SUPPLEMENTARY METHODS

Design and construction of gRNA library

gRNA sequences were designed to target the 5’ exons of genes according to standard protocols that reduce off target effects and target common exons between gene isoforms (1-3). gRNAs were cloned into the pLKO2 (Sigma; SHC201) lentiviral plasmid as oligonucleotides or by site directed mutagenesis of 19-mer sequences (Genscript). Each gRNA clone was sequence verified.

Lentivirus production and infection

Lentivirus production and infection was carried out as previously described (4). For high throughput production of the gRNA library, lentivirus was generated in 96-well plates. 293T cells were seeded at 55,000 cells/well 18 hours prior to transfection. Cells were subsequently transfected with the pLKO2-gRNA constructs along with pCMV-VSVG and pCMV-Δ8.9 using Lipofectime 2000 (Invitrogen). Six hours after transfection, the media was changed to 200 µl of fresh growth media. Viral supernatant was harvested 48 hours after transfection.

For RNAi experiments involving constitutive BRD4 and cMYC knockdown, the following pLKO-shRNA target sequences were utilized: shNTC (Sigma: SHC002), shBRD4-1 (CCACCCAAGTCAGTTCTTC), shBRD4-2 (CAGTGACAGTGACTGAGTGA), shBRD4-3 (GCCTATGCTCTTATGAGGAGA), shMYC-1 (CCTGAGACAGATCAGCAACAA), shMYC-2 (CGCAGGTGTCAATGGATAAT). Infected cells were selected with 2 µg/ml
puromycin 48 hours after transduction and harvested for immunoblot (3.5 days) or monitored for cell proliferation with CellTiter-Glo (Promega; 6 days).

**CRISPR high throughput arrayed cell viability screen**

RKO-Cas9 cells were seeded (750 cells/well) 18 hours prior to infection in 96 well plates in media containing 8 µg/ml polybrene. Virus particles were added to cells at an average MOI of 3-5 and spin infected at room temperature (1800 rpm, 30 minutes). Stable integration of gRNAs was selected with 2 µg/ml puromycin starting 48 hours after infection. The primary screen was performed in duplicate for cells receiving puromycin and in singlet for cells not receiving puromycin. Cell proliferation was measured 7 days after infection using CellTiter-Glo (Promega). Negative controls (Firefly Luciferase gRNAs) and positive controls (PLK1 gRNAs) were present on each plate. Data was normalized on each plate by dividing CTG values of each experimental gRNA by the average of 8 Luciferase gRNA controls that were on the same plate. Z scores were calculated from non-targeting control normalized values. PLK1 and Luciferase gRNA controls were not included for Z score calculations. Hit criteria were established using the following guidelines: 1 gRNA with a Z score < -2 and an additional gRNA with a Z score < -1.5.

**Generation of BRD4 clonal knockout cells using CRISPR**

CRISPR knockout clones were generated similarly to that described (5). Briefly, BRD4 gRNAs (target sequences: TTGGTACCGTGGAAACGCC and AAGATCATTTAAACGCCTA) were co-transfected with Cas9 using Lipofectamine
LTX (Invitrogen) to generate BRD4 knockout cells. Untransfected cells were eliminated with a 24 hour Puromycin treatment (2 µg/ml) two days after transfection. Clonal populations were isolated by FACS in 96 well format, screened by immunofluorescence (using Epitomics: 5716-1 antibody) and validated by immunoblot. For BRD4 long-isoform truncation experiments, a different gRNA (target sequence: AAAGAAGGGGCACCCCGGG) was utilized. Knockout cells were screened by immunofluorescence (using Bethyl: A301-985A100) and confirmed by immunoblot. Homozygous deletion frequency ranged from 10-40% depending on cell line and gRNA.

**Identification of genomic features predictive for JQ1 sensitivity**

Using EC50 values as a measure for the response to JQ1 treatment, we defined cell lines as sensitive or resistant based on a 30%- and 70%-quantile cutoff, respectively. To create genomic feature profiles for the classified sensitive (ATRFLOX, HT-29, RKO, CW-2, HCT-15, SW 48) and resistant (LS-180, SW 480, LS-174T, GP5d, SW 948, COLO 741) cell lines, total copy number (26,347 features), expression (26,225 features) and mutation (314 features) data were retrieved (6). The DESeq R package was applied to estimate size factors and obtain dispersion estimates for the associated RNA-seq data (7). Gene expression was quantified with variance-stabilized counts. DNA methylation data from Illumina microarrays yielded additional 35,788 features per cell line. For each transcript, the promoter region was identified (-2000 bp to + 500 bp around the transcriptional start site) and the mean expression across all contained probes was calculated (mean M-score). For both expression and methylation data, we restricted our classification approach to the 2000 most variable instances. Furthermore, 14 different
measures for CIMP classification including established gene expression signatures (8) were employed. EMT and MSI status, doubling time, seeding density and alteration status of canonical pathways formed eight additional features.

We applied the Lasso approach to identify the most reliable features for predicting either the class or the underlying EC50 values of sensitive versus resistant lines (9). Using the ‘cv.glmnet’ function from the R package glmnet, we trained lasso based models and used 10-fold cross validations to determine tuning parameter $\lambda$ yielding minimum cross-validated errors. Features were defined as reliable for predicting JQ1 sensitivity, if the associated $\beta$ coefficients were not zero.

**RNA sequencing and analysis**

Cells were treated for 24 hours with DMSO or 0.5 $\mu$M JQ1 in triplicate. For shRNA experiments cells were treated for 3.5 days with 0.5 $\mu$g/ml doxycycline. Total RNA was isolated with the Qiagen RNAeasy kit and subjected to oligo (dT) capture and enrichment. The resulting mRNA fraction was used to construct complementary DNA libraries and Transcriptome sequencing (RNA-seq) was performed as described (10). For analysis of differentially expressed genes, the following cutoffs were used: 2 fold change, median RPKM >0.1, and adjusted p value < 0.01.

**Gene set enrichment analysis (GSEA)**

GSEA (http://www.broadinstitute.org/gsea/index.jsp) (11) was performed using JQ1/DMSO log2 fold change for each cell line. CIMP(+) classified cell lines (n=4) were
compared to CIMP(-) classified cell lines (n=2) using difference of classes as the ranking metric and the Hallmark MSigDB gene set collection. An FDR cutoff of 5% was used.

**Gene expression analysis**

For qRT-PCR experiments total RNA was isolated with the RNeasy mini kit (Qiagen). Reverse transcription followed by qPCR was performed with the TaqMan one-step RT-PCR master mix (Applied Biosystems). Samples were normalized to expression of GAPDH mRNA.

**ChIP-sequencing and analysis**

Cells were fixed with 1% formaldehyde for 5-15 minutes and quenched with 0.125 M glycine. Chromatin was sonicated to an average length of 100-500 bp. H3K27ac ChIP was performed using 2 µg of rabbit anti-H3K27ac antibody (Abcam, ab4729, Lot: GR183919-2) with the Diagenode LowCell# ChIP kit with minor modifications. BRD4 ChIP was performed using 30 µg of precleared chromatin and 4 µg of antibody against BRD4 (Bethyl Laboratories, A301-895, Lot. A301-985A100-3). 200 ng of Drosophila S2 chromatin and 0.4 µg of drosophila-specific H2A.v antibody (Active Motif: 39715) were also added to each reaction for normalization purposes. Complexes were washed and eluted from beads with SDS buffer. ChIP and input samples were subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation at 65°C and DNA was purified.

For BRD4 ChIP-qPCR, qPCR was performed on ChIP and input DNA with TaqMan Universal PCR Master Mix (Applied Biosystems) with custom designed primers and probe for CCAT1 (forward: 5’- CAAAGGTCCCAATTTCACACT; reverse: 5’- ACAACTGTGCTCCTGAATGC; probe: 5’-TCCAGTTGGTTCTCTTTTCCTTTGCT)
and a negative control locus on chromosome 4 (forward: 5’-GATGGCCCAGTGTAAGCATT; reverse: 5’-TGACTCTGACGATAGCTCTCAAA; probe: 5’-AATGT CCTAGTTT CATAAA TACGGTC ACTCTATCTGG).

Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on a HiSeq2500 or a NexSeq 500 (75 bp, single end reads).

Bioinformatics analysis includes de-multiplexing and filtering followed by alignment to the human genome (hg19) using the BWA algorithm (default settings) (12). Duplicate reads were removed. For H3K27ac ChIP-seq, uniquely mapped reads were normalized for total read number per sample by random downsampling. For BRD4 ChIP-seq normalization to drosophila chromatin spike was performed as follows. Reads were aligned to the Drosophila genome (dm3) using the BWA algorithm (default settings). Normalization was performed by equalizing the Drosophila tag counts across all samples so that the final tag counts were based off of the sample containing the lowest number of Drosophila tags. Then the human tags counts for all samples were proportionally scaled based on the ratios used to adjust the drosophila tag counts. Scaling to the target tag number was performed by randomly removing excess tags.

To identify super-enhancers, peak locations for BRD4 were determined using the MACS algorithm (v1.4.2) (13) with a cutoff of p-value = 1x10^{-7}. MACS peak locations and BAM files were used as input into the ROSE software to identify super-enhancers (14, 15). The default stitching distance of 12.5 kb was used and promoters were not
excluded. Super-enhancers were annotated to genes if they fell within 50 kb upstream or downstream of the gene.

Heatmap representation of drosophila chromatin normalized BRD4 binding at super-enhancer regions was generated using seqplots (http://github.com/przemol/seqplots). Overlapping super-enhancer regions were grouped into active regions defined by the most upstream start position and the most downstream end position (the union of overlapping intervals). When a super-enhancer was present in only one cell line the active region was defined by that interval.

To calculate input-normalized and averaged ChIP-seq signal, coverage was calculated across 10 bp bins using BEDTools (16) and the ratio of ChIP/input across bins was used directly or averaged across CIMP(+) and CIMP(-) lines. Tracks were visualized in the Integrative Genomics Viewer (IGV) (17, 18) (http://www.broadinstitute.org/igv/).

**Human tissues**

To evaluate CCAT1 expression, a population-based series of patients who had undergone surgical resections for colorectal adenocarcinoma was compiled retrospectively from the pathology archives at St James' University Hospital (Leeds, UK) from 1988 to 2003. Patient demographics and treatment information were obtained from clinical records. Morphology, site, Dukes' stage, grade of differentiation, number of lymph nodes retrieved and number of positive lymph nodes were compiled from the surgical pathology reports. Survival data were obtained from the Northern and Yorkshire Cancer Registry and Information Service (St James' University Hospital; Leeds, UK). Inclusion of patients into this cohort was dependent on the availability of archival material and outcome data.
TMAs were constructed as above with one core of normal mucosa and three cores of adenocarcinoma per patient.

Biostatistical analysis of Leeds tumor collection

Patients were classified as CCAT1+ if their CCAT1 ISH score was >1 and CCAT1- otherwise. Any patients with missing CCAT1 expression score were removed from the analysis. Patients with tumor location specified as ‘ASCENDING COLON’, ‘CAECUM’, or ‘HEPATIC FLEXURE OF COLON’ were classified as having ‘proximal’ colon cancer, patients with tumor location specified as 'DESCENDING COLON', 'RECTOSIGMOID JUNCTION', 'SPLENIC FLEXURE OF COLON', 'SIGMOID COLON', or 'TRANSVERSE COLON' were classified as having ‘distal’ colon cancer and patients with tumor location specified as ‘RECTUM’ where classified as having ‘rectal’ colon cancer. A single patient with tumor location specified as ‘APPENDIX’ was removed from the analysis. Patients with cMYC expression scores <1 were classified as having ‘low’ cMYC expression, while patients with cMYC expression scores ≥1 and <2 were classified as having ‘moderate’ cMYC expression and patients with the scores ≥2 were classified as having ‘high’ cMYC expression. Patients with relative cMYC copy number >0.6 and <1.6 were classified as having cMYC ‘gain’ events, while patients with relative cMYC copy number ≥1.6 were classified as having cMYC ‘amplification’ events, with the rest classified as having low cMYC copy number. All deceased patients had between 1 and 4 entries specifying their cause of death (COD). Any patient with at least one COD entry containing colon cancer description was classified as having died from colon cancer. Further, any patient for whom the only COD
given was unspecified malignant neoplasm was assumed to have died from colon (and not another) cancer and classified accordingly. All other patients were classified as having died from causes other than colon cancer. Upon manual examination of the COD table, a handful (<10) of patients were reclassified to account for uncommon combinations of COD entries. Any patients who died within 30 days of diagnosis or had surgery more than a year removed from diagnosis date were excluded from the analysis. The additional filtering steps described here resulted in removal of 51/689 patients with available CCAT1 scores, leaving 638 patients subject to further analyses.

Analysis of pairwise associations between CCAT1 expression status (+/-) and other covariates (Table 1 and Supplemental Table 6) was carried out using R (http://www.R-project.org/) package Epi (http://CRAN.R-project.org/package=Epi) with Fisher Exact Test p-value reported in the table. For each individual association test, patients with missing values of the particular covariate were excluded.

Survival analysis was carried out using R package survival (http://CRAN.R-project.org/package=survival). Cox proportional hazards model was fitted using covariates given in Supplemental Table 6. The analysis was carried out for 5 year survival for both colon cancer and overall causes of death (Table 1, Supplemental Table 6, and Figure 7C). Initial model fit was performed including cMYC expression score and cMYC relative copy number category as two additional covariates. Due to large number of patients with missing cMYC CN information, only 147/638 patients contributed to parameter estimation. Further, due to lack of colon cancer deaths for patients with stage I tumors within this smaller patient subset, parameters for the effects of tumor stage could not be estimated. Exclusion of cMYC covariates did not result in significant increase of
lack of fit (Chi-square p-value > 0.9 for both types of COD). We therefore excluded cMYC covariates from the final analysis, resulting in 491/638 patients contributing to the reported parameter estimates.

Non-isotopic in situ hybridization

Non-isotopic in situ hybridization (ISH) was performed on 4 µm FFPE sections using QuantiGene® ViewRNA ISH Tissue Assay (Affymetrix/Panomics) following the manufacturer's protocol on a Tecan platform equipped to carry out non-isotopic in situ hybridization. Gene-specific probe for detection of human CCAT1 RNA (Affymetrix; VA1-17802) target region 2-2696 in Genbank accessions NR_108049.1 was used on tissue samples. A probe set to Bacillus subtilis dihydropicolinate reductase (dapB) (VF1-11712), target region 1363-2044 in Genbank accession L38424 was used as a negative control. Horseradish peroxidase (HRP) conjugated label probe was used, followed by TSA™ (tyramide signal amplification) to increase sensitivity (Perkin Elmer NEL748001KT). Briefly, TSA Plus DIG stock solution (digoxigenin) was diluted 1:50 in 1x Plus Amplification Diluent and applied to sections and incubated for 10 minutes at room temperature. This was followed by incubation with anti-DIG-AP (Roche 11093274910) diluted 1:500 in TNB blocking buffer with 4% lamb serum (Gibco, 16070-096) for 30 minutes at room temperature. Vulcan Fast Red substrate (Biocare, FR805S) was used for chromogenic detection.

Genomic copy number analysis
For quantitative copy number analysis, genomic DNA was isolated with the DNeasy blood & tissue kit (Qiagen). Quantitative PCR for cMYC copy number was performed with the TaqMan genotyping master mix using Taqman copy number probes (Applied Biosystems). For colon cell lines and tumors, samples were normalized to copy number at the TERT locus, which is rarely amplified or deleted in colon cancer (Tumorscape, http://www.broadinstitute.org/tumorscape).

**Immunohistochemistry**

Immunohistochemistry (IHC) was performed on 4um thick formalin-fixed, paraffin-embedded tissue sections mounted on glass slides. All IHC steps were carried out on the Ventana Discovery XT automated platform (Ventana Medical Systems; Tucson, AZ). Sections were treated with Cell Conditioner 1, standard time, and then incubated in primary antibodies: cleaved caspase 3 (Cell Signaling Technologies), phospho-Histone H3(Ser10) (Upstate Biotechnology), BRD4, clone EPR5150 (AbCam), and cMYC, clone Y69 (Ventana Medical Systems; Tucson, AZ). Cleaved caspase 3, phospho-Histone H3 and BRD4 were detected by OmniMap anti-Rabbit-HRP (Ventana Medical Systems; Tucson, AZ). cMYC was detected by UltraMap anti-Rabbit-HRP. The sections were counterstained with hematoxylin, dehydrated, and coverslipped.
SUPPLEMENTARY REFERENCES


Supplemental Figure 1: Validation of Epi200 gRNA library hits.

A) KAT8 gRNAs

B) AURKB gRNAs

C) CHD1 gRNAs

D) HDAC1 gRNAs

E) Protein remaining after gRNA targeting (%)

- KAT8: NTC, 1, 2, 3, 4, 5
- AURKB: NTC, 1, 2, 3, 4, 5
- CHD1: NTC, 1, 2, 3, 4, 5
- HDAC1: NTC, 1, 2, 3, 4, 5

TUBULIN: 100% <1% <1% <1% <1%

KAT8: 39% 36% 41% 29% 27% 23%

AURKB: 13% 3% 7% 3% <1%

CHD1: 1% 2% 18% 29% 36%

HDAC1: <1% 3% 3% 3% 3% <1%
Supplemental Figure 2: BRD4 is required for colon cancer growth.
Supplemental Figure 3: JQ1 sensitivity profiling and associating features.

A

Sensitive
Resistant

JQ1 Relative EC50

ATRFLOX
HT-29
RKO
CW-2
HCT-15
SW 48
COLO 205
SW 948
GP5d
LS-174T
SW 480
LS-180

B

DNA methylation intensity

gene
CACNA1G
CDKN2A
CRABP1
IGF2
MLH1
NEUROG1
RUNX3
SOCS1

JQ1:

CIMP(+)

CIMP (-)

S
S
R
S
S
S
R
S

R
R
R
R
R
R
R
R
R
R
R
R
R
R
R

Features per Sample:

RNAseq 26,225
DNA methylation 35,788
Mutation 314
Copy Number 26,374
CIMP measures 14
Other (MSI, EMT, ...)
Total 88,723

C

POU3F2 p<0.01
ρ=−0.769

FAM65B p<0.01
ρ=−0.649

GABRB2 p<0.01
ρ=−0.640

D

ZNF606 p<0.001
ρ=0.757

ZNF606 p<0.05
ρ=−0.479
Supplemental Figure 4: Colon cancer I-BET-762 sensitivity.
Supplemental Figure 5: BRD4 ChIP-seq analysis and gene expression changes following BET inhibition in colon cancer cells.
Supplemental Figure 6: Differential gene expression analysis following JQ1 treatment in colon cancer cells.

RNA-seq: JQ1 down-regulated genes

ChIP-seq: BRD4 super-enhancer associated genes

CIMP(+)

HCT 116
HT-29
COLO 205
HCT-15

CIMP(-)

COLO 320
SW 480

CIMP(+)

CIMP(-)

HALLMARK_MYC_TARGETS_V2
58
0.70
2.19
0.0000

HALLMARK_MYC_TARGETS_V1
200
0.57
2.15
0.0000

HALLMARK_MTORC1_SIGNALING
195
0.54
2.02
0.0000

HALLMARK_E2F_TARGETS
200
0.52
1.96
0.0000

HALLMARK_UNFOLDED_PROTEIN_RESPONSE
111
0.52
1.80
0.0009

HALLMARK_APICAL_SURFACE
39
0.56
1.65
0.0041

HALLMARK_G2M_CHECKPOINT
199
0.43
1.63
0.0043

HALLMARK_SPERMATOGENESIS
91
0.45
1.50
0.0173

GSEA: CIMP(+) enriched terms

NAME
SIZE
ES
NES
FDR q-val
HALLMARK_MYOGENESIS
155
-0.50
-1.71
0.0137

HALLMARK_COAGULATION
103
-0.51
-1.65
0.0187

HALLMARK_EPITHELIAL_MESenchymal_TRANSITION
142
-0.48
-1.62
0.0217

GSEA: CIMP(-) enriched terms

NAME
SIZE
ES
NES
FDR q-val
HALLMARK_HALLMARK_E2F_TARGETS
200
0.57
2.15
0.0000

HALLMARK_MTORC1_SIGNALING
195
0.54
2.02
0.0000

HALLMARK_E2F_TARGETS
200
0.52
1.96
0.0000

HALLMARK_UNFOLDED_PROTEIN_RESPONSE
111
0.52
1.80
0.0009

HALLMARK_APICAL_SURFACE
39
0.56
1.65
0.0041

HALLMARK_G2M_CHECKPOINT
199
0.43
1.63
0.0043

HALLMARK_SPERMATOGENESIS
91
0.45
1.50
0.0173
Supplemental Figure 7: cMYC transcriptional response to JQ1 in CIMP(+) and CIMP(-) cell lines.
Supplemental Figure 8: The cMYC super enhancer, marked by BRD4 and Histone H3K27ac, is primarily enriched in CIMP(+) colon cancer cells.
Supplemental Figure 9: Transcriptional profiling following BRD4 knockdown and JQ1 treatment reveals that CCAT1 is a BET specific target.

A

HT-29

JQ1

Up-regulated genes (>2-fold, p<0.001)

601

166

182

Down-regulated genes (>2-fold, p<0.001)

1027

35

16

shRNA

B

HCT 116

JQ1

Up-regulated genes (>2-fold, p<0.001)

234

53

142

Down-regulated genes (>2-fold, p<0.001)

640

51

28

shRNA
Supplemental Figure 10: CCAT1 is a BET inhibitor target, but not exclusively regulated by BRD4.
Supplemental Figure 11: CCAT1, PCAT1 and LOC728724, IncRNAs associated with cMYC super-enhancers, are exquisitely sensitive to BET inhibition.
Supplemental Figure 12: Cells with elevated CCAT1 and PCAT1 expression show a marked decrease in cMYC RNA following JQ1 treatment.
Supplemental Figure 13: CCAT1 is superior to cMYC as a BET inhibitor predictive biomarker.

A

Gastric

CCAT1 Relative EC50

CCAT1 Low
CCAT1 High

B

CCAT1 log2 RPKM

JQ1 Relative EC50 (uM)

C

cMYC log2 RPKM

JQ1 Relative EC50 (uM)

r = -0.40

P = 0.0116

D

CCAT1 log2 RPKM

C

r = -0.15

P = 0.3609

E

Cell Number (CTG)

-f- cMYC

-f- ACTIN

F

CCAT1-
CCAT1+

CCAT1-
CCAT1+

P = 0.77

P = 0.59
Supplemental Figure 14: Development of a CCAT1 in situ hybridization assay.

A

HT-29 (CCAT1 high)

SW 620 (CCAT1 low)

CCAT1 ISH

RKO (+ 1 μM JQ1)

RKO (DMSO)

50 μm

B

CCAT1 ISH

50 μm

C

CCAT1 expression (nRPKM) in normal human tissue

 Colon  Gastric  Pancreatic  Lung

0  1  2  3  4

D

Normal Colon

CCAT1 ISH

100 μm
Supplemental Figure 15: CCAT1 expression correlates to poor outcome independent of clinical stage.

A

Survival by CCAT1 score category
All deaths

Survival by CCAT1 score category
Cancer-related deaths

B

Survival by CCAT1 score category
Cancer-related deaths; Stage I

Survival by CCAT1 score category
Cancer-related deaths; Stage II

Survival by CCAT1 score category
Cancer-related deaths; Stage III

Survival by CCAT1 score category
Cancer-related deaths; Stage IV