

Identification of *IGF1*, *SLC4A4*, *WWOX*, and *SFMBT1* as Hypertension Susceptibility Genes in Han Chinese with a Genome-Wide Gene-Based Association Study

Hsin-Chou Yang¹, Yu-Jen Liang^{1,3}, Jaw-Wen Chen⁴, Kuang-Mao Chiang^{2,5}, Chia-Min Chung², Hung-Yun Ho⁶, Chih-Tai Ting⁶, Tsung-Hsien Lin⁷, Sheng-Hsiung Sheu⁷, Wei-Chuan Tsai⁸, Jyh-Hong Chen⁸, Hsin-Bang Leu⁴, Wei-Hsian Yin⁹, Ting-Yu Chiu¹⁰, Ching-Iuan Chern¹⁰, Shing-Jong Lin⁴, Brian Tomlinson¹¹, Youling Guo¹², Pak C. Sham^{12,13}, Stacey S. Cherny^{12,13}, Tai Hing Lam¹⁴, G. Neil Thomas¹⁵, Wen-Harn Pan^{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, *IGF1*, *SLC4A4*, *WWOX*, and *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. *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. *SLC4A4*, which encodes the electrogenic sodium bicarbonate cotransporter 1, is associated with decreased body weight/size and abnormal ion homeostasis in mice. *WWOX*, which encodes the WW domain-containing protein, is related to hypoglycemia and hyperphosphatemia. *SFMBT1*, which encodes the scm-like with four MBT domains protein 1, is a novel hypertension gene. *GRB14*, *TMEM56* and *KIAA1797* exhibited highly significant differential allelic and expressed distributions between hypertensive patients and normotensive controls. *GRB14* was also found relevant to blood pressure in a previous genetic association study in East Asian populations. *TMEM56* and *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.

Citation: Yang H-C, Liang Y-J, Chen J-W, Chiang K-M, Chung C-M, et al. (2012) Identification of *IGF1*, *SLC4A4*, *WWOX*, and *SFMBT1* as Hypertension Susceptibility Genes in Han Chinese with a Genome-Wide Gene-Based Association Study. PLoS ONE 7(3): e32907. doi:10.1371/journal.pone.0032907

Editor: Leonardo Mariño-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-001-029), National Science Council grants (NSC 97-2314-B-001-006-MY3, NSC 100-2314-B-001-005-MY3, and NSC 100-2319-B-010-002) of 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 complex 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 *ECE* (1p36.1) [4], *AGT* (1q42-43) [5,6], *AGTR1A* (3q21-q25) [7], *HYT1* (17q) [8,9], *HYT2* (15q) [9], *HYT3* (2p25-p24) [10], *HYT4* (12p) [11], *HYT5* (20q), *HYT6* (5p) [12], *HYT7* (3p14.1-q12.3) [13], and *HYT8* (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 [35], 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, 0.435, 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 “CLR_N”, “CLR_C”, “CLR_{N,BMI}”, and “CLR_{C,BMI}”, 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 without or with a covariate adjustment for body mass index (BMI). No significant SNPs ($pFDR < 10^{-3}$ or $p < 10^{-7}$) 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 *KIAA1797* between 20,638,805 bp and 20,993,372 bp on chromosome 9p21 contains 84 SNPs probed on the HumanHap550-Duo BeadChip. The proportions of SNPs with marginal significance ($p < 5 \times 10^{-2}$) in the four conditional logistic regression analyses (CLR_N, CLR_C, CLR_{N,BMI}, and CLR_{C,BMI}) were 10%, 26%, 20%, and 23%, with the smallest values of p-values on *KIAA1797* of 7.56×10^{-4} , 7.38×10^{-4} , 4.31×10^{-5} , and 7.69×10^{-6} , 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.

Ref	Population	Analysis (n = sample size)	Blood pressure or hypertension- associated genes
[19]	Multi-ethnicity	Single-stage: GWAS (n = 1,260, 1,233 and 1,327 individuals in different analyses)	<i>CCL20</i> , <i>WDR69</i> , <i>CDH13</i> , <i>LPP</i>
[20]	Asian ancestry – Japan	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)	<i>ADD2</i> , <i>KIAA0789</i> , <i>M6PR</i>
[22]	Amish Caucasian	Stage 1: GWAS (n = 542 Amish Caucasians); Stage 2: Replication (n = 1,367 Amish Caucasians and 5,804 non-Amish Caucasians)	<i>STK39</i>
[23]	European ancestry	Stage 1: GWAS (n = 34,433); Stage 2: Replication (n = 71,225 subjects of European ancestry and 12,889 subjects of Indian Asian ancestry)	<i>CYP17A1</i> , <i>CYP17A2</i> , <i>FGF5</i> , <i>SH2B3</i> , <i>MTHFR</i> , <i>C10orf107</i> , <i>ZNF652</i> , <i>PLCD3</i>
[24]	European ancestry – Germany	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)	<i>CDH13</i>
[25]	Asian ancestry – Korea	Stage 1: GWAS (n = 8,842 Koreans); Stage 2: Replication (n = 7,861 Koreans)	<i>ATP2B1</i>
[26]	African ancestry – America	Stage 1: GWAS (n = 509 African American cases+508 controls); Stage 2: Replication (n = 366 West African cases+614 controls)	<i>SLC24A4</i> , <i>CACNA1H</i>
[27]	European ancestry	Stage 1: GWAS (n = 29,136 European descent subjects); Stage 2: Replication (n = 34,433 European descent subjects)	<i>ATP2B1</i> , <i>CYP17A1</i> , <i>PLEKHA7</i> , <i>SH2B3</i> , <i>CACNB2</i> , <i>CSK</i> ~ <i>ULK3</i> , <i>TBX3</i> ~ <i>TBX5</i> , <i>ULK4</i>
[28]	Asian ancestry	Stage 1: GWAS (n = 19,608 East Asian ancestry subjects); Stage 2: Replication (n = 10,518 East Asian ancestry subjects)	<i>ST7L</i> ~ <i>CAPZA1</i> , <i>FIGN</i> ~ <i>GRB14</i> , <i>ENPEP</i> , <i>NPR3</i> , <i>TBX3</i> , <i>ALDH2</i>
[29]	Asian ancestry – Taiwan	Stage 1: GWAS (n = 175 cases+175 controls); Stage 2: Replication (n = 833 cases+833 controls)	<i>LOC344371</i> , <i>MYADM1</i> , <i>FAM98A</i> , <i>RASGRP3</i> , <i>IMPG1</i>
[30]	African ancestry – America	Stage 1: GWAS (n = 8,591 African Americans); Stage 2: Replication (n = 11,882 African Americans); Stage 3: Replication (n = 69,899 European Americans)	<i>GPR98</i> ~ <i>ARRDC3</i> , <i>C21orf91</i> , <i>SLC25A42</i> , <i>HLA-B</i>
[31]	European ancestry – Sweden	Stage 1: GWAS (n = 1,621 cases+1,699 controls); Stage 2: Replication (n = 19,845 cases+16,541 controls)	<i>UMOD</i>
[32]	European ancestry – Great Britain	Single-stage: GWAS (n = 2,000 cases+3,000 controls)	<i>GPR39</i> , <i>XRCC4</i> , <i>MYO6</i> , <i>ZFAT</i> , <i>MACROD2</i>
[33]	Asian ancestry – Korea	Stage 1: GWAS (n = 7,551 Koreans); Stage 2: Replication (n = 3,703 Koreans)	<i>AKAP13</i>
[34]	Asian ancestry – Japan	Stage 1: GWAS (n = 936 Japanese); Stage 2: Replication (n = 3,228 Japanese); Stage 3: Replication (n = 2,895 Japanese)	<i>CCBE1</i> , <i>MAP7</i> , <i>ZFP64</i> , <i>PCDH18</i> , <i>CDH2</i> , <i>WWOX</i> ^a

^aGenes (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 ($pFDR < 10^{-3}$) 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 replications. 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 *FMO5*) were successfully validated ($pFDR < 5 \times 10^{-2}$).

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 ($p < 5 \times 10^{-2}$) (**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

Analysis flow

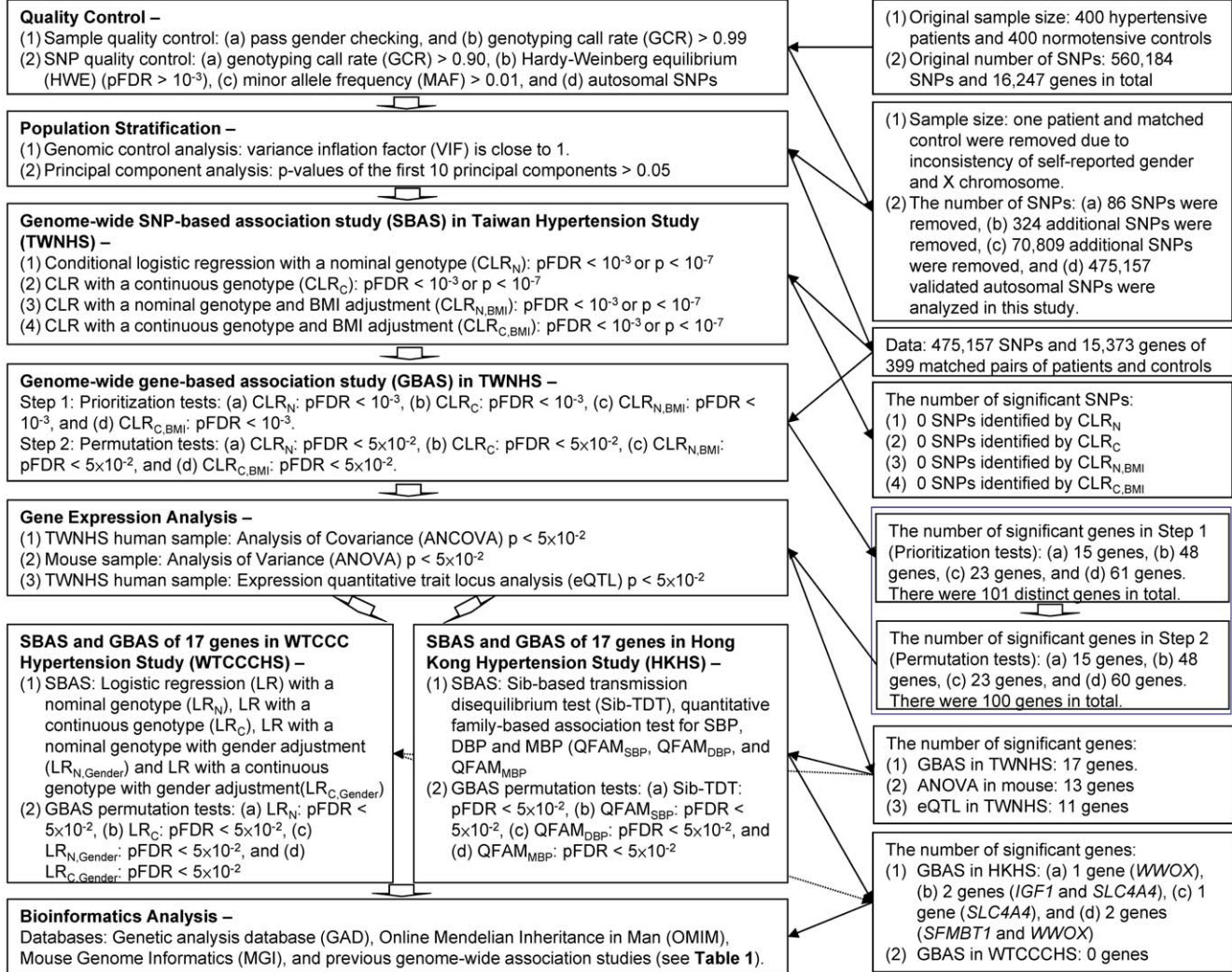


Figure 1. An analysis flowchart of this study.

doi:10.1371/journal.pone.0032907.g001

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 ($p < 5 \times 10^{-2}$) (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 ($p < 5 \times 10^{-2}$) 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*, *KIAA1797*, *LARS*, *GRB14*, *IGF1*, *WVVOX*, *HLC5*, *CSF1*, *CSNK1G3*, *TTC12* and *SYNE2*) associated with YOH in our TWNHS study ($p < 5 \times 10^{-2}$), where the disease associated alleles of 40 cis-eQTL increased genetic risk of YOH as well as up-regulated gene expression (Table S4). In addition to the

17 genes in Table 2, we also examined cis-eQTL of the hypertension-associated genes identified by previous hypertension genomic studies [16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34] in Table 1 where *GRB14* and *WVVOX* were excluded here because they were also identified by the present study and their cis-eQTL were already examined. Cis-eQTL of 54 genes were studied, although 6 of these genes (*CYP17A2*, *HLA-B*, *KIAA0789*, *LOC344371*, *MYADML* and *ULK4*) did not have cis-eQTL in Genevar database [37]. The gene-centric cis-eQTL analysis found 1,902 cis-eQTL ($p < 5 \times 10^{-2}$) in 2-Mb windows around the 48 genes, where 1,628 cis-eQTL on 33 genes were interrogated in the Illumina HumanHap550-Duo BeadChip. Finally, 31 cis-eQTL on 9 genes (*AKAP13*, *CACNB2*, *CCBE1*, *CDH13*, *CSK*, *FIGN*, *LPP*, *SLC24A4* and *TBX3*) were associated with YOH in our TWNHS study ($p < 5 \times 10^{-2}$), 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”, “QFAM_{SBP}”, “QFAM_{DBP}”, and “QFAM_{MBP}” for analyzing dichotomous disease status and three quantitative traits

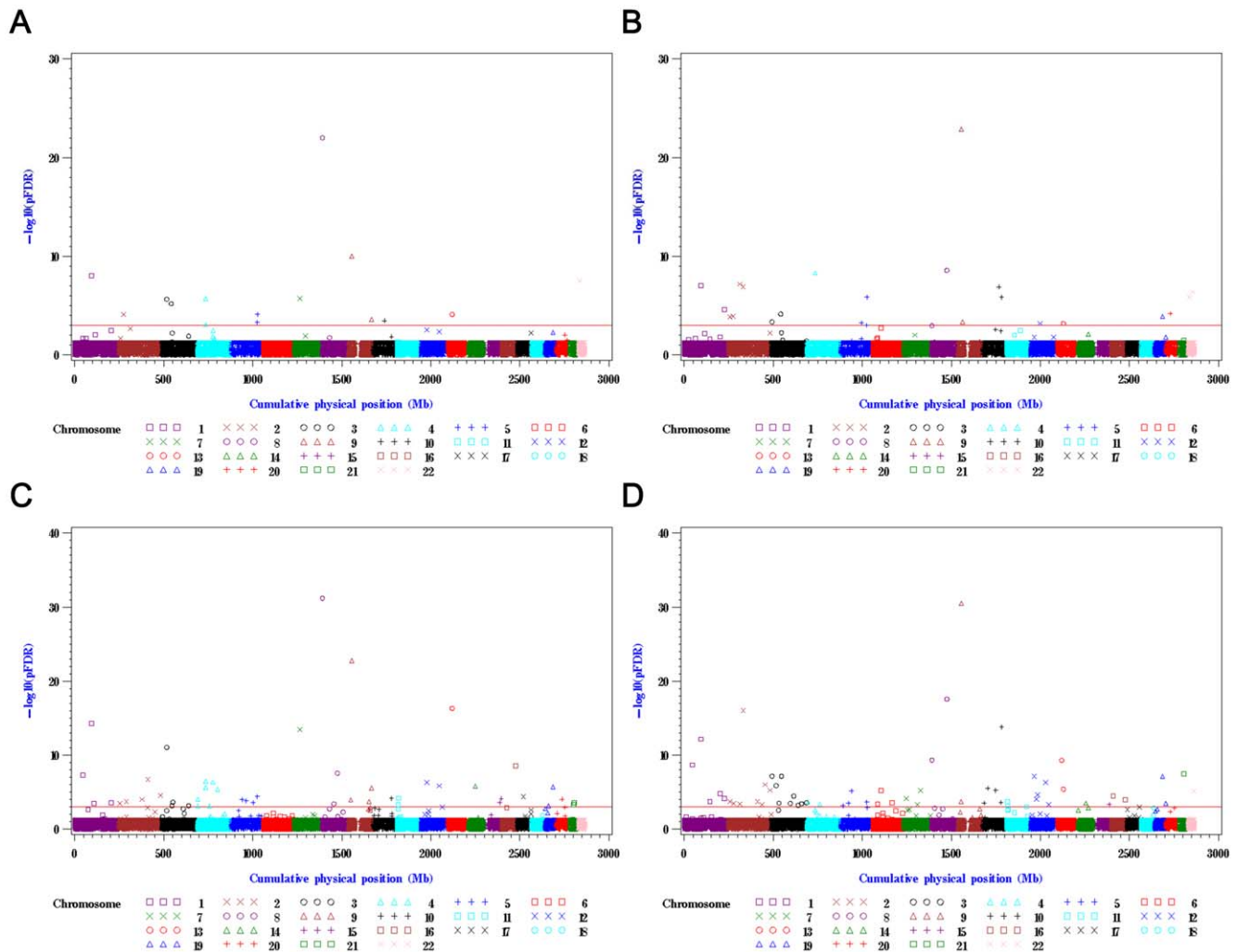


Figure 2. Gene-based association tests based on p-values from the four conditional logistic regression models. (A) model CLR_N , (B) model $CLR_{N,BMI}$, (C) model CLR_C , and (D) model $CLR_{C,BMI}$. In each figure, the vertical axis is the FDR-adjusted p-values ($-\log_{10}$ scale) of gene-based association tests, and the horizontal axis is cumulative physical position (Mb scale). The red reference line signifies $pFDR = 10^{-3}$. doi:10.1371/journal.pone.0032907.g002

(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 “ LR_N ”, “ LR_C ”, “ $LR_{N,Gender}$ ”, and “ $LR_{C,Gender}$ ” 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 [$p_{CLR_N} = 3.10 \times 10^{-7}$ ($pFDR_{CLR_N} = 2.47 \times 10^{-2}$); $p_{CLR_C} = 6.97 \times 10^{-11}$ ($pFDR_{CLR_C} = 2.20 \times 10^{-2}$); $p_{CLR_{C,BMI}} = 9.90 \times 10^{-8}$ ($pFDR_{CLR_{C,BMI}} = 2.95 \times 10^{-2}$)] and gene expression analysis [$p_{PANCova} = 3.21 \times 10^{-3}$] but was also identified by a GBAS of the HKHS [$p_{QFAMSBP} = 1.21 \times 10^{-4}$ ($pFDR_{QFAMSBP} = 3.73 \times 10^{-2}$)]. Moreover, association between gene expression and cis-eQTL markers of *IGF1* was also found in our

gene-centric cis-eQTL analysis (**Table S4**). Two cis-eQTL of *IGF1* were found (rs10860862 and rs5742632) where the latter SNP is an intronic enhancer at a transcriptional factor binding site. *IGF1* is a well-documented gene that plays a role in cardiovascular disorders [38,39,40,41,42] and metabolic disorders [43,44,45,46,47,48,49,50,51,52]. *IGF1* variants are associated with decreased body weight/size, abnormal glucose homeostasis, increased/decreased circulating insulin levels, decreased circulating *IGF1* levels, and increased DBP/SBP values in mice [53,54,55].

SLC4A4, which encodes the electrogenic sodium bicarbonate cotransporter 1 on 4q21, was not only identified by our GBAS [$p_{CLR_{N,BMI}} = 3.81 \times 10^{-5}$ ($pFDR_{CLR_{N,BMI}} = 2.99 \times 10^{-2}$); $p_{CLR_{C,BMI}} = 7.26 \times 10^{-8}$ ($pFDR_{CLR_{C,BMI}} = 2.64 \times 10^{-2}$)] and gene expression analysis [$p_{PANCova} = 3.14 \times 10^{-2}$] but was also identified by a GBAS of the HKHS [$p_{QFAMSBP} = 9.19 \times 10^{-6}$ ($pFDR_{QFAMSBP} = 1.63 \times 10^{-2}$); $p_{QFAMDDBP} = 1.19 \times 10^{-3}$ ($pFDR_{QFAMDDBP} = 4.79 \times 10^{-2}$)]. 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].

WVVOX, which encodes the WW domain-containing protein on 16q23.3-q24.1, is the third gene that was simultaneously identified by our genome-wide GBAS [$p_{\text{CLR}_{\text{C}}}=7.47 \times 10^{-14}$ ($p\text{FDR}_{\text{CLR}_{\text{C}}}=5.77 \times 10^{-3}$) and $p_{\text{CLR}_{\text{C},\text{BMI}}}=1.18 \times 10^{-8}$ ($p\text{FDR}_{\text{CLR}_{\text{C},\text{BMI}}}=1.84 \times 10^{-2}$)], gene expression analysis [$p_{\text{ANCOVA}}=5.96 \times 10^{-3}$], and a GBAS of the HKHS [$p_{\text{TDT}}=1.05 \times 10^{-6}$ ($p_{\text{TDT}}=1.37 \times 10^{-2}$) and $p_{\text{QFAMMBP}}=8.82 \times 10^{-7}$ ($p_{\text{QFAMMBP}}=1.05 \times 10^{-2}$)]. Moreover, association between gene expression and cis-eQTL markers of *WVVOX* was also found in our gene-centric cis-eQTL analysis (Table S4). In total, 24 cis-eQTL were found, where 14 cis-eQTL are intronic enhancers at transcriptional factor binding sites, 1 cis-eQTL locates in *WVVOX* upstream, and the remaining 9 cis-eQTL are in an intronic region with no known function. *WVVOX* variants are associated with hypoglycemia, hyperphosphatemia, decreased circulating cholesterol levels and triglyceride levels, but with increased blood uric acid levels in mice [59]. *WVVOX* variants were also found to be associated with blood pressure in a previous Japanese GWAS [23].

SFMBT1, which encodes the drosophila sex comb on middle (scm)-like protein on 3p21.31 and contains four malignant brain tumor repeat domains, was simultaneously identified by our genome-wide GBAS [$p_{\text{CLR}_{\text{N}}}=1.82 \times 10^{-10}$ ($p\text{FDR}_{\text{CLR}_{\text{N}}}=2.47 \times 10^{-2}$) and $p_{\text{CLR}_{\text{N},\text{BMI}}}=3.59 \times 10^{-9}$ ($p\text{FDR}_{\text{CLR}_{\text{N},\text{BMI}}}=2.81 \times 10^{-2}$)], gene expression analysis [$p_{\text{ANCOVA}}=4.74 \times 10^{-2}$], and a GBAS of the HKHS [$p_{\text{QFAMMBP}}=1.36 \times 10^{-5}$ ($p\text{FDR}_{\text{QFAMMBP}}=3.44 \times 10^{-2}$)]. This gene may have a novel role in hypertension, and its role in disease etiology should be studied further.

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 hyperten-

sion susceptibility genes, *IGF1*, *SLC4A4*, *WVVOX*, and *SFMBT1*, 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 *WVVOX* 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. *SFMBT1* 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 ($p\text{FDR}_{\text{LRC}}=7.46 \times 10^{-2}$). *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, *TMEM56* on 1p21.3 and *KIAA1797* 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.

Gene	Raw p-values of GBAS of conditional logistic model ^{a,b}				Raw p-values of the human and mouse GE analysis ^{a,b}				
	CLR _N	CLR _C	CLR _{N,BMI}	CLR _{C,BMI}	Human	Mouse aorta	Mouse heart	Mouse kidney	Mouse liver
<i>TMEM56</i>	9.83E-14	6.78E-20	1.74E-12	1.19E-17	5.77E-03	3.81E-03	4.48E-02	5.31E-02	1.39E-02
<i>KIAA1797</i>	6.78E-16	1.13E-28	4.78E-29	1.03E-36	7.97E-03	NA ^c	NA ^c	NA ^c	NA ^c
<i>FMBT1</i>	1.82E-10	6.93E-01	3.59E-09	6.93E-01	4.74E-02	7.53E-03	7.18E-01	1.00E-01	6.60E-01
<i>LARS</i>	2.80E-09	1.44E-05	5.28E-11	2.24E-05	6.83E-03	8.54E-01	7.77E-02	6.03E-01	1.38E-01
<i>GRB14</i>	2.56E-04	7.04E-12	8.64E-04	2.50E-08	7.15E-03	5.04E-01	3.79E-02	3.57E-02	2.21E-01
<i>IGF1</i>	3.10E-07	6.97E-11	3.10E-04	9.90E-08	3.21E-03	3.29E-02	5.70E-02	1.09E-01	1.71E-01
<i>FURIN</i>	3.62E-05	3.34E-08	4.76E-05	8.66E-08	2.26E-03	4.96E-01	1.72E-01	6.05E-01	6.51E-01
<i>WVVOX</i>	1.70E-02	7.47E-14	2.23E-02	1.18E-08	5.96E-03	2.19E-02	9.94E-02	2.37E-01	4.22E-03
<i>HLC5</i>	5.88E-05	3.51E-08	5.15E-06	1.09E-12	1.38E-02	4.02E-02	1.28E-02	1.66E-01	1.90E-03
<i>CSF1</i>	5.66E-05	4.67E-08	8.36E-02	9.98E-06	2.76E-02	6.44E-03	7.60E-01	1.54E-02	3.20E-01
<i>CSNK1G3</i>	3.23E-02	3.39E-08	3.82E-01	1.06E-01	2.71E-03	1.10E-03	4.99E-03	2.85E-05	1.00E-01
<i>COMMD7</i>	2.50E-04	8.74E-09	1.18E-01	6.81E-04	2.62E-02	4.48E-02	6.55E-01	5.31E-01	7.99E-01
<i>AP3S1</i>	1.38E-02	5.17E-02	3.92E-08	3.22E-02	8.03E-03	NA ^c	NA ^c	NA ^c	NA ^c
<i>LC4A4</i>	7.58E-01	1.90E-02	3.81E-05	7.26E-08	3.14E-02	3.32E-04	1.62E-03	4.42E-02	2.69E-01
<i>LACE1</i>	5.37E-01	4.82E-01	1.46E-03	4.24E-08	1.86E-02	4.14E-03	6.17E-01	2.96E-01	5.80E-01
<i>TTC12</i>	1.85E-03	1.01E-03	1.23E-02	2.03E-07	2.33E-02	4.69E-02	5.78E-07	2.87E-01	6.00E-01
<i>SYNE2</i>	9.62E-01	6.71E-01	2.63E-01	4.52E-08	1.61E-02	1.55E-01	4.71E-01	2.48E-02	2.18E-02

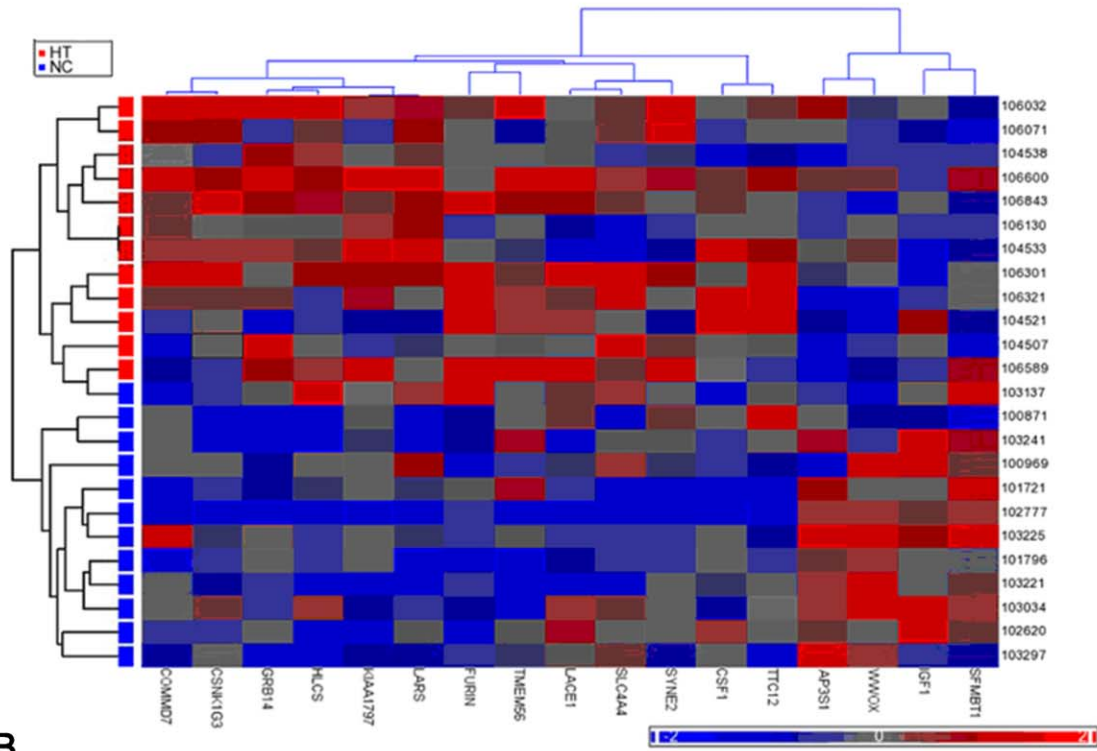
^aTest results include p-values of the gene-based association analysis (CLR_N, CLR_C, CLR_{N,BMI}, and CLR_{C,BMI}) and $-\log_{10}(p)$ of gene expression [GE] analysis.

^bRaw 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.

^cNA denotes that expression data of a gene were not available in the mouse gene expression analysis.

doi:10.1371/journal.pone.0032907.t002

A



B

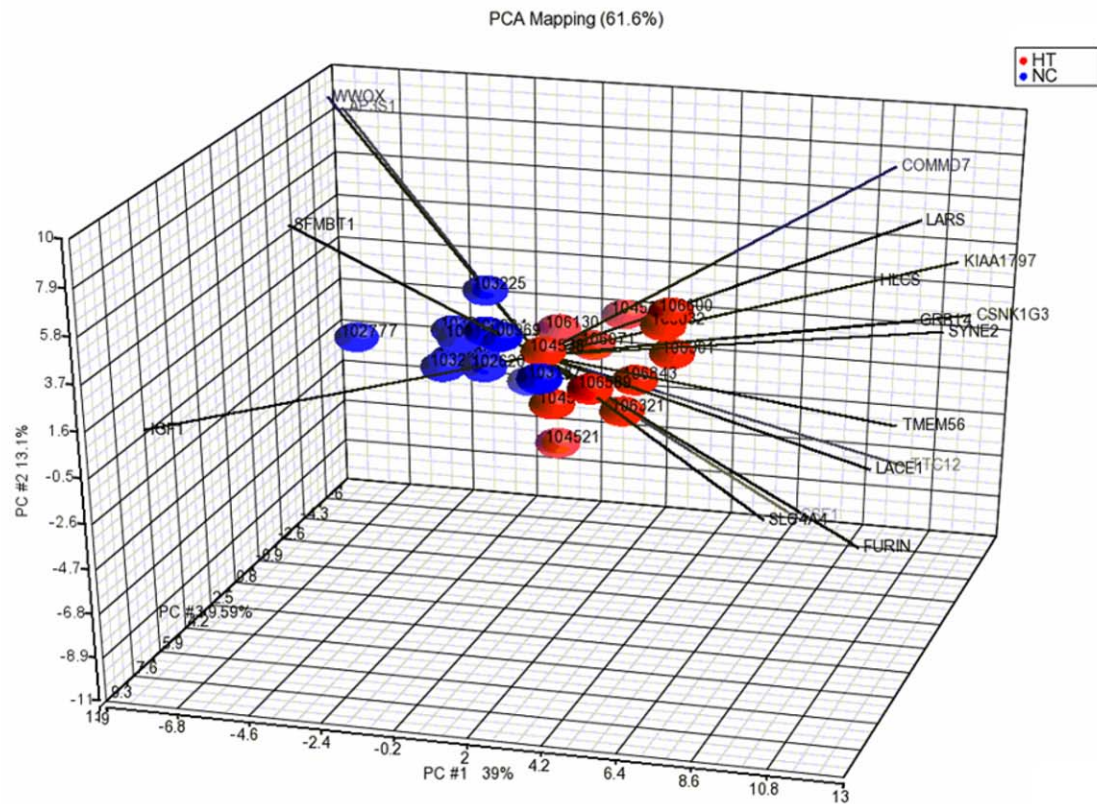


Figure 3. Evaluation of the 17 differentially expressed genes in hypertension. (A) Cluster analysis of transcriptional data from 12 pairs of patients with hypertension (HT) and normotensive controls (NC). (B) Principal component analysis (PCA) of transcriptional data from 12 pairs of patients with hypertension and normotensive controls, where the variation explained by the first two principal components (PC #1 and PC #2) is 39% and 13.1%.

doi:10.1371/journal.pone.0032907.g003

respectively. These two genes were validated by our human gene expression analysis [*TMEM56*: $p_{\text{ANCOVA}} = 5.77 \times 10^{-3}$; *KIAA1797*: $p_{\text{ANCOVA}} = 7.97 \times 10^{-3}$] and *TMEM56* was also validated by the mice gene expression analysis [$p_{\text{ANOVAaorta}} = 3.81 \times 10^{-3}$; $p_{\text{ANOVAheart}} = 4.48 \times 10^{-2}$; $p_{\text{ANOVA liver}} = 1.39 \times 10^{-2}$] (**Table 2**). In addition, *FURIN* on 15q26.1, which encodes a subtilisin-like pro-protein convertase family protein [GBAS: $p_{\text{CLRn}} = 3.62 \times 10^{-5}$ ($p_{\text{FDR}_{\text{CLRn}}} = 2.04 \times 10^{-2}$); $p_{\text{CLRc}} = 3.34 \times 10^{-8}$ ($p_{\text{FDR}_{\text{CLRc}}} = 1.52 \times 10^{-2}$); $p_{\text{CLRn,BMI}} = 4.76 \times 10^{-5}$ ($p_{\text{FDR}_{\text{CLRn,BMI}}} = 2.16 \times 10^{-2}$); $p_{\text{CLRc,BMI}} = 8.66 \times 10^{-8}$ ($p_{\text{FDR}_{\text{CLRc,BMI}}} = 1.52 \times 10^{-2}$); gene expression: $p_{\text{ANCOVA}} = 2.26 \times 10^{-3}$] (**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 WTCCCHS.

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*, *KIAA1797*, *LARS*, *GRB14*, *IGF1*, *WVWX*, *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.

Gene	Raw p-values of GBAS of the HKHS ^{a,b}				Raw p-values of GBAS of the WTCCCHS ^{a,b}			
	sTDT	QFAM _{SBP}	QFAM _{DBP}	QFAM _{MBP}	LR _N	LR _C	LR _{N,Gender}	LR _{C,Gender}
<i>TMEM56</i>	5.60E-01	5.60E-01	5.60E-01	5.60E-01	8.06E-01	7.15E-01	7.69E-01	6.61E-01
<i>KIAA1797</i>	9.75E-01	9.86E-01	4.84E-01	9.86E-01	7.99E-01	9.34E-01	6.58E-01	9.34E-01
<i>SFMBT1</i>	3.04E-01	6.93E-01	2.91E-01	1.36E-05	5.12E-01	4.83E-01	5.12E-01	5.12E-01
<i>LARS</i>	2.26E-01	2.26E-01	2.26E-01	2.26E-01	3.37E-01	3.37E-01	3.37E-01	3.27E-01
<i>GRB14</i>	6.93E-01	6.93E-01	6.93E-01	6.93E-01	8.58E-01	8.58E-01	8.58E-01	8.58E-01
<i>IGF1</i>	2.02E-01	1.21E-04	6.03E-01	6.03E-01	8.06E-01	7.46E-01	7.84E-01	8.51E-01
<i>FURIN</i>	2.26E-01	2.26E-01	2.26E-01	2.26E-01	5.00E-02	5.00E-02	5.00E-02	5.00E-02
<i>WVWX</i>	1.05E-06	1.11E-03	1.54E-02	8.82E-07	3.83E-01	8.68E-01	3.85E-01	9.35E-01
<i>HLCs</i>	6.93E-01	5.08E-01	9.29E-01	9.34E-01	8.06E-01	4.36E-01	7.64E-01	6.16E-01
<i>CSF1</i>	4.31E-01	4.31E-01	4.26E-01	3.88E-01	5.66E-02	1.98E-02	6.00E-02	2.15E-02
<i>CSNK1G3</i>	5.82E-01	5.82E-01	5.82E-01	5.82E-01	1.44E-02	3.22E-01	1.19E-02	3.21E-01
<i>COMMD7</i>	2.26E-01	2.26E-01	2.26E-01	2.26E-01	6.71E-02	1.52E-03	3.06E-01	1.11E-02
<i>AP3S1</i>	3.37E-01	3.37E-01	3.37E-01	3.37E-01	2.26E-01	2.26E-01	2.26E-01	2.26E-01
<i>SLC4A4</i>	4.33E-02	9.19E-06	1.19E-03	4.61E-03	9.43E-01	6.76E-01	9.43E-01	6.87E-01
<i>LACE1</i>	4.58E-01	5.37E-01	5.37E-01	5.37E-01	7.38E-02	4.89E-02	7.23E-02	4.73E-02
<i>TTC12</i>	6.03E-01	7.20E-03	6.03E-01	2.01E-01	6.42E-01	6.42E-01	6.42E-01	6.42E-01
<i>SYNE2</i>	9.08E-01	8.78E-01	8.83E-01	9.61E-01	9.54E-01	8.76E-01	9.54E-01	8.85E-01

^aTest results include p-values of GBAS of the HKHS (Method: sTDT, QFAM_{SBP}, QFAM_{DBP}, and QFAM_{MBP}) and GBAS of the WTCCCHS (Method: LR_N, LR_C, LR_{N,Gender}, and LR_{C,Gender}) of the 17 genes characterized by both differential allelic distributions and transcriptional distributions between patient and control groups.

^bRaw p-values of genes that reached $p_{\text{FDR}} < 5 \times 10^{-2}$ in the GBAS of the HKHS and the WTCCCHS are marked in bold.

doi:10.1371/journal.pone.0032907.t003

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 [38,39,40,41,42] and metabolic disorders [43,44,45,46,47,48,49,50,51,52]. Some hypertension-associated cis-eQTL act as intronic enhancers at transcriptional factor binding sites on this gene were also found. *WWOX* 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 intronic enhancers at transcriptional factor binding site on this gene were also found. *WWOX* encoded protein is more than 90 percent identical to the mouse protein responsible for modulation of tumor necrosis factor- α -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 *WWOX* 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 ($p_{CLR,N} = 1.79 \times 10^{-3}$, $p_{CLR,C} = 1.31 \times 10^{-4}$, and $p_{CLR,BMI} = 2.96 \times 10^{-3}$), but genetic interaction between rs1886985 and rs6129969 was not significant in this TWNHS.

The original WTCCCCHS study only performed SBAS and no significant SNPs or genes attained a genome-wide significance level of 5×10^{-7} [17]. In our WTCCCCHS data analysis, we performed both SBAS and GBAS to examine the 100 significant genes identified by the genome-wide GBAS in our TWNHS. We made a comparison of SBAS and GBAS in our WTCCCCHS 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*, *CYP7B1* and *PIP5K3*) and SBAS identified 2 genes (*ACTN4* and *TCF7L1*). The concordance rate of gene mapping results in SBAS and GBAS in our WTCCCCHS data analysis is 96.55%.

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 multilocus 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 WTCCCCHS). 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 WTCCCCHS. Second, the three studies used different genome-wide oligonucleotide SNP arrays. Therefore, the SNP markers that were analyzed were different. Our study used the Illumina HumanHap550-Duo BeadChip; HKHS used the Illumina HumanHap610-quad BeadChip; WTCCCCHS used the Affymetrix Human Mapping 500 K Set. Some association signals identified by TWNHS could not be examined in the other two studies due to a lack of SNPs in the regions of interest. Third, the controls in the WTCCCCHS data analysis were the 3,004 shared community controls for seven complex diseases in WTCCC [17]. Some controls may not be normotensive and, therefore, the association signal may be diluted due to a misclassification of hypertensive patients to normotensive status. In other words, significant association identified by TWNHS may be not replicable in WTCCCCHS due to the reduced power of WTCCCCHS. Finally, our current gene expression analysis aimed to prioritize hypertension susceptibility genes for further confirmation. The sample size was somewhat small, although the selected samples had extremely discordant genotype distributions. The small sample size could lead to false negatives. Therefore, we used a liberal multiple-testing correction in the expression data analysis. In our next gene expression study, we will conduct a large-scale gene expression study of 400 patients with hypertension

and 400 normotensive controls using the Human OneArray (Phalanx Biotech, Hsin-Chu, Taiwan), which provides ~300,000 genome probes to achieve more power.

Materials and Methods

Ethics statement

Written informed consent was obtained from each participant at his/her initial clinic visit. The study was approved by the Internal Review Board of Academia Sinica (Permit Number: AS-IRB01-08012).

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 (WTCCCCHS) [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 WTCCCCHS 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/ μ l 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 WTCCCCHS 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 TotalPrepTM RNA Amplification Kit (Illumina, Inc.). The cRNA concentration was quantified and adjusted to 150 ng/ μ l 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

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].

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 WTCCCCHS. 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-based 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 (“CLR_N”), (2) conditional logistic regression with a continuous genotype without BMI adjustment (“CLR_C”), (3) conditional logistic regression with a nominal genotype and BMI adjustment (“CLR_{N,BMI}”), and (4) conditional logistic regression with a continuous genotype and BMI adjustment (“CLR_{C,BMI}”). For each SNP locus, p-values of a one degree-of-freedom Wald test were calculated for model CLR_C and CLR_{C,BMI}, and p-values of a two degree-of-freedom Wald test were calculated for model CLR_N and CLR_{N,BMI}.

Fourth, genome-wide GBAS was performed. According to SNP- and gene-annotation provided by the SNPper 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, 11–100, 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 ($p\text{FDR} < 10^{-3}$) 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, $pFDR < 10^{-3}$. 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 ($pFDR < 5 \times 10^{-2}$) 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 were calculated and adjusted by FDR, where FDR was calculated across four models (“ CLR_N ”, “ CLR_C ”, “ $CLR_{N,BMI}$ ” and “ $CLR_{C,BMI}$ ”) jointly. For association validation of the identified genes, statistical significance was claimed under a significance threshold of $pFDR < 5 \times 10^{-2}$. 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 $p_{ANCOVA} < 5 \times 10^{-2}$. 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 *AP3S1*, 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 between genetically hypertensive and normotensive mouse strains. Confirmation of genetic association was claimed if $p_{ANOVA} < 5 \times 10^{-2}$.

Sixth, we performed gene-centric cis-acting expression quantitative trait loci (cis-eQTL) analysis by using the lymphoblastoid cell line datasets in Geneva GenCord [94] and MuTHER [95] provided in the Genevar (GENe Expression VARIation) [37]. The analysis examined the association between gene expressions and SNPs in 2-Mb windows around the differentially expressed genes identified in our gene expression study (**Table 2**) and hypertension-associated genes identified in previous hypertension genomic

studies (**Table 1**); Spearman’s correlation coefficient and correlation test of gene expressions and continuous SNP coding were calculated. Cis-eQTL with $p < 5 \times 10^{-2}$ 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 $p < 5 \times 10^{-2}$ were selected. Finally, we examined the direction of genetic effect (risky 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 $pFDR < 5 \times 10^{-2}$. We also analyzed genotype data from the WTCCCCHS [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 (“ LR_N ”), (2) logistic regression with a continuous genotype without gender adjustment (“ LR_C ”), (3) logistic regression with a nominal genotype with gender adjustment (“ $LR_{N,Gender}$ ”), and (4) logistic regression with a continuous genotype with gender adjustment (“ $LR_{C,Gender}$ ”). 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 (“ LR_N ”, “ LR_C ”, “ $LR_{N,Gender}$ ”, and “ $LR_{C,Gender}$ ”) jointly. Replication of genetic association was claimed if $pFDR < 5 \times 10^{-2}$. Note that 13 genes out of 100 were not examined in the WTCCCCHS 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://geneticassociationdb.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 ($-\log_{10}$ 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 ($-\log_{10}$ scale) of gene-based association tests, and the horizontal axis is cumulative physical position (Mb scale). (DOC)

Table S1 Raw p-values ($-\log_{10}$ 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 $-\log_{10}$ 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 $pFDR < 5 \times 10^{-2}$ 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 that the positive/negative of Spearman correlation coefficients from Geneva GenCord and MuTHER is not consistent. (XLS)

Table S5 The cis-eQTL analysis of 54 genes in Table 1 identified by previous GWAS in hypertension. 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 that 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 JWC HYH CTT T. Lin SHS WCT JHC HBL WHY TYC CIC SJL GNT BT PCS SSC T. Lam WHP. Wrote the paper: HCY WHP.

References

- Pickering G (1978) Normotension and hypertension: the mysterious viability of the false. *The American Journal of Medicine* 65: 561–563.
- Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, et al. (2005) Global burden of hypertension: analysis of worldwide data. *Lancet* 365: 217–223.
- Pan WH, Chang HY, Yeh WT, Hsiao SY, Hung YT (2001) Prevalence, awareness, treatment and control of hypertension in Taiwan: results of Nutrition and Health Survey in Taiwan (NAHSIT) 1993–1996. *Journal of Human Hypertension* 15: 793–798.
- Funke-Kaiser H, Reichenberger F, Kopke K, Herrmann SM, Pfeifer J, et al. (2003) Differential binding of transcription factor E2F-2 to the endothelin-converting enzyme-1b promoter affects blood pressure regulation. *Human Molecular Genetics* 12: 423–433.
- Jeunemaitre X, Soubrier F, Kotelevtsev YV, Lifton RP, Williams CS, et al. (1992) Molecular basis of human hypertension: role of angiotensinogen. *Cell* 71: 169–180.
- Lifton RP (1996) Molecular genetics of human blood pressure variation. *Science* 272: 676–680.
- Bonnardeaux A, Davies E, Jeunemaitre X, Fery I, Charru A, et al. (1994) Angiotensin II type 1 receptor gene polymorphisms in human essential hypertension. *Hypertension* 24: 63–69.
- Julier C, Delepine M, Keavney B, Terwilliger J, Davis S, et al. (1997) Genetic susceptibility for human familial essential hypertension in a region of homology with blood pressure linkage on rat chromosome 10. *Human Molecular Genetics* 6: 2077–2085.
- Xu X, Rogus JJ, Terwedow HA, Yang J, Wang Z, et al. (1999) An extreme-sib-pair genome scan for genes regulating blood pressure. *American Journal of Human Genetics* 64: 1694–1701.
- Angius A, Petretto E, Maestrale GB, Forabosco P, Casu G, et al. (2002) A new essential hypertension susceptibility locus on chromosome 2p24-p25, detected by genomewide search. *American Journal of Human Genetics* 71: 893–905.

11. Gong M, Zhang H, Schulz H, Lee YA, Sun K, et al. (2003) Genome-wide linkage reveals a locus for human essential (primary) hypertension on chromosome 12p. *Human Molecular Genetics* 12: 1273–1277.
12. Wallace C, Xue MZ, Newhouse SJ, Marcano AC, Onipinla AK, et al. (2006) Linkage analysis using co-phenotypes in the BRIGHT study reveals novel potential susceptibility loci for hypertension. *American Journal of Human Genetics* 79: 323–331.
13. Koivukoski L, Fisher SA, Kaminen T, Lewis CM, von Wöern F, et al. (2004) Meta-analysis of genome-wide scans for hypertension and blood pressure in Caucasians shows evidence of susceptibility regions on chromosomes 2 and 3. *Human Molecular Genetics* 13: 2325–2332.
14. Guzman B, Cormand B, Ribases M, Gonzalez-Nunez D, Botey A, et al. (2006) Implication of chromosome 18 in hypertension by sibling pair and association analyses: putative involvement of the RKHD2 gene. *Hypertension* 48: 883–891.
15. Cowley AW (2006) The genetic dissection of essential hypertension. *Nature Reviews Genetics* 7: 829–840.
16. Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. (2009) Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nature Genetics* 41: 35–46.
17. The Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678.
18. Saxena R, Voight BF, Lyssenko V, Burt NP, de Bakker PIW, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 316: 1331–1336.
19. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, et al. (2007) Framingham Heart Study 100 K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC Medical Genetics* 8 Suppl 1: S3.
20. Kato N, Miyata T, Tabara Y, Katsuya T, Yanai K, et al. (2008) High-density association study and nomination of susceptibility genes for hypertension in the Japanese National Project. *Human Molecular Genetics* 17: 617–627.
21. Psaty BM, O'Donnell CJ, Gudnason V, Lunetta KL, Folsom AR, et al. (2009) Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circulation: Cardiovascular Genetics* 2: 73–80.
22. Wang Y, O'Connell JR, McArdle PF, Wade JB, Dorff SE, et al. (2009) Whole-genome association study identifies STK39 as a hypertension susceptibility gene. *Proceedings of the National Academy of Sciences of the United States of America* 106: 226–231.
23. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, et al. (2009) Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics* 41: 666–676.
24. Org E, Eyheramendy S, Juhanson P, Gieger C, Lichtner P, et al. (2009) Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Human Molecular Genetics* 18: 2288–2296.
25. Cho YS, Go MJ, Kim YJ, Heo JY, Oh JH, et al. (2009) A large-scale genome-wide association study of Asian populations uncovers genetic factors influencing eight quantitative traits. *Nature Genetics* 41: 527–534.
26. Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, et al. (2009) A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genetics* 5: e1000564.
27. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, et al. (2009) Genome-wide association study of blood pressure and hypertension. *Nature Genetics* 41: 677–687.
28. Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, et al. (2011) Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians. *Nature Genetics* 43: 531–538.
29. Yang H-C, Liang Y-J, Wu Y-L, Chung C-M, Chiang K-M, et al. (2009) Genome-wide association study of young-onset hypertension in the Han Chinese population of Taiwan. *PLoS One* 4: e5459.
30. Fox ER, Young JH, Li Y, Dreisbach AW, Keating BJ, et al. (2011) Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study. *Human Molecular Genetics* 20: 2273–2284.
31. Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, et al. (2010) Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS Genetics* 6: e1001177.
32. Slavin TP, Feng T, Schnell A, Zhu X, Elston RC (2011) Two-marker association tests yield new disease associations for coronary artery disease and hypertension. *Human Genetics*.
33. Hong K-W, Lim J-E, Oh B (2011) A regulatory SNP in AKAP13 is associated with blood pressure in Koreans. *Journal of Human Genetics* 56: 205–210.
34. Hiura Y, Tabara Y, Kokubo Y, Okamura T, Miki T, et al. (2010) A genome-wide association study of hypertension-related phenotypes in a Japanese population. *Circulation* 121: 2353–2359.
35. Yang H-C, Liang Y-J, Chung C-M, Chen J-W, Pan W-H (2009) Genome-wide gene-based association study. *BMC Proceedings* 3: S135.
36. Fang Y-J, Deng H-B, Thomas GN, Tzang CH, Li C-X, et al. (2010) Linkage of angiotensinogen gene polymorphisms with hypertension in a sibling study of Hong Kong Chinese. *Journal of Hypertension* 28: 1203–1209.
37. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* 26: 2474–2476.
38. Schut AF, Janssen JA, Deinum J, Vergeer JM, Hofman A, et al. (2003) Polymorphism in the promoter region of the insulin-like growth factor I gene is related to carotid intima-media thickness and aortic pulse wave velocity in subjects with hypertension. *Stroke* 34: 1623–1627.
39. Yazdanpanah M, Rietveld I, Janssen JA, Njajou OT, Hofman A, et al. (2006) An insulin-like growth factor-I promoter polymorphism is associated with increased mortality in subjects with myocardial infarction in an elderly Caucasian population. *The American Journal of Cardiology* 97: 1274–1276.
40. Ruigrok YM, Slooter AJ, Bardoel A, Frijns CJ, Rinkel GJ, et al. (2005) Genes and outcome after aneurysmal subarachnoid haemorrhage. *Journal of Neurology* 252: 417–422.
41. Telgmann R, Dordelmann C, Brand E, Nicaud V, Hagedorn C, et al. (2009) Molecular genetic analysis of a human insulin-like growth factor I promoter P1 variation. *FASEB Journal* 23: 1303–1313.
42. Horio T, Kamide K, Takiuchi S, Yoshii M, Miwa Y, et al. (2010) Association of insulin-like growth factor-I receptor gene polymorphisms with left ventricular mass and geometry in essential hypertension. *Journal of Human Hypertension* 24: 320–326.
43. te Velde SJ, van Rossum EF, Voorhoeve PG, Twisk JW, Delemarre van de Waal HA, et al. (2005) An IGF-I promoter polymorphism modifies the relationships between birth weight and risk factors for cardiovascular disease and diabetes at age 36. *BMC Endocrine Disorders* 5: 5.
44. San Millan JL, Corton M, Villuendas G, Sancho J, Peral B, et al. (2004) Association of the polycystic ovary syndrome with genomic variants related to insulin resistance, type 2 diabetes mellitus, and obesity. *Journal of Clinical Endocrinology & Metabolism* 89: 2640–2646.
45. 't Hart LM, Fritsche A, Rietveld I, Dekker JM, Nijpels G, et al. (2004) Genetic factors and insulin secretion: gene variants in the IGF genes. *Diabetes* 53 Suppl 1: S26–30.
46. Voorhoeve PG, van Rossum EF, Te Velde SJ, Koper JW, Kemper HC, et al. (2006) Association between an IGF-I gene polymorphism and body fatness: differences between generations. *European Journal of Endocrinology* 154: 379–388.
47. Sweeney C, Murtaugh MA, Baumgartner KB, Byers T, Giuliano AR, et al. (2005) Insulin-like growth factor pathway polymorphisms associated with body size in Hispanic and non-Hispanic white women. *Cancer Epidemiology, Biomarkers & Prevention* 14: 1802–1809.
48. Rietveld I, Janssen JA, van Rossum EF, Houwing-Duistermaat JJ, Rivadeneira F, et al. (2004) A polymorphic CA repeat in the IGF-I gene is associated with gender-specific differences in body height, but has no effect on the secular trend in body height. *Clinical Endocrinology* 61: 195–203.
49. Rietveld I, Hofman A, Pols HA, van Duijn CM, Lamberts SW, et al. (2006) An insulin-like growth factor-I gene polymorphism modifies the risk of microalbuminuria in subjects with an abnormal glucose tolerance. *European Journal of Endocrinology* 154: 715–721.
50. Landmann E, Geller F, Schilling J, Rudloff S, Foeller-Gaudier E, et al. (2006) Absence of the wild-type allele (192 base pairs) of a polymorphism in the promoter region of the IGF-I gene but not a polymorphism in the insulin gene variable number of tandem repeat locus is associated with accelerated weight gain in infancy. *Pediatrics* 118: 2374–2379.
51. Knight J, Sham P (2006) Design and analysis of association studies using pooled DNA from large twin samples. *Behavior Genetics* 36: 665–677.
52. Dupuis J, Langenberg C, Prokopenko I, Saxena R, Soranzo N, et al. (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* 42: 105–116.
53. Liu JP, Baker J, Perkins AS, Robertson EJ, Efstratiadis A (1993) Mice carrying null mutations of the genes encoding insulin-like growth factor I (Igf-1) and type I IGF receptor (Igf1r). *Cell* 75: 59–72.
54. Ueki I, Ooi GT, Tremblay ML, Hurst KR, Bach LA, et al. (2000) Inactivation of the acid labile subunit gene in mice results in mild retardation of postnatal growth despite profound disruptions in the circulating insulin-like growth factor system. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6868–6873.
55. Stratikopoulos E, Szabolcs M, Dragatsis I, Klinakis A, Efstratiadis A (2008) The hormonal action of IGF1 in postnatal mouse growth. *Proceedings of the National Academy of Sciences of the United States of America* 105: 19378–19383.
56. Gawenis LR, Bradford EM, Prasad V, Lorenz JN, Simpson JE, et al. (2007) Colonic anion secretory defects and metabolic acidosis in mice lacking the NBC1 Na⁺/HCO₃⁻ cotransporter. *Journal of Biological Chemistry* 282: 9042–9052.
57. Hunt SC, Xin Y, Wu LL, Cawthon RM, Coon H, et al. (2006) Sodium bicarbonate cotransporter polymorphisms are associated with baseline and 10-year follow-up blood pressures. *Hypertension* 47: 532–536.
58. Taylor JY, Maddox R, Wu CY (2009) Genetic and environmental risks for high blood pressure among African American mothers and daughters. *Biological Research For Nursing* 11: 53–65.
59. Aqeilan RI, Hassan MQ, de Bruin A, Hagan JP, Volinia S, et al. (2008) The WWOX tumor suppressor is essential for postnatal survival and normal bone metabolism. *Journal of Biological Chemistry* 283: 21629–21639.

60. Peterson AJ, Mallin DR, Francis NJ, Ketel CS, Stamm J, et al. (2004) Requirement for sex comb on midleg protein interactions in *Drosophila* polycomb group repression. *Genetics* 167: 1225–1239.
61. Burstein E, Hoberg JE, Wilkinson AS, Rumble JM, Csomos RA, et al. (2005) COMMD proteins, a novel family of structural and functional homologs of MURR1. *Journal of Biological Chemistry* 280: 22222–22232.
62. Wiemuth D, Lott JS, Ly K, Ke Y, Teesdale-Spittle P, et al. (2010) Interaction of serum- and glucocorticoid regulated kinase 1 (SGK1) with the WW-domains of Nedd4-2 is required for epithelial sodium channel regulation. *PLoS One* 5: e12163.
63. Bansal V, Libiger O, Torkamani A, Schork NJ (2010) Statistical analysis strategies for association studies involving rare variants. *Nature Reviews Genetics* 11: 773–785.
64. Kos CH, Le TC, Sinha S, Henderson JM, Kim SH, et al. (2003) Mice deficient in alpha-actinin-4 have severe glomerular disease. *The Journal of Clinical Investigation* 111: 1683–1690.
65. Jakobsson J, Karypidis H, Johansson JE, Roh HK, Rane A, et al. (2004) A functional C-G polymorphism in the CYP7B1 promoter region and its different distribution in Orientals and Caucasians. *The Pharmacogenomics Journal* 4: 245–250.
66. Li-Hawkins J, Lund EG, Turley SD, Russell DW (2000) Disruption of the oxysterol 7alpha-hydroxylase gene in mice. *Journal of Biological Chemistry* 275: 16536–16542.
67. Hanasaki H, Takemura Y, Fukuo K, Ohishi M, Onishi M, et al. (2009) Fas promoter region gene polymorphism is associated with an increased risk for myocardial infarction. *Hypertension Research* 32: 261–264.
68. Zhang MC, Furukawa H, Tokunaka K, Saiga K, Date F, et al. (2008) Mast cell hyperplasia in the skin of Dsg4-deficient hypotrichosis mice, which are long-living mutants of lupus-prone mice. *Immunogenetics* 60: 599–607.
69. Pineda LL, Nakano A, Iijima M, Wada M, Yamasaki Y, et al. (2003) A new mutation, ataxia and male sterility (ams), of autoimmune-prone MRL/lpr mouse is not linked to lpr gene but associated with reduction of spleen size and alteration of lymphocyte subpopulations. *Pathology International* 53: 390–395.
70. Cooney GJ, Lyons RJ, Crew AJ, Jensen TE, Molero JC, et al. (2004) Improved glucose homeostasis and enhanced insulin signalling in Grb14-deficient mice. *EMBO Journal* 23: 582–593.
71. Goenaga D, Hampe C, Carre N, Cailliau K, Browaeys-Poly E, et al. (2009) Molecular determinants of Grb14-mediated inhibition of insulin signaling. *Molecular Endocrinology* 23: 1043–1051.
72. Wang K, Li WD, Zhang CK, Wang ZH, Glessner JT, et al. (2011) A Genome-Wide Association Study on Obesity and Obesity-Related Traits. *PLoS ONE* 6.
73. Roy D, Sin SH, Damania B, Dittmer DP (2011) Tumor suppressor genes FHIT and WWOX are deleted in primary effusion lymphoma (PEL) cell lines. *Blood* 118: E32–E39.
74. Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics* 5: e1000529.
75. Buil A, Martinez-Perez A, Perera-Lluna A, Rib L, Caminal P, et al. (2009) A new gene-based association test for genome-wide association studies. *BMC Proceedings* 3: S130.
76. Neale BM, Sham PC (2004) The future of association studies: Gene-based analysis and replication. *American Journal of Human Genetics* 75: 353–362.
77. Pan WH, Fann CSJ, Wu JY, Hung YT, Ho MS, et al. (2006) Han Chinese cell and genome bank in Taiwan: Purpose, design and ethical considerations. *Human Heredity* 61: 27–30.
78. The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426: 789–796.
79. The International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437: 1299–1320.
80. The International HapMap Consortium (2004) Integrating ethics and science in the international HapMap project. *Nature Reviews Genetics* 5: 467–475.
81. The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
82. Puig O, Wang IM, Cheng P, Zhou P, Roy S, et al. (2010) Transcriptome profiling and network analysis of genetically hypertensive mice identifies potential pharmacological targets of hypertension. *Physiological Genomics* 42: 24–32.
83. Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361–372.
84. Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55: 997–1004.
85. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38: 904–909.
86. Wilcoxon F (1945) Individual comparisons by ranking methods. *Biometrics Bulletin* 1: 80–83.
87. Riva A, Kohane IS (2004) A SNP-centric database for the investigation of the human genome. *BMC Bioinformatics* 5: 33.
88. Zaykin DV, Zhivotovsky LA, Westfall PH, Weir BS (2002) Truncated product method for combining P-values. *Genetic Epidemiology* 22: 170–185.
89. Yang HC, Hsieh HY, Fann CSJ (2008) Kernel-based association test. *Genetics* 179: 1057–1068.
90. Yang HC, Lin CY, Fann CSJ (2006) A sliding-window weighted linkage disequilibrium test. *Genetic Epidemiology* 30: 531–545.
91. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Methodological* 57: 289–300.
92. Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19: 185–193.
93. Kendall MG (1938) A new measure of rank correlation. *Biometrika* 30: 81–93.
94. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325: 1246–1250.
95. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genetics* 7: e1002003.
96. Horvath S, Laird NM (1998) A discordant-sibship test for disequilibrium and linkage: No need for parental data. *American Journal of Human Genetics* 63: 1886–1897.
97. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* 81: 559–575.