**Figure titles and legends**

**Figure 1. MYC overexpression selectively dysregulates embryonic stem cell genes and tissue-lineage genes in tumorigenesis**

**a**, Workflow for identifying genes differentially expressed (DE) in MYC-induced transgenic mouse tumor models. **b**, Venn diagram of the DE genes among five MYC transgenic mouse models. **c**, Enrichment of mouse tissue-lineage genes among DE upregulated genes from transgenic mouse tumors. Mouse tissue-lineage genes are genes highly expressed in mouse liver, kidney, spleen, lung, or embryonic stem cell tissue based on BioGPS Mouse Gene Atlas data. For LAC, KRASG12D‑induced (KRAS) and MYC/KRASG12D co-induced (MYC+KRAS) tumors were also included in addition to the MYC-induced tumors. **d,** Same as **c** except enrichment of mouse tissue-lineage genes was assessed among DE downregulated genes from transgenic mouse tumors. \*p < 0.000595 (Bonferroni correction for 84 tests, α = 0.05, Fisher's exact test), n.s. = not significant.

**Figure 2. Gene expression changes in MYC-induced tumorigenesis are associated with epigenetic changes**

**a**, Metagene plots and heatmaps showing changes in H3K4me3 and H3K27ac in the promoters of genes upregulated and downregulated in MYC‑induced mouse HCC (left) and BCL (right) (GEO accession numbers: GSE76042 and GSE51004, respectively). Log2 fold changes reflect ChIP-Seq signal changes in tumor relative to normal tissue, both normalized to an input sample. **b**, ChIP‑Seq H3K27ac signal changes in tumor relative to normal for superenhancers (obtained from dbSUPER) associated with genes upregulated and downregulated in the MYC‑induced mouse HCC (left) and BCL (right) model. P-values were determined by Welch’s t-test.

**Figure 3. Pathways associated with genes differentially expressed in MYC-driven tumorigenesis reflect tissue dedifferentiation**

Gene ontology (GO) gene set enrichment analysis (GSEA) of genes, ranked by log2 fold change, was performed for each of the five MYC-induced transgenic mouse tumor models. Heatmap shows the Fisher’s exact test adjusted p-values for the top GO terms among the different models. Each column of the heatmap represents a distinct GO term and similar GO terms were grouped together into pathways. The bar graphs at the bottom depict pathway enrichment among lung, liver, kidney, spleen, and embryonic stem cell tissue-lineage genes (BioGPS Mouse Gene Atlas); each bar shows the median enrichment across the GO terms in the pathway.

**Figure 4. A tumorigenesis gene signature that is highly associated with MYC expression in primary human cancers**

**a**, One-sided volcano plot representing the pairwise Pearson correlation between each gene’s expression and MYC expression across 33 cancer types in The Cancer Genome Atlas (TCGA)’s RNA-seq dataset (n = 9354 primary human cancers). Robust rank aggregation (RRA) was performed across the 33 cancer types by ranking each gene by its Pearson correlation with MYC expression within each cancer type. **b**, Venn diagram showing the overlap between tumorigenesis-associated genes (genes considered upregulated in at least 4 out of 5 MYC-driven mouse tumor models) and MYC-correlated genes (genes with median Pearson’s r > 0.30 and RRA adjusted p‑value < 0.05). This 67-gene overlap constitutes the combined MYC gene signature. **c**, Barcode plot showing the representation of two distinct pathways (based on gene ontology) and embryonic stem cell genes (based on BioGPS Mouse Gene Atlas data) among the tumorigenesis-associated genes when the genes are ordered from low correlation with MYC expression (left) to high correlation with MYC expression (right). **d**, Protein-protein interaction (PPI) network of genes from the 67-gene MYC gene signature. **e**, Changes in mRNA expression upon MYC inactivation of five selected signature genes plus SLC46A3 in P493‑6 cells (left) and EC4 cells (right), a cell line derived from a transgenic mouse HCC tumor. Error bars represent mean ± SD. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, two-tailed one-sample Student's t test (n = 3 independent RT-qPCR experiments).

**Figure 5**. **Evaluation of the heterogeneity, MYC expression association, and prognostic significance among MYC gene signatures**

**a**, t‑distributed stochastic neighbor embedding (t-SNE) plot showing the clustering of cancer cell lines (from CCLE RNA-seq data) based on this study’s 67-gene signature. MYC expression tertiles are colored. **b**, Comparison of the empirical cumulative distribution of survival meta-z scores for the genes in the 67-gene signature with that for the reference distribution (i.e. the whole genome). The p-value was determined by the Wilcoxon rank sum test. **c**, UpSet plot showing the overlap among this study’s 67-gene MYC signature and four previously published MYC gene signatures. The four signatures shown are the signatures which shared the most genes with this study’s MYC gene signature. **d**, Box plots showing the quartiles of the Pearson correlation coefficients, which were computed pairwise between MYC expression and expressions of all genes in a given MYC gene set (or the whole genome for the reference distribution), using RNA‑seq data from CCLE (n = 1019 cell lines). **e**, Box plots showing the quartiles of TCGA survival meta‑z scores for various MYC gene signature gene sets.

**Figure 6. Proposed model for how MYC overexpression results in tissue-specific selective gene expression changes in tumorigenesis**

Graphical representation showing that MYC overexpression results in a tissue state which involves higher expression of embryonic stem cell-like genes and lower expression of tissue-lineage genes. This gene expression program change is likely the result of both direct effects of MYC on gene expression in its capacity as a transcription factor as well as epigenetic changes.

**Table 1. Transgenic mouse models used in gene expression profiling in this study**

**Supplemental Information titles and legends**

**Supplementary Figure 1.** Overview of the tetracycline-regulatory systems used to conditionally activate MYC in a tissue-specific manner. In these systems, the expression of the MYC transgene can be controlled by the administration of doxcycyline (dox).

**Supplementary Figure 2.** UpSet plots of the significantly differentially expressed genes among five different cancer types from MYC-induced transgenic mouse models.

**Supplementary Figure 3.** ChIP‑Seq H3K27ac signal changes, relative to normal tissue, in HCC tumors for superenhancers associated with BCL-differentially expressed genes and in BCL tumors for superenhancers associated with HCC-differentially expressed genes.

**Supplementary Figure 4.** Median log fold changes in gene expression for TCGA primary cancers with matched normal tissue plotted against MYC expression correlation.

**Supplementary Figure 5.** RT-qPCR validation of MYC inactivation after 24 hours and 48 hours of tetracycline administration in human P493-6 cells and mouse EC4 cells.

**Supplementary Table 1.** Differentially expressed genes among the transgenic mouse tumor models used in gene expression analysis in this study.

**Supplementary Table 2.** Full results from BioGPS Mouse Gene Atlas cell/tissue-type enrichment analysis of the differentially expressed genes in each of the transgenic mouse tumor models used in the study.

**Supplementary Table 3.** Genomic coordinate information and tissue type information about the superenhancers associated with the differentially expressed genes in the HCC and BCL transgenic mouse tumor models.

**Supplementary Table 4.** Full results from gene ontology GSEA performed on the expression profiles of the five MYC-induced transgenic mouse tumor models.

**Supplementary Table 5.** Pathway groupings of the top gene ontology terms selected from GSEA analysis of the five MYC-induced transgenic mouse tumor models.

**Supplementary Table 6.** Full results from enrichment of gene ontology terms among lung, liver, kidney, spleen, and embryonic stem cell tissue-lineage genes derived from the BioGPS Mouse Gene Atlas.

**Supplementary Table 7.** Pairwise Pearson correlation between each gene’s expression and MYC expression median-computed across 33 cancer types in TCGA’s RNA-seq dataset.

**Supplementary Table 8.** The MYC gene signatures that are highly correlated with MYC expression in TCGA human cancers and/or are upregulated in tumorigenesis in at least 4 out of 5 MYC-induced transgenic mouse tumor models.

**Supplementary Table 9.** Gene ontology analysis and cell/tissue-type enrichment analysis of the MYC tumorigenesis-associated genes.

**Supplementary Table 10.** Z-scores and meta-z scores derived from Cox proportional hazards survival analysis of TCGA pan-cancer clinical data and RNA-seq data.

**Supplementary Table 11.** Pairwise Pearson correlation between each gene’s expression and MYC expression median-computed across cancer cell lines in the CCLE RNA-seq dataset.

**Supplementary Table 12.** Published MYC signature gene sets analyzed in this study.

**Supplementary Table 13.** Primers used for RT-qPCR.