CANCER

Arginine limitation drives a directed codon-dependent DNA sequence evolution response in colorectal cancer cells

Dennis J. Hsu[1,2†], Jenny Gao[1], Norihiro Yamaguchi[1‡], Alexandra Pinzaru[1], Qiushuang Wu[1], Nandan Mandayam[1], Maria Liberti[1], Søren Heissel[3], Hanan Alwaseem[3], Saeed Tavazoie[4,5,6*], Sohail F. Tavazoie[1*]

Utilization of specific codons varies between organisms. Cancer represents a model for understanding DNA sequence evolution and could reveal causal factors underlying codon evolution. We found that across human cancer, arginine codons are frequently mutated to other codons. Moreover, arginine limitation—a feature of tumor microenvironments—is sufficient to induce arginine codon–switching mutations in human colon cancer cells. Such DNA codon switching events encode mutant proteins with arginine residue substitutions. Mechanistically, arginine limitation caused rapid reduction of arginine transfer RNAs and the stalling of ribosomes over arginine codons. Such selective pressure against arginine codon translation induced an adaptive proteomic shift toward low-arginine codon–containing genes, including specific amino acid transporters, and caused mutational evolution away from arginine codons—reducing translational bottlenecks that occurred during arginine starvation. Thus, environmental availability of a specific amino acid can influence DNA sequence evolution away from its cognate codons and generate altered proteins.

INTRODUCTION

Genomes of organisms are enriched in certain codons over others. The origins of such codon usage biases have been attributed to both sequence-specific mutational biases that are thought to dominate over long time scales and to organism-specific transfer RNA (tRNA) availabilities as encoded in the genomes of difference species (*1*). However, the challenge inherent to observing the emergence of such a long time scale process has precluded definitive support for various proposed models. The mechanisms underlying the emergence of codon usage bias, including the extent to which tRNAs shape genomic evolution, have also remained poorly defined. We reasoned that for cancer cells, which divide rapidly, acquire mutations more frequently relative to normal cells, and are exposed to a variety of selective pressures, the evolution of DNA sequence biases would be expedited. This would allow us to detect the emergence of codon-based sequence changes and search for potential underlying mechanisms.

[1]Laboratory of Systems Cancer Biology, The Rockefeller University, New York, NY, USA. [2]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Proteomics Resource Center, The Rockefeller University, New York, NY, USA. [4]Department of Systems Biology, Columbia University Medical Center, New York, NY, USA. [5]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. [6]Department of Biological Sciences, Columbia University, New York, NY, USA.
*Corresponding author. Email: st2744@cumc.columbia.edu (S.T.); sohail.tavazoie@rockefeller.edu (S.F.T.)
†Present address: UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, USA.
‡Present address: National Cancer Institute, National Institutes of Health, Bethesda, MD, USA.

RESULTS

Arginine codons and residues are frequently lost through mutation across human cancers

To determine whether specific codons or amino acids are favored in cancer genomes, we computationally assessed codon-switching events—defined as the gain or loss of a codon via mutation—across all cancers in The Cancer Genome Atlas (TCGA) (*2*). Although dozens of mutational signatures have been shown to be operant with varying weights in different cancers (*3*), we observed that the majority of cancers displayed unexpectedly similar patterns of codon gains and losses (fig. S1A). Notably, when collapsed onto their cognate amino acids, we observed that arginine codons were universally depleted across all cancer types (Fig. 1A and fig. S1B). Thus, mutagenic events affecting arginine codons are extremely frequent in cancer.

Mutational processes have been shown to act more frequently at specific nucleotides based on their surrounding contexts (*3*, *4*). It is therefore important to distinguish between observed codon changes that simply resulted from sequence-specific mutational biases versus codon-switching events that arose from evolutionary selection for or against a given codon. To distinguish between these possibilities, we devised a series of computational simulations that used cancer-specific mutational signatures to model expected codon-switching events (fig. S2). We extracted mutational signatures from the noncoding regions of cancer genomes to build an unbiased model, reasoning that mutations arising in nontranslated regions would be less affected by selective pressures, such as tRNA or amino acid availability, which would be unique to protein-coding genes. Consistent with this possibility, we observed notable disparities in the frequencies of different mutations between coding and noncoding regions of the genome (fig. S3). Modeling codon changes using mutational spectra derived from the noncoding genome further highlighted arginine codon–switching events as

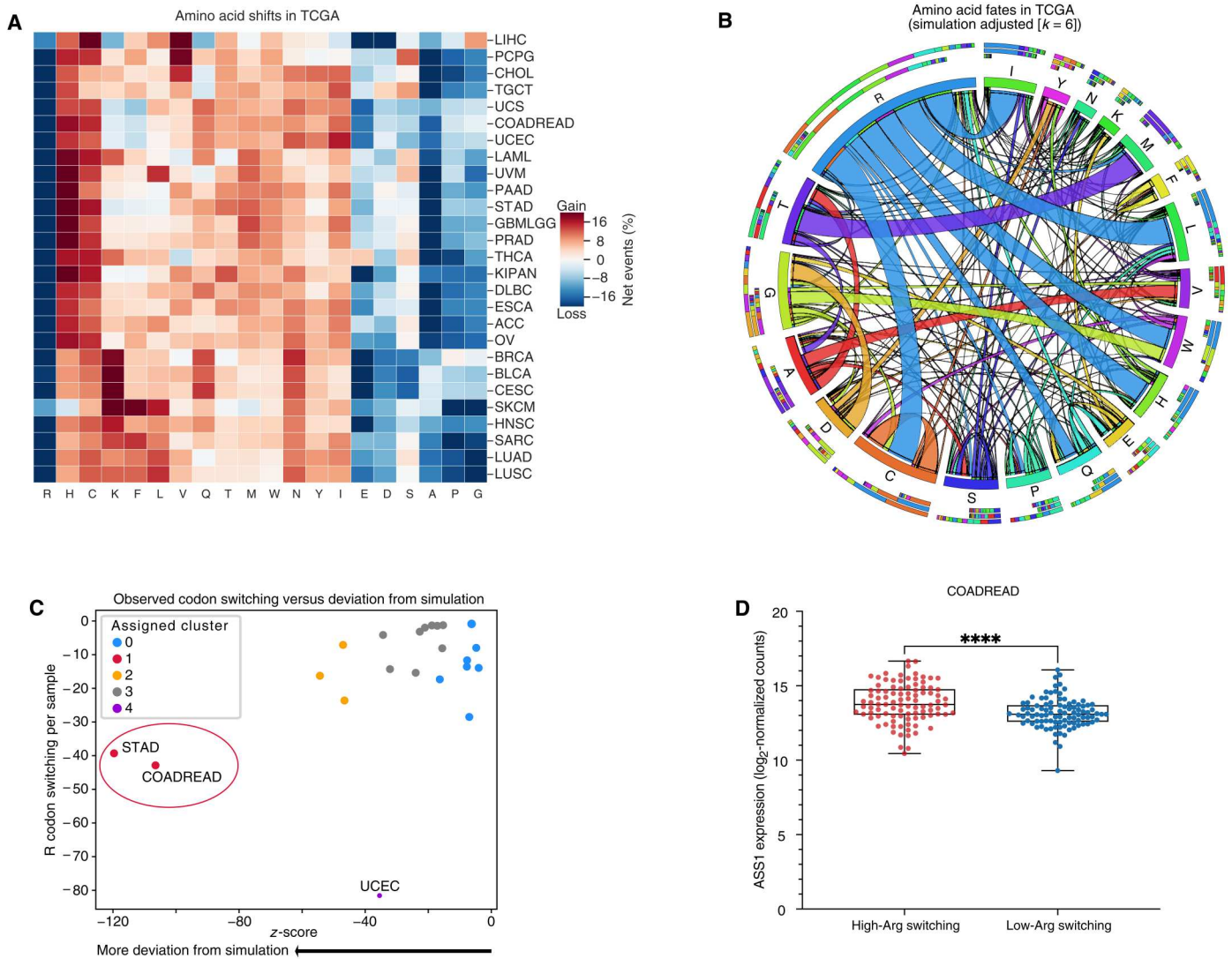

Fig. 1. Arginine codons and residues are frequently lost and are associated with an increase in *ASS1* expression. (**A**) Heatmap depicting codons gained (red) and lost (blue) across the TCGA. Gains and losses are normalized to the total number of missense and silent mutation events per sample for each cancer type. (**B**) Qualitative chord diagram showing amino acid switching events in cancer after adjustment from simulations. Ribbons that directly touch a column segment indicate loss of that specific amino acid codon during a mutational event and gain of the corresponding amino acid codon in which the ribbon terminates. Ribbons that begin and end at the same amino acid represent synonymous mutations. (**C**) Arginine codon–switching events observed versus predicted. Clusters were assigned with Affinity Propagation. (**D**) *ASS1* expression in colorectal cancer (CRC) samples with either a high-degree or low-degree of arginine codon–switching events (n = 96 per group) with whiskers denoting minimum and maximum values. (DESeq2, ****$P_{adjusted} < 0.0001$).

being especially overrepresented in coding genes (Fig. 1B and fig. S4). These findings support the possibility that certain codon-switching mutational events in the coding genome may confer selective fitness to cells. At the codon level, the most frequently lost codons across all cancers were arginine codons: CGG, CGA, CGC, and AGA, with frequent conversions to CAC (histidine), TGC (cysteine), ATA (isoleucine), and CTA (leucine) (fig. S4). Thus, arginine codon–switching events in the coding genome are generally overrepresented even when one considers mutational biases.

Next, to identify the tumor types and potential stresses associated with arginine codon mutational loss, we first compared our computational predictions with biological observations for each tumor type. This revealed that stomach adenocarcinoma (STAD) and colorectal adenocarcinoma (COADREAD) tumors were the most enriched in arginine codon–switching events and most notably deviated from the simulated background expectation compared to other cancer types (Fig. 1C). In contrast, although endometrial carcinomas (UCEC) exhibited the highest degree of arginine codon–switching, these events were relatively well-accounted for by sequence-specific mutational biases. Thus, the extent to which arginine codon–switching mutations occur in excess varies based on tumor type and is most overrepresented in colorectal and stomach cancers.

Increased arginine codon loss associates with expression of bioenergetic pathways

Because mutations involving arginine codons are especially overrepresented in colorectal and gastric cancers, we sought to identify a common pattern between the two. We hypothesized that a potential association with arginine codon loss could be extracellular arginine availability, because arginine is known to become limiting in tumor microenvironments (*5*, *6*) and the decoding of arginine codons requires this amino acid. Consistent with this, we observed that arginosuccinate synthetase 1 (*ASS1*), a critical gene that catalyzes the penultimate step of arginine biosynthesis, was overexpressed in both colorectal and gastric adenocarcinoma samples that exhibited high-arginine codon–switching events relative to those exhibiting low-arginine codon–switching events (Fig. 1D and fig. S5A). ASS1 has been shown to be variably expressed in tumors and can be induced when arginine becomes depleted from the tumor microenvironment (*7*, *8*). These data suggest that tumors with increased arginine codon–switching events may have experienced reduced extracellular arginine bioavailability during their development, which would have necessitated the expression of arginine biosynthesis pathway components such as *ASS1* for survival.

We next asked whether tumors that underwent a high frequency of arginine codon–switching events share common transcriptional programs beyond arginine metabolism. To answer this, we analyzed tumor transcriptomes at a global level using a mutual information-based framework (*9*). We found that in both colorectal and stomach adenocarcinomas, expression levels of genes belonging to S phase of the cell cycle, DNA replication, nucleotide metabolism, mitochondrial translation, and energetics (glycolysis, the citric acid cycle, and electron transport) pathways were significantly correlated with

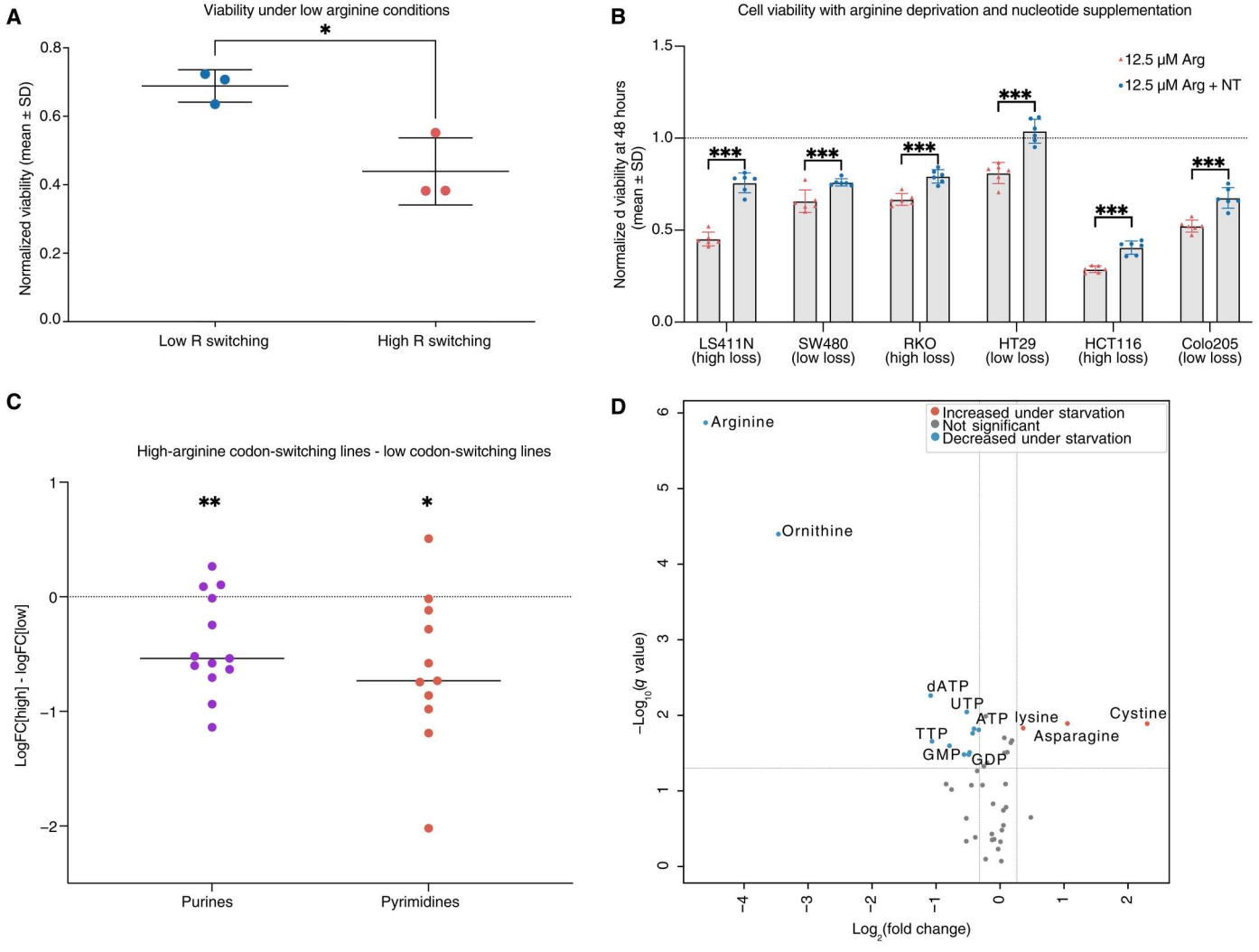

Fig. 2. Arginine codon losses are associated with increased dependence on extracellular arginine and nucleotide pool instability during starvation. (**A**) Cell line viability (means ± SD) under low-arginine (12.5 μM) conditions. (n = 3 per group). (**B**) Effect of nucleotide supplementation on colon cancer cell viability with arginine deprivation (n = 6 per group, two-tailed t test). (**C**) Metabolite profiling differences after exposure to low-arginine concentrations for 24 hours. Each point represents a purine/pyrimidine pathway metabolite and is the average $\log_2$ fold change ($\log_2$FC) difference between high-arginine codon–mutated lines and low-arginine codon mutated lines (one-sample t test with μ0 = 0). (**D**) Volcano plot of metabolite changes following arginine deprivation. Only detected citric acid cycle, urea cycle, amino acids, and nucleotide intermediates are labeled. (*P < 0.05, **P < 0.01, and ***P < 0.001). NT, nucleotides; TTP, thymidine 5'-triphosphate; UTP, uridine 5′-triphosphate; GMP, guanosine 5′-monophosphate; GDP, guanosine diphosphate.

increased arginine codon–switching events (fig. S6). To determine which of these pathways and processes are relevant to the in vivo microenvironment, where arginine levels can be substantially limiting (*6*, *10*, *11*), we conducted a similar analysis on colon and gastric adenocarcinoma cells in the cancer cell line encyclopedia (CCLE) (*12*), where cells were cultured in media containing excess arginine at least over five times circulating plasma levels—0.399 mM versus 0.074 mM, respectively (table S1) (*13*). Expression of genes belonging to nucleotide metabolism and bioenergetic pathways was selectively modulated in in vivo arginine codon–switching tumors but not in cancer cells growing in vitro with excess arginine (fig. S6). Consistent with this, the essentiality of genes belonging to bioenergetic and purine metabolism pathways was correlated with increased arginine codon loss in cancer cell lines (fig. S7). In sum, our findings suggest that provision of arginine to supraphysiologic levels, as is the case for in vitro culture of CCLE cells, may reduce cellular dependence on expression of certain bioenergetic (mitochondrial translation and electron transport) and nucleotide metabolism pathways relative to the in vivo arginine-limiting tumor context.

Arginine limitation causes nucleotide pool imbalances

Our observations collectively support a model whereby a subset of tumors facing arginine restriction experience perturbations to energy metabolism and nucleotide synthesis. Perturbed nucleotide synthesis can give rise to nucleotide imbalance and, in turn, increase base misincorporation rates, thereby accelerating mutagenesis and potentiating codon-switching events. Arginine metabolism derangements have been shown to affect nucleotide biosynthesis and potentially result in DNA damage (*14*–*16*). To further define the relationship between arginine codon–switching events and arginine and nucleotide metabolism, we collected a panel of colorectal cancer (CRC) cell lines that were either of the high-arginine codon loss type or the low-arginine codon loss type based on mutational sequence analysis of the CCLE (table S2). We observed that at low concentrations of arginine, within the range reported for tumor core levels (*5*), high-arginine codon loss cell lines exhibited significantly lower viability than low-arginine codon loss lines (Fig. 2A). To test whether arginine deprivation results in nucleotide metabolism stress, we performed rescue experiments with extracellular nucleotide supplementation. We observed that while CRC cell lines exhibited variably impaired growth at low-arginine concentrations, there was universal partial rescue of cell viability with nucleotide supplementation with purine supplementation being dominant for this effect (Fig. 2B and fig. S5, B and C). Thus, increased arginine codon switching is associated with heightened dependence on arginine availability with viability being partially rescued by provision of exogenous nucleotides. Because nucleotide supplementation conferred a survival advantage under low-arginine conditions, we sought to define how arginine metabolism affects intracellular nucleotide concentrations. Metabolomic profiling of CRC cells revealed that arginine deprivation caused depletion of both purine and pyrimidine nucleotides with a greater reduction in high-arginine codon–mutated lines relative to low-arginine codon–mutated lines (Fig. 2, C and D, and fig. S8). It has been suggested that arginine deprivation can affect nucleotide pools through induction via activating transcription factor 4 (ATF4) of asparagine synthetase (ASNS), which converts aspartate to asparagine (*16*). In support of this, arginine restriction induced ASNS in CRC cells (fig. S9). Because aspartate is a critical precursor for nucleotide synthesis, its shunting toward asparagine under arginine starvation would impair nucleotide synthesis. These findings reveal that arginine restriction causes nucleotide pool imbalance in CRC cells that contributes to impaired survival.

Arginine limitation causes an acute arginyl tRNA repression response

Our findings reveal that arginine limitation of CRC cells that are more reliant on extracellular arginine alters nucleotide pool balance, which can potentially cause mutations. These findings, however, do not explain why arginine codon–switching events are enriched in specific tumors. We thus focused on the association between arginine availability and arginine codon–switching events. Metabolic perturbations such as oxidative stress and glutamine deprivation were recently shown to reduce the levels of specific charged tRNAs—inhibiting translation of downstream genes (*17*, *18*). Furthermore, complete elimination of arginine from the environment has been shown to induce ribosome pausing at arginine codons in bacteria and in mammalian cells in vitro—repressing global protein synthesis (*19*, *20*). Because arginine codon–switching mutations would theoretically lessen the requirement for arginine tRNAs during protein translation, we hypothesized that arginine codon–switching events may facilitate gene expression under conditions where arginine becomes limiting. Arginine codon–switching mutations tended to occur in higher-expressed genes in patient samples, highlighting the possibility that arginine codon–switching events might have an outsized influence on gene translation (fig. S10). In such tumors, switching to non–arginine codons may facilitate gene translation in contexts where environmental arginine is scarce. We therefore sought to quantify how arginine deprivation affects availability of arginine tRNAs. We assessed tRNA levels in colorectal and gastric cancer cells following arginine limitation through Northern blotting. Arginine limitation acutely and markedly depleted arginine tRNA levels (Fig. 3A). We detected a significant reduction in multiple arginine tRNA isodecoders including $tRNA^{Arg}_{UCG}$, $tRNA^{Arg}_{UCU}$, and $tRNA^{Arg}_{CCG}$ within 24 hours of arginine deprivation. We did not observe reduced levels of other tRNAs such as $tRNA^{Leu}_{UAG}$, $tRNA^{Tyr}_{GUA}$, or $tRNA^{His}_{GTG}$ upon arginine restriction. These results were further confirmed through the use of tRNA quantitative polymerase chain reaction (qPCR) (fig. S11) as well as in gastric cancer and breast cancer cell lines (fig. S12). Moreover, time course analyses revealed that the affected arginine tRNAs became repressed within 2 hours of arginine limitation and generally reach a steady state within 4 hours (fig. S13). Thus, extracellular arginine restriction causes an acute and substantial reduction of arginine tRNA levels in CRC cells.

We hypothesized that one mechanism that could contribute to arginine tRNA repression may be reduced tRNA aminoacylation in the setting of arginine limitation. Reduced aminoacylation of certain tRNAs has been shown to destabilize tRNAs (*21*). To test this, we inhibited aminoacylation in CRC cells by depleting the arginyl-tRNA synthetase (RARS) and quantified tRNA levels. Suppressing arginine aminoacylation substantially suppressed expression of multiple arginyl tRNAs (fig. S14). These findings are consistent with arginine limitation causing reduced arginyl-tRNA charging and consequently contributing to degradation or destabilization of arginyl tRNAs.

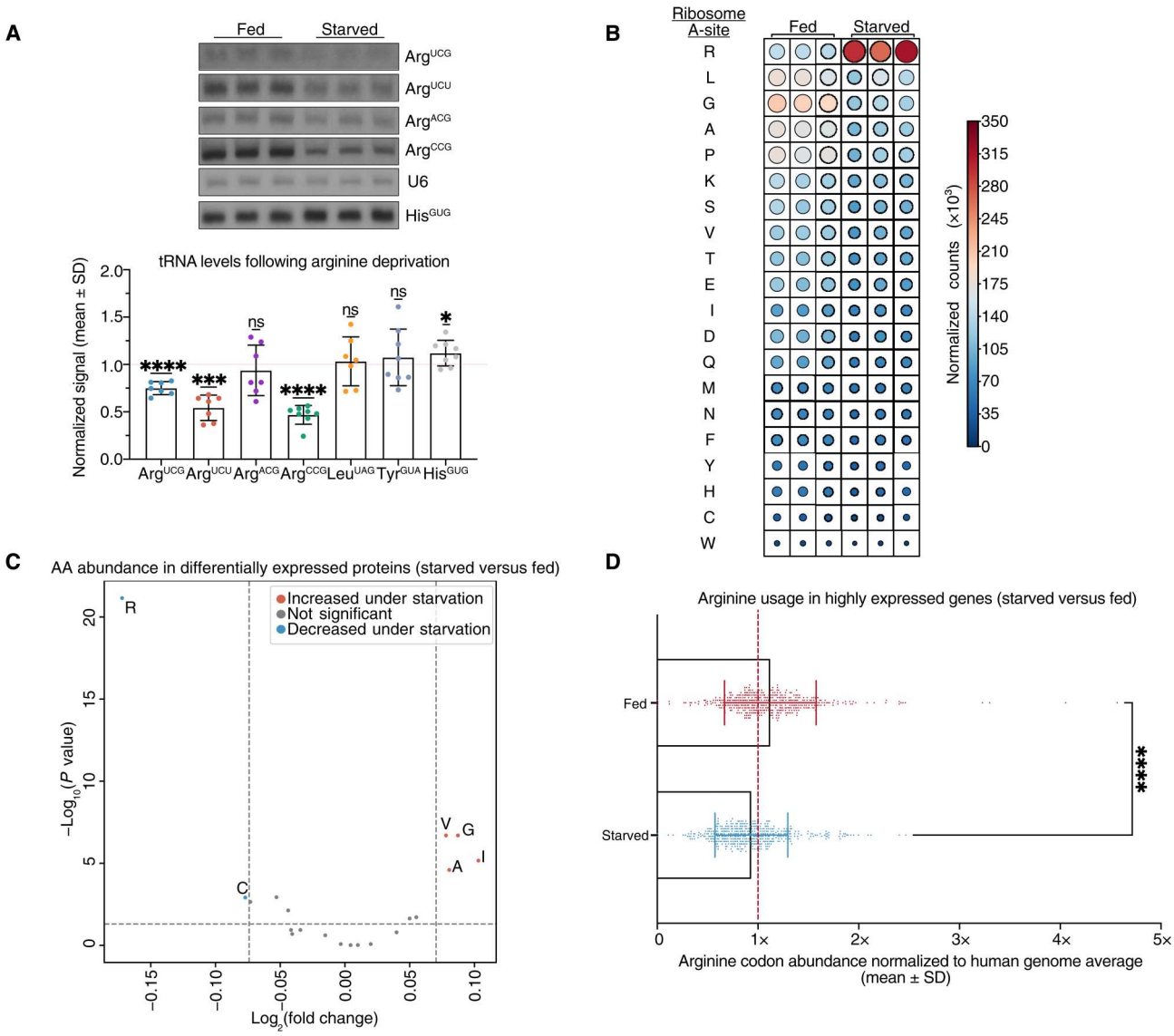

Fig. 3. Arginine deprivation reduces arginine tRNA availability, increases arginine ribosome localization, and reduces arginine usage in the tumor proteome. (**A**) tRNA quantification as assessed via Northern blot. Each dot represents the average abundance in an independent colon cancer or gastric cancer cell line ($n = 7$ per group, one-sample t test with $\mu 0 = 0$). (**B**) Ribosome A-site localization counts from ribosomal profiling experiments under starved or fed conditions. Circle size is scaled to counts. (**C**) Amino acid (AA) usage in genes that are highly expressed under fed or starved states. (**D**) Arginine codon abundance in genes expressed in fed or starved states ($n \geq 450$ per group, two-tailed Mann-Whitney test). Proteins are stratified on the basis of the top 10% most changed in either fed or starved states. (*$P < 0.05$, ***$P < 0.001$, and ****$P < 0.0001$). ns, not significant.

Arginine restriction causes ribosomal stalling at specific arginine codons

Marked reductions in arginine tRNA availability would be expected to impair arginine codon–dependent translation. To quantify how arginine deprivation–mediated tRNA changes affect gene translation, we performed ribosomal profiling (*22*, *23*). As expected, arginine starvation substantially increased ribosomal occupancy at arginine codons under starvation conditions (Fig. 3B). Such increased ribosomal A-site localization over arginine codons upon arginine restriction is consistent with increased stalling at arginine codons. As orthogonal approaches for assessing ribosomal dynamics, we used two additional metrics to quantify ribosome stalling events. We first calculated Consistent Excess of Loess Predictions (CELP) coefficients to measure the degree of stalling at all codons (*24*). This analysis further confirmed global and marked increases in stalling at arginine codons upon arginine deprivation (fig. S15A). Second, we calculated the frequency of amino acid appearances immediately upstream and downstream of maximal ribosome stalling sites during arginine deprivation and observed that arginine codons were notably overrepresented near the global stalling maxima of transcripts, on average appearing more than twice as often as expected (fig. S15B). In contrast, we found no such evidence of arginine enrichment near stalling sites during arginine-replete conditions (fig. S15C). These findings reveal that arginine

limitation at pathophysiologic levels substantially increases ribosome stalling events at arginine codons.

Next, to understand how codon-switching events influence ribosome dynamics, we compared ribosome localization in specific genes that were heterozygous for single-nucleotide variants (SNVs) (due to arginine codon–switching at one allele) under arginine-fed and arginine-deplete conditions. This experimental model provided us wild-type and mutant arginine codon endogenous "reporters" for specific genes. Variant alleles that underwent codon switching away from arginine codon usage showed significantly less ribosome stalling under arginine limitation at those specific codon positions compared to their corresponding wild-type alleles in the same cell (figs. S16A and S17). By comparison, codon-switching events that only involved non–arginine codons showed no differences in ribosome stalling at the wild-type versus variant alleles (fig. S16B). Consistent with these observations, genes that harbored arginine codon changes generally showed less stalling at multiple arginine codons under arginine starvation conditions (fig. S18A) and consequently higher translational efficiency compared to genes that were wild type with respect to arginine codon mutation status (fig. S18B). These results demonstrate that codon-switching events—specifically, the loss of rate-limiting codons—can directly influence ribosome localization dynamics under amino acid limitation. Therefore, whereas a direct consequence of arginine starvation–mediated tRNA changes is increased ribosome stalling at arginine codons, mutations that result in a loss of an arginine codon tend to relieve this translational bottleneck.

Arginine limitation causes a proteomic shift from arginine rich to arginine low proteins

Substantial stalling of arginine translation would be predicted to alter arginine utilization in the tumor proteome. We thus performed tandem mass tags (TMT)–based quantitative proteomics under conditions of arginine excess versus limitation and found that arginine deprivation resulted in a shift in the tumor proteome toward proteins with substantially lower arginine content, findings that were validated by Western blotting for multiple differentially regulated proteins (Fig. 3C and fig. S19). Moreover, arginine usage in highly expressed genes was highly significantly reduced upon arginine restriction (Fig. 3D). Pathway enrichment analysis of proteomic changes revealed that proteins related to amino acid transport and DNA damage–induced senescence were increased, whereas proteins related to DNA strand elongation and interferon-α/β signaling were reduced upon arginine restriction (fig. S20A). Notably, proteins that were up-regulated in these pathways upon starvation also tended to show reduced arginine codon content (fig. S20B). At the codon level, the three most affected codons with respect to underutilization in proteins were all arginine codons (fig. S21). Thus, arginine deprivation promotes induction of multiple gene expression programs that use arginine less frequently. The substantially decreased need for arginine within multiple gene sets that respond to arginine limitation, such as amino acid transport and synthesis, suggests that the amino acid requirements for expression of these gene sets may have undergone prior evolutionary selection to allow the continued expression of specific adaptive stress response programs when arginine is limiting. To assess whether this arginine restriction induced proteomic shift is adaptive for cells to respond to reduced arginine availability, we performed loss-of-function experiments targeting of multiple amino acid transporters that were up-regulated upon arginine limitation (fig. S22). RNA interference–mediated depletion of solute carrier family 7 member 1 (SLC7A1) significantly reduced colon cancer cell growth in the context of arginine restriction relative to arginine-replete conditions (fig. S22, B and C). SLC7A1, also known as CAT-1, has been shown to transport amino acids including arginine (*25*). Its translational induction upon arginine restriction and the positive impact of this induction on fitness of cells upon arginine restriction support an adaptive role for this tRNA repression–mediated translational response to arginine limitation. When coupled with the ribosomal profiling data, these findings suggest that during arginine scarcity, when arginine tRNAs become limiting, there may be an evolutionary advantage for tumors that have undergone additional codon-switching events from arginine codons to codons for which cognate tRNAs remain available for usage in translation.

Our findings thus far reveal that arginine restriction causes an acute response whereby arginyl tRNAs become repressed, leading to ribosomal stalling at rate-limiting arginyl codons of highly expressed genes. This is associated with a proteomic shift away from arginine-rich proteins toward arginine-low proteins, which includes amino acid transporters. Arginine restriction also causes nucleotide imbalance, accelerating mutagenesis. We hypothesized that over longer time scales, this context selects for cancer cells that have undergone arginine codon mutational switching events, which enables translation of proteins that are adaptive for survival.

Arginine restriction is sufficient to cause arginine codon–switching evolution in vitro

We next sought to determine whether arginine restriction is sufficient to causally drive codon-switching events away from arginine. To do this, we conducted laboratory evolution experiments by culturing colon cancer cells (RKO, SW480, and HT29) under reduced arginine conditions and assessing genomic codon-switching events using whole-exome sequencing (Fig. 4A). Iterative passaging of multiple CRC cell lines over eight passages (~24 population doublings lasting ~2 to 3 months) caused a significant increase in arginine codon–switching events in arginine-restricted cells relative to control cells that were passaged the same number of times under arginine-rich conditions (Fig. 4B). Consistent with our prior observations that arginine deprivation results in nucleotide pool imbalances, potentially accelerating mutational rate, arginine deprivation was associated with an increase in general mutational load (fig. S23). Notably, arginine codon mutations were increased in genes up-regulated during arginine deprivation, as identified from our prior proteomic experiments, compared to genes that were highly expressed during the arginine fed state (Fig. 4C). In contrast, the rate of histidine mutational losses between the gene sets was not significantly different (Fig. 4D). Thus, arginine deprivation in vitro is sufficient to increase the frequency of arginine codon–switching mutational events.

Arginine limitation causes arginine codon–switching evolution in vivo

We next asked whether we could recapitulate arginine codon–switching events in vivo and whether tumor propagation in a microenvironment low in arginine could also elicit such codon-switching events. We specifically focused on the liver microenvironment due to the liver being the organ in which the arginine

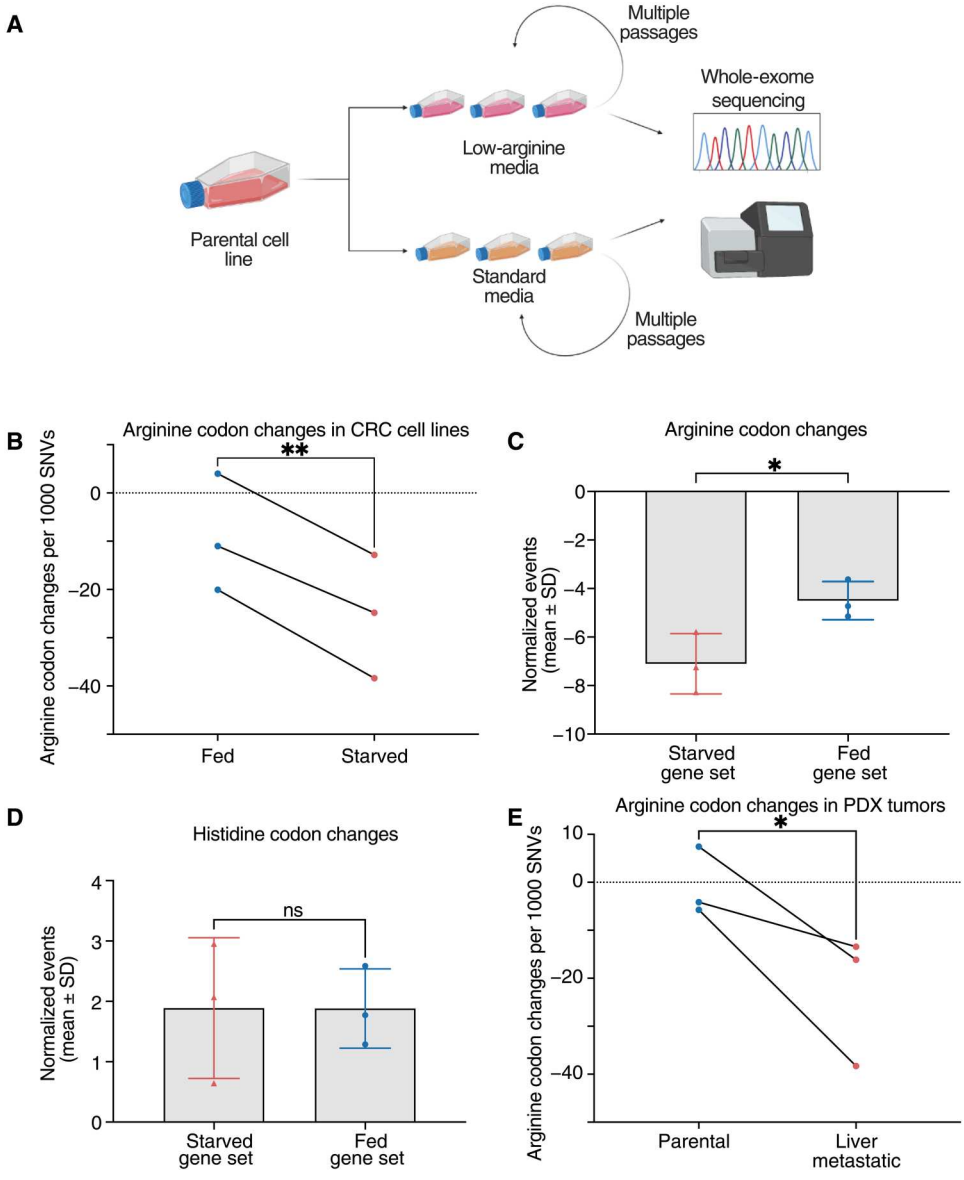

Fig. 4. Arginine deprivation promotes arginine-losing mutations. (**A**) Schematic of arginine deprivation experiments. (**B**) Arginine codon changes in cells serially passaged in either full media or low-arginine media ($n = 3$ per group, two-tailed paired t test). (**C**) Arginine codon changes in proteins that are increased during the fed or arginine-starved states ($n = 3$ per group, two-tailed t test). (**D**) Histidine codon changes in proteins that are increased during the fed or arginine-starved states ($n = 3$ per group, two-tailed t test). (**E**) Arginine codon changes in patient-derived xenograft (PDX) tumors that underwent multiple rounds of in vivo liver metastatic selection ($n = 3$ per group, one-tailed paired t test). (*$P < 0.05$ and **$P < 0.01$).

degrading enzyme, arginase, is most highly expressed (*26*) and also because the liver is a frequent and pathophysiologically relevant site of distant organ metastatic relapse in both colorectal and gastric cancers (*27*, *28*). We first analyzed metabolite profiling data of highly liver metastatic patient-derived xenograft (PDX) tumors that had undergone at least five rounds of in vivo selection for liver colonization (*29*) and observed that arginine was indeed the lowest abundance-free amino acid in the highly liver metastatic tumors compared to the parental tumors (fig. S24). We next conducted whole-exome sequencing of additional PDX tumors and observed that the rate of acquisition of arginine codon mutations was significantly increased in tumors that had undergone serial rounds of in vivo liver colonization selection compared to the rate measured in the parental tumors (Fig. 4E). Thus, reduced arginine bioavailability in vivo is associated with an increased rate of arginine codon mutations, mirroring our observations in vitro. These findings as a whole reveal that limitation of a single amino acid, arginine, results in multiple consequences in CRC cells. First, arginine limitation causes nucleotide pool imbalances and increased mutational rate. Concurrently, arginine tRNA levels are reduced, resulting in ribosomal stalling at arginine codons, providing a selective pressure against arginine codon usage and providing an evolutionary advantage to cancer cells whose coding genomes require less arginine for translation of highly expressed genes

required for growth. This context selects for cancer cells that have undergone arginine codon mutational switching events in the coding regions of such growth-promoting genes. On the basis of the totality of these observations, we propose that limitation of an amino acid (arginine) can causally increase the rate of mutations of its cognate codons in the cancer genome—facilitating the continued translation of proteins that can be adaptive for responding to the specific amino acid restriction and leading to the generation of proteins with arginine substitutions (Fig. 5).

DISCUSSION

The acquisition of somatic mutations contributes to the development of cancers de novo, the emergence of treatment resistance, and can predict response to immunotherapy (*30*–*33*). Understanding the mechanisms that drive the acquisition of mutations remains an important problem in cancer biology and oncology. Environmental contributions to mutational processes have generally been thought of as foreign additions to a system: Examples include ultraviolet radiation, tobacco smoke, or aristolochic acid. Our work suggests that environmental limitation, i.e., absence or restriction, of just a single amino acid can drive switching away from specific codons in the human cancer genome by simultaneously enhancing mutagenesis and altering specific cognate tRNA availability. In yeast, genetic defects in nitrogen metabolism can increase mutational rates in strains with heightened mutagenic backgrounds (*34*). Moreover, genetic defects in the urea cycle, a critical downstream pathway in the utilization of intracellular arginine, can result in altered rates of pyrimidine synthesis and affect mutational spectra (*14*, *15*). While these studies have focused on genetically driven defects in metabolism resulting in mutagenesis, our findings reveal that availability of a specific environmental nutrient, arginine, can filter the mutational landscape of cancer cells in a codon-dependent manner and drive them toward acquisition of arginine codon mutations. Our findings of a rapid repression of arginine tRNAs upon arginine limitation and the induction of an arginine-low tumor proteome suggest the existence of an acute tRNA-mediated stress response to arginine restriction that promotes translation of genes with reduced arginine codons. Others have also shown that for a given tRNA, distinct isodecoders associate with proliferation versus differentiation states (*35*). While mammalian target of rapamycin (mTOR) signaling and the integrated stress response (ISR) pathways may certainly become activated upon arginine limitation and contribute to global translational deregulation, our perturbations yielded arginine limitation at physiological levels similar to that observed in tumors rather than eliminating extracellular arginine entirely. Moreover, while these other responses may certainly become activated upon arginine limitation, the mTOR and ISR pathways do not mediate codon-specific response and instead mediate global translational repression responses (*36*). Thus, while these other stress pathways likely contribute to alterations in protein translation, the codon-specific effects and the DNA evolution response we are observing reveal a codon-specific pathway being involved—an arginine tRNA repression/ribosomal pausing/DNA evolution response caused by arginine limitation that is adaptive. Specifically, in colon cancer cells, limitation of arginine causes an acute translational shift toward an arginine low proteome. We provide evidence that this shift is remarkably adaptive by identifying an arginine transporter that becomes translationally up-regulated and that its induction provides a fitness advantage to cells under arginine-limiting conditions. Prior work in bacteria and in mammalian cells in vitro had shown that complete elimination of arginine from the environment can cause ribosome pausing and global translational repression that was proposed to be caused by reduced tRNA aminoacylation (*19*, *20*). Our findings across a series of colon cancer cell lines reveals that physiological limitation of arginine represses the levels of arginine tRNAs, an effect that could also be elicited upon repression of arginyl tRNA aminoacylation. Collectively, our findings are consistent with a combination of arginyl tRNA repression and reduced aminoacylation contributing to ribosomal pausing at cognate arginine codons and inducing a proteomic shift in response to arginine deprivation in colon cancer cells.

To our knowledge, this is the first demonstration of directed DNA evolution and selection against specific codons in response to a specific environmental perturbation. Our observations imply that over time, cancer cells growing in an arginine-scarce environment are likely to lose more arginine codons and suggest that in vitro systems currently used to study cancer and other diseases, for example, cells growing in tissue culture, are potentially susceptible to evolving away from arginine codons at different rates depending on their level of arginine supplementation, the fidelity of their DNA repair mechanisms, and the robustness of their arginine tRNA pool. Further work is required to understand whether other nutrient limitations also elicit DNA sequence evolution and to understand how other genetic and environmental factors, such as competition with the surrounding microbiome or the presence of inflammatory states interact with nutritional availability to affect DNA sequence evolution. With respect to arginine, it has already been suggested that free arginine is especially critical for regulating

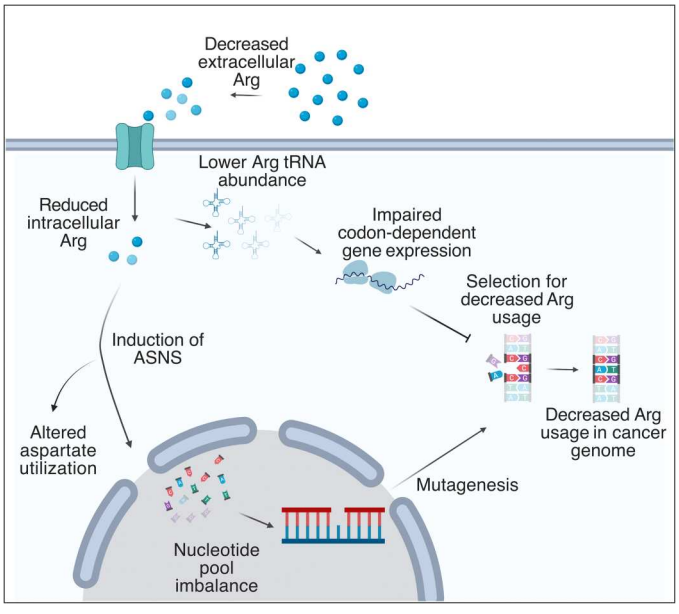

Fig. 5. Arginine deprivation drives a codon-dependent DNA sequence evolution response. A model depicting how arginine deprivation results in multiple consequences including nucleotide pool imbalances and impaired translation of specific arginine codons, ultimately resulting in the loss of arginine codons in CRC genomes.

cancer immune responses (*37*), thus competition for this common substrate may influence the evolution of the cancer genome, especially in contexts of tumors with high immune infiltration. Inflammatory bowel disease and *Helicobacter pylori* infections, precursor disease states with established epidemiological and pathophysiological links to the development of colorectal and gastric cancer, respectively, have both been shown to modulate arginine availability in affected tissues (*38*, *39*). Recent work has elegantly demonstrated that amino acid limitation can be so substantial under inflammatory signaling that cancer cells use alternative translational decoding for specific amino acids (*40*), leading to the production of altered proteins and neo-antigens. Our findings reveal that limitation of an amino acid can also elicit protein sequence changes and perhaps neo-antigen load via an alternative mechanism—DNA sequence codon-switching events. Notably, arginine limitation superimposed on a background of increased base misincorporation rates, such as in mismatch repair deficiency, would increase the probability of stochastically acquiring arginine codon mutations that may then confer a survival advantage and may partially contribute to the increased signal in some tumors over others. However, our experiments reveal that this process occurs in both mismatch repair–proficient and mismatch repair–deficient tumors and cell lines. Last, our findings reveal that codon-based mutations can potentially identify subsets of cancers that are more sensitive to restrictions of a specific amino acid. These findings have implications for dietary amino acid restriction approaches that have been tested in tumor models as well probiotic engineering approaches that can modulate tumoral amino acids (*41*–*45*). The codon-based genotype-dependent vulnerability described here suggests potential for use of codon-centric mutational spectra as biomarkers for emerging cancer metabolism oncologic therapies.

MATERIALS AND METHODS

Experimental design

Sample sizes were selected back on knowledge of intragroup variation and expected effect size. For in vitro experiments, sample sizes were chosen on the basis of prior knowledge on intragroup variation. Data were collected on the basis of predetermined endpoints (in vitro assays) or tumor burden exceeding 2000 mm^3. Experiments were carried out as biological replicates as noted in the text and figure legends and were generally repeated at least twice. Samples were allocated randomly if possible. No blinding was performed.

Codon mutation analyses

Mutation annotation files (.maf) corresponding to TCGA studies were downloaded from the Broad Firehose platform (http://firebrowse.org). When possible, we used combined cancer datasets (i.e., COADREAD, KIPAN, and GBMLGG). A script was written in Python (v. 3.8.5) to manually count codons lost and gained across the coding regions of cancer types and samples. For each cancer sample, the count for a codon was subtracted if it was lost through a missense or silent mutation and added if it were gained instead. An "event" was defined as the gain and loss of a pair of codons. We also used a similar framework to count the total flux between codons or amino acids to determine flux between codons and amino acids. Events were plotted using circos plots (*46*).

Derivation of null distributions

We used a Monte Carlo approach to derive various null distributions of codon and amino acid usage shifts across cancer types and different samples. Briefly, our algorithm scatters nucleotide mutations across reference gene sequences downloaded from ENSEMBL. Before input into the simulation, genes with multiple splice variants were filtered against the annotation of principle and splice isoforms (APPRIS) database to include only the highest-ranking principle splice variant for simulation (*47*). Probabilities of specific nucleotide mutations were weighted on the basis of the 5′ and 3′ contexts of each nucleotide (*3*).

For inputs into the analysis, we first downloaded the mutation calls from International Cancer Genome Consortium (last accessed February 2018) (*48*) and then cross-referenced the intergenic and intronic mutational calls with the reference genome to extract the 5′- and 3′-nucleotide contexts to infer mutational probabilities of different nucleotides under different contexts. Each cancer type was assigned its own unique mutational matrix. For each tumor sample in the TCGA, we created a corresponding in silico sample and constrained potential mutations to the same set of genes that are mutated in each specific TCGA sample. For each in silico sample, candidate genes were randomly mutated the same number of times as was observed in its matched TCGA sample. These specific constraints were placed to prevent the model from deviating due mutations being simulated on lowly mutated genes or genes with wildly different codon content compared to the original sample. For each gene, nucleotide positions are first hashed by 5′- and 3′ contexts and selected for mutation using a vectorized approach to randomly select possibilities along the entire transcript. The effects on the gene (codon change and amino acid change) were calculated and used for downstream analyses. Each sample was simulated a thousand times (n = 1000).

For statistical inference, we created a "null distribution" mean for different codon/amino acid gains/losses by populating the dataset with mean inferences from each individual TCGA sample. A log-rank test was then performed to determine the extent to which the observed TCGA dataset was different from the simulated dataset. Heatmaps were generated using the Seaborn library in Python (*49*). Chord diagrams were generated using both the observed datasets and simulation data using circos (*46*). Qualitative circos plots were generated using scaled values following the formula provided by the developer using the following equation: $(e^{k*x/\max(x)} - 1)/(e^k - 1)$, where k is the scaling factor, x is the ratio of the observed shift to the simulation mean for the specific shift, and max (x) is the maximum test statistic across the entire simulation.

Gene expression analysis in arginine codon switching samples

Tumors from TCGA were assigned as high- or low-arginine codon–switching using the in silico model described above and assigned a z score based on the number of deviations from expectation for each sample. The top and bottom 20% of samples were assigned as high switching and low switching, respectively. To determine whether *ASS1* expression is differentially expressed between tumors with high- or low-arginine codon–switching, raw counts from RNA sequencing were obtained using the TCGAbiolinks package in R (*50*–*52*), and subsequent normalization and differential gene expression analysis between high- and low-arginine codon–switching groups

were performed using DESeq2 (*53*, *54*). For graphical purposes, DESeq2 log-normalized counts were plotted with the DESeq2 *P* value (adjusted for multiple comparisons across all genes).

To contrast gene expression patterns between in vivo and in vitro cancer cells, Spearman correlation coefficients were calculated between median-of-ratios normalized count data and the codon-switching scores assigned from simulation and then ranked on the basis of strength of correlation for mutual information analysis with information-theoretic pathway-level analysis of gene expression (iPAGE) (*9*). For CCLE samples, RNA sequencing count data were obtained from the Cancer Dependency Map version (most recently processed with release 22Q2) (*55*) and also normalized with median of ratios using DESeq2. Codon changes in colorectal and gastric cancer cell lines were calculated on the basis of corresponding mutational data that were obtained from the Cancer Dependency Map. Spearman correlation coefficients between gene expression and arginine codon loss were calculated and then input for mutual information analysis identical to how the TCGA samples were processed. For the iPAGE program, the independence flag was set to zero to allow for calculation of overrepresentation in the maximum number of pathways, and the ebins parameter was set to four. To graphically depict shared pathways, only genes in the top bin (corresponding to the top 25% of correlated genes) with pathway overrepresentation in both colorectal and gastric cancer datasets were selected for graphing, with pathways collapsed onto the most top-level statistically significant pathway in the Reactome hierarchy (*56*).

Analysis of common pathway dependencies in vitro

Gene essentiality scores from the Cancer Dependency Map version (22Q2) (*55*) were correlated with the arginine codon loss (based on whole-exome sequencing data from the same data release) for each cell line, and similar to above, Spearman correlation coefficients were used to rank genes for mutual information analysis.

Analysis of mutational events in TCGA RNA sequencing data

Raw counts from TCGA were obtained using the TCGAbiolinks package in R (*50*–*52*). Gene size was estimated using the GenomicFeatures package in R (*57*) to calculate gene size from exon length using the GRCh38.105 gene transfer format data file from ENSEMBL and used to calculate transcripts per million for each gene in each sample. Genes were then sorted and ranked within each sample. Mutational events in the top or bottom half of gene expression in each sample were counted by cross-referencing and matching sample barcodes to whole-exome sequencing data collected in the TCGA. High-arginine mutated and low-arginine groups were assigned as previously specified.

Arginine viability studies

Colon cancer cell lines were grown in arginine free media (Thermo Fisher Scientific, catalog no. A2493901; US Bio, catalog no. D9803-07B) supplemented with 10% dialyzed fetal bovine serum (FBS) (Thermo Fisher Scientific, catalog no. 26400044) and with arginine supplementation (Sigma-Aldrich, catalog no. A8094) to desired levels. Lysine (Sigma-Aldrich, catalog no. A8094) and bicarbonate were supplemented to Dulbecco's modified Eagle's medium (DMEM) reference levels (table S1). Cells were plated into 96-well plates (3000 cells per well), and cell viability was assessed with a luminescence-based assay (CellTiter Glo, Promega, catalog no. G7572) at 48 hours on a SpectraMax M3 plate reader (Molecular Devices). For nucleotide rescue experiments, nucleobases were supplemented at concentrations up to 10× reported physiologic concentrations (Sigma-Aldrich, catalog nos. A2786, C3506, G11950, and T0895) (*58*). All cell lines were periodically assessed for mycoplasma contamination by PCR for genomic DNA.

Cancer evolution experiments

Cell lines were grown under periods of intermittent arginine deprivation (12.5 μM) using arginine-free media (Thermo Fisher Scientific, catalog no. A2493901; US Bio, catalog no. D9803-07B) supplemented with L-arginine to desired concentrations (Thermo Fisher Scientific, catalog no. A2493901) and dialyzed FBS (Gibco, catalog no. 26400044). Starvation cycles consisted of 4-day starvation followed by rescue with standard DMEM and dialyzed FBS. Cell lines were starved for a total of 8 cycles and passaged at 1:10 ratios. In parallel, cell lines were maintained under standard tissue culture conditions and passaged to control for underlying genetic drift. At the end of the starvation cycles, DNA was extracted from both the starved and unstarved cancer cell lines (QIAGEN DNeasy Blood and Tissue Kit, catalog no. 69506) with RNAse A treatment (QIAGEN, catalog no. 19101) and sent for whole-genome sequencing at the New York Genome Center.

Arginine deprivation experiments

DMEM media with varying levels of arginine were prepared as described above. Cells were plated to approximately 20% confluence in standard DMEM media supplemented with 10% (v/v) FBS. At approximately 40% confluence, cells were washed three times with equal volume phosphate-buffered saline (PBS), and media were replaced with either control media [DMEM with standard amino acid concentrations and 10% (v/v) dialyzed FBS] or treatment media [DMEM with 12.5 μM arginine with 10% (v/v) dialyzed FBS]. Sample collection methods for respective experiments, i.e., Western blots, Northern blots, etc., are described in the respective sections. Unless specified otherwise, cell samples were collected at 24 hours after initiating starvation for downstream experiments.

Knockdown experiments

For knockdown of RARS, SMARTPool (Horizon Discovery, L-009820-02), small interfering RNAs (siRNAs) were used with Lipofectamine RNAiMAX transfection reagent (Invitrogen). Transfections were carried out with 20 nM siRNA following the manufacturer's instructions with Opti-MEM I (Invitrogen). Transfections were incubated for 4 days before RNA and protein collection. For amino acid transporter knockdown experiments, cells were transfected with SMARTPool siRNA targeting SLC7A5 (Horizon Discovery, L-004953-01), SLC7A1 (Horizon Discovery, L-007610-01), or SLC7A11 (Horizon Discovery, L-007612-01). For all knockdown experiments, a negative control was carried out using nontargeting siRNA (Horizon Discovery, catalog no. D-001810-10). Validation was performed using Western blot with the antibodies listed in the section "Western blots" below. For SLC7A1, SLC7A5, and SLC7A11, qPCR was also used to validate on-target specificity using the following primers: SLC7A1: 5′- CATCGCCTACTTTGGGGTGT, 3′- TAACCCGAGGCATGGGAAAC; SLC7A5: 5′- AACCCCTACAGAAACCTGCC, 3′- CATGACGCCCAGGTGATAGT; SLC7A11: 5′-

ATGCTGGCTGGTTTTACCTCA, 3′- CGCTCAGAAAAGGTCACTGC; glyceraldehyde-3-phosphate dehydrogenase (GAPDH): 5′- GAAGGTGAAGGTCGGAGTC, 3′- GAAGATGGTGATGGGATTTC. Gene expression was calculated using the $\Delta\Delta C_t$ method relative to GAPDH. Effects on cell viability were measured by first transfecting cells, subjecting to arginine limitation 24 hours later, and subsequently measuring cell viability using bioluminescence (CellTiter Glo, Promega, catalog no. G7572) 48 hours into arginine limitation.

Western blots

Protein lysates were extracted with ice-cold radioimmunoprecipitation assay buffer supplemented with protease and phosphatase inhibitors (Roche). Thirty micrograms of protein lysate was separated using SDS–polyacrylamide gel electrophoresis and transferred to a polyvinylidene difluoride membrane (Immobilion-P, Millipore, IPVH00010). After blocking the membranes in 5% bovine serum albumin (BSA) in tris-buffered saline with 0.1% Tween (TBST) [1× TBS (Cell Signaling Technology); 0.1% Tween 20 (Sigma-Aldrich)], the membranes were incubated overnight at 4°C. Antibodies used in this study were rabbit anti-ASNS antibody (Proteintech, 14681-1-AP) 1:1000 in 5% BSA (Sigma-Aldrich), mouse anti–β-actin antibody (Millipore Sigma, A5441) 1:5000 in 5% BSA, rabbit anti-RARS (Proteintech, 27344-1-AP) 1:1000 in 5% BSA, SLC7A1 antibody (LS Bio, LS-C749764) diluted 1:2000 in 5% milk, hypoxanthine phosphoribosyltransferase 1 (HPRT1) (Proteintech, 15059-1-AP) 1:4000 in 5% milk, SLC7A5 (Proteintech, 28670-1-AP) 1:8000 in 5% BSA, SLC7A11 (Proteintech, 26864-1-AP) 1:2000 in 5% BSA, and phosphoserine phosphatase (PSPH) (Proteintech, 14513-1-AP) 1:1000 in 5% BSA. Primary antibodies were incubated in 5% BSA in TBST overnight at 4°C. Secondary antibodies used included horseradish peroxidase (HRP)–conjugated goat anti-rabbit immunoglobulin G (IgG; H + L) or HRP-conjugated goat anti-mouse IgG (H + L) secondary antibody (Invitrogen). Membranes were incubated with enhanced chemiluminescence Western blot substrate (Thermo Fisher Scientific) for 1 min and then exposed to x-ray films (Fujifilm) that were then developed with a film processor (SRX-101A, Konica Minolta).

RNA isolation and purification

RNA was extracted from cells using TRIzol (Invitrogen) and isopropanol precipitation according to the manufacturer's instructions. After precipitation, the RNA pellet was washed twice with ice-cold freshly prepared 75% ethanol and then subsequently air-dried and resuspended in tris-EDTA buffer.

Northern blots

Purified RNA was run on 10% tris/borate/EDTA-urea gels at 200 V for 1 hour and transferred to a Hybond-N^+ membrane (GE Healthcare) at 150A for 1 hour. RNA was cross-linked to the membrane using ultraviolet radiation at 240 mJ/cm^2. Membranes were blocked with Oligo Hybridization Buffer (Ambion) for 1 hour at 42°C. Northern probes were labeled with ^{32}P adenosine triphosphate with T4 polynucleotide kinase (New England Biolabs) and purified with a G25 column (GE Healthcare). Probes were hybridized in Oligo Hybridization Buffer overnight at 42°C. Membranes were washed with 2× saline-sodium citrate(SSC), 0.1% SDS buffer, and 1× SSC 0.1% SDS before exposing film. Films were developed with exposure times adjusted based on the probe signal strength. Probe sequences were as follows: $tRNA^{Arg}_{UCG}$: 5′-GCCTTATCCATTAGGCCACGT-3′; $tRNA^{Arg}_{UCU}$: 5′-ATCCATTGCGCCACAGAGCC-3′; $tRNA^{Arg}_{ACG}$: 5′-CCGTAGTCAGACGCGTTA-3′; $tRNA^{Arg}_{CCG}$: 5′-CCGGAATCAGACGCCTTAT-3′; $tRNA^{His}_{GUG}$: 5′-AACGCAGAGTACTAACCACTATACG-3′; $tRNA^{Tyr}_{GUA}$: 5′-ACAGTCCTCCGCTCTACCAGCTGA-3′; $tRNA^{Leu}_{UAG}$: 5′-CTCCGAAGAGACTGGAGCCTAAA-3′; and U6: 5′-CACGAATTTGCGTGTCATCCTT-3′. Multiple probes were tried for $tRNA^{Arg}_{CCU}$; however, none yielded any detectable signal despite large yields of total RNA and strong signals for the other arginine tRNA. There is currently no clearly identified gene for $tRNA^{Arg}_{GCG}$, and therefore, Northern blots were not attempted for this tRNA (*59*, *60*). Membranes were stripped by washing with 0.1% SDS in boiling water followed by equilibration to room temperature. Subsequent probes were applied starting with reincubation with Oligo Hybridization Buffer and repeating all downstream steps with freshly labeled probes. Band intensity quantification was performed using ImageJ with the signal in each lane being normalized to U6.

tRNA qPCR

The tRNA qPCR protocol was adapted from previously published protocols for Y-shaped adapter-ligated mature tRNA sequencing (*61*). Briefly, cells were plated and then starved (12.5 μM arginine) or kept under fed conditions for 24 hours. Total RNA was subsequently extracted and then deacylated with 20 mM tris-HCl (pH 9.0) at 37°C for 40 min. Y-shaped adapters were ligated with mature tRNA in the total RNA by T4 RNA ligase 2. Adapters used in this study were as follows: Y-3-AD_UMI: 5′-P-GTATCCAGTNNNNTGGAATTCTCGGGTGCCAAGG-3′-ddC and Y-5-AD_UMI: GTTCAGAGTTCTACAGTCCGACGATCNNNNACTGGATACTGrGrN. Ligation reactions were carried out with 1 μg of total RNA at 37 °C for 2 hours and then 4 °C overnight. cDNA was synthesized by SuperScript IV Reverse Transcriptase (RT) (Thermo Fisher Scientific) with the common RT primer GCCTTGGCACCCGAGAATTCCA. The qPCR was performed with common RT primer and unique primers for different tRNAs. Fold gene expression was calculated relative to $tRNA^{His}_{GUG}$ using the $\Delta\Delta C_t$ method. Specific primers used in this study were as follows: $tRNA^{His}_{GUG}$: 5′-AGTGGTTAGTACTCTGCGTT-3′; $tRNA^{Leu}_{GAG}$: 5′- TAAGGCGCTGGATTTAGGCT-3′; mt-Arg: 5′-CAAAACGAATGATTTCGACTCA-3′; $tRNA^{Arg}_{CCG}$: 5′-ATAAGGCGTCTGATTCCGG-3′; $tRNA^{Arg}_{UCG}$: 5′-GCCTAATGGATAAGGCGTCTGACT-3′; $tRNA^{Arg}_{CCU}$: 5′-TGGCCTCCTAAGCCAGGGAT-3′; $tRNA^{Arg}_{ACG}$: 5′-AGTGGCGCAATGGATAACG-3′; $tRNA_{Arg}{}^{UCU}$: 5′- GGCTCTGTGGCGCAATGGAT-3′.

PDX propagation

PDXs were propagated with the methods similar to those previously published (*29*). Briefly, within 2 hours of surgical resection, CRC tumor tissue that was not needed for diagnosis was implanted subcutaneously into NSG mice (RRID: IMSR_JAX:005557), aged 6 to 10 weeks, at the Memorial Sloan Kettering Cancer Center (MSKCC) Antitumor Assessment Core facility in compliance with MSKCC IRB protocol 10-018A and The Rockefeller University IRB protocol STA-0681. When the tumor reached the predetermined endpoint of 1000 mm, the tumor was excised and transferred to the Rockefeller University. Xenograft tumor pieces of 20 to 30 mm^3 were reimplanted. When the subcutaneous tumor reached 1000 mm^3, the tumor was excised. The rest of the tumor was chopped finely with a

scalpel and placed in a 50-ml conical tube with a solution of DMEM (Gibco) supplemented with 10% (v/v) FBS (Corning), L-glutamine (2 mM; Gibco), penicillin-streptomycin (100 U/ml; Gibco), amphotericin (1 μg/ml; Lonza), sodium pyruvate (1 mM; Gibco), and collagenase type IV (200 U/ml; Worthington) and placed in a 37°C shaker at 220 RPM for 30 min. After centrifugation and removal of the supernatant, the sample was subjected to ammonium-chloride-potassium (ACK) lysis buffer (Lonza) for 3 min at room temperature to remove red blood cells. After centrifugation and removal of ACK lysis buffer, the sample was subjected to a density gradient with OptiPrep (1114542, Axis-Shield) to remove dead cells. The sample was washed in media and subjected to a 100-μm cell strainer and followed by a 70-μm cell strainer. Mouse cells were removed from the single-cell suspension via magnetic-associated cell sorting using the Mouse Cell Depletion Kit (Miltenyi, catalog no. 130-104-694), resulting in a single-cell suspension of predominantly CRC cells of human origin.

PDX metabolite profiling analysis

Metabolite profiling results from PDX samples were acquired from previously published data from our group (*29*) and processed identically to the publication.

Metabolite extraction and profiling

Metabolite profiling experiments were performed in collaboration with the Rockefeller University's Proteomics Resource Center (RRID:SCR_017797). Some of the following methods are similar to those previously published (*24*). Metabolite extraction and subsequent liquid chromatography (LC) coupled to high-resolution mass spectrometry (MS) for polar metabolites of cells were carried out using a Q Exactive Plus (Thermo Fisher Scientific). For all metabolite profiling, cells were washed with ice cold 0.9% NaCl and harvested in ice-cold 80:20 LC-MS methanol:water (v/v). Samples were vortexed vigorously and centrifuged at 20,000*g* at maximum speed at 4°C for 10 min. The supernatant was transferred to clean tubes. Samples were dried to completion using a nitrogen dryer.

Dried polar samples were resuspended in 60 μl of prechilled 50% (v/v) acetonitrile/water resuspension solvent, vortexed for 10 s, and centrifuged for 10 min at 4°C and 13,200 resolution/min, and then 14 μl from each sample was transferred to create a pooled sample. This pooled sample was further diluted with 1:3 and 1:10 dilution factors and used as biological quality control. Samples were analyzed in randomized order and at 5-μl injection volume via LC-MS system.

Polar metabolites were separated on a SeQuant ZIC-pHILIC 5-μm polymer (150 mm by 2.1 mm) column (EMD Millipore) connected to a Thermo Vanquish ultrahigh-pressure LC coupled to a Q Exactive Plus Hybrid Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) with a heated electrospray ionization source. Chromatographic separation was achieved by mixing mobile phase A consisted of 20 mM ammonium carbonate with 0.1% (v/v) ammonium hydroxide (adjusted to pH 9.3 with formic acid) and mobile phase B of acetonitrile in the following gradients: 90 to 40% B (0 to 22 min), held at 40% B (22 to 24 min), 40 to 90% B (24 to 24.1 min), and reequilibrated at 90% B (24.1 to 30 min) at a flow rate of 0.15 ml/min. MS data were acquired in polarity switching mode for both MS1 (full MS) and MS2 (data-dependent acquisition) with the following parameters: spray voltage, 3.0 kV; capillary temperature, 275°C; source temperature, 250°C; sheath gas flow, 40 arbitrary units (a.u.); auxiliary gas flow, 15 a.u. The full MS scans were acquired with 70,000 resolution, 1×10^6 ACG target, 80-ms max injection time, and a scan range of 55 to 825 mass/charge ratio. The data-dependent tandem MS scans were acquired at a resolution of 17,500, 1×10^5 ACG target, 50-ms max injection time, 1.6-Da isolation width, and stepwise normalized collision energy of 20, 30, and 40 U, with 8-s dynamic exclusion and a loop count of 2. Relative quantification of polar metabolites was performed in Skyline Daily (v.21.2.1.403) (https://skyline.ms/project/home/software/Skyline/begin.view) with the maximum mass error and retention time tolerance set to 2ppm and 12s respectively, referencing in-house retention time for polar metabolite standards.

Ribosome profiling

Cell lysis, ribosome footprint purification, and downstream library construction were performed according to previously published protocols 24 hours after starvation (*23*) with minor modifications. Namely, for harvesting the cells, the dishes were flash-frozen on liquid nitrogen after washing them with prechilled 1× PBS, and subsequently, the frozen cells were scraped off in cold lysis buffer on ice. In addition, because of the phasing out of the legacy Ribo-Zero Gold rRNA Removal Kit from Illumina, we used the RiboCop rRNA Depletion Kit for Human/Mouse/Rat (HMR) V2 (Legoxen, catalog no. 144.24) to deplete ribosomal RNA (rRNA), following the manufacturer's instructions to retrieve small RNAs after rRNA depletion using alcohol precipitation instead of the kit purification steps. In parallel, total RNAs were also isolated for downstream RNA sequencing and normalization for translational efficiency analysis using the TruSeq RNA Library Prep Kit v2 (Illumina).

Bioinformatic processing of ribosomal profiling experiments

For bioinformatics processing, cutadapt was used to remove the linker sequence AGATCGGAAGAGCAC (*62*), and the FastX-Toolkit (RRID: SCR_005534) was used to split reads by their barcodes (*63*). Before further downstream analysis, reads aligning to rRNA sequences were discarded using STAR (*64*), and the remaining reads were aligned to the transcriptome (GRCh38.p13). UMI-tools was used to extract unique molecular identifiers introduced during the sequencing steps and deduplicate the reads (*65*). The riboWaltz library was used to quantify ribosome A-site localization (*66*). Ribosomal-protected footprint (RPF) counts were normalized using median of ratio normalization before calculating differences in A-site codon abundances in each group.

For loess regression and quantification of stalling bias, the Ribolog package was used to quantify stalling coefficients under fed and starvation conditions separately using the default spanning parameter (*24*). In this framework, local regression is used to smooth out peaks introduced by stalling events and measure the degree to which ribosomal stalling occurs at any given position along a transcript. Larger coefficients correspond to more stalling, with values centered around 1. All transcripts with less than three aligned RPFs were removed before downstream analysis. To compare amino acid or codon-specific stalling coefficients, each gene was assigned its maximum stalling coefficient for each amino acid or codon to estimate the greatest degree of stalling at

a specific codon or amino acid for each gene. Bias coefficient ratios were calculated by taking the ratio between coefficients under starvation and fed conditions. To calculate amino acid identities in close proximity to maximal sites of stalling, gene positions were ranked by CELP bias coefficient to identify regions where maximal stalling was predicted to occur and then amino acid frequencies within the first three codons upstream and downstream of the maximum stalled site were counted. To restrict observations to the open reading frame, only positions that were at least 10 codons downstream of the start codon and 5 codons upstream of the stop codon were considered for analysis. Observations were normalized for gene-specific codon content and scaled to window width to account for codon composition of each individual gene. Translational efficiency analysis was performed using CELP-corrected RPF counts.

For allele-specific ribosome profiling, mutations from whole-exome sequencing were used to identify genes with heterozygous SNVs. Corresponding variant sequences were created from the wild-type sequence using a custom Python transcript and subsequently added to the original reference transcriptome fasta file. Alignment, ribosome A-site localization, and quantification of stalling coefficients were performed following the same procedure as above. Comparison of ribosome stalling around SNVs were then calculated by comparing stalling coefficients upstream and downstream of the SNV and statistical analysis performed only on genes containing that were heterozygous for a SNV.

Proteomics

The RKO cell line was selected for proteomic experiments due to its high frequency of arginine codon–switching mutations. At 24 hours after growth under either arginine full (400 μM) or limited (12.5 μM) conditions, cells were washed with PBS and treated with 0.25% trypsin to detach cells. Suspensions were immediately placed on ice with full media and centrifuged at 4°C and then resuspended in PBS twice. Following washing steps, cell pellets were resuspended in lysis buffer consisting of 0.02 M tris-HCl (pH 7.4), 0.1 M KCl, 0.001 M EDTA (pH 8), and 0.5 M NP-40. One tablet of 1× cOmplete protease inhibitor (Roche) was added to 10 ml of fresh lysis buffer. Samples were incubated on ice and vortexed every 5 s for a total of 15 min and sonicated on ice with a 4× 5-s pulse at 40% amplitude with a 30-s break between samples. Following sonication, samples were transferred to clean tubes and spun down at max speed at 4°C for 10 min. The supernatant was transferred to an additional set of clean tubes for further proteomic analysis. Protein was then quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific, catalog no. 23225). Twenty-five micrograms of protein was aliquoted from each sample and run at 200 V for 50 min in a 4 to 12% bis-tris gel with Mops buffer. The gel was stained and visualized with SimplyBlue Safestain (Thermo Fisher Scientific, catalog no. LC6060) following the manufacturer's instructions to ensure distinct protein bands before proceeding to downstream quantification.

Further sample processing was then performed in collaboration with the Proteomics Resource Center at Rockefeller University: Fifty micrograms of protein from each sample was reduced and alkylated using dithiothreitol and iodoacetamide. Proteins were precipitated using chloroform/water/methanol extraction, and pellets were digested with Endopeptidase LysC (Wako Chemicals) and sequencing grade modified trypsin (Promega). Peptides were labeled with TMTpro isobaric tags (Thermo Fisher Scientific), pooled, purified using an Oasis HLB cartridge (Waters), and fractionated using a high-pH fractionation spin column kit (Pierce). Fractionated peptides were separated across a 2.5-hour linear gradient on a 250 mm by 75 μm EasySpray column using a Dionex 3000 HPLC system operating at 300 nl/min and analyzed by a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) operating in positive data-dependent acquisition mode. Raw data were queried against the human proteome (downloaded from uniprot.org on 2/12/2019) at 1% false discovery rate (FDR) using MaxQuant v. 1.6.1.0. Data were searched using the standard settings. Further statistical analysis was performed within the Perseus framework using version 1.6.5.0. Protein group intensities were $\log_2$-transformed and normalized by subtraction of the median. Statistical significance was tested for using FDR-corrected (permutation-based with 250 randomizations) t test (q = 0.05). Validation of proteomic findings were carried out in distinct experiments using freshly starved samples and Western blots using the standard procedures and antibodies described above.

Pathway enrichment for proteomic changes was performed with iPAGE (*9*). For downstream quantification of mutational events in the proteome, proteins that were significantly increased or decreased under either fed or starvation conditions were selected with FDR $\leq$ 0.05. Mutational status in each gene set was then cross-referenced to previously collected whole-exome sequencing data to determine mutational status of differentially abundant proteins. Mutational changes were normalized to the total number of unique mutations per cell relative to the unselected cell lines.

Whole-exome sequencing and analysis

DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's instructions. Before sequencing, DNA was subjected to quality control with Picogreen and Fragment Analyzer to determine DNA integrity. Cancer samples were then sent for sequencing at the New York Genome Center (NYGC). Whole-exome sequencing libraries were prepared using the Agilent SureSelect XT library preparation kit in accordance with the manufacturer's instructions. Briefly, DNA was sheared using a Covaris LE220. DNA fragments were end-repaired, adenylated, ligated to SureSelect oligo adapters, and amplified by PCR. Exome capture was performed using the Agilent SureSelect XT Human All Exome v6 (60 Mb) capture probe set, and captured exome libraries were ligated to Agilent Sequencing adapters during target selection and enriched by PCR. The final libraries were quantified using the Qubit Fluorometer (Life Technologies) or Spectromax M2 (Molecular Devices) and Fragment Analyzer (Advanced Analytical) or Agilent 2100 BioAnalyzer and were sequenced on an Illumina NovaSeq 6000 sequencer run across two lanes of an S4-300 cycle flow cell.

For cancer cell lines, base calling and filtering were performed using current Illumina software; sequences were aligned to National Cancer for Biotechnology Information (NCBI) genome build 37 using Burrows-Wheeler Aligner (*67*). Picard was used to mark duplicate reads (Picard v1.83; http://broadinstitute.github.io/picard/) local realignment around insertions and deletions, and base quality scores were recalibrated using GATK (Genome Analysis Toolkit v3.5, PMID: 21478889). Variants were called using GATK HaplotypeCaller, which generates a single-sample genomic variant call format (GVCF) file. To improve variant call accuracy, multiple

single-sample GVCF files were jointly genotyped using GATK GenotypeGVCFs, which generates a multisample VCF. Variant Quality Score Recalibration (VQSR) was performed on the multisample VCF, which adds quality metrics to each variant that can be used in downstream variant filtering.

For PDX exome sequencing, base calling and filtering were performed using current Illumina software. Mouse reads were then detected and removed from the FASTQ files by aligning the data to a combined reference of mouse (GRCm38) and human (NCBI genome build 37). All read pairs with both reads mapping to mouse or one read mapping to mouse and the other unmapped were excluded from subsequent processing and analysis steps. The samples were then processed through NYGC's somatic preprocessing and variant-calling pipelines. The samples were aligned to build 37 using Burrows-Wheeler Aligner (BWA-MEM v0.7.15) (*67*); NYGC's ShortAlignmentMarking (v2.1) is used to mark short reads as unaligned (https://github.com/nygenome/nygc-short-alignment-marking). GATK (v4.1.0) FixMateInformation is run to verify and fix mate-pair information, followed by Novosort (v1.03.01) markDuplicates to merge individual lane BAM files into a single BAM file per sample. Duplicates are then sorted and marked, and GATK's base quality score recalibration is performed. The final result of the preprocessing pipeline is a coordinate-sorted BAM file for each sample. Variants were called using GATK HaplotypeCaller, which generates a single-sample GVCF. To improve variant call accuracy, the GVCF files were genotyped using GATK GenotypeGVCFs, and VQSR was performed which adds quality metrics to each variant that was used in downstream variant filtering.

Variants were annotated using Annotate Variation (ANNOVAR) (*68*). Any mutation appearing in the majority (at least two of three) control cell lines were considered parental mutations. Variant calls that did not fall into the former category were used for further analysis. In situations where mutational events were predicted to affect multiple transcripts and result in more than one possible amino acid/codon switching event, all mutational events were first counted and then normalized to the total number of transcripts affected from the single mutation. For cell line analyses, mutational counts were averaged across triplicates in each cell line, and amino acid/codon switching events were normalized to the total number of SNVs not considered parental mutations. For PDX experiments, we made the following adjustment: Because the parental tumors were generally not propagated across multiple mice in contrast to the highly liver metastatic derivative, unique mutations were filtered out using the matched tumors as a reference. We then calculated the rate of events in shared mutations and compared this to the frequency in the mutations unique to only liver-metastatic tumors.

Software libraries used

Experiment and model schematics were created with BioRender.com. Specialized software libraries used for gene expression, ribosome profiling, and whole-exome sequencing analyses are cited in their respective methods sections. For statistical analysis and plotting not specifically referenced, we used the following Python libraries: NumPy (1.19.2), pandas (1.13), SciPy (1.5.2), bioinfokit (2.0.3), and Seaborn (0.11.0) as well as the following R libraries: corrplot (0.92). GraphPad Prism (9.1.2) was also used to assist with statistical analysis and figure creation.

Statistical analysis

Statistical analyses (*t* tests and Mann-Whitney tests) were carried out using Prism 9 or SciPy (*69*) and were two-tailed tests unless otherwise specified in the text. Bioinformatic analyses for gene expression and ribosome profiling were carried out using specialized software packages as described under their corresponding sections. Throughout all figures, $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. Significance was concluded at $P < 0.05$.

Supplementary Materials

This PDF file includes:

Figs. S1 to S24

Tables S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. R. Hershberg, D. A. Petrov, Selection on codon bias. *Annu. Rev. Genet.* **42**, 287–299 (2008).
2. The Cancer Genome Atlas Research Network, J. N. Weinstein, E. A. Collisson, G. B. Mills, K. R. M. Shaw, B. A. Ozenberger, K. Ellrott, I. Shmulevich, C. Sander, J. M. Stuart, The Cancer Genome Atlas pan-cancer analysis project. *Nat. Genet.* **45**, 1113–1120 (2013).
3. L. B. Alexandrov, S. Nik-Zainal, D. C. Wedge, S. A. J. R. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A.-L. Børresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjörd, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jäger, D. T. W. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. López-Otín, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. J. Tutt, R. Valdés-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates; Australian Pancreatic Cancer Genome Initiative; ICGC Breast Cancer Consortium; ICGC MMML-Seq Consortium; ICGC PedBrain, J. Zucman-Rossi, P. A. Futreal, U. M. Dermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell, M. R. Stratton, Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
4. L. B. Alexandrov, M. R. Stratton, Mutational signatures: The patterns of somatic mutations hidden in cancer genomes. *Curr. Opin. Genet. Dev.* **24**, 52–60 (2014).
5. M. Pan, M. A. Reid, X. H. Lowman, R. P. Kulkarni, T. Q. Tran, X. Liu, Y. Yang, J. E. Hernandez-Davies, K. K. Rosales, H. Li, W. Hugo, C. Song, X. Xu, D. E. Schones, D. K. Ann, V. Gradinaru, R. S. Lo, J. W. Locasale, M. Kong, Regional glutamine deficiency in tumours promotes dedifferentiation through inhibition of histone demethylation. *Nat. Cell Biol.* **18**, 1090–1101 (2016).
6. M. R. Sullivan, L. V. Danai, C. A. Lewis, S. H. Chan, D. Y. Gui, T. Kunchok, E. A. Dennstedt, M. G. Vander Heiden, A. Muir, Quantification of microenvironmental metabolites in murine cancers reveals determinants of tumor nutrient availability. *Elife* **8**, e44235 (2019).
7. W.-B. Tsai, I. Aiba, S.-y. Lee, L. Feun, N. Savaraj, M. T. Kuo, Resistance to arginine deiminase treatment in melanoma cells is associated with induced argininosuccinate synthetase expression involving c-Myc/HIF-1α/Sp4. *Mol. Cancer Ther.* **8**, 3223–3233 (2009).
8. W.-B. Tsai, I. Aiba, Y. Long, H.-K. Lin, L. Feun, N. Savaraj, M. T. Kuo, Activation of Ras/PI3K/ERK pathway induces c-Myc stabilization to upregulate argininosuccinate synthetase, leading to arginine deiminase resistance in melanoma cells. *Cancer Res.* **72**, 2622–2633 (2012).
9. H. Goodarzi, O. Elemento, S. Tavazoie, Revealing global regulatory perturbations across human cancers. *Mol. Cell* **36**, 900–911 (2009).
10. A. H. Zea, P. C. Rodriguez, M. B. Atkins, C. Hernandez, S. Signoretti, J. Zabaleta, D. McDermott, D. Quiceno, A. Youmans, A. O'Neill, J. Mier, A. C. Ochoa, Arginase-producing myeloid suppressor cells in renal cell carcinoma patients: A mechanism of tumor evasion. *Cancer Res.* **65**, 3044–3048 (2005).
11. M. Pan, H. A. Choudry, M. J. Epler, Q. Meng, A. Karinch, C. Lin, W. Souba, Arginine transport in catabolic disease states. *J. Nutr.* **134**, 2826S–2829S (2004).
12. J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehár, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jané-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. A. Jr, M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. M. Conaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris,

M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, L. A. Garraway, The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

13. A. Le, A. Ng, T. Kwan, K. Cusmano-Ozog, T. M. Cowan, A rapid, sensitive method for quantitative analysis of underivatized amino acids by liquid chromatography-tandem mass spectrometry (LC-MS/MS). *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **944**, 166–174 (2014).
14. S. Rabinovich, L. Adler, K. Yizhak, A. Sarver, A. Silberman, S. Agron, N. Stettner, Q. Sun, A. Brandis, D. Helbling, S. Korman, S. Itzkovitz, D. Dimmock, I. Ulitsky, S. C. S. Nagamani, E. Ruppin, A. Erez, Diversion of aspartate in ASS1-deficient tumours fosters *de novo* pyrimidine synthesis. *Nature* **527**, 379–383 (2015).
15. J. S. Lee, L. Adler, H. Karathia, N. Carmel, S. Rabinovich, N. Auslander, R. Keshet, N. Stettner, A. Silberman, L. Agemy, D. Helbling, R. Eilam, Q. Sun, A. Brandis, S. Malitsky, M. Itkin, H. Weiss, S. Pinto, S. Kalaora, R. Levy, E. Barnea, A. Admon, D. Dimmock, N. Stern-Ginossar, A. Scherz, S. C. S. Nagamani, M. Unda, D. M. Wilson, R. Elhasid, A. Carracedo, Y. Samuels, S. Hannenhalli, E. Ruppin, A. Erez, Urea cycle dysregulation generates clinically relevant genomic and biochemical signatures. *Cell* **174**, 1559–1570.e22 (2018).
16. C.-T. Cheng, Y. Qi, Y.-C. Wang, K. K. Chi, Y. Chung, C. Ouyang, Y.-R. Chen, M. E. Oh, X. Sheng, Y. Tang, Y.-R. Liu, H. H. Lin, C.-Y. Kuo, D. Schones, C. M. Vidal, J. C.-Y. Chu, H.-J. Wang, Y.-H. Chen, K. M. Miller, P. Chu, Y. Yen, L. Jiang, H.-J. Kung, D. K. Ann, Arginine starvation kills tumor cells through aspartate exhaustion and mitochondrial dysfunction. *Commun. Biol.* **1**, 178 (2018).
17. N. N. Pavlova, B. King, R. H. Josselsohn, S. Violante, V. L. Macera, S. A. Vardhana, J. R. Cross, C. B. Thompson, Translation in amino-acid-poor environments is limited by $tRNA^{Gln}$ charging. *Elife* **9**, e62307 (2020).
18. D. Huh, M. C. Passarelli, J. Gao, S. N. Dusmatova, C. Goin, L. Fish, A. M. Pinzaru, H. Molina, Z. Ren, E. A. McMillan, H. Asgharian, H. Goodarzi, S. F. Tavazoie, A stress-induced tyrosine-tRNA depletion response mediates codon-based translational repression and growth suppression. *EMBO J.* **40**, e106696 (2021).
19. A. R. Subramaniam, T. Pan, P. Cluzel, Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 2419–2424 (2013).
20. A. M. Darnell, A. R. Subramaniam, E. K. O'Shea, Translational control through differential ribosome pausing during amino acid limitation in mammalian cells. *Molecular Cell.* **71**, 229–243.e11 (2018).
21. M. C. Passarelli, A. M. Pinzaru, H. Asgharian, M. V. Liberti, S. Heissel, H. Molina, H. Goodarzi, S. F. Tavazoie, Leucyl-tRNA synthetase is a tumour suppressor in breast cancer and regulates codon-dependent translation dynamics. *Nat. Cell Biol.* **24**, 307–315 (2022).
22. N. T. Ingolia, S. Ghaemmaghami, J. R. S. Newman, J. S. Weissman, Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
23. N. J. McGlincy, N. T. Ingolia, Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112–129 (2017).
24. J. Blaze, A. Navickas, H. L. Phillips, S. Heissel, A. Plaza-Jennings, S. Miglani, H. Asgharian, M. Foo, C. D. Katanski, C. P. Watkins, Z. T. Pennington, B. Javidfar, S. Espeso-Gil, B. Rostandy, H. Alwaseem, C. G. Hahn, H. Molina, D. J. Cai, T. Pan, W. D. Yao, H. Goodarzi, F. Haghighi, S. Akbarian, Neuronal Nsun2 deficiency produces tRNA epitranscriptomic alterations and proteomic shifts impacting synaptic signaling and behavior. *Nat Commun.* **12**, 4913 (2021).
25. E. I. Closs, CATs, a family of three distinct mammalian cationic amino acid transporters. *Amino Acids* **11**, 193–208 (1996).
26. E. B. Spector, S. C. Rice, S. Moedjono, B. Bernard, S. D. Cederbaum, Biochemical properties of arginase in human adult and fetal tissues. *Biochem. Med.* **28**, 165–175 (1982).
27. M. Riihimäki, A. Hemminki, K. Sundquist, J. Sundquist, K. Hemminki, Metastatic spread in patients with gastric cancer. *Oncotarget.* **7**, 52307–52316 (2016).
28. N. Hugen, C. J. H. van de Velde, J. H. W. de Wilt, I. D. Nagtegaal, Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. *Ann. Oncol.* **25**, 651–657 (2014).
29. N. Yamaguchi, E. M. Weinberg, A. Nguyen, M. V. Liberti, H. Goodarzi, Y. Y. Janjigian, P. B. Paty, L. B. Saltz, T. P. Kingham, J. M. Loo, E. de Stanchina, S. F. Tavazoie, PCK1 and DHODH drive colorectal cancer liver metastatic colonization and hypoxic growth by promoting nucleotide synthesis. *Elife* **8**, e52135 (2019).
30. E. R. Fearon, Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* **6**, 479–507 (2011).
31. J. F. Gainor, L. Dardaei, S. Yoda, L. Friboulet, I. Leshchiner, R. Katayama, I. Dagogo-Jack, S. Gadgeel, K. Schultz, M. Singh, E. Chin, M. Parks, D. Lee, R. H. DiCecca, E. Lockerman, T. Huynh, J. Logan, L. L. Ritterhouse, L. P. Le, A. Muniappan, S. Digumarthy, C. Channick, C. Keyes, G. Getz, D. Dias-Santagata, R. S. Heist, J. Lennerz, L. V. Sequist, C. H. Benes, A. J. Iafrate, M. Mino-Kenudson, J. A. Engelman, A. T. Shaw, Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discov.* **6**, 1118–1133 (2016).
32. P. A. Watson, V. K. Arora, C. L. Sawyers, Emerging mechanisms of resistance to androgen receptor inhibitors in prostate cancer. *Nat. Rev. Cancer* **15**, 701–711 (2015).
33. R. M. Samstein, C.-H. Lee, A. N. Shoushtari, M. D. Hellmann, R. Shen, Y. Y. Janjigian, D. A. Barron, A. Zehir, E. J. Jordan, A. Omuro, T. J. Kaley, S. M. Kendall, R. J. Motzer, A. A. Hakimi, M. H. Voss, P. Russo, J. Rosenberg, G. Iyer, B. H. Bochner, D. F. Bajorin, H. A. Al-Ahmadie, J. E. Chaft, C. M. Rudin, G. J. Riely, S. Baxi, A. L. Ho, R. J. Wong, D. G. Pfister, J. D. Wolchok, C. A. Barker, P. H. Gutin, C. W. Brennan, V. Tabar, I. K. Mellinghoff, L. M. DeAngelis, C. E. Ariyan, N. Lee, W. D. Tap, M. M. Gounder, S. P. D'Angelo, L. Saltz, Z. K. Stadler, H. I. Scher, J. Baselga, P. Razavi, C. A. Klebanoff, R. Yaeger, N. H. Segal, G. Y. Ku, R. P. DeMatteo, M. Ladanyi, N. A. Rizvi, M. F. Berger, N. Riaz, D. B. Solit, T. A. Chan, L. G. T. Morris, Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
34. T. T. Schmidt, G. Reyes, K. Gries, C. Ü. Ceylan, S. Sharma, M. Meurer, M. Knop, A. Chabes, H. Hombauer, Alterations in cellular metabolism triggered by *URA7* or *GLN3* inactivation cause imbalanced dNTP pools and increased mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E4442–E4451 (2017).
35. H. Gingold, D. Tehler, N. R. Christoffersen, M. M. Nielsen, F. Asmar, S. M. Kooistra, N. S. Christophersen, L. L. Christensen, M. Borre, K. D. Sørensen, L. D. Andersen, C. L. Andersen, E. Hulleman, T. Wurdinger, E. Ralfkiær, K. Helin, K. Grønbæk, T. Ørntoft, S. M. Waszak, O. Dahan, J. S. Pedersen, A. H. Lund, Y. Pilpel, A dual program for translation regulation in cellular proliferation and differentiation. *Cell* **158**, 1281–1292 (2014).
36. K. Pakos-Zebrucka, I. Koryga, K. Mnich, M. Ljujic, A. Samali, A. M. Gorman, The integrated stress response. *EMBO Rep.* **17**, 1374–1395 (2016).
37. T. M. Grzywa, A. Sosnowska, P. Matryba, Z. Rydzynska, M. Jasinski, D. Nowis, J. Golab, Myeloid cell-derived arginase in cancer immune response. *Front Immunol.* **11**, 938 (2020).
38. L. A. Coburn, S. N. Horst, M. M. Allaman, C. T. Brown, C. S. Williams, M. E. Hodges, J. P. Druce, D. B. Beaulieu, D. A. Schwartz, K. T. Wilson, L-arginine availability and metabolism is altered in ulcerative colitis. *Inflamm. Bowel Dis.* **22**, 1847–1858 (2016).
39. R. Chaturvedi, M. Asim, N. D. Lewis, H. M. S. Algood, T. L. Cover, P. Y. Kim, K. T. Wilson, l-arginine availability regulates inducible nitric oxide synthase-dependent host defense against *Helicobacter pylori*. *Infect. Immun.* **75**, 4305–4315 (2007).
40. A. Pataskar, J. Champagne, R. Nagel, J. Kenski, M. Laos, J. Michaux, H. S. Pak, O. B. Bleijerveld, K. Mordente, J. M. Navarro, N. Blommaert, M. M. Nielsen, D. Lovecchio, E. Stone, G. Georgiou, M. C. de Gooijer, O. van Tellingen, M. Altelaar, R. P. Joosten, A. Perrakis, J. Olweus, M. Bassani-Sternberg, D. S. Peeper, R. Agami, Tryptophan depletion results in tryptophan-to-phenylalanine substitutants. *Nature* **603**, 721–727 (2022).
41. L. Willems, N. Jacque, A. Jacquel, N. Neveux, T. T. Maciel, M. Lambert, A. Schmitt, L. Poulain, A. S. Green, M. Uzunov, O. Kosmider, I. Radford-Weiss, I. C. Moura, P. Auberger, N. Ifrah, V. Bardet, N. Chapuis, C. Lacombe, P. Mayeux, J. Tamburini, D. Bouscary, Inhibiting glutamine uptake represents an attractive new strategy for treating acute myeloid leukemia. *Blood* **122**, 3521–3532 (2013).
42. O. D. K. Maddocks, D. Athineos, E. C. Cheung, P. Lee, T. Zhang, N. J. F. van den Broek, G. M. Mackay, C. F. Labuschagne, D. Gay, F. Kruiswijk, J. Blagih, D. F. Vincent, K. J. Campbell, F. Ceteci, O. J. Sansom, K. Blyth, K. H. Vousden, Modulating the therapeutic response of tumours to dietary serine and glycine starvation. *Nature.* **544**, 372–376 (2017).
43. M. Tajan, A. K. Hock, J. Blagih, N. A. Robertson, C. F. Labuschagne, F. Kruiswijk, T. J. Humpton, P. D. Adams, K. H. Vousden, A role for p53 in the adaptation to glutamine starvation through the expression of SLC1A3. *Cell Metab.* **28**, 721–736.e6 (2018).
44. F. P. Canale, C. Basso, G. Antonini, M. Perotti, N. Li, A. Sokolovska, J. Neumann, M. J. James, S. Geiger, W. Jin, J.-P. Theurillat, K. A. West, D. S. Leventhal, J. M. Lora, F. Sallusto, R. Geiger, Metabolic modulation of tumours with engineered bacteria for immunotherapy. *Nature* **598**, 662–666 (2021).
45. P. Thandapani, A. Kloetgen, M. T. Witkowski, C. Glytsou, A. K. Lee, E. Wang, J. Wang, S. E. LeBoeuf, K. Avrampou, T. Papagiannakopoulos, A. Tsirigos, I. Aifantis, Valine tRNA levels and availability regulate complex I assembly in leukaemia. *Nature* **601**, 428–433 (2022).
46. M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, M. A. Marra, Circos: An information aesthetic for comparative genomics. *Genome Res.* **19**, 1639–1645 (2009).
47. J. M. Rodriguez, P. Maietta, I. Ezkurdia, A. Pietrelli, J.-J. Wesselink, G. Lopez, A. Valencia, M. L. Tress, APPRIS: Annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* **41**, D110–D117 (2013).
48. J. Zhang, R. Bajari, D. Andric, F. Gerthoffert, A. Lepsa, H. Nahal-Bose, L. D. Stein, V. Ferretti, The international cancer genome consortium data portal. *Nat. Biotechnol.* **37**, 367–369 (2019).
49. M. Waskom, seaborn: Statistical data visualization. *JOSS.* **6**, 3021 (2021).
50. T. C. Silva, A. Colaprico, C. Olsen, F. D'Angelo, G. Bontempi, M. Ceccarelli, H. Noushmehr, *TCGA workflow*: Analyze cancer genomics and epigenomics data using Bioconductor packages. *F1000Res.* **5**, 1542 (2016).

51. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
52. M. Mounir, M. Lucchetta, T. C. Silva, C. Olsen, G. Bontempi, X. Chen, H. Noushmehr, A. Colaprico, E. Papaleo, New functionalities in the TCGAbiolinks package for the study and integration of cancer data from GDC and GTEx. *PLOS Comput. Biol.* **15**, e1006701 (2019).
53. S. Anders, W. Huber, Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
54. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
55. A. Tsherniak, F. Vazquez, P. G. Montgomery, B. A. Weir, G. Kryukov, G. S. Cowley, S. Gill, W. F. Harrington, S. Pantel, J. M. Krill-Burger, R. M. Meyers, L. Ali, A. Goodale, Y. Lee, G. Jiang, J. Hsiao, W. F. J. Gerath, S. Howell, E. Merkel, M. Ghandi, L. A. Garraway, D. E. Root, T. R. Golub, J. S. Boehm, W. C. Hahn, Defining a cancer dependency map. *Cell* **170**, 564–576.e16 (2017).
56. B. Jassal, L. Matthews, G. Viteri, C. Gong, P. Lorente, A. Fabregat, K. Sidiropoulos, J. Cook, M. Gillespie, R. Haw, F. Loney, B. May, M. Milacic, K. Rothfels, C. Sevilla, V. Shamovsky, S. Shorser, T. Varusai, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D'Eustachio, The reactome pathway knowledgebase. *Nucleic Acids Res.* **48**, D498–D503 (2020).
57. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
58. T. W. Traut, Physiological concentrations of purines and pyrimidines. *Mol. Cell Biochem.* **140**, 1–22 (1994).
59. P. P. Chan, T. M. Lowe, GtRNAdb: A database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93–D97 (2009).
60. P. P. Chan, T. M. Lowe, GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184–D189 (2016).
61. M. Shigematsu, S. Honda, P. Loher, A. G. Telonis, I. Rigoutsos, Y. Kirino, YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res.* **45**, e70 (2017).
62. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
63. G. J. Hannon, FASTX-Toolkit, (available at http://hannonlab.cshl.edu/fastx_toolkit).
64. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
65. T. Smith, A. Heger, I. Sudbery, UMI-tools: Modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
66. F. Lauria, T. Tebaldi, P. Bernabò, E. J. N. Groen, T. H. Gillingwater, G. Viero, riboWaltz: Optimization of ribosome P-site positioning in ribosome profiling data. *PLOS Comput. Biol.* **14**, e1006169 (2018).
67. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
68. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
69. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. V. Plas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt; SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).

Acknowledgments: We are grateful to members of our laboratory for comments on previous versions of the manuscript as well as K. Birsoy at Rockefeller University and B. Santhanam at Columbia University for suggestions. We thank Rockefeller University resource centers for assistance: V. Francis and other veterinary staff of the Comparative Bioscience Center for animal husbandry and care and H. Molina and staff at the Proteomics Resource Center for assistance with proteomic quantification and metabolite profiling. We also acknowledge collaborators at the New York Genome Center for support in processing whole-exome sequencing samples. We would also like to acknowledge A. Azzam for assistance in processing proteomic samples as a medical rotation student. The results published here are in part based on data generated by the TCGA Research Network: www.cancer.gov/tcga. **Funding:** This work was supported by a Damon Runyon Physician Scientist Training Award PST 31-20 (D.J.H.), National Institutes of Health Grant 5R01CA257153 (S.T.), National Institutes of Health Grant R35CA274446 (S.F.T.), the Black Family Metastasis Research Center (S.F.T.), a Pershing Square Innovation Fund (S.F.T.), and a Reem-Kayden Award (S.F.T.). **Author contributions:** Conceptualization: D.J.H., S.T., and S.F.T. Methodology: D.J.H., S.T., and S.F.T. Validation: D.J.H., N.M., J.G., and Q.W. Formal analysis: D.J.H., S.H., and H.A. Investigation: D.J.H., J.G., N.Y., A.P., M.L., S.H., H.A., and Q.W. Supervision: D.J.H., S.F.T., and S.T. Writing—original draft: D.J.H. and S.F.T. Writing—review and editing: D.J.H., S.T., and S.F.T. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. RNA sequencing and ribosome profiling data are publicly available on the NCBI Gene Expression Omnibus (#GSE213472). Proteomic data have been made publicly available on the Proteomics Identifications Database (#PXD036737). Whole-exome sequencing results have been made available at the NCBI Sequence Read Archive (#PRJNA880559).

Submitted 16 September 2022
Accepted 1 December 2022
Published 6 January 2023
10.1126/sciadv.ade9120