Cellular and Molecular Neurobiology (2019) 39:1093–1114
https://doi.org/10.1007/s10571-019-00704-5

ORIGINAL RESEARCH

Hypoxia Regulated Gene Network in Glioblastoma Has Special Algebraic Topology Structures and Revealed Communications Involving Warburg Effect and Immune Regulation

Xing-gang Mao[1] · Xiao-yan Xue[2] · Ling Wang[3] · Liang Wang[4] · Liang Li[1] · Xiang Zhang[1]

Received: 24 January 2019 / Accepted: 10 June 2019 / Published online: 15 June 2019

Abstract

Hypoxia regulated genes (HRGs) formed a complex molecular interaction network (MINW), contributing to many aspects of glioblastoma (GBM) tumor biology. However, little is known about the intrinsic structures of the HRGs–MINW, mainly due to a lack of analysis tools to decipher MINWs. By introducing general hyper-geometric distribution, we obtained a statistically reliable gene set of HRGs (SR-HRGs) from several datasets. Next, MINWs were reconstructed from several independent GBM expression datasets. Algebraic topological analysis was performed to quantitatively analyze the amount of equivalence classes of cycles in various dimensions by calculating the Betti numbers. Persistent homology analysis of a filtration of growing networks was further performed to examine robust topological structures in the network by investigating the Betti curves, life length of the cycles. Random networks with the same number of node and edge and degree distribution were produced as controls. As a result, GBM–HRGs–MINWs reconstructed from different datasets exhibited great consistent Betti curves to each other, which were significantly different from that of random networks. Furthermore, HRGs–MINWs reconstructed from normal brain expression datasets exhibited topological structures significantly different from that of GBM–HRGs–MINWs. Analysis of cycles in GBM–HRGs–MINWs revealed genes that had clinical implications, and key parts of the cycles were also identified in reconstructed protein–protein interaction networks. In addition, the cycles are composed by genes involved in the Warburg effect, immune regulation, and angiogenesis. In summary, GBM–HRGs–MINWs contained abundant molecular interacting cycles in different dimensions, which are composed by genes involved in multiple programs essential for the tumorigenesis of GBM, revealing novel interaction diagrams in GBM and providing novel potential therapeutic targets.

Keywords Protein–protein interaction · Molecular network · Mutual information · Degree distribution · Betti number · Persistent homology · Drug resistance · Angiogenesis

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10571-019-00704-5) contains supplementary material, which is available to authorized users.

Xing-gang Mao and Xiao-yan Xue contributed equally to the study.

✉ Xing-gang Mao
xgmao@hotmail.com

✉ Xiang Zhang
xzhang@fmmu.edu.cn

[1] Department of Neurosurgery, Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi, People's Republic of China

[2] Department of Pharmacology, School of Pharmacy, Fourth Military Medical University, Xi'an, Shaanxi, People's Republic of China

[3] State Key Laboratory for Manufacturing System Engineering, School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710054, People's Republic of China

[4] Department of Neurosurgery, Tangdu Hospital, Fourth Military Medical University, Xi'an, Shaanxi, People's Republic of China

Abbreviations

ARACNe	Algorithm for the reconstruction of accurate cellular networks
CI	Confidential interval

CSC	Cancer stem cell
DD	Degree distribution
DPI	Data processing inequality
D_{gene}	Connectivity degree of the gene
GBM	Glioblastoma
GBM–HRGs–MINWs	Molecular interaction network of hypoxia regulated gene reconstructed from GBM expression datasets
GSCs	Glioblastoma stem like cells
HRGs	Hypoxia regulated genes
MI	Mutual information
MINW	Molecular interaction network
N_{edge}	Number of edges
N_{node}	Number of nodes
NC_{gene}	Number of cycles involved for a gene
$NC_{q\text{-clique}}$	Number of cycles involved for a q dimensional clique (or simplex)
Normal-HRGs–MINWs	Molecular interaction network of hypoxia regulated gene reconstructed from normal brain expression datasets
NSC	Neural stem cells
PH	Persistent homology
PPI	Protein–protein interaction
PPI–MINWs	Protein–protein interaction MINWs
PPI–HRGs–MINWs	PPI–MINWs reconstructed with the HRGs
q-simplex	q dimensional simplex (fully connected $(q+1)$ nodes)
q-cycle	q dimensional cycle
q-sig-cycle	q dimensional cycle whose life is statistically long (longer than the 95% bound of q-cycle life in random-MINWs)
q-end-cycles	q dimensional cycle in the end point of filtration
RandomExp	Random gene expression datasets
RandomSameDD-MINWs	Random-MINWs with the same degree distribution (DD) to the corresponding MINWs
SR-HRGs	Statistically reliable gene set of HRGs
SRExp	Rank values ordered by expression levels for each gene
sub-cycles-3genes	Sub-cycles containing three genes that are presented in both GBM–HRG–MINWs and PPI–HRGs–MINW
TF	Transcription factor
Up-HRGs	Up-regulated HRGs

Introduction

Glioblastoma (GBM), the most lethal primary malignant brain tumor in adult, still lacks ideal treatment, and patients with GBM only have a median survival of about 14.6 months after combined treatment including surgery resection and adjuvant chemo- and radiotherapies (Stupp et al. 2005). A lot of potential molecule targets have been proposed, while effective treatment strategies are not available. The recently approved anti-VEGF mono-antibody bevacizumab has limited effect which would result in drug resistance and ultimately recurrence of the tumor (Mahase et al. 2017; Michaelsen et al. 2018; Tamura et al. 2017). Therefore, efforts were made to find more promising therapeutic targets. However, given the fact that cancers including GBM are diseases driven by multiple genetic and epigenetic alterations, single or few molecule targets may not be sufficient to cure this kind of tumor (Brennan et al. 2013; TCGA 2008). Based on this assumption, and along with the progress in high-throughput data acquisition technologies, it is reasonable that a cohort of genes, and more precisely, a molecular interacting network may fundamentally drive the disease (Chen et al. 2014; Plaisier et al. 2016). For example, the existence of drug resistance when targeting certain key genes implied a hidden unknown molecular network with ingenious organizational structures developed by the tumor (Piao et al. 2013; Rizzolio et al. 2018; Tamura et al. 2017; Urup et al. 2017), which have great adaptability to compensate the missing of certain key genes. Therefore, great efforts have been made to reconstruct the molecular interaction network (MINW) for cancers (Carro et al. 2010; Kim et al. 2010; Kleaveland et al. 2018; Li and Izpisua Belmonte 2018; Nakagawa et al. 2018; Shi et al. 2018). The reconstruction and investigation of the structure and function of the MINW not only revealed the intrinsic nature of the tumor biology, but also provide potential therapeutic avenues to change or even reverse the global biological features of the tumorigenic potential of the cancer. A lot of experimental and theoretical approaches have been developed to reconstruct the network (Markmiller et al. 2018; Shi et al. 2018), and some achieved excellent results (Wang et al. 2009). As has been reported previously, hypoxia is a critical factor promoting the tumorigenesis of

GBM (Li et al. 2009; Man et al. 2018; Wang et al. 2018), and is involved in many biological processes, including tumorigenic potential, cancer stem cell (CSC) maintenance (Li et al. 2009), invasion (Du et al. 2008; Joseph et al. 2015), angiogenesis or vascular mimicry (Joseph et al. 2015; Mao et al. 2013c), and anti-VEGF therapy resistance (Mahase et al. 2017; Mao et al. 2016; Tamura et al. 2017; Xu et al. 2015), et al.

However, little is known about the global characteristics of the MINW, especially in the high dimensional space containing thousands of genes, given that each gene represents a dimension. Obviously, the MINW is a network defined by the relationships between each gene, and is metric independent. As has been described and developed in mathematics, topology studied the integral features of the geometric space without consideration of the measures such as length and angle quantity (Giusti et al. 2015, 2016; Sizemore et al. 2018). To more rigorously describe the characteristics of different geometric objects, especially the more complicated cases such as molecular networks, algebraic topology was established and homology group analysis was used to quantitatively describe different topological objects (Adams et al. 2014; Giusti et al. 2016; Sizemore et al. 2018). Although traditional graph analysis provides certain information, little, if any, is reported to deeply decipher the integral structures of MINW quantitatively, especially the loop structures, by algebraic topology analysis. In this study, we reconstructed the MINW of hypoxia regulated genes (HRGs) in GBM, and analyzed its algebraic topology features by calculating the number of simplexes, Betti numbers, and Betti curves by persistent homology (PH) analysis. Our results first implied that GBM–HRGs–MIWN had special robust intrinsic topological structures, which are statistically different from that of random networks, and importantly, different from that of normal brains.

Materials and Methods

Identification of Hypoxia Regulated Genes (HRGs)

To get a reliable HRGs set, we searched and screened high-throughput datasets of hypoxia regulated genes in GBM from Pubmed Dataset website by searching the item "(glioblastoma or glioma) and (hypoxia or oxygen)". At last, HRGs were identified from 5 cell results, including two U87 cells data [GSE32100 (Nissou et al. 2013) and GSE45301 (Kucharzewska et al. 2013)], one GSC cell data [GSE32100 (Nissou et al. 2013)], one serum cultured GSC cell data [GSE32100 (Nissou et al. 2013)], and one LN229 cell line data [GSE27523 (Koh et al. 2011)]. The genes which are statistically significantly up-regulated or down-regulated ($p < 0.05$) are chosen as candidate HRGs in GBM. Given the relatively great variation of high-throughput data, we selected overlapped genes in the 5 datasets to get most reliable HRGs by taking advantages of the general hypergeometric distribution (GHG) (Mao and Xue 2018). The GHG was used because if we use the genes overlapped in at least 5 or 4 datasets, then we only get a few genes and would lose too many HRGs, while if we use genes overlapped in too few sets then the result HRGs would contain too many false positive genes. By using the formulas of mathematical expectations and variances of the GHG (Mao and Xue 2018), the 95% confidential interval (CI) of the GHG can be estimated with Chebyshev's inequality, which give an upper bound of number of randomly overlapped genes (random_Up_NOGs). Then the number of statistically significant overlapped genes (sig_NOGs) can be deduced by sig_NOGs = observed_NOGs − random_Up_NOGs (Mao and Xue 2018). Then total number of sig_NOGs genes ranked by change folds will be taken as statistically reliable HRGs (SR-HRGs).

Reconstruction of MINW with ARACNe Based on Information Theory

To reconstruct a MINW, the first step is to get the interactome of each gene in the HRGs. We also used several datasets (Mao et al. 2013a, b, 2016) to reconstruct the HRGs–MINW (GBM–HRGs–MINWs), by using Algorithm for the Reconstruction of Accurate Cellular Networks (ARACNe), an information-theoretic algorithm for inferring transcriptional interactions (Margolin et al. 2006a), to identify a repertoire of candidate transcriptional regulators of interesting genes. First, candidate interactions between a transcription factor (TF, x) and its potential target (y) are identified by computing pairwise mutual information, MI[x; y], using a Gaussian kernel estimator and by thresholding the mutual information based on the null hypothesis of statistical independence ($p < 0.05$, Bonferroni corrected for the number of tested pairs). Here mutual information (MI) is a concept in information theory which measures the degree of statistical dependency between two variables (Margolin et al. 2006a; Shi et al. 2018), like the commonly used Pearson correlation. However, correlation coefficients are not feasible to deal with nonlinear associations, and may be zero even for manifestly dependent variables. MI is reparameterization invariant and is nonzero if and only if any kind of statistical dependence exists (Margolin et al. 2006a). Then, indirect interactions are removed using the data processing inequality (DPI), a well-known property of the mutual information. The DPI states that if genes g1 and g3 interact only through a third gene g2, then MI(g1, g3) ≤ min [MI(g1, g2); MI(g2, g3)] (Cover and Thomas 1991; Margolin et al. 2006a). To remove potential indirect interactions, during DPI process, in each gene triplet for which all three MIs are greater than the MI threshold, the edge with the smallest value

would be removed. Therefore, the networks identified with DPI (DPI-MINWs) do not contain simplexes with a dimension higher than 1, because the DPI process removed potential 2-simplexes. The DPI-MINWs contained more reliable direct interactions and was used to analyze 1-cycles, while the NoDPI-MINWs were used to find higher-dimensional cycles of the networks.

Calculation of Homology Group and Betti Numbers

Here a simplicial complex K was made up of vertices and simplexes (Giusti et al. 2016). A p-dimensional simplex (p-simplex) is defined in terms of collections of $(p+1)$ full connected vertices. For example, a point is a 0-dimensional simplex, an object of two points with a connection in a network is a 1-dimensional simplex, and an object of three points which are connected to each other is a 2-dimensional simplex, et al. It should be noted that a p-simplex is not simply the collection of its verities and edges, it is actually a p-dimension object, and therefore the edges are actually its "sub-simplex", which is called a "face" of the simplex. Actually, any subset of the vertices of one simplex is a face of the simplex. For all p-simplexes, assign a value (or an element) in a *group* for each of the p-simplex (coefficient *group*; here *group* is a conception in algebra, which is defined as follows. A group is an algebraic structure consisting of a set of elements equipped with a binary operation which combines any two elements to form a third element. To be a group, this operation must satisfy four conditions called the group axioms, namely closure, associativity, identity and invertibility). An operation for the p-simplex was defined as the same operation of its corresponding element in the *group*. Therefore, a finite number of p-simplex with the above defined operation formed a *chain* with p-dimension (p-chain). For example, for a set of p-simplexes, s_1, s_2, …, s_l, each s_i can be represented by its vertices: $\sigma_k = \sigma\left[v_{i_1}, v_{i_2}, \ldots, v_{i_p}\right]$, and the p-chain is:

$$C_p(K) = \sum_{i=0}^{l} \sigma\left[v_{i_1}, v_{i_2} \ldots v_{i_p}\right]$$

Therefore, each p-chain has a value belonging to the group. All of the p-chains with the above defined operation forms a group $C_p(K)$.

The simplex can be oriented according to the order of the vertices. Next define the boundary map for each simplex $C_p(K) \to C_{p-1}(K)$ ($\widehat{v_{i_j}}$ indicates omission of the vertex v_{i_j}):

$$\partial\sigma\left[v_{i_1}, v_{i_2}, \ldots, v_{i_k}\right] = \sum_{j=0}^{k} (-1)^j \sigma\left[v_{i_1}, v_{i_2} \ldots \widehat{v_{i_j}} \ldots v_{i_k}\right]$$

It is clear that the boundary map ∂ transforma a p-simplex to a series of $(p-1)$-simplexes, and the result $(p-1)$-simplexes, which form a $(p-1)$-chian, are denoted as a boundary for the p-simplex. For a given p, the image from a upper dimension $\partial_{p+1}\left(C_{p+1}(K)\right)$ ($\mathrm{im}\partial_{p+1}$, that is boundaries of the $p+1$ chains) is a subgroup of the $C_p(K)$,, and the kernal of the $\partial_p\left(C_p(K)\right)$ ($\ker\partial_p$, closed chains in the $C_p(K)$, which are termed as cycles) is also a subgroup of $C_p(K)$. Here a p-dimensional cycle (p-cycle) is a "closed p-chain", that is, all the p-simplexes constituting the p-cycle have a 0 value in the above defined operation. In addition, considering that $\partial_{p+1}\partial_p = 0$, any boundary from an upper dimension is a cycle, and therefore the $\mathrm{im}\partial_{p+1}$ is a subgroup of $\ker\partial_p$. Based on these definitions, the homology group for dimension q $H_q(K)$ is defined as the quotient group:

$$H_p(K) = \ker\partial_p / \mathrm{im}\partial_{p+1}$$

Behind the highly abstract definition process of the homology group, the $H_p(K)$ have special geometric meanings. For finite simplicial complexes, which is the main topic for gene networks, the $H_p(K)$ is a finitely generated Abel group, and the rank of the group is called Betti number Betti_q. Intuitively, betti_0 indicates the number of connected graphs, betti_1 indicates the number of 1-dimensional cycles, while betti_2 indicates the number of 2-dimensional cycles (geometrical structures similar to the surface of a ball, but not containing the inside). In the present study, the Z/Z2 *group* (a 2-order cyclic group; here Z is an abelian group consisted of all integers in the operation addition "+"; Z2 is the group consisted of all even integers; Z/Z2 is the quotient group of Z and Z2, which is a 2-order cyclic group containing two elements) was used as the coefficients *group*. Calculation of Betti numbers were performed by using the JPlex program (Adams et al. 2014).

Persistence Homology Analysis of the MINWs

Given the great fluctuations of biological data, especially high-throughput data, the MINWs reconstrued with the static data would not reflect their intrinsic structures. Therefore, persistent homology (PH) analysis was introduced to examine the stable intrinsic algebraic topological structures (Giusti et al. 2015; Sizemore et al. 2018). PH was performed on a series of simplicial complex, which is called a filtration. A filtration is formed by a growing simplicial complex, that is, simplexes are gradually added to the complex, and therefore cycles are formed and diminished. During the process, a filtration value was assigned for each step, and the length between the values of appear and disappear of a cycle was defined as the life for that cycle. For example, as shown in Fig. S1, the filtration started in value 1 and ended in value 5. At value 1, there are five 0-simplexes (points) [1], [2], [3], [4], [5] and four 1-simplexes [1,2], [2,3], [4,5], [1,5]. At value 2, one more 1-simplex [3,4] was added into the complex and thus a 1-cycle (cycle 1) was formed, hence the life

of the 1-cycle started at value 2. At value 3, two more 0-simplexes [6], [7] and three more 1-simplex [2,6], [3,7], [6,7] were added and another 1-cycle (cycle 2) was formed whose life started as value 3. At value 4, a 2-simplex δ1 [3,6,7] was added, and the two 1-cycles were persisted. At value 5, another 2-simplex δ2 [2,3,6] was added and thus cycle 2 was disappeared whose life ended at value 5, while cycle 1 still persisted. Therefore, there are totally two 1-cycles, and the life length of cycle 1 was infinite − 2 (here infinite indicates the cycle persists longer than the observed filtration values), while that of cycle 2 was 5 − 3 = 2.

In the present MINW, for PH analysis, at value 0, all of the 0-simplex were added to the MINW complex. Then the mutual information MI[x; y] for each paired genes were ranked in descending order, and the edges formed by these two genes [x,y] were added gradually to the complex according to their order. For each step, the value of rank order for the edge was assigned as filtration value. Therefore, any high-order simplexes (interconnected nodes) formed due to the addition of the edge were also added into the complex in this step. When all of the edges in the above reconstructed MINW were added, a filtration formed and PH analysis was performed by using the JPlex program (Adams et al. 2014).

Construction of Random Networks

To get topological features of random networks (Random-MINWs), networks with the defined nodes and edges were constructed. For PH analysis, the filtration began at value 0 with all of the nodes in the complex, and edges were randomly added to the simplicial complex (SC). One edge was added in each step, until the total number of edges reached to the defined amount. For DPI-networks which did not contain 2 or greater-dimension simplexes, edges were randomly added only when it did not form 2 or greater-dimension simplexes with other edges. When the number of edges reached to the defined amount, a random filtration with defined nodes and edges was produced. Then the above algebraic topological and PH analysis were performed to get betti number in each dimension. To get the distribution properties of the topological features of Random-MINWs, 1000 filtrations were produced, and the betti number, life lengths, et al. were calculated with the above process, and the statistical distribution of these parameters were established. In our practice, only 100 or fewer random filtrations would produce quite consistent distribution curves of the Betti numbers, life lengths, et al.

To produce Random-MINWs with same degree distribution (DD) (RandomSameDD–MINWs) of the GBM–HRGs–MINWs, during the Random-MINWs production, a random edge was evaluated before adding to the network. Only the edge meeting the following criteria would be added to the network: the network with the added edge have a DD that is not larger than that of the GBM–HRGs–MINWs. The definition of comparison between two DDs (for example DD1 and DD2) are defined as follows: the two networks should have same number of nodes (N_{node}), let the degrees for all nodes are D_{nodes} (here D_{nodes} is a integer array with N_{node} elements; let the two D_{nodes} are $D1_{nodes}$ and $D2_{nodes}$), sort the both D_{node} descending, if each value in $D1_{node}$ is not larger than the corresponding value in $D2_{node}$ then DD1 is not larger than DD2. For example, let $D1_{nodes} = \{5,4,4,3,1,1\}$, $D2_{nodes} = \{4,3,4,3,1,1\}$, $D3_{nodes} = \{5,5,3,3,1,1\}$, then DD1 is not larger then DD2, while DD2 and DD3 can not be compared. Because each GBM–HRGs–MINWs reconstructed from the 5 GBM datasets has its own DD, corresponding RandomSameDD–MINWs were produced. Totally 100 RandomSameDD–MINWs for each DD were produced for statistical analysis.

Ranking Values of the Cycles

First, all of the cycles during PH analysis were listed. Then, the degree of each node (gene; here degree indicate the number of edges containing the node) in the MINW were calculated, which is denoted as degree of the gene's (D_{gene}). The number of q-cycles containing the node was defined as number of q-cycles involved a gene (q-NC_{gene}), and the q-NC_{gene} for each gene was calculated by examining all of the q-cycles during PH analysis. Next, a rank value for a q-cycle representing its importance was defined as the average value of all q-NC_{gene} for each gene in the q-cycle. Then the cycles can be ordered by their rank values and cycles in the top rank represent relatively more important ones for the network.

Culture of Primary GSCs and NSCs

Isolation and culture of cells, including GSCs, NSCs, U87, and U251 cells, were performed as described in our previous papers (Mao et al. 2009a, Mao et al. 2010, 2011, 2013a, b). Briefly, for primary GSC culturing, GBM samples were dissociated to a single cell suspension using a fire-polished Pasteur pipette and cultured in serum-free medium (SFM) consisting of DMEM-F12 medium, EGF (20 ng/mL; Invitrogen, Carlsbad, CA), bFGF (20 ng/mL; Invitrogen, Carlsbad, CA) and B27 (1:50; Invitrogen, Carlsbad, CA). Human glioblastoma U251MG and U87MG cells (U87 and U251) were purchased from GENECHEM Company (Shanghai, China) and cultured in DMEM medium supplemented with 10% newborn calf serum (GIBCO BRL, Invitrogen). U87 and U251 cell lines were authenticated with short tandem repeat in 2018. Two human fetal cortical NSCs were isolated from spontaneous aborted fetuses (8–12 weeks), which were dissociated into single cells as described above and then cultured in SFM. Primary human astrocytes (Life Technologies,

Carlsbad, CA, USA; lot number 802268) were cultured in DMEM medium supplemented with 10% newborn calf serum, 1% N-2 Supplement (Life Technologies), and EEF (20 ng/mL). The study protocol was approved by the institutional review board of Xijing Hospital of the Fourth Military Medical University, and written informed consent was obtained from patients.

qPCR

RNA was extracted from cultured cells and brain tumor tissues using Trizol Reagent (Invitrogen). The extracted RNA was then reverse transcribed into cDNA using Superscript II RT (Gibco-BRL, Gaithersburg, Maryland) and a mixture of oligo dT primers and random hexamers, according to the manufacturers' instructions. qPCR analysis was performed on an ABI7700 using SYBR Green PCR Core Reagents in 20 µL volume (Applied Biosystems, Warrington, UK). qPCR using water instead of template was used as negative controls. All samples were assayed in triplicate and the relative amount of target transcripts normalized to the number of human β-actin transcripts found in the same sample. Specificity was verified by melting curve analysis and agarose gel electrophoresis. Relative fold changes were calculated using the $\Delta\Delta C_t$ method by using the threshold cycle values of each sample.

Results

Definition of Statistically Reliable Hypoxia Regulated Genes (SR-HRGs)

To reconstruct the HRGs–MINW, the first step is to identify a gene set of SR-HRGs. We screened the up-regulated and down-regulated HRGs (Up-HRGs and Down-HRGs) under hypoxia conditions, by using five cell datasets. The genes which are statistically up-regulated or down-regulated ($p < 0.05$) are chosen as candidate HRGs in GBM. Next, we used the general hyper-geometric distribution (GHG) (Mao and Xue 2018) to get genes that are overlapped in the five lists of HRGs with statistically significance as the bona fide SR-HRGs. The GHG distribution is used because we are going to find a set of genes, where "all of the genes" rather than "existence of genes" are HRGs with statistical significance (Mao and Xue 2018). Therefore, to quantitatively determine how reliability of these genes can be used as SR-HRGs, we should examine the statistical distribution of the "number of genes overlapped in at least t times" ($N_{OL\geq t}$, and the corresponding genes were denoted as $G_{OL\geq t}$, here $1 < t \leq 5$). By taking advantage of the GHG, the 95% confidential interval (CI) of $N_{OL\geq t}$ are calculated, and the HRG set with statistical significance can be deduced (Mao and Xue 2018) (Fig. 1, Table 1, and Supplementary Table S1).

As a result, we got 1424, 545, 986, 1205, 380 genes that are significantly up-regulated in the five datasets respectively, and there are 42, 108, 276 genes that are overlapped at least 5, 4, and 3 times ($N_{OL\geq 5}$=42, $N_{OL\geq 4}$ = 108, $N_{OL\geq 3}$ = 276). According to the GHG analysis, all of the $G_{OL\geq 5}$ are statistically significant, while for the $G_{OL\geq 4}$ and $G_{OL\geq 3}$, there are at most 2 and 32 false positive genes that may be caused by random, respectively. At last, we used the $G_{OL\geq 3}$ as the SR-HRGs by excluding bottom 32 genes ranked by their change folds, resulting in 244 (= 276 − 32) Up-SR-HRGs (Table 1). Similarly, 208 Down-HRGs were identified with the same procedure (Supplementary Table S1). We would focus on the Up-SR-HRGs, which would be simplified as HRGs thereafter, if not confused.

Reconstruction of Hypoxia Regulated Gene Networks

Next, the 244 HRGs were used to reconstruct the HRGs–MINWs. To reconstruct a network, we used the

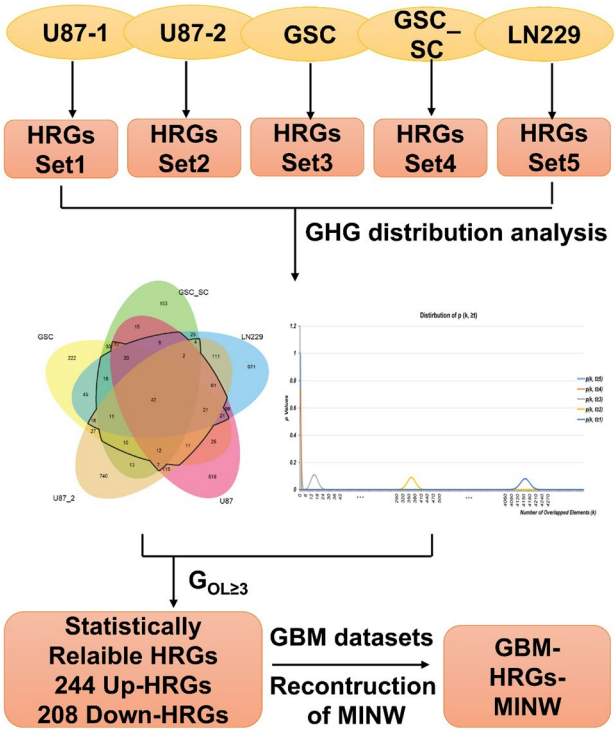


Fig. 1 Procedures of obtaining reliable HRGs and reconstruction of the HRG-MINWs. Hypoxia regulated genes were identified from five datasets, resulting in five HRG lists. The significance of genes overlapped in these five lists ($G_{OL\geq t}$, $5 \geq t \geq 1$) were examined with GHG distribution, and at last $G_{OL\geq 3}$ excluding the potential random ones were chosen as reliable HRGs, which were used to reconstruct the MINW with or without DPI

Table 1 Statistical analysis of the overlapped Up-SR-HRGs derived from the five datasets according to the GHG distribution

Random variable	Mean	Variance	95% CI[a]	Observed values	*N* of observed values with significance	% (observed values with significance)
$N_{OL\geq5}$	0.00212	0.0022	[0, 0.2]	42	42	100%
$N_{OL\geq4}$	0.29	0.29	[0, 2.7]	108	106	98.15%
$N_{OL\geq3}$	15.15	14.71	[0, 32.3]	276	244	88.41%
$N_{OL\geq2}$	361.74	254.04	[290.4, 433.0]	780	347	44.49%
$N_{OL\geq1}$	4162.80	286.46	[4087.1, 4238.5]	3334	–	–

[a]The 95% CI was estimated by using the Chebyshev's inequality

mutual information (MI) based dependency to define the relationships between paired genes. By using the MI based ARACNe algorithm (Margolin et al. 2006a), the HRGs–MINWs were reconstructed from five different GBM sample expression datasets (GBM–HRGs–MINWs): a unified GBM data from TCGA (Verhaak et al. 2010), a validation GBM data from Verhaak et al. (2010), the GBM data from Rembrendt (GSE68848) (Madhavan et al. 2009), a GBM dataset from Lee et al. (2008) (GSE13041), and a GBM data from Gravendeel et al. (2009) (GSE16011), which contained 197, 260, 228, 191, and 159 samples, respectively. A threshold of 10^{-8} was used to obtain candidate genes that interacted with each gene in the 244 HRGs with or without a DPI process to remove potential indirectly interacting genes. It should be noted that there is still no ideal approach to reconstruct the MINW exactly, either by experimental or theoretical approaches, and any reconstructed network is an approximation of the realistic one with certain deviation. In recent years, information-theoretic approaches were more and more widely recognized as powerfully tools to reconstruct the MINW (Shi et al. 2018). There has been several MI based approaches to reconstruct a MINW based on gene chip data, and some achieved excellent results with recall comparable to that of high-throughput assays from protein–protein interactions (Basso et al. 2005; Margolin et al. 2006a, b, Shi et al. 2018; Wang et al. 2009).

Algebraic Topology Features of the Reconstructed GBM–HRGs–MINWs

For each reconstructed network, denote the number of nodes as N_{node} and the number of edges as N_{edge}. To analyze the algebraic topological features of the HRGs–MINWs, we first introduced the concept of algebraic topology (Giusti et al. 2016; Sizemore et al. 2018). Traditional graph theory normally deals with partial or local parameters of the network structure, such as degree of vertices, simplex distribution, maximal simplex distribution, efficiency, and path length, et al. Unlike the graph theory approach, algebraic topology further seeks the intrinsic property of the global structure of the graph quantitatively (Curto 2016; Giusti et al. 2016). The most important and commonly used measurement is the homology group $H_q(K)$ (here q is the dimension of simplex) of the simplicial complex, for which the Betti numbers is the most important feature of the group (detailed information about the definition of homology group and Betti numbers is provided in methods). Simplified, the homology group quantitatively describes the integral structure of a geometric object, and the Betti numbers specifying the amount of "holes" or "cycles" in each dimension. Intuitively, the betti_1 represents the number of 1-dimensional cycles in a graph, while betti_2 represents the number of 2-dimensional cycle (similar to a hollow spherical structure). Similarly, for $q>2$, the Betti_q reflects the number of high-dimension holes. It should be noted that Betti numbers represent the amount of equiveillance classes of cycles, implying that for a same "hole", there would be more than one "cycles" surrounding it. For a class of k-cycles surrounding a same $(k+1)$-dim hole, they are equivalent but differ from one to another by the boundaries of some $(k+1)$-cliques.

For algebraic topological analysis, a network is considered as a simplicial complex, that is, a set of $(q+1)$ fully connected nodes (each pair in the $(q+1)$ nodes are connected) is considered as a q-simplex. By using a Z/Z2 *group* as coefficients *group*, we calculated the Betti numbers of the network with the JPlex algorithm (Adams et al. 2014; Lockwood and Krishnamoorthy 2015). By mapping the q-simplex to the coefficients group, finite number of q-simplexes formed a chain (q-chain) with the operation in the coefficients group. Closed chains are called cycles (q-cycle, see detail for the definition of chains and cycles in methods).

We then analyzed the algebraic topological features of the GBM–HRGs–MINW. Because during DPI process the simplexes with dimension $q\geq2$ are removed, the MINWs reconstructed with DPI only have 0- and 1-simplexes. We first investigated the $H_0(K)$ and $H_1(K)$ of the reconstructed networks. Take the Rembrandt GBM data set as a representative in the 5 datasets. There are totally 244 nodes (genes) and 406 edges (gene–gene interactions) in the HRGs–MINW. Algebraic topological analysis demonstrated that $\text{Betti}_0=19$ and

$Betti_1 = 181$. Because $Betti_0$ reflects the number of connected graphs, there are totally 19 connected graphs which are not connected together. $Betti_1 = 181$, indicating there are 181 loops in the HRGs–MINW (Table S2). The Betti numbers were similarly calculated in the other GBM–HRGs–MINWs.

Any network has its own topological structure. Therefore, it should be stringently examine that the topological features observed in the GBM–HRGs–MINW have significance rather than just random structures. To this end, Betti numbers in the random networks were investigated as controls (Giusti et al. 2015). Random networks with the same N_{node} and N_{edge} were constructed and their Betti numbers were calculated. A total of 1000 random networks (Random-MINWs) were produced and the distribution of Betti numbers was calculated. As a result, the 95% CI were obtained by examining the Betti number distribution (Table S2). According to the null hypothesis, $betti_1$ (= 181) in the Rembrandt HRGs–MINW is beyond the range of 95% CI (162.99–174.91) in Random-MINWs, indicating the observed structures had statistical significance and are not randomly produced ones. Similarly, although with different nodes and edges, most of the five databases had similar statistical results for $Betti_1$ results (Table S2).

Persistence Homolog Analysis Revealed Robust Algebraic Topological Structures of the GBM–HRGs–MINWs

Given the relatively large variances in biological data, the fluctuations of the MI values may apparently influence the network structure. To get more robust intrinsic topological structure of the network, we further performed persistence homology (PH) analysis by constructing filtrations of simplicial complexes (Curto 2016; Giusti et al. 2016; Sizemore et al. 2018). A filtration is defined by a growing network along with gradually addition of more simplexes to the simplicial complex (See detail in methods, Supplementary Fig. S1). As indicated by the name, "persistent" revealed the length of the cycles' life, and, reasonably, cycles with longer life would reflect more robust features of the network structure.

For the reconstructed GBM–HRGs–MINW, filtrations were built by adding edges according to their indexes ranked by MI values. Similarly, to rule out that the simplexes and cycles appeared in the GBM–HRGs–MINW are random structures, 1000 Random-MINWs with the same N_{node} and N_{edge} were analyzed as controls.

Take the Rembrandt dataset as a representative. During the PH analysis, Betti numbers for each step (or filtration value) were obtained, and at last curves of Betti number versus the filtration values were plotted (Betti curves, Fig. 2). In addition, the statistical distribution of Betti numbers and the corresponding 95% CI in the random networks in each step were also calculated and shown as Betti curves with upper and bottom bounds defined by the 95% CI (Fig. 2a). It was clearly that, the GBM–HRGs–MINW had Betti curves significantly different from that of random networks, as revealed by the fact they were deviated from the upper bounds of the curves from random networks (Fig. 2b). Analysis of other GBM databases exhibited similar results (Fig. 2b, Table 2; Supplementary Fig. S2). Strikingly, although with different N_{node} and N_{edge}, the Betti curves derived from the five independent GBM datasets showed great consistence to each other, implying this topological feature is intrinsic for GBM–HRGs–MINWs. The datasets used different gene ID list according to their gene Chip platform, resulting in different N_{node} and N_{edge} values in the datasets. To reduce the impact of different N_{node} values for the results and compare their Betti curves together, we used the common Up-HRGs in the five datasets to re-perform the analysis, and, as a result, the Betti curves of the five GBM datasets still exhibited great consistence to each other (Fig. 2c). All of these data implied that the GBM–HRGs–MINW have intrinsic rather than random algebraic topological structures.

PH Analysis Revealed Cycles with Significantly Long Life in the GBM–HRGs–MINWs

PH analysis revealed robust topological features of HRGs–MINW intrinsic for GBM. Next, we explored how to find cycles with statistical significance. To this end, the life lengths of all cycles produced in the filtration process were calculated, and the statistical distribution of life length in the Random-MINWs were analyzed as controls (Fig. 2d). Similarly, the distribution and 95% CI of cycle life lengths in the Random-MINWs were obtained. As revealed by the density curve of life length distribution (Fig. 2d), most of the cycles have very short life, and therefore the 95% bound in the accumulated distribution was used as thresholds for significantly long life cycles (Sig.-cycles, Table 2, Fig. 2e) in the GBM–HRGs–MINW. Taking the results from Rembrandt data as representative (Table 2), the 95% bound value is 226 in the Random-MINWs, and thus cycles with a life ≥ 226 were considered as Sig.-cycles. As a result, there are totally 65 Sig.-1-cycles (life ≥ 226), which would be potential molecular cycles contributing to the hypoxia regulated biological in GBM. Similarly, the number of simplexes, total Betti numbers and number of Sig.-cycles were also investigated in all of the five databases (Table 2). Although with different N_{node} and N_{edge} values, the five GBM datasets resulted in relatively comparable results, indicating the topological structures uncovered by our PH analysis reflected stable intrinsic features of the GBM–HRGs–MINW. For example, the numbers of Sig.-1-cycles ($N_{Sig.\text{-}1\text{-}cycles}$) in all of the datasets ranged from 27 to 65, while their normalized

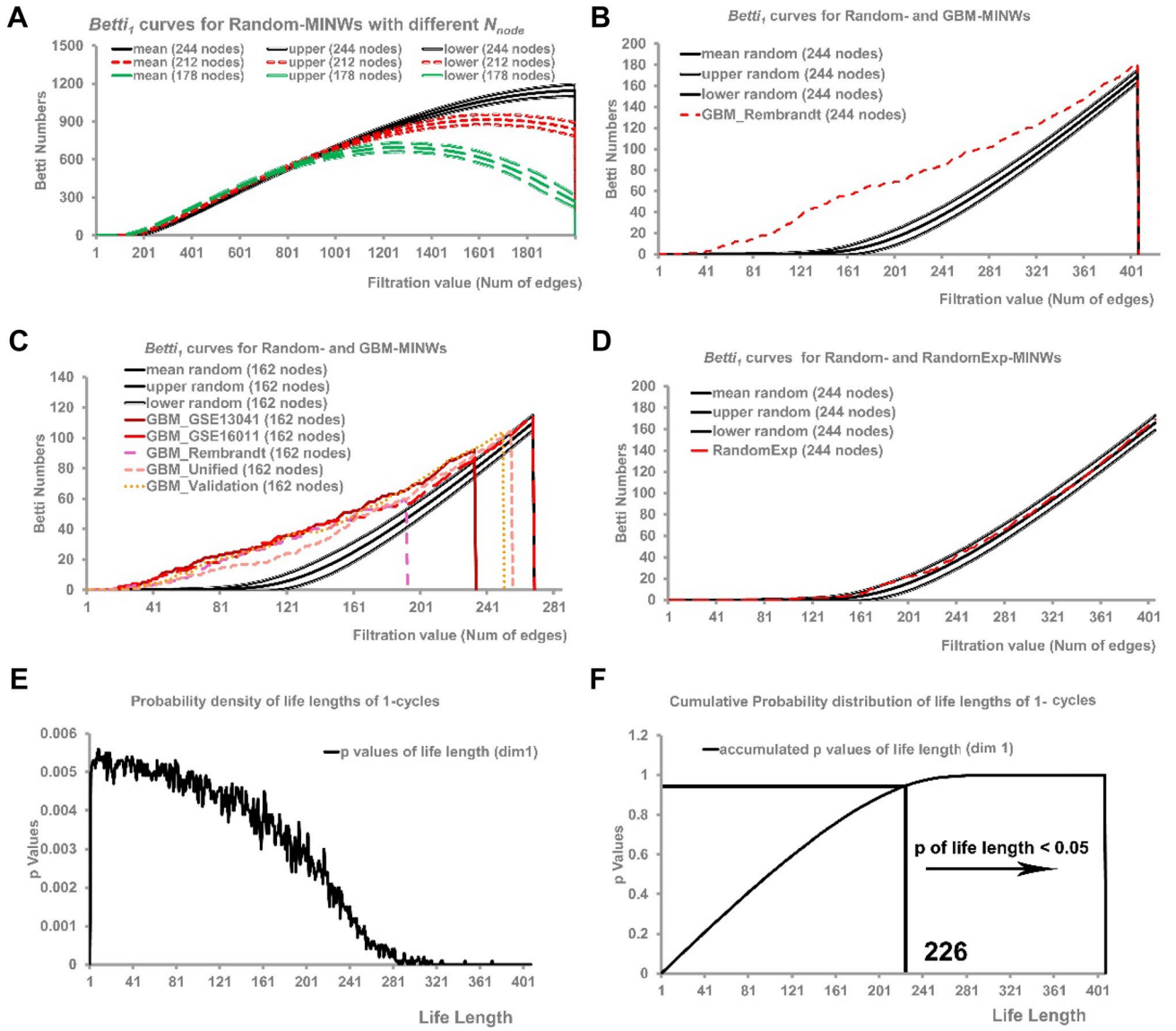
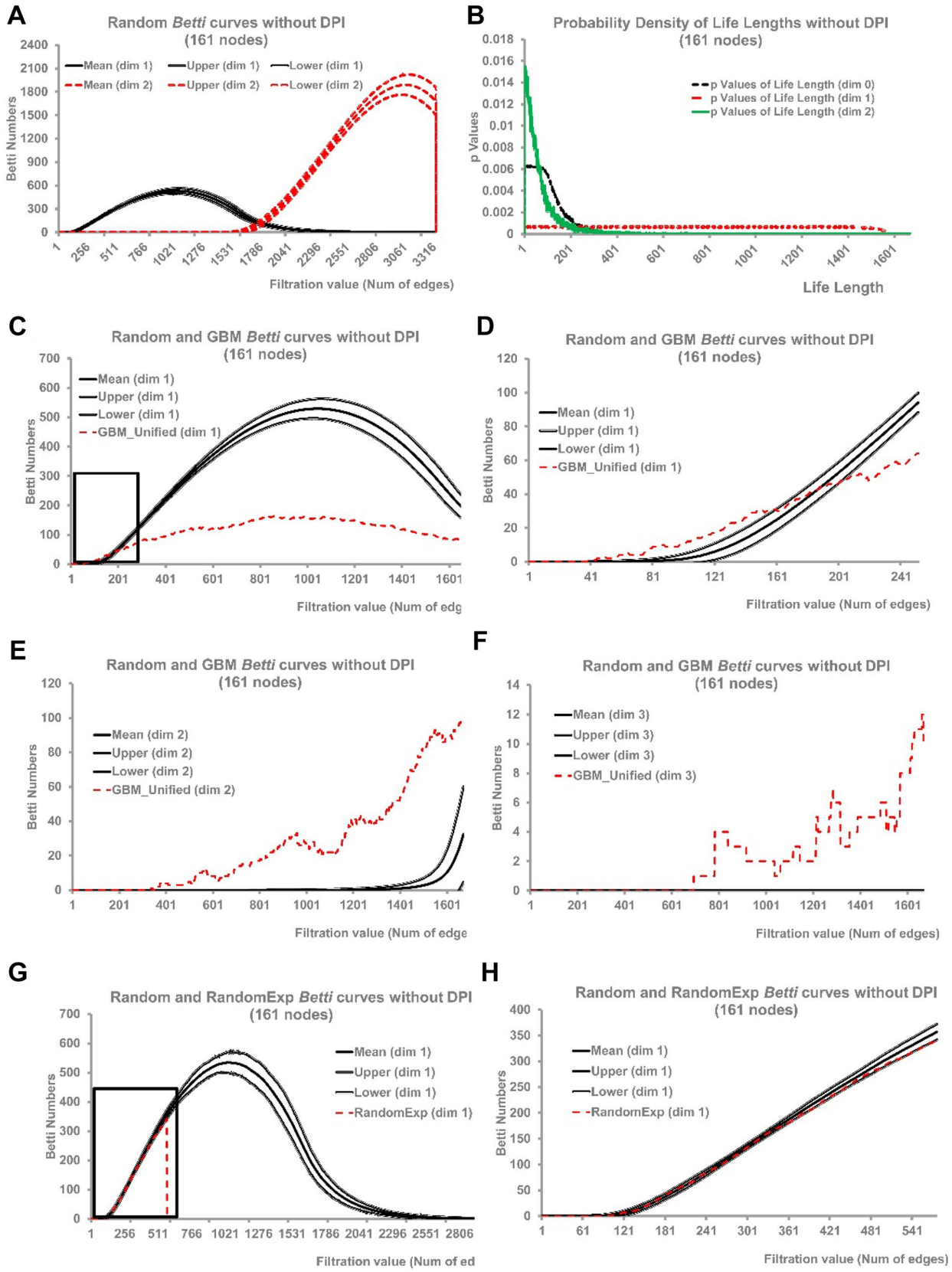


Fig. 2 Betti curves and life length distribution of the random network and GBM–HRG–MINWs reconstructed with DPI. **a** $Betti_1$ curves with different N_{node} values in the random networks. **b** $Betti_1$ curves of the GBM–HRG–MINWs reconstructed from Rembrandt GBM expression datasets. **c** All of the $Betti_1$ curves of GBM–HRG–MINW adjusted to have same N_{node} values exhibited great consistence to each other. **d**, **e** Probability density curve (**d**) and cumulative probability distribution (**e**) of life lengths for 1-cycles in random networks. **f** Betti curve of the HRGs–MINW derived from the random-expression data exhibited high consistence to that of random network, which is located exactly in the 95% CI range of the Betti curve from random network. The vertical red lines indicate the end point of the Betti curves

Table 2 Number of simplexes, cycles, and Sig.-cycles during the filtration process revealed persistently robust topological structures of the Up-HRGs–MINWs in GBM

Database	0-simplex	1-simplex	0-cycles	1-cycles	0-sig. cycles	1-sig. cycles	95% threshold of life length for 0 cycles	95% threshold of life length for 1 cycles
GBM_GSE13041	178	282	177	116	15	38	275	153
GBM_GSE16011	212	425	211	217	21	54	329	260
GBM_Rembrandt	244	406	243	181	28	65	375	226
GBM_Unified	174	285	173	120	12	27	274	157
GBM_Validation	174	285	173	127	19	38	268	158


A
Random Betti curves without DPI
(161 nodes)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
Mean (dim 2)
Upper (dim 2)
Lower (dim 2)
Betti Numbers
Filtration value (Num of edges)
B
Probability Density of Life Lengths without DPI
(161 nodes)
p Values of Life Length (dim 0)
p Values of Life Length (dim 1)
p Values of Life Length (dim 2)
p Values
Life Length
C
Random and GBM Betti curves without DPI
(161 nodes)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
GBM_Unified (dim 1)
Betti Numbers
Filtration value (Num of edg
D
Random and GBM Betti curves without DPI
(161 nodes)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
GBM_Unified (dim 1)
Betti Numbers
Filtration value (Num of edges)
E
Random and GBM Betti curves without DPI
(161 nodes)
Mean (dim 2)
Upper (dim 2)
Lower (dim 2)
GBM_Unified (dim 2)
Betti Numbers
Filtration value (Num of edge
F
Random and GBM Betti curves without DPI
(161 nodes)
Mean (dim 3)
Upper (dim 3)
Lower (dim 3)
GBM_Unified (dim 3)
Betti Numbers
Filtration value (Num of edges)
G
Random and RandomExp Betti curves without DPI
(161 nodes)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
RandomExp (dim 1)
Betti Numbers
Filtration value (Num of ed
H
Random and RandomExp Betti curves without DPI
(161 nodes)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
RandomExp (dim 1)
Betti Numbers
Filtration value (Num of edges)

◂**Fig. 3** Betti curves and life length distribution of the random network and GBM–HRG–MINWs reconstructed without DPI. **a** $Betti_1$ and $Betti_2$ curves ($N_{node}=161$) in the random networks. **b** Probability density curves for 0, 1, and 2-cycles in random networks ($N_{node}=161$). **c**–**f** Representative $Betti_1$, $Betti_2$ and $Betti_3$ curves for GBM–HRG–MINWs reconstructed from the Unified TCGA dataset without DPI. Image in (**d**) showed the detailed information of the boxed part in the curve in (**c**). **g**, **h** $Betti_1$ curve of HRG-MINW reconstructed from random-expression dataset, and image in (**h**) showed detailed information of the boxed part in the curve in (**g**)

values to the N_{node} ($N_{Sig.\text{-}1\text{-}cycles}/N_{node}$) ranged from 0.16 to 0.25, exhibiting a great consistence with each other.

Adjusted GBM–HRGs–MINWs Have Consistent Algebraic Topological Structures

Different genome genes, N_{node} and N_{edge} values would have impacts in the topological structures of reconstructed HRGs–MINW. To reduce the variances introduced by these differences, we next re-performed the analysis by using the adjusted datasets which contained the common genome set in all databases. The adjusted networks have about 228–341 1-simplexes and 26–37 1-cycles. To further reduce the influence of different N_{edge} values in the network topology, we used the same number of top edges ranked by their MI values to construct the filtration. As a result, all of the networks have quite consistent parameters of topological structures, such as total number of 1-simplex, 1-cycles, and Sig.-cycles (Table S3), further validating the results that the topological structures uncovered here are intrinsic features of the GBM–HRGs–MINW.

Topological Structures in the GBM–HRGs–MINWs are Not Caused by Algorithm Biases and Specific Degree Distributions (DD)

To further examine whether the consistent topological features of all the GBM–HRGs–MINW would be caused by biases originated from the algorithm, we generated random gene expression datasets (RandomExp), and reconstructed their HRGs–MINW (RandomExp-HRGs–MINWs) with the same procedure. As a result, the Betti curves of the RandomExp-HRGs–MINWs are located almost exactly in the range of the 95% CI of Random-MINWs (Fig. 2f), indicating the reconstruction algorithm did not bring perceptual biases to the topological structures.

In addition, to further examine whether the GBM–HRGs–MINWs have significant rather than random produced structures, we studied the Random-MINWs with the same degree distribution (DD) to the GBM–HRGs–MINWs (RandomSameDD–MINWs) with DPI. First, the DD of all the GBM–HRGs–MINWs with DPI exhibited consistency to each other (Supplementary Fig. S4A). Totally five kinds of RandomSameDD–MINWs were produced with identical DD to each of the five GBM–HRGs–MINWs with DPI. As a result, strikingly, the $Betti_1$ curves of the RandomSameDD–MINWs with DPI are almost completely overlapped with that of Random-MINWs (Fig. S4B-C). Together, these results indicated that the observed topological structures in GBM–HRGs–MINWs are not produced by Random-MINWs with specific DDs, and these data further strengthened the conclusion that GBM–HRGs–MINW had its own intrinsic topological structures.

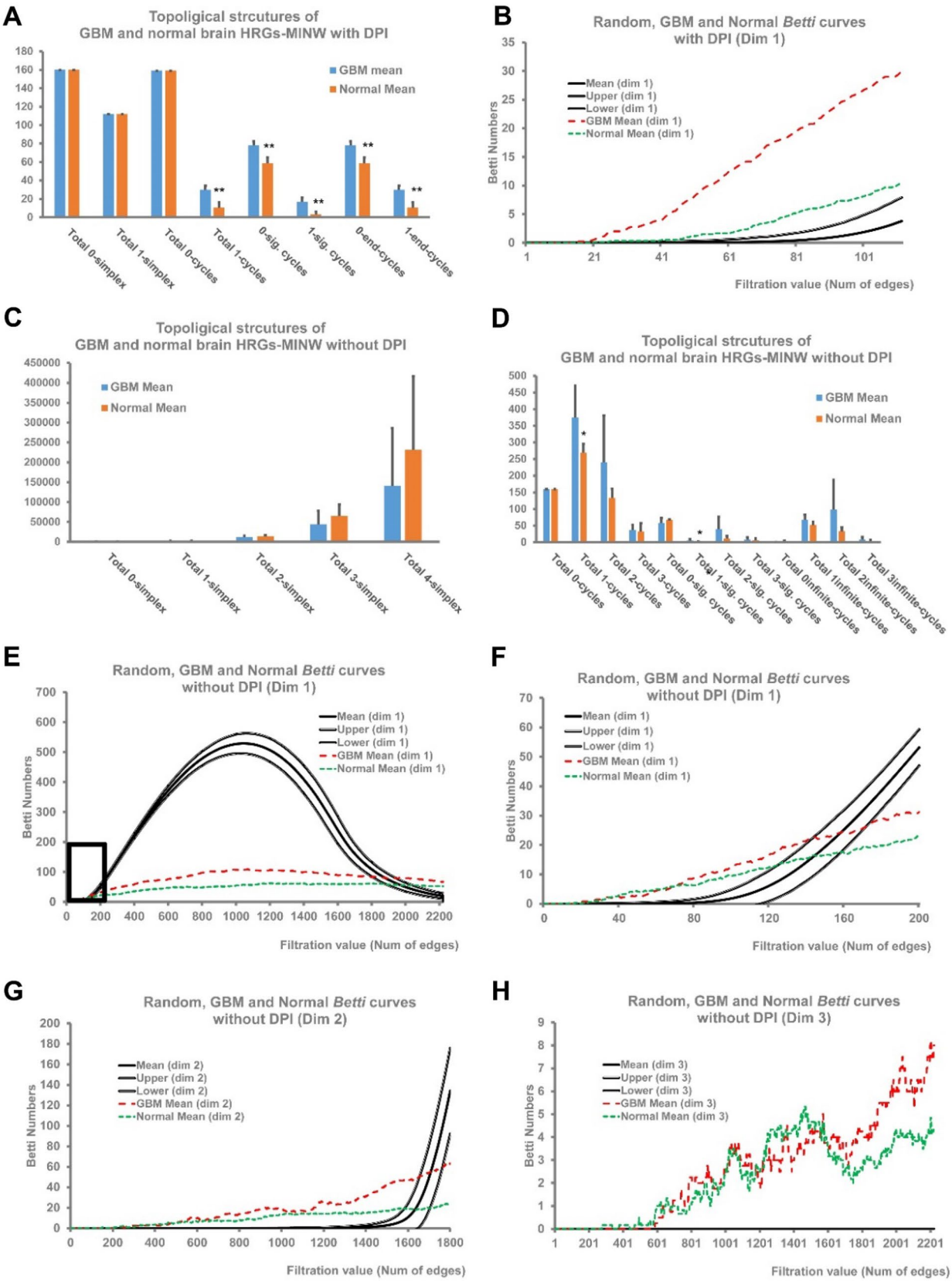
Higher Dimension Topological Structures of the GBM–HRGs–MINWs Reconstructed Without DPI

To investigate higher dimension topological structures in the GBM–HRGs–MINW, we used the network reconstructed by ARACNe without DPI (GBM–HRGs–MINW-NoDPI). In this scenario, the network described a relationship of the genes where the gene–gene interactions can be direct or indirect. The networks were constructed from the adjusted expression data, and 1000 random networks were produced as controls to get statistical distributions of the parameters of random network (Fig. 3a, b).

First, the GBM–HRGs–MINW-NoDPI showed much richer structures, and the highest dimension of cycles are 3 (Supplementary Table S4). In this situation, the Betti curves are apparently deviated from the 95% CI of the Random-MINWs (Fig. 3c–f). The $Betti_1$ curve in the Random-MINWs had an apparent peak along with the increase of edge numbers, while that from the GBM network exhibited relatively gentle curves with small amount of Betti numbers. When investigating the $Betti_1$ curves in detail, at the beginning of the filtration process (edges < 150, Fig. 3c), the Betti numbers of the GBM–HRGs–MINW-NoDPI are larger than that of the Random-MINWs, but it kept relative stable values across the whole filtration process (Fig. 3c, d).

The $Betti_2$ curves of GBM–HRGs–MINW-NoDPI exhibited more apparent differences from that of Random-MINWs (Fig. 3e). Similar to the situation of $Betti_1$ curves, ahead the ascending section of the curve, the $Betti_2$ curve in the GBM–HRGs–MINW-NoDPI are significantly larger than that of Random-MINWs (Fig. 3e). Similar situation was observed in the $Betti_3$, with no 3-cycles observed in the Random-MINWs until the observed edge numbers (Fig. 3f), demonstrating the GBM–HRGs–MINW-NoDPI also have significant high-dimension topological structures.

Next, further investigation of Betti curves derived from RandomExp-MINWs showed that they are located exactly in the 95% CI range of Betti curves of the Random-MINWs (Fig. 3g, h), ruling out the possibility that the observed topological structures are biases caused by the algorithm.


A
Topoligical strcutures of
GBM and normal brain HRGs-MINW with DPI
GBM mean
Normal Mean
Total 0-simplex
Total 1-simplex
Total 0-cycles
Total 1-cycles
0-sig. cycles
1-sig. cycles
0-end-cycles
1-end-cycles
B
Random, GBM and Normal Betti curves
with DPI (Dim 1)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
GBM Mean (dim 1)
Normal Mean (dim 1)
Betti Numbers
Filtration value (Num of edges)
C
Topoligical strcutures of
GBM and normal brain HRGs-MINW without DPI
GBM Mean
Normal Mean
Total 0-simplex
Total 1-simplex
Total 2-simplex
Total 3-simplex
Total 4-simplex
D
Topoligical strcutures of
GBM and normal brain HRGs-MINW without DPI
GBM Mean
Normal Mean
Total 0-cycles
Total 1-cycles
Total 2-cycles
Total 3-cycles
Total 0-sig. cycles
Total 1-sig. cycles
Total 2-sig. cycles
Total 3-sig. cycles
Total 0infinite-cycles
Total 1infinite-cycles
Total 2infinite-cycles
Total 3infinite-cycles
E
Random, GBM and Normal Betti curves
without DPI (Dim 1)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
GBM Mean (dim 1)
Normal Mean (dim 1)
Betti Numbers
Filtration value (Num of edges)
F
Random, GBM and Normal Betti curves
without DPI (Dim 1)
Mean (dim 1)
Upper (dim 1)
Lower (dim 1)
GBM Mean (dim 1)
Normal Mean (dim 1)
Betti Numbers
Filtration value (Num of edges)
G
Random, GBM and Normal Betti curves
without DPI (Dim 2)
Mean (dim 2)
Upper (dim 2)
Lower (dim 2)
GBM Mean (dim 2)
Normal Mean (dim 2)
Betti Numbers
Filtration value (Num of edges)
H
Random, GBM and Normal Betti curves
without DPI (Dim 3)
Mean (dim 3)
Upper (dim 3)
Lower (dim 3)
GBM Mean (dim 3)
Normal Mean (dim 3)
Betti Numbers
Filtration value (Num of edges)

◂**Fig. 4** GBM derived HRGs–MINW had algebraic topological structures significantly different from that of normal brains. **a**, **b** For the MINW reconstructed with DPI, GBM–HRGs–MINW had more total cycles, Sig-cycles and end-cycles than that of normal brains (**a**). PH analysis revealed that the averaged $Betti_1$ curve of GBM–HRGs–MINW was apparently different from that of normal brains (**b**). **c**, **d** For the MINW reconstructed without DPI, GBM–HRGs–MINW had a smaller number of simplexes, but more cycles in dimensions 1–3, compared to that of normal brains. **e**–**h** Averaged $Betti_1$, $Betti_2$, and $Betti_3$ curves for GBM and normal brain-HRG-MINWs. The image **f** showed the detailed information in the boxed part in (E)

Furthermore, the $Betti_1$ curves of RandomSameDD–MINWs exhibited only slight deviation from that of Random-MINWs after the peak of the curves (Fig. S4D), implying the observed topological structures in GBM–HRGs–MINW-NoDPI are not produced by specific DDs.

The Topological Structures of GBM–HRGs–MINW are Significantly Different from that Derived from Normal Brains

A further question is, whether the GBM–HRGs–MINWs have structures specific for the GBM disease. To this end, we analyzed the HRGs–MINW reconstructed by genome expression data derived from normal brains (Normal-HRGs–MINWs), by using six datasets [one from GSE11882 (Berchtold et al. 2008), one from GSE8919 (Myers et al. 2007), and four datasets of different brain regions (including cerebellar, frontal cortex, ponts, and temporal cortex) from GSE15745 (Gibbs et al. 2010)]. By analyzing the topological structures of the Normal-HRGs–MINWs, we found they had topological structures significantly different from that of GBM–HRGs–MINWs.

Similar to the above analysis, we first analyzed the network reconstructed with DPI. Interestingly, although with the same value of N_{node}, Normal-HRGs–MINWs had less amount of 1-simplexes, total cycles and Sig.-cycles, and cycles in the simplicial complex at the endpoint of filtration (end-cycles) (Supplementary Table S5).

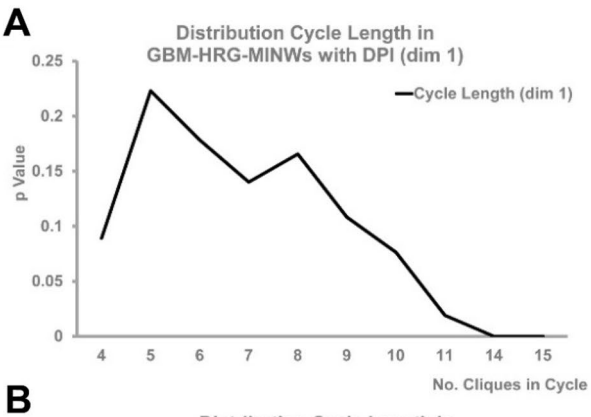
Next, we adjusted all of the GBM and normal brain MINWs to have the same N_{edge} value, by taking the same number of topped edges ranked by their MI values. As a result, the GBM–HRGs–MINWs and Normal-HRGs–MINWs showed significantly different topological structures, such as number of cycles, Sig.-cycles, and end-cycles ($p < 0.05$, Table S6). It should be noted that Normal-HRGs–MINWs derived from different datasets also exhibited consistent results to each other. Interestingly, GBM had significantly more 0-cycles and 1-cycles than normal brains (Table S5, Fig. 4a, b).

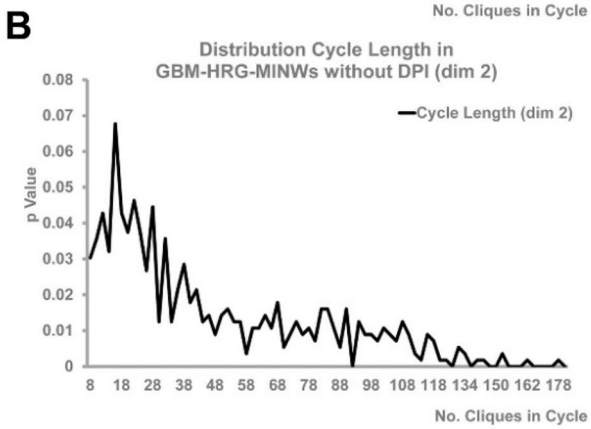
At last, we further analyzed networks reconstructed without DPI (Normal-HRGs–MINW-NoDPI) to investigate high dimensional structures. Normal-HRGs–MINW-NoDPI tend to have more simplexes in all dimensions (although not statistically significant, $p = 0.22$–0.44, Fig. 4c), while, like the situation in the DPI results, they had less cycles than the GBM–HRGs–MINW-NoDPI, especially in dimension-1 (Fig. 4d). Investigation of the averaged Betti curves for GBM and normal brain demonstrated that GBM-HRGs–MINW-NoDPI have more cycles than Normal-HRGs–MINW-NoDPI in dimension 1 and 2, and slightly in dimension 3 (Fig. 4e–h). Quite interestingly, when examining the Betti curves in detail (Fig. 4f), Betti curves of normal brain showed larger values at beginning and kept more stable values which are gradually become smaller than that of GBM–HRGs–MINWs (Fig. 4e, f), indicating that Betti curves of normal brains more deviated from Betti curves of Random-MINWs than GBMs did.

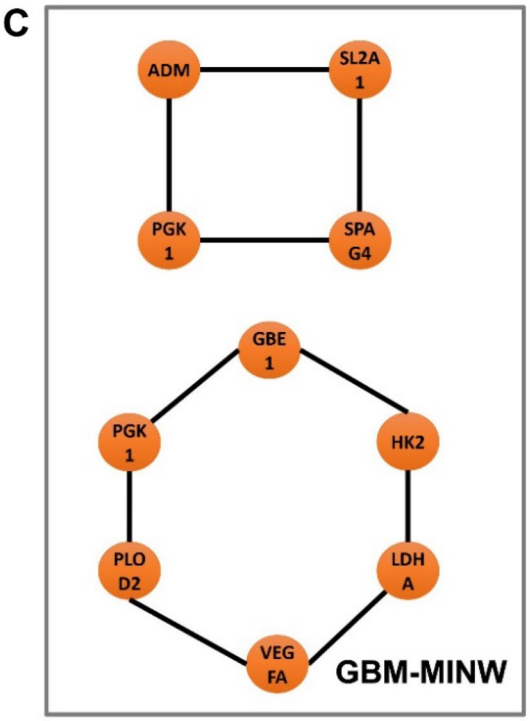
Algebraic Topological Analysis Revealed Richer Information About the GBM–HRGs–MINW Structures and Identified Key Cycles for GBM Tumorigenesis

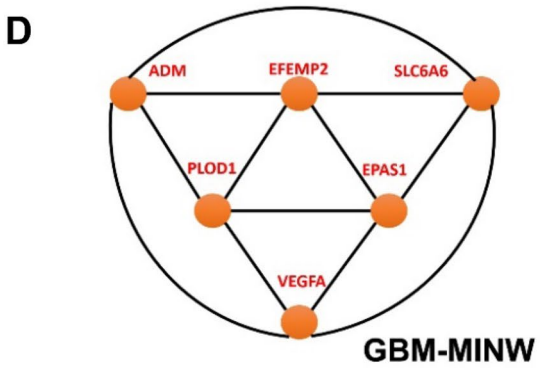
Different from results obtained by traditional graph analysis, algebraic topological analysis revealed much richer information about the integral features of network. As a representative result, we investigated the number of cycles involved for a gene (NC_{gene}), and the connectivity degree of the gene (D_{gene}). Genes with high values of D_{gene} did not necessarily had high value of NC_{gene}. Interestingly, some genes had very low D_{gene} but are involved in large number of cycles (Supplementary Table S7), and there are also genes exhibiting opposite feature. Nevertheless, the topology analysis identified potential novel targets that are more frequently involved in loops of cell signal transmission, rather than those that only have more interactions. For example, some hypoxia responsive genes contributing to the tumorigenesis of GBM are identified, including ADM, VEGFA, EPAS1, PGK1, et al. In addition, a cohort of novel potential targets for hypoxia in GBM were identified, such as PLOD1, PLOD2, GPRC5A, LDHA, GBE1, SPAG4, et al. Because one gene can be viewed as 0-clique, we can also investigate the NC_{clique} of 1- or higher dimension cliques. For example, the $NC_{1\text{-clique}}$ were identified and revealed the most frequently appeared paired gene interactions (Supplementary Table S8), containing many unreported potential interactions, such as ADM-SPAG4, GBE1-PLOD2, et al.

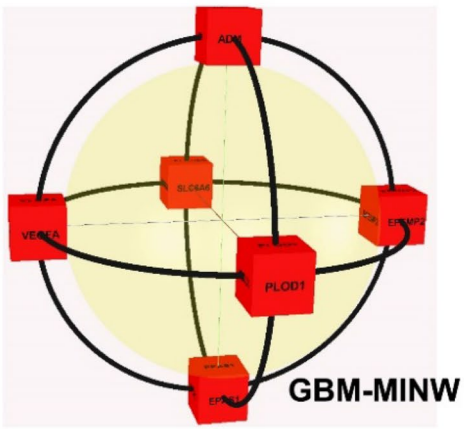
To quantitatively evaluate the importance of the cycles, we assigned a rank value to each cycle, by calculating the average values of the NC_{gene} of all the nodes in the cycle. Analysis of the distribution of 1-cycle lengths revealed that the cycle lengths are ranged from 4 to 10 and peaked at 5 (Fig. 5a). For higher topological structures, the 2-cycles in the GBM–HRGs–MINW-NoDPI were studied (Fig. 5b). The 1-cycles, such as the representative ones in

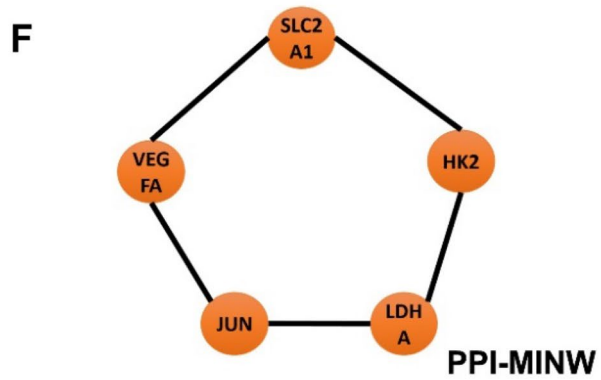

A
Distribution Cycle Length in
GBM-HRG-MINWs with DPI (dim 1)
Cycle Length (dim 1)
p Value
0.25
0.2
0.15
0.1
0.05
0
4 5 6 7 8 9 10 11 14 15
No. Cliques in Cycle

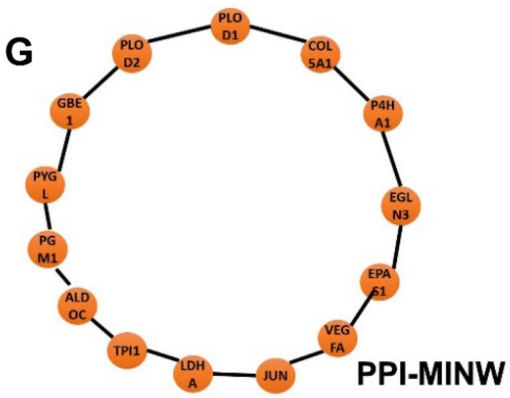

B
Distribution Cycle Length in
GBM-HRG-MINWs without DPI (dim 2)
Cycle Length (dim 2)
p Value
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
8 18 28 38 48 58 68 78 88 98 108 118 134 150 162 178
No. Cliques in Cycle


C
ADM
SL2A 1
PGK 1
SPA G4
GBE 1
PGK 1
HK2
PLO D2
LDH A
VEG FA
GBM-MINW


D
ADM
EFEMP2
SLC6A6
PLOD1
EPAS1
VEGFA
GBM-MINW


E
ADM
SLC6A6
VEGFA
EFEMP2
PLOD1
EPAS1
GBM-MINW


F
SLC2 A1
VEG FA
HK2
JUN
LDH A
PPI-MINW


G
PLO D1
PLO D2
COL 5A1
GBE 1
P4H A1
PYG L
EGL N3
PG M1
EPA S1
ALD OC
VEG FA
TPI1
LDH A
JUN
PPI-MINW

◂**Fig. 5** Representative 1- and 2-cycles (1- and 2-dimension cycles) in the reconstructed GBM–HRGs–MINWs and PPI-HRGs–MINWs. **a** 1-Cycles length distribution in the GBM–HRGs–MINWs reconstructed with DPI. **b** 2-Cycles length distribution in the GBM–HRGs–MINWs reconstructed without DPI. **c** Representative 1-cycles, which can be viewed as traditional loops. **d**, **e** Representative 2-cycles (**d**), which can be considered as extensions of 1-dimenional loops and viewed as 2-dimensional sphere surfaces (**e**). **f**, **g** The PPI-HRGs–MINWs contained sub-cycles-3genes [HK2, LDHA, VEGFA] (**f**) and [GBE1-PLOD2-VEGFA] that are identified in the GBM–HRGs–MINWs 1-cycles

the top rank (ADM–SLC2A1–PGK1–SPAG4–ADM) and (GBE1–HK2–LDHA–VEGFA–PLOD2–PGK1–GBE1) (Fig. 5c), can be viewed as loops. The 2-cycles can be viewed as spherical structures, where signals can transfer in any directions along the surface but not inside of them (Fig. 5d, e), reflecting certain rules controlling the information flow. Differentially, Normal-HRGs–MINWs not only have different topological structures, but also have different gene compositions of the top ranked cycles (Supplementary Table S9).

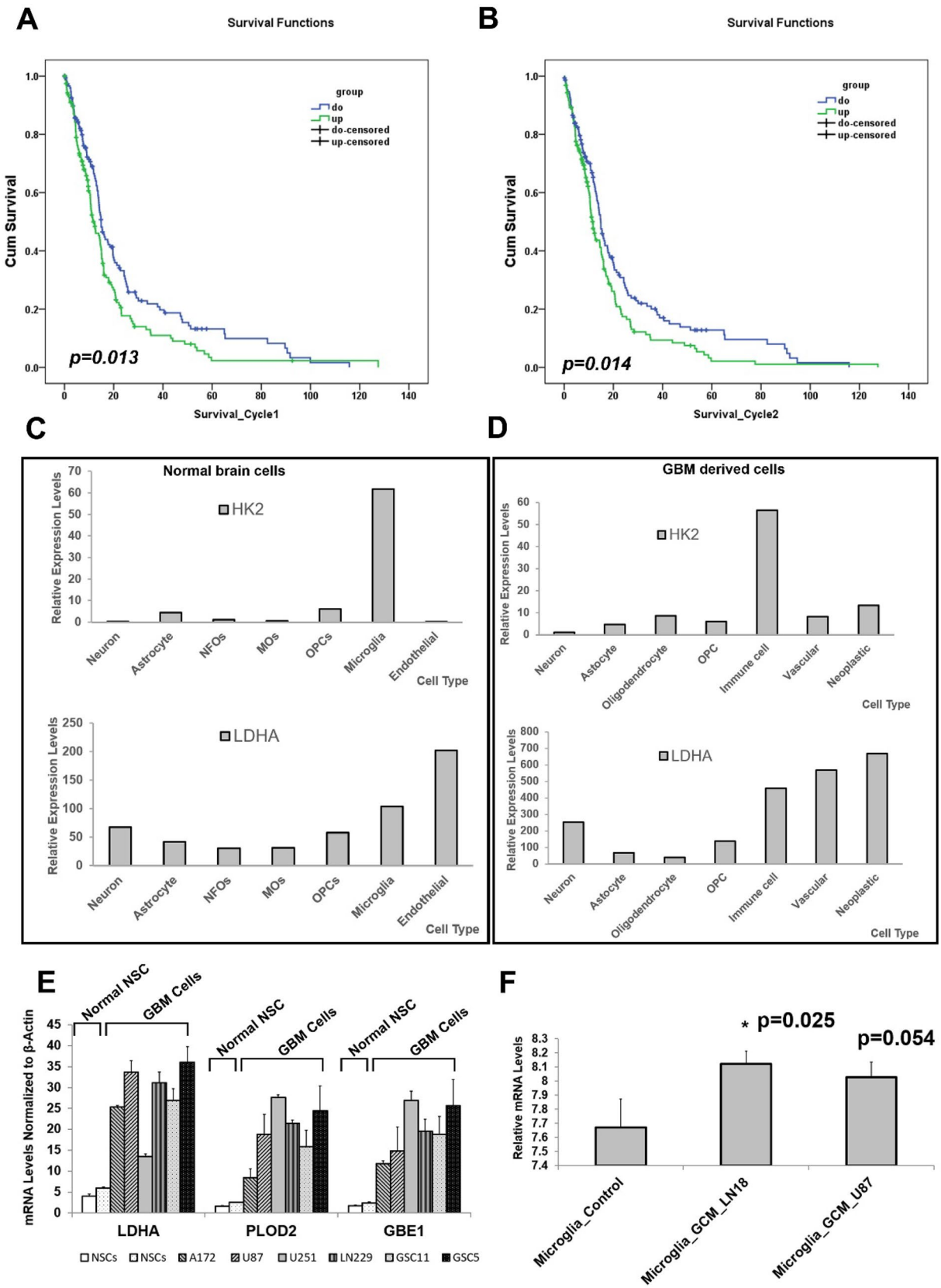
Topologically Identified Cycles Have Biological and Clinical Indications

What are the biological or clinical implications of the topological structures identified above? An apparent significance is that the topological structures revealed novel interactions between genes or, more interestingly, biological modules. Firstly, we observed the clinical significance of the above representative 1-cycles. These genes are poor prognostic factors for the patient's survival (Fig. S5, A–D). In addition, to investigate the influence of cycles for survival, the sum of the rank values ordered by expression levels for each gene (SRExp) were defined as the expression index for the cycle. Analysis of the survival of patients with high and low SRexp revealed that high level of SRExp of the 1-cycles were significantly predictors for poor survival (Fig. 6a, b). By interrogating the Rembrandt database, almost all of the genes were significantly highly expressed in gliomas than in normal brain, and especially more highly expressed in GBMs (Fig. S6). In addition, all of the genes in the representative 1-cycles exhibited positive correlation to each other (Fig. S7, Table S10), implying potential interactions among these genes.

To further validate the biological indications of our identified cycles, we reconstructed protein–protein interaction (PPI) MINW (PPI-MINW) of the HRGs (PPI-HRGs–MINW) by using the STRING database (Szklarczyk et al. 2015) (Fig. S8). Next, algebraic topological analysis were performed to identify 1-cycles and 2-cycles of the PPI-HRGs–MINW. Then we examined the overlaps between the cycles of the PPI-HRGs–MINW and our identified GBM–HRGs–MINWs. As a result, we totally identified 48 sub-cycles containing 3 genes (sub-cycles-3genes, Figs. 5c, f, g, S9) and 129 sub-cycles-2genes that were presented in both of the MINWs. As representatives, the sub-cycles-3genes included [HK2, LDHA, VEGFA] (Fig. 5c, f, Tables S11, S12), [GBE1–PLOD2–VEGFA] (Fig. 5c, g, Tables S11, S12), which are key parts of the identified representative 1-cycles (GBE1–HK2–LDHA–VEGFA–PLOD2–PGK1–GBE1) (Figs. 5c, S9). It should be noted that, the PPI-HRGs–MINW is not disease specific and may missing other kinds of direct interactions, such as protein or RNA binding to promoters or in other DNA/RNA regions. Nevertheless, the identified sub-cycles in the PPI-HRGs–MINW provided biological evidences that the cycles in the GBM-GRGs-MINWs identified by PH analysis would played certain biological roles for the GBM.

GBM–HRGs–MINW Contained Novel Loops Implying Interplays Between Warburg Effect, Immune Regulation, and Angiogenesis

We further analyzed the potential functions of the genes in the loops. Interestingly, these genes exhibited different biological modules and/or features. For example, the top ranked representative loop (GBE1–HK2–LDHA–VEGFA–PLOD2–PGK1–GBE1) contained key genes for the Warburg effect, including LDHA (Liu et al. 2018; Pathria et al. 2018; Valvona et al. 2016), PGK1 (Li et al. 2016), HK2 (Lis et al. 2016), and gene GBE1, which has not been investigated in GBM. In addition, there are also potential genes that are involved in the immune response of GBM, such as HK2 (Li et al. 2018). By examining the expression profile of HK2 in different cell types in normal (GSE52564) (Zhang et al. 2014) and GBM datasets (GSE84465) (Darmanis et al. 2017), we found that HK2 was almost exclusively highly expressed in microglial cells in normal brain and immune cells in GBM (Fig. 6c, d). Interestingly, the genes involved in the Warburg effect also showed specific expression in certain cell types: for example, LDHA was enriched in both microglial and vascular cells (Fig. 6c, d). In addition, some of the genes (including LDHA, PLOD2, GBE1) were highly expressed in GBM cell lines, including U87, U251, and glioblastoma stem like cells (GSCs), than in normal neural stem cells (NSCs) (Fig. 6e), indicating specific roles of these genes for GBM cells. Furthermore, by examining a database investigating the glioma-induced polarization of microglia (Walentynowicz et al. 2018), HK2 was significantly up-regulated in microglial cells after treatment with GBM cell culture medium (Fig. 6f). Another loop (ADM–SLC2A1–PGK1–SPAG4–ADM) revealed similar multi-components interplays in GBM: ADM and


A
Survival Functions
group
do
up
do-censored
up-censored
Cum Survival
p=0.013
Survival_Cycle1
B
Survival Functions
group
do
up
do-censored
up-censored
Cum Survival
p=0.014
Survival_Cycle2
C
Normal brain cells
HK2
Relative Expression Levels
Neuron
Astrocyte
NFOs
MOs
OPCs
Microglia
Endothelial
Cell Type
LDHA
Relative Expression Levels
Neuron
Astrocyte
NFOs
MOs
OPCs
Microglia
Endothelial
Cell Type
D
GBM derived cells
HK2
Relative Expression Levels
Neuron
Astocyte
Oligodendrocyte
OPC
Immune cell
Vascular
Neoplastic
Cell Type
LDHA
Relative Expression Levels
Neuron
Astocyte
Oligodendrocyte
OPC
Immune cell
Vascular
Neoplastic
Cell Type
E
Normal NSC
GBM Cells
mRNA Levels Normalized to β-Actin
LDHA
PLOD2
GBE1
NSCs
NSCs
A172
U87
U251
LN229
GSC11
GSC5
F
* p=0.025
p=0.054
Relative mRNA Levels
Microglia_Control
Microglia_GCM_LN18
Microglia_GCM_U87

◂**Fig. 6** **a**, **b** The expression index of a cycle is defined as the sum of the rank values ordered by expression levels for each gene (SRExp) in the cycle. For the two 1-cycles indicated in Fig. 5c, GBM patients with high SRExp values had a significantly poor prognosis. **c**, **d** Genes in the 1-cycles, such as HK2 and LDHA are specifically expressed in certain cell types in both normal brain cells (**c**) and GBM cells (**d**). **e** Genes in the 1-cycles such as LDHA, PLOD2 and GBE1 are more highly expressed in GBM cells than in normal NSCs. **f** Treatment of microglial cells with GBM cell culture medium resulted in significant up-regulation of HK2 expressions

SLC2A1 were both specifically highly expressed in vascular cells, while SPAG4 was enriched in neuronal, vascular and immune cells (Supplementary Figs. S6 and S10), and they are all involved in Warburg effect or aerobic glycolysis by interacting with PGK1. These identified loops revealed intrinsic complicated cycled communications among multiple cell components (including tumor cells, vascular cells, immune cells, neuronal cells, et al.) and with Warburg effect. All of the cells together with Warburg effect are essential cell types or programs for the tumorigenesis of GBM, while the detailed interplays have not been well understood. There are still a lot of potential loops identified, which give clues to further decipher the complicated interplays of multiple programs in GBM.

Discussion

A considerable number of potential targets and results about gene interactions in GBM have been obtained by experimental research. However, it is more and more widely recognized that gliomas are caused by multiple genetic or epigenetic changes (Brennan et al. 2013; Chen et al. 2014; Mao et al. 2009b; TCGA 2008). Therefore, targeting several key genes may not be sufficient to reverse the disease. In addition, acquisition of resistance to anti-angiogenesis treatment indicates that the MINW in GBM have much great elastic adaptability to the therapeutic interventions. For example, anti-VEGF monoclonal antibody Bevacizumab only have limited effect on GBM patients and caused drug resistance rapidly (Piao et al. 2013). In addition, anti-Notch therapy with a gamma-secretase inhibitor can lead to therapy resistance by suppressing Hedgehog via Hes1 mediated inhibition of Gli1 transcription (Schreck et al. 2010), disclosing an ingenious adaptability of the molecular network. Therefore, there are increasing requirements to study the global gene interaction network in tumors, especially with the rapid advancement of high-throughput data acquisition technologies. However, although a lot of research revealed gene to gene interactions and the significance of certain genes for tumors, little is known about the global structures of the MINW of GBM.

There are two major obstacles to study the global structures of gene interaction networks. First, ideal approach to get exact gene interaction networks is lacking, especially the MINW in special contexts (Kleaveland et al. 2018; Markmiller et al. 2018; Shi et al. 2018). Progressions are made in recent years in experimental and theoretical approaches (Kermany et al. 2018; Kleaveland et al. 2018; Li and Izpisua Belmonte 2018; Markmiller et al. 2018; Shi et al. 2018). Given the large number of genes in a network, it is not feasible to get the exact interactions between all paired genes or molecules experimentally, and theoretical methods are especially important for the reconstruction of MINW (Plaisier et al. 2016; Shi et al. 2018). Information theory based analysis is proved to have important roles in MINW reconstruction in recent years (Plaisier et al. 2016; Shi et al. 2018; Wang et al. 2009). Actually, in some studies, information theory based analysis can get results comparable to that from high-throughput experiments (Basso et al. 2005; Wang et al. 2009). In addition, besides paired gene interactions, there are still triplet and multiple gene interactions in the MINW, which are more difficult to get the interaction patterns by pure biological experiments.

Another difficult is the analysis of the reconstructed MINW. For a given reconstructed network, how to decipher its structure and its relevance to tumor biological is also a, probably more, difficult task (Bullmore and Sporns 2009; Sporns and Betzel 2016). A direct difficult issue is how to describe the structure of MINW, especially when it contains thousands of genes. The similar problem is encountered in mathematics about 100 years ago, when the algebraic geometry is established to quantitatively study the integral structure of a geometry independent of metric features such as length, angle et al. The precise mathematical language to study the structure of geometries is based on the abstract algebraic concept—group, an essential tool in modern mathematics and physics. Interestingly, MINWs are geometries that are determined by the structures rather than their measures. Therefore, they are useful tools to study the intrinsic features of MINW in depth. For example, algebraic topological studies have been performed in neuronal activity correlation networks in the brain (Giusti et al. 2015), and intrinsic organization patterns of the network were detected. Nevertheless, little, if any, is known about the geometric organization of the MINWs in tumor.

Given the huge amount of computation for a network consisted of thousands of genes, a prominent problem for the MINW study at first is how to determine a gene set responsible for certain functions, such as hypoxia regulated genes. Because of the great variance in biological study, we introduced GHG to get a significant gene set from several independent datasets. The particular advantage of the GHG is that it can quantitatively tell about exactly how many genes with certain overlapping criteria can reach the statistical

significance (Mao and Xue 2018). Based on the GHG, we got more than 200 genes that are up-regulated or down-regulated under hypoxia, which is then used for further study.

To get intrinsic topological structures of the HRGs–MINW in GBM, one fundamental issue is how to distinguish special topological structures in the network from nonspecific noise. To this end, we studied random networks with the same N_{node} and N_{edge} parameters of the GBM–HRG–MINW, and even RandomSameDD–MINWs with the same DD of the GBM–HRG–MINW. Strikingly, both the Random-MINWs and the RandomSameDD-MINWs generated structures with quite stable Betti curves, while, in contrast, GBM–HRGs–MINWs had significantly different Betti curves which kept stable Betti values across the filtration process. Importantly, by examining five independent GBM datasets, the HRGs–MINW in all of the datasets exhibited great consistent Betti curves and Sig.-cycles. Therefore, the topological structures represented by Betti numbers and Sig.-cycles reflected invariant intrinsic features of the GBM–HRGs–MINW. Next, to exclude the possibility that the algorithms used to reconstruct the network would generate biases, we further investigated a randomly expressed gene dataset. Interestingly, the RandomExp-MINWs exhibited high consistent topological features to the Random-MINWs, confirming the unbiasedness of the algorithm, and further validated that the detected topological structures are intrinsic for GBM. In addition, analysis of the MINWs reconstructed with Down-HRGs (GBM-Down-HRGs–MINWs) also revealed that the Betti curves of these MINWs exhibited high consistency to each other and are significantly different from that of Random-MINWs (Fig. S11A). Furthermore, GBM-Down-HRGs–MINWs and GBM-Up-HRGs–MINWs had different Betti curves (Fig. S11B), indicating that PH analysis also revealed function (or gene set) specific structures.

To further explore whether the HRGs–MINW have topological structures specific for the disease, we next studied the Normal-HRGs–MINW. Similarly, several independent datasets were used to reconstruct the Normal-HRGs–MINWs. First of all, all of the Normal-HRGs–MINW also have consistent topological structures to each other. An apparent characteristic of the Normal-HRGs–MINW compared to that of GBM–HRGs–MINWs is that they have much less number of simplexes and cycles in dimension 1 or higher, indicating less loops in the Normal-HRGs–MINWs. To get comparable results to GBM, we used adjusted networks by using top edges ranked by the MI values to make all GBM- and Normal-HRGs–MINW to have same N_{node} and N_{edge}. As a result, GBM–HRGs–MINWs have more cycles than that of Normal-HRGs–MINW, indicating GBM changed their MINW organization pattern during progression into malignant tumors. Besides altered expression levels of tumorigenic genes or tumor-suppressor genes as has been widely studied, the GBM also exhibited apparent changes of the organization of gene information flow. However, as discussed above, some key genes can be compensated by redundant loops in the gene network (Rizzolio et al. 2018; Schreck et al. 2010; Tamura et al. 2017). It is possible that the topological nature of MINW in GBM is the more essential and invariant feature of the disease, and interventions to change the topological structure may reverse its tumorigenic behavior.

Different from traditional graph analysis, algebraic topological analysis revealed more abundant and intrinsic characteristics of the network (Curto, 2016; Giusti et al. 2015, 2016). Higher dimension cycles can be viewed as extensions of traditional loops. For example, a 2-cycle reflects gene information flow in a 2-dimensional closed manifold (Fig. 5d, e), where genes can propagate their information flow in all directions in the surface but not inside of the sphere, uncovering a more flexible but controlled gene interaction pattern that can't be described by traditional 1-dimension loops. In the present study, the HRGs–MINW reconstructed without DPI exhibited apparent two and higher dimension persistent cycles.

It should be noted that algebraic topology used in the present work identified equivalence classes of cycles in various dimensions, different from traditional loop-finding algorithms which extract all loops (Sizemore et al. 2018; Tiernan 1970). Considering the high variability of biological data, it would be questioned that small changes of the edge weights or their ordering would change the results. It has been demonstrated that persistent homology is quite robust to noise, as small changes in the order of the edges will results in small changes in the persistence diagram (Cohen-Steiner et al. 2007; Sizemore et al. 2018). These results, together with that our results were based on several independent datasets, heavily confirmed the validation of the present data, revealing intrinsic structures of the GBM–HRGs–MINW.

Limitations of the Present Results and Future Perspectives

First, as discussed above, reconstruction of MINW is still a main obstacle for network analysis. Any network, no matter how precise the method is used, is an approximation of the real one. The methods used in the present study have been proven to be useful to reconstruct networks (Basso et al. 2005; Margolin et al. 2006a, b; Wang et al. 2009). More accurate approaches, including experimental and theoretical methods to detect direct physical interactions between genes or molecules, would provide more accurate MINW, and analysis based on these networks would reveal more intrinsic topological features of the GBM–HRGs–MINW. Notably, our PPI analysis provided important biological basis that the identified cycles would have functional roles

in tumor (Figs. 5f, g, S8, S9). Therefore, based on the methods adopted here, we have first clearly addressed that GBM have a HRGs–MINW with significant typological structures different from that of random networks and normal brains.

Second, given the importance of hypoxia for the tumorigenicity of GBM, the current study investigated the HRGs–MINW containing more than 200 genes, while it would be constructive to study bigger networks, and ultimately, the whole genome network. Nevertheless, because our gene set is derived from stringent procedure, and the gene interaction relationship is calculated from whole genome dataset, we firstly uncovered intrinsic topological features of the hypoxia driven program in GBM.

Third, how to validate the biological relevance of the topological structures of the HRG-MINW and, more interestingly, how to target it to reverse the disease into a normal pattern as a therapeutic strategy are all open questions for future research. Possible perspectives include: by changing the global gene interaction pattern, or by targeting novel genes which are essential for topological structure of the MINW in GBM, the GBM can be reprogramed into normal cells.

For the many novel potential loops identified in our analysis, we have examined the biological significance of a few representative ones in clinical samples and GBM cell lines. There are still a lot of loops providing clues of novel gene interactions in GBM, and future studies based on these results would be useful to fundamentally decipher the tumorigenic mechanisms of GBM.

Conclusion

As a conclusion, by stringently select a set of HRGs in GBM, we reconstructed the HRGs–MINW in GBM and normal brains. Algebraic topological analysis showed that GBM-MINW had special topological structures, significantly different from random networks and networks reconstructed from random gene expressions. PH analysis further revealed Sig-cycles in the GBM-MINW. In addition, GBM–HRGs–MINW have more simplexes and cycles compared to that of Normal-HRGs–MINW, indicating GBM changed the network organization significantly after malignant transformation. A lot of potential interactions between different cell types or programs in GBM are uncovered and some of them were validated by PPI analysis. In summary, GBM–HRG–MINWs have significantly special topological structures, revealing intrinsic natures of the disease.

Funding This study was funded by National Natural Science Foundation of China (Grant Numbers: 81502143, 51675411, 81671302) and foundation of fourth military medical university (2018JSTS05).

Compliance with Ethical Standards

Conflicts of interest All authors declare that they have no conflicts of interest.

Ethical Approval The study protocol was approved by the institutional review board of Xijing Hospital of the Fourth Military Medical University. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Research Involving Human Participants and/or Animals This article does not contain any studies with animals performed by any of the authors.

Informed Consent Informed consent was obtained from all individual participants included in the study.

References

Adams H, Tausz A, Vejdemo-Johansson M (2014) javaPlex: a research software package for persistent (Co) homology. Springer, Berlin

Basso K, Margolin AA, Stolovitzky G, Klein U, Dalla-Favera R, Califano A (2005) Reverse engineering of regulatory networks in human B cells. Nat Genet 37:382–390

Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, Beach T, Miller C, Troncoso J, Trojanowski JQ, Zielke HR, Cotman CW (2008) Gene expression changes in the course of normal brain aging are sexually dimorphic. Proc Natl Acad Sci USA 105:15605–15610

Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu CJ, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, Yung WK, Zhang W, Sougnez C, Mikkelsen T, Aldape K, Bigner DD, Van Meir EG, Prados M, Sloan A, Black KL, Eschbacher J, Finocchiaro G, Friedman W, Andrews DW, Guha A, Iacocca M, O'Neill BP, Foltz G, Myers J, Weisenberger DJ, Penny R, Kucherlapati R, Perou CM, Hayes DN, Gibbs R, Marra M, Mills GB, Lander E, Spellman P, Wilson R, Sander C, Weinstein J, Meyerson M, Gabriel S, Laird PW, Haussler D, Getz G, Chin L, Network TR (2013) The somatic genomic landscape of glioblastoma. Cell 155:462–477

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nat Rev Neurosci 10:186–198

Carro MS, Lim WK, Alvarez MJ, Bollo RJ, Zhao X, Snyder EY, Sulman EP, Anne SL, Doetsch F, Colman H, Lasorella A, Aldape K, Califano A, Iavarone A (2010) The transcriptional network for mesenchymal transformation of brain tumours. Nature 463:318–325

Chen JC, Alvarez MJ, Talos F, Dhruv H, Rieckhof GE, Iyer A, Diefes KL, Aldape K, Berens M, Shen MM, Califano A (2014) Identification of causal genetic drivers of human disease through systems-level analysis of regulatory networks. Cell 159:402–414

Cohen-Steiner D, Edelsbrunner H, Harer J (2007) Stability of persistence diagrams. Discret Comput Geom 37:103–120

Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York

Curto C (2016) What can topology tell us about the neural code? arXiv :160501905

Darmanis S, Sloan SA, Croote D, Mignardi M, Chernikova S, Samghababi P, Zhang Y, Neff N, Kowarsky M, Caneda C, Li G, Chang SD, Connolly ID, Li Y, Barres BA, Gephart MH, Quake SR (2017) Single-cell RNA-Seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. Cell Rep 21:1399–1410

Du R, Lu KV, Petritsch C, Liu P, Ganss R, Passegue E, Song H, Vandenberg S, Johnson RS, Werb Z, Bergers G (2008) HIF1alpha induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion. Cancer Cell 13:206–220

Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, Arepalli S, Dillman A, Rafferty IP, Troncoso J, Johnson R, Zielke HR, Ferrucci L, Longo DL, Cookson MR, Singleton AB (2010) Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 6:e1000952

Giusti C, Pastalkova E, Curto C, Itskov V (2015) Clique topology reveals intrinsic geometric structure in neural correlations. Proc Natl Acad Sci USA 112:13455–13460

Giusti C, Ghrist R, Bassett DS (2016) Two's company, three (or more) is a simplex: algebraic-topological tools for understanding higher-order structure in neural data. J Comput Neurosci 41:1–14

Gravendeel LA, Kouwenhoven MC, Gevaert O, de Rooi JJ, Stubbs AP, Duijm JE, Daemen A, Bleeker FE, Bralten LB, Kloosterhof NK, De Moor B, Eilers PH, van der Spek PJ, Kros JM, Sillevis Smitt PA, van den Bent MJ, French PJ (2009) Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res 69:9065–9072

Joseph JV, Conroy S, Pavlov K, Sontakke P, Tomar T, Eggens-Meijer E, Balasubramaniyan V, Wagemakers M, den Dunnen WF, Kruyt FA (2015) Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the HIF1alpha-ZEB1 axis. Cancer Lett 359:107–116

Kermany DS, Goldbaum M, Cai W, Valentim CCS, Liang H, Baxter SL, McKeown A, Yang G, Wu X, Yan F, Dong J, Prasadha MK, Pei J, Ting MYL, Zhu J, Li C, Hewett S, Dong J, Ziyar I, Shi A, Zhang R, Zheng L, Hou R, Shi W, Fu X, Duan Y, Huu VAN, Wen C, Zhang ED, Zhang CL, Li O, Wang X, Singer MA, Sun X, Xu J, Tafreshi A, Lewis MA, Xia H, Zhang K (2018) Identifying medical diagnoses and treatable diseases by image-based deep learning. Cell 172(1122–1131):e1129

Kim J, Woo AJ, Chu J, Snow JW, Fujiwara Y, Kim CG, Cantor AB, Orkin SH (2010) A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell 143:313–324

Kleaveland B, Shi CY, Stefano J, Bartel DP (2018) A network of noncoding regulatory RNAs acts in the mammalian brain. Cell 174(350–362):e317

Koh MY, Lemos R Jr, Liu X, Powis G (2011) The hypoxia-associated factor switches cells from HIF-1alpha- to HIF-2alpha-dependent signaling promoting stem cell characteristics, aggressive tumor growth and invasion. Cancer Res 71:4015–4027

Kucharzewska P, Christianson HC, Welch JE, Svensson KJ, Fredlund E, Ringner M, Morgelin M, Bourseau-Guilmain E, Bengzon J, Belting M (2013) Exosomes reflect the hypoxic status of glioma cells and mediate hypoxia-dependent activation of vascular cells during tumor development. Proc Natl Acad Sci USA 110:7312–7317

Lee Y, Scheck AC, Cloughesy TF, Lai A, Dong J, Farooqi HK, Liau LM, Horvath S, Mischel PS, Nelson SF (2008) Gene expression analysis of glioblastomas identifies the major molecular basis for the prognostic benefit of younger age. BMC Med Genom 1:52

Li M, Izpisua Belmonte JC (2018) Deconstructing the pluripotency gene regulatory network. Nat Cell Biol 20:382–392

Li Z, Bao S, Wu Q, Wang H, Eyler C, Sathornsumetee S, Shi Q, Cao Y, Lathia J, McLendon RE, Hjelmeland AB, Rich JN (2009) Hypoxia-inducible factors regulate tumorigenic capacity of glioma stem cells. Cancer Cell 15:501–513

Li X, Jiang Y, Meisenhelder J, Yang W, Hawke DH, Zheng Y, Xia Y, Aldape K, He J, Hunter T, Wang L, Lu Z (2016) Mitochondria-translocated PGK1 functions as a protein kinase to coordinate glycolysis and the TCA cycle in tumorigenesis. Mol Cell 61:705–719

Li Y, Lu B, Sheng L, Zhu Z, Sun H, Zhou Y, Yang Y, Xue D, Chen W, Tian X, Du Y, Yan M, Zhu W, Xing F, Li K, Lin S, Qiu P, Su X, Huang Y, Yan G, Yin W (2018) Hexokinase 2-dependent hyperglycolysis driving microglial activation contributes to ischemic brain injury. J Neurochem 144:186–200

Lis P, Dylag M, Niedzwiecka K, Ko YH, Pedersen PL, Goffeau A, Ulaszewski S (2016) The HK2 dependent "Warburg effect" and mitochondrial oxidative phosphorylation in cancer: targets for effective therapy with 3-bromopyruvate. Molecules 21:1730

Liu J, Chen G, Liu Z, Liu S, Cai Z, You P, Ke Y, Lai L, Huang Y, Gao H, Zhao L, Pelicano H, Huang P, McKeehan WL, Wu CL, Wang C, Zhong W, Wang F (2018) Aberrant FGFR tyrosine kinase signaling enhances the Warburg effect by reprogramming LDH isoform expression and activity in prostate cancer. Cancer Res 78:4459–4470

Lockwood S, Krishnamoorthy B (2015) Topological features in cancer gene expression data. In: Pacific symposium on biocomputing. pp 108–119

Madhavan S, Zenklusen JC, Kotliarov Y, Sahni H, Fine HA, Buetow K (2009) Rembrandt: helping personalized medicine become a reality through integrative translational research. Mol Cancer Res 7:157–167

Mahase S, Rattenni RN, Wesseling P, Leenders W, Baldotto C, Jain R, Zagzag D (2017) Hypoxia-mediated mechanisms associated with antiangiogenic treatment resistance in glioblastomas. Am J Pathol 187:940–953

Man J, Yu X, Huang H, Zhou W, Xiang C, Huang H, Miele L, Liu Z, Bebek G, Bao S, Yu JS (2018) Hypoxic induction of vasorin regulates Notch1 turnover to maintain glioma stem-like cells. Cell Stem Cell 22(104–118):e106

Mao XG, Xue XY (2018) General hypergeometric distribution: a basic statistical distribution for the number of overlapped elements in multiple subsets drawn from a finite population. arXiv :1808.06924

Mao XG, Zhang X, Xue XY, Guo G, Wang P, Zhang W, Fei Z, Zhen HN, You SW, Yang H (2009a) Brain tumor stem-like cells identified by neural stem cell marker CD15. Transl Oncol 2:247–257

Mao XG, Zhang X, Zhen HN (2009b) Progress on potential strategies to target brain tumor stem cells. Cell Mol Neurobiol 29:141–155

Mao XG, Guo G, Wang P, Zhang X, Xue XY, Zhang W, Fei Z, Jiang XF, Yan M (2010) Maintenance of critical properties of brain tumor stem like cells after cryopreservation. Cell Mol Neurobiol 30:775–786

Mao XG, Yan M, Xue XY, Zhang X, Ren HG, Guo G, Wang P, Zhang W, Huo JL (2011) Overexpression of ZNF217 in glioblastoma contributes to the maintenance of glioma stem cells regulated by hypoxia-inducible factors. Lab Investig 91:1068–1078

Mao XG, Song SJ, Xue XY, Yan M, Wang L, Lin W, Guo G, Zhang X (2013a) LGR5 is a proneural factor and is regulated by OLIG2 in glioma stem-like cells. Cell Mol Neurobiol 36:851–865

Mao XG, Xue XY, Wang L, Zhang X, Yan M, Tu YY, Lin W, Jiang XF, Ren HG, Zhang W, Song SJ (2013b) CDH5 is specifically activated in glioblastoma stem like cells and contributes to vasculogenic mimicry induced by hypoxia. Neuro oncology 15:865–879

Mao XG, Xue XY, Wang L, Zhang X, Yan M, Tu YY, Lin W, Jiang XF, Ren HG, Zhang W, Song SJ (2013c) CDH5 is specifically

activated in glioblastoma stemlike cells and contributes to vasculogenic mimicry induced by hypoxia. Neuro-oncology 15:865–879

Mao XG, Wang C, Liu DY, Zhang X, Wang L, Yan M, Zhang W, Zhu J, Li ZC, Mi C, Tian JY, Hou GD, Miao SY, Song ZX, Li JC, Xue XY (2016) Hypoxia upregulates HIG2 expression and contributes to bevacizumab resistance in glioblastoma. Oncotarget 7:47808–47820

Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, Califano A (2006a) ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinform 7(Suppl 1):S7

Margolin AA, Wang K, Lim WK, Kustagi M, Nemenman I, Califano A (2006b) Reverse engineering cellular networks. Nat Protoc 1:662–671

Markmiller S, Soltanieh S, Server KL, Mak R, Jin W, Fang MY, Luo EC, Krach F, Yang D, Sen A, Fulzele A, Wozniak JM, Gonzalez DJ, Kankel MW, Gao FB, Bennett EJ, Lecuyer E, Yeo GW (2018) Context-dependent and disease-specific diversity in protein interactions within stress granules. Cell 172(590–604):e513

Michaelsen SR, Staberg M, Pedersen H, Jensen KE, Majewski W, Broholm H, Nedergaard MK, Meulengracht C, Urup T, Villingshoj M, Lukacova S, Skjoth-Rasmussen J, Brennum J, Kjaer A, Lassen U, Stockhausen MT, Poulsen HS, Hamerlik P (2018) VEGF-C sustains VEGFR2 activation under bevacizumab therapy and promotes glioblastoma maintenance. Neuro Oncol 20:1462–1474

Myers AJ, Gibbs JR, Webster JA, Rohrer K, Zhao A, Marlowe L, Kaleem M, Leung D, Bryden L, Nath P, Zismann VL, Joshipura K, Huentelman MJ, Hu-Lince D, Coon KD, Craig DW, Pearson JV, Holmans P, Heward CB, Reiman EM, Stephan D, Hardy J (2007) A survey of genetic human cortical gene expression. Nat Genet 39:1494–1499

Nakagawa M, Shaffer AL 3rd, Ceribelli M, Zhang M, Wright G, Huang DW, Xiao W, Powell J, Petrus MN, Yang Y, Phelan JD, Kohlhammer H, Dubois SP, Yoo HM, Bachy E, Webster DE, Yang Y, Xu W, Yu X, Zhao H, Bryant BR, Shimono J, Ishio T, Maeda M, Green PL, Waldmann TA, Staudt LM (2018) Targeting the HTLV-I-regulated BATF3/IRF4 transcriptional network in adult T cell leukemia/lymphoma. Cancer Cell 34(286–297):e210

Nissou MF, El Atifi M, Guttin A, Godfraind C, Salon C, Garcion E, van der Sanden B, Issartel JP, Berger F, Wion D (2013) Hypoxia-induced expression of VE-cadherin and filamin B in glioma cell cultures and pseudopalisade structures. J Neurooncol 113:239–249

Pathria G, Scott DA, Feng Y, Sang Lee J, Fujita Y, Zhang G, Sahu AD, Ruppin E, Herlyn M, Osterman AL, Ronai ZA (2018) Targeting the Warburg effect via LDHA inhibition engages ATF4 signaling for cancer cell survival. EMBO J 37:e99735

Piao Y, Liang J, Holmes L, Henry V, Sulman E, de Groot JF (2013) Acquired resistance to anti-VEGF therapy in glioblastoma is associated with a mesenchymal transition. Clin Cancer Res 19:4392–4403

Plaisier CL, O'Brien S, Bernard B, Reynolds S, Simon Z, Toledo CM, Ding Y, Reiss DJ, Paddison PJ, Baliga NS (2016) Causal mechanistic regulatory network for glioblastoma deciphered using systems genetics network analysis. Cell Syst 3:172–186

Rizzolio S, Battistini C, Cagnoni G, Apicella M, Vella V, Giordano S, Tamagnone L (2018) Downregulating neuropilin-2 triggers a novel mechanism enabling EGFR-dependent resistance to oncogene-targeted therapies. Cancer Res 78:1058–1068

Schreck KC, Taylor P, Marchionni L, Gopalakrishnan V, Bar EE, Gaiano N, Eberhart CG (2010) The Notch target Hes1 directly modulates Gli1 expression and Hedgehog signaling: a potential mechanism of therapeutic resistance. Clin Cancer Res 16:6060–6070

Shi J, Zhao J, Li T, Chen L (2018) Detecting direct associations in a network by information theoretic approaches. Sci China Math 62:823–838

Sizemore AE, Giusti C, Kahn A, Vettel JM, Betzel RF, Bassett DS (2018) Cliques and cavities in the human connectome. J Comput Neurosci 44:115–145

Sporns O, Betzel RF (2016) Modular brain networks. Annu Rev Psychol 67:613–640

Stupp R, Mason WP, van den Bent MJ, Weller M, Fisher B, Taphoorn MJ, Belanger K, Brandes AA, Marosi C, Bogdahn U, Curschmann J, Janzer RC, Ludwin SK, Gorlia T, Allgeier A, Lacombe D, Cairncross JG, Eisenhauer E, Mirimanoff RO (2005) Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352:987–996

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452

Tamura R, Tanaka T, Miyake K, Yoshida K, Sasaki H (2017) Bevacizumab for malignant gliomas: current indications, mechanisms of action and resistance, and markers of response. Brain Tumor Pathol 34:62–77

TCGA (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455:1061–1068

Tiernan JC (1970) An efficient search algorithm to find the elementary circuits of a graph. Commun ACM 13:722–726

Urup T, Staunstrup LM, Michaelsen SR, Vitting-Seerup K, Bennedbaek M, Toft A, Olsen LR, Jonson L, Issazadeh-Navikas S, Broholm H, Hamerlik P, Poulsen HS, Lassen U (2017) Transcriptional changes induced by bevacizumab combination therapy in responding and non-responding recurrent glioblastoma patients. BMC Cancer 17:278

Valvona CJ, Fillmore HL, Nunn PB, Pilkington GJ (2016) The regulation and function of lactate dehydrogenase A: therapeutic potential in brain tumor. Brain Pathol 26:3–17

Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, Sarkaria JN, Brennan C, Kahn A, Spellman PT, Wilson RK, Speed TP, Gray JW, Meyerson M, Getz G, Perou CM, Hayes DN (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17:98–110

Walentynowicz KA, Ochocka N, Pasierbinska M, Wojnicki K, Stepniak K, Mieczkowski J, Ciechomska IA, Kaminska B (2018) In search for reliable markers of glioma-induced polarization of microglia. Front Immunol 9:1329

Wang K, Saito M, Bisikirska BC, Alvarez MJ, Lim WK, Rajbhandari P, Shen Q, Nemenman I, Basso K, Margolin AA, Klein U, Dalla-Favera R, Califano A (2009) Genome-wide identification of post-translational modulators of transcription factor activity in human B cells. Nat Biotechnol 27:829–839

Wang Q, He Z, Huang M, Liu T, Wang Y, Xu H, Duan H, Ma P, Zhang L, Zamvil SS, Hidalgo J, Zhang Z, O'Rourke DM, Dahmane N, Brem S, Mou Y, Gong Y, Fan Y (2018) Vascular niche IL-6 induces alternative macrophage activation in glioblastoma through HIF-2alpha. Nat Commun 9:559

Xu H, Rahimpour S, Nesvick CL, Zhang X, Ma J, Zhang M, Zhang G, Wang L, Yang C, Hong CS, Germanwala AV, Elder JB, Ray-Chaudhury A, Yao Y, Gilbert MR, Lonser RR, Heiss JD, Brady RO, Mao Y, Qin J, Zhuang Z (2015) Activation of hypoxia signaling induces phenotypic transformation of glioma cells: implications for bevacizumab antiangiogenic therapy. Oncotarget. https://doi.org/10.18632/oncotarget.3592

Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, Phatnani HP, Guarnieri P, Caneda C, Ruderisch N, Deng S, Liddelow SA, Zhang C, Daneman R, Maniatis T, Barres BA, Wu JQ (2014) An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J Neurosci 34:11929–11947