RESEARCH

Binding of NF-Y to transposable elements in mouse and human cells

Mirko Ronzio[1], Andrea Bernardini[1], Alberto Gallo[1], Roberto Mantovani[1] and Diletta Dolfini[1*]

Abstract

Background Transposable Elements (TEs) represent a sizeable amount of mammalian genomes, providing regulatory sequences involved in shaping gene expression patterns. NF-Y is a Transcription factor -TF- trimer that binds to the CCAAT box, belonging to a selected group implicated in determining initiation of coding and noncoding RNAs.

Results We focus on NF-Y TE locations in 8 human and 8 mouse cells. Binding is exclusive for retroviral LTR12, MLT1 and MER in human and RLTR10 and IAPLTR in mouse cells. Cobinding and analysis of the DNA matrices signal enrichment of distinct TFs neighboring CCAAT in the three TE classes: MAFK/F/G in LTR12 and USF1/2 in MLT1 with precise alignment of sites, PKNOX1, MEIS2, PBX2/3 TALE TFs in MER57. The presence of "epigenetic" marks in human cells indicate prevalent co-association with open chromatin in MER, closed in LTR12 and mixed in MLT1. Based on chromatin features, these locations are mostly marked as enhancers, as confirmed by analysis of *loci* predicted to generate eRNAs.

Conclusions These results are discussed in the context of functional data, suggesting a complex -positive and potentially-negative role of NF-Y on distinct classes of repetitive sequences.

Keywords Transcription factor, NF-Y, Repetitive sequences, LTR12, MLT1, MER

Background

A large portion of the human genome is made of different classes of repetitive DNA, which include Transposable Elements (TEs). A sizeable amount -some 8% in humans- is constituted by TEs resulting from insertions of RNA viruses and retro-transpositions. They are typically species-specific and many are used as cis-acting elements -CREs- that regulate transcription as enhancers and promoters [1–5]. These TEs contain individual short DNA elements recognized by sequence-specific TFs governing the recruitment of the RNA Polymerase II machinery and of cofactors, often empowered with chromatin modifying enzymatic activities. Distinct TF Binding Sites -TFBS- are associated to individual classes of repetitive sequences in vivo [6–10].

The CCAAT box is one of the first DNA element described in mammalian promoters, whose precise matrix was formalized by unbiased searches and confirmed by functional studies of RefSeq genes (Reviewed by [11]). Recently, this element was identified in studies aimed at finding sequence determinants driving selection of Transcription Start Site(s) (TSS) in coding, non-coding genes and units of enhancer RNAs (eRNAs) [12–16]. The matrix identified by these studies matches precisely the binding site of NF-Y, a trimer composed of the histone-like NF-YB/NF-YC dimer and NF-YA, conferring strict sequence-specificity [17]. NF-Y/CCAAT is important for

*Correspondence:
Diletta Dolfini
diletta.dolfini@unimi.it
[1]Dipartimento di Bioscienze, Università degli Studi di Milano, Milano, Italy

promoter function, as assessed by mutation of CCAAT in functional experiments, the use of an NF-YA dominant negative mutant or RNAi of the subunits (Reviewed by [18]). NF-Y was shown to maintain the borders of the core TSS region free of nucleosomes, and its removal causes relocation of TSS and appearance of extended, aberrantly initiated transcripts [19]. These data implicate the trimer in shaping the chromatin and transcriptional architecture of core promoters, including determining TSS selection by the General Transcription Factors and RNA Pol II.

Following inclusion of NF-Y in ENCODE, ChIP-seq location analysis in human K562, HeLa-S3 and GM12878 cell lines identified TEs, notably families of retroviral origin [7, 10, 20, 21]. In humans, ATAC-seq analysis found enrichment of NF-Y sites in Spermatogonial Stem Cells (SSCs), absent in differentiating cKIT+ spermatogonia, specifically at LTR12C/D/E [22]; this was confirmed by scRNA-seq analysis of stem and differentiated populations [23, 24]. As for mouse cells, CCAAT boxes were reported in RLTR10, TEs associated to genes activated during spermatogenesis [25]. Mouse embryo primordial germ cells -PGCs- gain CCAAT accessibility in active regions between d13.5 and d14.5, in enhancers of genes repressed by H3K27me3; these units are postnatally activated and remain expressed as hallmarks of adult SSCs [26].

Knowledge on the role of CCAAT in specific TEs is limited to sites located in enhancers: the ERV-9/LTR12 of the globin locus control region mediates long range interactions with the developmentally regulated ε-, γ- and β-globin CCAAT promoters [27–29] and an LTR12 enhancer requires NF-Y to drive RAE1 expression [30]. More recent experiments highlighted a widespread function of NF-Y on LTRs. The use of "epigenetic" drugs blocking the enzymatic activity of DNA Methyltransferases -DNMTs- and Histone Deacetylates -HDACs- entailed induction of cryptic RNAs (TINATs, Treatment Induced Non-Annotated Transcripts): more than 80% are generated by LTR12 (notably LTR12C), downstream of conserved NF-Y and Sp1 sites [31]. This study formalized and extended reports focusing on HDACi, signaling widespread activation of LTR12-driven promoters [32], including in a NF-Y-dependent way [33]. Additional reports on DNMTi and HDACi confirmed this point [34–38], including a study based on RNA-mediated Cas9 activation domains specifically directed to LTR12C sites [39]. LTR12-driven transcripts from enhancers were identified upon inactivation of BRD4, a cofactor "reader" of acetylated histones: NF-YA/NF-YB inactivation by RNAi confirmed functional impairment of LTR12C/D enhancers [40]. As for mouse, in embryonic stem cells CCAAT-driven RNAs generated from IAPLTR TEs are part of condensates that squelch the basal transcriptional machinery from enhancers involved in driving expression of stemness and fate determination genes [41]. Unlike humans, analysis of genomic binding of NF-Y to TEs in mouse cells has yet to be reported.

We recently extended the analysis of NF-Y in vivo locations across 8 human and 8 mouse cell types, by exploiting available datasets and performing additional ChIP-seq experiments with a common pipeline of analysis [42]. In parallel, we inactivated NF-YB in HeLa cells focusing on genomic binding of USF1, a NF-Y partner: removal of NF-Y from promoters entailed elimination of USF1 binding and decreased function. The growing interest on repetitive elements of retroviral origin as functional promoters/enhancers, as well as the activities of NF-Y in determining positioning of RNAs initiation, spurred us to provide a detailed analysis of NF-Y binding to TEs in human and mouse cells.

Results

NF-Y binding to repetitive sequences

We recently evaluated NF-Y peaks in 16 cell lines and tissues of human and mouse origin: in addition to ENCODE cell lines, we considered data from mouse cells, performing further ChIP-Seq experiments in several mouse and human cells. To standardize the data, we used the ENCODE pipeline, adjusted for different lengths of the reads gathered from the ChIP-seq datasets [42], integrated by further analysis with Repbase, as indicated in Fig. S1. We focus here on peaks located on TEs. A graphic representation that highlights the prevalence in specific repetitive sequences in terms of overall number (Fig. 1A, B) or percentages with respect to the total number of peaks is shown (Fig. 1C, D). The peaks are mainly in LTR families (Dark blue) and the number appears to be higher in human cells, especially in transformed cells. For comparison, we analyzed TBP -a General Transcription Factor binding to the TATA element- and USF1, a b-HLH TF binding to the E-box. TBP shows less skewing toward repeats, USF1 confirms enrichment in LTR repeats, vast in HepG2 and K562, less in mouse MEL and C2C12 myocytes. NF-Y peaks in repetitive DNA range from 9% (B cells) to 55% (K562) of the total (Fig. 1E). In comparison with previous analysis of ENCODE lines [20], the fraction is identical in GM12878, somewhat higher in HeLa-S3 (34% vs. 23%) and K562 (55% vs. 40%). In mouse cells (Fig. 1F), a maximum is scored in mESCs (14%), a minimum in fetal myoblasts/myocytes (4%). USF1 is in the same range in human cells, higher in mouse, TBP is at < 20% (Fig. 1E, F). Overall, these results confirm and extend previous analysis: NF-Y can be placed at the high end among TFs binding to TEs in human cells [9].

To verify the above results, we also analyzed the data according to a modification of the protocol we previously employed [20], which consists of mapping reads on
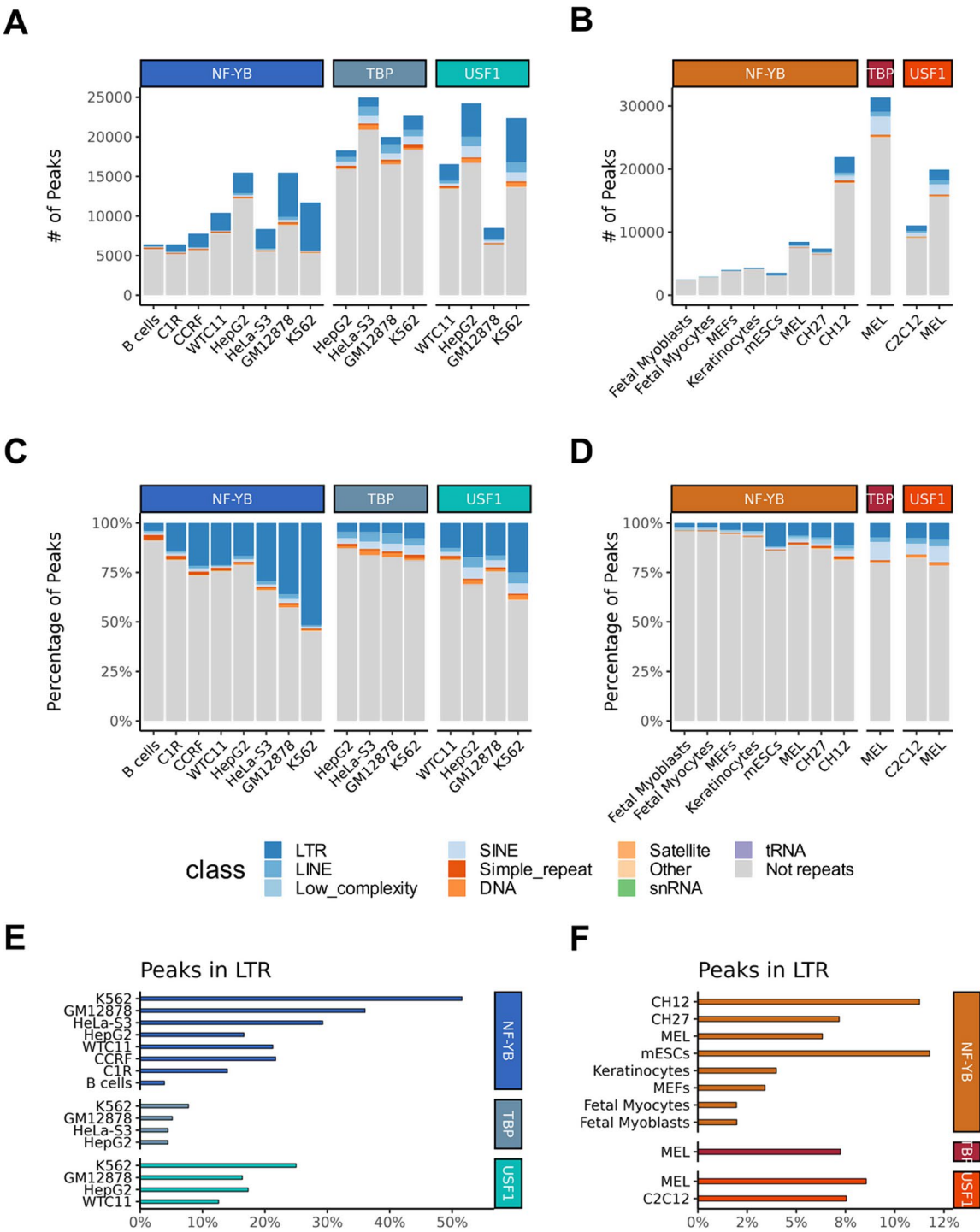
Fig. 1 NF-Y peaks in repetitive elements. **A**, **B**. Absolute distribution of human (**A**) and mouse (**B**) cells ChIP-seq peaks between non-repetitive genome (gray) and repetitive regions as annotated in RepeatMasker (colored). **C**, **D**. Same as **A** except that the data are shown in percentages over the total. **E**, **F**. Percentage of total peaks specifically falling in LTR sequences in human (**E**) and mouse (**F**) cells

Repbase repeat consensus sequences and then evaluating their enrichment over Input DNAs. The results confirm the LTR enrichment, allowing us to identify LTR12, MLT1 and MER families in human cells, and RLTR10 and IAPLTR1 in mouse (Fig. 2A and B). In Fig. 2C and

D, we considered examples of signal on individual repetitive elements consensus of Immunoprecipitated DNA (in red) compared to Input DNA (Blue) -and the relative fold change ratio (Log2FC, in yellow)- in human K562 and mouse CH12: LTR12C, MER122, RLTR10B2

A

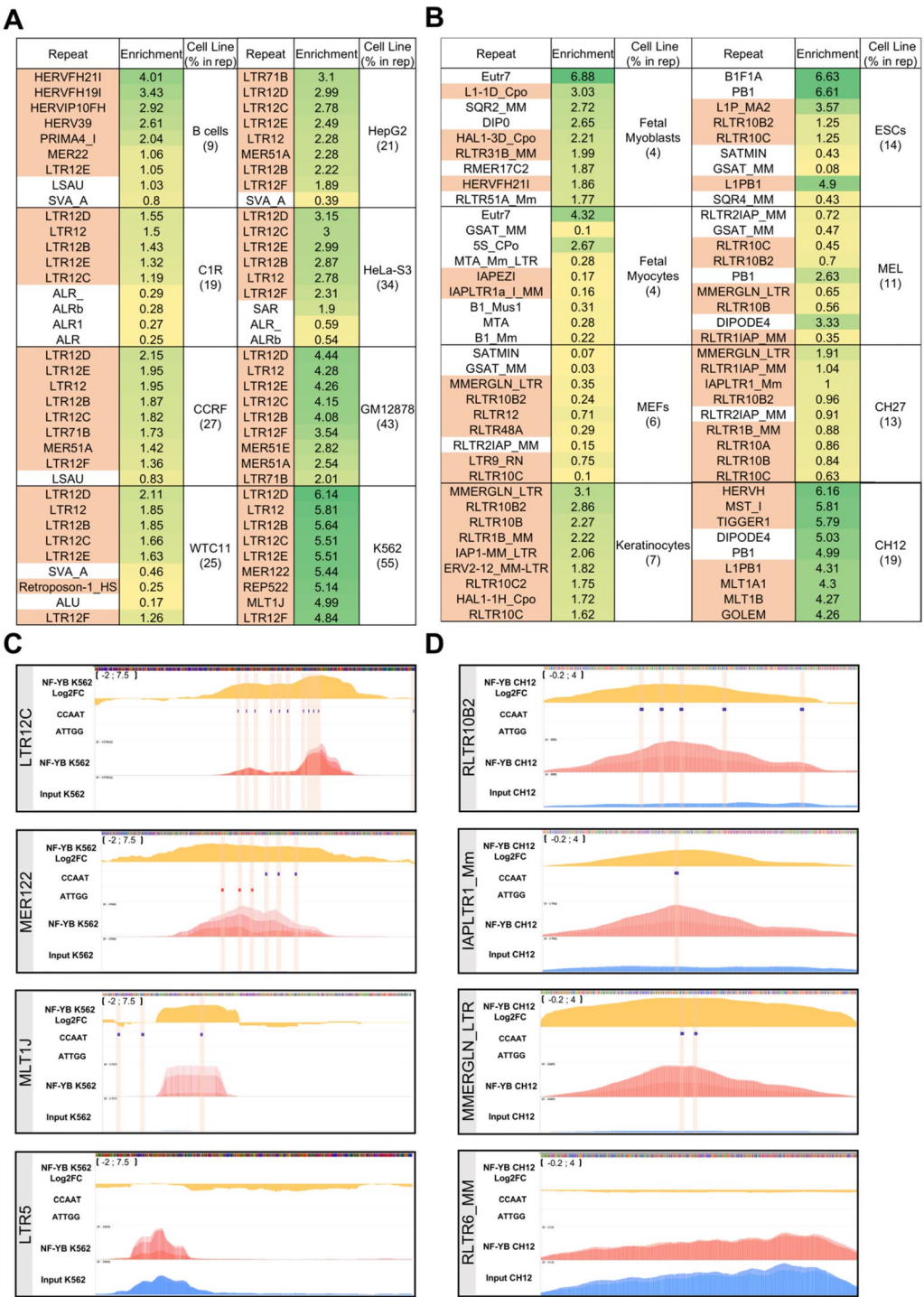| Repeat	Enrichment	Cell Line (% in rep)	Repeat	Enrichment	Cell Line (% in rep)
HERVFH21I	4.01		LTR71B	3.1	
HERVFH19I	3.43		LTR12D	2.99	
HERVIP10FH	2.92		LTR12C	2.78	
HERV39	2.61	B cells (9)	LTR12E	2.49	HepG2 (21)
PRIMA4_I	2.04		LTR12	2.28	
MER22	1.06		MER51A	2.28	
LTR12E	1.05		LTR12B	2.22	
LSAU	1.03		LTR12F	1.89	
SVA_A	0.8		SVA_A	0.39	
LTR12D	1.55		LTR12D	3.15	
LTR12	1.5		LTR12C	3	
LTR12B	1.43		LTR12E	2.99	
LTR12E	1.32	C1R (19)	LTR12B	2.87	HeLa-S3 (34)
LTR12C	1.19		LTR12	2.78	
ALR_	0.29		LTR12F	2.31	
ALRb	0.28		SAR	1.9	
ALR1	0.27		ALR_	0.59	
ALR	0.25		ALRb	0.54	
LTR12D	2.15		LTR12D	4.44	
LTR12E	1.95		LTR12	4.28	
LTR12	1.95		LTR12E	4.26	
LTR12B	1.87	CCRF (27)	LTR12C	4.15	GM12878 (43)
LTR12C	1.82		LTR12B	4.08	
LTR71B	1.73		LTR12F	3.54	
MER51A	1.42		MER51E	2.82	
LTR12F	1.36		MER51A	2.54	
LSAU	0.83		LTR71B	2.01	
LTR12D	2.11		LTR12D	6.14	
LTR12	1.85		LTR12	5.81	
LTR12B	1.85		LTR12B	5.64	
LTR12C	1.66	WTC11 (25)	LTR12C	5.51	K562 (55)
LTR12E	1.63		LTR12E	5.51	
SVA_A	0.46		MER122	5.44	
Retroposon-1_HS	0.25		REP522	5.14	
ALU	0.17		MLT1J	4.99	
LTR12F	1.26		LTR12F	4.84	

B

Repeat	Enrichment	Cell Line (% in rep)	Repeat	Enrichment	Cell Line (% in rep)
Eutr7	6.88		B1F1A	6.63	
L1-1D_Cpo	3.03		PB1	6.61	
SQR2_MM	2.72		L1P_MA2	3.57	
DIP0	2.65	Fetal Myoblasts (4)	RLTR10B2	1.25	ESCs (14)
HAL1-3D_Cpo	2.21		RLTR10C	1.25	
RLTR31B_MM	1.99		SATMIN	0.43	
RMER17C2	1.87		GSAT_MM	0.08	
HERVFH21I	1.86		L1PB1	4.9	
RLTR51A_Mm	1.77		SQR4_MM	0.43	
Eutr7	4.32		RLTR2IAP_MM	0.72	
GSAT_MM	0.1		GSAT_MM	0.47	
5S_CPo	2.67		RLTR10C	0.45	
MTA_Mm_LTR	0.28	Fetal Myocytes (4)	RLTR10B2	0.7	MEL (11)
IAPEZI	0.17		PB1	2.63	
IAPLTR1a_I_MM	0.16		MMERGLN_LTR	0.65	
B1_Mus1	0.31		RLTR10B	0.56	
MTA	0.28		DIPODE4	3.33	
B1_Mm	0.22		RLTR1IAP_MM	0.35	
SATMIN	0.07		MMERGLN_LTR	1.91	
GSAT_MM	0.03		RLTR1IAP_MM	1.04	
MMERGLN_LTR	0.35		IAPLTR1_Mm	1	
RLTR10B2	0.24	MEFs (6)	RLTR10B2	0.96	CH27 (13)
RLTR12	0.71		RLTR2IAP_MM	0.91	
RLTR48A	0.29		RLTR1B_MM	0.88	
RLTR2IAP_MM	0.15		RLTR10A	0.86	
LTR9_RN	0.75		RLTR10B	0.84	
RLTR10C	0.1		RLTR10C	0.63	
MMERGLN_LTR	3.1		HERVH	6.16	
RLTR10B2	2.86		MST_I	5.81	
RLTR10B	2.27		TIGGER1	5.79	
RLTR1B_MM	2.22	Keratinocytes (7)	DIPODE4	5.03	CH12 (19)
IAP1-MM_LTR	2.06		PB1	4.99	
ERV2-12_MM-LTR	1.82		L1PB1	4.31	
RLTR10C2	1.75		MLT1A1	4.3	
HAL1-1H_Cpo	1.72		MLT1B	4.27	
RLTR10C	1.62		GOLEM	4.26	

C

D

Fig. 2 Repbase alignment enrichment of IP NF-YB experiments. **A.** Top 9 significant enriched TE repeats for each human cell line in the Repbase alignment according to the statistical significance (Poisson). CCAAT-containing elements are highlighted in orange. **B.** same as A, for mouse cell lines. **C.** Representative NF-YB signal enrichment in four human repeat consensus *loci*. Log2FC of NF-YB signal in K562 cell lines over the input depicted in yellow. Overlay of the three biological replicates are shown in red hues and overlay of input replicates in blue. **D.** same as C for NF-YB in mouse

and MMERGLN, which have multiple CCAAT, have a high Log2FC, lower in MLT1J and IAPLTR1 with single or double CCAAT. As controls, the CCAAT-less LTR5 and RLTR6 show bound signals similar to those of background Input DNA. We take this as a further indication that specific TEs of retroviral origin are selectively bound by NF-Y in human and mouse cells.

Factors in NF-Y-bound repetitive sequences

We analyzed NF-Y locations in the different subfamilies of TEs in human cells. LTR12 are ubiquitously bound by NF-Y -except in B cells- whereas MLT1 are enriched in K562, GM12878 and, to a lesser extent, HepG2 and WTC11 (Fig. 3A, B, upper Panels). MER are either ubiquitously enriched -MER51A/E, MER101, MER57C1 with the exception of B cells- or tissue-specific: MER57A1/F/ E1 in GM12878, MER57B1 in GM12878 and K562, MER52C in HepG2, WTC11 and B cells (Fig. 3C, Upper Panel). Next, we assessed the quality -adherence to the consensus- and number of CCAAT boxes in these TEs: they are optimal and numerous (>1/100 bp; quality

scores close to 1) in LTR12, MERs, and MLT1K, but not in other MLT1 (Fig. 3A-C, middle Panels). Note that MLT1H2 has excellent CCAAT scores, yet less than 1% are bound by NF-Y, suggesting that there might be additional features on TEs that affect binding, such as the "epigenomic" environment and the presence of additional TFs bound nearby.

Binding to TEs is a common feature for TFs. We previously computed co-association of NF-Y genomic locations with those of the TFs and Cofactors present in ENCODE [20, 42–44]. We then investigated the subclasses of TEs matching binding of NF-Y with ENCODE TFs and cofactors. The heatmaps show degrees of association distinct for the three families (Fig. 3A-C, Lower Panels). In LTR12, notably in LTR12B/D/F, we find small MAFs -K, F, G- ZNF316, ZBTB33, MBD2 and C11Orf30/ EMSY (Fig. 3A): all factors are associated to repressive activities, in line with recent findings [21]. In MLT1, b-HLH TFs -USF1/2, MITF, MAX and TFE3- are prevalent, while other factors are subfamily-specific (Fig. 3B). Note that the global fraction of MLT1s bound by NF-Y
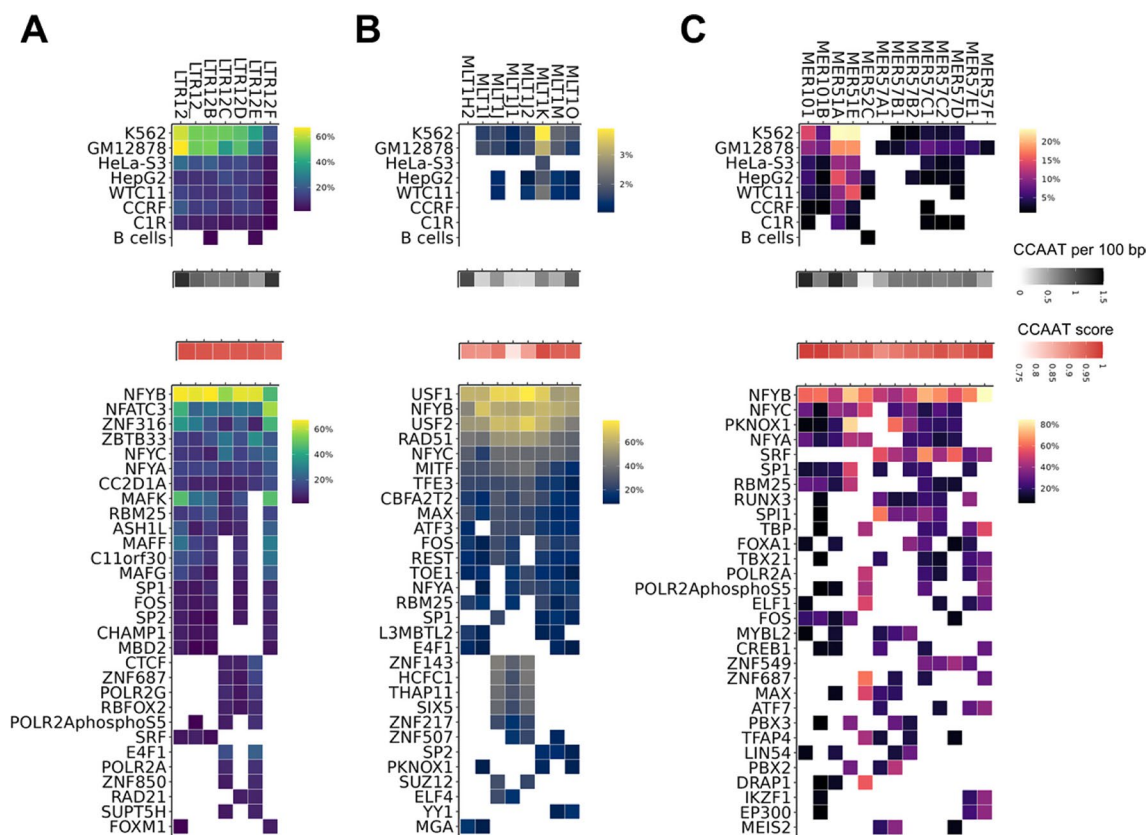

Fig. 3 Cobinding of TFs on LTR12, MLT1 and MER. **A.** Heatmap showing the percentage (if higher than 1%) of LTR12 genomic sites bound by NF-YB in human cell lines. Middle panels show the number (top) and the quality score (down) of CCAAT boxes present in the corresponding consensus. If more than one CCAAT box is present in the consensus, the score refers to the best one. Lower Panel shows the percentage of LTR12 genomic sites cobound by NF-Y and TFs present in ENCODE calculated over the total LTR sites bound by NF-YB. If more than one experiment is present in one or more cell lines, the maximum value is plotted. Top 30 TFs showing binding to more than 2% sites and for more than one repeat member are shown. **B**. same as A, for enriched MLT1 repetitive elements. **C**. Same as A, for enriched MERs repetitive elements

-and other factors- is considerably lower (2/6%) than the 10/60% scored for LTR12. MERs have an intermediate level of TFs bound (5/20%) and NF-Y is with the TALE PKNOX1/PBX2/PBX3/MEIS2 and SP1 (Fig. 3C); Zn Fingers TFs are specific for MER57 (ZNF549) or MER52 (ZNF687). We conclude that NF-Y co-binds with selected companions in the three TE families.

Modules in NF-Y-bound repetitive sequences
We retrieved LTR12, MLT1 and MER sequences bound by NF-Y and some of the TFs identified above, notably small MAFs for LTR12, USF1/2 for MLT1 and TALE for

MER57B1. The sites of these TFs were thereafter aligned by using the CCAAT box as anchor, with surroundings of 100 bp. The PKNOX1/PBX2/PBX3 sites were previously shown to be aligned according to a precise geometry – 5'TALE-10 bp-CCAAT3'- in promoters and enhancers [45]: Fig. 4 shows that the TALE site is indeed present, but located at the 3' and relatively distant from CCAAT. Between CCAAT and TALE, MEME analysis found a sequence corresponding to a STAT1/2 site (p value 10^{-99}). As for LTRs, representative results are shown for LTR12D in Fig. S2: MAREs (MAFs-Responsive Elements) are found, with a precise positioning immediately
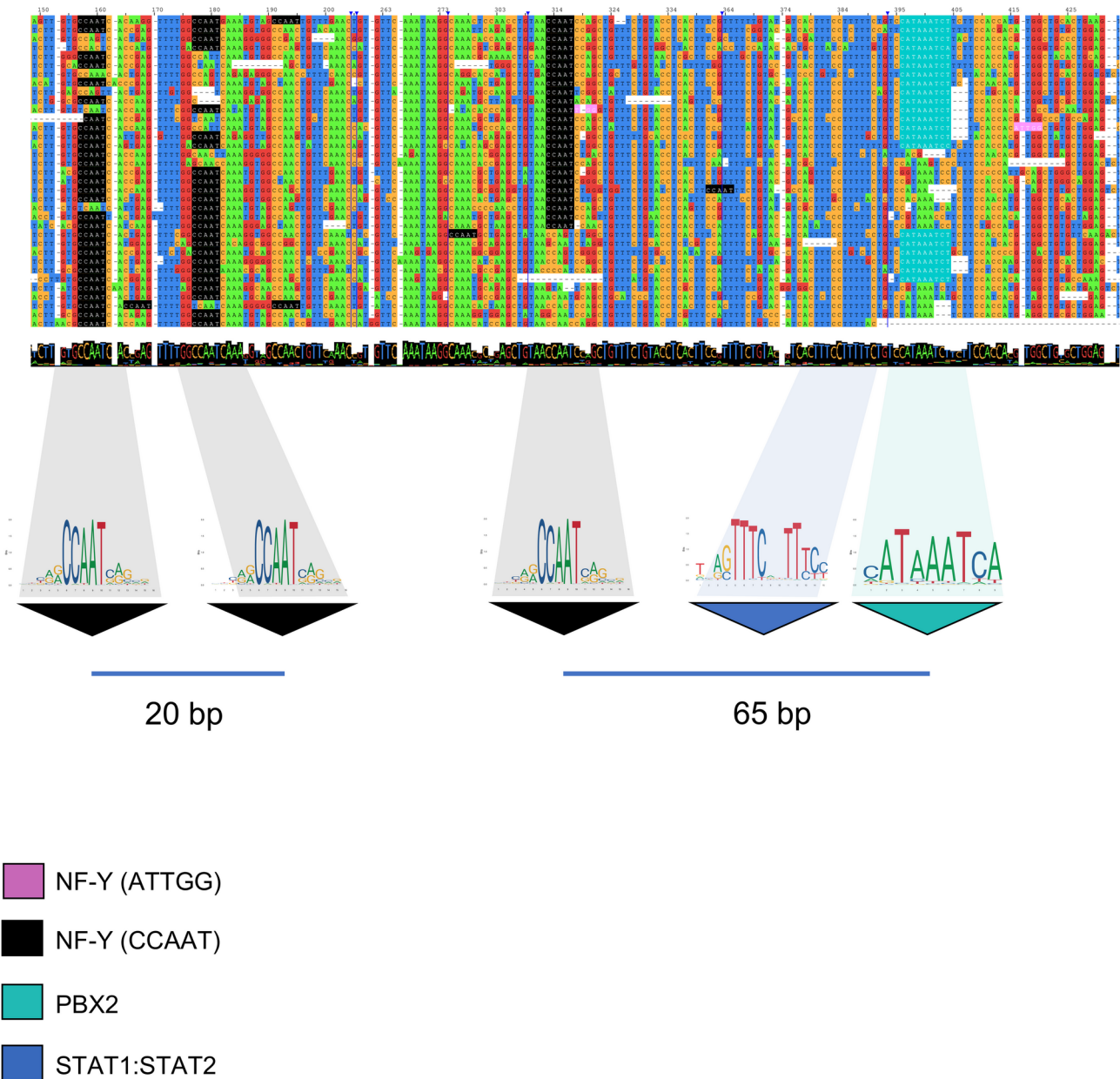
Fig. 4 Alignment of MER57B1 sequences around CCAAT boxes. Alignment of MER57B1 sequencing focusing on CCAAT boxes. Sequences retrieved from genomic sites co-bound by NF-Y and PKNOX/PBX2/PBX3. PWMs derived from JASPAR 2024 Redundant database

3' of CCAAT: indeed, the first nucleotide of MARE -a T- is shared by CCAA*T* and the CAG core immediately thereafter represents an optimal 3' bp for NF-Y binding. This configuration is striking because it is found at least 3/4 times within each repeat. Finally, we considered the large overlap between NF-Y and USF1/2 in MLT1K: the distance of 16/18 bp between the center of the two sites -E box and CCAAT- corresponds to the expected E-box-10 bp-CCAAT configuration (Fig. S3), based on previous biochemical experiments [46]. In summary, the TFs identified by genomic analysis appear to be bound to their respective sites in the different TEs.

Chromatin in NF-Y *loci*
Another relevant feature that might impact on NF-Y binding to TEs is the chromatin *status*. ENCODE provides data on several epigenomic marks in the cell lines analyzed above, allowing distinction of 18 chromatin states: it is important to note that this analysis does not match annotated RNAs, and therefore the indicated "TSS" or "Enhancer" labels are mere extrapolations from the presence of epigenomic marks. Analysis of the three families bound by NF-Y is shown in Fig. 5. Most LTR12 repeats are marked as heterochromatin (Light blue) or quiescent/low (Light grey) in K562 and GM12878; HeLa-S3 and HepG2 have a higher fraction of *loci* classified as weak transcription (Dark green) or enhancer (Yellow) (Fig. 5 Left Panel). The MLT1 profile is similar in K562, GM12878 and HepG2, only with more active promoters and enhancers, specifically MLT1M in HepG2 (Fig. 5 Middle Panel). As for MERs, many sites of HepG2 and GM12878 are active promoters or enhancers, especially in MER57; they are less "active" in HeLa-S3 and K562, but still comparatively more active than LTR12 or MLT1 (Fig. 5 Right Panel).

TEs in promoters and enhancers
TEs are spread throughout genomes and it is relevant to establish the locations of those bound by NF-Y: a first classification of genomic annotations in human and mouse cells is shown in Fig. S4: there are several in Promoters, but the vast majority are in Intergenic or Intronic regions: this is in contrast to the result of analysis of all NF-Y sites in most cell types, showing that NF-Y locates preferably in promoters [42].

To verify NF-Y TE locations based on functional features, we exploited the ENCODE definition of cCREs -candidate *cis*-Regulatory Elements- based on the mapping in several cell lines of H3K4me3, H3K27ac, DNase I hypersensitivity and CTCF binding [47]. This catalogue comprises 5 categories: Promoters (PLS), annotated as such if they are within -/+200 bp from a TSS; DNase-H3K4me3 (without H3K27ac) are promoters but without a functional annotation nearby; CTCF-only; Unmarked; Enhancers (ELS), in which we included both Distal and Proximal enhancer locations. First, we checked the overall presence of cCREs as a percentage of the total number of the TE subfamilies: the majority are in un-marked areas, with 10/20% having Enhancers signatures, very few in Promoters, present exclusively in LTR12E/D/C (Fig. S5). Thereafter, we matched NF-Y locations to cCREs within TEs in the different cell lines, only considering subfamilies with more than 50 locations bound (Fig. 6). As expected, K562 and GM12878 show a higher number of peaks, particularly in LTR12 and MLT1 (Fig. 6A Top and Middle Panel). LTR12, LTR12C, LTR12D and MLT1K are numerous, as expected from data of Fig. 2. As to functional categories, the most abundant is Unmarked across the board: this is in line with previous data showing that binding to most LTR sequences is not associated to histone PTMs marks [20]. The other LTR12 and MLT1 sites are essentially within Enhancers, not Promoters, although a sizeable number of sites are marked as
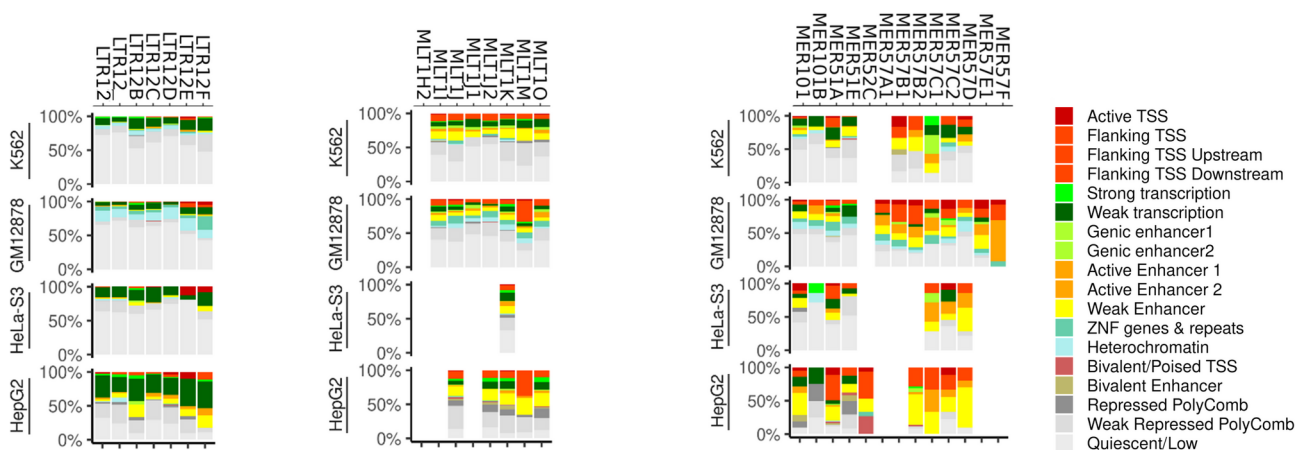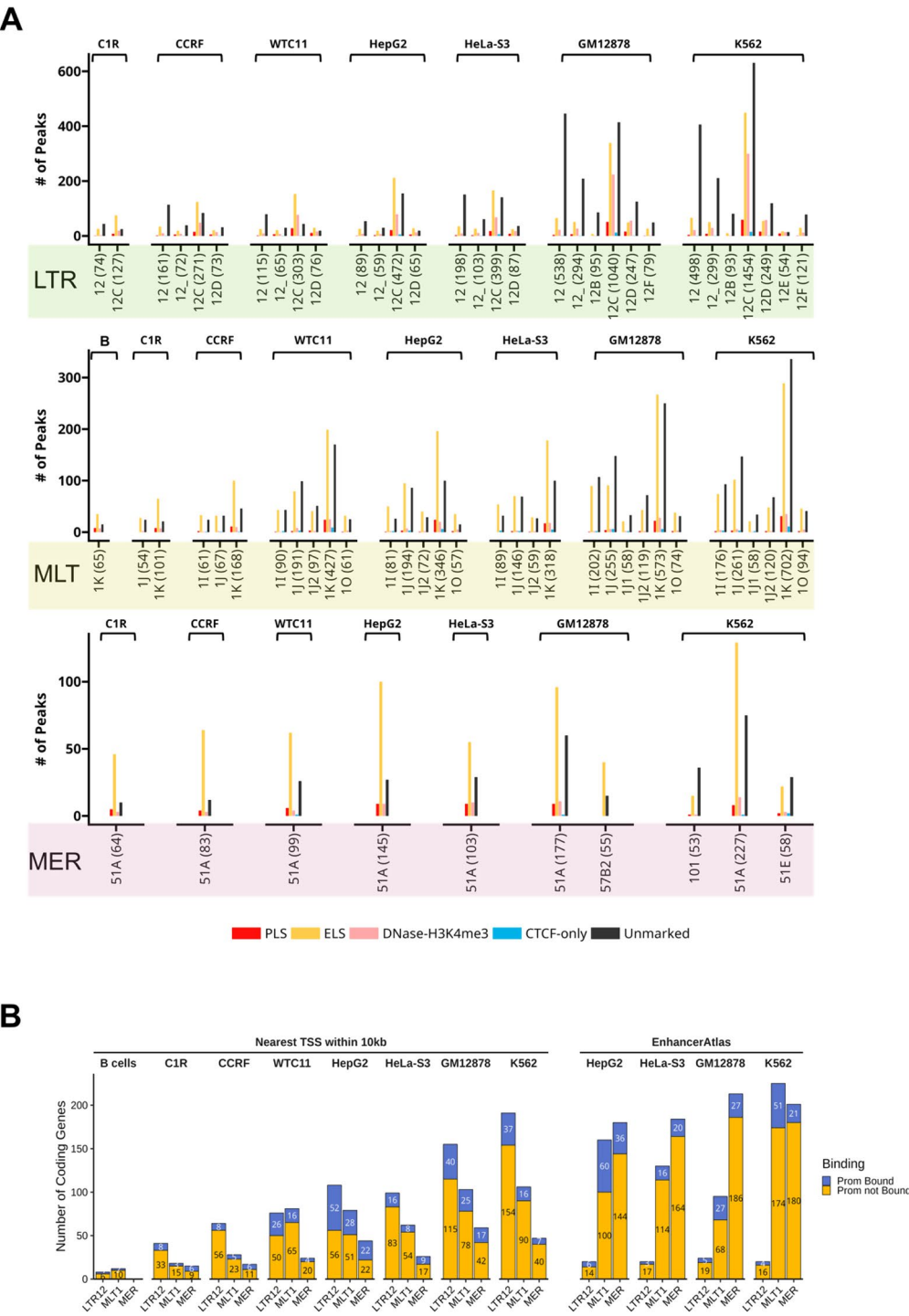

Fig. 5 Chromatin features of NF-Y-bound TEs. Percentage of chromatin states associated with NF-Y bound LTR12 (Left Panel), MLT1 (Middle Panel) and MERs (Right Panel) repetitive elements. The different colors designate the chromatin states as derived from Roadmap Epigenomics

Fig. 6 cCREs annotations of NF-Y-bound TEs. **A**. Number of NF-Y peaks in cCREs falling within LTR12 (Top Panel), MLT1 (Middle Panel) and MER (Bottom Panel) sites in human cell lines. PLS = Promoter-Like Signature; ELS = Enhancer-Like Signature. In parenthesis the total number of sites bound for each specific repetitive element. **B.** Left Panel: analysis of the binding of NF-Y on the nearest TSS within 10 kb of the NF-Y-LTR considered (ELS and Unmarked). Right Panel: analysis of NF-Y binding on promoters interacting with NF-Y-LTR considered (ELS and Unmarked) according to Enhancer Atlas 2.0 data [48]

DNase-H3K4me3, features of promoters without a TSS annotation. In MER, the numbers are relatively balanced in the different cell lines, with MER51A being widespread and the Enhancer category prevailing (Fig. 6A Bottom Panel). Globally, these data indicate that NF-Y TEs locations are few in promoters, most residing in chromatin without positive marks -especially LTR12 and MLT1- or in enhancers (MER51).

Finally, we matched NF-Y-LTR classified as ELS and Unmarked to promoters across all human cell lines, considering the nearest TSS located within a distance of 10 kb from an NF-Y-LTR: the results in Fig. 6B (Left Panel) show that only 20/40% of LTR considered are close to an NF-Y-bound promoter. To substantiate this finding, we made use of EnhancerAtlas 2.0 [48], a database that contains predicted, potentially multiple, enhancer-promoter interactions: analysis of ENCODE cell lines indicates that the percentage of cobound Enhancer/Promoters is consistently low, with unexpectedly few LTR12 units, compared to the high number of LTR12 bound by NF-Y (Fig. 6B Right Panel). We conclude that NF-Y-LTR enhancers do not necessarily match to NF-Y-bound promoters.

RNA production in TEs bound by NF-Y

To measure the transcriptional activity of the TEs bound by NF-Y, we used the ENCODE total RNA-seq data of K562, HepG2 and GM12878. Independently from the levels of RNAs, we partitioned the TE *loci* as Transcribed (T, TPM > 0), Not Transcribed (NT), NF-Y Bound (B) and Not Bound (NB): transcribed and bound *loci* are significantly more numerous than the expected in all tests, therefore NF-Y is more bound to transcribed *loci* for all TEs (Fig. 7A). We considered expression levels, by filtering out all non-transcribed TEs: this exercise returns higher levels of LTR12 transcribed and bound by NF-Y in GM12878 and HepG2 (Fig. 7B left Panel), but not in the other TEs, in which NF-Y binding is associated to lower expression (Fig. 7B Central and Right Panel). This is consistent with NF-Y binding being a positive factor only for LTR12 in GM12878 and HepG2, but not in the other TEs nor in K562.

A recent global map of enhancer RNAs (eRNAs) was derived in 20 cell lines, based on multiple parameters [49]. Focusing exclusively on eRNAs *loci* in LTR12, MLT1 and MER sites, we computed binding of NF-Y in K562, GM12878 and HepG2: the whole set of LTR, MLT1 and MER TEs contain eRNAs producing *loci* at 47%, 40% and 12% of the total; in those bound by NF-Y, we score a significant increase in all TEs, at 73%, 58% and 36%, respectively (Fig. 7C). The p-values at $10^{-5/17}$ are significant; NF-Y binding in Hela-S3, C1R, CCRF and WTC11 have lower significance, especially for MLT1 and MER (Fig. S6). To check the effect of the removal of NF-Y

transcription in NF-Y-LTR, we chose 4 regions positive for NF-Y binding showing mapped RNA-seq reads in all ENCODE cell lines, and we analyzed the expression by qRT-PCR on HeLa inactivated for NF-YB [44]. First, we checked expression of NF-YB and NF-YA to verify the downregulation of the former gene NF-YB and upregulation of the latter (Fig. 7D). Four of the six regions are modestly upregulated after the NF-YB inactivation (Fig. 7D). Importantly, the MLTK1 is one of the regions classified as eRNA. RT- samples were negative.

We conclude that NF-Y binding to TEs in enhancers is associated to eRNAs production, but not universally, depending on the cell types and on the repetitive sequence considered. In addition, NF-Y could be part of a repressive mechanism in selected TEs.

Discussion

The present study sheds light on the extent of NF-Y binding to specific classes of repetitive sequences of retroviral origin in human and mouse cells.

TEs contribute substantially to the landscape of cis-acting elements in mammals, providing several promoters and enhancers. Studies on RefSeq genes indicated that *cell-cycle*, *metabolism* and *transcriptional regulation* units are a major part of the "core" NF-Y regulome (Reviewed by [18]), specifically in promoters, where CCAAT is positioned at -60/-100 relative to the TSS [11].

Recent evidence further expand the role of NF-Y in promoters in helping to determine TSS locations [13–16] (reviewed by [50]). Some of these AI-based studies predict the presence of multiple TSS in CCAAT promoters, and the elimination of CCAAT/NF-Y located upstream of the TSS has a negative impact on transcription, as validated with wet experiments [13, 16]. Consequent with these findings, the most prominent localization of NF-Y is in promoters of human and mouse cells [42]. MER, and to a lesser extent MLT1, have epigenomic marks typical of promoters (Fig. 5).

Distal locations are essentially associated to tissue-specific enhancers, a sizable number of which are of retroviral origin. A stringent classification based on the presence of nearby (200 bp) annotated transcripts, as well as analysis of eRNA *loci* (Fig. 7), suggests that only a small fraction of cCREs can be considered "promoters". On LTR12, "epigenomic" and cCREs analysis concur that promoters bound by NF-Y are rare. Globally, we conclude that NF-Y binding to *bona fide* TE promoter sequences is an exception, whereas bound enhancers, or "middle-of-nowhere" sites without common histone marks, could generate enhancer RNAs -eRNAs- or cryptic transcripts.

The function of NF-Y in LTR12 appears to be complex. The data of Figs. 5 and 6 indicate that most LTR12 sites bound by NF-Y lie in locations marked as "repressed" or "quiescent". On MLT1 and MERs, NF-Y-binding is
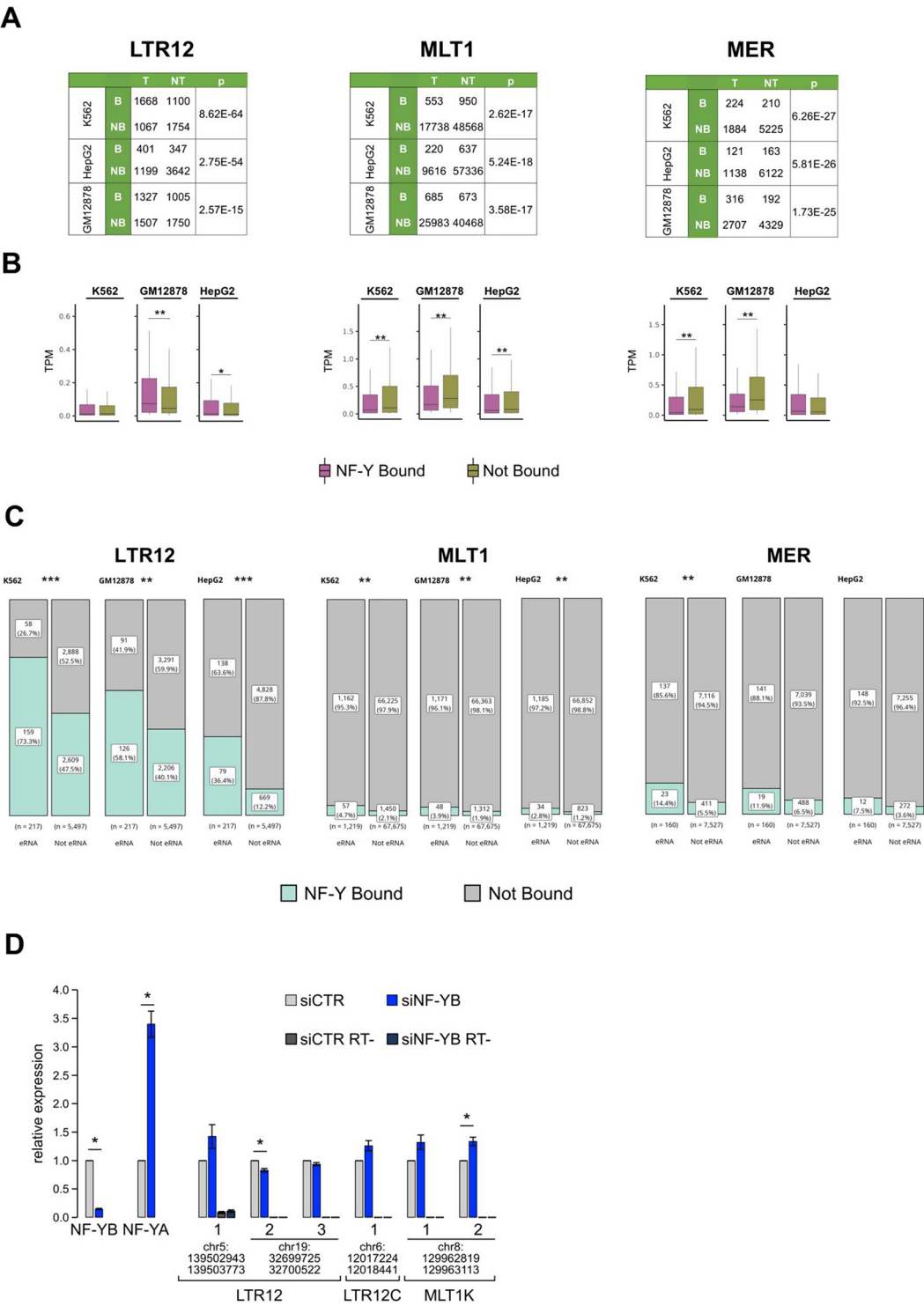
Fig. 7 RNA production in TEs bound by NF-Y. **A.** Contingency table of NF-Y binding to repetitive elements and repetitive transcripts in ENCODE K562, GM12878 and HepG2 cells. Transcribed (T) and Bound (**B**) repeats number higher than the relative expected number in all cases. **B**. Distribution of transcription levels of transcribed repetitive sequences bound or not bound by NF-YB. **C.** Comparison of eRNA and not eRNA *loci* within TEs bound or not bound by NF-Y in K562 (Left Panel), GM12878 (Middle Panel) and HepG2 (Right Panel) cell lines. Statistical significance assessed with Fisher's exact test (**A**, **C**) and Wilcoxon rank-sum test (**B**). Significance levels are given by stars: * -10^{-2}, ** -10^{-3} and *** -10^{-10}. **D.** RT-qPCR quantification of the expression of several LTR regions upon NF-YB siRNA-treatment in HeLa cells. Relative expression was expressed relative to control siRNA treated cells (siCTR). For LTR targets also RT- control results are shown to evaluate genomic DNA carryover. The genomic regions assessed are indicated below the plot. Bar plots correspond to the average of three biological replicas ($n=3$). Error bars correspond to the SEM. * pvalue < 0.05 according to one-sample t-test

associated to lower levels of transcripts in all cells, except in GM12878 and HepG2, in which transcripts generated from bound sites are higher than from unbound ones (Fig. 7). The functional experiments on specific RNAs generated from TEs -shown in Fig. 7D- indicate that removal of NF-Y has either a neutral effect, or a -modest- positive one, suggesting a repressive role. On one hand, many of the LTR12 *loci* are indeed enhancers, based on functional annotations, extending the notion originally described in enhancers of the globin gene cluster [27–29, 51]. Possibly, these are tissue-restricted, as suggested by the findings of LTR12 enhancers driving expression of endoderm-specific genes [37] and of RLTR10B enhancers in mouse spermatocytes [22].

The present data should consider experiments indicating that modules of retroviral origin become de-repressed/activated upon treatment of -cancer- cells with inhibitors of HDACs or DNMTs, pointing at NF-Y as important for those originating specifically from LTR12B/C [31–33, 36]. The original observation of induction of the pro-apoptotic TNFRSF10B gene by the LTR12 enhancer [33] was later extended in two directions: the use of additional HDAC inhibitors and hints that LTR12 enhancers mediate induction preferentially of pro-apoptotic genes [38].

Another study supports enhancer repression of LTR12: the repressor ZNF676 -and the ZNF728 paralogue- are bound to LTR12C, together with their corepressor KAP1/TRIM28 and NF-Y [21]; these sites are possibly the ones we find abundantly bound in K562 and GM12878 embedded in inactive (Fig. 5), unmarked (Fig. 6) chromatin. Interestingly, ZNF676 is hominoid-specific, as are ERV9/LTR12 [52], implying that classes of Zn-finger repressors have co-evolved along the invasion of LTR12 in *Hominidae.* Indeed, among ENCODE-tested TFs, we identify cobinding of repressors -MBD2, ZBTB33/Kaiso, C11Orf30/EMSY- or TFs lacking activation domains, such as small MAFs (MAFG, MAFK, MAFF); this could be interpreted as NF-Y being part of a repressive mechanism, since these b-ZIP TFs have been associated to repression [53]

Mechanistically, the partial overlap of the NF-Y/MAFs locations (Fig. S2) could result in mutually exclusive binding to the respective sequences, or a tight relationship between the TFs. Further biochemical experiments with recombinant TFs should discriminate this point, but we remark the following: (i) sequence-specificity conferred by the NF-YA subunit involves exclusively minor groove interactions [54], leaving available the major groove, where b-ZIP proteins interact with DNA. (ii) An equally intricated interplay was dissected on *ER R*esponsive *E*lement (ERSE) II, between NF-Y and the b-ZIP ATF6, overlapping at the 5' end of CCAAT: cobinding was indeed reported [55–57].

A further twist has been recently proposed concerning the activity of the U7 snRNA involved in 3' processing of histone pre-mRNAs, which represses LTR12-driven transcription of lncRNAs through HDE-like motifs [58]: the importance of NF-Y was tested by RNAi-inactivation of the three subunits, and while expression is unchanged under normal conditions, the induction driven by U7 snRNA elimination is abolished, mirroring what is reported with HDACi and DNMTi treatments. Intriguingly, the U7 snRNA was previously shown to interact with NF-YA [59], possibly suggesting that this ncRNA blocks NF-Y function directly.

Studying the role of specific cofactors on enhancers, Neumayr et al. identified LTR12C/D *loci* as induced by removal of BRD4 [40], a cofactor impacting on the regulatory pause-release step of RNA Pol II elongation [60, 61]. These sites are embedded in closed chromatin with mild positivity for H3K4me3/H3K27me3, but no marks associated to enhancers. CCAAT boxes in these TEs are functionally important, as assessed by RNAi of NF-YA/NF-YB [40]. In summary, NF-Y could play an active role of some LTR12-based enhancers, as well as a repressive one on others, becoming derepressed upon HDAC/DNMT/U7 inhibition.

As for MLT1 sequences, an important NF-Y partnership is with USF1/2, which comes at the top of the list of co-bound TFs, as shown before [7, 20]. Cooperative DNA-binding, mediated by the USF1 USR domain, was dissected by biochemical assays [46]. NF-YB inactivation evicts NF-Y from MLT1 sequences, whereas USF1 binding is still substantial, and the opposite happens on promoters [42]. We take this as an indication that USF1 has recruiting activities in these TEs independently from NF-Y, which is therefore not playing a "pioneering" role. Finally, we identify MER sites in which NF-Y is co-bound with TALE PBX2/PBX3/PKNOX1 containing homeobox DNA-binding domains. This partnership was originally discovered in mouse and Zebrafish developmentally controlled genes, and a common, discrete alignment of sites characterized: TALE-10 bp-CCAAT (Reviewed by [45]). Note that this configuration is similar to the E-box-12 bp-CCAAT shown for USF1/NF-Y (Fig. S3). Instead, the sites found in MER are quite different, with TALE being at the 3' of CCAAT and at a considerable (50 bp) distance, which is not consistent with predictions of direct interactions. Furthermore, a STAT1/2 site is located in between.

We provide here a first comprehensive outlook of NF-Y binding to mouse TEs, pointing at selected classes of retroviral origin, RLTR10B and intracisternal A-type particles (IAPs), rich in CCAAT boxes.

By looking at changes of gene expression of TE-driven RNAs during mouse spermatogenesis, Sakashita et al. found a specific role for RLTR10B in enhancers driving expression of genes involved in the maturation of

pachytene spermatocytes, notably in mitosis-to-meiosis transitions [25]. These Authors noticed NF-Y/CCAAT and A-MYB sites in these elements and went on to prove the role of A-MYB in the activation of neighboring germ-line genes.

We find here that RLTR10B -and the related RLTR10A, RLTR10B2, RLTR10C- are indeed bound by NF-Y in several cell types. As for human LTR12, degradation of the KAP1/TRIM28 repressors led to activation of RLTRIAPs in mouse ESCs: in this context, IAPs-generated ncRNAs apparently hijack the condensates formed by RNA Pol II, Mediator and cofactors normally recruited at super-enhancers; in turn, the resulting changes in gene expression cause depletion of pluripotent lineages in vivo [41]. Interestingly, ncRNAs generated by IAPs are part of the condensates, and the Authors showed that the intrinsically disordered region -IDR- of NF-YC enhances the capacity of IAPLTR-ncRNAs to form droplets with Pol II CTD in vitro. By confirming binding of NF-Y to LTR10B -widespread- and IAP -in transformed MEL and CH27 cells (Fig. 2)- we extend these findings beyond the spermatocytes and mESC contexts.

The current data converge on two non-mutually exclusive scenarios regarding binding to apparently non-active sites.

First, TEs could serve as *reservoir* of NF-Y molecules that can be readily mobilized to keep growth-promoting genes up and running in cycling cells. In this respect, we remark a higher percentage of TEs bound in transformed human cells compared to normal B lymphocytes; in mouse, there is a general trend of lower fractions of TEs bound (Fig. 1), but higher percentages are in transformed MEL, CH27, CH12 and mESCs, compared to normal myoblasts/myocytes, MEFs, keratinocytes. Whether this is a biological trend should be assessed in systems of step-wise cell transformation.

Second, NF-Y could take an active part in repression of some of these locations or keep their transcriptional activity low; inhibition of the repressive HDACs, DNMTs, or removal of KAP1 or ZNF676, is sufficient to unleash a positive role for NF-Y. Note, however, that this was not observed in U7 snRNA-dependent units. It remains to be studied whether the repressive complexes are recruited in part *via* NF-Y/CCAAT -or the NF-Y/MAF combination- and what are the relationships with U7 snRNAs.

Conclusions

Our results represent a useful resource of genomic data on NF-Y binding to repetitive sequences, widening our knowledge in human cells, and demonstrating its presence in mouse cells for the first time. Put in the context of functional data, and on recent findings in the literature, we propose that NF-Y has a complex role -positive or negative- in three major classes of TEs bound.

Methods

Overlap analysis and annotation

Overlap analysis was conducted as follows (Fig. S1). ChIP-seq peaks were analyzed with ChIPseeker R package [62] and overlaps were performed with plyranges R package [63]. Bound repetitive regions were calculated intersecting peaks summit to repeatmasker annotation: hg38 and mm10 for human and mouse respectively (Fig. 1).

Genomic annotation within TEs was obtained by running annotatePeak function on peaks summit, setting -450 and 50 as 'TSS' region, and using UCSC knownGene database (Fig. S4). The genomic features annotation was collapsed for each experiment as previously described [42].

Analogously, chromatin states of TEs were obtained by overlapping peaks summit to the Roadmap Epigenomics Project mnemonics bed files (Fig. 5).

ENCODE SCREEN platform was used to recover the human cCREs locations. cCREs were collapsed in fewer categories merging distal Enhancer (dELS) and proximal Enhancer (pELS) in Enhancer (ELS). Resulted regions were intersected with selected TEs bound by NF-Y in Fig. 3 (Figs. 6, S5). Enhancer-target gene links of human ENCODE cell lines were retrieved from Enhancer Atlas 2.0 [48] and were overlapped with repetitive elements bound by NF-Y annotated as ELS or Unmarked, according to the cCREs (Fig. 6).

eRNA regions were downloaded from the TCeA database [49] and overlapped with the selected TEs bound by NF-Y (Figs. 7, S6).

Repbase annotation enrichment

Raw data were homogenized in read length as previously described (Fig. S1) [42].

Reads of each ChIP-seq experiment were aligned against the collection of repeats consensus sequences available in the Repbase 26.10 database [64] with bowtie2, allowing for multi-mapping events and reporting primary alignments. To assess enrichment in repeats with respect to the rest of the genome, ChIP-seq reads were also aligned against the reference genome, keeping only uniquely mapping reads.

For each repeat consensus of Repbase, the enrichment score was then calculated as follows:

- Given X reads mapping on a repeat consensus over a total of N mapping on the genome for the IP experiment;

- Given Y reads mapping on a repeat consensus over a total of M mapping on the genome for the corresponding control (Input or IgG);

The enrichment score s was computed as $s = X/E$, where the $E = N(Y/M)$ represents the expected number of reads mapping on the repeat by chance, estimated from the control experiment. Enrichment values were finally log-corrected. The corresponding statistical significance was assessed according to a Poisson distribution, with $\lambda = E$. N and M were calculated as the overall number of reads uniquely mapping on the genome and to repeats, respectively.

Genomic alignment
Repetitive elements classes with high enrichment in binding for NF-Y and the indicated TF in the correlation heatmap in Fig. 3 were selected for genomic alignments. Sequences were first filtered for the binding of NF-Y and the indicated TF in at least one experiment. Genomic alignments were performed with Jalview v. 2.11.4.1 [65] running Muscle algorithm set with default parameters.

Motif analysis
The number of CCAAT boxes present within repetitive elements (using the Repbase 26.10 humrep.ref file) shown in Fig. 3 was computed by a custom Python script calling BioPython [66] library functions. Quality scores of the CCAAT box instances were computed with Pscan [67].

Motif enrichment in sequences between conserved motifs in genomic alignments was run with XSTREME v.5.5.7 from MEME suite [68], giving as input the aligned sequences and filtering out gaps.

RNA-seq data processing
Sequence reads for each experiment were aligned onto the human genome (assembly hg38) with STAR [69], allowing multimapping up to 100 positions. Feature-Counts [70] was employed to estimate read counts for the repetitive sequences providing the repeatmasker-based gtf file available on http://hammelllab.labsites.cshl.edu/software/ website. Resulting counts on repetitive *loci* were normalized to TPM for downstream analyses.

Statistical analyses and plots
Statistical analyses were run -and plots generated- in R environment (v. 4.2.0) using tidyverse [71] suite and ggstatsplot [72]. Statistical tests were indicated in each Figure legend.

RT-qPCR
Total RNA obtained from siRNA treated HeLa cells [44] was purified using RNeasy Plus Micro (Qiagen) kit, including the genomic DNA elimination column step.

One microgram of RNA was reverse-transcribed using SuperScript II reverse transcriptase (Invitrogen) and random primers according to the manufacturers' protocol, including RT- controls. cDNA was diluted 1:3 and used in qPCR reactions using SsoAdvanced Universal SYBR Green Supermix (Biorad) in a CFX Duet Real-Time PCR System (Biorad). Primers used are listed in Supplementary Table S1. Two technical replicate reactions were run for each of the three biological replicates ($n = 3$). RT- samples were analyzed to assess the extent of genomic DNA carryover. Relative expression was calibrated to the ribosomal protein gene RPS20 and normalized to the control siRNA treated sample using the $2^{-\Delta\Delta Ct}$ method.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13100-025-00358-9.

Supplementary Material 1

Supplementary Material 2

Data availability
ChIP-seq raw sequencing data of NF-YB were retrieved from [42] (PRJNA861853), [73](ERR637890),[19](SRR1239520), [74] (SRR18486079/80/81) and ENCODE repository (experiments ENCSR146UIC, ENCSR935GZV, ENCSR000DNR, ENCSR000DNM, ENCSR000EGQ). IDR thresholded peaks of NF-YB, TBP and USF1 were downloaded from[42]. Genomic locations of the factors used to build co-binding clusters of Fig. 3 were downloaded from ENCODE repository as IDR thresholded peaks. Total RNA-seq of K562, GM12878 and HepG2 cell lines (experiment identifiers ENCSR000CPY, ENCSR000CVT, ENCSR181ZGR) are from ENCODE. Any additional information required to reanalyze the data reported in this paper is available upon request.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

References

1. Jordan IK, Rogozin IB, Glazko GV, Koonin EV. Origin of a substantial fraction of human regulatory sequences from transposable elements. Trends Genet. 2003;19:68–72. https://doi.org/10.1016/S0168-9525(02)00006-9.

2. Jacques P-É, Jeyakani J, Bourque G. The majority of Primate-Specific regulatory sequences are derived from transposable elements. PLOS Genet. 2013;9:e1003504. https://doi.org/10.1371/journal.pgen.1003504.

3. Thompson PJ, Macfarlan TS, Lorincz MC. Long terminal repeats: from parasitic elements to Building blocks of the transcriptional regulatory repertoire. Mol Cell. 2016;62:766–76. https://doi.org/10.1016/j.molcel.2016.03.029.

4. Sundaram V, Wysocka J. Transposable elements as a potent source of diverse cis-regulatory sequences in mammalian genomes. Philos Trans R Soc Lond B Biol Sci. 2020;375:20190347. https://doi.org/10.1098/rstb.2019.0347.

5. Kellner M, Makałowski W. Transposable elements significantly contributed to the core promoters in the human genome. Sci China Life Sci. 2019;62:489–97. https://doi.org/10.1007/s11427-018-9449-0.

6. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, et al. Evolution of the mammalian transcription factor binding repertoire via transposable elements. Genome Res. 2008;18:1752–62. https://doi.org/10.1101/gr.080663.108.

7. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. 2012;22:1798–812. https://doi.org/10.1101/gr.139105.112.

8. Wang J, Zhuang J, Iyer S, Lin X-Y, Greven MC, Kim B-H, et al. Factorbook.org: a Wiki-based database for transcription factor-binding data generated by the ENCODE consortium. Nucleic Acids Res. 2012;41:D171–6. https://doi.org/10.1093/nar/gks1221.

9. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, et al. Widespread contribution of transposable elements to the innovation of gene regulatory networks. Genome Res. 2014;24:1963–76. https://doi.org/10.1101/gr.168872.113.

10. Ito J, Sugimoto R, Nakaoka H, Yamada S, Kimura T, Hayano T, et al. Systematic identification and characterization of regulatory elements derived from human endogenous retroviruses. PLoS Genet. 2017;13:e1006883. https://doi.org/10.1371/journal.pgen.1006883.

11. Dolfini D, Zambelli F, Pavesi G, Mantovani R. A perspective of promoter architecture from the CCAAT box. Cell Cycle. 2009;8:4127–37. https://doi.org/10.4161/cc.8.24.10240.

12. Santana JF, Collins GS, Parida M, Luse DS, Price DH. Differential dependencies of human RNA polymerase II promoters on TBP, TAF1, TFIIB and XPB. Nucleic Acids Res. 2022;50:9127–48. https://doi.org/10.1093/nar/gkac678.

13. Dudnyk K, Cai D, Shi C, Xu J, Zhou J. Sequence basis of transcription initiation in the human genome. Science. 2024;384:eadj0116. https://doi.org/10.1126/science.adj0116.

14. He AY, Danko CG. Dissection of core promoter syntax through single nucleotide resolution modeling of transcription initiation 2024:2024.03.13.583868. https://doi.org/10.1101/2024.03.13.583868

15. Cochran K, Yin M, Mantripragada A, Schreiber J, Marinov GK, Shah SR et al. Dissecting the cis-regulatory syntax of transcription initiation with deep learning 2024:2024.05.28.596138. https://doi.org/10.1101/2024.05.28.596138

16. Duttke SH, Guzman C, Chang M, Delos Santos NP, McDonald BR, Xie J, et al. Position-dependent function of human sequence-specific transcription factors. Nature. 2024;631:891–8. https://doi.org/10.1038/s41586-024-07662-z.

17. Dolfini D, Mantovani R. Targeting the Y/CCAAT box in cancer: YB-1 (YBX1) or NF-Y? Cell Death Differ. 2013;20:676–85. https://doi.org/10.1038/cdd.2013.13.

18. Dolfini D, Gnesutta N, Mantovani R. Expression and function of NF-Y subunits in cancer. Biochim Biophys Acta BBA - Rev Cancer. 2024;1879:189082. https://doi.org/10.1016/j.bbcan.2024.189082.

19. Oldfield AJ, Henriques T, Kumar D, Burkholder AB, Cinghu S, Paulet D, et al. NF-Y controls fidelity of transcription initiation at gene promoters through maintenance of the nucleosome-depleted region. Nat Commun. 2019;10. https://doi.org/10.1038/s41467-019-10905-7.

20. Fleming JD, Pavesi G, Benatti P, Imbriano C, Mantovani R, Struhl K. NF-Y coassociates with FOS at promoters, enhancers, repetitive elements, and inactive chromatin regions, and is stereo-positioned with growth-controlling transcription factors. Genome Res. 2013;23:1195–209. https://doi.org/10.1101/gr.148080.112.

21. Iouranova A, Grun D, Rossy T, Duc J, Coudray A, Imbeault M, et al. KRAB zinc finger protein ZNF676 controls the transcriptional influence of LTR12-related endogenous retrovirus sequences. Mob DNA. 2022;13:1–17. https://doi.org/10.1186/s13100-021-00260-0.

22. Guo J, Grow EJ, Yi C, Mlcochova H, Maher GJ, Lindskog C, et al. Chromatin and Single-Cell RNA-Seq profiling reveal dynamic signaling and metabolic transitions during human spermatogonial stem cell development. Cell Stem Cell. 2017;21:533–e5466. https://doi.org/10.1016/j.stem.2017.09.003.

23. Lu X, Luo Y, Nie X, Zhang B, Wang X, Li R, et al. Single-cell multi-omics analysis of human testicular germ cell tumor reveals its molecular features and microenvironment. Nat Commun. 2023;14:8462.

24. Wu X, Lu M, Yun D, Gao S, Chen S, Hu L, et al. Single-cell ATAC-Seq reveals cell type-specific transcriptional regulation and unique chromatin accessibility in human spermatogenesis. Hum Mol Genet. 2022;31:321–33.

25. Sakashita A, Maezawa S, Takahashi K, Alavattam KG, Yukawa M, Hu Y-C, et al. Endogenous retroviruses drive species-specific germline transcriptomes in mammals. Nat Struct Mol Biol. 2020;27:967–77. https://doi.org/10.1038/s41594-020-0487-4.

26. Li J, Shen S, Chen J, Liu W, Li X, Zhu Q, et al. Accurate annotation of accessible chromatin in mouse and human primordial germ cells. Cell Res. 2018;28:1077–89.

27. Yu X, Zhu X, Pi W, Ling J, Ko L, Takeda Y, et al. The long terminal repeat (LTR) of ERV-9 human endogenous retrovirus binds to NF-Y in the assembly of an active LTR enhancer complex NF-Y/MZF1/GATA-2. J Biol Chem. 2005;280:35184–94. https://doi.org/10.1074/jbc.M508138200.

28. Pi W, Zhu X, Wu M, Wang Y, Fulzele S, Eroglu A, et al. Long-range function of an intergenic retrotransposon. Proc Natl Acad Sci U S A. 2010;107:12992–7. https://doi.org/10.1073/pnas.1004139107.

29. Zhu X, Wang Y, Pi W, Liu H, Wickrema A, Tuan D. NF-Y recruits both transcription activator and repressor to modulate tissue- and developmental stage-specific expression of human γ-globin gene. PLoS ONE. 2012;7:e47175. https://doi.org/10.1371/journal.pone.0047175.

30. Jung Y-D, Lee H-E, Jo A, Hiroo I, Cha H-J, Kim H-S. Activity analysis of LTR12C as an effective regulatory element of the RAE1 gene. Gene. 2017;634:22–8. https://doi.org/10.1016/j.gene.2017.08.037.

31. Brocks D, Schmidt CR, Daskalakis M, Jang HS, Shah NM, Li D, et al. DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. Nat Genet. 2017;49:1052–60. https://doi.org/10.1038/ng.3889.

32. Beyer U, Krönung SK, Leha A, Walter L, Dobbelstein M. Comprehensive identification of genes driven by ERV9-LTRs reveals TNFRSF10B as a re-activatable mediator of testicular cancer cell death. Cell Death Differ. 2016;23:64–75. https://doi.org/10.1038/cdd.2015.68.

33. Krönung SK, Beyer U, Chiaramonte ML, Dolfini D, Mantovani R, Dobbelstein M. LTR12 promoter activation in a broad range of human tumor cells by HDAC Inhibition. Oncotarget. 2016;7:33484–97. https://doi.org/10.18632/oncotarget.9255.

34. Daskalakis M, Brocks D, Sheng Y-H, Islam MS, Ressnerova A, Assenov Y, et al. Reactivation of endogenous retroviral elements via treatment with DNMT- and HDAC-inhibitors. Cell Cycle Georget Tex. 2018;17:811–22. https://doi.org/10.1080/15384101.2018.1442623.

35. Ohtani H, Liu M, Zhou W, Liang G, Jones PA. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. Genome Res. 2018;28:1147–57. https://doi.org/10.1101/gr.234229.118.

36. White CH, Beliakova-Bethell N, Lada SM, Breen MS, Hurst TP, Spina CA, et al. Transcriptional modulation of human endogenous retroviruses in primary CD4 + T cells following Vorinostat treatment. Front Immunol. 2018;9:603. https://doi.org/10.3389/fimmu.2018.00603.

37. Karttunen K, Patel D, Xia J, Fei L, Palin K, Aaltonen L, et al. Transposable elements as tissue-specific enhancers in cancers of endodermal lineage. Nat Commun. 2023;14:5313. https://doi.org/10.1038/s41467-023-41081-4.

38. Gualandi N, Minisini M, Bertozzo A, Brancolini C. Dissecting transposable elements and endogenous retroviruses upregulation by HDAC inhibitors in leiomyosarcoma cells: implications for the interferon response. Genomics. 2024;116:110909. https://doi.org/10.1016/j.ygeno.2024.110909.

39. Ohtani H, Liu M, Liang G, Jang HJ, Jones PA. Efficient activation of hundreds of LTR12C elements reveals cis-regulatory function determined by distinct epigenetic mechanisms. Nucleic Acids Res. 2024;52:8205–17. https://doi.org/10.1093/nar/gkae498.

40. Neumayr C, Haberle V, Serebreni L, Karner K, Hendy O, Boija A, et al. Differential cofactor dependencies define distinct types of human enhancers. Nature. 2022;606:406–13. https://doi.org/10.1038/s41586-022-04779-x.

41. Asimi V, Sampath Kumar A, Niskanen H, Riemenschneider C, Hetzel S, Naderi J, et al. Hijacking of transcriptional condensates by endogenous retroviruses. Nat Genet. 2022;54:1238–47. https://doi.org/10.1038/s41588-022-01132-w.

42. Ronzio M, Bernardini A, Taglietti V, Ceribelli M, Donati G, Gallo A, et al. Genomic binding of NF-Y in mouse and human cells. Genomics. 2024;116:110895. https://doi.org/10.1016/j.ygeno.2024.110895.

43. Dolfini D, Zambelli F, Pedrazzoli M, Mantovani R, Pavesi G. A high definition look at the NF-Y regulome reveals genome-wide associations with selected transcription factors. Nucleic Acids Res. 2016;44:4684–702. https://doi.org/10.1093/nar/gkw096.

44. Ronzio M, Bernardini A, Pavesi G, Mantovani R, Dolfini D. On the NF-Y regulome as in ENCODE (2019). PLOS Comput Biol. 2020;16:e1008488. https://doi.org/10.1371/journal.pcbi.1008488.

45. Dolfini D, Imbriano C, Mantovani R. The role(s) of NF-Y in development and differentiation. Cell Death Differ 2024:1–12. https://doi.org/10.1038/s41418-024-01388-1

46. Bernardini A, Lorenzo M, Chaves-Sanjuan A, Swuec P, Pigni M, Saad D, et al. The USR domain of USF1 mediates NF-Y interactions and cooperative DNA binding. Int J Biol Macromol. 2021;193:401–13. https://doi.org/10.1016/j.ijbiomac.2021.10.056.

47. Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shoresh N, Adrian J, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. Nature. 2020;583:699–710. https://doi.org/10.1038/s41586-020-2493-4.

48. Gao T, Qian J. EnhancerAtlas 2.0: an updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. 2020;48:D58–64. https://doi.org/10.1093/nar/gkz980.

49. Chen H, Liang HA, High-Resolution. Map of human enhancer RNA loci characterizes Super-enhancer activities in Cancer. Cancer Cell. 2020;38:701–15. https://doi.org/10.1016/j.ccell.2020.08.020..e5.

50. Bernardini A, Mantovani R. Q-rich activation domains: flexible 'rulers' for transcription start site selection? Trends Genet 2024;0. https://doi.org/10.1016/j.tig.2024.11.008

51. Hu T, Pi W, Zhu X, Yu M, Ha H, Shi H, et al. Long non-coding RNAs transcribed by ERV-9 LTR retrotransposon act in cis to modulate long-range LTR enhancer function. Nucleic Acids Res. 2017;45:4479–92. https://doi.org/10.1093/nar/gkx055.

52. López-Sánchez P, Costas JC, Naveira HF. Paleogenomic record of the extinction of human endogenous retrovirus ERV9. J Virol. 2005;79:6997–7004. https://doi.org/10.1128/JVI.79.11.6997-7004.2005.

53. Moyers BA, Partridge EC, Mackiewicz M, Betti MJ, Darji R, Meadows SK, et al. Characterization of human transcription factor function and patterns of gene regulation in HepG2 cells. Genome Res. 2023;33:1879–92. https://doi.org/10.1101/gr.278205.123.

54. Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, et al. Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell. 2013;152:132–43. https://doi.org/10.1016/j.cell.2012.11.047.

55. Kokame K, Kato H, Miyata T. Identification of ERSE-II, a new *cis*-Acting element responsible for the ATF6-dependent mammalian unfolded protein response**. J Biol Chem. 2001;276:9199–205. https://doi.org/10.1074/jbc.M010486200.

56. Ma Y, Brewer JW, Diehl JA, Hendershot LM. Two distinct stress signaling pathways converge upon the CHOP promoter during the mammalian unfolded protein response. J Mol Biol. 2002;318:1351–65. https://doi.org/10.1016/s0022-2836(02)00234-6.

57. Yamamoto K, Yoshida H, Kokame K, Kaufman RJ, Mori K. Differential contributions of ATF6 and XBP1 to the activation of Endoplasmic reticulum stress-responsive cis-acting elements ERSE, UPRE and ERSE-II. J Biochem (Tokyo). 2004;136:343–50. https://doi.org/10.1093/jb/mvh122.

58. Plewka P, Szczesniak MW, Stepien A, Pasieka R, Wanowska E, Makalowska I, et al. Novel function of U7 SnRNA in the repression of HERV1/LTR12s and LincRNAs in human cells. Nucleic Acids Res. 2024;52:10504–19. https://doi.org/10.1093/nar/gkae738.

59. Higuchi T, Anzai K, Kobayashi S. U7 SnRNA acts as a transcriptional regulator interacting with an inverted CCAAT sequence-binding transcription factor NF-Y. Biochim Biophys Acta. 2008;1780:274–81. https://doi.org/10.1016/j.bbagen.2007.11.005.

60. Winter GE, Mayer A, Buckley DL, Erb MA, Roderick JE, Vittori S, et al. BET bromodomain proteins function as master transcription elongation factors independent of CDK9 recruitment. Mol Cell. 2017;67:5–e1819. https://doi.org/10.1016/j.molcel.2017.06.004.

61. Muhar M, Ebert A, Neumann T, Umkehrer C, Jude J, Wieshofer C, et al. SLAM-seq defines direct gene-regulatory functions of the BRD4-MYC axis. Science. 2018;360:800–5. https://doi.org/10.1126/science.aao2793.

62. Yu G, Wang L-G, He Q-Y. ChIPseeker: an R/Bioconductor package for chip peak annotation, comparison and visualization. Bioinforma Oxf Engl. 2015;31:2382–3. https://doi.org/10.1093/bioinformatics/btv145.

63. Lee S, Cook D, Lawrence M. Plyranges: a grammar of genomic data transformation. Genome Biol. 2019;20:4. https://doi.org/10.1186/s13059-018-1597-8.

64. Bao W, Kojima KK, Kohany O. Repbase update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:1–6. https://doi.org/10.1186/s13100-015-0041-9.

65. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics. 2009;25:1189–91. https://doi.org/10.1093/bioinformatics/btp033.

66. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25:1422–3. https://doi.org/10.1093/bioinformatics/btp163.

67. Zambelli F, Pesole G, Pavesi G. Pscan: finding over-represented transcription factor binding site motifs in sequences from co-regulated or co-expressed genes. Nucleic Acids Res. 2009;37:W247–52. https://doi.org/10.1093/nar/gkp464.

68. Grant CE, Bailey TL. XSTREME: comprehensive motif analysis of biological sequence datasets 2021:2021.09.02.458722. https://doi.org/10.1101/2021.09.02.458722

69. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

70. Liao Y, Smyth GK, Shi W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30. https://doi.org/10.1093/bioinformatics/btt656.

71. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4:1686. https://doi.org/10.21105/joss.01686.

72. Patil I. Visualizations with statistical details: the Ggstatsplot approach. J Open Source Softw. 2021;6:3167. https://doi.org/10.21105/joss.03167.

73. Völkel S, Stielow B, Finkernagel F, Stiewe T, Nist A, Suske G. Zinc finger independent Genome-Wide binding of Sp2 potentiates recruitment of Histone-Fold protein Nf-y distinguishing it from Sp1 and Sp3. PLOS Genet. 2015;11:e1005102. https://doi.org/10.1371/journal.pgen.1005102.

74. Zhao Y, Vartak SV, Conte A, Wang X, Garcia DA, Stevens E, et al. Stripe transcription factors provide accessibility to co-binding partners in mammalian genomes. Mol Cell. 2022;82:3398–e341111. https://doi.org/10.1016/j.molcel.2022.06.029.

Publisher's note