



Article

NAGS, CPS1, and SLC25A13 (Citrin) at the Crossroads of Arginine and Pyrimidines Metabolism in Tumor Cells

Melissa Owusu-Ansah^{1,2}, Nikita Guptan¹, Dylon Alindogan¹, Michio Morizono³ and Ljubica Caldovic^{4,5,*} 

¹ Columbian College of Arts and Sciences, George Washington University, Washington, DC 20052, USA

² Department of Microbiology, Immunology, and Tropical Medicine, School of Medicine and Health Sciences, George Washington University, Washington, DC 20052, USA

³ School of Mathematics, College of Science and Engineering, University of Minnesota, Minneapolis, MN 55455, USA

⁴ Department of Genomics and Precision Medicine, School of Medicine and Health Sciences, George Washington University, Washington, DC 20052, USA

⁵ Center for Genetic Medicine Research, Children's National Research Institute, Children's National Hospital, Washington, DC 20010, USA

* Correspondence: lcaldovic@childrensnational.org or ljubica@gwu.edu; Tel.: +1-202-545-2804

Abstract: Urea cycle enzymes and transporters collectively convert ammonia into urea in the liver. Aberrant overexpression of carbamylphosphate synthetase 1 (*CPS1*) and *SLC25A13* (citrin) genes has been associated with faster proliferation of tumor cells due to metabolic reprogramming that increases the activity of the CAD complex and pyrimidine biosynthesis. N-acetylglutamate (NAG), produced by NAG synthase (*NAGS*), is an essential activator of *CPS1*. Although *NAGS* is expressed in lung cancer derived cell lines, expression of the *NAGS* gene and its product was not evaluated in tumors with aberrant expression of *CPS1* and citrin. We used data mining approaches to identify tumor types that exhibit aberrant overexpression of *NAGS*, *CPS1*, and citrin genes, and evaluated factors that may contribute to increased expression of the three genes and their products in tumors. Median expression of *NAGS*, *CPS1*, and citrin mRNA was higher in glioblastoma multiforme (GBM), glioma, and stomach adenocarcinoma (STAD) samples compared to the matched normal tissue. Median expression of *CPS1* and citrin mRNA was higher in the lung adenocarcinoma (LUAD) sample while expression of *NAGS* mRNA did not differ. High *NAGS* expression was associated with an unfavorable outcome in patients with glioblastoma and GBM. Low *NAGS* expression was associated with an unfavorable outcome in patients with LUAD. Patterns of DNase hypersensitive sites and histone modifications in the upstream regulatory regions of *NAGS*, *CPS1*, and citrin genes were similar in liver tissue, lung tissue, and A549 lung adenocarcinoma cells despite different expression levels of the three genes in the liver and lung. Citrin gene copy numbers correlated with its mRNA expression in glioblastoma, GBM, LUAD, and STAD samples. There was little overlap between *NAGS*, *CPS1*, and citrin sequence variants found in patients with respective deficiencies, tumor samples, and individuals without known rare genetic diseases. The correlation between *NAGS*, *CPS1*, and citrin mRNA expression in the individual glioblastoma, GBM, LUAD, and STAD samples was very weak. These results suggest that the increased cytoplasmic supply of either carbamylphosphate, produced by *CPS1*, or aspartate may be sufficient to promote tumorigenesis, as well as the need for an alternative explanation of *CPS1* activity in the absence of *NAGS* expression and NAG.

Keywords: N-acetylglutamate synthase; carbamylphosphate synthetase 1; citrin; urea cycle; pyrimidine biosynthesis; CAD complex; tumor; tumorigenesis



Citation: Owusu-Ansah, M.; Guptan, N.; Alindogan, D.; Morizono, M.; Caldovic, L. *NAGS*, *CPS1*, and *SLC25A13* (Citrin) at the Crossroads of Arginine and Pyrimidines Metabolism in Tumor Cells. *Int. J. Mol. Sci.* **2023**, *24*, 6754. <https://doi.org/10.3390/ijms24076754>

Academic Editor: Roberto Ronca

Received: 9 February 2023

Revised: 13 March 2023

Accepted: 22 March 2023

Published: 4 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The reprogramming of cellular metabolism to favor anabolic pathways can contribute to the uncontrolled growth and proliferation of tumor cells. The metabolic reprogramming that drives tumor development can be achieved through gain-of-function somatic

mutations that result in the increased activity of enzymes such as isocitrate dehydrogenase 1 and 2 [1–5]. Heterozygous germline sequence variants that result in a lower activity of fumarate hydratase and succinate dehydrogenase can increase the risk of tumor development [6–8]. Aberrant expression of enzymes belonging to catabolic pathways such as urea cycle can contribute to the increased proliferation of tumor cells and unfavorable patient outcomes [9].

The urea cycle comprises six enzymes and two mitochondrial solute carriers that convert ammonia, a product of amino acid catabolism and a potent neurotoxin, into non-toxic urea in the liver [10,11]. Urea cycle enzymes are partitioned between the mitochondria and cytoplasm of liver cells. N-acetylglutamate synthase (NAGS), carbamylphosphate synthetase I (CPSI), and ornithine transcarbamylase (OTC) are located in the mitochondrial matrix [11]. CPSI and OTC collectively convert ammonia, bicarbonate, ATP, and ornithine into citrulline, which is then converted into urea by the cytosolic enzymes argininosuccinate synthetase 1 (ASS1), argininosuccinate lyase (ASL), and arginase 1 (ARG1) [11]. NAGS uses glutamate and acetyl coenzyme A to produce NAG, an obligatory allosteric activator of CPSI [11–18]. NAGS, CPSI, and OTC are also expressed in the small intestine, where they synthesize citrulline which is then transported to the kidneys for the biosynthesis of arginine by ASS1 and ASL [11]. ASS1 and ASL also function in NO signaling, where they catalyze the formation of arginine which is a substrate of NO-synthetase [19]. Mitochondrial solute carriers that function in urea cycle are ornithine transporter 1 (ORNT1), encoded by the *SLC25A15* gene, and aspartate-glutamate carrier 2, also known as citrin or ARALAR 2 and encoded by the *SLC25A13* gene [11,20–22]. ORNT1 transports ornithine from the cytoplasm into mitochondria and citrulline out of mitochondria into the cytoplasm [23]. Citrin transports aspartate from mitochondria into the cytoplasm where it is used as an ASS1 substrate [24–26]. Citrin also participates in the malate-aspartate shuttle, which maintains the redox balance by transporting electrons produced in the cytoplasm by glycolytic enzymes into mitochondria for use in energy production [22,26].

Aberrant expression of urea cycle enzymes and mitochondrial solute carriers can contribute to the metabolic reprogramming of tumor cells [9]. Silencing of ASS1 and ASL in hepatocellular carcinoma, bladder cancer, and lung adenocarcinoma can lead to an increased supply of cytoplasmic aspartate which is a substrate of the CAD complex (carbamylphosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase) that catalyzes the first three reactions of de novo pyrimidine biosynthesis [27,28]. Overexpression of citrin can contribute to the metabolic reprogramming of tumor cells in two ways. First, citrin overexpression can contribute to the increased supply of cytoplasmic aspartate, increased de novo biosynthesis of pyrimidine nucleotides, faster cell proliferation, and unfavorable outcomes in patients with lung adenocarcinoma, bladder cancer, and uterine tumors [22,28]. Second, citrin overexpression can disrupt the function of the malate-aspartate shuttle leading to a redox imbalance and the disruption of energy metabolism in tumor cells [29,30]. Aberrant expression of *CPS1* in tumors that originate in tissues that normally do not express the gene has been associated with an increased cytoplasmic supply of carbamylphosphate (CP), a substrate of the CAD complex, leading to increased de novo pyrimidine biosynthesis, tumor cell proliferation, and unfavorable patient outcomes [31–35]. Although these studies provided evidence of CPS1 enzymatic activity in tumor cells, only one study showed the presence of NAGS in cell lines that model non-small cell lung carcinoma and express CPS1 [36]. Since the aberrant expression of two or more urea cycle genes is needed for the metabolic reprogramming of tumor cells [9], we used data mining approaches to identify tumor types with aberrant expression of *NAGS*, *CPS1*, and citrin genes that can increase the cytoplasmic concentration of CP and aspartate leading to higher CAD activity, biosynthesis of pyrimidine nucleotides, and dysregulation of the cellular pyrimidine–purine balance. We queried repositories of tumor genomic data to determine whether the higher expression of *NAGS*, *CPS1*, and citrin genes correlate with unfavorable patient outcomes. We also investigated molecular mechanisms that could contribute to overexpression of the three genes in tumor cells, assessed differences between

somatic and germline sequence variants, and determined if there is a correlation between expression levels of the three genes in individual tumor samples.

2. Results

2.1. Aberrant Expression of *NAGS*, *CPS1*, and *Citrin* in Tumors

The Genome Data Analysis Center (GDAC) at the Broad Institute hosts processed gene expression data from The Cancer Genome Atlas (TCGA) Project and makes them available to the research community through the Firehose Data portal [37]. We queried the GDAC Firehose Data Portal for the mRNA expression of urea cycle genes in tumors and matched normal tissues. GDAC uses the Log₂ transform of RSEM values [38] to quantify abundance of mRNA. The difference in expression of urea cycle genes between tumor samples and matched normal tissues, expressed as the fold-change of the median log₂RSEM values, was available for 27 tumor types (Figure 1A). *NAGS* and *CPS1* genes were markedly overexpressed in esophageal carcinoma and kidney chromophobe samples, respectively (Figure 1A). We focused on tumors in which the median expression of *NAGS*, *CPS1*, and citrin mRNA was higher than in the matched normal tissue. The median *NAGS*, *CPS1*, and citrin expression was between 1.4- and 4-fold higher in glioblastoma multiforme (GBM, n = 166), glioma (GBMLGG, n = 696), stomach adenocarcinoma (STAD, n = 415), and stomach and esophageal carcinoma (STES, n = 600) (Figure 1A,B). *NAGS* and *CPS1* had a similar median mRNA expression in lung adenocarcinoma (LUAD, n = 517) and matched normal tissue, while citrin was overexpressed in LUAD (Figure 1A,B).

NAGS and *CPS1* have no known functions in tissues other than liver and small intestine [10,11]. The citrin gene is expressed more broadly since it functions in ureagenesis and the malate-aspartate shuttle, but the liver has the highest citrin expression [24–26]. Therefore, we evaluated the *NAGS*, *CPS1*, and citrin mRNA and protein expression in glioblastoma multiforme, glioma, lung adenocarcinoma, stomach adenocarcinoma, and stomach and esophageal carcinoma in comparison to the expression of the three genes in the liver and small intestine. The GTEx portal was queried for *NAGS*, *CPS1*, and citrin mRNA expression levels. The abundance of the *NAGS*, *CPS1*, and citrin proteins in the liver, small intestine, esophagus, stomach, brain, and lung tissues was obtained from ProteomicsDB [39,40] and the GTEx project [41]. First, we compared the mRNA and protein expression of the three genes in the liver and small intestine to determine whether they are good proxies for *NAGS*, *CPS1*, and citrin protein expression. Differences in the abundance of *NAGS*, *CPS1*, and citrin mRNA and *NAGS*, *CPS1*, and citrin proteins between the liver and small intestine are similar (Figure 1C–K). This suggests that the abundances of *NAGS*, *CPS1*, and citrin mRNA are good proxies for the abundance of the *NAGS*, *CPS1*, and citrin proteins. *NAGS* mRNA is 1–2 orders of magnitude less abundant in the stomach, esophagus, brain, and lung tissues than in the liver (Figure 1C), while the *NAGS* protein was detectable only in the lung (Figure 1D,E). Abundance of the *CPS1* mRNA and protein is 2–3 orders of magnitude lower in the esophagus, stomach, brain, and lung tissues than in the liver (Figure 1F–H). This suggests markedly lower *NAGS* and *CPS1* expression in glioblastoma multiforme, glioma, lung adenocarcinoma, stomach adenocarcinoma, and stomach and esophageal carcinoma than in the liver and small intestine. Citrin mRNA and protein expression are similar in the esophagus, stomach, brain, lung, and small intestine, and is approximately one tenth of the citrin expression levels in the liver (Figure 1I–K). Up to three-fold higher citrin expression in the tumors than in the matching normal tissue (Figure 1A,B) can affect the flux of metabolites in tumor cells.

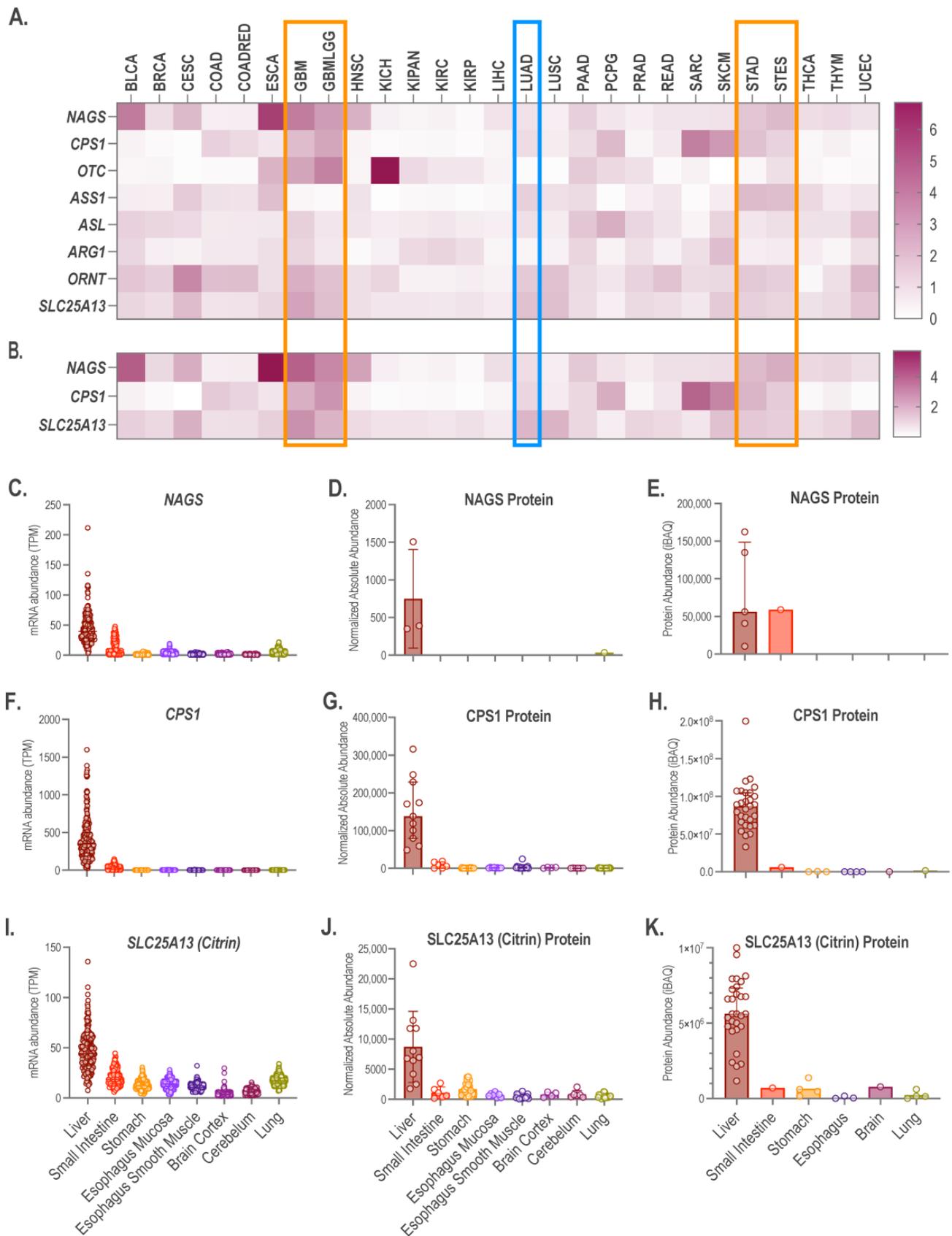


Figure 1. Expression of urea cycle genes in tumors and normal tissues. Fold-change of the median mRNA expression of all eight urea cycle genes (A) and NAGS, CPS1 and citrin mRNA (B) in tumor

samples and the matched normal tissues. Depending on the tumor type, the number of tumor samples considered here was between 36 and 1100. Orange boxes—tumors that overexpress *NAGS*, *CPS1*, and citrin. Blue box—lung adenocarcinoma overexpressing citrin. Urea cycle gene symbols: *NAGS*—N-acetylglutamate synthase; *CPS1*—carbamylphosphate synthetase; *OTC*—ornithine transcarbamylase; *ASS1*—argininosuccinate synthetase 1; *ASL*—argininosuccinate lyase; *ARG1*—arginase 1; *ORNT*—ornithine transporter encoded by the *SLC25A15* gene; *SLC25A13*—glutamate/aspartate transporter, also known as citrin. Abundance of *NAGS* (C), *CPS1* (F) and citrin (I) mRNA in the liver, small intestine, stomach, esophagus, brain, and lung tissues in GTEx project. TPM—Transcripts per kilobase per million reads. Abundance of *NAGS* (D), *CPS1* (G) and citrin (J) proteins in the liver, small intestine, stomach, esophagus, brain, and lung tissues in the GTEx project. Abundance of *NAGS* (E), *CPS1* (H) and citrin (K) in the liver, small intestine, stomach, esophagus, brain, and lung tissues reported in ProteomicsDB. iBAQ—total intensity of precursor peptide ions divided by the number of theoretically observable peptides of the protein. Tumor type symbols: BLCA—bladder urothelial carcinoma; BRCA—breast invasive carcinoma; CESC—cervical and endocervical cancers; COAD—colon adenocarcinoma; COADREAD—colorectal adenocarcinoma; GBM—glioblastoma multiforme; GBMLGG—glioma; HNSC—head and neck squamous cell carcinoma; KICH—kidney chromophobe; KIPAN—pan-kidney cohort (KICH + KIRC + KIRP); KIRC—kidney renal clear cell carcinoma; KIRP—kidney renal papillary cell carcinoma; LIHC—liver hepatocellular carcinoma; LUAD—lung adenocarcinoma; LUSC—lung small cell carcinoma; PPAD—pancreatic adenocarcinoma; PCPG—pheochromocytoma and paraganglioma; PRAD—prostate adenocarcinoma; READ—rectum adenocarcinoma; SARC—sarcoma; SKCM—skin cutaneous melanoma; STAD—stomach adenocarcinoma; STES—stomach and esophageal carcinoma; THCA—thyroid carcinoma; THYM—thymoma; UCEC—uterine corpus endometrial carcinoma.

2.2. Expression of *NAGS*, *CPS1*, and Citrin in Tumors and Patient Outcomes

We queried the cBioPortal database [42,43] to determine whether the expression of *NAGS*, *CPS1*, and citrin in glioblastoma multiforme, glioma, stomach adenocarcinoma, and lung adenocarcinoma correlated with the patient outcomes for these four tumor types. Patient outcome data for stomach and esophageal carcinoma are not available in the cBioPortal; cBioPortal does have the outcome data for patients with glioblastoma, which is a type of glioma. Therefore, we queried cBioPortal for the association between *NAGS*, *CPS1*, and citrin mRNA expression levels and outcomes of patients with glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma. Glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples were ranked according to expression level of *NAGS*, *CPS1*, or citrin mRNA and divided into quartiles based on the expression levels of the three genes, followed by comparison of survival times of patients with the highest and the lowest expression of the three genes (Figure 2). The highest *NAGS* mRNA expression was associated with a worse outcome, i.e., shorter survival time, in patients with glioblastoma (Figure 2A). There was no association between *CPS1* and citrin mRNA expression levels and the outcome of patients with glioblastoma (Figure 2B,C). There was a trend ($p = 0.07$) towards a worse outcome (shorter survival time) for patients with glioblastoma multiforme exhibiting the highest expression of *NAGS* mRNA, compared to glioblastoma multiforme patients exhibiting the lowest expression of *NAGS* mRNA (Figure 2D). Expression levels of *CPS1* and citrin mRNA were not associated with the outcome of patients with glioblastoma multiforme (Figure 2E,F). There was a trend towards a worse outcome ($p = 0.07$) in patients exhibiting the lowest quartile of *NAGS* mRNA expression in lung adenocarcinoma (Figure 2G), while the outcome was worse for patients with the highest *CPS1* mRNA expression in lung adenocarcinoma (Figure 2H). There was no association between citrin mRNA expression and outcome of patients with lung adenocarcinoma (Figure 2I), and there was no association between expression levels of *NAGS*, *CPS1*, or citrin mRNA and outcome of patients with stomach adenocarcinoma (Figure 2J–L).

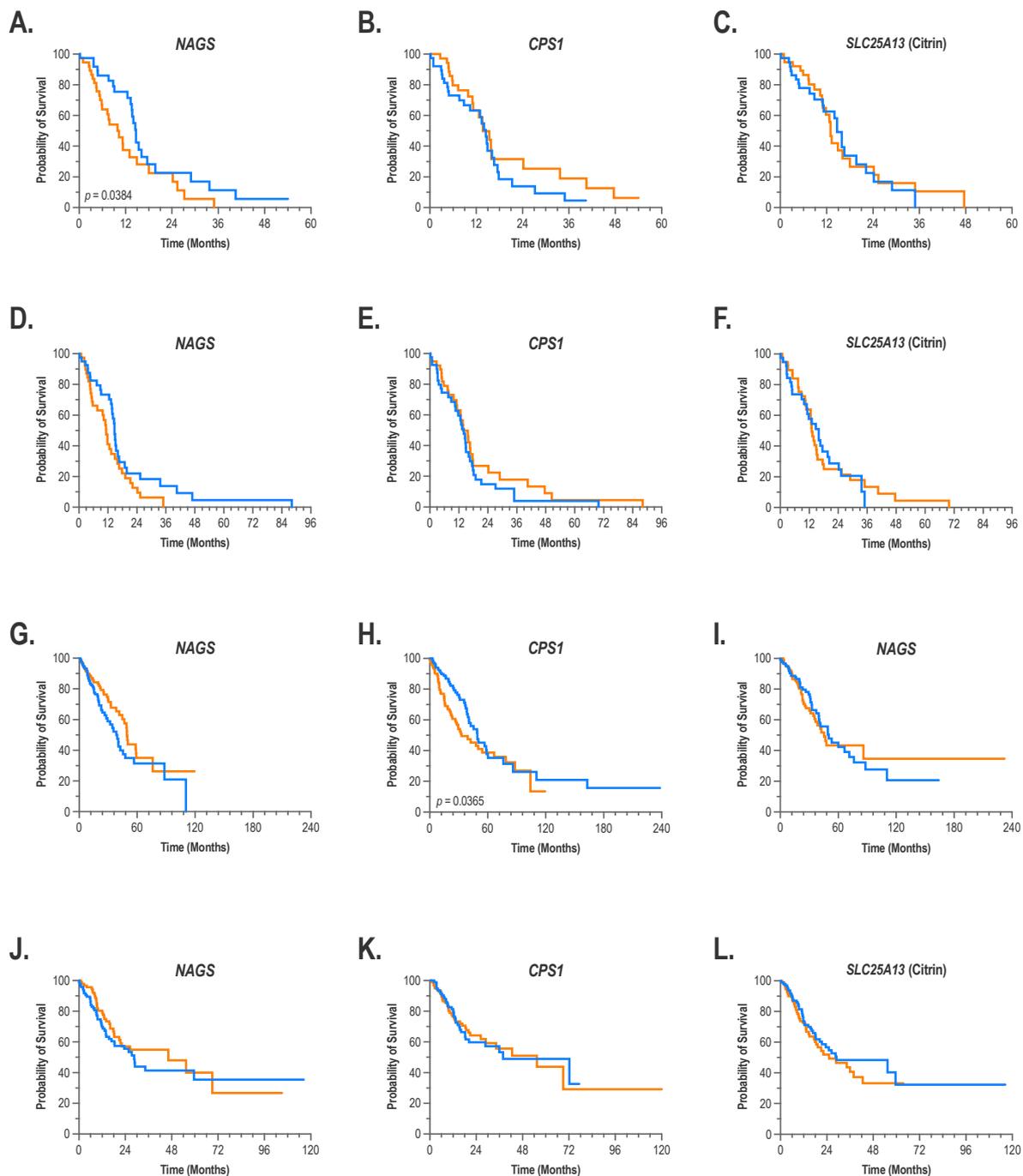


Figure 2. Association between expression of *NAGS*, *CPS1*, and citrin mRNA in tumors and patient outcomes. Kaplan–Meier curves showing survival of patients with glioblastoma (A–C), glioblastoma multiforme (D–F), lung adenocarcinoma (G–I) and stomach adenocarcinoma (J–L), exhibiting highest and lowest levels of *NAGS*, *CPS1*, and citrin mRNA expression. Orange—tumors in the highest quartile of mRNA expression; blue—tumors in the lowest quartile of mRNA expression.

2.3. Epigenetic Regulation of *NAGS*, *CPS1*, and Citrin in Lung Adenocarcinoma

Overexpression of *NAGS* and *CPS1* genes in the brain, lungs, and stomach tissues—where these two genes are not highly expressed—would require epigenetic changes that allow the transcription of the two genes. The citrin gene is expressed in most tissues due to its functioning in the malate-aspartate shuttle [26]. Therefore, the epigenetic regulation of citrin ought to be similar in the liver, brain, lung and stomach. We sought to compare

chromatin accessibility to transcription factors, histone modifications, binding of CTCF (a component of the cohesin complex that regulates chromatin domain organization), and a POLR2A subunit of RNA polymerase II as indicators of molecular mechanisms that enable gene expression in the liver, where *NAGS*, *CPS1*, and citrin are highly expressed; brain, lung and stomach tissues; and cancer cell lines that model glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma. DNase-Seq and ATAC-Seq data, which indicate chromatin accessibility to transcription factors; ChIP-Seq data for CTCF; POLR2A; H3K4me3, which indicate active gene transcription; and H3K27ac, which indicates active enhancers, were available in the ENCODE database for the liver and lung tissues, and for the A549 lung adenocarcinoma cell line.

The *NAGS* promoter, −3 kb enhancer, and a regulatory element in the first intron of the *NAGS* gene regulate its expression in the liver [44,45] and are among candidate *cis*-regulatory elements (cCRE) of the *NAGS* gene (Figure 3A–C, highlighted in green). DNase-Seq and ATAC-Seq data indicate that the *NAGS* promoter and −3 kb enhancer are accessible to transcription factor binding in the liver (Figure 3A), but not in the A549 cell line and lung tissue (Figure 3B,C). An ATAC-Seq peak located approx. 9.5 kb upstream of the *NAGS* transcription start site (TSS) coincides with CTCF binding in the liver, A549 cells, and lung, suggesting that this region may represent a boundary between the *NAGS* locus and the upstream *PYY* locus (Figure 3A–C). The −3 kb *NAGS* enhancer and intronic regulatory elements are associated with strong H3K27ac signals in the liver, indicating that these two regulatory elements are active enhancers in the liver (Figure 3A). The H3K27ac signal is absent at the −3 kb enhancer in A549 cells and lung tissue (Figure 3B,C), consistent with the liver-specific activity of this regulatory element [45]. The H3K27ac signal at the *NAGS* intronic regulatory element is markedly weaker in A549 cells and lung tissue, indicating low *NAGS* expression (Figure 3B,C). The H3K4me3 and POLR2A signals were lower in A549 cells and lung tissue than in the liver (Figure 3A–C). Taken together, these data are consistent with low *NAGS* expression in the A549 cells, lung adenocarcinoma, and lung tissue.

Expression of *CPS1* is regulated by a proximal enhancer and promoter located immediately upstream of *CPS1* TSS, and a distal enhancer located approx. 7.5 kb upstream of the *CPS1* TSS [45,46]. These known *CPS1* regulatory elements are associated with strong ATAC-Seq and DNase-Seq signals in the liver (Figures 3D and S1, highlighted in green). Upstream of the *CPS1* distal enhancer are several cCREs, of which, some are associated with ATAC-Seq and DNase-Seq signals in the liver (Figures 3D and S1A). The distal *CPS1* enhancer is associated with strong ATAC-Seq and DNase-Seq signals in A549 cells (Figures 3E and S1B), but not in the lung (Figures 3F and S1C). The other known and predicted *CPS1* upstream regulatory elements are associated with weak ATAC-Seq and DNase-Seq signals in the lung and A549 cells (Figures 3E,F and S1). Approximately 35 kb upstream of the *CPS1* TSS is a site that binds CTCF in the liver and is predicted to be a chromatin insulator (Figures 3D and S1A). This predicted chromatin insulator does not bind CTCF in the lung and A549 cells (Figure 3E,F). Known *CPS1* regulatory elements and some of the predicted *CPS1* regulatory elements are associated with strong H3K27ac, H3K4me3, and POLR2A signals in the liver (Figure 3D), which are weak in A549 cells (Figure 3E) and absent in the lung tissue (Figure 3F). This is consistent with aberrant *CPS1* expression in lung adenocarcinoma.

Expression of the citrin gene is regulated by a recently characterized promoter [47,48]. This promoter is associated with strong ATAC-Seq, DNase-Seq, H3K27ac, H3K4me3, and POLR2A signals in the liver, A549 cells, and lung tissue, consistent with ubiquitous citrin gene expression (Figures 3G–I and S2). ATAC-Seq signals were associated with some of the cCREs in the first intron of the citrin gene and its upstream regulatory region (Figures 3G–I and S2). CTCF binding was not detected in the first intron of the citrin gene and within 25 kb upstream of the citrin TSS (Figure 3G–I). However, CTCF binding sites were present approx. 100 kb upstream of the citrin TSS and within intron 4 of the citrin gene (Figure S2). Similar epigenetic regulation of citrin in the liver and lung is consistent

with its expression in many tissues, and could contribute to its aberrant expression in lung adenocarcinoma.

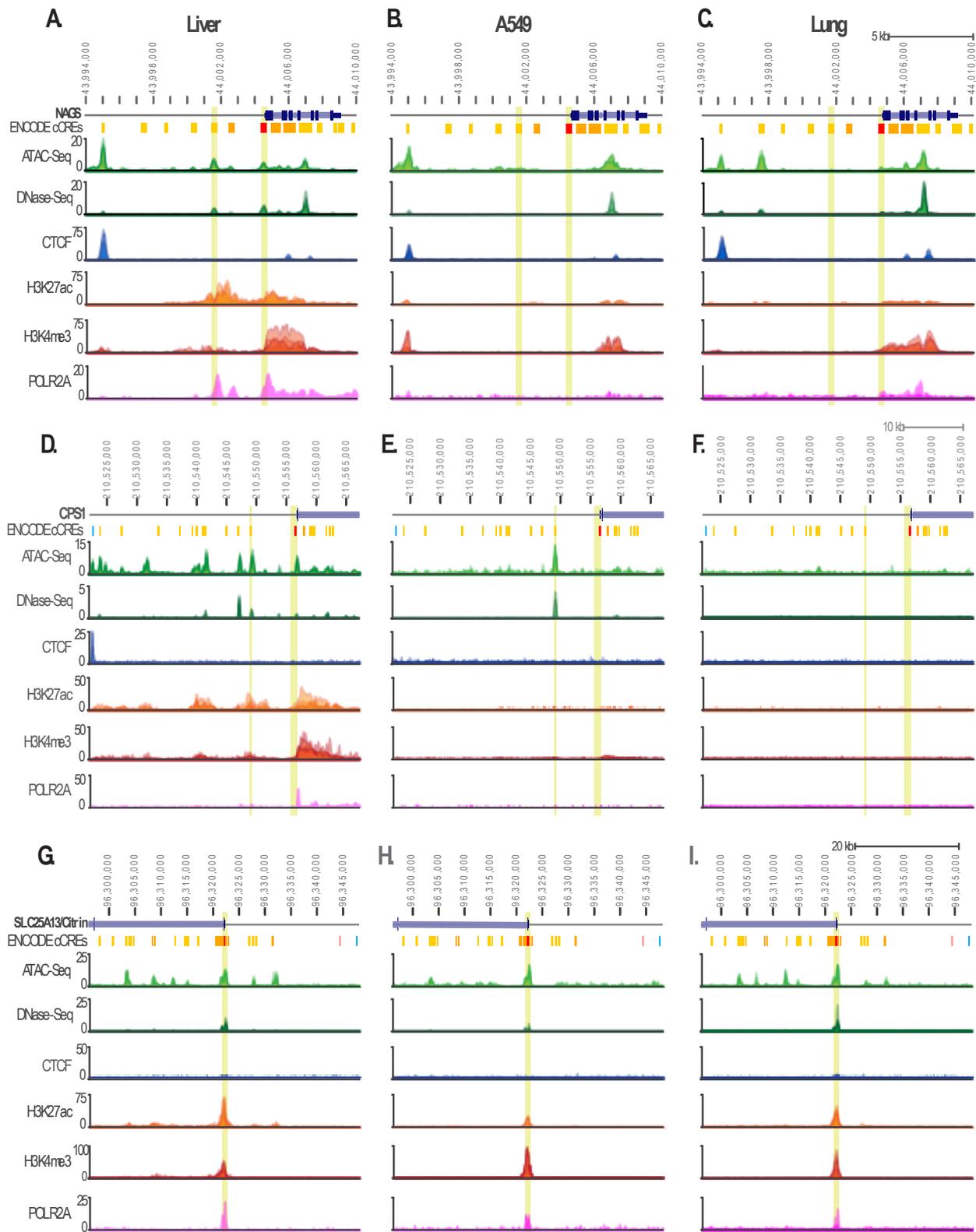


Figure 3. Epigenetic regulation of *NAGS* (A–C), *CPS1* (D–F), and citrin (G–I) in the liver tissue (A,D,G), A549 cell line (B,E,H), and lung tissue (C,F,I). Chromatin accessibility (ATAC-Seq and DNase-Seq shown in light and dark green, respectively), CTCF binding (blue), acetylation of histone

H3 lysine 27 (H3K27ac, shown in light orange), tri-methylation of histone H3 lysine 4 (H3K4me3, shown in dark orange), and binding of the 2A subunit of RNA polymerase II (POLR2A, shown in magenta) are shown for the following genomic regions: whole *NAGS* gene, approx. 10.5 kb of the upstream regulatory region and 1 kb downstream of the *NAGS* (chr17: 43,994,000–44,010,000); first exon of the *CPS1* gene, approx. 35 kb of the upstream regulatory region and approx. 10 kb of the first *CPS1* intron (chr2: 210,522,000–210,567,000); first and second exons of the citrin gene, first intron of the citrin gene, and approx. 25 kb of the upstream regulatory region (chr7: 96,296,000–96,348,000). Predicted cCREs: promoters–red, proximal enhancers–orange, distal enhancers–yellow, chromatin insulators–blue.

2.4. Sequence Variation and Expression in Tumor and Normal Tissues

Since *NAGS*, *CPS1*, or citrin genes are not highly expressed in the brain, lung, and stomach, we explored molecular mechanisms that can contribute to elevated *NAGS*, *CPS1*, and citrin activities in tumors. Increased copy numbers of *NAGS*, *CPS1*, and citrin genes and/or activating sequence variants could result in higher activity of the three proteins in glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma.

2.4.1. Copy Number Variation of *NAGS*, *CPS1*, and Citrin Genes in Glioblastoma, Glioblastoma Multiforme, Lung Adenocarcinoma and Stomach Adenocarcinoma

Structural variants resulting in increased copy numbers of *NAGS*, *CPS1*, and/or citrin loci could explain the overexpression of the three genes. Therefore, we queried cBioPortal for the number of glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples exhibiting *NAGS*, *CPS1*, and citrin gene copy number variation (CNV), and whether there is an association between CNVs and mRNA expression of the three genes in the four tumor types (Figure 4). CNVs are classified as deep deletions, shallow deletions, diploid, gain, and amplifications in cBioPortal, and they correspond to 0, 1, 2, a few additional copies, and more than a few additional copies of the gene/region of interest, respectively [49].

Glioblastoma and glioblastoma multiforme had similar patterns of *NAGS*, *CPS1*, and citrin CNV (Figure 4). *NAGS* and *CPS1* loci were diploid in the 82 and 89%, respectively, of glioblastoma samples while 5–10% of glioblastoma samples exhibited either shallow deletions or gains of *NAGS* and *CPS1* genes (Figure 4A). Most of the glioblastoma samples (80%) exhibited gains of the citrin gene (Figure 4A). The citrin gene was diploid in 18% of glioblastoma samples (Figure 4A). The remaining glioblastoma samples had amplifications of the citrin gene (Figure 4A). There was no association between CNVs and the expression of *NAGS* and *CPS1* mRNA in glioblastoma (Figure 4B,C). The expression of citrin mRNA was higher in glioblastoma samples with a higher copy number of the citrin gene (Figure 4D). The majority of glioblastoma multiforme samples were diploid for *NAGS* and *CPS1* loci and exhibited gains of the citrin gene (Figure 4E). The remaining glioblastoma multiforme samples exhibited either shallow deletions or gains of the *NAGS* and *CPS1* genes, and either gains or amplifications of the citrin gene (Figure 4E). There was no association between CNVs and the expression of *NAGS* and *CPS1* mRNA in glioblastoma multiforme samples (Figure 4F,G). The expression of citrin mRNA was higher in the glioblastoma multiforme samples with a higher copy number of the citrin gene (Figure 4H).

More than 90% of the lung adenocarcinoma samples were diploid for *NAGS*, *CPS1*, and citrin genes (Figure 4I). The remaining lung adenocarcinoma samples exhibited deep and shallow deletions, gains, and amplifications of the three genes (Figure 4I–L). The expression of *NAGS*, *CPS1*, and citrin mRNA correlated with the CNV of each gene (Figure 4J–L).

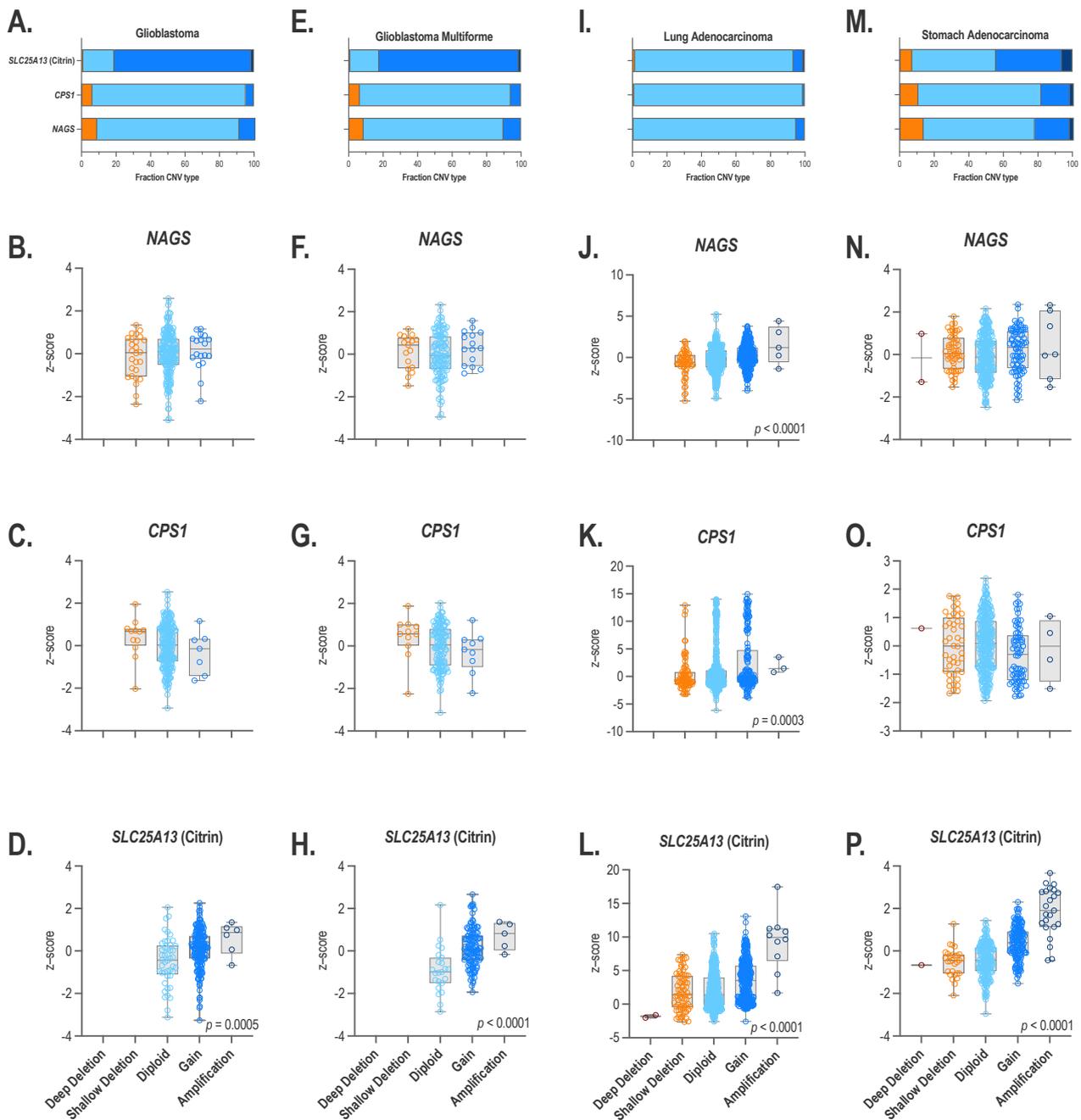


Figure 4. Copy number variation and *NAGS*, *CPS1*, and citrin mRNA expression. Fraction of glioblastoma (A), glioblastoma multiforme (E), lung adenocarcinoma (I) and stomach adenocarcinoma (M) samples with deep deletions, shallow deletions, two copies, gain and amplification of *NAGS*, *CPS1*, and citrin genes. Relationship between copy number and *NAGS* (B,F,J,N), *CPS1* (C,G,K,O) and citrin (D,H,L,P) mRNA expression. Brown—deep deletion, orange—shallow deletion, light blue—diploid, medium blue—gain, dark blue—amplification of the genomic regions of interest.

The CNV of *NAGS* and *CPS1* were similar in stomach adenocarcinoma samples. The majority of the stomach adenocarcinoma samples, 64% and 71%, were diploid for *NAGS* and *CPS1* genes, respectively; 20% and 16% of the stomach adenocarcinoma samples exhibited gains of *NAGS* and *CPS1*, respectively; 13% and 10% of stomach adenocarcinoma samples exhibited shallow deletions of *NAGS* and *CPS1*, respectively; and the remaining stomach adenocarcinoma samples had either deep deletions or amplifications of the two genes (Figure 4M). The majority of the stomach adenocarcinoma samples were either diploid

(48%) or exhibited a gain (38%) of the citrin gene; the remaining samples exhibited shallow deletions, amplifications, and a loss of the citrin gene (Figure 4M). There was no association between the CNV and *NAGS* mRNA expression in the stomach adenocarcinoma samples (Figure 4N). The expression of *CPS1* mRNA was lower in the stomach adenocarcinoma samples with higher copy numbers of the *CPS1* gene (Figure 4O). The expression of citrin mRNA was higher in the stomach adenocarcinoma samples with higher copy number of the citrin gene (Figure 4P).

2.4.2. Germline and Somatic *NAGS*, *CPS1*, and Citrin Sequence Variants

Metabolic dysregulation contributing to increased de novo pyrimidine biosynthesis could result from sequence variants that increase the activity of *NAGS*, *CPS1*, and citrin without the need for the overexpression of the corresponding genes. Therefore, we intended to compare functional effects of *NAGS*, *CPS1*, and citrin somatic sequence variants found in glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma with functional effects of germline variants found in patients with respective urea cycle disorders, and in individuals without known rare diseases. However, the fraction of glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples with sequence variants in *NAGS*, *CPS1*, and citrin was low (Table 1). Therefore, we compared the types of *NAGS*, *CPS1*, and citrin sequence variants found in all tumor samples with sequence variants found in individuals without known rare diseases, and patients with *NAGS*, *CPS1*, or citrin deficiencies to increase the power of comparison.

Table 1. Number and fraction of glioblastoma, glioblastoma multiforme (GBM), lung adenocarcinoma (LUAD), and stomach adenocarcinoma (STAD) samples with sequence variants in *NAGS*, *CPS1*, and citrin genes.

	Glioblastoma and GBM	LUAD	STAD
<i>NAGS</i>	7 ^a /2176 ^b (0.32%) ^c	11/1737 (0.63%)	6/1208 (0.50%)
<i>CPS1</i>	41/2660 (1.54%)	229/2313 (9.90%)	74/1261 (5.87%)
Citrin	42/1471 (2.86%)	44/2153 (2.04%)	31/1240 (2.50%)

^a Number of tumor samples with SNVs and short InDels in COSMIC and cBioPortal. ^b Total number of analyzed tumor samples in COSMIC and cBioPortal. ^c Percent of tumor samples with SNVs and Indels.

NAGS, *CPS1*, and citrin sequence variants found in tumor samples were collected from the TCGA [50], COSMIC [51], and cBioPortal [42,43] databases of tumor genomic information. *NAGS*, *CPS1*, and citrin sequence variants present in individuals without rare genetic disorders were collected from the gnomAD database [52]. Pathogenic, likely pathogenic sequence variants, variants of uncertain significance, and variants with conflicting interpretation found in patients with *NAGS*, *CPS1*, or citrin deficiency were collected from ClinVar and LOVD [53] databases as well as from published case reports (Figure 5A). For all three genes, there was little overlap between variants found in tumors, patients, and gnomAD (Figure 5B). This is not surprising since mechanisms of mutagenesis differ in germline and tumor cells that often exhibit aberrant mismatch repair and proofreading [54].

Although there was little overlap between sequence variants found in patients with *NAGS*, *CPS1*, or citrin deficiencies, tumors and gnomAD, the frequencies of different types of sequence variants were similar in the three groups of samples (Table 2 and Figure 5C). Variants were classified based on their location in genes (regulatory, 5'-UTR, intron, and 3'-UTR) and based on their effect on the coding region and splicing (missense, nonsense, frameshift, synonymous, splicing variants of canonical splice sites, and splice region variants that affect base pairs flanking canonical splice sites).

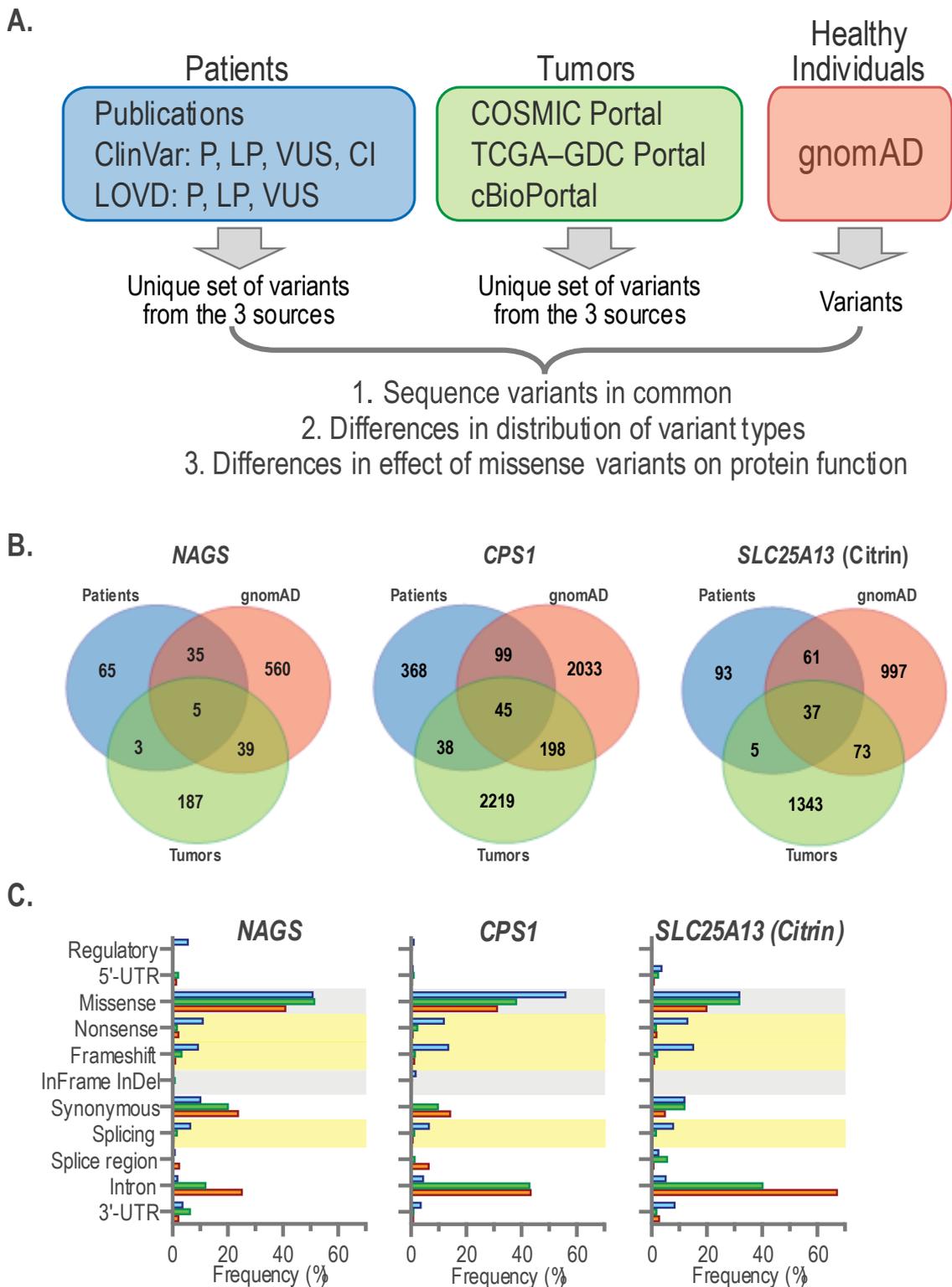


Figure 5. *NAGS*, *CPS1*, and citrin single nucleotide variants and small indels in urea cycle patients, tumors, and individuals without rare genetic diseases. **(A)** Strategy used to collect and analyze sequence variants found in the three groups of samples. **(B)** Venn diagrams showing overlap between *NAGS*, *CPS1*, or citrin sequence variants found in patients with respective deficiencies, tumors and gnomAD. **(C)** Distribution of *NAGS*, *CPS1*, and citrin variant types and their functional effects in patients (blue), tumors (green) and gnomAD (orange).

Table 2. Number and type of sequence variants found in *NAGS*, *CPS1*, and citrin genes in patients with *NAGS*, *CPS1*, or citrin deficiency, tumor samples, and individuals without rare genetic disorders.

<i>NAGS</i>	Patients Count (%)	Tumors Count (%)	GnomAD Count (%)
Regulatory	6 (5.6)	0 (0.0)	0 (0.0)
5'-UTR	0 (0.0)	5 (2.1)	9 (1.4)
Missense	55 (50.9)	120 (51.5)	262 (41.0)
Nonsense	12 (11.1)	4 (1.7)	14 (2.2)
Frameshift	10 (9.3)	8 (3.4)	7 (1.1)
InFrame InDel	0 (0.0)	2 (0.9)	1 (0.2)
Synonymous	11 (10.2)	47 (20.2)	152 (23.8)
Splicing	7 (6.5)	4 (1.7)	3 (0.5)
Splice region	1 (0.9)	0 (0.0)	16 (2.5)
Intron	2 (1.9)	28 (12.0)	161 (25.2)
3'-UTR	4 (3.7)	15 (6.4)	14 (2.2)
Total	108	233	639
<i>CPS1</i>	Patients	Tumors	gnomAD
Regulatory	6 (1.1)	0 (0.0)	0 (0.0)
5'-UTR	4 (0.7)	28 (1.1)	13 (0.5)
Missense	312 (56.0)	956 (38.2)	743 (31.3)
Nonsense	67 (12.0)	61 (2.4)	16 (0.7)
Frameshift	76 (13.6)	38 (1.5)	30 (1.3)
InFrame InDel	10 (1.8)	4 (0.2)	8 (0.3)
Synonymous	0 (0.0)	244 (9.8)	340 (14.3)
Splicing	37 (6.6)	33 (1.3)	17 (0.7)
Splice region	0 (0.0)	35 (1.4)	155 (6.5)
Intron	25 (4.5)	1076 (43.0)	1029 (43.4)
3'-UTR	20 (3.6)	25 (1.0)	22 (0.9)
Total	557	2500	2373
<i>SLC25A13</i> (Citrin)	Patients	Tumors	gnomAD
Regulatory	0 (0.0)	0 (0.0)	0 (0.0)
5'-UTR	7 (3.7)	29 (2.5)	13 (0.9)
Missense	61 (31.9)	372 (31.9)	291 (20.0)
Nonsense	25 (13.1)	20 (1.7)	28 (1.9)
Frameshift	29 (15.2)	24 (2.1)	14 (1.0)
InFrame InDel	0 (0.0)	3 (0.3)	1 (0.1)
Synonymous	23 (12.0)	140 (12.0)	71 (4.9)
Splicing	15 (7.9)	20 (1.7)	8 (0.5)
Splice region	5 (2.6)	66 (5.7)	11 (0.8)
Intron	10 (5.2)	470 (40.3)	980 (67.2)
3'-UTR	16 (8.4)	22 (1.9)	41 (2.8)
Unknown	191	1166	1458

Loss-of-function (LOF) variants that include nonsense, frameshift, and splicing variants were more frequent in patients with *NAGS*, *CPS1*, or citrin deficiency than in tumors and gnomAD (Figure 5C and Table 2). A higher frequency of LOF variants in patients with *NAGS*, *CPS1*, or citrin deficiency than in tumors and gnomAD is expected, since carriers of LOF variants in any of the three genes have normal ureagenesis and phenotype [10]. Very small fractions of sequence variants were found in the 5'-UTRs, splice regions and 3'-UTRs of all three genes (Figure 5C and Table 2). This is not surprising since these regions are minor portions of *NAGS*, *CPS1*, and citrin genes. Sequence variants found in *NAGS*, *CPS1*, and citrin introns were common in tumors and gnomAD (Figure 5C and Table 2). This is likely due to the whole genome sequencing of DNA from tumor samples and gnomAD, while sequencing of patient DNA focuses on the coding regions and canonical splice

sites where most disease-causing sequence variants are found. Both absolute and relative numbers of sequence variants found in *CPS1* and citrin introns were higher than in *NAGS* introns (Figure 5C and Table 2), likely due to the small size of *NAGS* introns compared to *CPS1* and citrin introns. Synonymous *NAGS* and *CPS1* sequence variants were more frequent in tumor samples and gnomAD than in patients with *NAGS* or *CPS1* deficiency, while citrin synonymous sequence variants were less frequent in gnomAD than in patients and tumor samples (Figure 5C and Table 2). Surprisingly, whole genome sequencing of DNA from tumor samples and gnomAD did not reveal many variants in the upstream regulatory regions (Figure 5C and Table 2). This could be due to the poor annotation of gene regulatory regions in the current assembly of the human genome. Missense variants found in all three genes were either the largest or the second largest fraction of all sequence variants in patients, tumors, and gnomAD (Figure 5C and Table 2). Similar fractions of *NAGS* and citrin variants found in patients and tumor samples were missense variants, while they represented a slightly lower fraction of *NAGS* and citrin variants in gnomAD (Figure 5C and Table 2). This pattern was different for *CPS1* missense variants. Similar fractions of *CPS1* variants found in tumor samples and gnomAD were missense variants, while they represented the highest fraction of *CPS1* variants found in patients with *CPS1* deficiency (Figure 5C and Table 2).

There was more overlap between missense variants found in patients and gnomAD than between somatic missense variants found in tumors and germline variants found in patients and gnomAD (Figure 6A). The REVEL functional effect predictor was used to evaluate the effects of missense variants on protein function. Missense variants with a REVEL score above 0.5 are considered damaging while missense variants with a REVEL score below 0.5 are considered tolerated [55]. The median REVEL scores of *NAGS*, *CPS1*, and citrin missense variants found in patients were all above 0.5 (Figure 6B–D). This is consistent with the reduced or absent enzymatic activity and/or decreased stability of mutant *NAGS*, *CPS1*, and citrin proteins found in patients with respective deficiencies.

The median REVEL scores of *NAGS* missense variants found in tumors and gnomAD are below 0.5, suggesting that most of these variants do not affect *NAGS* protein function (Figure 6B). REVEL scores of *CPS1* and citrin missense variants found in tumors and gnomAD are above 0.5 but lower than the median REVEL score in patients (Figure 6C,D). This likely reflects a higher conservation of *CPS1* and citrin proteins across phyla [56] and the weight given to amino acid conservation by REVEL [55].

The strength of evidence for damaging and tolerated/benign effects of a missense variant on protein function can be inferred from its REVEL score [57]. The distribution of *NAGS* REVEL scores differs in patients vs. gnomAD and tumor samples. REVEL scores for the majority of *NAGS* missense variants in all three groups of samples indicate an uncertain effect on *NAGS* function. The second highest fraction (29%) of *NAGS* missense variants have REVEL scores indicative of moderate support for the damaging effect on *NAGS* function (Figure 6E).

The majority of *CPS1* missense variants found in patients with *CPS1* deficiency (40%) have REVEL scores higher than 0.932, suggesting strong evidence for damaging effects on *CPS1* function (Figure 6F). Between 25 and 30% of sequence variants in all three groups of samples have REVEL scores that indicate moderate evidence for damaging effects on *CPS1* function (Figure 6F). The majority of *CPS1* missense variants in gnomAD and tumor samples have REVEL scores that indicate an uncertain effect on *CPS1* function (Figure 6F).

Two-thirds of citrin missense variants found in patients with a citrin deficiency have REVEL scores that suggest either a moderate or uncertain effect on citrin function (Figure 6G). The majority of citrin missense variants in gnomAD and tumor samples have REVEL scores that indicate an uncertain effect on citrin function (Figure 6F). None of the missense variants found in *NAGS*, *CPS1*, and citrin genes had REVEL scores that indicate either strong or very strong evidence for a benign effect on protein function.

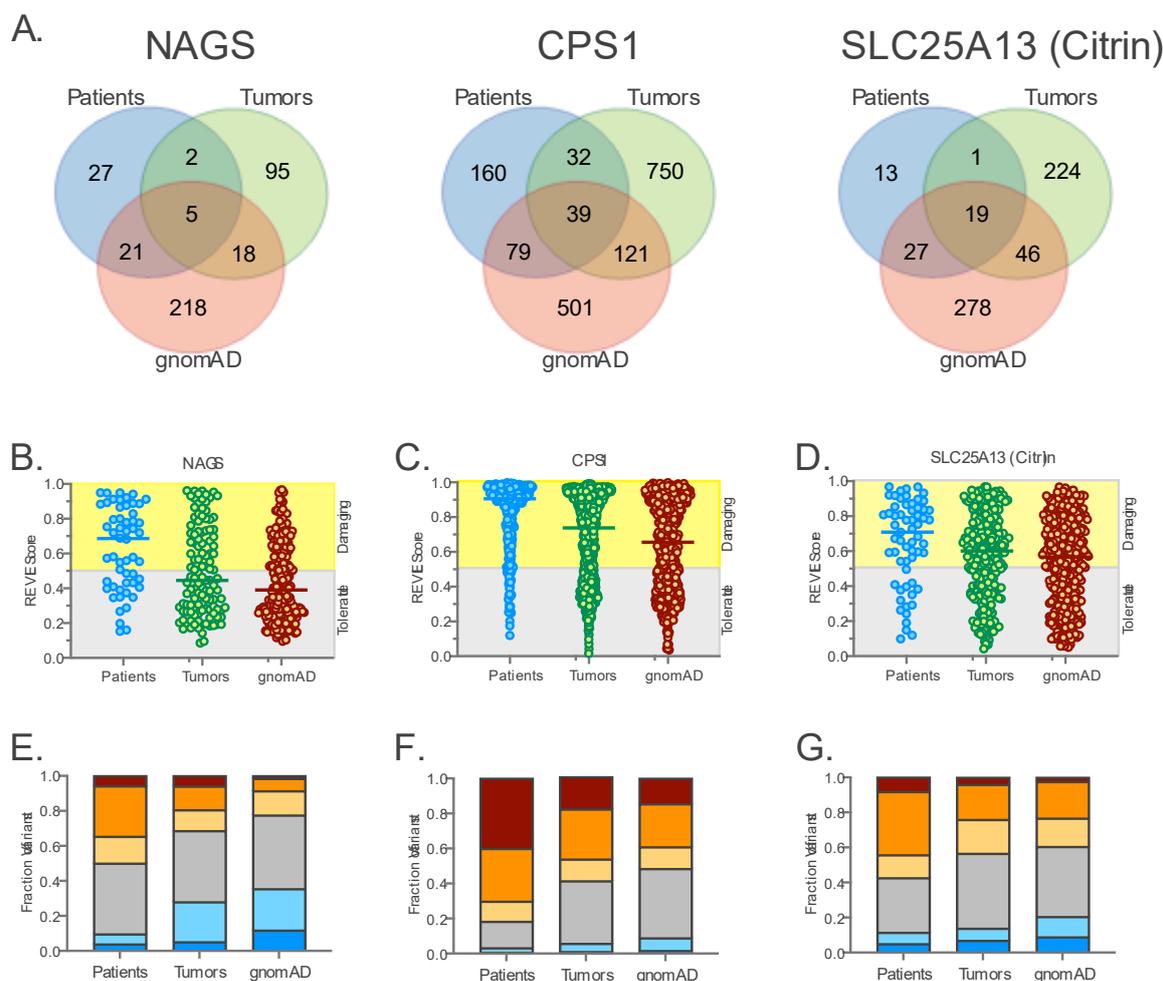


Figure 6. *NAGS*, *CPS1*, and citrin missense variants in urea cycle patients, tumors, and individuals without rare genetic disorders. (A) Venn diagrams showing overlap between *NAGS*, *CPS1*, or citrin missense variants found in patients with respective deficiencies, tumors, and gnomAD. (B–D) REVEL scores of *NAGS* (B), *CPS1* (C) and citrin (D) missense variants found in patients (blue), tumors (green), and gnomAD (orange). Yellow—REVEL scores that indicate damaging effect of missense variants on protein function; gray—REVEL scores indicating that missense variants are tolerated and do not affect protein function. (E–G) Fraction of sequence variants with REVEL scores indicating strength of evidence for either damaging or benign effect of missense variant on *NAGS* (E), *CPS1* (F) and citrin (G) function in patients with respective deficiencies, tumor samples, and gnomAD. Brown—strong evidence for damaging effect on protein function; orange—moderate evidence for damaging effect on protein function; light orange—supporting evidence for damaging effect on protein function; gray—uncertain effect on protein function; light blue—supporting evidence for benign effect on protein function; dark blue—moderate evidence for benign effect on protein function.

2.5. *NAGS*, *CPS1*, and Citrin Expression in Individual Tumor Samples

NAG, produced by *NAGS*, is required for *CPS1* function in vivo [11–18,58,59]. Increased production of carbamylphosphate (CP) due to the overexpression of *CPS1*, together with higher cytoplasmic abundance of aspartate due to the overexpression of citrin could contribute to a higher CAD activity and de novo pyrimidine synthesis. Therefore, we examined whether individual glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples exhibit the overexpression of *NAGS*, *CPS1*, and citrin mRNA. The correlation between *NAGS* and *CPS1*, citrin and *NAGS*, and *CPS1* and citrin mRNA expression in individual tumor samples was either weak or negligible (Figure 7).

CPS1 and citrin protein abundance data were available for individual glioblastoma and lung adenocarcinoma samples [60,61]. The correlation between CPS1 and citrin protein abundance was weak in individual glioblastoma and lung adenocarcinoma samples (Figure 8). These results suggest that an increased cytoplasmic supply of either CP or aspartate may be sufficient for a higher de novo pyrimidine biosynthesis, as well as the need for an alternative explanation of CPS1 activity in the absence of NAGS expression and NAG.

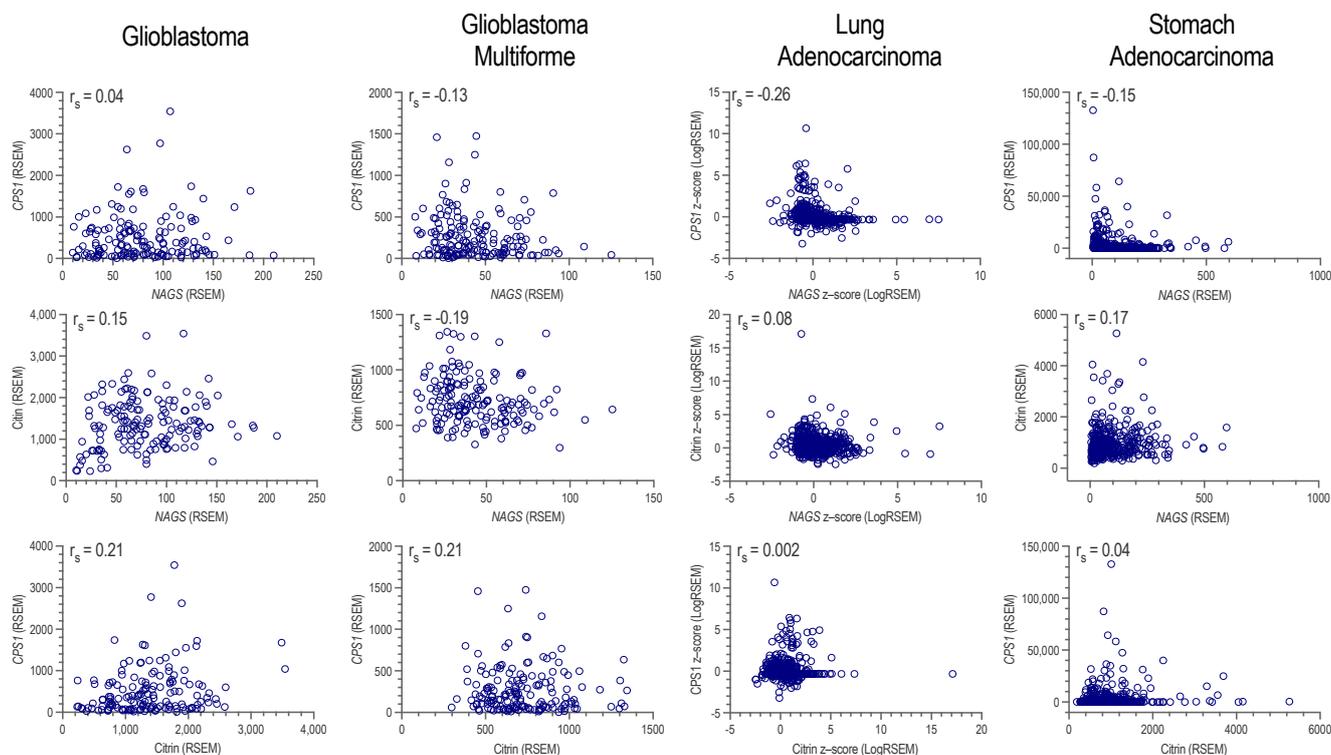


Figure 7. Correlation between *NAGS* and *CPS1* mRNA expression (**top**), *NAGS* and citrin mRNA expression (**middle**), and citrin and *CPS1* mRNA expression (**bottom**) in individual glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach glioblastoma samples. Spearman correlation coefficient was calculated to infer the relationship between mRNA expression levels in individual tumor samples.

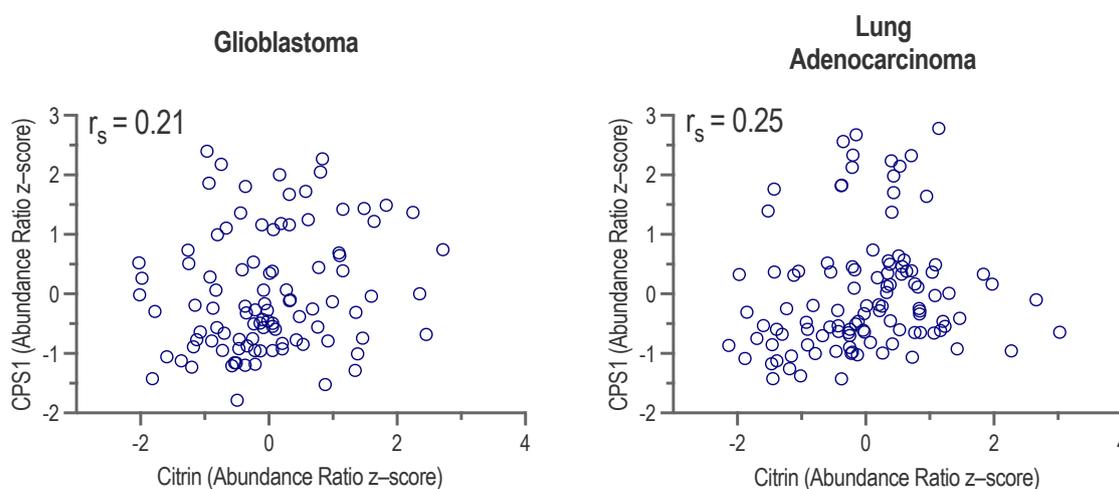


Figure 8. Correlation between citrin and CPS1 protein abundance in individual glioblastoma (**left**) and lung adenocarcinoma (**right**) samples. Spearman correlation coefficient was calculated to infer the relationship between protein abundance in individual tumor samples.

3. Discussion

This study used data mining to assess whether the aberrant expression of *NAGS* and biosynthesis of NAG can explain the increased activity of *CPS1* that contributes to higher de novo pyrimidine production and tumor cell proliferation. We also assessed whether tumors that overexpress *NAGS* and *CPS1* exhibit a higher expression of citrin, which can increase the cytoplasmic supply of aspartate for the de novo biosynthesis of pyrimidine nucleotides. The project was conceived and carried out during the COVID-19 pandemic when access to laboratory research was limited. We asked a series of questions that could be answered using data mining and bioinformatic approaches and used the answers to generate experimentally testable hypotheses.

The urea cycle can catalyze the formation of approximately 20 g of urea nitrogen, which corresponds to 43 g of urea per day in an average man [62]. Urea cycle enzymes and transporters are among the most abundant proteins in periportal hepatocytes [63]. The abundances of *NAGS*, *CPS1*, and citrin mRNA and proteins in the lung, brain, esophagus, and stomach tissues are 10–1000 times lower than in the liver, and the expression of the three genes is two- to four-fold higher in glioblastoma, glioblastoma multiforme, lung adenocarcinoma, stomach adenocarcinoma, and stomach and esophagus carcinoma samples. Although markedly lower than in the liver, the expression of *CPS1* in lung cancers and p53-depleted tumor cell lines can sufficiently alter cellular homeostasis towards the increased de novo biosynthesis of pyrimidine nucleotides via a higher CAD activity [31,32,34]. The *NAGS* protein is expressed in two cell lines commonly used to model non-small cell lung carcinoma [36]; this provides a molecular mechanism for *CPS1* activity and increased production of CP in the two cell lines. Since *CPS1* is a mitochondrial enzyme, CP produced by *CPS1* in the mitochondria of tumor cells would have to be translocated into the cytoplasm to enter de novo pyrimidine biosynthesis. Excess CP, produced in hepatocytes of patients with OTC deficiency, can exit mitochondria and enter pyrimidine biosynthesis leading to the accumulation of orotic acid, which is an intermediate of pyrimidine biosynthesis and a biomarker of OTC deficiency [10,11].

ARALAR1 is a citrin paralog with identical biochemical function [22]. ARALAR1 is highly expressed in the brain, skeletal muscle and heart, present in other tissues and absent from the liver, the site of high citrin expression [22]. An increased supply of cytoplasmic aspartate due to the ectopic overexpression of citrin could disrupt the balance of cytoplasmic metabolites, resulting in a higher CAD activity, higher de novo pyrimidine biosynthesis, and metabolic reprogramming that leads to increased cell proliferation. In addition to supplying cytoplasmic aspartate for increased CAD activity, the overexpression of citrin in tumor cells could lead to metabolic reprogramming in tumor cells through the disruption of the malate-aspartate shuttle and energy production [29,30]. Therefore, we hypothesize that the overexpression of *NAGS*, *CPS1*, and citrin promotes the proliferation of glioblastoma, glioblastoma multiforme, stomach adenocarcinoma, and stomach and esophagus carcinoma cells through the increased biosynthesis of pyrimidine nucleotides and/or dysregulated malate-aspartate shuttle. This hypothesis could be tested in glioblastoma, stomach, and esophageal cancer cell lines by determining whether they express *NAGS*, *CPS1*, and/or citrin genes and proteins, followed by measuring the rate of cell proliferation and metabolite concentrations after knocking down the expression of the three genes.

Copy number variation, aberrant epigenetic regulation, and gain-of-function sequence variants are molecular mechanisms that could contribute to the high activity of *NAGS*, *CPS1*, and citrin in tumor cells that originate from tissues that normally do not express the three genes. An increased gene copy number appears to be responsible for citrin overexpression in glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples. Unlike the four tumor types examined here, amplification of the chromosomal region that harbors the citrin locus has been observed in hepatocellular carcinoma samples, but the abundance of citrin mRNA was similar in the tumors and neighboring normal tissue [64].

An altered chromatin structure of *NAGS*, *CPS1*, and citrin loci in tumors originating from the brain, lung, stomach, or esophagus tissues would allow the binding of transcription factors and RNA polymerase II to promoters and enhancers of the three genes. Histone modifications, CTCF, and RNA polymerase II binding to *NAGS*, *CPS1*, and citrin regulatory regions in the liver, lung, and A549 lung adenocarcinoma cells suggest that all three genes appear to be poised for expression in the lung tissue. Reporter gene assays have shown that the *NAGS* promoter functions in the A549 cell line, but the -3 kb *NAGS* enhancer does not [45]. This could explain the low level of *NAGS* expression in the lung. Therefore, we hypothesize that *CPS1* and citrin promoters could function in the glioblastoma, glioblastoma multiforme, stomach adenocarcinoma, and stomach and esophagus carcinoma cells. This hypothesis could be tested with expression constructs in which *NAGS*, *CPS1*, or citrin promoters and enhancers control reporter gene expression in cell lines that model glioblastoma, glioblastoma multiforme, stomach adenocarcinoma, and stomach and esophagus carcinoma.

There was little overlap between somatic sequence variants found in tumor samples and germline sequence variants found in patients with *NAGS*, *CPS1*, and citrin deficiencies, and in individuals without rare genetic diseases. This is not surprising, because defective DNA proofreading and mismatch repair as well as the imbalance of purine and pyrimidine pools in tumor cells results in different types of nucleotide replacements [54]. Gain-of-function sequence variants in gene regulatory elements can cause the increased expression and/or activity of *NAGS*, *CPS1*, and citrin in tumors. The data in the databases of somatic sequence variants was insufficient to evaluate whether sequence variants in *NAGS*, *CPS1*, and citrin promoters and enhancers contribute to the overexpression of the three genes in tumors. *NAGS*, *CPS1*, and citrin missense variants found in tumors are predicted to be less likely to damage protein function than missense variants found in patients with respective gene deficiencies. This could reflect the limitation of computational methods to predict gain-of-function effects of amino acid replacements. Gain-of-function missense variants that result in a higher enzymatic activity of mutant proteins have been observed in urea cycle enzymes. A high-throughput functional assay of all single nucleotide variant (SNV)-accessible amino acid replacements in human OTC revealed that around 5% of all OTC variants have up to 30% higher activity than wild-type human OTC [65]. Furthermore, human OTC with two common p.R46K and p.Q270R sequence variants had 40% higher enzymatic activity than the wild-type or either of the single mutant proteins [66]. Therefore, we hypothesize that some of the *NAGS*, *CPS1*, and citrin missense variants found in tumors can increase the activity of mutant proteins. This hypothesis could be tested in a high-throughput functional screen in *S. cerevisiae*, since this organism has homologs of *NAGS*, *CPS1*, and citrin [67,68].

Our analysis did not show an association between aberrant citrin expression and unfavorable outcomes in patients with glioblastoma, glioblastoma multiforme, and lung adenocarcinoma. There was no association between the expression of *NAGS*, *CPS1*, or citrin and stomach adenocarcinoma patient outcomes. The number of tumor samples with aberrant expression of two out of the three genes analyzed in this study was insufficient for evaluation of patient outcomes. Association between *NAGS* overexpression and unfavorable outcomes in patients with glioblastoma and glioblastoma multiforme suggests that the aberrant overexpression of *NAGS* and production of NAG could promote tumor development by activating *CPS1* without the requirement for a correlated expression of *NAGS* and *CPS1* in tumor cells. Aberrant overexpression of *NAGS* could contribute to poor outcomes for patients with glioblastoma and glioblastoma multiforme independently of *CPS1*. NAG has been detected in the cytoplasm of mammalian and avian brain cells [69,70] but the function of NAG in the brain remains poorly understood. The abundance of NAG in the mammalian brain and liver are similar [69]. NAG that is produced in the liver mitochondria is transported to the cytoplasm for degradation [71]. A similar transport mechanism may operate in the brain cells allowing NAG, produced in the mitochondria of brain cells by the

aberrantly overexpressed NAGS, to be transported in the cytoplasm, increase the pool of cytoplasmic NAG, and contribute to metabolic reprogramming and tumorigenesis.

Our analysis replicated the association between *CPS1* overexpression and unfavorable outcomes in patients with lung adenocarcinoma [31]. However, the association between lower *NAGS* expression and unfavorable outcomes in patients with lung adenocarcinoma is consistent with the absence of correlated *NAGS* and *CPS1* expression in this tumor type. These observations were supported by the very weak correlation between *NAGS*, *CPS1*, and citrin expression in the individual glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma samples. This raises the question regarding the mechanism behind aberrant *CPS1* activity that contributes to metabolic reprogramming towards increased de novo pyrimidine biosynthesis in tumor cells. One possibility is that *CPS1* missense variants found in tumors enable *CPS1* enzymatic activity without NAG. This hypothesis could be tested in a high-throughput functional activity screen in yeast cells that do not produce NAG. Another possibility is the activation of *CPS1* in tumor cells by N-carbamylglutamate (NCG) produced by the human microbiota. NCG is a structural analog of NAG that can cross plasma and mitochondrial membranes to bind and activate *CPS1* in patients with a *NAGS* deficiency [59]. NCG is an intermediate of histidine catabolism in bacteria and has been detected in metabolomes from gut and oral bacteria [72–74]. Therefore, it is possible that NCG of microbial origin can activate *CPS1* in tumor cells.

The connection between increased de novo biosynthesis of pyrimidine nucleotides, tumor cell proliferation, and aberrant *CPS1* expression and activity made *CPS1* an attractive target for antitumor drugs. AT067-H09 and H3B-120 are two inhibitors of *CPS1* in vitro, in cultured hepatocytes and in cultured non-small cell lung carcinoma cell lines that appear to be promising antitumor drug candidates [36,75]. In addition to inhibiting aberrantly expressed *CPS1* in tumor cells, AT067-H09 and H3B-120 could inhibit hepatic *CPS1* and block the urea cycle leading to hyperammonemia, which can cause irreversible brain injury [10]. Therefore, preclinical and clinical testing of AT067-H09 and H3B-120 will require close collaboration between oncologists and physicians who treat patients with urea cycle disorders, as well as patients with acute and chronic liver failure to avoid hyperammonemia as a serious side effect of tumor treatment with drugs such as AT067-H09 and H3B-120.

4. Methods and Materials

4.1. Expression Patterns of Urea Cycle Genes

The Firehose Data portal, developed by the Genome Data Analysis Center (GDAC) at the Broad Institute, was used to query gene expression profiles of tumors and matched normal tissues for the expression levels of all six urea cycle enzymes and two transporters. Median Log_2RSEM values [37,38] for tumor samples and the matched normal tissues were extracted using the Firebrowse API (Supplementary Table S1). The median Log_2RSEM values for urea cycle genes in tumor and matched normal tissue samples were determined by GDAC. The median Log_2RSEM values from 28 tumor samples and matched normal tissues were used to calculate the fold-change values. For each urea cycle gene in each tumor type, the median Log_2RSEM value for the matched normal tissue samples was subtracted from the median Log_2RSEM value in tumor samples. The fold-change values were then calculated as two raised to the power of the difference between the median Log_2RSEM values in tumor samples and matched normal tissues.

The abundance of urea cycle gene transcripts in the liver, small intestine, stomach, esophageal mucosa and smooth muscle, cerebellum, cerebral cortex, and lung tissues were obtained from the GTEx Project database (Supplementary Table S2). The GTEx database (<https://gtexportal.org/home/datasets>) was accessed on 26 June 2022. The abundance of urea cycle enzymes and transporters in the liver, small intestine, stomach, esophagus, brain, and lung tissues were obtained from the GTEx Project [41] and Proteomics DB, which were accessed on 23 June 2022 [39,40] (Supplementary Tables S3 and S4, respectively).

The transcript abundance data for individual samples of lung adenocarcinoma [60,76–80], glioblastoma [61,79,81,82], glioblastoma multiforme [79,83], and stomach adenocarcinoma [50,79,83–89] were obtained from the cBioPortal for Cancer Genomics [42,43]. Transcript abundance was expressed either as RSEM in glioblastoma, glioblastoma multiforme, and stomach adenocarcinoma samples, or z-score relative to diploid cells in lung adenocarcinoma, to include the maximal number of samples for comparison (Supplementary Table S5).

4.2. Patient Outcome Data

The correlation between patient outcomes and *NAGS*, *CPS1*, or citrin mRNA expression levels was available for select studies of glioblastoma [81], glioblastoma multiforme [79], lung adenocarcinoma [79,80], and stomach adenocarcinoma [79] at the cBioPortal. *NAGS*, *CPS1*, and citrin gene-specific charts were selected for each study. Gene-specific expression data for *NAGS*, *CPS1*, or citrin were used to define patient groups based on the quartiles of gene expression levels. Kaplan–Meier survival curves were compared for patients in the highest and the lowest expression quartile for each gene (Supplementary Table S6).

4.3. Sequence Variant Data

CNV data for *NAGS*, *CPS1*, and citrin were available at the cBioPortal for select studies of glioblastoma [61,81], glioblastoma multiforme [50], lung adenocarcinoma [60,78,80], and stomach adenocarcinoma [50]. The number of samples harboring deep deletions, shallow deletions, two copies, gain and amplification of *NAGS*, *CPS1*, or citrin genes (Supplementary Table S7), as well as the correlation between the type of CNV and *NAGS*, *CPS1*, or citrin expression levels (Supplementary Table S8) were obtained for the four tumor types.

The following databases and websites were queried in September 2021 to collect *NAGS*, *CPS1*, and citrin single nucleotide sequence variants and small indels found in (1) patients with *NAGS*, *CPS1*, or citrin deficiency; (2) tumor samples; and (3) individuals without rare genetic diseases: ClinVar (<https://www.ncbi.nlm.nih.gov/clinvar/> (accessed on 1 May 2021)), Leiden Open Variant Database (LOVD) [53], COSMIC [51], cBioPortal [42,43], TCGA Data Portal [50], and gnomAD [52]. Published reports of sequence variants found in patients with *NAGS*, *CPS1*, and citrin deficiencies were also included in the analysis. *NAGS*, *CPS1*, and citrin sequence variants annotated as pathogenic, likely pathogenic, variants of unknown significance, or variants with conflicting interpretation in ClinVar and/or LOVD were included in the analysis. Variant validator [90] was used for the formatting of all sequences found in patients, tumors, and healthy individuals (Supplementary Table S9). Sequence variants were classified by type and location in the *NAGS*, *CPS1*, and citrin genes: regulatory (located upstream of transcription start site), 5'-UTR, missense, nonsense, frameshift, in-frame indel, synonymous, splicing (affecting canonical GT and AG splice signals at the 5' and 3'-ends of introns), splice region (located up to 10 bp downstream and upstream of the GT and AG splice signals, respectively), intronic (located more than 10 bp away from the canonical splice sites), 3'-UTR, and unknown. Nonsense, frameshift, and splicing sequence variants were considered loss-of-function variants. The REVEL functional effect predictor was used to evaluate the effects of missense variants on protein function. REVEL [55] scores of *NAGS*, *CPS1*, and citrin missense variants obtained through the Ensembl Variant Effect Prediction tool [91] were used to evaluate the effects of missense variants on protein function (Supplementary Table S10). Missense variants with a REVEL score above 0.5 are considered damaging, while missense variants with a REVEL score below 0.5 are considered tolerated [55]. The ClinGen recommendations for relating the REVEL score of a missense variant to the strength of evidence that the missense variant has either a damaging or benign effect on protein function [57] were used to refine the functional effect predictions for *NAGS*, *CPS1*, and citrin variants found in patients, tumors, and gnomAD. A REVEL score greater than or equal to 0.932 is considered strong evidence for the damaging effect of a sequence variant on protein function [57]. A REVEL score between 0.932 and 0.773 is considered moderate evidence for the damaging effect of a sequence

variant on protein function [57]. A REVEL score between 0.773 and 0.664 is considered supporting evidence for the damaging effect of a sequence variant on protein function [57]. The effect of a sequence variant with REVEL score between 0.644 and 0.290 on protein function is uncertain [57]. REVEL scores between 0.290 and 0.183, 0.193 and 0.016, and 0.016 and 0.003 are considered respectively as supporting, moderate, and strong evidence for benign effect on protein function, while a REVEL score lower than 0.003 is considered very strong evidence for benign effect on protein function [57].

4.4. Epigenetic Regulation Data

The ENCODE Project Functional Genomics Portal [92,93] was queried for the availability of data for DNase sensitivity and hypersensitivity sites, histone modifications, CTCF binding sites, and RNA polymerase II binding sites in the liver, lung, adult brain, and stomach tissues, as well as cancer cell lines that model glioblastoma, glioblastoma multiforme, lung adenocarcinoma, and stomach adenocarcinoma. The following filters were applied to the ENCODE Experimental Matrix: DNA binding and DNA accessibility for assay type; TF-ChIP-seq, Histone ChIP-seq, DNase-seq and ATAC-seq for assay title; Homo sapiens for organism; cell line and tissue for biosample classification; brain, liver, lung, stomach, A549, H54, A172 and M059J for biosample; liver, lung, brain, and stomach for organ. Histone ChIP-seq, TF-ChIP-seq, DNase-seq, and ATAC-seq data were available for liver, stomach, and lung tissues, and for the A549 lung adenocarcinoma cell line. Histone ChIP-seq, TF-ChIP-seq, and DNase-seq data were available for embryonic brain samples and brain samples from patients with Alzheimer's disease. BigWig fold-change files for biological and technical replicates for the following experiments were downloaded: CTCF ChIP-seq, RNAP2A ChIP-seq, DNase-seq, ATAC-seq, H3K27Ac ChIP-seq, and H3K4me3 ChIP-seq. Genomic coordinates corresponding to *NAGS*, *CPS1*, and citrin loci were determined based either on the location of CTCF binding sites or the adjacent upstream and downstream gene.

A custom Python script (<https://github.com/MIMOR02/bigwig-file-visualizations> (accessed on 10 June 2022)) was used to extract and graph the fold-change signal from BigWig files for the *NAGS*, *CPS1*, and citrin loci. There are three notebooks in the GitHub repository which can be run in JupyterLab or Google Colab: *Data_Visualization-GITHUB_PUSH.ipynb*, *ENCODE_JSON_parser-GITHUB_PUSH.ipynb*, and *Format_Directory-GITHUB_PUSH.ipynb*. *ENCODE_JSON_parser-GITHUB_PUSH.ipynb* scrapes an experiment matrix from the ENCODE Project website for the desired bigwig file download links. Bigwig files were downloaded manually. *Format_Directory-GITHUB_PUSH.ipynb* creates a nested file system that all downloaded bigwig files will be sorted into. *Data_Visualization-GITHUB_PUSH.ipynb* reads in bigwig files, generates appropriate graphs to the user's specifications, and moves the visualizations into the file system created by *Format_Directory-GITHUB_PUSH.ipynb*.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24076754/s1>.

Author Contributions: M.O.-A., N.G. and D.A. contributed equally to this work. M.O.-A.—performed collection and analysis of *NAGS* sequence variants, collected *NAGS* expression data in tumors and normal tissues, wrote and edited the manuscript. N.G.—performed collection and analysis of *CPS1* sequence variants, collected *CPS1* expression data in tumors and normal tissues, wrote and edited the manuscript. D.A.—performed collection and analysis of citrin sequence variants and collected citrin expression data in tumors and normal tissues, wrote and edited the manuscript. M.M.—wrote Python scripts that were used for data collection from the ENCODE Project database and visualization of DNase-Seq, ATAC-Seq, and ChIP-Seq data, and edited the manuscript. L.C.—conceived and directed the project, collected patient outcome data, wrote and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Recordati Rare Disease, Inc. special purpose fund.

Institutional Review Board Statement: Not applicable. Patient consent was waived due to the data collection from public databases.

Informed Consent Statement: Not applicable. Patient consent was waived due to the data collection from public databases.

Data Availability Statement: All data used in this study have been included in Supplementary Files. Custom Python scripts used for acquiring and analysis of the data from the ENCODE Project are available at <https://github.com/MIMOR02/bigwig-file-visualizations> (accessed on 10 June 2022).

Acknowledgments: We are grateful to Hiroki Morizono for the helpful discussions and advice related to this project.

Conflicts of Interest: Ljubica Caldovic's work was supported by the Recordati Rare Diseases, Inc. that manufactures and sells NCG as Carbaglu® (carglumic acid), which is used for the treatment of NAGS deficiency. Other manuscript authors declare no conflict of interest.

References

1. Jin, G.; Reitman, Z.J.; Duncan, C.G.; Spasojevic, I.; Gooden, D.M.; Rasheed, B.A.; Yang, R.; Lopez, G.Y.; He, Y.; McLendon, R.E.; et al. Disruption of wild-type IDH1 suppresses D-2-hydroxyglutarate production in IDH1-mutated gliomas. *Cancer Res.* **2013**, *73*, 496–501. [[CrossRef](#)]
2. Jin, G.; Pirozzi, C.J.; Chen, L.H.; Lopez, G.Y.; Duncan, C.G.; Feng, J.; Spasojevic, I.; Bigner, D.D.; He, Y.; Yan, H. Mutant IDH1 is required for IDH1 mutated tumor cell growth. *Oncotarget* **2012**, *3*, 774–782. [[CrossRef](#)]
3. Sonoda, Y.; Kumabe, T.; Nakamura, T.; Saito, R.; Kanamori, M.; Yamashita, Y.; Suzuki, H.; Tominaga, T. Analysis of IDH1 and IDH2 mutations in Japanese glioma patients. *Cancer Sci.* **2009**, *100*, 1996–1998. [[CrossRef](#)]
4. Hartmann, C.; Meyer, J.; Balss, J.; Capper, D.; Mueller, W.; Christians, A.; Felsberg, J.; Wolter, M.; Mawrin, C.; Wick, W.; et al. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: A study of 1010 diffuse gliomas. *Acta Neuropathol.* **2009**, *118*, 469–474. [[CrossRef](#)] [[PubMed](#)]
5. Yan, H.; Parsons, D.W.; Jin, G.; McLendon, R.; Rasheed, B.A.; Yuan, W.; Kos, I.; Batinic-Haberle, I.; Jones, S.; Riggins, G.J.; et al. IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.* **2009**, *360*, 765–773. [[CrossRef](#)]
6. Selak, M.A.; Armour, S.M.; MacKenzie, E.D.; Boulahbel, H.; Watson, D.G.; Mansfield, K.D.; Pan, Y.; Simon, M.C.; Thompson, C.B.; Gottlieb, E. Succinate links TCA cycle dysfunction to oncogenesis by inhibiting HIF- α prolyl hydroxylase. *Cancer Cell* **2005**, *7*, 77–85. [[CrossRef](#)]
7. Letouze, E.; Martinelli, C.; Loriot, C.; Burnichon, N.; Abermil, N.; Ottolenghi, C.; Janin, M.; Menara, M.; Nguyen, A.T.; Benit, P.; et al. SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell* **2013**, *23*, 739–752. [[CrossRef](#)] [[PubMed](#)]
8. Sullivan, L.B.; Martinez-Garcia, E.; Nguyen, H.; Mullen, A.R.; Dufour, E.; Sudarshan, S.; Licht, J.D.; Deberardinis, R.J.; Chandel, N.S. The proto-oncometabolite fumarate binds glutathione to amplify ROS-dependent signaling. *Mol. Cell.* **2013**, *51*, 236–248. [[CrossRef](#)] [[PubMed](#)]
9. Lee, J.S.; Adler, L.; Karathia, H.; Carmel, N.; Rabinovich, S.; Auslander, N.; Keshet, R.; Stettner, N.; Silberman, A.; Agemy, L.; et al. Urea Cycle Dysregulation Generates Clinically Relevant Genomic and Biochemical Signatures. *Cell* **2018**, *174*, 1559–1570.e1522. [[CrossRef](#)]
10. Ah Mew, N.; Simpson, K.L.; Gropman, A.L.; Lanpher, B.C.; Chapman, K.A.; Summar, M.L. Urea Cycle Disorders Overview. In *GeneReviews ((R))*; Adam, M.P., Ardinger, H.H., Pagon, R.A., Wallace, S.E., Bean, L.J.H., Stephens, K., Amemiya, A., Eds.; University of Washington: Seattle, WA, USA, 1993.
11. Brusilow, S.W.; Horwich, A.L. Urea Cycle Enzymes. In *The Metabolic & Molecular Bases of Inherited Disease*; Scriver, C.R., Beaudet, A.L., Sly, W.S., Valle, D., Eds.; McGraw-Hill: New York, NY, USA, 2001; Volume 2, pp. 1909–1963.
12. Caldovic, L.; Tuchman, M. N-acetylglutamate and its changing role through evolution. *Biochem. J.* **2003**, *372*, 279–290. [[CrossRef](#)]
13. Meijer, A.J.; Lamers, W.H.; Chamuleau, R.A.F.M. Nitrogen Metabolism and Ornithine Cycle Function. *Physiol. Rev.* **1990**, *70*, 701–748. [[CrossRef](#)]
14. Powers-Lee, S.G. N-Acetylglutamate synthase. *Methods Enzymol.* **1985**, *113*, 27–35.
15. De Cima, S.; Polo, L.M.; Diez-Fernandez, C.; Martinez, A.I.; Cervera, J.; Fita, I.; Rubio, V. Structure of human carbamoyl phosphate synthetase: Deciphering the on/off switch of human ureagenesis. *Sci. Rep.* **2015**, *5*, 16950. [[CrossRef](#)]
16. Guthohrlein, G.; Knappe, J. Structure and function of carbamoylphosphate synthase. I. Transitions between two catalytically inactive forms and the active form. *Eur. J. Biochem.* **1968**, *7*, 119–127. [[CrossRef](#)]
17. Hall, L.M.; Metzberg, R.L.; Cohen, P.P. Isolation and characterization of a naturally occurring cofactor of carbamyl phosphate biosynthesis. *J. Biol. Chem.* **1958**, *230*, 1013–1021. [[CrossRef](#)] [[PubMed](#)]
18. McCudden, C.R.; Powers-Lee, S.G. Required allosteric effector site for N-acetylglutamate on carbamoyl-phosphate synthetase I. *J. Biol. Chem.* **1996**, *271*, 18285–18294. [[CrossRef](#)] [[PubMed](#)]
19. Erez, A.; Nagamani, S.C.; Shchelochkov, O.A.; Premkumar, M.H.; Campeau, P.M.; Chen, Y.; Garg, H.K.; Li, L.; Mian, A.; Bertin, T.K.; et al. Requirement of argininosuccinate lyase for systemic nitric oxide production. *Nat. Med.* **2011**, *17*, 1619–1626. [[CrossRef](#)]

20. Begum, L.; Jalil, M.A.; Kobayashi, K.; Iijima, M.; Li, M.X.; Yasuda, T.; Horiuchi, M.; del Arco, A.; Satrustegui, J.; Saheki, T. Expression of three mitochondrial solute carriers, citrin, aralar1 and ornithine transporter, in relation to urea cycle in mice. *Biochim. Biophys. Acta* **2002**, *1574*, 283–292. [[CrossRef](#)]
21. Camacho, J.A.; Obie, C.; Biery, B.; Goodman, B.K.; Hu, C.A.; Almashanu, S.; Steel, G.; Casey, R.; Lambert, M.; Mitchell, G.A.; et al. Hyperornithinaemia-hyperammonaemia-homocitrullinuria syndrome is caused by mutations in a gene encoding a mitochondrial ornithine transporter. *Nat. Genet.* **1999**, *22*, 151–158. [[CrossRef](#)] [[PubMed](#)]
22. Amoedo, N.D.; Punzi, G.; Obre, E.; Lacombe, D.; De Grassi, A.; Pierri, C.L.; Rossignol, R. AGC1/2, the mitochondrial aspartate-glutamate carriers. *Biochim. Biophys. Acta* **2016**, *1863*, 2394–2412. [[CrossRef](#)]
23. Morizono, H.; Woolston, J.E.; Colombini, M.; Tuchman, M. The use of yeast mitochondria to study the properties of wild-type and mutant human mitochondrial ornithine transporter. *Mol. Genet. Metab.* **2005**, *86*, 431–440. [[CrossRef](#)] [[PubMed](#)]
24. Palmieri, L.; Pardo, B.; Lasorsa, F.M.; del Arco, A.; Kobayashi, K.; Iijima, M.; Runswick, M.J.; Walker, J.E.; Saheki, T.; Satrustegui, J.; et al. Citrin and aralar1 are Ca(2+)-stimulated aspartate/glutamate transporters in mitochondria. *EMBO J.* **2001**, *20*, 5060–5069. [[CrossRef](#)]
25. Saheki, T.; Kobayashi, K.; Iijima, M.; Nishi, I.; Yasuda, T.; Yamaguchi, N.; Gao, H.Z.; Jalil, M.A.; Begum, L.; Li, M.X. Pathogenesis and pathophysiology of citrin (a mitochondrial aspartate glutamate carrier) deficiency. *Metab. Brain Dis.* **2002**, *17*, 335–346. [[CrossRef](#)] [[PubMed](#)]
26. Saheki, T.; Kobayashi, K. Mitochondrial aspartate glutamate carrier (citrin) deficiency as the cause of adult-onset type II citrullinemia (CTLN2) and idiopathic neonatal hepatitis (NICCD). *J. Hum. Genet.* **2002**, *47*, 333–341. [[CrossRef](#)] [[PubMed](#)]
27. Nagamani, S.C.; Erez, A. A metabolic link between the urea cycle and cancer cell proliferation. *Mol. Cell. Oncol.* **2016**, *3*, e1127314. [[CrossRef](#)]
28. Rabinovich, S.; Adler, L.; Yizhak, K.; Sarver, A.; Silberman, A.; Agron, S.; Stettner, N.; Sun, Q.; Brandis, A.; Helbling, D.; et al. Diversion of aspartate in ASS1-deficient tumours fosters de novo pyrimidine synthesis. *Nature* **2015**, *527*, 379–383. [[CrossRef](#)]
29. Miyo, M.; Konno, M.; Nishida, N.; Sueda, T.; Noguchi, K.; Matsui, H.; Colvin, H.; Kawamoto, K.; Koseki, J.; Haraguchi, N.; et al. Metabolic Adaptation to Nutritional Stress in Human Colorectal Cancer. *Sci. Rep.* **2016**, *6*, 38415. [[CrossRef](#)] [[PubMed](#)]
30. Rabinovich, S.; Silberman, A.; Adler, L.; Agron, S.; Levin-Zaidman, S.; Bahat, A.; Porat, Z.; Ben-Zeev, E.; Geva, I.; Itkin, M.; et al. The mitochondrial carrier Citrin plays a role in regulating cellular energy during carcinogenesis. *Oncogene* **2020**, *39*, 164–175. [[CrossRef](#)]
31. Celiktas, M.; Tanaka, I.; Tripathi, S.C.; Fahrman, J.F.; Aguilar-Bonavides, C.; Villalobos, P.; Delgado, O.; Dhillon, D.; Dennison, J.B.; Ostrin, E.J.; et al. Role of CPS1 in Cell Growth, Metabolism and Prognosis in LKB1-Inactivated Lung Adenocarcinoma. *J. Natl. Cancer Inst.* **2017**, *109*, 1–9. [[CrossRef](#)] [[PubMed](#)]
32. Kim, J.; Hu, Z.; Cai, L.; Li, K.; Choi, E.; Faubert, B.; Bezwada, D.; Rodriguez-Canales, J.; Villalobos, P.; Lin, Y.F.; et al. CPS1 maintains pyrimidine pools and DNA synthesis in KRAS/LKB1-mutant lung cancer cells. *Nature* **2017**, *546*, 168–172. [[CrossRef](#)]
33. Li, L.; Mao, Y.; Zhao, L.; Li, L.; Wu, J.; Zhao, M.; Du, W.; Yu, L.; Jiang, P. p53 regulation of ammonia metabolism through urea cycle controls polyamine biosynthesis. *Nature* **2019**, *567*, 253–256. [[CrossRef](#)]
34. Pham-Danis, C.; Gehrke, S.; Danis, E.; Rozhok, A.I.; Daniels, M.W.; Gao, D.; Collins, C.; Paola, J.T.D.; D’Alessandro, A.; DeGregori, J. Urea Cycle Sustains Cellular Energetics upon EGFR Inhibition in EGFR-Mutant NSCLC. *Mol. Cancer Res.* **2019**, *17*, 1351–1364. [[CrossRef](#)]
35. Wu, G.; Yan, Y.; Zhou, Y.; Wang, X.; Wei, J.; Chen, X.; Lin, W.; Ou, C.; Zhou, J.; Xu, Z. Expression and clinical significance of CPS1 in glioblastoma multiforme. *Curr. Res. Transl. Med.* **2019**, *67*, 123–128. [[CrossRef](#)]
36. Makris, G.; Kayhan, S.; Kreuzer, M.; Rufenacht, V.; Faccin, E.; Underhaug, J.; Diez-Fernandez, C.; Knobel, P.A.; Poms, M.; Gougeward, N.; et al. Impact of small molecule-mediated inhibition of ammonia detoxification on lung malignancies and liver metabolism. *Cancer Commun.* **2023**, *ahead of print*. [[CrossRef](#)]
37. Cheng, P.F.; Dummer, R.; Levesque, M.P. Data mining The Cancer Genome Atlas in the era of precision cancer medicine. *Swiss Med. Wkly.* **2015**, *145*, w14183. [[CrossRef](#)]
38. Li, B.; Dewey, C.N. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **2011**, *12*, 323. [[CrossRef](#)]
39. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509*, 582–587. [[CrossRef](#)]
40. Lautenbacher, L.; Samaras, P.; Muller, J.; Grafberger, A.; Shraideh, M.; Rank, J.; Fuchs, S.T.; Schmidt, T.K.; The, M.; Dallago, C.; et al. ProteomicsDB: Toward a FAIR open-source resource for life-science research. *Nucleic Acids Res.* **2022**, *50*, D1541–D1552. [[CrossRef](#)] [[PubMed](#)]
41. Jiang, L.; Wang, M.; Lin, S.; Jian, R.; Li, X.; Chan, J.; Dong, G.; Fang, H.; Robinson, A.E.; Consortium, G.T.; et al. A Quantitative Proteome Map of the Human Body. *Cell* **2020**, *183*, 269–283.e219. [[CrossRef](#)] [[PubMed](#)]
42. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **2013**, *6*, p11. [[CrossRef](#)] [[PubMed](#)]
43. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; et al. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2012**, *2*, 401–404. [[CrossRef](#)]

44. Heibel, S.K.; Ah Mew, N.; Caldovic, L.; Daikhin, Y.; Yudkoff, M.; Tuchman, M. N-carbamylglutamate enhancement of ureagenesis leads to discovery of a novel deleterious mutation in a newly defined enhancer of the NAGS gene and to effective therapy. *Hum. Mutat.* **2011**, *32*, 1153–1160. [[CrossRef](#)]
45. Heibel, S.K.; Lopez, G.Y.; Panglao, M.; Sodha, S.; Marino-Ramirez, L.; Tuchman, M.; Caldovic, L. Transcriptional regulation of N-acetylglutamate synthase. *PLoS ONE* **2012**, *7*, e29527. [[CrossRef](#)]
46. Chen, Z.; Tang, N.; Wang, X.; Chen, Y. The activity of the carbamoyl phosphate synthase 1 promoter in human liver-derived cells is dependent on hepatocyte nuclear factor 3-beta. *J. Cell. Mol. Med.* **2017**, *21*, 2036–2045. [[CrossRef](#)] [[PubMed](#)]
47. Chen, J.L.; Zhang, Z.H.; Li, B.X.; Cai, Z.; Zhou, Q.H. Bioinformatic and functional analysis of promoter region of human SLC25A13 gene. *Gene* **2019**, *693*, 69–75. [[CrossRef](#)] [[PubMed](#)]
48. Convertini, P.; Todisco, S.; De Santis, F.; Pappalardo, I.; Iacobazzi, D.; Castiglione Morelli, M.A.; Fondufe-Mittendorf, Y.N.; Martelli, G.; Palmieri, F.; Infantino, V. Transcriptional Regulation Factors of the Human Mitochondrial Aspartate/Glutamate Carrier Gene, Isoform 2 (SLC25A13): USF1 as Basal Factor and FOXA2 as Activator in Liver Cells. *Int. J. Mol. Sci.* **2019**, *20*, 1888. [[CrossRef](#)] [[PubMed](#)]
49. Shao, X.; Lv, N.; Liao, J.; Long, J.; Xue, R.; Ai, N.; Xu, D.; Fan, X. Copy number variation is highly correlated with differential gene expression: A pan-cancer study. *BMC Med. Genet.* **2019**, *20*, 175. [[CrossRef](#)] [[PubMed](#)]
50. Ding, L.; Bailey, M.H.; Porta-Pardo, E.; Thorsson, V.; Colaprico, A.; Bertrand, D.; Gibbs, D.L.; Weerasinghe, A.; Huang, K.L.; Tokheim, C.; et al. Perspective on Oncogenic Processes at the End of the Beginning of Cancer Genomics. *Cell* **2018**, *173*, 305–320.e310. [[CrossRef](#)]
51. Tate, J.G.; Bamford, S.; Jubb, H.C.; Sondka, Z.; Beare, D.M.; Bindal, N.; Boutselakis, H.; Cole, C.G.; Creatore, C.; Dawson, E.; et al. COSMIC: The Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **2019**, *47*, D941–D947. [[CrossRef](#)]
52. Karczewski, K.J.; Francioli, L.C.; Tiao, G.; Cummings, B.B.; Alfoldi, J.; Wang, Q.; Collins, R.L.; Laricchia, K.M.; Ganna, A.; Birnbaum, D.P.; et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **2020**, *581*, 434–443. [[CrossRef](#)]
53. Fokkema, I.F.; Taschner, P.E.; Schaafsma, G.C.; Celli, J.; Laros, J.F.; den Dunnen, J.T. LOVD v.2.0: The next generation in gene variant databases. *Hum. Mutat.* **2011**, *32*, 557–563. [[CrossRef](#)] [[PubMed](#)]
54. Alexandrov, L.B.; Kim, J.; Haradhvala, N.J.; Huang, M.N.; Tian Ng, A.W.; Wu, Y.; Boot, A.; Covington, K.R.; Gordenin, D.A.; Bergstrom, E.N.; et al. The repertoire of mutational signatures in human cancer. *Nature* **2020**, *578*, 94–101. [[CrossRef](#)] [[PubMed](#)]
55. Ioannidis, N.M.; Rothstein, J.H.; Pejaver, V.; Middha, S.; McDonnell, S.K.; Baheti, S.; Musolf, A.; Li, Q.; Holzinger, E.; Karyadi, D.; et al. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am. J. Hum. Genet.* **2016**, *99*, 877–885. [[CrossRef](#)]
56. Lawson, F.S.; Charlebois, R.L.; Dillon, J.A. Phylogenetic analysis of carbamoylphosphate synthetase genes: Complex evolutionary history includes an internal duplication within a gene which can root the tree of life. *Mol. Biol. Evol.* **1996**, *13*, 970–977. [[CrossRef](#)] [[PubMed](#)]
57. Pejaver, V.; Byrne, A.B.; Feng, B.J.; Pagel, K.A.; Mooney, S.D.; Karchin, R.; O'Donnell-Luria, A.; Harrison, S.M.; Tavtigian, S.V.; Greenblatt, M.S.; et al. Calibration of computational tools for missense variant pathogenicity classification and ClinGen recommendations for PP3/BP4 criteria. *Am. J. Hum. Genet.* **2022**, *109*, 2163–2177. [[CrossRef](#)]
58. Senkevitch, E.; Cabrera-Luque, J.; Morizono, H.; Caldovic, L.; Tuchman, M. A novel biochemically salvageable animal model of hyperammonemia devoid of N-acetylglutamate synthase. *Mol. Genet. Metab.* **2012**, *106*, 160–168. [[CrossRef](#)]
59. Caldovic, L.; Morizono, H.; Daikhin, Y.; Nissim, I.; McCarter, R.J.; Yudkoff, M.; Tuchman, M. Restoration of ureagenesis in N-acetylglutamate synthase deficiency by N-carbamylglutamate. *J. Pediatr.* **2004**, *145*, 552–554. [[CrossRef](#)]
60. Gillette, M.A.; Satpathy, S.; Cao, S.; Dhanasekaran, S.M.; Vasaiakar, S.V.; Krug, K.; Petralia, F.; Li, Y.; Liang, W.W.; Reva, B.; et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* **2020**, *182*, 200–225.e235. [[CrossRef](#)]
61. Wang, L.B.; Karpova, A.; Gritsenko, M.A.; Kyle, J.E.; Cao, S.; Li, Y.; Rykunov, D.; Colaprico, A.; Rothstein, J.H.; Hong, R.; et al. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **2021**, *39*, 509–528.e520. [[CrossRef](#)]
62. Rudman, D.; DiFulco, T.J.; Galambos, J.T.; Smith, R.B., 3rd; Salam, A.A.; Warren, W.D. Maximal rates of excretion and synthesis of urea in normal and cirrhotic subjects. *J. Clin. Investig.* **1973**, *52*, 2241–2249. [[CrossRef](#)] [[PubMed](#)]
63. Wang, D.; Eraslan, B.; Wieland, T.; Hallstrom, B.; Hopf, T.; Zolg, D.P.; Zecha, J.; Asplund, A.; Li, L.H.; Meng, C.; et al. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* **2019**, *15*, e8503. [[CrossRef](#)]
64. Dong, H.; Zhang, H.; Liang, J.; Yan, H.; Chen, Y.; Shen, Y.; Kong, Y.; Wang, S.; Zhao, G.; Jin, W. Digital karyotyping reveals probable target genes at 7q21.3 locus in hepatocellular carcinoma. *BMC Med. Genomics* **2011**, *4*, 60. [[CrossRef](#)] [[PubMed](#)]
65. Lo, R.S.; Cromie, G.A.; Tang, M.; Teng, K.; Owens, K.; Sirr, A.; Kutz, J.N.; McLaughlin, R.N., Jr.; Morizono, H.; Caldovic, L.; et al. The functional impact of 1,570 SNP-accessible missense variants in human OTC. *bioRxiv* **2022**. [[CrossRef](#)]
66. Lopes-Marques, M.; Pacheco, A.R.; Peixoto, M.J.; Cardoso, A.R.; Serrano, C.; Amorim, A.; Prata, M.J.; Cooper, D.N.; Azevedo, L. Common polymorphic OTC variants can act as genetic modifiers of enzymatic activity. *Hum. Mutat.* **2021**, *42*, 978–989. [[CrossRef](#)]
67. Abadjieva, A.; Pauwels, K.; Hilven, P.; Crabeel, M. A new yeast metabolon involving at least the two first enzymes of arginine biosynthesis: Acetylglutamate synthase activity requires complex formation with acetylglutamate kinase. *J. Biol. Chem.* **2001**, *276*, 42869–42880. [[CrossRef](#)]

68. Jauniaux, J.C.; Urrestarazu, L.A.; Wiame, J.M. Arginine metabolism in *Saccharomyces cerevisiae*: Subcellular localization of the enzymes. *J. Bacteriol.* **1978**, *133*, 1096–1107. [[CrossRef](#)]
69. Alonso, E.; Garcia-Perez, M.A.; Bueso, J.; Rubio, V. N-acetyl-L-glutamate in brain: Assay, levels, and regional and subcellular distribution. *Neurochem. Res.* **1991**, *16*, 787–794. [[CrossRef](#)] [[PubMed](#)]
70. Reichelt, K.L.; Kvamme, E. Acetylated and peptide bound glutamate and sapartate in brain. *J. Neurochem.* **1967**, *14*, 987–995. [[CrossRef](#)]
71. Reglero, A.; Rivas, J.; Mendelson, J.; Wallace, R.; Grisolia, S. Deacylation and transacetylation of acetyl glutamate and acetyl ornithine in rat liver. *FEBS Lett.* **1977**, *81*, 13–17. [[CrossRef](#)] [[PubMed](#)]
72. Bowerman, K.L.; Rehman, S.F.; Vaughan, A.; Lachner, N.; Budden, K.F.; Kim, R.Y.; Wood, D.L.A.; Gellatly, S.L.; Shukla, S.D.; Wood, L.G.; et al. Disease-associated gut microbiome and metabolome changes in patients with chronic obstructive pulmonary disease. *Nat. Commun.* **2020**, *11*, 5886. [[CrossRef](#)]
73. Menni, C.; Zierer, J.; Pallister, T.; Jackson, M.A.; Long, T.; Mohny, R.P.; Steves, C.J.; Spector, T.D.; Valdes, A.M. Omega-3 fatty acids correlate with gut microbiome diversity and production of N-carbamylglutamate in middle aged and elderly women. *Sci. Rep.* **2017**, *7*, 11079. [[CrossRef](#)] [[PubMed](#)]
74. Pei, J.; Li, F.; Xie, Y.; Liu, J.; Yu, T.; Feng, X. Microbial and metabolomic analysis of gingival crevicular fluid in general chronic periodontitis patients: Lessons for a predictive, preventive, and personalized medical approach. *EPMA J.* **2020**, *11*, 197–215. [[CrossRef](#)]
75. Yao, S.; Nguyen, T.V.; Rolfe, A.; Agrawal, A.A.; Ke, J.; Peng, S.; Colombo, F.; Yu, S.; Bouchard, P.; Wu, J.; et al. Small Molecule Inhibition of CPS1 Activity through an Allosteric Pocket. *Cell. Chem. Biol.* **2020**, *27*, 259–268.e255. [[CrossRef](#)]
76. Imielinski, M.; Berger, A.H.; Hammerman, P.S.; Hernandez, B.; Pugh, T.J.; Hodis, E.; Cho, J.; Suh, J.; Capelletti, M.; Sivachenko, A.; et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **2012**, *150*, 1107–1120. [[CrossRef](#)]
77. Caso, R.; Sanchez-Vega, F.; Tan, K.S.; Mastrogiacomo, B.; Zhou, J.; Jones, G.D.; Nguyen, B.; Schultz, N.; Connolly, J.G.; Brandt, W.S.; et al. The Underlying Tumor Genomics of Predominant Histologic Subtypes in Lung Adenocarcinoma. *J. Thorac. Oncol.* **2020**, *15*, 1844–1856. [[CrossRef](#)]
78. Chen, J.; Yang, H.; Teo, A.S.M.; Amer, L.B.; Sherbaf, F.G.; Tan, C.Q.; Alvarez, J.J.S.; Lu, B.; Lim, J.Q.; Takano, A.; et al. Genomic landscape of lung adenocarcinoma in East Asians. *Nat. Genet.* **2020**, *52*, 177–186. [[CrossRef](#)]
79. Liu, J.; Lichtenberg, T.; Hoadley, K.A.; Poisson, L.M.; Lazar, A.J.; Cherniack, A.D.; Kovatich, A.J.; Benz, C.C.; Levine, D.A.; Lee, A.V.; et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **2018**, *173*, 400–416.e411. [[CrossRef](#)]
80. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. [[CrossRef](#)]
81. Brennan, C.W.; Verhaak, R.G.; McKenna, A.; Campos, B.; Noushmehr, H.; Salama, S.R.; Zheng, S.; Chakravarty, D.; Sanborn, J.Z.; Berman, S.H.; et al. The somatic genomic landscape of glioblastoma. *Cell* **2013**, *155*, 462–477. [[CrossRef](#)] [[PubMed](#)]
82. Zhao, J.; Chen, A.X.; Gartrell, R.D.; Silverman, A.M.; Aparicio, L.; Chu, T.; Bordbar, D.; Shan, D.; Samanamud, J.; Mahajan, A.; et al. Immune and genomic correlates of response to anti-PD-1 immunotherapy in glioblastoma. *Nat. Med.* **2019**, *25*, 462–469. [[CrossRef](#)] [[PubMed](#)]
83. Taylor, A.M.; Shih, J.; Ha, G.; Gao, G.F.; Zhang, X.; Berger, A.C.; Schumacher, S.E.; Wang, C.; Hu, H.; Liu, J.; et al. Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell* **2018**, *33*, 676–689.e673. [[CrossRef](#)] [[PubMed](#)]
84. Wang, K.; Yuen, S.T.; Xu, J.; Lee, S.P.; Yan, H.H.; Shi, S.T.; Siu, H.C.; Deng, S.; Chu, K.M.; Law, S.; et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **2014**, *46*, 573–582. [[CrossRef](#)] [[PubMed](#)]
85. Hoadley, K.A.; Yau, C.; Hinoue, T.; Wolf, D.M.; Lazar, A.J.; Drill, E.; Shen, R.; Taylor, A.M.; Cherniack, A.D.; Thorsson, V.; et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **2018**, *173*, 291–304.e296. [[CrossRef](#)] [[PubMed](#)]
86. Gao, Q.; Liang, W.W.; Foltz, S.M.; Mutharasu, G.; Jayasinghe, R.G.; Cao, S.; Liao, W.W.; Reynolds, S.M.; Wyczalkowski, M.A.; Yao, L.; et al. Driver Fusions and Their Implications in the Development and Treatment of Human Cancers. *Cell. Rep.* **2018**, *23*, 227–238.e223. [[CrossRef](#)] [[PubMed](#)]
87. Bonneville, R.; Krook, M.A.; Kautto, E.A.; Miya, J.; Wing, M.R.; Chen, H.Z.; Reeser, J.W.; Yu, L.; Roychowdhury, S. Landscape of Microsatellite Instability Across 39 Cancer Types. *JCO Precis. Oncol.* **2017**, *2017*, PO.17.00073. [[CrossRef](#)]
88. Kakiuchi, M.; Nishizawa, T.; Ueda, H.; Gotoh, K.; Tanaka, A.; Hayashi, A.; Yamamoto, S.; Tatsuno, K.; Katoh, H.; Watanabe, Y.; et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.* **2014**, *46*, 583–587. [[CrossRef](#)] [[PubMed](#)]
89. Wang, K.; Kan, J.; Yuen, S.T.; Shi, S.T.; Chu, K.M.; Law, S.; Chan, T.L.; Kan, Z.; Chan, A.S.; Tsui, W.Y.; et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **2011**, *43*, 1219–1223. [[CrossRef](#)]
90. Freeman, P.J.; Hart, R.K.; Gretton, L.J.; Brookes, A.J.; Dagleish, R. VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Hum. Mutat.* **2018**, *39*, 61–68. [[CrossRef](#)]
91. Cunningham, F.; Allen, J.E.; Allen, J.; Alvarez-Jarreta, J.; Amode, M.R.; Armean, I.M.; Austine-Orimoloye, O.; Azov, A.G.; Barnes, I.; Bennett, R.; et al. Ensembl 2022. *Nucleic Acids Res.* **2022**, *50*, D988–D995. [[CrossRef](#)]

92. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. [[CrossRef](#)]
93. Luo, Y.; Hitz, B.C.; Gabdank, I.; Hilton, J.A.; Kagda, M.S.; Lam, B.; Myers, Z.; Sud, P.; Jou, J.; Lin, K.; et al. New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic Acids Res.* **2020**, *48*, D882–D889. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.