





 Open access • Posted Content • DOI:10.1101/077222

## **Neolithic super-grandfather Y haplotypes, their related surnames, and autism spectrum disorder — [Source link](#)**

He P, Chen N, Zhi Hu, Zuobin Zhu ...+2 more authors

**Institutions:** Central South University

**Published on:** 30 Sep 2016 - bioRxiv (Cold Spring Harbor Laboratory)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/neolithic-super-grandfather-y-haplotypes-their-related-12fupokpza>

1 **Neolithic super-grandfather Y haplotypes, their related surnames, and autism**  
2 **spectrum disorder**

3

4 Pei He, Na Chen, Zhengmao Hu, Zuobin Zhu<sup>1</sup>, Kun Xia, and Shi Huang\*

5

6 Laboratory of Medical Genetics, School of Life Sciences, Central South University

7 110 Xiangya Road, Changsha, Hunan, 410078, China

8

9 <sup>1</sup> Present address: Department of Genetics, Xuzhou Medical University, Xuzhou, Jiangsu

10 221004, China

11

12 \*Corresponding author: [huangshi@sklmg.edu.cn](mailto:huangshi@sklmg.edu.cn);

13

14 **Key words:** Neolithic super-grandfather Y haplotypes, Yan and Huang, assortative mating,  
15 social economic status (SES), autism spectrum disorder (ASD)

16 **Running title:** Super-grandfather Y haplotypes, their related surnames, and autism

17

18

19

20 **Abstract**

21           Recent studies found three Neolithic super-grandfather Y haplotypes among Han  
22 Chinese, consistent with the legend of Yan and Huang Emperors. Individuals of royal and noble  
23 ancestry or high social economic status (SES) are known to practice assortative mating and  
24 consanguineous marriages, which can produce offspring of both higher and lower than average  
25 fitness. However, the roles of the super-grandfather Y haplotypes and their descendant lines in  
26 history, fitness, and the male biased autism spectrum disorder (ASD) remain unknown. Here we  
27 show a link between the super-grandfathers and the legend of *Yan-Huang* Emperors and  
28 between their descendant haplotypes and ASD. We found that subjects carrying the O3a1c and  
29 O3a2c1a super-grandfather haplotypes were enriched with *Yan* and *Huang* related surnames,  
30 respectively, in two independent datasets of 1564 and 772 male Han subjects. We identified  
31 high and low SES descendant haplotypes of the super-grandfathers using the Han dataset of  
32 the 1000 genomes project based on two criteria: more descendant branches and fewer  
33 mutations before star-like expansions. By genotyping 505 fathers of ASD affected male children  
34 from the Autism Clinical and Genetic Resources in China with surnames either closely related to  
35 Huang (Ying group) or less related (Ji group), we found the high SES haplotypes within the  
36 O3a2c1a clade at ~2 fold lower (odds ratio 2.05, 95% CI 1.28-3.26, P=0.0026) while the low  
37 SES haplotypes at ~2 fold higher frequency (odds ratio 1.92, 95% CI 1.01-3.64, P = 0.046) in  
38 the fathers relative to 505 normal subjects. The fraction of low SES haplotypes was greater than  
39 that of high SES in ASD fathers of the Ying group, in contrast to Ying controls or the Ji fathers  
40 and Ji controls. Consistently, analysis of 2366 ASD affected children showed higher male to  
41 female ratio for Ying versus Ji group (6.52 +/-1.11 v 4.59 +/-0.41, P = 0.028, one tailed). These  
42 results provide evidence for the Yan-Huang legend and suggest a role for Y in ASD.

43

## 44 **Introduction**

45           The Han Chinese population uses hereditary surnames that are thought to be first  
46 established ~5000 years ago (1-3). Modern Han Chinese people are thought to be largely  
47 descended from *Yan Di* (Yan Emperor) and *Huang Di* (*Huang* Emperor) who lived ~5000 years  
48 ago. There were “Eight Great *Xings* of High Antiquity” from ~4000 years ago that are believed to  
49 be ancestors of most of today’s ~23813 surnames of Chinese people (4). Although these Eight  
50 Great *Xings* are thought to originate in matriarchal societies, it is expected that certain males  
51 may be more dominant than others in such societies. *Yan* belonged to one of the Eight Great  
52 *Xings* (*Jiang*) and the other 7 Great *Xings* are all related to *Huang*. Of these, *Ji* is thought to be  
53 the original surname of *Huang* and has the most descendant surnames today. *Ying* is special  
54 because one of its related contemporary surnames, *Huang*, is also the same as the commonly  
55 used name for the Yellow Emperor or *Huang Di*.

56           Consistent with Neolithic individuals matching the legendary status of *Yan-Huang*, there  
57 were three Neolithic super-grandfathers (5). Their Y haplotypes originated ~5.4 Kya (thousand  
58 years ago) for O3a2c1a-Page23 or M117 (O2a2b1a1, ISOGG 2017), ~6.5 Kya for O3a2c1-F46  
59 (O2a2b1a2a1), and ~6.8 Kya for O3a1c-F11 (O2a1c1a1a), and represent 16%, 11%, and 14%  
60 of present Han Chinese, respectively. Based on the estimated age and frequency, O3a2c1a-  
61 Page23 could be a good candidate for *Huang* and O3a1c-F11 for *Yan*. Therefore, we here  
62 tested whether contemporary Han males with surnames or *Xings* more closely related to *Yan*  
63 and *Huang* are enriched with O3a1c and O3a2c1a, respectively.

64           Individuals of royal and noble ancestry or high social economic status (SES) are known  
65 to practice assortative mating and consanguineous marriages (6, 7), which can produce  
66 offspring of both higher and lower than average fitness (8). One of the diseases associated with  
67 assortative mating is autism spectrum disorders (ASD) (9-11). Parents with ASD children often

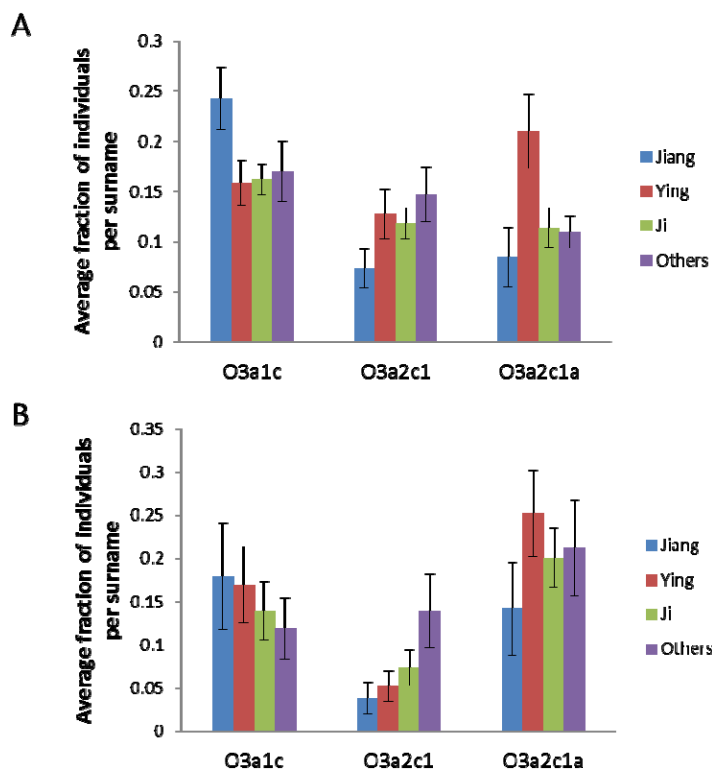
68 have mild forms of autistic-like characteristics (12). The male to female ratio is 4:1 in the global  
69 ASD population, but is 23:1 in ASD subjects without physical or brain abnormalities (13). Little is  
70 known about the male bias in ASD (14). We hypothesized that descendant haplotypes of a  
71 super-grandfather may show dimorphism, with some associated with high SES and better  
72 fitness while others the opposite. In conjunction with studying high SES haplotypes and  
73 surnames among normal Han Chinese subjects, we studied a large ASD cohort from the Autism  
74 Clinical and Genetic Resources in China (ACGC) (15).

75

## 76 **Results:**

77 We first made use of the Y haplotype data of surnames representing 1564 males as  
78 reported on the Website “One Surname a Week” maintained by researchers from Fudan  
79 University. To determine Y haplotype distribution among the Eight Great *Xings*, we divided  
80 contemporary surnames into 4 groups of Great *Xings*, Jiang, Ying, Ji, and Others according to  
81 popular surnames literatures (Supplementary Table S1). We obtained the average fraction of  
82 individuals per surname for each of the 3 super-grandfather haplotypes (Figure 1A). The Jiang-  
83 group has more O3a1c than each of the other groups ( $P < 0.05$ , Student’s t test, one tailed).  
84 The Ying-group has more O3a2c1a than each of the other groups ( $P < 0.02$ , Student’s t test,  
85 one tailed). To verify the above result, we collected peripheral blood samples from healthy  
86 subjects in the Hunan area in China and did PCR-sequencing on the 3 haplotypes. The results  
87 on 772 males again showed similar patterns of O3a1c enrichment in the Jiang-group, and  
88 O3a2c1a enrichment in the Ying-group (Figure 1B and Supplementary Table S2). The  
89 combined results from these two surveys were significant, as the probability of getting both  
90 haplotypes correctly matched to their respective candidate groups is 1/144 or 0.007 (the chance  
91 of randomly matching a haplotype to its surname group is 1/16 [4 groups and 2 surveys] and the

92 chance of getting a second haplotype correctly matched is 1/9 [3 remaining groups and 2  
93 surveys]).



94  
95 **Figure 1. Distribution of the three super-grandfather Y haplotypes.** The fraction of each  
96 haplotype in a surname was calculated and the average fractions per surname with standard  
97 error of the mean are shown in the plot. Shown are results from 1564 male subjects with  
98 surname and Y data from the website "One Surname a Week" (A) and from 772 male subjects  
99 collected in this study (B).

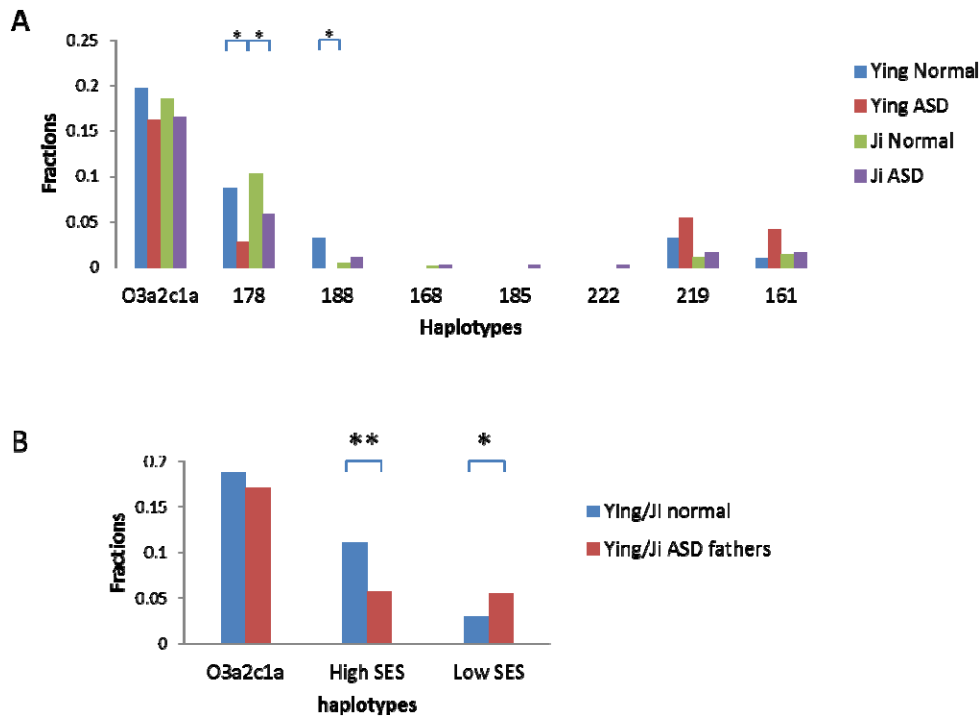
100  
101 We made use a previous study on the 1000 genomes project (1kGP) dataset that  
102 showed star like expansion for the three super-grandfathers (Supplementary Figure S1) (16).  
103 We focused on Han Chinese in this dataset and assigned high SES Y haplotypes based on two  
104 criteria: more descendant branches and fewer mutations before star-like expansions since a

105 super-grandfather's early descendants were more likely to be high SES individuals. Within the  
106 O3a2c1a-Page23 clade, branch 225 as indexed by Poznik et. al. (2016) (Supplementary Figure  
107 S1) was identified as the super-grandfather haplotype (16). Under branch 225, 4 branches (168,  
108 178, 185, and 188) appear to be high SES as each of these has only 1 SNP before further  
109 branching splits with branch 178 having the largest number of descendant lines. The remaining  
110 3 branches (161, 219, and 222) were identified as low SES haplotypes among Han Chinese.

111 Within the O3a2c1-M1561 clade, the super-grandfather haplotype was identified as  
112 branch 150 defined by F46, which has 2 sub-branches with the sub-branch 149 as the more  
113 super-grandfather like. Of the 3 descendant lines from branch 149, the high SES branch was  
114 identified as 131, followed by 149 and 148. Within the O3a1c-002611 clade, the super-  
115 grandfather haplotype was identified as branch 69 defined by F11, which also has 2 sub-  
116 branches with the sub-branch 68 as the more super-grandfather like. Of the descendant lines  
117 from branch 68, the high SES branch was identified as 57, followed by 64 and 40.

118 To study the roles of Y and surnames in ASD, we focused on the O3a2c1a clade and  
119 the Huang related surname groups Ying and Ji. We genotyped 91 Ying and 414 (or 355 for  
120 some haplotypes) Ji normal subjects and 202 Ying and 303 Ji fathers of ASD affected male  
121 children (Supplementary Table S3). No significant differences in frequencies were observed  
122 between the normal and ASD subjects in the O3a2c1a-page23 haplotype (Figure 2). However,  
123 relative to normal Ying subjects, Ying fathers showed >3 fold lower fraction of the two high SES  
124 haplotypes 178 and 188 ( $P < 0.05$ ). Relative to normal Ji subjects, Ji fathers showed 1.75 fold  
125 lower fraction of 178 ( $P < 0.05$ ). As the low SES haplotypes each had too few samples to be  
126 informative, we combined Ying and Ji subjects and low or high SES haplotypes for further  
127 analysis. We found the high SES haplotypes (178, 188, 168 and 185 combined) at ~2 fold  
128 lower (Odds ratio 2.05, 95% CI 1.28-3.26,  $P=0.0026$ ) while the low SES haplotypes (222, 219  
129 and 161) at ~2 fold higher frequency (odds ratio 1.92, 95% CI 1.01-3.64,  $P = 0.046$ ) in the

130 fathers relative to normal Ying and Ji subjects (Figure 2B). The results indicated a relationship  
131 between ASD and Y haplotypes, especially for the Ying subjects relative to the Ji subjects,  
132 consistent with the above noted association between Ying and the O3a2c1a super-grandfather.



133

134 **Figure 2. Frequency of O3a2c1a related haplotypes.** **A.** Fractions of O3a2c1a and related  
135 downstream haplotypes in normal or fathers of ASD affected male children with Ying or Ji  
136 related surnames. **B.** Fractions of O3a2c1a and related downstream high SES (178, 188, 168  
137 and 185) and low SES haplotypes (222, 219 and 161) in subjects of combined Ying and Ji  
138 groups. \*\*,  $P < 0.01$ , \*,  $P < 0.05$ , Chi square test, 2 tailed.

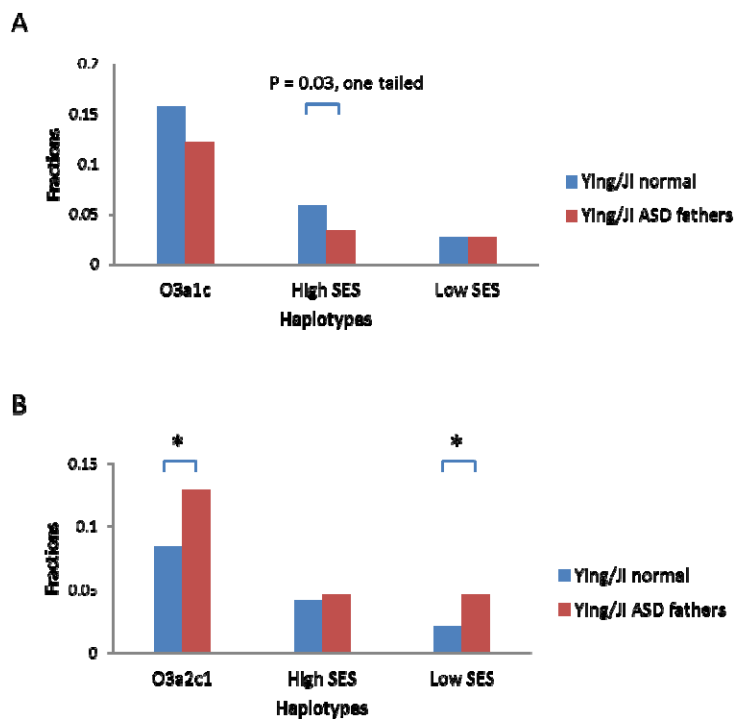
139

140 To confirm this pattern of ASD link with Y SES status, we next studied the other two  
141 super-grandfather haplotypes in the combined Ying and Ji subjects. For the O3a1c-002611  
142 clade, ASD fathers showed 1.76 fold ( $P < 0.03$ , one tailed Chi square test) lower frequency in



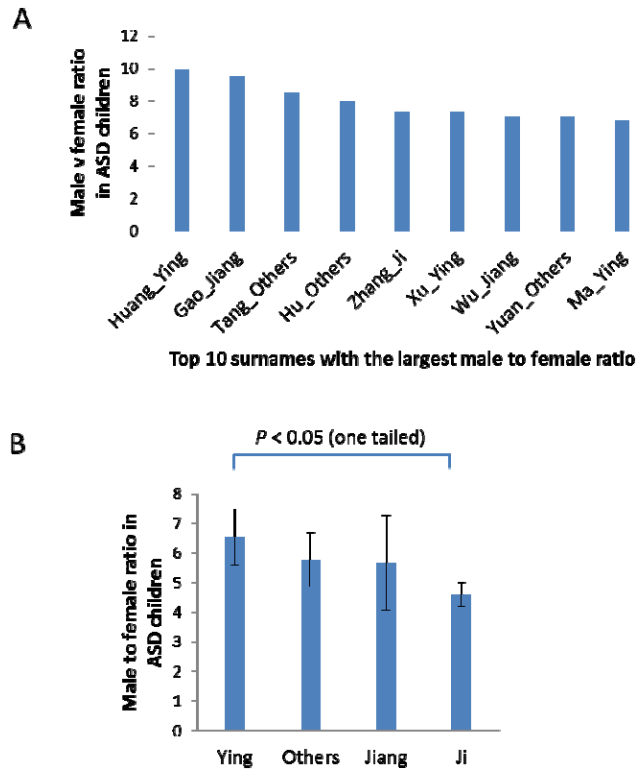
143 the high SES haplotype (branch 57) but no significant difference in the low SES haplotypes  
144 (branches 64 and 40) or the O3a1c-002611 haplotype relative to normal Ying and Ji subjects  
145 (Figure 3A). For the O3a2c1-M1561 clade, ASD fathers showed 2.22 fold ( $P < 0.05$ , Chi square  
146 test, 2 tailed) and 1.52 fold ( $P < 0.05$ , Chi square test, 2 tailed) higher frequency for the low SES  
147 haplotypes (branches 148 and 140) and the O3a2c1-M1561 haplotype, respectively (Figure 3B).  
148 These results provide additional data for a consistent link between ASD and SES status of Y  
149 haplotypes and ASD.

150



151

152 **Figure 3. Frequency of O3a1c and O3a2c1 related haplotypes.** **A.** Fractions of O3a1c and  
153 related downstream high SES (57) and low SES haplotypes (64 and 40) in subjects of combined  
154 Ying and Ji groups. **B.** Fractions of O3a2c1 and related downstream high SES (131) and low  
155 SES haplotypes (148 and 140) in subjects of combined Ying and Ji groups. \*,  $P < 0.05$ , Chi  
156 square test, 2 tailed.



157

158 **Figure 4. Male to female ratio in ASD children.** Only top 40 surnames of China were counted  
159 among the 2362 ASD children studied. The minimal number of female cases was 2 in order to  
160 for a surname to be considered. **A.** Male to female ratio of ASD children with surnames among  
161 the top 10 most sex-biased surnames. **B.** Average male to female ratio of ASD children for each  
162 of the 4 groups of surnames. Also shown are the Standard Errors of the Mean and P value from  
163 Student's t test, one tailed.

164

165 If certain low SES Y haplotypes are enriched in ASD, one would expect surnames linked  
166 with those haplotypes to be also enriched in ASD. The above analysis showed a more  
167 consistent enrichment of low SES Y haplotypes in ASD for the Ying relative to the Ji group  
168 (Figure 2A). We further confirmed this by finding that, while the majority of normal Ying subjects  
169 carried the top ranked SES haplotypes 178, 188, 168, 185, 131, and 57, most Ying ASD cases

170 carried low SES haplotypes 161, 219, 148, 140, 64, and 40 (fractions of high and low SES were  
171 0.25 and 0.12 for Ying normal and 0.11 and 0.17 for Ying ASD, respectively), which was in  
172 contrast to the Ji group (0.2 and 0.07 for Ji normal and 0.16 and 0.1 for Ji ASD, respectively).  
173 Therefore one expects ASD cases with Ying surnames to be more affected by Y and show more  
174 extreme male bias than Ji-related cases. We analyzed 2362 ASD affected children for their  
175 distribution among the top 40 surnames in China (with at least 2 female cases in our dataset  
176 here) and found 3 of 6 Ying-related surnames (Huang, Xu and Ma) ranked among the top 10  
177 surnames in male to female ratio whereas only 1/13 Ji-related surnames did ( $P < 0.05$ , Chi  
178 square test, one tailed, Figure 4A). The average ratio of Ying group was higher than the other  
179 three groups and significantly higher than the Ji group ( $P < 0.05$ , Student's t test, one tailed,  
180 Figure 4B, Supplementary Table S4).

181

## 182 **Discussion**

183 The association of the Jiang-group and Ying-group of surnames with O3a1c and  
184 O3a2c1a respectively suggests *Yan* and *Huang* as the candidate ancestors of these haplotypes,  
185 respectively. That Ji-group of surnames has less representation of O3a2c1a than Ying-group  
186 indicates more admixture for Ji related populations. Since the *Huang*-related haplotype  
187 O3a2c1a is the youngest among the three, and yet has claimed similar if not more descendants  
188 than the *Yan*-related O3a1c, the oldest of the three, the pace of expansion for the *Huang*  
189 lineage appears to be the fastest among the three Y haplotypes, consistent with *Huang* being  
190 the ultimate victor among the legendary leaders.

191 Previous studies on Europeans have produced either weak or no associations between  
192 Y haplotypes and ASD (17, 18). However, those studies did not consider surnames and finer  
193 classifications of Y haplotypes. The top ranked high SES descendant haplotype (178, 131, and

194 57) under each of the 3 super-grandfathers consistently showed a trend of lower fraction in ASD  
195 subjects with Ying related surnames, while the bottom ranked haplotypes (219, 161, 140, 40)  
196 showed the opposite. Also, ASD children of the Ying group showed expected higher male to  
197 female ratio. Therefore, population stratification is unlikely to account for these results.

198 Our results suggest a role for Y haplotypes in the male bias in ASD. The results also  
199 indicate that certain haplotypes may confer fitness advantages. A super-grandfather may leave  
200 many descendants for at least two reasons, fitter Y and more partners. Descendants with Y  
201 haplotypes more similar to the super-grandfather would continue to enjoy fitter traits and leave  
202 more descendants. Future studies along this line of investigation may help understand sex  
203 dimorphism in other diseases (19).

204

## 205 **Acknowledgements**

206 Supported by the National Natural Science Foundation of China (81171880, 81330027,  
207 81525007 and 31400919) and the National Basic Research Program of China (2011CB51001,  
208 2012CB517900).

209

## 210 **Author contributions:**

211 SH and PH conceived the project. PH, NC, and ZZ performed DNA analysis. ZH and KX  
212 contributed the ASD and some normal DNA samples. SH and PH wrote the manuscript and all  
213 authors provided comments on the manuscript.

214 **References:**

- 215 1. Y. D. Yuan, Science and culture of surnames. *Chinese Nat. Geog.* **2**, 38 (2007).
- 216 2. Y. Liu, L. Chen, Y. Yuan, J. Chen, A study of surnames in China through isonymy. *Am J*  
217 *Phys Anthropol* **148**, 341 (Jul, 2012).
- 218 3. M. A. Jobling, In the name of the father: surnames and genetics. *Trends Genet* **17**, 353  
219 (Jun, 2001).
- 220 4. Y. D. Yuan, J. R. Qiu, *Dictionary of Chinese Surnames*. (Jiangxi People's Press, Jiangxi,  
221 China, 2010).
- 222 5. S. Yan *et al.*, Y chromosomes of 40% Chinese descend from three Neolithic super-  
223 grandfathers. *PLoS ONE* **9**, e105691 (2014).
- 224 6. B. W. Domingue, J. Fletcher, D. Conley, J. D. Boardman, Genetic and educational  
225 assortative mating among US adults. *Proc Natl Acad Sci U S A* **111**, 7996 (Jun 03,  
226 2014).
- 227 7. D. A. Roff, D. J. Fairbairn, The evolution of phenotypes and genetic parameters under  
228 preferential mating. *Ecology and evolution* **4**, 2759 (Jul, 2014).
- 229 8. A. Helgason, S. Palsson, D. F. Gudbjartsson, T. Kristjansson, K. Stefansson, An  
230 association between the kinship and fertility of human couples. *Science* **319**, 813 (Feb  
231 08, 2008).
- 232 9. Z. Zhu, X. Lu, D. Yuan, S. Huang, Close genetic relationships between a spousal pair  
233 with autism-affected children and high minor allele content in cases in autism-associated  
234 SNPs. *Genomics*, 10.1016/j.ygeno.2016.12.001 (Dec 09, 2016).
- 235 10. W. J. Peyrot, M. R. Robinson, B. W. Penninx, N. R. Wray, Exploring Boundaries for the  
236 Genetic Consequences of Assortative Mating for Psychiatric Traits. *JAMA psychiatry* **73**,  
237 1189 (Nov 01, 2016).

- 238 11. D. L. Christensen *et al.*, Prevalence and Characteristics of Autism Spectrum Disorder  
239 Among Children Aged 8 Years--Autism and Developmental Disabilities Monitoring  
240 Network, 11 Sites, United States, 2012. *MMWR Surveill Summ* **65**, 1 (Apr 01, 2016).
- 241 12. D. V. Bishop *et al.*, Using self-report to identify the broad phenotype in parents of  
242 children with autistic spectrum disorders: a study using the Autism-Spectrum Quotient.  
243 *Journal of child psychology and psychiatry, and allied disciplines* **45**, 1431 (Nov, 2004).
- 244 13. J. H. Miles, R. E. Hillman, Value of a clinical morphology examination in autism.  
245 *American journal of medical genetics* **91**, 245 (Apr 10, 2000).
- 246 14. S. Baron-Cohen *et al.*, Why are autism spectrum conditions more prevalent in males?  
247 *PLoS Biol* **9**, e1001081 (Jun, 2011).
- 248 15. T. Wang *et al.*, De novo genic mutations among a Chinese autism spectrum disorder  
249 cohort. *Nature communications* **7**, 13316 (Nov 08, 2016).
- 250 16. G. D. Poznik *et al.*, Punctuated bursts in human male demography inferred from 1,244  
251 worldwide Y-chromosome sequences. *Nat Genet* **48**, 593 (Jun, 2016).
- 252 17. F. J. Serajee, A. H. Mahbubul Huq, Association of Y chromosome haplotypes with  
253 autism. *Journal of child neurology* **24**, 1258 (Oct, 2009).
- 254 18. S. Jamain *et al.*, Y chromosome haplogroups in autistic subjects. *Mol Psychiatry* **7**, 217  
255 (2002).
- 256 19. F. J. Charchar *et al.*, Inheritance of coronary artery disease in men: an analysis of the  
257 role of the Y chromosome. *Lancet* **379**, 915 (Mar 10, 2012).

258

259

260

261 **Supplementary Information:**

262 **Supplementary Information text:**

263 1. Chinese surnames and the Eight Great Xings of High Antiquity

264 2. Supplementary Figure S1

265 **Supplementary Table S1.** Y chromosome haplotype distribution among Chinese surnames

266 based on data from "One surname a week" website.

267

268 **Supplementary Table S2.** Distribution of the three Neolithic super-grandfather Y haplotypes

269 among 772 male samples collected in this study.

270

271 **Supplementary Table S3.** Haplotype profiles of ASD and normal subjects.

272

273 **Supplementary Table S4.** Surname profiles of ASD children in the Autism Clinical and Genetic

274 Resources in China.

275 **Supplementary Table S5.** List of SNPs for haplotype genotyping.

276