



Pan-Cancer Analysis Reveals the Functional Importance of Protein Lysine Modification in Cancer Development

Li Chen^{1†}, Yanyan Miao^{1†}, Mengni Liu^{1†}, Yanru Zeng¹, Zijun Gao¹, Di Peng¹, Bosu Hu¹, Xu Li², Yueyuan Zheng¹, Yu Xue³, Zhixiang Zuo¹, Yubin Xie^{1*} and Jian Ren^{1*}

¹ State Key Laboratory of Oncology in South China, Cancer Center, Collaborative Innovation Center for Cancer Medicine, School of Life Sciences, Sun Yat-sen University, Guangzhou, China, ² Spine Center, Department of Orthopaedics, Anhui Provincial Hospital, The First Affiliated Hospital of USTC, Hefei, China, ³ Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, China

OPEN ACCESS

Edited by:

Amit Kumar Yadav,
Translational Health Science and
Technology Institute, India

Reviewed by:

Sudipto Saha,
Bose Institute, India
Marco Vanoni,
Università degli Studi di Milano
Bicocca, Italy

*Correspondence:

Yubin Xie
xieyb6@mail.sysu.edu.cn
Jian Ren
renjian.sysu@gmail.com

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted to
Systems Biology,
a section of the journal
Frontiers in Genetics

Received: 30 March 2018

Accepted: 25 June 2018

Published: 17 July 2018

Citation:

Chen L, Miao Y, Liu M, Zeng Y, Gao Z,
Peng D, Hu B, Li X, Zheng Y, Xue Y,
Zuo Z, Xie Y and Ren J (2018)
Pan-Cancer Analysis Reveals the
Functional Importance of Protein
Lysine Modification in Cancer
Development. *Front. Genet.* 9:254.
doi: 10.3389/fgene.2018.00254

Large-scale tumor genome sequencing projects have revealed a complex landscape of genomic mutations in multiple cancer types. A major goal of these projects is to characterize somatic mutations and discover cancer drivers, thereby providing important clues to uncover diagnostic or therapeutic targets for clinical treatment. However, distinguishing only a few somatic mutations from the majority of passenger mutations is still a major challenge facing the biological community. Fortunately, combining other functional features with mutations to predict cancer driver genes is an effective approach to solve the above problem. Protein lysine modifications are an important functional feature that regulates the development of cancer. Therefore, in this work, we have systematically analyzed somatic mutations on seven protein lysine modifications and identified several important drivers that are responsible for tumorigenesis. From published literature, we first collected more than 100,000 lysine modification sites for analysis. Another 1 million non-synonymous single nucleotide variants (SNVs) were then downloaded from TCGA and mapped to our collected lysine modification sites. To identify driver proteins that significantly altered lysine modifications, we further developed a hierarchical Bayesian model and applied the Markov Chain Monte Carlo (MCMC) method for testing. Strikingly, the coding sequences of 473 proteins were found to carry a higher mutation rate in lysine modification sites compared to other background regions. Hypergeometric tests also revealed that these gene products were enriched in known cancer drivers. Functional analysis suggested that mutations within the lysine modification regions possessed higher evolutionary conservation and deleteriousness. Furthermore, pathway enrichment showed that mutations on lysine modification sites mainly affected cancer related processes, such as cell cycle and RNA transport. Moreover, clinical studies also suggested that the driver proteins were significantly associated with patient survival, implying an opportunity to use lysine modifications as molecular markers in cancer diagnosis or treatment. By searching within protein-protein interaction networks using a random walk with restart (RWR) algorithm, we further

identified a series of potential treatment agents and therapeutic targets for cancer related to lysine modifications. Collectively, this study reveals the functional importance of lysine modifications in cancer development and may benefit the discovery of novel mechanisms for cancer treatment.

Keywords: lysine modifications, cancer, somatic mutations, clinical analysis, pathway and network analysis

INTRODUCTION

Somatic mutations have a crucial role in the regulation of cancer progression, and therefore, interpreting the functional consequences of somatic mutations on gene products will be essential for developing potential targets for cancer therapies. As a benefit of the recent advances in next-generation sequencing technology and reduced analysis costs, the amount of data regarding somatic mutations in various cancer types has increased enormously in the past few years. A complex landscape of somatic mutations in cancers of multiple types and tissues has been revealed in large-scale cancer genomic datasets, such as TCGA and The International Cancer Genome Consortium (ICGC). However, among these massive amounts of mutations, not all of them are real drivers for cancer; instead, the majority of the mutations do not have a noticeable effect. Therefore, distinguishing only a few driver mutations from the majority of passenger mutations present in a cohort of patients is still a key challenge in the analysis of cancer genomes.

According to previous studies (Diaz-Cano, 2012; Morris et al., 2016), there is a significant genetic heterogeneity within the driver mutations presented in various cancer types. One possible explanation for this phenomenon is that the behavior of a cancer cell depends not only on genetic mutations but also on the dynamic regulation of non-genetic information. Therefore, combining mutations with other non-genetic regulations is an effective approach for predicting novel cancer driver genes and may provide extra guidance for cancer studies compared to traditional frequency-based methods (Vandin et al., 2012; Chen et al., 2013; Gonzalez-Perez et al., 2013; Leiserson et al., 2013; Tamborero et al., 2013; Cheng et al., 2014; Zhang et al., 2014). Protein post-translational modifications (PTMs), which are known to play critical roles in cancer development (Bode and Dong, 2004; Krueger and Srivastava, 2006; Jin and Zangar, 2009; Silvera et al., 2010), are an important functional feature that can be used in the prediction of novel cancer drivers. Among various protein amino acid residues, lysine has comparatively extensive and important modifications, such as acetylation, methylation and ubiquitination (Freiman and Tjian, 2003). The modification of lysine by ubiquitin and SUMO on specific proteins can reshape the binding interface between the modified protein and other biological macromolecules, regulating their affinity for DNA, proteins, plasma membranes or other endomembrane systems (Bergink and Jentsch, 2009; Dikic et al., 2009). Therefore, mutations of these modification sites may cause malfunction in PTM process, altering the subcellular location or activity of the modified protein and leading to abnormal functionalities (Ea et al., 2006). Recent research has reported that aberrant levels

of histone acetylation can promote oncogenic transformation and tumorigenesis by deregulating chromatin-based processes (Lee et al., 2014; Di Martile et al., 2016). As research moves forward, growing evidence has shown that the acetylation process on non-histone proteins (Glozak et al., 2005; Singh et al., 2010), especially transcription factors, is highly related to cancer phenotype. In addition to lysine acetylation, SUMOylation, and ubiquitination were also found to be involved in cancer progression. For example, the mutation of E318K on melanoma-lineage-specific microphthalmia-associated transcription factor (MITF) can disrupt a SUMO consensus site, and lack of SUMOylation increased the transcriptional activity of MITF, thereby increasing the levels of other tumor promoting factors, such as HIF1 α (Bertolotto et al., 2011; Yokoyama et al., 2011). Thus, aberrant SUMOylation of MITF promotes tumor initiation and progression. In addition to MITF, another key cancer driver that is regulated by lysine modification is the androgen receptor (AR). According to previous evidence (Heinlein and Chang, 2004; Balk and Knudsen, 2008; Tan et al., 2015), AR is the main driver of prostate cancer development and progression. Recent research has revealed that the ubiquitination of AR at position K311 is critical for its proper function, regulating both AR protein stability and AR transcriptional activity. When such an ubiquitination site loses its function, the expression of over a thousand downstream genes will be altered, possibly leading to misregulation in chromatin organization, cellular adhesion, motility, and signal transduction (McClurg et al., 2016). In this regard, the annotation of known cancer mutations based on the effects on lysine modification and the discovery of novel lysine modification-related drivers may be important for providing potential guidance in the development of new therapeutic strategies and drugs for cancer patients.

To uncover the potential mechanism of lysine modification in cancer development, here, we collected 1,085,623 somatic mutations and 103,248 lysine modification sites from existing databases and published literatures. Combining the above data together, we identified 164,884 lysine modification-related mutations. To further predict driver proteins that carried recurrent mutations on lysine modification sites, we developed a hierarchical Bayesian model and applied a Markov Chain Monte Carlo (MCMC) method for analysis. Furthermore, based on the identified driver proteins, we performed comprehensive downstream analysis to reveal their regulatory roles in biological pathways and molecular interaction networks. In addition, their potential clinical utilities in cancer diagnosis and treatment were also evaluated in this study. We expect that the above information will help in the discovery of novel mutagenic mechanisms and therapeutic targets for cancer studies.

MATERIALS AND METHODS

Collection of Lysine Modification Sites and Non-synonymous Mutations

Experimentally identified lysine modification sites in human proteins and their exact sequences were manually collected from published literatures in PubMed by searching for keywords such as “ubiquitination,” “acetylation,” “sumoylation,” “methylation,” “succinylation,” “malonylation,” “glutarylation,” “glycation,” “formylation,” “hydroxylation,” “butyrylation,” “propionylation,” “crotonylation,” “pupylation,” “neddylation,” “2-hydroxyisobutyrylation,” “phosphoglycerylation,” “carboxylation,” “lipoylation,” and “biotinylation.” To ensure an adequate amount of data, only the modifications with more than 1,000 sites were retained for subsequent analysis. Ultimately, seven types of lysine modifications (ubiquitination, acetylation, SUMOylation, glycation, malonylation, methylation, and succinylation) were collected as the final data set. All modified proteins were annotated with Ensembl transcripts and HGNC symbols using the UniProt database.

Somatic mutations of 12 cancer types (BLCA, UCEC, LAML, STAD, SKCM, GBM, THCA, LIHC, HNSC, COAD, LUSC, and THYM) were downloaded from the TCGA data portal (<https://portal.gdc.cancer.gov/>) on 16 March 2017. To obtain a complete set of mutation data, we also downloaded somatic mutations of these cancer types from the ICGC data portal (ICGC, <https://dcc.icgc.org/>, downloaded on 21 November 2017) and the Catalog of Somatic Mutations in Cancer (COSMIC Forbes et al., 2017, downloaded on 21 November 2017). The original mutation sites were then combined from the above three databases, and redundant mutations in the same patients in the same cancer type were removed. After annotation by ANNOVAR (Wang et al., 2010), only non-synonymous single nucleotide variants (SNVs) were retained.

Identification of Lysine Modification-Related Mutations From Mutation Data

We applied a k-means clustering algorithm to extract the corresponding motif regions of lysine modifications (Supplementary Methods, Supplementary Table 1). We merged the motifs of the same type of lysine modification in order at integrated PTM sites with cancer mutations for each protein and then denoted them as the modification regions. Correspondingly, the remaining sequences were merged separately and denoted as background regions (Figure 2). Non-synonymous mutations were mapped to the protein sequences and divided into lysine modification-related mutations, which were located in the modification region, and modification-irrelevant mutations, which were located in the background region. We only focused on proteins with at least one lysine modification and discarded other non-modified proteins to avoid systematic biases.

Analysis of Lysine Modification-Related Driver Proteins by Hierarchical Bayesian Model

To identify proteins with significantly altered numbers of lysine modification-related mutations, we then constructed the following hierarchical Bayesian model. In our model, we first assumed that mutations on the motif regions would probably damage the lysine modification process, thereby influencing the function of their corresponding proteins via PTM-related pathways. If such mutations are highly correlated with tumor proliferation, they will probably undergo strong positive selection during the cancer development process, and therefore, unexpectedly high mutation rates will be observed in these regions. In view of this assumption, we can identify lysine modification-related driver proteins by comparing the mutation rates in both motif regions and modification-free regions. Accordingly, a null hypothesis that the mutation rate in the motif region is same as the mutation rate in the modification-free region is proposed. More formally, we describe the detailed computational process below.

First, for a given protein, let Y_1, Y_2, \dots, Y_k represent the number of somatic mutations in each position in the modification region, and $Y_{k+1}, Y_{k+2}, \dots, Y_n$ be the same count in the background region. According to this definition, the observed counts Y can be described by a Poisson distribution as shown in Equations (1) and (2), where λ_1 and λ_2 are the mutation rates of the modification region and the background region, respectively.

$$Y_i \sim \text{Poisson}(\lambda_1) \quad i = 1, 2, \dots, k \quad (1)$$

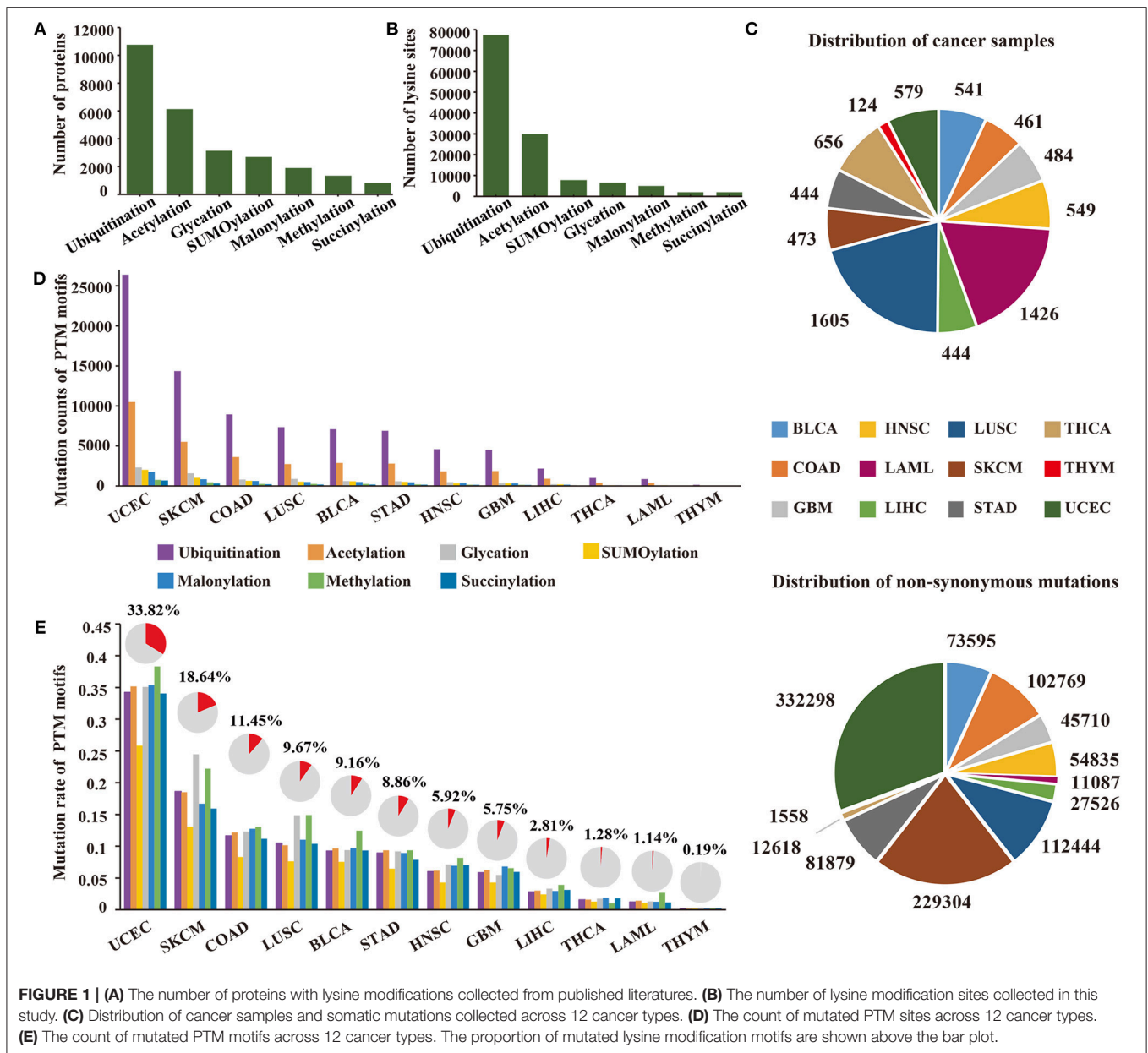
$$Y_i \sim \text{Poisson}(\lambda_2) \quad i = k + 1, k + 2, \dots, n \quad (2)$$

However, due to heterogeneity in the mutational spectrum of tumors, the mutation rate may vary markedly within different regions across different cancer types (Lawrence et al., 2013). To capture this fluctuation, a prior distribution was applied on λ_1 and λ_2 to build a double hierarchical model. As stated in the theory of probability, a gamma distribution is the conjugate prior to the Poisson distribution. Therefore, two gamma distributions with different shape parameters α and scale parameters β were used to describe the distribution of λ_1 and λ_2 in Equations (3) and (4):

$$\lambda_1 \sim \text{Gamma}(\alpha_1, \beta_1) \quad (3)$$

$$\lambda_2 \sim \text{Gamma}(\alpha_2, \beta_2) \quad (4)$$

To compare the mutation rates in the modification and background regions, we first need to compute the marginal distribution of λ_1 and λ_2 given the observed data Y in our hierarchical model, i.e., calculating $P(\lambda_1|Y)$ and $P(\lambda_2|Y)$. Therefore, a concrete form of the full joint probability should be obtained. According to Bayesian theory, the full joint probability can be written as shown in Equation (5) (see Supplementary Methods for detail deviation process).



$$\begin{aligned}
 P(\lambda_1, \lambda_2 | Y) &= P(Y_1:k | \lambda_1)P(Y_{k+1}:n | \lambda_2)P(\lambda_1)P(\lambda_2) \\
 &= \prod_{i=1}^k e^{-\lambda_1} \frac{\lambda_1^{Y_i}}{Y_i!} \prod_{i=k+1}^n e^{-\lambda_2} \frac{\lambda_2^{Y_i}}{Y_i!} \frac{\beta_1^{\alpha_1} \lambda_1^{\alpha_1-1} e^{-\beta_1 \lambda_1}}{\Gamma(\alpha_1)} \\
 &\quad \frac{\beta_2^{\alpha_2} \lambda_2^{\alpha_2-1} e^{-\beta_2 \lambda_2}}{\Gamma(\alpha_2)} \quad (5)
 \end{aligned}$$

However, computing the marginal distribution from the above full joint probability required integrating over other unrelated variables in Equation (5), which was generally a formidable analytic problem and could hardly be done manually. Rather than mathematically computing the integration, estimating the marginal distributions by the MCMC method, i.e., Gibbs

sampling, is a more straightforward approach. Therefore, in order to implement Gibbs sampling (Supplementary Figure 3), the full conditional posterior probability of every parameter should be calculated. As shown in the Supplementary Methods, the final full conditional posterior probability of λ_1 and λ_2 were obtained in

$$P(\lambda_1 | \lambda_2, Y) = \text{Gamma}\left(\sum_{i=1}^k Y_i + \alpha_1, k + \beta_1\right) \quad (6)$$

$$P(\lambda_2 | \lambda_1, Y) = \text{Gamma}\left(\sum_{i=k+1}^n Y_i + \alpha_2, n - k + \beta_2\right) \quad (7)$$

Equations (6) and (7).

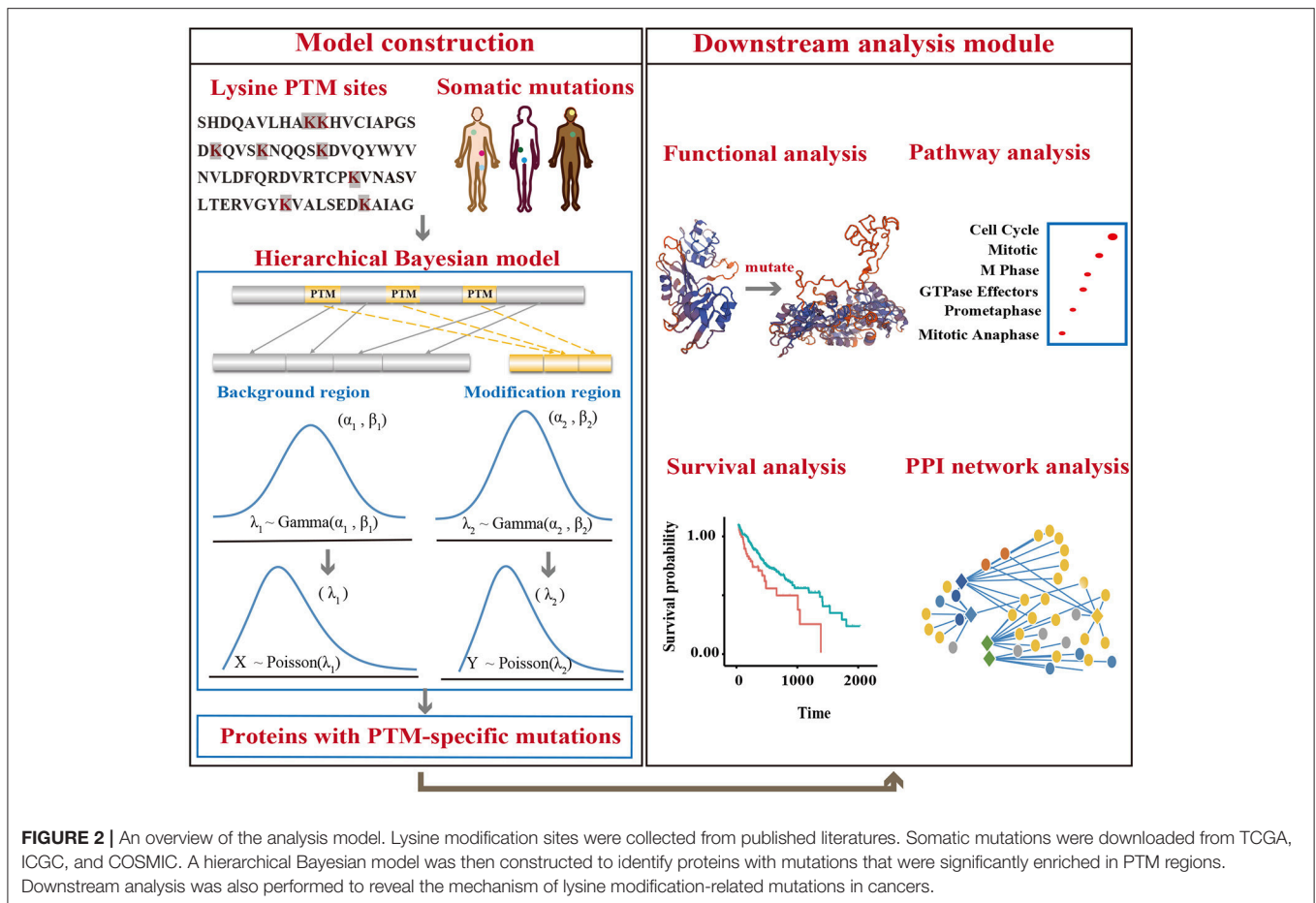


FIGURE 2 | An overview of the analysis model. Lysine modification sites were collected from published literatures. Somatic mutations were downloaded from TCGA, ICGC, and COSMIC. A hierarchical Bayesian model was then constructed to identify proteins with mutations that were significantly enriched in PTM regions. Downstream analysis was also performed to reveal the mechanism of lysine modification-related mutations in cancers.

To test the difference between the mutation rates of the background and modification regions, a variable of interest might be the relative mutation rate, which is defined as $R = \frac{\lambda_1}{\lambda_2}$. Similarly, the full conditional posterior probability can be calculated as shown in Equation (8) (Supplementary Methods).

$$P(R | \lambda_1, \lambda_2, Y) = \text{Gamma}\left(\sum_{i=1}^k Y_i + \alpha_1, \lambda_2 k + \lambda_2 \beta_1\right) \quad (8)$$

After calculating all full conditional probabilities of each variable, we can now use a Gibbs sampling algorithm to estimate the marginal distribution of these parameters. During the calculation, we performed 5,200 iterations in total and removed the first 200 iterations as a burn-in process. Finally, the marginal distribution of λ_1 , λ_2 and R was estimated by the data sampled from the last 5,000 iterations. Given the null hypothesis raised at the very beginning of this section, we can rewrite the hypothesis as shown in Equation (9).

$$\begin{aligned} H_0 : R &\leq 1 \\ H_1 : R &> 1 \end{aligned} \quad (9)$$

The p -value under the null hypothesis is then calculated from the marginal distribution of R . For each tested protein, the probability of observing a relative mutation rate < 1 can be

calculated. To control false positives, the Benjamini-Hochberg procedure is applied to each p -value. If the corrected p -value for a given protein is lower than the significance level, i.e., 0.05, we identify it as a significantly mutated protein.

Domain Association Analysis of Lysine Modification-Related Driver Proteins

The functional domains of each candidate driver protein were first predicted from InterProScan (Jones et al., 2014) using the Pfam (Finn et al., 2014) and SMART databases (Schultz et al., 1998; Letunic et al., 2009). The predicted regions of each protein were then merged together to construct a domain region, and the remaining sequences were merged as a disorder region. To examine whether the lysine modification-related mutations occurred preferentially in the domain region than in the disorder region, we designed a two-tailed bootstrap tests to compare the number of lysine modification-related mutations in the domain and disorder region. The bootstrap test was performed according to the following steps.

First, for each protein, we used Equations (10) and (11) to calculate the number of mutations that occurred per thousand amino acids in the domain region and disorder region. Specifically, we denoted the above mutation number in the

domain region and disorder region as x and y , respectively.

$$x = \frac{X}{l_1} \times 1000 \quad (10)$$

$$y = \frac{Y}{l_2} \times 1000 \quad (11)$$

where X and Y are the exact number of lysine modification-related mutations observed in the domain region and disorder region, respectively. l_1 and l_2 are the length of the domain region and disorder region, respectively.

Next, we tested the null hypothesis that x was equal to y in our observed data. To test this hypothesis, the probability of observing x not equal to y under the null distribution must be calculated. Therefore, we used the transformation in Equations (12) and (13) to estimate the null distribution. After the transformation in Equations (12) and (13), we can let the distribution of x and y be the same and constrain them to have the same center z .

$$\tilde{x}_i = x_i - \bar{x} + z \quad (12)$$

$$\tilde{y}_j = y_j - \bar{y} + z \quad (13)$$

$$z = \frac{\sum_{i=1}^m x_i + \sum_{j=1}^n y_j}{m + n} \quad (14)$$

In the above equation, x_i is the number of mutations located in the domain region for the i -th protein, whereas y_j is the number of mutations for disorder regions in the j -th protein. \bar{x} and \bar{y} are the average number of mutations located in all domain regions and disorder regions, respectively. m and n represent the total number of mutations in the domain and disorder region, respectively.

Based on the above definition, we then constructed B bootstrap data sets (x^*, y^*) by sampling x^* with replacement from \tilde{x} and y^* with replacement from \tilde{y} . The test statistic t_b^* was calculated as shown in Equation (15).

$$t_b^* = \frac{\bar{x}_b^* - \bar{y}_b^*}{\sqrt{\frac{\hat{\sigma}_{\tilde{x}_b}^{*2}}{n} + \frac{\hat{\sigma}_{\tilde{y}_b}^{*2}}{m}}} \quad (15)$$

where \bar{x}_b^* is the mean and $\hat{\sigma}_{\tilde{x}_b}^{*2}$ is the variance of the b th bootstrap sample x_b^* . The probability of observing x not equal to y under the null distribution can now be approximated by Equations (16) and (17).

$$p = \frac{\sum_{b=1}^B I(|t_b^*| \geq |t_{\text{obs}}|)}{B} \quad (16)$$

$$I(x) = \begin{cases} 1 & x = \text{True} \\ 0 & x = \text{False} \end{cases} \quad (17)$$

If the calculated p -value is lower than a pre-defined significance level, e.g., 0.05, then we should reject the null hypothesis and accept that the lysine modification mutations are more likely to be enriched in the domain region.

Conservation and Deleteriousness Analysis of Lysine Modification-Related Mutations

The sequence conservation of each mutation site was quantified using the 100 way phastCons score calculated in ANNOVAR. The phastCons score was originally designed to identify conserved elements in multiply aligned sequences. Using a phylogenetic hidden Markov model (phylo-HMM), the probability of nucleotide substitutions occur at each site in a genome was quantitatively measured (Siepel et al., 2005). And based on such probability profile (i.e., phastCons score profile), one can calculate the conservation degree of a given mutation site. In this study, we used the phastCons scores to quantify the conservation degree of all the lysine modification-related mutations and other non-lysine modification mutations. Their cumulative distribution functions (CDF) were also plotted to present the differences.

To investigate that whether our identified lysine modification-related mutations was more probable to damage specific protein functions, we next introduced a deleterious scores in our study for measuring their deleteriousness. We defined that if a given SNV was found to disrupt the functional domains or regulation regions in a specific protein, such mutation would be deleterious to protein functions. Therefore, five pieces of software, including SIFT (Ng and Henikoff, 2001), PolyPhen2 HVAR, PolyPhen2 HDIV (Adzhubei et al., 2010), LRT (Chun and Fay, 2009), and FATHMM (Shihab et al., 2013), were adopted to predict the functional consequences of our identified lysine modification-related mutation sites. To ensure prediction accuracy, we further defined a deleterious score by integrating the prediction results from the above five software. Specifically, the deleterious score was calculated by counting the number of the above methods that considered a mutation to be deleterious. A deleterious score of 0 means that the mutation is predicted to be tolerated in all methods, whereas a deleterious score of 5 means that the corresponding mutation is predicted to be deleterious in all five predictors. As a result, the deleterious score may range from 0 to 5, and a higher score indicates a higher probability of deleterious. Next, a two-tailed proportion test was then applied to compare the deleterious difference between lysine modification-related mutations and other mutations.

Subcellular Location Analysis

To annotate the subcellular location of our identified driver proteins, we first downloaded the data set from Thul's paper (Thul and Akesson, 2017). The identified driver proteins were then mapped to their corresponding subcellular locations according to the downloaded data set. Specifically, we categorized our driver proteins into 13 basic cellular compartments, which were the cytosol, mitochondria, microtubules, actin filaments, intermediate filaments, centrosome, nucleus, nucleoli, vesicles, plasma membrane, Golgi apparatus, ER, and secreted. The final annotation was summarized in a Venn diagram.

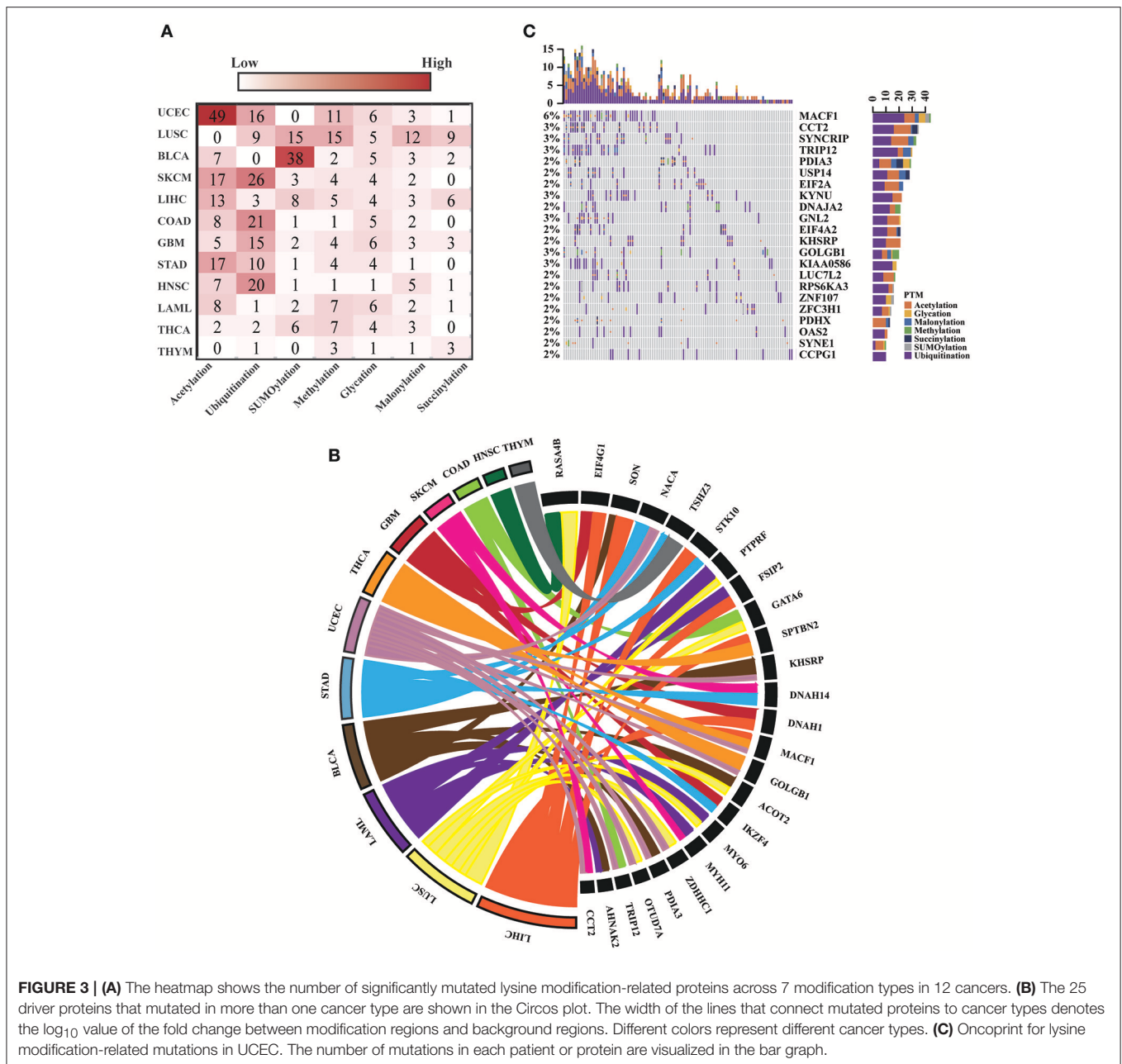


FIGURE 3 | (A) The heatmap shows the number of significantly mutated lysine modification-related proteins across 7 modification types in 12 cancers. **(B)** The 25 driver proteins that mutated in more than one cancer type are shown in the Circos plot. The width of the lines that connect mutated proteins to cancer types denotes the \log_{10} value of the fold change between modification regions and background regions. Different colors represent different cancer types. **(C)** OncoPrint for lysine modification-related mutations in UCEC. The number of mutations in each patient or protein are visualized in the bar graph.

Pathway Enrichment Analysis

To uncover the regulation roles of our identified driver proteins in cancers, we performed KEGG pathway and Gene Ontology enrichment analysis using the “clusterProfiler” (Yu et al., 2012) and “ReactomePA” (Yu and He, 2016) package in R. The analysis results were illustrated using bubble plots or Cytoscape (Demchak et al., 2014).

Survival Analysis

We downloaded survival data from the TCGA data portal (<https://tcga-data.nci.nih.gov/docs/publications/tcga/>) and employed the R package “survival” (<https://CRAN.R-project.org/package=survival>) to obtain the distribution of overall survival

time using Kaplan-Meier estimation. A log-rank test was used to compare the survival distributions of two groups: patients with mutations exactly located in PTM modification regions and patients with other mutations. Survival curves were plotted by the R package “survminer” (<http://www.sthda.com/english/rpkgs/survminer>).

RESULTS

Global Analysis Reveals Recurrent Cancer Mutations in Lysine Modification Sites

To investigate specific cancer mutations in lysine modifications, we collected 103,248 experimentally identified lysine

modification sites in 13,378 proteins in total from published literatures (Figure 1A). The collected modification sites consisted of 77,364 ubiquitination sites, 29,942 acetylation sites, 7,821 SUMOylation sites, 6,568 glycation sites, 5,013 malonylation sites, 2,018 methylation sites, and 2,014 succinylation sites (Figure 1B, Supplementary Table 2). Considering the fact that modifications of lysine residues were mainly catalyzed by specific enzymes and that each enzyme has a quite different recognition motif, we first applied a modified k-means clustering algorithm to divide the modification sites into different consensus groups (Supplementary Figure 1). To determine the optimal recognition motif for each consensus group, we then carried out a PSSM-based method on the grouped data and visualized the amino acid preference with the Seq2Logo software (Thomsen and Nielsen, 2012). According to the calculated amino acid profiles (Supplementary Figure 2), we empirically selected the optimal length of the recognition motif and constructed the motif region for each consensus group.

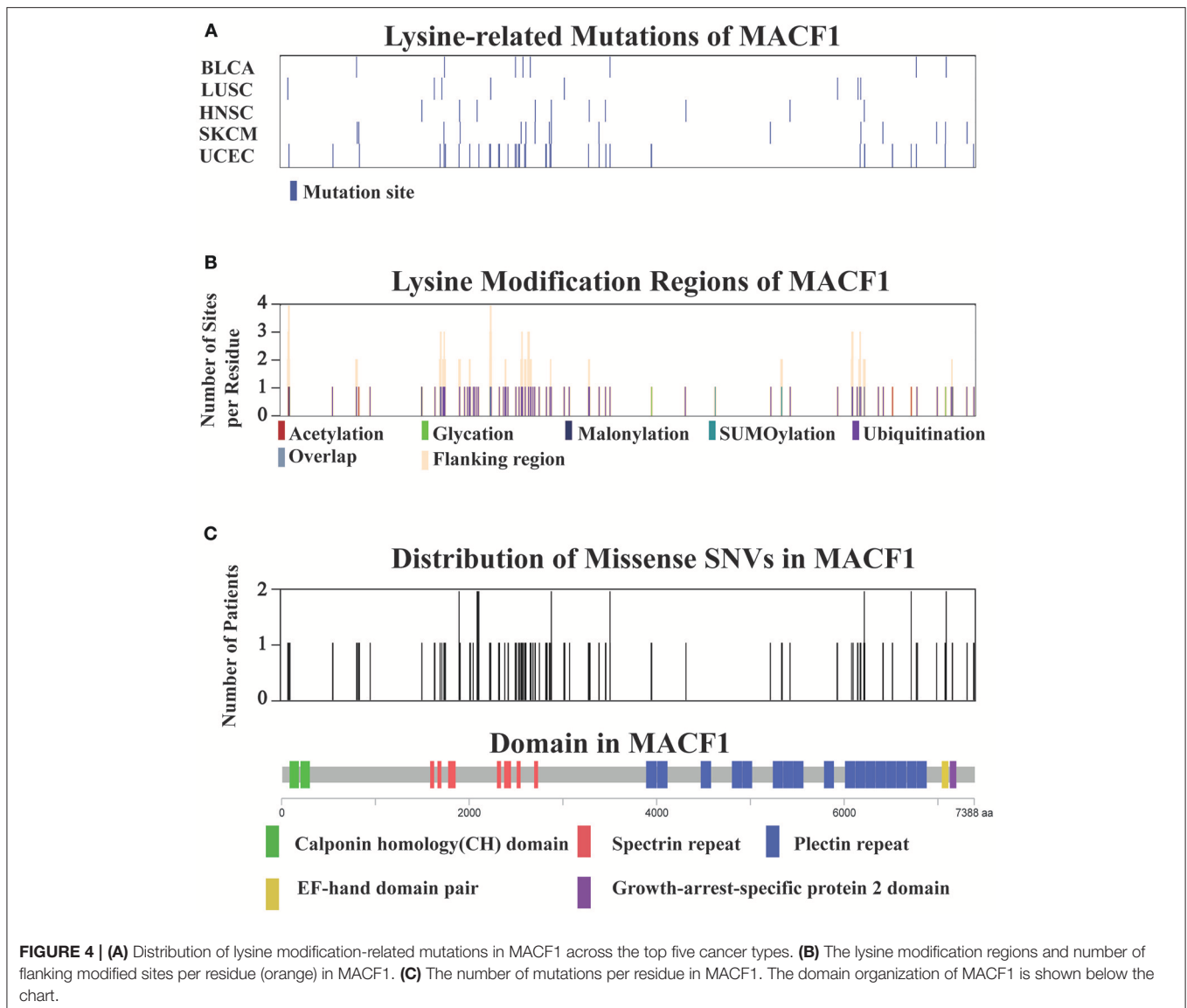
Recent publications have revealed that amino acid mutations within the modification motif can disrupt the interaction between modification enzymes and specific amino acid residues, thereby altering the level of post-translational modification on specific proteins (Taverna et al., 2007; Reimand and Bader, 2013; Hornbeck et al., 2015). Therefore, a total of 1,085,623 non-synonymous mutations (Figure 1C) from 7,786 patients (Figure 1C) were collected from TCGA for subsequent analysis. By mapping the non-synonymous mutations to the motif regions, we finally obtained 164,884 lysine modification-related mutations from 12 selected cancers in 7 modification types (ubiquitination, acetylation, SUMOylation, glycation, malonylation, methylation, and succinylation) (Supplementary Table 3), which amounted to 68,401 damaged lysine modification sites (Figure 1D). Surprisingly, of the 12 selected cancer types, we observed that uterine corpus endometrial carcinoma carried the largest number of lysine modification-related mutations in its samples, and more than 33.8% of the modification sites were mutated in this cancer type (Figure 1E). These results demonstrated that abnormal lysine modification is a general mechanism of cancer cell regulation, implying its functional importance in different cancer types.

Driver Proteins With Significant Lysine Modification-Related Mutations

To identify driver proteins carrying significant lysine modification-related mutations in multiple cancer types, we developed a hierarchical Bayesian model and applied the MCMC method to estimate the mutation frequency in modification regions (Figure 2, Supplementary Methods). We assumed that for a given protein, if the mutation frequency observed in the motif region was higher than that the non-motif region, the modification process in this protein may undergo obvious positive selection and the corresponding mutations may have a significant effect on protein function. Therefore, identifying proteins that carried a higher mutation rate in the lysine modification site could assist with finding targets

that potentially drive cancer progression. In this regard, a null hypothesis that the mutation rate in motif regions is equal to that in non-motif regions was proposed in our Bayesian model. For each tested protein, the p -value of observing a higher mutation rate in motif regions than in non-motif regions was calculated. In addition, the Benjamini-Hochberg method was then applied to control the false discover rate in the statistical test.

For all 12 selected cancer types, we applied this model to identify potential driver proteins regarding the 7 types of lysine modification. Of the 13,378 mutated proteins, a total of 473 proteins were found to have significantly higher substitution rates in lysine modification motifs than in background regions (Figure 3A). Of these 473 proteins, 45 are known cancer drivers according to the Cancer Gene Census in COSMIC database where there are 699 cancer drivers in total, highlighting that our identified proteins had significant functionality in tumorigenesis (Supplementary Table 4, p -value = 1.9×10^{-5} , Fisher's exact test). Among these driver proteins, 25 were found to be significantly mutated in more than one cancer type (Figure 3B), suggesting a general driver mechanism of lysine modification in multiple cancer types. In our tested cancer types, endometrial carcinoma had the most striking number of lysine modification-related mutations. In total, 86 proteins were identified as significantly mutated in the region of modification motifs across 7 types of lysine modifications (Figure 3A). More than 20 proteins in endometrial carcinoma had a mutation rate in lysine modification motifs that was higher than 2% (Figure 3C). Moreover, we found that in endometrial carcinoma, the most frequently mutated gene, *MACF1*, also had a high lysine modification-related mutation rate in other cancers (Figure 4A), including BLCA, LUSC, HNSC, and SKCM. According to published literature (Karakesisoglou et al., 2000), the coding product of *MACF1* can facilitate actin-microtubule interactions and couple the microtubule network to cellular junctions. Some related works indicated that *MACF1* was an important signaling molecule with various functions in cell processes, embryo development, tissue-specific functions, and human diseases (Hu et al., 2016). Since *MACF1* can act as a positive regulator in the Wnt receptor signaling pathway and function through the oncogenic MAPK signaling pathway (Chen et al., 2006), it has been selected as a novel potential target in several cancers (Miao et al., 2017). In our studies, various types of lysine modifications were mapped to *MACF1*, indicating an important function of post-translational modification in regulating the formation and interaction of cytoskeletal networks (Figure 4B). Interestingly, in our analysis, we observed a remarkable distribution of amino acid mutations around the lysine modification sites across 12 cancer types (Figure 4C). Moreover, most of the lysine modification-related mutations were found to be located in important functional domains, such as plectin repeats and growth-arrest-specific domains (Figure 4C). The above results suggested that lysine modification-related mutations in *MACF1* may interfere with its proper function and cause the appearance of cancer phenotypes.



Lysine Modification-Related Mutations Are Highly Conserved and Highly Deleterious

To explore the potential function of our identified lysine modification-related mutations, we conducted a series of proteome-wide analysis in this study. First, a bootstrap test was applied to examine the functional impact of lysine modification-related mutations on protein domains. Interestingly, among the 12 tested cancer types, lysine modification-related mutations more preferably occurred in known domain regions than in other regions (Figure 5A), suggesting that these kinds of mutations may have underlying effects on protein functions in different cancer types. Furthermore, using hypergeometric test, we also filtered out a set of protein domains that were more concentrated with lysine modification-related mutations. Of which, the Myotonic dystrophy protein kinase domain in *CDC42BPB* (Zhao and Manser, 2015) and AT

hook domain in *SETBP1* (Piazza et al., 2013; Coccaro et al., 2017) stand out as two most representative examples (Supplementary Table 5). We expected that this candidate list of mutated domains may help to discover novel mechanisms of abnormal lysine modifications on regulating protein domain functions.

In addition to domain analysis, we also examined the evolutionary conservation of these lysine modification-related mutations. Using the phastCons 100-way scheme (Siepel et al., 2005), we calculated the conservation scores in both lysine modification-related mutations and other non-synonymous mutations and found that mutations related to lysine modification were more functionally conserved than other mutations (p -value $< 2.2 \times 10^{-16}$, Kolmogorov-Smirnov test) (Figure 5B). Moreover, the deleteriousness level of both lysine modification-related mutations and

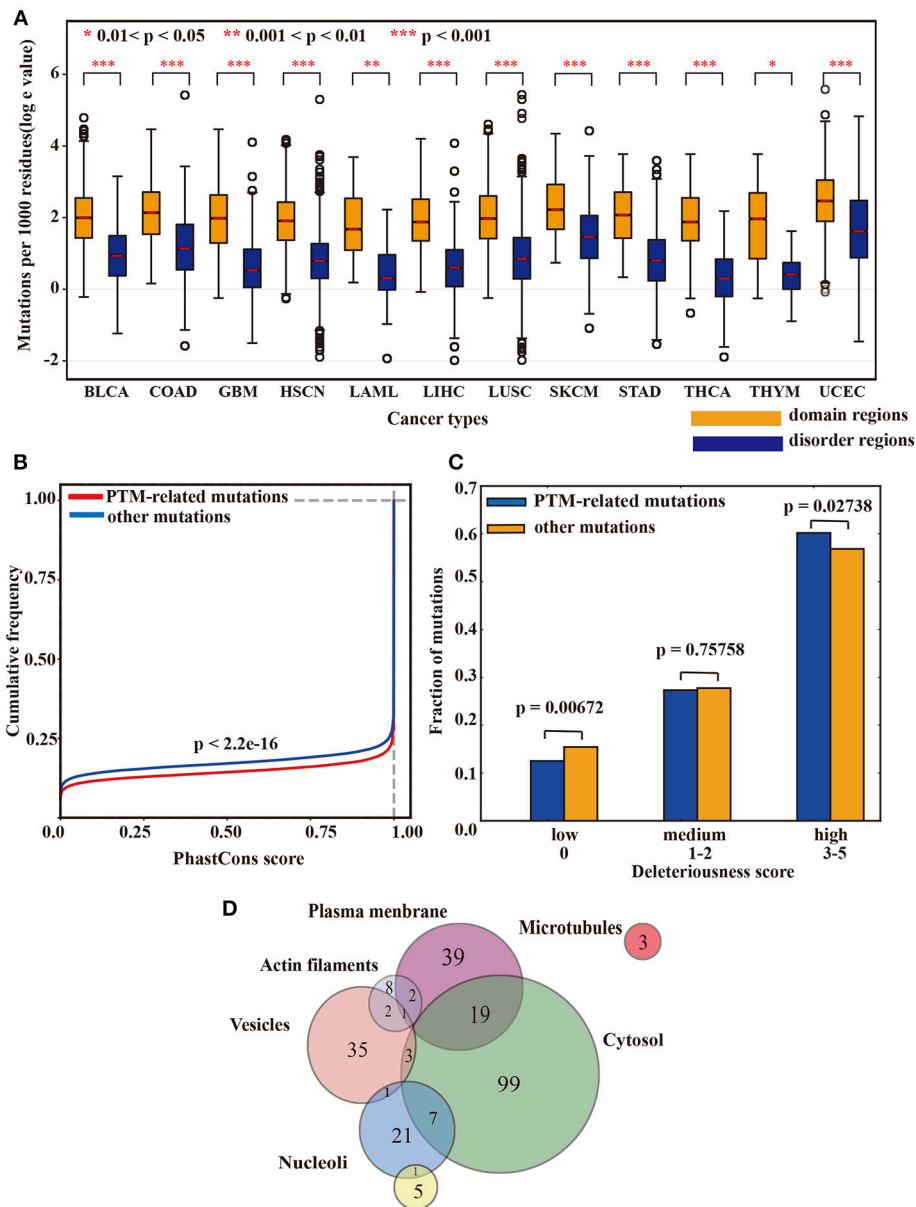


FIGURE 5 | (A) The box plot shows the differences in mutation rates in the domain regions and disorder regions. **(B)** The cumulative distribution function of the predicted conservation scores in lysine modification-related mutations and other mutations. A Kolmogorov-Smirnov test was applied to examine their statistical differences. **(C)** The deleteriousness of lysine modification-related mutations and other mutations. A two-tailed population test was applied to evaluate the differences. **(D)** The subcellular localization of the driver proteins that carried a high rate of lysine modification-related mutations.

other ordinary mutations was measured in **Figure 5C**. As expected, lysine modification-related mutations were predicted to be more deleterious than other ordinary mutations with a two-tailed population test (**Figure 5C**). Taking together the above three analyses, we can conclude that lysine modification-related mutations may have important roles in regulating many hallmark cancer pathways (Knittle et al., 2017; Chen et al., 2018; Cho et al., 2018) and may be driven by strong positive selection during the development of cancer cachexia.

In addition, to further characterize the cellular function of our identified driver proteins, an analysis of subcellular location was also carried out according to Thul's paper (Thul and Akesson, 2017). For the 473 identified driver proteins, 173 (36.57%) were localized in at least one cell compartment. Among them, 128 (27.06%) were localized in the cytosol, 61 (12.89%) were found in the plasma membrane and 42 (8.87%) were in the vesicles (**Figure 5D**). Specifically, a majority of our identified proteins (443 out of 473, 93.66%) were outside the nucleus, revealing that identified driver proteins involved

in tumorigenesis mostly are non-histone and supporting the idea that abnormal lysine modifications on non-histone proteins also played an indispensable role in cancer development (Singh et al., 2010; Carlson and Gozani, 2016). Further annotation with HistoneDB2.0 revealed that only one protein named H2B1M was histone.

Pathway Analysis Reveals Underlying Roles of Lysine Modification-Related Mutations

Based on the identified driver genes, we next performed pathway analysis to explore network signaling driven by lysine modifications in multiple cancer types. Interestingly, in KEGG annotation, we found that lysine modification-related mutations were significantly enriched in pathways such as cell cycle and RNA transport (Figure 6A). According to published literature (Senderowicz, 2002; van Kouwenhove et al., 2011; Williams and Stoeber, 2012), these pathways were known to have important regulatory roles in the proliferation and apoptosis of cancer tissue. Recently, some driver proteins in these pathway, for example the *MYC* and *EGFR* antagonists, were also being developed as therapeutic agents for prostate and colorectal cancer (Moroni et al., 2005; Vita and Henriksson, 2006; Ciardiello and Tortora, 2008; Perez et al., 2011). Similarly, GO enrichment analysis also indicated that these driver proteins are more likely to participate in cellular processes related to tumorigenesis, including those involved in cell-cell adhesion, negative regulation of transcription, cellular response to DNA damage stimulus and transcription from RNA pol II promoter. As reported in previous studies, the abnormal cell-cell adhesion can serve as one of the 10 special characteristics of cancer and reduced intercellular adhesiveness is indispensable for cancer invasion and metastasis (Hirohashi and Kanai, 2003; Farahani et al., 2014; Lin and Gregory, 2015). Besides, the negative regulation of transcription pathway has been reported to be related with proliferation and apoptosis of cancer cells (Lin and Gregory, 2015; Özeş et al., 2016). Also, the cellular response to DNA damage stimulus is important for maintaining genome stability (Siveen et al., 2014; Lin and Gregory, 2015; Özeş et al., 2016). For the enriched GO term, the transcription from RNA pol II, has been proved as a highly regulated process for tumorigenesis, and can regulate the transcript level of some known oncogenes such as *MYC* (Jonkers and Lis, 2015). A consistent result were also observed in the enrichment analysis of molecular function and cellular component aspects. Our identified driver proteins mainly located in Nucleoplasm, Nucleus Cytosol, Cytoplasm and cell adherens junction, and mainly regulated the RNA-binding or cell adhesion functions. Moreover, Reactome pathway analysis further suggested that mutation of lysine modifications on these proteins can affect cell cycle and mitotic process, especially those associated with cell apoptosis (i.e., the G1/S and G2/M checkpoints) (Bannister and Kouzarides, 2011; Fiandalo and Kyprianou, 2012; Williams and Stoeber, 2012; Figure 6B).

In summary, the above pathway analyses confirmed that proteins with significant lysine modification mutations take part in several critical regulation processes related to cancer

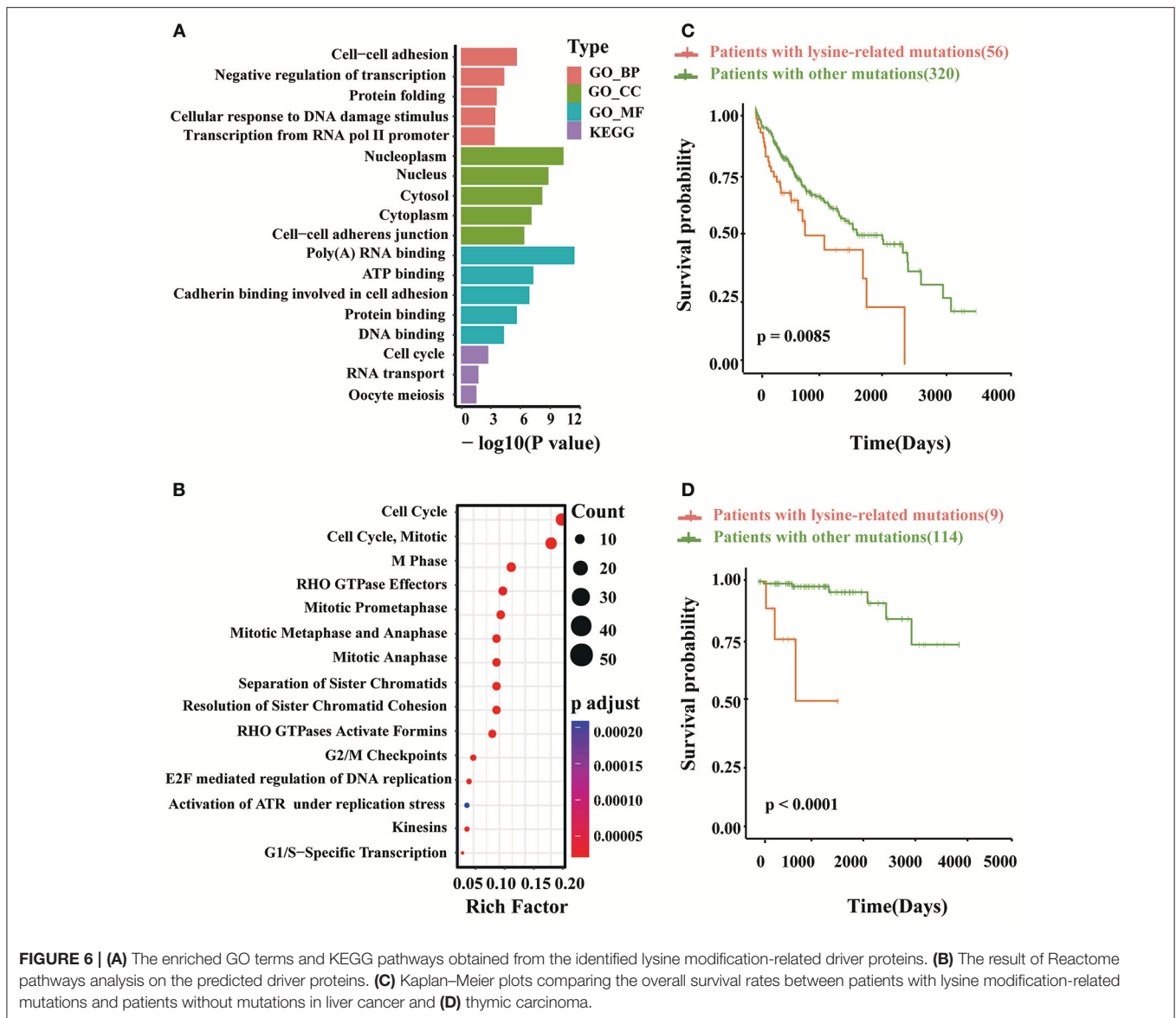
phenotypes, such as cell cycle progression and transcript regulation. Particularly, several important driver proteins were found to be dysfunctional in the above cancer-related pathways, which was mainly due to the mutation of specific lysine modification processes as reported by previous experimental studies. Polyubiquitination mutation in PCNA (Cazzalini et al., 2014) and abnormal acetylation in ATM (Bakkenist and Kastan, 2003; Sun et al., 2007) are two representative examples. These results have further confirmed the validity of our computational models.

Clinical Implications of Lysine Modification Dysregulation in Cancer Patients

Currently, the clinical implications of lysine modifications were largely unknown in multiple cancer types; therefore, here, we performed a systematic investigation to explore the clinical significance of lysine modification processes across 12 cancer types. To perform this investigation, 7,786 clinical data were first collected from TCGA patients. The identified lysine modification-related mutations were then mapped to their corresponding patients according to their recorded barcodes, and the prognosis of both the mutated group and non-mutated group were compared using Kaplan-Meier curves. Of the 12 selected cancer types, we found that patients with lysine modification-related mutations had significantly worse prognosis in liver cancer (LIHC) (Figure 6C) and thymus cancer (THYM) (Figure 6D). The significant correlation between lysine modification-related driver proteins and clinical prognosis observed in these cancer types further indicated the critical implications of lysine modifications in tumorigenesis. Given the above results, we further analyzed the implications of the identified proteins in these two cancer types. Specifically, in liver cancer, we predicted 41 driver proteins that were significantly mutated at lysine modification motifs. Eight (19.5%) of these proteins were reported previously as cancer drivers (Friedenson, 2005; Silvera et al., 2010; Zhou et al., 2011; Chang et al., 2013; Yu et al., 2013; Pandey et al., 2016; Miao et al., 2017; Tang et al., 2017). In particular, *HOXC10* was known to be associated with a decrease in the overall survival rate of liver cancer (Tang et al., 2017). Similar results were also observed in patients with thymus cancer. In total, we identified 8 lysine modification-related driver proteins, and 37.5% (3 out of 8) of them were proven by previous publication (Heyd and Lynch, 2011; Blachly and Baiocchi, 2014; Park et al., 2014; Papoudou-Bai et al., 2016). Again, the *CCAR2* in our identified list is also known to correlate with patient prognosis (Park et al., 2014; Papoudou-Bai et al., 2016). In this regard, we can conclude that our method is sensitive to finding potential gene products that have strong clinical implications in cancer patients. Our computational model may provide useful candidates for cancer diagnosis and therapies.

Network Analysis Identified Potential Downstream Targets for Our Predicted Driver Proteins

As the lysine modification-related driver proteins may function crucially in several cancer hallmark pathways, it is indispensable



to exploit their potential downstream targets through regulation networks in cancer samples. We believed that this step was critical for understanding complex biological networks in cancer patients and could help identify novel drugs and targets for cancer treatments. In this view, we applied a random walk with restart (RWR) algorithm in this section. First, we collected 199,734 pairs of experimentally validated protein-protein interactions from the STRING database (Szklarczyk et al., 2015) and 17,800 pairs of drug-target interactions from the DrugBank database (Law et al., 2013). These interactions were then combined into a heterogeneous network for subsequent RWR search. Starting from our identified driver proteins, RWR searched through the whole network and determined neighboring targets and drugs that were potentially regulated by the inputted proteins (Supplementary Methods). The search results from 7 types of lysine modifications were presented using

Cytoscape software. In total, 2,511 pairs of interactions were selected from the original heterogeneous network.

To comprehensively investigate the functional role of the selected network, we applied the Enrichment Map (Zhang et al., 2018) to cluster pathways from the RWR results. Interestingly, 14 pathways were identified as significantly enriched in our lysine modification-mediated network. A majority of these pathways were important cellular processes known to be associated with cancer misregulation (Figure 7A). For example, the VEGF signaling pathway (Hartsough et al., 2013), apoptosis (Lee et al., 2015), and T cell receptor signaling pathway (Friend et al., 2014; Hu and Sun, 2016) are three identified processes that regulate cancer cell proliferation.

Next, to understand network details, we further extracted the subnetwork of each studied modification type by reserving the top 10 most relevant downstream targets (Fouss et al., 2012;

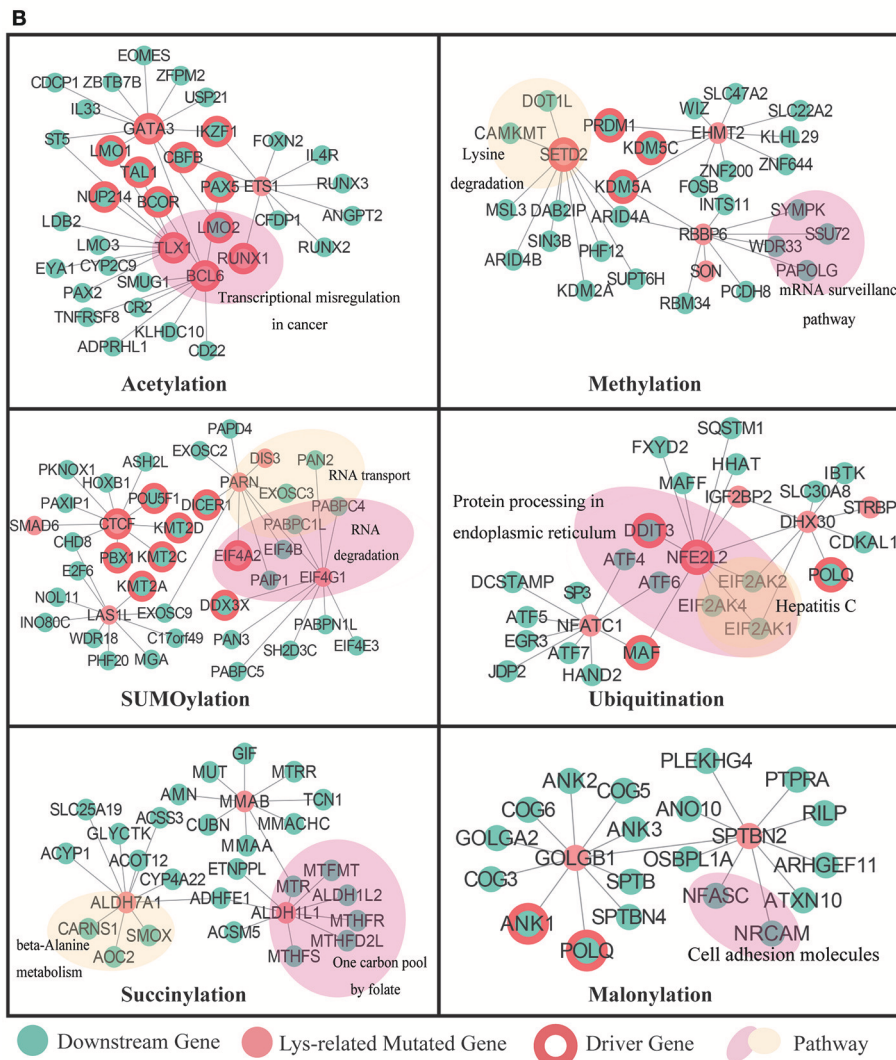
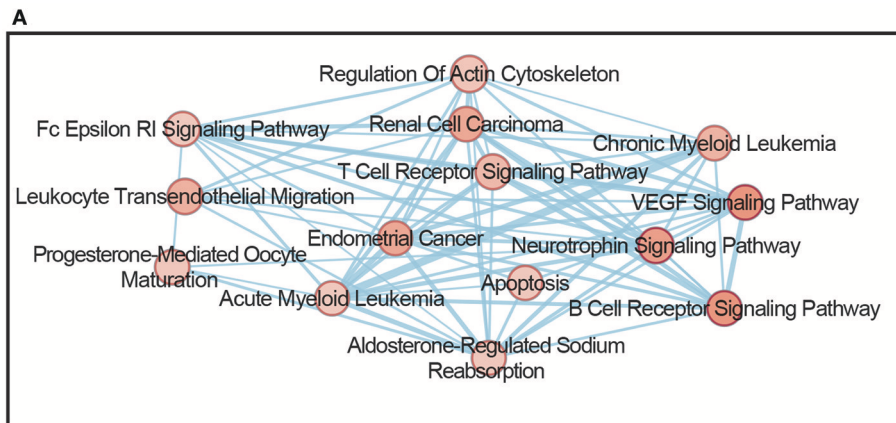


FIGURE 7 | (A) Enrichment Map showing the annotated pathways in the whole network. Nodes represent a specific pathway, and edges connect pathways with common genes. **(B)** The RWR analysis result for 7 types of lysine modifications. The identified driver proteins were taken as initial seeds in the RWR process. The predicted targets were labeled in green, the known cancer driver genes were labeled with red circles, and enriched pathways were labeled with a colored shading.

Zhu et al., 2013). Specifically, in acetylation, RWR identified 32 downstream targets that may interact with the inputted driver proteins. Among these 32 downstream targets, 12 were well-known cancer drivers. Moreover, 10 of the known cancer drivers were transcription factors related to various cancer types, which agreed with our functional enrichment analysis results that acetylation mutations may affect translational misregulation in cancer (**Figure 7B**).

Interesting results were also obtained from the analysis of methylation-mutated proteins. Methylation mutations were found to be related to the lysine degradation pathway and mRNA surveillance pathway (**Figure 7B**). According to published literature (Dimitrova et al., 2015; Shen et al., 2017), lysine demethylases mainly regulate chromatin organization to influence transcriptional processes, and cellular differentiation. Therefore, abnormalities in the lysine degradation pathway may cause serious diseases, such as cancer. In addition, the mRNA surveillance pathway was also reported to be critical in cancer development (Popp and Maquat, 2017). Under normal circumstances, the mRNA surveillance pathway can ensure the quality of transcripts and fine-tune transcript abundance in the process of cell metabolism. However, in some cases, tumors will exploit this pathway to downregulate gene expression by apparently selecting for mutations that cause the destruction of key tumor-suppressor mRNAs.

As for SUMOylation, we identified 9 downstream targets that were known to drive cancer development. Further functional analysis revealed that these downstream targets mainly functioned in pathways that controlled RNA metabolism, such as RNA transport and RNA degradation processes. It is well known that in tumor samples, the malignant phenotype is largely the consequence of dysregulated gene expression (Raza et al., 2015). Of all the molecules that affect gene expression, the dysfunction of *EIF4A2* was already known as a key mechanism that may regulate malignant transformation (Eberle et al., 1997). In addition, because of that knowledge, there is a developing focus on targeting *EIF4A* in cancer therapy. In our computational study, we found that aberrant SUMOylation on *EIF4A2* may contribute to the degradation process of RNA transcripts, representing an interesting candidate for further experimental verification.

For other lysine modifications, such as ubiquitination, succinylation, and malonylation, the RWR method also identified many potential downstream targets that may play crucial roles in cancer-related pathways. For example, in ubiquitination, protein processing in the endoplasmic reticulum was identified as related to ubiquitination mutations in cancer patients. Two critical driver genes were found to be associated with upstream ubiquitination-related mutations. Additionally, succinylation-related mutations were predicted to regulate the downstream one carbon metabolism pathway. As illustrated in published papers (Kalhan, 2013; Baggott and Tamura, 2015; Pirouzpanah et al., 2015), this cancer pathway is not only essential for the de novo synthesis of purines but also significantly related to the expression of driver genes in breast cancer patients. Furthermore, malonylation mutations were shown to influence cell adhesion molecules and components in the Golgi complex

(**Figure 7B**), which may correlate to the metastasis of cancers. In summary, our method identified a series of potential downstream proteins that were expected to correlate to lysine modification mutations. Some of these proteins were identified in previous publications, whereas others may be good candidates for follow-up experimental studies. We hope that a deeper investigation of these candidates will help illuminate novel mechanisms in cancer biology.

In addition to lysine modification-mediated downstream targets, RWR analysis also identified 13 corresponding drugs that were considered to be affected by lysine modification mutations (Supplementary Table 6). Of the 13 identified drugs, 7 were reported to be antineoplastic agents. They are azacitidine (Cortvrindt et al., 1987), acadesine (Montraveta et al., 2014), Cytidine (Periyasamy et al., 2015), mizoribine (Franchetti et al., 2005), titanium dioxide (Tyagi et al., 2016), Cytarabine (Xie et al., 2015), and Zebularine (Sabatino et al., 2013). As a remarkable therapeutic drug, Zebularine can produce an impressive therapeutic effect through the induction of apoptosis in several cancers, such as lymphoma (Montraveta et al., 2014, 2015), leukemia (Robert et al., 2009), retinoblastoma (Theodoropoulou et al., 2013), and colorectal cancer (Ly et al., 2013).

DISCUSSION

Since mutations in cancer-driven genes perform a crucial promoting effect in the process of cancer development, the prediction of such genes is of great importance in both the theoretical study of complex diseases and the clinical diagnosis of cancer patients (Kelly et al., 2010; Dawson and Kouzarides, 2012; Di Martile et al., 2016). Instead of simply identifying cancer drivers that carried exceeded number of mutations, our studies further considered the corresponding functional consequences of a given mutation in the form of lysine modifications. A hierarchical Bayesian model was therefore established to predict mutations that can alter lysine modification level. Unlike other frequency-based methods, our model did not require to accurately interpret the background mutation rate, which make it more appropriate for identifying rare mutations in the modification motif regions. By using our computational model, we have identified numbers of amino acid mutations that located in the lysine modification motifs. Based on the lysine modification-related mutations, we also identified a set of proteins that may probably drive the development of cancer. The subsequent pathway analysis revealed the functional importance of our identified driver proteins. And the survival analysis also confirmed that these proteins may have clinical implications on cancer patients. When annotating the subcellular location of these proteins, we found that vast majority of them were distributed outside nucleus. This is mainly because that most of the identified drivers were non-histone proteins, and may participate in various cellular processes across cytoplasm. For lysine modified proteins, one of the most special type was histones. In previous literatures (Strahl and Allis, 2000; Tan et al., 2011), a huge

catalog of histone modifications have been described, especially lysine modification. In our data sets, 376 lysine modification sites were collected from histones, and 349 of them were found to have lysine modification-related mutations in cancer patients. The remaining 99.8% mutation sites were located in 472 non-histone proteins. This result indicated that the abnormal lysine modification on non-histone proteins may also have critical role on regulating cancer progression. Further experimental verification on these non-histone proteins will assist the discovery of novel mechanisms for the pathogenesis of multiple cancers. Based on the above lysine modification-related mutations and cancer driver proteins, we further explore their downstream targets through a heterogeneous network using the RWR algorithm. As we expected, searching the PPI and drug-target network can help us identify potential treatment agents for cancer therapeutics. For instance, the azacitidine and acadesine. Azacitidine has been studied as antiproliferative agent in murine B16 melanoma by effecting several cellular metabolic pathways (Cortvrindt et al., 1987), including the activities of S-adenosylmethionine methyltransferase and orotidine-5'-monophosphate decarboxylase (Cihák, 1974; Christman et al., 1983). Acadesine has shown antitumoral effects in the majority of MCL cell lines and primary MCL samples via modulating immune response, actin cytoskeleton organization and metal binding (Montraveta et al., 2014). We believed that the integration of our computational resources with other downstream analysis methods or experimental studies may contribute to expand the prospects of lysine modification in cancer studies, and may also open new avenues for cancer therapeutics.

Although satisfying results were obtained in our analysis, several considerations still limit the interpretation of our discoveries. First, our current analysis only involved non-synonymous point mutations, which were the simplest to interpret, and removed other mutation types, such as frameshift variations, deletions and insertions. However, these other types of mutations can also induce carcinogenesis, so these mutations must be included in our computational model to comprehensively interpret the functional role of lysine modification processes. Second, as somatic mutations are rare in some proteins, bias will be introduced and false positives will increase in such cases. Expanding the data volumes and covering as many cancer patients as possible are efficient ways to solve this problem. In the near future, more mutation samples will be included in the analysis, and the accuracies of our computational model can be improved. Third, at

this stage, our original computational model only evaluated mutations located in modification motifs. As post-translational modification processes are mainly catalyzed by specific enzymatic systems, another valuable factor for interpreting mechanisms in cancer development will be considering upstream mutations that can alter enzyme activities (Li et al., 2010; Prabakaran et al., 2012). Therefore, constructing a complete model that not only identifies substrate mutations but also analyzes enzymatic alterations is a priority task in future studies.

In summary, the above analysis highlights a new tumorigenesis mechanism through the misregulation of lysine modifications in cancer-relevant pathways. We found that mutations at lysine modification sites significantly correlated with worse overall survival in several cancers, indicating that mutated proteins identified by our model can function as novel potential cancer drivers and can be used as diagnostic biomarkers in clinical practice. Overall, we expect that the integration of PTM data and cancer mutations by our proposed method can provide further functional evidence not available from traditional methods to the research community.

AUTHOR CONTRIBUTIONS

JR and YBX conceived, designed, and supervised all phases of the project. LC, YM, ML, YRZ, ZG, DP, BH, and YYZ performed the bioinformatics analysis. LC and YBX wrote the manuscript. XL, YX, and ZZ contributed to data interpretation, discussions, and editing of the paper. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China [31471252, 31771462, 81772614, and U1611261]; National Key Research and Development Program [2017YFA0106700]; Guangdong Natural Science Foundation [2014TQ01R387 and 2017A030313134]; Science and Technology Program of Guangzhou, China [201604020003 and 201604046001]; and China Postdoctoral Science Foundation [2017M622864 and 2018T110907].

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00254/full#supplementary-material>

REFERENCES

- Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., et al. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods* 7:248. doi: 10.1038/nmeth0410-248
- Baggott, J. E., and Tamura, T. (2015). Folate-dependent purine nucleotide biosynthesis in humans. *Adv. Nutr.* 6, 564–571. doi: 10.3945/an.115.008300
- Bakkenist, C. J., and Kastan, M. B. (2003). DNA damage activates ATM through intermolecular autophosphorylation and dimer dissociation. *Nature* 421, 499–506. doi: 10.1038/nature01368
- Balk, S. P., and Knudsen, K. E. (2008). AR, the cell cycle, and prostate cancer. *Nucl. Recept. Signal.* 6:e001. doi: 10.1621/nrs.06001
- Bannister, A. J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* 21, 381–395. doi: 10.1038/cr.2011.22
- Bergink, S., and Jentsch, S. (2009). Principles of ubiquitin and SUMO modifications in DNA repair. *Nature* 458, 461–467. doi: 10.1038/nature07963
- Bertolotto, C., Lesueur, F., Giuliano, S., Strub, T., De Lichy, M., Bille, K., et al. (2011). A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma. *Nature* 480, 94–98. doi: 10.1038/nature10539

- Blachly, J. S., and Baiocchi, R. A. (2014). Targeting PI3-kinase (PI3K), AKT and mTOR axis in lymphoma. *Br. J. Haematol.* 167, 19–32. doi: 10.1111/bjh.13065
- Bode, A. M., and Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nat. Rev. Cancer* 4, 793–805. doi: 10.1038/nrc1455
- Carlson, S. M., and Gozani, O. (2016). Nonhistone lysine methylation in the regulation of cancer pathways. *Cold Spring Harb. Perspect. Med.* 6:a026435. doi: 10.1101/cshperspect.a026435
- Cazzalini, O., Sommatis, S., Tillhon, M., Dutto, I., Bachi, A., Rapp, A., et al. (2014). CBP and p300 acetylate PCNA to link its degradation with nucleotide excision repair synthesis. *Nucleic Acids Res.* 42, 8433–8448. doi: 10.1093/nar/gku533
- Chang, X., Xu, B., Wang, L., Wang, Y., Wang, Y., and Yan, S. (2013). Investigating a pathogenic role for TXNDC5 in tumors. *Int. J. Oncol.* 43, 1871–1884. doi: 10.3892/ijo.2013.2123
- Chen, H. J., Lin, C. M., Lin, C. S., Perez-Olle, R., Leung, C. L., and Liem, R. K. (2006). The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev.* 20, 1933–1945. doi: 10.1101/gad.1411206
- Chen, H., Xu, Z., Li, X., Yang, Y., Li, B., Li, Y., et al. (2018). α -catenin SUMOylation increases IkappaBalpha stability and inhibits breast cancer progression. *Oncogenesis* 7:28. doi: 10.1038/s41389-018-0037-7
- Chen, Y., Hao, J., Jiang, W., He, T., Zhang, X., Jiang, T., et al. (2013). Identifying potential cancer driver genes by genomic data integration. *Sci. Rep.* 3:3538. doi: 10.1038/srep03538
- Cheng, W. C., Chung, I. F., Chen, C. Y., Sun, H. J., Fen, J. J., Tang, W. C., et al. (2014). DriverDB: an exome sequencing database for cancer driver gene identification. *Nucleic Acids Res.* 42, D1048–D1054. doi: 10.1093/nar/gkt1025
- Cho, Y., Kang, H. G., Kim, S.-J., Lee, S., Jee, S., Ahn, S. G., et al. (2018). Post-translational modification of OCT4 in breast cancer tumorigenesis. *Cell Death Differ.* doi: 10.1038/s41418-018-0079-6. [Epub ahead of print].
- Christman, J. K., Mendelsohn, N., Herzog, D., and Schneiderman, N. (1983). Effect of 5-azacytidine on differentiation and DNA methylation in human promyelocyte leukemia cells (HL-60). *Cancer Res.* 43, 763–769.
- Chun, S., and Fay, J. C. (2009). Identification of deleterious mutations within three human genomes. *Genome Res.* 19, 1553–1561. doi: 10.1101/gr.092619.109
- Ciardello, F., and Tortora, G. (2008). EGFR antagonists in cancer treatment. *New Engl. J. Med.* 358, 1160–1174. doi: 10.1056/NEJMra0707704
- Cihák, A. (1974). Biological effects of 5-azacytidine in eukaryotes. *Oncology* 30, 405–422. doi: 10.1159/000224981
- Coccaro, N., Tota, G., Zagaria, A., Anelli, L., Specchia, G., and Albano, F. (2017). SETBP1 dysregulation in congenital disorders and myeloid neoplasms. *Oncotarget* 8, 51920–51935. doi: 10.18632/oncotarget.17231
- Cortvrindt, R., Bernheim, J., Buysens, N., and Roobol, K. (1987). 5-Azacytidine and 5-aza-2'-deoxycytidine behave as different antineoplastic agents in B16 melanoma. *Br. J. Cancer* 56, 261–265. doi: 10.1038/bjc.1987.187
- Dawson, M. A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell* 150, 12–27. doi: 10.1016/j.cell.2012.06.013
- Demchak, B., Hull, T., Reich, M., Liefeld, T., Smoot, M., Ideker, T., et al. (2014). Cytoscape: the network visualization tool for GenomeSpace workflows. *F1000Research* 3:151. doi: 10.12688/f1000research.4492.2
- Diaz-Cano, S. J. (2012). Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. *Int. J. Mol. Sci.* 13, 1951–2011. doi: 10.3390/ijms13021951
- Dikic, I., Wakatsuki, S., and Walters, K. J. (2009). Ubiquitin-binding domains— from structures to functions. *Nat. Rev. Mol. Cell Biol.* 10, 659–671. doi: 10.1038/nrm2767
- Di Martile, M., Del Bufalo, D., and Trisciuglio, D. (2016). The multifaceted role of lysine acetylation in cancer: prognostic biomarker and therapeutic target. *Oncotarget* 7, 55789–55810. doi: 10.18632/oncotarget.10048
- Dimitrova, E., Turberfield, A. H., and Klose, R. J. (2015). Histone demethylases in chromatin biology and beyond. *EMBO Rep.* 16, 1620–1639. doi: 10.15252/embr.201541113
- Ea, C.-K., Deng, L., Xia, Z.-P., Pineda, G., and Chen, Z. J. (2006). Activation of IKK by TNF α requires site-specific ubiquitination of RIP1 and polyubiquitin binding by NEMO. *Mol. Cell* 22, 245–257. doi: 10.1016/j.molcel.2006.03.026
- Eberle, J., Krasagakis, K., and Orfanos, C. E. (1997). Translation initiation factor eIF-4A1 mRNA is consistently overexpressed in human melanoma cells *in vitro*. *Int. J. Cancer* 71, 396–401. doi: 10.1002/(SICI)1097-0215(19970502)71:3<396::AID-IJC16>3.0.CO;2-E
- Farahani, E., Patra, H. K., Jangamreddy, J. R., Rashedi, I., Kawalec, M., Rao Pariti, R. K., et al. (2014). Cell adhesion molecules and their relation to (cancer) cell stemness. *Carcinogenesis* 35, 747–759. doi: 10.1093/carcin/bgu045
- Fiandalo, M. V., and Kyprianou, N. (2012). Caspase control: protagonists of cancer cell apoptosis. *Exp. Oncol.* 34, 165–175.
- Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Forbes, S. A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783. doi: 10.1093/nar/gkw1121
- Fouss, F., Francoise, K., Yen, L., Pirotte, A., and Saerens, M. (2012). An experimental investigation of kernels on graphs for collaborative recommendation and semisupervised classification. *Neural Netw.* 31, 53–72. doi: 10.1016/j.neunet.2012.03.001
- Franchetti, P., Pasqualini, M., Cappellacci, L., Petrelli, R., Vita, P., Grifantini, M., et al. (2005). Ribose-modified mizoribine analogues: synthesis and biological evaluation. *Nucleos. Nucleot. Nucleic Acids* 24, 2023–2027. doi: 10.1080/15257770500334673
- Freiman, R. N., and Tjian, R. (2003). Regulating the regulators: lysine modifications make their mark. *Cell* 112, 11–17. doi: 10.1016/S0092-8674(02)01278-3
- Friedenson, B. (2005). BRCA1 and BRCA2 pathways and the risk of cancers other than breast or ovarian. *MedGenMed* 7, 60–60.
- Friend, S. F., Deason-Towne, F., Peterson, L. K., Berger, A. J., and Dragone, L. L. (2014). Regulation of T cell receptor complex-mediated signaling by ubiquitin and ubiquitin-like modifications. *Am. J. Clin. Exp. Immunol.* 3, 107–123.
- Glozak, M. A., Sengupta, N., Zhang, X., and Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. *Gene* 363, 15–23. doi: 10.1016/j.gene.2005.09.010
- Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Tamborero, D., Schroeder, M. P., Jene-Sanz, A., et al. (2013). IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* 10, 1081–1082. doi: 10.1038/nmeth.2642
- Hartsough, E. J., Meyer, R. D., Chitalia, V., Jiang, Y., Marquez, V. E., Zhdanova, I. V., et al. (2013). Lysine methylation promotes VEGFR-2 activation and angiogenesis. *Sci. Signal.* 6:ra104. doi: 10.1126/scisignal.2004289
- Heinlein, C. A., and Chang, C. (2004). Androgen receptor in prostate cancer. *Endocr. Rev.* 25, 276–308. doi: 10.1210/er.2002-0032
- Heyd, F., and Lynch, K. W. (2011). PSF controls expression of histone variants and cellular viability in thymocytes. *Biochem. Biophys. Res. Commun.* 414, 743–749. doi: 10.1016/j.bbrc.2011.09.149
- Hirohashi, S., and Kanai, Y. (2003). Cell adhesion system and human cancer morphogenesis. *Cancer Sci.* 94, 575–581. doi: 10.1111/j.1349-7006.2003.tb01485.x
- Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* 43, D512–D520. doi: 10.1093/nar/gku1267
- Hu, H., and Sun, S.-C. (2016). Ubiquitin signaling in immune responses. *Cell Res.* 26, 457–483. doi: 10.1038/cr.2016.40
- Hu, L., Su, P., Li, R., Yin, C., Zhang, Y., Shang, P., et al. (2016). Isoforms, structures, and functions of versatile spectraplakins MACF1. *BMB Rep.* 49, 37–44. doi: 10.5483/BMBRep.2016.49.1.185
- Jin, H., and Zangar, R. C. (2009). Protein modifications as potential biomarkers in breast cancer. *Biomark. Insights* 4, 191–200. doi: 10.4137/BMIS.2557
- Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., et al. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30, 1236–1240. doi: 10.1093/bioinformatics/btu031
- Jonkers, I., and Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* 16, 167–177. doi: 10.1038/nrm3953
- Kalhan, S. C. (2013). One carbon metabolism, fetal growth and long term consequences. *Nestle Nutr. Inst. Workshop Ser.* 74, 127–138. doi: 10.1159/000348459
- Karakesisoglou, I., Yang, Y., and Fuchs, E. (2000). An epidermal plakin that integrates actin and microtubule networks at cellular junctions. *J. Cell Biol.* 149, 195–208. doi: 10.1083/jcb.149.1.195
- Kelly, T. K., De Carvalho, D. D., and Jones, P. A. (2010). Epigenetic modifications as therapeutic targets. *Nat. Biotechnol.* 28, 1069–1078. doi: 10.1038/nbt.1678

- Knittle, A. M., Helkkula, M., Johnson, M. S., Sundvall, M., and Elenius, K. (2017). SUMOylation regulates nuclear accumulation and signaling activity of the soluble intracellular domain of the ErbB4 receptor tyrosine kinase. *J. Biol. Chem.* 292, 19890–19904. doi: 10.1074/jbc.M117.794271
- Krueger, K. E., and Srivastava, S. (2006). Posttranslational protein modifications: current implications for cancer detection, prevention, and therapeutics. *Mol. Cell. Proteomics* 5, 1799–1810. doi: 10.1074/mcp.R600009-MCP200
- Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., et al. (2013). DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42, D1091–D1097. doi: 10.1093/nar/gkt1068
- Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer genes. *Nature* 499, 214–218. doi: 10.1038/nature12213
- Lee, J. V., Carrer, A., Shah, S., Snyder, N. W., Wei, S., Venneti, S., et al. (2014). Akt-dependent metabolic reprogramming regulates tumor cell histone acetylation. *Cell Metab.* 20, 306–319. doi: 10.1016/j.cmet.2014.06.004
- Lee, S. H., Lee, E. H., Lee, S.-H., Lee, Y. M., Kim, H. D., and Kim, Y. Z. (2015). Epigenetic role of histone 3 lysine methyltransferase and demethylase in regulating apoptosis predicting the recurrence of atypical meningioma. *J. Korean Med. Sci.* 30, 1157–1166. doi: 10.3346/jkms.2015.30.8.1157
- Leiserson, M. D., Blokh, D., Sharan, R., and Raphael, B. J. (2013). Simultaneous identification of multiple driver pathways in cancer. *PLoS Comput. Biol.* 9:e1003054. doi: 10.1371/journal.pcbi.1003054
- Letunic, I., Doerks, T., and Bork, P. (2009). SMART 6: recent updates and new developments. *Nucleic Acids Res.* 37, D229–D232. doi: 10.1093/nar/gkn808
- Li, S., Iakoucheva, L. M., Mooney, S. D., and Radivojac, P. (2010). “Loss of post-translational modification sites in disease,” in *Pacific Symposium on Biocomputing (Kohala Coast)*, 337–347.
- Lin, S., and Gregory, R. I. (2015). MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer* 15, 321–333. doi: 10.1038/nrc3932
- Ly, P., Kim, S. B., Kaisani, A. A., Marian, G., Wright, W. E., and Shay, J. W. (2013). Aneuploid human colonic epithelial cells are sensitive to AICAR-induced growth inhibition through EGFR degradation. *Oncogene* 32, 3139–3146. doi: 10.1038/onc.2012.339
- McClurg, U. L., Cork, D. M. W., Darby, S., Ryan-Munden, C. A., Nakjang, S., Mendes Côrtes, L., et al. (2016). Identification of a novel K311 ubiquitination site critical for androgen receptor transcriptional activity. *Nucleic Acids Res.* 45, 1793–1804. doi: 10.1093/nar/gkw1162
- Miao, Z., Ali, A., Hu, L., Zhao, F., Yin, C., Chen, C., et al. (2017). Microtubule actin cross-linking factor 1, a novel potential target in cancer. *Cancer Sci.* 108, 1953–1958. doi: 10.1111/cas.13344
- Montraveta, A., Xargay-Torrent, S., López-Guerra, M., Rosich, L., Pérez-Galán, P., Salaverria, I., et al. (2014). Synergistic anti-tumor activity of acadesine (AICAR) in combination with the anti-CD20 monoclonal antibody rituximab in *in vivo* and *in vitro* models of mantle cell lymphoma. *Oncotarget* 5, 726–739. doi: 10.18632/oncotarget.1455
- Montraveta, A., Xargay-Torrent, S., Rosich, L., López-Guerra, M., Roldán, J., Rodríguez, V., et al. (2015). Bcl-2(high) mantle cell lymphoma cells are sensitized to acadesine with ABT-199. *Oncotarget* 6, 21159–21172. doi: 10.18632/oncotarget.4230
- Moroni, M., Veronese, S., Benvenuti, S., Marrapese, G., Sartore-Bianchi, A., Di Nicolantonio, F., et al. (2005). Gene copy number for epidermal growth factor receptor (EGFR) and clinical response to antiEGFR treatment in colorectal cancer: a cohort study. *Lancet Oncol.* 6, 279–286. doi: 10.1016/S1470-2045(05)70102-9
- Morris, L. G., Riaz, N., Desrichard, A., Senbabaoglu, Y., Hakimi, A. A., Makarov, V., et al. (2016). Pan-cancer analysis of intratumor heterogeneity as a prognostic determinant of survival. *Oncotarget* 7, 10051–10063. doi: 10.18632/oncotarget.7067
- Ng, P. C., and Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Res.* 11, 863–874. doi: 10.1101/gr.176601
- Özeş, A. R., Miller, D. F., Özeş, O. N., Fang, F., Liu, Y., Matei, D., et al. (2016). NF-κB-HOTAIR axis links DNA damage response, chemoresistance and cellular senescence in ovarian cancer. *Oncogene* 35, 5350–5361. doi: 10.1038/onc.2016.75
- Pandey, P., Sliker, B., Peters, H. L., Tuli, A., Herskovitz, J., Smits, K., et al. (2016). Amyloid precursor protein and amyloid precursor-like protein 2 in cancer. *Oncotarget* 7, 19430–19444. doi: 10.18632/oncotarget.7103
- Papoudou-Bai, A., Barbouti, A., Galani, V., Stefanaki, K., Rontogianni, D., and Kanavaros, P. (2016). Expression of cell cycle and apoptosis regulators in thymus and thymic epithelial tumors. *Clin. Exp. Med.* 16, 147–159. doi: 10.1007/s10238-015-0344-7
- Park, J. H., Lee, S. W., Yang, S. W., Yoo, H. M., Park, J. M., Seong, M. W., et al. (2014). Modification of DBC1 by SUMO2/3 is crucial for p53-mediated apoptosis in response to DNA damage. *Nat. Commun.* 5:5483. doi: 10.1038/ncomms6483
- Perez, E. A., Jenkins, R. B., Dueck, A. C., Wiktor, A. E., Bedroske, P. P., Anderson, S. K., et al. (2011). C-MYC alterations and association with patient outcome in early-stage HER2-positive breast cancer from the north central cancer treatment group N9831 adjuvant trastuzumab trial. *J. Clin. Oncol.* 29, 651–659. doi: 10.1200/JCO.2010.30.2125
- Periyasamy, M., Patel, H., Lai, C.-F., Nguyen, V. T., Nevedomskaya, E., Harrod, A., et al. (2015). APOBEC3B-mediated cytidine deamination is required for estrogen receptor action in breast cancer. *Cell Rep.* 13, 108–121. doi: 10.1016/j.celrep.2015.08.066
- Piazza, R., Valletta, S., Winkelmann, N., Redaelli, S., Spinelli, R., Pirola, A., et al. (2013). Recurrent SETBP1 mutations in atypical chronic myeloid leukemia. *Nat. Genet.* 45, 18–24. doi: 10.1038/ng.2495
- Pirouzpanah, S., Taleban, F. A., Mehdipour, P., and Atri, M. (2015). Association of folate and other one-carbon related nutrients with hypermethylation status and expression of RARβ, BRCA1, and RASSF1A genes in breast cancer patients. *J. Mol. Med.* 93, 917–934. doi: 10.1007/s00109-015-1268-0
- Popp, M. W., and Maquat, L. E. (2017). Nonsense-mediated mRNA decay and cancer. *Curr. Opin. Genet. Dev.* 48, 44–50. doi: 10.1016/j.gde.2017.10.007
- Prabakaran, S., Lippens, G., Steen, H., and Gunawardena, J. (2012). Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* 4, 565–583. doi: 10.1002/wsbm.1185
- Raza, F., Waldron, J. A., and Quesne, J. L. (2015). Translational dysregulation in cancer: eIF4A isoforms and sequence determinants of eIF4A dependence. *Biochem. Soc. Trans.* 43:1227–1233. doi: 10.1042/BST20150163
- Reimand, J., and Bader, G. D. (2013). Systematic analysis of somatic mutations in phosphorylation signaling predicts novel cancer drivers. *Mol. Syst. Biol.* 9:637. doi: 10.1038/msb.2012.68
- Robert, G., Ben Sahara, I., Puissant, A., Colosetti, P., Belhacene, N., Gounon, P., et al. (2009). Acadesine kills Chronic Myelogenous Leukemia (CML) cells through PKC-dependent induction of autophagic cell death. *PLoS ONE* 4:e7889. doi: 10.1371/journal.pone.0007889
- Sabatino, M. A., Geroni, C., Ganzinelli, M., Ceruti, R., and Broggin, M. (2013). Zebularine partially reverses GST methylation in prostate cancer cells and restores sensitivity to the DNA minor groove binder brostallicin. *Epigenetics* 8, 656–665. doi: 10.4161/epi.24916
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* 95, 5857–5864. doi: 10.1073/pnas.95.11.5857
- Senderowicz, A. M. (2002). The cell cycle as a target for cancer therapy: basic and clinical findings with the small molecule inhibitors flavopiridol and UCN-01. *Oncologist* 7(Suppl. 3), 12–19. doi: 10.1634/theoncologist.7-suppl_3-12
- Shen, H., Xu, W., and Lan, F. (2017). Histone lysine demethylases in mammalian embryonic development. *Exp. Mol. Med.* 49:e325. doi: 10.1038/emm.2017.57
- Shihab, H. A., Gough, J., Cooper, D. N., Stenson, P. D., Barker, G. L. A., Edwards, K. J., et al. (2013). Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* 34, 57–65. doi: 10.1002/humu.22225
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050. doi: 10.1101/gr.3715005
- Silvera, D., Formenti, S. C., and Schneider, R. J. (2010). Translational control in cancer. *Nat. Rev. Cancer* 10, 254–266. doi: 10.1038/nrc2824
- Singh, B. N., Zhang, G., Hwa, Y. L., Li, J., Dowdy, S. C., and Jiang, S.-W. (2010). Nonhistone protein acetylation as cancer therapy targets. *Expert Rev. Anticancer Ther.* 10, 935–954. doi: 10.1586/era.10.62
- Siveen, K. S., Nguyen, A. H., Lee, J. H., Li, F., Singh, S. S., Kumar, A. P., et al. (2014). Negative regulation of signal transducer and activator of transcription-3 signalling cascade by lupeol inhibits growth and induces

- apoptosis in hepatocellular carcinoma cells. *Br. J. Cancer* 111, 1327–1337. doi: 10.1038/bjc.2014.422
- Strahl, B. D., and Allis, C. D. (2000). The language of covalent histone modifications. *Nature* 403, 41–45. doi: 10.1038/47412
- Sun, Y., Xu, Y., Roy, K., and Price, B. D. (2007). DNA damage-induced acetylation of lysine 3016 of ATM activates ATM kinase activity. *Mol. Cell. Biol.* 27, 8502–8509. doi: 10.1128/MCB.01382-07
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandath, C., Reimand, J., et al. (2013). Comprehensive identification of mutational cancer driver genes across 12 tumor types. *Sci. Rep.* 3:2650. doi: 10.1038/srep02650
- Tan, M. H. E., Li, J., Xu, H. E., Melcher, K., and Yong, E.-I. (2015). Androgen receptor: structure, role in prostate cancer and drug discovery. *Acta Pharmacol. Sin.* 36, 3–23. doi: 10.1038/aps.2014.18
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J. S., Montellier, E., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* 146, 1016–1028. doi: 10.1016/j.cell.2011.08.008
- Tang, X. L., Ding, B. X., Hua, Y., Chen, H., Wu, T., Chen, Z. Q., et al. (2017). HOXC10 promotes the metastasis of human lung adenocarcinoma and indicates poor survival outcome. *Front. Physiol.* 8:557. doi: 10.3389/fphys.2017.00557
- Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D., and Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* 14, 1025–1040. doi: 10.1038/nsmb1338
- Theodoropoulou, S., Brodowska, K., Kayama, M., Morizane, Y., Miller, J. W., Gragoudas, E. S., et al. (2013). Aminoimidazole Carboxamide Ribonucleotide (AICAR) Inhibits the growth of retinoblastoma *in vivo* by decreasing angiogenesis and inducing apoptosis. *PLoS ONE* 8:e52852. doi: 10.1371/journal.pone.0052852
- Thomsen, M. C., and Nielsen, M. (2012). Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* 40, W281–W287. doi: 10.1093/nar/gks469
- Thul, P. J., and Akesson, L. (2017). A subcellular map of the human proteome. *Science* 356:aal3321. doi: 10.1126/science.aal3321
- Tyagi, N., Srivastava, S. K., Arora, S., Omar, Y., Ijaz, Z. M., Al-Ghadhban, A., et al. (2016). Comparative analysis of the relative potential of silver, zinc-oxide and titanium-dioxide nanoparticles against UVB-induced DNA damage for the prevention of skin carcinogenesis. *Cancer Lett.* 383, 53–61. doi: 10.1016/j.canlet.2016.09.026
- Vandin, F., Upfal, E., and Raphael, B. J. (2012). *De novo* discovery of mutated driver pathways in cancer. *Genome Res.* 22, 375–385. doi: 10.1101/gr.120477.111
- van Kouwenhove, M., Kedde, M., and Agami, R. (2011). MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer* 11, 644–656. doi: 10.1038/nrc3107
- Vita, M., and Henriksson, M. (2006). The Myc oncoprotein as a therapeutic target for human cancer. *Semin. Cancer Biol.* 16, 318–330. doi: 10.1016/j.semcancer.2006.07.015
- Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164–e164. doi: 10.1093/nar/gkq603
- Williams, G. H., and Stoeber, K. (2012). The cell cycle and cancer. *J. Pathol.* 226, 352–364. doi: 10.1002/path.3022
- Xie, C., Edwards, H., Caldwell, J. T., Wang, G., Taub, J. W., and Ge, Y. (2015). Obatoclax potentiates the cytotoxic effect of cytarabine on acute myeloid leukemia cells by enhancing DNA damage. *Mol. Oncol.* 9, 409–421. doi: 10.1016/j.molonc.2014.09.008
- Yokoyama, S., Woods, S. L., Boyle, G. M., Aoude, L. G., MacGregor, S., Zismann, V., et al. (2011). A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma. *Nature* 480, 99–103. doi: 10.1038/nature10630
- Yu, G., and He, Q.-Y. (2016). ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12, 477–479. doi: 10.1039/C5MB00663E
- Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16, 284–287. doi: 10.1089/omi.2011.0118
- Yu, Q., Zhou, C., Wang, J., Chen, L., Zheng, S., and Zhang, J. (2013). A functional insertion/deletion polymorphism in the promoter of PDCD6IP is associated with the susceptibility of hepatocellular carcinoma in a Chinese population. *DNA Cell Biol.* 32, 451–457. doi: 10.1089/dna.2013.2061
- Zhang, J., Wu, L. Y., Zhang, X. S., and Zhang, S. (2014). Discovery of co-occurring driver pathways in cancer. *BMC Bioinformatics* 15:271. doi: 10.1186/1471-2105-15-271
- Zhang, N., Bao, Y.-J., Tong, A. H.-Y., Zuyderduyn, S., Bader, G. D., Malik Peiris, J. S., et al. (2018). Whole transcriptome analysis reveals differential gene expression profile reflecting macrophage polarization in response to influenza A H5N1 virus infection. *BMC Med. Genomics* 11:20. doi: 10.1186/s12920-018-0335-0
- Zhao, Z., and Manser, E. (2015). Myotonic dystrophy kinase-related Cdc42-binding kinases (MRCK), the ROCK-like effectors of Cdc42 and Rac1. *Small GTPases* 6, 81–88. doi: 10.1080/21541248.2014.100699
- Zhou, Y.-M., Cao, L., Li, B., Zhang, R.-X., Sui, C.-J., Yin, Z.-F., et al. (2011). Clinicopathological significance of ZEB1 protein in patients with hepatocellular carcinoma. *Ann. Surg. Oncol.* 19, 1700–1706. doi: 10.1245/s10434-011-1772-6
- Zhu, J., Qin, Y., Liu, T., Wang, J., and Zheng, X. (2013). Prioritization of candidate disease genes by topological similarity between disease and protein diffusion profiles. *BMC Bioinformatics* 14(Suppl. 5):S5. doi: 10.1186/1471-2105-14-S5-S5

Conflict of Interest Statement: 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 © 2018 Chen, Miao, Liu, Zeng, Gao, Peng, Hu, Li, Zheng, Xue, Zuo, Xie and Ren. 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.