bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Close genetic relationships between a spousal pair with autism-affected children and high minor allele content in cases in autism-associated SNPs

Zuobin Zhu[a,b], Xitong Lu[b], Dejian Yuan[b] and Shi Huang[b]*

[a] Department of Genetics, Research Center for Neurobiology, Xuzhou Medical University, Xuzhou, Jiangsu 221004, China.

[b] State Key Laboratory of Medical Genetics, School of Life Sciences, Xiangya Medical School, Central South University.110 Xiangya Road, Changsha, Hunan 410078, China

*Corresponding author: Email: huangshi@sklmg.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Highlights

- The overall properties of the genomes of ASD trios show ASD specific features.
- A spousal pair with ASD-affected children showed smaller genetic distance than that of a randomly selected pair among the spouses in the ASD trios.
- Minor allele contents in ASD-associated SNPs in ASD cases were greater than those in the parents of cases.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

Parents of children affected with autism spectrum disorders (ASD) often have mild forms of autistic-like characteristics. Past studies have focused on searching for individual genetic risk loci of ASD. Here we studied the overall properties of the genomes of ASD trios by using previously published genome-wide data for common SNPs. The pairwise genetic distance (PGD) between a spousal pair with ASD-affected children was found smaller than that of a random pair selected among the spouses in the ASD trios, and spousal relatedness correlated with severe forms of ASD. Furthermore, for a set of 970 ASD associated SNPs, cases showed higher homozygous minor allele content than parents. These results indicate new genetic elements in the broad phenotypes of parents with ASD-affected offspring and in ASD pathogenesis.

Key words: Pairwise genetic distance (PGD), Minor allele content (MAC).

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Autism spectrum disorders (ASD) is a common disease today. The prevalence of ASD is increasing and reached 14.6 per 1,000 (one in 68) children aged 8 years in the United States at 2012[1]. About four times as many males as females are autism [2; 3]. Twin and family studies show that siblings of children with ASD are at a significant higher risk for autism than the general population[4; 5; 6]. Strict (STR) and spectrum (SPC) definition of ASD differ mainly in social deficits[7]. Parents of ASD children are in general of high social economic status (SES) with similar background in science and engineering fields but often have mild forms of autistic-like characteristics or the 'broad phenotype' of autism, such as social and communicative difficulties[8; 9; 10; 11]. High SES and educational attainment are strongly correlated, both of which are also correlated with general intelligence[12; 13]. There is also a genetic component to educational attainment[14]. ASD children also show wide distribution in general intelligence with high functioning individuals performing better than the general population or even their high SES parents in certain tasks[15; 16; 17].

ASD remains poorly understood but may have a strong genetic component with a heritability of 40–80%[18; 19; 20; 21]. ASD are genetically highly heterogeneous, with no single gene accounting for more than 1% of cases[22]. Recent work has shown a substantial contribution of de novo variations[23; 24; 25; 26]. However, genome-wide association studies have revealed few replicable common polymorphisms associated

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

with ASD[27; 28; 29; 30].

Theories of ASD are numerous. According to the hyper-systemizing theory[31; 32], people with ASD have an unusually strong drive to systemize. A comprehensive hypothesis of ASD, taking intelligence and nearly all aspects of ASD into account, has emphasized the role of an optimum level of a suppressive force of innate traits[12].

It has been reported that similar phenotypes and genetic parameters influence preferential mating[33; 34; 35]. Consanguineous marriages appear to increase the prevalence of ASDs[36; 37; 38]. The prevalence of autism might increase by 1.5-fold after 1 generation of assortative mating (≥2.4-fold in the long term) depending on several assumptions[39]. Common genetic variants are individually of little effect but may be a major source of risk for autism[40]. Preferential mating may bring about additive genetic influences in concentrating inherited ASD susceptibility[41]. These observations suggest a potential role for combination of common variants in ASD.

Consistent with the notion of a collective and additive effects of common variants, recent studies indicate a role for genome wide minor allele content (MAC) of an individual in a variety of complex traits and diseases[42; 43; 44; 45]. The more the number of minor alleles of common SNPs in an individual (i.e., the higher the MAC values), the higher the risk in general for many complex diseases such as lung cancer and Parkinson's disease[42; 43; 44; 45]. Such findings indicate an optimum level of genetic variations that an individual can tolerate. Too much lower or higher than the optimum may result in lower fitness and complex diseases[46; 47]. In this study, we investigated whether spousal pairs with ASD-affected children are more genetically alike and whether changes in MAC values may be linked with ASD.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

Genetic similarities between a spousal pair with ASD-affected children

To determine whether the spousal pairs with ASD-affected children may be genetically more related to each other, we calculated the pairwise genetic distance (PGD) between a spousal pair with ASD-affected children relative to the PGD between a randomly chosen pair of male and female within the parents population in the Autism Genome Project (AGP) cohort[48; 49] and in the Miami dataset[50]. The AGP consists of two cohorts, Stage 1 and Stage 2. As controls for the spousal pairs in the ADS trios, we used the control cohort from the Australian twin-family study of alcohol use disorder (OZALC study)[51] and the European family in 1000 genomes project. As general population controls for the AGP ASD trios. We made use of the published SNP genotyping data from these autism and controls studies, and only used shared autosome SNPs for comparative analyses involving different datasets. We performed PCA analysis of control cohorts in combination with the autism cohorts (Supplementary Fig. S1). To minimize population heterogeneity, we only selected individuals that form a relatively tight cluster as shown in Supplementary Fig. S1.

We used two methods to calculate pairwise genetic relationships. The first method measures PGD and made use of a previously described custom script that does not

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

take into account the frequency of the SNPs in the population: all the SNPs contributed equally to genetic distance[45]. For homozygous (Hom) vs Hom mismatch, a difference of 1 was scored. For Hom vs Het or for Het vs Het, a difference of 0.5 was scored. The score of 0.5 for Het vs Het match can be intuitively understood as described in detail in the Methods section. Common software for calculating genetic distance such as PLINK and PEAS typically score Het vs Het as 0 distance, which fail to take into account of haplotype differences and are hence not realistic[52; 53].

To generate random pair distance, we calculated the distance between a father in a parental pair and each of the female individuals in the parents cohort. We then compared the parental pair distance with the middle ranked distance among the randomly paired distances. If the parental pair is indeed closer than a random pair, its distance should be smaller rather than larger or similar to middle ranked random pair distance. We did this for all parental pairs and performed pairwise T test to determine whether the parental pair distance is significantly smaller than the middle ranked random pair. We found that the PGD of a spousal pair was significantly smaller than that of a random pair in both Stage 1 (P = 1.27E-9) and Stage 2 (P = 8.71E-18) cohorts of the AGP. Similar trend was found in the Miami dataset (P = 0.08) (Fig. 1a). However, the same type of analysis did not reveal a difference between spousal pairs and random pairs in the general population OZLAC and CEU (Fig. 1a).

We repeated the experiments by employing an alternative measure of pairwise

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genetic relatedness, the equation 3 of genome wide complex traits analysis(GCTA) that takes into account the frequency of alleles in the population: sharing a rare allele is given a higher weight for genetic relatedness compared to sharing a common allele[54]. This method generates relatedness value rather than distance value with higher value meaning closer genetic distance. Using this relatedness index, we again obtained the same results showing parental pairs having higher relatedness values than random pairs (Stage 1 dataset, P = 0.0009; Stage 2 dataset, P =0.0007; Miami dataset, P =0.06) (Fig 1b). Here, different datasets from independent studies gave different relatedness values, which may be related to sample size variations since the GCTA method makes use of allele frequency the reliability or accuracy of which is dependent on sample size. Small population size would gave inaccurate allele frequency estimation and may make more alleles appear as rare, which would lead to overestimation of genetic relatedness by the GCTA method. The results in Fig. 1b did show a trend of more relatedness in datasets with small population size. Likewise, sample size would also affect estimation of average PGD as shown in Fig. 1a with small sample size being less accurate. So, average PGD between independent datasets may not be identical.

To study if the relatedness of parents correlates with disease status, which would also serve to verify that it was not a result of trivial bias during genotyping experiments such as batch effects, we compared the PGD of a spousal pair with strictly defined (STR) ASD-affected children with that of a spousal pair with spectrum defined (SPC)

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ASD-affected children, expecting SPC cases to be relatively more control like. We also compared the PGD of a spousal pair with multiple incidence (multiplex, MPX) of ASD-affected children with that of a spousal pair with single incidence (simplex, SPX) of ASD-affected children, expecting the SPX cases to be relatively more like the controls. For Stage 1 data of AGP, by using the first method, we found the PGD of a spousal pair with STR (P = 0.08) or MPX (P = 0.04) affected children to be weakly smaller than that with SPC or SPX affected children (Fig. 2a). For Stage 2 data, we did not find significant difference regarding STR and SPC comparisons and there were not enough number of cases to perform MPX versus SPX comparisons (Fig. 2a). The Miami dataset was not suitable for this type of analysis as it had few numbers of relevant cases. By using the second method, we did find the same trend although insignificant (Fig 2b), indicating that allele frequency based distance method may not be sensitive enough in cases where the cohort was relatively small as in the case here with allele frequency less accurate or realistic. Overall, these results indicate close genetic relationship between a spousal pair with ASD children, and a specific association of this relationship with ASD severity.

Minor allele contents in ASD-associated SNPs in ASD cases

It has been reported that Parkinson's disease (PD) population has higher MAC than control population[44] and that mice with more heterozygosity (Het) show more

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sociability than mice with less[55]. Here we tested these genome wide characteristics in the ASD trios and control populations. We calculated the amount of homozygous SNPs, homozygous minor alleles content (homoMAC) and MAC in the ASD trios of AGP and the Miami dataset and the control populations. However, we found no difference in these values between the parents and children in the datasets of either cases or controls (Fig. 3).

To test whether a set of risk SNPs may reveal a difference between parents and their case children, we next studied the amount of homozygous SNPs, homoMAC and MAC using the risk SNPs located in candidate ASD-associated genes as reported by the literature or the SFARI database[56]. We found 1114 SNPs located in the ASD-associated genes. Using the pairwise linkage disequilibrium (LD) option of PLINK, we calculated the LD of each neighboring SNPs, and found 124 SNPs in LD with their neighboring SNPs with R2 > 0.5. Most of these LD groups involved on 2 SNPs and most of 1114 SNPs examined here were not in LD. We randomly selected one among SNPs in LD in order to use only 1 SNP to represent each haplotype. Thereby, we obtained a final set of 970 SNPs as ASD-associated SNPs (Supplementary Table S1). We found no difference in the amount of homozygous SNPs between parents and children (Fig 4a). However, homoMAC values in these SNPs were higher in ASD cases relative to that of their parents (significant in Stage 1 data, P =0.006; and Stage 2 data, P = 0.003; insignificant in the Miami data) (Fig 4b). Also, ASD case children in the Stage 1 (P = 0.0003) and 2 (P = 0.002) and the Miami

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

dataset (*P* = 0.006) had higher MAC than their parents (Fig 4c). We did not find a relationship between homoMAC or MAC and disease severity. Thus, MAC in a small set of risk SNPs located in candidate ASD genes may play a role in ASD disease onset but not in progression.

Discussion

In this study, we found closer genetic similarities between a spousal pair with ASD children than that between randomly paired males and females. Such relatedness of parents correlated with disease status. Furthermore, relative to parents, cases were linked with higher homozygous MAC in ASD-associated SNPs. These results were robust and repeated using two different distance scoring methods and 2 or sometimes 3 ASD datasets.

A caveat with our study involving different datasets from different sources is the potential bias introduced by non-uniform genotyping analyses. We have controlled for this by using only shared SNPs among different datasets and only SNPs with normal frequency distribution consistent with the Hardy Weinberg equation. We used middle ranked distance among all possible random pairs rather than average distance. This should minimize the impact of population outliers on comparing parental pair distance and the distance between random pairs. That ASD spousal pairs are more related is unlikely a result of trivial genotyping batch effects, since spousal relatedness also

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

correlated with disease status. The key findings of this study did not involve any control dataset and was based on the same set of uniformly genotyped data, which include closer ASD spousal pair distance relative to random pairs, correlation of spousal relatedness with disease status, and higher homozygous MAC in ASD-associated SNPs in the ASD-affected children relative to their parents.

A previous study has found a closer genetic similarity between a spousal pair in the general population relative to between randomly paired males and females[34]. However, we did not repeat this observation in our control population. It is likely that even if spousal pairs in the general population are more related than random, it may not be as robust as the ASD spousal pairs. Consistently, ASDs have been found to be more common in consanguinity marriages[36; 37; 57]. The AGP samples were mostly from Europe and some from Canada (about 85% of AGP samples were European ancestry with the remainder other ancestries) while the controls we used were mostly European Americans who may be expected in general to be more genetically diverse than Europeans. We have corrected for this by using PCA analysis to exclude outliers. Regardless, however, these demographic factors may not affect the key findings as listed above that did not involve controls and used only AGP samples.

Our previous studies show increased genome wide MAC values in patients with Parkinson's disease and several other types of complex disease[42; 43; 44; 45]. However, the genome wide MAC values of ASD-affected children were not found to

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

be significantly different from their parents. Nonetheless, they did show increased homoMAC than their parents in ASD-associated SNPs, consistent with a more deleterious role for minor alleles in ASD pathogenesis. The difference in average homoMAC between cases and parents seems small and yet significant. This is consistent with previous findings on PD and other complex traits and diseases[44; 45; 58]. The optimum MAC level may be very finely balanced and slightly abnormal traits such as ASD or PD may result with just a slight excess in MAC.

The finding here strengthens the concept of a genome wide optimum level of genetic diversity[46; 47]. There may exist an optimum level of genetic variations for a set of genes responsible for a particular trait. An increase over such a trait-specific optimum threshold may not necessarily bring about significant change in overall genome wide genetic diversity or MAC in an individual, which would depend on how many loci are involved in controlling a trait, but could nonetheless be sufficient to result in abnormality in the trait and hence disease. Overall, our results here suggest new genetic elements for ASD and the broad phenotypes of parents with ASD-affected children.

Materials and methods

Cohort description

Two GWAS family datasets of ASD, Stage 1 and Stage 2, from the Autism Genome Project (phs000267.v4.p2) and one GWAS family dataset from the Miami study

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(phs000436.v1.p1) were downloaded from database of Genotypes and Phenotypes (dbGaP)[48; 49; 50]. Controls derived from convenience samples. All the controls used in this study were more than 18 years old healthy adults. Similar to previous studies that used convenience controls[40], we reasoned that ASD is sufficiently rare (approximately 1.46%) that screened and unscreened controls would yield similar results[1]. As controls for the spousal pairs in the ADS trios, we downloaded from dbGaP the data on the spouses in the Australian twin-family study of alcohol addiction study (OZALC) (phs000181.v1.p1)[51]. As general population controls for the AGP ASD trios, we downloaded from dbGaP a control populations based on overlapping SNP genotypes and similar demographic profiles. All analyses were done with autosomal SNPs which ~300K were shared with all the datasets used here. Genotype distributions for each SNP were consistent with Hardy-Weinberg equilibrium ($P > 0.01$).

We only used shared SNPs for comparative studies between ASD and controls. The OZALC samples were genotyped on Illumina HumanCNV370v1 containing 370K SNPs, of which ~300K were shared with the AGP data and Miami data.

Population Stratification

The AGP Stage 1 dataset comprised 1471 families, of which 1141 were previously identified to be of European ancestry[48; 49]. AGP Stage 2 dataset had 1301 families, of which 1108 were previously identified to be of European ancestry[48; 49]. The

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Miami dataset comprised of 377 Caucasian families[50]. Principal component analysis (PCA) using the GCTA tool was used to estimate the genetic relatedness[54]. The shared ~300K SNPs that have passed quality control filters were used here. We excluded the outliers by principle component 1-3 (Supplementary Fig. S1). Duplicated individuals were excluded from the analysis. After outlier exclusion, there were 320 spouses in AGP stage 1 trios, 283 spouses in AGP stage 2 trios, 133 spouses in Miami ASD trios, 125 spouses in OZALC data and 22 spouses in CEU data (Supplementary Fig S1).

Comparison of parental pair distance with a random pair distance

To generate random pair distance for comparison with a parental pair, we calculated the distance between a father in a parental pair and each of the female individuals in the parents cohort. We then compared the parental pair distance with the middle ranked distance among the randomly paired distances. We did this for all parental pairs and performed pairwise T test to determine whether the parental pair distance is significantly smaller than the middle ranked random pair. The middle ranked distance among the random pairs, relative to average distance of random pairs, should be less sensitive to influence by very large or very small distances from certain random pairs that may be due to paring of demographically very distant or very close individuals that may represent a small fraction of our selected cohort despite our best effort to select homogeneous populations. For example, 20 outliers in an otherwise

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

homogeneous cohort of 300 may significantly raise the average random pair distance but would only marginally affect the middle ranked random pair distance.

Selection of SNPs within ASD-associated genes

The ASD associated genes were obtained from the published papers[49]. We first selected the ASD-associated SNPs as those that are located within the ASD associated genes, or within 1,000 base pairs upstream or downstream and the UTR regions of these genes. Using the pairwise linkage disequilibrium (LD) option of PLINK, we calculated the LD of each neighboring SNPs (the LD window was 1 million base pairs on the same chromosome). If there were two or more SNPs in LD with $R^2 > 0.5$, we randomly selected one among these SNPs.

Statistical analysis

The population used for calculating the pairwise genetic distance (PGD) were homogeneous groups with outliers excluded by “GCTA” (genome-wide complex trait analysis). PGD were scored using a software as described in previous studies[43,44]. Every non-repetitive pair within a population was scored to produce the average PGD. The PGD software measures genetic distance between two individuals by the number of mismatched SNPs. For homozygous (Hom) vs Hom mismatch such as CC vs TT, a difference of 1 was scored. For Hom vs Het such as CC vs CT, a difference of 0.5 was scored. For Het vs Het such as CT vs CT, a difference of 0.5 was scored, which is based on the following reasoning. When there is AB v AB match, there are two

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

situations depending on the haplotypes. First, if haplotypes are matched, the two hets would be identical (AB matched with AB) and it would take 0 mutation to convert AB to AB. Second, if haplotypes are not matched, AB would be matched with BA. It would take 2 mutations to convert AB into BA. Since only 50% of het vs het matches would be AB vs BA, so the overall number of mutations required to make AB and BA equal to AB in terms of haplotype matches is 0.5 x 2 = 1, which is 50% lower than that required for changing AA to BB. Since we score AA v BB as a difference of 1, the score for AB v AB is naturally 0.5. We verified this approach by comparing the PGD in X chromosome for CEU females vs CEU males using HapMap SNP data and found them to be similar as expected. In contrast, a software based on IBS (identical by status) such as PEAS that score A/B vs A/B as 0 showed the males to have much greater PGD in X than females[53]. For the missing genotypes N/N, N/N vs Hom was scored as 0 and N/N vs Het as 0.5. All the PGD (or the ratio of homozygous genotype) were expressed as total number of the distance (or homozygous SNPs) divided by the total number of SNPs that actual used except the N/N.

The MAF of each SNP was calculated by PLINK and SNP Tools for Microsoft Excel[52; 59]. From MAF data of controls we obtained the MA set, which excluded non-informative SNPs with MAF = 0 in both cases and controls or with MAF = 0.5 in controls.

Minor allele content (MAC) means the ratio of the number of minor alleles divided by

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the total number of SNPs scanned.

Acknowledgements

This work was supported by the National Natural Science Foundation of China grant 81171880 and the National Basic Research Program of China grant 2011CB51001 (S.H.).

The datasets used in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?Db=gap. The dbGaP accession numbers include phs000267.v4.p2, phs000436.v1.p1and phs000181.v1.p1. We thank the Contributing Investigator(phs000267.v4.p2, Stephen W. Scherer, Bernie Devlin,etc; phs000436.v1.p1, Margaret A. Pericak-Vance; phs000181.v1.p1, Andrew C. Heath, Richard D. Todd, Pamela A. Madden, Alexandre Todorov etc. The Autism Genome Project (AGP) Consortium were supported by the National Institutes of Health, Bethesda, MD, USA, (phs000267.v1.p1). (HD055751,HD055782, HD055784, HD35465, MH52708, MH55284, MH57881, MH061009, MH06359, MH066673, MH080647, MH081754, MH66766, NS026630, NS042165 andNS049261). The Miami study were supported by the National Institutes of Health, Bethesda, MD, USA(AUT R01 R01MH080647 and AUT PPG P01NS026630) and Hussman Foundation(Gift). The Australian twin-family study of alcohol use disorder (OZALC study) were supported by HHSN268200782096C. NIH contract "High throughput genotyping for studying the genetic contributions to human disease". National Institutes of Health, Bethesda, MD, USA(phs000181.v1.p1). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References:

[1] D.L. Christensen, J. Baio, K. Van Naarden Braun, D. Bilder, J. Charles, J.N. Constantino, J. Daniels, M.S. Durkin, Robert T. Fitzgerald, M. Kurzius-Spencer, L. Lee, S. Pettygrove, C. Robinson, E. Schulz, C. Wells, M.S. Wingate, W. Zahorodny, and M. Yeargin-Allsopp, Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. Surveillance Summaries 65 (2016) 1-23.

[2] C. Lord, E. Schopler, and D. Revicki, Sex differences in autism. J Autism Dev Disord 12 (1982) 317-30.

[3] L. Wing, Diagnosis and definition. In Early Childhood Autism (2nd edition) (Edited by Wing L.). Pergamon, Oxford (1976).

[4] C.L. Wood, F. Warnell, M. Johnson, A. Hames, M.S. Pearce, H. McConachie, and J.R. Parr, Evidence for ASD recurrence rates and reproductive stoppage from large UK ASD research family databases. Autism Res 8 (2015) 73-81.

[5] A. Bailey, A. Le Couteur, I. Gottesman, P. Bolton, E. Simonoff, E. Yuzda, and M. Rutter, Autism as a strongly genetic disorder: evidence from a British twin study. Psychol Med 25 (1995) 63-77.

[6] D.H. Geschwind, Advances in autism. Annu Rev Med 60 (2009) 367-80.

[7] M.C. Lai, M.V. Lombardo, B. Chakrabarti, and S. Baron-Cohen, Subgrouping the autism "spectrum": reflections on DSM-5. PLoS Biol 11 (2013) e1001544.

[8] D.V. Bishop, M. Maybery, A. Maley, D. Wong, W. Hill, and J. Hallmayer, Using self-report to identify the broad phenotype in parents of children with autistic spectrum disorders: a study using the Autism-Spectrum Quotient. J Child Psychol Psychiatry 45 (2004) 1431-6.

[9] J. Piven, P. Palmer, D. Jacobi, D. Childress, and S. Arndt, Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. Am J Psychiatry 154 (1997) 185-90.

[10] R. Landa, S.E. Folstein, and C. Isaacs, Spontaneous narrative-discourse performance of parents of autistic individuals. J Speech Hear Res 34 (1991) 1339-45.

[11] S. Narayan, B. Moyes, and S. Wolff, Family characteristics of autistic children: a further report. J Autism Dev Disord 20 (1990) 523-35.

[12] R.P.M. Clarke, A theory of general impairment of gene-expression manifesting as autism . Personality and Individual Differences 14 (1993) 465–482.

[13] I.J. Deary, S. Strand, P. Smith, and C. Fernandes, Intelligence and educational achievement . Intelligence 35 (2007) 13–21.

[14] C.A. Rietveld, S.E. Medland, J. Derringer, J. Yang, T. Esko, N.W. Martin, H.J. Westra, K. Shakhbazov, A. Abdellaoui, A. Agrawal, E. Albrecht, B.Z. Alizadeh, N. Amin, J. Barnard, S.E. Baumeister, K.S. Benke, L.F. Bielak, J.A. Boatman, P.A. Boyle, G. Davies, C. de Leeuw, N. Eklund, D.S. Evans, R. Ferhmann, K. Fischer, C. Gieger, H.K. Gjessing, S. Hagg, J.R. Harris, C. Hayward, C. Holzapfel, C.A. Ibrahim-Verbaas, E. Ingelsson, B. Jacobsson, P.K. Joshi, A. Jugessur, M. Kaakinen, S.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Kanoni, J. Karjalainen, I. Kolcic, K. Kristiansson, Z. Kutalik, J. Lahti, S.H. Lee, P. Lin, P.A. Lind, Y. Liu, K. Lohman, M. Loitfelder, G. McMahon, P.M. Vidal, O. Meirelles, L. Milani, R. Myhre, M.L. Nuotio, C.J. Oldmeadow, K.E. Petrovic, W.J. Peyrot, O. Polasek, L. Quaye, E. Reinmaa, J.P. Rice, T.S. Rizzi, H. Schmidt, R. Schmidt, A.V. Smith, J.A. Smith, T. Tanaka, A. Terracciano, M.J. van der Loos, V. Vitart, H. Volzke, J. Wellmann, L. Yu, W. Zhao, J. Allik, J.R. Attia, S. Bandinelli, F. Bastardot, J. Beauchamp, D.A. Bennett, K. Berger, L.J. Bierut, D.I. Boomsma, U. Bultmann, H. Campbell, C.F. Chabris, L. Cherkas, M.K. Chung, F. Cucca, M. de Andrade, P.L. De Jager, J.E. De Neve, I.J. Deary, G.V. Dedoussis, P. Deloukas, M. Dimitriou, G. Eiriksdottir, M.F. Elderson, and J.G. Eriksson, et al., GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science 340 (2013) 1467-71.

[15] R.L. Young, and T. Nettelbeck, The abilities of a musical savant and his family. J Autism Dev Disord 25 (1995) 231-48.

[16] D.A. Treffert, The savant syndrome: an extraordinary condition. A synopsis: past, present, future. Philos Trans R Soc Lond B Biol Sci 364 (2009) 1351-7.

[17] D.A. Treffert, The savant syndrome and autistic disorder. CNS Spectr 4 (1999) 57-60.

[18] D.H. Geschwind, Genetics of autism spectrum disorders. Trends Cogn Sci 15 (2011) 409-16.

[19] J. Hallmayer, S. Cleveland, A. Torres, J. Phillips, B. Cohen, T. Torigoe, J. Miller, A. Fedele, J. Collins, K. Smith, L. Lotspeich, L.A. Croen, S. Ozonoff, C. Lajonchere, J.K. Grether, and N. Risch, Genetic heritability and shared environmental factors among twin pairs with autism. Arch Gen Psychiatry 68 (2011) 1095-102.

[20] E.B. Robinson, K.C. Koenen, M.C. McCormick, K. Munir, V. Hallett, F. Happe, R. Plomin, and A. Ronald, A multivariate twin study of autistic traits in 12-year-olds: testing the fractionable autism triad hypothesis. Behav Genet 42 (2012) 245-55.

[21] T. Gaugler, L. Klei, S.J. Sanders, C.A. Bodea, A.P. Goldberg, A.B. Lee, M. Mahajan, D. Manaa, Y. Pawitan, J. Reichert, S. Ripke, S. Sandin, P. Sklar, O. Svantesson, A. Reichenberg, C.M. Hultman, B. Devlin, K. Roeder, and J.D. Buxbaum, Most genetic risk for autism resides with common variation. Nat Genet 46 (2014) 881-5.

[22] M.H. Chahrour, T.W. Yu, E.T. Lim, B. Ataman, M.E. Coulter, R.S. Hill, C.R. Stevens, C.R. Schubert, M.E. Greenberg, S.B. Gabriel, and C.A. Walsh, Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. PLoS Genet 8 (2012) e1002635.

[23] I. Iossifov, B.J. O'Roak, S.J. Sanders, M. Ronemus, N. Krumm, D. Levy, H.A. Stessman, K.T. Witherspoon, L. Vives, K.E. Patterson, J.D. Smith, B. Paeper, D.A. Nickerson, J. Dea, S. Dong, L.E. Gonzalez, J.D. Mandell, S.M. Mane, M.T. Murtha, C.A. Sullivan, M.F. Walker, Z. Waqar, L. Wei, A.J. Willsey, B. Yamrom, Y.H. Lee, E. Grabowska, E. Dalkic, Z. Wang, S. Marks, P. Andrews, A. Leotta, J. Kendall, I. Hakker, J. Rosenbaum, B. Ma, L. Rodgers, J. Troge, G. Narzisi, S. Yoon, M.C. Schatz, K. Ye, W.R. McCombie, J. Shendure, E.E. Eichler, M.W. State, and M. Wigler, The contribution of de novo coding mutations to autism spectrum disorder. Nature 515 (2014) 216-21.

[24] S.J. Sanders, M.T. Murtha, A.R. Gupta, J.D. Murdoch, M.J. Raubeson, A.J. Willsey, A.G. Ercan-Sencicek, N.M. DiLullo, N.N. Parikshak, J.L. Stein, M.F. Walker, G.T. Ober, N.A. Teran, Y. Song, P. El-Fishawy, R.C. Murtha, M. Choi, J.D. Overton, R.D. Bjornson, N.J. Carriero, K.A. Meyer, K. Bilguvar, S.M. Mane, N. Sestan, R.P. Lifton, M. Gunel, K. Roeder, D.H. Geschwind, B. Devlin, and M.W. State, De novo mutations revealed by whole-exome sequencing are strongly associated with autism. Nature 485 (2012) 237-41.

[25] B.J. O'Roak, L. Vives, S. Girirajan, E. Karakoc, N. Krumm, B.P. Coe, R. Levy, A. Ko, C. Lee, J.D.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Smith, E.H. Turner, I.B. Stanaway, B. Vernot, M. Malig, C. Baker, B. Reilly, J.M. Akey, E. Borenstein, M.J. Rieder, D.A. Nickerson, R. Bernier, J. Shendure, and E.E. Eichler, Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485 (2012) 246-50.

[26] B.M. Neale, Y. Kou, L. Liu, A. Ma'Ayan, K.E. Samocha, A. Sabo, C.F. Lin, C. Stevens, L.S. Wang, V. Makarov, P. Polak, S. Yoon, J. Maguire, E.L. Crawford, N.G. Campbell, E.T. Geller, O. Valladares, C. Schafer, H. Liu, T. Zhao, G. Cai, J. Lihm, R. Dannenfelser, O. Jabado, Z. Peralta, U. Nagaswamy, D. Muzny, J.G. Reid, I. Newsham, Y. Wu, L. Lewis, Y. Han, B.F. Voight, E. Lim, E. Rossin, A. Kirby, J. Flannick, M. Fromer, K. Shakir, T. Fennell, K. Garimella, E. Banks, R. Poplin, S. Gabriel, M. DePristo, J.R. Wimbish, B.E. Boone, S.E. Levy, C. Betancur, S. Sunyaev, E. Boerwinkle, J.D. Buxbaum, E.J. Cook, B. Devlin, R.A. Gibbs, K. Roeder, G.D. Schellenberg, J.S. Sutcliffe, and M.J. Daly, Patterns and rates of exonic de novo mutations in autism spectrum disorders. Nature 485 (2012) 242-5.

[27] B. Devlin, N. Melhem, and K. Roeder, Do common variants play a role in risk for autism? Evidence and theoretical musings. Brain Res 1380 (2011) 78-84.

[28] Y. Pan, J. Chen, H. Guo, J. Ou, Y. Peng, Q. Liu, Y. Shen, L. Shi, Y. Liu, Z. Xiong, T. Zhu, S. Luo, Z. Hu, J. Zhao, and K. Xia, Association of genetic variants of GRIN2B with autism. Sci Rep 5 (2015) 8296.

[29] K. Xia, H. Guo, Z. Hu, G. Xun, L. Zuo, Y. Peng, K. Wang, Y. He, Z. Xiong, L. Sun, Q. Pan, Z. Long, X. Zou, X. Li, W. Li, X. Xu, L. Lu, Y. Liu, Y. Hu, D. Tian, L. Long, J. Ou, Y. Liu, X. Li, L. Zhang, Y. Pan, J. Chen, H. Peng, Q. Liu, X. Luo, W. Su, L. Wu, D. Liang, H. Dai, X. Yan, Y. Feng, B. Tang, J. Li, Z. Miedzybrodzka, J. Xia, Z. Zhang, X. Luo, X. Zhang, C.D. St, J. Zhao, and F. Zhang, Common genetic variants on 1p13.2 associate with risk of autism. Mol Psychiatry 19 (2014) 1212-9.

[30] K. Wang, H. Zhang, D. Ma, M. Bucan, J.T. Glessner, B.S. Abrahams, D. Salyakina, M. Imielinski, J.P. Bradfield, P.M. Sleiman, C.E. Kim, C. Hou, E. Frackelton, R. Chiavacci, N. Takahashi, T. Sakurai, E. Rappaport, C.M. Lajonchere, J. Munson, A. Estes, O. Korvatska, J. Piven, L.I. Sonnenblick, R.A. Alvarez, E.I. Herman, H. Dong, T. Hutman, M. Sigman, S. Ozonoff, A. Klin, T. Owley, J.A. Sweeney, C.W. Brune, R.M. Cantor, R. Bernier, J.R. Gilbert, M.L. Cuccaro, W.M. McMahon, J. Miller, M.W. State, T.H. Wassink, H. Coon, S.E. Levy, R.T. Schultz, J.I. Nurnberger, J.L. Haines, J.S. Sutcliffe, E.H. Cook, N.J. Minshew, J.D. Buxbaum, G. Dawson, S.F. Grant, D.H. Geschwind, M.A. Pericak-Vance, G.D. Schellenberg, and H. Hakonarson, Common genetic variants on 5p14.1 associate with autism spectrum disorders. Nature 459 (2009) 528-33.

[31] S. Baron-Cohen, and S. Wheelwright, 'Obsessions' in children with autism or Asperger syndrome. Content analysis in terms of core domains of cognition. Br J Psychiatry 175 (1999) 484-90.

[32] S. Baron-Cohen, The hyper-systemizing, assortative mating theory of autism. Prog Neuropsychopharmacol Biol Psychiatry 30 (2006) 865-72.

[33] D.A. Roff, and D.J. Fairbairn, The evolution of phenotypes and genetic parameters under preferential mating. Ecol Evol 4 (2014) 2759-76.

[34] B.W. Domingue, J. Fletcher, D. Conley, and J.D. Boardman, Genetic and educational assortative mating among US adults. Proc Natl Acad Sci U S A 111 (2014) 7996-8000.

[35] J.H. Fowler, J.E. Settle, and N.A. Christakis, Correlated genotypes in friendship networks. Proc Natl Acad Sci U S A 108 (2011) 1993-7.

[36] S.M. Al-Salehi, E.H. Al-Hifthy, and M. Ghaziuddin, Autism in Saudi Arabia: presentation, clinical correlates and comorbidity. Transcult Psychiatry 46 (2009) 340-7.

[37] E. Hoodfar, and A.S. Teebi, Genetic referrals of Middle Eastern origin in a western city:

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

inbreeding and disease profile. J Med Genet 33 (1996) 212-5.

[38] M.P. Mamidala, M.K. Kalikiri, K.P. Praveen, N. Rajesh, O.R. Vallamkonda, and V. Rajesh, Consanguinity in India and its association with autism spectrum disorder. Autism Res 8 (2015) 224-8.

[39] W.J. Peyrot, M.R. Robinson, B.W. Penninx, and N.R. Wray, Exploring Boundaries for the Genetic Consequences of Assortative Mating for Psychiatric Traits. JAMA Psychiatry 73 (2016) 1189-1195.

[40] L. Klei, S.J. Sanders, M.T. Murtha, V. Hus, J.K. Lowe, A.J. Willsey, D. Moreno-De-Luca, T.W. Yu, E. Fombonne, D. Geschwind, D.E. Grice, D.H. Ledbetter, C. Lord, S.M. Mane, C.L. Martin, D.M. Martin, E.M. Morrow, C.A. Walsh, N.M. Melhem, P. Chaste, J.S. Sutcliffe, M.W. State, E.J. Cook, K. Roeder, and B. Devlin, Common genetic variants, acting additively, are a major source of risk for autism. Mol Autism 3 (2012) 9.

[41] K. Lyall, J.N. Constantino, M.G. Weisskopf, A.L. Roberts, A. Ascherio, and S.L. Santangelo, Parental social responsiveness and risk of autism spectrum disorder in offspring. JAMA Psychiatry 71 (2014) 936-42.

[42] Z. Zhu, X. Man, M. Xia, Y. Huang, D. Yuan, and S. Huang, Collective effects of SNPs on transgenerational inheritance in Caenorhabditis elegans and budding yeast. Genomics 106 (2015) 23-9.

[43] Z. Zhu, Q. Lu, J. Wang, and S. Huang, Collective effects of common SNPs in foraging decisions in Caenorhabditis elegans and an integrative method of identification of candidate genes. Sci Rep (2015).

[44] Y.D.L.D. Zhu Zuobin, Enrichment of minor alleles of common SNPs and improved risk prediction for Parkinson's disease. PLoS One (2015).

[45] D. Yuan, Z. Zhu, X. Tan, J. Liang, C. Zeng, J. Zhang, J. Chen, L. Ma, A. Dogan, G. Brockmann, O. Goldmann, E. Medina, A.D. Rice, R.W. Moyer, X. Man, K. Yi, Y. Li, Q. Lu, Y. Huang, and S. Huang, Scoring the collective effects of SNPs: association of minor alleles with complex traits in model organisms. Sci China Life Sci 57 (2014) 876-88.

[46] T. Hu, M. Long, D. Yuan, Z. Zhu, Y. Huang, and S. Huang, The genetic equidistance result: misreading by the molecular clock and neutral theory and reinterpretation nearly half of a century later. Sci China Life Sci 56 (2013) 254-61.

[47] S. Huang, New thoughts on an old riddle: what determines genetic diversity within and between species? arXiv: 1510.05918 (2015).

[48] R. Anney, L. Klei, D. Pinto, J. Almeida, E. Bacchelli, G. Baird, N. Bolshakova, S. Bolte, P.F. Bolton, T. Bourgeron, S. Brennan, J. Brian, J. Casey, J. Conroy, C. Correia, C. Corsello, E.L. Crawford, M. de Jonge, R. Delorme, E. Duketis, F. Duque, A. Estes, P. Farrar, B.A. Fernandez, S.E. Folstein, E. Fombonne, J. Gilbert, C. Gillberg, J.T. Glessner, A. Green, J. Green, S.J. Guter, E.A. Heron, R. Holt, J.L. Howe, G. Hughes, V. Hus, R. Igliozzi, S. Jacob, G.P. Kenny, C. Kim, A. Kolevzon, V. Kustanovich, C.M. Lajonchere, J.A. Lamb, M. Law-Smith, M. Leboyer, A. Le Couteur, B.L. Leventhal, X.Q. Liu, F. Lombard, C. Lord, L. Lotspeich, S.C. Lund, T.R. Magalhaes, C. Mantoulan, C.J. McDougle, N.M. Melhem, A. Merikangas, N.J. Minshew, G.K. Mirza, J. Munson, C. Noakes, G. Nygren, K. Papanikolaou, A.T. Pagnamenta, B. Parrini, T. Paton, A. Pickles, D.J. Posey, F. Poustka, J. Ragoussis, R. Regan, W. Roberts, K. Roeder, B. Roge, M.L. Rutter, S. Schlitt, N. Shah, V.C. Sheffield, L. Soorya, I. Sousa, V. Stoppioni, N. Sykes, R. Tancredi, A.P. Thompson, S. Thomson, A. Tryfon, J. Tsiantis, H. Van Engeland, J.B. Vincent, F. Volkmar, J.A. Vorstman, S. Wallace, K. Wing, K. Wittemeyer, S. Wood, D. Zurawiecki, L. Zwaigenbaum, and A.J. Bailey, et al., Individual common variants exert weak effects on the risk for autism spectrum disorderspi. Hum Mol Genet 21 (2012) 4781-92.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[49] R. Anney, L. Klei, D. Pinto, R. Regan, J. Conroy, T.R. Magalhaes, C. Correia, B.S. Abrahams, N. Sykes, A.T. Pagnamenta, J. Almeida, E. Bacchelli, A.J. Bailey, G. Baird, A. Battaglia, T. Berney, N. Bolshakova, S. Bolte, P.F. Bolton, T. Bourgeron, S. Brennan, J. Brian, A.R. Carson, G. Casallo, J. Casey, S.H. Chu, L. Cochrane, C. Corsello, E.L. Crawford, A. Crossett, G. Dawson, M. de Jonge, R. Delorme, I. Drmic, E. Duketis, F. Duque, A. Estes, P. Farrar, B.A. Fernandez, S.E. Folstein, E. Fombonne, C.M. Freitag, J. Gilbert, C. Gillberg, J.T. Glessner, J. Goldberg, J. Green, S.J. Guter, H. Hakonarson, E.A. Heron, M. Hill, R. Holt, J.L. Howe, G. Hughes, V. Hus, R. Igliozzi, C. Kim, S.M. Klauck, A. Kolevzon, O. Korvatska, V. Kustanovich, C.M. Lajonchere, J.A. Lamb, M. Laskawiec, M. Leboyer, A. Le Couteur, B.L. Leventhal, A.C. Lionel, X.Q. Liu, C. Lord, L. Lotspeich, S.C. Lund, E. Maestrini, W. Mahoney, C. Mantoulan, C.R. Marshall, H. McConachie, C.J. McDougle, J. McGrath, W.M. McMahon, N.M. Melhem, A. Merikangas, O. Migita, N.J. Minshew, G.K. Mirza, J. Munson, S.F. Nelson, C. Noakes, A. Noor, G. Nygren, G. Oliveira, K. Papanikolaou, J.R. Parr, B. Parrini, T. Paton, A. Pickles, J. Piven, D.J. Posey, A. Poustka, and F. Poustka, et al., A genome-wide scan for common alleles affecting risk for autism. Hum Mol Genet 19 (2010) 4072-82.

[50] D. Ma, D. Salyakina, J.M. Jaworski, I. Konidari, P.L. Whitehead, A.N. Andersen, J.D. Hoffman, S.H. Slifer, D.J. Hedges, H.N. Cukier, A.J. Griswold, J.L. McCauley, G.W. Beecham, H.H. Wright, R.K. Abramson, E.R. Martin, J.P. Hussman, J.R. Gilbert, M.L. Cuccaro, J.L. Haines, and M.A. Pericak-Vance, A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. Ann Hum Genet 73 (2009) 263-73.

[51] J.D. Grant, A.C. Heath, K.K. Bucholz, P.A. Madden, A. Agrawal, D.J. Statham, and N.G. Martin, Spousal concordance for alcohol dependence: evidence for assortative mating or spousal interaction effects? Alcohol Clin Exp Res 31 (2007) 717-28.

[52] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, and P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81 (2007) 559-75.

[53] S. Xu, S. Gupta, and L. Jin, PEAS V1.0: a package for elementary analysis of SNP data. Mol Ecol Resour 10 (2010) 1085-8.

[54] J. Yang, S.H. Lee, M.E. Goddard, and P.M. Visscher, GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88 (2011) 76-82.

[55] S.S. Moy, J.J. Nadler, A. Perez, R.P. Barbaro, J.M. Johns, T.R. Magnuson, J. Piven, and J.N. Crawley, Sociability and preference for social novelty in five inbred strains: an approach to assess autistic-like behavior in mice. Genes Brain Behav 3 (2004) 287-302.

[56] S.N. Basu, R. Kollu, and S. Banerjee-Basu, AutDB: a gene reference resource for autism research. Nucleic Acids Res 37 (2009) D832-6.

[57] A.H. Bittles, and M.L. Black, Evolution in health and medicine Sackler colloquium: Consanguinity, human evolution, and complex diseases. Proc Natl Acad Sci U S A 107 Suppl 1 (2010) 1779-86.

[58] D. Yuan, Z. Zhu, X. Tan, J. Liang, C. Zeng, J. Zhang, J. Chen, L. Ma, A. Dogan, and G. Brockmann, Minor alleles of common SNPs quantitatively affect traits/diseases and are under both positive and negative selection. 1209.2911 (2012).

[59] B. Chen, S. Wilkening, M. Drechsel, and K. Hemminki, SNP_tools: A compact tool package for analysis and conversion of genotype data for MS-Excel. BMC Res Notes 2 (2009) 214.

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends:

Fig 1. Pairwise genetic distance (PGD) between spousal pairs and between randomly selected pairs. PGD of spousal pairs vs random pairs using either Stage 1 or Stage 2 dataset of AGP and the Miami dataset. Control spouses were from the Australian twin-family study of alcohol addiction study (OZALC) and the CEU data from 1000 genomes. (a) PGD calculated by the custom script. (b) Genetic relatedness calculated by the GCTA method. *** $P < 0.001$, NS $P > 0.05$, paired T test. Data are means ± S.E.M.

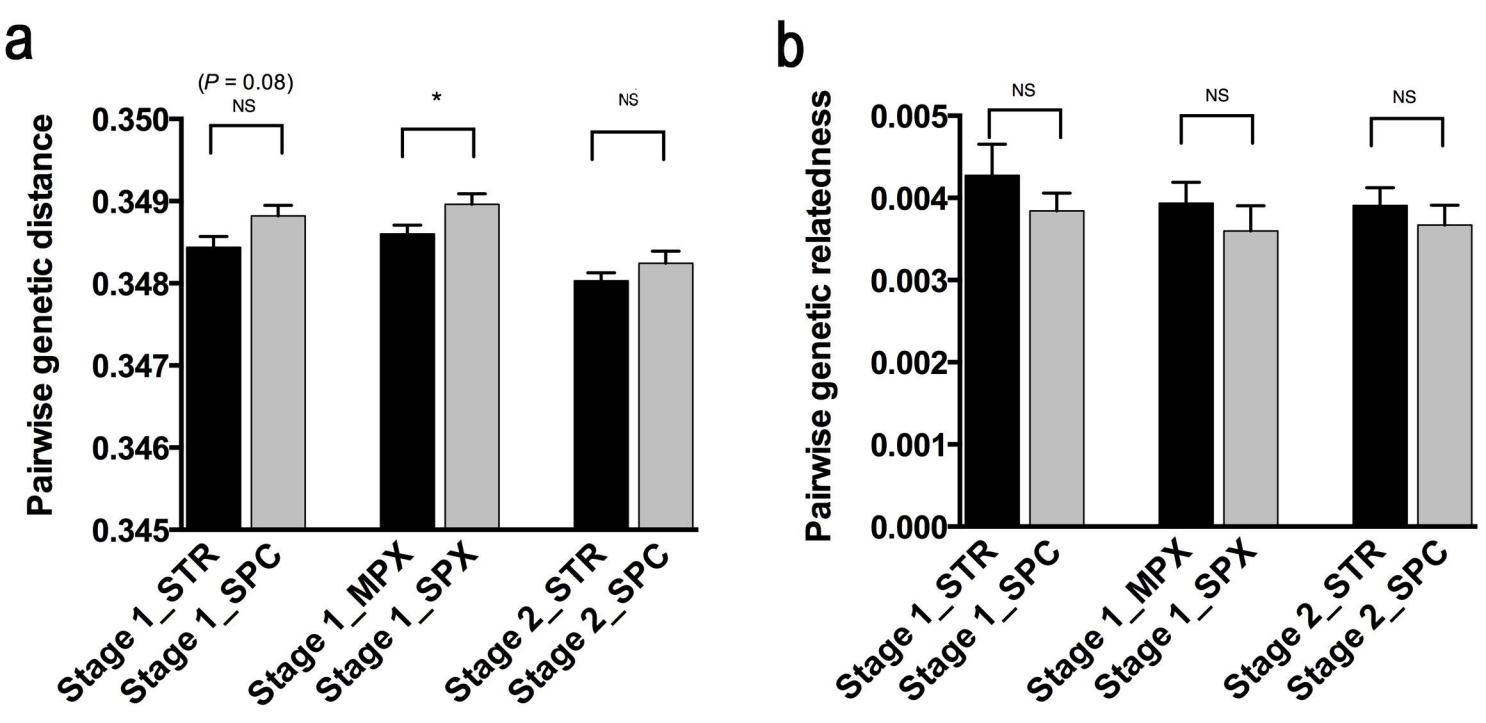
Fig 2. Genetic distance of spousal pairs with STR, SPC, MPX or SPX-affected children. STR: strictly defined ASD, SPC: spectrum defined ASD, MPX: multiplex families, SPX: simplex families. (a) Pairwise genetic distance (PGD) calculated by the custom script. (b) Genetic relatedness calculated by the GCTA method. *$P < 0.05$, NS $P > 0.05$, Z test. Data are means ± S.E.M.

Fig 3. Genomic characteristics in ASD cases and their parents. (a) Shown are

bioRxiv preprint doi: https://doi.org/10.1101/073452; this version posted December 12, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

proportions of homozygous SNPs of parents or ASD-affected children in either Stage 1 or Stage 2 dataset of AGP, the Miami dataset, and the two control datasets. (b) Shown are proportions of homozygous minor allele genotype of parents or children. (c) Shown are minor allele content (MAC) of parents or children. NS $P > 0.05$., Z test. Data are means ± S.E.M.

Fig 4. The proportion of homozygous risk SNPs, homozygous minor allele contents and MAC in ASD trios. (a) Shown are proportions of homozygous SNPs of parents or ASD-affected children in either Stage 1 or Stage 2 dataset of AGP, the Miami dataset, and the two control datasets. (b) Shown are proportions of homo MAC of parents or children. (c) Shown are MAC of parents or children. ** $P < 0.01$, ** $P < 0.01$, NS $P > 0.05$, Z test. Data are means ± S.E.M.


a
(P = 0.08)
NS
*
NS
Pairwise genetic distance
0.350
0.349
0.348
0.347
0.346
0.345
Stage 1_STR
Stage 1_SPC
Stage 1_MPX
Stage 1_SPX
Stage 2_STR
Stage 2_SPC
b
NS
NS
NS
Pairwise genetic relatedness
0.005
0.004
0.003
0.002
0.001
0.000
Stage 1_STR
Stage 1_SPC
Stage 1_MPX
Stage 1_SPX
Stage 2_STR
Stage 2_SPC