**RESEARCH ARTICLE**
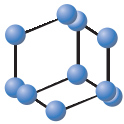
# Recursive Feature Elimination-based Biomarker Identification for Open Neural Tube Defects

Kadhir Velu Karthik[1], Aruna Rajalingam[1], Mallaiah Shivashankar[1] and Anjali Ganjiwale[1,*]

[1]*Department of Life Science, Bangalore University, Bangalore, India*

**Abstract:** *Background*: Open spina bifida (myelomeningocele) is the result of the failure of spinal cord closing completely and is the second most common and severe birth defect. Open neural tube defects are multifactorial, and the exact molecular mechanism of the pathogenesis is not clear due to disease complexity for which prenatal treatment options remain limited worldwide. Artificial intelligence techniques like machine learning tools have been increasingly used in precision diagnosis.

*Objective*: The primary objective of this study is to identify key genes for open neural tube defects using a machine learning approach that provides additional information about myelomeningocele in order to obtain a more accurate diagnosis.

*Materials and Methods*: Our study reports differential gene expression analysis from multiple datasets (GSE4182 and GSE101141) of amniotic fluid samples with open neural tube defects. The sample outliers in the datasets were detected using principal component analysis (PCA). We report a combination of the differential gene expression analysis with recursive feature elimination (RFE), a machine learning approach to get 4 key genes for open neural tube defects. The features selected were validated using five binary classifiers for diseased and healthy samples: Logistic Regression (LR), Decision tree classifier (DT), Support Vector Machine (SVM), Random Forest classifier (RF), and K-nearest neighbour (KNN) with 5-fold cross-validation.

*Results*: Growth Associated Protein 43 (GAP43), Glial fibrillary acidic protein (GFAP), Repetin (RPTN), and CD44 are the important genes identified in the study. These genes are known to be involved in axon growth, astrocyte differentiation in the central nervous system, post-traumatic brain repair, neuroinflammation, and inflammation-linked neuronal injuries. These key genes represent a promising tool for further studies in the diagnosis and early detection of open neural tube defects.

*Conclusion*: These key biomarkers help in the diagnosis and early detection of open neural tube defects, thus evaluating the progress and seriousness in diseases condition. This study strengthens previous literature sources of confirming these biomarkers linked with open NTD's. Thus, among other prenatal treatment options present until now, these biomarkers help in the early detection of open neural tube defects, which provides success in both treatment and prevention of these defects in the advanced stage.

## 1. INTRODUCTION

Neural tube defects, or NTD's are the second most common and severe birth defects whose number of cases in pregnant women varies from 0.5 to more than 10 per 1,000 women (1 in 1000 live birth) worldwide [1]. During embryogenesis, a few cells of the ectoderm form a neural tube which gives rise to the central nervous system (CNS), and this is the most important structure observed in the embryo and is essential for the developing organism [2]. Any defect in this process affects the neurulation process, which results in brain and spinal cord defects [3], affecting the morphogenetic process of the neural tube closure. Thus, anencephaly and open spina bifida (myelomeningocele) are such neural tube defects observed in children due to the abnormal closure of neural tubes in the brain and spinal cord. Evidence suggests that the failure of neural tube closure is the key feature resulting in open neural tube defects [4]. On the other hand, ample evidence convinces us that even genetic and environmental factors also play an important role in the process of open NTD's [5, 6].

After the process of successful initiation of neural tube formation, failure of the closure of the neural tube at the de

*Address correspondence to this author at the Department of Life Science, Bangalore University, Bangalore, India; E-mail: anjalig@bub.ernet.in

novo initiation site results in open NTDs. Defects in the brain and spinal cord result in lethal conditions shortly after birth. If there is any defect in the brain, it results in anencephaly, encephalocele, craniorachischisis (severe condition), and a defect in the spine and spinal cord results in spina bifida. Babies with this condition are found to have neural tissues within meninges covered sac that protrudes through the open vertebrae (myelomeningocele; spina bifida cystica) or are exposed directly to the amniotic fluid (myelocele). Some having open spina bifida need intensive medical care due to various medical conditions and neurological presentations [5].

It is found that most of the NTDs are multifactorial in which each gene plays an important role individually for the risk factor and is sufficient in causing the neural tube closure defects in humans [7]. Epidemiologic studies and experimental models have identified some novel candidate genes that play an important role in causing NTDs in humans. However, exact gene-gene interaction and gene environmental factors are yet to be understood due to their complexity [5].

The foetal surgery to improve motor function is the only therapeutic aid available for myelomeningocele. The understanding of molecular mechanisms of the disease is limited not only by the availability of biological samples but also due to rapidly changing gene expression patterns in the developing foetus. Whole-genome mRNA expression profiles from amniotic fluid samples (GSE4182: five spina bifida and four healthy control samples) have been reported. All the spina bifida samples in the study lack closure of the neural tube [8]. The study highlights the contribution of *SLA*, *LST1*, and *BENE* genes in the development of neural tube defects. Transcriptomic analysis of amniotic fluid with myelomeningocele has been reported by Tarui and colleagues [9]. The study reports 284 differentially regulated genes and includes 10 control and 10 myelomeningocele amniotic fluid supernatant samples (GSE101141). Genes associated with neurodevelopment and neuronal regeneration (GAP43 and ZEB1) along with known genes associated with myelomeningocele PRICKLE2, GLI3, RAB23, HES1, FOLR1 have been reported to play an important role. Recently, Li *et al*. have reported hub gene analysis and miRNA–mRNA network for the GSE4182 dataset. A total of 967 candidate genes were identified to be plausibly involved in the pathophysiology of spina bifida (open neural tube defect) [10]. miRNA–mRNA network was also confirmed recently by Sun and his colleagues with experiments in a mouse model showing miR-222-3p to be an important biomarker in NTDs. The study reports the downregulation of miR-222-3p in the Wnt/β-catenin signaling pathway by inhibition of DNA damage transcript 4 (Ddit4), resulting in abnormal cell proliferation and apoptosis, leading to neural tube closure defects [11].

Neurotrophic factors present in the brain, for example, Brain-derived neurotrophic factor (BDNF), have been shown to bind to a receptor tyrosine kinase B (TrkB), which leads to the migration of neural crest cells for a normal sympathetic nervous system (SNS) development. Any defect in the migration of neural crest cells facilitated by TrkB sig-

nalling leads to abnormal sympathetic nervous system (SNS) development, and also its overexpression leads to neuroblastoma cancer [12]. BDNF has also been shown to be linked with caffeine metabolism involving Sirtuin 1/BDNF regulation and synaptic plasticity. Studies also suggest that anti-aging genes and caffeine metabolism are linked to Non-Alcoholic Fatty Liver Disease (NAFLD), diabetes, and chronic diseases [13], and therefore, overconsumption of caffeine during pregnancy results in a lot of open neural tube defect complications [14].

GSE4182 and GSE101141 represent one of the few studies on amniotic fluid samples with open neural tube defects. In the present study, we report the use of machine learning algorithms to identify important gene markers contributing to the pathophysiology of open neural tube defects from the existing gene expression data. Due to high dimensional and low sample size datasets of myelomeningocele, our approach involves a combination of the traditional fold change method along with feature selection using recursive feature elimination (RFE) to filter the most important genes to distinguish between healthy and diseased samples and derive insightful inferences on biological functions and pathways involved.

## 2. MATERIALS AND METHODS

### 2.1. Data Collection and Pre-processing

The gene expression profile of amniotic fluid for open neural tube defects was downloaded from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/) with *Homo sapiens* as the inclusion criteria. GSE4182 reports microarray analysis of foetal mRNA isolated from amniocytes collected by amniocentesis. The dataset represents one of the few studies available on human foetus microarray detected using Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133_Plus_2] and represents 4 Spina Bifida (open neural tube defect) and 5 Normal human foetus samples. GSE101141 reports amniotic fluid transcriptomics of foetuses with myelomeningocele detected using Human Genome U133 Plus 2.0 arrays and represents 10 myelomeningocele and 10 control samples.

Affymetrix CEL files were downloaded from NCBI GEO, accession number GSE4182, GSE101141, and renormalized using the RMA method in AltAnalyze software [15, 16]. Outliers in the datasets were detected and removed using the inter-quartile range. Principal component analysis (PCA) was used for clustering and outlier analysis in the diseased and control samples [17].

### 2.2. Differentially Expressed Gene Analysis by Fold Change Method

Differentially expressed genes were calculated with log2 transformed data of GSE4182 and GSE101141individually. A gene was 'Downregulated' for the log2 fold change $\leq$ -2 and 'Upregulated' for the log2 fold change $\geq$ 2 when the adjusted *p* values were $\leq$ 0.05 and the data was visualized using heatmaps. The structure of the data in the set of 2-fold differentially expressed genes was analysed using the complete-linkage (farthest neighbour) algorithm with 'Euclidean' metric measure.

## 2.3. Machine Learning Binary Classifiers

For all the further studies, myelomeningocele dataset (GSE101141) was selected as the training data. The dataset was scaled with MinMax Scalar and Power Transformed to obtain the normal Gaussian distribution. Scikit-learns "test-train split ()" class was used to split the data into train and test sets with a ratio of 80: 20 with 5-fold cross-validation. Logistic Regression (LR), Decision tree classifier (DT), Support Vector Machine (SVM), Random Forest classifier (RF), and K-nearest neighbour (KNN) classifier [18] were used as binary classifiers between diseased and control samples. All the five algorithms were implemented using the Scikit-learn machine learning library (Version 0.21.2) [19].

## 2.4. Marker Gene Selection using Machine Learning

Two different machine learning algorithms, Recursive feature elimination (RFE) and Random Forest classifier (RFC) [20], for feature selection were applied to the whole dataset of myelomeningocele using the Scikit-learn python module [19]. Recursive feature elimination [21] is basically a backward selection of the predictors based on model performance. Since RFE trains the whole dataset, every time it drops a feature, it requires heavy computational time. Random Forest classifier for feature selection is a tree-based strategy and is applicable to "large p, small n" problems. Each tree in the RFC algorithm is trained on a random subset of features, and ranks are assigned to the features by decreasing the mean impurity at the node [20].

## 2.5. Identification and Validation of Marker Genes

The performance of selected marker genes was validated using LR, DT, SVM, RF, and KNN with test-train split () and 5-fold cross-validation as used in binary classifiers between diseased and control samples. The performance of all the 5 classifiers was measured using accuracy, precision, and area under the ROC curve (AUC) as the performance metrics [22-24].

## 2.6. Functional Enrichment Analysis

Enrichment analysis, including Over-Representation Analysis (ORA), and Gene Set Enrichment Analysis (GSEA), were performed using WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) [25]. Enrichment analysis was done separately for upregulated and downregulated features. All the results were validated with a statistical significance of *p*-value <0.05. Gene-Gene interaction network was performed for the selected biomarker genes through the Gene-MANIA Cytoscape plugin command-line tool [26].

## 3. RESULTS

### 3.1. Data Preparation

The raw data files (CEL) of GSE101141 (10 diseased and 10 healthy controls) and GSE4182 (4 diseased and 5 healthy controls) were downloaded, and the data was processed using AltAnalyze v.2.0 0 (http://www.altanalyze.org). The RMA method (Robust Multiple Array Average) was applied to all Affymetrix CEL files, detection above background (DABG) with *p*-value < 0.05, and the probe sets were annotated. All the GSM files were renamed as 'C' for Control/healthy samples and 'S' for diseased samples. Data with missing 'Gene Symbols' was dropped, and a total of 41,597 gene features were selected for further analysis. Outlier detection and treatment in both datasets were performed using interquartile range (IQR) (Fig. **S1**).

Principal Component Analysis is a linear dimensionality reduction technique and can be used for extracting information from high dimensional space by projecting it as a vector (a lower-dimensional sub-space). PCA of GSE4182 shows that sample S3 (GSM94601) is very close to that of healthy controls. Similarly, in GSE101141, control C7 (GSM2699731) is very close to that of disease samples. Also, sample clustering of GSE4182 showed that S3 is clustered with other healthy control samples and clustering of GSE101141 showed that C7 is clustered along with diseased samples (Fig. **1**). Therefore, S3 and C7 samples were eliminated from further analysis.

### 3.2. Differential Gene Expression and Functional Enrichment Analysis

In GSE4182, a total of 840 genes were differentially regulated with an absolute fold change value of ± 2 and an adjusted *p*-value of <0.05 (Table **S1**). 5 Genes were upregulated, and 10 genes were downregulated amongst the top 15 genes sorted by absolute fold change value (Table **1**). The heatmap plotted for differentially regulated genes shows a clear distinction between diseased and control samples (Fig. **2A**). Genes GPM6A (Glycoprotein M6A) and TUBB2B (Tubulin Beta 2B Class IIb) are upregulated and known to be involved in the neuron migration pathway (GO: 0001764) and neuron projection morphogenesis pathway (GO: 0048812) along with PTPRZ1 (protein tyrosine phosphatase, receptor type Z1) gene. The KEGG pathway analysis of upregulated genes in this dataset are involved in Osteoclast differentiation (hsa04380), Staphylococcus aureus infection (hsa05150), Transcriptional misregulation in cancer (hsa05202), Adipocytokine signaling pathway (hsa04920), Phagosome (hsa04145) , MicroRNAs in cancer (hsa05206) , Systemic lupus erythematosus (hsa05322), Cytokine-cytokine receptor interaction (hsa04060), JAK-STAT signaling pathway (hsa04630), Amyotrophic lateral sclerosis (ALS) (hsa05014) and downregulated genes are involved in Tight junction(hsa04530), Hippo signaling pathway(hsa04392), Systemic lupus erythematosus (hsa05322), NF-kappa B signaling pathway(hsa04064), Mucin type O-glycan biosynthesis(hsa00512), Alcoholism (hsa05034), FoxO signaling pathway(hsa04068), ErbB signaling pathway(hsa04012), Non-small cell lung cancer (hsa05223) and Axon guidance(hsa04360) (Fig. **3A**, Table **S3**).

A total of 29 genes were differentially regulated with an absolute fold change value of ±1.5 and an adjusted *p*-value of <0.05 (Table **S2**). In contrast, only one gene, PDZRN3 (PDZ domain containing ring finger 3), was observed to be downregulated and the remaining 14 were upregulated amongst the top 15 genes sorted by absolute fold change for GSE101141 (Table **2**). Like the GSE4182 dataset, the differentially expressed genes for
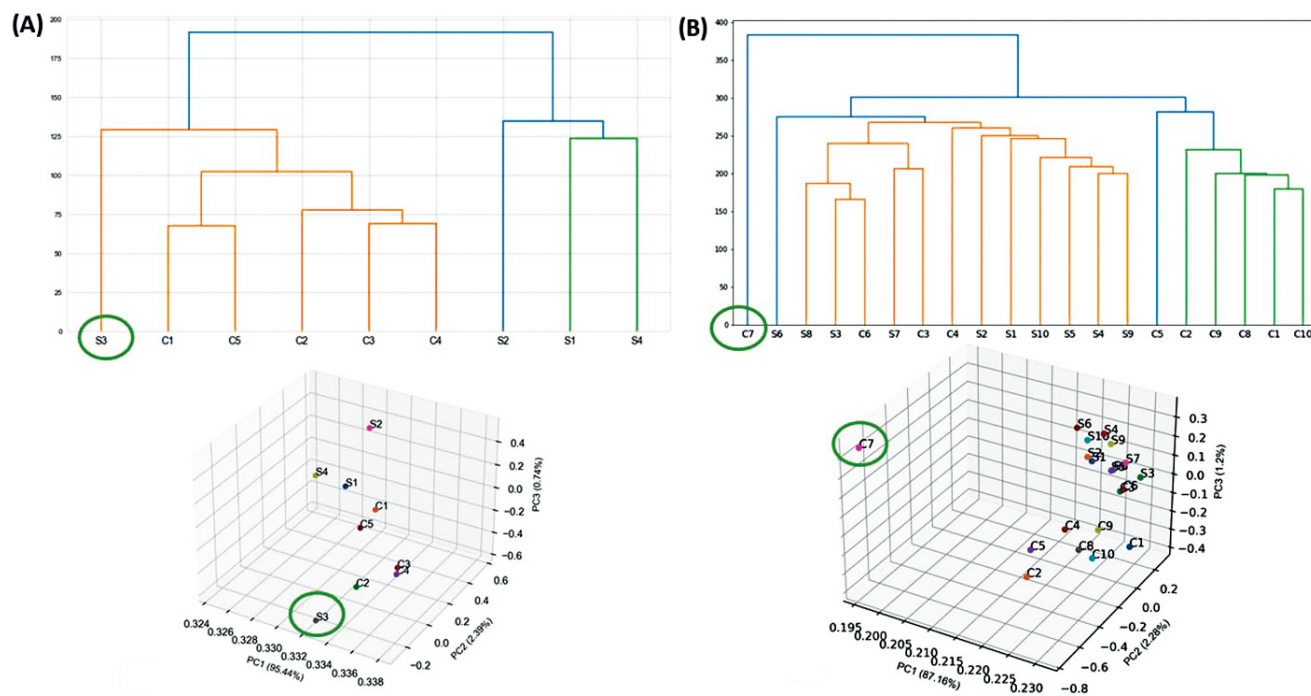
**Fig. (1).** Clustering and principal component analysis of (**A**) GSE4182 (**B**) GSE101141. Samples S3 and C7 were removed from further analysis. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

**Table 1. List of top 15 differentially expressed genes ranked by the absolute value of the fold change for GSE4182.**

| Probeset | Symbol | logFC | *p*-value | Adj. *p*-value | Regulation |
|---|---|---|---|---|---|
| 205000_at | DDX3Y | -5.85 | 1.46013E-05 | 0.004272919 | Down |
| 201909_at | LOC100133662 /// RPS4Y1 | -5.40 | 7.42371E-09 | 0.000301974 | Down |
| 221008_s_at | AGXT2L1 | -5.11 | 0.000513329 | 0.015471965 | Down |
| 209469_at | GPM6A | 4.98 | 8.6895E-05 | 0.007223776 | Up |
| 204469_at | PTPRZ1 | 4.82 | 3.66137E-05 | 0.005409601 | Up |
| 214023_x_at | TUBB2B | 4.55 | 0.000513718 | 0.015471965 | Up |
| 225895_at | SYNPO2 | -4.30 | 0.004473876 | 0.048679104 | Down |
| 228170_at | OLIG1 | 4.28 | 0.000174858 | 0.009334244 | Up |
| 223720_at | SPINK7 | -4.16 | 0.000583189 | 0.016382868 | Down |
| 226803_at | CHMP4C | -4.11 | 0.001338642 | 0.025023867 | Down |

**(Table 1) contd….**

| Probeset | Symbol | logFC | *p*-value | Adj. *p*-value | Regulation |
|----------|--------|-------|-----------|----------------|------------|
| 1564960_at | KRTAP7-1 | -4.06 | 0.000336396 | 0.012669992 | Down |
| 206912_at | FOXE1 | -4.04 | 0.000267143 | 0.011474723 | Down |
| 209470_s_at | GPM6A | 4.03 | 0.000231394 | 0.010638966 | Up |
| 219995_s_at | ZNF750 | -4.03 | 1.30607E-06 | 0.002380474 | Down |
| 205964_at | ZNF426 | -3.99 | 3.91642E-05 | 0.005493382 | Down |
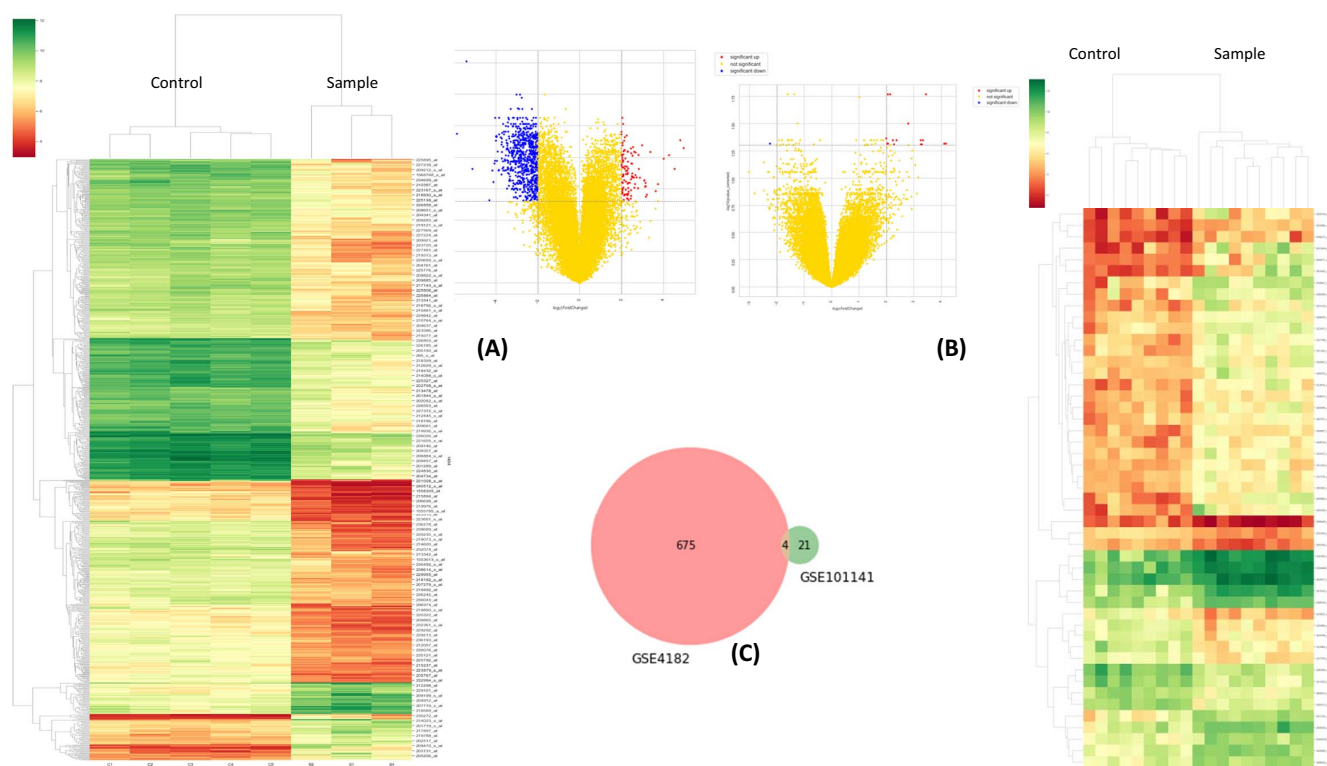


**Fig. (2).** Differential gene expression analysis represented by heatmap and volcano plots for (**A**) GSE4182 and (**B**) GSE101141 datasets. (**C**) 4 genes are common in differentially expressed genes for GSE4182 and GSE101141. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

GSE101141 show a clear distinction between healthy controls and diseased samples (Fig. **2B**). The KEGG pathway analysis of this dataset's upregulated genes showed involvement in biosynthesis of unsaturated fatty acids (hsa01040), vasopressin-regulated water reabsorption (hsa04962), fatty acid metabolism (hsa01212), shigellosis (hsa05131), drug metabolism (hsa00982), bile secretion (hsa04976), PPAR signaling pathway (hsa03320), ECM-receptor interaction (hsa04512), salivary secretion (hsa04970), hematopoietic cell lineage (hsa04640) and downregulated genes showed involvement of axon guidance (hsa04360), bladder cancer (hsa05219), VEGF signaling pathway (hsa04370), mitophagy (hsa04137), shigellosis (hsa05131), epithelial cell signaling in Helicobacter pylori infection (hsa05120), prolactin signaling pathway (hsa04917), adherens junction (hsa04520), bacterial invasion of epithelial cells (hsa05100), and EGFR tyrosine kinase inhibitor resistance (hsa01521) (Fig. **3B**, Table **S4**). Though the FDR is $\leq 0.05$ for functional enrichment analysis, the *p*-value is $\leq 0.05$ for all the enriched pathways. Correlation analysis in the differentially expressed gene sets shows a clear distinction between the diseased and the healthy controls (Fig. **2**). 4 Genes 'GAP43' (growth-associated protein 43), 'RPTN' (Repetin), 'CEACAM6' (carcinoembryonic antigen-related cell adhesion molecule 6 (non-specific cross-reacting antigen),
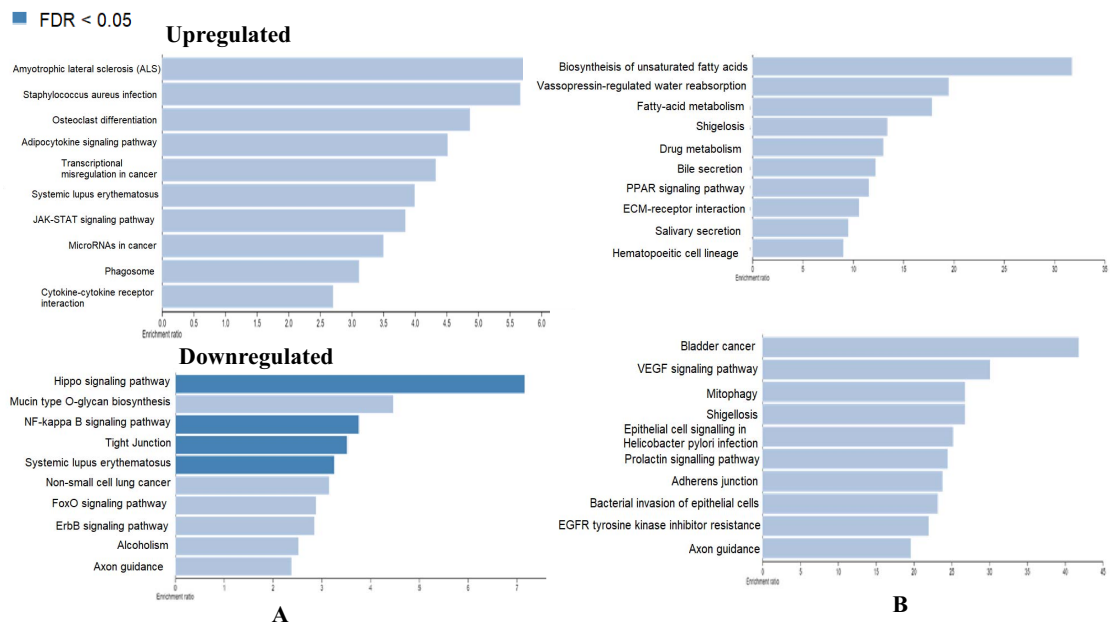
**Fig. (3).** Functional enrichment analysis of KEGG pathways of up and down-regulated genes (**A**) GSE4182 (**B**) GSE101141. GO analysis has been provided in the supplemental information. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

**Table 2. List of top 15 differentially expressed genes ranked by the absolute value of the fold change for GSE101141.**

| Probeset | Symbol | logFC | *p*-value | Adj. *p*-value | Regulation |
|----------|--------|-------|-----------|----------------|------------|
| 1553454_at | RPTN | 4.18 | 4.39725E-05 | 0.048500241 | Up |
| 203540_at | GFAP | 4.12 | 4.23228E-05 | 0.048500241 | Up |
| 204489_s_at | CD44 | 3.45 | 2.10527E-06 | 0.017124244 | Up |
| 204471_at | GAP43 | 3.34 | 1.98423E-05 | 0.045128701 | Up |
| 210067_at | AQP4 | 3.30 | 5.4523E-05 | 0.049276697 | Up |
| 232578_at | CLDN18 | 3.27 | 5.23796E-05 | 0.049276697 | Up |
| 226228_at | AQP4 | 3.26 | 3.04098E-05 | 0.045128701 | Up |
| 213975_s_at | LYZ | 2.80 | 6.23675E-06 | 0.031706064 | Up |
| 211726_s_at | FMO2 | 2.38 | 2.12628E-05 | 0.045128701 | Up |
| 211657_at | CEACAM6 | 2.29 | 5.13881E-05 | 0.049276697 | Up |
| 212915_at | PDZRN3 | -2.25 | 3.75728E-05 | 0.048500241 | Down |

**(Table 2) contd….**

| Probeset | Symbol | logFC | *p*-value | Adj. *p*-value | Regulation |
|----------|--------|-------|-----------|----------------|------------|
| 211708_s_at | SCD | 2.24 | 2.82754E-05 | 0.045128701 | Up |
| 209835_x_at | CD44 | 2.14 | 1.8977E-06 | 0.017124244 | Up |
| 204526_s_at | TBC1D8 | 2.06 | 4.41237E-05 | 0.048500241 | Up |
| 224568_x_at | MALAT1 | 2.06 | 1.42846E-06 | 0.017124244 | Up |

'TFAP2A' (transcription factor AP-2 alpha (activating enhancer-binding protein 2 alpha) were common in both the datasets. These genes are involved in tight junction (hsa04530) and Axon guidance (hsa04360). Details of biological processes involved are provided in the supplemental information (Fig. **S2**).

### 3.3. Machine Learning-based Selection of Marker Genes

The normalized GSE101141 dataset with 41,957 gene features was used for further analysis. The outliers in the dataset were detected and dropped using interquartile range (IQR), reducing the number of features to 40,670. The 'Control' sample was encoded to '0', and the 'Diseased' sample was encoded to '1' to convert the categorical features to the numeric features. Binary classification algorithms with gene features as predictor variables were used to classify control '0' *vs* diseased '1' as target variables. Comparative performance of five different classifier algorithms logistic regression (LR), decision tree classifier (DT), support vector machine (SVM), random forest classifier (RF), and K-nearest neighbour (KNN) classifier is as shown in Fig. **4A**. The bias in the models was reduced by applying a 5-fold cross-validation approach. Logistic regression showed an accuracy of 86.67% and AUC_ROC of 0.90, followed by KNN with 80% accuracy and AUC_ROC of 0.95. The dataset used in the study has a high dimensionality and low sample count. Dimensionality reduction and selection of important features were performed using two different feature selection algorithms, Recursive Feature Elimination (RFE) and Random Forest classifier (RFC). RFE can be implemented using two different configuration options: the number of features to be selected and the choice of the classification algorithm. In this study, the top 15 gene features were selected by RFE with LR as the classifier (Table **S5**). Since RFE works by removing the features step by step and fitting the model on the training data until the desired number remains, it becomes computationally expensive.

Random Forests are the tree-based strategies and rank the features by reducing the mean impurity called 'Gini impurity' on the overall trees. The top 15 features were selected by RFC (Table **S6**). The selected features by RFE and RFC were used to check the classification performance by retraining the selected features (Fig. **4A**). Features selected by RFE improved the accuracy and AUC of all the five classifiers; in contrast, RFC reduced the accuracy and AUC considerably.

It was of interest to look for common genes selected by RFE and the fold change method as important biomarker genes (Fig. **4B**). Our study could identify 4 common 'Biomarkers' for open neural tube defects. The top 4 genes selected by absolute fold change value appear to be selected by the RFE algorithm as well. 'GAP43' (Growth associated protein 43) is a neuron-specific phosphoprotein selected by both the fold change method and RFE method and is known to play a critical role during neurogenesis in axon growth and synapsis functions [27]. 'GFAP' (Glial fibrillary acidic protein) is an intracellular protein found in astrocytes of the central nervous system and one of the classical biomarkers used for the diagnosis of neural tube defects in prenatal diagnosis through amniotic fluid [28]. 'RPTN' (Repetin) is a protein-coding gene and is involved in the developmental biology and keratinization pathways. 'CD44' is known to be involved in neuroinflammation and inflammation-linked neuronal injuries [29]. CD44 has been shown to be involved in axon guidance, synaptic transmission, astrocyte differentiation, and post-traumatic brain repair [30]. All the 4 biomarkers identified are upregulated (Fig. **4C**). GAP43 and RPTN are also differentially regulated in GSE4182 spina bifida datasets.

Gene-Gene interaction network predicted for proposed biomarker genes through GeneMANIA Cytoscape plugin command-line tool is shown in Fig. (**5A**). The network shows 20 related genes and 145 total links based on protein-protein interaction data collected from BioGRID and PathwayCommons, genetic interaction data, shared protein domains, co-localization, pathway data along with predicted functional relationships between the genes [26]. GAP43 shows a shared protein domain with the MYO5A protein-coding gene for Myosin V, a class of actin-based motor proteins involved in cytoplasmic vesicle transport and anchorage. Mutations in MYO5A cause Griscelli syndrome type-1 (GS1), Griscelli syndrome type-3 (GS3), and neuroectodermal melanolysosomal disease [31]. The network shows genetic interaction between GAP43, GFAP, and CD44 genes along with ABCC5, MYO5A, ERBB4, CALM1, TIAM2, and PRNP. Notably, ABBC5 is a protein-coding gene associated with glycosaminoglycan metabolism and has been reported in Primary Angle-Closure Glaucoma disease [32]. PRNP gene observed in the network is plausibly involved in neuronal myelin sheath maintenance and might play a role in neuronal development [33]. ERBB4 gene belongs to the EGFR subfamily of receptor tyrosine kinases, and defective signalling of tyrosine kinase leads to abnormal sympathetic
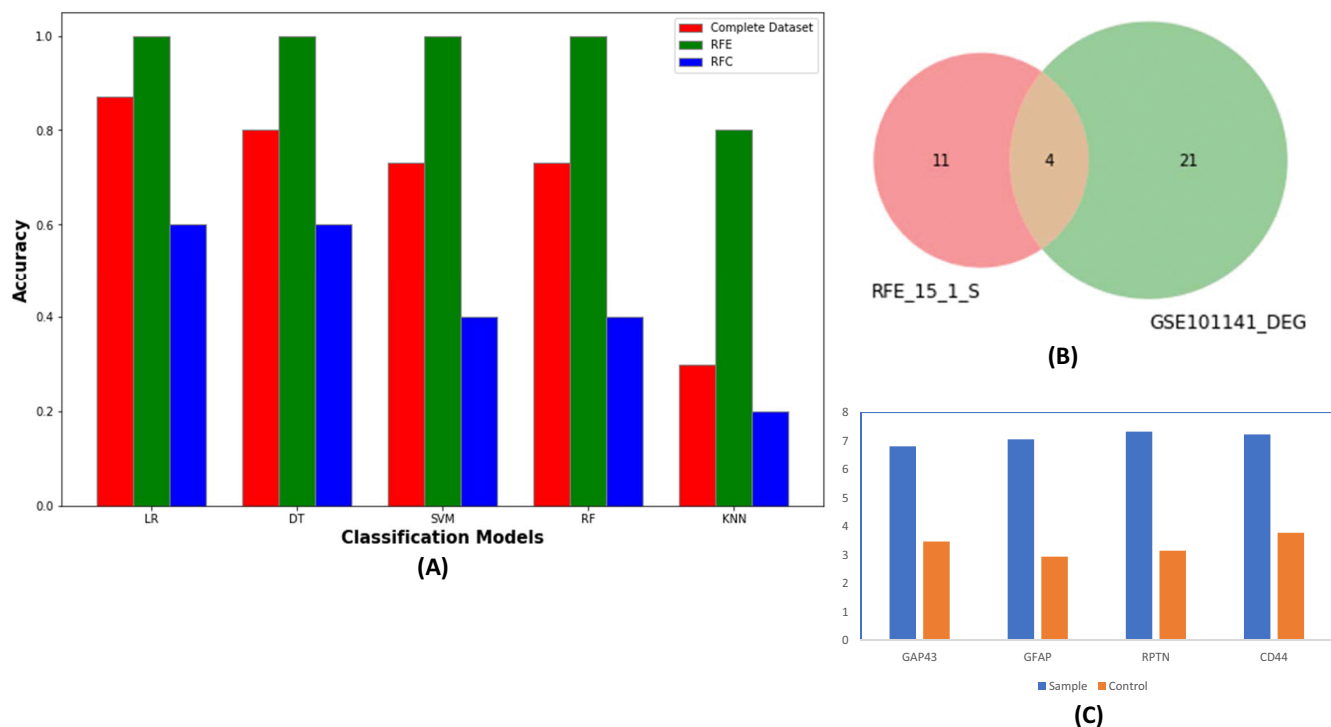
**Fig. (4).** (**A**) Classification accuracy with feature selection. Recursive feature elimination (RFE) shows an increase in training accuracy as compared to random forest classifier. , LR: Logistic Regression, DT: Decision Tree, SVM: Support Vector Machine, RF: Random Forest, KNN: K-Nearest Neighbour (**B**) Common gene features selected with fold change and RFE (**C**) All the 4 gene features are upregulated as compared to control. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).
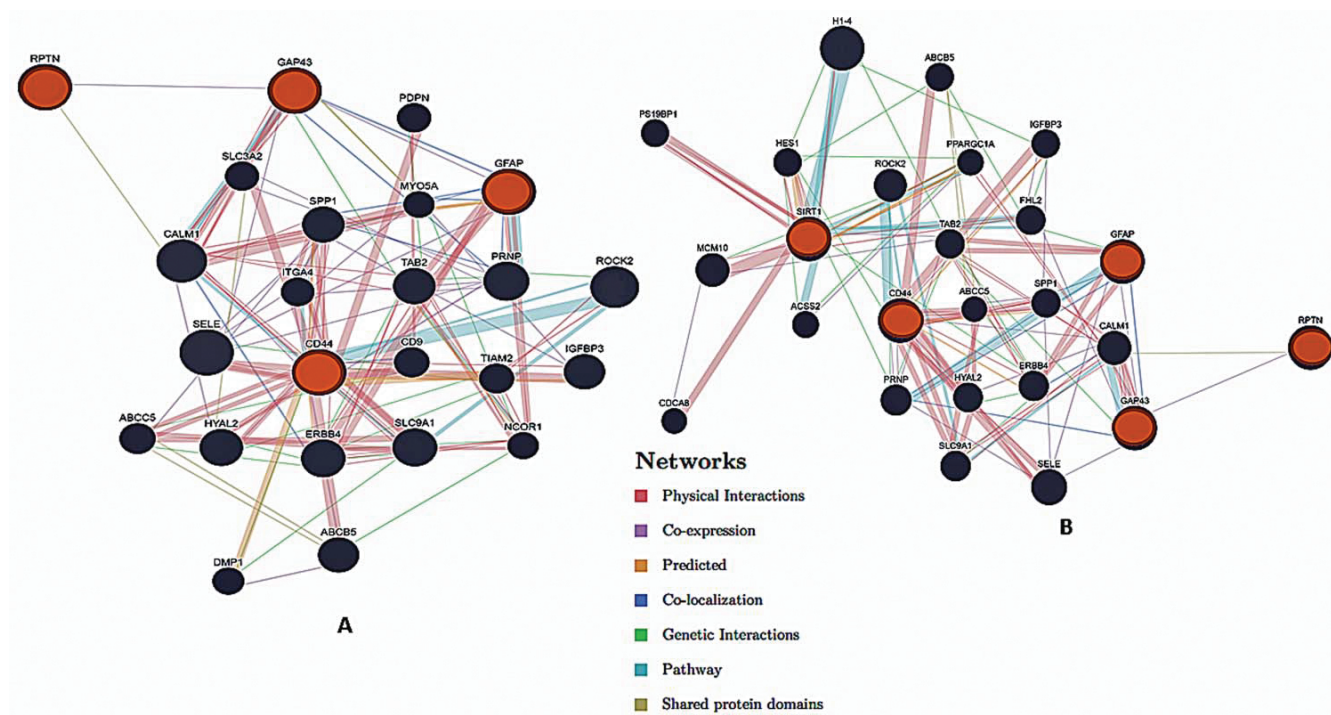


**Fig. (5).** (**A**) GeneMANIA Cytoscape composite network of RPTN, GAP43, GFAP, and CD44 showing 20 related genes and 145 total links. (**B**) GeneMANIA Cytoscape composite network of GAP43, GFAP, and CD44 with Sirtuin 1 anti-aging gene. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

nervous system (SNS) development [12]. An extended gene interaction network shows a possible interaction with Sirtuin 1, an anti-aging gene that plays a very important role in disease progress, treatment, prevention, and seriousness of the open neural tube defects (Fig. **5B**). This gene is also found to be linked with various chronic diseases such as obesity, cardiovascular disease, NAFLD, and neurodegenerative diseases. Environmental factors are also associated with Sirtuin 1 downregulation, and Sirtuin 1/OCT6 has been shown to be linked to environmental stress induction neural tube defects, and its activators play a crucial role in the treatment of NTDs [34-36].

## 4. DISCUSSION

Spina bifida, also known as myelomeningocele, is one of the most serious congenital anomalies diagnosed worldwide, for which prenatal treatment options remain limited [1, 25, 37]. Open neural tube defects in amniotic fluid samples are reported by two different datasets from the GEO database. The fold change analysis of differential gene expression shows that the two datasets, GSE101141 and GSE4182, have differences in gene expression dynamics, the plausible reason being the sample collection time is 24-weeks for GSE101141 and less than 20 weeks for GSE4182. However, a comparison of the differentially expressed genes from two datasets representing open neural tube defects from amniotic fluid samples shows that GAP43, RPTN, CEACAM6, and TFAP2A are common and are involved in the tight junction (hsa04530) and axon guidance (hsa04360) pathways. The gene-gene interaction network of the 4 biomarker genes clearly indicates the involvement of tyrosine kinase signaling, Calomodulin, and Prion protein (PRNP) genes that are linked to neural tube defects. Apart from gene-gene interactions, there are also certain environmental factors like folic acid, maternal insulin-dependent diabetes, and maternal use of certain anticonvulsant (antiseizure) medications that are linked with multiple genes causing neural tube closure effects. [38-40]. However, direct linkages to these environmental factors are beyond the scope of the current study.

In the present study, we propose a data-driven approach for the identification of marker genes. Since the two datasets cannot be merged, only GSE101141, due to its higher sample count, was used for machine learning algorithms. The top 15 genes ranked by recursive feature elimination-based machine learning approach by the validation with 5 different binary classifiers (LR, DT, SVM, RF, and KNN) showed an increase in accuracy with RFE selected features. A total of four biomarker genes, GAP43, RPTN, CD44, and GFAP, were identified to be common in both RFE-based and fold change-based gene feature selection in the study. All four genes are upregulated in GSE101141 (Fig. **4C**). Out of the 4 genes, GAP43 and RPTN are also differentially expressed in GSE4182.

As a member of the neuromodulin family, Growth Associated Protein 43 (GAP43) is a major growth cone protein and is a substrate for protein kinase C, which in response to extracellular guidance cues, can regulate F-actin behaviour. It is required for commissural axon guidance in the developing vertebrate nervous system [41]. The protein product of this gene is associated with nerve growth which is the major

component of motile growth cones that form the tip of cones of the elongating axon. Apart from these functions, it also plays an important role in axonal and dendritic filopodia [42]. For a long time, GAP43 gene product protein has been termed as growth or plasticity protein [43, 44]. This protein is termed so because it is expressed at high levels in neuronal growth cones during development and axonal regeneration. This protein is considered a crucial component of an effective regenerative response in the nervous system.

Mascaro and his colleagues reported that downregulating GAP43 causes a significant increase in the turnover of presynaptic boutons using the RNA interference approach. In addition, silencing hampers the generation of reactive sprouts, and also their findings showed the requirement of GAP43 in sustaining synaptic stability and promoting the initiation of axonal regrowth. Thus, this protein plays a very important role in the regenerative process in the nervous system [45]. Basi *et al.* found that among several tissues and cells examined, GAP43 mRNA was expressed only in neurons. Developmental and regeneration-associated changes in GAP43 synthesis appeared to be mediated largely at the level of transcription of a single gene [46]. Nguyen *et al.* reported that GAP43 interactions with neurotubulin in growing neurites extend the role of GAP43 at the mitotic spindle, and also modification of GAP43 at its PKC phosphorylation site directs its distribution to different membrane microdomains that have distinct roles in the regulation of intrinsic and extrinsic behaviours in growing neurons [47, 48]. Studies conducted by Kawasaki *et al.* in the developing rat spinal cord showed the spatiotemporal expression of GAP43 in neural development, axonal regeneration, and modulation of synaptic function [49]. Thus, GAP43 serves as one of the biomarkers in NTDs and is confirmed in our study.

Repetin (RPTN) protein is a member of the S100 family. The human repetin gene is a member of the "fused" gene family and localized in the epidermal differentiation complex on chromosome 1q21. The "fused" gene family comprises profilaggrin, trichohyalin, repetin, hornerin, the profilaggrin-related protein, and a protein encoded by c1orf10. Functionally, these proteins are associated with keratin intermediate filaments and partially crosslinked to the cell envelope [50]. For a long time, it has been known to be expressed in the normal epidermis. In 2015, a study conducted by Wang *et al.* showed that RPTN is ubiquitously expressed with relatively high levels in the choroid plexus, hippocampus, and prefrontal cortex in both mouse and the human brain. This study was the first to report RPTN expression in the brain [51]. RPTN protein contains two EF-hands, which structurally resemble the calcium-binding domain of Parvalbumin (PV). Calcium-binding proteins play important roles in the brain and in psychiatric disorders in which PV is regarded as a neuronal marker and has a potential key role in the pathophysiology of schizophrenia. Thus, RPTN plays a potential role in emotional and cognitive processing in which it is involved in cornified cell envelope formation and has an important role in the pathogenesis of schizophrenia and bipolar disorder [51].

Amniotic fluid glial fibrillary acidic protein (AF-GFAP) is a brain-specific protein. As a part of an investigation for a more sensitive and specific biochemical marker of NTD,

Petzold *et al*. focused on a group of proteins specifically expressed in the brain called brain-specific proteins (BSPs), in which glial fibrillary acidic protein (GFAP) is one among them and is considered as a potentially useful biomarker for the specific diagnosis of open NTDs [52]. Of note, this protein had already been studied as a potential marker for NTDs in the 1980s, but preceding works had failed to establish its interest in NTD diagnosis [53]. Among these BSPs, GFAP is considered the major intermediate filament protein of the astrocyte and is reported to indicate astrogliosis and astrocytic activation [54, 55]. Thus, studies on the immuno-histochemical profile of the myelomeningocele placode also showed that GFAP is a standard marker showing strongly positive in normal spinal cord development [56], and this serves as a useful biomarker for the specific diagnosis of open NTDs.

CD44 is a receptor for hyaluronan, among other extra-cellular matrix components, and acts as a co-factor for growth factors and cytokines [57]. CD44, which is the principal cell membrane receptor for hyaluronic acid (HA), is expressed in embryos on the surface of differentiated embryonic stem cells at different stages of development [58]. Some growth factors bind to glycosaminoglycans (GAGs), and HA action through the CD44 receptor is involved in regulating these growth factors [59], which facilitates the proliferation and migration of cells [60]. CD44 binds to HA and participates in the migration of mesenchymal cells [61]. Wheatley and their team reported that in mouse embryos, CD44 is strongly expressed on days 9.5−12.5 in heart, somite, and extremity mesenchyme [62]. Previously, it has been reported that CD44 was first observed in cephalic neural folding in chicken embryos and the mesenchyme and ectoderm in the caudal region at a later stage [60]. Reduced amounts of HA in NTDs may interfere with the migration and proliferation of the stem cells in the neural tube and prevent its closure [63]. The reduced amount of CD44 localization due to HA depletion indicates that sufficient mesenchymal stem cells cannot migrate to the cranial and caudal regions of the embryo, which prevents the neural folds from fusing. So, the mesenchymal stem cells could not migrate sufficiently to the cranial and caudal portions of the mesenchyme, which caused the failure of the neural folds to unite [64]. Thus, CD44 also serves as a potential marker for the open NTDs.

## CONCLUSION

To conclude, recursive feature elimination (RFE), along with absolute fold change, has resulted in identifying four biomarker genes, GAP43, RPTN, CD44, and GFAP, and the role of these biomarkers in the early prediction of the neural tube defects. Also, this study strengthens previous literature sources of confirming these biomarkers linked with open NTDs. Thus, prenatal treatment options that remain today are limited, and there are emerging successes in the treatment and prevention of these defects. Present advances in our understanding of the human genome, gene expression, and genetic architecture are making a path for new opportunities to understand the root causes of this disorder and advance the vision of developing improved strategies for the prevention of open NTDs.

## ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Not applicable.

## HUMAN AND ANIMAL RIGHTS

No animals/humans were used for studies that are the basis of this research.

## CONSENT FOR PUBLICATION

Not applicable.

## AVAILABILITY OF DATA AND MATERIALS

Datasets GSE101141 and GSE4182 can be downloaded from the Gene Expression Omnibus database (GEO, https://www.ncbi.nlm.nih.gov/geo/).

## FUNDING

## CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

## ACKNOWLEDGEMENTS

## SUPPLEMENTARY MATERIAL

Supplementary material is available on the publisher's website along with the published article.

## REFERENCES

[1]     Juriloff, D.M.; Harris, M.J. Hypothesis: The female excess in cranial neural tube defects reflects an epigenetic drag of the inactivating X chromosome on the molecular mechanisms of neural fold elevation. *Birth Defects Res. A Clin. Mol. Teratol.,* **2012**, *94*(10), 849-855.
        http://dx.doi.org/10.1002/bdra.23036 PMID: 22753363
[2]     Wu, Y.; Peng, S.; Finnell, R.H.; Zheng, Y. Organoids as a new model system to study neural tube defects. *FASEB J.,* **2021**, *35*(4), e21545.
        http://dx.doi.org/10.1096/fj.202002348R PMID: 33729606
[3]     Sadler, T.W. Embryology of neural tube development. *Am. J. Med. Genet. C. Semin. Med. Genet.,* **2005**, *135C*(1), 2-8.
        http://dx.doi.org/10.1002/ajmg.c.30049 PMID: 15806586
[4]     Copp, A.J.; Adzick, N.S.; Chitty, L.S.; Fletcher, J.M.; Holmbeck, G.N.; Shaw, G.M. Spina bifida. *Nat. Rev. Dis. Primers,* **2015**, *1*(1), 15007.
        http://dx.doi.org/10.1038/nrdp.2015.7 PMID: 27189655
[5]     Greene, N.D.; Copp, A.J. Neural tube defects. *Annu. Rev. Neurosci.,* **2014**, *37*(1), 221-242.
        http://dx.doi.org/10.1146/annurev-neuro-062012-170354     PMID: 25032496
[6]     Copp, A.J.; Greene, N.D. Genetics and development of neural tube defects. *J. Pathol.,* **2010**, *220*(2), 217-230.
        http://dx.doi.org/10.1002/path.2643 PMID: 19918803

[7]     Harris, M.J.; Juriloff, D.M. Mouse mutants with neural tube clo-sure defects and their role in understanding human neural tube de-fects. *Birth Defects Res. A Clin. Mol.,* **2007**, *79*(3), 187-210.
http://dx.doi.org/10.1002/bdra.20333

[8]     Nagy, G.R.; Györffy, B.; Galamb, O.; Molnár, B.; Nagy, B.; Papp, Z. Use of routinely collected amniotic fluid for whole-genome ex-pression analysis of polygenic disorders. *Clin. Chem.,* **2006**, *52*(11), 2013-2020.
http://dx.doi.org/10.1373/clinchem.2006.074971 PMID: 17008366

[9]     Tarui, T.; Kim, A.; Flake, A.; McClain, L.; Stratigis, J.D.; Fried, I.; Newman, R.; Slonim, D.K.; Bianchi, D.W. Amniotic fluid transcriptomics reflects novel disease mechanisms in fetuses with myelomeningocele. *Am. J. Obstet. Gynecol.,* **2017**, *217*(5), 587.e1-587.e10.
http://dx.doi.org/10.1016/j.ajog.2017.07.022 PMID: 28735706

[10]    Li, Z.; Feng, J.; Yuan, Z. Key modules and hub genes identified by coexpression network analysis for revealing novel biomarkers for spina bifida. *Front. Genet.,* **2020**, *11*, 583316.
http://dx.doi.org/10.3389/fgene.2020.583316 PMID: 33343629

[11]    Sun, Y.; Zhang, J.; Wang, Y.; Wang, L.; Song, M.; Khan, A.; Zhang, L.; Niu, B.; Zhao, H.; Li, M.; Luo, T.; He, Q.; Xie, X.; Liu, Z.; Xie, J. miR-222-3p is involved in neural tube closure by di-rectly targeting Ddit4 in RA induced NTDs mouse model. *Cell Cycle,* **2021**, *20*(22), 2372-2386.
http://dx.doi.org/10.1080/15384101.2021.1982506       PMID: 34779712

[12]    Kasemeier-Kulesa, J.C.; Spengler, J.A.; Muolo, C.E.; Morrison, J.A.; Woolley, T.E.; Schnell, S.; Kulesa, P.M. The embryonic trunk neural crest microenvironment regulates the plasticity and invasion of human neuroblastoma *via* TrkB signaling. *Dev. Biol.,* **2021**, *480*, 78-90.
http://dx.doi.org/10.1016/j.ydbio.2021.08.007 PMID: 34416224

[13]    Martins, I.J. Nutrition therapy regulates caffeine metabolism with relevance to NAFLD and induction of type 3 diabetes. *J. Diabetes Metab. Disord.,* **2017**, *4*(1), 1-9.

[14]    Schmidt, R.J.; Romitti, P.A.; Burns, T.L.; Browne, M.L.; Dru-schel, C.M.; Olney, R.S. Maternal caffeine consumption and risk of neural tube defects. *Birth Defects Res. A Clin. Mol. Teratol.,* **2009**, *85*(11), 879-889.
http://dx.doi.org/10.1002/bdra.20624 PMID: 19711421

[15]    Emig, D.; Salomonis, N.; Baumbach, J.; Lengauer, T.; Conklin, B.R.; Albrecht, M. AltAnalyze and DomainGraph: Analyzing and visualizing exon expression data. *Nucleic Acids Res.,* **2010**, *38*(Suppl. 2), W755-62.
http://dx.doi.org/10.1093/nar/gkq405 PMID: 20513647

[16]    Irizarry, R.A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y.D.; An-tonellis, K.J.; Scherf, U.; Speed, T.P. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics,* **2003**, *4*(2), 249-264.
http://dx.doi.org/10.1093/biostatistics/4.2.249 PMID: 12925520

[17]    Lenz, M.; Müller, F.J.; Zenke, M.; Schuppert, A. Principal com-ponents analysis and the reported low intrinsic dimensionality of gene expression microarray data. *Sci. Rep.,* **2016**, *6*(1), 25696.
http://dx.doi.org/10.1038/srep25696 PMID: 27254731

[18]    Hassan, C.A.; Khan, M.S.; Shah, M.A. Comparison of machine learning algorithms in data classification. IEEE, **2018**, *2018*, 8748995.
http://dx.doi.org/10.23919/IConAC.2018.8748995

[19]    Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.,* **2011**, *12*, 2825-2830.

[20]    Díaz-Uriarte, R.; Alvarez de Andrés, S. Gene selection and classi-fication of microarray data using random forest. *BMC Bioinf.,* **2006**, *7*(1), 3.
http://dx.doi.org/10.1186/1471-2105-7-3 PMID: 16398926

[21]    Guyon, I.; Weston, J.; Barnhill, S.; Vapnik, V. Gene selection for cancer classification using support vector machines. *Mach. Learn.,* **2002**, *46*(1), 389-422.
http://dx.doi.org/10.1023/A:1012487302797

[22]    Baldi, P.; Brunak, S.; Chauvin, Y.; Andersen, C.A.; Nielsen, H. Assessing the accuracy of prediction algorithms for classification: An overview. *Bioinformatics,* **2000**, *16*(5), 412-424.
http://dx.doi.org/10.1093/bioinformatics/16.5.412       PMID: 10871264

[23]    Bradley, A.P. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit.,* **1997**, *30*(7), 1145-1159.
http://dx.doi.org/10.1016/S0031-3203(96)00142-2

[24]    Abbas, M.; El-Manzalawy, Y. Machine learning based refined differential gene expression analysis of pediatric sepsis. *BMC Med. Genom.,* **2020**, *13*(1), 122.
http://dx.doi.org/10.1186/s12920-020-00771-4 PMID: 32859206

[25]    Liao, Y.; Wang, J.; Jaehnig, E.J.; Shi, Z.; Zhang, B. WebGestalt 2019: Gene set analysis toolkit with revamped UIs and APIs. *Nu-cleic Acids Res.,* **2019**, *47*(W1), W199-W205.
http://dx.doi.org/10.1093/nar/gkz401 PMID: 31114916

[26]    Warde-Farley, D.; Donaldson, S.L.; Comes, O.; Zuberi, K.; Bad-rawi, R.; Chao, P.; Franz, M.; Grouios, C.; Kazi, F.; Lopes, C.T.; Maitland, A.; Mostafavi, S.; Montojo, J.; Shao, Q.; Wright, G.; Bader, G.D.; Morris, Q. The GeneMANIA prediction server: Bio-logical network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.,* **2010**, *38*(Suppl. 2), W214-20.
http://dx.doi.org/10.1093/nar/gkq537 PMID: 20576703

[27]    Zhao, J.C.; Zhang, L.X.; Zhang, Y.; Shen, Y.F. The differential regulation of Gap43 gene in the neuronal differentiation of P19 cells. *J. Cell. Physiol.,* **2012**, *227*(6), 2645-2653.
http://dx.doi.org/10.1002/jcp.23006 PMID: 21938722

[28]    Van Regemorter, N.; Gheuens, J.; Noppe, M.; Vamos, E.; Seller, M.J.; Lowenthal, A. Value of glial fibrillary acidic protein deter-mination in amniotic fluid for prenatal diagnosis of neural tube de-fects. *Clin. Chim. Acta,* **1987**, *165*(1), 83-88.
http://dx.doi.org/10.1016/0009-8981(87)90221-X PMID: 2440621

[29]    Pinner, E.; Gruper, Y.; Ben Zimra, M.; Kristt, D.; Laudon, M.; Naor, D.; Zisapel, N. CD44 splice variants as potential players in Alzheimer's disease pathology. *J. Alzheimers Dis.,* **2017**, *58*(4), 1137-1149.
http://dx.doi.org/10.3233/JAD-161245 PMID: 28550248

[30]    Dzwonek, J.; Wilczynski, G.M. CD44: Molecular interactions, signaling and functions in the nervous system. *Front. Cell. Neuro-sci.,* **2015**, *9*, 175.
http://dx.doi.org/10.3389/fncel.2015.00175 PMID: 25999819

[31]    Pastural, E.; Ersoy, F.; Yalman, N.; Wulffraat, N.; Grillo, E.; Ozkinay, F.; Tezcan, I.; Gediköglu, G.; Philippe, N.; Fischer, A.; de Saint Basile, G. Two genes are responsible for Griscelli syn-drome at the same 15q21 locus. *Genomics,* **2000**, *63*(3), 299-306.
http://dx.doi.org/10.1006/geno.1999.6081 PMID: 10704277

[32]    Tang, F.Y.; Ma, L.; Tam, P.O.S.; Pang, C.P.; Tham, C.C.; Chen, L.J. Genetic association of the PARL-ABCC5-HTR3D-HTR3C locus with primary angle-closure glaucoma in Chinese. *Invest. Ophthalmol. Vis. Sci.,* **2017**, *58*(10), 4384-4389.
http://dx.doi.org/10.1167/iovs.17-22304 PMID: 28813580

[33]    Scalabrino, G.; Veber, D.; Tredici, G. Relationships between coba-lamin, epidermal growth factor, and normal prions in the myelin maintenance of central nervous system. *Int. J. Biochem. Cell Biol.,* **2014**, *55*, 232-241.
http://dx.doi.org/10.1016/j.biocel.2014.09.011 PMID: 25239885

[34]    Martins, I.J. Anti-aging genes improve appetite regulation and reverse cell senescence and apoptosis in global populations. *Adv. Aging Res.,* **2016**, *5*(1), 9-26.
http://dx.doi.org/10.4236/aar.2016.51002

[35]    Martins, I.J. Single gene inactivation with implications to diabetes and multiple organ dysfunction syndrome. *J. Clin. Epigenet.,* **2017**, *3*(3), 24.
http://dx.doi.org/10.21767/2472-1158.100058

[36]    Li, G.; Jiapaer, Z.; Weng, R.; Hui, Y.; Jia, W.; Xi, J.; Wang, G.; Zhu, S.; Zhang, X.; Feng, D.; Liu, L.; Zhang, X.; Kang, J. Dysre-gulation of the SIRT1/OCT6 axis contributes to environmental stress-induced neural induction defects. *Stem Cell Reports,* **2017**, *8*(5), 1270-1286.
http://dx.doi.org/10.1016/j.stemcr.2017.03.017 PMID: 28434941

[37]    Boulet, S.L.; Yang, Q.; Mai, C.; Kirby, R.S.; Collins, J.S.; Rob-bins, J.M.; Meyer, R.; Canfield, M.A.; Mulinare, J. Trends in the postfortification prevalence of spina bifida and anencephaly in the United States. *Birth Defects Res. A Clin. Mol. Teratol.,* **2008**, *82*(7), 527-532.
http://dx.doi.org/10.1002/bdra.20468 PMID: 18481813

[38]    Geisel, J. Folic acid and neural tube defects in pregnancy: A re-view. *J. Perinat. Neonatal Nurs.,* **2003**, *17*(4), 268-279.
http://dx.doi.org/10.1097/00005237-200310000-00005       PMID: 14655787

[39] Salbaum, J.M.; Kappen, C. Neural tube defect genes and maternal diabetes during pregnancy. *Birth Defects Res. A Clin. Mol. Teratol.,* **2010**, *88*(8), 601-611.
http://dx.doi.org/10.1002/bdra.20680 PMID: 20564432

[40] Steele, J.W.; Lin, Y.L.; Chen, N.; Wlodarczyk, B.J.; Chen, Q.; Attarwala, N.; Venkatesalu, M.; Cabrera, R.M.; Gross, S.S.; Finnell, R.H. Embryonic hypotaurine levels contribute to strain-dependent susceptibility in mouse models of valproate-induced neural tube defects. *Front. Cell Dev. Biol.,* **2022**, *10*, 832492.
http://dx.doi.org/10.3389/fcell.2022.832492 PMID: 35265619

[41] Shen, Y.; Mani, S.; Donovan, S.L.; Schwob, J.E.; Meiri, K.F. Growth-associated protein-43 is required for commissural axon guidance in the developing vertebrate nervous system. *J. Neurosci.,* **2002**, *22*(1), 239-247.
http://dx.doi.org/10.1523/JNEUROSCI.22-01-00239.2002 PMID: 11756507

[42] Arstikaitis, P.; Gauthier-Campbell, C.; Huang, K.; El-Husseini, A.; Murphy, T.H. Proteins that promote filopodia stability, but not number, lead to more axonal-dendritic contacts. *PLoS One,* **2011**, *6*(3), e16998.
http://dx.doi.org/10.1371/journal.pone.0016998 PMID: 21408225

[43] Gispen, W.H.; Nielander, H.B.; De Graan, P.N.; Oestreicher, A.B.; Schrama, L.H.; Schotman, P. Role of the growth-associated protein B-50/GAP-43 in neuronal plasticity. *Mol. Neurobiol.,* **1991**, *5*(2-4), 61-85.
http://dx.doi.org/10.1007/BF02935540 PMID: 1840422

[44] Strittmatter, S.M.; Vartanian, T.; Fishman, M.C. GAP-43 as a plasticity protein in neuronal form and repair. *J. Neurobiol.,* **1992**, *23*(5), 507-520.
http://dx.doi.org/10.1002/neu.480230506 PMID: 1431834

[45] Allegra Mascaro, A.L.; Cesare, P.; Sacconi, L.; Grasselli, G.; Mandolesi, G.; Maco, B.; Knott, G.W.; Huang, L.; De Paola, V.; Strata, P.; Pavone, F.S. *In vivo* single branch axotomy induces GAP-43-dependent sprouting and synaptic remodeling in cerebellar cortex. *Proc. Natl. Acad. Sci. USA,* **2013**, *110*(26), 10824-10829.
http://dx.doi.org/10.1073/pnas.1219256110 PMID: 23754371

[46] Basi, G.S.; Jacobson, R.D.; Virág, I.; Schilling, J.; Skene, J.H. Primary structure and transcriptional regulation of GAP-43, a protein associated with nerve growth. *Cell,* **1987**, *49*(6), 785-791.
http://dx.doi.org/10.1016/0092-8674(87)90616-7 PMID: 3581170

[47] Nguyen, L.; He, Q.; Meiri, K.F. Regulation of GAP-43 at serine 41 acts as a switch to modulate both intrinsic and extrinsic behaviors of growing neurons, *via* altered membrane distribution. *Mol. Cell. Neurosci.,* **2009**, *41*(1), 62-73.
http://dx.doi.org/10.1016/j.mcn.2009.01.011 PMID: 19249369

[48] Mishra, R.; Gupta, S.K.; Meiri, K.F.; Fong, M.; Thostrup, P.; Juncker, D.; Mani, S. GAP-43 is key to mitotic spindle control and centrosome-based polarization in neurons. *Cell Cycle,* **2008**, *7*(3), 348-357.
http://dx.doi.org/10.4161/cc.7.3.5235 PMID: 18235238

[49] Kawasaki, T.; Nishio, T.; Kawaguchi, S.; Kurosawa, H. Spatiotemporal distribution of GAP-43 in the developing rat spinal cord: A histological and quantitative immunofluorescence study. *Neurosci. Res.,* **2001**, *39*(3), 347-358.
http://dx.doi.org/10.1016/S0168-0102(00)00234-0 PMID: 11248375

[50] Huber, M.; Siegenthaler, G.; Mirancea, N.; Marenholz, I.; Nizetic, D.; Breitkreutz, D.; Mischke, D.; Hohl, D. Isolation and characterization of human repetin, a member of the fused gene family of the epidermal differentiation complex. *J. Invest. Dermatol.,* **2005**, *124*(5), 998-1007.
http://dx.doi.org/10.1111/j.0022-202X.2005.23675.x PMID: 15854042

[51] Wang, S.; Ren, H.; Xu, J.; Yu, Y.; Han, S.; Qiao, H.; Cheng, S.; Xu, C.; An, S.; Ju, B.; Yu, C.; Wang, C.; Wang, T.; Yang, Z.; Taylor, E.W.; Zhao, L. Diminished serum repetin levels in patients with schizophrenia and bipolar disorder. *Sci. Rep.,* **2015**, *5*(1), 7977.
http://dx.doi.org/10.1038/srep07977 PMID: 25613293

[52] Lopez, J.; Mikaelian, I.; Gonzalo, P. Amniotic fluid glial fibrillary acidic protein (AF-GFAP), a biomarker of open neural tube defects. *Prenat. Diagn.,* **2013**, *33*(10), 990-995.
http://dx.doi.org/10.1002/pd.4181 PMID: 23784867

[53] Petzold, A.; Stiefel, D.; Copp, A.J. Amniotic fluid brain-specific proteins are biomarkers for spinal cord injury in experimental myelomeningocele. *J. Neurochem.,* **2005**, *95*(2), 594-598.
http://dx.doi.org/10.1111/j.1471-4159.2005.03432.x PMID: 16190875

[54] Petzold, A.; Eikelenboom, M.J.; Gveric, D.; Keir, G.; Chapman, M.; Lazeron, R.H.; Cuzner, M.L.; Polman, C.H.; Uitdehaag, B.M.; Thompson, E.J.; Giovannoni, G. Markers for different glial cell responses in multiple sclerosis: Clinical and pathological correlations. *Brain,* **2002**, *125*(Pt 7), 1462-1473.
http://dx.doi.org/10.1093/brain/awf165 PMID: 12076997

[55] O'Callaghan, J.P.; Sriram, K. Glial fibrillary acidic protein and related glial proteins as biomarkers of neurotoxicity. *Expert Opin. Drug Saf.,* **2005**, *4*(3), 433-442.
http://dx.doi.org/10.1517/14740338.4.3.433 PMID: 15934851

[56] George, T.M.; Cummings, T.J. The immunohistochemical profile of the myelomeningocele placode: Is the placode normal? *Pediatr. Neurosurg.,* **2003**, *39*(5), 234-239.
http://dx.doi.org/10.1159/000072867 PMID: 14512686

[57] Yan, Y.; Zuo, X.; Wei, D. Concise review: Emerging role of CD44 in cancer stem cells: A promising biomarker and therapeutic target. *Stem Cells Transl. Med.,* **2015**, *4*(9), 1033-1043.
http://dx.doi.org/10.5966/sctm.2015-0048 PMID: 26136504

[58] Haegel, H.; Dierich, A.; Ceredig, R. CD44 in differentiated embryonic stem cells: Surface expression and transcripts encoding multiple variants. *Dev. Immunol.,* **1994**, *3*(4), 239-246.
http://dx.doi.org/10.1155/1994/25484 PMID: 7542511

[59] Jackson, R.L.; Busch, S.J.; Cardin, A.D. Glycosaminoglycans: Molecular properties, protein interactions, and role in physiological processes. *Physiol. Rev.,* **1991**, *71*(2), 481-539.
http://dx.doi.org/10.1152/physrev.1991.71.2.481 PMID: 2006221

[60] Corbel, C.; Lehmann, A.; Davison, F. Expression of CD44 during early development of the chick embryo. *Mech. Dev.,* **2000**, *96*(1), 111-114.
http://dx.doi.org/10.1016/S0925-4773(00)00347-6 PMID: 10940630

[61] Zhu, H.; Mitsuhashi, N.; Klein, A.; Barsky, L.W.; Weinberg, K.; Barr, M.L.; Demetriou, A.; Wu, G.D. The role of the hyaluronan receptor CD44 in mesenchymal stem cell migration in the extracellular matrix. *Stem Cells,* **2006**, *24*(4), 928-935.
http://dx.doi.org/10.1634/stemcells.2005-0186 PMID: 16306150

[62] Wheatley, S.C.; Isacke, C.M.; Crossley, P.H. Restricted expression of the hyaluronan receptor, CD44, during postimplantation mouse embryogenesis suggests key roles in tissue formation and patterning. *Development,* **1993**, *119*(2), 295-306.
http://dx.doi.org/10.1242/dev.119.2.295 PMID: 7507029

[63] Sahin Inan, Z.D.; Unver Saraydin, S. Immunohistochemical profile of CD markers in experimental neural tube defect. *Biotech. Histochem.,* **2019**, *94*(8), 617-627.
http://dx.doi.org/10.1080/10520295.2019.1622783 PMID: 31184499

[64] Zöller, M. CD44: Can a cancer-initiating cell profit from an abundantly expressed molecule? *Nat. Rev. Cancer,* **2011**, *11*(4), 254-267.
http://dx.doi.org/10.1038/nrc3023 PMID: 21390059