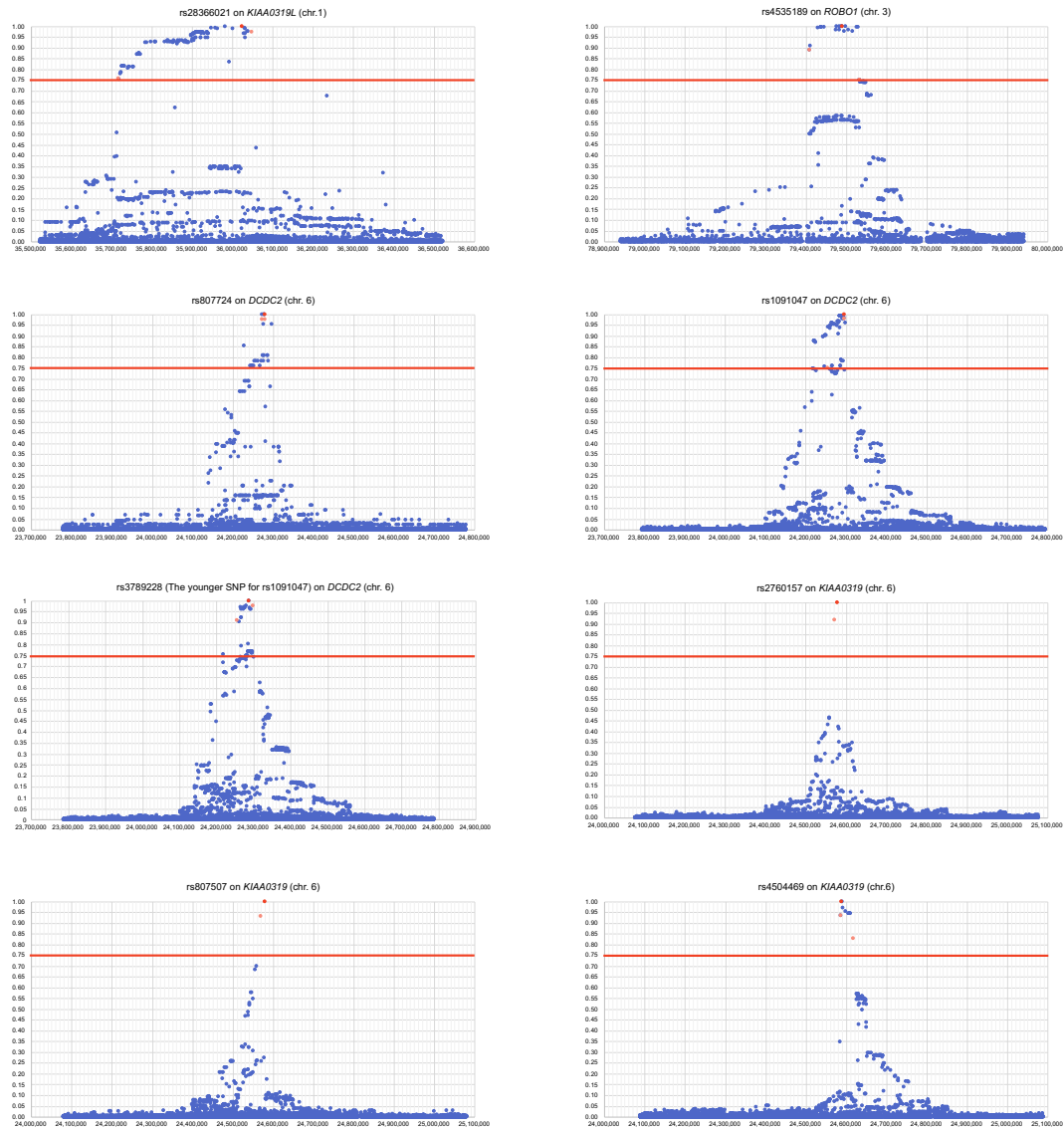# Supplementary Figures

**Do genes associated with dyslexia of Chinese characters evolve neutrally?**

Kumiko V. Nishiyama, Yoko Satta and Jun Gojobori

Figure S1 r^2 with the core SNP in a 1-Mb region, where the core SNP is centered (red dot). Pink dots represent SNPs that are boundaries of core regions. In some cases, core SNPs themselves are one of the boundaries. Red line represents $r^2 = 0.75$. The SNP is not taken as the boundary in the case when a derived allele at the core SNP is linked to ancestral alleles in the linked SNPs, and when ancestral allele at the core SNP is linked to derived alleles in linked SNPs, even if $r^2 > 0.75$. The surrounding region of rs2255526 on *DIP2A* is less than 1-Mb because the core SNP is located near the end of chromosome 21.
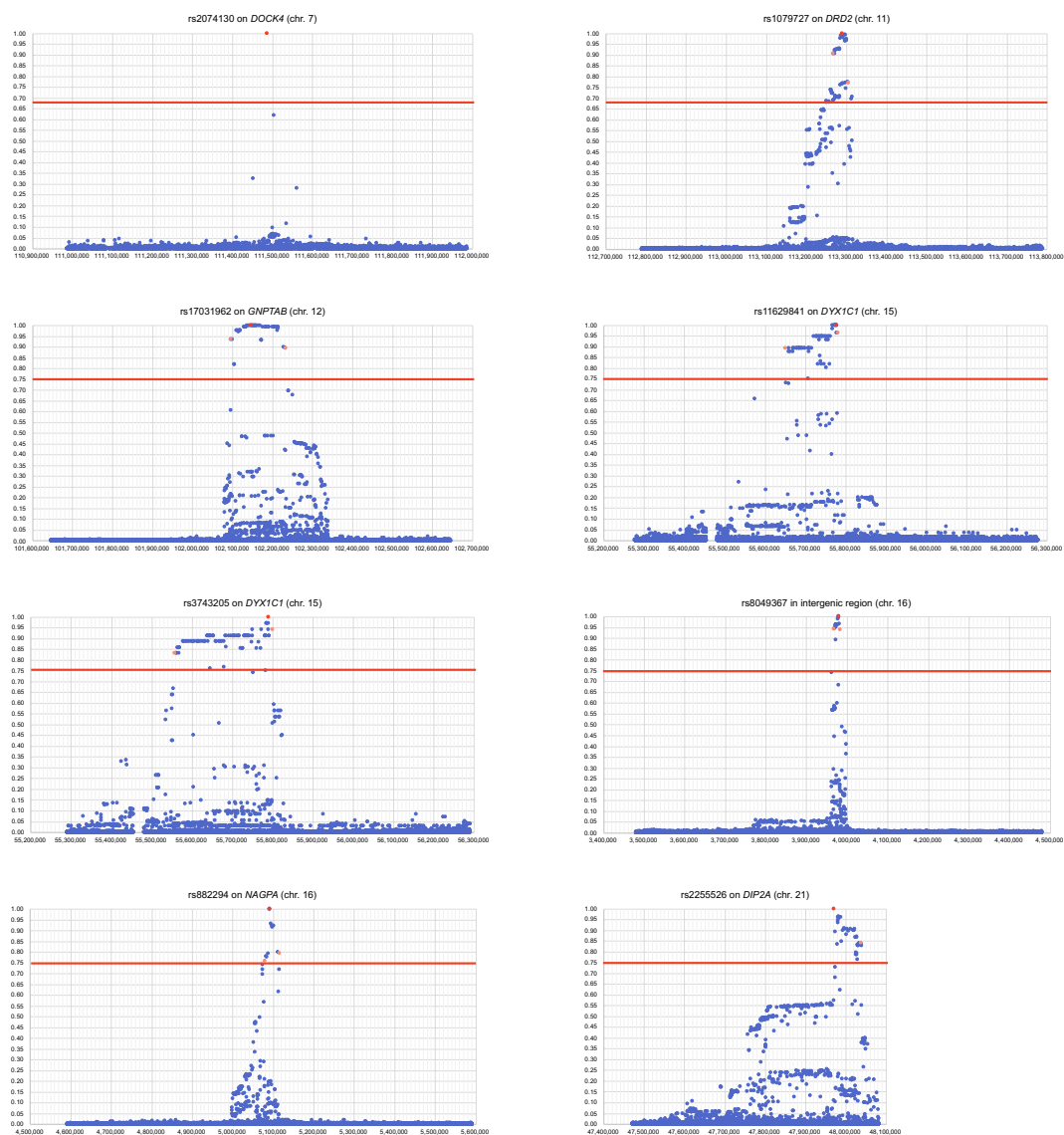
Figure S1 (Continued)
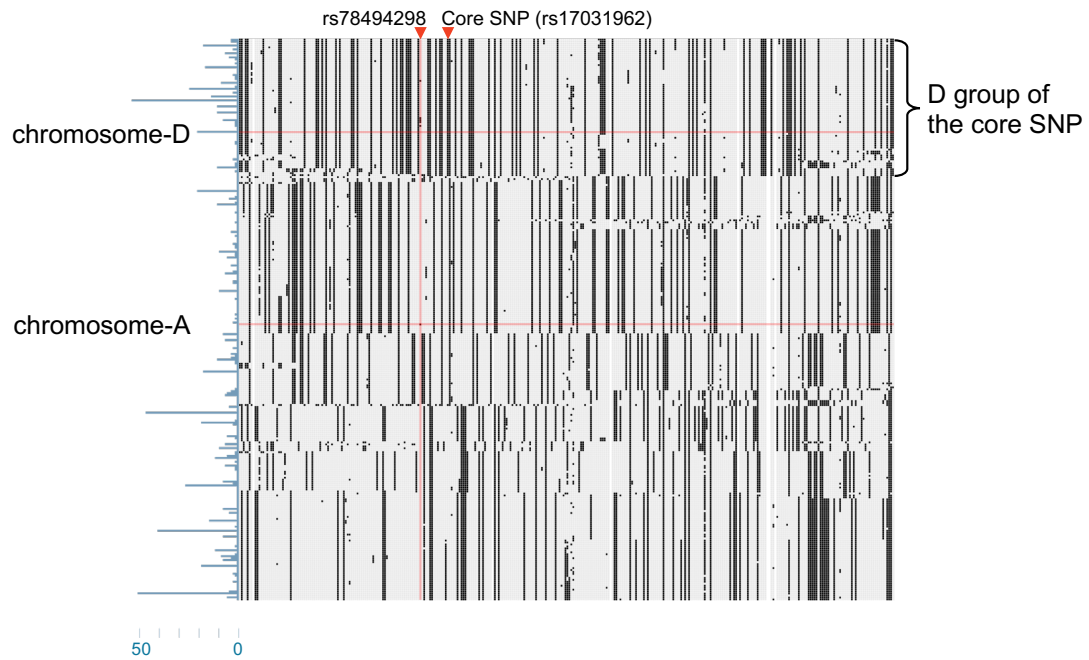
Figure S2 Haplotypes observed in EAS for the core region of rs17031962, using SNPs with minor allele frequency (MAF) in EAS ≥ 0.01. Columns represent genomic positions, and rows represent haplotypes. Haplotypes were sorted according to neighbor joining tree constructed by MEGA7 [1]. For each haplotype, a black or grey cell represents derived or ancestral allele, respectively, at the position. The positions of rs17031962 and rs78494298 are indicated by red triangles. The vertical red line indicates the position of rs78494298. The lengths of blue bars on the left side display the counts of each haplotype. "chromosome-D" indicates the one chromosome of HG00707 carrying the derived allele of the core SNP, and "chromosome-A" indicates the other chromosome of HG00707 carrying the ancestral allele of the core SNP. Horizontal red lines represent haplotypes containing each of the two chromosomes of HG00707.
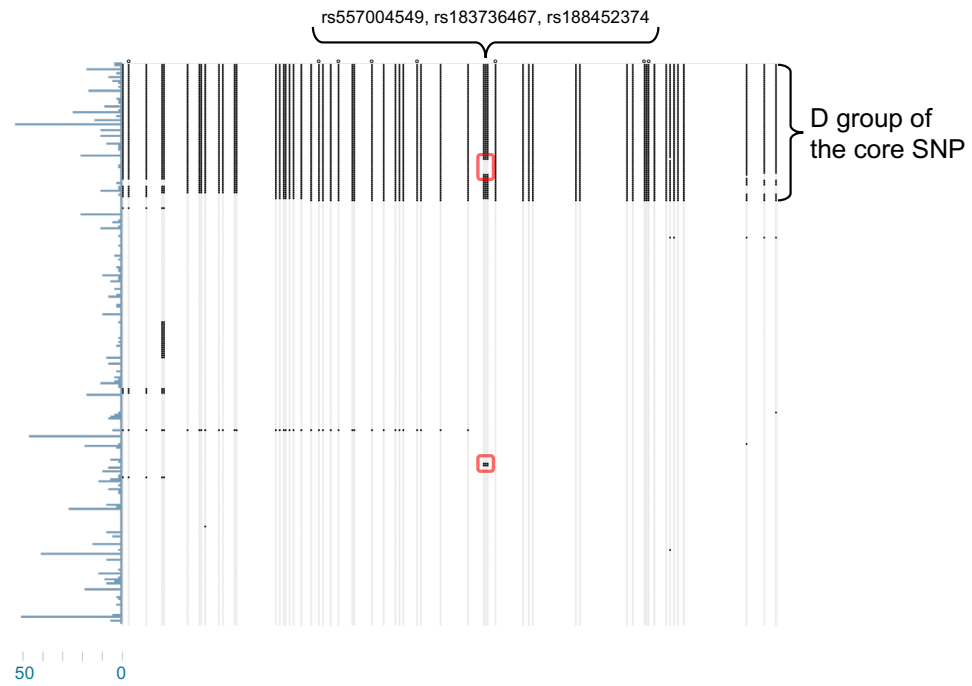
Figure S3 rs17031962 and its linked SNPs. SNPs with "o" above indicate linked SNPs in globally older frequency classes than rs17031962, which were ignored in the analysis. Three consecutive SNPs (rs557004549, rs183736467, rs188452374), which were excluded from the analysis, are noted. Different pattern from the other linked SNPs are marked with red rectangles: some of their derived alleles seemed to be linked to ancestral alleles of the other candidate SNPs, and vice versa.
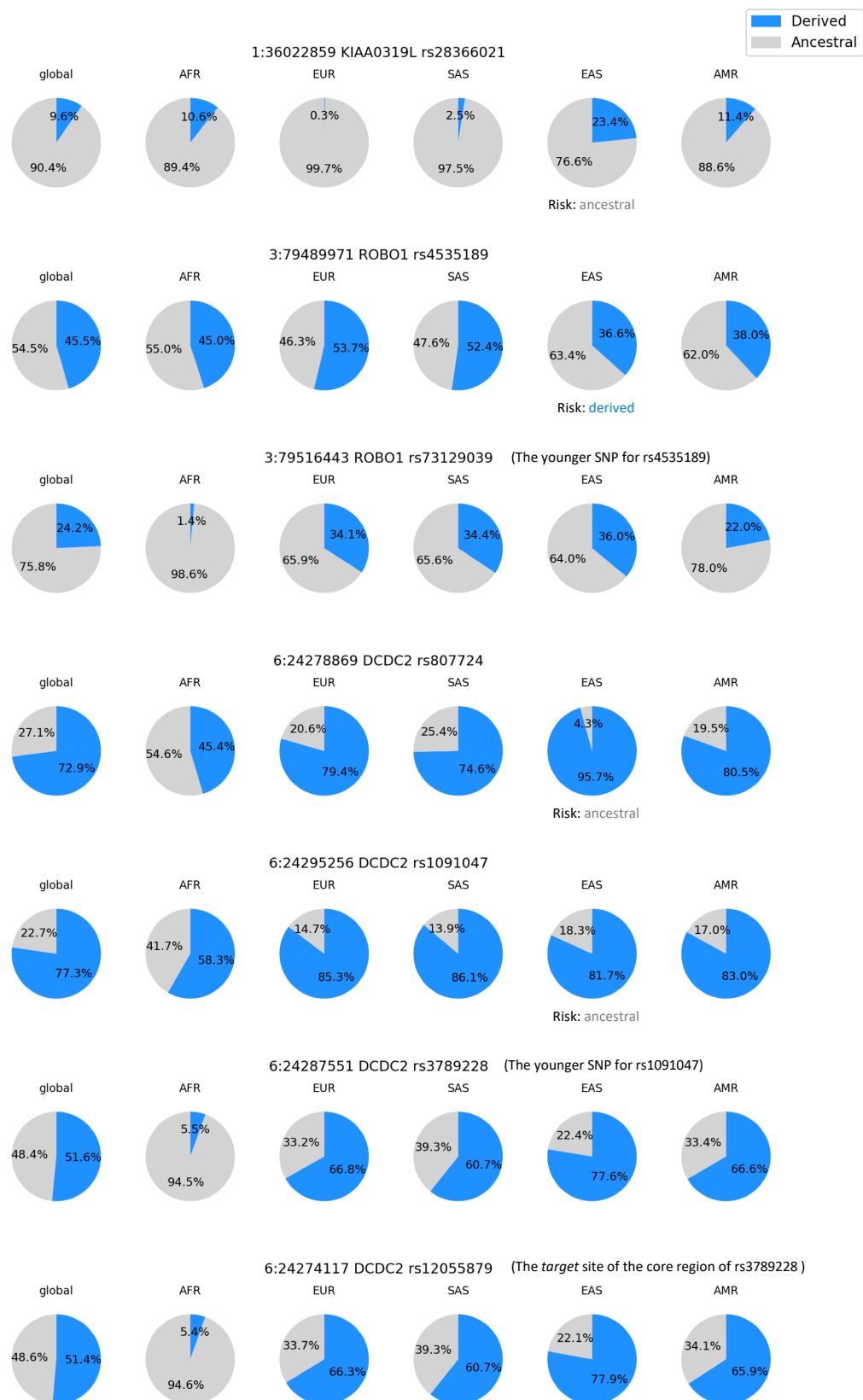
Figure S4 The derived allele frequencies of the 15 core SNPs in global, African (AFR), European (EUR), South Asian (SAS), East Asian (EAS), and American (AMR) populations of 1KG. Three younger SNPs and two target sites are also displayed after their core SNPs. For the 15 core SNPs, the risk alleles reported in the previous studies in Table 1 are noted below the pie charts of EAS.

6:24578272 KIAA0319 rs2760157

global: 35.8% / 64.2%
AFR: 44.7% / 55.3%
EUR: 19.7% / 80.3%
SAS: 18.2% / 81.8%
EAS: 54.4% / 45.6%
AMR: 39.9% / 60.1%
Risk: ancestral

6:24579867 KIAA0319 rs807507

global: 74.0% / 26.0%
AFR: 87.7% / 12.3%
EUR: 55.2% / 44.8%
SAS: 71.2% / 28.8%
EAS: 81.2% / 18.8%
AMR: 68.7% / 31.3%
Risk: derived

6:24588884 KIAA0319 rs4504469

global: 79.4% / 20.6%
AFR: 96.3% / 3.7%
EUR: 57.8% / 42.2%
SAS: 70.6% / 29.4%
EAS: 88.8% / 11.2%
AMR: 77.5% / 22.5%
Risk: derived

7:111487098 DOCK4 rs2074130

global: 96.9% / 3.1%
AFR: 100.0% / 0.0%
EUR: 99.9% / 0.1%
SAS: 99.8% / 0.2%
EAS: 89.9% / 10.1%
AMR: 93.1% / 6.9%
Risk: derived

11:113289182 DRD2 rs1079727

global: 77.3% / 22.7%
AFR: 92.0% / 8.0%
EUR: 85.3% / 14.7%
SAS: 71.4% / 28.6%
EAS: 58.4% / 41.6%
AMR: 73.6% / 26.4%
Risk: derived

12:102146558 GNPTAB rs17031962

global: 90.6% / 9.4%
AFR: 99.9% / 0.1%
EUR: 99.2% / 0.8%
SAS: 83.3% / 16.7%
EAS: 70.6% / 29.4%
AMR: 99.3% / 0.7%
Risk: ancestral

12:102131564 GNPTAB rs3751248 (The *target* site of the core region of rs17031962)

global: 90.4% / 9.6%
AFR: 99.1% / 0.9%
EUR: 99.2% / 0.8%
SAS: 83.3% / 16.7%
EAS: 70.7% / 29.3%
AMR: 99.3% / 0.7%

Figure S4 (Continued)

15:55777638 DYX1C1 rs11629841

global — 17.0% / 83.0%
AFR — 9.2% / 90.8%
EUR — 35.4% / 64.6%
SAS — 18.7% / 81.3%
EAS — 5.8% / 94.2%
AMR — 19.0% / 81.0%
Risk: derived

15:55790530 DYX1C1 rs3743205

global — 10.3% / 89.7%
AFR — 24.1% / 75.9%
EUR — 5.9% / 94.1%
SAS — 6.9% / 93.1%
EAS — 3.5% / 96.5%
AMR — 5.5% / 94.5%
Risk: derived

15:55679447 DYX1C1 rs79024225 (The younger SNP for rs3743205)

global — 0.6% / 99.4%
AFR — 0.0% / 100.0%
EUR — 0.0% / 100.0%
SAS — 0.0% / 100.0%
EAS — 3.1% / 96.9%
AMR — 0.0% / 100.0%

16:3980445 intergenic region rs8049367

global — 56.6% / 43.4%
AFR — 69.7% / 30.3%
EUR — 63.3% / 36.7%
SAS — 46.8% / 53.2%
EAS — 33.9% / 66.1%
AMR — 68.6% / 31.4%
Risk: derived

16:5092118 NAGPA rs882294

global — 25.6% / 74.4%
AFR — 23.9% / 76.1%
EUR — 34.5% / 65.5%
SAS — 22.1% / 77.9%
EAS — 18.9% / 81.1%
AMR — 30.4% / 69.6%
Risk: derived

21:47971539 DIP2A rs2255526

global — 25.7% / 74.3%
AFR — 12.5% / 87.5%
EUR — 29.2% / 70.8%
SAS — 32.3% / 67.7%
EAS — 26.4% / 73.6%
AMR — 35.2% / 64.8%
Risk: derived
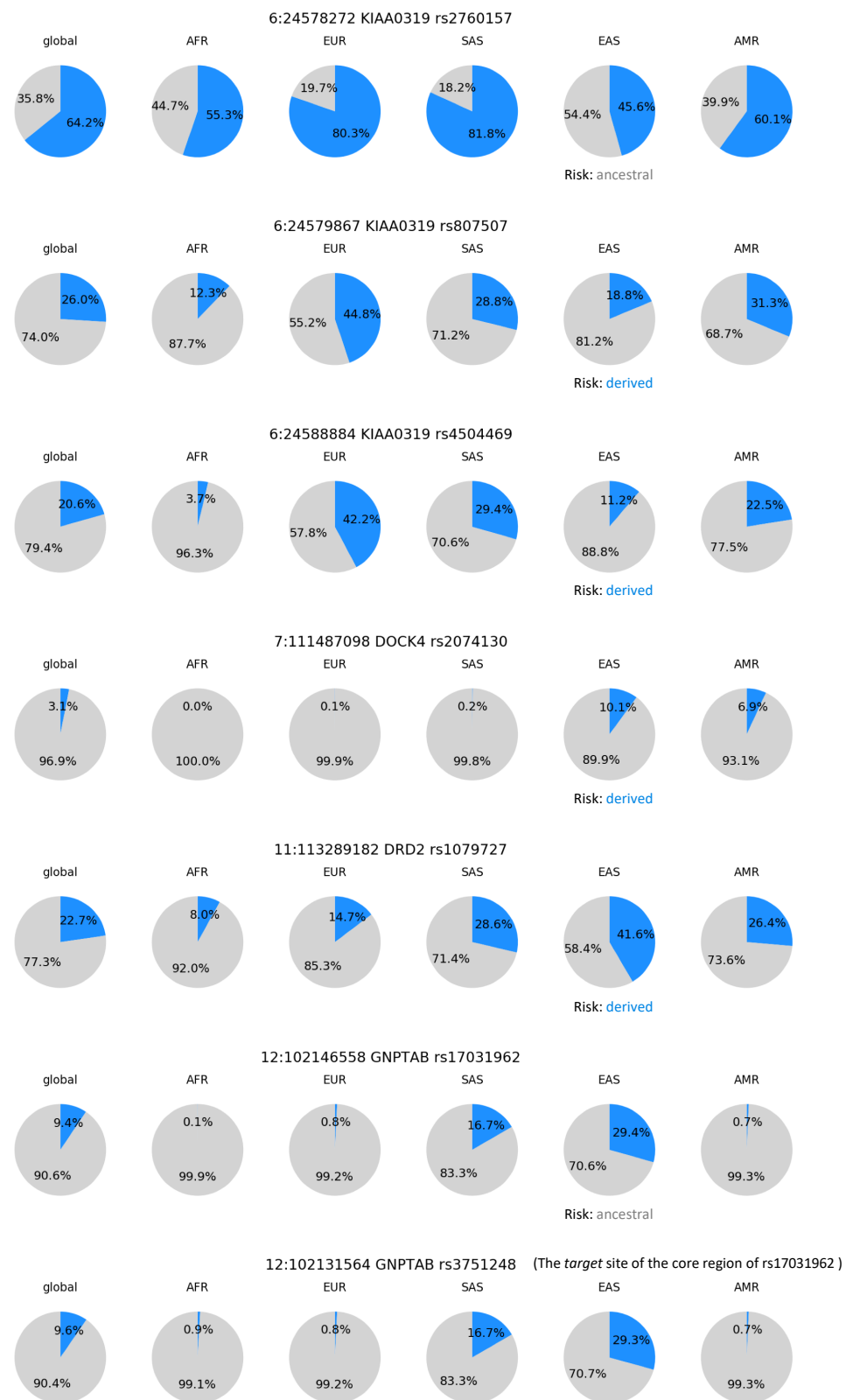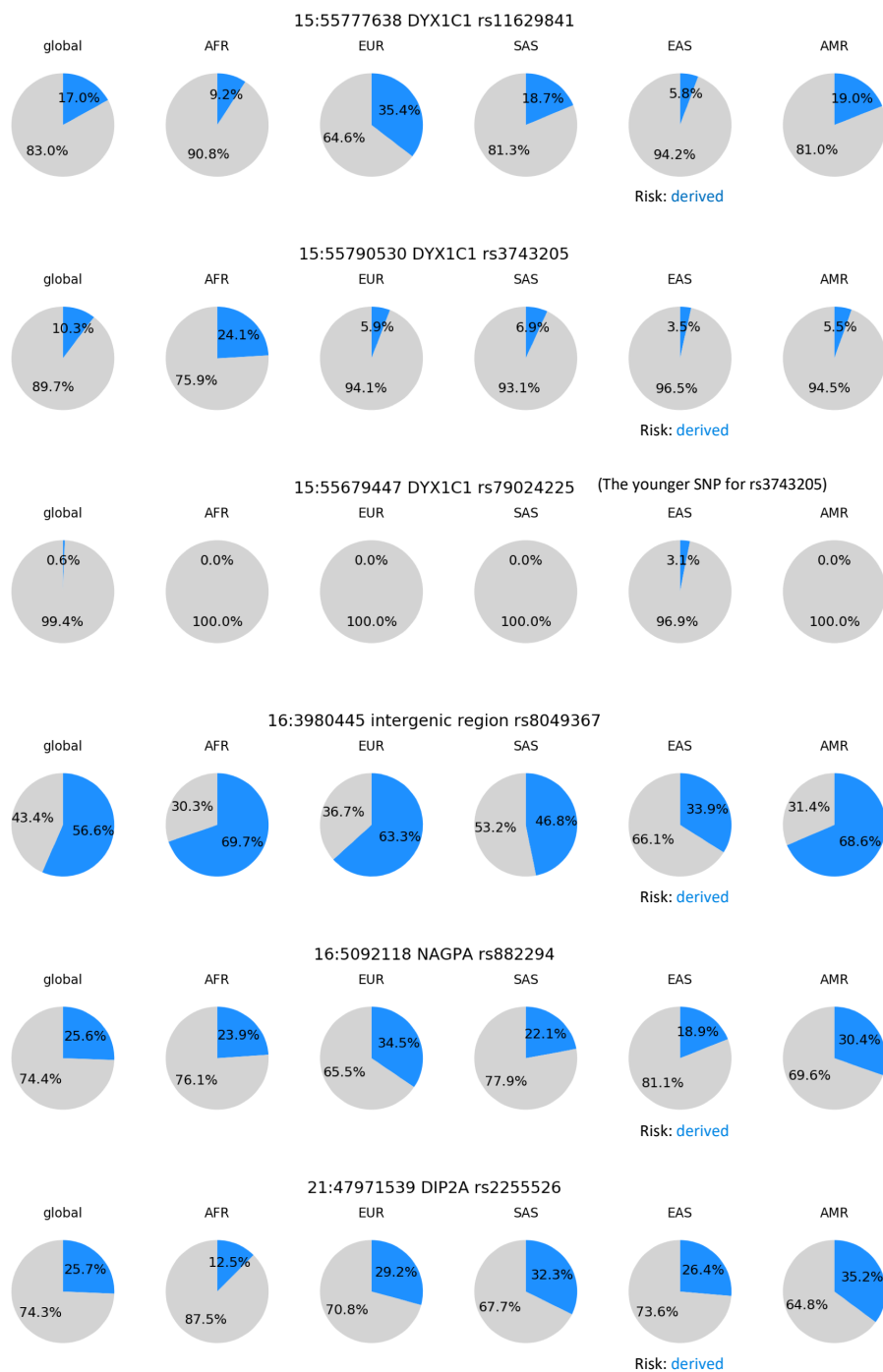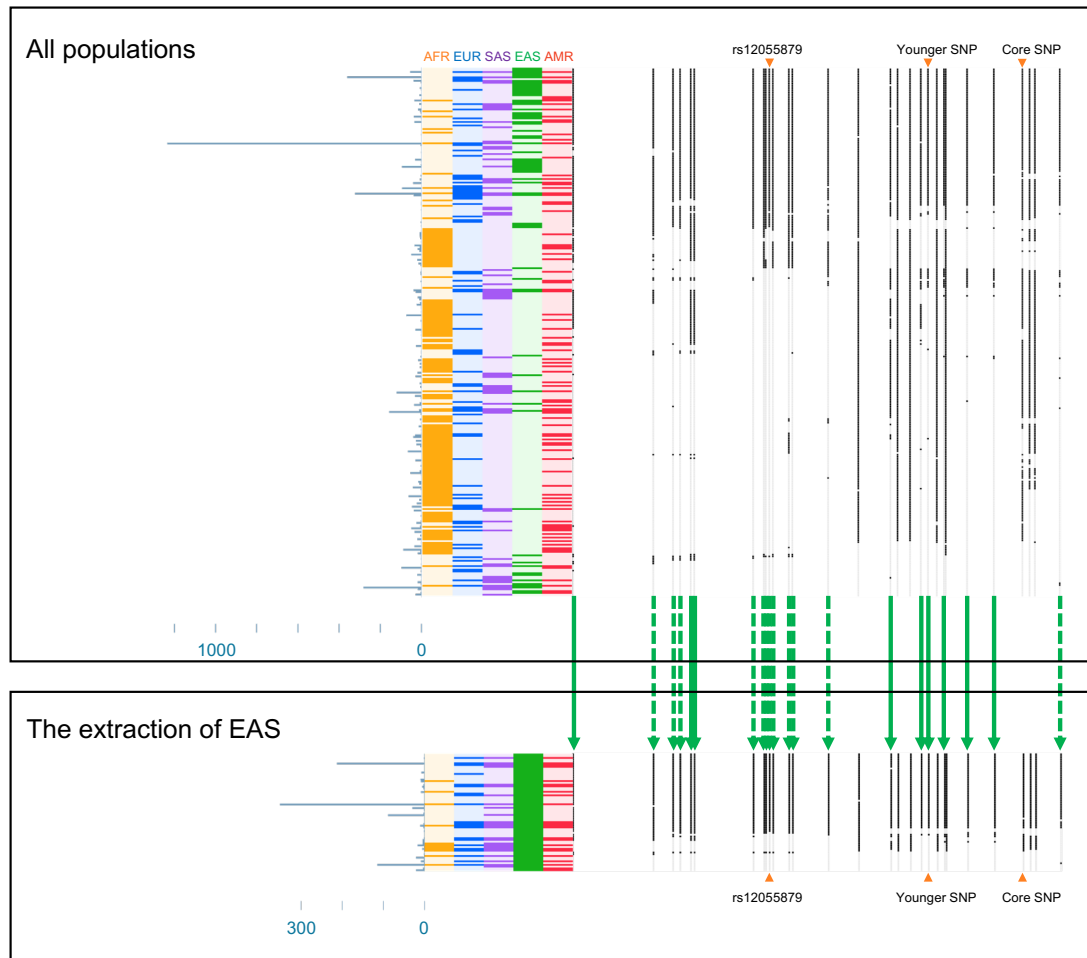
Figure S4 (Continued)

Figure S5 Top: haplotypes observed in all populations of 1KG for the core region of rs3789228, defined by SNPs with global MAF ≥ 0.005. Columns represent genomic positions, and rows represent haplotypes. Haplotypes were sorted according to neighbor joining tree constructed by MEGA7 [1]. For each haplotype, a black or grey cell represents derived or ancestral allele, respectively, at the position. Only rs3789228 and rs1091047 and their linked SNPs are indicated. Bottom: haplotypes observed in EAS. SNPs with solid green arrows indicate SNPs linked to both rs3789228 and rs1091047 in global frequency class 9; SNPs with broken green arrows indicate SNPs linked only to rs3789228 in global frequency class 9; SNPs with no arrows indicate linked SNPs in global frequency class 10. The five colored columns represent presence/absence of sequences from the five populations of 1KG for each haplotype. In the colored columns, dark tone indicates at least one sequence from the population is present in the haplotype while light color indicates no sequence from the population is observed in the haplotype. The blue bar lengths on the left side indicate the counts of each haplotype.

References

1.   Kumar, S.; Stecher, G.; Tamura, K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **2016**, *33*, 1870–1874.