Supplementary Material:

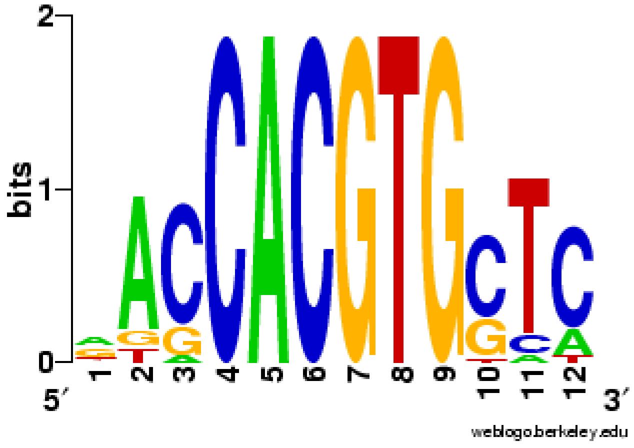
A c-Myc regulatory subnetwork from human transposable element sequences

Jiangrong Wang, Nathan J. Bowen, Leonardo Mariño-Ramírez and I. King Jordan

Supplementary Figure 1. **c-Myc binding site position frequency matrices (PFMs) and sequence logos.** A) V$CMYC_01 is the TRANSFAC PFM for c-Myc alone. B) V$CMYC_01 is the TRANSFAC PFM for the c-Myc heterodimer.

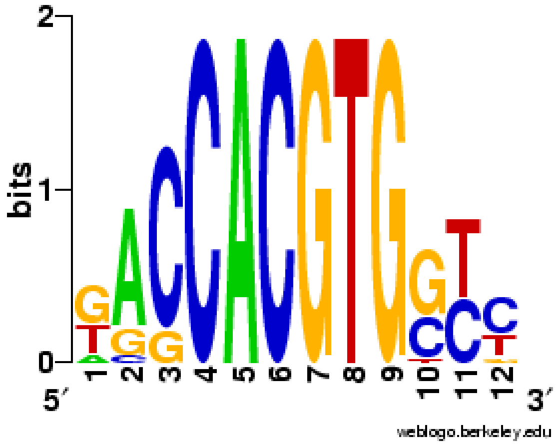
A) V$MYC_01

Position	**Residue Count**			
	A	**C**	**G**	**T**
1	7	1	7	3
2	16	0	2	2
3	1	16	4	0
4	0	21	0	0
5	21	0	0	0
6	0	21	0	0
7	0	0	21	0
8	0	0	0	21
9	0	0	21	0
10	0	13	6	1
11	1	2	0	15
12	3	11	0	1



B) V$MYC_02

Position	**Residue Count**			
	A	**C**	**G**	**T**
1	2	0	8	6
2	13	1	3	0
3	0	16	3	0
4	0	19	0	0
5	19	0	0	0
6	0	19	0	0
7	0	0	19	0
8	0	0	0	19
9	0	0	19	0
10	0	7	10	1
11	0	6	0	7
12	0	5	1	3



Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 1. **Number of TEs that contain c-Myc binding sites for each TE class/family.** Based on Clover run with P<0.001.

TE class/family[a]	Observed number[b]	Observed percent[c]	Expected percent[d]
L1	173	21.2%	21.9%
L2	115	14.1%	9.7%
LINE other	8	1.0%	1.6%
Alu	108	13.2%	28.1%
MIR	158	19.4%	13.9%
SINE other	9	1.1%	0.1%
DNA	90	11.0%	9.3%
LTR	155	19.0%	15.5%
Total	816	100.00%	100.0%

[a] Name of TE class or family. LINE other means LINE elements excluding L1 and L2; SINE other means SINE elements excluding Alu and MIR

[b] Observed number of TEs in each class/family

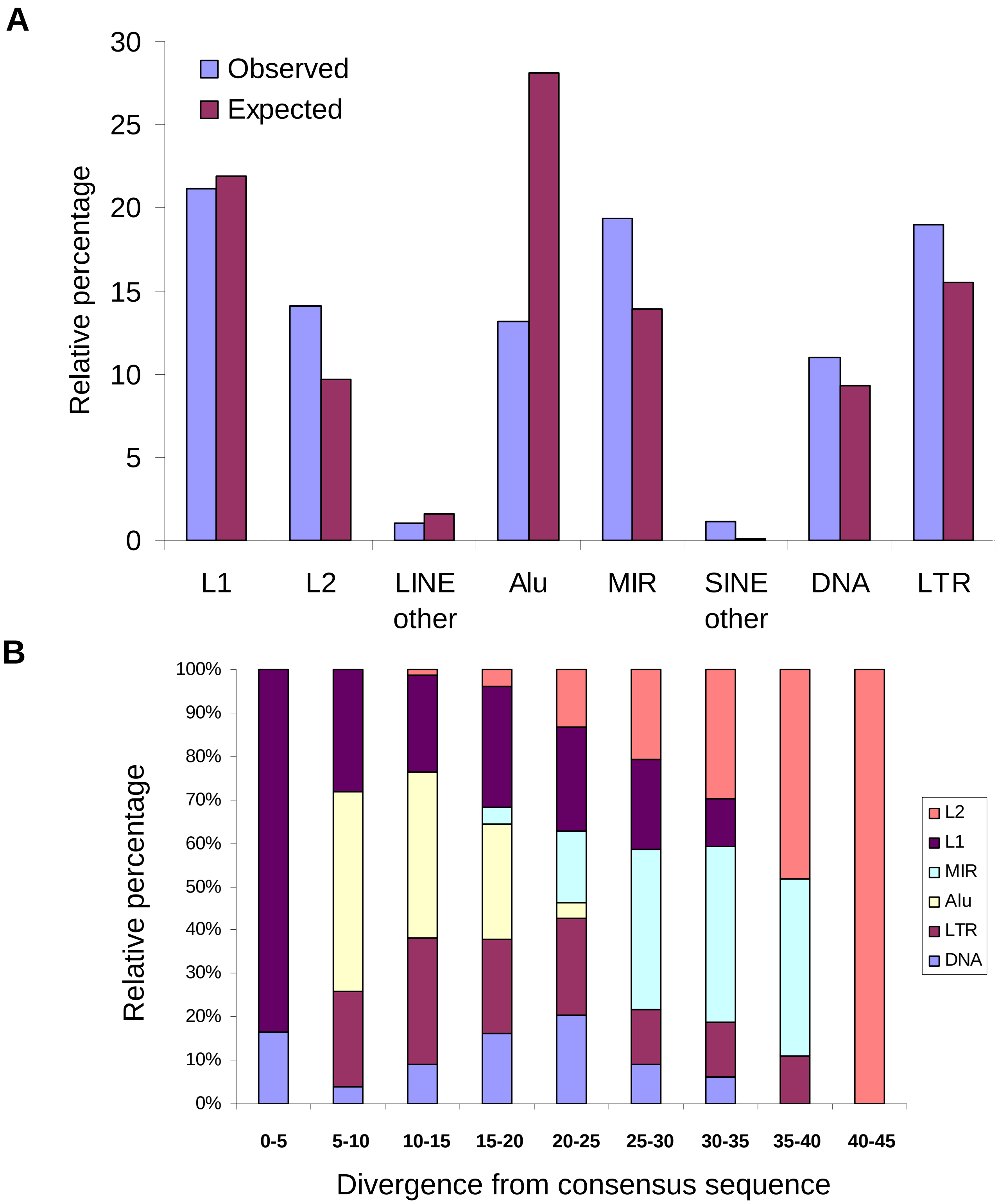
[c] Observed percent: the observed number of TEs in each class/family divided by the total observed number (816) of TEs containing c-Myc binding sites * 100

[d] Expected percent: the total number of TEs in each class/family in human genome divided by the total number of all TEs in human genome * 100

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Figure 2. **Family origins and relative ages for human TEs bound by c-Myc.** A) Observed versus expected relative percentages of c-Myc binding sites derived from different classes/families of human TEs. Based on Clover run with *P*<0.001. B) Percent divergence from subfamily consensus sequences for human TEs that are bound by c-Myc.



Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 2. **Over-expressed tissues of c-Myc and TE-derived target genes.** Based on permutations test (*P* value = 0.05).

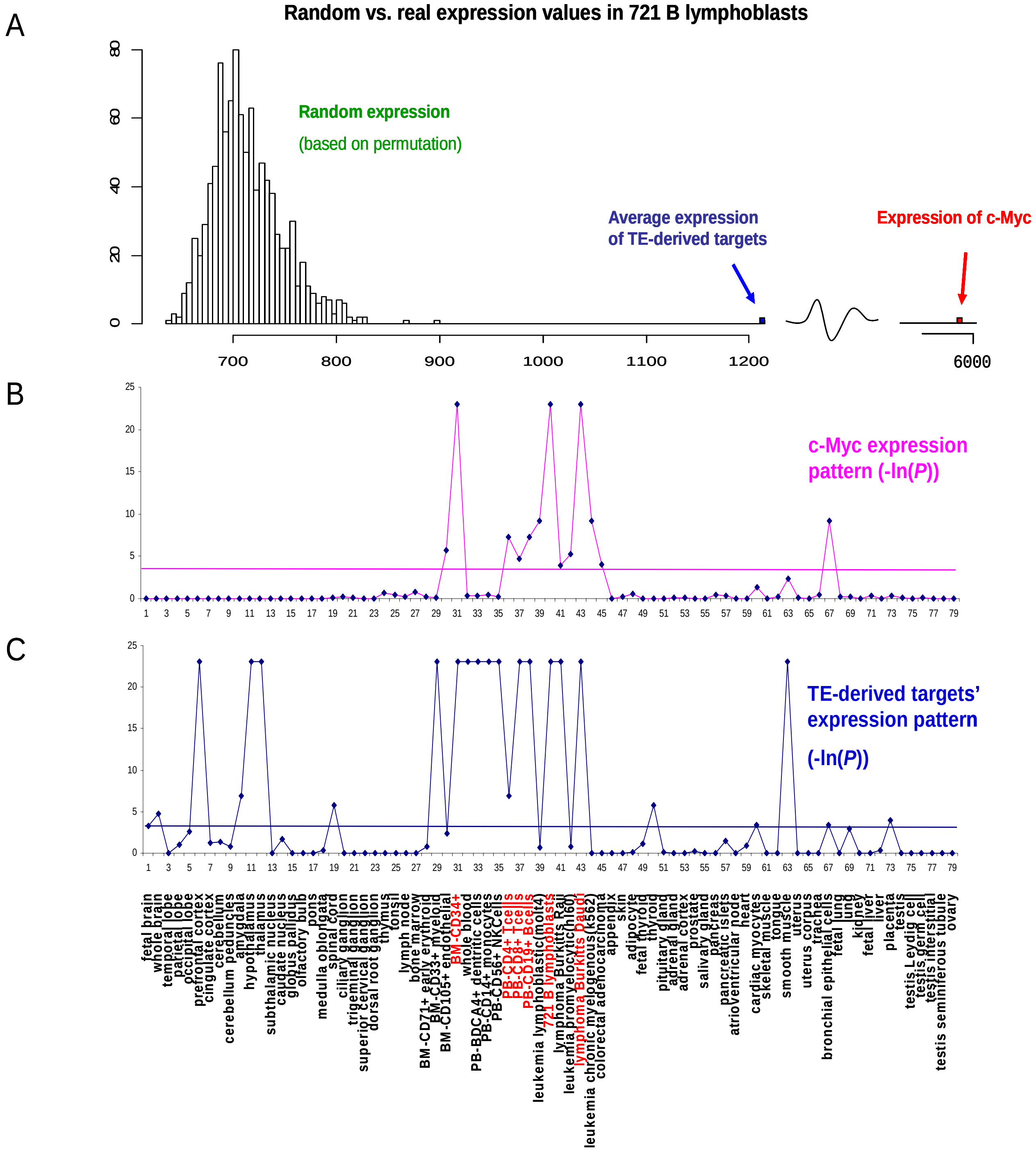
c-Myc over-expressed tissues	***P* value**
BM-CD105+ endothelial	0.0034
BM-CD34+	0
PB-CD4+ Tcells	0.0007
PB-CD8+ Tcells	0.0086
PB-CD19+ Bcells	0.0007
leukemia lymphoblastic(molt4)	0.0001
721 B lymphoblasts	0
leukemia promyelocytic(hl60)	0.0052
lymphoma Burkitts Daudi	0
leukemia chronic myelogenous(k562)	0.0001
colorectal adenocarcinoma	0.0174
bronchial epithelial cells	0.0001

TE-derived c-Myc target genes' over-expressed tissues	***P* value**
fetal brain	0.036
whole brain	0.009
prefrontal cortex	0
amygdala	0.001
hypothalamus	0
thalamus	0
spinal cord	0.003
BM-CD33+ myeloid	0
BM-CD34+	0
whole blood	0
PB-BDCA4+ dentritic cells	0
PB-CD14+ monocytes	0
PB-CD56+ NKCells	0
PB-CD4+ Tcells	0.001
PB-CD8+ Tcells	0
PB-CD19+ Bcells	0
721 B lymphoblasts	0
lymphoma Burkitts Raji	0
lymphoma Burkitts Daudi	0
thyroid	0.003
smooth muscle	0
placenta	0.019

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Figure 3. **Over-expression of c-Myc and TE-derived target genes.** Based on permutations test. A) Example of a comparison between randomly permuted 721 B lymphoblast expression levels and observed c-Myc and TE-derived c-Myc target genes 721 B lymphoblast expression levels. B) Results of the permutation test for c-Myc across all 79 tissues [-ln*P*-values]. C) Results of the permutation test for TE-derived c-Myc taget genes across all 79 tissues [-ln*P*-values]. Significance threshold levels are indicated for panels B & C.



Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 3. **Enriched annotations of target genes derived from specific TE-class/family.** Based on gene set enrichment analysis using MsigDB.

Genes with Alu derived c-Myc binding sites	
Functional annotations	***P* value**
Biopolymer metabolic process	4.78E-19
Nucleus	1.58E-17
Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	8.04E-14
Biopolymer modification	4.48E-12
Cytoplasm	4.92E-12
Protein metabolic process	1.05E-11
Protein modification process	1.91E-11
RNA metabolic process	3.22E-9
Post translational protein modification	4.45E-9
Stem cell embryonic up	3.08E-8
Genes with L1 derived c-Myc binding sites	
Functional annotations	***P* value**
Targets of MicroRNA TTGCACT, MIR-130A, MIR-301, MIR-130B	2.42E-9
Down-regulated at any timepoint following treatment of XPB/CS fibroblasts with 3 J/m^2 UVC	1.49E-8
Targets of MicroRNA GCACTTT, MIR-17-5P, MIR-20A, MIR-106A, MIR-106B, MIR-20B, MIR-519D	1.59E-7
Up-regulated in mouse hematopoietic late progenitors from adult bone marrow	1.09E-6
Kinase activity	1.16E-5
Nucleus	1.36E-5
Chromatin modification	1.98E-5
Protein serine threonine kinase activity	2.88E-5
Transferase activity transferring phosphorus containing groups	3.06E-5
Targets of MicroRNA GTGCAAT, MIR-25, MIR-32, MIR-92, MIR-363, MIR-367	3.25E-5
Targets of MicroRNA CTATGCA, MIR-153	4.14E-5
Membrane	4.64E-5

Supplementary Table 3 (Continued). **Enriched annotations of target genes derived from specific TE-class/family.** Based on gene set enrichment analysis using MsigDB.

Genes with L2 derived c-Myc binding sites	
Functional annotations	***P* value**
Lymphoma and immune response expression clusters (Module_126)	1.63E-7
Targets of MicroRNA AAGTCCA, MIR-422B, MIR-422A	5.44E-7
Cytoplasm	7.07E-7
Downregulated in correlation with overt Alzheimer’s Disease, in the CA1 region of the hippocampus	9.83E-7
Enriched in mouse neural stem cells, compared to differentiated brain and bone marrow cells	2.08E-6
Plasma membrane	2.55E-6
Targets of MicroRNA ACATTCC, MIR-1, MIR-206	4.7E-6
Membrane	5.85E-6
Genes with promoter regions around transcription start site containing the motif which matches annotation for MAZ: MYC-associated zinc finger protein	9.7E-6
Enriched in mouse embryonic stem cells, compared to differentiated brain and bone marrow cells	3.34E-5
Plasma membrane	4.25E-5
Neighborhood of MAX	4.47E-5
Transport	5.92E-5

Genes with MIR derived c-Myc binding sites	
Functional annotations	***P* value**
Cytoplasm	2.81E-6
Membrane	3.77E-6
Module_188 (Clinical annotaion: B lymphoma is enriched in this module)	5.25E-6
Receptor activity	1.6E-5
Biopolymer metabolic process	1.91E-5
Protein metabolic process	3.02E-5
Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.4E-5
Nucleus	3.91E-5

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 3 (Continued). **Enriched annotations of target genes derived from specific TE-class/family.** Based on gene set enrichment analysis using MsigDB.

Genes with DNA derived c-Myc binding sites	
Functional annotations	***P* value**
Down-regulated at any timepoint following treatment of XPB/TTD fibroblasts with 3 J/m^2 UVC	3.96E-7
Targets of MicroRNA AATGTGA, MIR-23A, MIR-23B	1.45E-5
Phosphoric monoester hydrolase activity	6.04E-5
Up-regulated in correlation with overt Alzheimer's Disease in the CA1 region of the hippocampus	7.91E-5
Downregulated by both Et-743 and Pt-650 in HCT116 cells	7.99E-5
Genes involved in MAPK signaling pathway	1.21E-4
Targets of MicroRNA TGCACTG, MIR-148A, MIR-152, MIR-148B	2.72E-4
Genes up-regulated in PMNs upon migration to skin lesions	4E-4
Double stranded RNA binding	4.67E-4
Enriched in mouse neural stem cells, compared to differentiated brain and bone marrow cells	4.92E-4

Genes with LTR derived c-Myc binding sites	
Functional annotations	***P* value**
Down-regulated at any timepoint following treatment of XPB/CS fibroblasts with 3 J/m^2 UVC	1.75E-8
Cytoplasm	8.3E-8
Targets of MicroRNA TGAATGT, MIR-181A, MIR-181B, MIR-181C, MIR-181D	9.77E-6
Targets of MicroRNA TAGCTTT, MIR-9	1.15E-5
Targets of MicroRNA AAGCACA, MIR-218	2.36E-5
Targets of MicroRNA CATTTCA, MIR-203	3.6E-5
Genes whose expression is induced in human mast cells by lipopolysaccharide treatment	4.77E-5
Targets of MicroRNA CAGTATT, MIR-200B, MIR-200C, MIR-429	6.2E-5
Module_136 (Clinical annotaion: B lymphoma is enriched in this module)	6.61E-5
Targets of MicroRNA ATGCAGT, MIR-217	1.08E-4
Targets of MicroRNA TGTGCCTT, MIR-506	1.5E-4
Genes with promoter regions around transcription start site containing the motif GGGCGGR which matches annotaion for SP1.	1.59E-4
Up-regulated in fibroblasts at 6 hours following treatment with iterferon-alpha	2.01E-4
Targets of MicroRNA TGTTTAC, MIR-30A-5p, MIR-30C, MIR-30D, MIR-30B, MIR-30E-5p	2.23E-4

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 4. **PCC between differentially expressed TE-derived target genes and c-Myc**

Gene Symbol & TE class/family	PCC	*t*	*P* value (Student's t-test)
PSMA5_Alu	0.63	6.03	1.54E-07
GORASP2_Alu	0.62	5.91	2.35E-07
XRCC5_Alu	0.61	5.79	3.71E-07
MAP4_Alu	0.43	3.51	925E-04
GNAI2_Alu	-0.63	-6.04	1.50E-07
ADD1_Alu	-0.54	-4.72	1.696E-05
SLC29A1_Alu	0.77	8.88	3.83E-12
DKC1_Alu	0.64	6.13	1.04E-07
VBP1_Alu	0.62	5.77	4.01E-07
KIAA0247_Alu	-0.58	-5.19	3.20E-06
OIP5_Alu	0.48	4.00	1.90E-04
AKAP1_Alu	0.79	9.60	2.84E-13
RBBP8_Alu	0.48	3.97	2.13E-04
SLC23A2_Alu	-0.40	-3.27	1.89E-03
DPM1_Alu	0.60	5.46	1.22E-06
LSM1_Alu	0.50	4.22	9.56E-05
JAK1_L1	-0.59	-5.43	1.38E-06
LRPPRC_L1	0.80	9.68	2.16E-13
TRIO_L1	-0.51	-4.37	5.77E-05
BAZ1B_L1	0.54	4.67	2.05E-05
MTAP_L1	0.70	7.19	2.06E-09
PRKCB1_L1	-0.52	-4.51	3.54E-05
RBBP8_L1	0.48	3.97	2.13E-04
ADARB1_L1	-0.46	-3.80	3.69E-04
ABCC4_L1	0.48	4.00	1.94E-04
MAN2A1_L1	-0.47	-3.96	2.19E-04
SPTLC2_L1	-0.38	-3.01	3.98E-03

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Table 4(Continued). **PCC between differentially expressed TE-derived target genes and c-Myc**

Gene Symbol & TE class/family	PCC	*t*	*P* value (Student's t-test)
SLC20A1_MIR	0.46	3.80	3.75E-04
ITPR1_MIR	-0.66	-6.50	2.65E-08
CTSO_MIR	-0.40	-3.21	2.25E-03
MCCC2_MIR	0.73	7.83	1.86E-10
PGK1_MIR	0.55	4.82	1.22E-05
UROS_MIR	0.60	5.58	8.11E-07
SLC7A1_MIR	0.66	6.47	2.95E-08
ERH_MIR	0.63	5.99	1.77E-07
AKT2_MIR	0.60	5.54	9.27E-07
HCLS1_MIR	-0.63	-5.89	2.53E-07
ZNF330_MIR	0.70	7.19	2.01E-09
NOLC1_L2	0.83	10.75	4.89E-15
PRKCB1_L2	-0.52	-4.51	3.54E-05
POLR2F_L2	0.53	4.58	2.82E-05
CD151_L2	0.23	1.75	8.66E-02
VAMP2_L2	-0.59	-5.41	1.47E-06
PPT1_DNA	0.59	5.31	2.15E-06
XRCC2_DNA	0.53	4.58	2.79E-05
DDX10_DNA	0.74	8.06	7.82E-11
HELZ_DNA	-0.59	-5.42	1.42E-06
AFG3L2_DNA	0.68	6.73	1.11E-08
ADARB1_DNA	-0.46	-3.80	3.68E-04
FCGR2A_LTR	-0.47	-3.86	3.04E-04
SELL_LTR	-0.59	-5.37	1.73E-06
SNX3_LTR	-0.45	-3.75	4.30E-04
LAIR1_LTR	-0.41	-3.32	1.60E-03

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

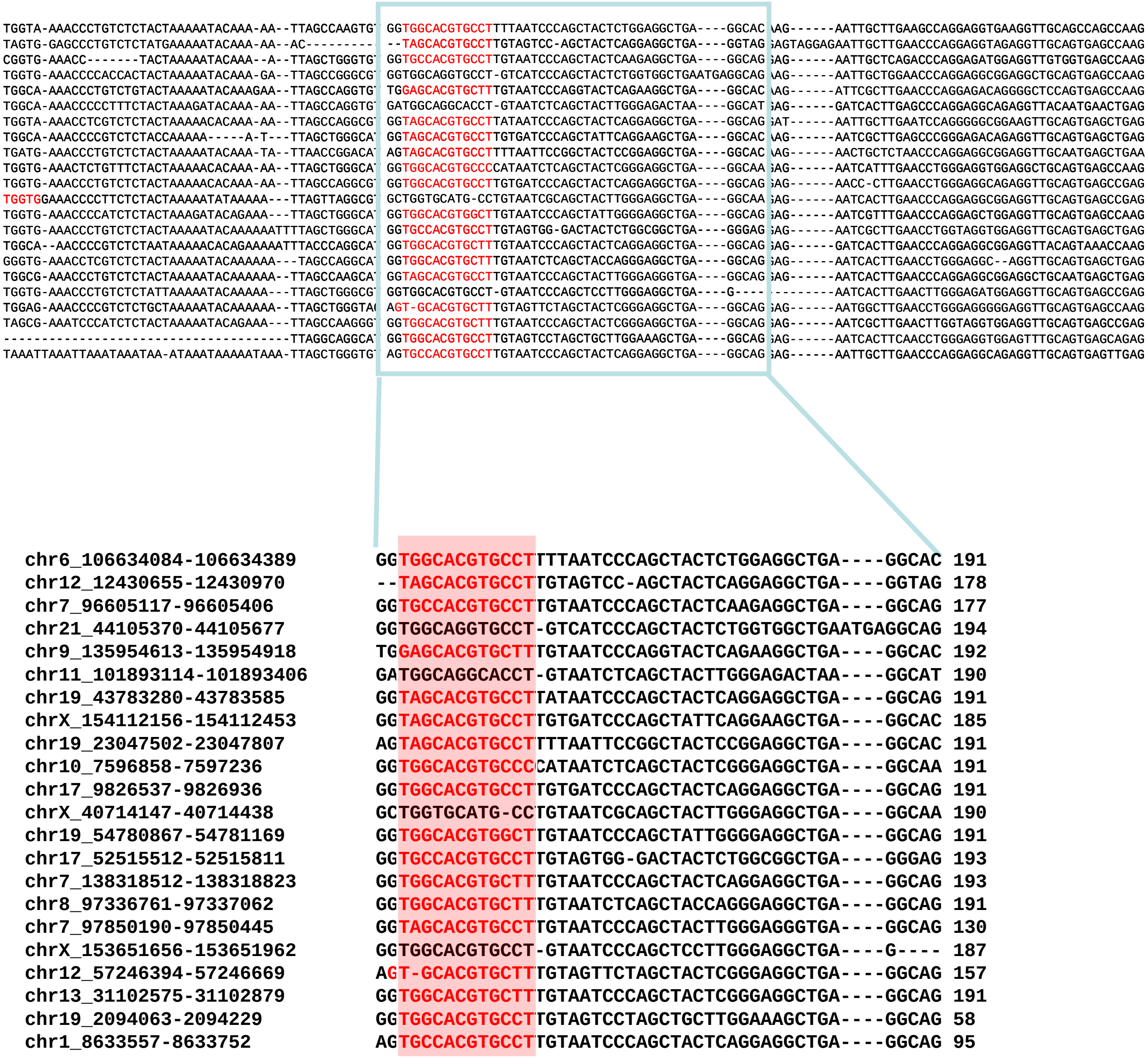
Supplementary Table 5. **Enriched annotations for differentially expressed target genes derived from specific TE class/family.** Based on gene set enrichment analysis using MsigDB.

Genes with MIR derived c-Myc binding sites		
Functional annotations	**Overlapping genes**	***P* value**
Up-regulated by UV-B light in normal human epidermal keratinocytes	ITPR1 PGK1 ZNF330	1.39E-5
G-alpha-q activates phospholipase C, resulting in calcium influx and increasing protein kinase C activity	ITPR1 AKT2	2.63E-5
Genes related to PIP3 signaling in B lymphocytes	ITPR1 AKT2	4.37E-5
Gi and Go proteins are members of the same family that transducer cell signals through both their alpha and beta subunits	ITPR1 AKT2	4.37E-5
Members of the BCR (B cell antigen receptor) signaling pathway	ITPR1 AKT2	7.17E-5
Genes with L1 derived c-Myc binding sites		
Functional annotations	**Overlapping genes**	***P* value**
Down-regulated following treatment of XPB/CS fibroblasts with 3J/m^2 UVC	LRPPRC TRIO ADARB1 MAN2A1 MTAP	4.89E-8
EGF pathway	PRKCB1 JAK1	2.44E-5
PDGF pathway	PRKCB1 JAK1	2.44E-5
Genes up-regulated in fatal/refractory diffuse B-cell lymphoma and down-regulated in cured diffuse B-cell lymphoma	LRPPRC PRKCB1	4.62E-5
Genes with Alu derived c-Myc binding sites		
Functional annotations	**Overlapping genes**	***P* value**
Down-regulated in response to glucose starvation	XRCC5 PSMA5 MAP4 OIP5 GORASP2	3.82E-9
Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	XRCC5 SLC29A1 SLC23A2 LSM1 RBBP8 DKC1	5.54E-6
Down-regulated in response to leusine starvation	XRCC5 PSMA5 SLC29A1	4.86E-5

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Supplementary Figure 4. **Multiple alignment of all AluSx elements containing c-Myc binding sites.**



```
TGGTA-AAACCCTGTCTCTACTAAAAATACAAA-AA--TTAGCCAAGTG GGTGGCACGTGCCTTTTAATCCCAGCTACTCTGGAGGCTGA----GGCAC AAG------AATTGCTTGAAGCCAGGAGGTGAAGGTTGCAGCCAGCCAAG
TAGTG-GAGCCCTGTCTCTATGAAAAATACAAA-AA--AC--------- --TAGCACGTGCCTTGTAGTCC-AGCTACTCAGGAGGCTGA----GGTAG GAGTAGGAGAATTGCTTGAACCCAGGAGGTAGAGGTTGCAGTGAGCCAAG
CGGTG-AAACC-------TACTAAAAATACAAA--A--TTAGCTGGGTG GGTGCCACGTGCCTTGTAATCCCAGCTACTCAAGAGGCTGA----GGCAG GAG------AATTGCTCAGACCCAGGAGATGGAGGTTGTGGTGAGCCAAG
TGGTG-AAACCCCACCACTACTAAAAATACAAA-GA--TTAGCCGGGCG GGTGGCAGGTGCCT-GTCATCCCAGCTACTCTGGTGGCTGAATGAGGCAG AAG------AATTGCTGGAACCCAGGAGGCGGAGGCTGCAGTGAGCCAAG
TGGCA-AAACCCTGTCTGTACTAAAAATACAAAGAA--TTAGCCAGGTG TGGAGCACGTGCTTTGTAATCCCAGGTACTCAGAAGGCTGA----GGCAC AAG------ATTCGCTTGAACCCAGGAGACAGGGGCTCCAGTGAGCCAAG
TGGCA-AAACCCCCTTTCTACTAAAGATACAAA-AA--TTAGCCAGGTG GATGGCAGGCACCT-GTAATCTCAGCTACTTGGGAGACTAA----GGCAT GAG------GATCACTTGAGCCCAGGAGGCAGAGGTTACAATGAACTGAG
TGGTA-AAACCTCGTCTCTACTAAAAACACAAA-AA--TTAGCCAGGCG GGTAGCACGTGCCTTATAATCCCAGCTACTCAGGAGGCTGA----GGCAG GAT------AATTGCTTGAATCCAGGGGGCGGAAGTTGCAGTGAGCTGAG
TGGCA-AAACCCCGTCTCTACCAAAAA-----A-T---TTAGCTGGGCA GGTAGCACGTGCCTTGTGATCCCAGCTATTCAGGAAGCTGA----GGCAC AAG------AATCGCTTGAGCCCGGGAGACAGAGGTTGCAGTGAGCTGAG
TGATG-AAACCCTGTCTCTACTAAAAATACAAA-TA--TTAACCGGACA AGTAGCACGTGCCTTTTAATTCCGGCTACTCCGGAGGCTGA----GGCAC AAG------AACTGCTCTAACCCAGGAGGCGGAGGTTGCAATGAGCTGAA
TGGTG-AAACTCTGTTTCTACTAAAAACACAAA-AA--TTAGCTGGGCA GGTGGCACGTGCCCCATAATCTCAGCTACTCGGGAGGCTGA----GGCAA GAG------AATCATTTGAACCTGGGAGGTGGAGGCTGCAGTGAGCCAAG
TGGTG-AAACCCTGTCTCTACTAAAAACACAAA-AA--TTAGCCAGGCG GGTGGCACGTGCCTTGTGATCCCAGCTACTCAGGAGGCTGA----GGCAG GAG------AACC-CTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAG
TGGTGGAAACCCCTTCTCTACTAAAAATATAAAAA---TTAGTTAGGCG GCTGGTGCATG-CCTGTAATCGCAGCTACTTGGGAGGCTGA----GGCAA GAG------AATCACTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCGAG
TGGTG-AAACCCCATCTCTACTAAAGATACAGAAA---TTAGCTGGGCA GGTGGCACGTGGCTTGTAATCCCAGCTATTGGGGAGGCTGA----GGCAG GAG------AATCGTTTGAACCCAGGAGCTGGAGGTTGCAGTGAGCCAAG
TGGTG-AAACCCTGTCTCTACTAAAAATACAAAAAATTTTAGCTGGGCA GGTGCCACGTGCCTTGTAGTGG-GACTACTCTGGCGGCTGA----GGGAG GAG------AATCGCTTGAACCTGGTAGGTGGAGGTTGCAGTGAGCTGAG
TGGCA--AACCCCGTCTCTAATAAAAACACAGAAAAATTTACCCAGGCA GGTGGCACGTGCTTTGTAATCCCAGCTACTCAGGAGGCTGA----GGCAG GAG------GATCACTTGAACCCAGGAGGCGGAGGTTACAGTAAACCAAG
GGGTG-AAACCTCGTCTCTACTAAAAATACAAAAAA---TAGCCAGGCA GGTGGCACGTGCTTTGTAATCTCAGCTACCAGGGAGGCTGA----GGCAG GAG------AATCACTTGAACCTGGGAGGC--AGGTTGCAGTGAGCTGAG
TGGCG-AAACCCTGTCTCTACTAAAAATACAAAAAA--TTAGCCAAGCA GGTAGCACGTGCCTTGTAATCCCAGCTACTTGGGAGGGTGA----GGCAG GAG------AATCACTTGAACCCAGGAGGCGGAGGCTGCAATGAGCTGAG
TGGTG-AAACCCTGTCTCTATTAAAAATACAAAAA---TTAGCTGGGCG GGTGGCACGTGCCT-GTAATCCCAGCTCCTTGGGAGGCTGA----G---- ---------AATCACTTGAACTTGGGAGATGGAGGTTGCAGTGAGCCGAG
TGGAG-AAACCCCGTCTCTGCTAAAAATACAAAAAA--TTAGCTGGGTA AGT-GCACGTGCTTTGTAGTTCTAGCTACTCGGGAGGCTGA----GGCAG GAG------AATGGCTTGAACCTGGGAGGGGGAGGTTGCAGTGAGCCAAG
TAGCG-AAATCCCATCTCTACTAAAAATACAGAAA---TTAGCCAAGGG GGTGGCACGTGCTTTGTAATCCCAGCTACTCGGGAGGCTGA----GGCAG GAG------AATCGCTTGAACTTGGTAGGTGGAGGTTGCAGTGAGCCGAG
--------------------------------------TTAGGCAGGCA GGTGGCACGTGCCTTGTAGTCCTAGCTGCTTGGAAAGCTGA----GGCAG GAG------AATCACTTCAACCTGGGAGGTGGAGTTTGCAGTGAGCAGAG
TAAATTAAATTAAATAAATAA-ATAAATAAAAATAAA-TTAGCTGGGTG AGTGCCACGTGCCTTGTAATCCCAGCTACTCAGGAGGCTGA----GGCAG GAG------AATTGCTTGAACCCAGGAGGCAGAGGTTGCAGTGAGTTGAG
```

```
chr6_106634084-106634389     GGTGGCACGTGCCTTTTAATCCCAGCTACTCTGGAGGCTGA----GGCAC 191
chr12_12430655-12430970      --TAGCACGTGCCTTGTAGTCC-AGCTACTCAGGAGGCTGA----GGTAG 178
chr7_96605117-96605406       GGTGCCACGTGCCTTGTAATCCCAGCTACTCAAGAGGCTGA----GGCAG 177
chr21_44105370-44105677      GGTGGCAGGTGCCT-GTCATCCCAGCTACTCTGGTGGCTGAATGAGGCAG 194
chr9_135954613-135954918     TGGAGCACGTGCTTTGTAATCCCAGGTACTCAGAAGGCTGA----GGCAC 192
chr11_101893114-101893406    GATGGCAGGCACCT-GTAATCTCAGCTACTTGGGAGACTAA----GGCAT 190
chr19_43783280-43783585      GGTAGCACGTGCCTTATAATCCCAGCTACTCAGGAGGCTGA----GGCAG 191
chrX_154112156-154112453     GGTAGCACGTGCCTTGTGATCCCAGCTATTCAGGAAGCTGA----GGCAC 185
chr19_23047502-23047807      AGTAGCACGTGCCTTTTAATTCCGGCTACTCCGGAGGCTGA----GGCAC 191
chr10_7596858-7597236        GGTGGCACGTGCCCCATAATCTCAGCTACTCGGGAGGCTGA----GGCAA 191
chr17_9826537-9826936        GGTGGCACGTGCCTTGTGATCCCAGCTACTCAGGAGGCTGA----GGCAG 191
chrX_40714147-40714438       GCTGGTGCATG-CCTGTAATCGCAGCTACTTGGGAGGCTGA----GGCAA 190
chr19_54780867-54781169      GGTGGCACGTGGCTTGTAATCCCAGCTATTGGGGAGGCTGA----GGCAG 191
chr17_52515512-52515811      GGTGCCACGTGCCTTGTAGTGG-GACTACTCTGGCGGCTGA----GGGAG 193
chr7_138318512-138318823     GGTGGCACGTGCTTTGTAATCCCAGCTACTCAGGAGGCTGA----GGCAG 193
chr8_97336761-97337062       GGTGGCACGTGCTTTGTAATCTCAGCTACCAGGGAGGCTGA----GGCAG 191
chr7_97850190-97850445       GGTAGCACGTGCCTTGTAATCCCAGCTACTTGGGAGGGTGA----GGCAG 130
chrX_153651656-153651962     GGTGGCACGTGCCT-GTAATCCCAGCTCCTTGGGAGGCTGA----G---- 187
chr12_57246394-57246669      AGT-GCACGTGCTTTGTAGTTCTAGCTACTCGGGAGGCTGA----GGCAG 157
chr13_31102575-31102879      GGTGGCACGTGCTTTGTAATCCCAGCTACTCGGGAGGCTGA----GGCAG 191
chr19_2094063-2094229        GGTGGCACGTGCCTTGTAGTCCTAGCTGCTTGGAAAGCTGA----GGCAG 58
chr1_8633557-8633752         AGTGCCACGTGCCTTGTAATCCCAGCTACTCAGGAGGCTGA----GGCAG 95
```

Supplement: TE-derived c-Myc regulatory subnetwork

Wang et al.

Additional files provided as Supplementary Information. Three BED format text files are provided so that the primary data can be visualized with the UCSC Genome Browser.

PotentialGeneTargets.bed Human genome reference sequence (hg18) coordinates and RefSeq identifiers for the 1,550 genes with proximal TE-derived c-Myc binding sites.

TEsBoundbycMyc_0.01.bed Human genome reference sequence (hg18) coordinates for 4,564 TE sequences experimentally characterized to be bound by c-Myc. 0.01 refers to the *P*-value threshold used by Clover to identify c-Myc binding site motifs.

TEsBoundbycMyc_0.001.bed Human genome reference sequence (hg18) coordinates for 816 TE sequences experimentally characterized to be bound by c-Myc. 0.001 refers to the *P*-value threshold used by Clover to identify c-Myc binding site motifs.