Behav Genet (2016) 46:754–762
DOI 10.1007/s10519-016-9804-4

ORIGINAL RESEARCH

The Unique Evolutionary Signature of Genes Associated with Autism Spectrum Disorder

Erez Tsur[1,2] · **Michael Friger**[1] · **Idan Menashe**[1,2]

Received: 16 March 2016 / Accepted: 4 August 2016 / Published online: 11 August 2016

Abstract Autism spectrum disorder (ASD) is a common heritable neurodevelopmental disorder, which is characterized by communication and social deficits that reduce the reproductive fitness of individuals with the disorder. Here, we studied the genomic characteristics of 651 ASD genes in a whole-exome sequencing dataset, to search for traces of the evolutionary forces that helped maintain ASD in the human population. We show that ASD genes are ~65 longer and ~20 % less variable than non-ASD genes. The mutational shortage in ASD genes was particularly eminent when considering only deleterious genetic variations, which is a hallmark of negative selection. We further show that these genomic characteristics are unique to ASD genes, as compared with brain-specific genes or with genes of other diseases. Our findings suggest that ASD genes have evolved under complex evolutionary forces, which have left a unique signature that can be used to identify new candidate ASD genes.

Keywords Autism spectrum disorder (ASD) · Evolution · Exome · Negative selection

Edited by Stacey Cherny.

Electronic supplementary material The online version of this article (doi:10.1007/s10519-016-9804-4) contains supplementary material, which is available to authorized users.

✉ Idan Menashe
idanmen@bgu.ac.il

[1] Department of Public Health, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beersheba, Israel

[2] Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev, Beersheba, Israel

Introduction

Autism spectrum disorders (ASDs) are a collection of neurodevelopmental conditions, which are characterized by impaired communication skills, avoidance of social interactions, and repetitive and stereotype behavior (American Psychiatric Association 2013). In the past three decades, ASD has become a major public health concern, with a substantial increase in the prevalence of ASD worldwide (Elsabbagh et al. 2012; Christensen et al. 2016; Davidovitch et al. 2013; Richards et al. 2015; Taylor et al. 2013). While this increase has largely been attributed to higher public awareness and changes in diagnostic criteria (Lord 2011; Posserud et al. 2010; Maenner et al. 2014), the contribution of other risk factors cannot be excluded. Research into the risk factors of ASD and the molecular mechanisms associated with them can shed light on the possible etiologies of these increasingly prevalent neurodevelopment conditions.

It is well accepted that genetic factors play a significant role in ASD susceptibility (Connolly and Hakonarson 2014; de la Torre-Ubieta et al. 2016; Devlin and Scherer 2012; Huguet et al. 2013; Muers 2012). During the past decade, genetic studies have suggested hundreds of ASD-susceptibility genes, but only a few of these genes have been robustly associated with the disorder (Abrahams et al. 2013; Iossifov et al. 2015; Sanders et al. 2015) and mutations in these genes can explain ASD in only a portion of affected individuals. Thus, an important endeavor in ASD research is to identify new ASD susceptibility genes and to study their role in the etiology of the disease.

Individuals with ASD show marked deficiencies in communication and social skills and a repetitive and stereotypic behavior, which complicate their integration into society and significantly reduce their reproductive

fitness (Power et al. 2013). Hence, one would expect that mutations that predispose to such a highly heritable and harmful trait would be eliminated from the population by natural selection (Keller and Miller 2006), and the fact that ASD has remained a relatively common trait in the population is, therefore, an evolutionary enigma. A number of theories have been postulated to explain this enigma (Ploeger and Galis 2011), all of which rely on the assumption that ASD is a polygenic trait. Therefore, certain mutations would have a deleterious effect only if they occur in combination with other mutations (Huguet et al. 2013). Other theories postulate that certain mutations in ASD-susceptibility genes may result in neurotypical development accompanied by an evolutionary advantageous trait (e.g., higher intelligence) when expressed under a certain genetic background, but in an ASD when expressed under a different genetic background (Ploeger et al. 2009). If these assumptions are true, then genes that are involved in ASD susceptibility are expected to have a unique allelic signature, which has possibly been shaped by both positive and negative selection.

In recent years, genetic research has shifted from small-scale studies, which often focus on a few genes in small samples, to large, high-throughput studies of massive 'omics' data obtained from large populations and from a wide range of species (Thorisson et al. 2009). The resulting high-resolution and rich genomic data is a "goldmine" for population genetics analyses (Platt and Novembre 2012). In the current study, we employed such a large whole-exome sequencing dataset to explore the genomic and evolutionary characteristics of genes associated with ASD.

Methods

Sample and Genomic Data

We studied the exome sequencing dataset previously generated by Tennessen et al. (2012). This dataset was derived from the genomes of 1351 European-American and 1088 African-Americans individuals, sampled randomly from fifteen North American cohorts who participated in a large genetic study of cardiovascular diseases. The exome sequencing dataset included data on 503,481 single-nucleotide variants (SNVs) distributed across 15,585 human genes. There was no indication of population- or phenotype-specific effects, or of other systematic biases, during the analysis of these data (Tennessen et al. 2012).

Disease Gene Datasets

To determine genes associated with ASD, we used the annotation from the human gene module of AutDB (Basu et al. 2009) (data freeze of December 2015), which is the most comprehensive genetic database of ASD to date. Of the 790 ASD genes in AutDB, 651 genes (82.4 %) were included in the exome sequencing dataset (Supplementary Table S1). This percentage is slightly higher than that of all genes in the human genome represented in this exome dataset [75.5 %; (Brown et al. 2014)], but the difference is not statistically significant ($P = 0.11$; two-sided Chi-Square test). Furthermore, we used the gene scoring module of AutDB (Abrahams et al. 2013) to construct a subset of 12 "high-confidence" and 18 "strong candidate" ASD genes (categories "1" and "2", respectively, in the SFARI gene scoring module) (Supplementary Table S1).

We used three other disease-specific gene sets and one brain-specific gene set as control groups for our analyses (Supplementary Table S1). Specifically, we used the gene sets of two neurological diseases: schizophrenia and Alzheimer's disease, which we downloaded from the "Schizophrenia Gene Resource (SZGR)" (258 genes) (Jia et al. 2010) and from the "AlzGene database" (570 genes) (Bertram et al. 2007) respectively. We used asthma-related genes from "InnateDB" (499 genes) (Lynn et al. 2008) as a gene set of an early-onset disease. Finally, we used the 7795 genes from Supplementary Table S1 of Ouwenga and Dougherty (2015) as a brain-specific gene set.

Statistical Analyses

We examined the coding sequence length, frequency of SNVs, and measures of nucleotide diversity of genes from the exome dataset of Tennessen et al. (2012). We calculated Tajima's D statistic (Tajima 1989) for all genes in our dataset to compare the observed frequency spectrum of SNVs with neutral model expectations. We compared these genomic characteristics between ASD genes and non-ASD genes using two-tailed t-tests. We also tested the significance of our findings by using a bootstrap procedure with 1000 replications. For genomic characteristics that were significantly different between ASD genes and non-ASD genes, we further compared ASD genes to the genes of three other disease-specific groups, namely, genes associated with schizophrenia, Alzheimer's disease, and asthma. To account for the overlap of genes between these diseases, we calculated the 95, 99 and 99.9 % confidence intervals for the difference between the means of the genomic characteristics of ASD and those of the other diseases (e.g., $\Delta\mu_{\text{ASD-Alzheimer's}}$). Consequently, a CI that does not include zero indicates a significant difference between the two groups (at $P < 0.05$, $P < 0.01$, and $P < 0.001$ for a 95, 99, and 99.9 % CI, respectively). Finally, we used the unique variables of ASD genes in a multivariate logistic regression model differentiate ASD genes from non-ASD genes. Then, we used the DAVID functional annotation analysis

(with its default parameters) (Huang da et al. 2009) to test which biological pathways (i.e., groups of genes that have similar biological ontologies) are enriched with the top-ranked genes by this model.

Results

Genomic Characteristics

We first compared several genomic and evolutionary characteristics of 651 ASD genes and 14,934 non-ASD genes (Table 1). On average, ASD genes were both longer than non-ASD genes (2.68 vs. 1.63 kbp, respectively) and less variable than non-ASD genes (nucleotide diversity: $\pi = 0.036$ % vs. $\pi = 0.046$ %, respectively). The lower variability in ASD genes was driven by their relative dearth in SNVs (20.61 vs. 23.57 per kbp, respectively) and by the lower allele frequencies of these SNVs (rare/common SNV ratio: 8.40 vs. 6.80, respectively). All differences between ASD and non-ASD genes were statistically significant ($P < 0.001$) and were more prominent when compared with high-confidence subset of ASD genes to the non-ASD genes (Table 1).

Next, we calculated Tajima's D statistic (Tajima 1989) for all genes in our dataset to compare the observed frequency spectrum of SNVs with neutral model expectations. The majority of the genes examined in this study (>99 %) had a negative Tajima's D, indicating an excess of low-frequency polymorphisms relative to expectation, an observation that is consistent with the recent expansion of the modern human population (McEvoy et al. 2011). On average, Tajima's D value was more negative in ASD genes than in non-ASD genes (−1.96 vs.−1.76, respectively; $P < 0.001$), and this difference was even greater between the high-confidence subset of the ASD genes and the non-ASD genes (Table 1).

We continued to examine the effects of evolutionary forces on loci with potential functional consequences. ASD genes had a significantly lower non-synonymous/

Table 1 Genomic and evolutionary characteristics

Genomic characteristic	ASD genes[a] (High-confidence ASD genes[b])	Non-ASD genes	P-values
Number of genes	651 (30)	14,934	–
Gene length[c] (kbp)	2.68 ± 2.54 (4.33 ± 2.78)	1.63 ± 1.38	<0.001 (<0.001)
No. of SNV (per kbp)	20.61 ± 8.90 (15.71 ± 5.49)	23.57 ± 10.70	<0.001 (<0.001)
Rare/common SNV ratio[d]	8.40 ± 8.06 (10.77 ± 7.46)	6.80 ± 6.13	<0.001 (0.007)
Nucleotide diversity (π %)	0.036 ± 0.060 (0.022 ± 0.021)	0.046 ± 0.060	<0.001 (<0.001)
Tajima's D	−1.96 ± 0.48 (−2.18 ± 0.39)	−1.76 ± 0.49	<0.001 (<0.001)
Non-synonymous/synonymous SNV ratio	1.34 ± 1.12 (0.87 ± 0.43)	1.74 ± 1.40	<0.001 (<0.001)
No. of missense SNVs (per kbp)	9.39 ± 5.68 (6.44 ± 3.73)	11.31 ± 6.89	<0.001 (<0.001)
No. of nonsense SNVs (per kbp)	0.15 ± 0.42 (0.03 ± 0.65)	0.30 ± 0.65	<0.001 (<0.001)
No. of functional SNV (per kbp) (Polyphen2)[e]	0.21 ± 0.43 (0.08 ± 0.18)	0.41 ± 0.81	<0.001 (<0.001)

[a] Genes that were associated with ASD in the literature according to the AutDB database {Basu, 2009#468}

[b] Genes with strong evidence of association with ASD according to the SFARI gene scoring module {Abrahams, 2013#904}

[c] Gene's coding sequence length

[d] Rare SNV—minor allele frequency (MAF) < 0.5 %; Common SNVs—MAF ≥0.5 %

[e] SNVs with potential functional effect were determined using Polyphen2 software {Adzhubei, 2013#1264}

synonymous SNVs ratio, as compared with non-ASD genes (1.34 vs. 1.74, respectively; $P < 0.001$). This difference was particularly noticeable when considering SNVs with potential deleterious effects [e.g., nonsense SNVs or deleterious SNVs according to PolyPhen-2 (Adzhubei et al. 2013)]. These findings suggest that negative selection has a stronger effect in removing potentially deleterious SNVs from ASD genes than from non-ASD genes. As in the previous analyses, the magnitude of all differences increased when comparing the high-confidence ASD genes to the non-ASD genes (Table 1).

Recently, the Exome Aggregation Consortium (ExAC), had released the exome sequencing data of over 60,000 people (http://exac.broadinstitute.org; Release 0.3.1). Examining the data from this dataset revealed similar differences in the genomic characteristics of ASD genes and non-ASD genes (Supplementary Table S2), thus providing further assurance for our finding.

Positive Selection

Tennessen et al. have identified signatures of positive selection in 114 genes from their dataset [Table S4 in Tennessen et al. (2012)]. None of these genes overlapped with our list of ASD genes. We further examined the representation of ASD genes in 722 autosomal genomic regions that have been implicated in several studies as targets of positive selection [reviewed in Akey (2009)]. The proportion of ASD genes in these regions was slightly higher than that of non-ASD genes (12.9 vs. 9.7 %, respectively), but this enrichment was not statistically significant after accounting for their relatively longer genomic sequence ($P = 0.096$).

ASD Genes Versus Genes of Other Diseases

To test whether our findings are specific to ASD genes, we examined the same genomic characteristics in a set of brain-specific genes (n = 7795) and in three sets of other disease-specific genes [namely, Alzheimer's disease (n = 570), schizophrenia (n = 258), and asthma (n = 499)]. Due to the overlap of genes in the different datasets (Supplementary Fig. S1), we calculated the CI of the difference between the means to evaluate significant differences between ASD genes and genes from the other datasets. ASD genes were longer ($P < 0.001$), had a higher proportion of rare SNVs ($P < 0.001$), and had a more negative Tajima's D ($P < 0.05$) than genes from any of the other four gene sets (Fig. 1a, d, and e, respectively). In addition, ASD genes had fewer SNV per kbp than the genes of schizophrenia and Alzheimer's disease ($P < 0.001$ for each comparison; Fig. 1c), a lower non-synonymous/synonymous ratio than brain-specific genes (Fig. 1f), and a lower rate of functional SNVs than Alzheimer's disease genes ($P < 0.05$ for each comparison; Fig. 1g).

A Multivariate Model for Classifying ASD Genes

We conducted a multivariate logistic regression analysis to test the combined classification ability of the above-mentioned characteristics of ASD genes (Table 2). The resulting model fitted the data very well (Homster–Lemshow test; $P = 0.989$), but had a moderate classification efficiency (AUC = 0.70; Fig. 2). However, applying the same model to the high-confidence ASD genes resulted in a remarkably high classification efficiency (AUC = 0.92), whereas, by contrast, applying the same genomic characteristics to classification models of the other disease-specific gene sets resulted in a poor classification efficiency (schizophrenia, AUC = 0.61; Alzheimer's disease, AUC = 0.58; and asthma, AUC = 0.63) (Fig. 2). Similar results were obtained when we tested the ability of these genomic characteristics to differentiate these disease-specific genes from brain-specific genes (Supplementary Table S3).

Finally, we used the 2 % top-ranked genes that were predicted by these genomic characteristics to be associated with ASD in a DAVID functional annotation analysis [309 genes overall, of which 91 are listed in AutDB (Basu et al. 2009)]. These top-ranked genes were enriched in one KEGG pathway ("long-term potentiation") and with one PANTHER pathway ("ionotropic glutamate receptor pathway") at a significance level of $P < 0.01$ (Bonferroni-corrected). In addition, these genes were enriched in 19 other gene ontology (GO) terms, most of which are related to neuronal/synaptic function and to chromatin remodeling (Supplementary Table S4).

Discussion

We examined the exome data of ASD susceptibility genes to look for signatures of the evolutionary forces that have shaped the genomic landscape of these genes. To the best of our knowledge, this is the first study to make such analysis by using a high-throughput whole-exome dataset from a large and diverse population (Tennessen et al. 2012). Our findings suggest that genes associated with ASD have genomic characteristics that are distinct from other genes in the genome, and they provide clues to the evolutionary forces that act on these genes.

Our analysis indicates that ASD genes are, on average, 65 % longer than other genes in the genome. This finding is not new, and it has already been shown by King et al. (2013), who attributed the exceptionally long sequences of

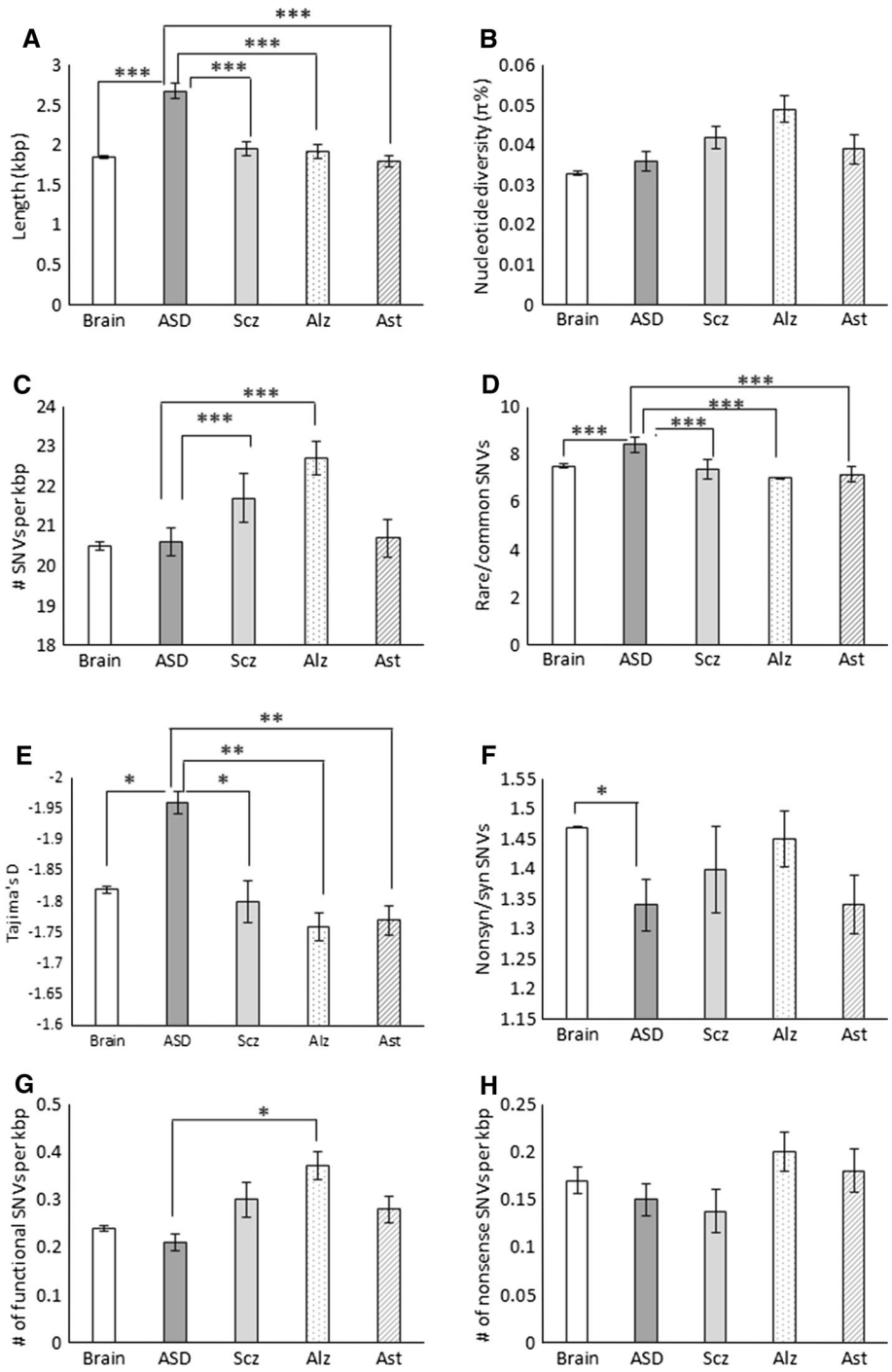

Fig. 1 Genomic characteristics of ASD genes relative to brain-specific genes and to three other sets of disease-specific genes. All genes associated with autism spectrum disorder (ASD) are compared with genes that are expressed in the brain and to genes that are specific to schizophrenia (Scz), Alzheimer's disease (Alz), and asthma (Ast). Comparisons regard: **a** protein coding sequence length; **b** nucleotide diversity (π); **c** number of SNVs per kbp; **d** rare/common SNV ratio; **e** Tajima's D; **f** non-synonymous/synonymous SNV ratio; **g** number of functional SNV per kbp; **h** number of nonsense SNV per kbp. Values indicate the mean ± SEM. Significant differences between the groups were calculated by using a shared confidence interval for the differences between the means (see Methods) and are indicated as *, **, and *** for $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively

ASD genes to their unique transcription regulation mechanism. This observation was then amended by Shohat et al. (2014), who implied that exceptionally long genomic sequences are not an inherent characteristic of all ASD risk genes, but, rather, that the long sequences appear only in genes that reside within copy-number variations (CNVs) associated with ASD. Notably, our analysis is slightly different from the analyses made in these two studies, as it focuses on exome sequences; whereas non-coding sequences may constitute a significant portion of some gene transcripts, exome sequences contain only tiny portions of non-coding sequences. We also show that ASD genes are significantly longer than other brain-expressed genes, which are thought to be longer than other genes in the genome (Ouwenga and Dougherty 2015), and that their exome length is unique, as compared with three other disease-specific gene sets, two of which are associated with other brain disorders (schizophrenia and Alzheimer's

Table 2 Results of multivariate logistic regression models

Variable	Aut			Scz			Alz			Ast		
	β	OR (95 % CI)	*P* value	β	OR (95 % CI)	*P*-value	β	OR (95 % CI)	*P*-value	β	OR (95 % CI)	*P*-value
Length (kbp)	0.17	1.18 (1.13–1.24)	<0.001	0.10	1.10 (1.02–1.18)	0.015	0.12	1.13 (1.07–1.19)	<0.001	0.03	1.03 (0.96–1.10)	0.437
SNV per kbp	−0.03	0.97 (0.96–0.98)	<0.001	−0.01	0.99 (0.97–1.00)	0.074	0	1.00 (0.99–1.01)	0.627	−0.03	0.98 (0.96–0.99)	<0.001
Rare/common SNV ratio	0.01	1.01 (0.99–1.02)	0.414	0.01	1.01 (0.99–1.03)	0.423	0.01	1.01 (0.99–1.02)	0.428	0	1.00 (0.99–1.02)	0.657
Tajima's D	−0.65	0.52 (0.40–0.67)	<0.001	0.02	1.02 (0.72–1.43)	0.922	0.23	1.26 (1.01–1.57)	0.038	−0.10	0.90 (0.70–1.17)	0.441
Non-synonymous/synonymous SNV ratio	−0.27	0.76 (0.69–0.84)	<0.001	0.22	0.80 (0.71–0.92)	0.002	−0.19	0.83 (0.76–0.90)	<0.001	−0.28	0.76 (0.68–0.84)	<0.001
Functional SNV per kbp	−0.20	0.82 (0.68–0.98)	0.031	0.02	0.98 (0.80–1.21)	0.867	0.04	1.04 (0.92–1.17)	0.547	−0.01	0.99 (0.85–1.16)	0.902

Results of multivariate logistic regression models to classify four diseases-specific gene-sets (*Aut* Autism, *Scz* Schizophrenia, *Alz* Alzheimer disease and *Ast* Asthma)

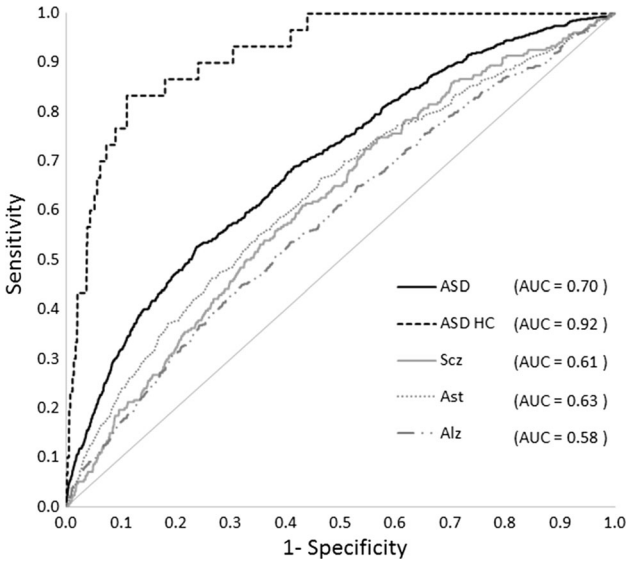

Fig. 2 Receiver operating characteristic(ROC). The plots demonstrate the classification efficiency of different sets of disease-specific genes [autism spectrum disorder (ASD), high-confidence ASD genes (HC-ASD), schizophrenia (Scz), Alzheimer's disease (Alz), and asthma (Ast)], based on multivariate logistic models with six genomic characteristics (gene length, number of SNVs per kbp, rare/common SNV ratio, non-synonymous/synonymous SNV ratio, functional SNV per kbp, and Tajima's D). A larger area under the *curve* indicates a better classification accuracy of the model

disease). It is not clear, at this point, why the protein-coding sequence of ASD genes is exceptionally long. One possible explanation is that it provides a large and versatile target for genetic manipulations, which could contribute to the development of complex and possibly diverse cognitive functions.

The lower nucleotide diversity—and, specifically, the paucity of SNVs with potentially deleterious consequences—that we observed in ASD genes as compared with non-ASD genes is a hallmark of negative selection (Hartl and Clark 2006). Myers et al. (2011) detected a similar signature of negative selection among 408 brain-expressed genes, which were studied in 240 ASD and schizophrenia cases and in a comparable number of control cases. In addition, results from multiple exome sequencing studies of families with a child diagnosed with ASD (Iossifov et al. 2012; De Rubeis et al. 2014) indicate a shift in the mutation spectrum toward deleterious rare variants among probands, as compared with their unaffected siblings, and these findings were further supported by genome-wide analyses of de novo mutations (Petrovski et al. 2013; Samocha et al. 2014; Uddin et al. 2014). Taken together, these results support the premise that negative selection removes damaging mutations from genes involved in the etiology of ASD.

Several theories have been raised regarding the mechanisms involved in maintaining ASD, a common

heritable low-reproductive trait, in the human population (Keller and Miller 2006; Ploeger and Galis 2011). Interestingly, some of these theories suggest that ASD genes have evolved under balancing selection, i.e., that their alleles may exert both positive and negative effects on human fitness (Keller and Miller 2006; Ploeger et al. 2009). Although our findings support the premise that ASD genes are subjected to strong negative selection, we could not find evidence for positive or balancing selection acting on these genes. Thus, the prevalence of ASD in the human population is likely maintained through a large number of rare susceptibility alleles (Krumm et al. 2015) and/or through genetic variations that exert deleterious effect only under certain genetic or environmental backgrounds (i.e., GxG or GxE interactions) (Tordjman et al. 2014; Chaste and Leboyer 2012; Corominas et al. 2014).

The utilization of genomic parameters for classifying ASD genes is an important aspect of this study. We show that employing a combination of ASD-specific genomic characteristics can reliably predict ASD susceptibility genes. However, despite the relatively good classification efficiency of such a multivariate model, the resulting predicted genes should be treated as candidate genes for ASD, and they should be further studied by other approaches, which may confirm or refute their involvement in the etiology of ASD.

Our study has several advantages and disadvantages. The main advantage is the size and comprehensiveness of the exome dataset; the large sample size, which includes individuals of both European and African ancestry, offers an adequate representation of the human population. In addition, the large sequencing depth of more than 75 % of the coding exons in the human genome has provided us with an unprecedented dataset for population genetic analysis of this kind. Considering all genes in the AutDB dataset as ASD genes is both an advantage and a disadvantage of the study. While AutDB is the most comprehensive genetic database of ASD, it likely includes many genes that are false positives (Abrahams et al. 2013) and might have introduced some bias to our results. In addition, some well-established ASD genes (e.g., *CHD8*, *SHANK3* and *NRXN1*) are not included in our analysis due to their absence from the exome dataset of Tennessen et al. (2012). The other disease-specific gene sets in our study may also lack important candidate genes. Nevertheless, the consistent differences between ASD and non-ASD genes in our analyses suggest that such a bias is likely negligible.

In summary, our results indicate that genes implicated in the etiology of ASD have explicit genomic characteristics, which separate them from other genes in the genome. These characteristics are likely a result of complex evolutionary forces that act on ASD genes and can be used as a signature to identify additional ASD candidate genes.

Acknowledgments The authors are thankful to Dr. Ram Gal for the editing of the manuscript and his useful comments.

Compliance with Ethical Standards

Conflict of Interest Mr. Erez Tsur declares that he has no conflict of interest. Prof. Michael Friger declares that he has no conflict of interest. Dr. Idan Menashe declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, Menashe I, Wadkins T, Banerjee-Basu S, Packer A (2013) SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism 4(1):36

Adzhubei I, Jordan DM, Sunyaev SR (2013) Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet Chapter 7:20

Akey JM (2009) Constructing genomic maps of positive selection in humans: where do we go from here? Genome Res 19(5):711–722

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, 5th Edition (DSM-V). American Psychiatric Publishing, Arlington

Basu SN, Kollu R, Banerjee-Basu S (2009) AutDB: a gene reference resource for autism research. Nucleic Acids Res 37:D832–D836

Bertram L, McQueen MB, Mullin K, Blacker D, Tanzi RE (2007) Systematic meta-analyses of alzheimer disease genetic association studies: the AlzGene database. Nat Genet 39(1):17–23

Brown GR, Hem V, Katz KS, Ovetsky M, Wallin C, Ermolaeva O, Tolstoy I, Tatusova T, Pruitt KD, Maglott DR, Murphy TD (2014) Gene: a gene-centered information resource at NCBI. Nucleic Acids Res 43:36–42

Chaste P, Leboyer M (2012) Autism risk factors: genes, environment, and gene-environment interactions. Dialogues Clin Neurosci 14(3):281–292

Christensen DL, Baio J, Braun KV, Bilder D, Charles J, Constantino JN, Daniels J, Durkin MS, Fitzgerald RT, Kurzius-Spencer M, Lee LC, Pettygrove S, Robinson C, Schulz E, Wells C, Wingate MS, Zahorodny W, Yeargin-Allsopp M (2016) Prevalence and characteristics of autism spectrum disorder among children aged 8 Years—autism and developmental disabilities monitoring network, 11 sites, United States 2012. MMWR Surveill Summ 65(3):1–23

Connolly JJ, Hakonarson H (2014) Etiology of autism spectrum disorder: a genomics perspective. Curr Psychiatry rep 16(11):501

Corominas R, Yang X, Lin GN, Kang S, Shen Y, Ghamsari L, Broly M, Rodriguez M, Tam S, Trigg SA, Fan C, Yi S, Tasan M, Lemmens I, Kuang X, Zhao N, Malhotra D, Michaelson JJ, Vacic V, Calderwood MA, Roth FP, Tavernier J, Horvath S, Salehi-Ashtiani K, Korkin D, Sebat J, Hill DE, Hao T, Vidal M, Iakoucheva LM (2014) Protein interaction network of alternatively spliced isoforms from brain links genetic risk factors for autism. Nat Commun 5:3650

da Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57

Davidovitch M, Hemo B, Manning-Courtney P, Fombonne E (2013) Prevalence and incidence of autism spectrum disorder in an Israeli population. J Autism Dev Disord 43(4):785–793

de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH (2016) Advancing the understanding of autism disease mechanisms through genetics. Nat Med 22(4):345–361

De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S, Singh T, Klei L, Kosmicki J, Shih-Chen F, Aleksic B, Biscaldi M, Bolton PF, Brownfeld JM, Cai J, Campbell NG, Carracedo A, Chahrour MH, Chiocchetti AG, Coon H, Crawford EL, Curran SR, Dawson G, Duketis E, Fernandez BA, Gallagher L, Geller E, Guter SJ, Hill RS, Ionita-Laza J, Jimenz Gonzalez P, Kilpinen H, Klauck SM, Kolevzon A, Lee I, Lei I, Lei J, Lehtimaki T, Lin CF, Ma'ayan A, Marshall CR, McInnes AL, Neale B, Owen MJ, Ozaki N, Parellada M, Parr JR, Purcell S, Puura K, Rajagopalan D, Rehnstrom K, Reichenberg A, Sabo A, Sachse M, Sanders SJ, Schafer C, Schulte-Ruther M, Skuse D, Stevens C, Szatmari P, Tammimies K, Valladares O, Voran A, Li-San W, Weiss LA, Willsey AJ, Yu TW, Yuen RK, Study DDD, Consortium UK, Cook EH, Freitag CM, Gill M, Hultman CM, Lehner T, Palotie A, Schellenberg GD, Sklar P, State MW, Sutcliffe JS, Walsh CA, Scherer SW, Zwick ME, Barett JC, Cutler DJ, Roeder K, Devlin B, Daly MJ, Buxbaum JD, Homozygosity mapping 428 collaborative for A (2014) Synaptic, transcriptional and chromatin genes disrupted in autism. Nature 515(7526):209–215

Devlin B, Scherer SW (2012) Genetic architecture in autism spectrum disorder. Curr Opin Genet Dev 22(3):229–237

Elsabbagh M, Divan G, Koh YJ, Kim YS, Kauchali S, Marcin C, Montiel-Nava C, Patel V, Paula CS, Wang C, Yasamy MT, Fombonne E (2012) Global prevalence of autism and other pervasive developmental disorders. Autism Res 5(3):160–179

Hartl DL, Clark AG (2006) Principles of Population Genetics. Sinauer Associates, Inc

Huguet G, Ey E, Bourgeron T (2013) The genetic landscapes of autism spectrum disorders. Annu Rev Genomics Hum Genet 14:191–213

Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A, Kendall J, Grabowska E, Ma B, Marks S, Rodgers L, Stepansky A, Troge J, Andrews P, Bekritsky M, Pradhan K, Ghiban E, Kramer M, Parla J, Demeter R, Fulton LL, Fulton RS, Magrini VJ, Ye K, Darnell JC, Darnell RB, Mardis ER, Wilson RK, Schatz MC, McCombie WR, Wigler M (2012) De novo gene disruptions in children on the autistic spectrum. Neuron 74(2):285–299

Iossifov I, Levy D, Allen J, Ye K, Ronemus M, Lee YH, Yamrom B, Wigler M (2015) Low load for disruptive mutations in autism genes and their biased transmission. Proc Natl Acad Sci USA 112(41):E5600–E5607

Jia P, Sun J, Guo AY, Zhao Z (2010) SZGR: a comprehensive schizophrenia gene resource. Mol Psychiatry 15(5):453–462

Keller MC, Miller G (2006) Resolving the paradox of common, harmful, heritable mental disorders: which evolutionary genetic models work best? Behav Brain Sci 29(4):385–404

King IF, Yandava CN, Mabb AM, Hsiao JS, Huang HS, Pearson BL, Calabrese JM, Starmer J, Parker JS, Magnuson T, Chamberlain SJ, Philpot BD, Zylka MJ (2013) Topoisomerases facilitate transcription of long genes linked to autism. Nature 501(7465):58–62

Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, Raja A, Coe BP, Stessman HA, He ZX, Leal SM, Bernier R, Eichler EE (2015) Excess of rare, inherited truncating mutations in autism. Nat Genet 47(6):582–588

Lord C (2011) Epidemiology: how common is autism? Nature 474(7350):166–168

Lynn DJ, Winsor GL, Chan C, Richard N, Laird MR, Barsky A, Gardy JL, Roche FM, Chan TH, Shah N, Lo R, Naseer M, Que J, Yau M, Acab M, Tulpan D, Whiteside MD, Chikatamarla A, Mah B, Munzner T, Hokamp K, Hancock RE, Brinkman FS (2008) InnateDB: facilitating systems-level analyses of the mammalian innate immune response. Mol Syst Biol 4(1):218

Maenner MJ, Rice CE, Arneson CL, Cunniff C, Schieve LA, Carpenter LA, Van Naarden Braun K, Kirby RS, Bakian AV, Durkin MS (2014) Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. JAMA psychiatry 71(3):292–300

McEvoy BP, Powell JE, Goddard ME, Visscher PM (2011) Human population dispersal "Out of Africa" estimated from linkage disequilibrium and allele frequencies of SNPs. Genome Res 21(6):821–829

Muers M (2012) Human genetics: fruits of exome sequencing for autism. Nat Rev Genet 13(6):377

Myers RA, Casals F, Gauthier J, Hamdan FF, Keebler J, Boyko AR, Bustamante CD, Piton AM, Spiegelman D, Henrion E, Zilversmit M, Hussin J, Quinlan J, Yang Y, Lafreniere RG, Griffing AR, Stone EA, Rouleau GA, Awadalla P (2011) A population genetic approach to mapping neurological disorder genes using deep resequencing. PLoS Genet 7(2):e1001318

Ouwenga RL, Dougherty J (2015) Fmrp targets or not: long, highly brain-expressed genes tend to be implicated in autism and brain disorders. Mol Autism 6(1):16

Petrovski S, Wang Q, Heinzen EL, Allen AS, Goldstein DB (2013) Genic intolerance to functional variation and the interpretation of personal genomes. PLoS Genet 9(8):e1003709

Platt A, Novembre J (2012) A new era of human population genetics. Genome Biol 13(12):182

Ploeger A, Galis F (2011) Evolutionary approaches to autism—an overview and integration. Mcgill J Med 13(2):38

Ploeger A, van der Maas HL, Raijmakers ME, Galis F (2009) Why did the savant syndrome not spread in the population? A psychiatric example of a developmental constraint. Psychiatry Res 166(1):85–90

Posserud M, Lundervold AJ, Lie SA, Gillberg C (2010) The prevalence of autism spectrum disorders: impact of diagnostic instrument and non-response bias. Soc Psychiatry Psychiatr Epidemiol 45(3):319–327

Power RA, Kyaga S, Uher R, MacCabe JH, Langstrom N, Landen M, McGuffin P, Lewis CM, Lichtenstein P, Svensson AC (2013) Fecundity of patients with schizophrenia, autism, bipolar disorder, depression, anorexia nervosa, or substance abuse vs their unaffected siblings. JAMA Psychiatry 70(1):22–30

Richards C, Jones C, Groves L, Moss J, Oliver C (2015) Prevalence of autism spectrum disorder phenomenology in genetic disorders: a systematic review and meta-analysis. Lancet Psychiatry 10:909–916

Samocha KE, Robinson EB, Sanders SJ, Stevens C, Sabo A, McGrath LM, Kosmicki JA, Rehnstrom K, Mallick S, Kirby A, Wall DP, MacArthur DG, Gabriel SB, DePristo M, Purcell SM, Palotie A, Boerwinkle E, Buxbaum JD, Cook EH Jr, Gibbs RA, Schellenberg GD, Sutcliffe JS, Devlin B, Roeder K, Neale BM, Daly MJ (2014) A framework for the interpretation of de novo mutation in human disease. Nat Genet 46(9):944–950

Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S, Goldberg AP, Jinlu C, Keaney JF 3rd, Klei L, Mandell JD, Moreno-De-Luca D, Poultney CS, Robinson EB, Smith L, Solli-Nowlan T, Su MY, Teran NA, Walker MF, Werling DM, Beaudet AL, Cantor RM, Fombonne E, Geschwind DH, Grice DE, Lord C, Lowe JK, Mane SM, Martin DM, Morrow EM, Talkowski ME, Sutcliffe JS, Walsh CA, Yu TW, Ledbetter DH, Martin CL, Cook EH, Buxbaum JD, Daly MJ, Devlin B, Roeder K, State MW (2015) Insights into autism spectrum disorder genomic architecture and biology from 71 risk Loci. Neuron 87(6):1215–1233

Shohat S, Shifman S (2014) Bias towards large genes in autism. Nature 512(7512):E1–E2

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3):585–595

Taylor B, Jick H, Maclaughlin D (2013) Prevalence and incidence rates of autism in the UK: time trend from 2004–2010 in children aged 8 years. BMJ Open 3(10):e003219

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G, Kang HM, Jordan D, Leal SM, Gabriel S, Rieder MJ, Abecasis G, Altshuler D, Nickerson DA, Boerwinkle E, Sunyaev S, Bustamante CD, Bamshad MJ, Akey JM, Broad GO, Seattle GO, Project NES (2012) Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science 337(6090):64–69

Thorisson GA, Muilu J, Brookes AJ (2009) Genotype-phenotype databases: challenges and solutions for the post-genomic era. Nat Rev Genet 10(1):9–18

Tordjman S, Somogyi E, Coulon N, Kermarrec S, Cohen D, Bronsard G, Bonnot O, Weismann-Arcache C, Botbol M, Lauth B, Ginchat V, Roubertoux P, Barburoth M, Kovess V, Geoffray MM, Xavier J (2014) Gene x environment interactions in autism spectrum disorders: role of epigenetic mechanisms. Front in Psychiatry 5:53

Uddin M, Tammimies K, Pellecchia G, Alipanahi B, Hu P, Wang Z, Pinto D, Lau L, Nalpathamkalam T, Marshall CR, Blencowe BJ, Frey BJ, Merico D, Yuen RK, Scherer SW (2014) Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder. Nat Genet 46(7):742–747