<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Identification of igf1, slc4a4, wwox, and sfmbt1 as hypertension susceptibility genes in han chinese with a genome-wide gene-based association study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Yang, HC; Liang, YJ; Chen, JW; Chiang, KM; Chung, CM; Ho, HY; Ting, CT; Lin, TH; Sheu, SH; Tsai, WC; Chen, JH; Leu, HB; Yin, WH; Chiu, TY; Chern, CI; Lin, SJ; Tomlinson, B; Guo, Y; Sham, PC; Cherny, SS; Lam, TH; Thomas, G; Pan, WH</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>Plos One, 2012, v. 7 n. 3</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2012</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/149135">http://hdl.handle.net/10722/149135</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
Identification of \textit{IGF1}, \textit{SLC4A4}, \textit{WWOX}, and \textit{SFMBT1} as Hypertension Susceptibility Genes in Han Chinese with a Genome-Wide Gene-Based Association Study

Hsin-Chou Yang\textsuperscript{1}, Yu-Jen Liang\textsuperscript{1,3}, Jaw-Wen Chen\textsuperscript{4}, Kuang-Mao Chiang\textsuperscript{2,5}, Chia-Min Chung\textsuperscript{2}, Hung-Yun Ho\textsuperscript{6}, Chih-Tai Ting\textsuperscript{6}, Tsung-Hsien Lin\textsuperscript{7}, Sheng-Hsiung Sheu\textsuperscript{7}, Wei-Chuan Tsai\textsuperscript{8}, Jyh-Hong Chen\textsuperscript{8}, Hsin-Bang Leu\textsuperscript{4}, Wei-Hsian Yin\textsuperscript{9}, Ting-Yu Chiu\textsuperscript{10}, Ching-Iuan Chern\textsuperscript{10}, Shing-Jong Lin\textsuperscript{4}, Brian Tomlinson\textsuperscript{11}, Youling Guo\textsuperscript{12}, Pak C. Sham\textsuperscript{12,13}, Stacey S. Cherry\textsuperscript{12,13}, Tai Hing Lam\textsuperscript{14}, G. Neil Thomas\textsuperscript{15}, Wen-Harn Pan\textsuperscript{2,16,*}

1 Institute of Statistical Science, Academia Sinica, Taipei, Taiwan, 2 Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan, 3 Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei, Taiwan, 4 National Yang-Ming University School of Medicine and Taipei Veterans General Hospital, Taipei, Taiwan, 5 School of Public Health, National Medical Defense Center, Taipei, Taiwan, 6 Cardiovascular Center, Taichung Veterans General Hospital, Taichung, Taiwan, 7 Department of Internal Medicine, Kaohsiung Medical University Hospital, Kaohsiung, Taiwan, 8 Department of Internal Medicine, College of Medicine, National Cheng Kung University, Tainan, Taiwan, 9 Division of Cardiology, Cheng-Hsin Rehabilitation Medical Center, Taipei, Taiwan, 10 Division of Cardiology, Min-Sheng General Hospital, Taoyuan, Taiwan, 11 Department of Medicine and Therapeutics, The Chinese University of Hong Kong, Hong Kong, China, 12 Department of Psychiatry, The University of Hong Kong, Hong Kong, China, 13 The State Key Laboratory of Brain and Cognitive Sciences, The University of Hong Kong, Hong Kong, China, 14 School of Public Health, The University of Hong Kong, Hong Kong, China, 15 Public Health, Epidemiology and Biostatistics, School of Health and Population Sciences, University of Birmingham, Birmingham, United Kingdom, 16 Division of Preventive Medicine and Health Services Research, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan

Abstract

Hypertension is a complex disorder with high prevalence rates all over the world. We conducted the first genome-wide gene-based association scan for hypertension in a Han Chinese population. By analyzing genome-wide single-nucleotide-polymorphism data of 400 matched pairs of young-onset hypertensive patients and normotensive controls genotyped with the Illumina HumanHap550-Duo BeadChip, 100 susceptibility genes for hypertension were identified and also validated with permutation tests. Seventeen of the 100 genes exhibited differential allelic and expression distributions between patient and control groups. These genes provided a good molecular signature for classifying hypertensive patients and normotensive controls. Among the 17 genes, \textit{IGF1}, \textit{SLC4A4}, \textit{WWOX}, and \textit{SFMBT1} were not only identified by our gene-based association scan and gene expression analysis but were also replicated by a gene-based association analysis of the Hong Kong Hypertension Study. Moreover, cis-acting expression quantitative trait loci associated with the differentially expressed genes were found and linked to hypertension. \textit{IGF1}, which encodes insulin-like growth factor 1, is associated with cardiovascular disorders, metabolic syndrome, decreased body weight/size, and changes of insulin levels in mice. \textit{SLC4A4}, which encodes the electrogenic sodium bicarbonate cotransporter 1, is associated with decreased body weight/size and abnormal ion homeostasis in mice. \textit{WWOX}, which encodes the WW domain-containing protein, is related to hypoglycemia and hyperphosphatemia. \textit{SFMBT1}, which encodes the scm-like with four MBT domains protein 1, is a novel hypertension gene. \textit{GRB14}, \textit{TMEM56} and \textit{KIAA1797} exhibited highly significant differential allelic and expressed distributions between hypertensive patients and normotensive controls. \textit{GRB14} was also found relevant to blood pressure in a previous genetic association study in East Asian populations. \textit{TMEM56} and \textit{KIAA1797} may be specific to Taiwanese populations, because they were not validated by the two replication studies. Identification of these genes enriches the collection of hypertension susceptibility genes, thereby shedding light on the etiology of hypertension in Han Chinese populations.


Editor: Leonardo Marín-Ramírez, National Institutes of Health, United States of America

Received February 10, 2011; Accepted February 7, 2012; Published March 29, 2012

Copyright: © 2012 Yang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Research Program for Genomic Medicine (NSC 97-3112-B-001-027, NSC 98-3112-B-001-013, NSC 99-3112-B-001-009, and NSC 99-3112-B-001-028), National Science Council grants (NSC 97-2314-B-001-006-MY3, NSC 100-2314-B-001-005-MY3, and NSC 100-2314-B-010-002), the Taiwan, the Career Development Award of Academia Sinica (AS-100-CDA-M03), the Academia Sinica Genomics and Proteomics Program (2003-2006) (521023-92-81, 521023-93-81, 521023-94-19 and 30-34-GSK), and the Academia Sinica Genomic Medicine Multicenter Studies (2007-present) (40-05-GMM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: pan@ibms.sinica.edu.tw

Introduction

Hypertension [OMIN #145500] is a common disorder that is associated with both environmental and genetic factors [1] and exhibits moderate to high estimates of heritability ranging from 20% to 60% [2]. The disease has important public health effects throughout the world. One of every four adults...
globally is hypertensive, and the number of adults with hypertension in 2025 is predicted to increase by about 60% to a total of 1.56 billion people [2]. In Taiwan, one-third of adult males and one-fifth of adult females have hypertension [3]. Hypertension is one of the most critical risk factors for cerebral hemorrhage and infarction, coronary heart disease, heart failure, and kidney disease. The combination of these diseases contributes to a considerable proportion of mortality in the world.

Identifying hypertension susceptibility genes is potentially useful for understanding the complex genetic mechanism. Screening high-risk individuals and preventing disease development by early lifestyle and pharmacological treatment can contribute to the reduction of hypertension complications and cardiovascular mortality. Efforts to identify hypertension susceptibility genes have been ongoing for several decades. Candidate gene linkage analyses and genome-wide linkage approaches have identified hypertension susceptibility genes such as \\textit{ECE} (1p36.1) [4], \\textit{AGT} (1q42-43) [5,6], \\textit{AGTR1} (3q21-q25) [7], \\textit{HHT7} (17q) [8,9], \\textit{HIT2} (15q) [9], \\textit{HIT3} (2p23-p24) [10], \\textit{HIT4} (12p) [11], \\textit{HIT5} (20q) [12], \\textit{HIT6} (5p) [12], \\textit{HIT7} (3p14.1-q12.3) [13], and \\textit{HIT8} (18q21) [14]. In recent years, many single nucleotide polymorphism (SNP) markers have been discovered and broadly applied to study the genetics of complex diseases [15]. Methods to map the genes of complex diseases have evolved from linkage approaches for rare Mendelian traits to genome-wide association studies (GWAS). Several GWAS have been conducted for blood pressure or hypertension in recent years [16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34]. Blood pressure and hypertension susceptibility genes identified using GWAS are summarized (Table 1). However, most studies identified genes with small-to-moderate effects, and the genetic variations identified so far collectively explain only a small fraction of the genetic variability attributed to essential hypertension. In addition, most studies were conducted with Caucasian samples. Previous Asian studies were carried out primarily in Japanese and Korean populations and the findings may or may not be applicable to Han Chinese populations.

In 2009, we published the first young-onset hypertension (YOH) genome-wide association mapping, which analyzed 175 patients with hypertension and 175 normotensive matched controls with the Affymetrix Human Mapping 100 K Set in the Han Chinese population of Taiwan [29]. We selected YOH as the phenotype of interest because it has a stronger genetic component than late-onset hypertension. Because we aimed to identify more disease genes responsible for YOH, we recruited more patient and control samples (400 YOH patients and 400 normotensive matched controls) and used a denser SNP genotyping platform (the Illumina HumanHap550-Duo BeadChip which provides 546,376 autosomal SNPs). There are 109 YOH patients and 98 normotensive controls used in both YOH GWAS. In addition to a genome-wide SNP-based association scan (genome-wide SBAS), based on an extension of our previously developed gene-based algorithms [33], we conducted a genome-wide gene-based association scan (genome-wide GBAS) to increase power, diminish the effect of locus heterogeneity, reduce the multiple testing burden, incorporate gene information, and provide a direct interpretation of association signals to genes. The genes identified with differential allelic associations in patient and control groups were then validated with gene expression analyses. We also studied expression quantitative trait loci (eQTL) associated with the differentially expressed genes identified in the present study and previous GWAS of hypertension [16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34] and linked them to hypertension. The analysis provides a deeper insight of phenotype-genotype relationship and illustrates that whether SNPs and genes influence YOH via gene regulation or via mechanisms rather than varied gene expression. Moreover, two replication analyses for this Taiwan Hypertension Study (TWNHS) were performed using genotype data from the Hong Kong Hypertension Study (HKHS) [36] and the Wellcome Trust Case-Control Consortium Hypertension Study (WTCCCHS) [17]. The HKHS collected samples from a Han Chinese population using a family-based study design, and the WTCCCHS collected samples from a Great Britain population using an independent case-control study design. A flowchart for the overall study design and analysis methods is provided in Figure 1 and elaborated in the Materials and Methods section.

## Results

### Quality control and population stratification

For the 400 YOH patients and the 400 normotensive matched controls genotyped with the Illumina HumanHap550-Duo BeadChip, quality examination showed that all samples satisfied the genotype call rate (GCR) >0.99. One hypertensive patient and the matched normotensive control were excluded from the analysis due to an inconsistency of self-reported gender and X chromosome. Following quality examination of the SNPs, 86 SNPs with a GCR<0.9 were removed, and 324 additional SNPs were removed because of a departure from Hardy-Weinberg equilibrium (HWE) (pFDR<10−5) (where FDR is false discovery rate). An additional 70,809 SNPs were removed because of a minor allele frequency (MAF)<0.01. Ultimately, 475,157 validated SNPs (~87% of autosomal SNPs on the HumanHap550-Duo BeadChip) were analyzed in this study. The impact of population substructures in our genome-wide SBAS and GBAS was evaluated. The variance inflation fraction (VIF) was 1.027, which is close to 1 and suggests that the population admixture with respect to association tests was not significant in this study. The principal component plot shows all hypertensive and normotensive samples randomly mixed together (Figure S1). Moreover, the results of non-parametric median tests also show no significant differences between patient and control groups on the first ten principal components (p-values for the first ten principal components were 0.435, 0.523, 0.620, 0.178, 0.177, 0.435, 0.355, 0.831, 0.055 and 0.435). The results illustrate no substructure in patient and control groups in this study.

### Genome-wide SBAS

Initially, genome-wide SBAS was performed with four conditional logistic regression models abbreviated “CLRN”, “CLRC”, “CLRN,BMI”, and “CLRC,BMI”, in which the subscript “C” stands for a nominal SNP coding for a genotype-based analysis, and the subscript “N” stands for a continuous SNP coding for a trend-based analysis without or with a covariate adjustment for body mass index (BMI). No significant SNPs (pFDR<10−5 or p<10−5) were found. However, we observed that some genomic regions contained a significant proportion of SNPs with marginal significance (Figure S2A–S2D). For example, the region in KLLA1797 between 20,638,805 bp and 20,993,372 bp on chromosome 9p21 contains 84 SNPs probed on the Human-Hap550−Duo BeadChip. The proportions of SNPs with marginal significance (p<5×10−5) in the four conditional logistic regression analyses (CLRN, CLRC, CLRN,BMI, and CLRC,BMI) were 10%, 26%, 20%, and 23%, with the smallest values of p-values on KLLA1797 of 7.56×10−4, 7.39×10−4, 4.31×10−5, and 7.69×10−5, respectively. P-values of genome-wide SBAS can be downloaded at the website, http://pan.ibms.sinica.edu.tw/hypertension/data.
Table 1. A summary table of blood pressure and hypertension susceptibility genes identified by using GWAS.

<table>
<thead>
<tr>
<th>Ref</th>
<th>Population</th>
<th>Analysis (n = sample size)</th>
<th>Blood pressure or hypertension- associated genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>[19]</td>
<td>Multi-ethnicity</td>
<td>Single-stage: GWAS (n = 1,260, 1,233 and 1,327 individuals in different analyses)</td>
<td>CCL20, WDRA69, CDH13, LPP</td>
</tr>
<tr>
<td>[20]</td>
<td>Asian ancestry – Japan</td>
<td>Stage 1: GWAS (n = 188 cases+1,054 controls); Stage 2: Replication (n = 752 cases+752 controls); Stage 3: Replication (n = 619 cases+1,406 controls)</td>
<td>ADD2, RIAA0789, MPFR</td>
</tr>
<tr>
<td>[22]</td>
<td>Amish Caucasian</td>
<td>Stage 1: GWAS (n = 542 Amish Caucasians); Stage 2: Replication (n = 1,367 Amish Caucasians and 5,804 non-Amish Caucasians)</td>
<td>STK39</td>
</tr>
<tr>
<td>[23]</td>
<td>European ancestry</td>
<td>Stage 1: GWAS (n = 34,433); Stage 2: Replication (n = 3,511 subjects of European ancestry and 12,889 subjects of Indian ancestry)</td>
<td>CYP17A1, CYP17A2, FG5S, SH2B3, MTHFR, C11orf107, ZNF652, PLEC3</td>
</tr>
<tr>
<td>[24]</td>
<td>European ancestry – Germany</td>
<td>Stage 1: GWAS (n = 1,017 Germany and 364 cases+596 controls); Stage 2: Replication (n = 1,551 Germany and 447 cases+1,119 controls); Stage 3: Replication (n = 1,097 Estonians and 596 cases+650 controls; 2,401 British cases+1,969 controls)</td>
<td>CDH13</td>
</tr>
<tr>
<td>[25]</td>
<td>Asian ancestry – Korea</td>
<td>Stage 1: GWAS (n = 8,842 Koreans); Stage 2: Replication (n = 7,861 Koreans)</td>
<td>ATP2B1</td>
</tr>
<tr>
<td>[26]</td>
<td>African ancestry – America</td>
<td>Stage 1: GWAS (n = 509 African American cases+508 controls); Stage 2: Replication (n = 366 West African cases+614 controls)</td>
<td>SLC2444, CACNA1H</td>
</tr>
<tr>
<td>[27]</td>
<td>European ancestry</td>
<td>Stage 1: GWAS (n = 29,136 European descent subjects); Stage 2: Replication (n = 34,433 European descent subjects)</td>
<td>ATP2B1, CYP17A1, PLEKH7, SH2B3, CACNB2, CSK ~ UKL, TBX5 ~ TBX5, UKL4</td>
</tr>
<tr>
<td>[28]</td>
<td>Asian ancestry – Taiwan</td>
<td>Stage 1: GWAS (n = 19,608 East Asian ancestry subjects); Stage 2: Replication (n = 10,518 East Asian ancestry subjects)</td>
<td>S7L ~ C7PA1, FIGN ~ GB14A, ENPEP, NPR3, TBX3, ALOH2</td>
</tr>
<tr>
<td>[29]</td>
<td>Asian ancestry – Taiwan</td>
<td>Stage 1: GWAS (n = 175 cases+175 controls); Stage 2: Replication (n = 833 cases+833 controls)</td>
<td>LOC344371, MYADM1, FAM89A, RASSGR3, IMPG1</td>
</tr>
<tr>
<td>[30]</td>
<td>African ancestry – America</td>
<td>Stage 1: GWAS (n = 8,591 African Americans); Stage 2: Replication (n = 11,882 African Americans); Stage 3: Replication (n = 69,899 European Americans)</td>
<td>GPR98 ~ ARRD3C, C21orf91, SLC25A42, HLA-B</td>
</tr>
<tr>
<td>[31]</td>
<td>European ancestry – Sweden</td>
<td>Stage 1: GWAS (n = 1,621 cases+1,699 controls); Stage 2: Replication (n = 19,845 cases+16,541 controls)</td>
<td>UMOD</td>
</tr>
<tr>
<td>[32]</td>
<td>European ancestry – Great Britain</td>
<td>Single-stage: GWAS (n = 2,000 cases+3,000 controls)</td>
<td>GRB39, XRCC4, MYO6, ZFAT, MACROD2</td>
</tr>
<tr>
<td>[33]</td>
<td>Asian ancestry – Korea</td>
<td>Stage 1: GWAS (n = 7,551 Koreans); Stage 2: Replication (n = 3,703 Koreans)</td>
<td>AKAP13</td>
</tr>
<tr>
<td>[34]</td>
<td>Asian ancestry – Japan</td>
<td>Stage 1: GWAS (n = 936 Japanese); Stage 2: Replication (n = 1,228 Japanese); Stage 3: Replication (n = 2,895 Japanese)</td>
<td>CCB1E1, MAP7, ZFP64, PCD18B, CDH2, WWOX</td>
</tr>
</tbody>
</table>

*Genes (marked in bold) which were also identified to show differential allelic distributions between patient and control groups by our GBAS. doi:10.1371/journal.pone.0032907.t001

Genome-wide GBAS

Genome-wide GBAS was then applied to increase the statistical power in the context of gene regions. The method of genome-wide GBAS is introduced in Method S1. The analysis identified 101 potential hypertension susceptibility genes (pFDR<10^{-5}) in at least one of the four genome-wide GBAS models, which combined p-values obtained from the genome-wide SBAS (Figure 2A–2D; Table S1). The 101 genes are classified into 12 categories (Category I to XII) according to gene significance in the four genome-wide GBAS models. In Table S1, three genes were identified by all four genome-wide GBAS regression models (Category I), six genes were identified by three of the four models (Categories II, III and VII), 25 genes were identified by two of the four models (Categories IV, V, VIII and X), and 67 genes were identified by only one model (Categories VI, IX, XI and XII). Empirical p-values of the 101 genes were calculated using a permutation procedure with 10,000 replicates. The gene-trait association information collected from the Genetic Association Database (GAD), Online Mendelian Inheritance in Man (OMIM), and Mouse Genome Informatics (MGI) databases for the 101 genes is summarized in Table S2. All the genes (except for FAM05) were successfully validated (pFDR<5x10^{-5}).

Gene expression analysis

As a validation process, gene expression analysis was conducted to compare transcriptional levels of the 100 identified genes for 12 matched patient–control pairs in which the 12 matched pairs of patients and controls were selected to have an extremely discordant genotypic distribution between groups. The method for selecting extremely discordant case-control pairs is introduced in Method S2. Among the 100 genes, gene expression data for 92 genes were provided by the Illumina HT12 Expression BeadChip, of which 17 genes were differentially expressed (p<0.05) (Table 2). Beta coefficients and their standard errors from ANCOVA are also provided (Table S3). Normalized gene expression data of the 17 genes can be downloaded at the website, http://pan.ibms.sinica.edu.tw/hypertension/data. These 17 genes showed not only differential allelic distributions but also differential transcriptional distributions between patient and control groups. Based on the gene expression data of the 17 differentially expressed genes, a cluster analysis of 12 pairs of hypertensive patients and normotensive controls only misclassified one normotensive control to the patient group (Figure 3A). The first three principal components of expression of the 17 genes explained ~62% of the total variation for the 12 matched pairs of
YOH patients and normotensive controls (Figure 3B). Fifteen of the 17 genes had data of gene expression in mouse. Tissue-specific analysis of gene expression showed that, except for LARS and FURIN, other genes exhibited differential transcriptional distributions between genetically hypertensive and normotensive inbred mouse strains in at least one tissue ($p < 10^{-2}$) (Table 2). Beta coefficients and their standard errors from ANOVA are also provided (Table S3).

We made use of the Genevar database version 3.0.1 [37] to investigate cis-acting expression quantitative trait loci (cis-eQTL) associated with the 17 differentially expressed genes identified in our human gene expression study (Table 2). Among the 17 genes, only AP3S1 did not contain cis-eQTL. The gene-centric cis-eQTL analysis found 361 cis-eQTL ($p < 10^{-2}$) in 2-Mb windows around the 16 genes, where 285 cis-eQTL on 13 genes were interrogated in the Illumina HumanHap550-Duo BeadChip. There were also 70 cis-eQTL on 11 genes (TMEM56, KIAA1977, LARS, GRB14, IGF1, WWOX, HECS, CSF1, CSNK1D3, TTC12 and SINE2) associated with YOH in our TWNHS study ($p < 10^{-2}$), where the disease associated alleles of 9 cis-eQTL increased genetic risk of YOH as well as up-regulated gene expression (Table S5).

Replication analysis
In the HKHS replication analysis, genome-wide SBAS contained ''sib-TDT'', ''QFAMSBP'', ''QFAMDBP'', and ''QFAMMBP'' for analyzing dichotomous disease status and three quantitative traits.

Figure 1. An analysis flowchart of this study.

doi:10.1371/journal.pone.0032907.g001
(quantitative systolic blood pressure [SBP], quantitative diastolic blood pressure [DBP], and quantitative mean blood pressure [MBP]). In the WTCCCHS, genome-wide SBAS was performed with four logistic regression models abbreviated “LRN”, “LRC”, “LRN,Gender”, and “LRC,Gender” in which the subscript “N” stands for a nominal SNP coding for a genotype-based analysis, and the subscript “C” stands for a continuous SNP coding for a trend-based analysis with or without a covariate adjustment for gender. Among the 17 genes that exhibited differential allelic distributions and differential transcriptional distributions between patient and control groups, four hypertension-associated genes (IGF1, SLC4A4, WWOX, and SFMBT1) were further confirmed by replication studies. The results are summarized in Table 3.

IGF1, which encodes insulin-like growth factor on 12q23.2, was not only identified by the GBAS in our TWNHS \( p_{\text{CLRN}} = 3.10 \times 10^{-7} \) \( p_{\text{FDR}_{\text{CLRN}}} = 2.47 \times 10^{-5} \); \( p_{\text{CLRC}} = 6.97 \times 10^{-11} \) \( p_{\text{FDR}_{\text{CLRC}}} = 2.20 \times 10^{-8} \); \( p_{\text{CLRC,BMI}} = 9.90 \times 10^{-10} \) \( p_{\text{FDR}_{\text{CLRC,BMI}}} = 2.95 \times 10^{-7} \) and gene expression analysis \( p_{\text{FAMCOVA}} = 3.21 \times 10^{-5} \) but was also identified by a GBAS of the HKHS \( p_{\text{QFAMSBP}} = 3.81 \times 10^{-7} \) \( p_{\text{FDR}_{\text{QFAMSBP}}} = 2.99 \times 10^{-5} \); \( p_{\text{QFAMDBP}} = 7.26 \times 10^{-5} \) \( p_{\text{FDR}_{\text{QFAMDBP}}} = 2.64 \times 10^{-2} \) and gene expression analysis \( p_{\text{FAMCOVA}} = 3.14 \times 10^{-7} \) but was also identified by a GBAS of the HKHS \( p_{\text{QFAMSBP}} = 9.19 \times 10^{-6} \) \( p_{\text{FDR}_{\text{QFAMSBP}}} = 1.63 \times 10^{-3} \); \( p_{\text{QFAMDBP}} = 4.19 \times 10^{-3} \) \( p_{\text{FDR}_{\text{QFAMDBP}}} = 4.79 \times 10^{-5} \). This gene is associated with decreased body weight/size, hematocrit, and abnormal ion homeostasis in mice [56]. In addition, SLC4A5 is a member of the same family of solute carriers as SLC4 and was also shown to be associated with blood pressure [57] and with hypertension [58].

Figure 2. Gene-based association tests based on p-values from the four conditional logistic regression models. (A) model CLRN, (B) model CLRN,BMI, (C) model CLRC, and (D) model CLRC,BMI. In each figure, the vertical axis is the FDR-adjusted p-values (−log10 scale) of gene-based association tests, and the horizontal axis is cumulative physical position (Mb scale). The red reference line signifies pFDR = 10−2.

doi:10.1371/journal.pone.0032907.g002
**Discussion**

We conducted the first genome-wide GBAS, using a novel region- and context-based method, to identify hypertension susceptibility genes in a Han Chinese population. Four hypertension susceptibility genes, *IGF1*, *SLC4A4*, *WWOX*, and *SEMBTI*, were found by a statistical permutation procedure, confirmed by gene expression analysis, and further replicated using data from the HKHS. Among them, *IGF1*, *SLC4A4*, and *WWOX* were previously implicated as hypertension candidate genes in humans and mice and are primarily associated with obesity, glucose metabolism, and ion homeostasis, which are well-known mechanisms of blood pressure regulation. *SEMBTI* encoding a drosophila scm-like protein is potentially involved in the transcriptional repression of homeotic genes in embryonic development via the histone methylation mechanism. Our study shows that it may play a novel role in hypertension, but its disease causing mechanism should be further elucidated [60]. Although no genes identified by us were significantly confirmed by the WTCCCHS data (Caucasian), *COMMD7*, which encodes COMM domain-containing protein 7 on 20q11.21, was replicated with a borderline significance (pFDR = 7.46 × 10^{-2}). *COMMD* and *COMM* domains 1–10 are extensively conserved in multicellular eukaryotic organisms and define a novel family of structural and functional homologs of *MURR1/COMMD1* that suppress NF-kappaB by affecting the association of NF-kappaB with chromatin [61]. *MURR1/COMMD1* is also a regulator of epithelial sodium channels, and therefore may regulate blood pressure [62].

Some potential hypertension susceptibility genes were identified by our genome-wide GBAS analysis and gene expression analysis but were not replicated in the GBAS analysis of HKHS or WTCCCHS. For example, *TMEM65* on 1p21.3 and *KIAA1779* on 9p21 were simultaneously identified by all four regression models of our genome-wide GBAS. For these two genes, the smallest p values in the genome-wide GBAS of the four regression models were 6.78 × 10^{-20} and 1.03 × 10^{-36}, respectively (Table 2), and the largest proportion that SNPs satisfy p < 10^{-2} in a SBAS among the four regression models is 58.82% and 27.38%.

**Table 2. Seventeen genes with differential allelic and transcriptional distributions identified by the TWNHS.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Raw p-values of GBAS of conditional logistic model&lt;sup&gt;ab&lt;/sup&gt;</th>
<th>Raw p-values of the human and mouse GE analysis&lt;sup&gt;ab&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CLR&lt;sub&gt;R&lt;/sub&gt;</td>
<td>CLR&lt;sub&gt;L&lt;/sub&gt;</td>
</tr>
<tr>
<td>TMEM56</td>
<td>9.83E-14</td>
<td>6.78E-20</td>
</tr>
<tr>
<td>KIAA1779</td>
<td>6.78E-16</td>
<td>1.13E-28</td>
</tr>
<tr>
<td>FMBT1</td>
<td>1.82E-10</td>
<td>6.83E-01</td>
</tr>
<tr>
<td>LARS</td>
<td>2.80E-09</td>
<td>6.44E-05</td>
</tr>
<tr>
<td>GRB14</td>
<td>5.04E-01</td>
<td>7.04E-12</td>
</tr>
<tr>
<td>IGF1</td>
<td>3.10E-07</td>
<td>6.97E-11</td>
</tr>
<tr>
<td>FURIN</td>
<td>3.62E-05</td>
<td>3.34E-08</td>
</tr>
<tr>
<td>WWOX</td>
<td>7.0E-02</td>
<td>7.47E-14</td>
</tr>
<tr>
<td>HLC5</td>
<td>5.88E-05</td>
<td>3.51E-08</td>
</tr>
<tr>
<td>CSF1</td>
<td>5.66E-05</td>
<td>4.67E-08</td>
</tr>
<tr>
<td>CSNK1G3</td>
<td>3.23E-02</td>
<td>3.39E-08</td>
</tr>
<tr>
<td>COMMD7</td>
<td>2.50E-04</td>
<td>8.74E-09</td>
</tr>
<tr>
<td>AP351</td>
<td>1.38E-02</td>
<td>5.04E-02</td>
</tr>
<tr>
<td>LC44</td>
<td>7.58E-01</td>
<td>1.90E-02</td>
</tr>
<tr>
<td>LACE1</td>
<td>5.37E-01</td>
<td>4.82E-01</td>
</tr>
<tr>
<td>TTC12</td>
<td>1.85E-03</td>
<td>1.01E-02</td>
</tr>
<tr>
<td>SYNE2</td>
<td>9.62E-01</td>
<td>6.71E-01</td>
</tr>
</tbody>
</table>

<sup>a</sup>Test results include p-values of the gene-based association analysis (CLR<sub>R</sub>, CLR<sub>L</sub>, CLR<sub>LSBMI</sub>, and CLR<sub>CBMI</sub>) and −log10(p) of gene expression (GE) analysis.

<sup>b</sup>Raw p-values of genes that reached pFDR < 0.05 in GBAS of the TWNHS and p-values of genes that reached p < 0.05 in gene expression studies are marked in bold.

<sup>NA</sup>NA denotes that expression data of a gene were not available in the mouse gene expression analysis.

doi:10.1371/journal.pone.0032907.t002
respectively. These two genes were validated by our human gene expression analysis \[ TMEM56\_pANCOVA = 4.48 \times 10^{-2}; \ PCLRN\_BMI = 4.76 \times 10^{-2} \] and \[ TMEM56 was also validated by the mice gene expression analysis \[ p_{PANOVA}\_WTCCHS = 3.62 \times 10^{-5}; \ pFDR_{pCLRN,BMI} = 3.34 \times 10^{-8}; \ pFDR_{pCLRN,BMI} = 1.52 \times 10^{-2}; \ p_{PANOVA}\_WTCCHS = 1.39 \times 10^{-2} \] (Table 2). In addition, \[ FURIN \] on 15q26.1, which encodes a subtilisin-like pro-protein convertase family protein [GBAS: \[ p_{PCLRN} = 3.62 \times 10^{-5}; \ p_{PANOVA} = 3.81 \times 10^{-3}; \ p_{PANOVA} = 4.48 \times 10^{-2}; \ PCLRN\_BMI = 1.39 \times 10^{-2} \] (Table 2), has been associated with blood pressure [63]. Due to the possibility of population specificity, these genes should be further studied even though they were not replicated in a GBAS of HKHS or WTCCHS.

Our GBAS also identified some hypertension-associated genes that showed differential allelic distributions between hypertensive patients and normotensive controls and the genes could be replicated by the GBAS of HKHS, even though no significant differences in gene expression were observed. \[ ACTN4 \] on 19q13 is associated with cortical renal glomerulopathies, abnormal kidney physiology, small kidney, proteinuria, and an abnormal blood urea nitrogen level [64]. \[ CYP7B1 \] on 8q21.3 is associated with lipoproteins in humans [65] and is associated with abnormal cholesterol homeostasis in mice [66]. \[ FAS \] on 10q24.1 is associated with hypertension [67], decreased body size, and increased body weight in mice [68,69]. These SNPs may exert their effects via mechanisms other than by varying gene expression.

This study investigated the association between DNA-level variants (SNPs) and disease status of YOH by examining differential allelic distributions between patient and control groups and also investigated the association between mRNA-level variants (gene expression) and disease status of YOH. An interesting question to answer is whether SNP variants influence YOH development via gene expression regulation. Therefore, we examined whether the genes with differential allelic and expression distributions between patient and control groups in the present study were regulated by SNPs in the 2-Mb windows around the differentially expressed genes (i.e., cis-eQTL mapping). Under a less stringent criterion, 70 SNPs on 11 genes (\[ TMEM56, KLA1797, LARS, GRB14, IGF1, WWOX, HLCS, CSF1, CSNK1G3, TTC12 and SYNE2 \]) were identified to be cis-eQTL, generating a testable hypothesis that the 70 cis-eQTL may regulate hypertension-associated genes in turn to influence disease status.

Among the 11 genes, \[ GRB14 \], which encodes a growth factor receptor binding protein on 2q22-q24, can interact with insulin receptors and insulin-like growth-factor receptors to modulate insulin signaling in mice [70,71]. This gene was found to be associated using GBAS and gene expression analyses of the TWNHS, and had been identified in a previous GWAS of blood pressure in East Asian populations [28]. Moreover, some hypertension-associated cis-eQTL act as intronic enhancers at transcriptional factor binding sites on this gene. This protein likely has an inhibitory effect on receptor tyrosine kinase signaling and thus on the insulin receptor signaling pathway, through which

### Table 3. Replication analysis of the seventeen genes with differential allelic and transcriptional distributions in cases and controls.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Raw p-values of GBAS of the HKHSa,b</th>
<th>Raw p-values of GBAS of the WTCCHSs,b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pANOVA</td>
<td>pFDR</td>
</tr>
<tr>
<td>TMEM56</td>
<td>5.60E-01</td>
<td>5.60E-01</td>
</tr>
<tr>
<td>SFMBT1</td>
<td>3.04E-01</td>
<td>6.93E-01</td>
</tr>
<tr>
<td>IGF1</td>
<td>2.02E-01</td>
<td>1.21E-04</td>
</tr>
<tr>
<td>FURIN</td>
<td>2.26E-01</td>
<td>2.26E-01</td>
</tr>
<tr>
<td>WWOX</td>
<td>1.05E-06</td>
<td>1.11E-03</td>
</tr>
<tr>
<td>HLCS</td>
<td>6.93E-01</td>
<td>5.08E-01</td>
</tr>
<tr>
<td>CSF1</td>
<td>4.31E-01</td>
<td>4.31E-01</td>
</tr>
<tr>
<td>CSNK1G3</td>
<td>5.82E-01</td>
<td>5.82E-01</td>
</tr>
<tr>
<td>COMMD7</td>
<td>2.26E-01</td>
<td>2.26E-01</td>
</tr>
<tr>
<td>SLCO4A4</td>
<td>4.33E-02</td>
<td>9.19E-06</td>
</tr>
<tr>
<td>LACE1</td>
<td>4.58E-01</td>
<td>5.37E-01</td>
</tr>
<tr>
<td>TTC12</td>
<td>6.03E-01</td>
<td>7.20E-03</td>
</tr>
<tr>
<td>SYNE2</td>
<td>9.08E-01</td>
<td>8.78E-01</td>
</tr>
</tbody>
</table>
GRB14 variants may exert their effect on the development of hypertension. *IGF1*, another insulin-related gene which encodes insulin-like growth factor on 12q23.2, was associated with hypertension by GBAS and gene expression analyses of the TWNHS, successfully validated by GBAS of the HKHS in this study, and was previously associated with cardiovascular disorders [30,39,40,41,42] and metabolic disorders [43,44,45,46,47,48,49,50,51,52]. Some hypertension-associated cis-eQTL act as intrinsic enhancers at transcriptional factor binding sites on this gene were also found. *HWOX* was found relevant by GBAS and gene expression analyses of the TWNHS, successfully validated by GBAS of the HKHS in this study, identified by a previous GWAS in hypertension in a Japanese population [34], and associated with hypertension-related traits in mice [59]. Moreover, some hypertension-associated cis-eQTL which act as intrinsic enhancers at transcriptional factor binding site on this gene were also found. *HWOX* encoded protein is more than 90 percent identical to the mouse protein responsible for modulation of tumor necrosis factor-alpha-induced apoptosis and therefore suggests an apoptosis correspondence for the human protein. This gene contains a short-chain dehydrogenase/reductase domain and is highly expressed in hormonally regulated normal tissues such as ovary, testis, and prostate. More recent evidence has suggested important roles of *HWOX* in the suppression of tumor growth, the regulation of a wide variety of cellular functions, steroid metabolism, and possibly hypertension [72,73].

There were other cis-eQTL with relevant biological functions. For example, five cis-eQTL on *SYNE2* were found, where rs9944035 is a missense variant leading to non-conservative change, rs10140978 locates in promoter/regulatory region, and rs1890908 is a missense variant leading conservative change. Other examples include that rs10988 on *LARS* is a missense variant leading to conservative change and rs333970 on *LARS* is a splicing site. These cis-eQTL may play relevant roles in basic genetics of gene regulation to confer susceptibility to YOH by suppressing, inhibiting, enhancing or activating transcriptional levels of the identified YOH associated genes.

In addition to examine cis-eQTL for our identified hypertension-associated genes in Table 2, we also used the Genevar database [37] to investigate cis-eQTL for the hypertension-associated genes in Table 1. Results showed that the genes identified by our integrative analysis of SNP and gene expression did contain a higher proportion of hypertension-associated cis-eQTL than a SNP-only GWAS; the proportions of genes containing hypertension-associated cis-eQTL were 0.846 versus 0.273 and the proportions of hypertension-associated cis-eQTL SNPs in genes were 0.140 versus 0.019. This suggests the importance of an integrative analysis of SNP and gene expression in the post-GWAS era.

Our previous genome-wide association study of YOH analyzed 175 matched pairs of patients and normotensive controls genotyped with the Affymetrix Human Mapping 100 K Set in the Han Chinese population of Taiwan [29]. The study identified a significant genetic interaction between rs1886985 on chromosome 6 and rs6129969 on chromosome 20. In addition, the study also identified a significant SNP quartet composed of rs9308945, rs6711736, rs6729869 and rs10495809 on chromosome 2, using a sliding-window p-value combination method. We also attempted to evaluate the genetic association of the six SNPs in the current study of 400 matched pairs of YOH patients and normotensive controls genotyped with the Illumina HumanHap550-Duo BeadChip in the same Han Chinese population of Taiwan. SNPs rs6711736 and rs10495809 are on the Illumina HumanHap550-Duo BeadChip and SNPs rs1886985, rs6129969, rs9308945 and rs6729869 were imputed based on the 1,000 Genomes Project reference panel using IMPUTE2 [74]. Results show that genetic association between YOH and the SNP quartet composed of rs9308945, rs6711736, rs6729869 and rs10495809 on chromosome 2 was confirmed (pCLRN = 1.79 × 10⁻⁷, pCLRC = 1.31 × 10⁻⁴, and pCLRBM = 2.96 × 10⁻⁷), but genetic interaction between rs1086903 and rs6129969 was not significant in this TWNHS.

The original WTCCCHS study only performed SBAS and no significant SNPs or genes attained a genome-wide significance level of 5 × 10⁻⁷ [17]. In our WTCCCHS data analysis, we performed both SBAS and GBAS to examine the 100 significant genes identified by the genome-wide SBAS in our TWNHS. We made a comparison of SBAS and GBAS in our WTCCCHS data analysis based on the 87 genes (SNP data on 13 out of 100 significant genes were not available on the Affymetrix Human Mapping 500 K Set). At a false discovery rate of pFDR < 0.05, GBAS identified three genes (*ACTN4, C17orf11* and *PPIP5K3*) and SBAS identified 2 genes (*ACTN4* and *TCF7L1*). The concordance rate of gene mapping results in SBAS and GBAS in our WTCCCHS data analysis is 96.5%.

GBAS has some merits compared with SBAS [35,75,76]. Because GBAS combines genetic contributions of multiple SNPs in a gene, the test power may be increased, and the effect of locus heterogeneity on gene mapping may be alleviated. In addition, the issue of over-correction with a large number of association tests in a genome-wide association study is mitigated because the number of association tests is reduced. Because the effect of the whole gene is evaluated, the association results can be directly interpreted using established gene functions and models. On the other hand, the gene is a natural unit in which a set of SNPs can be combined. Therefore, GBAS provides a convenient marker selection alternative to LD-based or sliding-window strategies for a genome-wide multifocus association scan. GBAS provides a convenient, feasible, and powerful tool for identifying genes involved in complex disorders.

Several limitations of this study should be noted when interpreting the results. First, the phenotype definition and sample ascertainment were not exactly the same in the three studies (TWNHS, HKHS, and WTCCCHS). For example, the YOH study in TWNHS recruited patients with hypertension with an age of onset ≤ 50, but this criterion was not considered in HKHS or WTCCCHS. Second, the three studies used different genome-wide multilocus association scan. GBAS provides a convenient marker selection alternative to LD-based or sliding-window strategies for a genome-wide multifocus association scan. GBAS provides a convenient, feasible, and powerful tool for identifying genes involved in complex disorders.
was described previously [17]. Samples in the WTCCCHS were genotyped with the Affymetrix which enables the whole-genome genotyping of 620,901 SNPs.

Module (Illumina, Inc.). Samples in the HKHS were genotyped calling was performed using the Illumina BeadStudio Genotyping phases I and II of the HapMap Project [78,79,80,81]. Genotype genotyping platform contains 560,186 tagging SNPs selected from Duo BeadChip (Illumina, Inc., San Diego, CA, USA). This SNP Genetics (Reykjavik, Iceland) with the Illumina HumanHap550-NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, DE, USA).

Gene expression experiments were performed by Genetech Biotech (Taipei, Taiwan) with the Illumina HT12 Expression BeadChip (Illumina, Inc.). This gene expression platform contains more than 25,000 annotated genes and more than 48,000 probes derived from the NCBI RefSeq (Build 36.2, Rel 22) and the UniGene (Build 199) databases. We also analyzed the gene expression data sets of genetically hypertensive and normotensive inbred mouse strains. Tissues of aorta, heart, kidney and liver from the two mouse strains were collected. The details about tissue collection, RNA preparation, and microarray experiment was described previously [82].

DNA samples
This TWNHS analyzed 400 matched pairs of patients with hypertension and normotensive controls. Patients were recruited from the Academia Sinica Multi-Center YOH Genetic Study. A one-to-one match strategy for age (±5 years) and gender was applied to select 400 normal controls from the Taiwan Han Chinese Cell and Genome Bank [77]. Immortalized cell lines from their lymphocytes have been established. The criteria for sample inclusion/exclusion and sample matching are described in our previous GWAS [29]. Genetic association replication analysis was performed using data from the other Han Chinese population in Hong Kong (HKHS) [36] and a Caucasian population (WTCCCHS) [17]. The HKHS collected 111 families with a total of 315 individuals, including 37 affected sib pairs and 195 discordant sib pairs, SBP, DBP, MBP off treatment, and disease status of hypertension were measured. The WTCCCHS collected 1,999 patients with hypertension and 3,004 normal controls consisting of 1,504 samples from the 1958 British Birth Cohort and 1,500 samples from the UK Blood Service Control Group [17].

SNP genotyping experiments
The genomic DNA of 800 Han Chinese samples was isolated from leukocytes using a Puregene kit (Gentra Systems, Minneapolis, MN, USA) for genomic DNA isolation. The DNA concentration was quantified and adjusted to 60 ng/μl using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, DE, USA). All samples were genotyped by deCODE Genetics (Reykjavik, Iceland) with the Illumina HumanHap550-Duo BeadChip (Illumina, Inc., San Diego, CA, USA). This SNP genotyping platform contains 560,186 tagging SNPs selected from phases I and II of the HapMap Project [78,79,80,81]. Genotype calling was performed using the Illumina BeadStudio Genotyping Module (Illumina, Inc.). Samples in the HKHS were genotyped with the Illumina HumanHap610-quad BeadChip (Illumina, Inc.), which enables the whole-genome genotyping of 620,901 SNPs. Samples in the WTCCCHS were genotyped with the Affymetrix Human Mapping 500 K Set (Affymetrix, Inc., San Diego, CA, USA), which provides 500,568 SNPs. The genotyping experiment was described previously [17].

Gene expression microarray experiments
A total of 12 matched pairs of patients with hypertension with the highest risk scores and normotensive controls with the lowest risk scores were selected. Total RNA was extracted using TRIzol Reagent (Life Technologies, Inc., CA, USA) and then reverse transcribed to cDNA and cRNA using Illumina TotalPrep™ RNA Amplification Kit (Illumina, Inc.). The cRNA concentration was quantified and adjusted to 150 ng/μl using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, DE, USA). Gene expression experiments were performed by Genetech Biotech (Taipei, Taiwan) with the Illumina HT12 Expression

Statistical analyses
A flowchart for the overall study design is provided (Figure 1). Quality of samples and SNPs were examined and population stratification in the TWNHS was evaluated. Genome-wide SBAS and GBAS were performed and confirmed by microarray gene expression analyses for the Han Chinese population of Taiwan. The identified genes were further examined by two association studies, the HKHS and the WTCCCHS. The detailed analysis procedures are as follows.

First, sample and SNP quality were examined. We checked the consistency of self-reported gender and X chromosome, calculated the GCR, estimated MAF using an allele counting approach, and examined HWE using an exact HWE test [83] with SAS GENETICS (SAS Institute, Inc., Cary, NC, USA) for all 546,376 autosomal SNPs from 560,184 tag SNPs probed on the HumanHap550-Duo BeadChip. Second, population stratification and admixture were examined. We evaluated population structure using a genomic control method [84] with SAS GENETICS (SAS Institute, Inc.) and a principal component method with EigenStrat software [85]. A non-parametric median test was used to examine differences of principal components in case and control groups [86].

Third, genome-wide SBAS was performed based on conditional logistic regression analyses using SAS STAT (SAS Institute, Inc.). The analysis regressed a dichotomous response on a nominal SNP coding for a genotype-type analysis or a continuous SNP coding for a trend-based analysis at each SNP locus with or without a covariate adjustment for BMI, resulting in four models: (1) conditional logistic regression with a nominal genotype without BMI adjustment (“CLRN”), (2) conditional logistic regression with a continuous genotype without BMI adjustment (“CLRC”), (3) conditional logistic regression with a nominal genotype and BMI adjustment (“CLRN,BMI”), and (4) conditional logistic regression with a continuous genotype and BMI adjustment (“CLRC,BMI”). For each SNP locus, p-values of a one degree-of-freedom Wald test were calculated for model CLRC and CLRC,BMI, and p-values of a two degree-of-freedom Wald test were calculated for model CLRN and CLRN,BMI.

Fourth, genome-wide GBAS was performed. According to SNP- and gene-annotation provided by the SNPer database [87], all 475,157 validated SNPs were divided into 216,994 intragenic SNPs on 15,373 genes and 258,163 intergenic SNPs. If a SNP mapped to more than one gene, it was assigned randomly. The mean, median, and standard deviation of the number of SNPs within a gene were 14.12, 6.00, and 32.19, respectively. In summary, 1,688 genes (10.98%), 9,000 genes (58.57%), 4,387 genes (28.55%), 291 genes (1.89%), and 1 gene (0.01%) contained 1, 2–10, 101–1000, and >1000 SNPs; CSMD1 on chromosome 8 contained the most SNPs (1,518) of any gene.

In the first step [prioritization test] of GBAS, we prioritized genes using a more stringent criterion (pFDR<10^-3) to reduce the potential inflated significance. For each of the 15,373 genes, with a specific regression model, p-values of intragenic SNPs within the same gene were combined by calculating a truncated product p-
value test statistic [88,89,90] with a p-value truncation threshold of 0.05 to represent a gene [35]. An FDR [91] was calculated using SAS GENETICS (SAS Institute, Inc.) to account for a multiple test correction. Genome-wide statistical significance of gene-based association tests was evaluated by adjusting for multiple tests of all genes using a stringent threshold, pFDR<10^{-5}. The identified candidate genes were further validated in the next step by using a permutation test for a stringent consideration. In the second step (permutation test) of GBAS, we used the threshold conventionally adopted (pFDR<5×10^{-2}) to further remove the inflated significance in the first test and take p-value dependency into consideration. Then, 10,000 permutations that we permuted for disease status but that maintained the matched relationship between patients and controls were performed for calculation of empirical p-values. In each permutation sample, we determined the statistical significance of the identified genes using the same gene-based association test. Empirical p-values of each identified gene was calculated and adjusted by FDR, where FDR was calculated across four models (“CLR_n”, “CLR_R”, “CLR_N,BMI” and “CLR_C,BMI”) jointly. For association validation of the identified genes, statistical significance was claimed under a significance threshold of pFDR<5×10^{-2}. The details of our genome-wide GBAS method are described in Method S1.

Fifth, microarray gene expression analysis was performed. Based on the genes identified in the second step, we selected 12 patients and 12 controls with extremely discordant genotypic distributions between the patient and control groups and conducted a microarray gene expression study using the Illumina HT-12 Expression Beadchip (Illumina, Inc.). Genotypes of all SNPs in the genes identified in the second step were assigned as “risky genotypes” or “protective genotypes” if the genotype frequencies in case group were higher or lower than in control group, respectively. The average proportion of risky genotypes in every patient and average proportion of protective genotypes in every control were calculated for every control over the identified genes. The 12 patients who carried the maximum proportion of risky genotypes and the 12 controls who carried the maximum proportion of protective genotypes among all samples were selected. The details of the selection of extremely discordant case-control pairs are described in Method S2. Quantile normalization [92] was applied to normalize gene expression, and analysis of covariance (ANCOVA) with an adjustment for BMI using PARTEK software (Partek, Inc., St. Louis, Missouri, USA) was performed to analyze the normalized data and identify the differentially expressed genes among the 100 identified genes. Confirmation of genetic association was claimed if \(p_{\text{ANCOVA}}<5×10^{-2}\). The identified differentially expressed genes were used to perform a cluster analysis with an average linkage in which distances of any two samples were measured by Kendall’s dissimilarity [93]. Except for two genes, KIAA1797 and AKAP5, not available in the mouse gene expression study [82], analysis of variance (ANOVA) using SAS STAT (SAS Institute, Inc.) was performed to discuss tissue-specific variation in gene expression, variance (ANOVA) using SAS STAT (SAS Institute, Inc.) was performed to discuss tissue-specific variation in gene expression, and four phenotypes from the HKHS GWAS. The four phenotypes were: (1) dichotomous disease status, (2) quantitative SBP, (3) quantitative DBP, and (4) quantitative MBP. A single-locus sib-based transmission disequilibrium test (sib-TDT) [96] was applied to examine the genetic association of the dichotomous disease status of hypertension with the SNPs. A quantitative family-based association test (QFAM) [97] was applied to examine the genetic association of quantitative SBP, DBP, MBP, and genes (“QFAM_SBP”, “QFAM_DBP”, and “QFAM_MBP”) by calculating p-values for a within-family component. All the analyses were done with PLINK software [97]. Moreover, we permuted the disease status to obtain 10,000 permutation samples for calculation of empirical p-values. In each permutation sample, we examined the statistical significance of the identified genes using the same gene-based association test described in the fourth step. The empirical p-value of each identified gene was calculated and adjusted by FDR, where FDR was calculated across four phenotypic models (“sib-TDT”, “QFAM_SBP”, “QFAM_DBP”, and “QFAM_MBP”) jointly.

Replication of genetic association was claimed if pFDR<5×10^{-2}. We also analyzed genotype data from the WTCCCHS [17]. Single-locus association tests were performed using logistic regression analyses, which regressed a dichotomous response on continuous or nominal-coding genotypes at each SNP locus with or without covariate adjustment for gender, resulting in four logistic regression models: (1) logistic regression with a nominal genotype without gender adjustment (“LRC_n”), (2) logistic regression with a continuous genotype without gender adjustment (“LRC_c”), (3) logistic regression with a nominal genotype with gender adjustment (“LRC_N,Gender”), and (4) logistic regression with a continuous genotype with gender adjustment (“LRC_C,Gender”). Again, we permuted the disease status to obtain 10,000 permutation samples for calculation of empirical p-values. In each permutation sample, we determined the statistical significance of the identified genes using the same gene-based association test described in the fourth step. The empirical p-value of each identified gene was calculated and adjusted by FDR, where FDR was calculated across four models (“LRC_n”, “LRC_c”, “LRC_N,Gender”, and “LRC_C,Gender”) jointly. Replication of genetic association was claimed if pFDR<5×10^{-2}. Note that 13 genes out of 100 were not examined in the WTCCCHS because the Affymetrix Human Mapping 500 K Set does not provide SNPs on these 13 genes.

Bioinformatics analyses

We used the Genetic Association Database (GAD, http://geneticsassociationdb.nih.gov/cgi-bin/index.cgi) and Online Mendelian Inheritance in Man (OMIM, http://www.ncbi.nlm.nih.gov/omim/) to determine whether the identified genes with differential allelic associations and differential expression in studies (Table 1); Spearman’s correlation coefficient and correlation test of gene expressions and continuous SNP coding were calculated. cis-eQTL with p<5×10^{-2} were selected. Furthermore, we examined if the cis-eQTL were also associated with YOH in the TWNHS; conditional logistic regression at each cis-eQTL was performed with or without a covariate adjustment for BMI by using SAS STAT (SAS Institute, Inc.), where a continuous SNP coding was considered. YOH-associated cis-eQTL with p<5×10^{-2} were selected. Finally, we examined the direction of genetic effect (risk effect or protective effect) and gene regulation (up-regulation or down-regulation) of the disease associated alleles of cis-eQTL.

Finally, we conducted GBAS for genomic data sets of a Han Chinese population of Hong Kong and a Caucasian population. We analyzed genotype data of SNPs on the 100 identified genes and four phenotypes from the HKHS GWAS. The four phenotypes were: (1) dichotomous disease status, (2) quantitative SBP, (3) quantitative DBP, and (4) quantitative MBP. A single-locus sib-based transmission disequilibrium test (sib-TDT) [96] was applied to examine the genetic association of the dichotomous disease status of hypertension with the SNPs. A quantitative family-based association test (QFAM) [97] was applied to examine the genetic association of quantitative SBP, DBP, MBP, and genes (“QFAM_SBP”, “QFAM_DBP”, and “QFAM_MBP”) by calculating p-values for a within-family component. All the analyses were done with PLINK software [97]. Moreover, we permuted the disease status to obtain 10,000 permutation samples for calculation of empirical p-values. In each permutation sample, we examined the statistical significance of the identified genes using the same gene-based association test described in the fourth step. The empirical p-value of each identified gene was calculated and adjusted by FDR, where FDR was calculated across four phenotypic models (“sib-TDT”, “QFAM_SBP”, “QFAM_DBP”, and “QFAM_MBP”) jointly.

Replication of genetic association was claimed if pFDR<5×10^{-2}. We also analyzed genotype data from the WTCCCHS [17]. Single-locus association tests were performed using logistic regression analyses, which regressed a dichotomous response on continuous or nominal-coding genotypes at each SNP locus with or without covariate adjustment for gender, resulting in four logistic regression models: (1) logistic regression with a nominal genotype without gender adjustment (“LRC_n”), (2) logistic regression with a continuous genotype without gender adjustment (“LRC_c”), (3) logistic regression with a nominal genotype with gender adjustment (“LRC_N,Gender”), and (4) logistic regression with a continuous genotype with gender adjustment (“LRC_C,Gender”). Again, we permuted the disease status to obtain 10,000 permutation samples for calculation of empirical p-values. In each permutation sample, we determined the statistical significance of the identified genes using the same gene-based association test described in the fourth step. The empirical p-value of each identified gene was calculated and adjusted by FDR, where FDR was calculated across four models (“LRC_n”, “LRC_c”, “LRC_N,Gender”, and “LRC_C,Gender”) jointly. Replication of genetic association was claimed if pFDR<5×10^{-2}. Note that 13 genes out of 100 were not examined in the WTCCCHS because the Affymetrix Human Mapping 500 K Set does not provide SNPs on these 13 genes.

Bioinformatics analyses

We used the Genetic Association Database (GAD, http://geneticsassociationdb.nih.gov/cgi-bin/index.cgi) and Online Mendelian Inheritance in Man (OMIM, http://www.ncbi.nlm.nih.gov/omim/) to determine whether the identified genes with differential allelic associations and differential expression in
patients with hypertension and normotensive controls were also reported. In addition, we used Mouse Genome Informatics (MGI, http://www.informatics.jax.org/) to determine if the identified genes were involved in hypertension-related phenotypes in mice and other mammals.

Supporting Information

Figure S1 Principal component analysis for evaluation of the population substructure in patient and control groups. Hypertensive patients (symbol: ○) and normotensive controls (symbol: ×) are projected onto a two-dimensional plane defined by the first two eigenvectors. (DOC)

Figure S2 P-values (−log10 scale) of genome-wide SBAS. (A) P-values of the conditional logistic regression with a nominal genotype without BMI adjustment (CLR_N). (B) P-values of the conditional logistic regression with a nominal genotype with BMI adjustment (CLR_N,BMI). (C) P-values of the conditional logistic regression with a continuous genotype without BMI adjustment (CLR_C). (D) P-values of the conditional logistic regression with a continuous genotype with BMI adjustment (CLR_C,BMI). In each figure, the vertical axis is the raw p-values (−log10 scale) of gene-based association tests, and the horizontal axis is cumulative physical position (Mb scale). (DOC)

Table S1 Raw p-values (−log10 scale) of 101 genes identified by the genome-wide GBAS. Gene name, number of SNPs, chromosome, physical position, and category (the 101 genes are classified into 12 categories (Category I to XII) according to gene significance in the four genome-wide GBAS models) are shown. The raw p-values in −log10 scale are shown for our GBAS of the TWNHS (Method: CLR_N, CLR_C, CLR_N,BMI, and CLR_C,BMI) and gene expression analysis (Method: ANCOVA) in order. Raw p-values that reached pFDR<5×10−2 in each model are marked in bold. (XLS)

Table S2 The information for the 101 identified hypertension susceptibility genes collected from the GAD, OMIM, and MGI databases. (XLS)

Table S3 Beta coefficients and their standard errors for the 17 differentially expressed genes in hypertension. (DOC)

Table S4 The cis-eQTL analysis of 17 genes in Table 2 identified in the present study. Gene name followed by chromosome number, RS number of cis-eQTL, beta coefficient in CLR_C, beta coefficient in CLR_C,BMI, p-values of the tests of beta coefficients, the maximal Spearman correlation coefficient from Geneva GenCord and MuTHER, p-value of the test of Spearman correlation, and the first and the second allele followed by an index for the direction of genetic effect and gene regulation of the YOH-associated risky alleles in the cis-eQTL (r/u, r/d, p/u and p/d denote that risky/up-regulated, risky/down-regulated, protective/up-regulated and protective/down-regulated, respectively). “NS” indicates that beta coefficient is not significant and p-value is not shown. “NC” indicates the positive/negative of Spearman correlation coefficients from Geneva GenCord and MuTHER is not consistent. (XLS)

Method S1 Genome-wide gene-based association tests. (DOC)

Method S2 Selection of extremely discordant case-control pairs. (DOC)

Acknowledgments

This study made use of data obtained from the Wellcome Trust Case Control Consortium. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk. We sincerely thank two anonymous reviewers for their very constructive and insightful comments that helped in preparing our revision.

Author Contributions

Conceived and designed the experiments: HCY WHP. Performed the experiments: JWC HYH CTT T. Lin SHS WCT JHC HBL WHY TYC CIC SJL. Analyzed the data: HCY YJL KMC CMC YG. Contributed reagents/materials/analysis tools: HCY WHP. Wrote the paper: HCY WHP.


