Principal Pathogenetic Components and Biological Endophenotypes in Autism Spectrum Disorders

Roberto Sacco, Paolo Curatolo, Barbara Manzi, Roberto Militerni, Carmela Bravaccio, Alessandro Frolli, Carlo Lenti, Monica Saccani, Maurizio Elia, Karl-Ludvig Reichelt, Tiziana Pascucci, Stefano Puglisi-Allegra, and Antonio M. Persico

Autism is a complex neurodevelopmental disorder, likely encompassing multiple pathogenetic components. The aim of this study is to begin identifying at least some of these components and to assess their association with biological endophenotypes. To address this issue, we recruited 245 Italian patients with idiopathic autism spectrum disorders and their first-degree relatives. Using a stepwise approach, patient and family history variables were analyzed using principal component analysis (''exploratory phase''), followed by intra- and inter-component cross-correlation analyses (''follow-up phase''), and by testing for association between each component and biological endophenotypes, namely head circumference, serotonin blood levels, and global urinary peptide excretion rates (''biological correlation phase''). Four independent components were identified, namely ''circadian & sensory dysfunction,'' ''immune dysfunction,'' ''neurodevelopmental delay,'' and ''stereotypic behavior,'' together representing 74.5% of phenotypic variance in our sample. Marker variables in the latter three components are positively associated with macrocephaly, global peptiduria, and serotonin blood levels, respectively. These four components point toward at least four processes associated with autism, namely (I) a disruption of the circadian cycle associated with behavioral and sensory abnormalities, (II) dysreactive immune processes, surprisingly linked both to prenatal obstetric complications and to excessive postnatal body growth rates, (III) a generalized developmental delay, and (IV) an abnormal neural circuitry underlying stereotypies and early social behaviors.

Keywords: autistic disorder; macrocephaly; neurodevelopment; pervasive developmental disorders; principal component analysis; serotonin

Introduction

Autism is a complex Pervasive Developmental Disorder (PDD) characterized by impaired language, communication, and social skills, as well as by repetitive and stereotypic patterns of behavior, appearing before 3 years of age [American Psychiatric Association, 1994]. Heritability estimates greater than 90% have spurred major efforts aimed at elucidating the genetic and neurobiological underpinnings of this disease [for reviews, see DiCicco-Bloom et al., 2006; Persico & Bourgeron, 2006]. Yet, our understanding of autism pathogenesis is far from satisfactory. Difficulties have arisen both from the clinical and from the genetic perspectives. On one hand, categorical diagnoses represent only an easy-to-use approximation to the real clinical complexity of ''autism spectrum disorder'' (ASD), a dimensional continuum which can range from minimal autistic traits to full-blown autism [Piven, Palmer, Jacobi, Childress, & Arndt, 1997]. An additional level of clinical complexity is contributed by co-morbidity with epilepsy and mental retardation, present in up to 30 and 65% of autistic patients, respectively [Fombonne, 2005; Tuchman & Rapin, 2002]. Likewise, an initially unexpected degree of genetic complexity stems from interindividual heterogeneity, rare disease-causing de novo mutations, common variants at numerous loci conferring vulnerability or protection, incomplete penetrance, phenocopies, genomic instability often accompanied by parental germline mosaicism, gene–gene and gene–environment interactions [for reviews, see Lintas & Persico, 2009; Persico & Bourgeron, 2006].

Additional Supporting Information may be found in the online version of this article.

From the Laboratory of Molecular Psychiatry and Neurogenetics, University ''Campus Bio-Medico'', Rome, Italy (R.S., A.M.P.); I.R.C.C.S. ''Fondazione Santa Lucia'', Rome, Italy (R.S., A.M.P.); Department of Child Neuropsychiatry, University of ''Tor Vergata'', Rome, Italy (P.C., B.M.); Department of Child Neuropsychiatry, II University of Naples, Naples, Italy (R.M., A.F.); Department of Pediatrics, University ''Federico II'', Naples, Italy (C.B.); Department of Child Neuropsychiatry, University of Milan, Milan, Italy (C.L., M.S.); Unit of Neurology and Clinical Neurophysiopathology, I.R.C.C.S. ''Oasi Maria S.S.'', Troina, EN, Italy (M.E.); Department of Pediatric Research, Rikshospitalet, University of Oslo, Oslo, Norway (K.L.R.); Department of Psychology, University ''La Sapienza'', Rome, Italy (T.P., S.P.A.); I.R.C.C.S. ''Fondazione Santa Lucia'', Rome, Italy (T.P., S.P.A.)

Received July 28, 2009; accepted for publication June 25, 2010

Address for correspondence and reprints: Antonio M. Persico, Laboratory of Molecular Psychiatry and Neurogenetics, University ''Campus Bio-Medico'', Via Alvaro del Portillo 21, I-00128 Rome, Italy. E-mail: a.persico@unicampus.it

Grant sponsor: Italian Ministry for University, Scientific Research and Technology; Grant number: PRIN n.2006058195; Grant sponsor: Italian Ministry of Health; Grant number: RFPS-2007-5-640174; Grant sponsor: Autism Speaks Foundation.
Published online in Wiley Online Library (wileyonlinelibrary.com)
DOI: 10.1002/aur.151

Despite the extraordinary challenge posed by this disease, different lines of evidence are starting to converge upon some basic features of the pathogenetic processes underlying autistic disorder. Developmental abnormalities later leading to autism begin in most cases prenatally, reducing apoptosis and/or boosting cell proliferation, and affecting cell migration, differentiation, and synaptogenesis [Bauman & Kemper, 2005; DiCicco-Bloom et al., 2006; Miller et al., 2005; Persico & Bourgeron, 2006]; postnatally, these abnormalities yield excessive head and body growth in a large subset of patients [Courchesne et al., 2007; Lainhart et al., 2006; Sacco et al., 2007]. Despite prominent involvement of the central nervous system, systemic signs and symptoms such as macrosomy, immune dysreactivity, and renal peptiduria are also present [Jyonouchi, Geng, Ruby, & Zimmerman-Bier, 2005; Lainhart et al., 2006; Reichelt, Knivsberg, Nodland, Stensrud, & Reichelt, 1997; Sacco et al., 2007]. In particular, contributions by the immune system have received strong support by in vivo and postmortem findings [Garbett et al., 2008; Vargas, Nascimbene, Krishnan, Zimmerman, & Pardo, 2005]. Unfortunately, immune abnormalities are not simple to connect with neuropathological findings and behavioral symptoms.

The present study was designed to further delineate the complex pathogenetic processes underlying autistic disorder. To this aim, we have implemented a multi-step statistical approach, whose outcome supports the existence of at least four principal components potentially contributing to autism pathogenesis, hereby named ''circadian and sensory dysfunction,'' ''immune dysfunction,'' ''neurodevelopmental delay,'' and ''stereotypic behavior.'' These components together represent 74.5% of phenotypic variance in our sample. Marker variables for the latter three components are significantly associated with head size, global peptiduria, and 5-HT blood levels, respectively.

Methods and Materials

Subjects

A total of 245 patients with idiopathic, non-syndromic ASD were consecutively recruited in Italy between 1998 and 2007. Their demographic and clinical characteristic are summarized in Table I. Compared with clinical samples assessed in studies employing similar approaches [for review see Happé & Ronald, 2008], this sample overlaps in distribution of DSM-IV diagnoses, while displaying higher M:F ratio and mental retardation rates. Syndromic autism was excluded by magnetic resonance imaging, electroencephalogram (EEG), audiometry, urinary amino acid and organic acid measurements, cytogenetic and fragile X testing. Patients with evident dysmorphic features were excluded even in the absence of detectable cytogenetic abnormalities. Patients with sporadic seizures (i.e., <1 every 6 months) were included ($N = 16/245$, 6.5%); patients with frequent seizures or focal neurological deficits possibly suggestive of undiagnosed syndromic forms were excluded. All parents gave written informed consent for themselves and for their children. The consent form was approved by the Institutional Review Board of University ''Campus Bio-Medico'' (Rome, Italy).

Table I. Participant Characteristics

	Mean/Median	SD	Range
Mean age (yr)	8.82	5.62	2–30
Median VABS scores ($N = 129$)			
Communication	71.0		19–104
Daily living skills	71.0		14–109
Socialization	66.0		30–103
Motor skills	81.0		25–128
Composite	65.0		19–115
	N	Percent	
Gender			
Male	216	88.2	
Female	29	11.8	
M/F ratio	7.4:1		
Family type			
Simplex	234	95.5	
Multiplex	11	4.5	
DSM-IV diagnosis			
Autistic disorder	210	85.7	
Asperger's disorder	14	5.7	
PDD-NOS	21	8.6	
IQ ($N = 75$)			
>70	11	14.7	
≤ 70	64	85.3	

$N = 245$, unless otherwise specified.

Measures

Our entire sample was characterized using a questionnaire designed by our group, encompassing 36 developmental, clinical, and family history variables, as previously reported [Sacco et al., 2007]. The English version of this questionnaire is available in the Supplemental Methods (item no. 1), variables are listed in Supplementary Table S1, responses to each variable are presented in Table II. The psychometric properties of this questionnaire were not rigorously validated, because it does not represent a diagnostic tool. It is simply a list of variables describing clinical features, as well as patient and family history. These variables were selected, based on the literature available in 1997, on clinical observation, and on parental reports. Demographic information and physical parameters, including head circumference, weight, and height were also recorded. Autistic behaviors, adaptive functioning, and IQ were assessed in subsets of

Table II. Outcome of the Questionnaire Employed in this Study

(A) Variables based on parental report to the clinician at intake				
(1) Age at non-verbal language development, in months ($N = 90$)	21.5 ± 14.6 (7–84)			
(2) Age at verbal language development, in months ($N = 119$)	27.1 ± 16.9 (8–96)			
(3) Level of verbal language development ($N = 199$)	Normal: 36 (18.1%)	Language delay: 79 (39.7%)	Regression: 7 (3.5%)	Never acquired: 77 (38.7%)
(4) Age at walking, in months ($N = 211$)	15.6 ± 4.5 (8–36)			
(5) Age at acquisition of bladder control at night, in months ($N = 111$)	40.3 ± 13.8 (18–96)			
(6) Age at first social smile, in months ($N = 123$)	4.7 ± 5.9 (1–60)			
(8) Sleep disorder at intake ($N = 196$)	NO: 105 (53.6%)		YES: 91 (46.4%)	
(10) History of allergies in the patient, including any form of asthma ($N = 225$)	NO: 154 (68.4%)		YES: 71 (31.6%)	
(11) Employment of a specific diet ($N = 231$)	NO: 201 (87.0%)		YES: 30 (13.0%)	
(12) History of regression ($N = 224$)	NO: 133 (59.4%)		YES: 91 (40.6%)	
(13) History of regression following vaccination ($N = 219$)	NO: 210 (95.9%)		YES: 9 (4.1%)	
(14) Prior or current presence of autoimmune disorders in the patient ($N = 102$; a list of disorders was provided to parents)	NO: 100 (98.0%)		YES: 2 (2.0%)	
(15) History of gastroenteritis at autism onset ($N = 225$)	NO: 202 (89.8%)		YES: 23 (10.2%)	
(16) History of chronic or recurrent otitis at autism onset ($N = 205$)	NO: 185 (90.2%)		YES: 20 (9.8%)	
(17) History of other infectious diseases at autism onset ($N = 189$)	NO: 178 (94.2%)		YES: 11 (5.8%)	
(18) History of obstetric complications affecting the patient at or immediately after birth[a] ($N = 228$)	NO: 171 (75.0%)		YES, minor: 43 (18.9%) YES, major: 14 (6.1%)	
(19) History of any complication affecting the mother or the foeto-maternal unit during pregnancy ($N = 225$)	NO: 148 (65.8%)		YES: 77 (34.2%)	
(20) History of recurrent spontaneous abortions in the mother ($N = 97$)	NO: 73 (75.3%)		YES: 24 (24.7%)	
(21) Pregnancy duration[b] ($N = 225$)	At term: 199 (88.4%)	Pre-term: 24 (10.7%)	Post-term: 2 (0.9%)	
(22) Patient weight at birth, in kg ($N = 225$)	3.33 ± 0.53 (1.85–4.75)			
(24) History of seizures ($N = 228$)	NO: 212 (93.0%)	YES, <1 every 6 months: 16 (7.0%)		
(29) Pain sensitivity ($N = 98$)	Normal: 62 (63.3%)	Decreased: 36 (36.7%)	Increased: 0	
(31) Dermatological allergies in first-degree relatives ($N = 161$)	NO: 134 (83.2%)		YES: 27 (16.8%)	
(32) Food allergies in first-degree relatives ($N = 161$)	NO: 143 (88.8%)		YES: 18 (11.2%)	
(33) Other allergies in first-degree relatives ($N = 135$)	NO: 93 (68.9%)		YES: 42 (31.1%)	
(34) Autoimmune disease in first-degree relatives ($N = 104$)	NO: 75 (72.1%)		YES: 29 (27.9%)	
(35) Autoimmune disease in second-degree relatives ($N = 92$)	NO: 85 (92.4%)		YES: 7 (7.6%)	
(36) Recurrent otitis in first-degree relatives ($N = 95$)	NO: 76 (80.0%)		YES: 19 (20.0%)	
(B) Variables either observed or assessed directly by the clinician upon patient intake				
(7) Abnormal lip conformation when kissing ($N = 43$)	NO: 33 (76.7%)		YES: 10 (23.3%)	
(9) Hyperactivity at intake ($N = 192$)	NO: 94 (49.0%)		YES: 98 (51.0%)	
(23) EEG pattern[c] ($N = 221$)	Normal: 149 (67.4%)	Dysregulated: 47 (21.3%)	Paroxysmal: 25 (11.3%)	

Table II. Continued

(25) Presence of muscle hypotonia at intake ($N = 206$)	NO: 185 (89.8%)		YES: 21 (10.2%)
(26) Motor stereotypies at intake ($N = 231$)	NO: 64 (27.7%)		YES: 167 (72.3%)
(27) Verbal or vocal stereotypies at intake ($N = 219$)	NO: 116 (53.0%)		YES: 103 (47.0%)
(28) Self-aggressive or self-injurious behavior at intake ($N = 99$)	NO: 65 (65.7%)		YES: 34 (34.3%)
(30) Presence of congenital anomalies ($N = 194$)	NO: 183 (94.3%)	Minor: 11 (5.7%)	Major: 0
(C) Variables obtained combining partly overlapping questionnaire items for the principal component analysis			
15–17: History of any infectious disease at autism onset ($N = 229$)	NO: 186 (81.2%)		YES: 43 (18.8%)
19–20: History of obstetric complications or recurrent spontaneous abortions in the mother ($N = 225$)	NO: 137 (60.9%)	Obs compl. only: 64 (28.4%)	Rec. abortions + obs compl.: 24 (10.7%)
31–36: History of any allergic and/or immune disease in the family ($N = 163$)	NO: 105 (64.4%)		YES: 58 (35.6%)

Variables are listed by questionnaire item number and distinguished between items (A) based exclusively on parental report and (B) observed or assessed directly by the clinician at intake. For the principal component analysis, 22 items highlighted in bold were selected (for IQ, see Table I), while partly overlapping items highlighted in italics were combined into the 3 variables listed under (C). Sample sizes refer to valid cases; whenever largely below $N = 245$, either the function under scrutiny was achieved by a smaller subset of patients or "don't know" answers were frequent and the item was dropped from the analysis (see *Methods*). Scores for categorical variables are ordered from least-to-most severe. Quantitative variables are reported as mean ± SD (range); categorical variables are reported as *N* (%).

[a]"Minor" obstetric complications = not requiring hospitalization and/or inpatient neonatological treatment.; "major" complications, requiring hospitalization and/or inpatient neonatological treatment.

[b]Pregnancy duration: 38–42 wks = at term; < 38 wks = pre-term; > 42 wks = post-term.

[c]Dysregulated = presence of slow waves, but no spikes; paroxysmal = presence of spikes.

patients using, respectively: the official Italian versions of the Autism Diagnostic Interview—Revised (ADI-R) [Rutter, Le Couter, & Lord, 2003], and the Autism Diagnostic Observation Schedule (ADOS) [Lord, Rutter, DiLavore, & Risi, 2002], available since 2005 (ADOS and/or ADI-R, $N = 95$ patients); the Vineland Adaptive Behavior Scales (VABS) [Sparrow, Balla, & Cicchetti, 1984] ($N = 129$ patients); either the Griffith Mental Developmental Scales, the Colored Raven Matrices, the Bayley Developmental Scales or the Leiter International Performance Scale, depending on the age of the patient and on the recruiting site, dichotomizing patients into two categories, either IQ > 70 or ≤ 70 ($N = 75$ patients). Importantly, ADOS and ADI-R cut-off scores for ASD were exceeded by 92.9 and 95.6% of patients enrolled here with a clinical DSM-IV diagnosis of an ASD, confirming the diagnostic reliability of our recruitment, which was carried out by the same clinicians both before and after 2005.

Three biological endophenotypes were also assessed. Head circumference was measured in all 231 ASD patients aged < 16 years old, and transformed into percentiles using the sex- and age-specific standard tables currently adopted in the vast majority of European countries and by the Italian Pediatric Association, as described [Sacco et al., 2007]. Serotonin blood levels were measured in the platelet-rich plasma of 138 ASD patients aged < 11 years old, as described [Piven et al., 1991], and are expressed in ng/ml. Analyses were restricted to pre-puberal children not undergoing treatment with selective serotonin reuptake inhibitors (SSRIs), to prevent the confounding effects of sex hormones and of drugs interfering with serotonin uptake into platelets [McBride et al., 1998; Persico et al., 2002]. Statistical adjustments by age, gender, and medication would be unreliable in this situation, because the extent of blood 5-HT decrease recorded after puberty or under SSRI treatment is not linearly related to these variables, but more likely depends on pharmacogenetic parameters, as well as on sex hormone blood levels and cellular responsiveness. Global amounts of oligopeptides present in the first morning urines were measured by HPLC in 151 ASD patients, by calculating the total area under the 215 nm absorption curve (AUC) in the region following the hippuric acid peak, as described [Reichelt, Ek, Stensrud, & Reichelt, 1998]. Only the global AUC expressed in μm^2, and not single peptide peaks, were used in the present study as an endophenotypic measure. A detailed protocol is provided in the Supplemental Methods (item no.2), together with a description of internal and external controls, as well as two exemplificative chromatograms (Supplementary Figure S1A and B). One outlier with 5-HT blood level at 1927.2 ng/ml (i.e., > 3 SD), confirmed by two independent HPLC measurements, was excluded

from further analysis. The clinical and demographic characteristics of the final subsets of 138 and 151 ASD patients analyzed for 5-HT blood levels and global peptiduria, respectively, are presented in Supplementary Table S2, and do not largely differ from the characteristics of the total sample presented in Table II.

Statistical Analyses

Patient and family history variables were analyzed implementing a step-wise approach, including: (1) an exploratory principal component analysis, (2) a follow-up inter- and intra-component cross-correlation analysis, and (3) an association analysis between component variables and three biological endophenotypes. Principal component analysis represents the statistical multivariate strategy of choice when aiming to reduce a large number of variables down to a small set of components [Tabachnick & Fidell, 1989]. By its very nature, principal component analysis is not an hypothesis-testing technique requiring control for multiple testing; it is an exploratory and descriptive approach, yielding stable and reliable models if a minimum of approximately ten subjects per variable is analyzed [Tabachnick & Fidell, 1989]. In order to satisfy this criterion, the number of variables was reduced from 37 (36 questionnaire items, plus presence/absence of mental retardation based on IQ) down to 22 by: (a) dropping items no. 7, 11, 13, 14, 22, 24, and 30 (see Table II), due to low informativeness (i.e., either $N<80$ cases, or minor class frequency $<10\%$, and/or factor scores consistently <0.1 for all principal components, regardless of data extraction and rotation technique); (b) combining items no. 15–17, 19–20, and 31–36 into three single items ''History of any infectious disease at autism onset,'' ''History of obstetric complications and/or recurrent spontaneous abortions in the mother,'' and ''History of any allergic and/or immune disease in the family,'' respectively (Table II), eliminating multicollinearity and singularity; and (c) dichotomizing answers to items no. 18 and 21. Our final data set thus included 22 quantitative variables (18 items from the questionnaire, 3 multi-item variables, and 1 item based on IQ score), with categorical and quantitative variables scored on an ordinal and interval scale, respectively. Five variables (items no. 1, 2, 4, 5, and 6) were log-transformed to achieve normality [Tabachnick & Fidell, 1989]. At this point, four different factor extraction techniques were tested (principal components, unweighted least squares, image factoring, and maximum likelihood factoring, followed by either oblimin or varimax rotations), as requested by this analytical approach. We are hereby reporting the outcome of a principal component analysis performed selecting principal components based on a threshold eigenvalue >1.7, followed by oblimin rotation with Kaiser normalization [Tabachnick & Fidell, 1989]. This threshold eigenvalue was chosen following examination of the scree plot shown in Supplementary Figure S2, whereas this overall design was chosen to maximize the variance explained by a small number of components, displaying relatively high intra-component and low inter-component correlations (see *Results*). As a term of comparison, a threshold eigenvalue >1 produced as many as nine principal components, altogether able to explain 59% of the variance compared to 85.7% of the variance explained by the five components identified in our analysis (see *Results*). The outcome of the principal component analysis presented here is practically superimposable to the model obtained applying unweighted least squares (Supplementary Table S3A), whereas image factoring and maximum likelihood factoring produced 4- and 6-component models which explain significantly lower percentages of variance and appear less biologically plausible (Supplementary Table S3B and S3C).

Follow-up analyses included inter-component cross-correlations using Spearman's ρ statistics to test for independence among putative principal components, and intra-component cross-correlations among variables with oblimin-rotated factor loadings >0.4 (negative factor loadings typically represent minimum levels of correlation and values distributed around zero represent intermediate, non-significant levels of correlation). Inter- and intra-component cross-correlations were performed using non-parametric Kendall's τ statistics. Principal components were confirmed if at least 50% of their variables with oblimin-rotated factor-loadings >0.4 were significantly correlated with each other and were not included into components explaining larger percentages of variance. Within each component, variables were retained if they were significantly correlated with at least one other variable.

Finally, the association between principal components and morphometric/biochemical endophenotypes (head circumference, 5-HT blood levels, global peptiduria) was used to begin verifying the existence of independent biological underpinnings for each component. Factor analytical techniques provided no statistically stable and scientifically meaningful model when incorporating three additional biological variables into our original 22-variable phenotypic data set, due to sample size limitations. Hence, factor scores for each component were obtained using the regression method (the Bartlett and the Anderson–Rubin methods gave superimposable results), and morphometric/biochemical endophenotypes were correlated with these cumulative factor scores and with single variables present in each component, especially with ''marker variables'' displaying the largest intra-component cross-correlation coefficients [Tabachnick & Fidell, 1989]. Correlation analyses were performed using non-parametric Kendall's τ statistics;

quantitative analyses were performed by Mann–Whitney U test and/or Kruskal–Wallis non-parametric ANOVA. Endophenotypic data are reported as median percentile ± semi-interquartile range (IQR) for head circumference, and as mean ± standard error of the mean (SEM) for 5-HT blood levels and peptiduria. Statistical significance is set at a nominal two-tail $P < 0.05$, unless otherwise specified. Statistical analyses were performed using SPSS software release 17.0 (SPSS INC, Chicago, IL).

Results

The outcome of all questionnaire items is summarized in Table II. A principal component analysis performed on these data, as described above, identifies five putative principal components (Table III). This five-component model represents 85.7% of the variance, with approximately 60% of the variance explained by the first three components (Table III). Inter-component correlation analyses unveil a lack of correlation among these five putative principal components, with ρ values ranging from −0.139 to +0.066 (mean ρ = −0.0276) (Supplementary Table S4). Intra-component correlation analyses confirm the existence of components 1–4 (Table IV). Putative component 5 does not qualify, because less than 50% of its variables are significantly correlated with each other; furthermore, the correlation between hyperactivity and reduced pain sensitivity is already included in component 1, which explains a much larger percentage of variance (Table IV). Also a limited number of variables, namely muscle hypotonia and level of verbal language development in components 2 and 4, respectively, are not correlated with any other intra-component variable and most likely represent false-positives (Table IV). These variables are dropped and four principal components are conclusively defined at the end of follow-up analyses, as summarized in Table V. The substructure found in component 1 (see Table IV) is recognized by distinguishing within this component two subcomponents, encompassing: (A) sleep disorders and self-injurious behaviors, on one hand, and (B) hyperactivity, decreased pain sensitivity, delayed non-verbal language development and abnormal level of language development, on the other hand (Table V). For simplicity, the four components are labelled "circadian and sensory dysfunction," "immune dysfunction," "neurodevelopmental delay," and "stereotypic behavior" (Table V). "Age at non-verbal

Table III. Principal Component Analysis

	Component				
	1: Circadian and sensory dysfunction	2: Immune dysfunction	3: Neuro-developmental delay	4: Stereotypic behavior	5
Sleep disorders	**0.891**	−0.092	−0.087	−0.270	−0.204
Self-injurious behaviors	**0.891**	−0.092	−0.087	−0.270	−0.204
Hyperactivity	**0.867**	0.334	0.046	0.017	**0.482**
Decreased pain sensitivity	**0.865**	−0.102	−0.159	−0.161	**0.438**
Age at non-verbal language development	**0.685**	−0.068	**0.508**	0.369	−0.181
Level of verbal language development	**0.572**	0.019	**0.546**	**0.422**	−0.252
Pregnancy duration	−0.121	**0.943**	−0.028	0.236	−0.053
Muscle hypotonia	0.073	**0.682**	0.310	0.268	0.105
History of any infectious disease at autism onset	−0.234	**0.565**	−0.348	0.045	−0.174
Obstetric complications or recurrent spontaneous abortions in the mother	−0.469	**0.541**	0.383	−0.140	−0.214
History of allergies in the patient	0.291	**0.490**	0.326	−0.386	−0.406
History of any allergic and/or immune disease in the family	−0.396	**0.435**	0.166	−0.583	**0.477**
Age at verbal language development	−0.231	−0.117	**0.983**	−0.035	0.028
Age at acquisition of bladder control	0.027	0.159	**0.914**	0.138	0.004
Age at walking	−0.313	−0.415	**0.448**	−0.139	**0.666**
Verbal or vocal stereotypies	−0.120	0.042	0.266	**0.875**	0.151
Age at first social smile	−0.225	0.075	−0.202	**0.839**	0.121
Motor stereotypies	−0.112	0.160	−0.574	**0.652**	−0.011
EEG pattern	0.119	−0.032	−0.123	0.116	**0.928**
Obstetric complications in the patient	0.119	−0.032	−0.123	0.116	**0.928**
Mental retardation	−0.189	−0.873	0.103	0.287	0.018
History of regression	−0.007	−0.049	−0.536	0.133	0.126
Percent of variance	23.2%	22.1%	15.2%	13.9%	11.3%
Cumulative percent of variance	23.2%	45.3%	60.5%	74.4%	85.7%

Factor loadings greater than 0.4 are highlighted in bold.

Table IV. Intra-Component Correlations

Component 1: circadian and sensory dysfunction					
Level of verbal language development	$\tau = 0.466$, $P = 4.5 \times 10^{-9}$, $N = 103$				
Self-injurious behaviors	$\tau = 0.127$, $P = 0.229$, $N = 69$	$\tau = 0.135$, $P = 0.130$, $N = 110$			
Sleep disorders	$\tau = -0.060$, $P = 0.489$, $N = 101$	$\tau = 0.113$, $P = 0.098$, $N = 191$	$\tau = 0.253$, $P = 0.009$, $N = 107$		
Reduced pain sensitivity	$\tau = 0.164$, $P = 0.117$, $N = 70$	$\tau = 0.114$, $P = 0.199$, $N = 111$	$\tau = 0.177$, $P = 0.081$, $N = 98$	$\tau = 0.129$, $P = 0.177$, $N = 110$	
Hyperactivity	$\tau = 0.250$, $P = 0.004$, $N = 102$	$\tau = 0.170$, $P = 0.011$, $N = 198$	$\tau = 0.096$, $P = 0.319$, $N = 108$	$\tau = 0.096$, $P = 0.189$, $N = 190$	$\tau = 0.438$, $P = 5.4 \times 10^{-6}$, $N = 109$
	Age at non-verbal language development	Level of verbal language development	Self-injurious behaviors	Sleep disorders	Reduced pain sensitivity
Component 2: immune dysfunction					
Any allergic and/or immune disease in the family	$\tau = 0.329$, $P = 1.3 \times 10^{-5}$, $N = 177$				
History of any infectious disease at autism onset	$\tau = 0.184$, $P = 0.004$, $N = 247$	$\tau = 0.159$, $P = 0.034$, $N = 179$			
Obstetric complications or recurrent abortions in the mother	$\tau = 0.199$, $P = 0.001$, $N = 242$	$\tau = 0.176$, $P = 0.014$, $N = 180$	$\tau = 0.236$, $P = 1.3 \times 10^{-4}$, $N = 246$		
Pregnancy duration	$\tau = 0.040$, $P = 0.535$, $N = 242$	$\tau = 0.109$, $P = 0.143$, $N = 180$	$\tau = 0.059$, $P = 0.355$, $N = 246$	$\tau = 0.183$, $P = 0.003$, $N = 247$	
Muscle hypotonia	$\tau = 0.061$, $P = 0.368$, $N = 220$	$\tau = -0.040$, $P = 0.611$, $N = 160$	$\tau = 0.028$, $P = 0.679$, $N = 225$	$\tau = 0.017$, $P = 0.787$, $N = 224$	$\tau = 0.015$, $P = 0.826$, $N = 224$
	History of allergies in the patient	Any allergic and/or immune disease in the family	History of any infectious disease at autism onset	Obstetric complications or recurrent abortions in the mother	Pregnancy duration
Component 3: neurodevelopmental delay					
Age at non-verbal language development	$\tau = 0.348$, $P = 2 \times 10^{-4}$, $N = 65$				
Level of verbal language development	$\tau = 0.372$, $P = 8.7 \times 10^{-8}$, $N = 142$	$\tau = 0.466$, $P = 4.5 \times 10^{-9}$, $N = 103$			

Table IV. Continued

Age at walking	$\tau = 0.104$, $P = 0.097$, $N = 138$	$\tau = 0.121$, $P = 0.099$, $N = 103$	$\tau = 0.114$, $P = 0.036$, $N = 218$	
Age at acquisition of bladder control at night	$\tau = 0.151$, $P = 0.066$, $N = 86$	$\tau = 0.216$, $P = 0.019$, $N = 70$	$\tau = 0.141$, $P = 0.059$, $N = 122$	$\tau = 0.167$, $P = 0.013$, $N = 123$
	Age at verbal language development	Age at non-verbal language development	Level of verbal language development	Age at walking
Component 4: stereotypic behavior				
Verbal or vocal stereotypies	$\tau = 0.280$, $P = 3 \times 10^{-4}$, $N = 136$			
Motor stereotypies	$\tau = 0.207$, $P = 0.007$, $N = 137$	$\tau = 0.371$, $P = 9.6 \times 10^{-9}$, $N = 240$		
Level of verbal language development	$\tau = 0.133$, $P = 0.067$, $N = 137$	$\tau = -0.042$, $P = 0.523$, $N = 202$	$\tau = 0.090$, $P = 0.166$, $N = 211$	
	Age at first social smile	Verbal or vocal stereotypies	Motor stereotypies	
Component 5				
Obstetric complications in the patient	$\tau = 0.010$, $P = 0.875$, $N = 236$			
Age at walking	$\tau = 0.024$, $P = 0.668$, $N = 223$	$\tau = 0.047$, $P = 0.413$, $N = 226$		
Hyperactivity	$\tau = -0.083$, $P = 0.227$, $N = 201$	$\tau = 0.030$, $P = 0.662$, $N = 210$	$\tau = -0.016$, $P = 0.790$, $N = 205$	
Any allergic and/or immune disease in the family	$\tau = -0.128$, $P = 0.088$, $N = 167$	$\tau = -0.021$, $P = 0.777$, $N = 179$	$\tau = -0.143$, $P = 0.029$, $N = 171$	$\tau = 0.080$, $P = 0.318$, $N = 156$
Reduced pain sensitivity	$\tau = 0.011$, $P = 0.908$, $N = 101$	$\tau = 0.085$, $P = 0.373$, $N = 110$	$\tau = -0.098$, $P = 0.229$, $N = 111$	$\tau = 0.438$, $P = 5.4 \times 10^{-6}$, $N = 109$	$\tau = 0.219$, $P = 0.023$, $N = 109$
	EEG pattern	Obstetric complications in the patient	Age at walking	Hyperactivity	Any allergic and/or immune disease in the family

Only variables yielding factor loadings > 0.4 in each putative principal component were included in this analysis. Kendall's τ *P*-values and sample sizes (*N*) are reported (missing data correspond to "don't know" answers); dark, middle- and light grey highlight nominal *P*-values $< .001$, $< .01$, and < 0.05, respectively.

Table V. Pathogenetic Component Structure of Autism Spectrum Disorders

1: Circadian and sensory dysfunction	2: Immune dysfunction	3: Neurodevelopmental delay	4: Stereotypic behavior
Sleep disorders	*History of allergies in the patient*	*Level of verbal language development*	*Verbal and vocal stereotypies*
Self-injurious behaviors	History of any allergic and/or immune disease in the family	Age at verbal language development	Motor stereotypies
- - - - - - - - - - -			
Hyperactivity	History of any infectious disease at autism onset	Age at non-verbal language development	Age at first social smile
Decreased pain sensitivity	Obstetric complications or recurrent abortions in the mother	Age at walking	
Age at non-verbal language development	Pregnancy duration	Age at acquisition of bladder control at night	
Level of verbal language development			

The hyphenated line distinguishes subcomponents 1A and 1B (see *Results*). Marker variables are highlighted in italics.

language development'' and ''level of verbal language development'' represent the only complex variables, comprised in components 1 and 3, while all remaining variables load onto a single component (Table V), confirming that these four components are largely independent from each other (Supplementary Table S4). A follow-up principal component analysis excluding also the five variables dropped after the intra-component cross-correlation step (EEG abnormalities, muscle hypotonia, obstetric complications in the patient, mental retardation, history of regression) yields a five-component structure, explaining 47.1% of the total variance. This component structure especially replicates the ''neurodevelopmental delay,'' ''immune dysfunction,'' and ''stereotypic behavior'' components, in accordance with the more complex substructure of the ''circadian and sensory dysfunction'' component (Supplementary Table S5).

To begin dissecting the biological underpinnings of this component model, we measured three morphological/biochemical endophenotypes, namely cranial circumference ($N = 231$, 75 ± 23.8 percentile), serotonin blood levels ($N = 138$, 381.6 ± 24.6 ng/ml) and global peptiduria ($N = 151$, $291.7 \pm 13.9\ \mu m^2$). The distributions of these parameters in autistic patients are depicted in Supplementary Figures S3, S4, and S5. Familiality was demonstrated for all these endophenotypes in our own sample, as head size, 5-HT blood levels, and peptiduria co-vary in first-degree relatives, depending on the level of these parameters recorded in the autistic proband (Table VI). Despite some complexity, the overall patterns of correlation between principal components and the three endophenotypes appear clearly non-random (Table VII). The highest discriminatory power is evident in components 3 and 4. The ''neurodevelopmental delay'' component displays significant positive correlation with peptiduria ($P = 0.014$), and negative correlation with 5-HT blood levels ($P = 0.04$). Importantly, single variables that were significantly correlated exhibited the same relationships with the endophenotypes as the whole component (Table VII). Instead, correlations with items belonging to the ''stereotypic behavior'' component display exactly the opposite trend in reference to peptiduria, and 5-HT blood levels, reaching overall P values of 0.052 and 0.063, respectively (Table VII). Correlation patterns are less consistent for the ''circadian and sensory dysfunction'' and the ''immune dysfunction'' components. The former is negatively correlated with 5-HT blood levels, but this result is essentially driven by the strong negative correlation of ''presence/absence of self-injury'' with serotoninemia. The latter component displays no significant overall correlation with any endophenotype. Two single variables belonging to the ''immune dysfunction'' component, namely the marker variable ''history of allergies in the patient'' and ''immune and/or allergic disease in the family,'' are significantly correlated with head circumference and with 5-HT blood levels, respectively (Table VII). The above-mentioned significant correlations for components 2–4 can be better visualized using quantitative analyses (Fig. 1 and Table S6). Interestingly, 15 patients undergoing a gluten- and/or casein-free diet (i.e., ''yes'' responses in Table II to item no.11 ''Employment of a special diet,'' which was dropped from the principal component analysis due to low informativeness) display significantly higher, and not lower, global peptiduria compared to 152 ASD patients undergoing no special diet (399.13 ± 48.75 vs. 301.79 ± 13.67, $U = 753.5$, $P < 0.05$). However, diets do not explain elevated peptiduria in children who never acquired

Table VI. Familiality of Biological Endophenotypes

	Biological endophenotypes measured in the autistic proband							
	Head circumference (percentile categories)			5-HT blood levels		Global peptiduria		
Proband class	**1–25**	**26–97**	**> 97**	**≤ 279**	**> 279**	**< 387**	**≥ 387**
Mothers ($N = 236$)	–	–	–	172.0 ± 12.4 ($N = 97$)	340.1 ± 20.1 ($N = 110$)	252.0 ± 13.0 ($N = 139$)	293.3 ± 17.4 ($N = 46$)
Fathers ($N = 220$)	–	–	–	179.9 ± 11.5 ($N = 89$)	388.4 ± 25.1 ($N = 99$)	197.3 ± 9.2 ($N = 125$)	274.2 ± 25.8 ($N = 39$)
Unaffected siblings ($N = 94$)	31.2 ± 41.9 ($N = 6$)	68.7 ± 20 ($N = 9$)	82.5 ± 14.4 ($N = 9$)	165.5 ± 24.7 ($N = 23$)	495.7 ± 33.1 ($N = 42$)	181.6 ± 12.0 ($N = 46$)	295.9 ± 41.0 ($N = 14$)

Descriptive statistics of head circumference, 5-HT blood levels, and global peptiduria in 550 first-degree relatives of the 245 autistic probands assessed in this study. First-degree relatives are divided into classes, which refer to parameters as recorded in their autistic family member; cut-offs for 5-HT blood levels and global peptiduria represent the upper confidence interval of the normal distribution, as reported by McBride et al. [1998], and Reichelt et al. [1998], respectively. Data are presented as median percentile ± semi IQR for head circumference, and mean ± SEM for 5-HT blood levels and global peptiduria.

spoken language (Supplementary Table S5), as only 13.5% of these children were dieting, vs. 18.3% of the remaining sample.

Restricting our entire stepwise statistical procedure either to patients satisfying DSM-IV diagnostic criteria for ''autistic disorder'' ($N = 210$, Table I), or to ASD patients younger than 18 years old ($N = 226$) yields results entirely overlapping with those summarized above (data not shown).

Discussion

This study outlines possible contributions to autism by at least four underlying pathophysiological components, here defined ''circadian and sensory dysfunction,'' ''immune dysfunction,'' neurodevelopmental delay,'' and ''stereotypic behavior.'' It further unveils the existence of correlations between at least some of these components and three biological endophenotypes, namely macrocephaly, peptiduria, and hyperserotoninemia, respectively. While the strength of these correlations is influenced by statistical power and stochastic factors, the consistency of their direction in the ''neurodevelopmental delay,'' and ''stereotypic behavior'' components is surprising, as they involve apparently distant clinical and patient-history variables. Despite several methodological limitations analyzed below, these results provide novel and unpredicted pieces of information, each bearing significant heuristic potential: (a) altered circadian rhythms and abnormal sensory processing could be pathophysiologically connected, albeit in a complex fashion (*component 1*); (b) converging lines of evidence, including but not limited to the present data set, support a dysreactive immune process or an autoimmune component in autism pathogenesis, which is likely shared by patients and first-degree relatives [Saresella et al., 2009] (*component 2*). This dysreactive immune process appears surprisingly correlated here with the obstetric complications frequently reported by mothers of ASD patients [Kolevzon, Gross, & Reichenberg, 2007]; (c) multiple neurodevelopmental milestones, including verbal and non-verbal language, walking and sphincter control, tend to co-vary and are often collectively delayed or hampered (*component 3*). This multifunctional developmental delay is significantly correlated with higher global amounts of oligopeptides present in the first morning urines; (d) the only neurodevelopmental milestone clearly distinguishable from the previous cluster is social smile, whose delayed appearance is significantly correlated with an increased probability of observing verbal and/or motor stereotypies at intake, and with elevated 5-HT blood levels (*component 4*).

Correlation does not imply causation. Even assuming that pathophysiological mechanisms do link variables

Table VII. Correlation Between Whole Components/Single Variables and Biological Endophenotypes in Autistic Patients

	Head circumference ($N = 231$)	5-HT blood levels ($N = 138$)	Peptiduria ($N = 151$)
(1) Circadian and sensory dysfunction	$\tau = -0.076$, $P = 0.101$, $N = 231$	$\tau = -0.171$, $P = 0.004$, $N = 138$	$\tau = 0.095$, $P = 0.054$, $N = 151$
Presence/absence of self-injury	$\tau = -0.075$, $P = 0.396$, $N = 99$	$\tau = -0.417$, $P = 0.003$, $N = 35$	$\tau = -0.088$, $P = 0.612$, $N = 24$
Presence/absence of sleep disorders	$\tau = -0.073$, $P = 0.240$, $N = 196$	$\tau = 0.002$, $P = 0.980$, $N = 119$	$\tau = -0.103$, $P = 0.191$, $N = 110$
Excessive pain tolerance	$\tau = -0.034$, $P = 0.707$, $N = 98$	$\tau = 0.030$, $P = 0.814$, $N = 44$	$\tau = -0.192$, $P = 0.206$, $N = 31$
Hyperactivity	$\tau = -0.029$, $P = 0.648$, $N = 192$	$\tau = -0.111$, $P = 0.167$, $N = 106$	$\tau = 0.131$, $P = 0.108$, $N = 102$
(2) Immune dysfunction	$\tau = 0.075$, $P = 0.105$, $N = 231$	$\tau = 0.081$, $P = 0.167$, $N = 133$	$\tau = -0.015$, $P = 0.776$, $N = 151$
History of allergies in the patient	$\tau = 0.171$, $P = 0.003$, $N = 225$	$\tau = -0.041$, $P = 0.579$, $N = 125$	$\tau = -0.103$, $P = 0.160$, $N = 125$
Immune and/or allergic disease in the family	$\tau = 0.071$, $P = 0.125$, $N = 161$	$\tau = 0.327$, $P = 0.002$, $N = 61$	$\tau = -0.137$, $P = 0.195$, $N = 62$
Presence/absence of pathology at onset	$\tau = -0.102$, $P = 0.075$, $N = 229$	$\tau = -0.002$, $P = 0.981$, $N = 131$	$\tau = -0.018$, $P = 0.799$, $N = 129$
Obstetric complications in the mother	$\tau = 0.086$, $P = 0.139$, $N = 225$	$\tau = 0.010$, $P = 0.895$, $N = 126$	$\tau = 0.141$, $P = 0.056$, $N = 124$
(3) Neurodevelopmental delay	$\tau = -0.024$, $P = 0.608$, $N = 231$	$\tau = -0.121$, $P = 0.04$, $N = 138$	$\tau = 0.121$, $P = 0.014$, $N = 151$
Age at verbal language development	$\tau = -0.108$, $P = 0.112$, $N = 119$	$\tau = -0.134$, $P = 0.122$, $N = 67$	$\tau = 0.073$, $P = 0.411$, $N = 65$
Age at non-verbal language development	$\tau = -0.195$, $P = 0.014$, $N = 90$	$\tau = -0.284$, $P = 0.003$, $N = 57$	$\tau = 0.333$, $P = 0.003$, $N = 44$
Level of verbal language development	$\tau = -0.097$, $P = 0.094$, $N = 199$	$\tau = -0.110$, $P = 0.132$, $N = 114$	$\tau = 0.203$, $P = 0.007$, $N = 108$
Age at walking	$\tau = -0.128$, $P = 0.012$, $N = 211$	$\tau = -0.126$, $P = 0.053$, $N = 118$	$\tau = 0.051$, $P = 0.435$, $N = 115$
Age at sphincter control	$\tau = -0.113$, $P = 0.115$, $N = 113$	$\tau = -0.154$, $P = 0.083$, $N = 68$	$\tau = 0.030$, $P = 0.735$, $N = 66$
(4) Stereotypic behavior	$\tau = -0.068$, $P = 0.141$, $N = 231$	$\tau = 0.109$, $P = 0.063$, $N = 138$	$\tau = -0.096$, $P = 0.052$, $N = 151$
Age at social smile	$\tau = 0.030$, $P = 0.675$, $N = 123$	$\tau = 0.184$, $P = 0.031$, $N = 82$	$\tau = -0.323$, $P = 0.001$, $N = 69$
Verbal/vocal stereotypies at intake	$\tau = -0.010$, $P = 0.858$, $N = 219$	$\tau = 0.203$, $P = 0.005$, $N = 132$	$\tau = -0.061$, $P = 0.392$, $N = 132$
Motor stereotypies at intake	$\tau = 0.069$, $P = 0.225$, $N = 231$	$\tau = 0.009$, $P = 0.895$, $N = 136$	$\tau = -0.131$, $P = 0.064$, $N = 136$

Marker variables are highlighted in italics. Kendall's τ *P*-values, and sample sizes (*N*) are reported; middle- and light grey highlight nominal *P*-values <0.01, and <0.05, respectively, as in Table IV.

within each component, the nature of these mechanisms remains open to interpretation. Nonetheless, our component structure does provide important clues pointing toward possible mechanisms which will deserve closer scrutiny in future investigations. The first component, ''circadian and sensory dysfunction,'' encompasses both signs and symptoms reminiscent of the old ''opioid hypothesis'' of autism, relating to endogenous or dietary peptides [Leboyer et al., 1994; Meisel & FitzGerald, 2000; Sahley & Panksepp, 1987]. However, multiple studies now refute these putative mechanisms [Cass et al., 2008; Dettmer, Hanna, Whetstone, Hansen, & Hammock, 2007; Hunter, O'Hare, Herron, Fisher, & Jones, 2003]. In a recent genome-wide expression study, as many as 15 circadian rhythm regulatory or responsive genes were identified as differentially expressed in lymphoblastoid cell lines derived from the most severely affected subgroup of autistic individuals [Hu et al., 2009]. Impaired melatonin synthesis and profound decreases in melatonin serum levels have been described in autistic children [Melke et al., 2008]. Among biological endophenotypes not assessed in our study, melatonin serum levels indeed represent the primary candidate to be tested for association with the ''circadian and sensory dysfunction'' component. The second component, ''immune dysfunction,'' provides further evidence of possible contributions to ASD by a dysreactive immune system, as discussed elsewhere [Garbett et al., 2008; Lintas et al., 2009; Sacco et al., 2007; Vargas et al., 2005]. Immune proteins belonging to the major histocompatibility complex class I and to the complement cascade are also expressed in the brain, where they play a critical role in axonal guidance, functional plasticity and synaptic remodelling [Huh et al., 2000; Stevens et al., 2007]. Importantly, the present study unveils that the obstetric complications and recurrent spontaneous abortions, frequently reported by mothers of autistic children [Kolevzon et al., 2007], could be related to this immune dysfunction. These and other results confirm that autism is not typically the consequence of obstetric complications [Glasson et al., 2004], but may rather share with them some dysimmune pathophysiological underpinnings triggered prenatally and affecting the foeto-maternal unit. The third component, ''neurodevelopmental delay,'' closely resembles the ''developmental milestones'' factor described by Tadevosyan-Leyfer et al. [2003], indicating that in many autistic children multiple developmental milestones indeed tend to be affected in parallel. This component correlates here with global amounts of oligopeptides found in the first morning urines of ASD patients (the sequence of the peptides contributing to specific HPLC peaks was not investigated here). Recent evidence suggests that abnormal calcium homeostasis, resulting in oxidative stress, could favor peptiduria at the kidney level (as occurs in diabetes), while affecting neurodevelopment at the brain level [Chauhan & Chauhan, 2006; Krey & Dolmetsch, 2007; Lintas et al., 2009; Palmieri et al., 2010]. Finally component 4, ''stereotypic behavior,'' suggests that social smile develops independently of all the remaining behavioral milestones included in component 3. The neurobiological

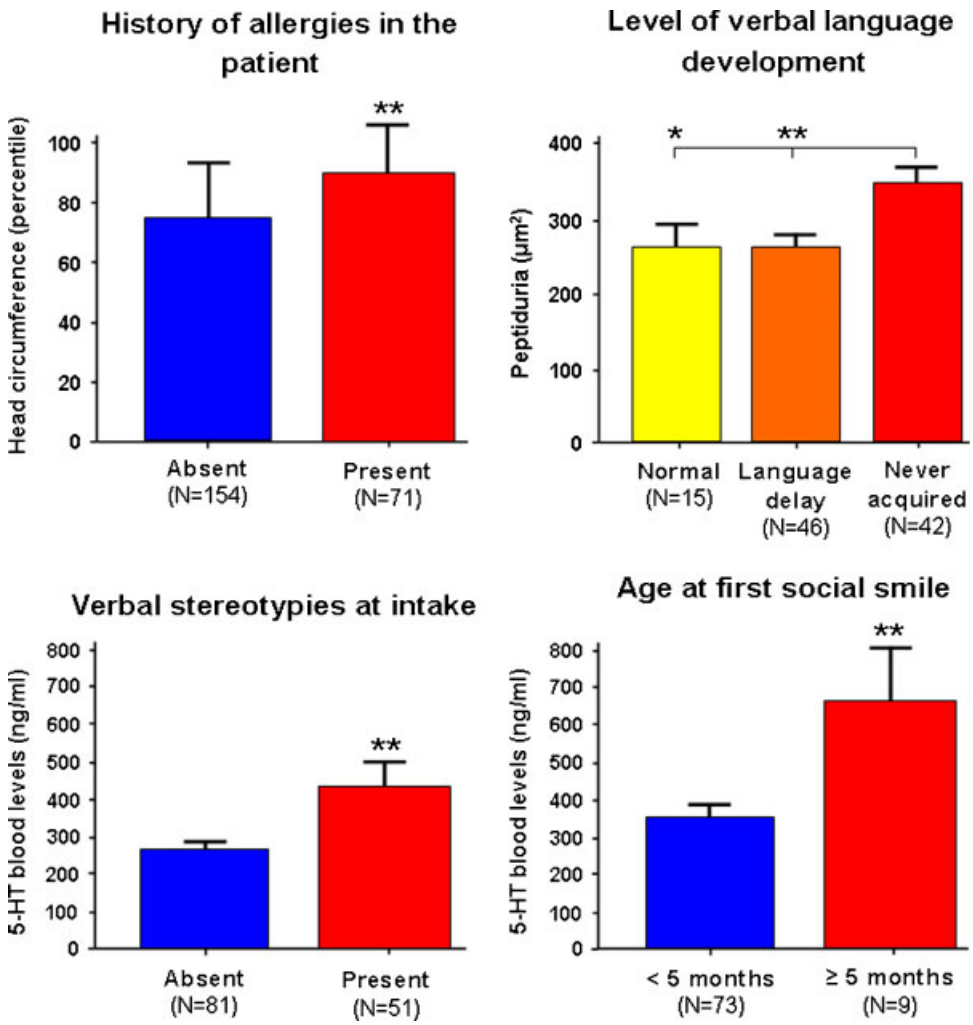

Figure 1. Quantitative biological endophenotypes by marker variable for components 2, 3, and 4, with the addition of "age at first social smile" for component 4. Data are presented as median percentile ± semi IQR for head circumference, and mean ± SEM for 5-HT blood levels and peptiduria. $^{*}P<0.05$, $^{**}P<0.01$. [Color figure can be viewed online at wileyonlinelibrary.com]

substrate underlying social smile in a neonate, responding to the smile he/she observes in his/her parent's face, is represented by "mirror neurons," frontocortical motoneurons which discharge to the same extent whether the subject is performing an action or whether he/she is observing the same action performed by another individual with the same purpose [for review see Rizzolatti & Fabbri-Destro, 2008]. The strong positive correlation between age at first social smile and stereotypic behaviors at intake (Table IV), as well as the link with serotonin, are difficult to interpret at this time, although 5-HT ought to be recognized as more than just a neurotransmitter, since all the way from invertebrates to humans it regulates neurodevelopmental processes as relevant as cell proliferation, apoptosis, migration, synaptogenesis, neurite sprouting and pruning, and neural plasticity [for review see Persico, 2008].

The methodology implemented in this study presents several limitations, in reference to psychometric measures, statistical strategy and methods to measure biological endophenotypes:

(a) *Psychometric measures*: the unavailability of ADOS and ADI-R in Italy until 2005 forced us to use a questionnaire designed by our group, which has not undergone rigorous instrument validation. Some variables, such as familial psychiatric disorders and parental age, proven to represent pathogenetic contributors in autism after 1997 [Durkin et al., 2008; Larsson et al., 2005], were not included. Also, the ADI-R would have indeed provided a more detailed clinical characterization: at least two clinical factors identified in recent years using the ADI-R, "savant skills" and "insistence in sameness" [Happé & Ronald, 2008], were not addressed here. Several items, based exclusively on parental report, should be viewed with due caution, although parental reports constitute the sole information source also for several standardized scales such as ADI-R and VABS. The variable extent to which parents felt confident in responding at

each item of our questionnaire generated large ranges in the completeness of clinical and family history data (see Table II). This dyshomogeneity, coupled with different age-related inclusion criteria for biological endophenotypic measures, may have reduced our statistical power enhancing the probability of false-negatives, despite the elimination of items with low informativeness. Notwithstanding these limitations, our questionnaire was unanimously judged by clinicians both simple to administer and reliable, two critical features in a multi-center study. Furthermore, several patient and family history-related items relevant to this study would not have been picked by a clinical assessment limited to standardized diagnostic interviews. For example, many items regarding normal and abnormal behaviors are not scored here on the basis of ''presence/absence'' or ''degree of severity,'' but asking parents the age at onset. This strongly limits the replicability of our findings using already available data sets based exclusively on the ADOS or the ADI-R, unless a similar questionnaire is applied. Nonetheless, a few items (no. 2, 9, 26, and 27), mostly pertaining to the domain of stereotypic behaviors, can be directly correlated with specific ADOS or ADI-R items, yielding in our sample τ coefficients ranging between 0.270 and 0.525 despite very small sample sizes (Supplementary Table S7). Also the diagnostic validity of our recruitment could be tested and confirmed after 2005, following the implementation of ADOS and ADI-R, since the same clinicians were involved in patient selection throughout the duration of this study.

(b) *Statistical analyses*: principal component analysis and similar statistical approaches have been frequently used to investigate autistic disorder. Researchers applying this strategy have typically administered standardized clinical assessment tools, such as the ADI-R, ADOS and VABS, seeking to achieve a detailed description of the patient's behavior and level of functioning ''here and now.'' This information was then used to dissect symptom and severity clusters out of the overall clinical phenotype [for review see Happé & Ronald, 2008]. This approach can be of paramount importance when attempting to improve the validity of symptom models implemented by standardized diagnostic protocols such as the DSM-IV [American Psychiatric Association, 1994]. However, it provides little or no clue regarding the pathophysiological processes underlying symptom clusters. The present study significantly extends previous findings and cannot be immediately placed within the context of published studies applying similar statistical approaches. Nonetheless, the general limitations of principal component analysis indeed apply also to our work and must be acknowledged, as there is no criterion beyond interpretability against which to test the solution [Tabachnick & Fidell, 1989]. We essentially defined our model based on the convergence and statistical consistency of principal component and unweighted least squares analyses, on the higher percentage of variance explained by this model compared to image factoring and maximum likelihood factoring, and on its higher biological plausibility. Component 4, ''stereotypic behavior,'' was maintained because its items show significant cross-correlations, a significant association with 5-HT blood levels and a strong biological plausibility, as discussed above. Component 5 was interpreted as a false-positive. Overall, the principal component structure presented here should be viewed with caution until independently replicated, especially in reference to the latter two components.

(c) *Biological endophenotypes*: an ''endophenotype'' can be best described as a familial and heritable quantitative trait associated with the disease [Gottesman & Gould, 2003]. In complex disorders, endophenotypes are used to define subgroups of patients likely sharing common genetic and pathophysiological underpinnings [Gottesman & Gould, 2003]. The endophenotypic measures employed in this study display several limitations: (I) endophenotypic and demographic data was available on subgroups of patients, and not on our entire sample. While we are striving to increase the number of patients with all three endophenotypic measures, currently this limitation is still significant, because the subgroups with physiological measures cannot be factor-analyzed due to small sample size. Since factors generated on a larger sample have then been correlated to physiological measures in smaller subsets, caution should be exercised in generalizing the results again to the entire group; (II) the sex- and age-specific standard tables used in most European countries as norms for head circumference, height, and weight in children were originally derived from Caucasian populations living in the US and the UK and are not nation-specific [Bushby, Cole, Matthews, & Goodship, 1992; Hamill et al., 1979]; (III) 5-HT was measured in platelet-rich plasma obtained by using a straightforward and standardized centrifugation protocol [Rolf, Haarmann, Grotemeyer, & Kehrer, 1993], and consequently 5-HT levels were not normalized by platelet numbers; (IV) global amounts of urinary peptides were measured by integrating the total AUC of the HPLC plot and single peaks were not measured, because the latter approach has not been replicated [Cass et al., 2008; Dettmer et al., 2007; Hunter et al., 2003]. We acknowledge that more precise methods would have conceivably yielded more accurate estimates of sample distributions. However, the appropriateness of a given methodology is strictly dependent on the specific experimental question that is being addressed. In the present study, our methods have been evidently able to pick correlations whose strength appears to largely outweigh the sum of naturally occurring variability plus spurious variability introduced by methodological limitations. Furthermore, our use of morphological and biochemical endophenotypes

provides at least two advantages over behavioral endophenotypes, namely enhanced reliability and greater validity. Biological parameters are measured using standardized and/or automated procedures, generally more reproducible than psychometric measures. Their lesser complexity and greater proximity to the genetic level facilitates the interpretation of the results. Finally, family members of the autistic patients enrolled in this study were used to demonstrate the existence of familiality for these three parameters, further enhancing confidence in the appropriateness of their use as an ''endophenotype'' and of our measurement methodologies.

In conclusion, our data suggest the existence of at least four components contributing to autistic disorder and their association with biological endophenotypes. These results open the exciting perspective of allowing the identification of homogeneous subgroups of ASD patients, as suggested by preliminary cluster analyses, indicating that the distribution of the cumulative component scores allows each patient to be categorized into one of several relatively homogeneous subgroups (Roberto Sacco and Antonio M. Persico, manuscript in preparation). The absence of other pathophysiological components in this study may simply indicate that the measures employed here were not designed to identify them, or that our sample was underpowered to detect these additional components. Until this principal component structure is confirmed in component-specific replication studies, caution should be exercised in applying our conclusions to other samples. If confirmed, these results will contribute to design better-targeted mechanistic studies aimed at achieving a more thorough understanding of the pathophysiological processes leading to autism. On a broader scale, they may also apply to normal cognitive diversity in the general population, where nocturnalism and weak circadian rhythms, allergies, motor delay and clumsiness often times characterize individuals distributed along the autism spectrum.

Acknowledgments

The authors gratefully acknowledge all the families who participated in this study, Louis Tao (Department of Biostatistics, University of Beijing, China) for providing statistical advise and for reviewing the manuscript, Giacomo Rizzolatti (Department of Physiology, University of Parma, Italy) for helpful discussion, and Roberto Rigardetto, Marina Gandione, Simona Trillo, and Maria Paola Santangelo for contributing to patient recruitment and data collection. This work was supported by the Italian Ministry for University, Scientific Research and Technology (PRIN n.2006058195), the Italian Ministry of Health (RFPS-2007-5-640174), and the Autism Speaks Foundation (Princeton, NJ).

Financial Disclosure: the Authors have no conflict of interest, financial or otherwise, related directly or indirectly to the submitted work.

References

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders, 4e. Washington, DC: American Psychiatric Press.

Bauman, M.L., & Kemper, T.L. (2005). Neuroanatomic observations of the brain in autism: A review and future directions. International Journal of Developmental Neuroscience, 23, 183–187.

Bushby, K.M.D., Cole, T., Matthews, J.N.S., & Goodship, J.A. (1992). Centiles for adult head circumference. Archives of Disease in Childhood, 67, 1286–1287.

Cass, H., Gringras, P., March, J., McKendrick, I., O'Hare, A., et al. (2008). Absence of urinary opioid peptides in children with autism. Archives of Disease in Childhood, 93, 745–750.

Chauhan, A., & Chauhan, V. (2006). Oxidative stress in autism. Pathophysiology, 13, 171–181.

Courchesne, E., Pierce, K., Schumann, C.M., Redcay, E., Buckwalter, J.A., et al. (2007). Mapping early brain development in autism. Neuron, 56, 399–413.

Dettmer, K., Hanna, D., Whetstone, P., Hansen, R., & Hammock, B.D. (2007). Autism and urinary exogenous neuropeptides: development of an on-line SPE-HPLC-tandem mass spectrometry method to test the opioid excess theory. Analytical and Bioanalytical Chemistry, 388, 1643–1651.

DiCicco-Bloom, E., Lord, C., Zwaigenbaum, L., Courchesne, E., Dager, S.R., et al. (2006). The developmental neurobiology of autism spectrum disorder. Journal of Neuroscience, 26, 6897–6906.

Durkin, M.S., Maenner, M.J., Newschaffer, C.J., Lee, L.C., Cunniff, C.M., et al. (2008). Advanced parental age and the risk of autism spectrum disorder. American Journal of Epidemiology, 168, 1268–1276.

Fombonne, E. (2005). Epidemiology of autistic disorder and other pervasive developmental disorders. Journal of Clinical Psychiatry, 66, 3–8.

Garbett, K., Ebert, P.J., Mitchell, A., Lintas, C., Manzi, B., et al. (2008). Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiology of Disease, 30, 303–311.

Glasson, E.J., Bower, C., Petterson, B., de Klerk, N., Chaney, G., & Hallmayer, J.F. (2004). Perinatal factors and the development of autism: a population study. Archives of General Psychiatry, 61, 618–627.

Gottesman, I.I., & Gould, T.D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. American Journal of Psychiatry, 160, 636–645.

Hamill, P.V.V., Drizd, T.A., Johnson, C.L., Reed, R.B., Roche, A.F., & Moore, W.M. (1979). Physical growth: National Canter for Health Statistics percentiles. American Journal of Clinical Nutrition, 32, 607–629.

Happé, F., & Ronald, A. (2008). The ''fractionable autism triad'': A review of evidence from behavioural, genetic, cognitive and neural research. Neuropsychology Review, 18, 287–304.

Hu, V.W., Sarachana, T., Kim, K.S., Nguyen, A., Kulkarni, S., et al. (2009). Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: Evidence for circadian rhythm dysfunction in severe autism. Autism Research, 2, 78–97.

Huh, G.S., Boulanger, L.M., Du, H., Riquelme, P.A., Brotz, T.M., & Shatz, C.J. (2000). Functional requirement for class I MHC in CNS development and plasticity. Science, 290, 2155–2159.

Hunter, L.C., O'Hare, A., Herron, W.J., Fisher, L.A., & Jones, G.E. (2003). Opioid peptides and dipeptidyl peptidase in autism. Developmental Medicine & Child Neurology, 45, 121–128.

Jyonouchi, H., Geng, L., Ruby, A., & Zimmerman-Bier, B. (2005). Dysregulated innate immune responses in young children with autism spectrum disorders: Their relationship to gastrointestinal symptoms and dietary intervention. Neuropsychobiology, 51, 77–85.

Kolevzon, A., Gross, R., & Reichenberg, A. (2007). Prenatal and perinatal risk factors for autism: A review and integration of findings. Archives of Pediatrics and Adolescent Medicine, 161, 326–333.

Krey, J.F., & Dolmetsch, R.E. (2007). Molecular mechanisms of autism: A possible role for Ca^{2+} signaling. Current Opinion in Neurobiology, 17, 112–119.

Lainhart, J.E., Bigler, E.D., Bocian, M., Coon, H., Dinh, E., et al. (2006). Head circumference and height in autism: A study by the collaborative program of excellence in autism. American Journal of Medical Genetics, 140, 2257–2274.

Larsson, H.J., Eaton, W.W., Madsen, K.M., Vestergaard, M., Olesen, A.V., et al. (2005). Risk factors for autism: Perinatal factors, parental psychiatric history, and socioeconomic status. American Journal of Epidemiology, 161, 916–925.

Leboyer, M., Bouvard, M.P., Recasens, C., Philippe, A., Guilloud-Bataille, M., et al. (1994). Difference between plasma N- and C-terminally directed beta-endorphin immunoreactivity in infantile autism. American Journal of Psychiatry, 151, 1797–1801.

Lintas, C., & Persico, A.M. (2009). Autistic phenotypes and genetic testing: State-of-the-art for the clinical geneticist. Journal of Medical Genetics, 46, 1–8.

Lintas, C., Sacco, R., Garbett, K., Mirnics, K., Militerni, R., et al. (2009). Involvement of the PRKCB1 gene in autistic disorder: significant genetic association and reduced neocortical gene expression. Molecular Psychiatry, 14, 705–718.

Lord, C., Rutter, M., DiLavore, P.C., & Risi, S. (2002). ADOS, Autism Diagnostic Observation Schedule. Los Angeles, CA: Western Psychological Services [Italian version by Tancredi, R., Saccani, M., Persico, A.M., Parrini, B., Igliozzi, R., & Faggioli, R., editors. (2005). Florence, Organizzazioni Speciali].

McBride, P.A., Anderson, G.M., Hertzig, M.E., Snow, M.E., Thompson, S.M., et al. (1998). Effects of diagnosis, race, and puberty on platelet serotonin levels in autism and mental retardation. Journal of the American Academy of Child and Adolescent Psychiatry, 37, 767–776.

Meisel, H., & FitzGerald, R.J. (2000). Opioid peptides encrypted in intact milk protein sequences. British Journal of Nutrition, 84, S27–S31.

Melke, J., Goubran Botros, H., Chaste, P., Betancur, C., Nygren, G., et al. (2008). Abnormal melatonin synthesis in autism spectrum disorders. Molecular Psychiatry, 13, 90–98.

Miller, M.T., Stromland, K., Ventura, L., Johansson, M., Bandim, J.M., & Gillberg, C. (2005). Autism associated with conditions characterized by developmental errors in early embryogenesis: A mini review. International Journal of Developmental Neuroscience, 23, 201–219.

Palmieri, L., Papaleo, V., Porcelli, V., Scarcia, P., Gaita, L., et al. (2010). Altered calcium homeostasis in autism-spectrum disorders: Evidence from biochemical and genetic studies of the mitochondrial aspartate/glutamate carrier AGC1. Molecular Psychiatry, 1, 38–52.

Persico, A.M. (2008). Developmental roles of the serotonin transporter. In: Kalueff AV, editor. Experimental models in serotonin transporter research. Cambridge: Cambridge University Press, pp 78–104.

Persico, A.M., & Bourgeron, T. (2006). Searching for ways out of the autism maze: genetic, epigenetic and environmental clues. Trends in Neurosciences, 29, 349–358.

Persico, A.M., Pascucci, T., Puglisi-Allegra, S., Militerni, R., Bravaccio, C., et al. (2002). Serotonin transporter promoter variants do not explain the hyperserotoninemia in autistic children. Molecular Psychiatry, 7, 795–800.

Piven, J., Tsai, G., Nehme, E., Coyle, J.T., Chase, G.A., & Folstein, S.E. (1991). Platelet serotonin, a possible marker for familial autism. Journal of Autism and Developmental Disorders, 21, 51–59.

Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: Evidence from a family history study of multiple-incidence autism families. American Journal of Psychiatry, 154, 185–190.

Reichelt, W.H., Knivsberg, A.M., Nodland, M., Stensrud, M., & Reichelt, K.L. (1997). Urinary peptide levels and patterns in autistic children from seven countries, and the effect of dietary intervention after 4 years. Developmental Brain Dysfunction, 10, 44–55.

Reichelt, W.H., Ek, J., Stensrud, M., & Reichelt, K.L. (1998). Peptide excretion in coeliac disease. Journal of Pediatric Gastroenterology and Nutrition, 26, 305–309.

Rizzolatti, G., & Fabbri-Destro, M. (2008). The mirror system and its role in social cognition. Current Opinion in Neurobiology, 18, 179–184.

Rolf, L.H., Haarmann, F.Y., Grotemeyer, K.-H., & Kehrer, H. (1993). Serotonin and amino acid content in platelets of autistic children. Acta Psychiatrica Scandinavica, 87, 312–316.

Rutter, M., Le Couter, A., & Lord, C. (2003). ADI-R, Autism Diagnostic Interview—Revised. Los Angeles, CA: Western Psychological Services [Italian version by Faggioli, R., Saccani, M., Persico, A.M., Tancredi, R., Parrini, B., & Igliozzi, R., editors. (2005). Florence: Organizzazioni Speciali].

Sacco, R., Militerni, R., Frolli, A., Bravaccio, C., Gritti, A., et al. (2007). Clinical, morphological, and biochemical correlates of head circumference in autism. Biological Psychiatry, 62, 1038–1047.

Sahley, T.L., & Panksepp, J. (1987). Brain opioids and autism: An updated analysis of possible linkages. Journal of Autism and Developmental Disorders, 17, 201–216.

Saresella, M., Marventano, I., Guerini, F.R., Mancuso, R., Ceresa, L., et al. (2009). An autistic endophenotype results in complex

immune dysfunction in healthy siblings of autistic children. Biological Psychiatry, 66, 978–984.
Sparrow, S.S., Balla, D.A., & Cicchetti, D.V. (1984). Vineland Adaptive Behavior Scales—Survey Form. Circle Pines, MN: American Guidance Service Inc.
Stevens, B., Allen, N.J., Vasquez, L.E., Howell, G.R., Christopherson, K.S., et al. (2007). The classical complement cascade mediates CNS synapse elimination. Cell, 131, 1164–1178.
Tabachnick, B.G., & Fidell, L.S. (1989). Using multivariate statistics, 2e. New York, NY: Harper & Row, pp 597–677.
Tadevosyan-Leyfer, O., Dowd, M., Mankoski, R., Winklosky, B., Putnam, S., et al. (2003). A principal components analysis of the Autism Diagnostic Interview-Revised. Journal of The American Academy of Child and Adolescent Psychiatry, 42, 864–872.
Tuchman, R., & Rapin, I. (2002). Epilepsy in autism. Lancet Neurology, 1, 352–358.
Vargas, D.L., Nascimbene, C., Krishnan, C., Zimmerman, A.W., & Pardo, C.A. (2005). Neuroglial activation and neuroinflammation in the brain of patients with autism. Annals of Neurology, 57, 67–81.