Research Article   **Open Access**

Reelin Associated With Restricted and Stereotyped Behavior Based on Principal Component Analysis on Autism Diagnostic Interview-Revised

Ulrika Roine[1]*, Samuli Ripatti[2,3,4], Karola Rehnström[3], Timo Roine[5], Helena Kilpinen[3], Ida Surakka[2], Juho Wedenoja[2], Tero Ylisaukko-oja[2,6], Elli Kempas[2], Jaana Wessman[2], Irma Moilanen[7], Marja-Leena Mattila[7], Marko Kielinen[7], Katja Jussila[7], Saara Suomalainen[8], Esko Pulkkinen[9], Lennart von Wendt[10] and Leena Peltonen[2,3]

[1]Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076 AALTO, Finland
[2]Institute for Molecular Medicine Finland, Human Genomics, Helsinki, Finland
[3]Wellcome Trust Sanger Institute, Human Genetics, Cambridge, United Kingdom
[4]Hjelt Institute, University of Helsinki, Finland
[5]Department of Automation and Systems Technology, Aalto University, Espoo, Finland
[6]Institute of Health Sciences (General Practice), University of Oulu, Finland
[7]Clinic of Child Psychiatry, University of Oulu, Finland
[8]Lapland Hospital District, Rovaniemi, Finland
[9]Clinic of Psychiatry, Pieksämäki, Finland
[10]Hospital for Children and Adolescents, Unit of Child Neurology, Helsinki, Finland

Abstract

Background: Twin and family studies have indicated a strong genetic component in autism spectrum disorders, and genetic studies have revealed highly heterogeneous risk factors. The range and severity of the symptom presentation also vary in the spectrum. Thus, symptom-based phenotypes are putatively more closely related to the underlying biology of autism than the end-state diagnosis.

Methods: We performed principal component analysis on Autism Diagnostic Interview-Revised algorithm for 117 Finnish families and 594 families from the Autism Genetic Research Exchange (AGRE). The resulting continuous component scores were used as quantitative phenotypes in family-based association analysis. In addition, K-means clustering was performed to cluster and visualize the results of the PCA. Unaffected siblings were included in the study.

Results: The components were interpreted as Social Component (SC), communication component and Restricted and Stereotyped Behavior Component (RSBC). K-means clustering showed that, especially in SC, the range of the symptom severity was increased by the siblings. The association of neuroligin 1 with SC was increased, compared to a previous study where only the end-state diagnosis was used. In RSBC, the range of the symptom severity of siblings overlapped greatly with that of patients, which could explain why no association of reelin was found in previous studies in which only the end-state diagnosis was used, but a significant association of reelin with RSBC was now found in the Finnish families (Bonferroni-corrected p=0.029 for rs362644). Although, the Finnish sample is isolated and genetically very homogeneous, compared to the heterogeneous background of AGRE families, many single-nucleotide polymorphisms in reelin, showed modest association with RSBC in the AGRE sample, too.

Conclusions: This study demonstrates how the quantitative phenotypes can affect the association analyses, and yields further support to the use of siblings in the study of complex neuropsychiatric disorders.

Keywords: Autism; ADI-R; Principal component analysis; K-means clustering; Neuroligin 1; Reelin; Quantitative phenotype; Association analysis

Abbreviations: ADI-R: Autism Diagnostic Interview-Revised; AGRE: Autism Genetic Research Exchange; ASDs: Autism Spectrum Disorders; CC: Communication Component; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, fourth edition; FIN: Finnish (data sample); ICD-10: International Classification of Diseases, tenth revision; NLGN1: Neuroligin 1; PC: Principal Component; PCA: Principal Component Analysis; PDD-NOS: Pervasive Developmental Disorder Not Otherwise Specified; RELN: Reelin; RSBC: Restricted and Stereotyped Behavior Component; SC: Social Component; SNP: Single-nucleotide Polymorphism

Background

Autism is a neurodevelopmental disorder characterized by deficits in reciprocal social interactions and communication, and restricted and stereotyped behavior. It has been estimated to affect 20.6 per 10,000 individuals, while Autism Spectrum Disorders (ASDs) altogether have been estimated to affect approximately 60-70 per 10,000 [1]. In Finland, the prevalence of autism is 41 per 10,000, and that of ASDs is 84 per 10,000, according to DSM-IV [2,3]. Although twin and family studies have indicated a strong genetic component in ASDs [4,5], resulting in a heritability estimate of over 90% [6], only 10-15% of the cases can be explained by a variety of single-gene disorders [7].

The symptoms of ASDs have been described to belong to a complex continuum, where each symptom can vary from mild to severe [8]. Constantino [7] also emphasized in his recent publication that

---

*Corresponding author: Ulrika Roine, M.D., M.Sc., Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076 AALTO, Finland, Tel: +358 50 571 0003; E-mail: ulrika.roine@helsinki.fi

understanding the social abnormality of autism as a quantitative trait, instead of a categorically defined disorder is important. Autism-related traits have been reported also in close relatives of individuals with ASDs [6], supporting the idea that ASDs are the other end of a continuous spectrum of symptoms present, also in general population. Genetic studies using quantitative traits have indeed shown enhancement in specific linkage signals, while also using siblings in the analysis [9,10]. In addition, it has been demonstrated that the same genes that predispose to complex disorders could be responsible also for the phenotypic variation in the general population [7]. Thus, the use of quantitative traits and siblings in the genetic analyses of complex disorders can be essential.

Autism Diagnostic Interview-Revised (ADI-R) is a widely used clinical diagnostic instrument in assessment of autism [11]. Therefore, it would be valuable to have a simple and practical way of using the ADI-R results as phenotypic data in genetic analyses. In ADI-R algorithm, the information received from the ADI-R is gathered together, and the diagnosis according to ADI-R is calculated from the sums of the different subareas of the algorithm. Although this is one way to reduce the data in the ADI-R algorithm, for the purposes of genetic analyses, there may be more efficient ways. Principal Component Analysis (PCA) is a computational tool, which reduces the dimensions of the data by maximizing the amount of variation preserved and forms orthogonal, independent principal components [12]. Thus, the few first principal components contain most of the variation in the data, and in addition, the components are independent of each other. This could be important, as also the predisposing genetic factors may be independent of each other. Furthermore, the data is not influenced by any assumptions made beforehand.

In some previous studies, PCA has been used to reduce the dimensions of ADI-R data. However, the studies have either focused only on a certain domain of ADI-R [13,14], or used ADI-R items to form clusters based on the principal components and intracorrelating items [15]. Those clusters or in some studies a part of the clusters have been used in genetic analyses with enhanced results [16-19]. Sutcliffe et al. [20] used the clusters to explore phenotypic correlates with the coding variants in their study. In addition, factor analysis and clustering methods have been used to form models, explaining the symptom spectrum of ASDs [21-24].

In this study, we have performed a principal component analysis for ADI-R algorithm data, resulting in independent principal components. The scores of the components were then used in genetic analyses. Our study differs from previous studies in many aspects: only ADI-R algorithm data was used, instead of the whole interview data; all subareas of the symptomatology were studied, the results of the PCA were directly used in genetic analyses without any modifications, and finally, siblings were included in the study. In no previous study all of these aspects were included. In addition, our aim was not to build a model to describe the whole spectrum of symptoms in ASDs, which perhaps would have required a different approach, as used by Van Lang et al. [22], Frazier et al. [23] and Lecavalier et al. [24], but to use a robust and practical method to construct independent quantitative phenotypes from the widely used ADI-R for genetic analyses.

In addition to performing PCA, we interpreted the principal components by calculating correlations between the components and ADI-R items. We also used the components in clustering individuals with k-means clustering [25], to visualize the results of PCA, and to further validate the interpretations of the components. In addition to test the reliability of the components, we calculated Cronbach's alphas

for the ADI-R items correlating with the components, and compared them with the Cronbach's alphas of the items belonging to the subareas of the ADI-R algorithm. Finally, we used the principal component scores, describing the positions of the individuals along the principal components, as quantitative phenotypes in family-based association analyses of previously genotyped candidate genes. All analyses were first conducted in a Finnish sample, after which they were repeated in a sample received from Autism Genetic Research Exchange (AGRE) [26]. However, it is noteworthy that the Finnish sample is isolated and genetically very homogeneous [27], whereas the AGRE sample has been gathered all over the United States, and it is only required that one of the parents is English-speaking [28].

Materials and Methods

Ethics statement

For the Finnish sample, the ethics committee of Helsinki University Central Hospital approved the research protocol. Also, all participants or their parents have signed a written consent form.

Finnish sample

The Finnish sample included 117 families with autism. They were recruited through Finnish university and central hospitals, and the final diagnoses were established by a multidisciplinary group headed by an experienced child neurologist. Only individuals fulfilling both ICD-10 and DSM-IV criteria were included. The ADI-R (version published in 1994 [11]) was administered in Finnish for all individuals included in the PCA. However, all interviewers did not have inter-rater reliability. Yet, all but one of the interviewers had been schooled together to the use of ADI-R, and they also did interviews together in the training phase. In total, the study sample included 177 individuals from 117 families, for whom ADI-R data was available. Of these individuals, 120 were diagnosed with autism, 10 with a Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), and five with Asperger syndrome. The remaining 42 individuals with available ADI-R data were siblings of the affected individuals, who were asked to participate in the study, whether or not any symptoms had been noticed earlier.

AGRE sample

AGRE is one of the world's largest shared genetic resources for the study of autism. DNA, clinical and medical information has been gathered from families throughout the United States. In the AGRE families, one of the parents must be English-speaking. Unfortunately, clinical diagnoses are not available for AGRE participants. Instead, the diagnostic assessment in the AGRE is based on the ADI-R, which is based on both ICD-10 and DSM-IV. All individuals in the AGRE sample with the same version of the ADI-R (referred to 1995 version in AGRE) were included in this study, resulting in a total of 1395 individuals from 594 families. 1075 individuals had a diagnosis of autism and the rest belonged to the "Not Quite Autism", "Broad Spectrum", or "Not Met" category. These categories are described in detail on AGRE website [29], except the "Not Met" category. It consists of individuals who did not meet the criteria for any of AGRE's affected status categories. In our study, all individuals in the "Not Met" category were siblings of affected individuals.

Autism Diagnostic Interview–Revised

ADI-R is a widely used structured diagnostic interview used both by clinicians and researchers. It is administered to the primary caregivers of individuals with suspected autism diagnosis. The interview

contains 111 questions concerning the behavior and development of the child, and provides data on all three symptom domains in ASDs. ADI-R algorithm has been designed to gather together the information received from the ADI-R, and it contains four subareas concerning social behavior, communication, restricted and repetitive behaviors, and development. The diagnosis according to ADI-R is then based on certain threshold values of the sums of the subareas of the ADI-R algorithm. Although in this study, the earlier version of ADI-R was used, algorithms for both versions [11,30] are comprised of the same items. In our study, the values of individual items were used exactly as they are coded in the algorithm, as explained in detail on AGRE website [31]. Missing values were coded as -1. ADI-R takes at least two hours to complete, which has limited its applicability in large studies. Therefore, we decided to use the ADI-R algorithm instead of the whole interview, as in the future it could be possible to use only the questions in the algorithm for research purposes.

Principal component analysis and interpretation of the principal components

Principal Component Analysis (PCA) was then performed for the data, consisting of all items in the ADI-R algorithm. PCA is a statistical method, which performs a linear transformation of the data into a new orthogonal basis. The new variables, called Principal Components (PCs), are linear combinations of the original variables selected to preserve as much of the variance as possible, and ordered by decreasing variance. PCA is performed by calculating the singular value decomposition of the data matrix. PCA is an excellent tool to reduce the dimensions of the data, as the first principal components often explain most of the variance in the data. In addition, the orthogonality of the new basis means that the PCs are independent of each other, which is often useful. Factor analysis has also been used to find the main factors of ADI-R data [24], and PCA and factor analysis are often confused. While PCA maximizes the amount of variation preserved, the aim in factor analysis is to estimate interdependence between variables, and thereby find common factors, which however, do not necessarily contain most of the variation in the data, and are not necessarily independent of each other.

PCA was performed for both Finnish and AGRE data. In the Finnish sample, the amount of principal components used in the sequential analyses was chosen according to commonly used Kaiser's criteria [32], where only principal components with an eigenvalue greater than one are included in the analyses. In the AGRE sample, the same amount of components was chosen. The correlations (c) between the algorithm items and the most significant (eigenvalue>1) principal components were calculated, and the components were interpreted according to the correlations. A correlation threshold of 0.2 was chosen as it is statistically significant for the Finnish sample ($p<0.01$), and of course, for the much larger AGRE sample too [33]. The statistical significance of the correlations is further increased as the use of ordinal data reduces the amount of correlation observed [34]. It should be noted that in PCA, the direction of the axes is chosen arbitrarily. Thus, if all algorithm items are correlated negatively to the first principal component in the Finnish data and the same items are correlated positively in the AGRE data, it means that the direction of the axis has been chosen differently in the two analyses, but actually the two data sets are correlated in the same way with the items. Therefore, the correlation threshold actually means that all items with a correlation coefficient above 0.2 or below -0.2 were chosen for the interpretation.

Visualization and validation of the data

In addition to the correlation analysis, the most significant PCs were used in clustering individuals into three clusters to visualize the data, and to further validate the interpretations of the components. The clustering was performed using the k-means algorithm, which minimizes the sum of the distances between the points, and the centroid of the cluster over all clusters. Squared Euclidean distances were used in this study. Some relevant information such as the number of individuals with an autism diagnosis according to ADI-R, sex, and age, were calculated for each of the clusters.

Principal components carry the maximal amount of data preserved, and are independent of each other. The ADI-R scores are computed for each domain, as listed in the diagnostic criteria for research (World Health Organization, 1992) [11]. Thus, they do not necessarily carry the maximal amount of variation and neither are they necessarily independent of each other. However, for genetic analyses, both of these properties are useful, and therefore, we performed PCA on ADI-R algorithm data. It is also important that the PCs have good reliability. We calculated Cronbach's alpha [35], a reliability measurement for the most significant PCs, by using the answers of the significantly correlating items, and compared them to the ADI-R algorithm domains by using the answers to the items that belong to the different subareas of the ADI R algorithm. MATLAB® [36] was used in PCA, correlation analysis, clustering and reliability measurements.

Family-based association analysis

Finally, we applied the information provided by PCA to genetic analyses. We performed family-based association analysis with quantitative phenotypes for previously genotyped candidate genes in Finnish population [37-41] (and unpublished data), including 290 Single Nucleotide Polymorphisms (SNPs) in 23 genes (Supplementary table 1). All SNPs in the Finnish sample were tested with each of the three PCs. The SNPs are located in known autism susceptibility genes, as well as genes in regions identified in several Finnish genome-wide linkage and association scans for ASDs. Genotyping was performed using allele specific primer extension on microarrays [42], Homogenous MassEXTEND (hME) and iPLEX (Sequenom Inc., San Diego, CA, USA) technology using the MassARRAY Platform, as specified by manufacturer's instructions. Association analyses in the AGRE sample were performed for the two genes, which resulted in most significant associations in the Finnish sample. The SNPs for the AGRE families were extracted from genome-wide SNP data from Affymetrix 5.0 chips genotyped at the Broad Institute [43].

For the family-based association analyses, a linear regression of phenotype on genotype was performed, and permutation was used to correct for the dependence between related individuals [44]. The permutation procedure is also robust for non-normal phenotype distributions. The principal component scores, describing the positions of the individuals along the PCs, were used in family-based association analysis as quantitative phenotypes. The association analyses were carried out with the PLINK software package, by using the QFAM option [45-49]. Physical positions of SNPs were obtained from NCBI dbSNP build 125. Haploview was used to check the haplotype blocks of the most significantly associating SNPs, and Bonferroni correction was used to correct for multiple testing of the data.

Results and Discussion

PCA of ADI-R algorithm data resulted in three PCs, with an eigenvalue greater than 1 in the Finnish sample. In the AGRE sample, the same amount of components as for the Finnish sample was selected, although the eigenvalue for the fourth component was slightly over 1.

In the Finnish sample, the selected components accounted for 61% of variation in the data, while in AGRE data they accounted for 47% of variation. Although, for practical reasons, only the first three PCs were used in this study, it is possible that also the PCs with smaller eigenvalues could contain relevant phenotypic information, and should be investigated in the future. The eigenvalues for both samples are presented in supplementary table 2.

Interpreting the principal components

The interpretations of the three PCs were formed based on the correlation analysis. Correlation coefficients which were greater than 0.20 are shown in bold in table 1. The PC correlating with items concerning social interactions was named Social Component (SC), and the PC correlating with items concerning communication was named Communication Component (CC). Respectively, the third principal component was named Restricted and Stereotyped Behavior Component (RSBC).

Despite the similar themes of the components in the two samples, there were some differences too. First, in the Finnish data, the SC accounted for 37% and CC for 20% of variation in the data, whereas in the AGRE sample the SC accounted for 17% and the CC for 26% of variation. RSBC was the third PC in both samples and accounted for 4% of total variance, both in the Finnish and AGRE samples. Second, there were some differences in the contents of the components, as can be seen in table 1. For example, more items correlated with the SC in the Finnish sample than in the AGRE sample, and different items in the two samples had the biggest correlation with the RSBC: hand and finger

ADI-R algorithm item	F-SC	A-SC	F-CC	A-CC	F-RSBC	A-RSBC	ADI-R domain
Direct gaze	-0.19	**0.21**	0.03	-0.05	-0.07	-0.03	SOC
Social smiling	**-0.22**	**0.22**	0.03	-0.08	-0.07	0.04	SOC
Range of facial expressions	**-0.22**	**0.21**	-0.00	-0.07	-0.05	0.05	SOC
Imaginative play with peers	-0.19	**0.20**	0.05	-0.01	-0.03	0.01	SOC
Interest in children	**-0.23**	0.19	-0.02	-0.10	-0.02	0.10	SOC
Response to other children's approaches	**-0.20**	0.18	-0.03	-0.09	-0.03	0.06	SOC
Group play with peers or friends	**-0.21**	0.17	0.03	-0.01	-0.00	-0.01	SOC
Showing and directing attention	**-0.23**	**0.20**	0.01	-0.12	-0.01	0.10	SOC
Offering to share	-0.17	0.13	-0.01	-0.08	0.06	0.05	SOC
Seeking to share own enjoyment with others	**-0.22**	**0.21**	-0.00	-0.10	-0.05	0.10	SOC
Use of other's body	-0.08	0.11	-0.02	-0.11	**0.28**	-0.13	SOC
Offers comfort	**-0.20**	0.17	0.00	-0.11	-0.08	0.09	SOC
Quality of social overtures	**-0.22**	**0.22**	0.03	-0.11	-0.17	-0.00	SOC
Inappropriate facial expressions	-0.10	0.14	0.04	-0.05	**0.20**	-0.17	SOS
Appropriateness of social responses	**-0.20**	0.19	0.02	-0.09	-0.10	0.05	SOS
Pointing to express interest	-0.19	0.17	0.04	-0.13	-0.12	0.06	COM
Conventional instrumental gestures	**-0.20**	**0.20**	0.02	-0.12	-0.15	0.07	COM
Nodding	-0.17	0.16	-0.02	-0.16	**-0.22**	0.15	COM
Headshaking	**-0.20**	0.17	-0.03	-0.15	-0.17	0.18	COM
Spontaneous imitation of actions	**-0.21**	0.12	-0.01	-0.08	-0.09	0.07	COM
Imaginative play	**-0.20**	0.15	0.05	-0.09	-0.03	0.03	COM
Imitative social play	-0.18	0.17	0.02	-0.08	-0.10	0.04	COM
Social chat	-0.00	**0.22**	**0.39**	**0.31**	0.01	**0.22**	COM
Reciprocal conversation	-0.02	**0.23**	**0.42**	**0.38**	0.03	**0.23**	COM
Stereotyped utterances	0.01	**0.21**	**0.50**	**0.41**	-0.00	-0.03	COM
Inappropriate questions	0.07	0.12	**0.34**	**0.34**	-0.16	-0.13	COM
Pronominal reversal	0.04	0.17	**0.36**	**0.36**	-0.01	-0.03	COM
Neologisms/idiosyncratic language	0.09	0.09	**0.28**	**0.27**	-0.04	-0.08	COM
Circumscribed interests	-0.01	0.11	0.15	0.13	0.14	-0.19	RSB
Unusual preoccupations	-0.12	0.06	0.04	-0.03	**0.29**	**-0.25**	RSB
Verbal rituals	-0.04	0.12	**0.21**	0.13	0.10	-0.15	RSB
Compulsions/rituals	-0.13	0.09	0.03	0.03	**0.24**	**-0.43**	RSB
Hand and finger movements	-0.13	0.13	-0.01	-0.09	**0.57**	**-0.27**	RSB
Complex mannerisms	-0.16	0.11	0.01	-0.10	**0.24**	**-0.33**	RSB
Repetitive use of objects	-0.17	0.14	0.01	-0.06	**0.29**	**-0.35**	RSB
Unusual sensory interests	-0.16	0.11	0.01	-0.09	0.10	**-0.29**	RSB
Age when parents first noticed	-0.05	0.02	-0.00	-0.02	-0.00	-0.02	DEV
Age when abnormality first evident	-0.08	0.04	0.00	-0.02	-0.04	-0.01	DEV
Interviewer's jugdement on age when developmental abnormalities first manifest	-0.04	0.02	-0.01	-0.01	0.01	-0.01	DEV
Age of first single words	-0.01	0.03	-0.04	-0.06	0.02	0.04	DEV
Age of first phrases	-0.07	0.03	-0.02	-0.06	0.02	0.04	DEV

F=Finnish sample, A=AGRE sample, SC=Social Component, CC=Communication Component, RSBC=Restricted and Stereotyped Behavior Component. (Correlation coefficients of absolute value over 0.20 are bold).

Table 1: ADI-R algorithm item correlations with the three most significant principal components in Finnish and AGRE samples.
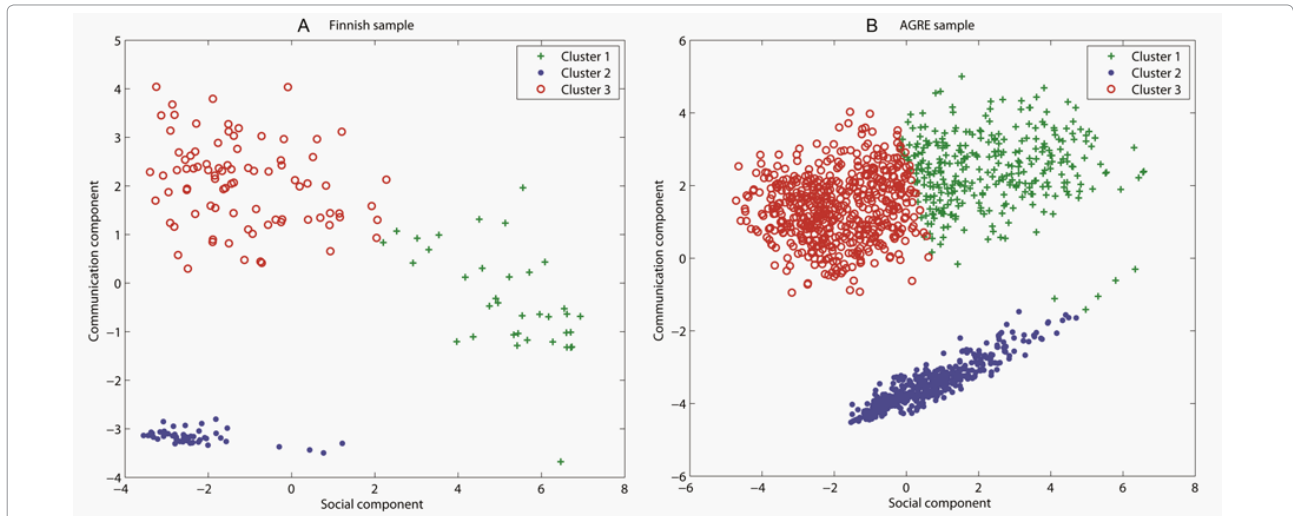
Figure 1: Clusters based on PCA and k-means clustering of ADI-R data in Finnish and AGRE samples. Although cluster 1 contains mostly unaffected individuals according to ADI-R scores (see table 2), especially in the Finnish sample, there is a lot of variation in the symptom severity in that cluster. Thus, the inclusion of unaffected siblings increases the range of the symptom severity, increasing also the power of association analyses with quantitative phenotypes. The direction of the SC in AGRE sample has been inverted for visualization purposes, as the SC was in the opposite direction in the two samples (in PCA the direction of the axes is chosen arbitrarily).

Variable	FIN C1	AGRE C1	FIN C2	AGRE C2	FIN C3	AGRE C3
1. Number of individuals	41	342	50	448	86	544
2. Speaking individuals (%)	97.56%	98.25%	0.00%	0.00%	100.00%	100.00%
3. Average age (years)	12.668	8.4229	13.201	7.4542	12.795	9.6887
4. Males (%)	60.97%	74.24%	75.99%	77.55%	75.58%	78.25%
5. Social interaction (avg)	3.9756	11.099	25.8	22.824	23.291	23.623
6. Verbal communication (avg)	4.2927	9.6637	-1	-2	17.93	18.007
7. Non-verbal communication (avg)	-0.97561	-1.8889	13.04	12.277	-1	-2
8. Repetitive/stereotyped behavior (avg)	1.6829	4.4298	5.78	5.0536	7.1512	6.5257
9. Abnormality of development (avg)	2.0488	3.3012	4.14	4.6116	3.4651	4.1452
10. Diagnosis according to ADI-R (%)	9.76%	42.98%	94.00%	91.07%	94.19%	95.59%

FIN=Finnish sample, AGRE=AGRE sample, C1=Cluster 1, C2=Cluster 2, C3=Cluster 3
Number of individuals, percent of speaking individuals, average age of individuals, and percent of men are provided for all clusters in both samples. In addition, the average score (of all individuals in a cluster) for all subareas in ADI-R algorithm is provided for all clusters in both samples.

Table 2: Descriptive phenotypic information of the three clusters in Finnish and AGRE samples (shown in figures 1 and 2).

movements in the Finnish data and compulsions/rituals in the AGRE data. Finally, only one of the items significantly correlated with more than one PC in the Finnish data, whereas in AGRE data three items correlated with two components.

The differences between the two samples could be explained by various things. Most importantly, the population of Finland is isolated and genetically homogeneous, whereas the AGRE sample has been gathered all over the United States, and the subjects are only required to have one English-speaking parent. Thus, it is reasonable to assume that the AGRE sample is both genetically and phenotypically more heterogeneous than the Finnish sample. This can lead to stronger cultural differences, both within AGRE sample and between Finnish and AGRE sample.

Visualizing the data by k-means clustering

The results of the k-means clustering are shown in figures 1 and 2, and the calculated properties of the clusters are shown in table 2. Visually, two clusters can be clearly recognized in the data, one consisting of verbal and other of non-verbal individuals. However, we

have selected to use three clusters instead of two, because especially in the Finnish sample the k-means clustering was then able to separate almost all unaffected individuals based on ADI-R scores into a separate cluster (cluster 1). Likewise, in AGRE sample, most of the unaffected individuals belonged to cluster 1, but also about 40% of the individuals in that cluster had a diagnosis of autism. Clusters 2 and 3 contained mostly affected individuals, according to ADI-R in both samples (Table 2).

The clustering supported our interpretations of the components, as the component that we had named social component could be used to distinguish subjects based on their social skills (according to the social domain of ADI-R algorithm), and the same could be applied to communication component (Figure 1). In RSBC, there was a big overlap in the range of values of all three clusters, as shown in figure 2, but according to the repetitive/stereotyped behavior domain of ADI-R algorithm the severity of symptoms in restricted and stereotyped behavior was lowest in cluster 1. Respectively, the range of the symptom presentation was smallest in cluster 1 (Figure 2). Due to the differences in the samples, the orientation of the principal components differs
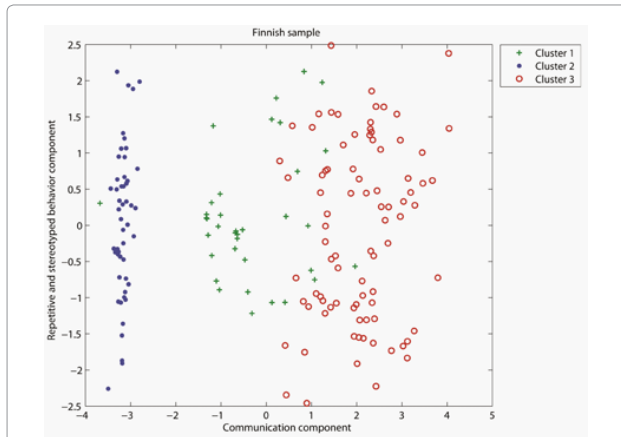
Figure 2: Visualization of the variation in the restricted and stereotyped behavior component in the Finnish sample. In RSBC, the range of the symptom severity of unaffected individuals (according to ADI-R) overlaps greatly with that of patients. If only the end-state diagnosis is used in family-based association analyses, this could lead to negative results, because the unaffected individuals do not differ much from the affected ones.

Component	FIN PCA	FIN ADI-R	Between FIN PCA & ADI-R	AGRE PCA	AGRE ADI-R	Between AGRE PCA & ADI-R
SC	0.9681	0.9545	0.7506	0.8474	0.8732	0.6434
CC	0.9197	0.8269	0.6331	0.7449	0.7278	0.6014
RSBC	0.7579	0.7263	0.4470	0.8053	0.4638	0.6785

FIN=Finnish sample, AGRE=AGRE sample, PCA=Cronbach's alpha for significantly correlating items in the principal components (PCA), ADI-R=Cronbach's alpha for the equivalent ADI-R algorithm subareas, Between=Cronbach's alphas between the principal component scores and equivalent ADI-R algorithm subarea scores

Table 3: Cronbach's alphas for the significantly correlating items in the principal components, and equivalent ADI-R algorithm subareas in Finnish and AGRE samples.

between the Finnish and AGRE samples, approximately 30 degrees.

Furthermore, K-means clustering showed that the unaffected individuals (according to ADI-R) also had variation in the symptom severity. Especially in SC, the range of the symptom severity was increased by the siblings (Figure 1), which could give more power to the association analysis with quantitative phenotypes. In RSBC, the range of the symptom severity of the unaffected individuals overlapped greatly with that of patients (Figure 2). In family-based association analyses where only the end-state diagnosis is used, this could lead to negative results because of the similar range of symptom severity for the affected and unaffected individuals.

Measuring the reliability of the principal components with Cronbach's alpha

Interestingly, the Cronbach's alphas for the ADI-R subareas were greater in the Finnish sample than in the AGRE sample, as shown in table 3. The weaker consistency of the subarea content could indicate that the AGRE sample is phenotypically more heterogeneous than the Finnish sample. For all three PCs, the Cronbach's alphas were greater in the Finnish sample than in the AGRE sample, indicating that the PCs also had better internal consistency in the Finnish sample than in the AGRE sample. In the Finnish families, for all of the three PCs, the alphas were greater for the items correlating with the PCs than for the ones included in the equivalent ADI-R algorithm subareas. This indicates that the data content of the principal components was internally more

consistent than the data in the ADI-R algorithm subareas, and thus, the reliability of the components was better. In the AGRE sample, the CC and RSBC also had greater alphas than the equivalent ADI-R algorithm subareas, whereas the alpha of SC was slightly lower.

In addition, the principal component scores and the equivalent ADI-R algorithm subarea totals were compared with Cronbach's alphas. They varied from 0.45 to 0.75, indicating that the principal components do not describe exactly the same thing as the subareas. Especially the RSBC component consisted of many different items than the equivalent subarea in the algorithm (Table 1). This explains why the range of symptom severity overlapped so much in RSBC in all the three clusters (Figure 2), although the cluster averages based on ADI R restricted and stereotyped behavior subarea differed from each other (Table 2).

Comparison of our methodology and results concerning PCA with others' studies

Tadevosyan-Leyfer et al. [15] performed PCA to common questions or items in ADI and ADI-R, and then used the results to cluster the items. However, after that they removed items based on multiple criteria and performed the PCA again to the remaining items. No reliability scores were used to show the superiority of the modified PCs, in comparison to the original ones. We did not want to remove any data, because it would lead to losing some of the variation in the data. Besides, the correlation of an item with many components does not mean that it would not contain important information to the components. Furthermore, the PCs are by definition, independent of each other. Actually, in our analysis, the first three principal components accounted for 61% of variation in the Finnish data and 47% of variation in the AGRE data, while in Tadevosyan-Leyfer's study, six clusters of items accounted for about 41% of variation. However, these values are not directly comparable, because they used the original ADI-R questions, while we used the questions in the ADI-R algorithm.

In some studies [22-24], the aim has been to find an optimal model describing the symptomatology of autism. Thus, the researchers wanted to make a model that would be simple to understand and would best describe the data, whereas we wanted to use existing phenotype data in a more efficient way, for the purposes of association analyses. Different goals lead to different methods, as a model that describes well the whole

CHR[a]	SNP[b]	EMP1[c]	BP POSITION[d]	SAMPLE
7	rs362731	0.003643	102996264	Finnish
7	rs362732	0.008088	102997304	Finnish
7	rs123712	0.003961	103018526	Finnish
7	rs2249372	0.004579	103025446	Finnish
7	rs362644	0.000101	103039099	Finnish
7	rs6958081	0.00439	103079061	Finnish
7	rs3919520	0.002824	103079723	Finnish
7	rs13239238	0.00345	103165523	Finnish
7	rs1963647	0.001698	103231266	Finnish
7	rs262375	0.006105	103240293	Finnish
7	rs4727583	0.006032	103386180	AGRE
7	rs4727582	0.006333	103386042	AGRE
7	rs2109697	0.00717	103020445	AGRE
7	rs6943822	0.007598	103385907	AGRE
7	rs10247439	0.008886	103143532	AGRE

[a]Chromosome
[b]Single Nucleotide Polymophism
[c]Empirical p-value given by PLINK based on one million permutations
[d]Base Pair Position from NCBI dbSNP build 125

Table 4: SNPs within RELN showing association with RSBC (p<0.01).

spectrum of the symptomatology does not necessarily describe as well a certain phenotype.

Factor analysis has also been used to find factors from ADI-R data [24]. However, we chose to use PCA for three reasons. We wanted to preserve the maximum amount of variation in the resulting components, and we wanted that the resulting components would be independent of each other. Both properties are important when the aim is to use the components in genetic analyses. In factor analysis, the aim is not to preserve the maximal amount of variation, but to find factors consisting of items that correlate with each other, as much as possible. The resulting factors are not necessarily independent either. In addition, in factor analysis, the factors are often rotated to find the best possible explanation of the factors. We did not want to subjectively influence the data by rotating the factors to correspond to any assumptions made beforehand.

Family-based association analysis

Finally, the continuous PC scores were used as quantitative phenotypes in family-based association analysis. It was first performed for the Finnish sample, where two genes could be distinguished: Neuroligin 1 (NLGN1) and Reelin (RELN). Then the association analysis was repeated in the AGRE sample for 144 SNPs within NLGN1, and 103 SNPs within RELN. As we used existing data, the SNP panels in the Finnish and AGRE data did not contain overlapping SNPs. However, both SNP sets have been designed to tag common variation in the genes examined. Particularly in RELN, where the associated SNPs were located in the same haplotype block in Finnish and AGRE sample, these SNPs tag the same genetic risk factors.

In the Finnish sample, the SNP rs1488545 in NLGN1 showed association with the SC (p=0.0004). Neuroligins are a family of genes encoding postsynaptic cell adhesion molecules, and mutations in the family have been reported before in ASDs and mental retardation [50,51]. The association of NLGN1 with SC improved from a previously reported association with the end-state diagnosis of autism in the same sample (formerly p=0.002 for rs1488545) [39], perhaps due to the use of the quantitative phenotype and inclusion of siblings. In contrast, the lack of association of NLGN1 in AGRE sample, and the fact that the result in the Finnish sample did not endure Bonferroni correction, indicates that it might be a false positive. However, the different results in the two samples could also be explained by differences in the SC and phenotypic and genetic backgrounds of the samples.

Several SNPs in RELN showed association with RSBC (p<0.01) in the Finnish sample. The SNP rs362644 also endured the Bonferroni correction (uncorrected p=0.0001 and Bonferroni-corrected p=0.029). In the embryonic brain, RELN is involved in neuron guidance and radial cell migration, whereas in the adult brain it is implicated in synaptic plasticity and neurotransmission [52]. It is located at a well replicated linkage peak for ASDs on chromosome 7q22 [53,54], and in addition to association between RELN and autism [55,56], association between RELN and schizophrenia and other neuropsychiatric disorders [52], has been reported before. Post-mortem studies of individuals with ASDs also indicate that RELN might have a role in the pathophysiology of autism [57,58].

Interestingly, in the same sample, no association of RELN was found in previous studies, in which only the end-state diagnosis was used. This could be due to the large overlap in the range of the symptom severity between patients and unaffected siblings. Our finding indicates that RELN is associated particularly to restricted and stereotyped behavior,

whether it takes place in affected individuals or their siblings. Thus, this is an interesting example where the use of quantitative phenotypes is extremely important, and further supports the demand for new analysis techniques in the research of complex neuropsychiatric disorders [59].

Many SNPs in RELN associated with the RSBC in the AGRE sample too, although none of them endured the Bonferroni correction (after correction p>0.05). However, it is not entirely unexpected that the association of RELN with RSBC was stronger in the Finnish sample, compared to AGRE sample. The Finnish population is isolated and genetically very homogeneous [27], while AGRE families have a heterogeneous background, as the families are from all over the United States, and it is only required that one of the parents is English-speaking [28]. The heterogeneity of the sample decreases the power of the association analyses and in addition, the genetic background of the disorder in the AGRE sample could be slightly different than that of the homogeneous Finnish sample. Indeed, in a recent review, autism was described as a behavioral manifestation of tens, or perhaps hundreds of genetic and genomic disorders [60]. Furthermore, the AGRE sample has a bias towards multiplex families [28], whereas the Finnish sample consists mostly of simplex families, and it has been reported that there are differences in the rates of copy number variations and autism-related traits in non-affected family members of simplex and multiplex families [28]. This gives further support to the possibility of slightly different genetic backgrounds of Finnish and AGRE samples. In addition, there were small differences in the content of RSBC between the two samples. In the AGRE data, the biggest correlation in RSBC was with the item compulsions/rituals, whereas in the Finnish sample, it was with the item hand and finger movements. Cuccaro et al. [61] have reported a study which suggests that these two items might belong to two different components, within restricted and repetitive behavior: repetitive sensory motor actions and resistance to change. All SNPs within RELN associated with RSBC (p<0.01), either in Finnish or AGRE sample, are shown in table 4.

Not much is known about the role of RELN in the neurobiological background of ASDs. Therefore, our findings indicating an association of RELN with restricted and stereotyped behavior, might prove important in planning future studies, both in molecular genetics and animal model studies of complex neuropsychiatric disorders. Age-dependent decrease in prepulse inhibition of startle has been reported in heterozygous reeler mice, a mouse model with an autosomal recessive mutation in RELN [62], and it has been demonstrated that adults with autism exhibit a decrease in prepulse inhibition, which is correlated with increased ratings of restricted and repetitive behavior [63].

Conclusions

We showed in this study that PCA of ADI-R algorithm items lead to relevant PCs, evaluated by correlation analysis, k-means clustering and reliability measurements. Especially in the Finnish sample, the unaffected individuals were clearly separated into one cluster by the k-means clustering. In addition, K-means clustering showed that siblings also had variation in the symptom severity. Especially in SC, the total range of the symptom severity was increased by siblings. The association of NLGN1 with SC was increased, compared to a previous study where only the end-state diagnosis was used. In RSBC, the range of the symptom severity of siblings overlapped greatly with that of patients, which could explain why no association of RELN was found in previous studies, in which only the end-state diagnosis was used, but a significant association of RELN with RSBC was now found in the Finnish families (Bonferroni-corrected p=0.029 for rs362644). Thus,

our study also shows that in some situations, it is extremely important to use quantitative phenotypes.

We also compared the findings in the isolated and homogeneous Finnish sample to a much larger, but also much more heterogeneous AGRE sample. The association of RELN with RSBC in the Finnish sample was stronger than in the AGRE sample, which could be explained by the heterogeneity of the AGRE sample and differences in the genetic backgrounds of the samples. In addition, the small differences in the content of the RSBC could also affect the results of the association analysis.

In future studies, the role of RELN in restricted and stereotyped behavior should be further studied, both in ASDs and other neuropsychiatric disorders like schizophrenia, Tourette syndrome and obsessive-compulsive disorder. In addition, also PCs with smaller eigenvalues can contain important phenotypic information, and should be investigated. Finally, the proposed approach should be used in broader genome-wide association analyses.

Acknowledgements

References

1. Fombonne E (2009) Epidemiology of pervasive developmental disorders. Pediatr Res 65: 591-598.

2. Mattila ML, Kielinen M, Linna SL, Jussila K, Ebeling H, et al. (2011) Autism spectrum disorders according to DSM-IV-TR and comparison with DSM-5 draft criteria: an epidemiological study. J Am Acad Child Adolesc Psychiatry 50: 583-592.

3. Mattila ML, Kielinen M, Jussila K, Linna SL, Bloigu R, et al. (2007) An epidemiological and diagnostic study of Asperger syndrome according to four sets of diagnostic criteria. J Am Acad Child Adolesc Psychiatry 46: 636-646.

4. Bolton P, Macdonald H, Pickles A, Rios P, Goode S, et al. (1994) A case-control family history study of autism. J Child Psychol Psychiatry 35: 877-900.

5. Bailey A, Le Couteur A, Gottesman I, Bolton P, Simonoff E, et al. (1995) Autism as a strongly genetic disorder: Evidence from a british twin study. Psychol Med 25: 63-77.

6. Folstein SE, Rosen-Sheidley B (2001) Genetics of autism: Complex aetiology for a heterogeneous disorder. Nature Rev Genet 2: 943-955.

7. Constantino JN (2011) The quantitative nature of autistic social impairment. Pediatr Res 69: 55R-62R.

8. Schopler E, Mesihov GM (1988) Diagnosis and assessment in autism. Springer, New York, USA.

9. Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, et al. (2007) A quantitative trait locus analysis of social responsiveness in multiplex autism families. Am J Psychiatry 164: 656-662.

10. Coon H, Villalobos ME, Robison RJ, Camp NJ, Cannon DS, et al. (2010) Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. Mol Autism 1: 8.

11. Lord C, Rutter M, Le Couteur A (1994) Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 24: 659-685.

12. Jolliffe IT (2002) Principal component analysis. (2nd edn), Springer, New York, USA.

13. Szatmari P, Georgiades S, Bryson S, Zwaigenbaum L, Roberts W, et al. (2006) Investigating the structure of the restricted, repetitive behaviours and interests domain of autism. J Child Psychol Psychiatry 47: 582-590.

14. Shao Y, Cuccaro ML, Hauser ER, Raiford KL, Menold MM, et al. (2003) Fine mapping of autistic disorder to chromosome 15q11-q13 by use of phenotypic subtypes. Am J Hum Genet 72: 539-548.

15. Tadevosyan-Leyfer O, Dowd M, Mankoski R, Winklosky B, Putnam S, et al. (2003) A principal components analysis of the autism diagnostic interview-revised. J Am Acad Child Psychiatry 42: 864-872.

16. Nurmi EL, Dowd M, Tadevosyan-Leyfer O, Haines JL, Folstein SE, et al. (2003) Exploratory subsetting of autism families based on savant skills improves evidence of genetic linkage to 15q11-q13. J Am Acad Child Adolesc Psychiatry 42: 856-863.

17. McCauley JL, Olson LM, Dowd M, Amin T, Steele A, et al. (2004) Linkage and association analysis at the serotonin transporter (SLC6A4) locus in a rigid-compulsive subset of autism. Am J Med Genet B Neuropsychiatr Genet 127B: 104-112.

18. Mulder EJ, Anderson GM, Kema IP, Brugman AM, Ketelaars CE, et al. (2005) Serotonin transporter intron 2 polymorphism associated with rigid-compulsive behaviors in Dutch individuals with pervasive developmental disorder. Am J Med Genet B Neuropsychiatr Genet 133B: 93-96.

19. Ma D, Jaworski J, Menold M, Donnelly S, Abramson R, et al. (2005) Ordered-subset analysis of savant skills in autism for 15q11-q13. Am J Med Genet B Neuropsychiatr Genet 135B: 38-41.

20. Sutcliffe JS, Delahanty RJ, Prasad HC, McCauley JL, Han Q, et al. (2005) Allelic heterogeneity at the serotonin transporter locus (SLC6A4) confers susceptibility to autism and rigid-compulsive behaviors. Am J Hum Genet 77: 265-79.

21. Hu VW, Steinberg ME (2009) Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. Autism Res 2: 67-77.

22. van Lang ND, Boomsma A, Sytema S, de Bildt AA, Kraijer DW, et al. (2006) Structural equation analysis of a hypothesised symptom model in the autism spectrum. J Child Psychol Psychiatry 47: 37-44.

23. Frazier TW, Youngstrom EA, Kubu CS, Sinclair L, Rezai A (2008) Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. J Autism Dev Disord 38: 474-80.

24. Lecavalier L, Aman MG, Scahill L, McDougle CJ, McCracken JT, et al. (2006) Validity of the autism diagnostic interview-revised. Am J Ment Retard 111: 199-215.

25. Seber GAF (1984) Multivariate observations. Hoboken, John Wiley & Sons, New Jersey, USA.

26. Geschwind DH, Sowinski J, Lord C, Iversen P, Shestack J, et al. (2001) The autism genetic resource exchange: A resource for the study of autism and related neuropsychiatric conditions. Am J Hum Genet 69: 463-466.

27. Peltonen L, Jalanko A, Varilo T (1999) Molecular genetics of the Finnish disease heritage. Hum Mol Genet 8: 1913-1923.

28. Lajonchere CM (2010) Changing the landscape of autism research: the autism genetic resource exchange. Neuron 68: 187-191.

29. http://research.agre.org/agrecatalog/algorithm.cfm

30. Rutter M, Le Couteur A, Lord C (2003) Autism diagnostic interview-revised. Western Psychological Services, Los Angeles, CA.

31. Autism Diagnostic Interview-Revised (ADI-R) Autism Genetic Resource Exchange.

32. Kaiser HF (1960) The application of electronic computers to factor analysis. Educ Psychol Meas 20: 141-151.

33. Nummenmaa L (2004) Behavioural statistical methods. In January, Helsinki, Finland.

34. Kolenikov S, Angeles G (2009) Socioeconomic status measurement with discrete proxy variables: is principal component analysis a reliable answer? Rev Income Wealth 55: 128-165.

35. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16: 297-334.

36. (2010) MATLAB version 7.11.0. The MathWorks Inc., Natick, Massachusetts.

37. Auranen M, Vanhala R, Varilo T, Ayers K, Kempas E, et al. (2002) A genomewide screen for autism-spectrum disorders: Evidence for a major susceptibility locus on chromosome 3q25-27. Am J Hum Genet 71: 777-790.

38. Ylisaukko-Oja T, Peyrard-Janvid M, Lindgren CM, Rehnstrom K, Vanhala R, et al. (2005) Family-based association study of DYX1C1 variants in autism. Eur J Hum Genet 13: 127-30.

39. Ylisaukko-Oja T, Rehnstrom K, Auranen M, Vanhala R, Alen R, et al. (2005) Analysis of four neuroligin genes as candidates for autism. Eur J Hum Genet 13: 1285-92.

40. Kilpinen H, Ylisaukko-Oja T, Hennah W, Palo OM, Varilo T, et al. (2008) Association of DISC1 with autism and Asperger syndrome. Mol Psychiatry 13: 187-96.

41. Rehnstrom K, Ylisaukko-Oja T, Vanhala R, von Wendt L, Peltonen L, et al. (2008) No association between common variants in glyoxalase 1 and autism spectrum disorders. Am J Med Genet B Neuropsychiatr Genet 147B: 124-127.

42. Pastinen T, Raitio M, Lindroos K, Tainola P, Peltonen L, et al. (2000) A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays. Genome Res 10: 1031-1042.

43. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, et al. (2008) Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med 358: 667-675.

44. Sung YJ, Dawson G, Munson J, Estes A, Schellenberg GD, et al. (2005) Genetic investigation of quantitative traits related to autism: use of multivariate polygenic models with ascertainment adjustment. Am J Hum Genet 76: 68-81.

45. Riva A, Kohane IS (2004) A SNP-centric database for the investigation of the human genome. BMC Bioinformatics 5: 33.

46. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.

47. Purcell S (2010) PLINK 0.99s.

48. Fulker DW, Cherny SS, Sham PC, Hewitt JK (1999) Combined linkage and association sib-pair analysis for quantitative traits. Am J Hum Genet 64: 259-267.

49. Abecasis GR, Cardon LR, Cookson WO (2000) A general test of association for quantitative traits in nuclear families. Am J Hum Genet 66: 279-292.

50. Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, et al. (2004) X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. Am J Hum Genet 74: 552-557.

51. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, et al. (2003) Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34: 27-29.

52. Fatemi SH (2005) Reelin glycoprotein: Structure, biology and roles in health and disease. Mol Psychiatry 10: 251-257.

53. Klauck SM (2006) Genetics of autism spectrum disorder. Eur J Hum Genet 14: 714-720.

54. Trikalinos TA, Karvouni A, Zintzaras E, Ylisaukko-oja T, Peltonen L, et al. (2006) A heterogeneity-based genome search meta-analysis for autism-spectrum disorders. Mol Psychiatry 11: 29-36.

55. Skaar DA, Shao Y, Haines JL, Stenger JE, Jaworski J, et al. (2005) Analysis of the RELN gene as a genetic risk factor for autism. Mol Psychiatry 10: 563-571.

56. Ashley-Koch AE, Jaworski J, Ma de Q, Mei H, Ritchie MD, et al. (2007) Investigation of potential gene-gene interactions between APOE and RELN contributing to autism risk. Psychiatr Genet 17: 221-226.

57. Fatemi SH, Snow AV, Stary JM, Araghi-Niknam M, Reutiman TJ, et al. (2005) Reelin signaling is impaired in autism. Biol Psychiatry 57: 777-787.

58. Fatemi SH, Stary JM, Halt AR, Realmuto GR (2001) Dysregulation of reelin and bcl-2 proteins in autistic cerebellum. J Autism Dev Disord 31: 529-535.

59. Randolph-Gips M, Srinivasan P (2012) Modeling autism: a systems biology approach. J Clin Bioinforma 2: 17.

60. Betancur C (2011) Etiological heterogeneity in autism spectrum disorders: More than 100 genetic and genomic disorders and still counting. Brain Res 1380: 42-77.

61. Cuccaro ML, Shao Y, Grubber J, Slifer M, Wolpert CM, et al. (2003) Factor analysis of restricted and repetitive behaviors in autism using the Autism Diagnostic Interview-R. Child Psychiatry Hum Dev 34: 3-17.

62. Tueting P, Costa E, Dwivedi Y, Guidotti A, Impagnatiello F, et al. (1999) The phenotypic characteristics of heterozygous reeler mouse. Neuroreport 10: 1329-1334.

63. Perry W, Minassian A, Lopez B, Maron L, Lincoln A (2007) Sensorimotor gating deficits in adults with autism. Biol Psychiatry 61: 482-486.