

## Multi-ancestry polygenic risk scores for the prediction of type 2 diabetes and complications in diverse ancestries.

Alicia Huerta-Chagoya <sup>\*1,2,3,4</sup>, Joohyun Kim <sup>\*5</sup>, Ravi Mandla & <sup>1,2,3</sup>, Yingchang Lu & <sup>5</sup>, Ken Suzuki & <sup>6,7,8</sup>, Lauren E. Petty & <sup>5</sup>, Hong Kiat Ng & <sup>9</sup>, Jaewon Choi & <sup>10</sup>, Simon Lee & <sup>11</sup>, Madhusmita Rout & <sup>12</sup>, Kuang Lin & <sup>13</sup>, Linda S. Adair <sup>14</sup>, Adebowale Adeyemo <sup>15</sup>, Habibul Ahsan <sup>16</sup>, Masato Akiyama <sup>17,18</sup>, Ping An <sup>19</sup>, Sonia S. Anand <sup>20,21,22</sup>, Diane M. Becker <sup>23</sup>, Alain G. Bertoni <sup>24</sup>, Zheng Bian <sup>25</sup>, Lawrence F. Bielak <sup>26</sup>, John Blangero <sup>27</sup>, Michael Boehnke <sup>28</sup>, Erwin P. Bottinger <sup>29,30,11</sup>, Donald W. Bowden <sup>31,32,33</sup>, Fiona Bragg <sup>13,34</sup>, Jennifer A. Brody <sup>35</sup>, Thomas A. Buchanan <sup>36</sup>, Brian E. Cade <sup>37,38</sup>, Jin-Fang Chai <sup>39</sup>, John C. Chambers <sup>9</sup>, Giriraj R. Chandak <sup>40,41</sup>, Li-Ching Chang <sup>42</sup>, Kyong-Mi Chang <sup>43,44</sup>, Miao-Li Chee <sup>45</sup>, Chien-Hsiun Chen <sup>42</sup>, Yuan-Tsong Chen <sup>42</sup>, Zhengming Chen <sup>13</sup>, Yii-Der I. Chen <sup>46</sup>, Ji Chen <sup>47</sup>, Guanjie Chen <sup>15</sup>, Shyh-Huei Chen <sup>48</sup>, Wei-Min Chen <sup>49</sup>, Ching-Yu Cheng <sup>50,51,45,52</sup>, Yoon Shin Cho <sup>53</sup>, Hyeok Sun Choi <sup>53</sup>, Lee-Ming Chuang <sup>54,55</sup>, Miguel Cruz <sup>56</sup>, Mary Cushman <sup>57</sup>, Swapan K. Das <sup>58</sup>, Ralph A. DeFronzo <sup>59</sup>, H Janaka deSilva <sup>60</sup>, Latchezar Dimitrov <sup>32</sup>, Ayo P. Doumatey <sup>15</sup>, Shufa Du <sup>14</sup>, Qing Duan <sup>61</sup>, Ravindranath Duggirala <sup>27</sup>, Leslie S. Emery <sup>62</sup>, James C. Engert <sup>63</sup>, Daniel S. Evans <sup>64</sup>, Michele K. Evans <sup>65</sup>, Sarah Finer <sup>66</sup>, Jose C. Florez <sup>1,4,67</sup>, James S. Floyd <sup>35</sup>, Myriam Fornage <sup>68</sup>, Elizabeth G. Frankel <sup>5</sup>, Barry I. Freedman <sup>69</sup>, Lourdes García-García <sup>70</sup>, Pauline Genter <sup>71</sup>, Hertzfel C. Gerstein <sup>20,21,22</sup>, Mark O. Goodarzi <sup>72</sup>, Penny Gordon-Larsen <sup>14</sup>, Mariaelisa Graff <sup>73</sup>, Myron Gross <sup>74</sup>, Yu Guo <sup>25</sup>, Xiuqing Guo <sup>46</sup>, Yang Hai <sup>46</sup>, Craig L. Hanis <sup>75</sup>, MGeoffrey Hayes <sup>76,77,78</sup>, Momoko Horikoshi <sup>79</sup>, Annie-Green Howard <sup>80</sup>, Sarah Hsu <sup>1,3</sup>, Willa Hsueh <sup>81</sup>, Wei Huang <sup>82</sup>, Mengna Huang <sup>83,84</sup>, Yi-Jen Hung <sup>85,86</sup>, Mi Yeong Hwang <sup>87</sup>, Chii-Min Hwu <sup>88,89</sup>, Sahoko Ichihara <sup>90</sup>, Michiya Igase <sup>91</sup>, Eli Ipp <sup>71</sup>, Mohammad T. Islam <sup>92</sup>, Masato Isono <sup>93</sup>, Hye-Mi Jang <sup>87</sup>, Farzana Jasmine <sup>16</sup>, Jost B. Jonas <sup>94,95,96,97,98</sup>, Yoonjung Y. Joo <sup>99,78</sup>, Edmond Kabagambe <sup>100</sup>, Takashi Kadowaki <sup>7</sup>, Yoichiro Kamatani <sup>101,102</sup>, Fouad R. Kandeel <sup>103</sup>, Sharon L.R. Kardia <sup>26</sup>, Elizabeth W. Karlson <sup>104</sup>, Anuradhani Kasturiratne <sup>105</sup>, Norihiro Kato <sup>93</sup>, Tomohiro Katsuya <sup>106,107</sup>, Varinderpal Kaur <sup>1,2,3</sup>, Takahisa Kawaguchi <sup>108</sup>, Jacob M. Keaton <sup>32,100</sup>, Abel N. Kho <sup>109,110</sup>, Chiea-Chuen Khor <sup>111,112</sup>, Muhammad Kibriya <sup>16</sup>, Bong-Jo Kim <sup>87</sup>, Woon-Puay Koh <sup>113,39</sup>, Katsuhiko Kohara <sup>114</sup>, Jaspal S. Kooner <sup>115</sup>, Charles Kooperberg <sup>116</sup>, Raymond J. Kreienkamp <sup>1,4,2,117</sup>, Amel Lamri <sup>20,21</sup>, Leslie A. Lange <sup>118</sup>, Nanette R. Lee <sup>119</sup>, Myung-Shik Lee <sup>120,121</sup>, Jung-Jin Lee <sup>122</sup>, Donna M. Lehman <sup>59</sup>, Liming Li <sup>123</sup>, Yun Li <sup>124</sup>, Victor JY. Lim <sup>39</sup>, Jianjun Liu <sup>111,125</sup>, Yongmei Liu <sup>126,24</sup>, Simin Liu <sup>127,128,129,130</sup>, Jirong Long <sup>100</sup>, Tin Louie <sup>62</sup>, Xi Luo <sup>131</sup>, Jun Lv <sup>123</sup>, Julie A. Lynch <sup>132,133</sup>, Shiro Maeda <sup>79,134,135</sup>, Anubha Mahajan <sup>136</sup>, Nisa M. Maruthur <sup>137</sup>, Fumihiko Matsuda <sup>108</sup>, Mark I. McCarthy <sup>136,138,139</sup>, Roberta McKean-Cowdin <sup>140</sup>, James B. Meigs <sup>1,4,141</sup>, Iona Y. Millwood <sup>13</sup>, Karen L. Mohlke <sup>61</sup>, Ayesha A. Motala <sup>142</sup>, Girish N. Nadkarni <sup>11,29</sup>, Jerry L. Nadler <sup>143</sup>, Masahiro Nakatochi <sup>144</sup>, Mike A. Nalls <sup>145,146,147</sup>, Uma Nayak <sup>148</sup>, Aude Nicolas <sup>145</sup>, Kari E. North <sup>73</sup>, Darryl Nossome <sup>140</sup>, Yukinori Okada <sup>149,150,151</sup>, Ian Pan <sup>84</sup>, James S. Pankow <sup>152</sup>, Guillaume Paré <sup>153,21,22</sup>, Jaehyun Park <sup>5</sup>, Kyong Soo Park <sup>154,155</sup>, Esteban J. Parra <sup>156</sup>, Sanjay R. Patel <sup>157</sup>, Mark A. Pereira <sup>152</sup>, Patricia A. Peyser <sup>26</sup>, Fraser J. Pirie <sup>142</sup>, Michael Preuss <sup>11</sup>, Michael A. Province <sup>19</sup>, Bruce M. Psaty <sup>158</sup>, Leslie J. Raffel <sup>159</sup>, Laura M. Raffield <sup>61</sup>, Laura J. Rasmussen-Torvik <sup>160</sup>, Susan Redline <sup>161,162,38</sup>, Alexander P. Reiner <sup>163,116</sup>, Stephen S. Rich <sup>164</sup>, Rebecca Rohde <sup>73</sup>, Kathryn Roll <sup>46</sup>, Rashedeh Roshani <sup>5</sup>, Charles N. Rotimi <sup>15</sup>, Charumathi Sabanayagam <sup>45,50</sup>, Danish Saleheen <sup>43</sup>, Kevin Sandow <sup>46</sup>, Claudia Schurmann <sup>29,30,11</sup>, Mohammad Shahriar <sup>16</sup>, Douglas M. Shaw <sup>5</sup>, Wayne H-H. Sheu <sup>86,88,89,165,166</sup>, Jinxiu Shi <sup>82</sup>, Xiao-Ou Shu <sup>100</sup>, Megan M. Shuey <sup>5</sup>, Moneeza K. Siddiqui <sup>66</sup>, Jennifer A. Smith <sup>26,167</sup>, Tamar Sofer <sup>168,169,4</sup>, Cassandra N. Spracklen <sup>61,170</sup>, Adrienne M. Stilp <sup>62</sup>, Meng Sun <sup>171</sup>, Yasuharu Tabara <sup>108</sup>, E-Shyong Tai <sup>125,39,112</sup>, Salman M. Tajuddin <sup>65</sup>, Atsushi Takahashi <sup>101,172</sup>, Fumihiko Takeuchi <sup>93</sup>, Jingyi Tan <sup>46</sup>, Kent D. Taylor <sup>46</sup>, Katherine Taylor <sup>1</sup>, Farook Thameem <sup>173</sup>, Lin Tong <sup>16</sup>, Fuu-Jen Tsai <sup>174</sup>, Philip S. Tsao <sup>175,176</sup>, Miriam S. Udler <sup>1,4,2,3</sup>, Adan Valladares-Salgado <sup>56</sup>, David A. van Heel <sup>177</sup>, Rob M. van Dam <sup>39,125</sup>, Rohit Varma <sup>178</sup>, Maheak Vora <sup>1,2</sup>, Niels Wachter-Rodarte <sup>179</sup>, Ya-Xing Wang <sup>180</sup>, Ellie Wheeler <sup>181,47</sup>, Eric A. Whitel <sup>73,182</sup>, Ananda R. Wickremasinghe <sup>105</sup>, Genevieve L. Wojcik <sup>183</sup>, Tien Y. Wong <sup>45,50,51</sup>, Jer-Yuarn Wu <sup>42</sup>, Yong-Bing Xiang <sup>184,185</sup>, Anny H. Xiang <sup>186</sup>, Chittaranjan S. Yajnik <sup>187</sup>, Ken Yamamoto <sup>188</sup>, Toshimasa Yamauchi <sup>7</sup>, Lisa R. Yanek <sup>23</sup>, Jie Yao <sup>46</sup>, Mitsuhiro Yokota <sup>188</sup>, Canqing Yu <sup>123</sup>, Jian-Min Yuan <sup>189,190</sup>, Salim Yusuf <sup>191,192,193</sup>, Eleftheria Zeggini <sup>194,195</sup>, Liang Zhang <sup>45</sup>, Weihua Zhang <sup>196</sup>, Wei Zheng <sup>197</sup>, Alan B. Zonderman <sup>65</sup>, ENSA Genomics Consortium <sup>198</sup>, Genes & Health Research Team <sup>177</sup>, VA Million Veteran Program <sup>199</sup>, Carlos A. Aguilar-Salinas <sup>200,201</sup>, Clicerio González-Villalpando <sup>202</sup>, Christopher A. Haiman <sup>203</sup>, Young Jin Kim <sup>87</sup>, Soo Heon Kwak <sup>154</sup>, Aaron Leong <sup>1,2,204,3,141,205</sup>, Ruth J.F. Loos <sup>11,206</sup>, Andres Moreno-Estrada <sup>207</sup>, Andrew P. Morris <sup>8,208</sup>, Lorena Orozco <sup>209</sup>, Jerome I. Rotter <sup>46</sup>, Dharambir Sanghera <sup>12</sup>, Teresa Tusie-Luna <sup>210</sup>, Benjamin F. Voight <sup>211,212,213,214</sup>, Marijana Vujkovic <sup>215,211</sup>, Robin G. Walters <sup>13</sup>, Tian Ge <sup>216</sup>, Alisa K. Manning <sup>217,1,4</sup>, Marie Loh <sup>218,219</sup>, Jennifer E. Below <sup>219</sup>, Xueling Sim <sup>219</sup>, Josep M. Mercader <sup>220</sup>, Maggie C.Y. Ng <sup>221</sup>, D-PRISM Consortium <sup>221</sup>

<sup>1</sup>Programs in Metabolism and Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, 2142, USA, <sup>2</sup>Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>3</sup>Diabetes Unit, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>4</sup>Department of Medicine, Harvard Medical School,

Boston, MA, 2215, USA, <sup>5</sup>Vanderbilt Genetics Institute, Division of Genetic Medicine, Vanderbilt University Medical Center, Nashville, TN, USA, <sup>6</sup>Department of Statistical Genetics, Osaka University, Osaka, 565-0871, Japan, <sup>7</sup>Department of Diabetes and Metabolic Diseases, Graduate School of Medicine, The University of Tokyo, Tokyo, 113-8654, Japan, <sup>8</sup>Centre for Genetics and Genomics Versus Arthritis, Centre for Musculoskeletal Research, Division of Musculoskeletal and Dermatological Sciences, University of Manchester, Manchester, UK, <sup>9</sup>Lee Kong Chian School of Medicine, Clinical Sciences Building, Nanyang Technological University, Singapore, 308232, Singapore, <sup>10</sup>Department of Internal Medicine, Seoul National University Hospital, Seoul, 3080, South Korea, <sup>11</sup>The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>12</sup>Department of Pediatrics, Section of Genetics, University of Oklahoma, Oklahoma, OK, USA, <sup>13</sup>Nuffield Department of Population Health, University of Oxford, Oxford, OX3 7LF, UK, <sup>14</sup>Department of Nutrition, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, <sup>15</sup>Center for Research on Genomics and Global Health, National Human Genome Institute, National Institutes of Health, Bethesda, MD, USA, <sup>16</sup>Biological Sciences Division, Public Health Sciences, The University of Chicago, Chicago, IL, 60637, USA, <sup>17</sup>Laboratory for Statistical Analysis, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, <sup>18</sup>Department of Ophthalmology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, 812-8582, Japan, <sup>19</sup>Division of Statistical Genomics, Washington University School of Medicine, St. Louis, MO, USA, <sup>20</sup>Department of Medicine, McMaster University, Hamilton, ON, Canada, <sup>21</sup>Population Health Research Institute, Hamilton, ON, Canada, <sup>22</sup>Hamilton Health Sciences, Hamilton, ON, Canada, <sup>23</sup>Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, 21205, USA, <sup>24</sup>Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>25</sup>Chinese Academy of Medical Sciences, Beijing, 100730, China, <sup>26</sup>Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA, <sup>27</sup>Human Genetics and STDOJ, University of Texas Rio Grande Valley, Edinburg, TX, USA, <sup>28</sup>Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA, <sup>29</sup>Digital Health Center, Digital Engineering Faculty of Hasso Plattner Institute and University Potsdam, Potsdam, Germany, <sup>30</sup>Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>31</sup>Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>32</sup>Center for Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>33</sup>Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>34</sup>Health Data Research UK Oxford, University of Oxford, Oxford, UK, <sup>35</sup>Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA, <sup>36</sup>Department of Medicine, Division of Endocrinology and Diabetes, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA, <sup>37</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, <sup>38</sup>Harvard Medical School, Boston, MA, USA, <sup>39</sup>Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, 117549, Singapore, <sup>40</sup>Genomic Research on Complex diseases, CSIR-Centre for Cellular and Molecular Biology (CSIR-CCMB), Hyderabad, India, <sup>41</sup>Lightening Lives, Hyderabad, India, <sup>42</sup>Institute of Biomedical Sciences, Academia Sinica, Taipei, 11529, Taiwan, <sup>43</sup>Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA, 19104, USA, <sup>44</sup>Division of Translational Medicine and Human Genetics, University of Pennsylvania, Philadelphia, PA, 19104, USA, <sup>45</sup>Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, 168751, Singapore, <sup>46</sup>The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA, <sup>47</sup>Wellcome Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK, <sup>48</sup>Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest University Health Sciences, Winston-Salem, NC, USA, <sup>49</sup>Department of Public Health Sciences and Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA, <sup>50</sup>Ophthalmology & Visual Sciences Academic Clinical Program (Eye ACP), Duke-NUS Medical School, Singapore, 169857, Singapore, <sup>51</sup>Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, 119228, Singapore, <sup>52</sup>Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, 119228, Singapore, <sup>53</sup>Biomedical Science, Hallym University, Chuncheon, 24252, South Korea, <sup>54</sup>Department of Internal Medicine, Division of Endocrinology and Metabolism, National Taiwan University Hospital, Taipei, Taiwan, <sup>55</sup>College of Medicine, National Taiwan University, Taipei, Taiwan, <sup>56</sup>Unidad de Investigación Médica en Bioquímica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, IMSS, Mexico City, Mexico, <sup>57</sup>Department of Medicine, University of Vermont, Colchester, VT, USA, <sup>58</sup>Department of Internal Medicine, Section on Endocrinology and Metabolism, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>59</sup>Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA, <sup>60</sup>Department of Medicine, Faculty of Medicine, University of Kelaniya, Sri Lanka, <sup>61</sup>Department of Genetics,

University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, <sup>62</sup>Department of Biostatistics, University of Washington, Seattle, WA, USA, <sup>63</sup>Department of Medicine, McGill University, Montreal, QC, Canada, <sup>64</sup>California Pacific Medical Center Research Institute, San Francisco, CA, USA, <sup>65</sup>Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA, <sup>66</sup>Wolfson Institute of Population Health, Queen Mary University of London, London, UK, <sup>67</sup>Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>68</sup>Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA, <sup>69</sup>Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA, <sup>70</sup>Instituto Nacional de Salud Pública, Cuernavaca, Mexico, <sup>71</sup>Division of Endocrinology & Metabolism, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, 90502, USA, <sup>72</sup>Department of Medicine, Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, CA, USA, <sup>73</sup>Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, <sup>74</sup>Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, MN, 55454, USA, <sup>75</sup>Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA, <sup>76</sup>Department of Anthropology, Northwestern University, Evanston, IL, USA, <sup>77</sup>Center for Genetic Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, <sup>78</sup>Division of Endocrinology, Metabolism, and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, <sup>79</sup>Laboratory for Endocrinology, Metabolism and Kidney Diseases, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, <sup>80</sup>Department of Biostatistics, Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA, <sup>81</sup>Department of Internal Medicine, Diabetes and Metabolism Research Center, Ohio State University Wexner Medical Center, Columbus, OH, USA, <sup>82</sup>Shanghai-MOST Key Laboratory of Health and Disease Genomics, Shanghai Institute for Biomedical and Pharmaceutical Technologies (SIBPT), Shanghai, 200032, China, <sup>83</sup>Center for Global Cardiometabolic Health, Brown University, Providence, RI, USA, <sup>84</sup>Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA, <sup>85</sup>Division of Endocrine and Metabolism, Tri-Service General Hospital Songshan Branch, Taipei, Taiwan, <sup>86</sup>School of Medicine, National Defense Medical Center, Taipei, Taiwan, <sup>87</sup>Department of Precision Medicine, Division of Genome Science, National Institute of Health, Chungcheongbuk-do, 28159, Republic of Korea, <sup>88</sup>Department of Medicine, Section of Endocrinology and Metabolism, Taipei Veterans General Hospital, Taipei, Taiwan, <sup>89</sup>School of Medicine, National Yang-Ming University, Taipei, Taiwan, <sup>90</sup>Department of Environmental and Preventive Medicine, Jichi Medical University School of Medicine, Shimotsuke, 329-0498, Japan, <sup>91</sup>Department of Anti-aging Medicine, Ehime University Graduate School of Medicine, Ehime, 791-0295, Japan, <sup>92</sup>U Chicago Research, Uttara, Dhaka, 1230, Bangladesh, <sup>93</sup>Department of Gene Diagnostics and Therapeutics, Research Institute, National Center for Global Health and Medicine, Tokyo, 162-8655, Japan, <sup>94</sup>Rothschild Foundation Hospital, Paris, France, <sup>95</sup>Singapore Eye Research Institute, Singapore National Eye Center, Singapore, <sup>96</sup>Privatpraxis Prof Jonas und Dr. Panda-Jonas, Heidelberg, Germany, <sup>97</sup>Beijing Visual Science and Translational Eye Research Institute (BERI), Beijing Tsinghua Changgung Hospital, Tsinghua Medicine, Tsinghua University, Beijing, China, <sup>98</sup>L V Prasad Eye Institute, Hyderabad, Telangana, India, <sup>99</sup>Department of Health and Biomedical Informatics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, <sup>100</sup>Department of Medicine, Division of Epidemiology, Vanderbilt University Medical Center, Nashville, TN, USA, <sup>101</sup>Laboratory for Statistical and Translational Genetics, RIKEN Centre for Integrative Medical Sciences, Yokohama, 230-0045, Japan, <sup>102</sup>Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, 108-8639, Japan, <sup>103</sup>Departments of Clinical Diabetes, Endocrinology & Metabolism and Translational Research and Cellular Therapeutics, City of Hope, Duarte, CA, USA, <sup>104</sup>Department of Medicine, Brigham and Women's Hospital, Boston, MA, 2215, USA, <sup>105</sup>Department of Public Health, Faculty of Medicine, University of Kelaniya, Sri Lanka, <sup>106</sup>Department of Clinical Gene Therapy, Osaka University, Osaka, 565-0871, Japan, <sup>107</sup>Department of Geriatric and General Medicine, Graduate School of Medicine, Osaka University, Osaka, 565-0871, Japan, <sup>108</sup>Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, 606-8507, Japan, <sup>109</sup>Center for Health Information Partnerships, Institute for Public Health and Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, <sup>110</sup>Department of Medicine, Division of General Internal Medicine and Geriatrics, Northwestern University Feinberg School of Medicine, Chicago, IL, USA, <sup>111</sup>Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, 138672, Singapore, <sup>112</sup>Duke-NUS Medical School, Singapore, 169857, Singapore, <sup>113</sup>Health Services and Systems Research, Duke-NUS Medical School, Singapore, 169857, Singapore, <sup>114</sup>Department of Regional Resource Management, Ehime University Faculty of Collaborative Regional Innovation, Ehime, 790-8577, Japan, <sup>115</sup>Department of Cardiology, Ealing Hospital, London North West Healthcare NHS Trust, Middlesex, UB1 3HW, UK, <sup>116</sup>Fred Hutchinson Cancer Research Center, Seattle, WA, USA, <sup>117</sup>Department of Pediatrics, Division of Endocrinology, Boston Children's Hospital, Boston, MA, USA, <sup>118</sup>Department of Medicine,

University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA, <sup>119</sup>Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, 6000, Philippines, <sup>120</sup>Severance Biomedical Science Institute and Department of Internal Medicine, Yonsei University College of Medicine, Seoul, 3722, South Korea, <sup>121</sup>Soonchunhyang Institute of Medi-bio Science, Cheonan, 31151, South Korea, <sup>122</sup>Translational Medicine and Human Genetics, Department of Medicine, University of Pennsylvania, Philadelphia, PA, USA, <sup>123</sup>Department of Epidemiology and Biostatistics, Peking University Health Science Centre, Beijing, 100191, China, <sup>124</sup>Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, <sup>125</sup>Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, 117549, Singapore, <sup>126</sup>Department of Medicine, Division of Cardiology, Duke University School of Medicine, Durham, NC, USA, <sup>127</sup>Department of Epidemiology and Biostatistics, The Joe Wen School of Population & Public Health, UCI, Irvine, CA, USA, <sup>128</sup>Center for Global Cardiometabolic Health and Nutrition, the University of California, Irvine (UCI), Irvine, CA, USA, <sup>129</sup>Division of Cardiology, Department of Medicine, School of Medicine, UCI, USA, <sup>130</sup>Center for Global Cardiometabolic Health, Brown University School of Public Health, Providence, RI, USA, <sup>131</sup>Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Houston, TX, USA, <sup>132</sup>Salt Lake City VA Medical Center, Salt Lake City, UT, USA, <sup>133</sup>University of Utah, School of Medicine, Salt Lake City, UT, USA, <sup>134</sup>Department of Advanced Genomic and Laboratory Medicine, Graduate School of Medicine, University of the Ryukyus, Okinawa, 903-0213, Japan, <sup>135</sup>Division of Clinical Laboratory and Blood Transfusion, University of the Ryukyus Hospital, Okinawa, 903-0213, Japan, <sup>136</sup>Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, OX3 7BN, UK, <sup>137</sup>Department of Medicine and Epidemiology, Johns Hopkins University, Baltimore, MD, 21205, USA, <sup>138</sup>Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK, <sup>139</sup>Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK, <sup>140</sup>Department of Preventative Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA, <sup>141</sup>Division of General Internal Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>142</sup>Department of Diabetes and Endocrinology, Nelson R Mandela School of Medicine, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa, <sup>143</sup>Department of Medicine and Endocrinology Division, University of California Davis School of Medicine, Sacramento, CA, USA, <sup>144</sup>Department of Integrated Health Sciences, Nagoya University Graduate School of Medicine, Nagoya, 461-8673, Japan, <sup>145</sup>Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA, <sup>146</sup>Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA, <sup>147</sup>DataTecnica LLC, Washington, DC, USA, <sup>148</sup>Department of Genome Sciences, University of Virginia School of Medicine, Charlottesville, VA, USA, <sup>149</sup>Department of Statistical Genetics, Osaka University Graduate School of Medicine, Tokyo, 565-0871, Japan, <sup>150</sup>Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka University, Osaka, 565-0871, Japan, <sup>151</sup>Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan, <sup>152</sup>Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, 55454, USA, <sup>153</sup>Department of Pathology and Molecular Medicine, McMaster University, Hamilton, ON, Canada, <sup>154</sup>Department of Internal Medicine, Seoul National University College of Medicine and Seoul National University Hospital, Seoul, 3080, South Korea, <sup>155</sup>Department of Internal Medicine, Konkuk University Medical Center, Seoul, 3080, South Korea, <sup>156</sup>Department of Anthropology, University of Toronto at Mississauga, Mississauga, Canada, <sup>157</sup>Department of Medicine, Division of Pulmonary, Allergy, Critical Care, and Sleep Medicine, University of Pittsburgh, Pittsburgh, PA, USA, <sup>158</sup>Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Systems and Population Health, University of Washington, Seattle, WA, USA, <sup>159</sup>Department of Pediatrics, Division of Genetic and Genomic Medicine, UCI Irvine School of Medicine, Irvine, CA, USA, <sup>160</sup>Department of Preventive Medicine, Northwestern University, Chicago, IL, USA, <sup>161</sup>Division of Pulmonary, Critical Care, and Sleep Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA, <sup>162</sup>Division of Sleep and Circadian Disorders, Brigham and Women's Hospital, Boston, MA, USA, <sup>163</sup>Department of Epidemiology, University of Washington, Seattle, WA, USA, <sup>164</sup>Department of Genome Sciences, University of Virginia, Charlottesville, VA, USA, <sup>165</sup>Department of Medicine, Division of Endocrinology and Metabolism, Taichung Veterans General Hospital, Taichung, Taiwan, <sup>166</sup>Institutes of Molecular and Genomic Medicine, National Health Research Institutes, Taiwan, <sup>167</sup>Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA, <sup>168</sup>Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA, <sup>169</sup>Division of Cardiovascular Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA, <sup>170</sup>Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, 1003, USA, <sup>171</sup>Nuffield Department of Surgical Sciences, University of Oxford, Oxford, OX3 9DU, UK, <sup>172</sup>Department of Genomic Medicine, National Cerebral and Cardiovascular Center, Osaka, 565-8565, Japan, <sup>173</sup>Department of Biochemistry, Faculty of Medicine, Health Science Center, Kuwait University, Safat, Kuwait, <sup>174</sup>Department of Medical Genetics and Medical Research, China Medical

University Hospital, Taichung, 40402, Taiwan, <sup>175</sup>Department of Medicine, Stanford University, Palo Alto, CA, USA, <sup>176</sup>VA Palo Alto Health Care System, Palo Alto, CA, USA, <sup>177</sup>Blizard Institute, Queen Mary University of London, London, UK, <sup>178</sup>Southern California Eye Institute, CHA Hollywood Presbyterian Medical Center, Los Angeles, CA, USA, <sup>179</sup>Unidad de Investigación Médica en Epidemiología Clínica, Hospital de Especialidades, Centro Médico Nacional Siglo XXI, IMSS, Mexico City, Mexico, <sup>180</sup>Beijing Institute of Ophthalmology, Ophthalmology and Visual Sciences Key Laboratory, Beijing Tongren Hospital, Capital Medical University, Beijing, 100005, China, <sup>181</sup>MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK, <sup>182</sup>Department of Medicine, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, <sup>183</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, 21205, USA, <sup>184</sup>State Key Laboratory of Oncogene and Related Genes, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200032, China, <sup>185</sup>Department of Epidemiology, Shanghai Cancer Institute, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, 200032, China, <sup>186</sup>Department of Research & Evaluation, Division of Biostatistics Research, Kaiser Permanente of Southern California, Pasadena, CA, USA, <sup>187</sup>Diabetes Unit, King Edward Memorial Hospital and Research Centre, Pune, India, <sup>188</sup>Department of Medical Biochemistry, Kurume University School of Medicine, Kurume, 830-0011, Japan, <sup>189</sup>Division of Cancer Control and Population Sciences, UPMC Hillman Cancer Center, University of Pittsburgh, Pittsburgh, PA, 15232, USA, <sup>190</sup>Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, 15232, USA, <sup>191</sup>Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada, <sup>192</sup>Population Health Research Institute, Hamilton, ON, Canada, <sup>193</sup>Hamilton Health Sciences, Hamilton, ON, Canada, <sup>194</sup>Institute of Translational Genomics, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany, <sup>195</sup>Technical University of Munich (TUM), TUM University Hospital, TUM School of Medicine and Health, Munich, 81675, Germany, <sup>196</sup>Department of Epidemiology and Biostatistics, Imperial College London, London, W2 1PG, UK, <sup>197</sup>Division of Epidemiology, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, 37203, USA, <sup>198</sup>Instituto Nacional de Salud Pública, Cuernavaca, Mexico, <sup>199</sup>Corporal Michael J Crescenz VA Medical Center, Philadelphia, PA, USA, <sup>200</sup>Unidad de Investigación de Enfermedades Metabólicas, Research Direction of the Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, <sup>201</sup>Tecnológico de Monterrey, Escuela de Medicina y Ciencias de la Salud, Mexico City, Mexico, <sup>202</sup>Centro de Estudios en Diabetes, Unidad de Investigación en Diabetes y Riesgo Cardiovascular, Centro de Investigación en Salud Poblacional, Instituto Nacional de Salud Pública, Mexico City, Mexico, <sup>203</sup>Department of Population and Public Health Sciences, Keck School of Medicine of USC, Los Angeles, CA, USA, <sup>204</sup>Department of Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>205</sup>Endocrine Division, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>206</sup>The Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA, <sup>207</sup>Aging Research Center, Cinvestav Sede Sur, Center for Research and Advanced Studies of the National Polytechnic Institute, Mexico City, Mexico, <sup>208</sup>NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK, <sup>209</sup>Instituto Nacional de Medicina Genómica, Mexico City, Mexico, <sup>210</sup>Unidad de Biología Molecular y Medicina Genómica, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México/Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City, Mexico, <sup>211</sup>Corporal Michael J. Crescenz Philadelphia VA Medical Center, Philadelphia, PA, USA, <sup>212</sup>Department of Genetics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, <sup>213</sup>Department of Systems Pharmacology and Translational Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA, <sup>214</sup>Institute for Translational Medicine and Therapeutics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA, <sup>215</sup>Department of Medicine, University of Pennsylvania, Philadelphia, PA, 19104, USA, <sup>216</sup>Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>217</sup>Clinical and Translational Epidemiology Unit, Mongan Institute, Massachusetts General Hospital, Boston, MA, 2114, USA, <sup>218</sup>Department of Epidemiology and Biostatistics, Imperial College London, St Mary's Campus, London, UK, <sup>219</sup>Research Division, National Skin Centre, Singapore, <sup>220</sup>Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute of MIT and Harvard, Cambridge, MA, USA.

\* These authors equally contributed to this work.

& These authors equally contributed to this work.

^ These authors jointly directed this work.

# Corresponding authors

## Corresponding authors

Josep M. Mercader  
Programs in Metabolism and Medical and Population Genetics  
Broad Institute of Harvard and MIT  
75 Ames St, Cambridge, MA, 02142, United States of America  
E-mail: [mercader@broadinstitute.org](mailto:mercader@broadinstitute.org)

Maggie C.Y. Ng  
Division of Genetic Medicine, Vanderbilt Genetics Institute  
Vanderbilt University Medical Center  
Nashville, TN, 37232, United States of America  
E-mail: [maggie.ng@vumc.org](mailto:maggie.ng@vumc.org)

## Summary

### Background

Polygenic risk scores (PRSs) improve type 2 diabetes (T2D) prediction beyond clinical risk factors but perform poorly in non-European populations, where T2D burden is often higher, undermining their global clinical utility.

### Methods

We conducted the largest global effort to date to harmonize T2D genome-wide association study (GWAS) meta-analyses across five ancestries—European (EUR), African/African American (AFR), Admixed American (AMR), South Asian (SAS), and East Asian (EAS)—including 360,000 T2D cases and 1.8 million controls (41% non-EUR). We constructed ancestry-specific and multi-ancestry PRSs in training datasets including 11,000 T2D cases and 32,000 controls, and validated their performance in independent datasets including 39,000 T2D cases and 126,000 controls of diverse ancestries. In the All of Us Research Program, we compared these PRSs to those from the Polygenic Score Catalog and assessed their ability to predict diabetes micro- and macrovascular complications.

### Findings

Ancestry-specific PRSs showed limited prediction power for T2D in AFR, AMR, and SAS compared to EUR and EAS. In contrast, multi-ancestry PRSs, built using GWAS data from five ancestries, substantially improved T2D prediction across all ancestries. Compared to those in the interquartile range, individuals at the 97.5<sup>th</sup> percentile of their PRSs had a 6-fold increased T2D risk in AMR, EAS, and EUR, and  $\geq 3$ -fold in AFR and SAS. These PRSs were also associated with the development of microvascular complications and outperformed all previously reported PRSs for all ancestries.

### Interpretation

We developed and extensively validated the most up-to-date T2D PRSs across diverse ancestry groups. These PRSs are publicly available to support further evaluation of their clinical utility in diverse ancestries.

## Introduction

Type 2 diabetes (T2D) represents one of the largest health problems of the 21st century, affecting 537 million people globally, and is predicted to increase to 783 million by 2045.<sup>1</sup> Both genetic and environmental factors contribute to T2D susceptibility. Genome-wide association studies (GWAS) have identified 1,289 genetic signals associated with T2D in diverse ancestries.<sup>2,3</sup> The aggregation of risk alleles in polygenic risk scores (PRSs) for T2D<sup>2,4</sup> may provide insights into disease progression and prognosis, or help identify people at risk for prioritization of therapeutic or lifestyle intervention.<sup>5</sup> Several initiatives are beginning to test their utility in clinical settings<sup>6,7</sup>, and an observational study in Electronic Health Record (EHR) data has shown that T2D PRSs provide data orthogonal to standard clinical risk factors and are therefore particularly valuable in identifying at-risk individuals among those perceived to be low risk based on standard clinical risk factors.<sup>8</sup>

T2D disproportionately affects people of ancestral genetic backgrounds other than European, with South Asian, African American, and Hispanic/Latin American populations having a higher prevalence of diabetes and related complications.<sup>9</sup> Yet, most PRSs are based on GWAS, including predominantly individuals of European ancestry, and

have poor performance in individuals of other ancestries, which may further exacerbate health disparities if PRS were to be deployed for disease prediction in ancestrally diverse populations.<sup>2,10</sup>

While efforts have been made to enhance PRS performance in diverse populations<sup>11</sup>, a comprehensive, standardized, and harmonized approach for the development and validation of PRSs across continental ancestries for T2D remains lacking. In particular, recent efforts have aimed at improving the transferability of PRS across diverse ancestries. First, the Polygenic Risk Methods in Diverse Populations (PRIMED) consortium and others have developed new methods, such as those incorporating GWAS and linkage disequilibrium (LD) data from multiple ancestries, showing improved prediction.<sup>11,12</sup> Second, large-scale T2D GWAS data for populations other than European have become available, which may improve the power of PRS development in these populations. Third, the emergence of large-scale biobank data from around the world — such as the All of Us Research Program (AoU) — offers valuable opportunities to develop, train, and validate novel PRSs, as data from these participants has not yet been included in GWAS meta-analyses.<sup>13,14</sup>

As part of the Type 2 Diabetes Global Genomics Initiative (T2DGGI), we published the largest multi-ancestry T2D GWAS meta-analysis to date, based on 2.5 million individuals, including 428,452 with T2D, and enumerating 1,289 signals for T2D risk.<sup>3</sup> Here, in collaboration with T2DGGI, we present results from the Diabetes Polygenic Risk Scores in Multiple Ancestries (D-PRISM)<sup>2,11</sup>, an international consortium focusing on improving PRS prediction of different types of diabetes and progression across the lifespan in diverse ancestries. We leveraged these two consortia to aggregate extensive T2D GWAS data from five major continental ancestry groups, trained and validated PRS models in independent cohorts to evaluate their performance in predicting T2D and diabetes complications using a unified pipeline. We tested the optimal PRS model for each of the five ancestries and made them available to the community for downstream analyses, as well as future testing for clinical utility and implementation.

## Methods

An overview of the overall strategy is shown in Fig.1 and detailed in the supplementary methods. We leveraged T2D GWAS from selected cohorts participating in three large consortia: the Diabetes Meta-analysis of Trans-ethnic Association Studies (DIAMANTE)<sup>15</sup>, the Million Veteran Program (MVP)<sup>16</sup>, and FinnGen<sup>17</sup>. Cohorts were categorized by genetic similarity to one or more of the five ancestries available in the 1000 Genomes (1KG) project<sup>18</sup>, or the country of origin: African/African American (AFR), Admixed American (AMR), East Asian (EAS), European (EUR), and South Asian (SAS). While ancestry labels do not represent or account for the full continuum of human genetic diversity, they were necessary for statistical analyses where no individual-level data were available. We included 2,185,548 individuals (359,819 cases, 1,825,729 controls) across 125 T2D GWAS datasets to conduct ancestry-specific meta-analyses, with summary statistics from EUR and EAS ancestry groups representing 86% of the total sample size (EUR=68%, EAS=18%, AFR=6%, AMR=4%, and SAS=4%; Fig.1a, Supplementary Tables 1,2).

We leveraged cohorts not included in previous meta-analyses and, when necessary, conducted held-out meta-analyses to allow some cohorts to train and validate PRS. We aggregated data from a total of 42,784 individuals (10,992 cases, 31,792 controls) from five ancestries for training (tuning PRS construction parameters; Fig.1b) and 164,972 individuals (39,148 cases, 125,824 controls) for validation (testing PRS performance; Fig.1c). We used the five ancestry-specific T2D GWAS summary statistics to construct single-ancestry and multi-ancestry PRSs, using the PRS-CS (Continuous Shrinkage)<sup>19</sup> and PRS-CSx<sup>20</sup> methods, respectively (Fig.1b). Both methods use LD panels from reference datasets to model the pairwise correlations between SNPs in GWAS summary statistics during PRS construction. The SNPs in the most widely used LD panels are based on the HapMap3 (HM3), which generally do not tag well in non-EUR populations and may miss ancestry-specific signals. In addition, the samples used to model LD structure are either from the 1KG project or the UK Biobank (UKBB). The former has a limited sample size, while the latter lacks sufficient representation of populations other than the EUR. To improve tagging, we developed new LD panels with an expanded set of SNPs generated using the Tag(ging) It(erative) of SNPs in multiple populations (TagIt) program<sup>21</sup> and variants with ancestry-specific minor allele frequency (MAF)  $\geq 0.01$  across samples from the 1KG. We also use >8,000 in-house samples to compute pairwise LD in each ancestry, enabling more accurate modeling of the LD structure (Fig.1a).

We compared the impact of several input parameters (GWAS ancestry, two SNP sets, and two LD sources) on PRS performance in the training cohorts. We defined the best models as maximizing the incremental AUC (iAUC) for predicting prevalent T2D, comparing a full model, including the PRS, sex, age, and genetic principal components

(PCs), to a model without the PRS (Fig.1d). We validated the best-performing models for association with T2D in at least four independent cohorts from each ancestry group (Fig.1c) and fitted secondary models with body mass index (BMI) as an additional covariate. We estimated the effect size of the PRSs as the odds ratio per standard deviation (OR per SD) unit of the PRS, and calculated OR for individuals at the 90<sup>th</sup>, 95<sup>th</sup>, and 97·5<sup>th</sup> PRS percentiles compared with the interquartile range as the reference. In the AoU cohort, we also compared the association of our best-performing T2D PRSs with other published T2D PRSs from the Polygenic Score (PGS) catalog, and tested their association with common diabetes complications (Fig.1d).

## Results

### Development of new LD reference panels to improve variant coverage and LD estimation in diverse populations.

Compared to the LD panels using the HM3 set of variants, our new TagIt-based LD panels increased the proportion of SNPs being tagged ( $r^2 \geq 0.8$ ) up to 2 folds. The best improvement was 7 folds for SNPs with MAF 0.01-0.05 in AFR ancestry (Supplementary Table 7, Supplementary Fig.1). In all ancestry groups, the best-performing PRSs were those constructed using the expanded TagIt set of variants and/or the recomputed pairwise LD with large sample sizes (Supplementary Table 8, Supplementary Fig.2), suggesting that the prediction performance benefited from increasing SNP coverage, better LD modeling, or both.

### Matched single-ancestry PRS performance is positively correlated with the sample size of GWAS summary statistics.

We first trained PRSs (Supplementary Tables 3,4) using GWAS summary statistics and four different LD panels, each matched to the ancestry of the validation cohorts (Supplementary Tables 5,6). When using a single-ancestry summary statistics, matching the GWAS and LD panel ancestries to that of the validation dataset resulted in the best prediction performance for the EUR (iAUC = 0.07-0.14) and EAS (iAUC = 0.02-0.16) ancestries. However, the prediction was poorer for AFR (iAUC = 0.02-0.03), AMR (iAUC = 0.02-0.04), and SAS (iAUC = 0.02-0.04) (Fig.2a-e, Supplementary Table 9).

The effect size of the PRS on T2D risk showed a similar pattern with larger OR per SD from the meta-analysis of the EUR (OR per SD [95% CI] = 2.31 [2.26-2.35]) and EAS (2.19 [2.12-2.27]) validation cohorts, compared to the AFR (1.38 [1.34-1.43]), AMR (1.64 [1.57-1.71]) and SAS (1.6 [1.5-1.74]). The EUR and EAS PRSs also showed better power to identify individuals at the highest T2D genetic risk than those trained for the other ancestries. For instance, individuals at the 90<sup>th</sup> percentile of the respective PRS distribution had ~4-fold increased risk of T2D in the EUR (OR [95% CI] = 4.01 [3.82-4.21]) and in the EAS (3.58 [3.24-3.94]) ancestries, compared to ~2-fold increased risk for individuals in the AFR (1.82 [1.64-2.01]), AMR (2.1 [1.84-2.39]) and SAS (2.22 [1.75-2.81]) ancestries (Fig.2f-j, Supplementary Table 9). We observed consistent results regardless of adjustment for BMI, indicating that the predictive performance of the PRSs was robust to the inclusion of this well-established T2D risk factor (Supplementary Table 10, Supplementary Fig.3).

### Using GWAS data from large non-matched ancestries improves PRS performance in ancestries with limited GWAS sample size.

We then tested whether constructing PRS using GWAS summary statistics from ancestries with larger sample sizes, such as EUR and EAS, could improve T2D prediction in other ancestries with limited GWAS data, despite the larger ancestral differences between discovery and validation cohorts. Compared to the matched single-ancestry PRSs, those based on the EUR GWAS improved the T2D prediction in the AFR, AMR, and SAS validation cohorts but had lower predictions in the EAS.

A PRS based on EAS GWAS showed modestly improved T2D prediction in the AMR and SAS validation cohorts, but worse performance in the AFR and EUR (Fig.2a-e, Supplementary Table 9). The effect sizes of the PRSs were consistent with their prediction performance. The best-performing single-ancestry PRSs were the one derived from EUR GWAS for AFR (OR per SD (95% CI) = 1.75 [1.67-1.82]), AMR (3.07 [2.89-3.27]), SAS (2.21 [2.03-2.42]), and EUR (2.31 [2.26-2.35]) validation cohorts, while for EAS cohorts (2.19 [2.12-2.27]), the best-performing single-ancestry PRS was based on EAS GWAS (Fig.2f-j, Supplementary Table 9).

### Multi-ancestry PRSs show the best prediction performance in all ancestries.

We further applied a multi-ancestry PRS method, PRS-CSx<sup>20</sup>, which jointly models ancestry-specific GWAS and LD panel data from multiple ancestries. This approach leverages the increased statistical power from GWAS of five continental ancestries, which are jointly modelled to maximize the power of variants that are present in all ancestries,

while still incorporating the effects of variants that are specific or enriched in specific ancestries, even if their sample size is modest. Compared to the best single-ancestry PRSs, the prediction performance of the multi-ancestry PRSs was the highest across validation cohorts from all five ancestries (iAUC ranging from 0.02-0.06 in AFR, 0.06-0.09 in AMR, 0.04-0.17 in EAS, 0.07-0.14 in EUR, and 0.06-0.10 in SAS) (Fig.2a-e, Supplementary Table 9).

The multi-ancestry PRSs also had the highest effect sizes along with smaller confidence intervals across all ancestries (OR per SD [95% CI] = 1.73 [1.67-1.80] in AFR, 2.82 [2.67-2.97] in AMR, 2.45 [2.36-2.54] in EAS, 2.36 [2.32-2.41] in EUR, and 2.23 [2.05-2.42] in SAS). The improvement was particularly notable for individuals at the extremes of the PRS distributions. For instance, the individuals in the 97.5<sup>th</sup> percentile of the multi-ancestry PRSs had 3 to 7-fold increased T2D risk compared to those in the interquartile range (OR [95% CI] = 3.43 [2.8-4.21] in AFR, 7.47 [5.64-9.89] in AMR, 6.62 [5.58-7.85] in EAS, 6.25 [5.72-6.82] in EUR and 4.50 [2.70-7.53] in SAS (Fig.2f-j, Supplementary Table 9).

#### Multi-ancestry T2D PRSs outperform previously published T2D PRSs.

We then leveraged the AoU cohort to compare this study's best-performing multi-ancestry PRSs against the published T2D PRSs from the PGS catalog.<sup>22</sup> We tested 55 out of 147 available PRS for T2D (accessed on October 07, 2024), after excluding pathway-specific PRSs or those for which the AoU cohort was used as a training dataset due to potential model overfitting. In all ancestry groups, our best-performing multi-ancestry PRSs showed better predictive performance than the previously published T2D PRSs from the PGS catalog. The differences with the best published PRSs from the PGS catalog were statistically significant for AFR, AMR, and EUR (De Long  $p < 9 \times 10^{-4}$ , Bonferroni-corrected threshold=0.05/55 PRSs tested; iAUC for D-PRISM multi-ancestry PRS-CSx vs. best iAUC from PGS Catalog: 0.041 vs. 0.029 in AFR, 0.086 vs. 0.073 in AMR, 0.143 vs. 0.123 in EUR) (Fig.3a,b,d). The improvement was nominally significant for EAS (0.074 vs. 0.058,  $p=0.03$ ) (Fig.3c), but not for SAS ( $p=0.58$ ) (Fig.3e), likely due in part to their limited sample sizes in the AoU cohort for these two ancestries (Supplementary Table 11, Supplementary Fig.4-8).

We also compared different strategies for PRS construction. Some studies leverage ancestry diversity using the inverse variance-weighted (IVW) meta-analysis results from multi-ancestry GWAS, typically applying a single LD panel — often of EUR ancestry— which may fail to model LD patterns and tag ancestry-specific variants accurately. In contrast, PRS-CSx jointly models GWAS from multiple ancestries while accounting for differences in allele frequencies and LD patterns across ancestries. To assess the value of PRS-CSx, we constructed a PRS-CS model using summary statistics from the IVW GWAS meta-analysis of the D-PRISM ancestry-specific GWAS summary statistics and tested its performance in the validation cohorts. In all ancestry groups, the T2D prediction performance of the multi-ancestry PRS-CSx was better than that of the PRS-CS models constructed using the IVW GWAS meta-analysis. The improvement was statistically significant in the AMR, EAS and EUR (iAUC for the multi-ancestry PRS-CSx vs. the multi-ancestry PRS-CS: 0.086 vs. 0.077 in AMR, 0.074 vs. 0.061 in EAS, and 0.143 vs. 0.126 in EUR, De Long  $p < 0.05$ ) (Fig.3b-d) but not significantly different for AFR and SAS (0.041 vs. 0.038 in AFR,  $p=0.07$ , 0.061 vs. 0.060 in SAS,  $p=0.91$ ) (Fig.3a,e, Supplementary Table 11, Supplementary Figures 4-8).

The largest multi-ancestry T2D GWAS meta-analysis to date, conducted by Suzuki et al.,<sup>3</sup> includes a sample size 16% larger than our D-PRISM GWAS datasets (N=2,535,601, 428,452 cases, 2,107,149 controls), as we intentionally held out several cohorts for PRS training and validation in this study. Nevertheless, since the AoU cohort was not included in the discovery GWAS by Suzuki et al., we also evaluated the performance of a PRS-CS model built using those summary statistics. Despite a smaller sample size, our D-PRISM multi-ancestry PRS-CSx showed better prediction performance in the AMR, EUR, and EAS ancestry groups (iAUC for D-PRISM multi-ancestry PRS-CSx vs. Suzuki et al. PRS-CS: 0.086 vs. 0.077 in AMR, 0.074 vs. 0.059 in EAS, and 0.143 vs. 0.128 in EUR, De Long  $p < 0.05$ ) (Fig.3b-d). In the AFR and SAS ancestry groups, the D-PRISM multi-ancestry PRS-CSx yielded prediction performance compared to the Suzuki et al. PRS-CS model (0.041 vs. 0.041,  $p=0.82$  in AFR, 0.061 vs. 0.068,  $p=0.34$  in SAS) (Fig.3a,e).

In addition, our D-PRISM multi-ancestry PRS-CSx models showed consistently better prediction performance in four of the five ancestries (i.e., AMR, EAS, EUR, and SAS,  $p < 0.05$ ) (Fig.3b-e) compared to a multi-ancestry PRS that was restricted to the 1,289 genome-wide significant variants (rsPRS) identified in the discovery GWAS by Suzuki et al. We did not observe significant improvement in the AFR ancestry (Fig.3a, Supplementary Table 11, Supplementary Figs 4-8).

### Multi-ancestry T2D PRSs are associated with the risk of microvascular diabetes complications.

T2D adversely affects the functioning of multiple organs, and the long-term complications accompanying the disease contribute the greatest morbidity for patients suffering from T2D.<sup>23</sup> We leveraged the AoU cohort to assess if the best-performing multi-ancestry T2D PRSs may be helpful to identify individuals who are at high risk of developing diabetes associated macrovascular [i.e., cardiovascular disease (CAD), ischemic stroke (IS)] and microvascular complications [i.e., diabetic retinopathy (DR), proliferative diabetic retinopathy (PDR), diabetic nephropathy (DN), and end-stage diabetic nephropathy (ESDN)]. Given the sample size constraints of the AoU cohort, we only assessed the AFR, AMR, and EUR ancestries. We restricted microvascular complication analyses to individuals with T2D, as these outcomes are largely diabetes specific. For macrovascular complications, which also occur in those without T2D, we included all individuals and adjusted for T2D status in the models.

In each of the three ancestries, the D-PRISM multi-ancestry PRS-CSx was associated with increased risk of developing DR among individuals with T2D: OR per SD [95% CI] = 1.24 [1.12-1.37] in AFR, 1.44 [1.27-1.63] in AMR, 1.32 [1.21-1.44] in EUR. For the more severe form, PDR, the PRSs were also associated with increased risk in AFR, AMR, and EUR ancestries: 1.35 [1.09-1.67] in AFR, 1.94 [1.51-2.49] in AMR, 1.55 [1.19-2.01] in EUR. Additionally, the multi-ancestry PRSs predicted the risk for developing DN: 1.49 [1.27-1.76] in AMR and 1.23 [1.12-1.35] in EUR, as well as ESDN: 1.75 [1.35-2.27] in AMR and 1.41 [1.1-1.8] in EUR. It was also associated with increased risk of CAD in EUR only: 1.05 [1.02-1.08] ( $p < 0.008$ , Bonferroni-corrected threshold from 0.05/6 diabetes complications tested) (Fig.4, Supplementary Table 12).

## **Discussion**

Identifying individuals with high T2D risk is essential to prioritize those who will benefit from lifestyle or therapeutic interventions to delay the disease or its complications.<sup>24</sup> While clinical risk factors can identify individuals at risk, polygenic risk scores can be estimated at birth and can identify at risk individuals who may clinically be perceived at low risk, e.g., without a family history of T2D, young, lean, etc.<sup>8</sup> Multiple efforts have been made to identify individuals at high genetic risk for T2D. More than a hundred PRS for T2D have been published<sup>22</sup>, yet they have poorer predictive performance in underrepresented populations. Several reasons influence the low transferability, including the overrepresentation of European populations in T2D GWAS and differential LD patterns between variants across ancestries. Importantly, populations in which PRSs are less predictive, including the AFR, AMR, and SAS, are disproportionately affected by diabetes and its complications, highlighting the risk of exacerbating health disparities by applying PRSs derived only from European genetic information.<sup>11</sup> To improve the transferability and predictive accuracy of PRSs across diverse populations, including those underrepresented in GWAS efforts, various multi-ancestry PRS methods have been developed, while in parallel, several efforts have focused on expanding the representation of diverse ancestry populations in GWAS.<sup>3,15,16,25,26</sup> Beyond GWAS discovery, multiple and sufficiently large sample sizes are also essential for training and validating PRSs, and rigorous data aggregation and harmonization following best practices are crucial to determine PRS accuracy.

In this study, we developed the most comprehensive PRSs for T2D across five continental ancestries and conducted extensive evaluations in at least four independent cohorts per ancestry. To accomplish this we leveraged nearly all available genetic datasets with T2D phenotype information to: i) maximize ancestry diversity by harmonizing 125 T2D GWAS datasets, including up to 2.2 million individuals; ii) enhance the representation and tagging of the genetic variants contributing to the PRS by generating new ancestry-specific LD reference panels, iii) train PRS models across ancestries and iv) thoroughly validate the PRSs in multiple independent harmonized cohorts and assess their association with diabetes-related complications.

As previously described<sup>12,27</sup>, we observed that PRSs have lower accuracy when the validation sample is genetically distant from the discovery GWAS sample. However, underrepresented populations in GWAS have limited power and imprecise variant effect size estimates. For this reason, PRSs derived from EUR GWAS continue to outperform those derived from ancestry-matched GWAS in AFR, AMR, and SAS ancestries, likely due to the EUR GWAS sample size being at least 11 times larger. In contrast, applying a EUR PRS yielded worse performance for EAS ancestry than using a matched-ancestry EAS GWAS, even though the latter was four times smaller than the EUR GWAS. This suggests that non-matched ancestry PRSs (i.e., PRSs based on GWAS from a different ancestry than the validation cohort) improve performance only when the GWAS sample size in the other ancestry is orders of magnitude larger than that of the matched-ancestry GWAS, which may lack power to capture ancestry-specific genetic effects. However, when power is sufficient, ancestry-matched GWAS can improve accuracy compared to larger cross-ancestry

GWAS. The minimum ancestry-specific GWAS sample size for powerful PRS prediction may vary depending on the trait and ancestry and may require further investigation. Regardless, we observe that for all the ancestry groups, including EUR, which has the largest GWAS sample size<sup>15</sup>, the optimal PRS strategy is to combine PRSs from multiple ancestries, here using PRS-CSx. This approach can leverage the most accurate estimates from variants present in all ancestries while still accounting for population-specific or enriched variants, improving the overall prediction accuracy.

For all ancestries, the multi-ancestry PRSs showed the strongest associations with T2D risk among individuals at the extremes of the risk distribution. For instance, individuals of EUR, EAS, or AMR ancestry in the top 97·5<sup>th</sup> percentile of the PRS distribution were associated with a seven-fold increased risk of developing T2D than those with average scores. Despite limited GWAS representation, the multi-ancestry PRS still shows the most predictive accuracy for individuals of AFR and SAS ancestry relative to the matched-ancestry PRS. However, overall performance remains lower than in other ancestry groups, with individuals in the 97·5<sup>th</sup> percentile associated with three-fold and four-fold increased risk of T2D. These results underscore that, despite our best efforts, there is still a substantial gap among the performance of PRSs, particularly in these two ancestry groups.

Using data from the AoU cohort, we demonstrate that our multi-ancestry PRSs outperformed all previously available PRSs for T2D, likely because we used the largest and most diverse GWAS data and multi-ancestry-based methodology for PRS development. Additionally, we constructed our multi-ancestry PRSs using a standardized and rigorous approach and tested them extensively across diverse ancestry groups, supporting their broader applicability. For example, the multi-ancestry PRSs outperformed PRSs derived from the largest trans-ancestry meta-analysis. These findings underscore that, in addition to increasing representation in GWAS, leveraging methods that jointly model GWAS and LD panels across multiple ancestries can enhance the tagging of causal variants, thereby improving predictive performance. In contrast, standard methods that use multi-ancestry IVW GWAS meta-analysis results rely on a single LD reference panel and may fail to capture ancestry-specific LD patterns, limiting their ability to model genetic risk prediction accurately.

Previous studies have reported significant associations between T2D PRS, proliferative retinopathy, and end-stage diabetic nephropathy across and within ancestries.<sup>15,16</sup> We confirm and extend previous findings by showing that multi-ancestry PRSs predict microvascular complications in T2D individuals of AFR, AMR, and EUR ancestries. Notably, our results reveal stronger associations and capture a broader range of diabetes complications severity than previously reported.

Our PRSs still face some limitations, mainly due to the lack of diversity in available genetic data. First, while constructing multi-ancestry PRS including GWAS from diverse ancestries—even those with smaller sample sizes—has proven beneficial to capture ancestry-specific effects, the SNP effect sizes are still strongly influenced by the largest European cohorts. Second, we acknowledge that using discrete population categories or restricting the analyses to groups defined by genetic similarity is sub-optimal, particularly in highly admixed populations where such groupings do not fully capture heterogeneous ancestry. While our current approach partially addresses these challenges, it remains limited by the need to group individuals into discrete ancestries, rather than a continuum. A critical next step will be adopting continuous genetic ancestry methods, which will require broader data access and new analytical strategies.

Despite criticism regarding PRSs' limited added value to clinical risk factors, it has been shown that incorporating PRSs can help identify high-risk individuals among people who are clinically perceived as low risk, such as the young, the lean, or those with sparse clinical data<sup>8</sup>, and that PRSs improve prediction over family history alone<sup>28</sup>. Ongoing randomized controlled trials are starting to evaluate the impact of PRS implementation into adult primary care as an additional risk factor for the primary care provider to consider.<sup>7,29</sup> Importantly, for any PRS-informed interventions to be viable, a state-of-the-art T2D PRS is a necessary first step. Among the characteristics to prioritize when using a PRS are its generalizability, portability, and implementation feasibility. For instance, Lennon *et al.*<sup>6</sup> prioritized a multi-ancestry T2D PRS<sup>30</sup> trained using EUR, EAS, and AFR GWAS data, and proposed a high-risk cut-off of 2%. Compared to the rest of the non-high-risk population, this PRS showed ORs ranging from 4·44 [3·60-5·49] in EUR to a maximum of 2·35 [1·54-3·60] in AFR cohorts. Notably, our multi-ancestry PRSs consistently improved the prediction performance of T2D irrespective of the individual's continental ancestry over any previously existing PRS. Compared to people with average genetic risk, those above the high genetic risk cut-off (2·5%) have ORs ranging

from 3·43 to 7·47 across ancestries, compared to the interquartile group. We therefore propose using the T2D PRS models presented here for future clinical applications as they show improved predictive ability across all ancestries.

In summary, this study addresses a critical gap by delivering the most comprehensive and rigorously tested set of multi-ancestry PRSs for T2D. These scores improve risk prediction across diverse populations and enhance the identification of individuals at high genetic risk of developing T2D and microvascular complications, including mild and severe forms. By making the PRS weights publicly available, we provide a valuable resource for researchers and clinicians seeking to advance genetic risk stratification and develop prevention strategies for T2D.

## Contributors

J.M.M. and M.C.Y.N. conceived the study. A.H.-C. and J.K. led the analyses. R.M., Y.L., K.S., L.E.P., H.K.N., J.C., S.L., M.R., and K.L. performed the training and/or validation of the PRS analyses. Y.J.K., R.G.W., D.S., S.H.K., and R.J.F.L. supervised the training and/or validation of the PRS analyses. T.G., A.K.M., M.L., J.E.B., X.S., J.M.M. and M.C.Y.N. jointly directed this work. The rest of the authors provided data for the different stages of the analyses. A.H.-C., J.K., J.M.M., and M.C.Y.N. wrote the first draft of the manuscript. All authors contributed to interpreting the data, and they read, revised, and approved the final manuscript.

## Declaration of interests

H.C.G. holds the McMaster-Sanofi Population Health Institute Chair in Diabetes Research and Care. He reports research grants from Eli Lilly, Novo Nordisk, and Hanmi Pharmaceutical; grants to support continuing education programs from Eli Lilly, Abbott, Sanofi, Novo Nordisk, and Boehringer Ingelheim; honoraria for speaking from AstraZeneca, Eli Lilly, Zuellig, and Jiangsu Hanson; and consulting fees from Abbott, Bayer, Biolinq, Eli Lilly, Novo Nordisk, Pfizer, Shionogi, and Zealand. M.S.U. has consulting activity and research funded in collaboration with Novo Nordisk. A.K.M. has research funded in collaboration with Novo Nordisk. M.S.U. has research funded in collaboration with Novo Nordisk and is an unpaid research collaborator with AstraZeneca. J.M.M. has research funded in collaboration with Novo Nordisk. S.R.P. has had research funded by Philips Respiroics and consulting fees from Apnimed, Bayer, Philips Respiroics, Mineralys, and SleepRes. M.A.N.'s participation in this project was part of a competitive contract awarded to DataTecnica LLC by the National Institutes of Health to support open science research. He also currently owns stock in Character Bio and Neuron23 Inc.

## Data sharing

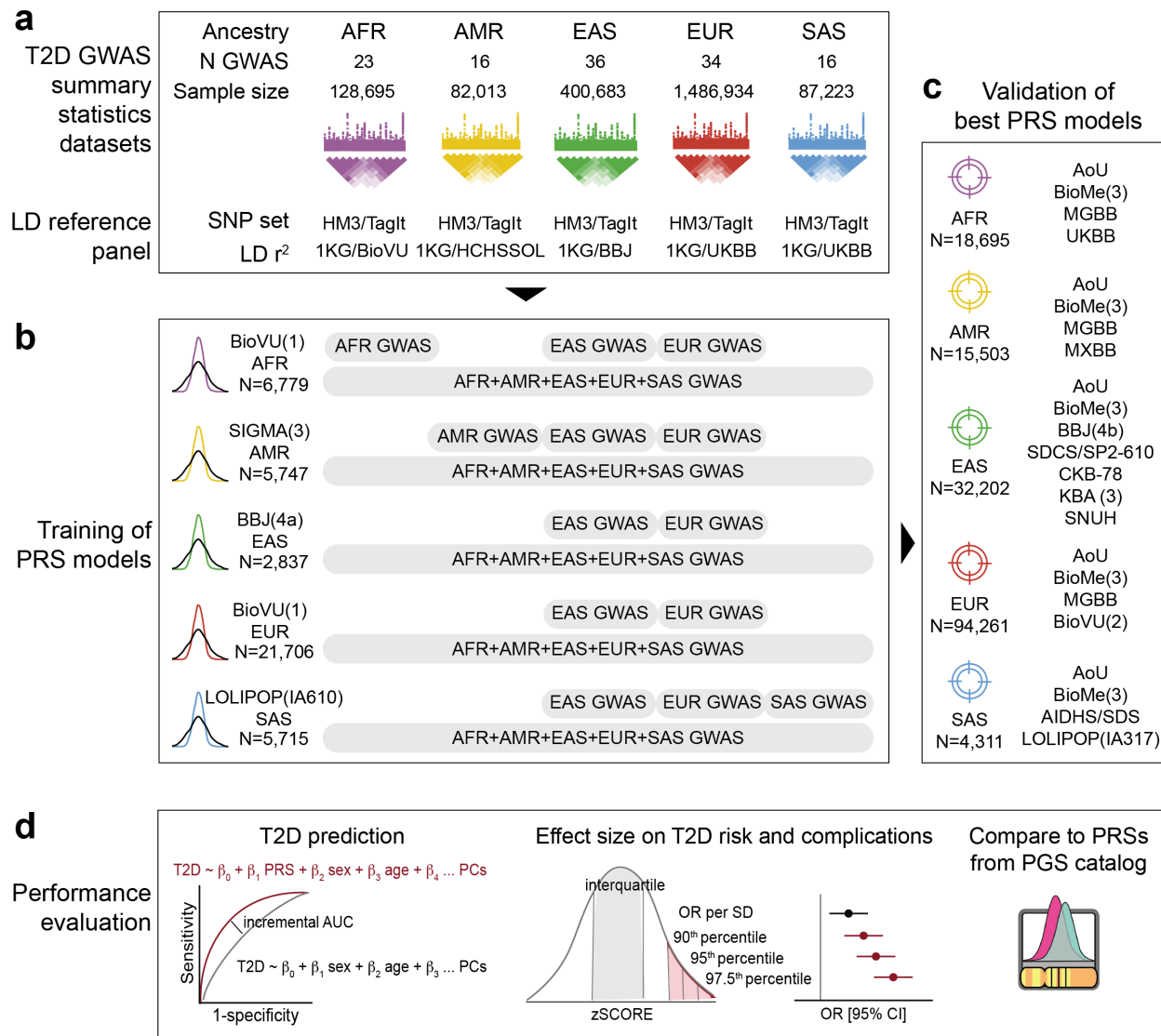
Individual participant data is not available because they are subject to data protection laws and restrictions imposed by the ethics committee to ensure study participants' privacy. The study protocol and the individual methods are included in the methods section. The developed PRS weights are available without restrictions through the PGS catalog (<https://www.pgscatalog.org>).

## Acknowledgments

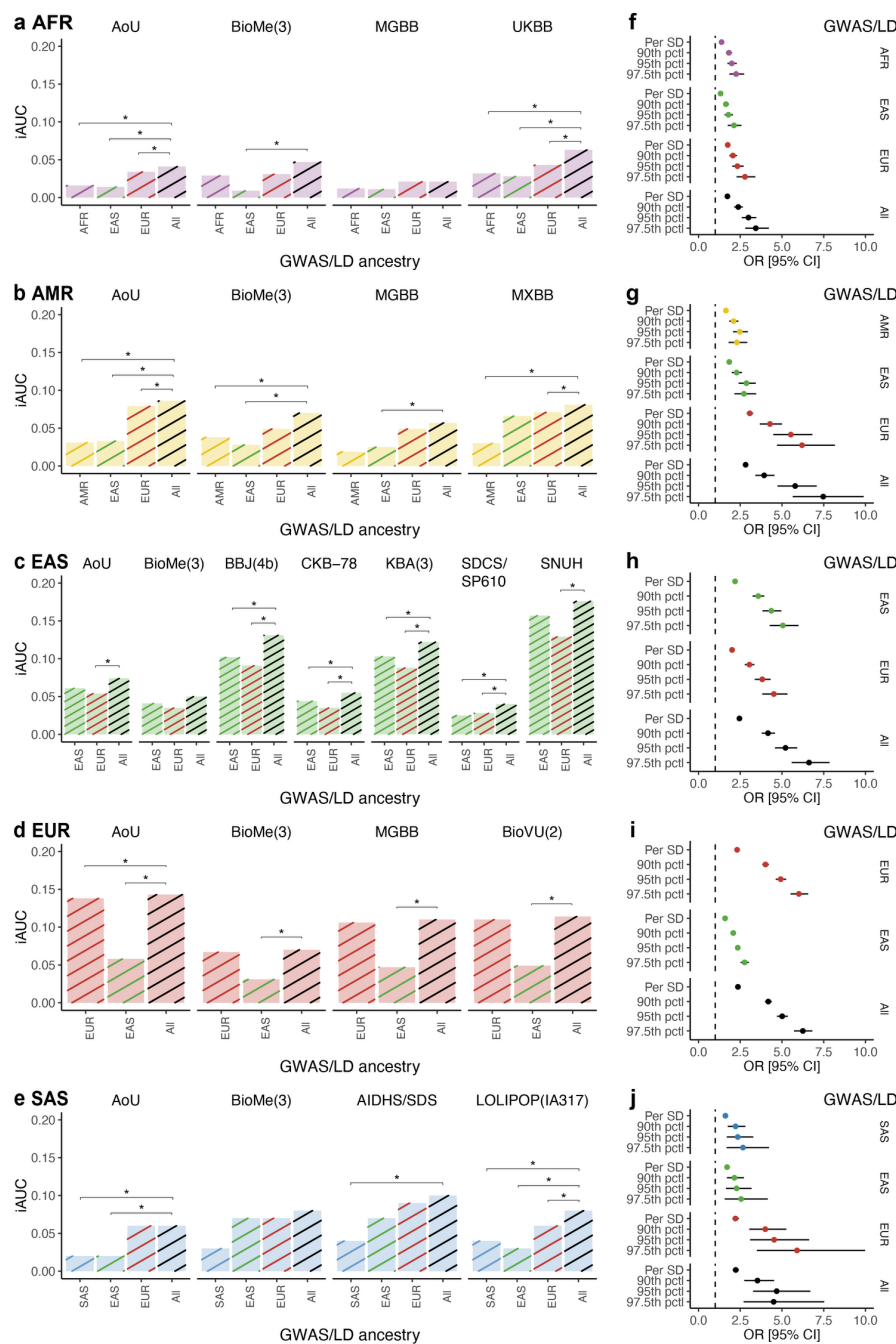
This work is supported by National Human Genome Research Institute (NHGRI) of the National Institutes of Health (NIH) U01HG011723. A.H.-C. is supported by the American Diabetes Association (ADA) grant 11-23-PDF-35. Y.L. is supported by R56HL150186, R01HL158884, and R01DK135938. M.O.G. was supported in part by NIH grants from the National Institute of Diabetes and Digestive and Kidney Disease (NIDDK) (P30-DK063491) and from the National Center for Advancing Translational Sciences (NCATS) (UL1TR001420, UL1TR001881) and the Eris M. Field Chair in Diabetes Research. S.S.R. is supported by NHGRI U01HG011723. M.A.N. research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), NIH, Department of Health and Human Services; project number Z01 AG000535, as well as the National Institute of Neurological Disorders and Stroke (NINDS). This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>). A.L. is supported by grant 2020096 from the Doris Duke Foundation, the ADA grant 7-22-ICTSPM-23, and NHGRI U01HG011723. A.P.M. acknowledges support from the NIHR Manchester Biomedical Research Centre (NIHR203308). C.N.S. was supported by the NIH (R01DK118011; R01DK136671) and the ADA (11-22-JDFPM-06). J.B.M. is supported by grant UMDK078616. B.F.V. is grateful for support from the NIDDK DK138521 and DK126194. A.K.M. is supported by grant UMDK078616. J.M.M. is supported by ADA grant #11-22-ICTSPM-16 and by NHGRI U01HG011723, by the NIDDK under Award Number R01DK137993 and U01 DK140757, AMP

CMD award from RFP 6 from the Foundation for the NIH, and a Medical University of Bialystok (MUB) grant from the Ministry of Science and Higher Education (Poland). This work is supported by the Novo Nordisk Foundation (NNF21SA0072102). M.C.Y.N. is supported by U01HG011723, R01DK066358 and U01DK105556.

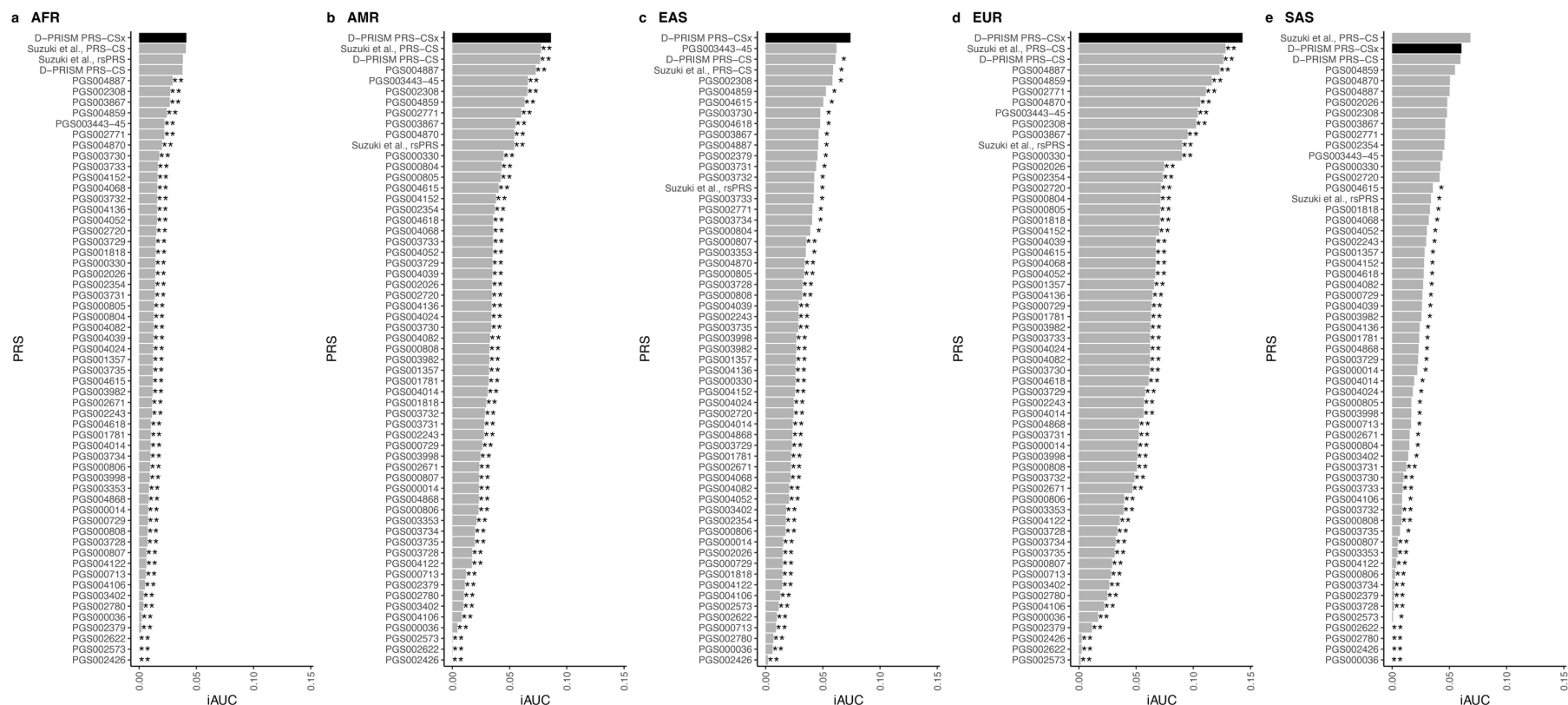
## Figures



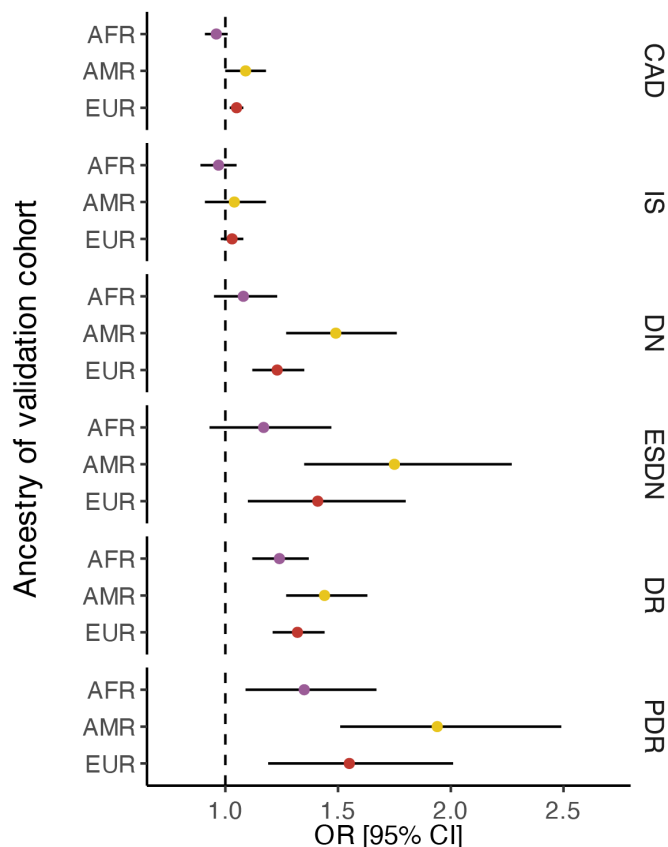
**Fig.1 | Overall analysis approach. a**, Overview of the 125 T2D GWAS datasets, variant sets, and pairwise LD information used to generate five ancestry-specific T2D GWAS meta-analyses and 20 LD reference panels for PRSs training. **b**, Independent, ancestry-specific cohorts used to train the PRS models and select the optimal continuous shrinkage prior from five phi values (i.e., 0.01, 0.001,  $1 \times 10^{-4}$ ,  $1 \times 10^{-5}$ ,  $1 \times 10^{-6}$ ) based on predictive performance. Single-ancestry PRSs using GWAS summary statistics and LD panels matched to the validation ancestry or using data from the EAS or EUR ancestries. Multi-ancestry PRSs jointly modeled GWAS summary statistics and LD panels from all five ancestry groups. **c**, Set of 23 ancestry-specific and independent cohorts used to validate the 18 best-performing PRS. **d**, Evaluation of PRS predictive performance, including: i) incremental AUC, calculated as the difference between the AUC of the full model (PRS + covariates) and the model without the PRS, ii) proportion of variation in T2D status explained by the PRS, estimated using Nagelkerke's  $r^2$ , iii) odds ratio per standard deviation (OR per SD) of the PRS distribution or odds ratio (OR) comparing PRS distribution extremes relative to the interquartile range.



**Fig. 2 | Performance of the T2D PRS in the validation cohorts across ancestry groups.** a-e: Incremental AUC (iAUC) of the T2D PRS in the validation cohorts across ancestry groups: a, AFR, b, AMR, c, EAS, d, EUR, e, SAS. For each ancestry, the best-performing single-ancestry and multi-ancestry PRSs were evaluated. Each bar represents a single cohort. Bar colors represent the ancestry group: purple for AFR, yellow for AMR, green for EAS, red for EUR, and blue for SAS. Line colors represent the ancestry of the T2D GWAS summary statistics and LD panels used to train the PRS, using the same color codes for single-ancestry PRSs, and black for multi-ancestry PRSs \*De Long test  $p < 0.05$ . f-j: Odds ratio from the meta-analysis of validation cohorts across ancestry groups: f, AFR, g, AMR, h, EAS, i, EUR, j, SAS. Points represent the odds ratio per standard deviation (OR per SD) of the PRS distribution or the odds ratio (OR) comparing different PRS distribution extremes relative to the interquartile range. Error bars show the 95% confidence intervals (95% CI). Point colors represent the ancestry of the T2D GWAS summary statistics and LD panels used to train the PRS. \*De Long  $p < 0.05$ .



**Fig. 3 | Performance of D-PRISM multi-ancestry PRSs compared to published T2D PRSs from the PGS Catalog and other sources in the All of Us cohort. a-e: Incremental AUC (iAUC) across ancestry groups: a, AFR, b, AMR, c, EAS, d, EUR, e, SAS. Black bars highlight this study's multi-ancestry PRSs. \*De Long  $p < 0.05$ ; \*\*Bonferroni-corrected De Long  $p < 9 \times 10^{-4}$ .**



**Fig. 4 | Association of D-PRISM multi-ancestry PRS with common complications of diabetes in the All of Us cohort.** Odds ratio of this study’s multi-ancestry PRS for the following outcomes: CAD (cardiovascular disease), IS (ischemic stroke), DN (diabetic nephropathy), ESDN (end-stage diabetic nephropathy), DR (diabetic retinopathy), and PDR (proliferative diabetic retinopathy). Points represent the odds ratios per standard deviation (OR per SD) and are colored according to the genetic ancestry of the individuals tested: purple for AFR, yellow for AMR, and red for EUR. Error bars show the 95% confidence intervals (95% CI).

## References

- 1 Dianna J. Magliano, Edward J. Boyko, IDF Diabetes Atlas 10th edition scientific committee. IDF DIABETES ATLAS. 2021.
- 2 Mercader JM, Ng MCY, Manning AK, Rich SS. Predicting diabetes risk in diverse populations: what next? *Lancet Diabetes Endocrinol* 2021; **9**: 808–10.
- 3 Suzuki K, Hatzikotoulas K, Southam L, *et al.* Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* 2024; **627**: 347–57.
- 4 Udler MS, McCarthy MI, Florez JC, Mahajan A. Genetic Risk Scores for Diabetes Diagnosis and Precision Medicine. *Endocr Rev* 2019; **40**: 1500–20.
- 5 Diabetes Prevention Program Research Group, Knowler WC, Fowler SE, *et al.* 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. *Lancet* 2009; **374**: 1677–86.
- 6 Lennon NJ, Kottyan LC, Kachulis C, *et al.* Selection, optimization and validation of ten chronic disease polygenic risk scores for clinical implementation in diverse US populations. *Nat Med* 2024; **30**: 480–7.
- 7 Vassy JL, Brunette CA, Lebo MS, *et al.* The GenoVA study: Equitable implementation of a pragmatic randomized trial of polygenic-risk scoring in primary care. *Am J Hum Genet* 2023; **110**: 1841–52.
- 8 Mandla R, Schroeder P, Porneala B, *et al.* Polygenic scores for longitudinal prediction of incident type 2 diabetes in an ancestrally and medically diverse primary care physician network: a patient cohort study. *Genome Med* 2024; **16**: 63.

- 9 Fortmann AL, Savin KL, Clark TL, Philis-Tsimikas A, Gallo LC. Innovative Diabetes Interventions in the U.S. Hispanic Population. *Diabetes Spectr* 2019; **32**: 295–301.
- 10 Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019; **51**: 584–91.
- 11 Kullo IJ, Conomos MP, Nelson SC, *et al.* The PRIMED Consortium: Reducing disparities in polygenic risk assessment. *Am J Hum Genet* 2024; **111**: 2594–606.
- 12 Kachuri L, Chatterjee N, Hirbo J, *et al.* Principles and methods for transferring polygenic risk scores across global populations. *Nat Rev Genet* 2024; **25**: 8–25.
- 13 Kidenya BR, Mboowa G. Inclusiveness of the All of Us Research Program improves polygenic risk scores and fosters genomic medicine for all. *Communications medicine* 2024; **4**: 227.
- 14 All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. *Nature* 2024; **627**: 340–6.
- 15 Mahajan A, Spracklen CN, Zhang W, *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nat Genet* 2022; **54**: 560–72.
- 16 Vujkovic M, Keaton JM, Lynch JA, *et al.* Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat Genet* 2020; **52**: 680–91.
- 17 Kurki MI, Karjalainen J, Palta P, *et al.* FinnGen provides genetic insights from a well-phenotyped isolated population. *Nature* 2023; **613**: 508–18.
- 18 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74.
- 19 Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019; **10**: 1776.
- 20 Ruan Y, Lin Y-F, Feng Y-CA, *et al.* Improving polygenic prediction in ancestrally diverse populations. *Nat Genet* 2022; **54**: 573–80.
- 21 Wojcik GL, Fuchsberger C, Taliun D, *et al.* Imputation-Aware Tag SNP Selection To Improve Power for Large-Scale, Multi-ethnic Association Studies. *G3 (Bethesda)* 2018; **8**: 3255–67.
- 22 Lambert SA, Wingfield B, Gibson JT, *et al.* Enhancing the Polygenic Score Catalog with tools for score calculation and ancestry normalization. *Nat Genet* 2024; **56**: 1989–94.
- 23 Farmaki P, Damaskos C, Garmpis N, Garmpi A, Savvanis S, Diamantis E. Complications of the Type 2 Diabetes Mellitus. *Curr Cardiol Rev* 2020; **16**: 249–51.
- 24 Crandall JP, Knowler WC, Kahn SE, *et al.* The prevention of type 2 diabetes. *Nat Clin Pract Endocrinol Metab* 2008; **4**: 382–93.
- 25 Loh M, Zhang W, Ng HK, *et al.* Identification of genetic effects underlying type 2 diabetes in South Asian and European populations. *Commun Biol* 2022; **5**: 329.
- 26 Spracklen CN, Horikoshi M, Kim YJ, *et al.* Identification of type 2 diabetes loci in 433,540 East Asian individuals. *Nature* 2020; **582**: 240–5.
- 27 Ding Y, Hou K, Xu Z, *et al.* Polygenic scoring accuracy varies across the genetic ancestry continuum. *Nature* 2023; **618**: 774–81.
- 28 Drzymalla E, Raffield L, Kolor K, *et al.* Additive Value of Polygenic Risk Score to Family History for Type 2 Diabetes Prediction: Results From the All of Us Research Database. *Diabetes Care* 2025; **48**: 212–9.
- 29 Linder JE, Allworth A, Bland HT, *et al.* Returning integrated genomic risk and clinical recommendations: The eMERGE study. *Genet Med* 2023; **25**: 100006.
- 30 Ge T, Irvin MR, Patki A, *et al.* Development and validation of a trans-ancestry polygenic risk score for type 2 diabetes in diverse populations. *Genome Med* 2022; **14**: 70.