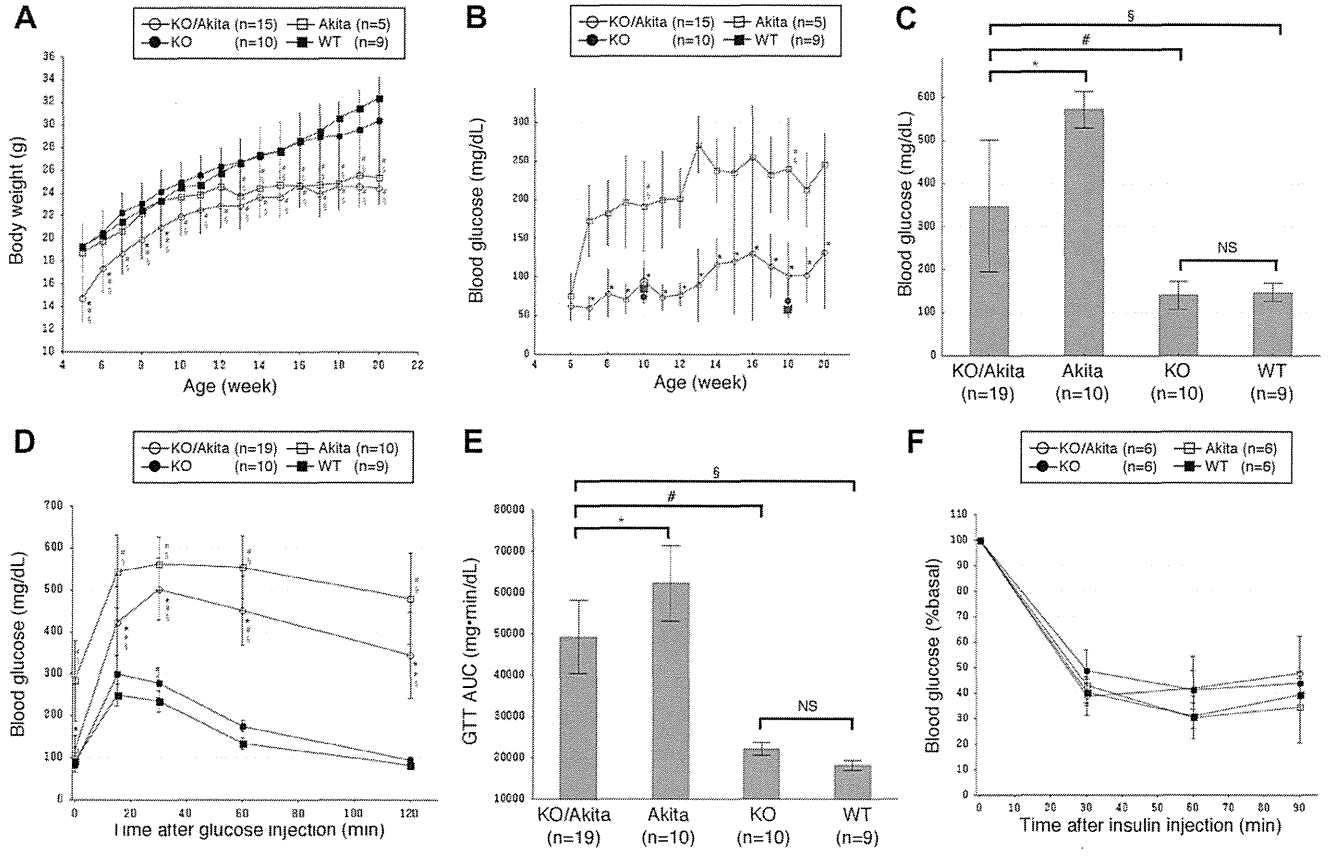

Fig. 2. Mouse growth curves, blood glucose concentrations, GTT and ITT. (A) Time course of body weight of KO/Akita, Akita, KO, and WT mice from 6 to 20 weeks of age. (B) Time course of 16 h fasting blood glucose concentrations in KO/Akita, Akita, KO and WT mice from 6 to 20 weeks of age. Glucose concentrations in KO and WT mice were measured at 10 and 18 weeks of age. (C) Six hours fasting blood glucose concentrations in 18 week old KO/Akita, Akita, KO, and WT mice. (D and E) GTT of 18 week old KO/Akita, Akita, KO, and WT mice. Blood glucose concentrations are shown at indicated times after glucose injections. Area under the curve was calculated for these mice. (F) ITT in 18 week old KO/Akita, Akita, KO, and WT mice. Blood glucose concentrations are shown at indicated times after insulin injections. Data are shown as mean ± SD. $^{*}p < 0.05$ vs Akita, $^{\#}p < 0.05$ vs KO, $^{\S}p < 0.05$ vs WT, NS, Not significant.

cantly higher in KO/Akita than in Akita mice (Supplemental Fig. 2B). Pancreas weight was similar in these 4 groups (Supplemental Fig. 3A).

3.8. Immunohistochemical assays of insulin and CHOP, and electron microscopy of islets

No morphological abnormalities were observed in the pancreas or islets of KO/Akita and KO mice. Immunohistochemical examination showed that a higher proportion of insulin-positive β cells was preserved in the islets of KO/Akita (0.141 ± 0.046 insulin positive cells/islet) than of Akita (0.088 ± 0.042 insulin positive cells/islet) mice, although both were lower than in KO (0.643 ± 0.080 insulin positive cells/islet) and WT (0.616 ± 0.076 insulin positive cells/islet) mice (Fig. 4B). Mean islet area did not differ among KO/Akita, Akita, KO and WT mice (Supplemental Fig. 3B).

CHOP is an ER stress-inducible transcription factor that promotes apoptosis [18] and that has been used as a marker of ER stress-mediated apoptosis in β cells of Akita mice [19]. To test whether ER stress occurs in the β cells of KO/Akita mice, we assayed for CHOP immunohistochemically. The percentage of CHOP-positive cells in islets was significantly lower in KO/Akita (0.102 ± 0.042 CHOP positive cells/islet) than in Akita (0.135 ± 0.037 CHOP positive cells/islet) mice, but were much lower in KO (0.002 ± 0.000 CHOP positive cells/islet) and WT (0.002 ± 0.000 (CHOP positive cells/islet) mice (Fig. 4C), indicating that ER stress is lower in the β cells of KO/Akita mice.

Electron microscopy of β cells in WT mice revealed abundant mature secretory granules in the cytoplasm, inconspicuous ER, and intact mitochondria with cristae (Fig. 4D, WT). KO mice showed no morphological abnormalities (Fig. 4D, KO). In contrast, examination of Akita mice showed a small number of secretory granules, a tubulovesicular structure comprised of markedly enlarged ER, and swelling or disruption of mitochondria (Fig 4D, Akita), indicators of insulin secretory pathway impairment and ER stress. Unlike Akita mice, KO/Akita mice showed mild ER enlargement and slight swelling of the mitochondria in β cells, although the number of secretory granules was markedly reduced (Fig. 4D, KO/Akita), suggesting less ER stress in the β cells of these mice than in Akita mice. The α cells of KO/Akita, Akita, KO and WT mice were morphologically similar (data not shown).

4. Discussion

We have shown here that targeted disruption of *Rnf213* unexpectedly improved glucose tolerance in Akita mice, although insulin sensitivity was not altered. These findings are consistent with results showing that plasma and pancreatic insulin levels were higher in KO/Akita than in Akita mice. Moreover, disruption of *Rnf213* reduced hyperphagia by elevating plasma insulin concentrations in KO/Akita, but did not alter plasma leptin concentrations in these mice. Taken together, these findings suggest that ablation of *Rnf213* may mitigate the diabetic phenotype by preserving β cell function.

Amelioration by *Rnf213* ablation contradicts a mechanistic link between MMD and diabetes [8], if variants were associated with MMD by loss-of-function or haploinsufficiency of *RNF213*. Alterna-

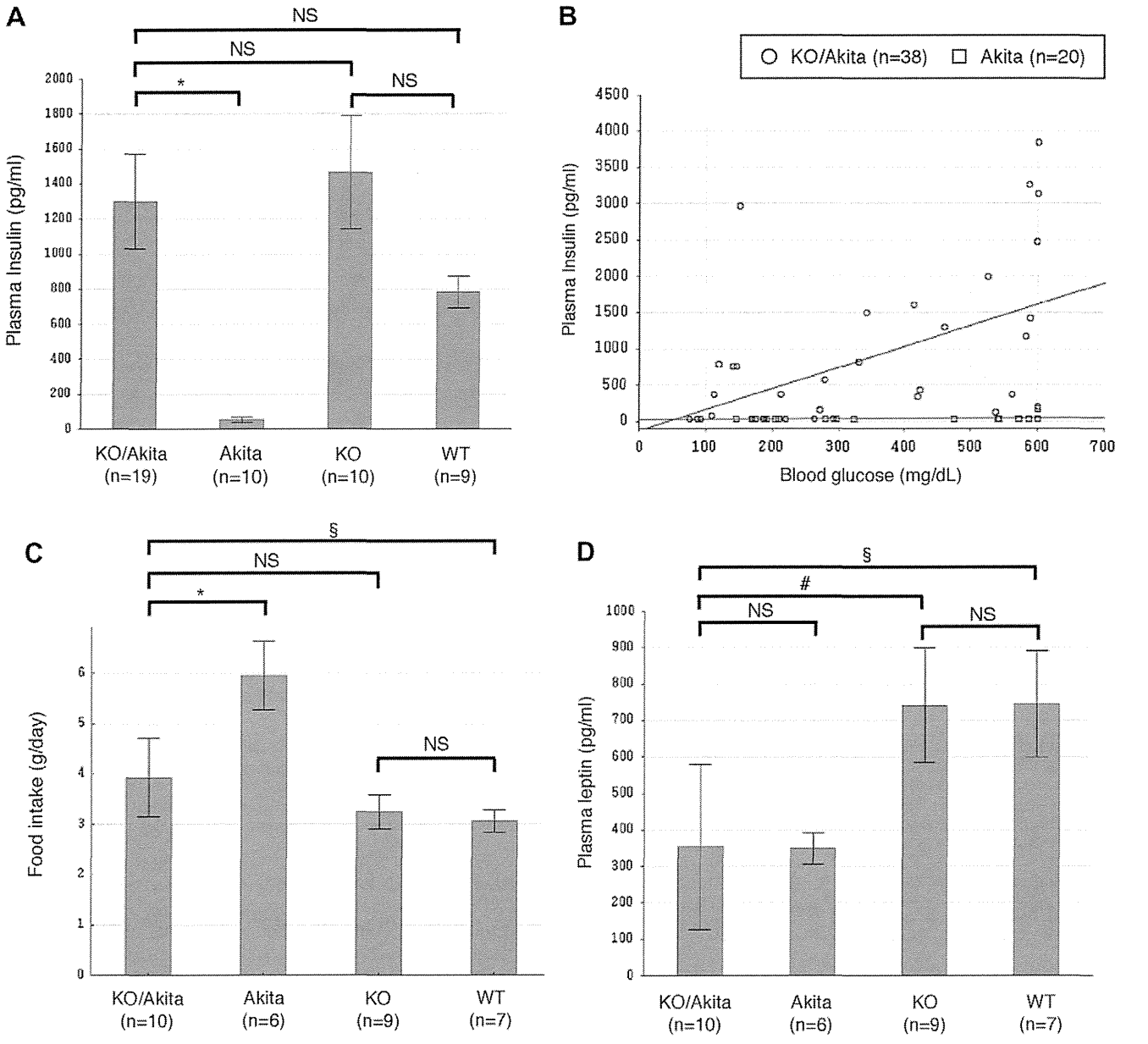

Fig. 3. Plasma insulin and leptin concentrations and food intake at 18 weeks of age. (A) Plasma insulin concentrations in KO/Akita, Akita, KO, and WT mice after a 6 h fast. Data are shown as mean ± SE. (B) Correlation between blood glucose and plasma insulin concentrations of KO/Akita and Akita mice after fasting for 6 h and 16 h (combined). (C) Food intake by KO/Akita, Akita, KO, and WT mice. (D) Plasma leptin concentrations of KO/Akita, Akita, KO, and WT mice after 16 h fasting. Data are shown as mean ± SD except for plasma insulin concentrations. $^{*}p < 0.05$ vs Akita, $^{\#}p < 0.05$ vs KO, $^{\S}p < 0.05$ vs WT, NS, not significant.

tively, pathological variants including R4810K of *RNF213* may cause MMD and diabetes by a gain-of-function or in a dominant-negative fashion. Among MMD predisposing diseases, diabetogenic mechanisms are well defined in MOPDII, a rare genetic disease characterized by severe growth retardation and early onset diabetes, as well as complication by MMD. Pericentrin, the causative gene for MOPDII, may regulate the intracellular distribution and secretion of insulin, and mutations of pericentrin may result in β-cell dysfunction [20]. The findings presented here indicate that β-cell dysfunction may have a mechanistic link with MMD.

Akita mice carrying a heterozygous C96Y mutation in the *Ins2* gene spontaneously develop hyperglycemia at an early age with reduced pancreatic β cell mass [12,13]. This C96Y mutation causes a conformational change in the insulin molecule, resulting in ER stress. ER stress, in turn, induces an unfolded protein response (UPR), indicating increased degradation of unfolded proteins by ER-associated degradation (ERAD), which is associated with E3 ligase and AAA + ATPase.

Recent studies [21,22] have demonstrated that the $Ins2^{C96Y}$ allele acts dominantly to enhance degradation of both the Akita and wild-type allele proinsulins by the ERAD pathway. We hypothesize that ablation of *Rnf213* may impair ERAD and lead to the sparing of wild-type proinsulin. Then we should explain how such preserved insulin secretion in KO/Akita mice reduced ER stress, as indicated by a reduction in the relative abundance of CHOP positive cells in these mice. Diabetes progresses more rapidly in male than female Akita mice [12]. This gender difference in susceptibility can be reversed by castration of males, thus suppressing hyperphagia [16]. Hyperphagia increases insulin demand due to elevated energy uptake, resulting in enhanced ER-stress with stimulated production of $Ins2^{C96Y}$. Such a vicious cycle may likely accelerate the progression of diabetes in male Akita mice. We found that

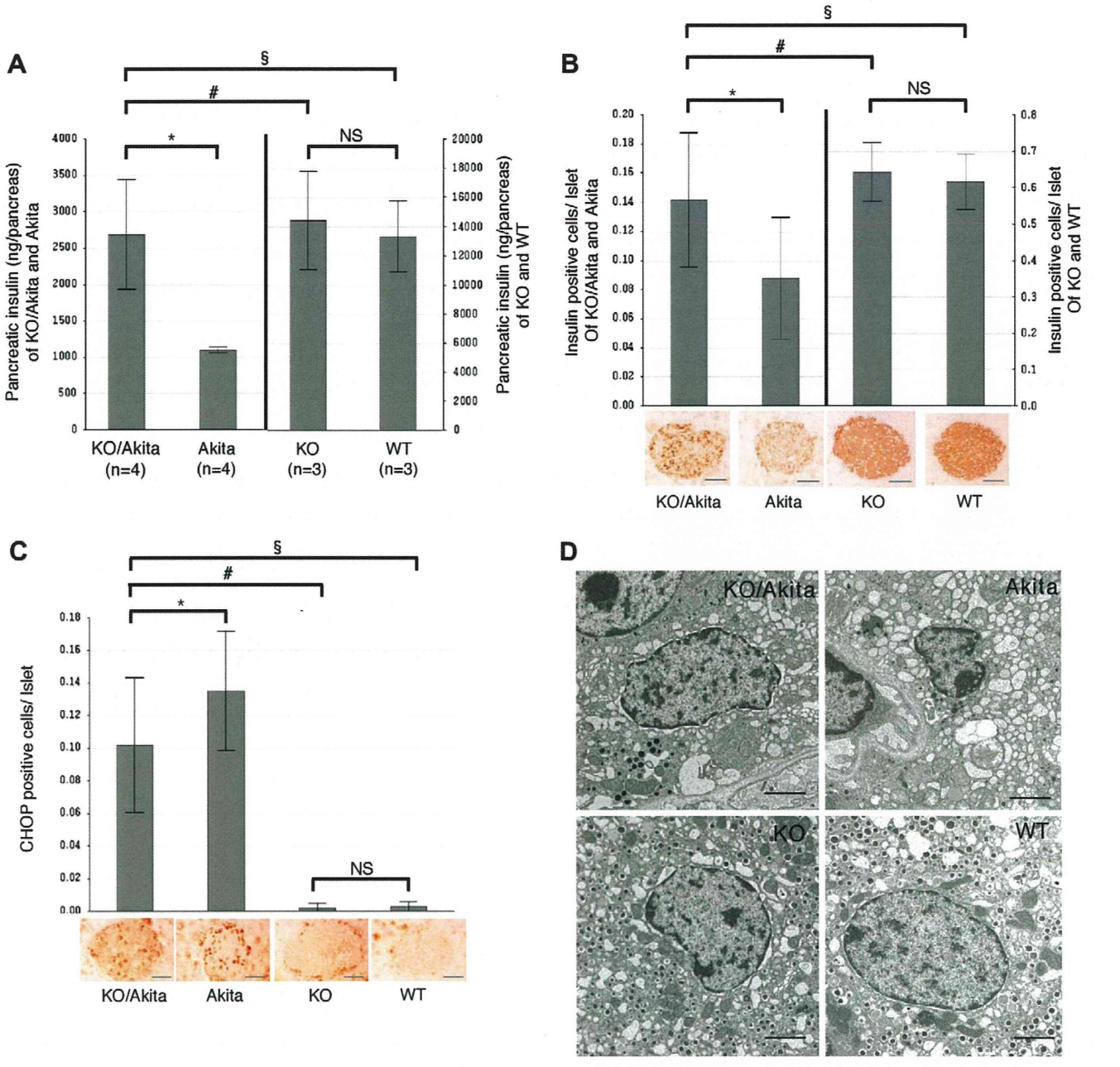

Fig. 4. Pancreatic insulin contents, insulin and CHOP immunohistochemistry, and electron microscopy of 18 week old mice. (A) Pancreatic insulin contents of KO/Akita, Akita, KO, and WT mice. (B) Representative images of islets stained with anti-insulin antibody (lower) and insulin positive cells per islet (upper) of KO/Akita ($n = 4$), Akita ($n = 4$), KO ($n = 3$), and WT ($n = 3$) mice. Quantification was performed on more than 20 islets from each mouse. Scale bar indicates 50 μm. (C) Representative images of islets stained with anti-CHOP antibody (lower) and CHOP positive cells per islet (upper) of KO/Akita ($n = 4$), Akita ($n = 4$), KO ($n = 3$), and WT ($n = 3$) mice. Quantification was performed on more than 20 islets from each mouse. Scale bar indicates 50 μm. (D) Electron micrographs of islets of KO/Akita, Akita, KO, and WT mice. Scale bar indicates 2 μm. Data are shown as mean ± SD. $^{*}p < 0.05$ vs Akita, $^{\#}p < 0.05$ vs KO, $^{\S}p < 0.05$ vs WT, NS, Not significant.

the higher serum insulin levels in KO/Akita mice were sufficient to suppress hyperphagia. Thus, *RNF213* ablation can spare wild-type insulin, thereby ameliorating this viscous cycle. Further study is warranted to test whether *RNF213* is involved in the ERAD pathway.

RNF213 is a single protein with two types of enzymatic activity, E3 ligase and AAA + ATPase [9]. AAA + ATPase is involved in various cellular processes, including vesicular transport, UPR, motor proteins and microtubule severing [11]. The association between *Rnf213* and β cell function is likely mediated by both E3 ligase and AAA + ATPase activities. The core assumption, that the normal allele of *Ins2* is also a target of degradation by ERAD, is intriguing and requires more quantitative assessment in the future. Future studies may help provide clues into a new therapeutic approach for diabetes as well as to gain insight into *RNF213* function.

Acknowledgments

This study was mainly supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Kiban Kenkyu A: 22249020) and from the Ministry of Health, Labour and Welfare of Japan (H23-Nanji-Ippan-01 and H23-Bio-Ippan-003) to AK and partially by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Tokubetukenkyuin

Syoreihi: 225192) to HK. We thank Ms. Emi Nakai for assistance with ES screening.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bbrc.2013.02.015.

References

[1] J. Suzuki, A. Takaku, Cerebrovascular "moyamoya" disease. Disease showing abnormal net-like vessels in base of brain, Arch. Neurol. 20 (1969) 288–299.
[2] R.M. Scott, E.R. Smith, Moyamoya disease and moyamoya syndrome, N. Engl. J. Med. 360 (2009) 1226–1237.
[3] D.S. Kainth, S.A. Chaudhry, H.S. Kainth, F.K. Suri, A.I. Qureshi, Prevalence and characteristics of concurrent down syndrome in patients with moyamoya disease, Neurosurgery 72 (2013) 210–215.
[4] K. Okazaki, A. Kakita, H. Tanaka, K. Kimura, M. Minagawa, T. Morita, H. Takahashi, Widespread ischemic brain lesions caused by vasculopathy associated with neurofibromatosis type 1, Neuropathology 30 (2010) 627–633.
[5] M.B. Bober, N. Khan, J. Kaplan, K. Lewis, J.A. Feinstein, C.I. Scott Jr., G.K. Steinberg, Majewski osteodysplastic primordial dwarfism type II (MOPD II): expanding the vascular phenotype, Am. J. Med. Genet. A 152A (2010) 960–965.
[6] A.J. Anwar, J.D. Walker, B.M. Frier, Type 1 diabetes mellitus and Down's syndrome: prevalence, management and diabetic complications, Diabet. Med. 15 (1998) 160–163.
[7] M. Kamoun, N. Charfi, N. Rekik, M.F. Mnif, F. Mnif, H. Kmiha, Z. Mnif, M. Abid, Neurofibromatosis and Type 1 diabetes mellitus: an unusual association, Diabet. Med. 26 (2009) 1180–1181.
[8] R.S. Bower, G.W. Mallory, M. Nwojo, F.B. Meyer, Y.C. Kudva, Diabetes mellitus and the moyamoya syndrome, Ann. Intern. Med. 157 (2012) 387–388.
[9] W. Liu, D. Morito, S. Takashima, Y. Mineharu, H. Kobayashi, T. Hitomi, H. Hashikata, N. Matsuura, S. Yamazaki, A. Toyoda, K. Kikuta, Y. Takagi, K.H. Harada, A. Fujiyama, R. Herzig, B. Krischek, L. Zou, J.E. Kim, M. Kitakaze, S. Miyamoto, K. Nagata, N. Hashimoto, A. Koizumi, Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development, PLoS One 6 (2011) e22542.
[10] R.J. Deshaies, C.A. Joazeiro, RING domain E3 ubiquitin ligases, Annu. Rev. Biochem. 78 (2009) 399–434.
[11] S.R. White, B. Lauring, AAA + ATPases: achieving diversity of function with conserved machinery, Traffic 8 (2007) 1657–1667.
[12] M. Yoshioka, T. Kayo, T. Ikeda, A. Koizumi, A novel locus, Mody4, distal to D7Mit189 on chromosome 7 determines early-onset NIDDM in nonobese C57BL/6 (Akita) mutant mice, Diabetes 46 (1997) 887–894.
[13] J. Wang, T. Takeuchi, S. Tanaka, S.K. Kubo, T. Kayo, D. Lu, K. Takata, A. Koizumi, T. Izumi, A mutation in the insulin 2 gene induces diabetes with severe pancreatic beta-cell dysfunction in the Mody mouse, J. Clin. Invest. 103 (1999) 27–37.
[14] J. Nozaki, H. Kubota, H. Yoshida, M. Naitoh, J. Goji, T. Yoshinaga, K. Mori, A. Koizumi, K. Nagata, The endoplasmic reticulum stress response is stimulated through the continuous activation of transcription factors ATF6 and XBP1 in Ins2^{+}/Akita pancreatic beta cells, Genes Cells 9 (2004) 261–270.
[15] J. Miyazaki, K. Araki, E. Yamato, H. Ikegami, T. Asano, Y. Shibasaki, Y. Oka, K. Yamamura, Establishment of a pancreatic beta cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms, Endocrinology 127 (1990) 126–132.
[16] M. Toyoshima, A. Asakawa, M. Fujimiya, K. Inoue, S. Inoue, M. Kinboshi, A. Koizumi, Dimorphic gene expression patterns of anorexigenic and orexigenic peptides in hypothalamus account male and female hyperphagia in Akita type 1 diabetic mice, Biochem. Biophys. Res. Commun. 352 (2007) 703–708.
[17] M.S. Martin-Gronert, S.E. Ozanne, Metabolic programming of insulin action and secretion, Diabetes Obes. Metab. 14 (Suppl. 3) (2012) 29–39.
[18] S.J. Marciniak, C.Y. Yun, S. Oyadomari, I. Novoa, Y. Zhang, R. Jungreis, K. Nagata, H.P. Harding, D. Ron, CHOP induces death by promoting protein synthesis and oxidation in the stressed endoplasmic reticulum, Genes Dev. 18 (2004) 3066–3077.
[19] S. Yamane, Y. Hamamoto, S. Harashima, N. Harada, A. Hamasaki, K. Toyoda, K. Fujita, E. Joo, Y. Seino, N. Inagaki, GLP-1 receptor agonist attenuates endoplasmic reticulum stress-mediated β-cell damage in Akita mice, J. Diabetes Invest. 2 (2011) 104–110.
[20] A. Jurczyk, S.C. Pino, B. O'Sullivan-Murphy, M. Addorio, E.A. Lidstone, P. Diiorio, K.L. Lipson, C. Standley, K. Fogarty, L. Lifshitz, F. Urano, J.P. Mordes, D.L. Greiner, A.A. Rossini, R. Bortell, A novel role for the centrosomal protein, pericentrin, in regulation of insulin secretory vesicle docking in mouse pancreatic beta-cells, PLoS One 5 (2010) e11812.
[21] J.R. Allen, L.X. Nguyen, K.E. Sargent, K.L. Lipson, A. Hackett, F. Urano, High ER stress in beta-cells stimulates intracellular degradation of misfolded insulin, Biochem. Biophys. Res. Commun. 324 (2004) 166–170.
[22] M. Liu, I. Hodish, C.J. Rhodes, P. Arvan, Proinsulin maturation, misfolding, and proteotoxicity, Proc. Natl. Acad. Sci. USA 104 (2007) 15841–15846.

Environ Health Prev Med (2013) 18:121–129
DOI 10.1007/s12199-012-0299-1

REGULAR ARTICLE

P.R4810K, a polymorphism of *RNF213*, the susceptibility gene for moyamoya disease, is associated with blood pressure

Akio Koizumi · Hatasu Kobayashi · Wanyang Liu · Yukiko Fujii · S. T. M. L. D. Senevirathna · Shanika Nanayakkara · Hiroko Okuda · Toshiaki Hitomi · Kouji H. Harada · Katsunobu Takenaka · Takao Watanabe · Shinichiro Shimbo

Received: 20 June 2012 / Accepted: 26 July 2012 / Published online: 10 August 2012

Abstract
Background Moyamoya disease—an idiopathic vascular disorder of intracranial arteries—is often accompanied by hypertension. *RNF213* has been identified as a susceptibility gene for moyamoya disease. In the present study, the association of p.R4810K (G>A) with blood pressure (BP) was investigated in a Japanese population.
Methodology/principal findings Three independent study populations, the Nyukawa ($n = 984$), Noshiro ($n = 2,443$) and Field ($n = 881$) studies, joined this study. BP, body weight and height were measured. Past and present symptoms and disease and medication histories were assessed by interview. Associations of p.R4810K (rs112735431, ss179362673) of *RNF213* with BP were investigated. Two linkage disequilibrium blocks were constructed for moyamoya patients with p.R4810K ($n = 140$) and the general population ($n = 384$) using 39 single nucleotide polymorphisms (SNPs) spanning 390 kb around *RNF213*. A total of 60 carriers (3 for AA genotype and 57 for GA genotype) were found in these samples, and the minor allele frequencies were 1.4 % in the Nyukawa and Field studies and 0.2 % in the Noshiro study. Regression analyses adjusted for age, sex and body mass index based on an additive model demonstrated significant associations with systolic BP (mmHg/allele): β (standard error) was 8.2 (2.9) in the Nyukawa study ($P = 4.7 \times 10^{-3}$), 18.7 (5.4) in the Noshiro study ($P = 4.6 \times 10^{-4}$) and 8.9 (2.0) ($P = 1.0 \times 10^{-5}$) in the three populations. In contrast, diastolic BP showed significant associations only in the Noshiro study. Linkage disequilibrium blocks contained none of the BP-associated proxy SNPs reported by previous studies.
Conclusions/significance Our study suggests that p.R4810K of *RNF213* is associated strongly with systolic BP.

Keywords *RNF213* · Moyamoya disease · P.R4810K · Systolic blood pressure · Hypertension

Electronic supplementary material The online version of this article (doi:10.1007/s12199-012-0299-1) contains supplementary material, which is available to authorized users.

A. Koizumi (✉) · H. Kobayashi · W. Liu · Y. Fujii · S. T. M. L. D. Senevirathna · S. Nanayakkara · H. Okuda · T. Hitomi · K. H. Harada
Department of Health and Environmental Sciences, Graduate School of Medicine, Kyoto University, Konoe-cho, Yoshida, Sakyo-ku, Kyoto 606-8501, Japan
e-mail: koizumi.akio.5v@kyoto-u.ac.jp

K. Takenaka
Department of Neurosurgery, Takayama Red Cross Hospital, Takayama, Japan

T. Watanabe
Department of Education, Tohoku Bunkyo College, Yamagata, Japan

S. Shimbo
Department of Food and Nutrition, Kyoto Women's University, Kyoto, Japan

Introduction

Moyamoya disease is an idiopathic occlusive vascular lesion that occurs at the terminal portion of internal carotid arteries in the circle of Willis [1, 2]. Familial clustering has been reported for moyamoya disease, with 15 % of cases being reported to have a family history [3]. This familial clustering led to searches for genetic factors, which identified several loci, including 3p24–p26 [4], 6q25 [5], 8q23 [6] and 17q25 [7, 8]. The locus on 17q25.3 was further refined by a linkage analysis [9] and an association study

[10]. Finally, *RNF213* was identified as the susceptibility gene of the 17q25.3 locus for moyamoya disease [11].

Several polymorphisms specific to Japanese, Korean, Chinese and Caucasian patients with moyamoya disease have been found in *RNF213* [11]. Of these, the p.R4810K (rs 112735431, ss179362673; G>A) polymorphism is a founder variant found commonly in Japanese, Korean and Chinese patients. In particular, it is found at rates of >90 % in Japanese patients and approximately 80 % in Korean patients. More surprisingly, 2–3 % of the Japanese and Korean general populations are carriers of this variant [11]. Given that the prevalence of moyamoya disease in patients is 10.5 per 10^5 people [12] and the rate of carriers of p.R4810K is 3 %, only a minor portion of carriers will develop moyamoya disease.

We conducted molecular cloning of a full-length cDNA of *RNF213* and found that it comprises 5,207 amino acids. *RNF213* has two well-known domains, a RING finger domain and a Walker motif that show ubiquitin ligase activity and ATPase activity, respectively [11]. *RNF213* was shown to be a novel functional E3 ligase [11]. Two splicing variants of *RNF213* were detected ubiquitously in cDNAs isolated from various types of cells, including vascular endothelial cells [11]. After knockdown of *RNF213* in zebrafish embryos, severely abnormal sprouting vessels were seen in the cranial vessels. However, the p.R4810K variant had no apparent detrimental effects on E3 ligase activity.

Moyamoya disease is often accompanied by hypertension [13, 14]. The pathogenesis of hypertension in patients with moyamoya disease has been attributed to stenosis of the renal arteries, i.e., renovascular hypertension. However, in some cases, hypertension was reported to occur without stenosis of the renal arteries [15]. Such observations suggest that hypertension in moyamoya disease may be caused by a novel mechanism associated with *RNF213* variants. If this is the case, it can be hypothesized that hypertension is a phenotype of p.R4810K carriers.

In the present study, we investigated whether p.R4810K elevates blood pressure (BP) in p.R4810K carriers. To examine this possibility, we conducted a genetic epidemiological study in three independent populations.

Methods

Ethical statement

Ethical approval for the three independent studies was given by the Institutional Review Board and Ethics Committee of Kyoto University School of Medicine, Kyoto University, Japan (Approval number: G140 approval date 10/18/2004). All the subjects provided written informed consent. All the studies involved the use of humans and adhered to the principles of the Declaration of Helsinki and Title 45, US Code of Federal Regulations, Part 46, Protection of Human Subjects.

Study population

Hypertension was defined as systolic blood pressure (SBP) of $\geq$140 mmHg and/or diastolic blood pressure (DBP) of $\geq$90 mmHg or treatment with antihypertensive medication. The pulse pressure (PP) was calculated as the difference between SBP and DBP.

Samples were donated by participants in three independent studies: the Nyukawa, Noshiro and Field studies. The Nyukawa and Noshiro studies are cohort studies to investigate common diseases. An annual health check-up is provided to the residents as a subsidy program run by the local autonomy at its public health center under the Industrial Safety and Health Act or Elderly Health law. Weight was measured in light clothes on an electrical balance, and height was measured with a stadiometer. Body mass index (BMI) was computed as weight in kilograms divided by the square of the height in meters. BP was measured on the right arm of seated participants by public health nurses after 5 min of rest at each examination using a mercury sphygmomanometer. BP was measured once or more than once when the readings were much higher than past records, i.e., 10 mmHg for SBP and 5 mmHg for DBP. In cases with more than one reading, the later BP reading was used. Clinical data such as past history of medication, past disease history, past and present symptoms, and demographic data (age and birth date) were collected by interview by public health nurses during the clinical examinations. The Nyukawa study started in 2004 in Nyukawa town in Gifu prefecture and has continued to the present, while the Noshiro study started in 1998 in Noshiro city in Akita prefecture and has continued to the present. All participants who underwent health check-ups at local health centers were recruited, and >95 % of the potential candidates agreed to join the present study. The participants donated blood samples at start of the studies and data collected at the start of the studies were used. Although these two studies are cohort studies, the study design for the present study was a cross-sectional association.

The Field study started in the early 1990s and ended in the mid-1990s as an environmental and nutritional study designed to evaluate exposure to environmental pollutants through dietary routes in nationwide general populations [16, 17]. Participants were recruited from among farmers by local community farming support stations. More than 90 % of subjects agreed to join the study. Owing to its focus on household information, such as dietary habits and

sources of vegetables, we selectively recruited more unrelated housewives than males. The researchers visited local towns and collected blood samples. Weight was measured in light clothes on a scale and height was measured with a stadiometer. BP was measured on the right arm of seated participants by physicians after 5 min of rest at each examination using a mercury sphygmomanometer. BP readings were measured once, and more than once when readings were much higher (i.e., 10 mmHg for SBP and 5 mmHg for DBP) than the values anticipated by the participants. In cases with more than one reading, the later BP reading was used. Clinical data such as past history of medication, past disease history, past and present symptoms, and demographic data (age and birth date) were collected by interview by physicians during the participant's visit. The samples and data were donated to the Kyoto Specimen sample bank [18, 19]. Since the minor allele frequency (MAF) of p.R4810K in the general population was significantly lower for the Noshiro population (0.2 %) than for the Nyukawa population (1.4 %), we selected populations in the western part of Japan in the Field study, i.e., Niigata, Ishikawa, Toyama, Tokyo, Gunma, Nagano, Aichi, Shiga, Kyoto, Kochi, Ehime, Shimane, Yamaguchi and Kagoshima.

We excluded from the analyses subjects with chronic kidney disease, polycystic kidney diseases, diabetic nephropathy and secondary hypertension based on the information obtained by interview. Blood-relatedness of the participants was checked by address, family names and interviews. If there was blood-relatedness among participants, one of the blood-related participants was selected and others were excluded from the analyses. For carriers, interview sheets describing blood-relatedness among carriers within the same communities were re-reviewed. Carriers or related family members were re-interviewed if needed.

Linkage disequilibrium around ss179362673

In the Japanese population, haplotypes carrying p.R4810K in *RNF213* were reported to be derived from a common founder [11]. We investigated linkage disequilibrium (LD) blocks in unrelated Japanese controls ($n = 384$) in the Kyoto and Osaka areas in western Japan, and in Japanese patients with p.R4810K and moyamoya disease ($n = 140$) [11]. Thirty-two single nucleotide polymorphisms (SNPs) (rs6565649, rs7216577, rs7406843, rs8078855, rs7217421, rs9902702, rs11869363, rs12451808, rs55996424, rs7222014, rs35968416, rs4890012, rs12150356, rs8070106, rs4889848, rs6565683, rs9913006, rs6565686, rs8065843, rs4074303, rs4890025, rs11869626, rs9898443, rs12601738, rs12185227, rs7502866, rs9911978, rs12950635 rs4890047, rs4889863, rs11655474, rs8080957) were selected on the basis of information in the Hapmap database (http://hapmap.ncbi.nlm.nih.gov/) to cover a region spanning 390 kb from the 5′ end of *SGSH* to the 5′ end of *Raptor*. These 32 SNPs were selected to illustrate the LD blocks using the Tagger program, with criteria of $r^2 > 0.65$ and a MAF of >0.05 in Japanese people. Typing for these 32 SNPs plus 7 rare SNPs (ss179362670, ss179362671, ss179362672, ss179362673, ss179362674, ss179362675, ss161110142) was selected. Each LD block was constructed using SNP & Variation Suite V7. LDs were evaluated by D'.

Genotyping and quality control

Genomic DNA was extracted from blood samples with a QIAamp DNA Blood Mini Kit (Qiagen, Tokyo, Japan). Genotyping of the 39 SNPs was conducted using TaqMan probes (TaqMan SNP Genotyping Assays; Applied Biosystems, Tokyo, Japan) using a 7300/7500 Real-Time PCR System (Applied Biosystems, Foster City, CA) as previously reported [11]. Data were cleaned using a quality control process. This process included evaluation of sample and marker call rates, mismatches of sex, age and BP readings, past histories, duplicates and batch effects. Sample relatedness and population stratification were not evaluated genetically. The genotyping results for p.R4810K were confirmed by direct sequencing by the Sanger method in 42 subjects out of 140 patients, as previously reported [11]. The genotype of p.R4810K in two independent genotyping procedures was perfectly matched for control 384 subjects. The call rates for the 39 SNPs were 100 %.

Statistical analysis

Hypertension was analyzed as a binary trait (cases vs controls) using a logistic model under an additive model. Association of p.R4810K with BP was also analyzed by an additive model as a quantitative trait using linear regression models. For subjects taking antihypertensive therapies, BPs were imputed by adding 10 and 5 mmHg to the SBP and DBP values, respectively [20]. Unless otherwise specified, analyses were based on the protocol used in the large studies [21–23]: adjustment for age, sex and BMI: inclusion of subjects taking antihypertensive medications and imputation of their BP. Genetic association analyses were performed using SNP & Variation Suite V7 (Golden Helix, Bozeman, MT). In the analyses, we tested additive and dominant models. Odds ratio was calculated by the SNP & Variation Suite for dichotomous phenotypes. Hardy-Weinberg equilibrium (HWE) was tested in entire populations. The exclusion criteria was HWE $<10^{-3}$.

A Bonferroni correction of $P = 0.05/2 = 0.025$ was applied to the regression analysis. This correction was selected because two parameters, SBP and DBP, were independent, while the other parameters (PP and adjusted

values) were derivatives of these values. Other statistical analyses were conducted using SAS software (SAS Institute, Cary, NC). A value of $P < 0.05$ was considered to indicate statistical significance.

Results

Demographic characteristics of three populations

The demographic characteristics of the participants in the three studies are summarized in Table 1. The numbers of participants were 984 for the Nyukawa study, 2,443 for the Noshiro study and 881 for the Field study. The percentages of patients taking antihypertensive agents or diuretics ranged from 10.8 % in the Field study to 15.8 % in the Noshiro study. The percentages of patients with hypertension ranged from 12.8 % in the Field study to 17.6 % in the Noshiro study.

The MAFs of p.R4810K were 1.4 % in the Nyukawa and Field studies, and 0.2 % in the Noshiro study (Table 2). The MAF was significantly lower in the Noshiro study than in the Nyukawa and Field studies (Fisher's exact test, $P < 1 \times 10^{-5}$). Based on interview, none of the carriers had any symptoms related to moyamoya disease or past history of stroke.

Blood-relatedness checked by address and family names did not exclude any subjects in three studies. None of the 60 carriers were confirmed to be blood-related by re-review of the interview sheets.

Table 1 Demographic characteristics of the three samples

Characteristics	Nyukawa study	Noshiro study	Field study
Observation period	2004-present	1998-present	1990s
Number of participants	984	2,443	881
Men:women	425:559	1,229:1,214	188:693
Age (years)	56.1 ± 15.1	49.9 ± 14.0	50.8 ± 10.4
Body mass index (BMI)	22.3 ± 2.9	23.3 ± 3.1	23.2 ± 3.0
SBP (mmHg)	125.2 ± 19.2	126.6 ± 18.2	129.1 ± 13.8
DBP (mmHg)	72.7 ± 12.1	77.5 ± 11.9	76.6 ± 10.3
% Taking antihypertensive or diuretics	10.9	15.8	10.8
% Hypertension	15.5	17.6	12.8

Data are shown as values or mean ± standard deviation (SD)

Table 2 Association of p.R4810K with hypertension by additive model

Study	p.R4810K genotype (G>A)						Dichotomous	
	AA	GA	GG	Total	MAF (%)		HT (+)	HT (−)
Nyukawa	2	23	959	984	1.4	AA	1	1
						AG	6	17
						GG	145	814
						Odds (95 % CI)	2.09 (0.96–4.58)	
	HWE	0.012				*P*	0.08	
Noshiro	0	11	2,432	2,443	0.2	AG	3	8
						GG	428	2004
						Odds (95 % CI)	1.76 (0.46–6.64)	
	HWE	1				*P*	0.368	
Field	1	23	857	881	1.4	AA	0	1
						AG	5	18
						GG	108	749
						Odds (95 % CI)	1.93 (0.70–5.30)	
	HWE	0.159				*P*	0.329	
Combined	3	57	4,248	4,308	0.7	AA	1	2
						AG	14	43
						GG	681	3567
						Odds (95 % CI)	1.71 (0.93–3.13)	
	HWE	1.39×10^{-3}				*P*	0.055	

MAF minor allele frequency, *HT* hypertension, *CI* confidence interval, *HWE* Hardy–Weinberg equilibrium

Table 3 Association of p.R4810K with blood pressures by additive model and dominant model

	SBP			DBP			PP		
	β	SE	p	β	SE	p	β	SE	p
Additive model study									
	Raw data (mm Hg)								
Nyukawa	8.2	2.9	$4.7 \times 10^{-3*}$	−0.4	2.0	0.83	8.5	2.1	$4.6 \times 10^{-5*}$
Noshiro	18.7	5.4	$4.6 \times 10^{-4*}$	11.4	3.4	$6.6 \times 10^{-4*}$	7.3	2.8	$1.0 \times 10^{-2*}$
Field studies	4.7	2.4	5.6×10^{-2}	−0.4	2.0	0.83	5.1	2.2	$1.8 \times 10^{-2*}$
Combined	8.9	2.0	$1.0 \times 10^{-5*}$	0.6	1.9	0.70	8.3	1.3	$5.4 \times 10^{-11*}$
	Adjusted data (mmHg)								
Nyukawa	10.0	3.2	$1.6 \times 10^{-3*}$	0.5	2.1	0.82	9.6	2.2	$1.7 \times 10^{-5*}$
Noshiro	19.1	5.8	$1.0 \times 10^{-3*}$	11.6	3.6	$1.2 \times 10^{-3*}$	7.5	3.0	$1.4 \times 10^{-2*}$
Field studies	6.1	2.6	$2.0 \times 10^{-2*}$	0.3	2.0	0.88	5.8	2.2	$8.6 \times 10^{-3*}$
Combined	10.0	2.2	$5.0 \times 10^{-5*}$	1.2	1.4	0.40	8.9	1.3	$2.3 \times 10^{-11*}$
Dominant model study									
	Raw data (mm Hg)								
Nyukawa	9.2	3.2	$4.4 \times 10^{-3*}$	−0.2	2.2	0.94	9.4	2.4	$7.6 \times 10^{-5*}$
Noshiro	18.7	5.4	$4.6 \times 10^{-4*}$	11.4	3.4	$6.6 \times 10^{-4*}$	7.3	2.8	$1.0 \times 10^{-2*}$
Field studies	6.3	1.7	$3.0 \times 10^{-4*}$	0.04	1.4	0.97	6.2	1.5	$3.1 \times 10^{-5*}$
Combined	9.9	2.2	$5.7 \times 10^{-6*}$	1.1	1.5	0.46	8.8	1.4	$1.0 \times 10^{-10*}$
	Adjusted data (mmHg)								
Nyukawa	10.8	3.5	$2.3 \times 10^{-3*}$	0.6	2.3	0.78	10.2	2.5	$4.0 \times 10^{-5*}$
Noshiro	19.1	5.8	$1.0 \times 10^{-3*}$	11.6	3.6	$1.2 \times 10^{-3*}$	7.5	3.0	$1.4 \times 10^{-2*}$
Field studies	6.7	2.8	$2.0 \times 10^{-2*}$	0.5	2.2	0.82	6.9	2.5	$5.4 \times 10^{-3*}$
Combined	10.8	2.4	$4.9 \times 10^{-6*}$	1.5	1.5	0.92	9.2	1.4	$8.5 \times 10^{-11*}$

* Significant association $p < 0.025$

Association of p.R4810K with hypertension

The associations of p.R4810K by the additive model for dichotomous traits (hypertension vs normal) were not significant in any of the populations (Table 2).

Regression analyses adjusted for age, sex and BMI revealed significant associations of p.R4810K with SBP and PP (Table 3). Specifically, the β [standard error (SE)] values for SBP were 8.2 (2.9) mmHg/allele for the Nyukawa study and 18.7 (5.4) mmHg/allele for the Noshiro study, while those for PP were 8.5 (2.1) mmHg/allele for the Nyukawa study, 7.3 (2.8) mmHg/allele for the Noshiro study and 5.1 (2.2) mmHg/allele for the Field study. The associations of p.R4810K with therapy-adjusted SBP and PP were also confirmed in the three studies. In the three combined populations, p.R4810K was highly associated with SBP [8.9 (2.0) mmHg/allele] and adjusted SBP [10.0 (2.2) mmHg/allele]; with PP [8.3 (1.3) mmHg/allele] and adjusted PP [8.9 (1.3) mmHg/allele]. The results by additive model are essentially the same by dominant model.

The characteristic features of carriers are shown in Table 4. SBP was significantly higher for the AA + GA genotype than for the GG genotype [136.4 (20.3) vs 126.4 (18.2) mmHg/allele, $P < 0.001$], while the male/female ratios, age and BMI did not differ significantly between any two groups. Adjusted SBP (mmHg/allele: mean ± SD) was 155.3 (19.7) for the AA genotype, 137.8 (22.7) for the GA genotype and 138.7 (22.7) for the AA + GA genotype, which is significantly higher than 127.7 (19.9) with the GG genotype ($P < 0.001$). On the other hand, neither DBP nor adjusted DBP differed between any two groups: the adjusted BP was 78.2 (13.2) and 76.9 (12.5) for GG genotypes, respectively. The % hypertension was marginally higher in the AA + AG group than in the GG group (Fisher's exact test, $P = 0.06$).

Effects on associations of protocol changes for statistical analysis

The present statistical analyses were based on the same protocol [21–23]. The protocol included subjects with antihypertensive medications and adjustment for age, sex and BMI. We investigate the effects of modifications of the protocol on the statistical analyses to determine to what degree the current results are dependent on the protocol. First, we eliminated subjects with antihypertensive

Table 4 Characteristics of p.R4810K carriers. Values are mean (standard deviation)

Genotype	*N*	M:F	Age (years)	BMI	Raw blood pressure (mmHg)		% Hypertension
					SBP	DBP	
AA	3	1:2	72.7 (5.1)	23.2 (4.8)	145.3 (19.7)	73.3 (5.8)	66.7
GA	57	14:43	51.9 (21.4)	23.5 (3.9)	135.9 (20.4)	77.3 (13.1)	24.6
AA + GA	60	15:45	52.4 (14.0)	23.5 (3.9)	136.4 (20.3)*	77.1 (12.8)	25.0
GG	4,248	1,827:2,421	51.1 (13.9)	23.0 (3.0)	126.4 (18.2)	76.2 (11.8)	16.0

* Significantly different from the values in GG genotype $P < 0.001$

medications to investigate the effect of genotypes without modifications by medications. Although reduced numbers of subjects decreased statistical power, there were significant associations of p.R4810K with SBP and PP (Table S1). Second, analyses were conducted without adjustment for BMI for the population including subjects with antihypertensive medications. The associations were reproduced without adjustment for BMI (Tables S2 and S3). Changes in the statistical analysis protocol did not alter the results by quantitative regression analyses extensively, suggesting that associations of p.R4810K are robust and reproducible.

LD blocks around p.R4810K of *RNF213*

Analyses of the LD blocks of 39 SNPs spanning 390 kb in patients with p.R4810K and moyamoya disease demonstrated that p.R4810K was in LD with the 3′ half of *RNF213* and *ENDOV* (Fig. 1a). Similar LD blocks were observed in the general control population (Fig. 1b). LD blocks of p.R4810K did not contain any proxy SNPs, which have been reported to be associated with BP in East Asian people or other ethnicities, even when less stringent statistical criteria are applied [21–23].

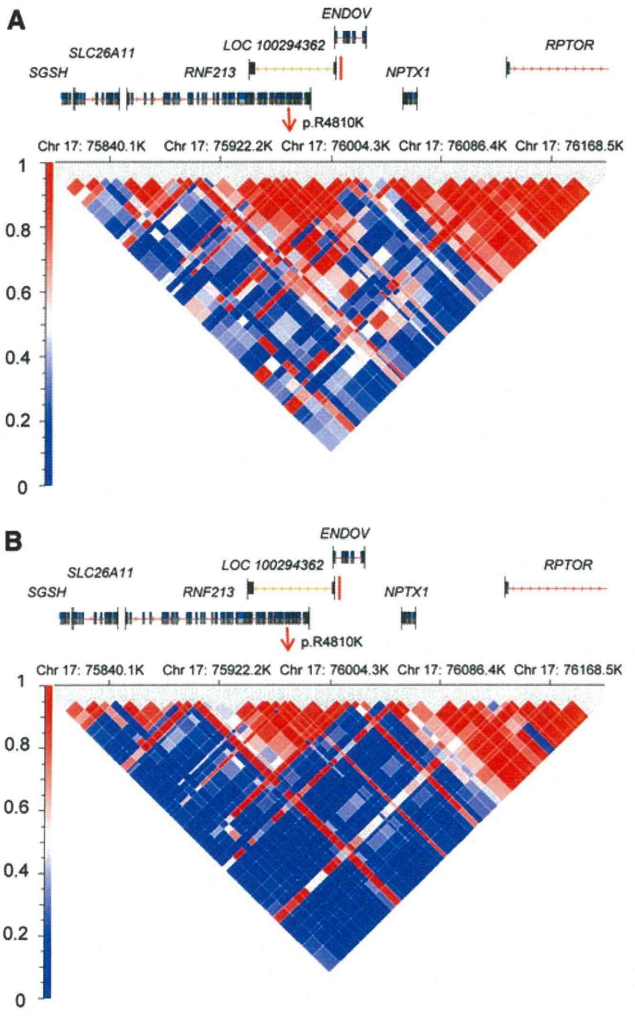

Fig. 1 Linkage disequilibrium (LD) blocks from a region spanning 390 kb from *SGSH* to *RAPTOR* in two independent populations. The LD blocks were constructed with 39 single nucleotide polymorphisms (SNPs). Scales are D'. The physical positions refer to Genome View Build 36 (http://www.ncbi.nlm.nih.gov/mapview/). **a** LD blocks for 140 unrelated Japanese patients with p.R4810K and moyamoya disease. **b** LD blocks for 384 unrelated Japanese subjects

Discussion

In the present study, quantitative trait locus analyses demonstrated that p.R4810K was associated significantly with SBP and PP in the general population. The p.R4810K allele resulted in consistently elevated SBP from 4.7 to 18.7 mmHg/allele and PP from 5.1 to 8.5 mmHg/allele. Such large and consistent contributions support our hypothesis that p.R4810K increases SBP.

Several large-scale studies have used the genome-wide association study (GWAS) approach [21–23]. However, none of these studies detected a locus on 17q25.3 for hypertension. Investigations of the LD blocks of p.R4810K failed to reveal any SNPs that have been reported to be associated with BP, even with less stringent *P* values [21–23]. There is no doubt that recent association studies have been successful; however, we consider that there are three major reasons for the failure to detect an association of *RNF213* polymorphisms with hypertension. First, the

critical assumption for the efficient detection of association through GWAS using LD mapping with TagSNPs ≥5 % is that, for the susceptibility locus, there is only low level allelic heterogeneity. In the presence of allelic heterogeneities, GWAS with TagSNPs may have missed potentially significant associations [24, 25]. Second, another GWAS assumption is that variants are common. The GWAS protocol usually discards minor alleles, i.e., with allele frequencies smaller than 10 %. Thus, minor alleles in *RNF213*, such as p.R4810K, might have been eliminated from the analysis. Third, even large studies are not free from sample size limitations to detect small population effects even using TagSNPs. For example, a simple genetic power calculation [26] on the assumption (HT prevalence 0.2; risk of the allele = 2.0; D' between a tag SNP and risk allele 0.9; tag SNP allele frequency = 0.5) reveals that to obtain alpha = 0.05 and power = 0.80, more than 50,000 HT Japanese cases would be needed for the low frequencies of risk allele ~0.01. Given the lower frequencies in the Tohoku areas of Japan, more cases would have been needed. Collectively, although p.R4810K has a discernible influence on BP in individuals, its small contribution to the population BP makes it extremely difficult to detect, even in large scale GWAS studies.

Given that the current observation can be generalized, carriers of p.R4810K are at risk of high SBP, irrespective of whether they have moyamoya disease. An increase of this magnitude overwhelms the population average of BP treatment effects for single antihypertensive agents [27]. Moreover, it has been suggested that differences in SBP of this magnitude elevate total mortality and the incidence of stroke [28]. Therefore, p.R4810K carriers may need special attention to control their SBP to prevent cerebrovascular events, regardless of whether they are affected by moyamoya disease. In this regard, it remains an enigma why p.R4810K has a strong association with SBP while it does not with DBP. Further studies are warranted to explain the preferential association with SBP.

Hypertension in patients with moyamoya disease has been postulated to be caused by renovascular hypertension [13, 14]. Similarly, it may occur in carriers without manifestation of moyamoya disease. There may also be other possibilities for the mechanisms of hypertension. For example, it remains to be addressed why hypertension occurs in some patients with moyamoya disease without stenosis of the renal arteries [15]. The physiological role of the p.R4810K variant of *RNF213* in the pathological process is unknown. *RNF213* comprises a novel class of E3 ligase and is involved in angiogenesis [11]. Therefore, its biochemical and physiological functions need to be elucidated. Further studies are required with a focus on the functional aspects of p.R4810K of *RNF213* in SBP.

Except for Down syndrome [29], various disorders associated with moyamoya syndrome are also known to be associated with BP, including Sickle cell disease [30, 31], neurofibromatosis I [32], Noonan syndrome [33, 34], Seckel syndrome [35], familial thoracic aortic aneurysm/dissection syndrome caused by *ACTA2* mutations [36] and X-linked moyamoya syndrome [37]. The genetic association between BP and p.R4810K of *RNF213* may not be unique to moyamoya disease, but could instead be common to the pathophysiology associated with moyamoya syndrome. The common pathological processes in these diseases are associated with occlusive vascular lesions. The biomedical mechanisms associated with occlusive vascular lesions may be attributable to endothelial dysfunction, as is the case for prehypertension [38]. Alternatively, it is interesting that patients with *ACTA2* mutations have excessive proliferation of vascular smooth muscle cells, and thereby elevated BP [36].

The present study has several limitations, the most major being the population sizes of the carriers of p.R4810K. Although we performed extensive genotyping in the three populations, we could not obtain large numbers of carriers because of the low prevalence of p.R4810K. It is of particular interest that the MAF of p.R4810K is significantly lower in the eastern part of mainland of Japan than in the western part of Japan. It should be investigated in future whether or not the prevalence of moyamoya disease has geographic differences in Japan. Specifically, it may be lower in the eastern part of mainland Japan than in the western part of Japan, as predicted by the low prevalence in Caucasians [11]. In this regard, although we cannot eliminate the effects of population stratification completely, the strength of the present study, namely a single ethnicity, can mitigate such biases. In addition, we did not test genetic relatedness among participants, although we checked it with indirect information. Thus, blood-related samples may be contaminated in associations and lead to inflation of associations. The second limitation of this study is that clinical data were not available for examining renovascular hypertension in carriers. Third, it could be argued p.R4810K is a proxy SNP or that *RNF213* is associated directly with BP, and that p.R4810K of *RNF213* may elevate BP. At present, we cannot say whether *RNF213* is associated with BP, and whether p.R4180K elevates BP directly or not. Further evidence is required to explain the association of p.R4810K with BP biologically. Finally, we cannot exclude a healthy participant effect in the current study. In fact, most of the participants are active both mentally and physically. Handicapped subjects with hypertension-associated cardiovascular diseases or subjects receiving clinical care might have been eliminated in the communities. The present findings may help to determine the pathological consequences of moyamoya disease in

carriers of p.R4810K. Furthermore, our findings suggest that high SBP may be a common phenotype of carriers of p.R4810K. Finally, although the MAF of p.R4810K in the general population is small, antihypertensive measures may be efficient for this high-risk population.

Acknowledgments We thank Dr. Kayoko Inoue, who died in 2007, Prof. Yasuhiko Wada (University of Kochi Prefecture) and Prof. Masayuki Ikeda (Kyoto Industrial Health Association). We are also grateful to Ms. Misato Umakosi (Kyoto University Graduate School of Medicine). This work was supported mainly by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan (Kiban Kenkyu A: 22249020) from the ministry of Health, Labour and Welfare of Japan (H23-Nanji-Ippan-019, chaired by Dr. Nobuo Hashimoto) and partially by a grant from Creative Scientific Research (19G50314).

Conflict of interest None.

References

1. Takeuchi K, Shimizu K. Hypogenesis of bilateral internal carotid arteries. Brain Nerve. 1957;9:37–43.
2. Suzuki J, Takaku A. Cerebrovascular "moyamoya" disease. Disease showing abnormal net-like vessels in base of brain. Arch Neurol. 1969;20:288–99.
3. Mineharu Y, Takenaka K, Yamakawa H, Inoue K, Ikeda H, et al. Inheritance pattern of familial moyamoya disease: autosomal dominant mode and genomic imprinting. J Neurol Neurosurg Psychiatry. 2006;77:1025–9.
4. Ikeda H, Sasaki T, Yoshimoto T, Fukui M, Arinami T. Mapping of a familial moyamoya disease gene to chromosome 3p24.2-p26. Am J Hum Genet. 1999;64:533–7.
5. Inoue TK, Ikezaki K, Sasazuki T, Matsushima T, Fukui M. Linkage analysis of moyamoya disease on chromosome 6. J Child Neurol. 2000;15:179–82.
6. Sakurai K, Horiuchi Y, Ikeda H, Ikezaki K, Yoshimoto T, et al. A novel susceptibility locus for moyamoya disease on chromosome 8q23. J Hum Genet. 2004;49:278–81.
7. Yamauchi T, Tada M, Houkin K, Tanaka T, Nakamura Y, et al. Linkage of familial moyamoya disease (spontaneous occlusion of the circle of Willis) to chromosome 17q25. Stroke. 2000;31: 930–5.
8. Mineharu Y, Liu W, Inoue K, Matsuura N, Inoue S, et al. Autosomal dominant moyamoya disease maps to chromosome 17q25.3. Neurology. 2008;70:2357–63.
9. Liu W, Hashikata H, Inoue K, Matsuura N, Mineharu Y, et al. A rare Asian founder polymorphism of Raptor may explain the high prevalence of moyamoya disease among East Asians and its low prevalence among Caucasians. Environ Health Prev Med. 2010; 15:94–104.
10. Kamada F, Aoki Y, Narisawa A, Abe Y, Komatsuzaki S, et al. A genome-wide association study identifies RNF213 as the first moyamoya disease gene. J Hum Genet. 2011;56:34–40.
11. Liu W, Morito D, Takashima S, Mineharu Y, Kobayashi H, et al. Identification of RNF213 as a susceptibility gene for moyamoya disease and its possible role in vascular development. PLoS ONE. 2011;6:e22542.
12. Kuroda S, Houkin K. Moyamoya disease: current concepts and future perspectives. Lancet Neurol. 2008;7:1056–66.
13. Yamada I, Himeno Y, Matsushima Y, Shibuya H. Renal artery lesions in patients with moyamoya disease: angiographic findings. Stroke. 2000;31:733–7.
14. Togao O, Mihara F, Yoshiura T, Tanaka A, Kuwabara Y, et al. Prevalence of stenoocclusive lesions in the renal and abdominal arteries in moyamoya disease. AJR Am J Roentgenol. 2004;183:119–22.
15. Limaye CS, Khude S, Pednekar SJ. Moyamoya disease with hypertension in a young adult. J Assoc Physicians India. 2011;59: 124–6.
16. Shimbo S, Zhang ZW, Moon CS, Watanabe T, Nakatsuka H, et al. Correlation between urine and blood concentrations, and dietary intake of cadmium and lead among women in the general population of Japan. Int Arch Occup Environ Health. 2000;73:163–70.
17. Watanabe T, Shimbo S, Nakatsuka H, Koizumi A, Higashikawa K, et al. Gender-related difference, geographical variation and time trend in dietary cadmium intake in Japan. Sci Total Environ. 2004;329:17–27.
18. Koizumi A, Yoshinaga T, Harada K, Inoue K, Morikawa A, et al. Assessment of human exposure to polychlorinated biphenyls and polybrominated diphenyl ethers in Japan using archived samples from the early 1980s and mid-1990s. Environ Res. 2005;99:31–9.
19. Koizumi A, Harada KH, Inoue K, Hitomi T, Yang HR, et al. Past, present, and future of environmental specimen banks. Environ Health Prev Med. 2009;14:307–18.
20. Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. Hypertension. 2003;41:207–10.
21. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, et al. Genome-wide association study of blood pressure and hypertension. Nat Genet. 2009;41:677–87.
22. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. Nature. 2011;478:103–9.
23. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, et al. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in East Asians. Nat Genet. 2011;43:531–8.
24. Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR, Amon CI. Shifting paradigm of association studies: value of rare single nucleotide polymorphisms. Am J Human Genet. 2008;82:100–12.
25. Slager SL, Huang J, Vieland VJ. Effect of allelic heterogeneity on the power of the transmission disequilibrium test. Genet Epidemiol. 2000;18:143–56.
26. Purcell S, Cherny SS, Sham PC. Genetic power calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics. 2003;19:149–50.
27. Burt VL, Cutler JA, Higgins M, Horan MJ, Labarthe D, et al. Trends in the prevalence, awareness, treatment, and control of hypertension in the adult US population. Data from the health examination surveys, 1960 to 1991. Hypertension. 1995;26:60–9.
28. Staessen JA, Gasowski J, Wang JG, Thijs L, Den Hond E, et al. Risks of untreated and treated isolated systolic hypertension in the elderly: meta-analysis of outcome trials. Lancet. 2000;355:865–72.
29. Draheim CC, Geijer JR, Dengel DR. Comparison of intima-media thickness of the carotid artery and cardiovascular disease risk factors in adults with versus without the Down syndrome. Am J Cardiol. 2010;106:1512–6.
30. Koduri PR, Agbemadzo B, Nathan S. Hemoglobin S-C disease revisited: clinical study of 106 adults. Am J Hematol. 2001;68:298–300.
31. Reid HL, Anah CO. Haemorheological parameters in hypertensive Nigerians with and without sickle-cell trait. Angiology. 1985;36:379–86.
32. Zinnamosca L, Petramala L, Cotesta D, Marinelli C, Schina M, et al. Neurofibromatosis type 1 (NF1) and pheochromocytoma: prevalence, clinical and cardiovascular aspects. Arch Dermatol Res. 2011;303:317–25.

33. Eu-ahsunthornwattana J, Trachoo O, Dejsuphong D, Tunteeratum A, Srichan K, et al. Noonan syndrome, metabolic syndrome and stroke-in-the-young: coincidence, causal or contribution? J Med Assoc Thai. 2010;93:1084–7.
34. Rokicki W, Rokicka A. Noonan syndrome coexisting with essential arterial hypertension in 8 year old boy. Wiad Lek. 2002;55:488–93.
35. Di Bartolomeo R, Polidori G, Piastra M, Viola L, Zampino G, et al. Malignant hypertension and cerebral haemorrhage in Seckel syndrome. Eur J Pediatr. 2003;162:860–2.
36. Guo DC, Papke CL, Tran-Fadulu V, Regalado ES, Avidan N, et al. Mutations in smooth muscle alpha-actin (ACTA2) cause coronary artery disease, stroke, and moyamoya disease, along with thoracic aortic disease. Am J Hum Genet. 2009;84:617–27.
37. Miskinyte S, Butler MG, Herve D, Sarret C, Nicolino M, et al. Loss of BRCC3 deubiquitinating enzyme leads to abnormal angiogenesis and is associated with syndromic moyamoya. Am J Hum Genet. 2011;88:718–28.
38. Giannotti G, Doerries C, Mocharla PS, Mueller MF, Bahlmann FH, et al. Impaired endothelial repair capacity of early endothelial progenitor cells in prehypertension: relation to endothelial dysfunction. Hypertension. 2010;55:1389–97.

LETTER

doi:10.1038/nature12873

Genetics of rheumatoid arthritis contributes to biology and drug discovery

A list of authors and their affiliations appears at the end of the paper

A major challenge in human genetics is to devise a systematic strategy to integrate disease-associated variants with diverse genomic and biological data sets to provide insight into disease pathogenesis and guide drug discovery for complex traits such as rheumatoid arthritis (RA)[1]. Here we performed a genome-wide association study meta-analysis in a total of >100,000 subjects of European and Asian ancestries (29,880 RA cases and 73,758 controls), by evaluating ~10 million single-nucleotide polymorphisms. We discovered 42 novel RA risk loci at a genome-wide level of significance, bringing the total to 101 (refs 2–4). We devised an *in silico* pipeline using established bioinformatics methods based on functional annotation[5], *cis*-acting expression quantitative trait loci[6] and pathway analyses[7–9]—as well as novel methods based on genetic overlap with human primary immunodeficiency, haematological cancer somatic mutations and knockout mouse phenotypes—to identify 98 biological candidate genes at these 101 risk loci. We demonstrate that these genes are the targets of approved therapies for RA, and further suggest that drugs approved for other indications may be repurposed for the treatment of RA. Together, this comprehensive genetic study sheds light on fundamental genes, pathways and cell types that contribute to RA pathogenesis, and provides empirical evidence that the genetics of RA can provide important information for drug discovery.

We conducted a three-stage trans-ethnic meta-analysis (Extended Data Fig. 1). On the basis of the polygenic architecture of RA[10] and shared genetic risk among different ancestry[3,4], we proposed that combining a genome-wide association study (GWAS) of European and Asian ancestry would increase power to detect novel risk loci. In stage 1, we combined 22 GWAS for 19,234 cases and 61,565 controls of European and Asian ancestry[2–4]. We performed trans-ethnic, European-specific and Asian-specific GWAS meta-analysis by evaluating ~10 million single-nucleotide polymorphisms (SNPs)[11]. Characteristics of the cohorts, genotyping platforms and quality control criteria are described in Extended Data Table 1 (overall genomic control inflation factor $\lambda_{GC} < 1.075$).

Stage 1 meta-analysis identified 57 loci that satisfied a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$, including 17 novel loci (Extended Data Fig. 2). We then conducted a two-step replication study (stage 2 for *in silico* and stage 3 for *de novo*) in 10,646 RA cases and 12,193 controls for the loci with $P < 5.0 \times 10^{-6}$ in stage 1. In a combined analysis of stages 1–3, we identified 42 novel loci with $P < 5.0 \times 10^{-8}$ in any of the trans-ethnic, European or Asian meta-analyses. This increases the total number of RA risk loci to 101 (Table 1 and Supplementary Table 1).

Comparison of 101 RA risk loci revealed significant correlations of risk allele frequencies (RAFs) and odds ratios (ORs) between Europeans and Asians (Extended Data Fig. 3a–c; Spearman's $\rho = 0.67$ for RAF and 0.76 for OR; $P < 1.0 \times 10^{-13}$), although five loci demonstrated population-specific associations ($P < 5.0 \times 10^{-8}$ in one population but $P > 0.05$ in the other population without overlap of the 95% confidence intervals (95% CIs) of the ORs). In the population-specific genetic risk model, the 100 RA risk loci outside of the major histocompatibility complex (MHC) region[12] explained 5.5% and 4.7% of heritability in Europeans and Asians, respectively, with 1.6% of the heritability explained by the novel loci. The trans-ethnic genetic risk model, based on the RAF from one population but the OR from the other population, could explain the majority (>80%) of the known heritability in each population (4.7% for Europeans and 3.8% for Asians). These observations support our hypothesis that the genetic risk of RA is shared, in general, among Asians and Europeans.

We assessed enrichment of 100 non-MHC RA risk loci in epigenetic chromatin marks[13] (Extended Data Fig. 3d). Of 34 cell types investigated, we observed significant enrichment of RA risk alleles with trimethylation of histone H3 at lysine 4 (H3K4me3) peaks in primary $CD4^+$ regulatory T cells (T_{reg} cells; $P < 1.0 \times 10^{-5}$). For the RA risk loci enriched with T_{reg} H3K4me3 peaks, we incorporated the epigenetic annotations along with trans-ethnic differences in patterns of linkage disequilibrium to fine-map putative causal risk alleles (Extended Data Fig. 3e, f).

We found that approximately two-thirds of RA risk loci demonstrated pleiotropy with other human phenotypes (Extended Data Fig. 4), including immune-related diseases (for example, vitiligo, primary biliary cirrhosis), inflammation-related or haematological biomarkers (for example, fibrinogen, neutrophil counts) and other complex traits (for example, cardiovascular diseases).

Each of 100 non-MHC RA risk loci contains on average ~4 genes in the region of linkage disequilibrium (in total 377 genes). To prioritize systematically the most likely biological candidate gene, we devised an *in silico* bioinformatics pipeline. In addition to the published methods that integrate data across associated loci[7,8], we evaluated several biological data sets to test for enrichment of RA risk genes, which helps to pinpoint a specific gene in each loci (Extended Data Figs 5, 6 and Supplementary Tables 2–4).

We first conducted functional annotation of RA risk SNPs. Sixteen per cent of SNPs were in linkage disequilibrium with missense SNPs ($r^2 > 0.80$; Extended Data Fig. 5a, b). The proportion of missense RA risk SNPs was higher compared with a set of genome-wide common SNPs (8.0%), and relatively much higher in the explained heritability (~26.8%). Using *cis*-acting expression quantitative trait loci (*cis*-eQTL) data obtained from peripheral blood mononuclear cells (5,311 individuals)[6] and from $CD4^+$ T cells and $CD14^+CD16^-$ monocytes (212 individuals), we found that RA risk SNPs in 44 loci showed *cis*-eQTL effects (false discovery rate (FDR) q or permutation $P < 0.05$; Extended Data Table 2).

Second, we evaluated whether genes from RA risk loci overlapped with human primary immunodeficiency (PID) genes[14], and observed significant overlap (14/194 = 7.2%, $P = 1.2 \times 10^{-4}$; Fig. 1a and Extended Data Fig. 5c). Classification categories of PID genes showed different patterns of overlap: the highest proportion of overlap was in 'immune dysregulation' (4/21 = 19.0%, $P = 0.0033$) but there was no overlap in 'innate immunity'.

Third, we evaluated overlap with cancer somatic mutation genes[15], under the hypothesis that genes with cell growth advantages may contribute to RA development. Among 444 genes with registered cancer somatic mutations[15], we observed significant overlap with genes implicated in haematological cancers (17/251 = 6.8%, $P = 1.2 \times 10^{-4}$; Fig. 1b and Extended Data Fig. 5d), but not with genes implicated in non-haematological cancers (6/221 = 2.7%, $P = 0.56$).

Table 1 | Novel rheumatoid arthritis risk loci identified by trans-ethnic GWAS meta-analysis in >100,000 subjects

SNP	Chr	Genes	A1/A2 (+)	Trans-ethnic OR (95% CI)	Trans-ethnic P	European OR (95% CI)	European P	Asian OR (95% CI)	Asian P
rs227163	1	*TNFRSF9*	C/T	1.04 (1.02–1.06)	3.9×10^{-4}	1.00 (0.97–1.03)	9.3×10^{-1}	1.11 (1.08–1.16)*	3.1×10^{-9}*
rs28411352	1	*MTF1-INPP5B*	T/C	1.11 (1.08–1.14)*	2.8×10^{-12}*	1.10 (1.07–1.14)*	5.9×10^{-9}*	1.12 (1.06–1.19)	7.8×10^{-5}
rs2105325	1	*LOC100506023*	C/A	1.12 (1.08–1.15)*	6.9×10^{-13}*	1.12 (1.08–1.15)*	3.3×10^{-11}*	1.13 (1.04–1.23)	5.2×10^{-3}
rs10175798	2	*LBH*	A/G	1.08 (1.06–1.11)*	1.1×10^{-9}*	1.09 (1.06–1.12)*	4.2×10^{-8}*	1.07 (1.02–1.13)	6.4×10^{-3}
rs6732565	2	*ACOXL*	A/G	1.07 (1.05–1.10)*	2.7×10^{-8}*	1.10 (1.07–1.14)*	9.4×10^{-9}*	1.04 (1.00–1.08)	4.0×10^{-2}
rs6715284	2	*CFLAR-CASP8*	G/C	1.15 (1.10–1.20)*	1.8×10^{-9}*	1.15 (1.10–1.20)*	2.5×10^{-9}*	-	-
rs4452313	3	*PLCL2*	T/A	1.09 (1.06–1.12)*	1.6×10^{-10}*	1.11 (1.08–1.15)*	5.2×10^{-11}*	1.04 (0.99–1.09)	9.2×10^{-2}
rs3806624	3	*EOMES*	G/A	1.08 (1.05–1.11)*	8.6×10^{-9}*	1.08 (1.05–1.12)*	2.8×10^{-8}*	1.06 (0.99–1.14)	1.0×10^{-1}
rs9826828	3	*IL20RB*	A/G	1.44 (1.28–1.61)*	8.6×10^{-10}*	1.44 (1.28–1.61)*	8.7×10^{-10}*	-	-
rs13142500	4	*CLNK*	C/T	1.10 (1.07–1.13)*	3.0×10^{-9}*	1.10 (1.06–1.15)	2.4×10^{-6}	1.10 (1.04–1.15)	2.8×10^{-4}
rs2664035	4	*TEC*	A/G	1.07 (1.04–1.10)	9.5×10^{-8}	1.08 (1.05–1.11)*	3.3×10^{-8}*	1.03 (0.97–1.08)	3.3×10^{-1}
rs9378815	6	*IRF4*	C/G	1.09 (1.06–1.12)*	1.7×10^{-10}*	1.09 (1.05–1.12)	1.4×10^{-7}	1.10 (1.04–1.15)	2.3×10^{-4}
rs2234067	6	*ETV7*	C/A	1.15 (1.10–1.20)*	1.6×10^{-9}*	1.14 (1.09–1.19)*	4.1×10^{-8}*	1.22 (1.06–1.41)	7.0×10^{-3}
rs9373594	6	*PPIL4*	T/C	1.09 (1.06–1.12)*	3.0×10^{-9}*	1.07 (1.02–1.12)	6.5×10^{-3}	1.11 (1.07–1.15)*	4.8×10^{-8}*
rs67250450	7	*JAZF1*	T/C	1.10 (1.07–1.14)*	3.7×10^{-9}*	1.11 (1.07–1.14)*	2.6×10^{-9}*	1.02 (0.84–1.23)	8.5×10^{-1}
rs4272	7	*CDK6*	G/A	1.10 (1.06–1.13)*	5.0×10^{-9}*	1.10 (1.07–1.14)*	1.2×10^{-8}*	1.06 (0.98–1.15)	1.3×10^{-1}
rs998731	8	*TPD52*	T/C	1.08 (1.05–1.11)*	1.9×10^{-8}*	1.09 (1.06–1.12)*	6.6×10^{-9}*	1.02 (0.96–1.10)	4.9×10^{-1}
rs678347	8	*GRHL2*	G/A	1.08 (1.05–1.11)*	1.6×10^{-8}*	1.10 (1.06–1.13)*	7.3×10^{-9}*	1.03 (0.98–1.10)	2.6×10^{-1}
rs1516971	8	*PVT1*	T/C	1.15 (1.10–1.20)*	1.3×10^{-10}*	1.16 (1.11–1.21)*	3.2×10^{-11}*	-	-
rs12413578	10	10p14	C/T	1.20 (1.13–1.29)*	4.8×10^{-8}*	1.20 (1.12–1.29)	7.5×10^{-8}	-	-
rs793108	10	*ZNF438*	T/C	1.08 (1.05–1.10)*	1.3×10^{-9}*	1.07 (1.04–1.10)	6.1×10^{-7}	1.09 (1.04–1.14)	4.4×10^{-4}
rs2671692	10	*WDFY4*	A/G	1.07 (1.05–1.10)*	2.8×10^{-9}*	1.06 (1.03–1.09)	2.6×10^{-5}	1.10 (1.05–1.14)	9.9×10^{-6}
rs726288	10	*SFTPD*	T/C	1.14 (1.07–1.20)	1.6×10^{-5}	0.96 (0.86–1.06)	4.1×10^{-1}	1.22 (1.14–1.31)*	8.8×10^{-9}*
rs968567	11	*FADS1-FADS2-FADS3*	C/T	1.12 (1.07–1.16)*	1.8×10^{-8}*	1.12 (1.07–1.16)*	1.8×10^{-8}*	-	-
rs4409785	11	*CEP57*	C/T	1.12 (1.09–1.16)*	1.2×10^{-11}*	1.12 (1.08–1.16)*	3.6×10^{-9}*	1.16 (1.07–1.27)	4.3×10^{-4}
chr11:107967350	11	*ATM*	A/G	1.21 (1.13–1.29)*	1.4×10^{-8}*	1.21 (1.13–1.29)*	1.1×10^{-8}*	-	-
rs73013527	11	*ETS1*	C/T	1.09 (1.06–1.12)*	1.2×10^{-10}*	1.08 (1.05–1.11)	1.0×10^{-6}	1.14 (1.08–1.21)	4.1×10^{-6}
rs773125	12	*CDK2*	A/G	1.09 (1.06–1.12)*	1.1×10^{-10}*	1.09 (1.06–1.12)*	2.1×10^{-8}*	1.10 (1.04–1.17)	1.1×10^{-3}
rs10774624	12	*SH2B3-PTPN11*	G/A	1.09 (1.06–1.13)*	6.8×10^{-9}*	1.09 (1.06–1.13)*	6.9×10^{-9}*	-	-
rs9603616	13	*COG6*	C/T	1.10 (1.07–1.13)*	1.6×10^{-12}*	1.11 (1.07–1.14)*	2.8×10^{-11}*	1.08 (1.02–1.14)	1.0×10^{-2}
rs3783782	14	*PRKCH*	A/G	1.14 (1.09–1.18)*	2.2×10^{-9}*	1.12 (0.96–1.31)	1.4×10^{-1}	1.14 (1.09–1.19)*	4.4×10^{-9}*
rs1950897	14	*RAD51B*	T/C	1.10 (1.07–1.13)*	8.2×10^{-11}*	1.09 (1.06–1.12)*	5.0×10^{-8}*	1.16 (1.08–1.25)	1.1×10^{-4}
rs4780401	16	*TXNDC11*	T/G	1.07 (1.05–1.10)*	4.1×10^{-8}*	1.09 (1.06–1.13)*	8.7×10^{-9}*	1.03 (0.98–1.08)	2.5×10^{-1}
rs72634030	17	*C1QBP*	A/C	1.12 (1.08–1.17)*	1.5×10^{-9}*	1.12 (1.06–1.19)	2.9×10^{-5}	1.12 (1.07–1.18)	9.6×10^{-6}
rs1877030	17	*MED1*	C/T	1.09 (1.06–1.12)*	1.9×10^{-8}*	1.09 (1.05–1.13)	1.3×10^{-5}	1.09 (1.04–1.14)	3.2×10^{-4}
rs2469434	18	*CD226*	C/T	1.07 (1.05–1.10)*	8.9×10^{-10}*	1.05 (1.02–1.08)	6.7×10^{-4}	1.11 (1.07–1.15)*	1.2×10^{-8}*
chr19:10771941	19	*ILF3*	C/T	1.47 (1.30–1.67)*	8.6×10^{-10}*	1.47 (1.30–1.67)*	8.8×10^{-10}*	-	-
rs73194058	21	*IFNGR2*	C/A	1.08 (1.05–1.12)	1.2×10^{-6}	1.13 (1.08–1.18)*	2.6×10^{-8}*	1.03 (0.98–1.08)	2.9×10^{-1}
rs1893592	21	*UBASH3A*	A/C	1.11 (1.08–1.14)*	7.2×10^{-12}*	1.11 (1.07–1.15)*	9.8×10^{-9}*	1.11 (1.05–1.18)	1.3×10^{-4}
rs11089637	22	*UBE2L3-YDJC*	C/T	1.08 (1.05–1.11)*	2.1×10^{-9}*	1.10 (1.06–1.15)	2.0×10^{-7}	1.06 (1.02–1.10)	8.9×10^{-4}
rs909685	22	*SYNGR1*	A/T	1.13 (1.10–1.16)*	1.4×10^{-16}*	1.11 (1.08–1.15)*	6.4×10^{-12}*	1.23 (1.14–1.33)	2.0×10^{-7}
chrX:78464616	X	*P2RY10*	A/C	1.11 (1.07–1.15)*	3.5×10^{-8}*	1.16 (0.78–1.75)	4.6×10^{-1}	1.11 (1.07–1.15)*	3.6×10^{-8}*

SNPs newly associated with $P < 5.0 \times 10^{-8}$ in the combined study of the stage 1 GWAS meta-analysis and the stages 2 and 3 replication studies of trans-ethnic (Europeans and Asians), European or Asian ancestry are indicated. SNPs, positions and alleles are based on the positive (+) strand of NCBI build 37. A1 represents an RA risk allele. Chr, chromosome; OR, odds ratio; 95% CI, 95% confidence interval. Full results of the studies are available in Supplementary Table 1. Hyphens between gene names indicate that several candidate RA risk genes were included in the region.
*Association results with $P < 5.0 \times 10^{-8}$.

Fourth, we evaluated overlap with genes implicated in knockout mouse phenotypes[16]. Among the 30 categories of phenotypes[16], we observed 3 categories significantly enriched with RA risk genes ($P < 0.05/30 = 0.0017$): 'haematopoietic system phenotype', 'immune system phenotype', and 'cellular phenotype' (Extended Data Fig. 5e).

Last, we conducted molecular pathway enrichment analysis (Fig. 1c and Extended Data Fig. 5f). We observed enrichment (FDR $q < 0.05$) for T-cell-related pathways, consistent with cell-specific epigenetic marks, as well as enrichment for B-cell and cytokine signalling pathways (for example, interleukin (IL)-10, interferon, granulocyte–macrophage colony-stimulating factor (GM-CSF)). For comparison, our previous RA GWAS meta-analysis[2] did not identify the B-cell and cytokine signalling pathways, thereby indicating that as more loci are discovered, further biological pathways are identified.

On the basis of these new findings, we adopted the following 8 criteria to prioritize each of the 377 genes from the 100 non-MHC RA risk loci (Fig. 2 and Extended Data Fig. 6a–c): (1) genes with RA risk missense variant ($n = 19$); (2) *cis*-eQTL genes ($n = 51$); (3) genes prioritized by PubMed text mining[7] ($n = 90$); (4) genes prioritized by protein–protein interaction (PPI)[8] ($n = 63$); (5) PID genes ($n = 15$); (6) haematological cancer somatic mutation genes ($n = 17$); (7) genes prioritized by associated knockout mouse phenotypes ($n = 86$); and (8) genes prioritized by molecular pathway analysis[9] ($n = 35$).

Ninety-eight genes (26.0%) had a score ≥2, which we defined as 'candidate biological RA risk genes'. Nineteen loci included multiple biological RA risk genes (for example, *IL3* and *CSF2* at chromosome 5q31), whereas no biological gene was selected from 40 loci (Supplementary Table 5).

To provide empirical evidence of the pipeline, we evaluated relationships of the gene scores to independent genomic or epigenetic information. Genes with higher biological scores were more likely to be the nearest gene to the risk SNP (18.6% for gene score <2 and 49.0% for gene score ≥2; $P = 2.1 \times 10^{-8}$), and also to be included in the region where RA risk SNPs were overlapping with H3K4me3 T_{reg} peaks (41.9% for gene score <2 and 57.1% for gene score ≥2; $P = 0.034$). Further, T_{reg} cells demonstrated the largest increase in overlapping proportions with H3K4me3 peaks for increase of biological gene scores compared with other cell types (Extended Data Fig. 6d).

Finally, we evaluated the potential role of RA genetics in drug discovery. We proposed that if human genetics is useful for drug target validation, then it should identify existing approved drugs for RA. To test this 'therapeutic hypothesis'[1], we obtained 871 drug target genes corresponding to approved, in clinical trials or experimental drugs for human diseases[17,18] (Supplementary Table 6). We evaluated whether any of the protein products from the identified biological RA risk genes, or any genes from a direct PPI network with such protein products

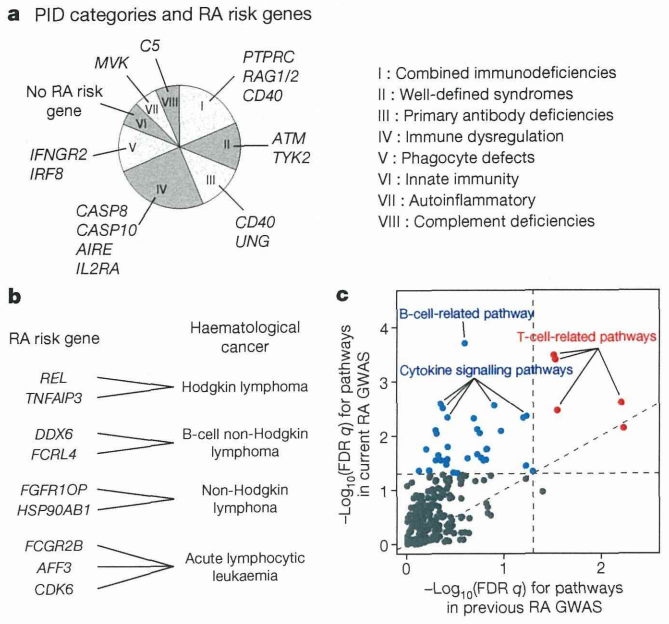

Figure 1 | Overlap of RA risk loci with PID genes, haematological cancer somatic mutations and molecular pathways. a, Overlap of RA risk genes with PID genes, subdivided by PID categories (I–VIII). **b**, Examples of overlap of haematological cancer somatic mutation genes with RA risk genes. **c**, Comparisons of molecular pathway analysis results between the current trans-ethnic meta-analysis (*y*-axis) and the previous meta-analysis for RA (*x*-axis)[2]. Each dot represents a molecular pathway. Dotted line represents FDR $q = 0.05$ or $y = x$.

(Fig. 3a), are the pharmacologically active targets of approved RA drugs (Extended Data Fig. 7a).

Twenty-seven drug target genes of approved RA drugs demonstrated significant overlap with 98 biological RA risk genes and 2,332 genes from the expanded PPI network (18 genes overlapped; 3.7-fold enrichment by permutation analysis, $P < 1.0 \times 10^{-5}$; Fig. 3b). For comparison, all drug target genes (regardless of disease indication) overlapped with 247 genes, which is 1.7-fold more enrichment than expected by chance, but less than 2.2-fold enrichment compared with overlap of the target genes of RA drugs ($P = 0.0035$). Examples of approved RA therapies identified by this analysis include tocilizumab[19,20] (anti-IL6R), tofacitinib[21] (JAK3 inhibitor) and abatacept[21] (CTLA4–immunoglobulin; Fig. 3c and Extended Data Fig. 8).

We also assessed how approved drugs for other diseases might be connected to biological RA risk genes. We highlight *CDK6* and *CDK4*, targets of three approved drugs for different types of cancer[22] (Fig. 3d).

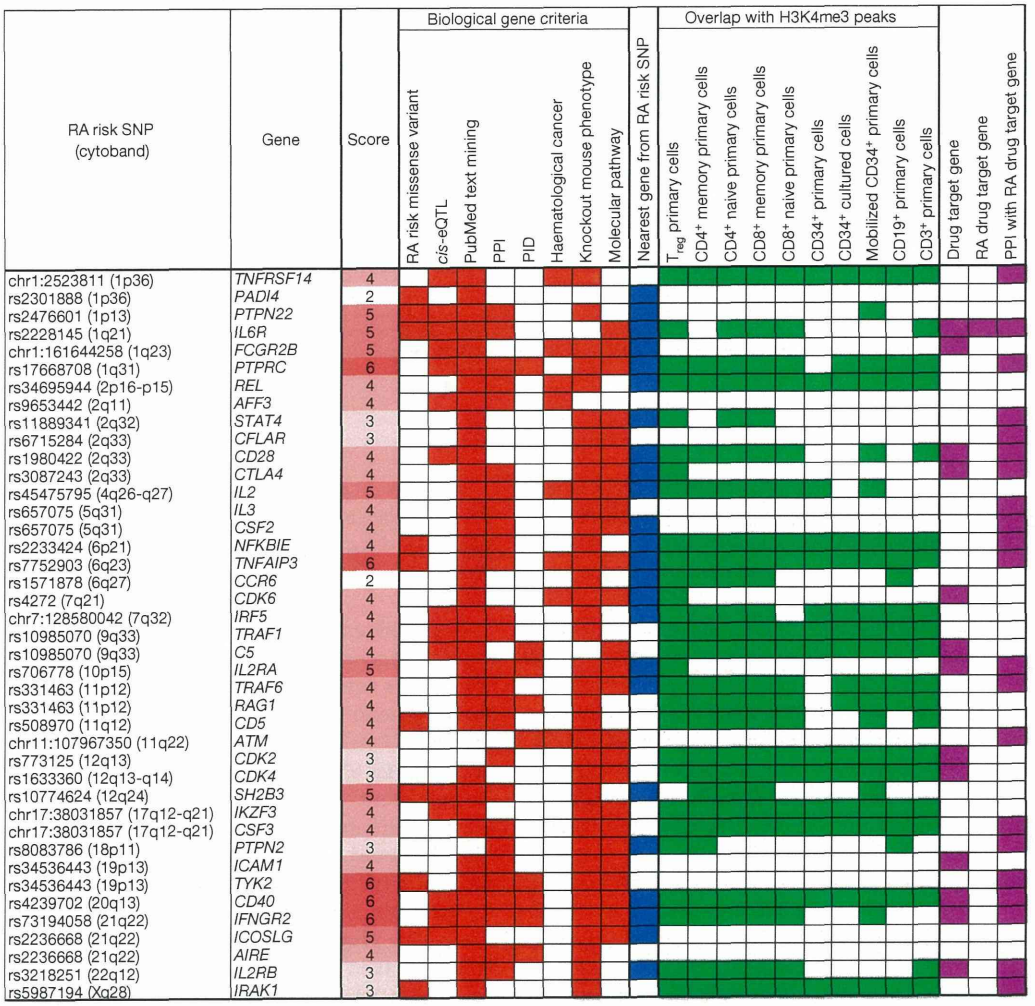

Figure 2 | Prioritized biological RA risk genes. Representative biological RA risk genes. We list the summary gene score derived from individual criteria (filled red box indicates criterion satisfied; 98 genes with a score ≥2 out of 377 genes included in the RA risk loci were defined as 'biological candidate genes'; see Extended Data Fig. 6). Filled blue boxes indicate the nearest gene to the RA risk SNP. Filled green boxes indicate overlap with H3K4me3 peaks in immune-related cells. Filled purple boxes indicate overlap with drug target genes. For full results, see Supplementary Table 5.

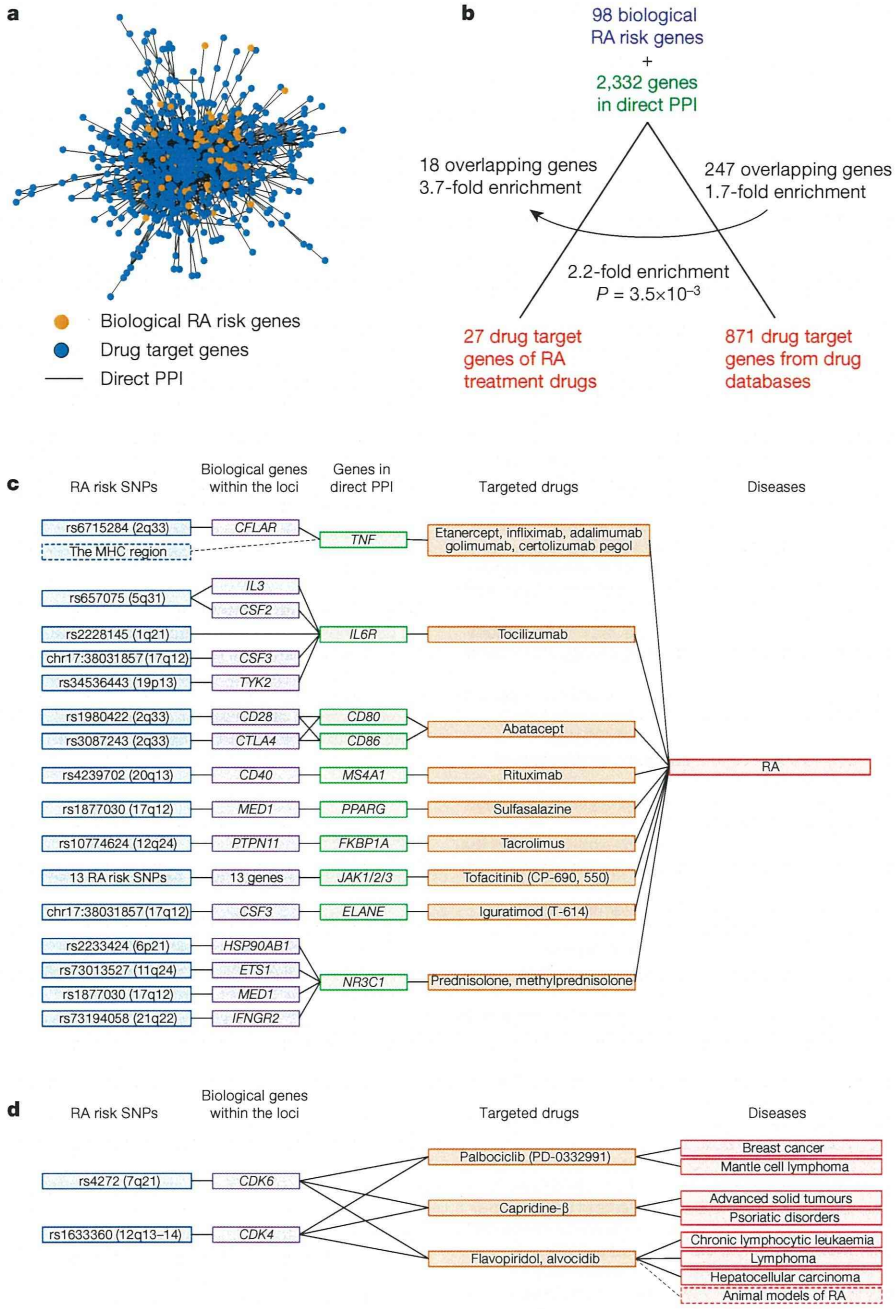

Figure 3 | Connection of biological RA risk genes to drug targets. **a**, PPI network of biological RA risk genes and drug target genes. **b**, Overlap and relative enrichment of 98 biological RA risk genes with targets of approved RA drugs and with all drug target genes. Enrichment was more apparent than that from all 377 RA risk genes (Extended Data Fig. 7c). **c**, Connections between RA risk SNPs (blue), biological genes (purple), genes from PPI (green) and approved RA drugs (orange). For full results, see Extended Data Fig. 8. **d**, Connections between RA genes and drugs indicated for other diseases.

In support for repurposing, one *CDK6/CDK4* inhibitor, flavopiridol, has been shown to ameliorate disease activity in animal models of RA[22]. Further, the biology is plausible, as several approved RA drugs were initially developed for cancer treatment and then repurposed for RA (for example, rituximab). Although further investigations are necessary, we propose that target genes/drugs selected by this approach could represent promising candidates for novel drug discovery for RA treatment.

We note that a non-random distribution of drug-to-disease indications in the databases could potentially bias our results. Namely, because RA risk genes are enriched for genes with immune function, spurious enrichment with drug targets could occur if the majority of drug indications in databases were for immune-mediated diseases or immune-related target genes. However, such enrichment was not evident in our analysis (~11% for drug indications and ~9% for target genes; Extended Data Fig. 7b).

Through a comprehensive genetic study with >100,000 subjects, we identified 42 novel RA risk loci and provided novel insight into RA pathogenesis. We particularly highlight the role of genetics for drug discovery. Although there have been anecdotal examples of this[1,23], our study provides a systematic approach by which human genetic data can be efficiently integrated with other biological information to derive biological insights and drive drug discovery.

METHODS SUMMARY

Details can be found in Methods, Extended Data Fig. 1, Extended Data Table 1 and Supplementary Information, including (1) information about the patient collections;

(2) genotyping, quality control and genotype imputation of GWAS data; (3) genome-wide meta-analysis (stage 1); (4) *in silico* and *de novo* replication studies (stages 2 and 3); (5) trans-ethnic and functional annotations of RA risk SNPs; (6) prioritization of biological candidate genes; and (7) drug target gene enrichment analysis.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 June; accepted 7 November 2013.
Published online 25 December 2013.

1. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nature Rev. Drug Discov.* **12,** 581–594 (2013).
2. Stahl, E. A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nature Genet.* **42,** 508–514 (2010).
3. Okada, Y. *et al.* Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nature Genet.* **44,** 511–516 (2012).
4. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nature Genet.* **44,** 1336–1340 (2012).
5. Ferreira, R. C. *et al.* Functional *IL6R* 358Ala allele impairs classical IL-6 receptor signaling and influences risk of diverse inflammatory diseases. *PLoS Genet.* **9,** e1003444 (2013).
6. Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genet.* **45,** 1238–1243 (2013).
7. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5,** e1000534 (2009).
8. Rossin, E. J. *et al.* Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet.* **7,** e1001273 (2011).
9. Segrè, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
10. Stahl, E. A. *et al.* Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nature Genet.* **44,** 483–489 (2012).
11. 1000 Genomes Project Consortium *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491,** 56–65 (2012).
12. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nature Genet.* **44,** 291–296 (2012).
13. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nature Genet.* **45,** 124–130 (2013).
14. Parvaneh, N., Casanova, J. L., Notarangelo, L. D. & Conley, M. E. Primary immunodeficiencies: a rapidly evolving story. *J. Allergy Clin. Immunol.* **131,** 314–323 (2013).
15. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39,** D945–D950 (2011).
16. Eppig, J. T., Blake, J. A., Bult, C. J., Kadin, J. A. & Richardson, J. E. The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. *Nucleic Acids Res.* **40,** D881–D886 (2012).
17. Knox, C. *et al.* DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.* **39,** D1035–D1041 (2011).
18. Zhu, F. *et al.* Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* **40,** D1128–D1136 (2012).
19. Smolen, J. S. *et al.* Consensus statement on blocking the effects of interleukin-6 and in particular by interleukin-6 receptor inhibition in rheumatoid arthritis and other inflammatory conditions. *Ann. Rheum. Dis.* **72,** 482–492 (2013).
20. Nishimoto, N. *et al.* Study of active controlled tocilizumab monotherapy for rheumatoid arthritis patients with an inadequate response to methotrexate (SATORI): significant reduction in disease activity and serum vascular endothelial growth factor by IL-6 receptor inhibition therapy. *Mod. Rheumatol.* **19,** 12–19 (2009).
21. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. *N. Engl. J. Med.* **365,** 2205–2219 (2011).
22. Sekine, C. *et al.* Successful treatment of animal models of rheumatoid arthritis with small-molecule cyclin-dependent kinase inhibitors. *J. Immunol.* **180,** 1954–1961 (2008).
23. Sanseau, P. *et al.* Use of genome-wide association studies for drug repositioning. *Nature Biotechnol.* **30,** 317–320 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements R.M.P. is supported by National Institutes of Health (NIH) grants R01-AR057108, R01-AR056768, U01-GM092691 and R01-AR059648, and holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund. Y.O. is supported by a grant from the Japan Society of the Promotion of Science. D.W. is supported by a grant from the Australian National Health and Medical Research Council (1036541). G.T. is supported by the Rubicon grant from the Netherlands Organization for Scientific Research. A.Z. is supported by a grant from the Dutch Reumafonds (11-1-101) and from the Rosalind Franklin Fellowship, University of Groningen. S.-C.B., S.-Y.B. and H.-S.L. are supported by the Korea Healthcare technology R&D project, Ministry for Health and Welfare (A121983). J.M., M.A.G.-G. and L.R.-R. are funded by the RETICS program, RIER, RD12/0009 from the Instituto de Salud Carlos III, Health Ministry. S.R.-D. and L.Á.'s work is supported by the Medical Biobank of Northern Sweden. H.K.C. is supported by NIH (NIAMS) grants R01-AR056291, R01-AR065944, R01-AR056768, P60 AR047785 and R21 AR056042. L.P. and L.K. are supported by a senior investigator grant from the European Research Council. S.R. is supported by NIH grants R01AR063759-01A1 and K08-KAR055688A. P.M.V. is a National Health and Medical Research Council Senior Principal Research Fellow. M.A.B. is funded by the National Health and Medical Research Foundation Senior Principal Research Fellowship, and a Queensland State Government Premier's Fellowship. H.X. is funded by the China Ministry of Science and Technology (973 program grant 2011CB946100), the National Natural Science Foundation of China (grants 30972339, 81020108029 and 81273283), and the Science and Technology Commission of Shanghai Municipality (grants 08XD1400400, 11410701600 and 10JC1418400). K.A.S. is supported by a Canada Research Chair, The Sherman Family Chair in Genomics Medicine, Canadian Institutes for Health Research grant 79321 and Ontario Research Fund grant 05-075. S.M. is supported by Health and Labour Sciences Research Grants. The BioBank Japan Project is supported by the Ministry of Education, Culture, Sports, Science and Technology of the Japanese government. This study is supported by the BE THE CURE (BTCure) project. We thank K. Akari, K. Tokunaga and N. Nishida for supporting the study.

Author Contributions Y.O. carried out the primary data analyses. D.W. managed drug target gene data. G.T. conducted histone mark analysis. T.R., H.-J.W., T.E., A.M., B.E.S., P.L.D. and L.F. conducted eQTL analysis. C.T., K.I., Y.K., K.O., A.S., S.Y., G.X., E.K. and K.A.S. conducted the *de novo* replication study. R.R.G., A.M., W.O., T.B., T.W.B., L.J., J. Yin, L.Y., D.-F.S., J. Yang, P.M.V., M.A.B. and H.X. conducted the *in silico* replication study. E.A.S., D.D., J.C., T.K., R.Y. and A.T. managed GWAS data. All other authors, as well as the members of the RACI and GARNET consortia, contributed to additional analyses and genotype and clinical data enrolments. Y.O. and R.M.P. designed the study and wrote the manuscript, with contributions from all authors on the final version of the manuscript.

Author Information Summary statistics from the GWAS meta-analysis, source codes, and data sources used in this study are available at http://plaza.umin.ac.jp/~yokada/datasource/software.htm. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.M.P. (robert.plenge@merck.com) or Y.O. (yokada.brc@tmd.ac.jp).

Yukinori Okada[1,2,3], Di Wu[1,2,3,4,5], Gosia Trynka[1,2,3], Towfique Raj[2,3,6], Chikashi Terao[7,8], Katsunori Ikari[9], Yuta Kochi[10], Koichiro Ohmura[8], Akari Suzuki[10], Shinji Yoshida[9], Robert R. Graham[11], Arun Manoharan[11], Ward Ortmann[11], Tushar Bhangale[11], Joshua C. Denny[12,13], Robert J. Carroll[12], Anne E. Eyler[13], Jeffrey D. Greenberg[14], Joel M. Kremer[15], Dimitrios A. Pappas[16], Lei Jiang[17], Jian Yin[17], Lingying Ye[17], Ding-Feng Su[18], Jian Yang[19,20], Gang Xie[21,22,23], Ed Keystone[24], Harm-Jan Westra[25], Tõnu Esko[3,26,27], Andres Metspalu[26], Xuezhong Zhou[28], Namrata Gupta[3], Daniel Mirel[3], Eli A. Stahl[29], Dorothée Diogo[1,2,3], Jing Cui[1,2,3], Katherine Liao[1,2,3], Michael H. Guo[1,3,27], Keiko Myouzen[10], Takahisa Kawaguchi[7], Marieke J. H. Coenen[30], Piet L. C. M. van Riel[31], Mart A. F. J. van de Laar[32], Henk-Jan Guchelaar[33], Tom W. J. Huizinga[34], Philippe Dieudé[35,36], Xavier Mariette[37], S. Louis Bridges Jr[38], Alexandra Zhernakova[25,34], Rene E. M. Toes[34], Paul P. Tak[39,40,41], Corinne Miceli-Richard[37], So-Young Bang[42], Hye-Soon Lee[42], Javier Martin[43], Miguel A. Gonzalez-Gay[44], Luis Rodriguez-Rodriguez[45], Solbritt Rantapää-Dahlqvist[46,47], Lisbeth Ärlestig[46,47], Hyon K. Choi[48,49,50], Yoichiro Kamatani[51], Pilar Galan[52], Mark Lathrop[53], the RACI consortium†, the GARNET consortium†, Steve Eyre[54,55], John Bowes[54,55], Anne Barton[54], Niek de Vries[56], Larry W. Moreland[57], Lindsey A. Criswell[58], Elizabeth W. Karlson[1], Atsuo Taniguchi[9], Ryo Yamada[59], Michiaki Kubo[60], Jun S. Liu[4], Sang-Cheol Bae[42], Jane Worthington[54,55], Leonid Padyukov[61], Lars Klareskog[61], Peter K. Gregersen[62], Soumya Raychaudhuri[1,2,3,63], Barbara E. Stranger[64,65], Philip L. De Jager[2,3,6], Lude Franke[25], Peter M. Visscher[19,20], Matthew A. Brown[19], Hisashi Yamanaka[9], Tsuneyo Mimori[8], Atsushi Takahashi[66], Huji Xu[17], Timothy W. Behrens[11], Katherine A. Siminovitch[21,22,23], Shigeki Momohara[9], Fumihiko Matsuda[7,67,68], Kazuhiko Yamamoto[10,69] & Robert M. Plenge[1,2,3]

[1]Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [2]Division of Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. [3]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [4]Department of Statistics, Harvard University, Cambridge, Massachusetts 02138, USA. [5]Centre for Cancer Research, Monash Institute of Medical Research, Monash University, Clayton, Victoria 3800, Australia. [6]Program in Translational NeuroPsychiatric Genomics, Institute for the Neurosciences, Department of Neurology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA. [7]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [8]Department of Rheumatology and Clinical immunology, Graduate School of Medicine, Kyoto University, Kyoto 606-8507, Japan. [9]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo 162-0054, Japan. [10]Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [11]Immunology Biomarkers Group, Genentech, South San Francisco, California 94080, USA. [12]Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. [13]Department of Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA. [14]New York University Hospital for Joint Diseases, New York, New York 10003, USA. [15]Department of Medicine, Albany Medical Center and The Center for Rheumatology, Albany, New York 12206, USA.

[16]Division of Rheumatology, Department of Medicine, New York, Presbyterian Hospital, College of Physicians and Surgeons, Columbia University, New York, New York 10032, USA. [17]Department of Rheumatology and Immunology, Shanghai Changzheng Hospital, Second Military Medical University, Shanghai 200003, China. [18]Department of Pharmacology, Second Military Medical University, Shanghai 200433, China. [19]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4072, Australia. [20]Queensland Brain Institute, The University of Queensland, Brisbane, Queensland 4072, Australia. [21]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. [22]Toronto General Research Institute, Toronto, Ontario M5G 2M9, Canada. [23]Department of Medicine, University of Toronto, Toronto, Ontario M5S 2J7, Canada. [24]Department of Medicine, Mount Sinai Hospital and University of Toronto, Toronto M5S 2J7, Canada. [25]Department of Genetics, University Medical Center Groningen, University of Groningen, Hanzeplein 1, Groningen 9700 RB, the Netherlands. [26]Estonian Genome Center, University of Tartu, Riia 23b, Tartu 51010, Estonia. [27]Division of Endocrinology, Children's Hospital, Boston, Massachusetts 02115, USA. [28]School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China. [29]The Department of Psychiatry at Mount Sinai School of Medicine, New York, New York 10029, USA. [30]Department of Human Genetics, Radboud University Medical Centre, Nijmegen 6500 HB, the Netherlands. [31]Department of Rheumatology, Radboud University Medical Centre, Nijmegen 6500 HB, the Netherlands. [32]Department of Rheumatology and Clinical Immunology, Arthritis Center Twente, University Twente & Medisch Spectrum Twente, Enschede 7500 AE, the Netherlands. [33]Department of Clinical Pharmacy and Toxicology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands. [34]Department of Rheumatology, Leiden University Medical Center, Leiden 2300 RC, the Netherlands. [35]Service de Rhumatologie et INSERM U699 Hôpital Bichat Claude Bernard, Assistance Publique des Hôpitaux de Paris, Paris 75018, France. [36]Université Paris 7-Diderot, Paris 75013, France. [37]Institut National de la Santé et de la Recherche Médicale (INSERM) U1012, Université Paris-Sud, Rhumatologie, Hôpitaux Universitaires Paris-Sud, Assistance Publique-Hôpitaux de Paris (AP-HP), Le Kremlin Bicêtre 94275, France. [38]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. [39]AMC/University of Amsterdam, Amsterdam 1105 AZ, the Netherlands. [40]GlaxoSmithKline, Stevenage SG1 2NY, UK. [41]University of Cambridge, Cambridge CB2 1TN, UK. [42]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 133-792, South Korea. [43]Instituto de Parasitologia y Biomedicina Lopez-Neyra, CSIC, Granada 18100, Spain. [44]Department of Rheumatology, Hospital Marques de Valdecilla, IFIMAV, Santander 39008, Spain. [45]Hospital Clinico San Carlos, Madrid 28040, Spain. [46]Department of Public Health and Clinical Medicine, Umeå University, Umeå SE-901 87, Sweden. [47]Department of Rheumatology, Umeå University, Umeå SE-901 87, Sweden. [48]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston 02115, Massachusetts, USA. [49]Section of Rheumatology, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [50]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, Massachusetts 02118, USA. [51]Centre d'Etude du Polymorphisme Humain (CEPH), Paris 75010, France. [52]Université Paris 13 Sorbonne Paris Cité, UREN (Nutritional Epidemiology Research Unit), Inserm (U557), Inra (U1125), Cnam, Bobigny 93017, France. [53]McGill University and Génome Québec Innovation Centre, Montréal, Québec H3A 0G1 Canada. [54]Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9NT, UK. [55]National Institute for Health Research, Manchester Musculoskeletal Biomedical Research Unit, Central Manchester University Hospitals National Health Service Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK. [56]Department of Clinical Immunology and Rheumatology & Department of Genome Analysis, Academic Medical Center/University of Amsterdam, Amsterdam 1105 AZ, the Netherlands. [57]Division of Rheumatology and Clinical Immunology, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, USA. [58]Rosalind Russell Medical Research Center for Arthritis, Division of Rheumatology, Department of Medicine, University of California San Francisco, San Francisco, California 94117, USA. [59]Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto 606-8507, Japan. [60]Laboratory for Genotyping Development, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [61]Rheumatology Unit, Department of Medicine (Solna), Karolinska Institutet, Stockholm SE-171 76, Sweden. [62]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York 11030, USA. [63]NIHR Manchester Musculoskeletal Biomedical, Research Unit, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester M13 9NT, UK. [64]Section of Genetic Medicine, University of Chicago, Chicago, Illinois 60637, USA. [65]Institute for Genomics and Systems Biology, University of Chicago, Chicago, Illinois 60637, USA. [66]Laboratory for Statistical Analysis, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan. [67]Core Research for Evolutional Science and Technology (CREST) program, Japan Science and Technology Agency, Kawaguchi, Saitama 332-0012, Japan. [68]Institut National de la Sante et de la Recherche Medicale (INSERM) Unite U852, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan. [69]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-0033, Japan. †Lists of participants and their affiliations appear in the Supplementary Information.

METHODS

Subjects. Our study included 29,880 RA cases (88.1% seropositive and 9.3% seronegative for anti-citrullinated peptide antibody (ACPA) or rheumatoid factor (RF), and 2.6% who had unknown autoantibody status) and 73,758 controls. All RA cases fulfilled the 1987 criteria of the American College of Rheumatology for RA diagnosis[24], or were diagnosed with RA by a professional rheumatologist. The 19,234 RA cases and 61,565 controls enrolled in the stage 1 trans-ethnic GWAS meta-analysis were obtained from 22 studies on people with European and Asian ancestries (14,361 RA cases and 43,923 controls from 18 studies of Europeans and 4,873 RA cases and 17,642 controls from 4 studies of Asians): BRASS[2], CANADA[2], EIRA[2], NARAC1[2], NARAC2[2], WTCCC[2], Rheumatoid Arthritis Consortium International for Immunochip (RACI)-UK[4], RACI-US[4], RACI-SE-E[4], RACI-SE-U[4], RACI-NL[4], RACI-ES[4], RACI-i2b2, ReAct, Dutch (including AMC, BeSt, LUMC and DREAM), anti-TNF response to therapy collection (ACR-REF: BRAGGSS, BRAGGSS2, ERA, KI and TEAR), CORRONA, Vanderbilt, three studies from the GARNET consortium (BioBank Japan Project[3], Kyoto University[3] and IORRA[3]), and Korea. Of these, GWAS data of 4,309 RA cases and 8,700 controls from six studies (RACI-i2b2, ReAct, Dutch, ACR-REF, CORRONA and Vanderbilt) have not been previously published.

The 3,708 RA cases and 5,535 controls enrolled in the stage 2 *in silico* replication study were obtained from two studies of Europeans (2,780 RA cases and 4,700 controls from Genentech and SLEGEN) and Asians (928 RA cases and 835 controls from China) (H.X. *et al.*, manuscript submitted). The 6,938 RA cases and 6,658 controls enrolled in the stage 3 *de novo* replication study were obtained from two studies of Europeans (995 RA cases and 1,101 controls from CANADAII[2]) and Asians (5,943 RA cases and 5,557 controls from BioBank Japan Project, Kyoto University and IORRA[3]).

All subjects in the stage 1, stage 2 and stage 3 studies were confirmed to be independent through analysis of overlapping SNP markers. Any duplicate subjects were removed from the stage 2 and stage 3 replication studies, leading to slightly different sample sizes compared with previous studies that used these same collections[2,3].

All participants provided written informed consent for participation in the study as approved by the ethical committees of each of the institutional review boards. Detailed descriptions of the study design, participating cohorts and the clinical characteristics of the RA cases are provided in detail in Extended Data Fig. 1 and Extended Data Table 1a, as well as in previous reports[2–4].

Genotyping, quality control and genotype imputation of GWAS data. Genotyping platforms and quality control criteria of GWAS, including cut-off values for sample call rate, SNP call rate, minor allele frequency (MAF), and Hardy–Weinberg equilibrium (HWE) *P* value, covariates in the analysis, and imputation reference panel information are provided for each study in Extended Data Table 1b. All studies were analysed based on the same analytical protocol, including exclusion of closely related subjects and outliers in terms of ancestries, as described elsewhere[3]. After applying quality control criteria, whole-genome genotype imputation was performed using 1000 Genomes Project Phase I (α) European ($n = 381$) and Asian ($n = 286$) data as references[11]. We excluded monomorphic or singleton SNPs or SNPs with deviation of HWE ($P < 1.0 \times 10^{-7}$) from each of the reference panels. GWAS data were split into ~300 chunks that evenly covered whole-genome regions and additionally included 300 kb of duplicated regions between neighbouring chunks. Immunochip data were split into ~2,000 chunks that included each of the targeted regions or SNPs on the array. Each chunk was pre-phased and imputed by using minimac (release stamp 2011-10-27). SNPs in the X chromosome were imputed for males and females separately. We excluded imputed SNPs that were duplicated between chunks, SNPs with MAF < 0.005 in RA cases or controls, or with low imputation score (Rsq < 0.5 for genome-wide array and < 0.7 for Immunochip) from each study. We found that imputation of Immunochip effectively increased the number of the available SNPs by 7.0 fold (from ~129,000 SNPs to ~924,000 SNPs) to cover ~12% of common SNPs (MAF > 0.05) included in the 1000 Genomes Project reference panel for European ancestry[11].

Stage 1 trans-ethnic genome-wide meta-analysis. Associations of SNPs with RA were evaluated by logistic regression models assuming additive effects of the allele dosages including top 5 or 10 principal components as covariates (if available) using mach2dat v.1.0.16 (Extended Data Table 1b). Allele dosages of the SNPs in X chromosome were assigned as 0/1/2 for females and 0/2 for males and analysed separately. Meta-analysis was performed for the trans-ethnic study (both Europeans and Asians), European study, and Asian study separately. The SNPs available in ≥3 studies were evaluated in each GWAS meta-analysis, which yielded ~10 million autosomal and X-chromosomal SNPs. Information about the SNPs, including the coded alleles, was oriented to the forward strand of the NCBI build 37 reference sequence. Meta-analysis was conducted by an inverse-variance method assuming a fixed-effects model on the effect estimates (β) and the standard errors of the allele dosages using the Java source code implemented by the authors[25]. Double GC correction was carried out using the inflation factor (λ_{GC}) obtained from the results of each GWAS and the GWAS meta-analysis[25] after removing the SNPs located ± 1 Mb from known RA loci or in the MHC region (chromosome 6, 25–35 Mb). Although there is not yet uniform consensus on the application of double GC correction, we note that potential effects of double GC correction would not be substantial in our study because of the small values of the inflation factors in the GWAS meta-analysis ($\lambda_{GC} < 1.075$ and λ_{GC} adjusted for 1,000 cases and 1,000 controls ($\lambda_{GC\_1,000}$) < 1.005; Extended Data Table 1b).

As for the definition of known RA risk loci in this study, we included the loci that showed significant associations in one of the previous studies ($P < 5.0 \times 10^{-8}$) or that had been replicated in independent cohorts. We consider the locus including multiple independent signals of associations as a single locus, such as the MHC locus[12] and *TNFAIP3* (ref. 4). Although 6 of these 59 loci previously identified as known RA risk loci did not reach a suggestive level of association (defined as $P < 5.0 \times 10^{-6}$) in our stage 1 meta-analysis, previous studies have gone on to replicate most of these associations in additional samples (Supplementary Table 1)[2,3]. Thus, the number of confirmed RA risk loci is 101 (including the MHC region).

Stage 2 and stage 3 replication studies. *In silico* (stage 2) and *de novo* (stage 3) replication studies were conducted using independent European and Asian subjects (Extended Data Table 1). The 146 loci that satisfied $P < 5.0 \times 10^{-6}$ in the stage 1 trans-ethnic, European or Asian GWAS meta-analysis were selected for the stage 2 *in silico* replication study. The SNPs that demonstrated the most significant associations were selected from each of the loci. When the SNP was not available in replication data sets, a proxy SNP with the highest linkage disequilibrium ($r^2 > 0.80$) was alternatively assessed. GWAS quality control, genotype imputation and association analysis were assessed in the same manner as in the stage 1 GWAS. For the 60 loci that demonstrated suggestive associations in the combined results of the stage 1 GWAS meta-analysis and the stage 2 *in silico* replication study but were not included as a known RA risk locus, we calculated statistical power to newly achieve a genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ for Europeans and Asians separately, which were estimated based on the allele frequencies, ORs and *de novo* replication sample sizes of the populations. We then selected the top 20 SNPs with the highest statistical power for Europeans and Asians separately (in total 32 SNPs), and conducted the stage 3 *de novo* replication study. Genotyping methods, quality control and confirmation of subject independence in the stage 3 *de novo* replication study were described previously[2,3]. The combined study of the stage 1 GWAS meta-analysis and the stages 2 and 3 replication studies was conducted by an inverse-variance method assuming a fixed-effects model[25].

Trans-ethnic and functional annotations of RA risk SNPs. Trans-ethnic comparisons of RAF (in the reference panels), ORs and explained heritability were conducted using the results of the stage 1 GWAS meta-analysis of Europeans and Asians. Correlations of RAF and OR were evaluated using Spearman's correlation test. ORs were defined based on minor alleles in Europeans. Explained heritability was estimated by applying a liability-threshold model assuming disease prevalence of 0.5% (ref. 10) and using the RAF and OR of the population(s) according to the genetic risk model. For the population-specific genetic risk model, the RAF and OR of the same population was used. For the trans-ethnic genetic risk model, the RAF of the population but the OR of the other population was used.

Details of the overlap enrichment analysis of RA risk SNPs with H3K4me3 peaks have been described elsewhere[13]. Briefly, we evaluated whether the RA risk SNPs (outside of the MHC region) and SNPs in linkage disequilibrium ($r^2 > 0.80$) with them were enriched in overlap with H3K4me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) assay peaks of 34 cell types obtained from the National Institutes of Health Roadmap Epigenomics Mapping Consortium, by a permutation procedure with $\times 10^5$ iterations.

Fine mapping of causal risk alleles. For fine mapping of the causal risk alleles, we selected the 31 RA risk loci where the risk SNPs yielded $P < 1.0 \times 10^{-3}$ in the stage 1 GWAS meta-analysis of both Europeans and Asians with the same directional effects of alleles (outside of the MHC region). For fine mapping using linkage-disequilibrium structure differences between the populations, we calculated average numbers of the SNPs in linkage disequilbrium ($r^2 > 0.80$) in Europeans, Asians, and in both Europeans and Asians, separately.

For fine mapping using H3K4me3 peaks of T_{reg} primary cells, we first evaluated H3K4me3 peak overlap enrichment of the SNPs in linkage disequilbrium (in Europeans and Asians) compared with the neighbouring SNPs (±2 Mb). We fixed the SNP positions but physically slid H3K4me3 peak positions by 1 kb bins within ±2 Mb regions of the risk SNPs, and calculated overlap of the SNPs in linkage disequilibrium with H3K4me3 peaks for each sliding step, and evaluated the significance of overlap in the original peak positions by a one-sided exact test assuming enrichment of overlap. For the 10 loci that demonstrated significant overlap ($P < 0.05$), we calculated the average number of the SNPs that were in linkage disequilibrium in both Europeans and Asians and also included in H3K4me3 peaks.