Identifying Key Somatic Copy Number Alterations Driving Dysregulation of Cancer Hallmarks in Lower-Grade Glioma

Yao Zhou†, Shuai Wang†, Haoteng Yan, Bo Pang, Xinxin Zhang, Lin Pang, Yihan Wang, Jinyuan Xu, Jing Hu*, Yujia Lan* and Yanyan Ping*

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, China

Somatic copy-number alterations (SCNAs) are major contributors to cancer development that are pervasive and highly heterogeneous in human cancers. However, the driver roles of SCNAs in cancer are insufficiently characterized. We combined network propagation and linear regression models to design an integrative strategy to identify driver SCNAs and dissect the functional roles of SCNAs by integrating profiles of copy number and gene expression in lower-grade glioma (LGG). We applied our strategy to 511 LGG patients and identified 98 driver genes that dysregulated 29 cancer hallmark signatures, forming 143 active gene-hallmark pairs. We found that these active gene-hallmark pairs could stratify LGG patients into four subtypes with significantly different survival times. The two new subtypes with similar poorest prognoses were driven by two different gene sets (one including *EGFR, CDKN2A, CDKN2B, INFA8,* and *INFA5,* and the other including *CDK4, AVIL,* and *DTX3*), respectively. The SCNAs of the two gene sets could disorder the same cancer hallmark signature in a mutually exclusive manner (including E2F_TARGETS and G2M_CHECKPOINT). Compared with previous methods, our strategy could not only capture the known cancer genes and directly dissect the functional roles of their SCNAs in LGG, but also discover the functions of new driver genes in LGG, such as *IFNA5, IFNA8,* and *DTX3*. Additionally, our method can be applied to a variety of cancer types to explore the pathogenesis of driver SCNAs and improve the treatment and diagnosis of cancer.

Keywords: somatic copy number alteration, driver genes, random walk with restart, cancer hallmark, LGG, regression analysis

INTRODUCTION

Lower-grade glioma is the most common malignant primary tumor associated with a wide range of survival times from 1 to 15 years (Ostrom et al., 2013; Tan et al., 2020; Wu et al., 2020). This tumor is made up of World Health Organization (WHO) grade II and grade III gliomas, showing slower disease progress compared with glioblastoma (GBM, grade IV)

Abbreviations: LGG, lower-grade glioma; SCNAs, somatic copy-number alterations; RWR, random walk with restart; PCGs, protein coding genes; NES, normalized enrichment score; GSEA, gene set enrichment analysis; ssGSEA, single-sample gene set enrichment analysis.

(Ceccarelli et al., 2016). Until recently, comprehensive treatments of lower-grade glioma (LGG) have mainly focused on neurosurgical resection, radiotherapy, and chemotherapy, but tumor recurrence and drug resistance are inevitable due to their highly invasive nature and extensive genetic heterogeneity (Ceccarelli et al., 2016; Hayes et al., 2018). Therefore, it is urgent to better understand the pathogenesis and molecular characteristics of LGG and develop more driver genes for the early diagnosis and prognostic prediction of LGG.

Somatic copy-number alteration (SCNA), as a crucial somatic genetic alteration event, is exceedingly common in cancer (Baudis, 2007; Lopez-Gines et al., 2010; Negrini et al., 2010; Hollander et al., 2011). Up to 15% of the human genome is located in the CNV area (Stankiewicz and Lupski, 2010). Analysis of copy number profiles from multiple cancers shows that each sample carries an average of 24 amplifications and 18 deletions, and even some cancer samples have an average of 17% amplifications and 16% deletions in the human genome, compared with an average of 0.35%, less than 0.1% in normal samples (Beroukhim et al., 2010). However, the extensive complexity and high degree of heterogeneity of the human cancer genome have posed challenges in identifying these key genes driving the initiation and progression of LGG (Tamborero et al., 2013; Alizadeh et al., 2015; Yadav and De, 2015). Hence, identifying the key driver SCNAs that play causal roles in oncogenesis is crucial for understanding the occurrence and development of the LGG.

Recently, many approaches have been proposed to identify driver SCNAs. The traditional method identified the drivers by executing functional experiments on the genes located in the region (Pon and Marra, 2015). For example, through functional tests, Hagerstrand et al. (2013) found that *SKIL* and *TLOC1* that frequently amplified in multiple cancers were identified as drivers of 3q26, leading to subcutaneous tumor growth. However, the traditional method was time-consuming and expensive. With the accumulation of multi-dimensional omic data, multiple computational algorithms for identifying driver SCNAs have emerged. Some computational algorithms discovered drivers based on the alteration frequencies in cancer populations. For instance, GISTIC detected significant SCNAs by calculating the significance of gene amplification or deletion across cancer samples (Mermel et al., 2011). The effects of copy number alterations on the expressions of other genes were introduced to reflect the functional influence of driver SCNAs. DriverNet is a computational framework to identify the minimum number of driver genes that explain transcriptome changes with the largest extent across cancer samples (Bashashati et al., 2012). Akavia et al. (2010) identified driver genes with higher frequencies by regulating expression to influence the expressions of other gene sets by integrating copy number and expression profiles in human cancers. However, it was extremely limited in making clear the functions of SCNAs in cancer and how they contributed to malignant phenotypes.

Random walk with restart (RWR) (Kohler et al., 2008), as one of the classic network propagation algorithms, can capture the global structure of the network and has the characteristics of robustness to the noise in the network (Zhang C. et al.,

2015). In the field of biology, RWR has been used to capture the disordered information that disease-related genes transmit through the topological structures of biological networks for identifying disease genes, mining disease modules, and predicting drug target (Robinson et al., 2008; Li and Patra, 2010; Cowen et al., 2017). HotNet2 and Hierarchical HotNet applied a similar RWR procedure but implemented different approaches to identify disease-related modules (Leiserson et al., 2015; Reyna et al., 2018). These studies indicate that the RWR algorithm could help us to capture the driver effects of SCNAs through topological structures of biological networks.

The present study developed an integrated computational framework to identify the key SCNAs driving the dysregulation of cancer hallmarks in LGG. We used the RWR algorithm to build the candidate gene-hallmark network by estimating the driver effects of seed genes on the cancer hallmarks based on the weighted co-expression protein interaction network. We also used linear regression analysis to identify the driver gene sets that cooperatively contributed to the dysregulation of hallmarks. We found that driver gene-hallmark pairs could identify two new LGG subtypes with a similarly poor prognosis.

MATERIALS AND METHODS

The copy number and mRNA expression data of LGG patients were obtained from the cBioPortal database[1]. Additionally, high confidence protein interaction data were obtained from STRING (Franceschini et al., 2013) which included 419,720 pairs of interactions involving 17,155 protein coding genes (PCGs). Finally, we downloaded 50 hallmark signature gensets from the Msigdb database[2].

Overview of the Method for Identifying the Key Copy Number Alterations Driving the Dysregulation of Cancer Hallmarks

We developed an integrated method to identify key SCNAs using network propagation algorithm and regression analysis through integrating the profiles of copy number alterations and gene expression and weighted co-expression protein interaction network (**Figure 1**).

Filtering the Candidate Genes With Somatic Copy-Number Alterations

Genes with SCNAs in the copy number profile were filtered as candidate genes by integrating profiles of copy number and gene expression if these genes met the following four criteria: (1) Genes should have a dominant SCNAs type (either high-level amplification or homogenous deletion) using the binomial test at $P < 0.05$ (Zhang et al., 2017). (2) Genes should be expressed in at least 10% of cancer samples (Zhou et al., 2017). (3) Genes with the dominant type should be altered in at least 3% of cancer samples. (4) The dominant SCNAs of genes should have a concordant

---

[1]https://www.cbioportal.org/

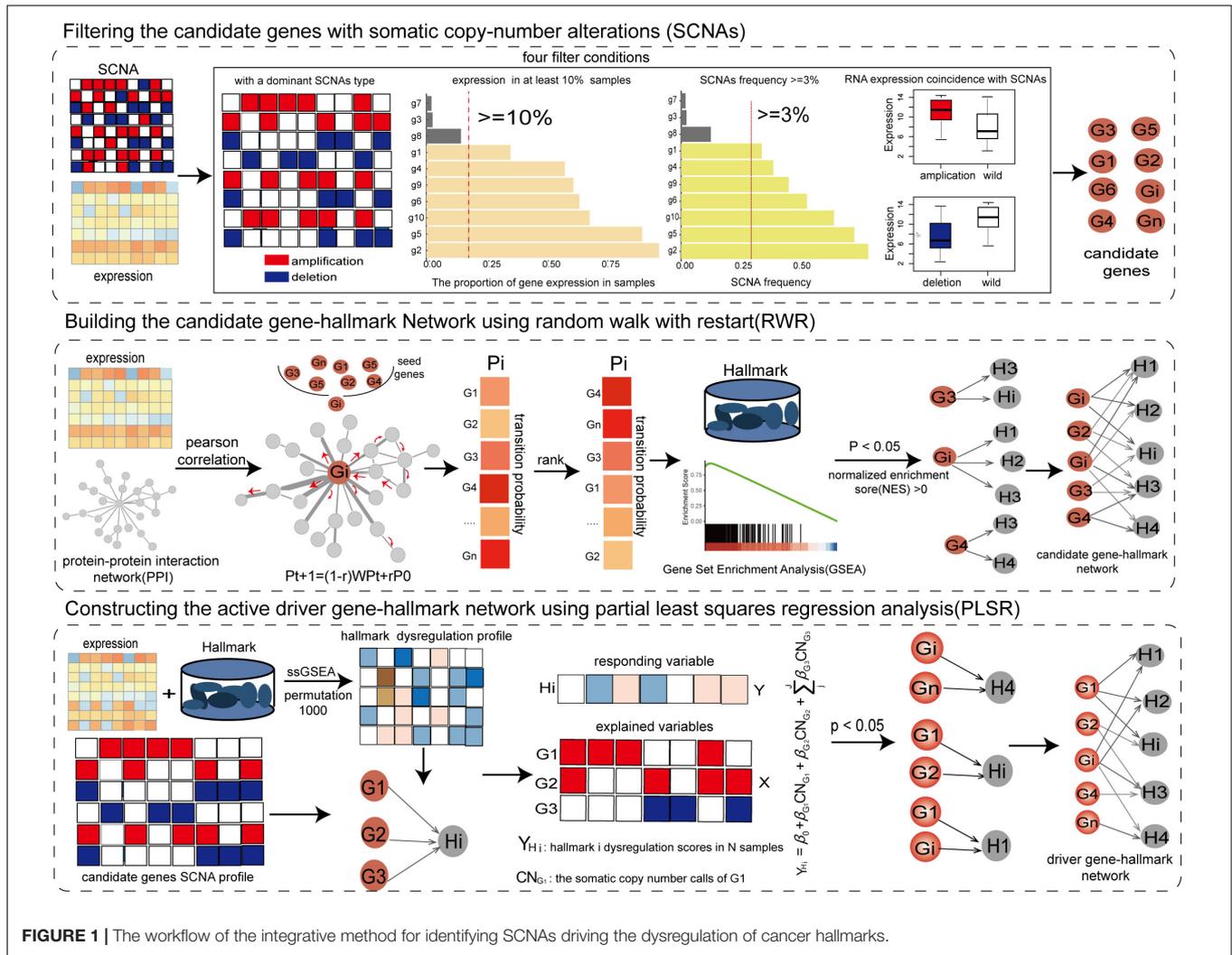[2]https://www.gsea-msigdb.org/gsea/msigdb/

FIGURE 1 | The workflow of the integrative method for identifying SCNAs driving the dysregulation of cancer hallmarks.

influence on their expression (one-tailed Wilcoxon rank-sum test with FDR < 0.05). The genes selected by the four criteria were used to identify driver genes.

Building the Candidate Gene-Hallmark Network Using Random Walk With Restart

By integrating the gene expression profile of cancer samples and protein interaction network, the weighted co-expression protein interaction network was constructed by calculating the expression correlation value for each pair of protein interactions. The weighted protein interaction network was unsigned, in which the weight of each edge represented the active extent between PCGs. According to the RWR principle, the higher the correlation coefficient between PCGs, the greater the probability of transmitting the imbalanced information. To measure the driver influence of candidate genes, each candidate gene was used as seed node to carry out a RWR (Kohler et al., 2008) on the weighted co-expression protein interaction network. The stable transfer probabilities from the seed gene reflected the driver

extent by the seed gene on the genes in the protein interaction network. The information flow could restart from the seed genes with probability r.

$$P_{t1} = (1 - r)WP_t + rP_0$$

where r was set to 0.3; P_0 were the initial probabilities of genes, in which the probabilities of seed genes were 1; P_t were the transfer probabilities of genes at the t step; The $P_t + 1$ characterized the stable transfer probabilities; W was the normalized transfer matrix of the weighted co-expression protein interaction network; the random walk process reached the steady-state when the maximum difference between $P_t + 1$ and P_t was less than <1e−8. Based on the stable transfer probabilities, we used gene set enrichment analysis (GSEA) (Subramanian et al., 2005) to identify the significant cancer hallmarks driven by the candidate gene at NES (normalized enrichment score) >0 and $P < 0.05$. The candidate gene and its driving cancer hallmarks formed candidate gene-hallmark pairs and were then used to build a candidate gene-hallmark network by collecting all candidate gene-hallmark pairs from all candidate genes.

Constructing the Active Driver Gene-Hallmark Network Using Partial Least Squares Regression Analysis

The gene expression profile was standardized by dividing the mean expression to eliminate the natural expression levels of genes in cancer samples. To assess dysfunctional activities of hallmark signatures in each sample, we built the significant dysfunctional profile of cancer hallmark signatures using single-sample GSEA (ssGSEA) (Hanzelmann et al., 2013). To measure the significance of dysfunctional activity, we permuted the gene expression profile of each sample 1,000 times, and recalculated 1,000 random scores of dysfunctional activities for each hallmark signature. The significance of activation was calculated as the frequency with which the random activity differences were greater than the actual activity score. Conversely, the significance of inactivation was calculated by the frequency of the random activity differences less than the real one. The significant activation/inactivation hallmark signatures in each sample were identified at FDR < 0.05.

To further determine the driver genes contributing to the variation of dysfunctional activities of hallmarks, a linear regression model was used to dissect the driver extents of driver genes on hallmarks based on driver relationships from the candidate gene-hallmark network. For each significant dysfunction hallmark in cancer samples, the candidate genes driving this hallmark were extracted. The partial least squares regression (PLSR) model was adopted to identify driver genes that cooperatively contributed to the dysfunctional activities of this hallmark (Bjornstad et al., 2004), in which the significant activity scores of the hallmark in cancer samples were the responding variables (Y), and the SCNAs statuses of candidate genes were explained variables (X).

The significance of the driver effects of candidate genes on dysfunctional activities of the hallmark was assessed using 10-fold cross-validation in R package "pls" (R functions of PLSR, RMSEP, and Jack.test). Subsequently, the driver genes that significantly contributed to the dysfunction of the hallmark with significant coefficients ($P < 0.05$) were identified, and active driver gene-hallmark pairs were formed. The active driver gene-hallmark network was constructed which consisted of all active driver gene-hallmark pairs. The genes in the active gene-hallmark network were identified as driver genes.

RESULTS

Identifying Cancer Hallmarks Driven by Somatic Copy Number Alterations Using Random Walk With Restart in LGG

In total, 511 LGG samples detected in both expression profile and copy number profile were used for subsequent analysis. Among the 20,530 PCGs (protein coding genes) with SCNAs, 19,059 PCGs were expressed in at least 10% of LGG samples. 10,280 PCGs showed one dominant SCNAs type (high-level amplification or homozygous deletion) (binomial test, $P < 0.05$), of which 1,736 P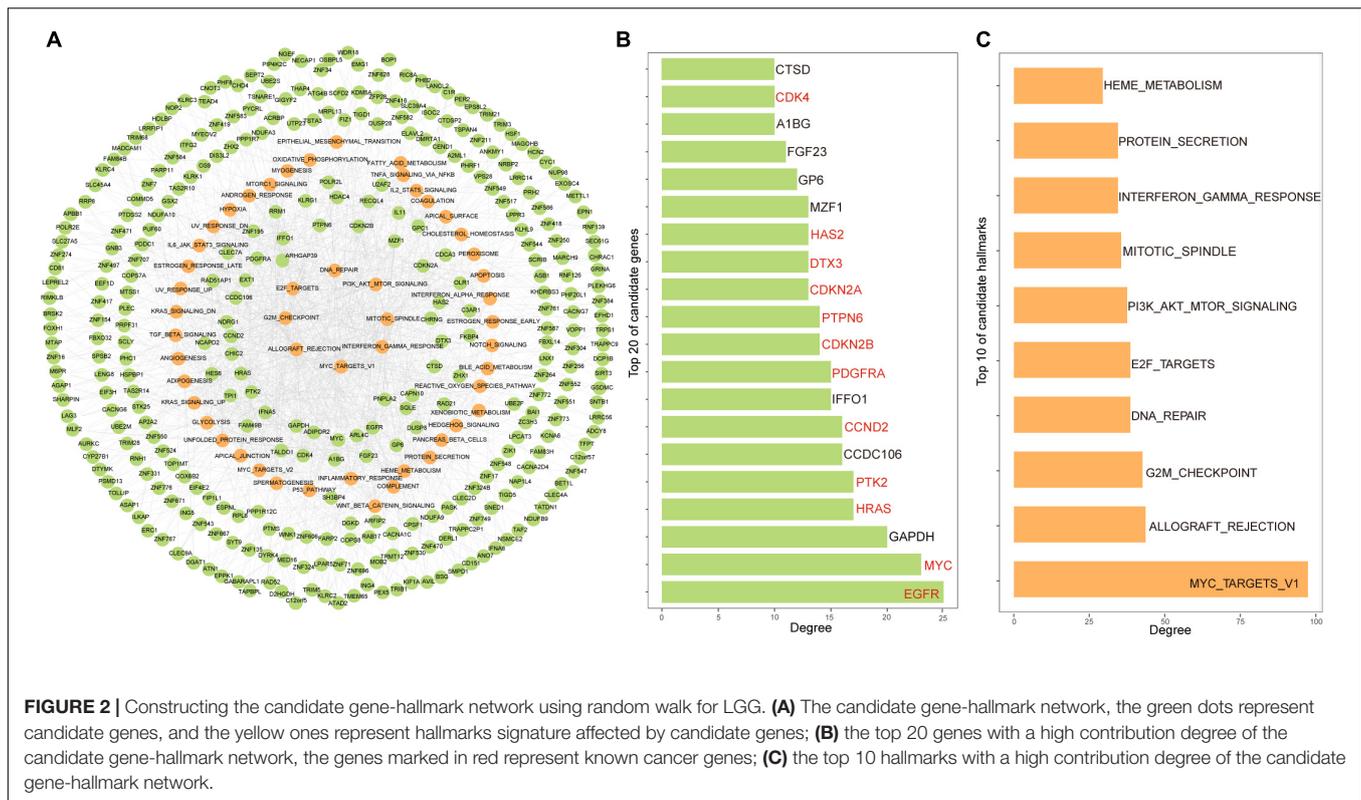CGs somatic copy number alterations significantly affected their expressions (FDR < 0.05). Eventually, 391 candidate PCGs with SCNAs frequency greater than 3% were identified, and only 40% of cancer samples occurred SCNAs of these candidate PCGs.

The weighted protein interaction network was constructed in LGG, in which each interaction was weighted by the expression correlation coefficient of the gene pair, representing the active extent of the interaction. The dysfunctional information of genes could be diffused following the topological structure of the weighted protein interaction network. The genes with higher expression correlation coefficients had higher probabilities to receive dysfunctional information. To estimate the driver effects of candidate genes on cancer hallmarks, each candidate gene was sowed as a seed node in the weighted protein interaction network, and the RWR algorithm was used to calculate the stable transition probability from the candidate gene to the other genes in the protein interaction network. The significant cancer hallmarks driven by the candidate gene were identified using GSEA (Subramanian et al., 2005) according to the rank of stable transition probabilities (NES > 0, $P < 0.05$). For instance, EGFR amplification was frequent in LGG, which significantly upregulated the EGFR expression (**Supplementary Figure 1**). We found that EGFR amplification significantly affected many cancer hallmarks, including development signature (EPITHELIAL_MESENCHYMAL_TRANSITION, $P = 0.012$), immune signature (ALLOGRAFT_REJECTION, $P = 0.004$; IL6_JAK_STAT3_SIGNALING, $P = 0.005$; INTERFERON_GAMMA_RESPONSE, $P = 0.013$) (**Supplementary Figure 2A**).

By assembling cancer hallmarks driven by all candidate genes, the candidate gene-hallmark network was constructed, in which nodes represented candidate genes and cancer hallmarks, and edges represented driver relationships among them. The candidate gene-hallmark network involved 1,177 gene-hallmark pairs including 329 PCGs and 50 hallmarks (**Figure 2A**). We ranked candidate genes in descending order according to their degree in the gene-hallmark network and found that 60% of the top 20 genes were known cancer genes including EGFR, MYC, HRAS, PTK2, CCND2, PDGFRA, CDKN2B, PTPN6, CDKN2A, DTX3, HAS2, and CDK4 (**Figure 2B**). For ranking hallmark signatures, among the top ten hallmarks, four were related to proliferation signatures, two were related to immune signatures, one was related to signal signature, one was related to pathway signature, one was related to DNA damage signature, and one was related to the metabolic signature (**Figure 2C**).

The Driver SCNAs in LGG Based on the Active Gene-Hallmark Driving Network

Single-sample gene set enrichment analysis algorithm was used to evaluate the active scores of cancer hallmarks for each LGG patient. In order to characterize dysfunctional activities of hallmark signatures in each patient, the dysfunctional profile of cancer hallmarks was constructed by permutating the expression profile of each sample 1,000 times at the threshold of FDR < 0.05 (**Figure 3A**). A total of 39 cancer hallmarks were dysregulated in at least one LGG patient. Among them, eight hallmarks were dysregulated in more

FIGURE 2 | Constructing the candidate gene-hallmark network using random walk for LGG. **(A)** The candidate gene-hallmark network, the green dots represent candidate genes, and the yellow ones represent hallmarks signature affected by candidate genes; **(B)** the top 20 genes with a high contribution degree of the candidate gene-hallmark network, the genes marked in red represent known cancer genes; **(C)** the top 10 hallmarks with a high contribution degree of the candidate gene-hallmark network.

than 50% of LGG samples including two proliferation signatures (E2F_TARGETS and G2M_CHECKPOINT), four immune signatures (INTERFERON_ GAMMA_ RESPONSE, INTERFERON_ALPHA_RESPONSE, INFLAMMATORY_ RESPONSE, and ALLOGRAFT_ REJECTION), one development signature (EPITHELIAL_MESENCHYMAL_ TRANSITION), one signaling signature (TNFA_SIGNALING_ VIA_ NFKB). None of these cancer hallmarks showed a consistent state of activity in the LGG population. For example, the proliferation signature E2F_TARGETS was dysregulated in 339 LGG samples, which was significantly activated in 131 samples and inactivated in 208 samples. These results showed the activity heterogeneities of cancer hallmarks across the LGG population.

To further identify the driver factors underlying the activity heterogeneities of cancer hallmarks, the linear regression analysis was used to identify the driver gene sets that cooperatively contributed to activity changes of hallmarks by integrating the hallmark dysfunctional profile, the SCNAs profile of the candidate PCGs and the candidate gene-hallmark network. For each dysregulated cancer hallmark, the driver genes were identified at FDR < 0.05. The SCNAs of the driver genes could affect the activities of cancer hallmarks. For example, *EGFR* amplification significantly influenced the activities of EPITHELIAL_MESENCHYMAL_TRANSITION (P = 5.7e-09), ALLOGRAFT_ REJECTION (P = 7.2e-06), IL6_JAK_STAT3_SIGNALING (P = 0.002), INTERFERON_ GAMMA_ RESPONSE (P = 2.7e-08) (**Supplementary Figure 2B**). Finally, we found that 29 cancer hallmark signatures

whose dysfunctional activities were significantly driven by 98 driver genes, forming an active gene-hallmark network with 143 gene-hallmark pairs (**Figure 3B**). Among 98 driver genes, 21 genes (including *CDKN2A, CDKN2B, FBXO32, SIRT3, ING4, PTPN6, CHIC2, RAD21, PDGFRA, EGFR, HRAS, ERC1, CDK4, HDLBP, DTX3, C3AR1, GPC1, TALDO1,* and *MTAP*) were recorded as cancer genes in at least one of four known databases [Cancer Gene Census in COSMIC (Sondka et al., 2018), TSGene (Zhao et al., 2016), Bushman[3], and DriverDBv3 (Liu et al., 2020)]. Specifically, based on the degree of driver genes in the active gene-hallmark network, we found that 70% of the top 10 driver genes were cancer genes.

After conducting a manual literature search, we also found that 97.4% (75/77) of the remaining driver genes identified by our method were reported as cancer-associated genes *in vivo* or *in vitro* (**Supplementary Table 1**). For instance, Menezes et al. (2020) found that MTAP overexpression was associated with proliferation, migration, and invasion of glioma cells *in silico* and *in vitro* models. The low expression of PTPN6 was significantly associated with poor overall survival in bladder cancer patients and co-expression with TNFRSF14 (tumor necrosis factor receptor superfamily member 14) had a close correlation in breast cancer (Menezes et al., 2020). METTL1 promoted the proliferation and migration of hepatocellular carcinoma cells by inhibiting the PTEN signaling pathway and was associated with poor prognosis (Tian et al., 2019). Additionally, the ratio of known cancer genes in the active gene-hallmark network was
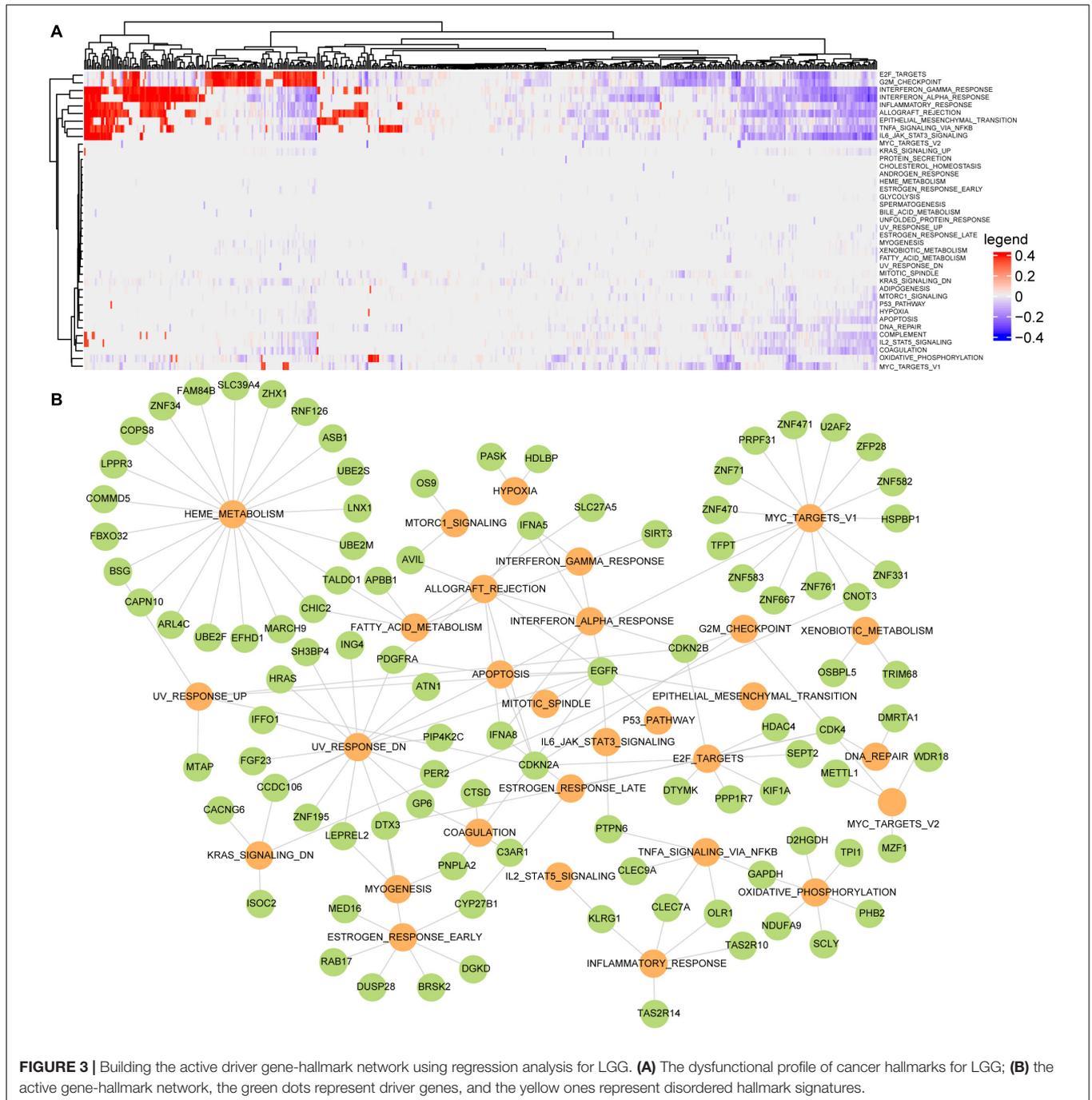
---

[3]http://www.bushmanlab.org/links/genelists

FIGURE 3 | Building the active driver gene-hallmark network using regression analysis for LGG. **(A)** The dysfunctional profile of cancer hallmarks for LGG; **(B)** the active gene-hallmark network, the green dots represent driver genes, and the yellow ones represent disordered hallmark signatures.

elevated to 21.4% (21/98) by comparing that ratio of 15.5% (51/329) in the candidate gene-pathway network. These findings confirmed the driving roles of the identified PCGs.

LGG Subtypes With Poor Prognosis Contributed by the Active Gene-Hallmark Network

To investigate whether the gene-hallmark pairs from the active gene-hallmark driving network were associated with LGG
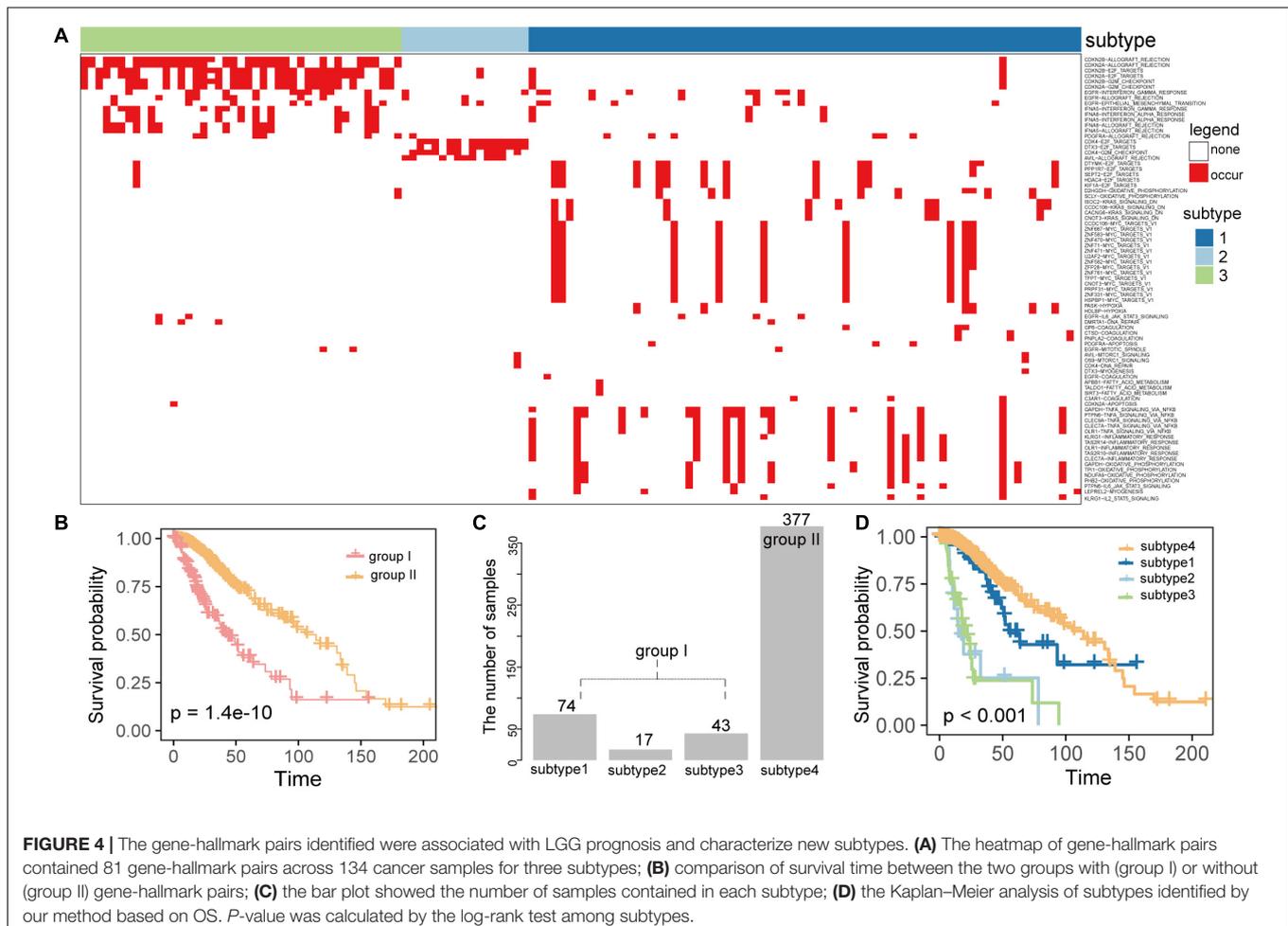
prognosis and characterized new subtypes, we mapped the active gene-hallmark network into LGG patients and constructed a binary profile of gene-hallmark pairs for the LGG population, where each row was a pair of gene-hallmark, each column was a cancer sample and value referred to whether this gene-hallmark pair occurred in a certain sample. A gene-hallmark pair was considered to be present if the gene showed an SCNA and the cancer hallmark was significantly activated or inactivated in this sample. The binary profile of gene-hallmark pairs contained 81 gene-hallmark pairs across 511 cancer samples. We observed that

134 samples carried at least one gene-hallmark pair (group I) (**Figure 4A**) while the other 377 samples did not carry any of these pairs (group II), mainly because 303 samples out of 377 did not harbor any SCNAs. Using survival analysis, we found the survival time of 134 samples were significantly shorter than that of 377 samples ($P = 1.4e-10$, log-rank test, **Figure 4B**). Based on the active gene-hallmark pairs, these 134 patients (group I) were classified using consensus clustering [50 resamplings, 80% item resampling, partitioning around medoids (PAM) clustering method]. The consensus heatmap and cumulative distribution function were used to determine the optimal *K*. We identified three subtypes (1, 2, and 3) with the largest relative change in area under the CDF curve ($K = 3$) (**Supplementary Figures 3A–C**; Wilkerson and Hayes, 2010; Senbabaoglu et al., 2014). Finally, 511 LGG patients were stratified into four subtypes and the sample numbers of the four subtypes were 74, 17, 43, 377 accounting for 14.5, 3.3, 8.4, 73.8% of the 511 samples, respectively (**Figure 4C**). We compared the overall survival times of the three subtypes with that of subtype 4 (without gene-hallmark pairs), found that there were significant differences in overall survival time between the four groups ($P < 0.001$, log-rank test, **Figure 4D**). Subtype 2 and 3 showed significantly poorer prognosis than subtype 4 ($P = 3.07e-13$ for subtype 2; $P = 0$ for subtype 3, log-rank

test, **Supplementary Table 2**), and subtype 1 showed shorter survival time with weaker significance. ($P = 0.089$, log-rank test, **Supplementary Table 2**). Furthermore, we compared the survival times among the three subtypes and found there were significant survival differences ($P = 2.37e-07$, log-rank test, **Supplementary Table 2**). The subtype 2 and 3 showed significantly shorter survival times than that of subtype 1 ($P = 1.77e-05$ for subtype 2; $P = 1.64e-07$ for subtype 3, log-rank test, **Supplementary Table 2**), but there was no significance in survival times between subtype 2 and 3($P = 0.89$, log-rank test, **Supplementary Table 2**). Together, these results proved that the active gene-hallmark pairs could identify new LGG subtypes with poor prognoses.

The Subtype-Specific Gene-Hallmark Pairs in LGG

Subtypes 2 and 3 had the poorest prognosis and showed no significance in survival time. To explore the molecular mechanisms underlying these subtypes, we investigated the distribution of gene-hallmark pairs across LGG samples and identified the subtype specific gene-hallmark pairs using Fisher's exact test at $P < 0.05$ for each subtype (**Supplementary Table 3**). For subtype 2, specific gene-hallmark pairs involving driver genes *CDK4, DTX3,* and *AVIL* were identified (**Supplementary Figure 4A**). For example, *CDK4* specifically

FIGURE 4 | The gene-hallmark pairs identified were associated with LGG prognosis and characterize new subtypes. **(A)** The heatmap of gene-hallmark pairs contained 81 gene-hallmark pairs across 134 cancer samples for three subtypes; **(B)** comparison of survival time between the two groups with (group I) or without (group II) gene-hallmark pairs; **(C)** the bar plot showed the number of samples contained in each subtype; **(D)** the Kaplan–Meier analysis of subtypes identified by our method based on OS. *P*-value was calculated by the log-rank test among subtypes.

and significantly drove two proliferation signatures (P = 6.02e-15 for E2F_TARGETS and P = 3.01e-12 for G2M_CHECKPOINT, Fisher's exact test). Amplification of *CDK4, DTX3,* and *AVIL* significantly upregulated their gene expression (P = 2.4e-14 for *CDK4*, P = 1.7e-08 for *DTX3*, and P = 4.6e-10 for *AVIL*, Wilcoxon rank-sum test), and showed significant association with LGG poor prognosis (P = 1.2e-10 for *CDK4*, P = 8.03e-10 for *DTX3*, and P = 3.86e-6 for *AVIL*, log-rank test, **Supplementary Figure 4B**). *AVIL*, as a novel oncogene gene in LGG, was reported to play important roles in functions of cancer *in vivo* or *in vitro* through literature searches. For example, *AVIL* overexpression in GBM promoted cell proliferation and metastasis, leading to poor prognosis in patients (Xie et al., 2020). In the subtype 3, 12 specific gene-hallmark pairs were identified including driver genes (*EGFR, CDKN2A, CDKN2B, INFA8,* and *INFA5*), and their affected hallmarks were significantly presented in 43 LGG individuals (P < 0.05, Fisher's exact test, **Supplementary Figure 4A**). *EGFR* amplification and the deletions of *CDKN2A, CDKN2B, INFA8*, and *INFA5* were also significantly associated with LGG poor prognosis (P < 0.001 for *CDKN2A, CDKN2B, P* = 1.42e-12 for *IFNA8*, P = 1.76e-12 for *IFNA5*, and P < 0.001 for *EGFR*, log-rank test, **Supplementary Figure 4B**).

Of note, we observed these specific gene-hallmark pairs in subtype 2 and 3, which showed obvious mutually exclusive patterns (**Supplementary Figure 4A**), but these driver genes drove the same dysregulated hallmark signatures such as proliferation (E2F_TARGETS and G2M_CHECKPOINT) and immunity signature (ALLOGRAFT_ REJECTION, INTERFERON_ GAMMA_ RESPONSE, and INTERFERON_ALPHA_ RESPONSE) in both subtype 2 and 3. These results implied that these two subtypes with the poorest prognosis were driven by different driver genes through dysregulating the same biological functions, further improving the potential of the identified active gene-hallmark pairs for identifying new LGG subtypes.

Performance Evaluation of the Method

Known cancer genes were collected from multiple sources (**Table 1**). The driver genes identified by our method significantly overlapped with these known cancer gene sets (hypergeometric test, P < 0.05, **Table 1**). We then further compared our method with the other three previous methods, which also identified driver copy number alterations [Ping et al., 2020; Zhou et al., 2017, and DriverDBv3 (Liu et al., 2020)]. The results showed that our method could also significantly capture the driver genes identified by other methods with the Ping et al. (2020), Zhou et al. (2017) (hypergeometric test, P = 1.84e-05 for Ping et al. (2020), P = 4.23e-12 for Zhou et al. (2017), **Table 1**), but not significantly with DriverDBv3(hypergeometric test, P = 0.175, **Table 1**). Among them, Ping et al. (2020) could directly dissect long non-coding RNAs (lncRNAs) functional roles in cancers based on dysregulated ceRNA network induced by SCNAs but was limited in identifying potential driver SCNAs with low expression levels and SCNA frequencies, however, Zhou et al. (2017) and DriverDBv3 cannot dissect driver roles of genes in cancers in a direct manner (**Supplementary Table 4**). Additionally, the driver genes identified by our method also showed a significantly

TABLE 1 | The performance of our method compared with six known cancer gene sources and three previous methods.

Type	Cancer gene sets	Number of cancer genes	Number of gene intersections	*P*-value
Known cancer gene sets	CGC	723	10	0.00103
	TSG	638	7	0.0193
	Bushman	2,579	18	0.0237
	Tamborero et al., 2013	435	5	0.0385
	Rahman, 2014	114	6	9.29e-06
	An et al., 2016	1,571	16	0.0427
Previous method	Ping et al., 2020	37	4	1.84e-05
	Zhou et al., 2017	88	10	4.23e-12
	DriverDBv3	45	1	0.175

higher degree than the non-drivers whose expression levels were concordant with SCNAs in the PPI network (**Supplementary Figure 5**, P < 0.05, Wilcoxon rank-sum test). Our method could complement other methods, which identified novel driver genes contributing to cancer development. For instance, *DTX3* was a gene that had not been identified by other methods, but it had the ability to drive E2F_TARGETS hallmark signature dysfunction and was associated with LGG prognosis. These results suggest that our method is useful in identifying driver SCNAs that contribute to dysfunction cancer hallmarks in LGG.

DISCUSSION

This article has proposed an integrative method to identify driver SCNAs and characterize the dysfunctional cancer hallmarks driven by these SCNAs in LGG. The active driver gene-hallmark pairs identified could stratify the LGG patients into four subtypes and identify two new subtypes with similar poor prognoses and with different underlying molecular mechanisms. We also discovered that different driver genes could disorder the same cancer hallmarks, leading to malignant phenotypes.

The dysfunctional effects of driver SCNAs were characterized by network propagation on the protein interaction networks. The selection of the specific forms of the protein interaction network was crucial. Our method used three forms of protein interaction network, including static protein interaction network, co-expression weighted PPI across only samples with SCNAs of the seed gene, weighted co-expression PPI across all cancer samples (our method). We found that the first two forms do not help to significantly capture the known cancer genes. For example, using the static PPI network, 147 driver PCGs dysregulating 27 hallmarks could not significantly overlap with TSGene (Zhao et al., 2016) (P = 0.225, hypergeometric test) and Bushman (P = 0.0622, hypergeometric test), which also did not account for the differences in co-expression patterns among genes from cancer to cancer. For the second, the identified active gene-hallmark network including 139

PCGs only weaker significantly overlapped with TSGene (Zhao et al., 2016) at the significance of $P = 0.0926$, which was not suitable for the PCGs with low SCNAs frequencies. The development of a comprehensive dynamic PPI network in the future could help us to better analyze the functional roles of SCNAs.

Two subtypes of poor prognosis were driven by different driver gene-hallmark pairs. The subtype 3 characterized by driver genes *CDKN2A* and *CDKN2B* was consistent with our previous studies (Ping et al., 2020), in which identified *CDKN2A* and *CDKN2B* could dysregulate the hallmark of G2M_CHECKPOINT based on miRNA-mediated ceRNA networks and further contributed to poor LGG prognosis. The SCNAs of *CDK4, AVIL,* and *DTX3* driven another new subtype with poor prognosis showed significant mutual exclusivity with that of *CDKN2A* and *CDKN2B* (**Supplementary Figure 6**). These phenomenons forming the hypothesis of functional redundancy were extensively used to identify the cancer genes (Sparks et al., 1998; Zhao et al., 2012; Babur et al., 2015). For example, Deng et al. (2018) identified cancer driver lncRNAs that were mutually exclusive with well-known driver genes based on functional redundancy hypothesis. Ping et al. (2014) showed that the glioma pathway was affected by 12 different genes (including *EGFR, PDGFRA, CAMK2B, AKT1, CDK4, MDM2, NRAS, PIK3CA, TGFA, SHC4, CDKN2A*, and *PDGFA*) with significant patterns of mutual exclusivity.

Another important finding of our method was that hallmark activity showed broad heterogeneity in the cancer population. For instance, proliferation (E2F_TARGETS and G2M_CHECKPOINT) and immunity signature (ALLOGRAFT_REJECTION) showed inconsistent activity statuses across the LGG population and dysregulated in more than 50% of LGG samples, but these hallmarks contributed malignant phenotypes driven by different genes in a mutually exclusive manner.

The four subtypes identified by our method were compared with WHO subtypes (*IDH* mutant and 1p/19q codeleted, *IDH* mutant and 1p/19q non-codeleted, *IDH* wild-type) (Louis et al., 2016). We found that subtype 2 and subtype 3 were mainly enriched in the *IDH* wild-type subtype, which showed the worst prognosis [58.8% (10/17) for subtype 2; 72.1% (31/43) for subtype3, **Supplementary Figures 7A,B**]. We identified the risk prognostic markers (including *CDKN2A/CDKN2B* deletion, *EGFR* amplification, *CDK4* amplification, IFNA5/IFNA8 deletion, *AVIL* amplification, and *DTX3* amplification) for *IDH* wild-type subtype from SCNAs level. For the *IDH* mutant and 1p/19q non-codeleted, and *IDH* wild-type subtypes, our subtypes could further divide them into four different subgroups with significant differences in survival times ($P = 0.0045$ for *IDH* mutant and 1p/19q non-codeleted, $P < 0.0001$ for *IDH* wild-type, log-rank test, **Supplementary Figure 7C**). Our subtypes were compared with the subtypes of LGG identified by eight different methods which had distinct molecular and clinical features (Ceccarelli et al., 2016). The comparison results were similar to that obtained from the comparison with WHO subtypes (**Supplementary Figures 8–15**). Our method provided the subtype and prognosis markers from the view of copy number alterations, which was served as an important complement for the existing methods.

To further verify our method, we applied our method to 507 lung adenocarcinoma (LUAD) samples and identified a total of 347 active gene-hallmark pairs containing 234 driver genes and 28 hallmarks. Based on the active gene-hallmark pairs, three subtypes were identified for LUAD samples, which contained 285, 170, 52 samples respectively (**Supplementary Figure 16A**). Survival analysis showed that there were significant differences in overall survival times among these three subtypes ($P = 0.038$, log-rank test, **Supplementary Figure 16B**). In the subtype 3 with the worst prognosis, specific gene-hallmark pairs involving driver genes *PAIP, RAP1B, AVIL, RAD1,* and *C1QTNF3* were identified. For example, amplification of *AVIL* specifically and significantly drove immunity signature ($P < 0.05$, for ALLOGRAFT_REJECTION, Fisher's exact test). These results proved that our strategy has the potential to be extended to dissecting the driver roles of SCNAs in other cancer types.

It is noteworthy that this study captured a subset of novel driver genes including *IFNA5, IFNA8,* and *DTX3* with low SCNAs frequencies in LGG populations. Among these, *IFNA5* deletion (5.09% of LGG samples), *IFNA8* deletion (5.28% of LGG samples), and *DTX3* amplification (3.52% of LGG samples) indicated poor survival in LGG. Kao et al. (2018) found that the upregulation of *IFNA5* activated the ERβ-Ube3a interaction which further facilitated hepatic progenitor cell differentiation. Human epidemiologic studies showed that SNPs in *IFNA8* were significantly associated with the overall survival of patients with WHO grade 2 to 3 gliomas (Fujita et al., 2010). Byun et al. (2018) found that SNPs located in the promoter of *IFNA8* affected GBM patient prognosis. Ding et al. (2020) found that *DTX3* played an important role in the progression, and acted as an anti-oncogene in esophageal carcinoma. Gatza et al. (2014) discovered that *DTX3* was essential for cell proliferation in breast tumors.

In this work, each candidate PCG was sowed as a seed in the weighted protein interaction network to identify driver SCNAs using RWR, so our method was limited to identifying PCGs rather than non-coding genes and there was no way to identify seed PCGs that were not in the network. In the future, this problem may be solved with continuous improvement and expansion of the network or through different network forms, such as co-expression network, dysregulated ceRNA network. Additionally, Pearson correlation coefficient (PCC) was used to measure correlations between PCGs, which could not distinguish between direct and indirect relationships and ignored nonlinear relations between two genes due to only relying on the information of co-occurring events. Instead of PCC, we could in future use partial correlation similarly, using conditional mutual information to construct a direct association network (Zhang et al., 2013; Zhang X. et al., 2015; Liu et al., 2016).

The present study developed an integrative strategy to discover the key SCNAs driving dysfunction of cancer hallmarks and investigate the functional roles of driver SCNAs based on the weighted protein interaction network in the LGG population, which complimented previous methods. Our strategy could be extended to explore other driver factors, with the accumulation of multi-omics of multiple cancers.

THE CODE FOR THE INTEGRATIVE METHOD

The code for the integrative method is available at https://github.com/zhouyao-max/Driver_SCNAs/tree/master/Driver_SCNAs.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

YP, YL, and JH designed and guided this work. YZ, YP, and SW participated in data processing, program implementation, and manuscript writing. HY, BP, XZ, LP, YW, and JX contributed to data collecting and organized the figures and tables. All authors provided critical advice for the final manuscript.

FUNDING

ACKNOWLEDGMENTS

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.654736/full#supplementary-material

REFERENCES

Akavia, U. D., Litvin, O., Kim, J., Sanchez-Garcia, F., Kotliar, D., Causton, H. C., et al. (2010). An integrated approach to uncover drivers of cancer. *Cell* 143, 1005–1017. doi: 10.1016/j.cell.2010.11.013

Alizadeh, A. A., Aranda, V., Bardelli, A., Blanpain, C., Bock, C., Borowski, C., et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat. Med.* 21, 846–853. doi: 10.1038/nm.3915

An, O., Dall'olio, G. M., Mourikis, T. P., and Ciccarelli, F. D. (2016). NCG 5.0: updates of a manually curated repository of cancer genes and associated properties from cancer mutational screenings. *Nucleic Acids Res.* 44, D992–D999. doi: 10.1093/nar/gkv1123

Babur, O., Gonen, M., Aksoy, B. A., Schultz, N., Ciriello, G., Sander, C., et al. (2015). Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome Biol.* 16:45. doi: 10.1186/s13059-015-0612-6

Bashashati, A., Haffari, G., Ding, J., Ha, G., Lui, K., Rosner, J., et al. (2012). DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. *Genome Biol.* 13:R124. doi: 10.1186/gb-2012-13-12-r124

Baudis, M. (2007). Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer* 7:226. doi: 10.1186/1471-2407-7-226

Beroukhim, R., Mermel, C. H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905. doi: 10.1038/nature08822

Bjornstad, A., Westad, F., and Martens, H. (2004). Analysis of genetic marker-phenotype relationships by jack-knifed partial least squares regression (PLSR). *Hereditas* 141, 149–165. doi: 10.1111/j.1601-5223.2004.01816.x

Byun, J., Schwartz, A. G., Lusk, C., Wenzlaff, A. S., De Andrade, M., Mandal, D., et al. (2018). Genome-wide association study of familial lung cancer. *Carcinogenesis* 39, 1135–1140. doi: 10.1093/carcin/bgy080

Ceccarelli, M., Barthel, F. P., Malta, T. M., Sabedot, T. S., Salama, S. R., Murray, B. A., et al. (2016). Molecular Profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* 164, 550–563. doi: 10.1016/j.cell.2015.12.028

Cowen, L., Ideker, T., Raphael, B. J., and Sharan, R. (2017). Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* 18, 551–562. doi: 10.1038/nrg.2017.38

Deng, Y., Luo, S., Zhang, X., Zou, C., Yuan, H., Liao, G., et al. (2018). A pan-cancer atlas of cancer hallmark-associated candidate driver lncRNAs. *Mol. Oncol.* 12, 1980–2005. doi: 10.1002/1878-0261.12381

Ding, X. Y., Hu, H. Y., Huang, K. N., Wei, R. Q., Min, J., Qi, C., et al. (2020). Ubiquitination of NOTCH2 by DTX3 suppresses the proliferation and migration of human esophageal carcinoma. *Cancer Sci.* 111, 489–501. doi: 10.1111/cas.14288

Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., et al. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808–D815. doi: 10.1093/nar/gks1094

Fujita, M., Scheurer, M. E., Decker, S. A., Mcdonald, H. A., Kohanbash, G., Kastenhuber, E. R., et al. (2010). Role of type 1 IFNs in antiglioma immune surveillance–using mouse studies to guide examination of novel prognostic markers in humans. *Clin. Cancer Res.* 16, 3409–3419. doi: 10.1158/1078-0432.CCR-10-0644

Gatza, M. L., Silva, G. O., Parker, J. S., Fan, C., and Perou, C. M. (2014). An integrated genomics approach identifies drivers of proliferation in luminal-subtype human breast cancer. *Nat. Genet.* 46, 1051–1059. doi: 10.1038/ng.3073

Hagerstrand, D., Tong, A., Schumacher, S. E., Ilic, N., Shen, R. R., Cheung, H. W., et al. (2013). Systematic interrogation of 3q26 identifies TLOC1 and SKIL as cancer drivers. *Cancer Discov.* 3, 1044–1057. doi: 10.1158/2159-8290.CD-12-0592

Hanzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform.* 14:7. doi: 10.1186/1471-2105-14-7

Hayes, J., Yu, Y., Jalbert, L. E., Mazor, T., Jones, L. E., Wood, M. D., et al. (2018). Genomic analysis of the origins and evolution of multicentric diffuse lower-grade gliomas. *Neuro Oncol.* 20, 632–641. doi: 10.1093/neuonc/nox205

Hollander, M. C., Blumenthal, G. M., and Dennis, P. A. (2011). PTEN loss in the continuum of common cancers, rare syndromes and mouse models. *Nat. Rev. Cancer* 11, 289–301. doi: 10.1038/nrc3037

Kao, T. L., Kuan, Y. P., Cheng, W. C., Chang, W. C., Jeng, L. B., Yeh, S., et al. (2018). Estrogen receptors orchestrate cell growth and differentiation to facilitate liver regeneration. *Theranostics* 8, 2672–2682. doi: 10.7150/thno.23624

Kohler, S., Bauer, S., Horn, D., and Robinson, P. N. (2008). Walking the interactome for prioritization of candidate disease genes. *Am. J. Hum. Genet.* 82, 949–958. doi: 10.1016/j.ajhg.2008.02.013

Leiserson, M. D., Vandin, F., Wu, H. T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., et al. (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* 47, 106–114. doi: 10.1038/ng.3168

Li, Y., and Patra, J. C. (2010). Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics* 26, 1219–1224. doi: 10.1093/bioinformatics/btq108

Liu, S. H., Shen, P. C., Chen, C. Y., Hsu, A. N., Cho, Y. C., Lai, Y. L., et al. (2020). DriverDBv3: a multi-omics database for cancer driver gene research. *Nucleic Acids Res.* 48, D863–D870. doi: 10.1093/nar/gkz964

Liu, X., Wang, Y., Ji, H., Aihara, K., and Chen, L. (2016). Personalized characterization of diseases using sample-specific networks. *Nucleic Acids Res.* 44:e164. doi: 10.1093/nar/gkw772

Lopez-Gines, C., Gil-Benso, R., Ferrer-Luna, R., Benito, R., Serna, E., Gonzalez-Darder, J., et al. (2010). New pattern of EGFR amplification in glioblastoma and the relationship of gene copy number with gene expression profile. *Mod. Pathol.* 23, 856–865. doi: 10.1038/modpathol.2010.62

Louis, D. N., Perry, A., Reifenberger, G., Von Deimling, A., Figarella-Branger, D., Cavenee, W. K., et al. (2016). The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol.* 131, 803–820. doi: 10.1007/s00401-016-1545-1

Menezes, W. P., Silva, V.a.O, Gomes, I. N. F., Rosa, M. N., Spina, M. L. C., Carloni, A. C., et al. (2020). Loss of 5'-Methylthioadenosine Phosphorylase (MTAP) is Frequent in High-Grade Gliomas; Nevertheless, it is Not Associated with Higher Tumor Aggressiveness. *Cells* 9, 492. doi: 10.3390/cells9020492

Mermel, C. H., Schumacher, S. E., Hill, B., Meyerson, M. L., Beroukhim, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* 12:R41. doi: 10.1186/gb-2011-12-4-r41

Negrini, S., Gorgoulis, V. G., and Halazonetis, T. D. (2010). Genomic instability–an evolving hallmark of cancer. *Nat. Rev. Mol. Cell. Biol.* 11, 220–228. doi: 10.1038/nrm2858

Ostrom, Q. T., Gittleman, H., Farah, P., Ondracek, A., Chen, Y., Wolinsky, Y., et al. (2013). CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. *Neuro Oncol.* 15 (Suppl 2), ii1–ii56. doi: 10.1093/neuonc/not151

Ping, Y., Zhang, H., Deng, Y., Wang, L., Zhao, H., Pang, L., et al. (2014). IndividualizedPath: identifying genetic alterations contributing to the dysfunctional pathways in glioblastoma individuals. *Mol. Biosyst.* 10, 2031–2042. doi: 10.1039/c4mb00289j

Ping, Y., Zhou, Y., Hu, J., Pang, L., Xu, C., and Xiao, Y. (2020). Dissecting the functional mechanisms of somatic copy-number alterations based on dysregulated ceRNA networks across cancers. *Mol. Ther. Nucleic Acids* 21, 464–479. doi: 10.1016/j.omtn.2020.06.012

Pon, J. R., and Marra, M. A. (2015). Driver and passenger mutations in cancer. *Annu. Rev. Pathol.* 10, 25–50. doi: 10.1146/annurev-pathol-012414-040312

Rahman, N. (2014). Realizing the promise of cancer predisposition genes. *Nature* 505, 302–308. doi: 10.1038/nature12981

Reyna, M. A., Leiserson, M. D. M., and Raphael, B. J. (2018). Hierarchical HotNet: identifying hierarchies of altered subnetworks. *Bioinformatics* 34, i972–i980. doi: 10.1093/bioinformatics/bty613

Robinson, P. N., Kohler, S., Bauer, S., Seelow, D., Horn, D., and Mundlos, S. (2008). The Human Phenotype Ontology: a tool for annotating and analyzing human hereditary disease. *Am. J. Hum. Genet.* 83, 610–615. doi: 10.1016/j.ajhg.2008.09.017

Senbabaoglu, Y., Michailidis, G., and Li, J. Z. (2014). Critical limitations of consensus clustering in class discovery. *Sci. Rep.* 4:6207. doi: 10.1038/srep06207

Sondka, Z., Bamford, S., Cole, C. G., Ward, S. A., Dunham, I., and Forbes, S. A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat. Rev. Cancer* 18, 696–705. doi: 10.1038/s41568-018-0060-1

Sparks, A. B., Morin, P. J., Vogelstein, B., and Kinzler, K. W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res.* 58, 1130–1134.

Stankiewicz, P., and Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu. Rev. Med.* 61, 437–455. doi: 10.1146/annurev-med-100708-204735

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102

Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C., Deu-Pons, J., Kandoth, 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, Y., Zhang, S., Xiao, Q., Wang, J., Zhao, K., Liu, W., et al. (2020). Prognostic significance of ARL9 and its methylation in low-grade glioma. *Genomics* 112, 4808–4816. doi: 10.1016/j.ygeno.2020.08.035

Tian, Q. H., Zhang, M. F., Zeng, J. S., Luo, R. G., Wen, Y., Chen, J., et al. (2019). METTL1 overexpression is correlated with poor prognosis and promotes hepatocellular carcinoma via PTEN. *J. Mol. Med. (Berl.)* 97, 1535–1545. doi: 10.1007/s00109-019-01830-9

Wilkerson, M. D., and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 26, 1572–1573. doi: 10.1093/bioinformatics/btq170

Wu, F., Wang, Z. L., Wang, K. Y., Li, G. Z., Chai, R. C., Liu, Y. Q., et al. (2020). Classification of diffuse lower-grade glioma based on immunological profiling. *Mol. Oncol.* 14, 2081–2095. doi: 10.1002/1878-0261.12707

Xie, Z., Janczyk, P. L., Zhang, Y., Liu, A., Shi, X., Singh, S., et al. (2020). A cytoskeleton regulator AVIL drives tumorigenesis in glioblastoma. *Nat. Commun.* 11:3457. doi: 10.1038/s41467-020-17279-1

Yadav, V. K., and De, S. (2015). An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinform.* 16, 232–241. doi: 10.1093/bib/bbu002

Zhang, C., Jiang, S., Chen, Y., Sun, Y., and Han, J. (2015). Fast Inbound Top-K query for random walk with restart. *Mach. Learn. Knowl. Discov. Databases* 9285, 608–624. doi: 10.1007/978-3-319-23525-7_37

Zhang, H., Deng, Y., Zhang, Y., Ping, Y., Zhao, H., Pang, L., et al. (2017). Cooperative genomic alteration network reveals molecular classification across 12 major cancer types. *Nucleic Acids Res.* 45, 567–582. doi: 10.1093/nar/gkw1087

Zhang, X., Liu, K., Liu, Z. P., Duval, B., Richer, J. M., Zhao, X. M., et al. (2013). NARROMI: a noise and redundancy reduction technique improves accuracy of gene regulatory network inference. *Bioinformatics* 29, 106–113. doi: 10.1093/bioinformatics/bts619

Zhang, X., Zhao, J., Hao, J. K., Zhao, X. M., and Chen, L. (2015). Conditional mutual inclusive information enables accurate quantification of associations in gene regulatory networks. *Nucleic Acids Res.* 43:e31. doi: 10.1093/nar/gku1315

Zhao, J., Zhang, S., Wu, L. Y., and Zhang, X. S. (2012). Efficient methods for identifying mutated driver pathways in cancer. *Bioinformatics* 28, 2940–2947. doi: 10.1093/bioinformatics/bts564

Zhao, M., Kim, P., Mitra, R., Zhao, J., and Zhao, Z. (2016). TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res.* 44, D1023–D1031. doi: 10.1093/nar/gkv1268

Zhou, W., Zhao, Z., Wang, R., Han, Y., Wang, C., Yang, F., et al. (2017). Identification of driver copy number alterations in diverse cancer types and application in drug repositioning. *Mol. Oncol.* 11, 1459–1474. doi: 10.1002/1878-0261.12112