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], Cédric Feschotte[2], 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

Abstract

The HML2 subfamily of HERV-K (henceforth HERV-K) represents the most recently endogenized retrovirus in the human genome. While the products of certain HERV-K genomic copies are expressed in normal tissues, they are upregulated in a number of pathological conditions, including various tumours. It remains unclear whether HERV-K(HML2)-encoded products overexpressed in cancer contribute to disease progression or are merely by-products of tumorigenesis. Here, we focus on the regulatory activities of the Long Terminal Repeats (LTR5_Hs) of HERV-K and on the potential role of the HERV-K-encoded Rec in melanoma. Our regulatory genomics analysis of LTR5_Hs loci indicates that Melanocyte Inducing Transcription Factor (MITF) binds to a canonical E-box motif (CA(C/T)GTG) within these elements in *proliferative* type of melanoma, and that depletion of MITF results in reduced HERV-K expression. In turn, experimentally depleting Rec in a *proliferative* melanoma cell line leads to lower mRNA levels of MITF and its predicted target genes. Furthermore, Rec knockdown leads to an upregulation of epithelial-to-mesenchymal associated genes and to an enhanced invasion phenotype of *proliferative* melanoma cells. Together these results suggest the existence of a regulatory loop between MITF and Rec that

may modulate the transition from *proliferative* to *invasive* stages of melanoma. Because HERV-K(HML2) elements are restricted to hominoid primates, these findings might explain certain species-specific features of melanoma progression and point to some limitations of animal models in melanoma studies.

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. Although some HERV-K(HML-2) insertions are polymorphic in the human population [2-5], suggesting relatively recent integration events, no fully-active provirus has been identified. Due to recombination events and other post-insertional mutations, the vast majority of HERV-K(HML2) members (~950-1200) are either solitary long terminal repeats (solo LTRs) or variously truncated copies [6, 7]. A subset of proviral copies (~90) are still capable of expressing retroviral proteins (e.g., Gag, Pro-Protease, Pol-Polymerase, Env-Envelope), and can apparently produce viral particles [8-10]. HERV-K(HML2) proviruses have been further classified as type 1 (~26%) or type 2 (~74%), depending on the presence or absence of a 292-bp deletion overlapping the Pol-Env junction [11]. Np9 and Rec are alternatively spliced products of Env [8, 12]. Type 1 proviruses may only encode Np9 [12], whereas type 2 proviruses may express the RNA isoforms encoding Env, Rec, and Np9 [13-15]. Based on the phylogeny of their LTR sequences, HERV-K(HML-2) can be further classified into LTR5_Hs, LTR5_A/LTR5_B subfamilies [16, 17]. Elements bearing LTR5_Hs represent the youngest integration events in the human genome; with 615 copies (45 proviruses and 570 solo LTRs), 90% of them being exclusive to humans, and absent in other great apes [16]. Type 1 and 2 proviruses are about evenly split among these elements, with 20 and 17 copies respectively [9].

Rec is a small RNA-binding protein considered to be a functional homolog of the HIV Rev accessory protein [13]. Rec contains a nuclear localization signal (NLS), which in addition to facilitating nuclear transport of the Rec protein, recognizes the Rec-responsive element (RcRE) within the viral RNAs [14, 18]. Binding of Rec to RcRE stabilizes incompletely spliced/unspliced viral transcripts and enhance their nuclear export [8, 19, 20] via the CRM1 export pathway [12, 13, 21]. Moreover, Rec interacts with Staufen-1 to further facilitate unspliced viral RNA transport [22]. Furthermore, 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 [23]. 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 [22, 24]. While it is evident that Rec must have first evolved to facilitate HERV-K replication, there is speculation that the protein may have been co-opted for physiological function(s), such as spermatogenesis [21, 25] and defence against viral infections in the early human embryo [26]. Although Np9 is expressed in various types of cancer [12, 27] and healthy tissues [28], unlike Rec, it has no identified function promoting the replication of HERV-K [29].

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 [28]. However, the landscape of transcribed HERV-K loci differs considerably between cell/tissue types [28]. Rec RNA/protein expression has been observed in human embryonic tissues, placenta and retina [15, 30]. Early embryonic cells appear favourable to HERV-K expression as proviral mRNAs and proteins are relatively abundant in both germinal- and pluripotent stem cells [26, 31]. While HERV-K expression has been observed at low levels in various somatic healthy tissues, it can be further induced by environmental stress, such as ultraviolet irradiation (e.g., UVB and UVC), starvation, or viral infection [32-36]. Furthermore, many studies have reported robust mRNA upregulation of several genomic loci of HERV-K in various disease states, including certain autoimmune diseases and several cancers, especially lung, breast cancers, germ cell tumours and melanoma [33, 37, 38] and reviewed in [39, 40]. Moreover, transcriptional activation and translation of various HERV-K products [37, 41] and even virus-like particles have been observed in specific cancer cells such as teratoma, embryonic carcinoma and melanomas [8, 10, 42, 43].

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. Around 50% of the HERV-K LTRs (including solo LTRs) retained some of their promoter/enhancer activity [44], and are differentially expressed in healthy and cancer tissues. The activated LTRs could act as alternative promoters and de-regulate tumour suppressor genes or proto-oncogenes (reviewed in [7, 45]). Furthermore, the Env protein of HERV-K, via its fusogenic property, is capable of inducing cell-cell fusion and could, therefore, contribute to tumour invasiveness [27, 33, 46]. Env could also have oncogenic properties through direct interference with cell signalling pathways (e.g., RSS/MEK/ERK) [47, 48]. Indeed, the overexpression of the HERV-K Env protein has been reported to induce an epithelial to mesenchymal transition (EMT)-like process [47, 49], a crucial event in oncogenesis leading to the more malignant phenotype. In addition, Env was suggested to promote tumorigenesis via modulation of the immune response [50-52]. Confusingly, HERV-K may have either a positive or negative effect on the immune system [53, 54]. Similar to Env, the expression of Rec and Np9 is significantly increased in several pathological conditions [55], including germ cell tumours and melanoma [12, 21, 37, 42, 43, 56-60], but again it remains unclear whether and how these factors could contribute to tumorigenesis and other disease states.

To begin answering this question, we investigate the regulatory genomic aspects of LTR5_Hs sequences, as well as 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 evaluate 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 [47] modulates the EMT-like process of cell transformation. However, in contrast to Env [47], Rec appears to inhibit the transition from proliferative to invasive state and as such may represent a protective factor in melanoma.

2. Materials and Methods

95 *2.1*

96 **Methods:**

97 **Generating the KD-Rec cell lines**

98 All 3 KD targeting the ERVK6(HML2.HOM) locus. KD1 targets 3'UTR, KD2 3'UTR, and
99 KD3 the gene body of Rec (**Figure S2**). **KD1** (as) 5-'ATCCATTCAACTCTGAGTGGA-
100 3', chr7:4,572,718-4,593,717; **KD2** (as) 5'-TAAGGCTGACTGTAGATGTAC-
101 3'chr7:4,580,334-4,601,333, chr7:4,599,327-4,599,347; **KD3** (as) 5'-
102 CAACGGTGCTCGATTGCGATG-3' chr7:4,585,379-4,585,399, chr7:4,593,883-
103 4,593,903. Note that due to the high similarity of the HML2 sequences, other loci are
104 also likely to be targeted.–The control (scrambled) KD was designed to target GFP
105 cDNA, GCGAAGTACGAATAGTTATCA from [61]. To generate stable KD clones, the
106 targeting sequences were cloned into pT2-MP71-(KD-Scr/1/2/3)-EGFP construct [62],
107 and co-transfected with SB100X transposase [63] into A375 and SKMel-28 cells. The
108 stable lines, which should also express a green fluorescent protein (GFP) reporter,
109 were selected using two-rounds of FACS sorting for the GFP-reporter signal.

110 **Cell Culture**

111 The melanoma cell line, A375 (ATCC CRL-1619) was cultivated in RPMI1640 (Thermo
112 Fisher Scientific) supplemented with 10% fetal bovine serum (FBS; PAA Laboratories,
113 Pasching, Austria), 80 U/mL penicillin and streptomycin (Lonza, Basel, Switzerland) at
114 37°C in a humidified atmosphere of 95% air and 5% CO_2. The melanoma cell line,
115 SKMel-28 (ATCC HTB-72) was maintained in Dulbecco's modified Eagle's medium
116 (DMEM, Life Technologies, England) containing GlutaMAX™-I supplement (Gibco,
117 catalogue number: 31966-021), with 10% FBS and penicillin and streptomycin in the
118 same condition as above.

119

120 **Transfection of melanoma cells and cell sorting**

121 To stably knock down (KD) Rec, A375 and SKMel-28 melanoma cell lines were
122 transfected with pT2-MP71-(KD1-Rec)-EGFP, pT2-MP71-(KD2-Rec)-EGFP, pT2-
123 MP71-(KD3-Rec)-EGFP plasmids in the presence of pT2-CAGGS-SB100X, using the
124 Neon transfection system (Life technologies). The electroporation setting was 1350 V
125 (pulse voltage), 20ms (pulse width), 2 pulses. 2×10^5 cells in 11 µl resuspension buffer
126 R were combined with 2 µl of purified plasmid mixture (50 ng transposase and 500 ng
127 transposon). Transfected cells were transferred to 6-well tissue culture plates

128 containing 2 ml culture medium. EGFP+ cells were sorted at weeks 1 and 3 post-
129 transfection by FACSAria II cell sorter (Becton).
130
131 **Quantitative Real-Time PCR and semi-quantitative PCR**
132 Total RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo Research)
133 according to the protocol of the manufacturer. Reverse transcription was performed
134 from 1 µg total RNA with the High Capacity RNA-to-cDNA kit (Applied Biosystems).
135 The isolated RNA would be treated with additional DNaseI (Invitrogen), to get rid of
136 potential DNA contamination. Quantitative real-time PCR (qRT-PCR) was carried out
137 on ABI7900 with PowerUp SYBR Green Master Mix kit (Applied Biosystems),
138 according to the recommendations of the manufacturer. For quantification of mRNA
139 expression levels, real-time PCR was run in triplicates for each cDNA sample, using
140 GAPDH and 18s RNA as the internal controls for SKMel-28 and A375 cells,
141 respectively. For primers and annealing temperatures, see **Table 1**. Data were
142 analysed using the comparative CT (2-ΔΔCT) method, which describes relative gene
143 expression. For RNA-seq validation, the qPCR control was transfected with pT2-CAG-
144 GFP instead of the KD constructs.
145
146 **Table 1**. The primers/oligos used in this study.

Name	**Application**	**Sequence (5' – 3')**	**Tm(°C)**
HML-2-Env	Real-Time qPCR	F: GCTGCCCTGCCAAACCTGAG R: CCTGAGTGACATCCCGCTTACC	60
HML-2-Gag	Real-Time qPCR	F: AGCAGGTCAGGTGCCTGTAACATT R: TGGTGCCGTAGGATTAAGTCTCCT	60
HML-2-Np9	Real-Time qPCR	F: AGATGTCTGCAGGTGTACCCA R: CTCTTGCTTTTCCCCACATTTC	60
HML-2-Rec	Real-Time qPCR	F: ATCGAGCACCGTTGACTCACAAGA R: GGTACACCTGCAGACACCATTGAT	60
MITF(M)	Real-Time qPCR	F:ATGCTGGAAATGCTAGAATATAATCACT R: GAATGTGTGTTCATGCCTGG	60
SPANXB1/B2	Real-Time qPCR	F: AGGCCAATGAGGCCAACAAGAC R: TCCTCCTGTAGCGAACCACTAG	60

18S RNA	Real-Time qPCR	F: GTAACCCGTTGAACCCCATT R: CCATCCAATCGGTAGTAGCG	60
GAPDH	Real-Time qPCR	F: CAATGACCCCTTCATTGACCTC R: AGCATCGCCCCACTTGATT	60

147

148 **RNA sequencing and data analysis**

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

155

156 Sample-specific barcoded sequencing reads were de-multiplexed from multiplexed
157 flow cells and by using CASAVA. 1.8.2 BCL files were converted to FASTQ format
158 files. The quality of the raw sequence reads was determined by using the FastQC.
159 Reads with a quality score below 30 were removed. We removed the highly variable
160 two nt from the ends of the remaining sequencing reads and mapped them over the
161 reference genome (Human hg19/GRCh37) and transcriptome model (hg19.refseq.gtf).
162 hg19/GRCh37 and hg19.refseq.gtf that were downloaded from USCS tables
163 (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/bigZips/). For mapping the reads,
164 we used our defined settings (i.e. *–alignIntronMin 20 –alignIntronMax 1000000 –*
165 *chimSegmentMin 15 –chimJunctionOverhangMin 15 –outFilterMultimapNmax 20*) to
166 STAR splice mapper. This approach generated STAR genome/transcriptome indices
167 of the reads and provided their respective RefSeq gtf (genes and rmsk) annotations.
168 Final read alignments having more than two mismatches per 100 bp were discarded.
169 We obtained uniquely mapped read counts over individual gene bodies using
170 featureCounts function from the subread package, at gene and transcript level using
171 RefSeq annotations. Scalable gene expression levels were calculated as transcript per
172 million (TPM) counts over the entire gene (defined as any transcript located between
173 transcription start and end sites). In the datasets generated in different batches, we
174 normalized the batch effect using the 'RUVSeq' package from R. To obtain differentially
175 expressed genes (DEGs). Firstly, we removed the extremely low expressing genes
176 (genes with less than five mapped reads) to avoid the unnecessary variation in

transcriptomes. We then used *DESeq2*, a R package, to determine the differentially expressed genes from replicated data. Data were normalized using the default functions of DESeq2. The package "*DESeq2"* provides methods to test for differential expression by use of generalized linear models; the estimates of dispersion and logarithmic fold changes incorporate data-driven prior distributions. Genes with adjusted p-value less than 5% (according to the FDR method from Benjamini-Hochberg), Log2 Foldchange |1| were assigned as differentially expressed. This strategy provided both quantification and statistical inference of systematic changes between conditions (with at least three replicates). The following samples were RNA sequenced: KD-Scr_A375, KD1-Rec_A375, KD2-Rec_A375, KD3-Rec_A375, KD-Scr_SKMel28, KD1-Rec_SKMel28, KD2-Rec_SKMel28, KD3-Rec_SKMel28 in two batches. Canonical pathways and biological function of the differentially expressed genes (DEGs) were further subjected to KEGG and Gene Set Enrichment Analyses (GSEA) using the webgestalt tool.

To estimate expression levels for repetitive elements, we calculated the CPM (counts normalized per million of total reads mappable on the human genome) for each Class II TE locus given in the repeat-masked annotations, downloaded from the UCSC data portal. We had removed those repetitive element's locus that overlays with the genic exons/UTRs or in the vicinity of 1 Kb from the TSS of annotated genes before mapping the sequencing reads over the individual locus. To calculate the family-wise expression of repetitive elements, we considered multi mapping reads only if they were mapping exclusively within a repetitive family. Then counted one alignment per read to calculate the CPM. The expression level of repeat families is calculated as Log2(CPM+1) prior to comparison. The locus-specific information was only calculated for those HML2-HERV-K loci that encodes for Rec transcripts. For this, we used "Salmon" that implements the expectation-maximization algorithm to re-distribute and assign the multi-mapping reads to the provided coordinates, based on evidence from uniquely mapped reads.

Single-cell RNA-seq data processing

We used the single-cell count matrix from GSE72056. We calculated the activity of genes in every cell at TPM expression levels. We considered samples expressing over 1000 genes with expression levels exceeding the defined threshold (Log2 TPM > 1).

211 We used Seurat_3.1.1, a package from R to normalize the datasets at the logarithmic
212 scale using "scale.factor = 10000". After normalization, we calculated scaled
213 expression (z-scores for each gene) for downstream dimension reduction. We used
214 the original annotations of datasets to classify the Malignant, Non-malignant and
215 Heterogenous cell-types. To define cell population clusters, we employed the
216 FindClusters function of "Seurat" using "PCA" as a reduction method. The specific
217 markers for each cluster identified by "Seurat" were determined by the
218 "FindAllMarkers" function, using "roc" as a test for significance. This provided us two
219 lists of gene sets. 1) Malignant vs. Non-Malignant differentially expressed genes and
220 2) genes differentially expressed between MITF-High and MITF-Low tumours, which
221 we applied for comparison with Rec-KD RNA-seq DEGs. Feature plots, violin plots and
222 heatmaps were constructed using default functions, except for the colour scale that
223 was set manually. The annotated cells were re-clustered using the methodologies
224 described above and visualized on the UMAP coordinates.
225
226 **ATAC-seq and ChIP-seq data analyses**
227 ATAC-seq and ChIP-seq raw datasets in sra format were downloaded from the listed
228 studies and converted to fastq format using sratools function fastq-dump –split-3.
229 Fastq reads were mapped against the hg19 reference genome with the modified
230 bowtie2 parameters: --very-sensitive-local. The modified settings are the following: -D
231 (number of matches) was set to 30, N (number of mismatches in seed) was set to 1,
232 and -R (number of re-seed) was set to 5. All unmapped reads with MAPQ < 10 and
233 PCR duplicates were removed using Picard and samtools. MACS2 called all the
234 ATAC-seq peaks with the parameters –nomodel -q 0.01 -B. Blacklisted regions were
235 excluded from called peaks
236 (https://www.encodeproject.org/annotations/ENCSR636HFF/). To generate a set of
237 unique peaks, we merged ATAC-seq peaks using the mergeBed function from
238 bedtools, where the distance between peaks was less than 50 base pairs. We then
239 intersected these peak sets with repeat elements from hg19 repeat-masked
240 coordinates using bedtools intersectBed with a 50% overlap. To calculate the
241 enrichment over the given repeat elements, we first extended 5 Kb upstream and 5 Kb
242 downstream coordinates from the left boundary (TSS) of respective elements in a
243 strand-specific manner. These 10 Kb windows were further divided into 100 bps bins
244 and tags (bedGraph output from MACS2) were counted in each bin. Tag counts in

245 each bin were normalized by the total number of tags per million in given samples and
246 presented as CPM per 100 bps. We averaged CPM values for each bin between
247 replicates before plotting the figures. To find the TF binding motifs, 25 bps sequences
248 were extended from either side of MITF ChIP-seq peak summits. The extended
249 sequences were analysed by RSAT tool (http://rsat.sb-roscoff.fr/). The TF binding
250 motifs were calculated from JASPAR libraries of human TF motifs.
251
252 **Cell invasion assay**
253 The cell invasion assays were performed using transwell chambers (8 μm pore size;
254 Corning Costar) according to the vendor's instructions. Briefly, the insert of the wells
255 was first coated with 50 μl BD matrigel. A375 cells were resuspended with RPMI1640
256 medium containing 0.5% FCS to reach a concentration of 5×10^5 cells/ml. Then 100 μl
257 cells from each group (stable KD-REC clones, pT2-CAG-GFP control) were seeded
258 into the upper chamber of the insert, with adding DMEM containing 10% FCS to the
259 lower chamber. After incubation at 37 °C in a humidified atmosphere of 5% CO_2 for 24
260 h, non-invaded cells were removed with gentle swabbing. In contrast, the invaded ones
261 were stained, and then images were taken with microscopy. Cell numbers were
262 calculated with ImageJ. The control was transfected with pT2-CAG-GFP instead of the
263 KD constructs.
264
265 **The accession of datasets used in this study:**
266 Encode TF ChiP-seq datasets:
267 ftp://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeSydhTfbs
268 GSE60663: H3K27Ac Primary Melanoma
269 GSE60663: H3K4Me3 Primary Melanoma
270 GSE60663: MITF ChIP-seq Primary Melanoma
271 GSE60664: RNA-seq Primary Melanoma
272 GSE46817: RNA-seq of 7 distinct melanoma lines
273 GSE46805: RNA-seq Melanocyte BRAF overexpression
274 GSE50681: MITF ChIP-seq Melanoma and Melanocyte post-treatment
275 GSE82330: ATAC-seq Melanoma A375
276 GSE72056: Patient's Single-cell RNA-seq data
277 GSE61966: MITF knock-down RNA-seq in 501 Melanoma and Melanocytes
278

3. Results

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

Cis-regulatory elements driving the transcription of the youngest subfamily of HERV-K are embedded within the human-specific LTR5 (LTR5_Hs) [64, 65]. Various transcription factors (TFs) have been reported to bind LTR5_Hs, including the pluripotency factors such as OCT4 [66] or the melanocyte inducing TF (MITF) [65, 67]. To examine more systematically the regulation of HERV-K transcription in different cell types, we mined publicly available data produced by the ENCODE consortium [68]. Briefly, we mapped the peaks from 142 ChIP-seq TF datasets from 7 cell lines (ENCODE) over 615 genomic sequences annotated as LTR5_Hs in the human genome (see **Methods**). Our integrative analysis of various TF occupancy revealed that multiple TFs bound a group of LTR5_Hs loci in a cell-type specific fashion. Notably, in H1 embryonic stem cells (H1ESCs) and leukemia K562 cells, LTR5_Hs elements tend to be bound by more TFs than in the other cell lines examined (**Figure 1**A). In the leukemia cell line K562, LTR5_Hs elements are enriched for binding of multiple TFs that control cell cycle and proliferation (e.g., cMYC, MAX, JUND, PU1 and P300, see **Figure 1**A), which might reflect the transformed phenotype of these cells. Curiously, NANOG binding appeared particularly prominent when H1 cells were cultured in naïve ("3iL") conditions [69] (**Figure 1**B).

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-specific isoform MITF(M) is expressed and promote the expression of HERV-K transcripts likely through direct binding to LTR5_Hs [67]. To further substantiate that MITF drives HERV-K expression in melanoma, we examined the expression levels of various transposable elements (TEs) in MITF-depleted melanocytes and melanoma cells using previously published RNA-seq datasets [70]. Knocking down MITF affected the expression of a small subsets of TE families that were differentially expressed in both melanocytes and melanoma, but there was only 5% overlap in differential TE expression between the two cell types (**Figure 1**C-D). Notably, KD MITF was

313 associated with a significant decrease in HERV-K expression (adjusted p-value < 0.01,
314 Benjamini-Hochberg (BH) correction) in melanoma, but not in melanocytes (**Figure**
315 **1**C), suggesting that MITF may regulate HERV-K expression specifically in
316 melanomas.
317
318 MITF(M) is a critical transcription factor associated with melanoma progression [71,
319 72], and is a sensitive marker of melanoma invasiveness: it is highly expressed (e.g.,
320 MITF-High) in the melanocytes and the *proliferative* state of melanoma, but lowly (e.g.,
321 MITF-Low) expressed in the *invasive* state [73]. To investigate how tightly HERV-K is
322 controlled in melanoma by MITF, we mined ChIP-seq and RNA-seq data from
323 melanocytes, various melanoma cell lines, and primary melanoma cultures (**Methods**).
324 Characterizing primary melanoma cultures according to their MITF levels (**Figure 1**E)
325 revealed an inverse correlation between LTR5_Hs/HERV-K RNA levels and
326 invasiveness (**Figure 2**A). To see if additional TE families respond to melanoma
327 proliferative/invasive status, we performed a more systematic analysis including other
328 TE families. This analysis revealed that the expression of certain ERVs, including
329 LTR5_Hs/HERV-K, LTR2C and LTR13, was correlated with that of genes
330 characterizing the *proliferative* state of melanoma (**Figure S1**A). When compared to
331 the overall expression of other TE families, LTR5_Hs/HERV-K was the most
332 abundantly expressed TE family and showed the strongest positive correlation with
333 MITF levels (rho = 0.43, corrected p-value < 0.2e-9) in the melanoma primary cultures
334 (n=10) (**Figure S1**B). These data suggested that similarly to MITF, LTR5_Hs/HERV-K
335 expression correlated negatively with melanoma invasiveness. Next, we characterized
336 the regulatory activation of LTR5_Hs in the *proliferative* type of melanoma cells by
337 layering ChIP-seq occupancy signals of MITF and H3K4Me3, H3K27Ac (**Figure 2**C
338 and **S1**C-D). Relative to the invasive state, the proliferative state showed enrichment
339 of LTR5_Hs for these “active” histone marks, coupled with MITF binding (**Figure 2**C
340 and **S1**C-D). We found that MITF binding at LTR5_Hs loci coincided significantly (p-
341 value = 0.026, chi-square test of Observed vs. Expected ratio) with the presence of the
342 canonical E-box motif (CA(C/T)GTG), which is recognized by MITF as well as c-MYC
343 and MAX TFs [74] (**Figure 2**C). Overall, our analysis suggests that the upregulation of
344 HERV-K expression in *proliferative* melanoma is likely driven at least in part by direct
345 binding of MITF over the LTR5_Hs through the E-box motif.
346

347 *3.2. Inhibition of BRAF^{V600E} mutant leads to the MITF binding over LTR5_Hs/HERV-K*
348 *in melanoma*

349 We also investigated the transcriptional regulation of HERV-K in the background
350 of the BRAF^{V600E} mutation, which characterizes an *invasive* type of melanoma [75].
351 PLX4032, a BRAF^{V600E} inhibitor commonly used in the clinic for the treatment of
352 melanoma, promotes the expression of MITF by several folds and causes a reversion
353 of melanoma from *invasive* to *proliferative* state [75]. To examine how PLX4032
354 treatment may affect HERV-K regulation, we reanalysed publicly available MITF ChIP-
355 seq datasets of primary melanocytes and a BRAF^{V600E} mutant melanoma cell line
356 (COLO829) with and without PLX4032 treatment (GSE50681). We observed that
357 PLX4032 treatment results in a substantial increase of MITF ChIP-seq peaks over
358 LTR5_Hs elements (123 out of 615 loci were bound under PLX4032 treatment but only
359 19 were bound in the control cells) (**Figure 2**D), suggesting that many LTR5_Hs loci
360 are bound by MITF when it is highly expressed. Among all TE families in the human
361 genome, the gain of MITF binding under PLX4032 treatment was unique to the
362 LTR5_Hs family (p-value < 0.01, chi-square test of observed vs. expected ratio)
363 (**Figure 2**E). While MITF is also expressed and binds to a subset of LTR5_Hs elements
364 in healthy pigmented melanocytes (n=33), there was no elevated MITF ChIP-seq
365 signal over LTR5_Hs elements upon PLX4032 treatment of those cells (**Figure 2**D).
366 Thus, the *proliferative* state of melanoma is associated with robust and widespread
367 binding of LTR5_Hs by MITF.

368 Next, we seek to gain insight into the transcriptional activity of individual HERV-
369 K loci in different melanoma states. Because available RNA-seq datasets were all
370 using short reads, which precludes locus-specific analysis of recent TEs like HERV-K,
371 we turned to chromatin ChIP-seq data (which is more mappable) as a proxy to predict
372 transcriptionally active HERV-K loci [76]. Namely, we used published dataset for
373 H3K27ac and MITF ChiP-seq (GSE60663) for low passage melanoma cells and also
374 ATAC-seq and H3K27Ac data for the A375 melanoma cell line (GSE82334). Using this
375 approach, we were able to identify twelve loci (nine HERV-K proviruses and 3 solo
376 LTR5_Hs) bound by MITF and/or with active chromatin marks (H3K27ac, ATAC peak)
377 in MITF-High primary cells (but not in MITF-Low cells) (**Table S2**). For example, **Figure**
378 **3**A presents an illustration of a region on chromosome 7 that includes the ERVK6
379 locus, which is structured as two nearly-identical tandemly arranged proviruses [77],
380 mirrored around a centrally located third LTR5_Hs, encoding for apparently intact

381 ORFs [28]. The flanking LTR5_Hs of the tandem proviruses showed higher H3K27Ac
382 signal in both *proliferative* MITF-High and A375 cells relative to MITF-Low cells, and
383 also significant ATAC-seq and H3K27ac signals in A375 (**Figure 3**A), indicating that
384 these loci are preferentially active in the *proliferative* type of melanoma. Moreover,
385 genome-wide analyses of ATAC-seq and active histone marks (H3K27ac and
386 H3K4Me1) in A375 cells further support the notion that multiple LTR5_Hs elements
387 (>50) may function as promoters or enhancers in this melanoma cell line (**Figure 3**B-
388 C). Collectively, these regulatory genomics data suggest that MITF binding contributes
389 to the activation of multiple LTR5_Hs loci copies (both solo and proviral) in the
390 *proliferative* type of melanoma.
391
392 *3.3. Depleting HERV-K-Rec may induce an EMT-like process in A375 melanoma cells*
393
394 From the nine identified actively transcribed proviruses, five were reported to
395 express Rec, from which three were also detected in various healthy tissues [28]. To
396 gain insight into a potential role of Rec in melanoma, we first used an RNAi approach
397 to deplete Rec expression in two melanoma cell lines (A375 and SKMel-28). A375 and
398 SKMEL-28 both carry the $BRAF^{V600E}$ mutation, but A375 has a higher proliferation rate,
399 whereas SKMEL-28 is more invasive [78], indicating their distinct invasion phenotypes.
400 To knockdown (KD) Rec, we designed three RNAi constructs (**Figure S2**) to target the
401 sequence of the ERVK6 locus, for which we observed marks of high transcriptional
402 activity in MITF-High and A375 melanoma cells (**Figure 3**A). Nevertheless, due to the
403 high level of Rec sequence similarity among several HERV-K proviruses [15], we
404 predicted our KD constructs would target Rec transcripts produced by most
405 transcriptionally-competent HERV-K proviruses [15, 28]. In addition, we generated
406 control cell lines expressing scrambled sequences. In total, we generated 10 stable
407 cell lines (6 KD and 4 controls) using *Sleeping Beauty* as a stable vector [62, 63]. Real-
408 time qRT-PCR experiments showed that Rec transcript levels were depleted in each
409 of the six independent KD cell lines isolated for A375 and SKMel-28 (**Figure 3**D). We
410 also observed that knocking down Rec also led to an increased level of alternative
411 HERV-K transcripts encoding Gag and Env in SKMel-28 (all three KD lines) and in
412 A375 (in the A375-KD1 line only) (**Figure 3**E).
413

414 To detect transcriptome changes upon Rec KD, we used RNA-seq to compare the
415 stable KD-Rec lines to their respective control lines. While SKMel-28 expresses Rec
416 from three genomic loci [43], we were able to unambiguously identify five Rec-
417 expressing loci in A375 (**Figure 3**F). Next, we computed the differentially expressed
418 genes (DEGs, |2|-fold and FDR < 0.05) between the Rec KD and their control lines
419 (**Methods**). We obtained 171 and 773 DEGs upon Rec KD in SKMel-28 and A375
420 respectively (**Figure 4**A-B**, S1**E**, Table S1,** see **Methods**). In SKMel-28, the 171 DEGs
421 were not significantly enriched for any Gene Ontology (GO) categories. In contrast, the
422 773 DEGs in A375 showed significant enrichment for several GO terms including
423 *pigmentation*, *cell migration, IL1 production* and *cell proliferation* (FDR < 0.05) (**Figure**
424 **4**A).
425
426 Notably, we identified a set of genes involved in epithelial-mesenchymal transition
427 (EMT)-like process, which play a critical role in phenotype switching from the
428 *proliferative* to *invasive* state [79-81]. Intriguingly, this set of DEGs included matrix
429 metallopeptidase 2 (MMP2), a classical marker of cell invasion along with additional
430 genes previously reported to drive the canonical EMT process [82], such as a zinc
431 finger protein (SNAI2), forkhead box C2 (FOXC2), cadherin 2 (CDH2) and goosecoid
432 Homeobox (GSC) (**Figure 4**B). Further, the genes upregulated upon Rec KD in A375
433 cells included factors such as WNT5B and GSK3B, which are known to induce the
434 EMT-like process and are activated when melanoma gets relapsed to metastasis after
435 failed treatment (**Figure 4**B) [83, 84]. During melanoma metastasis, a TF regulatory
436 network characterized by SNAI2^{low}/ZEB2^{low}/ZEB1^{high} is established [85-88]. The same
437 TF expression signature is observed upon KD of Rec in A375 cells (**Figure 4**B). Beside
438 canonical EMT markers, we also observe a strong upregulation of several members of
439 the SPANX (sperm protein associated with the nucleus in the X chromosome) family
440 of genes upon Rec KD in A375 cells (**Figure 4**B). The expression of SPANX genes is
441 normally confined primarily to the germline, but upregulation of SPANXB1/2 has been
442 previously reported in invasive melanoma [89]. Furthermore, we detected a robust
443 decrease in the transcript level of MITF(M), the master regulator of the melanocyte
444 differentiation program, upon Rec KD in A375 cells (**Figure 4**B). In addition to MITF,
445 we also observed a decreased expression of other melanocyte differentiation markers
446 (e.g., TYR, PMEL) [90] (**Figure 4**C). Together, the transcriptome changes we observed

upon Rec KD in A375 cells suggests that Rec expression is part of a regulatory cascade inhibiting the EMT-like process.

To corroborate these results, we chose two of the most DEGs in our transcriptomic analysis (SPANXB1/2 and MITF(M)) and performed RT-qPCR to compare the mRNA levels of these two genes in the three independent Rec KD A375 cell lines and their control lines. The results were consistent with the RNA-seq analysis: SPANXB1/2 was strongly upregulated (**Figure 4**D) and MITF(M) was robustly downregulated upon Rec KD (**Figure 4**E).

To further explore the idea that Rec KD in A375 cells recapitulates part of the expression profile of *invasive* melanoma, we examined the correlation between DEG upon Rec KD and those differentially expressed between *invasive* and *proliferative* primary melanoma cultures. We found a significant positive correlation pattern between the two sets of DEGs (Spearman's rho = 0.20, p-value < 0.05, asymptotic t approximation) (**Figure 4**F). Overall, these analyses suggest that depletion of Rec in A375 cells may result in an increased invasion phenotype.

3.4. Depletion of Rec results an enhanced cell invasion in A375 melanoma

To test whether Rec expression modulates the invasiveness of melanoma cells, we performed a trans-well invasion assay to compare the invasiveness of Rec KD A375 cells to their control lines. The results showed that all three Rec KD lines exhibited significantly elevated invasiveness (**Figure 5**A-B). These data support the hypothesis suggested by the transcriptome analyses that Rec may function as a suppressor of the EMT-like transition process, which modulates the invasiveness of melanoma.

3.5. Comparison of the Rec KD A375 with the malignant cells in melanoma patients

Next, we sought to examine whether a set of dysregulated genes in Rec KD A375 cells were also differentially expressed in malignant cell types. To assess this, we first performed a single-cell (sc)RNA-seq analysis (n=4645 cells) derived from

481 melanoma patients (n=19), which includes both malignant and non-malignant cells
482 [91]. Consistent with the original study, our analysis distinguishes the *malignant* from
483 the *non-malignant* cells based on the expression of the most variables genes (**Figure**
484 **6**A) and identifies four major *malignant* cell types, marked by MAGEC2, MITF, APOE
485 (Apolipoprotein E) and SPP1 (Osteopontin) expression, respectively (**Figure 6**A-C). It
486 is known that high levels of MITF, APOE and SPP1 mark different aspects of
487 melanoma progression, i.e. melanocyte differentiation, angiogenesis [92] and
488 inflammation [93], respectively. Interestingly, those three genes were all significantly
489 downregulated in the Rec KD A375 cells (**Figure 4**B, **7**A), suggesting that the KD of
490 Rec in those cells recapitulates multiple phases of melanoma.

491 Having characterized the expression profile of *malignant* and *non-malignant* cells from
492 melanoma patients, we next compared the ~700 DEGs (**Figure 7**B, see **Methods**)
493 between *non-malignant* and *malignant* cell types to those detected in our Rec KD A375
494 experiments (n=773). There were 94 overlapping genes, showing a significant
495 enrichment over a random expectation (p-value <0.05, chi-square test). Genes
496 showing malignant-specific expression tend to be upregulated in KD-Rec (p-value =
497 0.09, Wilcoxon-test, **Figure 7**C).

498 Lastly, we sought to assess whether genes predicted as downstream targets of MITF
499 from the same scRNA-seq analysis [91], were also affected in our Rec KD
500 experiments. We found that 41 out of the top 100 MITF predicted downstream targets
501 were included in our list of 773 DEGs, a significant enrichment over the random
502 expectation [p-value, test]. Strikingly, thirty-nine out of the 41 overlapping genes were
503 downregulated by Rec KD (p-value < $2.2e^{-16}$, chi-square test of observed vs expected).
504 These observations are consistent with the finding that MITF is downregulated in our
505 Rec KD lines, thereby affecting MITF downstream targets. (n=41 out of 100, Wilcoxon-
506 test, p-value < $1.2e^{-11}$) (**Figure 7**D). This analysis suggests that the downstream
507 targets of MITF might be depleted as the consequence of MITF downregulation upon
508 KD Rec. Thus, our KD Rec A375 model recapitulates MITF-associated aspects of
509 phenotype switching in melanoma patients at some degree.

510

511

512 **4. Discussion**

513

514 HERV-K resembles "complex" retroviruses such as HIV-1 by its ability to
515 encode multiple proteins, including Rec. While Env has the typical structure of a
516 transmembrane protein (e.g., extracellular, transmembrane, cytoplasmic domains),
517 Rec lacks the domains required for a membrane protein function (**Figure 8**A). Still, it
518 carries nuclear localization/export signals [8, 19]. Although Env and Rec are different
519 in sequence and function, our results suggest that Rec, like Env [47], exerts a
520 modulatory effect on the EMT-like process in cancer progression. However, the two
521 proteins appear to act in opposite direction. While Env overexpression (both RNA and
522 protein) may promote tumour progression through ERK1/2 activation [47], our KD
523 experiments indicate that Rec may inhibit the EMT process and oppose melanoma
524 metastasis. Interestingly, similar to Env, expression of the Gag protein encoded by
525 HERV-K, also increases the tumorigenic potential of melanoma cells [46]. Because we
526 observed that KD Rec leads to the upregulation of Gag and Env in one of the three KD
527 lines we established using 3 different RNAi constructs, we cannot exclude that some
528 of the transcriptomic and phenotypic (invasiveness) changes we observed are due to
529 indirect effects on Gag and Env levels. Importantly, however, we found that all three
530 KD lines behaved similarly in terms of some of the key differences in gene expression
531 (e.g. MITF(M), compare KD lines in **Figure 4**E) or invasiveness (**Figure 5**). Thus, it
532 appears that Rec's oncogenic potential may oppose that of Env or Gag proteins
533 encoded by HERV-K, and when HERV-K is overexpressed, the balance between these
534 different products may be key to maintain cellular homeostasis.

535

536 In about 50% of melanomas, the protein kinase BRAF is mutated ($BRAF^{V600E}$) [94].
537 MITF is considered as a crucial factor in melanoma invasiveness [95], and plays a
538 critical role in controlling $BRAF^{V600E}$ melanoma [96, 97]. Accordingly, the expression of
539 MITF has been widely used as a marker to distinguish between the *proliferative* and
540 *invasive* types of melanoma cells [95, 96, 98, 99]. Reduced MITF expression generates
541 *invasive* melanomas, with metastasis-promoting properties [96]. Our data suggest that
542 MITF activates LTR5_Hs in melanoma (not in melanocytes); thus, Rec expression
543 specifically marks the *proliferative* type of melanoma cells. Notably, the *proliferative*
544 phase of melanoma is still a controllable stage of melanoma progression, by contrast
545 to the *invasive* stage. If Rec expression is involved in maintaining a delicate balance
546 between cell proliferation and invasion, as this study suggests, then Rec activation

547 might be used as a sensitive marker to distinguish between the *proliferative* vs.
548 *invasive* phenotypes of melanoma.
549
550 Our study indicates that depletion of Rec in $BRAF^{V600E}$ mutant melanoma cells (A375)
551 results in reduced MITF level. Thus, our data suggest a model whereby a MITF-
552 dependent regulatory loop is disrupted upon Rec KD through (i) decreased MITF
553 binding to LTR5_Hs, which leads to (ii) reduced Rec levels and (iii) upregulation of
554 EMT markers, contributing to (iv) escalated tumour aggressiveness (**Figure 8**B).
555 Future research is required to decipher the mechanism by which MITF is
556 downregulated upon Rec KD. It would also be interesting to find out whether Rec
557 overexpression can lead to enhanced MITF expression in *invasive* melanoma, and
558 revert the *invasive* phenotype to a *proliferative* one.
559
560 Under healthy physiological conditions, apparently a limited number (n=18, or often
561 less) of HERV-K loci is active [28], whereas, in melanoma patients and/or cell lines,
562 the number of transcribed HERV-K loci is higher (n~24) [43]. These studies indicate
563 that there are specific HERV-K loci transcribed under pathological conditions. Although
564 forced expression of MITF(M) in non-malignant cells has been shown to enhance the
565 levels of various HERV-K-derived transcripts (e.g., Gag, Env, and Rec) [67], the
566 activated HERV-K loci might differ in both the level of expression and composition of
567 their HERV-K-derived transcripts. Furthermore, a variable set of HERV-K loci seems
568 to be expressed in different melanoma patients [43, 100], suggesting that variability in
569 the amount of HERV-K products may contribute to melanoma heterogeneity [37, 42],
570 and possibly the progression and outcome of the disease.
571
572 Our data suggest that, in addition to the Rec-expressing loci, MITF targets also solo
573 LTR5_Hss that could function as cis-regulatory sequences (enhancers, promoters) for
574 nearby genes [64]. Notably, some of the these LTRs harbour the E-box CA(C/T)GTG
575 motif, which is also recognized by cMYC and MAX oncoproteins [72], suggesting
576 another connection between HERV-K-derived sequences and oncogenesis. Because
577 several LTR5_Hs elements are insertionally polymorphic in the human population [2-
578 5], this type of variation could also have implications in melanoma progression and
579 deserves future work.
580

581 It has been widely discussed how precisely melanoma can be recapitulated in animal
582 models (reviewed in [101]). Interestingly, analysis of single-cell transcriptome data
583 from multiple human patients suggests a distinct separation of *proliferative* and
584 *invasive* types of cells within a patient [5]. However, these two phases do not co-occur
585 in mice [91, 102] As Rec (and HERV-K(HML2)) is hominid-specific, our study might
586 explain certain species-specific features of melanoma progression in humans, and
587 stress the potential limitations of using animal models in melanoma studies.
588
589
590 **Figure legends**
591
592 Figure 1:
593 A. Heatmap is summarizing the transcription factor (TF) occupancy (ChIP-seq
594 peaks, ENCODE) over individual LTR5_Hs loci in five human cell lines. This plot
595 includes 234 out of 615 LTR5_Hs loci annotated in the human Refseq genome,
596 which is occupied by at least one ChIP-seq peak. Each row represents an
597 individual locus of LTR5_Hs. Each column shows TF ChIP-seq peak occupancy
598 in a different cell line. Various TFs encompassing LTR5_Hs loci are manually
599 annotated in a given cell line. The cluster of TFs binding over a subset of
600 LTR5_Hs loci is shown in the red box.
601 B. Line plots showing the distribution of normalized ChIP-seq signal counts of
602 cMYC (left panel) and NANOG (right panel) over the annotated LTR5_Hs
603 sequences in the human genome. The comparison is made between primed
604 hESCs (H1) and K562 cells (left panel), as well as primed hESCs (H1) and
605 hESCs (H1), cultured in 3iL "naive" medium (right panel). The ChIP-seq signal
606 counts are calculated relative to primed ESCs, in a 3 Kb genomic window at the
607 left boundaries of LTR5_Hs in the human genome.
608 C. Heatmap displays the expression of differentially expressed TE families (at
609 counts per million level) upon MITF knockdown in melanocytes (left panel) and
610 melanoma (right panel). (N) denotes the number of differentially expressed TE
611 families (DESeq2 adjusted p-value < 0.05). Note that HERVK expression is
612 affected in melanoma (marked), but not in melanocytes.
613 D. Venn diagrams illustrating shared and unique differentially expressed TE
614 families between MITF-depleted (KD-MITF) melanocytes and melanoma cells.

615 The upper panel shows down-regulated TEs, whereas the lower panel shows
616 up-regulated TEs upon KD MITF.
617 E. Characterizing primary melanoma cultures according to their MITF expression
618 levels. Bar plots showing the level of MITF expression (Log2 TPM) in primary
619 melanoma cultures (n=10). According to their MITF expression level, cultures
620 are defined as MITF-High (*proliferative*) and MITF-Low (*invasive*) melanoma
621 primary cultures. The annotations indicating the *proliferative* and *invasive* states
622 are taken from the original study [35]. Note that in the following analyses we
623 used the data of MITF-High and MITF-Low as *proliferative* (n=3) and *invasive*
624 (n=2).
625
626 Figure 2:
627 A. Bar plot showing the LTR5_Hs_HERV-K expression (Log2 CPM) levels in
628 *proliferative* melanoma (MITF-High) and *invasive* melanoma (MITF-Low) cell
629 lines (GSE60664).
630 B. Scatter plot visualizes the average expression (Log2 CPM) (X-axis) and
631 standard deviation from mean (Y-axis) of transposable element (TE) families in
632 various primary melanoma cultures (n=10).
633 C. Line plot showing the MITF ChIP-seq signal occupancy in *proliferative*
634 melanoma (MITF-High) and *invasive* melanoma (MITF-Low) cell lines averaged
635 over all of the LTR5_Hs loci (n=615). Note the E-Box motif, identified in the
636 sequences of LTR5_Hs loci with MITF peaks. The E-Box motif was present in
637 57 out of 115 MITF bound LTR5_Hs sequence in A375 (P-value indicates the
638 probability of the presence of the E-Box motif in MITF bound vs unbound
639 LTR5_Hs loci determined by Chi-square test).
640 D. The effect of PLX4032 treatment on MITF binding at LTR5_Hs loci. Upper
641 panel: Line plot showing the MITF ChIP-seq signal averaged over LTR5_Hs loci
642 in primary melanocytes and melanoma cells with and without the treatment of
643 PLX4032. Lower panel: Heatmap of MITF ChIP-seq signals. Note: MITF
644 occupies LTR5_Hs genomic loci upon PLX4032 treatment, specifically in
645 melanoma (123 unique loci were bound, but not in melanocytes (over 50 loci).
646 E. Bar plot showing the number of MITF peaks detected over the genomic loci of
647 various TE families in melanoma cells treated with PLX4032. This plot shows
648 only those families which have at least five peaks over the individual locus.

649

650 Figure 3:

651 A. Genome browser tracks showing the H3K27Ac, ATAC-seq, and MITF signals at
652 specific LTR5_Hs-HERV-K sequences at the ERVK6 locus on chr7, encoding
653 for Gag, Env, and known to produce Rec (UniProt). Note (i) that ERVK6 is a
654 duplicated, tandemly arranged provirus, flanked by two LTR5_Hs sequences,
655 mirrored around a centrally located third LTR5_Hs (left and right panels, see also
656 **Figure S2**); (ii) that the signals are detectable at the flanking LTRs only; (iii) that
657 the signals are more robust in both the MITF-High primary melanoma cultures
658 and A375 cells; (iv) the slight differences in signal intensities over the two
659 flanking LTR5_Hs. The phylogenetic conservation of the locus is demonstrated
660 on the middle panel. The schematic structure of the tandemly repeated locus is
661 shown at the bottom.

662 B. Normalized ATAC-seq (nucleosome-associated) signals averaged over the
663 annotated LTR5_Hs sequences in the human genome in A375 cells.

664 C. Normalized ChIP-seq signals of H3K27Ac (left panel) and H3K4Me1 (right panel)
665 histone marks (corresponding to active chromatin states) averaged over the
666 annotated LTR5_Hs sequences in the human genome in A375 cells. Note that
667 the analysis shows only loci that are mappable from the analysed ChIP-seq data
668 (likely underestimated).

669 D. Knocking down (KD) Rec in A375 and SKMel-28 cells using RNA interference.
670 Stable KD lines were selected in two-rounds by FACS sorting for the GFP-
671 reporter signal. Bar plots are showing the normalized RT-qPCR quantification of
672 Rec mRNA levels with three KD-Rec constructs. Control, KD-Scr (scrambled).

673 E. Knocking down Rec affects the transcription of alternative HERV-K products,
674 primarily in SKMel-28. Bar plot is showing the effect of knocking down Rec (KD-
675 Rec) on the expression of alternative HERV-K products, determined by real-time
676 qPCR. Note: KD-Rec led to an increased level of Gag in SKMel-28 (all the three
677 biological replicates), whereas Env in A375 (KD1 only). Control, KD-Scr
678 (scrambled).

679 F. Five HERV-K loci express Rec in A375. The bar plots demonstrate the locus
680 specificity of the RNAi strategy against Rec. Relative quantification of Rec
681 expression (CPM) upon KD at specific genomic loci, compared with scrambled

682 control (KD1-3 vs. KD-Scr), showing the significantly affected Rec expressing
683 loci in A375 cells.
684
685 Figure 4:
686 A. Significant Gene Ontology (GO) Biological Process terms enrichment analysis
687 of (differentially expressed genes) DEGs in Rec KD A375 vs. KD-Scr_A375
688 (scrambled control).
689 B. Volcano plot showing the comparison of Log2 fold-changes (x-axis) and
690 adjusted p-value (Log10, y-axis) of DEGs upon KD-Rec vs. KD-Scr in A375
691 cells. The annotated genes are among the top dysregulated genes that are
692 known to mark either the *invasive* (red) or the *proliferative* (blue) melanoma.
693 C. Scatter plot showing the pairwise comparison of DEGs between primary
694 melanoma cultures and A375 cells. Primary melanoma culture data derive from
695 the comparison between MITF-high (*proliferative*) and MITF-low (*invasive*) type
696 of melanomas. A375 data derive from the comparison between KD-Rec and
697 KD-Scr (scrambled control). The annotated genes are affected in both
698 comparisons in a similar fashion. Note (i) that MITF-Low (*invasive*) melanoma
699 and KD-Rec share the upregulation of inflammation-associated genes; (ii) The
700 downregulated genes are classical markers of *proliferative* melanomas.
701 D. Validating depleted SPANXB1/B2 expression upon Rec KD in A375 cells vs.
702 control, using real-time qPCR (previously detected as DEG in RNA-seq
703 experiments).
704 E. Validating depleted MITF-M expression (previously detected as DEG in RNA-
705 seq experiments) upon Rec KD in A375 vs. control, using real-time qPCR (see
706 **Methods**).
707 F. Bar plot representing means ± SEM from RNA-seq expression values (Log2
708 TPM) for melanocyte markers (e.g., TYR and PMEL) upon Rec-KD in A375 vs.
709 KD-Scr.
710
711 Figure 5:
712 A. Representative inverted light microscope images of trans-well invasion assay
713 performed in A375 melanoma cells upon knocking-down Rec by three KD
714 constructs (KD1-3). The Ctrl expresses pT2-CAG-GFP instead of the KD
715 constructs. Scale bars, 400 μm.

716 B. Quantification of cells invading in the trans-well invasion assay. ImageJ counted
717 the number of invaded cells. Columns indicate the average number of invaded
718 cells per field from two independent experiments. Error bars represent the
719 means±s.d. of two independent experiments. Mann Whitney test; **P ≤ 0.01,
720 ****p ≤ 0.0001 (See **Methods** for details).
721
722 Figure 6:
723 A. UMAP visualisation of the thirteen distinct cell clusters identified by
724 10XGenomics scRNA-seq in patient-derived melanoma [91]. The identified
725 clusters are annotated as *Malignant* and *Non-Malignant* (left panel) or according
726 to their origin (right panel), following original annotations [91]). Note that (i)
727 despite their proximity to malignant cell clusters, CD248-labelled cells in
728 Custer9 are most probably *Non-Malignant* cells [103]; (ii) This approach
729 distinguishes four *Malignant* clusters (e.g., Clusters 10-13).
730 B. Feature plots based on the UMAP plot shown on (A) visualizing the expression
731 of selected markers used to identify the distinct cell-types of melanoma. The
732 Colour intensity gradient indicates the expression of the depicted marker gene.
733 Each dot represents an individual cell. Dots in green denote lower, whereas in
734 dark-red, a higher level of gene expression in a given single cell.
735 C. Violin plots visualize the density and distribution of expression of distinct marker
736 genes in *Malignant* melanoma clusters (10-13).
737
738 Figure 7:
739 A. Heatmap visualization of scaled (Log2 TPM) expression of SPP1, MITF, APOE
740 and MAGEC2 upon knockdown (KD) of Rec in A375 cells (three replicates).
741 B. Heatmap visualization of scaled expression [log TPM (transcripts per million)]
742 values of distinctive set of ~ 1500 genes (Log2 fold change > 1 and > 50% of
743 cells expressing either in *Non-malignant* or *Malignant* cells), which are
744 differentially expressed in scRNA seq data. Colour scheme is based on Z-score
745 distribution from -2.0 (gold) to 2.0 (purple). Top colour bars highlight
746 representative gene sets specific to the respective clusters.
747 C. Box plot showing the comparison of differentially expressed gene sets in
748 *Malignant* and *Non-malignant* cells (patient-derived scRNA-seq) and KD-Rec
749 vs. KD-Scr in A375.

750 D. Box plot showing the comparison differential expression of a gene set of "MITF-
751 High" (consisting of MITF itself and the top 100 genes correlated with MITF
752 expression) between *Malignant* and *Non-malignant* cells (patient-derived
753 scRNA-seq) and KD-Rec vs. KD-Scr in A375.
754
755 Figure 8:
756 A. Schematic structure of the HERV-K-derived alternative splice products, Env and
757 Rec, and their effect on the endothelial-mesenchymal transition (EMT)-like
758 process of cancer progression (EMT(L). The signal peptide (SP);
759 transmembrane subunit (TM); Splice acceptor (Sa); Splice donor (Sd).
760 B. Modeling the interaction loop between MITF and LTR5_Hs and Rec in
761 melanoma. MITF binds to LTR5_Hs and drives Rec expression from certain
762 Type 2 proviruses. Rec marks the proliferative stage melanoma cells and,
763 similarly to MITF, inhibits the endothelial-mesenchymal transition (EMT)-like
764 process (EMT(L). Depletion of Rec results in MITF downregulation and the
765 failure to maintain the balance between cell proliferation and invasion.
766
767 Figure S1:
768 A. (Left panel) Ranked correlation matrix of genes and 42 transposable elements
769 (TE) families, which are differentially expressed between *proliferative* (MITF-
770 High) and *invasive* (MITF-Low) primary melanoma cultures. Genes and TE
771 families are manually annotated. Colour codes on the left side of the heat map
772 panel represent the unsupervised clusters of most variable genes and TEs
773 across the samples. On the right, selected genes are listed that are positively
774 or negatively correlated with LTR5_Hs expression across the melanoma
775 primary cultures.
776 B. Distribution of ChIP-seq signal of the H3K4Me3 histone mark occupancy,
777 corresponding to active promoters averaged over LTR5_Hs loci in MITF-High
778 (*proliferative*) vs. MITF-Low (*invasive*) state of primary melanoma cultures.
779 C. Distribution of ChIP-seq signal of the H3K27Ac histone mark occupancy (lower
780 panel) averaged over LTR5_Hs loci in MITF-High (*proliferative*) vs. MITF-Low
781 (*invasive*) state of primary melanoma cultures.
782 D. Heatmap displays the DEGs upon KD-Rec in A375 cells vs. KD-Scr.
783
784 Figure S2:

785 Schematic of the positions targeted by the knockdown (KD1,2,3) constructs on the
786 ERVK6(HML2.HOM) locus. ERVK6 is a tandemly organized HERV-K(HML2)
787 provirus, possessing three LTRs. The genomic coordinates of the locus on chr
788 seven are indicated. Both are type 2 and encode identical Rec proteins (UniProt).
789 A human-specific provirus encodes these Rec proteins. Note that due to the high
790 conservation of Rec sequences [15], the KD was targeted multiple loci in the human
791 genome.
792

793 **Acknowledgement**: M.S. was supported by a presidential postdoctoral fellowship
794 from Cornell University. H.C. was supported by from the China Scholarship Council
795 (CSC). Z.I. was funded by European Research Council, ERC Advanced [ERC-2011-
796 ADG 294742]. Work on host-virus interactions in the Feschotte lab is supported by R35
797 GM122550 from the National Institutes of Health.
798

799 **Author Contributions:** Conceptualization, M.S, H.C. M.B. and Z.I.; methodology,
800 H.C.; analyses, M.S.; supervision, original draft preparation, Z.I.; writing, Z.I., M.S., and
801 C.F.; review and editing, all authors; All authors have read and agreed to the published
802 version of the manuscript.
803
804
805
806 References1.
807

808 1. **Lander ES, Linton LM, Birren B, Nusbaum C, et al.,** 2001. Initial sequencing and
809 analysis of the human genome. *Nature,* **409**: 860-921.
810 2. **Hughes JF and Coffin JM,** 2004. Human endogenous retrovirus K solo-LTR formation
811 and insertional polymorphisms: implications for human and viral evolution. *Proc Natl Acad Sci*
812 *U S A,* **101**: 1668-72.
813 3. **Turner G, Barbulescu M, Su M, Jensen-Seaman MI, et al.,** 2001. Insertional
814 polymorphisms of full-length endogenous retroviruses in humans. *Curr Biol,* **11**: 1531-5.
815 4. **Belshaw R, Dawson AL, Woolven-Allen J, Redding J, et al.,** 2005. Genomewide
816 screening reveals high levels of insertional polymorphism in the human endogenous retrovirus
817 family HERV-K(HML2): implications for present-day activity. *J Virol,* **79**: 12507-14.
818 5. **Macfarlane CM and Badge RM,** 2015. Genome-wide amplification of proviral
819 sequences reveals new polymorphic HERV-K(HML-2) proviruses in humans and chimpanzees
820 that are absent from genome assemblies. *Retrovirology,* **12**: 35.
821 6. **Subramanian RP, Wildschutte JH, Russo C, and Coffin JM,** 2011. Identification,
822 characterization, and comparative genomic distribution of the HERV-K (HML-2) group of
823 human endogenous retroviruses. *Retrovirology,* **8**: 90.
824 7. **Babaian A and Mager DL,** 2016. Endogenous retroviral promoter exaptation in human
825 cancer. *Mob DNA,* **7**: 24.
826 8. **Lower R, Tonjes RR, Korbmacher C, Kurth R, et al.,** 1995. Identification of a Rev-
827 related protein by analysis of spliced transcripts of the human endogenous retroviruses
828 HTDV/HERV-K. *J Virol,* **69**: 141-9.

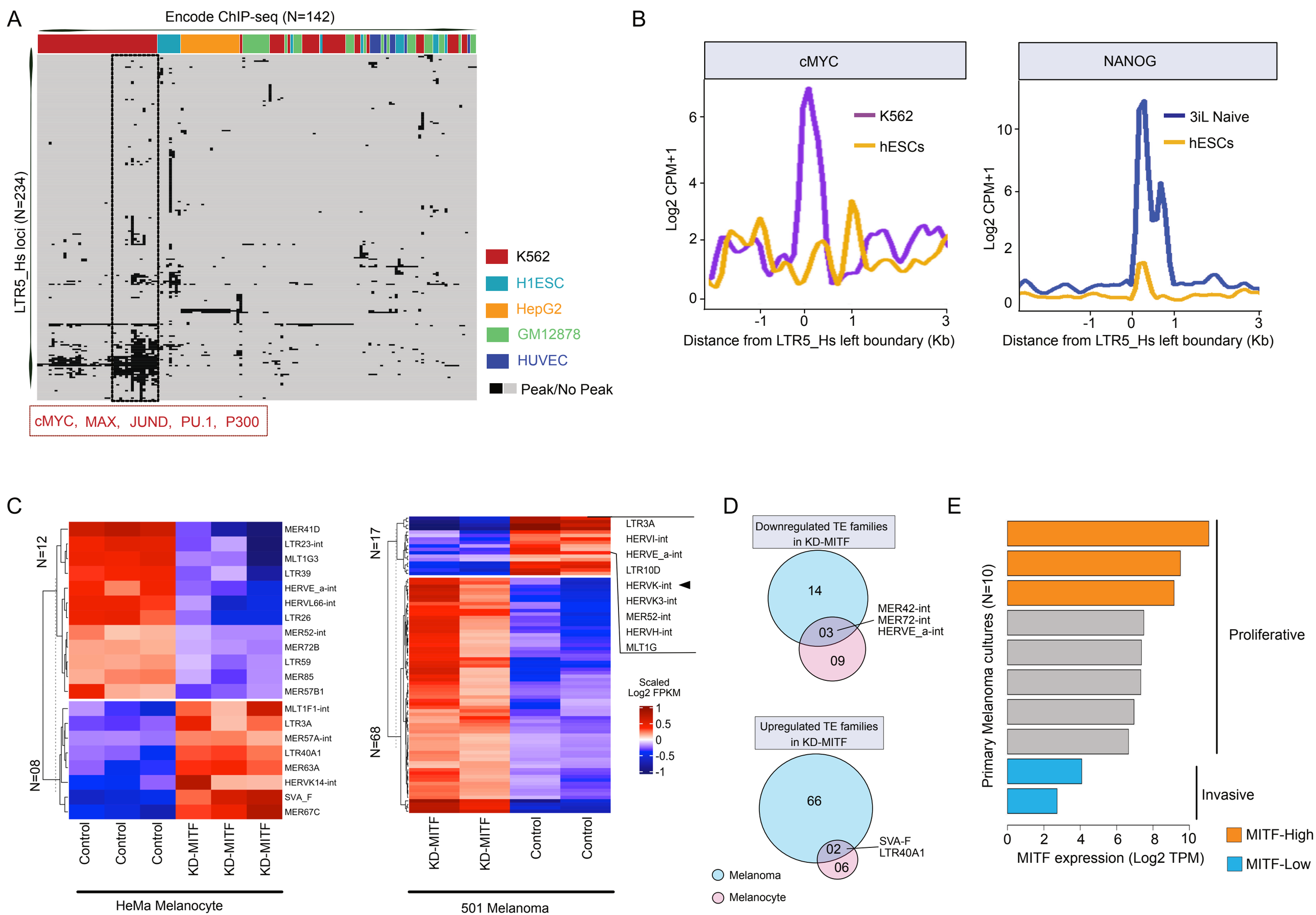
829 9. **Dewannieux M, Blaise S, and Heidmann T,** 2005. Identification of a functional
830 envelope protein from the HERV-K family of human endogenous retroviruses. *J Virol,* **79**:
831 15573-7.
832 10. **Contreras-Galindo R, Kaplan MH, Dube D, Gonzalez-Hernandez MJ, et al.,** 2015.
833 Human Endogenous Retrovirus Type K (HERV-K) Particles Package and Transmit HERV-K-
834 Related Sequences. *J Virol,* **89**: 7187-201.
835 11. **Lower R, Boller K, Hasenmaier B, Korbmacher C, et al.,** 1993. Identification of
836 human endogenous retroviruses with complex mRNA expression and particle formation. *Proc*
837 *Natl Acad Sci U S A,* **90**: 4480-4.
838 12. **Armbruester V, Sauter M, Krautkraemer E, Meese E, et al.,** 2002. A novel gene from
839 the human endogenous retrovirus K expressed in transformed cells. *Clin Cancer Res,* **8**: 1800-
840 7.
841 13. **Magin C, Lower R, and Lower J,** 1999. cORF and RcRE, the Rev/Rex and RRE/RxRE
842 homologues of the human endogenous retrovirus family HTDV/HERV-K. *J Virol,* **73**: 9496-
843 507.
844 14. **Magin-Lachmann C, Hahn S, Strobel H, Held U, et al.,** 2001. Rec (formerly Corf)
845 function requires interaction with a complex, folded RNA structure within its responsive
846 element rather than binding to a discrete specific binding site. *J Virol,* **75**: 10359-71.
847 15. **Mayer J, Ehlhardt S, Seifert M, Sauter M, et al.,** 2004. Human endogenous retrovirus
848 HERV-K(HML-2) proviruses with Rec protein coding capacity and transcriptional activity.
849 *Virology,* **322**: 190-8.
850 16. **Buzdin A, Ustyugova S, Khodosevich K, Mamedov I, et al.,** 2003. Human-specific
851 subfamilies of HERV-K (HML-2) long terminal repeats: three master genes were active
852 simultaneously during branching of hominoid lineages. *Genomics,* **81**: 149-56.
853 17. **Macfarlane C and Simmonds P,** 2004. Allelic variation of HERV-K(HML-2)
854 endogenous retroviral elements in human populations. *J Mol Evol,* **59**: 642-56.
855 18. **Langner JS, Fuchs NV, Hoffmann J, Wittmann A, et al.,** 2012. Biochemical analysis
856 of the complex between the tetrameric export adapter protein Rec of HERV-K/HML-2 and the
857 responsive RNA element RcRE pck30. *J Virol,* **86**: 9079-87.
858 19. **Yang J, Bogerd HP, Peng S, Wiegand H, et al.,** 1999. An ancient family of human
859 endogenous retroviruses encodes a functional homolog of the HIV-1 Rev protein. *Proc Natl*
860 *Acad Sci U S A,* **96**: 13404-8.
861 20. **Wodrich H and Krausslich HG,** 2001. Nucleocytoplasmic RNA transport in retroviral
862 replication. *Results Probl Cell Differ,* **34**: 197-217.
863 21. **Boese A, Sauter M, Galli U, Best B, et al.,** 2000. Human endogenous retrovirus
864 protein cORF supports cell transformation and associates with the promyelocytic leukemia zinc
865 finger protein. *Oncogene,* **19**: 4328-36.
866 22. **Hanke K, Hohn O, Liedgens L, Fiddeke K, et al.,** 2013. Staufen-1 interacts with the
867 human endogenous retrovirus family HERV-K(HML-2) rec and gag proteins and increases
868 virion production. *J Virol,* **87**: 11019-30.
869 23. **Denne M, Sauter M, Armbruester V, Licht JD, et al.,** 2007. Physical and functional
870 interactions of human endogenous retrovirus proteins Np9 and rec with the promyelocytic
871 leukemia zinc finger protein. *J Virol,* **81**: 5607-16.
872 24. **Kaufmann S, Sauter M, Schmitt M, Baumert B, et al.,** 2010. Human endogenous
873 retrovirus protein Rec interacts with the testicular zinc-finger protein and androgen receptor. *J*
874 *Gen Virol,* **91**: 1494-502.
875 25. **Costoya JA, Hobbs RM, Barna M, Cattoretti G, et al.,** 2004. Essential role of Plzf in
876 maintenance of spermatogonial stem cells. *Nat Genet,* **36**: 653-9.
877 26. **Grow EJ, Flynn RA, Chavez SL, Bayless NL, et al.,** 2015. Intrinsic retroviral
878 reactivation in human preimplantation embryos and pluripotent cells. *Nature,* **522**: 221-5.
879 27. **Chen T, Meng Z, Gan Y, Wang X, et al.,** 2013. The viral oncogene Np9 acts as a
880 critical molecular switch for co-activating beta-catenin, ERK, Akt and Notch1 and promoting
881 the growth of human leukemia stem/progenitor cells. *Leukemia,* **27**: 1469-78.
882 28. **Schmitt K, Heyne K, Roemer K, Meese E, et al.,** 2015. HERV-K(HML-2) rec and np9
883 transcripts not restricted to disease but present in many normal human tissues. *Mob DNA,* **6**:
884 4.

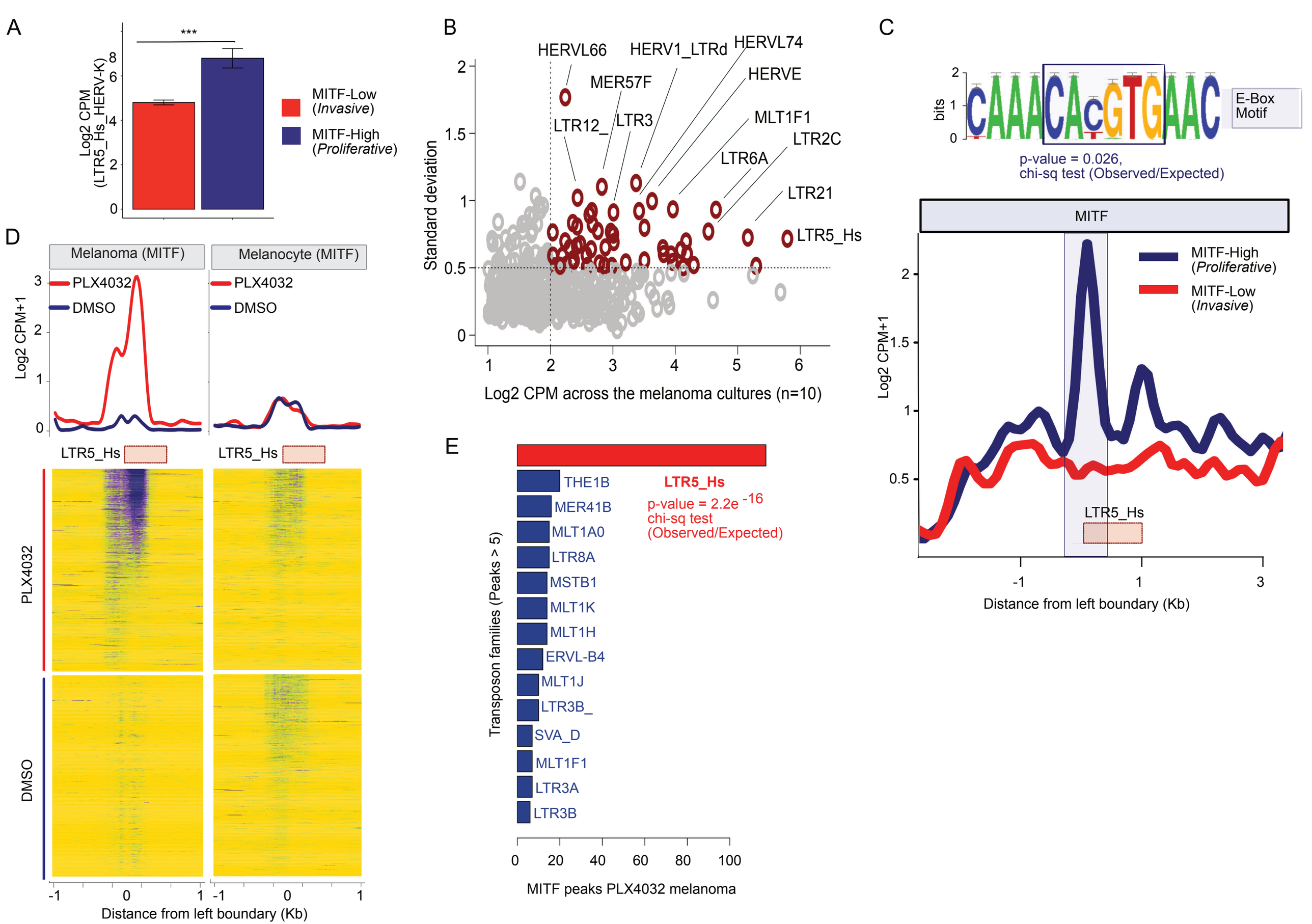
885 29. **Garcia-Montojo M, Doucet-O'Hare T, Henderson L, and Nath A,** 2018. Human
886 endogenous retrovirus-K (HML-2): a comprehensive review. *Crit Rev Microbiol,* **44**: 715-738.
887 30. **Andersson AC, Venables PJ, Tonjes RR, Scherer J, et al.,** 2002. Developmental
888 expression of HERV-R (ERV3) and HERV-K in human tissue. *Virology,* **297**: 220-5.
889 31. **Izsvak Z, Wang J, Singh M, Mager DL, et al.,** 2016. Pluripotency and the endogenous
890 retrovirus HERVH: Conflict or serendipity? *Bioessays,* **38**: 109-17.
891 32. **Kreimer U, Schulz WA, Koch A, Niegisch G, et al.,** 2013. HERV-K and LINE-1 DNA
892 Methylation and Reexpression in Urothelial Carcinoma. *Front Oncol,* **3**: 255.
893 33. **Serafino A, Balestrieri E, Pierimarchi P, Matteucci C, et al.,** 2009. The activation of
894 human endogenous retrovirus K (HERV-K) is implicated in melanoma cell malignant
895 transformation. *Exp Cell Res,* **315**: 849-62.
896 34. **Balestrieri E, Argaw-Denboba A, Gambacurta A, Cipriani C, et al.,** 2018. Human
897 Endogenous Retrovirus K in the Crosstalk Between Cancer Cells Microenvironment and
898 Plasticity: A New Perspective for Combination Therapy. *Front Microbiol,* **9**: 1448.
899 35. **Reiche J, Pauli G, and Ellerbrok H,** 2010. Differential expression of human
900 endogenous retrovirus K transcripts in primary human melanocytes and melanoma cell lines
901 after UV irradiation. *Melanoma Res,* **20**: 435-40.
902 36. **Schanab O, Humer J, Gleiss A, Mikula M, et al.,** 2011. Expression of human
903 endogenous retrovirus K is stimulated by ultraviolet radiation in melanoma. *Pigment Cell*
904 *Melanoma Res,* **24**: 656-65.
905 37. **Muster T, Waltenberger A, Grassauer A, Hirschl S, et al.,** 2003. An endogenous
906 retrovirus derived from human melanoma cells. *Cancer Res,* **63**: 8735-41.
907 38. **Stengel S, Fiebig U, Kurth R, and Denner J,** 2010. Regulation of human endogenous
908 retrovirus-K expression in melanomas by CpG methylation. *Genes Chromosomes Cancer,* **49**:
909 401-11.
910 39. **Clausen J,** 2003. Endogenous retroviruses and MS: using ERVs as disease markers.
911 *Int MS J,* **10**: 22-8.
912 40. **Matteucci C, Balestrieri E, Argaw-Denboba A, and Sinibaldi-Vallebona P,** 2018.
913 Human endogenous retroviruses role in cancer cell stemness. *Semin Cancer Biol,* **53**: 17-30.
914 41. **Hahn S, Ugurel S, Hanschmann KM, Strobel H, et al.,** 2008. Serological response
915 to human endogenous retrovirus K in melanoma patients correlates with survival probability.
916 *AIDS Res Hum Retroviruses,* **24**: 717-23.
917 42. **Buscher K, Trefzer U, Hofmann M, Sterry W, et al.,** 2005. Expression of human
918 endogenous retrovirus K in melanomas and melanoma cell lines. *Cancer Res,* **65**: 4172-80.
919 43. **Schmitt K, Reichrath J, Roesch A, Meese E, et al.,** 2013. Transcriptional profiling of
920 human endogenous retrovirus group HERV-K(HML-2) loci in melanoma. *Genome Biol Evol,*
921 **5**: 307-28.
922 44. **Buzdin A, Kovalskaya-Alexandrova E, Gogvadze E, and Sverdlov E,** 2006. At least
923 50% of human-specific HERV-K (HML-2) long terminal repeats serve in vivo as active
924 promoters for host nonrepetitive DNA transcription. *J Virol,* **80**: 10752-62.
925 45. **Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, et al.,** 2007. Distinct
926 classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate
927 cancer. *Nature,* **448**: 595-9.
928 46. **Oricchio E, Sciamanna I, Beraldi R, Tolstonog GV, et al.,** 2007. Distinct roles for
929 LINE-1 and HERV-K retroelements in cell proliferation, differentiation and tumor progression.
930 *Oncogene,* **26**: 4226-33.
931 47. **Lemaitre C, Tsang J, Bireau C, Heidmann T, et al.,** 2017. A human endogenous
932 retrovirus-derived gene that can contribute to oncogenesis by activating the ERK pathway and
933 inducing migration and invasion. *PLoS Pathog,* **13**: e1006451.
934 48. **Li M, Radvanyi L, Yin B, Li J, et al.,** 2017. Downregulation of Human Endogenous
935 Retrovirus Type K (HERV-K) Viral env RNA in Pancreatic Cancer Cells Decreases Cell
936 Proliferation and Tumor Growth. *Clin Cancer Res,* **23**: 5892-5911.
937 49. **Zhou F, Li M, Wei Y, Lin K, et al.,** 2016. Activation of HERV-K Env protein is essential
938 for tumorigenesis and metastasis of breast cancer cells. *Oncotarget,* **7**: 84093-84117.

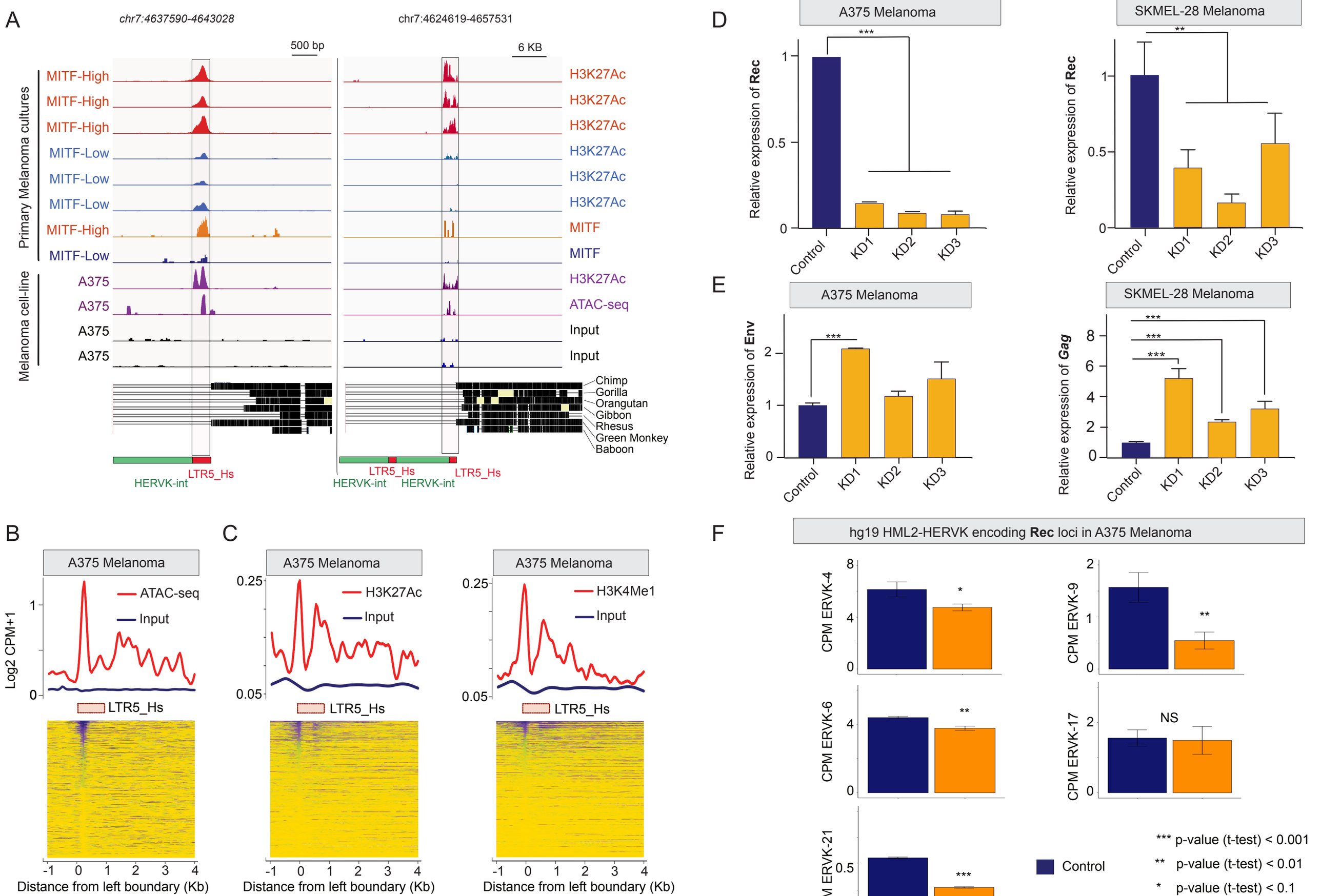
939 50. **Ishida T, Obata Y, Ohara N, Matsushita H, et al.,** 2008. Identification of the HERV-K
940 gag antigen in prostate cancer by SEREX using autologous patient serum and its
941 immunogenicity. *Cancer Immun,* **8**: 15.
942 51. **Riker AI, Enkemann SA, Fodstad O, Liu S, et al.,** 2008. The gene expression profiles
943 of primary and metastatic melanoma yields a transition point of tumor progression and
944 metastasis. *BMC Med Genomics,* **1**: 13.
945 52. **Wallace TA, Downey RF, Seufert CJ, Schetter A, et al.,** 2014. Elevated HERV-K
946 mRNA expression in PBMC is associated with a prostate cancer diagnosis particularly in older
947 men and smokers. *Carcinogenesis,* **35**: 2074-83.
948 53. **Dong J, Sedrak M, Rincon L, Sun H, et al.,** 2014. K-type human endogenous
949 retroviral elements in human melanoma. *Advances in Genomics and Genetics*: 153.
950 54. **Chiappinelli KB, Strissel PL, Desrichard A, Li H, et al.,** 2015. Inhibiting DNA
951 Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous
952 Retroviruses. *Cell,* **162**: 974-86.
953 55. **Nelson PN, Lever AM, Smith S, Pitman R, et al.,** 1999. Molecular investigations
954 implicate human endogenous retroviruses as mediators of anti-retroviral antibodies in
955 autoimmune rheumatic disease. *Immunol Invest,* **28**: 277-89.
956 56. **Hirschl S, Schanab O, Seppele H, Waltenberger A, et al.,** 2007. Sequence variability
957 of retroviral particles derived from human melanoma cells melanoma-associated retrovirus.
958 *Virus Res,* **123**: 211-5.
959 57. **Ruprecht K, Mayer J, Sauter M, Roemer K, et al.,** 2008. Endogenous retroviruses
960 and cancer. *Cell Mol Life Sci,* **65**: 3366-82.
961 58. **Hohn O, Hanke K, and Bannert N,** 2013. HERV-K(HML-2), the Best Preserved Family
962 of HERVs: Endogenization, Expression, and Implications in Health and Disease. *Front Oncol,*
963 **3**: 246.
964 59. **Galli UM, Sauter M, Lecher B, Maurer S, et al.,** 2005. Human endogenous retrovirus
965 rec interferes with germ cell development in mice and may cause carcinoma in situ, the
966 predecessor lesion of germ cell tumors. *Oncogene,* **24**: 3223-8.
967 60. **Singh S, Kaye S, Francis N, Peston D, et al.,** 2013. Human endogenous retrovirus K
968 (HERV-K) rec mRNA is expressed in primary melanoma but not in benign naevi or normal skin.
969 *Pigment Cell Melanoma Res,* **26**: 426-8.
970 61. **Lu X, Sachs F, Ramsay L, Jacques PE, et al.,** 2014. The retrovirus HERVH is a long
971 noncoding RNA required for human embryonic stem cell identity. *Nat Struct Mol Biol,* **21**: 423-
972 5.
973 62. **Clauss J, Obenaus M, Miskey C, Ivics Z, et al.,** 2018. Efficient Non-Viral T-Cell
974 Engineering by Sleeping Beauty Minicircles Diminishing DNA Toxicity and miRNAs Silencing
975 the Endogenous T-Cell Receptors. *Hum Gene Ther,* **29**: 569-584.
976 63. **Mates L, Chuah MK, Belay E, Jerchow B, et al.,** 2009. Molecular evolution of a novel
977 hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.
978 *Nat Genet,* **41**: 753-61.
979 64. **Fuentes DR, Swigut T, and Wysocka J,** 2018. Systematic perturbation of retroviral
980 LTRs reveals widespread long-range effects on human gene regulation. *Elife,* **7**.
981 65. **Manghera M and Douville RN,** 2013. Endogenous retrovirus-K promoter: a landing
982 strip for inflammatory transcription factors? *Retrovirology,* **10**: 16.
983 66. **Sun Q, Yang J, Xing G, Sun Q, et al.,** 2008. Expression of GSDML Associates with
984 Tumor Progression in Uterine Cervix Cancer. *Translational Oncology,* **1**: 73-IN1.
985 67. **Katoh I, Mirova A, Kurata S, Murakami Y, et al.,** 2011. Activation of the long terminal
986 repeat of human endogenous retrovirus K by melanoma-specific transcription factor MITF-M.
987 *Neoplasia,* **13**: 1081-92.
988 68. **Consortium EP,** 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project.
989 *Science,* **306**: 636-40.
990 69. **Chan YS, Goke J, Ng JH, Lu X, et al.,** 2013. Induction of a human pluripotent state
991 with distinct regulatory circuitry that resembles preimplantation epiblast. *Cell Stem Cell,* **13**:
992 663-75.

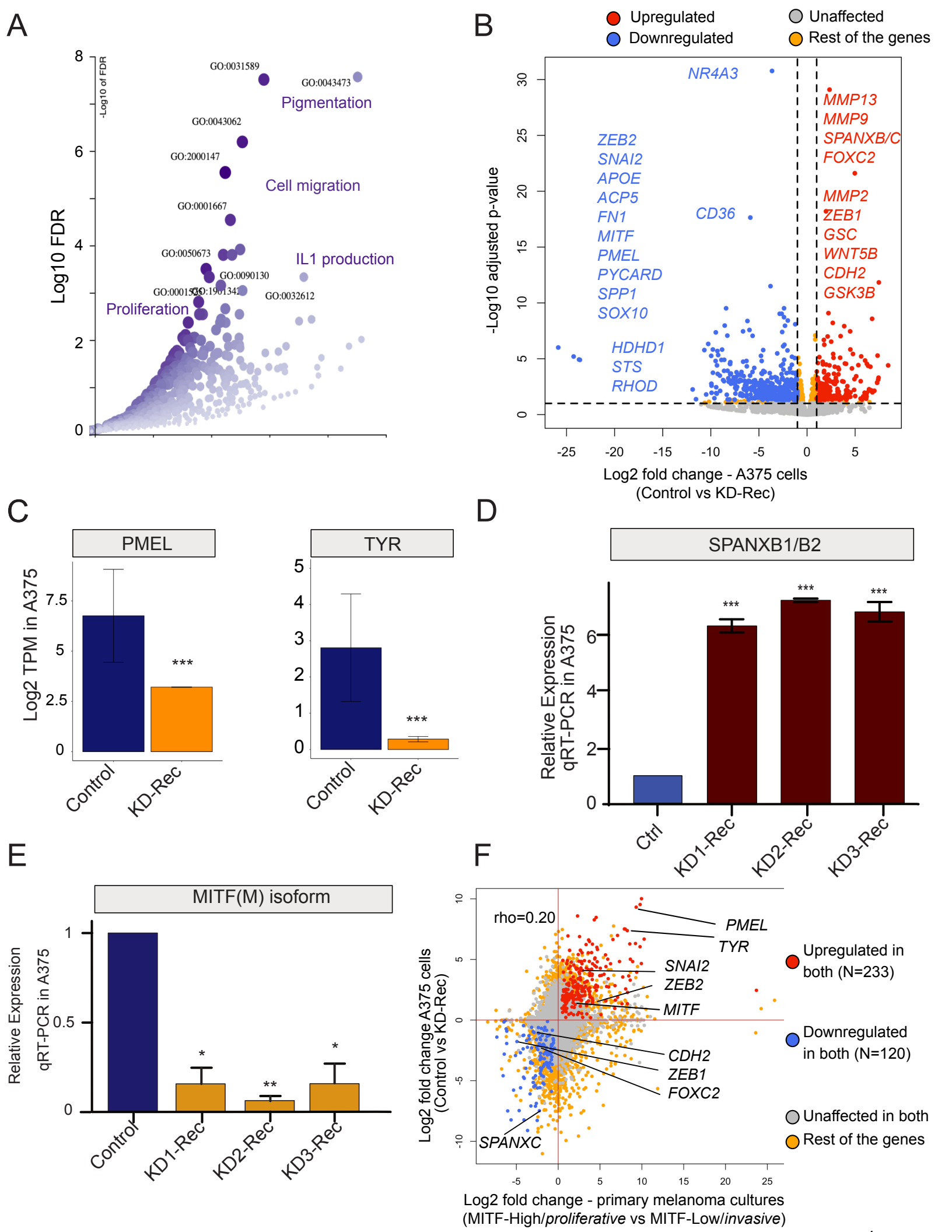
70. **Laurette P, Strub T, Koludrovic D, Keime C, et al.,** 2015. Transcription factor MITF and remodeller BRG1 define chromatin organisation at regulatory elements in melanoma cells. *Elife,* **4**.
71. **Levy C, Khaled M, and Fisher DE,** 2006. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med,* **12**: 406-14.
72. **Steingrimsson E, Copeland NG, and Jenkins NA,** 2004. Melanocytes and the microphthalmia transcription factor network. *Annu Rev Genet,* **38**: 365-411.
73. **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.
74. **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.
75. **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.
76. **Sultana T, Zamborlini A, Cristofari G, and Lesage P,** 2017. Integration site selection by retroviruses and transposable elements in eukaryotes. *Nat Rev Genet,* **18**: 292-308.
77. **Reus K, Mayer J, Sauter M, Scherer D, et al.,** 2001. Genomic organization of the human endogenous retrovirus HERV-K(HML-2.HOM) (ERVK6) on chromosome 7. *Genomics,* **72**: 314-20.
78. **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.
79. **Li FZ, Dhillon AS, Anderson RL, McArthur G, et al.,** 2015. Phenotype switching in melanoma: implications for progression and therapy. *Front Oncol,* **5**: 31.
80. **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.
81. **Nieto MA, Huang RY, Jackson RA, and Thiery JP,** 2016. Emt: 2016. *Cell,* **166**: 21-45.
82. **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.
83. **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.
84. **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.
85. **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.
86. **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.
87. **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.
88. **Denecker G, Vandamme N, Akay O, Koludrovic D, et al.,** 2014. Identification of a ZEB2-MITF-ZEB1 transcriptional network that controls melanogenesis and melanoma progression. *Cell Death Differ,* **21**: 1250-61.
89. **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.

1048 90. **Goldstein NB, Koster MI, Hoaglin LG, Wright MJ, et al.,** 2016. Isolating RNA from
1049 precursor and mature melanocytes from human vitiligo and normal skin using laser capture
1050 microdissection. *Exp Dermatol,* **25**: 805-11.
1051 91. **Tirosh I, Izar B, Prakadan SM, Wadsworth MH, 2nd, et al.,** 2016. Dissecting the
1052 multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science,* **352**: 189-
1053 96.
1054 92. **Pencheva N, Tran H, Buss C, Huh D, et al.,** 2012. Convergent multi-miRNA targeting
1055 of ApoE drives LRP1/LRP8-dependent melanoma metastasis and angiogenesis. *Cell,* **151**:
1056 1068-82.
1057 93. **Castello LM, Raineri D, Salmi L, Clemente N, et al.,** 2017. Osteopontin at the
1058 Crossroads of Inflammation and Tumor Progression. *Mediators Inflamm,* **2017**: 4049098.
1059 94. **Cheng L, Lopez-Beltran A, Massari F, MacLennan GT, et al.,** 2018. Molecular
1060 testing for BRAF mutations to inform melanoma treatment decisions: a move toward precision
1061 medicine. *Mod Pathol,* **31**: 24-38.
1062 95. **Goding CR,** 2011. Commentary. A picture of Mitf in melanoma immortality. *Oncogene,*
1063 **30**: 2304-6.
1064 96. **Hoek KS and Goding CR,** 2010. Cancer stem cells versus phenotype-switching in
1065 melanoma. *Pigment Cell Melanoma Res,* **23**: 746-59.
1066 97. **Vachtenheim J and Ondrusova L,** 2015. Microphthalmia-associated transcription
1067 factor expression levels in melanoma cells contribute to cell invasion and proliferation. *Exp*
1068 *Dermatol,* **24**: 481-4.
1069 98. **Carreira S, Goodall J, Denat L, Rodriguez M, et al.,** 2006. Mitf regulation of Dia1
1070 controls melanoma proliferation and invasiveness. *Genes Dev,* **20**: 3426-39.
1071 99. **Hoek KS, Eichhoff OM, Schlegel NC, Dobbeling U, et al.,** 2008. In vivo switching of
1072 human melanoma cells between proliferative and invasive states. *Cancer Res,* **68**: 650-6.
1073 100. **Litvin O, Schwartz S, Wan Z, Schild T, et al.,** 2015. Interferon alpha/beta Enhances
1074 the Cytotoxic Response of MEK Inhibition in Melanoma. *Mol Cell,* **57**: 784-96.
1075 101. **Perez-Guijarro E, Day CP, Merlino G, and Zaidi MR,** 2017. Genetically engineered
1076 mouse models of melanoma. *Cancer,* **123**: 2089-2103.
1077 102. **Tuncer E, Calcada RR, Zingg D, Varum S, et al.,** 2019. SMAD signaling promotes
1078 melanoma metastasis independently of phenotype switching. *J Clin Invest,* **129**: 2702-2716.
1079 103. **MacFadyen JR, Haworth O, Roberston D, Hardie D, et al.,** 2005. Endosialin (TEM1,
1080 CD248) is a marker of stromal fibroblasts and is not selectively expressed on tumour
1081 endothelium. *FEBS Lett,* **579**: 2569-75.
1082

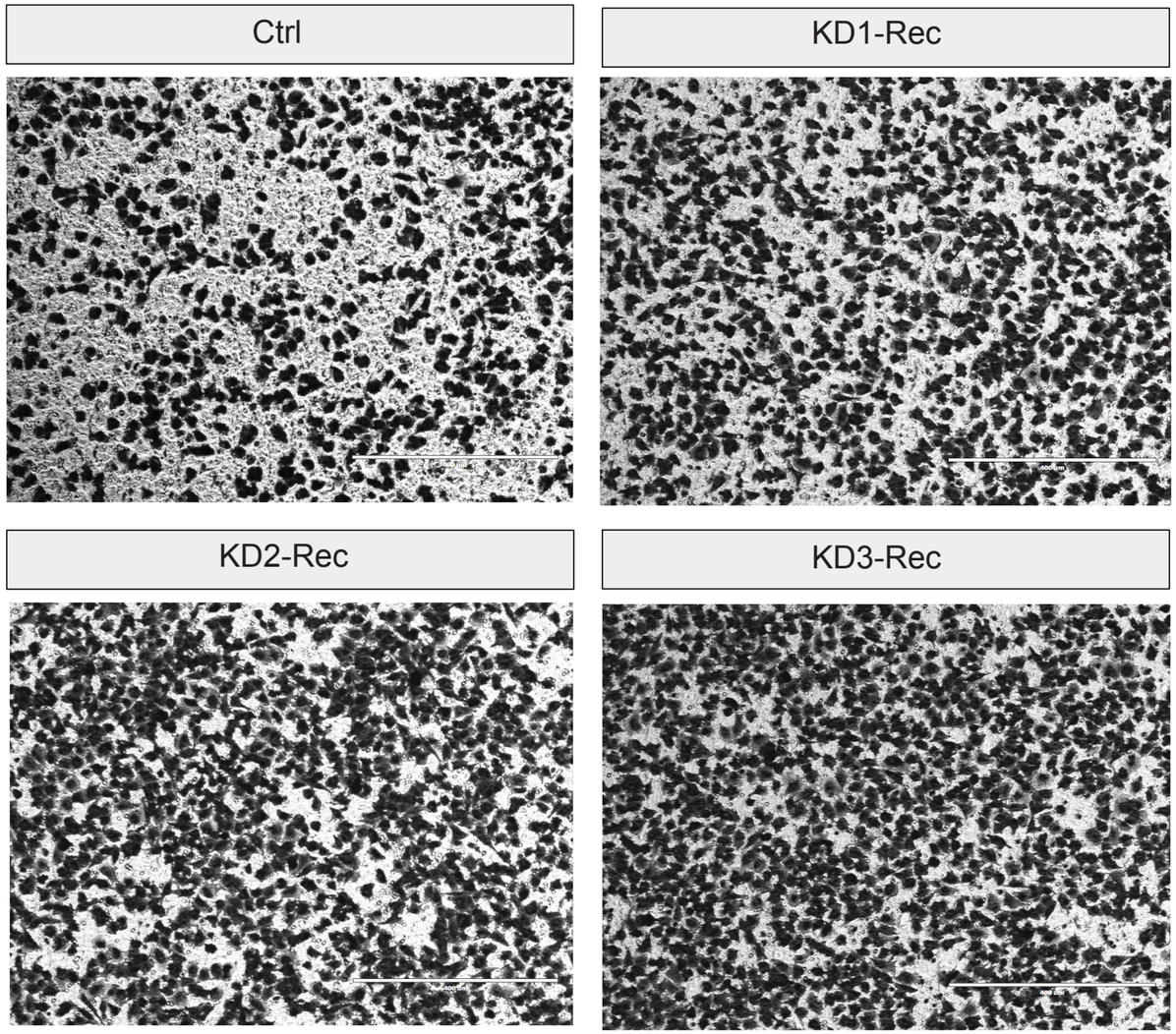

A
Encode ChIP-seq (N=142)
LTR5_Hs loci (N=234)
K562
H1ESC
HepG2
GM12878
HUVEC
Peak/No Peak
cMYC, MAX, JUND, PU.1, P300
B
cMYC
K562
hESCs
Log2 CPM+1
Distance from LTR5_Hs left boundary (Kb)
NANOG
3iL Naive
hESCs
Log2 CPM+1
Distance from LTR5_Hs left boundary (Kb)
C
N=12
N=08
MER41D
LTR23-int
MLT1G3
LTR39
HERVE_a-int
HERVL66-int
LTR26
MER52-int
MER72B
LTR59
MER85
MER57B1
MLT1F1-int
LTR3A
MER57A-int
LTR40A1
MER63A
HERVK14-int
SVA_F
MER67C
Control
Control
Control
KD-MITF
KD-MITF
KD-MITF
HeMa Melanocyte
N=17
N=68
LTR3A
HERVI-int
HERVE_a-int
LTR10D
HERVK-int
HERVK3-int
MER52-int
HERVH-int
MLT1G
Scaled Log2 FPKM
KD-MITF
KD-MITF
Control
Control
501 Melanoma
D
Downregulated TE families in KD-MITF
14
03
09
MER42-int
MER72-int
HERVE_a-int
Upregulated TE families in KD-MITF
66
02
06
SVA-F
LTR40A1
Melanoma
Melanocyte
E
Primary Melanoma cultures (N=10)
Proliferative
Invasive
MITF expression (Log2 TPM)
MITF-High
MITF-Low


A
***
Log2 CPM (LTR5_Hs_HERV-K)
MITF-Low (Invasive)
MITF-High (Proliferative)
B
HERVL66
HERV1_LTRd
HERVL74
MER57F
HERVE
LTR12_
LTR3
MLT1F1
LTR2C
LTR6A
LTR21
LTR5_Hs
Standard deviation
Log2 CPM across the melanoma cultures (n=10)
C
bits
E-Box Motif
p-value = 0.026,
chi-sq test (Observed/Expected)
MITF
MITF-High (Proliferative)
MITF-Low (Invasive)
Log2 CPM+1
LTR5_Hs
Distance from left boundary (Kb)
D
Melanoma (MITF)
Melanocyte (MITF)
PLX4032
DMSO
Log2 CPM+1
LTR5_Hs
PLX4032
DMSO
Distance from left boundary (Kb)
E
LTR5_Hs
p-value = 2.2e -16
chi-sq test
(Observed/Expected)
THE1B
MER41B
MLT1A0
LTR8A
MSTB1
MLT1K
MLT1H
ERVL-B4
MLT1J
LTR3B_
SVA_D
MLT1F1
LTR3A
LTR3B
Transposon families (Peaks > 5)
MITF peaks PLX4032 melanoma

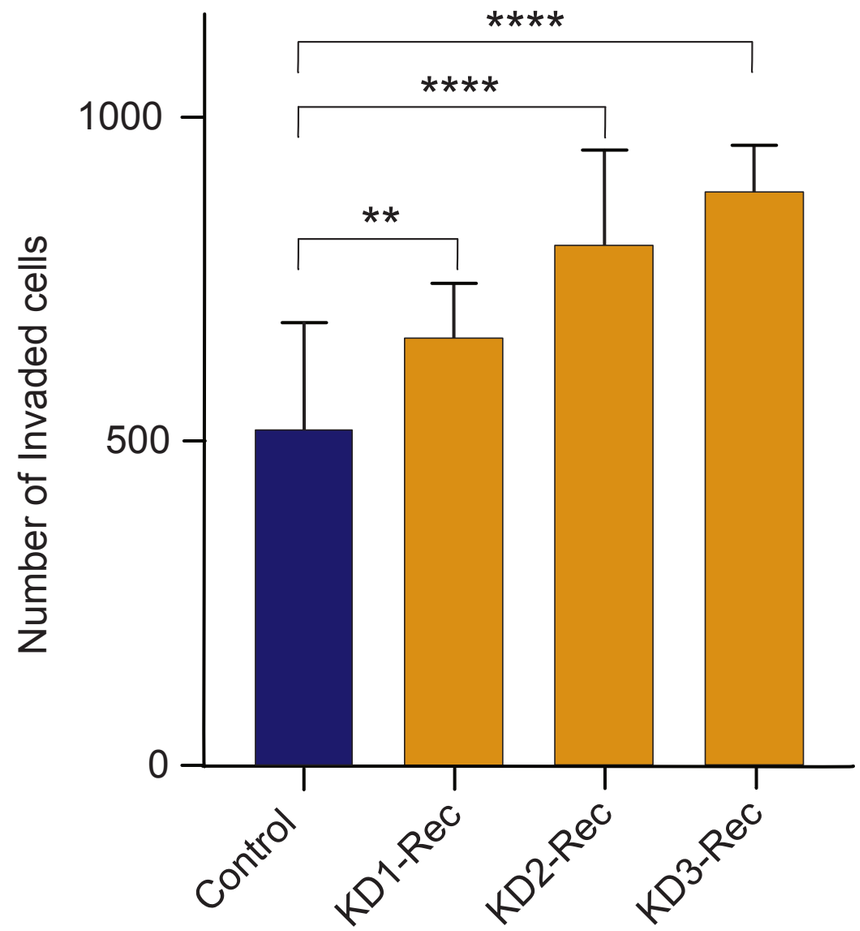

A
chr7:4637590-4643028
chr7:4624619-4657531
500 bp
6 KB
Primary Melanoma cultures
Melanoma cell-line
MITF-High
MITF-Low
A375
H3K27Ac
MITF
ATAC-seq
Input
Chimp
Gorilla
Orangutan
Gibbon
Rhesus
Green Monkey
Baboon
HERVK-int
LTR5_Hs
B
A375 Melanoma
ATAC-seq
Input
Log2 CPM+1
LTR5_Hs
Distance from left boundary (Kb)
C
A375 Melanoma
H3K27Ac
Input
H3K4Me1
D
A375 Melanoma
SKMEL-28 Melanoma
Relative expression of Rec
Control
KD1
KD2
KD3
E
Relative expression of Env
Relative expression of Gag
F
hg19 HML2-HERVK encoding Rec loci in A375 Melanoma
CPM ERVK-4
CPM ERVK-9
CPM ERVK-6
CPM ERVK-17
CPM ERVK-21
NS
Control
KD-Rec
*** p-value (t-test) < 0.001
** p-value (t-test) < 0.01
* p-value (t-test) < 0.1
NS Not significant


A
Log10 FDR
-Log10 of FDR
GO:0031589
GO:0043473
Pigmentation
GO:0043062
GO:2000147
Cell migration
GO:0001667
GO:0050673
IL1 production
GO:0090130
GO:0032612
Proliferation
B
Upregulated
Downregulated
Unaffected
Rest of the genes
-Log10 adjusted p-value
NR4A3
MMP13
MMP9
SPANXB/C
FOXC2
ZEB2
SNAI2
APOE
ACP5
FN1
MITF
PMEL
PYCARD
SPP1
SOX10
CD36
MMP2
ZEB1
GSC
WNT5B
CDH2
GSK3B
HDHD1
STS
RHOD
Log2 fold change - A375 cells
(Control vs KD-Rec)
C
PMEL
TYR
Log2 TPM in A375
Control
KD-Rec
D
SPANXB1/B2
Relative Expression
qRT-PCR in A375
Ctrl
KD1-Rec
KD2-Rec
KD3-Rec
E
MITF(M) isoform
Relative Expression
qRT-PCR in A375
Control
KD1-Rec
KD2-Rec
KD3-Rec
F
rho=0.20
PMEL
TYR
SNAI2
ZEB2
MITF
CDH2
ZEB1
FOXC2
SPANXC
Log2 fold change A375 cells
(Control vs KD-Rec)
Upregulated in
both (N=233)
Downregulated
in both (N=120)
Unaffected in both
Rest of the genes
Log2 fold change - primary melanoma cultures
(MITF-High/proliferative vs MITF-Low/invasive)

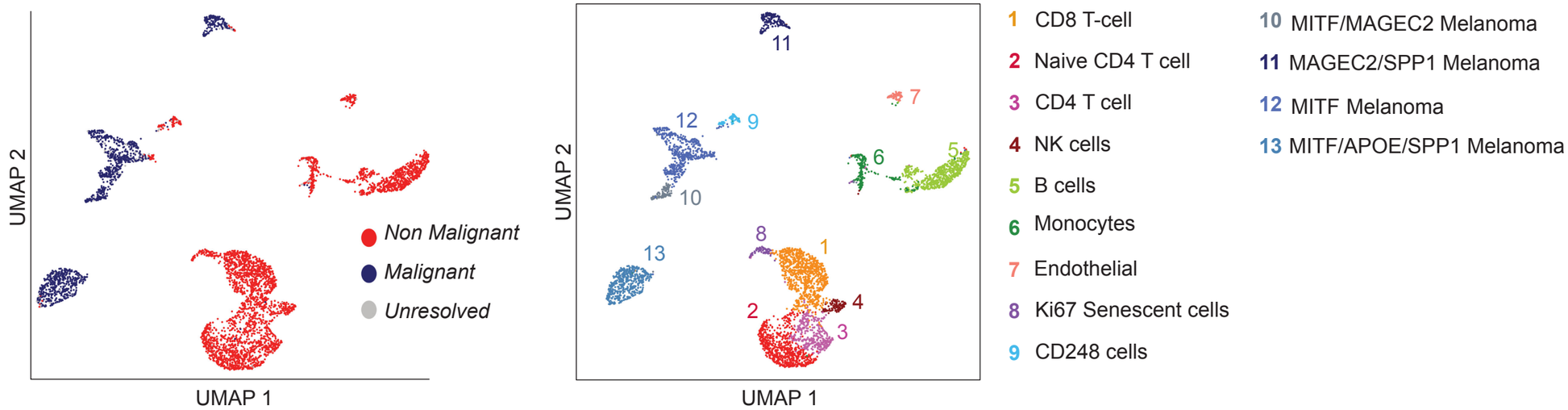
A


Ctrl
KD1-Rec
KD2-Rec
KD3-Rec

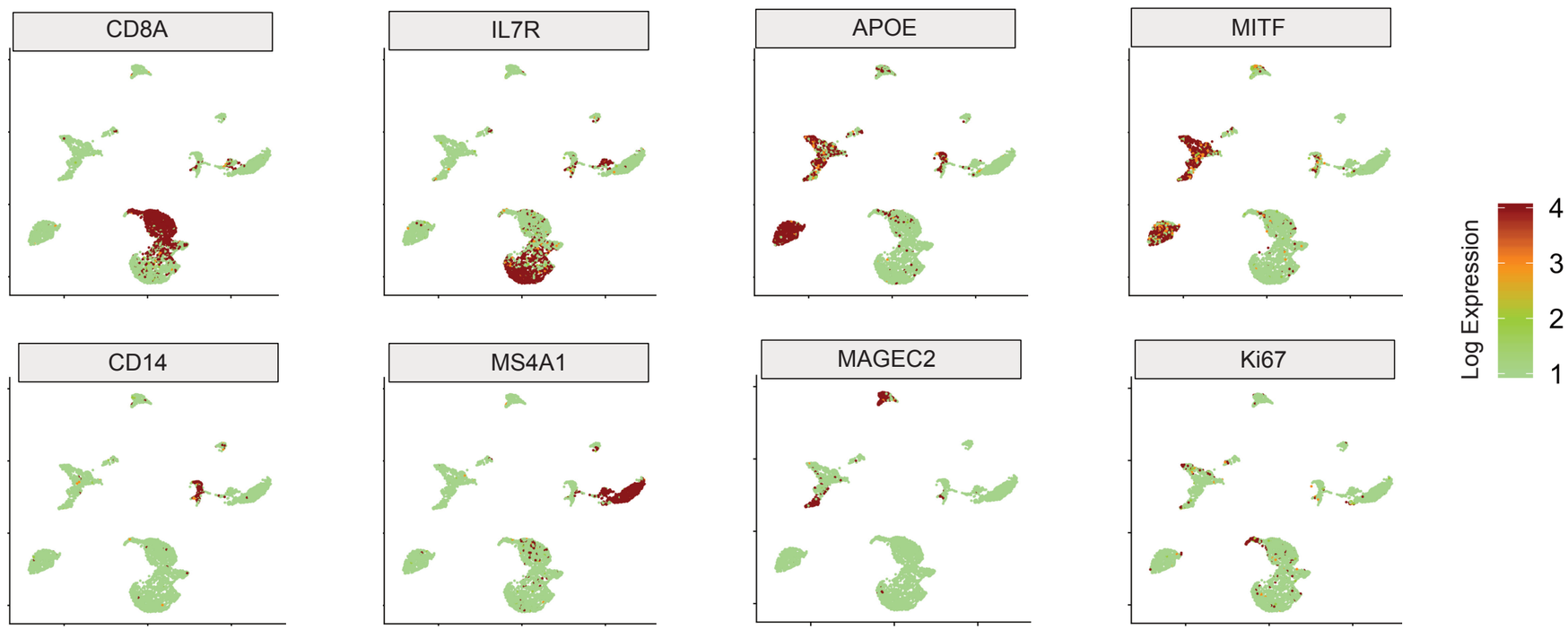
B


****
****
**
1000
500
0
Number of Invaded cells
Control
KD1-Rec
KD2-Rec
KD3-Rec

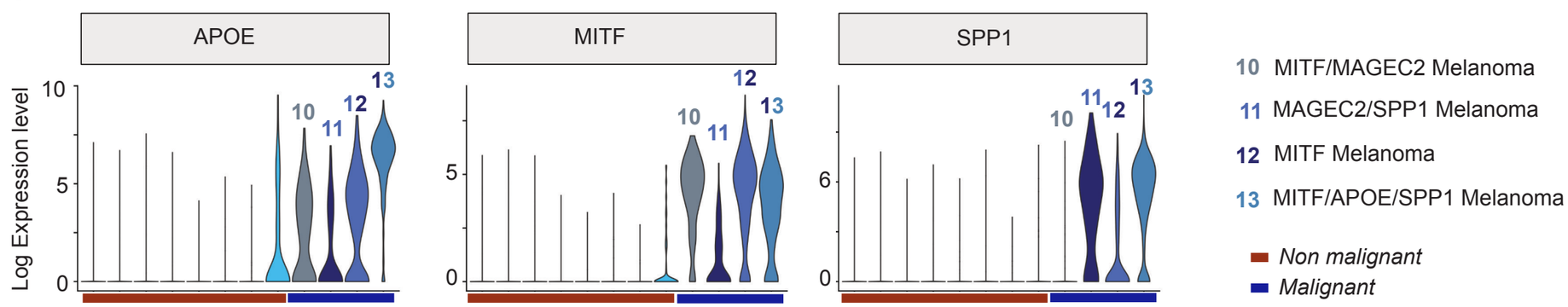
A

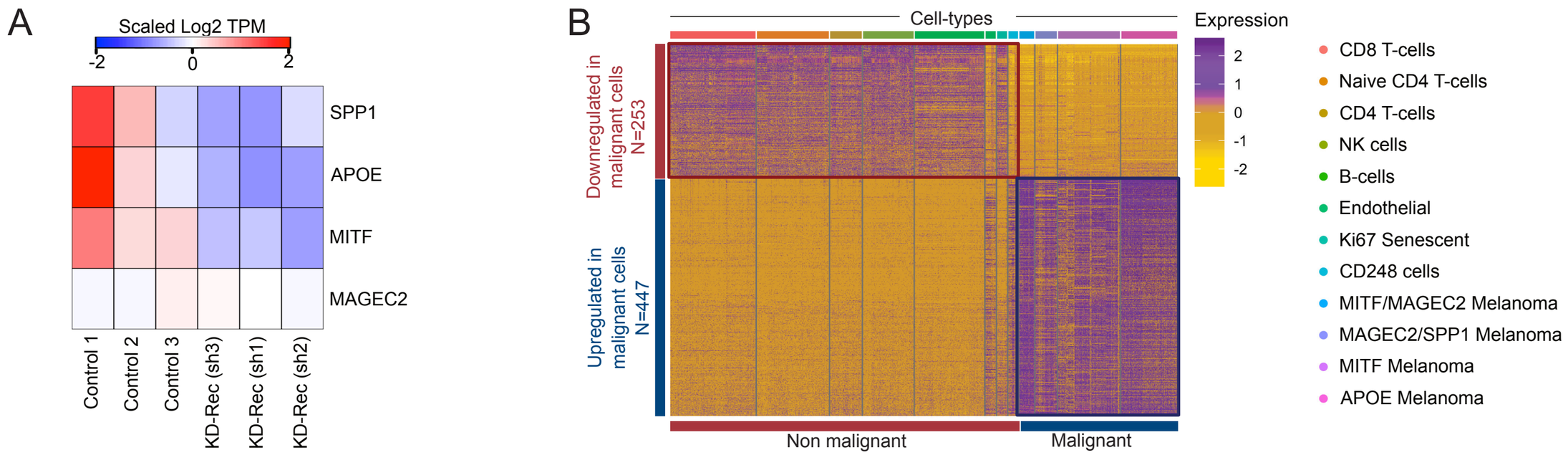

UMAP 2
UMAP 1
Non Malignant
Malignant
Unresolved
UMAP 2
UMAP 1
11
12
9
7
6
5
10
8
1
13
4
2
3
1 CD8 T-cell
2 Naive CD4 T cell
3 CD4 T cell
4 NK cells
5 B cells
6 Monocytes
7 Endothelial
8 Ki67 Senescent cells
9 CD248 cells
10 MITF/MAGEC2 Melanoma
11 MAGEC2/SPP1 Melanoma
12 MITF Melanoma
13 MITF/APOE/SPP1 Melanoma

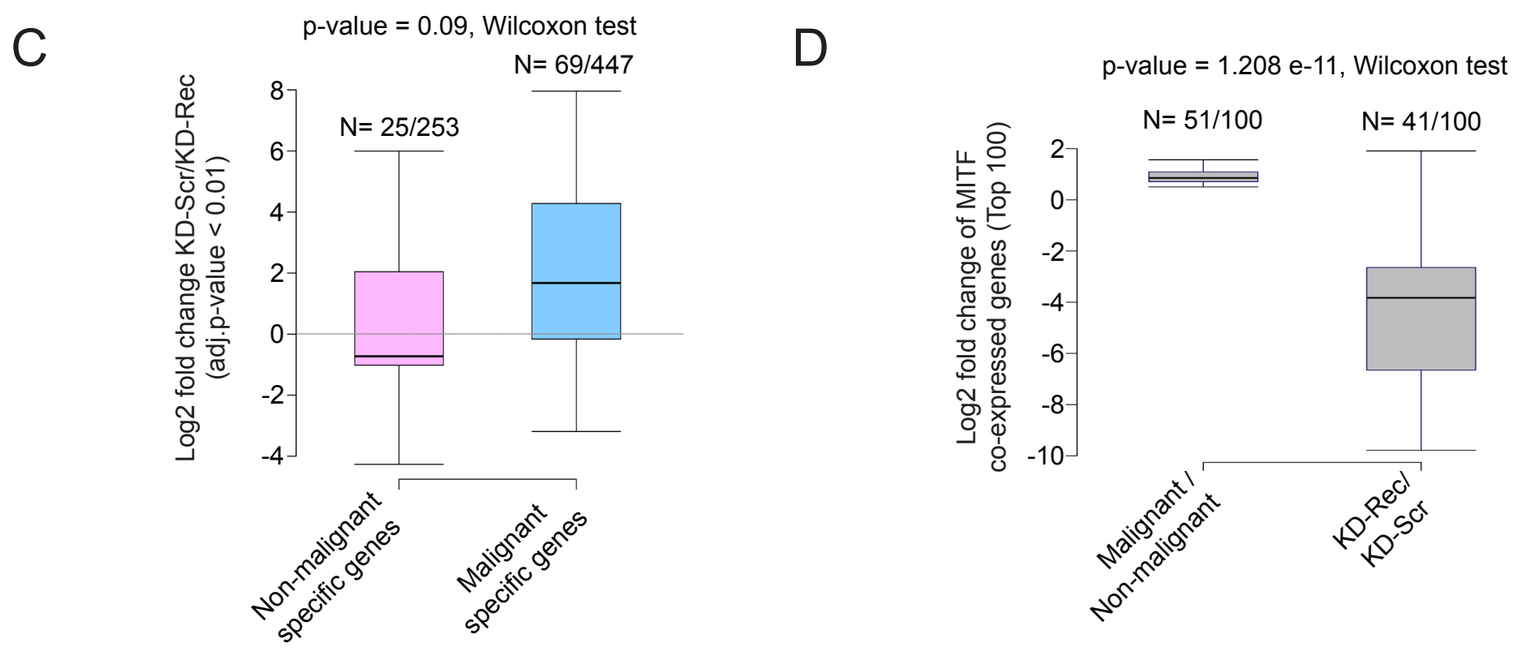
B


CD8A
IL7R
APOE
MITF
CD14
MS4A1
MAGEC2
Ki67
Log Expression
4
3
2
1

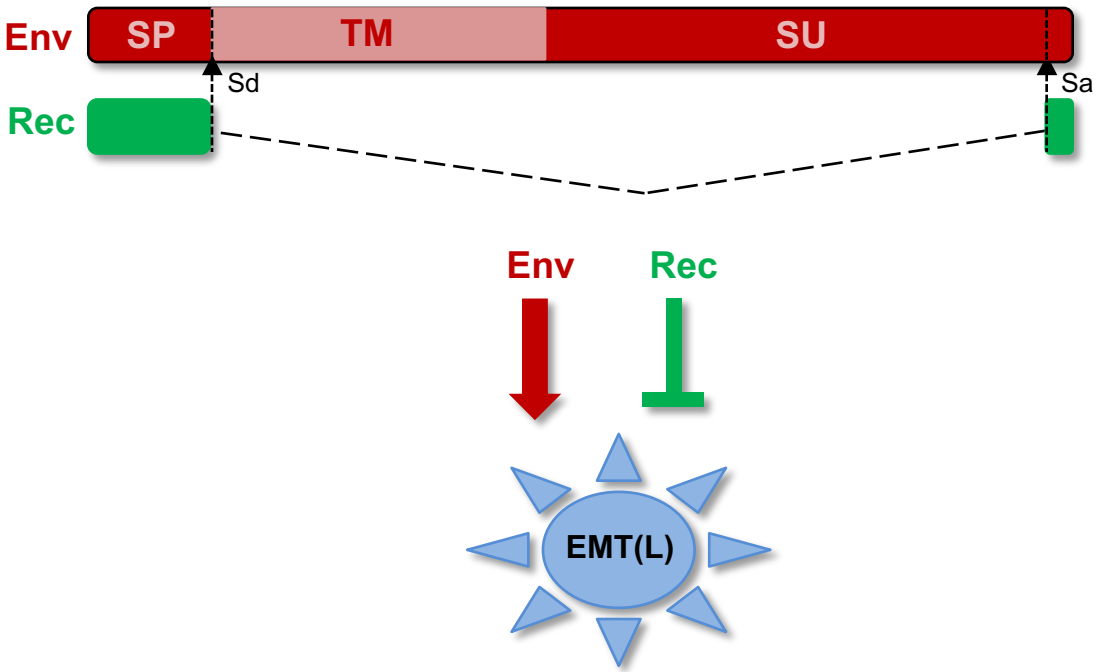
C


APOE
MITF
SPP1
Log Expression level
10
5
0
5
0
6
0
10
11
12
13
10 MITF/MAGEC2 Melanoma
11 MAGEC2/SPP1 Melanoma
12 MITF Melanoma
13 MITF/APOE/SPP1 Melanoma
Non malignant
Malignant

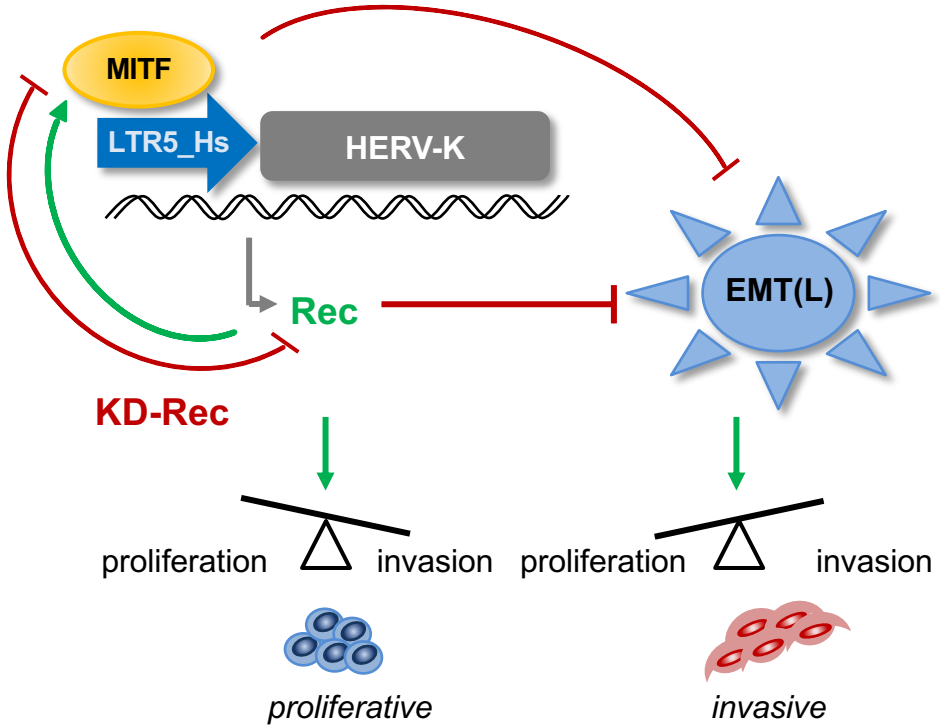

A
Scaled Log2 TPM
-2
0
2
SPP1
APOE
MITF
MAGEC2
Control 1
Control 2
Control 3
KD-Rec (sh3)
KD-Rec (sh1)
KD-Rec (sh2)
B
Cell-types
Expression
2
1
0
-1
-2
Downregulated in malignant cells N=253
Upregulated in malignant cells N=447
Non malignant
Malignant
CD8 T-cells
Naive CD4 T-cells
CD4 T-cells
NK cells
B-cells
Endothelial
Ki67 Senescent
CD248 cells
MITF/MAGEC2 Melanoma
MAGEC2/SPP1 Melanoma
MITF Melanoma
APOE Melanoma

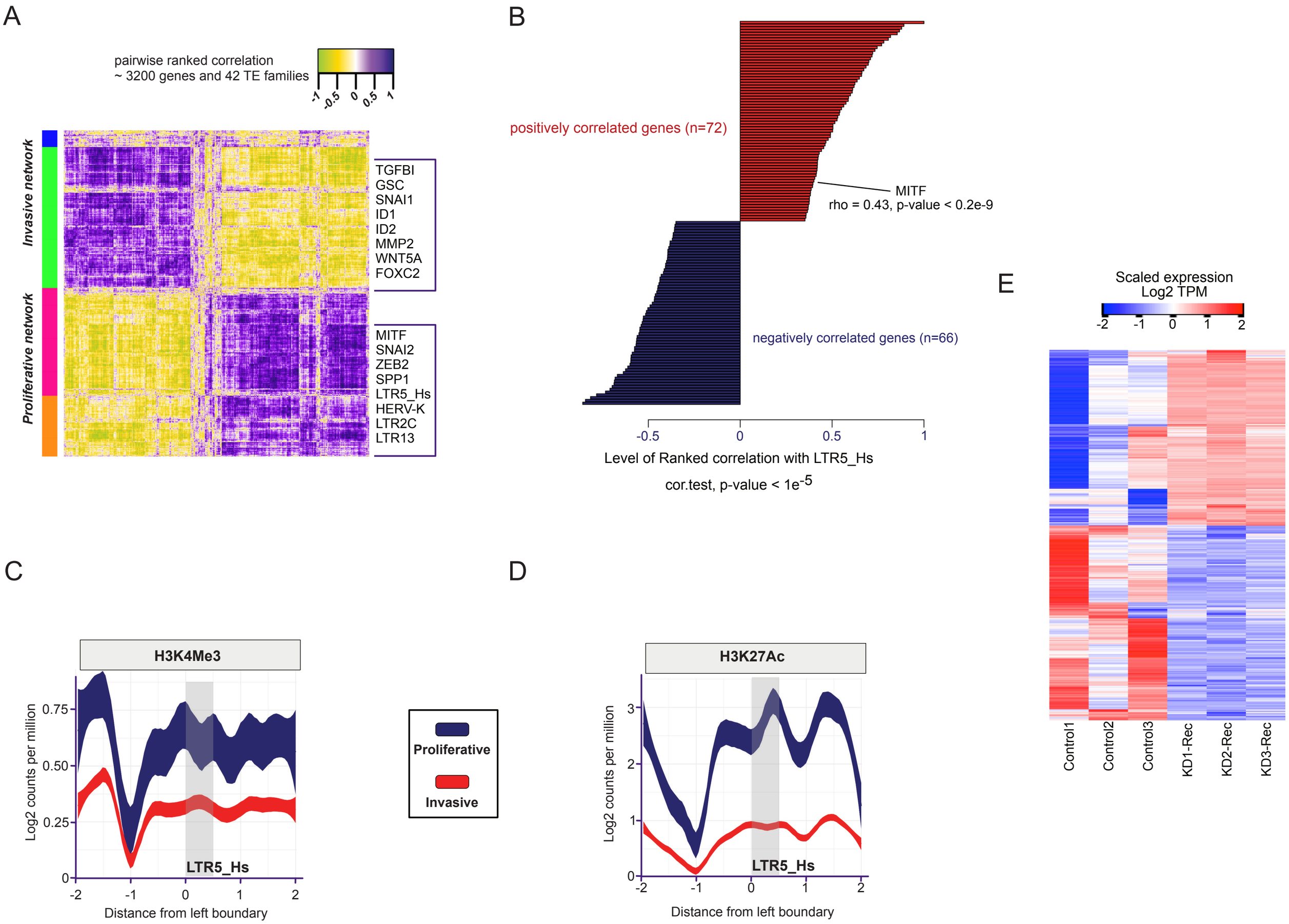

C
p-value = 0.09, Wilcoxon test
N= 25/253
N= 69/447
Log2 fold change KD-Scr/KD-Rec (adj.p-value < 0.01)
8
6
4
2
0
-2
-4
Non-malignant specific genes
Malignant specific genes
D
p-value = 1.208 e-11, Wilcoxon test
N= 51/100
N= 41/100
Log2 fold change of MITF co-expressed genes (Top 100)
2
0
-2
-4
-6
-8
-10
Malignant / Non-malignant
KD-Rec/ KD-Scr

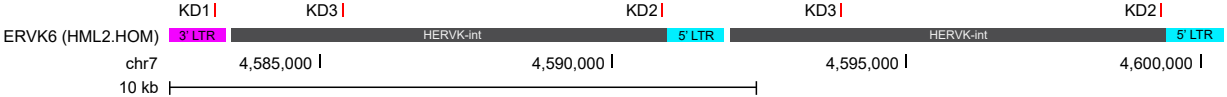
A



B


MITF
LTR5_Hs
HERV-K
Rec
KD-Rec
EMT(L)
proliferation
invasion
proliferation
invasion
proliferative
invasive


A
pairwise ranked correlation
~ 3200 genes and 42 TE families
-1
-0.5
0
0.5
1
Invasive network
Proliferative network
TGFBI
GSC
SNAI1
ID1
ID2
MMP2
WNT5A
FOXC2
MITF
SNAI2
ZEB2
SPP1
LTR5_Hs
HERV-K
LTR2C
LTR13
B
positively correlated genes (n=72)
MITF
rho = 0.43, p-value < 0.2e-9
negatively correlated genes (n=66)
-0.5
0
0.5
1
Level of Ranked correlation with LTR5_Hs
cor.test, p-value < 1e-5
E
Scaled expression
Log2 TPM
-2
-1
0
1
2
Control1
Control2
Control3
KD1-Rec
KD2-Rec
KD3-Rec
C
H3K4Me3
Log2 counts per million
0.75
0.50
0.25
0
LTR5_Hs
-2
-1
0
1
2
Distance from left boundary
Proliferative
Invasive
D
H3K27Ac
Log2 counts per million
3
2
1
0
LTR5_Hs
-2
-1
0
1
2
Distance from left boundary


KD1
KD3
KD2
KD3
KD2
ERVK6 (HML2.HOM)
3' LTR
HERVK-int
5' LTR
HERVK-int
5' LTR
chr7
4,585,000
4,590,000
4,595,000
4,600,000
10 kb