1 **Structural basis of ethnic-specific variants of PAX4 associated with type 2** 2 **diabetes**

3

4 Jun Hosoe¹, Ken Suzuki^{1,2}, Takashi Kato¹, Yukinori Okada², Momoko Horikoshi³,
5 Nobuhiro Shojima^{*1}, Toshimasa Yamauchi^{*1}, Takashi Kadowaki^{*1,4,5}

6

7 1. Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The
8 University of Tokyo, Tokyo, Japan

9 2. Department of Statistical Genetics, Osaka University Graduate School of Medicine,
10 Osaka, Japan

11 3. Laboratory for Genomics of Diabetes and Metabolism, RIKEN Center for Integrative
12 Medical Sciences, Kanagawa, Japan

13 4. Department of Prevention of Diabetes and Lifestyle-Related Diseases, Graduate
14 School of Medicine, The University of Tokyo, Tokyo, Japan

15 5. Department of Metabolism and Nutrition, Teikyo University Mizonokuchi Hospital,
16 Kawasaki, Kanagawa, Japan

17 *Corresponding author:

18 Correspondence should be addressed to Nobuhiro Shojima, Toshimasa Yamauchi, and
19 Takashi Kadowaki.

20

21 **Abstract**

22 Recently, we conducted genome-wide association studies of type 2 diabetes (T2D) in
23 a Japanese population, which identified 20 novel T2D loci that were not associated
24 with T2D in Europeans. Moreover, nine novel missense risk variants, such as those of
25 *PAX4*, were not rare in the Japanese population, but rare in Europeans. We report *in*
26 *silico* structural analysis of ethnic-specific variants of *PAX4*, which suggests the
27 pathogenic effect of these variants.

29 **Main Text**

30 Large-scale meta-analyses of genome-wide association studies (GWAS) of type 2
31 diabetes (T2D) have been performed recently¹⁻⁴. In particular, Mahajan et al. reported
32 T2D GWAS meta-analysis data on approximately 900,000 individuals of European
33 ancestry, increasing the association signals in over 240 loci associated with T2D³.
34 Recently, we conducted an extensive T2D GWAS meta-analysis in a Japanese
35 population and identified 28 new loci⁴, highlighting the value of genetic research in
36 ethnically diverse populations. As an example of clinical differences between ethnic
37 populations, East Asians are more likely to develop T2D at a lower body mass index
38 than Europeans⁵, who have been reported to show a higher insulin response and lower
39 insulin sensitivity.

40 We demonstrated that the biological pathway of the maturity onset diabetes of the
41 young (MODY), monogenic diabetes characterized by reduced β -cell function, showed
42 the most significant association with T2D in both Japanese and European populations⁴,
43 by performing pathway analysis of GWAS summary data in the two populations^{1,4} using
44 Pascal software. Here we focused on protein-coding genes involved in the MODY

45 pathway mapping nearest to lead variants at T2D loci, as previously reported⁶, and
 46 investigated ethnic differences in the associations of these loci with T2D between
 47 Japanese and European populations^{3,4}. As suggested by a previous study⁴, most of
 48 these lead variants at T2D loci in either the Japanese or European population showed
 49 at least nominally significant ($P < 0.05$) associations in the alternative population, even if
 50 not genome-wide significant, except for variants that were rare or monomorphic in one
 51 population such as *NKX6-1* rs201597274, *HNF1A* rs187150787, and *PAX4* rs2233580
 52 in the Japanese population and *HNF1A* rs56348580 in the European population.
 53 Interestingly, there were several loci such as *GCK* where lead variants in these
 54 populations were independent of each other (**Supplementary Table 1 and 2**). As an
 55 example of a locus specific to East Asians, a previously unreported missense T2D
 56 variant of *PAX4* NM_006193:c.574C>A:p.(Arg192Ser) reached genome-wide
 57 significance⁴. This variant was located at the same amino acid as another established
 58 independent T2D variant NM_006193:c.575G>A:p.(Arg192His) (Table 2 in ref. ⁴). *PAX4*
 59 encodes a transcription factor that is important for β -cell development, and rare
 60 mutations of this gene are suggested to cause a subtype of MODY (OMIM #612225).
 61 Arg192 of *PAX4* is located in the homeodomain, which is a DNA-binding domain
 62 conserved in a large family of transcription factors. We conducted an *in silico* structural
 63 analysis of the pathogenic effect of these *PAX4* variants, which revealed a decrease of
 64 the *PAX4* homeodomain stability and reduced DNA binding by this domain in both
 65 variants (**Fig. 1 and 2, Supplementary Note**).

66 To physiologically classify T2D loci, cluster analyses have been performed using
 67 data on metabolic traits^{1,2,7}. Mahajan et al. performed a hierarchical cluster analysis of
 68 loci associated with T2D in ethnically diverse populations, generating three main

69 clusters associated with BMI/dyslipidemia, insulin secretion, and insulin action². The
70 cluster related to reduced insulin secretion contained a number of important loci
71 involved in β -cell function as shown in their Supplementary Fig. 6², with several lead
72 variants, such as those of *PAX4* (p.Arg192His) and *HNF1A*, that showed distinct minor
73 allele frequency (MAF) spectra and ethnically different associations (Supplementary
74 Table 10 and 11 in ref. ²). Such differences could lead to distinct proportions of the
75 clusters of T2D loci with different underlying biological pathways^{2,7}, which may result in
76 variable proportions of subgroups of T2D patients classified based on their genetics in
77 ethnically diverse populations. For example, the *PAX4* variants are specific to East
78 Asians⁴, which presumably increase the proportion of the cluster of T2D loci with
79 biological pathways influencing β -cell function in East Asian populations.

80 In a recent review article⁸, the pronounced ethnic heterogeneity associated with
81 ethnic-specific risk variants (e.g., *PAX4* p.Arg192His and *HNF1A* p.Glu508Lys⁹) was
82 suggested to be relatively unusual and not to explain observed ethnic differences in
83 presentation of T2D. It was also suggested that rare ethnic-specific variants identified
84 through sequencing studies may be more important in this regard⁸. Whereas T2D
85 GWAS meta-analysis recently performed in a Japanese population identified no less
86 than 20 novel T2D loci⁴ that were not associated with T2D in Europeans^{1,3}
87 (Supplementary Table 3 in ref. ⁴). Ethnic differences in the associations of a number of
88 new loci with T2D may be affected by several factors, such as differences in allele
89 frequency and diverse patterns of linkage disequilibrium in these populations as
90 suggested recently¹⁰. For example, we identified nine novel missense variants such as
91 those of *PAX4* (p.Arg192Ser) and *GLP1R* (p.Arg131Gln)⁴ that were in linkage
92 disequilibrium with the lead variants of T2D loci and not rare in Japanese population, but

93 rare or monomorphic in Europeans (Supplementary Table 6 in ref. ⁴).

94 Considering the differences in sample size between these studies^{3,4}, more
 95 large-scale genetic studies of East Asian populations are needed to fully elucidate
 96 similarities and differences in the pathogenesis of T2D between them. In conclusion,
 97 our new GWAS data⁴ suggested that large-scale single ethnic genetic studies could be
 98 useful for identifying ethnic-specific risk variants including relatively common ones,
 99 providing new insights into the genetic heterogeneity of T2D in diverse ethnic groups.
 100 The new information thus obtained would facilitate development of tailored therapy for
 101 this disease in various populations.

102

103 **Acknowledgements**

104 We acknowledge all authors of our GWAS meta-analysis of T2D in a Japanese
 105 population⁴ for their contribution to advancing genetic research. This study was
 106 supported by a grant-in-aid for scientific research from the Ministry of Education, Culture,
 107 Sports, Science and Technology of Japan (MEXT) (grant 19K16534 to J.H.). This
 108 research was also supported by Advanced Genome Research and Bioinformatics Study
 109 to Facilitate Medical Innovation (GRIFIN) in Platform Program for Promotion of Genome
 110 Medicine (P3GM) of AMED.

111

112 **Author contributions**

113 J.H., T.Kato, and N.S. designed the study, performed statistical analysis, and wrote the
 114 manuscript. J.H., K.S., T.Kato, and N.S. and contributed to data acquisition. Y.O., M.H.,
 115 N.S., T.Y., and T.Kadowaki supervised the study. All authors contributed to and
 116 approved the final version of the manuscript.

117

118 **Competing interests**

119 The authors declare no competing interests.

120

121 Supplementary information is available at Human Genome Variation's website.

122

123 **References**

- 124 1. Scott, R. A. et al. *Diabetes* **66**, 2888-2902 (2017).
- 125 2. Mahajan, A. et al. *Nat. Genet.* **50**, 559-571 (2018).
- 126 3. Mahajan, A. et al. *Nat. Genet.* **50**, 1505-1513 (2018).
- 127 4. Suzuki, K. et al. *Nat. Genet.* **51**, 379-386 (2019).
- 128 5. Ntuk, U. E., Gill, J. M., Mackay, D. F., Sattar, N., & Pell, J. P. *Diabetes Care* **37**,
- 129 2500-2507 (2014).
- 130 6. Morris, A. P. et al. *Nat. Genet.* **44**, 981-990 (2012).
- 131 7. Udler, M. S. et al. *PLoS Med.* **15**, e1002654 (2018).
- 132 8. Barroso, I., & McCarthy, M. I. *Cell* **177**, 146-161 (2019).
- 133 9. Estrada, K. et al. *JAMA* **311**, 2305-2314 (2014).
- 134 10. Sirugo, G., Williams, S. M., & Tishkoff, S. A. *Cell* **177**, 26-31 (2019).

135

136 **Figure legends**

137

138 **Fig. 1 Structural model of the homeodomain in human PAX4.** **a** The human PAX4
 139 homeodomain is represented as cylinders, with the Arg192 and Glu203 residues being
 140 shown as sticks. The PAX4 homeodomain consists of three helices: helix-1 (light-blue),

141 helix-2 (green), and helix-3 (yellow). Arg192 located on helix-2 is predicted to form
 142 salt-bridges with Glu203 located on helix-3. **b** Binding of the human PAX4
 143 homeodomain to double-stranded DNA. DNA is shown by ribbon representation in blue.
 144 Arg192 of PAX4 (sticks) is predicted to bind with the phosphate backbone (sticks) of
 145 DNA.

146

147 **Fig. 2 Structural analysis of two PAX4 variants, p.Arg192Ser and p.Arg192His.**

148 PAX4 are shown by ribbon representation, with the key residues of PAX4 as sticks. The
 149 helix-2 and the helix-3 of PAX4 are shown in green and yellow, respectively. Wild-type
 150 Arg192 is green, while the Ser192 and His192 mutants are light red. **a** In wild-type PAX4,
 151 Arg192 forms salt-bridges with Glu203 (dotted blue line). In both mutants (p.Arg192Ser
 152 and p.Arg192His), formation of salt-bridges with Glu203 is disrupted. **b** DNA is shown by
 153 ribbon representation in blue, with the phosphate backbone as sticks. In wild-type PAX4,
 154 Arg192 directly binds to the phosphate backbone of DNA (dotted blue line). In both
 155 mutants (p.Arg192Ser and p.Arg192His), binding with the phosphate backbone of DNA
 156 is disrupted.

a

b