Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases

Makoto Hirata, Yoichiro Kamata, Akiko Nagai, Yutaka Kiyohara, Toshiharu Ninomiya, Akiko Tamakoshi, Zentaro Yamagata, Michiaki Kubo, Kaori Muto, Taisei Mushiroda, Yoshinori Murakami, Koichiro Yuji, Yoichi Furukawa, Hitoshi Zembutsu, Toshihiro Tanaka, Yozo Ohnishio, Yusuke Nakamura, Koichi Matsuda,

BioBank Japan Cooperative Hospital Group, Koichi Matsuda,*

Laboratory of Genome Technology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
Department of Public Policy, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Hisayama Research Institute for Lifestyle Diseases, Fukuoka, Japan
Department of Epidemiology and Public Health, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
Department of Public Health, Hokkaido University Graduate School of Medicine, Sapporo, Japan
Department of Health Sciences, University of Yamanashi, Yamanashi, Japan
RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
Laboratory for Pharmacogenomics, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan
Division of Molecular Pathology, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Project Division of International Advanced Medical Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Division of Clinical Genome Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Laboratory of Molecular Medicine, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan
SNP Research Center, RIKEN Yokohama Institute, Yokohama, Japan
Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan
Bioresource Research Center, Tokyo Medical and Dental University, Tokyo, Japan
Shinko Clinic, Medical Corporation Shinkokai, Tokyo, Japan
Section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, USA
Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan

ARTICLE INFO

Article info:
Received 22 October 2016
Accepted 15 December 2016
Available online 9 February 2017

Keywords:
BioBank Japan Project
BioBank
Common disease
Clinical information
Family history

ABSTRACT

Background: To implement personalized medicine, we established a large-scale patient cohort, BioBank Japan, in 2003. BioBank Japan contains DNA, serum, and clinical information derived from approximately 200,000 patients with 47 diseases. Serum and clinical information were collected annually until 2012. Methods: We analyzed clinical information of participants at enrollment, including age, sex, body mass index, hypertension, and smoking and drinking status, across 47 diseases, and compared the results with the Japanese database on Patient Survey and National Health and Nutrition Survey. We conducted multivariate logistic regression analysis, adjusting for sex and age, to assess the association between family history and disease development.

Results: Distribution of age at enrollment reflected the typical age of disease onset. Analysis of the clinical information revealed strong associations between smoking and chronic obstructive pulmonary disease, drinking and esophageal cancer, high body mass index and metabolic disease, and hypertension and cardiovascular disease. Logistic regression analysis showed that individuals with a family history of keloid exhibited a higher odds ratio than those without a family history, highlighting the strong impact of host genetic factor(s) on disease onset.

* Corresponding author. Laboratory of Clinical Genome Sequencing, Graduate School of Frontier Sciences, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan.
E-mail address: kmatsuda@k.u-tokyo.ac.jp (K. Matsuda).
Peer review under responsibility of the Japan Epidemiological Association.

http://dx.doi.org/10.1016/j.je.2016.12.003
0917-5040/© 2017 The Authors. Publishing services by Elsevier B.V. on behalf of The Japan Epidemiological Association. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Conclusions: Cross-sectional analysis of the clinical information of participants at enrollment revealed characteristics of the present cohort. Analysis of family history revealed the impact of host genetic factors on each disease. BioBank Japan, by publicly distributing DNA, serum, and clinical information, could be a fundamental infrastructure for the implementation of personalized medicine.

Introduction

BioBank Japan (BBJ) was established with the cooperation of 12 medical institutes, consisting of over 60 hospitals, as a leading project of the Ministry of Education, Culture, Sports, Science and Technology in 2003.1,2 As a disease-oriented biobank, BBJ collected DNA and serum samples from approximately 200,000 patients with 47 diseases. BBJ annually updates clinical information, which is another essential element of biobanks.3 The clinical information associated with the biospecimens was utilized in previous studies to select or stratify the participant group. Samples and their clinical information were used for over 200 studies.4 However, so far, a comprehensive analysis of the clinical information of the BBJ cohorts has not been conducted. Here, we analyzed clinical

Table 1
Baseline characteristics of participants with 47 diseases in the present cohort.

| 47 Diseases                      | Number of Subjects | Mean (SD) age at registration (y) | % of male subjects | % of male patients (Patient survey) |
|----------------------------------|--------------------|-----------------------------------|--------------------|------------------------------------|
| Whole cohort                     | 199,982            | 62.66 14.66 61.55 16.02           | 53.05              | N/A                                |
| Lung cancer                      | 3779               | 67.64 9.54 66.07 9.81             | 64.25              | 50.51                              |
| Esophageal cancer                | 1291               | 65.66 8.06 65.56 10.44            | 86.29              | 84.00                              |
| Gastric cancer                   | 6322               | 67.01 9.90 65.18 11.77            | 73.39              | 66.27                              |
| Colorectal cancer                | 6759               | 67.10 9.95 66.42 10.86            | 62.76              | 55.54                              |
| Liver cancer                     | 1924               | 67.37 8.41 69.97 8.15             | 75.68              | 58.18                              |
| Pancreatic cancer                | 392                | 66.02 9.80 66.21 11.02            | 64.54              | 50.85                              |
| Gallbladder/cholangiocarcinoma    | 392                | 67.71 9.22 68.75 9.05             | 62.50              | 51.02                              |
| Prostate cancer                  | 5066               | 72.60 7.46 N/A 100.00             | 100.00             | 100.00                             |
| Breast cancer                    | 6336               | 63.74 11.21 57.67 11.98           | 0.73               | 1.33                               |
| Uterine cervical cancer          | 1218               | N/A 100.00 51.83 13.33            | 0.00               | 0.00                               |
| Uterine corpus cancer            | 1026               | N/A 100.00 58.93 10.65           | 0.00               | 0.00                               |
| Ovarian cancer                   | 888                | N/A 100.00 56.39 11.91           | 0.00               | 0.00                               |
| Hematological cancer             | 1307               | 60.99 13.50 60.26 16.65           | 54.32              | 53.97                              |
| Cerebral infarction              | 16,534             | 68.82 9.90 71.68 10.60           | 62.27              | 44.37                              |
| Cerebral aneurysm                | 2710               | 60.52 11.51 62.84 10.78           | 35.24              | N/A                                |
| Epilepsy                         | 2303               | 46.56 5.41 63.31 11.31           | 52.77              | 54.42                              |
| Bronchial asthma                 | 8700               | 51.89 13.11 57.67 11.98           | 49.32              | 51.51                              |
| Pulmonary tuberculosis           | 863                | 62.14 16.82 62.43 19.34           | 71.38              | 64.10                              |
| Chronic obstructive pulmonary disease | 2774            | 72.33 8.57 72.71 9.82           | 86.81              | 86.28                              |
| Intestinal lung disease/pulmonary fibrosis | 808         | 68.74 14.11 68.11 11.97           | 58.04              | 55.32                              |
| Myocardial infarction            | 13,272             | 65.92 10.37 71.19 9.90           | 80.98              | 64.32                              |
| Unstable angina                  | 4330               | 66.76 9.71 71.26 9.15           | 73.70              | 55.20                              |
| Stable angina                    | 14,807             | 67.86 9.81 71.05 9.71           | 69.39              | 55.20                              |
| Arrhythmia                       | 15,912             | 67.03 11.67 69.27 12.52           | 64.38              | 52.24                              |
| Heart failure                    | 7610               | 66.01 12.63 71.46 12.72           | 61.81              | 38.18                              |
| Peripheral arterial diseases     | 2683               | 70.84 9.02 71.70 9.97           | 78.12              | 61.97                              |
| Chronic hepatitis B              | 1346               | 54.57 13.21 55.62 14.97           | 62.63              | 62.50                              |
| Chronic hepatitis C              | 5819               | 63.37 11.84 64.64 11.92           | 53.70              | 52.92                              |
| Liver cirrhosis                  | 2519               | 62.74 11.50 65.38 14.17           | 62.29              | 49.52                              |
| Nephrotic syndrome               | 1056               | 47.45 22.88 48.23 21.74           | 60.32              | 58.06                              |
| Urolithiasis                     | 6307               | 53.02 13.72 56.90 14.42           | 75.60              | 67.42                              |
| Osteoporosis                     | 6743               | 72.28 12.89 73.77 9.57           | 7.59               | 7.23                               |
| Diabetes mellitus                | 39,697             | 63.31 11.33 65.80 12.00           | 63.23              | 52.74                              |
| Dyslipidemia                     | 43,812             | 62.15 11.97 66.26 10.79           | 50.76              | 33.55                              |
| Graves’ disease                  | 2321               | 49.86 14.23 49.04 15.75           | 27.85              | 22.22                              |
| Rheumatoid arthritis             | 4139               | 64.05 12.12 62.39 12.29           | 20.25              | 18.87                              |
| Hay fever                        | 5658               | 46.39 17.63 44.94 15.84           | 42.93              | 46.74                              |
| Drug eruption                     | 585                | 60.53 16.17 54.82 17.46           | 45.81              | N/A                                |
| Atopic dermatitis                | 2938               | 29.98 14.85 29.74 13.54           | 53.13              | 51.61                              |
| Keloid                           | 809                | 48.53 19.97 43.31 19.60           | 38.94              | N/A                                |
| Uterine fibroid                  | 5904               | N/A 100.00 44.69 9.49           | 0.00               | 0.00                               |
| Endometriosis                    | 1843               | N/A 100.00 38.93 8.22           | 0.00               | 0.00                               |
| Fehlere seizure                  | 333                | 4.16 3.57 4.35 5.09           | 60.96              | N/A                                |
| Glioma                           | 4755               | 66.87 12.43 70.03 10.95           | 46.79              | 41.98                              |
| Cataract                         | 20,002             | 70.43 13.31 72.91 9.52           | 44.81              | 36.83                              |
| Periodontitis                    | 3898               | 58.20 15.92 56.59 16.00           | 43.69              | 41.03                              |
| Amyotrophic lateral sclerosis    | 782                | 60.86 10.21 61.03 10.76           | 64.32              | N/A                                |
information including age, sex, body mass index (BMI), hypertension, smoking, and drinking status across 47 diseases, and compared the results with the Japanese database. In addition, we assessed the association between target diseases and positive family history.

Materials and methods

Study design

In the present cohort, we focused on 47 common diseases (Table 1). Patients diagnosed with any one of the 47 diseases were recruited from 66 hospitals affiliated with 12 medical institutes between the fiscal year of 2003 and 2007. The detailed protocol of the recruitment process has been described elsewhere. Written informed consent was obtained from all participants. The study protocol was reviewed and approved by the Ethics Committees of all participating institutions, including the Institute of Medical Science, the University of Tokyo, and the Center for Integrative Medical Sciences, RIKEN.

We included patients who had been diagnosed with the diseases by physicians at the cooperating hospitals (eTable 1). As this project registered not only patients with newly developed diseases but also patients who were diagnosed and treated before starting the project, some participants were enrolled several years after disease onset or diagnosis. We excluded patients who had...
received a bone marrow transplant and those who were not of East Asian descent.

Clinical information

Clinical information including common clinical variables, disease-specific variables, prescriptions, and drug side-effect information, was collected from each participant. The detailed methods of the collection of clinical information has been described elsewhere. The clinical database was updated every year until 2012. After a thorough review and data-cleansing of clinical variables, clinical information of 199,982 participants with 47 diseases at enrollment was established on March 31 2015 and used in the current study.

Japanese database

The Ministry of Health, Labour and Welfare in Japan conducts a Patient Survey every three years and a National Health and Nutrition Survey every year. We obtained the results of the Patient Survey of 2005 and those of the National Health and Nutrition Survey of 2006. Table 65 in the Patient Survey was used to

Table 3
Baseline alcohol intake status of participants with 47 diseases in the present cohort.

| Diseases                        | Alcohol intake |                                        |                                        |
|---------------------------------|----------------|-----------------------------------------|---------------------------------------|
|                                 | Male subjects  |                                        | Female subjects                       |
|                                 | Never drinker  | Ex-drinker                              | Current drinker                       | Drinker with unknown status |
| Whole cohort                    | 30.32          | 13.35                                   | 52.24                                 | 4.09                       |
| Lung cancer                     | 26.73          | 15.85                                   | 54.21                                 | 3.21                       |
| Esophageal cancer               | 8.29           | 25.59                                   | 61.38                                 | 4.74                       |
| Gastric cancer                  | 25.19          | 18.96                                   | 51.49                                 | 4.36                       |
| Colorectal cancer               | 23.53          | 15.88                                   | 56.15                                 | 4.45                       |
| Liver cancer                    | 23.90          | 34.83                                   | 33.01                                 | 8.27                       |
| Pancreatic cancer               | 25.70          | 29.72                                   | 42.97                                 | 1.61                       |
| Gallbladder/cholangiocarcinoma   | 30.58          | 27.27                                   | 38.43                                 | 3.72                       |
| Prostate cancer                 | 29.33          | 13.05                                   | 51.47                                 | 6.16                       |
| Breast cancer                   | 28.89          | 11.11                                   | 60.00                                 | 0.00                       |
| Japanese database               |                |                                        |                                        |

Table 65 in the Patient Survey was used to compare the distributions of male (A) and female (B) participants with a smoking history in the BBJ cohort and in the National Health and Nutrition Survey (Japan, 2006) were compared. Age-adjustment was performed according to the age distribution of the National Health and Nutrition Survey (Japan, 2006).
estimate Japanese patient numbers, stratified by sex and age for each disease. Distributions of BMI categories, hypertension prevalence, smoking history, and alcohol intake history in the general Japanese population were calculated from Tables 23, 49-2, 97, and 91 of the National Health and Nutrition Survey, respectively.

Analysis of clinical information

The distributions of BMI, hypertension prevalence, smoking history, and alcohol intake history in the BBJ cohort were adjusted for sex and age group for each table in the national public survey when we compared the distributions among the 47 diseases and Japanese database. BMI category and hypertension were defined according to World Health Organization (WHO) criteria as follows: BMI < 18.5 was defined as underweight, 18.5 ≤ BMI < 25 as normal, 25 ≤ BMI < 30 as overweight, and BMI ≥ 30 as obese; hypertension was defined as systolic blood pressure ≥ 140 mmHg and diastolic blood pressure ≥ 90 mmHg or when participants were prescribed antihypertensive drugs. Multivariate logistic regression analyses were performed to assess the association between each target disease and positive family history associated with the target disease, adjusted for sex and age. SAS 9.4 software was used for the data analysis. A p-value of < 0.05 was considered statistically significant.

Table 4
Baseline BMI and hypertension of participants with 47 diseases in the present cohort.

| 47 Diseases | Male subjects | Female subjects |
|-------------|---------------|-----------------|
|              | Mean (SD)     | Mean (SD)       | Male subjects | Female subjects |
| Whole cohort | 23.51 3.47    | 22.94 3.89      | 51.52 41.11  |
| Lung cancer  | 22.29 3.05    | 22.05 3.37      | 36.74 33.83  |
| Esophageal cancer | 20.53 2.96   | 19.77 3.25      | 27.29 24.86  |
| Gastric cancer | 21.25 3.04    | 20.34 3.26      | 30.91 24.26  |
| Colorectal cancer | 22.66 3.17    | 22.00 3.51      | 38.00 30.31  |
| Liver cancer  | 22.68 3.27    | 22.82 3.96      | 44.64 45.51  |
| Pancreatic cancer | 20.44 3.19    | 19.90 3.03      | 30.83 29.50  |
| Gallbladder/cholangiocarcinoma | 21.46 3.29    | 22.20 3.89      | 33.47 31.29  |
| Prostate cancer | 23.28 2.86    | N/A             | 38.00 43.11  |
| Breast cancer | 23.87 3.75    | 22.74 3.60      | 52.17 22.82  |
| Cervical cancer | N/A           | 21.93 3.29      | N/A 25.89   |
| Uterine cancer | 23.74 4.37    | N/A             | N/A 25.89   |
| Ovarian cancer | N/A           | 22.04 3.38      | N/A 19.21   |
| Hematopoietic tumor | 23.11 3.23    | 21.87 3.33      | 30.94 26.71  |
| Cerebral infarction | 23.53 3.19    | 23.39 3.86      | 67.12 65.34  |
| Cerebral aneurysm | 23.86 3.36    | 23.11 3.64      | 65.45 59.52  |
| Epilepsy       | 23.47 3.84    | 22.70 4.19      | 36.26 25.64  |
| Bronchial asthma | 23.79 3.71    | 23.78 4.55      | 41.19 35.14  |
| Pulmonary tuberculosis | 20.82 3.28    | 20.26 3.26      | 32.31 37.80  |
| Chronic obstructive pulmonary disease | 21.30 3.37    | 20.33 4.08      | 46.05 44.66  |
| Intestinal lung disease/pulmonary fibrosis | 23.02 3.21    | 22.62 3.75      | 42.37 40.65  |
| Myocardial infarction | 24.04 3.23    | 23.40 3.74      | 73.11 77.50  |
| Unstable angina | 24.02 3.21    | 23.74 3.74      | 74.33 74.80  |
| Stable angina  | 23.85 3.16    | 23.63 3.66      | 77.17 77.45  |
| Arrhythmia     | 23.53 3.30    | 22.90 3.80      | 66.75 65.33  |
| Heart failure  | 23.50 3.89    | 22.64 4.31      | 78.70 78.10  |
| Peripheral arterial diseases | 22.52 3.25    | 22.44 3.81      | 70.21 69.74  |
| Chronic hepatitis B | 23.32 3.11    | 22.55 3.51      | 40.67 33.69  |
| Chronic hepatitis C | 22.86 3.15    | 22.54 3.65      | 46.19 40.71  |
| Liver cirrhosis | 22.88 3.51    | 23.03 4.05      | 52.18 49.41  |
| Nephrotic syndrome | 23.00 3.34    | 22.50 3.94      | 62.32 50.68  |
| Urolithiasis   | 24.43 3.39    | 23.59 4.16      | 37.27 35.61  |
| Osteoporosis   | 21.96 3.52    | 22.26 3.62      | 49.22 44.46  |
| Diabetes mellitus | 24.03 3.72    | 24.57 4.41      | 60.32 62.82  |
| Dyslipidemia   | 24.78 3.45    | 24.11 3.88      | 64.47 59.48  |
| Graves' disease | 23.55 3.63    | 22.35 3.61      | 41.03 32.40  |
| Rheumatoid arthritis | 22.49 3.29    | 21.85 3.69      | 40.57 33.56  |
| Hayward fever  | 23.66 3.17    | 22.06 3.51      | 24.70 14.77  |
| Drug eruption  | 23.27 3.39    | 22.62 4.07      | 49.81 30.87  |
| Atopic dermatitis | 23.01 3.51    | 21.48 3.70      | 13.24 5.75   |
| Keloid         | 23.95 3.29    | 22.71 4.11      | 27.60 18.82  |
| Uterine fibroid | N/A           | 22.29 3.51      | N/A 14.09   |
| Endometriosis  | N/A           | 21.43 3.25      | N/A 7.93    |
| Febrile seizure | 28.73 0.00    | 21.26 4.80      | N/A N/A     |
| Glaucoma       | 23.05 3.18    | 22.90 3.66      | 43.87 41.08  |
| Cataract       | 23.08 3.15    | 23.05 3.85      | 49.53 45.68  |
| Periodontitis  | 23.27 3.14    | 22.25 3.37      | 27.33 15.90  |
| Amyotrophic lateral sclerosis | 21.14 1.68    | 28.00 3.14      | N/A N/A     |

Fig. 2. Age-adjusted ratio of participants with alcohol history in each disease. The distributions of male (A) and female (B) participants with a drinking history in the BBJ cohort and in the National Health and Nutrition Survey (Japan, 2006) were compared. Age-adjustment was performed according to the age distribution of the National Health and Nutrition Survey (Japan, 2006).
Results

Basic characteristics (Age and sex)

We characterized the BioBank Japan cohort at enrollment by analyzing common clinical variables of age and sex across the target diseases. Mean age at enrollment, across the entire cohort or for each disease, was comparable between both sexes, but varied among the diseases (Table 1). The highest mean age was observed in men with prostate cancer and in women with osteoporosis (72.60 and 73.77 years, respectively), while the youngest mean age was observed in men and women with febrile seizures (4.16 and 4.35 years, respectively), reflecting the typical age of onset of each disease. A greater number of men were registered in the BBJ cohort compared to women (53.05% vs. 46.95%), while sex ratios varied according to the diseases (Table 1).

To highlight sex and age characteristics of the BBJ cohort, we further compared the sex and age distribution for each disease with the Patient Survey. We included participants with 42 out of the 47 diseases for the comparison, as we obtained the relevant clinical data from the Patient Survey (eTable 2). Almost all diseases displayed equivalent age distributions, while lower proportions of participants <20 years of age were observed in three diseases (bronchial asthma, atopic dermatitis, and hay fever), which are likely to occur in younger populations (eFigs. 1.1 and eTable 3). The proportion of male participants with dyslipidemia was considerably higher in the BBJ cohort (50.76%) than in the Patient Survey (33.55%), although both age distributions appeared equivalent. The low proportion of female patients with heart failure aged ≥80 years resulted in a lower proportion of female participants in the BBJ cohort. We also observed a low proportion of elderly female participants with cerebral infarction, chronic obstructive pulmonary disease (COPD), peripheral arterial diseases (PAD), unstable angina, stable angina, and myocardial infarction in the BBJ cohort. Varyed distributions between the BBJ cohort and Patient Survey were observed in pulmonary tuberculosis and nephrotic syndrome.

Basic characteristics (Lifestyle and physical status)

We also evaluated lifestyle including smoking and alcohol intake history, and physical status including BMI and blood pressure, at enrollment in the BBJ cohort. We included participants >20 years of age in this analysis because the frequency of smoking, alcohol intake, and hypertension among individuals under 20 years of age is quite low, and the criteria for underweight or obesity according to BMI in children and teenagers are different from those applied to adults. Furthermore, we compared the BBJ cohort and the National Health and Nutrition Survey 2006 for physical and lifestyle, after adjusting for sex and age, because sex- and age-distribution varied among diseases.

Smoking history at enrollment (including subjects both with and without information on current smoking status) was positive in 74.98% of male subjects and 21.24% of female subjects in the BBJ cohort, while current smokers accounted for 27.78% of male subjects and 10.45% of female subjects (Table 2). The highest frequency of positive smoking history in both sexes was observed in COPD, followed by PAD in male subjects, and esophageal cancer in female subjects (Table 2). The highest proportion of ex-smokers for both sexes was observed in participants with lung cancer, esophageal cancer and COPD (71.45%, 64.88% and 64.68% in male subjects, and 21.75%, 30.86% and 39.44% in female subjects, respectively), while the highest proportion of current smokers for both sexes was observed in participants with Graves’ disease (49.84% in male subjects and 24.92% in female subjects) (Table 2). We then compared age-adjusted smoking history among the 47 diseases. The frequency of smokers was highest among participants with COPD, esophageal cancer, interstitial lung diseases/pulmonary fibrosis, pancreatic cancer, and cardiovascular diseases, in which smoking was shown to be a critical risk factor (Fig. 1 and eTable 4).

A positive alcohol history at enrollment (including those with and without current drinking status) was found in 69.68% of male subjects and 28.20% of female subjects (Table 3). The proportion of current drinkers in the whole cohort was much higher than that of ex-drinkers in both sexes: 52.24% and 13.35% of male subjects and 21.70% and 3.99% of female subjects were current and ex-drinkers, respectively. Among the 47 diseases, the proportion of ex-drinkers was relatively high among participants with liver cirrhosis (34.05% in male subjects and 10.80% in female subjects), liver cancer (34.83% and 10.94%), pulmonary tuberculosis (33.17% and 11.20%), esophageal cancer (25.39% and 16%), and pancreatic cancer (29.72% and 10.14%) (Table 3). Age-adjusted alcohol intake history showed that the frequency of drinkers in esophageal cancer was remarkably higher than that in other diseases for male and female subjects (Fig. 2 and eTable 5). To highlight the smoking and drinking status in the BBJ cohort, the frequency of smokers or drinkers, stratified by sex and age group, was compared between the BBJ and the National Health and Nutrition Survey. The BBJ cohort had a higher frequency of smokers among female subjects across all age groups and among elderly male subjects, particularly among those >60 years of age; the frequency of drinkers was almost equivalent between the BBJ and the National Health and Nutrition Survey for both sexes and across all age groups (Fig. 2A and B and eTables 6 and 7).

Mean BMI at enrollment in the BBJ cohort was 23.51 in male subjects and 22.94 in female subjects. Analysis of BMI in each disease revealed that underweight participants (BMI <18.5) had an increased association of various cancers, while overweight or obese participants (BMI ≥25) had an increased association of metabolic and cardiovascular diseases (Table 4, Fig. 3 and eTable 8). When comparing the National Health and Nutrition Survey and the BBJ, there was a greater proportion of participants with overweight or obesity in the BBJ, among male and female subjects and across all age-groups; conversely, similar distribution patterns were found when comparing the BBJ cohort and the Survey, by sex and age-group (eFig. 2C and eTables 6 and 7). In contrast, in the BBJ cohort, there were fewer underweight participants in their twenties (for both sexes) but more overweight participants >60 years (among male subjects) and >50 years (among female subjects) (eFig. 2D and eTables 6 and 7).

Nearly half of the participants of the BBJ cohort had hypertension (51.52% of male subjects and 41.11% of female subjects, Table 4) at enrollment. The frequency of hypertension in cardiovascular diseases, particularly in coronary diseases, was higher than that in other diseases, while the frequency of hypertension among cancer participants tended to be low (Table 4, Fig. 4 and eTable 9). The frequency of hypertension increased with age, similarly to the increase observed in the Survey. However, the frequency of hypertension among subjects <50 years of age was higher and subjects >60 years of age was lower in the BBJ cohort than in the Survey (eFig. 2E and eTables 6 and 7).
Fig. 4. Age-adjusted ratio of participants with hypertension in each disease. The distributions of male (A) and female (B) participants with hypertension in the BBJ cohort and in the National Health and Nutrition Survey (Japan, 2006) were compared. Age-adjustment was performed according to the age distribution of the National Health and Nutrition Survey (Japan, 2006). Participants with a systolic blood pressure ≥135-mmHg, a diastolic blood pressure ≥90-mmHg, or participants prescribed antihypertensive medication, were diagnosed with hypertension.

Fig. 5. Sex- and age-adjusted odds ratios in family history, related with the 47 diseases. Dots represent odds ratios and bars represent 95% CIs by logistic regression analysis. The list of family histories, associated with the 47 diseases, is set out in eTable 2.
Family history

Finally, we performed multivariate logistic-regression analysis using age and sex status as covariates to assess the association between positive family history and disease risk. We were able to obtain the questionnaire-based information regarding family history of 45 diseases out of the 47 diseases (eTable 10). For all the diseases, except for PAD, there was a significant association with a positive family history, with an odds ratio of greater than 1.7 (Fig. 5 and eTable 11). Notably, the odds ratios for keloid, chronic hepatitis B, and Grave’s disease were relatively high (149.417, 53.474, and 23.751, respectively) indicating the strong impact of genetic and familial factors on disease onset.

Discussion

We analyzed common clinical variables at enrollment, across the whole BBJ cohort, as well as for each target disease, and we compared these results with those of the Japanese database to highlight the characteristics of the BBJ cohort. Statistical analyses were not conducted in this study, as the large-scale cohort sample in the BBJ would yield relatively low p-values, even when absolute differences were very small. The distribution of age, life style, and physical status, showed that the characteristics of each disease group could generally be explained.

It is an established fact that smoking and/or alcohol intake are risk factors for various diseases including cancer, cardiovascular disease, hepatic disease, and respiratory disease. In fact, these diseases showed a higher frequency of participants with a positive smoking or drinking history at enrollment in the BBJ cohort (Figs. 1 and 2 and eTables 4 and 5). Although we cannot estimate the odds ratios of smoking and drinking status due to the lack of control data in the present cohort, age-adjusted distributions of the smoking and drinking histories of participants suggest that these lifestyle factors have a significant impact on disease onset.

Analysis of BMI at enrollment indicated that lower BMI was more prevalent among participants with malignant tumors, while higher BMI was common among participants with metabolic and cardiovascular disease (Fig. 3 and eTable 8). Obesity could be a risk factor for dyslipidemia, type 2 diabetes, coronary disease, while cancer can induce weight loss. Therefore, we need to be cautious in the interpretation of the association between diseases and lifestyle or physical factors.

To highlight the characteristics of the BBJ cohort, we compared the age and sex distributions of the BBJ cohort with those of the Patient Survey for each disease, and the distributions of smoking and drinking history, BMI and hypertension in the BBJ cohort with those of the National Health and Nutrition Survey. It is difficult to discuss the discrepancy or consistency between the BBJ cohort and the Japanese database, because backgrounds of the subjects and methods to determine the numbers of patients or the distributions of life style and physical status were different. However, the comparisons between the BBJ cohort and the Japanese database gave us better insight about the characteristics of the BBJ cohort, contributing to utmost utilization of the biobank samples.

As one of our main aims was to identify genetic factors causing susceptibility to diseases, we analyzed the association between positive family history and disease onset to evaluate the impact of host genetic factors. It has been reported that a positive family history is an important risk factors for many common chronic diseases and keloid, chronic hepatitis B, and Graves’ disease showed the highest odds ratios for a positive family history (Fig. 5). While it is important to consider the possibility that perinatal transmission, a major route of hepatitis B virus transmission, resulted in the high odds ratio observed in chronic hepatitis B, several genome-wide association studies (GWAS), which identified some single nucleotide polymorphism loci significantly associated with these diseases in Japan, support the finding that genetic factors are associated with these diseases. However, the odds ratios, calculated in the previous genomic studies, were not as high as in the present analysis, suggesting the possibility that further genomic analysis could identify novel genomic loci. In addition, the fact that common clinical variables were consistently identified across the 47 diseases enabled us to evaluate and compare the risk significance of the positive family history on the diseases and to perform further genomic or other ‘omics’ analyses based on these results.

This study has some limitations. We could not eliminate the possibility of reporting bias, causing significantly higher odds ratio of positive family history in almost all target diseases, as the information on family history was mainly based on participants’ interviews, although this was completed by certified medical coordinators. Another limitation of this analysis is that the reference population for each logistic analysis was not the disease-free general population but the participants with the other diseases in the cohort. Therefore, again, we need to take into account selection bias.

In conclusion, we have established a large biobank cohort, consisting of approximately 200,000 patients with 47 diseases. Analysis of the clinical dataset and comparisons between the present cohort and the Japanese database largely revealed consistent trends in common clinical variables, particularly among participants aged ≥40 years, suggesting that the sampling is representative for the general patient population in Japan. Further analysis, combined with various high-throughput ‘omics’ technologies, using their DNA and serum samples, will aid us to identify novel genomic variants or biomarkers associated with disease progression or drug efficacy, contributing to the implementation of personalized medicine.

Conflicts of interest

None declared.

Acknowledgements

We express our gratitude to all the participants in the BioBank Japan Project. We thank all the medical coordinators of the cooperating hospitals for collecting samples and clinical information, as well as Yasushi Yamashita and staff members of the BioBank Japan Project for administrative support. We also thank Dr. Kumao Toyoshima for his overall supervision of the BioBank Japan Project. This study was supported by funding from the Tailor-Made Medical Treatment with the BB Project from Japan Agency for Medical Research and Development, AMED (since April 2015), and the Ministry of Education, Culture, Sports, Science, and Technology (from April 2003 to March 2015).

Appendix A. Author list for the BioBank Japan Cooperative Hospital Group

Members of medical institutions cooperating on the BioBank Japan Project who coauthored this paper include Masaki Shiono, Kazuo Misumi, Reiji Kaieda, Hiromasa Harada (Tokushukai Hospital); Shiro Minami, Mitsuru Emi, Naoya Emoto (Nippon Medical School), Hajime Arai, Ken Yamaji, Yoshimune Hiratsuka (Juntendo University), Satoshi Asai, Mitsuhiro Moriyama, Yasuo Takahashi (Nihon University), Tomoaki Fujioka, Wataru Obara (Iwate Medical University), Seiji Mori, Hideki Ito (Tokyo Metropolitan Institute of
References

1. Nakamura Y. The BioBank Japan project. Clin Adv Hematol Oncol. 2007;5:696–697.
2. Nagai A, Hirata M, Kamatani Y, et al. Overview of the BioBank Japan project: study design and profile. J Epidemiol. 2017;27:52–58.
3. Aissabet M, Zlatevsk K. Biobanks: transnational, European and global networks. Brief Funct Genomic Proteomic. 2007;6:193–201.
4. BioBank Japan. Publications from BioBank Japan. https://biobankjp.org/work/public.html; Updated 30.06.16. Accessed 25 July 2016.
5. Ministry of Health, Labour and Welfare, Japan. Patient Survey; 2005. http://www.e-stat.go.jp/SG1/estat/List.do?lid=000001047095 (in Japanese) Accessed 6 June 2016.
6. Ministry of Health, Labour and Welfare, Japan. National Health and Nutrition Survey; 2006. http://www.mhlw.go.jp/henyou/kensa/eryou08/01.html (in Japanese) Accessed 6 June 2016.
7. World Health Organization. BMI Classification; 2016. http://apps.who.int/bmi/index.jsp?introPage=intro_3.html Accessed 6 June 2016.
8. Kuczynski RJ, Ogden CL, Guo SS, et al. 2000 CDC growth charts for the United States: methods and development. Vital Health Stat. 2002;11:1–190.
9. Centers for Disease Control and Prevention. Health Effects of Cigarette Smoking; 2016. http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects_cig_smoking/. Accessed 16 June 2016.
10. Centers for Disease Control and Prevention. Fact Sheets — Alcohol Use and Your Health; 2016. http://www.cdc.gov/alcohol/fact-sheets/alcohol-use.htm. Accessed 16 June 2016.
11. Centers for Disease Control and Prevention. Adult Obesity Causes & Consequences; 2016. http://www.cdc.gov/obesity/adult/causes.html. Accessed 16 June 2016.
12. Reid GT, Walter FM, Brisbane JM, Emery JD. Family history questionnaires designed for clinical use: a systematic review. Public Health Genomics. 2009;12:73–83.
13. Brandi ML, Gennari L, Cerinic MM, et al. Genetic markers of osteoarticular disorders: facts and hopes. Arthritis Res. 2001;3:270–280.
14. Cole Johnson C, Ownby DR, Havstad SL, Peterson EL. Family history, dust mite exposure in early childhood, and risk for pediatric atopy and asthma. J Allergy Clin Immunol. 2004;114:105–110.
15. Collaborative Group on Hormonal Factors in Breast C. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 females with breast cancer and 101,986 females without the disease. Lancet. 2001;358:1389–1399.
16. Harrison TA, Hindorff LA, Kim H, et al. Family history of diabetes as a potential public health tool. Am J Prev Med. 2003;24:152–159.
17. Hawe E, Talmud PJ, Miller GJ, Humphries SE, Second Northwick Park Heart Study. Family history is a coronary heart disease risk factor in the Second Northwick Park Heart Study. Ann Hum Genet. 2003;67:97–106.
18. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. Am J Gastroenterol. 2001;96:2992–3003.
19. Pharoah PD, Ponder BA. The genetics of ovarian cancer. Best Pract Res Clin Obstet Gynaecol. 2002;16:449–468.
20. World Health Organization, Geneva. Guidelines for the Prevention, Care, and Treatment of Persons with Chronic Hepatitis B Infection; 2015. http://www.ncbi.nlm.nih.gov/books/NBK305553/. Accessed 16 June 2016.
21. Ban Y, Taniyama M, Ban Y. Vitamin D receptor gene polymorphism is associated with Graves’ disease in the Japanese population. J Clin Endocrinol Metab. 2000;85:4639–4643.
22. Ban Y, Tozaki T, Taniyama M. The replication of the association of the rs935610 within 6p27 with Graves’ disease. Autoimmunity. 2013;46:395–398.
23. Ban Y, Tozaki T, Taniyama M, Tomita M, Ban Y. Association of a C/T single-nucleotide polymorphism in the 5′ untranslated region of the CD40 gene with Graves’ disease in Japanese. Thyroid. 2006;16:443–436.
24. Furugaki K, Shirasawa S, Ishikawa N, et al. Association of the T-cell regulatory gene CTLA4 with Graves’ disease and autoimmune thyroid disease in the Japanese. J Hum Genet. 2004;49:166–168.
25. Hiratani H, Bowden DW, Ikegami S, et al. Multiple SNPs in intron 7 of thyrotopin receptor are associated with Graves’ disease. J Clin Endocrinol Metab. 2005;90:2898–2903.
26. Kamatani Y, Wattanapokayakit S, Ochi H, et al. A genome-wide association study identifies variants in the HLA-DP locus associated with chronic hepatitis B in Asians. Nat Genet. 2009;41:591–595.
27. Komatsu H, Murakami J, Inui A, Tsunoda T, Sogo T, Fujisawa T. Association between single-nucleotide polymorphisms and early spontaneous hepatitis B virus e antigen seroconversion in children. BMC Res Notes. 2014;7:795.
28. Kumar V, Yi Lo PH, Sawai H, et al. Soluble MICA and a MICA variation as possible prognostic biomarkers for HBV-induced hepatocellular carcinoma. PLoS One. 2012;7:e44743.
29. Mukai T, Hiromatsu Y, Fukutani T, et al. A C/T polymorphism in the 5′ untranslated region of the CD40 gene is associated with later onset of Graves’ disease in Japanese. Endocr J. 2005;52:471–477.
30. Nakashima M, Chung S, Takahashi A, et al. A genome-wide association study identifies four susceptibility loci for keloid in the Japanese population. Nat Genet. 2010;42:768–771.
31. Nishida N, Ohashi J, Khor SS, et al. Understanding of HLA-conferred susceptibility to chronic hepatitis B infection requires HLA genotyping-based association analysis. Sci Rep. 2016;6:24767.
32. Okada Y, Momozawa Y, Ashikawa K, et al. Construction of a population-specific HLA imputation reference panel and its application to Graves’ disease risk in Japanese. Nat Genet. 2015;47:798–802.
33. Trivedi G, Jager KJ, Dekker PW, Zoccali C. Selection bias and information bias in clinical research. Nephron Clin Pract. 2011;115:c4–c9S.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.je.2016.12.003.