

RESEARCH ARTICLE

Open Access



Differential correlation analysis of glioblastoma reveals immune ceRNA interactions predictive of patient survival

Yu-Chiao Chiu^{1,2,3†}, Li-Ju Wang^{3†}, Tzu-Pin Lu⁴, Tzu-Hung Hsiao^{3*}, Eric Y. Chuang^{2,5*} and Yidong Chen^{1,6*}

Abstract

Background: Recent studies illuminated a novel role of microRNA (miRNA) in the competing endogenous RNA (ceRNA) interaction: two genes (ceRNAs) can achieve coexpression by competing for a pool of common targeting miRNAs. Individual biological investigations implied ceRNA interaction performs crucial oncogenic/tumor suppressive functions in glioblastoma multiforme (GBM). Yet, a systematic analysis has not been conducted to explore the functional landscape and prognostic significance of ceRNA interaction.

Results: Incorporating the knowledge that ceRNA interaction is highly condition-specific and modulated by the expressional abundance of miRNAs, we devised a ceRNA inference by differential correlation analysis to identify the miRNA-modulated ceRNA pairs. Analyzing sample-paired miRNA and gene expression profiles of GBM, our data showed that this alternative layer of gene interaction is essential in global information flow. Functional annotation analysis revealed its involvement in activated processes in brain, such as synaptic transmission, as well as critical tumor-associated functions. Notably, a systematic survival analysis suggested the strength of ceRNA-ceRNA interactions, rather than expressional abundance of individual ceRNAs, among three immune response genes (*CCL22*, *IL2RB*, and *IRF4*) is predictive of patient survival. The prognostic value was validated in two independent cohorts.

Conclusions: This work addresses the lack of a comprehensive exploration into the functional and prognostic relevance of ceRNA interaction in GBM. The proposed efficient and reliable method revealed its significance in GBM-related functions and prognosis. The highlighted roles of ceRNA interaction provide a basis for further biological and clinical investigations.

Keywords: Competing endogenous RNA, Immune genes, Survival prediction, Differential correlation analysis, Glioblastoma multiforme

Background

microRNAs (miRNAs) are crucial players in tumorigenesis and patient prognosis in cancers [1, 2]. Upon complementary binding to target mRNAs, they lead to mRNA degradation or translational repression [3, 4]. Recently, an alternative role of miRNA as a “bridge” of gene

interaction, namely the competing endogenous RNA (ceRNA) interaction, was proposed and verified using in vitro and in vivo models [5–7]. In the scenario of ceRNA interaction, genes (ceRNAs) achieve coexpression by competing for a common pool of targeting miRNAs (acting as sponges; illustrated in Fig. 1a). ceRNA interaction has been shown to play essential roles in the development and progression of cancers, and to potentially serve as new therapeutic targets [8–10]. Recent studies also revealed its role in propagating the effects of alterations in 3′ untranslated regions of genes, including single nucleotide polymorphisms and alternative polyadenylation [11–13]. Strength of such miRNA-mediated gene crosstalk can even outperform the one achieved by direct transcriptional regulation [14]. In vitro and in silico investigations

* Correspondence: thsiao@vghtc.gov.tw; chuangey@ntu.edu.tw; ChenY8@uthscsa.edu

†Equal contributors

³Department of Medical Research, Taichung Veterans General Hospital, No. 1650, Section 4, Taiwan Blvd., Xitun District, Taichung City 40705, Taiwan

²Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, No. 1, Section 4, Roosevelt Rd., Taipei 10617, Taiwan

¹Greehey Children’s Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA

Full list of author information is available at the end of the article



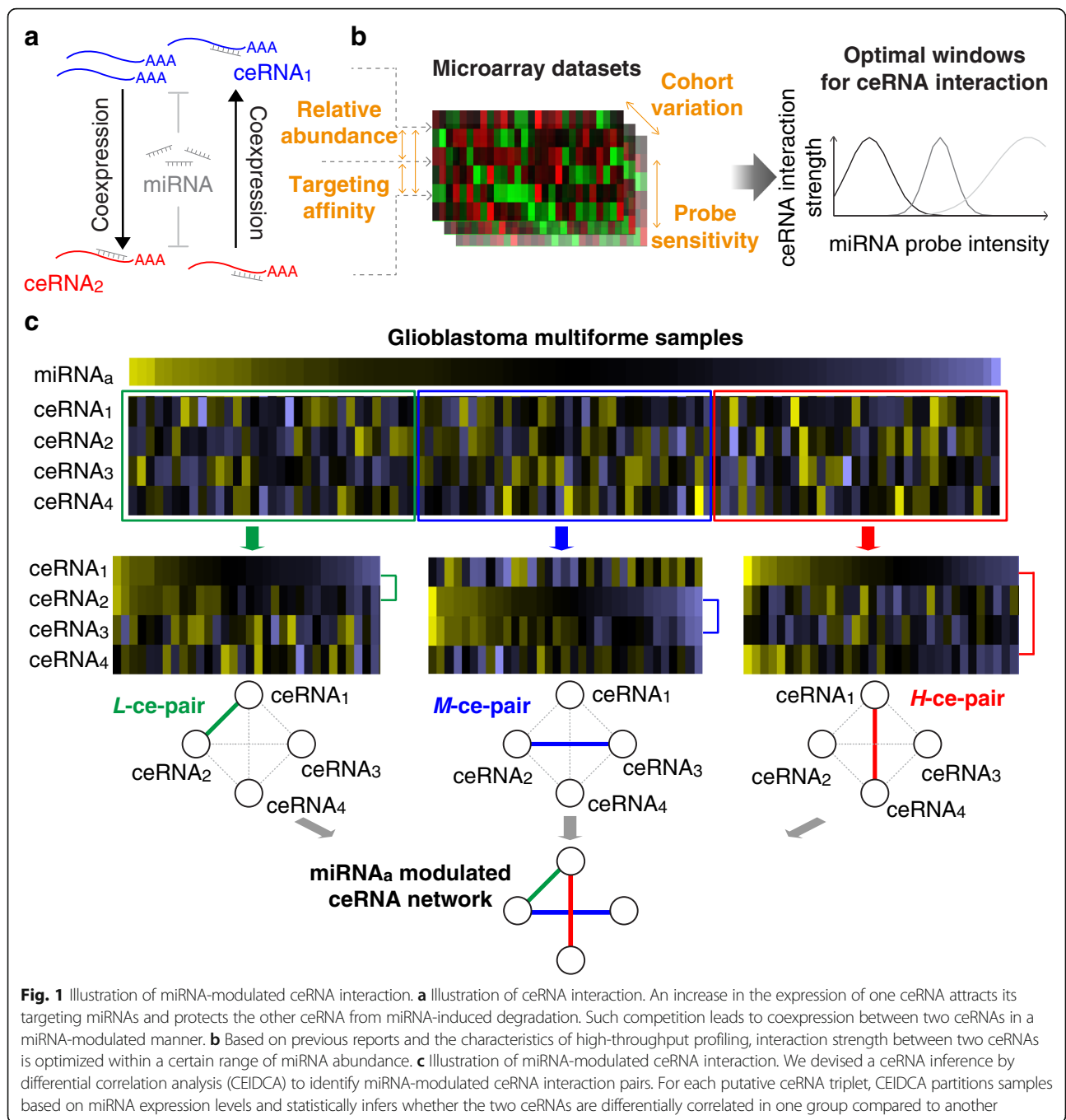


Fig. 1 Illustration of miRNA-modulated ceRNA interaction. **a** Illustration of ceRNA interaction. An increase in the expression of one ceRNA attracts its targeting miRNAs and protects the other ceRNA from miRNA-induced degradation. Such competition leads to coexpression between two ceRNAs in a miRNA-modulated manner. **b** Based on previous reports and the characteristics of high-throughput profiling, interaction strength between two ceRNAs is optimized within a certain range of miRNA abundance. **c** Illustration of miRNA-modulated ceRNA interaction. We devised a ceRNA inference by differential correlation analysis (CEIDCA) to identify miRNA-modulated ceRNA interaction pairs. For each putative ceRNA triplet, CEIDCA partitions samples based on miRNA expression levels and statistically infers whether the two ceRNAs are differentially correlated in one group compared to another

suggest that a balance among miRNA and ceRNA abundance and targeting affinities is essential for optimal ceRNA interaction [15–19]. Taking into consideration the variations from high-throughput profiling (probe sensitivity, heterogeneity and variations of sampled cohorts, etc.), ceRNA interaction appears to be modulated by (*i.e.*, dependent on) the expression of bridging miRNAs [7, 20, 21] (illustrated in Fig. 1b). In other words, coexpression between a pair of ceRNAs (ce-pair) is optimized when its bridging miRNA is expressed within a

certain range (Fig. 1b, right panel), suggesting the necessity to consider miRNA modulation in a systematic screening of ceRNA interaction.

Glioblastoma multiforme (GBM) is the most aggressive type of brain tumors. Previous works have extensively explored the molecular alterations and heterogeneity in order to better understand its biology and develop effective prognostic biomarkers [22–24]. However, 2-year and 3-year survival rates remain lower than 50% and 20%, respectively [25], suggestive of the urgent need for new

therapeutic strategies. GBM was one of the cancers in which the ceRNA hypothesis was postulated and experimentally validated [6, 7], where ceRNA interaction was demonstrated to play important roles in tumor initiation and progression. Existing systematic methods for ceRNA analysis, e.g., Hermes [6], are mostly built based on mutual information (MI) that requires heavy computation for statistical inference. Thus, a comprehensive investigation into the functional and prognostic significance of ceRNA interaction on a genome-wide scale that may reveal crucial biological and clinical clues to GBM remains an uncharted territory.

Addressing these needs, in the present study we devised a ceRNA inference by differential correlation analysis (CEIDCA) for a systematic identification of miRNA-modulated ce-pairs. Facilitated by its vastly improved computational efficiency, we conducted a genome-wide analysis and unveiled the essential involvement of ceRNA interaction in the information flow of global gene signaling, cellular signaling, and GBM-related crucial processes. Furthermore, incorporating survival data, we discovered the prognostic significance of ceRNA interactions among three immune response genes (*CCL22*, *IL2RB*, and *IRF4*), which was verified in two independent cohorts. Taken together, our data highlight the crucial roles of this alternative layer of gene interaction in GBM. The proposed method is broadly applicable to systematically explore ceRNA interaction in cancers with novel applications to regulatory biology and prognosis prediction.

Results

Overview of the CEIDCA algorithm

The study is aimed to comprehensively explore the functional landscape and prognostic significance of ceRNA interaction in GBM. We devised the CEIDCA algorithm to systematically identify ceRNA interaction pairs from sample-paired miRNA and gene expression profiles. Incorporating the findings from previous in vitro and in silico studies that ceRNA interaction is modulated by miRNA abundance, CEIDCA identifies a pair of ceRNAs (ce-pair) that are significantly differentially correlated with each other with changes in miRNA expression levels (Fig. 1). We defined putative ce-pair as two genes (say, i and j) that share at least one predicted targeting miRNA (m). For each putative ce-triplet ($i-j-m$) we tested whether the interaction strength, measured by Pearson correlation transformed to the t -domain (Eq. 1), between i and j changes among samples partitioned by the level of m . Out of simplicity, samples were stratified into high (H), medium (M), or low (L) expression of m (see the Discussion section). Statistical significance of an inter-group change (an interaction score, ΔI_{ij}^m) was evaluated by a simulation-based lookup table. A ce-pair showing significantly intensified coexpression from one

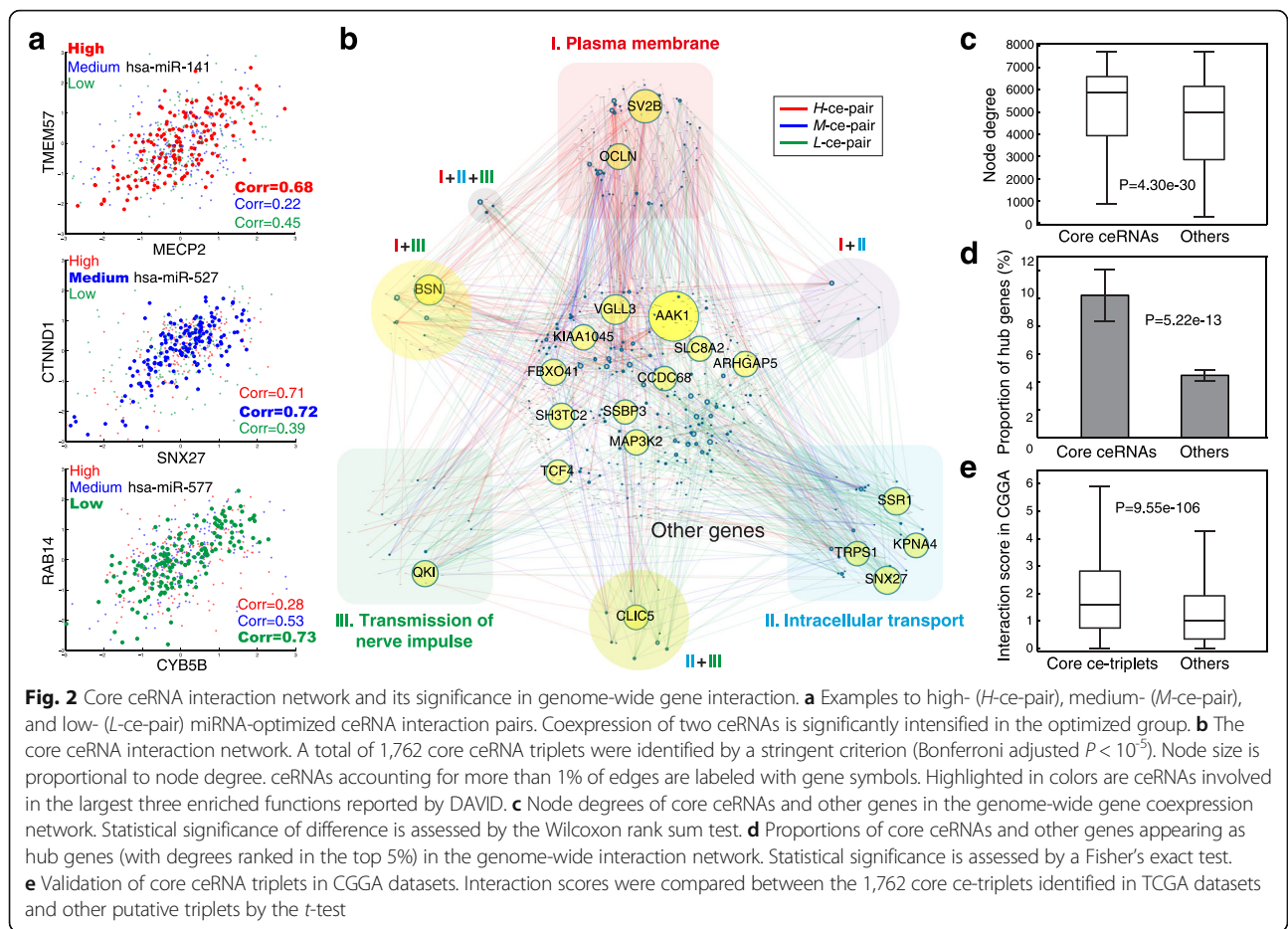
state of m (namely, optimized group G^{opt}) to another was identified as a miRNA-modulated ce-pair (G^{opt} -ce-pair). Details of CEIDCA are described in the Methods section (flowchart in Additional file 1: Figure S1).

Analysis of miRNA-modulated ceRNA interaction

In sample-paired miRNA and gene expression datasets of 520 GBM samples profiled by The Cancer Genome Atlas project (TCGA) [26, 27], we defined a total of 2,756,415 putative ce-triplets (ceRNA-miRNA-ceRNAs) that corresponded to 1,546,640 unique ce-pairs (ceRNA-ceRNAs), composed of 4,334 ceRNAs and 314 miRNAs. CEIDCA called 537,304 (3,968 ceRNAs and 304 miRNAs) significant miRNA-modulated ce-triplets ($P < 0.01$, corresponding $\Delta I > 2.36$). Among them, 27.0% (144,848), 21.5% (115,646), and 51.5% (276,810) were optimized in samples with high (H -ce-pair), medium (M -ce-pair), and low (L -ce-pair) expression of bridging miRNAs, respectively (examples in Fig. 2a). A highly interconnected ceRNA interaction network was constructed by merging these ce-pairs using Cytoscape [28] (connectivity = 270.8; Additional file 2: Figure S2A). Interestingly, the network was also highly centralized; the top 479 (12.1%) hub ceRNAs or the top 18 (5.9%) bridging miRNAs accounted for over one-half ce-triplets (Additional file 2: Figure S2B-C). The most well-studied ceRNA, *PTEN*, had an overrepresented number of ce-triplets (591, Fisher's exact test $P = 5.99 \times 10^{-26}$). In line with findings of previous in vitro investigations (reviewed in Ref. [29]), its partner ceRNAs were associated tumor suppressive functions, such as cell death (117 ceRNAs, Fisher's exact test $P = 6.49 \times 10^{-12}$) and apoptosis (93 ceRNAs, $P = 2.08 \times 10^{-10}$), reported by the knowledge-based Ingenuity Pathway Analysis software (IPA, Qiagen Inc.).

Functional landscape of core ceRNA interactions

To further investigate the functional landscape of ceRNA interactions in GBM, we extracted 1,762 core ce-triplets (Fig. 2b and Additional file 3: Table S1) by a stringent criterion (Bonferroni adjusted $P < 10^{-5}$, the most significant order of cutoff achievable from our simulation; corresponding $\Delta I > 7.36$). These ce-triplets were composed of 1,019 ceRNAs. We evaluated the role of core ceRNA interactions in global gene interaction. Notably, they had much higher node degrees in the genome-wide coexpression network than other genes (mean degrees, 5,208 vs. 4,601; Wilcoxon rank sum test $P = 4.30 \times 10^{-30}$ and t -test one-tailed $P = 3.92 \times 10^{-26}$; Fig. 2c), where the network was composed of gene pairs with Pearson correlation coefficient > 0.099 ($P = 0.012$) that includes all core ceRNA pairs. Furthermore, a higher proportion of core ceRNAs appeared to be hub genes, defined as the top 5% genes by degree [30], in the genome-wide network (10.2% vs. 4.5%; Fisher's exact test $P = 5.22 \times 10^{-13}$; Fig. 2d), suggesting the essential roles in maximizing information exchanges of gene interactions.



Interestingly, the top overrepresented ceRNAs, *AAK1*, *BSN*, and *SV2B* (Fisher's exact test $P = 1.61 \times 10^{-47}$, 1.10×10^{-32} , and 8.05×10^{-31} ; Additional file 4: Table S2), are all known to participate in cellular signaling of membrane-bound receptors or synapse [31–33]. Functional annotation analysis of the 1,019 core ceRNAs by the Database for Annotation, Visualization and Integrated Discovery (DAVID) [34] revealed concordant enrichments in synaptic/cellular signaling, accordingly plasma membrane (206 ceRNAs), cell/synapse junction (76), cytoplasmic vesicle (74), intracellular transport (108), and transmission of nerve impulse (104) (Fig. 2b and Additional file 5: Figure S3; Additional file 6: Table S3). Our findings substantiate a recent report that ceRNA is involved in routine functions of the nervous system [35], and further imply that ceRNA interaction, similar to other dynamic gene-gene interactions [36, 37], enables flexibility to the genome-wide interaction networks and facilitates cells to transiently respond to cellular stimulus and organize communication and signaling.

We also analyzed the involvement in physiology and disease by IPA. The core ceRNAs were enriched in the canonical pathway of “molecular mechanisms of cancer” (Fisher's exact test $P = 6.66 \times 10^{-9}$) and other

functions, such as “tumorigenesis of malignant tumor”, “growth of tumor”, and “epithelial-mesenchymal transition of tumor” ($P = 5.59 \times 10^{-5}$, 6.25×10^{-4} , and 3.17×10^{-3}).

Validation analysis of core ceRNA interactions

We validated the core ce-triplets identified in TCGA (Fig. 2b) by an independent dataset of sample-matched miRNA and gene expression profiles (Chinese Glioma Genome Atlas (CGGA); $n = 64$) [38, 39]. We note that for cohort variability, tumor heterogeneity, and difference in profiling technologies (illustrated in Fig. 1), the validation analysis was conducted with respect to interaction scores, not examining the consistency of G^{opt} between the two datasets. The 1,762 core ce-triplets showed concordantly increased interaction scores than other putative ce-triplets with respect to density functions (*t*-test $P = 9.55 \times 10^{-106}$; Fig. 2e) and cumulative distributions (Kolmogorov–Smirnov test statistic = 0.18, $P = 8.78 \times 10^{-46}$; Additional file 7: Figure S4). Our data suggest the reliability of CEIDCA and the stability of miRNA-modulated ceRNA interaction among cohorts.

Comparison with MI-based methods

We compared CEIDCA to two MI-based methods, namely SMI and CMI (see the Methods section for details), with respect to identified ceRNA triplets (edges in the ceRNA interaction network), ceRNAs (nodes), and computation time. Since CEIDCA utilizes a random simulation in the *t*-domain, high precision of *P*-values was achieved (to the order of 10^{-12} , while of 10^{-3} in SMI and CMI methods), enabling a statistically stringent inference and theoretically lowering false-positive rates. Generally, results of the three methods were highly comparable (Fig. 3a–b; Fisher’s exact test *P*-values $\leq 5.15 \times 10^{-15}$ and $\leq 4.01 \times 10^{-51}$ for identified ce-triplets and ceRNAs, respectively). While only a moderate proportion of CEIDCA-reported core ce-triplets (20.5%, 361 out of 1,762; Fig. 3a) were reported by SMI or CMI, reflecting the distinct mathematical characteristics of correlation and MI in detecting ceRNA-ceRNA interactions, we noted the two methods covered up to 96.7% of CEIDCA-identified ceRNAs (985 of 1,019; Fig. 3b). The data correspond to our previous study that suggests a massive rewiring among a stably maintained set of ceRNAs underlie ceRNA interaction networks in different cancer settings [20]. Since CEIDCA pre-generated a lookup table for the significance of interaction scores, the evaluation of ~2.76 million ce-triplets cost less than 1.4 h on a Xeon X7350 server with full 4-core 2.93 GHz processors (Fig. 3c). However, MI-based methods, for the need to permute datasets for each ceRNA triplet, are computationally expensive (4.7 and 5.7 days for SMI and CMI, respectively; Fig. 3c), limiting their applications to genome-wide analyses. More general simulation comparisons regarding MI-based vs. correlation-based methods can be found in [40].

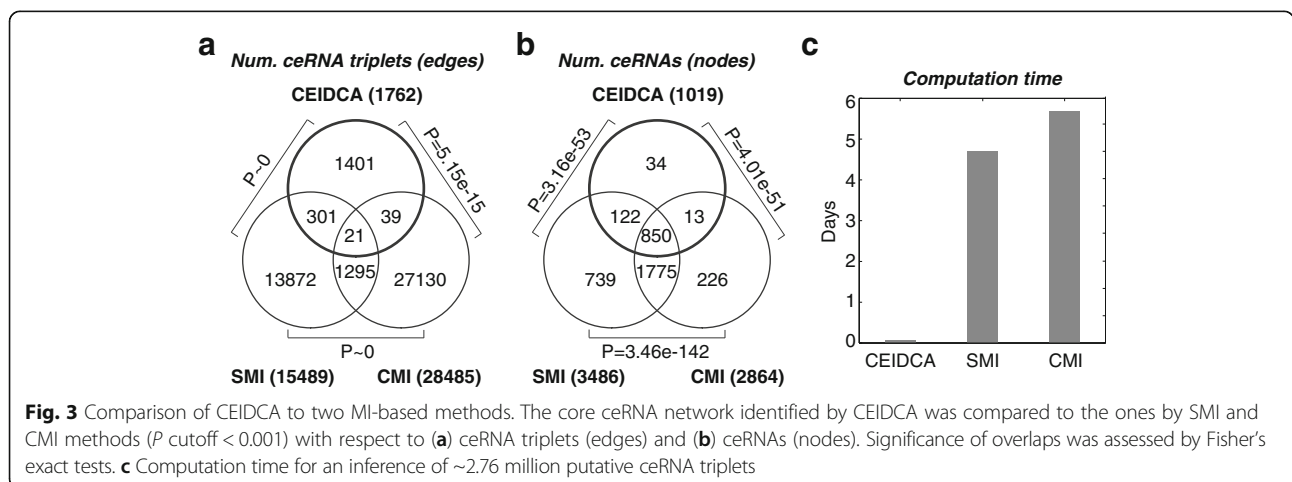
Prognostic significance of ceRNA interactions

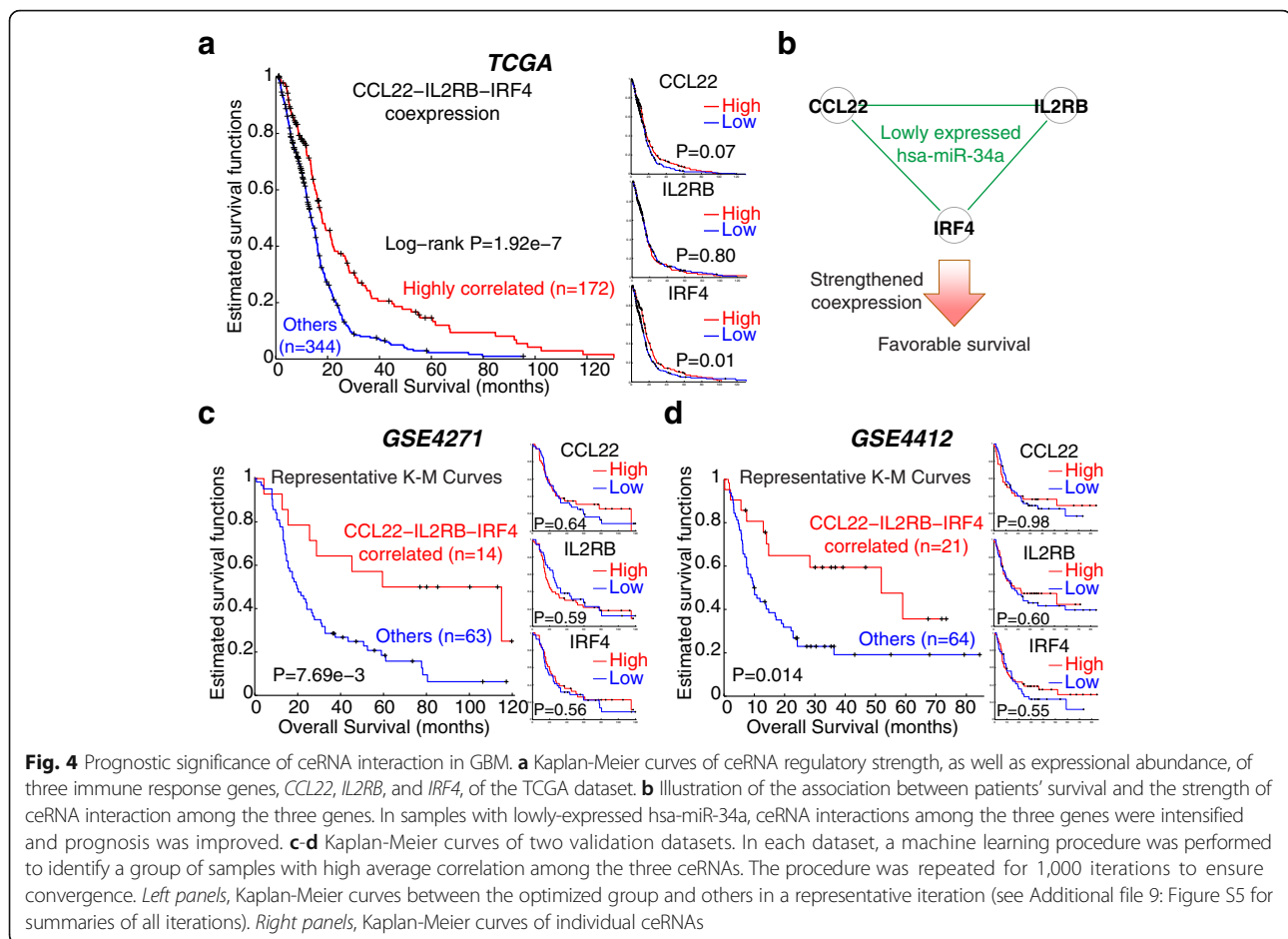
For each of the core ce-pairs, we compared the survival distributions between G^{opt} samples and others. Interestingly, pairwise interactions among three immune response-

related genes, *IRF4 – CCL22*, *CCL22 – IL2RB*, and *IRF4 – IL2RB*, were found with the highest significance (log-rank test *P*-values = 1.92×10^{-7} ; Additional file 8: Table S4). The three *L*-ce-pairs were optimally coexpressed and associated with favorable survival in patients with low hsa-miR-34a (Fig. 4a–b). Expression levels of *CCL22* and *IRF4* were also associated with patients’ overall survival (OS) (right panel, Fig. 4a). We analyzed two mRNA-only datasets, GSE4271 (*n* = 77) [41] and GSE4412 (*n* = 85) [42], to see whether the prognostic associations were attributed to the coexpression *per se* among the three ceRNAs. Of note, expression levels of none of the three ceRNAs were significantly associated with OS (right panels of Fig. 4c–d). We designed a machine learning procedure to identify an optimized subset of samples for the three ce-pairs by iteratively including samples until the average *t*-domain correlation cannot be further increased (detailed in the Methods section). Concordantly, patients with strong positive correlation among the three ceRNAs showed significantly prolonged OS (log-rank *P* = 7.69×10^{-3} and 0.014 in the two datasets; left panels, Fig. 4c–d). To test the convergence of our machine learning procedure, we repeated the entire learning procedure for 1,000 times. Overall, patients with strong positive correlation (average transformed correlation, 12.7 vs. 3.2 and 8.9 vs. 1.1; Additional file 9: Figure S5A) showed significantly extended median OS by 9.8 (30.9 vs. 21.1) and 6.3 (18.9 vs. 12.6) months (paired *t*-test *P* = 2.70×10^{-115} and 1.24×10^{-46} ; Additional file 9: Figure S5B). Also, patients with strong correlation achieved higher estimated 2-year survival rates than the complete cohorts in 89.5% and 65.9% repeats. The data strongly suggest the prognostic value of modulated ceRNA interaction and warrant further investigations into *IRF4 – CCL22 – IL2RB* in GBM.

Discussion

Immunotherapy is an emerging field in cancer biology and therapeutics. For the nature of the immune system to respond to cancer-specific or -associated antigens, it





can be employed to attack, and even prevent, cancer cells and achieve a durable response in cancer patients [43, 44]. Regulatory T cells (Tregs) are known to infiltrate brain tumors and preferentially accumulate in high-grade ones, such as GBM [45, 46]. An investigation into GBM patients' molecular profiles reported the association between Tregs and adverse survival [47]. Interestingly, *CCL22* (C-C motif chemokine ligand 22), one of the three ceRNAs we found to jointly predict survival, is a crucial mediator of Treg migration towards brain tumors [48–50] and specifically expressed in GBM (not in low-grade tumors) [47]. Although the other two genes in the 3-ceRNA signatures, *IRF4* and *IL2RB*, are not yet functionally characterized in GBM, they play essential roles in immune responses and cancer. *IRF4* encodes interferon regulatory factor 4, a member of the interferon regulatory factor family of transcription factors that are essential in interferon regulation in response to infection. The expression of *IRF4* is restricted to the immune system [51]. It is a critical player in the development/differentiation and the adaptive responses of B and T lymphocytes [52–55]. On the other hand, *IL2RB* (interleukin 2 receptor subunit beta) is a crucial player

in T cell-mediated immune responses. It was implied to be associated with apoptosis of GBM cells [56] and to participate in IL-15 induced activation of tumor-specific gamma delta T cells [57], a subset of T cells that express a unique T cell receptor and are multi-functional in cancer [58]. Taken together with our findings, miRNA-bridged crosstalk among the 3 immune genes is implied in cell-mediated immunity. Such crosstalk may enhance cellular immune response against cancer and thus has favorable prognostic effects on GBM patients. Further investigations are necessary to delineate the underlying mechanisms.

Several interaction-based approaches to predict survival were previously developed. For instance, the dynamics in protein-protein interactions was predictive of breast cancer outcome [36]. An association-based biomarker was proposed to distinguish patients harboring strong miRNA-gene interactions from others [59]. The biomarker was significantly associated with patient survival in both GBM and breast cancer. Recently, by incorporating a potential modulator feature, *NPM1* mutation, we proposed a prognostic predictor for acute myeloid leukemia based on the interaction strengths

between several miRNAs and their modulated target genes [60]. These reports showed the value to consider, in addition to expressional abundance, the interaction strength between genomic features for the development of prognostic markers. Our discovery of the *CCL22* – *IL2RB* – *IRF4* signature illuminated the potential of ceRNA interaction as a novel candidate for the interaction-based biomarkers. Though the signature was identified by stratifying patients according to the expression levels of hsa-miR-34a in TCGA data, we note that our further analysis verified the prognostic effect of the interaction strength *per se* by using two independent mRNA-only datasets. The finding echoes a central concept of ceRNA scenario that bridging miRNAs serve as only “buffers” or “sponges” of protein-coding genes that execute biological functions.

CeRNA interaction requires the correlation between two mRNAs modulated by miRNA expression levels. To systematically analyze ceRNA interaction, Hermes [6] was developed by comparing the MI between a miRNA and one ceRNA against the conditional MI given the expression profile of another ceRNA. Although MI is a widely used measure of coexpression in genomic interaction networks [61, 62], it poses a heavy computational burden due to the permutation-based statistical inference and is limited in the application to genome-wide studies. Instead, for the mathematical transformability, correlation-based methods are computationally efficient and biologically straightforward alternatives. Studies have confirmed its efficiency and comparable, or even better, performance to MI-based methods in simulated and patients’ datasets [40, 63]. Concordantly, we showed that CEIDCA identified results highly comparable to those achieved by MI-based methods (Fig. 3a–b), while greatly improving computation efficiency and statistical stringency, identify a core set of ceRNA triplets from a genome-wide study. We also note that the 3 prognostic ce-pairs cannot be identified by the MI methods (SMI *P*-values ranged from 0.18 to 0.86; CMI *P*-values falling between 0.20 and 0.95). Incorporating findings from a synthetic gene circuit or mass-action modeling [17, 18, 21], CEIDCA was designed based on a stratification of samples according to the expression level of a bridging miRNA. Out of simplicity, we set the number of groups at three (high, medium, and low). While an increase in the number of groups (*k*) equips the method the capability to identify ceRNA pairs with subtle changes, it may compromise sensitivity to systematic noises, statistical power, and computational efficiency. To understand the effect of this parameter on CEIDCA, we compared the core ceRNA interaction pairs identified by different settings ($k \in [3, 10]$). Notably, the results seemed to be quite stable (Fisher’s exact test *P*-values < precision of double-precision floating point, and Jaccard indices > 0.39, pairwise comparisons with the

setting of $k = 3$), suggestive of the robustness and wide applicability of CEIDCA, while computation time rose roughly linearly from 1.4 to 4.7 h with *k*.

Future studies may address several limitations of this work. First, since the TCGA gene expression dataset was derived by DNA microarrays, ceRNAs analyzed in this study are mostly protein-coding mRNAs. However, some long non-coding RNAs (lncRNAs), such as a well-known oncogene, *HOTAIR*, and pseudogenes (*e.g.*, *PTENP1*), perform their functions, at least partially, by acting as ceRNA partners of crucial mRNAs [64–66]. Such lncRNA-miRNA-mRNA scenario was overlooked in this study. Next, CEIDCA analyzes each of these putative triplets independently. Realizing those interactions among miRNAs and mRNAs are highly complex in living cells, ceRNA triplets may be to some extent associated with each other and form higher-order ceRNA modules, which poses great computation burden and mathematical complexity to further investigations. Future studies may, based on our findings, conduct network-based analysis to dissect higher-order graph properties among ceRNA triplets. Furthermore, we corroborated the prognostic value of the 3-ceRNA signature by stratifying patients based on the average pairwise correlation. A per-sample prediction (*e.g.*, a prediction score) was not developed. Further studies may incorporate mathematical advances in interaction-based prediction and translate our results into a personalized biomarker. Also, though we have validated its prognostic value in two independent cohorts, analysis of other large datasets is needed before it can be applied clinically.

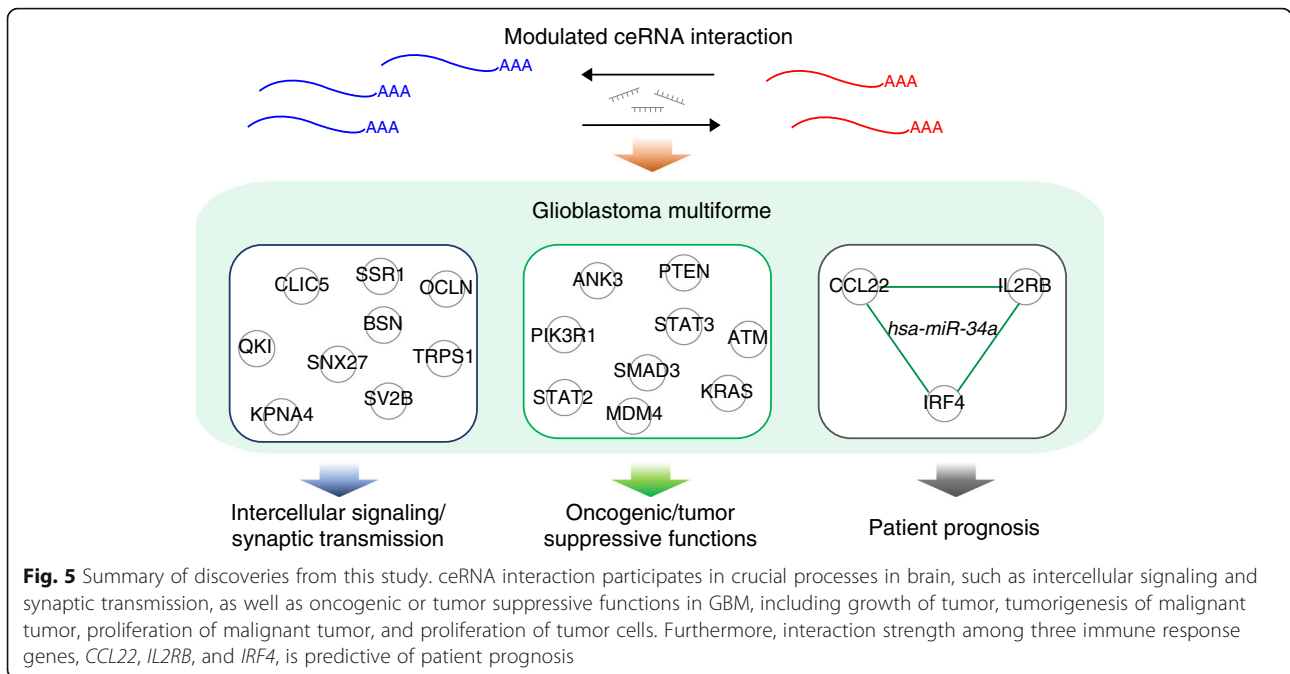
Conclusions

This work addresses the lack of a comprehensive exploration into the functional and clinical relevance of ceRNA interaction in GBM. We devised a novel and efficient algorithm that integrates miRNA and gene expression profiles of patients. As summarized in Fig. 5, by the proposed comprehensive and efficient algorithm, we showed that miRNA-modulated ceRNA interaction is involved in synaptic transmission as well as tumor-related functions in GBM. Furthermore, this is, to our knowledge, the first study to show that the interaction strength *per se* of three immune ceRNAs, *CCL22*, *IL2RB*, and *IRF4*, is predictive of patient prognosis. Overall, our findings illuminate the potential of ceRNA interaction in prognostication and therapeutics of the malignancy and warrant further biological and clinical investigations.

Methods

GBM datasets

We analyzed sample-paired datasets of miRNA and gene expression of 520 GBM patients profiled by TCGA [26, 27]. The two datasets were profiled by Agilent 8 × 15K Human



miRNA-specific microarrays and Affymetrix HT Human Genome U133 Array Plate Sets, respectively, and we used pre-normalized level-3 data. The identified core ce-triplets were validated by independent CGGA datasets of miRNA and gene expression profiles of 64 GBMs [38, 39]. To verify the results of survival analysis, we included two additional gene expression datasets from the Gene Expression Omnibus (accession numbers, GSE4271 ($n = 77$) [41] and GSE4412 ($n = 85$) [42]). For genes with multiple probes, the one with the largest coefficient of variation was chosen as a representative probe.

The CEIDCA algorithm

The CEIDCA algorithm was developed to systematically identify miRNA-modulated ceRNA interaction pairs. It contains three major components: i) definition of putative ce-pairs, ii) measuring miRNA-modulated interaction strengths of putative ce-pairs, and iii) statistical inference. In the first component (Additional file 1: Figure S1A), we processed TargetScan [67] prediction data for miRNA targets (downloaded from the miRSystem database [68]) as previously described [69]. A putative ce-pair was defined as two genes sharing at least one targeting miRNAs; a ce-triplet referred to the set of such ce-pair and the targeting miRNA.

For each putative ce-triplet, say ceRNAs i and j and miRNA m , in the second component of CEIDCA (Additional file 1: Figure S1B), samples were partitioned into k equally-sized groups ($G \in \{G_1, \dots, G_k\}$) based on the expression level of m . We measured the interaction strength between i and j within group G by a Pearson correlation coefficient $\rho_{ij,m}^G$. We eliminated non-informative

ceRNAs of which coefficients of variation were less than 5% in any group [40]. Since correlation coefficients are biased by population correlation (the change of correlation coefficients from 0 to 0.2 is statistically inequivalent to one from 0.8 to 1), we transformed the coefficients to the t -domain [70]:

$$I_{ij}^m = \left\{ t_{ij,m}^G \right\}_{G \in \{G_1, \dots, G_k\}} = \left\{ \rho_{ij,m}^G \sqrt{\frac{N-2}{1-\rho_{ij,m}^G{}^2}} \right\}_{G \in \{G_1, \dots, G_k\}}, \tag{1}$$

where N denotes the group size. We measured the inter-group change in normalized correlation coefficients as an “interaction score” by

$$\Delta I_{ij}^m = \max\left(\left\{ I_{ij}^m(G_P) \mid I_{ij}^m(G_P) > 0 \right\}\right) - \min\left(\left| I_{ij}^m \right|\right), \tag{2}$$

and tested the following hypotheses:

$$\begin{cases} H_0 : \Delta I_{ij}^m = 0 \\ H_1 : \Delta I_{ij}^m > 0 \end{cases} \tag{3}$$

We note that since only positive interaction is considered in the ceRNA scenario, ΔI_{ij}^m is set as 0 if I_{ij}^m has no positive element. Lastly, the significance level was assessed against an empirical density function $D^{\Delta I}$ generated by a trillion-time simulation of t -distributions (Additional file 1: Figure S1C):

$$P(\Delta I_{i,j}^m) = \frac{|D^{\Delta I} > \Delta I_{i,j}^m|}{|D^{\Delta I}|}. \quad (4)$$

We employed such a large number of simulations to enable Bonferroni correction addressing the multiple comparisons problem. The group G^{opt} that corresponds to $\max(I_{i,j}^m)$ for a significant ce-pair is referred to as the “optimized group” and the ce-pair as a “ G^{opt} -ce-pair”. Out of simplicity, analysis of this study was conducted under the setting of $k=3$ (i.e., $G \in \{H, M, L\}$, representing high-, medium-, and low-miRNA expression).

Implementation of MI-based methods

We compared CEIDCA to two MI-based methods, SMI and CMI. SMI was implemented by simply substituting t -domain correlation by MI in the calculation of an interaction score, with other procedures remained identical to CEIDCA. The CMI method (slightly adapted from a previous report [6]) tests the improvement of MI between m and one ceRNA i given the other ceRNA j , i.e., $\Delta I_{i,j}^m = CMI(i, m|j) - MI(i, m)$. We used the “MIToolbox for C and MATLAB” [71] for the calculation of MI and CMI. Statistical significance was assessed by 1,000 random permutations.

Survival analysis

To infer whether a miRNA-modulated ceRNA interaction is associated with prognosis, for each ce-pair we compared the survival curves between samples belonging to the optimized group and the others by a log-rank test. For each validation dataset, we used a machine-learning procedure to identify the optimized group of samples where a set of ceRNAs are (globally or locally) maximally correlated. We started by repeatedly randomly selecting 10% samples for 50,000 times and chose the selection achieving the largest average pairwise correlation (in the t -domain) as a seed. Subsequently, we iteratively added one sample into the seed that increased the average correlation by the most. The iterative procedure was terminated when the addition of any sample could not improve the correlation any more. Comparison of survival curves between the identified optimized group and the other samples, along with a comparison between samples with high and low expression of a ceRNA, directly verifies whether the predictive value of a ce-pair is confounded by the expression level *per se* of component ceRNAs. The entire machine-learning procedures were repeated for 1,000 times to ensure convergence.

Additional files

Additional file 1: Figure S1. Flowchart of CEIDCA. The algorithm is mainly built on three components. (A) Definition of putative ceRNA triplets. We defined putative ceRNA triplets by reprocessing prediction miRNA-target data of TargetScan. (B) Measuring interaction strengths of ceRNA pairs. For each putative ceRNA triplet, GBM samples were sorted and divided into three equally-sized groups based on the expression of miRNA. We employed Pearson correlation coefficients and conversion to the t -domain to measure the interaction strength between two ceRNAs. (C) Statistical inference of miRNA-modulated ceRNA pairs. We tested whether a putative ceRNA pair exhibited intensified correlation in one group compared to another. The statistical significance was assessed against a trillion-time simulation. (PDF 367 kb)

Additional file 2: Figure S2. Complete ceRNA interaction network. (A) ceRNA interaction network constructed by merging 537,304 significant ceRNA triplets ($P < 0.01$). (B) Histogram of number of ceRNA partners for a ceRNA in the network. (C) Histogram of number of bridged ceRNA triplets by a miRNA. (PDF 2423 kb)

Additional file 3: Table S1. List of 1,762 core ceRNA pairs. (PDF 535 kb)

Additional file 4: Table S2. Top overrepresented ceRNAs in the core ceRNA network. (PDF 67 kb)

Additional file 5: Figure S3. Functional subnetworks of the core ceRNA interaction network. (A) Subnetwork of the cluster of plasma membrane, extracted from Fig. 2b of main text. (B) Subnetwork of the cluster of intracellular transport, extracted from Fig. 2b of main text. (PDF 489 kb)

Additional file 6: Table S3. Top functional clusters of core ceRNAs. (PDF 2280 kb)

Additional file 7: Figure S4. Validation of core ce-triplets using CGGA dataset. We compared the cumulative distribution curves of interaction scores in the CGGA dataset between 1,762 core ce-triplets identified in TCGA and all other putative ce-triplets. Statistical significance was assessed by the Kolmogorov–Smirnov test. (PDF 162 kb)

Additional file 8: Table S4. Prognostic ceRNA triplets. (PDF 925 kb)

Additional file 9: Figure S5. Prognostic significance of *CCL22 – IL2RB – IRF4* in validation datasets. (A) Distributions of average t -scores of the three ceRNA pairs in the optimized group of samples and others in 1,000 iterations. (B) Distributions of median survival in the optimized group of samples and others in 1,000 iterations. (PDF 1295 kb)

Abbreviations

CEIDCA: ceRNA inference by differential correlation analysis; ce-pair: ceRNA pair; ceRNA: Competing endogenous RNA; ce-triplet: ceRNA triplet; CMI: Conditional mutual information; DAVID: Database for Annotation, Visualization and Integrated Discovery; GBM: Glioblastoma multiforme; IPA: Ingenuity Pathway Analysis; MI: Mutual information; miRNA: microRNA; OS: Overall survival; SMI: Substituted mutual information; TCGA: The Cancer Genome Atlas

Acknowledgements

Not applicable.

Funding

This work was supported in part by the Greehey Children’s Cancer Research Institute intramural research fund and the Ministry of Science and Technology of Taiwan (103-2917-I-002-166). The funding sources had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

All expression and clinical datasets used in the study are publicly available at The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>) and Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>).

Authors' contributions

YCC, LJW, TPL, THH, EYC, and YC conceived and designed research. YCC, THH, and YC designed analysis workflow. YCC and LJW analyzed data. YCC, LJW, TPL, THH, EYC, and YC wrote and approved the final version of paper.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable. All expression and clinical datasets used in this study were generated from previous studies and are publicly available at The Cancer Genome Atlas (TCGA; <https://cancergenome.nih.gov/>) and Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>).

Author details

¹Greehey Children's Cancer Research Institute, University of Texas Health Science Center at San Antonio, San Antonio, TX 78229, USA. ²Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, No. 1, Section 4, Roosevelt Rd., Taipei 10617, Taiwan. ³Department of Medical Research, Taichung Veterans General Hospital, No. 1650, Section 4, Taiwan Blvd., Xitun District, Taichung City 40705, Taiwan. ⁴Department of Public Health, National Taiwan University, Taipei 10055, Taiwan. ⁵Bioinformatics and Biostatistics Core, Center of Genomic Medicine, National Taiwan University, Taipei 10055, Taiwan. ⁶Department of Epidemiology and Biostatistics, University of Texas Health Science Center at San Antonio, 8403 Floyd Curl Dr., San Antonio, TX 78229, USA.

Received: 26 October 2016 Accepted: 21 February 2017

Published online: 28 February 2017

References

- Joshi P, Jeon YJ, Lagana A, Middleton J, Secchiero P, Garofalo M, Croce CM. MicroRNA-148a reduces tumorigenesis and increases TRAIL-induced apoptosis in NSCLC. *Proc Natl Acad Sci U S A*. 2015.
- Chuang MK, Chiu YC, Chou WC, Hou HA, Chuang EY, Tien HF. A 3-microRNA scoring system for prognostication in de novo acute myeloid leukemia patients. *Leukemia*. 2015;29(5):1051–9.
- Bartel DP. MicroRNAs: target recognition and regulatory functions. *Cell*. 2009;136(2):215–33.
- Shukla GC, Singh J, Barik S. MicroRNAs: Processing, Maturation, Target Recognition and Regulatory Functions. *Mol Cell Pharmacol*. 2011;3(3):83–92.
- Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, DeNicola G, Webster KA, Weiss D, Perez-Mancera PA, et al. In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*. 2011;147(2):382–95.
- Sumazin P, Yang X, Chiu HS, Chung WJ, Iyer A, Llobet-Navas D, Rajbhandari P, Bansal M, Guarnieri P, Silva J, et al. An extensive microRNA-mediated network of RNA-RNA interactions regulates established oncogenic pathways in glioblastoma. *Cell*. 2011;147(2):370–81.
- Tay Y, Kats L, Salmena L, Weiss D, Tan SM, Ala U, Karreth F, Poliseno L, Provero P, Di Cunto F, et al. Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell*. 2011;147(2):344–57.
- Karreth FA, Pandolfi PP. ceRNA cross-talk in cancer: when ce-bling rivalries go awry. *Cancer Discov*. 2013;3(10):1113–21.
- Qu L, Ding J, Chen C, Wu ZJ, Liu B, Gao Y, Chen W, Liu F, Sun W, Li XF, et al. Exosome-Transmitted IncARSR Promotes Sunitinib Resistance in Renal Cancer by Acting as a Competing Endogenous RNA. *Cancer Cell*. 2016;29(5):653–68.
- Thomson DW, Dinger ME. Endogenous microRNA sponges: evidence and controversy. *Nat Rev Genet*. 2016;17(5):272–83.
- Li MJ, Wu J, Jiang P, Li W, Zhu Y, Fernandez D, Ryan RJH, Chen Y, Wang J, Liu JS, et al. Exploring functional variation affecting ceRNA regulation in humans. *bioRxiv*. 2015. <http://biorxiv.org/content/early/2015/03/22/016865>.
- Masamha CP, Xia Z, Yang J, Albrecht TR, Li M, Shyu AB, Li W, Wagner EJ. CFIm25 links alternative polyadenylation to glioblastoma tumour suppression. *Nature*. 2014;510(7505):412–6.
- Xia Z, Donehower LA, Cooper TA, Neilson JR, Wheeler DA, Wagner EJ, Li W. Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3'-UTR landscape across seven tumour types. *Nat Commun*. 2014;5:5274.
- Martirosyan A, Figliuzzi M, Marinari E, De Martino A. Probing the Limits to MicroRNA-Mediated Control of Gene Expression. *PLoS Comput Biol*. 2016;12(1), e1004715.
- Denzler R, Agarwal V, Stefano J, Bartel DP, Stoffel M. Assessing the ceRNA hypothesis with quantitative measurements of miRNA and target abundance. *Mol Cell*. 2014;54(5):766–76.
- Bosson AD, Zamudio JR, Sharp PA. Endogenous miRNA and target concentrations determine susceptibility to potential ceRNA competition. *Mol Cell*. 2014;56(3):347–59.
- Ala U, Karreth FA, Bosia C, Pagnani A, Tauli R, Leopold V, Tay Y, Provero P, Zecchina R, Pandolfi PP. Integrated transcriptional and competitive endogenous RNA networks are cross-regulated in permissive molecular environments. *Proc Natl Acad Sci U S A*. 2013;110(18):7154–9.
- Yuan Y, Liu B, Xie P, Zhang MQ, Li Y, Xie Z, Wang X. Model-guided quantitative analysis of microRNA-mediated regulation on competing endogenous RNAs using a synthetic gene circuit. *Proc Natl Acad Sci U S A*. 2015;112(10):3158–63.
- Sanchez-Taltavull D, MacLeod M, Perkins TJ. On cross-conditional and fluctuation correlations in competitive RNA networks. *Bioinformatics*. 2016;32(17):i790–7.
- Chiu YC, Hsiao TH, Chen Y, Chuang EY. Parameter optimization for constructing competing endogenous RNA regulatory network in glioblastoma multiforme and other cancers. *BMC Genomics*. 2015;16 Suppl 4:S1.
- Lai X, Wolkenhauer O, Vera J. Understanding microRNA-mediated gene regulatory networks through mathematical modelling. *Nucleic Acids Res*. 2016.
- Hoelzinger DB, Mariani L, Weis J, Woyke T, Berens TJ, McDonough WS, Sloan A, Coons SW, Berens ME. Gene expression profile of glioblastoma multiforme invasive phenotype points to new therapeutic targets. *Neoplasia*. 2005;7(1):7–16.
- Liang Y, Diehn M, Watson N, Bollen AW, Aldape KD, Nicholas MK, Lamborn KR, Berger MS, Botstein D, Brown PO, et al. Gene expression profiling reveals molecularly and clinically distinct subtypes of glioblastoma multiforme. *Proc Natl Acad Sci U S A*. 2005;102(16):5814–9.
- Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Siu IM, Gallia GL, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008;321(5897):1807–12.
- Cho DY, Yang WK, Lee HC, Hsu DM, Lin HL, Lin SZ, Chen CC, Harn HJ, Liu CL, Lee WY, et al. Adjuvant immunotherapy with whole-cell lysate dendritic cells vaccine for glioblastoma multiforme: a phase II clinical trial. *World Neurosurg*. 2012;77(5-6):736–44.
- Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061–8.
- Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98–110.
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498–504.
- de Giorgio A, Krell J, Harding V, Stebbing J, Castellano L. Emerging roles of competing endogenous RNAs in cancer: insights from the regulation of PTEN. *Mol Cell Biol*. 2013;33(20):3976–82.
- Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H. Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. *Nat Commun*. 2014;5:3231.
- Ricotta D, Conner SD, Schmid SL, von Figura K, Honing S. Phosphorylation of the AP2 mu subunit by AAK1 mediates high affinity binding to membrane protein sorting signals. *J Cell Biol*. 2002;156(5):791–5.
- Dick O, tom Dieck S, Altmock WD, Ammermuller J, Weiler R, Garner CC, Gundelfinger ED, Brandstatter JH. The presynaptic active zone protein bassoon is essential for photoreceptor ribbon synapse formation in the retina. *Neuron*. 2003;37(5):775–86.
- Dong M, Liu H, Tepp WH, Johnson EA, Janz R, Chapman ER. Glycosylated SV2A and SV2B mediate the entry of botulinum neurotoxin E into neurons. *Mol Biol Cell*. 2008;19(12):5226–37.
- da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4(1):44–57.

35. Xu J, Feng L, Han Z, Li Y, Wu A, Shao T, Ding N, Li L, Deng W, Di X, et al. Extensive ceRNA-ceRNA interaction networks mediated by miRNAs regulate development in multiple rhesus tissues. *Nucleic Acids Res.* 2016.
36. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL. Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nat Biotechnol.* 2009;27(2):199–204.
37. Ideker T, Krogan NJ. Differential network biology. *Mol Syst Biol.* 2012;8:565.
38. Sun Y, Zhang W, Chen D, Lv Y, Zheng J, Lilljebjorn H, Ran L, Bao Z, Soneson C, Sjogren HO, et al. A glioma classification scheme based on coexpression modules of EGFR and PDGFRA. *Proc Natl Acad Sci U S A.* 2014;111(9):3538–43.
39. Yan W, Liu Y, Yang P, Wang Z, You Y, Jiang T. MicroRNA profiling of Chinese primary glioblastoma reveals a temozolomide-chemoresistant subtype. *Oncotarget.* 2015;6(13):11676–82.
40. Hsiao TH, Chiu YC, Hsu PY, Lu TP, Lai LC, Tsai MH, Huang TH, Chuang EY, Chen Y. Differential network analysis reveals the genome-wide landscape of estrogen receptor modulation in hormonal cancers. *Sci Rep.* 2016;6:23035.
41. Phillips HS, Kharbanda S, Chen R, Forrester WF, Soriano RH, Wu TD, Misra A, Nigro JM, Colman H, Soroceanu L, et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell.* 2006;9(3):157–73.
42. Freije WA, Castro-Vargas FE, Fang Z, Horvath S, Cloughesy T, Liau LM, Mischel PS, Nelson SF. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* 2004;64(18):6503–10.
43. Finn OJ. Cancer immunology. *N Engl J Med.* 2008;358(25):2704–15.
44. Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. *Nature.* 2011;480(7378):480–9.
45. El Andaloussi A, Lesniak MS. An increase in CD4+ CD25+ FOXP3+ regulatory T cells in tumor-infiltrating lymphocytes of human glioblastoma multiforme. *Neuro-Oncology.* 2006;8(3):234–43.
46. Jacobs JF, Idema AJ, Bol KF, Nierkens S, Grauer OM, Wesseling P, Grotenhuis JA, Hoogerbrugge PM, de Vries IJ, Adema GJ. Regulatory T cells and the PD-L1/PD-1 pathway mediate immune suppression in malignant human brain tumors. *Neuro-Oncology.* 2009;11(4):394–402.
47. Jacobs JF, Idema AJ, Bol KF, Grotenhuis JA, de Vries IJ, Wesseling P, Adema GJ. Prognostic significance and mechanism of Treg infiltration in human brain tumors. *J Neuroimmunol.* 2010;225(1-2):195–9.
48. Ishida T, Ueda R. CCR4 as a novel molecular target for immunotherapy of cancer. *Cancer Sci.* 2006;97(11):1139–46.
49. Jordan JT, Sun W, Hussain SF, DeAngulo G, Prabhu SS, Heimerlberger AB. Preferential migration of regulatory T cells mediated by glioma-secreted chemokines can be blocked with chemotherapy. *Cancer Immunol Immunother.* 2008;57(1):123–31.
50. Crane CA, Ahn BJ, Han SJ, Parsa AT. Soluble factors secreted by glioblastoma cell lines facilitate recruitment, survival, and expansion of regulatory T cells: implications for immunotherapy. *Neuro-Oncology.* 2012;14(5):584–95.
51. Shaffer AL, Emre NC, Romesser PB, Staudt LM. IRF4: Immunity. Malignancy! Therapy? *Clin Cancer Res.* 2009;15(9):2954–61.
52. Mittrucker HW, Matsuyama T, Grossman A, Kundig TM, Potter J, Shahinian A, Wakeham A, Patterson B, Ohashi PS, Mak TW. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science.* 1997;275(5299):540–3.
53. Lu R, Medina KL, Lancki DW, Singh H. IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. *Genes Dev.* 2003;17(14):1703–8.
54. Nayar R, Schutten E, Bautista B, Daniels K, Prince AL, Enos M, Brehm MA, Swain SL, Welsh RM, Berg LJ. Graded levels of IRF4 regulate CD8+ T cell differentiation and expansion, but not attrition, in response to acute virus infection. *J Immunol.* 2014;192(12):5881–93.
55. Huber M, Lohoff M. IRF4 at the crossroads of effector T-cell fate decision. *Eur J Immunol.* 2014;44(7):1886–95.
56. Berghoff J, Jaisimha AV, Duggan S, MacSharry J, McCarthy JV. Gamma-secretase-independent role for cadherin-11 in neurotrophin receptor p75 (p75(NTR)) mediated glioblastoma cell migration. *Mol Cell Neurosci.* 2015;69:41–53.
57. Yamaguchi T, Suzuki Y, Katakura R, Ebina T, Yokoyama J, Fujimiyama Y. Interleukin-15 effectively potentiates the in vitro tumor-specific activity and proliferation of peripheral blood gamma delta T cells isolated from glioblastoma patients. *Cancer Immunol Immunother.* 1998;47(2):97–103.
58. Silva-Santos B, Serre K, Norell H. gamma delta T cells in cancer. *Nat Rev Immunol.* 2015;15(11):683–91.
59. Ben-Hamo R, Efroni S. MicroRNA-gene association as a prognostic biomarker in cancer exposes disease mechanisms. *PLoS Comput Biol.* 2013;9(11), e1003351.
60. Chiu YC, Tsai MH, Chou WC, Liu YC, Kuo YY, Hou HA, Lu TP, Lai LC, Chen Y, Tien HF, et al. Prognostic significance of NPM1 mutation-modulated microRNA-mRNA regulation in acute myeloid leukemia. *Leukemia.* 2016; 30(2):274–84.
61. Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics.* 2006;7 Suppl 1:S7.
62. Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, et al. Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. *Nat Biotechnol.* 2009;27(9):829–39.
63. Song L, Langfelder P, Horvath S. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC Bioinformatics.* 2012;13:328.
64. Poliseno L, Salmena L, Zhang J, Carver B, Haveman WJ, Pandolfi PP. A coding-independent function of gene and pseudogene mRNAs regulates tumour biology. *Nature.* 2010;465(7301):1033–8.
65. Liu XH, Sun M, Nie FQ, Ge YB, Zhang EB, Yin DD, Kong R, Xia R, Lu KH, Li JH, et al. Lnc RNA HOTAIR functions as a competing endogenous RNA to regulate HER2 expression by sponging miR-331-3p in gastric cancer. *Mol Cancer.* 2014;13:92.
66. Eades G, Wolfson B, Zhang Y, Li Q, Yao Y, Zhou Q. lincRNA-RoR and miR-145 regulate invasion in triple-negative breast cancer via targeting ARF6. *Mol Cancer Res.* 2015;13(2):330–8.
67. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell.* 2005;120(1):15–20.
68. Lu TP, Lee CY, Tsai MH, Chiu YC, Hsiao CK, Lai LC, Chuang EY. miRSystem: an integrated system for characterizing enriched functions and pathways of microRNA targets. *PLoS One.* 2012;7(8), e42390.
69. Chiu Y-C, Chuang EY, Hsiao T-H, Chen Y. Modeling competing endogenous RNA regulatory networks in glioblastoma multiforme. In: *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on: 18-21 Dec. 2013*. 201-204.
70. Kendall MG, Stuart A. *The advanced theory of statistics: Inference and relationship*, vol. 2. London: Charles Griffin; 1961.
71. Brown G, Pocock A, Zhao M-J, Luján M. Conditional likelihood maximisation: a unifying framework for information theoretic feature selection. *J Mach Learn Res.* 2012;13(1):27–66.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

