ARTICLE   **Open Access**

Transcriptome signatures from discordant sibling pairs reveal changes in peripheral blood immune cell composition in Autism Spectrum Disorder

Michele Filosi[1], Tony Kam-Thong[2], Laurent Essioux[2], Pierandrea Muglia[3], Elisabetta Trabetti[4], Will Spooren[2], Bertram Müller-Myshok[5] Italian Autism Network, Enrico Domenici [ID][1,6]

Abstract

Notwithstanding several research efforts in the past years, robust and replicable molecular signatures for autism spectrum disorders from peripheral blood remain elusive. The available literature on blood transcriptome in ASD suggests that through accurate experimental design it is possible to extract important information on the disease pathophysiology at the peripheral level. Here we exploit the availability of a resource for molecular biomarkers in ASD, the Italian Autism Network (ITAN) collection, for the investigation of transcriptomic signatures in ASD based on a discordant sibling pair design. Whole blood samples from 75 discordant sibling pairs selected from the ITAN network where submitted to RNASeq analysis and data analyzed by complementary approaches. Overall, differences in gene expression between affected and unaffected siblings were small. In order to assess the contribution of differences in the relative proportion of blood cells between discordant siblings, we have applied two different cell deconvolution algorithms, showing that the observed molecular signatures mainly reflect changes in peripheral blood immune cell composition, in particular NK cells. The results obtained by the cell deconvolution approach are supported by the analysis performed by WGCNA. Our report describes the largest differential gene expression profiling in peripheral blood of ASD subjects and controls conducted by RNASeq. The observed signatures are consistent with the hypothesis of immune alterations in autism and an increased risk of developing autism in subjects exposed to prenatal infections or stress. Our study also points to a potential role of NMUR1, HMGB3, and PTPRN2 in ASD.

Introduction

Autism Spectrum Disorder (ASD) is a group of neuro-developmental disorders with onset in early childhood, characterized by a triad of core symptoms (impaired social interaction, poor language development and communication and repetitive and narrow pattern of behaviors and interests. The heritability of ASD is relatively well established, and it is now thought that 10–20% of the cases are genetically defined[1]. Yet, the genetic etiology is not fully understood, and genome-wide investigations support a complex genetic architecture based on major genes and polygenic factors having different extent of contribution across the ASD spectrum[2]. Several etiological hypotheses for ASD exist, as for example altered synaptic dysfunction leading to an imbalance of excitatory and inhibitory neurotransmission[3], although a unifying etiological theory is still missing. Abnormalities in brain tissue at the molecular level, including transcriptional and splicing dysregulations, have been shown to correlate with neuronal dysfunctions[4–6]. The investigations on post-mortem tissue from ASD patients have shed light on the molecular mechanisms underlying the disorder at brain level, confirming the importance of transcriptional analysis in disease characterization. However, the search for a

Correspondence: Enrico Domenici (enrico.domenici@unitn.it)
[1]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento (TN), Italy
[2]Roche Pharmaceutical Research and Early Development (pRED), Roche Innovation Center, Grenzacherstrasse 124, Basel, Switzerland
Full list of author information is available at the end of the article
These authors contributed equally: Michele Filosi, Tony Kam-Thong

reliable molecular signature for ASD based on peripheral samples, which might help clinicians in early diagnosis and in the identification of ASD subgroups, is still ongoing. Several attempts in this direction have been performed by gene expression analysis of lymphoblastoid cell lines[7–15] and blood samples[16–25] (see for a review ref. [26]). Overall, these studies suggest the implication of several signaling pathways and the immune response in ASD, but a consistent set of diagnostic biomarkers remains elusive. A recent meta-analysis of blood-based transcriptome investigations in ASD remarks the hypothesis of implication of the immunologic function[27]. Indeed, the predominant signature observed in the ASD blood transcriptome was characterized by reduced expression of transcripts related to innate immune and inflammatory signaling, including type I and type II interferon-stimulated signaling cascades. Additional findings were under-expression of EGF-, PDGF-, PI3K-AKT-mTOR-, and RAS-MAPK-signaling cascades, and over-expression of modules enriched in ribosomal translation and NK-cell-related functions.

Q1 – Q4

The Italian Autism Network (ITAN) consists of a collection of families formed by ASD probands, unaffected siblings when available, and parents recruited through thirteen clinical centers across Italy[28]. The network has an associated repository of genomic DNA, blood RNA, plasma, and lymphoblastoid cell lines to enable integrated genetic and biomarker research.

Here we present the first transcriptome analysis conducted on the ITAN cohort aimed at identifying peripheral signatures for ASD. The study was conducted on a selected subset of the available discordant sibling pairs, consisting in a total of 150 subjects. Blood samples were subjected to RNA Sequencing and data were analyzed by exploiting a matched pairs design. In order to disentangle the complexity of the molecular changes potentially arising from differences in blood composition in disease vs healthy state, we applied two different cell deconvolution algorithms and validated our findings with a complimentary WGCNA analysis.

Material and methods
Participants

The ITAN collection comprehends more than 800 subjects belonging to 256 families recruited across thirteen centers[28]. The study protocol was in first instance approved by the Verona Hospital Ethical Review board (study protocol AUT-SFK001, CE1419), and subsequently by the Ethical Review Committees of each recruiting site. All adult subjects participating in this project gave their written consent (or the consent for their children); assent to participate to the study from the children was obtained whenever obtainable. Diagnosis of autism spectrum disorder according to DSM IV[29], was assessed by experienced child psychiatrists using standard tools: Broader

Table 1 Demographic information on the subset of the ITAN collection used in this study.

Sibling pairs		Autism	PDD-NOS	ASP	Total
Gender concordance (N of subjects)	Male	27	5	6	38
	Female	4	1	0	5
	Discordant	24	4	4	32
Ethnicity (N of subjects)	CEU	41	8	6	55
	Other	14	2	4	20
Age	Cases	10.02	10.00	12.7	10.33
	CTRL	11.34	10.5	15.1	11.7

PDD-NOS Pervasive development disorder not otherwise specified, *ASP* Asperger syndrome.

Phenotype Autism Symptom Scale (BPASS), Autism Diagnostic Interview—Revised (ADI-R), Autism Diagnostic Observation Schedule (ADOS), and Krug Asperger Disorder Index screening. We selected families with at least two siblings discordant for ASD diagnosis, with the affected children between 4 and 18 years of age. After the selection process, 75 sibling pairs were sequenced, for a total of 150 subjects. Demographic parameters, age, gender, and ethnicity for each subject are shown in Table 1.

RNA sequencing

Total RNA from blood samples, randomized by Roche statisticians, was submitted to Poly-A RNA sequencing on Illumina RNASeq Platform. Pre-alignment/mapping quality control (using Illumina sequencing accuracy quality scoring to estimate the base calling error probabilities, thresholds of all bases >Q30) was performed to confirm that key laboratory quality metrics criteria were met. To estimate expression at gene level, paired-end RNAseq reads were mapped to the human genome (hg19) by using the short-read aligner GSNAP[30]. The number of mapped reads for all RefSeq transcript variants of a gene counts were combined into a single value by using SAMTOOLS[31]. Technical features such as RIN, sequencing batch number, sequencing lane and pool, RIN, µgRNA were collected for subsequent statistical analysis. The RNA Sequencing data that support the findings of this study are available from the ITAN Foundation (see https://www.fondazioneitan.org for details) upon submission of an official request. Data are released only for research purposes, upon assessment of a project proposal by the ITAN Scientific Committee.

HLA typing and HLA and KIR variants expression from RNASeq

We inferred HLA genotypes and quantified HLA allele expression from RNA Sequencing data we used a recently

developed computational pipeline called HLAPers[32]. The pipeline consists in two steps: (i) the HLA typing step, where reads are aligned on a curated database of HLA variants available as transcript sequences to infer the subject genotype; (ii) the quantification step, where expression is estimated based on the number of reads aligning to each reference, using a model accounting for the occurrence of reads mapping to multiple HLA references corresponding to different HLA alleles or genes. We thus realigned original ITAN RNASeq data onto the International Immunogenetics/HLA (IMGT) database release 3.31.0[33] to extract highly specific HLA transcripts and, based on the best aligned sequences, inferred the subject genotype for the HLA variants available in the IMGT database.

For KIR genes, we used the gene counts combined into single value by using SAMTOOLS to quantify specific the expression of KIR genes, since the KIR sequences database[34] does not provide transcript sequences for different genotypes.

Statistical analysis

A diagram representing our full analysis workflow is shown in Supplementary Fig. 1. After the preprocessing step, we filtered genes with count-per-million (CPM) greater than 1 in 25% of the samples, then gene expression was normalized and log-transformed using functions implemented in edgeR[35]. To account for differences in sequencing depths and RNA composition across samples, gene counts were normalized using the trimmed mean normalization method, while for the estimation of the biological coefficient of variation (BCV) under the assumption of a negative binomial distribution we used the estimateDisp function from edgeR (see Supplementary Fig. 2 where the BCV estimate is reported). A preliminary multivariate analysis was conducted to identify data structure, outliers and other factors potentially affecting expression levels. We used the non-affected sibling as the control, thus exploiting the high degree of kinship with the ASD subject, by using a paired design to correct for sample dependence. We estimated the beta coefficients for each technical and demographic feature, including diagnosis, using a generalized linear model. The technical features included in the model are described in the previous section, while the demographic features are: gender, ethnicity, age, family id and diagnosis. A moderated gene-wise variance was then computed based on an empirical Bayes procedure[36] to extract the differential gene expression (DGE) signature. The same model was also run by including an estimation score for blood cell composition for each sample (see below) as covariate in the DGE analysis. For all DGE analyses we used a paired design comparing the affected subjects with

the related discordant siblings. All analyses were conducted with R 3.4.3.

Enrichment analysis of the differentially expressed genes (DEGs) was performed by using the online web-service KOBAS 3.0[37]. We run the analysis with default parameters on five human pathway databases (KEGG, BioCyc, Reactome, Panther, GO) using all blood expressed genes from our RNASeq dataset as background list for the enrichment. Additional enrichment analysis was conducted by using Open Target[38] and StringDB[39] to identify diseases, drugs and literature associated with the DEG list.

Cell composition estimate

Two complementary deconvolution methods were used to estimate blood cell composition, ie CIBERSORT and xCell. CIBERSORT allows the extraction of cell mixture proportions from gene expression profiles based on support vector regression of a cell-specific signature matrix[40]. We have used the on-line version of the tool[41], which allows the estimation of a cell mixture based on a signature matrix validated with curated signatures from 22 human hematopoietic cell types (LM22). The algorithm takes the gene expression profile matrix as an input and returns a proportion between 0 and 1 for the 22 cell types and an empirical p-value. xCell[42] improves previous deconvolution algorithms by correcting the biases due to overlapping signatures between similar cell types. The signature matrix is based on lists of genes extracted from multiple gene expression analysis and validated by cell sorting studies. Differently from previous methods, xCell uses single sample enrichment to estimate a score linearly correlated with the abundances of cell types in the cell mixture. We adapted a R package publicly available on GitHub[43] by using a subset of the original xCell signature consisting in 14 cell types relative to blood (based on the studies SDY311 and SDY420[44] from the ImmPort database[45]). Input data of the cell enrichment algorithm was the voom normalized gene expression[46]. xCell returns a cell-specific enrichment score (linearly correlated with cell abundance in the mixture) and an associated FDR adjusted p-value. As suggested by the original paper[42], we selected cell types with a corrected p-value < 0.2. For both methods, differences in cell composition between discordant siblings were assessed by t-test with matched pairs design.

WGCNA

We followed the pipeline proposed by Langfelder et al.[44] in their CRAN package to infer gene co-expression networks and identify network modules within R 3.4.3 statistical environment. Further details of the methods used throughout this analysis can be found in the original manuscript[47] and on the website[48]. Networks were

inferred using the TOMsimilarityFromExpr function with "bicor" as correlation measure. The soft-threshold parameter was optimized with the function pickSoftThreshold and the best threshold ($\alpha = 12$) selected by visual inspection, as suggested by the WGCNA pipeline. Correlations between modules eigengenes, diagnosis and cell mixture estimation score were computed. Modules with the highest correlation and significant p-value ($\alpha < = 0.05$) were selected for further analysis.

Results

Identification of DEGs

An exploratory analysis of the RNASeq data through the PCA analysis did not reveal any noticeable outliers (see Supplementary Fig. 3). A close similarity of expression was observed for siblings, supporting the choice of a discordant sib-pair experimental design. When assessing the fraction of the variance explained by the first five PCA components for each covariate, the family component was indeed confirmed to be the predominant factor (see Supplementary Fig. 4 and Supplementary Table 1). The DGE analysis revealed subtle differences in expression between discordant siblings (see Supplementary Table 2 for the full list of DEGs). Setting a threshold on the FDR corrected p-value to 0.1 or 0.05 result with a shorter list of 21 and 10 genes, respectively (see Table 2a). The volcano plot in Fig. 1(a) shows that the DEGs are predominantly downregulated in ASD, in accordance with previous findings[18]. The QQ plot shows a deviation from the null model even at p-values below significance, with a calculated lambda value of 1.23, suggesting an inflation of the type I error rate, possibly due to hidden variability. Based

on prior evidence of sex differences in blood transcriptomic signatures of ASD[15,49], we performed a separate analysis for male sibling pairs, for which an adequate number was available (see Supplementary Fig. 5, and Supplementary Table 3). The results suggest a strong correlation between the gender-covariate analysis conducted on the full datasets and the male-specific analysis, with most genes (86%) maintaining the same direction of expression change, and no discordant sign for the first 900 genes of the ranked list of DGE in the full datasets (see also Supplementary Fig. 6).

Enrichment analysis

We performed enrichment analysis by submitting the full list of nominally significant DEGs to the KOBAS algorithm[37]. Enrichment analysis with KOBAS shows five statistically significant enriched pathways with FDR < 0.05 (See Supplementary Table 4). The first two pathways are Natural Killer-cell mediated cytotoxicity and immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell, suggesting a dysregulation in the immune functions in ASD.

When submitting the list of significant genes at FDR < 0.25 to Open Target batch search to look for diseases associated to our signature, the top finding was cytomegalovirus infection (see Supplementary Table 5). Cytomegalovirus infection was still the top finding when restricting the analysis to the shorter 21 gene list (FDR < 0.1). Similarly, when searching for published signatures similar to our list of FDR < 0.25 significant genes (by using StringDB) we identified the enrichment in CMV-specific CD4+ T cells signatures from chronically infected healthy donors[50] as the top hit (P value $= 8.86$ E−10).

Cell composition

Despite the limited number of DEGs between discordant siblings, our enrichment analysis indicates a clear dysregulation in immune functions. To verify if such differences are related to differences in immune cell populations between the autism and control groups, we submitted the gene expression dataset to two complimentary methods to estimate cell composition in blood. By using CIBERSORT and the LM 22 signature matrix to estimate blood cell proportion, and comparing discordant siblings with a paired design, we were able to show statistically significant differences (p-value 0.05) for four cell types, including NK cells, Tgd, B cells and CD4+ TEM cells (see Table 3a). The results from the analysis with xCell are consistent with CIBERSORT, in particular for NK cells and B cells (see Table 3b). Indeed, we found high degree of correlation between the scores for NK cells ($r2 = 0.7$, $p = 1e-24$) and naïve B cells ($r2 = 0.8$, $p = 1e-56$) from the two algorithms (see Supplementary Fig. 7).

Table 2 DGE model results ordered by FDR.

	(a) Standard model			(b) xCell enrichment		
Name	LogFC	P Value	FDR	LogFC	P Value	FDR
HMGB3	−0.306	6.15e−08	**8.99e−04**	−0.250	2.02e−05	**1.54e−01**
NMUR1	−0.345	3.30e−06	**2.41e−02**	−0.224	2.11e−05	**1.54e−01**
PTPRN2	0.294	6.47e−06	**2.84e−02**	0.290	3.36e−05	**1.63e−01**
NKG7	−0.355	8.01e−06	**2.84e−02**	−0.207	6.71e−04	9.93e−01
PIF1	−0.413	105e−05	**2.84e−02**	−0.299	7.54e−04	9.93e−01
KLRD1	−0.240	1.23e−05	**2.84e−02**	−0.145	7.39e−04	9.93e−01
FKBP11	−0.210	1.36e−05	**2.84e−02**	−0.159	1.31e−03	9.93e−01
GLNY	−0.451	1.75e−05	**3.20e−02**	−0.302	5.76e−04	9.93e−01
CLIC3	−0.415	2.07e−05	**3.36e−02**	−0.269	1.69e−03	9.98e−01
MANF	−0.176	2.69e−05	**3.93e−02**	−0.141	1.36e−03	9.93e−01

(a) Results obtained by the paired design, including technical and demographic covariates ("standard model").
(b) Results obtained by including in the standard model the cell composition score derived by cell deconvolution analysis with xCell as additional covariate. Only results for FDR < 0.05 from the standard model are shown in the table.
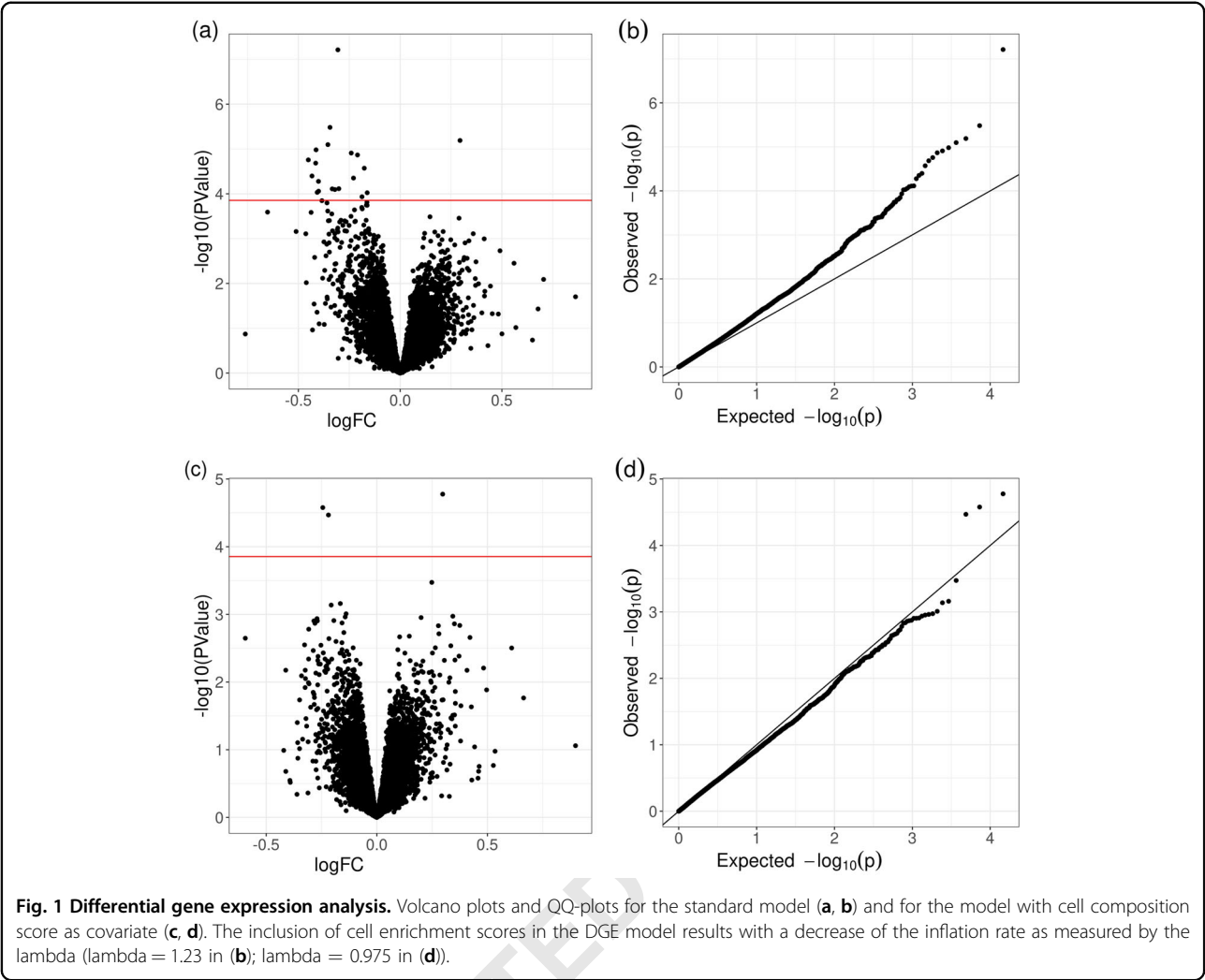
Fig. 1 Differential gene expression analysis. Volcano plots and QQ-plots for the standard model (**a**, **b**) and for the model with cell composition score as covariate (**c**, **d**). The inclusion of cell enrichment scores in the DGE model results with a decrease of the inflation rate as measured by the lambda (lambda = 1.23 in (**b**); lambda = 0.975 in (**d**)).

Table 3   Differential composition and cell type enrichment by using two different deconvolution approaches.

(a) CIBERSORT differential cell composition				(b) xCell differential cell composition			
Cell types	logFC enrich	*P* Value	FDR	Cell types	logFC enrich	*P* Value	FDR
NK.cells.resting	−0.0215673	3.66e−04	8.05e−03	NK cells	−0.0030001	7.45e−03	6.81e−02
Neutrophils	0.0340265	1.11e−02	1.22e−01	Tgd cells	−0.0008546	9.08e−03	6.81e−02
B.cells.naïve	0.0084004	4.30e−02	2.68e−01	naïve B cells	0.0038983	3.69e−02	1.39e−01
T.cells.CD8	−0.0134798	5.95e−02	2.68e−01	CD4+ Tem	−0.0029231	3.70e−02	1.39e−01
Macrophages.M0	0.0006095	6.10e−02	2.68e−01	B cells	0.0064601	5.44e−02	1.63e−01

(a) On the left, using CIBERSORT with LM 22 base matrix.
(b) On the right, using xCell algorithm FDR correction with the Benjamini–Hochberg method.

Effect of HLA alleles and KIR gene expression on NK cell proportion

Since a number of investigations has shown a potential relation between specific HLA allele groups and NK cell activity in the context of ASD[51−53], we have applied a recently developed computational tool which allows to perform in silico HLA typing (reported in Supplementary Table 6), and use the inferred HLA genotype to create a personalized index to quantify HLA expression[32]. We have then tested the correlation between HLA allele

Fig. 2 WGCNA Module analysis. a Module eigengene correlation with sample traits. b Expression heatmap for module 10—"brown". At the bottom, the eigengene values for each sample. Hierarchical clustering on the top shows three groups for NK cell composition, with the right group enriched for ASD samples. c Distribution of the module eigengene for brown module 10 and red module 13. The top panel shows a shift of the distribution peak on the left for autism and Asperger, while PDD-NOS are closer to controls.

expression and the enrichment score for NK cells obtained by XCell, focusing on the major HLA alleles for which expression was detected for more than 30% of the subjects. The results (supplementary Fig. 8a) suggest that the alteration in NK proportion has a poor correlation, and in most cases not significant, for every HLA allele tested. We explored in more details HLA-Cw7, an HLA allele previously associated with ASD[51], but we could not detect significant differences in NK cell estimate within diagnostic categories based on specific allele groups (see Supplementary Fig. 8b) nor in allele frequency between ASD subjects and unaffected related siblings (see Supplementary Table 6).

Finally, since genes coding for activating Killer-cell immunoglobulin-like receptor (KIR) proteins influence NK cell activity, and their frequency has been found to be significantly increased in ASD[54,55], we tested for the correlation between estimated NK cell proportion and the expression of KIR genes extracted from the original RNASeq dataset. The results (shown in Supplementary Figs. 9, 10) are consistent with the positive regulatory effect of KIR genes on NK cells, but do not suggest an increase frequency of specific KIR genes in ASD based on their expression. Therefore, putative differences in KIR gene frequency are unlikely to be at the basis of the observed altered NK cell proportion in ASD.

DGE analysis with cell composition estimate as covariate

We hypothesized that part of the unexplained variance in the previous DGE model could be driven by altered immune cell composition in ASD and not uniquely by transcriptome dysregulation. Thus, including an estimation of cell composition in the model would enable to uncover dysregulated gene expression withstanding the systematic imbalance in blood cell proportion between the discordant sibling pairs. Based on the results obtained with xCell, we selected the most differential enriched cell types ($p$value < 0.05, i.e. NK cells, Tgd, naïve β cells and CD4+ Tem), and included their scores for each individual sample as additional covariates in the model. As a result, the number of significantly de-regulated genes was largely decreased compared with the standard model, and the inflation previously observed the drastically reduced (see QQ plot in Fig. 1b, lambda = 0.975). As shown in Table 2 (b), when accounting for cell composition, only three genes (PTPRN2 and HGMB and NMUR1) remained significantly different at FDR < 0.25 (see also Supplementary Fig. 11). We then compared our results with a recently published meta-analysis on brain gene expression in autism[56]. Notably, whilst for NMUR1 and PTPRN2 we could not identify a statistically significant signal, for our top finding, HMGB3, we found a significant down-regulation in human brain tissues of autistic patients

(FDR = 0.03), with consistent downregulation in both cortex and cerebellum (See Supplementary Fig. 12).

WGCNA analysis

To corroborate the evidence of an effect of the imbalance in immune cell composition on the ASD gene expression signature, we conducted a further analysis based on WCGNA. Out of the network modules extracted by WGCNA, we identified two clusters of genes that are highly correlated with ASD diagnosis, and, at the same time, with the abundance of NK cell (ME10) or naïve Bcell (ME13) (see Fig. 2a). Whilst the expression heatmap constructed by using all expressed genes did not reveal any sample structure (not shown), the expression heatmap for ME10 obtained by unsupervised hierarchical clustering identifies three clusters, with the middle one (in blue) significantly enriched in ASD subjects (p < 0.003). The three clusters correspond to three groups with different NK cell proportion, as shown in Fig. 2b, top panel.

Interestingly, by looking at the distribution of the eigengene values for ME10 and ME13 across samples, differences between the three diagnostic subgroups can be observed (see Fig. 2c). In particular, a shift of the distribution peak for ME10 on the left was found for autism and Asperger, while PDD-NOS are closer to controls, suggesting that the decrease in NK cell proportion observed in ASD is not driven by PDD-NOS subjects. Similarly, the distribution of the eigengene values for ME13 is shifted towards the right for autism samples (but not for PDD-NOS and only marginally for Asperger), suggesting that the main difference in B cell composition is driven by autism subjects.

Discussion

Although previous applications of RNASeq efforts have been described in an integrated analysis of blood transcriptome[57] or in LCL[15], our report, to the best of our knowledge, describes the largest differential gene expression profiling in peripheral blood of ASD subjects conducted by RNASeq. In this study, we exploited the availability of a well characterized family collection and a discordant sibling design to uncover peripheral transcriptional signatures for ASD. The high degree of kinship between siblings and the shared environment should minimize differences in transcriptome not strictly related to ASD. We have used a robust model accounting for numerous demographic and technical covariates, and we have found relatively subtle differences between ASD cases and unaffected siblings, with most variance in gene expression being driven by the family component. Nevertheless, we were able to identify a gene expression signature, with ten DEGs below FDR threshold of 0.05. Our results are consistent with most published gene expression investigations by microarrays in the blood or

in blood cells of ASD subjects, where the number of DEG surviving to multi test correction is usually small, with some exceptions[21]. A recent meta-analysis[27], reports a list of more than 1500 DEGs significant at FDR level, however a more accurate inspection shows a substantial inflation (with a calculated lambda of 1.87), likely due to unexplained variance, which is not corrected by SVA analysis (lambda = 2.32, ca 2300 DEGs). A cell-specific investigation in leukocytes of ASD subjects[25] reports a list of ca 110 DEGs with FDR < 005, a relatively high number which might be somehow explained by the lack of covariates in the model. Indeed, when assessing differential expression in blood leukocytes in a sample of 118 subjects including part of the previous cohort, and adjusting for batch effects, sex, and RIN, no DEG resisted to multiple test correction for any pairwise group comparison[58]. A previous blood transcriptome analysis by microarrays on ASD discordant siblings reported the identification of a single gene with FDR < 0.05 after controlling for age, gender, and difference among families[22], in line with our findings.

Enrichment analysis on our ASD signature suggests a dysregulation of immune-inflammatory pathways, consistently with previous evidence of immune imbalance in ASD[59–62]. In particular, the top-enriched pathways are NK cell mediated cytotoxicity, and Immunoregulatory interactions between a lymphoid and a non-lymphoid cell, which also involves receptors and cell adhesion molecules playing a key role in modifying the response of B, T, and NK cells to antigens and pathogenic organisms. Interestingly, when searching for disease gene sets associated with signature, we identified Cytomegalovirus infection as the top hit. CMV infections have been associated with ASD since long, and recent evidences suggest a high prevalence of congenital CMV infection in ASD cases as measured by PCR using dried blood spots on filter paper[63]. In the Italian population, the infection rate was found to be about 10-fold higher in ASD than in the general population[64]. Given the known involvement of NK cells in congenital CMV infection[65], is it tempting to speculate a possible relationship between our findings and the increased prenatal exposure to viral infections, or other pathogens triggering similar mechanism, in ASD subjects.

The enrichment in immune-inflammatory pathways in blood gene expression signatures in ASD has been previously observed by other investigators[16,19,21,22,24], and confirmed by two recent meta-analyses of previous studies[27,66]. Studies conducted on whole blood RNA extracted with our same procedure (PAXgene) have shown a deregulation of immunoinflammatory pathways, such as chemokine signaling and FC gamma R-mediated phagocytosis[21,22]. In particular, chemokine signaling was among the top downregulated pathways when comparing blood samples from 20 discordant sibling pairs of the Simons Simplex Collection[21], consistent with our finding of a network of downregulated chemokines among top DEGs (See Supplementary Fig. 13). Of note, gene-expression classifiers of ASD identified based on microarray studies are enriched with genes related to immune response[19] or immune/inflammatory functions[25]. The ASD signature identified by the most comprehensive meta-analysis of blood transcriptome investigations is also characterized by downregulation of innate immune and inflammatory signaling transcripts[27], supporting the notion of overall decreased peripheral immune-inflammatory functions in ASD.

Deregulation of immune-inflammatory functions have been also reported in the brain of ASD subjects, in particular with the identification of an over-expressed module, named asdM16, enriched in astrocyte and activated microglial markers, and genes involved in immune and inflammatory responses[4], which was related to "activated" M2 microglial and "immune response" genes by a subsequent RNASeq-based investigation in ASD cortical brain tissue[5]. More recently, in the largest transcriptome analysis of post-mortem brain tissue in ASD so far, upregulated modules in ASD were found to be enriched in genes associated with inflammatory pathways[6]. Different hypotheses have been postulated to explain why blood and brain signatures in ASD converge; however with opposite deregulation, on immunological functions, including the contribution of decoupling mechanisms, tissue-related differences in feedback response or post-mortem related mechanisms, warranting further investigations[27].

To explore the immune cell repertoire in ASD subjects and estimate the effect of its dysregulation on the blood transcriptome, we exploited recent methods to infer cellular proportions from expression signatures of complex tissues. Using two complimentary methods, we were able to show a subtle, but statistically significant, decrease in NK cells paralleled by an increase in naïve B cells in ASD. The evidence from the cell deconvolution approach is corroborated by WGCNA analysis, which clearly identifies three clusters based on blood cell composition, possibly driven by different ASD subtypes.

Based on previous evidence of the effect of specific HLA allele groups and KIR gene variants on NK cell activity, we tested the hypothesis that differences in their frequency in ASD vs healthy controls could contribute to the alteration in NK cell proportion detected by cell deconvolution. We inferred HLA genotypes form RNASeq data and quantified expression of specific HLA alleles and KIR genes using recently developed pipelines. We could not correlate the alteration in NK cellular proportion between ASD and healthy siblings with any of the HLA alleles or KIR genes tested. Regarding HLA-Cw7, and other HLA allele groups having a role in stimulating NK cells, we could not

find a significant difference in frequency between ASD subject and matched siblings in our collection, differently from Harville et al.[51] who found an increased frequency in ASD vs unrelated subjects. Whilst we cannot rule out an increased overall increased activation of NK cell in ASD subjects, our data suggest a decrease in NK cell proportion over total blood cells according to signature deconvolution results irrespective from the HLA alleles or the KIR genes tested.

Our findings are consistent with the evidence of immune abnormalities in ASD, and the hypothesis of specific cellular immune-phenotypes related to different ASD subtypes[62,67]. Previous studies on NK cells in ASD, in particular on their proportion in blood and their activity, have produced mixed results, possibly due to the different CD markers used to determine cell counts. Two studies identified an increase in absolute number of NK cells in children with autism[17,68], whilst one report provides evidence of a decrease in CD57+ NK cells[69]. On the other hand, when investigating NK cell cytotoxic activity ex-vivo, there seems to be a convergence on findings on a reduction of NK cell activity[17,68,70], in agreement with early findings[71]. Of note, a recent work conducted in a group of adults with high functioning (hf) ASD reported no difference in the frequency of NK cells with respect to healthy controls; however, flow cytometric analysis revealed a hf-ASD signature based on NK cell-specific phenotypic markers, which was also associated with core ASD clinical dimensions[65]. Our work brings further evidence for a dysfunction in NK cell-related mechanisms in children with ASD. As suggested previously[17], abnormalities in NK cells may play a role in ASD etiology by predisposing to adverse neuroimmune interactions and/or autoimmunity mechanisms during critical development periods. Likewise, we observed an increase in naïve B cells in ASD children compared with their unaffected siblings. Increase in mature or memory B cells in ASD children was previously reported[72,73] even though most literature on B-cell number and function in ASD does not support significant abnormalities[62].

Whilst numerous peripheral blood gene expression investigations have highlighted dysregulation in immunoinflammatory pathways in ASD[68], little work has been conducted so far to investigate the consequence of immune cell imbalance in ASD signatures. In a recent transcriptome analysis of leukocytes in ASD, differences in proportion estimates of different leukocyte cell types were tested by using a deconvolution approach step[58]. No differences across diagnostic groups were identified; however, the authors used a different deconvolution algorithm and RNA extracted from whole blood after a cellular fractionation step[58]. A deconvolution analysis was also reported in a recent meta-analysis of blood ASD transcriptome investigations[27], which suggest increased expression of genes specific to NK cells and T-helper cells, consistent with earlier finding by Gregg et al.[16] showing gene expression differences in genes predominately associated with NK and CD8+ cells. The reported increase of NK cell-specific genes[27] is not consistent with our finding of reduced NK cells component. We are unable to provide further insights on the above discrepancy, since details on the results of the deconvolution analysis are not reported for the meta-analysis[27]. Given that our findings have been supported by two independent approaches, the disagreement with the previous report is unlikely to be due to methodological issues, but rather to differences in the demographic characteristics of our cohort or in the sibling pair design.

Finally, to identify genes that are differentially expressed irrespective from ASD-related differences in the proportion of circulating blood cells, we included the cell composition estimates from deconvolution analysis as covariates in our generalized linear model. Three genes survived to the DGE analysis when conditioning to cell composition scores, namely HMGB3, PTPRN2, and NMUR1.

HMGB3 was significantly downregulated in our study as well in human brain tissues of autistic patients. HMGB3 is a member of the high mobility group superfamily, and it has never been reported to be related to ASD, although there are several evidences for a role of a member of the same family, HMGB1, in ASD-associated inflammation[74].

We found a significant up-regulation of PTPRN2, coding for a Receptor-Type Protein Tyrosine Phosphatase N2, which is required for normal accumulation of secretory vesicles in hippocampus pancreatic islets and in the hippocampus[75]. PTPRN2 is also part of a set of protein complexes of which are tightly co-regulated during neuronal development[76]. A role for PTPRN2 in regulating brain development and function has been suggested, based on genetic disruptions linked to attention deficits, addiction and mood disorders, Down Syndrome and HOXA1 spectrum disorder[75] suggesting a pivotal role in regulating brain development and function. More recently, copy-number variations in PTPRN2 have been identified in children with developmental coordination disorder[77]. Of note, PTPRN2 has been shown to be a target of autoantibodies in Type I diabetics and a regulator of insulin secretion[78]. Since intrauterine hyperglycemia and neonatal hypoglycemia have been shown to be a risk factor for ASD[79], further investigations in cohorts with information on neonatal glycemia, maternal lifestyle or history of type I diabetes would be needed to test the hypothesis of a correlation between altered PTPRN2 expression in ASD and abnormalities in glucose or in prediabetes markers.

NMUR1 is the peripheral receptor for neuromedin, a neuropeptide with pleiotropic roles with multiple

functions. In particular, a key role for neuromedin in regulating food intake, circadian rhythms and inflammatory response has been described[80]. Of note, NMUR1 has been found to be downregulated in children from mothers exposed to psychosocial stress during pregnancy, with concomitant methylation of an enhancer region in the NMR1 gene dependent on the maternal stress score and altered immune response at birth[81]. Since neuromedin U activation of NMUR1 elicits production of cytokines by T-cells[82], further investigations can be envisaged to address a potential role for neuromedin U in mediating the interplay between neuroendocrine system and immune response in ASD.

Our study has a number of limitations and our results must be interpreted with caution. Although the discordant sibling design may help minimizing environmental or family-related confounders, our sample is rather heterogeneous, both in terms of ASD diagnosis and demographic factors, given our choice to include all available discordant pairs to maximize statistical power. Our statistical model, based on an empirical Bayes procedure, allowed us to estimate the beta coefficients for each technical and demographic covariate separately, including gender, ethnicity and diagnosis, resulting with a limited portion of unexplained variance, which can be fully ascribed to cell proportion according to QQ plot analysis. However, we cannot completely rule out subtle age, gender or ethnicity specific effects. This issue might be addressed by introducing gender-diagnosis, ethnicity-diagnosis or age-diagnosis interaction terms, which we opted out given the high complexity reached by of our generalized linear model. As far as gender is concerned, we have conducted a separate analysis for male sibling pairs, which corroborates our findings on the full cohort (the limited number of sibling pairs has hindered us to conduct a similar analysis on female subjects). Among demographic factors, a key element limiting the generalizability of our findings is the broad age range of our subjects, compared to some other investigations which have focused more specifically on toddlers or children within a narrow age range[19,24,25]. An Additional limitation is the lack of cell count data, which would have allowed an ultimate validation of our convergent findings of altered blood cell proportion in ASD, and placing full reliance on RNA Sequencing for the quantification of gene expression and for HLA genotyping. Finally, and more importantly, our findings should be seen as the result of an original investigation on a family cohort, which would need to be replicated in independent collections.

Despite the above limitations, taken together, our results suggest that subtle changes occur in the ASD blood transcriptome, which can be ascribed mainly to shifts in cellular composition, seemingly the major component driving gene expression changes when working with whole blood RNA. We identified three genes surviving cell deconvolution analysis which could be seen as novel suggestive biomarkers to be investigated in independent cohorts. Cell deconvolution analysis supports the hypothesis of a role of NK cells in the pathophysiology of autism, possibly related to increased prenatal exposure to infections in ASD subjects. While exploratory in nature, our findings are consistent with a growing body of evidence supporting immune-pathologies in ASD. The possibility of identifying subgroups with predominant immune system dysregulation, or to associate cellular immuno-phenotypes to different symptom dimensions might be a promising path forward in the development of non-invasive ASD biomarkers.

Author details
[1]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento (TN), Italy. [2]Roche Pharmaceutical Research and Early Development (pRED), Roche Innovation Center, Grenzacherstrasse 124, Basel, Switzerland. [3]UCB Pharma, Brussels, Belgium. [4]Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy. [5]Max Planck Institute of Psychiatry, Munich, Germany. [6]Fondazione The Microsoft Research - University of Trento Centre for Computational and Systems Biology (COSBI), Rovereto (TN), Italy

References
1. Abrahams, B. S. & Geschwind, D. H. Advances in autism genetics: on the threshold of a new neurobiology. *Nat. Rev. Genet.* **9**, 341–355 (2008).

2. de la Torre-Ubieta, L., Won, H., Stein, J. L. & Geschwind, D. H. Advancing the understanding of autism disease mechanisms through genetics. *Nat. Med.* **22**, 345–361 (2016).

3. Bourgeron, T. A synaptic trek to autism. *Curr. Opin. Neurobiol.* **19**, 231–234 (2009).

4. Voineagu, I. et al. Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).

5. Gupta, S. et al. Transcriptome analysis reveals dysregulation of innate immune response genes and neuronal activity-dependent genes in autism. *Nat. Commun.* **5**, 5748 (2014).

6. Parikshak, N. N. et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* **540**, 423–427 (2016).

7. Baron, C. A., Liu, S. Y., Hicks, C. & Gregg, J. P. Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. *J. Autism Dev. Disord.* **36**, 973–982 (2006).

8. Hu, V. W., Frank, B. C., Heine, S., Lee, N. H. & Quackenbush, J. Gene expression profiling of lymphoblastoid cell lines from monozygotic twins discordant in severity of autism reveals differential regulation of neurologically relevant genes. *BMC Genomics.* **7**, 118 (2006).

9. Nishimura, Y. et al. Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways †. *Hum. Mol. Genet.* **16**, 1682–1698 (2007).

10. Hu, V. W. et al. Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Reses.* **2**, 78–97 (2009).

11. Hu, V. W. et al. Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. *PLoS ONE* **4**, e5775 (2009).

12. Luo, R. et al. Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. *Am. J. Hum. Genet.* **91**, 38–55 (2012).

13. Chien, W.-H. et al. Increased gene expression of FOXP1 in patients with autism spectrum disorders. *Mol. Autism* **4**, 23 (2013).

14. Talebizadeh, Z., Aldenderfer, R. & Wen Chen, X. A proof-of-concept study. *Psychiatr. Genet.* **24**, 1–9 (2014).

15. Tylee, D. S. et al. RNA sequencing of transformed lymphoblastoid cells from siblings discordant for autism spectrum disorders reveals transcriptomic and functional alterations: evidence for sex-specific effects. *Autism Res.* **10**, 439–455 (2017).

16. Gregg, J. P. et al. Gene expression changes in children with autism. *Genomics* **91**, 22–29 (2008).

17. Enstrom, A. M. et al. Altered gene expression and function of peripheral blood natural killer cells in children with autism. *Brain Behav. Immun.* **23**, 124–133 (2009).

18. Alter, M. D. et al. Autism and increased paternal age related changes in global levels of gene expression regulation. *PLoS ONE* **6**, e16715 (2011).

19. Glatt, S. J. et al. Blood-based gene expression signatures of infants and toddlers with autism. *J. Am. Acad. Child Adolesc. Psychiatry* **51**, 934–944.e2 (2012).

20. Kuwano, Y. et al. Autism-associated gene expression in peripheral leucocytes commonly observed between subjects with autism and healthy women having autistic children. *PLoS ONE* **6**, e24723 (2011).

21. Kong, S. W. et al. Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PLoS ONE* **7**, e49475 (2012).

22. Kong, S. W. et al. Peripheral blood gene expression signature differentiates children with autism from unaffected siblings. *Neurogenetics* **14**, 143–152 (2013).

23. Ivanov, H., Stoyanova, V., Popov, N., Bosheva, M. & Vachev, T. Blood-based gene expression in children with autism spectrum disorder. *Biodiscovery* **17**, 2 (2015).

24. Pramparo, T. et al. Cell cycle networks link gene expression dysregulation, mutation, and brain maldevelopment in autistic toddlers. *Mol. Syst. Biol.* **11**, 841 (2015).

25. Pramparo, T. et al. Prediction of autism by translation and immune/inflammation coexpressed genes in toddlers from pediatric community practices. *JAMA Psychiatry* **72**, 386 (2015).

26. Ansel, A., Rosenzweig, J. P., Zisman, P. D., Melamed, M. & Gesundheit, B. Variation in gene expression in autism spectrum disorders: an extensive review of transcriptomic studies. *Front Neurosci.* **10**, 601 (2016).

27. Tylee, D. S. et al. Blood transcriptomic comparison of individuals with and without autism spectrum disorder: a combined-samples mega-analysis. *Am. J. Med. Genet. B: Neuropsychiatr. Genet.* **174**, 181–201 (2016).

28. Muglia, P. et al. The Italian Autism Network (ITAN): a resource for molecular genetics and biomarker investigations. *BMC Psychiatry* **18**, 369 (2018).

29. American Psychiatric Association. *Diagnostic and Statistical manual of Mental Disorders* (American Psychiatric Association, 1994).

30. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).

31. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

32. Aguiar, V. R. C., César, J., Delaneau, O., Dermitzakis, E. T., Meyer, D. Expression estimation and eQTL mapping for HLA genes with a personalized pipeline. *PLOS Genet.* **15**, e1008091 (2019).

33. IMGTHLA database [Internet]. Available from: https://github.com/ANHIG/IMGTHLA

34. KIR genotype databse [Internet]. Available from: https://www.ebi.ac.uk/ipd/kir/

35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).

36. Robinson, M. D. & Smyth, G. K. Moderated statistical tests for assessing differences in tag abundance. *Bioinformatics* **23**, 2881–2887 (2007).

37. Xie, C. et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Res.* **39**, W316–W322 (2011).

38. Open Target Portal [Internet]. Available from: www.targetvalidation.org

39. Consortium S. String, D. B. [Internet]. Available from: https://string-db.org

40. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).

41. Newman, A. M., Liu, C. L., Alizadeh, A. A. CIBERSORT [Internet]. Available from: https://cibersort.stanford.edu/

42. Aran, D., Hu, Z. & Butte, A. J. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* **18**, 220 (2017).

43. Aran, D. xCell [Internet]. Available from: https://github.com/dviraran/xCell

44. Bhattacharya, S. et al. ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.* **58**, 234–239 (2014).

45. ImmPort Database [Internet]. Available from: https://www.immport.org

46. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

47. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* **9**, 559 (2008).

48. Langfelder, P., Horvath, S. WGCNA: an R package for weighted correlation network analysis [Internet]. Available from: https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/

49. Lee, S. C. et al. Solving for X: evidence for sex-specific autism biomarkers across multiple transcriptomic studies. *Am. J. Med. Genet. B: Neuropsychiatr. Genet.* **180**, 377–389 (2018)

50. Pachnio, A. et al. Cytomegalovirus infection leads to development of high frequencies of cytotoxic virus-specific CD4+ T cells targeted to vascular endothelium. *PLOS Pathog.* **12**, e1005832 (2016).

51. Harville, T. et al. Inheritance of HLA-Cw7 associated with autism spectrum disorder (ASD). *Front Psychiatry* **10**, 612 (2019)

52. Torres, A. R., Westover, J. B. & Rosenspire, A. J. HLA immune function genes in autism. *Autism Res. Treat.* **2012**, 959073 (2012).

53. Bennabi, M. et al. HLA-class II haplotypes and autism spectrum disorders. *Sci. Rep.* **8**, 7639 (2018).

54. Torres, A. R., Westover, J. B., Gibbons, C., Johnson, R. C. & Ward, D. C. Activating killer-cell immunoglobulin-like receptors (KIR) and their cognate HLA ligands are significantly increased in autism. *Brain Behav. Immun.* **26**, 1122–1127 (2012).

55. Torres, A. R. et al. Common genetic variants found in HLA and KIR immune genes in autism spectrum disorder. *Front Neurosci.* **10**, 463 (2016).

56. Zhang, S. et al. dbMDEGA: a database for meta-analysis of differentially expressed genes in autism spectrum disorder. *BMC Bioinform.* **18**, 494 (2017).

57. Codina-Solà, M. et al. Integrated analysis of whole-exome sequencing and transcriptome profiling in males with autism spectrum disorders. *Mol. Autism* **6**, 21 (2015).

58. Lombardo, M. V. et al. Large-scale associations between the leukocyte transcriptome and BOLD responses to speech differ in autism early language outcome subtypes. *Nat. Neurosci.* **21**, 1680–1688 (2018).

59. Gesundheit, B. et al. Immunological and autoimmune considerations of autism spectrum disorders. *J. Autoimmun.* **44**, 1–7 (2013).

60. Masi, A., Glozier, N., Dale, R. & Guastella, A. J. The immune system, cytokines, and biomarkers in autism spectrum disorder. *Neurosci. Bull.* **33**, 194–204 (2017).

61. Meltzer, A. & Van de Water, J. The role of the immune system in autism spectrum disorder. *Neuropsychopharmacology* **42**, 284–298 (2017).

62. Gładysz, D., Krzywdzińska, A., Hozyasz, K. K. Immune abnormalities in autism spectrum disorder—could they hold promise for causative treatment? *Mol. Neurobiol.* **55**, 6387–6435 (2018).

63. Maeyama, K. et al. Congenital cytomegalovirus infection in children with autism spectrum disorder: systematic review and meta-analysis. *J. Autism Dev. Disord.* **48**, 1483–1491 (2018).

64. Gentile, I. et al. Prevalence of congenital cytomegalovirus infection assessed through viral genome detection in dried blood spots in children with autism spectrum disorders. *Vivo* **31**, 467–473 (2017).

65. Goodier, M. R., Jonjić, S., Riley, E. M. & Juranić Lisnić, V. CMV and natural killer cells: shaping the response to vaccination. *Eur. J. Immunol.* **48**, 50–65 (2018).

66. Ning, L. F. et al. Meta-analysis of differentially expressed genes in autism based on gene expression data. *Genet Mol. Res.* **14**, 2146–2155 (2015).

67. Careaga, M. et al. Immune endophenotypes in children with autism spectrum disorder. *Biol. Psychiatry* **81**, 434–441 (2017).

68. Ashwood, P. et al. In search of cellular immunophenotypes in the blood of children with autism. *PLoS ONE* **6**, e19299 (2011).

69. Siniscalco, D. et al. Decreased numbers of CD57+CD3− cells identify potential innate immune differences in patients with autism spectrum disorder. *Vivo (Brooklyn).* **30**, 83–89 (2016).

70. Vojdani, A. et al. Low natural killer cell cytotoxic activity in autism: the role of glutathione, IL-2 and IL-15. *J. Neuroimmunol.* **205**, 148–154 (2008).

71. Warren, R. P., Foster, A. & Margaretten, N. C. Reduced natural killer cell activity in autism. *J. Am. Acad. Child Adolesc. Psychiatry* **26**, 333–335 (1987).

72. Hollander, E. et al. B lymphocyte antigen D8/17 and repetitive behaviors in autism. *Am. J. Psychiatry* **156**, 317–320 (1999).

73. Wasilewska, J., Kaczmarski, M., Stasiak-Barmuta, A., Tobolczyk, J. & Kowalewska, E. Low serum IgA and increased expression of CD23 on B lymphocytes in peripheral blood in children with regressive autism aged 3-6 years old. *Arch. Med Sci.* **8**, 324–331 (2012).

74. Dipasquale, V. et al. Neuroinflammation in autism spectrum disorders: role of high mobility group box 1 protein. *Int J. Mol. Cell Med.* **6**, 148–155 (2017).

75. Ramírez-Franco, J. J., Munoz-Cuevas, F. J., Luján, R. & Jurado, S. Excitatory and inhibitory neurons in the hippocampus exhibit molecularly distinct large dense core vesicles. *Front Cell Neurosci.* **10**, 202 (2016).

76. Frese, C. K. et al. Quantitative map of proteome dynamics during neuronal differentiation. *Cell Rep.* **18**, 1527–1542 (2017).

77. Mosca, S. J. et al. Copy-number variations are enriched for neurodevelopmental genes in children with developmental coordination disorder. *J. Med. Genet.* **53**, 812 LP–812819 (2016).

78. Caromile, L. A., Oganesian, A., Coats, S. A., Seifer, R. A. & Bowen-Pope, D. F. The neurosecretory vesicle protein phogrin functions as a phosphatidylinositol phosphatase to regulate insulin secretion. *J. Biol. Chem.* **285**, 10487–10496 (2010).

79. Hoirisch-Clapauch, S., Nardi, A. E. Autism spectrum disorders: let's talk about glucose? *Transl. Psychiatry* **9**, 51 (2019).

80. Martinez, V. G., O'Driscoll, L. & Neuromedin, U. A multifunctional neuropeptide with pleiotropic roles. *Clin. Chem.* **61**, 471–482 (2015).

81. Trump, S. et al. Prenatal maternal stress and wheeze in children: novel insights into epigenetic regulation. *Sci. Rep.* **6**, 28616 (2016).

82. Johnson, E. N. et al. Neuromedin elicits cytokine release in murine Th2-type T cell clone D10.G4.1. *J. Immunol.* **173**, 7230–7238 (2004).

Author details

Italian Autism Network

Giovanni Alibrio[7], Laura Anchisi[8], Milena Andruccioli[9], Arianna Benvenuto[10], Pier Antonio Battistella[11], Flavio Boscaini[12], Carmela Bravaccio[13], Elisa Ceppi[14], Diego Cosentino[15], Paolo Curatolo[16], Lucio Da Ros[17], Bernardo Dalla Bernardina[18], Andrea De Giacomo[19], Giuseppa Di Vita[20], Enrico Domenici[21], Massimo Elia[22], Filippo Gitti[23], Serenella Grittani[24], Anna Linda Lamanna[25], Elisa Mani[26], Barbara Manzi[27], Lucia Margari[28], Gabriele Masi[29], Massimo Molteni[30], Pierandrea Muglia[31], Franco Nardocci[32], Antonio Pascotto[33], Antonia Parmeggiani[34], Pier Franco Pignatti[35], Tiziana Piroddi[36], Paola Prandini[37], Emiliangelo Ratti[38], Paolo Rizzini[39], Sebastiano Russo[40], Renato Scifo[41], Raffaella Tancredi[42], Alessandra Tiberti[43], Elisabetta Trabetti[44], Leonardo Zoccante[45] & Alessandro Zuddas[46]

[7]Azienda Ospedaliera Spedali Civili di Brescia, Brescia, Italy. [8]Università degli Studi di Cagliari, Centro per lo Studio delle Terapie Farmacologiche in Neuropsichiatria dell'infanzia e dell'adolescenza, Cagliari, Italy. [9]Ospedale Infermi, Divisione Neuropsichiatria Infantile, Centro per l'autismo, Rimini, Italy. [10]Neuropsichiatria Infantile, Università Tor Vergata, Policlinico Tor Vergata, Roma, Italy. [11]Neurologia Pediatrica, Dipartimento di Pediatria, Università di Padova, Padova, Italy. [12]Servizio di Neuropsichiatria Infantile, Azienda Ospedaliera, Istituti Ospitalieri di Verona, Policlinico G.B. Rossi, Verona, Italy. [13]Neuropsichiatria Infantile - Seconda Università degli Studi di Napoli, Napoli, Italy. [14]IRCCS Eugenio Medea, Monza, Italy. [15]Fondazione SmithKline, Verona, Italy. [16]Neuropsichiatria Infantile, Università "Tor Vergata", Policlinico Tor Vergata, Roma, Italy. [17]Fondazione FSK, Verona, Italy. [18]Servizio di Neuropsichiatria Infantile, Azienda Ospedaliera, Istituti Ospitalieri di Verona, Policlinico G.B. Rossi, Verona, Italy. [19]Unità Operativa di Neuropsichiatria Infantile, Dipartimento di Scienze Neurologiche e Psichiatriche, Università degli Studi Bari, Bari, Italy. [20]IRCCS Associazione Oasi Maria SS., Troina (Enna), Italy. [21]Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy. [22]IRCCS Associazione Oasi Maria SS., Troina (Enna), Italy. [23]Azienda Ospedaliera Spedali Civili di Brescia, Brescia, Italy. [24]Ospedale Infermi - Divisione Neuropsichiatria Infantile, Centro per l'autismo, Rimini, Italy. [25]Unità Operativa di Neuropsichiatria Infantile, Dipartimento di Scienze Neurologiche e, Università degli Studi di Bari, Bari, Italy. [26]IRCCS Eugenio Medea, Monza, Italy. [27]Neuropsichiatria Infantile, Università Tor Vergata, Policlinico Tor Vergata, Roma, Italy. [28]Unità Operativa di Neuropsichiatria Infantile, Dipartimento di Scienze Neurologiche e Psichiatriche, Università degli Studi di Bari, Bari, Italy. [29]IRCCS Stella Maris, Pisa, Italy. [30]IRCCS Eugenio Medea, Monza, Italy. [31]UCB Pharma, Brussels, Belgium. [32]Neuropsichiatria infantile, Modena, Italy. [33]Neuropsichiatria Infantile, Seconda Università degli Studi di Napoli, Napoli, Italy. [34]Dipartimento di Scienze Mediche e Chirurgiche, Università di Bologna, Bologna, Italy. [35]Universita' di Verona, Verona, Italy. [36]Ospedale Infermi, Divisione Neuropsichiatria Infantile, Centro per l'autismo, Rimini, Italy. [37]Universita' di Verona, Verona, Italy. [38]Central Nervous System Therapeutic Area Unit, Takeda, Boston, USA. [39]Fondazione FSK, Verona, Italy. [40]Neuropsichiatra Infantile, AUSL 3 Catania, Acireale (Catania), Catania, Italy. [41]Neuropsichiatria Infantile, AUSL 3 Catania, Acireale (Catania), Catania, Italy. [42]IRCCS Stella Maris, Pisa, Italy. [43]Azienda Ospedaliera Spedali Civili di Brescia, Brescia, Italy. [44]Department of Neurosciences, Biomedicine and Movement Sciences, Section of Biology and Genetics, University of Verona, Verona, Italy. [45]Servizio di Neuropsichiatria Infantile, Azienda Ospedaliera, Istituti Ospitalieri di Verona, Policlinico G.B. Rossi, Verona, Italy. [46]Università degli Studi di Cagliari, Centro per lo Studio delle Terapie Farmacologiche in Neuropsichiatria dell'infanzia e dell'adolescenza, Cagliari, Italy

QUERY FORM

TRANSLPSYCHIATRY	
Manuscript ID	**[Art. Id: 778]**
Author	
Editor	
Publisher	

Journal: TRANSLPSYCHIATRY

Author :- The following queries have arisen during the editing of your manuscript. Please answer by making the requisite corrections directly in the e.proofing tool rather than marking them up on the PDF. This will ensure that your corrections are incorporated accurately and that your paper is published as quickly as possible.

Query No.	Description	Author's Response
AQ1	Please check your article carefully, coordinate with any co-authors and enter all final edits clearly in the eproof, remembering to save frequently. Once corrections are submitted, we cannot routinely make further changes to the article.	
AQ2	Note that the eproof should be amended in only one browser window at any one time; otherwise changes will be overwritten.	
AQ3	Author surnames have been highlighted. Please check these carefully and adjust if the first name or surname is marked up incorrectly. Note that changes here will affect indexing of your article in public repositories such as PubMed. Also, carefully check the spelling and numbering of all author names and affiliations, and the corresponding email address(es).	
AQ4	You cannot alter accepted Supplementary Information files except for critical changes to scientific content. If you do resupply any files, please also provide a brief (but complete) list of changes. If these are not considered scientific changes, any altered Supplementary files will not be used, only the originally accepted version will be published.	