HHS Public Access
Author manuscript
Mol Cell. Author manuscript; available in PMC 2017 February 18.

Published in final edited form as:
Mol Cell. 2016 February 18; 61(4): 520–534. doi:10.1016/j.molcel.2016.01.015.

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,¥], **Lawrence H Cheung**[1], **Gilbert Yuanjay 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], **Andre LB 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, Texas 77030, USA [2]Department of Biology & Biochemistry, University of Houston, Houston, Texas 77204, USA [3]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA [4]Department of Gastrointestinal Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 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, Texas 77030, USA [7]Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA [8]Department of Head & Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas 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 33006Spain [11]Department of Surgery, Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA [12]Department of Biochemistry and Molecular Biology, The

[#]Corresponding author: George A. Calin, gcalin@mdanderson.org, Tel – 713-792-5461; Fax – 713-745-4528.
[¥]Current affiliation: Radboud Institute for Molecular Life Sciences (RIMLS) Department of molecular biology, Nijmegen 6525, The Netherlands.

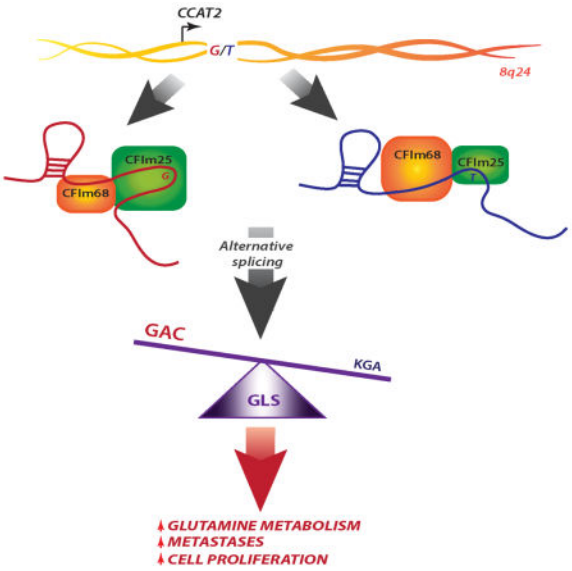
Author Contributions:
Conceived and designed the experiments: RSR, LEV, WL, JFO, AFF, MSR, MI, GB, MFF, SH, WRW, ALBA, AMGS and GAC.
Performed the experiments: RSR, LEV, WL, JFO, CRA, DA, BP, AT, YC, KVR, SC, MS, YA, LHC, GYH, PM, MSR, TCI, LV and HL.
Analyzed the data: RSR, LEV, WL, JFO, CI, CRA, DA, AT, AFF, LV, GK, JAB, MR, ALBA, SMGD and GAC.
Contributed reagents/materials/analysis tools: LH, SK, RG, GL, SYL, RCB, PH, MGR, IBN and GLB.
Wrote the first draft of the manuscript: RSR, SMGD, ALBA and GAC.
Contributed to the writing of the manuscript: RSR, SMGD, ALBA and GAC.
Agree with manuscript results and conclusions: All authors.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical Analysis: CI, LH, GK, MR and HL.
Study Supervision: SMGD and GAC.

University of Texas Health Science Center at Houston McGovern Medical School, Houston, Texas 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, Texas 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, Texas 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

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 alleles of a long ncRNA.

Graphical Abstract



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, transcriptional and post-transcriptional processing (Wilusz et al., 2009).

The revival of Warburg's theory of cancer (Warburg, 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 (Fig. 1A). These results were further confirmed in KM12SM cells with *CCAT2* downregulated expression (Fig. 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 [1-13C]pyruvate to [1-13C]lactate for the xenograft tumors derived from the *CCAT2*-overexpressing cells compared to the tumors derived from control cells (Fig. 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 *CCAT*2-overexpression and control cells. We found higher levels of both intra- and extracellular glutamate correspondent to higher levels of *CCAT2* (Fig. S1C and S1D), suggesting *CCAT2* is boosting glutamine metabolism (glutaminolysis). Surprisingly, the glutamine uptake was not significantly different between the three clones (Fig. 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 (Fig. 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-allele 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 (Fig. 1C and S1E). Moreover, the glutamine consumption was not significantly different between the clones, similar to our previous results (Fig. 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 (Fig. 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) both for the *in vitro* (Fig. 1E) and *in vivo* analysis (Fig. 1F), and similarly for the Principal Component Analysis (PCA) analysis (Fig. 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 (Fig. S1I, Table S1B). We then compared the

pathway analysis for both datasets and identified forty common pathways for the G-allele and five common pathways for the T-allele (Fig. 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 (Table S1A and S1B). We evaluated the expression 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) (Fig. 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 towards 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 towards 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-terminals) (Cassago et al., 2012) and/or a common antibody, recognizing the N-terminal 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 (Fig. 2A, S2B and S2C). Analogously, the mRNA expression pattern for GAC and KGA reflected the protein expression (Fig. 2B). Moreover, downregulation of *CCAT2* in KM12SM cells, reduced GAC protein expression with 34% and 26%, respectively (Fig. 2C, S2D and S2E), whereas KGA protein expression was either unaltered or slightly increased (Fig. 2C). Similar results were obtained when measuring the mRNA expression of the two isoforms in the same cellular model (Fig. 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 both at the mRNA and protein level, we found unanimously higher expression of GAC when *CCAT2* G-allele is upregulated (Fig. 2D, S2G, S2H and S2I). However, changes in the *KGA* mRNA expression pattern did not concur with the protein expression (Fig. 2D, S2G, S2H and S2I). The discordance observed for the KGA isoform probably is due to the regulation by the MYC - miR-23 axis (Gao et al., 2009 - additional data available upon request from authors). These results alluded 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 GFP-tagged protein, which is the equivalent of GAC, otherwise a dsRED-tagged protein will be produced, which is the equivalent of KGA

(Fig. 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 50% significantly higher EGFP/dsRED ratio in the *CCAT2*-overexpressing cells, corresponding to 50% more alternative splicing events (Fig. 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-allele, using the RG6 reporter, we found significantly higher EGFP/dsRED ratio for the *CCAT2* G-allele compared to the *CCAT2* T-allele (Fig. 2F, 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 (Fig. 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. Of note, 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 (Fig. S3B). We screened the *GLS* intron 14 for potential splicing and/or alternative polyadenylation (APA) sites using ASTRA (Alternative Splicing and TRanscription Archives database at http://alterna.cbrc.jp/) (Nagasaki et al., 2006) and identified a type 2 (skipping exon) poly(A) site (Fig. S3C) (Lutz and Moreira, 2008) 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 significant decrease in GAC protein expression in both cellular models and a clear increase the KGA protein expression (Fig. 3A). We also measured mRNA expression for the two isoforms and observed a significant reduction of the *GAC/KGA* mRNA ratio with *NUDT21* knockdown (Fig. 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 oligonucleotides (ASOs) to block the binding of the 25 kDa subunit to these motifs (Fig. 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 (Fig. 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 (RT-qPCR) (Fig. 3D, 3E and 3F, S3D, S3E and 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* (Fig 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 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) (Fig. 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) (Fig. 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 (Fig. 3D and S3D). The negative control (*GAS5*) revealed only 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 (Fig. 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 Q et al, 2011) (Fig. 3E and 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 both in the same cellular model and in KM12SM cell line, with a GT-heterozygous genotype (Fig. 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 significantly lower (28%) EGFP/dsRED ratio in the cells with *NUDT21* downregulation, correspondent to less splicing events and consequently lower expression of the GAC-equivalent (Fig. 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 (Fig. S4C). Such changes may plausibly translate into distinct tertiary folds that

could in principle explain the binding of the G- and T-allele 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 -Fig. 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 (Fig. 4A). This implied that the T-allele preferentially binds the 68 kDa subunit, but not in the region of the SNP. Of note, neither of the two mutated *CCAT2* alleles (A and C) interacted with CFIm complex and interestingly, both present secondary structures different from the G- and the T-alleles (Fig. 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, Fig. 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 identified in the pull-down also the 600 nucleotide-long region of the *GLS* pre-mRNA intron 14, containing one type-2 poly(A) site, with affinities for the protein complex corresponded to the two alleles (Fig. 4B). This suggested that the intron 14 may interact also 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. Not only did we confirm that *CCAT2* G-allele preferentially binds CFIm25, but also that *CCAT2* interacts with the intron 14 fragment in a SNP independent fashion (Fig. 4C and S4D). To ensure the specificity of the interaction, we repeated the biotin RNA pull-down assay to include biotinylated *CCAT2* C-allele as a negative control. We added also 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-allele displayed increased binding affinity to the CFIm complex, compared to the C-allele (Fig. 4D, S4E and Table S3). The G- and T-allele 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 (Fig. 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 (Fig. 4D, S4E and Table S3). We confirmed that *CCAT2* G- and T-allele 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 (Fig. S4F). We observed multiple short fragments (13 to 18 nt) of sequence complementarity spanning the entire *GLS* sequence, present in both

introns and exons (Fig. 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 (Fig. 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) (Fig. 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 (Fig. 4E and 4F), while the individual components (CFIm protein complex, *CCAT2* RNA and intron 14 RNA) appeared to be evidently smaller (Fig. S5B, S5C and 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 approximately corresponds 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 observed a reduction by half of the migration in the cells where GLS was either inhibited or downregulated (Fig. 5A). We next assessed the individual contribution of GAC to the migration, in KM12SM cells with stable downregulation of the isoform (Fig. S6A) and found that 60% less cells migrated when GAC expression was reduced (Fig. 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 (Fig. 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 (Fig. 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* (Fig. 5E). Additional confirmation of the higher dependency of the *CCAT2* G-allele on GAC was provided by the colony formation assay for NIH3T3 cells transfected with the *CCAT2*-overexpressing vectors (Fig. S6B and S6C).

We next sought to determine if the shift in GLS isoforms expression and CFIm25 are 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 (Fig. S6D and 5F). However these metabolic changes appeared to mirror only partly 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 (Fig. 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 and 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 for *GLS* a negative splicing index (SI=−1.19) 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 (Fig. 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 (Fig. 6A, 6B, 6C and 6D). We further analyzed the associations between *CCAT2*, *NUDT21, CPSF6, GAC* and *KGA* in the TCGA dataset of colon cancer samples (2012) and detected a direct correlations of *CCAT2, CPSF6, 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, Fig. 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, $P<0.0001$, Fig. S7A), as well as between *KGA* and *GAC* (r_s= −0.590, $P<0.0001$) (Fig. 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,* Fig. 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 genotype. We observed a significant association with the genotypes for *GAC*, having the highest expression in patients with GG genotype, but not for *KGA* (Fig. 6E and S7B). Similarly, we did not find any association of *NUDT21* or *MYC* with the genotypes (Fig. 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 (Fig. 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 (Fig. 7A and 7B). We further inquired whether the genes that were found significantly correlated with *CCAT2* expression were also associated with the genotypes of rs6983267 SNP. We repeated the GSEA analysis and screened for pathways that correlated with either of the genotypes (FDR q-val<0.25 and Nominal *P*-val<0.05). We identified various metabolic and RNA processing pathways that were associated with certain genotypes (Fig. 7C). Interestingly, several of the pathways were related with the ones identified by the Affymetrix HTA 2.0 array (highlighted in green in Fig. 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 same high protein levels of CFIm25 and GAC in tumor tissue compared to normal mucosae for 61.5% (8/13) of the pairs (Fig. 6F). As for the KGA isoform, protein levels were mostly lower in tumor samples or comparable to the levels in normal samples (Fig. 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 (Fig. S7F). We also measured *CCAT2* expression by RT-qPCR in the same set of samples and confirmed a positive correlation between *CCAT2,* CFIm25 and GAC for 69.2% of the samples (Fig. 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) (Fig. S7I). We repeated the measurements in a second set of paired normal mucosae and CRC samples (*Patient Cohort #2*) and obtained similar results (Fig. 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 *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

only in 62.5% of patients (5/8) (Fig. 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 a lncRNA to alter the function of the partner RNA-binding protein/complex. The enrichment on metabolites related to TCA cycle that we observed for tumors derived from HCT116 cell overexpressing *CCAT2* 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 in particular the cancer-specific GAC isoform.

We reveal also 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 Experimental Procedures are presented in Supplementary Experimental Procedures section.

Glucose uptake assay

Cells were plated in 96 well plates (25,000 cells/well) 16 h before performing the assay. The medium was removed and cells were washed twice with PBS. To the wells containing the blanks 50 μl PBS was added, while for the wells with the samples 50 μl 2-NBDG (100 μM) (Sigma) was added and the mixture was incubated 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 ul 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 h) 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) and following manufacturer's protocol. Briefly, 1 × 106 cells per tested sample were homogenized in 100 μl assay buffer and centrifuged to remove insoluble material. One hundred μl 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Dr Calin is The Alan M. Gewirtz Leukemia & Lymphoma Society Scholar. Work in Dr. Calin'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,. Dr. Berindan-Neagoe was financed by a POSCCE grant (709/2010) entitled Clinical and Economical Impact of Proteome and Transcriptome Molecular Profiling in Neoadjuvant Therapy of Triple Negative Breast Cancer (BREASTIMPACT). Dr. Gomes Dias, Dr. Ambrosio and Dr Adamoski are supported by the São Paulo Research Foundation FAPESP, under grants 2014/15968-3, 2014/20673-2 and 2014/17820-3, respectively. Dr. Lu 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. Dr Bankson 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. Dr. Liang 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 (Univ. of Ferrara) and Linda Ulazzi (Univ. of Ferrara) for the technical support. We also would like to thank the Biological Imaging Facility, the Protein Purification and the Bioassay Laboratories (LNBio/CNPEM/Brazil) for the access to the equipment.

References

Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012; 487:330–337. [PubMed: 22810696]

Bester AC, Roniger M, Oren YS, Im MM, Sarni D, Chaoat M, Bensimon A, Zamir G, Shewach DS, Kerem B. Nucleotide deficiency promotes genomic instability in early stages of cancer development. Cell. 2011; 145:435–446. [PubMed: 21529715]

Boroughs LK, DeBerardinis RJ. Metabolic pathways promoting cancer cell survival and growth. Nature Cell Biology. 2015; 17:351–359. [PubMed: 25774832]

Carroll PA, Diolaiti D, McFerrin L, Gu H, Djukovic D, Du J, Cheng PF, Anderson S, Ulrich M, Hurley JB, et al. Deregulated Myc Requires MondoA/Mlx for Metabolic Reprogramming and Tumorigenesis. Cancer Cell. 2015; 27:271–285. [PubMed: 25640402]

Cassago A, Ferreira AP, Ferreira IM, Fornezari C, Gomes ER, Greene KS, Pereira HM, Garratt RC, Dias SM, Ambrosio AL. 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. 2012; 109:1092–1097. [PubMed: 22228304]

Chen JQ, Russo J. Dysregulation of glucose transport, glycolysis, TCA cycle and glutaminolysis by oncogenes and tumor suppressors in cancer cells. Biochim Biophys Acta. 2012; 1826:370–384. [PubMed: 22750268]

Elkon R, Ugalde AP, Agami R. Alternative cleavage and polyadenylation: extent, regulation and function. Nature Reviews Genetics. 2013; 14:496–506.

Gao P, Tchernyshyov I, Chang TC, Lee YS, Kita K, Ochi T, Zeller KI, De Marzo AM, Van Eyk JE, Mendell JT, Dang CV. c-Myc suppression of miR-23a/b enhances mitochondrial glutaminase expression and glutamine metabolism. Nature. 2009; 458:762–765. [PubMed: 19219026]

Huang F, Zhang Q, Ma H, Lv Q, Zhang T. Expression of glutaminase is upregulated in colorectal cancer and of clinical significance. International Journal of Clinical and Experimental Pathology. 2014; 7:1093–1100. [PubMed: 24696726]

Jacque N, Ronchetti AM, Larrue C, Meunier G, Birsen R, Willems L, Saland E, Decroocq J, Thiago TT, Lambert M, et al. Targeting glutaminolysis has antileukemic activity in acute myeloid leukemia and synergizes with BCL-2 inhibition. Blood. 2015; 126:1346–1356. [PubMed: 26186940]

Jeong SM, Xiao C, Finley LW, Lahusen T, Souza AL, Pierce K, Li YH, Wang X, Laurent G, German NJ, et al. SIRT4 has tumor-suppressive activity and regulates the cellular metabolic response to DNA damage by inhibiting mitochondrial glutamine metabolism. Cancer Cell. 2013; 23:450–463. [PubMed: 23562301]

Kaldma A, Klepinin A, Chekulayev V, Mado K, Shevchuk I, Timohhina N, Tepp K, Kandashvili M, Varikmaa M, Koit A, et al. 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. The International Journal of Biochemistry & Cell Biology. 2014; 55C:171–186. [PubMed: 25218857]

Katt WP, Ramachandran S, Erickson JW, Cerione RA. Dibenzophenanthridines as inhibitors of glutaminase C and cancer cell proliferation. Molecular Cancer Therapeutics. 2012; 11:1269–1278. [PubMed: 22496480]

Le A, Lane AN, Hamaker M, Bose S, Gouw A, Barbi J, Tsukamoto T, Rojas CJ, Slusher BS, Zhang H, et al. Glucose-independent glutamine metabolism via TCA cycling for proliferation and survival in B cells. Cell Metabolism. 2012; 15:110–121. [PubMed: 22225880]

Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. Nature Reviews Drug Discovery. 2013a; 12:847–865. [PubMed: 24172333]

Ling H, Spizzo R, Atlasi Y, Nicoloso M, Shimizu M, Redis RS, Nishida N, Gafa R, Song J, Guo Z, et al. CCAT2, a novel noncoding RNA mapping to 8q24, underlies metastatic progression and chromosomal instability in colon cancer. Genome Research. 2013b; 23:1446–1461. [PubMed: 23796952]

Lutz CS, Moreira A. Alternative mRNA polyadenylation in eukaryotes: an effective regulator of gene expression. Wiley Interdiscip Rev RNA. 2008; 2:23–31. [PubMed: 21278855]

Lyubchenko YL, Shlyakhtenko LS, Ando T. Imaging of nucleic acids with atomic force microscopy. Methods. 2011; 54:274–283. [PubMed: 21310240]

Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. Nature. 2014; 510:412–416. [PubMed: 24814343]

Mercer TR, Dinger ME, Mattick JS. Long non-coding RNAs: insights into functions. Nat Rev Genet. 2009; 10:155–159. [PubMed: 19188922]

Millevoi S, Loulergue C, Dettwiler S, Karaa SZ, Keller W, Antoniou M, Vagner S. An interaction between U2AF 65 and CF I(m) links the splicing and 3′ end processing machineries. The EMBO Journal. 2006; 25:4854–4864. [PubMed: 17024186]

Naganuma T, Nakagawa S, Tanigawa A, Sasaki YF, Goshima N, Hirose T. Alternative 3′-end processing of long noncoding RNA initiates construction of nuclear paraspeckles. The EMBO Journal. 2012; 31:4020–4034. [PubMed: 22960638]

Nagasaki H, Arita M, Nishizawa T, Suwa M, Gotoh O. Automated classification of alternative splicing and transcriptional initiation and construction of visual database of classified patterns. Bioinformatics. 2006; 22:1211–1216. [PubMed: 16500940]

Orengo JP, Bundman D, Cooper TA. A bichromatic fluorescent reporter for cell-based screens of alternative splicing. Nucleic Acids Research. 2006; 34:e148. [PubMed: 17142220]

Rathore MG, Saumet A, Rossi JF, de Bettignies C, Tempe D, Lecellier CH, Villalba M. The NF-kappaB member p65 controls glutamine metabolism through miR-23a. Int J Biochem Cell Biol. 2012; 44:1448–1456. [PubMed: 22634383]

Redis RS, Sieuwerts AM, Look MP, Tudoran O, Ivan C, Spizzo R, Zhang X, de Weerd V, Shimizu M, Ling H, et al. CCAT2, a novel long non-coding RNA in breast cancer: expression study and clinical correlations. Oncotarget. 2013; 4:1748–1762. [PubMed: 24077681]

Stine ZE, Walton ZE, Altman BJ, Hsieh AL, Dang CV. MYC, Metabolism, and Cancer. Cancer Discovery. 2015; 5:1024–1039. [PubMed: 26382145]

Tian B, Pan Z, Lee JY. Widespread mRNA polyadenylation events in introns indicate dynamic interplay between polyadenylation and splicing. Genome Res. 2007; 17:156–165. [PubMed: 17210931]

Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. Nature Genetics. 2007; 39:984–988. [PubMed: 17618284]

Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymaki I, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nature Genetics. 2009; 41:885–890. [PubMed: 19561604]

van den Heuvel AP, Jing J, Wooster RF, Bachman KE. Analysis of glutamine dependency in non-small cell lung cancer: GLS1 splice variant GAC is essential for cancer cell growth. Cancer Biology & Therapy. 2012; 13:1185–1194. [PubMed: 22892846]

Vander Heiden MG. Targeting cancer metabolism: a therapeutic window opens. Nature Reviews Drug Discovery. 2011; 10:671–684. [PubMed: 21878982]

Wang J, Liu X, Wu H, Ni P, Gu Z, Qiao Y, Chen N, Sun F, Fan Q. CREB up-regulates long non-coding RNA, HULC expression through interaction with microRNA-372 in liver cancer. Nucleic Acids Res. 2010a; 38:5366–5383. [PubMed: 20423907]

Wang JB, Erickson JW, Fuji R, Ramachandran S, Gao P, Dinavahi R, Wilson KF, Ambrosio AL, Dias SM, Dang CV, Cerione RA. Targeting mitochondrial glutaminase activity inhibits oncogenic transformation. Cancer Cell. 2010b; 18:207–219. [PubMed: 20832749]

Warburg O, Posener K, Negelein E. Über den Stoffwechsel der Carcinomzelle. Biochem Zeitschr. 1924; 152:309–344.

Ward PS, Thompson CB. Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. Cancer Cell. 2012; 21:297–308. [PubMed: 22439925]

Wilusz JE, Sunwoo H, Spector DL. Long noncoding RNAs: functional surprises from the RNA world. Genes Dev. 2009; 23:1494–1504. [PubMed: 19571179]

Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nature Communications. 2014; 5:5274.

Yang F, Zhang H, Mei Y, Wu M. Reciprocal regulation of HIF-1alpha and lincRNA-p21 modulates the Warburg effect. Molecular Cell. 2014; 53:88–100. [PubMed: 24316222]

Yang Q, Gilmartin GM, Doublie S. 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 Biology. 2011; 8:748–753. [PubMed: 21881408]

Yoon JH, Srikantan S, Gorospe M. MS2-TRAP (MS2-tagged RNA affinity purification): tagging RNA to identify associated miRNAs. Methods. 2012; 58:81–87. [PubMed: 22813890]

Zhou Z, Licklider LJ, Gygi SP, Reed R. Comprehensive proteomic analysis of the human spliceosome. Nature. 2002; 419:182–185. [PubMed: 12226669]

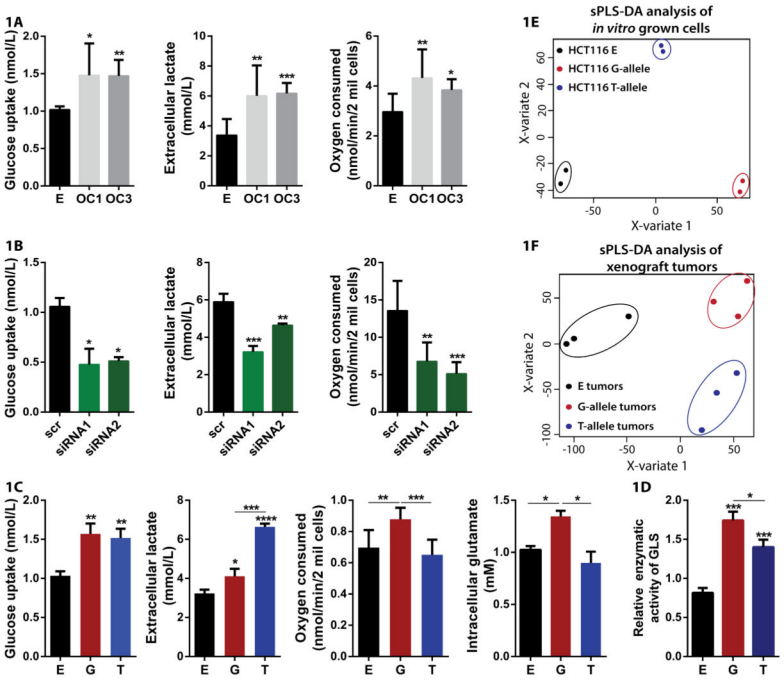


Figure 1. *CCAT2* regulates cancer metabolism *in vitro* and *in vivo*

(**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-allele, and control cells (E – empty vector) *in vitro* allowed an adequate classification of the different cell lines according to its metabolome*. (**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 Fig. S1 and Table S1.

*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.

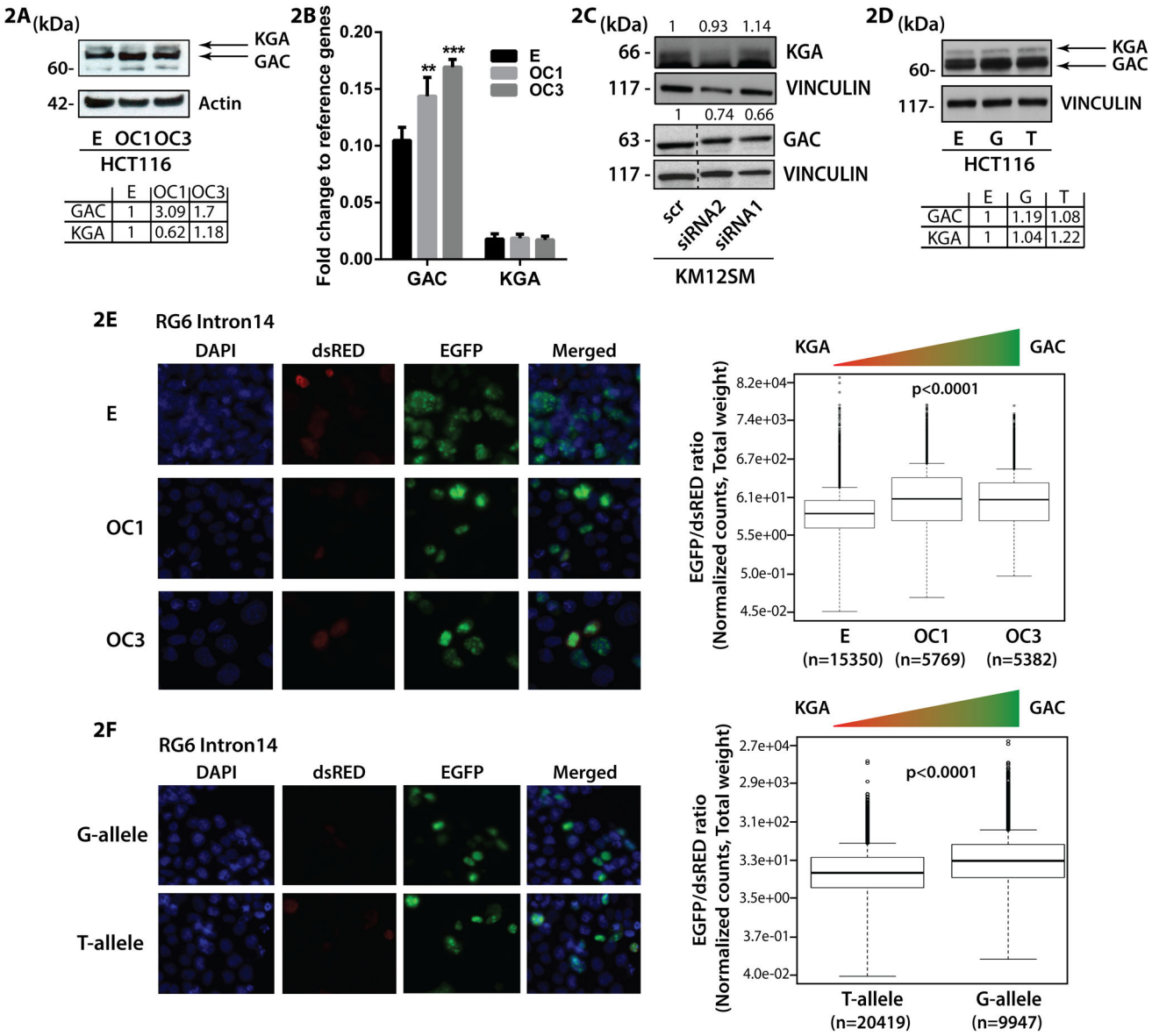


Figure 2. *CCAT2* induces the preferential splicing of GAC

(**A**) Western Blot analysis of GAC, KGA in HCT116 *CCAT2*-overexpressing (OC1 and OC3) and control cells. (**B**) RT-qPCR 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-allele 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. (**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. Results are presented as normalized mean values ± SD. See also Fig. S2 and **Table 2.**

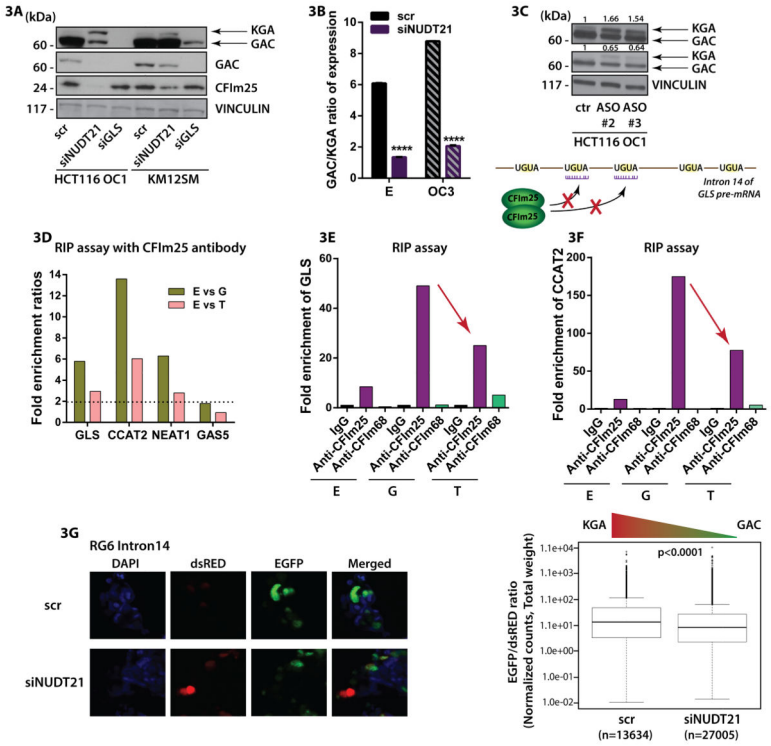


Figure 3. CFIm protein complex binds *GLS* pre-mRNA

(**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**) RT-qPCR 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**) RT-qPCR 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. RT-qPCR assessing the fold enrichment of *GLS* mRNA (**E**) and *CCAT2* (**F**) bound to CFIm25 and CFIm68 in HCT116 cells *CCAT2*-overexpressing G- or T-allele 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. Results are presented as normalized mean values ± SD. See also Fig. S3 and **Table 2.**

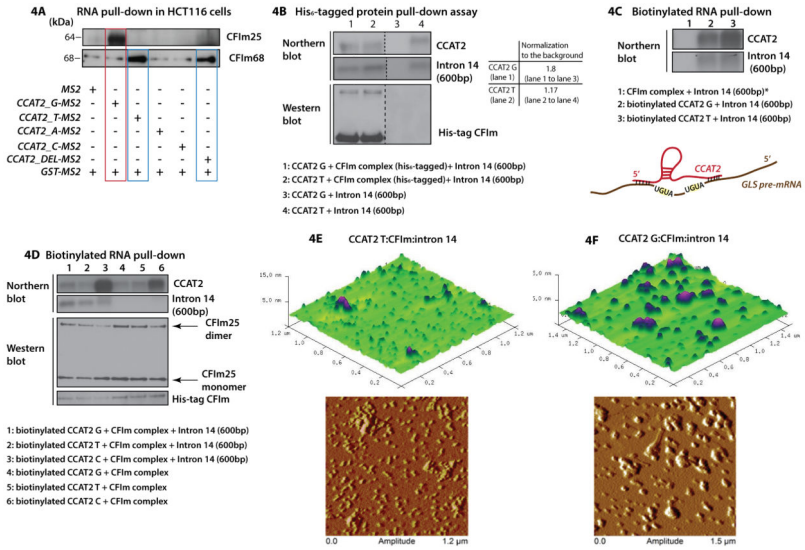


Figure 4. The rs6983267 SNP afftects the interaction of *CCAT2* with the CFIm protein complex

(**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 (600bp 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 (600bp fragment) in the lysate pulled down with Streptavidin beads. Lane 1 marked with the star symbol is identical to lane 3 in Fig. 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 (600bp 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**, **F**) AFM images of CCAT2:CFIm:intron 14 quaternary complex including either *CCAT2* T-allele (**E**) or *CCAT2* G-allele (**F**). See also Fig. S4, S5 and **Table 3.**

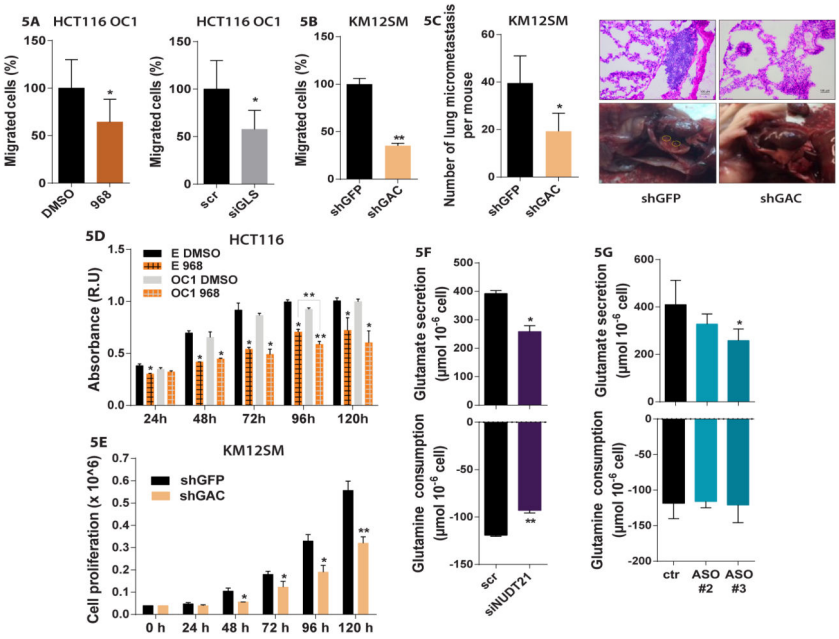


Figure 5. GLS promotes *in vivo* metastases and *in vitro* cell proliferation and migration

(**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 and 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, 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**). Glutamine and glutamate concentration in the media relative to the empty well 24 hours 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 Fig. S6 and **Table 4.**

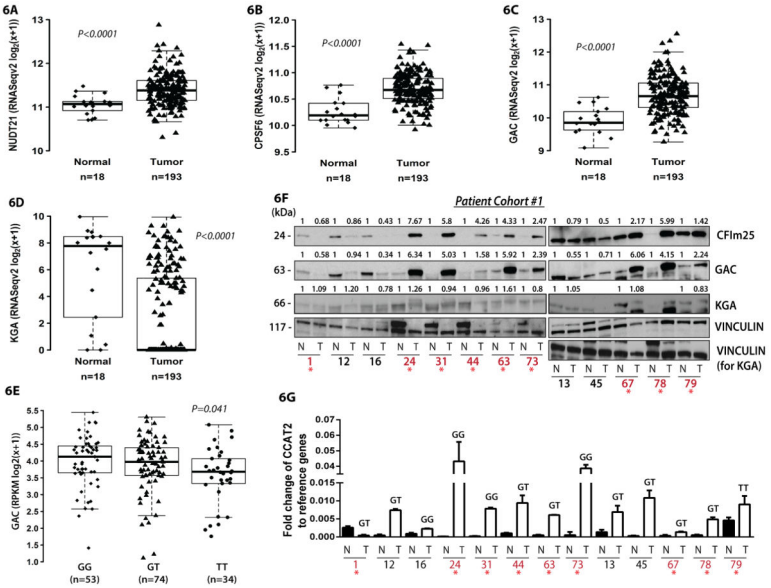


Figure 6. *CCAT2*, *NUDT21* (CFIm25), *CPSF6* (CFIm68) and *GLS* (GAC and KGA) expression pattern in TCGA dataset and CRC patient samples

(**A, B, C** and **D**) Analysis of *NUDT21, CPSF6, GAC* and *KGA* 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**) RT-qPCR analysis for *CCAT2* in the same paired CRC samples (*Patient cohort #1).* Results are presented as normalized mean values ± SD. See also Fig. S7.

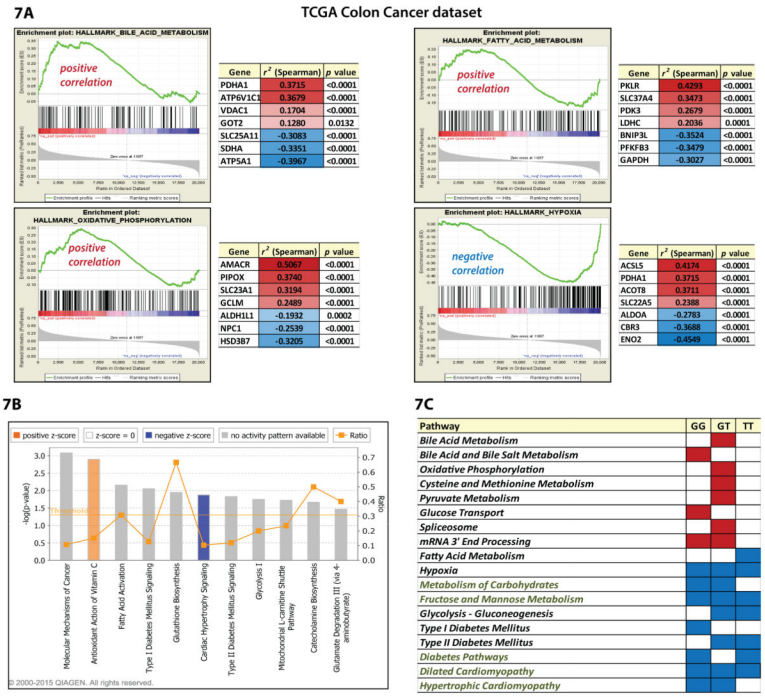


Figure 7. *CCAT2* gene signature in colon cancer patients (TCGA dataset)

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 panels (**A**) and (**B**). (**C**) Table containing the pathways significantly associated with *CCAT2* expression and rs6983267 genotype (FDR q-val<0.25 and Nom p-val<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 Fig. S7 and Table S4.