DOI: 10.1002/bdr2.1336

ORIGINAL ARTICLE

WILEY Birth Defects Research THE TERATOLOGY SOCIETY EST. 1960

Network correlation analysis revealed potential new mechanisms for neural tube defects beyond folic acid

Xiaoya Gao[1,2] | Richard H. Finnell[3,4] | Hongyan Wang[2,5] | Yufang Zheng[1,2,5]

[1] Institute of Developmental Biology & Molecular Medicine, School of Life Sciences, Fudan University, Shanghai, China

[2] State Key Laboratory of Genetic Engineering, Collaborative Innovation Center of Genetics and Development, School of Life Sciences, Fudan University, Shanghai, China

[3] Departments of Molecular and Cellular Biology and Medicine, Baylor College of Medicine, Houston, Texas

[4] Collaborative Innovation Center for Genetics & Development, School of Life Sciences, Fudan University, Shanghai, China

[5] Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai, China

Correspondence
Yufang Zheng, Obstetrics and Gynecology Hospital, Institute of Reproduction and Development, Fudan University, Shanghai, China.
Email: zhengyf@fudan.edu.cn

Funding information
NSFC, Grant/Award Number: 81741048, 31771669, and 81430005; the Commission for Science and Technology of Shanghai Municipality, Grant/Award Number: 17JC1400902 (to Drs. Zheng and Wang); NIH, Grant/Award Number: HD081216, HD08380, and HD067244; the Changjiang Scholarship (for Dr. Finnell)

Background: Neural tube defects (NTDs) are clinically significant congenital malformations which are known to be folic acid (FA) responsive, such that supplementation significantly reduces the prevalence of NTDs. Nonetheless, some individuals fail to respond to FA supplementation; hence NTDs remain a significant public health concern. The mechanisms that underlie the beneficial effects of FA supplementation remain poorly understood. Mouse models have been used extensively to study the mechanisms driving neural tube closure (NTC).

Methods: Microarray data of GSE51285 was downloaded from the NCBI GEO database, which contains the RNA expression profiles of livers from five NTD mouse mutants (heterozygous females) and their corresponding wildtype (WT) controls. Those five NTD mutants have different responsiveness to FA supplementation. The differentially expressed genes (DEGs) between NTD heterozygous and WT mice, as well as the DEGs between FA-responsive and FA-resistant mutants were carefully examined. Weighted gene correlation network analysis (WGCNA) was performed in order to identify genes with high correlations to either FA responsiveness or NTDs, respectively.

Results: In total, we identified 18 genes related to the pathogenesis of NTDs, as well as 55 genes related to FA responsiveness. Eight more candidate genes (*Abcc3, Gsr, Gclc, Mthfd1, Gart, Bche, Slc25a32*, and *Slc44a2*) were identified by examining the DEGs of those genes involved in the extended folate metabolic pathway between FA-responsive and FA-resistant mutants.

Conclusions: Those genes are involved in mitochondrial choline metabolism, de novo purine synthesis, and glutathione generation, suggesting that formate, choline, and manipulating antioxidant levels may be effective interventions in FA-resistant NTDs.

KEYWORDS
bioinformatics, folic acid, microarray, neural tube defects, WGCNA

1 | INTRODUCTION

Neural tube defects (NTDs) are a type of human birth defects with variable severity, ranging from relatively mild cases of spina bifida occulta to severe cases of craniorachischisis (Greene & Copp, 2014). The prevalence of NTDs is 0.2 to 14 per 1000 live births worldwide (Mitchell, 2005). The causes of NTDs are many, including both genetic and environmental factors. Many nutrition and maternal metabolic conditions have been demonstrated to contribute to the

etiology of NTDs, including folate deficiency, maternal hyperthermia, maternal diabetes and obesity (Wilde, Petersen, & Niswander, 2014).

Folic acid (FA) supplementation has been reported to reduce the prevalence of NTDs by approximately 70% (MRC_Vitamin_Study_Research_Group, 1991). Despite the fact that it has been more than fifty years since Hibbard and Smithells implicated maternal folate deficiency as a risk factor for NTDs (Hibbard et al., 1965); and 20 years or more since FA has been recognized as a crucial nutritional supplement for pregnant women (Czeizel & Dudas, 1992), the mechanism underlying FA's ability to prevent NTDs is not well understood. Furthermore, there remain 30% of NTDs that cannot be prevented by FA supplementation (Crider, Bailey, & Berry, 2011), and again the etiology of these non-FA responsive NTDs remains unknown (Shimoji et al., 2013).

Mouse models are widely used in the study of NTDs, as the neural tube closure (NTC) processes in mice is very similar to that in humans, and mechanisms underlying NTC and NTDs may be informed by use of experimental models (Greene & Copp, 2005; Greene et al., 2014; Harris & Juriloff, 2007, 2010; Juriloff & Harris, 2000). Currently, there are close to 400 mouse mutants reported to exhibit NTD phenotypes (Greene et al., 2005, 2014; Harris et al., 2007, 2010). As is the case in humans, the mouse mutants also have different sensitivity to FA supplementation. Some mutants, such as *Pax3*$^{Sp/Sp}$, *Folr1*$^{-/-}$, and *Cart1*$^{-/-}$, are reported to be FA responsive (Juriloff et al., 2000), meaning that pregnant dams receiving FA supplementation have fewer NTD affected progeny and the severity of the NTD affected fetuses were lessened. On the other hand, some mutants, including *Grhl3*$^{Ct/Ct}$, *Fkbp8*$^{-/-}$, *Nog*$^{-/-}$, are not FA responsive, and no beneficial effect of maternal FA treatment was observed (Harris, 2009).

In the present study, we downloaded the microarray data GSE51285, which contained the RNA expression profiles from livers of five groups of NTD mutants and their corresponding wildtype (WT) controls. Among the NTD mutants, *Pax3*$^{Sp/Sp}$ (Wlodarczyk, Tang, Triplett, Aleman, & Finnell, 2006) and *Lrp6*$^{Cd/Cd}$ (Carter, Ulrich, Oofuji, Williams, & Ross, 1999) were reported to be FA-responsive mutants, while *Grhl3*$^{ct/ct}$(Greene & Copp, 1997), *Apob*^{tm1Unc}(Homanics et al., 1995) and *Vangl2*$^{Lp/Lp}$ (Nakouzi & Nadeau, 2014) mutants were considered to be FA-resistant. The differentially expressed genes (DEGs) between NTDs heterozygous and WT mice, as well as the DEGs between FA-responsive and FA-resistant mutants were evaluated. A network based gene screening method, Weighted Gene Correlation Network Analysis (WGCNA), was applied to identify genes with high correlations to FA responsiveness or NTD risk, respectively. The aim of this study was to identify new candidate genes that might reveal the underlying mechanism of how FA prevents NTDs.

2 | MATERIALS AND METHODS

2.1 | Data source

The gene expression profile of GSE51285 was downloaded from NCBI Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) database. The gene expression assays for GSE51285 were performed on the Affymetrix GPL1261 platform (Affymetrix Mouse Genome 430 2.0 Array). Series GSE51285 contains five groups of NTD mutants (Supporting Information Table S1; *Pax3*^{Sp} [Splotch], *Apob*^{tm1Unc}, *Lrp6*^{Cd} [Crooked tail], *Grhl3*^{ct} [Curly tail], *Vangl2*^{Lp} [Loop tail]): each group contains four biological replicates of heterozygous females and their corresponding wild-type (WT) female pups (also with four biological replicates) at 6–8 weeks of age. Since *Pax3*^{Sp} and *Apob*^{tm1Unc} are both maintained on the C57BL/6J (B6) inbred strain background, there are five mutant groups but only four WT groups. The raw data (CEL files) were downloaded and analyzed with R packages (v.3.3.0).

2.2 | Quality check, data preprocessing, and differentially expressed genes screening

Initial quality checks were conducted using QCplot of Simpleaffy package in R (Wilson & Miller, 2005). Background correction and noise filtration were performed using the GC Robust Multichip Averaging (GCRMA) method in the Affy package in R (Irizarry et al., 2003). Following the normalization, the Empirical Bayes method (Smyth, 2004) was used to carry out gene specific *t* test, which identified differentially expressed genes (DEGs) between heterozygous and WT control samples within each group. This process was performed using the Limma package in R (Smyth, 2004).

2.3 | Construction of weighted gene co-expression network

The weighted gene correlation network analysis (WGCNA) package (v1.51) of R was used to describe the correlation pattern among genes across microarray samples (Langfelder & Horvath, 2008). This analysis was performed twice using corresponding DEGs (NTDs vs WT; FA-responsive vs FA-resistant) that were identified using the GCRMA method with a p value $< .05$. First, we calculated Pearson correlation coefficients for all gene pairs across all microarray samples and constructed a similarity matrix. Second, we converted the matrix of correlation into an adjacency matrix by a power function based on a scale-free topology criterion (Zhao et al., 2010). The soft-threshold power used in the criterion was 12

TABLE 1 The list of NTD candidate genes by WGCNA

Gene	P.Weighted_value	Target %id	WGA score	Mouse fetal CNS[a]	Human fetus CNS[b]
Hdac5	0.00015393	93.67	100	+	+
Atg12	0.000291589	88.65	95.54	+	+
Atp6v1e1	0.000550413	98.67	94.51	+	+
Bnip3	1.57E-06	88.77	87.04	+	+
D1Ertd622e	0.000385217	0	0	NR	/
Eef1e1	0.00010165	87.93	100	+	+
Fam126b	0.000331096	86.18	100	NR	NR
Lasp1	0.000330514	95.44	100	+	+
Lztfl1	0.000773938	90.97	88.03	NR	+
Nupl1	0.000515276	91.92	98.10	NR	+
Plekhm1	0.000443156	82.22	100	NR	+
Ppp1cc	0.000761654	96.90	100	+	+
Rwdd4a	2.94E-06	90.96	100	NR	+
Sc5d	0.000906533	83.61	100	+	+
Snip1	0.000243768	82.51	100	+	+
Thumpd3	0.000591776	80.20	100	+	+
Vps26a	5.99E-06	89.69	100	+	+
Zbtb41	0.000201526	93.72	100	NR	+

Note. WGCNA, weighted gene correlation network analysis. Target %id, percentage of the orthologues sequence matching the Human sequence, 0 means there is no human homolog identified. WGA score, whole genome alignment (WGA) score, 0 means there is no human homolog identified. CNS, central nervous system. + means this gene is expressed in CNS during E8.5 to E12.5 in mouse or carnegie stage 13–23 in human fetus. NR means not reported in database. / means this gene does not have human homolog, so the expression information is invalid.

[a]Mouse embryonic stage E8.5 to E12.5, based on MGI database.

[b]Human fetus at carnegie stage 13–23, based on EMBL_EBI Expression Atlas.

for the NTD network, and 16 for the FA response network. Third, the adjacency matrix was transformed into the topological overlap matrix (TOM), and a measure of dissimilarity was generated using 1-TOM. Finally, we employed the Dynamic Tree Cut method (Cockroft, Brook, & Copp, 1992) to define modules with parameters minCluasterSize = 30 and deepSplit = 2. Modules with Pearson correlation coefficient 0.75 or greater were merged.

2.4 | Identification of candidate genes from co-expression network

The GS (gene significance) value is defined by a mediated *p* value, and refers to the connection between a gene and the corresponding phenotype. The average GS of all the genes in a module constitute the value of module significance, which is used to measure the degree of correlation between modules and phenotypes. At the same time, based on the most representative gene in the module (module eigengene), Pearson correlation and Cox regression are used to evaluate the relevance between modules and phenotypes. Seven modules were defined in both NTD and FA networks. To improve the reliability of our results, only the module that was observed to have the most significant degree of correlation to a corresponding phenotype by both methods at the same time was selected for further analysis.

To reduce the false positive rate of the newly identified candidate genes generated by the above mentioned procedures, the "networkScreening" method (Langfelder et al., 2008) was used to identify genes that are most highly correlated with our phenotype of interest. This method computes the gene selection score "p.Weighted value" based on both GS and module membership (MM). MM measures the role a

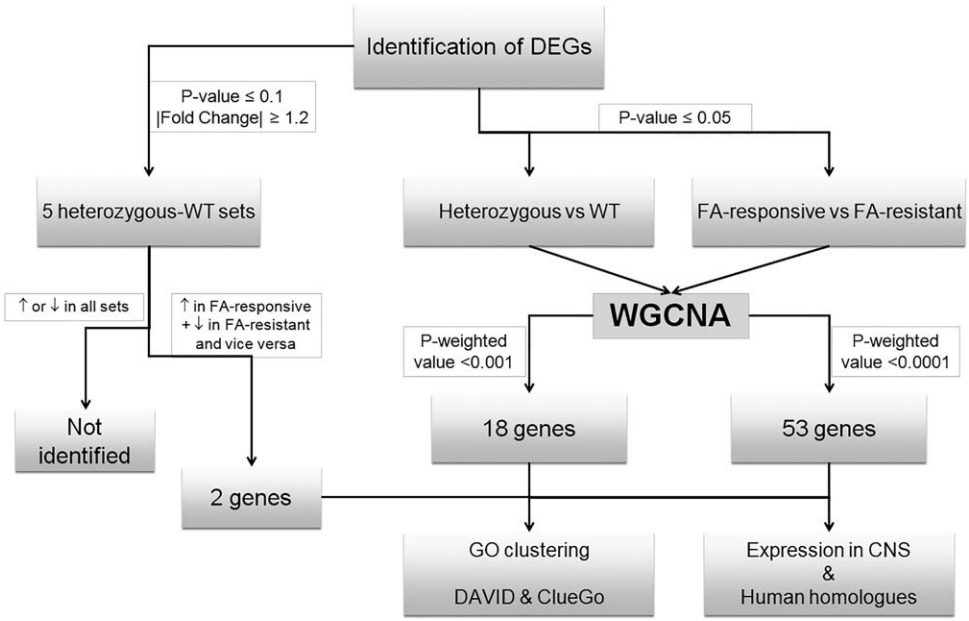

FIGURE 1 The pipeline for DEGs analysis in our study. DEGs, differentially expressed genes; WT, wild-type; WGCNA, weighted gene correlation network analysis; FA, folic acid; GO, gene ontology; CNS, central nervous system

gene plays in the whole network. For each gene, the smaller the p.Weighted value is, the stronger the correlation between this gene and its corresponding phenotype.

All the methods and calculation were carried out following the instructions of each function in R. The candidate genes represent the overlaps between the module result list and "networkScreening" result list (based on a specific p. Weighted value threshold, which is 0.001 for NTD network and 0.0001 for FA network).

2.5 | GO enrichment analysis

DAVID v6.8 (Database for Annotation Visualization and Integrated Discovery) database (Huang da, Sherman, & Lempicki, 2009) was used for GO enrichment analysis. A gene count ≥ 2 and p values $< .05$ were set as the cut-off criterion for this procedure.

ClueGO plug-in (Bindea et al., 2009) and CluePedia plugin (Bindea, Galon, & Mlecnik, 2013) of Cytoscape software were also used to process GO terms and KEGG pathway analysis. During this process, Kappa score ≥ 0.4, gene count ≥ 3, and p values $< .05$ were selected as threshold values, and the p value was corrected by the Bonferroni method.

2.6 | Expression profiles and human homologs information extraction

The expression profiles of candidate genes were examined based on the Mouse Genome Informatics (MGI) database and EMBL_EBI Expression Atlas. The human homologs were extracted from Ensembl database.

3 | RESULTS

3.1 | Data preprocessing, sample classification, and analysis pipeline

Following the quality check for all 36 samples of GSE51285 (five mutant groups plus four WT groups), GSM1241883, the biological replicate four from WT CBA mouse was discarded as an outlier (Supporting Information Table S1). The remaining 35 samples were used for further analyses. The pipeline for our analysis is shown in Figure 1. The same pipeline was used twice with different phenotypes. The first time it was focused on the difference between NTD mutants and WT controls. The second time, it was used to compare FA-responsive (*Pax3*^{Sp} and *Lrp6*^{Cd}) to FA-resistant (*Grhl3*^{ct}, *Apob*^{tm1Unc}, and *Vangl2*^{Lp}) groups.

We identified DEGs between each paired heterozygous-WT set. |Fold Change| is calculated by comparing the expression levels of a gene between heterozygous and WT. When $p \leq .05$ and |Fold Change| ≥ 1.2 criteria were used, the number of DEGs within each set was too limited to carry on further analysis among five heterozygous-WT paired sets. Therefore, we relaxed the p-value cutoff to 0.1, while the |Fold Change| cutoff remained the same. Under these conditions, 2,263 DEGs between *Pax3*$^{Sp/WT}$ and WT, 779 between *Lrp6*$^{Cd/WT}$ and WT, 3,575 between *Apob*^{tm1Unc}/*Apob*^{WT} and WT, 4,246 between *Grhl3*$^{ct/WT}$ and WT, and 2,569 between *Vangl2*$^{Lp/WT}$ and WT groups were identified.

We also reanalyzed the DEGs according to different phenotypes (NTDs vs WT; FA-responsive vs. FA-resistant) with $p \leq .05$ criteria. Those DEGS were used as input for co-expression analysis using the R package of WGCNA in order

TABLE 2 *Ywhaz* & *Matr3* are related to FA responsiveness

	FA-responsive		FA-resistant		
Gene	$Pax3^{Sp}$	$Lrp6^{Cd}$	$Grhl3^{ct}$	$Apob^{tm1Unc}$	$Vangl2^{Lp}$
Ywhaz	0.66979405	0.41297753	−0.6766863	−0.6287431	−0.5631201
Matr3	0.71434791	0.34668861	−0.7043957	−0.8942448	−0.302949

Note. The number of Log_2(fold change) was shown in the table. Fold change is calculated by the expression levels of a gene between heterozygous to wild-type in each group.

to identify genes that were correlated with each phenotype. Then, after checking the expression profiles of those genes, we focused on genes that were expressed in the central nervous system.

3.2 | Eighteen genes were identified related to NTD

As our goal was to identify new candidate NTDs-related genes, we directly compared the DEGs between each heterozygous-WT paired set in order to identify those genes that shared similar patterns among all five NTD groups. If a gene is either up-regulated or down-regulated in all NTD groups compared to their controls, then this gene may be located within a crucial point in the NTDs' pathogenic pathway shared by all five mutants. However, we failed to find a single gene that fit these rigorous criteria. Next, we combined all of the heterozygous data into a single set and all the WT data into another set and conducted WGCNA. 1095 DEGs were identified and were used as input data to build this co-expression network. Within this network structure, module "Brown" was significantly related to NTDs. After applying "networkScreening", 18 genes with a *p*-weighted value $< .001$ were identified within this module (Table 1).

We checked the human homologs of these 18 genes (Table 1) in Ensemble, and 17 of them have human homologs, and all 17 genes are highly conserved between mouse and human. We also compared their expression in the mouse and human based on MGI and EMBL_EBI Expression Atlas, respectively. Interestingly, although the original data were from liver samples, 16 genes are expressed in the central nervous system (CNS) in humans during the early fetal stages (Carnegie stage 13 to 23, equivalent to 8 to 20 post conception weeks) (Table 1).

3.3 | Fifty five genes were identified related to FA responsiveness

To identify genes that may contribute to FA sensitivity or resistance, we compared the gene expression profile between FA-responsive ($Pax3^{Sp}$ and $Lrp6^{Cd}$) and FA-resistant ($Grhl3^{ct}$, $Apob^{tm1Unc}$, and $Vangl2^{Lp}$) groups.

We initially tried to identify genes that were up-regulated in FA responsive and down-regulated in FA resistant mutants and vice versa using DEGs directly from the five heterozygous-WT paired sets. Only two genes were identified in this comparison. *Ywhaz* and *Matr3* were both up-regulated in FA-responsive mutants while being down-regulated in FA-resistant mutants (Table 2). There were no genes up-regulated in FA-resistant mutants while being down-regulated in FA-responsive mutants

We identified 2856 DEGs between FA-resistant and FA-responsive groups which were included as input in the WGCNA method to identify a FA responsiveness network. The correlation results showed that the merged Brown module, which combined six highly similar modules (correlation coefficient larger than 0.75) in this network had particularly significant relationships to FA responsiveness. In total, there were 53 genes in this module that had *p*-weighted value $< .0001$ based on the "networkScreening" method.

Together with *Ywhaz* and *Matr3*, we examined the human homologs of the aforementioned 55 genes. There are 51 out of 55 genes with human homologs, and 48 of those 51 genes shared over 80% identity between mouse and human sequences (Table 3). Fifty of the 51 human homologs are expressed in the CNS during early human fetal development (Carnegie stage 13 to 23, equivalent to 8 to 20 post conception weeks) (Table 3). We performed GO-term analysis for these 55 FA responsiveness-related genes with both DAVID and ClueGO methods. Using DAVID, we identified transcriptional regulation as the most highly enriched biological process, and the most enriched molecular function was protein binding (Figure 2a). Interestingly, the ClueGO method showed the regulation of epithelial to mesenchymal transition and TGFβ signaling is enriched in those 55 genes, with Tgfbr3 and Smad7 at the center of the network (Figure 2b).

3.4 | Eight folate metabolism genes showed significant difference between FA-responsive and FA-resistant mutants

To further explore the mechanism underlying FA responsiveness, we checked the expression of folate one carbon metabolism related genes in both FA responsive and resistant mutants. Eighty-nine FA-related genes (Supporting

TABLE 3 The list of FA responsiveness candidate genes by WGCNA

Gene	P.Weighted_value	Target %id	WGA score	Mouse fetal CNS[a]	Human fetus CNS[b]
Abcb1a	1.41E-05	0	0	+	/
Ankrd27	7.59E-05	81.87	100	NR	+
Apex1	4.02E-05	93.69	100	+	+
Atad3a	1.60E-05	90.86	96.36	+	+
Cadm1	4.57E-05	98.10	99.61	+	+
Cd2ap	4.16E-05	86.19	100	NR	+
Ces1g	8.56E-05	0	0	NR	/
Clasp2	6.47E-06	95.26	99.25	+	+
Cnot6	1.98E-05	96.23	100	+	+
Dazap1	4.45E-05	98.52	97.2	+	+
Ddah1	5.28E-05	93.68	100	+	+
Ddx19a	2.28E-05	97.49	100	+	+
Dnaja4	9.05E-05	93.20	100	+	+
Erdr1	4.55E-05	0	0	+	/
Fam117a	7.99E-06	84.48	100	NR	NR
Fbxo3	8.11E-05	93.33	99.54	NR	+
Fbxo32	7.54E-05	96.62	100	+	+
Fkbp9	3.13E-05	93.51	99.88	+	+
Fnip2	3.57E-05	77.89	100	NR	+
Grk6	7.11E-05	95.93	100	NR	+
Hars	6.95E-05	94.89	99.94	+	+
Hmgcr	2.79E-05	93.35	100	+	+
Kank2	7.57E-06	83.87	100	NR	+
Kdm2b	8.51E-05	93.96	100	NR	+
Ktn1	8.54E-05	83.42	99.28	NR	+
Large	2.56E-05	97.75	98.72	+	+
Lrrk1	4.89E-05	88.88	99.38	+	+
Map3k2	1.14E-05	96.28	100	+	+
Myo1c	8.91E-05	96.26	100	NR	+
Nodal	4.20E-06	78.81	98.96	NR	+
Pcbp2	1.75E-05	0	0	+	/
Pcgf3	7.11E-05	97.93	93.43	+	+
Pcyox1	9.72E-05	78.61	100	+	+
Phlpp2	1.40E-05	88.49	98.83	NR	+

(Continues)

TABLE 3 (Continued)

Gene	P.Weighted_value	Target %id	WGA score	Mouse fetal CNS[a]	Human fetus CNS[b]
Pkp4	9.20E-06	96.64	100	NR	+
Plcb3	5.81E-05	92.22	100	+	+
Pnisr	5.64E-05	91.77	100	+	+
Pparg	1.72E-05	96.24	100	+	+
Prkca	4.80E-06	99.40	99.17	+	+
Ptp4a2	9.75E-05	100.0	100	NR	+
Ralgps2	9.26E-06	96.44	99.83	+	+
Rnf26	1.65E-05	87.97	100	+/-	+
Rogdi	7.10E-05	94.43	100	+	+
Sgk1	5.70E-05	94.47	97.28	+	+
Smad7	7.93E-05	98.12	100	+	+
Tgfbr3	2.88E-05	81.41	99.15	+	+
Ttc7	1.01E-05	88.00	98.3	NR	+
Txnrd3	4.61E-06	87.48	92.3	-	+
Ube2i	9.52E-05	100.0	86.31	+	+
Usp12	2.32E-05	98.38	99.96	NR	+
Usp34	9.12E-05	96.54	99.74	NR	+
Vasn	9.23E-05	83.36	100	+	+
Zfp467	5.55E-05	84.01	98.9	-	+
Ywhaz	NA	99.59	100	+	+
Matr3	NA	97.87	100	+	+

Note. WGCNA, weighted gene correlation network analysis. Target %id, percentage of the orthologues sequence matching the Human sequence, 0 means there is no human homolog identified. WGA score, whole genome alignment (WGA) score, 0 means there is no human homolog identified. CNS, central nervous system. + means this gene is expressed in CNS during E8.5 to E12.5 in mouse or carnegie stage 13–23 in human fetus. **+/-** means the expression of this gene is ambiguous in CNS during E8.5 to E12.5 in mouse or carnegie stage 13–23 in human fetus. NR means not reported in database. **/** means this gene does not have human homolog, so the expression information is invalid.

[a]Mouse embryonic stage E8.5 to E12.5, based on MGI database.

[b]Human fetus at carnegie stage 13–23, based on EMBL_EBI Expression Atlas.

Information Table S2) were extracted based largely on a review by Au, Findley, and Northrup (2017), and information from the MGI database, which covers genes involved in one-carbon metabolism, choline metabolism, and transsulfuration pathways. Furthermore, the DEGs of those folate metabolism genes that differed between FA-responsive and FA-resistant mutants were examined. We focused primarily on the FA-resistant mutants, as this set contains three different mutants. DEGs from five heterozygous-WT paired sets with $p \leq .1$ and $|\text{Fold Change}| \geq 1.2$ were compared and those genes that are up- or down-regulated in at least two FA-resistant mutants were identified. At the same time, the expression in the FA-responsive mutants should not trend the same direction as in FA-resistant mutants. By these criteria, eight genes were identified (Table 4). *Gsr*, *Mthfd1*, *Gclc*, *Gart*, and *Abcc3* all have lower expression levels in at least two FA-resistant mutants; while *Slc44a2, Slc25a32*, and *Bche* have higher expression levels in at least two FA-resistant NTD mutants.

4 | DISCUSSION

It is well known that some mouse NTD mutants are FA responsive, meaning FA protects the embryos from developing NTDs by enhancing normal morphogenetic processes

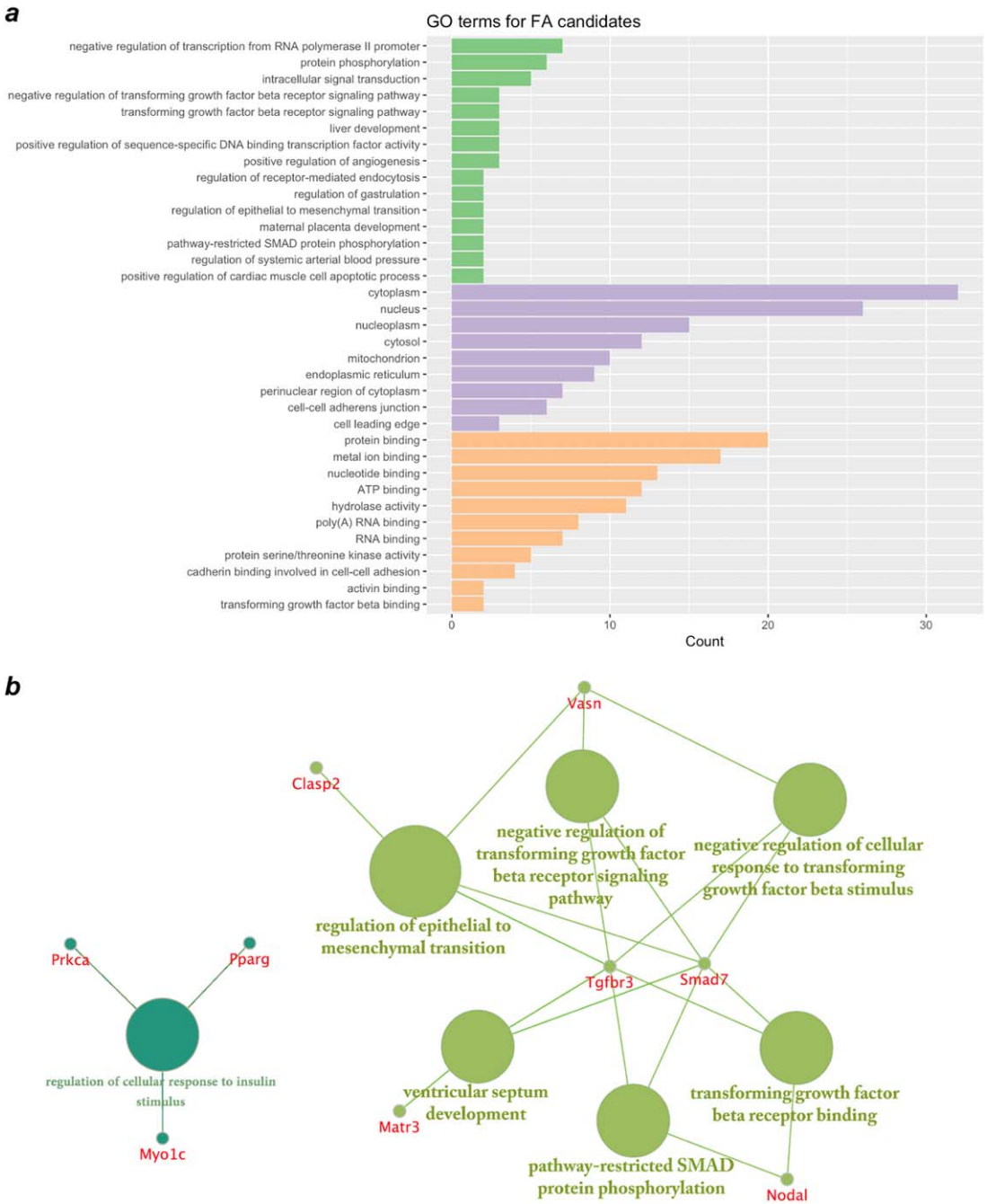

FIGURE 2 GO term analysis results in 55 FA responsiveness related genes by (a) DAVID and (b) ClueGO methods. Using DAVID, the categories are divided into biological process (green), cellular component (purple), and molecular function (orange)

during early embryogenesis, while some mutants are FA resistant. In this study, microarray data from five NTDs mutants with different FA responsiveness were evaluated. Four of them, *Pax3*^{Sp}(Epstein, Vogan, Trasler, & Gros, 1993), *Lrp6*^{Cd}(Carter et al., 1999), *Vangl2*^{Lp}(Torban, Wang, Groulx, & Gros, 2004), and *Grhl3*^{ct}(Cockroft et al., 1992), were reported to have various NTD penetrance in heterozygous mice. Since it is seldom, if ever, that homozygous mutations are reported in human NTD cases, the comparison between heterozygote and WT mice may be an effective way to parallel the human situations. Furthermore, we adapted the WGCNA method to analyze the NTD or FA responsiveness related genes based on network screening methodologies. A total of 18 NTD candidate genes and 55 genes related to FA responsiveness in NTDs were identified in our study. Within these genes, mice homozygous for *Lasp1* and *Kdm2b* had previously been reported to exhibit NTD phenotypes (Boulard, Edwards, & Bestor, 2015; Fukuda, Tokunaga, Sakamoto, & Yoshida, 2011; Hermann-Kleiter et al., 2009), supporting the effectiveness of our screening approach to identify new potential NTD candidate genes. Although the original samples were adult mouse liver, 16 out of 18 NTD

TABLE 4 Folate metabolism genes related to FA responsiveness

	FA-responsive		FA-resistant		
Gene	***Pax3*^{Sp}**	***Lrp6*^{Cd}**	***Grhl3*^{ct}**	***Apob*^{tm1Unc}**	***Vangl2*^{Lp}**
Abcc3	NS	NS	−0.9836972	−0.9142913	−1.1288219
Gart	NS	NS	−0.6221493	NS	−0.4605329
Gclc	0.63752375	NS	−0.6770293	−0.5545005	−1.00861
Gsr	0.35248121	NS	−0.5065363	−0.4846741	−0.3555942
Mthfd1	0.44960304	NS	−0.8434009	NS	−0.4201585
Bche	NS	NS	NS	1.59353157	0.99608566
Slc25a32	NS	NS	NS	0.28285351	1.16170874
Slc44a2	NS	−0.52207	0.7259365	0.58865228	NS

Note. The number of Log_2(fold change) was shown in the table. Fold change is calculated by the expression levels of a gene between heterozygous to wild-type in each group. NS, No significant difference in expression between heterozygous and wild-type mice.

candidate genes and 50 out of 55 FA responsiveness genes were expressed in the CNS during early human fetal development. These results support the idea that those genes identified herein could be important in human NTD pathogenesis.

A total of ten genes were identified by comparing FA-responsive to FA-resistant mutants. *Ywhaz* and *Matr3*, were up-regulated in all FA-responsive mutants and down-regulated in all FA-resistant mutants. Currently, there is very little known about these two genes' function during embryonic development. However, Ywhaz (tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta), also called 14-3-3zeta, interacts with, and facilitates, the activation of β-catenin in the WNT pathway (Tian et al., 2004), which is an essential signaling pathway for neural tube development. Matr3 (Matrin 3) is highly expressed in the neural tube at E8.5 (Quintero-Rivera et al., 2015), a critical early stage in NTC. It is likely these two genes are involved in NTD pathology and folate responsiveness. Further investigations into the function of these two genes during NTC should be highly informative.

The other eight genes were identified from 89 extended folate metabolism pathway genes (Figure 3). When same criteria were applied on 89 random selected genes, less positive genes were filtered out than from extended folate metabolism pathway genes (2/5/6 positive results in three repeats), suggesting the importance of extended folate metabolism pathway in NTD. However, other GSE datasets with more NTD

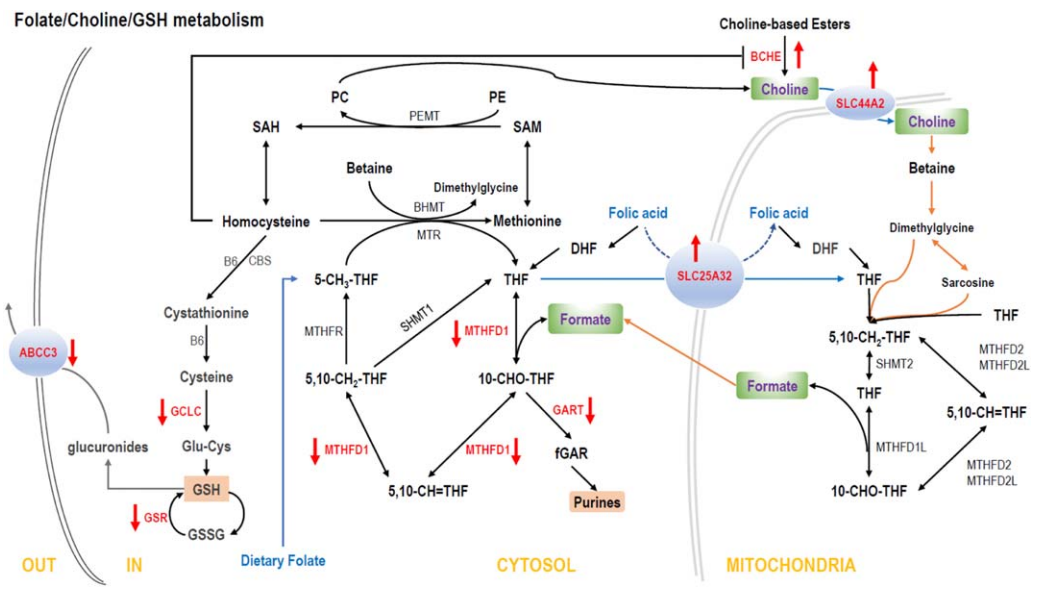

FIGURE 3 Simplified folate/choline/GSH metabolism pathways. Genes in red are those whose expression levels differ significantly between FA-resistant and FA-responsive mutants. Arrows indicate whether the expression levels are up or down-regulated in FA-resistant mutants

mice models in future will be needed to confirm such significance. These eight genes are primarily involved in three different functions: de novo purine biosynthesis, oxidative stress, and the mitochondrial folate/choline pathway. *Mthfd1* and *Gart* are involved in de novo purine biosynthesis. Mthfd1 (methylenetetrahydrofolate dehydrogenase 1) is essential for formate to enter folate-based one-carbon metabolism, as it transforms formate and tetrahydrofolate (THF) to 10-formyltetrahydrofolate (10-f-THF) (MacFarlane et al., 2009). Gart (phosphoribosylglycinamide formyltransferase) transforms the formyl group from 10-fTHF to GAR to generate fGAR, which is the second step in the *de novo* purine biosynthesis pathway (Connelly, DeMartino, Boger, & Wilson, 2013). As both *Mthfd1* and *Gart* have lower expression levels in FA-resistant mutants, even in the presence of FA supplementation, the *de novo* purine biosynthesis probably is still insufficient to rescue the FA-resistant mutants by promoting proper NTC.

Gsr (glutathione reductase) and Gcl (glutamate-cysteine ligase) are two key enzymes that regulate the level of glutathione (GSH), an important antioxidant in cells. Gclc is the catalytic subunit of Gcl, which is the first rate-limiting enzyme of GSH synthesis. Gsr catalyzes the reduction of glutathione disulfide (GSSG) to the sulfhydryl form GSH. The decreased levels of *Gclc* and *Gsr* in FA-resistant mutants could increase the levels of oxidative stress and lead to NTD-affected offspring. A previously published human study also supports this idea, as women with NTD-affected pregnancies had significantly higher plasma concentrations of GSSG than did controls (Zhao et al., 2006). The Abcc3 (ATP binding cassette subfamily C member 3) protein, also called Mrp3 (multidrug resistance-associated protein3), is not only involved in cellular folate homeostasis (Hooijberg et al., 2003), but also could export glucuronides and glutathione conjugates (Kruh, Belinsky, Gallo, & Lee, 2007). Therefore, the decreased level of Abcc3 in those FA-resistant mutants may further increase the toxic effects of oxidative stress in cells.

Unlike those genes mentioned above, *Slc25a32*, *Slc44a2*, and *Bche* have higher expression levels in FA-resistant mutants than FA-responsive mutants. Bche (butyrylcholinesterase) is a plasma cholinesterase which hydrolyzes choline-based esters to generate choline and Slc44a2 (solute carrier family 44 member 2), which is a choline transporter. Slc25a32 (solute carrier family 25 member 32), also called the mitochondrial folate transporter (Findley et al., 2017; Gutierrez-Aguilar & Baines, 2013), is also higher in FA-resistant mutants. Both folate and choline transported into mitochondria enter the mitochondrial folate metabolic pathway to produce formate, which is transported back to cytoplasm as an important carbon source for the one-carbon pool (Brosnan & Brosnan, 2016). Mice that are homozygous for the inactivated *Slc25a32* allele all express NTDs (Finnell, unpublished).

Based on these results, it appears as if the FA-resistant mutants have not only an unbalanced mitochondrial/cytosol folate metabolism, but also an increased oxidative stress level (Figure 3), which cannot be corrected by FA supplementation. However, a previously published report showed that dietary formate supplementation to the pregnant dam could rescue NTDs in the FA-resistant *Grhl3*$^{ct/ct}$ embryos (Sudiwala et al., 2016). Further, periconceptional dietary intake of choline and betaine decreases NTD risks in humans, and this is independent of folate intake (Shaw, Carmichael, Yang, Selvin, & Schaffer, 2004).

There are several limitations to our study. First, the expression profile review of our candidate genes was limited by the existing database for human fetal tissues. NTC occurs by the fourth week of pregnancy in humans (day 28 post conception); therefore, gene expression during 8–20 weeks post conception is too late in development to affect NTC. Second, since all of our results are based on bioinformatic methods, given the limitation of our sample numbers, these results all require experimental validation in future studies. Nevertheless, taken together, our data suggests that antioxidant, formate and/or choline supplementation together, represent viable interventions that might be more effective than simply providing FA alone, in the prevention of NTDs.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

ORCID

Yufang Zheng http://orcid.org/0000-0002-0488-2624

REFERENCES

Au, K. S., Findley, T. O., & Northrup, H. (2017). Finding the genetic mechanisms of folate deficiency and neural tube defects-Leaving no stone unturned. *American Journal of Medical Genetics Part A*, *173*(11), 3042–3057.

Bindea, G., Galon, J., & Mlecnik, B. (2013). CluePedia Cytoscape plugin: Pathway insights using integrated experimental and in silico data. *Bioinformatics*, *29*(5), 661–663.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., … Galon, J. (2009). ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*, *25*(8), 1091–1093.

Boulard, M., Edwards, J. R., & Bestor, T. H. (2015). FBXL10 protects Polycomb-bound genes from hypermethylation. *Nature Genetics*, *47*(5), 479–485.

Brosnan, M. E., & Brosnan, J. T. (2016). Formate: The neglected member of one-carbon metabolism. *Annual Review of Nutrition*, *36*, 369–388.

Carter, M., Ulrich, S., Oofuji, Y., Williams, D. A., & Ross, M. E. (1999). Crooked tail (Cd) models human folate-responsive neural tube defects. *Human Molecular Genetics*, *8*(12), 2199–2204.

Cockroft, D. L., Brook, F. A., & Copp, A. J. (1992). Inositol deficiency increases the susceptibility to neural tube defects of genetically predisposed (curly tail) mouse embryos in vitro. *Teratology*, *45*(2), 223–232.

Connelly, S., DeMartino, J. K., Boger, D. L., & Wilson, I. A. (2013). Biological and structural evaluation of 10R- and 10S-methylthio-DDACTHF reveals a new role for sulfur in inhibition of glycinamide ribonucleotide transformylase. *Biochemistry*, *52*(30), 5133–5144.

Crider, K. S., Bailey, L. B., & Berry, R. J. (2011). Folic acid food fortification-its history, effect, concerns, and future directions. *Nutrients*, *3*(3), 370–384.

Czeizel, A. E., & Dudas, I. (1992). Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation. *The New England Journal of Medicine*, *327*(26), 1832–1835.

Epstein, D. J., Vogan, K. J., Trasler, D. G., & Gros, P. (1993). A mutation within intron 3 of the Pax-3 gene produces aberrantly spliced mRNA transcripts in the splotch (Sp) mouse mutant. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(2), 532–536.

Findley, T. O., Tenpenny, J. C., O'Byrne, M. R., Morrison, A. C., Hixson, J. E., Northrup, H., & Au, K. S. (2017). Mutations in folate transporter genes and risk for human myelomeningocele. *American Journal of Medical Genetics. Part A*, *173*(11), 2973–2984.

Fukuda, T., Tokunaga, A., Sakamoto, R., & Yoshida, N. (2011). Fbxl10/Kdm2b deficiency accelerates neural progenitor cell death and leads to exencephaly. *Molecular and Cellular Neurosciences*, *46*(3), 614–624.

Greene, N. D., & Copp, A. J. (1997). Inositol prevents folate-resistant neural tube defects in the mouse. *Nature Medicine*, *3*(1), 60–66.

Greene, N. D., & Copp, A. J. (2005). Mouse models of neural tube defects: investigating preventive mechanisms. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *135C*(1), 31–41.

Greene, N. D., & Copp, A. J. (2014). Neural tube defects. *Annual Review of Neuroscience*, *37*, 221–242.

Gutierrez-Aguilar, M., & Baines, C. P. (2013). Physiological and pathological roles of mitochondrial SLC25 carriers. *Biochemical Journal*, *454*(3), 371–386.

Harris, M. J. (2009). Insights into prevention of human neural tube defects by folic acid arising from consideration of mouse mutants. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *85*(4), 331–339.

Harris, M. J., & Juriloff, D. M. (2007). Mouse mutants with neural tube closure defects and their role in understanding human neural tube defects. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *79*(3), 187–210.

Harris, M. J., & Juriloff, D. M. (2010). An update to the list of mouse mutants with neural tube closure defects and advances toward a complete genetic perspective of neural tube closure. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *88*(8), 653–669.

Hermann-Kleiter, N., Ghaffari-Tabrizi, N., Blumer, M. J., Schwarzer, C., Mazur, M. A., & Artner, I. (2009). Lasp1 misexpression influences chondrocyte differentiation in the vertebral column. *The International Journal of Developmental Biology*, *53*(7), 983–991.

Hibbard, B. M., Hibbard, E. D., & Jeffcoate, T. N. (1965). Folic acid and reproduction. *Acta Obstetricia Et Gynecologica Scandinavica*, *44*(3), 375–400.

Homanics, G. E., Maeda, N., Traber, M. G., Kayden, H. J., Dehart, D. B., & Sulik, K. K. (1995). Exencephaly and hydrocephaly in mice with targeted modification of the apolipoprotein B (Apob) gene. *Teratology*, *51*(1), 1–10.

Hooijberg, J. H., Peters, G. J., Assaraf, Y. G., Kathmann, I., Priest, D. G., Bunni, M. A., … Jansen, G. (2003). The role of multidrug resistance proteins MRP1, MRP2 and MRP3 in cellular folate homeostasis. *Biochemical Pharmacology*, *65*(5), 765–771.

Huang da, W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., & Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, *4*(2), 249–264.

Juriloff, D. M., & Harris, M. J. (2000). Mouse models for neural tube closure defects. *Human Molecular Genetics*, *9*(6), 993–1000.

Kruh, G. D., Belinsky, M. G., Gallo, J. M., & Lee, K. (2007). Physiological and pharmacological functions of Mrp2, Mrp3 and Mrp4 as determined from recent studies on gene-disrupted mice. *Cancer and Metastasis Reviews*, *26*(1), 5–14.

Langfelder, P., & Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, *9*, 559.

MacFarlane, A. J., Perry, C. A., Girnary, H. H., Gao, D., Allen, R. H., Stabler, S. P., … Stover, P. J. (2009). Mthfd1 is an essential gene in mice and alters biomarkers of impaired one-carbon metabolism. *Journal of Biological Chemistry*, *284*(3), 1533–1539.

Mitchell, L. E. (2005). Epidemiology of neural tube defects. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics*, *135C*(1), 88–94.

MRC_Vitamin_Study_Research_Group. (1991). Prevention of neural tube defects: Results of the Medical Research Council vitamin study. MRC Vitamin Study Research Group. *Lancet*, *338*(8760), 131–137.

Nakouzi, G. A., & Nadeau, J. H. (2014). Does dietary folic acid supplementation in mouse NTD models affect neural tube development or gamete preference at fertilization? *BMC Genetics* , *15*(1), 91.

Quintero-Rivera, F., Xi, Q. J., Keppler-Noreuil, K. M., Lee, J. H., Higgins, A. W., Anchan, R. M., … Maas, R. L. (2015). MATR3 disruption in human and mouse associated with bicuspid aortic valve, aortic coarctation and patent ductus arteriosus. *Human Molecular Genetics*, *24*(8), 2375–2389.

Shaw, G. M., Carmichael, S. L., Yang, W., Selvin, S., & Schaffer, D. M. (2004). Periconceptional dietary intake of choline and betaine and neural tube defects in offspring. *American Journal of Epidemiology*, *160*(2), 102–109.

Shimoji, K., Kimura, T., Kondo, A., Tange, Y., Miyajima, M., & Arai, H. (2013). Genetic studies of myelomeningocele. *Child's Nervous System*, *29*(9), 1417–1425.

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.

Statistical Applications in Genetics and Molecular Biology, *3*(1), 1. Article3.

Sudiwala, S., De Castro, S. C., Leung, K. Y., Brosnan, J. T., Brosnan, M. E., Mills, K., . . . Greene, N. D. (2016). Formate supplementation enhances folate-dependent nucleotide biosynthesis and prevents spina bifida in a mouse model of folic acid-resistant neural tube defects. *Biochimie*, *126*, 63–70.

Tian, Q., Feetham, M. C., Tao, W. A., He, X. C., Li, L., Aebersold, R., & Hood, L. (2004). Proteomic analysis identifies that 14-3-3zeta interacts with beta-catenin and facilitates its activation by Akt. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(43), 15370–15375.

Torban, E., Wang, H. J., Groulx, N., & Gros, P. (2004). Independent mutations in mouse Vangl2 that cause neural tube defects in looptail mice impair interaction with members of the Dishevelled family. *Journal of Biological Chemistry*, *279*(50), 52703–52713.

Wilde, J. J., Petersen, J. R., & Niswander, L. (2014). Genetic, epigenetic, and environmental contributions to neural tube closure. *Annual Review of Genetics*, *48*, 583–611.

Wilson, C. L., & Miller, C. J. (2005). Simpleaffy: A BioConductor package for Affymetrix Quality Control and data analysis. *Bioinformatics*, *21*(18), 3683–3685.

Wlodarczyk, B. J., Tang, L. S., Triplett, A., Aleman, F., & Finnell, R. H. (2006). Spontaneous neural tube defects in splotch mice supplemented with selected micronutrients. *Toxicology and Applied Pharmacology*, *213*(1), 55–63.

Zhao, W., Langfelder, P., Fuller, T., Dong, J., Li, A., & Hovarth, S. (2010). Weighted gene coexpression network analysis: State of the art. *Journal of Biopharmaceutical Statistics*, *20*(2), 281–300.

Zhao, W., Mosley, B. S., Cleves, M. A., Melnyk, S., James, S. J., & Hobbs, C. A. (2006). Neural tube defects and maternal biomarkers of folate, homocysteine, and glutathione metabolism. *Birth Defects Research Part A: Clinical and Molecular Teratology*, *76*(4), 230–236.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Gao X, Finnell RH, Wang H, Zheng Y. Network correlation analysis revealed potential new mechanisms for neural tube defects beyond folic acid. *Birth Defects Research*. 2018;00:1–12. https://doi.org/10.1002/bdr2.1336