MIT Open Access Articles

Peripheral blood gene expression signature differentiates children with autism from unaffected siblings

Massachusetts Institute of Technology

Peripheral blood gene expression signature differentiates children with autism from unaffected siblings

S. W. Kong · Y. Shimizu-Motohashi · M. G. Campbell ·
I. H. Lee · C. D. Collins · S. J. Brewster · I. A. Holm ·
L. Rappaport · I. S. Kohane · L. M. Kunkel

Abstract Autism spectrum disorder (ASD) is one of the most prevalent neurodevelopmental disorders with high heritability, yet a majority of genetic contribution to pathophysiology is not known. Siblings of individuals with ASD are at increased risk for ASD and autistic traits, but the genetic contribution for simplex families is estimated to be less when compared to multiplex families. To explore the genomic (dis-) similarity between proband and unaffected sibling in simplex families, we used genome-wide gene expression profiles of blood from 20 proband-unaffected sibling pairs and 18 unrelated control individuals. The global gene expression profiles of unaffected siblings were more similar to those from probands as they shared genetic and environmental background. A total of 189 genes were significantly differentially expressed between proband-sib pairs (nominal $p<0.01$) after controlling for age, sex, and family effects. Probands and siblings were distinguished into two groups by cluster analysis with these genes. Overall, unaffected siblings were equally distant from the centroid of probands and from that of unrelated controls with the differentially expressed genes. Interestingly, five of 20 siblings had gene expression profiles that were more similar to unrelated controls than to their matched probands. In summary, we found a set of genes that distinguished probands from the unaffected siblings, and a subgroup of unaffected siblings who were more similar to probands. The pathways that characterized probands compared to siblings using peripheral blood gene expression profiles were the up-regulation of ribosomal, spliceosomal, and mitochondrial pathways, and the down-regulation of neuroreceptor-ligand, immune response and calcium signaling pathways. Further integrative study with structural genetic variations such as de novo mutations, rare variants, and copy number variations would clarify whether these transcriptomic changes are structural or environmental in origin.

Keywords Autism spectrum disorders · Unaffected sibling · Simplex family · Gene expression

S. W. Kong · M. G. Campbell · I. H. Lee · I. S. Kohane
Informatics Program at the Harvard–Massachusetts Institute of Technology Division of Health Sciences and Technology, Boston Children's Hospital, Boston, MA 02115, USA

Y. Shimizu-Motohashi · C. D. Collins · S. J. Brewster ·
I. A. Holm · L. M. Kunkel (✉)
Division of Genetics, Program in Genomics, Boston Children's Hospital, Boston, MA 02115, USA
e-mail: kunkel@enders.tch.harvard.edu

L. Rappaport
Division of Developmental Medicine, Boston Children's Hospital, Boston, MA 02115, USA

I. S. Kohane
Center for Biomedical Informatics, Harvard Medical School, Boston, MA 02115, USA

Introduction

Autism spectrum disorders (ASD) comprise a group of neurodevelopmental disorders with high heritability. An increased risk for siblings and twins has been well reported in previous studies; however, the genetics of ASD has become increasingly complex from genome-wide association, linkage analysis, copy number variation (CNV), and whole-genome and -exome sequencing studies [1–4]. Altogether, the genetic

contribution from known de novo mutations, single nucleotide polymorphisms (SNPs) and CNVs explains no more than 20 % of cases. Moreover, shared environmental effects seem to play a more significant role in co-morbid fraternal twins compared to genetic factors [5].

The sibling recurrence risk of autism has been estimated to be between 3 % and 10 % [6, 7], and a recent prospective study revealed that the sibling recurrence rate of ASD is higher than suggested by previous estimates [5]. In that study, a total of 18.7 % of infant siblings developed ASD. Specifically, male gender and presence of at least one affected sibling were independent and significant predictors of an ASD outcome, with a 2.8-fold increase in the risk for ASD for male infants compared to female infants from simplex families (families with only one affected child, and unaffected parents and siblings) and an additional 2.2-fold increase for all children, regardless of gender, in multiplex families (families with more than one affected child, and unaffected parents and siblings) [8].

Phenotypically, autistic traits and endophenotypes of ASD are more frequently observed in unaffected siblings and parents of children with ASD in simplex families than in the unrelated control population [9–11]. Together with recent results from CNV and exome sequencing studies showing an increase in the rate of gene disrupting de novo mutation in probands compared to their unaffected siblings in simplex families [12–14], this has led to a genetic model of ASD risk that posits the combinatorial effect of common and rare variants including de novo CNVs and mutations [15]. In this model, common variants constitute genetic background that is shared in unaffected siblings and relatives, and de novo genetic events or environmental effects trigger the pathophysiology of ASD. Importantly, this model allows for a spectrum of ASD phenotypes in relatives due to the contribution of common variants.

Gene expression studies using peripheral blood cells and lymphoblastic cell lines have shown that genome-wide gene expression profiles differ between ASD cases and non-cases [16–23], suggesting transcriptomic signatures from peripheral blood could be used as a surrogate for understanding the genetics of ASD. To this end, we recently reported that a blood-based gene expression signature was able to classify the males with ASD from unrelated controls with greater than 70 % of accuracy in two independently collected cohorts [16]. Glatt and colleagues [17] reported a transcriptomic diagnostic signature of ASD compared to typically developing children using peripheral blood mononuclear cells. Luo and colleagues [1] found that outlier expression levels from lymphoblastic cell lines were highly correlated with structural genomic changes such as CNVs, but did not find significant differences in overall numbers of outlier genes between simplex cases and unaffected siblings. Based on the above evidence, we explored whether probands have a different functional genomic signature that is a snapshot of the combined effect of genetic and environmental factors compared to their unaffected siblings. We used peripheral blood gene expression profiles of the probands and unaffected siblings from the Simons Simplex Collection (SSC) to explore the (dis-) similarity of probands and siblings compared to unrelated controls, and to identify what genes and pathways differentiate probands from their unaffected siblings.

Materials and methods

Probands and siblings from the Simons Simplex Collection

Blood samples of 20 probands and their unaffected sibling pairs were collected from the SSC (Table 1). Five proband-sib pairs were of the same gender, i.e., males, while 15 pairs were of the opposite gender including 12 male and three female probands. These blood samples were prepared for genome-wide gene expression profiling using microarrays in one batch to minimize batch effects. The age range was 4.1–17 years (mean age 10.15 years) in probands and 6–17 years (mean age 10.96 years) in unaffected siblings, and there was no significant difference in age distribution between the two groups (paired t-test, $p=0.36$). All subjects were evaluated at Boston Children's Hospital and probands were diagnosed upon Autism Diagnostic Observation Schedule and Autism Diagnostic Interview Revised data according to the SSC criteria. No known genetic or syndromic disorders were described in these individuals. Blood samples from 18 individuals (11 males and seven females) unrelated to the SSC proband-sibling families were enrolled in this study as controls (Table 1). The age ranged from 4.1 to 16.1 years (mean age 9.69 years) and there was no significant difference in age distribution compared to probands or sibs. These control samples were collected from healthy patients seen in Children's Hospital Primary Care Center for a well child visit who had a routine blood draw. Gene expression profiling of these blood samples were processed in a different batch from the SSC samples. If the child had a chronic disease such as infectious disease, diabetes, cardiovascular disease, and developmental disorder or neurological disorder, they were excluded from this study. A male control individual had a history of delayed speech; however, he did not meet the diagnostic criteria of any language disorders. All SSC participants and control individuals were recruited with Institutional Review Board approval.

Table 1 Clinical and demographic characteristics of 20 simplex families and 18 unrelated controls

Characteristic	Probands	Unaffected siblings	Unrelated controls
Number	20	20	18
Age (years)			
Mean	10.2	11.0	9.7
Range	4.1–17.0	6.0–17.0	4.1–16.1
Gender			
Male	17	8	11
Female	3	12	7
Race			
White	20	20	3
Black	0	0	3
Other	0	0	12
Ethnicity			
Hispanic	2	2	9
Non-Hispanic	18	18	9
Diagnosis			
Autistic disorder	9	–	–
PDD-NOS	11	–	–
Developmental delays	0	4	1
Neurological disorders	0	1	0
Psychiatric disorders	4	1	0
Immune disorders	3	4	0
Gastrointestinal disorders	1	2	0

Genome-wide gene expression profiling with microarrays

Gene expression profiles were prepared using Affymetrix Gene 1.0 ST (GeneST) arrays (Affymetrix, California). Total RNA was extracted from 2.5 ml of whole venous blood using the PAXgene Blood RNA System (PreAnalytix) according to the manufacturer's instructions. Quality and quantity of these RNAs was assessed using the Nanodrop spectrophotometer (Thermo Scientific) and Bioanalyzer System (Agilent). Fragmented cRNA was hybridized to the appropriate Affymetrix array and scanned on an Affymetrix GeneChip scanner 3000.

Preprocessing of microarray data and statistical analysis

Gene expression levels were calculated using Affymetrix Power Tools version 1.10 (Affymetrix). We used the Probe Log Iterative ERror (PLIER) algorithm, and subsequent normalization using a locally weighted scatterplot smoothing (LOWESS) algorithm was performed to reduce the variability between microarrays in two batches of SSC and unrelated control samples. Differentially expressed genes between probands and sib pairs were identified by fitting a linear model for each probeset after controlling for age, sex,

and difference among families. We used the Storey–Tibshirani's method to estimated false discovery rates (FDRs) for multiple hypothesis tests [24].

To explore global gene expression patterns across all samples, we applied a variance filtering to exclude probesets for which expression levels were relatively consistent across all samples. There were 8,326 probesets that had higher than 75 percentile of variance across all individuals. Using these probesets, to measure similarity between the gene expression profiles from all individuals, the squared Mahalanobis distances (MDs) from the centroid of probands and that of unrelated controls were calculated for each sib with the pooled covariance matrix estimated using cov.shrink function in the corpcor R library [25]. We used the squared MDs as surrogates to quantitatively measure the similarity of sibs to probands and unrelated controls separately. Then, the distances from each centroid were compared using a paired t-test. After identifying subgroups of proband-sibling pairs that were either similar or divergent using unrelated controls as a benchmark, we performed a series of paired and unpaired t-tests with unequal variance to determine sets of differentially expressed genes. Principal component analysis (PCA) of differentially expressed genes was performed on centered (mean = 0) and scaled (variance 1) expression values. Hierarchical clustering on differentially expressed genes was performed with complete linkage on Euclidean distance. To determine enriched pathways among differentially expressed genes, we applied Gene Set Enrichment Analysis (GSEA) [26]. With the exception of GSEA, all statistical analyses were performed using the R statistical programming language, version 2.15 (http://cran.r-project.org/).

Results

Global gene expression patterns of probands, siblings, and unrelated controls

PCA of 58 samples using all probesets showed no significant batch effect or gender bias in global gene expression profiles (Fig. S1); however, there was significant difference between probands and the other groups. We calculated overall similarity of gene expression patterns using the squared MD. Similar to the PCA result, MDs from the common pooled centroid of all 58 samples did not show any difference between groups; however, the distribution of MDs from the centroid of probands and from that of unrelated controls revealed that the probands, unaffected siblings, and unrelated controls formed distinct clusters (Fig. 1). Unaffected siblings seemed to form a separate cluster from that of probands when the centroid of probands was used to calculate MD of each sample. Similarly, the distribution of MDs from the centroid of unrelated controls showed that probands and siblings were similar relative to

Fig. 1 Differences in global gene expression profiles using squared Mahalanobis distances from the centroids of probands and unrelated controls. Overall gene expression profiles were not different between probands, siblings, and unrelated controls by principal component analysis (Fig. S1). We calculated a multivariate distance, i.e., squared Mahalanobis distance (*MD*), of each individual sample from the centroid of unrelated controls and that of probands. The density plot on top of the scatter plot shows that probands and siblings are equally distant from unrelated controls. Interestingly, the distributions of MDs from the centroid of probands show significant difference ($p=6.6\times10^{-9}$). This result suggests that there seems to be subtle but significant difference in transcriptomic signature between proband and siblings
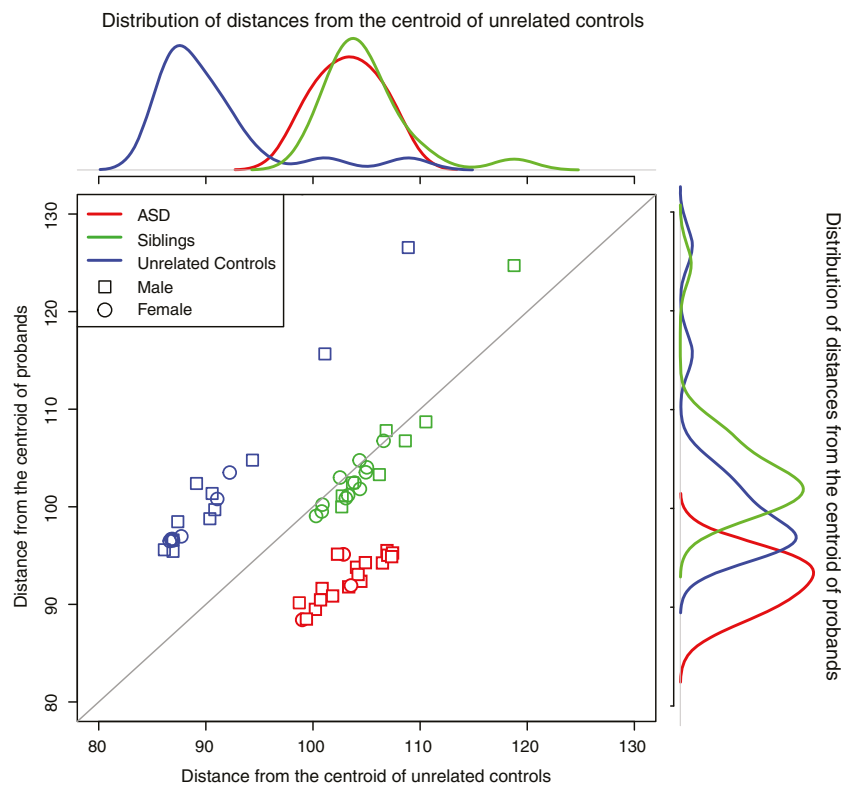
unrelated controls. Together, these observations suggested a difference between probands and siblings in gene expression patterns, and we sought to find the genes that differentiated probands from their unaffected siblings.

Differentially expressed genes in probands compared to their unaffected siblings

We fit a linear model to the gene expression levels for each probeset to prioritize consistently differentially expressed genes between proband and sibling across SSC families after controlling for age and sex. We used uncorrected p values to prioritize significant genes since only one gene — *IL15RA* — was significant at FDR <1 %. Among 269 differentially expressed genes between probands and sibs significant at uncorrected $p<0.01$, there were genes significant by the other factors such as age (29 genes), gender (28 genes), and between families (35 genes) at the same significance level. After excluding the genes changed by these other factors, 189 probesets that represent 163 unique genes including two previously reported ASD candidate genes — *CTNNB1* and *XPO1* — were significantly changed between probands and siblings (uncorrected $p<0.01$) (Table S1). Eighty-four probesets were up-regulated in probands compared to unaffected siblings.

These 189 probesets were able to cluster probands and siblings except for one sibling using hierarchical clustering suggesting that these genes could be informative in classifying probands from unaffected siblings (Fig. 2a). We further explored whether this clustering of probands was confounded by the other factors such as gender and age. PCA of 20 proband-sibling pairs with 189 probesets found probands and unaffected siblings were mostly separated by the first principal component that explained 32.8 % of total variance (Fig. 2b). Age, gender, and familial difference were not significantly correlated with any principal components. Thus, the gene expression profile of 189 probesets was able to distinguish probands from that of unaffected siblings.

Gene expression profiles reveal proband-like and control-like siblings compared to probands

Previous studies report that the recurrence risk of ASD in siblings could be as high as 10–20 % [27], and unaffected siblings and relatives may have autistic traits [9–11]. However, in simplex families, the prevalence of autistic phenotype in unaffected siblings and parents is not yet known. We explored whether a subgroup of unaffected siblings were similar to probands in transcriptomic changes compared to unrelated controls.

To estimate the similarity of global gene expression profiles between probands, unaffected siblings, and unrelated controls, we calculated the multivariate distances of each sibling with respect to the two centroids of probands and unrelated controls clusters (see Materials and methods). The unaffected siblings were more similar to probands overall (paired *t*-test, $p=0.052$). However, as shown in Fig. 3a, five

a

b

PCA of 20 proband-sibling pairs
with the differentially expressed genes
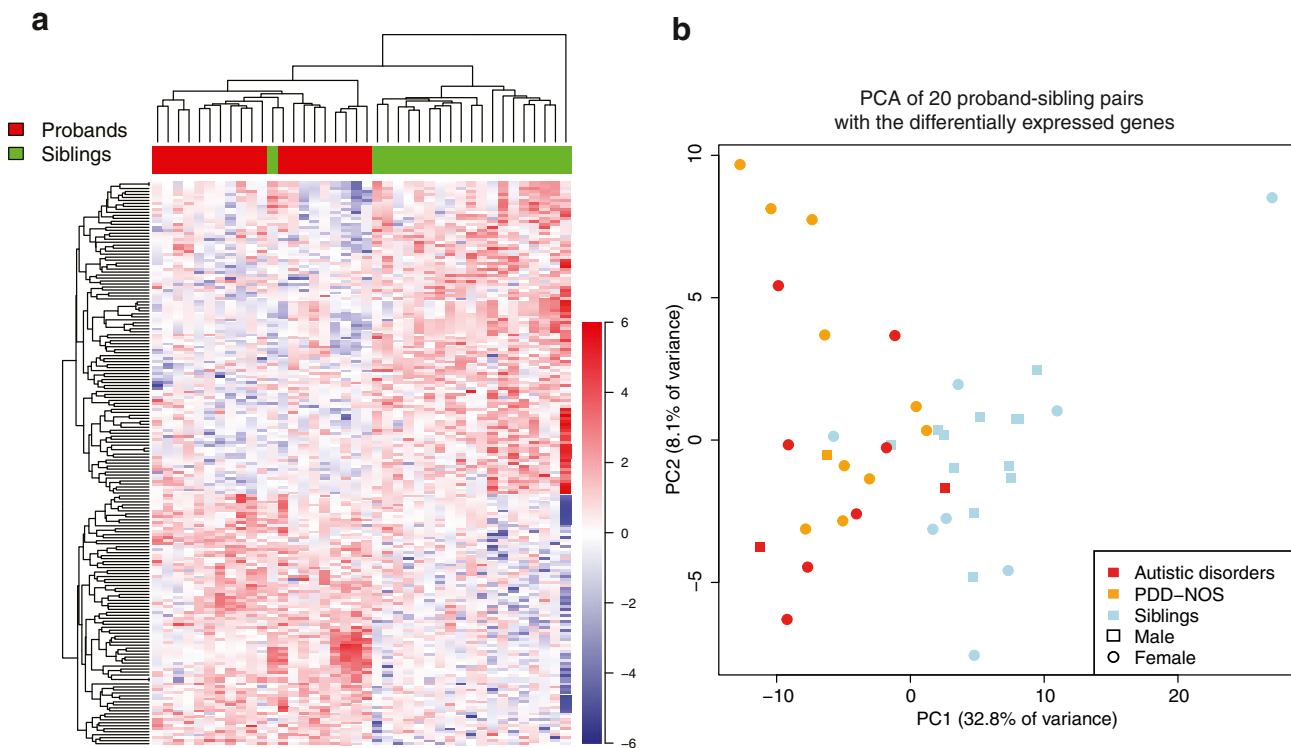
Fig. 2 Cluster analysis with the 189 differentially expressed probesets. **a** Unsupervised hierarchical clustering of probands and siblings with the 189 significant probesets that represent 163 genes successfully clusters probands and siblings except for the one sibling who is clustered with probands. The up- (*r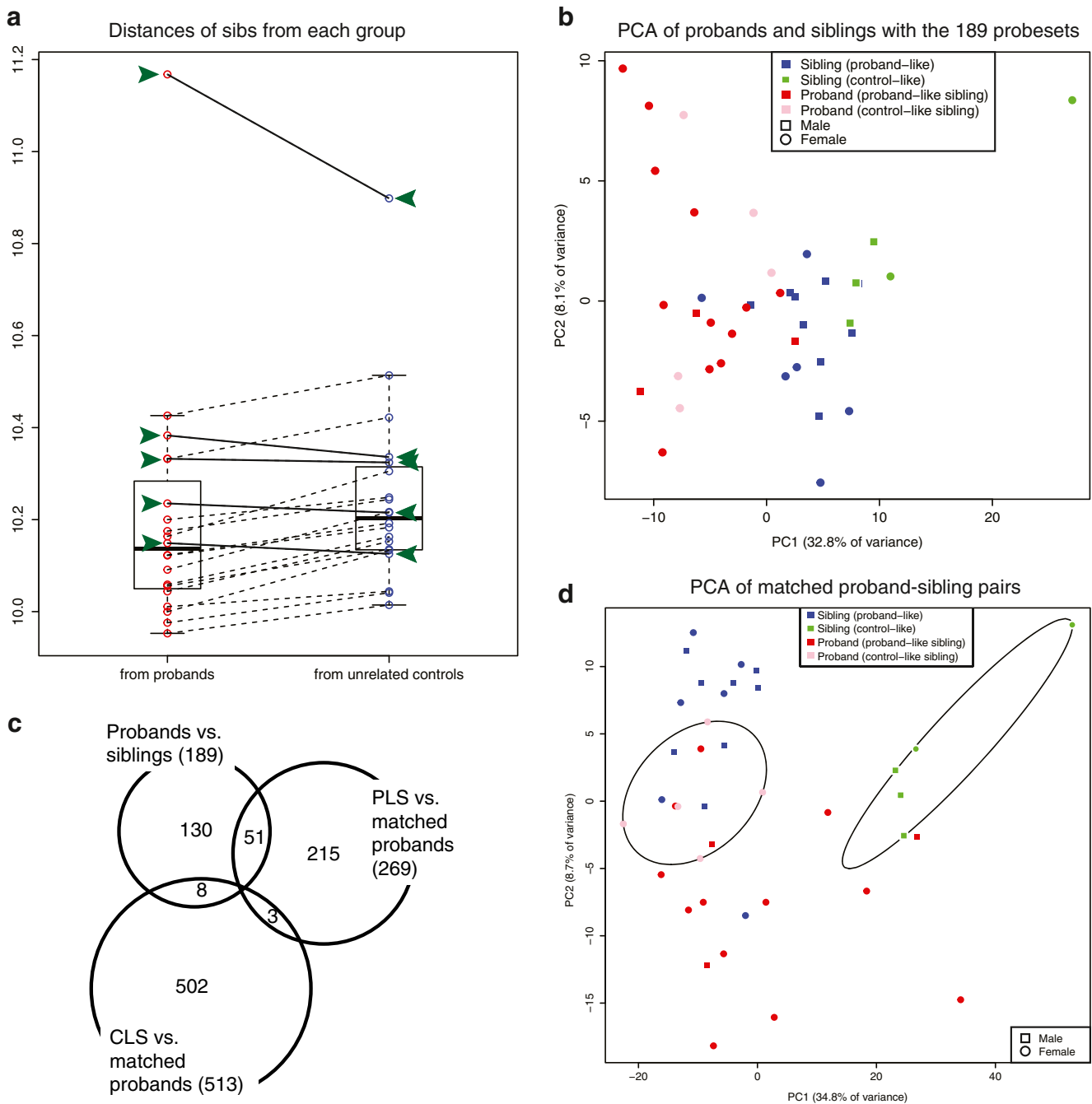ed*) and down- (*blue*) regulated genes in probands compared to siblings are also grouped in the heatmap. **b** Principal component analysis of 20 proband-sib pairs also reveals that two groups are well separated with the 189 probesets, and no significant bias presents for gender or subtype of autism spectrum disorders

unaffected siblings (green arrow heads) were more similar to unrelated controls while the other 15 siblings were more similar to probands or equidistant from proband and unrelated control groups. We designated the 15 siblings that were closer to probands as "proband-like" siblings (PLS), and the other siblings as "control-like" siblings (CLS). Interestingly, CLS individuals could be identified in the PCA plot using 189 probesets as they were located on the opposite side of probands compared to PLS individuals (Fig. 3b). We tested whether the difference between two sibling groups was due to the fact that they were from different families by permuting the group labels of the 20 siblings with replacement (permutation $B=10,000$). At the same statistical significance threshold (uncorrected $p<0.01$), the average number of differentially expressed probesets between random groupings of unaffected siblings into five and 15 was 316 (standard deviation (SD) 459.3, range 42–7,910). Thus, the differential expression of 3,126 probesets between PLS and CLS was significantly greater than random expectation (permutation $p=0.0056$).

After classifying the 20 unaffected siblings into five CLS and 15 PLS, we determined the differentially expressed genes for the two comparisons (CLS vs. matched probands and PLS vs. matched probands) by fitting linear models as described above. The numbers of differentially expressed probesets for these two comparisons were 513 and 269, respectively.

Compared to 189 significant probesets between 20 proband-sibs pairs, overall numbers of significant ones were greater after grouping CLS and PLS with their matched pairs (Fig. 3c). *PTPRC* (protein tyrosine phosphatase, receptor type, C) was differentially expressed between PLS and their matched probands, and nine genes — *CACNA1F*, *CDH22*, *EN2*, *FOXG1*, *HTR1B*, *PRKD1*, *SETDB1*, and *XPC* — were significant between CLS and matched their probands among the known ASD candidate genes in the Simons Foundation Autism Research Initiative (SFARI) Gene 2.0 database (available at http://gene.sfari.org).

Remarkably, PCA of the union of these differentially expressed genes (779 probesets total) revealed that CLS formed a distinct cluster, whereas PLS overlapped with probands (Fig. 3d). Taken together, these findings support the conclusion that the 15 PLS individuals, as measured by comparing squared MDs to probands and unrelated controls, approached the transcriptomic signature of ASD, whereas the CLS were farther from their matched probands in gene expression space.

Significant pathways between proband-like siblings and probands

To identify the pathways that distinguish probands from their unaffected siblings, we performed GSEA. GSEA ranks the

a Distances of sibs from each group

b PCA of probands and siblings with the 189 probesets

c

d PCA of matched proband-sibling pairs

Fig. 3 Identification of proband-like and control-like siblings in gene expression patterns. **a** Boxplot showing the squared Mahalanobis distances (*MDs*) from unaffected siblings to probands (*red circles*) and unrelated controls (*blue circles*). A *pair of red and blue circles connected by line* represents the distances to probands and unrelated controls for each unaffected sibling. That is, each *pair of connected red and blue circles* represents one sibling sample. Five control-like siblings (*CLS*) are equidistant to unrelated controls and matched probands (highlighted in *green arrowheads* and *solid*

connecting lines), whereas 15 proband-like siblings (*PLS*) are closer to matched probands (*black connecting lines*). **b** Five CLS individuals are highlighted in *green* in Fig. 2b, where CLS group is on the opposite side of probands relative to PLS. **c** Venn diagram of overlapping significant probesets from three comparisons. **d** Principal components analysis of probands and siblings with the differentially expressed genes from CLS vs. probands and PLS vs. probands. Ellipsoids are the minimum spanning ellipsoid for CLS and matched probands

pre-defined biological pathways by the concordant difference between two groups across multiple genes in a pathway while single-gene level change may not be significant. To facilitate a meaningful biological interpretation, we tested only pathways

from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [28]. We found 13 KEGG pathways were significantly changed in probands compared to unaffected siblings (nominal p<0.01, corresponding FDR 0.34) (Fig. 4a). The positive

a

NAME	SIZE	prob vs. sib	prob vs. cls	prob vs. pls	pls vs. cls
RIBOSOME	78	red	red	red	red
RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY	63	red		red	
OXIDATIVE_PHOSPHORYLATION	104	red	red		red
RNA_DEGRADATION	52	red	red		red
SPLICEOSOME	102	red	red		red
SYSTEMIC_LUPUS_ERYTHEMATOSUS	107	red			red
NEUROACTIVE_LIGAND_RECEPTOR_INTERACTION	257	blue			blue
MATURITY_ONSET_DIABETES_OF_THE_YOUNG	24	blue			blue
CALCIUM_SIGNALING_PATHWAY	171	blue			blue
REGULATION_OF_ACTIN_CYTOSKELETON	194	blue			
ECM_RECEPTOR_INTERACTION	78	blue			blue
GAP_JUNCTION	78	blue			
ARRHYTHMOGENIC_RIGHT_VENTRICULAR_CARDIOMYOPATHY_ARVC	71	blue			
DRUG_METABOLISM_CYTOCHROME_P450	53			red	blue
NUCLEOTIDE_EXCISION_REPAIR	38		red		red
CHEMOKINE_SIGNALING_PATHWAY	170			red	red
LYSOSOME	119				red
ENDOCYTOSIS	167		blue		
TOLL_LIKE_RECEPTOR_SIGNALING_PATHWAY	88		red		
HUNTINGTONS_DISEASE	158		red		
PEROXISOME	70		red		
BASE_EXCISION_REPAIR	31		red		
PARKINSONS_DISEASE	101		red		red
GRAFT_VERSUS_HOST_DISEASE	27		red		
AMINO_SUGAR_AND_NUCLEOTIDE_SUGAR_METABOLISM	43		red		red
CELL_CYCLE	106		red		
CITRATE_CYCLE_TCA_CYCLE	29		red		
BASAL_TRANSCRIPTION_FACTORS	30		red		
CYTOSOLIC_DNA_SENSING_PATHWAY	44		red		
PROTEIN_EXPORT	21		red		red
DNA_REPLICATION	34		red		red
RNA_POLYMERASE	23		red		red
MISMATCH_REPAIR	20		red		red
PROTEASOME	43		red		red
PYRIMIDINE_METABOLISM	87		red		red
GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	23		red		red
VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	41		red		red
UBIQUITIN_MEDIATED_PROTEOLYSIS	123		red		red
AMINOACYL_TRNA_BIOSYNTHESIS	39		red		red
BASAL_CELL_CARCINOMA	54				blue
AXON_GUIDANCE	125		blue		
METABOLISM_OF_XENOBIOTICS_BY_CYTOCHROME_P450	54		blue		
LONG_TERM_POTENTIATION	65		blue		
RETINOL_METABOLISM	46		blue		
VASCULAR_SMOOTH_MUSCLE_CONTRACTION	109		blue		
CYTOKINE_CYTOKINE_RECEPTOR_INTERACTION	239		blue		
MELANOGENESIS	94		blue		blue
NOD_LIKE_RECEPTOR_SIGNALING_PATHWAY	56				red
HOMOLOGOUS_RECOMBINATION	25				red

b
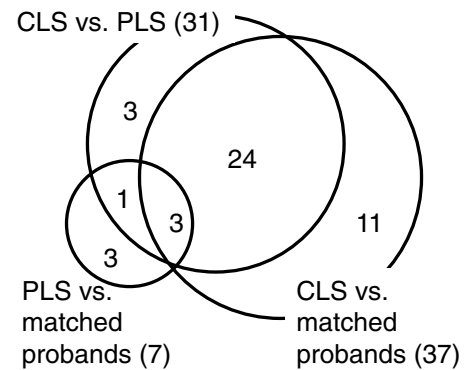
Fig. 4 Significant biological pathways between probands and siblings. **a** KEGG pathways identified by GSEA for the three comparisons at nominal $p<0.01$ (see Materials and methods) for the comparisons of probands (*prob*) and siblings (*sib*), control-like siblings (*cls*) vs. matched prob, proband-like siblings (*pls*) vs. matched prob, and pls vs. cls. The normalized enrichment score (NES) in Table 2 represents that a majority of genes in that pathway were up-regulated in probands compared to unaffected siblings. The two most significantly up-regulated pathways in probands were ribosome and spliceosome pathways. Seven pathways including neuroactive ligand receptor interaction pathway, calcium signaling pathway, and gap junction were down-regulated in probands.

In PLS vs. matched probands comparison, two immune pathways were dysregulated in opposite directions. RIG-I-like

pathways that are significant in any of four comparisons are listed and *red* (up-regulated in prob or pls) and *blue boxes* (down-regulated) represent significantly changes. Ribosomal genes were consistently up-regulated in all four comparisons (see Results). **b** The Venn diagram shows the overlap of significant pathways listed in panel **a**

receptor signaling pathway describes the signaling pathways that lead to the synthesis of type I interferon and other inflammatory cytokines in response to dsRNA due to viral replication in human cells, and chemokine signaling pathway activates various downstream pathways such as cellular polarization and actin reorganization in response to chemokines. In the matched probands of PLS group, RIG-I-like receptor signaling pathway was up-regulated while chemokine signaling pathway was down-regulated (Table S2).

Table 2 Significant biological pathways in probands compared to unaffected siblings

Pathway	Number of genes	Normalized enrichment score	Nominal p value	False discovery rate
Up-regulated in probands				
Ribosome	78	2.46	0.00000	0.00000
Spliceosome	102	1.82	0.00064	0.08867
RIG-I-LIKE receptor signaling pathway	63	1.81	0.00000	0.06492
Systematic Lupus Erythematosus	107	1.72	0.00128	0.11253
Oxidative Phosphorylation	104	1.68	0.00000	0.11198
RNA degradation	52	1.68	0.00644	0.09754
Down-regulated in probands				
Neuroactive ligand receptor pathway	257	−2.02	0.00000	0.00519
Maturity Onset Diabetes of the Young (MODY)	24	−2.00	0.00000	0.00409
Arrhythmogenic right ventricular cardiomyopathy	71	−1.84	0.00069	0.02690
ECM Receptor Interaction	78	−1.68	0.00136	0.11579
Gap Junction	78	−1.57	0.00814	0.27062
Calcium Signaling Pathway	171	−1.51	0.00561	0.24813
Regulation Of Actin Cytoskeleton	194	−1.42	0.00966	0.34729

We found the smallest number of significant pathways between PLS and matched probands. Ribosome pathway was up-regulated in the other comparison — CLS vs. matched probands, PLS vs. matched probands, and CLS vs. PLS — while chemokine signaling pathway and endocytosis were down-regulated in PLS vs. matched probands, but not in the other comparisons (Fig. 4a). Interestingly, 27 pathways were found in both CLS vs. matched probands and CLS vs. PLS comparisons (27/37 and 27/31 respectively). The Venn diagram of significant pathways from three comparisons suggested that CLS group seemed to be the most different from PLS and probands overall (Fig. 4b). Except for ribosome pathway, two overlapping pathways — nucleotide excision repair and drug metabolism cytochrome P450 — changed to opposite directions. Oxidative phosphorylation, Parkinson's disease, and Huntington's disease pathways all contained mitochondrial electron transport chain (ETC) genes. Except for these pathways, all other pathways had unique sets of genes that are specific for each pathway (Table S3). By splitting probands into two groups of PLS matched and CLS matched probands, we could identify additional interesting pathways such as long-term potentiation and axon guidance that were down-regulated in probands compared to matched CLS.

Discussion

The increased prevalence of ASD and autistic traits in relatives of individuals with ASD is well established [8, 10] and is likely due to, in part, shared genetic background of common and rare variants that predispose for ASD [15]. Simplex cases involve an increased de novo mutation rate and extent of de novo CNVs, although support for this hypothesis is not consistent across studies [2, 12, 14, 29–31] (reviewed by Berg and Geschwind [15]). In this paper, we found the gene expression signature of 189 genes that successfully distinguished probands from unaffected siblings. We also presented evidence for variable similarity to probands across 20 siblings from simplex families using blood gene expression profiling. Comparison of 20 unaffected siblings to matched probands and to 18 unrelated controls using squared MDs revealed that 15 were closer to probands (PLS), whereas five were closer to unrelated controls (CLS).

Pathway analysis identified that ribosomal, spliceosomal, and mitochondrial genes were up-regulated in PLS compared CLS, and neuroactive ligand receptor interaction, ECM receptor interaction, calcium signaling, and MODY (Maturity Onset Diabetes of the Young) pathways were down-regulated. Moreover, ribosomal and mitochondrial genes were up-regulated in probands compared to CLS. Two immune response pathways, RIG-I-like receptor signaling and chemokine signaling pathways were changed to opposite direction in probands compared to PLS. Immune pathways have been implicated in ASD throughout multiple tissues. CNS cytopathology [32], serum and CSF proteomics [33], a gene expression study using postmortem CNS tissue [34], and several gene expression studies using LCL revealed a converging signature of a perturbed immune system[21, 22]. Down-regulation of calcium signaling pathway is interesting since

recent collaborative effort of meta-analysis of five major neuropsychiatric disorders including ASD suggests that two calcium channel coding genes — *CACNA1C* and *CACNB2* — are significantly associated with all five diseases [35].

Multiple independent studies also support the role of mitochondrial function and pyrimidine metabolism in ASD. Giulivi and colleagues [36] discovered a subgroup of ASD with mtDNA CNVs and found that mitochondrial-dependent oxygen consumption was impaired in peripheral blood lymphocytes from children with autism compared with control children. Anitha and colleagues [37] reported that a group of mitochondrial ETC genes from Complex I–V was down-regulated in post-mortem brains from patients with ASD compared to controls. Moreover, functional brain imaging studies using magnetic resonance spectroscopy also found altered levels of ATP in ASD cases [38, 39].

There are potentially important limitations of the current study. We did not focus on the differentially expressed genes and pathways between unrelated controls and simplex families due to two reasons. First, families share environmental influences and most genetic material, including material unrelated to ASD, so siblings are better controls than unrelated individuals. Second, though we tried our best to remove the batch effect during microarray preparation, batch effects remained between individuals from simplex families and unrelated controls. The number of families and unrelated controls recruited for our current study was limited as we analyzed the gene expression profiles from 20 proband-sibling pairs and 18 unrelated controls. We recruited simplex families and unrelated controls; however, several studies have reported that subclinical ASD phenotypes were more frequently observed in multiplex families than in simplex families [10, 11]. Further investigation including multiplex families would be an interesting direction.

To summarize, we explored peripheral blood gene expression profiles from 20 children with ASD compared with unaffected siblings. We asked two simple questions: first, whether blood gene expression profiles distinguish probands from unaffected siblings, and second, what genes or pathways push unaffected siblings towards probands. We successfully identified 189 genes that could cluster probands and unaffected siblings into two separated groups, and two groups of unaffected siblings — PLS and CLS — using global gene expression patterns. The transcriptomic changes could be pathophysiologically linked or possibly reflect compensatory changes. As we cannot draw any conclusion regarding the causal relationship of these pathways to ASD, further investigation using integrative analysis of whole-genome or exome sequences with gene expression profiles including parents will be beneficial.

References

1. Luo R et al (2012) Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in autism spectrum disorders. Am J Hum Genet 91(1):38–55
2. Sanders SJ et al (2011) Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. Neuron 70:863–885
3. Anney R et al (2010) A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet 19(20):4072–82
4. Alarcon M et al (2008) Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. Am J Hum Genet 82(1):150–9
5. Hallmayer J et al (2011) Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 68:1095–1102
6. Ritvo ER et al (1989) The UCLA–University of Utah epidemiologic survey of autism: recurrence risk estimates and genetic counseling. Am J Psychiatry 146(8):1032–6
7. Chakrabarti S, Fombonne E (2001) Pervasive developmental disorders in preschool children. JAMA 285(24):3093–9
8. Ozonoff S et al (2011) Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. Pediatrics 128(3):e488–95
9. Szatmari P et al (2000) The familial aggregation of the lesser variant in biological and nonbiological relatives of PDD probands: a family history study. J Child Psychol Psychiatry 41(5):579–86
10. Constantino JN et al (2006) Autistic social impairment in the siblings of children with pervasive developmental disorders. Am J Psychiatry 163(2):294–6
11. Losh M et al (2008) Defining key features of the broad autism phenotype: a comparison across parents of multiple- and single-incidence autism families. Am J Med Genet B Neuropsychiatr Genet 147B(4):424–33
12. Sebat J et al (2007) Strong association of de novo copy number mutations with autism. Science 316:445–449
13. Neale BM et al (2012) Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485:242–245
14. Pinto D et al (2010) Functional impact of global rare copy number variation in autism spectrum disorders. Nature 466:368–372
15. Berg JM, Geschwind DH (2012) Autism genetics: searching for specificity and convergence. Genome Biol 13(7):247
16. Kong SW et al (2012) Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. PLoS One 7(12):e49475

17. Glatt SJ et al (2012) Blood-based gene expression signatures of infants and toddlers with autism. J Am Acad Child Adolesc Psychiatry 51(9):934–44 e2

18. Enstrom AM et al (2009) Altered gene expression and function of peripheral blood natural killer cells in children with autism. Brain Behav Immun 23(1):124–33

19. Hu VW 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 Res 2(2):78–97

20. Hu VW et al (2009) Gene expression profiling of lymphoblasts from autistic and nonaffected sib pairs: altered pathways in neuronal development and steroid biosynthesis. PLoS One 4(6):e5775

21. Gregg JP et al (2008) Gene expression changes in children with autism. Genomics 91(1):22–9

22. Nishimura Y et al (2007) Genome-wide expression profiling of lymphoblastoid cell lines distinguishes different forms of autism and reveals shared pathways. Hum Mol Genet 16(14):1682–98

23. Baron CA et al (2006) Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. J Autism Dev Disord 36(8):973–82

24. Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. Proc Natl Acad Sci U S A 100(16):9440–5

25. Schäfer J, Strimmer K (2005) A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. Statist Appl Genet Mol Biol 4:(32)

26. Reich M et al (2006) GenePattern 2.0. Nat Genet 38(5):500–1

27. Constantino JN et al (2010) Sibling recurrence and the genetic epidemiology of autism. Am J Psychiatry 167(11):1349–56

28. Kanehisa M et al (2012) KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 40(Database issue):D109–14

29. Levy D et al (2011) Rare de novo and transmitted copy-number variation in autistic spectrum disorders. Neuron 70(5):886–97

30. Szatmari P et al (2007) Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nat Genet 39:319–328

31. Marshall CR et al (2008) Structural variation of chromosomes in autism spectrum disorder. Am J Hum Genet 82:477–488

32. Vargas DL et al (2005) Neuroglial activation and neuroinflammation in the brain of patients with autism. Ann Neurol 57(1):67–81

33. Corbett BA et al (2007) A proteomic study of serum from children with autism showing differential expression of apolipoproteins and complement proteins. Mol Psychiatry 12(3):292–306

34. Garbett K et al (2008) Immune transcriptome alterations in the temporal cortex of subjects with autism. Neurobiol Dis 30(3):303–11

35. Cross-Disorder Group of the Psychiatric Genomics, C. (2013) Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet

36. Giulivi C et al (2010) Mitochondrial dysfunction in autism. JAMA 304(21):2389–96

37. Anitha A et al (2012) Downregulation of the expression of mitochondrial electron transport complex genes in autism brains. Brain Pathol. 23(3):294–302

38. Chugani DC et al (1999) Evidence of altered energy metabolism in autistic children. Prog Neuropsychopharmacol Biol Psychiatry 23(4):635–41

39. Minshew NJ et al (1993) A preliminary 31P MRS study of autism: evidence for undersynthesis and increased degradation of brain membranes. Biol Psychiatry 33(11–12):762–73