High-density association study and nomination of susceptibility genes for hypertension in the Japanese National Project

Norihiro Kato1,*, Toshiyuki Miyata3, Yasuharu Tabara4, Tomohiro Katsu5, Kazuyuki Yanai1, Hironori Hanada3, Kei Kamide9, Jun Nakura5, Katsuhiko Kohara5, Fumihiko Takeuchi2, Hiroyuki Mano7, Michio Yamasaki8,9, Akinori Kimura8,9, Yoshikuni Kita10, Hirotaki Ueshima10, Tomohiro Nakayama11, Masayoshi Soma12, Akira Hata13, Akihiro Fujikawa14, Yuhei Kawano3, Kazuwa Nakas5, Akihiro Sekine16, Teruhiko Yoshida17, Yusuke Nakamura16,18, Takao Saruta19, Toshio Ogihara6, Sumio Sugano20, Tetsuro Miki5 and Hitonobu Tomotake3

1Department of Gene Diagnostics and Therapeutics and 2Department of Medical Ecology and Informatics, Research Institute, International Medical Center of Japan, Tokyo, Japan, 3National Cardiovascular Center, Suita, Osaka, Japan, 4Department of Basic Medical Research and Education and 5Department of Geriatric Medicine, Ehime University Graduate School of Medicine, Toon, Ehime, Japan, 6Department of Geriatric Medicine, Osaka University Graduate School of Medicine, Suita, Osaka, Japan, 7Division of Functional Genomics, Jichi Medical University, Shimotsuke-shi, Tochigi, Japan, 8Department of Molecular Pathogenesis, Medical Research Institute and 9Laboratory of Genome Diversity, School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan, 10Department of Health Science, Shiga University of Medical Science, Otsu, Shiga, Japan, 11Division of Molecular Diagnostics, Advanced Medical Research Center and 12Division of Nephrology and Endocrinology, Department of Internal Medicine, Nihon University School of Medicine, Tokyo, Japan, 13Department of Public Health, Graduate School of Medicine, Chiba University, Chiba, Japan, 14Amagasaki Health Medical Foundation, Amagasaki, Hyogo, Japan, 15Department of Medicine and Clinical Science, Kyoto University Graduate School of Medicine, Kyoto, Japan, 16SNP Research Center, Institute of Physical and Chemical Research, Tokyo, Japan, 17Genetics Division, National Cancer Center Research Institute, Tokyo, Japan, 18Human Genome Center, Institute of Medical Science, University of Tokyo, Tokyo, Japan, 19Department of Internal Medicine, School of Medicine, Keio University, Tokyo, Japan and 20Division of Bioscience, Graduate School of Frontier Sciences, University of Tokyo, Tokyo, Japan

Received August 23, 2007; Revised and Accepted November 12, 2007

Essential hypertension is one of the most common, complex diseases, of which considerable efforts have been made to unravel the pathophysiological mechanisms. Over the last decade, multiple genome-wide linkage analyses have been conducted using 300–900 microsatellite markers but no single study has yielded definitive evidence for ‘principal’ hypertension susceptibility gene(s). Here, we performed a three-tiered, high-density association study of hypertension, which has been recently made possible. For tier 1, we genotyped 80 795 SNPs distributed throughout the genome in 188 male hypertensive subjects and two general population control groups (752 subjects per group). For tier 2 (752 hypertensive and 752 normotensive subjects), we genotyped a panel of 2676 SNPs selected (odds ratio ≥1.4 and P ≤0.015 in tier 1) and identified 75 SNPs that showed similar tendency of association in tier 1 and tier 2 samples (P ≤0.05 for allele frequency and P ≤0.01 for genotype distribution tests). For tier 3 (619 hypertensive and 1406 normotensive subjects), we genotyped the 75 SNPs and found nine SNPs from seven genomic loci to be associated with hypertension.

*To whom correspondence should be addressed at: Department of Gene Diagnostics and Therapeutics, Research Institute International Medical Center of Japan, 1-21-1 Toyama, Shinjuku-ku, Tokyo 162-8655, Japan. Tel: +81 332027181; Fax: +81 332027364; Email: nokato@ri.imcj.go.jp

© The Author 2007. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
INTRODUCTION

Essential hypertension (MIM 145500) is a multifactorial trait, in which interactions among genetic, environmental and demographic factors are involved. Substantial contribution of genetic factors to the overall disease etiology has been documented by a number of epidemiological studies. For example, family studies controlling for a common environment indicate that blood pressure heritability is in the range of 15–35% (1–3). Accordingly, considerable efforts have been made in the study of molecular genetics of hypertension, but the inherently complex nature has hampered progress in the elucidation of the genes involved (4). Over the last decade, multiple genome-wide linkage analyses have been conducted by using microsatellite markers to localize genes influencing hypertension status and/or blood pressure levels in a number of populations derived from various ethnic groups. Although no single study has so far yielded definitive evidence for ‘principal’ hypertension susceptibility gene(s), some of these studies provide consistency of linkage results in a few chromosomal regions (5–7). It is therefore assumed that multiple genes contribute to the etiology of hypertension independently or synergistically, with each gene exerting small effects under a certain environmental condition.

In parallel with family-based linkage analyses across the entire genome, population-based association studies have been performed, particularly focusing on individual candidate genes to search for genetic influences on hypertension. Association studies for mapping disease-related genes have recently gained popularity over traditional family-based linkage analyses mainly because of their far greater statistical power to detect the presence of genes with relatively ‘minor’ effects (8,9). Some researchers criticize the liability to false-positive or non-replicable claims. Nevertheless, population-based association studies have become an alternative and complementary approach to family-based linkage analyses in practice.

Given the limitation of statistical power that can be achieved by family-based linkage analyses with sample size practically collectable, population-based association studies are now underway in a genome-wide scale for a number of multifactorial diseases (10). Here, we performed a high-density association study of hypertension with a three-tiered genotyping approach in the Japanese population (Fig. 1).

RESULTS

Multi-tiered case–control study

We performed a large-scale case–control association study of hypertension using SNP markers selected from the Japanese SNP (JSNP) database (11,12). These SNP markers were distributed throughout the genome (Table 1). Only male hypertensive individuals were tested in tier 1, and a total of 80,795 SNPs distributed on 22 autosomes were used for the association study. Details of the high-throughput genotyping were same as previously described (13,14), and technical evaluation of our genotyping assay (e.g. overall success rate and accuracy of the genotyping assay) is shown in the supplementary material (Supplementary Explanation). JSNP had been developed as a database for the SNP discovery project with particular focus on common gene variations in the Japanese population. Although SNP marker resources used in the current study showed a certain degree of diversity in terms of the number of typed SNPs per gene locus, this partially reflected the variable size of re-sequenced fragments depending on the individual gene structure (12).

The gene-centered genome-wide exploratory test in tier 1 identified 2676 SNPs with odds ratio (OR) ≥1.4 and \( P \leq 0.015 \) in at least one test comparing allele frequency and/or genotype distribution (dominant or recessive models) between 188 hypertensive patients and 752 population control subjects in either of two panels (see Materials and Methods). In this exploratory test, the SNPs showing inverted tendency of OR between two pairs of case–control comparisons and significant deviations from Hardy–Weinberg equilibrium (HWE) in any panel \( (P \leq 0.01) \) were excluded. Subsequently, we performed a screening of these 2676 SNPs with 752 hypertensive patients and 752 normotensive controls in tier 2, which constituted the first ‘case versus unaffected control’ study panel, i.e. comparison between 940 cases and 752 controls, together with the 188 cases in tier 1. On the basis of relatively stringent criteria, we identified 75 SNPs that showed \( P \)-values of \( \leq 0.01 \) for genotype distribution and \( P \)-values of \( \leq 0.05 \) for allele frequency in the \( \chi^2 \)-test statistic. To further examine the association signals, we performed a replication study of these 75 SNPs with another panel of 619 hypertensive subjects and 1406 normotensive controls in tier 3. Cases and unaffected controls collected in tiers 2 and 3 were enrolled according to the identical criteria and their baseline characteristics are shown in Table 2. There were some trait differences in cases between tiers 2 and 3, such as blood pressure measurements and percentages of the subjects taking anti-hypertensive medication. This could be largely attributed to differences in sample enrollment settings between tiers 2 and 3; that is, cases in tier 3 were enrolled from either the annual medical checkup of a medical institution or the clinic practices of general practitioners, whereas a major part of cases in tier 2 were from the clinic practices of university hospitals. Among the 75 SNPs showing \( P \)-values between 0.05 and \( 4.4 \times 10^{-5} \) in the first ‘case versus unaffected control’ study, only nine SNPs showed borderline association (at the level of \( P \leq 0.05 \)) in
the second ‘case versus unaffected control’ study (Fig. 2). Of these, we found six SNPs that showed P-values of ≤0.05 for both genotype distribution and allele frequency in the x²-test statistic (Table 3). rs3755351 and rs3771426 were located within the assumed intron 1 of ADD2, rs3787240 and rs3787241 were located in the same intron of EYA2, and the remaining two SNPs—rs3794260 and rs1805762—each located in KIAA0789 and M6PR. To adjust for three covariates—age, gender and body mass index (BMI), we also performed logistic regression analysis for the significant SNPs (Supplementary Material, Table S1). With consideration of genetic model consistency, an SNP (rs3755351) showed the strongest association in the identical model (an additive model by logistic regression analysis) among three tiers. Further details of the association results are described in the Discussion.

### SNP discovery and further test of association in three selected genes

Because a group of SNPs from three genes, ADD2, KIAA0789 and M6PR, were particularly noted for their significant association with hypertension (Table 3), we searched for potentially functional SNPs by re-sequencing the 5'– and 3'–untranslated regions, all exons and exon–intron borders of the individual loci, on the basis of the gene structure deposited in the human genome database (http://www.ncbi.nlm.nih.gov/). We detected a total of 74 SNPs—25 SNPs in ADD2, 40 SNPs in KIAA0789 and 9 SNPs in M6PR—and thereby selected 25 tag SNPs for genotyping 2025 subjects in tier 3 (see Supplementary Material, Table S2A, B and C). Apart from four SNPs which had been already included in the JSNP screening marker set, we found four additional SNPs, two in ADD2 (rs2024453 and rs10084293) and one each in KIAA0789 (rs9739493) and M6PR (rs1805725), to be significantly associated with hypertension (Table 3). Thus, in each gene, we identified at least two SNPs showing modest evidence of association with hypertension (P ≤ 0.05 level in tier 3) but these SNPs did not necessarily belong to the same linkage disequilibrium (LD) block (Fig. 3 and LD group in Supplementary Material, Table S2A, B and C). The analysis of haplotypes inferable from tag SNPs did not show more significant disease association than the analysis of individual SNPs in any of three genes tested (data not shown).

### Consideration of study power and multiple testing

We first estimated a type I error probability for the three-tiered screening to be $6.8 \times 10^{-5}$: 0.036 for tier 1, 0.0009 for tiers 1
and 2 combined and 0.076 for tier 3 screening. Then, we estimated overall sensitivities (which could represent the statistical power) to be 0.10–0.45, 0.04–0.23 and 0.01–0.08 for a disease-associated SNP of OR = 1.4, 1.3 and 1.2, respectively, assuming the disease allele frequency within 0.1–0.9, the disease prevalence of 0.25 and the multiplicative genotype model. Since we had adopted relatively generous criteria for screening association signals, we evaluated the false discovery rate (FDR) to account for multiple testing (15). FDR for the nine SNPs found as significant was 0.69. A multi-staged screening in the current study could be largely categorized into two steps: tiers 1 and 2 (which constitute the first ‘case versus unaffected control’ study) and tier 3 (which constitutes the second ‘case versus unaffected control’ study). We therefore assessed experiment-wise type I errors with particular focus on the last-stage screening in tier 3. By permutation, the chance of observing a $P$-value of 0.0019 (for allele frequency test at rs3755351 in ADD2) in tier 3 was estimated to be 0.13.

**DISCUSSION**

With the recent advent of high-throughput genotyping technologies and high-resolution maps of SNP markers, it is expected that genome-wide association studies allow us to identify systematically the contributions of common genetic variations to human multifactorial diseases (16–18). In this line, our study has attempted to discover common hypertension susceptibility gene variants via a gene-centered genome-wide association design for the first time. Despite the modest genetic impacts assumed for hypertension, e.g. the $\lambda$-values (the relative risk for siblings of the affected probands) have been reported to be approximately 4 (19), we have nominated several susceptibility genes for hypertension (Table 3). Among these genes, findings for ADD2 and KIAA0789 are particularly noteworthy, because the former has been known to be a physiological candidate gene for hypertension and the latter is a novel gene with as-yet unknown physiological function.

Through a multi-tiered screening, nine SNPs derived from seven distinct gene loci have remained to show some evidence of association out of the 80 795 SNPs initially screened. Although the selection criteria were arbitrarily defined in the present study, a small percentage of the SNPs have passed the criteria in transitions from tier 1 to tier 2 (3.3%) and from tier 2 to tier 3 (2.8%). In the ADD2 gene, for example, the minor allele frequency (MAF) of rs3755351 is lower in tier 3 than that in control groups (0.21–0.22) throughout three tiers. A $P$-value of $1.7 \times 10^{-5}$ and an OR of 1.30 (95% CI 1.15–1.46) are attained for allele frequency comparison of rs3755351 when the subjects studied in different tiers are combined and finally categorized into...
Table 2. Clinical characteristics of participants

<table>
<thead>
<tr>
<th>Variables</th>
<th>Case group Tier 2 panel</th>
<th>Tier 3 panel</th>
<th>Control group Tier 2 panel</th>
<th>Tier 3 panel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects (female/male)</td>
<td>752 (353/399)</td>
<td>619 (280/339)</td>
<td>752 (366/386)</td>
<td>1406 (650/756)</td>
</tr>
<tr>
<td>Present age, year</td>
<td>62.4 ± 10.3</td>
<td>54.1 ± 8.4(^a)</td>
<td>62.0 ± 8.7</td>
<td>58.4 ± 6.6</td>
</tr>
<tr>
<td>Age of onset, year</td>
<td>47.3 ± 10.2</td>
<td>43.2 ± 9.9</td>
<td>42.5 ± 9.0</td>
<td>—</td>
</tr>
<tr>
<td>Current BMI, kg/m(^2)</td>
<td>23.9 ± 3.2</td>
<td>25.1 ± 3.6(^a)</td>
<td>22.5 ± 2.8</td>
<td>22.4 ± 2.7</td>
</tr>
<tr>
<td>Smoking(^b)</td>
<td>None, %</td>
<td>61.6</td>
<td>66.0</td>
<td>58.6</td>
</tr>
<tr>
<td>Previous smoker, %</td>
<td>—</td>
<td>17.0</td>
<td>—</td>
<td>10.2</td>
</tr>
<tr>
<td>Current smoker, %</td>
<td>51.4</td>
<td>21.4</td>
<td>34.0</td>
<td>31.2</td>
</tr>
<tr>
<td>Blood pressure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Systolic blood pressure, mmHg</td>
<td>146.4 ± 19.5(^a)</td>
<td>150.9 ± 19.3(^a)</td>
<td>113.8 ± 9.8</td>
<td>114.4 ± 10.1</td>
</tr>
<tr>
<td>Diastolic blood pressure, mmHg</td>
<td>86.4 ± 13.0(^a)</td>
<td>91.4 ± 12.2(^a)</td>
<td>69.8 ± 7.7</td>
<td>70.3 ± 7.2</td>
</tr>
<tr>
<td>Treatment of hypertension, %</td>
<td>92.6</td>
<td>75.4</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Blood chemistry</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum creatinine, mg/dl</td>
<td>0.87 ± 0.69(^a)</td>
<td>0.75 ± 0.50(^a)</td>
<td>0.73 ± 0.18</td>
<td>0.70 ± 0.23</td>
</tr>
<tr>
<td>Fasting plasma glucose, mg/dl</td>
<td>105.3 ± 28.7</td>
<td>109.0 ± 31.0(^a)</td>
<td>104.0 ± 41.9</td>
<td>99.2 ± 22.7</td>
</tr>
<tr>
<td>Serum total cholesterol, mg/dl</td>
<td>204.4 ± 31.1(^c)</td>
<td>213.4 ± 33.8</td>
<td>209.2 ± 33.7</td>
<td>215.6 ± 34.1</td>
</tr>
<tr>
<td>Serum triglyceride, mg/dl</td>
<td>129.8 ± 82.2(^a)</td>
<td>141.3 ± 124.6(^a)</td>
<td>108.1 ± 67.1</td>
<td>110.1 ± 71.6</td>
</tr>
<tr>
<td>Serum HDL cholesterol, mg/dl</td>
<td>56.2 ± 16.7(^a)</td>
<td>61.9 ± 19.5</td>
<td>60.6 ± 16.0</td>
<td>63.5 ± 17.5</td>
</tr>
</tbody>
</table>

Values are means ± SD.

For some variables, subjects with insufficient information are not included in the calculation.

\(^a\)P < 0.001, case group versus control group by the unpaired t-test in each tier.

\(^b\)Because of differences in the questionnaire, smoking status is categorized into two groups (non-smoker or smoker) in the tier 2 panel.

\(^c\)P < 0.01, case group versus control group by the unpaired t-test in each tier.

The candidacy of \(ADD2\) as a hypertension susceptibility gene has been supported by several physiological and biochemical findings (20–22), together with some evidence from the studies of molecular genetics (23–27). Adducin is a ubiquitously expressed membrane-skeleton heteromeric protein composed of different subunits, \(\alpha\)-, \(\beta\)- and \(\gamma\)-subunits. It is known to play a substantial role in the regulation of membrane ion transport. Point mutations of the \(ADD2\) gene located on chromosome 12q23.3. This gene is polymorphic in Asian populations (http://www.ncbi.nlm.nih.gov/SNP/). Also, it has to be noted that one previous study (28) showed significant evidence for hypertension linkage in the 2p13 region (a peak of 2.84 LOD at 93 cM), where the \(ADD2\) locus is located. However, our investigation in the \(ADD2\) locus, we could not find either a clear LD block-like structure or potentially functional SNPs in the vicinity of three disease-associated SNPs (rs2024453, rs3755351 and rs3771426), which are located in the putative promoter region and intron 1, apart from rs10084293 located within a disease-related LD block of \(ADD2\) (Fig. 3). We have assessed the independence of multiple associated SNPs in \(ADD2\) by logistic regression analysis and have found that the observed association in this gene could be explained principally by the most significant SNP (rs3755351) (see Supplementary Explanation). Once these associations are validated in an independent study panel, further extensive searches of functional SNPs in the \(ADD2\) locus are warranted.

Our high-density association study has also highlighted the \(KIAA0789\) gene located on chromosome 12q23.3. This gene encodes a putative protein, LOC9671, which is expressed principally in the central nervous system and modestly in the pancreas (unpublished data). The predicted gene structure of \(KIAA0789\) involves 9 exons, spanning 120 kb. There is a clear LD block in the 5’ region of the putative exon 1 (~3.8 kb in size), whereas we have found two other LD block-like structures within the \(KIAA0789\) gene (Fig. 3). Two disease-associated SNPs (rs3794260 and rs9739493) have turned out to reside in different LD blocks, and the construction of their haplotypes does not seem to provide much additional information on disease association. Although the precise gene structure and gene function remain unknown, \(KIAA0789\) appears to contain a carboxy-binding WSC domain, and its homologs are likely to exist in mice and rats according to the database information (http://www.ncbi.nlm.nih.gov/). Again, detailed investigation including independent
replication of disease association will lead us to clarify the etiological relevance of \textit{KIAA0789} to hypertension.

Another, potential disease association, though modest statistical significance, has been found for \textit{M6PR}. The \textit{M6PR} gene encodes a cation-dependent receptor for mannose-6-phosphate groups on lysosomal enzymes and plays a critical role in the segregation and targeting of lysosomal enzymes to lysosomes. Thus far, no functional relation between \textit{ADD2}, \textit{KIAA0789} and \textit{M6PR}—are depicted with solid circles as mentioned earlier.

We should bear in mind several limitations inherent in the present study. First, the level of genome coverage is an issue of heated debate (10,18). Some people may argue that our \textit{a priori} marker selection strategy is gene-centric without utilizing LD information and hence it is not sufficient to pick up as many modest associations as possible in genome-wide searches of hypertension susceptibility genes. A comprehensive framework of common variations throughout the human genome has been made available by the recent completion of the International HapMap Project (29). On the basis of our assessment, the JSNP screening markers in this study cover 20.6% of the HapMap SNPs, whereas a substantial proportion (~30%) of SNPs appear to be unique to JSNP (Table 1 and Supplementary Material, Fig. S1). Under these circumstances, an ideal set of SNPs for our study would encompass deliberately selected tag SNPs (principally common genetic variants) and additional ‘singleton’ SNPs (sometimes rare genetic variants). Besides this argument of tag SNPs, there are two points of weakness regarding genome coverage as follows: (i) sex chromosome markers have been excluded from the analysis because of the pre-determined policy of multi-disease collaborative study in the Japanese Millennium Genome Project, and (ii) a substantial part of the expressed human genes is not covered by the JSNP database (11), in which the fundamental SNP data were almost fixed in the middle of 2003. Second, the statistical power attainable by our study panel needs to be taken into consideration. For the last few years, genotype costs have fallen dramatically, yet present economic and experimental conditions make it necessary, in practice, to reduce the number of genotyped samples down to a moderately sized case group (188 subjects in our study) at the initial screening with approximately 80 000 SNPs. We arbitrarily set the selection criteria of OR $\geq 1.4$ and $P \leq 0.015$ in transition from tier 1 to tier 2, where the overall statistical power is estimated to be 10–45% for a disease-associated SNP of OR $= 1.4$ and 1–8% for that of OR $= 1.2$, assuming the disease allele frequency within 0.1–0.9 and the disease prevalence of 0.25. Thus, it is likely that our study design allows for capturing less than half of the true disease associations particularly with regard to modest genetic susceptibility. Third, ethnic diversity has not been tested within the scope of the present study. Instead of using commercially available SNP sets aimed at full genomic coverage, we have attempted to focus on potentially functional variants and also relatively common SNPs (MAF $\geq 0.1$) in the Japanese population. Accordingly, some of disease-associated SNPs listed in Table 3 may be rare or not polymorphic in the other ethnic groups. To clarify allele frequency representation of individual loci and etiological impacts attributable to them, further examination is required in the context of ethnic diversity.

During our preparation of this report, two genome-wide association studies for hypertension and/or blood pressure have been performed in Caucasians (30,31). When our results are compared with public data sets for these association statistics, a few SNPs in the regions of interest appear to show a tendency of association with hypertension or blood pressure;
<table>
<thead>
<tr>
<th>dbSNP number</th>
<th>Gene symbol</th>
<th>Orientation</th>
<th>dbSNP number</th>
<th>Gene symbol</th>
<th>Orientation</th>
<th>dbSNP number</th>
<th>Gene symbol</th>
<th>Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3755351</td>
<td>ADD2</td>
<td>C/A</td>
<td>0.14</td>
<td>0.19</td>
<td>0.17</td>
<td>0.022/0.21</td>
<td>0.22</td>
<td>0.21</td>
</tr>
<tr>
<td>rs3771426</td>
<td>ADD2</td>
<td>T/C</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.19/NA</td>
<td>0.2</td>
<td>0.19</td>
</tr>
<tr>
<td>rs204458</td>
<td>ADD2</td>
<td>T/C</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.29</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>rs10084293</td>
<td>ADD2</td>
<td>G/A</td>
<td>0.14</td>
<td>0.17</td>
<td>0.17</td>
<td>0.21/0.18</td>
<td>0.21</td>
<td>0.20</td>
</tr>
<tr>
<td>rs794260</td>
<td>KIAA0789</td>
<td>G/A</td>
<td>0.14</td>
<td>0.17</td>
<td>0.17</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>rs9739491</td>
<td>KIAA0789</td>
<td>T/C</td>
<td>0.14</td>
<td>0.17</td>
<td>0.17</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>rs1065762</td>
<td>M6PR</td>
<td>C/G</td>
<td>0.21</td>
<td>0.22</td>
<td>0.22</td>
<td>0.25/0.24</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>rs1065762</td>
<td>M6PR</td>
<td>T/G</td>
<td>0.14</td>
<td>0.17</td>
<td>0.17</td>
<td>0.44</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td>rs3787240</td>
<td>EYA2</td>
<td>C/T</td>
<td>0.25</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19/0.18</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>rs3787241</td>
<td>EYA2</td>
<td>G/A</td>
<td>0.24</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19/0.19</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>rs3787241</td>
<td>EYA2</td>
<td>G/A</td>
<td>0.24</td>
<td>0.21</td>
<td>0.22</td>
<td>0.19/0.19</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>rs3781987</td>
<td>THAP2</td>
<td>T/A</td>
<td>0.43</td>
<td>0.37</td>
<td>0.37</td>
<td>0.36/0.36</td>
<td>0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>rs3741691</td>
<td>THAP2</td>
<td>A/C</td>
<td>0.26</td>
<td>0.22</td>
<td>0.22</td>
<td>0.21/0.20</td>
<td>0.19</td>
<td>0.21</td>
</tr>
<tr>
<td>rs1298463</td>
<td>CCDC131</td>
<td>A/G</td>
<td>0.26</td>
<td>0.22</td>
<td>0.22</td>
<td>0.21/0.19</td>
<td>0.19</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Two SNPs of EYA2, rs3787240 and rs3787241, are located closely (only 296 bp apart) and are in complete LD (r² = 1.00) to each other. Also, rs3741691 and rs1298463 are located closely (44 kb apart) and have turned out to be in strong LD (r² = 0.99–1.00) to each other.

In tier 1 control subjects, the figures in parentheses are minor allele frequencies (MAFs) calculated separately in the Ref. 1 panel—the other disease patients who can be regarded as arbitrary general controls; Ref. 2 panel—752 individuals from the Japanese general population (see Materials and Methods).

The OR was calculated as the ratio of the odds of disease in chromosomes with major alleles relative to those without them.
for example, an SNP (rs17006246) in ADD2, which is in strong LD ($r^2 = 0.806$ and $D' = 1$ in the HapMap JPT population) with rs3755351, the most significant SNP in our study, is modestly associated with hypertension status ($P = 0.029$) in the Diabetes Genetics Initiative (DGI) study but the direction of effect is opposite between rs3755351 typed in this study and rs17006246 typed in the DGI study. On the other hand, rs1805740, in strong LD with an SNP (rs1805762) in M6PR, is modestly associated with hypertension status ($P = 0.036$) in the Wellcome Trust Case Control Consortium study with the same direction of effect as in this study (see Supplementary Material, Table S3).

In summary, our high-density association study provides a list of gene loci potentially predisposing people to hypertension, which awaits replication across populations. With the available samples, we have observed an association of SNPs including three SNPs clusters (or gene loci) in the Japanese populations. In face of the complex nature of disease etiology, it seems to be a formidable task but worth challenging that we eventually apply the SNPs information to improved prevention, diagnosis and treatment of hypertension.

**MATERIALS AND METHODS**

**Study design**

We performed a large-scale association study for genes susceptible to hypertension by using a three-tiered genotyping approach (tiers 1, 2 and 3) as depicted in Figure 1. All methods of the study were approved by the review committees of the individual institutions involved in the present study. All subjects provided written informed consent for participation.

In the gene-centered genome-wide exploratory test in tier 1, we carried out genotyping of 83,802 SNPs (3007 of which were excluded from the analysis because they are on sex chromosomes or in the unknown locations) using genomic DNAs from 188 Japanese male hypertensive patients and 752 unrelated Japanese individuals (referred to as general population controls) and another panel of 752 Japanese subjects (referred to as arbitrarily defined controls) who were affected with any of the other four common diseases including gastric cancer, diabetes mellitus, bronchial asthma and Alzheimer’s disease; each of these was investigated as the ‘Japanese Millennium Genome Project’ (Fig. 1). The theoretical basis of adopting this exploratory test scheme was previously reported elsewhere (32). Cases were enrolled from the clinical practice or the annual medical checkup of university hospitals and medical institutions according to the uniformly defined criteria. These included (i) systolic blood pressure $\geq 160$ mmHg, diastolic blood pressure $\geq 95$ mmHg, or both on two consecutive visits for untreated subjects; (ii) patients receiving long-term antihypertensive treatments; (iii) no secondary form of hypertension as evaluated by an extensive workup; (iv) family history of hypertension, i.e. at least one hypertensive subject detectable among parents and siblings of the participants; (v) an age of onset known to be between 30 and 59 years. Moreover, only male subjects with BMI $< 25$ kg/m$^2$ were selected in tier 1. We compared allele frequencies and/or genotype distributions in hypertensive patients and two population control panels and evaluated deviation from HWE at each of the genotyped loci. For the subsequent screening
in tier 2, we selected SNPs (i) with OR ≥ 1.4 and P ≤ 0.015 against either of two population control panels and with concordant OR tendency against two control panels; (ii) with MAF ≥ 0.1 and (iii) not showing significant deviations (P = 0.01 level) from Hardy–Weinberg expectations in the patient or control panels.

In tier 2 (which comprised 752 hypertensive patients and 752 normotensive controls), we further tested the SNPs thus screened in tier 1, which effectively constituted the first ‘case (tiers 1 and 2) versus unaffected control (tier 2)’ study. Here, cases in tier 2 were selected according to the criteria (i)–(v) mentioned earlier for tier 1. Normotensive controls, on the other hand, were defined as follows: (i) systolic blood pressure ≤130 mmHg and diastolic blood pressure ≤85 mmHg without receiving antihypertensive treatments; (ii) age ≥ 50 years and (iii) no family history of hypertension. Both males and females were included in tier 2 without reference to BMI. We selected SNPs (i) with P-value ≤ 0.05 when comparing allele frequency; and (ii) with P-value ≤ 0.01 when comparing genotype distribution between (tiers 1 and 2) cases and (tier 2) controls by χ² test statistics.

In tier 3 (which comprised 619 hypertensive patients and 1406 normotensive controls), we performed the second ‘case versus unaffected control’ study to examine significant associations observed in tiers 1 and 2. The diagnostic criteria in tier 3 were identical to those in tier 2. For the assessment of assumptions when using statistical models in the present study, quantile–quantile plots of P-values were depicted for each stage of association test described in Supplementary Explanation.

No significant population stratification was observed for samples in tier 1 when it was assessed with the methods reported by Patterson et al. (33). However, the presence of population stratification was indicated for samples in the first stage ‘case (tiers 1 and 2) versus unaffected control’ study. We observed moderate bias in genotype frequency of some SNPs between the two tiers, which may have resulted from technical/experimental artifacts between genotyping of cases in tiers 1 and 2. Therefore, the trend test statistic at this analytical stage was corrected according to the significant eigenvector (see Supplementary Explanation). Stratification in tier 3 was not detected but could not be ruled out because of the relatively small number of SNPs (n = 75) genotyped in tier 3. As for the nine SNPs that showed significant disease association after multi-stage screening, they were not correlated with the significant eigenvector detected in tiers 1 and 2 cases and tier 2 controls. The P-values for nine SNPs were similar between the nominal and the EIGENSTRAT-corrected ones; for example, the nominal P-value was 0.0029 and the EIGENSTRAT-corrected P-value was 0.0069 at rs3755351 in ADD2.

**SNP discovery in the selected genes**

Approximately 38 kb of genomic sequence spanning the exons and the 5' and 3'-untranslated regions of three genes, ADD2, KIAA0789 and M6PR, was re-sequenced in 48 Japanese control individuals to identify potentially functional SNPs. Since KIAA0789 had not been fully annotated, the arbitrary positions of translation initiation sites were estimated according to the human genome database. From the SNPs thus identified, tag SNPs were selected for the three genes with the algorithm that we previously reported (35). These tag SNPs were then used for the case–control analysis in tier 3 to further examine association signals seen throughout the multi-staged screening. We deposited the identified SNP information in the NCBI’s SNP database and also in our own database, JMDBase (Japan Metabolic Disease Database).

**Statistical analysis**

The SNPs were tested individually for the statistical significance of disease association with the χ²-test statistic, which evaluated three inheritance models—(2 × 3) contingency table, dominant and recessive models—for genotype distributions and independence on [2 × 2] contingency table for allele frequencies. Here, the most significant P-values among three inheritance models were adopted for genotype distributions when we selected SNPs for screening in tier 3. The criteria for declaring suggestive evidence of disease association were arbitrarily set at each analytical stage as summarized in Figure 1, and they are described in the Results section. SNPs’ genotype departures from HWE were tested using the χ²-test with 1 degree of freedom.

In the three genes showing significant association signals, the extent of LD was measured in terms of an LD coefficient r² before the analysis of haplotype structure. Within each LD block, haplotypes were inferred from genotype data by the SNPHAP software for the case and control groups, respectively.

We randomly permuted the genotype of individuals across different panels, 100 times per SNP, and counted the ratio of permutations that fulfill the screening criteria. This ratio indicates the specificity of the study. According to the P-value distribution of the permutations, we evaluated the probability of observing an SNP with P-value no larger than the actual minimum. This probability indicates the experiment-wise P-value. For the specific prevalence and penetrance, we calculated genotype frequency and randomly generated genotypes according to their frequency. We generated genotypes for 1000 simulations of each panel and computed the ratio of simulations that could pass the screening. This ratio is considered the sensitivity of the study.
Values were expressed as means ± SD unless otherwise indicated.

Uniform resource locators

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG Online.

ACKNOWLEDGEMENTS
We acknowledge the outstanding contributions of the SNP Research Center, the Institute of Physical and Chemical Research (RIKEN), the National Cancer Center (NCC) and the International Medical Center of Japan (IMCJ) employees who provided technical and infrastructural support for this work. Above all, we thank the patients and study subjects who made this work possible and who give it value. We thank all the people who continuously support the Suita Study. We also thank Drs Yukio Yamam, Yoshinori Kobu, Masakuni Serizawa, Ken Sugimoto, Hiroshi Akasaka, Chikanori Makibayashi and many physicians in Amagasaki Medical Association, Sadahiko Iwamoto, Yoichi Ohno, Mikano Hishiki, Kotoko Kosegu, Noriko Aoi, Akira Haketa, Yoichi Izumi, Toru Izumi, Taishi Sasaoka, Shigeru Hohda and Takeharu Hayashi, and Ms Kaoru Sugama for their contribution to collecting DNA samples and clinical information accompanied; Drs Hiromi Rakugi and Masafumi Hohda and Takeharu Hayashi, and Ms Kaoru Sugama for their advice and assistance.

Conflict of Interest statement. None declared.

FUNDING
This work was supported by the Grant-in-Aid for Scientific Research on Priority Areas (Millennium Genome Project) from the Ministry of Education, Science, Sports, and Culture of Japan, the grants from the Program for Promotion of Fundamental Studies in Health Sciences of Pharmaceuticals and Medical Devices Agency (PMDA) and a grant for the Core Research for the Evolutional Science and Technology, from the Japan Science Technology Agency (JST).

REFERENCES


