Quantitative Estimation of Oxidative Stress in Cancer Tissue Cells Through Gene Expression Data Analyses

Liyang Liu¹,², Haining Cui ¹* and Ying Xu²,³*

¹ College of Physics, Jilin University, Changchun, China, ² Department of Biochemistry and Molecular Biology, Institute of Bioinformatics, The University of Georgia, Athens, GA, United States, ³ Cancer Systems Biology Center, The China-Japan Union Hospital, Jilin University, Changchun, China

Quantitative assessment of the intracellular oxidative stress level is a very important problem since it is the basis for elucidation of the fundamental causes of metabolic changes in diseased human cells, particularly cancer. However, the problem proves to be very challenging to solve in vivo because of the complex nature of the problem. Here a computational method is presented for predicting the quantitative level of the intracellular oxidative stress in cancer tissue cells. The basic premise of the predictor is that the genomic mutation level is strongly associated with the intracellular oxidative stress level. Based on this, a statistical analysis is conducted to identify a set of enzyme-encoding genes, whose combined expression levels can well explain the mutation rates in individual cancer tissues in the TCGA database. We have assessed the validity of the predictor by assessing it against genes that are known to have anti-oxidative functions for specific types of oxidative stressors. Then the applications of the predictor are conducted to illustrate its utility.

Keywords: oxidative stress, genomic mutation, transcriptomic data, cancer, TCGA data analysis, computational prediction

INTRODUCTION

Oxidative stress has long been recognized as a cellular stress associated with cancer formation and development (Cerutti and Trump, 1993; Wangpaichitr et al., 2009; Bellezza et al., 2010; Chan et al., 2010; Conti et al., 2010; Dayem et al., 2010; Gibellini et al., 2010; Henkler et al., 2010; Karlenius and Tonissen, 2010; Massi et al., 2010; Ortega et al., 2010; Pizzimenti et al., 2010; Reuter et al., 2010; Schultz et al., 2010; Soory, 2010; Sun et al., 2010; Valle et al., 2010; Gorrini et al., 2013; Sosa et al., 2013; Leone et al., 2017). It refers to the gap between the intracellular oxidizing power and the anti-oxidation capacity in a human cell. Numerous authors have pointed out that oxidative stress may be responsible for the induction of a variety of altered metabolisms in cancer, which include (i) considerably increased lipid metabolism (Santos and Schulze, 2012; Zhang and Du, 2012; Alfaradhi et al., 2014; Zhang et al., 2014), (ii) altered metabolisms of sulfur-containing amino acids (Schulz et al., 2000; Zhang et al., 2012; Campbell et al., 2016; Martinez et al., 2017), (iii) reprogrammed sugar metabolisms (Takeyama et al., 2000; Wu et al., 2009; Dewald et al., 2016), among other reprogrammed metabolisms. Increased genomic mutations in cancer have also been...
attributed to oxidative stress (Limoli and Giedzinski, 2003; Doudican et al., 2005; Slane et al., 2006; Xu et al., 2014; Fitzgerald et al., 2017; Markkanen, 2017; Rao et al., 2017; Tubbs and Nussenzweig, 2017). However, it has proven to be very challenging to accurately estimate the level of intracellular oxidative stress (Selvaraj et al., 2002; Farah, 2005). Because existing techniques are mostly designed to detect quantitatively specific oxidizing molecular species (e.g., H$_2$O$_2$ and O$_2$) rather than detecting the overall level of oxidative stress. Computational techniques could potentially play a key role in estimating the quantitative level of oxidative stress.

Unlike hypoxia or acid-base imbalance whose related stress can be measured/estimated in terms of the concentration of one or a few molecular species such as O$_2$ or H+, there are many types of oxidizing molecules such as reactive oxygen species (Barash et al., 2010; Reczek and Chandel, 2017), reactive nitrogen species (Fionda et al., 2016; Kruk and Aboul-Enein, 2017), reactive lipid species (Higdon et al., 2012; Graham, 2016) among others. Furthermore, to estimate the stress level, it also requires to know the total cellular reducing capacity. This is also challenging since human cells not only have a basic set of anti-oxidation (or reducing) capacities consisting of (1) glutathione and associated enzymes (e.g., glutathione transferase and peroxidase), (2) vitamin A, C, E and derivatives like beta-carotene, and (3) anti-oxidation enzymes like SOD (superoxide dismutase), PRX (peroxisin), and TRX (thioredoxin) (Chang et al., 2008; Mantovani et al., 2010), but also rely on anti-oxidation capacities of various other molecules including some enzymes and fatty acids. Among them, the main functions of the enzymes may not be for anti-oxidation (Osmundsen et al., 1982; Wieczorek et al., 2017), while the anti-oxidative properties of fatty acids have long been well-established (Richard et al., 2008; Freitas et al., 2017). All these make it very difficult to pin down on what molecular species should be used when assessing the level of oxidative stress.

A few proposals have been made regarding possible biomarkers for intracellular level of oxidative stress such as the carbonylation level (aldehydes and ketones) of proteins (Dalle-Donne et al., 2003; Haciseyki et al., 2012; Fernando et al., 2016), the level of oxidized low-density lipoprotein (Itabe, 2012; Osman et al., 2016), oxidized products of lipids such as 4-HNE (4-hydroxynonenal) and MDA (malondialdehyde) (Niki, 2008; Halder and Bhattacharyya, 2014; Teppner et al., 2015), and protein thios (Giustarini et al., 2012, 2017). There are two general issues with these biomarkers: (1) they tend to reflect the level of oxidation by specific oxidizing molecules; and (2) more importantly, they are not high-throughput, hence it is impractical for large-scale analyses, such as analyses of TCGA tissue samples (https://portal.gdc.cancer.gov/) to elucidate possible causes of specific metabolic changes in such tissues.

The goal here is to identify a set of genes whose mRNA expression levels can collectively reflect the overall level of intracellular oxidative stress present in a cancer tissue. The strategy is: (1) the somatic point-mutation rate in each cancer tissue sample is used as an indicator for the level of intracellular oxidative stress, an idea that has been well-established and applied (Doudican et al., 2005; Slane et al., 2006; Xu et al., 2014; Fitzgerald et al., 2017; Markkanen, 2017; Rao et al., 2017; Tubbs and Nussenzweig, 2017); (2) a few enzyme classes are selected, which are known to have anti-oxidation activities such as EC 3.1.1., EC 3.6.1., EC 2.4.1., and EC 2.7.7. (Kato et al., 1995; Kong et al., 2007); and (3) a subset of genes are selected from these enzyme classes, whose combined expressions correlate strongly with the mutation rates in the matching genomes, determined through regression analyses. One assumption used in the analysis is that the oxidative stress levels in cytosol and nucleus are the same, which is reasonable knowing that the nuclear pores are large enough to allow most of the oxidizing molecules to go through freely between the two compartments; and increased mutation rates in cancer are known to be related to the nucleus oxidative stress level (Chung et al., 2014; Markkanen, 2017).

By applying this strategy to gene-expression data and matching genomic mutation data of cancer tissues of 14 cancer types in TCGA, representing all those with sufficiently large sample sizes, we have trained a predictor for the intracellular level of oxidative stress for each of the 14 cancer types. They are BLCA (bladder urothelial carcinoma), BRCA (breast invasive carcinoma), COAD (colon adenocarcinoma), ESCA (esophageal carcinoma), HNSC (head and neck squamous cell carcinoma), KICH (kidney renal papillary cell carcinoma), KIRC (kidney chromophobe), KIRP (kidney renal clear cell carcinoma), LICH (liver hepatocellular carcinoma), LUAD (lung squamous cell carcinoma), LUSC (lung adenocarcinoma), PRAD (prostate adenocarcinoma), STAD (stomach adenocarcinoma), and THCA (thyroid carcinoma). We have then validated the predictor against data with known oxidative stress related information. A key advantage in having a gene-expression data-based predictor is that RNA-seq data tend to be collected in general for cancer research; and such a predictor does not require a user to know the mutation rate distribution of the cancer type under study since such information is already encoded in the predictor for each cancer type.

RESULTS
Genomic Mutation Profiles
We have calculated and plotted the distribution of the number of point-mutations in coding regions per genome across all cancer genomes for each of the 14 cancer types, as shown in Figure 1. From the figure, LUAD has the highest average mutation rate at 347.20 per genome, and KICH has the lowest one at 19.75 per genome. The following lists the average numbers of the remaining 12 cancer types in the descending order: 288.95 mutations in LUSC, 272.02 in COAD, 257.50 in STAD, 234.71 in BLCA, 213.26 in HNSC, 124.59 in ESCA, 147.65 in LIHC, 303.75 in BRCA, 111.14 in KIPR, 113.94 in KIRC, 138.75 in PRAD, and 134.18 in THCA.

Selection of Antioxidant Enzyme-Encoding Genes
Some classes of enzymes have long been known to have anti-oxidative functions as their main function such as the antioxidant enzymes mentioned earlier while some others have anti-oxidation as their secondary function such as cytochromes
P450 and mitogen activated protein kinases (Limón-Pacheco and Gonsebatt, 2009). And yet, increasingly more proteins have been found to have antioxidant roles in addition to their main functions such as translocases (Tang et al., 2017), hydrolase (Liu et al., 2018), and hexokinase (Heneberg, 2019).

Based on such information, we have conducted a preliminary regression analysis of the mutation rates against the expression data of all the enzyme-encoding genes in the same samples for each of the 14 cancer types (with a Lasso penalty). Interestingly, genes that give rise to good regression results across all 14 cancer types generally fall into a small set of enzyme subclasses, particularly EC 3.- (hydrolases) and EC 2.- (transferases). Hence, we have conducted a second round of regression against expression data of genes only in these two EC classes.

From genes selected for each of the 14 regression models, we note the following: (1) in all 14 cancer types, genes in each regression model fall into exactly four sub-subclasses of EC 2.- and EC 3.-; (2) in five cancer types: BLCA, KIRC, KIRP, LUAD, and LUSC, all genes fall into four sub-subclasses of EC 3.-; (3) in three cancer types, BRCA, ESCA and KICH, all genes fall into four sub-subclasses of EC 2.-; and in the remaining six cancer types: COAD, HNSC, LIHC, PRAD, STAD, and THCA, all genes fall into two subclasses of EC 2.- and two subclasses of EC 3.-; (3) the two most commonly used EC 3.- subclasses are EC 3.4.21.- and EC 3.1.3.- while the two mostly used EC 2.- subclasses are EC 2.4.1.- and EC 2.7.1.-. Table S1 gives the gene names selected in the regression model for each cancer type.

### Linear Regression Analyses

For each cancer type, a linear regression model is trained to predict the mutation rate using expression data in the same cancer sample, of selected genes from some EC subclasses as discussed above. The detailed objective function is described in the Methods section. To ensure the quality of each trained model, we have randomly selected 2/3 of the samples as the training data, and used the remaining 1/3 as the test data. Figure 2 shows the predicted values by the best trained model vs. the actual mutation rates across all samples for each cancer type. Similar plots for the test samples are given in Figure S1. Table 1 summarizes the level of contribution by genes of each EC subclass to the regression result for each cancer type, with the detailed information of the 14 models being given in Table S2.

### Prediction Validation

We have conducted the following analyses to provide supporting data to our prediction model.

### Validation Against Fatty Acid Synthesis Genes

It is well-known that fatty acids serve as antioxidants in cancer and other disease cells (Richard et al., 2008; Freitas et al., 2017).
Hence, we anticipate that our predictor should have some level of correlation with the fatty acid synthesis process. Table 2 shows the best correlation coefficient between our predictor and one of the four fatty acid synthesis genes (see Methods) across the 14 cancer types, with Table 3 giving the statistical significance of the observed correlations.

**Validation Against Mucin Genes**
Mucins have been found to have elevated levels across numerous cancer types and are known to have anti-oxidation roles (Takeyama et al., 2000; Wu et al., 2009; Dewald et al., 2016). We have examined their expression levels and our predictor. Table 4 shows the best correlation coefficient between our predictor and one of the mucin genes across the 14 cancer types, with Table 5 giving the statistical significance of the observed correlations. And the statistical significance of the observed correlations for each Mucin genes is given in Table S3 in detail.

**Validation Against Glutathione Synthesis Genes**
Glutathione is a molecule human cells use as a key antioxidant. We have examined their expression levels and our predictor. Table 6 shows the R2 values between our predictor and
TABLE 1 | Estimated contribution by genes in each EC subclass to the linear model for each of the 14 cancer types.

| Cancer name | EC subclasses used | Contributions by each EC class |
|-------------|--------------------|--------------------------------|
| BLCA        | 3.1.1              | 2.05E-12                       |
|             | 3.1.4              | 2.02E-26                       |
|             | 3.4.21             | 4.88E-13                       |
|             | 3.2.1              | 1.46E-21                       |
| BRCa        | 2.3.1              | 3.52E-21                       |
|             | 2.4.1              | 2.11E-24                       |
|             | 2.7.1              | 2.41E-14                       |
|             | 2.7.7              | 2.85E-18                       |
| COAD        | 2.7.1              | 1.18E-14                       |
|             | 2.4.1              | 5.97E-13                       |
|             | 3.1.3              | 2.44E-20                       |
|             | 3.4.21             | 3.02E-16                       |
| ESCA        | 2.1.1              | 4.56E-71                       |
|             | 2.3.1              | 7.06E-04                       |
|             | 2.4.1              | 5.07E-07                       |
|             | 2.6.1              | 9.62E-22                       |
| HNSC        | 2.7.1              | 2.94E-13                       |
|             | 2.4.1              | 7.11E-19                       |
|             | 3.1.3              | 5.68E-14                       |
|             | 3.4.21             | 2.51E-16                       |
| KICH        | 2.7.7              | 2.09E-05                       |
|             | 2.1.1              | 1.78E-16                       |
|             | 2.5.1              | 4.37E-05                       |
|             | 2.8.2              | 6.96E-06                       |
| KIRC        | 3.1.1              | 1.16E-10                       |
|             | 3.1.3              | 9.65E-17                       |
|             | 3.2.1              | 2.20E-10                       |
|             | 3.4.21             | 5.91E-28                       |
| KIRP        | 3.1.1              | 8.78E-14                       |
|             | 3.1.3              | 6.68E-07                       |
|             | 3.4.21             | 1.60E-23                       |
|             | 3.6.3              | 6.96E-16                       |
| LIHC        | 2.4.1              | 1.41E-21                       |
|             | 3.1.3              | 1.44E-12                       |
|             | 3.4.21             | 2.33E-12                       |
|             | 3.2.1              | 6.07E-25                       |
| LUAD        | 3.1.3              | 1.65E-63                       |
|             | 3.1.4              | 3.48E-07                       |
|             | 3.4.21             | 5.10E-18                       |
|             | 3.6.3              | 3.57E-35                       |
| LUSC        | 3.1.1              | 5.77E-43                       |
|             | 3.1.3              | 2.09E-24                       |
|             | 3.4.21             | 1.78E-26                       |
|             | 3.6.3              | 8.40E-22                       |
| PRAD        | 2.7.1              | 2.74E-41                       |
|             | 2.4.1              | 1.66E-11                       |
|             | 3.4.21             | 5.10E-14                       |
|             | 3.1.3              | 5.27E-05                       |
| STAD        | 2.7.1              | 5.47E-23                       |
|             | 2.4.1              | 1.34E-15                       |
|             | 3.1.3              | 5.06E-16                       |
|             | 3.4.21             | 8.40E-08                       |
| THCA        | 2.7.1              | 2.27E-31                       |
|             | 2.4.1              | 5.53E-12                       |
|             | 3.1.3              | 6.47E-11                       |
|             | 3.4.21             | 1.05E-06                       |

TABLE 2 | Correlation coefficients between our oxidative-stress predictor and the fatty acid synthesis gene across 14 cancer types.

| Cancer name | Correlation coefficient | Gene name |
|-------------|-------------------------|-----------|
| BLCA        | 0.77                    | FASN      |
| BRCa        | 0.64                    | ACAT2     |
| COAD        | 0.89                    | ACAT1     |
| ESCA        | 0.99                    | ACAT2     |
| HNSC        | 0.92                    | MCAT      |
| KICH        | 0.98                    | ACAT2     |
| KIRP        | 0.68                    | ACAT1     |
| LIHC        | 0.73                    | FASN      |
| LUAD        | 0.81                    | ACAT2     |
| LUSC        | 0.69                    | FASN      |
| PRAD        | 0.72                    | FASN      |
| STAD        | 0.66                    | MCAT      |
| THCA        | 0.79                    | ACAT1     |

It is noteworthy that our predictor is designed to predict the level of oxidative stress; and each of the above three groups of genes is known to be associated with the level of oxidative stress and not involved in the training dataset. The observed strong correlations between our predicted oxidative stress levels and the expression levels of each such group, along with significant p-values, provide strong support for that our predictor captures the anti-oxidative level from different independent aspects, hence indicating the validity of our trained predictor as an indicator for the level of oxidative stress.

Application
To demonstrate the utility of our predictor, we have calculated the average oxidative stress levels for each of the four stages...
TABLE 4 | Correlation coefficients between our oxidative-stress predictor and mucin genes across 14 cancer types.

| Cancer name | Correlation coefficient | Gene name |
|-------------|-------------------------|-----------|
| BLCA        | 0.69                    | MUC15     |
| BRCA        | 0.8                     | MUC20     |
| COAD        | 0.75                    | MUC5B     |
| ESCA        | 0.87                    | MUC17     |
| HNSC        | 0.86                    | MUC3A     |
| KICH        | 0.91                    | MUC12     |
| KIRC        | 0.83                    | MUC16     |
| KIRP        | 0.97                    | MUC12     |
| LIHC        | 0.72                    | MUC20     |
| LUAD        | 0.94                    | MUC22     |
| LUSC        | 0.92                    | MUC4, MUC20 |
| PRAD        | 0.76                    | MUC6      |
| STAD        | 0.88                    | MUC7      |
| THCA        | 0.88                    | MUC15     |

TABLE 5 | Statistical significance for the observed correlation coefficients in Table 4 across 14 cancer types.

| Cancer name | Statistics significance | Gene name |
|-------------|-------------------------|-----------|
| BLCA        | 2.40E-13                | MUC15     |
| BRCA        | 4.01E-11                | MUC20     |
| COAD        | 2.04E-12                | MUC5B     |
| ESCA        | 8.74E-07                | MUC17     |
| HNSC        | 8.74E-07                | MUC3A     |
| KICH        | 1.10E-04                | MUC12     |
| KIRC        | 3.95E-07                | MUC6      |
| KIRP        | 9.38E-06                | MUC12     |
| LIHC        | 2.96E-08                | MUC1      |
| LUAD        | 6.28E-07                | MUC22     |
| LUSC        | 8.48E-11                | MUC1      |
| PRAD        | 2.85E-04                | MUC6      |
| STAD        | 3.40E-07                | MUC17     |
| THCA        | 2.27E-03                | MUC15     |

TABLE 6 | Correlation coefficients between our predictor and the glutathione synthesis genes across 14 cancer types.

| Cancer name | Correlation coefficient | Gene name |
|-------------|-------------------------|-----------|
| BLCA        | 0.63                    | GCLLM     |
| BRCA        | 0.57                    | GCLLM     |
| COAD        | 0.94                    | GSS       |
| ESCA        | 0.78                    | GCLC      |
| HNSC        | 0.77                    | GCLM      |
| KICH        | 0.45                    | GCLC      |
| KIRC        | 0.74                    | GSS       |
| KIRP        | 0.75                    | GSS       |
| LIHC        | 0.66                    | GCLC      |
| LUAD        | 0.74                    | GCLM      |
| LUSC        | 0.75                    | GSS       |
| PRAD        | 0.82                    | GCLC      |
| STAD        | 0.79                    | GCLM      |
| THCA        | 0.75                    | GCLC      |

TABLE 7 | Statistical significance for observed correlation coefficient in Table 6 across 14 cancer types.

| Cancer name | GCLC | GSS | GCLM |
|-------------|------|-----|------|
| BLCA        | 3.66E-01 | 7.67E-03 | 1.12E-07 |
| BRCA        | 3.07E-01 | 4.49E-01 | 5.16E-09 |
| COAD        | 8.20E-01 | 1.53E-05 | 3.14E-02 |
| ESCA        | 3.61E-14 | 4.09E-01 | 5.25E-01 |
| HNSC        | 4.62E-01 | 3.05E-04 | 1.47E-06 |
| KICH        | 1.38E-01 | 9.81E-01 | 7.89E-01 |
| KIRC        | 5.90E-01 | 3.65E-05 | 3.67E-03 |
| KIRP        | 7.88E-01 | 3.42E-07 | 7.88E-01 |
| LIHC        | 5.50E-01 | 6.13E-06 | 9.30E-02 |
| LUAD        | 1.88E-03 | 3.17E-02 | 6.50E-02 |
| LUSC        | 4.62E-02 | 2.50E-29 | 3.15E-01 |
| PRAD        | 1.11E-01 | 7.47E-02 | 4.29E-04 |
| STAD        | 3.24E-04 | 1.74E-07 | 1.21E-08 |
| THCA        | 9.80E-04 | 1.92E-01 | 2.08E-02 |

vs. the matching controls in each of the 14 cancer types, as detailed in Figure 3. From the figure, we can see: (1) cancer tissue cells have elevated oxidative-stress levels than matching controls in all of 14 cancer types; and (2) the oxidative stress level tends to progressively increase as a cancer advances from stage 1 through stage 4, in 10 out of 14 cancer types with at most one predicted level of oxidative stress being out of order. This is clearly consistent with our general understanding about cancer progression.

We have also conducted co-expression analyses to find genes that strongly correlate with our predictor in each cancer type and analyzed the pathways enriched by such genes. Table S4 gives the up-to top 100 enriched pathways with \( p < 0.05 \) in each cancer. We note that the enriched pathways are highly consistent across different cancer types. Furthermore, we note based on extensive literature search that a majority of the enriched pathways, marked in bold letters, is: (1) involved in anti-oxidation activities, (2) induced by oxidative stress, (3) induces oxidative stress, (4) involved in reactive oxidative stress signaling, and/or (5) is altered by oxidative stress. While these data provide supporting data to our predictor, pathways not known to be oxidative stress related may provide useful candidates for elucidation of the overall oxidative-stress responding mechanisms in cancer cells.

**DISCUSSION**

Quantitative assessment of the intracellular oxidative stress will prove to be an invaluable tool for elucidation of the possible causes of various changes in cancer cells, including extensive metabolic reprogramming. However, the problem proves to be very challenging because there are large numbers of contributors to both the total oxidizing power and the anti-oxidizing capacity in human cells. Previous studies tend to focus on oxidative stress...
induced by specific molecular species such as H2O2 or lipid radicals rather than the overall oxidative stress level. To the best of our knowledge, our work is the first computational tool for estimating the oxidative stress level. The basic premise of our tool is that the genomic mutation level is strongly associated with the overall intracellular oxidative stress level. The second premise is that many enzymes have anti-oxidation capacity as reported by numerous authors (Sies, 1997; Rajput et al., 2013).

One surprising result in our regression analyses is that genes in two EC classes, EC 2.- (transferases) and EC 3.- (hydrolases), specifically four subclasses of these two, can be used to well interpret the mutation rate in each cancer sample by

![Boxplots of predicted oxidative stress levels across 4 cancer stages (when data are available) along with matching controls for each of the 14 cancer types.](Figure3)
a linear combination of their expressions with high statistical significance. This strongly suggests that all the enzymes encoded by these genes play roles in cellular anti-oxidation, which is clearly a novel discovery and warrants further investigation. Furthermore, different cancer types tend to use a distinct combination of genes from four subclasses of enzymes strongly suggest that these cancer types may encounter different types of oxidative stress, hence using different combinations of anti-oxidation enzymes. This also warrants further study regarding why genes in different enzyme classes show strong correlations with mutation rates in different cancer types, therefore to understand the detailed mechanisms of their anti-oxidative functions possibly for different types of oxidants.

As mentioned in the Introduction, oxidative stress may arise from different molecular species such reactive oxygen species, reactive nitrogen species, reactive lipid species and various free radicals. Different molecular species might be utilized to consume specific types of oxidizing molecules. By showing that three independent classes of molecules all have statistically significant correlation coefficients with our general-purpose predictor for oxidative stress, we clearly have strong support for the validity of our predictor.

The availability of this new tool make open new doors for studying impact of oxidative stress on various chronic inflamed diseases, including cancer, including (i) elucidation of all metabolic processes, particularly reprogrammed metabolisms observed in cancer and other diseases, that are statistically associated with the level of oxidative stress, hence providing a new capability for detection of the possible causes of various altered metabolisms, and (ii) systematic analyses of different classes of enzymes in terms of their anti-oxidative roles, which could provide potentially powerful and new targets for treating various chronic diseases, including cancer.

We anticipate that our predictor will prove a powerful tool useful for elucidation of causes of variety of systematic changes, including metabolic reprogramming to gain in-depth understanding of why specific metabolic pathways are reprogrammed and certain cellular functions tend to be repressed or hyper-activated in cancer.

**DATA AND METHODS**

**Gene-Expression Data and Genomic Mutation Data**

Gene expression and genomic mutation data of 14 cancer types: BLCA, BRCA, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, STAD, and THCA were downloaded from the TCGA data portal. The detailed information of these data is summarized in Table 8.

| Cancer name | Number of cancer samples | Number of control samples |
|-------------|--------------------------|--------------------------|
| BLCA        | 408                      | 20                       |
| BRCA        | 1,092                    | 114                      |
| COAD        | 456                      | 42                       |
| ESCA        | 164                      | 12                       |
| HNSC        | 501                      | 45                       |
| KICH        | 66                       | 25                       |
| KIRC        | 530                      | 73                       |
| KIRP        | 289                      | 33                       |
| LIHC        | 371                      | 51                       |
| LUAD        | 515                      | 60                       |
| LUSC        | 501                      | 50                       |
| PRAD        | 495                      | 53                       |
| STAD        | 360                      | 33                       |
| THCA        | 502                      | 59                       |

Genomic mutation data were derived from the whole-exome sequencing data. Specifically, somatic changes are identified through comparing allele frequencies in the aligned DNA sequences of each cancer and the matching normal samples, using the GDC DNA-Seq analysis pipeline (GDC DNA-Seq analysis pipeline: https://docs.gdc.cancer.gov/). MuTect2 (GDC MuTect2: https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline). MuTect2 (GDC MuTect2: https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/DNA_Seq_Variant_Calling_Pipeline/#somatic-variant-calling-workflow) is used to call each mutation using the default parameters.

Table 8 summarizes the validity of our predictor.

**Synthesis of Fatty Acids**

We have used the four fatty-acid synthesis genes: FASN (Fatty Acid Synthase), ACAT1 (Acetyl-CoA Acetyltransferase 1), ACAT2 (Acetyl-CoA Acetyltransferase 2), and MCAT (Malonyl-CoA-Acyl Carrier Protein Transacylase) to reflect the level of fatty acid synthesis, and have calculated the Pearson correlation coefficient between the expressions of one of them and our predictor.

**Mucin Genes**

Human has 20 mucin genes, namely MUC1, MUC2, MUC8, MUC12, MUC13, MUC15, MUC16, MUC17, MUC19, MUC20, MUC21, MUC22, MUC3A, MUC4, MUC5AC, MUC5B, MUC6, MUC7, and MUC11. To assess their overall correlation with our predictor, we have conducted the Pearson correlation coefficient between the expression of each mucin gene and our oxidative stress predictor.

**Glutathione Synthesis**

We have used the three biosynthesis genes: GCLC (glutamate-cysteine ligase catalytic subunit), GCLM (glutamate-cysteine ligase modifier subunit) and GSS (glutathione synthase) to reflect the level of glutathione synthesis. As above, we have calculated the Pearson correlation coefficient between the expressions of the glutathione synthesis genes and our oxidative stress predictor.

---

1. TCGA data portal: https://portal.gdc.cancer.gov/
Differential Expression Analyses
We have used edgeR in the R package (edgeR package: https://www.r-project.org/) to determine if a gene is differentially expressed in cancer vs. control samples of the same cancer type. T-test is applied to estimate the statistical significance of each gene considered as differentially expressed, using 0.05 as the cut-off.

Linear Regression Analysis
We have conducted a linear regression analysis of the observed mutation rates, Y_n, in n samples of a specified cancer type against the expressions of m selected genes across n samples, X_{n,m}, so that residual ||ε|| is as small as possible as defined below:
\[ Y = XB + \varepsilon \]

where B_m is a coefficient vector with its m values to be determined through solving this optimization problem. To avoid using too many genes in the regression analysis, we have included a penalty for penalizing using more variables than necessary.
\[ Y = XB + \varepsilon + \lambda_m \]

where \( \lambda \) is an (adjustable) constant. This problem can be solved using a least squared regression analyses in the following form:
\[ \arg\min_{B \in R^m} (Y - XB - \lambda_m)^2 \]

For each Y, we have retrieved the number of point mutations per cancer genome. Then for the n cancer genomes of each of the 14 cancer types, we have n numbers, namely the mutation number divided by the number of genes (20,000), which gives rise to the Y. For the ith row of X, we have retrieved the expression data of m selected genes in the ith sample of the same cancer type. For each regression analysis, we have used the R package to solve the minimization problem. The regression result has a p-value associated with each of the m values of B. We then remove those genes with insignificant p-values, i.e., \( p > 0.05 \), for the second round of regression analyses mentioned in the Results section.

Correlation Analyses
We have used Pearson correlation coefficient to calculate the linear correlation between two lists of numbers, with one being the combined gene-expression data and the other being mutation rates of the same samples.

Pathway Enrichment Analysis
We have conducted a pathway enrichment analysis over a given set of genes found to be strongly correlated with our predictor for oxidative stress using DAVID tool (David links: https://david.ncifcrf.gov/) against the combined database of GO/Biological Process, KEGG and Reactome pathways. A pathway is considered enriched if its adjusted \( p < 0.05 \).

Assessment of the Level of Contribution by Each Enzyme Subclass in Regression Model
For each derived regression model against selected enzymes, the following is used to estimate the level of contribution by each subclass of enzymes. The package used for linear repression construction provides a p-value for each selected gene, indicating the level of the gene's contribution to the regression result with smaller p-value representing higher level of contribution. We then use the product of the p-values of the selected genes from the same enzyme subclass as the p-value of the subclass.

SOFTWARE AVAILABILITY
Our predictor is written in R language and is available upon request.

DATA AVAILABILITY STATEMENT
BLCA, BRCA, COAD, ESCA, HNSC, KICH, KIRC, KIRP, LIHC, LUAD, LUSC, PRAD, STAD, and THCA were downloaded from the TCGA data portal (TCGA data portal: https://portal.gdc.cancer.gov/). Human enzyme information, including gene names and the relevant enzyme classes, is downloaded from BRENDA (BRENDA links: https://www.brenda-enzymes.org/).

AUTHOR CONTRIBUTIONS
LL designed the method, carried out computational analyses, analyzed computational results, and wrote the article. HC supported in data analysis. YX conceived the project, reviewed computational analyses, and revised the article.

ACKNOWLEDGMENTS
LL wants to extend her thanks to members of the Computational Systems Biology Lab (CSBL) in the University of Georgia, for their helpful discussions related to the study here.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2020.00494/full#supplementary-material

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal obesity. Am. J. Physiol. Regul. Integr. Comp. Physiol. 307, R26–R34. doi: 10.1152/ajpregu.00049.2014
Barash, H. R., Gross, E., Edrei, Y., Ella, E., Israel, A., Cohen, I., et al. (2010). Accelerated carcinogenesis following liver regeneration is associated with chronic inflammation-induced double-strand DNA

AUTHOR CONTRIBUTIONS
LL designed the method, carried out computational analyses, analyzed computational results, and wrote the article. HC supported in data analysis. YX conceived the project, reviewed computational analyses, and revised the article.

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal obesity. Am. J. Physiol. Regul. Integr. Comp. Physiol. 307, R26–R34. doi: 10.1152/ajpregu.00049.2014
Barash, H. R., Gross, E., Edrei, Y., Ella, E., Israel, A., Cohen, I., et al. (2010). Accelerated carcinogenesis following liver regeneration is associated with chronic inflammation-induced double-strand DNA

AUTHOR CONTRIBUTIONS
LL designed the method, carried out computational analyses, analyzed computational results, and wrote the article. HC supported in data analysis. YX conceived the project, reviewed computational analyses, and revised the article.

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal

REFERENCES
Alfaradhi, M. Z., Fernandez-Twinn, D. S., Martin-Gronert, M. S., Musial, B., Fowden, A., and Ozanne, S. E. (2014). Oxidative stress and altered lipid homeostasis in the programming of offspring fatty liver by maternal
breaks. *Proc Natl Acad Sci USA.* 107, 2207–2212. doi: 10.1073/pnas.0909888

Bellezza, I., Miera, A. L., and Minelli, A. (2010). Nf2 and NF-kB and their concerted modulation in cancer pathogenesis and progression. *Cancers, 2,* 490–505. doi: 10.3390/cancers2020383

Campbell, K., Vörös, D., Keller, M. A., and Ralser, M. (2016). Methionine metabolism alters oxidative stress resistance via the pentose phosphate pathway. *Antioxid. Redox Signal.* 24, 543–547. doi: 10.1089/ars.2015.6516

Cerutti, P. A., and Trump, B. F. (1993). Inflammation and oxidative stress in carcinogenesis. *Cancer Cells 3,* 1–7. doi: 10.1007/978-1-4615-3520-1_75

Chan, H. P., Lewis, C., and Thomas, P. S. (2010). Oxidative stress and exhaled breath analysis: a promising tool for detection of lung cancer. *Cancers 2,* 32–42. doi: 10.3390/cancers2010032

Chang, D., Wang, F., Zhao, Y. S., and Pan, H. Z. (2008). Evaluation of oxidative stress in colorectal cancer patients. *Biomed. Environ. Sci.* 21, 286–289. doi: 10.1658/0898-9988.060403-4

Chung, Y. I., Robert, C., Gough, S. M., Rassool, F. V., and Aplan, P. (2010). Oxidative stress leads to increased mutation frequency in a murine model of myelodysplastic syndrome. *Leuk. Res.* 38, 95–102. doi: 10.1016/j.leukres.2013.07.008

Conti, A., Guli, C., La Torre, D., Tommasi, C., Angileri F. F., and M’Hammed, A. (2014). Oxidative stress and nitric oxide involved in the antitumor activity of cannabinoids on gliomas. *Cancers 6,* 609–635. doi: 10.3390/cancers6030609

Dalle-Donne, I., Rossi, R., Giustarini, D., Milzani, A., and Colombo, R. (2003). Protein carbonyl groups as biomarkers of oxidative stress. *Clin. Chim. Acta* 329, 23–38. doi: 10.1016/S0021-9927(03)00003-2

Dewald, J. H., Colomb, F., Bobowski-Gerard, M., Groux-Degroote, S., and Tomasello, C. (2014). Oxidative stress leads to increased mutation frequency in a murine model of myelodysplastic syndrome. *Leuk. Res.* 38, 95–102. doi: 10.1016/j.leukres.2013.07.008

Dewald, J. H., Colomb, F., Bobowski-Gerard, M., Groux-Degroote, S., and Tomasello, C. (2014). Oxidative stress leads to increased mutation frequency in a murine model of myelodysplastic syndrome. *Leuk. Res.* 38, 95–102. doi: 10.1016/j.leukres.2013.07.008

Dayem, A. A., Choi, H. Y., Kim, J. H., and Cho, S. G. (2010). Role of oxidative stress in stem, cancer, and cancer stem cells. *Cancers 2,* 859–884. doi: 10.3390/cancers2020859

De Vleeschouwer, M., Benndorf, C., and Reiter, M. (2016). Protein carbonyl as a biomarker of oxidative stress in severe leptospirosis, and its usefulness in differentiating leptospirosis from other diseases. *Proc Natl Acad Sci USA.* 113, 11938–11943. doi: 10.1073/pnas.1615497113

Dewald, J. H., Colomb, F., Bobowski-Gerard, M., Groux-Degroote, S., and Tomasello, C. (2014). Oxidative stress leads to increased mutation frequency in a murine model of myelodysplastic syndrome. *Leuk. Res.* 38, 95–102. doi: 10.1016/j.leukres.2013.07.008

De Vleeschouwer, M., Benndorf, C., and Reiter, M. (2016). Protein carbonyl as a biomarker of oxidative stress in severe leptospirosis, and its usefulness in differentiating leptospirosis from other diseases. *Proc Natl Acad Sci USA.* 113, 11938–11943. doi: 10.1073/pnas.1615497113
Sosa, V., Moliné, T., Somoza, R., Paciucci, R., Kondoh, H., and Leonart, M. E.

Frontiers in Genetics | www.frontiersin.org

11
May 2020 | Volume 11 | Article 494

Quantitative Estimation of Oxidative Stress

Liu et al.

acetylcarnitines by isolated rat liver mitochondria. Biochem. J. 208, 749–757. doi: 10.1042/bj2080749

Pizzimenti, S., Toaldo, C., Pettazzoni, P., Dianzani, M. U., and Barrera, G. (2010). Effects of reactive oxygen species and the lipid peroxidation product 4-hydroxynonenal in the hallmark of cancer. Cancers 2, 338–363. doi: 10.3390/cancers2020338

Rajput, I. R., Li, Y. L., Xu, X., Huang, Y., Zhi, W. C., Yu, D. Y., et al. (2013). Supplementary effects of Saccharomyces boulardii and Bacillus subtilis B10 on digestive enzyme activities, antioxidation capacity and blood homeostasis in broiler. Int. J. Agric. Biol. 15, 1560–8530. Available online at: https://www.researchgate.net/publication/236872023_Supplementary_Effects_of_Saccharomyces_boulardii_and_Bacillus_subtilis_B10_on_Digestive_Enzyme,...Activities_Antioxidation_Capacity_and_Blood_Homeostasis_in_Broiler

Rao, C. V., Asch, A. S., and Yamada, H. Y. (2017). Frequently mutated genes/pathways as prevention targets in liver cancer. Carcinogenesis 38, 2–11. doi: 10.1093/carcin/bgw118

Reczek, C. R., and Chandel, N. S. (2017). The two faces of reactive oxygen species in cancer. Ann. Rev. Cancer Biol. 1, 79–98. doi: 10.1146/annurev-cancerbio-041916-065808

Reuter, S., Gupta, S. C., Chaturvedi, M. M., and Aggarwal, B. B. (2010). Oxidative stress, inflammation, and cancer: how are they linked? Free Radic. Biol. Med. 49, 1603–1616. doi: 10.1016/j.freeradbiomed.2010.09.006

Richard, D., Kefi, K., Barbe, U., Bausero, P., and Visoli, F. (2008). Polynsaturated fatty acids as antioxidants. Pharmacol. Res. 57, 451–453. doi: 10.1016/j.phrs.2008.05.002

Santos, C. R., and Schulz, A. (2012). Lipid metabolism in cancer. FEBS J. 279, 2610–2623. doi: 10.1111/j.1742-4658.2012.08644.x

Schultz, M. A., Abdel-Mageed, A. B., and Mondal, D. (2010). The Nrf1 and Nrf2 balance in oxidative stress regulation and androgen signaling in prostate cancer cells. Cancers 2, 1354–1378. doi: 10.3390/cancers201354

Schulz, J. B., Lindenau, J., Seyfried, J., and Dichgans, J. (2000). Glutathione, oxidative stress and neurodegeneration. Eur. J. Biochem. 267, 4904–4911. doi: 10.1046/j.1432-1327.2000.01595.x

Selvaraj, N., Bobby, Z., Das, A. K., Ramesh, R., and Koner, B. C. (2002). An evaluation of level of oxidative stress and protein glycation in non-diabetic Undiagnosed chronic renal failure patients. Clin. Chim. Acta 324, 45–50. doi: 10.1016/S0021-9926(01)00211-5

Sies, H. (1997). Oxidative stress: oxidants and antioxidants. Exp. Physiol. Transl. Integr. 82, 291–295. doi: 10.1131/epspsi.1997.sp000424

Slane, B. G., Aikin-Burns, N., Smith, B. J., Kalen, A. L., Goswami, P. C., Domann, F. E., et al. (2006). Mutation of succinate dehydrogenase subunit C results in increased O2–oxidative stress, and genomic instability. Cancer Res. 66, 7615–7620. doi: 10.1158/0008-5472.CAN-06-0833

Soory, M. Oxidative Stress Induced Mechanisms in the Progression of Periodontal Diseases and Cancer: A Common Approach to Redox Homeostasis? Cancers (2010), 2, 670–692. doi: 10.3390/cancers2020670

Sosa, V., Moliné, T., Somoza, R., Paciucci, R., Kondoh, H., and Leonart, M. E. (2013). Oxidative stress and cancer: an overview. Ageing Res. Rev. 12, 376–390. doi: 10.1016/j.arr.2012.10.004

Sun, W. G., Weydert, C. J., Zhang, Y., Yu, L., Liu, J., Spitz, D. R., et al. (2010). Superoxide enhances the antitumor combination of AdMnSOD plus BCU1 in breast cancer. Cancers 2, 68–87. doi: 10.3390/cancers2010068

Takeyama, K., Dabbagh, K., Jeong Shim, J., Dao-Pick, T., Ueki, I. F., and Nadel, J. A. (2000). Oxidative stress causes mucin synthesis via transactivation of epidermal growth factor receptor: role of neutrophils. J. Immunol. 164, 1546–1552. doi: 10.4049/jimmunol.164.3.1546

Tang, K., Zhao, Y., Li, H., Zhu, M., Li, W., Liu, W., et al. (2017). Translocase of inner membrane 50 functions as a novel protective regulator of pathological cardiac hypertrophy. J. Am. Heart Assoc. 6:e004346. doi: 10.1161/JAHA.116.004346

Teppner, M., Böss, F., Ernst, B., and Pähler, A. (2015). Application of lipid peroxidation products as biomarkers for flutamid-induced oxidative stress in vitro. Toxicol. Lett. 238, 53–59. doi: 10.1016/j.toxlet.2015.08.005

Tuibs, A., and Nussenzwieg, A. (2017). Endogenous DNA damage as a source of genomic instability in cancer. Cell 168, 644–656. doi: 10.1016/j.cell.2017.01.002

Valle, A., Oliver, J., and Roca, P. (2010). Role of uncoupling proteins in cancer. Cancers 2, 567–591. doi: 10.3390/cancers2020567

Wangxuichit, M., Wu, C., You, M., Maher, J. C., Dinh, V., Feun, L. G., et al. (2009). N1′, N3′-dimethyl-N1′, N3′-bis (phenylcarboxyothio) propanedihydroxideazalone (Elesclomol) selectively kills cisplatin resistant lung cancer cells through reactive oxygen species (ROS). Cancers 1, 23–38. doi: 10.3390/cancers1010023

Wieczorek, E., Jablonowski, Z., Tomasik, B., Gromadzinska, J., Jablonska, E., Konecki, T., et al. (2017). Different gene expression and activity pattern of antioxidant enzymes in bladder cancer. Anticancer Res. 37, 841–848. doi: 10.21873/anticancer.11387

Wu, Y. M., Nowack, D. D., Omenn, G. S., and Haab, B. B. (2009). Mucin glycoylation is altered by pro-inflammatory signaling in pancreatic-cancer cells. J. Proteome Res. 8, 1876–1886. doi: 10.1021/pr8008379

Xu, B., Wang, W., Guo, H., Sun, Z., Wei, Z., Zhang, X., et al. (2014). Oxidative stress preferentially induces a subtype of micrornuclei and mediates the genomic instability caused by p53 dysfunction. Mutat. Res. 770, 1–8. doi: 10.1016/j.mrfmmm.2014.08.004

Zhang, F., and Du, G. (2012). Dysregulated lipid metabolism in cancer. World J. Biol. Chem. 3, 167–174. doi: 10.4317/wjbc.v3i8.167

Zhang, H., Lei, Y., Yuan, P., Li, L., Luo, C., Gao, R., et al. (2014). ROS-mediated autophagy induced by dysregulation of lipid metabolism plays a protective role in colorectal cancer cells treated with gabmiccic acid. PLoS ONE 9:e96418. doi: 10.1371/journal.pone.0096418

Zhang, W., Trachootham, D., Liu, J., Chen, G., Pelicano, H., Garcia-Prieto, C., et al. (2012). Stromal control of cystine metabolism promotes cancer cell survival in chronic lymphocytic leukaemia. Nat. Cell Biol. 14, 276–286. doi: 10.1038/ncb2432

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Liu, Cui and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.