Samantha Kuwada Teixeira

Identificação e análise estrutural e funcional de genes candidatos do cromossomo 4 de ratos *SHR* que possam influenciar a hipertensão essencial

Dissertação apresentada à Faculdade de Medicina da Universidade de São Paulo para a obtenção do título de Mestre em Ciências

Programa de Ciências Médicas

Área de concentração: Distúrbios Genéticos de Desenvolvimento e Metabolismo

Orientador: Prof. Dr. José Eduardo Krieger

São Paulo

2013

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Teixeira, Samantha Kuwada

Identificação e análise estrutural e funcional de genes candidatos do cromossomo 4 de ratos *SHR* que possam influenciar a hipertensão essencial / Samantha Kuwada Teixeira. -- São Paulo, 2013.

Dissertação(mestrado)--Faculdade de Medicina da Universidade de São Paulo. Programa de Ciências Médicas. Área de concentração: Distúrbios Genéticos de Desenvolvimento e Metabolismo.

Orientador: José Eduardo Krieger.

Descritores: 1.Hipertensão/genética 2.Predisposição genética para doença 3.Ratos endogâmicos SHR 4.Animais congênicos 5.Identificação de genes candidatos 6.Expressão gênica

USP/FM/DBD-356/13

Este trabalho foi realizado no LABORATÓRIO DE GENÉTICA E CARDIOLOGIA MOLECULAR – InCor/ HC-FMUSP e recebeu apoio financeiro da Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP. Nº Processo: 2010/12871-8), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação Zerbini.

Aos meus pais e à minha irmã que sempre apoiaram todas as minhas decisões com amor e com compreensão e me deram todo o suporte necessário, emocional e intelectual, para crescer, desenvolver e buscar a excelência tanto na vida pessoal como profissional.

Ao meu companheiro, André Ramos Vaquero, que me incentivou e me acolheu em momentos de aflição e dúvidas, sempre com carinho e dedicação.

Dedico este trabalho.

AGRADECIMENTOS

Agradeço em especial ao Professor Dr. José Eduardo Krieger, que me deu a chance de fazer parte de seu laboratório e de desenvolver este projeto. Agradeço também à oportunidade de trabalhar ao seu lado, aos ensinamentos passados sobre a área de pesquisa e pela confiança, que contribuíram com meu desenvolvimento profissional e intelectual e que me fizeram amar não só este projeto, mas a pesquisa como um todo.

Ao Dr. Alexandre Pereira, por todo o conhecimento passado, principalmente, neste último ano, que me permitiu visualizar áreas da pesquisa antes desconhecidas para mim. Muito obrigada pela paciência, pela compreensão e pelos ensinamentos.

À Mariliza Rodrigues, por todos os ensinamentos técnicos e intelectuais, pela paciência, pela compreensão, pelo companheirismo e pela amizade.

À Camila Bonin, por toda a ajuda no desenvolvimento deste projeto, pelas discussões que contribuíram e contribuem para meu desenvolvimento como pesquisadora e pela amizade.

Ao Professor Dr. Marcelo Morales, pela colaboração que possibilitou a realização do experimento de micro dissecação.

Ao Dr. Pedro Galante e a Bruna Corrêa, que realizaram as análises do RNA-seq.

Ao Edson Moreira, à Mônica Nunes e à Simone Fernandes, pelo cuidado com os animais, pela ajuda nos procedimentos cirúrgicos, pelo carinho e pela disposição em ajudar.

A todos os funcionários do biotério, pelo cuidado com os animais.

À Renata Carmona, por toda ajuda e auxílio no laboratório, pelo carinho e pelos conselhos quando necessários.

À Amanda Brasil e Cristina Thomazella, por sempre estarem dispostas a resolver qualquer problema, pelo carinho, pela atenção e pela amizade.

À Silvana Campos, Maria de Lurdes Junqueira, Luciana Soares, Felícia Ciglio, Ana Maria Piesco, Lúcia Mariano e Sileide Lemos, por toda a ajuda e auxílio durante o desenvolvimento deste projeto, pela atenção e carinho.

Ao Arruda, à Andrea e ao Marcio, pela dedicação e pelo auxílio sempre que necessário.

A todos os colegas de laboratório: Dra. Adriana e ao grupo renal, Dra. Ayumi e ao grupo vascular, ao grupo terapia celular, ao grupo proteômica, ao grupo HFE, ao grupo da genética e ao grupo do Professor Bryan, pela convivência, amizade e considerações relevantes para o desenvolvimento deste projeto.

A minha família, em especial à Rosa, ao Edson, à Giovanna, ao André e ao Alexander, e aos meus amigos, em especial à Dani Oxer, à Karina, à Elisa, à Camila, à Thais, à Fernanda, à Maíra, à Tarsila, à Natalia e às Lauras, pelo companheirismo, amizade, carinho, paciência, compreensão e conselhos.

Agradeço à FAPESP, a CAPES e à Fundação Zerbini, pelo apoio financeiro para realização deste projeto de pesquisa.

Muito obrigada.

“I think there is only one way to do science (...): to meet a problem, to see its beauty and fall in love with it; to get married to it and to live with it happily, till death do ye part – unless you should meet another and even more fascinating problem or unless, indeed, you should obtain a solution.” - Karl Popper, *Realism and the Aim of Science*.

“Happiness only real when shared” – Jon Krakauer, *Into the Wild*.

Esta dissertação está de acordo com as seguintes normas, em vigor no momento desta publicação:

Referências: adaptado de *International Committee of Medical Journals Editors* (Vancouver).

Universidade de São Paulo. Faculdade de Medicina. Divisão de Biblioteca e Documentação. Guia de apresentação de dissertações, teses e monografias. Elaborado por Anneliese Carneiro da Cunha, Maria Julia de A. L. Freddi, Maria F. Crestana, Marinalva de Souza Aragão, Suely Campos Cardoso, Valéria Vilhena. 3^a ed. São Paulo: Divisão de Biblioteca e Documentação; 2011.

SUMÁRIO

LISTA DE SÍMBOLOS	x
LISTA DE ABREVIACÕES	xi
LISTA DE FIGURAS	xiii
LISTA DE TABELAS	xiv
RESUMO	xv
ABSTRACT	xvi
1 – INTRODUÇÃO	1
1.1. Hipertensão Arterial	1
• 1.1.1. Prevalência	1
• 1.1.2. Fatores de risco	2
• 1.1.3. Classificação	2
• 1.1.4. O papel central do rim na regulação a longo prazo da pressão arterial e na fisiopatologia da hipertensão essencial	3
1.2. Estratégias para identificar os determinantes genéticos da hipertensão essencial em humanos	5
1.3. Estratégias para identificar os determinantes genéticos da hipertensão essencial em modelos experimentais	7
2 – OBJETIVOS	13
2.1. Objetivos específicos	13
3 – MATERIAL E MÉTODOS	14
3.1. RNA-seq de amostras de rim de animais congênicos e <i>SHR</i> após sobrecarga salina..	14
• 3.1.1. Extração de RNA Total	14
• 3.1.2. Sequenciamento do RNA	15
• 3.1.3. Análise do RNA-seq por meio de bioinformática	20
3.2. Busca de alterações estruturais entre <i>SHR</i> e <i>BN</i> com base no sequenciamento de seus DNAs genômicos	22
3.3. Combinação das estratégias baseadas em genes diferencialmente expressos entre as linhagens e busca de <i>SNVs</i> e Indels	22
4 – RESULTADOS	24

4.1.	RNA-seq de amostras de rim de animais congênicos e <i>SHR</i> após sobrecarga salina..	24
4.2.	Busca de alterações estruturais entre <i>SHR</i> e <i>BN</i> com base no sequenciamento de seus DNAs genômicos.....	30
4.3.	Combinação das estratégias baseadas em genes diferencialmente expressos entre as linhagens e busca de <i>SNVs</i> e Indels	31
4.4.	Comparação dos resultados do RNA-seq com dados de expressão gênica ao longo do néfron	33
5	– DISCUSSÃO	35
6	– CONCLUSÃO	47
7	– REFERÊNCIAS BIBLIOGRÁFICAS.....	48
APÊNDICES		

LISTA DE SÍMBOLOS

bp	<i>Base pair</i>
Mbp	<i>Mega base pair</i>
%	Por cento
<	Menor que
>	Maior que
°C	Graus Celsius
µg	Micrograma
µl	Microlitro
cM	Centi Morgan
L	Litro
ml	Mililitro
mm Hg	Milímetros de mercúrio
ng/µl	Nanograma por microlitro
nm	Nanômetro
rpm	Rotações por minuto
U/µl	Unidades por microlitro
V	Volts

LISTA DE ABREVIACÕES

BN	<i>Brown Norway</i>
BP-QTLs	<i>Blood Pressure Quantitative Trait Loci</i>
cDNA	Ácido desoxirribonucleico complementar
CDS	<i>Coding DNA Sequence</i>
DEPC	<i>Diethylpyrocarbonate</i>
DNA	Ácido desoxirribonucleico
FDR	<i>False Discovery Rate</i>
FPKM	<i>Fragments Per Kilobase of exon model per Million mapped reads</i>
GWAS	<i>Genome-Wide Association Study</i>
HECT	<i>Homologous to E6-AP (E6-associated protein) C- Terminus</i>
Indels	Inserções e Deleções
miRNAs	microRNAs
NCBI	<i>National Center of Biotechnology Information</i>
NCC	<i>Sodium/chloride transporters</i>
NGS	<i>Next Generation Sequencing</i>
NKCC2	<i>Sodium/potassium/chloride transporters</i>
PAD	Pressão Arterial Diastólica
PAS	Pressão Arterial Sistólica
PCR	Reação de polimerase em cadeia
PPIase	Peptidilprolil cis/trans isomerase
qRT-PCR	Real-Time PCR quantitativo
RCC1	Regulador de Condensação do Cromossomo 1
RNA	Ácido ribonucleico
RNA _m	Ácido ribonucleico mensageiro
RNA _r	Ácido ribonucleico ribossômico
RNA-seq	Sequenciamento do RNA mensageiro
SAGE	<i>Serial Analysis of Gene Expression</i>
SHR	<i>Spontaneous Hypertensive Rats</i>
SNPs	<i>Single Nucleotide Polymorphisms</i>
SNVs	<i>Single Nucleotide Variants</i>

SOLiD	<i>Sequencing by Oligonucleotide Ligation and Detection</i>
SRAA	Sistema Renina-Angiotensina-Aldosterona
U.V.	Ultra Violeta
UCSC	<i>University of California, Santa Cruz</i>
UTR	<i>Untranslated Region</i>
WNK	<i>Lysine deficient protein kinase</i>

LISTA DE FIGURAS

Figura 1. Esquema demonstrando o desenvolvimento de linhagem consômica a esquerda (15) e congênica a direita (48)	9
Figura 2. A - Diferença de pressão arterial basal entre congênico do cromossomo 4 e <i>SHR</i> e manutenção de sensibilidade ao sal no <i>SHR.BN4</i> . B - Localização dos marcadores moleculares no cromossomo 4.....	11
Figura 3. Esquema do método de sequenciamento por ligação com 4 cores utilizado pelo <i>SOLiD</i>	19
Figura 4. Os 310 genes diferencialmente expressos entre <i>SHR</i> e <i>SHR.BN4</i> no RNA-seq segundo sua localização por cromossomo	25
Figura 5. Conjunto de 28 Genes diferencialmente expressos entre <i>SHR</i> e <i>SHR.BN4</i> no intervalo do cromossomo 4 transferido do animal <i>BN</i> para o animal <i>SHR</i>	26
Figura 6. Diferença de expressão gênica em diferentes segmentos do néfron entre congênico <i>SHR.BN4</i> e <i>SHR</i> em amostras adultas de genes selecionados pelo RNA-seq	34

LISTA DE TABELAS

Tabela 1 - Total de genes expressos nas diferentes linhagens, considerando diferentes categorias de corte.....	24
Tabela 2 - Genes diferentemente expressos entre <i>SHR.BN4</i> e <i>SHR</i> no intervalo do cromossomo 4 transferido do animal <i>BN</i> para o animal <i>SHR</i> com base no RNA-seq ²⁷	
Tabela 3 - Vias metabólicas canônicas, rede de processos, doenças e “ <i>Gene Ontology Processes</i> ” obtidos no Metacore a partir dos genes diferencialmente expressos entre congênico e <i>SHR</i> pelo RNA-seq	28
Tabela 4 - Número de <i>SNVs</i> por região genômica	30
Tabela 5 - Número de <i>SNVs</i> no intervalo do congênico por região genômica	31
Tabela 6 - <i>SNVs</i> presentes em regiões promotoras, intrônicas e 3’ <i>UTR</i> em genes que apresentaram diferença de expressão entre as linhagens	32
Tabela 7 - Indels presentes em regiões promotoras, intrônicas e 3’ <i>UTR</i> em genes que apresentaram diferença de expressão entre as linhagens	32
Tabela 8 - <i>SNVs</i> não sinônimos em genes do intervalo expressos no rim	33

RESUMO

Teixeira SK. Identificação e análise estrutural e funcional de genes candidatos do cromossomo 4 de ratos *SHR* que possam influenciar a hipertensão essencial [Dissertação]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2013.

O emprego de “*Total Genome Scan*” em modelos genéticos de doenças complexas tem sido fundamental para seleção de regiões cromossômicas envolvidas com traços complexos. Em nosso laboratório, identificamos cinco regiões cromossômicas associadas ao traço quantitativo pressão arterial (*BP-QTL*) que explicam 43% da variação da pressão arterial numa progênie obtida a partir de animais espontaneamente hipertensos (*SHR*) e “*Brown Norway*” (*BN*). Os *BP-QTLs* foram, então, validados por desenvolvimento de linhagens congênicas, incluindo uma para o cromossomo 4 (*SHR.BN4*) cuja substituição das sequências *SHR* pelo do animal *BN* levou a redução da pressão arterial sistólica basal (~14 mm Hg). O objetivo deste trabalho foi identificar as variantes genéticas candidatas neste intervalo cromossômico com base em diferenças no padrão de expressão gênica e na presença de alterações genéticas não sinônimas “*missense*” ou em regiões regulatórias conservadas que possam estar envolvidas na gênese da hipertensão. Identificamos 533 genes com expressão renal, dentre os 682 do intervalo, sendo que 28 apresentaram padrão de expressão diferente entre amostras de animais adultos (congênico vs. *SHR*) e seis apresentaram alterações não sinônimas “*missense*”. É importante salientar que dos genes diferentemente expressos, encontramos alterações estruturais em regiões conservadas com potencial de participar na regulação em 11. Em conjunto, utilizamos uma plataforma integrada para selecionar 34 genes candidatos no cromossomo 4, dos quais 17 genes serão priorizados, para ser investigados quanto sua contribuição na hipertensão arterial do *SHR* e na hipertensão primária humana.

Descritores: 1. Hipertensão/genética 2. Predisposição genética para doenças 3. Ratos endogâmicos *SHR* 4. Animais congênicos 5. Identificação de genes candidatos 6. Expressão gênica.

ABSTRACT

Teixeira SK. Identification and structural and functional analysis of candidate genes on chromosome 4 in SHR that may influence essential hypertension [Dissertation]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2013.

Total genome scan in genetic models of complex diseases have been instrumental to select candidate genes underlying complex traits. We previously mapped 5 blood pressure related quantitative trait loci (BP-QTLs) that explain about 43% of the BP variance in a progeny derived from Spontaneous Hypertensive Rat (SHR) and Brown Norway (BN) rats. The BP-QTLs were then validated by derivation of congenic strains, including one for chromosome 4 (SHR.BN4) in which a segment from BN replaced the SHR sequences reducing basal systolic BP (~14 mm Hg). The aim of this project is to identify the candidate genetic variants within the chromosome interval based on differences in renal gene expression patterns and structural changes in both non-synonymous missense or within adjacent regulatory sequences that may contribute to hypertension. We identified 533 genes with renal expression, out of 682 in the interval, in which 28 presented differences in expression pattern in adult samples (congenic vs. SHR) and six presented non-synonymous missense alterations. In addition, 11 out of 28 differentially expressed genes showed structural alterations in adjacent conserved regions that potentially contribute to gene regulation. Taken together, using the proposed combination of strategies, we selected 34 hypertensive candidate genes in chromosome 4, in which 17 will be prioritized, to be further explored to assess their contribution to hypertension in the SHR and to essential hypertension in humans.

Descriptors: 1. Hypertension/genetics 2. Genetic predisposition to disease 3. Rats, inbred SHR 4. Congenic strains 5. Candidate genes identification 6. Gene expression, genomic alterations.

1 – INTRODUÇÃO

1.1. Hipertensão Arterial

A pressão sanguínea em humanos é um traço quantitativo que apresenta variações contínuas. Os elevados níveis da pressão arterial estão diretamente relacionados com a predisposição a doenças cardiovasculares, cerebrovasculares e renais, tais como acidente vascular cerebral, doença coronariana arterial, insuficiência cardíaca, doença vascular periférica e insuficiência renal (1)(2)(3)(4). Segundo Lewington e colaboradores, em indivíduos entre 40 e 89 anos, a elevação da pressão arterial a partir de níveis de pressão arterial sistólica (PAS) de 115 mm Hg e diastólica (PAD) de 75 mm Hg está associada com o aumento progressivo e linear do risco de mortalidade por doença coronariana arterial, acidente vascular cerebral, doença renal e outras doenças vasculares (5). A Organização Mundial da Saúde estabelece que pressões arteriais sub ótimas ($PAS > 115$ mm Hg) são responsáveis por 62% das doenças cerebrovasculares e por 49 % das doenças coronarianas isquêmicas, com pequena variação entre os gêneros (2).

Pessoas que apresentam média de PAS maior ou igual a 140 mm Hg, média de PAD maior ou igual a 90 mm Hg e/ou usam medicação anti-hipertensiva são definidos como portadores de hipertensão arterial (2).

• 1.1.1. Prevalência

Considerada pela Organização Mundial da Saúde como a 1ª causa de morte no mundo (1), a hipertensão arterial é uma doença que afeta 1 bilhão de pessoas no mundo e é responsável por, aproximadamente, 7,5 milhões de mortes por ano (13% do total de mortes) (6). No Brasil, ela também configura como um problema de saúde pública, apresentando uma prevalência de 28% (7).

Contudo, a frequência dessa condição varia dependendo de uma série de fatores como idade, etnia, padrões geográficos, sexo ou status socioeconômico (8). Sua prevalência é inversamente proporcional à escolaridade e renda, isto é, quanto maior o grau de instrução e capacidade econômica, menor a incidência devido ao maior cuidado com a saúde. É estimado que o número de pessoas com hipertensão em países em desenvolvimento é quase o dobro em relação aos países desenvolvidos (4).

Mesmo com esforços de muitas organizações, estima-se que metade dos hipertensos desconhece sua condição de doente, dos outros 50% que estão cientes, metade não faz tratamento e dentre os que fazem 55% controla adequadamente a sua pressão arterial. Dessa forma podemos concluir que somente 15% dos hipertensos têm controle efetivo da pressão arterial (2)(3).

- **1.1.2. Fatores de risco**

Diversos fatores ambientais estão implicados no aumento da pressão arterial, dentre estes, os principais fatores são: o tabagismo; a dislipidemia; o diabetes melito; idade superior a 60 anos; sexo: homens e mulheres em fase pós-menopáusicas; histórico familiar para a doença cardiovascular em mulheres com menos de 65 anos e homens menores de 55 anos, entre outros, como a obesidade, a resistência à insulina, o consumo aumentado de álcool, o consumo aumentado de sal, o sedentarismo, o estresse (8).

Além disso, há evidências substanciais da influência genética no fenótipo da hipertensão arterial. Estudos entre gêmeos documentaram maior concordância entre a pressão sanguínea entre gêmeos monozigóticos do que entre dizigóticos (9) e estudos populacionais demonstraram maior similaridade de pressão sanguínea entre familiares do que entre famílias (10). Esta agregação familiar de valores de pressão arterial não é simplesmente atribuída a fatores ambientais, como demonstra estudos de segregação familiar: há correlação significativa entre os níveis pressóricos de irmãos biológicos quando comparados com irmãos adotivos (11)(12).

Estima-se que a herdabilidade da pressão arterial varia de 25 a 68%, dependendo da população estudada (13)(14).

- **1.1.3. Classificação**

As causas que levam o aumento da pressão sanguínea são desconhecidas em 95% dos casos. Indivíduos cuja causa não foi encontrada são descritos como portadores da Hipertensão Primária ou Essencial, que é uma patogenia complexa, onde vários genes são afetados simultaneamente; multifatorial, pois resulta da interação ainda desconhecida entre diversos fatores ambientais e genéticos. Nos restantes 5%, cujos processos que levam a doença são conhecidos, os indivíduos são categorizados como portadores de Hipertensão Secundária ou Monogênica (15). Nestes casos, há mutação em um único gene e os fenótipos se manifestam

precocemente durante a vida. Este tipo de hipertensão difere em muitos aspectos da hipertensão essencial que apresenta alta prevalência, ocorre predominantemente a partir da terceira década de vida e depende da interação entre fatores ambientais e vários genes exercendo efeitos individuais modestos sobre a pressão arterial (16).

- **1.1.4. O papel central do rim na regulação a longo prazo da pressão arterial e na fisiopatologia da hipertensão essencial**

A pressão arterial é determinada por uma gama de fatores físicos, como o volume do sangue, o débito cardíaco e a resistência vascular. Estes elementos são regulados por um conjunto de fatores neurais, endócrinos e parácrinos (17)(18). Envolvidos a eles, há mecanismos de regulação da pressão arterial de ação rápida, como aquelas promovidas pelos barorreceptores, quimiorreceptores e receptores cardiopulmonares que são quase que inteiramente reflexos neurais agudos ou respostas nervosas que agem segundos após as alterações de pressão; de ação a médio prazo, que se deve às propriedades mecânicas dos vasos que são capazes de acomodar seu diâmetro em resposta à mudança de volume sanguíneo e ao deslocamento de líquido para dentro e fora da circulação através das paredes de capilares em resposta à osmose, assim como o sistema renina-angiotensina-aldosterona (SRAA) que influencia a homeostase do volume vascular e também o tônus vascular; e de ação a longo prazo, que demonstram a grande importância dos rins na regulação dos fluídos e do balanço eletrolítico como o maior determinante da manutenção da pressão sanguínea (15)(18)(19)(20).

Segundo Guyton, os sistemas de regulação a curto e a médio prazo da pressão arterial mantêm a pressão próxima ao normal enquanto mecanismos renais controlariam os desvios da pressão arterial inteiramente e por isso exerceriam papel fundamental e seriam os mais eficientes a longo prazo. Simplificadamente, quando a pressão arterial sobe acima do normal, os rins excretam mais água e sal do que a quantidade ingerida, levando a uma diminuição do volume sanguíneo. O contrário ocorre quando há queda da pressão arterial, levando a um aumento do volume sanguíneo.(18). Assim, considerando a teoria Guytoniana, os rins estariam intimamente ligados à fisiopatologia da hipertensão essencial.

Dois conjuntos de evidências corroboram esta ideia de que os rins têm um importante papel na gênese da hipertensão primária ou essencial. A primeira delas se

baseia em estudos de formas mendelianas de hipertensão/hipotensão arterial. Lifton e colaboradores foram pioneiros ao elucidar as bases moleculares e fisiopatológicas dessas doenças, identificando várias mutações específicas em genes envolvidos diretamente com a função de transporte de eletrólitos no túbulo renal. A Síndrome de Gordon ou Pseudoaldosteronismos tipo II, a Doença de Giltelman e a Doença de Bartter tipo I são alguns exemplos de doenças hipertensivas ou hipotensivas com mutações encontradas em genes presentes na via das *WNKs* (do inglês *Lysine deficient protein kinase*), responsável pela regulação de transportadores de sais no túbulo renal, como o *NCC* (do inglês *Sodium/chloride transporters*) no túbulo distal e o *NKCC2* (do inglês *Sodium/potassium/chloride transporters*) na Alça de Henle espessa (21)(22)(23)(24)(25).

O segundo conjunto de evidências baseia-se em estudos clínicos e experimentais de transplante renal, que foram realizados nas últimas três décadas. Na maioria dos casos, o “*background*” genético do rim do doador tem principal impacto na determinação da pressão arterial do receptor. Ou seja, o rim de um doador geneticamente hipertenso aumenta a pressão de um receptor geneticamente normotenso e o rim de um doador geneticamente normotenso reduz a pressão arterial de um receptor geneticamente hipertenso (26).

Em meados dos anos 80, 3 estudos de transplantes foram conduzidos em humanos, um deles cujos receptores tiveram uma redução da pressão arterial após receberem rins de doadores normotensos (27), e dois onde os pacientes receberam rins de doadores com histórico familiar e predisposição a hipertensão e apresentaram a longo prazo hipertensão severa e necessidade de tratamentos com anti-hipertensivos mais rigorosos (28)(29).

Em modelos experimentais, foram utilizadas diferentes cepas de ratos, e em geral os resultados corroboraram a ideia de que a pressão arterial é determinada pelo genótipo do doador e não do receptor (26). Além disso, estudos em animais geneticamente normotensos que receberam rins de animais geneticamente hipertensos na fase pré-hipertensa (30) ou tratados a longo prazo com anti-hipertensivos (26) demonstraram que os receptores assim mesmo receberam o programa genético completo do animal hipertenso, esclarecendo que o aparecimento da hipertensão em animais normotensos submetidos a transplante renal não é uma

resposta a alta pressão de perfusão renal nos doadores antes da remoção do rim e, por consequência, uma resposta ao dano causado ao rim antes do transplante.

1.2. Estratégias para identificar os determinantes genéticos da hipertensão essencial em humanos

A dificuldade em identificar as causas da hipertensão essencial somente a partir de estudos fisiológicos e o conhecimento da contribuição genética nessa fisiopatologia motivou a aplicação de técnicas de genômica para identificar seus determinantes genéticos. Na última década, com o sequenciamento completo do genoma humano e o desenvolvimento de ferramentas de genômica, abrem-se novas oportunidades para entendimento dos determinantes genéticos da hipertensão arterial.

Classicamente três estratégias têm sido utilizadas para mapear as variantes genéticas em humanos: estudos de associação em genes candidatos, estudos de ligação e estudos de associação por meio da varredura completa do genoma.

Nos estudos de associação com genes candidatos, buscam-se evidências de associação entre variantes em genes envolvidos na regulação da pressão arterial e a hipertensão arterial. Partindo dos achados de mutações em genes relacionados com o transporte de eletrólitos no rim em doenças hipertensivas ou hipotensivas monogênicas, pesquisadores não mediram esforços para identificar variantes genéticas associadas à hipertensão essencial não só em canais, transportadores e trocadores de íons no rim, mas também em genes presentes na sinalização da aldosterona, na via das catecolaminas, em genes relacionados com o sistema nervoso simpático, com o sistema renina-angiotensina, com a vasoconstrição e com a inflamação. No entanto, observou-se que, para obter resultados de associação, é necessário um tamanho amostral extenso, sendo que a contribuição de cada variante é bastante modesta. Além disso, a partir dessa estratégia, perde-se a possibilidade de identificar novos genes, tornando-a bastante limitada (15)(31)(32).

Nos estudos de ligação e de associação por meio da varredura completa do genoma (do inglês *Genome-Wide Association Study – GWAS*), não há uma hipótese a priori e há uma investigação sistemática do genoma para identificar genes ou vias associadas com o traço estudado. No entanto, nos *GWAS*, assim como nos estudos de associação em genes candidatos, testa-se a associação entre um alelo e um traço nos indivíduos, neste caso, a variação na pressão arterial e/ou a hipertensão. No caso dos

estudos de ligação, testa-se a co-segregação de regiões gênicas e um traço em um núcleo familiar (31). Dispondo-se de núcleos familiares informativos e marcadores polimórficos, procura-se identificar aqueles marcadores que acompanham os indivíduos afetados e assim localizar uma região cromossômica que deve conter o(s) gene(s) responsável(eis) pelo fenótipo estudado (15). Com essa estratégia, muitas regiões gênicas ligadas ao traço de pressão arterial (do inglês *Blood Pressure Quantitative Trait Loci - BP-QTLs*) foram identificadas, mas em geral eram regiões bastante extensas (entre 15 a 40 *Mbp*), com muitos genes candidatos a serem estudados (31). Esta estratégia foi muito útil para mapear genes responsáveis por doenças monogênicas, cujas mutações resultam em defeitos de grande magnitude que explicam a totalidade do fenótipo encontrado, porém ela é limitada para investigar doenças complexas, onde, aparentemente, estão envolvidos muitos genes com contribuições individuais modestas e que interagem entre si.

Nos *GWAS*, como falado anteriormente, busca-se uma associação entre variantes genéticas, que na maioria dos casos são polimorfismo de um único nucleotídeo (do inglês *single nucleotide polymorphisms – SNPs*), e o traço estudado. Este tipo de estratégia ganhou força com o surgimento de metodologias que permitem a genotipagem de uma grande quantidade de variantes dispersas no genoma em um tamanho amostral extenso, como os “*chips*” e lâminas de genotipagem (por exemplo, os “*chips*” de genotipagem da Affymetrix ou as lâminas de Open Array da Applied Biosystems). Em um estudo publicado em 2007, onde se buscou identificar variantes genéticas com associação em sete doenças comuns que afetam a população mundial, pesquisadores falharam em identificar *SNPs* associados à hipertensão essencial, evidenciando que para esta patologia, tal estratégia teria suas limitações (33). Posteriormente, outros *GWAS* foram conduzidos em diferentes grupos populacionais, existindo até o momento mais de 50 genes identificados por apresentarem variantes associadas à hipertensão essencial e a regulação da pressão arterial. No entanto, essas variantes em geral têm um pequeno efeito sobre o fenótipo de pressão arterial (contribuem com <1 mm Hg) e, em conjunto, elas explicam menos de 2,5% da variação da pressão arterial em populações humanas (33)(34)(35)(36)(37)(38)(39)(40).

Dessa forma, o sucesso na identificação de determinantes genéticos para hipertensão em humanos tem sido pouco expressivo e os resultados conflitantes. As razões para isto são várias e incluem desde problemas metodológicos, como o número inadequado de indivíduos estudados, problemas na fabricação dos “chips” utilizados na genotipagem dos indivíduos e a má classificação dos controles; como a heterogeneidade genética e a diversidade étnica na população humana, que impedem a replicação dos resultados e diminuem a probabilidade de sucesso dos estudos; sem contar com as interações gênicas existentes e a influência ambiental presente nesta doença complexa (31). Dessa forma, faz-se necessário melhorar as estratégias vigentes ou testar alternativas para identificação dos genes e vias envolvidas com a hipertensão essencial.

1.3. Estratégias para identificar os determinantes genéticos da hipertensão essencial em modelos experimentais

Devido à grande dificuldade em identificar as variantes genéticas que contribuem para a hipertensão em humanos, muitos pesquisadores passaram a investigar esses fatores em alguns modelos experimentais, tais como ratos, camundongos e hamsters. Muitos modelos de animais foram desenvolvidos para a hipertensão arterial, como por exemplo, os ratos espontaneamente hipertensos (do inglês *Spontaneous Hypertensive Rats - SHR*), ratos sensíveis à sobrecarga salina (*Dahl-S*), entre outros, onde as características ligadas à hipertensão arterial são geneticamente determinadas. A utilização de modelos experimentais consiste de uma ferramenta importante, pois estes mimetizam certos aspectos da doença em humanos, solucionam o problema da heterogeneidade presente na população humana, que mascara alguns resultados, e permitem a produção em larga escala de progênie de interesse. Atualmente, há disponível mais de 10 linhagens de ratos hipertensos, que carregam combinações de traços observados em distúrbios de hipertensão essencial em humanos (15)(41).

A partir da década de 90, a disponibilidade de marcadores genéticos para o genoma do rato, tais como microssatélites (repetições de sequência simples) e *SNPs*, aumentou consideravelmente, levando a viabilização da utilização da estratégia de varredura total do genoma (do inglês *total genome scan*). Nesta, assim como nos estudos de ligação em humanos, marcadores polimórficos permitem que o genoma

inteiro de uma espécie possa ser interrogado com o objetivo de localizar um intervalo cromossômico associado a um traço quantitativo (*QTLs*) (42). Geralmente, duas subespécies isogênicas que apresentam diferenças significativas em um traço quantitativo (como, por exemplo, a pressão arterial) são avaliadas utilizando-se como base cruzamentos controlados para mistura aleatória dos genomas parentais (tal como uma população F2 de um cruzamento entre uma subespécie de rato normotensa e uma hipertensa). Nos indivíduos desta população, testa-se a associação entre os marcadores genéticos e o fenótipo estudado, ou seja, as regiões cromossômicas e a variação do fenótipo de interesse. Muitos pesquisadores obtiveram sugestivos e/ou significativos *QTLs* de pressão sanguínea em ratos em muitos cromossomos utilizando-se dessa técnica (15)(41)(43)(44)(45)(46)(47).

No entanto, essa estratégia revela uma região do cromossomo muito ampla para identificação de genes candidatos à hipertensão essencial. Além disso, o suporte estatístico para a existência do *QTL* para pressão sanguínea é bastante problemático, já que o “*background*” genético não é estabelecido e, dessa forma, o fenótipo está sob a influência de outras regiões cromossômicas não controladas (41).

Para resolver este impasse, foram construídas linhagens consômicas e congênicas. Estas linhagens são obtidas através de cruzamentos sucessivos entre duas linhagens de animais que possuem fenótipos relevantes para o estudo da hipertensão essencial (como, por exemplo, uma cepa hipertensa e outra normotensa, em geral, as mesmas utilizadas para identificar os *QTLs* para pressão arterial) e da técnica de seleção de animais assistida por marcadores moleculares. Com esta metodologia, permite-se à transferência de um cromossomo de uma linhagem doadora (como, por exemplo, a normotensa) a outra receptora (como, por exemplo, a hipertensa), gerando uma linhagem consômica, ou somente de uma parte específica do cromossomo, gerando uma linhagem congênica (Figura 1) (15).

Figura 1. Esquema demonstrando o desenvolvimento de linhagem consômica a esquerda (15) e congênica a direita (48)

A criação desses animais é muito importante, pois é um excelente passo para obter evidências fisiológicas do *QTL* investigado, fixar os limites que carregam o *QTL* e estreitar o tamanho da região para produção de um mapa físico adequado. Além disso, ela oferece uma oportunidade única de identificar variantes genéticas envolvidas com a hipertensão essencial, já que, caso a linhagem consômica ou congênica apresente diferenças fenotípicas em relação à linhagem receptora do fragmento, geneticamente, essa diferença se deve aos genes presentes no cromossomo ou parte dele transferido da linhagem doadora (49)(50)(51)(52)(53)(48)(54).

A partir de uma linhagem consômica ou congênica, três estratégias são possíveis: 1) diminuir a região do *QTL* em fragmentos menores àqueles presentes na linhagem consômica ou congênica, a partir do desenvolvimento de sublinhagens, a fim de adquirir um mapa físico dos genes relacionados à hipertensão essencial; 2) o estudo detalhado dos possíveis genes candidatos presentes no *QTL*; 3) o estudo fisiológico, a partir do qual fenótipos intermediários diretamente associados a genes envolvidos na regulação da pressão arterial são explorados (48).

Em um estudo realizado em 1995 em nosso laboratório (55), utilizando a técnica de “total genome scan”, foram analisados 235 animais da população F2 proveniente do inter cruzamento da linhagem de ratos normotensos “*Brown Norway*” (BN) com a cepa de ratos espontaneamente hipertensa (SHR). Nesse estudo, buscou-se relacionar 24 fenótipos cardiovasculares e 336 marcadores moleculares, sendo possível a identificação de cinco regiões cromossômicas (duas no cromossomo 2 e

uma nos cromossomos 4, 8 e 16) que explicavam 43% da variação na pressão arterial sistólica destes animais após sobrecarga salina.

A fim de testar a hipótese de que estas cinco regiões descritas anteriormente contêm genes associados ao desenvolvimento da hipertensão essencial, iniciou-se o desenvolvimento de cinco linhagens de animais congênicos que contêm o “background” genético do animal *SHR*, exceto pela região mapeada que foi transferida do animal normotenso *BN*. Este processo foi realizado por meio de retrocruzamentos sucessivos entre um animal que era heterozigoto *BN/SHR* para todas as regiões do genoma com um animal *SHR*. Através da seleção dos animais assistida por marcadores moleculares, transferiu-se a região desejada, proveniente do animal normotenso para o genoma do animal hipertenso, obtendo-se, no final, animais fundadores homozigotos para a região transferida. Em aproximadamente 13 gerações foram estabelecidas 4 linhagens denominadas de acordo com a região cromossômica transferida (*SHR.BN2A*, *SHR.BN2C*, *SHR.BN4* e *SHR.BN16*) sendo que a linhagem de congênicos para o cromossomo 8 foi perdida ao longo do processo de reprodução e não foi completada. A descrição fenotípica inicial desses animais quanto ao comportamento da pressão arterial basal e após sobrecarga salina (1% na água de beber), demonstrou que aqueles que receberam o fragmento cromossômico da cepa normotensa apresentaram alterações de fenótipo quando comparado com animais *SHR* na pressão arterial basal, na sensibilidade ao sal ou em ambos (56). Este dado é muito importante, pois confirma a hipótese de que as regiões mapeadas de fato contêm genes associados ao desenvolvimento da hipertensão.

Entre as quatro linhagens de ratos congênicos por nós desenvolvidos, os animais *SHR.BN4* referentes ao cromossomo 4 apresentam redução da pressão arterial basal (~14 mm Hg) quando comparados ao animais controle (*SHR*), mantendo a sensibilidade ao sal (Figura 2A). Por meio de informações contidas em bancos de dados públicos, ficou confirmado que a região transferida do animal normotenso para o hipertenso possui 20,94 cM e compreende 56 *Mbp*, sendo delimitada pelos seguintes marcadores moleculares: D4rat163, D4rat87 (ambos externos ao intervalo), D4rat128 e D4rat185 (ambos internos ao intervalo) (Figura 2B).

Figura 2. A - Diferença de pressão arterial basal entre congênico do cromossomo 4 e *SHR* e manutenção de sensibilidade ao sal no *SHR.BN4*. * p<0,05 vs. *SHR*; # p<0,05 vs. condição basal. B - Localização dos marcadores moleculares no cromossomo 4. A região em vermelho representa a região cromossômica previamente mapeada e transferida, por meio de cruzamentos, do animal normotenso (*BN*)

Dessa forma, no presente projeto, adotamos 2 estratégias principais para **selecionar genes/variantes candidatas no intervalo do congênico *SHR.BN4***: 1) Diferença de expressão gênica; 2) Análises de variantes na região de interesse com base no sequenciamento do DNA genômico dos parentais *SHR* e *BN*.

A primeira estratégia se baseou na análise da diferença de expressão gênica, utilizando uma ferramenta inovadora: o sequenciamento de próxima geração (do inglês *Next Generation Sequencing* - *NGS*). Durante o desenvolvimento do projeto as plataformas de *NGS* evoluíram rapidamente e se tornaram uma alternativa poderosa para estudos de genômica estrutural e funcional. Essas tecnologias promovem o sequenciamento de DNA e são capazes de gerar informações sobre milhões de pares de bases em uma única corrida. Dessa forma, realizou-se o RNA-seq de amostras renais dos animais congênicos e *SHR*. Esta tecnologia permite a identificação de genes diferencialmente expressos entre as linhagens estudadas, assim como a confirmação de alterações estruturais nas sequências de RNAm previamente identificadas por meio do sequenciamento do DNA genômico.

Na segunda estratégia, com base no sequenciamento do DNA genômico de animais *SHR* e *BN* disponíveis em bancos de dados públicos, buscou-se identificar variantes de um único nucleotídeo (do inglês *Single Nucleotide Variants* - *SNVs*) e inserções e deleções (Indels) entre os animais na região de interesse. Buscava-se, com isso, identificar alterações estruturais entre as sequências dos animais *BN* e *SHR* nos genes selecionados pelo RNA-seq que explicassem suas diferenças de expressão, assim como alterações estruturais em regiões codificadoras de proteínas em genes presentes nas regiões das linhagens congênicas.

Acredita-se que utilizando estas metodologias serão selecionados alguns genes candidatos, que acumularão evidências de estarem envolvidos no desenvolvimento da hipertensão essencial nesses animais.

2 – OBJETIVOS

Selecionar genes candidatos no intervalo do cromossomo 4 que influenciam a gênese da hipertensão arterial no *SHR* utilizando o congênico do cromossomo 4.

2.1. Objetivos específicos

1. Identificar os genes diferencialmente expressos no intervalo do cromossomo 4 referente ao QTL de pressão arterial mapeado utilizando o RNA-seq do rim dos animais *SHR* e congênicos por meio de técnica da *NGS* por meio da plataforma *SOLiD* (do inglês *Sequencing by Oligonucleotide Ligation and Detection*);
2. Identificar alterações genéticas estruturais que expliquem as diferenças de expressão encontradas nos genes candidatos identificados no objetivo 2.1.1 e aquelas que alterem as regiões codificadoras de genes do intervalo independente do padrão de expressão gênica entre o congênico e o *SHR*.

3 – MATERIAL E MÉTODOS

3.1. RNA-seq de amostras de rim de animais congênicos e SHR após sobrecarga salina

Foram coletadas amostras de rins de 8 animais adultos submetidos à dieta hipersódica (1% de NaCl na água de beber durante duas semanas) de ambas as linhagens (congênico e *SHR*).

• 3.1.1. Extração de RNA Total

A extração e a preparação do RNA total de amostras renais dos animais compreenderam quatro fases:

1) 100 mg de tecido renal armazenado congelado em freezers -80 °C foram transferidos para um tubo de 50 ml (Falcon) e homogeneizados com 1 ml de TRIzol[®] (Gibco/BRL, Invitrogen, Carlsbad, CA) por aproximadamente 30 segundos no gelo;

2) Depois de homogeneizadas, as amostras foram colocadas em tubos de 1,5 ml (Eppendorf) e receberam 200 µl de clorofórmio, sendo vigorosamente misturadas. Estas, então, foram centrifugadas a 10.000 rpm por 15 minutos a 4 °C para que o RNA total fosse separado do DNA e das proteínas. A fase aquosa superior que continha o RNA total foi retirada e transferida para um novo tubo de 1,5 ml previamente identificado;

3) Para a precipitação do RNA foi adicionado a cada tubo 500 µl de isopropanol que, em seguida, foram invertidos suavemente, para serem centrifugados a 10.000 rpm por 45 minutos a 4 °C. Após a centrifugação, o sobrenadante foi desprezado e as amostras foram lavadas com 1 ml de etanol 75% gelado diluído em água isenta de RNase, sendo novamente centrifugadas a 10.000 rpm por 10 minutos a 4 °C. O sobrenadante foi novamente desprezado e os tubos foram invertidos e colocados para secar a 4 °C;

4) Depois de seco, o precipitado de RNA foi ressuscitado em 150 µl de água livre de RNase, sendo sua concentração determinada a partir da medida de absorbância a 260 nm em espectrofotômetro (ND-1000 spectrophotometer - NanoDrop Technologies, Inc.). As soluções de RNA foram aliquotadas e armazenadas a -80 °C até o uso.

A água livre de RNase (água *DEPC*) utilizada nos procedimentos acima descritos foi previamente preparada, sendo que para 1L de H₂O deionizada, adicionou-se 1 ml de Dicarbonato de dietila (do inglês *Diethylpyrocarbonate* – *DEPC*). Esta foi incubada a temperatura ambiente por 12 horas, seguida por autoclavagem.

A integridade das amostras de RNA foi verificada através da eletroforese em gel de agarose a 0,8% contendo 0,5 µg/ml de Brometo de etídio. O gel foi imerso em tampão TAE 1X e a eletroforese realizada a 80 Volts por aproximadamente 20 minutos. Após o término da corrida, as bandas foram visualizadas por um sistema que combina exposição à luz U.V. acoplada a uma câmara digital de aquisição de imagem (Quantum ST4-1100/26MX).

Apenas as preparações íntegras e sem indício de degradação, como falta de integridade das bandas 28S RNAr e 18S RNAr e diminuição da razão de fluorescência das bandas ribossomais 28S RNAr/18S RNAr (razão para o RNA intacto igual aproximadamente a 2), foram utilizadas no projeto.

Amostras de RNA total de 4 animais de cada grupo foram unidas para formar 2 pools de amostras de cada linhagem. Estas, então, foram submetidas ao sequenciamento do RNAm.

• 3.1.2. Sequenciamento do RNAm

Quatro etapas são necessárias para realização do RNA-seq: a fragmentação do RNA mensageiro; construção da biblioteca de cDNA; ligação do cDNA a “beads” e PCR de emulsão; e o sequenciamento baseado na ligação e detecção de oligonucleotídeos.

3.1.2.1. Fragmentação do RNAm

Inicialmente, foi utilizada uma quantidade de RNA total de 80 µg por pool. As 4 amostras, 2 da linhagem congênica *SHR.BN4* e 2 da linhagem *SHR*, foram submetidas à depleção de RNA ribossômico com o kit RiboMinus™ Eukaryote para RNA-Seq (Invitrogen) e seleção de RNAm maduro pela cauda poli-A com o kit Applied Biosystems MicroPoly(A)Purist™ segundo recomendações dos fabricantes. Após a depleção do RNAr e da seleção de RNAm maduro, 500 ng de cada RNAm purificado foi submetido a fragmentação com o kit SOLiD™ Total RNA-Seq, onde foi utilizado 8 µl de RNAm purificado, 1 µl de 10X RNase III Buffer e 1 µl de

RNAse III, totalizando uma reação de 10 µl. As amostras, então, foram para o termociclador, onde ficaram 10 minutos a 37 °C. Imediatamente após a incubação foi adicionado 90 µl de H₂O livre de nuclease em cada pool de amostras, sendo que estas foram mantidas no gelo.

Após a fragmentação, as amostras foram purificadas e concentradas com o kit RiboMinus™ Concentration Module (Invitrogen) segundo recomendações do fabricante. O rendimento do RNAm fragmentado foi, então, analisado através do kit Quant-iT RNA Assay no Qubit Fluorometer (Invitrogen) e a distribuição dos fragmentos foi avaliada usando o Agilent 2100 bioanalyzer com o kit RNA 6000 Pico Chip (Agilent).

3.1.2.2. Construção da biblioteca de cDNA

Os seguintes procedimentos são necessários para a construção da biblioteca de cDNA: hibridização e ligação do RNA, reação de transcrição reversa, purificação do cDNA, seleção de cDNA com 150-250 nucleotídeos, amplificação do cDNA, purificação do DNA amplificado e avaliação do rendimento e da distribuição dos fragmentos de DNA amplificados.

A hibridização e ligação do RNA fragmentado, assim como a reação de transcrição reversa foram realizadas utilizando o kit SOLiD™ Total RNA-Seq, seguindo as especificações do fabricante. Para purificação do cDNA, foi utilizado o kit MinElute® PCR Purification (Qiagen). Para seleção de cDNAs com 150 a 250 nucleotídeos, foi utilizado o kit Novex® pre-cast gel products (Invitrogen), 50 bp DNA Ladder (Invitrogen) e SYBR® Gold nucleic acid gel stain (Invitrogen), sendo que as amostras correram em gel em cuba de eletroforese a 180 V a, aproximadamente, 30 minutos para separação dos diferentes fragmentos. Utilizando SYBR® Gold nucleic acid gel stain e um sistema que combina exposição à luz U.V. acoplada a uma câmara digital de aquisição de imagem (Quantum ST4-1100/26MX), cortou-se a região do gel que possuía os cDNAs com tamanhos entre 150-250 nucleotídeos.

A amplificação do cDNA foi realizada utilizando os kits SOLiD™ Total RNA-Seq e SOLiD® RNA Barcoding, Module 1-16, já que queríamos realizar o sequenciamento de todas as amostras conjuntamente. O sistema de código de barras nada mais é que a inserção de uma sequência de nucleotídeos específica antes do

primer 3' comum utilizado na amplificação do cDNA. Cada pool de amostras recebeu um primer 3' específico antes da reação. Como foi realizado um sequenciamento “*paired-end*”, ou seja, sequenciou-se no sentido 5'-3' e no sentido 3'-5', tanto na sequência do primer 3', como do primer 5' havia um adaptador onde se ligou o iniciador do sequenciamento.

Para purificação dos pools de amostras utilizou-se o kit PureLink™ PCR Micro Kit (Invitrogen). A concentração do DNA purificado foi determinada em espectrofotômetro (ND-1000 spectrophotometer- NanoDrop Technologies, Inc.) e a distribuição dos fragmentos foi avaliada utilizando-se o Agilent 2100 Bioanalyzer com o kit DNA 1000.

3.1.2.3. Ligação do cDNA a “*beads*” e PCR de emulsão

Cada pool de amostra foi submetido ao PCR de emulsão separadamente. A fase aquosa e a fase oleosa da emulsão foram preparadas separadamente. A fase oleosa continha emulsificantes e a fase aquosa continha os componentes necessários para uma reação de PCR: a amostra, primers para ligação aos adaptadores (P1 e P2), DNA polimerase e os SOLiD™ P1 DNA “*beads*”, “*beads*” ligadas a primers P1. A emulsão foi criada utilizando um ULTRA-TURRAX® da IKA®. Nela, formou-se microbolhas (micro reatores) contendo um “*bead*”, uma molécula de DNA e os reagentes necessários para uma PCR convencional. A emulsão então foi submetida à amplificação em termociclador convencional, de maneira que ocorreu a amplificação de uma única molécula de DNA em cada micro reator. Durante o PCR de emulsão, 30.000 ou mais cópias da fita de DNA presa a um “*bead*” é amplificada. Após o PCR de emulsão, os micro reatores foram quebrados adicionando solução de 2-butanol e as “*beads*” ligadas ou não ao DNA foram, então, lavadas para a retirada do óleo e dos emulsificadores. A seguir, foi realizado o enriquecimento de “*beads*” ligadas a fitas de DNA. Neste passo, “*beads*” de poliestireno ligadas ao primer P2 (chamadas de “*beads*” de enriquecimento) foram adicionadas as soluções e usadas para captura de “*beads*” com fitas de DNA. As soluções contendo “*beads*” de enriquecimento, “*beads*” enriquecidas com DNA e “*beads*” sem DNA foram, então, centrifugadas após o acréscimo de glicerol 60%, fazendo com que os “*beads*” sem DNA fossem precipitados, enquanto que os “*beads*” com DNA permaneceram numa camada do sobrenadante ligados aos “*beads*” de poliestireno. Essa camada foi, então, extraída e

desnaturada para a dissociação dos “*beads*” de enriquecimento e dos “*beads*” ligados a DNA. Com a finalidade de preparar as “*beads*” enriquecidas com fitas de DNA para a deposição e sequenciamento, foi adicionada desoxiuridina trifosfato (dUTP) na porção 3’ dos primers P2 ligados às fitas de DNA usando reação de transferase terminal. Todo o procedimento foi realizado segundo recomendações do fabricante, seguido o guia Applied Biosystems SOLiD™4 System Templated Bead Preparation.

3.1.2.4. Sequenciamento baseado na ligação e detecção de oligonucleotídeos

As “*beads*” ligadas a fitas de DNA com a região 3’ modificada foram depositadas em uma lâmina de vidro. Esta foi sequenciada pelo SOLiD4. O sequenciamento de cada amostra (de cada pool) ocorreu em triplicata, obtendo, ao final, doze sequenciamentos (replicata biológica=2, replicata técnica=3).

No *SOLiD* o sequenciamento baseia-se na ligação de oligos marcados com corantes diferentes. Neste caso, a fita complementar ao DNA molde não é sintetizada através de polimerização, mas sim por sucessivas ligações de oligos contendo 8 nucleotídeos cada. A ligação dos oligos à fita que está pareada ao DNA molde depende da complementaridade entre os mesmos e a fita molde. Para tanto, uma biblioteca contendo oligos com todas as combinações possíveis de nucleotídeos nas 5 primeiras posições é utilizada em cada reação de ligação ($4^5 = 1024$ oligos diferentes; Figura 3A). Já as últimas três bases são nucleotídeos modificados que fazem o pareamento com qualquer outro nucleotídeo presente na fita molde. Além disso, são estes os nucleotídeos marcados com os diferentes corantes. A reação inicia-se com a utilização de um iniciador específico que se associa à fita molde e posteriormente apenas um oligo se liga ao mesmo (Figura 3A). O sinal do corante presente em tal oligo é captado e em seguida as 3 bases marcadas da porção 5’ do oligo são removidas enzimaticamente, restando assim apenas os 5 nucleotídeos que apresentam complementaridade com a fita molde. O processo se repete com novas ligações até um número máximo de 10 ligações por iniciador. Uma vez que o corante utilizado na marcação dos oligos é definido apenas pelos 2 primeiros nucleotídeos de cada oligo (Figura 3B), ao final deste primeiro ciclo de ligações, o que se obtém são informações apenas dos nucleotídeos: 1 e 2; 6 e 7; 11 e 12; 16 e 17; e assim sucessivamente até os nucleotídeos 46 e 47.

Figura 3. Esquema do método de sequenciamento por ligação com 4 cores utilizado pelo *SOLiD*. A - Exemplo das sondas (1,2-probes) constituídas por 8 nucleotídeos, sendo os 3 últimos bases modificadas (z). Esquema de um “bead” com uma fita molde, o iniciador (primer) de sequenciamento e a ligação de uma sonda, seguida pela detecção da fluorescência e clivagem das últimas 3 bases. Ao final, podemos observar o resultado obtido após o primeiro ciclo de sequenciamento. No segundo ciclo, liga-se a fita molde um iniciador uma posição acima do anterior. Diferente da polimerização, a ligação da sonda ao iniciador pode ocorrer bidirecionalmente, tanto no final 5’-PO₄, quanto no final 3’-OH, permitindo assim o sequenciamento “paired-end”. B - Código de cores dos corantes que são determinados pelas duas primeiras bases das sondas e montagem da sequência de nucleotídeos (57).

Para a obtenção da sequência de todos os nucleotídeos, outros ciclos de ligações são feitos, mas utilizando-se iniciadores que diferem na posição em que se associam ao adaptador utilizado na construção da biblioteca que está sendo sequenciada. No total são utilizados 5 iniciadores diferentes com os quais é possível obter as informações sobre todos os nucleotídeos da fita molde (Figura 3A) (57).

• 3.1.3. Análise do RNA-seq por meio de bioinformática

Para a análise dos resultados obtidos a partir do sequenciamento, foi utilizado como genoma referência o rn4 (do animal *BN*) para alinhamento das sequências, obtido no *UCSC Genome Browser* (<http://genome.ucsc.edu>). Para anotar os transcritos, foi utilizado o banco de dados Ensemble Release 69 (<http://www.ensembl.org>).

O alinhamento foi realizado com o programa TopHat 2 (versão 2.0.8) (<http://tophat.cbcb.umd.edu/>) (58)(59), que utiliza a maquinaria do programa Bowtie 2 (versão 2.0.0) (<http://bowtie-bio.sourceforge.net/index.shtml>) (60). Este permite o armazenamento e a pesquisa do genoma referência de forma econômica e rápida. Para o alinhamento, foi utilizado somente sequências com qualidade mínima (Q) de 20, ou seja, cuja probabilidade de erro de uma base é de 1 em 100. As replicatas técnicas de cada pool foram unidas no alinhamento (61).

Em seguida, os arquivos resultantes do alinhamento foram fornecidos ao Cufflinks 2 (versão 2.1.0) (<http://cufflinks.cbcb.umd.edu/>) (62) para gerar a montagem do transcriptoma para cada pool de amostras. Os dois pools de cada linhagem foram, então, unidos, utilizando a ferramenta Cuffmerge, gerando um arquivo para o *SHR* e um arquivo para o *SHR.BN4*. A união dos arquivos é necessária para fornecer uma base uniforme para calcular a expressão gênica em cada grupo de animais (experimento e controle) e se há diferenças entre eles.

Para calcular o nível de expressão gênica, assim como identificar genes diferencialmente expressos entre *SHR* e *SHR.BN4*, foram utilizadas duas ferramentas, o Cuffdiff do pacote Cufflinks e o DEseq (63), e foram considerados somente resultados que eram concordantes entre as ferramentas. Foram contemplados como genes expressos aqueles que apresentaram uma ou mais sequências em pelo menos um exon do gene.

Para as análises realizadas pelo Cuffdiff, a expressão gênica foi dada em *FPKM* (do inglês *Fragments Per Kilobase of exon model per Million mapped reads*), onde $FPKM = 10^9 \times (\text{número de fragmentos mapeados no gene}) / (\text{total de fragmentos mapeados na amostra}) \times (\text{somatória dos exons do gene em bp})$ (64). Foram considerados fragmentos duas sequências que se sobrepunham, já que o sequenciamento foi realizado em “*paired-end*”. Para as análises realizadas pelo DEseq, a expressão gênica foi dada em “*baseMean*”, que é uma média normalizada de sequências de cada gene.

Foram considerados genes diferencialmente expressos entre as linhagens aqueles que apresentaram p-valor menor ou igual a 0,05 e *FDR* (do inglês *False Discovery Rate*) menor ou igual a 0,05, através do teste-t de Student, em ambas as análises. A diferença de expressão gênica é dada em “*fold change*”, que é igual a razão da expressão gênica do *SHR.BN4* ($FPKM_{SHR.BN4}$ ou $baseMean_{SHR.BN4}$) pela expressão gênica do *SHR* ($FPKM_{SHR}$ ou $baseMean_{SHR}$).

Após a análise dos genes diferencialmente expressos foram selecionados dentre estes quais estariam no intervalo transferido do animal *BN* para o *SHR* na linhagem congênica do cromossomo 4.

Com base nos genes diferencialmente expressos entre congênico e *SHR* no intervalo e nos demais cromossomos e com o auxílio do software Metacore (Thomson Reuters) (<http://www.genego.com/metacore.php>), foi analisado quais vias e redes estariam significativamente alterados entre as linhagens. Isso porque esta ferramenta possibilita acessar todos os bancos públicos de interação gene-gene/gene-proteína disponíveis até o momento, permitindo a visualização de vias relacionadas com os genes em estudo com todas as informações publicadas até o presente (tipo de proteína, tipo de interações, ação de uma droga sobre a proteína de estudo, regulação, etc.).

Dessa forma, inserindo no programa os dados de genes diferencialmente expressos, realizou-se a análise de enriquecimento de vias, onde se buscou as dez mais afetadas vias metabólicas, redes de processos, doenças e “*Gene Ontology Processes*” e que possuíam genes que estavam no intervalo. As vias e redes foram consideradas significativamente alteradas entre as linhagens quando o *FDR* calculado pelo programa era menor ou igual a 0,05, ou seja, considerando o dado inserido e o

número de componentes das vias e redes encontradas, quais tinham alta probabilidade de estarem afetadas com uma taxa de falso positivo de 5%.

3.2. Busca de alterações estruturais entre *SHR* e *BN* com base no sequenciamento de seus DNAs genômicos

Para a busca de alterações estruturais entre os animais *SHR* e *BN*, foram utilizados dados do sequenciamento do DNA genômico tanto do animal *BN*, quanto do animal *SHR*, presentes em banco de dados públicos (*UCSC Genome Browser*) (<http://genome.ucsc.edu>). As sequências do animal *SHR* foram alinhadas com base no genoma do animal *BN* (rn4, o mesmo utilizado para o alinhamento do RNA-seq), por meio da ferramenta Bowtie 2. Assim como realizado nas análises do RNA-seq, somente sequências para alinhamento com qualidade (Q) igual ou superior a 20 foram consideradas. As chamadas de *SNVs* e Indels foram feitas por meio da ferramenta Samtools (versão 0.1.18) (65) e foram consideradas variantes trocas suportadas por pelo menos 5 sequências. Para os *SNVs*, buscou-se aqueles já conhecidos utilizando o dbSNP (Versão 136), obtido no banco de dados Ensemble (<http://www.ensembl.org>)

Tanto os *SNVs* como os Indels foram classificados segundo a região genômica a quais pertencem e se identificou aqueles presentes do intervalo de interesse do congênico *SHR.BN4*.

3.3. Combinação das estratégias baseadas em genes diferencialmente expressos entre as linhagens e busca de *SNVs* e Indels

Dessa forma, foram selecionados genes que demonstraram diferença de expressão significativa entre as linhagens a partir do RNA-seq e que apresentaram alterações estruturais entre o DNA genômico de animais *SHR* e *BN* na região promotora (1500 *bp* anteriores ao primeiro exon do gene), no primeiro intron e na região 3' não traduzida (do inglês *Untranslated Region – UTR*) conservadas ao longo da evolução e cuja troca de variante não foi identificada entre mamíferos (camundongo, humano, cachorro e gambá). Para tanto, utilizou-se de informações de conservação presentes no “*UCSC Genome Browser*” (<http://genome.ucsc.edu>).

Além disso, buscou-se *SNVs* e Indels em regiões codificadoras de genes (do inglês *Coding DNA Sequence – CDS*) presentes no intervalo do animal congênico e que são expressos no tecido renal que não necessariamente apresentaram diferença

de expressão significativa entre as linhagens. Para os *SNVs*, foram consideradas trocas não sinônimas, conservadas ao longo da evolução, cujo aminoácido trocado não está presente em outras espécies e a característica química difere. Para os Indels, foram selecionados aqueles que levam a uma mudança do quadro de leitura do gene.

4 – RESULTADOS

4.1. RNA-seq de amostras de rim de animais congênicos e SHR após sobrecarga salina

A partir do RNA-seq de amostras renais, foram geradas 160.114.754 de sequências e 8.005.737.700 de bases para a linhagem *SHR* e 176.522.994 de sequências e 8.826.149.700 de bases para a linhagem *SHR.BN4*, considerando todas as replicatas técnicas e biológicas. Destas, 68.042.734 de sequências (42,50%) do animal *SHR* foram mapeadas no genoma referência, sendo que 61.525.707 de sequências (38,43%) tiveram mapeamento confiável, ou seja, tiveram sequências que alinharam em uma única região do genoma referência, o que corresponde a 3.076.285.650 de bases (38,43%). Para a linhagem congênica, 74.606.724 de sequências (42,26%) foram mapeadas, sendo que 66.342.505 de sequências (37,58%) tiveram mapeamento confiável, o que corresponde a 3.317.125.550 de bases (37,58%).

Em uma primeira análise, foi verificada a quantidade de genes expressos em tecido renal nas diferentes linhagens. A tabela 1 mostra a quantidade de genes considerados como expressos com uma ou mais sequências, dez ou mais sequências ou trinta ou mais sequências em pelo menos um exon do gene. Segundo esses resultados, cerca de 5.000 genes apresentaram baixa expressão no tecido renal, já que apresentaram entre 1 a 10 sequências no RNA-seq. Seguindo o mesmo raciocínio, cerca de 2.000 genes apresentaram nível de expressão intermediário e 11.000 são altamente expressos no rim.

Tabela 1 - Total de genes expressos nas diferentes linhagens, considerando diferentes categorias de corte

Linhagem	Total de genes expressos		
	≥ 1 sequência	≥ 10 sequências	≥ 30 sequências
SHR	18 351	13 457	11 623
SHR.BN4	18 671	13 605	11 766

Considerando o intervalo de 56 *Mbp* do congênico *SHR.BN4* transferido do animal *BN* para o *SHR*, com base no banco de dados do Ensemble, existem nesta região 682 genes conhecidos. Destes, 513 apresentaram expressão renal segundo os dados de RNA-seq em animais *SHR* (considerando como expressos genes com uma

ou mais sequências em pelo menos um exon do gene), enquanto que 533 foram expressos no congênico *SHR.BN4*. Quanto aos genes diferencialmente expressos entre o congênico e o *SHR* no RNA-seq com $FDR \leq 0,05$, 310 genes apresentaram diferença quando considerados todos os cromossomos (Figura 4), sendo que 28 genes estão presentes no intervalo do congênico do cromossomo 4 (Figura 5 e Tabela 2). Com a finalidade de diminuir a chance de erro, consideramos genes diferencialmente expressos somente aqueles que foram identificados pelas duas ferramentas de análise, ou seja, Cufflinks e DEseq.

Figura 4. Os 310 genes diferencialmente expressos entre *SHR* e *SHR.BN4* no RNA-seq segundo sua localização por cromossomo. Os dados são apresentados em “*Fold Change*”, ou seja, quantas vezes o gene é mais ou menos expresso no congênico em comparação com o *SHR*, considerando um $FDR \leq 0,05$

Figura 5. Conjunto de 28 Genes diferencialmente expressos entre *SHR* e *SHR.BN4* no intervalo do cromossomo 4 transferido do animal *BN* para o animal *SHR*. Os dados são apresentados em “*Fold Change*”, ou seja, quantas vezes o gene é mais ou menos expresso no congênico em comparação com o *SHR*, considerando um $FDR \leq 0,05$. O gráfico mostra resultados concordantes utilizando o método de análise Cufflinks e DEseq.

Tabela 2 - Genes diferentemente expressos entre *SHR.BN4* e *SHR* no intervalo do cromossomo 4 transferido do animal *BN* para o animal *SHR* com base no RNA-seq

Ensemble ID	Gene	Locus	Fold Change ¹	
			Cufflinks	DEseq
ENSRNOG00000034549	Y_RNA	chr4: 75 915 530-75 915 632	3,509	3,411
ENSRNOG00000035205	Y_RNA	chr4: 75 918 786-75 918 898	6,769	6,734
ENSRNOG00000007707	Znf467	chr4: 76 521 930-76 528 420	1,540	1,538
ENSRNOG00000024705	Rarres2	chr4: 76 640 814-76 662 603	1,394	1,929
ENSRNOG00000008465	Tmem176b	chr4: 76 734 381-76 977 649	1,209	1,190
ENSRNOG00000010594	LOC500124	chr4: 78 828 164-78 862 257	1,863	1,838
ENSRNOG00000009172	Cpvl	chr4: 82 238 756-82 344 130	3,470	3,835
ENSRNOG00000010205	MGC95152	chr4: 83 087 756-83 108 673	1,202	1,291
ENSRNOG00000010629	Nod1	chr4: 83 398 817-83 449 207	1,666	1,666
ENSRNOG00000005478	Fkbp9	chr4: 85 816 059-85 865 955	0,725	0,727
ENSRNOG00000006646	Ecop	chr4: 87 116 145-87 186 828	1,831	1,836
ENSRNOG00000023969	Herc6	chr4: 87 278 853-87 320 930	1,622	1,544
ENSRNOG00000006893	Ppm1k	chr4: 87 366 152-87 393 063	1,317	1,431
ENSRNOG00000007041	Abcg2	chr4: 87 500 740-87 560 447	1,167	1,199
ENSRNOG00000008656	Snca	chr4: 89 613 730-89 722 807	2,134	2,165
ENSRNOG00000031796	FILYY3_RAT	chr4: 97 364 279-97 364 570	0,000	0,000
ENSRNOG00000043369	F1M5H9_RAT	chr4: 101 181 862-101 182 150	0,000	0,000
ENSRNOG00000009484	Ptcd3	chr4: 105 159 843-105 251 459	2,234	1,339
ENSRNOG00000012659	Vamp5	chr4: 105 673 707-105 684 835	1,611	1,352
ENSRNOG00000014090	Retsat	chr4: 106 099 967-106 117 949	2,129	2,138
ENSRNOG00000005587	Suc1g1	chr4: 106 572 135-106 794 369	0,626	0,648
ENSRNOG00000011617	Dguok	chr4: 117 697 250-117 725 383	0,639	0,788
ENSRNOG00000013573	Atp6v1b1	chr4: 117 931 925-117 951 966	0,774	0,812
ENSRNOG00000015589	Smyd5	chr4: 119 686 587-119 701 841	0,752	0,739
ENSRNOG00000015799	Cml1	chr4: 119 978 870-120 190 938	2,870	2,758
ENSRNOG00000017720	Mxd1	chr4: 120 800 057-120 822 346	1,299	1,256
ENSRNOG00000016316	Mcm2	chr4: 122 977 521-122 991 936	1,320	1,328
ENSRNOG00000017808	Klf15	chr4: 124 714 394-124 726 962	1,287	1,285

¹ Expressão congênico/expressão *SHR*, considerando um *FDR*<0,05.

Considerando que 1. o conhecimento das vias metabólicas a que pertencem estes genes pode ser útil na identificação de processos biológicos candidatos associados a gênese da hipertensão arterial e 2. a gênese da hipertensão arterial no homem possa ter em comum a mesma via metabólica e não necessariamente o mesmo gene afetado no rato, buscou-se identificar vias metabólicas ou redes de processos onde genes do intervalo participavam e que poderiam estar alteradas e envolvidas com fenótipo de aumento da pressão arterial no *SHR*. A tabela 3 mostra o

resultado desta análise, que inclui vias metabólicas canônicas, redes de processos, doenças e “*Gene Ontology Processes*” candidatos. Esta análise foi realizada por meio da ferramenta de enriquecimento de vias do software Metacore.

Tabela 3 - Vias metabólicas canônicas, rede de processos, doenças e “*Gene Ontology Processes*” obtidos no Metacore a partir dos genes diferencialmente expressos entre congênico e *SHR* pelo RNA-seq

#	Mapas de Vias metabólicas	Total ¹	p-valor	FDR ²	Genes dif. Expressos ³	Genes dif. Exp. do intervalo ⁴
1	LRRK2 nos neurônios na Doença de Parkinson	33	0,001	0,029	4	Snca
2	Proteólise: Papel da parquina na via Ubiquitina-Proteossomal	24	0,004	0,047	3	Snca
3	Ciclo Celular: Transição e finalização da replicação do DNA	28	0,020	0,047	1	Mcm2
4	Desenvolvimento: Regulação transcricional no desenvolvimento de granulócito	32	0,022	0,047	2	Mxd1
5	Ciclo celular: Início da replicação do DNA no início da fase S	32	0,022	0,047	1	Mcm2
6	Resposta imune: Inflamasoma na resposta inflamatória	34	0,024	0,047	1	Nod1
7	Desenvolvimento: Papel da activina A na diferenciação e proliferação celular	40	0,028	0,048	2	Mxd1
8	Desenvolvimento: Diferenciação de queratinócitos	56	0,039	0,050	2	Mxd1
9	Metabolismo de dATP/dITP	95	0,065	0,065	2	Dguok
#	Redes de Processos	Total ¹	p-valor	FDR ²	Genes dif. Expressos ³	Genes dif. Exp. do intervalo ⁴
1	Dobramento de proteína: Retículo endoplasmático e citoplasma	45	0,000	0,003	6	Fkbp9
2	Ciclo celular: Núcleo	115	0,005	0,047	3	Mcm2
3	Ciclo celular: Fase S	149	0,008	0,047	3	Mcm2
4	Ciclo celular: Regulação da fase G1-S pelo fator de crescimento	195	0,029	0,169	8	Mxd1
5	Desenvolvimento: Diferenciação de queratinócitos	58	0,055	0,153	2	Mxd1
6	Processo neurofisiológico: Potenciação de longa duração	82	0,078	0,171	2	Snca
7	Inflamação: Inflamasoma	118	0,110	0,180	3	Nod1
8	Apoptose: Receptores com domínios de morte e caspases na apoptose	123	0,115	0,180	4	Nod1
9	Transporte: Exocitose de vesículas sinápticas	176	0,160	0,194	2	Snca
10	Proliferação: Regulação negativa da proliferação celular	184	0,167	0,194	2	Mxd1
#	Doenças	Total ¹	p-valor	FDR ²	Genes dif. Expressos ³	Genes dif. Exp. do intervalo ⁴
1	Neoplasias colorretais	12 494	1,1E-16	1,1E-13	110	Nod1, Mcm2, Abcg2, Ecop
2	Neoplasias	21 268	1,4E-14	6,4E-12	105	Nod1, Mxd1, Abcg2

#	Doenças	Total ¹	p-valor	FDR ²	Genes dif. Expressos ³	Genes dif. Exp. do intervalo ⁴
3	Neoplasias de estômago	5 713	5,7E-14	1,8E-11	131	Nod1, Mcm2, Abcg2, Snca, LOC500124
4	Artrite	1 782	3,1E-13	7,3E-11	30	Nod1
5	Neoplasias do Colo	1 647	6,3E-13	1,1E-10	58	Nod1, Mcm2, Abcg2
6	Mieloma múltiplo	2 126	6,8E-13	1,1E-10	68	Mcm2, Abcg2, Herc6
7	Asma	1 206	8,8E-13	1,2E-10	48	Nod1, Cpv1
8	Lúpus eritematoso sistêmico	679	3,1E-12	3,7E-10	34	-
9	Melanoma cutâneo maligno	937	5,2E-12	5,5E-10	41	Abcg2, Mcm2
10	Neoplasias de seios	11 108	6,8E-12	6,4E-10	196	Nod1, Cpv1, Smyd5, Mxd1, Abcg2, Mcm2, Snca, Retsat, Ptcd3
#	Gene Ontology Processes	Total ¹	p-valor	FDR ²	Genes dif. Expressos ³	Genes dif. Exp. do intervalo ⁴
1	Processo metabólico de pequenas moléculas	2 999	3,0E-13	5,7E-10	60	Dguok, Abcg2, Suclg1
2	Resposta imune inata	928	4,5E-08	1,9E-05	31	Nod1, Herc6
3	Resposta ao Interferon-gamma	164	7,8E-07	1,2E-04	7	Snca
4	Dobramento de proteínas	261	2,6E-06	2,5E-04	14	Fkbp9
5	Envelhecimento	403	3,3E-06	2,7E-04	16	Snca
6	Resposta a estímulo de glicocorticoides	294	5,5E-06	3,9E-04	12	Abcg2
7	Processo metabólico	10 910	5,8E-06	3,9E-04	37	Suclg1, Cml1
8	Homeostase de íon de cálcio	437	6,2E-06	8,9E-04	3	ATP6V1B1, Snca
9	Gastrulação com formação da boca	41	8,6E-06	8,9E-04	3	Cml1
10	Ativação das células da microglia	23	1,5E-05	7,4E-04	4	Snca

¹ Número de componentes (genes/proteínas/fármacos) na via/rede.

² “False Discovery Rate” (buscou-se vias e processos com $FDR < 0,05$).

³ Número de componentes da via que apresentaram diferença de expressão entre congênico e *SHR*.

⁴ Genes que apresentaram diferença de expressão entre as linhagens e que estão no intervalo.

De acordo com a análise, identificou-se genes diferencialmente expressos que participariam de vias metabólicas e redes de processos importantes no funcionamento e homeostase do organismo: dobramento de proteínas, ciclo celular, processos inflamatórios, metabolismos, desenvolvimento, diferenciação, proliferação e morte celular. Destes, a resposta a estímulo de glicocorticoide e o controle da homeostase de íons cálcio são potencialmente importantes para o controle da pressão arterial e especificamente incluem os genes *Abcg2*, *ATP6V1B1* e *Snca*.

Em conjunto, os resultados mostram que 533 dos 682 genes presentes no intervalo sob investigação no cromossomo 4 são expressos no tecido renal, sendo que 28 se apresentam diferencialmente expressos entre as 2 linhagens e 17 participam de

vias metabólicas ou redes de processos que poderiam estar alterados nestes animais e com potencial de influenciar o controle da pressão arterial e a hipertensão.

4.2. Busca de alterações estruturais entre *SHR* e *BN* com base no sequenciamento de seus DNAs genômicos

Com base no sequenciamento do DNA genômico de animais *SHR* e *BN* disponíveis em bancos de dados públicos, foram encontrados 2.819.762 *SNVs* entre o *SHR* e o *BN*, sendo que 2.699.022 já eram conhecidos segundo o dbSNP (95,72%). Entre os 2.819.762 *SNVs* encontrados, 2.741.135 estavam em homozigose no animal *SHR* quando comparado com o *BN* e 78.627 em heterozigose. A Tabela 4 sumariza a quantidade de *SNVs* presentes em cada uma das regiões genômicas.

Tabela 4 - Número de *SNVs* por região genômica

Região genômica	Número de <i>SNVs</i>
CDS ¹	9 316
UTR ²	21 659
Intron	820 381
Intron 1	207 596
Splice Site	135
miRNA ³	1
Região Intergênica	1 937 501
Região Promotora ⁴	30 769

¹ Do inglês “*Coding DNA Sequences*”.

² Do inglês “*Untranslated Regions*”.

³ MicroRNA.

⁴ Região de 1500 *bp* anterior ao primeiro exon do gene.

Considerando apenas os *SNVs* encontrados em regiões codificadoras, 6.004 são alterações sinônimas, ou seja, a troca da base não promove alteração do aminoácido codificado, e 3.312 são alterações não sinônimas, onde 3.285 são “*missense*” (a troca do nucleotídeo promove a troca do aminoácido codificado) e 27 são “*non-sense*” (a troca do nucleotídeo leva a um códon de parada).

No intervalo da linhagem congênica do cromossomo 4, foram encontrados 62.966 *SNVs*, sendo que 95,60% já eram conhecidos segundo o dbSNP (60.193 *SNVs*). A tabela 5 sumariza a quantidade de *SNVs* no intervalo de 56 *Mbp* do *SHR.BN4* presentes em cada uma das regiões genômicas.

Tabela 5 - Número de *SNVs* no intervalo do congênico por região genômica

Região genômica	Número de <i>SNVs</i>
CDS¹	210
UTR²	430
Intron	18 789
Intron 1	4 447
Sítio de splicing	2
miRNA³	0
Intergênica	42 614
Promotora⁴	450

¹ Do inglês “*Coding DNA Sequences*”.

² Do inglês “*Untranslated Regions*”.

³ MicroRNA.

⁴ Região de 1500 *bp* anterior ao primeiro exon do gene.

Considerando apenas os *SNVs* encontrados em regiões codificadoras, 127 são alterações sinônimas e 83 são alterações não sinônimas presentes em 47 genes, onde 82 são “*missense*” e 1 é “*non-sense*”. Destes, 67 *SNVs* “*missense*” estão presentes em 36 genes expressos no rim. A única alteração “*non-sense*” esta presente no gene *Vomlr70* que não é expresso no rim. Quanto às análises de inserções e deleções, foram identificados 413.871 Indels entre o genoma do *SHR* e do *BN*, onde 14.835 estão em regiões codificadoras de proteínas, 2.333 estão em regiões não traduzidas, 108.775 estão em introns, 52 estão em sítios de splicing, 282.808 estão em regiões intergênicas e 5.068 estão em regiões promotoras (1500 *bp* anteriores ao primeiro exon do gene).

No intervalo da linhagem congênica do cromossomo 4, foram encontrados 8.604 Indels, onde 319 estão em regiões codificadoras de proteínas, 48 estão em regiões não traduzidas, 2.382 estão em regiões intrônicas, 5.750 estão em regiões intergênicas e 105 estão em regiões promotoras. Considerando apenas os Indels em regiões codificadoras, 277 estão em 54 genes e causam mudança de quadro de leitura. Destes, 266 estão em 50 genes expressos no rim.

4.3. Combinação das estratégias baseadas em genes diferencialmente expressos entre as linhagens e busca de *SNVs* e Indels

Dentre os 28 genes que apresentaram diferença de expressão gênica significativa pelo RNA-seq (Tabela 2), 8 apresentaram *SNVs* (Tabela 6) e 4 apresentaram Indels (Tabela 7) em regiões promotoras, intrônicas e 3' *UTR*

conservadas ao longo da evolução segundo o banco de dados do “UCSC Genome Browser”, sendo que o gene *Abcg2* apresentou tanto *SNVs* quanto Indels.

Tabela 6 - *SNVs* presentes em regiões promotoras, intrônicas e 3’UTR em genes que apresentaram diferença de expressão entre as linhagens

SNP ID	Posição	Alelo Ref.	Alelo Var.	N alelo ref. (SHR)	N alelo var (SHR)	Região	Gene
rs197986969	87 121 768	T	A	0	8	Intron1	Ecop
rs197508516	87 122 889	G	A	0	7	Intron1	Ecop
rs198795083	87 322 756	G	A	0	10	Intron1	Herc6
rs199387865	87 322 852	C	G	0	12	Intron1	Herc6
rs199257895	87 508 614	C	T	0	9	Intron1	Abcg2
rs198788200	87 510 717	C	T	0	8	Intron1	Abcg2
rs197078739	87 512 018	T	G	0	5	Intron1	Abcg2
rs198528199	87 512 066	C	T	0	9	Intron1	Abcg2
rs198771647	87 513 840	A	G	0	9	Intron1	Abcg2
rs105654027	106 576 541	C	T	0	6	Intron1	Suclg1
rs198683616	85 814 655	T	C	0	10	Promoter	Fkbp9
rs198573098	89 722 876	C	A	0	8	Promoter	Snca
rs199300301	87 390 081	T	C	0	5	3' UTR	Ppm1k
rs8151046	89 613 910	T	C	0	6	3' UTR	Snca
rs197471458	120 803 023	G	A	0	11	3' UTR	Mxd1

Tabela 7 - Indels presentes em regiões promotoras, intrônicas e 3’UTR em genes que apresentaram diferença de expressão entre as linhagens

Seq ref.	Seq var	Posição	N seq ref. (SHR)	N seq var (SHR)	Região	Gene
GAAAAAAAAAA	GAAAAAAAAAA	78 859 835	0	5	Intron1	LOC500124
GAAA	GAA	83 426 063	0	10	Intron1	Nod1
ATCT	ATCTCT	87 511 440	0	6	Intron1	Abcg2
TAAAAA	TAAAAAA	87 513 757	0	5	Intron1	Abcg2
AGG	AG	120 064 449	0	12	Intron1	Cml1

Com relação a genes que apresentaram alterações em regiões codificadoras que são expressos no rim, 6 genes apresentaram *SNVs* que são trocas não sinônimas, conservadas ao longo da evolução e cujos aminoácidos trocados têm características químicas diferentes (Tabela 8) e 50 apresentaram Indels que causam mudança em quadro de leitura.

Tabela 8 - SNVs não sinônimos em genes do intervalo expressos no rim

SNP ID	Posição	Alelo Ref.	Alelo Var.	N alelo ref. (SHR)	N alelo var (SHR)	gene	Códon ref.	Códon var	aa ref.	aa var		
rs13448732	75 931 470	C	T	0	6	Pdia4	CGA	CAA	R	Q	polar básico	polar neutro
rs198490537	83 035 059	A	T	0	14	Plekha8	GAA	GTA	E	V	polar ácido	apolar
rs106590595	87 257 387	G	A	0	7	Pigy	GCA	ACA	A	T	apolar	polar neutro
rs197110744	105 209 635	T	C	0	13	Polr1a	ATC	ACC	I	T	apolar	polar neutro
rs198576319	120 884 704	G	C	0	9	Gmcl1	GAC	GAG	D	E	polar ácido	polar ácido
rs199147785	125 687 984	C	A	0	14	Xpc	GCC	TCC	A	S	apolar	polar neutro

Dentre os genes que apresentaram Indels, 6 estão diferencialmente expressos entre as linhagens: o Fkbp9, o Ecop, o Ppm1k, o Abcg2, Suc1l e o Atp6v1b1.

Em conjunto, as análises de alterações estruturais mostram:

1. Alterações (SNVs e/ou Indels) em regiões conservadas ao longo da evolução potencialmente envolvidas com a regulação de 11 dos 28 genes diferencialmente expressos entre as linhagens. Dentre estes, apenas os genes Ppm1k e LOC500124 não participam de nenhuma via metabólica ou rede de processo alterado entre as linhagens nas análises do Metacore;
2. Alterações não sinônimas “*missense*” em 6 genes expressos no rim.

4.4. Comparação dos resultados do RNA-seq com dados de expressão gênica ao longo do néfron

Dentre os 28 genes diferencialmente expressos no RNA-seq, 6 tiveram sua expressão previamente testada ao longo de segmentos do néfron de animais adultos submetidos a dieta normossódica: LOC500124, MGC95152, Fkbp9, Retsat, Dguok e o Cml1 (Apêndice I).

Os genes Retsat e Cml1 tiveram a diferença de expressão observada entre as linhagens em amostra de rim total, amplificadas no túbulo proximal. O gene Fkbp9 não apresentou diferença de expressão em nenhum segmento estudado e os genes LOC500124, MGC95152 e Dguok apresentaram diferença de expressão significativa, mas no sentido contrário do observado nas análises do RNA-seq (Figura 6 e Tabela Suplementar 9). Dentre esses 6 genes, apenas o LOC500124 e o

Cml1 possuem alterações estruturais em regiões conservadas ao longo da evolução, potencialmente regulatórias.

Figura 6. Diferença de expressão gênica em diferentes segmentos do néfron entre congênico *SHR.BN4* e *SHR* em amostras adultas de genes selecionados pelo RNA-seq. A figura mostra a expressão gênica nos animais congênicos com base nos valores de expressão obtidos no *SHR*. TPC – túbulo proximal convoluto, TPR – túbulo proximal reto, AHE – Alça de Henle espessa, DCC – duto coletor cortical e DCM – duto coletor medular. * representa genes diferencialmente expressos entre as linhagens com $p < 0,05$

5 – DISCUSSÃO

A proposta deste projeto foi selecionar genes candidatos que possam resultar em hipertensão arterial presentes no intervalo do cromossomo 4 de linhagem congênica de ratos previamente desenvolvida em nosso laboratório (56). Para isso, utilizamos uma abordagem integrada com base na análise das diferenças de padrão de expressão gênica no rim e na busca de alterações genéticas estruturais que expliquem essas diferenças ou que causem alterações na região codificadora de genes expressos no rim entre animais da linhagem parental *SHR* e congênica. Utilizando esta abordagem, identificamos 34 genes candidatos que deverão ser objeto de maior investigação e que apresentam padrão de expressão diferenciada (N=28) ou alterações não sinônimas em região codificadora (N=6) entre as linhagens.

Durante o desenvolvimento do projeto, as plataformas de sequenciamento de próxima geração evoluíram rapidamente e se tornaram uma alternativa poderosa para identificação de genes candidatos com base em estudos de genômica estrutural e funcional. Essas tecnologias promovem o sequenciamento de DNA capaz de gerar informações sobre bilhões de pares de bases em uma única corrida.

Existem atualmente no mercado três modelos de sequenciadores de alto rendimento com características distintas (incluindo-se a química de sequenciamento) que estão sendo amplamente utilizados em projetos de sequenciamento em larga escala de diferentes organismos. São eles: o 454 da Roche, o *SOLiD* da Applied Biosystems e o HiSeq da Illumina. Dentre estes, os dois últimos são semelhantes no que tange o número de sequências obtidas: ambos geram número muito grande de sequências (centenas de milhões por corrida) relativamente curtas, de 50 a 100 nucleotídeos. Por outro lado, o exemplar da Roche é capaz de gerar sequências mais longas, de até 700 nucleotídeos, mas em número bem menor por corrida: um milhão de sequências. Sendo assim, cada uma destas plataformas apresenta suas vantagens e desvantagens dependendo da aplicação desejada. Ainda assim, uma característica compartilhada pelos três equipamentos é a detecção clonal de moléculas únicas de DNA isoladas fisicamente (66).

As leituras produzidas pela plataforma *SOLiD* a tornaram uma excelente ferramenta destinada principalmente aos estudos de RNA-seq, devido a sua metodologia de sequenciamento de duas bases, que promove uma acurácia de

99,85%, após filtrar as sequências de baixa qualidade, e do tamanho das sequências geradas (66).

Neste projeto, utilizamos o RNA-seq de amostras de rim de animais adultos *SHR* e congênicos submetidos à dieta hipersódica. Considerando os altos custos desta metodologia (66), optamos por maximizar os resultados por meio de 1) utilização de pool de amostras de animais, já que, além de reduzir o número de amostras a serem sequenciadas, ao juntar amostras de animais diferentes da mesma linhagem, minimiza-se o efeito das diferenças biológicas encontradas entre eles; 2) utilização de animais submetidos à sobrecarga salina, pois, nesta circunstância, o animal está sob estresse e, alterações antes não perceptíveis, serão amplificadas, conforme demonstrado durante o mapeamento dos QTLs no genoma completo do *SHR* (55).

Alguns dos dados gerados por essa metodologia corroboram resultados previamente obtidos em nosso laboratório (67), utilizando a técnica de Real Time PCR quantitativo (qRT-PCR) e *in silico* “*chromosome walking*” para identificar genes do intervalo com perfil de expressão diferencial entre o congênico e o *SHR* possivelmente envolvidos com a hipertensão arterial (Apêndice I).

No entanto, como os experimentos de expressão gênica por qRT-PCR e RNA-seq foram realizados em condições diferentes (o qRT-PCR em amostras renais de animais submetidos à dieta normossódica e o RNA-seq em amostra renais de animais submetidos em dieta hipersódica), os resultados obtidos não são diretamente comparáveis. É importante salientar que a estratégia empregada neste projeto é mais ampla e robusta, já que analisa a expressão de todos os genes renais, até mesmo aqueles cuja expressão neste órgão era desconhecida.

Sendo assim, os 34 genes que apresentaram diferença de expressão gênica pelo RNA-seq e/ou alterações estruturais (*SNVs* e/ou *Indels*) entre as linhagens foram considerados candidatos. Dentre esses, aqueles que não foram previamente testados, serão avaliados por qRT-PCR em condições utilizadas anteriormente (adulto basal, jovem basal, 5 segmentos do néfron). Apesar de estudos demonstrarem a alta acurácia na quantificação de expressão gênica no RNA-seq, especialmente devido à forte correlação com qRT-PCR (68), como realizamos o sequenciamento com pool de amostras, decidimos validar também em amostras individuais de animais adultos pós-sobrecarga salina.

Além disso, a expressão desses genes candidatos será também avaliada em amostras de córtex e medula renal de animais *SHR* e congênicos, como forma de grosseiramente inferir se as diferenças de expressão observadas em amostra de rim total ocorrem em parte do sistema glomérulo-tubular ou de outras estruturas renais não avaliadas, já que nem sempre essas diferenças são observadas nos 5 segmentos tubulares anteriormente estudados.

Vale salientar que os *SNVs* e Indels foram encontrados com base no sequenciamento de animais *SHR* e *BN* disponíveis em banco de dados públicos e que não necessariamente são idênticos aos nossos animais. Apesar do *SHR* e do *BN* utilizados para o desenvolvimento da linhagem congênica e dos animais *SHR* utilizados como grupo controle serem originários da mesma cepa de animais, mutações podem ter sido acumuladas tanto em nossos animais, como naqueles sequenciados. Assim, todas as alterações aqui reportadas estão sendo validadas por sequenciamento do DNA.

Além disso, as alterações encontradas em regiões potencialmente regulatórias (promotor, primeiro intron e região 3' *UTR*) deverão, antes de serem validadas pelo sequenciamento, ser comparadas com bancos de dados públicos de fatores de transcrição e miRNAs por meio de ferramentas de bioinformática para evidenciar se a troca, a inserção ou a deleção de um ou mais nucleotídeos levam a quebra de um sítio de ligação de um fator de transcrição ou de um miRNA (69)(70)(71).

Os genes que apresentaram alterações em regiões codificadoras, que são expressos no rim segundo os dados do RNA-seq, mas que não demonstraram diferença de expressão significativa entre congênico e *SHR* também foram considerados candidatos. No entanto, em um estudo publicado recentemente na "*Cell*", onde foram sequenciadas 27 linhagens de ratos utilizadas para recapitular doenças complexas em humanos, entre elas a hipertensão essencial, nenhum dos Indels identificados pelas nossas análises foi encontrado nesta publicação (72), sendo estes genes desconsiderados em primeiro momento.

Nesta publicação, foram apresentados também os *SNVs* que levavam formação ou perda de códon de parada, sendo que apenas um gene foi identificado em nossas análises. No entanto, este também será desconsiderado em primeiro momento por não apresentar expressão renal.

Os *SNVs* não sinônimos “*missense*” encontrados em nossa análise, em regiões conservadas evolutivamente e cujos aminoácidos trocados diferem quimicamente são novos candidatos importantes, mas que não puderam ser confirmados nos dados da referência recém publicada, já que esta informação não estava disponível.

Dentre os genes diferencialmente expressos que apresentaram alterações (*SNVs* e Indels) em regiões potencialmente regulatórias estão o *Fkbp9*, *Ecop*, *Herc6*, *Ppm1k*, *Abcg2*, *Snca*, *Suclg1*, *Mxd1*, *LOC500124*, *Nod1* e *Cml1*.

O gene *Fkbp9* foi testado tanto no RNA-seq quanto no qRT-PCR e em ambas as situações ele apresentou diminuição da expressão no congênico quando comparado com o *SHR* (de 0,7 vezes no RNA-seq e de 0,6 vezes no qRT-PCR em amostras de animais adultos submetidos à dieta normossódica). No entanto, esse gene não apresentou diferença de expressão em animais jovens, onde a pressão arterial entre as duas linhagens é igual, nem nos 5 segmentos no néfron estudados quando avaliado pelo qRT-PCR (Tabela Suplementar 9. Apêndice I). Segundo as análises de alterações estruturais em regiões regulatórias entre animais *SHR* e *BN*, o *Fkbp9* apresentou um *SNV* na região promotora conservada evolutivamente (rs198683616).

O gene *Fkbp9* faz parte da família de genes que produzem imunofilinas, proteínas altamente conservadas que se ligam a drogas imunossupressoras (tacrolimus). Elas exibem uma atividade de peptidilprolil cis/trans isomerase (PPIase), que é responsável por catalisar a isomeração cis-trans da ligação peptídica da amida das prolinas. Esta atividade está intimamente ligada ao processo de dobramento lento de proteínas, que é conservada em todos os organismos, de bactérias a primatas. Além disso, sabe-se que a família *Fkbp* está envolvida com uma série de processos bioquímicos, como a sinalização a receptores, o tráfego de proteínas e a transcrição (73). Em camundongos, foi observado que o *Fkbp9* possui 4 domínios PPIase e 2 domínios de ligação de cálcio, sendo que este gene é altamente expresso no coração, músculo, pulmão e rim (74).

Segundo as análises do Metacore, a rede de processo de dobramento de proteína no citoplasma e no retículo endoplasmático está alterada entre as linhagens

congênicas e *SHR*, sendo que além do *Fkbp9*, outros cinco genes cujas proteínas participam dessa rede estão diferencialmente expressos entre as linhagens (Tabela 3).

O gene *Herc6* não foi testado pela técnica de qRT-PCR, pois nas análises baseadas em bancos de dados públicos e na estratégia de *SAGE* (do inglês *Serial Analysis of Gene Expression*) realizada em nosso laboratório (67), este gene não foi classificado por apresentar expressão renal. No RNA-seq, ele apresentou uma superexpressão no congênico quando comparado com o *SHR* (1,8 vezes) e *SNVs* em regiões conservadas do seu intron 1 (rs198795083 e rs199387865).

O *Herc6* faz parte da família de proteínas envolvidas com a ubiquitinação. A ubiquitinação demanda um conjunto de ações de uma cascata enzimática envolvendo uma enzima ativadora de ubiquitina E1, uma conjugase de ubiquitina E2 e uma ligase de ubiquitina E3, que garante o reconhecimento específico do substrato. A família *Herc* exibe uma atividade de ligase de ubiquitina E3, conhecida como *HECT* (do inglês *homologous to E6-AP (E6-associated protein) C-terminus*) (75). Estas formam um tioéster com a ubiquitina por meio de um resíduo de cisteína ativo antes de transferi-lo para o substrato. Além disso, essa família de proteína possui também um domínio regulador de condensação do cromossomo 1 (*RCC1*) (76).

O gene *Herc6* foi recentemente caracterizado, juntamente com o *Herc4*, e ambas apresentam muitas isoformas. Mostrou-se também que este gene é mais expresso em cérebro e testículos, e que, intracelularmente, a proteína se localiza no citoplasma ou em estruturas vesiculares (75).

Segundo o Metacore, o *Herc6* faz parte do processo de resposta imune inata que está alterada entre as linhagens. Além do *Herc6*, outros 30 genes diferencialmente expressos entre as linhagens codificam proteínas envolvidas neste processo, incluindo o gene *Nod1*, também presente no intervalo (Tabela 3).

Assim como o gene *Herc6*, os genes *Nod1*, *Ecop*, *Ppm1k*, *Snca* e *Mxd1* não foram testados até este momento por qRT-PCR pelos mesmos motivos.

O gene *Nod1* apresentou superexpressão na linhagem congênica quando comparada com o *SHR* nas análises de RNA-seq (1,6 vezes) e uma deleção de um nucleotídeo A nos animais *SHR* quando comparados com os animais *BN* no seu primeiro intron, em uma região conservada evolutivamente segundo o “*USCS Genome Browser*”.

O Nod1 faz parte da subfamília NLR (proteínas com domínio de oligomerização de nucleotídeos que são ricas em repetições de leucina) que constitui a maior subfamília de proteínas da imunidade inata. Estas agem primeiramente como receptores intracelulares de reconhecimento citoplasmático de padrões moleculares associados a patógenos e de produtos endógenos de dano tecidual. O reconhecimento desses sinais inicia uma cascata de reações de defesa através da ativação do NF- κ B, estresse de quinases, fatores regulados por interferon, ativação de caspases e autofagia. O Nod1 é amplamente expresso em todos os tipos celulares e em todos os tecidos e é ativada através do reconhecimento de motivos específicos presentes em bactérias Gram-positivas (peptídeoglicanos) (77).

A desregulação da função do Nod1 tem sido descrita em uma grande variedade de doenças, tais como inflamação crônica, autoimunidade e predisposição a câncer. Polimorfismos no gene Nod1 foram associados em humanos com o aumento na predisposição a câncer e, recentemente, a ativação da sinalização por Nod1 tem sido relacionada com a progressão da inflamação cardiovascular (77). No entanto, nenhum estudo relacionou este gene/proteína com a hipertensão essencial.

O gene *Ecop* apresentou uma superexpressão nos animais congênicos quando comparado com os animais *SHR* nas análises de RNA-seq (1,8 vezes) e 2 *SNVs* em regiões conservadas evolutivamente de seu intron 1 (rs197986969 e rs197508516).

Este gene está intimamente ligado a tumorigênese e confere um fenótipo pró-sobrevivência. Recentemente, foi dado uma nova nomenclatura ao gene, *Vopp1* (Proteína vesicular pró-sobrevivência superexpressa no câncer 1), já que foi observado que esta proteína não é secretada e se localiza intracelularmente em vesículas citoplasmáticas, ao contrário do que se acreditava (78).

Em 2011, Baras e colaboradores evidenciaram que a depleção do gene *Vopp1* induz morte celular por meio do estresse oxidativo e disfunção mitocondrial (79).

O gene *Ppmlk* está superexpresso nos animais congênicos quando comparados com os animais *SHR* na condição pós-sobrecarga salina no RNA-seq (1,3 vezes). Ele apresenta um *SNV* na região 3'UTR conservada evolutivamente no animal *SHR* (rs199300301), sendo um possível sítio de ligação de reguladores pós-transcricionais.

Alguns trabalhos já levantaram a possível influência de um *SNV* na região alvo de ligação de um miRNA. Nesses casos, essas variantes estariam impossibilitando a ligação do miRNA ou favorecendo a ligação de outros miRNAs, afetando, assim, a regulação pós-transcricional de um gene (80)(71).

O gene *Ppm1k* codifica a proteína fosfatase mitocondrial 2C (PP2Cm), que é membro da família das proteínas PP2 fosfatase serina/treonina (81). Ela é descrita por apresentar um papel essencial na ativação das vias de sinalização a estresse, regulação na permeabilidade de transição dos poros mitocondriais e outras funções essenciais no desenvolvimento e na função cardíaca, sendo que foi sugerido que a perda do PP2Cm contribui significativamente com a fisiopatologia da insuficiência cardíaca (82). Em um estudo recente, foi identificada uma relação entre a expressão do gene *Ppm1k* entre indivíduos hiperglicêmicos e normoglicêmicos, sugerindo a participação desse gene numa rede de interação envolvida com diabetes tipo 2 (83).

O gene *Snca* também apresentou superexpressão no animal congênico, quando comparado com o animal *SHR* no RNA-seq (2,1 vezes). Além disso, foi encontrado *SNVs* em sua região promotora e 3'UTR (rs198573098 e rs8151046, respectivamente).

Este gene, também chamado *Park1/4*, codifica a proteína chamada α -sinucleína. Ela é expressa pré-sinápticamente por todo o sistema nervoso central, mas ainda não se sabe ao certo qual é sua função. Alguns trabalhos discutem seu papel pré-sináptico no controle da fusão, regulação e clusterização de vesículas, sendo que, em culturas celulares, a superexpressão de α -sinucleína selvagem ou mutante leva a deficiência no ciclo das vesículas e na liberação de transmissores (84)(85)(86)(87)(88)(89).

Formas mutadas e duplicadas ou triplicadas do gene estão envolvidas com formas familiares raras de Doença de Parkinson (90)(91)(92)(93) e alguns polimorfismos no promotor do gene apresenta associação com Doença de Parkinson esporádica (94).

Segundo as análises do Metacore, a α -sinucleína participa de vários processos alterados entre a linhagem congênica e o *SHR*, como a potenciação de longa duração, a exocitose de vesículas sinápticas, resposta a interferon gama, envelhecimento,

homeostase do íon cálcio e ativação de células microgлияis. No entanto, não há nenhuma evidencia da participação deste gene/proteína com hipertensão essencial.

O gene *Mxd1* apresentou superexpressão no congênico quando comparado com o *SHR* (1,2 vezes) no RNA-seq e um *SNV* na região 3'UTR altamente conservado (rs197471458).

A proteína de dimerização MAX 1 (MAD1), codificada pelo gene *Mxd1*, é um supressor transcricional que antagoniza a ativação transcricional mediada pelo MYC, sendo que o mecanismo de inibição ocorre principalmente através da competição entre o MAD1 e o MYC pelos sítios de ligação nos promotores de genes alvos (95). É observado que o alto nível de expressão de MYC é encontrado em células em proliferação, enquanto que o nível de expressão do MAD1 é elevado em células em diferenciação (96).

Segundo as análises no Metacore, esta proteína está envolvida com a regulação transcricional de uma série de genes. No entanto, a partir dos dados de genes diferencialmente expressos entre as linhagens em todos os cromossomos, os genes por ele regulados não apresentaram diferença de expressão entre as linhagens. Sendo assim, o MAD1 não é um forte candidato.

O gene *Suc1g1* apresentou uma diminuição na expressão no congênico quando comparado com o *SHR* no RNA-seq (0,6 vezes), sendo que quando avaliado pelo qRT-PCR em pool de amostras de rim total de animais adultos submetidos à dieta normossódica, nenhuma diferença significativa foi observada (Tabela Suplementar 9. Apêndice I). Além disso, ele possui um *SNV* no seu primeiro intron conservado evolutivamente (rs105654027).

Este gene codifica uma enzima dimérica responsável por catalisar uma reação reversível onde a succinil-CoA, juntamente com uma guanósina difosfato (GDP) é convertido em succinato e GTP (97). Esta enzima está localizada na matriz mitocondrial e é expressa na maioria dos tecidos humanos, com predominância em tecidos anabólicos, como fígado e rim (98). Esta enzima está intimamente relacionada com a manutenção de nucleotídeo trifosfato, sendo essencial no metabolismo mitocondrial.

Pacientes com mutações no gene *Suc1g1* apresentam severa acidose láctica e elevados níveis de metilmalonato, metilcitrato e piruvato sanguíneo e urinário, resultando em morte neonatal.

O gene *Abcg2* apresentou uma superexpressão no congênico quando comparado com o *SHR* no RNA-seq (1,1 vezes), sendo que quando avaliado pelo qRT-PCR em pool de amostras de rim total de animais adultos submetidos à dieta normossódica, nenhuma diferença significativa foi observada (Tabela Suplementar 9. Apêndice I). Foi observado com base no genoma dos animais *SHR* e *BN*, que este gene possui uma série de *SNVs* (5) e *Indels* (2) no seu primeiro intron em regiões conservadas evolutivamente.

Segundo a análise do Metacore, esta proteína está envolvida com a resposta ao estímulo de glicocorticoides e este processo está alterado entre as linhagens, sendo que outros 11 genes cujas proteínas participam desse processo estão diferencialmente expressos entre as linhagens.

Esta proteína faz parte da família de transportadores com cassete de ligação a ATP (ABC transporters) que contém proteínas de membrana que estão envolvidas com o tráfego de uma ampla variedade de substratos através das membranas extra e intracelulares. Elas também são conhecidas por permitir o efluxo de drogas das células, já que limitam a concentração intracelular de um substrato através do bombeamento deste para fora da célula por meio de um mecanismo dependente de energia. Devido a essas características, muitas dessas proteínas são associadas com o fenômeno de resistência a multidrogas (99). Além da sua função de efluxo de substâncias, os transportadores ABC estão envolvidos na proteção celular contra apoptose e dano em resposta a hipóxia, assim como na proliferação e diferenciação celular na regeneração de fígado, coração e músculo esquelético (100).

A proteína ABCG2 é também conhecida por BCRP, ou proteína resistente do câncer de mama. Ela tem alta capacidade de bombear para fora da célula moléculas com carga positiva e negativa, ânions orgânicos e conjugados de sulfato. Ela é expressa em diversos tecidos, entre eles a placenta, o fígado, o intestino, o cérebro, a mama e os rins, sendo que nos rins ele contribui com a excreção de urina (99).

Huls e colaboradores demonstraram que o gene *Abcg2* é expresso na membrana apical das células do túbulo proximal de camundongos, ratos e humanos

(100)(101) e que, após a isquemia e reperfusão dos rins de camundongos, seu gene apresenta expressão reduzida enquanto que a proteína é altamente expressa. Segundo os autores, esta é uma resposta adaptativa ao acúmulo de compostos prejudiciais gerados pela isquemia, que deverão ser excretados na urina. Além disso, acredita-se que este gene exerce um papel na defesa e regeneração do tecido renal através da diferenciação celular.

O gene LOC500124 apresentou superexpressão no congênico em comparação com o *SHR* tanto no RNA-seq, quanto no qRT-PCR (em rim total e no ducto coletor de animais adultos submetidos à dieta normossódica) (Tabela Suplementar 9. Apêndice I). Este gene foi identificado em 2004 (102) e sua função ainda não é conhecida.

O gene *Cml1* apresentou superexpressão no congênico em comparação com o *SHR* tanto no RNA-seq, quanto no qRT-PCR (Tabela Suplementar 9. Apêndice I). Segundo o banco de dados do *NCBI* (*National Center of Biotechnology Information*), este gene pode agir como uma N-acetiltransferase, mas não existem artigos discutindo sua função. Segundo a literatura, parece haver uma confusão entre os genes *Cml1* e *Nat8* (103), mas segundo o *NCBI*, apesar de estarem no mesmo cromossomo no rato este genes possuem sítios de início de transcrição independentes.

Além dos genes presentes no intervalo que apresentaram diferença de expressão entre as linhagens e alterações em possíveis regiões regulatórias (*SNVs* e *Indels*) entre as cepas *SHR* e *BN*, outros 6 genes presentes no intervalo da linhagem congênica possuem *SNVs* em região *CDS* conservada evolutivamente, levando uma troca de aminoácido. Apesar da troca não ter sido confirmada quanto a sua influência na funcionalidade da proteína, estes genes também são considerados possíveis candidatos. Entre eles estão os genes *Pdia4*, *Plekha8*, *Pigy*, *Polr1a*, *Gmcl1* e *Xpc*.

Esses genes estão envolvidos com dobramento de proteínas (104)(105), transporte de proteínas e lípidos através do complexo de golgi (106), biossíntese de ancoradouros de membrana para proteínas de superfície celular (107), transcrição gênica (108), manutenção da integridade do núcleo celular, remodelamento da cromatina (109) e reconhecimento a dano no DNA (110). No entanto, nenhum desses genes apresentou trabalhos relacionando-os a hipertensão essencial ou a regulação da pressão arterial.

Desde o surgimento da penicilina e, por consequência, dos avanços no tratamento de doenças infecciosas, as doenças cardiovasculares se tornaram a principal causa de morte nos EUA e no mundo desenvolvido. O início da década de 60 ficou marcado pelas primeiras publicações advindas do estudo populacional de Framingham iniciado há mais de uma década (111)(112). Os resultados do estudo longitudinal permitiram a identificação dos principais fatores de risco para doenças cardiovasculares que incluem a hipertensão arterial, a dislipidemia, o diabetes, o tabagismo, a obesidade e o sedentarismo.

Neste contexto, o diagnóstico e o desenvolvimento de estratégias terapêuticas para o tratamento da hipertensão arterial se tornaram uma grande preocupação de saúde pública e nestas últimas seis décadas um avanço muito grande foi realizado no conhecimento dos sistemas de controle da pressão arterial e suas relações com diversos fatores ambientais, como a dieta rica em cloreto de sódio e o tabagismo (113)(114). O conhecimento profundo dos aspectos fisiológicos destes sistemas serviu de base para o desenvolvimento de várias classes de fármacos com capacidade de reduzir a pressão arterial na ausência de efeitos colaterais importantes. Ainda assim, as bases genéticas da hipertensão arterial primária permanecem até hoje basicamente desconhecidas.

O início da década de 90 marca um importante momento, pois o desenvolvimento e uso amplo das técnicas de biologia molecular por fisiologistas e clínicos permitiu indagar a gênese de doenças complexas por meio de estratégias que, em primeiro momento, dependiam de conhecimentos parciais de vias celulares e metabólicas envolvidas (estudos de associação). Dessa forma, muitos estudos foram conduzidos em genes/proteínas/vias metabólicas conhecidas por estar envolvidas com a regulação da pressão arterial para tentar identificar as causas genéticas da hipertensão arterial. No entanto, todas as evidências encontradas foram insatisfatórias para explicar a grande variação da pressão arterial observada na população humana e as causas da hipertensão arterial. Alternativamente, foram efetivadas estratégias que se baseiam no mapeamento no genoma completo para identificar regiões cromossômicas associadas à gênese das doenças complexas, discutidas na Introdução deste trabalho.

É neste contexto que nossos resultados se inserem. Desta forma, não é surpresa que os eventuais genes candidatos propostos neste trabalho aparentemente não estejam relacionados a processos de controle da pressão arterial e da hipertensão. Por outro lado, os genes propostos estão envolvidos com funções e processos biológicos importantes, cuja desregulação podem ocasionar, de maneira ainda desconhecida, fenótipos críticos, como a hipertensão arterial.

É importante salientar que nesta fase será fundamental o uso de estratégias “analíticas” e “sintéticas” para validar e caracterizar os 34 genes candidatos propostos. Nosso laboratório vem fazendo investimentos importantes há anos no desenvolvimento destas estratégias com base em modelos celulares e modificação genética de modelos complexos, como o rato, para teste destas hipóteses. Além disto, contamos com amostras de estudos transversais da população brasileira, bem como estudos tipo caso-controle para hipertensão e estudos populacionais longitudinais com núcleos familiares (115), que permitirão estudos translacionais, onde eventuais genes candidatos ou vias metabólicas selecionadas neste projeto poderão ser avaliados quanto sua contribuição na gênese da hipertensão arterial no *SHR* e no homem.

O estudo de determinantes genéticos da hipertensão arterial já está sendo conduzido com eficiência nessas amostras populacionais, sendo que SNPs associados com esse fenótipo em outras populações humanas foram testados, a fim de validar os dados encontrados e desenvolver um melhor modelo preditivo de risco de hipertensão arterial na população geral (Apêndice II).

Dessa forma, acreditamos que esse conjunto de abordagens exploratórias e translacionais será fundamental para a validação dos genes candidatos propostos e avaliação dos mesmos na gênese da hipertensão arterial no rato *SHR* e na hipertensão arterial primária no homem.

6 – CONCLUSÃO

Utilizando-se de um conjunto de estratégias, com base no padrão diferencial de expressão gênica em rim e na presença de alterações estruturais gênicas, selecionamos 34 genes candidatos que deverão ser investigados no futuro para estabelecer a sua influência na hipertensão arterial do *SHR* e sua relação com a hipertensão arterial primária no homem. Destes, 11 foram inicialmente selecionados por apresentarem padrão de expressão diferencial e presença de alterações estruturais com potencial de explicar as diferenças no padrão de expressão gênica (*Fkbp9*, *Ecop*, *Herc6*, *Ppm1k*, *Abcg2*, *Snca*, *Suclg1*, *LOC500124*, *Nod1*, *Cml1* e *Mxd1*) e 6 por possuírem alterações não sinônima “*missense*” (*Pdia4*, *Plekha8*, *Pigy*, *Polr1a*, *Gmcl1* e *Xpc*).

7 – REFERÊNCIAS BIBLIOGRÁFICAS

1. WHO. Global health risks. WHO Library Cataloguing-in-Publication Data. 2009;
2. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, Jr., Jones DW, Materson BJ, Oparil S, Wright JT, Jr. and RE. The Seventh Report of the Joint National Committee on Prevention , Detection , Evaluation , and Treatment of High Blood Pressure: the JNC 7 report. JAMA. 2003;289:2560–72.
3. Hajjar I, Kotchen TA. Trends in Prevalence , Awareness , Treatment, and Control of Hypertension in the United States , 1988-2000. JAMA. 2013;290(2):199–206.
4. Kearney PM, Whelton M, Reynolds K, Muntner P, Whelton PK, He J. Global burden of hypertension: analysis of worldwide data. Lancet [Internet]. 2005;365(9455):217–23. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15652604>
5. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet [Internet]. 2002 Dec 14;360(9349):1903–13. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12493255>
6. Lawes CMM, Vander Hoorn S, Rodgers A. Global burden of blood-pressure-related disease, 2001. Lancet [Internet]. 2008 May 3;371(9623):1513–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18456100>
7. Picon R V, Fuchs FD, Moreira LB, Riegel G, Fuchs SC. Trends in prevalence of hypertension in Brazil: a systematic review with meta-analysis. PloS one [Internet]. 2012 Jan [cited 2013 Aug 23];7(10):e48255. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3485225&tool=pmcentrez&rendertype=abstract>
8. Carretero O a., Oparil S. Essential Hypertension : Part I: Definition and Etiology. Circulation [Internet]. 2000 Jan 25 [cited 2013 May 22];101(3):329–35. Available from: <http://circ.ahajournals.org/cgi/doi/10.1161/01.CIR.101.3.329>
9. Feinleib M, Garrison R, Fabsitz R, Christian J, Hrubec Z, Borhani N, et al. The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results. Am J Epidemiol. 1977;106:284–5.
10. Longini IJ, Higgins M, Hinton P, Moll P, Keller J. Environmental and genetic sources of familial aggregation of blood pressure in Tecumseh, Michigan. Am J Epidemiol. 1984;120:131–44.

11. Biron P, Mongeau JG, Bertrand D. Familial aggregation of blood pressure in 558 adopted children. *Canadian Medical Association journal* [Internet]. 1976 Oct 23;115(8):773–4. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1878814&tool=pmcentrez&rendertype=abstract>
12. Rice T, Vogler GP, Perusse L, Bouchard C, Rao DC. Cardiovascular risk factors in a French Canadian population: resolution of genetic and familial environmental effects on blood pressure using twins, adoptees, and extensive information on environmental correlates. *Genetic epidemiology* [Internet]. 1989 Jan;6(5):571–88. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2591728>
13. De Oliveira CM, Pereira AC, de Andrade M, Soler JM, Krieger JE. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC medical genetics* [Internet]. 2008 Jan [cited 2013 Jun 28];9:32. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2386446&tool=pmcentrez&rendertype=abstract>
14. Ehret GB. Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Curr Hypertens Rep*. 2010;12(1):17–25.
15. Cowley AW. The genetic dissection of essential hypertension. *Nature reviews. Genetics* [Internet]. 2006 Nov [cited 2013 Aug 8];7(11):829–40. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17033627>
16. Kotchen T a., Kotchen JM, Grim CE, George V, Kaldunski ML, Cowley a. W, et al. Genetic Determinants of Hypertension: Identification of Candidate Phenotypes. *Hypertension* [Internet]. 2000 Jul 1 [cited 2013 Aug 23];36(1):7–13. Available from: <http://hyper.ahajournals.org/cgi/doi/10.1161/01.HYP.36.1.7>
17. Cowley Jr. A. Long-term control of arterial blood pressure. *Physiol Rev*. 1992;72:231–300.
18. Guyton a C. Blood pressure control--special role of the kidneys and body fluids. *Science (New York, N.Y.)* [Internet]. 1991 Jun 28;252(5014):1813–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2063193>
19. Guyton A. *Tratado da fisiologia médica*.
20. Zatz R. *Fisiopatologia Renal*.
21. Flatman PW. Cotransporters, WNKs and hypertension: an update. *Current opinion in nephrology and hypertension* [Internet]. 2008 Mar;17(2):186–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18277153>

22. 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. *Nature genetics* [Internet]. 2008 May [cited 2013 Aug 14];40(5):592–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18391953>
23. Monette MY, Rinehart J, Lifton RP, Forbush B. Rare mutations in the human Na-K-Cl cotransporter (NKCC2) associated with lower blood pressure exhibit impaired processing and transport function. *American journal of physiology. Renal physiology* [Internet]. 2011 Apr [cited 2013 Aug 23];300(4):F840–7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3074999&tool=pmcentrez&rendertype=abstract>
24. Simon DB, Karet FE, Hamdan JM, DiPietro A, Sanjad SA, Lifton RP. Bartter’s syndrome, hypokalaemic alkalosis with hypercalciuria, is caused by mutations in the Na-K-2Cl cotransporter NKCC2. *Nature Genetics*. 1996;13:183–8.
25. Wilson FH, Disse-Nicodème S, Choate K a, Ishikawa K, Nelson-Williams C, Desitter I, et al. Human hypertension caused by mutations in WNK kinases. *Science (New York, N.Y.)* [Internet]. 2001 Aug 10 [cited 2013 Aug 23];293(5532):1107–12. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11498583>
26. Rettig R, Grisk O. The kidney as a determinant of genetic hypertension: evidence from renal transplantation studies. *Hypertension* [Internet]. 2005 Sep [cited 2013 Aug 23];46(3):463–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16103270>
27. Curtis J, Luke R, Dustan H, Kashgarian M, Whelchel J, Jones P, et al. Remission of essential hypertension after renal transplantation. *N Engl J Med*. 1983;309:1009–15.
28. Guidi E, Bianchi G, Rivolta E, Ponticelli C, Quarto di Palo F, Minetti L, et al. Hypertension in man with a kidney transplant: role of familial versus other factors. *Nephron*. 1985;41:14–21.
29. Strandgaard S, Hansen U. Hypertension in renal allograft recipients may be conveyed by cadaveric kidneys from donors with subarachnoid haemorrhage. *British medical journal (Clinical research ed.)* [Internet]. 1986 Apr 19;292(6527):1041–4. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1340109&tool=pmcentrez&rendertype=abstract>
30. Kopf D, Waldherr R, Rettig R. Source of kidney determines blood pressure in young renal transplanted rats. *Am J Physiol*. 1993;265:104–11.

31. Simino J, Rao DC, Freedman BI. Novel findings and future directions on the genetics of hypertension. *Current opinion in nephrology and hypertension* [Internet]. 2012 Sep [cited 2013 Aug 23];21(5):500–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22614628>
32. Wang X, Prins BP, Söber S, Laan M, Snieder H. Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension. *Current hypertension reports* [Internet]. 2011 Dec [cited 2013 Aug 8];13(6):442–51. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3212682&tool=pmcentrez&rendertype=abstract>
33. Wellcome T, Case T, Consortium C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* [Internet]. 2007 Jun 7 [cited 2013 May 21];447(7145):661–78. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2719288&tool=pmcentrez&rendertype=abstract>
34. Basson J, Simino J, Rao DC. Between candidate genes and whole genomes: time for alternative approaches in blood pressure genetics. *Current hypertension reports* [Internet]. 2012 Feb [cited 2013 Jun 17];14(1):46–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22161147>
35. 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* [Internet]. Nature Publishing Group; 2011 Oct 6 [cited 2013 May 24];478(7367):103–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3340926&tool=pmcentrez&rendertype=abstract>
36. Guo Y, Tomlinson B, Chu T, Fang YJ, Gui H, Tang CS, et al. A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits. *PloS one* [Internet]. 2012 Jan [cited 2013 Jun 17];7(2). Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3286457&tool=pmcentrez&rendertype=abstract>
37. Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, et al. Genome-wide association study of blood pressure and hypertension. *Nature genetics*. 2009;41(6):677–87.
38. Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang S-J, et al. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC medical genetics* [Internet]. 2007 Jan [cited 2013 May 25];8 Suppl 1:S3. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1995621&tool=pmcentrez&rendertype=abstract>

39. Newton-Cheh C, Johnson T, Gateva V, Tobin MD, Bochud M, Coin L, et al. Genome-wide association study identifies eight loci associated with blood pressure. *Nature Genetics* [Internet]. 2009 Jun [cited 2013 Aug 8];41(6):666–76. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2891673&tool=pmcentrez&rendertype=abstract>
40. Wain L, Verwoert G, O'Reilly P, Shi G, Johnson T, Johnson A, et al. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature genetics*. 2012;43(10):1005–11.
41. Rapp JP. Genetic analysis of inherited hypertension in the rat. *Physiological reviews* [Internet]. 2000 Jan;80(1):135–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10617767>
42. Deng AY. Positional cloning of quantitative trait Loci for blood pressure: how close are we?: a critical perspective. *Hypertension* [Internet]. 2007 Apr [cited 2013 Aug 22];49(4):740–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17296871>
43. Bilusic M, Bataillard A, Tschannen MR, Gao L, Barreto NE, Vincent M, et al. Mapping the genetic determinants of hypertension, metabolic diseases, and related phenotypes in the lyon hypertensive rat. *Hypertension* [Internet]. 2004 Nov [cited 2013 Aug 11];44(5):695–701. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15452030>
44. Deng AY, Dene H, Rapp JP. Mapping of a Quantitative Trait Locus for Blood Pressure Rat Chromosome 2. *J Clin Invest*. 1994;94:431–6.
45. Herrera VLM, Tsikoudakis A, Ponce LRB, Matsubara Y, Ruiz-Opazo N. Sex-specific QTLs and interacting loci underlie salt-sensitive hypertension and target organ complications in Dahl S/jrHS hypertensive rats. *Physiological genomics* [Internet]. 2006 Aug 16 [cited 2013 Aug 23];26(3):172–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16720678>
46. Hilbert P, Lindpaintner K, Beckmann JS, Serikawa T, Soubrier F, Dubay C, et al. Chromosomal mapping of two genetic loci associated with blood-pressure regulation in hereditary hypertensive rats. *Nature*. 1991;353:521–9.
47. Jacob HJ, Lindpaintner K, Lincoln SE, Kusumi K, Bunker RK, Mao Y, et al. Genetic Mapping of a Gene Causing Hypertension in the Stroke-Prone Spontaneously Hypertensive Rat. *Cell*. 1991;67:213–24.
48. Nabika T, Kobayashi Y, Yamori Y. CONGENIC RATS FOR HYPERTENSION: HOW USEFUL ARE THEY FOR THE HUNTING OF HYPERTENSION GENES? *Clinical and Experimental Pharmacology and Physiology*. 2000;27(July):251–6.

49. Deng A, Dene H, Rapp J. Congenic strains for the blood pressure quantitative trait locus on rat chromosome 2. *Hypertension*. 1997;30:199–202.
50. Dukhanina OI, Dene H, Deng a Y, Choi CR, Hoebee B, Rapp JP. Linkage map and congenic strains to localize blood pressure QTL on rat chromosome 10. *Mammalian genome: official journal of the International Mammalian Genome Society* [Internet]. 1997 Apr;8(4):229–35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9096100>
51. Garrett MR, Joe B, Yerga-Woolwine S. Genetic linkage of urinary albumin excretion in Dahl salt-sensitive rats: influence of dietary salt and confirmation using congenic strains. *Physiological genomics* [Internet]. 2006 Mar 13 [cited 2013 Aug 22];25(1):39–49. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16534143>
52. Jeffs B, Negrin CD, Graham D, Clark JS, Niall H, Gauguier D, et al. Applicability of a “Speed” Congenic Strategy to Dissect Blood Pressure Quantitative Trait Loci on Rat Chromosome 2. *Hypertension*. 2000;35:179–87.
53. McBride MW, Carr FJ, Graham D, Anderson NH, James S, Lee WK, et al. Microarray Analysis of Rat Chromosome 2 Congenic Strains. *Hypertension*. 2003;41:847–53.
54. Rapp JP, Garrett MR, Dene H, Meng H, Hoebee B, Lathrop GM. Linkage analysis and construction of a congenic strain for a blood pressure QTL on rat chromosome 9. *Genomics* [Internet]. 1998 Jul 15;51(2):191–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/9722941>
55. Schork NJ, Krieger JE, Trolliet MR, Franchini KG, Koike G, Krieger EM, et al. A biometrical genome search in rats reveals the multigenic basis of blood pressure variation. *Genome Research* [Internet]. 1995 Sep 1 [cited 2013 Sep 18];5(2):164–72. Available from: <http://www.genome.org/cgi/doi/10.1101/gr.5.2.164>
56. Aneas I, Rodrigues M V, Pauletti B a, Silva GJJ, Carmona R, Cardoso L, et al. Congenic strains provide evidence that four mapped loci in chromosomes 2, 4, and 16 influence hypertension in the SHR. *Physiological genomics* [Internet]. 2009 Mar 3 [cited 2013 Aug 22];37(1):52–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19126752>
57. Metzker ML. Sequencing technologies - the next generation. *Nature reviews. Genetics* [Internet]. Nature Publishing Group; 2010 Jan [cited 2013 Aug 6];11(1):31–46. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19997069>
58. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions

- and gene fusions. *Genome biology* [Internet]. BioMed Central Ltd; 2013 Apr 25 [cited 2013 Aug 7];14(4):R36. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23618408>
59. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* (Oxford, England) [Internet]. 2009 May 1 [cited 2013 Aug 7];25(9):1105–11. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2672628&tool=pmcentrez&rendertype=abstract>
 60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2013;9(4):357–9.
 61. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*. 2013;7(3):562–78.
 62. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Baren MJ Van, et al. Transcript assembly and abundance estimation from RNA-Seq reveals thousands of new transcripts and switching among isoforms. *Nature Biotechnology*. 2011;28(5):511–5.
 63. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biology* [Internet]. BioMed Central Ltd; 2010 Jan [cited 2013 Aug 6];11(10):R106. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3218662&tool=pmcentrez&rendertype=abstract>
 64. Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC bioinformatics* [Internet]. 2010 Jan;11:94. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2838869&tool=pmcentrez&rendertype=abstract>
 65. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* (Oxford, England) [Internet]. 2009 Aug 15 [cited 2013 Aug 7];25(16):2078–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2723002&tool=pmcentrez&rendertype=abstract>
 66. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology* [Internet]. 2012 Jan [cited 2013 Aug 6];2012:251364. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3398667&tool=pmcentrez&rendertype=abstract>

67. Pauletti B a. Identificação de genes candidatos determinantes da hipertensão arterial no cromossomo 4 de ratos SHR. 2007. p. 142.
68. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, Gerstein M, et al. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* (New York, N.Y.) [Internet]. 2008 Jun 6 [cited 2013 Aug 10];320(5881):1344–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2951732&tool=pmcentrez&rendertype=abstract>
69. Harismendy O, Notani D, Song X, Rahim NG, Tanasa B, Heintzman N, et al. 9p21 DNA variants associated with coronary artery disease impair interferon- γ signalling response. *Nature* [Internet]. Nature Publishing Group; 2011 Feb 10 [cited 2013 Aug 12];470(7333):264–8. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3079517&tool=pmcentrez&rendertype=abstract>
70. Musunuru K, Strong A, Frank-Kamenetsky M, Lee NE, Ahfeldt T, Sachs K V, et al. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* [Internet]. Nature Publishing Group; 2010 Aug 5 [cited 2013 Aug 8];466(7307):714–9. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3062476&tool=pmcentrez&rendertype=abstract>
71. Georges M, Coppieters W, Charlier C. Polymorphic miRNA-mediated gene regulation: contribution to phenotypic variation and disease. *Current opinion in genetics & development* [Internet]. 2007 Jun [cited 2013 Aug 22];17(3):166–76. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17467975>
72. Atanur SS, Diaz AG, Maratou K, Sarkis A, Rotival M, Game L, et al. Genome Sequencing Reveals Loci under Artificial Selection that Underlie Disease Phenotypes in the Laboratory Rat. *Cell* [Internet]. The Authors; 2013 Jul [cited 2013 Jul 26];1–13. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867413007794>
73. Kang CB, Hong Y, Dhe-Paganon S, Yoon HS. FKBP family proteins: immunophilins with versatile biological functions. *Neuro-Signals* [Internet]. 2008 Jan [cited 2013 Aug 14];16(4):318–25. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18635947>
74. Jo D, Lyu M, Cho E, Park D, Kozak C, Kim M. Identification and genetic mapping of the mouse Fkbp9 gene encoding a new member of FK506-binding protein family. *Mols Cells*. 2001;12:272–5.
75. Hochrainer K, Mayer H, Baranyi U, Binder B, Lipp J, Kroismayr R. The human HERC family of ubiquitin ligases: novel members, genomic organization, expression profiling, and evolutionary aspects. *Genomics*

- [Internet]. 2005 Feb [cited 2013 Aug 23];85(2):153–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15676274>
76. Huibregtse JM, Scheffner M, Beaudenon S, Howley PM. A family of proteins structurally and functionally related to the. *Biochemistry*. 1995;92(March):2563–7.
 77. Correa RG, Milutinovic S, Reed JC. Roles of NOD1 (NLRC1) and NOD2 (NLRC2) in innate immunity and inflammatory diseases. *Bioscience reports* [Internet]. 2012 Dec [cited 2013 Aug 1];32(6):597–608. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3497720&tool=pmcentrez&rendertype=abstract>
 78. Baras A, Moskaluk C a. Intracellular localization of GASP/ECOP/VOPP1. *Journal of molecular histology* [Internet]. 2010 Apr [cited 2013 Aug 19];41(2-3):153–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20571887>
 79. Baras AS, Solomon A, Davidson R, Moskaluk C a. Loss of VOPP1 overexpression in squamous carcinoma cells induces apoptosis through oxidative cellular injury. *Laboratory investigation; a journal of technical methods and pathology* [Internet]. Nature Publishing Group; 2011 Aug [cited 2013 Aug 19];91(8):1170–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21519330>
 80. Endale Ahanda M-L, Fritz ER, Estellé J, Hu Z-L, Madsen O, Groenen M a M, et al. Prediction of altered 3'-UTR miRNA-binding sites from RNA-Seq data: the swine leukocyte antigen complex (SLA) as a model region. *PloS one* [Internet]. 2012 Jan [cited 2013 Aug 22];7(11):e48607. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3490867&tool=pmcentrez&rendertype=abstract>
 81. Wynn RM, Li J, Brautigam C a, Chuang JL, Chuang DT. Structural and biochemical characterization of human mitochondrial branched-chain α -ketoacid dehydrogenase phosphatase. *The Journal of biological chemistry* [Internet]. 2012 Mar 16 [cited 2013 Aug 23];287(12):9178–92. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3308798&tool=pmcentrez&rendertype=abstract>
 82. Lu G, Ren S, Korge P, Choi J, Dong Y, Weiss J, et al. A novel mitochondrial matrix serine / threonine protein phosphatase regulates the mitochondria permeability transition pore and is essential for cellular survival and development. *Genes & Development*. 2007;21:784–96.
 83. Taneera J, Lang S, Sharma A, Fadista J, Zhou Y, Ahlqvist E, et al. A systems genetics approach identifies genes and pathways for type 2 diabetes in human islets. *Cell metabolism* [Internet]. 2012 Jul 3 [cited 2013 Aug 11];16(1):122–34. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22768844>

84. Burré J, Sharma M, Tsetsenis T, Buchman V, Etherton MR, Südhof TC. Alpha-synuclein promotes SNARE-complex assembly in vivo and in vitro. *Science* (New York, N.Y.) [Internet]. 2010 Sep 24;329(5999):1663–7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3235365&tool=pmcentrez&rendertype=abstract>
85. Cabin DE, Shimazu K, Murphy D, Cole NB, Gottschalk W, McIlwain KL, et al. Synaptic Vesicle Depletion Correlates with Attenuated Synaptic Responses to Prolonged Repetitive Stimulation in Mice Lacking alpha -Synuclein. *The Journal of Neuroscience*. 2002;22(20):8797–807.
86. Larsen KE, Schmitz Y, Troyer MD, Mosharov E, Dietrich P, Quazi AZ, et al. Alpha-synuclein overexpression in PC12 and chromaffin cells impairs catecholamine release by interfering with a late step in exocytosis. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [Internet]. 2006 Nov 15 [cited 2013 Aug 23];26(46):11915–22. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17108165>
87. Murphy DD, Rueter SM, Trojanowski JQ, Lee VM. Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [Internet]. 2000 May 1;20(9):3214–20. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10777786>
88. Nemani VM, Lu W, Berge V, Nakamura K, Onoa B, Michael K, et al. Increased expression of alpha-synuclein reduces neurotransmitter release by inhibiting synaptic vesicle reclustering after endocytosis. *Neuron*. 2011;65(1):66–79.
89. Yavich L, Tanila H, Vepsäläinen S, Jäkälä P. Role of alpha-synuclein in presynaptic dopamine recruitment. *The Journal of neuroscience : the official journal of the Society for Neuroscience* [Internet]. 2004 Dec 8 [cited 2013 Aug 23];24(49):11165–70. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15590933>
90. Chartier-Harlin M-C, Kachergus J, Roumier C, Mouroux V, Douay X, Lincoln S, et al. alpha-synuclein locus duplication as a cause of familial Parkinson ' s disease. *The Lancet*. 2004;07:1167–9.
91. Fuchs J, Nilsson C, Kachergus J, Munz M, Larsson E, Schule B, et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. *Neurology*. 2006;25:39–49.
92. Krüger R, Kuhn W, Müller T, Woitalla D, Graeber M, Kösel S, et al. Ala30Pro mutation in the gene encoding alpha-synuclein in Parkinson's disease. *Nature*. 1998;18:106–8.

93. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha -Synuclein Locus Triplication Causes Parkinson ' s Disease. *Science*. 2003;302(October):2003.
94. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, Beecham GW, et al. Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. *Am Hum Genet*. 2011;74(2):97–109.
95. Wu S, Hultquist A, Hyldbring P, Cetinkaya C, Oberg F, Larsson L-G. TGF-beta enforces senescence in Myc-transformed hematopoietic tumor cells through induction of Mad1 and repression of Myc activity. *Experimental cell research* [Internet]. Elsevier Inc.; 2009 Nov 1 [cited 2013 Aug 7];315(18):3099–111. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19766114>
96. Cultraro C, Bino T, Segal S. Regulated expression and function of the c-Myc antagonist, Mad1, during a molecular switch from proliferation to differentiation. *Curr Top Microbiol Immunol*. 1997;224:149–58.
97. Lambeth DO. Reconsideration of the significance of substrate-level phosphorylation in the citric acid cycle*. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* [Internet]. 2006 Jan;34(1):21–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21638627>
98. Johnson JD. Genetic Evidence for the Expression of ATP- and GTP-specific Succinyl-CoA Synthetases in Multicellular Eucaryotes. *Journal of Biological Chemistry* [Internet]. 1998 Oct 16 [cited 2013 Aug 23];273(42):27580–6. Available from: <http://www.jbc.org/cgi/doi/10.1074/jbc.273.42.27580>
99. Staud F, Pavek P. Breast cancer resistance protein (BCRP/ABCG2). *The international journal of biochemistry & cell biology* [Internet]. 2005 Apr [cited 2013 Aug 13];37(4):720–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15694832>
100. Huls M, van den Heuvel JJMW, Dijkman HBPM, Russel FGM, Masereeuw R. ABC transporter expression profiling after ischemic reperfusion injury in mouse kidney. *Kidney international* [Internet]. 2006 Jun [cited 2013 Aug 21];69(12):2186–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16612327>
101. Huls M, Brown CD a, Windass a S, Sayer R, van den Heuvel JJMW, Heemskerk S, et al. The breast cancer resistance protein transporter ABCG2 is expressed in the human kidney proximal tubule apical membrane. *Kidney international* [Internet]. 2008 Jan [cited 2013 Aug 21];73(2):220–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17978814>

102. The MGC Project Team. The Status , Quality , and Expansion of the NIH Full-Length cDNA Project: The Mammalian Gene Collection (MGC). *Genome Research*. 2004;2121–7.
103. Ding Y, Ko MH, Pehar M, Kotch F, Peters NR, Luo Y, et al. Biochemical inhibition of the acetyltransferases ATase1 and ATase2 reduces β -secretase (BACE1) levels and A β generation. *The Journal of biological chemistry* [Internet]. 2012 Mar 9 [cited 2013 Aug 22];287(11):8424–33. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3318698&tool=pmcentrez&rendertype=abstract>
104. Ellgaard L, Ruddock LW. The human protein disulphide isomerase family: substrate interactions and functional properties. *EMBO reports* [Internet]. 2005 Jan [cited 2013 Aug 13];6(1):28–32. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1299221&tool=pmcentrez&rendertype=abstract>
105. Maattanen P, Kozlov G, Gehring K, Thomas DY. ERp57 and PDI: multifunctional protein disulfide isomerases with similar domain architectures but differing substrate – partner associations 1. *Biochem. Cell Biol*. 2006;57:881–9.
106. D’Angelo G, Uemura T, Chuang C-C, Polishchuk E, Santoro M, Ohvo-Rekilä H, et al. Vesicular and non-vesicular transport feed distinct glycosylation pathways in the Golgi. *Nature* [Internet]. 2013 Aug 4 [cited 2013 Aug 8];1–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23913272>
107. Murakami Y, Siripanyaphinyo U, Hong Y, Tashima Y, Maeda Y, Kinoshita T. The Initial Enzyme for Glycosylphosphatidylinositol Biosynthesis Requires PIG-Y , a Seventh Component. *Molecular Biology of the Cell*. 2005;16(November):5236–46.
108. Tsai Y-C, Greco TM, Boonmee A, Miteva Y, Cristea IM. Functional proteomics establishes the interaction of SIRT7 with chromatin remodeling complexes and expands its role in regulation of RNA polymerase I transcription. *Molecular & cellular proteomics: MCP* [Internet]. 2012 Feb [cited 2013 Aug 22];11(2):M111.015156. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3277772&tool=pmcentrez&rendertype=abstract>
109. Kimura T, Ito C, Watanabe S, Takahashi T, Ikawa M, Yomogida K, et al. Mouse Germ Cell-Less as an Essential Component for Nuclear Integrity. *Molecular and Cellular Biology*. 2003;23(4):1304–15.
110. Dai Q-S, Hua R-X, Zeng R-F, Long J-T, Peng Z-W. XPC gene polymorphisms contribute to bladder cancer susceptibility: a meta-analysis. *Tumour biology: the journal of the International Society for*

Oncodevelopmental Biology and Medicine [Internet]. 2013 Aug 6 [cited 2013 Aug 22]; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23918308>

111. KANNEL WB, DAWBER TR, KAGAN A, REVOTSKIE N, STOKES III J. Factors of Risk in the Development of Coronary Heart Disease—Six-Year Follow-up Experience: The Framingham Study. *Annals of Internal Medicine*. 1961;55(1):33–50.
112. KANNEL WB. Blood Pressure as a Cardiovascular Risk Factor: Prevention and Treatment. *JAMA*. 1996;275(20):1571–6.
113. Krieger JE, Dzau VJ. Molecular biology of hypertension. *Hypertension* [Internet]. 1991 Sep 1 [cited 2013 Sep 22];18(3_Suppl):I3–I3. Available from: http://hyper.ahajournals.org/cgi/content/long/18/3_Suppl/I3
114. Krieger JE. New contributions to clinical hypertension from molecular biology. *Current Opinion in Cardiology*. 1998;13(5).
115. De Oliveira CM, Pereira AC, de Andrade M, Soler JM, Krieger JE. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC medical genetics* [Internet]. 2008 Jan [cited 2013 Aug 22];9:32. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2386446&tool=pmcentrez&rendertype=abstract>
116. Moriyama T, Murphy H, Martin B, Garcia-Perez A. Detection of specific mRNAs in single nephron segments by use of the polymerase chain reaction. *Am J Physiol*. 1990;258:1470–4.

APÊNDICES

Apêndice I

Identificação de genes candidatos determinantes da hipertensão arterial no cromossomo 4 de ratos *SHR* utilizando-se de qRT-PCR

Para a identificação de genes do intervalo envolvidos com o fenótipo aqui estudado, em fases anteriores a este projeto de mestrado (67), foi adotado a análise de padrão de expressão gênica diferencial entre o congênico e o *SHR* a partir da metodologia de qRT-PCR. Inicialmente, foram identificados quais dos genes do intervalo possuem expressão renal e, posteriormente, quais apresentavam diferença de expressão entre as linhagens em pool de amostras (n=8) de rim proveniente de animais congênicos e *SHR* em fase adulta (12 a 14 semanas de vida). Os genes que apresentaram diferença significativa de expressão entre as linhagens foram, então, confirmadas em amostras individuais de animais sob a mesma condição. Dentre os 105 genes considerados expressos no rim até aquele momento, 16 foram significativamente diferentes nas amostras de tecido renal de animais adultos em amostras individuais (Figura Suplementar 7A).

Entretanto, como a pressão arterial dos animais *SHR* é maior do que a dos congênicos na fase adulta e este fator per se pode interferir no controle da expressão gênica secundariamente, a expressão dos 16 genes foi avaliada em amostras individuais de rim de animais jovens dos 2 grupos no que é considerado a fase pré-hipertensa (5-7 semanas de vida), onde os níveis de pressão arterial são semelhantes. Apesar dessa estratégia não levar em conta alterações na expressão gênica associadas aos estágios de desenvolvimento, que podem estar exercendo papel determinante na hipertensão, acredita-se que esta é uma estratégia razoável para excluir efeitos secundários grosseiros.

Os resultados obtidos na análise da expressão gênica dos 16 genes em amostras de rim de animais pré-hipertensos mostraram que somente 7 permaneceram diferencialmente expressos quando a pressão era a mesma (Figura Suplementar 7B).

Posteriormente, esses 16 genes foram também avaliados em 5 segmentos do néfron: túbulo proximal convoluto, túbulo proximal reto, Alça de Henle espessa, duto coletor cortical e duto coletor medular (dados não publicados). Para isso, os rins foram micro dissecados segundo metodologia de Moriyama et al (116) e a expressão quantificada pela metodologia de qRT-PCR anteriormente utilizada (67). A análise da expressão desses genes em segmentos do néfron é uma informação fundamental, já que segue a ideia de que genes expressos no túbulo renal possam estar participando do metabolismo de sal e água e, conseqüentemente, de sua disfunção. A análise mostrou que todos os 7 genes diferencialmente expressos em amostra de rim total de animais jovens mantiveram este perfil em 1 ou mais segmentos do néfron. Foi possível observar também que algumas das diferenças de expressão gênica obtidas na análise de amostras de rim total entre animais adulto congênicos e *SHR* foram amplificadas quando analisadas em diferentes segmentos do néfron (como, por exemplo, os genes *Retsat*, *Dguok* e *Tex261* no túbulo proximal). Além disso, alguns genes não apresentaram expressão em nenhum dos fragmentos estudados (*Tra2a/RGD1562563* e *Dusp11*) ou apresentaram diferenças de expressão nos segmentos do néfron com direções contrárias àquelas obtidas quando analisadas em amostras de rim total (Figura Suplementar 8).

Figura Suplementar 2. Diferença de expressão gênica em diferentes segmentos do néfron entre congênico *SHR.BN4* e *SHR* em amostras adultas. A figura mostra a expressão gênica nos animais congênicos com base nos valores de expressão obtidos no *SHR*. TPC – túbulo proximal convoluto, TPR – túbulo proximal reto, AHE – Alça de Henle espessa, DCC – duto coletor cortical e DCM – duto coletor medular. * representa genes diferencialmente expressos entre as linhagens com $p < 0,05$

Dentre esses 16 genes previamente selecionados, 6 são comuns a análise do RNA-seq, sendo que em geral os dados corroboram entre as duas metodologias (Tabela Suplementar 9).

Tabela Suplementar 1 - Genes diferencialmente expressos entre linhagens por RNA-seq e por qRT-PCR

gene	RNAseq		Pool	adulto	pre	qRT-PCR				
	fold change Cufflinks	fold change DEseq				Basal	ao longo do nefron			
						TPC	TPR	AHE	DCC	DCM
Y_RNA	3,509397818	3,411006537				Não testado				
Y_RNA	6,769080982	6,733617571				Não testado				
Zfp467	1,540430277	1,538282863				Não testado				
Rarres2	1,394379507	1,929292939				Não testado				
Tmem176b	1,209043434	1,189628962	1,33							
LOC500124	1,862870478	1,8379147	1,81	1,82	0,96	0,74	0,78	0,31	0,33	0,16
Cpvl	3,470331142	3,83518604				Não testado				
MGC95152	1,202023322	1,291463581	2,19	1,63	2,01	1,70	1,52	0,41	0,41	0,67
Nod1	1,666135666	1,665812615				Não testado				
Fkbp9	0,725106546	0,727480791	0,59	0,56	0,87	1,67	1,53	0,83	0,84	0,78
Ecop	1,831413297	1,836433676				Não testado				
Herc6	1,621851367	1,543973592				Não testado				
Ppm1k	1,317089262	1,430601835				Não testado				
Abcg2	1,167311611	1,199193885	0,67							
Snca	2,133632796	2,165198203				Não testado				
FILYY3_RAT	0	0				Não testado				
F1M5H9_RAT	0	0				Não testado				
Ptcd3	2,234372233	1,338799612	1,40	1,27						
Vamp5	1,611451315	1,351847249				Não testado				
Retsat	2,129231993	2,138147267	2,12	3,71	1,77	15,42	10,39	1,36	0,89	0,32
Suclg1	0,626068192	0,647663178	1,07							
Dguok	0,638728401	0,788124528	1,95	2,47	2,12	8,52	8,26	4,04	1,45	0,53
Atp6v1b1	0,774009707	0,811520479				Não testado				
Smyd5	0,752406296	0,739344572				Não testado				
Cml1	2,870470849	2,758280383	2,75	3,64	3,80	1,95	4,38	1,26	1,13	0,66
Mxd1	1,298585991	1,255677792				Não testado				
Mcm2	1,319537425	1,327569077				Não testado				
Klf15	1,287234687	1,28538619				Não testado				

NOTA: No qRT-PCR, foram representados em vermelho os “*fold change*” significativos, com $p < 0,05$.

Apêndice II

The role of blood pressure SNPs on the risk prediction of hypertension

Samantha Kuwada Teixeira¹

Livia Serra Selvatici¹

Rafael Ferreira Reis¹

Camila Bonin Pinto¹

José Geraldo Mill²

Alexandre da Costa Pereira¹

José Eduardo Krieger¹

¹ Laboratory of Genetics and Molecular Cardiology, Heart Institute (InCor), University of São Paulo Medical School, São Paulo, Brazil.

² Department of Physiology, Espirito Santo Federal University, São Paulo, Brazil.

Abstract

Genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNPs) associated with blood pressure (BP) and hypertension (HTN). Unfortunately, these SNPs explain maximally 2% of the BP variation. In this study, we validated the association of these SNPs and tested their use to predict hypertension in combination with clinical data. Fifty eight SNPs previously associated with BP phenotypes were genotyped in 1504 individuals from a cross-sectional population-based sample from the Brazilian urban population. The influence of genetics on predictive models for hypertension was assessed by the area under receiver operating characteristic curves (ROC-AUC) from baseline and 5 year follow-up data. Individuals were stratified in 3 groups (high, intermediate and low risk of hypertension) using the cutoffs generated by coordinates of ROC-AUCs. Ten SNPs were associated with at least one BP related phenotype and seventeen showed interaction with age and/or body mass index (BMI). The discriminatory ability of Model 1 (genetic markers without interaction) to predict hypertension was weak (ROC-AUC: 0.553; 95% CI: 0.521-0.586), while the model with genetic markers interacting with age and/or BMI (Model 2) showed comparable discrimination as the model employing clinical variables (Model 4) (ROC-AUC: 0.766; 95% CI: 0.740-0.792 vs. ROC-AUC: 0.762; 95% CI: 0.735-0.788, respectively). Interestingly, the association of all genetic markers and clinical variables (Model 3) displayed the best discriminatory ability (ROC-AUC: 0.785; 95% CI: 0.760-0.810). Model 3 was also significantly able to predict 5-year hypertensive incidence (ROC-AUC: 0.659; 95% CI: 0.588-0.729) and showed a better stratification of hypertension risk, once it reclassified mainly individuals previously identified as being intermediate risk by Model 4. These findings suggest that genetic variants, that isolated explain a minute difference in BP, may be useful for early stratification of hypertensive risk in the general population when incorporated into more complex algorithms including clinical information.

Introduction

High Blood Pressure affects one billion people worldwide and is a major risk factor for mortality in the world, that contributes to 7.5 million deaths annually^{1,2}. Nearly 95 per cent of hypertensive patients have essential hypertension, a complex disease in which the causes remain largely unknown³. It is widely recognized the genetic influence on the disease with heritability estimation for systolic and diastolic blood pressure from 25 to 68 per cent^{4,5}.

Over the past decade, genome-wide association studies (GWAS) have identified several genetic variants associated with blood pressure (BP) and hypertension (HTN)⁶⁻¹⁶. Unfortunately, these results have been somewhat disappointing since these variants have individually or collectively small effect sizes, explaining only 1-2 percent on BP variation. Moreover, the association of these variants with hypertension has been poorly replicated among populations, and little is known about their actual contribution in the disease and their usefulness in hypertension prediction and prevention.

Therefore, the present study was undertaken to test the hypothesis that SNPs associated with hypertension published previously and validated for this study population can contribute to improve the prediction ability of clinical information.

Methods

Study Population

A cross-sectional study to determine the prevalence and severity of cardiovascular risk factors was performed in an urban population of Vitória, Brazil. Data were collected following the general guidelines of the WHO MONICA Project between 1999 and 2001¹⁷. A sample of 2044 individuals (from an eligible population of 137330) of either gender, 25 to 65 years of age, was chosen according to the nearest birthday after a random selection of domiciles.

In total, 1507 subjects attended the clinic visit and were evaluated for height, weight, smoking status, blood pressure measurements, and use of medicines. Blood glucose, total cholesterol, lipoproteins fractions, and triglycerides were assessed by standard techniques in 12-hour fasting blood samples.

Subjects were classified as white or non-white according to a set of phenotypes (skin color, shape of the nose, hair texture, aspect of the lip, and jaw position).

The project was approved by Espírito Santo Federal University ethic committee and all subjects provided written informed consent.

After five years, 619 out of 1507 individuals selected at baseline were revisited to evaluate the same variables.

Procedures

Blood pressure was measured in the sitting position using standard mercury sphygmomanometer on the left arm after a 5 to 10 minutes rest period. The first and fifth phases of Korotkoff sound were used to indicate systolic and diastolic blood pressure, respectively. The systolic and diastolic blood pressure for each patient were obtained from the mean values of two measurements carried out by two independent expert examiners, within an interval of 10 minutes between the two assessments. Pulse pressure was considered the difference between systolic and diastolic blood pressure.

For biochemical analysis, blood samples were collected soon after blood pressure measurements. Glucose, total cholesterol, lipoproteins fractions, and triglycerides were determined by using validated commercial analytical kits.

Anthropometric parameters, as weight and height, were obtained by technicians as previously reported¹⁸. BMI was calculated as body weight (kg) divided by the square height (m²).

Subjects were considered hypertensive with a systolic blood pressure ≥ 140 mmHg and/or a diastolic blood pressure ≥ 90 mmHg or if they were on anti-hypertensive drugs, including diuretics¹⁹. Hyperglycemia was defined when the fasting blood glucose was ≥ 110 mg/dl.

Genotyping

We selected 58 SNPs previously associated with BP in GWAS or with phenotypes involved with hypertension, as successful aging and stiffness (Table S1)^{6-14,20}. We were unable to design primers and probes for 16 of the 58 SNPs selected for genotyping using Taqman technology (Applied Biosystem, Foster City, CA). For those, we selected SNPs in strong linkage disequilibrium using HaploReg tool (v. 1,

Broad Institute/IMT), setting CEU population as reference (Table S1). To identify the corresponding risk allele of the LD SNP, we compared the allele's frequency of the two SNPs using HaploView tool or information from the 1000 Genomes Project database.

Statistical Analysis

Association analysis between SNPs and SBP, DBP or PP was performed using general linear model, adjusting for age, gender, and ethnicity. Association analysis between SNPs and hypertension was performed using logistic regression, adjusting for the same covariates above.

Association analysis between SNPs with significant interaction effects with age or BMI and SBP, DBP or PP was performed using general linear model, adjusting for age or BMI, respectively, in which the interaction between SNP and age or BMI was inputted in the model. Association analysis between SNPs in interaction with age or BMI with hypertension was performed using logistic regression, and was analyzed and adjusted as described above.

The incremental value for hypertension prediction of the validated genetic markers was investigated with several different prediction models using baseline information. The discriminatory ability of each model was determined by the area under receiver operating characteristic curves (ROC-AUCs), using the probabilities estimated from the final multivariate logistic regression models. In our models, we considered clinical information, genetic markers, and/or the interaction between them that showed association with one of the BP phenotypes tested and presented p-value $\leq 0,1$ in the univariate logistic regression (Figure S1 and Tables S2-5). To assess the improvement in prediction of model 2 and 3 in comparison with model 4, differences in ROC-AUCs and 95% confidence intervals were calculated with Delong test²¹. The integrated discrimination improvement (IDI) was calculated with predicted risks from logistic regression and the net reclassification improvement (NRI) was calculated using the cutoff generated by coordinates of the ROC-AUCs²².

Considering only individuals without hypertension at baseline, the discriminatory ability of models 3 and 4 was assessed by ROC-AUCs in incident cases identified during follow-up.

Using the cutoffs generated by coordinates of ROC-AUCs from models 3 and 4 at baseline and follow-up analysis, individuals were stratified in 3 groups in each model: high risk, intermediate risk and low risk of hypertension.

All the statistical analyses were performed with SPSS software version 13.0.

Results

The characteristics of the Vitória Population Study are given in Table 1. We studied and genotyped 1504 individuals, which included 638 cases of hypertension.

Ten of the 58 SNPs evaluated were associated with at least one BP phenotype in this population. Six SNPs were associated with SBP, four with DBP, three with PP and five with hypertension. Moreover, three of risk alleles previously shown to increase BP or risk of hypertension displayed the opposite effect in the present analysis (rs1799945, rs198833, and rs805303) (Table 2).

On the other hand, seventeen of the 58 SNPs were associated with at least one of BP phenotype (SBP, DBP, PP and hypertension) in interaction with age (N=7, Table 3) and/or BMI (N=14, Table 4).

Figure 2 and Table 5 depict the comparisons between the discriminatory ability to predict hypertension among the models (1-4). The discrimination ability of the genetic model including SNPs without interaction with BMI and/or age (Model 1) was weak. On the other hand, the model including SNPs that displayed interaction with age and/or BMI (Model 2) showed comparable discrimination ability to predict hypertension as the model employing clinical variables (Model 4). Interestingly, the association of all genetic markers and clinical variables (Model 3) displayed the best discrimination ability to predict hypertension. Moreover, IDI and NRI were significantly different between Model 3 and 4, inferring that Model 3 improved risk classification (Table 6).

As only Model 3 showed a significant difference in the discriminatory ability compared with Model 4, these two models had their discriminatory ability also compared in incident cases identified during follow-up (Figure S2 and Table S6). Both models showed a significant discriminatory ability (Model 3 ROC-AUC: 0.659; 95% CI: 0.588-0.729, and Model 4 ROC-AUC: 0.688; 95% CI: 0.618-0.757), but no difference was found between them (p-value=0.0639).

In contrast, the risk strata score based on Model 4 was less precise and showed more individuals in intermediate risk than Model 3 (Figure 2) considering the risk stratification obtained from cutoffs of ROC-AUCs coordinates at baseline and follow-up analysis (Table S7). Moreover, accuracy of model 3 was superior of that of model 4 both at baseline and at follow up (Table 7 and Table S8).

Discussion

Taken together, our data showed that combination of clinical and genetic information can improve the stratification of hypertension risk in the general population, even though the genetic polymorphisms individually or collectively explain a minute amount of the variance of BP or risk of hypertension. In our study, we investigated the usefulness of SNPs previously associated with BP phenotypes to predict hypertension and whether the genetic information could improve the prediction ability of using only clinical information. We tested 58 SNPs in a cross-sectional population-based sample from Brazil. The validation of GWAs in an admixed population is important since there is a true variability in association among different populations.

Over the past decade, with the completion of the human genome sequence, the development of dbSNP database, the International HapMap Project, and the improvements in the SNP genotyping technologies, GWAS became a powerful tool to map the genetic basis of common human diseases²³. In this context, in 2007, The Wellcome Trust Case Control Consortium published a joint GWA study identifying genetic variants associated with the seven major common human diseases, among them the essential hypertension. Surprisingly, contrary to other diseases, no strong association was found for hypertension⁶.

From these findings, other researchers did not measure efforts to identify genetic determinants associated with hypertension using this methodology and again dozens of polymorphisms have been associated with hypertension and/or blood pressure phenotypes that individually or collectively explain 1-2 percent of the blood pressure variance or the risk of hypertension. Furthermore, little is known about their actual contribution to experimentally investigate their influence on the relevant cell

processes and therefore their usefulness to develop novel diagnostic and therapeutic strategies⁶⁻¹⁶.

Several studies that have analyzed the contribution of SNPs associated with other complex diseases and their risk prediction²⁴⁻²⁷. Recently, Havulina et al and Ehret et al published studies that estimated genetic risk score (GRS) based on SBP and DBP and identified an association between them and hypertension. It is important to emphasize that no attempt was made to analyze their contribution to hypertension risk prediction^{11,26}.

Ehret et al demonstrated that among 28 loci that contain 29 SNPs associated with BP in European descendants, only nine loci showed association in individuals of East Asian ancestry and six in individuals of South Asian ancestry¹¹. A recent study by Xu et al found that SNPs in STK39 gene previously associated with hypertension in Caucasian population²⁸ and further validated in African Americans subjects⁷ were not associated in the Chinese population²⁹. Our findings are consistent with these data, since only ten out of 58 SNPs showed significant association with BP phenotypes in Brazilian population. Furthermore, three of them demonstrated the risk allele contrary to previously reported. The lack of significant association for individual SNPs and the differences in the risk alleles may reflect differences in the genetic architecture between populations due to differences in allele frequencies or in linkage disequilibrium patterns, as indicated, even though we cannot rule out issues regarding sample size and the imprecise ancestry characterization.

Thus, 26 out of 58 previously reported SNPs were validated and used to develop three models incorporating genetic information for determining risk prediction of hypertension. These models were compared with a model designed without genetic contribution, but with clinical information known to be associated with high blood pressure. As expected the prediction power of the pure genetic model was weak compared to the clinical model (Model 4) and the model with polymorphisms with interaction with BMI and/or age (Model 2). Interestingly, combination of all genetic information with clinical data rendered Model 3, which enabled 1. improvement in stratification of individuals in regard to hypertension risk and 2. higher concordance of hypertension prevalence with empirical data. This type of analyses was possible since our experimental design contemplated a subset of

follow up data. On the other hand, the analysis failed to improve prediction ability to assess hypertension risk.

Although disappointing, these findings are supported by the results of similar studies in complex diseases, such as, type 2 diabetes and cardiovascular diseases^{25,26}. According to Schulze et al (2009), the contribution of 20 SNPs was not informative for predicting future diabetes²⁴. In another study, Mühlenbruch (2013) showed that the genetic risk score calculated through genotyping of 42 SNPs did not improve the diabetes risk prediction above classic and metabolic markers. Although, they suggested that genetic variants could be useful in diabetes risk prediction within subgroups that already manifest risk factors²⁵.

Havulinna et al have demonstrated that a GRS based on 32 SNPs previously associated with BP phenotypes were associated with risk of cardiovascular disease. However, they demonstrated that the genetic information did not improve the cardiovascular risk prediction above Framingham Risk Score. Nevertheless, they concluded that Framingham Risk Score predict 10-years risk while the GRS estimate a lifetime risk, once genotype do not change over time²⁶.

In agreement with these results, we did not show significant difference between models prediction in the follow-up, but Model 3 displayed a better stratification of hypertension risk in general population, since it reclassified mainly individuals previously identified within intermediate risk by Model 4. One may speculate that this type of data provides the basis for development of improved strategies to manage these patients by public health care programs, postponing a re-evaluation of individuals reclassified as low risk or present monitoring blood pressure and adopting preventive measures for hypertension of individuals reclassified as high risk.

These findings suggest that genetic variants that isolated explain a minute difference in BP may be useful for early stratification of hypertensive risk in the general population when incorporated into more complex algorithms including clinical information.

References

- 1 WHO. Global health risks. *WHO Library Cataloguing-in-Publication Data*. 2009.
- 2 Lawes CMM, Vander Hoorn S, Rodgers A. Global burden of blood-pressure-related disease, 2001. *Lancet*. 2008; 371:1513–8.
- 3 Carretero OA, Oparil S. Essential Hypertension: Part I: Definition and Etiology. *Circulation*. 2000; 101:329–335.
- 4 De Oliveira CM, Pereira AC, de Andrade M, Soler JM, Krieger JE. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study. *BMC medical genetics*. 2008; 9:32.
- 5 Ehret GB. Genome-Wide Association Studies: Contribution of Genomics to Understanding Blood Pressure and Essential Hypertension. *Curr Hypertens Rep*. 2010; 12:17–25.
- 6 Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007; 447:661–78.
- 7 Adeyemo A, Gerry N, Chen G, Herbert A, Doumatey A, Huang H, Zhou J, Lashley K, Chen Y, Christman M, Rotimi C. A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS genetics*. 2009; 5(7):e1000564.
- 8 Levy D, Ehret GB, Rice K, Verwoert GC, Launer LJ, Dehghan A, *et al*. Genome-wide association study of blood pressure and hypertension. *Nature genetics*. 2009; 41:677–687.
- 9 Newton-Cheh C, Johnson T, Gateva V, Martin D, Bochud M, Coin L, *et al*. Eight blood pressure loci identified by genome-wide association study of 34,433 people of European ancestry. *Nature genetics*. 2010; 41:666–676.
- 10 Padmanabhan S, Melander O, Johnson T, Di Blasio AM, Lee WK, Gentilini D, *et al*. Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension. *PLoS genetics*. 2010; 6(10):e1001177.

- 11 International Consortium for Blood Pressure Genome-Wide Association Studies. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature*. 2011; 478:103–109.
- 12 Kato N, Takeuchi F, Tabara Y, Kelly TN, Go MJ, Sim X, *et al*. Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in East Asians. *Nature genetics*. 2011; 43:531–538.
- 13 Wain L, Verwoert G, O'Reilly P, Shi G, Johnson T, Johnson A, *et al*. Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure. *Nature genetics*. 2012; 43:1005–1011.
- 14 Guo Y, Tomlinson B, Chu T, Fang YJ, Gui H, Tang CS, Yip BH, Cherny SS, Hur YM, Sham PC, Lam TH, Thomas NG. A genome-wide linkage and association scan reveals novel loci for hypertension and blood pressure traits. *PloS one*. 2012; 7(2):e31489.
- 15 Org E, Eyheramendy S, Juhanson P, Gieger C, Lichtner P, Klopp N, *et al*. Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations. *Human molecular genetics*. 2009; 18:2288–2296.
- 16 Levy D, Larson MG, Benjamin EJ, Newton-Cheh C, Wang TJ, Hwang S-J, Vasani RS, Mitchell GF. Framingham Heart Study 100K Project: genome-wide associations for blood pressure and arterial stiffness. *BMC medical genetics*. 2007; 8 Suppl 1:S3.
- 17 Monica WHO. THE WORLD HEALTH ORGANIZATION MONICA PROJECT (MONITORING TRENDS AND DETERMINANTS IN CARDIOVASCULAR INTERNATIONAL DISEASE): A MAJOR INTERNATIONAL COLLABORATION. *J Clin Epidemiol*. 1988; 41:105–114.
- 18 Rodrigues SL, Baldo MP, Lani L, Nogueira L, Mill JG, Sa Cunha R De. Body mass index is not independently associated with increased aortic stiffness in a Brazilian population. *American journal of hypertension*. 2012; 25:1064–9.

- 19 The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA*. 2003; 289:2560–2572.
- 20 Kumari M, Lunetta KL, Milaneschi Y. A Genome-Wide Association Study of Aging. *Neurobiol Aging*. 2011; 32:1–21.
- 21 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988; 44:837–45.
- 22 Pencina MJ, D'Agostino Sr. RB, D'Agostino Jr. RB, Vasan RS. Evaluating the added predictive ability of a new marker : From area under the ROC curve to reclassification and beyond. *Statistic in Medicine*. 2008; 27:157–172.
- 23 Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nature reviews Genetics*. 2005; 6:95–108.
- 24 Schulze MB, Weikert C, Pischon T, Bergmann MM, Al-Hasani H, Schleicher E, Fritsche A, Häring H-U, Boeing H, Joost H-G. Use of Multiple Metabolic and Genetic Markers to Improve the Prediction of Type 2 Diabetes : the EPIC-Potsdam Study. *Diabetes Care*. 2009; 32:2116-2119.
- 25 Mühlenbruch K, Jeppesen C, Joost H-G, Boeing H, Schulze MB. The value of genetic information for diabetes risk prediction - differences according to sex, age, family history and obesity. *PloS one*. 2013; 8(5):e64307.
- 26 Havulinna AS, Kettunen J, Ukkola O, Osmond C, Eriksson JG, Kesäniemi YA, Jula A, Peltonen L, Kontula K, Salomaa V, Newton-Cheh C. A blood pressure genetic risk score is a significant predictor of incident cardiovascular events in 32,669 individuals. *Hypertension*. 2013; 61:987–94.
- 27 Park J-H, Wacholder S, Gail MH, Peters U, Jacobs KB, Chanock SJ, Chatterjee N. Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics*. 2010; 42:570–5.
- 28 Wang Y, O'Connell JR, McArdle PF, Wade JB, Dorff SE, Shah SJ, *et al*. Whole-genome association study identifies STK39 as a hypertension susceptibility gene. *Proceedings of the National Academy of Sciences of the United States of America*. 2009; 106:226–31.

- 29 Xu J, Ji L-D, Zhang L-N, Dong C-Z, Fei L-J, Hua S, Tsai J-Y, Zhang Y-P.
Lack of association between STK39 and hypertension in the chinese
population. *Journal of Human Hypertension* 2013; 27:294–297.

Tables

Table 1. Characteristics of the Vitória population.

N° individuals		1504
Anthropologic characteristics		
Age (years)		44.80±10.84
Gender (%)	Male	45.58
	Female	54.42
Ethnicity (%)	White	37.4
	others	62.6
Body mass index (kg/m ²)		26.34±4.88
Smoking Status (%)	Nonsmoker	77.86
	Smoker	22.14
Blood pressure		
SBP (mmHg)		127.97±21.94
DBP (mmHg)		84.32±14.10
Pulse Pressure (mmHg)		43.66±13.70
Hypertension (%)		42.4
Dysglycemia (%)		21.4

Table 2. SNPs associated with SBP, DBP, PP or hypertension adjusted for age, gender and ethnicity. *Alleles associated with BP or hypertension risk decrease.

SBP							
SNP	Reported SNP-risk allele	Observed SNP-risk allele	Mapped gene(s)	Beta	SE	p-value	
rs11014166	A	T	CACNB2	-1.642	0.817	0.045	
rs12046278	T*	T	CASZ1	-2.970	0.778	0.000	
rs13107325	T*	T	SLC39A8	-3.544	1.685	0.036	
rs1799945	G	G	HFE	-2.830	1.121	0.012	
rs198833	A*	G	HIST1H1T	-2.521	1.010	0.013	
rs805303	C	T	BAT3	1.552	0.754	0.040	
DBP							
SNP	Reported SNP-risk allele	Observed SNP-risk allele	Mapped gene(s)	Beta	SE	p-value	
rs12046278	T*	T	CASZ1	-1.305	0.523	0.013	
rs1425609	A	G	TOMM22P6 - RPS6P4	-1.206	0.541	0.026	
rs1550576	T	T	GRINL1A - ALDH1A2	1.415	0.712	0.047	
rs2683696	G	G	ULK4	1.164	0.542	0.032	
PP							
SNP	Reported SNP-risk allele	Observed SNP-risk allele	Mapped gene(s)	Beta	SE	p-value	
rs12046278	T*	T	CASZ1	-1.663	0.499	0.001	
rs1799945	G	G	HFE	-2.147	0.720	0.003	
rs198833	A*	G	HIST1H1T	-1.893	0.645	0.003	
HTN							
SNP	Reported SNP-risk allele	Observed SNP-risk allele	Mapped gene(s)	Exp(Beta)	Exp(Beta) 95% CI lower	Exp(Beta) 95% CI upper	p-value
rs1799945	G	G	HFE	0.779	0.608	0.999	0.049
rs198833	A*	G	HIST1H1T	0.796	0.637	0.994	0.044
rs2683696	G	G	ULK4	1.302	1.094	1.550	0.003
rs805303	C	T	BAT3	1.249	1.059	1.473	0.008
rs871606	A	G	LNK1 - RPL21P44	0.759	0.606	0.950	0.016

Table 3. SNPs with significant interaction effects with age associated with SBP, DBP, PP or hypertension adjusted for age.

SBP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs12046278	CASZ1	-0.247	0.072	0.010	0.001
rs319689	MAP4	-0.173	0.070	0.025	0.014
DBP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs35650857	UMOD	0.126	0.062	0.085	0.042
PP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs12046278	CASZ1	-0.174	0.046	0.004	0.000
rs17030613	CAPZA1	0.110	0.054	0.052	0.041
rs319689	MAP4	-0.102	0.045	0.045	0.024
rs9889262	ZNF652	0.112	0.053	0.056	0.035
HTN					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs12046278	CASZ1	-0.024	0.008	0.015	0.005
rs12508308	GUCY1A3	-0.018	0.009	0.029	0.033
rs16982520	MRPS16P - ZNF831	-0.032	0.012	0.005	0.009

Table 4. SNPs with significant interaction effects with BMI associated with SBP, DBP, PP or hypertension adjusted for BMI.

SBP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs1458038	PRDM8 - FGF5	-0.412	0.190	0.036	0.030
rs16982520	MRPS16P - ZNF831	0.603	0.254	0.022	0.018
rs17367504	MTHFR	0.467	0.221	0.012	0.035
DBP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs12046278	CASZ1	0.206	0.100	0.022	0.040
rs13082711	UBA52P4 - RPS27P11	-0.284	0.136	0.029	0.036
rs1378942	CSK	-0.247	0.091	0.012	0.006
rs1458038	PRDM8 - FGF5	-0.275	0.121	0.023	0.023
rs16982520	MRPS16P - ZNF831	0.482	0.164	0.005	0.003
rs1887320	FAT1P1 - RPS11P1	-0.197	0.100	0.082	0.050
rs6495122	CPLX3 - ULK3	-0.227	0.098	0.022	0.021
PP					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs12946454	ACBD4 - PLCD3	-0.274	0.120	0.033	0.022
rs1425609	TOMM22P6 - RPS6P4	-0.279	0.110	0.014	0.012
rs17367504	MTHFR	0.472	0.145	0.001	0.001
HTN					
SNP	Mapped gene(s)	Beta interaction	SE	p-value SNP	p-value
rs11191548	NT5C2	0.060	0.029	0.064	0.039
rs12046278	CASZ1	0.045	0.018	0.010	0.014
rs13082711	UBA52P4 - RPS27P11	-0.068	0.026	0.009	0.009
rs1378942	CSK	-0.039	0.017	0.055	0.019
rs17367504	MTHFR	0.064	0.031	0.017	0.036
rs2071518	NOV	-0.037	0.018	0.024	0.035
rs381815	PLEKHA7	0.049	0.023	0.032	0.031
rs9889262	ZNF652	-0.048	0.023	0.042	0.037

Table 5. Comparison between the discriminatory ability of each model determined by the area under receiver operating characteristic curves (ROC-AUCs).

ROC-AUC					
	Area (95% CI)	SE	Asymptotic Sig.	delta ROC-AUC (95% CI)	Delong test p-value
Model 1	0.553 (0.521-0.586)	0.017	0.001	-	-
Model 2	0.766 (0.740-0.792)	0.013	1.086E-57	0.004(-0.013-0.021)	0.648
Model 3	0.785 (0.760-0.810)	0.013	3.970E-66	0.023(0.010-0.036)	0.0004
Model 4	0.762 (0.735-0.788)	0.013	4.902E-56	Ref.	Ref.

Table 6. Integrated discrimination improvement of models 2 and 3 compared with model 4.

IDI		NRI (%)		
	Absolute IDI (95% CI)	SE	p-value	
Model 1	-	-	-	-
Model 2	0.008 (-0.006-0.022)	0.007	0.262	-1,45
Model 3	0.036 (0.025-0.047)	0.006	2.844E-10	5.21
Model 4	ref.	ref	ref	ref

Table 7. Comparison of risk strata score from Model 3 and 4 with the prevalence of hypertension at baseline.

Risk strata score Model 3									
				Low Risk		Intermediate Risk		High Risk	
				HTN		HTN		HTN	
				No	Yes	No	Yes	No	Yes
Risk strata score Model 4	Low Risk	HTN	No	432	0	14	0	27	0
			Yes	0	113	0	7	0	30
	Intermediate Risk	HTN	No	18	0	3	0	23	0
			Yes	0	11	0	3	0	30
	High Risk	HTN	No	20	0	7	0	157	0
			Yes	0	11	0	2	0	336

Figures

Figure 1. Comparison between the discriminatory ability of each model determined by the area under receiver operating characteristic curves (ROC-AUCs).

Figure 2. Risk stratification of hypertension based on Model 3 and 4.

SUPPLEMENTARY ONLINE CONTENT

THE ROLE OF BLOOD PRESSURE SNPs ON THE RISK PREDICTION OF HYPERTENSION

Table S1. SNPs previously associated with BP phenotype that were tested in Brazilian population. Linked disequilibrium SNPs (LD SNPs - column 5) replaced the reported SNP (column 1) due to the inability to design primers to Taqman assay.

*Alleles associated with BP or hypertension risk decrease.

SNP	Alleles	SNP- risk allele	Mapped gene(s)	LD SNPs	Alleles	LD SNP- risk allele
rs1004467	[A/G]	A	CYP17A1			
rs11014166	[A/T]	A	CACNB2			
rs11066280	[A/T]	T	C12orf51			
rs1110183	[A/G]	G	IGFBPL1 - C9orf51			
rs11191548	[T/C]	T	NT5C2			
rs11222084	[A/T]	T	DDX18P5 - ADAMTS8	rs747249	[A/G]	A
rs1173766	[C/T]	C	C5orf23 - RPS8P8	rs1173747	[A/C]	A
rs11775334	[A/G]	A	MSRA			
rs11953630	[C/T]	T*	RPLP2P2 - EBF1			
rs12046278	[C/T]	T*	CASZ1			
rs12940887	[C/T]	T	ZNF652	rs9889262	[A/T]	A
rs12946454	[A/T]	T	ACBD4 - PLCD3			
rs13002573	[A/G]	G*	FIGN - GRB14			
rs13082711	[C/T]	T*	UBA52P4 - RPS27P11			
rs13107325	[C/T]	T*	SLC39A8			
rs13139571	[C/A]	C	GUCY1A3	rs12508308	[G/T]	G
rs1327235	[G/A]	G	FAT1P1 - RPS11P1	rs1887320	[A/G]	A
rs13333226	[A/G]	A	UMOD	rs35650857	[C/T]	C
rs1378942	[G/T]	C	CSK			
rs1425609	[A/G]	A	TOMM22P6 - RPS6P4			
rs1446468	[C/T]	T*	FIGN - GRB14	rs2675149	[A/C]	A
rs1458038	[A/G]	T	PRDM8 - FGF5			
rs1530440	[C/T]	T	C10orf107			
rs1550576	[C/T]	T	GRINL1A - ALDH1A2			
rs16948048	[A/G]	G	ZNF652			
rs16982520	[A/G]	A*	MRPS16P - ZNF831			
rs16998073	[A/T]	T	PRDM8 - FGF5			
rs17030613	[A/C]	C	CAPZA1			

SNP	Alleles	SNP- risk allele	Mapped gene(s)	LD SNPs	Alleles	LD SNP- risk allele
rs17249754	[A/G]	G	ATP2B1 - MRPL2P1			
rs17367504	[A/G]	G*	MTHFR			
rs1799945	[C/G]	G	HFE			
rs1918974	[C/T]	T*	MECOM			
rs198846	[G/A]	G*	HIST1H1T	rs198833	[A/G]	A
rs2071518	[C/T]	T	NOV			
rs2384550	[G/A]	G	TBX3 - UBA52P7	rs1991391	[C/T]	C
rs2398162	[A/G]	A	RNU2-3P - NR2F2	rs72751246	[A/T]	A
rs2521501	[A/T]	T	FES	rs1894400	[C/T]	T
rs2681472	[A/G]	A	POC1B - ATP2B1	rs4842666	[C/T]	T
rs2820037	[A/T]	T	RPL39P10 - CHRM3			
rs2932538	[A/G]	G	MOV10			
rs3184504	[C/T]	T	SH2B3			
rs319690	[T/C]	T	MAP4	rs319689	[C/T]	C
rs35444	[C/T]	A	TBX3 - UBA52P7			
rs3774372	[C/T]	T	ULK4			
rs381815	[C/T]	T	PLEKHA7			
rs419076	[C/T]	T	MECOM			
rs4373814	[C/G]	G*	SLC39A12 - CACNB2			
rs4590817	[C/G]	G	C10orf107			
rs4936894	[A/G]	A	VWA5A			
rs633185	[G/C]	G*	ARHGAP42	rs2055450	[A/T]	T
rs6495122	[A/C]	A	CPLX3 - ULK3			
rs6825911	[C/T]	C	ZBED1P - ENPEP			
rs7129220	[G/A]	G*	ADM - AMPD3	rs1450276	[C/T]	C
rs805303	[C/T]	G	BAT3			
rs871606	[A/G]	T	LNK1 - RPL21P44			
rs9663362	[C/G]	G*	PLCE1			
rs9815354	[G/A]	A	ULK4	rs2683696	[A/G]	G
rs991316	[A/G]	T	ADH1C - ADH7			

Table S2. Final Model 1

	Beta	SE	p-value	Exp(Beta)	Exp(Beta) 95% CI lower	Exp(Beta) 95% CI upper
rs805303	0.203	0.076	0.007	1.226	1.057	1.421
rs2683696	0.241	0.078	0.002	1.273	1.092	1.484
Constant	-0.671	0.102	5.8E-11	0.512		

Table S3. Final Model 2.

	Beta	SE	p-value	Exp(Beta)	Exp(Beta) 95% CI lower	Exp(Beta) 95% CI upper
rs805303	0.169	0.089	0.058	1.184	0.994	1.411
rs2683696	0.316	0.094	0.001	1.371	1.140	1.650
rs11014166	-0.171	0.096	0.074	0.842	0.698	1.017
rs12046278	-0.411	0.680	0.545	0.663	0.175	2.512
Age by rs12046278	-0.022	0.009	0.021	0.979	0.961	0.997
rs319689	0.811	0.426	0.057	2.251	0.977	5.189
Age by rs319689	-0.018	0.009	0.052	0.983	0.965	1.000
rs16982520	1.522	0.653	0.020	4.580	1.273	16.483
Age by rs16982520	-0.030	0.014	0.034	0.971	0.944	0.998
rs17367504	-0.321	0.142	0.023	0.725	0.549	0.957
BMI by rs12046278	0.047	0.020	0.020	1.048	1.008	1.090
rs13082711	1.796	0.762	0.018	6.025	1.352	26.848
BMI by rs13082711	-0.070	0.028	0.014	0.932	0.882	0.986
rs6495122	1.003	0.528	0.058	2.725	0.968	7.669
BMI by rs6495122	-0.041	0.020	0.037	0.960	0.924	0.998
BMI by rs381815	0.056	0.025	0.025	1.058	1.007	1.111
rs381815	-1.414	0.665	0.033	0.243	0.066	0.895
Age	0.115	0.015	0.000	1.122	1.090	1.155
BMI	0.207	0.059	0.000	1.230	1.095	1.381
Constant	-10.930	1.767	0.000	0.000		

Table S4. Final Model 3.

	Beta	SE	p-value	Exp(Beta)	Exp(Beta) 95% CI lower	Exp(Beta) 95% CI upper
Age	0.129	0.018	0.000	1.126	1.088	1.165
Gender(1)	-2.022	0.626	0.001	7.552	2.215	25.744
BMI	0.209	0.097	0.032	1.233	1.018	1.492
Ethnicity (1)	2.346	0.715	0.001	0.096	0.024	0.389
Age by Gender (1)	0.029	0.013	0.027	0.971	0.946	0.997
Ethnicity (1) by Age	-0.040	0.015	0.007	1.041	1.011	1.071
rs12046278	-0.502	0.729	0.492	0.606	0.145	2.530
rs2683696	0.293	0.100	0.003	1.340	1.101	1.631
Age by rs12046278	-0.023	0.010	0.019	0.977	0.958	0.996
rs9889262	1.025	0.693	0.139	2.788	0.717	10.842
rs12508308	0.897	0.483	0.064	2.451	0.951	6.321
Age by rs12508308	-0.017	0.010	0.090	0.983	0.964	1.003
Age by rs16982520	-0.036	0.015	0.017	0.965	0.937	0.994
rs16982520	1.660	0.701	0.018	5.261	1.331	20.805
rs17367504	-0.273	0.148	0.065	0.761	0.569	1.017
BMI by rs12046278	0.052	0.022	0.016	1.054	1.010	1.099
rs13082711	2.048	0.841	0.015	7.755	1.493	40.279
BMI by rs13082711	-0.081	0.032	0.010	0.922	0.867	0.981
BMI by rs6495122	-0.037	0.020	0.070	0.964	0.926	1.003
rs6495122	0.934	0.554	0.092	2.544	0.860	7.528
rs11191548	-1.365	0.925	0.140	0.255	0.042	1.563
BMI by rs11191548	0.060	0.035	0.081	1.062	0.993	1.137
BMI by rs9889262	-0.044	0.026	0.090	0.957	0.910	1.007
Constant	-11.366	2.764	0.000	0.000		

Table S5. Final Model 4.

	B	SE	p-value	Exp(Beta)	Exp(Beta) 95% CI lower	Exp(Beta) 95% CI upper
Age	0.079	0.010	0.000	1.074	1.054	1.095
Gender(1)	-1.590	0.566	0.005	4.905	1.617	14.877
BMI	0.137	0.014	0.000	1.147	1.116	1.179
Ethnicity (1)	1.925	0.647	0.003	0.146	0.041	0.519
Age by Gender (1)	0.021	0.012	0.083	0.979	0.957	1.003
Ethnicity (1) by Age	-0.028	0.013	0.034	1.029	1.002	1.056
Constant	-7.592	0.593	0.000	0.001		

Table S6. Comparison between the discriminatory ability of Model 3 and 4 determined by the area under receiver operating characteristic curves (ROC-AUCs) in incident cases identified during follow-up.

ROC-AUC					
	Area (95% CI)	SE	Asymptotic Sig.	delta ROC-AUC (95% CI)	Delong test p-value
Model 3	0.659 (0.588-0.729)	0.036	6.659E-5	0.030(-0.002-0.062)	0.064
Model 4	0.688 (0.618-0.757)	0.035	2.328E-6	Ref.	Ref.

Table S7. Cutoffs obtained from ROC-AUCs coordinates (sensitivity and 1-specificity) used in risk stratification based on Models 3 and 4.

	Baseline				Follow-up			
	Sensitivity	Specificity	Sensitivity + Specificity	Probability of HTN	Sensitivity	Specificity	Sensitivity + Specificity	Probability of HTN
Model 3	0,7281	0,7046	1,4327	0,4263	0,5571	0,6912	1,2484	0,4058
Model 4	0,7207	0,6729	1,3936	0,4302	0,5286	0,7788	1,3074	0,4776

Table S8. Comparison of risk strata score from Model 3 and 4 with the prevalence of hypertension in the follow-up.

Risk strata score Model 3									
			Low Risk		Intermediate Risk		High Risk		
			HTN		HTN		HTN		
			No	Yes	No	Yes	No	Yes	
Risk strata score Model 4	Low Risk	HTN	No	136	0	2	0	11	0
			Yes	0	65	0	3	0	11
	Intermediate Risk	HTN	No	9	0	1	0	10	0
			Yes	0	4	0	1	0	14
	High Risk	HTN	No	5	0	2	0	41	0
			Yes	0	5	0	2	0	154

Figure S1. Schematic representation of variables inputted in each developed model.

Figure S2. Comparison between the discriminatory ability of Model 3 and 4 determined by the area under receiver operating characteristic curves (ROC-AUCs) in incident cases identified during follow-up.