

Identification and DepMap CRISPR Dependency Analysis of a Pan-Cancer Epigenomic Signature of MYC-Driven Activation

Mohammed A H Alsaedi

Mohammedhalsaedi@gmail.com

Newcastle University <https://orcid.org/0009-0007-2776-2290>

Gordon Strathee

Newcastle University <https://orcid.org/0000-0001-9681-8429>

Research Article

Keywords: MYC, DNA methylation, DMRcate, epigenomics, TCGA, pan-cancer, DepMap, CRISPR dependency, survival

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11405571/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

While MYC is among the most frequently deregulated oncogenes in cancer, the extent to which MYC-associated gene activation is accompanied by DNA hypomethylation across diverse tumour types has remained an important, unresolved question. To address this, we integrated RNA-seq and DNA methylation data from MYC-high and MYC-low breast, liver, and colorectal tumours. Rigorously adjusting for proliferation, we successfully identified a conserved set of genes that are consistently upregulated and adjacent to hypomethylated regions. We further evaluated their clinical impact on survival and utilized public CRISPR screening data to determine their essentiality for cell viability. Our analysis uncovered a robust fifteen-gene signature, including MYC, characterized by concurrent upregulation and hypomethylation across all three malignancies. Highlighting their clinical significance, elevated expression of four of these genes significantly correlated with poorer overall survival in liver cancer. Furthermore, CRISPR dependency profiling revealed that MYC, BYSL, and BOP1 represent critical, broad vulnerabilities, as they were required for survival across nearly all cancer cell lines. Additionally, FOXK1 and SLC7A1 demonstrated essentiality in specific cellular contexts, revealing selective dependencies. Overall, this study establishes a strong, conserved fifteen-gene epigenomic signature associated with high MYC expression across multiple cancer types. These findings provide a compelling, evidence-based foundation to guide future functional assessments and delineate their precise roles in MYC-driven carcinogenesis.

1. Introduction

MYC family proto-oncogenes (MYC, MYCN, MYCL) are dysregulated in an estimated 70% of human cancers and represent one of the most studied yet therapeutically recalcitrant targets in oncology [1]. As a transcription factor, MYC drives a broad transcriptional programme encompassing cell cycle progression, ribosome biogenesis, nucleotide synthesis and metabolic reprogramming [2]. MYC also interacts with the epigenome: it recruits histone acetyltransferases such as TRRAP/GCN5 to target genes [3]. MYC-driven transcriptional activation is mechanistically coupled with targeted DNA demethylation at cis-regulatory elements to establish a transcriptionally permissive chromatin landscape [4]. This localized removal of repressive epigenetic marks is achieved by MYC orchestrating the recruitment and upregulation of Ten-Eleven Translocation (TET) dioxygenases, which sequentially oxidize 5-methylcytosine (5mC) into 5-hydroxymethylcytosine (5hmC) to facilitate active demethylation via base excision repair pathways [5]. A functional demonstration of this mechanism occurs during metabolic reprogramming, where oncogenic MYC actively forces the demethylation of specific promoter regions to ensure sustained accessibility and robust expression of target genes [6]. DNA methylation, predominantly at CpG dinucleotides, is one of the most stable and heritable epigenetic marks [7]. Hypomethylation at promoter-associated CpGs and cis-regulatory elements facilitates chromatin accessibility and transcription factor binding, thereby amplifying gene expression programmes [8,9]. In cancer, global DNA hypomethylation is a near-universal feature that promotes genomic instability, reactivation of transposable elements and aberrant expression of oncogenes [10]. Whether MYC-high

tumours show a reproducible pattern of focal hypomethylation that is linked to a shared set of upregulated genes in different cancer types has not been systematically examined.

MYC-associated expression programmes and methylation changes have each been studied, mostly within single cancer types. Adjustment for proliferation is important because MYC-high tumours tend to be more proliferative, and proliferation per se drives passive DNA demethylation, which confounds differential methylation analyses unless explicitly controlled for [11].

Transcriptional and epigenetic co-activation does not by itself show that a gene is functionally required for tumour cell fitness; a gene may be upregulated and hypomethylated as a bystander of a shared regulatory architecture (for example, proximity to an amplified locus). Large-scale functional genomic resources, such as the Cancer Dependency Map (DepMap) project [12,13], make it possible to assess this directly.

Here we compared MYC-high and MYC-low tumours in three TCGA cohorts (breast, hepatocellular and colorectal adenocarcinoma) to identify genes that are both upregulated and located near hypomethylated regions, using the DMRcate algorithm [14] for region-based methylation analysis. We examined the genes shared by all three cancers, and used public DepMap data to ask whether these genes are cell-line dependencies and whether any dependency is related to MYC expression.

2. Materials and Methods

2.1 Expression and DNA methylation datasets

Gene-level RNA-seq data and Illumina HumanMethylation450 beta values for breast invasive carcinoma (BRCA), liver hepatocellular carcinoma (LIHC) and colon adenocarcinoma (COAD) were downloaded from the TCGA GDC Data Portal (<https://portal.gdc.cancer.gov>). For expression files, the first column matching 'unstranded' (unstranded read counts) was used. Gene symbols were taken from the file annotation or, where only Ensembl identifiers were present, mapped with org.Hs.eg.db; where a symbol was duplicated the first occurrence was kept. Methylation beta values were assembled across files over the union of probe identifiers. File UUIDs were converted to sample barcodes through the GDC REST API (batches of 200). Samples with sample codes 11 to 13 (normal tissue) were removed. Expression and methylation samples were matched at patient level using the first 12 characters of the barcode; where a patient had several samples the primary tumour (code 01) was preferred, otherwise the first available sample was used. After matching, 789 (BRCA), 371 (LIHC) and 295 (COAD) patients had both expression and methylation data. Vital status, days to death and days to last follow-up were taken from the GDC clinical files.

2.2 MYC stratification, proliferation score and differential expression

Within each cancer type, tumours were ranked by MYC expression. Tumours at or above the 75th percentile (MYC-high) and at or below the 25th percentile (MYC-low) were retained and the middle 50% excluded. A proliferation score was calculated for each tumour as the mean raw count of 26 E2F and S-phase genes (E2F1, E2F2, E2F3, E2F7, E2F8, CDC6, CDK1, CCNE1, CCNE2, CCNA2, MCM2 to MCM7, PCNA, RRM1, RRM2, TK1, TYMS, CDC25A, CDC25C, DHFR, POLA1, POLE). Tumour purity was not included as a covariate in the expression or methylation models; consensus purity estimates (CPE) from the TCGAblinks Tumor.purity dataset were matched to the stratified tumours by sample barcode and are summarised for reference in Supplementary Table S5. Genes with a mean count of 1 or less across the stratified tumours were removed and counts were transformed as $\log_2(\text{count} + 1)$. No library-size normalisation was applied. Differential expression was tested with limma (lmFit, eBayes) using the design $\sim \text{MYC group} + \text{proliferation score}$, extracting the MYC-group coefficient with Benjamini-Hochberg (BH) adjustment. Genes with $\log_2$ fold change of at least 0.5 and adjusted $P < 0.05$ were classed as upregulated in MYC-high tumours.

2.3 Differentially methylated region analysis

Probes with 20% or more missing values were removed (84,134 of 486,435 probes in BRCA, 92,071 of 486,437 in LIHC and 75,115 of 486,451 in COAD, i.e. 17.3%, 18.9% and 15.4%, calculated in the stratified tumours), remaining missing beta values were replaced by the probe mean, and only probes present in IlluminaHumanMethylation450kanno.ilmn12.hg19 were retained. Beta values were converted to M-values (logit transformation with an offset of 1×10^{-5}). CpG sites were tested with DMRcate::cpg.annotate (array, 450K, differential analysis, M-values) using the same samples and design matrix as the expression model, with a CpG-level FDR of 0.05. DMRs were called with DMRcate::dmrcate ($\lambda = 1,000$, $C = 2$, default minimum of two CpGs, no minimum methylation-difference threshold) and converted to genomic ranges on hg19 with extractRanges. If the FDR-based call failed, as implemented in the deposited script, CpG sites were re-annotated and DMRs called with an unadjusted P cutoff of 0.05; in all three cancer types every reported DMR had a minimum smoothed FDR below 0.001. DMRs with a negative mean methylation difference (lower methylation in MYC-high tumours) were classed as hypomethylated. Genes were assigned to a DMR if they were named in the DMRcate overlap annotation or if their gene body in TxDb.Hsapiens.UCSC.hg19.knownGene overlapped the DMR extended by 10 kb on each side (GenomicRanges; Entrez identifiers converted to symbols with org.Hs.eg.db).

2.4 Intersection of expression and methylation results

Within each cancer type, candidate genes were those that were upregulated in MYC-high tumours and assigned to a hypomethylated DMR (intersection at gene-symbol level). The signature comprised genes that were candidates in all three cancer types. To test whether overlaps exceeded chance, the universe was defined as the genes tested in the expression analysis of that cancer type. For each cancer type, overlap between upregulated and DMR-proximal genes was tested with a one-sided hypergeometric test. The DMR-proximal gene lists could not be restricted to genes that were also tested in the expression analysis, so all DMR-proximal genes were counted when calculating the expected overlap. This overstates the expected overlap and makes the test conservative. Pairwise overlaps between cancer

types were tested in the same way. For the three-way overlap, the number of genes expected by chance was calculated assuming independence, and significance was assessed by Monte Carlo permutation (10,000 random draws of gene sets of the observed sizes from the genes tested in all three cohorts). Because MYC was selected by study design, the three-way overlap was also assessed for the fourteen other signature genes.

2.5 Co-expression analysis

Pearson correlations were calculated between all pairs of the fifteen signature genes in primary tumours (sample code 01) using log₂-transformed unstranded counts (log₂[count + 1]), visualised as hierarchically clustered heatmaps with pheatmap, and summarised as the mean absolute correlation over the upper triangle (mean |r|).

2.6 Survival analysis

Overall survival (OS) was analysed in primary tumours where data for the diagnostic sample were available, with OS time greater than 0 days and at least 20 patients per analysis. Patients were dichotomised at the median expression of each gene (at or above the median = high). Kaplan–Meier curves were fitted with the survival package, groups were compared by log-rank test, and univariate Cox proportional hazards models were fitted on continuous log₂ transformed expression (log₂[count + 1]). P-values across all 45 gene-cancer tests were adjusted by the Benjamini–Hochberg (BH) method.

2.7 DepMap CRISPR dependency analysis

Genome-wide CRISPR-Cas9 knockout data (CRISPRGeneEffect.csv for Chronos gene-effect scores and CRISPRGeneDependency.csv for dependency probabilities), the common-essential gene reference list (CRISPRInferredCommonEssentials.csv), cell-line metadata (Model.csv) and RNA-seq expression were obtained from the public DepMap release 26Q3. Chronos scores estimate the fitness effect of gene knockout (0 = no effect; -1 = approximately the median effect of pan-essential genes). All DepMap cell lines classified as breast (n = 53), colon (n = 63) or liver (n = 53) that had CRISPR data were analysed (169 cell lines). A gene was classed as a dependency in a line if Chronos < -0.5, and dependency is reported as the fraction of cell lines meeting this threshold. Ten of the fifteen signature genes are represented in the CRISPR library; the five non-coding RNA genes (PVT1, CASC11, SNHG16, SNHG17, SNORA71B) were absent. For context, mean Chronos was also calculated in the 1,039 DepMap CRISPR-profiled cell lines from other cancer types.

To determine whether dependency was associated with MYC expression, MYC mRNA levels (log₂[TPM + 1]) were obtained for the 163 cell lines profiled in both CRISPR and expression datasets. For each gene, Spearman rank correlations (ρ) between Chronos scores and MYC expression were calculated across the 163 cell lines and within each cancer type, with P-values adjusted using the Benjamini–Hochberg (BH) method. In addition, Chronos scores were compared between the highest and lowest quartiles of MYC expression using Wilcoxon rank-sum tests. A MYC-selective dependency was expected to demonstrate significantly more negative Chronos scores at higher MYC expression.

2.8 Tumour-versus-normal discrimination

Receiver operating characteristic (ROC) analysis was performed for each signature gene to discriminate primary tumours from solid tissue normal samples (unpaired), using the pROC package with automatic direction (yielding AUC > 0.5) and 95% confidence intervals calculated via ci.auc. To benchmark the mean AUC of the fifteen signature genes, 1,000 random sets of fifteen genes were sampled, each without replacement, from all genes in the expression matrix using a fixed random seed (set.seed(123)) for reproducibility. The empirical p-value was calculated as the proportion of random gene sets achieving a mean AUC at least as high as that observed for the signature.

2.9 Functional enrichment analysis

Cancer-specific candidate gene lists (upregulated in MYC-high tumours and adjacent to hypomethylated DMRs) were converted to Entrez identifiers with org.Hs.eg.db and tested for over-representation of Gene Ontology Biological Process (BP) and Molecular Function (MF) terms (clusterProfiler::enrichGO) and KEGG pathways (enrichKEGG, organism hsa). P values were adjusted by the BH method (pvalueCutoff = qvalueCutoff = 0.05) using the clusterProfiler default gene-set size limits (10 to 500 genes), and gene symbols were restored with readable = TRUE or setReadable. No custom background was supplied, so all annotated genes formed the universe, and no redundancy reduction was applied. Fold enrichment is the GeneRatio divided by the BgRatio. Terms with adjusted $p < 0.05$ were considered significant. Enrichment was applied to the cancer-specific candidate lists rather than to the fifteen-gene signature, which is too small for over-representation testing.

2.10 Statistical analysis and reproducibility

All computational analyses were performed in R (version 4.3.1) using two-sided statistical tests and Benjamini–Hochberg (BH) false discovery rate correction unless otherwise specified. Analyses utilized Bioconductor and CRAN packages including TCGAbiolinks (v2.28.0), limma (v3.56.0), DMRcate (v2.14.0), minfi (v1.46.0), GenomicRanges (v1.52.0), org.Hs.eg.db (v3.17.0), TxDb.Hsapiens.UCSC.hg19.knownGene (v3.2.2), clusterProfiler (v4.8.3), survival (v3.5-5), survminer (v0.4.9), pheatmap (v1.0.12), and pROC (v1.18.4). Random permutation workflows incorporated fixed random seeding (set.seed(123)) for exact reproducibility. All analysis code is openly available on Zenodo (<https://doi.org/10.5281/zenodo.23352617>). Primary TCGA dataset files are accessible via the GDC Data Portal (<https://portal.gdc.cancer.gov>) and DepMap functional genomics data via the DepMap Data Portal (<https://depmap.org/portal/>).

3. Results

3.1 Upregulated and hypomethylated gene sets in MYC high expressing tumours

We compared MYC-high (top quartile) and MYC-low (bottom quartile) tumours within each cancer type, adjusting for proliferation score, to identify genes with higher expression and DMRs with lower methylation in MYC-high tumours, and then intersected the two lists (Fig. 1). The stratified groups comprised 198 MYC-high and 198 MYC-low tumours in BRCA, 93 and 93 in LIHC, and 74 and 74 in COAD (Supplementary Table S5). Median proliferation score was significantly higher in MYC-high tumours in BRCA and COAD ($P < 0.001$), but not in LIHC; the adjustment was applied identically in all three cohorts (Supplementary Fig. 6). While MYC is upregulated in MYC-high tumours by construction, it was included in the intersected lists if it was identified as lying near a hypomethylated DMR.

BRCA. Of 34,920 genes tested in 198 MYC-high and 198 MYC-low tumours, 6,500 were upregulated in MYC-high tumours ($\log_2FC \geq 0.5$, $FDR < 0.05$) and 1,323 were downregulated (Fig. 1A). Methylation analysis identified 9,571 DMRs that were hypomethylated in MYC-high tumours; these overlapped or lay within 10 kb of 12,013 genes. A total of 2,588 genes (39.8% of the upregulated genes) were present in both lists, 1.16-fold more than the 2,236 expected by chance (hypergeometric $P = 3 \times 10^{-24}$).

LIHC. Of 29,928 genes tested in 93 MYC-high and 93 MYC-low tumours, 5,905 were upregulated and 1,080 downregulated. There were 739 hypomethylated DMRs, mapping to 1,441 genes, and 366 genes (6.2% of the upregulated genes) were present in both lists, 1.29-fold more than the 284 expected ($P = 4 \times 10^{-8}$).

COAD. Of 32,368 genes tested in 74 MYC-high and 74 MYC-low tumours, 3,561 were upregulated and 553 downregulated. There were 6,911 hypomethylated DMRs, mapping to 8,473 genes, and 994 genes (27.9% of the upregulated genes) were present in both lists, 1.07-fold more than the 932 expected ($P = 0.007$).

The number of hypomethylated DMRs differed markedly between cancer types (739 to 9,571). This may reflect differences in statistical power, as the number of DMRs increased with cohort size, and in the extent of methylation differences associated with MYC expression in each tissue, but it was not investigated further. All DMRs were significant (minimum smoothed $FDR < 0.001$ for every region), although the methylation differences were generally modest (median mean beta-value difference of 0.04 in BRCA, 0.05 in LIHC and 0.09 in COAD), as no minimum methylation-difference threshold was applied. Within each cancer the overlap between upregulated genes and DMR-proximal genes was only modestly greater than expected by chance (1.07- to 1.29-fold). This is because a large proportion of the genes tested lay near a hypomethylated DMR, particularly in BRCA, so that a substantial overlap would be expected by chance. The expected overlaps are also likely to be overestimated (Methods 2.4), so these enrichments are conservative.

Shared genes. Among the 27,396 genes tested in all three cohorts, pairwise overlaps between the cancer-specific candidate lists were 92 (BRCA-LIHC), 215 (BRCA-COAD) and 44 (LIHC-COAD) genes, compared with 33, 86 and 13 expected by chance (all $P < 1 \times 10^{-12}$; Fig. 1C). Fifteen genes (MYC plus fourteen others) were present in the candidate lists of all three cancers. Because MYC is included largely by study design, we also considered the fourteen other genes, which alone were still about 13-fold more

than the 1.1 expected by chance (permutation $P < 1 \times 10^{-4}$; the largest overlap in 10,000 random draws was 7). These fifteen genes form the signature analysed below (Table 1). The signature includes genes involved in ribosome biogenesis (BYSL, BOP1), chromatin remodelling (CHD7, MLLT10, FOXK1), amino acid transport (SLC7A1), transcriptional regulation (ZBTB10, FOXK1), long non-coding RNAs (CASC11, SNHG16, SNHG17, PVT1) and a small nucleolar RNA (SNORA71B). Four genes (MYC, PVT1, CASC11 and BOP1) map to the 8q24 locus, so their co-detection may partly reflect locus-level copy-number gain or shared chromatin state rather than independent regulation.

Table 1

The fifteen genes upregulated in, and adjacent to hypomethylated regions in, MYC-high tumours of all three cancer types.

Gene	Functional Category	Chromosome	Notable Role
BYSL	Ribosome biogenesis	6p21.1	rRNA processing; proliferation marker
BOP1	Ribosome biogenesis	8q24.3	Pre-rRNA processing; MYC co-amplification region
PVT1	Long non-coding RNA	8q24.21	MYC co-amplification; oncogenic lncRNA
MYC	Transcription factor	8q24.21	Master oncogenic regulator; index gene
CASC11	Long non-coding RNA	8q24.21	Cancer susceptibility lncRNA (8q24 locus)
SNHG16	snoRNA host gene	17q25.1	Oncogenic lncRNA; promotes proliferation
SNHG17	snoRNA host gene	20q11.23	Promotes tumour growth and invasion
SNORA71B	Small nucleolar RNA	20q11.22	snoRNA; rRNA modification
CHD7	Chromatin remodelling	8q12.2	SWI/SNF-related helicase; super-enhancer regulation
MLLT10	Chromatin remodelling	10p12.31	MLL fusion partner; H3K4 methylation
FOXK1	Transcription factor	7p22.1	Represses autophagy; promotes glycolysis
ZBTB10	Transcription factor	8q21.13	Sp/KLF repressor
SLC7A1	Amino acid transporter	13q12.3	Cationic amino acid transporter; arginine uptake
PRKAR1B	Signalling regulator	7p22.3	PKA regulatory subunit
FCHSD1	Cytoskeletal regulation	5q35.1	Actin dynamics; endocytosis

3.2 Expression of the fifteen genes in MYC-high versus MYC-low tumours and in normal tissue

To assess how consistent and how large the differences between MYC-high and MYC-low tumours were, we compared the expression of each signature gene between the two groups separately in each cancer type (Fig. 2). MYC itself, which defines the groups, showed by far the largest difference in all three cancers. For the other fourteen genes the differences were much smaller, and for some genes (FO XK1, MLLT10 and FCHSD1) they were close to the selection threshold in all three cancers. Differences were generally largest in LIHC, and the ranking of genes by fold change was only partly conserved across cancer types. The non-coding genes CASC11 and SNORA71B showed some of the larger fold changes but were expressed at low absolute levels, so the absolute differences are small and their biological relevance is uncertain. By contrast, highly expressed genes such as BOP1, SLC7A1 and CHD7 showed consistent increases at high absolute expression.

We also asked how the signature genes behaved in tumour versus normal tissue, using primary tumours and solid tissue normal samples. Most signature genes discriminated tumours from normal samples, although performance varied widely between genes and cancer types (Supplementary Fig. 2 and Supplementary Table S2): BOP1, PVT1 and BYSL were consistently strong in all three cancers, whereas PRKAR1B discriminated poorly. As a set, the fifteen genes discriminated tumour from normal tissue better than random sets of fifteen genes in all three cancers (Supplementary Fig. 3 and Supplementary Table S3). This is not specific to MYC status: because the genes were selected for upregulation in MYC-high tumours and many genes differ between tumour and normal tissue, these results show that the genes are tumour-associated but do not establish diagnostic value, which would require independent cohorts.

3.3 Co-expression of the signature genes

We next asked whether the fifteen genes behave as a coordinated module in tumours and how closely each tracks MYC expression. Correlations were calculated across all primary tumours with RNA-seq data in each cohort ($n = 1,111$ in BRCA, 371 in LIHC, and 481 in COAD), rather than being restricted to the MYC-high/MYC-low groups (Fig. 3).

The mean absolute correlation between the signature genes was highest in COAD (0.428), followed by LIHC (0.338) and BRCA (0.189). As expected, all fifteen genes were positively correlated with MYC expression in every cohort ($r = 0.01$ to 0.69), although for some genes the correlation was relatively weak when all samples were included (as opposed to just the highest and lowest MYC quartiles), such as SNORA71B ($r = 0.01$ to 0.25 across the three cancer types). Differences between the cohorts, for example in tumour purity, could contribute to the differences in co-expression but were not assessed. Co-expression was not tested against a random background, and we do not interpret it as validation of the signature.

3.4 Association of signature gene expression with overall survival

To test whether expression of the signature genes is associated with outcome, patients in each cohort were dichotomised at median expression and overall survival compared by log-rank test (15 genes in each of three cancers; 45 tests). Nine associations had log-rank $P < 0.05$ before correction for multiple testing (Table 2; all 45 results in Supplementary Table S1), and in every case high expression was associated with poorer survival. After BH correction for multiple testing across all 45 tests, four associations in LIHC remained significant: FOXK1, SNHG17, BYSL and BOP1 (Fig. 4). Continuous univariate Cox proportional hazards regression on log2-transformed expression ($\log_2[\text{count} + 1]$) confirmed that each log2 unit increase in expression of these four genes in LIHC was associated with significantly elevated mortality risk (hazard ratios 1.51 to 1.74). MYC mRNA expression itself was not significantly associated with overall survival in any cancer type (Supplementary Table S1), indicating that these prognostic associations do not simply act as surrogates for MYC transcript levels.

We next evaluated whether these prognostic associations were specific to MYC-high tumours. Stratified analysis within the MYC-high and MYC-low LIHC subgroups revealed that elevated expression of FOXK1, SNHG17, BYSL, and BOP1 conferred increased mortality risk in both subgroups, and no significant gene x MYC-group interaction was observed (P interaction > 0.05 for all four genes). These associations are therefore not exclusive to MYC-high tumours.

Table 2

Genes with log-rank $p < 0.05$ for overall survival (median dichotomisation) in each cancer type. q: BH-adjusted p across all 45 gene-cancer tests.

Cancer	Gene	N	Log-rank P	BH q (45 tests)	Direction
LIHC	SNHG17	366	0.0004	0.013	Higher expression, worse OS
LIHC	FOXK1	366	0.0006	0.013	Higher expression, worse OS
LIHC	BYSL	366	0.0017	0.025	Higher expression, worse OS
LIHC	BOP1	366	0.0036	0.040	Higher expression, worse OS
LIHC	CASC11	366	0.007	0.063	Higher expression, worse OS
LIHC	SNHG16	366	0.017	0.125	Higher expression, worse OS
LIHC	MLLT10	366	0.021	0.133	Higher expression, worse OS
COAD	BYSL	457	0.026	0.15	Higher expression, worse OS
BRCA	CHD7	1,089	0.041	0.21	Higher expression, worse OS

3.5 CRISPR dependency on signature genes and relationship to MYC expression

We used DepMap CRISPR screens in 169 breast (n = 53), colon (n = 63) and liver (n = 53) cell lines to begin to assess if any of the signature genes are required in MYC high expressing cell lines. Ten signature genes (MYC, BYSL, BOP1, SLC7A1, CHD7, MLLT10, FOXK1, ZBTB10, PRKAR1B, FCHSD1) are

protein coding and are included in the DepMap CRISPR library; the five non-coding RNA genes are not included.

MYC, BYSL and BOP1 were dependencies in almost every line: MYC in 53/53 breast, 62/63 colon and 53/53 liver cell lines (mean Chronos – 1.99, – 2.12 and – 2.09), BYSL in 53/53, 63/63 and 53/53 (– 1.26, – 1.35, – 1.38), and BOP1 in 52/53, 62/63 and 53/53 (– 1.21, – 1.40, – 1.31). Dependency for these genes was also not limited to these cancer types and mean Chronos in the 1,039 cell lines from other cancer types was – 2.01 for MYC, –1.31 for BYSL and – 1.33 for BOP1, indicating these genes are essential across most cancer cell lines. Consequently, their dependency is not specifically selective for cell lines with high MYC expression. Consistent with this, Chronos scores for BYSL and BOP1 were not significantly correlated with MYC expression across the 163 cell lines (BYSL rho = 0.02, BOP1 rho = 0.03; BH-adjusted P = 0.84), and MYC dependency itself was also not related to MYC expression levels (rho = 0.08, P = 0.31).

SLC7A1 and FOXK1 were dependencies in a minority of cell lines: SLC7A1 in 6/53 breast, 13/63 colon and 7/53 liver cell lines (26/169, 15.4%) and FOXK1 in 16/53 breast, 7/63 colon and 11/53 liver cell lines (34/169, 20.1%). SLC7A1 dependency was unrelated to MYC expression (Spearman rho = – 0.04, P = 0.57). FOXK1 dependency showed a significant relationship in the direction opposite to a MYC-driven dependency: higher MYC expression was associated with less negative FOXK1 Chronos scores (rho = + 0.26, P = 7.5×10^{-4} ; BH-adjusted P = 0.007 across the nine genes tested, with none of the other genes reaching P < 0.05) (Fig. 6). FOXK1 dependency is therefore not explained by high MYC expression and, if anything, is more evident in cell lines with lower MYC expression. In addition, comparing FOXK1 Chronos scores between the highest and lowest MYC expression quartiles confirmed stronger dependency in MYC-low cell lines (– 0.28 in top quartile vs. –0.46 in bottom quartile, P = 0.0021).

The remaining five genes were not identified as dependencies in any or only a very small number of cell lines (CHD7–2/169).

3.6 Functional annotation of cancer-specific candidate genes

To characterise the genes that were both upregulated and adjacent to hypomethylated regions, each cancer-specific candidate list was subjected to Gene Ontology terms and KEGG pathway analysis (Supplementary Figs. 4 and 5, Supplementary Table S4). In LIHC, the candidates were most strongly enriched for translation and ribosome biogenesis, the theme to which BYSL, BOP1, CHD7 and MYC contribute. COAD candidates were enriched for transport, stem-cell maintenance and transcription-factor activity, and BRCA candidates mainly for developmental and signalling processes. The wide variety of pathways highlighted in this analysis is likely a reflection of the broad transcriptional programme regulated by MYC and reflective of tissue specific differences in MYC function.

4. Discussion

We combined RNA-seq and 450K methylation data from three TCGA cohorts, rigorously adjusting for proliferation in both the expression and methylation models, to investigate whether MYC-associated gene activation is accompanied by local DNA hypomethylation. We successfully identified a robust signature of fifteen genes, including MYC, that were consistently upregulated and adjacent to hypomethylated regions in MYC-high tumours of breast, liver, and colorectal cancer. While a large fraction of genes generally resides near hypomethylated DMRs, the specific overlap of this fifteen-gene signature across all three distinct cancer types was significantly greater than expected by chance, highlighting a conserved pan-cancer mechanism. This is in line with earlier reports that MYC activity is linked to DNA demethylation at its target loci [4–6], and suggests that a limited set of genes shows this combination of upregulation and nearby hypomethylation in breast, liver and colorectal cancer. Although the broader landscape of hypomethylation exhibits tissue-specific variation across these cancers, this core fifteen-gene set represents a powerful, conserved epigenomic convergence in MYC-driven malignancies.

The shared genes point to biology that is familiar from MYC studies. BYSL and BOP1 are pre-rRNA processing factors, consistent with ribosome biogenesis being a major MYC effector [2,4], and CHD7, MLLT10 and FOXK1 connect the signature to chromatin and transcriptional regulation. Four genes (MYC, PVT1, CASC11 and BOP1) lie at 8q24 [15,16], a region that is often gained in cancer. Future integration of copy number data will be highly valuable to delineate the precise contributions of chromosomal gain versus coordinate epigenetic upregulation at this critical locus. Importantly, across the full cohorts, all fifteen genes were positively correlated with MYC, demonstrating that these associations extend robustly across the entire spectrum of MYC expression, with varying degrees of correlation. Furthermore, these genes demonstrated notable co-expression, particularly in colorectal cancer, providing a strong rationale for exploring shared regulatory mechanisms in future functional studies.

Through DepMap analysis, we evaluated the functional essentiality of the signature genes across cancer cell lines. MYC, BYSL and BOP1 were required in nearly all breast, colon and liver cell lines, and also in other lineages, which supports ribosome biogenesis as a core requirement of cancer cells and fits the reported role of BOP1 in colorectal tumorigenesis [17]. Rather than being restricted exclusively to MYC-high contexts, these near-universal dependencies highlight fundamental vulnerabilities inherent to general cancer cell survival. Furthermore, FOXK1 and SLC7A1 emerged as dependencies in a distinct subset of cell lines independent of high MYC expression, indicating they serve as important context-dependent vulnerabilities. While these dependencies are not exclusively selective for MYC-driven cells, their essentiality underscores their broader, critical roles within oncogenic networks. While CHD7, MLLT10, ZBTB10, PRKAR1B and FCHSD1 did not show acute essentiality in two-dimensional single-gene knockout screens, they remain compelling candidates for in vivo investigation where complex tumour microenvironment interactions are preserved. The five non-coding genes in our signature await functional characterization in future targeted screens, as they are not currently represented in the DepMap CRISPR library. One of these, PVT1, has previously been reported to be required in cancers with MYC copy-number gain [16]. This provides a clear mandate for mapping the functional roles of the remaining four non-coding transcripts. Highlighting the clinical significance of our signature, four genes

(FO XK1, SNHG17, BYSL and BOP1) were strongly associated with poorer overall survival in liver cancer, yielding notable hazard ratios of 1.51 to 1.74 per log₂ unit of expression. Crucially, all nine nominally significant associations indicated that higher expression correlates with worse survival. Because these robust prognostic associations were present in both MYC-high and MYC-low LIHC tumours without a significant interaction and MYC mRNA itself was not associated with survival these signature genes hold distinct, independent prognostic value beyond standard MYC transcript quantification. This is consistent with reports that FO XK1 promotes hepatocellular carcinoma progression [18]. These compelling univariate findings lay a strong foundation for future validation in independent cohorts utilizing multivariable models adjusted for stage, age, and aetiology.

Our tumour-versus-normal analysis demonstrated that this signature gene set is significantly more discriminatory than random gene sets, reflecting a powerful tumour-associated expression profile. Among these, PVT1 and BOP1 displayed exceptionally consistent performance, establishing them as highly promising candidates for diagnostic evaluation in independent clinical cohorts.

The rigorous methodological design of this study reinforces the validity of the findings. By effectively adjusting for proliferation, we successfully isolated genes associated with MYC independently of its core proliferative functions. While this highly conservative approach provides a highly specific target set, the broader landscape of MYC-regulated genes is likely even more extensive. Tumour purity was exceptionally well-documented, with estimates available for over 99% of the stratified cohorts. While purity did not differ significantly between MYC-high and MYC-low tumours in BRCA or LIHC, we noted a marginal but statistically significant purity difference in MYC-high COAD tumours (median CPE 0.795 vs 0.751, $P = 0.002$; Supplementary Table S5). This offers an intriguing avenue to explore how variations in tumour cellularity and stromal content might further shape expression and methylation profiles in colorectal cancer.

The associative strength of this study paves the way for targeted mechanistic follow-up. Future experimental models will be essential to definitively delineate the direct causality between hypomethylation and coordinate upregulation, particularly distinguishing epigenetic drivers from copy number gains at loci like 8q24. Furthermore, leveraging higher-resolution methylation sequencing beyond the standard 450K array coverage will undoubtedly uncover additional regulatory elements and further refine the biological impact of the observed methylation variances. Ultimately, this study establishes a strong, proliferation-adjusted foundational framework. These findings provide a compelling rationale for copy-number-aware analyses in independent cohorts and targeted in vitro perturbations of MYC and DMR methylation. The identification of this fifteen-gene signature successfully highlights BYSL and BOP1 as critical ribosome biogenesis dependencies and elevates FO XK1 and SLC7A1 as valuable, context-dependent therapeutic candidates.

5. Conclusions

We successfully defined a robust, conserved fifteen-gene epigenomic signature characterized by concurrent upregulation and hypomethylation in MYC-driven breast, liver, and colorectal cancers, independent of proliferation status. Highlighting the clinical relevance of this signature, elevated expression of FO XK1, SNHG17, BYSL, and BOP1 strongly correlated with poorer overall survival in liver cancer. Crucially, CRISPR dependency profiling revealed that BYSL and BOP1, alongside MYC, represent fundamental, broad-spectrum vulnerabilities essential for survival across diverse cancer cell lineages. Furthermore, FO XK1 and SLC7A1 were required for survival in a distinct subset of cell lines, independent of high MYC expression, indicating they act as context-dependent vulnerabilities within broader oncogenic networks.

Declarations

Funding

M.A.H.A. is supported by a PhD scholarship from the Ministry of Education, Saudi Arabia, to study at Newcastle University.

Competing interests

The authors declare no competing interests.

Author contributions

M.A.H.A. conceived the analysis, performed all computational analyses, and wrote the manuscript. G.S. supervised the study, contributed to the study design and interpretation, and edited the manuscript. Both authors read and approved the final manuscript.

Data availability

All data analysed are publicly available. TCGA RNA-seq, 450K methylation and clinical data were obtained from the GDC Data Portal (<https://portal.gdc.cancer.gov>), and CRISPR dependency and expression data from the DepMap Data Portal (<https://depmap.org/portal/>). Derived results are provided in the Supplementary Information.

Code availability

All analysis code is openly available on Zenodo (<https://doi.org/10.5281/zenodo.23352617>).

Ethics statement

This study used only publicly available, de-identified data from TCGA and DepMap; no new data were collected from human participants and no additional ethical approval was required.

References

1. Dang C V. MYC on the Path to Cancer. *Cell* 149(1), 22–35 (2012).
2. Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang C V. MYC, Metabolism, and Cancer. *Cancer Discov.* 5(10), 1024–1039 (2015).
3. Kenneth NS, Ramsbottom BA, Gomez-Roman N, Marshall L, Cole PA, White RJ. TRRAP and GCN5 are used by c-Myc to activate RNA polymerase III transcription. *Proc. Natl. Acad. Sci. U. S. A.* 104(38), 14917–14922 (2007).
4. Poole CJ, van Riggelen J. MYC—Master Regulator of the Cancer Epigenome and Transcriptome. *Genes* 2017, Vol. 8, Page 142 8(5), 142 (2017).
5. Poole CJ, Lodh A, Choi JH, Van Riggelen J. MYC deregulates TET1 and TET2 expression to control global DNA (hydroxy)methylation and gene expression to maintain a neoplastic phenotype in T-ALL. *Epigenetics & Chromatin* 2019 12:1 12(1), 41- (2019).
6. Bott AJ, Peng IC, Fan Y et al. Oncogenic Myc Induces Expression of Glutamine Synthetase through Promoter Demethylation. *Cell Metab.* 22(6), 1068–1077 (2015).
7. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13(7), 484–492 (2012).
8. Aran D, Hellman A. DNA methylation of transcriptional enhancers and cancer predisposition. *Cell* 154(1), 11 (2013).
9. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015 518:7539 518(7539), 317–330 (2015).
10. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* 1983 301:5895 301(5895), 89–92 (1983).
11. Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. *Nature Reviews Cancer* 2011 11:10 11(10), 726–734 (2011).
12. Dempster JM, Boyle I, Vazquez F et al. Chronos: a cell population dynamics model of CRISPR experiments that improves inference of gene fitness effects. *Genome Biol.* 22(1) (2021).
13. Tsherniak A, Vazquez F, Montgomery PG et al. Defining a Cancer Dependency Map. *Cell* 170(3), 564–576.e16 (2017).
14. Peters TJ, Buckley MJ, Statham AL et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin* 2015 8:1 8(1), 6- (2015).
15. Dang C V. MYC, metabolism, cell growth, and tumorigenesis. *Cold Spring Harb. Perspect. Med.* 3(8) (2013).

16. Tseng YY, Moriarity BS, Gong W *et al.* PVT1 dependence in cancer with MYC copy-number increase. *Nature* 2014 512:7512 512(7512), 82–86 (2014).
17. Killian A, Sarafan-Vasseur N, Sesboüé R *et al.* Contribution of the BOP1 gene, located on 8q24, to colorectal tumorigenesis. *Genes Chromosomes Cancer* 45(9), 874–881 (2006).
18. Liu J, Zhang M, Li J, Tian Y, Zhao J, Liu X. FOXK1 induced upregulation of KIF20A promotes hepatocellular carcinoma progression via Wnt/ β -Catenin/EMT signaling. *Cell. Mol. Life Sci.* 83(1) (2026).

Figures

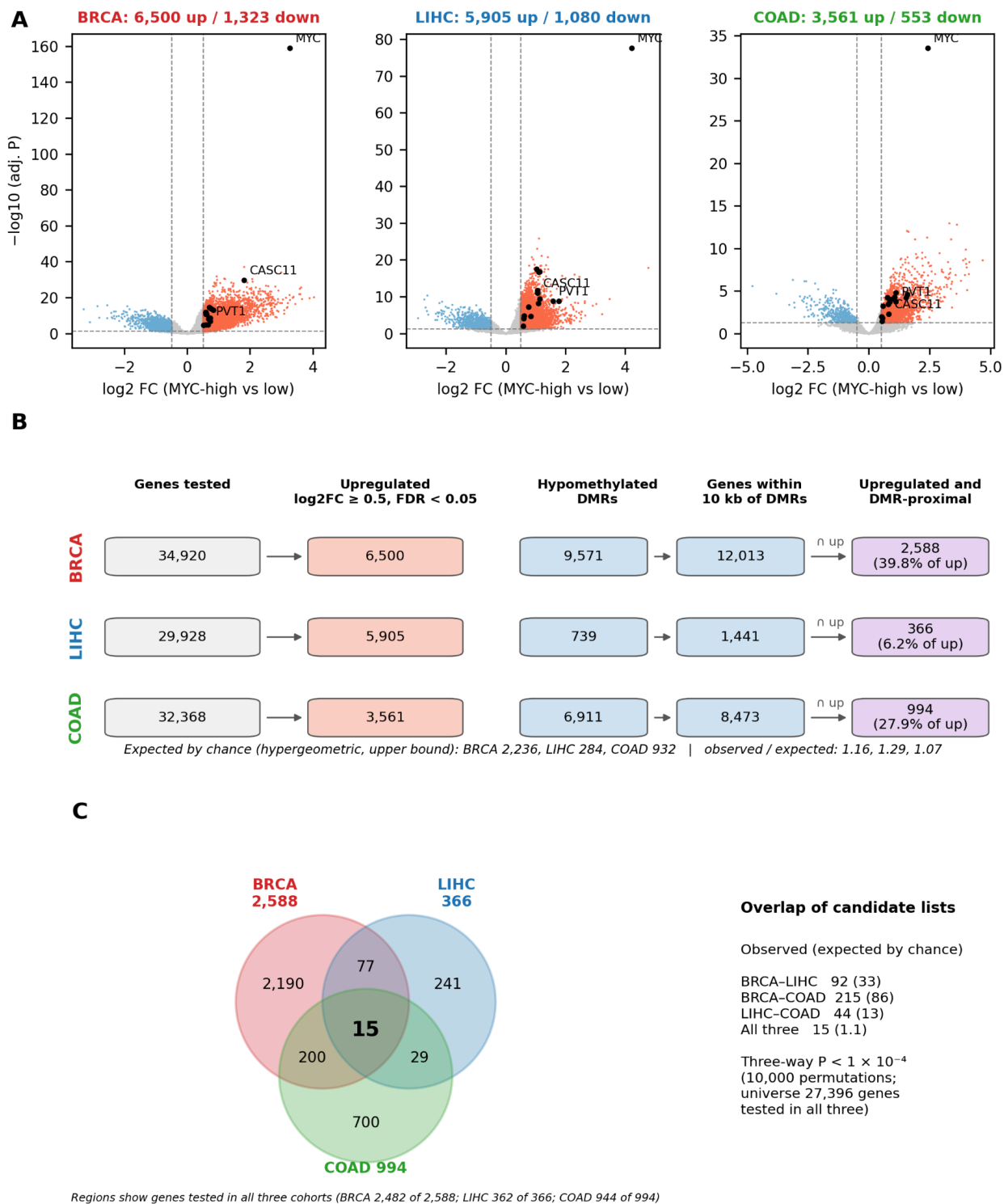


Figure 1

Identification of the MYC high expression signature. (A) Volcano plots of differential expression between MYC-high and MYC-low tumours in each cancer type (limma, adjusted for proliferation score); red, upregulated ($\log_2\text{FC} \geq 0.5$, $\text{FDR} < 0.05$); blue, downregulated; black, the fifteen signature genes; dashed lines mark the thresholds. (B) Number of genes selected at each stage of the analysis in each cancer

type. (C) Overlap of the three cancer-specific candidate lists (regions show genes tested in all three cohorts); the central region is the fifteen-gene signature.

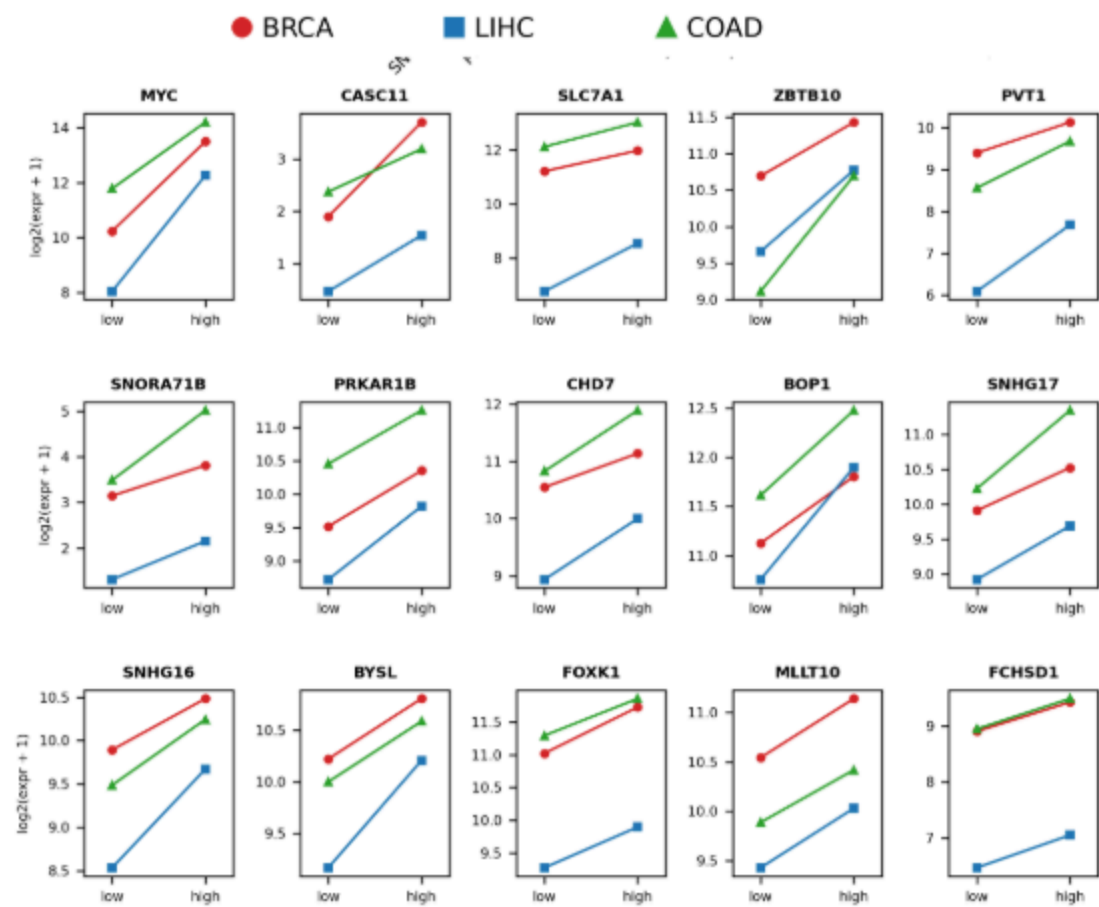


Figure 2

Expression of the fifteen signature genes in MYC-low and MYC-high tumours. Estimated mean $\log_2(\text{expression} + 1)$ in BRCA (red circles), LIHC (blue squares) and COAD (green triangles), calculated as the average expression minus or plus half the model $\log_2$ fold change (limma, adjusted for proliferation score; assuming equal group sizes). Genes are ordered by mean fold change across the three cancers.

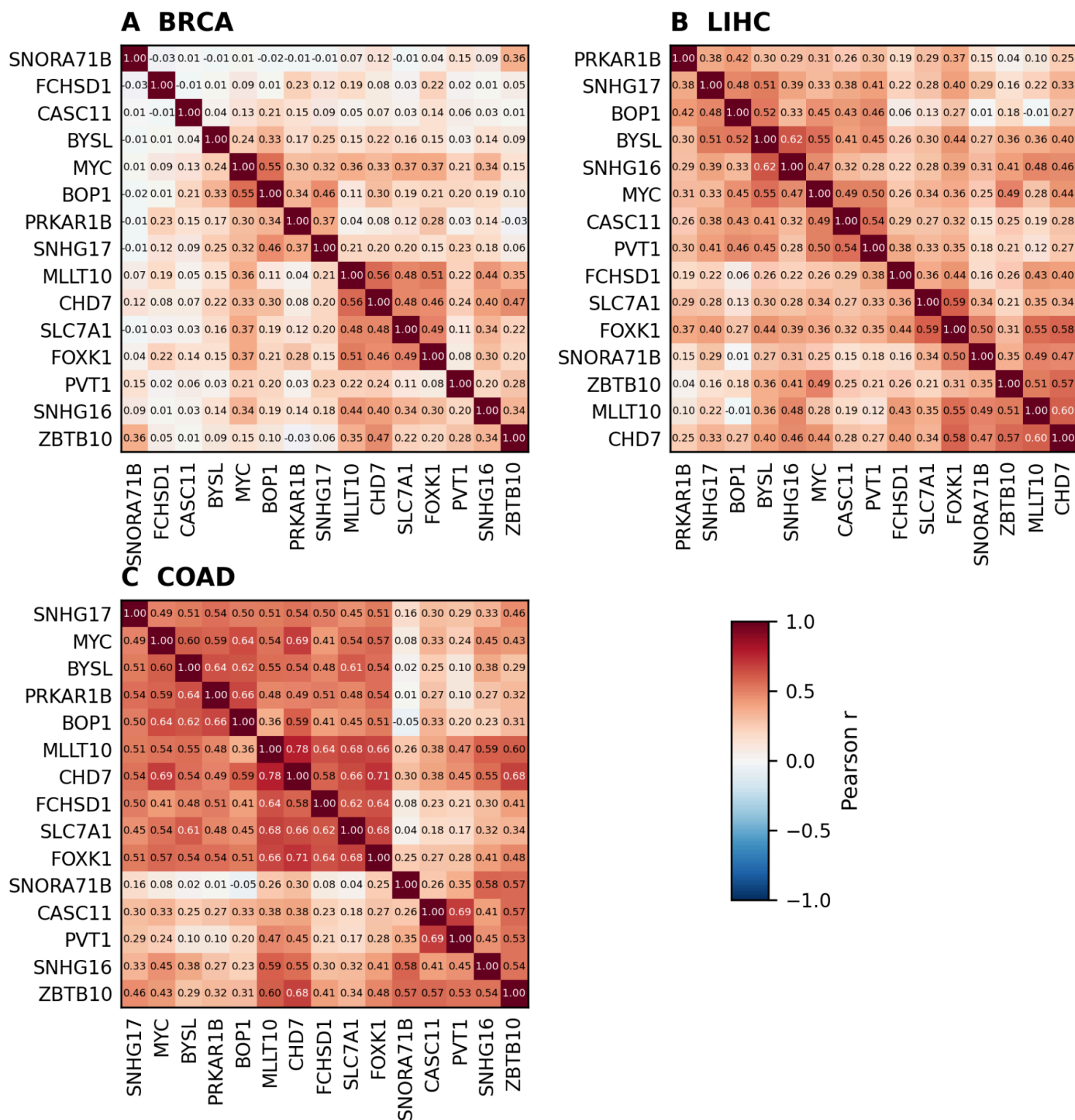


Figure 3

Gene-gene correlations among the fifteen signature genes in primary tumours of (A) BRCA, (B) LIHC and (C) COAD. Colour indicates the Pearson correlation coefficient and numbers in cells give its value. Genes are ordered by hierarchical clustering independently in each panel, so gene order differs between panels.

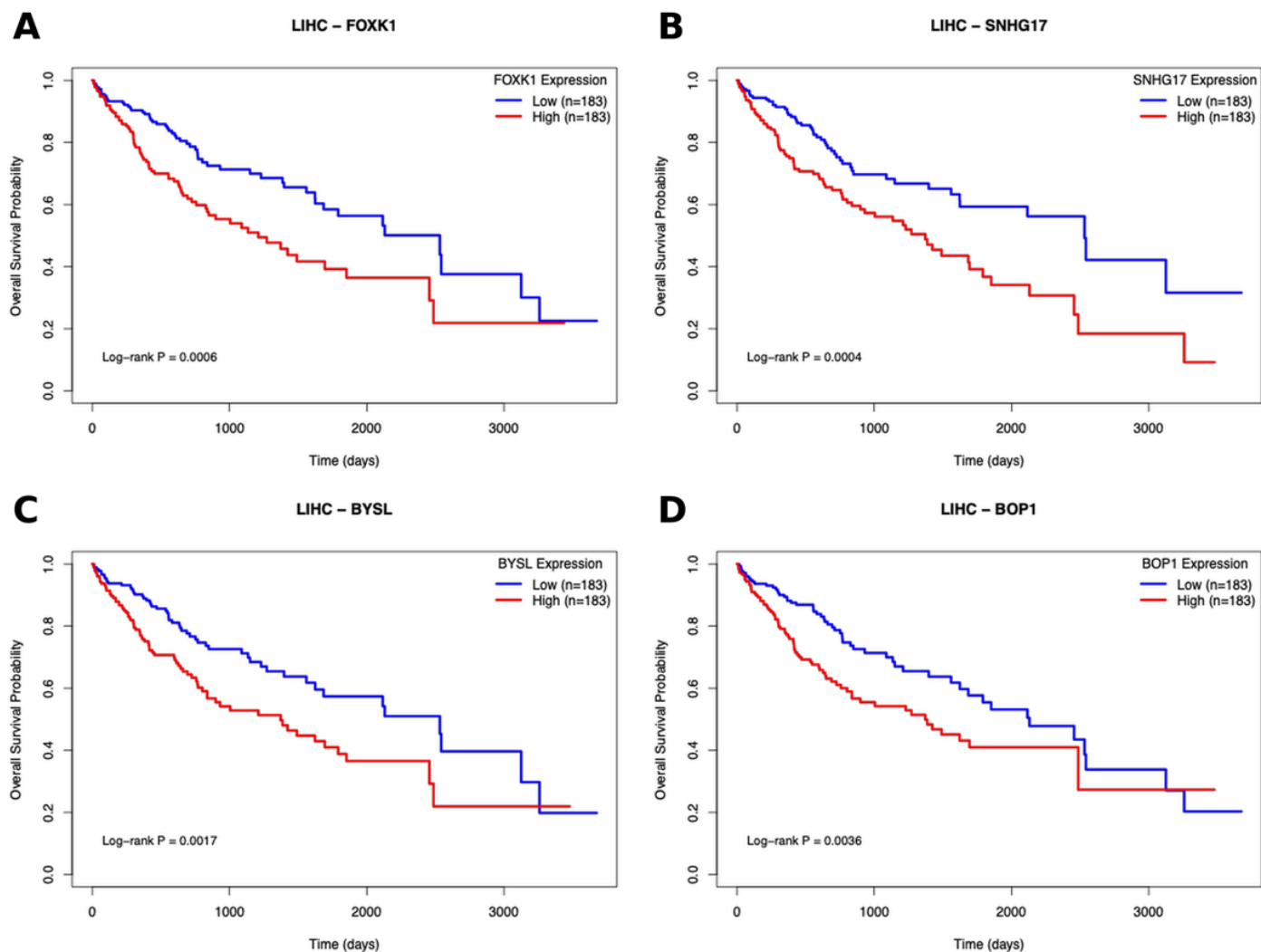


Figure 4

Overall survival in relation to signature gene expression in LIHC. Kaplan-Meier curves of overall survival for the four genes that remained significant after BH correction: FOXK1 (A), SNHG17 (B), BYSL (C) and BOP1 (D). Patients were split at the cohort median expression of each gene (blue, low; red, high).

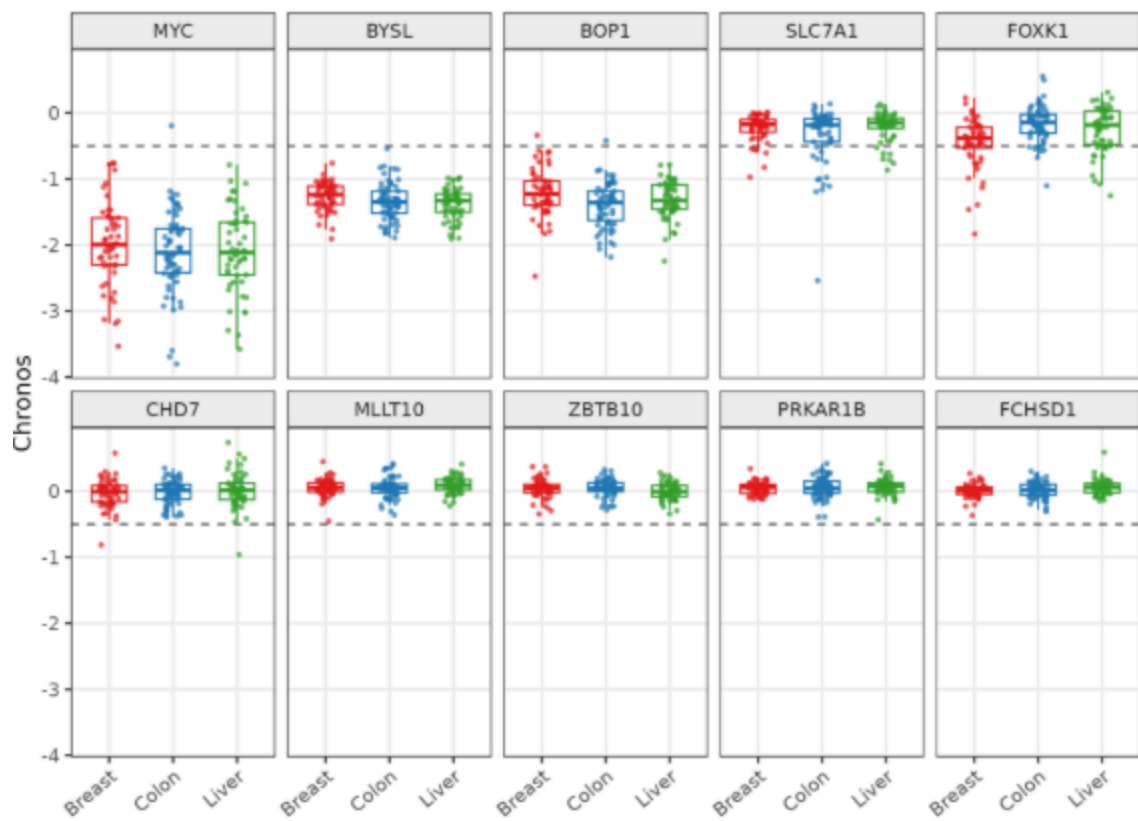


Figure 5

CRISPR dependency of the ten protein-coding signature genes in 169 breast (n = 53), colon (n = 63) and liver (n = 53) cell lines. Each point is a cell line; the dashed line marks Chronos = -0.5 and more negative scores indicate stronger dependency. Per-gene rankings are in Supplementary Figure 1.

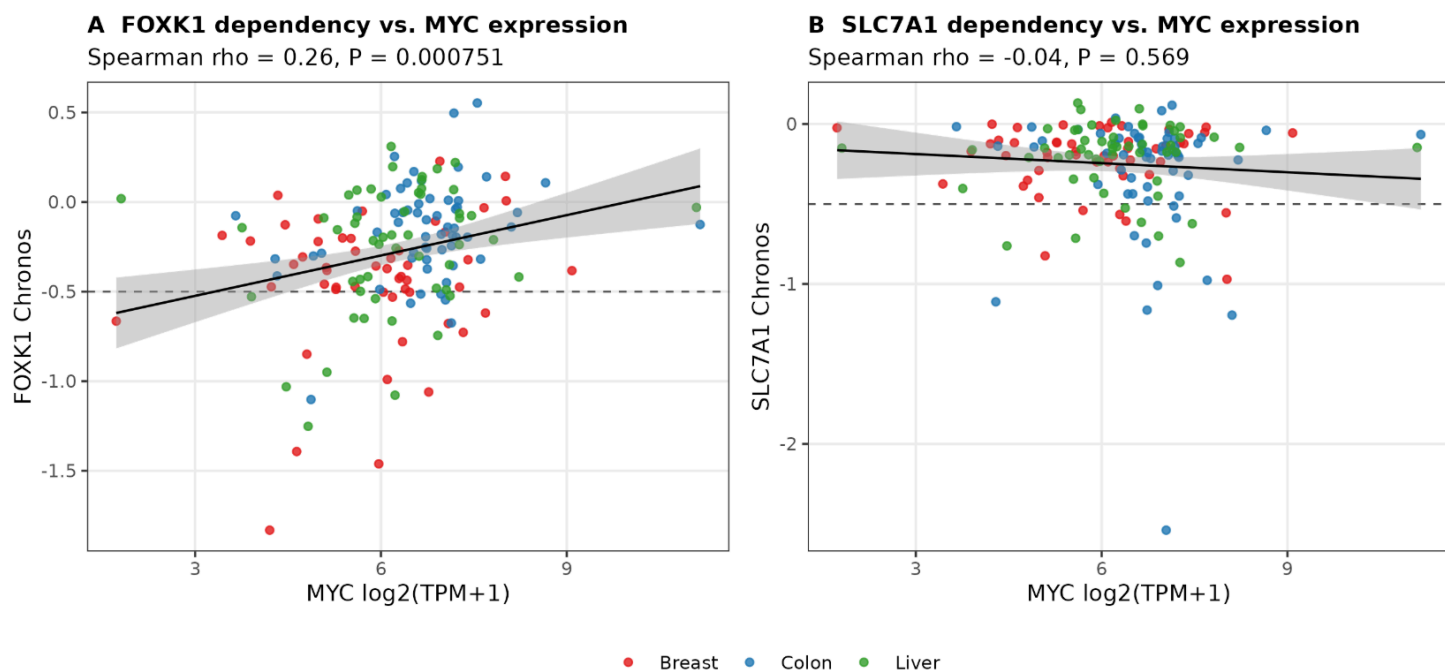


Figure 6

Relationship between MYC expression ($\log_2[\text{TPM} + 1]$) and dependency on FOXP1 (A) or SLC7A1 (B) in 163 cell lines profiled in both datasets. Each point is a cell line; the line is a pooled linear fit, and Spearman correlations are shown. More negative Chronos scores indicate stronger dependency, so a MYC-driven dependency would give a positive slope toward the lower right.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementary.docx](#)