Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Multi-Omics Integration Reveals Three Potential Therapeutic Targets for Spina Bifida: NPC2, PCDH12 and IL10RB

Chensong Sun
Medical Research Public Service Center, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University

Ga Liu
Department of Ultrasound, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University

Ziwei Zhong
Center for Molecular Diagnosis and Precision Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 Dongyue Dadao

Xiao Lei
Center for Molecular Diagnosis and Precision Medicine, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University, 1519 Dongyue Dadao

Haitao Deng
Center for Assisted Reproduction, The First Affiliated Hospital of Nanchang University

Fang Huang
Infection Control Division, The First Affiliated Hospital, Jiangxi Medical College, Nanchang University

Jin Zeng
ndyfy02763@ncu.edu.cn

Department of Orthopedics, The First Affiliated Hospital of Nanchang University

Research Article

Keywords: Spina bifida, Mendelian randomization, pQTL, Biomarkers, Molecular docking, Molecular dynamics simulation

Posted Date: June 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9882445/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Italian Journal of Pediatrics on October 3rd, 2026. See the published version at https://doi.org/10.1186/s13052-026-02356-w.

Abstract

Background

Neural tube defects may cause major congenital malformations, including spina bifida (SB). Despite advances in the prevention, prenatal diagnosis and surgical intervention, reliable biomarkers and therapeutic targets remain limited.

Methods

This study integrated Mendelian randomization (MR) and GEO transcriptomic data from placental and fetal tissues to identify potential biomarkers and therapeutic targets for SB. Three large-scale pQTL datasets (Decode, UKB_PPP, and Prot_a) and two tissue gene expression datasets (GSE33111 and GSE252552) were included. MR analyses were performed using both cis- and trans-pQTLs, followed by differential expression analysis, GO/KEGG enrichment, GSMR, and single-gene GSEA to prioritize genes. Finally, PheWAS, drug prediction, molecular docking and molecular dynamics (MD) simulations were employed to evaluate druggability of the targets.

Results

By integrating MR and transcriptomic results, three downregulated candidate genes were identified in SB cases: NPC2 (OR = 0.753, 95% CI: 0.646–0.879), PCDH12 (OR = 0.636, 95% CI: 0.507–0.798), and IL10RB (OR = 0.665, 95% CI: 0.549–0.805). These genes are involved in cholesterol metabolism, cell adhesion, and immune regulation, respectively. Enrichment analysis implicated the JAK–STAT, CCRI, and CAMs pathways. Docking and MD simulations confirmed NPC2-cholesterol binding as an endogenous positive control, while docking supported tamibarotene as a candidate IL10RB-targeting compound, although its binding stability requires further validation.

Conclusion

Through a multi-omics integrative strategy, this study identified NPC2, PCDH12, and IL10RB as novel biomarkers and druggable targets for SB. Cholesterol acted as an endogenous control validating our approach, while tamibarotene emerged as a promising therapeutic compound.

Introduction

Spina bifida (SB) is a common and clinically significant subtype of neural tube defects (NTDs), primarily resulting from neural tube closure failure during early embryonic development. SB occupies a prominent position among congenital malformations of the central nervous system[1], with a global prevalence of approximately 1 per 1,000 live births, varying regionally due to differences in nutrition, genetics, and

healthcare[2]. The etiology of SB is complex and multifactorial, involving disturbances in folate metabolism, the planar cell polarity (PCP) signaling pathway, inositol-related metabolic processes, and environmental factors like polycyclic aromatic hydrocarbons[3–6]. While ultrasound and maternal serum alpha-fetoprotein (AFP) tests are common diagnostic tools, their sensitivity and specificity are limited, complicating early detection[7]. Current therapies mainly focus on surgical correction, which cannot reverse neurological damage or address the underlying pathophysiology[8]. Therefore, identifying reliable biomarkers and novel therapeutic targets for SB is urgently needed.

Quantitative trait loci (QTL) mapping links genetic markers like single nucleotide polymorphisms (SNPs) to complex traits, offering valuable insights when integrated with genomics, proteomics, and phenotypic data such as genome-wide association studies (GWAS)[9]. Combining QTL with GWAS overcomes some limitations of GWAS alone[10]. Protein quantitative trait loci (pQTL) are critical genetic determinants of protein function, revealing disease biology and serving as valuable resources for identifying biomarkers and therapeutic targets[11]. Despite the identification of biomarkers like PCSK9[12] for SB through proteomics, phenotype-to-protein observations alone are insufficient for uncovering causal relationships or mechanisms.

Mendelian randomization (MR) uses genetic variants as instruments to estimate the causal impact of exposures on outcomes, simulating randomized trials. The random inheritance of these variants helps MR address confounding and reverse causation, common biases in observational studies[13]. MR has been widely applied to explore causal relationships in various diseases and identify novel biomarker candidates, such as those for multiple sclerosis. For example, MR analysis of pQTL data from brain, cerebrospinal fluid, and plasma identified 30 genes linked to Alzheimer's disease risk[14]. In congenital disorders, pyruvate dehydrogenase kinase isozyme 1 and mitochondrial ATP synthase subunit β have been causally associated with congenital heart disease and identified as protective factors[15]. The Gene Expression Omnibus (GEO) is the largest publicly accessible repository of transcriptomic data, encompassing gene expression profiles across species, tissues, and disease states[16]. Comparative analysis of differential expression between SB cases and healthy controls in GEO datasets identifies differentially expressed genes (DEGs), providing insights into disease mechanisms, potential biomarkers, and therapeutic targets.

In this study, we conducted a MR analysis integrating pQTL-based proteomic data with gene expression data from the GEO database to explore the causal relationships and differential gene patterns associated with SB. Furthermore, we employed generalized summary-data-based Mendelian randomization (GSMR), single-gene gene set enrichment analysis (GSEA), and phenome-wide association study (PheWAS) to identify potential biomarkers of SB and assess their potential as druggable targets. A detailed schematic overview of the study design is presented in Fig. 1.

Methods

Data sources

Three large-scale pQTL datasets were utilized. The Decode dataset, generated using the SomaScan platform, included protein quantification for 4,719 targets in 35,559 Icelandic individuals[17]. The UK Biobank Pharma Proteomics Project (UKB_PPP) dataset, based on the Olink platform, provided measurements of 2,923 proteins in 54,219 UK participants[18]. The Prot_a (SUN_BB) dataset, also using SomaScan, contained measurements of 2,994 proteins in 3,301 individuals of European ancestry[19]. Genetic association data for SB were obtained from PAN_UKB (https://pan.ukbb.broadinstitute.org/downloads), comprising 210 cases and 408,394 controls. An overview of all datasets used is provided in **Table S1**.

The GSE252552 and GSE33111 datasets, derived from placental and fetal tissues, were retrieved from the GEO database. For GSE252552, case and control samples were matched based on gestational age, including 12 cases and 12 controls. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE252552). For GSE33111, brain and spinal cord samples were selected, comprising 6 cases and 8 controls, skin tissue samples were excluded from the analysis (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE33111).

Proteome-wide MR

Instrumental variables (IVs) were determined following standard GWAS criteria. SNPs that achieved genome-wide significance ($\mathbf{p} < 5\times10^{-8}$) were initially included, followed by LD pruning with a cutoff of $R^2 < 0.1$ and a 10,000 kb distance threshold[20]. Each instrument was assessed with the F statistic; SNPs with F values above 10 were retained to avoid weak-instrument bias[21, 22]. To assess heterogeneity between instruments, we utilized Cochran's Q test. Additionally, the Steiger directionality test was employed to confirm the direction of causal relationships[23]. For single SNPs, the Wald ratio method was used23. For multiple SNPs, the inverse variance weighted (IVW) method was employed, supplemented by weighted median, MR-Egger regression, and simple mode analyses[24–26]. Sensitivity checks, including MR-PRESSO for outlier detection, were performed[27]. Significance thresholds were adjusted for multiple comparisons using the Bonferroni method. All MR analyses were conducted using the “TwoSampleMR” package[28] in R.

Transcriptomic analyses (DEG & GO/KEGG)

Differential expression was performed with the “limma” package[29] on GSE33111 and GSE252552, using normalized probe-level data. DEGs were defined as |log2 fold change (FC)| > 0.25 and $\mathbf{p} < 0.05$. Overlaps between datasets were visualized with Venny 2.1.0 (https://bioinfogp.cnb.csic.es/tools/venny/index.html).

Functional enrichment was conducted with the “clusterProfiler” package[30], including Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses. For enriched terms, significance was set at $\mathbf{p} < 0.05$.

Candidate gene prioritization

Genes identified through MR with Bonferroni corrected **p** < 0.05 were intersected with DEGs to identify genes with concordant expression directions. Volcano plots were created to visualize differential expression, highlighting genes consistent with MR findings.

GSMR analyses

GSMR was performed using the GSMR2 R package (version 1.1.1)[31], assessing adjusted causal associations between pQTLs and SB while minimizing pleiotropic effects. GSMR improves statistical power by using multiple independent SNPs, minimizing horizontal pleiotropy. Associations with **p** < 0.05 were considered statistically significant.

Single-gene GSEA analysis

To investigate biological mechanisms associated with the prioritized genes, we performed single-gene GSEA[32] using the “c2.cp.kegg.Hs.symbols” gene set from the Molecular Signatures Database (MSigDB). The most significantly enriched KEGG pathways were identified for each gene, and the top six pathways were selected based on the Normalized Enrichment Score (NES). The analysis was conducted using the “clusterProfiler” package and the org.Hs.eg.db30 annotation package in R..

PheWAS

Public data from the UK Biobank 500K WGS v2 cohort (~ 500,000 participants) were used to evaluate prioritized genes related to SB. A gene-level phenome-wide association study (PheWAS) was conducted using the AzPheWAS platform (https://azphewas.com/) was used to assess associations. Associations with $\mathbf{p} \leq 1\times10^{-8}$ were considered genome-wide significant, and those with $\mathbf{p} \leq 1\times10^{-6}$ were considered suggestive[33].

Drug target prediction

Potential drug-gene interactions were explored using the Enrichr platform (https://maayanlab.cloud/Enrichr/enrich). Candidate gene sets were uploaded to the Drug Signature Database (DSigDB, http://dsigdb.tanlab.org/DSigDBv1.0), and compounds linked to these gene sets were screened. Enrichment results were prioritized within Enrichr, and compounds with **p** < 0.05 were considered potential therapeutic candidates.

Molecular docking

Candidate small-molecule structures were first retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov/) in SDF format, imported into SYBYL-X 2.1.1[34] for conversion to mol2, and subjected to energy minimization. Protein coordinates for the human NPC1-NPC2 complex (PDB ID: 6W5V) and the IL10RB complex (PDB ID: 9BPV) were downloaded from the Protein Data Bank (https://www2.rcsb.org/). Using PyMOL[35], the NPC2 and IL10RB chains were isolated; crystallographic ligands and solvent were removed, hydrogen atoms were introduced, and incomplete side chains were completed. Docking pockets were delineated around the native binding cavities. Protein-ligand docking was then carried out with Surflex-Dock GeomX in SYBYL-X to estimate interaction profiles and compare binding affinities.

Molecular dynamics (MD) simulation

To evaluate the stability of the docking results, MD simulations were performed. The CHARMM36m force field was applied to proteins, while ligands were parameterized using the CHARMM General Force Field (CGenFF)[36]. The system was solved using the TIP3P water model, and a 1.2 nm cutoff radius was applied using the Verlet list algorithm.

Energy minimization was conducted with the steepest descent method, followed by canonical (NVT) ensemble and isothermal–isobaric (NPT) ensemble equilibration under 303.15 K and 1 bar, using the stochastic velocity-rescaling (V-rescale) thermostat and the stochastic cell-rescaling (C-rescale) barostat. A 50 ns production run was then carried out under the same conditions. Structural stability was assessed by calculating root-mean-square deviation (RMSD), root-mean-square fluctuation (RMSF), and solvent-accessible surface area (SASA). All simulations and analyses were performed using GROMACS 2022.2[37].

Results

Proteome-wide MR results

Analyses were performed across three pQTL datasets using both cis- and trans-acting variants, with SNP details provided in **Tables S2–S4**. Using SB data from the PAN_UKB dataset as the discovery outcome, 34 proteins showed statistically significant causal associations after Bonferroni correction (IVW for multiple instruments or Wald ratio for single instruments). These findings are shown in Fig. 2, where positively associated proteins are labeled and positioned on the volcano plot. Full MR estimates are provided in **Table S5**. For proteins with ≥ 3 instrumental SNPs, heterogeneity was assessed using Cochran's Q test and horizontal pleiotropy was evaluated via MR-PRESSO; all Q values and Global Test **p** were > 0.05 (**Tables S6 and S7**), indicating no evidence of heterogeneity or pleiotropy. Steiger directionality tests further supported the direction from protein levels to SB (**Table S8**).

Transcriptomic findings (DEG overlap & GO/KEGG).

After platform-normalized differential expression, we identified 3,422 up / 4,073 down DEGs in GSE252552 and 1,492 up / 1,380 down in GSE33111. Venn diagrams (Fig. 3A-B) revealed 229 overlapping upregulated and 362 overlapping downregulated genes (**Tables S9-S12**). GO enrichment of the common DEGs (**p** < 0.05) yielded 999 terms (796 BP, 103 MF, 100 CC) predominantly related to immune regulation, cytoskeletal organization, vesicle trafficking, and signal transduction (Fig. 3C; **Table S13**); KEGG analysis highlighted NOD-like receptor, PI3K–Akt, and MAPK pathways, with additional signals in neurodegenerative disease pathways (e.g., Alzheimer's, ALS) (Fig. 3D; **Table S14**).

Concordant prioritization of candidate genes

DEGs from GSE252552 and GSE33111 are shown in Figs. 4A and 4B. We intersected MR-significant proteins with DEGs, selecting upregulated genes with OR > 1 and downregulated genes with 0 < OR < 1. Based on IVW results (**Table S5)**, NPC2 (OR = 0.753, 95% CI: 0.646–0.879), PCDH12 (OR = 0.636, 95% CI: 0.507–0.798), and IL10RB (OR = 0.665, 95% CI: 0.549–0.805) overlapped with the downregulated DEGs, and were prioritized as candidate genes (Fig. 4**)**. Scatter and funnel plots are provided in **Supplementary Figs. 1 and 2**.

GSMR results

GSMR analyses supported significant associations for all three genes (**p** < 0.05) (Table 1**)**, supporting robust causal relationships between their expression or functional alteration and the risk of SB, even after adjusting for potential pleiotropic effects. These findings further reinforce the evidence for their involvement in disease pathogenesis and highlight their potential as druggable targets, warranting further functional studies and therapeutic development.

Table 1
GSMR estimates for NPC2, IL10RB, and PCDH12

Phenotype	Beta	SE	OR	P value
NPC2	-0.401	0.163	0.670	0.010
PCDH12	-0.392	0.189	0.676	0.038
IL10RB	-0.318	0.152	0.728	0.037
Notes: Beta is the GSMR estimate on the log-odds scale. SE, standard error. OR, odds ratio.				

Single-gene GSEA results

Single-gene GSEA results are summarized in **Tables S15**-**S17**, and the top six enriched KEGG pathways for each gene are shown in Figs. 5A-C. NPC2 was enriched in 45 pathways, PCDH12 in 49, and IL10RB in 78. NPC2 and IL10RB were enriched in cytokine-cytokine receptor interaction (CCRI) and Leishmaniasis pathways, while PCDH12 and IL10RB shared enrichment in cell adhesion molecules (CAMs) and viral myocarditis pathways. Additional signals included NPC2 in chemokine signaling and JAK-STAT

pathways, PCDH12 in antigen processing and presentation, and IL10RB in the peroxisome pathway. These results provide new insights into the potential biological roles of these genes in the pathogenesis of SB.

PheWAS results

Using AzPheWAS with UK Biobank WGS v2 data, no strong associations were observed for binary traits. For continuous traits, NPC2 expression was negatively associated with multiple metabolic proteins and respiratory-related phenotypes; PCDH12 was negatively correlated with neural protein expression; IL10RB showed negative associations with inflammatory protein levels and a modest association with tumor-related proteins (**Table S18**). Overall, these findings provide preliminary safety signals for downstream prioritization.

Predicted drug candidates

The top 10 compounds ranked by significance are summarized in Table 2. Drug associations were identified for NPC2 and IL10RB, while no eligible drugs were found for PCDH12. Tamibarotene (CTD 00002527) showed the most significant predicted interaction with both NPC2 and IL10RB, and decitabine (CTD 00000750) also overlapped across both genes. Additional IL10RB-linked compounds included fluoride (CTD 00005982) and ferrous sulfate (CTD 00001009); for NPC2, NSC25485 (CTD 00006443) showed relatively strong significance.

Table 2
Signature-based compound enrichment and candidate ranking

Candidate	P value	Odds ratio	Combined score	Gene(s)
Tamibarotene	0.002	69.803	424.160	NPC2; IL10RB
Fluoride	0.008	184.657	886.403	IL10RB
Ferrous sulfate	0.009	178.045	848.318	IL10RB
0179445-0000 PC3 DOWN	0.013	118.530	517.553	IL10RB
NSC25485	0.019	77.008	303.704	NPC2
Isotretinoin HL60 UP	0.02	73.563	286.821	NPC2
Decitabine	0.023	20.244	76.513	NPC2; IL10RB
Cholesterol BOSS	0.028	52.402	186.935	NPC2
Sodium dichromate	0.031	47.802	166.241	NPC2
1-NITROPYRENE	0.035	42.412	142.557	IL10RB

Notes: Enrichment against DSigDB (via Enrichr). Entries such as “HL60 UP” or “PC3 DOWN” denote expression signatures (not compound names).

Molecular docking results

Docking results are summarized in Table 3. A total score > 7 was considered highly reliable, and scores between 5 and 7 indicated good binding affinity. Based on predicted targets, top compounds were selected for structural inspection. NPC2 docked strongly with cholesterol (total score = 7.262), consistent with its endogenous ligand role and serving as a positive control (Fig. 6A). IL10RB docking with tamibarotene yielded a total score of 6.019, suggesting favorable binding (Fig. 6B). Together, these results support the workflow and highlight tamibarotene as a candidate compound.

Table 3
Molecular docking results for NPC2 (6W5V) and IL10RB (9BPV).

	Drug	Total score	Crash	Polar
NPC2	Cholesterol	7.262	-3.810	0
	Tamibarotene	6.091	-1.202	0.003
	NSC25485	6.024	-3.043	1.036
	Decitabine	5.174	-0.944	1.924
	Sodium dichromate	2.312	-0.308	1.296
IL10RB	Tamibarotene	6.019	-1.032	2.359
	Decitabine	4.694	-0.572	5.283
	Ferrous sulfate	3.099	-0.015	2.371
	1-NITROPYRENE	2.961	-1.128	1.389
	Fluoride	NA	NA	NA

Notes: NA, Not applicable (not suitable for docking). Surflex-Dock total score is a dimensionless affinity score; >7 indicates highly reliable, 5–7 indicates good affinity. Crash refers to the calculated energy, and Polar is the polarity of the molecule.

MD simulation results

RMSD, RMSF, and SASA were used to evaluate stability and interface properties (Fig. 7A, C, **and E**). For NPC2 and the NPC2-cholesterol complex, both systems stabilized around ~ 5 ns, with RMSD fluctuations within 0-0.02 nm. RMSF remained low across most residues, and SASA was stable, suggesting a compact binding interface with minimal conformational disturbance upon ligand binding. For the IL10RB-tamibarotene complex (Fig. 7B, D, **and F**), the structure was relatively stable during 0–8 ns and remained stable from 8–33 ns, followed by a secondary RMSD increase between 34–50 ns. RMSF showed limited residue-level fluctuations, while SASA changes aligned with the later RMSD increase, suggesting conformational adjustments during the late simulation stage.

Discussion

As a common neural tube defect, SB arises from failure of posterior neural tube closure and poses a substantial global health burden[38].Although progress has been made in the diagnosis and treatment of NTDs-including biomarker identification in maternal serum[39], amniotic fluid[40], fetal cerebrospinal fluid[41], and exosomes[42]-the causal mechanisms and directionality of these associations remain poorly understood. In this study, we integrated proteome-wide MR, microarray-based differential gene expression, and drug target prediction to identify biomarkers for SB, clarify causal relationships, and explore potential therapeutic targets. MR identified 34 proteins with potential causal associations with SB. Differential expression analyses based on GEO datasets indicated that overlapping DEGs were enriched in immune responses, inflammatory pathways, and neurodegenerative disease-related processes. By intersecting downregulated genes with MR-identified protective factors ($0 < OR < 1$), we prioritized three downregulated genes-NPC2, PCDH12, and IL10RB-as candidate genes. We further evaluated potential adverse effects using PheWAS and assessed druggability through compound screening, molecular docking, and MD simulations, providing a foundation for future drug development in SB.

Previous studies have implicated the JAK-STAT pathway as a central mediator of neuroinflammation in SB[43]. Our findings support this association by demonstrating enrichment of the JAK-STAT pathway in single-gene GSEA for NPC2. The PD-1/PD-L1 pathway is an essential immune regulatory mechanism in neurodevelopment, and JAK-STAT signaling can mediate PD-1 upregulation on immune cells[44]. JAK-STAT signaling is also involved in neuronal activity regulation and has been implicated in congenital disorders, including Noonan syndrome[45], mitochondrial diseases[46], and Down syndrome[47]. NPC2 encodes a soluble lysosomal protein crucial for intracellular cholesterol transport. NPC2 deficiency disrupts cholesterol storage and alters sphingolipid metabolism in neural tissues, leading to cholesterol accumulation and progressive neurodegeneration[48], characteristic of Niemann-Pick disease type C. Reduced NPC2 expression has also been linked to cholesterol accumulation in skeletal muscle in ALS, and NPC2 dysfunction may occur before clinical onset[49]. In brain tissues from adolescent patients with neuronal ceroid lipofuscinosis (NCL), endosomal proteomic and lipidomic profiling revealed mechanisms resembling NPC, with NPC2 playing a pivotal role[50]. Consistent with these reports, our MR and differential expression analyses indicate that downregulation of NPC2 is causally associated with SB; however, experimental work is needed to clarify the underlying mechanisms.

PCDH12, a gene belonging to the protocadherin (PCDH) family, encodes one of the CAMs[51] that is abundantly expressed in the central nervous system and is crucial for brain development and the formation of synapse-specific memories[52]. Prenatal MRI and postnatal assessments have linked PCDH12 mutations with congenital microcephaly and hypothalamic–midbrain dysplasia, alongside marked reductions in PCDH12 expression[53]. Biallelic loss-of-function mutations in PCDH12 have been associated with cerebral calcification, severe epilepsy, and major brain malformations[54]. Additionally, patient-derived induced pluripotent stem cells (iPSCs) from individuals with diencephalic-mesencephalic junction dysplasia-accompanied by craniofacial malformations and motor neuron deficits-demonstrated

loss of PCDH12 expression, impaired axonal growth, and endothelial dysfunction[55]. In both human cerebral organoids and 2D neural progenitor models, PCDH12 deficiency disrupted organoid morphology, induced premature neural progenitor differentiation, impaired neuronal migration, and resulted in abnormal cortical structure formation[56]. In our data, PCDH12 was enriched in the CAMs pathway, together with IL10RB. CAMs mediate cell–cell and cell–matrix interactions[57] and are essential during embryonic neural tube closure. Prior studies report that NCAM1 rs2298526 is associated with SB in familial NTDs, suggesting CAM-related polymorphisms may increase susceptibility[58]. Mechanistically, Vangl2 interacts with CDH2 during convergent extension, and disruption of either leads to SB[59]. Similarly, CAMs proteins FAT1 and FAT4 co-regulate apical constriction and neural fold closure; loss of FAT1 leads to anterior neural tube closure failure and forebrain malformations[60]. Together, these data support a role for the CAMs pathway-particularly PCDH12-in the pathogenesis of SB.

Although direct evidence linking IL10RB to the CAMs pathway is limited, IL-10 signaling through the IL-10RA/IL-10RB receptor complex activates anti-inflammatory JAK-STAT signaling and suppresses ICAM-1 and VCAM-1 expression, reducing inflammation-induced adhesion[61]. In streptococcal infection models, IL10RB-dependent signaling may indirectly modulate CAMs expression and immune cell migration[62]. In our study, IL10RB and NPC2 were both significantly enriched in the CCRI pathway-a signaling axis previously reported to be upregulated in embryonic tissues affected by SB[63]. For example, SB embryos with GPR161 mutations show changes not only in cAMP signaling but also in CCRI-related genes[64]. Among CCRI genes, LIFR has been proposed to regulate neural tube formation; consistent with our findings, reduced LIFR expression in SB embryos has been associated with SB pathogenesis[65]. Similarly, transcriptomic analyses of amniotic fluid from fetuses with SB have also shown enrichment of the CCRI pathway[66]. Functionally, IL10RB has been linked to neuroprotection, as IL10RB-deficient $CD4^+$ T cells fail to exert neuroprotective effects and instead exacerbate inflammation and neuronal loss[67].In Down syndrome, the interferon receptor cluster including IL10RB is upregulated and may disrupt immune homeostasis, contributing to congenital heart defects[68]. The rs2834167G allele in IL10RB has been associated with congenital heart defects in Down syndrome and may influence development by enhancing IL-10RB expression and downstream signaling[69].

To investigate druggability, we queried DSigDB via Enrichr, then performed molecular docking in SYBYL-X. As expected, NPC2 showed strong binding to cholesterol, consistent with its established role as a key cholesterol transporter within the endosomal-lysosomal system[70]. This finding not only underscores the critical role of NPC2-cholesterol interactions in maintaining intracellular lipid homeostasis but also validates the reliability of our computational approach as a methodological positive control. In contrast, tamibarotene emerged as a novel of interest. Docking analyses suggested favorable binding to IL10RB (and also NPC2), and MD simulations indicated that, while the NPC2-cholesterol complex rapidly reached a stable equilibrium, the IL10RB-tamibarotene complex exhibited structural fluctuations at later stages, reflecting lower stability compared with the endogenous ligand. To our knowledge, prior reports linking tamibarotene to IL10RB or NPC2 are limited. Therefore, further in vitro assays and functional validation studies are warranted to determine its potential pharmacological activity.

Limitations

Despite the integration of MR and GEO data providing insights into SB biomarkers and drug targets, several limitations should be acknowledged. First, statistical power for the MR outcome was limited (210 cases), which may increases the risk of false negatives. To include a broader set of SNPs, we adopted a relaxed LD threshold ($R^2 < 0.1$), which may have weakened the strength of instrumental variables and potentially affected the robustness and sensitivity of the causal estimates. Second, some datasets had limited sample sizes-for example, GSE33111 included only 6 cases and 8 controls-which may have reduced the statistical power. Furthermore, the GEO datasets were derived from both European and Chinese populations, whereas the MR outcome data were exclusively derived from European individuals. The inclusion of SB data from Asian populations would enhance the generalizability and credibility of the findings. From a drug development perspective, although tamibarotene showed high docking scores with both NPC2 and IL10RB, no prior evidence from the literature or drug-target databases supports a functional interaction with these genes. As such, its clinical applicability remains speculative and requires further investigation. Finally, this study is primarily based on bioinformatics analyses and lacks in vitro or in vivo experimental validation of the identified candidate genes and their mechanistic roles in SB. Future studies are needed to functionally confirm these findings and explore their translational potential.

Conclusion

This study integrated MR and GEO transcriptomic data to identify NPC2, PCDH12, and IL10RB as downregulated protective factors in SB. These genes are enriched in key pathways including JAK-STAT, CAMs, and CCRI. Docking and MD simulations validated NPC2-cholesterol binding as an endogenous positive control, while highlighting tamibarotene as a promising IL10RB-targeting compound. The findings shed light on the molecular mechanisms and targeted therapies of SB, though further experimental work is needed to confirm gene functions and assess the bioactivity of candidate compounds.

Abbreviations

AFP
Alpha-fetoprotein
ALS
Amyotrophic lateral sclerosis
API
Application programming interface
BP
Biological process
CAMs
Cell adhesion molecules

CC
Cellular component
CCRI
Cytokine-cytokine receptor interaction
CI
Confidence interval
CSF
Cerebrospinal fluid
DEG
Differentially expressed gene
GEO
Gene expression omnibus
GO
Gene ontology
GSEA
Gene set enrichment analysis
GSMR
Generalized summary-data-based mendelian randomization
GWAS
Genome-wide association study
IV
Instrumental variable
IVW
Inverse variance weighted
KEGG
Kyoto encyclopedia of genes and genomes
LD
Linkage disequilibrium
MD
Molecular dynamics
MF
Molecular function
MR
Mendelian randomization
NES
Normalized enrichment score
NTDs
Neural tube defects
OR
Odds ratio

PCP
Planar cell polarity
PD
1-Programmed cell death protein 1
PD
L1-Programmed death-ligand 1
PDB
Protein data bank
PheWAS
Phenome-wide association study
RMSD
Root-mean-square deviation
RMSF
Root-mean-square fluctuation
SASA
Solvent-accessible surface area
SB
Spina bifida
SNP
Single nucleotide polymorphism
WGS
Whole-genome sequencing
iPSCs
Induced pluripotent stem cells
mQTL
Methylation quantitative trait locus
pQTL
Protein quantitative trait locus

Declarations

Funding

This work was supported by the National Natural Science Foundation of China (No. 82160592).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Clinical trial number

Not applicable.

Data availability

The datasets supporting the conclusions of this article are included within the article and its supplementary information files (**Supplementary Table S1**).

Authors' contributions

CS and GL contributed equally to this work. CS, FH, and JZ conceived and designed the study. CS and ZZ conducted the Mendelian randomization analyses and integrated multi-omics evidence. CS and GL performed data curation and transcriptomic analyses, including differential expression and functional enrichment analyses. XL carried out drug candidate prioritization, molecular docking, and molecular dynamics simulations. ZZ performed the mQTL-based mediation Mendelian randomization analyses. HD contributed to interpretation of results and critical revision of the manuscript. FH and JZ jointly supervised the study as co-corresponding authors. All authors read and approved the final manuscript.

Acknowledgments

We sincerely thank all the researchers and participants who provided publicly available GWAS data, as well as the platforms that made the data available and the developers of the software platforms.

References

1. Avagliano L, Massa V, George TM, Qureshy S, Bulfamante GP, Finnell RH. Overview on neural tube defects: From development to physical characteristics. Birth Defects Res. 2019;111(19):1455–67.
2. Okon II, Temitope AE, Ogundele IO, Akpan U, Mbong EO, Kasimieh O, et al. The current state of Spina Bifida in low- and middle-income countries: where does Africa stand? Neurochirurgie. 2025;71(1):101616.
3. Dean JH, Pauly R, Stevenson RE. Neural Tube Defects and Associated Anomalies before and after Folic Acid Fortification. J Pediatr. 2020;226:186–94. e4.
4. Wang M, Marco P, Capra V, Kibar Z. Update on the Role of the Non-Canonical Wnt/Planar Cell Polarity Pathway in Neural Tube Defects. Cells. 2019;8(10).
5. Guan Z, Liang Y, Zhu Z, Yang A, Li S, Wang X, et al. Genetic Effects of ITPK1 Polymorphisms on the Risk of Neural Tube Defects: a Population-Based Study. Reprod Sci. 2023;30(5):1585–93.

6. Wang B, Jin L, Ren A, Yuan Y, Liu J, Li Z, et al. Levels of polycyclic aromatic hydrocarbons in maternal serum and risk of neural tube defects in offspring. Environ Sci Technol. 2015;49(1):588–96.
7. Kim GJ, Seong JS, Oh JA. Prenatal screening for neural tube defects: from maternal serum alpha-fetoprotein to ultrasonography. Obstet Gynecol Sci. 2023;66(1):1–10.
8. Lee SY, Papanna R, Farmer D, Tsao K. Fetal Repair of Neural Tube Defects. Clin Perinatol. 2022;49(4):835–48.
9. Molendijk J, Parker BL. Proteome-wide Systems Genetics to Identify Functional Regulators of Complex Traits. Cell Syst. 2021;12(1):5–22.
10. Zhang Y, Du X, Yang Y, Ren Y, Zhou L, Hua J, et al. A multi-omics and mediation-based genetic screening approach identifies STX4 as a key link between epigenetic regulation, immune cells, and childhood asthma. Clin Epigenetics. 2025;17(1):101.
11. Suhre K. Genetic associations with ratios between protein levels detect new pQTLs and reveal protein-protein interactions. Cell Genom. 2024;4(3):100506.
12. An D, Wei X, Li H, Gu H, Huang T, Zhao G, et al. Identification of PCSK9 as a novel serum biomarker for the prenatal diagnosis of neural tube defects using iTRAQ quantitative proteomics. Sci Rep. 2015;5:17559.
13. Zhu X. Mendelian randomization and pleiotropy analysis. Quant Biol. 2021;9(2):122–32.
14. Hu T, Liu Q, Dai Q, Buchman AS, Bennett DA, Tasaki S, et al. Proteome-wide association studies using summary pQTL data of brain, CSF, and plasma identify 30 risk genes of Alzheimer's disease dementia. Alzheimers Res Ther. 2025;17(1):135.
15. Li XY, Li DT, Li YY, Xu ZC, Zhang Q, Chen X, et al. Mitochondrial proteins and congenital birth defect risk: a mendelian randomization study. BMC Pregnancy Childbirth. 2025;25(1):444.
16. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res. 2013;41(Database issue):D991–5.
17. Ferkingstad E, Sulem P, Atlason BA, Sveinbjornsson G, Magnusson MI, Styrmisdottir EL, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet. 2021;53(12):1712–21.
18. Sun BB, Chiou J, Traylor M, Benner C, Hsu YH, Richardson TG, et al. Plasma proteomic associations with genetics and health in the UK Biobank. Nature. 2023;622(7982):329–38.
19. Sun BB, Maranville JC, Peters JE, Stacey D, Staley JR, Blackshaw J, et al. Genomic atlas of the human plasma proteome. Nature. 2018;558(7708):73–9.
20. Shi K, Chen X, Zhao Y, Li P, Chai J, Qiu J, et al. Identification of potential therapeutic targets for nonischemic cardiomyopathy in European ancestry: an integrated multiomics analysis. Cardiovasc Diabetol. 2024;23(1):338.
21. Levin MG, Judy R, Gill D, Vujkovic M, Verma SS, Bradford Y, et al. Genetics of height and risk of atrial fibrillation: A Mendelian randomization study. PLoS Med. 2020;17(10):e1003288.

22. Gill D, Efstathiadou A, Cawood K, Tzoulaki I, Dehghan A. Education protects against coronary heart disease and stroke independently of cognitive function: evidence from Mendelian randomization. Int J Epidemiol. 2019;48(5):1468–77.
23. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D et al. The MR-Base platform supports systematic causal inference across the human phenome. Elife. 2018;7.
24. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304–14.
25. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985–98.
26. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.
27. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693–8.
28. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. Genet Epidemiol. 2013;37(7):658–65.
29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):e47.
30. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
31. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. Nat Commun. 2018;9(1):224.
32. Subramanian A, Kuehn H, Gould J, Tamayo P, Mesirov JP. GSEA-P: a desktop application for Gene Set Enrichment Analysis. Bioinformatics. 2007;23(23):3251–3.
33. Wang Q, Dhindsa RS, Carss K, Harper AR, Nag A, Tachmazidou I, et al. Rare variant contribution to human disease in 281,104 UK Biobank exomes. Nature. 2021;597(7877):527–32.
34. Xu W, Hu B, Cheng Y, Guo Y, Yao W, Qian H. Material basis research for Echinacea purpurea (L.) Moench against hepatocellular carcinoma in a mouse model through integration of metabonomics and molecular docking. Phytomedicine. 2022;98:153948.
35. Lill MA, Danielson ML. Computer-aided drug design platform using PyMOL. J Comput Aided Mol Des. 2011;25(1):13–9.
36. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: a web-based graphical user interface for CHARMM. J Comput Chem. 2008;29(11):1859–65.
37. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 2015;1–2:19–25.

38. Copp AJ, Adzick NS, Chitty LS, Fletcher JM, Holmbeck GN, Shaw GM. Spina bifida. Nat Rev Dis Primers. 2015;1:15007.
39. Dong N, Gu H, Liu D, Wei X, Ma W, Ma L, et al. Complement factors and alpha-fetoprotein as biomarkers for noninvasive prenatal diagnosis of neural tube defects. Ann N Y Acad Sci. 2020;1478(1):75–91.
40. Guilbaud L, Roger K, Schmidt A, Chhuon C, Breimann S, Lipecka J, et al. Molecular insights into myelomeningocele via proteomic analysis of amniotic fluid. J Proteom. 2025;313:105372.
41. Yan X, Mai L, Lin C, He W, Yin G, Yu J, et al. CSF-Based Analysis for Identification of Potential Serum Biomarkers of Neural Tube Defects. Neurosci Bull. 2017;33(4):436–44.
42. Wang Y, Ma L, Jia S, Liu D, Gu H, Wei X, et al. Serum exosomal coronin 1A and dynamin 2 as neural tube defect biomarkers. J Mol Med (Berl). 2022;100(9):1307–19.
43. Wang W, Ji Y, Dong Z, Liu Z, Chen S, Dai L, et al. Characterizing neuroinflammation and identifying prenatal diagnostic markers for neural tube defects through integrated multi-omics analysis. J Transl Med. 2024;22(1):257.
44. Camille H, Pierre G. Glycosylation in neurodevelopment: What oncology teaches? Neurobiol Dis. 2025;211:106945.
45. Altmüller F, Pothula S, Annamneedi A, Nakhaei-Rad S, Montenegro-Venegas C, Pina-Fernández E, et al. Aberrant neuronal activity-induced signaling and gene expression in a mouse model of RASopathy. PLoS Genet. 2017;13(3):e1006684.
46. Chen A, Yangzom T, Hong Y, Lundberg BC, Sullivan GJ, Tzoulis C, et al. Hallmark Molecular and Pathological Features of POLG Disease are Recapitulated in Cerebral Organoids. Adv Sci (Weinh). 2024;11(18):e2307136.
47. Huang T, Fakurazi S, Cheah PS, Ling KH. REST Targets JAK-STAT and HIF-1 Signaling Pathways in Human Down Syndrome Brain and Neural Cells. Int J Mol Sci. 2023;24(12).
48. Vanier MT. Niemann-Pick disease type C. Orphanet J Rare Dis. 2010;5:16.
49. Sapaly D, Cheguillaume F, Weill L, Clerc Z, Biondi O, Bendris S, et al. Dysregulation of muscle cholesterol transport in amyotrophic lateral sclerosis. Brain. 2025;148(3):788–802.
50. Chen J, Soni RK, Xu Y, Simoes S, Liang FX, DeFreitas L, et al. Juvenile CLN3 disease is a lysosomal cholesterol storage disorder: similarities with Niemann-Pick type C disease. EBioMedicine. 2023;92:104628.
51. Wu Q, Maniatis T. Large exons encoding multiple ectodomains are a characteristic feature of protocadherin genes. Proc Natl Acad Sci U S A. 2000;97(7):3124–9.
52. Hilschmann N, Barnikol HU, Barnikol-Watanabe S, Götz H, Kratzin H, Thinnes FP. The immunoglobulin-like genetic predetermination of the brain: the protocadherins, blueprint of the neuronal network. Naturwissenschaften. 2001;88(1):2–12.
53. Aran A, Rosenfeld N, Jaron R, Renbaum P, Zuckerman S, Fridman H, et al. Loss of function of PCDH12 underlies recessive microcephaly mimicking intrauterine infection. Neurology.

2016;86(21):2016–24.
54. Nicolas G, Sanchez-Contreras M, Ramos EM, Lemos RR, Ferreira J, Moura D, et al. Brain calcifications and PCDH12 variants. Neurol Genet. 2017;3(4):e166.
55. Guemez-Gamboa A, Çağlayan AO, Stanley V, Gregor A, Zaki MS, Saleem SN, et al. Loss of Protocadherin-12 Leads to Diencephalic-Mesencephalic Junction Dysplasia Syndrome. Ann Neurol. 2018;84(5):638–47.
56. Rakotomamonjy J, Rylaarsdam L, Fares-Taie L, McDermott S, Davies D, Yang G, et al. PCDH12 loss results in premature neuronal differentiation and impeded migration in a cortical organoid model. Cell Rep. 2023;42(8):112845.
57. Kozlova I, Sytnyk V. Cell Adhesion Molecules as Modulators of the Epidermal Growth Factor Receptor. Cells. 2024;13(22).
58. Deak KL, Boyles AL, Etchevers HC, Melvin EC, Siegel DG, Graham FL, et al. SNPs in the neural cell adhesion molecule 1 gene (NCAM1) may be associated with human neural tube defects. Hum Genet. 2005;117(2–3):133–42.
59. Nagaoka T, Katsuno T, Fujimura K, Tsuchida K, Kishi M. Functional interaction between Vangl2 and N-cadherin regulates planar cell polarization of the developing neural tube and cochlear sensory epithelium. Sci Rep. 2023;13(1):3905.
60. Badouel C, Zander MA, Liscio N, Bagherie-Lachidan M, Sopko R, Coyaud E, et al. Fat1 interacts with Fat4 to regulate neural tube closure, neural progenitor proliferation and apical constriction during mouse brain development. Development. 2015;142(16):2781–91.
61. Krakauer T. IL-10 inhibits the adhesion of leukocytic cells to IL-1-activated human endothelial cells. Immunol Lett. 1995;45(1–2):61–5.
62. Liu M, Fang L, Tan C, Long T, Chen H, Xiao S. Understanding Streptococcus suis serotype 2 infection in pigs through a transcriptional approach. BMC Genomics. 2011;12:253.
63. Huang Q, Yang L, Nong B, Gan H, Wu H, Li M, et al. Identification of immune-related biomarkers in embryos with neural tube defects via a bioinformatics analysis. Ann Transl Med. 2022;10(9):521.
64. Kim SE, Lei Y, Hwang SH, Wlodarczyk BJ, Mukhopadhyay S, Shaw GM, et al. Dominant negative GPR161 rare variants are risk factors of human spina bifida. Hum Mol Genet. 2019;28(2):200–8.
65. An D, Wei XW, Zhang HN, Liu D, Ma W, Yuan ZW. Spatiotemporal expression of leukemia inhibitory factor receptor protein during neural tube development in embryos with neural tube defects. Neural Regen Res. 2020;15(4):705–11.
66. 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.
67. Runge EM, Iyer AK, Setter DO, Kennedy FM, Sanders VM, Jones KJ. CD4 + T cell expression of the IL-10 receptor is necessary for facial motoneuron survival after axotomy. J Neuroinflammation. 2020;17(1):121.

68. Liu C, Morishima M, Jiang X, Yu T, Meng K, Ray D, et al. Engineered chromosome-based genetic mapping establishes a 3.7 Mb critical genomic region for Down syndrome-associated heart defects in mice. Hum Genet. 2014;133(6):743–53.
69. Balistreri CR, Ammoscato CL, Scola L, Fragapane T, Giarratana RM, Lio D et al. Susceptibility to Heart Defects in Down Syndrome Is Associated with Single Nucleotide Polymorphisms in HAS 21 Interferon Receptor Cluster and VEGFA Genes. Genes (Basel). 2020;11(12).
70. McCauliff LA, Langan A, Li R, Ilnytska O, Bose D, Waghalter M, et al. Intracellular cholesterol trafficking is dependent upon NPC2 interaction with lysobisphosphatidic acid. eLife. 2019;8:e50832.

Figures

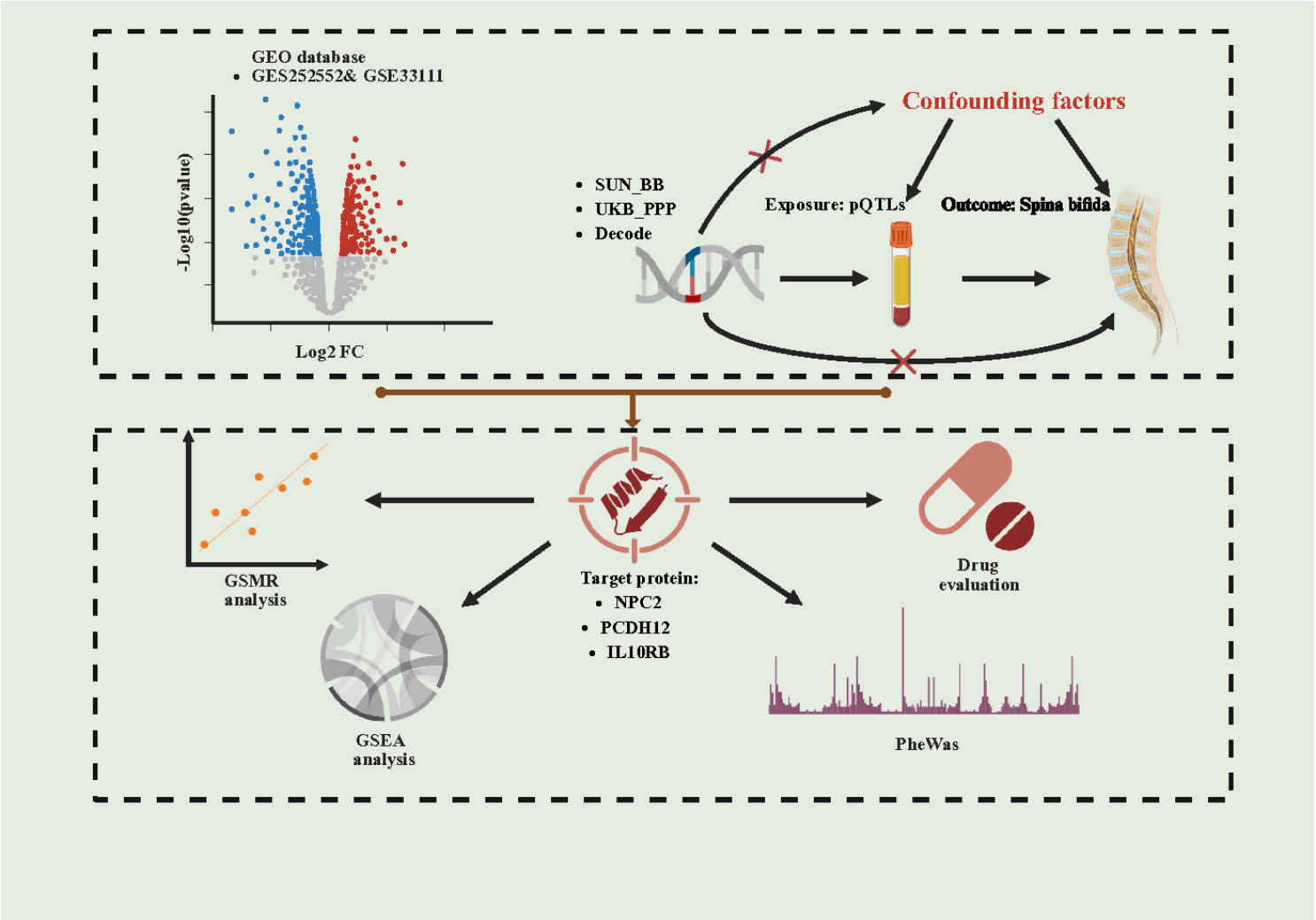


Figure 1

Legend not included with this version.

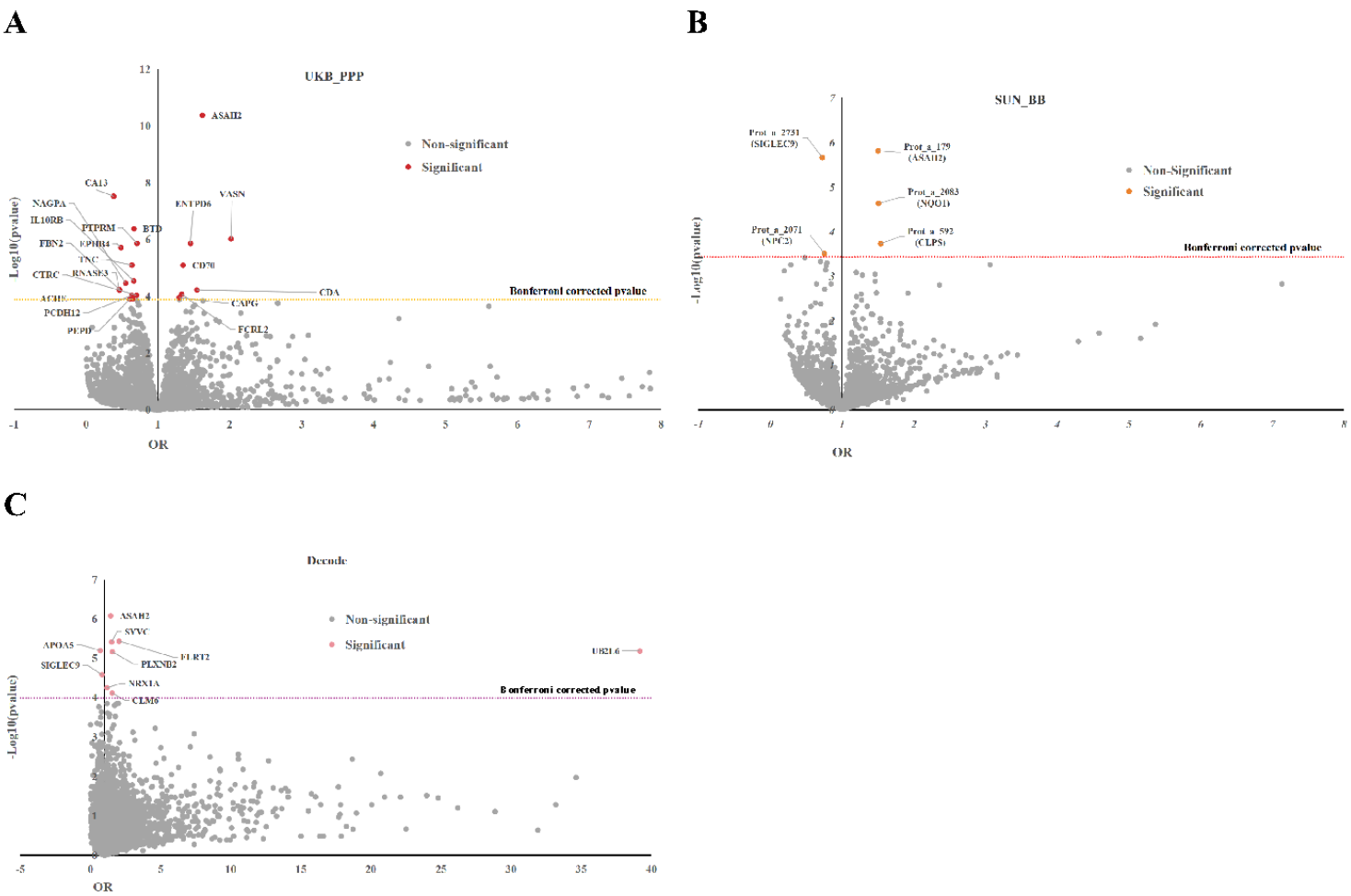


Figure 2

Legend not included with this version.

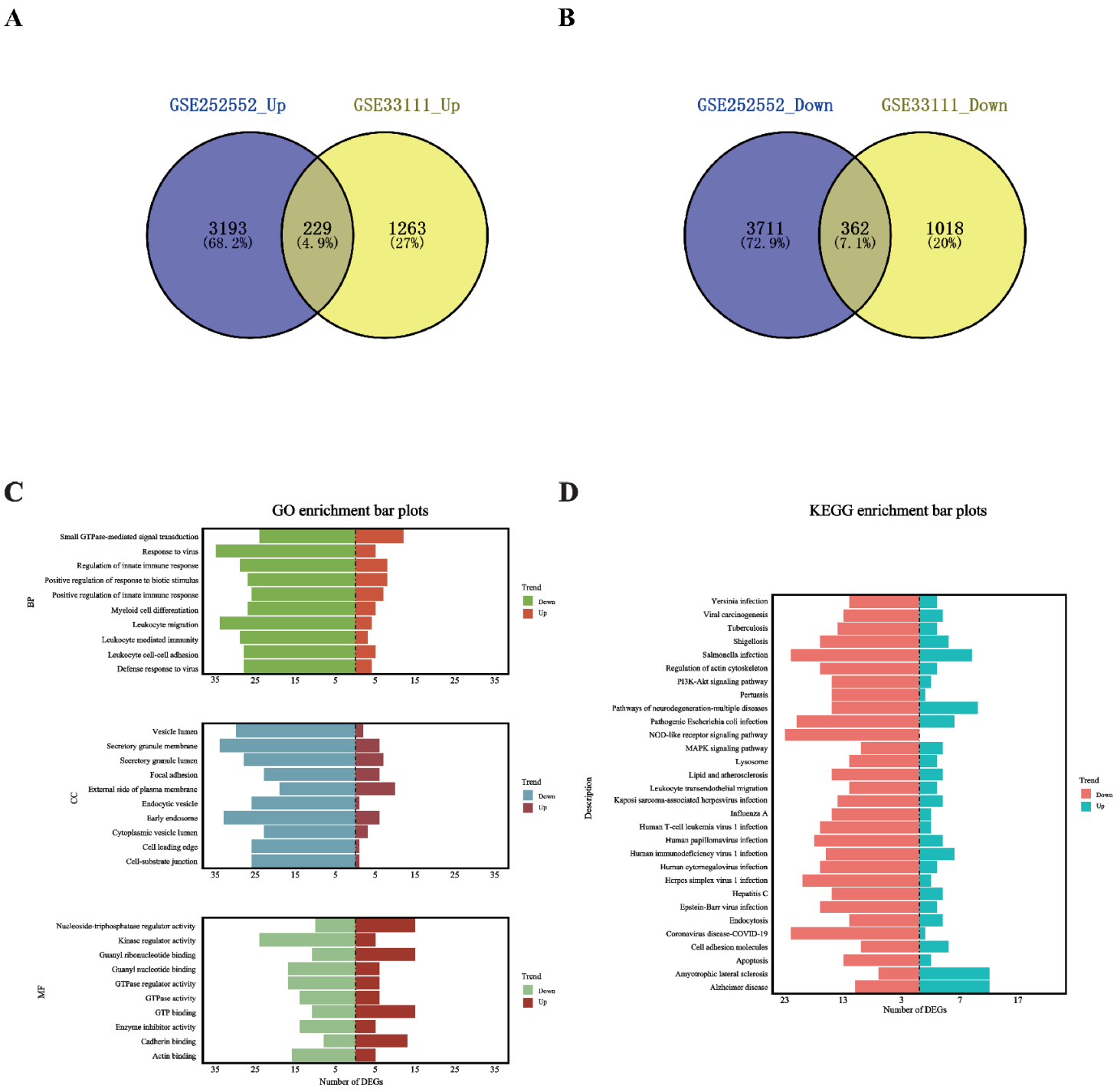


Figure 3

Legend not included with this version.

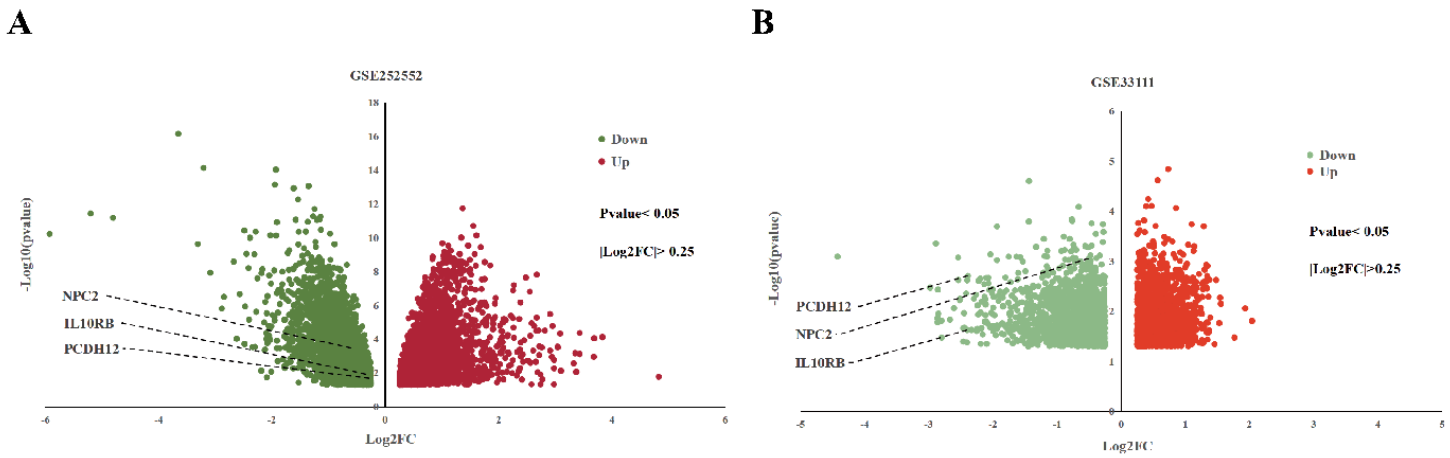


Figure 4

Legend not included with this version.

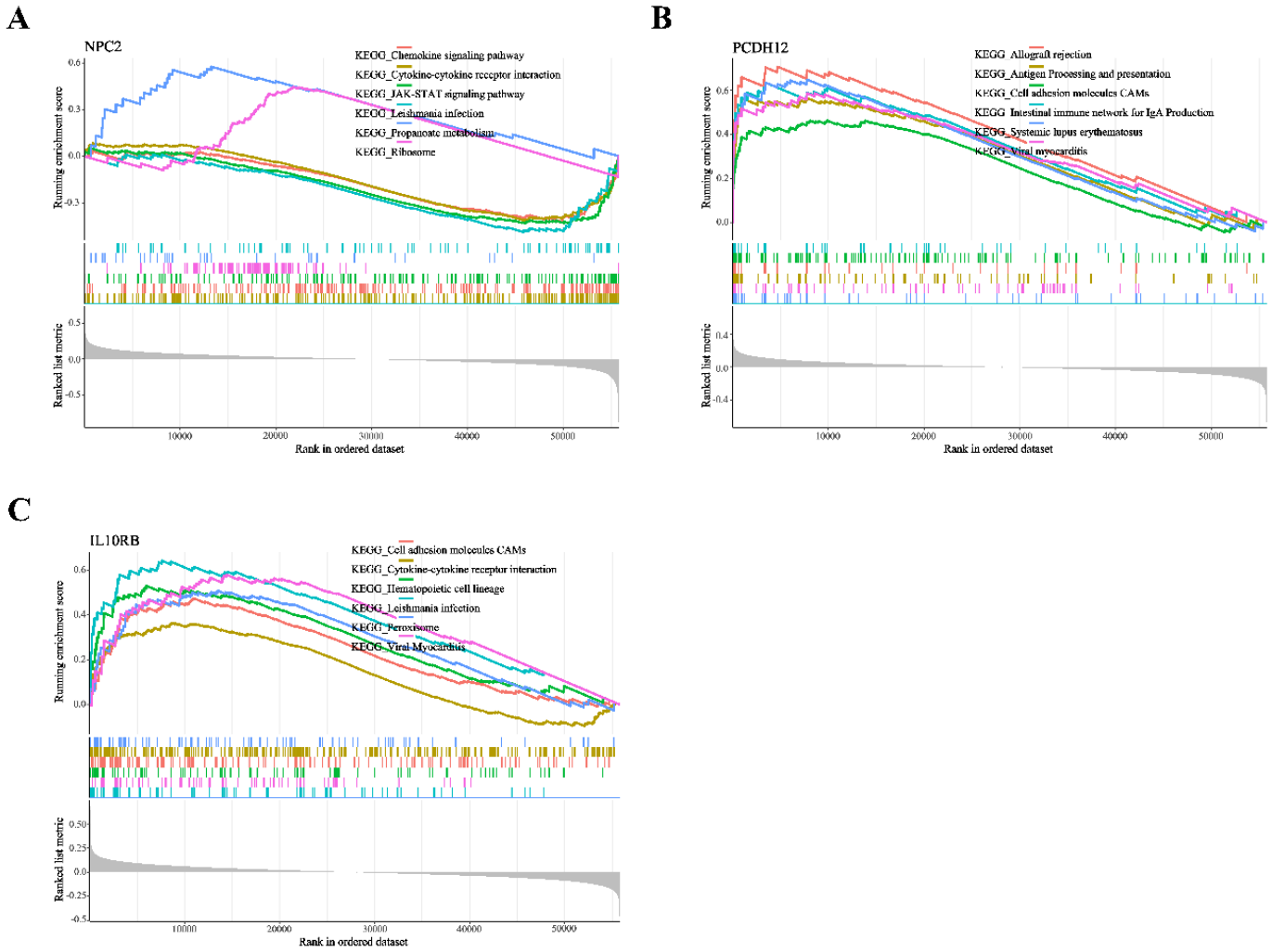


Figure 5

Legend not included with this version.

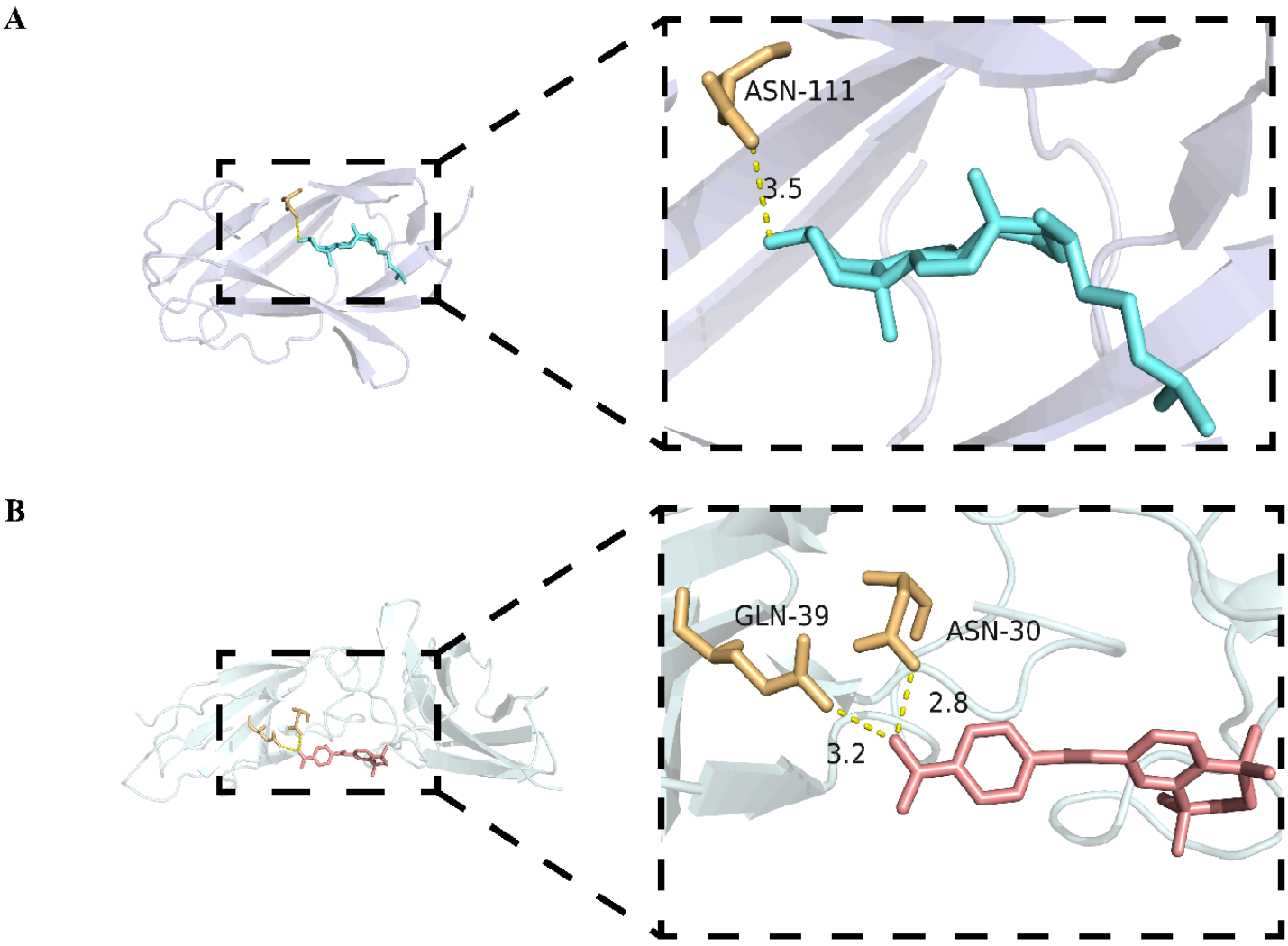


Figure 6

Legend not included with this version.

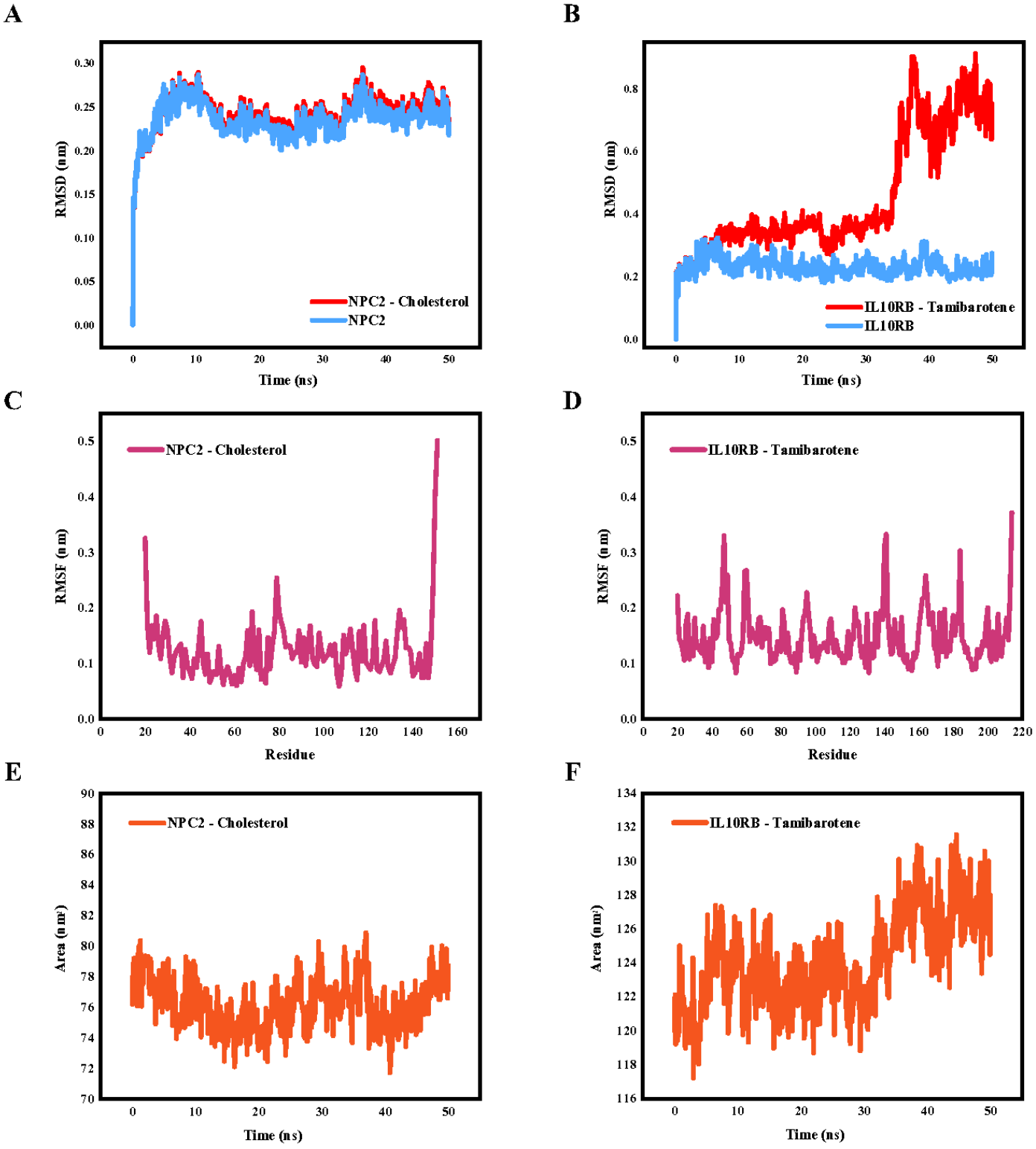


Figure 7

Legend not included with this version.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- STROBEMR.docx
- Supplementaryfig1.pdf
- Supplementaryfig2.pdf
- Supplementaryfigurelegends.docx
- Supplementarytables.xlsx