SI Appendix

Interaction of the oncoprotein transcription factor MYC with its chromatin co-factor WDR5 is essential for tumor maintenance

Lance R. Thomas, Clare M. Adams, Jing Wang, April M. Weissmiller, Joy Creighton, Shelly L. Lorey, Qi Liu, Stephen W. Fesik, Christine M. Eischen, and William P. Tansey

Corresponding author: William Tansey (william.p.tansey@vanderbilt.edu)

This PDF file includes:

Other supplementary datasets for this manuscript include the following Excel files:

Materials and Methods

Plasmids. pcDNA-Hygro-CRE-ERT2 was made by PCR-amplification of CRE-ERT2 from pCAG-ERT2-CRE–ERT2 (gift from Connie Cepko, Addgene #13777) to introduce a Kozak sequence and initiation codon; this cassette was then inserted into the NheI and ApaI sites of pcDNA3.1-Hygro(+). pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene #42230). pUC57, containing a U6 promoter, scaffold, and terminator was synthesized by *Genscript* and the gRNA sequence TCAACGTTAGCTTCACCAAC inserted by whole plasmid PCR. WT–MYC Exon 3 targeting vector was made by assembly of the following fragments into pBluescript II SK+: (i) [Intron 2 (Chr8: 127739145–127740190) as SacI/XbaI], (ii) [loxP–Intron 2 (Chr8: 127740207-127740395)–Recoded Exon 3–P2A (synthesized by *Genscript*) as XbaI/BamHI], (iii) [Puromycin resistance–SV40 terminator (from pcDNA3.1-Puro) as Bam/HindIII], (iv) [loxP–Intron 2–Exon 3 (Chr8: 127740207-127740955)–2xHA tags as HindIII/XhoI], (v) [P2A-GFP as XhoI/BglII], and (v) [Intron 3 (Chr8: 127740955-127741963) as BglII/ApaI]. WBM–MYC Exon 3 targeting vector was made by whole plasmid PCR of the WT–MYC Exon 3 vector to introduce the WBM mutation. ∆264–Exon 3 targeting vector was made using the WT plasmid as template for PCR amplification of a 2xHA-P2A-GFP fragment which was ligated into the ClaI and BglII sites of the WT plasmid.

Chromatin immunoprecipitation (ChIP) primers. Primers used for qPCR on ChIP DNA were *SNHG15* (cgccactgaacccaatcc, tctagtcatccaccgccatc), GB (aattatgtgtccagggttgc, caccggcttctatattccac), *RNPS1* (tgaggagtggaccggcttc, cggcgggaagatgtaagttg), *RPS14* (gaaagacccccgtctctcgt, gagacgacgtgcaggtaggag), *EIF4G3* (cctttcacggcaatatcctc, tgagtgaagaaaatccaccg), *RPL5*(gctttagatgcatggaggtt, ccaagtcagagctagcaatc), *RPL17* (cagcgaggatttagtagcc, gtttctcctgcgttgctc), POU6F1 (acacaactgtcctcctctta, ctggtgtcctgtttccttttc), *ZFPM1* (ctcctcgatttccagcct, tttctcttgggccagactt), and *ZNF771* (gcagaggagataaagaggga, ctgctccgcccgaaatag).

Antibodies. α-HA (C29F4; #3724), α-WDR5 (D9E1I; #13105), α-MYCN (D4B2Y; #51705), α-cMYC (#9402) antibodies, and Normal Rabbit IgG, were from *Cell Signaling Technology*. Other antibodies were: α-MAX (*Santa Cruz*; sc-765), α-cMYC (*Abcam*; Y69), α-GAPDH (*Thermo Fisher*; PA1-987-HRP), α-HA (CSHL Monoclonal Antibody Shared Resource; 12CA5), α-HCF-1 (*Bethyl*; A301-400A), anti-rabbit IgG Fc-HRP (*Thermo Fisher*; 31463), and anti-Rabbit IgG-HRP, Light Chain Specific (*Jackson ImmunoResearch*; 211-032-171).

Cell culture and transfections. Ramos cells were obtained from the ATCC (CRL-1596) and maintained in RPMI media supplemented with 10% FBS, 100 IU/ml Penicillin, and 100 µg/ml streptomycin. For generation of cells stably expressing CRE–ER, 10^7 Ramos cells were electroporated with 40 µg of PvuI-linearized pcDNA-Hygro-Cre-ERT2, and selected with 50 µg/ml Hygromycin. For generation of switchable cell lines, a CRE–ER-positive clone was electroporated with 10 µg of the appropriate targeting vector, 15 µg of gRNA vector, and 15 µg of pX330-U6-Chimeric_BB-CBh-hSpCas9. After a two-day recovery, stable cells were selected with 200 ng/ml Puromycin. Clones were obtained by limited dilution, and screened by PCR prior to Southern blotting. Cells were treated with 20 nM (Z)-4-Hydroxytamoxifen (OHT: *Tocris*) to activate CRE–ER. Ramos cells permanently expressing HA-tagged WT MYC were made by treating the WT MYC line with 20 nM 4-OHT for 24 hours followed by limited dilution cloning.

Southern blotting. For each clone, 10 µg of genomic DNA was digested with XbaI and run on an 0.8% TAE agarose gel. DNA was transferred to Hybond N+ (*GE Lifesciences*). Probe 1 was generated by PCR off Ramos gDNA (primers: CGATTTCGATTCCTCTGC and CCAGATCTGCTATCTCTCC). Probe 2 was generated by PCR off pcDNA3.1-Puro (primers: ATGACCGAGTACAAGCCC and GGCACCGGGCTTGCGGGT).

Precision run on (PRO)–Seq. For PRO-Seq analysis, 3 x 10^7 cells were switched for 24 hr prior to analysis and nuclei prepared according to published protocols (1), with minor changes. Briefly, nuclei were by dounce homogenization in 10 mM Tris-HCl (pH 7.4), 300 mM sucrose, 3 mM $CaCl_2$, 2 mM $MgCl_2$, 0.1% Triton-X 100, 0.5 mM DTT, RNAse inhibitor and protease inhibitor cocktail (*Roche*) and then snap frozen in 10 mM Tris-HCl, pH 8.0, 25% glycerol, 5 mM $MgCl_2$, 0.1 mM EDTA, 5 mM DTT with protease inhibitor cocktail and stored at -80°C until ready to use. Samples of proteins from the nuclear and cytosolic extracts were probed by Western blotting to confirm the integrity of nuclei preparations and to determine the location of switched MYC proteins. Biotin run-on reactions were performed on thawed nuclei in a reaction buffer containing Biotin-11-CTP (*PerkinElmer*) for 3 minutes at 30°C. Reactions were stopped by adding Trizol LS (*Thermo Scientific*) and RNA purified by chloroform and isopropanol extraction. Resuspended RNA pellets were heated at 65°C for 40 seconds and 1 M NaOH was added, followed by incubation on ice for 10 minutes. Base hydrolysis was neutralized by addition of 1 M Tris (pH 6.8) and the sample ran over a Micro Bio-Spin P-30 gel column (*BioRad*). Streptavidin Dynabeads (*ThermoFisher*) were incubated with collected material and biotinylated RNA bound to beads using their standard protocol. Elution of bound, biotinylated, RNA was achieved by extracting beads with Trizol and subsequently purified using chloroform and isopropanol. RNA adaptors (*IDT*) were added to the 3' side biotinylated RNA and following a second round of biotin-RNA purification, 5' RNA caps were removed using CAP CLIP (*CellScript*). 5' RNA adaptors (*IDT*) were then added. One additional biotin-RNA purification was performed, and purified RNA was used in a reverse transcriptase reaction to generate cDNA. Library amplification was performed with Phusion high fidelity polymerase (*NEB*) and customized Illumina-based index primers (*IDT*). Sequencing data was obtained on an Illumina NextSeq500 with 75 bp single reads. Sequencing was performed at the VU VANTAGE Shared Resource.

Bioinformatic analysis. ChIP-Seq reads were aligned to the human genome (hg19) using Bowtie2 (2). Peaks in each sample were called using MACS2 with q-value of 0.05 (3). Differential binding peaks were identified using DiffBind (4) based on consensus peaks occurring at least two replicates. Peaks were annotated using Homer command annotatePeaks, and enriched motifs were identified by Homer command findMotifsGenome with the default region size and the motif length (-size 200 and -len 8, 10, 12) (http://homer.ucsd.edu/homer/). After adapter trimming and low quality sequence removal by Cutadapt (5), PRO-Seq reads longer than 15bp were reversed complemented using FastX tools (6). Reversed complemented reads were aligned to human genome using Bowtie2. Reads mapped to rRNA loci and reads with mapping quality less than 10 were removed. NRSA (7), a tool to provide a comprehensive analysis on nascent transcriptional profiles for known genes and novel enhancers, was used to estimate RNA polymerase abundance in proximal-promoter and gene body regions of genes. RNA-Seq reads were aligned to human genome using STAR (8) after adapter trimming, and quantified by featureCounts (9). Differential analysis for RNA-Seq data were performed by DESeq2 (10). Gene set enrichment analysis (GSEA) (11) was performed on the list of expressed genes pre-ranked by the log_2-fold change in expression between MYC mutations compared with wild type. A false discovery rate of 0.05 was used as the cutoff to select significantly enriched gene sets.

In vivo studies. Athymic nude mice (*Foxn1$^{nu/nu}$*; female 6-8 weeks old; *Envigo*) were subjected to whole body irradiation (6 Gy) 24 hours prior to subcutaneous flank injection of 10^7 switched or unswitched WT, Δ264, or WBM Ramos cells (described above). For tumor engraftment studies, one cohort of mice was injected with cells that were cultured *in vitro* with OHT. To maximize the proportion of switched cells that were injected into mice, cultures went through three successive rounds of OHT treatment (on days one, three, and five) during expansion for these experiments. For tumor maintenance studies a second cohort of mice was injected with unswitched cells and allowed to grow palpable tumors. Once tumors reached approximately 200 mm^3 (15 days after injection), mice were

administered tamoxifen (2 mg in corn oil, intraperitoneal) once daily for 3 consecutive days to induce the switchable MYC cassette *in vivo*. A portion of the mice with size-matched tumors prior to tamoxifen were sacrificed 48 and 96 hours following the first administration of tamoxifen, and lymphoma cells analyzed for apoptosis. Cells stained with propidium iodide for fragmented (subG1) DNA and Annexin V/7AAD. Tumors were measured with calipers and volumes calculated using the ellipsoid formula. Mice were sacrificed at humane endpoints based on tumor volume, and log-rank tests used to determine statistical significance for survival. Two-tailed *t*-tests were used to determine significance when comparing two groups. All studies complied with state and federal guidelines and were approved by the Institutional Animal Care and Use Committee at Thomas Jefferson University.

Tumor switch quantification. Tumor DNA was isolated using PureLink Genomic DNA Mini Kit (*Invitrogen*). Primers AAGAAAACGTGAAACGCA and GAGAGCGAAGAAGGATCTC were used in qPCR to detect retention of recoded wild-type exon 3 ($Ex3^R$). ΔCt values were calculated against a genomic loading control (*SNHG15* gene body primer) then $\Delta\Delta$Ct was calculated using genomic DNA isolated from WT cells grown in the presence of puromycin (100% $Ex3^R$) or a permanently-switched subclone (0% $Ex3^R$).

Tumor RNA-Sequencing. RNA extraction from frozen tumor shavings and sequencing was performed by Genewiz. RNA was isolated by PolyA selection and libraries were prepared for paired-end 150 bp reads. Sequencing was performed on an Illumina HiSeq 4000, with 20-30 million reads per sample.

Supplementary References

1. D. B. Mahat *et al.*, Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**, 1455-1476 (2016).
2. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
3. J. Feng, T. Liu, B. Qin, Y. Zhang, X. S. Liu, Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**, 1728-1740 (2012).
4. R. B. Stark, G.D., DiffBind: differential binding analysis of ChIP-seq peak data. *Bioconductor* (2011).
5. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, 10-12 (2011).
6. W. R. Pearson, T. Wood, Z. Zhang, W. Miller, Comparison of DNA sequences with protein sequences. *Genomics* **46**, 24-36 (1997).
7. J. Wang *et al.*, Nascent RNA sequencing analysis provides insights into enhancer-mediated gene regulation. *BMC genomics* **19**, 633 (2018).
8. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
9. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
10. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
11. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

Figure S1

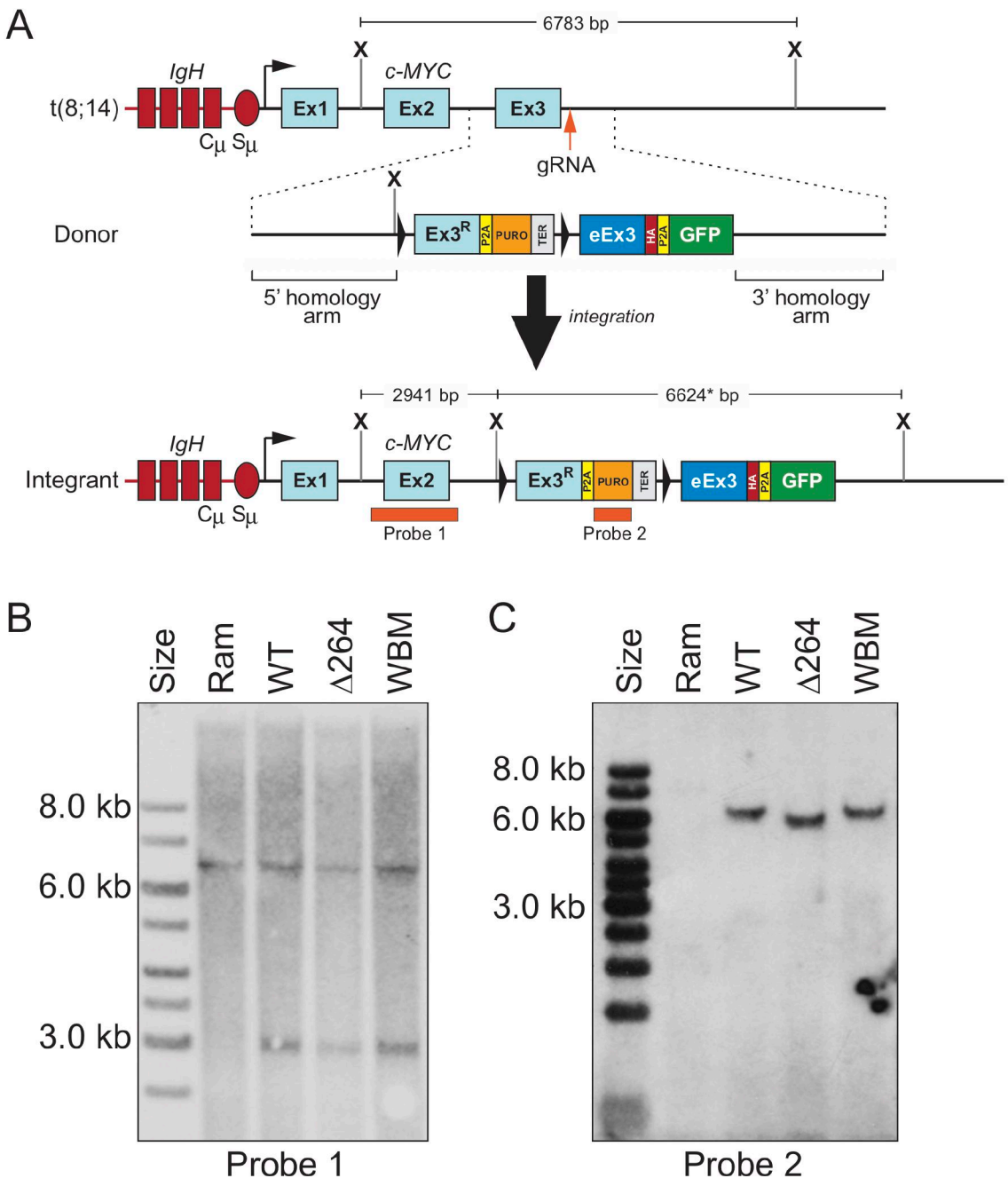


Fig. S1. Construction of modified Ramos cells to permit inducible exchange of exon three of *c-MYC*. (A) Schematic of the t(8;14) translocated *c-MYC* gene in Ramos cells (top). Shown are the location of XbaI sites (X) and the distance between them, and the guide RNA (gRNA) used to induce a double-stranded DNA break. The donor cassette used for homologous recombination is displayed beneath (middle), which carries a unique XbaI site near the 5' LoxP sequence. Structure of the modified locus after integration (bottom), representing the modified XbaI cleavage pattern, and the position of the two probes used for Southern blotting. (B) Southern blot with Probe 1 (c-MYC exon 2; left) reveals the expected sizes of XbaI digestion products (~2.9 and 6.6 kb) for parental Ramos cells (Ram) and the successful integrant cell lines, respectively. (C) Southern blot with Probe 2 (PURO resistance gene; right) reveals the expected XbaI fragment present only in correctly targeted cell lines.

Figure S2

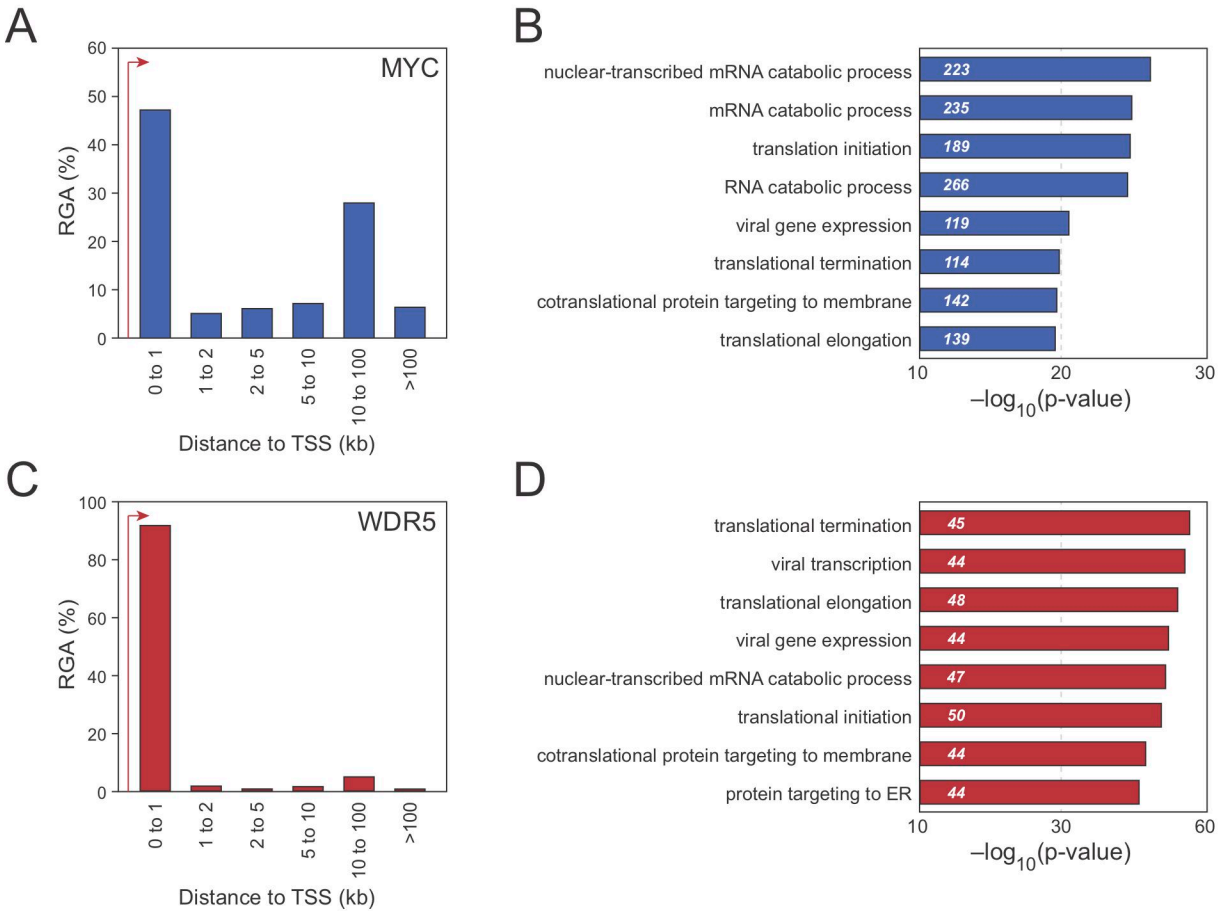


Fig. S2. Characteristics of WT–MYC and WDR5 binding sites in Ramos cells. (A) Distribution of HA-tagged WT–MYC binding sites in switched WT cells, binned according to distance from the annotated TSS. RGA refers to region-gene association. (B) Top eight GO enrichment categories for genes bound by HA-tagged WT–MYC (within 1 kb of a TSS). Numbers in italics represent the number of genes enriched in each category. (C) As in A) except for WDR5 binding sites. (D) As in (B) except for WDR5.

Figure S3

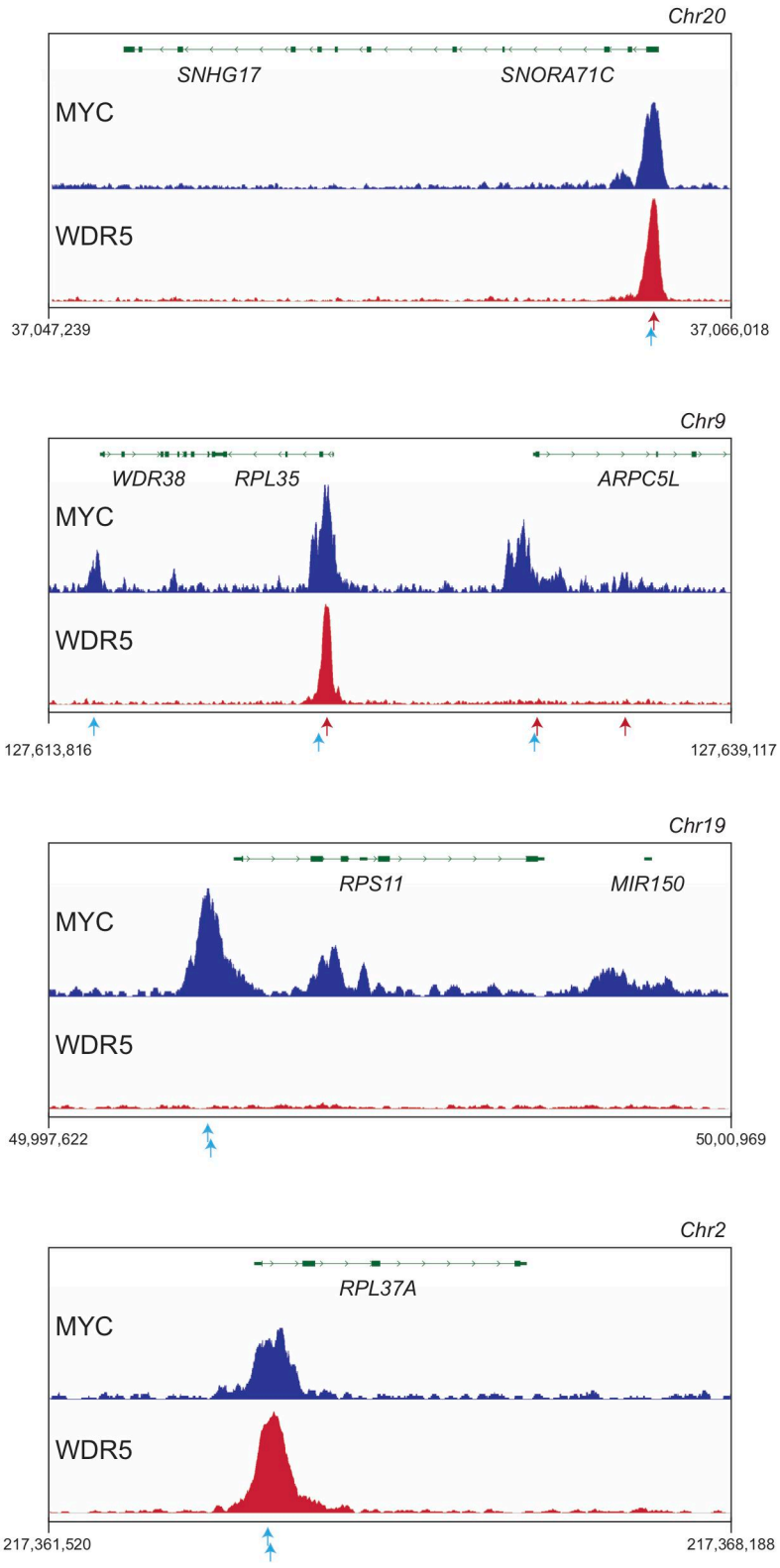


Fig. S3. MYC and WDR5 binding at representative genes. The figure shows four IGV screenshots of normalized representative ChIP-Seq data for HA-tagged MYC (blue) and WDR5 (red). Red arrows indicate positions of perfect E-box sequences; blue arrows indicate position of imperfect E-box sequences.

Figure S4

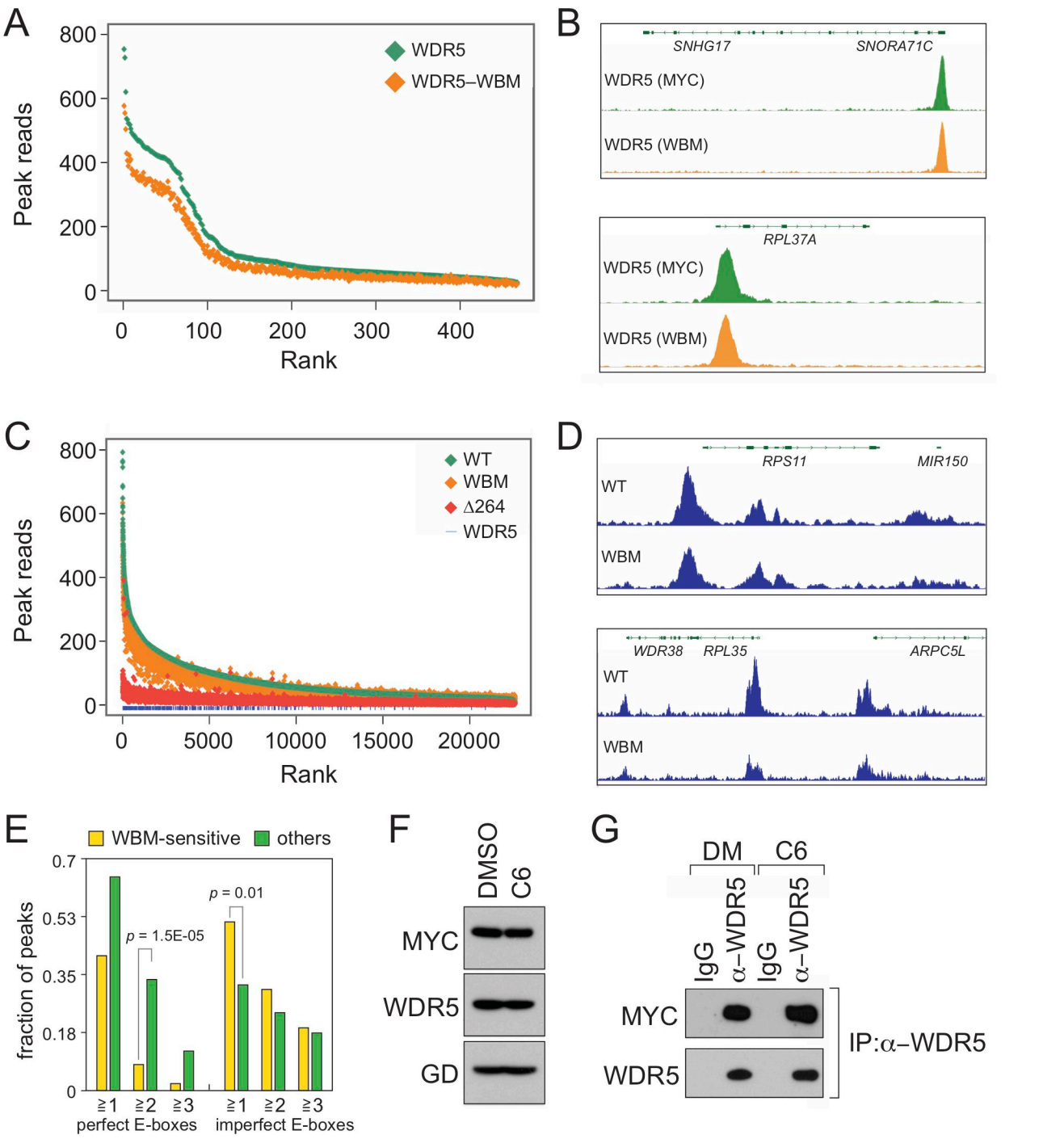


Fig. S4. Impact of MYC mutants on WDR5 and MYC binding to chromatin in engineered Ramos cells, and controls for C6 treatment. (A) Scatterplot of normalized average read counts for WDR5 binding peaks (as determined by ChIP-Seq) in either parental Ramos cells (WDR5) or switched WBM cells (WDR5–WBM). (B) Two IGV screenshots of normalized representative ChIP-Seq data for WDR5 in parental Ramos cells [WDR5 (MYC)] or switched WBM cells [WDR5 (WBM)]. (C) Scatterplot of normalized average read counts for HA-tagged MYC proteins in the indicated cell lines after switching. Loci co-bound by WDR5 are indicated by a vertical blue line at the bottom of the plot. (D) Two IGV screenshots of normalized representative ChIP-Seq data for HA-tagged MYC (WT or the WBM mutant) in the indicated cell lines after switching. *RPS11* and *ARPC5L* do not bind WDR5. *RPL35* is a WDR5-bound gene. (E) Graph shows the fraction of peaks with the listed number of perfect or imperfect E-boxes, broken down to the 88 WBM-sensitive genes (yellow) or the remainder of shared MYC/WDR5 binding sites (green). *p* values were calculated by hypergeometric analysis. (F) Western blot, showing MYC, WDR5, and GAPDH (GD) in "WT" engineered Ramos cells treated with either DMSO or 25 μM C6 for four hours. (G) "WT" engineeredRamos cells were treated with DMSO, or 25 μM C6 for four hours, lysate prepared, and IP carried out with an anti-WDR5 antibody. Immune complexes were probed for MYC and WDR5 by Western blotting.

Figure S5

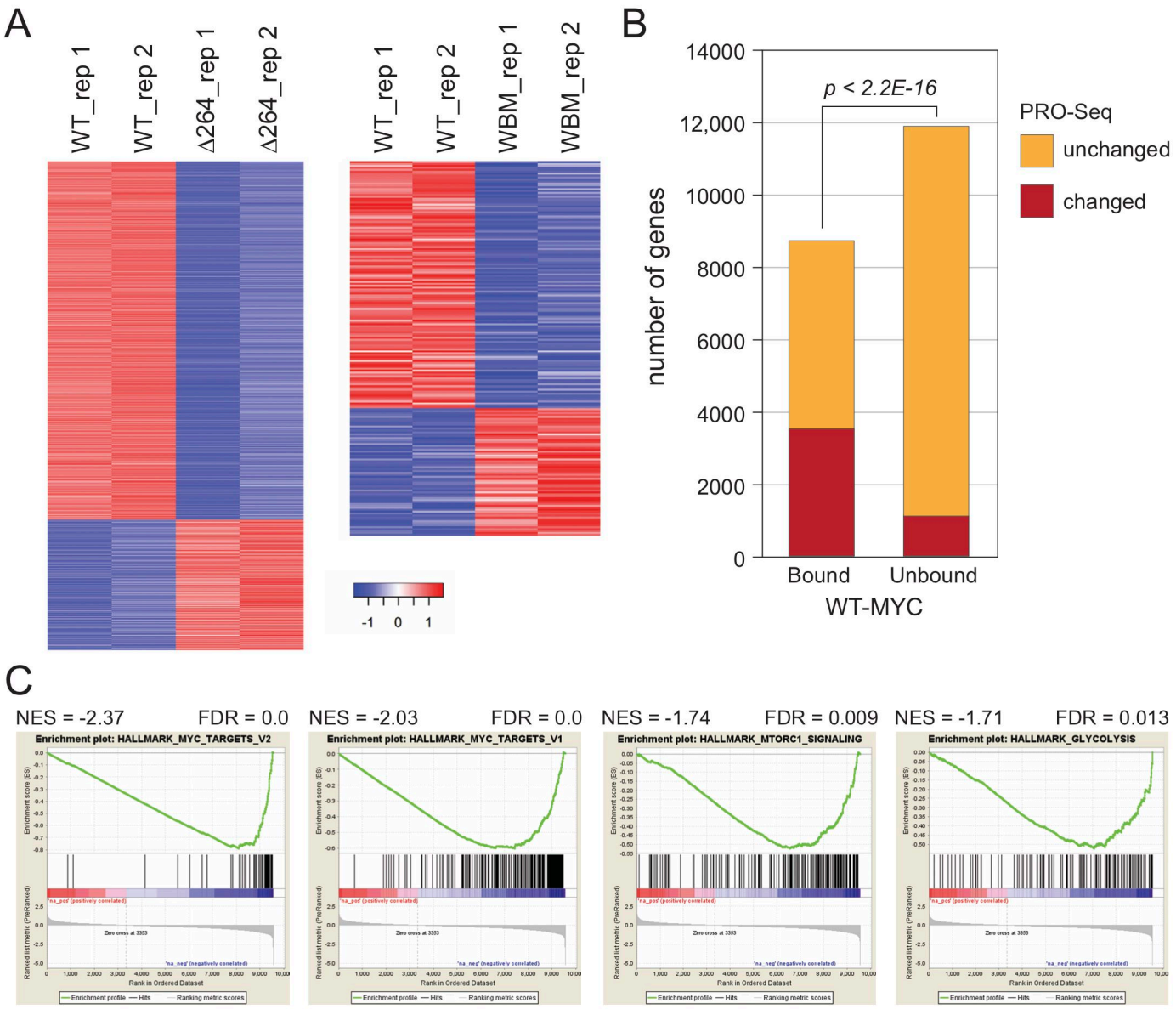


Fig. S5. Transcriptional consequences of switching Ramos cells to express MYC mutants, as determined by PRO-Seq. (A) Heatmaps, showing changes in the individual replicates for PRO-Seq analysis. The scale bar indicates z-transformed normalized read density (Z-score). (B) Stacked bar graph, comparing MYC bound and unbound genes (as determined by ChIP-Seq in WT switched cells) with those genes that display a significant change (either direction) in gene body-associated RNA polymerases in the Δ264 switched cells. A significantly higher proportion of MYC bound genes display transcriptional changes in PRO-Seq (red) than the unbound gene set. (C) Highly significantly enriched Reactome gene sets (defined in the Molecular Signatures Database), determined by GSEA of PRO-Seq from Δ264 switched cells, compared to WT switched counterparts. Shown also are the normalized enrichment scores (NES) and the corresponding FDR values.

Figure S6

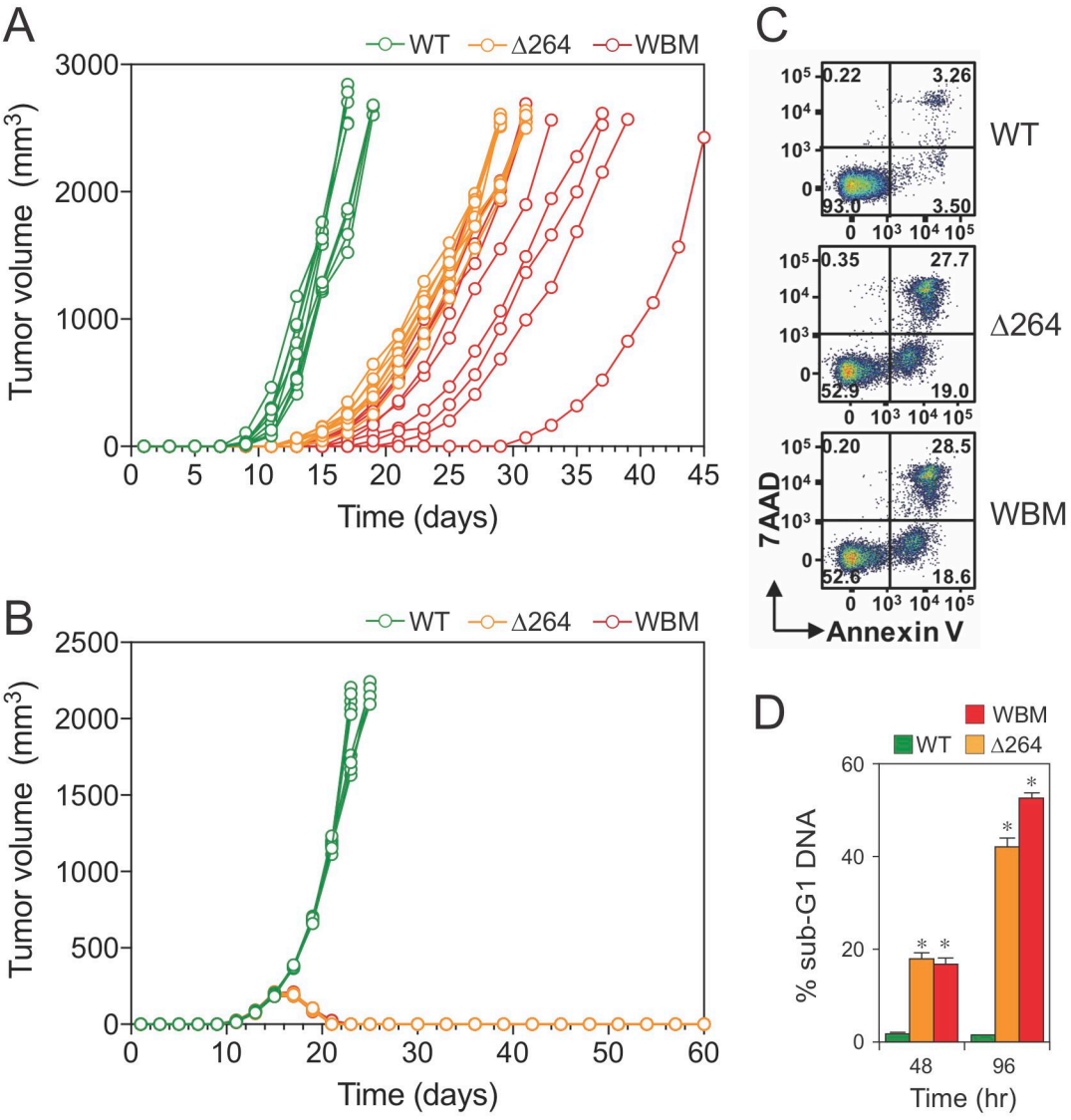
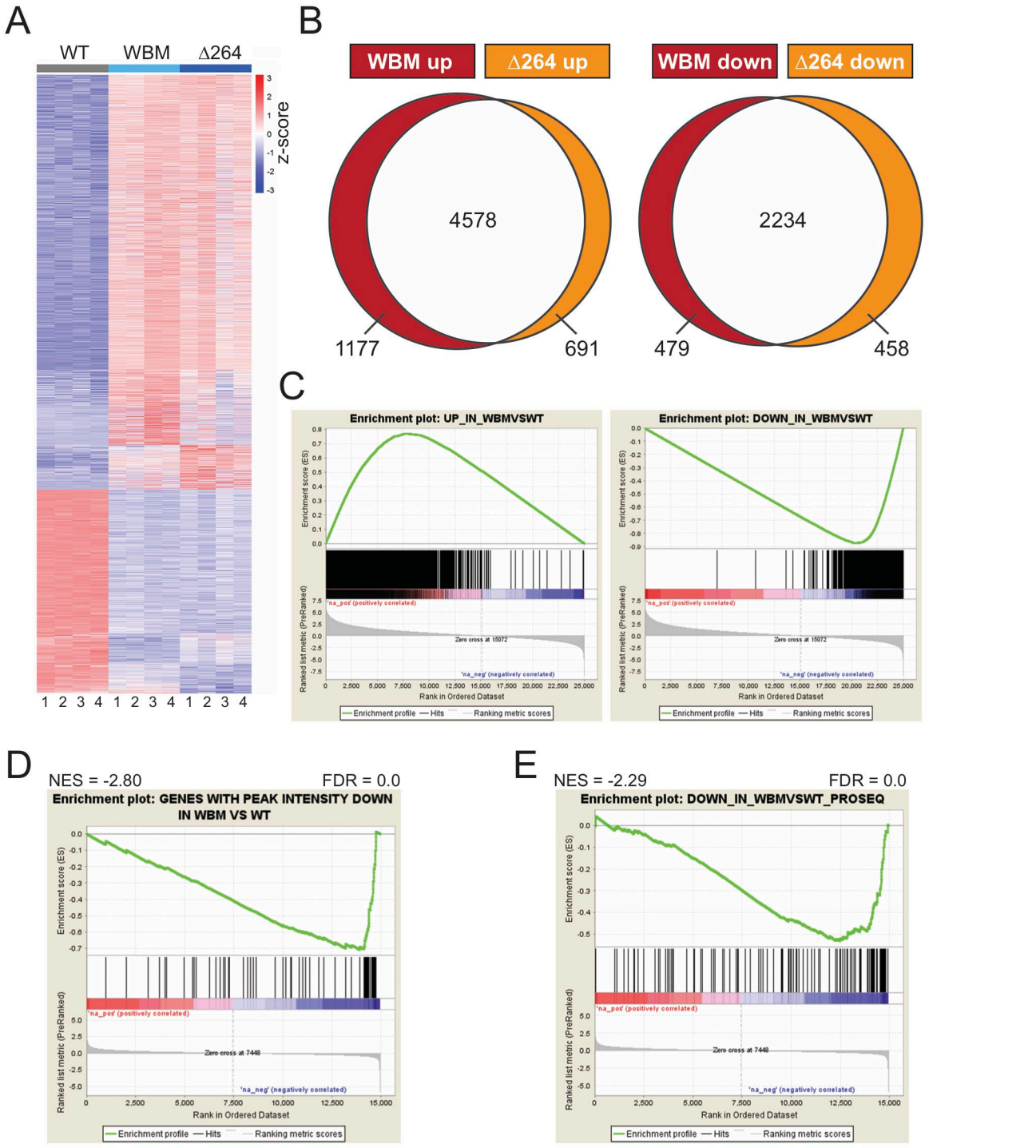


Fig. S6. *In vivo* consequences of switching Ramos cells to express the Δ264 and WBM MYC mutants. (A) Individual tumor volumes in mice injected with the indicated switched Ramos cells over the entire course of the experiment. Average data are presented in Fig. 5B. (B) Individual tumor volumes in mice injected with the indicated unswitched Ramos cells, before and after switching was induced beginning at day 15. (C) Representative flow cytometry data, staining for Annexin V and 7AAD (non-viable cell stain), in the indicated tumors harvested 48 hours following tamoxifen administration. (D) For the indicated cell types, tumors were harvested (n=4 per group) 48 and 96 hours following tamoxifen administration. Apoptosis was evaluated in the isolated lymphoma cells by staining with propidium iodide and scoring sub-G1 DNA content by flow cytometry. The extent of sub-G1 in Δ264 and WBM tumor cells was significantly different (*) from WT tumors ($p < 1.56E-4$; two-tailed t-test).

Figure S7



Supplementary Figure 7. Comparative transcriptomics of tumor material after in vivo switching. (A) Heatmap, displaying z-transformed gene expression for significantly changed genes in tumors 48 hours after switching, as determined by RNA-Seq. Data for all four mice for each switch (WT, WBM, or Δ264) are shown. (B) Venn diagrams, showing overlap of significantly (FDR < 0.05, |log2-fold change| >1) up- (left) or down- (right) regulated genes in the WBM versus Δ264 tumors. (C) GSEA comparing genes that had increased (left) or decreased (right) transcription in Δ264 tumors against a gene list ranked by expression alteration in WBM tumors vs. wild-type, as determined by RNA-Seq. (D) GSEA comparing genes with a reduction in MYC binding in WBM mutation determined by ChIP-Seq against a gene list ranked by expression alterations in WBM vs. Δ264 tumors, as determined by RNA-Seq. (E) GSEA comparing genes with a reduced gene body-associated polymerases in WBM mutation determined by PRO-Seq against a gene list ranked by expression alterations in WBM vs. Δ264 tumors, as determined by RNA-Seq.