SUBJECTS AND METHODS

The Nagahama cohort project is a population-based prospective cohort survey of a broad range of chronic illness, which was conducted in Nagahama City, Shiga prefecture, Japan.[7] This is a prospective study composed of a questionnaire survey, anthropometric measures, physiological measures, biochemical measures from blood samples, and genomic information. The baseline data of the study population were obtained from August 2008 to November 2010. All protocols and informed consent procedures were approved by the Kyoto University Graduate School and Faculty of Medicine Ethics Committee, the Ad hoc Review Board of the Nagahama Cohort Project, and the Nagahama Municipal Review Board of Personal Information Protection.

The International Prostate Symptom Score (I-PSS)[8,9] and the Overactive Bladder Symptom Score[10] were used to assess LUTS. The Japanese versions of these questionnaires have been validated.[9,10] Frequency of nocturnal voiding was assessed using the seventh question of the I-PSS and the second question of the OABSS. For analyses, we only used data from those participants whose answers for these two questions were compatible with each other. Self-reported information on medical history, major comorbidities, current medication use, lifestyle and psychosocial factors were also collected. Researchers asked a participant to answer appropriately to minimize the lack of data, if an answer sheet was incomplete with some blanks.

For the analysis of the prevalence of nocturia, we defined nocturic subjects as those who voided urine twice or more during the night. If some other data, including fasting blood sugar levels, were lacking, we excluded these data for analyses. We also stratified the subjects into those without nocturnal voiding, those with one nocturnal voiding, those with two nocturnal voidings, and those with three or more nocturnal voidings in the analysis for severity of nocturnal voiding. Independent variables investigated by questionnaire in this study were menopause, smoking, urgency (the third question of the OABSS), time in bed, sleeplessness, daytime sleepiness, history of hypertension, history of diabetes, history of dyslipidemia, history of cerebrovascular disease, and history of CVD.

Height, weight, and waist circumference were measured, and BMI was calculated. Physiological variables applied in this study were blood pressure (systolic, diastolic, and central), heart rate, AI, CAVI, and ABI. These variables were obtained twice 5-min apart, and were averaged. Central blood pressure and AI were measured using the HEM-9000AI® (Omron Corp., Kyoto, Japan), and CAVI and ABI were measured using the VaSera® VS-1000 (Fukuda Denshi Inc., Tokyo, Japan). The value of AI was modified under the setting of heart rate as 75 beats/min. Central blood pressure and AI represent the stiffness of the systemic arterial tree and the cardiac load, and they were reported to predict prognosis in subjects with end-stage renal disease or coronary heart disease.[11] CAVI is a marker of central arterial stiffness, and several studies showed this parameter can predict the prognosis of subjects with acute coronary syndrome, end-stage renal disease, or heart failure.[11,12] ABI is an index of stenosis of the central artery, and can predict CVDs.[13]

Blood samples were obtained in the fasting period. Data of participants whose blood samples were obtained within 8 hr after final ingestion were excluded from analysis because of inappropriate fasting blood sugar. Biochemical variables applied in this study other than fasting blood sugar were hemoglobin A1c, insulin, triglycerides, low density lipoprotein-cholesterol (LDLC), high-density lipoprotein-cholesterol (HDLC), B-type natriuretic peptide (BNP), hypersensitive C-reactive protein (CRP), creatinine, and cholinesterase. Since the distributions of BNP and CRP were skewed, natural log transformations of these two variables were used for analyses. The estimated glomerular filtration rate (eGFR) was calculated according to the Modification of Diet in Renal Diseases Study equations.[14,15]

Statistics

The total number of independent variables was 32, and of those, 22 were continuous variables and 9 were nominal variables. Only the parameter of daytime sleepiness was ordinal, ranked 1–5 (Table I). Dependent variables were a nocturnal voiding frequency of two or more times as defined for the analysis of the prevalence of nocturnal voiding, and the stratification of nocturnal voiding frequency in the analysis of its severity.

For analyzing correlates of the prevalence of nocturnal voiding, statistical analyses were performed in three steps using a logistic regression model. First, bivariate analysis was performed to confirm the basic relationship between each independent variable and nocturnal voiding, and if $P < 0.25$, the variable was applied to the next step. In step 2, multivariate logistic regression analysis was performed using the parameters included in the same category; for example, diabetic history, blood sugar, hemoglobin A1c, and serum insulin levels were used in the category of diabetes, to remove confounding factors. The types of categories with multiple variables were sleep, obesity, hypertension, diabetes, dyslipidemia, cardiac function, arteriosclerosis, and renal function. Using the variables with $P < 0.05$ in step 2, final multivariate logistic regression analysis was performed. If a category included only one variable, such as cerebral vascular disease, step 2 analysis was skipped, and $P < 0.25$ in step 1 would be required for final analyses. For analyzing correlates of the severity of nocturnal voiding, we used multiple regression analysis. All P-values were two-sided and $P < 0.05$ was considered significant. SPSS® version 13.0 was used for all calculations.

RESULTS

A total of 6,581 female residents participated in the Nagahama project. Of these, 482, 131, and 18 residents were excluded from this study because of inconsistency of answers for two questions regarding nocturnal voiding, a non-fasting blood sugar examination, and insufficient data, respectively. The data of the remaining 5,980 (91%) residents were used in this study.

Characteristics of Participants

The characteristics and the distributions of the participants are shown in Table I. The prevalence of none, one, two, and three or more nocturnal voiding episodes was 49.0%, 38.5%, 9.9%, and 2.6%, respectively.

Correlates for the Prevalence of Nocturnal Voiding

In step 1 of the analysis, only one variable, serum insulin level, was eliminated from step 2. In step 2, the four variables of history of hypertension, history of diabetes, HDLC, and serum creatinine level, were eliminated from step 3. Finally, age, urgency, time in bed, sleeplessness, waist circumference, log(BNP), BMI, menopause, and history of hypertension were significantly and independently associated with nocturnal

TABLE I. Background of Participants and Types of Variables

Category	Factor	Type of variable	Number (rate)	Mean (SD)
Subjects without nocturnal voiding			2,932 (49.0%)	
Subjects with nocturnal voiding			3,048 (51.0%)	
One nocturnal voiding			2,303 (38.5%)	
Two nocturnal voidings			591 (9.9%)	
Three or more nocturnal voidings			154 (2.6%)	
Age	Age	Continuous		52.6 (13.2)
Hormonal status	Menopause (yes)	Nominal	3,573 (59.7%)	
Smoking	Smoking (yes)	Nominal	385 (6.4%)	
Urgency	Urgency (yes)	Nominal	2,727 (54.6%)	
Sleep	Time in bed (hr)	Continuous		6.48 (1.03)
	Sleeplessness (yes)	Nominal	1,051 (17.6%)	
	Daytime sleepiness (yes)	Ordinal		
Obesity	Body mass index (kg/m/m)	Continuous		21.7 (3.2)
	Waist circumference (cm)	Continuous		78.6 (9.2)
Hypertension	History (yes)	Nominal	972 (16.3%)	
	Systolic pressure (mmHg)[a]	Continuous		120.4 (17.7)
	Diastolic pressure (mmHg)[a]	Continuous		73.8 (10.9)
	Central systolic pressure (mmHg)[a]	Continuous		125.8 (20.0)
Diabetes	History (yes)	Nominal	182 (3.0%)	
	Blood sugar (mg/dl)	Continuous		88.7 (12.4)
	Hemoglobin A1c (%)	Continuous		5.08 (0.45)
	Insulin (μU/ml)	Continuous		5.24 (9.03)
Dyslipidemia	History (yes)	Nominal	663 (11.1%)	
	Triglycerides (mg/dl)	Continuous		86.5 (52.5)
	Low-density lipoprotein-cholesterol (mg/dl)	Continuous		122.8 (31.2)
	High-density lipoprotein-cholesterol (mg/dl)	Continuous		68.8 (16.3)
Cerebrovascular disease	History (yes)	Nominal	39 (0.7%)	
Cardiac functions	History of cardiovascular disease (yes)	Nominal	270 (4.5%)	
	Heart rate (/min)[a]	Continuous		64.4 (10.0)
	B-type natriuretic peptide (pg/ml)	Continuous		17.5 (18.7)
Arterioscrelosis	C-reactive protein (ng/ml)	Continuous		762.5 (290.3)
	Augmentation index (%)[a]	Continuous		80.8 (11.5)
	Cardio-ankle vascular index[a]	Continuous		7.20 (1.04)
	Ankle-brachial index[a]	Continuous		1.08 (0.08)
Renal functions	Serum creatinine (mg/dl)	Continuous		0.62 (0.18)
	Estimated glomerular filtration rate (ml/min)	Continuous		77.6 (15.3)
Hepatic function	Serum cholinesterase (IU/L)	Continuous		328.8 (78.4)

[a]The average value of two measurements per individual.

voiding ≥2 times. BMI had an inverse association with the prevalence of nocturnal voiding after the final step. While bivariate analysis showed a positive correlation of BMI with the prevalence of nocturnal voiding [odds ratio (OR): 1.073, $P < 0.001$], waist circumference-adjustment changed the correlation from positive to negative. Central blood pressure, AI, CAVI, ABI, and CRP had no independent association (Table II).

Correlates for the Severity of Nocturnal Voiding

In this analysis, the adjusted r square value, the standard error, and the *F*-value were 0.271, 0.674, and 71.1, respectively. Age, urgency, sleeplessness, length of sleep, waist circumference, log(BNP), history of hypertension, and LDLC were significantly correlated with the severity of nocturnal voiding. LDLC had an inverse association with the severity of nocturnal voiding. While bivariate analysis showed a positive correlation of LDLC with the severity of nocturnal voiding ($\beta = 0.097$, $P < 0.001$), age-adjustment changed the correlation from positive to negative. Similar to the prevalence of nocturnal voiding, central blood pressure, AI, CAVI, ABI, and CRP had no independent association (Table III).

DISCUSSION

Many epidemiological studies have revealed a close association between nocturnal voiding and hypertension, diabetes, and obesity, which are components of METS.[3] Several other studies have reported associations between LUTS and vascular risk factors.[5,6] Therefore, it has been suggested that nocturnal voiding might have a relationship with cardiovascular conditions. Currently, several types of cardiovascular conditions are able to be assessed with various parameters. Cardiac load is represented by BNP and AI.[11,16] AI and central blood pressure are closely associated with each other, and it has been reported that central blood pressure is an independent risk factor of cardiovascular events.[11] Arterial stiffness, which results from fibrosis of the media of elastic arteries, can be estimated by pulse wave velocity, and CAVI is one of the indices to assess pulse wave velocity.[12] AI also reflects arterial stiffness. ABI is an index for arteriostenosis, which occurs at elastic and muscular arteries.[13] Hypersensitive-CRP is considered to mainly originate from endothelial and adipose cells, and to represent systemic inflammation inducing atherosclerosis.[17]

TABLE II. Correlates for the Prevalence of Nocturnal Voiding (≥2 Time)

		Step 1	Step 2	Step 3	
Category	Factor	*P*-value	*P*-value	*P*-value	OR (95% CI)
Age	Age	<0.001	—	<0.001	1.058 (1.044–1.079)
Hormonal status	Menopause (yes/no)	<0.001	—	0.043	1.503 (1.013–2.232)
Smoking	Smoking (yes/no)	<0.001	—	0.33	
Urgency	Urgency (yes/no)	<0.001	—	<0.001	1.528 (1.416–1.648)
Sleep	Time in bed (hr)	0.143	0.001	<0.001	1.194 (1.096–1.300)
	Sleeplessness (yes/no)	<0.001	0.002	<0.001	2.841 (2.347–3.436)
	Daytime sleepiness (yes/no)	0.009	<0.001	0.59	
Obesity	Body mass index (kg/m/m)	<0.001	<0.001	0.007	0.935 (0.890–0.982)
	Waist circumference (cm)	<0.001	<0.001	0.002	1.037 (1.019–1.055)
Hypertension	History (yes/no)	<0.001	<0.001	0.049	1.225 (1.001–1.500)
	Systolic pressure (mmHg)	<0.001	0.66	Con	
	Diastolic pressure (mmHg)	<0.001	<0.001	0.46	
	Central systolic pressure (mmHg)	<0.001	<0.001	0.85	
Diabetes	History (yes/no)	<0.001	0.36	Con	
	Blood sugar (mg/dl)	<0.001	<0.001	0.15	
	Hemoglobin A1c (%)	<0.001	<0.001	0.88	
	Insulin (μU/ml)	0.49	NA	NA	
Dyslipidemia	History (yes/no)	<0.001	<0.001	0.70	
	Triglycerides (mg/dl)	<0.001	<0.001	0.17	
	Low-density lipoprotein-cholesterol (mg/dl)	<0.001	<0.001	0.09	
	High-density lipoprotein-cholesterol (mg/dl)	0.004	0.32	Con	
Cerebrovascular disease	History (yes/no)	0.070		0.42	
Cardiac functions	History of cardiovascular disease (yes/no)	<0.001	<0.001	0.23	
	Heart rate (/min)	0.062	<0.001	0.17	
	B-type natriuretic peptide (pg/ml)[a]	<0.001	<0.001	<0.001	2.031 (1.529–2.698)
Arterioscrelosis	C-reactive protein (ng/ml)[a]	<0.001	<0.001	0.22	
	Augmentation index (%)	<0.001	0.004	0.38	
	Cardio-ankle vascular index	<0.001	<0.001	0.15	
	Ankle-brachial index	<0.001	<0.001	0.56	
Renal functions	Serum creatinine (mg/dl)	0.078	<0.001	0.88	
	Estimated glomerular filtration rate (ml/min)	<0.001	<0.001	0.06	
Hepatic function	Serum cholinesterase (IU/L)	<0.001	—	0.65	

NA, not analyzed; Con, not included in the final analysis because of severe confounding in the step 2 analysis.
[a]Natural log transformations of the BNP and CRP levels were used for analysis.

In this study, several variables, such as age, urgency, history of hypertension, obesity, and sleep-related factors, remained independent, as reported previously. As well as these known correlates, our study revealed that BNP was another strong independent correlate of nocturnal voiding. Plasma BNP levels have previously been reported to be closely related to the nocturnal polyuria index,[18] and this observation is compatible with that of another study, which found that nocturnal polyuria results from natriuresis during the night.[19] Since BNP represent cardiac load, as mentioned above, this suggests that clinical and preclinical heart failure might be an important cause of nocturnal voiding, and could also be a potential target of treatment for nocturnal voiding.

Unlike the significant correlation of BNP with nocturnal voiding in our study, CAVI or ABI did not have an independent association with the prevalence or severity of nocturnal voiding, although they exhibited a significant relationship until step 2 in the analysis of correlates for the prevalence of nocturnal voiding. This observation suggests that current status of arterial stiffness and arteriostenosis of elastic and muscular arteries have little, if any, impact on the frequency of nocturnal voiding.

Among the biochemical examinations investigated in this study, CRP and LDLC are associated with atherosclerosis.[17] While more than 30 epidemiological studies have reported that higher levels of CRP are a significant independent risk factor of ischemic stroke and cardiovascular events, some recent studies have reported that elevated serum CRP levels also have an association with several types of LUTS.[20–22] Although log(CRP) remained significantly associated with the prevalence of nocturnal voiding until step 2 of the analysis, it lost statistical significance in the final step of our study. Meanwhile, LDLC had an inverse association with the severity of nocturnal voiding. Since bivariate analysis demonstrated a positive correlation between LDLC and the severity of nocturnal voiding, this observation is probably due to the strong correlation between LDLC and aging.

Similarly, BMI was independently and inversely associated with prevalence of nocturnal voiding ≥2 times. There are many previous epidemiological studies that have shown a relationship between a higher BMI and nocturnal voiding. Most of these studies examined only BMI, but not waist circumference. BMI had a positive association with the prevalence of nocturnal voiding on bivariate analysis, and the outcome in the final step resulted from a strong correlation between BMI and waist circumference. Based on our observations of the prevalence and severity of nocturnal voiding, it is suggested that central obesity assessed by waist circumference plays a more important role in the development of nocturnal voiding than total obesity assessed by BMI of Japanese women.

TABLE III. Correlates for the Severity of Nocturnal Voiding

Factor	Beta-coefficient	*P*-value
(Constant)		<0.001
Age	0.256	<0.001
Menopause (yes/no)	—	0.12
Smoking (yes/no)	—	0.59
Urgency (yes/no)	0.195	<0.001
Time in bed (hr)	0.088	<0.001
Sleeplessness (yes/no)	0.181	<0.001
Daytime sleepiness (yes/no)	—	0.79
Body mass index (kg/m/m)	—	0.06
Waist circumference (cm)	0.086	<0.001
History of hypertension (yes/no)	0.051	<0.001
Systolic pressure (mmHg)	—	0.11
Diastolic pressure (mmHg)	—	0.06
Central systolic pressure (mmHg)	—	0.61
History of diabetes (yes/no)	—	0.57
Blood sugar (mg/dl)	—	0.20
Hemoglobin A1c (%)	—	0.75
Insulin (μU/ml)	—	0.09
History of dyslipidemia (yes/no)	—	0.29
Triglycerides (mg/dl)	—	0.56
Low-density lipoprotein-cholesterol (mg/dl)	−0.038	0.003
High-density lipoprotein-cholesterol (mg/dl)	—	0.67
History of cerebrovascular disease (yes/no)	—	0.32
History of cardiovascular disease (yes/no)	—	0.25
Heart rate (/min)	—	0.12
B-type natriuretic peptide (pg/ml)[a]	0.072	<0.001
C-reactive protein (ng/ml)[a]	—	0.13
Augmentation index (%)	—	0.10
Cardio-ankle vascular index	—	0.58
Ankle-brachial index	—	0.67
Serum creatinine (mg/dl)	—	0.88
Estimated glomerular filtration rate (ml/min)	—	0.40
Serum cholinesterase (IU/L)	—	0.50

[a]Natural log transformations of the BNP and CRP levels were used for analysis.

The strengths of this study include highly reliable data, with up to 91% of availability among all participants, as well as novel parameters being assessed for an epidemiological study, such as central blood pressure, AI, CAVI, ABI, BNP, and hypersensitive-CRP. However, there are several limitations to this study. First, several parameters did not show statistical significance because of a low prevalence. Although a history of stroke, for example, has been previously reported as a correlate of nocturnal voiding, only 39 individuals (0.7%) in this study suffered from the disease. Second, our study could not accurately link the correlates of nocturnal voiding to potential mechanisms of nocturnal voiding, which included decreased bladder capacity, increased urine production during the night, and sleep problems. Third, care should be taken to interpret the observations, because our study did not demonstrate risk factors of nocturnal voiding but correlates of nocturnal voiding instead. A longitudinal prospective study is warranted to reveal the risk factors and the consequences of nocturnal voiding. Fourth, several factors that may affect nocturnal voiding, such as history of pelvic organ surgery, were lacking in the analyses in this study. Finally, it is unknown whether similar results are observed in communities outside Japan or in other races, and whether these results are true in nocturic patients under the care of doctors. Despite these limitations, our observations are clinically important since they suggest that clinical and preclinical heart failure represented by high serum BNP levels might be a potential cause or marker of nocturnal voiding.

ACKNOWLEDGMENTS

We would like to express special thanks to the people of Nagahama city and Kyoto University who had contributed to launching this project, particularly the NPO Kenkou-Zukuri Zeroji Club. We also thank the Fukuda Denshi Co., Ltd. and Omron Healthcare Co., Ltd. for providing devices and personnel support.

APPENDIX: THE NAGAHAMA COHORT RESEARCH GROUP

Yuri Dowa, Department of Medical-Ethics/Medical Genetics, School of Public Health, Kyoto University

Tasuro Ishizaki, Department of Health Informatics, School of Public Health, Kyoto University

Hiroshi Kadotani, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University

Takahisa Kawaguchi, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University

Michiaki Mishima, Department of Respiratory Medicine, Graduate School of Medicine, Kyoto University

Koichi Miyaki, Department of Health Informatics, School of Public Health, Kyoto University

Seiho Mizusawa, Center for Genomic Medicine, Graduate School of Medicine, Kyoto University

Yoshimitsu Takahashi, Department of Health Informatics, School of Public Health, Kyoto University

REFERENCES

1. Homma Y, Yamaguchi O, Hayashi K. Epidemiologic survey of lower urinary tract symptoms in Japan. Urology 2006;68:560–4.
2. Resnick NM, Yalla SV. Geriatric incontinence and voiding dysfunction. In: Wein AJ, Kavoussi LR, Novick AC, et al., editors. Campbell-Walsh urology. 9th edition. Philadelphia: Saunders Elsevier; 2007. 2305–21.
3. Yoshimura K. Correlates for nocturia: A review of epidemiological studies. Int J Urol 2012; DOI: 10.111/j.1442-2042.2011.02956.x [Epub ahead of print].
4. Cameron A. The metabolic syndrome: Validity and utility of clinical definitions for cardiovascular disease and diabetes risk prediction. Maturitas 2009;65:117–21.
5. Ponholzer A, Temml C, Wehrberger C, et al. The association between vascular risk factors and lower urinary tract symptoms in both sexes. Eur Urol 2006;50:582–6.
6. Kim S, Jeong JY, Choi YJ, et al. Association between lower urinary tract symptoms and vascular risk factors in aging men: The Hallym Aging Study. Korean J Urol 2010;51:477–82.
7. Nakayama T. Implementation of a genomic epidemiology and public health activities in a community: Establishing a "rule" for the protection and utilization of individual information. Jap J Genet Counsel 2010;31: 117–21.
8. Barry MJ, Fowler FJ, O'Leary MP, et al. The American Urological Association symptom index for benign prostatic hyperplasia. J Urol 1992;148:1549–57.
9. Homma Y, Tsukamoto T, Yasuda K, et al. Linguistic validation of Japanese version of International Prostate Symptom Score and BPH Impact Index [in Japanese]. Nippon Hinyokika Gakkai Zasshi 2002;93:669–80.
10. Homma Y, Yoshida M, Seki N, et al. Symptom assessment tool for overactive bladder syndrome—Overactive Bladder Symptom Score. Urology 2006;68: 318–23.
11. Tomiyama H, Yamashina A. Non-invasive vascular function tests: Their pathophysiological background and clinical application. Circ J 2010;74: 24–33.
12. Yamabe T, Yoshizawa M, Saijo Y, et al. Brachio-ankle pulse wave velocity and cardio-ankle vascular index. Biomed Pharmacother 2004;58:S95–8.
13. Spacil J, Spacabilova J. The ankle-brachial blood pressure index as a risk indicator of generalized atherosclerosis. Semin Vasc Med 2002;2:441–5.
14. Levey A, Bosch JP, Lewis JB, et al. A more accurate method to estimate glomerular filtration rate from serum creatinine: A new prediction equation. Ann Intern Med 1999;130:461–70.
15. Japanese Society of Nephrology: Clinical guidebook for diagnosis and treatment of chronic kidney disease. Tokyo: Tokyo Igakusha Co. Ltd.; 2007. 33. [in Japanese].

16. McGrath MF, de Bold MLK, de Bold AJ. The endocrine function of the heart. Trends Endocrinol Metab 2005;16:469–77.
17. Calabro P, Golia E, Yeh ETH. CRP and the risk of atherosclerotic events. Semin Immnopathol 2009;31:79–94.
18. Sone A, Kondo N, Kobayashi T, et al. Association with relative nocturnal polyuria using BNP (brain natriuretic peptide) in elderly patients with nocturia [in Japanese]. Nippon Hinyokika Gakkai Zasshi 2007;98: 558–64.
19. Matthiesen TB, Rittig S, Norgaard JP, et al. Nocturnal polyuria and natriuresis in male patients with nocturia and lower urinary tract symptoms. J Urol 1996;156:1292–9.
20. Rohrmann S, De Marzo AM, Smit E, et al. Serum C-reactive protein concentration and lower urinary tract symptoms in older men in the Third National Health and Nutrition Examination Survey (NHANES III). Prostate 2006;62:27–33.
21. Kupelian V, McVary KT, Barry MJ, et al. Association of C-reactive protein and lower urinary tract symptoms in men and women: Results from Boston Area Community Health Survey. Urology 2009;73:950–7.
22. Lu Z, Gao Y, Tan A, et al. Increased high-sensitivity C-reactive protein predicts a high risk of lower urinary tract symptoms in Chinese male: Results from the Fangchenggang Area Male Health and Examination Survey. Prostate 2012;72:193–200.

Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population

Yukinori Okada[1–3,40], Chikashi Terao[4,5,40], Katsunori Ikari[6,40], Yuta Kochi[1,2,40], Koichiro Ohmura[5], Akari Suzuki[1], Takahisa Kawaguchi[4], Eli A Stahl[7,8], Fina A S Kurreeman[7–9], Nao Nishida[10], Hiroko Ohmiya[3], Keiko Myouzen[1], Meiko Takahashi[4], Tetsuji Sawada[11], Yuichi Nishioka[12], Masao Yukioka[13], Tsukasa Matsubara[14], Shigeyuki Wakitani[15], Ryota Teshima[16], Shigeto Tohma[17], Kiyoshi Takasugi[18], Kota Shimada[17], Akira Murasawa[19], Shigeru Honjo[20], Keitaro Matsuo[21], Hideo Tanaka[21], Kazuo Tajima[22], Taku Suzuki[6,23], Takuji Iwamoto[6,23], Yoshiya Kawamura[24], Hisashi Tanii[25], Yuji Okazaki[26], Tsukasa Sasaki[27], Peter K Gregersen[28], Leonid Padyukov[29], Jane Worthington[30], Katherine A Siminovitch[31], Mark Lathrop[32,33], Atsuo Taniguchi[6], Atsushi Takahashi[3], Katsushi Tokunaga[10], Michiaki Kubo[34], Yusuke Nakamura[35], Naoyuki Kamatani[36], Tsuneyo Mimori[5], Robert M Plenge[7,8], Hisashi Yamanaka[6], Shigeki Momohara[6,41], Ryo Yamada[37,41], Fumihiko Matsuda[4,38,39,41] & Kazuhiko Yamamoto[1,2,41]

Rheumatoid arthritis is a common autoimmune disease characterized by chronic inflammation. We report a meta-analysis of genome-wide association studies (GWAS) in a Japanese population including 4,074 individuals with rheumatoid arthritis (cases) and 16,891 controls, followed by a replication in 5,277 rheumatoid arthritis cases and 21,684 controls. Our study identified nine loci newly associated with rheumatoid arthritis at a threshold of $P < 5.0 \times 10^{-8}$, including *B3GNT2, ANXA3, CSF2, CD83, NFKBIE, ARID5B, PDE2A-ARAP1, PLD4* and *PTPN2*. *ANXA3* was also associated with susceptibility to systemic lupus erythematosus ($P = 0.0040$), and *B3GNT2* and *ARID5B* were associated with Graves' disease ($P = 3.5 \times 10^{-4}$ and 2.9×10^{-4}, respectively). We conducted a multi-ancestry comparative analysis with a previous meta-analysis in individuals of European descent (5,539 rheumatoid arthritis cases and 20,169 controls). This provided evidence of shared genetic risks of rheumatoid arthritis between the populations.

Rheumatoid arthritis is a complex autoimmune disease characterized by inflammation and the destruction of synovial joints and affects up to 1% of the population worldwide. To date, more than 35 rheumatoid arthritis susceptibility loci, including *HLA-DRB1*, *PTPN22*, *PADI4*, *STAT4*, *TNFAIP3* and *CCR6*, among others, have been identified by GWAS in multiple populations[1–12] and by several meta-analyses of the original GWAS[13–16]. In particular, each meta-analysis of these GWAS uncovered a number of loci that were not identified in the single GWAS, leading to recognition of the enormous power of the meta-analysis approach for detecting causal genes in disease. However, these previous meta-analyses have been performed solely in European populations[13–16] and not in Asian ones. As multi-ancestry studies on validated rheumatoid arthritis susceptibility loci showed the existence of both population-specific and shared genetic components of rheumatoid arthritis[10,17], additional studies in Asian populations might provide useful insight into the underlying genetic architecture of rheumatoid arthritis, which would otherwise be difficult to capture using the studies in a single population. Here, we report a meta-analysis of GWAS and a replication study for rheumatoid arthritis in a Japanese population that was conducted by the Genetics and Allied research in Rheumatic diseases NETworking (GARNET) consortium[10,12]. We subsequently performed a multi-ancestry comparative analysis that incorporated results from a previously conducted meta-analysis of individuals of European ancestry[15].

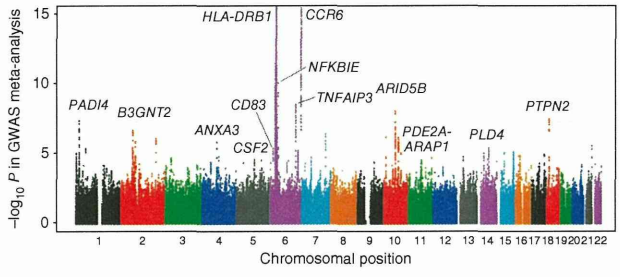

Figure 1 Manhattan plots of the GWAS meta-analysis for rheumatoid arthritis in the Japanese population. The genetic loci that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ (gray line) in the meta-analysis or in the combined study of the meta-analysis and the replication study are presented. The *y* axis shows the $-\log_{10} P$ values of the SNPs in the meta-analysis. The SNPs for which the *P* values were smaller than 1.0×10^{-15} are indicated at the upper limit of the plot.

A full list of author affiliations appears at the end of the paper.

Received 24 October 2011; accepted 1 March 2012; published online 25 March 2012; doi:10.1038/ng.2231

Table 1 Results of the GWAS meta-analysis and the replication studies for rheumatoid arthritis

					Associations in Japanese								Associations in Europeans[c]			
					GWAS meta-analysis				Replication study		Combined study		GWAS meta-analysis			
						Allele 1 freq.							Allele 1 Freq.			
rsID[a]	Chr.	Position (bp)	Cytoband	Gene(s)	Allele 1/2	RA	Control	OR (95% CI)[b]	*P*	OR (95% CI)[b]	*P*	OR (95% CI)[b]	*P*	RA	Control	OR (95% CI)[b]	*P*
SNPs with significant associations ($P < 5.0 \times 10^{-8}$ in the combined study)																	
rs11900673	2	62306165	2p15	*B3GNT2*	T/C	0.31	0.28	1.15 (1.08–1.21)	3.5×10^{-6}	1.09 (1.04–1.14)	6.0×10^{-4}	1.11 (1.07–1.15)	1.1×10^{-8}	0.13	0.13	1.05 (0.98–1.13)	0.17
rs2867461	4	79732239	4q21	*ANXA3*	A/G	0.46	0.44	1.13 (1.08–1.19)	4.7×10^{-6}	1.12 (1.08–1.17)	1.2×10^{-7}	1.13 (1.09–1.17)	1.2×10^{-12}	0.37	0.37	0.98 (0.92–1.04)	0.52
rs657075	5	131458017	5q31	*CSF2*	A/G	0.38	0.36	1.12 (1.06–1.18)	3.2×10^{-5}	1.11 (1.06–1.16)	3.8×10^{-6}	1.12 (1.08–1.15)	2.8×10^{-10}	0.10	0.10	1.04 (0.95–1.13)	0.37
rs12529514	6	14204637	6p23	*CD83*	C/T	0.16	0.14	1.19 (1.10–1.27)	6.8×10^{-6}	1.11 (1.05–1.18)	6.0×10^{-4}	1.14 (1.09–1.19)	2.0×10^{-8}	0.055	0.053	1.11 (0.99–1.24)	0.074
rs2233434	6	44340898	6p21.1	*NFKBIE*	G/A	0.24	0.21	1.23 (1.16–1.31)	9.2×10^{-11}	1.17 (1.11–1.23)	2.2×10^{-9}	1.19 (1.15–1.24)	5.8×10^{-19}	0.059	0.040	1.57 (1.11–2.21)	0.0099
rs10821944	10	63455095	10q21	*ARID5B*	G/T	0.39	0.36	1.17 (1.11–1.23)	1.0×10^{-8}	1.15 (1.10–1.20)	3.0×10^{-10}	1.16 (1.12–1.20)	5.5×10^{-18}	0.29	0.26	1.11 (1.05–1.17)	1.9×10^{-4}
rs3781913	11	72051144	11q13	*PDE2A-ARAP1*	T/G	0.71	0.69	1.11 (1.05–1.17)	3.2×10^{-4}	1.13 (1.08–1.18)	6.7×10^{-7}	1.12 (1.08–1.16)	5.8×10^{-10}	0.45	0.43	1.04 (0.99–1.09)	0.13
rs2841277	14	104462050	14q32	*PLD4*	T/C	0.72	0.69	1.11 (1.05–1.18)	2.8×10^{-4}	1.18 (1.13–1.24)	7.0×10^{-12}	1.15 (1.11–1.19)	1.9×10^{-14}	0.47	0.46	1.02 (0.96–1.09)	0.54
rs2847297	18	12787694	18p11	*PTPN2*	G/A	0.37	0.33	1.16 (1.11–1.23)	3.5×10^{-8}	1.06 (1.01–1.11)	0.013	1.10 (1.07–1.14)	2.2×10^{-8}	0.36	0.34	1.10 (1.05–1.15)	9.2×10^{-5}
SNPs with suggestive associations ($5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-6}$ in the combined study)																	
rs4937362	11	127997949	11q24	*ETS1-FLI1*	T/C	0.71	0.68	1.13 (1.07–1.19)	2.0×10^{-5}	1.07 (1.02–1.12)	0.0061	1.09 (1.06–1.13)	7.5×10^{-7}	0.46	0.44	1.06 (1.01–1.11)	0.015
rs3783637	14	54417868	14q22	*GCH1*	C/T	0.76	0.74	1.13 (1.07–1.20)	6.5×10^{-5}	1.07 (1.02–1.13)	0.0062	1.10 (1.06–1.14)	2.0×10^{-6}	0.88	0.88	0.99 (0.88–1.11)	0.87
rs1957895	14	60978085	14q23	*PRKCH*	G/T	0.40	0.39	1.12 (1.06–1.18)	4.1×10^{-5}	1.07 (1.02–1.12)	0.0022	1.09 (1.05–1.13)	3.6×10^{-7}	0.093	0.089	1.01 (0.95–1.07)	0.73
rs6496667	15	88694672	15q26	*ZNF774*	A/C	0.38	0.35	1.13 (1.07–1.19)	4.7×10^{-5}	1.07 (1.02–1.11)	0.0050	1.09 (1.05–1.13)	1.4×10^{-6}	0.21	0.20	1.07 (1.01–1.13)	0.031
rs7404928	16	23796341	16p12	*PRKCB1*	T/C	0.65	0.62	1.13 (1.07–1.19)	1.5×10^{-5}	1.05 (1.01–1.10)	0.026	1.08 (1.05–1.12)	4.0×10^{-6}	0.75	0.75	1.01 (0.94–1.09)	0.79
rs2280381	16	84576134	16q24	*IRF8*	T/C	0.86	0.84	1.16 (1.08–1.25)	1.0×10^{-4}	1.09 (1.03–1.15)	0.0049	1.12 (1.07–1.17)	2.4×10^{-6}	0.62	0.60	1.05 (0.99–1.11)	0.081
SNPs in previously reported rheumatoid arthritis susceptibility loci ($P < 5.0 \times 10^{-8}$ in the GWAS)																	
rs766449	1	17547439	1p36	*PADI4*	T/C	0.44	0.40	1.17 (1.11–1.24)	4.6×10^{-8}	–	–	–	–	0.38	0.37	1.09 (1.03–1.05)	0.0022
rs2157337	6	32609122	6p21.3	*HLA-DRB1*	C/T	0.59	0.44	1.99 (1.88–2.11)	2.6×10^{-118}	–	–	–	–	0.69	0.46	2.50 (2.39–2.62)	$< 1.0 \times 10^{-300}$
rs6932056	6	138284130	6q23	*TNFAIP3*	C/T	0.092	0.073	1.35 (1.23–1.49)	3.2×10^{-9}	–	–	–	–	0.044	0.034	1.41 (1.24–1.60)	1.3×10^{-7}
rs1571878	6	167460832	6q27	*CCR6*	C/T	0.54	0.48	1.31 (1.24–1.39)	3.2×10^{-19}	–	–	–	–	0.47	0.43	1.13 (1.08–1.19)	5.9×10^{-7}

Chr., chromosome; Freq., frequency; RA, rheumatoid arthritis; OR, odds ratio; CI, confidence interval.
[a]SNPs with $P < 5.0 \times 10^{-6}$ in the combined study of the GWAS meta-analysis and the replication study or SNPs with $P < 5.0 \times 10^{-8}$ in the GWAS meta-analysis are annotated according to forward strand and NCBI Build 36.3. Full results of the replication study are provided in **Supplementary Table 3**. [b]Odds ratio of allele 1. [c]Associations in the previous meta-analysis in European populations[15].

The meta-analysis included 4,074 rheumatoid arthritis cases (with 81.4% and 80.4% of the subjects being positive for antibody to cyclic citrullinated peptide (anti-CCP) and rheumatoid factor, respectively) and 16,891 controls from three GWAS of Japanese subjects (from the BioBank Japan Project[10,18], Kyoto University[12] and the Institute of Rheumatology Rheumatoid Arthritis (IORRA)[19]; **Supplementary Table 1**). After the application of stringent quality control criteria, including principal-component analysis (PCA; **Supplementary Fig. 1**) for each GWAS, the meta-analysis was conducted by evaluating ~2.0 million autosomal SNPs with minor allele frequencies (MAFs) ≥0.01, which were obtained through whole-genome imputation of genotypes on the basis of the HapMap Phase 2 East Asian panels (Japanese in Tokyo (JPT) and Han Chinese in Beijing (CHB)). The inflation factor of the test statistics in the meta-analysis λ_{GC} was as low as 1.036, suggesting no substantial effects of population structure (**Supplementary Table 2**). The quantile-quantile plot of *P* values showed a marked discrepancy in the values in its tail from those anticipated under the null hypothesis that there is no association—even after removal of the SNPs located in the human leukocyte antigen (HLA) region, the major rheumatoid arthritis susceptibility locus—thereby showing the presence of significant associations in the meta-analysis (**Supplementary Fig. 2**).

We identified seven loci in the current meta-analysis that satisfied the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$. These included previously known rheumatoid arthritis susceptibility loci, such as *PADI4* at 1p36, *HLA-DRB1* at 6p21.3, *TNFAIP3* at 6q23 and *CCR6* at 6q27 (refs. 1,3,6,10,15) (the smallest $P = 2.6 \times 10^{-118}$ was found at the *HLA-DRB1* locus; **Fig. 1** and **Table 1**). To our knowledge, the other three loci identified, *NFKBIE* at 6p21.1, *ARID5B* at 10q21 and *PTPN2* at 18p11, are newly associated ($P = 9.2 \times 10^{-11}$, 1.0×10^{-8} and 3.5×10^{-8}, respectively).

To validate the associations identified in the meta-analysis, we conducted a replication study of two independent Japanese rheumatoid arthritis case-control cohorts (cohort 1: 3,830 rheumatoid arthritis cases and 17,920 controls, cohort 2: 1,447 rheumatoid arthritis cases and 3,764 controls; **Supplementary Table 1**). To increase the number of subjects and enhance statistical power, genotype data obtained from other GWAS projects conducted for non-autoimmune diseases in Japanese using Illumina platforms were used for the replication control panels. For each of the 46 loci that exhibited $P < 5.0 \times 10^{-4}$ in

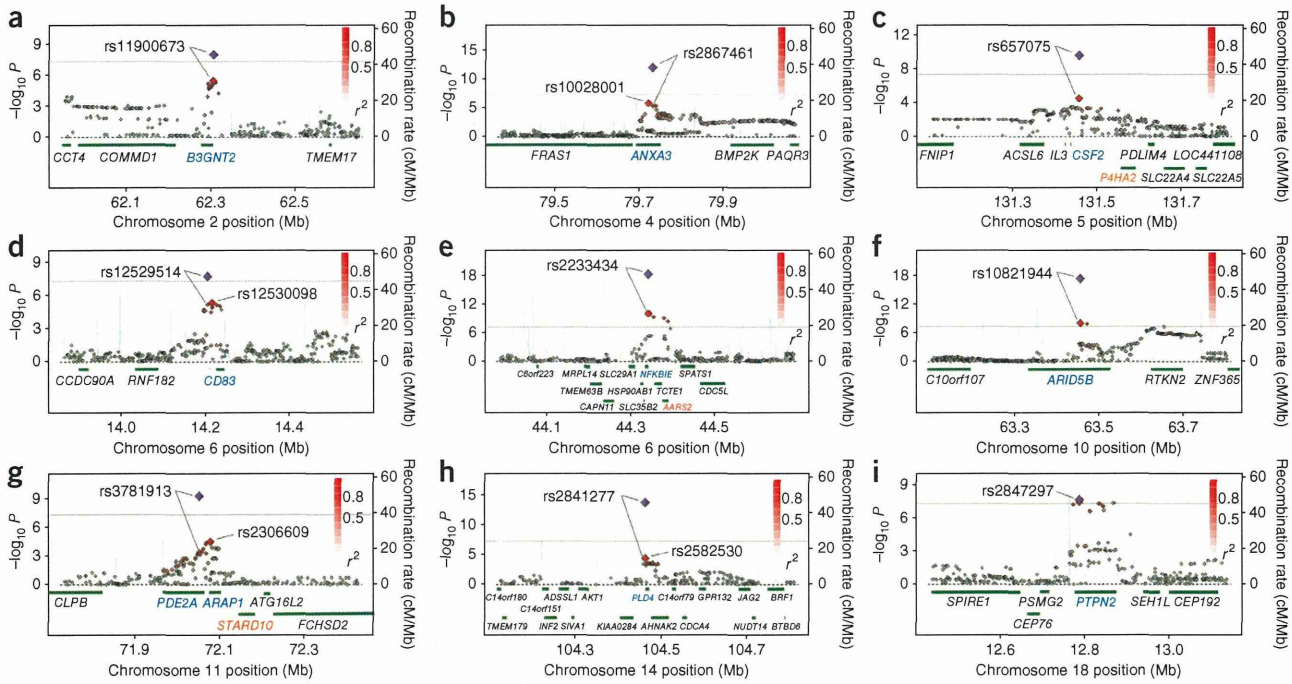

Figure 2 Regional plots of the loci newly associated with rheumatoid arthritis at the genome-wide significance threshold of $P < 5.0 \times 10^{-8}$ in the combined study of the meta-analysis and the replication study. (**a–i**) Regional plots are shown at *B3GNT2* (**a**), *ANXA3* (**b**), *CSF2* (**c**), *CD83* (**d**), *NFKBIE* (**e**), *ARID5B* (**f**), *PDE2A-ARAP1* (**g**), *PLD4* (**h**) and *PTPN2* (**i**). Diamonds represent the $-\log_{10} P$ values of the SNPs, and the red diamonds represent the $-\log_{10} P$ values of the SNPs in the meta-analysis. Red color for the smaller circles represents the r^2 value with the most significantly associated SNP (larger red circle). The purple circle represents the *P* value in the combined study. The blue line shows the recombination rates given by the HapMap Phase 2 east Asian populations (release 22). RefSeq genes at the loci are indicated below. Genes nearest to the marker SNPs at the loci are colored blue (**Supplementary Note**), and genes implicated in eQTL analysis are colored red (**Supplementary Table 4**). At 11q13, two genes (*PDE2A* and *ARAP1*) that are nearest to the SNP selected for the replication study and the most significant SNP in the meta-analysis are highlighted. The plots were drawn using SNP Annotation and Proxy Search (SNAP) version 2.2.

the meta-analysis and had not been reported as rheumatoid arthritis susceptibility loci[1–16], we selected a marker SNP for the replication study (Online Methods and **Supplementary Table 3**).

In the combined analyses of the meta-analysis and the replication study, including a total of 9,351 rheumatoid arthritis cases and 38,575 controls, we identified six newly associated loci, in addition to the *NFKBIE*, *ARID5B* and *PTPN2* loci, that satisfied the significance threshold of $P < 5.0 \times 10^{-8}$, including *B3GNT2* at 2p15, *ANXA3* at 4q21, *CSF2* at 5q31, *CD83* at 6p23, *PDE2A-ARAP1* at 11q13 and *PLD4* at 14q32 (**Figs. 1** and **2** and **Table 1**). Of these loci, *NFKBIE* had the smallest *P* value (5.8×10^{-19}). Although association with rheumatoid arthritis has been described for the *CSF2* and *PTPN2* loci[11,15,16,20,21], ours is the first report to our knowledge validating these associations with a threshold of $P < 5.0 \times 10^{-8}$. Suggestive associations were also observed in *ETS1-FLI1* at 11q24, *GCH1* at 14q22, *PRKCH* at 14q23, *ZNF774* at 15q26, *PRKCB1* at 16p12 and *IRF8* at 16q24 ($5.0 \times 10^{-8} \leq P < 5.0 \times 10^{-6}$). A summary of the genes in the newly associated loci and the results of *cis* expression quantitative trait locus (*cis* eQTL) analysis of the marker SNPs are provided (**Supplementary Table 4** and **Supplementary Note**).

Previous studies have reported associations of rheumatoid arthritis susceptibility loci with other autoimmune diseases[4,10,15,16]. Therefore, we assessed the association of these newly identified susceptibility loci with systemic lupus erythematosus (SLE) by examining the results of an SLE GWAS in the Japanese population (891 cases and 3,384 controls)[22] and in Graves' disease by genotyping 1,783 cases[10] (the controls from the SLE analysis were used for testing for Graves' disease). We observed significant associations of the *ANXA3* locus with SLE and of the *B3GNT2* and *ARID5B* loci with Graves' disease, which showed the same directional effects of the alleles as in rheumatoid arthritis ($P < 0.05/9 = 0.0056$, Bonferroni correction of the number of loci; **Supplementary Table 5**). It should be noted that relatively small sample sizes in the SLE and Graves' disease cohorts might yield limited statistical power, and further evaluations enrolling larger numbers of subjects would be desirable.

To highlight genetic backgrounds of rheumatoid arthritis that are common and divergent in different ancestry groups, we conducted a multi-ancestry comparative analysis of the present study in Japanese and a previous GWAS meta-analysis in Europeans that included 5,539 rheumatoid arthritis cases and 20,169 controls[15] (**Fig. 3a–c**). First, we compared associations in the reported[1–16] or newly identified rheumatoid arthritis susceptibility loci (**Fig. 3a** and **Supplementary Table 6**). Of the 46 rheumatoid arthritis risk variants evaluated, 6 were monomorphic in Japanese, and all were polymorphic in Europeans. We observed significant associations at 22 loci in Japanese and at 36 loci in Europeans (false discovery rate (FDR) < 0.05, $P < 0.0030$), with 14 loci being shared between the populations. Of the newly associated rheumatoid arthritis susceptibility loci identified in our Japanese meta-analysis, significant associations were also observed in the European meta-analysis at the *ARID5B* and *PTPN2* loci ($P = 1.9 \times 10^{-4}$ and 9.2×10^{-5}, respectively; **Table 1**). Significant positive correlation of odds ratios was observed between the studies ($r = 0.822$, $P = 8.1 \times 10^{-11}$; **Fig. 3b**), suggesting that a substantial proportion of genetic factors are shared between

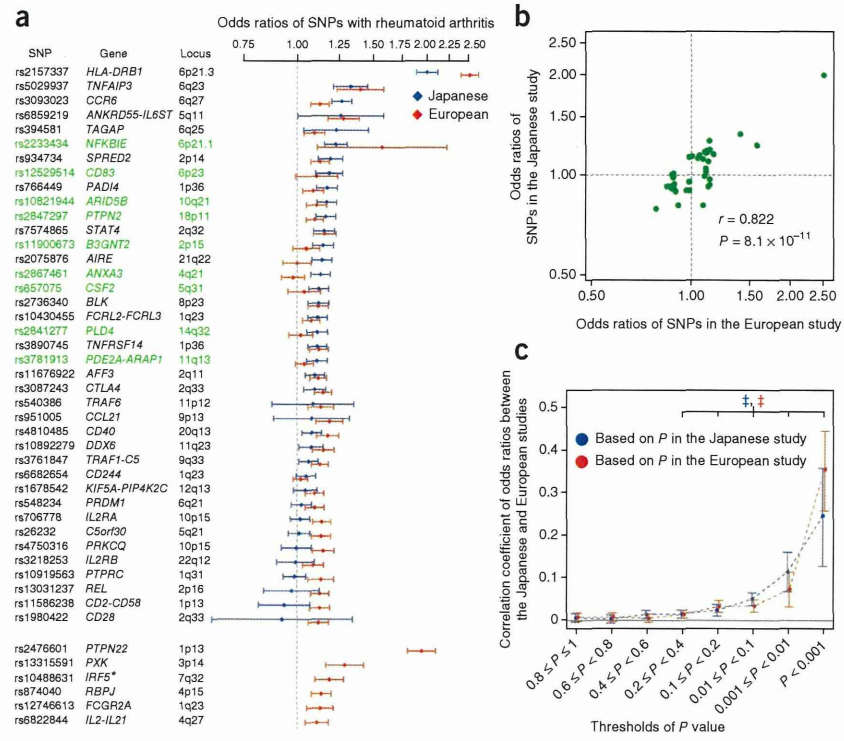

Figure 3 Overlap of the associations with rheumatoid arthritis between Japanese and European populations. (**a**) Forest plots of SNPs in the rheumatoid arthritis susceptibility loci (**Supplementary Table 6**). We selected the genetic loci that have been validated to be associated with rheumatoid arthritis susceptibility by showing associations in the reports of multiple cohorts or satisfying the genome-wide significant threshold ($P < 5.0 \times 10^{-8}$) in previous studies, including in the meta-analysis and replication phases[1–16]. For each of the loci, the most significant SNP among those reported in the previous or present study were selected[1–16]. SNPs in the newly identified rheumatoid arthritis susceptibility loci are colored green. Odds ratios and 95% confidence interval (CI) values are based on rheumatoid arthritis risk alleles, and the SNPs are ordered according to the odds ratios in the Japanese study. Several SNPs were monomorphic in the Japanese population. The odds ratios of these SNPs in the European study are presented below. The asterisk indicates that an association of another variant at the *IRF5* locus was reported in the Japanese population[24]. (**b**) Correlation of the odds ratios of the SNPs in the validated rheumatoid arthritis susceptibility loci between the two populations. SNPs that were polymorphic in both populations were used; odds ratios were based on the minor allele in the Japanese population. (**c**) Correlation of the odds ratios of the genome-wide SNPs, excluding the rheumatoid arthritis susceptibility loci. Correlations were evaluated for sets of SNPs stratified by the thresholds based on the meta-analysis *P* values in each population after pruning of the SNPs by LD ($r^2 < 0.3$). Correlation coefficient and 95% CI are indicated on the *y* axis. Significant correlation of the odds ratios was observed (‡, $P < 0.005$), even for the SNPs that showed moderate associations with rheumatoid arthritis (meta-analysis $P < 0.4$ in each population).

the two ancestry groups[17]. When the rheumatoid arthritis cases of the Japanese GWAS meta-analysis were stratified into anti-CCP–positive or rheumatoid factor–positive cases ($n = 3{,}209$) and controls ($n = 16{,}891$), similar results were observed (data not shown). Nevertheless, most of the SNPs assessed here are not necessarily causal variants, and further fine mapping of the loci is warranted to precisely evaluate the shared genetic predisposition between the populations.

Next, we compared regional associations within each of the loci and identified unique patterns in the *ARID5B* locus at 10q21 (**Supplementary Fig. 3**). In Japanese, three peaks of association were observed ($P = 1.0 \times 10^{-8}$ at rs10821944, $P = 5.7 \times 10^{-8}$ at rs10740069 and $P = 8.5 \times 10^{-6}$ at rs224311). These three variants were in weak linkage disequilibrium (LD) in Japanese ($r^2 < 0.10$), indicating independent associations with each of the other SNPs that satisfied a region-wide significance threshold of $P < 3.5 \times 10^{-5}$ (conditional $P = 4.3 \times 10^{-6}$, 1.7×10^{-5} and 1.8×10^{-5}, respectively) (**Supplementary Fig. 3**). In contrast, there was only one peak of association in Europeans ($P = 1.2 \times 10^{-6}$ at rs12764378; $r^2 = 0.59$ with rs10821944 in Europeans), and no additional association was observed in conditional analysis with rs12764378 (the smallest conditional $P = 2.2 \times 10^{-4}$), suggesting that the number of independent associations may be different at this locus in the two populations.

Finally, we conducted polygenic assessment for common variants showing modest associations to rheumatoid arthritis (those not meeting the genome-wide association threshold). This approach has been recognized to be a means to explain a substantial proportion of genetic risk[23]. For the SNPs that were shared between the two meta-analyses but not included in the validated rheumatoid arthritis susceptibility loci, we adopted LD pruning of the SNPs ($r^2 < 0.3$). We then evaluated the correlation of odds ratios of the SNPs between the two meta-analyses and observed a significant positive correlation ($r = 0.023$, $P < 1.0 \times 10^{-300}$). When the SNPs were stratified according to the *P* values in each meta-analysis, significant positive correlations of odds ratios were observed for the SNPs, even for those showing modest association ($P < 0.4$ in the meta-analysis of Japanese or Europeans; $r = 0.014$–0.36 for each *P* value range, $P < 0.005$ for each correlation test) (**Fig. 3c**). Correlations (r) of odds ratios observed herein suggest substantial overlap of the genetic risk of rheumatoid arthritis between the two populations, not only in the validated rheumatoid arthritis susceptibility loci but also at the loci showing nonsignificant associations. This suggests the usefulness of a meta-analysis approach involving multiple ancestry groups in identifying additional susceptibility loci.

In summary, we identified multiple new loci associated with rheumatoid arthritis through a large-scale meta-analysis of GWAS in Japanese. Multi-ancestry comparative analysis provided evidence of significant overlap in the genetic risks of rheumatoid arthritis between Japanese and Europeans. Thus, findings from the present study should contribute to the further understanding of the etiology of rheumatoid arthritis.

URLs. GARNET consortium, http://www.twmu.ac.jp/IOR/garnet/home.html; The BioBank Japan Project (in Japanese), http://biobankjp.org/; International HapMap Project, http://www.hapmap.org/; PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/; EIGENSTRAT, http://genepath.med.harvard.edu/~reich/Software.htm; MACH and mach2dat, http://www.sph.umich.edu/csg/abecasis/MACH/index.

html; R statistical software, http://cran.r-project.org/; SNAP, http://www.broadinstitute.org/mpg/snap/index.php; NCBI GEO database, http://www.ncbi.nlm.nih.gov/geo/.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

The authors acknowledge the essential role of the GARNET consortium in developing the study. In this study, the following GARNET members are included: CGM of RIKEN, University of Tokyo, the BioBank Japan Project, Kyoto University and IORRA. We would like to thank all the doctors and staff who participated in sample collection for the RIKEN cohort and the BioBank Japan Project. We thank K. Kobayashi and M. Kitazato for their technical assistance. We thank T. Raj for calculation of composite of multiple signals (CMS). We thank M. Kokubo for DNA extraction, GWAS genotyping and secretarial assistance. We would also like to thank H. Yoshifuji, N. Yukawa, D. Kawabata, T. Nojima, T. Usui and T. Fujii for collecting DNA samples. We thank Y. Katagiri for her technical efforts. We also appreciate the contribution of E. Inoue and other members of the Institute of Rheumatology, Tokyo Women's Medical University, for their efforts on the IORRA cohort. This study was supported in part by grants-in-aid from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, the Ministry of Health, Labour and Welfare (MHLW) in Japan, the Japan Society for the Promotion of Science (JSPS), Core Research for Evolutional Science and Technology (CREST), Solution-Oriented Research for Science and Technology (SORST), INSERM and the Okawa Foundation for Information and Telecommunications.

AUTHOR CONTRIBUTIONS

Y. Okada, C.T., K.I., Y. Kochi and K.O. designed the study and drafted the manuscript. Y. Okada, C.T., K.I., T.K., H.O., N.N., M.T., M.L., K. Tokunaga and M.K. managed genotyping and manipulation of GWAS data. Y. Okada, Y. Kochi, C.T. and K.I. managed genotyping of replication cohorts. Y. Okada, T.K., H.O., E.A.S., A. Takahashi and R.Y. performed statistical analysis. Y. Kochi, A.S., K. Myouzen, T. Sawada, Y. Nishoka, M.Y., T. Matsubara, S.W., R.T. and S.T. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from the BioBank Japan Project and CGM, RIKEN. C.T., K.O., T.K., M.T., K. Takasugi, K.S., A.M., S.H., K. Matsuo, H. Tanaka, K. Tajima and M.L. collected samples and managed phenotype data for the rheumatoid arthritis cohorts from Kyoto University. K.I., T. Suzuki, T.I., Y. Kawamura, H. Tanii, Y. Okazaki and T. Sakaki collected samples and managed phenotype data for the rheumatoid arthritis cohorts from IORRA. Y. Kochi managed the data for the SLE and Graves' disease cohorts. A.S., C.T. and K.I. analyzed the sera of subjects with rheumatoid arthritis. E.A.S., F.A.S.K., P.K.G., J.W., K.A.S., L.P. and R.M.P. managed the data for the rheumatoid arthritis cohorts in European populations. A. Taniguchi, A. Takahashi, K. Tokunaga, M.K., Y. Nakamura, N.K., T. Minori, R.M.P., H.Y., S.M., R.Y., F.M. and K.Y. supervised the overall study.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Suzuki, A. *et al.* Functional haplotypes of *PADI4*, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat. Genet.* **34**, 395–402 (2003).
2. Kochi, Y. *et al.* A functional variant in *FCRL3*, encoding Fc receptor–like 3, is associated with rheumatoid arthritis and several autoimmunities. *Nat. Genet.* **37**, 478–485 (2005).
3. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
4. Remmers, E.F. *et al. STAT4* and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.* **357**, 977–986 (2007).
5. Plenge, R.M. *et al. TRAF1–C5* as a risk locus for rheumatoid arthritis—a genomewide study. *N. Engl. J. Med.* **357**, 1199–1209 (2007).
6. Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
7. Barton, A. *et al.* Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat. Genet.* **40**, 1156–1159 (2008).
8. Suzuki, A. *et al.* Functional SNPs in *CD244* increase the risk of rheumatoid arthritis in a Japanese population. *Nat. Genet.* **40**, 1224–1229 (2008).
9. Gregersen, P.K. *et al. REL*, encoding a member of the NF-κB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. *Nat. Genet.* **41**, 820–823 (2009).
10. Kochi, Y. *et al.* A regulatory variant in *CCR6* is associated with rheumatoid arthritis susceptibility. *Nat. Genet.* **42**, 515–519 (2010).
11. Freudenberg, J. *et al.* Genome-wide association study of rheumatoid arthritis in Koreans: population-specific loci as well as overlap with European susceptibility loci. *Arthritis Rheum.* **63**, 884–893 (2011).
12. Terao, C. *et al.* The human *AIRE* gene at chromosome 21q22 is a genetic determinant for the predisposition to rheumatoid arthritis in Japanese population. *Hum. Mol. Genet.* **20**, 2680–2685 (2011).
13. Raychaudhuri, S. *et al.* Common variants at *CD40* and other loci confer risk of rheumatoid arthritis. *Nat. Genet.* **40**, 1216–1223 (2008).
14. Raychaudhuri, S. *et al.* Genetic variants at *CD28*, *PRDM1* and *CD2/CD58* are associated with rheumatoid arthritis risk. *Nat. Genet.* **41**, 1313–1318 (2009).
15. Stahl, E.A. *et al.* Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.* **42**, 508–514 (2010).
16. Zhernakova, A. *et al.* Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.* **7**, e1002004 (2011).
17. Kurreeman, F. *et al.* Genetic basis of autoantibody positive and negative rheumatoid arthritis risk in a multi-ethnic cohort derived from electronic health records. *Am. J. Hum. Genet.* **88**, 57–69 (2011).
18. Nakamura, Y. The BioBank Japan Project. *Clin. Adv. Hematol. Oncol.* **5**, 696–697 (2007).
19. Yamanaka, H. *et al.* Influence of methotrexate dose on its efficacy and safety in rheumatoid arthritis patients: evidence based on the variety of prescribing approaches among practicing Japanese rheumatologists in a single institute-based large observational cohort (IORRA). *Mod. Rheumatol.* **17**, 98–105 (2007).
20. Yamada, R. *et al.* Association between a single-nucleotide polymorphism in the promoter of the human interleukin-3 gene and rheumatoid arthritis in Japanese patients, and maximum-likelihood estimation of combinatorial effect that two genetic loci have on susceptibility to the disease. *Am. J. Hum. Genet.* **68**, 674–685 (2001).
21. Tokuhiro, S. *et al.* An intronic SNP in a RUNX1 binding site of *SLC22A4*, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat. Genet.* **35**, 341–348 (2003).
22. Okada, Y. *et al.* A genome-wide association study identified *AFF1* as a susceptibility locus for systemic lupus eyrthematosus in Japanese. *PLoS Genet.* **8**, e1002455 (2012).
23. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics* **187**, 367–383 (2011).
24. Shimane, K. *et al.* A single nucleotide polymorphism in the *IRF5* promoter region is associated with susceptibility to rheumatoid arthritis in the Japanese patients. *Ann. Rheum. Dis.* **68**, 377–383 (2009).

[1]Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Yokohama, Japan. [2]Department of Allergy and Rheumatology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [3]Laboratory for Statistical Analysis, CGM, RIKEN, Yokohama, Japan. [4]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [5]Department of Rheumatology and Clinical Immunology, Graduate School of Medicine, Kyoto University, Kyoto, Japan. [6]Institute of Rheumatology, Tokyo Women's Medical University, Tokyo, Japan. [7]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [8]Broad Institute, Cambridge, Massachusetts, USA. [9]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands. [10]Department of Human Genetics, Graduate School of Medicine, University of Tokyo, Tokyo, Japan. [11]Department of Rheumatology, Tokyo Medical University Hospital, Tokyo, Japan. [12]Yamanashi Prefectural Central Hospital, Yamanashi, Japan. [13]Department of Orthopaedic Surgery, Yukioka Hospital, Osaka, Japan. [14]Matsubara Mayflower Hospital, Hyogo, Japan. [15]Osaka Minami National Hospital, Osaka, Japan. [16]Department of Orthopedic Surgery, Tottori University, Tottori, Japan. [17]Department of Rheumatology, National Hospital Organization, Sagamihara Hospital, Kanagawa, Japan. [18]Center for Rheumatic Diseases, Dohgo Spa Hospital, Ehime, Japan. [19]Department of Rheumatology, Niigata Rheumatic Center, Niigata, Japan. [20]Saiseikai Takaoka Hospital, Toyama, Japan. [21]Division of Epidemiology and Prevention, Aichi Cancer Center Research Institute, Aichi, Japan. [22]Aichi Cancer Center Hospital and Research Institute, Aichi, Japan. [23]Department of Orthopaedic Surgery, Keio University, Tokyo, Japan. [24]Yokohama Clinic, Warakukai Medical Corporation, Yokohama, Japan. [25]Department of Psychiatry, Mie University School of Medicine, Mie, Japan. [26]Metropolitan Matsuzawa Hospital, Tokyo, Japan. [27]Graduate School of Education, University of Tokyo, Tokyo, Japan. [28]The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA. [29]Rheumatology Unit,

Department of Medicine in Solna, Karolinska Institutet and Karolinska University Hospital, Stockholm, Sweden. [30]Arthritis Research Campaign–Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [31]Division of Medicine, University of Toronto, Mount Sinai Hospital and University Health Network, Toronto, Ontario, Canada. [32]Commisariat a l'Energie Atomique (CEA), Institut Genomique, Centre National de Genotypage, Evry, France. [33]Fondation Jean Dausset, Centre d'Etude du Polymorphisme Humain, Paris, France. [34]Laboratory for Genotyping Development, CGM, RIKEN, Yokohama, Japan. [35]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan. [36]Laboratory for International Alliance, CGM, RIKEN, Yokohama, Japan. [37]Unit of Statistical Genetics, Center for Genomic Medicine Graduate School of Medicine Kyoto University, Kyoto, Japan. [38]Core Research for Evolutional Science and Technology (CREST) Program, Japan Science and Technology Agency, Kawaguchi, Japan. [39]Institut National de la Santé et de la Recherche Médicale (INSERM), Unité U852, Kyoto University Graduate School of Medicine, Kyoto, Japan. [40]These authors contributed equally to this work. [41]These authors jointly directed this work. Correspondence should be addressed to Y.K. (ykochi@src.riken.jp) or K.O. (ohmurako@kuhp.kyoto-u.ac.jp).

ONLINE METHODS

Subjects. The Japanese participants in the meta-analysis (4,074 rheumatoid arthritis cases and 16,891 controls) and the replication study (5,277 rheumatoid arthritis cases and 21,684 controls) were obtained through the collaborations of the GARNET consortium (**Supplementary Table 1**)[10,12]. The meta-analysis was conducted on three independent GWAS (from the BioBank Japan Project[18] with 2,414 rheumatoid arthritis cases and 14,245 controls[10], Kyoto University with 1,237 rheumatoid arthritis cases and 2,087 controls[12] and IORRA[19] with 423 rheumatoid arthritis cases and 559 controls). The replication study consisted of two independent cohorts (cohort 1 included 3,830 rheumatoid arthritis cases and 17,920 controls, and cohort 2 included 1,447 rheumatoid arthritis cases and 3,764 controls). We employed a case-control cohort of SLE (891 cases and 3,384 controls)[22] and 1,783 cases with Graves' disease[10]. Details of 5,539 rheumatoid arthritis cases and 20,169 controls included in the meta-analysis in European populations were described elsewhere[15]. All participants provided written informed consent for participation in the study, as approved by the ethical committees of the institutional review boards. Detailed descriptions of the participating subjects are provided (**Supplementary Note**).

Genotyping and quality control in the GWAS. Genotyping platforms and quality control criteria for the GWAS, including cutoff values for sample call rates, SNP call rates, MAF and Hardy-Weinberg *P* values, are given (**Supplementary Table 2**). For the subjects enrolled in each of three GWAS, we excluded closely related subjects with first- or second-degree kinship, which was estimated using PLINK version 1.06 (see URLs). We also excluded the subjects determined to be ancestry outliers from East Asian populations using PCA performed by EIGENSTRAT version 2.0 (see URLs) along with HapMap Phase 2 panels (release 24; **Supplementary Fig. 1**). Genotype imputation was performed on the basis of the HapMap Phase 2 East Asian populations, using MACH version 1.0.16 (see URLs) in a two-step procedure as described elsewhere[25]. We excluded imputed SNPs with MAF < 0.01 or *Rsq* < 0.5 from each of the GWAS. Associations of the SNPs with rheumatoid arthritis were assessed by logistic regression models assuming additive effects of the allele dosages of the SNPs using mach2dat software (see URLs).

Meta-analysis. We included 1,948,139 autosomal SNPs that satisfied quality control criteria in all three GWAS (**Supplementary Table 2**). SNP information was based on a forward strand of the NCBI build 36.3 reference sequence. The meta-analysis was performed using an inverse variance method assuming a fixed-effects model from the study-specific effect sizes (logarithm of odds ratio) and the standard errors of the coded alleles of the SNPs determined with the Java source code implemented by the authors[25]. Genomic control corrections[26] were carried out on test statistics of the GWAS using the study-specific inflation factor (λ_{GC}) and was applied or reapplied to the results of our current meta-analysis (**Supplementary Fig. 2**).

Replication study. We selected a SNP for the replication study from each of the loci that exhibited $P < 5.0 \times 10^{-4}$ in the meta-analysis that had not previously been reported as rheumatoid arthritis susceptibility loci[1–16] (**Supplementary Table 3**). For control subjects, we used genotype data obtained from additional GWAS for non-autoimmune diseases or healthy controls, genotyped using Illumina HumanHap550 BeadChips or HumanHap610-Quad BeadChips, and the cases for rheumatoid arthritis and Graves' disease were genotyped with the TaqMan genotyping system (Applied Biosystems; **Supplementary Table 1**). Selection of the SNP was conducted according to the following criteria: if the SNP with the most significant association in the locus was genotyped in the replication control panel, then that SNP was selected; otherwise, a tag SNP in the replication control panel with the strongest LD was selected (mean $r^2 = 0.89$). For the three SNPs that yielded low call rates (<90%), we alternatively selected proxy SNPs with the second strongest LD. As a result, average genotyping call rates of the SNPs were 99.9% and 99.0% for the controls and cases, respectively. We then evaluated concordance rates between the assayed genotypes by applying these two different methods to samples from 376 subjects who were randomly selected. This procedure yielded high concordance rates of ≥99.9%. Associations of the SNPs were evaluated using logistic regression assuming an additive-effects model of genotypes in R statistical software version 2.11.0 (see URLs). The combined study of the meta-analysis and replication study was performed using an inverse variance method assuming a fixed-effects model[25].

***Cis* eQTL analysis.** For each marker SNP of the newly identified rheumatoid arthritis susceptibility locus, correlations between SNP genotypes and expression levels of genes located 300 kb upstream or downstream of the SNP measured in B-lymphoblastoid cell lines (GSE6536) were evaluated using data from the HapMap Phase 2 east Asian populations[27].

Multi-ancestry analysis of the meta-analyses in Japanese and Europeans. We evaluated the associations of the variants in the validated rheumatoid arthritis susceptibility loci by comparing the results from the current meta-analysis in Japanese with those from a previous meta-analysis in Europeans[15]. We assessed two variants in the *IRF5* locus, where different causal variants were identified in the two populations[24]. For the conditional analysis of the regional associations in the *ARID5B* locus (**Supplementary Fig. 3**), we repeated the meta-analysis at that locus by incorporating genotypes of the referenced SNP(s) as additional covariate(s). For comparison of the odds ratios of the SNPs, we first selected SNPs that were shared between the meta-analyses in Japanese and Europeans. Next, we removed the SNPs located more than 1 Mb away from each of the marker SNPs in the validated rheumatoid arthritis susceptibility loci, except for in the HLA region, where we removed the SNPs located between 24,000,000 bp to 36,000,000 bp on chromosome 6 because of the existence of long-range haplotypes with rheumatoid arthritis susceptibility in this region[28]. LD pruning of the SNPs was conducted for the SNP pairs that were in LD ($r^2 \geq 0.3$) in both HapMap Phase 2 East Asian and Utah residents of Northern and Western European ancestry (CEU) populations (release 24). Correlations of the odds ratios were evaluated using R statistical software version 2.11.0.

25. Okada, Y. *et al.* Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. *PLoS Genet.* **7**, e1002067 (2011).
26. de Bakker, P.I. *et al.* Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum. Mol. Genet.* **17**, R122–R128 (2008).
27. Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
28. Okada, Y. *et al.* Contribution of a haplotype in the HLA region to anti–cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA-DRB1. *Arthritis Rheum.* **60**, 3582–3590 (2009).

Vascular Endothelial Growth Factor Gene Polymorphisms and Choroidal Neovascularization in Highly Myopic Eyes

Yumiko Akagi-Kurashige,[1,2,3] *Kyoko Kumagai,*[1,3] *Kenji Yamashiro,*[1] *Hideo Nakanishi,*[1,2] *Isao Nakata,*[1,2] *Masahiro Miyake,*[1,2] *Akitaka Tsujikawa,*[1] *Muka Moriyama,*[4] *Kyoko Ohno-Matsui,*[4] *Manabu Mochizuki,*[4] *Ryo Yamada,*[2] *Fumihiko Matsuda,*[2] *and Nagahisa Yoshimura*[1]

Purpose. To investigate a potential association between VEGF gene polymorphisms and the occurrence and/or the size of choroidal neovascularization (CNV) in highly myopic eyes.

Methods. In the case-control study for CNV occurrence, 327 highly myopic Japanese patients were enrolled. One hundred and eighty-four patients had CNV in at least one eye, and 143 did not have CNV in either eye. Of the 184 patients with CNV, 83 patients were used to evaluate an association with CNV size, and an additional 76 patients with CNV were used to confirm the association. We genotyped four tag single nucleotide polymorphisms (SNPs) and four functional SNPs previously reported to be correlated with VEGF gene expression to evaluate the associations of these eight SNPs with CNV occurrence and size. To confirm the association between CNV size and VEGF gene polymorphism, the associated SNP was genotyped in 76 additional patients with myopic CNV.

Results. There was no significant association between the occurrence of myopic CNV and the SNPs in the VEGF gene ($P > 0.16$). Of the eight SNPs evaluated, however, rs2010963 showed significant association with CNV area ($P = 0.0047$). This association was successfully replicated in the additional 76 eyes with myopic CNV, and pooled analysis revealed significant association of rs2010963 with CNV size ($P = 0.00078$).

Conclusions. VEGF gene polymorphisms were not associated with CNV occurrence in highly myopic eyes but were significantly associated with the size of CNV, suggesting roles in the growth rather than the emergence of CNV. (*Invest Ophthalmol Vis Sci.* 2012;53:2349-2353) DOI:10.1167/iovs.11-9405

From the [1]Department of Ophthalmology, and the [2]Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan; and the [4]Department of Ophthalmology and Visual Science, Tokyo Medical and Dental University, Tokyo, Japan.

[3]These authors contributed equally to this work and should therefore be regarded as equivalent authors.

Supported in part by Grants-in-Aid for Scientific Research (Nos. 21249084 and 200791294) from the Japan Society for the Promotion of Science, Tokyo, Japan, and the Japan National Society for the Prevention of Blindness, Tokyo, Japan.

Submitted for publication December 27, 2011; revised February 14, 2012; accepted February 17, 2012.

Disclosure: **Y. Akagi-Kurashige**, None; **K. Kumagai**, None; **K. Yamashiro**, None; **H. Nakanishi**, None; **I. Nakata**, None; **M. Miyake**, None; **A. Tsujikawa**, None; **M. Moriyama**, None; **K. Ohno-Matsui**, None; **M. Mochizuki**, None; **R. Yamada**, None; **F. Matsuda**, None; **N. Yoshimura**, None

Corresponding author: Kenji Yamashiro, Department of Ophthalmology and Visual Sciences, Kyoto University Graduate School of Medicine, 54 Kawahara, Shogoin, Sakyo, Kyoto 606-8507, Japan; yamashro@kuhp.kyoto-u.ac.jp.

Myopia is one of the most common ocular disorders worldwide. The prevalence of myopia is much higher in Asian populations, with a reported incidence of roughly 40% in the Japanese and Chinese population and 25% in Caucasians.[1-3] Pathological myopia, also called high myopia, is defined as a spherical equivalent refractive error of at least −6 diopters or an axial length ≥26.5 mm. Myopic axial length elongation can lead to chorioretinal atrophy and choroidal neovascularization (CNV), which is the most vision-threatening complication in highly myopic eyes.[4] Since the long-term visual outcomes of myopic CNV are extremely poor,[5] it is critical to determine in which highly myopic patients CNV will occur. CNV usually occurs in young adults with high myopia in the fourth and fifth decades of life. However, many eyes with high myopia do not have CNV even after 60 years of age. Furthermore, the size of the CNV seriously affects the visual prognosis because it determines the size of the scotoma, and some smaller CNVs can regress without treatment.[6] Since it is difficult to prevent the development of myopia, it is important to investigate the mechanisms underlying CNV occurrence and growth in myopic eyes; this may lead to the prevention of CNV development and the subsequent visual disturbance.

Genetic backgrounds may affect the development of high myopia; recently, we have determined a susceptible locus for pathological myopia using a genome-wide association study (GWAS).[7] Furthermore, recent GWASs reveal that myopia susceptibility loci exist in chromosome 15.[8-10] The occurrence of CNV in highly myopic eyes might also depend on genetic variations. Thus far, however, few studies have investigated the genetic background of patients with CNV in highly myopic eyes.

Since anti-VEGF treatment has been developed for neovascular AMD, it has become a popular treatment for ocular neovascularization. Anti-VEGF drugs have been shown to be effective in treating CNV secondary to high myopia.[11-13] In contrast to neovascular AMD, myopic CNV is easily inactivated with anti-VEGF treatment. In this study, we evaluated the associations between VEGF gene polymorphisms and CNV development in highly myopic eyes in Japanese patients.

Methods

This study was performed in accordance with the tenets of the Declaration of Helsinki. The Institutional Review Board/Ethics Committee of each institution approved the study protocols. All patients were fully informed of the study purpose and procedures, and written consent was obtained from each patient. For the case-control study of CNV occurrence, 327 highly myopic, unrelated Japanese patients with axial lengths of >26.0 mm in both eyes and who were ≥60 years of age were recruited from Kyoto University Hospital and Tokyo Medical and

Table 1. Characteristics of the Study Population

	With CNV	Without CNV	*P* Value
Number	184	143	
Mean age ± SD (years)	69.97 ± 6.35	69.23 ± 6.74	0.52*
Axial length ± SD (mm)			
Right	28.97 ± 1.72	29.11 ± 1.72	0.49*
Left	28.75 ± 1.72	28.84 ± 1.86	0.68*
Sex (male/female)	32/152	58/85	3.27×10^{-6}†

* Unpaired *t*-test.
† χ^2 test.

Dental University Hospital. The number of patients with macular CNV in at least one eye was 184, and the number of patients without macular CNV in either eye was 143 (Table 1). All patients underwent detailed ophthalmologic examinations, including dilated indirect and contact lens slit-lamp biomicroscopy, automatic objective refraction, measurement of the axial length by A-scan ultrasound (UD-6000; Tomey, Nagoya, Japan) or partial coherence interferometry (IOLMaster; Carl Zeiss Meditec, Dublin, CA), color fundus photography, optical coherence tomography, and fluorescein angiography. Individuals with a history of ocular surgery, with the exception of cataract surgery, were excluded from the study. Patients with secondary choroidal neovascular diseases, such as angioid streaks, presumed ocular histoplasmosis syndrome, and ocular trauma, were also excluded.

Of the 184 patients with myopic CNV, 83 patients underwent angiography with HRA2 (Heidelberg Engineering, Heidelberg, Germany) in Kyoto University Hospital. To evaluate the association between VEGF gene polymorphisms and CNV size, the area of CNV (mm^2) in these 83 patients was measured with the HRA-2 software. An additional 76 patients with myopic CNV were enrolled from Kyoto University Hospital to confirm the aforementioned associations. The average age of these patients was 63.8 ± 12.6 years, and the average axial length was 30.1 ± 1.1 mm.

For selecting tag single nucleotide polymorphisms (SNPs), we used the public dbSNP database build 126 (NCBI build 36.1) and HapMap database phase 2, release 22,[35] to extract the relevant sequencing information for the *VEGFA* gene and the genotyping information for the SNPs. A set of four tagging VEGF SNPs were selected for investigation: two SNPs on the promoter region, named rs699946 and rs699947, and two intronic SNPs, rs3025033 and rs3025035. This set of four tagging SNPs provided 100% coverage for all 14 common HapMap SNPs within a 26.3 kb region (16.3 kb gene length; 10 kb upstream) spanning the VEGF gene on chromosome 6 (r^2 threshold of 0.95). Furthermore, we evaluated four functional SNPs (rs1570360, rs2010963, rs833061, and rs3025039). Since these SNPs have been shown to affect VEGF expression,[14-17] many studies have evaluated the association of these SNPs with various diseases such as AMD, diabetic retinopathy, Behçet's disease, Alzheimer's disease, and diabetes.[18-26]

Genomic DNA was prepared from peripheral blood by a DNA extraction kit (QuickGene-610L; Fujifilm, Minato, Tokyo, Japan). VEGF-tagged SNPs (rs699946, rs699947, rs3025033, and rs3025035) and functional SNPs (rs1570360, rs2010963, rs833061, and rs3025039) were genotyped by a Taqman SNP assay with the ABI PRISM 7700 system (Applied Biosystems, Foster, CA). Deviations in genotype distributions from the Hardy-Weinberg equilibrium (HWE) were assessed with the HWE exact test. A χ^2 test for trend or its exact counterpart was used to compare the genotype distributions of the two groups. To adjust for age and sex, we performed logistic regression analysis. Mean age and axial length were compared using unpaired *t*-test or ANOVA, and sex ratio was compared with the χ^2 test. The associations between genotype and CNV size were evaluated using the Jonckheere-Terpstra trend test. *P* values of less than 0.05 were considered statistically significant.

Results

The demographics of the study population are shown in Table 1; there was no significant difference between patients with CNV and patients without CNV with respect to either age or axial length. The mean age of each group was 70.0 ± 6.4 years and 69.2 ± 6.7 years, respectively ($P = 0.52$). However, CNV is more predominant in women compared with men ($P = 3.27 \times 10^{-6}$) with an odds ratio (OR) of 3.24 (95% confidence interval [CI] = 2.27–4.64).

The genotype counts, associations, and ORs for the eight SNPs are shown in Table 2. The genotype distributions were not significantly different between patients with CNV and patients without CNV (nominal $P > 0.16$). Evaluation of the associations in a recessive model and a dominant model also showed no associations ($P > 0.10$). Even when adjusted for age and sex, the genotype distributions were not significantly different ($P > 0.10$).

In addition, we performed subset analysis for patients aged 70 years or older. In our cohort, 86 patients with CNV and 63 patients without CNV were ≥70 years of age. Associations between the eight SNPs with the occurrence of CNV were not statistically significant ($P > 0.17$).

Of the 184 patients with myopic CNV, the area of CNV was measured in 83 patients who underwent angiography with HRA2 in Kyoto University Hospital. The genotype distribution of rs2010963 was significantly correlated with CNV area ($P = 0.0047$), while the other seven SNPs did not show significant associations with CNV area (Fig. 1). The size of CNV was largest (1.71 ± 1.29 mm^2) in patients with a CC genotype of rs2010963, intermediate (0.98 ± 0.84 mm^2) with a CG genotype, and smallest (0.78 ± 0.78 mm^2) with a GG genotype. There was no significant difference in axial length, age of patients, or male/female ratio among the three

Table 2. Genotype Counts, Associations, and Odds Ratios for VEGF SNPs

		CNV (+)			CNV (−)				Age- and Sex-Adjusted	
SNP	Genotype	Genotype Count	MAF	HWE *P*	Genotype Count	MAF	HWE *P*	Nominal *P*	*P*	OR (95% CI)
rs699946	AA/AG/GG	64/82/33	G, 0.41	0.399	40/73/23	G, 0.44	0.250	0.543	0.10	0.80 (0.62–1.04)
rs699947	AA/AC/CC	22/77/85	A, 0.33	0.477	17/60/63	A, 0.34	0.626	0.856	0.68	0.93 (0.66–1.31)
rs3025033	AA/AG/GG	125/53/4	G, 0.17	0.286	90/44/8	G, 0.21	0.151	0.160	0.60	0.94 (0.73–1.20)
rs3025035	CC/CT/TT	90/71/17	T, 0.29	0.391	79/49/12	T, 0.26	0.200	0.355	0.34	1.13 (0.88–1.44)
rs1570360	AA/AG/GG	11/42/130	A, 0.17	0.005	8/32/102	A, 0.17	0.020	0.858	0.79	0.94 (0.60–1.47)
rs2010963	CC/GC/GG	34/84/62	C, 0.42	0.547	23/73/42	C, 0.43	0.348	0.820	0.42	0.88 (0.65–1.20)
rs833061	CC/CT/TT	22/75/82	C, 0.33	0.451	17/60/66	C, 0.33	0.554	0.922	0.69	0.93 (0.66–1.31)
rs3025039	CC/CT/TT	116/56/5	T, 0.19	0.402	87/45/8	T, 0.22	0.298	0.328	0.81	0.97 (0.76–1.24)

MAF, minor allele frequency.

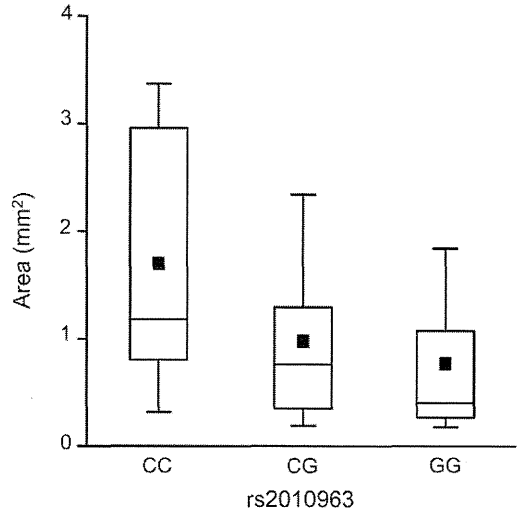

FIGURE 1. The area (mm²) of choroidal neovascularization among the three genotypes of rs2010963 in 83 patients. The area was significantly associated with the genotype ($P = 0.0047$).

genotypes of rs2010963 ($P = 0.54$, 0.98, and 0.69, respectively). To confirm the aforementioned association between rs2010963 and CNV size, we genotyped rs2010963 in an additional 76 patients with myopic CNV (20 male and 56 female). The genotype distribution of rs2010963 was significantly correlated with the CNV area ($P = 0.032$), while there was no significant difference in the axial length, age of patients, or male/female ratio among the three genotypes of rs2010963 ($P = 0.91$, 0.15, and 0.20, respectively). When these two cohorts were pooled for further evaluation of this association, the genotype distribution of rs2010963 was significantly correlated with the CNV area (Fig. 2, $P = 0.00078$).

DISCUSSION

In the present study, we found no association between VEGF gene polymorphisms and the occurrence of CNV in highly myopic eyes in Japanese patients, although rs2010963 was significantly associated with the size of CNV. To evaluate factors associated with CNV occurrence in highly myopic eyes, the age of the cohort is of critical importance. Therefore, when a younger cohort is used, some patients assigned to the group without CNV may eventually develop CNV, which can obscure potential differences between the two groups. Fernandez-Robredo et al. have evaluated the association of CFH Y402H and ARMS2 A69S polymorphisms with myopic CNV using 196 myopic patients who were aged ≥30 years.[27] We have previously evaluated the same association using 353 myopic patients who were ≥50 years of age,[28] and the present study consisted of 327 myopic patients who were aged ≥60 years. However, the association of VEGF gene polymorphism with CNV occurrence was not statistically significant. Furthermore, we evaluated the association using a cohort of patients older than 70 years, but statistical significance was still not found.

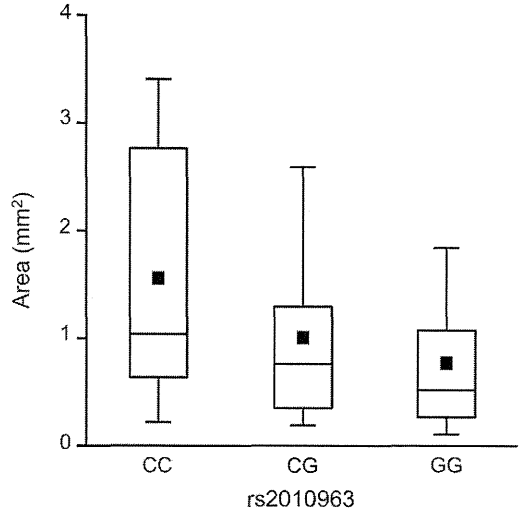

FIGURE 2. The area (mm²) of choroidal neovascularization among the three genotypes of rs2010963 in 159 patients. The area was significantly associated with the genotype ($P = 0.00078$).

Genetic associations with myopia have been investigated for several decades. Linkage studies have identified 18 possible loci for myopia (MYP1-18). Numerous candidate genes have been evaluated, and we have recently completed a GWAS study.[7] Furthermore, recent GWAS studies have revealed myopia susceptibility loci on chromosome 15, and we have successfully reproduced the association of these susceptibility loci with high myopia.[8–10] However, susceptibility genes for myopia have not been revealed; this makes it difficult to determine how to prevent myopia. Compared with the prevention of myopia, prevention and/or control of CNV occurrence and growth in highly myopic eyes might be a more practical approach. Since CNV is one of the most vision-threatening complications in highly myopic eyes, it is of great value to investigate the mechanism underlying CNV development in these eyes.

Although anti-VEGF treatments have been developed for the management of neovascular AMD, they are also substantially effective in treating myopic CNV.[11–13] Considering the effectiveness of these anti-VEGF treatments, we had hypothesized that VEGF is associated with the occurrence of CNV in highly myopic eyes. The present study, however, suggests that VEGF gene variations do not affect the occurrence of CNV in these eyes. In contrast with CNV occurrence, VEGF gene polymorphism rs2010963 was significantly associated with CNV size. Thus, it appears that VEGF contributes to CNV growth rather than CNV occurrence in highly myopic eyes. Experimental studies have shown that inhibition of VEGF leads to smaller CNV in laser-induced CNV models.[29–31] However, inhibition of VEGF does not always completely suppress CNV occurrence after laser photocoagulation to disrupt Bruch's membrane. This evidence suggests that VEGF only affects CNV size/growth, and other factors are responsible for triggering CNV occurrence, partly by interacting with Bruch's membrane.

The size of CNV is critical for visual prognosis in highly myopic eyes. Smaller CNVs can lead to smaller scotomas and spare the visual functions of the surrounding retina. Furthermore, very small CNVs can disappear completely after treatment.[6] Our findings suggest that development of larger CNVs in highly myopic eyes can be prevented by targeting VEGF, while prevention of CNV occurrence might be accomplished by targeting other factors.

Watson et al. reported that the amount of lipopolysaccharide-induced VEGF production from peripheral blood mononuclear cells (PBMNCs) is highest in individuals with a GG genotype of rs2010963, intermediate with a CG genotype, and lowest with a CC genotype.[17] In contrast to the findings of this study, we discovered that the size of CNV was largest in patients with a CC genotype, intermediate with a CG genotype, and smallest with a GG genotype. Considering that VEGF is a pro-angiogenic factor, these two findings seem contradictory. However, an evaluation of PBMNC function in in-vitro studies

does not always reflect their function in in-vivo situations. Furthermore, PBMNCs include several cell types such as lymphocytes, monocytes, and macrophages, and we have performed in vivo experiments that show that PBMNCs induce endothelium apoptosis[32] and that lymphocytes are negative regulators of pathological neovascularization, while monocytes are positive regulators in an ischemic retinopathy model.[33] Further studies are required to evaluate the roles of VEGF produced individually by monocytes or lymphocytes during myopic CNV development. In addition to VEGF produced from PBMNCs, VEGF produced from the RPE could also affect the growth of CNV in highly myopic eyes. Although we cannot evaluate the VEGF-producing ability of the RPE in an in-vivo situation, elucidation of the roles of the RPE in myopic CNV development might lead to better control of CNV size. It is also important to consider that VEGF can have several isoforms with different properties; we have demonstrated that VEGF165 is associated with pathological neovascularization, while VEGF121 is associated with physiological neovascularization.[33] Furthermore, recent studies have shown that some isoforms of VEGF are anti-angiogenic.[34] Additional studies on the role of different VEGF isoforms in myopic CNV development may lead to prevention of larger CNV secondary to high myopia.

Limitations of the present study include the age of the cohort and the small sample size. Although we used a cohort older than 60 years of age and performed a subanalysis using samples with patients older than 70 years, some participants included in the group without CNV might develop CNV in the future. Furthermore, our study is retrospective in nature, and the associations discovered herein need to be evaluated in prospective studies.

In conclusion, we have shown that VEGF gene polymorphisms have no association with the occurrence of CNV in highly myopic eyes in Japanese individuals; however, VEGF rs2010963 affects the size of CNV. Treatments that target VEGF may prevent large CNV formation in highly myopic eyes and help achieve better visual prognosis. To prevent CNV occurrence, further studies are needed to clarify the mechanism and/or background causes of CNV occurrence in highly myopic eyes.

References

1. Sawada A, Tomidokoro A, Araie M, Iwase A, Yamamoto T. Refractive errors in an elderly Japanese population: the Tajimi study. *Ophthalmology.* 2008;115:363-370, e363. http://www.ophsource.org/periodicals/ophtha/article/S0161-6420%2807%2900379. Accessed March 29, 2012.
2. Wong TY, Foster PJ, Hee J, et al. Prevalence and risk factors for refractive errors in adult Chinese in Singapore. *Invest Ophthalmol Vis Sci.* 2000;41:2486-2494.
3. Kempen JH, Mitchell P, Lee KE, et al. The prevalence of refractive errors among adults in the United States, Western Europe, and Australia. *Arch Ophthalmol.* 2004;122:495-505.
4. Hayashi K, Ohno-Matsui K, Shimada N, et al. Long-term pattern of progression of myopic maculopathy a natural history study. *Ophthalmology.* 2010; 117:1595-1611.
5. Yoshida T, Ohno-Matsui K, Yasuzumi K, et al. Myopic choroidal neovascularization: a 10-year follow-up. *Ophthalmology.* 2003; 110:1297-1305.
6. Hayashi K, Ohno-Matsui K, Yoshida T, et al. Characteristics of patients with a favorable natural course of myopic choroidal neovascularization. *Graefes Arch Clin Exp Ophthalmol.* 2005; 243:13-19.
7. Nakanishi H, Yamada R, Gotoh N, et al. A genome-wide association analysis identified a novel susceptible locus for pathological myopia at 11q24.1. *PLoS Genet.* 2009;5:e1000660. http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.100. Accessed March 29, 2012.
8. Solouki AM, Verhoeven VJ, van Duijn CM, et al. A genome-wide association study identifies a susceptibility locus for refractive errors and myopia at 15q14. *Nat Genet.* 2010;42:897-901.
9. Hysi PG, Young TL, Mackey DA, et al. A genome-wide association study for myopia and refractive error identifies a susceptibility locus at 15q25. *Nat Genet.* 2010;42:902-905.
10. Hayashi H, Yamashiro K, Nakanishi H, et al. Association of 15q14 and 15q25 with high myopia in Japanese. *Invest Ophthalmol Vis Sci.* 2011;52:4853-4858.
11. Chan WM, Lai TY, Liu DT, Lam DS. Intravitreal bevacizumab (Avastin) for myopic choroidal neovascularization: six-month results of a prospective pilot study. *Ophthalmology.* 2007;114:2190-2196.
12. Gharbiya M, Allievi F, Mazzeo L, Gabrieli CB. Intravitreal bevacizumab treatment for choroidal neovascularization in pathologic myopia: 12-month results. *Am J Ophthalmol.* 2009; 147:84-93, e81. http://www.ajo.com/. Accessed March 29, 2012.
13. Wu PC, Chen YJ. Intravitreal injection of bevacizumab for myopic choroidal neovascularization: 1-year follow-up. *Eye (Lond).* 2009;23:2042-2045.
14. Renner W, Kotschan S, Hoffmann C, Obermayer-Pietsch B, Pilger E. A common 936 C/T mutation in the gene for vascular endothelial growth factor is associated with vascular endothelial growth factor plasma levels. *J Vasc Res.* 2000;37:443-448.
15. Shahbazi M, Fryer AA, Pravica V, et al. Vascular endothelial growth factor gene polymorphisms are associated with acute renal allograft rejection. *J Am Soc Nephrol.* 2002;13:260-264.
16. Stevens A, Soden J, Brenchley PE, Ralph S, Ray DW. Haplotype analysis of the polymorphic human vascular endothelial growth factor gene promoter. *Cancer Res.* 2003;63:812-816.
17. Watson CJ, Webb NJ, Bottomley MJ, Brenchley PE. Identification of polymorphisms within the vascular endothelial growth factor (VEGF) gene: correlation with variation in VEGF protein production, Cytokine. 2000;12:1232-1235.
18. Mori K, Horie-Inoue K, Gehlbach PL, et al. Phenotype and genotype characteristics of age-related macular degeneration in a Japanese population. *Ophthalmology.* 2010;117:928-938.
19. Lin JM, Wan L, Tsai YY, et al. Vascular endothelial growth factor gene polymorphisms in age-related macular degeneration. *Am J Ophthalmol.* 2008;145:1045-1051.
20. Fang AM, Lee AY, Kulkarni M, Osborn MP, Brantley MA Jr. Polymorphisms in the VEGFA and VEGFR-2 genes and neovascular age-related macular degeneration. *Mol Vis.* 2009; 15:2710-2719.
21. Janik-Papis K, Zaras M, Krzyzanowska A, et al. Association between vascular endothelial growth factor gene polymorphisms and age-related macular degeneration in a Polish population. *Exp Mol Pathol.* 2009;87:234-238.
22. Kangas-Kontio T, Vavuli S, Kakko SJ, et al. Polymorphism of the manganese superoxide dismutase gene but not of vascular endothelial growth factor gene is a risk factor for diabetic retinopathy. *Br J Ophthalmol.* 2009;93:1401-1406.
23. Chun MY, Hwang HS, Cho HY, et al. Association of vascular endothelial growth factor polymorphisms with nonproliferative and proliferative diabetic retinopathy. *J Clin Endocrinol Metab.* 2010;95:3547-3551.
24. Salvarani C, Boiardi L, Casali B, et al. Vascular endothelial growth factor gene polymorphisms in Behcet's disease. *J Rheumatol.* 2004;31:1785-1789.
25. Landgren S, Palmer MS, Skoog I, et al. No association of VEGF polymorphisms with Alzheimer's disease. *Neuromolecular Med.* 2010; 12:224-228.

26. Freathy RM, Weedon MN, Shields B, et al. Functional variation in VEGF is not associated with type 2 diabetes in a United Kingdom Caucasian population. *JOP.* 2006;7:295–302.
27. Fernandez-Robredo P, Maestre SR, Zarranz-Ventura J, Mulero HH, Salinas-Alaman A, Garcia-Layana A. Myopic choroidal neovascularization genetics. *Ophthalmology.* 2008;115:1632, 1632. e1. http://www.ophsource.org/periodicals/ophtha/article/S0161-6420(08)00226-1/. Accessed March 29, 2012.
28. Nakanishi H, Gotoh N, Yamada R, et al. ARMS2/HTRA1 and CFH polymorphisms are not associated with choroidal neovascularization in highly myopic eyes of the elderly Japanese population. *Eye (Lond).* 2010;24:1078–1084.
29. Seo MS, Kwak N, Ozaki H, et al. Dramatic inhibition of retinal and choroidal neovascularization by oral administration of a kinase inhibitor. *Am J Pathol.* 1999;154:1743–1753.
30. Kwak N, Okamoto N, Wood JM, Campochiaro PA. VEGF is major stimulator in model of choroidal neovascularization. *Invest Ophthalmol Vis Sci.* 2000;41:3158–3164.
31. Krzystolik MG, Afshari MA, Adamis AP, et al. Prevention of experimental choroidal neovascularization with intravitreal anti-vascular endothelial growth factor antibody fragment. *Arch Ophthalmol.* 2002;120:338–346.
32. Ishida S, Yamashiro K, Usui T, et al. Leukocytes mediate retinal vascular remodeling during development and vaso-obliteration in disease. *Nat Med.* 2003;9:781–788.
33. Ishida S, Usui T, Yamashiro K, et al. VEGF164-mediated inflammation is required for pathological, but not physiological, ischemia-induced retinal neovascularization. *J Exp Med.* 2003;198:483–489.
34. Bates DO, Cui TG, Doughty JM, et al. VEGF165b, an inhibitory splice variant of vascular endothelial growth factor, is down-regulated in renal cell carcinoma. *Cancer Res.* 2002;62:4123–4131.
35. The International HapMap Project. International HapMap Consortium. *Nature.* 2003;426:789–796.

Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes

Lucia Kato[a], Nasim A. Begum[a], A. Maxwell Burroughs[b], Tomomitsu Doi[a,1], Jun Kawai[b], Carsten O. Daub[b], Takahisa Kawaguchi[c], Fumihiko Matsuda[c], Yoshihide Hayashizaki[b], and Tasuku Honjo[a,2]

[a]Department of Immunology and Genomic Medicine and [c]The Center for Genomic Medicine, Graduate School of Medicine, Kyoto University, Kyoto 606-8501, Japan; and [b]RIKEN Omics Science Center (OSC), RIKEN Yokohama Institute, Yokohama, Kanagawa 230-0045, Japan

Contributed by Tasuku Honjo, December 28, 2011 (sent for review December 5, 2011)

Activation-induced cytidine deaminase (AID) is required for both somatic hypermutation and class-switch recombination in activated B cells. AID is also known to target nonimmunoglobulin genes and introduce mutations or chromosomal translocations, eventually causing tumors. To identify as-yet-unknown AID targets, we screened early AID-induced DNA breaks by using two independent genome-wide approaches. Along with known AID targets, this screen identified a set of unique genes (*SNHG3, MALAT1, BCL7A,* and *CUX1*) and confirmed that these loci accumulated mutations as frequently as *Ig* locus after AID activation. Moreover, these genes share three important characteristics with the Ig gene: translocations in tumors, repetitive sequences, and the epigenetic modification of chromatin by H3K4 trimethylation in the vicinity of cleavage sites.

deep sequencing | end labeling by biotin oligonucleotide | microarray

Activation-induced cytidine deaminase (AID) is expressed in germinal center (GC) B cells upon antigen stimulation and is essential for two types of genetic alteration in the Ig gene: class switch recombination (CSR) and somatic hypermutation (SHM), which provide the genetic basis for antibody memory (1, 2). CSR produces antibodies with different effector functions by recombination at Ig heavy chain (H) switch (S) regions, so that the μ-chain constant (Cμ) region is replaced by a downstream C_H region. SHM introduces nontemplated point mutations in the rearranged variable (V) region genes, resulting in incremented antigen receptor affinity after clonal selection (3, 4).

Functional studies on AID mutants have shown that distinct AID domains are required for SHM and CSR, although AID has a single catalytic center (cytidine deaminase motif) in the middle of the molecule. Deletions and alterations in the N-terminal region affect both the CSR and SHM activities (5). However, AID C-terminal mutants almost completely lose CSR activity but retain or even increase SHM activity (6, 7). Although C-terminally truncated AID mutants cleave both V and S regions and induce enhanced c-myc-IgH translocations, they cannot mediate CSR, suggesting that the C-terminal domain is not required for DNA cleavage but is required to correctly pair cleaved ends (8).

The DNA cleavage of targets in CSR and SHM (the S region and V region, respectively) requires their transcription (9–12). Indeed, AID-induced mutations (SHM) are generally detected in a region within 2 kb downstream of the transcription start site (TSS) (13, 14). Transcription appears to play two roles in the targeting of cleavage sites. First, transcription is associated with the epigenetic marking of the target locus, particularly by H3K4 trimethylation (H3K4me3). The histone chaperone complex FACT is required to regulate H3K4me3 in the target S region, and FACT knockdown abolishes H3K4me3 and DNA cleavage in this region (15). Second, transcription is probably required to induce non-B structures in highly repetitive sequences such as S regions (16–18), due to excessive negative supercoiling induced immediately downstream of transcription. V regions have also been shown to form stem-loop structures under these conditions (19, 20). Non-B structure involvement has recently been reported in transcription-associated mutations in repetitive sequences such as the dinucleotide repeat hot spots or triplet repeat expansion/contractions causing Huntington's disease (17, 21, 22).

AID-dependent DNA cleavage is, in general, specific to the Ig locus. However, a number of reports have shown that AID can induce DNA cleavage in non-Ig loci. AID non-Ig targets were first demonstrated by studies on AID transgenic mice that produce numerous T lymphomas, in which vast numbers of mutations accumulate in the genes encoding the T-cell receptor, CD4, CD5, c-myc, and PIM1 (23, 24). This finding was followed by the observations that AID deficiency abolishes c-myc-Ig translocation and reduces the incidence of plasmacytoma (25, 26). AID expression is specific to activated B cells under normal conditions. However, AID expression has also been found in non-B cells, especially in cells stimulated by infection with pathogens such as human T-cell leukemia virus type 1 (HTLV1), hepatitis C virus (HCV), Epstein–Barr (EB) virus, and *Helicobacter pylori* (27–30). Based on these observations, AID is postulated to induce tumorigenesis, especially in B lymphomas and leukemias—and AID is expressed in many GC-derived human B-cell lymphomas (31–33). The prognosis of acute lymphocytic leukemia (ALL) and chronic myeloid leukemia (CML) is linked with AID expression (34, 35). It is therefore important to determine which non-Ig genes can be targeted by AID, and what features, if any, they share with Ig genes.

Several approaches have been used to explore AID non-Ig target genes in B cells. Candidate approaches involving the direct sequencing of proto-oncogenes, genes involved in translocations, or genes transcribed in normal GC B cells have shown that AID mutates several non-Ig genes, including *BCL6*, *MYC*, *PIM1*, and *PAX5* (24, 32, 36, 37). More recently, several efforts have been made to identify AID targets in a whole genome. These approaches have used chromatin immunoprecipitation (ChIP) of CSR-related proteins in combination with genome-wide tiling microarrays (ChIP-chip) or deep sequencing (ChIP-seq) on the assumption that proteins involved in CSR bind to AID targets. RPA, Nbs1, AID itself, and Spt5 have been used as marking proteins in this type of study (38–40). However, these approaches did not necessarily show that all of the protein-bound targets are cleaved or mutated by AID. There are indications that some genes identified by such approaches are not tran-

Author contributions: L.K., T.D., and T.H. designed research; L.K., N.A.B., and A.M.B. performed research; T.K. and F.M. contributed new reagents/analytic tools; L.K., N.A.B., A.M.B., J.K., C.O.D., and Y.H. analyzed data; and L.K. and T.H. wrote the paper.

The authors declare no conflict of interest.

[1]Present address: Laboratory Animal Research Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.

[2]To whom correspondence should be addressed. E-mail: honjo@mfour.med.kyoto-u.ac.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1120791109/-/DCSupplemental.

scribed (39). Therefore, it is important to reexamine non-Ig AID target genes by using a different strategy.

Here, we report four AID targets, identified by a combination of unique techniques. After directly labeling the DNA breakage ends from AID-induced cleavage with a biotinylated linker, we isolated the labeled fragments with streptavidin beads and analyzed them by a combination of promoter arrays and genome-wide sequencing. The candidates identified were then confirmed by quantitative PCR (qPCR) and the actual demonstration of mutations. With these methods, we identified at least four previously unknown AID targets—*SNHG3, MALAT1, BCL7A,* and *CUX1*. We found that these targets share important characteristics with Ig genes, namely, repetitive sequences that can form non-B structures upon efficient transcription, and the accumulation of H3K4me3 histone modifications on the chromatin.

Results

AID-Induced DNA Cleavage Detected by Labeling DNA Break Ends with a Biotinylated Linker. To detect genome-wide AID-induced DNA breaks, we used a modified in situ DNA end-labeling technique as described (8, 41) in BL2 cells, a Burkitt's lymphoma cell line that serves as an in vitro model for studying the SHM mechanism (31, 42, 43). We used the BL2 clone BL2-ΔC-AIDER, which expresses JP8Bdel, an AID mutant lacking the C-terminal 16 residues, fused with the hormone-binding domain of the estrogen receptor (ER) (JP8Bdel-ER). Tamoxifen (4-OHT) treatment induces DNA breakage in the Sμ and Sα regions but not in the Sγ region of JP8Bdel-ER-expressing CH12 cells, which switch almost exclusively from IgM to IgA (8).

BL2-ΔC-AIDER cells were treated with 4-OHT only for 3 h to minimize cell death and DNA break ends were labeled with a biotinylated linker, and the break-enriched biotinylated DNA was used as a PCR template (Fig. 1*A*). In agreement with previous reports (8, 42), we detected DNA breakage in the 5′ Sμ region of the IgH locus only in 4-OHT–treated cells. No breakage was detected in the *B2M* gene, which is expressed in BL2 cells but was shown not to accumulate mutations in activated B cells (Fig. 1*B*).

AID Targets Identified by Promoter Array and Whole Genome Sequencing. Because SHM is normally detected close to the TSS (13, 14), biotin linker-enriched DNA fragments were analyzed by a promoter array to identify unknown AID targets. Table S1 lists the genes whose signals increased after 3 h of 4-OHT treatment, compared with untreated samples with false discovery rate (FDR) values <0.3. We also looked for genes with increased signals after 4-OHT treatment that are known to be targets of chromosomal translocation or genes that had multiple breakage peaks, and we identified >50 genes, among which we found that *BCL7A* and *CUX1* are enriched in the original breakage-enriched library by qPCR (see below). We confirmed by RT-PCR and expression array that *SNHG3*, *MALAT1*, *NIN*, *C9orf72*, *CFLAR*, *SNX25*, *BCL7A,* and *CUX1* were transcribed in BL2 cells (Table S1). Fig. S1 shows the peak signals in a 10-kb segment surrounding the breakage area of *SNHG3, MALAT1, BCL7A,* and *CUX1*. We could not map the breakage in the Ig locus because of the absence of array probes in this region.

Because the promoter array does not detect DNA fragments outside of regions containing probes, we further analyzed the breakage-enriched DNA by direct sequencing of the biotin linker-enriched library. DNA breakage sites in both control and 4-OHT–treated libraries were identified by aligning sequenced tags to the genome, and significantly enriched regions were identified by comparing the local breakage density (*SI Materials and Methods*). Regions were identified in the genes listed in Table S2. Interestingly, *SNHG3* and *MALAT1*, which were identified by the promoter array, appear at the top of the list in the genome-wide sequencing as well.

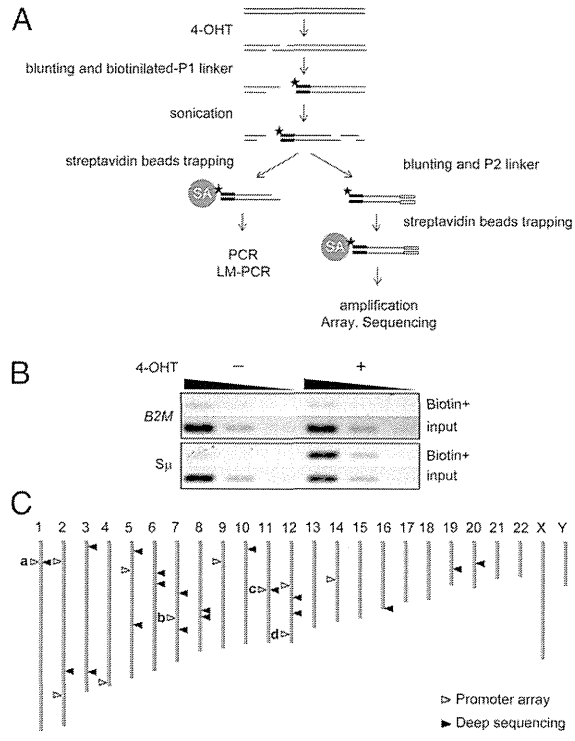

Fig. 1. (*A*) Schematic of the labeling technique. 4-OHT is added to activate AID, and DNA break ends are labeled in situ by biotinylated linker ligation. After genomic DNA is extracted and sonicated, biotinylated fragments are captured by streptavidin beads and used for PCR, array, or sequencing. (*B*) Detection of DNA breaks by PCR. BL2-ΔC-AIDER cells were treated with or without 4-OHT for 3 h, and the break ends were labeled. PCR of Sμ and *B2M* was performed with biotin-labeled DNA or input DNA by using fivefold serially diluted templates. (*C*) Chromosomal distribution of AID targets. a, *SNHG3*; b, *CUX1*; c, *MALAT1*; d, *BCL7A*. White arrowhead, promoter array (FDR < 0.3 plus *BCL7A* and *CUX1*); black arrowhead, whole genome sequencing (FDR < 0.01 and/or remarkable numbers of *P* value clusters).

Fig. 1*C* shows the chromosomal distribution of AID target candidates identified by promoter array or whole-genome sequencing. Breakage seemed to be distributed through the genome without any apparent bias. Surprisingly, of the 29 candidates identified by whole-genome sequencing with strict statistical parameters, only two matched candidates obtained from the promoter array. This discrepancy might be explained in part because most of the breakage-rich regions detected by whole genome sequencing are located in regions that do not contain promoter array probes.

Results may also be limited because of possible bias by PCR amplification of the primary library for microarray and whole-genome sequencing, which could affect the relative genome coverage. To avoid this bias, we relied on the original library and confirmed all candidates by qPCR.

qPCR Analyses of Linker Libraries. To confirm the AID-induced breakage candidates detected by the promoter array and whole-genome sequencing, we used qPCR assays with gene-specific primers to amplify the vicinity of the identified breakage regions in biotin linker-enriched DNA from cells treated with 4-OHT for 3 h (Fig. 2). We examined whether candidate genes were enriched in the 4-OHT–treated DNA library compared with the nontreated library. Among the 29 candidates identified by whole-genome sequencing, only *SNHG3* and *MALAT1* were strongly enriched ($P < 0.0001$ and $P < 0.001$, respectively). Besides these, *BCL7A*, *CUX1,* and *CFLAR*, which were picked up only by the

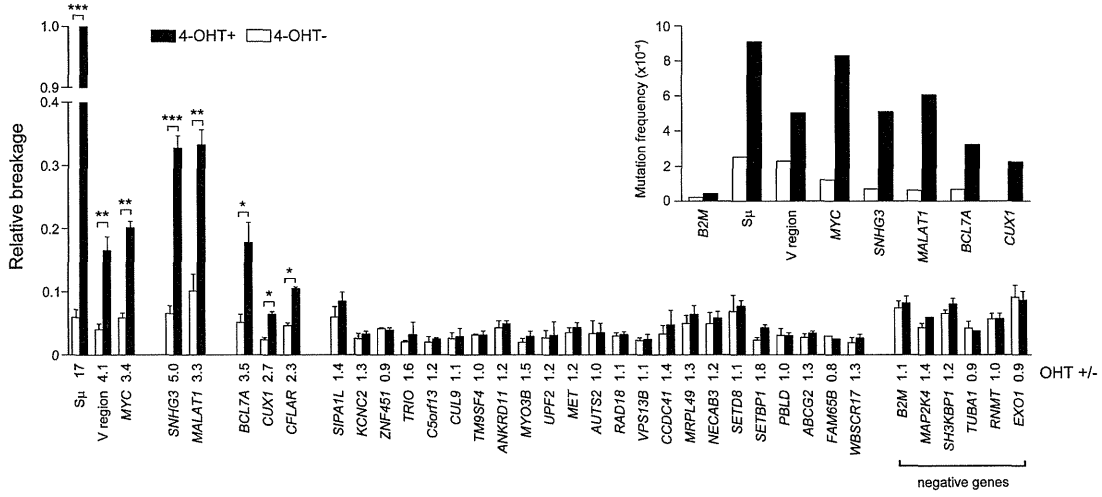

Fig. 2. qPCR measurement of DNA breaks. Break signals are presented relative to Sμ. SD values were derived from at least three independent experiments, and *P* values were calculated by a two-tailed *t* test. *$P < 0.01$, **$P < 0.001$, ***$P < 0.0001$. Numbers below the *x* axis indicate the ratio between samples treated and not treated with 4-OHT. (*Inset*) Mutation analysis of genes with significantly increased break signals after AID activation. Cells were treated with or without 4-OHT for 24 h. Only unique mutations were counted. Detailed mutation profiles can be found in Fig. S2 and Table S3.

promoter array, also showed significant enrichment ($P < 0.01$) in the 4-OHT–treated library.

We also confirmed that the Sμ and V regions in BL2 cells were cleaved, because they were enriched in the 4-OHT–treated library. Although *MYC*, which is translocated in an AID-dependent manner in human Burkitt's lymphoma (44), was not identified by either promoter array or whole-genome sequencing, qPCR of the 4-OHT–treated samples clearly revealed *MYC* gene enrichment (Fig. 2). The difference in cleavage detection between the direct candidate qPCR and genome-wide arrays and sequencing suggests that the amplification step required for microarray and whole-genome sequencing methods may introduce bias, either for or against many genes. In the case of sequencing, this bias can lead to low mapping coverage of certain regions, hampering efforts to identify significant enrichment. Therefore, we cannot exclude genes that were not identified by the present methods from being AID targets.

AID Targets Accumulate Somatic Mutations near Cleavage Sites. To test whether the newly identified target genes are mutated upon AID activation, we treated BL2-ΔC-AIDER cells with 4-OHT for 24 h and sequenced regions of ≈600 bp around each area with abundant breakage (Fig. S2 and Table S3). Mutations increased in all of the qPCR-confirmed AID target genes after 4-OHT treatment (Fig. 2, *Inset*), with mutation frequencies ranging from 6.1×10^{-4} for *MALAT1* to 2.2×10^{-4} for *CUX*. These frequencies are comparable to those of the V region (5.0×10^{-4}), the Sμ region (9.1×10^{-4}), and the *MYC* gene (8.3×10^{-4}), and are far higher than that of the control *B2M* gene (4.3×10^{-5}). We also detected mutations in the *CFLAR* gene; however, the mutation frequency (9.2×10^{-5}) was not as high as other AID target genes, although mutations increased significantly in 4-OHT–treated sample ($P = 0.004$) (Table S3).

To compare the distribution profiles of mutated bases and AID-induced DNA breaks in the biotin linker-enriched DNA, we mapped the linker positions by performing ligation-mediated (LM)-PCR with the linker primer and gene-specific primers. These PCR fragments were subsequently cloned and sequenced. Break ends identified by the linker were plotted, together with mutation positions (Fig. 3 and Fig. S2). The results clearly showed that the DNA cleavage marks (biotin linker) were closely associated with mutations, indicating that the DNA cleavage sites identified are functionally relevant to SHM by AID. We used RT-PCR and expression arrays to confirm that the regions

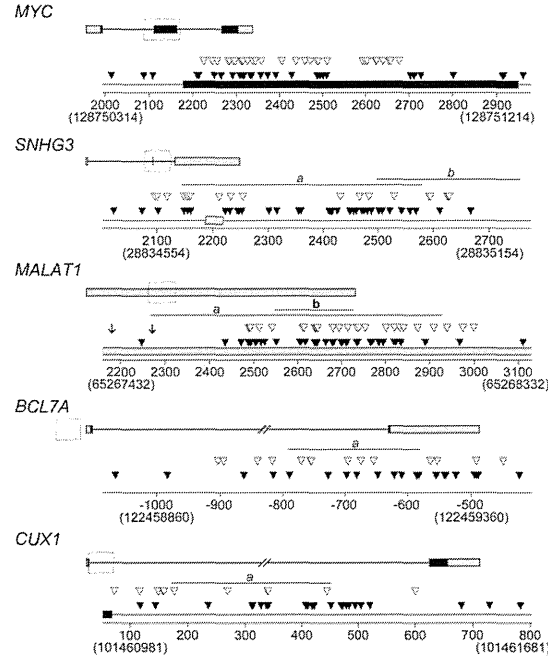

Fig. 3. Somatic mutations and breakpoint distribution in AID target loci. Mutations (open triangles) and breakpoints (filled triangles) detected by LM-PCR (Fig. S2) were plotted on the respective genomic sequences. The top scheme represents exons (rectangles) and introns (bars). Genomic loci are shown in untranslated and translated sequences (gray and black boxes, respectively). The horizontal lines *a* and *b* represent breakage regions identified by promoter array and sequencing, respectively. Regions outlined by dotted boxes are shown in more detail below each genomic locus. For the *MALAT1* locus, the translocation breakpoints reported by Davis et al. (45) are represented by arrows. *x* axis numbers indicate base positions according to RefSeq: NM_002467 (*MYC*), NR_002909 (*SNHG3*), NR_002819 (*MALAT1*), NM_020993 (*BCL7A*), and NM_181552 (*CUX1*). Numbers in parentheses indicate the corresponding base position according to hg19 assembly.