MYC overexpression leads to increased chromatin interactions at superenhancers and MYC binding sites

Yi Xiang See[1,2], Kaijing Chen[1,2], Melissa J. Fullwood[1,2,3,*]

[1]School of Biological Sciences, Nanyang Technological University, Singapore.

[2]Cancer Science Institute of Singapore, National University of Singapore.

[3]Institute of Molecular and Cell Biology, Agency for Science, Technology and Research (A*STAR), Singapore.

* Corresponding author (mfullwood@ntu.edu.sg)

Contents

Supplemental Tables

Supplemental Table S1 | SIQHiC sequencing statistics for Low *MYC* and High *MYC* cells.

	Low *MYC* R1	Low *MYC* R2	High *MYC* R1	High *MYC* R2
Total Reads	466,371,873	473,073,105	489,552,289	466,324,553
Total Mapped Reads	449,798,985	457,408,641	483,099,602	456,793,366
(Human hg19 ONLY)	96.4%	96.7%	98.7%	98.0%
Total Mapped Reads	463,092,913	469,974,590	486,308,738	462,613,290
(Human hg19 + Mouse mm10)	99.3%	99.3%	99.3%	99.2%
	1,097,259	1,085,736	826,538	927,729
Discarded Reads	0.2%	0.2%	0.2%	0.2%
	448,297,483	455,970,215	482,156,004	455,706,205
Human-Human Paired Reads	96.1%	96.4%	98.5%	97.7%
	13,698,171	12,918,639	3,326,196	5,979,356
Mouse-Mouse Paired Reads	2.9%	2.7%	0.7%	1.3%
Human-Human Paired Reads	448,297,483	455,970,215	482,156,004	455,706,205
	310,510,965	322,597,046	302,156,580	328,864,586
Human Unique Reads	69.3%	70.7%	62.7%	72.2%
Intra-Fragment Reads	721,931	383,939	1,659,119	455,654
Below MAPQ30 Threshold	43,677,206	47,649,671	39,799,686	47,840,823
	266,111,828	274,563,436	260,697,775	280,568,109
Hi-C Contacts	85.7%	85.1%	86.3%	85.3%
Inter-chromosomal	22,188,336	29,460,191	11,484,151	25,330,251
	243,923,492	245,103,245	249,213,624	255,237,858
Intra-chromosomal	91.7%	89.3%	95.6%	91.0%
Short (<20kb)	131,525,754	106,418,212	183,310,972	113,572,967
Long (>20kb)	112,397,575	138,684,836	65,902,548	141,664,640
Mouse-Mouse Paired Reads	13,698,171	12,918,639	3,326,196	5,979,356
	9,591,745	9,195,144	2,137,464	4,361,700
Mouse Unique Reads	70.0%	71.2%	64.3%	72.9%
Intra-Fragment Reads	34,711	16,340	20,531	9,946
Below MAPQ30 Threshold	1,584,670	1,644,047	342,276	757,343
	7,972,364	7,534,757	1,774,657	3,594,411
Hi-C Contacts	83.1%	81.9%	83.0%	82.4%
Inter-chromosomal	841,107	819,222	143,395	380,435
	7,131,257	6,715,535	1,631,262	3,213,976
Intra-chromosomal	89.4%	89.1%	91.9%	89.4%
Short (<20kb)	3,898,627	3,186,260	1,153,069	1,574,659
Long (>20kb)	3,232,626	3,529,269	478,192	1,639,311

Supplemental Table S2 | Calculation of Low *MYC* and High *MYC* SIQHiC Ratio.

	Total reads	Human Contacts	Mouse Contacts	Human Contacts / Mouse Contacts (HMR)		HMR Fold Change vs. Low *MYC*		SIQHiC Ratio	
Low *MYC* Replicate 1	466,371,873	266,111,828	7,972,364	33.379	34.866	1.000	1.000	0.227	0.346
Low *MYC* Replicate 2	473,073,105	274,563,436	7,534,757	36.440		1.000		0.467	
High *MYC* Replicate 1	489,552,289	260,697,775	1,774,657	146.900	100.812	4.401	2.891	1.000	1.000
High *MYC* Replicate 2	466,324,553	280,568,109	3,594,411	78.057		2.142		1.000	

Supplemental Table S3 | SIQHiC sequencing statistics for siControl and siCTCF cells.

	siControl_R1	siControl_R2	siCTCF_R1	siCTCF_R2
Total Reads	486,430,445	471,799,212	494,076,632	464,953,616
Total Mapped Reads	478,571,646	465,466,367	478,635,584	452,965,855
(Human hg19 ONLY)	98.4%	98.7%	96.9%	97.4%
Total Mapped Reads	484,412,681	469,633,466	492,190,679	463,577,777
(Human hg19 + Mouse mm10)	99.6%	99.5%	99.6%	99.7%
	8,946,230	8,676,392	9,467,542	8,385,251
Discarded Reads	1.8%	1.8%	1.9%	1.8%
	456,579,962	446,730,331	444,058,460	421,007,523
Human-Human Paired Reads	94.3%	95.1%	90.2%	90.8%
	18,886,489	14,226,743	38,664,677	34,185,003
Mouse-Mouse Paired Reads	3.9%	3.0%	7.9%	7.4%
Human-Human Paired Reads	456,579,962	446,730,331	444,058,460	421,007,523
	350,524,868	349,173,808	339,280,375	324,857,931
Human Unique Reads	76.8%	78.2%	76.4%	77.2%
Intra-Fragment Reads	1,432,009	1,406,054	791,344	867,189
Below MAPQ30 Threshold	48,226,797	47,048,482	50,259,186	47,455,721
	300,866,062	300,719,272	288,229,845	276,535,021
Hi-C Contacts	85.8%	86.1%	85.0%	85.1%
Inter-chromosomal	20,419,370	20,936,012	24,006,698	22,347,537
	280,446,692	279,783,260	264,223,147	254,187,484
Intra-chromosomal	93.2%	93.0%	91.7%	91.9%
Short (<20kb)	174,860,580	172,296,248	132,033,769	127,643,655
Long (>20kb)	105,584,341	107,485,122	132,188,499	126,543,115
Mouse-Mouse Paired Reads	18,886,489	14,226,743	38,664,677	34,185,003
	14,686,344	11,244,958	29,608,422	26,428,013
Mouse Unique Reads	77.8%	79.0%	76.6%	77.3%
Intra-Fragment Reads	77,091	59,623	84,404	89,673
Below MAPQ30 Threshold	2,291,250	1,723,683	5,061,344	4,473,166
	12,318,003	9,461,652	24,462,674	21,865,174
Hi-C Contacts	83.9%	84.1%	82.6%	82.7%
Inter-chromosomal	792,851	629,814	1,868,727	1,664,169
	11,525,152	8,831,838	22,593,947	20,201,005
Intra-chromosomal	93.6%	93.3%	92.4%	92.4%
Short (<20kb)	7,247,274	5,494,816	11,458,406	10,285,651
Long (>20kb)	4,277,798	3,336,949	11,135,460	10,285,651

Supplemental Table S4 | Calculation of siControl and siCTCF SIQHiC Ratio.

	Total reads	Human Contacts	Mouse Contacts	Human Contacts / Mouse Contacts		HMR Fold Change vs. siCTCF		SIQHiC Ratio	
siControl Replicate 1	486,430,444	300,866,062	12,318,003	24.425	27.621	2.073	2.266	1.000	1.000
siControl Replicate 2	471,799,212	300,719,272	9,461,652	31.783		2.513		1.000	
siCTCF Replicate 1	494,076,632	288,229,845	24,462,674	11.782	12.191	1.000	1.000	0.482	0.441
siCTCF Replicate 2	464,953,616	276,535,021	21,865,174	12.647		1.000		0.398	

Supplemental Table S5 | RT-qPCR and 4C primer sequences.

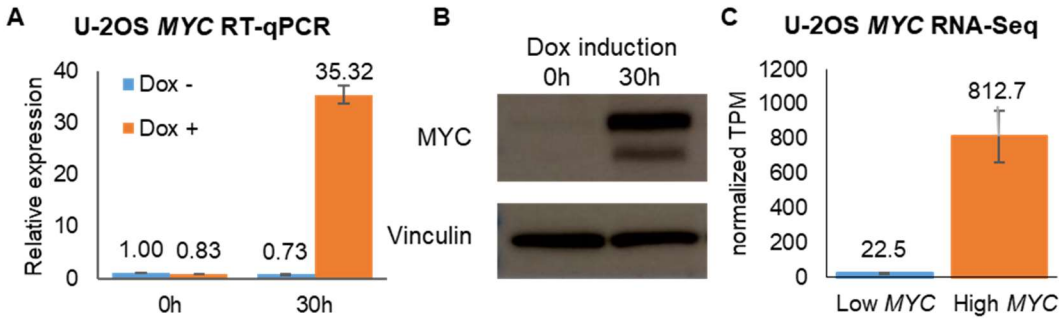
RT-qPCR Primers	Sequence
TBP _F	TAATCCCAAGCGATTTGCTG
TBP _R	CAGTTGTCCGTGGCTCTCTT
MYC _F	CACCAGCAGCGACTCTGA
MYC _R	GATCCAGACTCTGACCTTTTGC
CTCF _F	CAAACACACCCACGAGAA
CTCF _R	CTCCAGTATGAGAGCGAATG

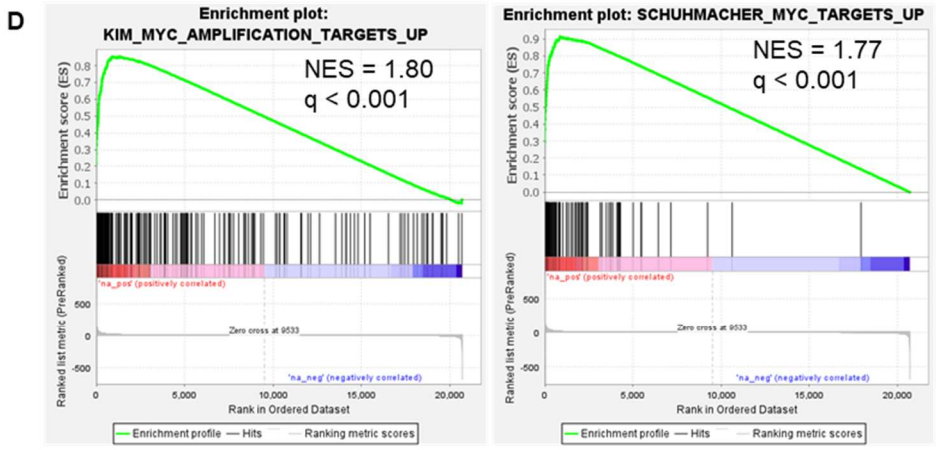
4C Viewpoint	4C Primer	Sequence
	Outer F	TTAGCCACGAAGTGTCCTGC
	Outer R	AGGAGAACTGTTTTTCATTGCCATG
	Nested F	AATGATACGGCGACCACCGAGATCTACAC**XXXXXXXX**TCG TCGGCAGCGTCAGATGTGTATAAGAGACAGGGTGGAGTT TCCTTCCCTGT
Chr2_PROC_SE	Nested R	CAAGCAGAAGACGGCATACGAGAT**YYYYYYYY**GTCTCGTG GGCTCGGAGATGTGTATAAGAGACAGTGATTGGATGCAG CACTTGG

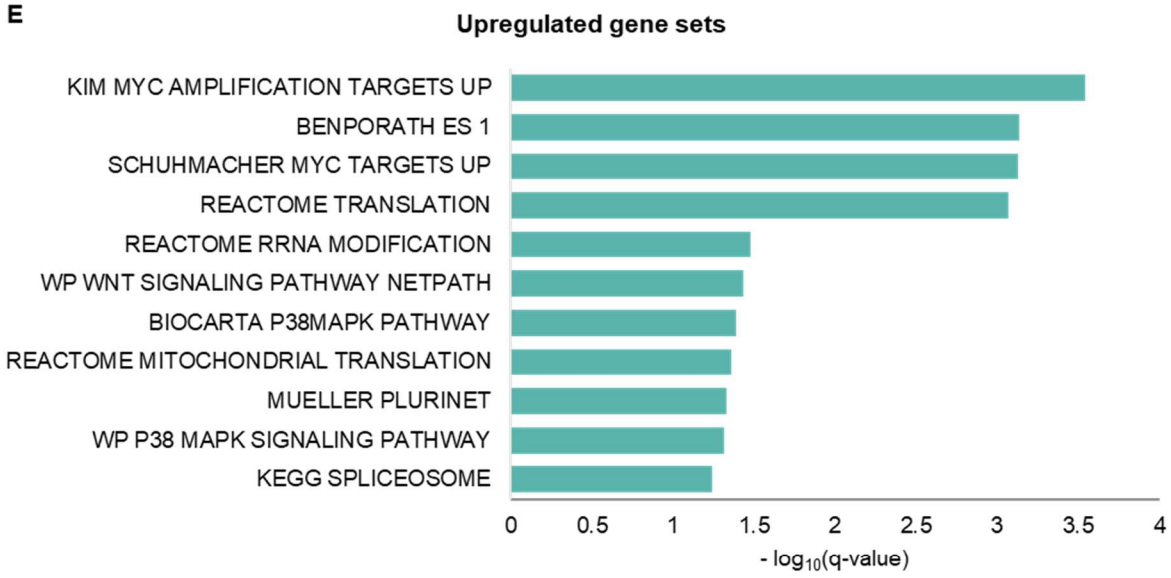
***XXXXXXXX**: Substitute with apropriate Illumina Nextera i5 index sequence

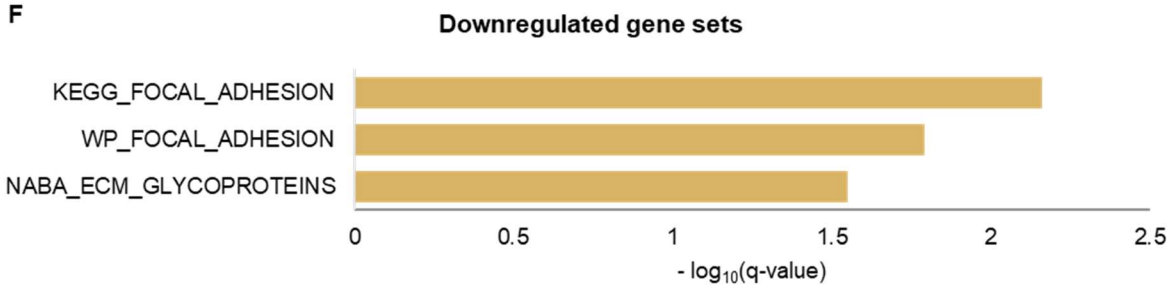
***YYYYYYYY**: Substitute with apropriate Illumina Nextera i7 index sequence

Supplemental Figures

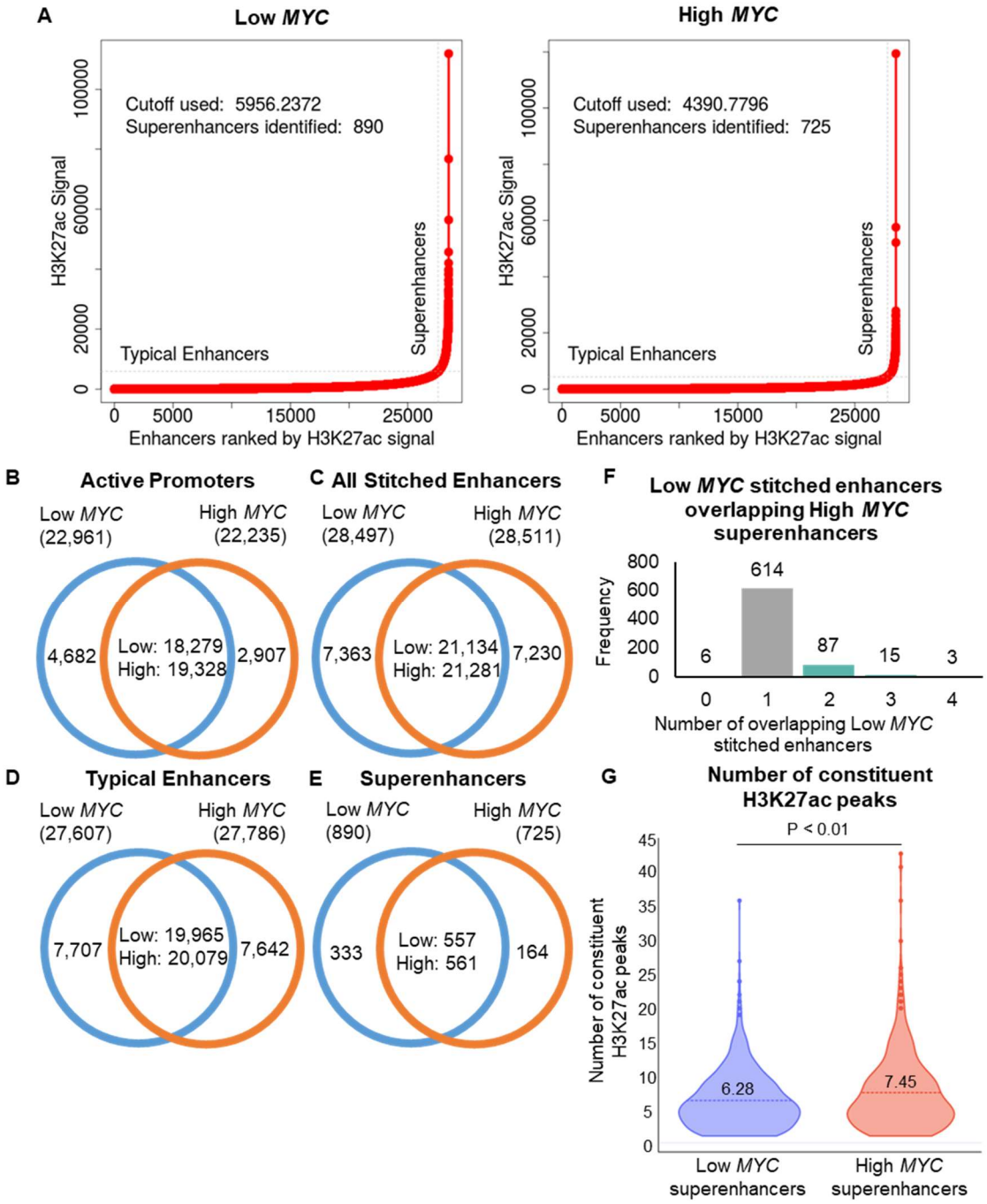






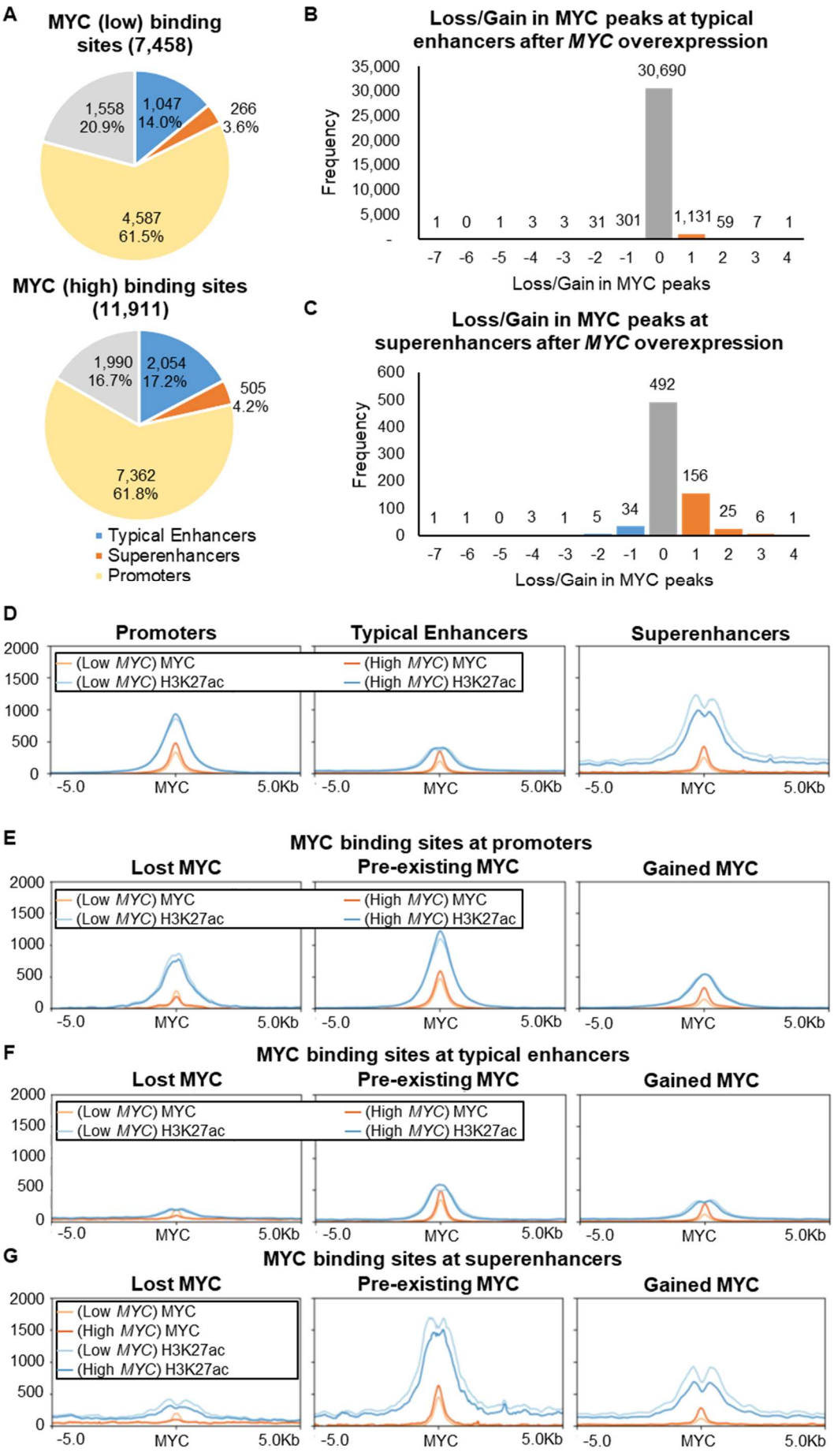


Supplemental Figure S1 | *MYC* expression and regulated genes. (**A**) *MYC* gene expression in U2OS cells before and after doxycycline (Dox) induction, normalized against *TBP* gene. Error bars show standard error across 3 biological replicates. (**B**) Western blot showing MYC protein expression with and without doxycycline (Dox) induction. Vinculin is shown as a loading control. (**C**) RNA-Seq normalized transcripts per million (TPM) expression of *MYC* in U2OS Low *MYC* and High *MYC* cells. (**D**) Gene set enrichment analysis (GSEA) plots of 2 previously published MYC regulated gene sets. (**E**) Top upregulated gene sets after *MYC* overexpression. (**F**) Top downregulated gene sets after *MYC* overexpression.

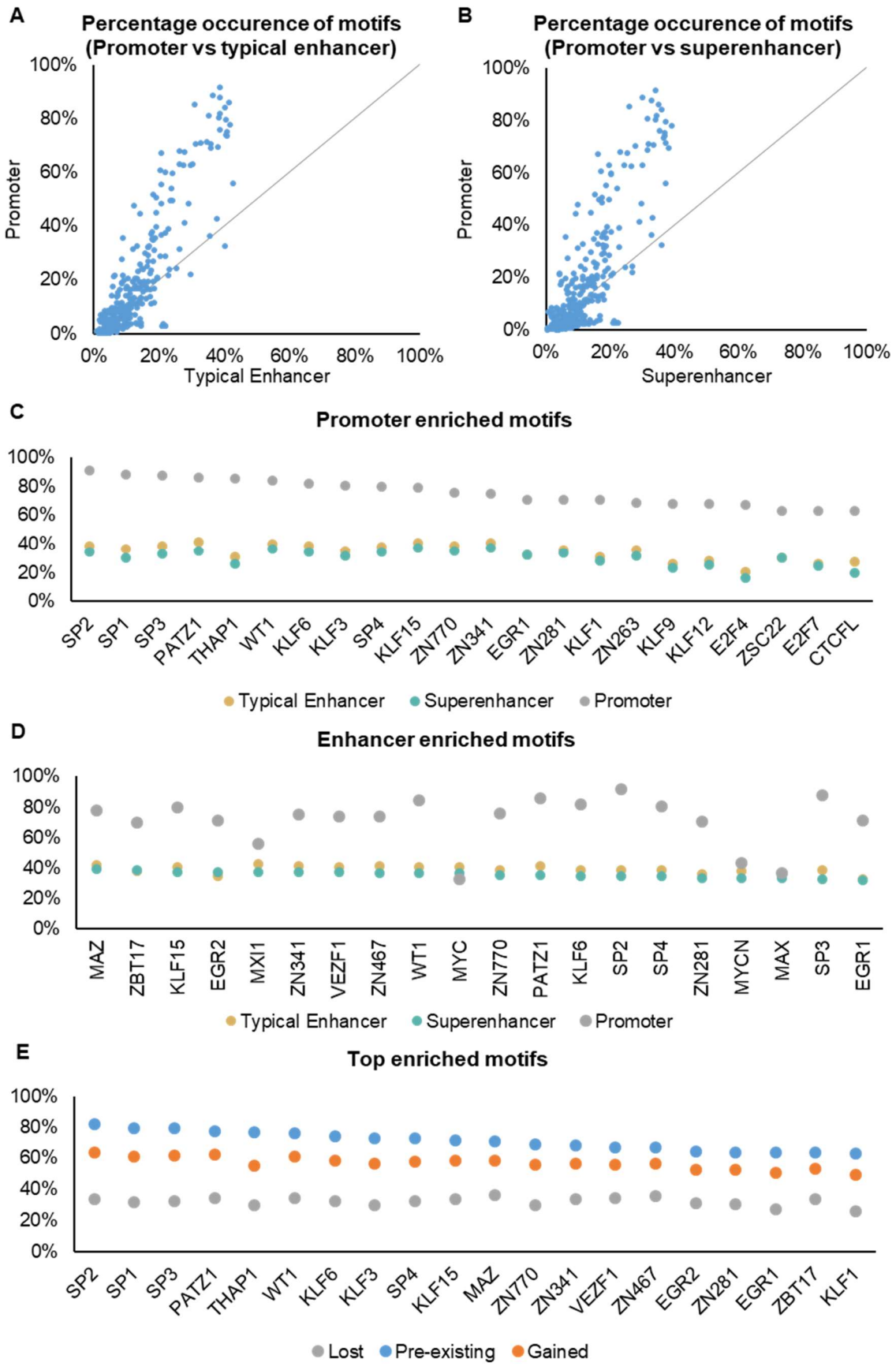


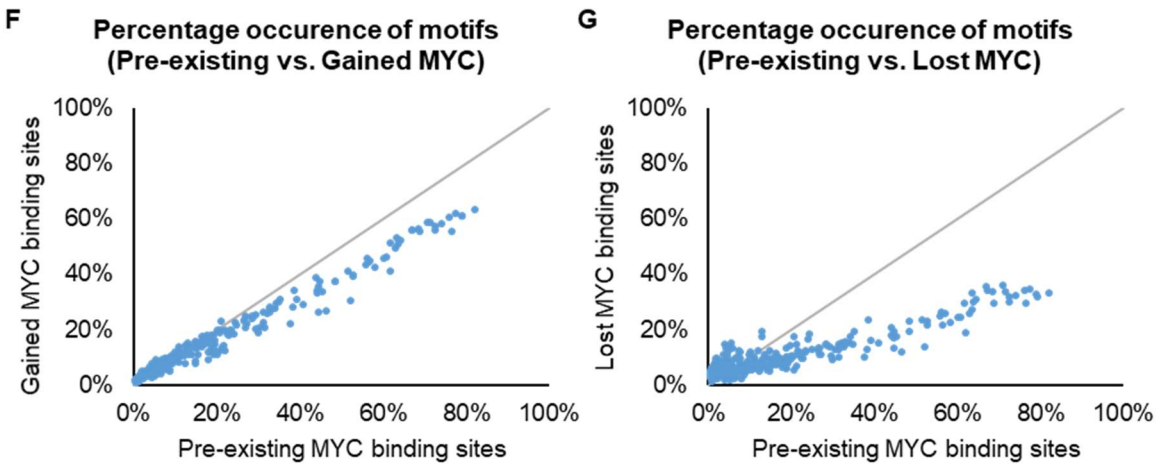
Supplemental Figure S2 | Enhancer and superenhancers. (**A**) Identification of Low *MYC* (left) and High *MYC* (right) superenhancers by ranking stitched enhancers based on H3K27ac (see Methods). (**B-E**) Overlap of (**B**) active promoter peaks, (**C**) all stitched enhancers, (**D**) typical enhancers and (**E**) superenhancers between Low *MYC* and High *MYC* cells. (**F**) Number of Low *MYC* stitched enhancers overlapping High *MYC* superenhancers. (**G**) Violin plot showing number of constituent H3K27ac

peaks merged with Low MYC and High MYC superenhancers. P values were calculated using two-tailed t-test assuming unequal variances.

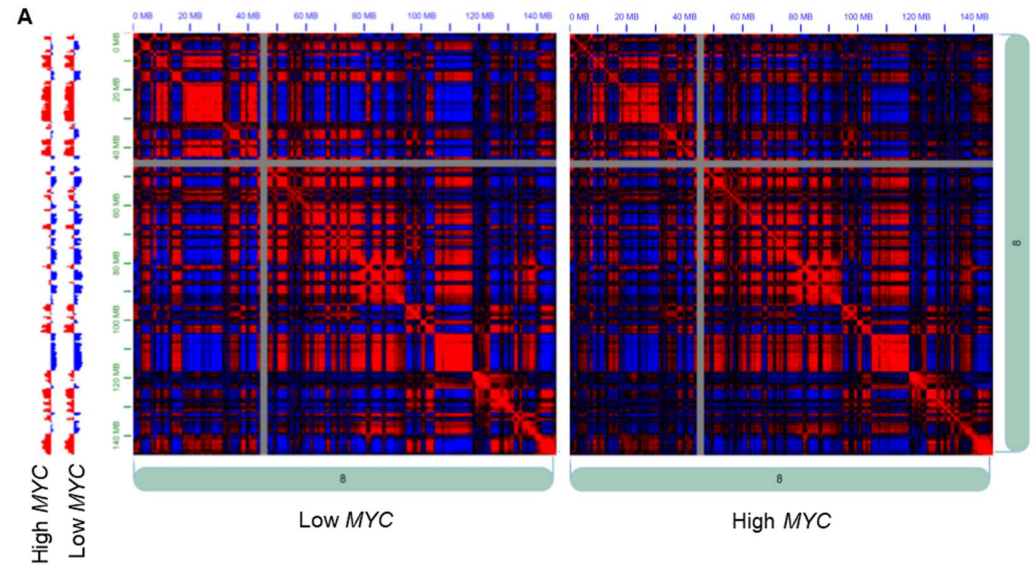

A
MYC (low) binding sites (7,458)
1,558 20.9%
1,047 14.0%
266 3.6%
4,587 61.5%
MYC (high) binding sites (11,911)
1,990 16.7%
2,054 17.2%
505 4.2%
7,362 61.8%
Typical Enhancers
Superenhancers
Promoters
B
Loss/Gain in MYC peaks at typical enhancers after MYC overexpression
Frequency
35,000
30,000
25,000
20,000
15,000
10,000
5,000
-
30,690
1 0 1 3 3 31 301 1,131 59 7 1
-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4
Loss/Gain in MYC peaks
C
Loss/Gain in MYC peaks at superenhancers after MYC overexpression
Frequency
600
500
400
300
200
100
0
492
156
1 1 0 3 1 5 34 25 6 1
-7 -6 -5 -4 -3 -2 -1 0 1 2 3 4
Loss/Gain in MYC peaks
D
Promoters
Typical Enhancers
Superenhancers
2000
1500
1000
500
0
(Low MYC) MYC
(Low MYC) H3K27ac
(High MYC) MYC
(High MYC) H3K27ac
-5.0
MYC
5.0Kb
E
MYC binding sites at promoters
Lost MYC
Pre-existing MYC
Gained MYC
F
MYC binding sites at typical enhancers
Lost MYC
Pre-existing MYC
Gained MYC
G
MYC binding sites at superenhancers
Lost MYC
Pre-existing MYC
Gained MYC
(Low MYC) MYC
(High MYC) MYC
(Low MYC) H3K27ac
(High MYC) H3K27ac

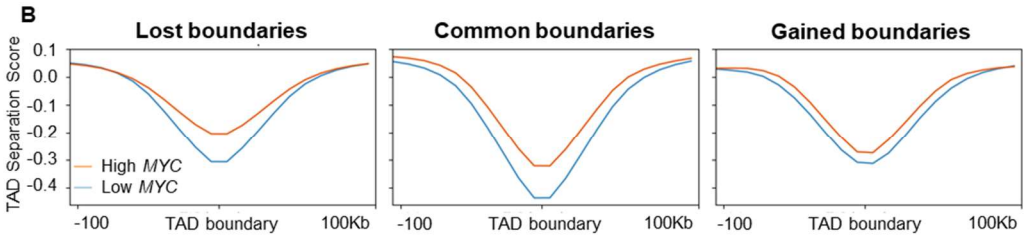
Supplemental Figure S3 | MYC binding sites at promoters and enhancers. (**A**) Distribution of MYC binding sites at cis-regulatory elements in Low *MYC* (left) and High *MYC* (right) cells. (**B-C**) Gain and loss of MYC peaks at (**B**) typical enhancers and (**C**) superenhancers after *MYC* overexpression. (**D**) H3K27ac and MYC ChIP-seq signal at MYC binding sites on promoters, typical enhancers and superenhancers. (**E-G**) H3K27ac and MYC ChIP-seq signal at lost, pre-existing and gained MYC binding sites on (**E**) typical enhancers, (**F**) superenhancers and (**G**) promoters.

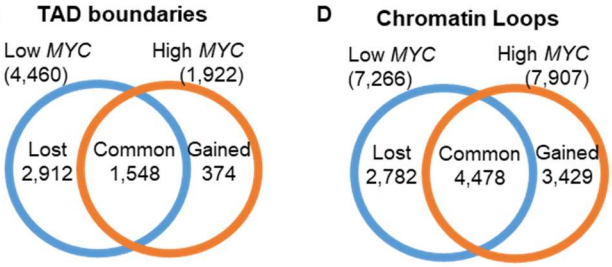

A
Percentage occurence of motifs
(Promoter vs typical enhancer)
Promoter
Typical Enhancer
B
Percentage occurence of motifs
(Promoter vs superenhancer)
Promoter
Superenhancer
C
Promoter enriched motifs
SP2 SP1 SP3 PATZ1 THAP1 WT1 KLF6 KLF3 SP4 KLF15 ZN770 ZN341 EGR1 ZN281 KLF1 ZN263 KLF9 KLF12 E2F4 ZSC22 E2F7 CTCFL
Typical Enhancer
Superenhancer
Promoter
D
Enhancer enriched motifs
MAZ ZBT17 KLF15 EGR2 MXI1 ZN341 VEZF1 ZN467 WT1 MYC ZN770 PATZ1 KLF6 SP2 SP4 ZN281 MYCN MAX SP3 EGR1
Typical Enhancer
Superenhancer
Promoter
E
Top enriched motifs
SP2 SP1 SP3 PATZ1 THAP1 WT1 KLF6 KLF3 SP4 KLF15 MAZ ZN770 ZN341 VEZF1 ZN467 EGR2 ZN281 EGR1 ZBT17 KLF1
Lost
Pre-existing
Gained

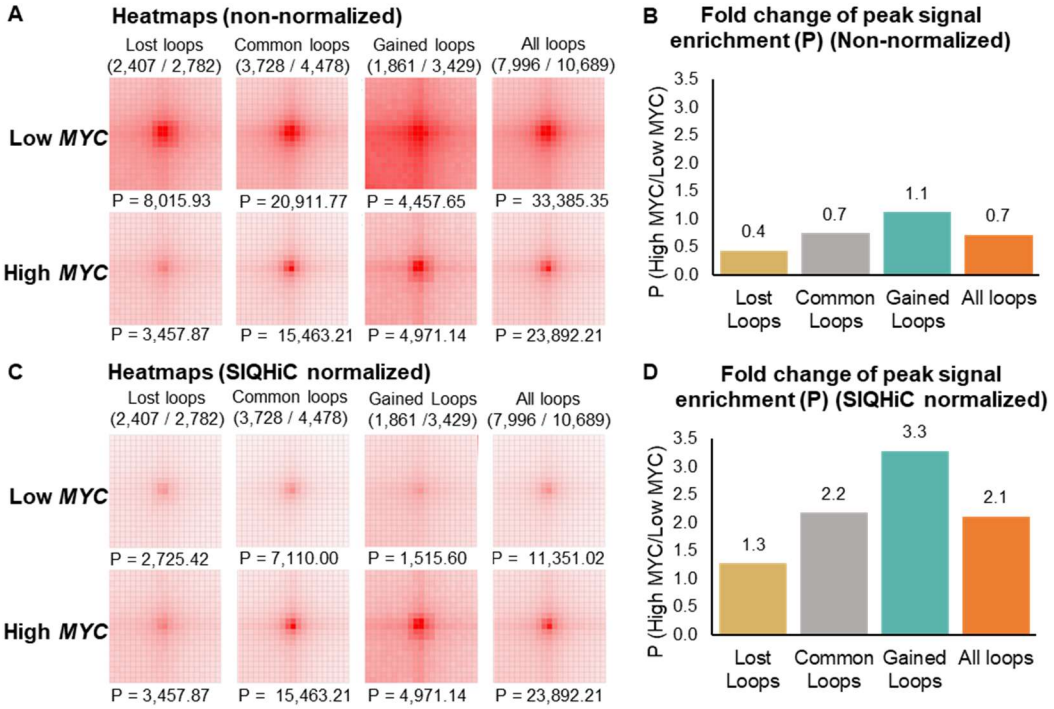


Supplemental Figure S4 | Motif occurrence at MYC binding sites. (**A-B**) Percentage occurrence of HOCOMOCO transcription factor motifs at promoter MYC binding sites versus (**A**) typical enhancer and (**B**) superenhancer MYC binding sites. Each point represents 1 motif. The grey line shows equal percentage occurrence. (**C-D**) Top enriched HOCOMOCO transcription factor motifs occurring at (**C**) promoter and (**D**) enhancer MYC binding sites. (**E**) Top enriched motifs at lost, pre-existing and gained MYC binding sites. (**F-G**) Percentage occurrence of HOCOMOCO transcription factor motifs at (**F**) gained and (**G**) lost MYC binding sites versus pre-existing MYC binding sites. Each point represents 1 motif. The grey line shows equal percentage occurrence.

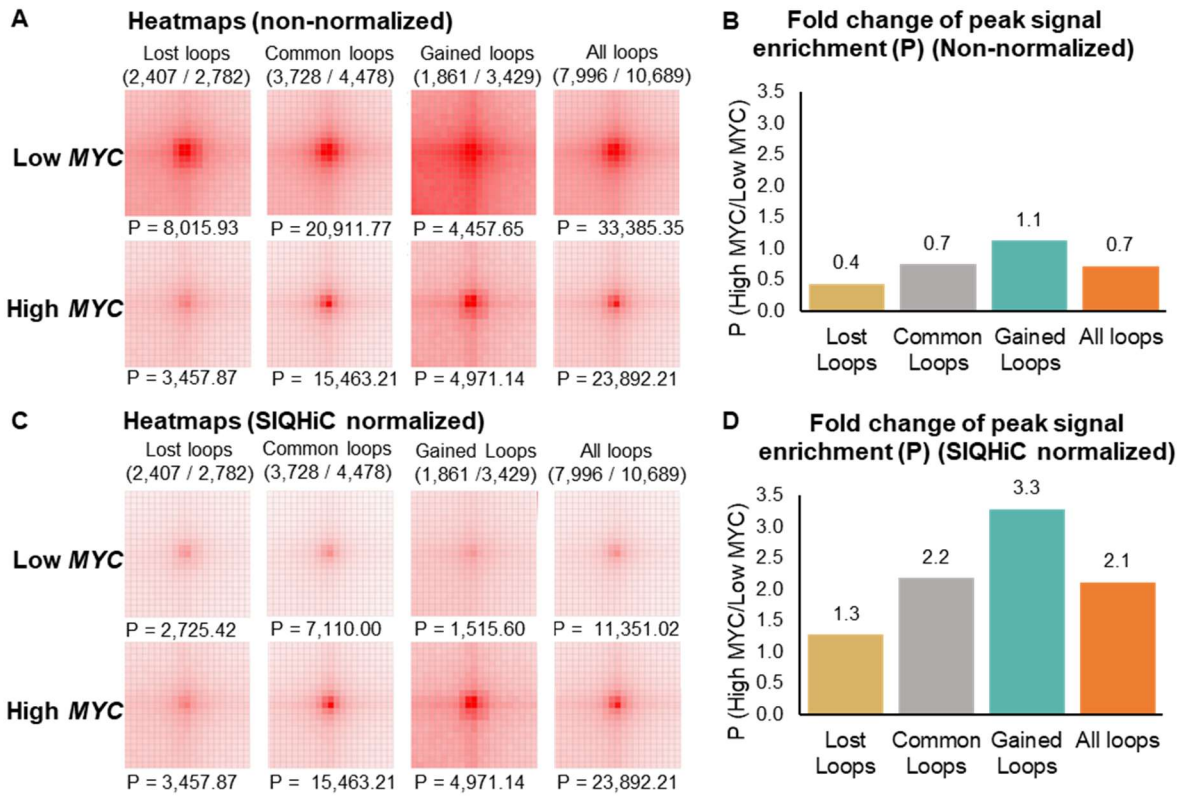

A
High MYC
Low MYC
Low MYC
High MYC


B
Lost boundaries
Common boundaries
Gained boundaries
TAD Separation Score
0.1
0.0
-0.1
-0.2
-0.3
-0.4
High MYC
Low MYC
-100
TAD boundary
100Kb

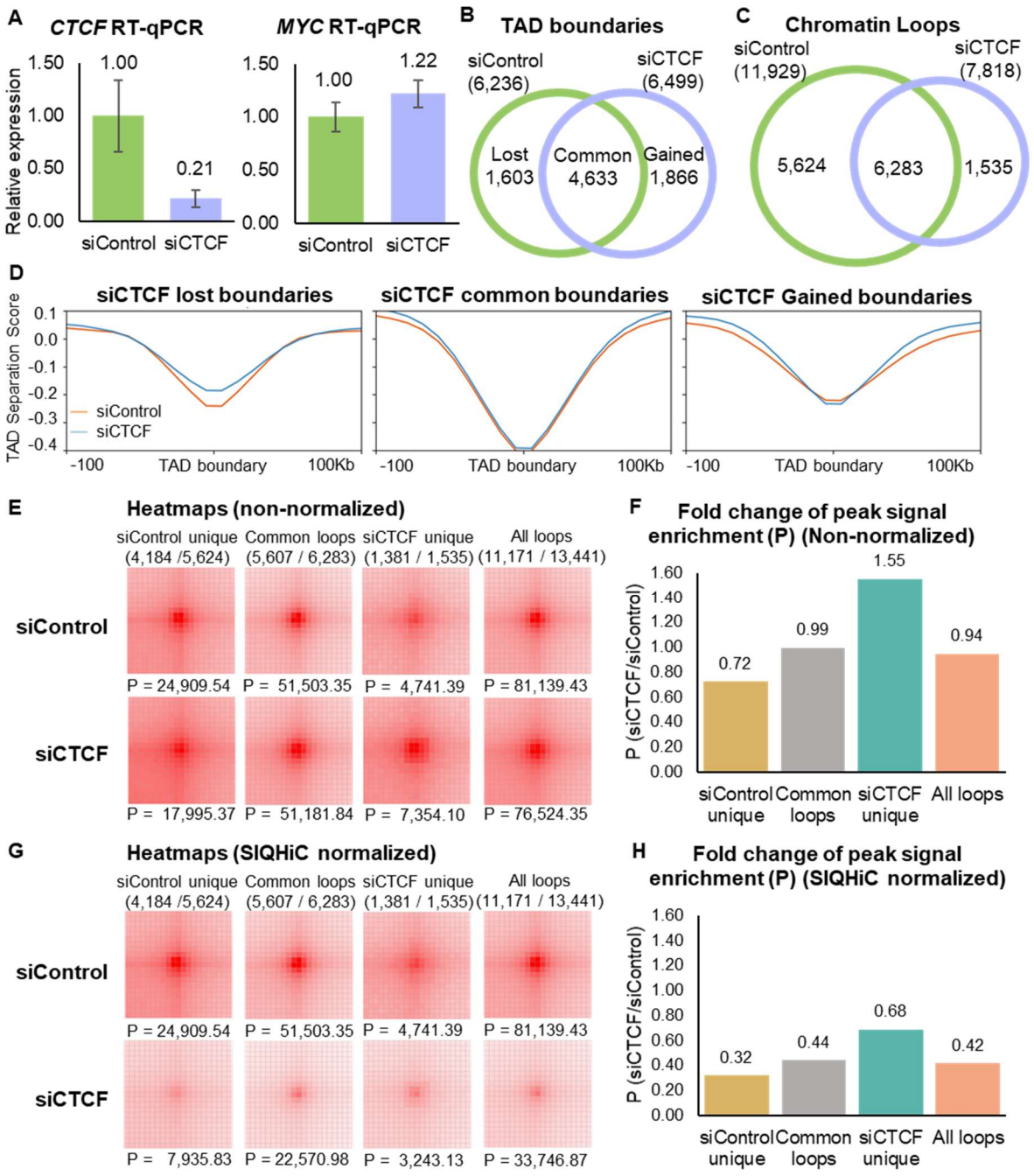

C
TAD boundaries
Low MYC
(4,460)
High MYC
(1,922)
Lost
2,912
Common
1,548
Gained
374
D
Chromatin Loops
Low MYC
(7,266)
High MYC
(7,907)
Lost
2,782
Common
4,478
Gained
3,429


A
Heatmaps (non-normalized)
Lost loops
(2,407 / 2,782)
Common loops
(3,728 / 4,478)
Gained loops
(1,861 / 3,429)
All loops
(7,996 / 10,689)
Low MYC
High MYC
P = 8,015.93
P = 20,911.77
P = 4,457.65
P = 33,385.35
P = 3,457.87
P = 15,463.21
P = 4,971.14
P = 23,892.21
B
Fold change of peak signal
enrichment (P) (Non-normalized)
P (High MYC/Low MYC)
0.4
0.7
1.1
0.7
Lost
Loops
Common
Loops
Gained
Loops
All loops
C
Heatmaps (SIQHiC normalized)
Lost loops
(2,407 / 2,782)
Common loops
(3,728 / 4,478)
Gained Loops
(1,861 /3,429)
All loops
(7,996 / 10,689)
Low MYC
High MYC
P = 2,725.42
P = 7,110.00
P = 1,515.60
P = 11,351.02
P = 3,457.87
P = 15,463.21
P = 4,971.14
P = 23,892.21
D
Fold change of peak signal
enrichment (P) (SIQHiC normalized)
P (High MYC/Low MYC)
1.3
2.2
3.3
2.1
Lost
Loops
Common
Loops
Gained
Loops
All loops

Supplemental Figure S5 | A/B compartments, topological associated domains and loops identified using SIQHiC. (**A**) A/B compartments are not altered after *MYC* overexpression, as shown in a representative view of chromosome 8. Pearson's correlation heatmaps were drawn using first principal component (PC1) at 100kb resolution, shown as tracks above the heatmaps. (**B**) Topologically associated domain (TAD) separation scores of lost, common and gained TAD boundaries in Low *MYC* and High *MYC* cells. (**C**) Overlap between TAD boundaries identified in Low *MYC* and High *MYC* cells. (**D**) Overlap between chromatin loops identified in Low *MYC* and High *MYC* cells.

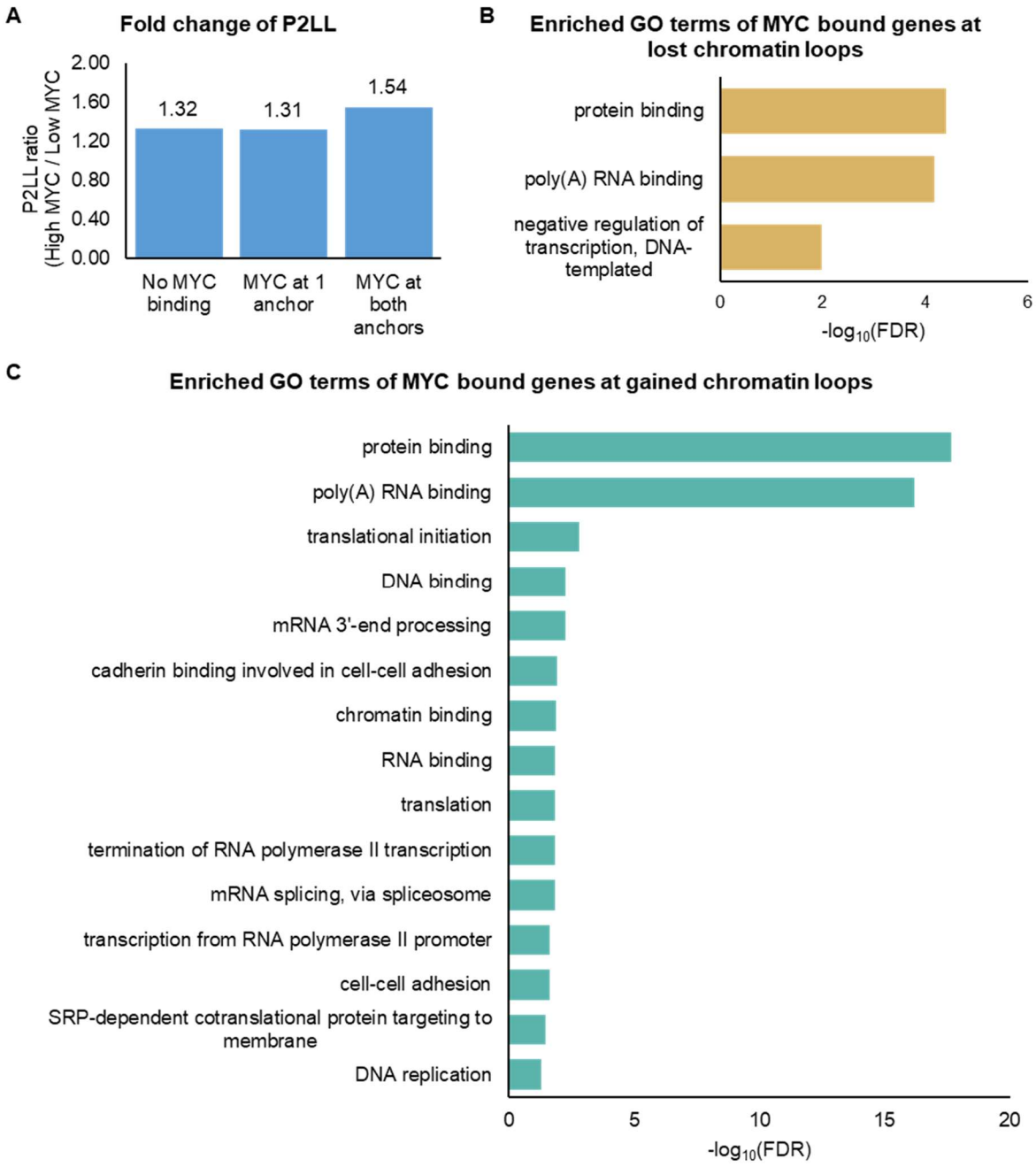


Supplemental Figure S6 | Aggregate peak analysis (APA). (**A**) Aggregate Peak Analysis (APA) plots at 5kb resolution showing the aggregate non-normalized signal of lost, common and gained chromatin loop sets in Low *MYC* and High *MYC* cells. Loop sets were filtered to remove short loops near the diagonal (shown above each APA plot). P: Peak signal at the centre pixel. (**B**) Fold change of non-normalized peak signal enrichment at lost, common, gained and all chromatin loop sets between High *MYC* and Low *MYC* cells. (**C**) Same APA analysis as in (**A**), but with SIQHiC normalization applied. (**D**) Fold change of SIQHiC normalized peak signal enrichment at lost, common and gained chromatin loop sets between High *MYC* and Low *MYC* cells.

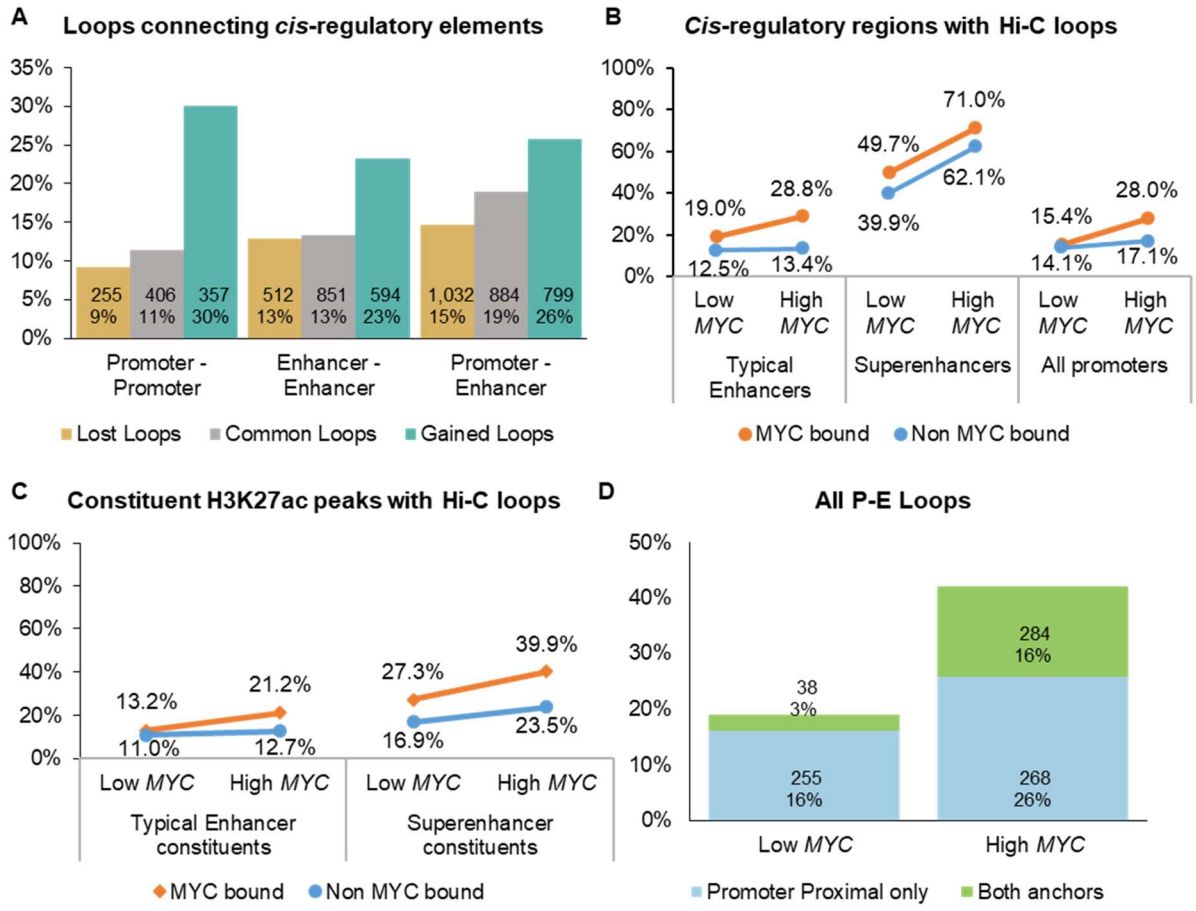


Supplemental Figure S7 | *CTCF* siRNA knockdown reduced insulation at selected TAD boundaries and reduced chromatin contact frequencies at all chromatin loops. (**A**) *CTCF* (left) and *MYC* (right) gene expression in *MYC* overexpressed U2OS cells transfected with control (siControl) and *CTCF* siRNA (siCTCF), normalized against the *TBP* gene. Error bars show standard error across 2 biological replicates. (**B**) Overlap between TAD boundaries identified in siControl and siCTCF cells. (**C**) Overlap between chromatin loops identified in siControl and siCTCF cells. (**D**)

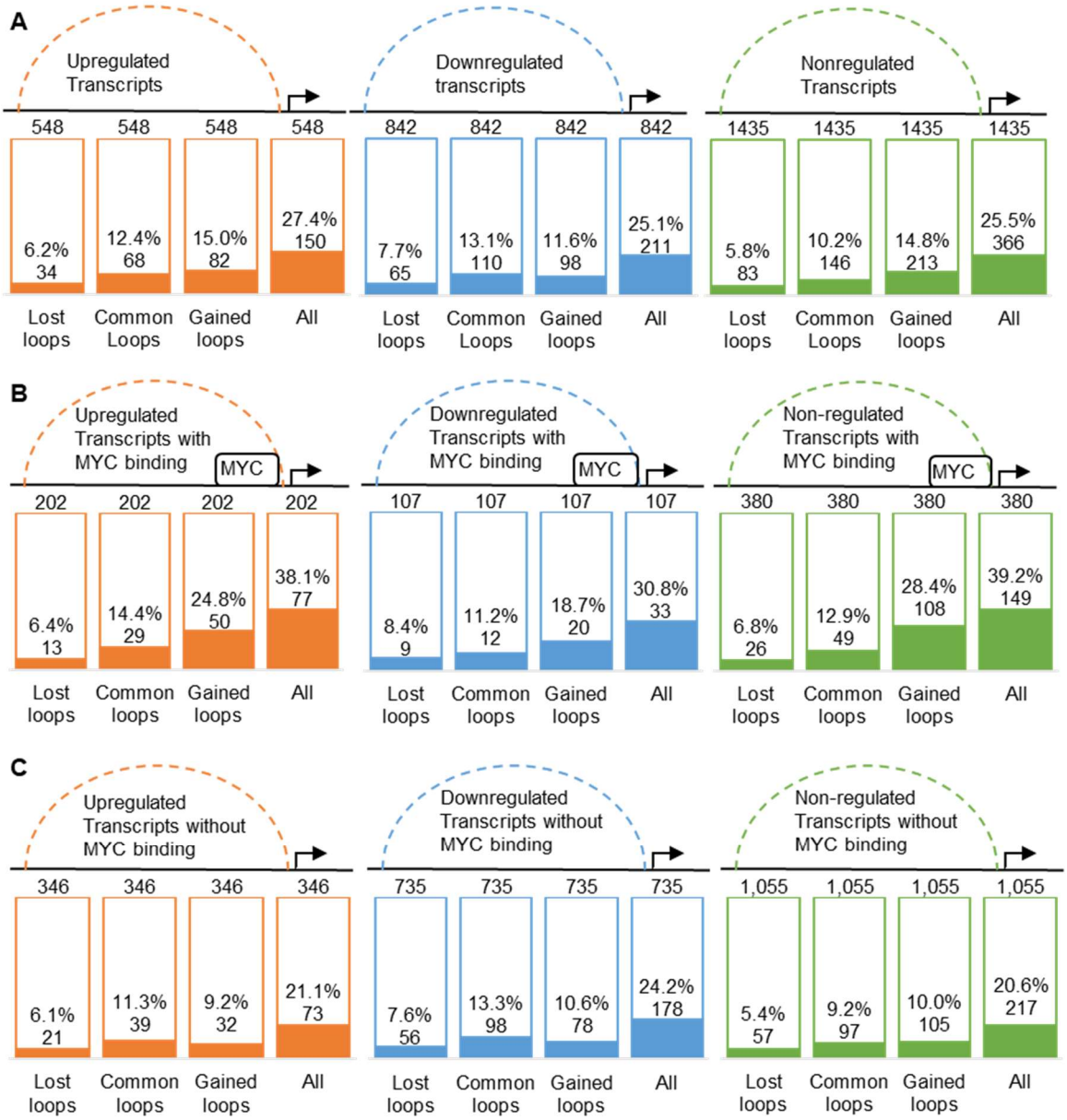
Topologically associated domain (TAD) separation scores of TAD boundaries that were lost, common and gained between siControl and siCTCF cells. (**E**) Aggregate Peak Analysis (APA) plots at 5kb resolution showing the aggregate non-normalized signal of lost, common and gained chromatin loop sets in siControl and siCTCF cells. Loop sets were filtered to remove short loops near the diagonal (shown above each APA plot). P: Peak signal at the centre pixel. (**F**) Fold change of non-normalized peak signal enrichment at lost, common, gained and all chromatin loop sets between siControl and siCTCF cells. (**G**) Same APA analysis as in (**E**), but with SIQHiC normalization applied. (**H**) Fold change of SIQHiC normalized peak signal enrichment at lost, common and gained chromatin loop sets between siControl and siCTCF cells.



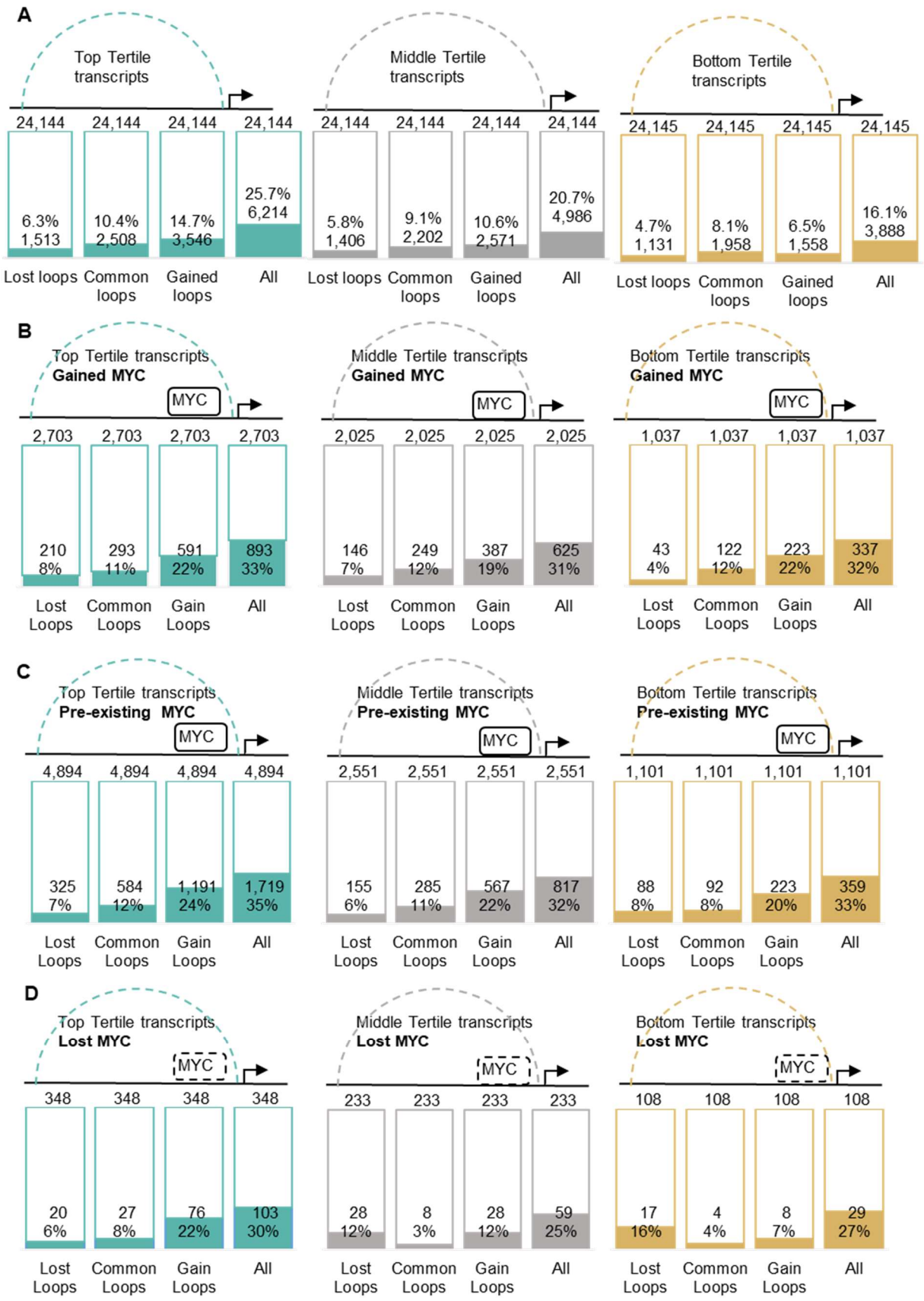
Supplemental Figure S8 | MYC binding at chromatin loops. (**A**) Fold change of P2LL signal between High *MYC* and Low *MYC* cells at chromatin loops with no MYC binding sites, loops with MYC binding sites at 1 loop anchor and loops with MYC at both loop anchors. P: peak enrichment signal. LL: Signal at lower left corner of the APA plot, representing local background signal. P2LL: Ratio of P to LL. (**B-C**) Gene ontology (GO) terms enriched in MYC bound genes at (**B**) lost and (**C**) gained chromatin loops.



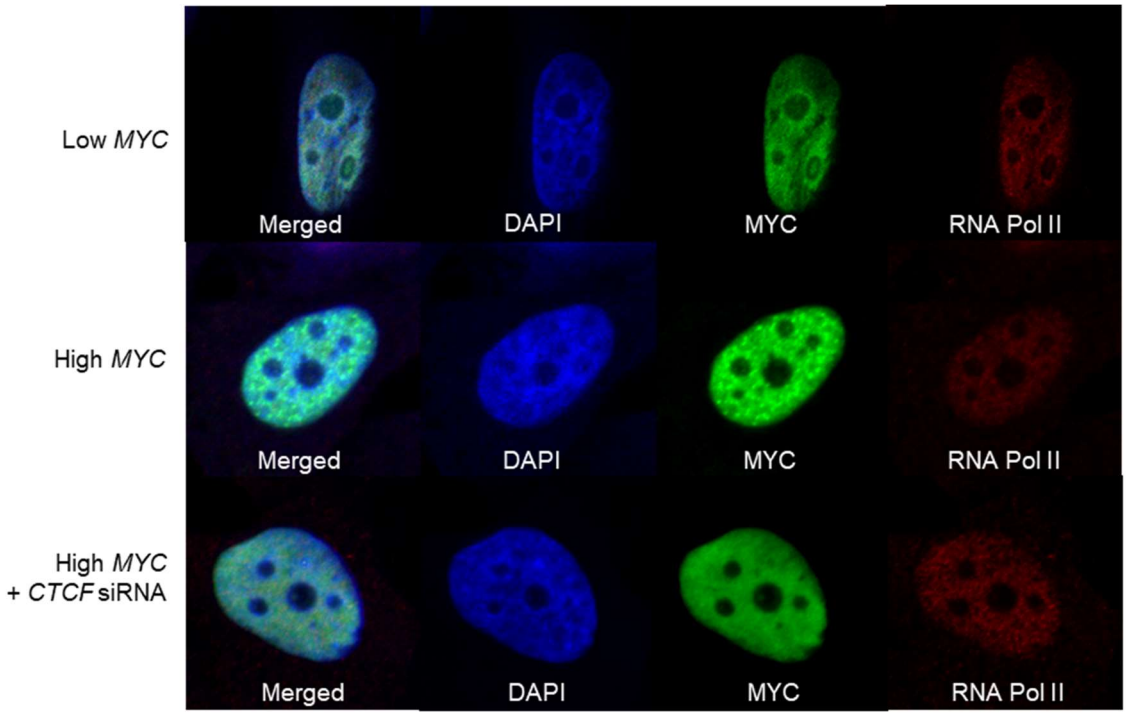
Supplemental Figure S9 | MYC binding at promoter-enhancer chromatin loops. (**A**) Percentage of promoter-promoter, enhancer-enhancer and promoter-enhancer loops within the lost, common and gained chromatin loop sets. (**B**) Percentage of MYC unbound (blue) and bound (orange) typical enhancers, superenhancers and promoters with Hi-C chromatin loops in Low *MYC* and High *MYC* cells. (**C**) Percentage of MYC unbound (blue) and bound (orange) typical enhancer and superenhancers H3K27ac constituent peaks. (**D**) Percentage of all promoter-enhancer chromatin loops bound by MYC at the promoter proximal loop anchor or at both anchors.



Supplemental Figure S10 | Correlation between *MYC* regulated transcripts and chromatin loops. (**A-C**) Percentage of (**A**) all, (**B**) MYC bound and (**C**) non MYC bound transcripts with Hi-C chromatin loops. Transcripts were stratified into upregulated (orange), downregulated (blue) and non-regulated (green) transcripts.


A
Top Tertile transcripts
24,144 24,144 24,144 24,144
6.3% 1,513
10.4% 2,508
14.7% 3,546
25.7% 6,214
Lost loops Common loops Gained loops All
Middle Tertile transcripts
24,144 24,144 24,144 24,144
5.8% 1,406
9.1% 2,202
10.6% 2,571
20.7% 4,986
Lost loops Common loops Gained loops All
Bottom Tertile transcripts
24,145 24,145 24,145 24,145
4.7% 1,131
8.1% 1,958
6.5% 1,558
16.1% 3,888
Lost loops Common loops Gained loops All
B
Top Tertile transcripts
Gained MYC
MYC
2,703 2,703 2,703 2,703
210 8%
293 11%
591 22%
893 33%
Lost Loops Common Loops Gain Loops All
Middle Tertile transcripts
Gained MYC
MYC
2,025 2,025 2,025 2,025
146 7%
249 12%
387 19%
625 31%
Lost Loops Common Loops Gain Loops All
Bottom Tertile transcripts
Gained MYC
MYC
1,037 1,037 1,037 1,037
43 4%
122 12%
223 22%
337 32%
Lost Loops Common Loops Gain Loops All
C
Top Tertile transcripts
Pre-existing MYC
MYC
4,894 4,894 4,894 4,894
325 7%
584 12%
1,191 24%
1,719 35%
Lost Loops Common Loops Gain Loops All
Middle Tertile transcripts
Pre-existing MYC
MYC
2,551 2,551 2,551 2,551
155 6%
285 11%
567 22%
817 32%
Lost Loops Common Loops Gain Loops All
Bottom Tertile transcripts
Pre-existing MYC
MYC
1,101 1,101 1,101 1,101
88 8%
92 8%
223 20%
359 33%
Lost Loops Common Loops Gain Loops All
D
Top Tertile transcripts
Lost MYC
MYC
348 348 348 348
20 6%
27 8%
76 22%
103 30%
Lost Loops Common Loops Gain Loops All
Middle Tertile transcripts
Lost MYC
MYC
233 233 233 233
28 12%
8 3%
28 12%
59 25%
Lost Loops Common Loops Gain Loops All
Bottom Tertile transcripts
Lost MYC
MYC
108 108 108 108
17 16%
4 4%
8 7%
29 27%
Lost Loops Common Loops Gain Loops All

Supplemental Figure S11 | Correlation between transcript expression and chromatin loops. (**A-D**) Percentage of (**A**) all transcripts, (**B**) transcripts with gained MYC binding sites, (**C**) transcripts with pre-existing MYC binding sites and (**D**) transcripts with lost MYC binding sites with Hi-C chromatin loops. Transcripts were stratified into top (teal), middle (grey) and bottom (khaki) tertiles based on transcript expression after *MYC* overexpression.



Supplemental Figure S12 | Immunofluorescence staining of MYC and RNA polymerase II. From left to right, immunofluorescence images show merged signal, DAPI staining of DNA in blue, MYC staining in green and RNA polymerase II staining in red. RNA Pol II: RNA polymerase II.