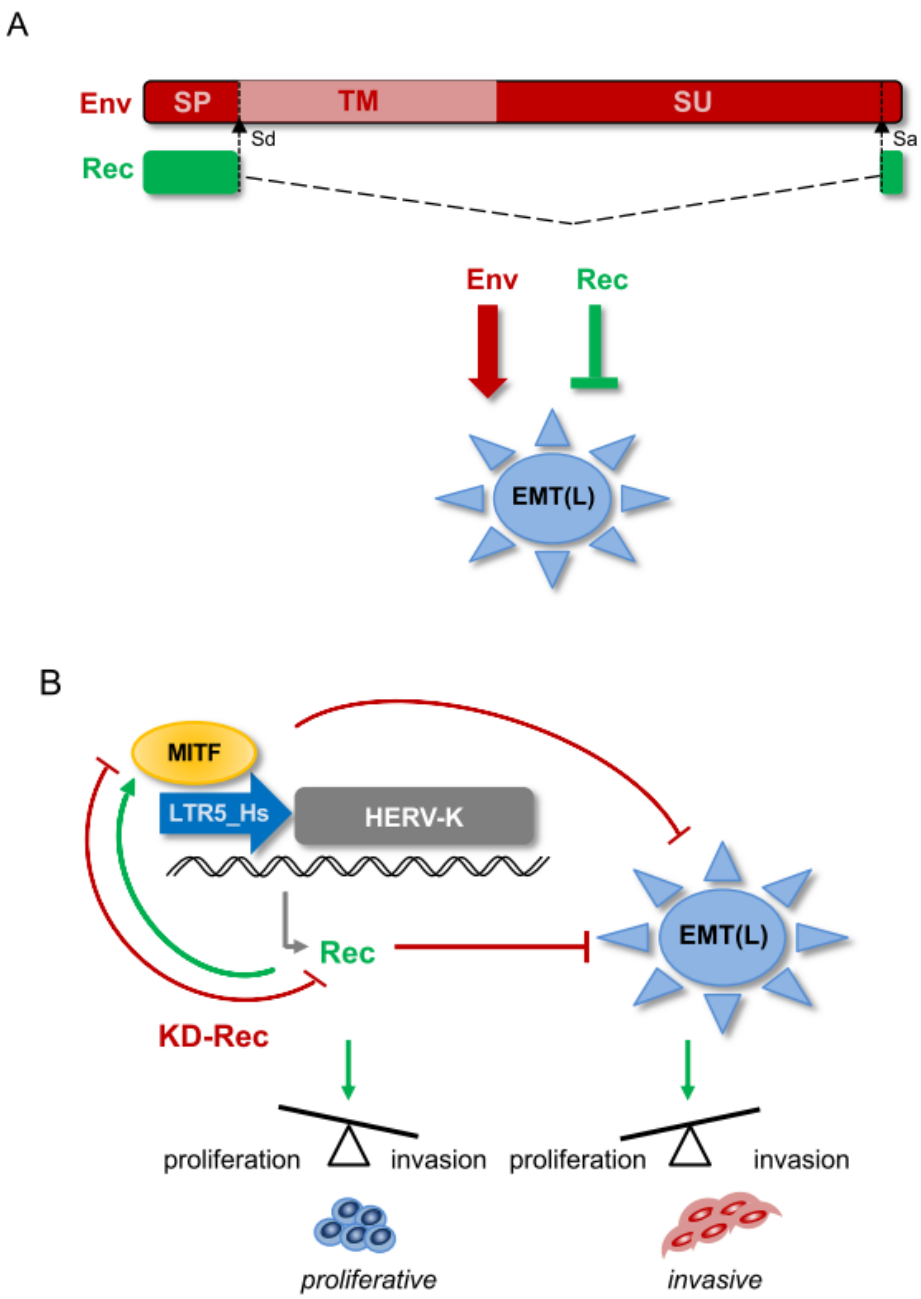


**Figure 8:** (A) Schematic structure of the HERV-K-derived alternative splice products, Env and Rec, and their effect on the endothelial-mesenchymal transition (EMT)-like process of cancer progression (EMT(L). Signal peptide (SP); transmembrane subunit (TM); Splice acceptor (Sa); Splice donor (Sd). (B) Modeling the interaction loop between MITF and LTR5_Hs and Rec in melanoma. MITF binds to LTR5_Hs and drives Rec expression. Rec marks the proliferative stage melanoma cells and, similarly to MITF, inhibits the endothelial-mesenchymal transition (EMT)-like process (EMT(L). Depletion of Rec results in MITF downregulation and the failure to maintain the balance between cell proliferation and invasion.

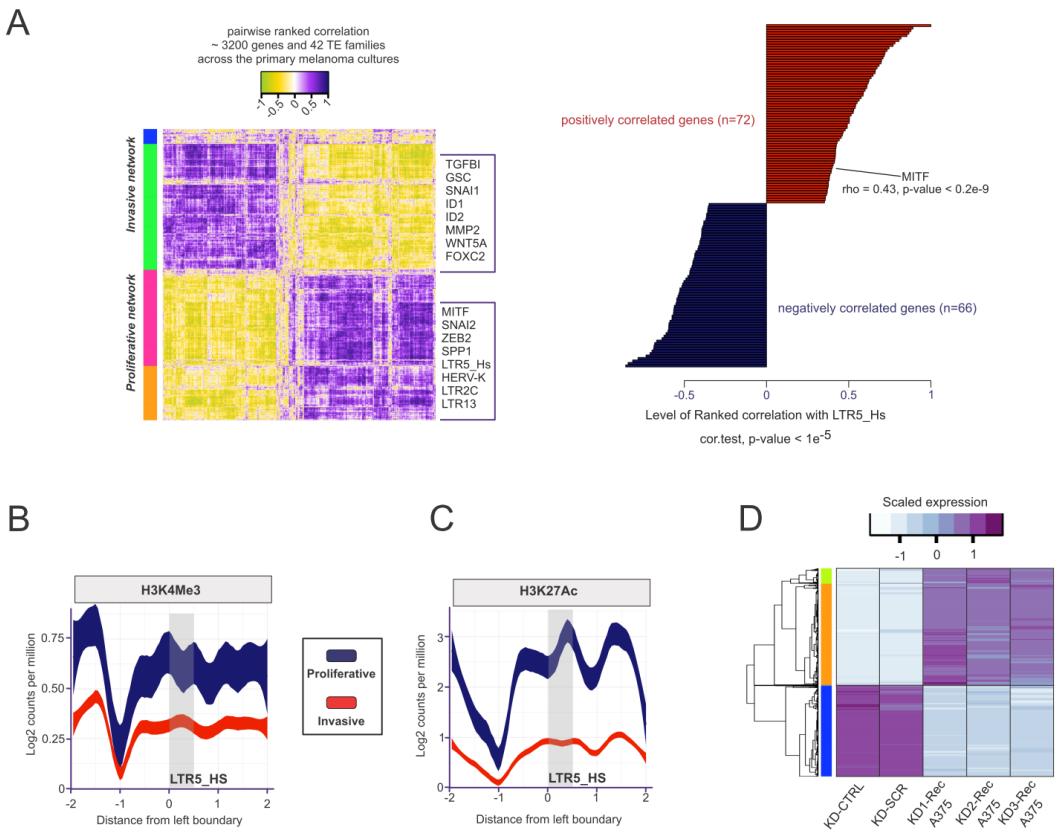


**Figure S1: (A)** (Left panel) Ranked correlation matrix of genes and 42 transposable element (TE) families, which are differentially expressed between *proliferative* and *invasive* primary melanoma cultures. Genes and TE families are manually annotated. Color codes on the left side represent the unsupervised clusters of most variable genes and TEs across the samples. On the right panel, selected genes are listed that are positively or negatively correlated with LTR5_Hs expression across the melanoma primary cultures. (B) Distribution of ChIP-seq signal of the H3K4Me3 histone mark occupancy, corresponding to active promoters averaged over LTR5_Hs loci in MITF-High (*proliferative*) vs MITF-Low (*invasive*) state of primary melanoma cultures. (C) Distribution of ChIP-seq signal of the H3K27Ac histone mark occupancy (lower panel) averaged over LTR5_Hs loci in MITF-High (*proliferative*) vs MITF-Low (*invasive*) state of primary melanoma cultures. (D) Heatmap displays the DEGs upon KD-Rec in A375 cells.

**Author Contributions:** Conceptualization, M.S, H.C. and Z.I.; methodology, H.C.; analyses, M.S.; writing, original draft preparation, Z.I.; writing—review and editing, all authors; visualization, A.K.; supervision, Z.I. All authors have read and agreed to the published version of the manuscript.

**Funding:**

**Acknowledgments:** M.S. H.C. was supported by from the China Scholarship Council (CSC). M.S. was supported by a presidential postdoctoral fellowship from Cornell University. Z.I. was funded by European Research Council, ERC Advanced [ERC-2011-ADG 294742].

**Conflicts of Interest:**

## References

1. Lander ES, Linton LM, Birren B, Nusbaum C, et al., 2001. Initial sequencing and analysis of the human genome. *Nature,* 409: 860-921.
2. Weiss RA, 2006. The discovery of endogenous retroviruses. *Retrovirology,* 3: 67.
3. Lower R, Tonjes RR, Korbmacher C, Kurth R, et al., 1995. Identification of a Rev-related protein by analysis of spliced transcripts of the human endogenous retroviruses HTDV/HERV-K. *J Virol,* 69: 141-9.

4. Armbruester V, Sauter M, Krautkraemer E, Meese E, et al., 2002. A novel gene from the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res,* 8: 1800-7.
5. Magin C, Lower R, and Lower J, 1999. cORF and RcRE, the Rev/Rex and RRE/RxRE homologues of the human endogenous retrovirus family HTDV/HERV-K. *J Virol,* 73: 9496-507.
6. Boese A, Sauter M, Galli U, Best B, et al., 2000. Human endogenous retrovirus protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc finger protein. *Oncogene,* 19: 4328-36.
7. Yang J, Bogerd HP, Peng S, Wiegand H, et al., 1999. An ancient family of human endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. *Proc Natl Acad Sci U S A,* 96: 13404-8.
8. Denne M, Sauter M, Armbruester V, Licht JD, et al., 2007. Physical and functional interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic leukemia zinc finger protein. *J Virol,* 81: 5607-16.
9. Kaufmann S, Sauter M, Schmitt M, Baumert B, et al., 2010. Human endogenous retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. *J Gen Virol,* 91: 1494-502.
10. Hanke K, Hohn O, Liedgens L, Fiddeke K, et al., 2013. Staufen-1 interacts with the human endogenous retrovirus family HERV-K(HML-2) rec and gag proteins and increases virion production. *J Virol,* 87: 11019-30.
11. Costoya JA, Hobbs RM, Barna M, Cattoretti G, et al., 2004. Essential role of Plzf in maintenance of spermatogonial stem cells. *Nat Genet,* 36: 653-9.
12. Grow EJ, Flynn RA, Chavez SL, Bayless NL, et al., 2015. Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature,* 522: 221-5.
13. Schmitt K, Heyne K, Roemer K, Meese E, et al., 2015. HERV-K(HML-2) rec and np9 transcripts not restricted to disease but present in many normal human tissues. *Mob DNA,* 6: 4.
14. Andersson AC, Venables PJ, Tonjes RR, Scherer J, et al., 2002. Developmental expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology,* 297: 220-5.
15. Mayer J, Ehlhardt S, Seifert M, Sauter M, et al., 2004. Human endogenous retrovirus HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity. *Virology,* 322: 190-8.
16. Izsvak Z, Wang J, Singh M, Mager DL, et al., 2016. Pluripotency and the endogenous retrovirus HERVH: Conflict or serendipity? *Bioessays,* 38: 109-17.
17. Kreimer U, Schulz WA, Koch A, Niegisch G, et al., 2013. HERV-K and LINE-1 DNA Methylation and Reexpression in Urothelial Carcinoma. *Front Oncol,* 3: 255.
18. Serafino A, Balestrieri E, Pierimarchi P, Matteucci C, et al., 2009. The activation of human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant transformation. *Exp Cell Res,* 315: 849-62.
19. Balestrieri E, Argaw-Denboba A, Gambacurta A, Cipriani C, et al., 2018. Human Endogenous Retrovirus K in the Crosstalk Between Cancer Cells Microenvironment and Plasticity: A New Perspective for Combination Therapy. *Front Microbiol,* 9: 1448.
20. Reiche J, Pauli G, and Ellerbrok H, 2010. Differential expression of human endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines after UV irradiation. *Melanoma Res,* 20: 435-40.
21. Schanab O, Humer J, Gleiss A, Mikula M, et al., 2011. Expression of human endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma. *Pigment Cell Melanoma Res,* 24: 656-65.
22. Muster T, Waltenberger A, Grassauer A, Hirschl S, et al., 2003. An endogenous retrovirus derived from human melanoma cells. *Cancer Res,* 63: 8735-41.
23. Stengel S, Fiebig U, Kurth R, and Denner J, 2010. Regulation of human endogenous retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer,* 49: 401-11.
24. Clausen J, 2003. Endogenous retroviruses and MS: using ERVs as disease markers. *Int MS J,* 10: 22-8.
25. Matteucci C, Balestrieri E, Argaw-Denboba A, and Sinibaldi-Vallebona P, 2018. Human endogenous retroviruses role in cancer cell stemness. *Semin Cancer Biol,* 53: 17-30.
26. Hahn S, Ugurel S, Hanschmann KM, Strobel H, et al., 2008. Serological response to human endogenous retrovirus K in melanoma patients correlates with survival probability. *AIDS Res Hum Retroviruses,* 24: 717-23.
27. Buscher K, Trefzer U, Hofmann M, Sterry W, et al., 2005. Expression of human endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res,* 65: 4172-80.

28. Schmitt K, Reichrath J, Roesch A, Meese E, et al., 2013. Transcriptional profiling of human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol,* 5: 307-28.
29. Contreras-Galindo R, Kaplan MH, Dube D, Gonzalez-Hernandez MJ, et al., 2015. Human Endogenous Retrovirus Type K (HERV-K) Particles Package and Transmit HERV-K-Related Sequences. *J Virol,* 89: 7187-201.
30. Verfaillie A, Imrichova H, Atak ZK, Dewaele M, et al., 2015. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun,* 6: 6683.
31. Bradish JR and Cheng L, 2014. Molecular pathology of malignant melanoma: changing the clinical practice paradigm toward a personalized approach. *Hum Pathol,* 45: 1315-26.
32. Bollag G, Hirth P, Tsai J, Zhang J, et al., 2010. Clinical efficacy of a RAF inhibitor needs broad target blockade in BRAF-mutant melanoma. *Nature,* 467: 596-9.
33. Li FZ, Dhillon AS, Anderson RL, McArthur G, et al., 2015. Phenotype switching in melanoma: implications for progression and therapy. *Front Oncol,* 5: 31.
34. Caramel J, Papadogeorgakis E, Hill L, Browne GJ, et al., 2013. A switch in the expression of embryonic EMT-inducers drives the development of malignant melanoma. *Cancer Cell,* 24: 466-80.
35. Babaian A and Mager DL, 2016. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA,* 7: 24.
36. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, et al., 2007. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature,* 448: 595-9.
37. Huang FW, Hodis E, Xu MJ, Kryukov GV, et al., 2013. Highly recurrent TERT promoter mutations in human melanoma. *Science,* 339: 957-9.
38. Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, et al., 2007. Distinct roles for LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression. *Oncogene,* 26: 4226-33.
39. Lemaitre C, Tsang J, Bireau C, Heidmann T, et al., 2017. A human endogenous retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and inducing migration and invasion. *PLoS Pathog,* 13: e1006451.
40. Li M, Radvanyi L, Yin B, Li J, et al., 2017. Downregulation of Human Endogenous Retrovirus Type K (HERV-K) Viral env RNA in Pancreatic Cancer Cells Decreases Cell Proliferation and Tumor Growth. *Clin Cancer Res,* 23: 5892-5911.
41. Zhou F, Li M, Wei Y, Lin K, et al., 2016. Activation of HERV-K Env protein is essential for tumorigenesis and metastasis of breast cancer cells. *Oncotarget,* 7: 84093-84117.
42. Ishida T, Obata Y, Ohara N, Matsushita H, et al., 2008. Identification of the HERV-K gag antigen in prostate cancer by SEREX using autologous patient serum and its immunogenicity. *Cancer Immun,* 8: 15.
43. Riker AI, Enkemann SA, Fodstad O, Liu S, et al., 2008. The gene expression profiles of primary and metastatic melanoma yields a transition point of tumor progression and metastasis. *BMC Med Genomics,* 1: 13.
44. Wallace TA, Downey RF, Seufert CJ, Schetter A, et al., 2014. Elevated HERV-K mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older men and smokers. *Carcinogenesis,* 35: 2074-83.
45. Dong J, Sedrak M, Rincon L, Sun H, et al., 2014. K-type human endogenous retroviral elements in human melanoma. *Advances in Genomics and Genetics*: 153.
46. Chiappinelli KB, Strissel PL, Desrichard A, Li H, et al., 2015. Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell,* 162: 974-86.
47. Nelson PN, Lever AM, Smith S, Pitman R, et al., 1999. Molecular investigations implicate human endogenous retroviruses as mediators of anti-retroviral antibodies in autoimmune rheumatic disease. *Immunol Invest,* 28: 277-89.
48. Hirschl S, Schanab O, Seppele H, Waltenberger A, et al., 2007. Sequence variability of retroviral particles derived from human melanoma cells melanoma-associated retrovirus. *Virus Res,* 123: 211-5.
49. Ruprecht K, Mayer J, Sauter M, Roemer K, et al., 2008. Endogenous retroviruses and cancer. *Cell Mol Life Sci,* 65: 3366-82.
50. Hohn O, Hanke K, and Bannert N, 2013. HERV-K(HML-2), the Best Preserved Family of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Front Oncol,* 3: 246.

51. Galli UM, Sauter M, Lecher B, Maurer S, et al., 2005. Human endogenous retrovirus rec interferes with germ cell development in mice and may cause carcinoma in situ, the predecessor lesion of germ cell tumors. *Oncogene,* 24: 3223-8.
52. Singh S, Kaye S, Francis N, Peston D, et al., 2013. Human endogenous retrovirus K (HERV-K) rec mRNA is expressed in primary melanoma but not in benign naevi or normal skin. *Pigment Cell Melanoma Res,* 26: 426-8.
53. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, et al., 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature,* 473: 43-9.
54. Sun Q, Yang J, Xing G, Sun Q, et al., 2008. Expression of GSDML Associates with Tumor Progression in Uterine Cervix Cancer. *Translational Oncology,* 1: 73-IN1.
55. Katoh I, Mirova A, Kurata S, Murakami Y, et al., 2011. Activation of the long terminal repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M. *Neoplasia,* 13: 1081-92.
56. Levy C, Khaled M, and Fisher DE, 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med,* 12: 406-14.
57. Steingrimsson E, Copeland NG, and Jenkins NA, 2004. Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet,* 38: 365-411.
58. Blackwood EM and Eisenman RN, 1991. Max: a helix-loop-helix zipper protein that forms a sequence-specific DNA-binding complex with Myc. *Science,* 251: 1211-7.
59. Webster DE, Barajas B, Bussat RT, Yan KJ, et al., 2014. Enhancer-targeted genome editing selectively blocks innate resistance to oncokinase inhibition. *Genome Res,* 24: 751-60.
60. Muller J, Krijgsman O, Tsoi J, Robert L, et al., 2014. Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nat Commun,* 5: 5712.
61. Hoek KS, Eichhoff OM, Schlegel NC, Dobbeling U, et al., 2008. In vivo switching of human melanoma cells between proliferative and invasive states. *Cancer Res,* 68: 650-6.
62. Kim HY, Lee H, Kim SH, Jin H, et al., 2017. Discovery of potential biomarkers in human melanoma cells with different metastatic potential by metabolic and lipidomic profiling. *Sci Rep,* 7: 8864.
63. Clauss J, Obenaus M, Miskey C, Ivics Z, et al., 2018. Efficient Non-Viral T-Cell Engineering by Sleeping Beauty Minicircles Diminishing DNA Toxicity and miRNAs Silencing the Endogenous T-Cell Receptors. *Hum Gene Ther,* 29: 569-584.
64. Mates L, Chuah MK, Belay E, Jerchow B, et al., 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet,* 41: 753-61.
65. Chae YK, Chang S, Ko T, Anker J, et al., 2018. Epithelial-mesenchymal transition (EMT) signature is inversely associated with T-cell infiltration in non-small cell lung cancer (NSCLC). *Sci Rep,* 8: 2918.
66. Alonso SR, Tracey L, Ortiz P, Perez-Gomez B, et al., 2007. A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. *Cancer Res,* 67: 3450-60.
67. Nieto MA, Huang RY, Jackson RA, and Thiery JP, 2016. Emt: 2016. *Cell,* 166: 21-45.
68. Lin K, Baritaki S, Militello L, Malaponte G, et al., 2010. The Role of B-RAF Mutations in Melanoma and the Induction of EMT via Dysregulation of the NF-kappaB/Snail/RKIP/PTEN Circuit. *Genes Cancer,* 1: 409-420.
69. Weiss MB, Abel EV, Mayberry MM, Basile KJ, et al., 2012. TWIST1 is an ERK1/2 effector that promotes invasion and regulates MMP-1 expression in human melanoma cells. *Cancer Res,* 72: 6382-92.
70. Dissanayake SK, Wade M, Johnson CE, O'Connell MP, et al., 2007. The Wnt5A/protein kinase C pathway mediates motility in melanoma cells via the inhibition of metastasis suppressors and initiation of an epithelial to mesenchymal transition. *J Biol Chem,* 282: 17259-71.
71. O'Connell MP, Marchbank K, Webster MR, Valiga AA, et al., 2013. Hypoxia induces phenotypic plasticity and therapy resistance in melanoma via the tyrosine kinase receptors ROR1 and ROR2. *Cancer Discov,* 3: 1378-93.
72. Derynck R and Zhang YE, 2003. Smad-dependent and Smad-independent pathways in TGF-beta family signalling. *Nature,* 425: 577-84.
73. Zavadil J and Bottinger EP, 2005. TGF-beta and epithelial-to-mesenchymal transitions. *Oncogene,* 24: 5764-74.
74. Morita T, Mayanagi T, and Sobue K, 2007. Dual roles of myocardin-related transcription factors in epithelial mesenchymal transition via slug induction and actin remodeling. *J Cell Biol,* 179: 1027-42.

75. Almanzar G, Olkhanud PB, Bodogai M, Dell'agnola C, et al., 2009. Sperm-derived SPANX-B is a clinically relevant tumor antigen that is expressed in human tumors and readily recognized by human CD4+ and CD8+ T cells. *Clin Cancer Res,* 15: 1954-63.
76. Yilmaz-Ozcan S, Sade A, Kucukkaraduman B, Kaygusuz Y, et al., 2014. Epigenetic mechanisms underlying the dynamic expression of cancer-testis genes, PAGE2, -2B and SPANX-B, during mesenchymal-to-epithelial transition. *PLoS One,* 9: e107905.
77. Woods K, Pasam A, Jayachandran A, Andrews MC, et al., 2014. Effects of epithelial to mesenchymal transition on T cell targeting of melanoma cells. *Front Oncol,* 4: 367.
78. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, et al., 2016. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science,* 352: 189-96.
79. Pencheva N, Tran H, Buss C, Huh D, et al., 2012. Convergent multi-miRNA targeting of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. *Cell,* 151: 1068-82.
80. Castello LM, Raineri D, Salmi L, Clemente N, et al., 2017. Osteopontin at the Crossroads of Inflammation and Tumor Progression. *Mediators Inflamm,* 2017: 4049098.
81. Cheng L, Lopez-Beltran A, Massari F, MacLennan GT, et al., 2018. Molecular testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision medicine. *Mod Pathol,* 31: 24-38.
82. Goding CR, 2011. Commentary. A picture of Mitf in melanoma immortality. *Oncogene,* 30: 2304-6.
83. Hoek KS and Goding CR, 2010. Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res,* 23: 746-59.
84. Vachtenheim J and Ondrusova L, 2015. Microphthalmia-associated transcription factor expression levels in melanoma cells contribute to cell invasion and proliferation. *Exp Dermatol,* 24: 481-4.
85. Carreira S, Goodall J, Denat L, Rodriguez M, et al., 2006. Mitf regulation of Dia1 controls melanoma proliferation and invasiveness. *Genes Dev,* 20: 3426-39.
86. Litvin O, Schwartz S, Wan Z, Schild T, et al., 2015. Interferon alpha/beta Enhances the Cytotoxic Response of MEK Inhibition in Melanoma. *Mol Cell,* 57: 784-96.
87. Shirley SH, Greene VR, Duncan LM, Torres Cabala CA, et al., 2012. Slug expression during melanoma progression. *Am J Pathol,* 180: 2479-89.
88. Tuncer E, Calcada RR, Zingg D, Varum S, et al., 2019. SMAD signaling promotes melanoma metastasis independently of phenotype switching. *J Clin Invest,* 129: 2702-2716.