

# Identifying Cancer Driver lncRNAs Bridged by Functional Effectors through Integrating Multi-omics Data in Human Cancers

Yong Zhang,<sup>1,3</sup> Gaoming Liao,<sup>1,3</sup> Jing Bai,<sup>1,3</sup> Xinxin Zhang,<sup>1,3</sup> Liwen Xu,<sup>1</sup> Chunyu Deng,<sup>1</sup> Min Yan,<sup>1</sup> Aimin Xie,<sup>1</sup> Tao Luo,<sup>1</sup> Zhilin Long,<sup>1</sup> Yun Xiao,<sup>1,2</sup> and Xia Li<sup>1,2</sup>

<sup>1</sup>College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China; <sup>2</sup>Key Laboratory of Cardiovascular Medicine Research, Harbin Medical University, Ministry of Education, Harbin, Heilongjiang 150086, China

**The accumulation of somatic driver mutations in the human genome enables cells to gradually acquire a growth advantage and contributes to tumor development. Great efforts on protein-coding cancer drivers have yielded fruitful discoveries and clinical applications. However, investigations on cancer drivers in non-coding regions, especially long non-coding RNAs (lncRNAs), are extremely scarce due to the limitation of functional understanding. Thus, to identify driver lncRNAs integrating multi-omics data in human cancers, we proposed a computational framework, DriverLncNet, which dissected the functional impact of somatic copy number alteration (CNA) of lncRNAs on regulatory networks and captured key functional effectors in dys-regulatory networks. Applying it to 5 cancer types from The Cancer Genome Atlas (TCGA), we portrayed the landscape of 117 driver lncRNAs and revealed their associated cancer hallmarks through their functional effectors. Moreover, lncRNA *RP11-571M6.8* was detected to be highly associated with immunotherapeutic targets (*PD-1*, *PD-L1*, and *CTLA-4*) and regulatory T cell infiltration level and their markers (*IL2RA* and *FCGR2B*) in glioblastoma multiforme, highlighting its immunosuppressive function. Meanwhile, a high expression of *RP11-1020A11.1* in bladder carcinoma was predictive of poor survival independent of clinical characteristics, and *CTD-2256P15.2* in lung adenocarcinoma responded to the sensitivity of methyl ethyl ketone (MEK) inhibitors. In summary, this study provided a framework to decipher the mechanisms of tumorigenesis from driver lncRNA level, established a new landscape of driver lncRNAs in human cancers, and offered potential clinical implications for precision oncology.**

## INTRODUCTION

In the human genome, cancer driver events, which confer a growth advantage to tumor cells, are crucial to tumor initiation, progression, and metastasis. They always destroy key biological pathways, such as P53 and mitogen-activated protein kinase (MAPK) signaling, and in turn disrupt the cell homeostasis.<sup>1</sup> Currently, great efforts on protein-coding drivers have largely deepened the understanding of tumorigenesis, and they have revealed many clinical predictive and targeted biomarkers.<sup>2,3</sup> However, investigations of

cancer drivers in non-coding regions, especially long non-coding RNAs (lncRNAs), are relatively scarce due to the lack of enough functional understanding.

Like protein-coding genes, a great deal of somatic genetic alterations also occurs on lncRNAs, such as copy number alteration (CNA).<sup>4</sup> It's reported that there are about 21.8% of lncRNAs located in regions with focal CNA in 12 major cancer types,<sup>5</sup> and the pervasiveness of CNA on lncRNAs was also confirmed by Zhang et al.<sup>6</sup> Moreover, some lncRNAs have been demonstrated to play driver roles and participate in cancer formation by impacting vital biological processes.<sup>7,8</sup> For instance, lncRNA *FALI*, recurrently amplified in ovarian cancer, represses p21 expression through regulating the stability of epigenetic repressor *BMI1* and then increases cell proliferation.<sup>5</sup> lncRNA *PRAL*, significantly deleted in hepatocellular carcinoma, enhances p53 stability via inhibiting *MDM2*-dependent p53 ubiquitination and further induces cell apoptosis.<sup>9</sup>

Nevertheless, it is still a huge challenge for researchers to distinguish driver lncRNAs from a large amount of passengers. As a representative method, OncodriveFML detects coding and non-coding drivers by estimating the accumulated functional impact bias of tumor somatic mutations based on signals of positive selection.<sup>10</sup> Here, we developed a computational framework, DriverLncNet, to identify driver lncRNAs in human cancer. It could dissect the functional impact of lncRNA CNA events on regulatory networks, and it could capture key functional effectors in dys-regulatory networks. Applying it to 5 cancer types, we presented a landscape of cancer driver lncRNAs, and we characterized their associated cancer hallmarks through functional effectors. Further, we revealed cancer immunity

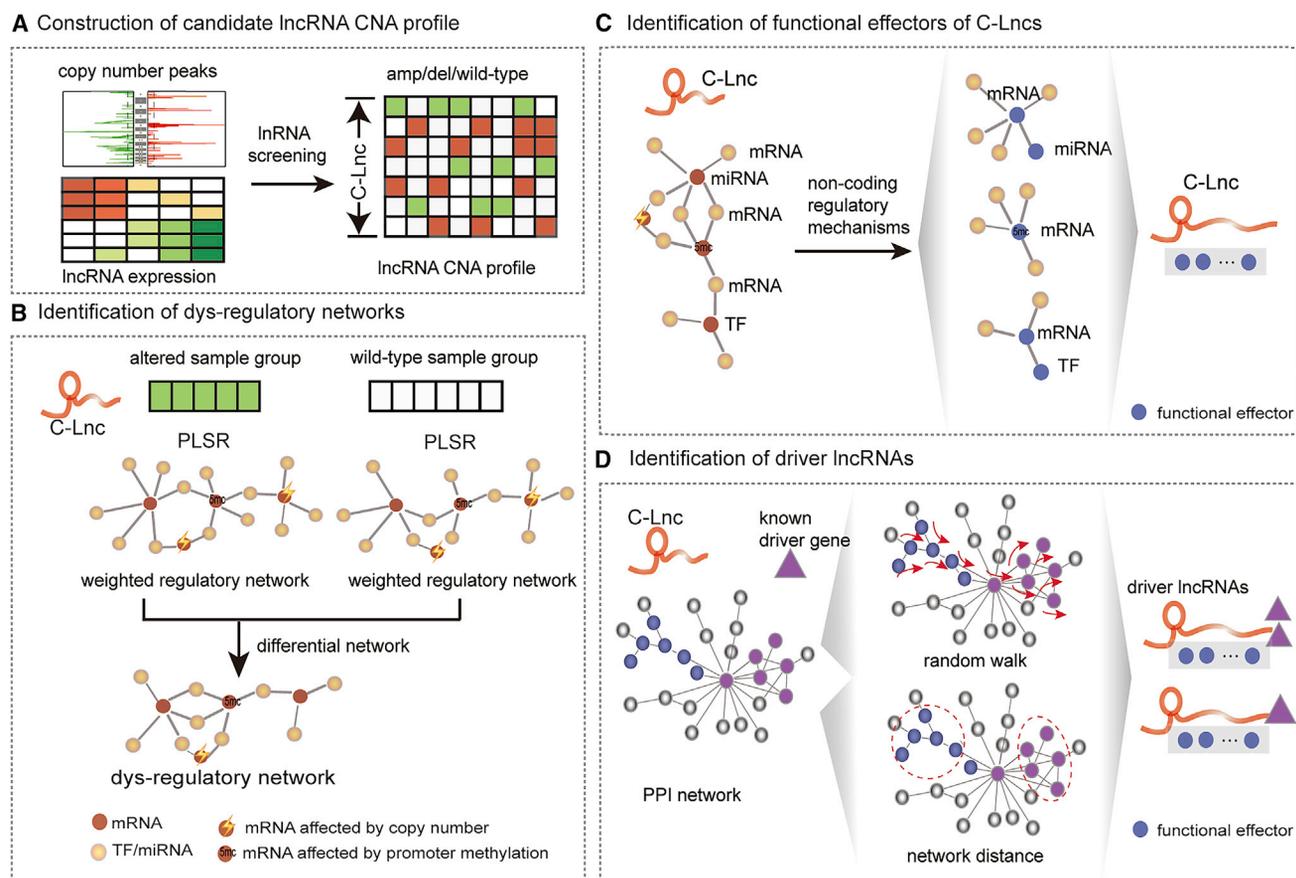
Received 15 February 2019; accepted 15 May 2019;  
<https://doi.org/10.1016/j.omtn.2019.05.030>.

<sup>3</sup>These authors contributed equally to this work.

**Correspondence:** Xia Li, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China.  
**E-mail:** [lixia@hrbmu.edu.cn](mailto:lixia@hrbmu.edu.cn)

**Correspondence:** Yun Xiao, College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang 150086, China.  
**E-mail:** [xiaoyun@ems.hrbmu.edu.cn](mailto:xiaoyun@ems.hrbmu.edu.cn)





**Figure 1. The Overview of DriverLncNet**

(A) Construction of binary copy number profile of candidate driver lncRNAs. (B) Identification of dys-regulatory networks by differential network analysis integrating multi-omics data. (C) Identification of functional effectors of candidate driver lncRNAs based on non-coding regulatory mechanisms. (D) Determination of final driver lncRNAs using random walk and network distance. C-Lnc, PLSR, and PPI networks indicate the candidate driver lncRNAs, partial least-squares regression, and protein-protein interaction network, respectively.

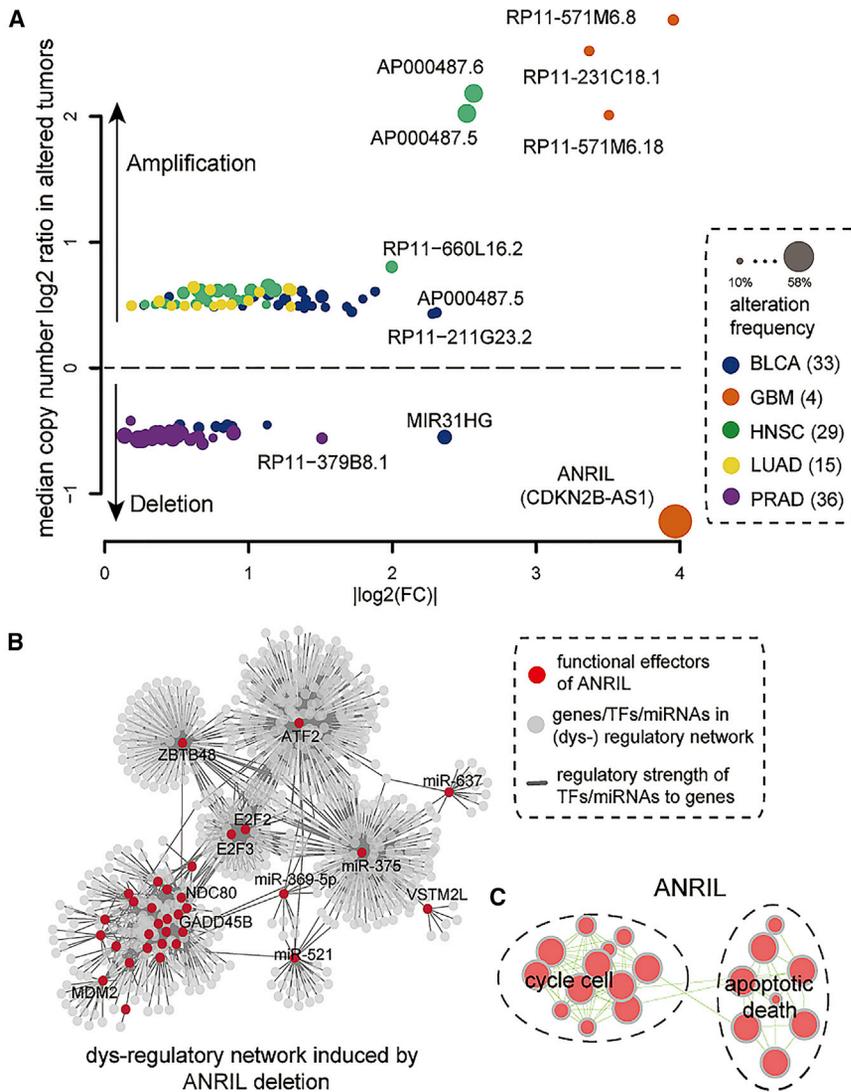
of driver lncRNAs, and we demonstrated their prognostic and drug response potential.

## RESULTS

### The Landscape of Driver lncRNAs in Human Cancers

We proposed a computational framework, DriverLncNet, to identify cancer driver lncRNAs integrating multi-omics data in human cancers (Figure 1; see the Materials and Methods for details). Applying DriverLncNet to 5 The Cancer Genome Atlas (TCGA) cancer types referring to 2,148 patients, we totally identified 117 driver lncRNAs (71 amplified and 46 deleted) with different numbers of functional effectors (Table S5). On average, there were 23 driver lncRNAs per cancer type (ranging from 4 to 36) (Figure S2). These driver lncRNAs were amplified or deleted in 10%–58% of patients, and their copy number level significantly affected their expression ( $p < 0.05$ , one-sided Wilcoxon test; Figure 2A; Table S6). Some have been demonstrated to be cancer drivers through biological experiments, such as *PVT1* and *ANRIL* (also known as *CDKN2B-AS1*) (Table S7).

Notably, *ANRIL* was copy number deleted in 58% of glioblastoma multiforme (GBM). It was located on chromosome 9p21 and was reported to be recurrently deleted in multiple cancer types.<sup>11</sup> *ANRIL* deletion significantly reduced its expression (16-fold decrease,  $p = 3.71e-17$ , one-sided Wilcoxon test; Figure S3A). Finally, we identified 33 functional effectors (4 microRNAs [miRNAs] and 29 genes) of *ANRIL* in the dys-regulatory network (Figure 2B). Among those, *miR-375* and *miR-637* have been demonstrated to be cancer-related miRNAs involved in cell cycle and apoptosis processes.<sup>12</sup> Also, the 29 functional effector genes participated in these processes ( $p < 0.05$ , hypergeometric test; Figure 2C), and they were significantly enriched in known cancer genes ( $p = 0.0073$ , hypergeometric test), supporting the driver role of *ANRIL* in cancer.<sup>11</sup> Also, *ANRIL* was closely associated with two known cancer driver genes, *PTEN* and *PDGFRA*, in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) network, mainly depending on several shared functional effector genes related to cell cycle and apoptosis, such as *MDM2*, *E2F2*, *ATF2*, *NDC80*, and *GADD45B* (Figure S3B). Actually *PTEN* and *PDGFRA* themselves and their functional effectors were



**Figure 2. Driver lncRNAs and Their Functional Effectors in Human Cancers**

(A) Copy number ratio in CNA samples (y axis), influence of copy number on expression (x axis;  $|\log_2(\text{FC})|$ , where  $\text{FC} = \text{mean expression [CNA samples]}/\text{mean expression [wild-type samples]}$ ), and alteration frequency (circle size) for each driver lncRNA. (B) The functional effectors in dys-regulatory network induced by *ANRIL* deletion in GBM. (C) Enrichment map for functional effectors of *ANRIL*.

$p = 0.031$ , and  $p = 0.0027$ , one-sided Wilcoxon test; Figure S5), suggesting underlying purifying selection of driver lncRNAs.<sup>16,17</sup> Also, driver lncRNAs showed significantly stronger pathogenicity and functionality ( $p = 6.646e-5$ ,  $p = 1.74e-6$ , and  $p = 0.0129$ , one-sided Wilcoxon test; Figure S6) using three popular methods, Funseq2,<sup>16</sup> EIGEN,<sup>18</sup> and CADD<sup>19</sup> (see the Supplemental Materials and Methods for details). Taken together, these driver lncRNAs form a new landscape of non-coding RNAs in human cancers, which is worthy of further functional characterization and experimental validation.

### Driver lncRNAs Contribute to Cancer Hallmarks through Functional Effectors

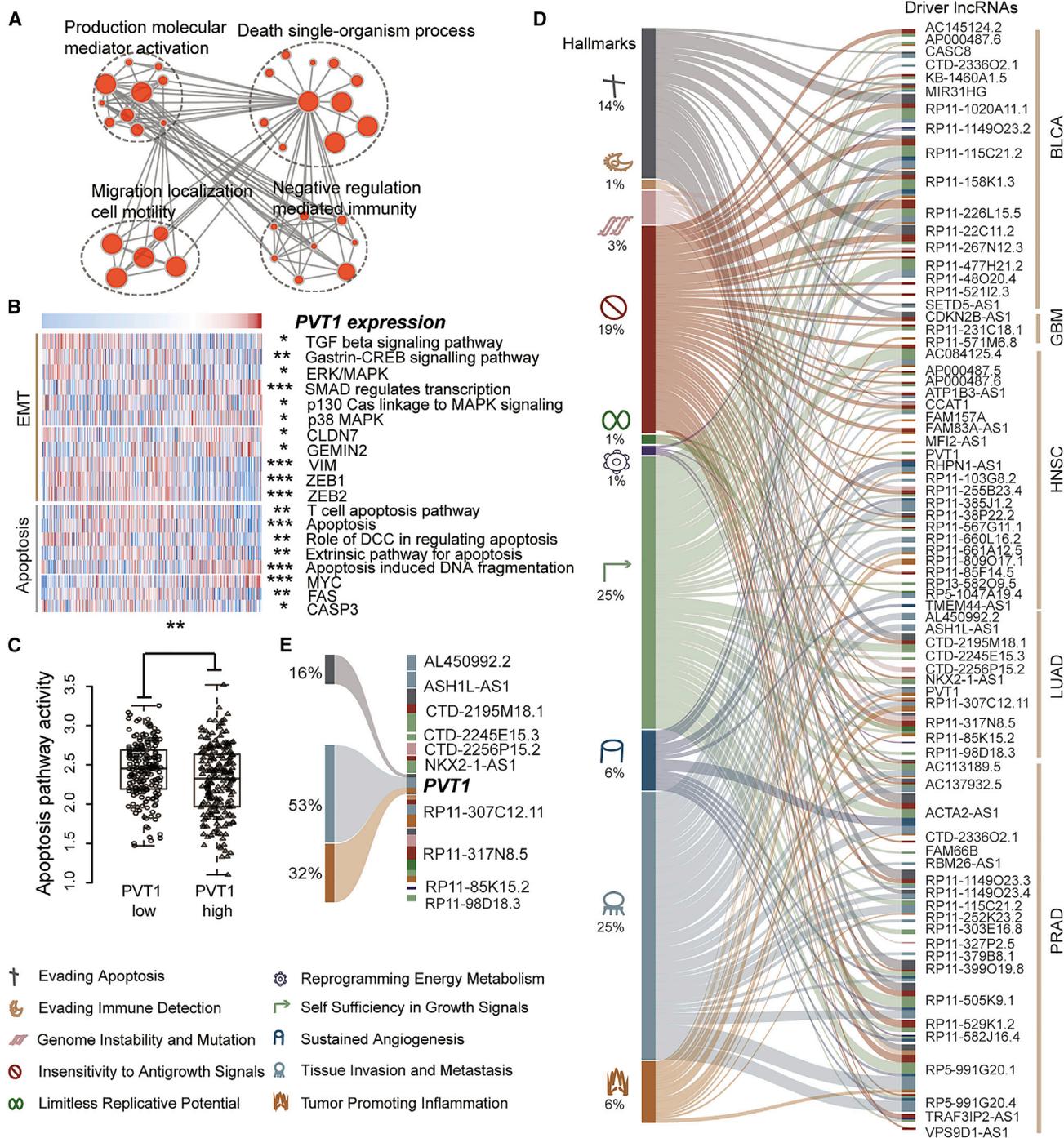
Our framework allowed us to identify driver lncRNAs and their functional effectors, which can greatly help us to further characterize the functional phenotypes of driver lncRNAs. To clarify the effectiveness of functional effectors, we took a known driver lncRNA *PVT1* in lung adenocarcinoma (LUAD) as an example for detailed illustration. Through a small interfering RNA (siRNA) knockdown experiment<sup>20</sup>

and CRISPR interference technology,<sup>21</sup> we utilized gene set enrichment analysis (GSEA) for functional effectors of *PVT1*, and we found that these functional effectors were significantly enriched in differentially expressed genes after perturbation of *PVT1* (Figures S7A and S7B), which indicated that *PVT1* can significantly influence these functional effectors. Furthermore, we observed that *PVT1* functional effectors were enriched in the functions of migration and death process (Figure 3A). We thus speculated that *PVT1* might be involved in the cancer hallmarks of metastasis and apoptosis.

To confirm this speculation, we calculated the pathway activity of the epithelial-mesenchymal transition (EMT) and apoptosis-associated pathways from The Molecular Signatures Database (MSigDB) using a popular method.<sup>22</sup> As a result, the activities of these pathways were significantly correlated with *PVT1* expression, such as transforming growth factor  $\beta$  (TGF- $\beta$ )-signaling pathway and apoptosis ( $p < 0.05$ ,  $p < 0.001$ , two-sided Wilcoxon test; Figures 3B and 3C).

also involved in cell cycle and apoptosis ( $p < 0.05$ , hypergeometric test; Figure S3C).

Furthermore, we assessed whether these driver lncRNAs exhibited similar properties with known cancer genes. Accumulating evidence has revealed that cancer genes tend to be at early stages of replication timing, which shaped the landscape of genetic alterations in the cancer genome.<sup>13</sup> Like known cancer genes, driver lncRNAs showed significantly earlier replicating time than others using Repli-seq data from Lawrence et al.<sup>14</sup> and 5 cancer cell lines in University of California, Santa Cruz (UCSC;  $p = 3.91e-10$ ,  $p = 1.4e-7$ , one-sided Wilcoxon test; Figure S4; see the Supplemental Materials and Methods for details). Through analyzing phastCons<sup>15</sup> conservation data from UCSC and variant call format file from the 1000 Genomes project, we observed that driver lncRNAs had higher exon conservation, higher fraction of rare SNPs, lower SNP density, and lower derived allele frequency (DAF) ( $p = 4.697e-5$ ,  $p = 0.0886$ ,



**Figure 3. Cancer Hallmarks and Driver IncRNAs with Functional Effectors**

(A) Enrichment map of functional effectors of *PVT1* using an enrichment tool gProfiler ( $p \leq 0.05$ ). The size of the red circle is proportional to the size of a functional gene set. (B) EMT- and apoptosis-associated pathways and molecular markers correlated to *PVT1* expression (blue for low expression or pathway activity, red for high; Pearson's correlation,  $*p \leq 0.05$ ,  $**p \leq 0.01$ ,  $***p \leq 0.001$ ). (C) Apoptosis pathway activity in *PVT1* low- and high-expression groups (two-sided Wilcoxon test). (D) BiPartite graph in five tumor types. The curve between a cancer hallmark and a driver IncRNA indicates that the IncRNA is associated with the cancer hallmark. The percent under the icon of cancer hallmark indicates the fraction of GO terms of each hallmark. (E) Three cancer hallmarks of driver IncRNA *PVT1*.

Also, some known EMT- and apoptosis-related markers,<sup>23,24</sup> such as *VIM*, *ZEB1/2*, and *MYC*, showed significant correlations with *PVT1* ( $p < 0.05$ , two-sided Wilcoxon test; Figure 3B). These findings were consistent with previous studies that *PVT1* could suppress cell apoptosis, migration, and invasion in non-small-cell lung cancer cells,<sup>25,26</sup> suggesting that functional effectors can effectively capture the functional mechanisms of driver lncRNAs.

Hence, we sought to characterize the cancer hallmark landscape of driver lncRNAs using their functional effectors in five tumor types. In brief, a driver lncRNA was considered to be associated with a specific cancer hallmark conservatively if its functional effectors were significantly enriched in hallmark gene ontology (GO) terms (Tables S8 and S9; see the Supplemental Materials and Methods for details). As a result, 74 of 117 (63.2%) driver lncRNAs were associated with various cancer hallmarks (Figure 3D). On average, more than two cancer hallmarks were found to be linked with a specific driver lncRNA, such as *PVT1* in LUAD linked with three cancer hallmarks, including evading apoptosis, tissue invasion and metastasis, and tumor-promoting inflammation (Figure 3E). These results indicated that the aberrant lncRNAs can widely influence cancer hallmarks through their functional effectors and, consequently, contribute to tumor development and progression.

#### Driver lncRNAs Are Involved in the Tumor Immune Microenvironment

We next focused on 17 driver lncRNAs related to the immune hallmarks evading immune detection and tumor-promoting inflammation. To characterize whether these lncRNAs were associated with an immunophenotype, we obtained 39 immunosuppressive genes comprising an immune signature from previous studies (Table S10). The 17 lncRNAs were strongly associated with these immunophenotype-related genes, with coincident patterns (Figure 4A). Through assessing the immunosuppressive score based on the signature using gene set variation analysis (GSVA),<sup>27</sup> we observed that 88.2% (15/17) of these lncRNAs showed significant associations with immunosuppressive scores ( $p < 0.05$ , two-sided Wilcoxon test; Figure 4B). For instance, patients with a high expression of lncRNA *RP11-571M6.8* in GBM had lower immunosuppressive scores than others ( $p < 0.001$ , two-sided Wilcoxon test; Figure 4C). As targets of cancer immunotherapy, immune checkpoint proteins *PD-1*, *PD-L1*, and *CTLA4*<sup>28</sup> also showed lower expression in patients with a high expression of *RP11-571M6.8* ( $p = 0.035$ ,  $p = 0.032$ , and  $p = 0.035$ , respectively; two-sided Wilcoxon test; Figure 4D).

Next, we explored the relationship between lncRNAs and immune cell infiltration that determined whether tumor cells would escape from immune-mediated destruction successfully.<sup>29</sup> We obtained three immune cell infiltration profiles of five immune cell types (including regulatory T [Treg] cells, macrophages, CD8+ T cells, natural killer [NK] cells, and neutrophils) from previous studies.<sup>30–32</sup> We found that 72.2% of lncRNAs were involved in different immune cell infiltrations in at least one dataset ( $p < 0.05$ , two-sided Wilcoxon test; Figure 4E). Among those, a high expression of *RP11-571M6.8* in

GBM was highly related to the exhaustion of Treg cells, which could suppress the immune response of cancer, in all three datasets ( $p < 0.05$ , two-sided Wilcoxon test; Figure 4F; Figure S8). Moreover, two Treg cell markers, *CD25* and *FCGR2B*, whose downregulation could indicate the exhaustion of Treg cells in the immune response,<sup>33,34</sup> showed significantly negative correlations with *RP11-571M6.8* expression ( $R = 0.269$ ,  $p = 0.001$ ;  $R = 0.349$ ,  $p < 0.001$ , respectively, Pearson's correlation test; Figure 4G). These results consistently reflected that a high expression of *RP11-571M6.8* was related to low immunosuppression and exhaustion of Treg cells, indicating an enhanced cancer immunity. Taken together, our results revealed the close connections of immune-related driver lncRNAs with the immunoregulation of cancer, which may be effective markers for immunotherapy.

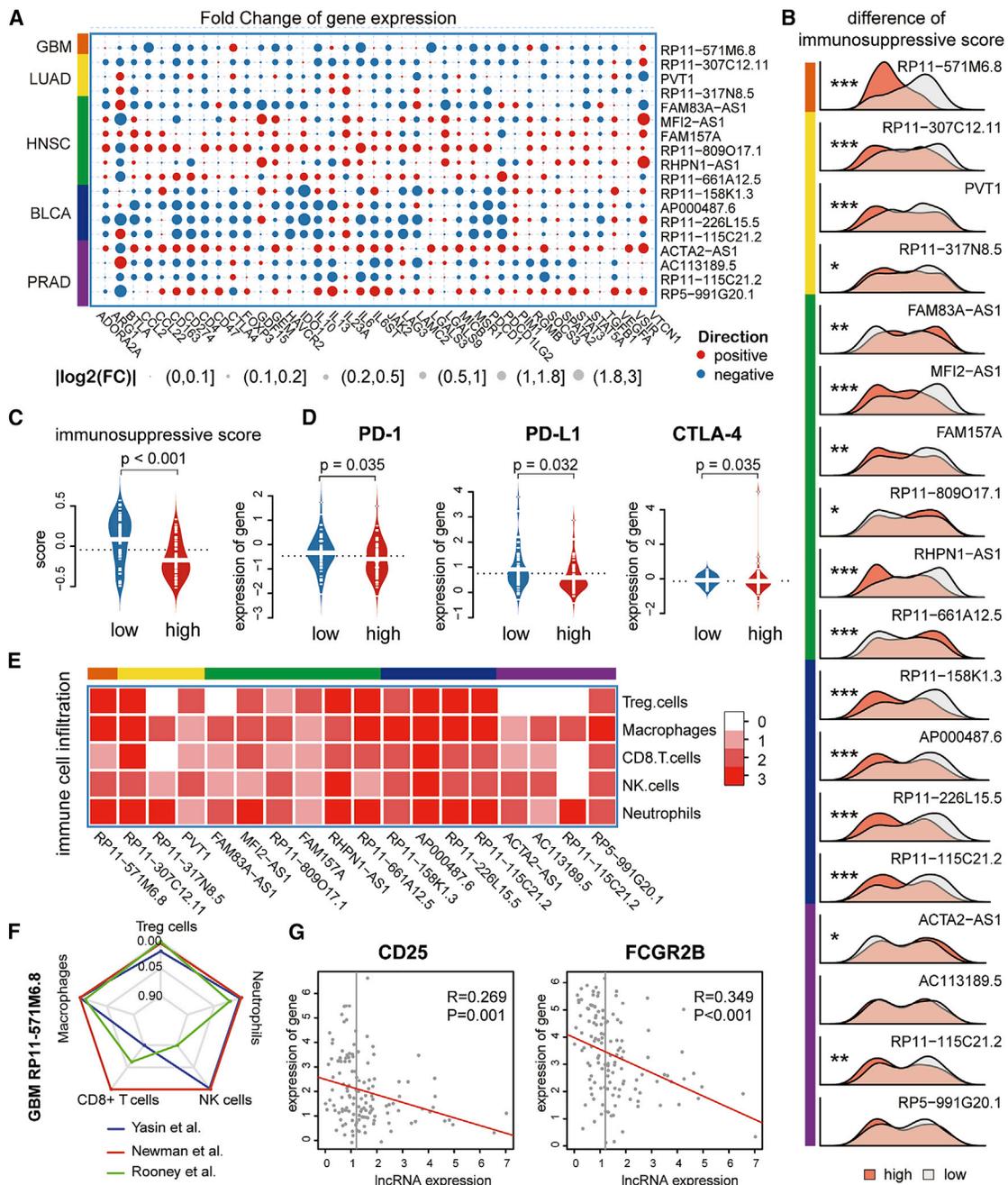
#### Clinical Prognosis of Driver lncRNAs

We wondered whether some driver lncRNAs had clinical prognostic implications. Since prostatic adenocarcinoma (PRAD) in TCGA had few dead or disease progression samples, we focused our attention on the other four cancer types. Through performing Kaplan-Meier and multivariate Cox proportional hazard regression analyses, we identified 12 driver lncRNAs that were significantly predictive of overall survival (OS), 4 of which were also significantly predictive of disease-free survival (DFS) (Tables S11 and S12).

For example, driver lncRNA *RP11-1020A11.1* was significantly predictive of OS and DFS in bladder carcinoma (BLCA) ( $p = 0.0228$ ,  $p = 0.0285$ , log rank test; Figure 5A). *RP11-1020A11.1*, located on chromosome 3p25, was recurrently amplified in BLCA, and its expression in amplified tumors was significantly higher compared to diploid tumors ( $p = 6.48 \times 10^{-8}$ , Wilcoxon test; Figure 5B). Also, functional enrichment analysis of its functional effectors showed that *RP11-1020A11.1* participated in cell proliferation, cell cycle, cell migration, and apoptosis ( $p = 0.0073$ , hypergeometric test; Figure 5C). Under Cox regression analyses with age at diagnosis, gender, and tumor stage as clinical covariates, we found that high *RP11-1020A11.1* expression was a favorable factor for clinical outcome (hazard ratio [HR] = 0.29, 95% confidence interval [CI] = 0.14–0.57,  $p = 0.0004$  for OS; HR = 0.53, CI = 0.27–1.03,  $p = 0.069$  for DFS; Figure 5D), independent of clinical features. Interestingly, we found that two other driver lncRNAs, *SETD5-AS1* and *RP11-380O24.1*, recurrently amplified in 3p25 in BLCA, had similar results to *RP11-1020A11.1* (Table S12), which suggested that the amplification of 3p25 may serve as a potential favorable prognostic biomarker in BLCA. Another driver lncRNA, *FAM83A-AS1* in LUAD, was significantly associated with patient OS and DFS independent of clinical features, such as age at diagnosis, gender, and tumor stage (Figures 5E–5H).

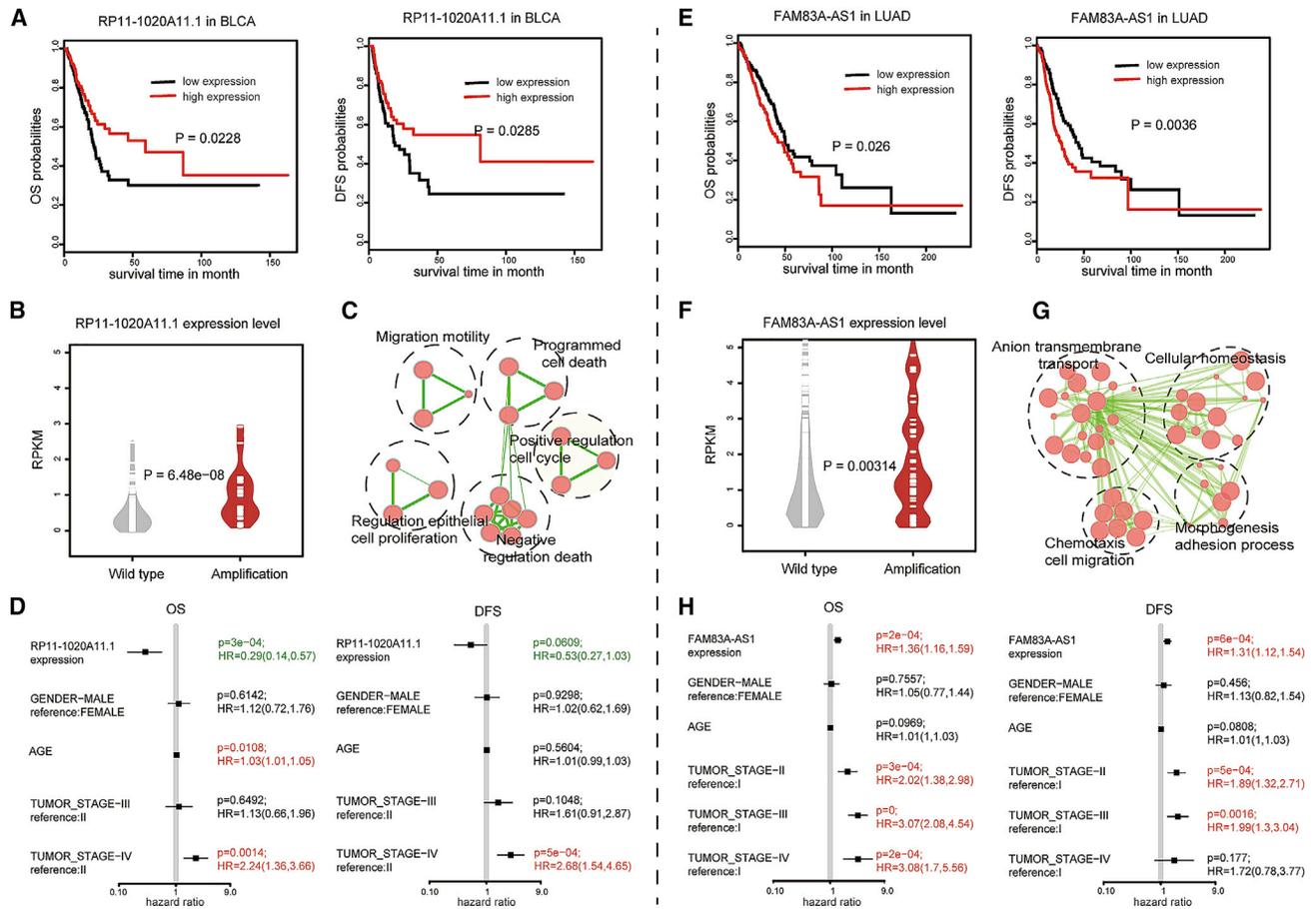
#### Anti-cancer Drug Sensitivity of Driver lncRNAs in LUAD

To explore the potential effects of driver lncRNAs on drug response, we focused on the 15 driver lncRNAs in LUAD, and we evaluated whether their CNA could influence drug response across 714 preclinical cell models from the Cancer Genome Project<sup>35</sup> (CGP). We found



**Figure 4. Cancer Immunity of Driver lncRNAs**

(A) Fold change (FC) of mean gene expression in patients with high or low expression of each lncRNA across 5 cancer types. The size of the red (or blue) circle is proportional to the positive (or negative) log-fold change between patients with high or low expression. (B) Distribution of immunosuppressive scores in patients with high (red) or low (gray) expression of each lncRNA ( $p \leq 0.05$ ,  $**p \leq 0.01$ ,  $***p \leq 0.001$ , two-sided Wilcoxon test). (C) Immunosuppressive score in patients with high (red) or low (blue) expression of *RP11-571M6.8* in GBM. (D) *PD-1*, *PD-L1*, and *CTLA4* expressions in patients with high (red) or low (blue) expression of *RP11-571M6.8* in GBM. (E) Immune cell infiltration level and immune-related driver lncRNAs in 3 independent datasets across 5 cancer types (two-sided Wilcoxon test for immune cell infiltration level of patients with high and low expression,  $p \leq 0.05$ ). (F) Radar chart of 5 immune cells' infiltrations in patients with high or low expression of *RP11-571M6.8* in GBM in 3 independent datasets (two-sided Wilcoxon test). (G) The correlation between expression of Treg cell markers and *RP11-571M6.8* expression in GBM.



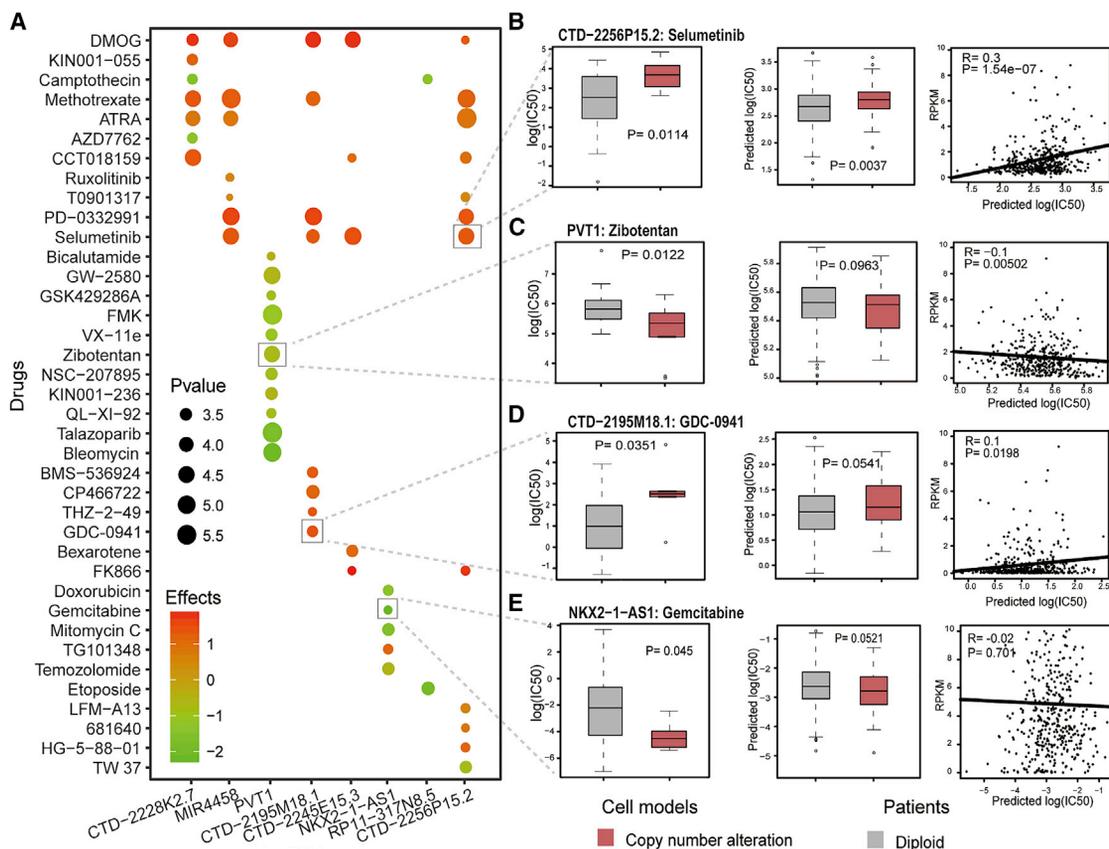
**Figure 5. The Prognosis Value of Driver lncRNAs**

(A and E) KM survival curve for tumor patients with high and low expressions of (A) *RP11-1020A11.1* or (E) *FAM83A-AS1* for OS and DFS in BLCA. (B and F) Distribution of (B) *RP11-1020A11.1* or (F) *FAM83A-AS1* expression in diploid and amplified patients. (C and G) Enrichment map for functional effectors of (C) *RP11-1020A11.1* or (G) *FAM83A-AS1*. (D and H) Results of multivariable cox regression model for OS and DFS in (D) BLCA or (H) LUAD. Red and green indicate risk factor and favorable factor, respectively.

that multiple driver lncRNAs presented strong correlations with the responses of drugs that are associated with lung cancer and other cancer treatments (Figure 6A). For example, driver lncRNA *CTD-2256P15.2* amplification can significantly enhance the resistance of selumetinib ( $p = 0.01$ , two-sided Wilcoxon test; Figure 6B, left), a methyl ethyl ketone (MEK) inhibitor, which can block the MAPK kinase and reduce the activity of the MAPK-ERK pathway.<sup>36</sup> Meanwhile, lncRNA *PVT1* copy number amplification enhanced the sensitivity of zibotentan ( $p = 0.01$ , two-sided Wilcoxon test; Figure 6C, left). Other anti-tumor drugs, such as GDC-0941 and gemcitabine, were associated with *CTD-2195M18.1* and *NKX2-1-AS1*, respectively (Figures 6D and 6E).

To further affirm the relationship between driver lncRNAs and anti-tumor drugs in patients, we built a predicted spectrum of drug response for LUAD patients using drug response data in the CGP through ridge regression.<sup>37</sup> We observed that the patients with *CTD-2256P15.2* amplification or high expression showed strong selumetinib resistance ( $p = 3.7e-3$ , two-sided Wilcoxon test for copy

number;  $R = 0.3$ ,  $p = 1.54e-7$ , Spearman rank correlation test for expression; Figure 6B, middle and right), while *PVT1* amplification or high expression showed strong zibotentan sensitivity ( $p = 9.63e-2$ , two-sided Wilcoxon test for copy number;  $R = -0.1$ ,  $p = 5.02e-3$ , Spearman rank correlation test for expression; Figure 6C, middle and right), which supported the results in cell models. Especially, similar results for *CTD-2256P15.2* were confirmed using another resource, the Cancer Therapeutics Response Portal<sup>38</sup> (CTRP version [v.2]). ( $p = 1.51e-3$  for copy number;  $R = 0.3$ ,  $p = 5.35e-10$  for expression; Figure S9A). Also, the functional effectors of *CTD-2256P15.2* were correlated with selumetinib response through GSEA (false discovery rate [FDR] =  $3.3e-2$ ; Figure S10A). In addition, *CTD-2256P15.2* amplification can mildly enhance resistance of another MEK inhibitor, trametinib, in CGP ( $p = 0.06$ , two-sided Wilcoxon test; Figure S9B, left), which was certified by its functional effectors through GSEA (FDR =  $2e-2$ ; Figure S10B) and predicted spectrum ( $p = 4.73e-4$  for copy number;  $R = 0.4$ ,  $p = 2.92e-14$  for expression in CGP; Figure S9B middle and right;  $p = 2.55e-2$  for copy number;  $R = 0.2$ ,  $p = 3.02e-6$  for expression



**Figure 6. The Association of Driver lncRNAs and Anti-cancer Drugs**

(A) Correlation between drug half-maximal inhibitory concentration ( $IC_{50}$ ) and lncRNA copy number status in CGP. Dot size is proportional to the p value of correlation; dot color indicates the drug response effect (drug resistance, red; and sensitivity, green) in CNA samples compared to diploid for a specific lncRNA. (B) Selumetinib has a high  $IC_{50}$  in samples with *CTD-2256P15.2* amplification in cell models (left) and predicted spectrum (middle and right). (C) Copy number status of lncRNA *PVT1* and drug zibotentan. (D) lncRNA *CTD-2195M18.1* and drug GDC-0941. (E) lncRNA *NKX2-1-AS1* and drug gemcitabine.

in CTRP; Figure S9C). Our results suggested that the driver lncRNAs could reflect the patterns of drug response and may be further investigated as markers for future drug guidance.

## DISCUSSION

To date, the identification of cancer drivers from the non-coding regions is quite challenging, due to the lack of enough functional understanding. In this study, we developed an integrated computational framework, DriverLncNet, for identifying driver lncRNAs and their functional effectors in human cancers. Applying it to 2,148 patients from 5 cancer types, we identified 117 driver lncRNAs, and we characterized their affected cancer hallmarks through analyzing their functional effectors. These driver lncRNAs shed new insights into molecular mechanisms, such as cancer immunity, and they provided novel prognostic and drug response potential for clinical practice. To our knowledge, this study represents the most extensive investigation of driver lncRNAs, and it proves that the integration of multi-omics data enables us to discover novel driver molecules and their dys-regulatory mechanisms during tumorigenesis.

Cancer genes generally induce deregulation of their functional effectors and exert driver roles in cancer. DriverLncNet utilizes dys-regulatory networks induced by lncRNA CNA events and regulatory principles of lncRNAs to identify functional effectors of lncRNAs. Our framework demonstrates that key functional nodes are quite effective at distinguishing cancer drivers from a large amount of passengers. Actually, through such a way, many known driver events are reported to deliver their functional influence from the genomic level to the functional level and, in turn, contribute to cancer hallmarks.<sup>39</sup> For example, tumor suppressor gene *TP53* could promote cell-cycle arrest by activating p21 expression in gliomas.<sup>40</sup>

In our work, *miR-195*, *miR-136*, *POU2F2*, and *WFDC2*, as key functional effectors of *PVT1* in LUAD, are involved in cell invasion and apoptosis (Table S13), which reconcile the biological functions of *PVT1*. Among them, *miR-195* is reported to directly interact with *PVT1* in multiple cancers through miRNA sponge and involved in tumor cell invasion and apoptosis. Notably, this strategy is quite dependent on the understanding of regulatory or interaction

mechanisms of lncRNAs. With the rapid advance of high-throughput technologies, more interaction partners with lncRNAs will be detected, such as DNA elements, RNAs, and proteins, which will enhance our method to further optimize functional effectors, greatly benefiting our understanding of the functional influence of driver lncRNAs.

Another advantage of our framework is determining novel non-coding drivers by capturing the functional association with known cancer driver genes through their functional effectors. Actually, tumor evolution is a process of somatic mutation and natural selection.<sup>41</sup> During this process, several driver events form evolutionary dependence, and they exhibit different combination mutational patterns to drive cancer formation. These evolutionary dependency drivers are always highly functionally associated, such as participating in similar biological processes and mediating pathway crosstalk, and they corporately promote clonal expansion or selective sweep.<sup>42–44</sup> To confirm that candidate lncRNAs indeed act as cancer drivers, we utilize network similarity to capture the potential functional connection with known cancer driver genes, which makes our framework more powerful.

In summary, our study presents a comprehensive landscape of cancer driver lncRNAs, which serves as a resource to extend our knowledge of non-coding driver events in cancer. Our analyses shed new insights into molecular mechanisms of lncRNAs underlying tumorigenesis, and they offer implications for prognosis prediction and drug selection strategies. With the large accumulation of whole-genome sequencing data, we expect our method to be applicable to the discovery of other non-coding drivers.

## MATERIALS AND METHODS

### Data Source

For 5 cancer types (including GBM, BLCA, PRAD, LUAD, and head and neck cancer [HNSC]), we obtained copy number (level 3), mutation (level 2), DNA methylation (level 3), gene expression (level 3), and miRNA expression (level 3) data, as well as clinical data from TCGA project. lncRNA expression data from TCGA was collected from The Atlas of ncRNA in Cancer (TANRIC).<sup>45</sup> The relationship of transcription factor (TF)-targeting mRNA was from Transfac, UCSC, and ChIPbase, while that of miRNA-regulating mRNA was from miTarbase and Starbase (see the [Supplemental Materials and Methods](#) for details; [Tables S1](#) and [S2](#)).

### DriverLncNet: A Network-Based Framework to Identify Driver lncRNAs

In general, cancer drivers abnormally regulate key molecules (termed as functional effectors) and, in turn, impact downstream signal pathways and networks, contributing to cancer formation<sup>7,8</sup> (see the [Supplemental Materials and Methods](#) for details; [Figure S1](#)). To identify driver lncRNAs in human cancers, we proposed a computational framework, DriverLncNet, through dissecting the functional impact of lncRNA CNA events by integrating multi-omics data.

### Building Binary CNA Profile of lncRNAs

Based on the copy number data, we built a binary CNA profile for potential driver lncRNAs ([Figure 1A](#)). Potential driver lncRNAs were selected through four filtering criteria: (1) lncRNAs should have a dominant CNA type (amplification or deletion,  $p \leq 0.05$ , binomial test) (see the [Supplemental Materials and Methods](#) for details); (2) for lncRNA expression, we used same threshold with TANRIC, i.e., lncRNAs should have detectable expression, which was defined as those with an average reads per kilobase per million mapped reads (RPKM)  $\geq 0.3$  across all samples; (3) lncRNAs should have concordant changes between CNA and expression; and (4) lncRNAs should harbor CNA in at least 10 samples, for robustly measuring the impact of CNA on regulatory networks in subsequent analysis.

### Constructing Regulatory Networks Associated with CNA and Wild-Type lncRNAs

To construct dys-regulatory networks associated with CNA of each lncRNA, we first used linear regression models to build two regulatory networks ([Figure 1B](#), top). One was constructed using CNA samples of this lncRNA, and the other was constructed using wild-type samples. Specially, in a given patient group, a linear regression model for each gene was constructed to explain gene expression variance. Only outlying genes (that is, the genes showing differential expression between cancer and normal samples and having high expression variability across cancer samples) were used to construct linear models (see the [Supplemental Materials and Methods](#) for details). In these models, multi-layer regulatory factors, including continuous DNA copy number level, promoter methylation level, and the expression of TFs and miRNAs regulating the corresponding genes, were included as covariates.

Given a gene  $G_i$  in a specific group containing  $N$  samples, there are  $J$  TFs ( $j = 1, 2, \dots, J$ ) and  $K$  miRNAs ( $k = 1, 2, \dots, K$ ) regulating  $G_i$ . A linear regression model is trained as

$$\exp_{G_i} \approx \beta_{CN} CN_{G_i} + \beta_{meth} meth_{G_i} + \sum_1^J \beta_{TF_j} \exp_{TF_j} + \sum_1^K \beta_{miRNA_k} \exp_{miRNA_k},$$

where  $\exp_{G_i}$  is the expression level of  $G_i$ ,  $CN_{G_i}$  is the copy number level of  $G_i$ ,  $meth_{G_i}$  is the promoter methylation level of  $G_i$ ,  $\exp_{TF_j}$  is the expression level of the  $j$ th TF regulating  $G_i$ , and  $\exp_{miRNA_k}$  is the expression level of the  $k$ th miRNA targeting  $G_i$ .  $\beta_{CN}$ ,  $\beta_{meth}$ ,  $\beta_{TF_j}$ , and  $\beta_{miRNA_k}$  represent regression coefficients of  $CN_{G_i}$ ,  $meth_{G_i}$ ,  $\exp_{TF_j}$ , and  $\exp_{miRNA_k}$ , respectively.

Considering the large number of variables and their possibly high collinearity, partial least-squares regression (PLSR) model was adopted.<sup>46</sup> Then, the significance of the effect of regulatory factors on gene expression of each gene was estimated using 10-fold cross-validation through the functions (plsr, RMSEP, and jack.test) in R package pls. Subsequently, we extracted statistically significant regulatory factors (copy number, promoter methylation, TFs, and miRNAs) with significant regression coefficients, which was adjusted by the Benjamini-Hochberg procedure to control the FDR  $\leq 0.05$ . Then, all regulatory relationships were integrated to form a regulatory network, in which

the weights of edges were the regression coefficients. Finally, for each candidate lncRNA, two regulatory networks were built separately using CNA samples of this lncRNA and wild-type samples. In addition, we filtered out some regulatory relationships to obtain regulatory networks with biological significance<sup>47</sup> (see the [Supplemental Materials and Methods](#) for details).

#### Constructing Dys-regulatory Networks Induced by the CNA of Candidate Driver lncRNAs

For each candidate driver lncRNA, we compared the two regulatory networks and got a dys-regulatory network using different network analysis with DiffK algorithm<sup>48</sup> (Figure 1B, bottom). Specifically, given a node  $v$  in a specific regulatory network, we set the node weight as the sum of the absolute edge weights connecting the node. The formula was shown as

$$S_{DiffK}^X(v) = \sum_{u \in X} w(u, v),$$

where  $X$  is direct neighbors of node  $v$ , and  $w(u, v)$  is the absolute weight of the edge between nodes  $u$  and  $v$ . Then, the absolute difference of the scaled weight of the corresponding node in two regulatory networks was calculated as follows:

$$DiffK(v) = \left| \frac{S_{DiffK}^A(v)}{\max_{u \in A} (S_{DiffK}^A(u))} - \frac{S_{DiffK}^B(v')}{\max_{u' \in B} (S_{DiffK}^B(u'))} \right|,$$

where  $A$  and  $B$  are direct neighbors of node  $v$  and  $v'$  in two regulatory networks. To determine the significance of the absolute weight differences for each node, the genetic alteration profile was randomly permuted 1,000 times using Bioconductor package BiRewire,<sup>49</sup> preserving alteration frequency of candidate driver lncRNAs and samples. Then, 1,000 random dys-regulatory networks were built, and the corresponding absolute weight differences were calculated. The significance of the absolute weight difference was calculated as the fraction of the 1,000 random permutations in which the random weight difference was greater than that observed in the actual data. Only the nodes with  $p$  value  $\leq 0.05$  were regarded as dys-regulatory factors.

#### Identifying Functional Effectors of Candidate Driver lncRNAs

Considering the fact that lncRNAs could regulate gene expression through several molecular mechanisms, such as competitive endogenous RNA (ceRNA)-based regulation and chromatin remodeling by recruiting chromatin complexes (see the [Supplemental Materials and Methods](#) for details), we utilized three criteria to identify key dys-regulatory factors for each candidate lncRNA (Figure 1C). For a candidate driver lncRNA, those dys-regulatory factors (1) forming miRNA response element (MRE) competitive relationships from a known ceRNA database lncCeDB<sup>50</sup> with this lncRNA or (2) co-expressing with this lncRNA ( $p \leq 0.05$ , Pearson's correlation test) were retained. In addition, (3) if the promoter methylation levels of dysregulated factors were significantly correlated with this lncRNA expression ( $p \leq 0.05$ , Pearson's correlation test), these mRNAs

were retained. These retained dysregulated factors were referred to as functional effectors of the lncRNA.

#### Detecting Driver lncRNAs through Network Analysis

Like cancer driver genes, driver lncRNAs also can provide tumor cells with a growth advantage, thus contributing to tumor initiation, progression, or metastasis. So we supposed that driver lncRNAs and driver genes should have closely functional associations on biological network (Figure 1D). We obtained high-confident cancer driver genes and then identified their functional effectors (see the [Supplemental Materials and Methods](#) for details; Table S3). Through mapping functional effectors of candidate driver lncRNAs and known cancer driver genes on the STRING network,<sup>51</sup> we measured the functional association between candidate driver lncRNAs and cancer driver genes by leveraging two strategies (random walk and network distance).

For random walk,<sup>52</sup> the functional effectors of a cancer driver gene were as seed nodes and the functional effectors of a candidate driver lncRNA were as response nodes. The formula was shown as

$$p_{t+1} = (1 - \alpha)wp_t + \alpha p_0,$$

where  $p_0$  is the vector of initial probabilities of genes in the network (the probability of the seed nodes are absolute Pearson's correlation coefficient between lncRNA expression and seed node expression and others are 0);  $p_t$  and  $p_{t+1}$  are the probabilities of the nodes at the  $t$ th and  $(t + 1)$ th steps, respectively;  $w$  represents the matrix of edge weight  $w_{ij}$ ; and  $\alpha$  is the restart probability of 0.3. If the maximum difference between  $p_{t+1}$  and  $p_t$  is less than  $10e-8$ , the random walk reaches the steady state. After random walk, every node in the network gets a probability  $P$  to reflect its association with seed nodes.

To measure the association of response nodes and seed nodes, we designed a statistic as follows:

$$a\_score = \sum_{m=1}^M \frac{1}{R_{p_m}},$$

where  $M$  denotes the number of functional effectors ( $m = 1, 2, \dots, M$ ) of the lncRNA,  $P_m$  denotes probability of functional effector  $m$  after random walk,  $R_{p_m}$  denotes the rank of  $P_m$  in all nodes in the decreasing order, and  $a\_score$  is the association score of the candidate driver lncRNA and the cancer driver gene.

For network distance, we used a pre-defined statistic<sup>53</sup> to integrally measure network association between two gene sets. The functional effectors of a cancer driver gene and a candidate driver lncRNA were treated as two gene sets,  $A$  and  $B$ . The formula was shown as

$$dis_{AB} = d_{AB} - \frac{d_{AA} + d_{BB}}{2},$$

where  $d_{AB}$  (or  $d_{BA}$ ) is the mean shortest distance of nodes in  $A$  (or  $B$ ) with nodes in  $B$  (or  $A$ ),  $d_{AB} = d_{BA}$ ;  $d_{AA}$  ( $d_{BB}$ ) is the mean shortest distance of nodes in  $A$  (or  $B$ ) with other nodes in  $A$  (or  $B$ ); and  $dis_{AB}$  is

the network distance of A and B. All weights  $w_{ij}$  are transformed to  $w_{ij}^2$  so that higher-confidence edges imply shorter paths. The distance of neighboring nodes is the transformed matrix of edge weight.

To determine the significance of functional association between the cancer driver gene and the candidate driver lncRNA, we randomly selected the same count of functional effectors 1,000 times, and we calculated 1,000 random statistics for two strategies. If the real statistic was less than 5% of frequency of random statistics, the lncRNA was considered to be significantly associated with the cancer driver gene. Combining with the two strategies, the lncRNAs that were associated with at least one cancer driver gene were identified as driver lncRNAs (Tables S4 and S5).

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.omtn.2019.05.030>.

## AUTHOR CONTRIBUTIONS

X.L. and Y.X. provided the scientific idea and designed the study. Y.Z., G.L., J.B., and X.Z. performed all data analysis, designed the figures, and drafted the manuscript. L.X., C.D., M.Y., A.X., T.L., and Z.L. performed preliminary data processing and modified the manuscript. All authors read and approved the final manuscript.

## CONFLICTS OF INTEREST

The authors declare no competing interests.

## ACKNOWLEDGMENTS

This work was supported in part by the National Key R&D Program of China (grant 2018YFC2000100); the National Program on Key Basic Research Project (973 Program) (973 Program, grant 2014CB910504); the National Natural Science Foundation of China (grants 61873075 and 61573122); the China Postdoctoral Science Foundation (2016M600260); the Wu lien-teh youth science fund project of Harbin Medical University (grant WLD-QN1407); special funds for the construction of higher education in Heilongjiang Province (grant UNPYSCT-2016049); and the Heilongjiang Postdoctoral Foundation (grant LBH-Z16098).

## REFERENCES

- Alexandrov, L.B. (2015). Understanding the origins of human cancer. *Science* 350, 1175.
- Bailey, M.H., Tokheim, C., Porta-Pardo, E., Sengupta, S., Bertrand, D., Weerasinghe, A., Colaprico, A., Wendl, M.C., Kim, J., Reardon, B., et al. (2018). Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* 173, 371–385.e18.
- Cheng, F., Hong, H., Yang, S., and Wei, Y. (2017). Individualized network-based drug repositioning infrastructure for precision oncology in the panomics era. *Brief. Bioinform.* 18, 682–697.
- Adelman, K., and Egan, E. (2017). Non-coding RNA: More uses for genomic junk. *Nature* 543, 183–185.
- Hu, X., Feng, Y., Zhang, D., Zhao, S.D., Hu, Z., Greshock, J., Zhang, Y., Yang, L., Zhong, X., Wang, L.P., et al. (2014). A functional genomic approach identifies *FAL1* as an oncogenic long noncoding RNA that associates with *BMI1* and represses *p21* expression in cancer. *Cancer Cell* 26, 344–357.
- Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S.D., Zhang, Y., Yang, L., Shan, W., He, Q., et al. (2015). Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. *Cancer Cell* 28, 529–540.
- Leucci, E., Vendramin, R., Spinazzi, M., Laurette, P., Fiers, M., Wouters, J., Radaelli, E., Eyckerman, S., Leonelli, C., Vanderheyden, K., et al. (2016). Melanoma addiction to the long non-coding RNA *SAMMSON*. *Nature* 531, 518–522.
- Tan, D.S.W., Chong, F.T., Leong, H.S., Toh, S.Y., Lau, D.P., Kwang, X.L., Zhang, X., Sundaram, G.M., Tan, G.S., Chang, M.M., et al. (2017). Long noncoding RNA *EGFR-AS1* mediates epidermal growth factor receptor addiction and modulates treatment response in squamous cell carcinoma. *Nat. Med.* 23, 1167–1175.
- Zhou, C.C., Yang, F., Yuan, S.X., Ma, J.Z., Liu, F., Yuan, J.H., Bi, F.R., Lin, K.Y., Yin, J.H., Cao, G.W., et al. (2016). Systemic genome screening identifies the outcome associated focal loss of long noncoding RNA *PRAL* in hepatocellular carcinoma. *Hepatology* 63, 850–863.
- Mularoni, L., Sabarinathan, R., Deu-Pons, J., Gonzalez-Perez, A., and López-Bigas, N. (2016). OncodriveFML: a general framework to identify coding and non-coding regions with cancer driver mutations. *Genome Biol.* 17, 128.
- Li, Z., Yu, X., and Shen, J. (2016). *ANRIL*: a pivotal tumor suppressor long non-coding RNA in human cancers. *Tumour Biol.* 37, 5657–5661.
- Yu, X., Zhao, W., Yang, X., Wang, Z., and Hao, M. (2016). *miR-375* Affects the Proliferation, Invasion, and Apoptosis of HPV16-Positive Human Cervical Cancer Cells by Targeting *IGF-1R*. *Int. J. Gynecol. Cancer* 26, 851–858.
- Woo, Y.H., and Li, W.H. (2012). DNA replication timing and selection shape the landscape of nucleotide variation in cancer genomes. *Nat. Commun.* 3, 1004.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., et al. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* 15, 1034–1050.
- Khurana, E., Fu, Y., Colonna, V., Mu, X.J., Kang, H.M., Lappalainen, T., Sboner, A., Lochovsky, L., Chen, J., Harman, A., et al.; 1000 Genomes Project Consortium (2013). Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* 342, 1235587.
- Ward, L.D., and Kellis, M. (2012). Evidence of abundant purifying selection in humans for recently acquired regulatory functions. *Science* 337, 1675–1678.
- Ionita-Laza, I., McCallum, K., Xu, B., and Buxbaum, J.D. (2016). A spectral approach integrating functional genomic annotations for coding and noncoding variants. *Nat. Genet.* 48, 214–220.
- Kircher, M., Witten, D.M., Jain, P., O’Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.* 46, 310–315.
- Zhang, Y., Dang, Y.W., Wang, X., Yang, X., Zhang, R., Lv, Z.L., and Chen, G. (2017). Comprehensive analysis of long non-coding RNA *PVT1* gene interaction regulatory network in hepatocellular carcinoma using gene microarray and bioinformatics. *Am. J. Transl. Res.* 9, 3904–3917.
- Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science* 355, aah7111.
- Ahn, T., Lee, E., Huh, N., and Park, T. (2014). Personalized identification of altered pathways in cancer using accumulated normal tissue data. *Bioinformatics* 30, i422–i429.
- Sohn, E.J., Li, H., Reidy, K., Beers, L.F., Christensen, B.L., and Lee, S.B. (2010). *EWS/FLI1* oncogene activates caspase 3 transcription and triggers apoptosis in vivo. *Cancer Res.* 70, 1154–1163.
- Zheng, H., and Kang, Y. (2014). Multilayer control of the EMT master regulators. *Oncogene* 33, 1755–1763.
- Guan, Y., Kuo, W.L., Stilwell, J.L., Takano, H., Lapuk, A.V., Fridlyand, J., Mao, J.H., Yu, M., Miller, M.A., Santos, J.L., et al. (2007). Amplification of *PVT1* contributes to the pathophysiology of ovarian and breast cancer. *Clin. Cancer Res.* 13, 5745–5755.

26. Yang, Y.R., Zang, S.Z., Zhong, C.L., Li, Y.X., Zhao, S.S., and Feng, X.J. (2014). Increased expression of the lncRNA PVT1 promotes tumorigenesis in non-small cell lung cancer. *Int. J. Clin. Exp. Pathol.* *7*, 6929–6935.
27. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* *14*, 7.
28. Maurel, J., and Postigo, A.; From the Preclinical Setting to Clinical Practice (2015). Prognostic and Predictive Biomarkers in Colorectal Cancer. *Curr. Cancer Drug Targets* *15*, 703–715.
29. Nagarsheth, N., Wicha, M.S., and Zou, W. (2017). Chemokines in the cancer micro-environment and their relevance in cancer immunotherapy. *Nat. Rev. Immunol.* *17*, 559–572.
30. Şenbabaoğlu, Y., Gejman, R.S., Winer, A.G., Liu, M., Van Allen, E.M., de Velasco, G., Miao, D., Ostrovskaya, I., Drill, E., Luna, A., et al. (2016). Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol.* *17*, 231.
31. Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* *160*, 48–61.
32. Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* *12*, 453–457.
33. Lei, H., Schmidt-Bleek, K., Dienelt, A., Reinke, P., and Volk, H.D. (2015). Regulatory T cell-mediated anti-inflammatory effects promote successful tissue repair in both indirect and direct manners. *Front. Pharmacol.* *6*, 184.
34. Arce Vargas, F., Furness, A.J.S., Solomon, I., Joshi, K., Mekkaoui, L., Lesko, M.H., Miranda Rota, E., Dahan, R., Georgiou, A., Sledzinska, A., et al.; Melanoma TRACERx Consortium; Renal TRACERx Consortium; Lung TRACERx Consortium (2017). Fc-Optimized Anti-CD25 Depletes Tumor-Infiltrating Regulatory T Cells and Synergizes with PD-1 Blockade to Eradicate Established Tumors. *Immunity* *46*, 577–586.
35. Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* *483*, 570–575.
36. Sun, C., Hobor, S., Bertotti, A., Zecchin, D., Huang, S., Galimi, F., Cottino, F., Prahallad, A., Grenrum, W., Tzani, A., et al. (2014). Intrinsic resistance to MEK inhibition in KRAS mutant lung and colon cancer through transcriptional induction of ERBB3. *Cell Rep.* *7*, 86–93.
37. Geeleher, P., Cox, N.J., and Huang, R.S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. *Genome Biol.* *15*, R47.
38. Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Cokol, M., Price, E.V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing Connectivity in a Large-Scale Small-Molecule Sensitivity Dataset. *Cancer Discov.* *5*, 1210–1223.
39. Jia, P., and Zhao, Z. (2017). Impacts of somatic mutations on gene expression: an association perspective. *Brief. Bioinform.* *18*, 413–425.
40. Yue, W.Y., Yu, S.H., Zhao, S.G., and Chen, Z.P. (2009). Molecular markers relating to malignant progression in Grade II astrocytoma. *J. Neurosurg.* *110*, 709–714.
41. Martincorena, I., and Campbell, P.J. (2015). Somatic mutation in cancer and normal cells. *Science* *349*, 1483–1489.
42. Gatenby, R.A., Cunningham, J.J., and Brown, J.S. (2014). Evolutionary triage governs fitness in driver and passenger mutations and suggests targeting never mutations. *Nat. Commun.* *5*, 5499.
43. Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. USA* *107*, 18545–18550.
44. Martincorena, I., Roshan, A., Gerstung, M., Ellis, P., Van Loo, P., McLaren, S., Wedge, D.C., Fullam, A., Alexandrov, L.B., Tubio, J.M., et al. (2015). Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* *348*, 880–886.
45. Li, J., Han, L., Roebuck, P., Diao, L., Liu, L., Yuan, Y., Weinstein, J.N., and Liang, H. (2015). TANRIC: An Interactive Open Platform to Explore the Function of lncRNAs in Cancer. *Cancer Res.* *75*, 3728–3737.
46. Bjørnstad, 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.
47. Ping, Y., Deng, Y., Wang, L., Zhang, H., Zhang, Y., Xu, C., Zhao, H., Fan, H., Yu, F., Xiao, Y., and Li, X. (2015). Identifying core gene modules in glioblastoma based on multilayer factor-mediated dysfunctional regulatory networks through integrating multi-dimensional genomic data. *Nucleic Acids Res.* *43*, 1997–2007.
48. Fuller, T.F., Ghazalpour, A., Aten, J.E., Drake, T.A., Lusk, A.J., and Horvath, S. (2007). Weighted gene coexpression network analysis strategies applied to mouse weight. *Mamm. Genome* *18*, 463–472.
49. Gobbi, A., Iorio, F., Dawson, K.J., Wedge, D.C., Tamborero, D., Alexandrov, L.B., Lopez-Bigas, N., Garnett, M.J., Jurman, G., and Saez-Rodriguez, J. (2014). Fast randomization of large genomic datasets while preserving alteration counts. *Bioinformatics* *30*, i617–i623.
50. Das, S., Ghosal, S., Sen, R., and Chakrabarti, J. (2014). lncDB: database of human long noncoding RNA acting as competing endogenous RNA. *PLoS ONE* *9*, e98965.
51. Franceschini, A., Szklarczyk, D., Frankild, S., Kuhn, M., Simonovic, M., Roth, A., Lin, J., Minguez, P., Bork, P., von Mering, C., and Jensen, L.J. (2013). STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* *41*, D808–D815.
52. Li, Y., and Patra, J.C. (2010). Genome-wide inferring gene-phenotype relationship by walking on the heterogeneous network. *Bioinformatics* *26*, 1219–1224.
53. Menche, J., Sharma, A., Kitsak, M., Ghiassian, S.D., Vidal, M., Loscalzo, J., and Barabási, A.L. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* *347*, 1257601.