Supplementary files

Supplementary figures S1-S7

Supplementary tables S1-S6

Supplementary Experimental Procedures

Supplementary references

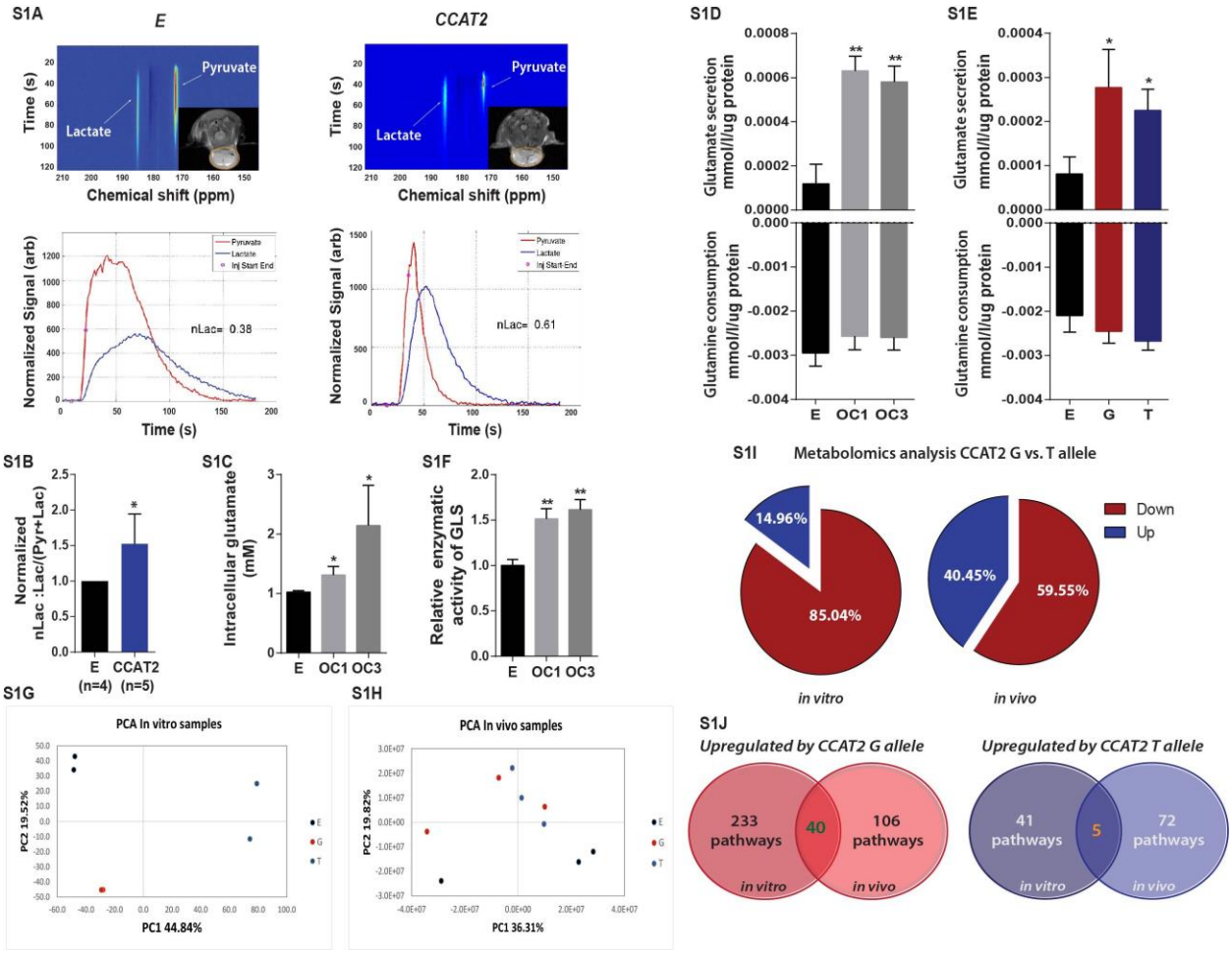
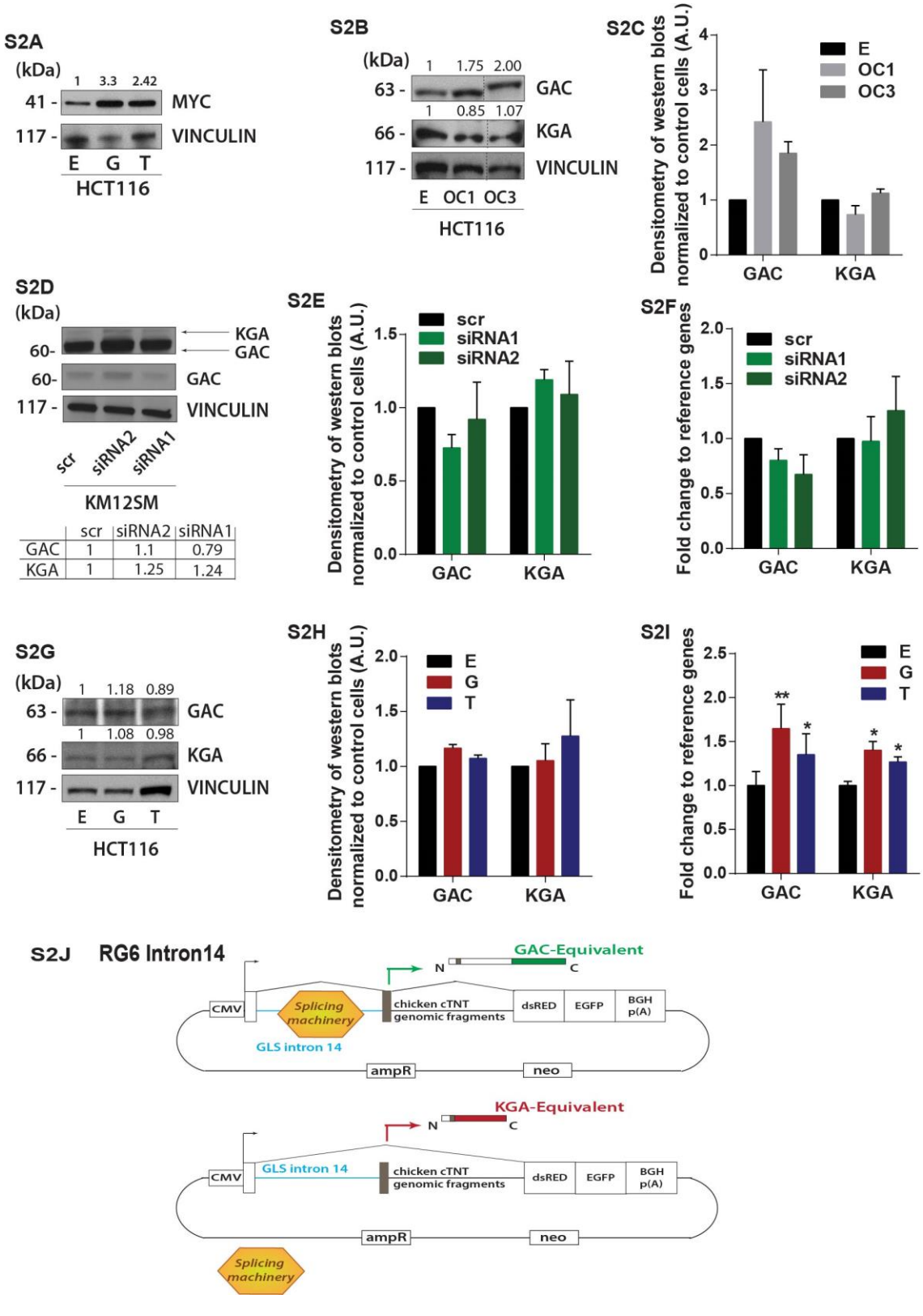


Fig S1. *CCAT2* regulates cancer metabolism *in vitro* and *in vivo* (see also Fig. 1 and Table S1) (A) Hyperpolarized ^{13}C-spectrum from xenograft tumors (E – control group; CCAT2 – *CCAT2* overexpressing group). The spectrum displays peaks from pyruvate and lactate (**upper panels**). Plot of the lactate and pyruvate peaks as a function of time. The lactate peak was normalized to the pyruvate peak integral at t=0 s corresponding to the injection time (**lower panels**) (**B**) The normalized mean lactate for the E and *CCAT2* xenograft tumors. The lactate produced by an individual tumor was calculated with the formula: nLac=Lac/(Pyr+Lac), representing the total area under the dynamic lactate curve divided by the sum of the areas under lactate and pyruvate curves. (**C**) Intracellular glutamate concentration measured in HCT116 cells with *CCAT2*-overexpression and control cells. (**D, E**) Glutamine and glutamate concentration in the media relative to the empty well 48 hours after seeding HCT116 *CCAT2*-overexpressiong and control cells (E, OC1 and OC3) (**D**) and *CCAT2*-overexpressing the G- or T-allele and control cells (E, G and T) (**E**). (**F**) Whole cell lysate *Glutaminase* activity measured in HCT116

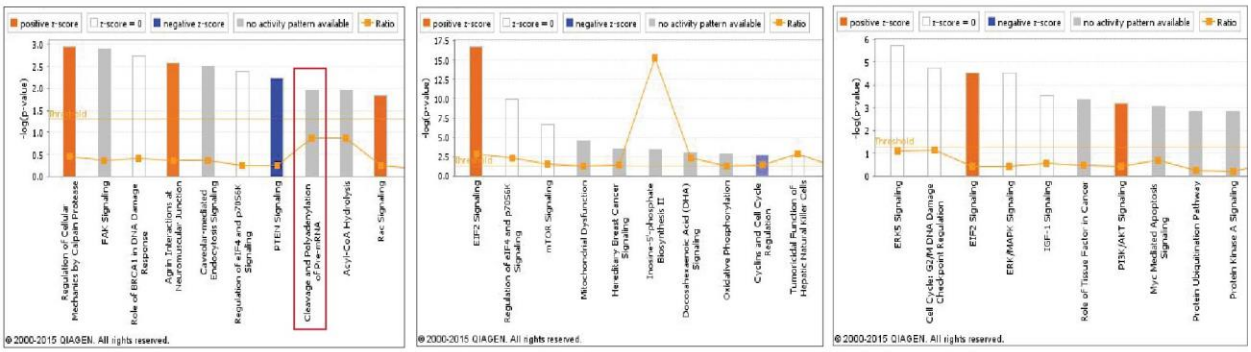
CCAT2-overexpressiong and control cells. (**G**) Principal component analysis (PCA) of HCT116 cells stably overexpressing *CCAT2* with either G - or T-allele, and control cells (E – empty vector) *in vitro* allowed an adequate classification of the different cell lines according to its metabolome. (**H**) Xenograft tumors derived from the same cell lines classified by PCA analysis. (**I**) Schematic representation of the percentage of metabolic pathways downregulated (marked in red) or upregulated (marked in blue) in T-allele cells compared to the G-allele cells. (**J**) Pathway analysis identified 233 and 106 metabolic pathways upregulated *in vitro* and *in vivo*, respectively, having 40 common pathways, for cells overexpressing the *CCAT2* G-allele (**left panel**) and 41 and 72 metabolic pathways upregulated *in vitro* and *in vivo*, respectively, having 5 common pathways, for cells overexpressing the *CCAT2* T-allele (**right panel**). Comparison was performed between the two alleles (see also **Table S1**). Results are normalized to the control and presented as mean values ± SD.



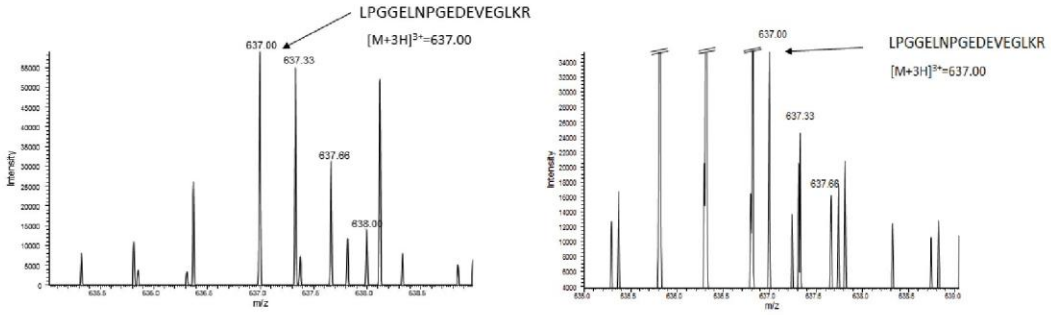
	scr	siRNA2	siRNA1
GAC	1	1.1	0.79
KGA	1	1.25	1.24

Fig. S2. *CCAT2* induces the alternative splicing of *GLS* (See also Fig. 2 and Table S2) (**A**) Western blot analysis of MYC in HCT116 stably overexpressing *CCAT2* G or T allele. (**B**) Western Blot analysis of GAC, KGA in HCT116 *CCAT2*-overexpressing (OC1 and OC3) and control cells. (**C**) Quantification of the three western blots is presented in the right panel. (**D**) Western Blot analysis of GAC, KGA in KM12SM cells with *CCAT2* downregulation. (**E**) Quantification of the three western blots is presented in the right panel. (**F**) RT-qPCR assessing the mRNA expression of GAC and KGA in KM12SM cells with *CCAT2* downregulation. (**G**) Western Blot analysis of GAC, KGA in HCT116 *CCAT2*-overexpressing the G- and the T-allele, and control cells. (**H**) Quantification of the three western blots is presented in the right panel. (**I**) RT-qPCR assessing the mRNA expression of GAC and KGA in HCT116 *CCAT2*-overexpressing the G- and the T-allele, and control cells**.** (**J**) Schematic representation of RG6 minigene with the *GLS* intron 14 and the principle of function. Alternative splicing of the intron shifts the reading frame between dsRED and EGFP. Results are normalized to the control and presented as mean values ± SD.

S3A Unique to CCAT2 G-allele | Unique to CCAT2 T-allele | Common for CCAT2 G- and T-allele



S3B



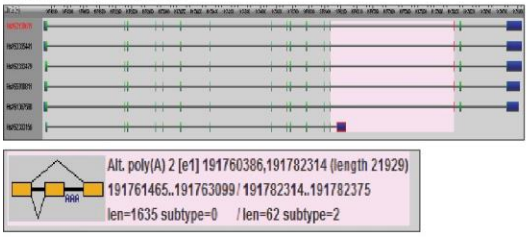
CCAT2 G-allele
m/z=636.99518, [M+H]+=1908.97098, Area=5.703E06
+3 charge, RT28.46, scan 7619

CCAT2 T-allele
m/z=636.99445, [M+H]+=1908.96878, Area=3.859E06
+3 charge, RT28.45, scan 7667

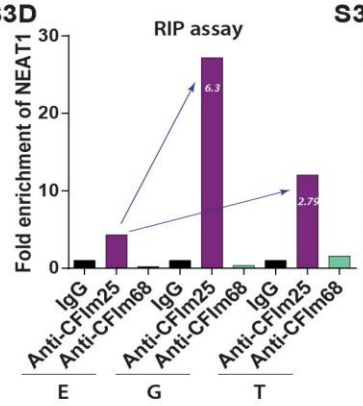
NUDT21 ratio: (CCAT2-G)/(CCAT2-T)=1.48

S3C

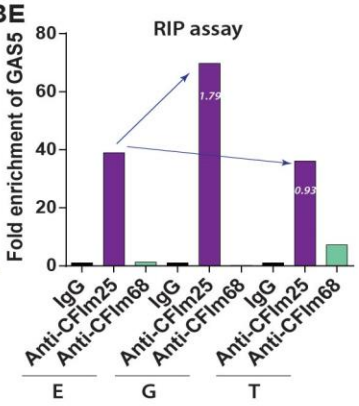
ASTRA database



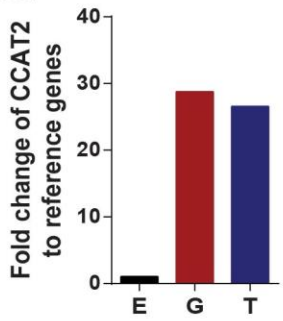
S3D



S3E



S3F



S3G

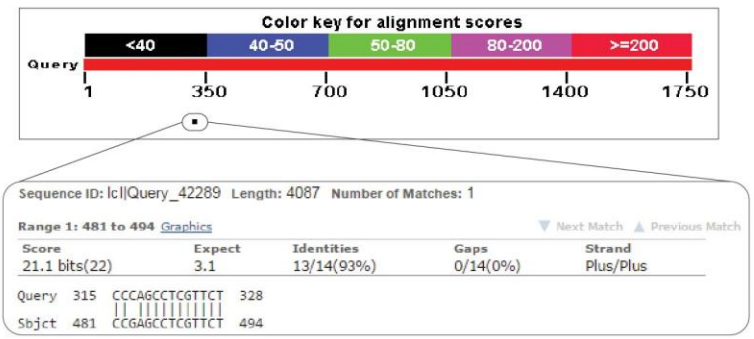


Fig. S3. The CFIm protein complex binds the *GLS* pre-mRNA (See also Fig. 3) (**A**) Pathways analysis of the proteomics data on Qiagen platform. Upper two graphs present the pathways having proteins unique either to the *CCAT2* G-allele or to the T-allele, and the lower graph presents the pathways with common proteins for the two alleles. (**B**) Spectrographic analysis of proteins interacting with *CCAT2* G-allele and T-allele presenting the peak corresponding to *NUDT21* (CFIm25). (**C**) Snap-shot of ASTRA database showing an alternative polyadenylation event in intron 14 of *GLS* pre-mRNA. RT-qPCR assessing the fold enrichment of *NEAT1* (**D**) and *GAS5* (**E**) bound to CFIm25 and CFIm68 in HCT116 cells *CCAT2*-overexpressing G- or T-allele and control cells (E). (**F**) RT-qPCR analysis of *CCAT2* expression in the cells used for RNA immunoprecipitation. (**G**) Snap-shot of NCBI Blast sequence alignment (http://blast.ncbi.nlm.nih.gov/BlastAlign.cgi) presenting the result of aligning *CCAT2* FASTA sequence (query) and *GAS5* FASTA sequence (subject). Results are normalized to the control and presented as mean values ± SD.

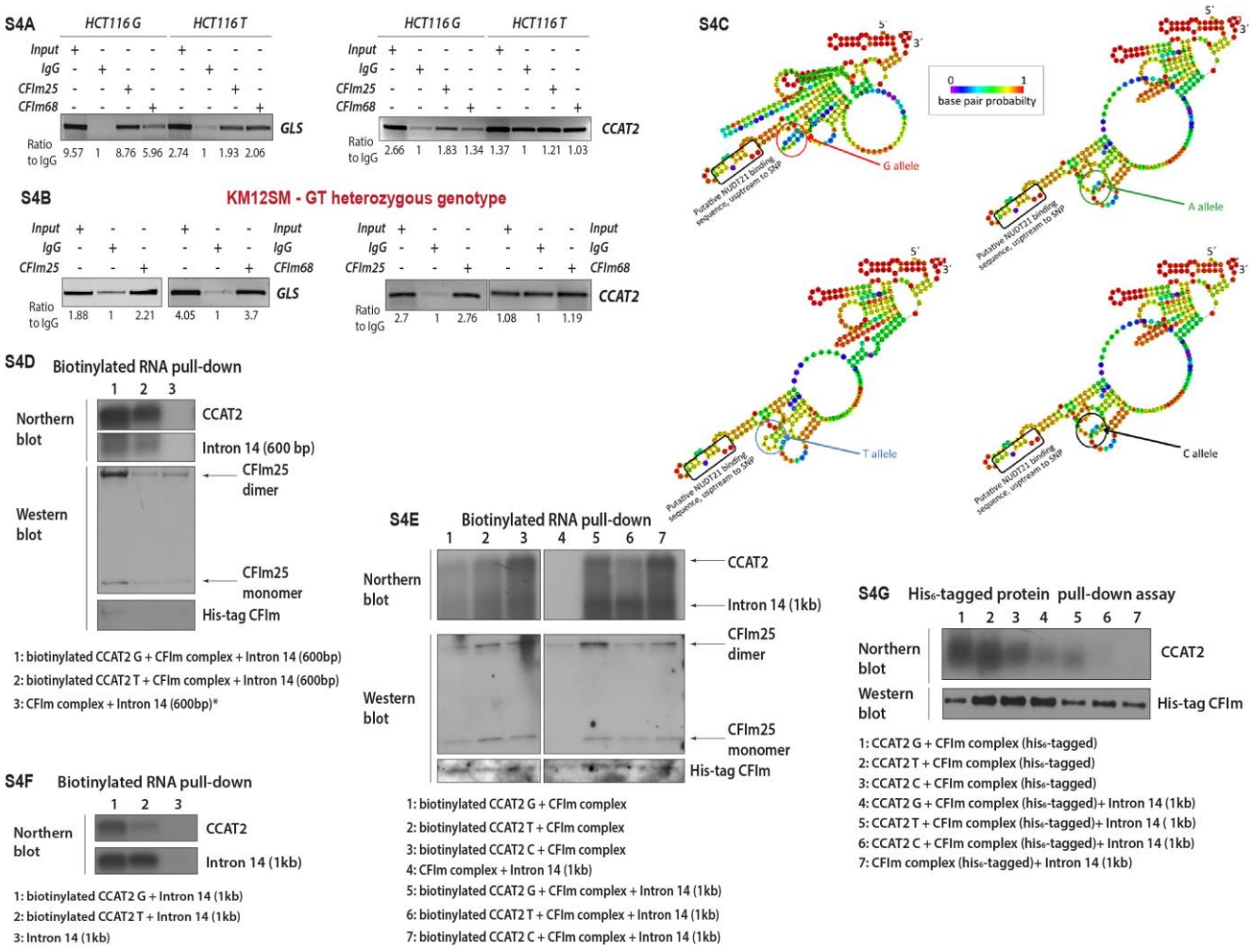


Fig. S4. The rs6983267 SNP affects the interaction of *CCAT2* with the CFIm protein complex (See also Fig. 4 and Table S3). PCR for *GLS* pre-mRNA (primers were designed to amplify the region in the intron 14 hosting the CFIm25 binding motifs) and *CCAT2* (primers were designed to amplify the region surrounding the SNP) for cDNA obtained from RNA immunoprecipitation with CFIm25 and CFIm68 antibodies. Experiments were performed in HCT116 *CCAT2*-overexpressing G or T allele cells (**A**) and KM12SM cells (**B**). (**C**) *CCAT2* G, T, A and C secondary structure prediction (RNAfold Webserver, http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) presenting the region around the SNP (at position 662) and one CFIm25 binding motif. The figure above depicts only the section between nucleotides 429 and 740. The predictions for the remaining *CCAT2* sequences are identical between each one of the alleles, and therefore not shown. (**D**) Northern Blot analysis showing the presence of *CCAT2* and intron 14 (600bp fragment) in the lysate pulled down with Streptavidin beads specific for biotin tags

(upper panel). Lane 3 marked with the star symbol is identical to lane 1 in **Fig. 4C**. Western Blot analysis showing the presence of CFIm25, monomer (26 kDa) and dimer (64 kDa), and His_6-tagged CFIm68 (38 kDa) (lower panel). (**E**) Northern Blot analysis showing the presence of *CCAT2* and intron 14 (1kb fragment) in the lysate pulled down with Streptavidin beads (upper panel). Western Blot analysis showing the presence of CFIm25, monomer (26 kDa) and dimer (64 kDa), and His_6-tagged CFIm68 (38 kDa) (lower panel). (**F**) Northern Blot analysis showing the presence of *CCAT2* and intron 14 (1kb fragment) in the lysate pulled down with Streptavidin beads specific for biotin tags. (**G**) Northern Blot analysis showing the presence of *CCAT2* in the lysate pulled down with the histidine affinity TALON resin (upper panel). Western Blot analysis showing the presence of the His_6-tagged CFIm complex in the lysate pulled down with the histidine affinity TALON resin (lower panel).

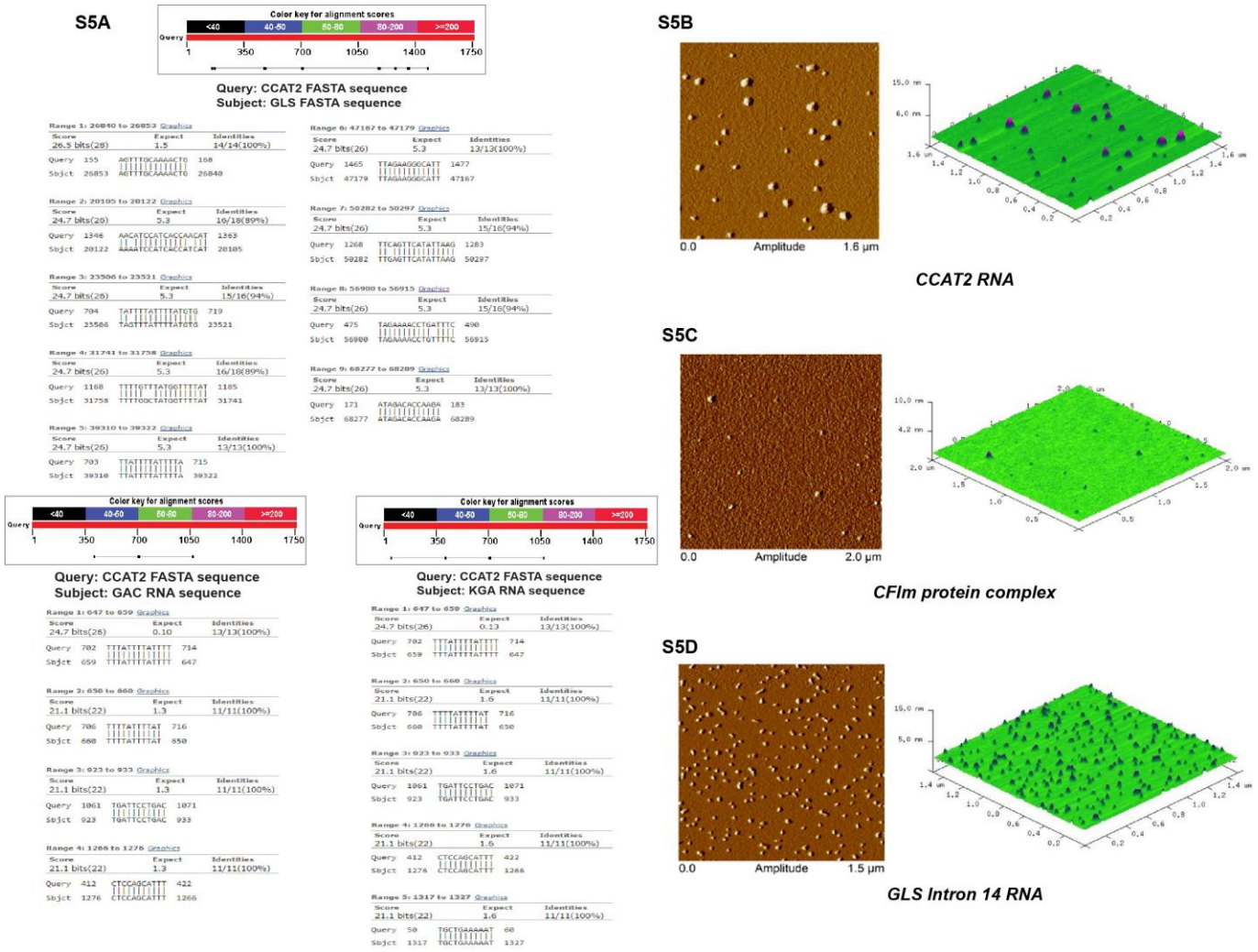
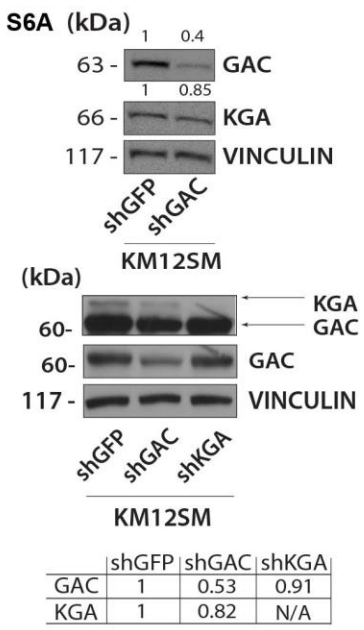
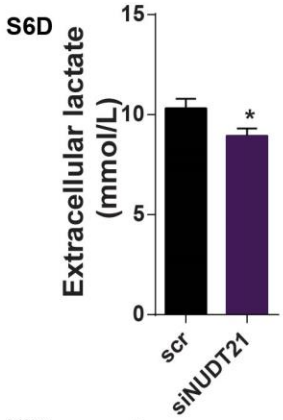
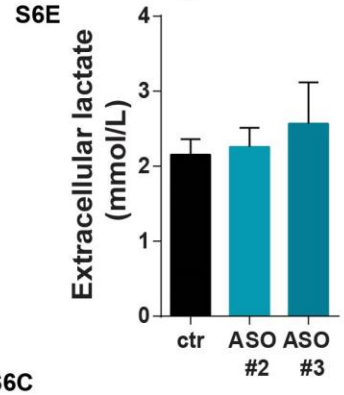


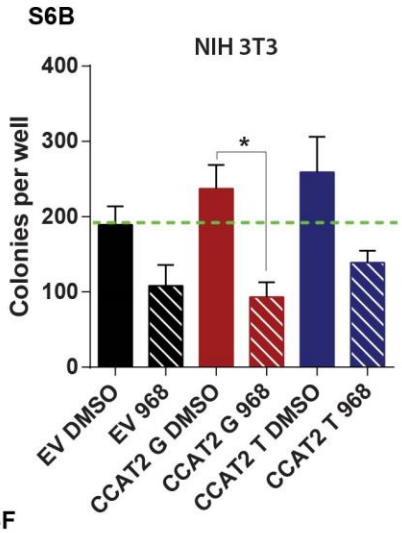
Fig. S5. ***CCAT2*** **interacts with both** ***GLS*** **pre-mRNA and the CFIm protein complex (See also Fig. 4).** (**A**) Snap-shot of NCBI Blast sequence alignment (http://blast.ncbi.nlm.nih.gov/BlastAlign.cgi) presenting the result of CCAT2-GLS alignments for the FASTA sequence for DNA (upper panel) and GAC/KGA RNA sequences (lower panels). AFM images of *CCAT2* RNA (**B**), *GLS* intron 14 RNA (**C**) and CFIm protein complex (**D**)

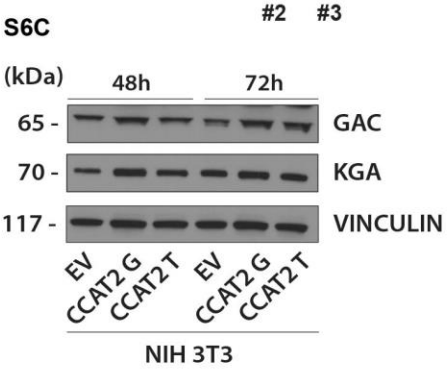


	shGFP	shGAC	shKGA
GAC	1	0.53	0.91
KGA	1	0.82	N/A









S6F

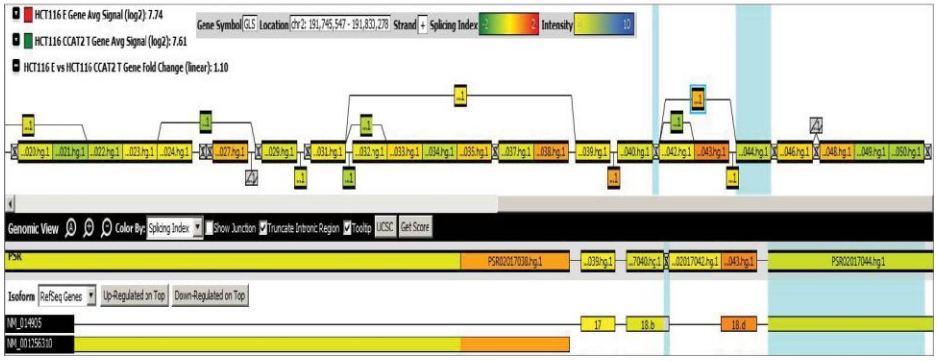


Fig. S6. The GLS isoforms contribute to the *CCAT2* induced metabolic phenotype and the *CCAT2* G-allele expressing cells depend on GAC for survival (See also Fig. 5 and Table S4) (**A**) Western blot analysis of GAC and KGA in KM12SM cells with stable downregulation of either GAC or KGA. The shKGA was generated as a control for the specificity of the shGAC construct. (**B**) Colony formation assay in soft agar of NIH3T3 cells transiently transfected with *CCAT2*-overexpressing G- or T-allele plasmids, and empty plasmid (E), and treated with GLS inhibitor, compound 968 and DMSO (left panel). (**C**) Western blot analysis for GAC and KGA in the same cellular model. Cells were collected for protein extraction 48h and 72h after transfection (right panel). Extracellular lactate concentration measured in HCT116 *CCAT2*-overexpressing cells (OC1 – GG genotype) transfected with siRNA against *NUDT21* and scrambled (**D**) and ASOs for inhibiting the binding sites of CFIm25 (**E**). (**F**) Snap-shot of Affymetrix HTA 2.0 software analysis (TAC) presenting the distribution of the two GLS isoforms between HCT116 *CCAT2*-overexpressing G- and T-allele cells. The junction connecting exon 14 and exon 15 is highlighted. The SI (splicing index) is presented above.

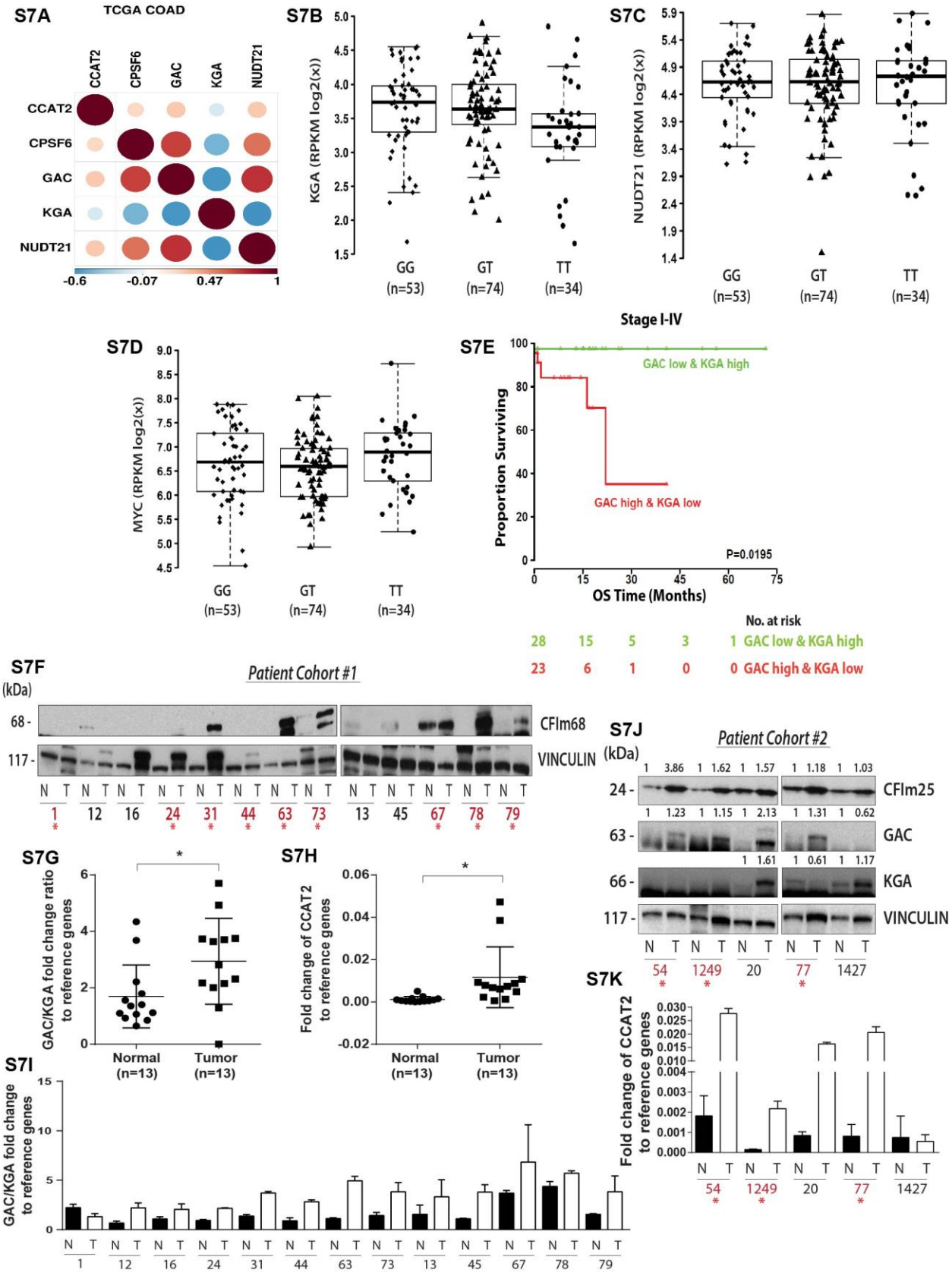

S7A
TCGA COAD
CCAT2
CPSF6
GAC
KGA
NUDT21
-0.6
-0.07
0.47
1
S7B
KGA (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7C
NUDT21 (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7D
MYC (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7E
Stage I-IV
GAC low & KGA high
GAC high & KGA low
Proportion Surviving
P=0.0195
OS Time (Months)
No. at risk
28 15 5 3 1 GAC low & KGA high
23 6 1 0 0 GAC high & KGA low
S7F
Patient Cohort #1
(kDa)
68 -
117 -
CFIm68
VINCULIN
N T N T N T N T N T N T N T N T N T N T N T N T N T N T
1 12 16 24 31 44 63 73 13 45 67 78 79
S7G
GAC/KGA fold change ratio to reference genes
Normal (n=13)
Tumor (n=13)
S7H
Fold change of CCAT2 to reference genes
Normal (n=13)
Tumor (n=13)
S7I
GAC/KGA fold change to reference genes
N T N T N T N T N T N T N T N T N T N T N T N T N T
1 12 16 24 31 44 63 73 13 45 67 78 79
S7J
Patient Cohort #2
(kDa)
1 3.86 1 1.62 1 1.57 1 1.18 1 1.03
24 -
CFIm25
1 1.23 1 1.15 1 2.13 1 1.31 1 0.62
63 -
GAC
1 1.61 1 0.61 1 1.17
66 -
KGA
117 -
VINCULIN
N T N T N T N T N T
54 1249 20 77 1427
S7K
Fold change of CCAT2 to reference genes
N T N T N T N T N T
54 1249 20 77 1427

Fig. S7. *CCAT2, NUDT21* (CFIm25), *CPSF6* (CFIm68) and *GLS* (GAC and KGA) expression pattern in patients with colon cancer. (See also Fig. 6 and 7) (**A**) Spearman's correlation between *CCAT2*, *CPSF6, NUDT21*, *GAC* and *KGA* mRNA expression for TCGA RNA-Seq colon cancer dataset. The size of the dots is directly correlated to the strength of the Spearman's coefficient and the corresponding scale is presented below the image. The red color represents direct a correlation and the blue color represents an indirect a correlation. *P* values < 0.02. Association of *KGA* (**B**), *NUDT21* (**C**) and *MYC* (**D**) mRNA expression with the genotypes (GG, GT and TT) of the rs6983267 SNP for CRC patients (TCGA RNA-Seq dataset)**.** (**E**) Kaplan-Meier survival curves for TCGA RNA-Seq colon cancer dataset as a function of GAC and KGA levels. (**F**) Western blot analysis of CFIm68 in *Patient Cohort #1* (13 paired normal and tumor tissue samples). Ratio of expression of GAC/KGA (**G**) and expression of *CCAT2* (**H**) in *Patient Cohort #1* by RT-qPCR. (**I**) RT-qPCR assessing the *GAC/KGA* mRNA expression ratio in *Patient Cohort #1*. (**J**) Western blot analysis of CFIm25, GAC and *KGA in Patient Cohort #2* (5 paired normal and tumor tissue samples). (**K**) *CCAT2* expression in *Patient Cohort #2* measured by RT-qPCR. The samples for which *CCAT2* expression correlates with CFIm25 and GAC protein expression are marked with red stars. Results are presented as mean value ± SD.

Table S1A. Metabolic pathways found to be differentially regulated by the two alleles using ConsensusPathDB-human. Results are presented as log2(G/T). If the calculated log2 is >0, the metabolic pathways is upregulated in G. See also **Fig.1** and **S1**.

Metabolic pathways	G vs. T*			
	in vivo	***P* value**	***in vitro***	***P* value**
Pyruvate metabolism and Citric Acid (TCA) cycle	0.4345	**0.0154**	1.2019	**0.0156**
Glucose metabolism	0.4997	**0.0400**	0.7431	**0.0167**
Glycolysis	0.7214	**0.0007**	1.0348	**0.0028**
The citric acid (TCA) cycle and respiratory electron transport	0.3956	**0.0209**	1.0342	**0.0070**
Activation of NMDA receptor upon glutamate binding and postsynaptic events	1.5608	**0.0285**	0.8210	**0.0200**
Lysine biosynthesis	0.4560	**0.0325**	0.9666	**0.0063**
Superpathway of conversion of glucose to acetyl CoA and entry into the TCA cycle	0.3197	**0.0170**	1.2859	**0.0112**

**Data are presented as log2 (G/T).*

Table S1B. List of metabolites found to be differentially regulated between the different experimental conditions using ConsensusPathDB-human. Results are presented as log2(G/T). See also **Fig.1** and **S1**.

	In vitro						***In vivo***					
	Log2 GvsT	***p* value**	**Log2 EvsT**	***p* value**	**Log EvsG**	***p* value**	**Log2 GvsT**	***p* value**	**Log2 EvsT**	***p* value**	**Log EvsG**	***p* value**
Glucose metabolism												
Biotin	0.538	0.326	-0.861	0.237	-1.399	0.340	-1.039	1.000	only in E	0.008	0.503	0.071
L-Glutamate	-0.122	0.142	-0.800	0.008	-0.678	0.007	-0.825	0.167	0.641	0.066	-0.184	0.187
Orthophosphate	-1.339	0.056	-1.388	0.079	-0.049	0.104	-1.253	0.167	1.114	0.002	0.143	0.242
ADP	0.746	0.678	-1.284	0.628	-2.030	0.384	-0.847	0.167	0.620	0.032	-0.227	0.181
NADH	-0.174	0.523	-2.404	0.095	-2.230	0.097	-0.467	0.667	0.594	0.262	0.154	0.250
NAD+	-1.389	0.145	-0.323	0.005	1.066	0.005						
3',5'-Cyclic AMP	-0.429	0.762	-1.348	0.481	-0.919	0.581	-1.440	0.167	1.374	0.060	0.045	0.337

Pyruvate	-2.761	0.038	0.188	0.689	2.949	0.032	only in T	0.979	only in T	0.002	0.786	0.005
UMP	1.070	0.071	-0.880	0.110	-1.949	0.103	only in G	0.979	1.292	0.023	-0.229	0.166
3-Phospho-D-glycerate							-0.978	0.110	1.547	0.052	0.438	0.269
Citrate	0.484	0.710	-3.453	0.073	-3.938	0.317	-0.226	0.167	2.699	0.002	-0.028	0.071
Diphosphate							-2.802	0.110	3.322	0.013	0.130	0.023
Glycerone phosphate	0.092	0.808	0.508	0.107	0.416	0.282	only in T	0.979	1.145	0.039	-0.425	0.059
UDP	-2.638	0.102	only in T	0.021	only in G	<0.0 0001						
Pyridoxal phosphate	0.813	0.532	only in T	0.423	only in G	0.021						
UDP-glucose							-0.585	1.000	-0.614	0.322	-1.199	0.206
D-Fructose 1,6-bisphosphate	0.065	0.906	-0.183	0.622	-0.248	0.680						
D-Fructose 6-phosphate							-0.056	0.333	0.730	0.094	0.806	0.032
L-Aspartate							6.958	0.110	0.558	0.119	0.073	0.123
Glycolysis												
Citrate	0.484	0.710	-3.453	0.073	-3.938	0.317	-0.226	0.167	2.699	0.002	-0.028	0.071
D-Fructose 6-phosphate							-0.056	0.333	0.730	0.094	0.806	0.032
D-Fructose 1,6-bisphosphate	0.065	0.906	-0.183	0.622	-0.248	0.680						
Glycerone phosphate	0.092	0.808	0.508	0.107	0.416	0.282	only in T	0.979	1.145	0.039	-0.425	0.059
Orthophosphate	-1.339	0.056	-1.388	0.079	-0.049	0.104	-1.253	0.167	1.114	0.002	0.143	0.242
ADP	0.746	0.678	-1.284	0.628	-2.030	0.384	-0.847	0.167	0.620	0.032	-0.227	0.181
NADH	-0.174	0.523	-2.404	0.095	-2.230	0.097	-0.467	0.667	0.594	0.262	0.154	0.250
NAD+	-1.389	0.145	-0.323	0.005	1.066	0.005						
3',5'-Cyclic AMP	-0.429	0.762	-1.348	0.481	-0.919	0.581	-1.440	0.167	1.374	0.060	0.045	0.337
Pyruvate	-2.761	0.038	0.188	0.689	2.949	0.032	only in T	0.979	only in T	0.002	0.786	0.005
3-Phospho-D-glycerate							-0.978	0.110	1.547	0.052	0.438	0.269
Superpathway of conversion of glucose to acetyl CoA and entry into the TCA cycle												
Citrate	0.484	0.710	-3.453	0.073	-3.938	0.317	-0.226	0.167	2.699	0.002	-0.028	0.071
Fumarate	0.845	0.501	-0.209	0.906	-1.054	0.489	-2.594	0.110	1.461	0.016	0.083	0.103
Glycerone phosphate	0.092	0.808	0.508	0.107	0.416	0.282	only in T	0.979	1.145	0.039	-0.425	0.059
Orthophosphate	-1.339	0.056	-1.388	0.079	-0.049	0.104	-1.253	0.167	1.114	0.002	0.143	0.242
ADP	0.746	0.678	-1.284	0.628	-2.030	0.384	-0.847	0.167	0.620	0.032	-0.227	0.181
D-Fructose 1,6-bisphosphate	0.065	0.906	-0.183	0.622	-0.248	0.680						
cis-Aconitate							-0.392	0.667	only in E	0.004	0.237	0.173
NAD+	-1.389	0.145	-0.323	0.005	1.066	0.005						

Succinate	0.245	0.873	1.535	0.588	1.290	0.615		0.979	1.297	0.061	0.578	0.106
NADH	-0.174	0.523	-2.404	0.095	-2.230	0.097	-0.467	0.667	0.594	0.262	0.154	0.250
Pyruvate	-2.761	0.038	0.188	0.689	2.949	0.032	only in T	0.979	only in T	0.002	0.786	0.005
3-Phospho-D-glycerate							-0.978	0.110	1.547	0.052	0.438	0.269
The citric acid (TCA) cycle and respiratory electron transport												
Citrate	0.484	0.710	-3.453	0.073	-3.938	0.317	-0.226	0.167	2.699	0.002	-0.028	0.071
Fumarate	0.845	0.501	-0.209	0.906	-1.054	0.489	-2.594	0.110	1.461	0.016	0.083	0.103
Orthophosphate	-1.339	0.056	-1.388	0.079	-0.049	0.104	-1.253	0.167	1.114	0.002	0.143	0.242
Pyruvate	-2.761	0.038	0.188	0.689	2.949	0.032	only in T	0.979	only in T	0.002	0.786	0.005
Lipoamide	-0.477	0.081	-1.908	0.059	-1.430	<0.0 0001						
NADH	-0.174	0.523	-2.404	0.095	-2.230	0.097	-0.467	0.667	0.594	0.262	0.154	0.250
NAD+	-1.389	0.145	-0.323	0.005	1.066	0.005						
Succinate	0.245	0.873	1.535	0.588	1.290	0.615	only in T	0.979	1.297	0.061	0.578	0.106
ADP	0.746	0.678	-1.284	0.628	-2.030	0.384	-0.847	0.167	0.620	0.032	-0.227	0.181
Pyruvate metabolism and Citric Acid (TCA) cycle												
Citrate	0.48	0.71	-3.45	0.07	-3.94	0.32	-0.226	0.167	2.699	0.002	-0.028	0.071
Fumarate	0.84	0.50	-0.21	0.91	-1.05	0.49	-2.594	0.110	1.461	0.016	0.083	0.103
Orthophosphate	-1.34	0.06	-1.39	0.08	-0.05	0.10	-1.253	0.167	1.114	0.002	0.143	0.242
ADP	0.75	0.68	-1.28	0.63	-2.03	0.38	-0.847	0.167	0.620	0.032	-0.227	0.181
Lipoamide	-0.48	0.08	-1.91	0.06	-1.43	<0.0 0001						
NADH	-0.17	0.52	-2.40	0.09	-2.23	0.10	-0.467	0.667	0.594	0.262	0.154	0.250
NAD+	-1.39	0.15	-0.32	0.00	1.07	0.00						
Succinate	0.25	0.87	1.54	0.59	1.29	0.61	only in T	0.979	1.297	0.061	0.578	0.106
Pyruvate	-2.76	0.04	0.19	0.69	2.95	0.03	only in T	0.979	only in T	0.002	0.786	0.005

Table S2. FACS analysis of cells transfected with the RG6 and RG6 Int 14 plasmids. EGFP/dsRED is normalized to the control (RG6). The efficiency of the NUDT21 knockdown in KM12SM cells at 72 hours after transfection, both for the VECTRA and the FACS analysis, is presented below. See also **Fig. 2** and **S2**.

Sample	*FITC-A mean*	*mCherry-A mean*	*EGFP/dsRED*	*Norm by empty vector*	*Norm by control*
HCT116 E RG6	5595	4115	1.36	1	-
HCT116 OC1 RG6	1926	1238	1.56	1	-
HCT116 E RG6 Int 14	6184	182	33.98	**24.98**	**1**
HCT116 OC1 RG6 Int 14	2565	55	46.64	**29.89**	**1.2**
HCT116 G-allele RG6	2281	1997	1.14	1	-
HCT116 T-allele RG6	2052	1684	1.22	1	-
HCT116 G-allele RG6 Int 14	3252	77	42.23	**37.04**	**1**
HCT116 T-allele RG6 Int14	2713	65	41.74	**34.2**	**0.92**
KM12SM scr RG6	3613	3670	0.98	1	-
KM12SM siNUDT21 RG6	3808	4165	0.91	1	-
KM12SM scr RG6 Int14	11761	494	23.81	**23.57**	**1**
KM12SM siNUDT21 RG6 Int14	9867	699	14.12	**15.51**	**0.65**

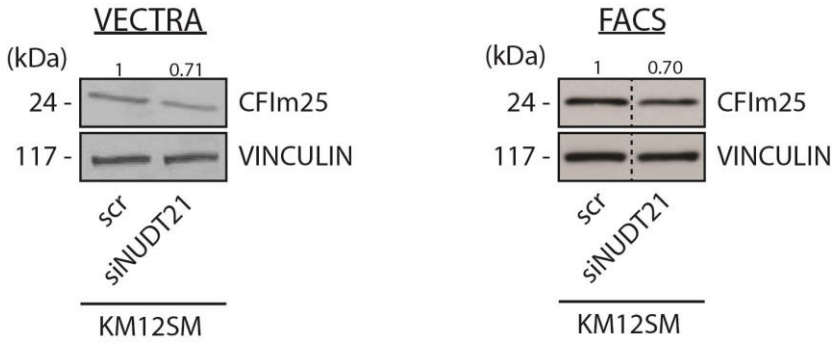


Table S3. Quantification of the Western Blot bands from **Fig. 4D**. Data is presented as ratio between the protein band and the corresponding *CCAT2* band (Northern Blot). See also **Fig. 4** and **S4**.

Intron 14	Binding partners	Ratio CCAT2:Protein
+	G + CFIm25monomer	0.1703
+	T + CFIm25monomer	0.1649
+	C + CFIm25monomer	0.0378
-	G + CFIm25monomer	0.1806
-	T + CFIm25monomer	0.1812
-	C + CFIm25monomer	0.1133
+	G + CFIm25dimer	0.1914
+	T + CFIm25dimer	0.1764
+	C + CFIm25dimer	0.0636
-	G + CFIm25dimer	0.3402
-	T + CFIm25dimer	0.2997
-	C + CFIm25dimer	0.1533
+	G + CFIm68	0.2807
+	T + CFIm68	0.2176
+	C + CFIm68	0.0404
-	G + CFIm68	0.2782
-	T + CFIm68	0.2115
-	C + CFIm68	0.1226

Table S4. Pathway analysis of genes found significantly spliced between HCT116 cells overexpressing *CCAT2* G- and T-allele by Affymetrix HTA 2.0 array. See also **Fig. 5, S6** and **7**.

Pathway	Count	*p* value	Genes	Fold Enrichment
Reversal of Insulin resistance by Leptin	3	0.0202	LEPR, LEPROT, CPT1A	12.605
Fructose and mannose metabolism	7	0.0394	MTMR2, AKR1B15, PFKFB3, ALDOB, PFKFB1, TNNI3K, PHPT1, FPGT	2.719
Dilated cardiomyopathy	16	0.0034	ADCY3, TNNC1, MYL3, ADCY5, MYBPC3, ITGA4, TPM1, PRKX, LAMA2, TNNT2, ITGA9, ITGAV, ITGB7, DMD, RYR2, CACNA1D	2.297
ECM-receptor interaction	13	0.0228	COL4A4, HSPG2, ITGA4, SDC2, COL4A6, COL5A1, LAMA2, LAMA1, ITGA9, LAMA4, ITGAV, ITGB7, COL1A2	2.044
Hypertrophic cardiomyopathy (HCM)	13	0.0248	TNNT2, LAMA2, ITGA9, MYL3, TNNC1, ITGB7, ITGAV, DMD, MYBPC3, RYR2, ITGA4, CACNA1D, TPM1	2.020
Cell adhesion molecules (CAMs)	19	0.0097	PTPRC, PTPRF, NLGN1, NFASC, HLA-A, CLDN22, NRXN1, ITGA4, CLDN11, HLA-DQA2, SDC2, HLA-F, ITGA9, ITGAV, ITGB7, CNTN1, MADCAM1, JAM2, CD226	1.901
Metabolism of carbohydrates	13	0.0408	PFKFB3, SI, PHKA1, PFKFB1, ALDOB, GALT, PCK2, SLC25A13, PYGL, MGAM, CALM3, AMY2B, CALM2, UGP2, CALM1	1.872
Calcium signaling pathway	22	0.0232	ADCY3, TNNC1, MYLK3, ERBB2, PHKA1, LHCGR, HTR4, PRKX, ITPR2, P2RX5, ATP2B1, AGTR1, P2RX7, PLCB4, CAMK4, CHRM1, CALM3, RYR2, CACNA1E, HTR2C, CACNA1D, CALM2, CACNA1A, CALM1	1.651

Table S5. Primers, oligoprobes and siRNA used in this study.

Gene	Sequence	Description
CCAT2 F	5’ CCCTGGTCAAATTGCTTAACCT 3’	PCR Primer
CCAT2 R	5’ TTATTCGTCCCTCTGTTTTATGGAT 3’	PCR Primer
U6 F	5’ CTCGCTTCGGCAGCACA 3’	PCR Primer
U6 R	5’ AACGCTTCACGAATTTGCGT 3’	PCR Primer
HPRT1 F	5’ TGACACTGGCAAAACAATGCA 3’	PCR Primer
HPRT1 R	5’ GGTCCTTTTCACCAGCAAGCT 3’	PCR Primer
GAC F	5’ CTTGATCCTCGAAGAGAAGGTG 3’	PCR Primer
GAC R	5’ ACAGTTGTAGAGATGTCCTCATTT 3’	PCR Primer
KGA F	5’ AAGAGAAGGTGGTGATCAAAGG 3’	PCR Primer
KGA R	5’ AGCTACATGGAGTGCTGTTC 3’	PCR Primer
ACTIN F	5’ AGCCTCGCCTTTGCCGA 3’	PCR Primer
ACTIN R	5’ GCGCGGCGATATCATCATC 3’	PCR Primer
NUDT21_PROM F	5’ GGACGCTGTTACGGAAGAAA 3’	ChiP Primer
NUDT21_PROM R	5’ GGTGGCCAAAGTCCTCTAAG 3’	ChiP Primer
GLS_intron14 F	5’ AGAAGTGCATTTGTTGGTCTTT 3’	RIP Primer
GLS_intron14 R	5’ AGAAGTACCTTCTATTGCCACTAA 3’	RIP Primer
CCAT2 siRNA 1	5’ AGGTGTAGCCAGAGTTAAT 3’	siRNA target sequence
CCAT2 siRNA 2	5’ AGGAAGAGGTTAAGCAATT 3’	siRNA target sequence
Intron141 F	5’ AGCTAATACGACTCACTATAGGGACACCACCATGTCGCAATCTTGGGTGCTGGAG 3’	*In vitro* transcription
Intron141 R	5’ TCTGACAGAACAGAGGAGTTGCAATAAAAAAA 3’	*In vitro* transcription
Intron 14 Probe F	5’ GGGTGCTGGAGCCATAAA 3’	Northern Blot
Intron 14 Probe R	5’ CTTCTATTGCCACTAAAGACATCAC 3’	Northern Blot
CCAT2 Probe F	5’ TTTAGCAGCTGCATCGCTCCATAG 3’	Northern Blot
CCAT2 Probe R	5’ CTGGGCAGTTGAGAAACGAGAACA 3’	Northern Blot
KGA RACE F	5’ GGACATGGAACAGCGGGACTATGA 3’	3’ RACE Primer
KGA RACE nPCR F	5’ GCTGCTGCAGAGGGTAATACAGGAACTA 3’	3’ RACE Primer
GAC RACE F	5’ GGACCATTGGACTATGAAAGTCTCCAACA 3’	3’ RACE Primer
GAC RACE nPCR F	5’ CCTGAGTCAAATGAGGACATCTCTAC 3’	3’ RACE Primer

KGA probe F	5’ TCATTGTGCACACAGGACAA 3’	Southern Blot Oligo Probe
KGA probe R	5’ TCATGGAAGACCACACAAGC 3’	Southern Blot Oligo Probe
GAC probe F	5’ TGGGTTCTAGTTTCAGAATGTTTC 3’	Southern Blot Oligo Probe
GAC probe R	5’ TTTTTAAAGACCAACAAATGCAC 3’	Southern Blot Oligo Probe

Table S6. Antibodies used in this study.

Antibodies	Product information
VINCULIN	Santa Cruz (sc-7649)
Actin	Abcam (AB3280)
MYC	Abcam (9E10)
NUDT21	Proteintech (10322-1-AP)
CPSF6 N-terminal	Aviva (OAAB15637)
KGA	Abcam (AB60709)
GAC	Genscript – custom antibody (Cassago et al., 2012)
GLS [EP7212]	Abcam (AB156876)
His-probe [H-3]	Santa Cruz (sc-8036)

Supplementary Experimental Procedures

Cell culture. Human colorectal cancer cell lines HCT116, KM12SM and COLO320DM were obtained from the American Type Culture Collection and grown as suggested by the supplier. Cells were cultured at 37^{o}C in 5% CO_2. All cell lines were validated by The Characterized Cell Line Core at MD Anderson Cancer Center using STR DNA fingerprinting.

Reverse Transcription Quantitative RT-PCR Analysis. cDNA was generated with Superscript III reverse transcriptase (Invitrogen) and random hexamers. RT-qPCR was performed with iQ SYBR Green SuperMix (Bio-rad) and the appropriate primers. Experiments were performed in duplicate and the data were normalized to the mRNA levels of *U6, HPRT1* or *Actin* using the $2^{-\Delta Ct}$ method. Primer sequences are available in **Table S5.**

Western Blot. Proteins were collected from cultured cells and lysed with RIPA buffer (SIGMA) freshly supplemented with a complete protease inhibitor cocktail (SIGMA). The Bradford assay was used to measure protein concentrations. Proteins were separated on polyacrylamide gel (Bio-rad) electrophoresis and transferred to a 0.2 µm nitrocellulose membrane (Bio-rad). The list of antibodies can be found in **Table S6.** Vinculin was used as loading control and quantification of protein expression was done with Adobe Photoshop CS6.

TaqMan microRNA assay. MiRNA expression was evaluated with the TaqMan microRNA assay (Applied Biosystems). Briefly, cDNA was synthesized with the TaqMan Reverse Transcription Reagents kit (Applied Biosystems) and employed for quantitative RT-PCR analysis together with TaqMan probes and SsoFast Supermix (Bio-rad). Primers and probes for miR-23a (000399), and internal controls U6 snRNA (001973) and RNU48 snRNA (001006) were purchased from Applied Biosystems. Experiments were performed in duplicate and normalized to the internal control. Relative expression levels were calculated using the $2^{-\Delta Ct}$ method.

RNAi experiments. MiRNA precursor molecules pre-miR-23a (PM10644) and pre-miRNA negative control #1 were purchased from Ambion. siRNAs targeting human *MYC* (siGENOME SMART pool M-003282-07-0005), *GLS* (ON-TARGETplus SMART pool L-004548-01-0005) and siGENOME Non-Targeting siRNA #2 (D-001210-02-05) were purchased from Dharmacon Research Inc, while *NUDT21* esiRNA (EHU041141-20UG) was obtained from SIGMA and

CCAT2 siRNAs were custom designed by SIGMA (**Table S5**) and. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol using a final concentration of 50 nM for miRNA precursor molecules and all siRNAs except *CCAT2* siRNAs for which the final concentration was 100 nM. RNA and proteins were collected after 48 h and 72 h, respectively. Transfection efficiencies were evaluated by reverse transcription RT-qPCR.

Antisense synthetic oligonucleotides (ASOs). The 22-27 nt antisense oligonucleotides (ASOs) blocking the binding sites of CFIm25 were designed and purchased from IDT. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol using a final concentration of 200 nM. Untreated cells were used as controls. Proteins were collected at 48 h after transfection.

Plasmid generation. The plasmid was generated as previously described (Ling et al., 2013). Briefly, the *CCAT2* genomic locus was amplified with specific primers from HCT116 genomic DNA with *Pfu* polymerase (Invitrogen), verified by sequencing and ligated into a pcDNA 3.1 vector. The pcDNA plasmid was then linearized with BglII and transfected into HCT116 cells using Lipofectamine 2000 (Invitrogen). Single clones were selected with G418 (0.5 mg/ml). For the Tet-ON inducible COLO320 clones, the Knockout Single Vector Inducible RNAi System was used according to the manufacturer's protocol.

Stable cell lines generation. The generation of stable cell lines was performed as previously described (Ling et al., 2013). HCT116 cells stably overexpressing *CCAT2* (OC1 and OC3) were generated by transfection of pcDNA 3.1 expression vector transfection (Invitrogen). Two clones overexpressing *CCAT2* (OC1 and OC3) having high expression of *CCAT2* were randomly chosen for further experiments. The E1 clone was obtained by transfecting HCT116 cells with the empty pcDNA 3.1 vector. The HCT116 G, T clones were generated by cloning the *CCAT2* sequence it into the pMX plasmid. The *CCAT2* containing retroviruses, as well as the empty pMX plasmid (E) were then transfected in 293 GP2 cell lines, and the virus-containing supernatant was used to infect HCT116 cells. After infection, HCT116 cells were grown in complete media containing Puromycin (1 μg/mL).

The GFP, GAC and KGA short-hairpin sequences were cloned into pLKO1.puro vectors and the 293FT cells were transfected to collect the collect the collect the virus-soup. KM12SM cells were infected with the virus soup and selected with Puromycin (0.5 μg/mL).

RNA extraction. Cells were harvested, washed with PBS and resuspended in 1 ml Trizol (Invitrogen). RNA extraction was performed according to manufacturer's protocol. RNA quantity was assessed with NanoDrop ND-1000 (Thermo Fisher Scientific) and the integrity was analyzed by gel electrophoresis. Samples that did not present clearly defined 28S and 18S ribosomal bands were considered of low quality and excluded from the study.

PCR. End point PCR was performed with Red Taq DNA Polymerase kit (Sigma) following manufacturer's instructions and using 10-20 ng DNA template, depending on the reaction. The products were run on 1 or 2% agarose gels (depending on the amplicon size). Primer sequences are available in **Table S4**.

Treatment with Compound 968 and proliferation assay. Compound 968 (5-(3-Bromo-4-(dimethylamino)phenyl)-2,2-dimethyl-2,3,5,6-tetrahydrobenzo[a]phenanthridin-4(1H)-one) was purchased from Calbiochem and resuspended according to manufacturer's protocol. HCT116 OC1 cells were seeded into 24 well plates 24 h before treatment. Cells were treated either with DMSO (as control) or with two different concentrations of 968, 5 μM and 10 μM. Proliferation was assessed by cell counting for 120 h. Each time point represents the mean value of a triplicate ± SD.

Luciferase reporter assay. Fragments containing miR-23a putative binding sites were amplified by PCR with primers containing the Xbal restriction site. The PCR products were purified, digested and directly cloned into the Xbal site of the pGL3 control vector (Promega Corporation), which is located downstream of the firefly luciferase reported gene. The QuickChange II XL site-directed mutagenesis kit (Agilent Technologies) was used to generate mutations in the miRNA-binding sites. HEK293 cells were seeded ($1x10^5$ cells/well) in 24-well plates and co-transfected with 50 nM scrambled, miR-23a and 0.4 μg pGL3-putative binding site plasmid or pGL3-mutated putative binding site plasmid, together with Rennilla luciferase construct, which was used as a normalization reference. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Cells were lysed 48 h after transfection and luciferase activity was measured using a dual-luciferase reporter assay

system (Promega) in the Veritas microplate luminometer (Turner Biosystems). Two independent experiments were performed in triplicate. The pLightswitch_prom vectors containing the NUDT21 promoter fragment and the NPM1 promoter fragment were purchased from SwitchGear. The luciferase repoter assay for NUDT21 and NPM1 promoters was performed in HCT116. Cells were seeded in 96-well plates (8000 cells) 24 h before co-transfection with 50nM siMYC or siRNA negative control and 0.102 µg pLightswitch_prom plasmids containing either NUDT21 and NPM1 promoters or the empty plasmid. Transfections were performed with Lipofectamine 2000 (Invitrogen) according to manufacturer's protocol. Cells were lysed 48 h after transfection and luciferase activity was measured using LightSwitch Luciferase Assay Reagent (SwitchGear) in the Veritas microplate luminometer (Turner Biosystems). Two independent experiments were performed in quadruplicates.

Colony formation in soft agar. NIH3T3 cells were transiently transfected with pcDNA 3.1 CCAT2-overexpressing G- or T-allele vectors and empty vector (E) using Lipofectamine 2000. Twenty-four hours after transfection, cells were harvested, counted and plated in soft agar. Cells were allowed to form colonies for fourteen days at 37^{o}C and 5% CO_2, after which the colonies were counted.

Chromatin immunoprecipitation (ChiP). Chromatin immunoprecipitation (ChIP) was performed with the EZ-ChiP kit (Millipore). Chromatin was prepared from HCT116 cells following crosslinking with 11% formaldehyde solution. DNA was sheared to an average size of ~500 bp using a sonicator. Lysates were pre-cleared with Protein G-conjugated agarose beads and subsequently immunoprecipitated overnight at 4°C using 5 µg of the following antibodies: anti-MYC (Millipore) and rabbit IgG (Millipore). Immunoprecipitates were sequentially washed with low salt, high salt, and TE buffer. Elutes were collected and reverse crosslinked at 65°C for 4 hours. ChIP DNA was treated with Proteinase K at 45°C for 2 hours, followed by purification with the phenol/chloroform extraction method.

3'RACE and Southern blot analysis. Amplification of 3' cDNA ends was performed using the SMARTer RACE cDNA Amplification kit (Clontech) following manufacturer's instruction, starting from 1 µg RNA for each HCT116 clone (E and OC1). For both PCRs (first PCR and nested PCR) we used the reverse primers provided with the kit (UPM primer – first PCR and NUP primer – nested PCR) together with isoform specific primers (KGA RACE F primer or

GAC RACE F Primer – first PCR and KGA nPCR F primer or GAC nPCR F primer – nested PCR). After the nested PCR, the samples were run on a 1% agarose gel. The DNA was then transferred overnight to the membrane by incubating in 0.4 N NaOH and UV cross-linked. The specific probes were prepared with the Prime-it II kit (Stratagene) and labeled with ^{32}Pα isotope. The membranes were hybridized overnight with each probe and after washing, the signal was assessed by autoradiography. Primer and probe sequences are available in **Table S5**.

RNA *in vitro* transcription and purification. The assay was performed using the MEGAscript kit (Ambion) and following the manufacturer's protocol. For the *CCAT2* sequences, the linearized pcDNA3.1 plasmids containing the G or the T allele were used as templates, while for the GLS intron 14 sequence, a PCR product of 600 bp was generated from gDNA to serve as template. The transcription reaction was optimized according to the length of the sequence, 9 h for GLS intron 14 and 4 h for *CCAT2*. To eliminate the remaining DNA template, 1 µl TURBO DNase was added at the end of transcription to the mix and incubated for 15 min at 37^{o}C. *In vitro* transcribed RNA was purified using the MEGAclear kit (Ambion) according to the manufacturer's recommendations and quantity was assessed with NanoDrop ND-1000 (Thermo Fisher Scientific).

RNA immunoprecipitation (RIP). The RNA-protein complexes were immunoprecipitated with the Magna RIP kit (Millipore) following the manufacturer's protocol. Briefly, 15 million cells were prepared in RIP lysis buffer and the RNA-protein complexes were immunoprecipitated using anti-NUDT21 (Proteintech) or anti-CPFS6 (Abviva) and normal rabbit IgG (as control) antibodies. RNA was purified using the phenol : chloroform : isoamyl alcohol method and further used for cDNA synthesis and PCR analysis.

Protein production and purification. Recombinant CFIm25 and CFIm68 proteins were purified as described previously (Coseno et al., 2008; Yang et al., 2011), respectively. Individually purified CFIm25 and CFIm68 were mixed in an equal molar ratio, concentrated and eluted on a Superdex 200 10/30 gel filtration column (GE Healthcare) with 20 mM Hepes pH 7.5, 50 mM KCl and 10% (v/v) glycerol.

His$_6$-tag pull-down assay. One µg purified CFIm complex (His$_6$-tagged) was mixed with 10 µg *CCAT2* G or T RNA and 15 µg intron 14 RNA, and 60 µl 50% slurry TALON metal affinity resin (Clontech) in a total volume of 300 µl buffer (20 mM TRIS-HCl pH 7.5, 50 mM KCl, 50

mM $MgCl_2$, 0.5 mM TCEP and 10% (v/v) glycerol). The mixture was incubated overnight at 4^{o}C with rotation. For each *CCAT2* allele one control sample was prepared consisting of same mixture without the protein complex. After incubation, the resin was first centrifuged to remove the supernatant and washed twice with buffer to remove excess RNA. For elution, imidazole was added to the buffer for a final concentration of 500 mM, and incubated for 30 min with the resin. After centrifugation, the elute containing the protein and RNAs, was divided in two, half was further used for western blot for protein detection and the other half was used for northern blot for RNA detection.

Biotinylated RNA pull-down assay. A biotin tag was added to the RNAs (*CCAT2* G, T and C) during in vitro transcription following manufacturer's protocol. RNA was purified as described above. Five hundred seven nanograms of biotinylated *CCAT2* RNA was incubated overnight with either same quantity of intron 14 RNA, or half the quantity of protein complex (CFIm) or with both intron 14 (570 ng) and CFIm (285 ng). One hundred ul Dynabeads MyOne Streptavidin C1 beads (Life Technologies) were prepared according to the manufacturer's protocol and incubated with the mix for 20 min at room temperature. After the washing steps, the RNAs were eluted from the beads with 25 ul elution solution (10 mM EDTA pH=8.2 in 95% formamide) by incubating 5 min at 65^{o}C. The elute was split in half for downstream Northern and Western blot analysis.

***In vitro* migration assay**. Sixty-five thousand cells were prepared in 100 µl serum-free media and seeded onto the gelatin coated insert and 500 µl of media with serum was added to the bottom of the well. Cells were left to migrate 16 h and 24 h depending on the cell line. The cells that migrated to the bottom of the insert were fixed, stained (hematoxylin staining reagents) and counted using a microscope. For each well, 7 different fields were counted and the average number of the cells per insert was determined. The experiment was performed twice in triplicate.

Whole cell lysate glutaminase activity assay. Cells were lysed in 25 mM HEPES (Sigma-Aldrich) pH=8.0, 150 mM NaCl (Merck), 1 mM EDTA (Merck), 0.01% Triton X-100 (USB), 10 mM Sodium Pyrophosphate (Sigma- Aldrich), 100 mM Sodium Fluoride (Sigma-Aldrich), 10 mM Sodium Orthovanadate (Sigma- Aldrich), 2 mM PMSF (Sigma-Aldrich), 10 µM Leupeptin (Sigma-Aldrich), 1 µM Pepstatin (Sigma-Aldrich), 2 µg/mL Aprotinin (Sigma-Aldrich) followed by 20 strokes through a 26 gauge needle. Glutaminase activity assay was performed

following an already published streamlined assay (Cassago et al., 2012). Briefly, 10 µg of bradford-quantified lysate cell lysates were assayed on a 96-well plate added of 50mM Tris-acetate pH=8.6, 0.5 units of bovine L-Glutamate Dehydrogenase (Sigma-Aldrich), 2 mM NAD (Sigma-Aldrich), 20 mM Dipotassium phosphate (Sigma-Aldrich) and 3.5 mM L-Glutamine (Sigma-Aldrich). NADH absorption at 340 nm was tracked over the time on the EnSpire Plate Reader (Perkin Elmer). The slope of the best r^2 fit was compared between the samples. The assay was performed with technical triplicates.

Glutamine/Glutamate quantification. For the Glutamate Quantification, a modified version of an existing protocol were used (Lund, 1986). After 48 or 72 hours of incubation with cells, 20 µL of culture medium were combined with 50mM Tris-acetate pH=8.6, 1.2 units of bovine L-Glutamate Dehydrogenase (Sigma-Aldrich), 2 mM NAD (Sigma-Aldrich) and 20 mM Dipotassium phosphate (Sigma-Aldrich) and incubated at 37°C for 90 minutes. NADH absorption at 340 nm was read on the EnSpire Plate Reader (Perkin Elmer) and compared to a standard curve prepared on media without glutamine. The glutamine plus glutamate total amount were quantified on the same reaction solution added of 30 nM of recombinant Glutaminase C. NADH absorption at 340 nm was tracked over the time on the EnSpire Plate Reader (Perkin Elmer) and the slope of the best r^2 compared to a standard curve. Final gutamine concentration was obtained after subtraction of glutamate concentration from total glutamine plus glutamate concentration. Data was normalized by cell mass as quantified by a Bradford reaction (BSA as standard). The assays were performed in technical triplicates.

O_2 consumption assay. Cells re-suspended in PBS were centrifuged in 500xg for 10 min. Cells were re-suspended in 1mL respiration buffer (130 mM sucrose, 50 mM KCl, 5 mM $MgCl_2$, 5 mM KH_2PO_4, 0.05 mM EDTA and 0.5mg/mL FAF BSA, pH 7.4) and centrifuged again at 500xg for 10 min. Final cell pellet was re-suspended in sufficient respiration buffer with a final concentration of $20x10^6$ cells/ml. Cells were placed in a 37^{o}C water bath while respiration assays were performed. Whole cell oxygen consumption was monitored at 37^{o}C using a Clarke-type electrode in a Strathkelvin Instrument Mitocell MT200 with an oxygen interface 929 system. The electrode was calibrated prior to each experiment. Respiration buffer was added to the oxygen chamber and electrode was allowed to stabilize before readings were recorded. Once stabilized, oxygen consumption was recorded for approximately 30 sec. Two million cells were added to the chamber and oxygen consumption was measured for 3-5 min. Ten µl of 100 mM potassium

cyanide (KCN) were added to chamber to inhibit respiration. Oxygen consumption was recorded for an additional 1-2 min. The rate of respiration was expressed as nmoles of O_2/min/2x10^6 cells (oxygen consumed/time/cells).

RG6 Assay. RG6 Splicing Reporter Vector (Orengo et al., 2006) (gift from Dr. Thomas A. Cooper) was modified for removal of the first cardiac troponin T intron and addition of the human GLS gene Intron 14. Twenty-four hours prior to transfection cells were plated in 6-well plates each containing two cover slips (Fisher Scientific). Transfections were performed with Lipofectamine 2000 according to the manufacturer's protocol. Forty-eight hours after transfection with the RG6 vectors, cells were washed with PBS, fixed with 4% Paraformaldehyde and stained with DAPI (Molecular Probes). The slides were imaged with the VECTRA Automated Imaging System (PerkinElmer) and analyzed with the inForm software.

RNA pull-down. The assay was performed according to the previously described protocol (Yoon et al., 2012). The *CCAT2* G and T sequences were cloned in the pMS2 vectors (pcDNA 3.1 plasmid containing 24 repeats of the MS2 tag – ACATGAGGATCACCCATGT) upstream of the MS2 tag in the EcoRI site. The *CCAT2* A, C and DEL vectors were generated by mutagenesis using the QuickChange II XL Mutagenesis kit (Stratagene, Agilent) and following manufacturer's protocol. HCT116 cells were co-transfected with *CCAT2* vector or empty vector and pMS2-GST vector. Forty-eight hours after transfection cells were harvested and proteins were lysed and quantified. Five hundred µl (2 µg/µl) lysate was incubated with 50 µl GSH agarose beads (GE Healthcare) for 3 hours at 4^{o}C, followed by 2 times washing with RIPA buffer to remove unspecific bound proteins. Beads were resuspended in SDS buffer, heated up to 95°C for 5 min and microcentrifuge for 1 min at 14,000 X g. The supernatant sample (30 µl) was loaded on SDS-PAGE gel (12–15%).

Proteomics. The silver stained bands were equally cut out into 8 pieces, and they were transferred to polypropylene tubes. The bands were destained by soaking the gel pieces in the destaining solution (SilverQuest staining kit, Invitrogen). Then the tubes were vigorously shaken on a micro-tube mixer (MixMate, Eppendorff) with adding 0.1 M ammonium bicarbonate containing DTT, and successively, 0.1 M ammonium bicarbonate containing acrylamide to reduce and alkylate the Cys containing peptides. Next, the buffer was discarded and the bands were extensively washed with methanol/water/acetic acid, 50:40:10, v/v/v for 30 min. After four

replacements of the washing solvent, the gel pieces were soaked with 400 μL of 100 mM ammonium bicarbonate solution for 5 min, then with 400 μL of acetonitrile for 5 min, and dried completely in a Speedvac evaporator (Thermo Savant, NY, USA). The dried gel pieces were re-swollen in 0.1 M ammonium bicarbonate containing trypsin and they were incubated at 37°C for overnight. After incubation, digested peptides were extracted twice with acetonitrile/water/trifluoroacetic acid (TFA) (75:25:0.1, v/v/v) by sonication for 15 min. The combined extracts were dried in a Speedvac. The samples were reconstituted with acetonitrile/water/trifluoroacetic acid (TFA) (97:3:0.1, v/v/v) and individually analyzed by LC-MS/MS in a Qexactive mass spectrometer coupled to an Easy nanoLC 1000 system (Thermo Scientific) using a 25 cm column (75um ID, C18 3um, column Technology Inc) as a separation column, and Symmetry C18 180 um ID x 20 mm trap column (Waters) over a 60 minutes gradient. Acquired mass spectrometry data were processed by Proteome Discover 1.4 (Thermo Scientific). The tandem mass spectra were searched against Uniprot human database using Sequest HT. A fixed modification of propionamide (+71.037114 Da) was added to Cys and a variable modification of oxidation (+15.994915 Da) was added to Met. The precursor mass tolerance was 10 ppm and the fragment mass tolerance 0.02 Da. The searched data was further processed with the Target Decoy PSM Validator function to filter with FDR 0.05.

Affymetrix HTA 2.0 array. Whole transcriptome alternative splicing pattern for HCT116 cells overexpressing *CCAT2*-G and T-allele was evaluated by Affymetrix HTA 2.0 array. Five hundred nanograms total RNA of each cell line (in duplicate) was provided to the Sequencing & Non-coding RNA Program, Center for Targeted Therapy at M.D. Anderson Cancer Center for performing the array on the Affymetrix platform. The Affymetrix Transcriptome Analysis Console (TAC) was used to analyze the SNP-related splicing event. A threshold of 1.5 (with a level of significance of at least $p < .01$) was set for the absolute value of the splicing index fold change. Integrated function and pathway analysis were performed using DAVID bioinformatics resources (http://david.abcc.ncifcrf.gov/), and significant features were clustered. The *p*-value and false discovery rate presented in the table are generated by a modified Fisher Exact test. Details on DAVID Functional Annotation Tool are given at: http://david.abcc.ncifcrf.gov/helps/functional_annotation.html#fisher".

Metabolomics. Sample preparation and metabolites extraction was performed as previously described with some minor modifications (Lorenz et al., 2011). Briefly, cells were grown up to 80% confluence. Medium was aspirated and rapidly rinsed by dispensing ~10 mL of 37 °C deionized water to the cell surface. Quenching was performed by directly adding ~15 mL of LN2 to the dish. For extraction, plates were transferred to a 4 °C cold room and 1.5 mL of ice cold 90% 9:1 MeOH: CHCl3 (MC) was added to each plate and cells were scraped with a cell lifter. Extracts were transferred to 1.5 mL microcentrifuge tubes and pelleted at 4 °C for 3 min at 16,100 g. Supernatants were transferred to tubes and dried by Speedvac. Metabolites were identified on a 1290 Infinity Series coupled to an Agilent 6550 iFunnel Q -TOF LC/MS (Agilent Technologies, Waldbronn, Germany) with dual electrospray ionization (ESI). Chromatographic separation was carried out on a C18 column (Agilent, Palo Alto, CA) (100 mm x 3.0 mm, 2.7 µm particle size) using 0.1% formic acid (A) and 0.1% formic acid in acetonitrile with (B) in a 22 min effective gradient (5-28% B) at a flow rate of 0.5 mL/min. Mass spectrometer was operated in negative mode, with a m/z dependent acquisition between 50 and 1100 m/z, and data was stored in centroid mode with an intensity threshold of 2000 counts. The operating conditions were as follows: gas temperature of 280 ºC, drying nitrogen gas of 9 L/min, nebulizer pressure of 45 psig, sheath gas temperature of 400 ºC, sheath gas flow of 12 L/min, capillary voltage of 3500 V, nozzle voltage of 500 V, fragmentor voltage of 100 V, skimmer of 65 V and octopole radiofrequency voltage of 750 V. TOF spectra acquisition rate/time was 1.5 spectra/s and 666.7 ms/spectrum respectively, and transients/spectrum were 5484. Internal calibration was performed using two lock masses (112.9855 and 1033.991) at a minimum height of 1000 counts. An initial RAW data processing by mzMine 2.10 software (Pluskal et al., 2010) allowed the definition of a normalized abundance matrix for each compound and sample. This table was used for the determination of the differential metabolites following the recommendations given by Valledor and Jorrín (2011) for the processing of omics datasets. Treatments were compared two by two by using t-test with a confidence interval of 99% and a 5% FDR. Graphical differences between whole samples were obtained after sPLS-DA analysis of the samples. All statistical analyses were performed in R 3.0.3 for Mac (R Core Team, 2014). Significantly different compounds were identified by mass homology by comparison to KEGG, MASSBANK, and METLIN databases, establishing a threshold of 10 ppm and a manual curation of the results.

Metabolites were clustered in different pathways using its KEGG id and the database ConsensusPathDB-mouse (Kamburov et al. 2011).

Northern Blot. Twenty µl elute from the histidine-tag pull-down assay was mixed with 30 µl RNA Sample Loading Buffer (Invitrogen) and 10 µl Blue/Orange Loading Dye (Promega), denatured for 10 min at 65^{o}C and ran on agarose gel. After complete running, the gel was washed and RNA was transferred overnight on BrightStar-Plus membrane (Ambion). Membrane was UV cross-linked and pre-hybridized with 10 ml Hybridyzation Buffer (ExpressHyb™) for 30 min at 65^{o}C. The specific probes for *CCAT2* and intron 14 were prepared with the Prime-it Ⅱ Random Primer Labeling Kit (Agilent) and labeled with α-^{32}P isotope. The membranes were hybridized overnight with each probe and after washing, the signal was assessed by autoradiography. Before re-probing the membrane was boiled for 15 min in 0.1% SDS solution and pre-hybridized with 10 ml Hybridyzation Buffer for 30 min at 65^{o}C. Primer sequences for generating the probes are available in **Table S4**.

Atomic Force Microscopy (AFM). Atomic Force Microscopy was conducted at the UT-Health Science Center AFM Core Facility using a BioScope II^{TM} Controller (Bruker Corporation; Santa Barbara, CA). The image acquisition was performed with the Research NanoScope software version 7.30 and analyzed with the NanoScope Analysis software version 1.40 (copyright 2013 Bruker Corporation). This system is integrated to a Nikon TE2000-E inverted optical microscope (Nikon Instruments Inc.; Lewisville, TX) to facilitate bright field/fluorescence imaging. High resolution images of RNA:protein complexes, RNAs and protein complex alone were obtained using RTESP cantilevers (*fo*=237-289 kHz, *k*=20-80 N/m, Bruker Corporation, Santa Barbara, CA). The nano-topography of the particles was determined using tapping mode operated in air to a scan rate of 0.5 - 0.7 Hz. Samples were prepared according to previously described methods (Lyubchenko et al., 2001; Lyubchenko et al., 2011). Briefly, freshly cleaved mica surface (highest grade V1 mica discs 12 mm, Ted Pella, Inc.) was treated with 10 µl of APTES (1µM in miliQ-water) for 5 min and rinsed with 2 ml of miliQ-water. A drop of 10 µl of the RNA alone, protein complex alone and RNA:complex suspension was incubated for 15 min on the functionalized mica (AP-mica), rinsed with 50 µl of miliQ-water and let it dry to be immediately scanned after preparation.

***In vivo* models and tissue processing.** Seventy male athymic nude mice were purchased from the National Cancer Institute, Frederick Cancer Research and Development Center (Frederick, MD) and were cared according to guidelines set forth by the American Association for Accreditation of Laboratory Animal Care and the U.S. Public Health Service policy on Human Care and Use of Laboratory Animals. All mouse studies were approved and supervised by the MD Anderson Cancer Center Institutional Animal Care and Use Committee. All animals were 8–12 weeks of age at the time of injection. For subcutaneous tumor models either HCT116-E or HCT116-G or HCT116-T cells were trypsinized, washed, and resuspended in Hanks' balanced salt solution (GIBCO, Carlsbad, CA) and injected subcutaneously into the dorsal side of the neck of mice (HCT116 $5x10^6$ cells/animal). For the HP-MRI analysis, fourteen days after tumor cell injection, mice were separated into two groups (n = 5 mice per group) and analyzed by Hyperpolarized [1-^{13}C]Pyruvate HP-MRI (described below). For the *in vivo* metabolomics analysis, fourteen days after injection mice were sacrificed and the tumors were collected (n = 3 mice per group). For lung metastasis model either KM12SM shGFP or KM12SM shGAC cells were injected into right-lateral tail vein and metastatic nodules were allowed to grow for 7-8 weeks. All mice in the experiment were then killed and necropsied, and their tumors were harvested. Lung weights, Tumor weights, numbers, and locations were recorded. Mean mouse body weights were similar among groups, suggesting that feeding and drinking habits were not affected. Tumor tissue was snap frozen. H&E staining were performed on frozen-embedded slides (**see below**).

H&E (Haematoxylin and Eosin) staining for frozen tissue sections. The frozen sections were air dried for several minutes to remove moisture and were stained with filtered 0.1% Mayers Hematoxylin (Sigma) for 10 min in a Coplin jar. Then they were rinsed in cool running ddH_2O for 5 min. Dipped in 0.5% Eosin 12 times. Dipped in distilled H_2O until the eosin stops streaking. Dipped in 50% $EtOH^-$ 10 times. Dipped in 70% $EtOH^-$ 10 times. Equilibrated in 95% $EtOH^-$ for 30 sec. Equilibrated in 100% $EtOH^-$ for 1 min and dipped in Xylene several times. The slides were cleaned off with a Kimwipes and mounted with universal mounting medium and coverslips.

Magnetic Resonance Imaging of Hyperpolarized [1-^{13}C]-Pyruvate. Hyperpolarized pyruvate was prepared as previously described (Ardenkjaer-Larsen et al., 2003; Sandulache et al., 2012) using a HyperSense dynamic nuclear polarizing system (Oxford Instruments). In brief, 26 mg of

[1-^{13}C]pyruvic acid containing 15 mM of the Ox63 polarizing radical (GE Healthcare) and 1.5mM ProHance (Bracco) was polarized at approximately 1.4K in a 3.35T magnetic field with 94.136 GHz microwave irradiation for up to an hour. The polarized substrate was dissolved in a heated base solution to yield an isotonic solution at body temperature (37°C) with pH of 7.6 containing 80 mM HP pyruvate. The hyperpolarized solution was immediately drawn into a syringe and 200 uL was administered into animals via tail-vein catheter.

All imaging was performed on a 7T Biospec small animal imaging system (Bruker Biospin Corp., Billerica, MA). Animals were anesthetized in isoflurane (0.5%-2% in oxygen) and placed head first and prone on an imaging sled. A custom ^{13}C surface coil with 20 mm outer diameter was placed immediately over the tumor. A dual-tuned ^{1}H/^{13}C volume resonator (72mm inner diameter) was used for anatomic imaging and for excitation of the ^{13}C signal. Animal placement was confirmed using a 3-plane gradient echo sequence (TE=3.6ms, TR=100ms) and the tumor was localized using axial and sagittal spin-echo images (TE_{eff}=16.7, TR=2500). Dynamic ^{13}C spectroscopy was acquired over for three minutes after dissolution using a slice-localized pulse-acquire sequence (TR=1.5s, 15° excitation) with a 12 mm axial slice encompassing the tumor. Spectra were phase adjusted, and the half-height full-width area of peaks associated with pyruvate and lactate were calculated for each point in time. Normalized lactate (nLac) was calculated as the total area under the dynamic lactate curve divided by the sum of the areas under lactate and pyruvate curves.

Statistical Analysis. Statistical analyses were performed in R (version 3.0.1) (http:///www.r-project.org/) or GraphPad Prism (version 5 or version 6). All tests were two-sided and considered statistical significant at the 0.05 level. The Spearman's rank-order correlation test was applied to measure the strength of the association between mRNA levels. The Shapiro-Wilk test was applied to determine whether data followed a normal distribution. The paired t-test was applied to normally distributed data otherwise the Wilcoxon rank test for paired data was applied to assess the association of mRNA levels with normal/tumor type. We downloaded and analyzed data publicly available from the Cancer Genome Atlas Project (TCGA; http://tcga-data.nci.nih.gov/) for Colon Adenocarcinoma (COAD). The mRNA expression of different isoforms was obtained from Level 3 Illumina RNASeqv2 “isoforms_normalized” files. The TCGA colon cancer population consists of 192 patients annotated for tumor stage, with overall survival information, and with RNA-Seq samples sequenced on GA platform. For the matched

normal-tumor pairs, RNA-Seq samples were sequenced on HiSeq platform. The putative copy number specific calls for MYC were obtained from the cBioPortal for Cancer Genomics (http://www.cbioportal.org/public-portal/). The nonparametric Kruskal-Wallis test followed by a Nemenyi post-hoc test, was applied to assess the relationship between NUDT21 expression and MYC amplification. Prediction of TFBS (transcription factor binding sites) was performed using TRANSFAC Match tool (accessed through Biobase, MDAnderson licence), which uses positional weight matrices (PWMs) from the TRANSFAC matrix library to search DNA sequences for potential TFBS. All the results presented in this manuscript represent the mean value of at least two biological replicates, each experiment was performed in triplicate or quadruplicate. Statistical significance is marked with the star symbol: '*' for $P < 0.05$, '**' for $P < 0.001$ and '***' for $P < 0.0001$.

Supplementary References:

Ardenkjaer-Larsen, J.H., Fridlund, B., Gram, A., Hansson, G., Hansson, L., Lerche, M.H., Servin, R., Thaning, M., and Golman, K. (2003). Increase in signal-to-noise ratio of > 10,000 times in liquid-state NMR. Proceedings of the National Academy of Sciences of the United States of America *100*, 10158-10163.

Cassago, A., Ferreira, A.P., Ferreira, I.M., Fornezari, C., Gomes, E.R., Greene, K.S., Pereira, H.M., Garratt, R.C., Dias, S.M., and Ambrosio, A.L. (2012). Mitochondrial localization and structure-based phosphate activation mechanism of Glutaminase C with implications for cancer metabolism. Proceedings of the National Academy of Sciences of the United States of America *109*, 1092-1097.

Coseno, M., Martin, G., Berger, C., Gilmartin, G., Keller, W., and Doublie, S. (2008). Crystal structure of the 25 kDa subunit of human cleavage factor Im. Nucleic Acids Research *36*, 3474-3483.

Gao, P., Tchernyshyov, I., Chang, T.C., Lee, Y.S., Kita, K., Ochi, T., Zeller, K.I., De Marzo, A.M., Van Eyk, J.E., Mendell, J.T., *et al.* (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature *458*, 762-765.

Haggerty, T.J., Zeller, K.I., Osthus, R.C., Wonsey, D.R., and Dang, C.V. (2003). A strategy for identifying transcription factor binding sites reveals two classes of genomic c-Myc target sites. Proceedings of the National Academy of Sciences of the United States of America *100*, 5313-5318.

Ling, H., Spizzo, R., Atlasi, Y., Nicoloso, M., Shimizu, M., Redis, R.S., Nishida, N., Gafa, R., Song, J., Guo, Z., *et al.* (2013). CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Research *23*, 1446-1461.

Lorenz, M.A., Burant, C.F., and Kennedy, R.T. (2011). Reducing time and increasing sensitivity in sample preparation for adherent mammalian cell metabolomics. Analytical Chemistry *83*, 3406-3414.

Lund, P. (1986). Methods Enzym. Anal. (Bergmeyer, H. U.: VCH Verlagsgesellschaft).

Lyubchenko, Y.L., Gall, A.A., and Shlyakhtenko, L.S. (2001). Atomic force microscopy of DNA and protein-DNA complexes using functionalized mica substrates. Methods Mol Biol *148*, 569-578.

Lyubchenko, Y.L., Shlyakhtenko, L.S., and Ando, T. (2011). Imaging of nucleic acids with atomic force microscopy. Methods *54*, 274-283.

Orengo, J.P., Bundman, D., and Cooper, T.A. (2006). A bichromatic fluorescent reporter for cell-based screens of alternative splicing. Nucleic Acids Research *34*, e148.

Pluskal, T., Castillo, S., Villar-Briones, A., and Oresic, M. (2010). MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics *11*, 395.

Sandulache, V.C., Skinner, H.D., Wang, Y., Chen, Y., Dodge, C.T., Ow, T.J., Bankson, J.A., Myers, J.N., and Lai, S.Y. (2012). Glycolytic inhibition alters anaplastic thyroid carcinoma

tumor metabolism and improves response to conventional chemotherapy and radiation. Molecular Cancer Therapeutics *11*, 1373-1380.

Yang, Q., Coseno, M., Gilmartin, G.M., and Doublie, S. (2011). Crystal structure of a human cleavage factor CFI(m)25/CFI(m)68/RNA complex provides an insight into poly(A) site recognition and RNA looping. Structure *19*, 368-377.

Yoon, J.H., Srikantan, S., and Gorospe, M. (2012). MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. Methods *58*, 81-87.