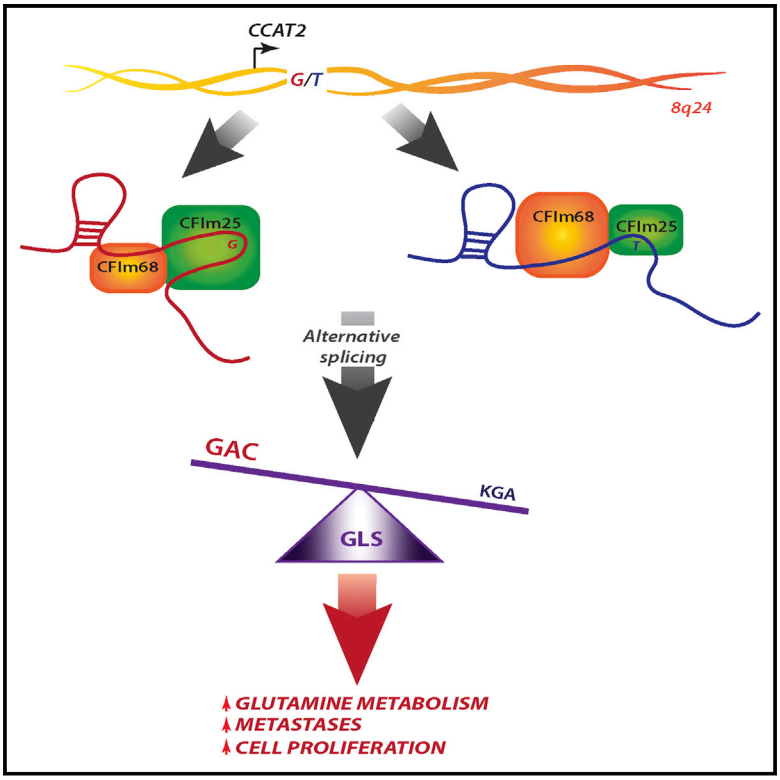
Article

Molecular Cell

Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2

Graphical Abstract



Highlights

- The lnRNA CCAT2 reprograms energy metabolism in an allele-specific manner
- CCAT2 alleles bind the CFIm complex with distinct affinities
- CCAT2-CFIm-GLS interaction regulates the alternative splicing of GLS
- This regulation axis is present in 61% of analyzed CRC cases

Authors

Roxana S. Redis, Luz E. Vela, Weiqin Lu, ..., Andre L.B. Ambrosio, Sandra M. Gomes Dias, George A. Calin

Correspondence

gcalin@mdanderson.org

In Brief

Redis et al. report that the two alleles of the lncRNA, CCAT2, induce distinct metabolic phenotypes. By interacting with the CFIm complex with allele-specific affinities, CCAT2 regulates the alternative splicing of GLS, resulting in the preferential expression of the more aggressive splice isoform.

Redis et al., 2016, Molecular Cell *61*, 1–15
February 18, 2016
http://dx.doi.org/10.1016/j.molcel.2016.01.015

CellPress

Please cite this article in press as: Redis et al., Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2, Molecular Cell (2016), http://dx.doi.org/10.1016/j.molcel.2016.01.015

Molecular Cell
Article

Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2

Roxana S. Redis,[1] Luz E. Vela,[2] Weiqin Lu,[3,4] Juliana Ferreira de Oliveira,[5] Cristina Ivan,[6] Cristian Rodriguez-Aguayo,[1] Douglas Adamoski,[5] Barbara Pasculli,[1] Ayumu Taguchi,[3] Yunyun Chen,[8] Agustin F. Fernandez,[9] Luis Valledor,[10] Katrien Van Roosbroeck,[1] Samuel Chang,[1] Maitri Shah,[1] Garrett Kinnebrew,[11] Leng Han,[12] Yaser Atlasi,[13,21] Lawrence H. Cheung,[1] Gilbert Y. Huang,[1] Paloma Monroig,[1] Marc S. Ramirez,[14] Tina Catela Ivkovic,[1,15] Long Van,[2] Hui Ling,[1] Roberta Gafà,[16] Sanja Kapitanovic,[15] Giovanni Lanza,[16] James A. Bankson,[14] Peng Huang,[3] Stephan Y. Lai,[8] Robert C. Bast,[1] Michael G. Rosenblum,[1] Milan Radovich,[11] Mircea Ivan,[11] Geoffrey Bartholomeusz,[1] Han Liang,[17] Mario F. Fraga,[18] William R. Widger,[2] Samir Hanash,[7] Ioana Berindan-Neagoe,[1,19,20] Gabriel Lopez-Berestein,[1,6] Andre L.B. Ambrosio,[5] Sandra M. Gomes Dias,[5] and George A. Calin[1,6,*]

[1]Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[2]Department of Biology & Biochemistry, University of Houston, Houston, TX 77204, USA
[3]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[4]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[5]Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas 13083–100, Brazil
[6]Center for RNA Interference and Non-coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[7]Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[8]Department of Head & Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[9]Cancer Epigenetics Laboratory, Institute of Oncology of Asturias (IUOPA), HUCA, Universidad de Oviedo, Oviedo 33006, Spain
[10]Department of Organisms and Systems Biology, University of Oviedo, Ovideo 33006, Spain
[11]Department of Surgery, Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN 46202, USA
[12]Department of Biochemistry and Molecular Biology, The University of Texas Health Science Center at Houston McGovern Medical School, Houston, TX 77030, USA
[13]Department of Pathology, Josephine Nefkens Institute, Erasmus Medical Center, Rotterdam 3015, the Netherlands
[14]Department of Imaging Physics, Division of Diagnostic Imaging, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[15]Laboratory for Personalized Medicine, Division of Molecular Medicine, Ruder Boskovic Institute, Zagreb 10000, Croatia
[16]Department of Morphology, Surgery and Experimental Medicine, University of Ferrara, Ferrara 44121, Italy
[17]Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[18]Nanomaterials and Nanotechnology Research Center (CINN-CSIC), Asturias 33424, Spain
[19]Research Center for Functional Genomics, Biomedicine and Translational Medicine, University of Medicine and Pharmacy Iuliu Hatieganu, Cluj-Napoca 400012, Romania
[20]Department of Functional Genomics, The Oncology Institute, Cluj-Napoca 400015, Romania
[21]Present address: Radboud Institute for Molecular Life Sciences (RIMLS) Department of Molecular Biology, Nijmegen 6525, the Netherlands
*Correspondence: gcalin@mdanderson.org
http://dx.doi.org/10.1016/j.molcel.2016.01.015

SUMMARY

Altered energy metabolism is a cancer hallmark as malignant cells tailor their metabolic pathways to meet their energy requirements. Glucose and glutamine are the major nutrients that fuel cellular metabolism, and the pathways utilizing these nutrients are often altered in cancer. Here, we show that the long ncRNA *CCAT2*, located at the 8q24 amplicon on cancer risk-associated rs6983267 SNP, regulates cancer metabolism in vitro and in vivo in an allele-specific manner by binding the Cleavage Factor I (CFIm) complex with distinct affinities for the two subunits (CFIm25 and CFIm68). The *CCAT2* interaction with the CFIm complex fine-tunes the alternative splicing of *Glutaminase* (*GLS*) by selecting the poly(A) site in intron 14 of the precursor mRNA. These findings uncover a complex, allele-specific regulatory mechanism of cancer metabolism orchestrated by the two alleles of a long ncRNA.

INTRODUCTION

Long non-coding RNAs (lncRNAs) form the largest part of the mammalian non-coding transcriptome (Mercer et al., 2009) and are generally expressed in a developmental-, tissue-, or disease-stage-specific manner, which makes them attractive therapeutic targets (Ling et al., 2013a). Although the underlying molecular mechanisms are not yet entirely understood, lncRNAs control gene expression at various levels including chromatin modification and transcriptional and post-transcriptional processing (Wilusz et al., 2009).

The revival of Warburg's theory of cancer (Warburg et al., 1924), complemented with novel discoveries in the field, has promoted cellular metabolism as an essential molecular mechanism for driving malignant transformation and progression (Boroughs

and DeBerardinis, 2015). Various studies have exposed the fine interplay between metabolic pathways orchestrated by protein-coding oncogenes and tumor suppressor genes (Chen and Russo, 2012) and, more recently, by ncRNAs (miRNAs and lncRNAs) (Gao et al., 2009; Yang et al., 2014). Glutamine, one of the essential nutrients, is deaminated by *Glutaminase* (*GLS*) to produce glutamate, which further serves as substrate for a variety of metabolic pathways (e.g., tricarboxylic cycle, TCA). Glutamine metabolism is modulated by MYC via miR-23a/b in prostate cancer and B cell lymphoma (Gao et al., 2009) and by NF-kB p65 subunit also through miR-23a downregulation in leukemic cells (Rathore et al., 2012; Wang et al., 2010b).

Colon Cancer-Associated Transcript 2 (*CCAT2*), a lncRNA that spans the highly conserved 8q24 region harboring the rs6983267 SNP (Ling et al., 2013b; Redis et al., 2013), is associated with increased risk for various types of cancer (Tomlinson et al., 2007; Tuupanen et al., 2009) and is specifically overexpressed in the microsatellite stable colorectal cancer (CRC MSS). The two alleles of the rs6983267 SNP present in the general population have been shown to render distinct risks of CRC; namely, the G allele was associated with greater predisposition to CRC than the T allele (Tomlinson et al., 2007). *CCAT2* induces chromosomal instability and metastases (Ling et al., 2013b) and regulates the expression levels of MYC oncogene, known to coordinate multiple molecular pathways supporting cell proliferation, metastases and cancer metabolism (Carroll et al., 2015; Stine et al., 2015). However, it is not clear how the two alleles are specifically involved in the malignant process. In this study, we demonstrate that the lncRNA, *CCAT2*, modulates cellular energy metabolism in an allele-specific manner by interacting with the Cleavage Factor I (CFIm) complex to regulate the alternative splicing of *GLS*.

RESULTS

CCAT2 Modulates Energy Metabolism in an Allele-Specific Manner In Vitro and In Vivo

We observed an unexpected change in the color of the media of in vitro grown cells when modulating the expression of *CCAT2* that suggested a possible shift in the energy metabolism consequent to *CCAT2* expression. We tested this hypothesis by measuring metabolic parameters in HCT116 colon cancer cells that stably overexpress *CCAT2* (OC1 and OC3) (Ling et al., 2013b) versus control cells and observed a significant and reproducible increase in glucose uptake, lactate secretion and oxygen consumption in the *CCAT2*-overexpressing clones (Figure 1A). These results were further confirmed in KM12SM cells with *CCAT2* downregulated expression (Figure 1B). Moreover, we explored whether these metabolic changes were occurring in vivo as well by injecting HCT116 *CCAT2*-overexpressing cells and control cells subcutaneously into nude mice and subjecting them to hyperpolarized magnetic resonance imaging (HP-MRI). We detected a significant increase in the flux of hyperpolarized [1-^{13}C]pyruvate to [1-^{13}C]lactate for the xenograft tumors derived from the *CCAT2*-overexpressing cells compared to the tumors derived from control cells (Figures S1A and S1B), consistent with our in vitro findings. These findings confirm that *CCAT2* alters metabolism, boosting glycolysis and cellular respiration. The coexistence of increased glycolysis with increased respiration, in highly proliferative cells, translates into enhanced anaplerotic reactions that replenish the TCA cycle intermediates (Ward and Thompson, 2012). Since glutamine is the main source for replenishing the intermediates of the TCA cycle, we measured the intra- and extracellular glutamate concentration, as well as the glutamine uptake in HCT116 cells with *CCAT2* overexpression and control cells. We found higher levels of both intra- and extracellular glutamate corresponding to higher levels of *CCAT2* (Figures S1C and S1D), suggesting *CCAT2* is boosting glutamine metabolism (glutaminolysis). Surprisingly, the glutamine uptake was not significantly different between the three clones (Figure S1D), implying the higher glutamate is not due to increased glutamine consumption. Therefore, we measured the enzymatic activity of GLS, the rate-limiting enzyme of glutaminolysis, in the whole lysate of the same cells and detected significantly higher activity in the cells with increased *CCAT2* expression (Figure S1F). In addition, both metabolic pathways (glycolysis and glutaminolysis) have been shown to be regulated by many factors, including the MYC oncogene (Carroll et al., 2015; Stine et al., 2015), a target of *CCAT2* by our previous report (Ling et al., 2013b).

We next explored whether the rs6983267 SNP influences these metabolic changes and assessed the glucose and glutamine uptake, oxygen consumption, lactate secretion, and intra- and extracellular glutamate concentration in HCT116 stably overexpressing either the G or the T allele of *CCAT2* and control HCT116 cells. Interestingly, we found on one hand higher glucose uptake and secreted glutamate in both G and T allele cells compared to control cells, while on the other hand we observed significant differences in lactate secretion, oxygen consumption, and intracellular glutamate production between the alleles (Figures 1C and S1E). Moreover, the glutamine consumption was not significantly different between the clones, similar to our previous results (Figure S1E). Consequently, we measured GLS enzymatic activity in these cells and observed that both *CCAT2* alleles induced a remarkable increase in activity compared to control, but the cells overexpressing the G allele displayed a significantly higher enzymatic activity compared to the T allele-overexpressing cells (Figure 1D). We also analyzed by mass spectroscopy the metabolites obtained from in vitro culturing of the HCT116 *CCAT2* G or T allele and control cells and from in vivo xenografted tumors derived from subcutaneous injection of the same cells. We observed contrasting distribution patterns when performing partial least squares discriminant analysis (PLS-DA) for both the in vitro (Figure 1E) and in vivo analyses (Figure 1F), and similarly for the principal component analysis (PCA) (Figures S1G and S1H). We detected 85.04% (in vitro) and 59.55% (in vivo) of metabolic pathways upregulated by *CCAT2* G allele compared to the T allele (Figure S1I, Table S1B). We then compared the pathway analysis for both datasets and identified 40 common pathways for the G allele and five common pathways for the T allele (Figure S1J). For these pathways, metabolic cluster distribution of differentially accumulated compounds revealed a significant overall enhancement of metabolic pathways related to glucose metabolism, TCA cycle, and glutamine metabolism for the G allele cells compared to the T allele cells (Tables S1A and S1B). We evaluated the expression

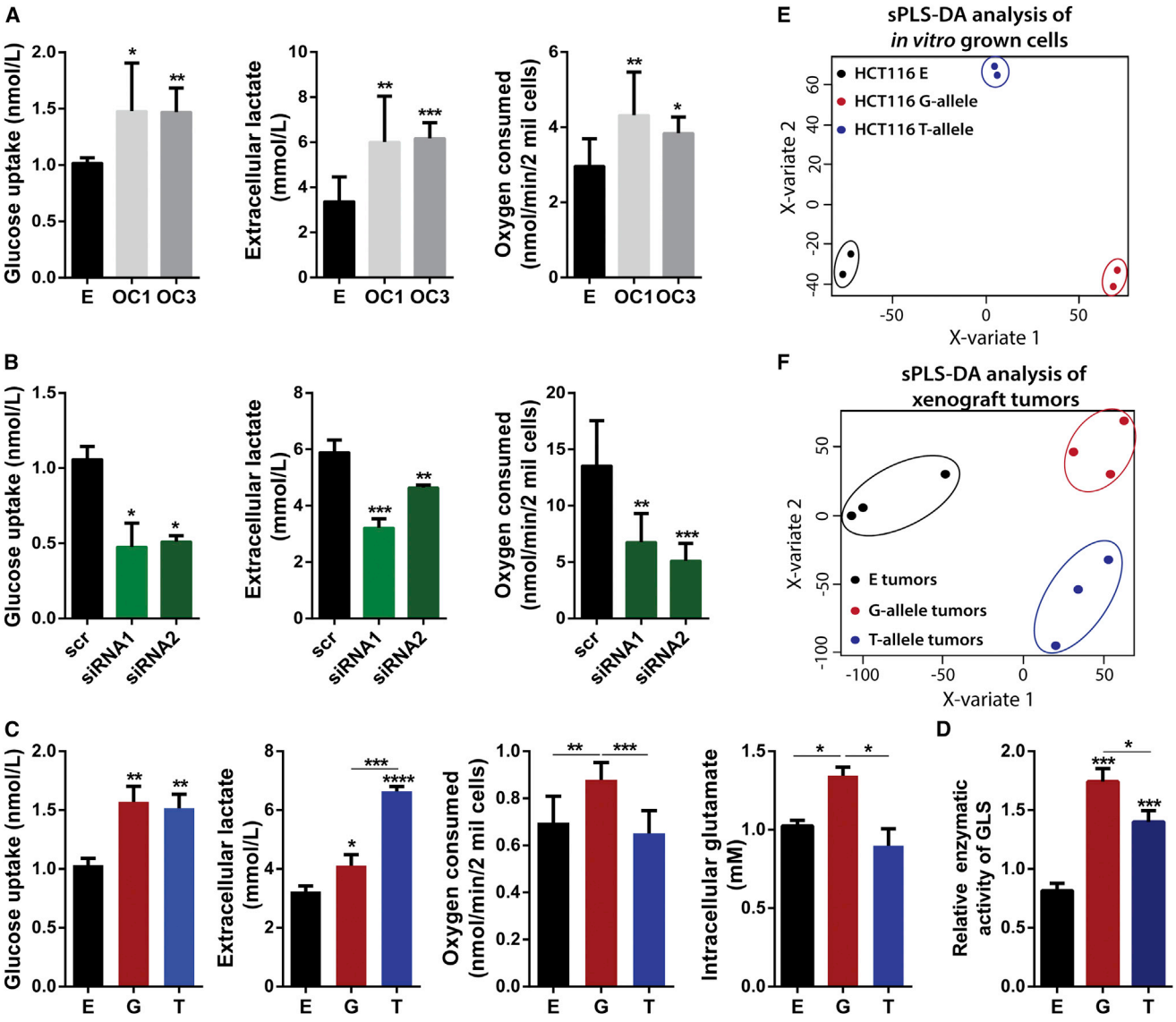


Figure 1. *CCAT2* Regulates Cancer Metabolism In Vitro and In Vivo

(A) Glucose uptake, lactate production, and oxygen consumption assays in HCT116 stable clones (E: empty control vector; OC1 and OC3: *CCAT2*-overexpressing, GG genotype).

(B) Glucose uptake, lactate production, and oxygen consumption assays in KM12SM cells with *CCAT2* downregulation (GT genotype).

(C) Whole-cell lysate *Glutaminase* activity measured in HCT116 *CCAT2*-overexpressiong G or T allele and control cells.

(D) Partial least-squares discriminant analysis (sPLS-DA) of HCT116 cells stably overexpressing *CCAT2* with either G or T alleles, and control cells (E, empty vector) in vitro allowed an adequate classification of the different cell lines according to its metabolome. sPLS-DA algorithm allows the classification of the samples based on the different abundances of each metabolite trying to find the maximum covariance between treatments and metabolome, in this way finding the most important metabolites for explaining the different effects of the treatments.

(E) Xenograft tumors derived from the same cell lines were also correctly classified by sPLS-DA analysis.

Results are presented as normalized mean values ± SD. See also Figure S1 and Table S1.

of MYC between the HCT116 *CCAT2* G and T allele cells but did not find impressive differences, consistent with our previous findings (Ling et al., 2013b) (Figure S2A). These results imply that in cells with high *CCAT2* expression, MYC activates energy metabolism in a general fashion; however, the fine-tuning of distinct metabolic pathways may occur through MYC-independent, but SNP-dependent, mechanisms. We therefore decided to direct our efforts toward exploring the *CCAT2* MYC-independent mechanism of regulation.

CCAT2 Regulates the Expression of GLS Isoforms

Our group has previously shown that *CCAT2* induces chromosomal instability, a process highly reliant on the supply of nucleotides (Bester et al., 2011) and intimately linked to glutamine

metabolism (Jeong et al., 2013). Further supported by the metabolic data, we directed our focus toward pathways metabolizing glutamine. We first assessed the protein expression of the two alternative splicing isoforms of *GLS*, KGA (glutaminase kidney isoform) and GAC (glutaminase isoform C) as a function of *CCAT2*. We used specific antibodies for each isoform (recognizing the distinct C termini) (Cassago et al., 2012) and/or a common antibody, recognizing the N terminus shared by the two isoforms, depending on the cell line. Although the two isoforms share the same active site, GAC has a higher catalytic activity than KGA and therefore may be more relevant for replenishing intermediates of the TCA cycle (Cassago et al., 2012; Le et al., 2012). While for GAC we observed an increase in protein expression in the HCT116 *CCAT2*-overexpressing cells with both antibodies, for KGA the protein expression presented an inconsistent variation with *CCAT2* upregulation (Figures 2A, S2B, and S2C). Analogously, the mRNA expression pattern for *GAC* and *KGA* reflected the protein expression (Figure 2B). Moreover, downregulation of *CCAT2* in KM12SM cells reduced GAC protein expression with 34% and 26%, respectively (Figures 2C, S2D, and S2E), whereas KGA protein expression was either unaltered or slightly increased (Figures 2C, S2D, and S2E). Similar results were obtained when measuring the mRNA expression of the two isoforms in the same cellular model (Figure S2F). Interestingly, when we evaluated the expression of the two isoforms in the HCT116 cells overexpressing the *CCAT2* G or T allele and control cells at both the mRNA and protein levels, we found unanimously higher expression of GAC when *CCAT2* G allele is upregulated (Figures 2D and S2G–S2I). However, changes in the *KGA* mRNA expression pattern did not concur with the protein expression (Figures 2D and S2G–S2I). The discordance observed for the KGA isoform is probably due to the regulation by the MYC-miR-23 axis (Gao et al., 2009; additional data available from authors upon request). These results alluded to the idea that *CCAT2* may preferentially induce the splicing of the GAC isoform. To determine this, we cloned the intron 14 of *GLS* precursor mRNA, known to encompass the alternative splicing site, in the RG6 bichromatic fluorescent reporter (Orengo et al., 2006). If the splicing machinery binds to the intron 14, it induces the splicing of a EGFP-tagged protein, which is the equivalent of GAC; otherwise, a dsRED-tagged protein will be produced, which is the equivalent of KGA (Figure S2J). We transfected the HCT116 control and *CCAT2*-overexpressing cells with the fluorescent reporter and determined the ratio between the expression levels of EGFP and dsRED fluorescence using the VECTRA automated imaging system. Although the EGFP-tagged protein, GAC-equivalent, was predominant in all models, we observed a 50% significantly higher EGFP/dsRED ratio in the *CCAT2*-overexpressing cells, corresponding to 50% more alternative splicing events (Figure 2E). Similar results were obtained when cells were analyzed by flow cytometry (Table S2). Moreover, when we compared the alternative splicing events occurring in the G and T alleles, using the RG6 reporter, we found significantly higher EGFP/dsRED ratio for the *CCAT2* G allele compared to the *CCAT2* T allele (Figure 2F and Table S2). Altogether, these data demonstrate that the *CCAT2* G allele is more efficient in boosting the alternative splicing of GAC isoform.

CFIm25 Governs the Switch between GLS Splicing Isoforms

To elucidate the underlying mechanism of *CCAT2*-induced regulation of the *GLS* alternative splicing, we introduced the MS2 tag (24 repeats) into vectors containing either the *CCAT2* G allele or T allele, pulled down the proteins that bind *CCAT2*, and analyzed them by mass spectrometry (Yoon et al., 2012). Pathway analysis on the QIAGEN platform identified Cleavage and Polyadenylation of Pre-mRNA among the top pathways associated with the G allele (Figure S3A), with CFIm25, the small (25 kDa) subunit of CFIm encoded by the *NUDT21* gene (Elkon et al., 2013; Yang et al., 2011), as the main protein correlated with the pathway. Notably, in the T allele pull down, both CFIm25 and the larger subunit (68 kDa) of the CFIm complex, CFIm68 (encoded by the *CPSF6* gene), were detected (CFIm68 was not detected in the G allele pull down); however, for CFIm25, the area under the peak had a 1.48 higher fold change in the G allele compared to the T allele (Figure S3B). We screened the *GLS* intron 14 for potential splicing and/or alternative polyadenylation (APA) sites using ASTRA (Alternative Splicing and TRanscription Archives database; http://dbarchive.biosciencedbc.jp/en/astra/desc.html) (Nagasaki et al., 2006) and identified a type 2 (skipping exon) poly(A) site (Figure S3C) (Lutz and Moreira, 2011) and multiple conserved binding motifs (UGUA) for CFIm25, consistent with a previous report mapping two poly(A) motifs within the same intron of the *GLS* pre-mRNA (Tian et al., 2007). Considering that CFIm25 and the CFIm heterotetramer complex have been previously linked to alternative splicing (Millevoi et al., 2006; Zhou et al., 2002), we first downregulated *NUDT21* and assessed the protein levels of GAC and KGA. We noticed a significant decrease in GAC protein expression in both cellular models and a clear increase in the KGA protein expression (Figure 3A). We also measured mRNA expression for the two isoforms and observed a significant reduction of the *GAC*/*KGA* mRNA ratio with *NUDT21* knockdown (Figure 3B). In order to evaluate whether the switch in isoform expression is a consequence of CFIm25 binding to the UGUA sequences within intron 14 of *GLS* pre-mRNA, we designed antisense synthetic oligonucleotides (ASOs) to block the binding of the 25 kDa subunit to these motifs (Figure 3C). Out of the four tested ASOs, we could identify two that were able to reverse the GAC to KGA protein expression ratio similar to the specific downregulation of *NUDT21* in HCT116 OC1 cells (Figure 3C). This result suggested that binding of CFIm25 to intron 14 is responsible for inducing the preferential expression of GAC. To further confirm the direct interaction of CFIm25 with *GLS* pre-mRNA and *CCAT2*, we immunoprecipitated the RNA bound to constituting proteins of the CFIm complex, CFIm25 and CFIm68, in HCT116 cells overexpressing either the G or T allele of *CCAT2* and control cells, and measured the differences in RNA enrichment between the distinct pull-down lysates by real-time quantitative PCR (qRT-PCR) (Figures 3D–3F and S3D–S3F). We included two lncRNAs as controls: *NEAT1* was chosen to be a positive control as it has been previously shown to interact with the CFIm complex (Naganuma et al., 2012), and *GAS5* was chosen as negative control due to minimal sequence similarity with *CCAT2* (Figure S3G). When assessing the fold enrichment of *GLS* and *CCAT2*, we detected, respectively, 5.77 and 13.6 times more RNA bound to CFIm25 in the

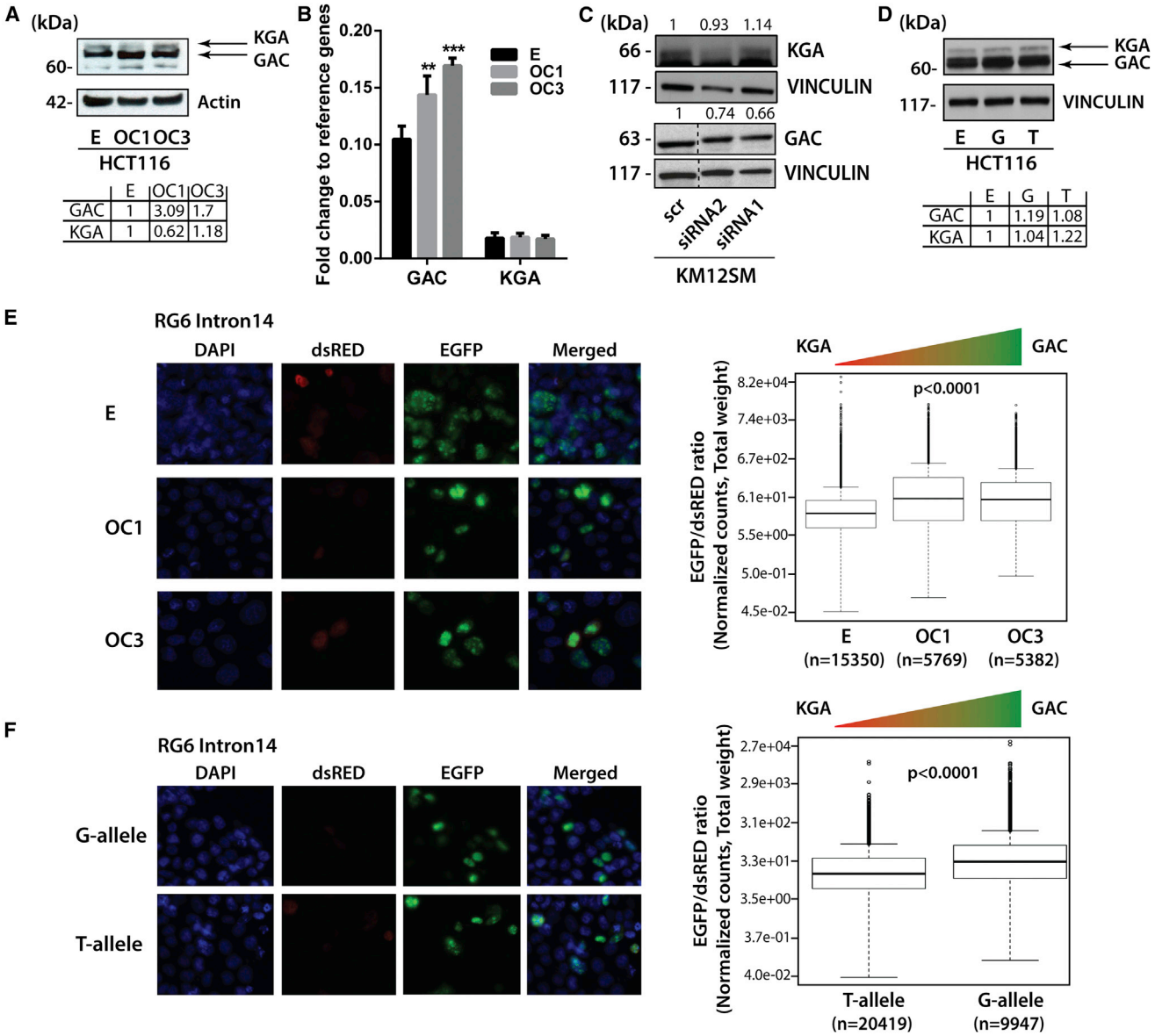


Figure 2. *CCAT2* Induces the Preferential Splicing of GAC

(A) Western blot analysis of GAC, KGA in HCT116 *CCAT2*-overexpressing cells (OC1 and OC3) and control cells.

(B) qRT-PCR assessing the mRNA expression of GAC and KGA in HCT116 *CCAT2*-overexpressing cells (OC1 and OC3) and control cells.

(C) Western blot analysis of GAC, KGA in KM12SM cells with *CCAT2* downregulation.

(D) Western blot analysis of GAC, KGA in HCT116 stably overexpressing *CCAT2* G or T alleles and control cells.

(E) Fluorescence microscopy images of HCT116 stable clones (E, empty control vector; OC1 and OC3, *CCAT2* overexpressing) transfected with the RG6 intron 14 vector and the analysis of the EGFP/dsRED ratio (n = number of analyzed cells).

(F) Fluorescence microscopy images of HCT116 *CCAT2* G allele and T allele transfected with the RG6 intron 14 vector and the analysis of the EGFP/dsRED ratio (n = number of analyzed cells).

Results are presented as normalized mean values ± SD. See also Figure S2 and Table S2.

cells overexpressing *CCAT2* G allele compared to control cells, while in the cells overexpressing *CCAT2* T allele compared to control cells, the fold enrichment ratios were only about half (2.95 and 6.02, respectively) (Figure 3D). Moreover, comparing the fold enrichment in the G-overexpressing cells to the T-overexpressing cells, we observed roughly double fold enrichment in the G-overexpressing cells for both *GLS* and *CCAT2* (1.95 and 2.26) (Figures 3E and 3F). The positive control (*NEAT1*) presented 6.3-fold more RNA bound to CFIm25 in the cells overexpressing *CCAT2* G allele compared to control cells and 2.79-fold more RNA enrichment in the cells overexpressing *CCAT2* T allele compared to control cells (Figures 3D and S3D). The negative

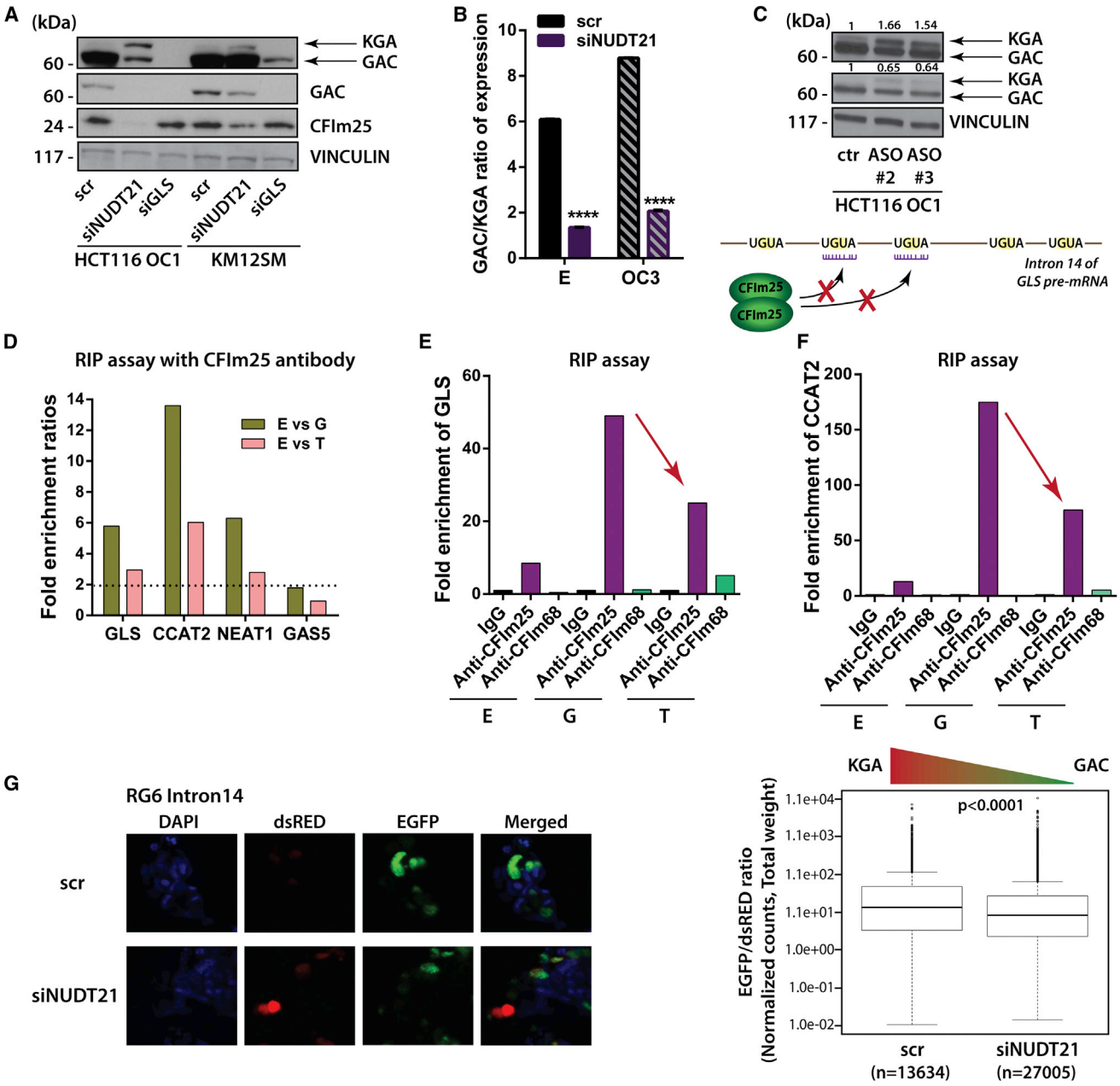


Figure 3. CFIm Protein Complex Binds *GLS* Pre-mRNA

(A) Western blot analysis of CFIm25, GAC, and KGA in HCT116 OC1 and KM12SM cells transiently transfected with siRNA for NUDT21, GLS (targeting the coding sequence shared by the two isoforms), and siRNA control.

(B) qRT-PCR assessing the GAC/KGA mRNA ratio in HCT116 *CCAT2*-overexpressing cells (OC1 and OC3) and control cells (E) with modulated CFIm25 expression.

(C) Western blot analysis of GAC and KGA in HCT116 OC1 cells with transient blockage of CFIm25 binding motifs (UGUA) by antisense oligonucleotides (ASOs). Schematic representation of the mechanism is presented below.

(D) qRT-PCR assessing the fold enrichment of *GLS*, *CCAT2*, *NEAT1*, and *GAS5* RNA bound to CFIm25 protein (RNA immunoprecipitation). Data are presented as fold enrichment ratios between control HCT116 cells (E) and *CCAT2*-overexpressing G or T allele.

(E and F) qRT-PCR assessing the fold enrichment of *GLS* mRNA (E) and *CCAT2* (F) bound to CFIm25 and CFIm68 in HCT116 cells, *CCAT2*-overexpressing G or T alleles, and control cells (E).

(G) Fluorescence microscopy images of KM12SM cells transfected with siNUDT21 and scr, followed by transfection with the RG6 intron 14 vector, and the analysis of the EGFP/dsRED ratio (n = number of analyzed cells).

Results are presented as normalized mean values ± SD. See also Figure S3 and Table S2.

control (*GAS5*) revealed only a 1.79-fold increase in the RNA bound to CFIm25 in the cells overexpressing *CCAT2* G allele compared to control cells and no difference in RNA enrichment between the overexpressing *CCAT2* T allele and control cells (Figures 3D and S3E). Thus, we concluded that, overall, in the cells overexpressing *CCAT2*, there is an augmentation of interaction between CFIm25 and *GLS* and *CCAT2*, with the highest levels in the G-overexpressing cells. The low binding of the RNAs to the CFIm68 was not surprising, as the main function of the protein is merely to enhance RNA binding and facilitate RNA looping, while the 25 kDa subunit has the leading role in binding the RNA via the UGUA elements (Yang et al., 2011) (Figures 3E, 3F, S3D, and S3E). Nonetheless, for our RNAs of interest, we observed an increase in enrichment in the T allele overexpressing cells compared to the G allele overexpressing and control cells, suggesting the secondary structure of *CCAT2* T allele may ease the interaction between the 68 kDa subunit and RNA molecules. In addition, we confirmed these results by end-point PCR in both the same cellular model and the KM12SM cell line, with a GT-heterozygous genotype (Figures S4A and S4B). As a further validation, we expressed the RG6 bichromatic fluorescent reporter with the intron 14 in KM12SM cells with knockdown of *NUDT21*. We assessed, by both VECTRA and FACS, the ratio of EGFP to dsRED and found a significantly lower (28%) EGFP/dsRED ratio in the cells with *NUDT21* downregulation, corresponding to fewer splicing events and consequently lower expression of the GAC equivalent (Figure 3G and Table S2).

Alternative Splicing of GLS Is Associated with the Interaction between CCAT2 and the CFIm Complex

We next aimed to investigate in more detail the mechanism leading to the difference in binding affinity of CFIm25 and CFIm68 to the distinct *CCAT2* alleles. We first scanned the *CCAT2* RNA sequence and identified two CFIm25 binding motifs surrounding the rs6983267, one upstream and the other downstream of the SNP. We performed secondary structure predictions using the RNAfold Webserver (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) and noticed major local structural changes induced near the putative upstream CFIm25 binding sequence, by the single nucleotide variation, especially among the G and T alleles (Figure S4C). Such changes may plausibly translate into distinct tertiary folds that could in principle explain the binding of the G and T alleles with different affinities. To further validate our results and evaluate the contribution of a single nucleotide to the structural changes and binding affinity, we mutated the SNP into an A or C (nucleotides never detected in the human population) and also deleted the 7 nucleotide region encompassing the SNP, which has been reported to have enhancer activity (Tuupanen et al., 2009). We repeated the RNA pull-down assay using different vectors containing the MS2 tags (*CCAT2*-G, -T, -A, -C, and DEL) and analyzed by western blot the proteins bound to RNAs. When CFIm25 antibody was hybridized on the blot, we detected a strong band of approximately 64 kDa unique to the G allele (band corresponds to the CFIm25 dimer; Figure 4A). When CFIm68 antibody was hybridized on the blot, we distinguished bands of approximately 68 kDa for both the T allele and the allele with the deleted region (Figure 4A). This implied that the T allele preferentially binds the 68 kDa subunit, but not in the region of the SNP. Notably, neither of the two mutated *CCAT2* alleles (A and C) interacted with the CFIm complex, and interestingly, both present secondary structures different from the G and T alleles (Figure S4C).

Additional evidence for the direct interaction between the *CCAT2* G and T alleles with the CFIm complex was provided by a His_6-tag pull-down assay using heterologously expressed CFIm68:CFIm25 complex (His_6-tagged CFIm68 subunit; Figure 4B) incubated with in vitro synthesized RNAs. We detected strong affinity of the CFIm complex for the *CCAT2* G allele, followed by moderate binding of the *CCAT2* T allele. We also identified in the pull down the 600 nt long region of the *GLS* pre-mRNA intron 14, containing one type 2 poly(A) site, with affinities for the protein complex corresponding to the two alleles (Figure 4B). This suggested that the intron 14 may also interact with *CCAT2*. To test this, we added a biotin tag to in vitro transcribed *CCAT2* RNAs, combined them with the intron 14 fragment and/or CFIm complex, and pulled down the complex with Streptavidin beads. We confirmed not only that the *CCAT2* G allele preferentially binds CFIm25, but also that *CCAT2* interacts with the intron 14 fragment in a SNP-independent fashion (Figures 4C and S4D). To ensure the specificity of the interaction, we repeated the biotin RNA pull-down assay to include the biotinylated *CCAT2* C allele as a negative control. We also added the whole intron 14, previously used for the RG6 splicing assay, to determine if *CCAT2* can interact with the entire region. We prepared mixes of *CCAT2* (G, T, and C) and the CFIm complex with and without the intron 14 fragment to evaluate how it impacts the interaction between *CCAT2* and CFIm complex. We discovered that in the presence of the intron 14, *CCAT2* G and T alleles displayed increased binding affinity to the CFIm complex, compared to the C allele (Figures 4D and S4E and Table S3). The G and T alleles presented a remarkable specificity for the CFIm25 dimer in the presence of the intron 14 (both the smaller fragment and the whole intron) with enhanced binding to the G allele (Figure S4E and Table S3), supporting the hypothesis that the secondary structure of *CCAT2* influences the interaction with the protein complex. In the absence of the intron, although the specificity of the interaction with the CFIm25 dimer is partly retained, it appeared to be revoked in the case of CFIm complex (Figures 4D and S4E and Table S3). We confirmed that *CCAT2* G and T alleles can bind the whole intron 14 as well and aligned the *CCAT2* genomic sequence with *GLS* genomic sequence to determine the extent of the interaction (Figure S4F). We observed multiple short fragments (13–18 nt) of sequence complementarity spanning the entire *GLS* sequence, present in both introns and exons (Figure S5A). For additional validation, we performed the His_6-tag pull-down assay using the G, T, and C alleles in the presence and absence of the whole intron 14. As expected, when the intron was included in the mix, we could detect only *CCAT2* G and T alleles in the pull-down lysates (Figure S4G).

Furthermore, the same mixes of *CCAT2* RNA (G/T), CFIm protein complex, and intron 14 RNA were prepared, as well as solutions of individual components, and subjected to atomic force microscopy (AFM) for visualizing the formation of the RNA:protein:RNA quaternary complex (Lyubchenko et al., 2011) (Figures

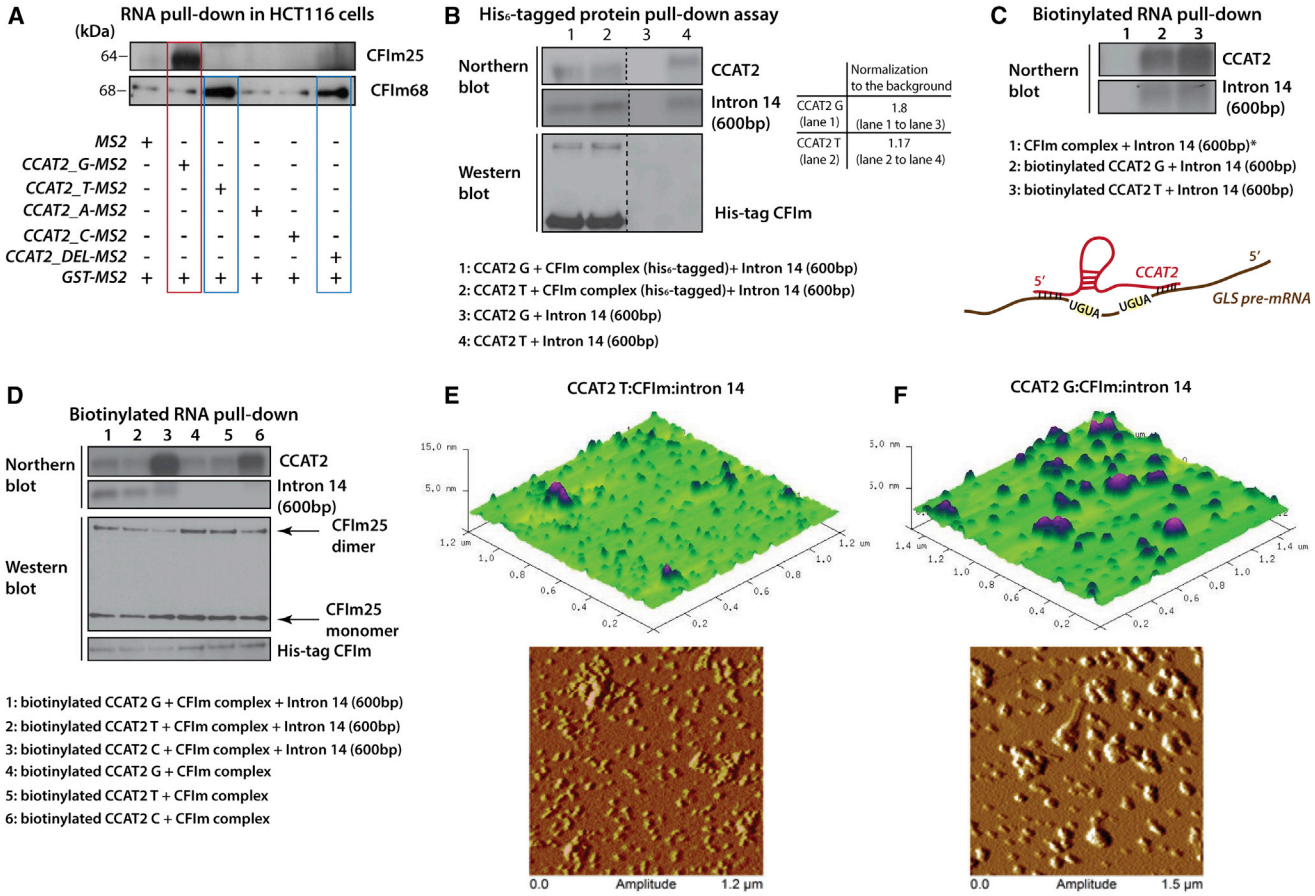


Figure 4. The rs6983267 SNP Affects the Interaction of *CCAT2* with the CFIm Protein Complex

(A) Western blot analysis of the proteins pulled down with the MS2-*CCAT2* vectors (G, T, A, C, and DEL) showing the presence of CFIm25 for the G allele and CFIm68 for the T allele.

(B) Northern blot analysis showing the presence of *CCAT2* and intron 14 (600 bp fragment) in the lysate pulled down with TALON resin (upper panel). Western blot analysis showing the presence of the His_6-tagged CFIm complex in the lysate pulled down with the TALON resin (lower panel).

(C) Northern blot analysis showing the presence of *CCAT2* and intron 14 (600 bp fragment) in the lysate pulled down with Streptavidin beads. Lane 1 marked with the star symbol is identical to lane 3 in Figure S4D. Schematic illustration of the interaction of *CCAT2* with *GLS* pre-mRNA (representation is not at scale).

(D) Northern blot analysis showing the presence of *CCAT2* and intron 14 (600 bp fragment) in the lysate pulled down with Streptavidin beads (upper panel). Western blot analysis showing the presence CFIm25, monomer (26 kDa) and dimer (64 kDa), and His_6-tagged CFIm68 (38 kDa) (lower panel).

(E and F) AFM images of CCAT2:CFIm:intron 14 quaternary complex including either *CCAT2* T allele (E) or *CCAT2* G allele (F).

See also Figures S4 and S5 and Table S3.

4E and 4F). Particles of substantial size were detected when scanning the mica surface, suggesting the formation of potential CCAT2:CFIm:intron14 complexes, with higher frequency for the G allele compared to the T allele (Figures 4E and 4F), while the individual components (CFIm protein complex, *CCAT2* RNA, and intron 14 RNA) appeared to be evidently smaller (Figures S5B–S5D). Moreover, when measuring the diameter of the complexes (G allele complex, 226.706 nm; T allele complex, 182.844 nm), we found that it corresponds approximately to the sum of the diameter of individual components (*CCAT2*, 81.865 nm; CFIm, 67.999 nm; intron 14, 48.360 nm).

Collectively, these findings suggested that: (1) *GLS* pre-mRNA impacts the interaction of *CCAT2* with CFIm complex; (2) *CCAT2* may be acting as a scaffold or assembly platform, promoting the selection of the poly(A) site in intron 14 of *GLS* pre-mRNA by directly binding both the pre-mRNA and the CFIm complex; and (3) the formation of the RNA:RNA:protein complex is dependent on the rs6983267 SNP and the secondary structure of *CCAT2*.

As an additional layer of regulation, we discovered that MYC is a transcription factor for *NUDT21* (data available from authors upon request).

GAC Promotes Metastases and Cell Proliferation

We proceeded with evaluating the contribution of GAC to aggressiveness of CRC. Considering that we have previously demonstrated that *CCAT2* promotes metastases (Ling et al., 2013b), we first assessed the in vitro migration potential of HCT116 *CCAT2*-overexpressing cells treated with either the allosteric inhibitor 968 or siRNA for *GLS* (Katt et al., 2012). We

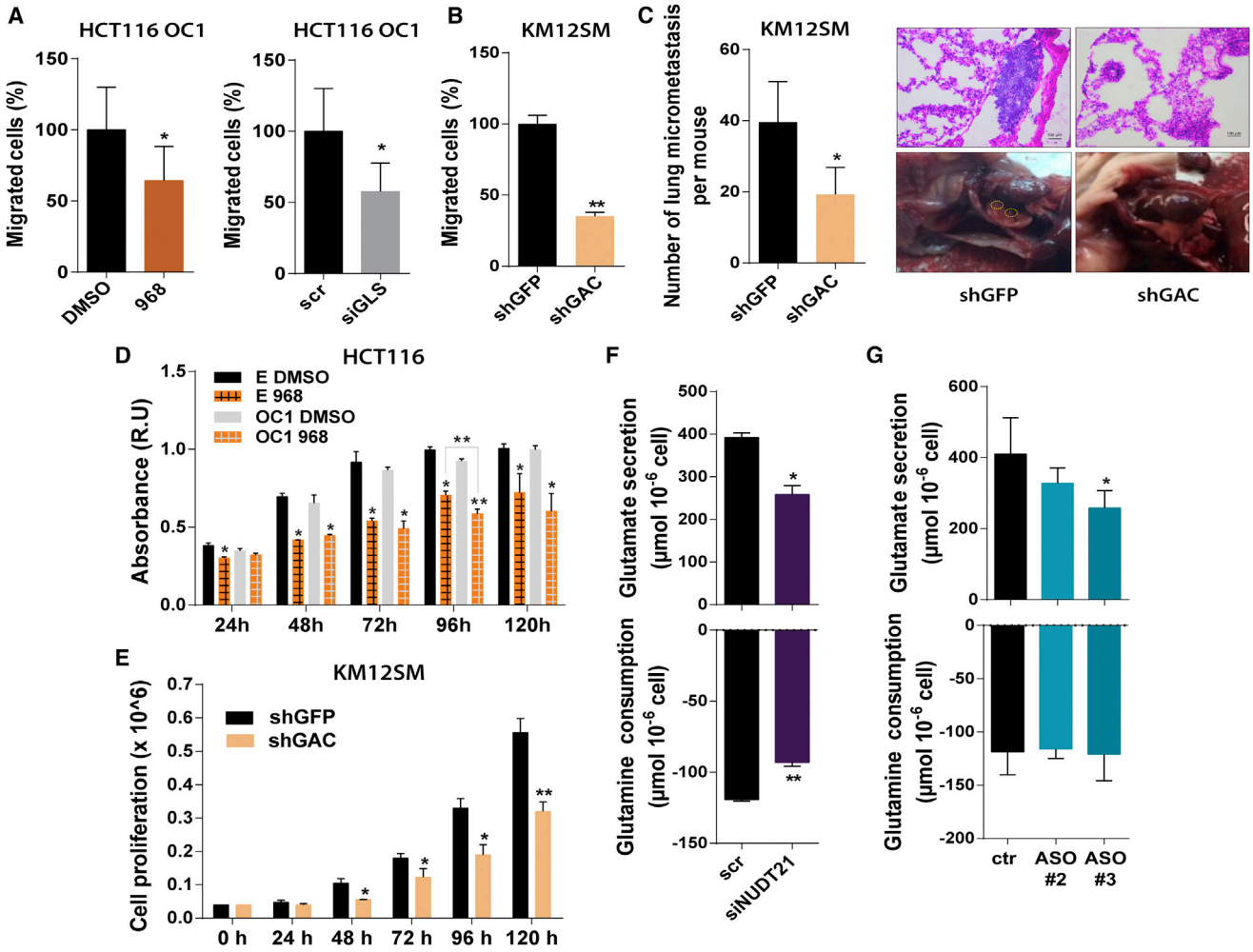


Figure 5. GLS Promotes In Vivo Metastases and In Vitro Cell Proliferation and Migration

(A) Migration of HCT116 OC1 cells (GG genotype) treated with the inhibitor 968 and DMSO (left panel) and with siGLS and scrambled siRNA (right panel).

(B) Migration of KM12SM cells (GT genotype) with stable downregulation of GAC. KM12SM shGFP cells represent the control cells.

(C) Number of the lung micrometastases for two groups (shGFP, 4 mice; shGAC, 4 mice) assessed by IHC (left panel). IHC images showing micrometastases in the three groups and images showing the presence or absence of lung metastases for mice injected in the tail vein with KM12SM shGFP and shGAC cells, respectively (right panel).

(D and E) Growth curves for HCT116 OC1 and control cells (GG genotype) treated with DMSO (control) or the GLS allosteric inhibitor 968 (10 μM) (D) and KM12SM cells (GT genotype) with stable downregulation of GAC (E).

(F and G) Glutamine and glutamate concentration in the media relative to the empty well 24 hr after seeding HCT116 *CCAT2*-overexpressing cells (OC1, GG genotype) transfected with siRNA against *NUDT21* and scrambled (F) and ASOs for inhibiting the binding sites of CFIm25 (G).

Results are presented as normalized mean values ± SD. See also Figure S6 and Table S4.

observed a reduction by half of the migration in the cells where *GLS* was either inhibited or downregulated (Figure 5A). We next assessed the individual contribution of GAC to the migration, in KM12SM cells with stable downregulation of the isoform (Figure S6A), and found that 60% fewer cells migrated when GAC expression was reduced (Figure 5B). We then injected the KM12SM stable clones in the tail vein of nude mice, sacrificed the mice 8 weeks after injections, and evaluated the in-lung macro- and micrometastases. Supporting our in vitro results, the incidence of metastases to the lung was 50% higher in the mice injected with the control cells (shGFP) compared to the shGAC group (Figure 5C). Moreover, when we assessed the proliferation of HCT116 control and *CCAT2*-overexpressing cells treated with the GLS inhibitor 968, we found that cells overexpressing *CCAT2* were more sensitive to GLS (GAC in this case) inhibition, implying that cells with high *CCAT2* expression are dependent on GAC for survival (Figure 5D) (Katt et al., 2012; Wang et al., 2010a). Similar results were obtained when using KM12SM cells with GAC downregulation compared to control cells in vitro (Figure 5E). Additional confirmation of the higher dependency of the *CCAT2* G allele on GAC was provided by the colony formation assay for NIH 3T3 cells transfected with the *CCAT2*-overexpressing vectors (Figures S6B and S6C).

We next sought to determine if the shift in *GLS* isoform expression and CFIm25 is responsible for the marked metabolic changes observed in cells overexpressing *CCAT2*. We therefore modulated the expression of the isoforms using the ASOs and downregulated *NUDT21* in HCT116 *CCAT2*-overexpressing cell line (OC1) and measured the extracellular lactate concentration and glutamine metabolism. We observed that downregulation of *NUDT21* significantly decreased lactate and glutamate secretion and glutamine consumption similarly to the effects of *CCAT2* overexpression in the same cell line (Figures S6D and 5F). However, these metabolic changes appeared to only partly mirror the metabolic effects of *CCAT2* overexpression, suggesting there are additional layers of regulations independent of *NUDT21*, most likely through the bona fide cancer metabolism modulator and *CCAT2* target, MYC. On the other hand, the switch in GLS isoforms reflected in lower secretion of glutamate with the glutamine consumption remaining relatively constant but had a modest effect on the extracellular levels of lactate (Figures 5G and S6E). These findings suggested that GLS is not the only contributor to *CCAT2*-induced metabolic profile. We hypothesized that other metabolic targets might be regulated by *CCAT2* via the same mechanism; thus, we performed Affymetrix HTA 2.0 array to compare the whole transcriptome splicing pattern in HCT116 *CCAT2*-overexpressing G and T allele cells. Pathway analysis revealed that several genes associated with two major metabolic pathways, ''Metabolism of Carbohydrates'' and ''Fructose and Mannose Metabolism,'' are significantly spliced between the *CCAT2* G and T alleles (Table S4). Moreover, validating our previous results, we identified a negative splicing index (SI = −1.19) for *GLS* for the exclusion junction connecting exon 14 and exon 15, suggesting higher signal for the exclusion junction in the T allele-expressing cells. This translates into less GAC being spliced in the T allele cells compared to the G allele cells, supportive of our findings (Figure S6F).

Collectively, these data demonstrate that GAC adds to the CRC phenotype; however, it is not solely responsible for the metabolic phenotype observed in cells overexpressing *CCAT2*.

CCAT2-CFIm-GLS Regulation Axis in CRC Tumors

We continued with evaluating the expression pattern of *CCAT2*, *GLS*, *NUDT21*, and *CPSF6* in colon tumors by analyzing the publicly available TCGA database of colon cancer (http://cancergenome.nih.gov/). We first compared 18 normal samples to 193 tumor samples and identified an enrichment of the *GAC* isoform, as well as *NUDT21* and *CPSF6* in tumor tissue, whereas *KGA* showed the opposite pattern (Figures 6A–6D). We further analyzed the associations between *CCAT2*, *NUDT21*, *CPSF6*, *GAC*, and *KGA* in the TCGA dataset of colon cancer samples (Cancer Genome Atlas Network, 2012) and detected direct correlations of *CCAT2*, *CPSF6*, and *NUDT21* with *GAC* ($r_s = 0.26$, $p = 0.0006$; $r_s = 0.68$, $p < 0.0001$; and $r_s = 0.72$, $p < 0.0001$, respectively; Figure S7A) and inverse correlations between *CCAT2*, *CPSF6*, *NUDT21*, and *KGA* ($r_s = -0.17$, $p = 0.0271$; $r_s = -0.47$, $p < 0.0001$; and $r_s = -0.60$, $p < 0.0001$; Figure S7A), as well as between *KGA* and *GAC* ($r_s = -0.590$, $p < 0.0001$) (Figure S7A). We also obtained significant direct associations of *CCAT2* with *CPSF6* and *NUDT21* ($r_s = 0.20$, $p = 0.0082$ and $r_s = 0.26$, $p = 0.0007$; Figure S7A). We did not find any significant correlations between the expression of *MYC* and *NUDT21*, *GAC*, or *KGA*. Since our in vitro findings advanced the concept that *CCAT2* G allele is promoting the expression of GAC, we compared the levels of *GAC* and *KGA* between patients having GG, GT, and TT genotypes. We observed a significant association with the genotypes for *GAC*, having the highest expression in patients with GG genotype, but not for *KGA* (Figures 6E and S7B). Similarly, we did not find any association of *NUDT21* or *MYC* with the genotypes (Figures S7C and S7D). These results suggested the molecular mechanism uncovered in vitro is present in CRC patients. Moreover, we analyzed the correlation between the expression of the two isoforms and the overall survival of patients from the TCGA dataset and observed a significant association of high *GAC* expression combined with low *KGA* expression with shorter overall survival (Figure S7E). This suggested that the *GAC* isoform may accelerate the progression of cancer. In addition, we interrogated the TCGA colon cancer dataset for genes that significantly correlate with the lncRNA and performed gene set enrichment analysis (GSEA) and ingenuity pathway analysis (IPA) (QIAGEN) to identify the *CCAT2* gene signatures. Various metabolically relevant pathways were found significantly associated with *CCAT2* expression by both analyses (Figures 7A and 7B). We further inquired as to whether the genes that were found significantly correlated with *CCAT2* expression were also associated with the genotypes of rs6983267 SNP. We repeated the GSEA and screened for pathways that correlated with either of the genotypes (FDR q value < 0.25 and nominal p value < 0.05). We identified various metabolic and RNA processing pathways that were associated with certain genotypes (Figure 7C). Interestingly, several of the pathways were related to the ones identified by the Affymetrix HTA 2.0 array (highlighted in green in Figure 7C), suggesting that *CCAT2* may modulate cellular metabolism in CRC patients through a network of metabolic genes regulated most likely via the same mechanism of alternative splicing (see Table S4).

To assess the protein expression of CFIm68, CFIm25, GAC, and KGA, we performed western blot analysis on paired normal mucosae and CRC samples (patient cohort #1) and identified the same high protein levels of CFIm25 and GAC in tumor tissue compared to normal mucosae for 61.5% (8/13) of the pairs (Figure 6F). As for the KGA isoform, protein levels were mostly lower in tumor samples or comparable to the levels in normal samples (Figure 6F). The 68 kDa subunit of the CFIm complex was either very low or not expressed in approximately half of the paired samples (7/13), while in the rest of the pairs it was clearly overexpressed in tumors (Figure S7F). We also measured *CCAT2* expression by qRT-PCR in the same set of samples and confirmed a positive correlation between *CCAT2*, CFIm25, and GAC for 69.2% of the samples (Figures 6F, 6G, S7G, and S7H). Moreover, the samples that had elevated levels of *CCAT2* displayed a high *GAC*/*KGA* ratio (*GAC*/*KGA* = 2–5.7) (Figure S7I). We repeated the measurements in a second set of paired normal mucosae and CRC samples (patient cohort #2) and obtained similar results (Figures S7J and S7K). Both CFIm25 and GAC proteins were overexpressed in 60% (3/5) of tumor samples. In all samples, elevated protein levels of CFIm25 and GAC matched the increased RNA levels of

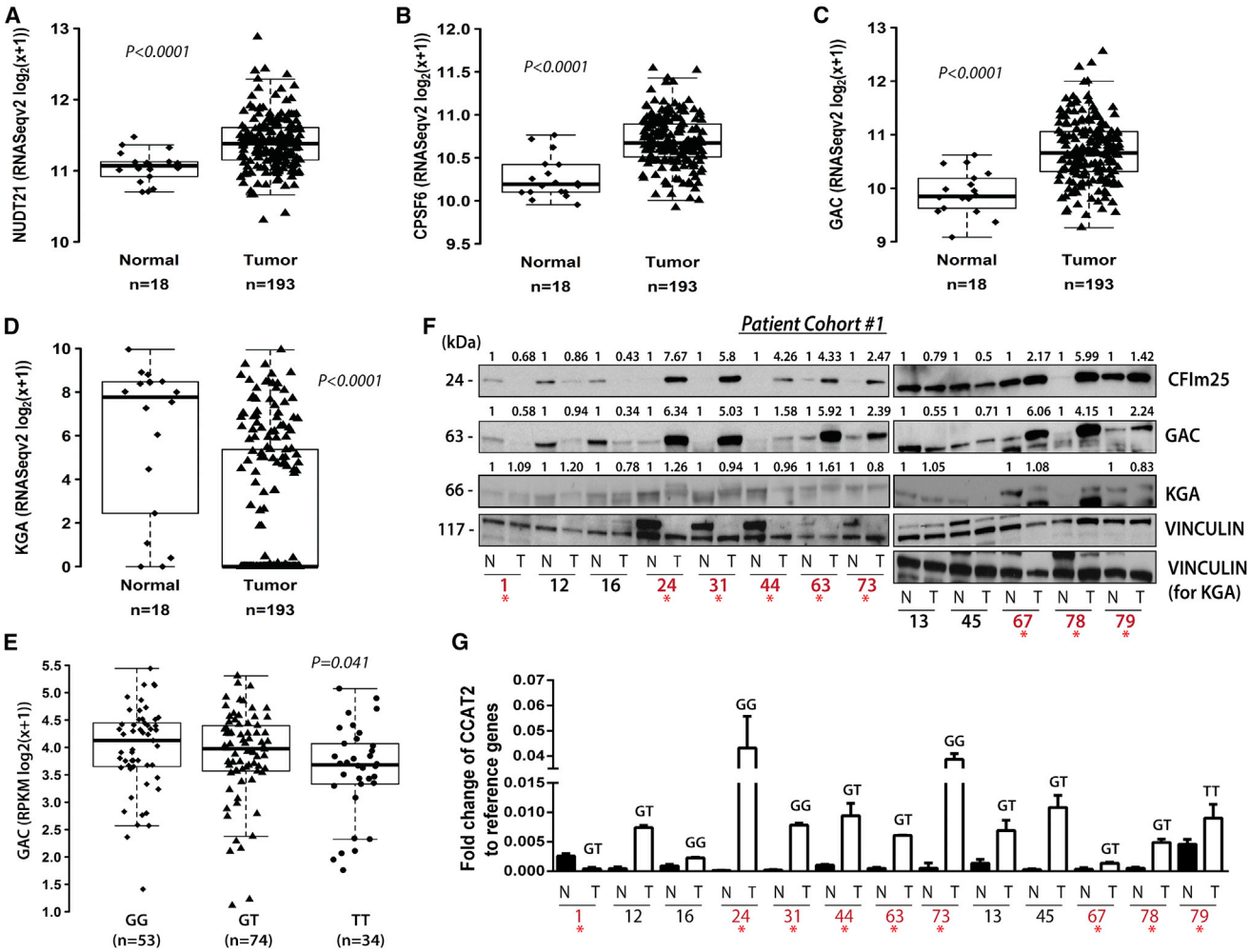


Figure 6. *CCAT2*, *NUDT21*, *CPSF6*, and *GLS* Expression Pattern in TCGA Dataset and CRC Patient Samples

(A–D) Analysis of *NUDT21* (A), *CPSF6* (B), *GAC* (C), and *KGA* (D) mRNA expression in TCGA RNA-seq colon cancer sample set.

(E) Association of GAC mRNA expression with the genotypes (GG, GT, and TT) of the rs6983267 SNP for CRC patients (TCGA RNA-seq dataset).

(F) Western blot analysis of CFIm25, GAC, and KGA expression in paired CRC samples (patient cohort #1).

(G) qRT-PCR analysis for *CCAT2* in the same paired CRC samples (patient cohort #1).

Results are presented as normalized mean values ± SD. See also Figure S7.

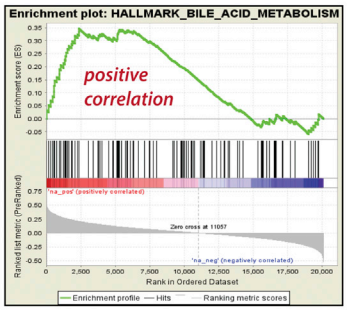
CCAT2. Additionally, we have genotyped the tumors from patient cohort #1 and found that the association of GG genotype with higher GAC and CFIm25 protein expression (samples marked in red) was consistent for 75% of patients (3/4), while for the GT genotype the association was present in only 62.5% of patients (5/8) (Figure 6G). No conclusion can be drawn for the TT genotype due to the limited representation of the genotype in this cohort (one patient). Altogether, this mechanism of *GLS* regulation was detected in the majority of the analyzed CRC cases (61%, 11/18).

DISCUSSION

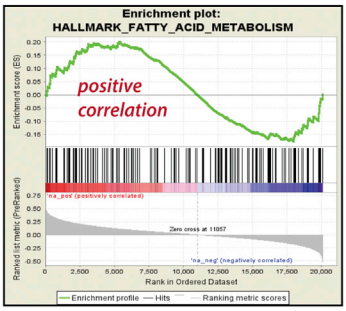
Our study demonstrates that the rs6983267 SNP (G/T) induces changes in the secondary structure of the lncRNA, *CCAT2*, initiating a domino effect mechanism, which leads to allele-specific reprograming of cellular energy metabolism. The consequence of the allele-specific interaction between *CCAT2*, CFIm, and *GLS* pre-mRNA appears to be the selection of the poly(A) site within intron 14 of *GLS*, resulting in the preferential splicing to the GAC isoform, the more catalytically active of the two GLS isoforms (Cassago et al., 2012). Although a recent study has described the negative regulation of GLS by CFIm25 in glioblastoma via 3′ UTR processing mechanisms (Masamha et al., 2014), suggesting a tumor-suppressive role for CFIm25, in our model, neither of the GLS isoforms is subjected to 3′ UTR shortening (data not shown). In the context of these findings and considering that *CCAT2* is not expressed in glioblastoma (data not shown), our results reveal an intriguing aspect of lncRNA mechanism of action, namely the ability of an lncRNA to alter the function of the partner RNA-binding protein/complex. The enrichment on metabolites related to TCA cycle that we

A

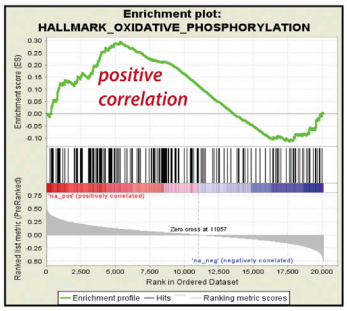
TCGA Colon Cancer dataset



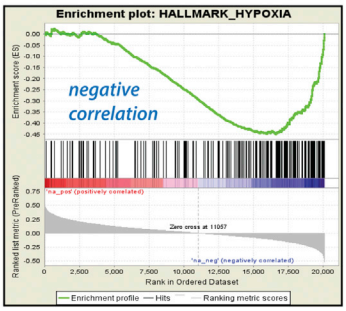
Gene	r^2 (Spearman)	p value
PDHA1	0.3715	<0.0001
ATP6V1C1	0.3679	<0.0001
VDAC1	0.1704	<0.0001
GOT2	0.1280	0.0132
SLC25A11	-0.3083	<0.0001
SDHA	-0.3351	<0.0001
ATP5A1	-0.3967	<0.0001



Gene	r^2 (Spearman)	p value
PKLR	0.4293	<0.0001
SLC37A4	0.3473	<0.0001
PDK3	0.2679	<0.0001
LDHC	0.2036	0.0001
BNIP3L	-0.3524	<0.0001
PFKFB3	-0.3479	<0.0001
GAPDH	-0.3027	<0.0001

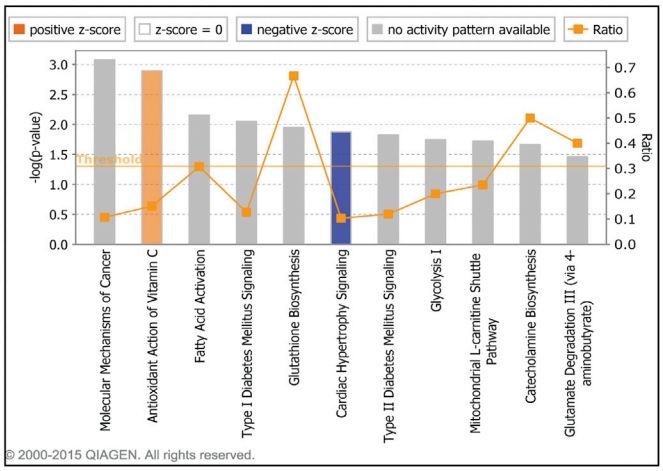


Gene	r^2 (Spearman)	p value
AMACR	0.5067	<0.0001
PIPOX	0.3740	<0.0001
SLC23A1	0.3194	<0.0001
GCLM	0.2489	<0.0001
ALDH1L1	-0.1932	0.0002
NPC1	-0.2539	<0.0001
HSD3B7	-0.3205	<0.0001



Gene	r^2 (Spearman)	p value
ACSL5	0.4174	<0.0001
PDHA1	0.3715	<0.0001
ACOT8	0.3711	<0.0001
SLC22A5	0.2388	<0.0001
ALDOA	-0.2783	<0.0001
CBR3	-0.3688	<0.0001
ENO2	-0.4549	<0.0001

B



C

Pathway	GG	GT	TT
Bile Acid Metabolism			
Bile Acid and Bile Salt Metabolism			
Oxidative Phosphorylation			
Cysteine and Methionine Metabolism			
Pyruvate Metabolism			
Glucose Transport			
Spliceosome			
mRNA 3' End Processing			
Fatty Acid Metabolism			
Hypoxia			
Metabolism of Carbohydrates			
Fructose and Mannose Metabolism			
Glycolysis - Gluconeogenesis			
Type I Diabetes Mellitus			
Type II Diabetes Mellitus			
Diabetes Pathways			
Dilated Cardiomyopathy			
Hypertrophic Cardiomyopathy			

Figure 7. *CCAT2* Gene Signature in Colon Cancer Patients—TCGA Dataset

(A and B) Genes associated with *CCAT2* were analyzed by gene set enrichment analysis (GSEA) (A) and ingenuity pathway analysis (IPA, QIAGEN) (B). Relevant examples for each analysis are presented in (A) and (B).

(C) Table containing the pathways significantly associated with *CCAT2* expression and rs6983267 genotype (FDR q value < 0.25 and nominal p value < 0.05). Pathways that were positively correlated are marked with red, and the ones that are negatively correlated are marked with blue. Highlighted in green are the pathways found common between GSEA analysis and the Affymetrix HTA 2.0 pathways analysis.

See also Figure S7 and Table S4.

observed for tumors derived from HCT116 cells overexpressing *CCAT2* the G allele are supported by Kaldma and colleagues' findings describing that CRC tumors are not purely glycolytic, but rather dependent on OXPHOS for ATP production (Kaldma et al., 2014).

The aberrant expression of *GLS* has been reported in many types of cancer, including CRC (Huang et al., 2014) Furthermore, various studies have described GAC as the more abundant isoform in lung adenocarcinoma, head and neck squamous cell carcinoma, kidney renal clear cell carcinoma, and AML (Jacque et al., 2015; van den Heuvel et al., 2012; Xia et al., 2014). Glutamine metabolism has also been associated with genomic instability (Jeong et al., 2013), commonly encountered in CRC and previously shown to be promoted by *CCAT2* (Ling et al., 2013b), and appears to be endorsed by the G risk allele. However, it must be stated that *CCAT2* is modulating energy metabolism in a general fashion via MYC and in an allele-specific manner via GLS and other metabolic enzymes and/or

metabolites, whose expression is finely regulated by the interaction of *CCAT2* with CFIm. Although the differences in regulation between the G and T alleles may not be impressive, the variation in expression of multiple enzymes/metabolites may have an additive effect towards a clear phenotypical change. The complex mechanism presented in this manuscript encompassing lncRNA, protein complexes, oncogenes, and transcription factors opens several windows for targeted therapy. The metabolic enzyme *GLS* is already considered a therapeutic target for cancer (Vander Heiden, 2011); however, our work introduces the opportunity of targeting the cancer-specific GAC isoform in particular.

Lastly, our study reveals the complexity and refinement of the interaction networks among the alleles of a non-coding RNA, the components of a protein complex, and the splicing isoforms of a metabolic enzyme that contribute to the malignant transformation and progression of CRC.

EXPERIMENTAL PROCEDURES

Full details of the Experimental Procedures are presented in the Supplemental Experimental Procedures. Primer sequences and information regarding the antibodies used in this study can be found in Tables S5 and S6.

Patient Samples

18 paired samples, normal colon mucosa and colon tumor, were used in this study. The samples were obtained from two different sources: The Ruder Boskovic Institute, Croatia (15 paired samples) and University of Ferrara, Italy (5 paired samples). Tissue samples were obtained from fresh surgical specimens frozen in liquid nitrogen and stored at −80°C. all the samples were obtained with the patients' informed consent and under the approval and supervision of the institutional review boards. The samples were histologically confirmed prior to use.

In Vivo Models and Tissue Processing

70 male athymic nude mice were purchased from the National Cancer Institute, Frederick Cancer Research and Development Center (Frederick, MD) and were cared for 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.

Glucose Uptake Assay

Cells were plated in 96-well plates (25,000 cells/well) 16 hr before performing the assay. The medium was removed and cells were washed twice with PBS. To the wells containing the blanks, 50 μl of PBS was added, while for the wells with the samples 50 μl of 2-NBDG (100 μM) (Sigma) was added and the mixture was incubated for 10 min at 37°C and 5% CO_2. After incubation, cells were washed twice with ice-cold PBS to stop the reaction, and 200 μl PBS was added to each well. Fluorescence measurements were performed at 485/520 nm with the PHERAstar *FS* (BMG Labtech).

Lactate Production Assay

To measure lactate production, cells that were 80% confluent were replenished with fresh medium. Aliquots of the medium were removed at the indicated time points (24 or 48 hr) for measurement of lactate using an Accutrend lactate analyzer (Roche). At each time point, cell numbers were also counted for normalization of lactate generation.

Intracellular Glutamate Assay

The glutamate concentration in cell lysates was measured using the Glutamate Colorimetric Assay Kit (Biovision) following the manufacturer's protocol. Briefly, 1 × 10^6 cells per tested sample were homogenized in 100 μl of assay buffer and centrifuged to remove insoluble material. 100 μl of reaction mix was added to the supernatant, standards, and background control samples, and after 30 min incubation at 37°C, absorbance was measured at 450 nm with a SpectraMax Plus384 MicroPlate Reader (Molecular Devices). The experiment was performed in quadruplicate.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and six tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.01.015.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: R.S.R., L.E.V., W.L., J.F.O., A.F.F., M.S.R., M.I., G.B., M.F.F., S.H., W.R.W., A.L.B.A., S.M.G.D., and G.A.C. Performed the experiments: R.S.R., L.E.V., W.L., J.F.O., C.R.-A., D.A., B.P., A.T., Y.C., K.V.R., S.C., M.S., Y.A., L.H.C., G.Y.H., P.M., M.S.R., T.C.I., L.V., and H.L. Analyzed the data: R.S.R., L.E.V., W.L., J.F.O., C.I., C.R.-A., D.A., A.T., A.F.F., L.V., G.K., J.A.B., M.R., A.L.B.A., S.M.G.D., and G.A.C. Contributed reagents/materials/analysis tools: L.H., S.K., R.G., G.L., S.Y.L., R.C.B., P.H., M.G.R., I.B.-N., and G.L.-B. Wrote the first draft of the manuscript: R.S.R., S.M.G.D., A.L.B.A., and G.A.C. Contributed to the writing of the manuscript: R.S.R., S.M.G.D., A.L.B.A., and G.A.C. Agree with manuscript results and conclusions: all authors. Critical revision of the manuscript for important intellectual content: all authors. Statistical analysis: C.I., L.H., G.K., M.R., and H.L. Study supervision: S.M.G.D. and G.A.C.

ACKNOWLEDGMENTS

G.A.C. is The Alan M. Gewirtz Leukemia & Lymphoma Society Scholar. Work in G.A.C.'s laboratory is supported in part by the NIH/NCI grants 1UH2TR00943-01 and 1 R01 CA182905-01, the UT MD Anderson Cancer Center SPORE in Melanoma grant from NCI (P50 CA093459), Aim at Melanoma Foundation and the Miriam and Jim Mulva research funds, the Brain SPORE (2P50CA127001), the Center for Radiation Oncology Research Project, the Center for Cancer Epigenetics Pilot project, a 2014 Knowledge GAP MDACC grant, a CLL Moonshot pilot project, the UT MD Anderson Cancer Center Duncan Family Institute for Cancer Prevention and Risk Assessment, a SINF grant in colon cancer, the Laura and John Arnold Foundation, the RGK Foundation, and the Estate of C.G. Johnson, Jr. I.B.-N. was financed by a grant entitled Non-Invasive Intelligent Systems for Colorectal Cancer Diagnosis and Prognosis Based on circulating miRNAs Integrated in the Clinical Workflow – INTELCOR. S.M.G.D., A.L.B.A., and D.A. are supported by the São Paulo Research Foundation FAPESP under grants 2014/15968-3, 2014/20673-2, and 2014/17820-3, respectively. W.L. was partly supported by grants from The University of Texas MD Anderson Cancer Center Sheikh Ahmed Bin Zayed Al Nahyan Center for Pancreatic Cancer Research. J.A.B. was supported by the Cancer Center Support Grant (P30 CA016672), and the HP imaging program of the Small Animal Facility (SAIF) was supported by the Cancer Prevention and Research Institutes of Texas grant RP-101243P5. H.L. was supported by NIH/NCI grant R01CA175486, a grant (RP140462) from the Cancer Prevention and Research Institute of Texas, and the R. Lee Clark Fellow Award from The Jeanne F. Shelby Scholarship Fund. We would like to thank Dr. Riccardo Fodde (Erasmus Medical Center) for the scientific support and advice, Dr. Sylvie Doublie for the generous gift of CFIm25 and CFIm68 plasmids, Dr. Riccardo Spizzo for generating the HCT116 *CCAT2* stable clones, and Dr. Thomas A. Cooper for the generous gift of RG6 plasmid. We would also like to thank the IM Bioscope II-UT core facility and Dr. Ana Maria Zaske for performing the AFM imaging. Additionally, we thank the members of Flow Cytometry & Cellular Imaging Core Facility (Department of Leukemia) and Dr. Jared Burks for performing the VECTRA imaging and analysis and Drs. Liuqing Yang (MDACC), Yibin Zhou (IBT), Clifford Stephan (IBT), Shawn Bratton (MDACC), Charles V. Kingsley (MDACC), Jorge Delacerda (MDACC), Iva Maestri (University of Ferrara), and Linda Ulazzi (University of Ferrara) for the technical

support. We also would like to thank the Biological Imaging Facility and the Protein Purification and Bioassay Laboratories (LNBio/CNPEM/Brazil) for access to the equipment.

Received: April 3, 2015
Revised: October 23, 2015
Accepted: January 8, 2016
Published: February 4, 2016

REFERENCES

Bester, A.C., Roniger, M., Oren, Y.S., Im, M.M., Sarni, D., Chaoat, M., Bensimon, A., Zamir, G., Shewach, D.S., and Kerem, B. (2011). Nucleotide deficiency promotes genomic instability in early stages of cancer development. Cell *145*, 435–446.

Boroughs, L.K., and DeBerardinis, R.J. (2015). Metabolic pathways promoting cancer cell survival and growth. Nat. Cell Biol. *17*, 351–359.

Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. Nature *487*, 330–337.

Carroll, P.A., Diolaiti, D., McFerrin, L., Gu, H., Djukovic, D., Du, J., Cheng, P.F., Anderson, S., Ulrich, M., Hurley, J.B., et al. (2015). Deregulated Myc requires MondoA/Mlx for metabolic reprogramming and tumorigenesis. Cancer Cell *27*, 271–285.

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. Proc. Natl. Acad. Sci. USA *109*, 1092–1097.

Chen, J.Q., and Russo, J. (2012). Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochim. Biophys. Acta *1826*, 370–384.

Elkon, R., Ugalde, A.P., and Agami, R. (2013). Alternative cleavage and polyadenylation: extent, regulation and function. Nat. Rev. Genet. *14*, 496–506.

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., and Dang, C.V. (2009). c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature *458*, 762–765.

Huang, F., Zhang, Q., Ma, H., Lv, Q., and Zhang, T. (2014). Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. Int. J. Clin. Exp. Pathol. *7*, 1093–1100.

Jacque, N., Ronchetti, A.M., Larrue, C., Meunier, G., Birsen, R., Willems, L., Saland, E., Decroocq, J., Thiago, T.T., Lambert, M., et al. (2015). Targeting glutaminolysis has antileukemic activity in acute myeloid leukemia and synergizes with BCL-2 inhibition. Blood *126*, 1346–1356.

Jeong, S.M., Xiao, C., Finley, L.W., Lahusen, T., Souza, A.L., Pierce, K., Li, Y.H., Wang, X., Laurent, G., German, N.J., et al. (2013). SIRT4 has tumor-suppressive activity and regulates the cellular metabolic response to DNA damage by inhibiting mitochondrial glutamine metabolism. Cancer Cell *23*, 450–463.

Kaldma, A., Klepinin, A., Chekulayev, V., Mado, K., Shevchuk, I., Timohhina, N., Tepp, K., Kandashvili, M., Varikmaa, M., Koit, A., et al. (2014). An in situ study of bioenergetic properties of human colorectal cancer: the regulation of mitochondrial respiration and distribution of flux control among the components of ATP synthasome. Int. J. Biochem. Cell Biol. *55*, 171–186.

Katt, W.P., Ramachandran, S., Erickson, J.W., and Cerione, R.A. (2012). Dibenzophenanthridines as inhibitors of glutaminase C and cancer cell proliferation. Mol. Cancer Ther. *11*, 1269–1278.

Le, A., Lane, A.N., Hamaker, M., Bose, S., Gouw, A., Barbi, J., Tsukamoto, T., Rojas, C.J., Slusher, B.S., Zhang, H., et al. (2012). Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. Cell Metab. *15*, 110–121.

Ling, H., Fabbri, M., and Calin, G.A. (2013a). MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nat. Rev. Drug Discov. *12*, 847–865.

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

Lutz, C.S., and Moreira, A. (2011). Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip. Rev. RNA *2*, 23–31.

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

Masamha, C.P., Xia, Z., Yang, J., Albrecht, T.R., Li, M., Shyu, A.B., Li, W., and Wagner, E.J. (2014). CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature *510*, 412–416.

Mercer, T.R., Dinger, M.E., and Mattick, J.S. (2009). Long non-coding RNAs: insights into functions. Nat. Rev. Genet. *10*, 155–159.

Millevoi, S., Loulergue, C., Dettwiler, S., Karaa, S.Z., Keller, W., Antoniou, M., and Vagner, S. (2006). An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. EMBO J. *25*, 4854–4864.

Naganuma, T., Nakagawa, S., Tanigawa, A., Sasaki, Y.F., Goshima, N., and Hirose, T. (2012). Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. EMBO J. *31*, 4020–4034.

Nagasaki, H., Arita, M., Nishizawa, T., Suwa, M., and Gotoh, O. (2006). Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns. Bioinformatics *22*, 1211–1216.

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

Rathore, M.G., Saumet, A., Rossi, J.F., de Bettignies, C., Tempé, D., Lecellier, C.H., and Villalba, M. (2012). The NF-κB member p65 controls glutamine metabolism through miR-23a. Int. J. Biochem. Cell Biol. *44*, 1448–1456.

Redis, R.S., Sieuwerts, A.M., Look, M.P., Tudoran, O., Ivan, C., Spizzo, R., Zhang, X., de Weerd, V., Shimizu, M., Ling, H., et al. (2013). CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. Oncotarget *4*, 1748–1762.

Stine, Z.E., Walton, Z.E., Altman, B.J., Hsieh, A.L., and Dang, C.V. (2015). MYC, Metabolism, and Cancer. Cancer Discov. *5*, 1024–1039.

Tian, B., Pan, Z., and Lee, J.Y. (2007). Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. *17*, 156–165.

Tomlinson, I., Webb, E., Carvajal-Carmona, L., Broderick, P., Kemp, Z., Spain, S., Penegar, S., Chandler, I., Gorman, M., Wood, W., et al.; CORGI Consortium (2007). A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nat. Genet. *39*, 984–988.

Tuupanen, S., Turunen, M., Lehtonen, R., Hallikas, O., Vanharanta, S., Kivioja, T., Björklund, M., Wei, G., Yan, J., Niittymäki, I., et al. (2009). The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat. Genet. *41*, 885–890.

van den Heuvel, A.P., Jing, J., Wooster, R.F., and Bachman, K.E. (2012). Analysis of glutamine dependency in non-small cell lung cancer: GLS1 splice variant GAC is essential for cancer cell growth. Cancer Biol. Ther. *13*, 1185–1194.

Vander Heiden, M.G. (2011). Targeting cancer metabolism: a therapeutic window opens. Nat. Rev. Drug Discov. *10*, 671–684.

Wang, J., Liu, X., Wu, H., Ni, P., Gu, Z., Qiao, Y., Chen, N., Sun, F., and Fan, Q. (2010a). CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. *38*, 5366–5383.

Wang, J.B., Erickson, J.W., Fuji, R., Ramachandran, S., Gao, P., Dinavahi, R., Wilson, K.F., Ambrosio, A.L., Dias, S.M., Dang, C.V., and Cerione, R.A. (2010b). Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. Cancer Cell *18*, 207–219.

Warburg, O., Posener, K., and Negelein, E. (1924). Über den Stoffwechsel der Carcinomzelle. Biochem. Zeitschr *152*, 309–344.

Ward, P.S., and Thompson, C.B. (2012). Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. Cancer Cell *21*, 297–308.

Wilusz, J.E., Sunwoo, H., and Spector, D.L. (2009). Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. *23*, 1494–1504.

Xia, Z., Donehower, L.A., Cooper, T.A., Neilson, J.R., Wheeler, D.A., Wagner, E.J., and Li, W. (2014). Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nat. Commun. *5*, 5274.

Yang, Q., Gilmartin, G.M., and Doublié, S. (2011). The structure of human cleavage factor I(m) hints at functions beyond UGUA-specific RNA binding: a role in alternative polyadenylation and a potential link to 5′ capping and splicing. RNA Biol. *8*, 748–753.

Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1α and lincRNA-p21 modulates the Warburg effect. Mol. Cell *53*, 88–100.

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.

Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. Nature *419*, 182–185.

Molecular Cell, Volume *61*

Supplemental Information

Allele-Specific Reprogramming of Cancer Metabolism by the Long Non-coding RNA CCAT2

Roxana S. Redis, Luz E. Vela, Weiqin Lu, Juliana Ferreira de Oliveira, Cristina Ivan, Cristian Rodriguez-Aguayo, Douglas Adamoski, Barbara Pasculli, Ayumu Taguchi, Yunyun Chen, Agustin F. Fernandez, Luis Valledor, Katrien Van Roosbroeck, Samuel Chang, Maitri Shah, Garrett Kinnebrew, Leng Han, Yaser Atlasi, Lawrence H. Cheung, Gilbert Y. Huang, Paloma Monroig, Marc S. Ramirez, Tina Catela Ivkovic, Long Van, Hui Ling, Roberta Gafà, Sanja Kapitanovic, Giovanni Lanza, James A. Bankson, Peng Huang, Stephan Y. Lai, Robert C. Bast, Michael G. Rosenblum, Milan Radovich, Mircea Ivan, Geoffrey Bartholomeusz, Han Liang, Mario F. Fraga, William R. Widger, Samir Hanash, Ioana Berindan-Neagoe, Gabriel Lopez-Berestein, Andre L.B. Ambrosio, Sandra M. Gomes Dias, and George A. Calin

Supplementary files

Supplementary figures S1-S7

Supplementary tables S1-S6

Supplementary Experimental Procedures

Supplementary references

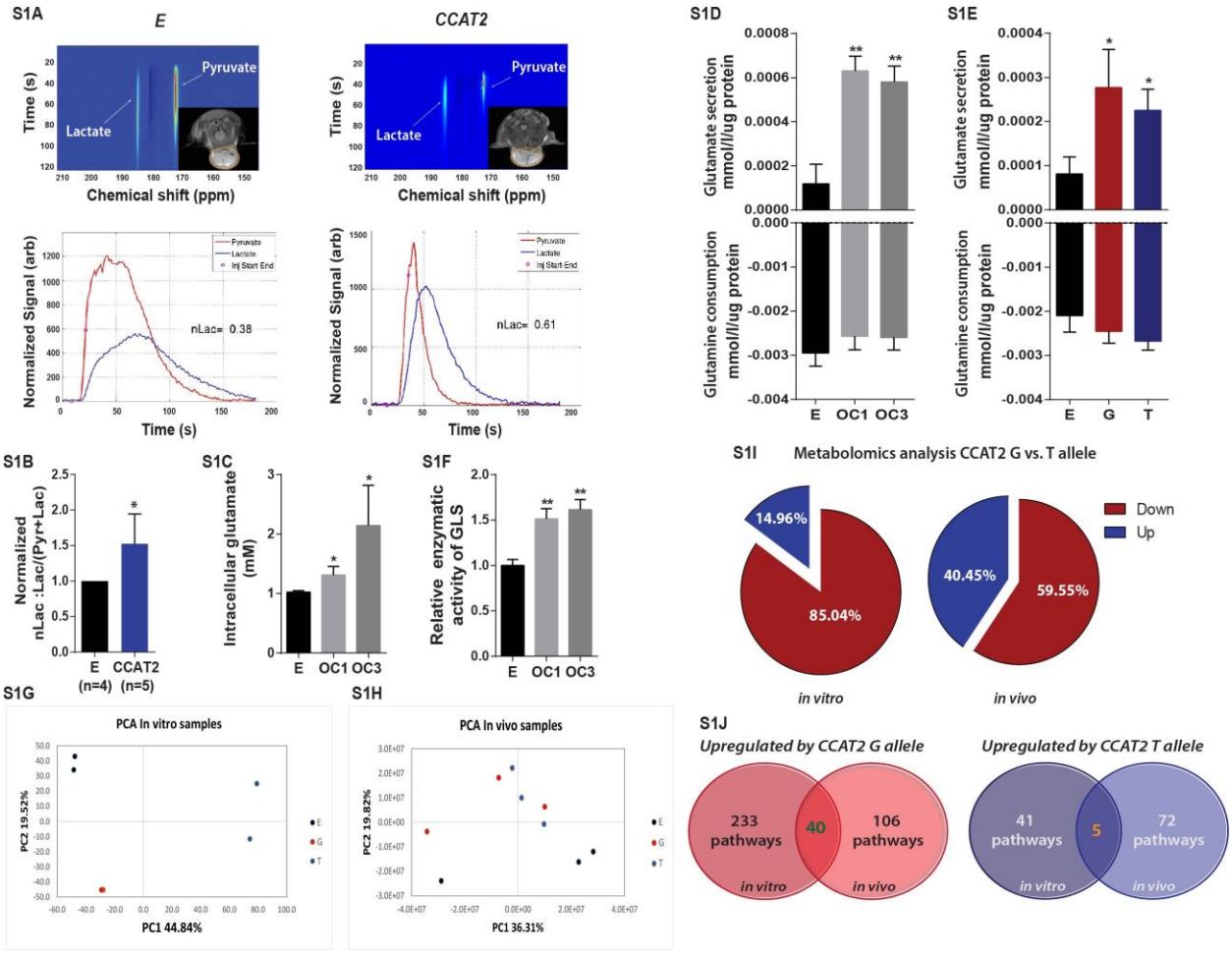


Fig S1. *CCAT2* regulates cancer metabolism *in vitro* and *in vivo* (see also Fig. 1 and Table S1) (A) Hyperpolarized ^{13}C-spectrum from xenograft tumors (E – control group; CCAT2 – *CCAT2* overexpressing group). The spectrum displays peaks from pyruvate and lactate (**upper panels**). Plot of the 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.

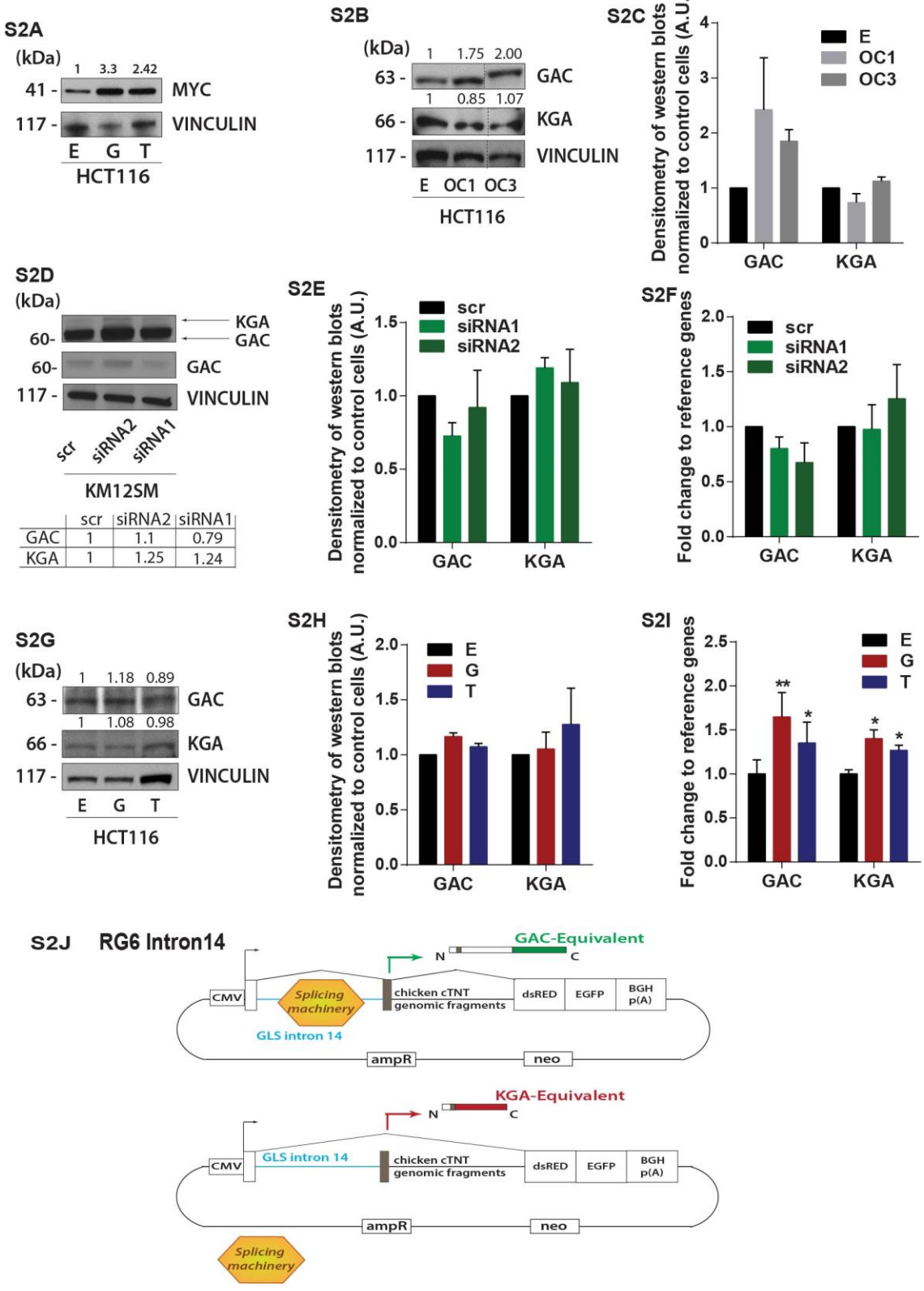
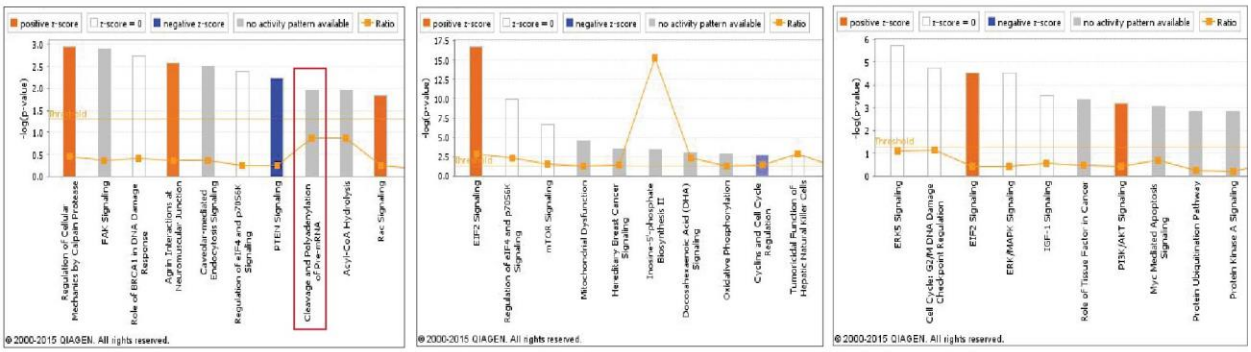
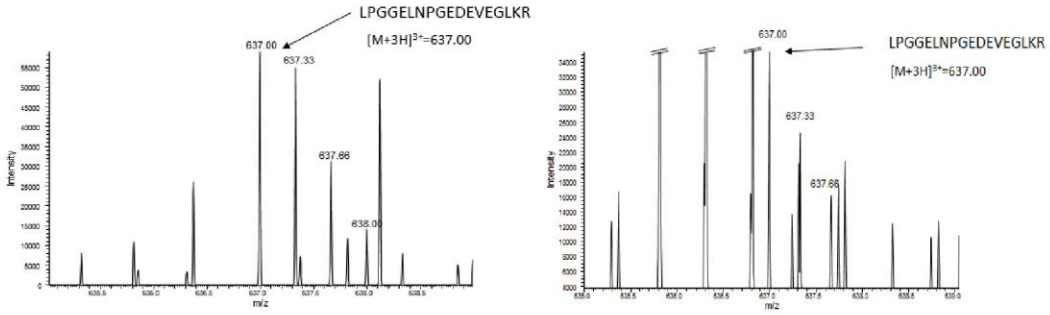

S2A
(kDa)
1 3.3 2.42
41 - MYC
117 - VINCULIN
E G T
HCT116
S2B
(kDa)
1 1.75 2.00
63 - GAC
1 0.85 1.07
66 - KGA
117 - VINCULIN
E OC1 OC3
HCT116
S2C
Densitometry of western blots normalized to control cells (A.U.)
E
OC1
OC3
GAC
KGA
S2D
(kDa)
60- KGA
GAC
60- GAC
117 - VINCULIN
scr siRNA2 siRNA1
KM12SM
scr siRNA2 siRNA1
GAC 1 1.1 0.79
KGA 1 1.25 1.24
S2E
Densitometry of western blots normalized to control cells (A.U.)
scr
siRNA1
siRNA2
GAC
KGA
S2F
Fold change to reference genes
scr
siRNA1
siRNA2
GAC
KGA
S2G
(kDa)
1 1.18 0.89
63 - GAC
1 1.08 0.98
66 - KGA
117 - VINCULIN
E G T
HCT116
S2H
Densitometry of western blots normalized to control cells (A.U.)
E
G
T
GAC
KGA
S2I
Fold change to reference genes
E
G
T
GAC
KGA
S2J RG6 Intron14
GAC-Equivalent
N
C
CMV
Splicing machinery
chicken cTNT genomic fragments
dsRED
EGFP
BGH p(A)
GLS intron 14
ampR
neo
KGA-Equivalent
N
C
CMV
GLS intron 14
chicken cTNT genomic fragments
dsRED
EGFP
BGH p(A)
ampR
neo
Splicing machinery

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



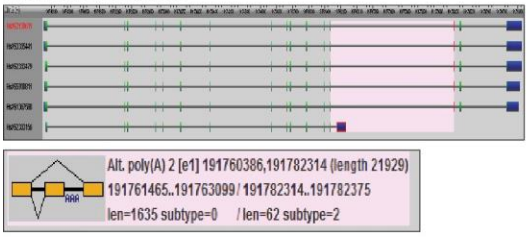
CCAT2 G-allele
m/z=636.99518, [M+H]+=1908.97098, Area=5.703E06
+3 charge, RT28.46, scan 7619

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

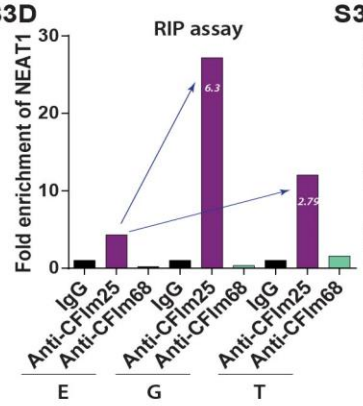
NUDT21 ratio: (CCAT2-G)/(CCAT2-T)=1.48

S3C

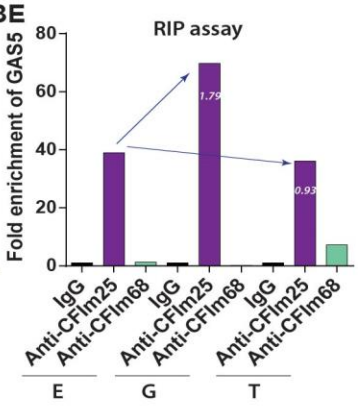
ASTRA database



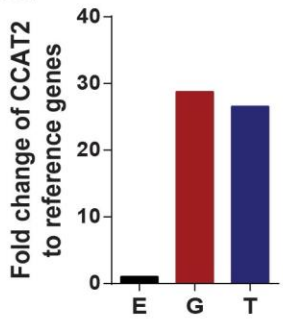
S3D



S3E



S3F



S3G

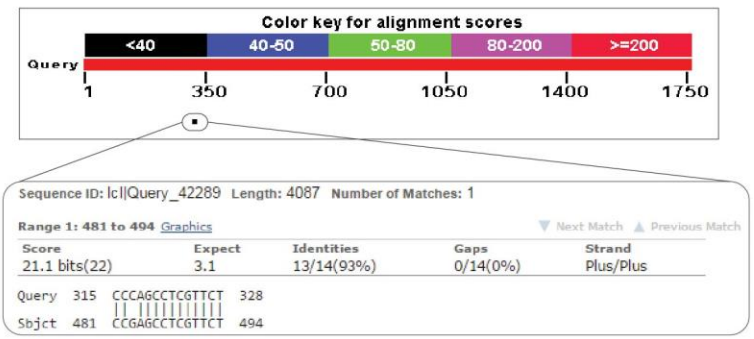


Fig. S3. The CFIm protein complex binds the *GLS* pre-mRNA (See also Fig. 3) (**A**) Pathways analysis of the proteomics data on Qiagen platform. Upper two graphs present the pathways having proteins unique either to the *CCAT2* G-allele or to the T-allele, and the lower graph presents the pathways with 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.

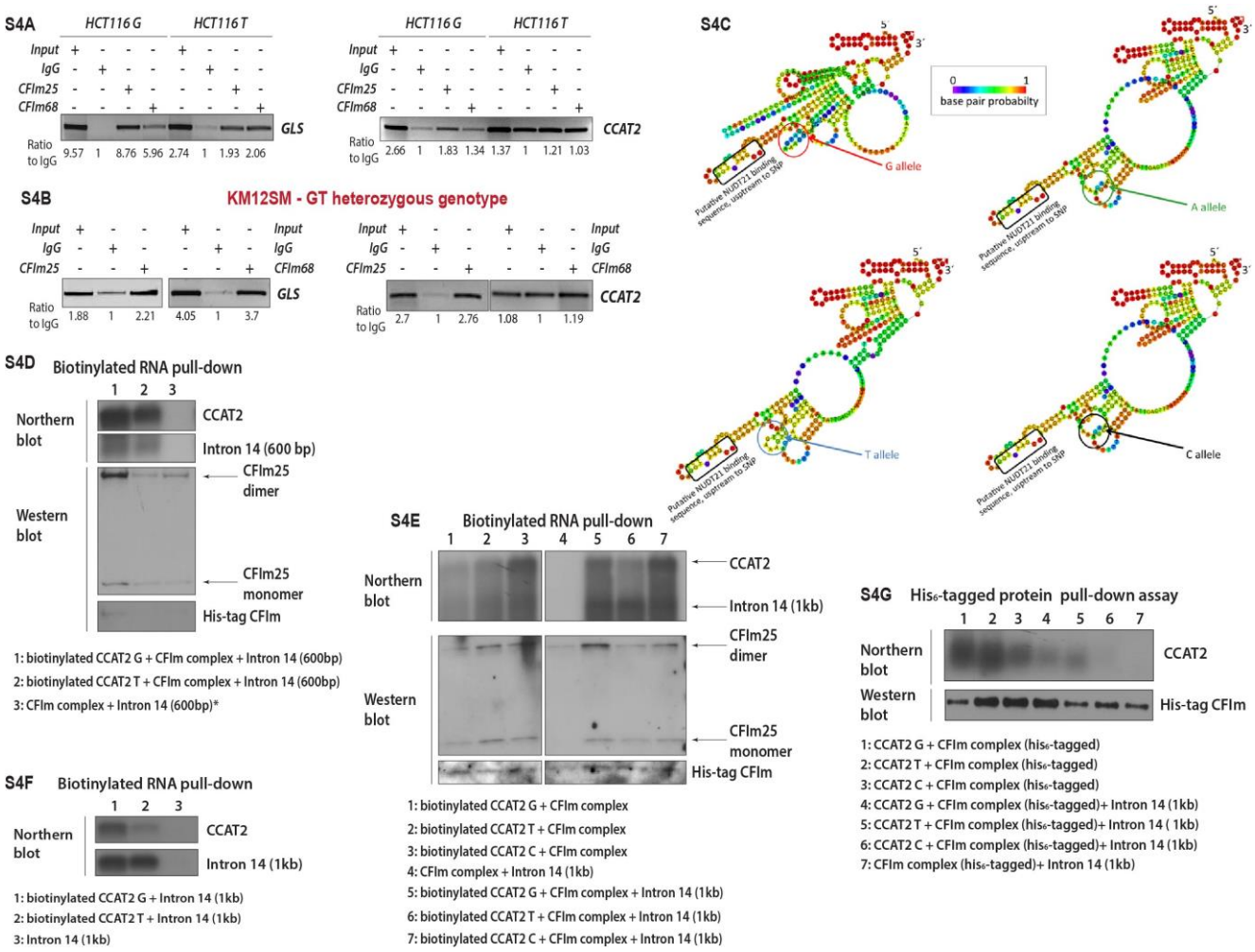


Fig. S4. The rs6983267 SNP affects the interaction of *CCAT2* with the CFIm protein complex (See also Fig. 4 and Table S3). PCR for *GLS* pre-mRNA (primers were designed to amplify the region in the intron 14 hosting the CFIm25 binding motifs) and *CCAT2* (primers were designed to amplify the region 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**).

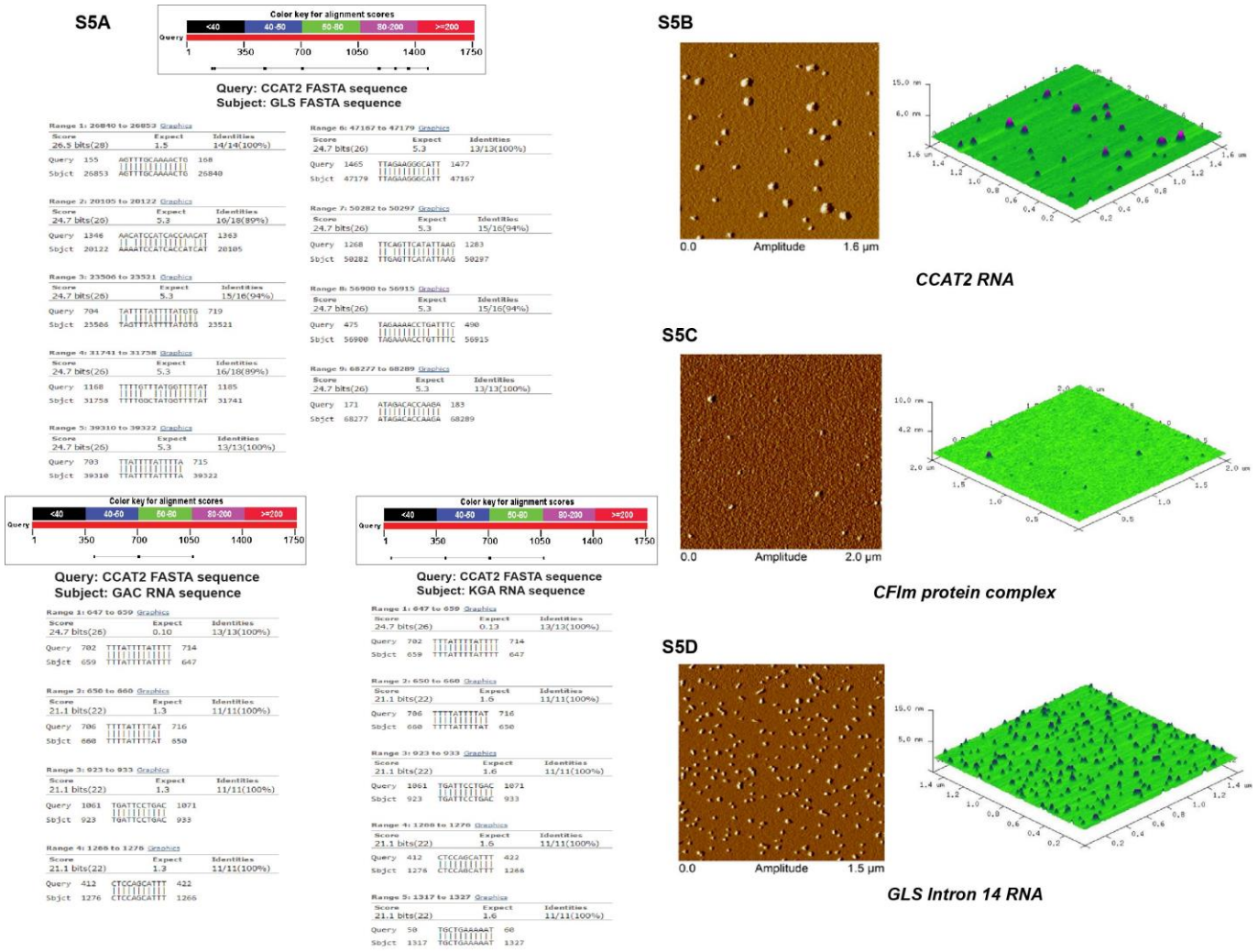
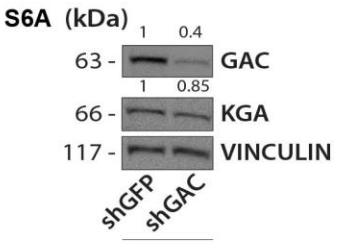
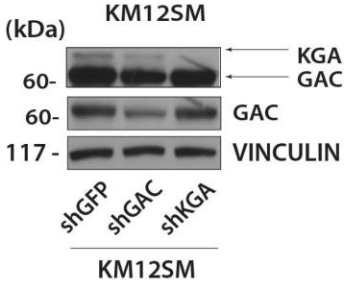
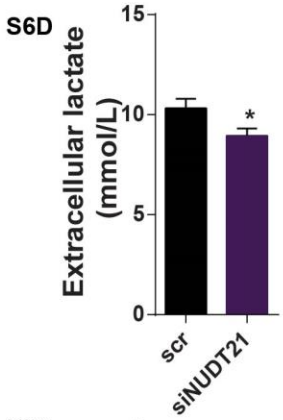


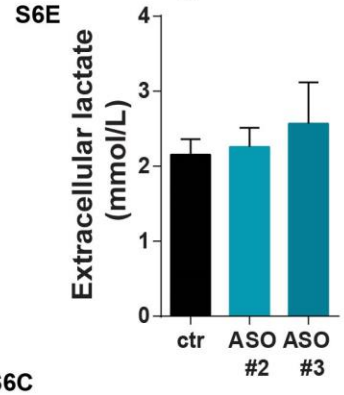
Fig. S5. ***CCAT2*** **interacts with both** ***GLS*** **pre-mRNA and the CFIm protein complex (See also Fig. 4).** (**A**) Snap-shot of NCBI Blast sequence alignment (http://blast.ncbi.nlm.nih.gov/BlastAlign.cgi) presenting the result of *CCAT2-GLS* alignments for the FASTA sequence for DNA (**upper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

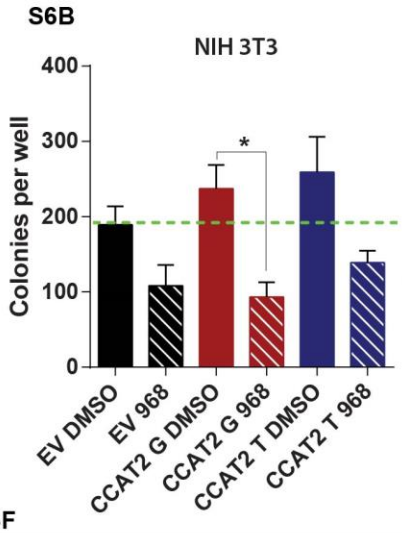


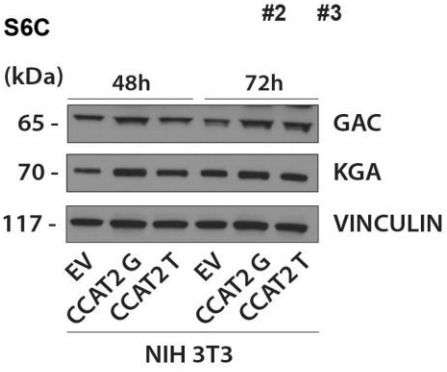


	shGFP	shGAC	shKGA
GAC	1	0.53	0.91
KGA	1	0.82	N/A









S6F

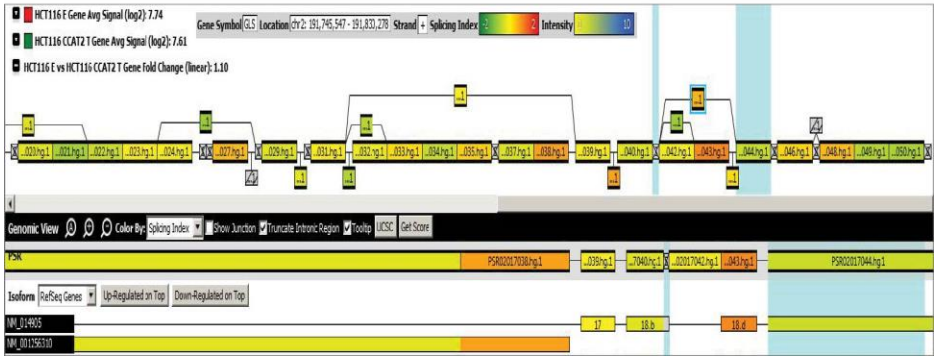


Fig. S6. The GLS isoforms contribute to the *CCAT2* induced metabolic phenotype and the *CCAT2* G-allele expressing cells depend on GAC for survival (See also Fig. 5 and Table S4) (**A**) Western blot analysis of GAC and KGA in KM12SM cells with stable downregulation of either *GAC* or *KGA*. The shKGA 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.

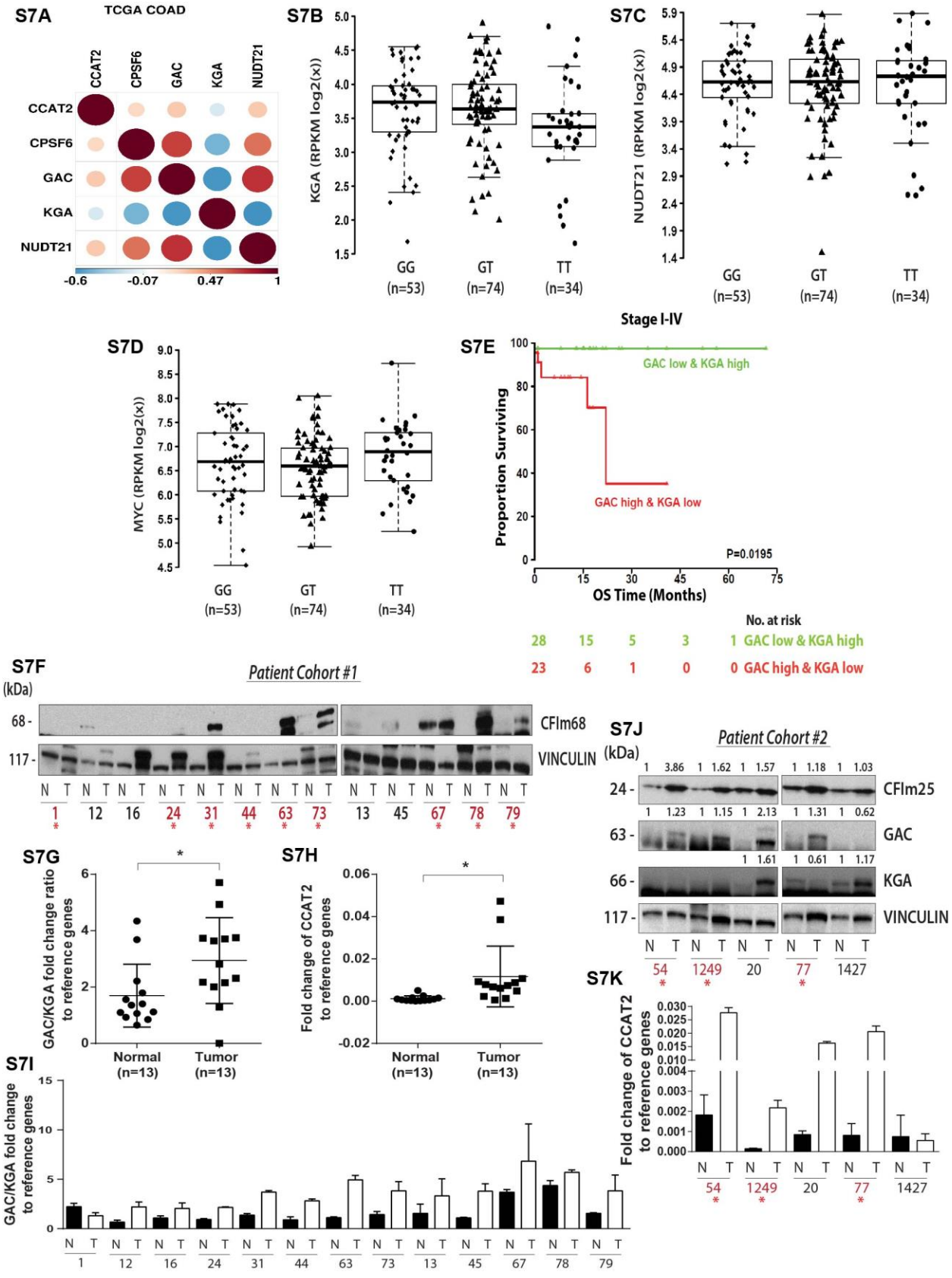

S7A
TCGA COAD
CCAT2
CPSF6
GAC
KGA
NUDT21
-0.6
-0.07
0.47
1
S7B
KGA (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7C
NUDT21 (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7D
MYC (RPKM log2(x))
GG (n=53)
GT (n=74)
TT (n=34)
S7E
Stage I-IV
GAC low & KGA high
GAC high & KGA low
Proportion Surviving
P=0.0195
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, CPSF6* and *GLS* isoforms 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**

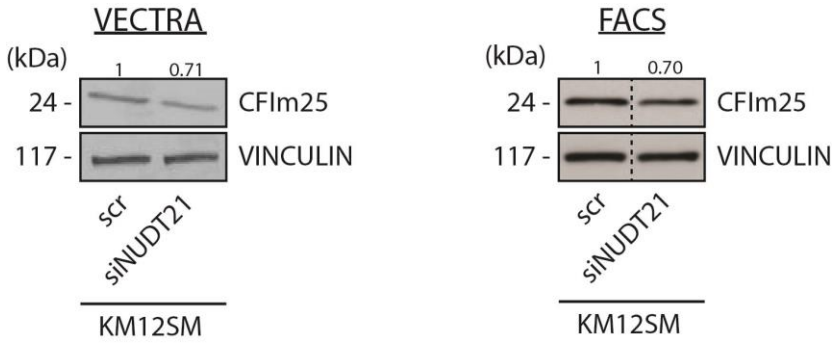


Table S3. Quantification of the Western Blot bands from **Fig. 4D**. Data is presented as ratio between the protein band and the corresponding *CCAT2* band (Northern Blot). See also **Fig. 4** and **S4**.

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 (extended).** 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.