Integrated Computational and Experimental Analysis of the Neuroendocrine Transcriptome in Genetic Hypertension Identifies Novel Control Points for the Cardiometabolic Syndrome

Ryan S. Friese, PhD; Chun Ye, PhD; Caroline M. Nievergelt, PhD; Andrew J. Schork, BS; Nitish R. Mahapatra, PhD; Fangwen Rao, MD; Philip S. Napolitan, BS; Jill Waalen, MD, MPH; Georg B. Ehret, MD; Patricia B. Munroe, PhD; Geert W. Schmid-Schönbein, PhD; Eleazar Eskin, PhD; Daniel T. O'Connor, MD

Background—Essential hypertension, a common complex disease, displays substantial genetic influence. Contemporary methods to dissect the genetic basis of complex diseases such as the genomewide association study are powerful, yet a large gap exists betweens the fraction of population trait variance explained by such associations and total disease heritability.

Methods and Results—We developed a novel, integrative method (combining animal models, transcriptomics, bioinformatics, molecular biology, and trait-extreme phenotypes) to identify candidate genes for essential hypertension and the metabolic syndrome. We first undertook transcriptome profiling on adrenal glands from blood pressure extreme mouse strains: the hypertensive BPH (blood pressure high) and hypotensive BPL (blood pressure low). Microarray data clustering revealed a striking pattern of global underexpression of intermediary metabolism transcripts in BPH. The MITRA algorithm identified a conserved motif in the transcriptional regulatory regions of the underexpressed metabolic genes, and we then hypothesized that regulation through this motif contributed to the global underexpression. Luciferase reporter assays demonstrated transcriptional activity of the motif through transcription factors HOXA3, SRY, and YY1. We finally hypothesized that genetic variation at *HOXA3*, *SRY*, and *YY1* might predict blood pressure and other metabolic syndrome traits in humans. Tagging variants for each locus were associated with blood pressure in a human population blood pressure extreme sample with the most extensive associations for *YY1* tagging single nucleotide polymorphism rs11625658 on systolic blood pressure, diastolic blood pressure, body mass index, and fasting glucose. Meta-analysis extended the *YY1* results into 2 additional large population samples with significant effects preserved on diastolic blood pressure, body mass index, and fasting glucose.

Conclusions—The results outline an innovative, systematic approach to the genetic pathogenesis of complex cardiovascular disease traits and point to transcription factor YY1 as a potential candidate gene involved in essential hypertension and the cardiometabolic syndrome. **(*Circ Cardiovasc Genet.* 2012;5:430-440.)**

Key Words: BPH mouse strain ■ complex trait ■ essential (genetic) hypertension ■ human genetics ■ metabolic syndrome

Essential hypertension, a common disease, displays substantial genetic influence with heritability estimates for blood pressure up to approximately 50%.[1] The complex, multifactorial nature of this disorder, however, has made it difficult to identify underlying genetic contributors. Investigation of such complex traits with novel and integrative methods might be required to completely understand their genetic basis.

Clinical Perspective on p 440

In the current report, we present a novel method that begins with genomewide transcriptome profiling and then

Received October 21, 2011; accepted May 23, 2012.

From the Departments of Bioengineering (R.S.F., G.W.S.-S.), Medicine (R.S.F., A.J.S., F.R., P.S.N., D.T.O.), Pharmacology (D.T.O.), and Psychiatry (C.M.N.), the Bioinformatics Program (C.Y.), and the Institute for Genomic Medicine (D.T.O.), University of California at San Diego; the VA San Diego Healthcare System, San Diego, CA (D.T.O.); the Departments of Computer Science & Human Genetics, University of California at Los Angeles (E.E.); the Department of Biotechnology, Indian Institute of Technology Madras, Chennai, India (N.R.M.); Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom (P.B.M.); Center for Complex Disease Genomics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD (G.B.E.); and Scripps Research Institute, La Jolla, CA (J.W.).

The online-only Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.111.962415/-/DC1.

Correspondence to Daniel T. O'Connor, MD, Department of Medicine, University of California at San Diego School of Medicine, VASDHS (0838), Skaggs (SSPPS) Room 4256, 9500 Gilman Drive, La Jolla, CA 92093-0838. E-mail doconnor@ucsd.edu

sequentially integrates several disciplines and tools (animal models, microarray technology, bioinformatics, molecular biology, extreme phenotypes, and human genetics) to create, filter, and refine a list of candidate genes for essential hypertension and the metabolic syndrome (a cluster of risk factors for cardiovascular disease, including essential hypertension, insulin resistance, dyslipidemia, and elevated plasma cholesterol).

Our analysis is based on 2 inbred, blood pressure extreme mouse strains, the hypertensive BPH (blood pressure high) and the hypotensive BPL (blood pressure low).[2] The use of extreme phenotypes has been suggested for the study of complex traits as a strategy to increase the power of detecting underlying genetic determinants as well as rare variants.[3–6] The BPH strain parallels human essential hypertension, with elevated blood pressure, increased heart rate, and early mortality.[7] Although BPH was developed through selection solely on blood pressure, the strain exhibits several metabolic syndrome-like abnormalities such as decreased plasma high-density lipoprotein (HDL) cholesterol, enhanced rate of body weight gain on a high-fat diet, and increased pulse rate.[8]

We chose the adrenal gland as the target of microarray analysis of BPH and BPL because its secretory products, both medullary and cortical, determine endocrine, cardiovascular, and sympathetic functions important in hypertension and the metabolic syndrome. For example, medullary catecholamines act through G-protein-coupled adrenergic receptors to modulate force of contraction of the heart and constriction of blood vessels and regulate carbohydrate and lipid metabolism through effects on glucagon and insulin secretion, glycogenolysis, glycolysis, and lipolysis. Cortical steroid hormones regulate reabsorption and secretion of sodium and potassium (mineralocorticoids) and modulate carbohydrate and lipid metabolism (eg, through effects on peripheral glucose uptake, gluconeogenesis, and lipolysis) as well as inflammation (glucocorticoids).

After microarray analysis, bioinformatic, molecular biology, and human genetic tools were applied sequentially to identify significant association (by meta-analysis) of *YY1* rs11625658 single nucleotide polymorphism (SNP) genotype with diastolic blood pressure (DBP), body mass index (BMI), and glucose across 3 large human cohorts, suggesting shared genetic determination for such metabolic syndrome traits. Uniquely, we focused on transcription factors as candidate gene "master switches" because functional changes in them are likely to be pleiotropic and, therefore, provide a unifying genetic mechanism for multiple traits of the metabolic syndrome.

Methods

Mouse Strains

Juvenile (approximately 5-week-old) and adult (approximately 12-week-old), hypertensive (strain=BPH) and hypotensive (strain=BPL) inbred male mice were obtained from colonies at the Jackson Laboratory (Bar Harbor, ME). Systolic blood pressure of the BPH increases early in life. At 7 weeks of age, BPH systolic blood pressure (110 mm Hg) is 40 mm Hg higher than BPL (70 mm Hg).[7] The systolic blood pressure differential increases to 60 mm Hg by 21 weeks of age (BPH=130 mm Hg; BPL=70 mm Hg).[7] Mice were studied according to a protocol approved by the Animal Subjects Committee of the University of California at San Diego, and research was conducted in accordance with institutional guidelines.

Microarray Experiments

We previously presented genomewide gene expression profiles of adrenal glands from BPH and BPL mice using standard Affymetrix (Santa Clara, CA) protocols and MG-U74Av2 GeneChips.[9,10] In short, adrenal gene expression of BPH and BPL mice (n=3, juvenile BPH; n=3, adult BPH; n=3, juvenile BPL; and n=3, adult BPL) was determined using MG-U74Av2 GeneChips, and statistically significant changes in gene expression were determined with 2-factor analysis of variance to assess the impact of strain (BPH versus BPL), age (juvenile versus adult), and strain-by-age interaction on gene expression.

The primary focus of the current report is the impact of strain on gene expression: BPH (3 juvenile BPH+3 adult BPH) versus BPL (3 juvenile BPL+3 adult BPL). Juvenile BPH mice (5 weeks old) are "prehypertensive"; they do not yet have maximal elevation of blood pressure. Prehypertensive animals are useful in genetic studies because the effects of confounding factors (eg, age, maximal blood pressure elevation) on gene expression are minimized. We also used 2-factor analysis of variance to compare BPH with BPL gene expression across 2 age groups (juvenile and adult) to identify genes demonstrating a consistent pattern of differential expression (ie, underexpressed in both juvenile and adult BPH or overexpressed in both juvenile and adult BPH) in both preliminary and advanced stages of hypertensive disease. This consistent level of expression across 2 age groups is a key part of the analysis. Such analysis might enrich the set of differentially expressed genes for underlying genetic determinants of hypertension with effects independent from (or resistant to) confounding changes in age, blood pressure, and hypertensive disease processes. Genes with consistent patterns of expression across both age groups might be stronger candidates for pathogenic drivers of disease than genes with changing or inconsistent expression patterns. Microarray data are available at the National Center for Biotechnology Information Gene Expression Omnibus database (www.ncbi.nlm.nih.gov/geo) under the following accession numbers: GSE1674, GSE19342.

Our analysis was designed to gain a broad perspective on which biochemical pathways and physiological systems exhibit global changes rather than focus on specific changes in individual candidate genes. Toward this end, we used GenMAPP Version 2.1 (www.genmapp.org)[11] to perform functional clustering and pathway analysis. Fisher exact tests were used to determine if significantly perturbed pathways (identified with GenMAPP) displayed patterns of global under- or overexpression.

Computational Motif Identification

We defined a gene "regulatory sequence" to be the sequence of DNA from −1000 bp upstream to +1000 bp downstream of the transcription start site to capture the majority of the proximal promoter transcriptional elements upstream of the TATA box as well as any transcriptional regulatory motifs within the 5′-untranslated region or first intron. Regulatory sequences were extracted from an in-house database of the *Mus musculus* genome, originally downloaded from the University of California, Santa Cruz, genome database (http://genome.ucsc.edu). We used MITRA[12,13] to discover conserved motifs within a set of regulatory sequences. The MITRA algorithm identifies common motifs in unaligned DNA sequences by maximizing a score that discriminates sequences between a positive set (ie, our set of regulatory sequences hypothesized to contain a common motif) and a negative set (ie, the set of regulatory sequences from all genes not represented in the positive set). The algorithm was constrained to search for motifs 10 bp in length on both the forward and reverse DNA strands. WebLogo[14,15] was used to create a consensus sequence representation of the motifs identified with MITRA.

Computational Prediction of Transcription Factor Binding

Predictions for transcription factors that bind the MITRA-identified motif were performed with CONSITE (which uses the JASPAR

database; www.phylofoot.org/consite)[16] and P-MATCH (which uses the TRANSFAC database; www.gene-regulation.com/cgi-bin/pub/programs/pmatch/bin/p-match.cgi),[17] 2 web-based tools for identifying *cis*-regulatory elements in genomic sequences. Predictions for transcription factors that bind the SV40 promoter of the pGL3-Promoter control vector (Promega, Madison, WI) were performed with CONSITE.

Construction of Promoter or Enhancer/Luciferase Reporter Plasmids

Cloning of the motif oligonucleotide sequences was accomplished at the unique XhoI (promoter) or SalI (enhancer) restriction enzyme sites in pGL3-Promoter. The 2 oligos (5′-[phosphate] TCGAGACC ATAGATAC-3′; 5′-[phosphate] TCGAGTATCTATGGTC-3′) were annealed in the following reaction mixture: 162.5 picomole of each oligo plus water for a total of 75 μL and 25 μL of annealing buffer (100 mmol/L Tris pH 7.5, 100 mmol/L $MgCl_2$, 10 mmol/L DTT). The annealing reaction mixture was heated at 90°C for 5 minutes and then allowed to cool to room temperature over a period of 45 minutes. To test the transcriptional effects of the motif when it is located in the proximal promoter region of a gene, the annealed oligos were inserted into the unique XhoI site just upstream of the SV40 promoter in the multiple cloning site of the firefly luciferase reporter vector, pGL3-Promoter (Promega). To allow for testing of transcriptional "enhancer" (distant) effects of the motif, the annealed oligos were inserted in the SalI restriction enzyme site just downstream of the luciferase gene poly-A signal in the pGL3-Promoter vector (Promega). The oligos inserted in the "enhancer" position were designed with the following sequence: 5′-(phosphate) CTAGCACC ATAGATAC-3′; 5′-(phosphate) TCGAGTATCTATGGTG-3′. A total of 4 types of luciferase reporter plasmids was constructed with the motif being inserted in either the "promoter" or "enhancer" position in either the forward or reverse orientation. Insertion of the motif in 2 directions enabled testing of orientation specific effects.

Adrenal Cell Promoter/Reporter Transfection and Luciferase Activity Assays

Rat PC12 pheochromocytoma cells (grown in Dulbecco's modified Eagle's medium high glucose [Invitrogen] with 5% heat-inactivated fetal bovine serum [Gemini Bioproducts, Woodland, CA], 10% heat-inactivated horse serum [Gemini Bioproducts], penicillin [100 U/mL], streptomycin [100 μg/mL], and L-glutamine [0.292 mg/mL]) were transfected (at 50%–60% confluence, 1 day after splitting 1:4) with motif-promoter (XhoI; forward or reverse orientation) or motif enhancer (SalI; forward or reverse orientation) reporter vector DNA (1 μg supercoiled DNA per well; 12-well polystyrene plates [coated with poly-L-lysine; Sigma], 2.2-cm diameter wells, Corning Inc, Corning, NY) using the liposome method (Superfect; Qiagen). Cotransfection of mammalian expression (CMV-promoter) transcription factor plasmids was performed with 50 ng of pcDNA3.1(-) (empty vector control; Invitrogen, Carlsbad, CA), HoxA3 (*Mus musculus*; MMM1013–98478526; Open Biosystems, Huntsville, AL), MEF2A[18] (*Homo sapiens*), RunX1 (*Mus musculus*; MMM1013–9498895; Open Biosystems), SOX9 (*Homo sapiens;* MHS1010–9205725; Open Biosystems), Sox17 (*Danio rerio*; MDR1734–97029554, Open Biosystems), Sry (*Rattus norvegicus*; Sry1/pcDNA3.1),[19] or YY1 (*Homo sapiens*; SC118004, OriGene, Rockville, MD). Cells were lysed 20 hours after transfection with lysis buffer (300 μL per well; 0.1 mol/L phosphate buffer [K_2HPO_4+KH_2PO_4; pH 7.8], 1 mmol/L DTT, and 0.1% Triton-X 100).

The bioluminescent activity of luciferase in 80 μL of cell lysate was determined with the AutoLumat LB 953 luminometer (EG&G Berthold, Nashua, NH) by measuring light emission (incubation time=0 seconds, measure time=10 seconds, temperature=25°C) after addition of assay buffer (100 μL per sample; 100 mmol/L Tris-acetate [pH 7.8], 10 mmol/L Mg-acetate, 1 mmol/L EDTA [pH 8.0], 3 mmol/L adenosine 5′-triphosphate, and 100 μmol/L luciferin [Sigma-Aldrich]). As a control for varying cell number between individual wells, the total protein content was measured in the cell lysate using the Bio-Rad Protein Assay (Coomassie blue dye absorbance shift; based on the Bradford method; Bio-Rad, Hercules, CA). Luciferase activity (n=4–5 wells/condition) is expressed as the normalized ratio of (luciferase activity)/(total protein) or (RLU/μg protein). Data were analyzed with analysis of variance followed by pairwise *t* tests corrected for multiple comparisons (Bonferroni).

Human Subjects

Population Blood Pressure Extremes

We previously published a detailed description of the human subjects and subject selection criteria used in this investigation.[20] Unrelated adults from the Kaiser-Permanente Medical Group (subscription-based Health Maintenance Organization) primary care population located in San Diego, CA, were selected from the lowest and highest (extreme) percentiles of diastolic blood pressure (DBP) distribution (≤63 mm Hg or ≥90 mm Hg); subjects were ascertained on the DBP trait, because twin and family studies provide evidence that DBP is substantially heritable[21–24] and SBP correlates highly with DBP. Two independent DBP extreme samples (cohort 1: n=996; cohort 2: n=1075) were obtained from different individuals within the source population. Blood pressure was measured in seated subjects using brachial sphygmomanometry. If DBP was elevated, repeat measurement was obtained for verification (only the initial value was reported; values were not averaged). Approximately 48% of the hypertensive subjects from the upper DBP extreme reported being prescribed and taking one or more antihypertensive medication. Subjects in the lower extreme DBP group reported no history of hypertension or antihypertensive medication. Subjects did not have renal failure (serum creatinine concentration was ≤1.5 mg/dL in 98.6% of subjects). Self-identified ethnicity (including that of both parents and all 4 grandparents) for all subjects was specified as white (European ancestry). Medical information was obtained from annual health appraisal visits and questionnaire. Blood for preparation of genomic DNA was obtained with informed consent, and samples were deidentified.

Extension of Human Genetic Association: ICBP-GWAS (International Consortium of Blood Pressure–Genome Wide Association Studies)

Extension of associations in our two blood pressure extreme cohorts was sought in the International Consortium of Blood Pressure (ICBP GWAS).[25] Complete details of ICBP methodology have previously been presented.[26] In short, ICBP data (www.igm.jhmi.edu/gehret/icbp32413ahsfd134/icbp_088023401234–9812599.html) were analyzed in separate genomewide meta-analyses for SBP and DBP. Before meta-analysis, the association results for each cohort were filtered to exclude SNPs not in HapMap, SNPs with alleles different from HapMap, and SNPs with observed/expected ratio of coded genotype scores <0.3. A genomic control correction was applied. Where studies stratified their analyses by sex, genomic control was applied within each sex stratum. For each SNP, the per-coded-allele effects were combined across studies (and across sex strata, when used) using inverse variance-weighted meta-analysis.

Genotyping and Statistical Analysis

The HapMap (http://hapmap.ncbi.nlm.nih.gov) was used to select common (minor allele frequency >5%), validated tagging SNPs that capture the linkage disequilibrium and haplotype block structure within candidate genes using CEU (European ancestry) subject data. By inspection of CEU linkage disequilibrium heat plots, each target locus (HOXA3, 7.8 kbp; YY1, 39.7 kbp; and SRY at 896 bp) was spanned by a single linkage disequilibrium block in CEU subjects. A total of 7 SNPs were selected to tag these 3 loci (at 2–3 SNPs per locus): 2 at HOXA3 (rs10085570, rs6948297), 2 at SRY (rs2058276, rs1865680), and 3 at YY1 (rs8021803, rs11625658, rs4905941). The HOXA3 and YY1 tagging SNPs were located within introns. SRY tagging SNPs were located adjacent to the SRY gene in the 5′-upstream region. SNP genotyping was performed on subjects' genomic DNA with the matrix-assisted laser desorption ionization

Figure 1. Computationally identified motif in the regulatory sequences of differentially expressed genes in the intermediary metabolism pathways of the BPH. MITRA identified a significantly overrepresented (P<0.05) consensus 10-bp motif in the set of regulatory sequences from differentially expressed genes in intermediary metabolism pathways of the BPH. The WebLogo consensus sequence consists of stacks of nucleotide symbols, one stack for each position in the sequence. The overall height of each stack indicates the sequence conservation at that position (measured in bits), whereas the height of symbols within the stack reflects the relative frequency of the corresponding nucleic acid at that position. Positions 8 and 10, which lack symbols, showed no nucleotide preference and can be represented as "N" in the consensus sequence: 5′-ACCATAGNTN-3′. BPH indicates blood pressure high mouse strain.

time-of-flight mass spectrometry system developed by Sequenom, as previously described.[20]

Statistical association between individual SNPs and individual metabolic syndrome traits was tested with univariate analysis of variance (independent variable=diploid genotype [additive model]; dependent variable=one metabolic syndrome trait; covariates=age, sex). A false discovery rate (<0.05) was used to control for testing of multiple genotypes and phenotypes, as described[27] and applied[28,29] in the literature. A total of 6 phenotypes were tested for association: SBP, DBP, body mass index (BMI), (fasting) plasma glucose, total cholesterol, and HDL cholesterol. Plasma triglyceride and low-density lipoprotein cholesterol data were not available in these subjects. A second control for testing of multiple phenotypes was also achieved using multivariate analysis of variance, wherein statistical association between individual SNPs and multiple metabolic syndrome traits was evaluated simultaneously using one test (multivariate analysis of variance: independent variable=diploid genotype [additive model]; dependent variable=SBP, DBP, BMI, plasma glucose, plasma cholesterol, and/or plasma HDL cholesterol; covariates=age, sex). We included age and sex as covariates to control for their effects on blood pressure and metabolic phenotypes. No adjustment was made for the use of antihypertensive medication. Blood pressure data can be adjusted for antihypertensive therapy (with, for example, the method described by Cui[30]); however, without adjustment, our analysis was likely more conservative, that is, biased toward the null (insignificance), because adjustment would tend to disproportionately increase the mean blood pressure of diploid genotype groups consisting of a higher proportion of hypertensive subjects. The effect size for significant SNPs is presented as partial η-squared (η^2): the proportion of total trait variance attributable to a specific SNP. Statistical analyses were performed using SPSS software (SPSS Inc, Chicago, IL). Data are presented as meantime±SEM.

To evaluate the cumulative effect of genetic variation at YY1 on metabolic syndrome traits in the 3 groups (San Diego cohorts 1 and 2; ICBP-GWAS), we turned to meta-analysis using the outcomes of regression analysis (additive model) in each group, focusing on the effect size β (slope per allele) and the SE of β using the command METAN within the program STATA (Stata Corporation, College Station, TX; www.stata.com) reporting results of fixed effect (ie, genotype as independent variable) models.

Results

Microarray Adrenal mRNA Expression Analysis

Previous statistical analysis of the BPH and BPL adrenal microarray data identified 2004 significantly differentially expressed genes—approximately 16% of all probe sets[9,10] (online-only Data Supplement Table I). The distribution of differential expression was split approximately evenly between overexpression (924 of 2004 genes [46%]) and underexpression (1080 of 2004 genes [54%]) in BPH versus BPL.

We used GenMAPP to functionally cluster the microarray data into statistically significant biochemical systems and pathways. GenMAPP identified 7 biochemical pathways within the intermediary metabolism domain (the set of canonical biochemical pathways responsible for intracellular energy production, ie, adenosine 5′-triphosphate formation resulting from carbohydrate or lipid metabolism) of the BPH that were significantly perturbed: the electron transport chain (P<0.001), fatty acid degradation (P<0.01), fatty acid synthesis (P<0.001), glycolysis/gluconeogenesis (P<0.001), mitochondrial long chain fatty acid β-oxidation (P<0.001), pentose phosphate (P<0.004), and tricarboxylic acid cycle (P<0.001) pathways (online-only Data Supplement Table II). In addition, all of the significantly perturbed intermediary metabolism pathways were also globally underexpressed in BPH: the electron transport chain (P<0.0001), fatty acid degradation (P=0.011), fatty acid synthesis (P=0.031), glycolysis/gluconeogenesis (P=0.006), mitochondrial long chain fatty acid β-oxidation (P=0.021), pentose phosphate (P=0.031), and tricarboxylic acid cycle pathways (P=0.0005; Fisher exact test; online-only Data Supplement Table II). The tricarboxylic acid cycle pathway exhibited a pattern of global underexpression representative of that observed in all of the intermediary metabolism pathways (online-only Data Supplement Figure I). In total, 82 genes of the intermediary metabolism pathways were differentially expressed with the directional pattern of expression significantly and globally shifted toward underexpression: 79 of 82 genes were underexpressed, whereas 3 of 82 genes were overexpressed (P<0.0001; Fisher exact test).

Computational Promoter Motif Identification

The MITRA algorithm was used to identify statistically overrepresented transcription factor binding motifs in regulatory sequences from the set of 82 differentially expressed genes (79 underexpressed, 3 overexpressed) in intermediary metabolism of the BPH adrenal gland. MITRA identified one significantly overrepresented (P<0.05) 10-bp motif that was present a total of 55 times in 38 of the 82 (38 of 82 [46%]) differentially expressed metabolic genes: 5′-ACCATAGNTN-3′ (Figure 1; online-only Data Supplement Table III). The motif was not localized to a limited set of subpathways of intermediary metabolism (eg, only in the tricarboxylic acid cycle and pentose phosphate subpathways) but instead was present in an approximately equal proportion of differentially expressed

genes (approximately 50%; P=0.97, χ^2 test) in each of the subpathways of intermediary metabolism.

Promoter/Luciferase Reporter Assays: Determination of Motif Transcriptional Activity

Promoter/luciferase reporter plasmid constructs were used to determine if the motif conferred transcriptional activity. Because positions 8 and 10 in the motif consensus sequence (5′-ACCATAGNTN-3′) lacked nucleotide preference ("N"), adenine ("A") was randomly and independently chosen to fill each "N" position in the motif (5′-ACCATAG**A**T**A**-3′). The motif was inserted in 2 regions of the pGL3-promoter luciferase reporter plasmid (driven by the SV-40 promoter): (1) in the "promoter" site just upstream of the SV-40 promoter; and (2) in the "enhancer" site downstream of the luciferase gene. The motif was inserted in either the forward (sense strand: 5′-ACCATAGATA-3′) or reverse (sense strand: 5′-TATCTATGGT-3′) orientation. Luciferase assays were performed in PC12 cells, a cell line derived from a rat adrenal medullary chromaffin cell tumor (pheochromocytoma).[31]

The motif significantly decreased transcriptional activity when inserted into the "promoter" position in both forward (0.50±0.04 RLU/μg protein; P=0.0003) and reverse orientations (0.66±0.06 RLU/μg protein; P=0.0055) compared with the control vector without motif insert (1.00±0.07 RLU/μg protein; Figure 2). There was no significant difference between luciferase activity of the promoter-forward and promoter-reverse orientation constructs. The motif did not significantly affect transcriptional activity when inserted into the "enhancer" position in either the forward (1.02±0.04 RLU/μg protein) or reverse orientations (0.94±0.07 RLU/μg protein) compared with the control vector without motif insert (1.00±0.07 RLU/μg protein; Figure 2).

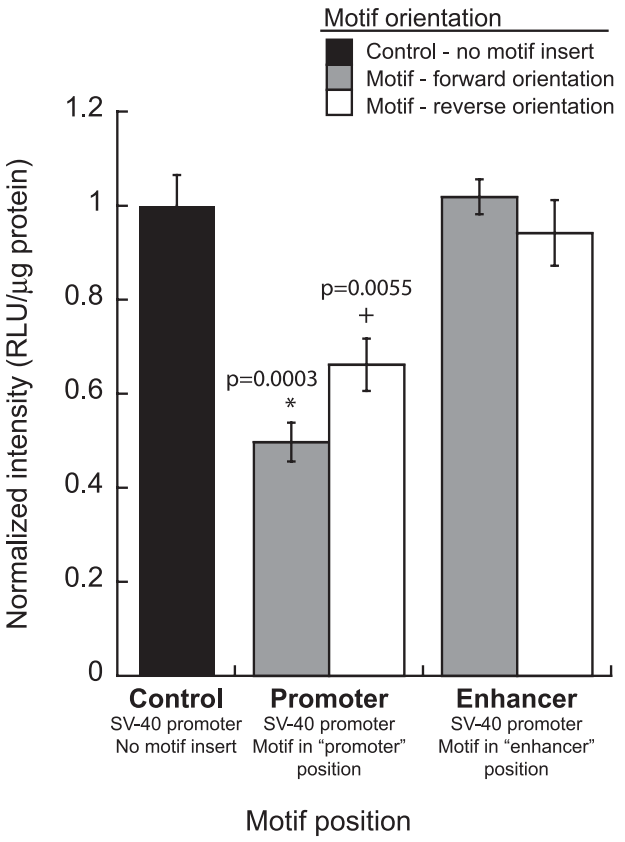

Figure 2. The motif confers transcriptional activity in the promoter position but not in the enhancer position. Orientation- and distance-specific transcriptional effects of the motif were determined using promoter/luciferase reporter assays. The motif was inserted into 2 regions of the pGL3-promoter luciferase reporter plasmid: (1) in the "promoter" (XhoI) site just upstream of the SV-40 promoter; and (2) in the "enhancer" (SalI) site downstream of the luciferase gene. The motif was inserted in either the forward (5′-ACCATAGATA-3′) or reverse (5′-TATCTATGGT-3′) orientation. The control plasmid was the pGL3-promoter vector (with eukaryotic transcription driven by the SV-40 early promoter) without a motif insert. The motif significantly altered transcriptional activity in the "promoter" position in both orientations at the same time as showing no effect in the "enhancer" position in either orientation. Data were analyzed with ANOVA followed by pairwise t tests corrected for multiple comparisons (Bonferroni). *P=0.0003 versus control. +P=0.0055 versus control. ANOVA indicates analysis of variance

Computational Prediction of Motif/Transcription Factor Binding

The CONSITE and P-MATCH algorithms were used to predict which transcription factors bind to the motif identified with MITRA (Table 1). CONSITE identified 6 putative transcription factors: MEF2A—myocyte enhancer factor 2A, RUNX1—runt-related transcription factor 1, SOX9—SRY-box 9, SOX17—SRY-box 17, SRY—sex-determining region Y, and YY1—yin yang 1. P-MATCH identified 4 putative transcription factors: HOXA3—homeobox A3, RUNX1, SRY, and YY1. The MEF2A, RUNX1, SOX9, and SOX17 transcription factors were also predicted to bind the SV-40 promoter, the promoter that drives expression of the pGL3-promoter vector (the control vector in the luciferase reporter experiments).

Luciferase Assays: Transactivation by Transcription Factor Cotransfection

Promoter/luciferase reporter cotransfection experiments were performed with the promoter-forward and promoter-reverse orientation constructs as well as cDNA expression plasmids for the transcription factors computationally predicted to bind the motif. One set of transcription factors (HOXA3, SRY, YY1) had no effect on the control vector at the same time as significantly increasing the expression of the promoter-forward or promoter-reverse constructs (Figure 3; Table 2). More specifically, HOXA3 increased expression of the promoter-reverse construct by 1.24-fold (P<0.05); SRY increased expression of the promoter-forward construct by 1.26-fold (P<0.05); and YY1 significantly increased expression of the promoter-forward construct by 1.49-fold (P<0.05) and the promoter-reverse construct by 1.63-fold (P<0.05).

Another set of transcription factors (RUNX1, SOX9, SOX17) significantly decreased expression of the insertless control vector (SV-40 promoter, pGL3-promoter) at the same time as also reducing expression of the promoter-forward and/or promoter-reverse constructs (Figure 4; Table 2). Specific changes included: RUNX1 reduced expression of the promoter-forward construct by 0.69-fold (P<0.05); SOX9 reduced expression of the promoter-forward construct by 0.64-fold (P<0.05); SOX17 reduced expression of the promoter-forward

Table 1. Transcription Factors Predicted to Bind the MITRA-Identified Motif

Transcription Factor	Accession No.	CONSITE Score	P-MATCH Score	Strand
SOX9	MA0077	6.194	…	+
SRY	MA0084; R07263	2.695	0.755	+
YY1	MA0095; R05992	7.027	0.964	+
HOXA3	R07263	…	0.938	–
MEF2A	MA0052	5.567	…	–
RUNX1	MA0002; R07891	5.491	1	–
SOX17	MA0078	4.273	…	–

The CONSITE and P-MATCH algorithms identified a total of 7 transcription factors predicted to bind the MITRA-identified motif. Three transcription factors (RUNX1, SRY, and YY1) were identified by both algorithms. The JASPAR accession no. for CONSITE predictions and the TRANSFAC site accession for P-MATCH predictions are listed. CONSITE and P-MATCH binding scores and the DNA strand on which the transcription factor was predicted to bind are also presented.

construct by 0.73-fold (P<0.05); and the promoter-reverse construct by 0.75-fold (P<0.05). MEF2A reduced expression of the control vector by 0.77-fold (P<0.05) but had no effect on the promoter-forward or promoter-reverse constructs.

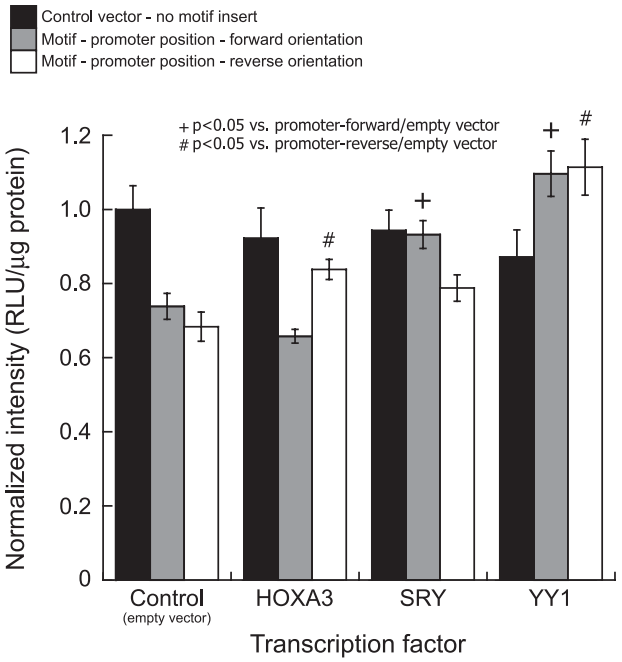

Figure 3. Transcriptional activity of the motif was enhanced by the transcription factors HOXA3, SRY, and YY1. Promoter/luciferase reporter cotransfection experiments were performed with cDNA expression plasmids (pCMV) for the HOXA3, SRY, and YY1 transcription factors. The effects of these transcription factors on the motif in the promoter position in the forward or reverse orientations were tested. HOXA3, SRY, and YY1 had no effect on the pGL3-promoter control vector (SV-40 promoter; no motif insert). The control for the transcription factor expression plasmids was an empty vector containing only the CMV promoter (pcDNA-3.1). HOXA3 increased expression of the promoter-reverse construct by 1.24-fold (24%). SRY significantly increased expression of the promoter-forward construct by 1.26-fold (26%). YY1 significantly increased expression of the promoter-forward construct by 1.49-fold (49%) and the promoter-reverse construct by 1.63-fold (63%). Data are presented as normalized intensity (RLU/µg protein) and were analyzed with ANOVA followed by pairwise *t* tests corrected for multiple comparisons. +P<0.05 versus promoter-forward/empty vector. #P<0.05 versus promoter-reverse/empty vector. ANOVA indicates analysis of variance.

Human Blood Pressure Extremes: Statistical Genetic Associations With the Metabolic Syndrome

Initial Study

The HapMap was used to select common (minor allele frequency >5%), validated tagging SNPs that capture the linkage disequilibrium and haplotype block structure within the human HOXA3, SRY, and YY1 loci. These tagging SNPs were genotyped in a sample from extremes of blood pressure in the San Diego population (all SNPs were genotyped in cohort 1): white (European ancestry) males and females from the lowest and highest percentiles of DBP distribution (≤63 mm Hg or ≥90 mm Hg). Statistical association was tested between tagging SNPs and several traits of the human metabolic syndrome: SBP, DBP, BMI, plasma glucose, plasma total cholesterol, and plasma HDL cholesterol (Table 3).

In the first cohort, *HOXA3* SNP rs10085570 showed significant association with DBP (P=0.017). The *SRY*-tagging SNP rs2058276 in males significantly associated with SBP (P=0.027), DBP (P=0.018), and BMI (P=0.004). *YY1* SNP rs8021803 showed a significant association with both SBP (P=0.004) and DBP (P=0.002). *YY1* SNP rs11625658 significantly associated with SBP (P=0.0002), DBP, (P=0.0006), BMI (P=0.006), and plasma glucose (P=0.024). Multivariate analysis of variance for *YY1* rs11625658 revealed significant association between *YY1* genotype and the following set of metabolic syndrome traits: SBP, DBP, BMI, and plasma glucose (P=0.0005; effect size=0.012 or 1.2% for the joint set of traits; specific individual trait effect size and directionality are shown in Tables 3 and 4). No significant associations were found for plasma total cholesterol or plasma HDL cholesterol for any of the HOXA3, SRY, or YY1 SNPs.

SNPs found to be significantly associated in cohort 1 were then examined within San Diego blood pressure extreme cohort 2. Only *YY1* SNP rs1625658 showed a significant association in this second cohort; *HOXA3* SNP rs10085570, *SRY* SNP rs2058276, and *YY1* SNP rs8021803 did not show significant association in cohort 2 (data not shown).

Potential interdependence of the statistical associations of SBP, DBP, BMI, and plasma glucose (4 correlated traits) with *YY1* SNP rs11625658 was examined using a modified statistical model wherein the 4 traits were systematically used as covariates for each other (using the false discovery rate <0.05 multiple testing correction; age and sex remained as covariates in all analyses). In cohort 1, associations with SBP and DBP remained significant when BMI and/or plasma glucose were added as covariates; and BMI retained significance when plasma glucose was added as a covariate (data not shown).

Meta-Analysis

Because rs11625658 showed the highest degree of association in cohort 1 and preliminary association in cohort 2,

Table 2. Effect of Candidate Transcription Factors on Motif Activity

Transcription Factor	Control Vector, RLU/μg Protein	Promoter-Forward, RLU/μg Protein	Promoter-Reverse, RLU/μg Protein
Control: empty vector	1.00±0.06	0.74±0.04	0.68±0.04
No effect on control vector			
HOXA3	0.92±0.08	0.66±0.02	0.84±0.03‡
SRY	0.94±0.05	0.93±0.04†	0.79±0.04
YY1	0.87±0.07	1.10±0.06†	1.11±0.08‡
Effect on control vector			
MEF2A	0.77±0.06*	0.73±0.04	0.62±0.07
RUNX1	0.62±0.03*	0.51±0.09†	0.65±0.06
SOX9	0.69±0.09*	0.47±0.02†	0.54±0.05
SOX17	0.62±0.02*	0.54±0.02†	0.51±0.03‡

Luciferase reporter assays were performed with cotransfection of cDNA expression plasmids for transcription factors computationally predicted to bind the motif. The effects of transcription factors on the motif in the promoter position in either the forward or reverse orientation were tested. Data are presented as normalized intensity (RLU/μg protein). The pGL3-promoter vector, which contains the SV-40 promoter driving expression of the luciferase gene, served as the control luciferase expression vector. The control for the transcription factor expression plasmids was an empty vector containing only the CMV promoter. Data were analyzed with analysis of variance followed by pairwise *t* tests corrected for multiple comparisons.

*P<0.05 versus control vector/empty vector.

†P<0.05 versus promoter-forward construct/empty vector.

‡P<0.05 versus promoter-reverse construct/empty vector.

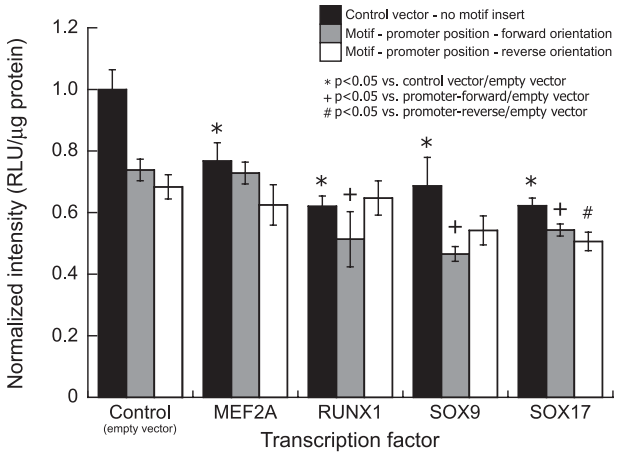

Figure 4. Transcriptional activity of the motif and the control vector were reduced by the transcription factors MEF2A, RUNX1, SOX9, and SOX17. Promoter/luciferase reporter cotransfection experiments were performed with cDNA expression plasmids (pCMV) for MEF2A, RUNX1, SOX9, and SOX17. The effects of these transcription factors on the motif in the promoter position in the forward or reverse orientation were tested. MEF2A, RUNX1, SOX9, and SOX17 significantly reduced expression of the pGL3-promoter control vector (SV-40 promoter; no motif insert). The control for the transcription factor expression plasmids was an empty vector containing only the CMV promoter (pcDNA-3.1). RUNX1 significantly reduced expression of the promoter-forward construct by 1.45-fold (31%). SOX9 significantly reduced expression of the promoter-forward construct by 1.57-fold (36%). SOX17 significantly reduced expression of the promoter-forward construct by 1.37-fold (27%) and the promoter-reverse construct by 1.33-fold (25%) (P<0.05). Data are presented as normalized intensity (RLU/μg protein) and were analyzed with ANOVA followed by pairwise *t* tests corrected for multiple comparisons. *P<0.05 versus control vector/empty vector. +P<0.05 versus promoter-forward/empty vector. #P<0.05 versus promoter-reverse/empty vector. ANOVA indicates analysis of variance.

we sought to extend the results of its significant SBP, DBP, BMI, and glucose associations into 2 additional population samples: San Diego blood pressure extreme cohort 2 (n=1075) and the ICBP-GWAS (n=66 741) for a total of n=68 812 subjects (Table 4). By meta-analysis, considering the effect size (β or slope per allele) and its SE for each subgroup, we found nominally significant effects across the 3 groups for YY1 tagging variant rs11625658 on DBP (P=0.033), BMI (P=0.004), and fasting glucose (P=0.038). Inspection of β (slope) values revealed directionally consistent effects in each subgroup, although inverse allelic effects (ie, positive versus negative β-slope values) were observed for DBP and BMI. SBP and cholesterol (total or HDL) were not significantly affected in the meta-analysis (each P>0.05). The previously reported, isolated effects of rs11625658 on SBP (P=0.502) and DBP (P=0.047) by the ICBP-GWAS parallel the results of our meta-analysis. The Genetic Investigation of ANthropometric Traits (GIANT) consortium GWAS of BMI in n=249 796 individuals reported P=0.0603 for SNP rs11625658.[32]

Discussion

Despite the substantial heritability of blood pressure, the genetic underpinnings of hypertension remain incompletely understood. Investigation of such complex traits may benefit from a comprehensive set of tools that includes not only the GWAS, but also integrative and novel methods that might ultimately be required to completely reveal complex trait genetic determinants. To this end, we developed a novel approach, integrating tools from several disciplines, including bioinformatics, molecular biology, and human genetics, to identify novel candidate genes for essential hypertension and the cardiometabolic syndrome (online-only Data Supplement Figure II).

Microarray Expression Analysis

We began with genomewide gene expression analysis of adrenal glands from the BPH and BPL blood pressure trait extreme inbred mouse strains. Although the widespread expression changes (up to approximately 16%) were initially surprising (online-only Data Supplement Table I), they likely reflected the polygenic nature of essential hypertension, the comparison of extreme phenotypes, and the diverse adaptations or responses to disease processes. Functional clustering uncovered a striking pattern of global underexpression of intermediary metabolism pathways in the BPH (online-only Data Supplement Table II), and we hypothesized that such consistent underexpression throughout interacting and functionally related pathways resulted from a common transcriptional mechanism.

Computational Motif Identification

The MITRA algorithm identified a conserved motif in the set of transcriptional regulatory sequences from the 82 differentially

Table 3. Initial (cohort 1; n=996) associations of candidate genes with hypertension and metabolic syndrome traits

Gene	RefSNP ID	Systolic BP	Diastolic BP	BMI	Plasma Glucose	HDL Cholesterol	Total Cholesterol
YY1	rs8021803	*P*=0.004* (1.3%)	*P*=0.002* (1.4%)	*P*=0.098	*P*=0.359	*P*=0.176	*P*=0.065
YY1	rs11625658	*P*=0.0002* (1.9%)	*P*=0.0006* (1.6%)	*P*=0.006* (1.2%)	*P*=0.024* (0.9%)	*P*=0.137	*P*=0.281
YY1	rs4905941	*P*=0.113	*P*=0.166	*P*=0.037	*P*=0.636	*P*=0.154	*P*=0.901
SRY	rs2058276	*P*=0.027* (1.5%)	*P*=0.018* (1.7%)	*P*=0.004* (2.3%)	*P*=0.079	*P*=0.191	*P*=0.149
SRY	rs1865680	*P*=0.299	*P*=0.371	*P*=0.113	*P*=0.651	*P*=0.272	*P*=0.733
HOXA3	rs10085570	*P*=0.050	*P*=0.017* (1.0%)	*P*=0.614	*P*=0.879	*P*=0.411	*P*=0.665
HOXA3	rs6948297	*P*=0.054	*P*=0.116	*P*=0.081	*P*=0.973	*P*=0.652	*P*=0.867

Statistical association between candidate gene tagging single nucleotide polymorphisms (SNPs) and the following metabolic syndrome traits was tested: systolic blood pressure (BP), diastolic BP, body mass index (BMI), plasma glucose, plasma total cholesterol, and plasma high-density lipoprotein (HDL) cholesterol. Data were analyzed with univariate analysis of variance using an additive model (independent variable=diploid genotype; dependent variable=metabolic syndrome trait; covariates=age, sex). Nominal *P* values that satisfy a false discovery rate (0.05) correction for testing of multiple genotypes and phenotypes are indicated with an asterisk (*). The effect size (partial eta-squared [η^2]: the proportion of total trait variance attributable to the specific SNP, expressed as percent of trait variance) for significant SNPs is presented in parenthesis following the *P* values.

expressed genes of the intermediary metabolism subpathways (Figure 1). The motif was present in approximately half (38 of 82) of the differentially expressed genes (online-only Data Supplement Table III) in an equal proportion in each of the metabolic subpathways, suggesting that if the motif altered transcription, it was likely to be important in regulation of not only each individual subpathway but also the function of intermediary metabolism as a whole.

Determination of Motif Transcriptional Activity

Promoter/luciferase reporter assays demonstrated that the motif was functional in the promoter position, in both the forward and reverse orientations, but lacked activity in the enhancer position (in either orientation; Figure 2). Classical *cis*-acting proximal promoter elements are typically dependent on both strand orientation and distance from the transcriptional start site, whereas classical *trans*-acting enhancer

Table 4. Meta-Analysis: Extension of Initial Genetic Associations Into Additional Population Samples

							Effect of SNP on Trait		
YY1 RefSNP	Group	Allele_1/ Allele_2	Freq_A1	No.	Trait	Model	Beta (Slope, per allele)	SE of Beta	*P* Value From Meta-Analysis
rs11625658	Cohort 1	A/C	25.3%	996	DBP	Additive	2.180	1.106	…
rs11625658	Cohort 2	A/C	26.2%	1075	DBP	Additive	0.153	0.816	…
rs11625658	ICBP	A/C	23.3%	66741	DBP	Additive	0.140	0.070	…
rs11625658	Meta-analysis	A/C	…	68812	DBP	Additive	0.148	0.069	0.033*
rs11625658	Cohort 1	A/C	25.3%	996	SBP	Additive	2.560	1.373	…
rs11625658	Cohort 2	A/C	26.2%	1075	SBP	Additive	0.540	1.169	…
rs11625658	ICBP	A/C	23.3%	66741	SBP	Additive	–0.074	0.110	…
rs11625658	Meta-analysis	A/C	…	68812	SBP	Additive	–0.052	0.109	0.634
rs11625658	Cohort 1	A/C	25.3%	996	BMI	Additive	–1.217	0.405	…
rs11625658	Cohort 2	A/C	26.2%	1075	BMI	Additive	–0.415	0.303	…
rs11625658	Meta-analysis	A/C	…	2071	BMI	Additive	–0.703	0.242	0.004*
rs11625658	Cohort 1	A/C	25.3%	996	Glucose	Additive	2.776	1.069	…
rs11625658	Cohort 2	A/C	26.2%	1075	Glucose	Additive	0.311	1.129	…
rs11625658	Meta-analysis	A/C	…	2071	Glucose	Additive	1.661	0.802	0.038*
rs11625658	Cohort 1	A/C	25.3%	996	HDL chol	Additive	–0.55	0.944	…
rs11625658	Cohort 2	A/C	26.2%	1075	HDL chol	Additive	0.589	0.821	…
rs11625658	Meta-analysis	A/C	…	2071	HDL chol	Additive	0.098	0.619	0.874
rs11625658	Cohort 1	A/C	25.3%	996	Total chol	Additive	1.28	2.015	…
rs11625658	Cohort 2	A/C	26.2%	1075	Total chol	Additive	–1.559	1.968	…
rs11625658	Meta-analysis	A/C	…	2071	Total chol	Additive	–0.173	1.407	0.902

Meta-analysis was conducted in STATA using fixed-effect models. Individual regression models were additive. Effect sizes (from regression) are given as beta (or slope per allele), ±SE of β. Significant effects ($P<0.05$) are indicated with an asterisk (*).

SNP indicates single nucleotide polymorphism; Cohort 1 and Cohort 2, independent samples from the extremes of a large primary care population; ICBP, International Collaboration on Blood Pressure Genome Wide Association Study; DBP, diastolic blood pressure; SBP, systolic blood pressure; BMI, body mass index; HDL, high-density lipoprotein; chol, cholesterol.

elements are independent of orientation and distance. If we assume the binding of only one transcription factor to the motif, the distance-dependent yet orientation-independent activity of the motif could represent an alternative mode of action for proximal promoter elements. It is also conceivable that the motif consensus sequence represents a composite of ≥2 distinct motifs with each one directing binding of different transcription factors in either the forward or reverse orientation.

Transcription Factor Identification

We then queried TRANSFAC and JASPAR to identify transcription factors that might bind the motif and, therefore, regulate its *cis*-acting transcriptional activity (Table 1). Expression plasmids for these transcription factors (HOXA3, MEF2A, RUNX1, SOX9, SOX17, SRY, and YY1) were cotransfected with the luciferase reporter constructs (with the motif inserted only in the promoter position) to determine which transcription factors could modulate transcription through the motif.

One important caveat of the luciferase experiments was that both the control vector (without motif insert) and the promoter-forward and -reverse constructs (with the motif insert) contained the SV-40 promoter (in pGL3-promoter) to direct basal eukaryotic transcription. Transcription factor effects on the motif were therefore difficult to distinguish from effects on the SV-40 promoter if transcriptional activity was altered in both the control vector (pGL3-promoter) and in the motif promoter-forward or -reverse constructs; such was the case for RUNX1, SOX9, and SOX17, which significantly decreased expression of both the control vector and the promoter-forward and/or promoter-reverse constructs (Figure 4; Table 2). Computational analysis of the SV-40 promoter revealed RUNX1, SOX9, and SOX17 binding sites, so their effects on the control vector were not unexpected.

Thus HOXA3, SRY, and YY1 emerged as the strongest candidates for transcriptional regulation of the motif, because they altered luciferase expression of the motif-containing constructs but lacked effects on the control vector (without a motif insert; Figure 3; Table 2). These 3 candidate transcription factors had distinct patterns of regulation: HOXA3 had no effect on the promoter-forward construct but increased expression of the promoter-reverse construct; SRY increased expression of the promoter-forward but had no effect on the promoter-reverse construct; YY1 increased expression of both the promoter-forward and the promoter-reverse constructs. It is conceivable that the motif consensus sequence might form a composite of HOXA3, SRY, and YY1 motifs, wherein binding specificity to target genes in vivo is dictated by strand orientation, deviation of the actual motif sequence from the consensus (shown in online-only Data Supplement Table III), and the relative abundance and/or activity of endogenous HOXA3, SRY, YY1, and their cofactors in the nucleus.

There is a precedent that inbred rodent models of genetic hypertension such as the BPH exhibit metabolic abnormalities. The spontaneously hypertensive rat (SHR), the most widely studied inbred model of genetic hypertension, was developed in a selection paradigm similar to that of the BPH (ie, selection only on the basis of elevated blood pressure), yet the SHR also exhibits dyslipidemia and insulin resistance and is widely studied as model of the metabolic syndrome.[33,34] Although the BPH has not been as thoroughly investigated as SHR as a model of the metabolic syndrome, the BPH strain exhibits several metabolic abnormalities, including decreased plasma HDL cholesterol, enhanced rate of body weight gain on a high-fat diet, and increased pulse rate.[8] Perhaps pleiotropic genetic variants that affect both blood pressure and metabolism were fixed during the blood pressure selection program for the BPH. If such "master switches" exist, transcription factors (such as HOXA3, SRY, and YY1) would be logical and indeed compelling candidates. Because we identified HOXA3, SRY, and YY1 as candidate transcriptional regulators of perturbed intermediary metabolism in an inbred rodent model of human genetic hypertension, we hypothesized that blood pressure and other metabolic syndrome traits in humans might stem from genetic variation in the HOXA3, SRY, and YY1 loci.

Human SNP Genotyping and Statistical Genetic Association

We then genotyped common HapMap tagging SNPs that captured the linkage disequilibrium structure within the human *HOXA3*, *SRY*, and *YY1* loci. These tagging SNPs were located in noncoding regions (ie, upstream of the gene or within an intron) of our candidate genes. Statistical association of the SNPs with several traits of the metabolic syndrome was tested in a sample of population blood pressure extremes (white males and females from the lowest and highest percentiles of DBP distribution). Although our blood pressure extreme groups were ascertained on a DBP criterion, recent evidence indicates that SBP is at least as important a risk factor for target organ damage; we plan future studies to explore the potential effect of polymorphism in isolated systolic hypertension.

Transcription factor YY1 emerged as the most compelling candidate gene for the human metabolic syndrome. In the BPH, *YY1* was differentially expressed across both age groups (in juvenile prehypertensives and adult hypertensives) as well as in only in the juvenile prehypertensives (data not shown). In cohort 1, *YY1* SNP rs11625658 significantly associated with not only blood pressure (SBP: P=0.0002; DBP: P=0.0006), but also BMI (P=0.006) and carbohydrate metabolism (plasma glucose: P=0.024), although not with lipid metabolism (neither plasma total cholesterol nor plasma HDL cholesterol; Table 3). Furthermore, multivariate analysis of variance demonstrated that rs11625658 genotype significantly associated with SBP, DBP, BMI, and plasma glucose as a joint set of traits (P=0.0005). Finally, meta-analysis (Table 4) indicated that the *YY1* variant rs11625658 continued to predict DBP, BMI, and glucose in a very large number of subjects (n=up to 68 812) with unexpectedly inverse (or opposite) allelic effects on DBP and BMI. Studies of rs11625658 in additional human populations would bolster *YY1* as a candidate gene for hypertension and the metabolic syndrome.

YY1 is a ubiquitous, multifunctional zinc-finger transcription factor with fundamental roles in biological processes such as embryogenesis, differentiation, cellular proliferation, and cell-cycle progression.[35] Indeed, homozygous knockout of the *YY1* gene is lethal.[36] The actions of YY1 are complex, because the transcription factor can directly or indirectly

(through cofactors) activate or repress transcription and can also disrupt binding sites by changing DNA conformation.[35] A large number of genes are regulated by YY1 including genes involved in physiological systems important to the metabolic syndrome such as the nervous system[37] and inflammatory system.[35,37,38] For example, YY1 regulates expression of p53,[35,39] a transcription factor and tumor suppressor that controls cell-cycle progression and the cellular stress response. Adipose expression of p53 plays a role in inflammation and the development of insulin resistance.[40] YY1 regulates the expression of other transcription factors (eg, CREB, c-MYC, and SP1),[35] which, by themselves, can regulate extensive physiological pathways. YY1 could modulate multiple metabolic syndrome traits (eg, SBP, DBP, BMI, and plasma glucose) through a network of direct or indirect gene-by-gene or protein–protein (transcription factor) interactions.

Genetic variation at the *YY1* locus in rats is associated with development of Type 1 diabetes, potentially through its actions on cytokine-related genes.[41] In mice, *YY1* gene expression significantly correlates with cardiovascular function, specifically the P-R wave interval of an electrocardiogram.[42] In addition, two quantitative trait loci for body weight and urinary albumin excretion in the rat have been mapped to the *YY1*-containing region of the human genome using stringently filtered cross-species alignments.[43] Knockout mice expressing graded amounts of YY1 (75%, 50%, 25%) displayed dose-dependent changes in sensitivity to apoptosis, a key component of target organ damage in hypertension.[44] RNA interference strategies have been used to generate transgenic mouse lines that express reduced levels of YY1 protein and exhibit decreased birth weight and size.[45]

Functional genetic variation at the *YY1* locus could thereby provide a unifying genetic mechanism for a portion of metabolic syndrome trait variation in the population.

Conclusions and Perspectives

We developed a novel sequential strategy for identification of candidate genes in the cardiometabolic syndrome. The method integrated several disciplines and tools (animal models, microarray technology, bioinformatics, molecular biology, extreme phenotypes, human genetics) and suggested a previously unexplored gene, *YY1,* as a candidate and potential mechanistic link for several traits of the human metabolic syndrome. Thus, development of novel methods to identify candidate genes might prove advantageous in the quest to understand the genetic basis of complex traits.

Acknowledgments

The authors thank the International Consortium for Blood Pressure Genome Wide Association Studies (ICBP-GWAS) for the data they contributed to this study.

Sources of Funding

Funding was provided by the National Center for Research Resources, M01RR000827, the Comprehensive Research Center of Excellence in Minority Health and Health Disparities (CRCHD), MD00020, the National Institutes of Health/National Institute of Diabetes and Digestive and Kidney Diseases DK007671 Nephrology Training Grant, Department of Veterans Affairs, and the National Institutes of Health (DK094894 and HL58120 to D.T.O.).

Disclosures

None.

References

1. O'Connor DT, Insel PA, Ziegler MG, Hook VY, Smith DW, Hamilton BA, et al. Heredity and the autonomic nervous system in human hypertension. *Curr Hypertens Rep*. 2000;2:16–22.
2. Schlager G. Selection for blood pressure levels in mice. *Genetics*. 1974;76:537–549.
3. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753.
4. Ahituv N, Kavaslar N, Schackwitz W, Ustaszewska A, Martin J, Hebert S, et al. Medical sequencing at the extremes of human body mass. *Am J Hum Genet*. 2007;80:779–791.
5. Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH. Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science*. 2004;305:869–872.
6. Ji W, Foo JN, O'Roak BJ, Zhao H, Larson MG, Simon DB, et al. Rare independent mutations in renal salt handling genes contribute to blood pressure variation. *Nat Genet*. 2008;40:592–599.
7. Schlager G, Sides J. Characterization of hypertensive and hypotensive inbred strains of mice. *Lab Anim Sci*. 1997;47:288–292.
8. Paigen B, Bouchard G, Carey MC. Diet effects on gallstone formation and the assessment of liver morphology, plasma lipids, and atherosclerosis in 43 inbred strains on high-fat atherogenic diet (not under pathogen-free conditions). Mouse Phenome Database Website. Bar Harbor, ME: Jackson Laboratory. Available at: http://phenome.jax.org. Accessed May 7, 2010.
9. Fries RS, Mahboubi P, Mahapatra NR, Mahata SK, Schork NJ, Schmid-Schoenbein GW, et al. Neuroendocrine transcriptome in genetic hypertension: multiple changes in diverse adrenal physiological systems. *Hypertension*. 2004;43:1301–1311.
10. Friese RS, Mahboubi P, Mahapatra NR, Mahata SK, Schork NJ, Schmid-Schönbein GW, et al. Common genetic mechanisms of blood pressure elevation in two independent rodent models of human essential hypertension. *Am J Hypertens*. 2005;18:633–652.
11. Salomonis N, Hanspers K, Zambon AC, Vranizan K, Lawlor SC, Dahlquist KD, et al. GenMAPP 2: new features and resources for pathway analysis. *BMC Bioinformatics*. 2007;8:217.
12. Eskin E, Pevzner PA. Finding composite regulatory patterns in DNA sequences. *Bioinformatics*. 2002;18(suppl 1):S354–363.
13. Eskin E, Keich U, Gelfand MS, Pevzner PA. Genome-wide analysis of bacterial promoter regions. *Pac Symp Biocomput*. 2003:29–40.
14. Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res*. 1990;18:6097–6100.
15. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004;14:1188–1190.
16. Sandelin A, Wasserman WW, Lenhard B. ConSite: web-based prediction of regulatory elements using cross-species comparison. *Nucleic Acids Res*. 2004;32:W249–252.
17. Chekmenev DS, Haid C, Kel AE. P-Match: transcription factor binding site search by combining patterns and weight matrices. *Nucleic Acids Res*. 2005;33:W432–437.
18. Ornatsky OI, Andreucci JJ, McDermott JC. A dominant-negative form of transcription factor MEF2 inhibits myogenesis. *J Biol Chem*. 1997;272:33271–33278.
19. Milsted A, Serova L, Sabban EL, Dunphy G, Turner ME, Ely DL. Regulation of tyrosine hydroxylase gene transcription by Sry. *Neurosci Lett*. 2004;369:203–207.
20. Rana BK, Insel PA, Payne SH, Abel K, Beutler E, Ziegler MG, et al. Population-based sample reveals gene–gender interactions in blood pressure in wWhite Americans. *Hypertension*. 2007;49:96–106.
21. Evans A, Van Baal GC, McCarron P, DeLange M, Soerensen TI, De Geus EJ, et al. The genetics of coronary heart disease: the contribution of twin studies. *Twin Res*. 2003;6:432–441.
22. Kupper N, Willemsen G, Riese H, Posthuma D, Boomsma DI, de Geus EJ. Heritability of daytime ambulatory blood pressure in an extended twin design. *Hypertension*. 2005;45:80–85.
23. Snieder H, Harshfield GA, Treiber FA. Heritability of blood pressure and hemodynamics in African- and European-American youth. *Hypertension*. 2003;41:1196–1201.

24. Seasholtz TM, Wessel J, Rao F, Rana BK, Khandrika S, Kennedy BP, et al. Rho kinase polymorphism influences blood pressure and systemic vascular resistance in human twins: role of heredity. *Hypertension*. 2006;47:937–947.
25. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nat Genet*. 2009;41:677–687.
26. Ehret GB, Munroe PB, Rice KM, Bochud M, Johnson AD, Chasman DI, et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011;478:103–109.
27. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003;100:9440–9445.
28. Lim CK, Ho C, Chou CH, Waye MM. Association of the rs3743205 variant of DYX1C1 with dyslexia in Chinese children. *Behav Brain Funct*. 2011;7:16.
29. Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA, et al. Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet*. 2007;16:36–49.
30. Cui JS, Hopper JL, Harrap SB. Antihypertensive treatments obscure familial contributions to blood pressure variation. *Hypertension*. 2003;41:207–210.
31. Greene LA, Tischler AS. Establishment of a noradrenergic clonal line of rat adrenal pheochromocytoma cells which respond to nerve growth factor. *Proc Natl Acad Sci USA*. 1976;73:2424–2428.
32. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249 796 individuals reveal 18 new loci associated with body mass index. *Nat Genet*. 2011;42:937–948.
33. Pravenec M, Zidek V, Landa V, Simakova M, Mlejnek P, Kazdova L, et al. Genetic analysis of 'metabolic syndrome' in the spontaneously hypertensive rat. *Physiol Res*. 2004;53(suppl 1):S15–22.
34. Shimamoto K, Ura N. Mechanisms of insulin resistance in hypertensive rats. *Clin Exp Hypertens*. 2006;28:543–552.
35. Gordon S, Akopyan G, Garban H, Bonavida B. Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. *Oncogene*. 2006;25:1125–1142.
36. Donohoe ME, Zhang X, McGinnis L, Biggers J, Li E, Shi Y. Targeted disruption of mouse Yin Yang 1 transcription factor results in peri-implantation lethality. *Mol Cell Biol*. 1999;19:7237–7244.
37. He Y, Casaccia-Bonnefil P. The Yin and Yang of YY1 in the nervous system. *J Neurochem*. 2008;106:1493–1502.
38. Rizvi AA. Cytokine biomarkers, endothelial inflammation, and atherosclerosis in the metabolic syndrome: emerging concepts. *Am J Med Sci*. 2009;338:310–318.
39. Sui G, Affar el B, Shi Y, Brignone C, Wall NR, Yin P, et al. Yin Yang 1 is a negative regulator of p53. *Cell*. 2004;117:859–872.
40. Minamino T, Orimo M, Shimizu I, Kunieda T, Yokoyama M, Ito T, et al. A crucial role for adipose tissue p53 in the regulation of insulin resistance. *Nat Med*. 2009;15:1082–1087.
41. Kloting N, Kloting I. Genetic variation in the multifunctional transcription factor Yy1 and type 1 diabetes mellitus in the BB rat. *Mol Genet Metab*. 2004;82:255–259.
42. Grubb SC, Maddatu TP, Bult CJ, Bogue MA. Mouse phenome database. *Nucleic Acids Res*. 2009;37:D720–730.
43. Rapp JP. Genetic analysis of inherited hypertension in the rat. *Physiol Rev*. 2000;80:135–172.
44. Affar el B, Gay F, Shi Y, Liu H, Huarte M, Wu S, et al. Essential dosage-dependent functions of the transcription factor yin yang 1 in late embryonic development and cell cycle progression. *Mol Cell Biol*. 2006;26:3565–3581.
45. Kim J, Kim JD. In vivo YY1 knockdown effects on genomic imprinting. *Hum Mol Genet*. 2008;17:391–401.

CLINICAL PERSPECTIVE

Essential hypertension is a common complex disease with a substantial yet incompletely understood genetic basis. Hypertension often clusters with metabolic abnormalities, which themselves also have genetic underpinnings, in a collection of disorders known as the cardiometabolic syndrome. It is unclear if these phenotypes share common genetic contributors. We developed a novel, integrative method (combining animal models, transcriptomics, bioinformatics, molecular biology, and trait-extreme phenotypes) to identify candidate genes for the cardiometabolic syndrome. Uniquely, we focused on transcription factors as "master switches" because functional changes in them are likely to be pleiotropic and, therefore, might provide a unifying genetic mechanism for multiple traits. A transcriptomic, bioinformatic, and molecular biological analysis of a murine model of genetic hypertension led us to hypothesize that genetic variation at the *HOXA3*, *SRY*, and *YY1* loci might predict blood pressure and other cardiometabolic syndrome traits in humans. Genetic variants for each locus were significantly associated in a human population blood pressure extreme sample with the most extensive associations for *YY1* single nucleotide polymorphism rs11625658 on systolic blood pressure, diastolic blood pressure, body mass index, and fasting glucose. Meta-analysis extended the *YY1* results into 2 additional large population samples with significant effects preserved on diastolic blood pressure, body mass index, and fasting glucose. The results suggest shared genetic contributors for multiple phenotypes of the cardiometabolic syndrome and specifically point to transcription factor YY1 as a potential candidate gene "master switch."