Water as a reactant in the differential expression of proteins in cancer

Jeffrey M. Dick[a,1]

[a]Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring, Ministry of Education, School of Geosciences and Info-Physics, Central South University, Changsha 410083, China

This manuscript was compiled on April 10, 2020

Abstract

The large amounts of proteomic data now available for cancer can used to investigate whether the physicochemical conditions of tumors are reflected in patterns of protein expression and chemical composition. Compositional analysis of more than 250 datasets for differentially expressed proteins compiled from the literature reveals a clear signal of higher stoichiometric hydration state (n_{H_2O}, derived from the theoretical formation reactions of proteins from particular basis species) in specific cancer types compared to normal tissue; this trend is also evident in pan-cancer transcriptomic and proteomic datasets from The Cancer Genome Atlas and Human Protein Atlas. In marked contrast to cancer, n_{H_2O} decreases for differentially expressed proteins in hyperosmotic stress (including high glucose) experiments and 3D cell culture compared to monolayer growth. Compositional analysis combined with gene ages (phylostrata) taken from the literature shows higher n_{H_2O} of human proteins earlier in evolution. Further analyses using amino acid biosynthetic reactions supports the conclusion that a net increase of water going into the reactions of protein synthesis is a biochemical characteristic shared by most cancer types. These findings raise the possibility of a basic physicochemical link between increased water content in tumors and the atavistic or embryonic patterns of gene and protein expression in cancer.

Keywords: cancer, proteomics, chemical composition, water content, hypoxia, gene ages, osmotic stress, 3D cell culture

Introduction

Although cancer is usually regarded as primarily a genetic disease (1), alterations in tumor microenvironments and metabolism provide conditions crucial for malignant progression (2). Differences in the abundances of many proteins are one manifestation of the combination of genetic, microenvironmental and metabolic alterations in cancer. In a gene-centric view of metabolism, the myriad reactions underlying changes to the proteome are catalyzed and regulated by the enzymatic products of the genome, but an adequate biochemical description should also account for the chemical compositions of the proteins themselves.

From a geochemical perspective, a natural question is to ask whether the chemical compositions of differentially expressed proteins are shaped by the physicochemical conditions of tumor microenvironments. Changes in water and oxygen content are major chemical characteristics of cancer. Hypoxia, or less than normal physiological concentration of oxygen, in tumor microenvironments plays a major role in the biochemistry, physiology and progression of cancer (3). Cancer tissue also has a relatively high water content (4), as consistently demonstrated by early desiccation experiments (5). Cell refractometry of tumor and normal tissue from livers of diseased rats indicates that much of the increase is due to intracellular water (6). More recent developments of spectroscopic methods further substantiate the generally higher water content of cancer tissue (7, 8). These observations are consistent with the hypothesis that higher cellular hydration in carcinogenesis is a major factor that is shared with embryonic conditions (9). Moreover, water content is a key player in other aspects of cell biology such as entry into dormancy (10). Nevertheless, the connections between cellular water content and biomolecular abundances in cells are not well understood.

Oxidation reactions, in which oxygen or another electron acceptor is a reactant, result in the formation of more oxidized biomolecules. Hydration reactions, in which water is a reactant, result in the formation of more hydrated biomolecules. If only water is added, hydration reactions do not involve the transfer of electrons, that is, they are redox-neutral. In a short section on "Water as a Reactant", a well known biochemistry textbook (11) describes a few types of reactions involving the release of H_2O as a product (oxidation of glucose, condensation reactions) or its consumption as a reactant (water splitting in photosynthesis, and hydrolysis, the reverse of condensation). Condensation reactions, in particular the polymerization of amino acids, are fundamental to protein synthesis, but the stoichiometry of the reactions depends only

Significance Statement

How the abundances of proteins are shaped by tumor microenvironments, such as hypoxic conditions and higher water content compared to normal tissues, is an important question for cancer biochemistry. I analyzed multiple sources of proteomic data, together with pan-cancer gene and protein expression data, and found that up-regulated proteins in most cancer types have a higher stoichiometric hydration state, meaning that water is consumed as a reactant if the differential expression of proteins is represented theoretically as a chemical reaction. Moreover, for all human proteins, those coded by older genes are relatively hydrated. These findings indicate that uptake of water may be a fundamental biochemical requirement at the proteome level in cancer and that high cellular hydration state is compatible with the activation of ancient gene expression patterns (atavism) in many cancer types.

J.M.D. designed and performed research, analyzed data, and wrote the paper.

The author declares no conflict of interest.

Data deposition: The compiled differential expression data are available in the canprot R package, version 0.1.6 (https://doi.org/10.5281/zenodo.3746998); Figs. S6–S15 are derived from the vignettes in this package. The code and other data used to make the figures in this paper, Tables 2, S1, and S3, and Figs. S1–S5 and S16–S17 are in the JMDplots package, version 1.2.0 (https://doi.org/10.5281/zenodo.3746999); the "canH2O" vignette in this package demonstrates the code to make the figures.

[1]To whom correspondence should be addressed. E-mail: jeff@chnosz.net

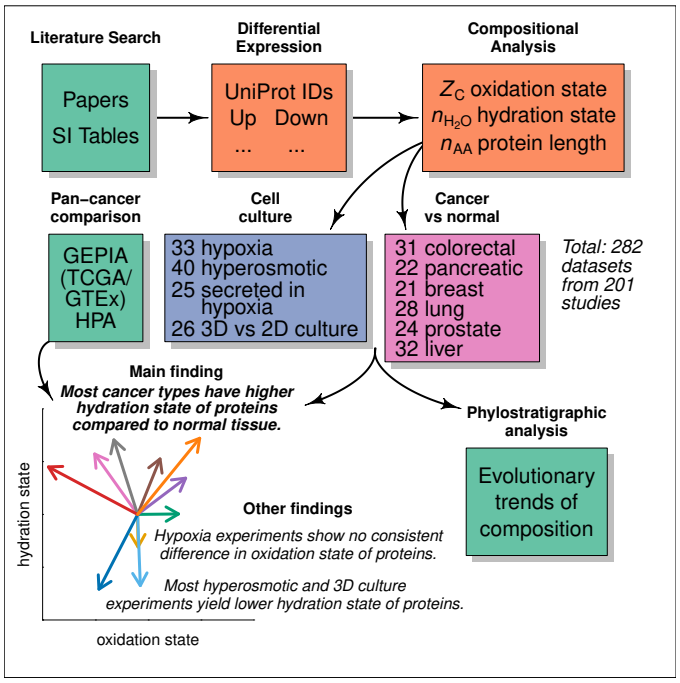

Fig. 1. Study overview. Abbreviations: SI – Supplementary Information; GEPIA – Gene Expression Profiling Interactive Analysis web server; TCGA – The Cancer Genome Atlas; GTEx – Genotype-Tissue Expression project; HPA – Human Protein Atlas. The arrow diagram represents mean values among datasets for differences of hydration state (Δn_{H_2O}) and oxidation state (ΔZ_C) (see Table 2).

on protein length; one water is lost for each peptide bond formed between any two amino acids. A more specific metric is needed to quantify the amount of H_2O gained or lost in the differential expression of proteins with different amino acid compositions.

Without considering individual biochemical reactions, it is still possible to use compositional metrics, which are derived from the elemental composition of proteins, to quantify the net differences in the degree of oxidation (oxidation state) and hydration (hydration state) of different proteins. This choice of variables follows from the altered oxygenation and hydration status of tumors and the observation that most biochemical transformations involve some combination of oxidation-reduction and hydration-dehydration reactions (12, 13).

The compositional analysis can be used to test the thermodynamic predictions of mass-action effects, specifically that more oxidizing or hydrating conditions favor the formation of more oxidized or hydrated proteins, and vice versa. The sensitivity of metabolic reactions to hypoxia is well documented; for example, the reduction of metabolites under hypoxic conditions is possible by running the TCA cycle in reverse (14), and hypoxic regions in tumors accelerate the reduction of nitroxide, a redox-sensitive probe used in magnetic resonance imaging (15, 16). However, hypoxia also induces the mitochondrial production of reactive oxygen species (17), so it would be an oversimplification to state that hypoxia leads to uniformly more reducing intracellular conditions. Cellular hydration state also has wide-ranging effects on cell metabolism (18), but no previous studies have systematically characterized chemical metrics of oxidation and hydration state at the proteome level in cancer.

My previous analysis of proteomic data provided preliminary evidence for a higher hydration state of proteomes in colorectal and pancreatic cancer (19). That compilation of differential expression data is expanded here to include breast, liver, lung and prostate cancer. Proteomic data are also considered for laboratory experiments of hypoxia, because of its relevance to cancer (3), and hyperosmotic stress, which has not been reported for cancer cell lines, but permits testing the sensitivity of the compositional analysis to changes in hydration state (see Ref. 20). Furthermore, I separately analyze proteomic data for both cellular and secreted proteins in hypoxia compared to normoxic controls. I also consider differential expression data for 3D culture conditions; compared to 2D or monolayer growth, the formation of cell aggregates, spheroids, or organoids in 3D culture more closely represents the tissue environment (21, 22). Finally, I combine the differential expression data with gene ages to get a picture of the evolutionary trajectories of chemical composition, and look at an alternative to the stoichiometric analysis using biosynthetic reactions for amino acids.

Through the compositional analysis of proteomic datasets for particular cancer types and cell culture conditions, as well as pan-cancer transcriptomic and proteomic data, I show that the differences in stoichiometric hydration state of proteins are specifically linked with different conditions. Hyperosmotic and 3D culture conditions in laboratory experiments induce the expression of proteins with a lower hydration state, whereas a higher hydration state characterizes proteins upregulated in most cancer types and those coded by phylogenetically older genes. Therefore, the differential expression of proteins in cancer appears to depend on the uptake of water as a reactant.

Table 1. Average oxidation state of carbon (Z_C), number of carbon atoms (n_C), and stoichiometric hydration state (n_{H_2O}) of amino acid residues computed using the rQEC derivation (see Materials and Methods and Ref. 20).

AA	Z_C	n_C	n_{H_2O}	AA	Z_C	n_C	n_{H_2O}
A	0	3	0.369	M	-2/5	5	0.046
C	2/3	3	-0.025	N	1	4	-0.122
D	1	4	-0.122	P	-2/5	5	-0.354
E	2/5	5	-0.107	Q	2/5	5	-0.107
F	-4/9	9	-2.568	R	1/3	6	0.072
G	1	2	0.478	S	2/3	3	0.575
H	2/3	6	-1.825	T	0	4	0.569
I	-1	6	0.660	V	-4/5	5	0.522
K	-2/3	6	0.763	W	-2/11	11	-4.087
L	-1	6	0.660	Y	-2/9	9	-2.499

Results

Extensive literature searches were performed to build a database of differentially expressed proteins in primary cancers of six organs compared to normal tissue and four cell culture conditions versus controls. In total, 282 datasets were obtained from 201 studies (Fig. 1).

The carbon oxidation state (Z_C) and stoichiometric hydration state (n_{H_2O}) (Table 1) are compositional metrics derived from the chemical formulas of amino acids; therefore, they do not denote any particular biological mechanisms for amino acid synthesis. It should also be emphasized that all of the calculations in this study are based on differences in the chemical composition of proteins as determined by their primary

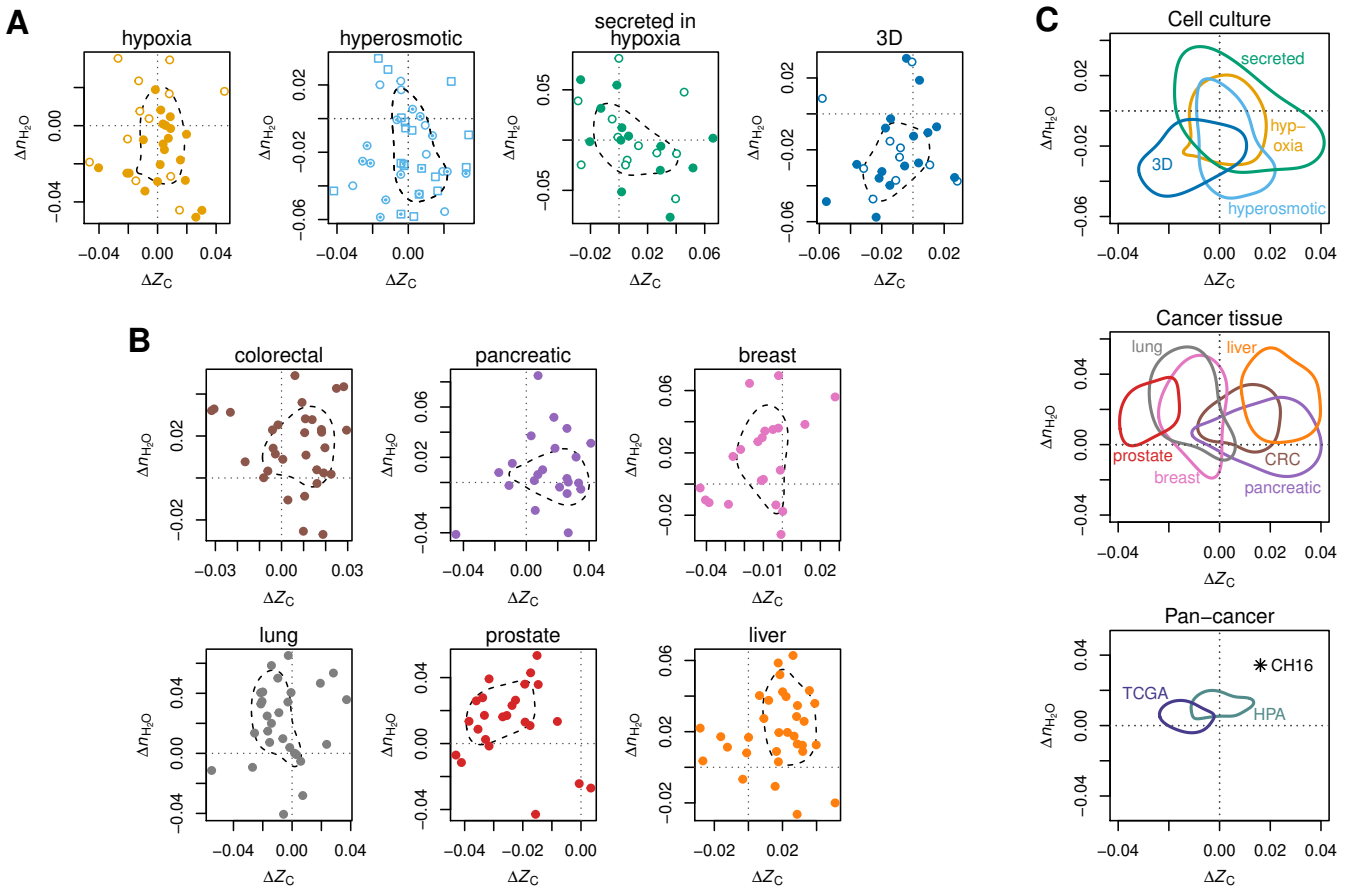

Fig. 2. Compositional analysis of proteins identified in differential expression datasets. Median differences of stoichiometric hydration state (n_{H_2O}) and average oxidation state of carbon (Z_C) in (**A**) cell culture experiments and (**B**) cancer tissues. Color-coded circles represent individual proteomics experiments. For cell culture experiments, open and filled symbols represent non-cancer and cancer cells, respectively; dotted symbols for hyperosmotic conditions indicate glucose treatment, and squares represent microbial (yeast or bacterial) cells. Dashed lines indicate the 50% credible region for highest probability density for all datasets for each condition. (**C**) Comparison of the 50% credible regions for cell culture and cancer tissue together with pan-cancer gene and protein expression datasets (TCGA and HPA); the latter are shown in detail in Fig. 3. The point labeled "CH16" indicates the median differences for proteins corresponding to 229 up- and 68 down-regulated genes with common expression changes across cancer types, defined by Chen and He (2016) as genes with unidirectional expression changes in at least 13 of 32 cancer data sets (data from Supplementary Table S3 of Ref. 23). In all plots, positive Δ values indicate a higher median value for the up-regulated proteins.

sequences, and do not take account of post-transcriptional modifications, like the oxidation of cysteine to make disulfide bonds, or the presence of water molecules in the hydration shell of folded proteins.

Carbon oxidation state for biomolecules lies between the extremes of -4 for CH_4 and +4 for CO_2 (see Figure 1 of Ref. 24). Because it is based on the relative electronegativities of elements, it can be calculated directly from the elemental composition of proteins (25, 26). On the other hand, a metric for hydration state depends on the stoichiometry of water in balanced chemical reactions. Since reactions involving only water do not involve the transfer of electrons, a useful metric for hydration state should not be correlated with oxidation state for proteins in general (i.e. all those coded by the genome). Following this reasoning, the basis species glutamine–glutamic acid–cysteine–H_2O–O_2 were selected to write theoretical formation reactions of amino acids; the number of water molecules in these reactions (Table S1) was input to a residual analysis to further reduce the covariation with Z_C (Fig. S1), giving the residual-corrected stoichiometric hydration state listed in Table 1. This derivation, denoted "rQEC", is briefly described in the Materials and Methods; see Ref. 20 for more details and conceptual background.

Differences Between Cell Culture and Cancer Tissue. The compositional analysis of differentially expressed proteins is summarized in scatterplots of median Δn_{H_2O} and ΔZ_C for individual datasets (Fig. 2*A*). The values for all datasets in each condition were used to compute the 50% credible regions for highest probability density using code adapted from the "HPDregionplot" function in the R package emdbook (27), which in turn uses 2-D kernel density estimates calculated with "kde2d" in the R package MASS (28). Plots with references and descriptions for all datasets are provided in Figs. S6–S15.

Several broad trends emerge from the compositional analysis of differentially expressed proteins in cell culture conditions. Differentially expressed proteins reported for cell extracts under hypoxia do not show consistent differences in oxidation state (Fig. 2*A*). However, differentially expressed proteins in many datasets for secreted proteins in hypoxia are somewhat oxidized ($\Delta Z_C > 0$). Although the wider credible region for secreted proteins indicates a larger variability

Table 2. Mean differences for all differential expression datasets in each condition, followed by $\log_{10}$ of *p*-value in parentheses. *p*-values less than 0.05 ($\log_{10}$ < -1.3) are shown in bold.

Condition	ΔZ_C	Δn_{H_2O} (rQEC)	Δn_{O_2} (biosynthetic)	Δn_{H_2O} (biosynthetic)	Δn_{AA}
Cell culture					
Hypoxia (cellular extracts)	0.000 (-0.0)	-0.009 (**-1.4**)	0.002 (-0.1)	0.007 (-0.3)	7.6 (-0.1)
Hyperosmotic	0.001 (-0.1)	-0.019 (**-3.6**)	0.016 (-0.7)	-0.015 (-0.6)	-0.1 (-0.0)
Secreted in hypoxia	0.011 (**-1.6**)	0.000 (-0.0)	0.031 (-1.1)	-0.018 (-0.5)	-19.6 (-0.2)
3D	-0.010 (**-1.9**)	-0.020 (**-4.0**)	-0.007 (-0.3)	0.001 (-0.0)	-4.1 (-0.0)
Cancer					
Colorectal	0.006 (-1.2)	0.015 (**-3.3**)	-0.013 (-1.0)	0.019 (**-2.1**)	35.9 (-0.8)
Pancreatic	0.013 (**-2.1**)	0.010 (-0.7)	0.014 (-0.6)	-0.007 (-0.3)	29.5 (-0.5)
Breast	-0.012 (**-2.0**)	0.016 (**-2.0**)	-0.058 (**-3.2**)	0.047 (**-5.9**)	-71.2 (**-1.6**)
Lung	-0.006 (-1.0)	0.020 (**-3.1**)	-0.020 (-0.9)	0.020 (-1.3)	-44.6 (-0.8)
Prostate	-0.024 (**-11.3**)	0.013 (**-1.4**)	-0.080 (**-8.8**)	0.048 (**-5.9**)	-16.1 (-0.2)
Liver	0.017 (**-6.1**)	0.021 (**-4.5**)	0.005 (-0.3)	0.004 (-0.2)	24.8 (-0.7)
Pan-cancer					
HPA	-0.000 (-0.0)	0.009 (**-3.6**)	-0.008 (-0.6)	0.013 (**-1.7**)	12.4 (-0.8)
TCGA/GTEx	-0.009 (**-5.8**)	0.006 (**-3.3**)	-0.034 (**-8.6**)	0.029 (**-8.8**)	-81.6 (**-6.3**)
*Secreted in hypoxia compared to cellular extracts in hypoxia **					
up-regulated	0.014 (**-2.8**)	-0.003 (-0.2)	0.048 (**-2.8**)	-0.034 (**-1.5**)	42.2 (-0.7)
down-regulated	0.003 (-0.3)	-0.012 (**-1.5**)	0.019 (-0.7)	-0.008 (-0.3)	69.4 (**-1.6**)

Mean differences were calculated as (mean of median values for up-regulated proteins in each dataset) - (mean of median values for down-regulated proteins in each dataset), except * (mean of median values for [up- or down-]regulated proteins secreted in hypoxia) - (mean of median values for [up-or down-]regulated proteins in cellular extracts in hypoxia). *p*-values were calculated with the Welch two-sample *t*-test by using R function "t.test"(29) with default options.

(Fig. 2*C*), the shift toward higher Z_C is statistically significant for these datasets (Table 2). Hyperosmotic stress results in the formation of proteins with predominantly lower hydration state ($\Delta n_{H_2O} < 0$). Lower hydration state also characterizes the majority of 3D cell culture experiments, which in addition tend to have more reduced proteins ($\Delta Z_C < 0$). Note that all of the 3D cell culture experiments analyzed here are for human or mouse cells, including some cancer cell lines, which are represented by filled circles in Fig. 2*A*. The experiments for cellular and secreted proteins in hypoxia include human and other mammalian cells. In contrast, the hyperosmotic stress experiments include mammalian as well as microbial (non-halophilic yeast and bacteria) cells; the latter are indicated by the squares in Fig. 2*A*. Hyperosmotic conditions in the experiments are generated by the addition of inorganic salts or organic osmolytes such as glucose or mannitol to the culture media; see Fig. S7 for details.

There is a clear trend of increased hydration state of proteins ($\Delta n_{H_2O} > 0$) for five of the six cancer types considered in detail here (Fig. 2*B* and *C*). The exception is pancreatic cancer, where the datasets are distributed more evenly among positive and negative Δn_{H_2O}. There are distinct trends in oxidation state of proteins for different cancer types: relatively oxidized proteins are up-regulated in colorectal, liver, and pancreatic cancer, whereas more reduced proteins are up-regulated in breast, lung, and prostate cancer.

The trends described above are also visible in the arrow diagram in Fig. 1. In this diagram, the lines are drawn from the origin to the mean difference of Z_C and n_{H_2O} in datasets for each cancer type and laboratory condition; see Table 2 for all mean differences and and *p*-values. All cancer types have positive mean Δn_{H_2O} across datasets, indicating greater hydration state of the up-regulated proteins, but the difference for pancreatic cancer is less statistically significant (*p*-value > 0.05). In contrast, hyperosmotic stress and 3D cell culture conditions, and to a lesser extent, hypoxia for cellular extracts, lead to the up-regulation of proteins with significantly lower hydration state.

Elevated Hydration State and Variable Oxidation State in Pan–Cancer Datasets. Some fundamental biological questions are: do different cancer types have similar patterns of protein expression, and are these patterns inherent in the expression of the genes that code for the proteins? To characterize pan-cancer transcriptomes and proteomes in terms of chemical composition, I obtained data for differential gene expression between normal tissue and cancer from GEPIA2 (32), which uses pre-compiled data files from UCSC Xena (33) that are in turn derived from the Genotype-Tissue Expression project (GTEx) (34) and The Cancer Genome Atlas (TCGA) (35). I used data from the Human Protein Atlas (HPA) (31, 36) to calculate differential protein expression as described in the Materials and Methods.

Except for prostate cancer, both pan-cancer datasets manifest a positive Δn_{H_2O} for the cancer types analyzed in detail in this study (color-coded circles in Fig. 3). Differential gene expression for all cancer types taken together corresponds to significantly more reduced proteins (Table 2, column ΔZ_C), but this is not evident in the HPA proteomics data. In a pairwise comparison of transcriptomic and proteomic data for cancer types, there is very little correlation in Z_C of proteins and even less in Δn_{H_2O} (Fig. S2). It is therefore remarkable that both the pan-cancer transcriptomic and proteomic datasets have a strong visible and statistically significant trend toward higher hydration state of the associated proteins (Table 2, column Δn_{H_2O} (rQEC)). Likewise, a set of genes with common

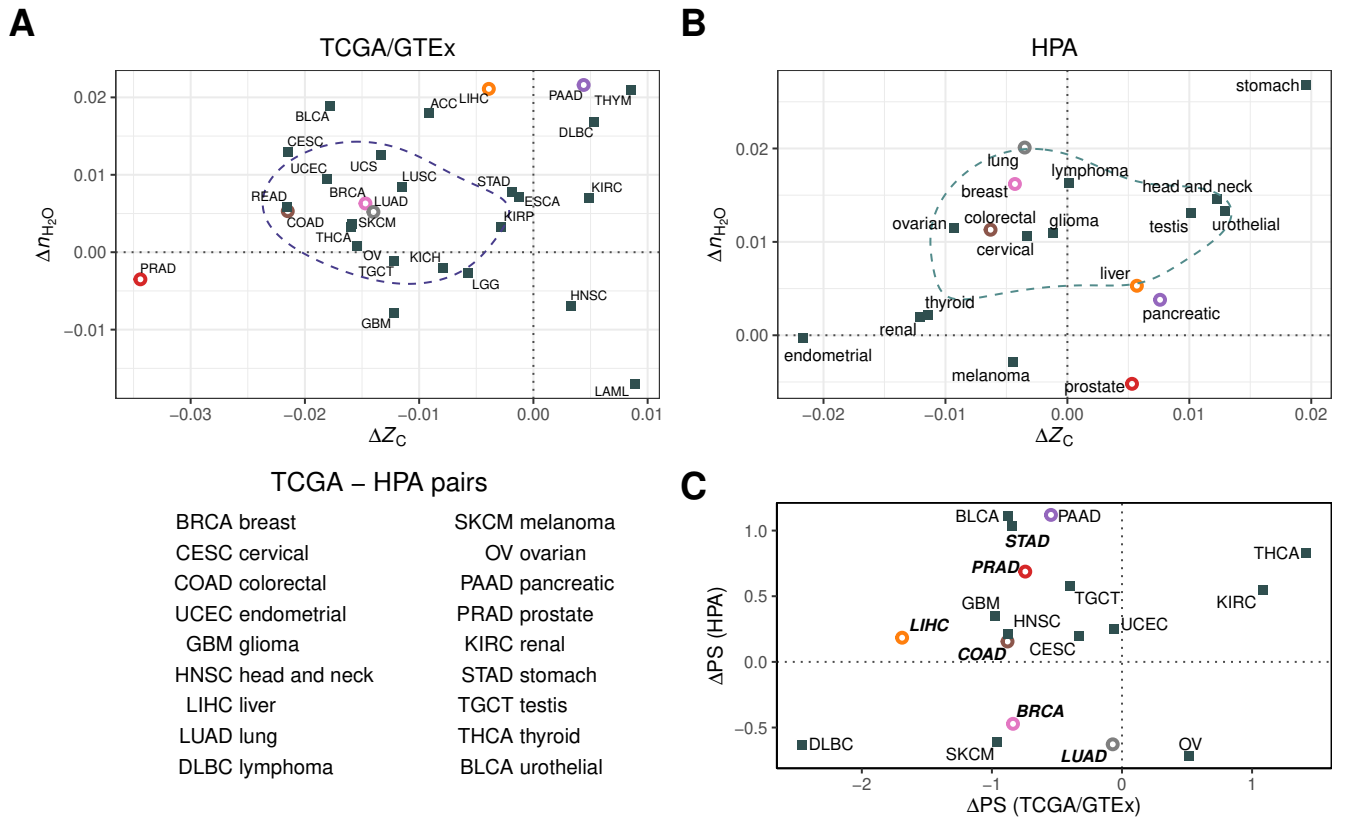

Fig. 3. Changes in chemical composition and phylostrata for differentially regulated proteins associated with large-scale transcriptomics and protein antibody studies. (**A**) Proteins coded by differentially expressed genes between normal tissue (GTEx) and cancer (TCGA). Abbreviations for cancer types are listed in Table S2 (from Ref. 30). (**B**) Differentially expressed proteins in the Human Protein Atlas (31). Dashed lines in panels A and B indicate the 50% credible region for highest probability density. (**C**) Cross-comparison of changes in mean phylostrata (PS) for differentially regulated proteins derived from TCGA and HPA (TCGA-HPA pairings used for this plot are indicated). See Figs. S16–S17 for ΔPS calculated for all TCGA and HPA datasets. Color-coded circles represent cancer types analyzed in greater detail in this study (see Fig. 2).

expression changes across cancer types (23) also shows a positive Δn_{H_2O} of the associated proteins (point labeled "CH16" in Fig. 2*C*).

The distribution of positive and negative ΔZ_C in the analysis of HPA datasets for different cancer types suggests a biological origin other than the nominally reducing effects of hypoxia. It is notable that membrane and extracellular proteins in yeast are relatively reduced and oxidized, respectively (26). Similarly, stoichiogenomic analysis of the proteomes of twelve eukaryotic organisms indicates that extracellular proteins have a relatively low hydrogen content (37), which would tend to increase the average carbon oxidation state. Furthermore, up-regulated proteins that are secreted in hypoxia are more oxidized than their counterparts in cellular extracts (Table 2). These fundamental subcellular differences might explain why tissues found by Uhlén et al. (31) to be enriched in membrane proteins (brain and kidney) host cancers with negative values of ΔZ_C (glioma and renal, respectively), while tissues with high levels of proteins known to be secreted (pancreas) or enriched in the transcripts of secreted proteins (liver, stomach) host cancers characterized by positive values of ΔZ_C. These patterns imply that the normal enrichment of subcellular protein classes in different tissue types is pushed to pathological levels in cancer.

It is informative to compare these results with experimental measurements of the hydration status of specific types of cancer. For instance, NMR T_1 relaxation times distinguish early pancreatic ductal adenocarcinoma in mice, but not later stages; this is likely a consequence of increased water and protein content in the early stages (38). The possible stage-specific variation of water content may help explain why the range of Δn_{H_2O} of proteins in pancreatic cancer is closer to zero, compared to other cancer types (Figs. 2*C*, 3*B*). In another study, optical measurements of gliomas in rats in the spectral range 350–1800 nm were used to infer increased water content in early stages, but decreased amounts in advanced stages, in conjunction with the formation of necrotic regions in the tumor (39). This appears to be consistent with the small decrease in Δn_{H_2O} between LGG (brain lower grade glioma) and GMB (glioblastoma multiforme) in the TCGA datasets (Fig. 3*A*), but the HPA data exhibit positive Δn_{H_2O} for a single glioma type (Fig. 3*B*).

Compared to other cancer types, prostate cancer has distinct trends in the chemical composition of differentially expressed proteins. The negative Δn_{H_2O} of proteins for prostate cancer in the TCGA and HPA datasets (Fig. 3) may be related to the lower water content of prostate cancer than surrounding normal tissue as measured using near infrared spectroscopy (40); this trend is not apparent in the compiled proteomic datasets (Fig. 2*B*), perhaps because of their lower number of proteins. Furthermore, the highly negative ΔZ_C of the proteins in many proteomic datasets (Fig. 2*B*) and those

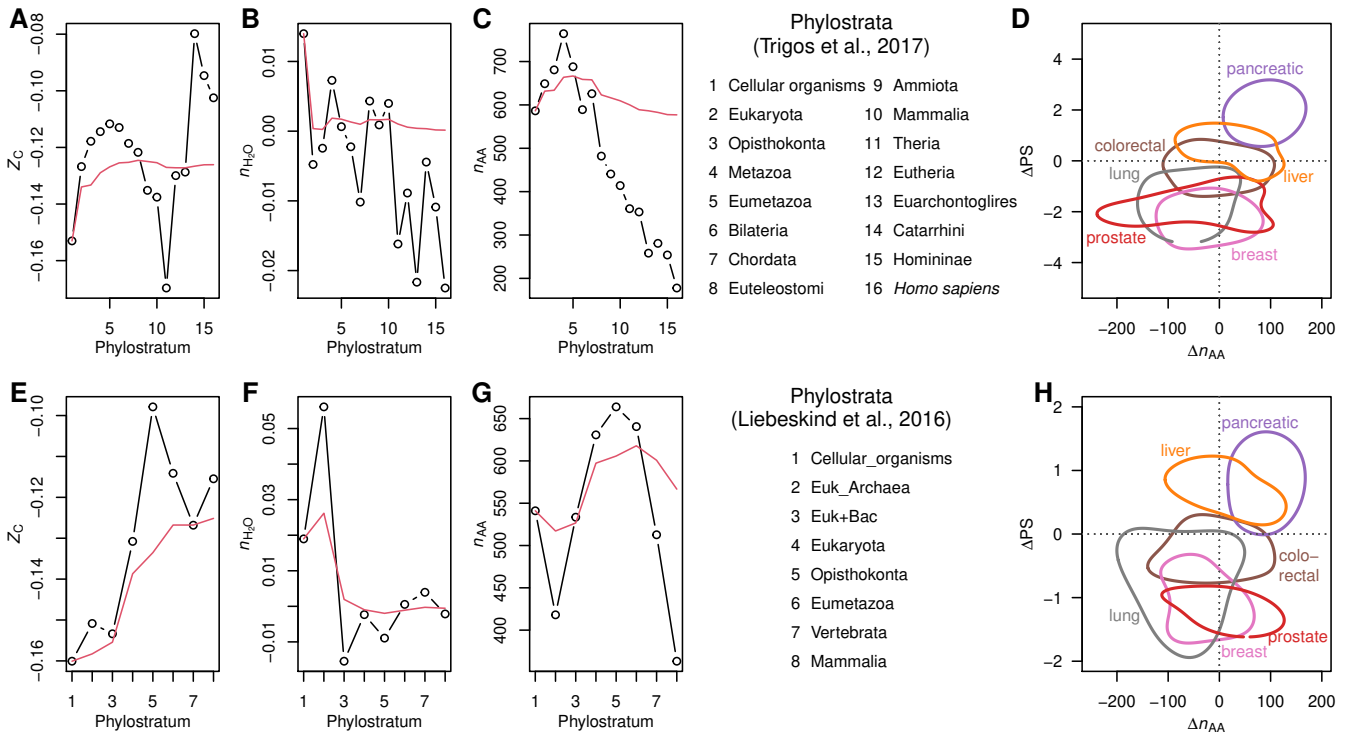

Fig. 4. Mean values of (**A**) Z_C, (**B**) n_{H_2O}, and (**C**) n_{AA} of proteins for all protein-coding genes in each phylostratum (PS) in the study of Trigos et al. (44). The points stand for the mean value in each phylostratum, and the red line indicates the cumulative mean starting from PS 1. The plot in (**D**) shows the 50% credible region for mean differences of PS and median differences of protein length (n_{AA}) for differential expression datasets compiled in this study for six types of cancer. Negative values of ΔPS correspond to older genes. Plots in (**E**)–(**H**) as above, except phylostrata assignments are based on gene ages given by Liebeskind et al. (45).

coded by differentially expressed genes in prostate cancer (Fig. 3*A*) might somehow be related to the hypoxic characteristics of normal prostate tissue (41) and the unusual metabolic profile of prostate cancer, in which the Warburg effect is absent except in late stages (42). Interestingly, the proteome of PC-3 prostate cancer cells under hypoxia is considerably reduced, unlike most other cell types under hypoxia, but combined treatment with sulindac (a non-steroidal anti-inflammatory drug) and radiation (43) reverses the trend (Fig. S6).

Relations between Phylostrata, Chemical Composition, and Protein Length. Several studies have linked gene expression in cancer to phylogenetically earlier genes (44, 46). The phylostratigraphic analysis used in these studies assigns ages of genes based on the latest common ancestor whose descendants have all the computationally detected homologs of that gene. To analyze the evolutionary trends of oxidation and hydration state of proteins, I used two sets of gene ages for human protein-coding genes: 16 phylostrata given by Trigos et al. (44), and eight gene ages based on consensus among different methods given by Liebeskind et al. (45). The phylostrata numbers are not identical in the two studies; the Liebeskind study has three steps between cellular organisms and Eukaryota, providing a greater resolution in earlier evolution, and stops at Mammalia, which corresponds to PS 10 in the Trigos phylostrata. In addition, the category named Euk+Bac in the Liebeskind compilation is not a phylogenetic lineage, but refers to genes present in eukaryotes and bacteria but not archaea; this category therefore represents the horizontal transfer of genes from bacteria to eukaryotes after the latter diverged from archaea (45).

Fig. 4 *A*–*B* shows distinct patterns of oxidation state and hydration state for proteins coded by genes aggregated according to the Trigos phylostrata. Z_C forms a strikingly smooth hump between PS 1 and 11 then increases rapidly to the maximum at PS 14, followed by a smaller decline to *Homo sapiens*. n_{H_2O} shows an overall decrease through time, but exhibits considerable small-scale variability. Keeping in mind the different resolutions and scales of the Trigos and Liebeskind gene ages, the two datasets show similar early maxima for Z_C and protein length and an overall evolutionary decrease of n_{H_2O} (Fig. 4 *A*–*C* and *E*–*G*).

Trigos et al. (44) used gene expression levels as weights to calculate the transcriptome age index (TAI), which was lower for seven cancer types compared to normal tissue, indicating higher expression of older genes. I used a different calculation, where ΔPS represents non-weighted mean differences between all up- and down-expressed genes, and obtained negative values for the same cancer types using the TCGA/GTEx data (Fig. S16). As shown in Fig. 3*C*, most cancer transcriptomes (TCGA/GTEx) are characterized by higher expression of older genes ($\Delta PS < 0$), whereas most proteomes (HPA) exhibit younger ages of the corresponding differentially expressed genes ($\Delta PS > 0$). The positive ΔPS for KIRC, which was not analyzed by Trigos et al. (44), is consistent with the large enrichment of vertebrate genes in this cancer type (46). The agreement with previous work suggests that ΔPS provides a reasonable metric for comparing gene ages in different cancer types.

Given the trends of mean protein length for phylostrata

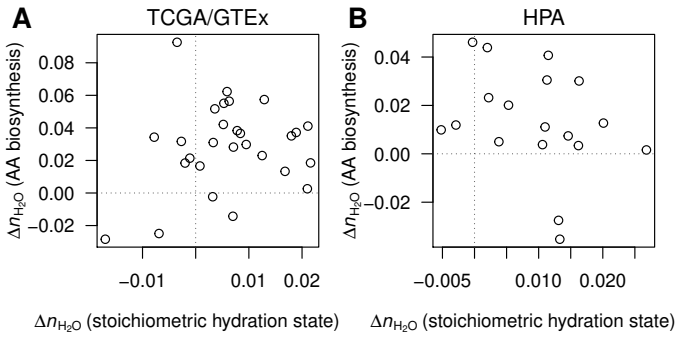

Fig. 5. Comparison of biosynthetic reaction coefficients and stoichiometric hydration state for calculating Δn_{H_2O}: (**A**) TCGA and (**B**) HPA data.

(Fig. 4*C*), the positive correlation between ΔPS and Δn_{AA} for differential expression in cancer (Fig. 4*D*) indicates that the corresponding genes are mostly present in Trigos PS 1 (cellular organisms) to 4 (metazoa). This agrees with previous studies in which differential gene expression in cancer was largely associated with genes spanning the unicellular–multicellular transition (44, 46).

The rise in protein length leading up to Eukaryota in both the Trigos and Liebeskind phylostrata is consistent with earlier reports that median protein length is greater in eukaryotes than prokaryotes (47). The decrease of protein length in later phylostrata is probably an artifact of BLAST-based homology searches (48), but does not greatly affect ages inferred for cancer-related genes. Fig. S3 shows that younger gene ages (i.e. higher ΔPS) in cancer are associated with relatively high oxidation state, as expected from the trends for the earliest phylostrata shown in Fig. 4; however, there is not a strong association between PS and hydration state for differentially expressed proteins in cancer.

Oxygen and Water in Amino Acid Biosynthetic Reactions. The analysis just described used compositional metrics that were formulated independently of any information about biosynthetic mechanisms. For comparison, it is useful to derive estimates of the amounts of water and oxygen in actual biosynthetic pathways for amino acid synthesis.

To examine simplified biosynthetic mechanisms, I used a standard depiction of pathways in the synthesis of amino acids (49) to identify six immediate precursor molecules for the amino acids: α-ketoglutarate (Glu, Gln, Pro, Arg), oxaloacetate (Asp, Asn, Met, Thr, Lys, Ile), pyruvate (Ala, Val, Leu), chorismate (Phe, Tyr, Trp), ribose-5-phosphate (His), and 3-phosphoglycerate (Ser, Gly, Cys). I wrote balanced reactions between the precursors and the amino acids in a 1:1 molar ratio by adding the appropriate amounts of CO_2, H_2O, NH_4^+, $H_2PO_4^-$, HS^-, O_2, and H^+. The derived reaction coefficients of H_2O and O_2 are listed in Table S3. The number of O_2 involved in the amino acid biosynthetic reactions is positively correlated with the Z_C of amino acids, but there is a weaker correlation between the number of H_2O in the biosynthetic reactions and in the rQEC derivation (Fig. S1 *C–D*).

The TCGA/GTEx and HPA datasets both exhibit predominantly positive values of Δn_{H_2O} calculated using either the biosynthetic reactions or stoichiometric hydration state, but the two metrics show little correlation with each other (Fig. 5). On the other hand, values of Δn_{O_2} in the biosynthetic reactions are positively correlated with ΔZ_C (Fig. S4 *A–B*), as expected for alternative metrics for oxidation state. More importantly, the biosynthetic reaction coefficients are characterized by a strong covariation of O_2 and H_2O (Fig. S4 *C–D*), which calls into question the uniqueness of hydration state calculated using this method. This covariation is reduced in the rQEC derivation of stoichiometric hydration state (20), making it more useful to distinguish condition-specific trends of oxidation and hydration state in differentially expressed proteins.

Further developments of genome-scale metabolic and macromolecular expression models (50) may lead to more precise estimates of the net water and oxygen demands for amino acid synthesis, uptake and incorporation into proteomes in different conditions. For the time being, the results of the present analyses of hydration state using both the rQEC stoichiometric derivation and biosynthetic reactions support the novel hypothesis that differential protein expression in most cancer types involves the consumption of water as a reactant.

Discussion

Despite the hypoxic nature of tumors, previous authors did not find significantly lower oxygen contents of proteins in glioma and stomach cancer compared to normal tissue (51, 52). Therefore, it is not surprising that in this study a range of differences in carbon oxidation state was documented, from very negative values for prostate cancer to positive values for colorectal, pancreatic, and liver cancer. These independent observations of biomolecular oxidation state should be compared with actual oxygen and redox measurements in tumors (e.g. 15, 16). Monitoring the levels of hypoxia-inducible factor (HIF-1) or its downstream targets (e.g. GLUT1 and VEGF) (53) or "hypoxia scores" for gene expression (54) is less suitable for these comparisons as they are not physicochemical measurements. Median hypoxia scores reported for 19 tumor types (55) are not correlated with differences of Z_C or n_{H_2O} from TCGA or HPA data (Fig. S5); therefore, hypoxia scores and chemical compositions of proteins likely reflect distinct physiological characteristics.

In marked contrast to the diverse trends of oxidation state, most cancer types are characterized by a higher stoichiometric hydration state of proteins. These results are complementary to experimental observations of elevated water content in tumors (7, 56, 57), and provide another line of evidence for a primary role for cellular hydration state in cancer (9). The results imply that sensitivity to water activity, as a thermodynamic indicator of of hydration potential, is deeply embedded in the network of metabolic reactions that maintains a dynamic proteome. More work is needed to elucidate the effects of water activity, which is modulated by solution composition and macromolecular crowding (58), on cellular metabolism in cancer. For instance, one question that can be asked is whether the proposed cancer-specific differences in water activity, which would affect the Gibbs energy of hydrolysis of ATP (59), may alter the extent of hydrolysis reactions that are thought to contribute to the production of protons in cancer (60). Similarly, a physicochemical link between protein length and water activity can be expected, as shorter proteins have more H_2O incorporated into the terminal groups in proportion to their mass. Interestingly, the compiled differential expression datasets for breast cancer and the pan-cancer tran-

scriptomes (TCGA/GTEx) both show a significant decrease in protein length (Table 2).

The general applicability of the compositional analysis is supported by proteomic data for cell culture in controlled laboratory experiments. In particular, the hydration state of differentially expressed proteins is significantly lower in the majority of hyperosmotic stress experiments (Fig. 2*C* and Table 2). In response to changing salinities, the interiors of most cells must be at least isosmotic with the environment to maintain a physiological water content (61). Nevertheless, the water content of cells grown in hyperosmotic NaCl solutions is actually substantially lowered, as shown for *E. coli* (62). The present results support the hypothesis that osmotically induced dehydration provides a thermodynamic drive for the preferential expression of proteins with lower stoichiometric hydration state.

Because many of the hyperosmotic stress experiments analyzed here are for yeast and microbial cells, more data are needed to ascertain whether these findings extend to tissue environments. It is notable that the hydration state of epidermal proteins decreases for wild-type mice kept in low humidity (40%) compared to high humidity (70%), but the trend is reversed (see Fig. S7) in mice that have a deficiency in fibroblast growth factor receptor (FGFR1/FGFR2) in keratinocytes (63), which is associated with a dysfunctional epidermal barrier. This suggests that not only physicochemical conditions but also tissue environments can influence the hydration state of cellular proteins.

An unexpected finding is that the hydration state of proteins is substantially lower in 3D culture, including spheroids and aggregates, compared to traditional 2D culture in monolayers (Fig. 2 *A* and *C*; see also Fig. S9). This finding might be linked with the less liquid-like state of the cytoplasm in 3D culture (64). These results are also congruent with metagenomes of particle-sized fractions compared to free-living microbes in river and marine samples; the former, which are more likely to harbor multicellular communities, are associated with lower n_{H_2O} of the coded proteins (20). In addition, up-regulated proteins in 3D culture also tend to be more reduced (Table 2), which might be a reflection of the hypoxic conditions that develop in the interiors of spheroids (21, 22, 65).

It is somewhat surprising that hypoxia experiments as a whole do not induce the up-regulation of more reduced proteins. Moreover, secreted proteins in hypoxia are more often relatively oxidized (Table 2). Hypoxia is often associated with the downregulation of mitochondrial proteins (66, 67). These have a relatively low Z_C compared to other subcellular fractions such as cytoplasm and nucleus (26), so their downregulation would tend to produce overall more oxidized proteins. The proteomes of specific subcellular fractions should be analyzed to better understand the complete cellular response. As noted above for PC-3 prostate cancer cells, treatment with drugs and radiation also strongly influences the differences in oxidation state of differentially expressed proteins.

Cancer has often been regarded as a reversion of both developmental (68, 69) and evolutionary (46) processes. The hypothesis of a major component of atavism in cancer (70) is supported by the up-regulation of older genes (44, 46). I obtained similar results from analysis of pan-cancer transcriptomic data, but found an opposite trend in many proteomics datasets (Fig. 3*C*), which may be an indication that post-transcriptional regulation masks the atavism signal. However, the issue is not settled, as among the six cancer types analyzed in detail here, only pancreatic cancer and to a lesser extent liver cancer are characterized by younger gene ages of the differentially expressed proteins (Fig. 4 *D* and *H*). In particular, the causes for the large discrepancy for prostate cancer between HPA data ($\Delta PS > 0$) and the compilation of differential expression datasets ($\Delta PS < 0$) are not clear.

An important area for future work is to document the relations between chemical composition of proteins and physiological oxygenation and hydration levels through development; according to the hypothesis that "oncogenesis recapitulates ontogenesis" (46), this might uncover deeper links among the patterns described above for cell culture and cancer datasets. The high water content characteristic of early fetuses in mammals declines significantly through gestation and continues to decline post-birth (71, 72). I would therefore predict that the hydration state of embryonic proteins is higher than that in adults; this can be tested using recent proteomic datasets for model organisms (73).

Concluding Remarks

This paper reports the first large-scale analysis of chemical compositions of proteins using differential expression data spanning multiple types of cancer and laboratory experiments. Differentially expressed proteins in hyperosmotic and 3D cell-culture experiments are on average shifted toward lower stoichiometric hydration state, but an increase in stoichiometric hydration state characterizes up-regulated proteins in five of six cancer types considered in detail (except pancreatic cancer) and most pan-cancer proteomic and transcriptomic data.

In contrast to hydration state, differences of carbon oxidation state calculated for pan-cancer proteomes are distributed more evenly among positive and negative values. This provides evidence against the hypothesis that hypoxic conditions in tumors drive changes in the oxidation state of proteins; instead, the compositional differences might be driven by the enrichment in different tissues of secreted (relatively oxidized) or membrane (reduced) proteins. On the other hand, pan-cancer transcriptomes are associated with generally more reduced proteins, so tumor hypoxia may have a stronger influence on chemical composition at the gene expression level.

The focus on oxidation and hydration state in this study was shaped by the dual observations that tumor microenvironments are typically hypoxic and have a relatively high water content, and that both oxidation-reduction and hydration-dehydration reactions have a major role in metabolism (12, 13). However, it remains challenging to interpret the proteome-level differences in chemical composition within the framework of molecular biology, structural biology and biochemistry, which are mainly concerned with the coding and regulation, structure, and enzymatic functions of proteins. As a counterpart to these established fields of investigation, further developments in the area of compositional biology are needed to advance our understanding of the genome–cell–environment interactions that drive changes in the abundances of biomacromolecules through the progression of cancer, including metastasis.

Materials and Methods

Proteomics Datasets. Differential protein expression data for cell culture experiments and cancer compared to normal tissue were located through literature searches and include studies using any proteomics techniques. Several review articles were also consulted in order to locate experimental data for breast cancer (74), lung cancer (75, 76) and 3D cell culture (65). In general, datasets were selected that have a minimum of 30 up-regulated and 30 down-regulated proteins in order to reduce random variation associated with small sample sizes, but smaller datasets (at least ca. 20 up-regulated and 20 down-regulated proteins) were included for hyperosmotic stress, secreted in hypoxia, lung cancer, and prostate cancer due to limited availability of data.

Previous compilations for hypoxia and and colorectal and pancreatic cancer (19) were updated in this study using more recently located datasets. Datasets related to prognosis, conditioned media, stromal samples, and adenoma were removed from the updated compilation for colorectal cancer. In addition, datasets for cellular and secreted proteins in hypoxia were considered separately, and datasets for reoxygenation after hypoxia were excluded. The previous compilation of data for hyperosmotic stress (19) was also expanded in this study, but a fish gill proteome and two transcriptomic datasets were excluded, and more high-glucose datasets were included.

Lists of significantly differentially expressed proteins were taken directly from the original publications if possible. In cases of datasets where mass spectrometric data but not lists of differentially expressed proteins were reported, quantile normalization using function "normalize.quantiles" in the R package preprocessCore (77) was performed on the intensities or peak areas in order to obtain normalized values that were used for differential expression analysis. Where needed, reported protein or gene identifiers were converted to UniProt IDs using the UniProt mapping tool (78). Protein sequences downloaded from UniProt were used to generate amino acid compositions using function "read.fasta" in the R package CHNOSZ (79). The canonical protein sequences in UniProt were used, unless isoforms were identified in the data sources. References for all data sources and details of additional processing steps are given with Figs. S6–S15.

Differential Expression from Pan-Cancer Datasets. Immunohistochemistry-based expression profiles of proteins in normal tissue and pathology samples were downloaded from the Human Protein Atlas version 19 (31, 36). Pathology and normal tissue datasets were paired based on information from the HPA web site (80): breast cancer / breast; cervical cancer / cervix, uterine; colorectal cancer / colon; endometrial cancer / endometrium 1; glioma / cerebral cortex; head and neck cancer / salivary gland; liver cancer / liver; lung cancer / lung; lymphoma / lymph node; melanoma / skin 1; skin cancer / skin 1; ovarian cancer / ovary; pancreatic cancer / pancreas; prostate cancer / prostate; renal cancer / kidney; stomach cancer / stomach 1; testis cancer / testis; thyroid cancer / thyroid gland; urothelial cancer / urinary bladder. Antibody staining intensities were converted to a semi-quantitative scale (not detected: 0, low: 1, medium: 3, high: 5). The expression level score for each protein was calculated by averaging the score for available samples, including "not detected" but excluding unavailable (NA) observations, and, for normal tissues, observations in all available cell types. Differences in expression score between normal and cancer $\geq$ 2.5 or $\leq$ -2.5 were considered to be differentially expressed proteins.

Differential gene expression values were obtained using version 2 of the Gene Expression Profiling Interactive Analysis web server (GEPIA2) (32) with default settings (ANOVA, $\log_2$ fold change cutoff = 1, q-value cutoff = 0.01). Pairings between source datasets for cancer (TCGA) and normal tissue (GTEx), as described on the GEPIA2 website (81) are: ACC / adrenal gland; BLCA / bladder; BRCA / breast; CESC / cervix uteri; COAD / colon; DLBC / blood; ESCA / esophagus; GBM / brain; KICH / kidney; KIRC / kidney; KIRP / kidney; LAML / bone marrow; LGG / brain; LIHC / liver; LUAD / lung; LUSC / lung; OV / ovary; PAAD / pancreas; PRAD / prostate; READ / colon; SKCM/ skin; STAD / stomach; TGCT / testis; THCA / thyroid; THYM / blood; UCEC / uterus; UCS / uterus. Gene expression data for both tumor and normal tissue for HNSC are from TCGA. Differential expression data were not available on GEPIA2 for five other cancer types in TCGA (CHOL, MESO, PCPG, SARC, UVM). Ensembl Gene IDs used in HPA and GEPIA were converted to UniProt accession numbers using the UniProt mapping tool (78).

Compositional Metrics. Values of average oxidation state of carbon (Z_C) of amino acids (Table 1) were calculated from the chemical formulas of the amino acids (25, 26). Values for Z_C of proteins were computed by combining the amino acid compositions of proteins with Z_C of amino acids and also weighting by carbon number (20). That is, $Z_C = \Sigma\ Z_{C,i} n_i n_{C,i} / \Sigma\ n_i n_{C,i}$, where the summation is over $i = 1..20$ amino acids and $Z_{C,i}$, n_i, and $n_{C,i}$ are the carbon oxidation state, frequency in the protein sequence, and number of carbon atoms of the ith amino acid, respectively.

Values of stoichiometric hydration state (n_{H_2O}) for amino acids (Table 1) were calculated using the rQEC derivation described by Dick et al. (20). Briefly, the numbers of H_2O in theoretical formation reactions for the 20 amino acid residues were obtained by projecting the elemental compositions of the amino acids into the basis species glutamine, glutamic acid, cysteine, H_2O, and O_2 (QEC basis species; see Table S1 and Fig. S1A). The stoichiometric hydration state was then obtained by calculating the residuals of a linear model fit to n_{H_2O} and Z_C for the amino acid residues, then subtracting a constant from the residuals to make the mean value for all human proteins equal to zero. The residual analysis ensures that there is no correlation between n_{H_2O} and Z_C of amino acids (Fig. S1B). To compute per-residue values for proteins, the values of n_{H_2O} for amino acid residues from Table 1 were combined with the amino acid compositions of the proteins. That is, $n_{H_2O} = \Sigma\ n_{H_2O,i} n_i / \Sigma\ n_i$, where $n_{H_2O,i}$ and n_i are the stoichiometric hydration state and frequency of the ith amino acid residue, respectively. Accordingly, $\Delta n_{H_2O} = 0.01$ corresponds to a difference of approximately 3 water molecules in the theoretical formation reaction of a typical 300-residue protein from the QEC basis species.

Phylostrata. Phylostrata were obtained from the supporting information of Trigos et al. (44) and the "main_HUMAN.csv" file of Liebeskind et al. (45, 82). Liebeskind et al. did not give phylostrata numbers, so Phylostrata 1–8 were assigned here based on the names in the "modeAge" column of the source file (see Fig. 4). The Ensembl gene identifiers in the Trigos dataset were converted to UniProt accession numbers (78); in the case of duplicate UniProt accession numbers, the first matching phylostratum was used. Phylostrata differences were not computed for non-human organisms.

ACKNOWLEDGMENTS. I am grateful to Alex Greenhough, Youngsoo Kim, Ming-Chih Lai, and Gordana Vunjak-Novakovic for providing data files. The results shown here are in part based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga) and the Human Protein Atlas (https://www.proteinatlas.org).

References

1. Vogelstein B, et al. (2013) Cancer genome landscapes. *Science* 339(6127):1546–1558.
2. Wang M, et al. (2017) Role of tumor microenvironment in tumorigenesis. *Journal of Cancer* 8(5):761–773.
3. Höckel M, Vaupel P (2001) Tumor hypoxia: Definitions and current clinical, biologic, and molecular aspects. *Journal of the National Cancer Institute* 93(4):266–276.
4. Winzler RJ (1959) The chemistry of cancer tissue in *The Physiopathology of Cancer*, ed. Homburger F. (Hoeber-Harper, New York), 2nd edition, pp. 686–706.
5. Downing JE, Christopherson WM, Broghamer WL (1962) Nuclear water content during carcinogenesis. *Cancer* 15(6):1176–1180.
6. Ross KFA, Gordon RE (1982) Water in malignant tissue, measured by cell refractometry and nuclear magnetic resonance. *Journal of Microscopy* 128(1):7–21.
7. Surmacki J, Musial J, Kordek R, Abramczyk H (2013) Raman imaging at biological interfaces: Applications in breast cancer diagnosis. *Molecular Cancer* 12(1):48.
8. Barroso EM, et al. (2015) Discrimination between oral cancer and healthy tissue based on water content determined by Raman spectroscopy. *Analytical Chemistry* 87(4):2419–2426.
9. McIntyre GI (2006) Cell hydration as the primary factor in carcinogenesis: A unifying concept. *Medical Hypotheses* 66(3):518–526.

10. Munder MC, et al. (2016) A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy. *eLife* 5:e09347.
11. Nelson DM, Cox MM (2005) *Lehninger Principles of Biochemistry*. (W. H. Freeman and Company, New York), 4th edition.
12. Morowitz HJ (1999) A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* 4(6):39–53.
13. Braakman R, Smith E (2013) The compositional and evolutionary logic of metabolism. *Physical Biology* 10(1):011001.
14. Filipp FV, Scott DA, Ronai ZA, Osterman AL, Smith JW (2012) Reverse TCA cycle flux through isocitrate dehydrogenases 1 and 2 is required for lipogenesis in hypoxic melanoma cells. *Pigment Cell & Melanoma Research* 25(3):375–383.
15. Kuppusamy P, et al. (1998) *In vivo* electron paramagnetic resonance imaging of tumor heterogeneity and oxygenation in a murine model. *Cancer Research* 58(7):1562–1568. https://cancerres.aacrjournals.org/content/58/7/1562.
16. Hyodo F, et al. (2012) The relationship between tissue oxygenation and redox status using magnetic resonance imaging. *International Journal of Oncology* 41(6):2103–2108.
17. Guzy RD, Schumacker PT (2006) Oxygen sensing by mitochondria at complex III: The paradox of increased reactive oxygen species during hypoxia. *Experimental Physiology* 91(5):807–819.
18. Häussinger D, Lang F, Gerok W (1994) Regulation of cell function by the cellular hydration state. *American Journal of Physiology* 267(3):E343–E355.
19. Dick JM (2017) Chemical composition and the potential for proteomic transformation in cancer, hypoxia, and hyperosmotic stress. *PeerJ* 5:e3421.
20. Dick JM, Yu M, Tan J (2020) Distinct trends in chemical composition of proteins from metagenomes in redox and salinity gradients. *bioRxiv* 10.1101/2020.04.01.020008.
21. Hirschhaeuser F, et al. (2010) Multicellular tumor spheroids: An underestimated tool is catching up again. *Journal of Biotechnology* 148(1):3–15.
22. Baker BM, Chen CS (2012) Deconstructing the third dimension – how 3D culture microenvironments alter cellular cues. *Journal of Cell Science* 125(13):3015–3024.
23. Chen H, He X (2016) The convergent cancer evolution toward a single cellular destination. *Molecular Biology and Evolution* 33(1):4–12.
24. Amend JP, LaRowe DE, McCollom TM, Shock EL (2013) The energetics of organic synthesis inside and outside the cell. *Philosophical Transactions of the Royal Society, B: Biological Sciences* 368(1622):20120255.
25. Dick JM, Shock EL (2011) Calculation of the relative chemical stabilities of proteins as a function of temperature and redox chemistry in a hot spring. *PLOS One* 6(8):e22782.
26. Dick JM (2014) Average oxidation state of carbon in proteins. *Journal of the Royal Society Interface* 11:20131095.
27. Bolker BM (2008) *Ecological Models and Data in R*. (Princeton University Press, Princeton, NJ).
28. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*. (Springer, New York), 4th edition. http://www.stats.ox.ac.uk/pub/MASS4.
29. R Core Team (2020) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria). https://www.R-project.org.
30. National Cancer Institute (2018) TCGA Study Abbreviations (https://gdc.cancer.gov/resources-tcga-users/tcga-code-tables/tcga-study-abbreviations last accessed on 2020-01-30).
31. Uhlén M, et al. (2015) Tissue-based map of the human proteome. *Science* 347(6220):1260419.
32. Tang Z, Kang B, Li C, Chen T, Zhang Z (2019) GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Research* 47(W1):W556–W560.
33. Goldman M, et al. (2019) The UCSC Xena platform for public and private cancer genomics data visualization and interpretation. *bioRxiv* 326470.
34. GTEx Consortium (2017) Genetic effects on gene expression across human tissues. *Nature* 550(7675):204–213.
35. The Cancer Genome Atlas Research Network, et al. (2013) The Cancer Genome Atlas Pan-Cancer analysis project. *Nature Genetics* 45(10):1113–1120.
36. Uhlen M, et al. (2017) A pathology atlas of the human cancer transcriptome. *Science* 357(6352):eaan2507.
37. Zhang YJ, et al. (2018) Subcellular stoichiogenomics reveal cell evolution and electrostatic interaction mechanisms in cytoskeleton. *BMC Genomics* 19(1):469.
38. Vohra R, et al. (2018) Evaluation of pancreatic tumor development in *KPC* mice using multi-parametric mri. *Cancer Imaging* 18(1):41.
39. Genina EA, et al. (2019) Optical properties of brain tissues at the different stages of glioma development in rats: pilot study. *Biomedical Optics Express* 10(10):5182–5197.
40. Ali JH, Wang WB, Zevallos M, Alfano RR (2004) Near infrared spectroscopy and imaging to probe differences in water content in normal and cancer human prostate tissues. *Technology in Cancer Research & Treatment* 3(5):491–497.
41. Parker C, et al. (2004) Polarographic electrode study of tumor oxygenation in clinically localized prostate cancer. *International Journal of Radiation Oncology*Biology*Physics* 58(3):750–757.
42. Eidelman E, Twum-Ampofo J, Ansari J, Siddiqui MM (2017) The metabolic phenotype of prostate cancer. *Frontiers in Oncology* 7:131.
43. Ross JA, et al. (2020) The influence of hypoxia on the prostate cancer proteome. *Clinical Chemistry and Laboratory Medicine*.
44. Trigos AS, Pearson RB, Papenfuss AT, Goode DL (2017) Altered interactions between unicellular and multicellular genes drive hallmarks of transformation in a diverse range of solid tumors. *Proceedings of the National Academy of Sciences* 114(24):6406–6411.
45. Liebeskind BJ, McWhite CD, Marcotte EM (2016) Towards consensus gene ages. *Genome Biology and Evolution* 8(6):1812–1823.
46. Zhou JX, et al. (2018) Phylostratigraphic analysis of tumor and developmental transcriptomes reveals relationship between oncogenesis, phylogenesis and ontogenesis. *Convergent Science Physical Oncology* 4(2):025002.
47. Brocchieri L, Karlin S (2005) Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Research* 33(10):3390–3400.
48. Moyers BA, Zhang J (2017) Further simulations and analyses demonstrate open problems of phylostratigraphy. *Genome Biology and Evolution* 9(6):1519–1527.
49. Wikipedia (2020) Amino acid synthesis (https://en.wikipedia.org/wiki/Amino_acid_synthesis last accessed on 2020-01-05).
50. Yang L, et al. (2016) Principles of proteome allocation are revealed using proteomic data and genome-scale models. *Scientific Reports* 6(1):36734.
51. Yin Y, et al. (2019) Stoichioproteomics reveal oxygen usage bias, key proteins and pathways in glioma. *BMC Medical Genomics* 12(1):125.
52. Zuo X, et al. (2019) Stoichiogenomics reveal oxygen usage bias, key proteins and pathways associated with stomach cancer. *Scientific Reports* 9(1):11344.
53. Stockwin LH, et al. (2006) Proteomic analysis of plasma membrane from hypoxia-adapted malignant melanoma. *Journal of Proteome Research* 5(11):2996–3007.
54. Harris BHL, Barberis A, West CML, Buffa FM (2015) Gene expression signatures as biomarkers of tumour hypoxia. *Clinical Oncology* 27(10):547–560.
55. Bhandari V, et al. (2019) Molecular landmarks of tumor hypoxia across cancer types. *Nature Genetics* 51(2):308–318.
56. Beall PT (1983) States of water in biological systems. *Cryobiology* 20(3):324–334.
57. Sun Q, et al. (2017) Recent advances in terahertz technology for biomedical applications. *Quantitative Imaging in Medicine and Surgery* 7(3):345–355.
58. Nakano Si, Miyoshi D, Sugimoto N (2014) Effects of molecular crowding on the structures, interactions, and functions of nucleic acids. *Chemical Reviews* 114(5):2733–2758.
59. de Meis L (1989) Role of water in the energy of hydrolysis of phosphate compounds: Energy transduction in biological membranes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 973(2):333–349.
60. Sun H, et al. (2020) Metabolic reprogramming in cancer is induced to increase proton production. *Cancer Research* 80(5):1143–1155.
61. Gunde-Cimerman N, Plemenitaš A, Oren A (2018) Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS Microbiology Reviews* 42(3):353–375.
62. Record, Jr. MT, Courtenay ES, Cayley DS, Guttman HJ (1998) Responses of *E. coli* to osmotic stress: large changes in amounts of cytoplasmic solutes and water. *Trends in Biochemical Sciences* 23(4):143–148.
63. Seltmann K, et al. (2018) Humidity-regulated CLCA2 protects the epidermis from hyperosmotic stress. *Science Translational Medicine* 10(440):eaao4650.
64. Staunton JR, So WY, Paul CD, Tanner K (2019) High-frequency microrheology in 3D reveals mismatch between cytoskeletal and extracellular matrix mechanics. *Proceedings of the National Academy of Sciences* 116(29):14448–14455.
65. Acland M, et al. (2018) Mass spectrometry analyses of multicellular tumor spheroids. *Proteomics: Clinical Applications* 12(3):1700124.
66. Wobma HM, et al. (2018) The influence of hypoxia and IFN-γ on the proteome and metabolome of therapeutic mesenchymal stem cells. *Biomaterials* 167:226–234.
67. Kugeratski FG, et al. (2019) Hypoxic cancer-associated fibroblasts increase NCBP2-AS2/HIAR to promote endothelial sprouting through enhanced VEGF signaling. *Science Signaling* 12(567):eaan8247.
68. Naxerova K, et al. (2008) Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biology* 9(7):R108.
69. Ma Y, et al. (2010) The relationship between early embryo development and tumourigenesis. *Journal of Cellular and Molecular Medicine* 14(12):2697–2701.
70. Davies PCW, Lineweaver CH (2011) Cancer tumors as Metazoa 1.0: tapping genes of ancient ancestors. *Physical Biology* 8(1):015001.
71. Moulton CR (1923) Age and chemical development in mammals. *Journal of Biological Chemistry* 57(1):79–97.
72. Lindower JB (2017) Water balance in the fetus and neonate. *Seminars in Fetal and Neonatal Medicine* 22(2):71–75.
73. Fabre B, et al. (2019) Comparison of *Drosophila melanogaster* embryo and adult proteome by SWATH-MS reveals differential regulation of protein synthesis, degradation machinery, and metabolism modules. *Journal of Proteome Research* 18(6):2525–2534.
74. Gromov P, Moreira JM, Gromova I (2014) Proteomic analysis of tissue samples in translational breast cancer research. *Expert Review of Proteomics* 11(3):285–302.
75. Fujii K, Nakamura H, Nishimura T (2017) Recent mass spectrometry-based proteomics for biomarker discovery in lung cancer, COPD, and asthma. *Expert Review of Proteomics* 14(4):373–386.
76. Calabrese F, et al. (2019) Are there new biomarkers in tissue and liquid biopsies for the early detection of non-small cell lung cancer? *Journal of Clinical Medicine* 8(3):414.
77. Bolstad B, Irizarry R, Åstrand M, Speed T (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19(2):185–193.
78. Huang H, et al. (2011) A comprehensive protein-centric ID mapping service for molecular data integration. *Bioinformatics* 27(8):1190–1191.
79. Dick JM (2019) CHNOSZ: Thermodynamic calculations and diagrams for geochemistry. *Frontiers in Earth Science* 7:180.
80. The Human Protein Atlas (2019) Dictionary: Pathology Overview (https://www.proteinatlas.org/learn/dictionary/pathology last accessed on 2020-01-31).
81. GEPIA2 (2019) Dataset Sources (http://gepia2.cancer-pku.cn/#dataset last accessed on 2020-01-31).
82. Liebeskind B, clairemcwhite, Hines K (2016) Gene-Ages v1.0 (https://doi.org/10.5281/zenodo.51708 last accessed on 2020-02-02).