

1 **Title:**

2 Trans-ancestry genome-wide association study identifies novel genetic mechanisms in rheumatoid  
3 arthritis

4  
5 **Authors:**

6 Kazuyoshi Ishigaki<sup>1,2,3\*</sup>, Saori Sakaue<sup>1,2,4,5\*</sup>, Chikashi Terao<sup>6,7,8\*</sup>, Yang Luo<sup>1,2,5,9,10</sup>, Kyuto Sonehara<sup>4,11</sup>,  
7 Kensuke Yamaguchi<sup>12,13</sup>, Tiffany Amariuta<sup>1,2,5,9,10,14</sup>, Chun Lai Too<sup>15,16</sup>, Vincent A Laufer<sup>17,18</sup>, Ian C  
8 Scott<sup>19,20</sup>, Sebastien Viatte<sup>21,22,35</sup>, Meiko Takahashi<sup>23</sup>, Koichiro Ohmura<sup>24</sup>, Akira Murasawa<sup>25</sup>, Motomu  
9 Hashimoto<sup>26,27</sup>, Hiromu Ito<sup>26,28</sup>, Mohammed Hammoudeh<sup>29</sup>, Samar Al Emadi<sup>29</sup>, Basel K Masri<sup>30</sup>, Hussien  
10 Halabi<sup>31</sup>, Humeria Badsha<sup>32</sup>, Imad W Uthman<sup>33</sup>, Xin Wu<sup>34</sup>, Li Lin<sup>34</sup>, Ting Lin<sup>34</sup>, Darren Plant<sup>21</sup>, Anne  
11 Barton<sup>21,35</sup>, Gisela Orozco<sup>21,35</sup>, Suzanne MM Verstappen<sup>35,36</sup>, John Bowes<sup>21,35</sup>, Alexander J MacGregor<sup>37</sup>,  
12 Suguru Honda<sup>38,39</sup>, Masaru Koido<sup>6</sup>, Kohei Tomizuka<sup>6</sup>, Yoichiro Kamatani<sup>6,40</sup>, Hiroaki Tanaka<sup>41</sup>, Eiichi  
13 Tanaka<sup>38,39</sup>, Akari Suzuki<sup>13</sup>, Yuichi Maeda<sup>11,42,43</sup>, Kenichi Yamamoto<sup>4,44,45</sup>, Satoru Miyawaki<sup>46</sup>, Gang Xie<sup>47</sup>,  
14 Jinyi Zhang<sup>47,48</sup>, Chris Amos<sup>49</sup>, Ed Keystone<sup>50</sup>, Gertjan Wolbink<sup>51</sup>, Irene van der Horst-Bruinsma<sup>52</sup>, Jing  
15 Cui<sup>9</sup>, Katherine P Liao<sup>9,10,53</sup>, Robert J Carroll<sup>54</sup>, Hye-Soon Lee<sup>55,56</sup>, So-Young Bang<sup>55,56</sup>, Katherine A  
16 Siminovitch<sup>47,57</sup>, Niek de Vries<sup>52</sup>, Lars Alfredsson<sup>58</sup>, Solbritt Rantapää-Dahlqvist<sup>59</sup>, Elizabeth W Karlson<sup>9</sup>,  
17 Sang-Cheol Bae<sup>55,56</sup>, Robert P Kimberly<sup>60</sup>, Jeffrey C Edberg<sup>60</sup>, Xavier Mariette<sup>61</sup>, Tom Huizinga<sup>62</sup>,  
18 Philippe Dieudé<sup>63,64</sup>, Matthias Schneider<sup>65</sup>, Martin Kerick<sup>66</sup>, Joshua C Denny<sup>54,67,68</sup>, The Biobank Japan  
19 Project<sup>69</sup>, Koichi Matsuda<sup>70,71</sup>, Keitaro Matsuo<sup>72,73</sup>, Tsuneyo Mimori<sup>24</sup>, Fumihiko Matsuda<sup>23</sup>, Keishi Fujio<sup>74</sup>,  
20 Yoshiya Tanaka<sup>41</sup>, Atsushi Kumanogoh<sup>11,21,42,43</sup>, Matthew Traylor<sup>75</sup>, Cathryn M Lewis<sup>75,76</sup>, Stephen  
21 Eyre<sup>21,35</sup>, Huji Xu<sup>34,77,78</sup>, Richa Saxena<sup>5</sup>, Thurayya Arayssi<sup>79</sup>, Yuta Kochi<sup>12,13</sup>, Katsunori Ikari<sup>38,80,81</sup>,  
22 Masayoshi Harigai<sup>38,39</sup>, Peter K Gregersen<sup>82</sup>, Kazuhiko Yamamoto<sup>13</sup>, S. Louis Bridges, Jr<sup>83,84</sup>, Leonid  
23 Padyukov<sup>16</sup>, Javier Martin<sup>66</sup>, Lars Klareskog<sup>16</sup>, Yukinori Okada<sup>4,11,44,85,86\*\*</sup>, Soumya  
24 Raychaudhuri<sup>1,2,5,9,10,21\*\*</sup>

25 (\*these authors contributed equally; \*\*corresponding authors)

26  
27 **Corresponding authors:** Yukinori Okada and Soumya Raychaudhuri.

28 **Affiliations:**

29 1, Center for Data Sciences, Brigham and Women's Hospital and Harvard Medical School, Boston, MA,  
30 USA.

31 2, Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical  
32 School, Boston, MA, USA.

33 3, Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences, Yokohama,  
34 Japan.

35 4, Department of Statistical Genetics, Osaka University Graduate School of Medicine, Suita, Japan.

36 5, Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA,  
37 USA.

38 6, Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences,  
39 Yokohama, Japan.

40 7, Clinical Research Center, Shizuoka General Hospital, Shizuoka, Japan.

41 8, The Department of Applied Genetics, The School of Pharmaceutical Sciences, University of Shizuoka,  
42 Shizuoka, Japan.

43 9, Division of Rheumatology, Inflammation, and Immunity, Department of Medicine, Brigham and  
44 Women's Hospital and Harvard Medical School, Boston, MA, USA.

45 10, Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

46 11, Integrated Frontier Research for Medical Science Division, Institute for Open and Transdisciplinary  
47 Research Initiatives, Osaka University, Suita, Japan.

48 12, Department of Genomic Function and Diversity, Medical Research Institute, Tokyo Medical and  
49 Dental University, Tokyo, Japan.

50 13, Laboratory for Autoimmune Diseases, RIKEN Center for Integrative Medical Sciences, Yokohama,  
51 Japan.

52 14, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

53 15, Immunogenetics Unit. Allergy and Immunology Research Center. Institute for Medical Research.  
54 National Institutes of Health Complex. Ministry of Health, Malaysia.

- 55 16, Department of Medicine, Division of Rheumatology, Karolinska Institutet and Karolinska University  
56 Hospital, Stockholm, Sweden.
- 57 17, Department of Clinical Immunology and Rheumatology, University of Alabama at Birmingham School  
58 of Medicine, AL, USA.
- 59 18, Michigan Medicine Department of Pathology, MI, USA.
- 60 19, Haywood Academic Rheumatology Centre, Haywood Hospital, Midlands Partnership NHS  
61 Foundation Trust, Burslem, UK.
- 62 20, Primary Care Centre Versus Arthritis, School of Medicine, Keele University, Keele, UK.
- 63 21, Centre for Genetics and Genomics Versus Arthritis, Division of Musculoskeletal and Dermatological  
64 Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of  
65 Manchester, Manchester Academic Health Science Centre, Manchester, UK.
- 66 22, Lydia Becker Institute of Immunology and Inflammation, Faculty of Biology, Medicine and Health, The  
67 University of Manchester, UK.
- 68 23, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan.
- 69 24, Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto  
70 University, Kyoto, Japan.
- 71 25, Department of Rheumatology, Niigata Rheumatic Center, Niigata, Japan.
- 72 26, Department of Advanced Medicine for Rheumatic Diseases, Kyoto University Graduate School of  
73 Medicine, Kyoto, Japan.
- 74 27, Department of Clinical Immunology, Graduate School of Medicine, Osaka City University, Osaka,  
75 Japan.
- 76 28, Department of Orthopaedic Surgery, Kurashiki Central Hospital, Okayama, Japan.
- 77 29, Department of Internal Medicine, Hamad Medical Corporation, Doha, Qatar.
- 78 30, Department of Internal Medicine, Jordan Hospital, Amman, Jordan.
- 79 31, Rheumatology Division, Department of Internal Medicine, King Faisal Specialist Hospital and  
80 Research Center, Jeddah6, Saudi Arabia.
- 81 32, Dr. Humeira Badsha Medical Center, Emirates Hospital, Dubai, United Arab Emirates.

- 82 33, Department of Rheumatology, American University of Beirut, Beirut, Lebanon.
- 83 34, Department of Rheumatology and Immunology, Shanghai Changzeng Hospital, The Second Military  
84 Medical University, China.
- 85 35, NIHR Manchester Biomedical Research Centre, Manchester University Foundation Trust,  
86 Manchester, UK.
- 87 36, Centre for Epidemiology Versus Arthritis, Centre for Musculoskeletal Research, Division of  
88 Musculoskeletal and Dermatological Sciences, the University of Manchester, Manchester, UK.
- 89 37, Norwich Medical School, University of East Anglia, NR4 7UQ. UK.
- 90 38, Institute of Rheumatology, Tokyo Women's Medical University Hospital, Tokyo, Japan.
- 91 39, Division of Rheumatology, Department of Internal Medicine, Tokyo Women's Medical University  
92 School of Medicine, Tokyo, Japan.
- 93 40, Laboratory of Complex Trait Genomics, Department of Computational Biology and Medical Sciences,  
94 Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.
- 95 41, The First Department of Internal Medicine, School of Medicine, University of Occupational and  
96 Environmental Health Japan, Kitakyushu, Japan.
- 97 42, Department of Respiratory Medicine and Clinical Immunology, Osaka University Graduate School of  
98 Medicine, Suita, Japan.
- 99 43, Department of Immunopathology, Immunology Frontier Research Center (WPI-IFReC), Osaka  
100 University, Suita, Japan.
- 101 44, Laboratory of Statistical Immunology, Immunology Frontier Research Center (WPI-IFReC), Osaka  
102 University, Suita, Japan.
- 103 45, Department of Pediatrics, Osaka University Graduate School of Medicine, Suita, Japan.
- 104 46, Department of Neurosurgery, Faculty of Medicine, the University of Tokyo, Tokyo, Japan.
- 105 47, Lunenfeld-Tanenbaum Research Institute, Toronto, Canada.
- 106 48, Department of Medicine, University of Toronto, Toronto, Canada.
- 107 49, Baylor College of Medicine, Houston, TX, USA.
- 108 50, University of Toronto, Toronto, Canada.

- 109 51, Department of Rheumatology, Amsterdam Rheumatology and Immunology Center (ARC), Reade,  
110 Amsterdam, The Netherlands.
- 111 52, Dept Rheumatology, Amsterdam University Medical Centers, AMC & ARC, Amsterdam, The  
112 Netherlands.
- 113 53, Massachusetts Veterans Epidemiology Research and Information Center, VA Boston Healthcare  
114 System, Boston, MA, USA.
- 115 54, Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA.
- 116 55, Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea.
- 117 56, Hanyang University Institute for Rheumatology Research, Seoul, Korea.
- 118 57, Departments of Medicine and Immunology, University of Toronto, Toronto, Canada.
- 119 58, Department of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.
- 120 59, Dept of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden.
- 121 60, Center for Clinical and Translational Science, Division of Clinical Immunology and Rheumatology,  
122 Department of Medicine, University of Alabama at Birmingham, AL, USA.
- 123 61, Department of Rheumatology, Université Paris-Saclay, Assistance Publique - Hôpitaux de Paris,  
124 Hôpital Bicêtre, INSERM UMR1184, Le Kremlin Bicêtre, France.
- 125 62, Leiden University Medical Center, Lieden, The Netherlands.
- 126 63, University of Paris, Paris, France.
- 127 64, Department of Rheumatology, INSERM U1152, Hôpital Bichat-Claude Bernard, APHP, Paris,  
128 France.
- 129 65, Department of Rheumatology & Hiller Research Unit Rheumatology, UKD, Heinrich-Heine University,  
130 Düsseldorf, Germany.
- 131 66, Institute of Parasitology and Biomedicine Lopez-Neyra, CSIC, Granada, Spain.
- 132 67, All of Us Research Program, Office of the Director, National Institutes of Health, Bethesda, MD, USA.
- 133 68, Department of Medicine, Vanderbilt University School of Medicine, Nashville, TN, USA.
- 134 69, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

- 135 70, Laboratory of Genome Technology, Human Genome Center, Institute of Medical Science, The  
136 University of Tokyo, Tokyo, Japan.
- 137 71, Laboratory of Clinical Genome Sequencing, Department of Computational Biology and Medical  
138 Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan.
- 139 72, Division of Cancer Epidemiology and Prevention, Department of Preventive Medicine, Aichi Cancer  
140 Center Research Institute, Nagoya, Aichi, Japan.
- 141 73, Department of Cancer Epidemiology, Nagoya University Graduate School of Medicine, Nagoya,  
142 Aichi, Japan.
- 143 74, Department of Allergy and Rheumatology, Graduate School of Medicine, The University of Tokyo,  
144 Japan.
- 145 75, Department of Medical & Molecular Genetics, King's College London, London, UK.
- 146 76, Social, Genetic and Developmental Psychiatry Centre, King's College London, London, UK.
- 147 77, School of Clinical Medicine Tsinghua University, Beijing, China.
- 148 78, Peking-Tsinghua Center for Life Sciences, Tsinghua University, Beijing, China.
- 149 79, Department of Internal Medicine, Weill Cornell Medicine-Qatar, Education City, Doha, Qatar.
- 150 80, Department of Orthopedics, Tokyo Women's Medical University School of Medicine, Tokyo, Japan.
- 151 81, Division of Multidisciplinary Management of Rheumatic Diseases, Tokyo Women's Medical  
152 University, Tokyo, Japan.
- 153 82, Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institutes for Medical  
154 Research, Northwell Health, Manhasset, NY, USA.
- 155 83, Department of Medicine, Hospital for Special Surgery, New York, NY, USA.
- 156 84, Division of Rheumatology, Weill Cornell Medicine, New York, NY, USA.
- 157 85, Laboratory for Systems Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan.
- 158 86, Center for Infectious Disease Education and Research (CiDER), Osaka University, Suita, Japan.
- 159

160 **Abstract**

161 Trans-ancestry genetic research promises to improve power to detect genetic signals, fine-mapping  
162 resolution, and performances of polygenic risk score (PRS). We here present a large-scale genome-wide  
163 association study (GWAS) of rheumatoid arthritis (RA) which includes 276,020 samples of five ancestral  
164 groups. We conducted a trans-ancestry meta-analysis and identified 124 loci ( $P < 5 \times 10^{-8}$ ), of which 34  
165 were novel. Candidate genes at the novel loci suggested essential roles of the immune system (e.g.,  
166 *TNIP2* and *TNFRSF11A*) and joint tissues (e.g., *WISP1*) in RA etiology. Trans-ancestry fine mapping  
167 identified putatively causal variants with biological insights (e.g., *LEF1*). Moreover, PRS based on trans-  
168 ancestry GWAS outperformed PRS based on single-ancestry GWAS and had comparable performance  
169 between European and East Asian populations. Our study provides multiple insights into the etiology of  
170 RA and improves genetic predictability of RA.

171

172 **Main text**

173 Rheumatoid arthritis (RA) is an autoimmune disease in which the immune system attacks the synovium  
174 in the joints, leading to chronic tissue inflammation, joint destruction, and disability. While recent  
175 therapeutic developments now alter the course of disease, RA mechanisms have yet to be fully  
176 elucidated and a cure has yet to be identified. RA can be divided into two major subtypes (seropositive  
177 and seronegative RA) based on the presence or absence of RA-specific serum antibodies (rheumatoid  
178 factor or anti-citrullinated peptide antibodies)<sup>1</sup>. Since RA is highly heritable<sup>2-4</sup>, genetic research has the  
179 potential to advance our understanding of its pathology. Indeed, previous studies successfully identified  
180 candidates of causal alleles, genes, pathways, and cell types<sup>2,5-7</sup>. For example, previous studies  
181 suggested that CD4<sup>+</sup> effector T cells play a central role and the T cell receptor signaling pathway drives  
182 autoimmunity in RA<sup>7-10</sup>.

183 Trans-ancestry genetic research has multiple advantages over single-ancestry analysis. First,  
184 genome-wide association study (GWAS) in a single ancestry can be underpowered to detect signals  
185 from a causal allele with low allele frequency in that ancestry. As notable examples, the causal alleles  
186 can be ancestry-specific, as shown in studies for other complex diseases<sup>11-13</sup>. Having multiple ancestries  
187 with different allele frequency spectrum can improve the power. Second, single-ancestry GWAS are  
188 hampered by the specific linkage disequilibrium (LD) structure in that ancestry, which could obscure the  
189 ability to effectively fine-map an associated locus<sup>14,15</sup>. Trans-ancestry GWAS can improve fine-mapping  
190 resolution by leveraging the distinct LD structures in each ancestry<sup>15-17</sup>. Third, PRS generally has limited  
191 transferability across ancestries. For example, when the PRS model is developed based on GWAS in  
192 European (EUR) populations, PRS performs poorly in non-EUR populations<sup>18</sup>. PRS based on trans-  
193 ancestry GWAS can potentially improve its performance in multiple ancestries<sup>19,20</sup>; this is a clinically  
194 important topic since PRS can benefit patients via precision medicine. Although many RA genetic studies  
195 were conducted in non-EUR populations<sup>2,3,17,21-24</sup>, they were relatively small in the sample sizes, and  
196 much larger research efforts have focused on EUR populations<sup>6,25-32</sup>.

197 Here, we report a large-scale trans-ancestry GWAS of RA, including individuals of EUR, East  
198 Asian (EAS), African (AFR), South Asian (SAS), and Arabian (ARB) ancestries. While seropositive and



199 seronegative RA are associated with phenotypic differences, they have shared heritability<sup>33</sup>, and their  
200 risk alleles appear to be similar outside of the major histocompatibility complex (MHC) locus<sup>34</sup>. Therefore,  
201 we initially focused on all RA, and then we restricted cases to seropositive patients. After identifying  
202 novel loci, we conducted fine-mapping to elucidate potential molecular mechanism of risk alleles. We  
203 examined the extent to which genetic signals are shared across ancestries while also investigating  
204 ancestry-specific genetic signals. We developed PRS models using our GWAS results and compared  
205 their performances across all ancestries. Our study provides multiple insights into the etiology of RA and  
206 highlights the importance and further needs of diversifying the ancestral background of GWAS  
207 participants.

208

## 209 **RESULTS**

### 210 **Trans- and single-ancestry GWAS**

211 We included 37 cohorts comprising 35,871 RA patients and 240,149 control individuals of EUR, EAS,  
212 AFR, SAS, and ARB ancestry (**Figure 1a; Supplementary Table 1 and 2**); 22,350 cases and 74,823  
213 controls in 25 EUR cohorts; 11,025 cases and 162,608 controls in eight EAS cohorts; 999 cases and  
214 1,108 controls in two AFR cohorts; 986 cases and 1,258 controls in one SAS cohort; and 511 cases and  
215 352 controls in one ARB cohort. RA-specific serum antibodies were measured in 31,963 (89%) of cases;  
216 among them 27,448 (86%) were seropositive and 4,515 (14%) were seronegative (**Supplementary**  
217 **Table 1; Methods**). To confirm the diversity of ancestral backgrounds, we projected each individual's  
218 genotype into principal component (PC) space which was calculated using all individuals in 1000  
219 Genomes Project Phase 3 (1KG Phase3). We further conducted uniform manifold approximation and  
220 projection (UMAP) using their top 20 PC scores. This revealed finer scale ancestral structure, and  
221 confirmed that our study represented many 1KG Phase 3 ancestries (**Figure 1b and 1c; Extended Data**  
222 **Figure 1**).

223 After quality control and imputation, we conducted GWAS in each cohort by logistic regression  
224 (**Methods**). We calculated genomic inflation using all variants outside of the MHC locus and observed  
225 little evidence of statistical inflation (mean of lambda = 1.01; S.D = 0.04; **Supplementary Table 1**).

226 Primary analysis included all cases, while we restricted cases to seropositive patients in a secondary  
227 analysis.

228 We then conducted a meta-analysis using all cohorts across five ancestries by the inverse-  
229 variance weighted fixed effect model (trans-ancestry GWAS). We observed almost identical effect sizes  
230 between this trans-ancestry GWAS and the previously reported 100 RA risk alleles<sup>2</sup> (Pearson's  $r = 0.98$   
231 and  $P = 2.8 \times 10^{-82}$ ; **Supplementary Figure 1; Supplementary Table 3**). We detected 108 genome-wide  
232 significant loci ( $P < 5 \times 10^{-8}$ ) in this trans-ancestry study: 106 autosomal loci and two loci on the X  
233 chromosome (**Supplementary Table 4 and 5**). For ancestries with multiple cohorts (EUR, EAS, and  
234 AFR), we also conducted a meta-analysis within each ancestry by the same strategy (EUR-, EAS-, and  
235 AFR-GWAS). EUR-GWAS identified three additional autosomal loci which were not significant in trans-  
236 ancestry GWAS (**Supplementary Table 4**). GWAS of seropositive RA additionally detected 13  
237 autosomal loci (**Supplementary Table 4**). In total, we detected significant signals at 122 autosomal loci  
238 outside of the MHC locus and two loci on the X chromosome (**Supplementary Table 4 and 5**; we  
239 provided Manhattan and QQ plots in **Supplementary Figure 2**). Among these 122, 34 autosomal loci  
240 were novel (**Table 1**). Notably, 25 novel loci had not been implicated in any other autoimmune diseases  
241 (**Supplementary Table 4; Methods**).

242 To quantify the heritability, we analyzed our GWAS results using stratified-linkage disequilibrium  
243 score regression (S-LDSC)<sup>10</sup> (**Supplementary Table 2**). Since S-LDSC assumes GWAS has samples  
244 from a single ancestral background and a sufficient sample size, we restricted this analysis to EUR- and  
245 EAS-GWAS. The heritability explained by non-MHC common variants was similar between EUR and  
246 EAS; the liability scale heritability was 0.14 (S.E. = 0.01) for EUR and 0.13 (S.E. = 0.01) for EAS  
247 (**Methods**). LDSC also confirmed that the amount of potential bias in the GWAS results was minimal;  
248 LDSC's intercept = 1.03 for EUR and 1.02 for EAS (**Supplementary Table 2**).

## 250 Fine-mapping analysis

251 We fine-mapped these 122 autosomal loci using approximate Bayesian factor (ABF)<sup>35</sup> (**Methods**). The  
252 95% credible sets included only one variant at seven loci (**Figure 2a**). Of these seven, six have not been

253 reported in prior studies that conducted trans-ancestry fine-mapping of RA<sup>17,36</sup> (**Supplementary Table**  
254 **6**). The 95% credible sets included less than ten variants at 43 loci (**Figure 2a**). We identified 35 fine-  
255 mapped variants with posterior inclusion probability (*PIP*) greater than 0.5, which agree with and largely  
256 subsume prior fine-mapping results<sup>6,17,36</sup>; in addition, nine novel loci are represented (**Figure 2b**;  
257 **Supplementary Figure 3; Supplementary Table 6**). The proportion of non-synonymous variants was  
258 higher in the credible set variants with high *PIP* (*PIP* > 0.5) than low *PIP* (odds ratio (OR) = 9.3; one-  
259 sided Fisher exact test *P* = 0.02; **Figure 2c**).

260 We quantified the 95% credible set variants within open chromatin regions in 18 hematopoietic  
261 populations using gchromVAR software<sup>37</sup>. Consistent with previous analyses, we observed the strongest  
262 enrichment in CD4<sup>+</sup> T cells (*P* = 5.4 × 10<sup>-4</sup>; **Extended Data Figure 2**). For example, rs58107865 at the  
263 *LEF1* locus (*PIP* > 0.99), rs7731626 at the *ANKRD55* locus (*PIP* > 0.99), and rs10556591 at the *ETS1*  
264 locus (*PIP* = 0.84) are located within CD4<sup>+</sup> T cell-specific open chromatin regions (*Z* score > 2;  
265 **Supplementary Table 6; Methods**). Among them, rs58107865 is a novel risk variant and suggested the  
266 importance of regulatory T cells (T-reg) in RA biology (**Figure 2d**); *LEF1* synergizes with *FOXP3* to  
267 reinforce the gene networks essential for T-reg<sup>38</sup>. These results recapitulated a critical role of CD4<sup>+</sup> T  
268 cells, especially T-regs, in RA biology.

269 As expected, compared with single-ancestry GWAS, trans-ancestry GWAS produced smaller  
270 sized credible sets and higher *PIP* (one-sided paired Wilcoxon test *P* < 3.1 × 10<sup>-11</sup> and < 1.1 × 10<sup>-9</sup>,  
271 respectively) (**Figure 2a; Supplementary Figure 3**). For example, the *WDFY4* locus included 6,391  
272 variants in the EUR 95% credible set, 64 variants in the EAS set, but only one in the trans-ancestry set, a  
273 missense variant of *WDFY4* (rs7097397; R1816Q; **Figure 2e**). Using a down-sampling experiment, we  
274 confirmed that this benefit was due to diversified LD structures as well as the increased sample size  
275 (**Supplementary Figure 3; Supplementary Note**).

## 276 277 **Conditional analysis**

278 We conducted conditional analyses in each cohort to explore associations independent from the lead  
279 variants and meta-analyzed the results using the same strategy. We detected 24 independent signals at

280 21 loci ( $P < 5.0 \times 10^{-8}$ ; **Supplementary Table 7**). Consistent with previous results<sup>6,31</sup>, we observed the  
281 largest number of independent associations at the *IL2RA*, *TYK2*, and *TNFAIP3* loci, where we observed  
282 three independent alleles (**Extended Data Figure 3**). The first and second lead variants at the *TYK2*  
283 locus were missense variants predicted to have damaging effects on *TYK2* protein function (SIFT score  
284  $< 0.01$ ; **Supplementary Table 8**).

285 In the *IL6R* locus, the conditional analysis identified two variants, rs12126142 (the first lead  
286 variant) and rs4341355 (the second lead variant), that were weakly correlated with each other but  
287 independently associated with RA ( $r^2 = 0.23$  and  $0.15$  in EUR and EAS of 1KG Phase 3, respectively;  
288 **Supplementary Table 7**). Interestingly, their protective alleles (rs12126142-A and rs4341355-C) almost  
289 always create a haplotype with the risk allele of the other variant (**Extended Data Figure 4**). Hence, the  
290 conditional analysis disentangled the independent yet mutually attenuating signals (**Figure 3a**). By  
291 analyzing expression quantitative trait loci (eQTL) and splicing quantitative trait loci (sQTL) in three  
292 immune cells from Blueprint consortium ( $CD4^+$  T cells, monocytes, and neutrophils)<sup>39</sup>, we found that  
293 rs12126142 and rs4341355 are likely to affect *IL6R* transcripts via different mechanisms. GWAS signals  
294 conditioned on rs4341355 colocalized with sQTL signals in monocytes (posterior probability of  
295 colocalization estimated by coloc software<sup>40</sup> ( $PP_{coloc}$ )  $> 0.99$ ; **Figure 3b**; **Supplementary Table 9**); this  
296 sQTL signal corresponds to a previously-reported splicing isoform of soluble *IL6R*<sup>41</sup>. On the other hand,  
297 GWAS signals conditioned on rs12126142 colocalized with eQTL signals in  $CD4^+$  T cells ( $PP_{coloc} = 0.97$ ;  
298 **Figure 3b**; **Supplementary Table 9**). Therefore, our results suggested that both splicing and total  
299 expression of *IL6R* independently influence the RA genetic risk.

300 Our conditional analyses also suggested interesting biology at the *PADI4-PADI2* locus. We found  
301 two independent associations at this locus: esv3585367 (the first lead variant at a *PADI4* intron) and  
302 rs2076616 (the second lead variant at a *PADI2* intron), consistent with previous studies<sup>17,42</sup> (**Figure 4a**;  
303 **Supplementary Table 7**). In sQTL results from the Blueprint consortium<sup>39</sup>, both *PADI4* and *PADI2* sQTL  
304 signals in neutrophils were colocalized with corresponding GWAS associations ( $PP_{coloc} = 0.98$  and  $0.79$ ,  
305 respectively; **Supplementary Table 9**), suggesting that alternative splicing of *PADI4* and *PADI2* likely  
306 increases the RA risk.

307 *PADI4* is critical for RA because it encodes an enzyme that citrullinates proteins, the main target  
308 for autoantibody in RA<sup>23,43–45</sup>. However, unlike *IL6R* with two functionally distinct isoforms<sup>41</sup>, full picture of  
309 *PADI4* splicing isoforms has not been extensively studied. To elucidate detailed molecular biology at the  
310 *PADI4* locus, we generated long-read sequencing datasets and inspected full-length *PADI4* transcripts.  
311 We identified a novel and probably non-functional splice isoform that produces a truncated protein-  
312 arginine deiminase (PAD) domain, an essential catalytic domain with two calcium-binding sites<sup>46</sup> (**Figure**  
313 **4b**). Next, we differentially quantified *PADI4* isoforms using RNA-seq data from 105 Japanese donors<sup>5</sup>.  
314 We found that the risk allele (esv3585367-A) was associated with the decrease of the non-functional  
315 novel isoform and the increase of the functionally intact isoform (**Figure 4c**). Notably, the allelic effect on  
316 the total expression, which had been conventionally used for eQTL studies, was not predictive of that on  
317 the functional isoform (the right panel in **Figure 4c**). Together, our analysis provided a novel genetic  
318 mechanism of *PADI4* and highlighted the importance of thoroughly investigating splice isoforms at the  
319 risk loci using long-read sequencing.

320

### 321 **Candidate causal genes at the associated loci**

322 We next inferred the possible molecular consequences of all 148 detected variants: 124 lead variants  
323 (including two variants on the X chromosome) and 24 secondary variants detected by the conditional  
324 analysis.

325 We first focused on coding variants in LD with the lead variants in this GWAS ( $r^2 > 0.6$  both in  
326 EUR and EAS samples in 1KG Phase 3; **Methods**). We found missense variants that may drive genetic  
327 signals at two novel loci (**Table 1; Supplementary Table 8**). An example is rs2269495 (A313V of  
328 *TNIP2*), in LD with a lead variant rs4690029 ( $r^2 = 0.65$  and  $0.89$  in EUR and EAS of 1KG Phase 3,  
329 respectively). rs2269495 is predicted to have a damaging effect on TNIP2 protein function (SIFT score =  
330  $0.02$ ; **Supplementary Table 8**). The protein product of *TNIP2* interacts with A20 (encoded by *TNFAIP3*)  
331 and inhibits NF- $\kappa$ B activation induced by TNF. Mice with a defective mutant *TNIP2* displayed intestinal  
332 inflammation and hypersensitivity to experimental colitis<sup>47</sup>. In addition, *TNIP1*, a homolog of *TNIP2*, was  
333 identified as one of the novel loci in this GWAS (**Supplementary Table 4**). Together, these results

334 suggested that *TNIP1* and *TNIP2* are novel candidate causal genes of RA. Combined with the well-  
335 established *TNFAIP3* locus (ref<sup>48</sup>; **Supplementary Table 4**), these findings further supported the  
336 importance of the TNFAIP3-axis in RA biology.

337 We next inferred the possible molecular consequences using QTLs. We analyzed eQTL and  
338 sQTL in three immune cells from Blueprint consortium (CD4<sup>+</sup> T cells, monocytes, and neutrophils) and  
339 multiple tissues from GTEx consortium<sup>39,49</sup>. We found colocalizing QTL signals at 11 novel loci ( $PP_{coloc} >$   
340  $0.7$ ; **Table 1; Supplementary Table 9 and 10**). Several novel loci with colocalizing QTL signals  
341 suggested the biology of the non-immune systems, including joint tissues. For example, the risk allele of  
342 rs55762233 was associated with the increased expression of *CILP2* ( $PP_{coloc} = 0.82$ ), which encodes a  
343 component of the cartilage extracellular matrix. Its homologous gene, *CILP1*, was recently reported as  
344 one of the candidate autoantigens of RA<sup>50</sup>. Therefore, the protein product of *CILP2* might also have a  
345 critical role in RA biology.

346 We then searched for other biologically plausible genes which might explain novel signals and  
347 found several genes whose importance was supported by previous knowledge (**Table 1;**  
348 **Supplementary Table 4**). First, *TNFRSF11A* encodes RANK, a key regulator of osteoclast. Its ligand  
349 *RANKL* has been investigated as a potential therapeutic target<sup>51,52</sup>. *TNFRSF11A* is a causal gene for  
350 several bone-related Mendelian disorders<sup>53–55</sup>. Second, *WISP1* encodes a protein essential for  
351 osteoblast differentiation and bone formation<sup>56,57</sup>. In addition, *WISP1* is highly expressed in *HLA-DRA*<sup>hi</sup>  
352 inflammatory sublining fibroblasts, which are dramatically expanded and pathogenic in RA synovium<sup>58</sup>.  
353 Third, *FLT3* encodes a tyrosine kinase that regulates hematopoiesis and knocking out of whose ligand  
354 suppressed arthritis in model mice<sup>59</sup>. A damaging variant of *FLT3* was suggested to be associated with  
355 RA ( $P = 4.3 \times 10^{-4}$ ) and other autoimmune diseases<sup>60</sup>.

356

### 357 **Differences and similarities of genetic risk across ancestries**

358 We next searched for ancestry-specific signals at 122 autosomal loci. We defined ancestry-specificity  
359 when the lead variant in each locus is monomorphic in EUR or EAS samples of 1KG Phase 3. We found  
360 five EUR-specific signals: rs2476601 (a *PTPN22* missense variant), rs9826420 located in *STAG1*



361 intronic region, rs7943728 (a *FADS2* eQTL), and rs34536443 and rs12720356 (both *TYK2* missense  
362 variants). EAS-GWAS also identified an EAS-specific signal at the *TYK2* locus: rs55882956, another  
363 *TYK2* missense variant. We thus detected two EUR-specific and one EAS-specific signal at the *TYK2*  
364 locus (**Extended Data Figure 5**). All these ancestry-specific signals were also reported by previous  
365 studies<sup>2,25,61</sup>. This study was underpowered to detect specific signals in non-EUR and non-EAS  
366 ancestries (**Supplementary Figure 4** and **Supplementary Note**). Although ancestry-specific signals are  
367 relatively few, they include predominantly large effect size variants, many of which are missense, and  
368 hence they are valuable resources to understand the etiology of RA.

369 Although we found these ancestry-specific signals, this study showed that genetic signals are  
370 generally shared across ancestries. We compared effect sizes between EUR-GWAS with non-EUR-  
371 GWASs at the 30 fine-mapped variants ( $PIP > 0.5$ ; **Methods**). We found that the effect sizes were  
372 strongly correlated among five ancestries (Pearson's  $r = 0.56-0.91$ ; **Supplementary Figure 5**;  
373 **Supplementary Note**). In addition, we targeted genome-wide variants and tested the trans-ancestry  
374 genetic correlation between EUR- and EAS-GWAS using popcorn software<sup>62</sup> (we restricted this analysis  
375 to EUR- and EAS-GWAS to avoid a biased correlation estimate caused by the small sample size). We  
376 again found a strong correlation ( $0.64$  (S.E.= $0.08$ );  $P = 4.4 \times 10^{-17}$ ;  $P$  value reported is for a test that the  
377 correlation is different from 0).

### 379 **Genetic risk differences between seropositive and seronegative RA**

380 The presence or absence of autoantibodies (rheumatoid factor and anti-citrullinated peptide antibodies)  
381 defines two major subgroups of RA: seropositive and seronegative RA<sup>1</sup>. We tested the differences in  
382 genetic signals between them at the 122 significant autosomal loci. Although their effect sizes were  
383 significantly correlated in general (Pearson's  $r = 0.76$ ;  $P = 3.2 \times 10^{-23}$ ), we found significant heterogeneity  
384 in effect size estimates at the nine loci: *CCR6*, *CTLA4*, *NFKBIE-TMEM151B*, *PADI4*, *PTPN22*, *RAD51B*,  
385 *SMIM20-RBPJ*, *TNFRSF14-AS1*, and *UBASH3A* ( $P_{het} < 0.05/122$ ; **Extended Data Figure 6**). Effect size  
386 estimates were larger in seropositive RA at all the nine loci, and these findings suggested etiological  
387 differences between seropositive and seronegative RA. For example, *CCR6* has critical roles in B cell

388 antibody production<sup>63,64</sup>. Together, these findings suggested generally shared genetic risks between  
389 seropositive and seronegative RA outside of the MHC locus, although substantial differences exist  
390 around biologically relevant genes.

391

### 392 **Genome-wide distributions of heritability**

393 To acquire insights into RA biology, we estimated the heritability enrichments within gene regulatory  
394 regions using S-LDSC<sup>10</sup>, a method to infer the genome-wide distribution of all causal variants irrespective  
395 of their effect sizes. We again restricted this analysis to EUR- and EAS-GWAS. We utilized 707 IMPACT  
396 regulatory annotations, which reflect comprehensive cell-type-specific transcription factor (TF)  
397 activities<sup>65</sup>. Briefly, IMPACT probabilistically annotates each nucleotide genome-wide on a scale from 0  
398 to 1, and we considered genomic regions scoring in the top 5% of each annotation. We detected  
399 significant enrichments in 114 annotations in either of EUR and EAS ( $P < 0.05/707 = 7.1 \times 10^{-5}$ ; **Figure**  
400 **5a and Supplementary Table 11**). The amount of heritability explained by each annotation was highly  
401 concordant between EUR and EAS (Pearson's  $r = 0.92$ ;  $P = 1.3 \times 10^{-290}$ ; **Figure 5b**), and we did not  
402 observe significant heterogeneities between EUR and EAS estimates ( $P_{het} > 0.05/707 = 7.1 \times 10^{-5}$ ).  
403 Among annotations with significant enrichments, the one which explained the largest fraction of EUR  
404 heritability was CD4<sup>+</sup> T cell T-bet annotation (90%; S.E. = 11%). This annotation explained 94% (S.E. =  
405 12%) of EAS heritability consistent with analyses on previous GWAS results<sup>66</sup>. Although four out of 114  
406 significant annotations were derived from non-immune cells, controlling the effect of CD4<sup>+</sup> T cell T-bet  
407 annotation canceled out all four enrichments. We also analyzed 396 histone mark annotations, but they  
408 were less enriched in RA heritability than CD4<sup>+</sup> T cell T-bet annotation (**Extended Data Figure 7**;  
409 **Supplementary Table 12; Supplementary Note**). In addition to identifying candidate critical TFs in RA  
410 pathology, these results also suggested that the distribution of causal variants is shared between EUR  
411 and EAS.

412 We next tested whether the findings in S-LDSC analysis can be recapitulated in fine-mapped  
413 variants from genome-wide significant loci. We analyzed credible set variants and found that high *PIP*  
414 variants ( $> 0.5$ ) were enriched in high IMPACT score variants for CD4<sup>+</sup> T cell T-bet annotation ( $> 0.5$ ),



415 compared with low *PIP* variants (**Figure 2c; Extended Data Figure 8**; OR = 8.7; one-sided Fisher exact  
416 test  $P = 1.5 \times 10^{-4}$ ). We found six variants which possess high *PIP* and high IMPACT score, and one of  
417 them was a novel association at the intronic region of *LEF1* (rs58107865; **Supplementary Table 6**).  
418 Together, both polygenic and fine-mapped signals support the critical roles of CD4<sup>+</sup> T cell's T-bet activity  
419 in RA pathology.

420

### 421 **PRS performance across five ancestries**

422 Our results showed that trans-ancestry GWAS can detect causal variants more efficiently than single-  
423 ancestry GWAS and these causal variants are strongly enriched within the CD4<sup>+</sup> T cell T-bet annotation.  
424 Therefore, we hypothesized that trans-ancestry GWAS and CD4<sup>+</sup> T cell T-bet annotation can improve  
425 PRS performances in non-EUR populations. To test this, we developed PRS models with six different  
426 conditions using combinations of two components: i) two variant selection settings (we used all variants  
427 or variants within the top 5% of the CD4<sup>+</sup> T cell T-bet annotation and refer PRS based on each of them  
428 as standard or functionally-informed PRS, respectively) and ii) three GWAS settings (we used trans-  
429 ancestry, EUR-, or EAS-GWAS and refer PRS based on each of them as trans-ancestry, EUR-, or EAS-  
430 PRS, respectively). We designed our study so that there were no overlapping samples; when we  
431 evaluated the PRS performance in a given cohort, we re-conducted meta-analysis excluding that cohort  
432 to develop PRS models (**Figure 6a; Methods**). We defined the performance of PRS by phenotypic  
433 variance (pseudo- $R^2$ ) explained by PRS. We tested PRS performances using ten different *P* value  
434 thresholds. We then selected the threshold with the best performance in each of the six PRS conditions  
435 separately and utilized this threshold for the following analyses.

436 We first evaluated the variant selection settings used for PRS (standard and functionally-informed  
437 PRS). Consistent with our recent study<sup>65</sup>, functionally-informed PRS improved the application of PRS  
438 constructed from different ancestries (EUR-PRS applied to non-EUR cohorts or EAS-PRS applied to  
439 non-EAS cohorts). Functionally-informed PRS increased  $R^2$  by 2.7 fold on average, and we observed the  
440 improvement in 32 of 41 applications (one-sided sign test  $P = 2.2 \times 10^{-4}$ ; **Figure 6b; Supplementary**  
441 **Table 13**). On the other hand, also consistent with our recent study<sup>65</sup>, this improvement was relatively

442 small in the application of PRS constructed from the same ancestry (EUR-PRS applied to EUR cohorts  
443 or EAS-PRS applied to EAS cohorts). Functionally-informed PRS increased  $R^2$  by 1.3 fold on average,  
444 and we observed the improvement in 21 out of 33 applications (one-sided sign test  $P = 0.08$ ; **Figure 6b**;  
445 **Supplementary Table 13**). PRS based on single-ancestry GWAS is affected by LD structures of GWAS  
446 participants' ancestral backgrounds; this can reduce performance in prediction when we apply this PRS  
447 to ancestries with different LD structures. These results confirmed that functionally-informed PRS can  
448 mitigate this problem. Expectedly, for trans-ancestry PRS, which prioritizes causal variants over variants  
449 solely associated through linkage, the benefit of functionally-informed PRS was very subtle; functionally-  
450 informed PRS increased  $R^2$  only by 1.02 fold on average, and we observed the improvement only in 16  
451 out of 37 applications (one-sided sign test  $P = 0.84$ ; **Figure 6b**; **Supplementary Table 13**). Since CD4<sup>+</sup>  
452 T cell T-bet annotation always had beneficial or neutral effects on PRS performances, we used  
453 functionally-informed PRS for the following analyses.

454 We next evaluated the GWAS settings used for PRS (trans-ancestry, EUR-, and EAS-PRS).  
455 Consistent with a recent study<sup>14</sup>, trans-ancestry PRS outperformed EUR-PRS and EAS-PRS; mean  $R^2$   
456 across 37 cohorts were 0.054, 0.041, and 0.022, in trans-ancestry, EUR-, and EAS-PRS, respectively  
457 (**Figure 6c**; **Supplementary Table 13**). Even for EUR cohorts for which the largest same-ancestry  
458 GWAS was available, trans-ancestry PRS outperformed EUR-PRS (one-sided paired Wilcoxon test  $P =$   
459  $3.3 \times 10^{-4}$ ; **Extended Data Figure 9**).

460 Finally, we compared the performance of trans-ancestry PRS in each population. The best  
461 performance was found in the EUR cohorts (mean  $R^2 = 0.059$ ; **Figure 6c**; **Supplementary Table 13**).  
462 Strikingly, the PRS explained around half of the heritability by the non-MHC common variants, which is  
463 the theoretical upper limitation (**Supplementary Table 2**). Even without the MHC region, we were able to  
464 identify 9.9% of the EUR population with an inherited genetic predisposition that conferred three times  
465 increased risk for RA (**Figure 6d**; **Methods**). The performance in the EAS cohorts was comparable with  
466 the EUR cohorts; the mean  $R^2$  was 0.057 (Wilcoxon test  $P = 0.67$  compared with EUR cohorts), and we  
467 were able to identify 5.5% of the EAS population with three times increased risk (**Figure 6c** and **6d**).  
468 However, we observed poor PRS performances in AFR, SAS, and ARB cohorts; the mean  $R^2$  was 0.018

469 (Wilcoxon test  $P = 0.002$  compared with EUR cohorts), and we were able to identify only 0.62% of these  
470 populations with three times increased risk (**Figure 6c** and **6d**; **Extended Data Figure 10**). Together,  
471 trans-ancestry PRS exhibited the best performance in all ancestries in this study. However, the PRS  
472 performance in each ancestry was substantially affected by its sample size in this trans-ancestry GWAS,  
473 which firmly claims that it is imperative to increase the sample sizes of underrepresented ancestries to  
474 equalize genetic predictability of disease status.

475

## 476 **DISCUSSION**

477 This study identified 34 novel genetic signals and less than ten 95% credible sets at 43 loci. By using  
478 multiple functional annotations and prior immunological knowledge, we provided their potential molecular  
479 consequences. In addition to the novel loci, our comprehensive analyses provided novel biological  
480 interpretations to the known loci (e.g., the *IL6R* and *PADI4* loci). We conducted detailed analyses on  
481 ancestry specificity; although most genetic signals are shared across ancestries, we observed some  
482 ancestry-specific signals. We also found several candidates of critical TFs contributing to RA biology.  
483 This trans-ancestry study thus substantially advanced our understanding of RA biology.

484 We utilized the molecular QTL database to infer risk allele's gene regulatory function. This  
485 approach is a standard approach but has a limited ability to explore the allelic role comprehensively, as  
486 reported in a previous study<sup>67</sup>. Since gene regulation is highly cell-type or cell-state specific, we need to  
487 extend the QTL experimental conditions to overcome this limitation. Single-cell QTL analysis may also  
488 represent a promising strategy<sup>68</sup>. Another promising approach is inducing risk alleles in target cell  
489 populations using gene-editing techniques; previous studies reported the feasibility of this approach<sup>69-71</sup>.  
490 Future advance in functional genetics would improve the biological insight from our GWAS.

491 Our study had insufficient power to detect significant signals for seronegative RA outside of the  
492 MHC region due to a limited sample size (4,515 cases of seronegative RA). Although we observed  
493 shared genetic risks between them (**Extended Data Figure 6**), this analysis was restricted to the loci  
494 detected in all RA or seropositive RA. To unveil the specific etiology of seronegative RA further, we need  
495 cohorts that are larger and have better representation of seronegative RA.

496            Poor PRS performance in non-EUR ancestries is becoming one of the major challenges in human  
497 genetics. Conducting a trans-ancestry GWAS on a large scale is a promising strategy to mitigate this  
498 issue. Indeed, trans-ancestry PRS performance in EAS cohorts was comparable to those in EUR  
499 cohorts, demonstrating that this study mitigated inequality of genetic benefit at least partially (**Figure 6c**).  
500 However, this study was underpowered to detect specific signals in non-EUR and non-EAS ancestries,  
501 resulting in poor PRS performance in these ancestries. To overcome these limitations, we need further  
502 efforts to diversify GWAS and increase sample sizes of underrepresented ancestries as in other common  
503 complex diseases.

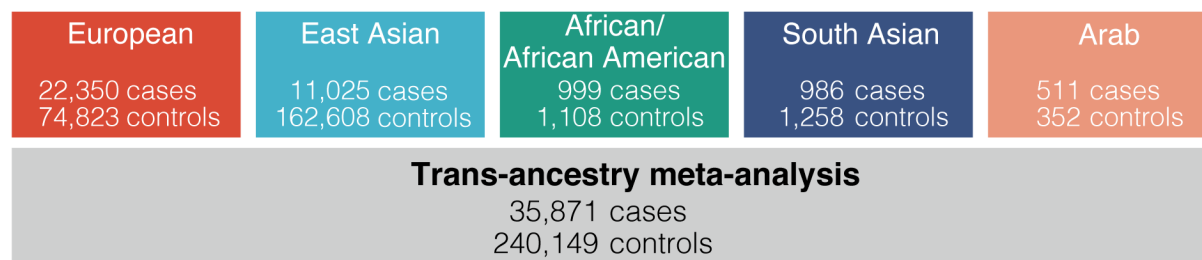
504 **Table 1. Novel RA risk loci detected in this study.**

505 Statistics in the GWAS setting with the lowest *P* values were provided (see **Supplementary Table 4** for  
 506 details). The genomic coordinate is according to GRCh37. Predicted causal gene, predicted molecular  
 507 consequences using eQTL or non-synonymous variants (see **Supplementary Table 8-10** for details);  
 508 OR, odds ratio (the effect allele is the alternative allele); L95, lower 95% confidence interval; U95, upper  
 509 95% confidence interval; allele freq., allele frequency of the effect allele.

Rs ID	Chr.	Position	Nearest gene	Predicted causal gene	OR	L95	U95	P value	Allele freq. in 1KG Phase 3			
									EAS	EUR	AFR	SAS
rs41269479	1	42,166,782	HIVEP3	NA	1.15	1.09	1.20	2.51E-08	0.26	0.28	0.08	0.42
rs41313373	1	92,940,411	GFI1	EVI5(eQTL)	1.12	1.08	1.16	1.08E-08	0.01	0.14	0.01	0.10
rs1188620266	1	235,800,357	GNG4	GNG4(eQTL)	0.91	0.88	0.94	2.06E-08	0.83	0.61	0.22	0.70
rs143259280	2	70,209,168	PCBP1-AS1	C2orf42(eQTL)	1.09	1.06	1.12	2.13E-08	0.46	0.32	0.89	0.32
rs77574423	3	11,984,744	TAMM41, SYN2	SYN2(eQTL)	1.10	1.07	1.14	1.35E-08	0.56	0.72	0.57	0.74
rs62264113	3	127,292,333	TPRA1	TPRA1(eQTL)	0.92	0.89	0.95	4.66E-08	0.27	0.08	0.01	0.21
rs4687070	3	189,306,650	TPRG1, TP63	NA	1.15	1.09	1.20	6.07E-09	0.02	0.07	0.03	0.14
rs4690029	4	2,722,815	FAM193A	TNIP2(p.A313V)	0.94	0.92	0.96	2.83E-09	0.40	0.41	0.59	0.47
rs138066321	4	80,952,409	ANTXR2	ANTXR2(eQTL)	0.93	0.91	0.95	4.48E-10	0.38	0.45	0.09	0.32
rs53107855	4	109,061,618	LEF1	NA	0.84	0.80	0.88	4.92E-14	0.21	0.01	0.00	0.03
rs56787183	5	40,499,290	LINC00603, PTGER4	NA	0.85	0.80	0.90	2.15E-09	0.09	0.00	0.04	0.02
rs244468	5	142,604,421	ARHGAP26	NA	1.07	1.05	1.10	8.19E-09	0.79	0.51	0.41	0.60
rs1422673	5	150,438,988	TNIP1	NA	1.10	1.06	1.14	1.56E-08	0.50	0.19	0.39	0.29
rs113532504	6	15,195,682	LINC01108, JARID2	JARID2(eQTL)	1.10	1.07	1.14	3.42E-08	0.06	0.10	0.36	0.10
rs67318457	6	23,925,021	LOC105374972, NRSN1	NA	1.08	1.05	1.11	1.10E-08	0.14	0.27	0.37	0.05
rs940825	7	17,207,164	AGR3, AHR	NA	1.11	1.07	1.16	3.39E-08	0.18	0.12	0.05	0.06
rs182199544	7	27,084,581	SKAP2, HOXA1	HOXA3(eQTL), HOXA4(eQTL)	0.87	0.84	0.91	3.61E-09	0.01	0.08	0.36	0.03
rs6583441	7	50,361,874	IKZF1	NA	0.95	0.93	0.97	4.69E-08	0.53	0.47	0.33	0.41
rs6979218	7	99,893,148	CASTOR3, SPDYE3	PILRA(p.R78G), PVRIG(p.N81D)	1.09	1.06	1.12	2.24E-11	0.38	0.75	0.91	0.72
rs11777380	8	134,211,965	WISP1	NA	0.92	0.90	0.95	3.00E-10	0.17	0.32	0.08	0.19
rs911760	9	5,438,435	PLGRK1	NA	1.15	1.09	1.20	2.15E-08	0.23	0.19	0.33	0.28
rs734094	11	2,323,220	C11orf21, TSPAN32	NA	1.08	1.05	1.10	3.40E-08	0.19	0.40	0.43	0.45
rs1427749	12	46,370,116	SCAF11	ARID2(eQTL), SLC38A1(eQTL)	0.93	0.90	0.95	1.17E-08	0.89	0.80	0.91	0.95
rs61944750	13	28,634,933	FLT3	NA	0.91	0.88	0.94	1.69E-08	0.05	0.21	0.09	0.09
rs2147161	13	42,982,302	AKAP11, LINC02341	NA	1.10	1.06	1.13	2.73E-08	0.13	0.21	0.02	0.28
rs175714	14	75,981,856	JDP2, BATF	NA	0.94	0.92	0.96	4.14E-08	0.43	0.60	0.30	0.64
rs115284761	15	77,326,836	PSTPIP1	NA	0.91	0.89	0.94	1.71E-11	0.28	0.25	0.15	0.23
rs11375064	17	25,904,074	KSR1	NA	1.08	1.05	1.11	3.92E-08	0.44	0.60	0.53	0.47
rs591549	18	3,542,247	DLGAP1	NA	0.91	0.88	0.94	9.14E-09	0.35	0.68	0.42	0.61
rs371734407	18	60,009,634	TNFRSF11A	NA	1.10	1.06	1.14	4.14E-08	0.44	0.59	0.50	0.52
rs10415976	19	941,603	ARID3A	NA	0.92	0.90	0.95	2.90E-08	0.47	0.08	0.35	0.25
rs55762233	19	19,367,319	HAPLN4	CILP2(eQTL), COMP(eQTL), HAPLN4(eQTL), LPAR2(eQTL), SUGP1(eQTL), TM6SF2(eQTL), TSSK6(eQTL), YJEFN3(eQTL), ZNF101(eQTL)	1.10	1.07	1.14	1.43E-09	0.02	0.17	0.32	0.14
rs28373672	19	36,213,072	KMT2B	LIN37(eQTL)	0.93	0.91	0.96	3.33E-08	0.24	0.23	0.62	0.29
rs8106598	19	52,017,940	SIGLEC12, SIGLEC6	NA	1.08	1.05	1.11	3.12E-08	0.08	0.22	0.32	0.17

510

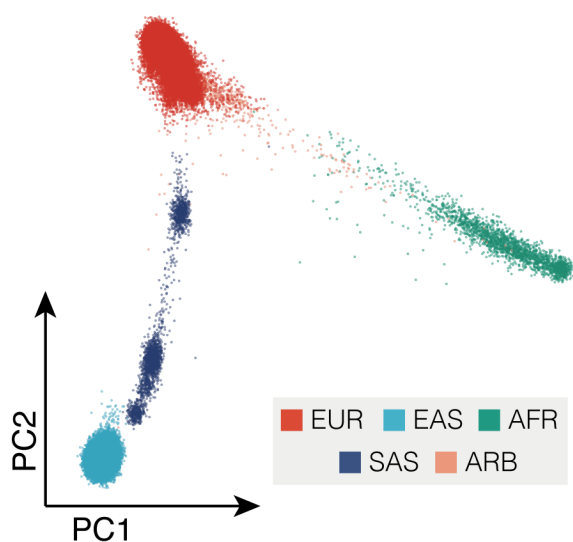
**a**



- ▶ Variant discovery (34 novel loci)
- ▶ Fine-mapping
- ▶ Population-specific + heterogeneous signals
- ▶ Distribution of causal variants
- ▶ Polygenic risk score

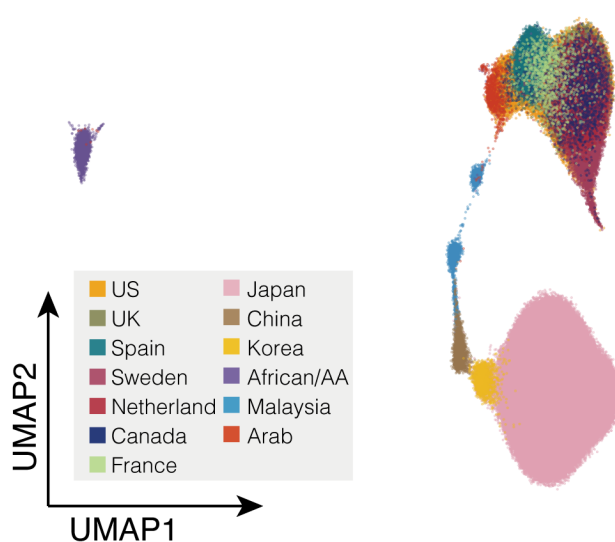
**b**

PCA



**c**

UMAP



511

512

513

514

515

516

517

518

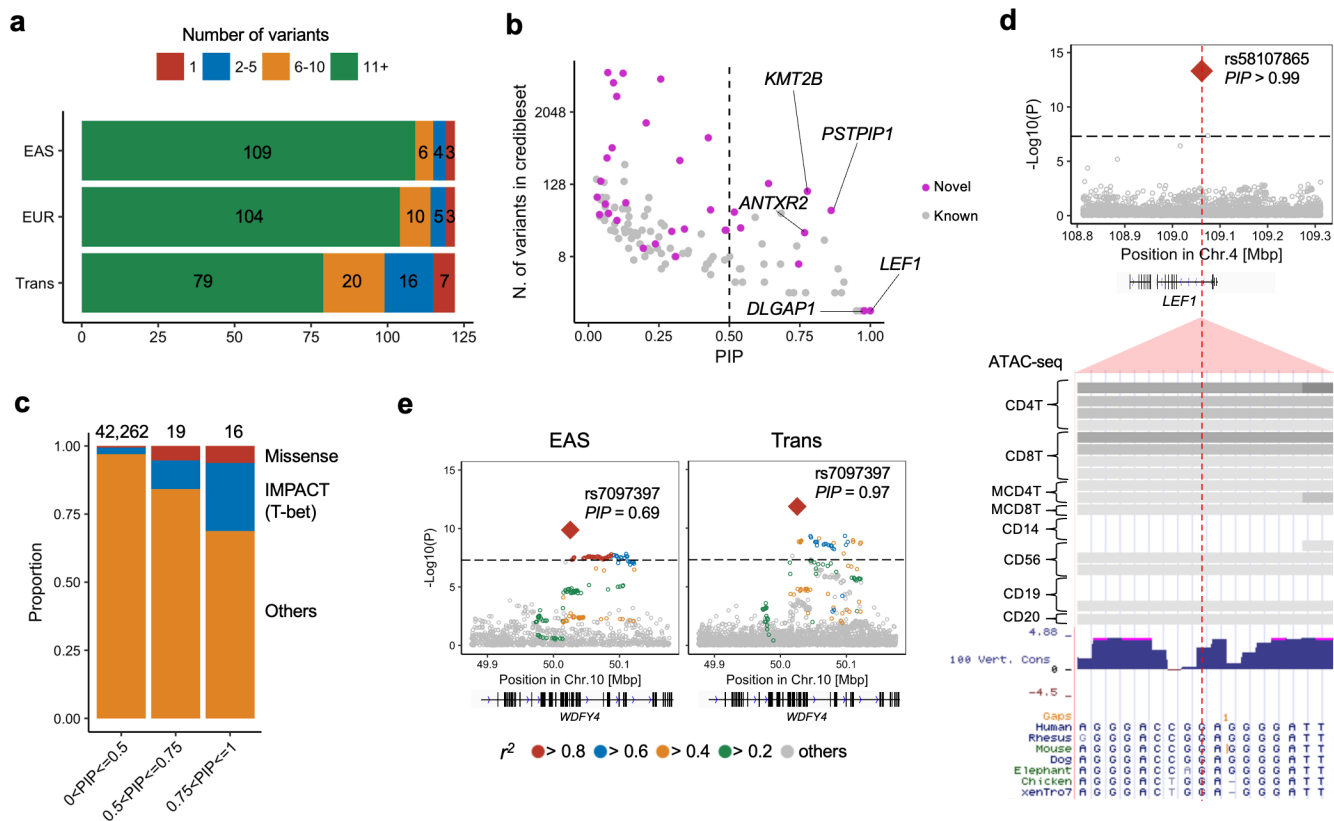
519

**Figure 1. Diverse ancestral background in this GWAS participants.**

**(a)** Study design of this GWAS. The total number of cases and controls are provided.

**(b)** PCA plot of all GWAS samples. We projected each individual's imputed genotype into a PC space which was calculated using all individuals in 1KG Phase3. The samples are colored by its ancestry group.

**(c)** UMAP plots of all GWAS samples. We conducted UMAP analysis using their top 20 PC scores. The samples in a cohort were colored by the country/region-level group of that cohort (**Supplementary Table 1**). When a cohort recruited participants from multiple countries, we did not plot its samples.



520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

**Figure 2. Fine-mapping analysis identified candidates of causal variants.**

(a) Among 122 autosomal loci analyzed, we counted the number of loci whose 95% credible set size was in a specified range. The results from EAS-, EUR- and trans-ancestry GWAS are provided.

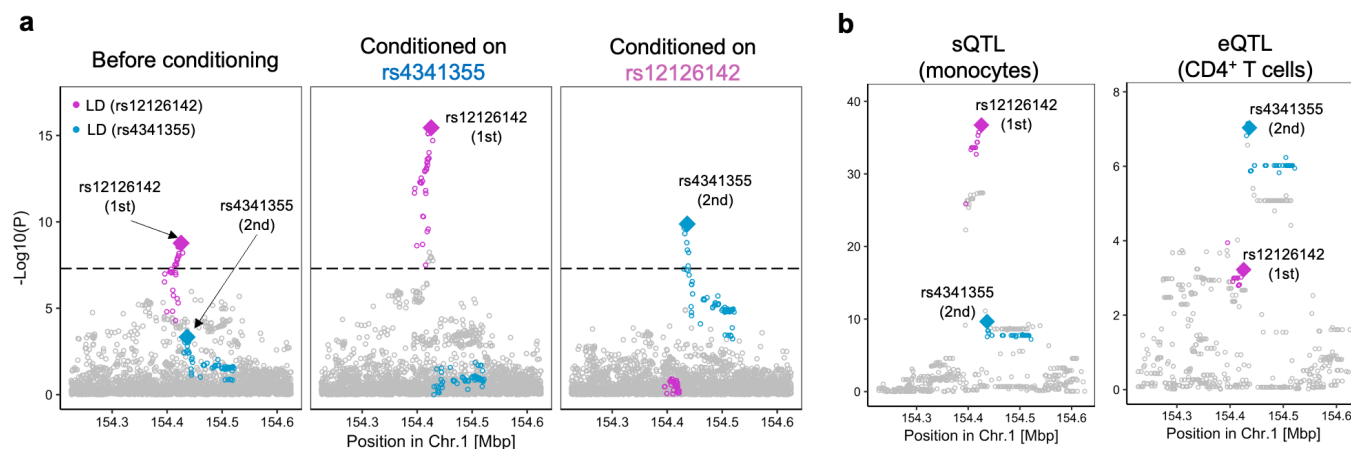
(b) The *PIP* of the lead variant and the size of 95% credible set at the 122 autosomal loci analyzed. The name of novel loci whose *PIP* was greater than 0.75 are labeled. We used trans-ancestry GWAS results.

(c) In each range of *PIP* (the total number of variants were provided on the top), we calculated the proportion of non-synonymous variants or variants with high IMPACT score ( $CD4^+$  T cell T-bet annotation > 0.5).

(d) A fine-mapped variant at the *LEF1* locus within  $CD4^+$  T cell specific open chromatin regions. *P* values of trans-ancestry GWAS are provided with dense view of immune cell ATAC-seq data (density indicate the read coverage) and vertebrate conservation data from UCSC genome browser (<http://genome.ucsc.edu>).

(e) *P* values in the *WDFY4* locus in EAS- and trans-ancestry GWAS. We provide  $r^2$  between each variant and the lead variant (*rs7097397*) by different colors. For trans-ancestry GWAS, we used intersection of LD variants in EUR and EAS ancestries.





537

538

539

540

541

542

543

544

545

546

547

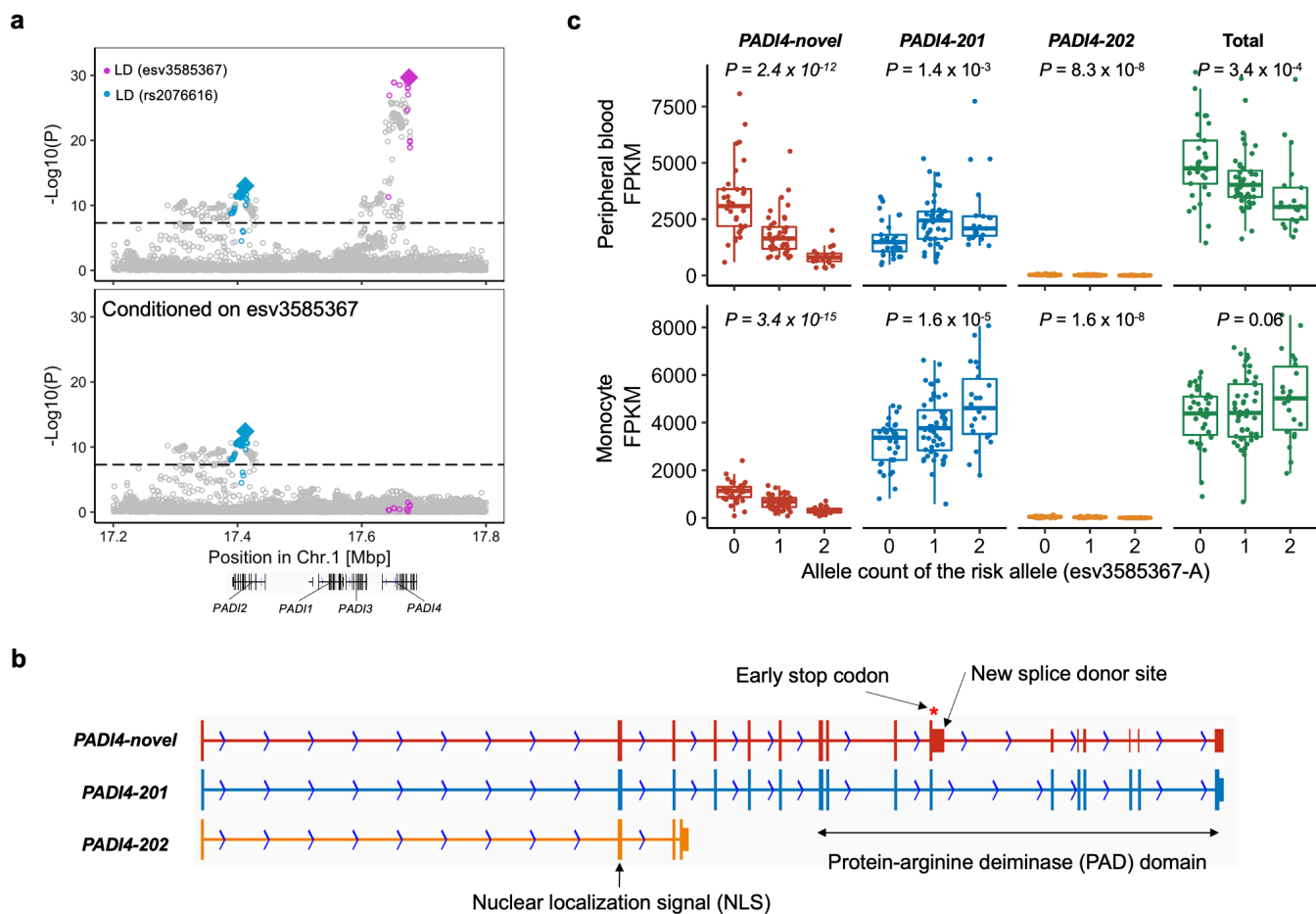
**Figure 3. Splicing and total expression of *IL6R* jointly contribute to RA risk.**

(a) The 1<sup>st</sup> lead variant (rs12126142; red) and the 2<sup>nd</sup> lead variant (rs4341355; blue) are mutually attenuating their signals (controlling the effect of the other increased their signals). Conditional analysis was conducted in each cohort and the results were meta-analyzed using the inverse-variance weighted fixed effect model. We used trans-ancestry GWAS results.

(b)  $P$  values of sQTL signals of *IL6R* (phenotype id: ENSG00000160712.8.17\_154422457 in Blueprint dataset) and eQTL signals of *IL6R* (total expression of *IL6R*).

We highlighted variants in LD with the lead variant by red or blue ( $r^2 > 0.6$  in both EUR and EAS ancestries).





548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

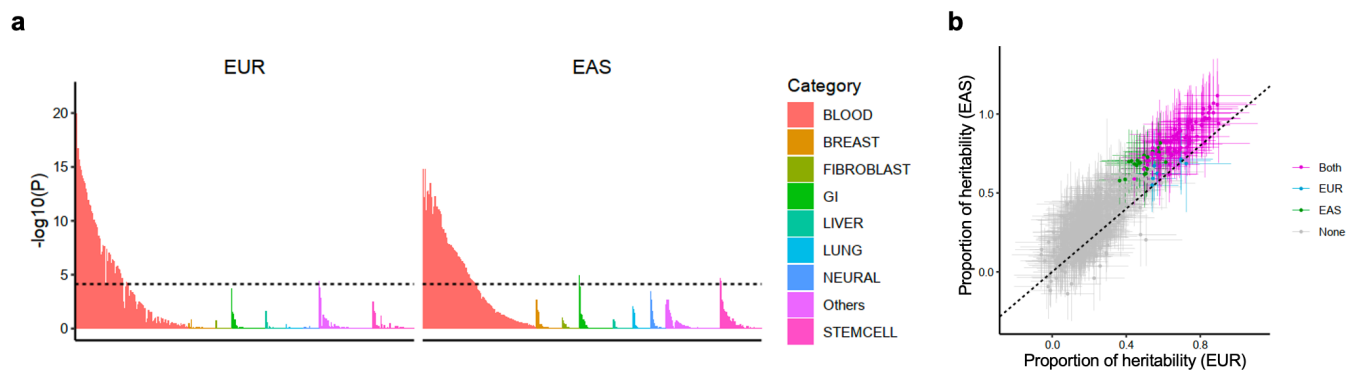
565

#### Figure 4. Splicing of *PADI4* contributes to RA risk.

(a) Conditional analyses identified two independent associations at the *PADI4* locus: esv3585367 (red) and rs2076616 (blue). We used trans-ancestry GWAS results. We highlighted variants in LD with the lead variant by red or blue ( $r^2 > 0.6$  in both EUR and EAS ancestries).

(b) A novel *PADI4* splice isoform confirmed by long-read sequencing datasets. *PADI4-novel*, a novel isoform we identified. *PADI4-201*, a functional isoform. *PADI4-novel* has an elongation of exon 10 which leads to an early stop codon and a truncated PAD domain at the C-terminus. PAD domain is an essential catalytic domain<sup>46</sup>, and highly conserved across other *PADI* genes.

(c) The total expression and the expression of three isoforms of *PADI4* were plotted with the imputed dosages of the risk allele (esv3585367-A). The isoform structures were shown in (b). We analyzed a RNA-seq dataset of 105 Japanese healthy individuals reported in a previous study<sup>5</sup>. We used peripheral blood leukocytes (neutrophils are its main component) and monocytes, both have high *PADI4* expression levels.  $P$  values from linear regression are provided. Within each boxplot, the horizontal lines reflect the median, the top and bottom of each box reflect the interquartile range (IQR), and the whiskers reflect the maximum and minimum values within each grouping no further than 1.5 x IQR from the hinge.



566

567

568

569

570

571

572

573

574

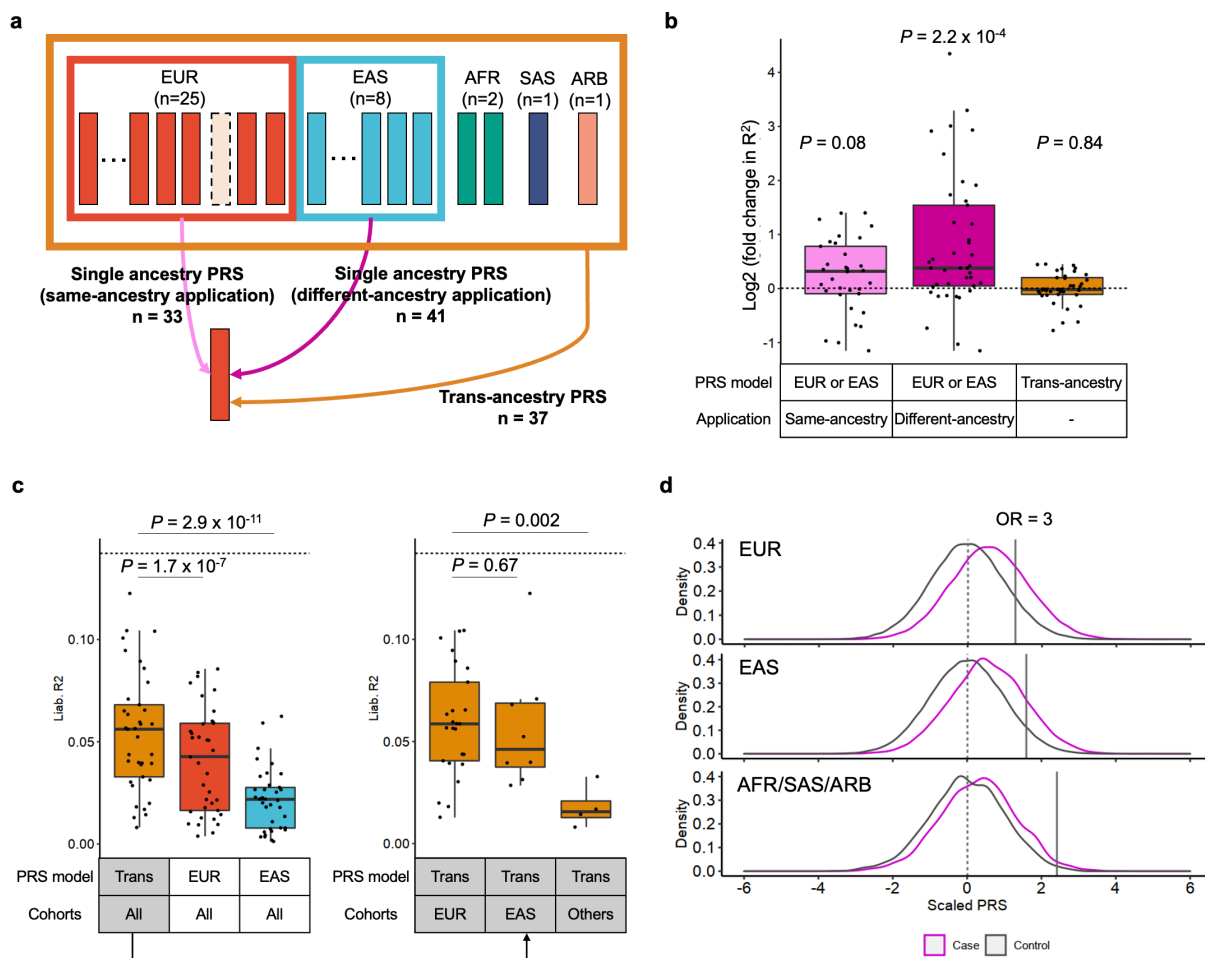
575

576

**Figure 5. S-LDSC analysis suggested similar causal variant distributions in EUR- and EAS-GWAS.**

(a) EUR- and EAS-GWAS results were analyzed by S-LDSC using 707 IMPACT annotations.  $P$  value indicates significance of non-zero tau (per variant heritability) of each annotation. Each annotation was colored by its cell type category. Horizontal dashed line indicates Bonferroni-corrected  $P$  value threshold ( $0.05/707 = 7.1 \times 10^{-5}$ ).

(b) The estimate and its 95% confidence interval of the heritability proportion explained by top 5% of IMPACT annotations. When a heritability enrichment was significant ( $P < 0.05/707 = 7.1 \times 10^{-5}$ ), that annotation was colored by the type of GWAS ("Both" indicates the annotation was significant both in EUR- and EAS-GWAS).



**Figure 6. Functional annotation and trans-ancestry GWAS improved PRS performances.**

**(a)** The strategy of our PRS analysis. We utilized PRS based on three GWAS settings; two single-ancestry PRS (EUR- and EAS-PRS) and a trans-ancestry PRS. Single ancestry PRS had two types of applications; same-ancestry application of PRS (EUR-PRS applied to EUR cohorts or EAS-PRS applied to EAS cohorts) and different-ancestry application of PRS (EUR-PRS applied to non-EUR cohorts or EAS-PRS applied to non-EAS cohorts). When we applied a PRS model to a cohort included in a GWAS setting, we re-conducted the meta-analysis excluding that cohort to avoid overlapped samples. The numbers of each application are provided.

**(b)** The improvements in PRS performance (*pseudo-R*<sup>2</sup>) by CD4<sup>+</sup> T cell T-bet IMPACT annotation. Fold change indicates *R*<sup>2</sup> in functionally-informed PRS divided by *R*<sup>2</sup> in the standard PRS. We compared three conditions: same-ancestry application of single ancestry PRS (n=33), different-ancestry application of single ancestry PRS (n=41), and trans-ancestry PRS (n=37). One-sided sign test *P* value is provided.

**(c)** The performance (liability-scale *R*<sup>2</sup>) of three different PRS models. The results of three PRS models in all cohorts are shown on the left panel. The results of trans-ancestry PRS in three cohort groups are shown on the right panel. The differences in *R*<sup>2</sup> were assessed by Wilcoxon test. In all conditions, CD4<sup>+</sup> T cell T-bet IMPACT annotation was utilized to select variants.

**(d)** PRS distribution differences between case and controls. Trans-ancestry PRS with CD4<sup>+</sup> T cell T-bet IMPACT annotation was used. In each cohort, PRS was scaled using mean and SD of the control samples, and individual level data were merged across cohorts in an ancestral group. For a given PRS value at the right tail of PRS distribution, we compared the case-control ratios between individuals whose PRS is higher than that value and individuals whose PRS is lower than that value, and calculated the odds ratio (OR). The PRS values with OR = 3 are shown by solid vertical lines.

Within each boxplot, the horizontal lines reflect the median, the top and bottom of each box reflect the interquartile range (IQR), and the whiskers reflect the maximum and minimum values within each grouping no further than 1.5 x IQR from the hinge.

## 603 **Methods**

### 604 **Study participants**

605 We included 35,871 RA patients and 240,149 control individuals of EUR, EAS, AFR, SAS, and ARB  
606 ancestry from 37 cohorts (**Supplementary Table 1**). All RA cases fulfilled the 1987 American College of  
607 Rheumatology (ACR) criteria<sup>72</sup> or the 2010 ACR/the European League Against Rheumatism criteria<sup>73</sup>, or  
608 were diagnosed with RA by a professional rheumatologist. Among 35,871 patients, seropositivity status  
609 was available for 31,963; 27,448 were seropositive and 4,515 were seronegative (**Supplementary Table**  
610 **1**). We defined seropositivity as the presence of rheumatoid factor or anti-citrullinated peptide antibodies.  
611 When a seropositive and seronegative GWAS has less than 50 cases, we excluded it from this study  
612 since GWAS with too few samples produces unstable statistics. All cohorts obtained informed consent  
613 from all participants by following the protocols approved by their institutional ethical committees. We  
614 have complied with all relevant ethical regulations.

615

### 616 **Genotyping and imputation**

617 Genotyping platform and all quality control (QC) parameters of each cohort were provided in  
618 **Supplementary Table 1**. For QC of samples, we excluded those with (i) low sample call rate, (ii) closely  
619 related individuals, and (iii) outliers in terms of ancestries identified by PCA using the genotyped samples  
620 and all 1KG Phase 3 samples. Since 1KG Phase 3 does not include ARB samples, we did not exclude  
621 individuals in ARB cohort based on ancestral outliers. For QC of genotypes, we excluded variants  
622 meeting any of the following criteria: (i) low call rate, (ii) low MAF, and (iii) low  $P$  value for Hardy  
623 Weinberg equilibrium (HWE). Post-QC genotype data of each cohort was pre-phased using Shapeit2 or  
624 Eagle2 software. For EUR, AFR, and ARB cohorts, we conducted imputation with Minimac3 or Minimac4  
625 software using the 1KG Phase 3 reference panel. For EAS and SAS cohorts, we conducted imputation  
626 with Minimac3 or Minimac4 using a reference panel which were generated by merging 1KG Phase 3  
627 panel and whole genome sequence (WGS) data of each population (**Supplementary Table 1**); we used  
628 WGS data of 1,037 Japanese<sup>74</sup>, 89 Korean<sup>75</sup>, 7 Chinese<sup>76</sup> or 96 Malaysian individuals<sup>77</sup>. For QC of  
629 imputed genotype data, we excluded low imputation quality variants (imputation  $r^2 < 0.3$ ) from each

630 cohort, excluded variants with minor allele count less than ten in the reference panel, and then we  
631 included variants which passed QC in at least five cohorts; we finally included 20,990,826 autosomal  
632 variants and 736,614 X chromosome variants. The genomic coordinate is according to GRCh37 in all  
633 analyses.

634

### 635 **PCA and UMAP using all GWAS participants**

636 To assess the ancestral background diversity of all GWAS participants, we projected them into the same  
637 PC space based on their imputed genotype data. From variants which passed QC criteria in all 37  
638 cohorts (imputation  $r^2 \geq 0.3$ ), we first identified 12,196 independent imputed variants ( $r^2 < 0.2$  in EUR  
639 samples of 1KG Phase 3). Due to data access restrictions, we were not able to transfer raw imputed  
640 genotype data across different institutes, and hence we were not able to conduct one PCA using all  
641 individuals. Therefore, we first conducted PCA using these variants and all 1KG Phase3 samples, and  
642 calculated the loadings of each variant for top 20 PCs. We then calculated each individual's PC scores  
643 using these loadings and imputed dosage of our GWAS samples. We further conducted UMAP by umap  
644 package in R using these top 20 PC scores ( $n\_neighbors = 30$  and  $min\_dist = 0.8$ ).

645

### 646 **Genome-wide association analysis**

647 We conducted GWAS in each cohort by a logistic regression model using PLINK2 software. We included  
648 age, sex, and genotype PCs within each cohort as covariates (details of covariates were provided in  
649 **Supplementary Table 1**). We then conducted meta-analysis using all cohorts by the inverse-variance  
650 weighted fixed effect model as implemented in METAL (trans-ancestry GWAS). For ancestries with  
651 multiple cohorts, we similarly conducted meta-analysis within each ancestry (EUR-, EAS-, and AFR-  
652 GWAS). When the seropositivity status was available, we also conducted GWAS only using seropositive  
653 RA samples and controls. We defined a locus as a genomic region within  $\pm 1$  Mb from the lead variant,  
654 and we considered a locus as novel when it did not include any variants previously reported for RA. For  
655 non-RA autoimmune diseases (systemic lupus erythematosus, systemic sclerosis, Sjögren's syndrome,  
656 dermatomyositis, juvenile dermatomyositis, and polymyositis), we used  $\pm 0.5$  Mb window from the lead

657 variant. We defined reported variants as significant variants ( $P < 5 \times 10^{-8}$ ) reported in the GWAS Catalog  
658 (<https://www.ebi.ac.uk/gwas/>; e104\_r2021-10-06) and those reported in previous literatures which we  
659 searched manually. Since we need a unique analytical strategy for the MHC locus, we excluded the  
660 MHC region (chr6:25Mb-35Mb) from this study, which will be reported in an accompanying project.

661 We performed stepwise conditional analysis within  $\pm 1$  Mb from the lead variant. We conducted  
662 the same logistic regression model but including the dosages of the lead variants (index variants in the  
663 first round of conditional analysis) as covariates in each cohort; when the lead variants did not exist in  
664 post-QC imputed genotype data of a cohort (imputation  $r^2 \geq 0.3$ ), we exclude that cohort from the  
665 analysis. We then conducted meta-analysis using the same strategy, and identified the 2<sup>nd</sup> lead variant.  
666 We repeated these processes by sequentially adding the identified lead variants as covariates until we  
667 did not detect any significant associations ( $P < 5 \times 10^{-8}$ ).

668

### 669 **Estimation of heritability and bias in GWAS results**

670 We estimated heritability and confounding bias in EUR- and EAS-GWAS results with S-LDSC (version  
671 1.0.0) using the baselineLD model (version 2.1). For EUR-GWAS, we utilized LD scores calculated in  
672 EUR samples in 1KG Phase3. For EAS-GWAS, we utilized LD scores calculated in EAS samples in 1KG  
673 Phase3. Since LDSC required a large sample size in GWAS (typically  $> 5K$  individuals), we restricted  
674 this analysis to EUR- and EAS-GWAS. We estimated prevalence of RA was 0.5% in both ancestries.

675

### 676 **Fine-mapping analysis**

677 We conducted fine-mapping analysis using approximate Bayesian factor (ABF) and constructed 95%  
678 credible set for each significant locus<sup>35</sup>. We used trans-ancestry GWAS results. We included all the 122  
679 autosomal loci ( $P < 5.0 \times 10^{-8}$ ). We calculated ABF of each variant according to equation (1):

680

$$ABF = \sqrt{\frac{SE^2}{SE^2 + \omega}} \exp \left[ \frac{\omega \beta^2}{2SE^2(SE^2 + \omega)} \right]$$

681 where  $\beta$  and  $SE$  are the variant's effect size and standard error, respectively;  $\omega$  denotes the prior  
682 variance in allelic effects (we empirically set this value to be  $0.04$ )<sup>78</sup>. For each locus, we calculated  
683 posterior inclusion probability (PIP) of variant  $k$  according to equation (2):

$$684 \quad PIP_k = \frac{ABF_k}{\sum_j ABF_j}$$

685 where  $j$  denotes all the variants included in the locus. We sorted all variants in order of decreasing PIP  
686 and constructed 95% credible set including variants from the top PIP until the cumulative PIP reached  
687 0.95. When we compared the fine-mapping resolution across different GWAS setting, we also applied  
688 this fine-mapping strategy to EUR-GWAS and EAS-GWAS for all the 122 autosomal loci.

689

### 690 **Functional interpretation of fine-mapped variants**

691 We quantified the enrichment of the 95% credible set variants at the 113 autosomal loci within ATAC-seq  
692 peaks in 18 hematopoietic populations using gchromVAR software<sup>37</sup>. We utilized the default parameters  
693 and ATAC-seq data processed by the developers. To access the specificity of a given ATAC-seq peak,  
694 we first normalized the read count of that peak in all 18 hematopoietic populations (each peak's read  
695 count was divided by the total read counts, scaled by 1000,000, added 1 as an offset value, and log2-  
696 transformed), and transformed these 18 normalized counts into Z-scores.

697

### 698 **Functional interpretation of associated variants**

699 We inferred the possible molecular consequences of all 148 variants detected in this study. We first  
700 focused on coding variants in LD with the lead variants in this GWAS ( $r^2 > 0.6$  both in EUR and EAS  
701 samples in 1KG Phase 3; when the lead variant is monomorphic in one ancestry, we only utilized the  
702 other ancestry). To annotate coding variants, we used ANNOVAR and assessed their potential impacts  
703 on protein function; we reported SIFT and Polyphen2 (HDIV) scores. To interpret their effects on gene  
704 regulation, we tested colocalization of our GWAS signals and eQTL or sQTL signals using coloc  
705 software<sup>40</sup>. We analyzed eQTL and sQTL results from Blueprint consortium database (CD4<sup>+</sup> T cells,  
706 monocytes, and neutrophils) and eQTL results from v7 GTEx project database (48 tissues)<sup>39,49</sup>. Since  
707 coloc assumes GWAS and QTL signals are obtained from the same ancestry group, we only used EUR-



708 GWAS results for this analysis. We defined GWAS and QTL signals are colocalized when the posterior  
709 probability estimated by coloc software  $> 0.7$ .

710

### 711 **Capture RNA-seq of *PADI4* isoforms**

712 We obtained total RNAs from THP-1 cells after stimulation with PMA for 72h, which induces the  
713 expression of *PADI4*<sup>79</sup>. We reverse-transcribed the RNA (10 ng) into cDNAs with Smart-seq2 primers<sup>80</sup>,  
714 and then amplified them by 10 cycles of polymerase chain reaction. We hybridized *PADI4* isoforms with  
715 xGen Lockdown Probes (5'-biotinylated 120-mer DNA probes synthesized by Integrated DNA  
716 Technologies) designed for all exons of *PADI4* main isoform. We captured the hybridized cDNAs with  
717 streptavidin-conjugated magnetic beads and then sequenced them with MinION sequencer using LSK-  
718 109 kit (Oxford Nanopore Technologies). We analyzed the sequenced reads with FLAIR  
719 (<https://github.com/BrooksLabUCSC/flair>).

720 We then performed sQTL analysis targeting *PADI4* using the eQTL data of peripheral blood  
721 subsets<sup>5</sup>. We reassembled and quantified the RNA-seq reads for *PADI4* isoforms including the newly  
722 discovered isoform using Cufflinks (<http://cole-trapnell-lab.github.io/cufflinks/>). We calculated the isoform  
723 ratio by dividing each isoform expression (FPKM) over total isoform expression. We used QTLtools  
724 (<https://qtltools.github.io/qtltools/>) for association testing.

725

### 726 **Stratified linkage disequilibrium score regression**

727 We conducted stratified LD score regression (S-LDSC) to partition heritability. For this analysis, we used  
728 707 cell-type-specific IMPACT annotations and 396 histone mark annotations<sup>65,81</sup>. IMPACT regulatory  
729 annotations were created by aggregating 5,345 epigenetic datasets to predict binding patterns of 142  
730 transcription factors across 245 cell types. We computed annotation-specific LD scores using the EUR  
731 samples in 1KG Phase3 to analyze EUR-GWAS results. Similarly, we used EAS samples in 1KG Phase3  
732 to analyze EAS-GWAS results. We estimated heritability enrichment of each annotation, while controlling  
733 for the 53 categories of the full baseline model. When we controlled the effect of an annotation, we



734 conducted the same S-LDSC analysis but additionally including that annotation in a single model. We  
735 excluded variants in the MHC region (chr6:25 Mb-35 Mb).

736

### 737 **Trans-ancestry comparison of genetic signals**

738 We first sought to compare the effect size estimates among GWAS results from each ancestry at the  
739 lead variants. However, the lead variants are not always the causal variants, and hence we restricted our  
740 targets to fine-mapped lead variants ( $PIP > 0.5$ ). In addition, we excluded rare variants from this analysis  
741 because the effect sizes could not be accurately estimated for rare variants ( $MAF < 0.01$  in either of the  
742 major ancestries in 1KG Phase 3). We finally included 30 fine-mapped variants for this analysis.

743 We next obtained trans-ancestry genetic-effect correlation using Popcorn software (version 1.0)<sup>62</sup>.  
744 We used summary statistics of EUR- and EAS-GWAS, and selected association statistics from variants  
745 with at least non-missing genotype from 5,000 individuals. We also excluded the MHC region from the  
746 analysis because of its complex LD structure. Using these post-QC summary statistics, we calculated the  
747 trans-ancestry genetic-effect correlation between EUR and EAS with precomputed cross-ancestry scores  
748 for EUR and EAS 1000 Genomes ancestries provided by the authors.

749

### 750 **Polygenic risk score**

751 We used the pruning and thresholding method to calculate PRS in this study. We developed PRS  
752 models with six different conditions using combinations of two components; i) two variant selection  
753 settings used for PRS ((a) all variants or (b) variants within top 5% of the CD4<sup>+</sup> T cell T-bet IMPACT  
754 annotation) and ii) three GWAS settings used for PRS (we used trans-ancestry, EUR-, or EAS-GWAS  
755 and refer PRS based on each of them as trans-ancestry, EUR-, or EAS-PRS, respectively). We designed  
756 our study so that the samples used in constructing PRS be independent from the samples in validation.  
757 When we evaluated the PRS performance in a given cohort, we re-conducted GWAS meta-analysis  
758 excluding that cohort to develop PRS models (**Figure 6a**). Before pruning, we removed rare variants  
759 from three GWAS results to reduce unstable effect estimates in PRS ( $MAF < 0.01$  in EUR samples of  
760 1KG Phase3 for EUR- and trans-ancestry GWAS; and  $MAF < 0.01$  in EAS samples of 1KG Phase3 for

761 EAS-GWAS). We also restricted this analysis to the variants which exist both in the GWAS results and  
762 post-QC imputed genotype of a cohort for which we apply PRS; and then we selected variants based on  
763 IMPACT annotation or utilized all variants (as described above). To LD-prune variants ( $r^2 < 0.2$ ), we used  
764 haplotype information in EUR samples of 1KG Phase3 for EUR- and trans-ancestry GWAS and EAS  
765 samples of 1KG Phase3 for EAS-GWAS. For each of six conditions, we used 10 different  $P$  value  
766 thresholds: 0.1, 0.03, 0.01, 0.003, 0.001,  $3.0 \times 10^{-4}$ ,  $1.0 \times 10^{-4}$ ,  $3.0 \times 10^{-5}$ ,  $1.0 \times 10^{-5}$ , and  $5.0 \times 10^{-8}$ ; we  
767 thus ended up having 60 different PRS models (6 conditions  $\times$  10  $P$  value thresholds). We applied these  
768 60 PRS models to 37 cohorts and applied a logistic regression model using per-individual PRS including  
769 the same covariates as used in GWAS; we evaluated PRS performances by Nagelkerke  $R^2$ . In each of  
770 six PRS conditions, we selected the  $P$  value threshold with the largest average Nagelkerke  $R^2$ , and used  
771 this  $P$  value threshold for the following analyses.

772 To discuss the PRS distribution in an ancestry, we first calculated PRS in each cohort using a  
773 specified condition; we next scaled those PRS values using the mean and the standard deviation of the  
774 PRS only of the control samples in that cohort; and we then merged PRS values across cohorts in an  
775 ancestry. We approximated the PRS distribution in general population by using that in control samples.  
776 For a given PRS value (at the right tail of PRS distribution), we compared the case-control ratios  
777 between individuals whose PRS is higher than that value and individuals whose PRS is lower than (or  
778 equal to) that value and calculated the odds ratio; and we then identified the minimum PRS value which  
779 showed odds ratio larger than three.

780

781 **Data availability:** The summary statistics and the PRS model with the best performance are publicly  
782 available at the following link:

783 [https://data.cyverse.org/dav-anon/iplant/home/kazuyoshiishigaki/ra\\_gwas/ra\\_gwas-10-28-2021.tar](https://data.cyverse.org/dav-anon/iplant/home/kazuyoshiishigaki/ra_gwas/ra_gwas-10-28-2021.tar).

784 The codes are available at our website:

785 [https://github.com/immunogenomics/RA\\_GWAS](https://github.com/immunogenomics/RA_GWAS).

786

787 **Acknowledgement:**

788 We thank the Director of Health Malaysia for supporting the work described in the South Asian (SAS):  
789 the Malaysian Epidemiological Investigation of Rheumatoid Arthritis (MyEIRA) study. The MyEIRA study  
790 was funded by grant from Ministry of Health Malaysia (NMRR-08-820-1975) and the Swedish National  
791 Research Council (DNR-348-2009-6468). The GENRA study and the CARDERA genetics cohort  
792 genotyping were funded by Versus Arthritis (Grant Reference 19739 to ICS). The Nurses' Health Study  
793 (NHS cohort) is funded by the NIH (R01 AR049880, UM1 CA186107, R01 CA49449, U01 CA176726,  
794 and R01 CA67262). The Swedish EIRA study was supported by the Swedish Research Council (to LK,  
795 LP and LA). SS was in part supported by The Mochida Memorial Foundation for Medical and  
796 Pharmaceutical Research, Kanae Foundation for the promotion of medical science, Astellas Foundation  
797 for Research on Metabolic Disorder, The JCR Grant for Promoting Basic Rheumatology, and Manabe  
798 Scholarship Grant for allergic and rheumatic diseases. ICS is funded by the NIHR Advanced Research  
799 Fellowship (Grant Reference NIHR300826). The views expressed are those of the author(s) and not  
800 necessarily those of the NIHR or the Department of Health and Social Care. KAS is supported by the  
801 Sherman Family Chair in Genomic Medicine and by a Canadian Institutes for Health Research  
802 Foundation grant (FDN 148457) and grants from the Ontario Research Fund (RE-09-090) and Canadian  
803 Foundation for Innovation(33374). S.Bae is supported by Basic Science Research Program through the  
804 NRF funded by the Ministry of Education (NRF-2021R1A6A1A03038899). RPK and JCE are funded by  
805 NIH (UL1 TR003096). CML is part-funded by the National Institute for Health Research (NIHR) Maudsley  
806 Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College  
807 London. T.Arayssi was partially supported by the National Priorities Research Program (grant 4-344-3-

808 105 from the Qatar National Research Fund, a member of Qatar Foundation). M.Kerick and JM are  
809 funded by Rheumatology Cooperative Research Thematic Network programme RD16/0012/0013 from  
810 the Instituto de Salud Carlos III (ISCIII, Spanish Ministry of Science and Innovation). YO is funded by  
811 JSPS KAKENHI (19H01021, 20K21834), AMED (JP21km0405211, JP21ek0109413, JP21ek0410075,  
812 JP21gm4010006, and JP21km0405217), JST Moonshot R&D (JPMJMS2021, JPMJMS2024), Takeda  
813 Science Foundation, Bioinformatics Initiative of Osaka University Graduate School of Medicine, Osaka  
814 University.

815

816 **Author contributions:**

817 K.Ishigaki, SS, C.Terao, YO, and S.Raychaudhuri conceived and designed the study. K.Ishigaki wrote  
818 the manuscript with critical inputs from SS, C.Terao, YL, YO, and S.Raychaudhuri. K.Ishigaki conducted  
819 meta-analysis and all GWAS downstream analyses with the help of SS, C.Terao, T.Amariuta, YL, YO,  
820 and S.Raychaudhuri. K.Yamaguchi and Y.Kochi conducted *PADI4* long-read sequencing and *PADI4*  
821 sQTL analysis. M.Koido, KT, Y.Kamatani, and C.Terao contributed to construction of population-specific  
822 reference panel. K.Ishigaki, SS, C.Terao, KS, VAL, ICS, SV, DP, JB, GX, JZ, CA, EK, RJC, KAS,  
823 M.Kerick, FM, M.Traylor, CML, HX, RS, T.Arayssi, JM, LK, YO, and S.Raychaudhuri conducted GWAS.  
824 C.Terao, CL.Too, VAL, SV, M.Takahashi, XW, LL, FL, DP, AB, GO, JB, SM, KPL, RJC, EWK, K.Matsuo,  
825 FM, SE, HX, K.Ikari, PKG, LP, YO, and S.Raychaudhuri contributed to genotyping experiments. C.Terao,  
826 KS, CL.Too, VAL, ICS, SV, KO, AM, MH, HI, M.Hammoudeh, SA, BKM, HH, HB, IWU, XW, LL, FL, DP,  
827 AB, GO, SMV, AJM, SH, HT, ET, AS, YM, Kenichi Yamamoto, SM, GX, JZ, CA, EK, GW, IvH, JC, KPL,  
828 RJC, HL, S.Bang, KAS, Nd, LA, S.Rantapää-Dahlqvist, EWK, S.Bae, RPK, JCE, XM, TH, PD, MS,  
829 M.Kerick, JCD, The Biobank Japan Project, K.Matsuda, K.Matsuo, TM, FM, KF, YT, AK, CML, SE, HX,  
830 RS, T.Arayssi, K.Ikari, M.Harigai, PKG, Kazuhiko Yamamoto, SLB, LP, JM, LK, YO, and S.Raychaudhuri  
831 contributed to collection of samples and management of genotype data and clinical information.

832

833 **Competing interests:**

834 The authors declare no competing interests.



836 **Reference:**

- 837 1. Ajejanova, S. & Huizinga, T. W. J. Seronegative and seropositive RA: alike but different? *Nat.*  
838 *Rev. Rheumatol.* 2014 111 **11**, 8–9 (2014).
- 839 2. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*  
840 **506**, 376–381 (2014).
- 841 3. Ishigaki, K. *et al.* Large-scale genome-wide association study in a Japanese population identifies  
842 novel susceptibility loci across different diseases. *Nat. Genet.* **52**, 669–679 (2020).
- 843 4. MacGregor, A. J. *et al.* Characterizing the quantitative genetic contribution to rheumatoid arthritis  
844 using data from twins. *Arthritis Rheum.* **43**, 30–37 (2000).
- 845 5. Ishigaki, K. *et al.* Polygenic burdens on cell-specific pathways underlie the risk of rheumatoid  
846 arthritis. *Nat. Genet.* **49**, 1120–1125 (2017).
- 847 6. Westra, H. J. *et al.* Fine-mapping and functional studies highlight potential causal variants for  
848 rheumatoid arthritis and type 1 diabetes. *Nat. Genet.* **50**, 1366–1374 (2018).
- 849 7. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants.  
850 *Nat. Genet.* **45**, 124–30 (2013).
- 851 8. Lamparter, D., Marbach, D., Rueedi, R., Kutalik, Z. & Bergmann, S. Fast and Rigorous  
852 Computation of Gene and Pathway Scores from SNP-Based Summary Statistics. *PLoS Comput.*  
853 *Biol.* **12**, 1–20 (2016).
- 854 9. Ishigaki, K. *et al.* HLA autoimmune risk alleles restrict the hypervariable region of T cell receptors.  
855 *medRxiv* (2020). doi:<https://doi.org/10.1101/2020.11.08.20227983>
- 856 10. Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide  
857 association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
- 858 11. Asgari, S. *et al.* A positively selected FBN1 missense variant reduces height in Peruvian  
859 individuals. *Nature* **582**, 234–239 (2020).
- 860 12. Estrada, K. *et al.* Association of a low-frequency variant in HNF1A with type 2 diabetes in a latino  
861 population the SIGMA Type 2 Diabetes Consortium. *JAMA* **311**, 2305–2314 (2014).
- 862 13. Moltke, I. *et al.* A common Greenlandic TBC1D4 variant confers muscle insulin resistance and

- 863 type 2 diabetes. *Nature* **512**, 190–193 (2014).
- 864 14. Koyama, S. *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct  
865 and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020).
- 866 15. Chen, M. H. *et al.* Trans-ethnic and Ancestry-Specific Blood-Cell Genetics in 746,667 Individuals  
867 from 5 Global Populations. *Cell* **182**, 1198–1213.e14 (2020).
- 868 16. Huang, H. *et al.* Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*  
869 **547**, 173–178 (2017).
- 870 17. Laufer, V. A. *et al.* Genetic influences on susceptibility to rheumatoid arthritis in African-  
871 Americans. *Hum. Mol. Genet.* **28**, 858–874 (2019).
- 872 18. Martin, A. R. *et al.* Clinical use of current polygenic risk scores may exacerbate health disparities.  
873 *Nat. Genet.* **51**, 584–591 (2019).
- 874 19. Ruan, Y. *et al.* Improving Polygenic Prediction in Ancestrally Diverse Populations. *medRxiv*  
875 2020.12.27.20248738 (2021). doi:10.1101/2020.12.27.20248738
- 876 20. Márquez-Luna, C. *et al.* Multiethnic polygenic risk scores improve risk prediction in diverse  
877 populations. *Genet. Epidemiol.* **41**, 811–823 (2017).
- 878 21. Leng, R. X. *et al.* Identification of new susceptibility loci associated with rheumatoid arthritis. *Ann.*  
879 *Rheum. Dis.* **79**, 1565–1571 (2020).
- 880 22. Kochi, Y. *et al.* A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with  
881 rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–85 (2005).
- 882 23. Suzuki, A. *et al.* Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine  
883 deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **10**, 520–527 (2003).
- 884 24. Okada, Y. *et al.* Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the  
885 Japanese population. *Nat. Genet.* **44**, 511–516 (2012).
- 886 25. Diogo, D. *et al.* TYK2 protein-coding variants protect against rheumatoid arthritis and  
887 autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits.  
888 *PLoS One* **10**, 1–21 (2015).
- 889 26. Traylor, M. *et al.* Genetic associations with radiological damage in rheumatoid arthritis: Meta-



- 890 analysis of seven genome-wide association studies of 2,775 cases. *PLoS One* **14**, (2019).
- 891 27. Márquez, A. *et al.* Meta-analysis of ImmunoChip data of four autoimmune diseases reveals novel  
892 single-disease and cross-phenotype associations. *Genome Med.* **10**, 97 (2018).
- 893 28. Wei, W. H., Viatte, S., Merriman, T. R., Barton, A. & Worthington, J. Genotypic variability based  
894 association identifies novel non-additive loci DHCR7 and IRF4 in sero-negative rheumatoid  
895 arthritis. *Sci. Rep.* **7**, (2017).
- 896 29. Márquez, A. *et al.* A combined large-scale meta-Analysis identifies COG6 as a novel shared risk  
897 locus for rheumatoid arthritis and systemic lupus erythematosus. *Ann. Rheum. Dis.* **76**, 286–294  
898 (2017).
- 899 30. Bossini-Castillo, L. *et al.* A genome-wide association study of rheumatoid arthritis without  
900 antibodies against citrullinated peptides. *Ann. Rheum. Dis.* **74**, (2015).
- 901 31. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid  
902 arthritis. *Nat. Genet.* **44**, 1336–1340 (2012).
- 903 32. acosta-Herrera, M. *et al.* Genome-wide meta-analysis reveals shared new loci in systemic  
904 seropositive rheumatic diseases. *Ann Rheum Dis* **78**, 311–319 (2019).
- 905 33. Frisell, T. *et al.* Familial risks and heritability of rheumatoid arthritis: Role of rheumatoid factor/anti-  
906 citrullinated protein antibody status, number and type of affected relatives, sex, and age. *Arthritis*  
907 *Rheum.* **65**, 2773–2782 (2013).
- 908 34. Padyukov, L. *et al.* A genome-wide association study suggests contrasting associations in ACPA-  
909 positive versus ACPA-negative rheumatoid arthritis. *Ann. Rheum. Dis.* **70**, 259–265 (2011).
- 910 35. Maller, J. B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases.  
911 *Nat. Genet.* **44**, 1294–1301 (2012).
- 912 36. Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping  
913 Studies. *Am. J. Hum. Genet.* **97**, 260–271 (2015).
- 914 37. Ulirsch, J. C. *et al.* Interrogation of human hematopoiesis at single-cell and single-variant  
915 resolution. *Nat. Genet.* **51**, 683–693 (2019).
- 916 38. Fu, W. *et al.* A multiply redundant genetic switch ‘locks in’ the transcriptional signature of

- 917 regulatory T cells. *Nat. Immunol.* **13**, 972–980 (2012).
- 918 39. Chen, L. *et al.* Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune  
919 Cells. *Cell* **167**, 1398-1414.e24 (2016).
- 920 40. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association  
921 Studies Using Summary Statistics. *PLoS Genet.* **10**, (2014).
- 922 41. Ferreira, R. C. *et al.* Functional IL6R 358Ala Allele Impairs Classical IL-6 Receptor Signaling and  
923 Influences Risk of Diverse Inflammatory Diseases. *PLoS Genet.* **9**, e1003444 (2013).
- 924 42. Okada, Y. *et al.* Significant impact of miRNA–target gene networks on genetics of human complex  
925 traits. *Sci. Reports 2016 61* **6**, 1–9 (2016).
- 926 43. Schellekens, G. A., De Jong, B. A. W., Van Den Hoogen, F. H. J., Van De Putte, L. B. A. & Van  
927 Venrooij, W. J. Citrulline is an essential constituent of antigenic determinants recognized by  
928 rheumatoid arthritis-specific autoantibodies. *J. Clin. Invest.* **101**, 273–281 (1998).
- 929 44. Suzuki, A. *et al.* Decreased severity of experimental autoimmune arthritis in peptidylarginine  
930 deiminase type 4 knockout mice. *BMC Musculoskelet. Disord.* **17**, (2016).
- 931 45. Seri, Y. *et al.* Peptidylarginine deiminase type 4 deficiency reduced arthritis severity in a glucose-  
932 6-phosphate isomerase-induced arthritis model. *Sci. Rep.* **5**, 1–10 (2015).
- 933 46. Arita, K. *et al.* Structural basis for Ca<sup>2+</sup>-induced activation of human PAD4. *Nat. Struct. Mol. Biol.*  
934 **11**, 777–783 (2004).
- 935 47. Nanda, S. K. *et al.* ABIN2 Function Is Required To Suppress DSS-Induced Colitis by a Tpl2-  
936 Independent Mechanism. *J. Immunol.* **201**, 3373–3382 (2018).
- 937 48. Matmati, M. *et al.* A20 (TNFAIP3) deficiency in myeloid cells triggers erosive polyarthritis  
938 resembling rheumatoid arthritis. *Nat. Genet.* **43**, 908–912 (2011).
- 939 49. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213  
940 (2017).
- 941 50. James, E. A. *et al.* Citrulline-Specific Th1 Cells Are Increased in Rheumatoid Arthritis and Their  
942 Frequency Is Influenced by Disease Duration and Therapy. *Arthritis Rheumatol.* **66**, 1712–1722  
943 (2014).

- 944 51. Takayanagi, H. *et al.* RANKL maintains bone homeostasis through c-fos-dependent induction of  
945 interferon- $\beta$ . *Nature* **416**, 744–749 (2002).
- 946 52. Takeuchi, T. *et al.* Effects of the anti-RANKL antibody denosumab on joint structural damage in  
947 patients with rheumatoid arthritis treated with conventional synthetic disease-modifying  
948 antirheumatic drugs (DESIRABLE study): A randomised, double-blind, placebo-controlled phase.  
949 *Ann. Rheum. Dis.* **78**, 899–907 (2019).
- 950 53. Nakatsuka, K., Nishizawa, Y. & Ralston, S. H. Phenotypic characterization of early onset Paget's  
951 disease of bone caused by a 27-bp duplication in the TNFRSF11A gene. *J. Bone Miner. Res.* **18**,  
952 1381–1385 (2003).
- 953 54. Osterberg, P. H. *et al.* Familial expansile osteolysis. A new dysplasia. *J. Bone Jt. Surg. - Ser. B*  
954 **70**, 255–260 (1988).
- 955 55. Guerrini, M. M. *et al.* Human Osteoclast-Poor Osteopetrosis with Hypogammaglobulinemia due to  
956 TNFRSF11A (RANK) Mutations. *Am. J. Hum. Genet.* **83**, 64–76 (2008).
- 957 56. French, D. M. *et al.* WISP-1 is an osteoblastic regulator expressed during skeletal development  
958 and fracture repair. *Am. J. Pathol.* **165**, 855–867 (2004).
- 959 57. Maeda, A. *et al.* WNT1-Induced Secreted Protein-1 (WISP1), a novel regulator of bone turnover  
960 and Wnt signaling. *J. Biol. Chem.* **290**, 14004–14018 (2015).
- 961 58. Zhang, F. *et al.* Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by  
962 integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 (2019).
- 963 59. Ramos, M. I. P. *et al.* Absence of Fms-like tyrosine kinase 3 ligand (Flt3L) signalling protects  
964 against collagen-induced arthritis. *Ann. Rheum. Dis.* **74**, 211–219 (2015).
- 965 60. Saevarsdottir, S. *et al.* FLT3 stop mutation increases FLT3 ligand level and risk of autoimmune  
966 thyroid disease. *Nature* **584**, 619–623 (2020).
- 967 61. Motegi, T. *et al.* Identification of rare coding variants in TYK2 protective for rheumatoid arthritis in  
968 the Japanese population and their effects on cytokine signalling. *Ann. Rheum. Dis.* **78**, 1062–1069  
969 (2019).
- 970 62. Brown, B. C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from

- 971 Summary Statistics. *Am. J. Hum. Genet.* **99**, 76–88 (2016).
- 972 63. Wiede, F. *et al.* CCR6 is transiently upregulated on B cells after activation and modulates the  
973 germinal center reaction in the mouse. *Immunol. Cell Biol.* **91**, 335–339 (2013).
- 974 64. Lee, B. O. *et al.* CD40, but Not CD154, Expression on B Cells Is Necessary for Optimal Primary B  
975 Cell Responses. *J. Immunol.* **171**, 5707–5717 (2003).
- 976 65. Amariuta, T. *et al.* Improving the trans-ancestry portability of polygenic risk scores by prioritizing  
977 variants in predicted cell-type-specific regulatory elements. *Nat. Genet.* **52**, 1346–1354 (2020).
- 978 66. Amariuta, T. *et al.* IMPACT: Genomic Annotation of Cell-State-Specific Regulatory Elements  
979 Inferred from the Epigenome of Bound Transcription Factors. *Am. J. Hum. Genet.* **104**, 879–895  
980 (2019).
- 981 67. Chun, S. *et al.* Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-  
982 disease-associated loci in three major immune-cell types. *Nat. Genet.* **49**, 600–605 (2017).
- 983 68. Van Der Wijst, M. G. P. *et al.* Single-cell RNA sequencing identifies celltype-specific cis-eQTLs  
984 and co-expression QTLs. *Nat. Genet.* **50**, 493–497 (2018).
- 985 69. Kumasaka, N., Knights, A. J. & Gaffney, D. J. High-resolution genetic mapping of putative causal  
986 interactions between regions of open chromatin. *Nat. Genet.* **51**, 128–137 (2019).
- 987 70. Gutierrez-Arcelus, M. *et al.* Allele-specific expression changes dynamically during T cell activation  
988 in HLA and other autoimmune loci. *Nat. Genet.* **52**, 247–253 (2020).
- 989 71. Baglaenko, Y., Macfarlane, D., Marson, A., Nigrovic, P. A. & Raychaudhuri, S. Genome editing to  
990 define the function of risk loci and variants in rheumatic disease. *Nat. Rev. Rheumatol.* **17**, 462–  
991 474 (2021).
- 992 72. Arnett, F. C. *et al.* The american rheumatism association 1987 revised criteria for the classification  
993 of rheumatoid arthritis. *Arthritis Rheum.* **31**, 315–324 (1988).
- 994 73. Aletaha, D. *et al.* 2010 Rheumatoid arthritis classification criteria: An American College of  
995 Rheumatology/European League Against Rheumatism collaborative initiative. *Arthritis and*  
996 *Rheumatism* **62**, 2569–2581 (2010).
- 997 74. Akiyama, M. *et al.* Characterizing rare and low-frequency height-associated variants in the

- 998 Japanese population. *Nat. Commun.* **10**, (2019).
- 999 75. Zhang, W. *et al.* Whole genome sequencing of 35 individuals provides insights into the genetic  
1000 architecture of Korean population. *BMC Bioinformatics* **15**, (2014).
- 1001 76. Lan, T. *et al.* Deep whole-genome sequencing of 90 Han Chinese genomes. *Gigascience* **6**,  
1002 (2017).
- 1003 77. Wong, L. P. *et al.* Deep whole-genome sequencing of 100 southeast Asian malays. *Am. J. Hum.*  
1004 *Genet.* **92**, 52–66 (2013).
- 1005 78. Wakefield, J. A bayesian measure of the probability of false discovery in genetic epidemiology  
1006 studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
- 1007 79. Rumble, J. M., Fackelman, E. M. & Mobley, J. L. Comparative analyses of PAD expression and  
1008 activity in myeloid cell lines. *J. Immunol.* **198**, (2017).
- 1009 80. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 2013 91 **9**,  
1010 171–181 (2014).
- 1011 81. Finucane, H. K. *et al.* Heritability enrichment of specifically expressed genes identifies disease-  
1012 relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).
- 1013