이학석사 학위논문

Identification of potential biomarkers of autism spectrum disorder via integrated analysis of animal model and patients' sera

동물 모델 및 환자 혈청의 통합분석을 통한 자폐 스펙트럼 장애 바이오마커 발굴 연구

2023 년 02 월

서울대학교 대학원

분자의학 및 바이오제약학과

김 지 연

Identification of potential biomarkers of autism spectrum disorder via integrated analysis of animal model and patients' sera

지도교수 Eugene C. Yi

이 논문을 이학석사 학위논문으로 제출함

2023 년 2 월

서울대학교 대학원

분자의학 및 바이오제약학과

김 지 연

김지연의 석사 학위논문을 인준함.

2023 년 2 월

위 원 장 _____ (인)

부위원장 _____ (인)

위 원 _____ (인)

Abstract

Identification of potential biomarkers of autism spectrum disorder via integrated analysis of animal model and patients' sera

Jiyeon Kim

Department of Molecular Medicine and Biopharmaceutical Sciences

Graduate School of Convergence Science and Technology

Seoul National University

Autism spectrum disorder (ASD) is one of the most common neurodevelopmental disorders (NDD) characterized by three core symptoms: (1) impaired social communication, (2) reciprocal interaction, and (3) the presence of repetitive behaviors and restricted interest. Although various studies have been conducted to discover pathophysiological mechanisms of ASD, the cause of disease is poorly understood, hence no reliable diagnostic biomarkers are available. Since the prevalence of ASD is rapidly increasing, there is an urgent need to identify ASD-related molecular biomarkers.

In this study, we conducted brain tissue and serum proteomic analysis of the Contactin-associated protein-like 2 (*Cntnap2*) KO mouse, a well-recognized genetic mouse model of ASD. We used a Data Dependent Acquisition (DDA) mass spectrometry to identify differentially expressed proteins in brain tissue samples followed by the analysis of serum samples to identify potential ASD biomarkers. Potential biomarkers were further verified by Data Independent

Acquisition (DIA) mass spectrometry with ASD patients' sera and identified three proteins TNC, TLN1 and SERPINC1 in patients' sera.

Keywords : Autism spectrum disorder (ASD), Contactic-associated protein-like 2 (*Cntnap2*), Proteomics, Biomarker, LC-MS/MS analysis

Student Number : 2020-26559

Abbreviation

The abbreviations are used:

ACN Acetonitrile

ADOS Autism diagnostic Observation Schedule

ASD Autism spectrum disorder

CARS Childhood Autism Rating Scale

CNTNAP2 Contactin-associated protein-like 2

CPTAC Clinical proteomic tumor analysis consortium

DDA Data-Dependent Acquisition

DEP Differentially expressed protein

DIAa Data-Independent Acquisition

DTT Dithiothreitol

FA Formic acid

FDR False discovery rate

HCD High-energy collision dissociation

IAA Iodoacetamide

LFQ Label free quantitation

MAI Moment adjusted imputation

NDD Neurodevelopmental disorder

STN Signal to noise

WT Wild type

Contents

List of tables and figures

1. Introduction

Autism spectrum disorder (ASD) is one of the most common neurodevelopmental disorders characterized by three core symptoms: (1) impaired social communication, (2) reciprocal interaction, and (3) the presence of repetitive behaviors and restricted interest (DSM-V) [1]. The Centers for Disease Control and Prevention (CDC) had reported that approximately 1 in 68 children have been diagnosed with ASD in the United States, and associated social cost is also increasing [2]. Because of the high heritability in autism, finding the underlying genetic causes has been a major focus of research in autism. Recently, studies that environmental factors, such as prenatal drug exposure, are related to autism are increasing, however the exact causes are still unknown [3].

Although the Food and Drug Administration (FDA) has approved two drugs (risperidone and aripiprazole) to treat hypersensitivity associated with autism, the drug which treats three key characteristics of autism has not yet been approved. Besides, current treatment options for the core symptoms of autism are limited to psychosocial treatments such as behavioral therapy [7]. Existing ASD diagnostic methods according to DSM-V criteria depend on the proficiency of the specialist, and biased views from parental or healthcare personnel can interfere with the accurate diagnosis. In other words, absence of practical biological molecular markers makes acute diagnosis difficult. For these reasons, demand for a blood-based biomarker which can reflect molecular alterations at both physiological and pathological levels is increasing [5].

Efforts to discover molecular biomarkers for ASD have focused on various biomolecules such as genes, mRNA transcripts, proteins. Among these, since proteins are the functional entities of biological processes, protein levels show a strong correlation with the phenotype of disease and could be a better indication of disease [6].

In this study, we used a mouse model, *Cntnap2* KO mouse, to identify potential biomarker of ASD. Among the identified ASD risk genes, such as *SHANK3, FMR1, and PTEN*, contactin-associated protein-like 2 (*CNTNAP2*) gene is one of the major susceptibility genes for NDDs, including ASD as well as schizophrenia and Gilles de la Tourette syndrome [7-9]. Neuropathological and physiological analyzes of *Cntnap2* KO mice revealed abnormal neuronal migration, reduced numbers of intervened neurons, and abnormal neuronal network activities [10]. Also, it mimics core ASD features, such as impaired social behavior and reduced vocalizations [11-12], learning and memory deficits [13], and epileptiform activity [12]. For these reasons, *Cntnap2* KO mouse is a one of widespread organisms for studying ASD and brain dysfunction.

Here, to identify molecular alteration and identify potential serum biomarker proteins for ASD, we performed reliable comparative proteomic analyses using brain tissues and sera from *Cntnap2* KO and wild-type mice. Mass spectrometry-based proteomics provides a powerful tool to study adaptation across multiple tissues and biological fluids in pathophysiological conditions, such as ASD [14]. We performed DDA (Data dependent acquisition)-MS for global profiling in brain tissues and sera in mouse. After that, we confirmed

whether the brain proteins are secreted into serum through the circulation systems. Biomarker candidates which translated from mouse data to human data verified by Data-Independent Acquisition (DIA) mass spectrometry especially in individual ASD patients' sera. As a result, we verified three proteins, TLN1, TNC and SERPINC1, which could serve as resources for future studies related to ASD biomarkers. (Figure 1)

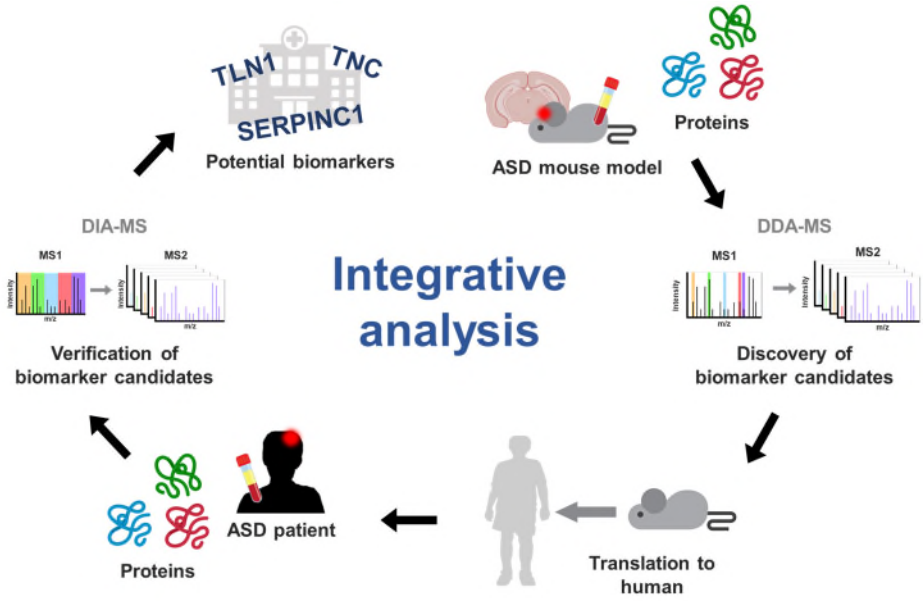

Figure 1. Overall scheme workflows of potential ASD biomarker identification

In this study, brain tissues and sera from Cntnap2 KO and wild type mice were used for global profiling, DDA-MS in discovery step. After discovery of potential biomarker candidates, we translated this biomarker candidate mouse data into human data. To verify them in individual ASD patients' serum, we used these two adult group and patients with ASD, we performed DIA-MS for targeted proteins.

2. Method

2.1 Materials

HPLC grade acetonitrile (ACN), formic acid (FA), HPLC grade water, and Tris-HCl (pH 8.0) were purchased from Thermo Scientific (Rockford, IL, USA). Urea dithiothreitol (DTT), iodoacetamide (IAA) and solid phase extraction C18 disks were purchased from Sigma-Aldrich (St. Luis, Missouri, USA). Trypsin/Lys-C Mix was obtained from Promega (Madison, Wi, USA). iRT kit are from Biognosys (Wagistrasse, Schlieren,Switzerland)

2.2 Sample collection

2.2.1 Transgenic mice

Cntnap2 KO and WT mice which used in this experiment were generated, bred from Samyuk university, as described previously. [15,16]

2.2.2 Study population

ASD patients were recruited from Seoul national university Children's hospital. Serum samples were obtained from ASD patients who exceeded CARS (Childhood Autism Rating Scale) and ADOS (Autism diagnostic Observation Schedule) scores. The population consisted of participants with ASD aged 4 years or above. And comparison group were consisted of adult group. The adult group population comprised participants with adults age ranges from 24 to 26 years.

2.3 Brain tissue lysis

Proteins from samples were extracted and were further dissolved with a lysis buffer, approximately 500 μL of RIPA and protease inhibitor, and phosphatase inhibitor. They were vortexed vigorously first, then repeatedly every 5 minutes during the incubation on ice for 15 minutes. Probe sonication (Sonics, vibra-cell™) was performed for each sample for 1 min with pulse 10 and amplitude 1, 25%. Thoroughly homogenized samples were centrifuged at 14,000 rpm for 20 minutes at 4°C to pellet down the cell debris. Then, Protein concentration was measured using a Micro BCA Protein Assay Kit (Thermo Fisher Scientific Rockford, IL, USA).

2.4 Depletion of high abundance proteins

2.4.1 Mouse serum

To remove the 3 high abundant proteins (albumin, IgG and transferrin) of serum were immunodepleted using the Multiple Affinity Removal Spin Cartridge (MARS Mouse-3, Agilent). The serum samples briefly sonicated for 30 sec. Then all of them were centrifuged at 3,000 xg for 5 min to remove a debris. The serum was transferred to a new tube and diluted with Agilent buffer A (Agilent Technologies, Wilmington, DE, USA) to final volume 200 μL. Diluted serums were filtered through 0.22 μm spin filters (Agilent Technologies, San Diego, CA, USA) and depleted following the spin cartridge manual. After cartridge equilibration with the buffer A, 200 μL of diluted serum was loaded on the cartridge. The bound proteins were eluted with buffer B (Agilent technologies, Wilmington, DE, USA). The cartridge was then washed and reequilibrated with buffer A. All of depleted fraction of serum samples were collected and proteins were concentrated with 3K Ultra Centrifugal filter (Millipore, Billerica, MA, USA) At the same time, the buffer was exchanged to 50mM Tris-HCL (pH 8.0). After that the protein concentration of samples was measured with micro BCA protein Assay Kit (Thermo Scientific, Rockford, IL USA)

2.4.2 Human serum

Each human sample were aliquoted to 30 μL. After 30 second brief sonication, all of serum samples were centrifuges at 3,000 xg for 5 min to remove debris in serum. And each samples injected into a High Select™ Depletion Spin Columns (Thermo Scientific, Rockford, IL, USA) for removing the 14 most

abundant proteins (Albumin, IgA, IgD, IgE, IgG, IgG (light chain), IgM, alpha-1-acid glycoprotein, alpha-1-antitryosin, alpha-2-macroglobulin, fibrinogen, apolipoprotein A1, haptoglobin and transferrin). Before depletion, equilibrate the depletion spin column to room temperature and add up to 10 μL of sample to the resin slurry in the column. After several inversions to make the solution complete homogenous, incubate the mixture 10 minutes at room temperature. After centrifugation at 1,000 xg for two minutes and filtrated mixture which depleted high abundant proteins were concentrated with Amicon Ultra-0.5mL Centrifugal Filter 3K device (MerckMillipore, Burlington, MA, USA) and exchange the buffer at 50mM Tris/HCL (pH=8.0). Each samples were approximately 150μl of total and the concentration was measured using a Micro BCA proteins Assay Kit (Thermo Scientific, Rockford, IL, USA). (Figure 2)

2.5 In-gel tryptic digestion of proteins

Proteins were denatured and reduced with LDS sample buffer by boiling them at 95℃ for 10 minutes before loading onto an SDS-PAGE gel. The protein samples were then fractionated on 4-12% gradient Bolt Bis-Tris gel (Invitrogen, MA, USA) by performing gel electrophoresis at 80V for 2 hours. The gel was stained with Instant Blue Coomassie Protein stain (Abcam, Waltham, Boston, USA) and destained with water. The in-gel tryptic digestion was conducted with the general protocol. Each gel lane was cut into 10 slices using a clean razor blade and diced into $1mm^3$ pieces on a glass plate. The excised gel pieces were washed and de-stained 2 times for 20 minutes with a sufficient amount of 50% (v/v) Acetonitrile (ACN) (Thermo Fisher Scientific, Waltham, Massachusetts, USA) in 25mM ammonium bicarbonate (NH_4HCO_3) (Sigma-Aldrich,

Burlington, MA, USA). In each step, the supernatant was discarded. The diced gels were shrunk by 100% ACN and dried under a speed-vac concentrator. The proteins in the gels were reduced by 25mM DTT in 100 mM NH_4HCO_3 at 60'C for 1 hour and alkylated with freshly prepared 55mM iodoacetamide (IAA) (Sigma Aldrich, Burlington, MA, USA) in NH_4HCO_3 in the dark for 45 min. To eliminate the remaining trypsin, the gels were thoroughly washed two times, alternating between 50% ACN in 25mM Ambic and 100mM Ambic. The particles were shrunk by 100% ACN and completely dried using the speed-vac concentrator. The shrunken gel pieces were saturated with 12.5 ng/μl trypsin (Promega, Madison, WI, USA) in 50mM of NH_4HCO_3 on ice for 45 min. The trypsin solutions were removed, and sufficient 50mM NH_4HCO_3 was added to cover the gel pieces. The proteins in the gel pieces were digested at 37℃ over-night, 16 hours. The digested peptide solutions were collected into a new clean tubes. The peptides were extracted after incubation with 10% formic acid (FA) for 15min on ice, and another extra extraction step was performed after incubation with 50% (v/v) ACN/ 0.1% (v/v) FA and 80% (v/v) ACN/ 0.1% (v/v) FA serially. The extracted peptides were dried under a concentrator and stored at -20℃ until the LC-MS/MS analysis. After drying out, digested peptides were desalted using manually packed C18 tips in accordance of modified CPTAC (Clinical Proteomic Tumor Analysis Consortium) protocol.

2.6 LC-MS/MS analysis

2.6.1 Data dependent acquisition method

Peptides were resuspended in 15μl Solvent A (0.1% FA in water) and 4 μL samples were loaded into an analytic column (PepMap, 75 μL ID*50cm 3μm, ES803, Thermo Fisher Scientific) and separated with a linear gradient of 5-32% Solvent B (0.1% FA in 98% ACN), for 125 min at a flow rate 300nL/min. Samples were analyzed in triplicate on a Q-Exactive (Thermo Fisher Scientific, San Jose, CA) hybrid quadrupole-Orbitrap mass spectrometry, interfaced with a HPLC system. The temperature of the heated capillary was set to 250℃, and the voltage of the spray was set to 2.2kV. The full scans were acquired in the mass analyzer at 400-1400 m/z with a resolution of 70,000 and the MS/MS scans were obtained with a resolution of 17,500 by normalized collision energy of 27eV for high-energy collisional dissociation fragmentation. The advanced gain control target was 5 x 104, maximum injection time was 120 ms, and the isolation window was set to 3 m/z. The Q-Exactive was operated in data-dependent mode with one survey MS scan followed by 10 MS/MS scans and the duration time of dynamic exclusion was 30s.

2.6.2 Data independent acquisition method

20 depleted serum samples were resuspended in 40µl of 0.1% formic acid and the same volume of peptides were transferred to a new tube by relative quantitation using Nanodrop 1000 Spectrophotometer (Thermo Scientific, Rockford, IL, USA).

After adjusting the volume to a total 8 µL using 0.1% formic acid, 2 µL of iRT peptide was added to make 10 µL of samples to injection. Of that, 4 µL of sample flows through trap column (PepMapTM RSLC, C18 column 75 µm ID*2 cm, 2 µm, Thermo Fisher scientific) for sample clean up and remove contaminations and then loaded onto an analytical column (PepMapTM RSLC, C18 column 75 µm ID*50 cm, 2 µm, Thermo Fisher scientific). The peptides were separated with linear gradient 5-35% solvent B for 165 min at a flow rate of 300nL/min; 0-5 min 2-5% of solvent B, 5-138 min 5-35% of solvent B, 138-140 min with 35-90% of solvent B, 140-148 min with 90% of solvent B, 148-150 min 90-2% of solvent B and 150-165 min 2% of solvent B for column re-equilibration. Standard mass spectrometric condition of the spray voltage was set to 1.8kV and the temperature of the heated capillary was set to 250°C. The full scans were acquired in range at 300-1650 m/z with 70,000 resolutions and the normalized collision energy was 32% and 17,500 resolutions for high-energy collisional dissociation (HCD) fragmentation. The data-independent acquisition (DIA) was operated with single survey MS scan according to a fixed isolation window followed by fragmentation of all peptides within MS scan.

2.7 Database search for protein identification and spectral library generation

Collected MS/MS raw data was converted into mzXML files through engine-based Peaks Studio. The processed MS/MS peaks lists were compared with Uniprot Mus Musculus reviewed database containing 17,131 entries. Precursor and fragment ion tolerance were set to 10 ppm and 0.8 Da, respectively. The enzyme was selected as trypsin with a maximum allowance of up to two missed cleavages. For post-translational modification, carbamidomethylation (+57.0215 Da) for cysteine as fixed modification and oxidation (+15.9949 Da) for methionine as dynamic modification. Peptide and protein validator was employed for peptide false discovery rate (FDR) which is between relaxed target FDR 5% and strict target FDR 1%.

Differentially expressed proteins (DEPs) were obtained by statistical processing, PLGEM-MAI. The normalized values along triplicate analyses of two different samples were analyzed using Power Law Global Error Model (PLGEM) (http://www.bioconductor.org) package within R program (version 4.2.0). PLGEM can control datasets to distinguish statistically significant DEPs and calculate expression level change by p-value and signal-to-noise (STN). We filtered statistically significant proteins using 0.01 as a p-value threshold. Then we refined spectral counts of DEPs within the range of $0.01 \leq$ p-value ≤ 0.05, which were excluded from first criteria p-value <0.01, using moment adjusted imputation (MAI) equation and finalized the list of DEPs with statistically significant change in their abundance.

To generate a spectral library for subsequent DIA database search, MS/MS raw data of pooled serum were loaded to Spectronaut™ version 14.10 (Biognosys, Wagistrasse, Schlieren, Switzerland). The generated serum database and the human serum standard spectral library provided by Spectronaut™ were combined to create a spectral library to be used.

2.8 Label free quantitation and statistical analysis

Relative quantitation analysis of identified protein identification and label-free quantitation were performed through Spectronaut™ (v 14.10), a DIA-based proteomics software. Using standard setting, interference correction on MS1 and MS2 level was enabled and FDR was set to 1% at peptide precursor and protein level using scrambled decoy generation and dynamic size at 0.1 fraction of library size. MS2 area used for quantitation enabling global normalization on median value. Further statistical analysis was performed by Perseus software (v 1.6.15, Max planck institute of biochemistry). Fist, Spectronaut™ search data of ASD and normal samples were input and all missing values were replaced with normal distribution. Statistical significance of changes in abundance between ASD and normal groups was calculated by paired two-sample t-tests ($p<0.05$, fold changed>1.2).

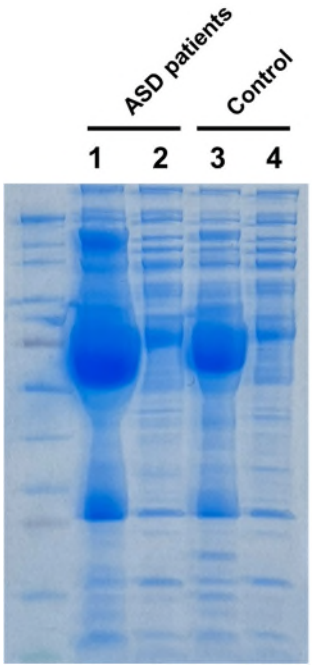

Figure 2. SDS-PAGE of the serum samples before and after depletion

The proteins were separated on Bolt 4-12% gradient Bis-Tris gel and stained using Coomassie brilliant blue. We could observe that the serum containing high abundant proteins (lane 1,3) were clearly depleted (lane 2,4)

3. Results

3.1 Identification and statistical analysis of DEPs in *Cntnap2* KO mice brain and serum

To assess the alteration of protein expression induced by *Cntnap2* KO, we performed a quantitative proteomic analysis with *Cntnap2* KO (n=5) and WT (n=5) hippocampus tissues and sera. As a result, we identified 6972 proteins and 1706, and they were analyzed with statistical analysis method to select statistically confident and significant DEPs. In detail, the PLGEM-STN analysis was conducted on identified proteins for p-value calculation, the spectral count value of each protein group acquired from mass spectrometry was refined by MAI normalization method, then the p-value of the refined spectral count was calculated with PLGEM-STN method once again. Finally, proteins with p-value <0.03, 1584 and 583 were selected as DEPs. Among these DEPs, 476 and 400 proteins were upregulated and 1108 and 182 were downregulated in brain and serum respectively. (Figure 3)

We performed functional enrichment analysis to look further into the functions of upregulated proteins in both brain tissues and sera in mouse models. Most proteins were engaged in carbon metabolism and transferase activity carbohydrate derivative binding, metabolism of proteins immune system and catalytic activity. (Figure 4)

3.2 The confirmation of ASD biomarker candidates in *Cntnap2* KO mice

To verify the DEPs generated from the *Cntnap2* KO mice proteomic analysis in the serum and brain samples, we had to select several upregulated and downregulated DEPs for further DIA verification.

Since the protein is secreted by the brain and released into the blood, its detection in both brain and serum may accurately represent a state of brain dysfunction. Verification of proteins commonly expressed in brain and serum will aid in the discovery of ASD biomarkers. Among the DEPs in the serum and brain samples of *Cntnap2* KO mice and the adult group, we found 144 common proteins.

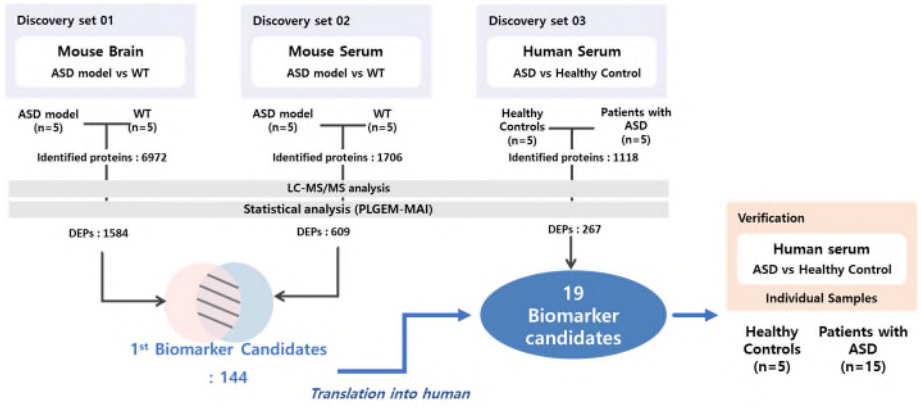

Figure 3. Analytic results of proteomic data

In discovery set, 6972, 1706 and 1118 proteins were identified in mouse brain, serum, and human serum, respectively. After statistical analysis and the steps for selection of proteins targets, final potential biomarker candidates for verification phase were 19.

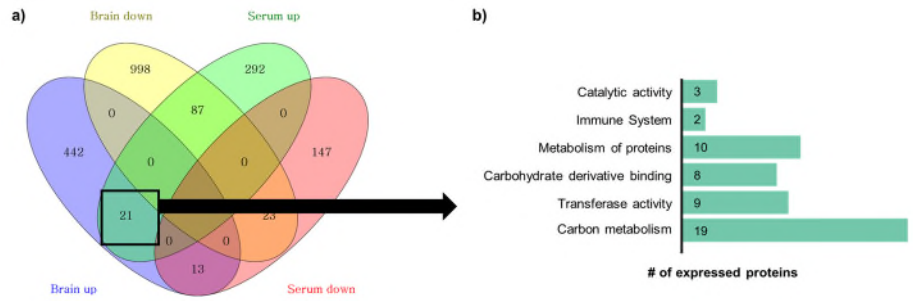

Figure 4. Functional enrichment of differentially expressed proteins in mouse model

a) Venn diagram which plotted Differentially expressed proteins found in brain and serum from Cntnap2 KO and WT

b) Functional enrichment analysis with upregulated in both brain and serum of Cntnap2 KO

3.3 Potential biomarker candidates translated into human

Before translating DEPs from mouse data to human data, we just want to compare and see is there any patterns between mouse and human proteome. We performed comparative analysis of canonical pathways. As you can see/ there are similar tendencies in both mouse models and human patients except the LXR/RXR activation. (Figure 5) Accordingly, it was found that there is no difficulty in translating from the mouse model data to human data. After translation, 19 biomarker candidates were selected. (Table 1)

These potential ASD biomarker candidates applied to STRING protein interaction analysis to understand further networks and biological process. It revealed that they were classified as 4 biological process or KEGG pathways including cell adhesion, establishment of localization, immune system, and cellular response to interleukin 4. (Figure 6)

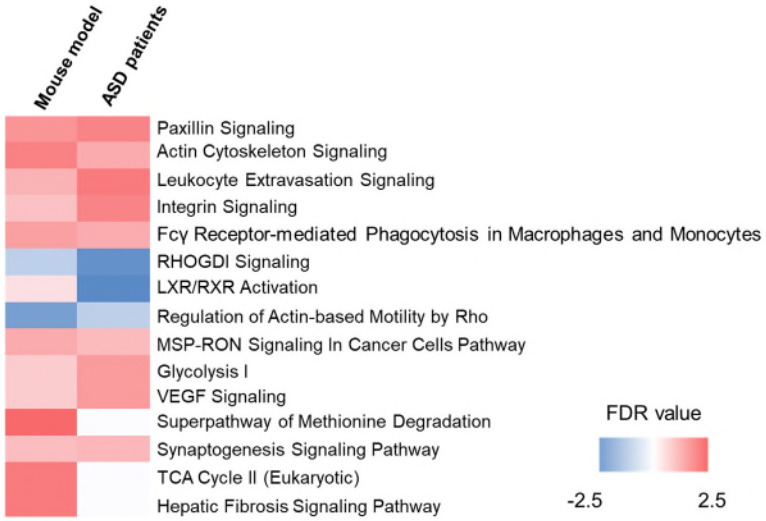

Figure 5. Comparison analysis of discovery in ASD mouse model and patients

The common canonical pathways between mouse and human proteomic findings based on IPA.

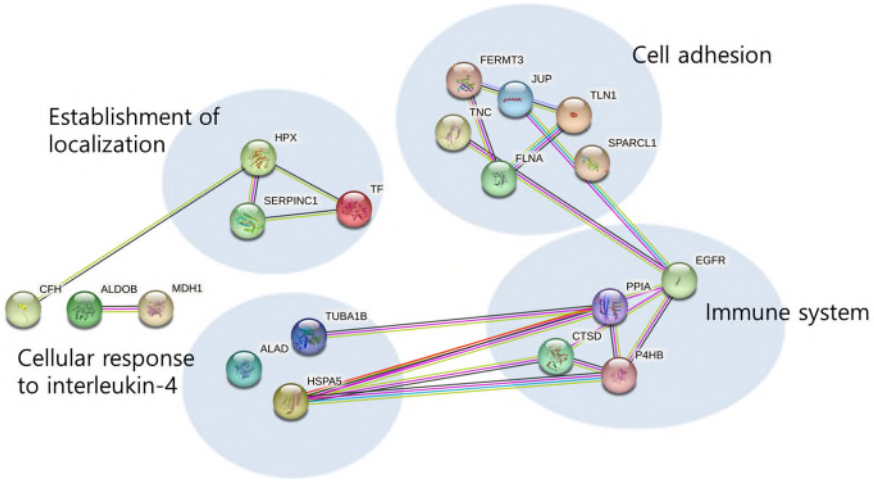

Figure 6. Functional classification of potential ASD biomarkers

19 Protein targets were visualized separately by the same biological process or KEGG pathway.

Table 1. The list of potential serum biomarkers in human

	Accession	Description	GN
1	P02790	Hemopexin	HPX
2	P01008	Antithrombin-III	SERPINC1
3	P05062	Fructose-bisphosphate aldolase B	ALDOB
4	P07237	Protein disulfide-isomerase	P4HB
5	P08603	Complement factor H	CFH
6	P13716	Delta-aminolevulinic acid dehydratase	ALAD
7	P40925	Malate dehydrogenase cytoplasmic	MDH1
8	P00533	Epidermal growth factor receptor	EGFR
9	Q14515	SPARC-like protein 1	SPARCL1
10	P11021	Endoplasmic reticulum chaperone BiP	HSPA5
11	P62937	Peptidyl-prolyl cis-trans isomerase A	PPIA
12	P68363	Tubulin alpha-1B chain	TUBA1B
13	Q86UX7	Fermitin family homolog 3	FERMT3
14	P07339	Cathepsin D	CTSD
15	P21333	Filamin-A	FLNA
16	P24821	Tenascin	TNC
17	P14923	Junction plakoglobin	JUP
18	Q9Y490	Talin-1	TLN1
19	P02787	Serotransferrin	TF

3.4 Verification of the biomarker candidate proteins using serum from human patients with Autism spectrum disease

We attempted to verify the potential utility of 19 selected ASD biomarker candidates for human specimens. To this end, we collected 20 blood specimens from normal and ASD patients classified CARS score. The biomarker candidates were subjected to DIA analysis and found that the presence of P4HB, HPX, HSPA5, TNC, TLN1, SERPINC1, CTSD, CFH, SPARCL1 and TF in human sera. Among them TNC and TLN1 were upregulated and SERPINC1 was downregulated in the Autism patients' serum. ($p<0.05$, fold change > 1.2) Thus, those 3 proteins are putative potential biomarkers of ASD. (Figure 7)

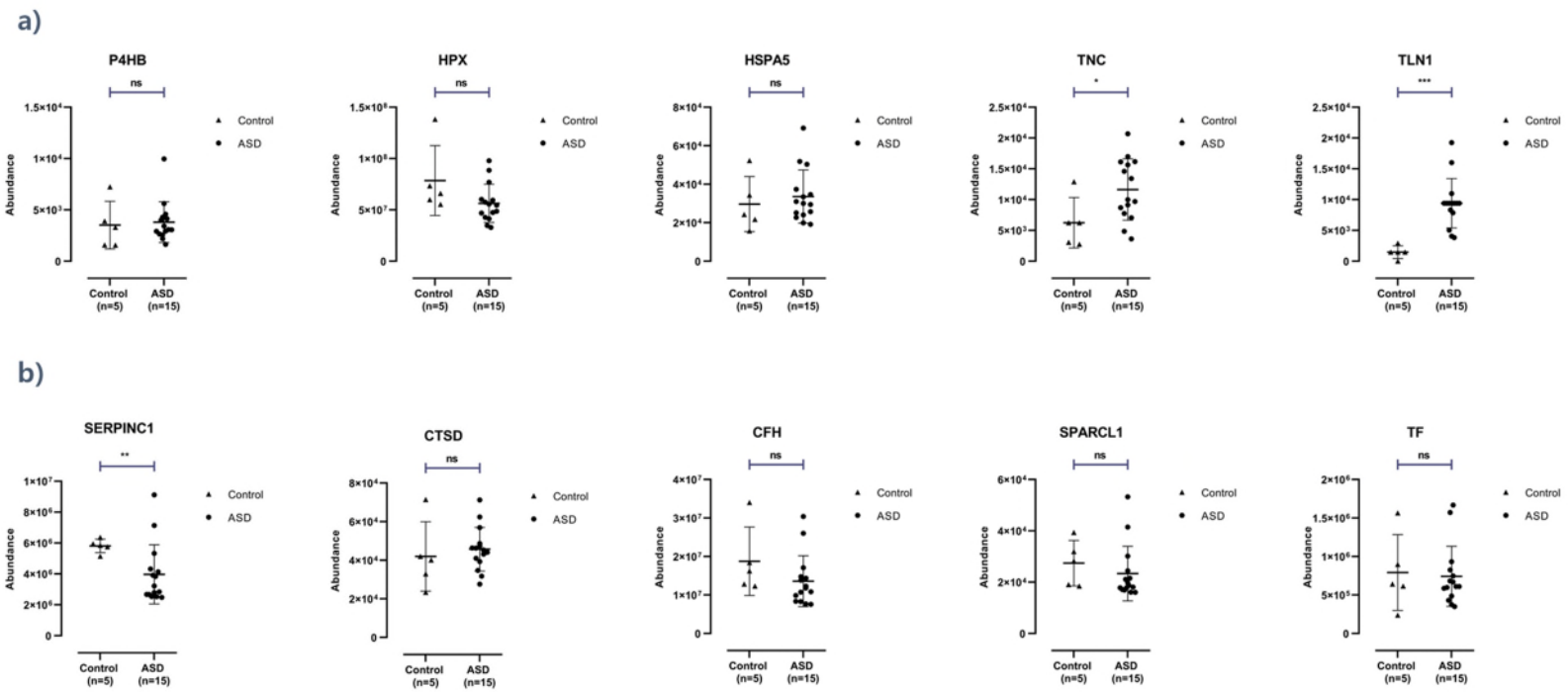

Figure 7. Comparison of protein expression levels in individual serums

a) Upregulated proteins in ASD patients and b) Downregulated proteins in ASD patients

4. Discussion

Increasing incidence of ASD, as well as other neurological dysregulation like attention deficit and hyperactivity has become a worldwide public health problem [17]. Multiple genes which involved in autism are reported, [18], however studies on proteins which could represent the status of ASD are insufficient. There is a demand for identifying practical molecular serum biomarkers for ASD.

Herein, we performed two sets of DDA-MS proteomic analyses based on the brains and sera of Cntnap2 KO mice, which is one of the optimal animal models that shows ASD core symptoms such as impaired sociability.

We examined proteins which altered its expression induced by Cntnap2 KO, and then tried to confirm whether the brain proteins which may involve in pathology of ASD are secreted into serum through the circulation systems. We have identified numerous biomarker candidates which differentially expressed in brain and serum of Cntnap2 KO mouse. We looked further into the functions of upregulated proteins in both brain tissues and sera in mouse models, because more likely it is to be selected as ASD potential biomarker. Most proteins were engaged in carbon metabolism and transferase activity carbohydrate derivative binding, metabolism of proteins immune system and catalytic activity.

We discovered total 144 potential ASD biomarker candidates using Cntnap2 KO mouse model, followed by mouse-to human translation. There is general hypothesis that biological processes in mice or rats mirror those in human under similar conditions. To confirm this hypothesis, we compared and investigated patterns between mouse and human proteome before translating DEPs mouse

models to human. We performed comparative analysis of canonical pathways. There were similar tendencies in both mouse models and human patients except the LXR/RXR activation. Accordingly, it was found that there was no difficulty in rendering from mouse datasets to humans and there were 19 DEPs after translation.

To find which types of proteins were enriched in both mouse and human, we performed functional classification of 19 potential ASD biomarkers. There were 4 categories: establishment of localization, cell adhesion, cellular response to interleukin-4 which could be related to brain-gut-axis and autism [19], immune system, which suggest a relationship between autoimmunity and autism [20].

For verification of these 19 selected DEPs, DIA-MS analysis of 20 individuals were performed and the protein alterations in human serum were compared. As a result, P4Hb, HPX, USPA5, TNC and TLN1 are upregulated and SERPINC1, CTSD, CFH, SPARCL1 and TF are downregulated in individual ASD patients' sera, respectively. Among these proteins, TNC, TLN1 and SEPRPINC1 were found to be statistically significant differential expression not only in humans but also in mouse models.

TNC, Tenascin-C, which contributes to extracellular matrix (ECM) remodeling, neuronal migration, and synaptic plasticity [21]. And it is associated with interleukin-1β (IL-1β), interleukin-8 (IL-8), tumor necrosis factor-α (TNFα) and the neuroinflammatory cascade through activation of Toll-like receptor 4 (TLR4) [22]. Elevated IL-6 levels promote excessive excitatory synaptogenesis and decrease inhibitory synaptogenesis. This is consistent with the imbalance

in excitatory inhibitory neuronal connections noted in ASD [23] and TNC is an important inducer to promote neuroinflammatory cascades.

TLN1, Talin-1, has emerged as a key cytoplasmic protein mediating integrin adhesion to the ECM and is well known to be associated with shank-complementary integrin activation pathways. Shank, one of the most prominent ASD risk genes, regulates cell function, the actin cytoskeleton for tissue organization. It is becoming increasingly clear that regulators of signaling to the actin cytoskeleton are disrupted by neuropsychiatric disorders [24]. We demonstrate a role for shank in controlling adhesion between cells and ECM and prevent TLN1 recruitment and integrin activation through binding and segregation of integrin activating GTPases [25].

SERPINC1, the only protein significantly downregulated in ASD patients, is a serine protease inhibitor of the serpin superfamily in which both intrinsic and extrinsic pathways control the proteolytic activity of procoagulant proteases, mainly depends on the activation of sensory neurons [26].

In conclusion, we present these three proteins, TNC, TLN1 and SERPINC1, as potential biomarkers, from animal model and ASD patients' cohorts. However, at a time ASD animal models are being used in various development studies internationally, it is necessary to use more diverse animal models to apply them to biological analysis studies and confirm in these biomarker data. And also, there is still a further way to go to translate into clinical evidence for validating effective and accurate biomarkers for ASD patients. Therefore, it is urgent to perform serum proteomic analysis relying on extensive independent large clinical ASD cohorts.

Reference

1. Faras, Hadeel, Nahed Al Ateeqi, and Lee Tidmarsh. "Autism spectrum disorders." *Annals of Saudi medicine 30.4* (2010): 295-300.

2. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z et al (2018) Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network, 11 sites, United States, 2014. MMWR *Surveill Summ* 67:1–23

3. Chaste, P., & Leboyer, M. (2012). Autism risk factors: genes, environment, and gene-environment interactions. *Dialogues in Clinical Neuroscience*, 14(3), 281–292.

4. DeFilippis, M., & Wagner, K. D. (2016). Treatment of Autism Spectrum Disorder in Children and Adolescents. *Psychopharmacology Bulletin*, 46(2), 18–41.

5. Shen, L. et al. (2019). Advances in Biomarker Studies in Autism Spectrum Disorders. In: Guest, P. (eds) Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders. *Advances in Experimental Medicine and Biology(), vol 1118*. Springer, Cham.

6. Ispas, C. R., Crivat, G., & Andreescu, S. (2012). Review: Recent Developments in Enzyme-Based Biosensors for Biomedical Analysis. *Analytical Letters, 45(2–3)*, 168–186.

7. Bourgeron, T. (2015). From the genetic architecture to synaptic plasticity in autism spectrum disorder. *Nat. Rev. Neurosci. 16*, 551–563

8. Rodenas-Cuadrado, P., Ho, J., & Vernes, S. C. (2013). Shining a light on CNTNAP2: complex functions to complex disorders. *European Journal of Human Genetics, 22*(2), 171–178.

9. Peñagarikano, O., & Geschwind, D. H. (2012). What does CNTNAP2 reveal about autism spectrum disorder? *Trends in Molecular Medicine, 18*(3), 156–163.

10. Peñagarikano, O., Abrahams, B., Herman, E., Winden, K., Gdalyahu, A., Dong, H., Sonnenblick, L., Gruver, R., Almajano, J., Bragin, A., Golshani, P., Trachtenberg, J., Peles, E., & Geschwind, D. (2011). Absence of CNTNAP2 Leads to Epilepsy, Neuronal Migration Abnormalities, and Core Autism-Related Deficits. *Cell, 147*(1), 235–246.

11. Poliak, S., Gollan, L., Martinez, R., Custer, A., Einheber, S., Salzer, J. L., Trimmer, J. S., Shrager, P., & Peles, E. (1999). Caspr2, a New Member of the Neurexin Superfamily, Is Localized at the Juxtaparanodes of Myelinated Axons and Associates with K+ Channels. *Neuron, 24*(4), 1037–1047.

12. Peñagarikano, O., Abrahams, B. S., Herman, E. I., Winden, K. D., Gdalyahu, A., Dong, H., Sonnenblick, L. I., Gruver, R., Almajano, J., Bragin, A., Golshani, P., Trachtenberg, J. T., Peles, E., & Geschwind, D. H. (2011b). Absence of CNTNAP2 Leads to Epilepsy, Neuronal Migration Abnormalities, and Core Autism-Related Deficits. *Cell, 147*(1), 235–246.

13. Brunner, D., Kabitzke, P., He, D., Cox, K., Thiede, L., Hanania, T., Sabath, E., Alexandrov, V., Saxe, M., Peles, E., Mills, A., Spooren, W., Ghosh, A., Feliciano, P., Benedetti, M., Luo Clayton, A., & Biemans, B. (2015). Comprehensive Analysis of the 16p11.2

Deletion and Null Cntnap2 Mouse Models of Autism Spectrum Disorder. *PLOS ONE, 10*(8), e0134572.

14. Savaryn, J. P., Toby, T. K., & Kelleher, N. L. (2016b). A researcher's guide to mass spectrometry-based proteomics. *PROTEOMICS, 16*(18), 2435–2443.

15. Park, G., Jeon, S. J., Ko, I. O., Park, J. H., Lee, K. C., Kim, M. S., Shin, C. Y., Kim, H., & Lee, Y. S. (2022). Decreased in vivo glutamate/GABA ratio correlates with the social behavior deficit in a mouse model of autism spectrum disorder. Molecular Brain, 15(1).
16. Cho, A., Haruyama, N., & Kulkarni, A. B. (2009). Generation of Transgenic Mice. Current Protocols in Cell Biology, 42(1).

17. Salari, N., Rasoulpoor, S., Rasoulpoor, S. et al. The global prevalence of autism spectrum disorder: a comprehensive systematic review and meta-analysis. *Ital J Pediatr 48*, 112 (2022).

18. Geschwind, D. H. (2008). Autism: Many Genes, Common Pathways? *Cell, 135*(3), 391–395.

19. Salim, S., Banu, A., Alwa, A., Gowda, S. B. M., & Mohammad, F. (2021). The gut-microbiota-brain axis in autism: what Drosophila models can offer? *Journal of Neurodevelopmental Disorders, 13*(1).

20. Enstrom, A. M., Van De Water, J. A., & Ashwood, P. (2009). Autoimmunity in autism. *Current Opinion in Investigational Drugs.*

21. Stamenkovic, V., Stamenkovic, S., Jaworski, T., Gawlak, M., Jovanovic, M., Jakovcevski, I., Wilczynski, G. M., Kaczmarek, L., Schachner, M., Radenovic, L., & Andjus, P. R. (2016). The extracellular matrix glycoprotein tenascin-C and matrix metalloproteinases modify cerebellar structural plasticity by exposure to an enriched environment. *Brain Structure and Function, 222*(1), 393–415.

22. Krasovska, V., & Doering, L. C. (2018). Regulation of IL-6 Secretion by Astrocytes via TLR4 in the Fragile X Mouse Model. *Frontiers in Molecular Neuroscience*, 11.

23. Uzunova, G., Pallanti, S., & Hollander, E. (2016). Excitatory/inhibitory imbalance in autism spectrum disorders: Implications for interventions and therapeutics. *World Journal of Biological Psychiatry, 17*(3), 174–186.

24. Monteiro, P., & Feng, G. (2017). SHANK proteins: roles at the synapse and in autism spectrum disorder. *Nature Reviews Neuroscience, 18*(3), 147–157.

25. Lilja, J., Zacharchenko, T., Georgiadou, M., Jacquemet, G., Franceschi, N., Peuhu, E., Hamidi, H., Pouwels, J., Martens, V., Nia, F., Beifuss, M., Boeckers, T., Kreienkamp, H. J., Barsukov, I., & Ivaska, J. (2017). SHANK proteins limit integrin activation by directly interacting with Rap1 and R-Ras. *Nature Cell Biology, 19*(4), 292–305.

26. Rezaie, A. (2014). Protease-activated receptor signalling by coagulation proteases in endothelial cells. *Thrombosis and Haemostasis, 112*(11), 876–882.

6. 국문초록

자폐 스펙트럼 장애(ASD)는 (1) 사회적 의사소통 장애, (2) 상호작용, (3) 반복적인 행동의 존재 및 제한된 관심의 세 가지 핵심 증상을 특징으로 하는 신경 발달 장애다. ASD 의 병태생리학적 기전을 밝히기 위한 다양한 연구가 진행되고 있지만, 질병의 원인에 대한 이해가 부족하여 신뢰할 수 있는 진단 바이오마커가 없는 실정이다. ASD 의 유병률이 급격히 증가함에 따라 ASD 관련 분자 바이오마커를 확인하는 것이 시급하다.

이 연구에서는 ASD 의 잘 알려진 유전 마우스 모델 인 Contactin 관련 단백질 유사 2 (Cntnap2) KO 마우스의 뇌 조직 및 혈청 단백질 분석을 수행했다. DDA(Data Dependent Acquisition) 질량 분석법을 사용하여 뇌 조직 샘플에서 차별적으로 발현된 단백질을 식별한 다음 혈액으로 분비된 단백질을 식별하기 위해 혈청 샘플을 분석했다. 선별된 단백질은 ASD 환자의 혈청을 사용하여 DIA(Data Independent Acquisition) 질량 분석법으로 추가 검증되었다. 우리는 환자의 혈청에서 3 개의 단백질 TNC, TLN1 및 SERPINC1 을 확인하였으며 이들을 잠재적 ASD 바이오마커 후보로 제안한다.