Instituto de Biociências
Universidade de São Paulo

Variabilidade fenotípica de um modelo murino para a
Síndrome de Marfan – Triagem de genes modificadores
do fenótipo.

Gustavo Ribeiro Fernandes

2012

Gustavo Ribeiro Fernandes

Variabilidade fenotípica de um modelo murino para a
Síndrome de Marfan – Triagem de genes modificadores do
fenótipo.

*Phenotypic variability in a mouse model for Marfan Syndrome
– Identification of phenotype modifier genes.*

Tese apresentada ao Instituto de
Biociências da Universidade de São
Paulo, para obtenção do Título de
Doutor em Ciências, na Área de
Biologia/Genética.

Orientadora: Dra. Lygia da Veiga
Pereira Carramaschi.

São Paulo
2012

Dados Catalográficos

Fernandes, Gustavo Ribeiro

Variabilidade fenotípica de um modelo murino para a Síndrome de Marfan – Triagem de genes modificadores do fenótipo.

151 páginas

Tese de Doutorado – Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva. Área de concentração: Biologia/Genética

1.Síndrome de Marfan 2.Genes modificadores 3.Mapeamento

Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva

Comissão Julgadora

Prof(a.) Dr(a).
Instituição

Prof(a.) Dr(a).
Instituição

Prof(a.) Dr(a).
Instituição

Prof(a.) Dr(a).
Instituição

Profa. Dra. Lygia da Veiga Pereira Carramaschi
Orientadora

**Aos meus pais, Maria José e Adelino,
por todo apoio e dedicação.**

Agracedimentos

À Dra. Lygia, por me ensinar tanto, por permitir que eu expandisse as áreas de interesse desse projeto para englobar os métodos matemáticos e computacionais envolvidos e ainda por confiar em mim para realizar todas as análises. Muito obrigado pela paciência, compreensão e confiança.

À Dra. Silva Massironi, por me ensinar a lidar com os animais e permitir que eu acompanhasse todas as etapas envolvidas na rotina do biotério e de seu laboratório, sem os quais este projeto não teria se iniciado, pelo apoio e amizade ao longo desses anos.

À Dra. Regina Mingroni e demais professores do departamento de genética de biologia evolutiva, por todo conhecimento e experiências compartilhadas.

Aos meus pais, por estarem sempre ao meu lado para conversar, apoiar, consolar e repreender quando preciso, sem os quais eu não seria quem sou hoje e não estaria onde estou. Os meus primeiros professores e a quem devo tanto. Também agradeço a minha irmã Eloísa,

À minha namorada Scheilla, que tanto me ajudou nessa reta final, por estar sempre ao meu lado e cuidar de mim mesmo nos momentos mais complicados, por ser tão compreensiva e nunca ter desistido de mim. Obrigado por tudo, nossos momentos juntos são inesquecíveis.

À minha prima Renata, que me ensinou e permitiu que eu usasse a sua versão do software AutoCAD para realizar e desenvolver o primeiro método de fenotipagem.

Ao Sensei Jorge Kishikawa, pelos ensinamento que não me deixaram esmorecer durante os momentos mais difíceis. A todos os senpais e colegas do Instituto niten.

À todos os colegas de laboratório, Alyne, Ana Claudia, Ana Maria, Ana Patrícia, Christian, Clarisse, Cláudia, Cynthia, Érica, Enrico, Fábio, Fernando, Giovana, Joana, Juliana, Karla, Lucas, Luciana, Mariana, Naja, Raquel, Roberto, Simone e Thiago. Em especial ao Bruno, que me recebeu como aluno de iniciação durante o seu doutorado e que compartilhou todos esses anos de trabalho, obrigado pelas discussões e conversas ao longo do caminho.

À Lilian e a todos os outros colegas de departamento, que mesmo não estando ligados diretamente ao projeto, ajudaram a entender e discutir os resultados encontrados.

À Deus, pelos dons concedidos e por colocar cada uma dessas pessoas na minha vida.

À FAPESP, pelo apoio financeiro concedido.

“São 4 os fatores determinandes para se vencer uma guerra:

Ki- A energia envolvida no combate, o quanto se esta determinado a entrar em combate e vencer a qualquer custo;

Ken- A técnica, o conhecimento envolvido no combate;

Tai- A postura, o condicionamento físico para aguentar as intempéries do combate;

Un- A sorte, para quando os outros não forem o suficiente.”

Sensei Jorge Kishikawa

Sumário

Resumo	14
Abstract	17
1. Introdução.....	19
1.1. Síndrome de Marfan	19
1.2. Fibrilina-1	22
1.3. Fibrilina-1 e TGF- β	24
1.4. Mutações e correlação fenótipo-genótipo	25
1.5. Modelos Animais	27
1.6. Genes Modificadores.....	32
1.7. Mapeamento genético.....	37
1.8. Técnicas matemáticas de mapeamento	39
1.9. Estratégias de Mapeamento.....	41
2. Objetivos.....	44
2.1. Objetivo Geral	44
2.2. Objetivos Específicos.....	44
3. Material e métodos.....	45
3.1. Extração de DNA.....	45
3.2. Genotipagem dos animais.....	45
3.3. Caracterização fenotípica.....	46
3.3.1. Análise radiológica	46
3.3.2. Análises histopatológicas	46
3.4. Avaliação da variabilidade e estratégia de cruzamentos	47
3.5. Mapeamento com microssatélites	48
3.6. Mapeamento com SNPs	48
3.7. <i>Microarray</i> de expressão	49
3.8. Real-time PCR.....	49
3.9. Análise estatística e Bioinformática	51
4. Resultados	53
4.1. Caracterizando a variabilidade clínica em animais híbridos.....	53
4.2. Métodos quantitativos de fenotipagem.....	55
4.2.1. Sistema esquelético.....	55
4.2.2. Sistema cardiovascular.....	59
4.2.3. Sistema pulmonar	61
4.3. Concordância fenotípica.....	63
4.4. Qualidade dos dados.....	66
4.5. Análise de ligação utilizando microssatélites	68
4.6. Análise de ligação utilizando SNPs	72
4.7. Efeito dos genes modificadores	78
4.8. Variabilidade explicada	83
4.9. Regiões candidatas.....	86
4.10. <i>Microarray</i> de Expressão	89
4.11. Validação dos resultados <i>microarray</i> expressão	102
5. Discussão	107
5.1. Genes modificadores e o modelo $mg\Delta^{loxPneo}$	109
5.2. Predição fenotípica	111
5.3. <i>Microarray</i> de Expressão	113
6. Conclusão	115

7. Bibliografia	119
8. Anexo 1 - Lista de microssatélites utilizados.....	126
9. Anexo 2 - Procedimento de mapeamento com microssatélites.....	127
10. Anexo 3 - Procedimento de identificação dos SNPs informativos e preparo do banco de dados.....	130
11. Anexo 4 - Procedimento de mapeamento com de SNPs.....	132
12. Anexo 5 - Procedimento de análise de Microarray de expressão e GSEA.....	136
13. Anexo 6 - Genes na região candidata do cromossomo 3	145
14. Anexo 7 - Genes na região candidata do cromossomo 4	146
15. Anexo 8 - Genes na região candidata do cromossomo 6	147
16. Anexo 9 - Genes na região candidata do cromossomo 13.....	148
17. Anexo 10 - Genes na região candidata do cromossomo X.....	149
18. Anexo 11 - Vias identificadas entre animais KO e selvagens	151

Lista de Figuras

Figura 1: Manifestações ósseas da SMF. (fontes: http://diseases-blog.blogspot.com.br/2012/03/marfan-syndrome.html ; http://www.primehealthchannel.com/marfan-syndrome-pictures-symptoms-treatment-and-life-expectancy.html ; http://www.hughston.com/hha/a_12_2_4.htm).....	19
Figura 2: Manifestações oculares da SMF. (fonte: http://www.medstudents.com.br/original/revisao/marfan/marfan2.jpg)	20
Figura 3: Manifestações cardiovasculares da SMF. (Fonte: http://www.health-reply.com/thoracic-aortic-dissection_9340/ e http://www.yalemedicalgroup.org/stw/Page.asp?PageID=STW022880)	20
Figura 4: Estrutura primária da fibrilina-1. 4 módulos EGF; 43 módulos CbEGF; 7 módulos TB (LTBP- <i>latent transforming growth factor-beta binding protein</i>); 3 módulos Fib (híbridos EGF e TB). (adaptado de Pereira e cols., 1993)	22
Figura 5: Formação da microfibrila e sua interação com LTBP para o ancoramento e inativação de TGF- β (adaptado de Byers, 2004).	25
Figura 6: Esquema de recombinação homologa para a criação do alelo $mg\Delta^{loxPneo}$. Retângulos azuis indicam a posição das sondas utilizadas para verificar o evento de recombinação. “h” e “b” indicam os sítios de restrição das enzimas <i>Hind III</i> e <i>BamHI</i> . Triângulos vermelhos são as sequências LoxP. (Adaptado de Lima, 2010)	29
Figura 7: Heredograma da família $mg\Delta^{loxPneo}$. Os heterozigotos com fenótipo grave estão indicados em vermelho. No detalhe, raio-X de um heterozigoto grave evidenciando manifestações ósseas. (Adaptado de Santos, 2005).....	30
Figura 8: Manifestações fenotípicas do modelo $mg\Delta^{loxPneo}$ observadas entre animais heterozigotos das linhagens 129/Sv (3 meses de idade) e C57BL/6 (6 meses de idade). Seções histológicas de pulmão e aorta coradas com HE, e radiografia da coluna vertebral	31
Figura 9: Possíveis efeitos de genes modificadores. Linhas contínuas representam a manifestação fenotípica na ausência de gene modificador, enquanto linhas pontilhadas representam o fenótipo na presença de gene modificador. A) Penetrância; B) Dominância; C) Expressividade e D) Pleiotropia.	34
Figura 10: Diferença entre fenótipos com característica poligênica (A) e influenciado por gene modificador (B). (Adaptado de Broman, 2009).	35
Figura 11: Frequência relativa dos QTLs já identificados de acordo com o poder explicativo de cada um. (Adaptado de Griffiths, 2006).	36
Figura 12: Cruzamentos controlados para realização de estudos de mapeamento. A) Retrocruzamento, B) Intercruzamento.	38
Figura 13: Equações derivadas por Harley e Knot em 1992 para determinar os efeitos aditivos e de dominância que um marcador teria dado o genótipo e a fração de recombinação entre os marcadores que o flanqueiam.	40
Figura 14: Genotipagem seletiva em um fenótipo com distribuição normal.....	41
Figura 15: Influência da quantidade de animais selecionados e da densidade de marcadores sobre a informação obtida do mapeamento.	42
Figura 16: Heredograma dos cruzamentos entre as linhagens 129/Sv e C57BL/6. Animais heterozigotos visualmente afetados e assintomáticos estão indicados de acordo com a legenda.....	53
Figura 17: Comparação do fenótipo dos sistemas ósseo, pulmonar e cardiovascular de animais F1 com três meses de idade. A - animal selvagem; B e C - animais heterozigotos.....	54

Figura 18: Comparação do fenótipo dos sistemas ósseo, pulmonar e cardiovascular de animais F2 com três meses de idade. A - animal selvagem; B, C e D - animais heterozigotos diferentemente afetados em cada um dos três sistemas.....	55
Figura 19: Quantificação do desvio da coluna com programa AutoCAD a partir de radiografias dos animais. Foram mensuradas a distancia linear e o comprimento da coluna de cada região. As cores vermelha, amarela e verde representam as regiões cervicais, torácicas e lombar, respectivamente	56
Figura 20: Comparação entre as distribuições obtidas da avaliação de cada região da coluna cervical dos animais F2 knock-outs. Distribuição entorno da média de cada conjunto de dados (observação/média obsevações).	57
Figura 21: Gráfico com o valores absolutos da razão (distância linear/coluna) para os animais selvagens e knock-outs das linhagens parentais e de cada geração híbrida aos 3 meses de idade. Dados	58
Figura 22: Método de quantificação das alterações cardio-vasculares. Estimação do raio externo, raio interno e espessura do vaso através da mensuração da área interna a cada linha vermelha.....	59
Figura 23: Comparação entre as distribuições obtidas da avaliação de cada uma das 5 medições realizadas nas aortas dos animais F2 knock-outs. Distribuição ao redor da média de cada conjunto de dados (observação/média observações).	60
Figura 24: Gráfico com a espessura da aorta (AWT) para os animais selvagens e knock-outs das linhagens parentais, e de cada geração dos animais híbridos aos 3 meses de idade..	61
Figura 25: Método de quantificação do fenótipo pulmonar.	62
Figura 26: Distribuição das alterações fenotípicas pulmonares	62
Figura 27: Gráfico com o as medidas de intercepto linear médio (Lm) para os animais selvagens e knock-outs das linhagens parentais, e de cada geração dos animais híbridos aos 3 meses de idade. Dados	63
Figura 28: Gráficos de correlação entre os fenótipos de animais KO das linhagens parentais 129/Sv (pontos vermelhos) e C57BL/6 (pontos pretos). A) Lm x KR; B) AWT x KR; C) AWT x Lm; D) Valores do coeficiente de correlação de Pearson (triangulo inferior) e níveis de significância (triangulo superior).....	64
Figura 29: Gráficos de correlação entre os fenótipos de animais KO dos animais F2 híbridos 129xB6. A) Lm x KR; B) AWT x KR; C) AWT x Lm; D) Valores do coeficiente de correlação de Pearson (triangulo inferior) e níveis de significância (triangulo superior).	64
Figura 30: Resumo dos dados usados para o mapeamento dos genes modificadores do fenótipo. A)mapa genético, B, C e D) distribuição dos fenótipos estudados, E) Sexo (0 para fêmea, 1 para macho) dos animais da amostra.	67
Figura 31: QQ plots para os traços estudados. A)Ósseo (KR). B) Cardíaco (AWT). C) Pulmonar (Lm). D) níveis de significância para o teste de normalidade de Shapiro.....	68
Figura 32: Nivel de recombinação estimado entre cada marcador (triangulo superior) e o nível de significância (triangulo inferior).	69
Figura 33:Perfil de Lod Score para o mapeamento com microssatélites do fenótipo Ósseo. .	70
Figura 34: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo cardiovascular.	71
Figura 35: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo pulmonar.	71
Figura 36: Mapa genético dos SNPs polimórficos entre 129/Sv e C57BL/6.....	73
Figura 37: Nivel de recombinação estimado entre cada marcador (triangulo superior) e o Lod Score observado para cada par (triangulo inferior).	74
Figura 38: Perfil de Lod Score para o mapeamento com SNPs do fenótipo Ósseo.	75

Figura 39: Perfil de Lod Score para o mapeamento com SNP do fenótipo cardiovascular..	76
Figura 40: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo pulmonar.	76
Figura 41: Efeito de cada classe genotípica sobre o fenótipo ósseo no marcador mais próximo a posição de cada QTLs identificados. “UNC030244865”, “UNC060308920” e “JAX00177060” são os marcadores mais próximos dos QTLs dos cromossomo 3, 6 e X, respectivamente.	79
Figura 42: Gráficos de interação par a par de cada QTL para o sistema ósseo. Os genótipos do primeiro QTL estão indicados na abcissa, o valor do fenótipo na ordenada, e os genótipos do segundo QTL são representados pelas cores azul (B6/B6), vermelha(129/B6) e preta (129/129). “UNC030244865”, “UNC060308920” e “JAX00177060” são os marcadores mais próximos da posição estimada dos QTLs dos cromossomo 3, 6 e X, respectivamente. A) QTLs cromossomos 3 e 6. B) 3 e X. C) 6 e X.	81
Figura 43: Efeito de cada classe genotípica sobre o fenótipo cardiovascular do marcador mais próximo à posição de cada QTL identificado. “UNC040316624” e “backupUNC130158494” são os marcadores mais próximos para o cromossomo 4 e o cromossomo 13, respectivamente.	82
Figura 44: Gráficos de interação par a par de cada QTL, os genótipos do primeiro QTL estão indicados na abcissa, o valor do fenótipo na ordenada, e os genótipos do segundo QTL são representados pelas cores azul (B6/B6), vermelha(129/B6) e preta (129/129). “UNC040316624” e “backupUNC130158494” são os marcadores mais próximos para o cromossomo 4 e o cromossomo 13, respectivamente.	82
Figura 45: Análise de qualidade de cada array realizado. CV% maiores que 10% indicam dados de qualidade duvidosa (pontilhado azul).	90
Figura 46: Dados antes e após normalizações, respectivamente. A e B) Gráficos RLE (logaritmo da expressão relativa) de todas as sondas para cada array (normalização interna de cada array) ;C e D) distribuição da intensidade do sinal (normalização entre arrays); E e F) Gráficos MVA (diferença do log da intensidade X média dos log da intensidade) para cada uma das sondas, exemplo da correção que cada array sofre ao longo do processo.	92
Figura 47: Desvio padrão x média de cada sonda para as três análises realizadas. O aumento/diminuição do desvio em função na média indicaria a necessidade de outros correções antes de se iniciar a análise.	93
Figura 48: Distribuição dos desvio padrões para cada uma das comparações feitas. Aquelas com desvios muito pequenos (a esquerda das linhas pontilhadas) foram excluídos na filtragem não específica.	94
Figura 49: Gráfico Vulcão WT x Leve. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.	95
Figura 50: Gráfico Vulcão WT x Grave. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.	96
Figura 51: Gráfico Vulcão Leve x Grave. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.	96
Figura 52: Heatmap para os genes diferencialmente expressos entre os animais leves e selvagens. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.	98

Figura 53: Heatmap para os genes diferencialmente expressos entre os animais graves e selvagens. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.	99
Figura 54: Heatmap para os genes diferencialmente expressos entre os animais leves e graves. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.	100
Figura 55: Gráficos de padronização dos genes de interesse (<i>Prrx2</i> e <i>Edn1</i>) contra os possíveis endógenos (<i>Rpl13a</i> , <i>Gapdh</i> e <i>Actb</i>). Abscissa $\log_{10}[\text{RNA}]$ e ordenada diferença entre Ct gene alvo e Ct endógeno. Cada gráfico contém a linha de tendência, a equação da reta e o coeficiente de explicação (R^2).	104
Figura 56: Expressão relativa para o gene <i>Edn1</i> em animais 129/Sv. A) Animais agrupados de acordo com a sua classe fenotípica. B) Expressão individual de cada animal.	105
Figura 57: Expressão relativa para o gene <i>Prrx2</i> em animais 129/Sv. A) animais agrupados de acordo com a sua classe fenotípica. B) Expressão individual de cada animal.	105

Lista de tabelas

Tabela 1: Critérios diagnósticos de Ghent.	21
Tabela 2: Coeficiente de herdabilidade (H^2) estimado.....	65
Tabela 3: Análise de Variância para detecção de covariáveis	70
Tabela 4: Regiões candidatas identificadas.....	72
Tabela 5: Regiões candidatas identificadas (SNPs).....	77
Tabela 6: Termos de regressão estimados para o fenótipo ósseo.	84
Tabela 7: Termos de regressão estimados para o fenótipo cardiovascular.....	85

Lista de Abreviações

AWT- *Aortic Wall Thickness*.
C57BL/6- B6.
cM- Centimorgam.
EGF- *Epidermal growth fator*.
eQTL- *expression QTL*.
Het- Heterozigoto.
Kb- Kilobases.
kDa- Kilodaltons.
KO- *Knockout*.
KR- *Kifosis ratio*.
kV- Kilovoltagem.
Lm- *Mean linear intercept*.
LTBP- *Latent TGF- β binding protein*.
 μ l- Microlitro.
 μ m- Micrômetro.
mA- Microamperes
MEC- Matriz extracelular.
neoR- Cassete de resistência à neomicina.
ng- Nanogramas.
nm- Nanomolar.
pb- Pares de base.
PCR- *Polimerase chain reaction*.
QTL- *Quantitative Trait Loci*.
RT-PCR- *Real-Time PCR*.
SMF- Síndrome de Marfan.
SNP- *Single Nucleotide Polimorphism*.
Tm- Temperatura de *melting*.
UV- Ultravioleta.
WT- *Wild-type*.
129- 129/Sv.

Resumo

A Síndrome de Marfan (SMF) (OMIM# 154700) é a mais comum das doenças genéticas do tecido conjuntivo. Herdada de forma autossômica dominante, ela apresenta incidência de 1 em cada 5.000 indivíduos. Apesar de apresentar grande variabilidade clínica inter e intrafamiliar, o fenótipo da SMF possui penetrância completa, e suas manifestações clínicas afetam primariamente os sistemas esquelético, ocular e cardiovascular.

A fim de estudar os mecanismos patogênicos da SMF foi desenvolvido um modelo murino, $mg\Delta^{neoLoxP}$, que reproduz as manifestações ósseas, cardiovasculares e pulmonares da síndrome. O modelo foi estabelecido nas linhagens isogênicas 129/Sv e C57BL/6, que apresentam diferenças tanto quanto a idade de acometimento quanto a gravidade das alterações, é possível que as diferenças alélicas existentes entre essas linhagens alterem a manifestação fenotípica, ou seja, que existam genes modificadores para a SMF.

Assim, objetivo deste projeto é utilizar este modelo experimental para identificar genes modificadores do fenótipo da SMF e tentar entender melhor a arquitetura genética da síndrome.

Ao todo foram gerados 82 animais 129xB6 F2 heterozigotos para o alelo $mg\Delta^{neoLoxP}$, a análise de ligação utilizando microssatélites e SNPs nos animais com fenótipos mais extremos mostraram ligação sugestiva do fenótipo ósseo com as regiões compreendidas entre as posições 56 cM e 68 cM do cromossomo 3; e 2 cM e 20 cM do cromossomo X; ligação significativa entre as posições 41cM e 49 cM do cromossomo 6; além de mostrar ligação sugestiva do fenótipo cardiovascular do 66 cM ao 70 cM do cromossomo 4; e do 44 cM ao 52 cM do cromossomo 13.

Além da variabilidade entre linhagens, os animais 129 apresentam uma grande variabilidade fenotípica interna, o que por se tratar de animais isogênicos causada por fatores aleatórios ou devido a modificações epigenéticas que alterem o nível de expressão de alguns

genes e assim o fenótipo. A comparação entre animais 129 leves e graves levou a identificação de 25 genes diferencialmente expressos dos quais 11 apresentavam funções relevantes para a SMF, entretanto foram aferidos os níveis de expressão de 2 destes que não validaram os resultados obtidos devido a uma grande variação observada entre os animais de todos as classes fenotípicas. Também foram identificadas 46 vias que se apresentavam mais frequentes nos conjuntos de genes obtidos entre as duas classes fenotípica de animais heterozigotos contra os animais selvagens.

Tanto as vias, quanto os genes identificados através de ligação quanto diferença de expressão mostram uma convergência para as funções dos genes de interesse, sendo que entre eles existem genes já associados com a SMF, controle de ativação de TGF-B e da biogênese das microfibras da matriz extracelular, quanto genes que ainda não foram associados mas são possíveis modificadores do fenótipo, tais como genes envolvidos nos processos de enovelamento e degradação protéico e nos processos de endocitose e exocitose de vesículas, que podem alterar a quantidade de fibrilina-1 truncada disponível e assim o fenótipo.

Abstract

The Marfan syndrome (MFS) (OMIM # 154700) is the most common genetic disorder of the connective tissue and is inherited in an autosomal dominant fashion, it has an incidence of 1 in 5,000 individuals. Despite a great clinical variability being one of the “trademarks” of the syndrome, the phenotype of the MFS has complete penetrance and its clinical manifestations primarily affect the skeletal, ocular and cardiovascular systems.

In order to study the pathogenic mechanisms of MFS was developed a mouse model, named $mg\Delta^{LoxPneo}$, which reproduces the skeletal, cardiovascular and pulmonary manifestations of the syndrome. The model was established in inbred mouse strains 129/Sv and C57BL / 6, which phenotypes differ as to age of onset and severity of manifestation. It is possible that allelic differences between these inbred strains alter the phenotypic manifestation of disease, leading to the conclusion that may exist modifier genes involved in the for MFS.

This study intends to use this experimental model to identify phenotype modifier genes of MFS so as to better understand the genetic architecture of the syndrome.

Altogether, 82 129xB6 F2 heterozygous animals were generated so that a linkage analysis using microsatellite and SNP could be conducted. The linkage analysis using a selective genotyping procedure showed a suggestive linkage of the skeletal phenotype with regions included between positions 56 cM and 68 cM on chromosome 3, and 2 cM and 20 cM on chromosome X; and a significant linkage between positions 41cM and 49 cM on chromosome 6; also showing suggestive linkage of the cardiovascular phenotype from 66 cM to 70 cM on chromosome 4, and 44 cM to 52 cM on chromosome 13.

Besides the variability between strains, 129 animals have a wide inner strain phenotypic variability, which in the case of isogenic animals should be caused by random factors or due to epigenetic modifications that may alter the expression level some genes and thus the phenotype. The comparison between animals of the 129 strain with mild and severe

alterations led to the identification of 25 differentially expressed genes of which 11 showed relevant functions to the MFS, however it was only possible to measure the expression levels of two genes using real-time PCR, although those did not validate the results obtained from the expression microarray due to a large expression variation in all phenotypic classes. It was also identified 46 pathways that were more frequent in the gene lists obtained from the comparison between the two phenotypic classes of heterozygous animals against 129 wild-type animals.

There is a similarity in the function of genes or pathways of interest found in pathways analysis and genes identified, either by differential expression or linkage analysis, and among them there are genes already associated with the MFS, such as in the control activity of TGF- β and biogenesis of microfibrils in the extracellular matrix, as also genes that were not associated with MFS but are possible phenotype modifier genes, such as genes involved in protein folding and degradation processes and of endocytosis and exocytosis processes of vesicle, which can change the amount of truncated fibrillin-1 available and thus the phenotype.

1. Introdução

1.1. Síndrome de Marfan

A Síndrome de Marfan (SMF) é a mais comum das doenças genéticas do tecido conjuntivo (OMIM# 154700). Herdada de forma autossômica dominante, essa síndrome tem uma incidência de, aproximadamente, 1-2 para cada 10000 indivíduos (Ades, 2007). O fenótipo da SMF é 100% penetrante e suas manifestações clínicas afetam primariamente os seguintes sistemas: esquelético - crescimento excessivo dos membros, hiperextensibilidade articular, escoliose e deformidades na região anterior do tórax - (fig. 1); ocular - ectopia lentis, miopia e deslocamento da retina - (fig. 2); e cardiovascular - dilatação da raiz da aorta, dissecção da aorta, prolapso da válvula mitral e refluxo mitral - (fig. 3).

Figura 1: Manifestações ósseas da SMF. (fontes: <http://diseases-blog.blogspot.com.br/2012/03/marfan-syndrome.html>; <http://www.primehealthchannel.com/marfan-syndrome-pictures-symptoms-treatment-and-life-expectancy.html>; http://www.hughston.com/hha/a_12_2_4.htm)

Figura 2: Manifestações oculares da SMF. (fonte: <http://www.medstudents.com.br/original/revisao/marfan/marfan2.jpg>)

Figura 3: Manifestações cardiovasculares da SMF. (Fonte: http://www.health-reply.com/thoracic-aortic-dissection_9340/ e <http://www.yalemedicalgroup.org/stw/Page.asp?PageID=STW022880>)

A sobrevida média de pacientes com a SMF é de 40 anos, dois terços do normal, embora venha se aproximando da média da população geral nas últimas décadas (Robinson e cols., 2006). Em 90% dos casos, a morte é causada por falha cardiovascular (McKusick, 1955). O diagnóstico precoce seguido de intervenção médica, tal como a utilização de bloqueadores para redução do stress hemodinâmico, juntamente ou não à realização de cirurgias profiláticas para a colocação de próteses na artéria aorta, possibilitam um aumento significativo na expectativa de vida dos pacientes (Silverman e cols, 1995).

Uma característica marcante da SMF é sua grande variabilidade clínica, observada tanto entre indivíduos afetados de famílias diferentes quanto entre afetados da mesma família. Isso, somado à sobreposição de alterações da SMF e de outras doenças que afetam o tecido conjuntivo, tornam o diagnóstico da SMF algo não trivial (Pyeritz, 2000).

O critério diagnóstico mais aceito para a SMF foi estabelecido em 1996 e ficou conhecido como nosologia de Ghent, que consiste na presença de duas manifestações principais e do envolvimento de um terceiro sistema para casos isolados ou uma manifestação principal e uma secundária para casos familiares ou que tenham uma mutação identificada (tabela 1). Entretanto, estes critérios foram modificados para atribuir um peso maior para alterações como aneurisma de aorta e ectopia lentis, e estabelecer uma pontuação sistêmica (Loeys e col., 2010). Em um trabalho recente o mesmo grupo de pacientes foi submetido a ambos os critérios diagnósticos. Apesar dos resultados terem sido similares, o sistema de pontuação implementado no novo critério facilita o diagnóstico da síndrome, sendo de grande utilidade para médicos pouco acostumados a lidar com SMF (Faivre e col., 2012).

Tabela 1: Critérios diagnósticos de Ghent.

<i>Sistema</i>	<i>Critérios Principais</i>	<i>Critérios Secundários</i>
Ósseo	<p>Ao menos 4 das seguintes alterações:</p> <ul style="list-style-type: none"> • <i>Pectus carinatum</i>; • <i>Pectus excavatum</i> (necessitando de cirurgia); • (Seguimento superior):(seguimento inferior) < 0,86 ou envergadura:altura > 1,05; • Alterações nos punhos e dedos polegares das mãos; • Escoliose > 20° ou espondilolistese; • Movimento de extensão do cotovelo reduzido (< 170°); • <i>Pes plenus</i>; • <i>Protusio acetabulae</i>. 	<p>2 dos critérios principais ou 1 dos critérios principais e 1 dos seguintes:</p> <ul style="list-style-type: none"> • <i>Pectus excavatum</i> (sem cirurgia); • Hiper mobilidade das juntas; • Palato alto com dentes; encavalados; • Rosto característico.
Ocular	<i>Ectopia lentis</i> (Deslocamento do cristalino).	<p>Achatamento da córnea;</p> <p>Miopia;</p> <p>Contração e expansão da pupila reduzidas.</p>
Cardiovascular	<p>Dilatação da base da aorta;</p> <p>Rompimento da aorta ascendente.</p>	<p>Prolápio da válvula mitral;</p> <p>Dilatação da artéria pulmonar antes dos 40 anos;</p> <p>Calcificação do anel mitral (músculo);</p> <p>Outras dilatações ou rompimentos da aorta.</p>
Pulmonar	Nenhum.	<p>Pneumotórax espontâneo;</p> <p>Bolhas apicais.</p>
Epitelial	Nenhum.	<p><i>Sriae atrophicae</i>;</p> <p>Hérnia recorrente.</p>
Achados genéticos	<p>Pai, filho ou irmão apresentarem tais critérios independentemente;</p> <p>Mutação na Fibrilina 1;</p> <p>Herança de marcadores moleculares ligados à Síndrome de Marfan na família.</p>	<p>Nenhum.</p>

Adaptado de Loeys e cols. 2009.

1.2. Fibrilina-1

A SMF é causada por mutações no gene *FBNI*, que codifica a fibrilina-1, a componente estrutural mais abundante das microfibrilas de 10-12 mm na matriz extracelular (Lee e cols., 1991; Dietz e cols., 1991; Sakai e cols., 1986).

A fibrilina-1 é uma glicoproteína que apresentam em sua estrutura diversas repetições de domínios homólogos, característica também encontrada em proteínas como as fibulinas-1 e 2 e a proteína de ligação ao TGF- β (LTBP). Os domínios que compõem a maior parte da fibrilina-1 são conhecidos como *calcium-binding EGF-precursor-like* (cbEGF), por serem semelhantes a módulos presentes no EGF (*epidermal growth factor*). Esses módulos apresentam funções críticas, tais como proteção contra as ações proteolíticas extracelulares, mediação para a interação entre monômeros, favorecimento para o empacotamento lateral de microfibras, e organização de macroagregados (Kielty e cols, 1995) (fig. 4). Cerca de 74% das mutações descritas no gene *FBNI* ocorrem em sequencias codificadoras de módulos cbEGF.

Figura 4: Estrutura primária da fibrilina-1. 4 módulos EGF; 43 módulos CbEGF; 7 módulos TB (LTBP- *latent transforming growth factor-beta binding protein*); 3 módulos Fib (híbridos EGF e TB). (adaptado de Pereira e cols., 1993)

A função das microfibrilas varia entre os tecidos nos quais estão presentes. Em tecidos não elásticos, elas apresentam função de ancoragem, como por exemplo, fazendo a ligação entre o cristalino e o corpo ciliar dos olhos. Em tecidos elásticos, as microfibras formam um arcabouço onde a elastina é depositada, formando assim a fibra elástica, que confere elasticidade aos tecidos. A relação funcional entre a fibrilina e o sistema esquelético pode estar associada à tensão que os tendões e ligamentos exercem para regular o crescimento ósseo (Zhang e cols, 1995).

Durante a sua secreção celular as moléculas de fibrilina-1 são processadas e se polimerizam para formar as microfibrilas que são fixadas à matriz extracelular (Kielty e cols, 2002). Esse processo ainda não é completamente compreendido, e por isso o mecanismo patogênico da SMF permanece desconhecido. Entretanto, o mecanismo mais aceito para a SMF é o dominante-negativo, onde proteínas mutantes e normais são polimerizadas para formar as microfibrilas, o que acaba perturbando a sua organização e integridade. Esse modelo também contempla a hipótese de que indivíduos portadores de um alelo mutante com expressão reduzida podem apresentar um fenótipo mais leve da doença devido à prevalência de fibrilina normal (Pereira e cols., 1997).

Estudos mais recentes indicam que o fenótipo da SMF pode ser gerado pela haploinsuficiência da fibrilina-1 normal, ou seja, a mutação de um dos alelos do gene *FBNI* causa uma redução da expressão de fibrilina-1 normal, e possivelmente o aparecimento do quadro clínico da SMF (Judge e cols, 2004 ; Hutchinson e cols, 2003).

Uma vez que os níveis de expressão da fibrilina-1 normal parecem apresentar o potencial de modular os efeitos patogênicos das proteínas mutantes, é possível que uma variação no nível de expressão dos alelos normais do gene *FBNI* possa explicar parte da

heterogeneidade encontrada em pacientes com SMF, principalmente naqueles indivíduos com mutações *nonsense* que, normalmente, apresentam baixa expressão de proteínas mutantes.

1.3. Fibrilina-1 e TGF- β

As proteínas de ligação ao TGF- β latente (LTBPs, do inglês *Latent TGF- β Binding Proteins*) apresentam um papel importante na sinalização celular por promoverem um sítio de ancoramento para as unidades latentes da citocina TGF- β , antes da sua liberação e ativação, devido a perturbações dos elementos da matriz extracelular (Annes e cols, 2003). As similaridades de estrutura e sequência das proteínas LTBPs e fibrilina-1 sugerem que a fibrilina-1 poderia propiciar um sítio de ligação para as unidades latentes de TGF- β (Nakajima e cols, 1999) (fig. 5).

Três achados importantes sustentam a hipótese de que a fibrilina-1 participa no controle da sinalização das unidades latentes de TGF- β . Primeiro, em 2003, Isogai e cols. mostraram que as moléculas de fibrilina-1 e LTBP interagem fisicamente, assim a fibrilina-1 ajudaria a estabilizar o complexo da LTBP para manter TGF- β na sua forma inativa. Como consequência dessa regulação negativa, foi sugerido que alterações na estrutura da fibrilina-1 ou na sua quantidade aumentariam os níveis de TGF- β ativos e que esses poderiam influenciar nas alterações fenotípicas da SMF. Segundo, o estudo do desenvolvimento pulmonar em camundongos mutantes, com baixa expressão de fibrilina-1 devido a mutações geradas no gene *Fbn1* (Pereira e cols., 1997; Pereira e cols., 1999), mostrou que existe uma maior atividade de TGF- β nos pulmões em desenvolvimento desses animais, que apresentam alvéolos aumentados, e desenvolvem enfisema nos pulmões quando envelhecem (Neptune e cols, 2003). Terceiro, o aumento da atividade e sinalização da proteína TGF- β foi

correlacionado com o prolapso da válvula mitral cardíaca em camundongos com deficiência da proteína fibrilina-1 (Ng e cols, 2004).

Figura 5: Formação da microfibra e sua interação com LTBP para o ancoramento e inativação de TGF- β (adaptado de Byers, 2004).

O papel do TGF- β na patogênese da SMF foi finalmente demonstrado quando mostrou-se que alterações pulmonares e cardíacas, observadas nos modelos animais, foram corrigidas em animais tratados com anticorpos neutralizantes de TGF- β , indicando que a maior atividade desta proteína é responsável pelas características fenotípicas, pulmonares e cardíacas, presentes na SMF (Neptune e cols, 2003; Ng e cols, 2004).

1.4. Mutações e correlação fenótipo-genótipo

O gene *FBNI* se expande ao longo de 230kb situados no cromossomo 15q21.1 e é dividido em 65 exons (Biery e cols, 1999; Pereira e cols, 1993), onde cada éxon codifica um

domínio diferente da fibrilina-1 (Collod-Beroud, 2003). A maioria das mutações que ocorrem nesse gene são mutações pontuais de troca de sentido em um dos 2.871 aminoácidos que compõem a proteína (Robinson PN, 2002), visto que muitas delas prejudicam a ligação de cálcio com ou sem alteração da estrutura secundária da fibrilina-1 (Downing, 1996). Já foram identificadas mais de 500 mutações do gene FBN1, quase todas únicas para cada indivíduo ou família afetados (Collod-Beroud, 2003).

Mutações do tipo *nonsense* resultantes de inserções ou deleções de bases na sequência do gene, e que levam a uma troca no quadro de leitura deste, correspondem a cerca de 25% de todas as mutações já identificadas.

Embora exista a dificuldade de se estabelecer uma correlacionar o genótipo e o fenótipo de cada mutação, devido a variabilidade clínica inerente à SMF (Hutchinson e cols, 2003), existe uma relação entre mutações localizadas entre os éxons 24 e 32 com a forma neonatal da SMF e manifestações mais graves. No entanto, esta relação não é biunívoca (Robinson e cols, 2000). Finalmente, mutações nos últimos sete éxons (59 a 65) estão associadas a uma forma menos grave da SMF.

A variabilidade clínica da SMF não se deve apenas ao tipo de mutação que o paciente carrega. Capotorti em 1959, Hutchinson em 2003 e Buoni em 2004 descreveram famílias distintas onde vários indivíduos afetados pela SMF possuíam diferentes graus de acometimento, a família descrita por Hutchinson apresentava 4 irmãos afetados, sendo que o mais velho apresentava desvios de coluna, dilatação da aorta (43 mm de espessura); o segundo possuía membros desproporcionais em relação ao corpo, dilatação e ruptura da aorta (70 mm); o terceiro apresentavam membros desproporcionais, desvios de coluna, aracnodactilia mas não apresentava nenhuma alteração cardiovascular, enquanto que o irmão mais jovem desenvolveu aracnodactilia, dilatação da aorta (>50 mm) e já havia passado por cirurgia para repor a base da aorta, mesmo os irmãos portando a mesma mutação cada um

desenvolveu o seu próprio conjunto de manifestações. Enquanto que na família relatada por Buoni e cols. mãe e filho portavam a mesma mutação (R2726W) porém apenas o filho apresentava alterações fenotípicas. Assim, deve existir uma influência do *background* genético na modulação da gravidade com a qual a SMF se manifesta, ou seja, a gravidade da SMF depende tanto do tipo de mutação que altere o gene *FBNI* quanto da interação de outros genes ao longo do genoma.

1.5. Modelos Animais

Em 1997 e em 1999, respectivamente, foram publicados dois trabalhos relatando a criação de dois modelos animais para a SMF, os modelos mg Δ e mgR (Pereira e cols., 1997; Pereira e cols., 1999). No modelo mg Δ os exons 19-24 do gene *Fbn1* foram substituídos, através de recombinação homóloga, por um cassete de expressão do gene de resistência a neomicina (neo). Animais heterozigotos para essa alteração não apresentam nenhuma alteração fenotípica, enquanto que os animais homozigotos sobrevivem por apenas duas semanas após o nascimento e morriam devido às alterações cardiovasculares que desenvolviam.

O modelo mgR é resultado de uma recombinação aberrante onde o cassete de neo foi inserido no intron 18, porém sem haver a deleção dos exons 19-24 – logo, o alelo mutante nos animais mgR produzia fibrilina normal. Para este modelo também só era possível identificar alterações fenotípicas em animais homozigotos, os quais possuíam um maior tempo de vida que os animais do modelo anterior, mas também morriam devido a falhas no sistema cardiovascular.

A análise da expressão desses alelos mutantes indicou que ambos apresentavam um nível transcricional muito inferior ao do alelo selvagem: 20% para o alelo mgR e 10% para o

alelo $mg\Delta$. Assim, concluiu-se que: (1) pouca quantidade de fibrilina-1 alterada não é capaz de desorganizar a microfibrila a ponto de causar alteração fenotípica esperadas da SMF, como observado nos animais heterozigotos $mg\Delta$; e (2) baixa quantidade de fibrilina normal leva a fenótipos da SMF, como observado nos animais mgR homozigotos.

A limitação desses modelos é que ambos só apresentam o fenótipo de SMF quando em homozigose, enquanto que a SMF é uma doença dominante, e a baixa expressão do alelo mutante $Fbn1^{mg\Delta}$, devido à presença do cassete *neo*, impede a manifestação do efeito dominante-negativo característico da SMF (Pereira e cols, 1997; Pereira e cols, 1999).

Assim para se obter um modelo para a SMF que realmente possua um alelo $Fbn1$ dominante-negativo foi criada uma variação do modelo $mg\Delta$ onde o cassete *neo* foi flanqueado por duas sequencias LoxP (fig. 6). Isso permitiria a posterior remoção desse cassete pela expressão da proteína CRE, e assim o controle da expressão do alelo mutante, e o estudo das alterações tecido específicas da SMF. Este novo modelo animal foi denominado $mg\Delta^{loxPneo}$ (Lima e cols., 2010) .

Figura 6: Esquema de recombinação homóloga para a criação do alelo $mg\Delta^{loxPneo}$. Retângulos azuis indicam a posição das sondas utilizadas para verificar o evento de recombinação. “h” e “b” indicam os sítios de restrição das enzimas *Hind III* e *BamHI*. Triângulos vermelhos são as sequências LoxP. (Adaptado de Lima, 2010)

Os animais pertencentes à geração F1, gerados do cruzamento da quimera portadora do alelo $Fbn1^{mg\Delta^{loxPneo}}$ com animais CD-1, não apresentavam nenhum fenótipo visível. Contudo, após algumas gerações de *backcross* com animais 129/Sv (129), alguns heterozigotos com manifestações ósseas e cardiovasculares graves foram identificados, sugerindo uma influência do *background* genético na manifestação da doença (figura 7).

Figura 7: Heredograma da família $mg\Delta^{loxPneo}$. Os heterozigotos com fenótipo grave estão indicados em vermelho. No detalhe, raio-X de um heterozigoto grave evidenciando manifestações ósseas. (Adaptado de Santos, 2005)

Outra evidência da importância do fundo genético na gravidade da doença surgiu quando o alelo $Fbn1^{mgAneoloxP}$ foi transferido para o background C57BL/6 (B6). Uma vez atingida a isogenia em ambas as linhagens, os animais heterozigotos, ou *knock-Out* (KO), apresentaram alterações esqueléticas (cifose), pulmonares (enfisema pulmonar, com aumento dos espaços aéreos alveolares e destruição das paredes dos bronquíolos) e cardiovasculares (espessamento da parede da aorta, com fibras elásticas pertencentes a túnica média do vaso desorganizadas e fragmentadas). Alterações similares são observadas em outros modelos para a SMF (Green *e cols.*, 1976 ; Pereira *e cols.*, 1997 ; Judge *e cols.*, 2004), e são representativas do fenótipo observado em humanos (Loeys *e cols.*, 2010).

Porém, as alterações observadas no fundo B6 são mais amenas e homogêneas, manifestando-se apenas aos 6 meses de idade, enquanto que heterozigotos da linhagem 129/Sv apresentam alterações mais graves e heterogêneas aos 3 meses de idade (fig. 8) (Lima *e cols.*, 2010).

Figura 8: Manifestações fenotípicas do modelo $mg\Delta^{loxPneo}$ observadas entre animais heterozigotos das linhagens 129/Sv (3 meses de idade) e C57BL/6 (6 meses de idade). Seções histológicas de pulmão e aorta corados com HE, e radiografia da coluna vertebral.

A figura 7 mostra ainda uma variabilidade fenotípica entre heterozigotos da linhagem 129/Sv. A diferença fenotípica entre animais isogênicos, como a observada entre os animais 129/Sv do modelo $mg\Delta^{loxPneo}$, não pode ser explicada por diferenças genéticas por definição, e já foi relatada em diferentes modelos animais para várias doenças (Rakyan e cols., 2002). Assim, é provável que fatores epigenéticos também estejam envolvidos na determinação fenotípica da SMF.

Logo o modelo $mg\Delta^{loxPneo}$ pode ajudar a elucidar a variabilidade intra-familiar da SMF, através da identificação de genes modificadores do fenótipo entre as linhagens 129/Sv e C57BL/6. Este modelo também permitirá identificar possíveis fatores epigenéticos que levem à expressão diferencial de genes entre os animais da linhagem 129/Sv e expliquem a variabilidade observada na linhagem.

1.6. Genes Modificadores

Os modelos de herança mendeliana tentam explicar como a transmissão de defeitos genéticos “simples” em apenas um gene pode estar associada com o aparecimento de um determinado quadro clínico. Esses modelos implicam na existência de uma relação genótipo-fenótipo que permite prever a quais alterações fenotípicas um indivíduo está sujeito, tornando possível a escolha de tratamentos e terapias apropriadas para cada caso.

Muitas vezes a descoberta dos genes causadores não consegue explicar toda a variabilidade clínica observada nos indivíduos acometidos pela doença. Assim, conceitos como penetrância incompleta, expressividade variável (revisados em Miko, 2008), modelos multifatoriais com diferentes limiares (Dipple e McCabe, 2000) e a existência de genes modificadores do fenótipo foram criados para explicar as distorções encontradas nas correlações entre o fenótipo clínico e o genótipo de cada doença.

Em 1941, John Haldane estudou a correlação entre as idades de acometimento de várias doenças entre pares de irmãos e entre pais e suas proles. Neste trabalho, Haldane postulou que se essas características fossem determinadas por apenas um gene, as correlações (ρ) obtidas entre os indivíduos aparentados estudados desviariam muito pouco de $\rho = 0.5$. Contudo, para quase todas as doenças analisadas, as correlações encontradas eram superiores a $\rho = 0.7$, o que levou à conclusão de que existiam um ou mais genes envolvidos na idade de manifestação dessas doenças e que contribuíam para o fenótipo observado. Assim, genes que modulam o fenótipo e não são genes causais da doença são considerados como genes modificadores do fenótipo (Haldane, 1941).

O conceito de gene modificador é um conceito relativamente vago. Várias definições foram criadas ao longo dos anos e abrem espaço para várias interpretações diferentes (Grüneberg 1963; Hall e Horton 1997). De forma mais abrangente do que a proposta originalmente por Haldane e por outras definições, podem ser considerados genes

modificadores aqueles cuja expressão é capaz de alterar a expressão de outro gene, sendo que o termo expressão pode se referir ao nível de transcrição ou, ainda, à manifestação fenotípica desse gene, desde níveis moleculares até alterações sistêmicas (Nadeau, 2001).

Um gene modificador pode influenciar um fenótipo alterando as seguintes características:

- **Penetrância:** altera o limiar de expressão para que mais ou menos indivíduos apresentem uma determinada alteração fenotípica. (fig. 9a);
- **Dominância:** altera o limiar de expressão para que características recessivas se tornem semi-dominantes ou o oposto. (fig. 9b);
- **Expressividade:** modula o nível de acometimento fenotípico (mais ou menos extremo). (fig. 9c);
- **Pleiotropia:** a interação entre o gene modificador e outros genes pode alterar fenótipos, ou conjuntos de fenótipos, diferentes em cada indivíduo. (fig. 9d) ;

Figura 9: Possíveis efeitos de genes modificadores. Linhas contínuas representam a manifestação fenotípica na ausência de gene modificador, enquanto linhas pontilhadas representam o fenótipo na presença de gene modificador. A) Penetrância; B) Dominância; C) Expressividade e D) Pleiotropia.

Apesar da grande variedade de efeitos modificadores, é importante diferenciar quando uma característica é influenciada por genes modificadores ou quando esta é determinada por um modelo poligênico. A diferença reside no fato de que, no modelo poligênico, cada alelo dos genes envolvidos contribui na determinação do fenótipo. Assim, a alteração do genótipo de um dos *locus* envolvidos (ex.: de aa para Aa ou de Aa para AA) acarreta uma alteração fenotípica para todas as outras categorias genotípicas, enquanto que a mesma alteração fenotípica em um modelo com genes modificadores só vai se manifestar em uma categoria genotípica (Nadeau, 2001) (fig. 10).

Figura 10: Diferença entre fenótipos com característica poligênica (A) e influenciado por gene modificador (B). (Adaptado de Broman, 2009).

Entretanto, as características descritas acima são específicas de cada gene modificador: cada traço será influenciado por um determinado número de genes modificadores com características e comportamentos específicos.

A quantidade de genes envolvidos na determinação de cada característica varia de traço a traço. Quando um único gene modificador explica grande parte ou toda a variabilidade fenotípica observada, esta característica segue um modelo similar ao monogênico. Porém, quando um conjunto de genes modificadores com efeitos individuais moderados mas que interagem entre si e com variáveis ambientais é usado para explicar toda a variabilidade, a característica assume um modelo multifatorial (Génin, 2008). Ou, ainda, quando a variabilidade fenotípica só pode ser explicada por uma imensa quantidade de genes modificadores com efeitos muito discretos, a característica apresenta um modelo poligênico ou “infinitesimal” (Mackay 2001).

Embora esses três modelos sejam possíveis, dois deles apresentam algumas contradições que os tornam menos aplicáveis. Apesar do modelo poligênico ser ótimo para

modelagem teórica de como traços quantitativos podem ser influenciados por muitos *loci*, este não faz muito sentido do ponto de vista genético uma vez que nele o efeito de cada alelo nos *loci* envolvidos é idêntico e, juntamente com a grande quantidade de genes envolvidos, tornaria o efeito individual tão ínfimo que seria impossível mensura-lo. Já o que torna o modelo monogênico menos aplicável é a baixa frequência de QTLs (Quantitative Trait Loci) que expliquem grande parte da variação fenotípica observada (fig. 11).

Figura 11: Frequência relativa dos QTLs já identificados de acordo com o poder explicativo de cada um. (Adaptado de Griffiths, 2006).

Apesar de genes modificadores se referirem a como a interação entre diferentes *loci* influencia a variabilidade clínica de doenças genéticas, e de QTLs serem *loci* que controlam traços quantitativos presentes em todos os indivíduos de uma população nos vários níveis fenotípicos possíveis, esses dois conceitos estão relacionados a variações de características complexas em uma população. Logo, os modelos e técnicas de mapeamento utilizados para a

identificação de QTLs também são úteis na identificação de genes modificadores, assim como o contrário também é válido (Génin e cols, 2008).

1.7. Mapeamento genético

As técnicas de mapeamento usadas na identificação de genes causadores de doenças, genes modificadores de fenótipo e de QTLs utilizam a análise de ligação para determinar as posições dos genes de interesse em relação à posição de sequências de DNA já conhecidas, chamadas de marcadores genéticos. A análise de ligação consiste em estimar o desequilíbrio de ligação existente entre dois *loci* diferentes: quanto mais próximos os *loci* estão maior a probabilidade de segregarem juntamente. Dessa forma, é possível utilizar desvios das proporções mendelianas esperadas para cada tipo de população sobre estudo para identificar regiões que segreguem com o fenótipo de interesse.

As linhagens de camundongos estabelecidas em laboratório são o fruto de décadas de seleção e de um rigoroso controle no cruzamento desses animais. Atualmente, já foram estabelecidas centenas de linhagens isogênicas, sendo que cada uma possui as suas particularidades (maior suscetibilidade a formação de tumores, maior massa corporal, resistência a certos agentes biológicos, entre outras) (<http://www.informatics.jax.org/external/festing/mouse/STRAINS.shtml>). Cada uma delas possuiu um conjunto único de alelos, uma vez que o contínuo acasalamento de animais com as características de interesse por muitas gerações acaba por fixar um alelo em cada loco, fazendo com que todos os animais de uma determinada linhagem sejam homozigotos para todas as regiões do genoma.

Quando utilizamos os animais gerados do cruzamento controlado de duas linhagens de camundongo, como o retrocruzamento (*backcross*) e o intercruzamento (*intercross*) (*vide* fig.

12), para realizar o mapeamento de genes responsáveis por doenças ou genes modificadores ou ainda QTLs, estamos garantindo uma série de propriedades que facilitam a identificação desses genes:

- A segregação do alelo de interesse entre os indivíduos estudados;
- A existência de desequilíbrio de ligação entre os marcadores e a região candidata;
- O número de estados que cada loco pode assumir é limitado à combinação dos alelos das linhagens parentais, logo o estudo se resume à análise de marcadores bialélicos em vez de marcadores multialélicos;

Figura 12: Cruzamentos controlados para realização de estudos de mapeamento. A) Retrocruzamento, B) Intercruzamento.

Independentemente da estratégia de cruzamento adotada para a realização do mapeamento, são utilizados marcadores moleculares polimórficos que se encontram por todo o genoma, podendo ser anônimos, como os altamente polimórficos microssatélites e a enorme quantidade de marcadores bialélicos SNPs (Single-Nucleotide Polymorphism) (Lin e cols., 2005)., ou ainda genes ou ESTs (Expressed Sequence Tags).

1.8. Técnicas matemáticas de mapeamento

O mapeamento com esses marcadores é feito através do ajuste de modelos de regressão para os efeitos aditivos e de dominância de cada marcador, sendo que o componente genético tem efeito fixo e o componente ambiental tem efeito aleatório $E \sim N(0, \sigma^2)$, do tipo:

$$Y = \mu + aX_a + dX_d + e$$

Onde,

$$X_{a_{ij}} = \begin{cases} 1 & \text{se } QQ \\ 0 & \text{se } Qq \\ -1 & \text{se } qq \end{cases} \quad X_{d_{ij}} = \begin{cases} 1 & \text{se } Qq \\ 0 & \text{cc} \end{cases}$$

Mesmo com a utilização de um grande número de marcadores e um grande número de animais, este procedimento necessita de muito tempo para ser concluído e é capaz de apenas indicar grandes regiões que contenham os *loci* de interesse, tornando necessária uma nova etapa de mapeamento para refinar a busca e identificar o gene modificador (Flint e cols.; 2005).

Uma possibilidade para se contornar a quantidade de marcadores limitada foi proposta por Harley e Knot em 1992, eles propuseram utilizar a função de Haldane para estimar a fração de recombinação entre uma certa posição cromossômica e os marcadores adjacentes a

ela e assim, estimar os efeitos aditivos e de dominância que um marcador em certa posição teria (fig. 13). Essa técnica ficou conhecida como Mapeamento Intervalar.

Efeitos Genéticos Esperados		
Genótipo do marcador	Expectativas em termos de:	
	a (efeitos aditivos)	d (efeitos de dominância)
$A_1A_1B_1B_1$	$[(1-r_A)^2(1-r_B)^2 - r_A^2r_B^2]/(1-r)^2$	$[2r_A(1-r_A)r_B(1-r_B)]/(1-r)^2$
$A_1A_1B_1B_2$	$[(1-r_A)^2r_B(1-r_B) - r_A^2r_B(1-r_B)]/r(1-r)$	$[r_A(1-r_A)(1-r_B)^2 + r_A(1-r_A)r_B^2]/r(1-r)$
$A_1A_1B_2B_2$	$[(1-r_A)^2r_B^2 - r_A^2(1-r_B)^2]/r^2$	$[2r_A(1-r_A)r_B(1-r_B)]/r^2$
$A_1A_2B_1B_1$	$[r_A(1-r_A)(1-r_B)^2 - r_A(1-r_A)r_B^2]/r(1-r)$	$[(1-r_A)^2r_B(1-r_B) + r_A^2r_B(1-r_B)]/r(1-r)$
$A_1A_2B_1B_2$	0	$[r_A^2r_B^2 + r_A^2(1-r_B)^2 + (1-r_A)^2r_B^2 + (1-r_A)^2(1-r_B)^2]/[r^2 + (1-r)^2]$
$A_1A_2B_2B_2$	$[r_A(1-r_A)r_B^2 - r_A(1-r_A)(1-r_B)^2]/r(1-r)$	$[(1-r_A)^2r_B(1-r_B) + r_A^2r_B(1-r_B)]/r(1-r)$
$A_2A_2B_1B_1$	$[r_A^2(1-r_B)^2 - (1-r_A)^2r_B^2]/r^2$	$[2r_A(1-r_A)r_B(1-r_B)]/r^2$
$A_2A_2B_1B_2$	$[r_A^2r_B(1-r_B) - (1-r_A)^2r_B(1-r_B)]/r(1-r)$	$[r_A(1-r_A)(1-r_B)^2 + r_A(1-r_A)r_B^2]/r(1-r)$
$A_2A_2B_2B_2$	$[r_A^2r_B^2 - (1-r_A)^2(1-r_B)^2]/(1-r)^2$	$[2r_A(1-r_A)r_B(1-r_B)]/(1-r)^2$

Figura 13: Equações derivadas por Harley e Knot em 1992 para determinar os efeitos aditivos e de dominância que um marcador teria dado o genótipo e a fração de recombinação entre os marcadores que o flanqueiam.

Entretanto, quando dois QTLs estão fisicamente próximos, a simples estimativa desses efeitos e a inclusão de mais marcadores no modelo de regressão gera um efeito conhecido como QTL fantasma, pois apesar de matematicamente existir a evidência de mais um QTL entre os dois primeiros, não é possível localizar geneticamente esse terceiro QTL. Isso se deve ao desequilíbrio de ligação que existe entre os dois QTLs externo e que leva a um falso positivo na posição intermediária a eles.

A proposta para solucionar o desequilíbrio existente entre dois QTLs ligados foi a inclusão de um termo no modelo de regressão que concentrasse a influência dos outros marcadores e não permitisse a interferência de um no outro. Logo, a fórmula de regressão ficou dessa forma:

$$Y = \mu + aX_a + dX_d + \sum \beta_j X_j + e$$

Assim, os valores β_j irão impedir que a proximidade dos marcadores influencie na estimativa dos efeitos aditivos e de dominância. Tal técnica de mapeamento recebeu o nome de Mapeamento Intervalar Composto.

1.9. Estratégias de Mapeamento

Independente do tipo de cruzamento e de marcadores utilizados o processo de fenotipagem e de genotipagem dos animais é bastante trabalhoso, demorado e dispendioso. Para se reduzir o custo e o tempo gasto nessa etapa do mapeamento é possível selecionar um sub-conjunto de animais com fenótipos extremos (fig. 13) capaz de retornar grande parte da informação que seria obtida com a genotipagem do conjunto completo (fig. 14).

Figura 14: Genotipagem seletiva em um fenótipo com distribuição normal

A eficiência da genotipagem seletiva depende da densidade de marcadores utilizados e da quantidade de animais a serem genotipados, como mostrado na figura 15. Esta abordagem é muito útil quando o custo de manutenção, criação e fenotipagem dos animais é inferior ao custo de genotipagem (Broman e cols., 2009).

Figura 15: Influência da quantidade de animais selecionados e da densidade de marcadores sobre a informação obtida do mapeamento.

Quando o custo de fenotipagem supera os custos de manutenção e de genotipagem é recomendável que seja feita uma fenotipagem seletiva, assim serão selecionados os animais que apresentarem a maior diversidade genética. A eficiência desse método é condicional ao conhecimento prévio da região candidata a conter o gene modificador ou QTL, caso contrário a fenotipagem seletiva é tão boa quanto fenotipar um sub-conjunto aleatório de animais (Broman e cols., 2009).

Quando escolhemos um sub-conjunto de animais que representam os extremos da distribuição fenotípica, a quantidade de animais com fenótipos mais graves é a mesma que de

animais com fenótipos mais leves. Assim, fenótipos que apresentam uma distribuição normal seriam melhores candidatos para a realização da genotipagem seletiva. Entretanto Sen e colaboradores publicaram um trabalho em 2009 onde a eficiência das duas estratégias descritas acima foram testadas para distribuições fenotípicas não-normais, e mostraram que essas estratégias podem ser úteis em fenótipos que fujam da normalidade, em detectar múltiplos QTLs e interações entre eles.

2. Objetivos

2.1. Objetivo Geral

Levando em consideração as diferenças fenotípicas apresentadas pelo modelo $mg\Delta^{loxPneo}$ entre os *backgrounds* 129/Sv e C57BL/6, acreditamos que exista pelo menos um gene modificador do fenótipo associado a variabilidade observada na SMF.

Assim, este trabalho pretende identificar genes modificadores do fenótipo da SMF, o que pode auxiliar na determinação de correlações genótipo-fenótipo mais precisas e a melhor compreensão da variabilidade clínica da síndrome.

2.2. Objetivos Específicos

- 1) Estabelecer um sistema de cruzamento do modelo $mg\Delta^{loxPneo}$ capaz de reproduzir a variabilidade clínica da SMF.
- 2) Estabelecer métodos quantitativos de fenotipagem das alterações esqueléticas, pulmonares e cardiovasculares para caracterizar fenotipicamente os animais das linhagens isogênicas 129/Sv e C57BL/6.
- 3) Identificar regiões candidatas a conterem genes modificadores do fenótipo, através de estudo de ligação utilizando microssatélites e SNPs.
- 4) Identificar genes e vias diferencialmente expressas entre animais 129/Sv com alterações leves e graves utilizando *microarray* de expressão. Tais genes são prováveis candidatos a genes modificadores, podendo assim confirmar ou restringir as regiões identificadas no mapeamento.

3. Material e métodos

3.1. Extração de DNA

Em torno dos vinte dias de idade os animais foram desmamados e um pedaço da cauda foi retirado para realizar a seleção dos animais portadores do alelo.

A amostra de tecido foi mantida em nitrogênio líquido ou em freezer -70°C até ser processada, sendo então macerada e embebida em 200 µL de solução de lise (50 mM Tris-HCl pH 8.0, EDTA 10 mM pH 8.0 e SDS 0,5%) com proteinase K (1,5 mg/mL); incubada à 65°C por duas horas; acrescentada de mais 200 µL de solução de lise; centrifugada a 12.000 rpm durante 20 minutos; 840 µL de etanol absoluto com acetato de sódio (150 mM) gelado foi adicionado ao sobrenadante; centrifugada a 5.000 rpm durante 5 minutos; desprezado o sobrenadante; lavada com 500 µL de etanol 70% gelado e ressuspender em 200 µL de TE (10mM TRIS, 1mM EDTA); a quantidade de DNA existente na amostra foi quantificada e dissolvida para uso.

3.2. Genotipagem dos animais

Após extraídas as amostras de DNA, cada animal foi genotipado através de técnica de PCR. Duas reações foram montadas para cada amostra de DNA genômico: a primeira utilizando os primers NeoF e NeoR (433bp; Tm 57°C) [amplificam o cassete neo do alelo mutado *FbnI*^{neoloxp}]; e a segunda usando os primers 22F e CC-R (600pb; Tm 55°C) [amplificam os exons 21 a 22].

NeoF: 5' -GAG GCT ATT CGG CTA TGA CT – 3'

NeoR: 5' – CTC TTC GTC CAG ATC ATC CT – 3'

20F: 5' – AAA CCA TCA AGG GCA CTT GC – 3'

CC-R: 5' – CAC ATT GCG TGC CTT TAA TTC – 3'

3.3. Caracterização fenotípica

3.3.1. Análise radiológica

Os animais foram sacrificados por inalação de CO₂ e posteriormente fixados em cassetes contendo filme para radiografia. Seus membros posteriores e anteriores foram distendidos, para que pudéssemos avaliar o grau de escoliose da coluna. Posteriormente foram expostos a 4,0 Kv de raio-x durante o período de 3,0 ms (milissegundos) (Tecno-Design – 500Ma/ 125 Kv).

Para quantificar o desvio de coluna, a imagem radiográfica de cada animal foi digitalizada. Através da utilização do *software* AutoCAD (Autodesk ; versão 2002) ou ainda com o auxílio *software ImageJ* (MURAISHI), foram calculadas a distância linear e o comprimento do espaço entre diferentes combinações de cada porção da coluna vertebral (cervical, torácica e lombar).

3.3.2. Análises histopatológicas

Amostras de pulmão e aorta foram submetidos às técnicas rotineiras de fixação, desidratação, diafanização, inclusão em parafina e microtomia (ANDRIKOPOULOS *et al.*, 1995). Cortes medindo 5 µm de espessura foram corados com hematoxilina-eosina (HE), e

cortes adjacentes foram corados por coloração de Weigert, específica para a visualização de fibras elásticas. Em seguida os cortes foram examinados e fotografados utilizando-se o equipamento Axiovert 200 (Carl Zeiss).

Parede aórtica: As amostras foram fotografadas em ampliações de 50x e 100x. Foram mensurados os perímetros interno e externo da aorta, com o auxílio do *software ImageJ* (MURAISHI), e ainda contabilizado o número de lâminas elásticas. A partir desses valores foram estimados o diâmetro interno e externo da aorta, bem como a espessura média de cada lâmina e de toda a parede.

Tecido pulmonar: As amostras foram fotografadas na ampliação de 50x. O diâmetro alveolar médio (Lm) foi mensurado como previamente descrito (DUNNILL, 1962).

3.4. Avaliação da variabilidade e estratégia de cruzamentos

Para se localizar genes candidatos a modificadores do fenótipo ou regiões candidatas, animais $mg\Delta^{loxPneo}$ das linhagens co-isogênicas C57BL/6J e 129/Sv foram cruzados para produzir a F1, ou seja, $Fbn1(+/-)^{C57}$ com $Fbn1(+/+)^{129}$ ou $Fbn1(+/-)^{129}$ com $Fbn1(+/+)^{C57}$. Os machos $Fbn1(+/-)$ da F1 foram inter cruzados com fêmeas $Fbn1(+/+)$ da F1 para se produzir a geração F2 (figura 12B).

A cada geração, foram coletadas amostras de DNA dos animais para identificar os portadores do alelo $Fbn1^{mg\Delta^{neoloxP}}$ e estes tiveram os fenótipos ósseo, cardiovascular e pulmonar quantificados aos três meses de idade, os animais foram fenotipados com essa idade devido a diferença fenotípica observada entre as linhagens isogênicas.. A partir daí, foi analisado um painel de 60 marcadores microssatélites polimórficos entre as linhagens

C57BL/6J e 129/Sv nos 10 animais mais representativos de cada extremo da distribuição fenotípica com o intuito de localizar regiões que possuam genes modificadores.

3.5. Mapeamento com microssatélites

O DNA extraído dos animais mutantes foi inicialmente analisado por PCR de 60 microssatélite polimórficos entre as linhagens parentais que cobrisse a maior parte do genoma murino. Os microssatélites utilizados foram amplificados por PCR (1,5 mM MgCl₂, 80 uM DNTP, 0,67 pM de cada primer, 5 U/ul de taq polimerase e 50 ng de DNA em 15ul), utilizando-se termociclador MJ Research PTC-200: inicialmente 95°C, 2 min. seguido por 30 ciclos de 95°C por 30 segundos, 57 °C por 1min e 20 seg., 72 °C por 1 min. e 15 seg.; finalizando com 95 C por 50 seg., 57 C por 1 min. e 20 seg., 72 C por 6 min. e 4 C por 30 min. Os produtos de pcr foram separados através de eletroforese em géis de agarose NuSieve 3:1 4% submetidos a uma d.d.p. de 100V pelo período de 1 à 3 horas, e visualizadas com brometo de etídio, Blue Green ou Gelred; ou ainda através de géis de poli-acrilamida 10% submetidas a uma d.d.p. de 200 V pelo período de 3 horas e coradas com prata, a escolha do gel a ser utilizado para a separação das amostras foi feita levando-se em consideração o tamanho do fragmento e na diferença de tamanho que os fragmentos das linhagens 129/Sv e C57BL/6 apresentam. Os géis foram interpretados e devidamente digitalizados para facilitar o armazenamento dessas informações. Vide anexo 1 para mapa dos marcadores utilizados

3.6. Mapeamento com SNPs

O mapeamento utilizando SNPs foi realizado utilizando os mesmo animais que no mapeamento com microssatélites. Para a genotipagem dos marcadores foram utilizadas

lâminas de *microarray Illumina Infinium Mouse Genotyping*, capaz de genótipar 7851 SNPs ao longo de todo o genoma murino. Todos os reagentes necessários e a genotipagem foram realizados pela empresa Hellixa. (<http://www.helixxa.com/ptbr/index.html>).

Os SNPs polimórficos entre as linhagens 129/Sv e C57BL/6 foram identificados comparando os resultados obtidos entre amostras de animais isogênicos. No anexo 2 encontra-se o script utilizado para identificar os SNPs e preparar o arquivo de entrada de dados.

3.7. Microarray de expressão

As amostras de tecido pulmonar de 8 animais 129/Sv (4 animais selvages, 2 heterozigotos com manifestações leves e 2 com manifestações graves) com 3 meses de idade foram coletas, o RNA foi extraído com TRIZOL, purificadas com o kit RNeasy, marcadas com cianina-5 fluorescente e hibridada com o array murino de 44 mil sondas produzidos pela Agilent (4x44k). Os dados foram extraídos utilizando o software *feature extraction*. Todos os passos foram realizados de acordo com o protocolo dos fabricantes e em parceria com Prof Dr. Rodrigo Panepucci do Hemocentro de Ribeirão Preto, USP.

3.8. Real-time PCR

A validação dos genes diferencialmente expressos identificados no microarray de expressão foi realizada utilizando o mesmo protocolo utilizado por Lima e cols. 2010. Foram utilizados 12 animais 129/Sv de três meses de idade, sendo 4 animais selvagens, 4 animais com manifestões pulmonares leves e 4 animais com alterações graves. A aferição do nível de expressão foi realizada num equipamento StepOnePlus da applied Biosystems utilizando o

reagente SYBR GREEN[®] (Applied biosystems), sendo que as análises foram realizadas no software que acompanha o equipamento.

Para garantir a especificidade dos primers, estes foram desenhados em junções exon-exon utilizando a ferramenta Primer_Blast (http://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=BlastHome) e estão listados abaixo.

<i>Cd109:</i>	5'- CGA GAT CCA TCC TGG AAG CAG AAG G-3'
	5'- CTG CTC AGA GGT AGT GCC GGA AG-3
<i>Cpb2:</i>	5'- CCC GTG ATG AAC GTG GAT GGC T -3'
	5'- GCA GAG CGG TTC TTC CTC CAC A -3
<i>Fcna:</i>	5'- GAG AAC GGG GCT TTC AGG GCA G -3'
	5'- CCA TGG TTC CTG GCT CTC CTT TGC -3
<i>Fbln7:</i>	5'- CAA GGC TGT GAG TGG CAA GAT GGG -3'
	5'- TGG TGG TGA GGA GCT GCT GTT TG -3
<i>S100b:</i>	5'- CAA CGA GCT CTC TCA CTT CCT GGA G -3'
	5'- CCC ATC CCC ATC TTC GTC CAG C -3
<i>Adamts13:</i>	5'- ACC AAG GAT GGA GCG CTG CTG -3'
	5'- TCG GAA CCC AGA TGA CTG GCT ACA -3
<i>Edn1:</i>	5'- CAG GGC CCA AAG TAC CAT GCA GA -3'
	5'- GCT GAT GGC CTC CAA CCT TCG T -3
<i>Prrx2:</i>	5'- CAG GCG CTG GAG CGT GTA TTT G -3'
	5'- CGG TTC TGG AAC CAG ACT TGG ACA C -3
<i>Eml6:</i>	5'- GCC TGG CTT TTG ACT CGG ATG GA -3'
	5'- GGC CAG TAG CTG AGG CCA GAA G -3
<i>Scgb3a1:</i>	5'- GCT CTG CTC AGT GAC TCT GGT GTT -3'
	5'- AAG GGC GGC CAC GGG TT C -3
<i>Syt6:</i>	5'- CAG TCC CCG GAG CAG AGT CCA -3'
	5'- GGG CCA CGC CAC ACA CAA TAA C -3

3.9. Análise estatística e Bioinformática

Todas as análises estatísticas realizadas nesse trabalho foram feitas no ambiente estatístico R (<http://cran.r-project.org/>) juntamente com as bibliotecas apropriadas para cada experimento.

Análise de Ligação: foi utilizada a biblioteca R/QTL para estimar a estatística LOD score e identificar as regiões em desequilíbrio de ligação com o fenótipo. Foram utilizados dois métodos robustos de mapeamento intervalar, ambos utilizando a equação de Haldane para converter distancias genéticas em fração de recombinação, sendo que o método de mapeamento de imputações múltiplas foi utilizado para aprimorar o mapeamento com microsatélites (1024 imputações), enquanto que para o mapeamento com SNPs foi utilizada a técnica de regressão desenvolvida por Harkey-Knott. Os limiares de sugestividade ($P < 0.63$) e significância ($P < 0.05$) foram estimados através de permutações (1000 permutação para cada fenótipo e cada experimento) conforme descrito em Broman e cols. 2009. Antes do processo de mapeamento do SNPs ter sido realizados, aqueles SNPs que não estavam em equilíbrio de ligação foram removidos da análise por serem possíveis erros de genotipagem. Vide anexo 3 e anexo 4 para os respectivos scripts de análise

Expressão diferencial: Foram utilizadas as bibliotecas Agi4x44PreProcess, limma, multtest, vsn, genefilter (pertencentes ao projeto BioConductor). Foram selecionados os arrays com menor coeficiente de variância e a variação observada entre esses array foi normalizada (vide anexo 5 para script utilizado). Devido as poucas amostras utilizadas os genes diferencialmente expressos foram identificados utilizando modelos de regressão linear, a correção para múltiplos testes foi feita estimando a taxa de falso positivo (FDR- false Discovery rate) para o nível de 5% (comparações de animais leves e graves contra animais

selvagens) e 10% (comparação entre animais leves graves) de significância estatística. Apenas aqueles genes que atingiram significâncias estatísticas e apresentavam uma diferença de expressão de pelo menos 40% foram considerados diferencialmente expressos. Como descrito previamente em Hahne e cols. 2008.

GSEA: Foi realizada utilizando as funções existentes nas bibliotecas GSEABase, KEGG e Category para identificar a que vias cada gene diferencialmente expresso pertencia e em seguida aquelas vias que apresentavam um nível de significância inferior a 5%, estimado através 1000 permutações do teste, foram consideradas como vias alteradas entre os grupos. Como descrito previamente em Hahne e cols. 2008.

4. Resultados

4.1. Caracterizando a variabilidade clínica em animais híbridos

Animais heterozigotos das linhagens 129/Sv e C57BL/6 foram cruzados para gerar animais híbridos (F1), e estes foram inter cruzados para gerar a população de mapeamento (geração de animais F2) (Figura 16).

Figura 16: Heredograma dos cruzamentos entre as linhagens 129/Sv e C57BL/6. Animais heterozigotos visualmente afetados e assintomáticos estão indicados de acordo com a legenda.

Aos três meses de idade os animais heterozigotos para o alelo *Fbn1^{mg ΔneoloxP}* da geração F1 apresentaram manifestações nos três sistemas estudados, com graus de acometimento similares (fig. 17). A manifestação de alterações aos 3 meses é uma característica presente apenas na linhagem isogênica 129/Sv, sugerindo que os alelos 129/Sv são dominantes e levam a um agravamento das alterações fenotípicas.

Figura 17: Comparação do fenótipo dos sistemas ósseo, pulmonar e cardiovascular de animais F1 com três meses de idade. A - animal selvagem; B e C - animais heterozigotos.

Entretanto, nos animais F2 a concordância entre a gravidade fenotípica de cada sistema afetado observada nos animais isogênicos e na geração F1 não foi mantida: os animais apresentaram diferentes combinações de gravidades entre os três sistemas (fig. 18).

Logo, é mais provável que cada sistema tenha um conjunto de genes modificadores do que apenas um único conjunto modulando todos os fenótipos observados. Por isso, para realizar um estudo de ligação nesses animais foi necessário agrupá-los de acordo com a gravidade fenotípica de cada um dos sistemas ósseo, cardíaco e pulmonar, e analisar que cada uma das manifestações separadamente.

A estratégia de cruzamento utilizada foi capaz de gerar bastante variabilidade nos três sistemas estudados, animais com diferentes graus de acometimento em cada sistema, o que é propício para a identificação de genes modificadores do fenótipo específicos de cada linhagem. Assim, foram obtidos 83 animais F2 portadores do alelo *Fbn1*^{mgΔneoloxP}, os quais

tiveram os fenótipos dos sistemas ósseo, pulmonar e vascular quantificados para o estudo de ligação.

Figura 18: Comparação do fenótipo dos sistemas ósseo, pulmonar e cardiovascular de animais F2 com três meses de idade. A - animal selvagem; B, C e D - animais heterozigotos diferentemente afetados em cada um dos três sistemas..

4.2. Métodos quantitativos de fenotipagem

4.2.1. Sistema esquelético

Para quantificar o desvio de coluna, a radiografia de cada animal foi digitalizada, e através da utilização do programa AutoCAD, foram calculadas a distancia linear de cada região (cervical, torácica e lombar) da coluna e o comprimento de cada uma delas para cada animal (Figura 19). Além disso, também foram mensuradas as distancias lineares da porção cérvico-torácica e de toda a coluna.

Figura 19: Quantificação do desvio da coluna com programa AutoCAD a partir de radiografias dos animais. Foram mensuradas a distancia linear e o comprimento da coluna de cada região. As cores vermelha, amarela e verde representam as regiões cervicais, torácicas e lombar, respectivamente

Para tentar separar os animais de acordo com o nível de acometimento, foi utilizada a razão entre essas medidas [distância linear/comprimento], sendo que a gravidade do desvio da coluna é inversamente proporcional ao valor desta razão (fig. 19).

A análise da região cervico-toracica é a que apresentou uma distribuição mais simétrica em torno da média, assemelhando-se mais a uma normal (fig. 20). Como cada gene segrega independentemente, é esperado que uma característica quantitativa, e portanto, complexa, determinada por poucos genes, possua uma distribuição fenotípica semelhante a uma distribuição normal. Assim foram escolhidos os valores da região cervicotorácica como variável fenotípica para a identificação de genes envolvidos na determinação do fenótipo ósseo.

Figura 20: Comparação entre as distribuições obtidas da avaliação de cada região da coluna cervical dos animais F2 knock-outs. Distribuição entorno da média de cada conjunto de dados (observação/média obsevações).

A figura 21 indica os valores obtidos quando comparados a distancia linear e o comprimento da região cervico-toracica de animais selvagens e KO das linhagens parentais e das duas gerações de animais híbridos.

Figura 21: Gráfico com o valores absolutos da razão (distância linear/coluna) para os animais selvagens e knock-outs das linhagens parentais e de cada geração híbrida aos 3 meses de idade. Dados

Observa-se uma variação quase contínua da razão da coluna nos animais F2 KO, desde animais com valores semelhantes aos animais selvagens até animais com valores inferiores aos observados nos animais KO das linhagens parentais. Portanto, para se avaliar a existência de genes modificadores influenciando este fenótipo, foi necessário utilizar um tratamento para traços quantitativos (QTLs) e não um tratamento discreto como imaginado originalmente.

Outra constatação importante feita a partir dos resultados expostos na figura 21, é que em média os animais F2 apresentaram alterações mais graves do que as observadas nas gerações parentais (teste-t, $p = 1,9 \times 10^{-3}$), indicando que cada uma das linhagens estudadas possuem diferentes locos/alelos de susceptibilidade, ou ainda de resistência, que quando

combinados na F2 são capazes de gerar alterações superiores àquelas encontradas nas linhagens parentais.

4.2.2. Sistema cardiovascular

Para quantificação das alterações observadas na estrutura da parede aórtica, foram mensuradas as áreas ocupadas pela aorta e pelo seu lúmen. Assim pode-se estimar o raio externo, o raio interno, e, através da diferença entre esses, a espessura da parede da aorta (fig. 22). Além dessas três medidas, também foi observado o número de lamelas elásticas presente no corte, e estimada a espessura média de cada lâmina (espessura aorta/número de lâminas) (fig. 23).

Figura 22: Método de quantificação das alterações cardio-vasculares. Estimação do raio externo, raio interno e espessura do vaso através da mensuração da área interna a cada linha vermelha.

Figura 23: Comparação entre as distribuições obtidas da avaliação de cada uma das 5 medições realizadas nas aortas dos animais F2 knock-outs. Distribuição ao redor da média de cada conjunto de dados (observação/média observações).

Diferentemente do fenótipo ósseo, nenhuma das cinco alterações vasculares observadas na parede aórtica apresentaram uma distribuição normal, sendo que algumas das distribuições apresentavam um padrão similar a uma distribuição de Poisson. Assim, para se realizar a análise do conjunto de genes modificadores que controlam as alterações aórticas seria necessário aplicar transformações aos dados obtidos e/ou utilizar métodos não-paramétricos de análise. Para esse sistema escolhemos como variável de análise aquela que apresentou maior variação: a espessura da parede da aorta.

Como observado com o fenótipo ósseo, temos que a espessura da parede aórtica também se comporta como uma variável contínua nos animais F2 KO, e que, essas alterações apresentam maior gravidade na F2 do que as observadas nas linhagens parentais (teste-t, $p=7 \cdot 10^{-4}$) (fig. 24).

Figura 24: Gráfico com a espessura da aorta (AWT) para os animais selvagens e knock-outs das linhagens parentais, e de cada geração dos animais híbridos aos 3 meses de idade.

4.2.3. Sistema pulmonar

Para as manifestações pulmonares foi utilizado um procedimento descrito por Dunnill e cols. (1962), que estima o espaçamento alveolar por meio do estimador intercepto linear médio (Lm). Este método consiste em, basicamente, contar o número de intersecções entre o tecido pulmonar e uma grade de tamanho conhecido, quanto maior for a razão $\frac{\text{tamanho grade}}{\text{número de intersecções}}$ maior será o espaço alveolar e mais afetado será o animal (fig. 25). Podemos ver na figura 26 que a distribuição das medidas obtidas com essa metodologia é similar a uma normal.

Figura 25: Método de quantificação do fenótipo pulmonar.

Figura 26: Distribuição das alterações fenotípicas pulmonares

Da mesma maneira que o fenótipo ósseo, o fenótipo pulmonar tem um comportamento que se assemelha a uma variável continua com distribuição normal. Porém, diferentemente dos outros dois fenótipos, a grau de acometimento dos animais F2 KO não é diferente da observada nos animais 129/Sv KO (teste-t; $p=0,12$), o que pode indicar que a linhagem 129/Sv

tenha os alelos de risco que estejam controlando a variabilidade do fenótipo pulmonar (fig. 27).

Figura 27: Gráfico com o as medidas de intercepto linear médio (Lm) para os animais selvagens e knock-outs das linhagens parentais, e de cada geração dos animais híbridos aos 3 meses de idade. Dados

4.3. Concordância fenotípica

Com a definição dos métodos de quantificação fenotípica, podemos confirmar que a concordância das gravidades dos fenótipos observada nas linhagens parentais (fig. 28) e a ausência desta nos animais F2 (fig. 29) são verdadeiras, e não se devem aos métodos subjetivos de análise.

Figura 28: Gráficos de correlação entre os fenótipos de animais KO das linhagens parentais 129/Sv (pontos vermelhos) e C57BL/6 (pontos pretos). A) Lm x KR; B) AWT x KR; C) AWT x Lm; D) Valores do coeficiente de correlação de Pearson (triângulo inferior) e níveis de significância (triângulo superior).

Figura 29: Gráficos de correlação entre os fenótipos de animais KO dos animais F2 híbridos 129xB6. A) Lm x KR; B) AWT x KR; C) AWT x Lm; D) Valores do coeficiente de correlação de Pearson (triângulo inferior) e níveis de significância (triângulo superior).

A falta de correlação entre os fenótipos e a maior variabilidade observada para os animais F2 nos sistemas cardiovascular e ósseo indicam que cada sistema deve possuir o seu próprio conjunto de genes modificadores, sendo que os mesmos não são mutuamente exclusivos, e que existem alelos de susceptibilidade diferentes nas linhagens 129/Sv e C57BL/6. Assim, optou-se por realizar o mapeamento utilizando uma genotipagem seletiva de cada um dos três sistemas de forma a identificar a maior quantidade possível de genes modificadores do fenótipo.

Além de possibilitar uma melhor caracterização dos animais, os métodos quantitativos de fenotipagem permitem estimar a herdabilidade de cada traço e assim, saber o quanto da variabilidade observada é explicado por componentes genéticos (tabela 2). Sendo a herdabilidade estimada por:

$$H^2 = \left(\sigma^2_{F2} - \frac{(\sigma^2_{129} + \sigma^2_{C57} + 2\sigma^2_{F1})}{4} \right) / \sigma^2_{F2}$$

Onde, σ^2_{129} , σ^2_{C57} , σ^2_{F1} e σ^2_{F2} representam a variância observada em cada um dos grupos de animais, 129/Sv, C57BL/6, animais híbridos da geração F1 e da geração F2, respectivamente.

Tabela 2: Coeficiente de herdabilidade (H^2) estimado

Fenótipo	Herdabilidade (H^2)
Ósseo (KR)	40%
Cardiovascular (AWT)	76%
Pulmonar (Lm)	32%

A alta herdabilidade encontrada para o fenótipo cardiovascular (AWT) mostra a importância dos fatores genéticos na determinação deste. Apesar da herdabilidade reduzida para os outros fenótipos, estes ainda apresentam significativa influência genética, mas estão sujeitos a maior influência de fatores ambientais e/ou epigenéticos.

4.4. Qualidade dos dados

Com os procedimentos de fenotipagem concluídos, optou-se por realizar uma genotipagem seletiva dos animais obtidos na F2. Esse padrão de mapeamento envolve a genotipagem dos extremos de cada fenótipo - usualmente os animais que representam o primeiro e o último decil. No total foram genotipados 49 animais para cobrir os extremos das 3 variáveis em estudo. Um resumo dos dados de todos os animais obtidos, genotipados ou não, se encontra na figura 30.

Figura 30: Resumo dos dados usados para o mapeamento dos genes modificadores do fenótipo. A) mapa genético, B, C e D) distribuição dos fenótipos estudados, E) Sexo (0 para fêmea, 1 para macho) dos animais da amostra.

Como os histogramas referentes aos fenótipos da amostra total se mostram pouco semelhantes à uma distribuição normal, foi necessário utilizar o teste de normalidade de Shapiro para confirmar a normalidade dos fenótipos (fig. 31). Como pode ser visto pelos QQ *plots* e pela tabela de p-valores, apenas a distribuição cardíaca (AWT) não pode ser aproximada por uma normal e precisou ser tratada com um método de análise não paramétrico.

Figura 31: QQ plots para os traços estudados. A) Ósseo (KR). B) Cardíaco (AWT). C) Pulmonar (Lm). D) níveis de significância para o teste de normalidade de Shapiro.

4.5. Análise de ligação utilizando microssatélites

Para a etapa inicial do mapeamento dos possíveis genes modificadores envolvidos na SMF foi utilizado um conjunto de 60 marcadores microssatélites polimórficos entre as duas linhagens para mapear os animais com manifestações fenotípicas mais extremas de cada um dos três sistemas estudados.

É importante anotar corretamente as posições dos marcadores utilizados, uma vez que um estudo de mapeamento utilizando um mapa genético impreciso ou com erros pode levar a associações errôneas de certas regiões. Por esta razão, a posição de cada microssatélite utilizado foi atualizada de acordo com a versão mais recente existente no banco de dados *Mouse Genome Informatics* (versão 4.4). Além disso, também foram averiguadas as frações

de recombinação entre os marcadores utilizados para evitar erros, e como pode ser observado na figura 32, não existe nenhum desvio significativo nas frações de recombinação encontradas.

Figura 32: Nivel de recombinação estimado entre cada marcador (triângulo superior) e o nível de significância (triângulo inferior).

Devida à baixa quantidade de marcadores para essa etapa de mapeamento, foi aplicada uma técnica de mapeamento intervalar que permite o ajuste de covariáveis, como sexo e idade, e usa múltiplas imputações para estabelecer os genótipos de pseudomarcadores intermediários ao conjunto de marcadores genotipados. Com isso, aumenta-se a densidade do mapa genético, aumentando, portanto, a sensibilidade do estudo.

Os limiares de sugestividade ($p < 0.63$) e significância ($p < 0.05$) para cada fenótipo foram determinados através de 1000 permutações dos valores de cada fenótipo em relação aos genótipos observados, fazendo com que os limiares variem de uma análise para a outra.

Como pode ser visto na tabela 3, ao se analisar as alterações do sistema cardiovascular, a análise de variância (ANOVA) indicou que a única covariável que precisaria ser incluída na análise de ligação é a correção para sexo de cada animal.

Tabela 3: Análise de Variância para detecção de covariáveis

Covariáveis\ Fenótipo	Análise de Variância (p-valor)		
	KR	AWT	Lm
Idade	0.34	0.53	0.6
Sexo	0.9	0.02	0.25
Interação Sexo-Idade	0.79	0.29	0.29

Os resultados do mapeamento de genes modificadores dos fenótipo ósseo (fig. 33), cardiovascular (fig. 34) e pulmonar (fig. 35) podem ser vistos nas figuras abaixo.

Figura 33: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo Ósseo.

Figura 34: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo cardiovascular.

Figura 35: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo pulmonar.

Como se vê, os únicos cromossomos que ultrapassaram o limiar de sugestividade são o cromossomo 6, para o fenótipo cardiovascular (fig. 34), e os cromossomos 1, 6 e 16, para o fenótipo pulmonar (fig. 35).

Desse modo, é possível que existam genes modificadores nestes cromossomos, mas como não temos evidências de ligação significativa é necessário realizar uma análise mais extensa destes para que realmente sejam excluídas ou confirmadas a ligação dos fenótipos com esses cromossomos.

Uma vez identificados os cromossomos com ligação, é possível estimar a posição do gene modificador de fenótipo, o seu intervalo de confiança (tabela 4) e ainda estabelecer modelos de regressão para estimar o quanto da variabilidade observada este gene consegue explicar.

Tabela 4: Regiões candidatas identificadas

Fenótipo	Cromossomo	95% IC (cM)	Posição Estimada (cM)	Lod Score
Pulmonar	2	[57,8-89,8]	57,8	2,21
Pulmonar	6	[1,8-39,8]	5,8	2,3
Pulmonar	16	[14,6-38,6]	28,6	3,05
Cardiovascular	6	[6,8-77,7]	77,7	3,98

Como todas regiões candidatas identificadas são muito extensas, é necessário aumentar o número de marcadores utilizados no mapeamento para restringir ao máximo tais regiões. Para isso optou-se em utilizar um chip de microarray capaz de genotipar mais de 7800 SNPs existentes no genoma murino.

4.6. Análise de ligação utilizando SNPs

A grande vantagem do mapeamento com SNPs é a grande quantidade de marcadores existentes e a rápida genotipagem desses marcadores, o que possibilita o rápido mapeamento de uma grande quantidade de animais e com uma cobertura quase completa do genoma.

O chip utilizado genotipa um total de 7851 SNPs. Desses, apenas 2711 se apresentaram informativos entre as linhagens utilizadas, enquanto que apenas 2045 respeitavam o equilíbrio de Hardy-weinberg e puderam ser utilizados para identificar regiões associadas a genes modificadores do fenótipo.

Considerando que o genoma murino apresenta cerca de 1450 cM (Copeland e cols, 1993), o mapa genético obtido dos SNPs informativos entre as linhagens 129/Sv e C57BL/6 (figura 36) se expande por aproximadamente 1410 cM, uma cobertura quase completa de todo o genoma e muito superior a cobertura obtida com o uso de microsatélites (750 cM).

Figura 36: Mapa genético dos SNPs polimórficos entre 129/Sv e C57BL/6.

A posição genética (cM) de cada marcador foi estimada a partir da posição física (bp) utilizando o aplicativo *Mouse Map Converter* (<http://cgd.jax.org/mousemapconverter/>), capaz

de interpolar a posição de cada marcador se estes não estiverem presentes nos mapas construído por Cox e cols. em 2009. Assim a prévia ordenação dos marcadores em função da sua posição física em cada cromossomo garante o correto posicionamento de cada marcador e a precisão do mapa genético estimado.

Conforme feito para a etapa anterior, é necessário averiguar as frações de recombinação entre os pares de SNPs a procura de possíveis erros de genotipagem (fig. 37). Entretanto, só são visíveis os desvios referentes aos marcadores mais próximos em cada cromossomo (pontos vermelhos ao longo da diagonal), assim o baixo nível de recombinação já era esperado devido ao desequilíbrio de ligação existente devido à proximidade.

Figura 37: Nível de recombinação estimado entre cada marcador (triângulo superior) e o Lod Score observado para cada par (triângulo inferior).

Novamente o mapeamento de cada fenótipo foi realizado utilizando as funções do pacote R/qtl. Levando em consideração a maior quantidade de marcadores presentes nessa etapa do mapeamento, optou-se por utilizar o método desenvolvido por Harley-Knott pela sua velocidade, o melhor aproveitamento da quantidade de marcadores, e também por permitir o uso de covariáveis.

Figura 38: Perfil de Lod Score para o mapeamento com SNPs do fenótipo Ósseo.

Figura 39: Perfil de Lod Score para o mapeamento com SNP do fenótipo cardiovascular..

Figura 40: Perfil de Lod Score para o mapeamento com microssatélites do fenótipo pulmonar.

O mapeamento do fenótipo ósseo identificou regiões candidatas nos cromossomos 3, 6 e X, sendo que existe uma forte ligação da região identificada no cromossomo 6 com este fenótipo (fig.38).

Para o fenótipo cardiovascular, as regiões identificadas nos cromossomos 4 e 13 apresentam evidências moderadas de ligação (fig.39), enquanto que nenhuma região foi identificada associada ao fenótipo pulmonar (fig. 40).

As regiões identificadas utilizando os SNPs (tabela 5) diferem das regiões identificadas pelo mapeamento com microssatélites devido à maior cobertura genômica utilizada nessa etapa, e não aos diferentes métodos utilizados para lidar com a informação obtidas de cada conjunto de marcadores.

Os intervalos de confiança foram estimados e expandidos até o próximo marcador genotipado para evitar que aproximações numéricas pudessem interferir no lod score estimado nas regiões entre os marcadores e alterassem o intervalo de confiança observado.

Tabela 5: Regiões candidatas identificadas (SNPs)

Fenótipo	Cromossomo	95% IC (cM)	Posição Estimada (cM)	Lod Score
Ósseo	3	[56,1-68,4]	64,49	3,85
Ósseo	6	[41-49,1]	45,85	6,12
Ósseo	X	[2,4-20,2]	5,91	3,84
Cardiovascular	4	[66,8-70,6]	68,74	7,19
Cardiovascular	13	[44,4-52,1]	47,79	7,19

A utilização de SNPs no mapeamento foi capaz de aprimorar significativamente os resultados obtidos na etapa anterior, fazendo com que novas regiões candidatas fossem identificadas e eliminando regiões identificadas anteriormente, mostrando a importância e influência da quantidade de marcadores utilizada em qualquer mapeamento realizado.

4.7. Efeito dos genes modificadores

Utilizando os genótipos observados do marcador mais próximo à posição estimada de cada gene modificador, é possível analisar o efeito que esses genes apresentam sobre o seu respectivo fenótipo e se existem também interações com os outros genes modificadores.

A figura 41 mostra o efeito individual de cada um dos QTLs identificados na gravidade do fenótipo ósseo. Podemos perceber efeitos de característica dominante/recessiva nos cromossomos 3 e 6, enquanto que o QTL do cromossomo X apresenta um efeito que pode ser tanto de dominância quanto aditivo.

A existência de um QTL no cromossomo X era esperado para o fenótipo cardiovascular, uma vez que apenas neste sistemas existe diferença fenotípica entre machos e fêmeas. Porém é possível que o efeito desse QTL no fenótipo ósseo não tenha sido detectado porque a média fenotípica de machos e fêmeas sejam parecidas. Assim, animais homozigotos para o alelo B6 nesse QTL podem apresentar uma fenótipo mais ameno que as outras classes genotípicas.

Figura 41: Efeito de cada classe genotípica sobre o fenótipo ósseo no marcador mais próximo a posição de cada QTLs identificados. “UNC030244865”, “UNC060308920” e “JAX00177060” são os marcadores mais próximos dos QTLs dos cromossomo 3, 6 e X, respectivamente.

De maneira semelhante, é possível verificar a presença de efeitos de epistasia entre os três locos identificados (fig. 42). A identificação desses efeitos é importante para uma melhor compreensão dos mecanismos e da complexidade da determinação de cada fenótipo, e bem como para calcular o poder de explicação de cada loco identificado.

Assim, é possível notar uma interação, ou epistasia, entre os QTLs dos cromossomos 3 e 6, pois o efeito do QTL 6 de tornar o fenótipo mais grave quando em homozigose do alelo B6 é visível apenas quando o QTL 3 se encontra em homozigose para o alelo 129. Também nota-se que os animais heterozigotos do QTL 6 são levemente mais graves que os outros animais, algo que é mascarado na análise do efeito individual desse marcador.

Quanto aos outros dois pares, não foi possível notar a existências de nenhuma interação entre eles por não observar alteração entre os perfis observados - assim só existem os efeitos aditivos individuais de cada loco.

Para as regiões candidatas a conterem os genes modificadores relacionados ao fenótipo cardiovascular, foi identificado que o QTL no cromossomo 4 apresenta um efeito aditivo que leva a um aumento da espessura da parede da aorta a cada alelo 129 presente, porém apenas os animais homozigotos apresentam um fenótipo fora da normalidade. Além disso, o QTL no cromossomo 13 apresenta um efeito dominante/recessivo que faz com que apenas os animais homozigotos para o alelo B6 apresentem uma espessura superior às observadas em animais normais (fig. 43).

A análise de interação entre esses dois locos indicou a presença de epistasia entre eles, sendo o QTL do cromossomo 4 só é capaz de influenciar o fenótipo quando o QTL do cromossomo 13 está em homozigose para o alelo B6 (fig. 44).

Figura 42: Gráficos de interação par a par de cada QTL para o sistema ósseo. Os genótipos do primeiro QTL estão indicados na abscissa, o valor do fenótipo na ordenada, e os genótipos do segundo QTL são representados pelas cores azul (B6/B6), vermelha(129/B6) e preta (129/129). “UNC030244865”, “UNC060308920” e “JAX00177060” são os marcadores mais próximos da posição estimada dos QTLs dos cromossomos 3, 6 e X, respectivamente. A) QTLs cromossomos 3 e 6. B) 3 e X. C) 6 e X.

Figura 43: Efeito de cada classe genotípica sobre o fenótipo cardiovascular do marcador mais próximo à posição de cada QTL identificado. “UNC040316624” e “backupUNC130158494” são os marcadores mais próximos para o cromossomo 4 e o cromossomo 13, respectivamente.

Figura 44: Gráficos de interação par a par de cada QTL, os genótipos do primeiro QTL estão indicados na abcissa, o valor do fenótipo na ordenada, e os genótipos do segundo QTL são representados pelas cores azul (B6/B6), vermelha (129/B6) e preta (129/129). “UNC040316624” e “backupUNC130158494” são os marcadores mais próximos para o cromossomo 4 e o cromossomo 13, respectivamente.

4.8. Variabilidade explicada

Apesar de identificados a existência e o tipo de efeito que cada um dos QTLs e suas interações apresentam, é importante quantificar esse efeito e avaliar o quanto da variabilidade de cada fenótipo é possível explicar com os QTLs identificados.

O estabelecimento desses modelos ainda auxiliam na determinação de quais efeitos de cada QTL são importantes para a determinação do fenótipo, e se existem QTLs adicionais a serem identificados.

Cada modelo consiste no uso de regressão linear para estimar os parâmetros aditivo (a) e dominante (d) de cada QTL e das interações entre cada QTL, 4 parâmetros para cada par interagindo. Assim é possível determinar quais efeitos são importantes para a compreensão de como cada sistema é afetado pelo seu conjunto de QTLs, ou seja, quais efeitos são importantes para a arquitetura genética de cada fenótipo.

Os modelos iniciais para cada fenótipo foram derivados utilizando as informações obtidas dos efeitos individuais e de interação de cada QTL. Assim o modelo referente ao fenótipo ósseo consiste de 3 QTLs aditivos e uma interação (cromossomos 3 e 6), enquanto que o modelo do fenótipo cardiovascular é composto pelos dois QTLs principais, a interação entre eles e o sexo de cada animal como variável aditiva e interativa com os QTLs.

O modelo inicial do fenótipo ósseo atinge um nível de significância referente a distribuição F de 10^{-6} , porém este modelo não é significativamente diferente do modelo sem os efeitos principais do QTL do cromossomo X, tornando esse modelo reduzido mais confiável, apresentando um p-valor de $7,4 * 10^{-8}$ e explicando 47,3% de toda a variabilidade observada entre os animais.

Os coeficientes estimados para cada termo do modelo, assim como os respectivos desvios padrões e valores da estatística *t-student*, estão indicados na tabela 6. As interações entre cada termo dos QTLs estão indicado da seguinte maneira: termo 1 * termo 2.

Tabela 6: Termos de regressão estimados para o fenótipo ósseo.

Termo	Estimativa	Desvio Padrão	T-score
Intercepto	84,16	0,61	138,7
Chr3 aditivo	-2,54	1,94	-1,31
Chr3 dominante	-5,4	2,68	-2,02
Chr6 aditivo	-6,85	2,37	-2,89
Chr6 dominante	2,16	1,83	1,18
Chr3 aditivo * Chr6 aditivo	10,88	3,79	2,87
Chr3 dominante * Chr6 aditivo	12,32	3,99	3,09
Chr3 aditivo * Chr6 dominante	10,23	3,97	2,58
Chr3 dominante* Chr6 dominante	16,59	5,51	3,01

Devido à quantidade de animais utilizada para estimar esses valores, 82 animais, é necessário utilizar uma distribuição T-student com 81 graus de liberdade para estabelecer quais termos são significativos para a construção do modelo. Assim, apenas os termos que obtiveram $|t\text{-score}| \geq 1,99$ foram mantidos no modelo final do fenótipo ósseo (eq. 4).

$$KR = 84,16 - 5,4 * X_{\text{Chr3,d}} - 6,85 * X_{\text{Chr6,a}} + 10,88 * X_{\text{Chr3,a}} * X_{\text{Chr6,a}} + 12,32 * X_{\text{Chr3,d}} * X_{\text{Chr6,a}} + 10,23 * X_{\text{Chr3,a}} * X_{\text{Chr6,d}} + 16,59 * X_{\text{Chr3,d}} * X_{\text{Chr6,d}}$$

A divergência entre a herdabilidade calculada e a variabilidade explicada pelo modelo proposto acima pode indicar que exista algum termo no modelo completo que não seja realmente necessário para se estimar o fenótipo, porém essa diferença deve ter pouco ou nenhum sentido biológico.

Para a modelagem do fenótipo cardiovascular foram utilizados os efeitos principais dos QTLs nos cromossomos 4 e 13, a interação entre eles, assim como o efeito aditivo do sexo como covariável e as interações desta com os QTLs, totalizando um como composto por 14 termos a serem estimados.

Como ocorrido com o fenótipo ósseo, existe um modelo reduzido que não é significativamente diferente do modelo completo. Entretanto os p-valores obtidos são muito próximos ao limiar de 5%, e como existe uma clara indicação de que existe interação entre os dois QLTs identificados optou-se por manter todos os termos e realizar a modelagem a partir do modelo completo.

Assim, o modelo do fenótipo cardiovascular possui um p-valor de 10^{-6} e é capaz de explicar 55% da variabilidade observada no animais. Os coeficiente se encontra na tabela 7 e o limiar de significância destes foi calculado referente à distribuição t-student com 71 graus de liberdade, devido a existência de animais não fenotipados, e estabelecido em $|t\text{-score}| \geq 1,99$. A equação 5 mostra o modelo final referente a esse fenótipo.

Tabela 7: Termos de regressão estimados para o fenótipo cardiovascular.

Termo	Estimativa	Desvio Padrão	T-score
Intercepto	91,09	6,44	14,14
Sexo	-20,02	7,61	-2,63
Chr4 aditivo	-41,89	17,5	-2,4
Chr4 dominante	-13,3	26,61	-0,5
Chr13 aditivo	41,91	23,43	1,79
Chr13 dominante	71,85	20,97	3,42
Chr4 aditivo * Chr13 aditivo	-56,072	29,85	-1,88
Chr4 dominante * Chr13 aditivo	-84,56	31,14	-2,72
Chr4 aditivo * Chr13 dominante	-79,02	33,02	-2,39
Chr4 aditivo * Chr13 dominante	-69,85	39,64	-1,76
Chr4 aditivo * Sexo	73,16	14,02	5,22
Chr4 dominante * Sexo	66,4	25,47	2,61
Chr13 aditivo * Sexo	-4,86	21,99	-0,22
Chr13 dominante * Sexo	-44,17	21,91	-2,02

$$\begin{aligned} \text{AWT } (\mu\text{m}) = & 91,09 - 20,02 * \text{Sexo} - 41,89 * X_{\text{Chr4,a}} + 71,85 * X_{\text{Chr13,d}} - \\ & 84,56 * X_{\text{Chr4,d}} * X_{\text{Chr13,a}} - 79,02 * X_{\text{Chr4,a}} * X_{\text{Chr13,d}} + 73,16 * X_{\text{Chr4,a}} * \text{Sexo} + \\ & 66,4 * X_{\text{Chr4,d}} * \text{Sexo} - 44,17 * X_{\text{Chr13,d}} * \text{Sexo} \end{aligned}$$

Novamente é encontrada uma diferença entre as herdabilidades calculadas, mas nesse caso a variância explicada pelo modelo é inferior à herdabilidade calculada. Essa perda de herdabilidade indica que, possivelmente, não foram encontrado todos os QTLs envolvidos na determinação do fenótipo cardiovascular, ou que alguma interação entre os QTLs observados tenha passado despercebida, ou ainda que a herdabilidade restante seja devida a alterações epigenéticas.

O aumento do número de animais genotipados, a análise de interação entre cada par dos utilizados que permita uma seleção apropriada dos termos a constituírem o modelo, ou ainda a transferência do alelo $mg\Delta^{loxPneo}$ para outros fundos genéticos e um novo processo de mapeamento realizado nessas novas linhagens poderão solucionar esse problema de perda de herdabilidade e propiciar uma nova visão da determinação da SMF.

4.9. Regiões candidatas

Devido ao tamanho dos intervalos já é possível visualizar os genes existentes em cada um delas e começar a procura pelos genes candidatos a modificadores do fenótipo. O QTL do cromossomo 3 se estende por 12 cM e contém 139 genes e miRNA (anexo 6). Uma análise preliminar da região foi capaz de identificar genes codificantes de fatores de transcrição da RNA polimerase II (*Gtf2b*, *Nfkb1*, *Lef1* e *Pitx2*); componentes da matriz extracelular (*Col25a1* e *Npmt*); componentes estruturais do ribossomo (*Rpl34* e *Gar1*); componentes do citoesqueleto (*Pdlim5*); genes envolvidos nas vias Wnt (*Dkk2* e *Defl*), TGF- β (*Bmpr1b*) e *Egf* ; além dos fatores de transcrição também foram identificados genes envolvidos em processos de metilação (*Rg9mtd2*), produção de histona (*H2af3*) e chaperona (*Dnajb14*).

A região identificada no cromossomo 4 se expande ao longo de 14 cM e contém 150 genes e miRNAs (anexo 7). Os genes que mais chamam atenção são mais fatores de transcrição da RNA polimerase II (*Runx3*, *Zfp46*, *E2f2* e *Grhl3*), sendo que *Runx3* está envolvido na diferenciação sexual, e dos sistemas esquelético e nervoso; componentes estruturais do ribossomo (*Rpl11*); componentes da matriz extracelular (*Hspg2*); fator de splicing (*Syf2*); componentes do citoesqueleto de actina (*Myom3*) e microtubulos (*Kif17*); assim como membros das vias Wnt (*Wnt4*) e de angiogenese (*Ephb2*).

A região no cromossomo 6 tem 8 cM, compreendendo 70 genes (anexo 8), dos quais 3 são fatores de transcrição da RNA polimerase II (*Foxp1*, *Mitf* e *Rybp*) - *Foxp1* está envolvido nas vias de Wnt, TGF- β , Fork e no estabelecimento dos eixos do organismo; 2 componentes estruturais do citoesqueleto de actina (*Prickl2*, *Frmd4b* e *Kbtb8*); 3 genes da mesma família envolvidos na angiogenese e desenvolvimento do sistema nervoso (*Cntn3*, *Cntn4* e *Cntn6*); ubiquitina (*Ube3*) e proteases envolvidas na organização do tecido conjuntivo (*Adamts9*).

No cromossomo 13 a região de interesse se estende por 8 cM, possuindo 121 genes (anexo 9). Igual às outras regiões, foram encontrados genes ligados à RNA polimerase (*Otp*, *Ssbp2* e *Ccnh*), mas também a uma RNA polimerase extensora da cauda poli-A (*Papd4*) e uma DNA polimerase (*Polk*); componente estrutural do ribossomo (*Mrps27*); componentes da via do TGF- β e Smads (*Foxd1* e *Zfyve16*), e Egf (*Edil3*); componentes do citoesqueleto de actina (*Enc1* e *Fchv2*) e uma chaperona envolvida no enovelamento da tubulina (*Tbca*).

A última região candidata identificada, do cromossomo X, possui 18 cM de extensão e compreende 398 genes e miRNAs (anexo 10). Também foram identificados genes de fatores de transcrição da RNA polimerase II (*Ssxb6*, *Ssxb12*, *Ssxa1*, *Med14*, *Pqbp1*, *Tte3* e *Phf16*); componentes do citoesqueleto de actina (*Prickl3*, *Was* e *Syn1*); ubiquitinas (*Usp27x* e *Uba1*); um gene envolvido na maquinaria de splicing (*Rbm3*); genes responsáveis pela

produção de histonas (histona H2A) e sua acetilação (*Hdac6*); um gene envolvido na maquinaria de endereçamento protéico à membrana (*Porcn*); assim como genes envolvidos nas vias de sinalização de TGF- β (*Bmp15* e *Foxp3*) e Wnt (*Mycs*).

Os genes citados de cada região são interessantes por estarem diretamente relacionados com alterações dos respectivos sistemas, como genes da via Wnt para o sistema esquelético por estarem relacionado com regeneração óssea (Chen e Alman, 2009); e de gene envolvidos no processo de angiogênese para o sistema cardiovascular; ou ainda genes envolvidos na organização da matriz extracelular e do tecido conjuntivo como um todo.

Também são de interesses genes da via do TGF- β e das Smads uma vez que a redução de TGF- β ativo foi vinculada com a melhora fenotípica da SMF em camundongos (Neptune e cols, 2003).

Genes que estejam de alguma forma envolvidos com alteração de expressão, sejam eles fatores de transcrição, componentes estruturais do ribossomo ou relacionados a maquinaria de splicing, também são interessantes pois podem levar a diferença de expressão gênica entre as duas linhagens estudadas e assim modular o fenótipo observado.

Da mesma maneira, foram considerados interessantes genes que pudessem estar envolvidos na regulação gênica através de alterações de metilação ou acetilação de histonas, e que assim pudessem modular a expressão gênica por meio de mecanismos epigenéticos (Degner e cols. 2012).

Genes envolvidos no processo de enovelamento protéico, endereçamento e controle do cito esqueleto foram considerados interessantes pois o modelo animal utilizado apresenta deposição de fibrilina-1 mutante no interior da célula (Lima e cols, 2010). Assim, alterações que favoreçam o correto enovelamento e endereçamento da proteína mutante e o seu transporte, ou ainda que levem os monômeros de fibrilina-1 alterados a serem degradados (ubiquitinas), podem ser de grande importância para a determinação fenotípica da SMF.

Como a quantidade de genes e possíveis mecanismos de ação é elevada, não é possível estipular quais sejam os genes responsáveis pelos QTLs identificados, pois é quase certo que a interação entre os QTLs de cada sistema tenha grande efeito na determinação fenotípica e compreensão da arquitetura genética e mecanismos por trás da SMF.

Entretanto, é possível notar que existe uma certa concordância entre os tipos de genes observados nas 5 regiões identificadas, pois quase todas apresentam fatores de transcrição afetando a RNA polimerase II, componentes do citoesqueleto celular, membros da via de TGF- β , da via Wnt, e genes envolvidos em processos de silenciamento gênico por mecanismos epigenéticos, fornecendo assim uma maior credibilidade as regiões identificadas.

4.10. Microarray de Expressão

Uma alternativa ao processo de mapeamento para levantar genes candidatos a modificadores de fenótipo seria utilizar microarrays de expressão para identificar quais genes estariam diferencialmente expressos entre os animais com alterações leves e animais com alterações graves na linhagem 129/Sv. Os genes identificados entre esses animais ou seriam uma resposta do organismo para se adaptar a uma nova condição ou estariam causando essa nova condição, possibilitando assim que se reduzisse a quantidade de genes candidatos que precisariam ser analisados.

Para isso foram coletadas amostras de pulmão, devido a sua facilidade de processamento e abundância de material quando comparado a aorta e aos ossos do camundongo, de animais selvagens e de animais KO com alterações extremas para esse sistema. Os RNAs dos pulmões foram isolados e utilizados nos experimentos de microarray.

Uma vez obtidos os resultados das intensidades de cada array, foi necessário realizar uma série de procedimentos de qualidade e normalizações antes de se obter os resultados propriamente ditos. A primeira etapa da análise consiste na identificação daqueles arrays com elevado coeficientes de variância (CV), o que indica problemas ou variação na hibridação das sondas e consequentemente nos níveis de expressão aferidos por esse array. Como pode-se ver pela fig. 45, de todos os 16 arrays feitos apenas o 5°, 9° e 13° array apresentam CV elevados, assim foram excluídos para não prejudicar os outros que obtiveram dados de melhor qualidade.

Figura 45: Análise de qualidade de cada array realizado. CV% maiores que 10% indicam dados de qualidade duvidosa (pontilhado azul).

Depois da seleção dos arrays de boa qualidade, foi necessário eliminar a variação de “background” de cada array e a variação da intensidade do sinal entre arrays, a qual pode ser causada por diferenças técnicas entre cada lâmina.

Para a normalização do background foi utilizado um método que estabelece uma intensidade mínima de sinal para o background, onde os sinais inferiores são elevadas até o mínimo, e acresce uma constante a cada valor para que as menores intensidades tendam a zero quando submetidas a transformação logarítmica. A normalização entre arrays consiste na equalização da distribuição da intensidade de sinal em cada um deles. Depois desse processo de normalização, o sinal de cada sonda é definido através da diferença entre o sinal inicial de cada sonda e do sinal de fundo. (fig. 46).

Figura 46: Dados antes e após normalizações, respectivamente. A e B) Gráficos RLE (logaritmo da expressão relativa) de todas as sondas para cada array (normalização interna de cada array) ;C e D) distribuição da intensidade do sinal (normalização entre arrays); E e F) Gráficos MVA (diferença do log da intensidade X média dos log da intensidade) para cada uma das sondas, exemplo da correção que cada array sofre ao longo do processo.

Essas sondas foram filtradas, retirando as sondas controle, aquelas cujos os genes não apresentam nenhum termo (função molecular, processo biológico ou componente celular) associado no banco de dados Gene Ontology, e aquelas que não possuíam um código ENTREZ. Caso um gene seja representado por mais de uma sonda, apenas a de maior variação é mantida nos próximos passos da análise.

Além da remoção das sondas que não estão localizadas em genes e/ou genes que sem função conhecida, também foi realizada uma filtragem não-específica dos sinais obtidos para cada sonda. Isso só foi possível porque não existe nenhuma associação entre o desvio padrão de cada sonda e o nível de expressão da mesma (fig. 47).

Figura 47: Desvio padrão x média de cada sonda para as três análises realizadas. O aumento/diminuição do desvio em função na média indicaria a necessidade de outras correções antes de se iniciar a análise.

Essa filtragem consiste na eliminação das sondas pouco variáveis entre os dois níveis comparados, isso diminui a quantidade de testes a serem feitos diminuindo o trabalho computacional necessário e a correção que tem que ser feita aos níveis de significância (fig. 48).

Figura 48: Distribuição dos desvio padrões para cada uma das comparações feitas. Aquelas com desvios muito pequenos (a esquerda das linhas pontilhadas) foram excluídos na filtragem não específica.

Assim, após eliminadas as fontes de variação não biológica, os controles e as sondas de pouca variação que não seriam informativas na identificação dos genes diferencialmente expressos, realizou-se a comparação dos níveis dois a dois através de uma série de teste-t que fornecem nível de significância e a diferença de expressão para cada sonda/gene questionado. Com esses dados foram montados gráficos vulcão que mensuram tanto o significado estatístico (p-valor) quanto o significado biológico (“fold-change” ou FC) dos genes deferencialmente expressos (figs.49 a 51).

Os genes de interesse são aqueles que estão acima do nível de significância estatístico (linha horizontal vermelha) e aqueles externos aos limiares biológicos (linhas verticais azuis). O nível de significância foi determinado através das correções de

múltiplos testes para que a taxa de falsos negativos (False Discovery Rate, FRD) fosse de 5%, método conhecido por TSBH e estabelecido por Benjamini & Hochber em 2006.

Ao todo foram identificados 1120 genes diferencialmente expressos entre animais selvagens e KO leves; 2524 genes entre animais selvagens e KO graves; e apenas 25 genes entre KO leves e KO graves.

Figura 49: Gráfico Vulcão WT x Leve. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.

Figura 50: Gráfico Vulcão WT x Grave. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.

Figura 51: Gráfico Vulcão Leve x Grave. Linha verticais azuis determinam o limiar de significância biológica e a linha horizontal vermelha o limiar de significância estatístico.

Como era esperado, a quantidades de genes diferencialmente expressos é maior quando comparamos animais selvagens e animais com fenótipo grave. Enquanto que a comparação entre animais graves e leves é a que produz a menor quantidade de genes diferencialmente expressos, esperava-se que a quantidade de genes identificados nesta fosse menor em relação a quantidade de genes nas comparações dos animais selvagens e qualquer classe de animal KO, porem um quantidade tão baixa que nem a encontrada foi inesperado.

Possivelmente a variação de expressão gênica já observada entre os animais 129/Sv heterozigotos não se restrinja ao gene *Fbn1*, o que faz com que os fatores de correção de múltiplos testes seja mais rigoroso nessa situação do que nas outras analisadas. Outra possibilidade seria que a diferença exigida de expressão para que os genes sejam considerados de interesse está muito grande e isso esteja excluindo genes relevantes para a análise.

Uma vez identificados os genes de interesse de cada comparação é possível utiliza-los para avaliar a distancia entre cada animal e agrupa-los de acordo com essa medida (fig. 52 a 54). Foi utilizado um método de agrupamento não supervisionado, ou seja, só é utilizada a distância estimada entre as amostras para se atribuir a qual grupo de estudo cada amostra pertencia.

Figura 52: Heatmap para os genes diferencialmente expressos entre os animais leves e selvagens. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.

Figura 53: Heatmap para os genes diferencialmente expressos entre os animais graves e selvagens. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.

Figura 54: Heatmap para os genes diferencialmente expressos entre os animais leves e graves. Os histogramas lateral e superior indicam o quão próximas as amostras ou genes são um do outro. As cores indicam o sentido e a intensidade da diferença de expressão de cada amostra em relação a média.

O agrupamento das amostras para as comparações leves contra graves e leves contra selvagens foi capaz de distinguir as duas classes fenotípicas, porém uma clara distinção de genes com expressão aumentada ou reduzida só pode ser visualizada na comparação entre animais leves e graves.

A falha em agrupar corretamente as amostras na comparação entre animais selvagens e graves, e a divergência no padrão de expressão de certas amostras com o que era esperado para a sua categoria, mostra que existe todo um espectro de variação para a expressão de cada gene, pois existem animais com alterações graves cujo o perfil de expressão é mais próximo do perfil de expressão de animais selvagens do que de animais KO (fig. 53), ou ainda animais que apresentem um perfil intermediário entre as classes (fig. 52).

Como já foi sugerido que a quantidade de fibrilina-1 disponível pudesse influenciar na gravidade de manifestação da SMF (Hutchinson e cols., 2003), também é possível que variações no nível de expressão de outros genes modificadores do fenótipo também sejam importantes para a manifestação fenotípica. Assim, além dos possíveis efeitos já descritos anteriormente para a função dos genes identificados, a simples expressão diferencial de um dos alelos talvez seja o suficiente para alterar o fenótipo.

Além de identificar os genes diferencialmente expressos entre animais leves e graves, que poderiam ser genes modificadores do fenótipo, é possível utilizar a lista de genes identificados nas outras duas comparações para realizar uma análise de quais vias se encontram enriquecidas ou empobrecidas nesse conjunto de genes identificados (GSEA).

As vias que apresentaram um aumento ou redução de genes diferencialmente expressos entre as comparações das duas classes de animais KO contra os animais selvagens são interessantes porque ou são vias responsáveis pelo aparecimento do fenótipo pulmonar ou são vias que estão respondendo as alterações causadas pelo alelo mutante do modelo.

Ao todo foram identificadas 46 vias alteradas e que são possíveis detentoras de genes modificadores do fenótipo pulmonar (anexo 11). Dentre elas se encontram as vias de sinalização de TGF- β , de Shh e de Wnt que são vias envolvidas em um ou mais fenótipos da SMF; nas vias de processamento de proteínas no retículo endoplasmático, de endocitose, do proteassoma, regulação do citoesqueleto de actina e de interação de receptores de membrana

com proteínas da matriz extracelular, que são vias envolvidas com o correto enovelamento de proteínas e degradação daquelas que não assumem a sua correta conformação além do processo de internalização de substâncias do meio extracelular e transmissão de sinais para o interior da célula e organização estrutural da célula; ainda envolvendo organização celular e comunicação entre células é possível identificar alterações nas vias responsáveis pelas junções comunicantes (“GAP junctions”) e junções oclusivas (“Tight junctions”); uma última via que chama atenção é a via que controla apoptose celular e que se encontra alterada nos animais KO.

4.11. Validação dos resultados microarray expressão

Como o experimento de microarray de expressão analisa uma grande quantidade de dados ao mesmo tempo, a sua sensibilidade de detecção é influenciada pela quantidade de comparações realizadas, necessitando assim que as diferenças encontradas sejam validadas por técnicas mais diretas e sensíveis.

Dentre a lista de 25 genes diferencialmente expressos identificados entre as duas categorias de animais KO foram escolhidos, baseando na função descrita para cada um, 11 genes para ter a sua expressão validada através de PCR em tempo real (*real-time PCR*): *Scgb3a1*, que codifica uma possível citocina que inibe fatores de crescimento; *Edn1*, que codifica um vasoconstritor envolvido na manutenção do tônus vascular e que é capaz de potencializar a ação de fatores de crescimento; *Fcna*, pela proteína ser capaz de se ligar a elastina e apresentar domínios similares ao colágeno; *Prrx2*, por estar envolvido no processo de cicatrização da epiderme e proliferação celular; *SI00b* que é capaz de inibir a polimerização de microtubulos; *Adamts13*, cuja família gênica está envolvida com a

gravidade do fenótipo da SMF (Saito e cols, 2011); *Cpb2* , que é capaz de regular a atividade de peptídeos livres na circulação sanguínea; *Cd109* , ligante de TGF- β que inativa a sua sinalização; *Syt6* , que codifica uma proteína de membrana que está envolvida nos processos de endocitose e fusão de membrana, principalmente nas regiões de sinapse; *Eml6* , que é capaz de alterar a dinâmica dos microtubulos; e *Fbln7* , que codifica uma proteína de adesão que interage com outras moléculas da matriz extracelular.

Apesar de desenhados para funcionar nas condições necessária para a reação de *real-time* PCR, apenas os primers de 2 dos 11 genes apresentavam amplificações limpas e com banda na altura esperada. Assim, apenas as expressões dos genes *Edn1* e *Prrx2* foram avaliadas.

Para encontrar os genes endógenos mais adequados para cada um desses primers foram testados três genes endógenos (*Atcb*, *Rpl13a* e *Gapdh*) que já eram usados pelo grupo e que funcionavam bem nas condições necessárias para o *real-time* PCR. Os endógenos que melhor se adequam aos genes de interesse são o *Gapdh* para o gene *Edn1* , e o *Actb* para o gene *Prrx2*. Como pode ser visto na figura 55, eles são os genes que apresentam a menor variação entre os diferentes pontos da diluição seriada realizada, ou seja, são os endógenos que produzem uma equação de reta com coeficiente angular mais próximo de zero.

Uma vez selecionados os genes alvos e os seus devidos endógenos, foram coletadas mais amostras de pulmão de animais com fenótipos leves e graves para aumentar o universo amostral utilizado no microarray de expressão e assim confirmar os resultados observados. Ao todo foram utilizados 12 animais, 4 representantes de cada classe, para a validação dos resultados. Os níveis de expressão foram estimados utilizando o nível apresentado por um animal selvagem escolhido aleatoriamente (3591).

As figuras 56 e 57 indicam os resultados obtidos para os genes *Edn1* e *Prrx2*, respectivamente.

Figura 55: Gráficos de padronização dos genes de interesse (*Prrx2* e *Edn1*) contra os possíveis endógenos (*Rpl13a*, *Gapdh* e *Actb*). Abscissa $\log_{10}[\text{RNA}]$ e ordenada diferença entre Ct gene alvo e Ct endógeno. Cada gráfico contém a linha de tendência, a equação da reta e o coeficiente de explicação (R^2).

Figura 56: Expressão relativa para o gene *Edn1* em animais 129/Sv. A) Animais agrupados de acordo com a sua classe fenotípica. B) Expressão individual de cada animal.

Figura 57: Expressão relativa para o gene *Prrx2* em animais 129/Sv. A) animais agrupados de acordo com a sua classe fenotípica. B) Expressão individual de cada animal.

O resultados esperado era que os animais com fenótipo grave apresentassem um aumento de expressão de *Prrx2* e uma expressão reduzida de *Edn1* em relação os níveis de expressão dos animais com fenótipo leve.

Apesar de observamos esse comportamento para o gene *Edn1*, não é possível afirmar que haja diferença de expressão entre as duas categorias fenotípicas para ambos os genes estudados, uma vez que os desvios padrões são grandes e há muita sobreposição entre eles.

Essa grande variação observada na expressão média de cada classe indica que a expressão de cada animal é diferente uma da outra (figuras 56-B e 57-B). Apesar dessa variação entre amostras com fenótipos semelhantes impossibilitar a validação dos resultados obtidos com o microarray de expressão, ela reforça a hipótese de que exista um espectro contínuo de expressão entre os animais KO e que ela estava relacionado com a gravidade fenotípica.

Como se tratam de animais isogênicos, a fonte de variação que venha a justificar a diferença de expressão entre esses animais pode estar em fatores epigenéticos ou ainda ser uma variação normal da linhagem que ainda não está completamente caracterizada.

5. Discussão

À medida que as correlações genótipo-fenótipo de doenças mendelianas são definidas e os seus genes causadores são identificados, se torna cada vez mais interessante o estudo de doenças que não são puramente mendelianas, ou seja, que apresentam fatores de complicação tais como penetrância incompleta e expressividade variável, como também o estudo de doenças complexas.

A junção das técnicas clássicas de mapeamento com a utilização de plataformas de obtenção de dados em larga escala, como microarrays de SNP e de expressão, e a capacidade de se criar modelos animais com mutações específicas capaz de mimetizar as alterações observadas em humanos permite que sejam identificados os genes modificadores envolvidos nas doenças que não são puramente mendelianas e também o conjunto de genes envolvidos nas doenças complexas.

O primeiro trabalho que utilizou as técnicas de mapeamento descritas acima e que demonstrou a importância de genes modificadores em camundongos foi publicado em 1996 por Rozmahel e colaboradores. Eles criaram um modelo murino com deficiência de proteína Cftr para simular as características fenotípicas da fibrose cística. Mutações no gene CFTR (*cystic fibrosis transmembrane conductance regulator gene*), que codifica um canal de cloreto ligado à membrana plasmática, causam a secreção de sais em vários tecidos. Camundongos mutantes homozigotos morrem logo após o nascimento devido ao acúmulo de muco no intestino. Entretanto, em algumas linhagens isogênicas, os camundongos Cftr deficientes apresentavam um maior tempo de vida, indicando que genes modificadores estão envolvidos na modulação da doença. Estudos de ligação mostraram um único locus, com efeito aditivo, como responsável pela maior parte da variação clínica e viabilidade dos animais. Assim, camundongos deficientes Cftr possuiriam uma maior expectativa de vida de acordo com o

número de alelos do gene modificador, pois este diminui a quantidade de muco no intestino e restabelece relativamente a eletrofisiologia normal da membrana.

O trabalho publicado por Rozmahel utilizou um mapeamento baseado em microssatélites e, apesar de identificar a existência de um gene modificador na região proximal do cromossomo 7, não foi possível chegar a apenas um gene candidato a modular o fenótipo nesse modelo.

Em um estudo recente publicado por Faber e colaboradores em 2009, foram utilizados tanto técnicas de mapeamento com SNPs e estudos de associação (sendo que estes procuram pela associação matemática entre duas variáveis quaisquer, neste caso o genótipo de cada SNP e os fenótipos de interesse, sem levar em consideração o desequilíbrio de ligação), quanto análise do perfil de expressão, modelagem de causalidade e sequenciamento para identificar QTLs, eQTLs (QTLs que influenciam expressão gênica) e inferir os possíveis genes causais que estariam influenciando a densidade de massa óssea entre três linhagens murinas. Essa abordagem integrada entre várias técnicas e bancos de dados conseguiu identificar nove QTLs, sendo 5 já descritos e 4 novos, cobrindo ao todo 300 cM do genoma murino, levando a uma lista de 6000 genes candidatos posicionais, e ainda reduzir essa lista para poucos genes candidatos funcionais, enquanto que 5 desses já haviam sido correlacionados com alterações na densidade óssea. Contudo, a combinação desses dois conjuntos de informações não é algo feito trivialmente (Nica e cols, 2008).

Dessa forma, a integração das técnicas de mapeamento com análise de perfil de expressão e sequenciamento são extremamente eficientes para identificar sejam QTLs ou genes modificadores em qualquer população, seja ela humana ou de algum modelo de estudo, que apresente uma variação fenotípica de interesse.

5.1. Genes modificadores e o modelo $mg\Delta^{loxPneo}$

Com a utilização de modelos animais é cada vez mais possível que os mecanismos e vias envolvidas na determinação do grau de acometimento de doenças genéticas sejam descobertas e o seu papel esclarecido. O modelo $mg\Delta^{loxPneo}$ é uma representação do quadro clínico observados em humanos (Lima, 2010), mimetizando também a variabilidade interfamiliar devido a variação fenotípica entre as linhas 129 e B6.

A mistura das duas linhagens nos animais F2 permitiu ver que o modelo é ainda mais semelhante ao quadro clínico humano, uma vez que podem ser encontrados animais com todas as combinações possíveis de gravidade fenotípica nos três sistemas estudados.

Apesar de ser um estudo utilizando apenas 82 animais, foi possível encontrar ligação significativa entre o fenótipo ósseo e uma região do cromossomo 6, indicando que o gene modificador existente nessa região deve ser capaz de controlar boa parte da variabilidade observada.

A interação entre o gene modificador do cromossomo 6 e o possível gene modificador do cromossomo 3 confirmam a hipótese levantada nas observações fenotípicas dos animais F2 de que ambas as linhagens possuem alelos de suscetibilidade e de proteção. A possível existência de um gene modificador no cromossomo X pode ser questionada uma vez que o modelo de regressão completo com as três regiões identificadas não é significativamente diferente do modelo matemático composto apenas pelos cromossomo 3 e 6 e suas interações. Por se tratar de um modelo mais simples e ser capaz de explicar toda a variabilidade para o sistema esquelético, este tende a ser o modelo mais plausível.

Encontrar ligação do cromossomo X, logo do sexo do animal, mesmo que apenas sugestiva com o fenótipo ósseo e não com o fenótipo cardiovascular é algo inusitado e que demanda mais atenção.

Os genes modificadores nos cromossomo 4 e 13, juntamente com o sexo dos animais, conseguem explicar 55% de toda a variabilidade do fenótipo cardiovascular, entretanto ainda existem 21% que não são explicadas. Essa perda de herdabilidade pode estar correlacionada com a incapacidade de detectar todos os genes envolvidos na determinação do fenótipo, principalmente aqueles que explicam uma parcela muito pequena da variabilidade fenotípica (Mackay, 2001) ou ainda com alguma interação entre locos que não tenha sido considerada (Eichler e cols. 2010; Fichler e cols. 2010; Makowsky e cols. 2011 e Zuk e cols. 2012).

Assim como a herdabilidade que falta ser explicada para o fenótipo cardiovascular, os genes que estiverem envolvidos no fenótipo pulmonar devem ter um poder explicativo inferior à capacidade de detecção do experimento, e por isso não foram detectados.

Mesmo não conseguindo identificar os genes modificadores em cada uma das regiões, é interessante notar que existem genes com funções similares nas 5 regiões, tornando ainda mais complexo o processo de escolher os genes candidatos, uma vez que as alterações podem ocorrer por várias vias diferentes e diferentes pontos delas.

Como o modelo $mg\Delta^{loxPneo}$ gera um monômero de fibrilina-1 truncado e que tende a acumular no meio intracelular, é possível que variações alélicas em proteínas que estejam envolvidas nos processos de enovelamento, exocitose ou ainda degradação possam levar a uma melhora ou piora dos fenótipos da SMF. Um exemplo seriam as proteínas do tipo chaperonas que estão envolvidas no enovelamento de proteínas, e inibem a agregação de proteínas que não assumiram a sua correta conformação ou que foram denaturadas para que possam ser processadas novamente. Assim são possíveis genes modificadores de fenótipos humanos decorrentes de instabilidade ou de agregação protéica (slavotenik e col., 2003).

Devido à diferença observada entre animais heterozigotos da linhagem 129, é possível que a simples variação de expressão de certos genes, e não uma outra variante alélica como se imagina, seja suficiente para alterar a manifestação fenotípica da SMF. Assim fatores de transcrição, ou proteínas envolvidas na maquinaria de inativação, formas variantes de histonas ou proteínas envolvidas em acetilação destas ou ainda de metiltransferases de ilhas CpG, podem ser importantes para modular as manifestações clínicas da SMF (Degner, 2012).

Além desses genes que ainda não estão associados com a SMF, existem genes que já foram envolvidos com a melhora do quadro clínico da SMF, tais como os genes envolvidos na via de TGF- β e da família Adamtsl. Entretanto estes genes/vias estão envolvidos com a SMF de maneiras diferentes. Enquanto a família Adamtsl está envolvida na biogênese do tecido conjuntivo, mais especificamente das microfibras (Gabriel e cols., 2012 e Saito e cols., 2011) e podem estar levando ao aparecimento do fenótipo devido a alteração a estrutura física da fibra elástica, TGF- β é regulado pela fibrilina, e o excesso de sua atividade leva às manifestações fenotípicas da SMF através de sinalização pela via das Smads ou Erk-1/2 (Neptune, 2003; Kaartinen, 2003 e Ng, 2004).

5.2. Predição fenotípica

A principal vantagem de se conhecer os genes modificadores envolvidos na arquitetura génica de um determinado traço é conseguir prever o valor que este traço irá assumir baseado unicamente no genoma de cada animal.

O problema desse tipo de abordagem é não levar em consideração os diferentes fatores ambientais que podem influenciar o fenótipo de interesse, os quais muitas vezes são desconhecido e a sua influencia pode mascarar completamente ou ainda alterar os efeitos dos

genes modificadores ou QTLs. Assim além de ser importante a identificação dos genes modificadores é essencial saber o comportamento destes em vários ambientes e durante as várias etapas de desenvolvimento do organismo (Griffiths, 2006).

Entretanto, nos estudos envolvendo animais de laboratório que são mantidos e criados em biotérios, essas condições ambientais são controladas e mantidas estáveis por todo o desenvolvimento de cada um dos animais. Logo, tais fatores ambientais são estáveis e não se alteram de um animal para outro, permitindo que não sejam fatores importantes para a modelagem dos fenótipos sobre estudo.

As equações derivadas para os fenótipos ósseo e cardiovascular são específicas para esse estudo, uma vez que os genes modificadores observados são apenas entre as linhagens 129 e B6, e que ainda precisam ser confirmadas a quantidade total de genes modificadores envolvidos e a posição exata destes para que os valores de cada monômio seja estimado corretamente.

Como o processo de mapeamento foi realizado utilizando uma genotipagem seletiva, os genótipos nas posições de interesse dos animais com manifestações fenotípicas não extremas tiveram que ser estimados através de múltiplas imputações, procedimento similar ao utilizado no mapeamento realizado com microssatélites.

Com a genotipagem desses outros animais é possível verificar se a diferença entre o fenótipo real e o valor estimado pelas formulas está controlada, ou seja, se respeita uma distribuição normal com média 0 (zero) e variância σ^2 , ou se existem desvios muito grandes entre os dois valores, indicando que esses parâmetros não são adequados.

Caso não sejam adequados a utilização de todos os 82 animais irá aumentar a sensibilidade de mapeamento possibilitando a identificação de genes modificadores que não tenham sido encontrados anteriormente, o melhor posicionamento dos genes já identificados

até que as regiões observadas sejam restringidas apenas um único gene. Melhorias que por sua vez permitem estimar equações mais precisas para predizer cada fenótipo.

Assim, o aumento do universo amostral não apenas irá melhorar a precisão do mapeamento como também a precisão com que os efeitos de cada QTL são estimados e das equações de predição fenotípica. Entretanto, estudos que levem em considerações diferenças ambientais seriam importantes para tentar elucidar quais fatores ambientais são importantes para cada fenótipo de interesse, bem como a sua interação com os outros QTLs que expliquem a herdabilidade restante do modelo (Wu, 2012).

Ainda é importante considerar que a herdabilidade observada nesse modelo é restrita apenas as linhagens 129 e B6, sendo que podem existir outros genes modificadores quando são comparadas outras linhagens murinas mais distantes que as utilizadas. Logo, os QTLs identificados podem ser apenas uma parcela de todos os existentes, mostrando que estudo de ligação com outras linhagens podem vir a confirmar a importância dos genes identificados, bem como identificar novos genes modificadores importantes para a arquitetura de cada um dos traços.

5.3. Microarray de Expressão

A integração de várias técnicas, assim como feito por Faber e cols. 2009, pode aprimorar o resultado obtido em estudos de mapeamento. Como resultados obtidos inicialmente para o mapeamento do sistema pulmonar não indicaram ligação com nenhuma região específica do genoma, os genes identificados na comparação de animais com fenótipo grave contra animais com fenótipos leves podem ser genes candidatos a modificarem o fenótipo pulmonar, uma vez que genes modificadores podem agir através da manipulação dos níveis de expressão de outros genes. É importante ressaltar que as alterações nos níveis de

expressão entre os animais 129 devem surgir devido a alteração em fatores epigenéticos, como nível de metilação de ilhas CpG ou acetilação de histonas, ou a eventos estocásticos.

Assim esperava-se encontrar genes que estivessem envolvidos nesses processos, porém nenhum dos genes com funções interessantes está envolvido neles. Contudo, os genes escolhidos para serem validados apresentam as mesmas funções que os genes identificados na regiões delimitadas através do mapeamento.

Essa concordância quanto às funções dos genes identificados também existe quando se considera as vias que se encontram enriquecidas usando como base os genes identificados entre os animais KO e os animais selvagens.

Logo, podem existir genes modificadores dos três fenótipos estudados da SMF que estejam envolvidos nos processos de enovelamento protéico e sua degradação; transporte de vesículas; na organização da matriz extracelular do tecido conjuntivo, sendo especificamente de mais importante a organização das microfibras e fibras elásticas; e de fatores de crescimento que sejam capaz de estimular/inibir o nível de ativação de TGF- β .

Como demonstrado nos estudos real-time PCR, a grande variação de expressão observada entre animais com fenótipos semelhantes sugere que exista uma faixa continua nos níveis de expressão dos genes modificadores envolvidos, que não deve ser exclusividade dos sistema pulmonar, e que a junção de várias alterações como essas levem à alteração observada. Assim, até as menores variações em algum gene possam ser importantes para a determinação fenotípica da SMF.

Uma vez que a manifestação fenotípica possa ser determinada por leves alterações de expressão, é possível que seja ainda mais complexo a identificação e elucidação de como e por quais vias esses genes interajam para levar a piora ou melhora das alterações fenotípicas observadas.

6. Conclusão

A hipótese deste trabalho era que de as diferenças fenotípicas observadas entre os animais do modelo murino $mg\Delta^{loxPneo}$ quando presente nas linhagens isogênicas 129 e B6 são devidas as diferenças alélicas entre as duas linhagens, e logo que existiam genes modificadores envolvidos na determinação de cada fenótipo. Assim para se identificar tais genes foram desenvolvidas técnicas de quantificação fenotípica; aplicados métodos clássicos de mapeamento através de cruzamentos controlados utilizando microssatélites, bem como a utilização de métodos modernos de genotipagem em larga escala de SNPs; e ainda análise de perfil de expressão gênica.

As técnicas de genotipagem estabelecidas são eficazes em separar os animais KO de animais selvagens e mostraram que existe uma variação contínua nos fenótipos dos animais KO, sendo que estes chegam a assumir valores mais extremos que os encontrados em ambas linhagens parentais, ou seja, indicando que existem alelos de predisposição ou proteção diferentes em cada uma das linhagens fazendo com que uma mistura das duas leve aos fenótipos mais graves. Elas também permitiram identificar que existe influência do sexo de cada animal na gravidade do fenótipo cardiovascular.

Apesar dos mapeamentos com microssatélites e SNP terem utilizado o mesmo conjunto de animais, os resultados gerados por cada conjunto de marcadores diverge um do outro, indicando que a densidade de marcadores e a cobertura do genoma é um fator essencial para a correta identificação de regiões ligadas a cada fenótipo. O conjunto que fornecia uma melhor cobertura do genoma, SNPs, identificou 3 regiões de ligação, uma significativa e duas sugestivas, com o fenótipo ósseo e outras 2 regiões sugestivas de ligação com o fenótipo cardiovascular, comprovando assim a hipótese levantada inicialmente.

Utilizando a região de maior valor de ligação como localização mais provável para o gene candidato, foi possível estimar o quanto da variabilidade de cada traço os genes modificadores encontrados podiam explicar. Toda a variabilidade do fenótipo ósseo foi explicada por dois dos três *loci* identificados, enquanto que ainda falta explicar uma parcela do fenótipo cardiovascular e toda a variabilidade do fenótipo pulmonar. Como era esperado, foram observados genes com efeitos transgressivos, ou seja, genes que na linhagem menos acometida levam a uma piora fenotípica; genes com efeitos aditivos; e genes com efeitos que só aparecem quando em heterozigose ou homozigose.

Também foi possível derivar fórmulas que permitem prever o fenótipo dos animais apenas baseado nos genótipos de cada um desses genes. Essas fórmulas ainda precisam ter a sua precisão verificada, mas podem ser utilizadas para estimar o fenótipo que um certo animal terá aos três meses de idade, possibilitando assim verificar se certo tratamento ou se determinadas condições ambientais são capazes de alterar cada um desses fenótipos.

Os genes identificados nas 5 regiões em desequilíbrio de ligação com os fenótipos ósseo e cardiovascular, os genes diferencialmente expressos entre animais 129 leves e graves, e as vias enriquecidas entre animais KO e selvagens mostram uma certa convergência para certos genes de interesse presentes em todas elas, mostrando que genes envolvidos na via de controle do TGF- β e genes da família *Adamtsl*, que já foram correlacionados com a recuperação dos fenótipos da SMF, são possíveis genes candidatos a modificadores do fenótipo. Também foram identificados genes com funções envolvendo o enovelamento de proteínas, tais como chaperonas, ou genes envolvidos na dinâmica de formação do citoesqueleto e assim no processo de endo- e exocitose celular e que podem desempenhar um papel importante no controle da quantidade de fibrilina-1 disponível para a polimerização das microfibras. Além das anteriores, é possível que genes envolvidos nos mecanismos de

expressão génica sejam importantes para a manifestação da SMF, logo os genes de metilação, produção de histonas e sua acetilação podem ter o seu papel no desenvolvimento da síndrome.

Como era esperado foram identificados genes já associados com a SMF, assim como também novas associações de genes como possíveis modificadores da SMF. A complexidade da arquitetura genética por trás das manifestações fenotípicas da SMF é maior do que o esperado inicialmente, uma vez que existem conjuntos de genes modificadores específicos para cada fenótipo que são constituídos por genes com funções semelhantes, que existe uma grande variação no nível de expressão desses genes entre cada animal e que mesmo a menor delas pode alterar o fenótipo dos animais.

Os próximos passos na busca por genes modificadores para a SMF seria aumentar o número de animais nos estudos de mapeamento, assim como realizar um mapeamento intervalar composto que leve em consideração a presença de mais de um gene modificador para cada fenótipo e que procure por esse tipo de interação entre dois locos do genoma ou ainda realizar um mapeamento que leve em consideração múltiplos fenótipos simultaneamente.

Além das melhorias citadas acima, seria interessante realizar uma análise de interação entre cada loco do genoma para estimar as penalidades de adição de fatores aditivos e de interação para que um procedimento iterativo de seleção pudesse ser realizado e assim estimar todos os parâmetros de cada modelo de predição fenotípica.

Os procedimentos envolvendo o perfil de expressão de cada classe fenotípica seriam melhorados se fosse realizada uma análise ANOVA que comparasse diretamente as três categorias em vez de se realizar três comparações par-a-par. Entretanto como só foram considerados os resultados da comparação entre animais KO para identificação de genes candidatos, não deve existir grande impacto sobre o nível de erro tipo 1 desse teste.

Logo o último passo seria redesenhar os *primers* utilizados para que os níveis de expressão dos outros 9 genes escolhidos sejam aferidos, e caso sejam mais constantes entre animais com fenótipos similares possam validar os achados os resultados obtidos no microarray de expressão.

7. Bibliografia

- Ades, L. **Guidelines for the diagnosis and management of Marfan syndrome**. Heart Lung Circ, v.16, n.1, Feb, p.28–30. 2007.
- Biery NJ, Eldadah ZA, Moore CS, Stetten G, Spencer F, Dietz HC.. **Revised genomic organization of FBN1 and significance for regulated gene expression**. Genomics. Feb 15;56(1):70–7. 1999
- Broman, K.W. e Saunak S. **A Guide to QTL Mapping with R/qtl**. Ed. M Gail et al. Springer, 2009.
- Buoni S., Zannolli R., Macucci F., Ansaldi S., Grasso M., Arbustini E., e Fois A. **The FBN1 (R2726W) mutation is not fully penetrant**. Annals of Human Genetics. 68, 633–638. 2004.
- Byers, P.H. **Determination of the molecular basis of Marfan syndrome: a growth industry**. Journal of Clinical Investigation, 114(2), 161–163. 2004. doi:10.1172/JCI200422399
- Capotorti, L.; Gaddini de Benedetti, R.; Rizzo, P. : **Contribution to the study of the heredity of Marfan's syndrome: description of a family tree of 4 generations with marriage between consanguineous parents**. Acta Genet. Med. Gemellol. 8: 455–482, 1959.
- Chen, J.J. **The Hardy–Weinberg principle and its applications in modern population genetics**. Frontiers in Biology, vol. 5, no. 4, pp. 348–353. 2010. <http://dx.doi.org/10.1007/s11515-010-0580-x>
- Chen, Y. e Alman, B. A. (Cellular Biochemistry, 362(November 2008), 353–362. 2009. doi:10.1002/jcb.22020
- Collod–Beroud G., Le Bourdelles S., Ades L., e cols. **Update of the UMD–FBN1 mutation database and creation of an FBN1 polymorphism database**. Hum Mutat;22:199208. 2003.
- Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J., Miller, J.C., Dietrich, W.F., e cols. **A genetic linkage map of the mouse: current applications and future prospects**. Science (New York, N.Y.), 262(5130), 57–66. 1993. <http://www.ncbi.nlm.nih.gov/pubmed/8211130>
- Degner, J.F., Pai, A., Pique–Regi, R., Veyrieras, J.B., Gaffney, D.J., Pickrell, J.K., De Leon, S., e cols. **DNase I sensitivity QTLs are a major determinant of human expression variation**. Nature, 482(7385), 390–4. 2012. doi:10.1038/nature10808

- Dietz, H.C., Cutting, C.R., Pyeritz, R.E., Maslen, C.L., Sakai, L.Y., Corson, G.M., Puffenberger, E.G., e cols. **Marfan syndrome caused by a recurrent de novo missense mutation.** Nature, 352(6333), 337–339. 1991. <http://dx.doi.org/10.1038/352337a0>
- Dipple K.M., e McCabe E.R.B. **Modifier Genes Convert “Simple” Mendelian Disorders to Complex Traits.** Molecular Genetics and Metabolism 71, 43–50. 2000.
- Eichler, E.E., Flint, J., Gibson, G., Kong, A., Leal, S.M., Moore, J.H., e Nadeau, J.H. **Missing heritability and strategies for finding the underlying causes of complex disease.** Nature Reviews Genetics, 11(6), 446–450. 2010. doi:10.1038/nrg2809
- Faivre, L., Collod-Beroud, G., Ades, L., Arbustini, E., Child, A., Callewaert, B., Loeys, B., e cols. **The new Ghent criteria for Marfan syndrome: what do they change?** Clinical Genetics, 81(5), 433–442. 2012. doi:10.1111/j.1399-0004.2011.01703.x
- Farber, C.R., Van Nas, A., Ghazalpour, A., Aten, J.E., Doss, S., Sos, B., Schadt, E.E., e cols. **An Integrative Genetics Approach to Identify Candidate Genes Regulating BMD: Combining Linkage, Gene Expression, and Association.** Journal of Bone and Mineral Research, 24(1), 105–116. 2009. doi:10.1359/jbmr.080908
- Flint J., Valdar W., Shifman S. e Mott R; **Strategies for mapping and cloning quantitative trait genes in rodents.** Nat Rev Genet 6:271–286. 2005.
- Gabriel, L.A.R., Wang, L.W., Bader, H., Ho, J.C., Majors, A.K., Hollyfield, J.G., Traboulsi, E.I., e cols. **ADAMTSL4, a Secreted Glycoprotein Widely Distributed in the Eye, Binds Fibrillin–1 Microfibrils and Accelerates Microfibril Biogenesis.** Investigative Ophthalmology & Visual Science, 53(1), 461–469. 2012. doi:10.1167/iovs.10–5955
- Génin, E., Feingold J., e Clerget-Darpoux F. **Identifying modifier genes of monogenic disease: strategies and difficulties.** Human Genetics 124.4: 357–368. 2008.
- Green, M.C., Sweet H.O., e cols. **Tight-skin, a new mutation of the mouse causing excessive growth of connective tissue and skeleton.** Am J Pathol, v.82, n.3, Mar, p.493–512. 1976.
- Griffiths, A.J.F., Wessler S.R., Lewontin R.C., Gelbart W., Suzuki D.T., Miller J.H.. Introdução à Genética Ed. Guanabara;2006 8ª Edição ISBN 9788527711104

- Grüneberg, H. **The Pathology of Development; a Study of Inherited Skeletal Disorders in Animals.** Wiley; New York;(1963).
- Hahne, F., Huber, W., Gentleman, R., Falcon, S. **Bioconductor Case Studies.** 2008, XII, 284 p. ISBN 978-0-387-77239-4.
- Haldane J. **The relative importance of principal and modifying genes in determining some human diseases.** Journal of Genetics 41:149-157. 1941.
- Haley C. e Knott S. **A simple regression method for mapping quantitative trait loci in line crosses using flanking markers.** Heredity 69: 315-324. 1992.
- Hall J, Horton W. **Genetics Glossary. Growth, Genetics and Hormones.** 1997.
- <http://www.informatics.jax.org/>
- http://www.uniscience.com.br/site/index.php?option=com_content&task=view&id=275&Itemid=5
- Hutchinson, S., e cols. **Allelic variation in normal human FBN1 expression in a family with Marfan syndrome: a potential modifier of phenotype?** Hum. Mol. Genet. 12:2269.2276. 2003.
- Index of Major Mouse Strains – Mouse Genome informatics. Versão 5.01.03
<http://www.informatics.jax.org/external/festing/mouse/STRAINS.shtml>
- Isogai, Z., Ono, R. N., Ushiro, S., Keene, D. R., Chen, Y., Mazzieri, R., Charbonneau, N. L., e cols. **Latent transforming growth factor beta-binding protein 1 interacts with fibrillin and is a microfibril-associated protein.** The Journal of biological chemistry, 278(4), 2750-7. 2003. <http://www.jbc.org/cgi/content/abstract/278/4/2750>
- Judge, D.P., Biery N.J., e cols. **Evidence for a critical contribution of haploinsufficiency in the complex pathogenesis of Marfan syndrome.** J Clin Invest, v.114, n.2, Jul, p.172-81. 2004.
- Kaartinen, V., e Warburton, D. **Fibrillin controls TGF-beta activation.** Nat. Genet. 33:331-332. 2003.
- Kielty C.M., Baldock C., Lee D., Rock M.J., Ashworth J.L., Shuttleworth C.A. **Fibrillin: from microfibril assembly to biomechanical function.** Philos Trans R Soc Lond B Biol Sci. Feb 28;357(1418):207-17. 2002.

- Kielty C.M. e Shuttleworth C.A. **Fibrillin-containing microfibrils: structure and function in health and disease.** Int J Biochem Cell Biol. Aug;27(8):747–60. 1995.
- Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M.G., Sarfarazi, M., Tsipouras, P., e cols. **Linkage of Marfan syndrome and a phenotypically related disorder to two different fibrillin genes.** Nature, 352(6333), 330–334. 1991. <http://dx.doi.org/10.1038/352330a0>
- Lima, B.L., Santos, E.J.C., Fernandes, G.R., Merkel, C., Mello, M.R.B., Gomes, J.P.A., Soukoyan, M., e cols. **A New Mouse Model for Marfan Syndrome Presents Phenotypic Variability Associated with the Genetic Background and Overall Levels of Fbn1 Expression.** PLoS ONE, 5(11), e14136. 2010. doi:10.1371/journal.pone.0014136.g005
- Lin, H.F., Juo, S.H., e Cheng, R. **Comparison of the power between microsatellite and single-nucleotide polymorphism markers for linkage and linkage disequilibrium mapping of an electrophysiological phenotype.** BMC Genetics, 6 (Suppl 1), S7. 2005. doi:10.1186/1471-2156-6-S1-S
- Lindblad-Toh, K., Winchester, E., Daly, M.J., Wang, D.G., Hirschhorn, J.N., Lavolette, J.P., Ardlie, K., e cols. **Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse.** Nature genetics, 24(4), 381–6. 2000. doi:10.1038/74215
- Loeys, B.L., Dietz, H.C., e cols. **The revised Ghent nosology for the Marfan syndrome.** J Med Genet, v.47, n.7, Jul, p.476–85. 2010.
- Mackay, T.F.C. **The Genetic Architecture of Quantitative Traits.** Annu. Rev. Genet. 35:303–39. 2001.
- Makowsky, R., Pajewski, N.M., Klimentidis, Y.C., Vazquez, A.I., Duarte, C.W., Allison, D.B., e de los Campos, G. **Beyond Missing Heritability: Prediction of Complex Traits.** PLoS Genetics, 7(4), e1002051. 2011. doi:10.1371/journal.pgen.1002051.t002
- McKusick, V. **The cardiovascular aspects of Marfan's syndrome: a heritable disorders of connective tissue.** Circulation, 11:321–42. 1955
- Mike, I. **Phenotype variability: penetrance and expressivity.** Nature Education 1(1). 2008.
- Nadeau, J.H. **Modifier genes in mice and humans.** Nature Reviews Genetics 2, 165–174. 2001.doi:10.1038/35056009.

- Nakajima, Y., Miyazono, K., e Nakamura, H. **Immunolocalization of latent transforming growth factor-beta binding protein-1 (LTBP1) during mouse development: possible roles in epithelial and mesenchymal cytodifferentiation.** Cell Tissue Res. 295:257-267. 1999.
- Neptune, E.R., e cols. **Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome.** Nat. Genet. 33:407-411. 2003.
- Ng C.M., Cheng A., Myers L.A., Martinez-Murillo F., Jie C., Bedja D., Gabrielson K.L., Hausladen J.M., Mecham R.P., Judge D.P., Dietz H.C. **TGF-beta-dependent pathogenesis of mitral valve prolapse in a mouse model of Marfan syndrome.** J Clin Invest. Dec;114(11):1586-92. 2004.
- Nica A.C., Dermitzakis E.T.; **Using gene expression to investigate the genetic basis of complex disorders.** Human molecular Genetics, vol 17, 129-134. 2008.
- Pereira, L., Andrikopoulos, K., Tian, j., lee, S.Y., keene, D.R., Ono, R., Reinhardt, D.P., Sakai, L.Y., Biery, N.J., Bunton, T., Dietz, H.C., Ramirez, F. **Targetting of the gene encoding fibrillin-1 recapitulates the vascular aspect of Marfan syndrome.** Nat. Genet.,17:218-22. 1997.
- Pereira, L., D'Alessio, M., Ramirez, F., Lynch J.R., Sykes, B., Pangilinan, T., Bonadio, J., **Genomic organization of the sequence coding for fibrillin, the defective gene product in Marfan syndrome.** Hum. Mol. Genet. 2(7): 961-968. 1993. doi:10.1093/hmg/2.7.961
- Pereira, L., Lee,S.Y., Gayraud,B., Andrikopoulos, K., Shapiro, S.D., Bunton, T., Biery, N.J., Dietz, H.C., Sakai, L.Y., and Ramirez, F. **Pathogenetic sequence for aneurysm revealed in mice underexpressing fibrillin-1.** Pros.Acad.Sci.USA, 96:3819-23. 1999.
- Pyeritz R.E. **The Marfan syndrome.** Am Fam Physician. Dec;34(6):83-94. 1986.
- Pyeritz, R.E. **The Marfan syndrome.** Annu Rev Med, v.51, p.481-510. 2000.
- Rakyan, V. K., Blewitt, M. E., e cols. **Metastable epialleles in mammals.** Trends Genet, v.18, n.7, Jul, p.348-51. 2002.
- Ramirez F, Pereira L. **The fibrillins.** Int J Biochem Cell Biol. Feb;31(2):255-9. 1999.
- Robinson P.N., Booms P., Katzke S., Ladewig M., Neumann L., Palz M., e cols. **Mutations of FBN1 and genotype-phenotype correlations in**

Marfan syndrome and related fibrillinopathies. Hum Mutat. 20:153-61. 2002.

- Robinson, P.N., Arteaga-Solis, E., e cols. **The molecular genetics of Marfan syndrome and related disorders.** J Med Genet, v.43, n.10, Oct, p.769-87. 2006.
- Robinson, P.N., Godfrey, M. **The Molecular Genetics of Marfan Syndrome and related microfibriliopathies.** J. Med. Genet. 37, 9-25. 2000.
- Rozmahel R., Wilschanski M., Matin A., Plyte S., Oliver M., e cols. **Modulation of disease severity in cystic fibrosis transmembrane conductance regulator deficient mice by a secondary genetic factor.** Nat Genet 12: 280-287. 1996.
- Saito, M., Kurokawa, M., Oda, M., Oshima, M., Tsutsui, K., Kosaka, K., Nakao, K., e cols. **ADAMTSL6 Protein Rescues Fibrillin-1 Microfibril Disorder in a Marfan Syndrome Mouse Model through the Promotion of Fibrillin-1 Assembly.** Journal of Biological Chemistry, 286(44), 38602-38613. 2011. doi:10.1074/jbc.M111.243451
- Sakai, L.Y., Keene, D.R., e Engvall, E. **Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils.** The Journal of cell biology, 103(6 Pt 1), 2499-509. 1986. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2114568&tool=pmcentrez&rendertype=abstract>
- Salvatore, F., Scudiero, O. e Castaldo, G. **Genotype-phenotype correlation in cystic fibrosis: the role of modifier genes.** Am. J. Med. Genet., 111, 88-95. 2002.
- Santos, E. **Desenvolvimento de um modelo animal para a Síndrome de Marfan através da manipulação do genoma do camundongo.** Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2005.
- Schadt E.E., Lamb J., Yang X., Zhu J., Edwards S., Guhathakurta D., Sieberts S.K., Monks S., Reitman M., Zhang C., Lum P.Y., Leonardson A., Thieringer R., Metzger J.M., Yang L., Castle J., Zhu H., Kash S.F., Drake T.A., Sachs A., Lusk A.J. **An integrative genomics approach to infer causal associations between gene expression and disease.** Nat Genet. 37:710-717. 2005.
- Sen, S., Johannes, F., e Broman, K. W. (2009). **Selective Genotyping and Phenotyping Strategies in a Complex Trait Context.** Genetics, 181(4), 1613-1626. 2009. doi: 10.1534/genetics.108.094607

- Shalon D., Smith S.J., Brown P.O.: **A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization.** *Genome Res.* 6:639–645. 1996.
- Silverman D.I., Burton K.J., Gray J., Bosner M.S., Kouchoukos N.T., Roman M.J., Boxer M., Devereux R.B., Tsipouras P. **Life expectancy in the Marfan syndrome.** *Am J Cardiol.* Jan 15;75(2):157–60. 1995.
- Silverman D.I., Gray J., Roman M.J., Bridges A., Burton K., Boxer M., Devereux R.B., Tsipouras P. **Family history of severe cardiovascular disease in Marfan syndrome is associated with increased aortic diameter and decreased survival.** *J Am Coll Cardiol.* Oct;26(4):1062–7. 1995.
- Slavotinek, A. **Genetic modifiers in human development and malformation syndromes, including chaperone proteins.** *Human Molecular Genetics*, 12(90001), 45R–50. 2003. doi:10.1093/hmg/ddg099
- Wu, R. **Predicting the Genotype–Phenotype Map of Complex Traits.** *Journal of Biometrics & Biostatistics*, 03(04). 2012. doi:10.4172/2155–6180.1000e109
- Xiao J., Lucas A., D'Andrade P., Visitacion M., **Performance of the Agilent microarray platform for one-color analysis of gene expression.** 2006. www.chem.agilent.com/Library/applications/5989–4486EN_72.pdf
- Zhang, H., Hu, W., e Ramirez, F. **Developmental expression of fibrilin genes suggest heterogeneity of extracellular microfibrils.** *J. Cell Biol.*, 123:1165–76. 1995.
- Zuk, O., Hechter, E., Sunyaev, S. R., & Lander, E. S. **The mystery of missing heritability: Genetic interactions create phantom heritability.** *Proceedings of the National Academy of Sciences*, 109(4), 1193–1198. 2012. doi:10.1073/pnas.1119675109

8. Anexo 1 – Lista de microssatélites utilizados

Marcadores	Cromossomo	Posição (cM)	Marcadores	Cromossomo	Posição (cM)
D1MIT3	1	6,27	D10MIT42	10	39,72
D1MIT211	1	10,59	D10MIT95	10	47,06
D1MIT132	1	39,52	D10MIT180	10	66,65
D1MIT286	1	56,25	D11MIT77	11	10,44
D2MIT274	2	57,72	D11MIT20	11	27,23
D2MIT277	2	60,69	D11MIT320	11	43,21
D2MIT285	2	75,41	D12MIT153	12	15,71
D2MIT265	2	97,97	D12MIT5	12	37,16
D3MIT203	3	10,82	D12MIT8	12	62,1
D3MIT22	3	32,25	D13MIT117	13	16,81
D3MIT10	3	43,03	D13MIT147	13	51,06
D3MIT44	3	72,43	D13MIT35	13	67,21
D4MIT181	4	4,23	D14MIT11	14	6,33
D4MIT2	4	10,61	D14MIT60	14	24,6
D4MIT254	4	83,75	D14MIT193	14	37,26
D5MIT57	5	28	D15MIT29	15	34,16
D5MIT259	5	44,32	D15MIT171	15	45,02
D6MIT138	6	1,81	D15MIT242	15	45,1
D6MIT36	6	48,93	D15MIT161	15	52,78
D6MIT294	6	77,7	D16MIT154	16	2,59
D7MIT96	7	54,36	D16MIT4	16	25,43
D7MIT238	7	63,78	D16MIT106	16	57,68
D7MIT71	7	65,2	D17MIT133	17	12,53
D8MIT155	8	2,14	D17MIT49	17	22,26
D8MIT224	8	19,38	D17MIT3	17	39,08
D8MIT249	8	38,64	D17MIT123	17	60,67
D9MIT67	9	20,71	D18MIT202	18	23,29
D9MIT191	9	25,52	D18MIT4	18	57,53
D9MIT182	9	53,48	D19MIT36	19	48,79
D10MIT2	10	9,78	D19MIT1	19	50,32

9. Anexo 2 – Procedimento de mapeamento com microsatélites

```
library(qtl)
library(snow)
library(rlecuyer)
setwd("/Users/apple/R/dados F2")
dados=read.cross("csv",dir="/Users/apple/R/dados F2", "dados F2
micro.csv")
dados.ko=subset(dados,ind=c(pull.pheno(dados,4)=="het"))
dados.geno=subset(dados.ko,ind=c(ntyped(dados.ko,"ind")>40))
plot(dados.ko,pheno.col=-c(4,6),alternate.chrid=FALSE)
plot(dados.geno)
plot(dados.ko)
summary.cross(dados)
dados.ko

par(mfrow=c(1,3))
boxplot(coluna/100~genotipo,data=dados$pheno,ylab="KR",xlab="genotip
o")
boxplot(coluna/100~sexo,data=dados.ko$pheno,ylab="KR",xlab="genotipo
")
anova(aov(coluna~idade*sexo,data=dados.ko$pheno))

boxplot(aorta~genotipo,data=dados$pheno,ylab="AWT(um)",xlab="genotip
o")
boxplot(aorta~sexo,data=dados.ko$pheno,ylab="AWT(um)",xlab="genotipo
")
anova(aov(aorta~idade*sexo,data=dados.ko$pheno))

boxplot(pulmao~genotipo,data=dados$pheno,ylab="Lm",xlab="genotipo")
boxplot(pulmao~sexo,data=dados.ko$pheno,ylab="Lm",xlab="genotipo")
anova(aov(pulmao~idade*sexo,data=dados.ko$pheno))

t.test(aorta~sexo,data=dados.ko$pheno)
t.test(coluna~genotipo,data=dados$pheno)
t.test(pulmao~genotipo,data=dados$pheno)

KR=pull.pheno(dados.ko,pheno.col=1)/100

AWT=pull.pheno(dados.ko,pheno.col=2)
Lm=pull.pheno(dados.ko,pheno.col=3)
hist(Lm,main="")

hist(log(AWT))
```

```

plot(KR,AWT)
kr1=KR/100
abline(lm(AWT~KR),col="RED")
cor.test(kr1,AWT,use="pairwise.complete.obs")

plot(KR,Lm)
co.kr.lm=coefficients(lm(Lm~KR))
abline(lm(Lm~KR),col="RED")
cor.test(KR,Lm,use="pairwise.complete.obs")

plot(Lm,AWT)
abline(lm(AWT~Lm),col="RED")
cor.test(Lm,AWT,use="pairwise.complete.obs")

shapiro.test(pull.pheno(dados.ko,pheno.col=1))
shapiro.test(pull.pheno(dados.ko,pheno.col=2))
shapiro.test(pull.pheno(dados.ko,pheno.col=3))

par(mfrow=c(2,2))
qqnorm(pull.pheno(dados.ko,pheno.col=1),xlab="KR")
qqline(pull.pheno(dados.ko,pheno.col=1))
qqnorm(pull.pheno(dados.ko,pheno.col=2),xlab="AWT")
qqline(pull.pheno(dados.ko,pheno.col=2))
qqnorm(pull.pheno(dados.ko,pheno.col=3),xlab="Lm")
qqline(pull.pheno(dados.ko,pheno.col=3))

gt=geno.table(dados.ko)
gt[gt$P.value<0.05,]

ko=est.rf(dados.geno)
checkAlleles(ko)
plot.rf(ko,mark.diagonal=TRUE)

sim.all=sim.geno(dados.ko,step=1,n.draws=1024,error.prob=0.01)
strata=(ntyped(dados.lod)>40)

kr.imp=scanone(sim.all,pheno.col=1,method="imp")
kr.imp.perm=scanone(sim.all,pheno.col=1,method="imp",
n.perm=1000,perm.strata=strata, n.cluster=8)
thres.kr.imp=summary(kr.imp.perm,alpha=c(0.63,0.05))

jpeg(file="perfil lodscore KR.jpeg")
plot.scanone(kr.imp,col=c("black"),main="KR",ylim=c(0,5))

```

```
abline(h=c(thres.kr.imp),lty=c(1,2),col="black")
cores=c("black")
legend("topright",legend=c("sugestivo","significativo"),col=cores,lt
y=c(1,2))
dev.off()
```

```
pulmao.imp=scanone(sim.all,pheno.col=3,method="imp")
pulmao.imp.perm <- scanone(sim.all,pheno.col=3, method="imp",
n.perm=1000,perm.strata=strata,n.cluster=8)
thres.pulmao.imp=summary(pulmao.imp.perm,alpha=c(0.63,0.05))
```

```
jpeg(file="perfil lodscore pulmao.jpeg")
plot.scanone(pulmao.imp,main="pulmão",ylim=c(0,5))
abline(h=c(thres.pulmao.imp),lty=c(1,2))
cores=c("black")
legend("topright",legend=c("sugestivo","significativo"),col=cores,lt
y=c(1,2))
dev.off()
```

```
sex.awt=pull.pheno(dados.ko,pheno.col=5)
cardiaco.imp.awt.cov=scanone(sim.all,pheno.col=2,method="imp",addcov
=sex.awt,intcov=sex.awt)
cardiaco.imp.awt.cov.perm=scanone(sim.all,pheno.col=2,method="imp",n
.perm=1000,perm.strata=strata,addcov=sex.awt,intcov=sex.awt,n.cluste
r=8)
thresh.card.imp.awt.cov=summary(cardiaco.imp.awt.cov.perm,alpha=c(0.
63,0.05))
```

```
jpeg(file="perfil lodscore AWT.jpeg")
plot(cardiaco.imp.awt.cov,ylim=c(0,8),main="AWT")
abline(h=c(thresh.card.imp.awt.cov),lty=c(1,2),col=c("black"))
cores=c("black")
legend("topright",legend=c("sugestivo","significativo"),col=cores,lt
y=c(1,2))
dev.off()
```

```
bayesint(pulmao.imp,2)
bayesint(pulmao.imp,6)
bayesint(pulmao.imp,16)
bayesint(cardiaco.imp.awt.cov,6)
```

10. Anexo 3 – Procedimento de identificação dos SNPs informativos e preparo do banco de dados

```
setwd("/Users/apple/R/SNP")
snp=read.table("resultado snps.txt", header=TRUE, sep="\t")
snp_red=snp[,c(1:4)]
Sv=which(snp_red[,2]==5480)
B6=which(snp_red[,2]==5481)

int.sv=which(snp_red[Sv,3]==snp_red[Sv,4])
int.b6=which(snp_red[B6,3]==snp_red[B6,4])
int.1=intersect(which(snp_red[Sv,3]!=snp_red[B6,3]),which(snp_red[Sv,4]!=snp_red[B6,4]))
int.2=intersect(which(snp_red[Sv,3]!=snp_red[B6,4]),which(snp_red[Sv,4]!=snp_red[B6,3]))

int=intersect(int.1,int.2)
int=intersect(int,int.sv)
int=intersect(int,int.b6)

mapa=read.table("SNP_Map.txt", header=TRUE, sep="\t")
mapa=mapa[int,c(2,3,4)]

write.table(mapa[,c(2,3)],file="SNPs_diferentes.txt",sep="\t")

mapa2=read.table("conversion-results.txt", header=FALSE, sep=" ")
index=order(mapa2[,4],mapa2[,5],mapa2[,2])
mapa2[index,-c(3)]

b=0
for (i in levels(as.factor(snp_red[,2]))) {
  b=cbind(b,assign(paste("a",i,sep="."),snp_red[which(snp_red[,2]==i),])
)
}
b=b[,-1]

d=matrix("-",length(int),48)
for(i in 1:length(int)){
  for(j in 1:length(int)){
    if (b[int[j],4*i-1]==b[int[j],4*i] && b[int[j],4*i-1]==b[int[j],4*47-1]) d[j,i]="A"
    if (b[int[j],4*i-1]==b[int[j],4*i] && b[int[j],4*i-1]==b[int[j],4*48-1]) d[j,i]="B"
    if (b[int[j],4*i-1]!=b[int[j],4*i]) d[j,i]="H"
    print(c(i,j))
  }
}
```

```
db=cbind(levels(as.factor(snp_red[,2])),t(d))
```



```
blah=c(1,index+1)
db=rbind(t(rbind(c("", ""),mapa[index,c(2,3)])),c("",mapa2[index,5]),db[,bla
h])
write.table(db,file="db snp.txt",sep="\t",dec=".")
```

11. Anexo 4 – Procedimento de mapeamento com de SNPs

```
library(qtl)
library(snow)
setwd("/Users/apple/R/SNP")
dados=read.cross("csv",dir="/Users/apple/R/SNP", "db snp final
cópia.txt")

dup=findDupMarkers(dados,exact.only=TRUE,adjacent.only=TRUE)
dados=drop.markers(dados,unlist(dup))

dados=jittermap(dados)
summary.cross(dados)
plot.cross(dados,pheno.col=-c(4,6,7,8),alternate.chrid=FALSE)
dados.ko=subset(dados,ind=c(pull.pheno(dados,4)=="het"))
plot.cross(dados.ko,pheno.col=-c(4,6,7,8))

cg=comparegeno(dados.ko)
hist(cg)
rug(cg)
which(cg>0.7,arr.ind=TRUE)

dados.ko=est.rf(dados.ko)
dados.ko=drop.markers(dados.ko,c("UNC030069155","UNC110212482"))
checkAlleles(dados.ko)
dados.ko=drop.markers(dados.ko,c("UNC030033636","UNC050482880","UNC1
20196132","UNC110105393","UNC140360737"))
checkAlleles(dados.ko)
dados.ko=drop.markers(dados.ko,c("UNC061512953"))
checkAlleles(dados.ko)

gt=geno.table(dados.ko)
gt[gt$P.value<0.05,]
dim(gt[gt$P.value>0.05,])
hwe=which(gt$P.value<0.05)

dados.hwe=drop.markers(dados.ko,c(markernames(dados.ko)[hwe]))
plot.map(dados.hwe)

pull.pheno(dados.hwe,pheno.col=1)
dat=pull.geno(dados.hwe)
image(1:ncol(dat),1:nrow(dat),t(dat),ylab="Individuals",xlab="Marker
s",
      col=c("red","yellow","blue","green","violet"))
```

```
abline(v=cumsum(c(0,nmar(dados.hwe)))+0.5)
abline(h=nrow(dat)+0.5)
```

```
plot.rf(dados.hwe,mark.diagonal=TRUE,main="Fração de Recombinação
par a par e valores de Lod Score")
```

```
dados.lod=calc.genoprob(dados.hwe, step=1, error.prob=0.01)
plot.info(dados.lod)
dados.errorlod=calc.errorlod(dados.geno,error.prob=0.01)
plot.errorlod(dados.errorlod)
top.errorlod(dados.errorlod)
```

```
strata=(ntyped(dados.lod)>40)
```

```
osseo.em <- scanone(dados.lod,pheno.col=1, method="hk")
osseoperm.em <- scanone(dados.lod,pheno.col=1, method="hk",
n.perm=1000,perm.Xsp=FALSE,n.cluster=8,perm.strata=strata)
thres.osseo.em=summary(osseoperm.em, alpha=c(0.63,0.05))
plot(osseoperm.em)
```

```
sex.all=as.numeric(pull.pheno(dados.lod,pheno.col=5)=="female")
cardiaco.np <- scanone(dados.lod,pheno.col=2,
model="normal",method="hk",addcovar=sex.all,intcovar=sex.all)
cardiacoperm.np <-
scanone(dados.lod,pheno.col=2,model="normal",method="hk",
n.perm=1000,perm.strata=strata,n.cluster=8,addcovar=sex.all,intcovar
=sex.all)
thres.cardiaco.np=summary(cardiacoperm.np, alpha=c(0.63,0.05))
```

```
pulmao.em <- scanone(dados.lod,pheno.col=3, method="hk")
pulmaoperm.em <- scanone(dados.lod,pheno.col=3, method="hk",
n.perm=1000,perm.strata=strata,n.cluster=8)
thres.pulmao.em=summary(pulmaoperm.em, alpha=c(0.63,0.05))
```

```
par(mfrow=c(3,1))
plot(osseo.em,col="black",ylim=c(0,6),main="KR",xlab="Cromossomo")
abline(h=c(thres.osseo.em),lty=c(1,2))
plot(cardiaco.np,col="black",ylim=c(0,11),main="AWT",xlab="Cromossom
o")
abline(h=c(thres.cardiaco.np),lty=c(1,2),col="black")
plot(pulmao.em,col="black",main="Lm",ylim=c(0,7),xlab="Cromossomo")
abline(h=c(thres.pulmao.em),lty=c(1,2),col=c("black"))
```

```
cores=c("black")
```

```
legend("topright",legend=c("sugestivo","significativo"),col=cores,lt
y=c(1,2))
```

```
bayesint(osseo.em,6,expandtomarkers=TRUE)
bayesint(osseo.em,3,expandtomarkers=TRUE)
bayesint(osseo.em,"X",expandtomarkers=TRUE)
```

```
bayesint(cardiaco.np,4,expandtomarkers=TRUE)
bayesint(cardiaco.np,13,expandtomarkers=TRUE)
```

```
sim.all=sim.geno(dados.hwe,step=1,n.draws=1024)
```

```
par(mfrow=c(3,1))
effect.kr.3=effectplot(sim.all,pheno.col=1,mname1="UNC030244865",geno1=c("129/129","129/B6","B6/B6"),main="Cromossomo 3")
effect.kr.6=effectplot(sim.all,pheno.col=1,mname1="UNC060308920",geno1=c("129/129","129/B6","B6/B6"),main="Cromossomo 6")
effect.kr.X=effectplot(sim.all,pheno.col=1,mname1="JAX00177060",geno1=c("129/129","129/B6","129/Y","B6/Y"),main="Cromossomo X")
```

```
effectplot(sim.all,pheno.col=1,mname2="UNC060308920",geno2=c("129/129","129/B6","B6/B6"),mname1="UNC030244865",geno1=c("129/129","129/B6","B6/B6"),main="Interação entre 3 e 6")
effectplot(sim.all,pheno.col=1,mname1="UNC060308920",geno1=c("129/129","129/B6","B6/B6"),mname2="JAX00177060",geno2=c("129/129","129/B6","129/Y","B6/Y"),main="Interação entre 6 e X")
effectplot(sim.all,pheno.col=1,mname2="JAX00177060",geno2=c("129/129","129/B6","129/Y","B6/Y"),mname1="UNC030244865",geno1=c("129/129","129/B6","B6/B6"),main="Interação entre 3 e X")
```

```
par(mfrow=c(1,2))
effect.awt.1=effectplot(sim.all,pheno.col=2,mname1="UNC040316624",geno1=c("129/129","129/B6","B6/B6"),main="Cromossomo 4")
effect.awt.2=effectplot(sim.all,pheno.col=2,mname1="backupUNC130158494",geno1=c("129/129","129/B6","B6/B6"),main="Cromossomo 13")
```

```
effect.awt.12=effectplot(sim.all,pheno.col=2,mname1="backupUNC130158494",geno1=c("129/129","129/B6","B6/B6"),mname2="UNC040316624",geno2=c("129/129","129/B6","B6/B6"),main="Interação 4 e 13")
```

```
qtl_kr=makeqtl(dados.lod,chr=c(3,6,"X"),pos=c(63.63,45.85,5.91),what="prob")
qtl_awt=makeqtl(dados.lod,chr=c(4,13),pos=c(68.74,63.03),what="prob")
```

```

plot(qtl_kr)
out.fitqtl.kr=fitqtl(dados.lod,pheno.col=1,method="hk",qtl=qtl_kr,ge
t.ests=TRUE,formula=y~Q1*Q2)
summary(out.fitqtl.kr)
step.kr=stepwiseqtl(dados.lod,pheno.col=1,qtl=qtl_kr,method="hk",for
mula=y~Q1*Q2,keeptrace=TRUE)

plot(qtl_awt)
sex.all=as.numeric(pull.pheno(dados.lod,pheno.col=5)=="female")
index=which(pull.pheno(dados.lod,pheno.col=2)>0)
out.fitqtl.awt=fitqtl(sim.all,pheno.col=2,method="hk",qtl=qtl_awt,ge
t.ests=TRUE,formula=y~Q1+Q1:sex.all,covar=as.data.frame(sex.all))
summary(out.fitqtl.awt)
step.awt=stepwiseqtl(dados.lod,pheno.col=2,qtl=qtl_awt,method="hk",f
ormula=y~Q1,keeptrace=TRUE,covar=as.data.frame(sex.all))

plotModel(qtl_kr,main="Modelo KR",formula=y~Q1*Q2+Q3)
plotModel(qtl_awt,main="Modelo AWT",formula=y~Q1+Q2+sex.all)

plot(osseo.em,chr=c(3,6,"X"))
abline(h=c(thres.osseo.em),lty=c(1,2))

```

12. Anexo 5 – Procedimento de análise de Microarray de expressão e GSEA

```
source("http://bioconductor.org/biocLite.R")
  biocLite("Agi4x44PreProcess")
  biocLite("mgug4122a.db")
  biocLite("affycoretools")

library("Agi4x44PreProcess")
library("mgug4122a.db")
library("affycoretools")
library("multtest")
library("vsn")
library("limma")
library("genefilter")
library("bioDist")
library("cluster")
library("RColorBrewer")
library("gplots")
library("GSEABase")
library("KEGG.db")
library("Category")
library("VennDiagram")
setwd("/Users/apple/R/array")

target=read.targets(infile="estrutura.txt")
data=read.AgilentFE(target,makePLOT=FALSE)
CV.rep.probes(data,"mgug4122a.db",foreground="ProcessedSignal",raw.d
ata=TRUE,writeR=FALSE,target)
genes.rpt.agi(data,"mgug4122a.db", raw.data = TRUE, WRITE.html =
TRUE, REPORT = TRUE)

a=c(3,4,6,7,8,11,12,14,15,16)
b=c(1,2,6,10,11,12,14,15,16)
c=c(1,2,3,4,7,8,10)

target.1=target[a,]
target.2=target[b,]
target.3=target[c,]

data.wt.leve=read.AgilentFE(target.1,makePLOT=FALSE)
data.wt.grave=read.AgilentFE(target.2,makePLOT=FALSE)
data.ko=read.AgilentFE(target.3,makePLOT=FALSE)

CV.rep.probes(data.wt.leve,"mgug4122a.db",foreground="ProcessedSigna
l",raw.data=TRUE,writeR=FALSE,target.1)
```

```

genes.rpt.agi(data.wt.leve, "mgug4122a.db", raw.data = TRUE,
WRITE.html = TRUE, REPORT = TRUE)
CV.rep.probes(data.wt.grave, "mgug4122a.db", foreground="ProcessedSignal", raw.data=TRUE, writeR=FALSE, target.2)
genes.rpt.agi(data.wt.grave, "mgug4122a.db", raw.data = TRUE,
WRITE.html = TRUE, REPORT = TRUE)
CV.rep.probes(data.ko, "mgug4122a.db", foreground="ProcessedSignal", raw.data=TRUE, writeR=FALSE, target.3)
genes.rpt.agi(data.ko, "mgug4122a.db", raw.data = TRUE, WRITE.html = TRUE, REPORT = TRUE)

```

```

dataNORM.1=
BGandNorm(data.wt.leve, BGmethod="half", NORMmethod="quantile", foreground="ProcessedSignal", background="BGUsed", offset=50, makePLOTpre=FALSE, makePLOTpost=FALSE)
dataNORM.2=
BGandNorm(data.wt.grave, BGmethod="half", NORMmethod="quantile", foreground="ProcessedSignal", background="BGUsed", offset=50, makePLOTpre=FALSE, makePLOTpost=FALSE)
dataNORM.3=
BGandNorm(data.ko, BGmethod="half", NORMmethod="quantile", foreground="ProcessedSignal", background="BGUsed", offset=50, makePLOTpre=FALSE, makePLOTpost=FALSE)

```

```

ddFILT.1 =
filter.probes(dataNORM.1, control=TRUE, wellaboveBG=TRUE, isfound=TRUE, wellaboveNEG=TRUE, sat=TRUE, PopnOL=TRUE, NonUnifOL=T, nas=TRUE, limWellabove=75, limISF=75, limNEG=75, limSAT=75, limPopnOL=75, limNonUnifOL=75, limNAS=100, makePLOT=TRUE, annotation.package="mgug4122a.db", flag.counts=T, target.1)
ddFILT.2 =
filter.probes(dataNORM.2, control=TRUE, wellaboveBG=TRUE, isfound=TRUE, wellaboveNEG=TRUE, sat=TRUE, PopnOL=TRUE, NonUnifOL=T, nas=TRUE, limWellabove=75, limISF=75, limNEG=75, limSAT=75, limPopnOL=75, limNonUnifOL=75, limNAS=100, makePLOT=TRUE, annotation.package="mgug4122a.db", flag.counts=T, target.2)
ddFILT.3 =
filter.probes(dataNORM.3, control=TRUE, wellaboveBG=TRUE, isfound=TRUE, wellaboveNEG=TRUE, sat=TRUE, PopnOL=TRUE, NonUnifOL=T, nas=TRUE, limWellabove=75, limISF=75, limNEG=75, limSAT=75, limPopnOL=75, limNonUnifOL=75, limNAS=100, makePLOT=TRUE, annotation.package="mgug4122a.db", flag.counts=T, target.3)

```

```

ddPROC.1 = summarize.probe(ddFILT.1, makePLOT = FALSE, target.1)
ddPROC.2 = summarize.probe(ddFILT.2, makePLOT = FALSE, target.2)

```

```
ddPROC.3 = summarize.probe(ddFILT.3, makePLOT = FALSE, target.3)
```

```
eset.wt.leve= build.eset(ddPROC.1, targets=target.1, makePLOT=FALSE,  
annotation.package = "mgug4122a.db")  
eset.wt.grave= build.eset(ddPROC.2, targets=target.2, makePLOT=FALSE,  
annotation.package = "mgug4122a.db")  
eset.ko= build.eset(ddPROC.3, targets=target.3, makePLOT=FALSE,  
annotation.package = "mgug4122a.db")
```

```
nomes=c("3399-grave", "3399-grave.2", "3401-leve", "3401-leve.2", "3591-  
wt", "3591-wt.2", "3592-leve", "3592-leve.2", "3599-grave", "3599-  
grave.2", "3605-wt", "3605-wt.2", "3611-wt", "3611-wt.2", "3613-  
wt", "3613-wt.2")
```

```
sampleNames(target.1[,1])=nomes[a]  
sampleNames(eset.wt.grave)=nomes[b]  
sampleNames(eset.ko)=nomes[c]
```

```
par(mfrow=c(1,3))  
sds.1=rowSds(exprs(eset.wt.leve))  
sh.1=shorth(sds.1)  
hist(sds.1,breaks=50,col="green",xlab="desvio  
padrão",xlim=c(0,2),main="wt x leve")  
abline(v=sh.1,col="black",lwd=3,lty=2)  
eset.wt.leve=eset.wt.leve[sds.1>=sh.1,]  
eset.wt.leve=nsFilter(eset.wt.leve, var.filter=FALSE,  
requite.entrez=TRUE, require.GOBP=TRUE, require.GOCC=TRUE,  
require.GOMF=TRUE, remove.dupEntrez=TRUE)
```

```
sds.2=rowSds(exprs(eset.wt.grave))  
sh.2=shorth(sds.2)  
hist(sds.2,breaks=50,col="blue",xlab="desvio  
padrão",xlim=c(0,2),main="wt x grave")  
abline(v=sh.2,col="black",lwd=3,lty=2)  
eset.wt.grave=eset.wt.grave[sds.2>=sh.2,]  
eset.wt.grave=nsFilter(eset.wt.grave,var.filter=FALSE,  
requite.entrez=TRUE, require.GOBP=TRUE, require.GOCC=TRUE,  
require.GOMF=TRUE, remove.dupEntrez=TRUE)
```

```
sds.3=rowSds(exprs(eset.ko))  
sh.3=shorth(sds.3)  
hist(sds.3,breaks=50,col="red",xlab="desvio  
padrão",xlim=c(0,2),main="leve x grave")
```



```

abline(v=sh.3,col="black",lwd=3,lty=2)
eset.ko=eset.ko[sds.3>=sh.3,]
eset.ko=nsFilter(eset.ko,var.filter=FALSE, require.entrez=TRUE,
require.GOBP=TRUE, require.GOCC=TRUE, require.GOMF=TRUE,
remove.dupEntrez=TRUE)

```

```

cl.1=as.numeric(eset.wt.leve$eset$Gerep)
cl.1=cl.1-1
design.1=cbind(mean=1,diff=cl.1)
fit.1=lmFit(exprs(eset.wt.leve$eset),design.1)
fit.1=eBayes(fit.1)
wt.leve=links(mgug4122aSYMBOL[featureNames(eset.wt.leve$eset)]),[2]
table.1=topTable(fit.1,coef="diff",
number=Inf,p.value=1,adjust.method="fdr",sort.by="none",,genelist=li
nks(mgug4122aSYMBOL[featureNames(eset.wt.leve$eset)]),[2])
table.1.sig=table.1[which(table.1[,6]<0.05),]
plot(table.1[,2],-
log10(table.1[,6]),pch=".",xlab=expression(log[2]~FC),ylab=expressio
n(-log[10]~p-value),main="Leve x WT")
abline(h=c(-log10(0.0499830345)),v=c(log2(1.5),-
log2(1.5)),col=c("red","blue","blue"))

```

```

int.1=which(table.1[,6]<0.05)

```

```

cl.2=as.numeric(eset.wt.grave$eset$Gerep)
cl.2=c(1,1,0,1,0,0,0,0,0)
design.2=cbind(mean=1,diff=cl.2)
fit.2=lmFit(exprs(eset.wt.grave$eset),design.2)
fit.2=eBayes(fit.2)
wt.grave=links(mgug4122aSYMBOL[featureNames(eset.wt.grave$eset)]),[2]
table.2=topTable(fit.2,coef="diff",
number=Inf,p.value=1,adjust.method="fdr",sort.by="none",genelist=lin
ks(mgug4122aSYMBOL[featureNames(eset.wt.grave$eset)]),[2])
table.2.sig=table.2[which(table.2[,6]<0.05),]
plot(table.2[,2],-
log10(table.2[,6]),pch=".",xlab=expression(log[2]~FC),ylab=expressio
n(-log[10]~p-value),main="Grave x WT")
abline(h=c(-log10(0.0499891418)),v=c(log2(1.5),-
log2(1.5)),col=c("red","blue","blue"))

```

```

int.2=which(table.2[,6]<0.05)

```

```

cl.3=as.numeric(eset.ko$eset$Gerep)-2

```

```

design.3=cbind(mean=1,diff=c1.3)
fit.3=lmFit(exprs(eset.ko$eset),design.3)
fit.3=eBayes(fit.3)
ko=links(mgug4122aSYMBOL[featureNames(eset.ko$eset)])
table.3=topTable(fit.3,coef="diff",
number=Inf,p.value=1,adjust.method="fdr",sort.by="none",genelist=links(mgug4122aSYMBOL[featureNames(eset.ko$eset)]),2)
table.3.sig=table.3[which(table.3[,6]<0.1),]
plot(table.3[,2],-
log10(table.3[,6]),pch=".",xlab=expression(log[2]~FC),ylab=expression(-log[10]~p-value),main="Leve x Grave")
abline(h=c(-log10(0.1)),v=c(log2(1.5),-
log2(1.5)),col=c("red","blue","blue"))
points(table.3[5936,2],-log10(table.3[5936,6]),pch=20,col="red")

int.3=which(table.3[,6]<0.05)

```

?meanSdPlot

```

par(mfrow=c(1,3))
meanSdPlot(eset.wt.leve$eset,main="WT x Mild")
meanSdPlot(eset.wt.grave$eset, main="WT x Severe")
meanSdPlot(eset.ko$eset, main= "Mild x Severe")

```

```

int.1r=which(abs(table.1[,2])>=log2(1.5) & table.1[,6]<=0.05)
int.2r=which(abs(table.2[,2])>=log2(1.5) & table.2[,6]<=0.05)
int.3r=which(abs(table.3[,2])>=log2(1.5) & table.3[,6]<=0.05)
blah=which(table.3[,6]<=0.1)

```

```

lista.3=table.3[blah,1]
lista.2=table.2.sig[,1]
lista.1=table.1.sig[,1]

```

```

comuns.1.2=intersect(lista.1,lista.2)
length(comuns.1.2)
intersect(lista.1,lista.3)
length(intersect(lista.1,lista.3))
intersect(lista.3,lista.2)
length(intersect(lista.3,lista.2))
intersect(comuns.1.2,lista.3)
length(intersect(comuns.1.2,lista.3))

```

```

draw.triple.venn(1120, 2524, 25, 364, 17, 4, 1, category = c("Leve x Wt", "Grave x Wt", "Leve x Grave"), rotation = 1, reverse = FALSE, euler.d = TRUE, scaled = TRUE, lwd = rep(2, 3),lty = rep("solid", 3),

```

```
col = rep("black",3), fill = c("green","dark blue","red"), alpha =
rep(0.5, 3), label.col = rep("black", 7), cex = rep(1, 7), fontface
= rep("plain", 7),fontfamily = rep("serif", 7), cat.pos = c(-40, 40,
180), cat.dist = c(0.05, 0.05, 0.025), cat.col = rep("black", 3),
cat.cex = rep(1, 3), cat.fontface = rep("plain", 3),cat.fontfamily =
rep("serif", 3), cat.just = list(c(0.5, 1), c(0.5, 1), c(0.5, 0)),
cat.default.pos = "outer", cat.prompts = FALSE, rotation.degree =
0,rotation.centre = c(0.5, 0.5), ind = TRUE, sep.dist = 0.05, offset
= 0, list.order = 1:3, main="Genes diferencialmente expressos")
```

```
links(mgug4122aSYMBOL[intersect(comuns.1.2,lista.3)])
links(mgug4122aSYMBOL[intersect(lista.2,lista.3)])
links(mgug4122aSYMBOL[lista.3])
```

```
hclust2 <- function(x, method="average")
  hclust(x, method=method)
dist2 <- function(x)
  as.dist(1-cor(t(x), method="spearman"))
```

```
heat.1=0
for (i in 1:length(lista.1)){
heat.1[i]=links(mgug4122aSYMBOL[lista.1[i]])[2]
}
```

```
heatmap.2(exprs(eset.wt.leve$eset[int.1r]), col=redgreen(75),
scale="row",key=TRUE, symkey=FALSE, density.info="none",
trace="none", cexRow=0.3,distfun=dist2,hclustfun=hclust2,
margins=c(7,7),labRow=heat.1, main="WT x Leve")
```

```
heat.2=0
for (i in 1:length(lista.2)){
heat.2[i]=links(mgug4122aSYMBOL[lista.2[i]])[2]
}
```

```
heatmap.2(exprs(eset.wt.grave$eset[int.2r]),
col=redgreen(75),scale="row", key=TRUE, symkey=FALSE,
density.info="none",
trace="none",cexRow=1,distfun=dist2,hclustfun=hclust2,margins=c(8,7)
,labRow=heat.2, main="WT x Grave")
```

```
heat.3=0
```

```

for (i in 1:length(lista.3)){
  heat.3[i]=links(mgug4122aSYMBOL[lista.3[i]])[2]
}

heatmap.2(exprs(eset.ko$eset[blah]), col=redgreen(75), scale="row",
  key=TRUE, symkey=FALSE, density.info="none",
  trace="none", cexRow=1, distfun=dist2, hclustfun=hclust2, margins=c(8,7),
  labRow=heat.3, main="Leve x Grave")

```

```

#GSEA
gsc.1=?GeneSetCollection(eset.wt.leve$eset[int.1r],
  setType=KEGGCollection())
inc.1=incidence(gsc.1)

nsF.1=eset.wt.leve$eset[colnames(inc.1),]
rtt.1=rowttests(nsF.1, "GErep")
rstat.1=rtt.1$statistic

tA.1=as.vector(inc.1%*%rstat.1)
tAadj.1=tA.1/sqrt(rowSums(inc.1))
names(tA.1)=names(tAadj.1)=rownames(inc.1)
qqnorm(tAadj.1)
qqline(tAadj.1)

gsea.1=tAadj.1[abs(tAadj.1)>5]
gsea.nomes.1=names(gsea.1)
gsea.lista.1=0
for(i in 1:length(gsea.nomes.1)){
  gsea.lista.1[i]=KEGGPATHID2NAME[[gsea.nomes.1[[i]]]
}

set.seed(1001)
nperm=1000
pvals=gseattperm(nsF.1, as.factor(nsF.1$GErep), inc.1, nperm)
pvalcut=0.025
lowP.1=names(which(pvals[,1]<=pvalcut))
highP.1=names(which(pvals[,2]<=pvalcut))
vias.1=union(lowP.1, highP.1)

gsea.lista.1=0
for(i in 1:length(vias.1)){
  gsea.lista.1[i]=KEGGPATHID2NAME[[vias.1[i]]]
}

```

```

}
gsea.lista.1

paths.1=inc.1[union(lowP.1,highP.1),]
mat=paths.1%*%t(paths.1)
mins.1=outer(diag(mat),diag(mat),pmin)
overlapIndex.1=mat/mins.1


gsc.2=GeneSetCollection(eset.wt.grave$eset[int.2r],
setType=KEGGCollection())
inc.2=incidence(gsc.2)

nsF.2=eset.wt.grave$eset[colnames(inc.2),]
rtt.2=rowttests(nsF.2,"GErep")
rstat.2=rtt.2$statistic

tA.2=as.vector(inc.2%*%rstat.2)
tAdj.2=tA.2/sqrt(rowSums(inc.2))
names(tA.2)=names(tAdj.2)=rownames(inc.2)
qqnorm(tAdj.2)
qqline(tAdj.2)


i=1

set.seed(1002)
nperm=1000
pvals=gseattperm(nsF.2,as.factor(nsF.2$GErep),inc.2,nperm)
pvalcut=0.025
lowP.2=names(which(pvals[,1]<=pvalcut))
highP.2=names(which(pvals[,2]<=pvalcut))

getPathNames(lowP.2)
getPathNames(highP.2)
vias.2=union(lowP.2,highP.2)

gsea.lista.2=0
for(i in 1:length(vias.2)){
gsea.lista.2[i]=KEGGPATHID2NAME[[vias.2[i]]]
}
gsea.lista.2

paths.2=inc.2[union(lowP.2,highP.2),]
mat.2=paths.2%*%t(paths.2)
mins.2=outer(diag(mat.2),diag(mat.2),pmin)

```

```

overlapIndex.2=mat.2/mins.2
gsc.3=GeneSetCollection(eset.ko$eset[blah],
setType=KEGGCollection())
inc.3=incidence(gsc.3)

nsF.3=eset.ko$eset[colnames(inc.3),]
rtt.3=rowttests(nsF.3,"GErep")
rstat.3=rtt.3$statistic

tA.3=as.vector(inc.3*%rstat.3)
tAadj.3=tA.3/sqrt(rowSums(inc.3))
names(tA.3)=names(tAadj.3)=rownames(inc.3)
qqnorm(tAadj.3)
qqline(tAadj.3)
vias.3=names(tA.3)

set.seed(1003)
nperm=1000
pvals.3=gseattperm(nsF.3,as.factor(nsF.3$GErep),inc.3,nperm)
pvalcut=0.025
lowP.3=names(which(pvals.3[,1]<=pvalcut))
highP.3=names(which(pvals.3[,2]<=pvalcut))

getPathNames(lowP.3)
getPathNames(highP.3)
vias.3=union(lowP.3,highP.3)

vias.comuns.low=intersect(lowP.1,lowP.2)
vias.comuns.high=intersect(highP.1,highP.2)
vias.comuns= intersect(vias.1,vias.2)
getPathNames(vias.comuns.low)
getPathNames(vias.comuns.high)
unlist(getPathNames(vias.comuns))

```

13. Anexo 6 - Genes na região candidata do cromossomo 3

1110002E22Rik	4930422G04Rik	4930579F01Rik	5730508B09Rik	5S_rRNA
AC098732.1	AC099576.1	AC102128.1	AC111097.1	AC114668.1
AC125040.1	AC135118.1	AC140486.1	AC140486.2	AC164408.1
AC165333.1	Adh1	Adh4	Adh5	Adh6-ps1
Adh6a	Adh6b	Adh7	Agxt2l1	Aimp1
Alpk1	Ank2	Ap1ar	B930007M17Rik	Bank1
Bdh2	Bmpr1b	Casp6	Ccbl2	Ccdc109b
Cenpe	Cfi	Cisd2	Col25a1	Cxxc4
Cyp2u1	D630013G24Rik	Dapp1	Ddit4l	Dkk2
Dnajb14	Egf	Eif4e	Elovl6	Emcn
Enpep	Gar1	Gbp1	Gbp2	Gbp3
Gbp5	Gbp7	Gm10650	Gm15460	Gm15540
Gm15688	Gm15689	Gm16060	Gm16100	Gm16238
Gm16559	Gm2574	Gm4392	Gm4861	Gm4862
Gm5105	Gm5549	Gm9387	Gm9396	Gm9799
Gstcd	Gtf2b	H2afz	Hadh	Ints12
Lamtor3	Larp7	Lef1	Manba	Metap1
Mir1895	Mir1956	Mir302a	Mir302b	Mir302c
Mir302d	Mir367	Mttp	Neurog2	Nfkb1
Nhedc1	Nhedc2	Npnt	Oaz2-ps	Ostc
Papss1	Pdha2	Pdlim5	Pitx2	Pkn2
Pla2g12a	Ppa2	Ppp3ca	Rap1gds1	Rg9mtd2
Rpl34	Rrh	Sec24b	Sgms2	Slc39a8
SNORA17	SNORA17	SNORA17	SNORA17	SNORA32
SNORA61	Tacr3	Tbck	Tet2	Tifa
Tspan5	U1	U6	U6	U6
U6	U6	U6	U6	U6
U6	U6	Ube2d3	Unc5c	

14. Anexo 7 - Genes na região candidata do cromossomo 4

1110049F12Rik	1700013G24Rik	1700021N21Rik	1700029M20Rik	1700037C06Rik
1700095J12Rik	1810058N05Rik	2310026L22Rik	2610002D18Rik	2610528B01Rik
4930549C01Rik	4930555I21Rik	6030445D17Rik	9130020K20Rik	A330049M08Rik
AB041806	AL645531.1	AL669982.1	AL670720.1	AL671011.1
AL671011.2	AL671173.1	AL672076.1	AL807764.1	Alpl
Asap3	BX000694.1	C1qa	C1qb	C1qc
Camk2n1	Catsper4	Ccdc21	Cda	Cdc42
Cela3b	Clic4	Cnksr1	Cnr2	D4Wsu53e
Ddost	E130218I03Rik	E2f2	Ece1	Eif4g3
Epha8	Ephb2	Extl1	Fam43b	Fam54b
Fuca1	Gale	Gm12982	Gm12983	Gm12984
Gm12986	Gm12987	Gm12988	Gm12989	Gm12990
Gm12991	Gm13000	Gm13001	Gm13002	Gm13003
Gm13005	Gm13006	Gm13007	Gm13008	Gm13009
Gm13010	Gm13011	Gm13012	Gm13013	Gm13030
Gm13131	Gm13195	Gm13250	Gm15979	Gm16224
Gm16225	Gm7534	Grhl3	Grrp1	Hmgcl
Hnrnpr	Hp1bp3	Hspg2	Htr1d	Id3
Il22ra1	Il28ra	Kdm1a	Kif17	Lactbl1
Ldlrad2	Ldlrap1	Luzp1	Lypla2	Man1c1
Mir700	Mul1	Myom3	Nipal3	Otud3
Pafah2	Paqr7	Pdik1l	Pink1	Pla2g2a
Pla2g2c	Pla2g2d	Pla2g2e	Pla2g2f	Pla2g5
Pnrc2	Rap1gap	Rcan3	Rhd	Rpl11
Rpl31-ps10	Rpl38-ps1	Runx3	Sepn1	Sh2d5
Sh3bgrl3	Slc30a2	SNORA17	Srrm1	Srsf10
Stmn1	Syf2	Tcea3	Tceb3	Tmem50a
Tmem57	Trim63	U1	U6	U6
U6	U6	Ubxn10	Ubxn11	Usp48
Vwa5b1	Wnt4	Zbtb40	Zfp46	Zfp593

15. Anexo 8 - Genes na região candidata do cromossomo 6

<i>1700010K10Rik</i>	<i>1700123L14Rik</i>	<i>5_8S_rRNA</i>	<i>7SK</i>	<i>7SK</i>
<i>A730049H05Rik</i>	<i>AC131676.1</i>	<i>AC131700.1</i>	<i>AC152987.1</i>	<i>AC157096.1</i>
<i>AC158665.1</i>	<i>AC162038.1</i>	<i>Adamts9</i>	<i>Arl6ip5</i>	<i>AY512915</i>
<i>Chl1</i>	<i>Cntn3</i>	<i>Cntn4</i>	<i>Cntn6</i>	<i>Eif4e3</i>
<i>Eogt</i>	<i>Fam19a1</i>	<i>Fam19a4</i>	<i>Foxp1</i>	<i>Frmd4b</i>
<i>Gm10234</i>	<i>Gm15531</i>	<i>Gm15576</i>	<i>Gm15631</i>	<i>Gm15737</i>
<i>Gm20696</i>	<i>Gm20705</i>	<i>Gm6565</i>	<i>Gm765</i>	<i>Gm9871</i>
<i>Gpr27</i>	<i>Gxylt2</i>	<i>Kbtbd8</i>	<i>Lmod3</i>	<i>Lrig1</i>
<i>Magi1</i>	<i>Mitf</i>	<i>Pdzrn3</i>	<i>Ppp4r2</i>	<i>Prickle2</i>
<i>Prok2</i>	<i>Rybp</i>	<i>Shq1</i>	<i>Slc25a26</i>	<i>SNORA17</i>
<i>SNORA17</i>	<i>SNORA17</i>	<i>SNORA17</i>	<i>SNORA36</i>	<i>SNORA44</i>
<i>SNORA48</i>	<i>SNORA63</i>	<i>Suc1g2</i>	<i>Tmf1</i>	<i>Tpt1-ps3</i>
<i>Trh</i>	<i>U1</i>	<i>U6</i>	<i>U6</i>	<i>U6</i>
<i>U6</i>	<i>U7</i>	<i>U7</i>	<i>Uba3</i>	<i>Zfyve20</i>

16. Anexo 9 - Genes na região candidata do cromossomo 13

1700024P04Rik	1700029F12Rik	1700119I11Rik	4833422C13Rik	5330416C01Rik
7SK	7SK	AC123043.1	AC124515.1	AC130536.1
AC131739.1	AC136976.1	AC137692.1	AC154225.1	AC154271.1
AC154424.1	AC154460.1	AC154460.2	AC160109.1	AC161588.1
Acot12	Aggf1	Ankdd1b	Ankra2	Ankrd34b
Ap3b1	Arsb	Atg10	Atp6ap1l	AW495222
Bhmt	Bhmt2	Btf3	C030017D09Rik	Ccnh
Ckmt2	Cmya5	Col4a3bp	Cox7c	Crhbp
Dhfr	Dmgdh	Edil3	Enc1	F2r
F2rl1	F2rl2	Fam151b	Fam169a	Fcho2
Foxd1	Gcnt4	Gfm2	Gm10320	Gm11019
Gm15620	Gm15622	Gm15907	Gm16243	Gm16318
Gm17190	Gm17352	Gm17450	Gm17622	Gm4117
Gm5453	Gm6169	Gm8526	Gm9776	Gm9828
Hapln1	Hexb	Hmgcr	Homer1	lqgap2
Jmy	Lhfpl2	Mrps27	Msh3	Mtx3
Nsa2	Otp	Papd4	Pde8b	Poc5
Polk	Ptcd2	Rasa1	Rasgrf2	Rgnef
RP24-76E2.4	Rps23	S100z	Scamp1	Serinc5
SNORA17	Snora47	SNORA68	snoU109	snoZ39
Spz1	Ssbp2	Sv2c	Tbca	Thbs4
Tmem167	Tmem171	Tmem174	Tnpo1	U1
U6	U6	U6	Utp15	Vcan
Wdr41	Xrcc4	Zbed3	Zcchc9	Zfp366
Zfyve16				

17. Anexo 10 – Genes na região candidata do cromossomo X

1700012L04Rik	1700054O13Rik	1810030O07Rik	2010204K13Rik	2900002K06Rik
2900008C10Rik	4930402K13Rik	4930453H23Rik	4930524L23Rik	4930557A04Rik
4930578C19Rik	4933416E14Rik	5730405O15Rik	5S_rRNA	5S_rRNA
A230072C01Rik	AA414768	Agtr2	Akap4	AL662925.1
AL671117.1	AL671211.1	AL671492.1	AL671885.1	AL671885.2
AL671922.1	AL732400.1	AL773580.1	Araf	Atp6ap2
AU022751	B630019K06Rik	BC049702	Bcor	Bmp15
BX571729.1	Cacna1f	Cask	Ccdc120	Ccdc22
Ccnb3	Cdk16	Cfp	Chst7	Clcn5
Csnk2a1-ps	CT867956.1	CT867960.1	CT868697.1	CT954250.1
CT963124.1	CT963124.2	CT963124.3	CU468592.1	Cybb
Cypt1	Ddx3x	Dgkk	Drr1	Dusp21
Dynlt3	E330010L02Rik	Ebp	Efhc2	Elk1
Eras	Foxp3	FQ976806.1	Fthl17	Ftsj1
Fundc1	Gata1	Glod5	Gm10058	Gm10096
Gm10147	Gm10230	Gm10487	Gm10488	Gm10490
Gm10491	Gm14363	Gm14365	Gm14366	Gm14367
Gm14369	Gm14370	Gm14373	Gm14374	Gm14375
Gm14379	Gm14380	Gm14381	Gm14394	Gm14395
Gm14448	Gm14450	Gm14451	Gm14452	Gm14457
Gm14458	Gm14459	Gm14470	Gm14473	Gm14474
Gm14475	Gm14476	Gm14477	Gm14478	Gm14479
Gm14480	Gm14481	Gm14482	Gm14483	Gm14484
Gm14485	Gm14491	Gm14493	Gm14499	Gm14500
Gm14501	Gm14502	Gm14503	Gm14504	Gm14505
Gm14506	Gm14509	Gm14510	Gm14511	Gm14512
Gm14513	Gm14514	Gm14515	Gm14516	Gm14517
Gm14518	Gm14519	Gm14520	Gm14521	Gm14523
Gm14524	Gm14525	Gm14527	Gm14529	Gm14530
Gm14531	Gm14533	Gm14534	Gm14535	Gm14536
Gm14537	Gm14538	Gm14539	Gm14558	Gm14561
Gm14562	Gm14563	Gm14632	Gm14634	Gm14635
Gm14636	Gm14703	Gm14720	Gm14799	Gm14820
Gm14862	Gm14863	Gm14864	Gm14865	Gm14867
Gm14883	Gm14884	Gm14885	Gm14974	Gm15027
Gm15028	Gm15029	Gm15030	Gm15276	Gm15280
Gm15432	Gm16265	Gm16479	Gm16480	Gm16481
Gm1848	Gm1989	Gm2003	Gm2005	Gm2012
Gm2030	Gm2066	Gm2092	Gm2101	Gm2117

<i>Gm2165</i>	<i>Gm2200</i>	<i>Gm2759</i>	<i>Gm2768</i>	<i>Gm2777</i>
<i>Gm2784</i>	<i>Gm2790</i>	<i>Gm2799</i>	<i>Gm2825</i>	<i>Gm3657</i>
<i>Gm3669</i>	<i>Gm3677</i>	<i>Gm3681</i>	<i>Gm4297</i>	<i>Gm4732</i>
<i>Gm4789</i>	<i>Gm4836</i>	<i>Gm4906</i>	<i>Gm4907</i>	<i>Gm4984</i>
<i>Gm4985</i>	<i>Gm5073</i>	<i>Gm5123</i>	<i>Gm5124</i>	<i>Gm5132</i>
<i>Gm5168</i>	<i>Gm5169</i>	<i>Gm5379</i>	<i>Gm5380</i>	<i>Gm5381</i>
<i>Gm5382</i>	<i>Gm5383</i>	<i>Gm5384</i>	<i>Gm5634</i>	<i>Gm5635</i>
<i>Gm5751</i>	<i>Gm5752</i>	<i>Gm5753</i>	<i>Gm5754</i>	<i>Gm5755</i>
<i>Gm5924</i>	<i>Gm5926</i>	<i>Gm5926</i>	<i>Gm5931</i>	<i>Gm5934</i>
<i>Gm5935</i>	<i>Gm6071</i>	<i>Gm6079</i>	<i>Gm6113</i>	<i>Gm6121</i>
<i>Gm6592</i>	<i>Gm6787</i>	<i>Gm6797</i>	<i>Gm6798</i>	<i>Gm6826</i>
<i>Gm6829</i>	<i>Gm6923</i>	<i>Gm6938</i>	<i>Gm6956</i>	<i>Gm7079</i>
<i>Gm7129</i>	<i>Gm7164</i>	<i>Gm7350</i>	<i>Gm7365</i>	<i>Gm7375</i>
<i>Gm7386</i>	<i>Gm7391</i>	<i>Gm7398</i>	<i>Gm7408</i>	<i>Gm7415</i>
<i>Gm7421</i>	<i>Gm7437</i>	<i>Gm7438</i>	<i>Gm9083</i>	<i>Gm9085</i>
<i>Gm9428</i>	<i>Gm9429</i>	<i>Gm9430</i>	<i>Gm9431</i>	<i>Gm9432</i>
<i>Gm9434</i>	<i>Gm9435</i>	<i>Gm9436</i>	<i>Gm9439</i>	<i>Gm9440</i>
<i>Gmcl1l</i>	<i>Gpkow</i>	<i>Gpr34</i>	<i>Gpr82</i>	<i>Gripap1</i>
<i>Hdac6</i>	<i>Kcnd1</i>	<i>Kdm6a</i>	<i>Klhl13</i>	<i>Lanc13</i>
<i>Magix</i>	<i>Maoa</i>	<i>Maob</i>	<i>Med14</i>	<i>Mid1ip1</i>
<i>Mir1198</i>	<i>Mir188</i>	<i>Mir221</i>	<i>Mir222</i>	<i>Mir362</i>
<i>Mir500</i>	<i>Mir501</i>	<i>Mir532</i>	<i>Mycs</i>	<i>n-R5s2</i>
<i>n-R5s3</i>	<i>n-R5s5</i>	<i>Ndp</i>	<i>Ndufb11</i>	<i>Nudt10</i>
<i>Nudt11</i>	<i>Nyx</i>	<i>Otc</i>	<i>Otud5</i>	<i>Pcsk1n</i>
<i>Phf16</i>	<i>Pim2</i>	<i>Plp2</i>	<i>Porcn</i>	<i>Ppp1r2-ps7</i>
<i>Ppp1r2-ps9</i>	<i>Ppp1r3f</i>	<i>Pqbp1</i>	<i>Praf2</i>	<i>Prickle3</i>
<i>Psemb6-ps</i>	<i>Psemb7-ps2</i>	<i>Rbm10</i>	<i>Rbm3</i>	<i>Rgn</i>
<i>RP23-367O24.1</i>	<i>Rp2h</i>	<i>Rpgr</i>	<i>Rpl19-ps12</i>	<i>Rpl22-ps1</i>
<i>Rpl23a-ps13</i>	<i>Rpl3-ps1</i>	<i>Rpl30-ps11</i>	<i>Rpl36-ps10</i>	<i>Shroom4</i>
<i>Slc35a2</i>	<i>Slc38a5</i>	<i>Slc6a14</i>	<i>Slc9a7</i>	<i>Slx</i>
<i>Smt3h2-ps</i>	<i>SNORA17</i>	<i>SNORA40</i>	<i>SNORA70</i>	<i>Spaca5</i>
<i>Srpx</i>	<i>Ssx9</i>	<i>Ssxa1</i>	<i>Ssxb1</i>	<i>Ssxb10</i>
<i>Ssxb2</i>	<i>Ssxb3</i>	<i>Ssxb5</i>	<i>Ssxb6</i>	<i>Ssxb8</i>
<i>Ssxb9</i>	<i>Suv39h1</i>	<i>Syn1</i>	<i>Syp</i>	<i>Syt15</i>
<i>Tbc1d25</i>	<i>Tfe3</i>	<i>Timm17b</i>	<i>Timp1</i>	<i>Tspan7</i>
<i>U1</i>	<i>U1</i>	<i>U2</i>	<i>U2</i>	<i>U6</i>
<i>U6</i>	<i>U6</i>	<i>U6</i>	<i>U6</i>	<i>U6</i>
<i>U6</i>	<i>U6</i>	<i>U6</i>	<i>U6</i>	<i>U6</i>
<i>U6</i>	<i>Uba1</i>	<i>Usp11</i>	<i>Usp27x</i>	<i>Usp9x</i>
<i>Uxt</i>	<i>Was</i>	<i>Wdr13</i>	<i>Wdr44</i>	<i>Wdr45</i>
<i>Xk</i>	<i>Zfp182</i>	<i>Zfp300</i>		

18. Anexo 11 – Vias identificadas entre animais KO e selvagens

Amebíase	Junção oclusiva
Apoptose	Junções Comunicantes
Biossíntese de Aminoacil-tRNA	Leucemia mielóide crônica
Biossíntese de hormônios esteróides	Lúpus eritematoso sistêmico
Biossíntese de O-glicano	Metabolismo de ácidos graxos
Biossíntese de terpenóides	Metabolismo de inositol fosfato
Biossíntese de valina, leucina e isoleucina	Metabolismo de pirimidina
Biossíntese pantotenato e CoA	Metabolismo de retinol
Câncer de bexiga	Metabolismo do ácido araquidônico
Carcinoma basocelular	Orientação de Axônio
Cardiomiopatia arritmogênica do ventrículo direito (ARVC)	Processamento de proteínas no retículo endoplasmático
Cardiomiopatia hipertrófica (CMH)	Proteassoma
Ciclo celular	Reabsorção de água regulada por vasopressina
Citotoxicidade mediada por células Nk	Regulação do citoesqueleto de actina
Doença autoimune da tireóide	Rejeição do Enxerto
Doença de Alzheimer	Via de sinalização de receptores de células T
Doença de Huntington	Via de sinalização do receptor de células B
Doenças de príons	Via de sinalização Fc epsilon RI
Endocitose	Via de sinalização Hedgehog
Fatores de transcrição basal	Via de sinalização Wnt
Imunodeficiência primária	Via de sinalização Jak-STAT
Interação ligante-receptor de neuroatividade	Via de sinalização TGF- β
Interação receptores-MEC	Vias metabólicas