UNIVERSIDADE DE SÃO PAULO

FACULDADE DE SAÚDE PÚBLICA

Aprendizado de Estruturas de Dependência entre Fenótipos da Síndrome Metabólica em Estudos Genômicos

Lilian Skilnik Wilk

Projeto para a defesa do Mestrado apresentado ao Programa de Pós-Graduação em Saúde Pública para obtenção do título de Mestre em Ciências

Área de Concentração: Bioestatística

Orientador: Gizelton Pereira Alencar

São Paulo

2017

Aprendizado de Estruturas de Dependência entre Fenótipos da Síndrome Metabólica em Estudos Genômicos

Lilian Skilnik Wilk

Projeto para a defesa do Mestrado apresentado ao Programa de Pós-Graduação em Saúde Pública para obtenção do título de Mestre em Ciências

Área de Concentração: Bioestatística

Orientador: Gizelton Pereira Alencar

Versão Original

São Paulo

2017

Ficha Catalográfica

Wilk, Lilian Skilnik
Aprendizado de Estruturas de Dependência entre fenótipos da Síndrome Metabólica em Estudos Genômicos.
186 páginas

Dissertação (mestrado) – Faculdade de Saúde Pública da Universidade de São Paulo.
Departamento de Epidemiologia.

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Para meu marido Marcelo e nossos filhos, Felipe e Beatriz, pelo incentivo diário e apoio incondicional.

“Descobrir consiste em olhar para o que todo mundo está vendo e pensar uma coisa diferente”. (Roger Von Oech)

AGRADECIMENTOS

Ao meu marido Marcelo, por acreditar e afirmar que eu chegaria lá.

Aos meus filhos, Felipe e Beatriz, por me apoiarem, entenderem e incentivarem diariamente.

Aos meus pais Joseph e Diana, que sempre me apoiaram, desde o início dessa jornada.

À minha irmã Helen, meu cunhado Júlio e meu sobrinho Ilan por me apoiarem, mesmo que à distância, todos os dias.

À família Wilk, Clarice, Henry, Ricardo, Eduardo, Luciana e Rafael pelo suporte emocional em todos os momentos.

Ao meu orientador, Prof. Gizelton P. Alencar, pela oportunidade de trabalhar neste grupo, por sua disponibilidade e todos os seus ensinamentos.

À Prof. Júlia Pavan por seus ensinamentos, pela colaboração, por ter me acolhido desde o início deste projeto e por todo o suporte técnico e emocional.

À Adèle Ribeiro pelo desenvolvimento das metodologias, pela implementação computacional das análises realizadas, pelo senso de urgência, pela incansável busca pela perfeição e por sua disponibilidade em todos os momentos.

À Dra. Núbia Esteban pelo auxílio na busca de informações relativas aos dados do InCor.

Ao amigo Dr. David Schlesinger, pela oportunidade de trabalharmos juntos novamente, pelas discussões sempre construtivas, pelos *inputs* desde o início do projeto e pelo fundamental auxílio na escolha do projeto.

Ao Dr. Alexandre Pereira por sua disponibilidade e pelas rápidas respostas às solicitações.

Ao Laboratório de Genética e Cardiologia Molecular pelo fornecimento dos dados e, especialmente, à Priscila e Marcelly pelo fornecimento das informações solicitadas.

RESUMO

Introdução: O número de estudos relacionados à Síndrome Metabólica (SM) vem aumentando nos últimos anos, muitas vezes motivados pelo aumento do número de casos de sobrepeso/obesidade e diabetes Tipo II levando ao desenvolvimento de doenças cardiovasculares e, como consequência, infarto agudo do miocárdio e AVC, dentre outros desfechos desfavoráveis. A SM é uma doença multifatorial composta de cinco características, porém, para que um indivíduo seja diagnosticado com ela, possuir pelo menos três dessas características torna-se condição suficiente. Essas cinco características são: Obesidade visceral, caracterizada pelo aumento da circunferência da cintura, Glicemia de jejum elevada, Triglicérides aumentado, HDL-colesterol reduzido, Pressão Arterial aumentada.

Objetivo: Estabelecer a rede de associações entre os fenótipos que compõem a Síndrome Metabólica através do aprendizado de estruturas de dependência, decompor a rede em componentes de correlação genética e ambiental e avaliar o efeito de ajustes por covariáveis e por variantes genéticas exclusivamente relacionadas à cada um dos fenótipos da rede.

Material e Métodos: A amostra do estudo corresponderá a 79 famílias da cidade mineira de Baependi, composta por 1666 indivíduos. O aprendizado de estruturas de redes será feito por meio da Teoria de Grafos e Modelos de Equações Estruturais envolvendo o modelo linear misto poligênico para determinar as relações de dependência entre os fenótipos que compõem a Síndrome Metabólica.

Palavras-chave: Síndrome Metabólica, Aprendizado de Estruturas, Grafos Não Direcionados, Grafos Acíclicos Direcionados, Propriedades de Markov, SNPs, Dados de Famílias, Modelo misto poligênico.

ABSTRACT

Introduction: The number of studies related to Metabolic Syndrome (MetS) has been increasing in the last years, encouraged by the increase on the overweight / obesity and Type II Diabetes cases, leading to the development of cardiovascular disease and, therefore, acute myocardial infarction and stroke, and others unfavorable outcomes. MetS is a multifactorial disease containing five characteristics, however, for an individual to be diagnosed with MetS, he/she may have at least three of them. These characteristics are: Truncal Obesity, characterized by increasing on the waist circumference, increasing on Fasting Blood Glucose, increasing on Triglycerides, decreasing on HDL cholesterol and increasing on Blood Pressure.

Aims: Establish the best association network between MetS phenotypes through structured dependency learning between phenotypes considering genetic variants exclusively related to each phenotype.

Materials and Methods: The study sample is composed of 79 families, 1666 individuals of a city in a rural area of Brazil, called Beapendi. Structured learning will use graph theory and Structural Equations Models to establish the dependency relations between MetS phenotypes.

Key Words: Metabolic Syndrome, Structured Learning, Undirected Graphs, Acyclic Directed Graphs, Structural Equation Models, Markov Properties, SNPs, Family Data.

SUMÁRIO

1 INTRODUÇÃO 25
1.1 SÍNDROME METABÓLICA NO CONTEXTO DA SAÚDE GLOBAL 25
1.2 FATORES DE RISCO E CONSEQUÊNCIAS DA SÍNDROME METABÓLICA 29
1.3 APRENDIZADO DE ESTRUTURAS 32
1.3.1 Modelos de Grafos no estudo da SM 33
1.3.2 Etapas do Estudo 40
1.4 GRAFO TEÓRICO 44

2 OBJETIVO 48
2.1 OBJETIVOS ESPECÍFICOS 48

3 MATERIAIS E MÉTODOS 49
3.1 MATERIAIS 49
3.2 METODOLOGIA 52
3.2.1 Embasamento Teórico 52
3.2.2 Condução do Trabalho 62
3.2.3 Fluxograma para a construção de redes de aprendizado de estruturas 66

4 RESULTADOS 67
4.1 RESULTADOS MLMP_MULTIVARIADO 74
4.1.1 Correlações parciais para o componente Poligênico 74
4.1.2 Correlações parciais para o componente Ambiental 75
4.1.3 Correlações Totais 76
4.2 RESULTADOS MLMP_UNIVARIADO 77
4.2.1 Correlações parciais para o componente Poligênico 77
4.2.2 Correlações parciais para o componente Ambiental 78
4.2.3 Correlações totais 79
4.3 RESULTADOS PC ALGORITMO 80
4.4 GRAFOS 80

5 VALIDAÇÃO DOS RESULTADOS 89

6 DISCUSSÃO 94

7 CONCLUSÕES 99

CRONOGRAMA 101

REFERÊNCIAS 102

ANEXO 1 110

ANEXO 2 120

APÊNDICE 1 125

APÊNDICE 2 131

APÊNDICE 3 133

APÊNDICE 4 140

APÊNDICE 5 176

APÊNDICE 6 179

APÊNDICE 7 184

LISTA DE FIGURAS

Figura 1: Grafo Não Direcionado (UDG)

Figura 2: Correspondência entre Grafos e Modelagem de Equações Estruturais

Figura 3: Quatro DAGs para três variáveis. Os dois primeiros (a e b) são chamados de cadeias; c é um garfo; d é um colisor. Se esses fossem grafos completos, ter-se-ia: Para os três primeiros $X \not\perp Y$ e $X \perp Y|Z$, para o colisor (d) $X \perp Y$ enquanto $X \not\perp Y|Z$. Em que $\not\perp$ significa que não são independentes

Figura 4: DAG considerando Aleatorização Mendeliana

Figura 5: Diagrama proposto da SM - Extraída de Muller, 2011

Figura 6: Diagrama proposto da SM - Diagrama proposto da SM - Extraído de Blann, 2010

Figura 7: Diagrama proposto da SM - Extraído de Karns, 2013.

Figura 8: Grafo Teórico da Síndrome Metabólica.

Figura 9: Exemplo de um heredograma

Figura 10: Exemplo de matriz de parentesco da família da figura 9

Figura 11: Esquema para inclusão de Covariáveis e SNPs

Figura 12: Fluxograma para construção de redes de aprendizado de estruturas.

Figura 13: Distribuição de frequências do número de membros de cada família na amostra de estudo.

Figura 14: Heredograma de uma família pertencente ao estudo

Figura 15: Matriz ϕ para a família representada pela Figura 14

Figura 16: Matriz 2ϕ para a família representada pela Figura 14

Figura 17: Histograma da Circunferência Abdominal

Figura 18: Histograma da Glicemia

Figura 19: Histograma de HDL-c

Figura 20: Histograma de Triglicérides

Figura 21: Histograma de PAS

Figura 22: Histograma de PAD

Figura 23: Grafos obtidos para as três abordagens (MLMP_Multivariado, MLMP_Univariado e Algoritmo PC) nas três etapas (sem covariáveis, com

covariáveis sexo, idade, PC1 e PC2 e com covariáveis sexo, idade, PC1, PC2 e SNPs)
Figura 24: Grafos obtidos a partir de indivíduos considerados fundadores usando-se as três metodologias e as três abordagens.
Figura 25: SNPs relativos a Obesidade Truncal
Figura 26: SNPs relativos a Glicose
Figura 27: SNPs relativos ao Triglicérides
Figura 28: SNPs relativos a cHDL
Figura 29: SNPs relativos a média da PAS
Figura 30: SNPs relativos a média da PAD
Figura 31: Grafos relativos às três abordagens (MLMP_Multivariado, MLMP_Univariado e Algortimo PC) para dois estudos feitos durante a Etapa 2 com a inclusão das covariáveis sexo e idade e sexo, idade e IMC.
Figura 32: Grafos comparativos entre as execuções do Algoritmo PC utilizando modo *Conservative* e modo *Major Rule*
Figura 33: Grafos comparativos utilizando 1 SNP por fenótipo e mais de 1 SNP por fenótipo através da execução do Algoritmo PC utilizando o modo Major Rule

LISTA DE TABELAS

Tabela 1: Dados da amostra comparados com os dados do Censo 2010 para a cidade de Baependi
Tabela 2: Dados das Famílias
Tabela 3: Estatística descritiva dos fenótipos da SM
Tabela 4: Herdabilidade dos fenótipos da SM
Tabela 5: SNPs mais significantes ($p \leq 10\text{-}7$) do banco de Baependi utilizados nas análises
Tabela 6: SNPs mais significantes de dados extraídos da literatura utilizados nas análises
Tabela 7: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângulo superior) e com covariáveis sexo, idade, PC1 e PC2 (triângulo inferior), considerando α = 5% para o caso sem correção e com correção por FDR.
Tabela 8: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casos sem correção com correção por FDR.
Tabela 9: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis (triângulo superior) e com covariáveis sexo, idade, PC1 e PC2 (triângulo inferior), considerando α = 5% para os casos sem correção e com correção por FDR.
Tabela 10: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.
Tabela 11: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior) considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.
Tabela 12: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.
Tabela 13: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2,

(triângular inferior) considerando α = 5% para os casos sem correção e com correção por FDR.
Tabela 14: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR.
Tabela 15: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis e com covariáveis sexo, idade, PC1 e PC2, considerando α = 5% para os casos sem correção e com correção por FDR.
Tabela 16: Correlações parciais do componente ambiental entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR.
Tabela 17: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior), considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.
Tabela 18: Correlações parciais totais entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR.
Tabela 19: Lista de SNPs significantes: Obesidade Truncal
Tabela 20: Lista de SNPs significantes: Glicose
Tabela 21: Lista de SNPs significantes: Triglicérides
Tabela 22: Lista de SNPs significantes: cHDL
Tabela 23: Lista de SNPs significantes: Média PAS
Tabela 24: Lista de SNPs significantes: Média PAD
Tabela 25: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo de 20.67 para o MLMP_Multivariado, componente genético, sem covariáveis (Oualkacha, 2012).
Tabela 26: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, sem covariáveis
Tabela 27: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, sem covariáveis (Oualkacha, 2012)

Tabela 28: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, sem covariáveis

Tabela 29: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, sem covariáveis (Yang, 2011)

Tabela 30: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, sem covariáveis

Tabela 31: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, sem covariáveis (Oualkacha, 2012)

Tabela 32: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, sem covariáveis

Tabela 33: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, sem covariáveis (Yang, 2011)

Tabela 34: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, sem covariáveis

Tabela 35: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo e idade (Oualkacha, 2012)

Tabela 36: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo e idade

Tabela 37: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade (Oualkacha, 2012)

Tabela 38: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade

Tabela 39: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade (Yang, 2011)

Tabela 40: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade

Tabela 41: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo e idade (Oualkacha, 2012)
Tabela 42: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo e idade
Tabela 43: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo e idade (Yang, 2011)
Tabela 44: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo e idade
Tabela 45: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade e IMC (Oualkacha, 2012)
Tabela 46: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade e IMC
Tabela 47: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC (Oualkacha, 2012)
Tabela 48: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC
Tabela 49: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC (Yang, 2011)
Tabela 50: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC
Tabela 51: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC (Oualkacha, 2012).
Tabela 52: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total com covariáveis sexo, idade e IMC
Tabela 53: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC (Yang, 2011)

Tabela 54: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC

Tabela 55: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012)

Tabela 56: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1 e PC2

Tabela 57: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012)

Tabela 58: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

Tabela 59: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2 (Yang, 2011)

Tabela 60: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

Tabela 61: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012).

Tabela 62: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2

Tabela 63: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2 (Yang, 2011)

Tabela 64: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2

Tabela 65: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi) (Oualkacha, 2012)

Tabela 66: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs(Baependi)

Tabela 67: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi) (Oualkacha, 2012)

Tabela 68: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 69: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi) (Yang, 2011)

Tabela 70: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 71: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi) (Oualkacha, 2012).

Tabela 72: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 73: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi) (Yang, 2011)

Tabela 74: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 75: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura) (Oualkacha, 2012)

Tabela 76: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 77: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura) (Oualkacha, 2012)

Tabela 78: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 79: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura) (Yang, 2011)

Tabela 80: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 81: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura) (Oualkacha, 2012).

Tabela 82: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 83: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura) (Yang, 2011)

Tabela 84: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 85: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, sem covariáveis

Tabela 86: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, sem covariáveis

Tabela 87: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, sem covariáveis
Tabela 88: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, sem covariáveis
Tabela 89: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, sem covariáveis
Tabela 78: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, sem covariáveis
Tabela 79: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado sem covariáveis
Tabela 80: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado sem covariáveis
Tabela 81: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo e idade
Tabela 82: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo e idade
Tabela 83: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo e idade
Tabela 84: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo e idade
Tabela 85: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo e idade
Tabela 86: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo e idade
Tabela 87: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo e idade
Tabela 88: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo e idade
Tabela 89: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade e IMC
Tabela 90: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade e IMC
Tabela 91: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade e IMC

Tabela 92: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade e IMC
Tabela 93: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade e IMC
Tabela 94: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade e IMC
Tabela 95: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade e IMC.
Tabela 96: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade e IMC.
Tabela 97: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1 e PC2
Tabela 98: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1 e PC2
Tabela 99: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2
Tabela 100: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2
Tabela 101: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1 e PC2
Tabela 102: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1 e PC2
Tabela 103: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1 e PC2
Tabela 104: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1 e PC2.
Tabela 105: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)
Tabela 106: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 107: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 108: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 109: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 110: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 111: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 112: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

Tabela 113: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 114: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 115: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 116: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 117: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 118: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 119: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 120: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

Tabela 121: Matriz de Correlações Parciais para o componente poligênico para o MLMP_Multivariado com as covariáveis sexo, idade e IMC

Tabela 122: Matriz de Correlações Parciais para o componente ambiental para o MLMP_Multivariado com as covariáveis sexo, idade e IMC

Tabela 123: Matriz de Correlações Parciais Total para o MLMP_Multivariado com as covariáveis sexo, idade e IMC

Tabela 124: Matriz de Correlações Parciais para o componente poligênico para o MLMP_Univariado com as covariáveis sexo, idade e IMC

Tabela 125: Matriz de Correlações Parciais para o componente ambiental para o MLMP_Univariado com as covariáveis sexo, idade e IMC

Tabela 126: Matriz de Correlações Parciais Total para o MLMP_Univariado com as covariáveis sexo, idade e IMC

LISTA DE QUADROS

Quadro 1: Critério Diagnóstico proposto para diagnóstico clínico de SM
Quadro 2: Variáveis Principais e Secundárias
Quadro 3: SNPs relacionados a Circunferência da Cintura
Quadro 4: SNPs relacionados a Glicemia
Quadro 5: SNPs relacionados a Dislipidemia
Quadro 6: SNPs relacionados a Pressão Arterial

1 INTRODUÇÃO

Esta seção apresentará a Síndrome Metabólica (SM) no Contexto da Saúde Global (Seção 1.1), em seguida trará uma visão geral sobre as causas e consequências associadas à SM (Seção 1.2) e, ainda, é apresentada uma abordagem sobre o Aprendizado de Estruturas de Redes, envolvendo a Teoria de Grafos e as Propriedades de Markov (Seção 1.3).

1.1 SÍNDROME METABÓLICA NO CONTEXTO DA SAÚDE GLOBAL

Nos últimos 50 anos, o mundo vem acompanhando o crescimento tanto da obesidade quanto de Diabetes Tipo 2. Em 2005, a estimativa era de que havia 190 milhões de pessoas no mundo com Diabetes Tipo 2 e até 2025 essa estimativa estaria em 324 milhões de pessoas (Zimmet, 2005). A SM torna-se, então, um sério problema de saúde pública e um desafio clínico no mundo com o crescimento da urbanização, consumo excessivo de energia, aumento da obesidade e hábitos de vida sedentários (Kaur, 2014).

A síndrome metabólica é uma desordem multifatorial complexa que pode ser definida de várias maneiras (Eckel, 2005). As definições mais comumente usadas são da Organização Mundial de Saúde (OMS), do grupo Europeu para o estudo da resistência à Insulina (EGIR), o do National Cholesterol Education Programme Adult Treatment Panel III (NCEP ATP III), da American Association of Clinical Endocrinologists (AACE), e do International Diabetes Federation (IDF) (Quadro 1) (Kaur, 2014).

Quadro 1: Critério Diagnóstico proposto para diagnóstico clínico de SM.

Medida Clínica	WHO (1998)	EGIR (1999)	ATP III (2001)	AACE (2003)	IDF (2005)
Resistência à Insulina	IGT, IFG, T2DM ou baixa sensibilidade à insulina[a] **Mais quaisquer dois abaixo**	Insulina plasmática > 75 percentil **Mais quaisquer dois abaixo**	Nenhuma, mas 3 das 5 abaixo	IGT ou IFG **Mais quaisquer das seguintes baseada em julgamento clínico**	Nenhuma
Peso Corporal	Homem: razão cintura/quadril > 0,90 Mulher: razão cintura/quadril > 0,85 e/ou IMC>30kg/m^2	CA ≥ 94 cm Homens e ≥ 80 cm Mulheres	CA ≥ 102 cm Homens e ≥ 88 cm Mulheres	IMC ≥ 25 kg/m^2	Aumento de CA (específico por população mais 2 dos seguintes)
Lipídeos	TGs ≥ 150mg/dL e/ou HDL-c < 35 mg/dL Homens e < 39 mg/dL Mulheres	TGs ≥ 150mg/dL e/ou HDL-c < 35 mg/dL (Homens e Mulheres)	TGs ≥ 150mg/dL HDL-c < 40 mg/dL Homens e < 50 mg/dL Mulheres	TGs ≥ 150mg/dL HDL-c < 40 mg/dL Homens e < 50 mg/dL Mulheres	TGs ≥ 150mg/dL Ou em tratamento medicamentoso HDL-c < 40 mg/dL Homens e < 50 mg/dL Mulheres ou em tratamento medicamentoso
Pressão Sanguínea	≥ 140/90 mmHg	≥ 140/90 mmHg	≥ 130/85 mmHg	≥ 130/85 mmHg	≥ 130 mmHg para sistólica ou ≥ 85 mmHg para diastólica ou em tratamento medicamentoso
Glicose	IGT, IFG ou T2DM	IGT ou IFG, mas não diabético	> 110 mg/dL (inclui diabéticos)	IGT ou IFG (mas não diabetes)	≥ 100 mg/dL (inclui diabéticos)[b]
Outros	Microalbuminúria: Taxa de Secreção urinária >20mg/min ou razão albumina/ creatinina >30mg/g.			Outras características de Resistência à Insulina[c]	

a Sensibilidade à Insulina medida sob condições hiperinsulinêmicas euglicêmicas, tomada da glicemia abaixo do menor quartil para população sob investigação.
b Em 2002, a Associação Americana de Diabetes (ADA) mudou o critério para tolerância de > 100 mg/dL para > 100 mg/dL.
c Inclui histórico familiar de diabetes tipo 2 mellitus, síndrome do ovário policístico, estilo de vida sedentário, idade avançada, e suscetibilidade do grupo étnico a diabetes tipo 2 mellitus.
IMC: Índice de massa corporal; IFG: glicemia de jejum; IGT: tolerância à glicose diminuída; TGs: triglicérides; T2DM: diabetes tipo 2 mellitus; Ca: circunferência abdominal. Fonte: Kaur, 2014

Apesar de as definições possuírem características comuns, existem diversos parâmetros que diferem e dificultam em temos de aplicabilidade,

uniformidade, etc. As definições de AACE, WHO e EGIR são mais focadas em resistência à insulina (Kaur, 2014). Porém, ATP III possui uma definição que utiliza medidas e resultados laboratoriais prontamente disponíveis para os médicos, facilitando sua aplicação.

No entanto, tanto a definição da OMS quanto da NCEP APT III enfrentam problemas relacionados aos grupos étnicos, especialmente no estabelecimento de cortes para a obesidade. Por essa razão, a IDF propôs um critério com cortes específicos para grupos étnicos (Kaur, 2014). Porém, esse tópico não será aprofundado nesse estudo.

Vale ressaltar que os cortes estabelecidos para todos os fenótipos da SM, não importando qual definição utilizada, são artificiais, e sofrem alterações com relativa frequência.

Mundialmente, entre a população adulta, constata-se uma elevada prevalência de SM, variando de 25% a 35%, com maior frequência em mulheres. Reflexo do aumento das prevalências dos componentes que a definem na população mundial (Bortoletto, 2014).

A estimativa da prevalência da SM pode variar bastante dependendo da região, ambiente rural ou urbano, composição da população estudada (sexo, idade, raça e etnia) e da definição da SM usada. Status socioeconômico elevado, estilo de vida sedentário, alto IMC estão significativamente associados com SM (Kaur, 2014).

Cameron et al concluíram que as diferenças do passado genético, dieta, níveis de atividade física, tabagismo, histórico familiar de diabetes, e escolaridade influenciam as prevalências da SM e de seus componentes. A prevalência observada da SM na Pesquisa de Saúde Nacional e Exame de Nutrição (NHANES) foi de 5% para indivíduos de peso normal, 22% com sobrepeso, e 60% em obesos. A prevalência aumenta também com a idade (10% em indivíduos entre 20 e 29 anos, 20% em indivíduos entre 40 e 49 anos e 45% em indivíduos entre 60 e 69 anos. A prevalência de SM (critério NCEP ATP III) variou de 8 a 43% em homens e de 7 a 56% em mulheres no mundo. Park et al observaram que existe um aumento na prevalência de indivíduos desde os 20 anos até a sexta e sétima década de vida para homens e mulheres, respectivamente. Ponholzer et al reportaram que existe alta prevalência de SM em mulheres pós menopausa, que varia de 32,6% a 41,5%. O estudo Framingham Heart Study indicou que um aumento de peso maior ou igual a 2,25 kg em um período de 16 anos está associado com um aumento em 45% do

risco de desenvolver SM e foi mostrado por Palaniappan et al que a cada aumento de 11 cm na Circunferência Abdominal está associado com um aumento do risco ajustado em 80% de desenvolver a síndrome dentro de 5 anos. As alterações metabólicas ocorrem simultaneamente de forma mais frequentemente do que o esperado e a concomitância de vários fatores aumenta o risco cardiovascular acima e além dos riscos associados com cada fator isolado. O risco aumenta com o número de componentes da SM presentes.

Bortoletto et al (2014) efetuaram uma revisão sistemática a partir das bases PubMed, Scielo, Lilacs e outras fontes, de janeiro de 2000 a junho de 2013, compreendendo estudos transversais com amostras representativas que apresentassem a SM segundo uma das definições a seguir: International Diabetes Federation (IDF); Third Report of the National Cholesterol Education Program Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol (ATPIII), com o intuito de traçar um perfil de prevalência da SM no contexto nacional (Brasil) podendo traçar estratégias de controle dos fatores de risco para Doenças Cardiovasculares, nas quais a SM está inserida. A prevalência de SM variou de 7,1% (adultos jovens – 23 anos) a 56,9% (idosos – 60 anos ou mais) observando-se aumento da SM com a idade.

Penalva (2008) também fez um levantamento de estudos nacionais sobre a SM. Em 2007, Salaroli et al. realizaram um estudo transversal em Vitória-ES, com 1.663 indivíduos com idades de 25 a 64 anos e mostraram que a prevalência da SM, de acordo com os critérios da ATPIII, é de quase 30% para a população geral, sendo maior com o avançar da idade: 15,5% na faixa entre 25 e 34 anos e 48,3% na faixa entre 55 e 64 anos. Também em 2007, foi publicado um estudo transversal de Marquezine et al, com 1.561 indivíduos de uma área urbana, que mostrou uma prevalência de SM de 25,4% nesta população estudada, sendo cada vez maior com o avançar da idade, especialmente em mulheres, e com a piora do nível sócio-econômico. Nakazone et al, em 2007, realizaram um estudo para analisar o perfil bioquímico e caracterizar a SM a partir de critérios propostos por ATPIII e IDF, com o intuito de verificar a predisposição para doença cardiovascular em 340 indivíduos (200 pacientes e 140 controles). A prevalência de SM no grupo de pacientes foi de 35,5%, segundo os critérios na ATPIII, e de 46%, segundo os critérios da IDF. Marcondes et al (2007) realizaram um estudo transversal com o objetivo de determinar a prevalência, características e preditores da síndrome metabólica em 73

mulheres com a síndrome dos ovários policísticos. A prevalência da síndrome metabólica foi de 38,4%, estando ausente nas mulheres com índice de massa corporal normal (n = 18) e presente em 23,8% das com sobrepeso (n = 17), 62,9% das obesas (n = 28) e 85,5% das obesas mórbidas (n = 7). Quando comparadas, as mulheres com síndrome metabólica apresentaram uma idade média mais avançada (27,3 anos (desvio-padrão=5,3) vs. 24,2 (4,6); p = 0,031) e um índice de massa corporal maior (36,3 (7,7) vs. 26,9 (5,4); p < 0,001) que as mulheres sem a síndrome (Penalva, 2008).

1.2 FATORES DE RISCO E CONSEQUÊNCIAS DA SÍNDROME METABÓLICA

A SM tem como principais consequências o desenvolvimento de Doenças Cardiovasculares, Diabetes Tipo 2 Mellitus e Acidente Vascular Cerebral. A SM aumenta em 5 vezes o risco de Diabetes Tipo 2 Mellitus, em 2 vezes o risco de Doenças Cardiovasculares nos 5 a 10 anos subsequentes. Além disso, pacientes com SM tem de 2 a 4 vezes mais risco de Acidente Vascular Cerebral (AVC), 3 a 4 vezes mais risco de infarto do miocárdio e 2 vezes mais risco de morrer desses eventos quando comparado àqueles sem a síndrome (Kaur, 2014).

Os fatores de risco são traços, condições ou hábitos que aumentam a chance de desenvolvimento da SM. Tratam-se de cinco fatores que desencadeiam a síndrome. Um indivíduo pode ter apenas um fator isolado, mas eles tendem a ocorrer em conjunto. Para que um indivíduo seja considerado com SM deve apresentar pelo menos três dos cinco fatores (NIH, 2016).

Um grande comprimento de cintura, denominada obesidade abdominal ou “formato de maçã” é um dos fatores de risco. Gordura em excesso da região estomacal é um alto fator de risco para doenças cardiovasculares quando comparado com excesso de gordura em outras partes do corpo, como os quadris, por exemplo. Um alto nível de triglicérides (ou sob tratamento medicamentoso para trata-lo) é outro fator de risco para SM. Triglicérides são um tipo de gordura encontrada no sangue. Um baixo HDL colesterol (ou sob tratamento medicamentoso para tratá-lo) também é um fator de risco para doenças cardiovasculares e, portanto, para o desenvolvimento da SM. O HDL é chamado de bom colesterol, porque ajuda

na remoção do colesterol das artérias. Alta pressão sanguínea (ou sob tratamento medicamentoso para tratá-la) compõe outro fator de risco para a SM. Pressão sanguínea é a força do sangue empurrando contra as paredes das artérias conforme o coração o bombeia. Se essa pressão aumenta e permanece alta, pode danificar o coração e levar a acúmulo de placa. Alta taxa de açúcar no sangue em jejum (ou em tratamento) pode ser um primeiro sinal de diabetes (NIH, 2016).

Com isso, o risco de desenvolver doenças cardiovasculares, diabetes e AVC aumenta com o número de fatores de risco que o indivíduo possui. O risco de desenvolver SM é bastante vinculado ao sobrepeso / obesidade e falta de atividade física (NIH, 2016).

A resistência à insulina também aumenta o risco de SM. A resistência à insulina é uma condição em que o corpo não consegue usar sua insulina de forma apropriada. A insulina é um hormônio que ajuda a mover o açúcar sanguíneo para dentro das células onde é usado para fornecer energia. Resistência à insulina pode levar ao aumento dos níveis de açúcar no sangue, o que está intimamente relacionado com sobrepeso e obesidade. Fatores genéticos (etnia e histórico familiar) e idade avançada são outros fatores que também desempenham papel com agentes causadores da SM (NIH, 2016).

SM tem se tornado mais comum devido ao aumento das taxas de obesidade em adultos. No futuro, SM pode superar o tabagismo como fator líder para doenças cardiovasculares. É possível prevenir ou retardar o desenvolvimento da SM, principalmente a partir de mudanças no estilo de vida. Um estilo de vida saudável é um compromisso de vida longa (NIH, 2016).

Assim, apesar de ser difícil determinar qual condição é a iniciadora do processo da Síndrome Metabólica, acredita-se que a Gordura Visceral / Abdominal, medida obtida a partir da circunferência abdominal, seja o fator pivotal das demais condições. Ela apresenta valores acima do corte quando existe, no mínimo, sobrepeso, de tal forma que o fator obesidade / sobrepeso pode contribuir para a resistência à insulina.

A SM, portanto, corresponde a um conjunto de fatores que agem em conjunto. É importante dizer que os indivíduos conseguem controlar algumas dessas causas como, por exemplo, sobrepeso e obesidade, estilo de vida sedentário e resistência à insulina. Porém, outros fatores que podem ser causadores da síndrome não são necessariamente controláveis como, por exemplo, o envelhecimento, visto

que o risco de SM aumenta com a idade. Também não é possível controlar a genética (fatores étnicos e histórico familiar), que também desempenham papel no desenvolvimento da SM. Outros possíveis fatores de risco que são alvo de estudos são gordura no fígado, Síndrome do Ovário Policístico, Cálculos Biliares e Apneia do Sono (NIH, 2016)

As pessoas com maior risco de desenvolverem SM possuem as seguintes características: obesidade abdominal, estilo de vida sedentário, resistência à insulina. Algumas pessoas apresentam maior risco por tomarem medicamentos que aumentam o peso ou alteram a pressão arterial, colesterol e níveis de açúcar no sangue. Esses medicamentos muitas vezes são usados para tratar inflamações, alergias, HIV e depressão e outros tipos de doenças mentais.

Em relação a etnias, alguns grupos raciais e étnicos nos Estados Unidos têm maior risco de SM do que outros. México Americanos têm a mais alta taxa de SM, seguida por brancos e negros (NIH, 2016)

Outros grupos com risco aumentado são as que possuem histórico familiar de diabetes, pessoas que têm um irmão ou pai/mãe com diabetes, mulheres quando comparadas com homens, mulheres com histórico de Síndrome do Ovário Policístico.

Pelas razões acima apresentadas, esse estudo tem como intuito principal entender a maneira pela qual os fenótipos da SM se relacionam entre si, evidenciando dessa forma, as relações de causalidade entre elas, na tentativa de demonstrar o(s) fator(es) pivotal(is), suas consequências, além de se tornar uma ferramenta clínica na tentativa de “atacar” os precursores no sentido de prevenção.

Dessa forma, a metodologia que será aplicada nesse estudo é o Aprendizado de Estruturas, que será introduzido a seguir.

1.3 APRENDIZADO DE ESTRUTURAS

Atualmente, vem crescendo na literatura a análise de doenças multifatoriais complexas, como é o caso da SM, caracterizando tanto a base ambiental de regulação da doença, como também sua base genética. Avaliar variação quantitativa mensurável em fenótipos relacionados com riscos de doenças, fornece mais poder para localizar e identificar genes suscetíveis relacionados a doenças do que pelos resultados obtidos através de exame clínicos. Um dos principais benefícios de se examinar variação quantitativa em fenótipos fisiológicos é que tanto indivíduos normais quanto indivíduos afetados contribuem com a informação genética (Blangero, 2003).

Estratégias poderosas de análise, como o mapeamento genético de doenças utilizando modelos de componentes de variância, que serão tratados neste estudo, têm sido desenvolvidas e amplamente utilizadas para localizar “genes” (SNPs) relacionados a doenças humanas em estudos com famílias estendidas através do escaneamento do genoma humano (Blangero, 2003). Contudo, além do mapeamento de genes, a informação genética pode ser usada para estabelecer possíveis relações causais entre variáveis (Burguess, 2014)

Assim, um dos pontos chave deste trabalho é analisar vários fenótipos simultaneamente fazendo uso de informação genética para o entendimento de doenças multifatoriais. Inicialmente, a descoberta de associações e correlações entre essas variáveis e, finalmente, o aprendizado das possíveis relações de causa e efeito entre elas são de fundamental importância no diagnóstico e tratamento das doenças relacionadas.

Em termos de metodologias estatísticas, os modelos de equações estruturais (SEM, do inglês S*tructural Equation Model*) têm contribuído para com o estudo das relações de causa entre variáveis à medida que se tem um diagrama de caminhos conhecido, ou que se conhece a relação existente entre as variáveis (Bollen, 1989). Para esses casos, parte-se de um modelo conhecido em que seu ajuste é feito a partir dos dados obtidos. Portanto, o ajuste do modelo é feito através de evidências amostrais. Contudo, quando não se conhece a estrutura das possíveis relações causais entre variáveis (o diagrama de caminhos), como é o caso do presente estudo, busca-se encontrar, estimar e aprender as relações estruturais entre variáveis de forma simultânea partindo-se de um conjunto de dados gerados a

partir de estudos observacionais. Através desses estudos, tem-se como objetivo construir a rede de relacionamentos e validar que as relações são possivelmente de causa e não espúrias.

Entretanto, pesquisadores consideram como “padrão-ouro” em termos de estabelecimento de causas os Ensaios Clínicos Randomizados Controlados em que se procura garantir que os “tratamentos” (as variáveis causais de um desfecho) são atribuídos de forma aleatória às unidades experimentais, consideradas homogêneas relativamente aos possíveis fatores de confundimento. Porém, quando se trata de estudos observacionais, não existe nenhuma garantia de que essas premissas estejam satisfeitas. Muitos esforços têm sido dedicados para a realização de inferências causais a partir de estudos observacionais. Destaca-se a teoria de Grafos (Pearl, 2009) que, inspirada em lições aprendidas dos Ensaios Clínicos Randomizados Controlados, tem permitido a implementação de algoritmos que permitem extrair relações orientadas entre variáveis livres de possíveis confundimentos. Adicionalmente, uma contribuição importante em termos de estudos observacionais tem sido dada por Burgess (2014), que introduziu o conceito de Aleatorização Mendeliana e o uso de variantes genéticas (SNPs) como variáveis instrumentais nas redes de associação entre variáveis fenotípicas. Sob essa estratégia de modelagem de dados pode-se mimetizar, isto é, estabelecer o mesmo padrão de relação entre variáveis e desfechos que os obtidos a partir dos Ensaios Clínicos Randomizados Controlados.

Assim, com o intuito de se construir redes de associação entre os fenótipos da SM, no presente estudo, combina-se SEM e teoria dos Grafos, juntamente com Aleatorização Mendeliana, como um esforço para a realização de inferência causal a partir de dados de estudos observacionais (Ribeiro et al, 2016).

1.3.1 Modelos de Grafos no estudo da SM

Devido ao fato de a SM ser caracterizada por uma desordem multifatorial complexa, uma forma interessante para que se possa compor as dependências condicionais para se obter as distribuições conjuntas, e verificar como as variáveis se relacionam entre si, ou não se relacionam, estabelecendo associações, correlações e possíveis relações causais dizem respeito aos Modelos dos Grafos.

O aprendizado de estruturas trata de uma metodologia em que se aprende um grafo (orientado) no qual se resume as estruturas de dependência entre variáveis de um conjunto de dados (Drton, M e Maathuis, MH, 2016). Tais variáveis, no contexto desse estudo, são os fenótipos da SM.

Um grafo é um Diagrama que contém vértices e arestas. Os vértices podem ou não estar conectados com as arestas e estas podem ter direção de um vértice para outro, ou não. A Figura 1 ilustra um grafo não direcionado.

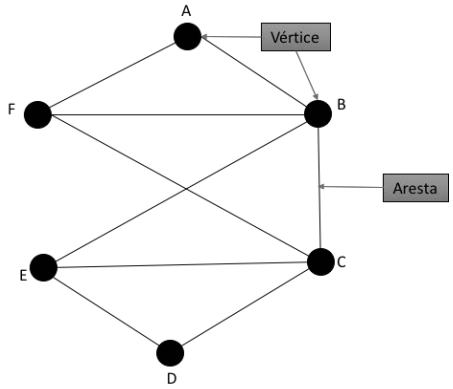


Figura 1: Grafo Não Direcionado (UDG)

Matematicamente, o grafo G é representado por G = (V,E), em que V são os vértices e E as arestas. Os vértices correspondem a variáveis ou fenótipos da SM.

No caso específico deste estudo, um comentário sobre os grafos deve ser feito, no que diz respeito ao fenótipo Hipertensão. No contexto da SM, este fenótipo é representado por uma elevação em conjunto da PAS e da PAD. Para efeito da determinação dos Grafos, esses dois fatores são dissociados, ou seja, a Hipertensão é formada por 2 vértices, PAS e PAD, nos grafos aprendidos.

Os grafos podem ser direcionados ou não direcionados (UDG), e cíclicos (DCG) ou acíclicos (DAG). Apesar de se saber que na SM podem ocorrer ciclos entre os fenótipos, para esse estudo assume-se a premissa de que os grafos, quando direcionados, são acíclicos.

Grafos não direcionados (UDG) são aqueles em que há conexão entre os vértices, mas não é estabelecida uma possível relação causal, não se tem um ordenamento das relações entre as variáveis. Já para os grafos direcionados estabelece-se posssíveis relações causais entre os vértices e essas são representadas por setas, indicando o sentido da relação descoberta, ou seja, de

onde a seta sai tem-se a possível causa e para onde ela aponta tem-se a possível consequência. Acíclico é o termo usado para explicar que não existe o estabelecimento de um ciclo fechado entre variáveis.

Para que se possa construir as redes de associação entre os fenótipos da SM, será feita a correspondência entre SEM (Modelagem de Equações Estruturais) e os Modelos de Grafos. Para que isso seja possível, assume-se a condição de Markov e a Condição de Fidelidade, que serão brevemente explicadas adiante, e mais detalhadamente no Anexo 1, além do mesmo ser feito para o conceito de d-separação e de independência condicional (Ribeiro et al, 2016)

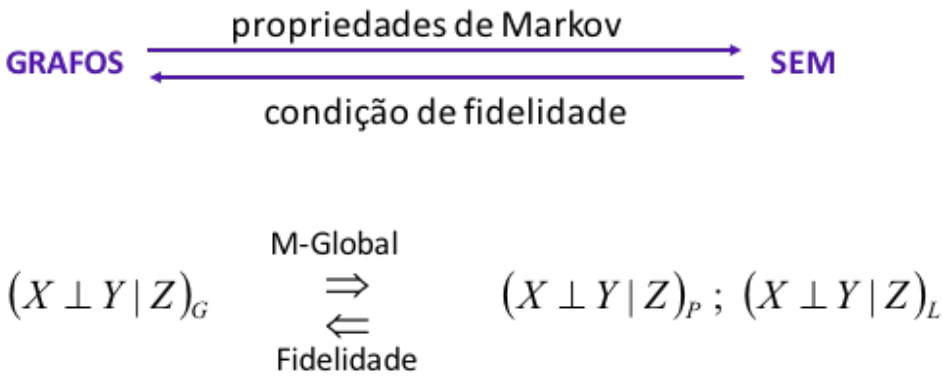


Figura 2: Correspondência entre Grafos e Modelagem de Equações Estruturais

Este estudo não chegará até a determinação das equações estruturais relacionadas aos diagramas de caminhos estabelecidos, porém esses diagramas serão construídos na forma dos grafos. Assim, a correspondência entre Teoria dos Grafos e Modelagem de Equações Estruturais seguirá como uma proposta de estudo futuro e um breve resumo teórico desta correspondência segue no Anexo 2.

Abaixo segue um resumo teórico sobre as premissas que precisam ser assumidas para a construção das redes de associação.

1.3.1.1 Independência Condicional e Propriedades de Markov

O principal conceito para que se possa aprender os grafos é o conceito probabilístico de independência condicional, em que se tem a independência de um fenótipo em relação a outro, dados os demais fenótipos. Ou seja, $(X \perp Y)|Z$, em que X representa um fenótipo, Y representa outro fenótipo e Z representa os fenótipos remanescentes (Shalize, 2015).

Assim, suponha duas variáveis, X_i e X_j, em que X_j não é um pai de X_i. Considere a distribuição de X_i condicionada em seus pais e X_j. Existem duas possibilidades:

i) X_j é um não-descendente de X_i. Então, pode-se notar que X_i e X_j são condicionalmente independentes. Isso é verdadeiro não importando quais funções de distribuição condicional estão envolvidas; isto é somente implicado pela distribuição conjunta em relação ao grafo e, neste caso, não existe uma conexão entre X_i e X_j no grafo (Shalize, 2015);

ii) Alternativamente, X_j é descendente de X_i, então, eles não são independentes, mesmo se condicionado nos pais de X_i, portanto, existe uma conexão entre eles no grafo (Shalize, 2015).

Então o grafo implica algumas relações em que a independência condicional é sustentada e outros em que não.

A propriedade de Markov utilizada para os grafos sugere que, ao se tomar uma coleção de vértices (V), eles são independentes de seus não-descendentes, dados seus pais:

$$X_v \perp X_{\text{não-descendentes}(V)} | X_{\text{pais}(V)} \quad (3)$$

Obs.: Por este trabalho tratar de um estudo envolvendo dados de famílias, o uso dos termos pais, filhos e descendentes na modelagem dos grafos pode gerar algum tipo de confusão. Para fins de clarificação, no caso específico da construção dos grafos, a relação entre filhos e pais pretende demonstrar relações causais entre os fenótipos, de modo que pais vêm antes dos filhos e, portanto, a seta aponta neste sentido. Porém, este resultado não tem relação com a estrutura familiar da amostra.

1.3.1.2 Independência Condicional e d-Separação

Torna-se muito importante ser capaz de deduzir quando dois conjuntos de variáveis são condicionalmente independentes entre si dado um terceiro conjunto. Um dos principais usos dos DAGs é que eles fornecem um critério relativamente simples para isso, em termos do próprio grafo. Todas as distribuições que estão em um DAG compartilham um conjunto de relações de independência condicional,

implicadas pelas propriedades de Markov, em geral, não importando quais os parâmetros ou formas de suas distribuições (Shalize, 2015).

A Figura 3, abaixo, mostra quatro grafos direcionados e as relações entre as variáveis.

Figura 3: Quatro DAGs para três variáveis. Os dois primeiros (a e b) são chamados de cadeias; c é um garfo; d é uma colisão. Se esses fossem grafos completos, ter-se-ia: Para os três primeiros $X \not\perp Y$ e $X \perp Y|Z$, para a colisão (d) $X \perp Y$ enquanto $X \not\perp Y|Z$. Em que $\not\perp$ significa que não são independentes.

Em todos os casos da Figura 3, condiciona-se X, que funciona como fonte de informação. Nos três primeiros casos, pode-se (em geral) propagar a informação de X para Z para Y. No último grafo, no entanto, que é chamado de colisor, não se pode propagar a informação porque Y não tem pais, e X não é seu descendente, portanto, são independentes (Shalize, 2015).

As cadeias (a e b), condicionadas nas variáveis intermediárias, bloqueiam o fluxo de informação de X para Y. Isso também é verdade no garfo (c) – condicionado em sua causa comum, os dois efeitos são não informativos sobre os outros. Mas no caso do colisor, condicionando no efeito comum Z faz com que X e Y sejam dependentes um do outro. De fato, se não se condicionar em Z, mas condicionar em um descendente de Z, também se cria dependência entre os pais de Z (Shalize, 2015).

Considerando-se o conceito de independência condicional em que se tem duas variáveis, X e Y, e condiciona-se ambas em um terceiro conjunto de variáveis S. Se S bloquear todos os caminhos não direcionados de X para Y, então eles devem ser condicionalmente independentes dado S. Um caminho não bloqueado também é chamado ativo. Um caminho é ativo quando todas as variáveis ao longo do caminho são ativas; se somente uma variável for bloqueada por S, todo o caminho é bloqueado. Uma variável Z ao longo do caminho é ativa, condicionada em S, se:

1. Z está na colisão ao longo do caminho, e em S; ou

2. Z é um descendente de uma colisão, e em S; ou
3. Z não está na colisão, e não está em S.

De maneira inversa, Z é bloqueado ou desativado pelo condicionamento em S se:

1. Z não está na colisão e está em S; ou
2. Z está na colisão, e nem Z ou nenhum de seus descendentes está em S.

Assim, S bloqueia um caminho quando bloqueia o fluxo de informação pelo condicionamento no vértice intermediário em uma cadeia ou garfo, e não cria dependência pelo condicionamento no vértice intermediário de uma colisão (ou descendente de uma colisão). Apenas um vértice em um caminho precisa estar bloqueado para bloquear todo o caminho. Quando S bloqueia todos os caminhos entre X e Y, diz-se que ele os d-separa. Um conjunto de variáveis U é d-separado de um outro conjunto V por S se todo $X \in U$ e $Y \in V$ forem d-separados. Em todas as distribuições que obedecem à Propriedade de Markov, d-separação implica independência condicional. Não é sempre o caso em que a implicação reversa está satisfeita, isto é, a de que independência condicional implica d-separação. Porém, quando a distribuição é “fiel” a um DAG, inferência causal é imensamente simplificada. Mas, sair de d-separação para independência condicional é verdade em qualquer DAG, tendo ou não interpretação causal (Shalize, 2015).

1.3.1.3 Fidelidade e Equivalência

Em modelos de Grafos, d-separação implica independência condicional: se S bloqueia todos os caminhos de U a V, então $U \perp V|S$. Para reverter isso, e concluir que se $U \perp V|S$ então S deve d-separar U e V, precisamos de uma premissa adicional chamada Fidelidade. Mais exatamente, se a distribuição é fiel ao grafo, então se S não d-separa U de V, $U \not\perp V|S$. A combinação de fidelidade e Propriedade de Markov significa que $U \perp V|S$ se e somente se S d-separa U e V. Esse resultado é extremamente promissor no aprendizado de grafos. Pode-se testar a

independência condicional, se $U \perp V|S$, para qualquer conjunto de variáveis que queremos. Podemos, em particular, testar se cada par de variáveis é independente, dadas todas as formas de condicionamento de conjuntos S. Se assumirmos fidelidade, quando encontrarmos que $U \perp V|S$, saberemos que S bloqueia todos os caminhos que ligam X a Y, então aprendemos algo sobre o grafo. Se $U \not\perp V|S$ para todo S, tem-se poucas escolhas, mas conclui-se que X e Y estão diretamente conectados. Contudo, há situações em que não é possível reconstruir ou descobrir o DAG correto pelo conhecimento de todas as independências condicionais e relações de dependência, como será mostrado abaixo:

$$X \rightarrow Y \Rightarrow X \not\perp Y \tag{4}$$

$$X \leftarrow Y \Rightarrow X \not\perp Y \tag{5}$$

Com apenas duas variáveis, existe somente uma relação de independência (ou dependência) para se preocupar, e é a mesma não importando para qual sentido a seta aponta.

Similarmente, considere os seguintes arranjos de três variáveis:

$$X \rightarrow Y \rightarrow Z \tag{6}$$

$$X \leftarrow Y \leftarrow Z \tag{7}$$

$$X \leftarrow Y \rightarrow Z \tag{8}$$

$$X \rightarrow Y \leftarrow Z \tag{9}$$

As primeiras duas são cadeias, a terceira um garfo e a última uma colisão. Não é difícil verificar que os três primeiros DAGs implicam no mesmo conjunto de relações de independência condicional, que são diferentes dos implicados pelo quarto (Shalize, 2015).

Esses exemplos ilustram um problema geral. Pode haver múltiplos grafos que implicam as mesmas relações de dependência, mesmo quando se assume fidelidade. Quando isso ocorre, exatamente a mesma distribuição de observações pode fatorar de acordo com e ser fiel a todos os grafos. Os grafos são então ditos equivalentes, equivalentes de Markov. Observação que, sozinha, não consegue distinguir entre DAGs equivalentes (Shalize, 2015).

1.3.2 Etapas do Estudo

O estudo foi construído em três Etapas para a determinação dos Modelos dos Grafos. O objetivo é partir do modelo mais simplificado até ser atingido um Modelo mais complexo e, portanto, final. A seguir são apresentadas as três etapas.

1.3.2.1 Etapa 1: Matriz de Precisão (Σ^{-1}) - Σ6x6

Esta etapa será simbolicamente representada por Σ6x6, pois a análise se baseia principalmente na estimativa da matriz de covariância fenotípica, uma matriz 6x6 no caso das variáveis da SM, e de sua inversa.

Dispondo dos fenótipos da SM será feito um estudo com os indivíduos e suas famílias cujo objetivo será obter uma estrutura de associação entre esses fenótipos, decomposta em relações de natureza genética e ambiental. Ou seja, essas análises irão mostrar se existe associação entre cada um dos fenótipos, gerando inicialmente Grafos Não Direcionados (UDG). Para gerar Grafos Acíclicos Direcionados (DAG) será executado o Algoritmo PC, cujo funcionamento será brevemente explicado abaixo, para a mesma base de dados, ou seja, os vértices (fenótipos) poderão ser conectados por setas estabelecendo uma relação orientada entre os fenótipos.

Para que seja possível evoluir no estudo e aprender estruturas da rede de fenótipos mais livres de efeitos de confundimento, na Etapa 2 serão utilizadas covariáveis/variáveis.

1.3.2.2 Etapa 2: Σ6X6 + Covariáveis

Da mesma maneira que na etapa anterior, utiliza-se a simbologia Σ6X6 + Covariáveis para representar que esta análise será baseada na estimativa da matriz de covariância fenotípica usando a informação de covariáveis.

A Etapa 2 compreende a adição de covariáveis, com o objetivo de explorar modelos mais gerais de associação e de direcionamento das conexões, estabelecendo a rede de associações menos influenciada por heterogeneidades da amostra.

1.3.2.3 Etapa 3: Σ6X6 + Covariáveis + SNPs

Da mesma maneira que na etapa anterior, utiliza-se a simbologia Σ6X6 + Covariáveis + SNPs já que a análise está baseada principalmente na estimativa da matriz de covariância fenotípica usando a informação adicional de SNPs como variáveis instrumentais.

No contexto deste estudo a variante genética (SNP) será a variável instrumental que tem por principal função eliminar possíveis efeitos de confundimento no aprendizado de um grafo direcionado (DAG).

A introdução da variante genética como variável instrumental busca induzir uma relação causal da variante genética para o fenótipo. Além disso, a introdução de no mínimo uma variável instrumental para cada fenótipo, e do fato de os erros gerados serem independentes, pode-se gerar uma solução única para o sistema, ou seja, definir um grafo acíclico direcionado (DAG) ou, minimamente, parcialmente direcionado.

Conforme o Teorema da Descoberta, assumindo-se algumas premissas, é possível aprender a estrutura causal, tomando-se como base a d-separação e, portanto, assumir as condições M_Global e Fidelidade.

As premissas comumente feitas são: suficiência causal, em que todas as causas comuns do sistema causal foram observadas e que os erros envolvidos são conjuntamente independentes; que a distribuição gerada pela estrutura causal satisfaz a propriedade M_Global; e a fidelidade causal, em que todas as relações de independência condicional presentes em um DAG são consequência de M_Global, implicando na obtenção de um Grafo fiel à estrutura causal verdadeira.

Isso significa que a introdução do SNP como variável instrumental associada à cada fenótipo da SM faz com que se obtenha um DAG (ou um grafo parcialmente direcionado) mais livre de fatores de confundimento.

É importante salientar o papel desempenhado pela Aleatorização Mendeliana no aprendizado de DAGs. A introdução das variantes genéticas (SNPs) como variáveis instrumentais exerce a função de eliminar os fatores de confundimento para possibilitar o estabelecimento de relações causais entre as variáveis (Burgess, 2014).

A Figura 4 ilustra um DAG, em que M é a variante genética atuando como variável instrumental, X e Y são fenótipos quantitativos e Z é causa comum e um possível fator de confundimento.

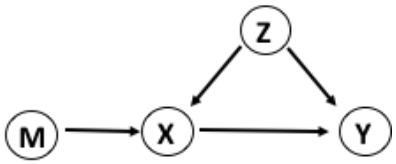


Figura 4: DAG considerando Aleatorização Mendeliana.

Uma variante genética usada como variável instrumental em uma análise de Aleatorização Mendeliana não precisa necessariamente estar relacionada de forma causal com o fenótipo (variável X), mas pode ser um representante da verdadeira variante causal. Qualquer variante em desequilíbrio de ligação (linkage disequilibrium – LD) com a variante causal que satisfaça as premissas de variável instrumental pode ser usada como uma variável instrumental (Burgess, 2014).

Dessa maneira, no presente estudo, a obtenção dos SNPs como variável instrumental, ocorre de duas formas:

- Revisão Bibliográfica

A partir de revisão da literatura foi obtida uma lista de SNPs relacionados com as variáveis da SM. Essa lista (apresentada no Apêndice 2) será confrontada com os SNPs presentes no banco de dados de Baependi pela sua “notação”. A identificação dos SNPs que serão usados como variáveis instrumentais poderá ser feita diretamente, no caso de notação concordante, ou via locos que estão no banco de dados de Baependi e em desequilíbrio de ligação com aqueles da lista.

- Mapeamento

Uma outra alternativa para se encontrar os SNPs a serem usados como variáveis instrumentais, será a busca no próprio banco de dados de Baependi através do Mapeamento Genético de cada um dos fenótipos da SM. Para esta finalidade o ajuste uniloco do Modelo Linear Misto Poligênico será considerado para cada um dos fenótipos da SM, e os SNPs significantes serão selecionados via o gráfico Manhattan.

1.3.2.4 Algoritmo PC

O algoritmo utilizado neste estudo é o pacote PC Algorithm, do software R. PC advém do nome de seus autores, Peter (Spirtes) e Clack (Glymour), e foi proposto em 2000, sendo um algoritmo com o mesmo princípio que o SGS (Spirtes, Glymour and Scheines, 2000), porém funciona de maneira mais rápida, principalmente pelo fato de fazer menos testes estatísticos. O PC Algorithm tem sido o mais indicado para descoberta causal (Shalize, 2015).

Segue abaixo uma breve explicação do funcionamento de ambos.

Os algoritmos iniciam com um grafo acíclico não direcionado completo (UDG), ou seja, todos os vértices estão conectados. Em cada etapa, ele visita todos os pares (i, j) de vértices adjacentes no grafo, e determina, baseado em testes de independência condicional, se a aresta i-j deve ser removida. Em particular, na etapa m (m=0, 1, 2,...) o algoritmo visita todos os pares (i,j) de vértices adjacentes no grafo atual, e a conexão entre i e j é mantida se e somente se a hipótese nula “i e j são condicionalmente independentes dado S” for rejeitada ao nível de significância α para todos os subconjuntos S de tamanho m das vizinhanças de i e das vizinhanças

de j no grafo. O algoritmo pára quando m é maior do que o maior tamanho de vizinhança de todos os vértices, ou quando m atingiu o limite m.max definido pelo usuário (Shalize, 2015).

De forma abstrata, o algoritmo funciona como segue:

- Comece com um grafo completamente não direcionado em todas as p variáveis, com arestas entre todos os vértices.
- Para cada par de variáveis X e Y, e cada conjunto de variáveis S, veja se $X \perp Y|S$; em caso positivo, remova a aresta entre X e Y.
- Encontre colisores checando para dependência condicional; oriente as arestas dos colisores.
- Tente orientar arestas não orientadas pela consistência com arestas já orientadas; faça isso de maneira recursiva até que não hajam mais arestas para serem orientadas.

1.4 GRAFO TEÓRICO

Atualmente existem alguns trabalhos envolvendo o diagrama causal entre os fenótipos da SM. Porém, muitas vezes, esses trabalhos envolvem outros fenótipos ou marcadores que vão além do objetivo deste estudo. Além disso, vale ressaltar que os estudos abaixo mencionados e sobre os quais o grafo teórico foi proposto se basearam em resultados obtidos a partir de estudos envolvendo indivíduos independentes sem qualquer informação de estrutura familiar.

O presente estudo tem como contribuição inovadora o fato de estar aprendendo a estrutura de dependência entre os fenótipos da SM tomando como base dados de famílias, de modo a permitir uma decomposição entre grafos genéticos e grafos ambientais.

Deve-se esclarecer, portanto, que os grafos obtidos podem diferir do modelo proposto justamente em decorrência do tipo de dados sob estudo.

A seguir, as três figuras (5, 6 e 7) trazem possíveis caminhos até que se chegue à SM.

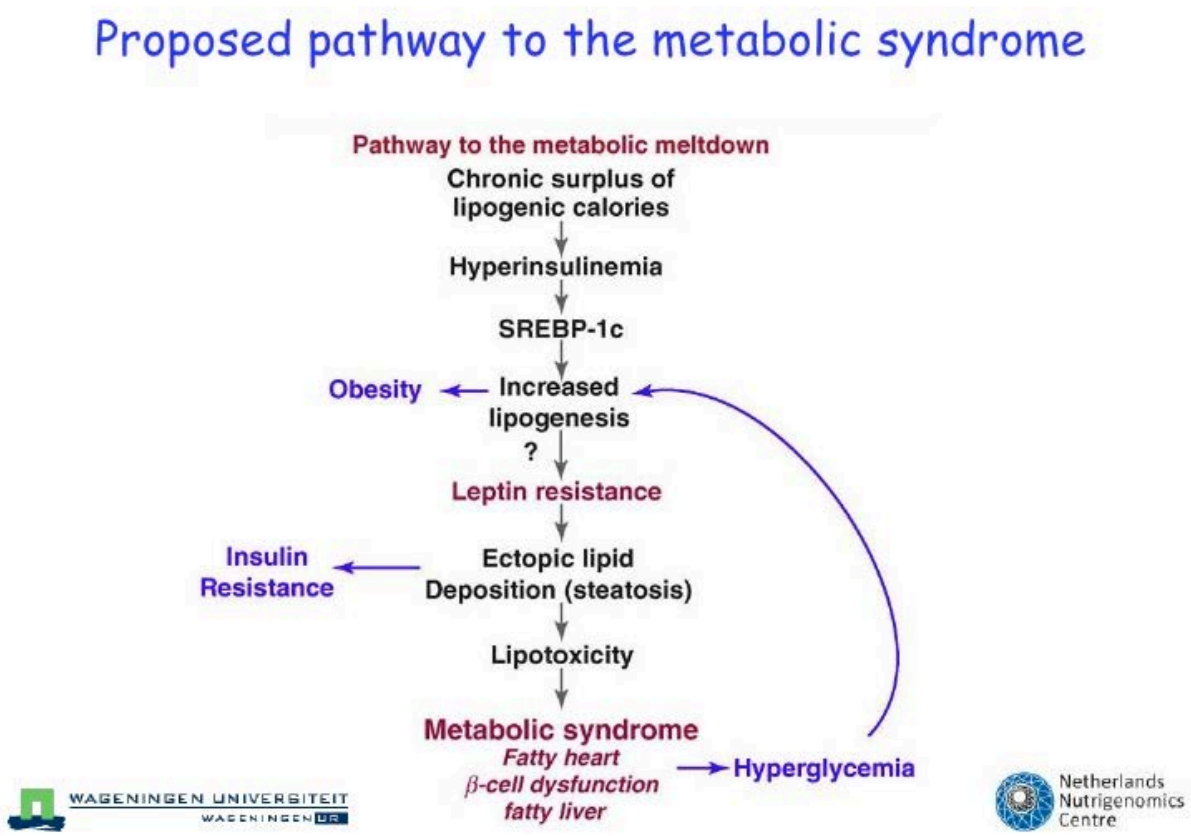


Figura 5: Diagrama proposto da SM - Extraída de Muller, 2011

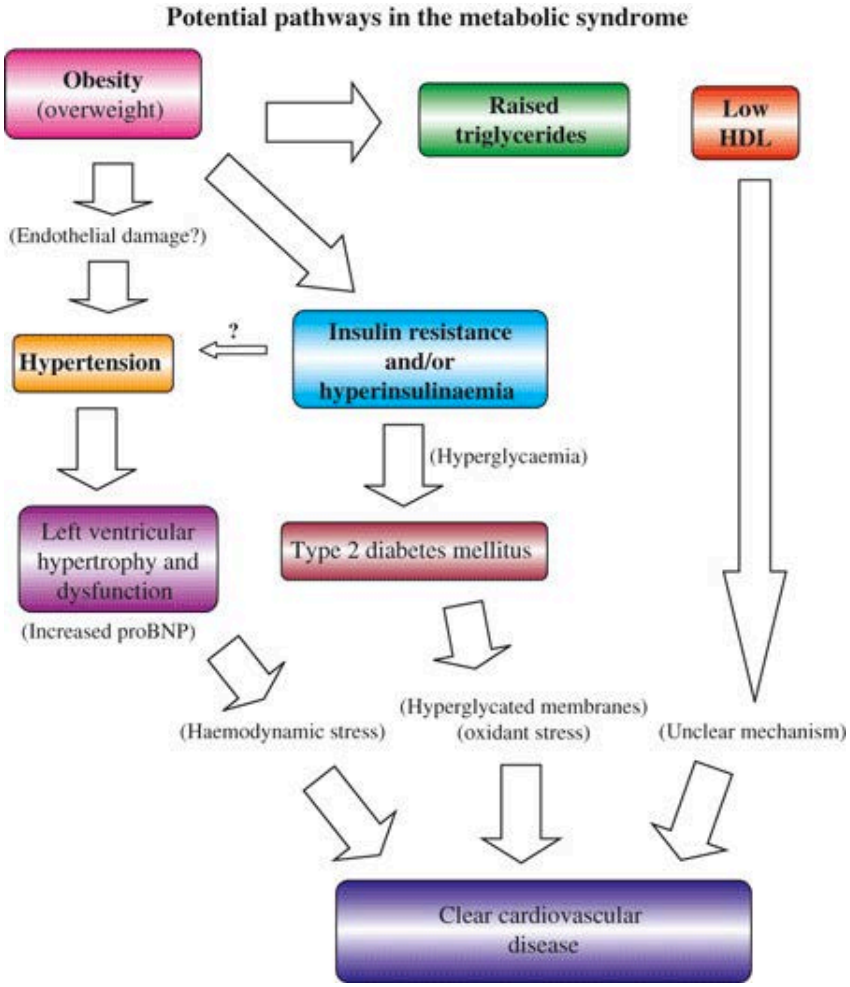
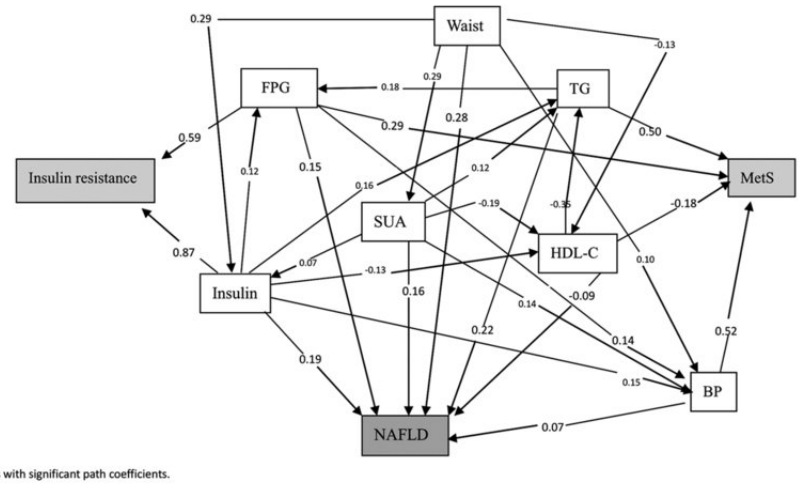


Figura 6: Diagrama proposto da SM - Extraído de Blann, 2010



Figures are presented as standardized path coefficients with corresponding p-value <.0.05
DWLS (diagonally weighted least squares) χ^2=57.367, df=14, p<0.001;
CFI (comparative fit index) =0.998; TLI (Tucker-Lewis fit index)=0.992; RMSEA (root mean square error of approximation) =0.048.
BP, blood pressure; FPG, fasting plasma glucose; HDL-C, high-density lipoprotein cholesterol; MetS, metabolic syndrome; NAFLD, non-alcoholic fatty liver disease; SUA, serum uric acid; TG, triglyceride.

Figura 7: Diagrama proposto da SM - Extraído de Karns, 2013.

Extraindo-se as informações dos três diagramas acima apresentados (Figuras 5, 6 e 7) contruiu-se um diagrama de caminhos, ou grafo, teórico. Para tornar essa construção possível, informações de vértices relativos a outros fenótipos diferentes dos aqui estudados, foram removidos, considerando-se um caminho contínuo entre os vértices da SM. A Figura 5 apresenta um modelo mais simplificado que se parte da ingestão calórica, que aumenta a produção de insulina, resultando em Obesidade, que por sua vez leva à resistência à insulina desencadeando a Síndrome. A Figura 6 parte da Obesidade, levando ao aumento da pressão, dos triglicérides, redução do HDLc, resistência à insulina, culminando em Diabetes Tipo 2 e Doenças Cardiovasculares. A Figura 7 apresenta um diagrama mais completo com outros fenótipos que não pertencem à SM, mas que estão a ela relacionados.

Com base no grafo teórico abaixo (Figura 8), os resultados obtidos da análise dos dados de Baependi serão comparados com esses.

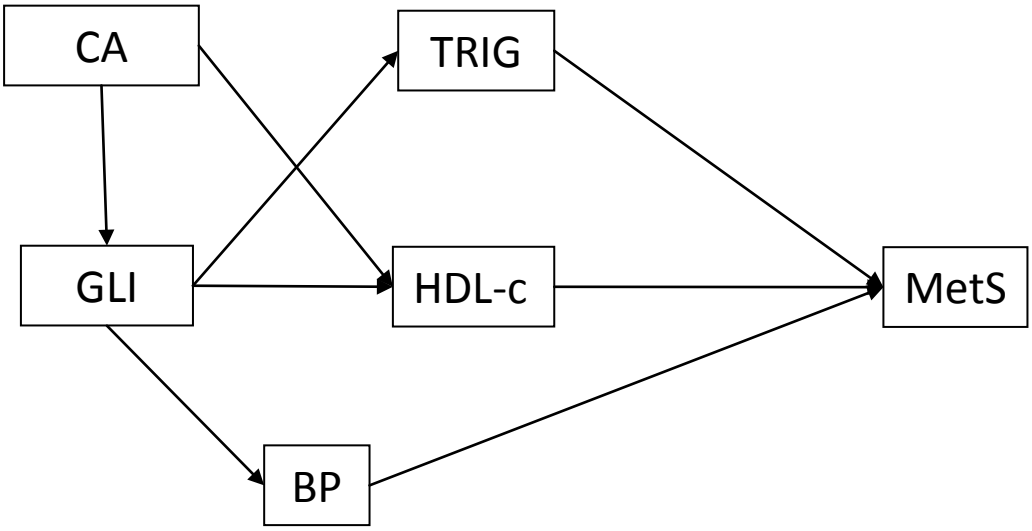


Figura 8: Grafo Teórico da Síndrome Metabólica.

No Grafo Teórico da Figura 8 pode-se dizer que o fator inicial é a Circunferência Abdominal (CA), que está relacionada como causa da Glicemia e do HDL-c. A Glicemia por sua vez seria um fator causal de Hipertensão (BP), Triglicérides e agiria em conjunto com a Circunferência Abdominal na causa do HD-c.

2 OBJETIVO

Estabelecer as associações entre os fenótipos que compõem a SM e introduzir genótipos, além de covariáveis, para contribuir com o estabelecimento de associações causais válidas entre as variáveis, por meio do aprendizado de estruturas de dependência entre os fenótipos da SM em estudos de famílias.

2.1 OBJETIVOS ESPECÍFICOS

- Construir modelos de grafos (UDG → DAG) para os fenótipos da SM;
- Introduzir de variáveis genéticas (SNPs) como variáveis instrumentais no aprendizado do Grafo da SM;
- Aprimorar a rede de associações introduzindo outras covariáveis;
- Analisar as associações entre os fenótipos que compõem a SM sob a ótica de relações causais.

3 MATERIAIS E MÉTODOS

Trata-se de um estudo de coorte transversal, tendo como desfecho de interesse o estabelecimento de uma rede de associações entre fenótipos da SM.

3.1 MATERIAIS

Este estudo fará uso dos dados gerados a partir do Projeto colaborativo liderado e gerenciado pelo Laboratório de Genética e Cardiologia Molecular do Instituto do Coração da Universidade de São Paulo, InCor/USP, denominado Projeto Corações de Baependi (Egan, 2016).

O protocolo de estudo original, em conformidade com os princípios da Declaração de Helsinki, foi submetido ao Comitê de Ética do Hospital das Clínicas, da Universidade de São Paulo (USP) e aprovado pela CAPPESQ (Comissão de Ética para Análise de Projetos de Pesquisa, do HCFMUSP) sob o número 0494/10. E o presente projeto foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Saúde Pública da USP, sob o número 66148217.8.0000.5421.

O Projeto Corações de Baependi teve início em 2005 como um estudo longitudinal baseado em uma coorte de famílias que refletia algumas peculiaridades genéticas e de estilo de vida da população brasileira para avaliar influencias genéticas e ambientais em fatores de risco para doenças cardiovasculares. O estudo focou na População de Baependi, uma cidade em uma área rural (752 km^2, 18.307 habitantes de acordo com o Censo de 2010) localizada no Estado de Minas Gerais, Brasil. Trata-se de uma comunidade tradicional, com cultura coesa e alto grau de mistura (interna) e migração bastante limitada. Os indivíduos participantes do estudo foram identificados da comunidade em 11 distritos censitários (de 12 possíveis). Endereços residenciais de cada distrito foram selecionados aleatoriamente

(primeiramente selecionando aleatoriamente a rua, e em seguida a casa na rua). Uma vez que o indivíduo nesta casa, com 18 anos ou mais, foi cadastrado, todos os seus parentes de primeiro grau (pais, irmãos ou descendentes), parentes de segundo grau (meio-irmãos, avós / netos, tios / tias, sobrinhos / sobrinhas e primos duplos) e de terceiro grau (primos, tios-avós e tias-avós, filhos / filhas de sobrinhos) e parentes do cônjuge, que tinham no mínimo 18 anos foram convidados a participar. Depois do primeiro contato com o indivíduo, parentes de primeiro grau foram convidados a participar por telefone; o que incluiu todos os parentes que viviam no município de Baependi e além. Um questionário geral foi inicialmente administrado para cada participante por um técnico treinado para cobrir atributos básicos como relações familiares e características demográficas. Para participantes analfabetos ou deficientes visuais, os questionários foram lidos em sua totalidade pelo pesquisador que estava escrevendo. Quando a capacidade de um indivíduo não permitia que ele completasse uma questão específica, ajuda era fornecida pelo pesquisador ou cuidador. A investigação também usou o instrumento epidemiológico WHO MONICA que já havia sido utilizado com sucesso em outros projetos epidemiológicos. As seguintes informações foram coletadas: estilo de vida geral e saúde, saúde cardiovascular, bioquímicos, antropométricos, saúde mental e sono e ritmo circadiano. Adicionalmente, amostras de DNA foram coletadas e foi usada a plataforma Affymetrix 6.0 arrays bem como dados foram imputados usando o 1000 Genomes Cosmopolitan Panel. Cada participante forneceu consentimento por escrito antes de participar (Egan, 2016).

A Tabela 1 a seguir mostra a distribuição da população de Baependi (dados do Censo de 2010) e da amostra (sem aplicar qualquer filtro aos dados) de acordo com a raça-cor. É possível observar que a amostra apresenta distribuição semelhante à da população do município a partir de dados do Censo Demográfico 2010 (Egan 2016).

Etnia	Baependi – Amostra		Baependi – Censo 2010	
	N	%	N	%
Brancos	1711	76,70	13669	74,70
Negros	124	5,60	1032	5,60
Mesclados	384	17,20	3554	19,40
Asiáticos	12	0,54	46	0,30
Indígenas	1	0,04	6	0,03

Tabela 1: Dados da amostra comparados com os dados do Censo 2010 para a cidade de Baependi. (Egan, 2016)

Em Oliveira et al (2008) foi feito um estudo sobre a herdabilidade dos fenótipos da SM utilizando como amostra os dados coletados na primeira onda do Projeto Corações de Baependi, e através dele foi possível verificar que a herdabilidade dos fenótipos associados a doenças cardiovasculares é elevada, e depende de covariáveis, como sexo e idade, principalmente. Quando comparados a outros estudos, tais como o Northern Manhattan Heart Study, o San Antonio Heart Study e o Framingham Heart Study, observa-se que os resultados obtidos são robustos tanto para a população brasileira quanto para a população global, com a ressalva de que para alguns casos a população de estudo já apresentava risco cardiovascular a priori, e por essa razão apresentou diferenças nos resultados.

Em Giolo et al (2011) trata-se dos aspectos da miscigenação e ancestralidade da população brasileira. Este estudo mostrou que a população brasileira descende dos Ameríndios, Africanos e Europeus, porém, o casamento entre indivíduos de diferentes origens étnicas exerce um importante papel na geração da variedade genética observada na população atual.

Geralmente os coeficientes de ancestralidade são estimados assumindo-se que os indivíduos não são correlacionados. Porém, quando estruturas familiares estão presentes e são ignoradas, pode-se ter resultados inválidos. Os resultados obtidos por de Andrade et al (2015) no trabalho sobre ancestralidade individual global mostraram que a estrutura de famílias desempenha um importante papel na estimação da ancestralidade global do indivíduo para famílias estendidas. Neste trabalho, fica clara a importância de se incorporar a estrutura familiar no cálculo da ancestralidade quando se trabalha com populações com alto grau de mistura, como é o caso da população brasileira. Em Pena et al (2011), também um estudo sobre ancestralidade da população brasileira, mostra-se a homogeneidade da ancestralidade da população brasileira nas diferentes regiões do país.

Dessa forma, considera-se que a amostra deste estudo, ou seja, selecionada a partir da população da cidade de Baependi, está de acordo com os padrões de ancestralidade da população brasileira.

As variáveis utilizadas neste estudo foram divididas em dois grupos: F = Fenótipos e C = Covariáveis. As covariáveis serão idade, sexo e ancestralidade. As ancestralidades foram extraídas do estudo de Andrade et al, 2015, através da

Análise dos Componentes Principais de Ancestralidade da População de Baependi. Dessa maneira, para este estudo, foram utilizados os Componentes Principais, PC1 e PC2 como covariáveis.

O Quadro 2 mostra as variáveis utilizadas neste estudo:

Fenótipos (F)	*Covariáveis (C)*
Circunferência da Cintura (cm)	Idade (anos)
Glicemia de Jejum (mg/dL)	Sexo (M/F)
HDL-colesterol (mg/dL)	Ancestralidade (PC1 e PC2)
Triglicérides (mg/dL)	
Pressão Arterial Sistólica (mmHg)	
Pressão Arterial Diastólica (mmHg)	

Quadro 2: Fenótipos e Covariáveis

Também em Oliveira et al (2008) essas variáveis são consideradas como fatores de risco cardiovascular.

3.2 METODOLOGIA

3.2.1 Embasamento Teórico

Muitos estudos envolvendo dados de famílias e os fenótipos da SM já foram e vêm sendo conduzidos na busca de relações genéticas que possam explicar possíveis relações entre os fenótipos (Kissebah, 2000; Souto, 2000; Blangero, 2003; Almasy, L e Blangero, J, 2010; Mei, 2010; Olvera, 2011; Almasy, 2012). Essas possíveis relações podem, muitas vezes, serem explicadas através da pleiotropia, ou seja, um mesmo conjunto de genes atuando sobre mais de um fenótipo.

O presente estudo buscou inspiração na teoria dos grafos para construir as redes de associação entre os fenótipos da SM, estratégia pioneira e inovadora na literatura no que diz respeito à pleiotropia.

Assim, para unificar a teoria de grafos e os modelos de equações estruturais lança-se mão das propriedades de Markov. No anexo é apresentada uma breve revisão de resultados importantes desta teoria baseados em Drton, M & Maathuis, MH (2016) e Ribeiro et al (2016). A ideia principal é a de que existe equivalência entre os modelos de grafos, densidades conjuntas e sua fatoração e o

modelo de equações estruturais. As propriedades de Markov permitem que essas equivalências sejam estabelecidas.

Além disso, considerando o uso de dados de famílias, o modelo estatístico usado para possibilitar o aprendizado das estruturas dos grafos será o modelo de componentes de variância poligênico. Primeiramente, tem-se o Modelo Linear Misto Poligênico Univariado (de Andrade, 1999, Almasy e Blangero, 1998 e Almasy e Blangero, 2010), ou seja, levando em conta somente um fenótipo e, em seguida, a extensão para o Modelo Linear Misto Poligênico Multivariado (Oualkacha, 2012 e de Andrade et al., 2015). A denominação de modelo multivariado corresponde às análises que envolvem múltiplas variáveis, ou seja, considera-se as relações entre os fenótipos (variáveis) da SM. Além disso, por se tratar de um estudo com dados de famílias, que envolve a modelagem das dependências entre os indivíduos, o termo multivariado também pode ser usado para estruturar os indivíduos. Assim, pode-se dizer que se trata de um estudo multi- multivariado, descrevendo dependências entre indivíduos e entre variáveis. Neste modelo a matriz de covariância do fenótipo é particionada em componentes de covariância poligênica, ou seja, atribuíveis ao conjunto de genes do genoma compartilhados entre indivíduos com grau de parentesco e componentes de variância devido ao erro, em geral, atribuíveis ao ambiente. Esta proposta se assemelha ao que ocorre nos estudos envolvendo pleiotropia, porém representada, neste estudo, através da decomposição dos grafos.

Este estudo considerará seis vértices como fenótipos característicos da Síndrome Metabólica (conforme ressalva feita sobre a Hipertensão): Circunferência Abdominal, Glicemia de Jejum, HDL-colesterol, Triglicérides, Pressão Arterial Sistólica e Pressão Arterial Diastólica.

Neste caso, tem-se p=6 e a matriz de observações *Y* corresponde aos seis fenótipos que caracterizam a SM avaliados em n indivíduos estruturados em famílias, sendo:

Y_1:= Circunferência abdominal

Y_2:= Glicemia de jejum

Y_3:= HDL-Colesterol

Y_4:= Triglicérides

Y_5:= Pressão Arterial Sistólica

Y_6:= Pressão Arterial Diastólica

Assume-se que Yi's são variáveis aleatórias com distribuição normal multivariada conforme será abordado mais adiante. Porém, deve-se considerar que podem existir situações em que a condição de normalidade não é sustentada para um fenótipo específico e, nesses casos, será aplicada a transformação logarítmica. Por outro lado, muitos dos resultados usados são robustos a desvios da normalidade, bastando que as relações entre as variáveis sejam lineares e os erros aleatórios envolvidos sejam independentes.

3.2.1.1 Propriedades de Markov:

Seja $X=(X_v: v \in V)$ um vetor aleatório indexado pelos vértices de um grafo não-direcionado G=(V,E).

X satisfaz a propriedade de Markov para pares de variáveis, denotada como M_Pares, em relação a G se:

$X_v \perp X_w | X_{V\setminus\{v,w\}}$ sempre que $\{v,w\} \notin E$. Ou seja, a variável X_v é condicionalmente independente de X_w dado $X_{V|\{v,w\}}$, o que significa que são condicionalmente independentes dadas todas as demais variáveis do sistema.

X satisfaz a propriedade de Markov local, denotada como M_Local, se:

$X_v \perp X_{V\setminus nbG(v)\cup\{v\}} | X_{nbG(v)}$ para todo $v \in V$, sendo $nb_G(v) = \{w \in V: \{w,v\} \in E\}$ um conjunto de vizinhos de v em G. M_Local analisa a vizinhança de cada fenótipo para avaliar se há ou não associação, ou seja, a presença de aresta entre as variáveis.

X satisfaz a propriedade de Markov global, denotada como M_Global, se:

$X_A \perp X_B | X_C$ para todos os trios de subconjuntos pareados disjuntos A, B, C $\subset$ V, isto é,

A propriedade M_Global engloba a M_Local que, por sua vez, engloba a M_Pares:

$$M_Global \Rightarrow M_Local \Rightarrow M_Pares$$

3.2.1.2 Modelo Linear Misto Poligênico Univariado (MLMP_Univariado)

O MLPM_Univariado diz respeito, primordialmente, a uma metodologia que envolve o modelo de Componentes de Variância que, por sua vez, corresponde a um particular modelo linear misto. Este tem sido adotado em estudos Genéticos na análise de dados de famílias e permite modelar a variância de um fenótipo em termos de componentes poligênicos e “ambientais” (devido ao erro), além de incorporar a dependência entre os indivíduos da amostra devido ao grau de parentesco. Na abordagem deste trabalho a denominação univariada refere-se à análise de um único fenótipo. Contudo, o ajuste deste modelo univariado (na estrutura da resposta) implica na suposição de um modelo multivariado para incorporar as dependências na estrutura dos indivíduos.

Em geral, no ajuste do modelo, assume-se que o fenótipo é uma variável aleatória normalmente distribuída e, assim, pode-se trabalhar com sua média e sua variância (Blangero, 2010), que são os parâmetros do modelo.

Pode-se escrever o modelo da seguinte forma (de Andrade e Almos, 1999; Almasy e Blangero, 1998):

$$Y_{f(n_f x 1)} \sim \mathcal{N}_{n_f}(1_f \mu; \Omega_f = 2\phi_f \sigma_g^2 + I_f \sigma_e^2) \qquad (10)$$

Em que:

$Y_{f(nfx1)}$ é o vetor de respostas da família f, de dimensão (n_fx1)

1_f é o vetor de 1’s de tamanho n_f

μ é a média do fenótipo

$2\phi_f$ é a matriz de parentesco da família f, de dimensão (n_fxn_f), assumida conhecida

I_f é a matriz identidade de dimensão (n_fxn_f)

Ω_f é a matriz de covariâncias de Y_f, de dimensão (n_fxn_f). (Os elementos da matriz são as covariâncias do fenótipo para cada par de indivíduos no conjunto de dados)

n_f é o número de indivíduos da família f

σ_g^2 é o componente da variância associado ao efeito aleatório genético (poligênico)

σ_e^2 é o componente da variância associado ao efeito aleatório devido ao erro

$\sigma_Y^2 = \sigma_g^2 + \sigma_e^2$ é a variância total do fenótipo.

Deste modo, no MLMP_Univariado (p=1) tem-se, para o fenótipo Y_{jf}, relativo ao indivíduo *j* da família *f*:

$$Y_{jf} = \mu + g_{jf} + e_{jf}. \tag{11}$$

Em que, g_{jf} é o efeito aleatório devido ao termo poligênico e e_{jf} é o efeito aleatório devido ao erro. Assume-se que os efeitos aleatórios g_{jf} e e_{jf} são não correlacionados e normalmente distribuídos com média zero e componentes de variâncias

$$\sigma_g^2 \text{ e } \sigma_e^2, \text{ respectivamente.} \tag{12}$$

O componente de erro é único para cada indivíduo, enquanto que o componente poligênico é compartilhado entre os indivíduos na proporção do coeficiente de relacionamento (*kinship*), tal que, a covariância entre os fenótipos para os indivíduos i e i' das famílias f e f', respectivamente, é dada por:

$$Cov(y_{if}, y_{i'f'}) = \begin{cases} \sigma_g^2 + \sigma_e^2, para\ i = i', f = f' \\ 2\phi_{ii'}\sigma_g^2, i \neq i', mas\ relacionados\ (f = f') \\ 0, para\ i \neq i'\ e\ não\ relacionados\ (f \neq f') \end{cases} \tag{13}$$

O parâmetro $2\phi ii'$ é o coeficiente de relacionamento entre os indivíduos i e i', cujo valor pode ser obtido diretamente da declaração de parentesco ou ainda

estimando de dados genômicos como SNPs. Neste trabalho a matriz de parentesco para as famílias foi construída conforme declaração de parentesco. A Figura 9 mostra um exemplo de heredograma, que é um tipo de diagrama que pode descrever as características de uma família em termos de ascendentes e descendentes, de uma família com 7 membros e três gerações. A Figura 10 apresenta a correspondente matriz de parentesco para esta família. A distribuição conjunta dos fenótipos dos membros da família é assumida como seguindo uma distribuição normal multivariada (expressão 11). Estimativas da média e dos componentes de variância são obtidos através de métodos de máxima verossimilhança (de Andrade e Amos, 1999; Almasy e Blangero, 1998).

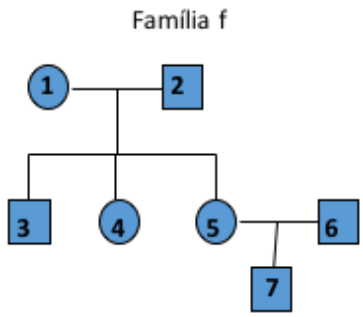


Figura 9: Exemplo de um heredograma

$$2\phi = \begin{array}{c|ccccccc} & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 1 & 1 & 0 & ½ & ½ & ½ & 0 & ¼ \\ 2 & 0 & 1 & ½ & ½ & ½ & 0 & ¼ \\ 3 & ½ & ½ & 1 & ½ & ½ & 0 & ¼ \\ 4 & ½ & ½ & ½ & 1 & ½ & 0 & ¼ \\ 5 & ½ & ½ & ½ & ½ & 1 & 0 & ½ \\ 6 & 0 & 0 & 0 & 0 & 0 & 1 & ½ \\ 7 & ¼ & ¼ & ¼ & ¼ & ½ & ½ & 1 \end{array}$$

Figura 10: Exemplo de matriz de parentesco da família da Figura 9.

O conceito de herdabilidade é bastante importante em Genética e, sob a formulação do modelo MLMP, representa o coeficiente de correlação intra-classe definido como:

$$h^2 = \frac{\sigma_g^2}{\sigma_Y^2}, \tag{14}$$

ou seja, a herdabilidade (h^2) é a proporção da variância fenotípica total que é explicada pelo componente poligênico, sendo assim uma medida da agregação familiar fenotípica.

3.2.1.3 Modelo Linear Misto Poligênico Multivariado (MLMP_Multivariado)

Após ter definido o MLMP_Univariado, pode-se fazer a extensão para o MLMP_Multivariado.

Neste caso, o que se deseja é, além de modelar a estrutura de dependências entre os indivíduos através da matriz $2\phi_f$, modelar a estrutura de dependência entre múltiplos fenótipos. Se sob o MLMP univariado tem-se dois componentes de variância (σ_g^2 e σ_e^2), sob a formulação do MLMP multivariado tem-se duas matrizes de covariância, sendo uma a matriz de covariância fenotípica relativa ao componente poligênico (Σ_g) e a outra relativa ao componente devido ao erro (Σ_e).

Assim, considerando p fenótipos avaliados em indivíduos e seus familiares, escreve-se o modelo linear misto poligênico (multivariado) como segue (Oualkacha et al., 2012; de Andrade et al., 2015):

$$Y_{f(n_f px1)} \sim \mathcal{N}_{n_f xp}(1_{n_f} \otimes \mu_{px1}, \Omega_f = 2\phi_f \otimes \Sigma_g + I_{n_f} \otimes \Sigma_e) \tag{15}$$

Em que:

$Y_{f(n_f px1)}$ é o vetor de dimensão (*n_fp x 1*) que corresponde às medidas de p fenótipos para os n_f indivíduos da família f

1_{n_f} é o vetor de 1's de dimensão ($n_f x1$)

μ_p é o vetor de dimensão (px1) das médias dos p fenótipos

I_{n_f} é a matriz identidade de dimensão ($n_f x n_f$)

$\otimes$ é o produto de Kronecker

Σ_g e Σ_e são matrizes de covariâncias fenotípicas de dimensão (pxp) específicas aos componentes poligênico e de erro, respectivamente.

$2\phi_f$ é a matriz de parentesco da família f, de dimensão ($n_f x n_f$), assumida conhecida

Ω_f é a matriz de covariâncias fenotípicas de dimensão ($n_f p x n_f p$) que combina as diferentes possibilidades de estabelecer covariâncias fenotípicas (entre diferentes fenótipos avaliados no mesmo indivíduo, diferentes fenótipos avaliados em diferentes indivíduos, bem como o mesmo fenótipo avaliado em indivíduos diferentes).

Assim, de posse do heredograma das famílias é possível obter a matriz ϕ. Existem diferentes recursos computacionais para atender a esta finalidade, por exemplo o pacote kinship do R. O próximo passo da análise é realizar os ajustes de acordo com os Modelos Lineares Mistos Poligênicos Univariado, Multivariado e a execução do Algoritmo PC (por exemplo, usando os recursos do R). Uma ilustração é mostrada na Figura 12 – Fluxograma para a construção da rede de aprendizado de estruturas.

Neste trabalho, para finalidade de construção dos Grafos, de posse da matriz de relacionamento ($2\phi_f$) e dos dados fenotípicos de cada indivíduo, faz-se, primeiramente, o ajuste do MLMP_Multivarido, seguindo os estimadores propostos em Oualkacha et al (2012) e de Andrade et al (2015). Obtém-se assim os dois componentes das matrizes de covariância, ou seja, obtém-se duas matrizes, a matriz de covariância relativa ao componente poligênico (estimativa de Σ_g) e a matriz de covariância relativa ao componente do erro (estimativa de Σ_g).

O objetivo é a obtenção das correlações parciais entre as variáveis, ou seja, as correlações entre duas variáveis condicionadas às demais. Sob a condição de M_Pares, torna-se possível obter as correlações parciais através do cálculo das inversas das matrizes de covariância, ou seja, as matrizes de precisão. Deste modo,

pode-se aplicar um teste de correlação parcial nula (igual a zero) a cada uma das entradas da matriz de precisão. O resultado obtido permite a construção dos grafos não direcionados.

Considere a matriz Σ representando a matriz de covariância total, isto é, $\Sigma = \Sigma_g + \Sigma_e$, que engloba tanto o componente genético quanto o residual ou ambiental.

A matriz Σ_g é a matriz de covariâncias do componente poligênico. Para esse estudo assume-se homocedasticidade, ou seja, a matriz Σ_g é igual para todas as famílias, o que significa que a relação entre os fenótipos é igual em todas as famílias.

A matriz Σ_e é a matriz de covariâncias do componente residual ou ambiental. Para este estudo também se assume homocedasticidade, ou seja, a matriz Σ_e não difere entre as famílias.

Um exemplo das matrizes Σ_g e Σ_e para p=3 segue abaixo:

$$\Sigma_g = \begin{pmatrix} \sigma_{g1}^2 & \sigma_{g12} & \sigma_{g13} \\ \sigma_{g21} & \sigma_{g2}^2 & \sigma_{g23} \\ \sigma_{g31} & \sigma_{g32} & \sigma_{g3}^2 \end{pmatrix}$$

$$\Sigma_e = \begin{pmatrix} \sigma_{e1}^2 & \sigma_{e12} & \sigma_{e13} \\ \sigma_{e21} & \sigma_{e2}^2 & \sigma_{e23} \\ \sigma_{e31} & \sigma_{e32} & \sigma_{e3}^2 \end{pmatrix}$$

Em que $\sigma_{g12} = \sigma_{g21}$; $\sigma_{g13} = \sigma_{g31}$; $\sigma_{g23} = \sigma_{g32}$ e $\sigma_{e12} = \sigma_{e21}$; $\sigma_{e13} = \sigma_{e31}$; $\sigma_{e23} = \sigma_{e32}$, ou seja, as matrizes Σ_g e Σ_e são matrizes simétricas. Os componentes σ_{gij} e σ_{eij} são covariâncias (poligênica e residual, respectivamente) entre os fenótipos i e j. Já os componentes σ^2_{gj} e σ^2_{ej} são as variâncias (poligênica e residual, respectivamente) do fenótipo j.

Situações em que as covariâncias (σ_{gij} e σ_{eij}) têm sinais contrários são mais informativas para a análise multivariada do conjunto de fenótipos (Oualkacha, 2012 e de Andrade, 2015) relativamente às análises univariadas de cada fenótipo.

Adicionalmente, é importante notar que o modelo de grafos não dirigidos é inicialmente construído a partir de uma estimativa da inversa da matriz de

covariâncias. Vale ressaltar que pelo fato de este estudo considerar um número pequeno de fenótipos (variáveis *Yi*), o cálculo da inversa da matriz de covariâncias é relativamente simples, não necessitando de soluções penalizadas. No entanto, quando p (número de fenótipos) é muito maior que n (número de indivíduos), denotado por p >> n, o cálculo dessa matriz se torna mais complexo, lançando-se mão de recursos computacionais, como o LASSO (Meinshausen e Bühlmann, 2006).

Tomando-se como base os Modelos Uni e Multivariados, os dados podem ser tratados como mostra a Figura 11.

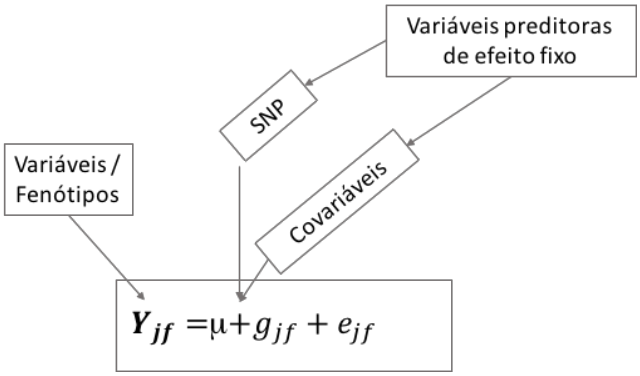


Figura 11: Esquema para inclusão de Covariáveis e SNPs

Os vértices (nós) que aparecem nos grafos correspondem ao componente à esquerda da equação (Variáveis / Fenótipos Y). Então, a inclusão de covariáveis (Etapa 2) ou inclusão de um SNP para cada fenótipo (Etapa 3) representa a inclusão de componentes de efeito fixo que podem eliminar efeitos de confundimento (Figura 11). Essa inclusão ocorre no componente da média ($E(Y)=\mu$), do lado direito da equação da Figura 11.

Na abordagem adotada neste trabalho, SNPs selecionados são assumidos terem relação (causal) com o fenótipo selecionado e não faz sentido introduzí-lo como uma nova variável, mas sim como variável instrumental.

Porém, com o objetivo de determinar não somente uma rede de fenótipos, mas também de SNPs, esses passariam para o lado esquerdo da equação e a rede de associações seria construída também com os SNPs sendo os vértices dos grafos. Neste caso, os modelos probabilísticos devem envolver conjuntos de variáveis multivariadas quantitativas (fenótipos) e discretas (SNPs, isto é, o número de alelos de risco, aa=0, Aa=1 ou AA=2).

3.2.2 Condução do Trabalho

O objetivo fundamental deste trabalho é a construção de redes de associação entre os fenótipos que compõem a SM. Três abordagens serão estudadas: Matriz de Precisão, Análise de Vizinhanças e construção dos DAGs através do PC Algoritimo (Spirtes, 2000). Por fim, será abordada a maneira como os resultados serão validados.

3.2.2.1 Matriz de Precisão (M_Pares) - Σ^{-1}

Conforme citado nas seções anteriores será utilizada M_Pares para construir a rede de associação dos fenótipos através de UDGs.

Com base em uma estimativa da inversa das matrizes de covariâncias serão extraídos os grafos (UDG). O procedimento consiste no teste de correlação parcial nula. Alternativas são o teste z de Fisher ou o teste de Spearman. Correções para múltiplos testes são recomendadas na tomada de decisão sobre a presença de uma aresta entre dois fenótipos. Alternativas úteis são Bonferroni e FDR.

Os testes de correlação parcial nula estão bem definidos na literatura para o caso de observações independentes. Na situação de delineamentos com famílias em que estes testes são independentemente aplicados a cada componente, genético e ambiental, uma limitação aparece sobre o número de graus de liberdade a ser usado nos testes. A recomendação da literatura, em situações similares, é usar o tamanho amostral efetivo (Bae et al., 2016; Raffa and Thompson, 2016). O uso desta abordagem neste trabalho teve grande impacto em reduzir o tamanho amostral usado nos testes do grafo para o componente genético.

Dessa forma, sob o modelo linear misto poligênico multivariado, tem-se três matrizes de covariâncias estimadas: Σ_Y (total), Σ_g (componente poligênico de covariâncias) e Σ_e (componente residual / ambiental de covariâncias). Desses resultados extrai-se dois sub-grafos das variáveis da SM, um relativo ao componente de variância poligênica e outro ao componente de variância residual / ambiental e um terceiro grafo total.

O entendimento e interpretação das configurações desses grafos é parte importante deste trabalho, ressaltando-se que este tipo de abordagem é pioneira na

literatura. Um enfoque interessante é introduzir o conceito de pleiotropia para justificar a rede poligênica.

Podem ocorrer divergências entre os dois grafos representando as influências de cada componente (poligênico ou ambiental) nas associações. Além disso, a depender de como se dá estes efeitos, o grafo (total) obtido de Σ^{-1} pode não fazer sentido. Dessa maneira, torna-se necessária uma interpretação cuidadosa dos resultados para que se tenha o melhor entendimento dessas relações.

A análise dos grafos gerados a partir das matrizes de precisão é aplicável para as três Etapas do estudo e para todos os casos foram gerados três grafos, apresentados na sessão de resultados e discutidos na etapa de discussão. Os grafos gerados levando em conta esta abordagem são grafos não–direcionados (UDGs).

3.2.2.2 Análise de Vizinhanças (M_Local)

Como já dito anteriormente, a M_Local implica a M_Pares e, portanto, possui uma força maior.

Nesta seção os UDGs serão construídos através da M_Local.

Aqui o intuito é estabelecer medidas de correlação parcial, ou seja, estabelecer a correlação entre duas variáveis dadas as demais através do ajuste de modelos de regressão para cada variável, utilizando-se o critério “AND” ou “OR” para combinar as regressões de dois fenótipos.

Neste procedimento tem-se:

$$Y_1 = \beta_2 Y_2 + \sum_{s \in S} \beta_s Y_s + g_1 + e_1 \tag{17}$$

$$Y_2 = \beta_1 Y_1 + \sum_{s \in S} \beta_s Y_s + g_2 + e_2 \tag{18}$$

Em que S é o conjunto condicionante comum aos ajustes de Y1 em função de Y2 e de Y2 em função de Y1. Se há evidência de que ambos $\beta 1$ e $\beta 2$ são não nulos, Y1 e Y2 são vizinhos e, portanto, existe conexão entre eles. O ajuste destes modelos univariados (MLMP Univariados) é feito considerando todas as possibilidades de

correlação parcial entre pares de variáveis. Apesar do esforço computacional ser grande há uma simplificação envolvida já que se trata de obter uma estimativa da inversa das matrizes de covariância sem ajustar o modelo MLMP Multivariado.

Nesta abordagem, a limitação envolvida diz respeito a obter estimativas não confundidas das matrizes de correlação parcial genética e ambiental. Neste trabalho foi adotada proposta similar à de Loy and Hofmann (2015). Como no caso multivariado anterior, o uso desta abordagem neste trabalho teve grande impacto em reduzir o tamanho amostral usado nos testes do grafo para o componente genético.

Vale ressaltar que a Análise de Vizinhanças também será executada para as 3 Etapas propostas neste trabalho e serão gerados e analisados os UDGs obtidos.

3.2.2.3 Grafos Acíclicos Direcionados

Para a obtenção dos grafos acíclicos direcionados será aplicado o pacote PC Algorithm (pcalg) do software R (Spirtes, 2000) como considerado na seção 1.3.2.4.

Para o Algoritmo PC assume-se Fidelidade e utiliza-se do conceito de independência condicional e d-separação para determinar se existe ligação entre os vértices. Em um segundo passo do algoritmo é que são testadas as direções dessas ligações utilizando o conceito de colisão e, portanto, o estabelecimento das relações causais.

Para este estudo o algoritmo será executado para as 3 Etapas.

3.2.2.4 Validação dos Resultados

Com o intuito de garantir a validade dos resultados obtidos para a população de estudo, os resultados serão submetidos à duas análises de validação.

Uma delas é o confronto com os dados da literatura sobre a SM. Ou seja, será feita uma comparação entre os modelos existentes na literatura sobre as relações da SM e os grafos obtidos. A outra alternativa a ser considerada é através da extração de sub-amostras da amostra total, seguida da execução dos testes para

construção dos Grafos com o objetivo de verificar se os resultados são sensíveis a perturbações nos dados amostrais.

Vale ressaltar que os modelos propostos para a SM atualmente disponíveis na literatura são escassos, além de terem considerado indivíduos independentes, ou seja, não levando em consideração a estrutura de famílias que aqui é proposta.

A metodologia utilizada para a Validação dos Grafos é de natureza exploratória e busca avaliara sensibilidade, ou seja, verificar a robustez do algoritmo.

Ressalta-se, que neste trabalho a validação dos resultados será conduzida de forma exploratória e descritiva, pois, apesar de haver na literatura medidas resumo para grafos as quais poderiam ser adotadas na comparação de resultados, como medidas de centralidade de grau, centralidade de proximidade e centralidade de intermediação, essas podem não ser diretamente aplicadas ao nosso estudo, que envolve observações correlacionadas e ajustes do modelo linear misto.

Sobre análise de poder do estudo, este é um tema de pesquisa ainda em aberto na literatura para modelos de Grafos. O maior desafio é devido ao grande número de testes envolvidos na construção do grafo e, dependendo do objetivo, há dificuldade em estabelecer uma medida que resuma o padrão de “verdadeiros positivos” no espaço ROC (Petterson, LE e Coleman, MA, 2008).

3.2.3 Fluxograma para a construção de redes de aprendizado de estruturas

A Figura 12 a seguir traz um fluxograma do emprego da metodologia, partindo dos dados até a obtenção dos Grafos (UDG e DAG).

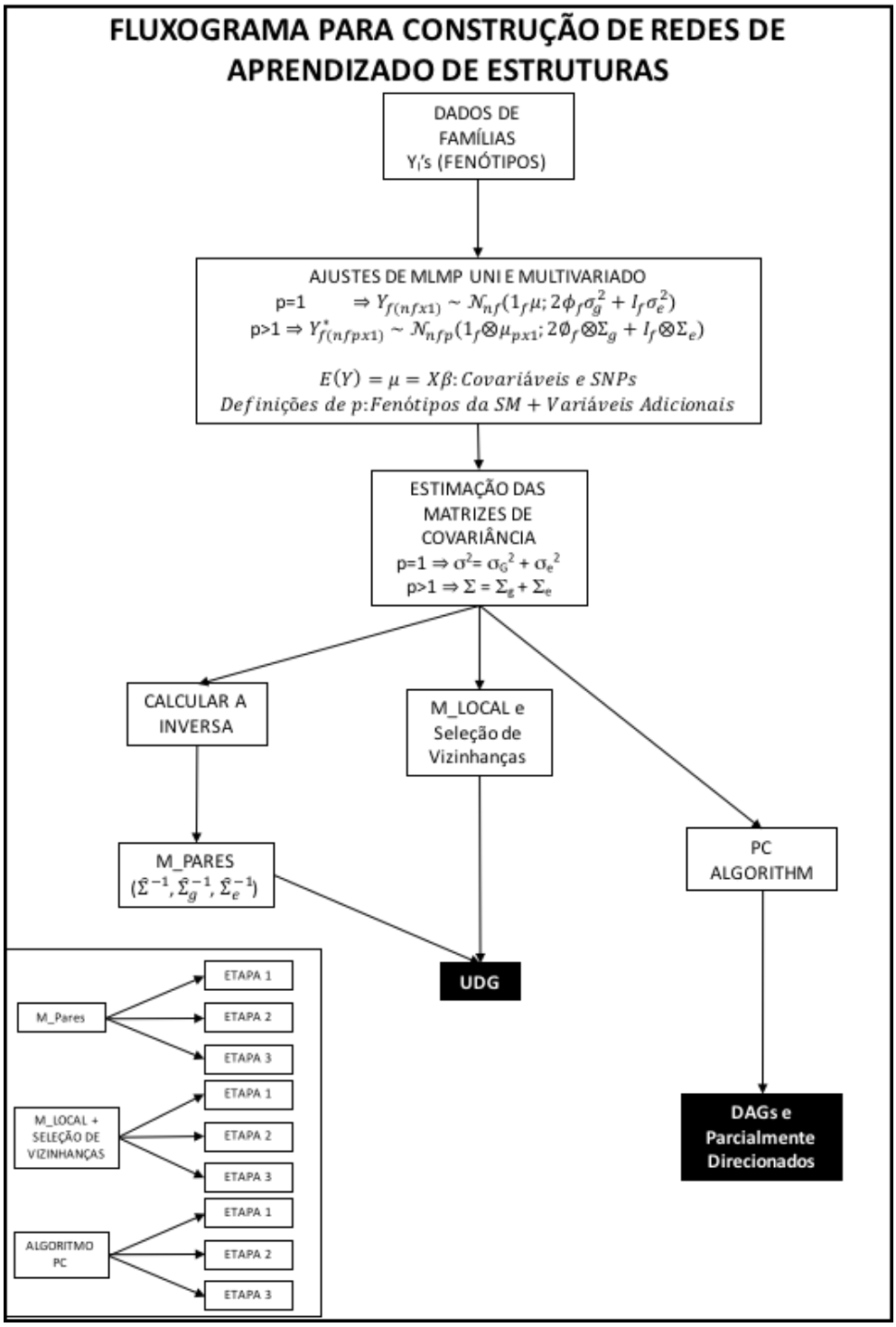


Figura 12: Fluxograma para construção de redes de aprendizado de estruturas.

4 RESULTADOS

O banco de dados de Baependi usado neste trabalho consiste de n = 1666 (indivíduos) de F = 79 (Famílias). O tamanho médio das famílias é de 21 membros (s = 26,52) e a mediana é de 14 membros. O Gráfico 1 e a Tabela 2 trazem essas informações:

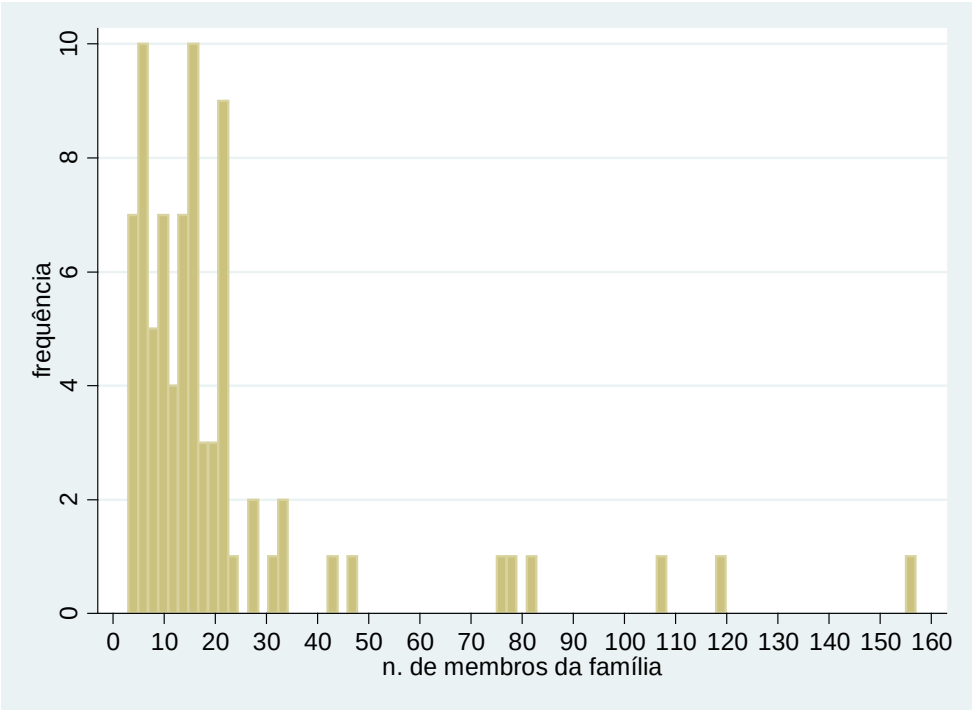


Figura 13: Distribuição de frequências do número de membros de cada família na amostra de estudo.

Dados das Famílias	
Min.	3,00
1st Qu.	7,50
Median	14,00
Mean	21,09
3rd Qu.	22,00
Max.	157,00
S.D	26,52294

Tabela 2: Dados descritivos do número de membros das Famílias

Observa-se, tanto no gráfico quanto na tabela que, apesar de algumas famílias possuírem um elevado número de membros, o que, de certa forma, puxa a

média para cima, a maioria das famílias conta com um número pequeno de membros.

Para exemplificar, tomamos o heredograma de uma das famílias do estudo, bem como a matriz de relacionamento correspondente, as quais estão apresentadas nas Figuras 14 a 16.

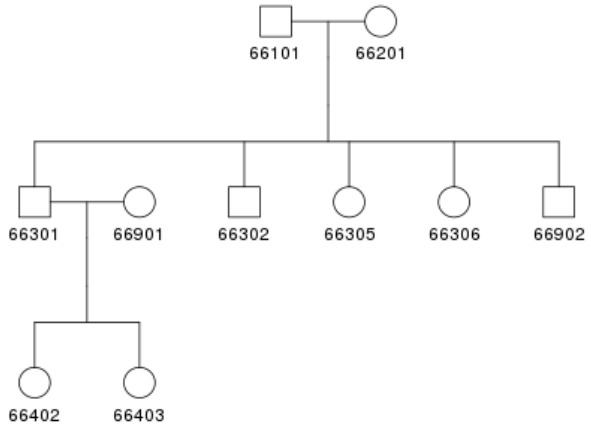


Figura 14: Heredograma de uma família pertencente ao estudo.

$\phi =$

	66101	66201	66301	66302	66305	66306	66402	66403	66901	66902
66101	0,500	0,000	0,250	0,250	0,250	0,250	0,125	0,125	0,000	0,250
66201	0,000	0,500	0,250	0,250	0,250	0,250	0,125	0,125	0,000	0,250
66301	0,250	0,250	0,500	0,250	0,250	0,250	0,250	0,250	0,000	0,250
66302	0,250	0,250	0,250	0,500	0,250	0,250	0,125	0,125	0,000	0,250
66305	0,250	0,250	0,250	0,250	0,500	0,250	0,125	0,125	0,000	0,250
66306	0,250	0,250	0,250	0,250	0,250	0,500	0,125	0,125	0,000	0,250
66402	0,125	0,125	0,250	0,125	0,125	0,125	0,500	0,250	0,250	0,125
66403	0,125	0,125	0,250	0,125	0,125	0,125	0,250	0,500	0,250	0,125
66901	0,000	0,000	0,000	0,000	0,000	0,000	0,250	0,250	0,500	0,000
66902	0,250	0,250	0,250	0,250	0,250	0,250	0,125	0,125	0,000	0,500

Figura 15: Matriz ϕ para a família representada pela Figura 14.

$2\phi =$

	66101	66201	66301	66302	66305	66306	66402	66403	66901	66902
66101	1,00	0,00	0,50	0,50	0,50	0,50	0,25	0,25	0,00	0,50
66201	0,00	1,00	0,50	0,50	0,50	0,50	0,25	0,25	0,00	0,50
66301	0,50	0,50	1,00	0,50	0,50	0,50	0,50	0,50	0,00	0,50
66302	0,50	0,50	0,50	1,00	0,50	0,50	0,25	0,25	0,00	0,50
66305	0,50	0,50	0,50	0,50	1,00	0,50	0,25	0,25	0,00	0,50
66306	0,50	0,50	0,50	0,50	0,50	1,00	0,25	0,25	0,00	0,50
66402	0,25	0,25	0,50	0,25	0,25	0,25	1,00	0,50	0,50	0,25
66403	0,25	0,25	0,50	0,25	0,25	0,25	0,50	1,00	0,50	0,25
66901	0,00	0,00	0,00	0,00	0,00	0,00	0,50	0,50	1,00	0,00
66902	0,50	0,50	0,50	0,50	0,50	0,50	0,25	0,25	0,00	1,00

Figura 16: Matriz 2ϕ para a família representada pela Figura 14.

Considerando os fenótipos da SM, a Tabela 3 e as Figuras 17 a 22 apresentam estatísticas descritivas e os histogramas destas variáveis e a Tabela 4 apresenta a herdabilidade de cada fenótipo.

Circunferência Abdominal (cm)		**Glicose (mg/dL)**		**CHDL (mg/dL)**		**Triglicérides (mg/dL)**		**Pressão Sistólica (mmHg)**		**Pressão Diastólica (mmHg)**	
Min.	54.00	Min.	44.60	Min.	13.00	Min.	34.20	Min.	79.70	Min.	50.00
1º Qua	78.00	1º Qua	76.60	1º Qua	44.88	1º Qua	84.25	1º Qua	112.30	1º Qua	70.70
Mediana	87.00	Mediana	89.00	Mediana	53.90	Mediana	113.30	Mediana	124.30	Mediana	77.70
Média	87.39	Média	93.82	Média	55.88	Média	133.37	Média	126.70	Média	78.75
3º Qua	95.00	3º Qua	103.50	3º Qua	66.00	3º Qua	160.35	3º Qua	137.00	3º Qua	85.70
Max.	166.00	Max.	395.00	Max.	125.40	Max.	719.00	Max.	216.30	Max.	129.70
S.D.	12.44	S.D.	29.70	S.D.	16.61	S.D.	75.36	S.D.	19.38	S.D.	11.36

Tabela 3: Estatística descritiva dos fenótipos da SM.

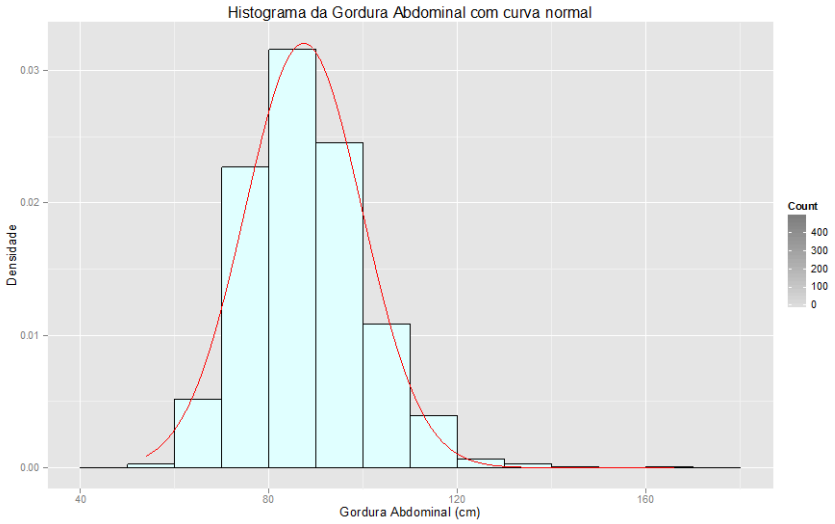


Figura 17: Histograma da Circunferência Abdominal

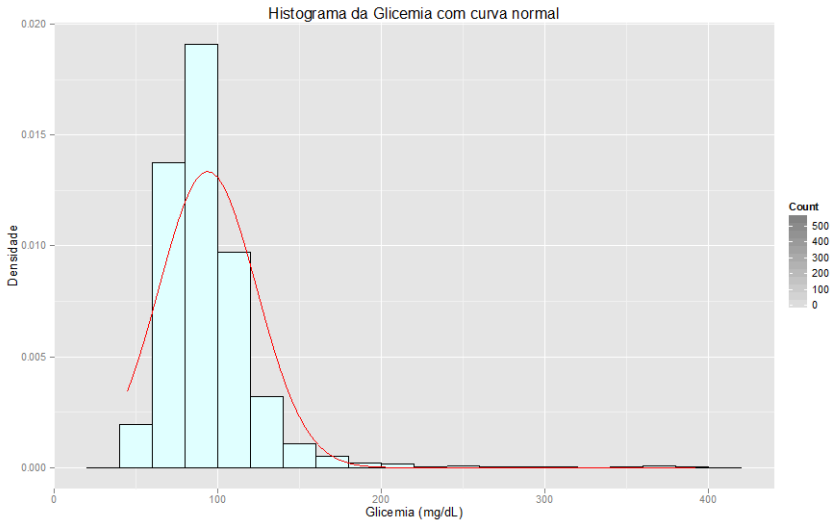


Figura 18: Histograma da Glicemia

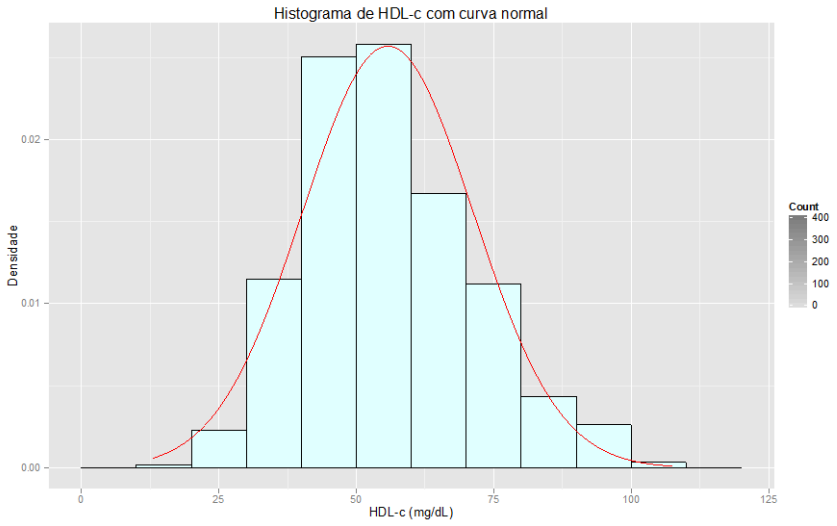


Figura 19: Histograma do HDL-c

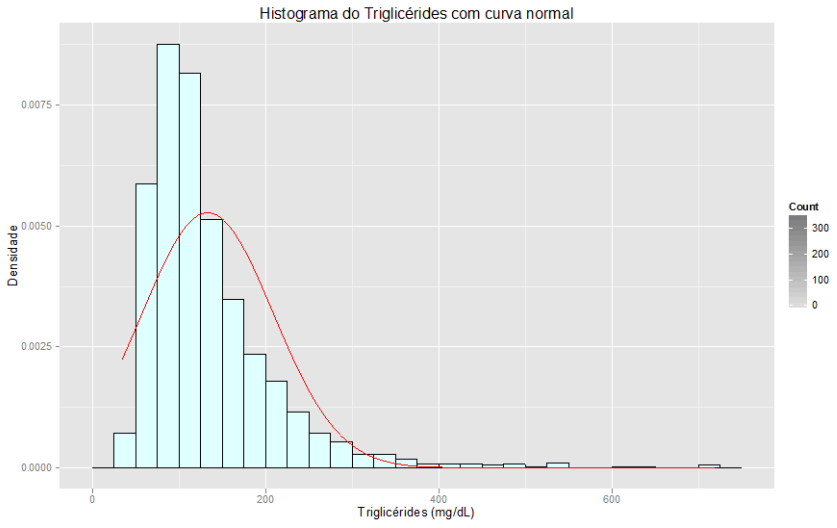


Figura 20: Histograma dos Triglicérides

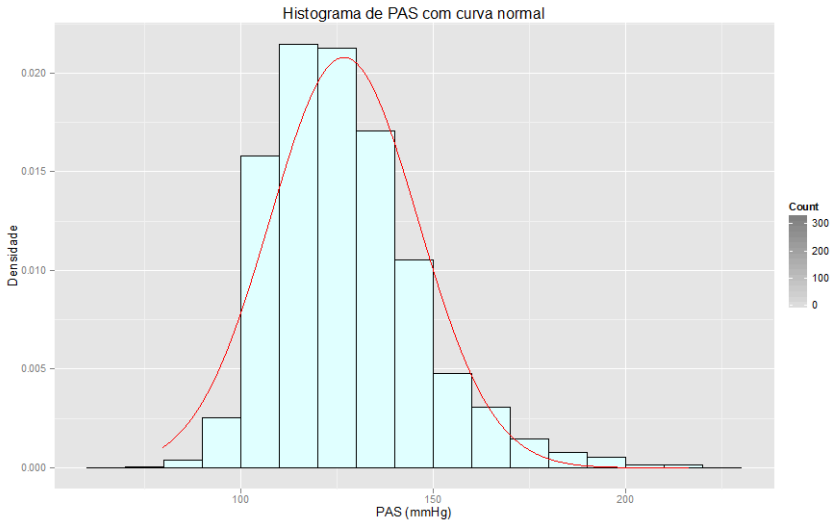


Figura 21: Histograma de PAS

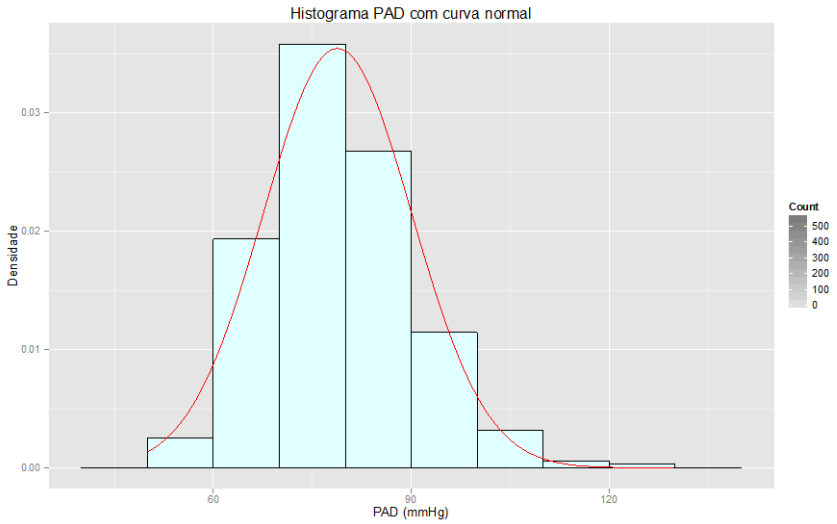


Figura 22: Histograma de PAD

Apesar de os histogramas aproximarem-se da distribuição normal, com o intuito de se fazer um melhor ajuste, todos os dados foram transformados para ln, o logaritmo natural.

Resposta	Obesidade Truncal	Glicose	cHDL	Trigligérides	PAS Média	PAD Média
h^2	0,164	0,118	0,302	0,352	0,177	0,132

Tabela 4: Herdabilidade dos fenótipos da SM.

A Tabela 4 mostra os resultados da herdabilidade univariada, ou seja, de cada um dos fenótipos individualmente, obtida para a amostra em estudo levando em conta ajustes para idade, idade^2, sexo e PC1 e PC2 (Componentes Principais da análise de ancestralidade desta amostra).

Os resultados das matrizes de correlação parcial e dos grafos serão apresentados de acordo com a abordagem adotada (Modelo Multivariado, Modelo Univariado e Algoritmo PC), mostrando os resultados obtidos em cada etapa (1 – sem covariáveis; 2 – com covariáveis sexo, idade, PC1 e PC2; 3 – com covariáveis sexo, idade, PC1, PC2 e SNPs).

Além disso, para a Etapa 3, em que há a inclusão de SNPs, o mapeamento de SNPs associados a cada fenótipo da SM foi realizado por meio de ajustes do modelo linear misto poligênico univariado e uniloco. Como, neste trabalho, os dados dos SNPs foram extraídos tanto do próprio banco de Baependi quanto da literatura, as duas bases de dados foram utilizadas. As tabelas abaixo mostram a lista dos SNPs mais significantes de cada base que foram utilizados nas

análises. O Apêndice 1 traz a lista completa dos SNPs relativos a cada fenótipo bem como os gráficos Manhattan para a base de dados de Baependi e o Apêndice 2 traz os quadros com a lista total de SNPs relativos a cada fenótipo de acordo com a busca feita na literatura.

CHR	***SNP***	***Pvalue***	***Fenótipo Relacionado***
11	rs12789896	1,1e-07	Circ. Abdominal
6	rs16889851	1,3e-07	Glicose
12	rs12302648	1,6e-07	Triglicérides
5	rs17526653	5,1e-07	HDLC
4	rs6835708	2,2e-07	PAS
3	rs17035693	9,9e-07	PAD
4	rs6850807	1,8e-07	PAD

Tabela 5: SNPs mais significantes ($p \leq 10^{-7}$) do banco de Baependi utilizados nas análises

CHR	***SNP***	***p-value***	***Fenótipo relacionado***
5	rs10061119	2,2 e-05	Circ. Abdominal
5	rs6883134	5,6 e-04	HDLC
5	rs16887451	7,9 e-04	Triglicérides
5	rs5745297	2,6 e-04	PAS
5	rs6450583	5,4 e-04	PAD
9	rs10757282	2,0 e-12	Glicose

Tabela 6: SNPs mais significantes de dados extraídos da literatura utilizados nas análises

4.1 RESULTADOS MLMP_MULTIVARIADO

4.1.1 Correlações parciais para o componente Poligênico

4.1.1.1 Etapas 1 e 2

	Circ. abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.6542 *1 *2	-0.1210	0.1848	0.3134	-0.3278
GLICOSE	0.4680	1.0000	0.1370	0.0695	-0.3852	0.4029
CHDL	-0.2064	-0.0769	1.0000	0.3280	-0.1264	0.1788
TRIG	0.1194	0.1911	0.3435	1.0000	0.3173	-0.3592
mediaPAS	0.1623	-0.6133	-0.2316	-0.2891	1.0000	0.9135 *1 *2
mediaPAD	-0.1395	0.5965	0.2256	0.2527	0.9402 *1 *2	1.0000

Tabela 7: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior) considerando α = 5% para os caso sem correção e com correção por FDR.

Para o cálculo dos p-valores do componente poligênico foi utilizado o número de graus de liberdade dos estimadores para a obtenção do tamanho amostral efetivo (Oualkacha, 2012).

4.1.1.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.4612	-0.2477	0.1696	0.1542	-0.1614
GLICOSE	0.4630	1.0000	-0.0884	0.2017	-0.6128	0.6091
CHDL	-0.2024	-0.0959	1.0000	0.3751	-0.2492	0.2235
TRIG	0.1120	0.2027	0.3306	1.0000	0.3187	-0.3322
mediaPAS	0.1670	-0.6254	-0.2507	0.2594	1.0000	0.9427*1*2
mediaPAD	-0.1401	0.6071	0.2432	-0.2956	0.9420*1*2	1.0000

Tabela 8: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casos sem correção com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

4.1.2 Correlações parciais para o componente Ambiental

4.1.2.1 Etapas 1 e 2

	Circ. Abdominal	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	-0.5648 *1 *2	-0.4788 *1 *2	-0.2147 *1 *2	0.5318 *1 *2	0.3635 *1 *2
GLICOSE	-0.3149 *1 *2	1.0000	-0.4285 *1 *2	-0.1319 *1 *2	0.5158 *1 *2	0.1605 *1 *2
CHDL	-0.1865 *1 *2	-0.1769 *1 *2	1.0000	-0.4894 *1 *2	0.2771 *1 *2	0.309 *1 *2
TRIG	0.0535	0.0272	-0.4014 *1 *2	1.0000	0.1866 *1 *2	0.4058 *1 *2
mediaPAS	0.3035 *1*2	0.3496 *1 *2	0.2180 *1 *2	0.1181 *1 *2	1.0000	0.2713 *1 *2
mediaPAD	0.1288 *1*2	-0.0467	0.1184 *1 *2	0.2277 *1 *2	0.4924 *1 *2	1.0000

Tabela 9: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior), considerando α = 5% para os casos sem correção e com correção por FDR.

Para o cálculo do tamanho amostral para o componente ambiental, utilizou-se o tamanho efetivo da amostra e o tamanho efetivo de Yang (Bae et al, 2016). O segundo valor, menor, foi utilizado, pois o tamanho efetivo da amostra acabou por tornar a análise extremamente sensível resultando em arestas mesmo quando a correlação parcial apresentou valores baixos, ou seja, tornou a análise muito sensível.

4.1.2.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	-0.2987*1*2	-0.1775*1*2	0.0309	0.3103*1*2	0.1487*1*2
GLICOSE	-0.2905*1*2	1.0000	-0.1456*1*2	0.0393	0.3456*1*2	-0.0555
CHDL	-0.1709*1*2	-0.1642*1*2	1.0000	-0.3887*1*2	0.2009*1*2	0.1223*1*2
TRIG	0.0650	0.0323	-0.3913*1*2	1.0000	0.0858	0.2433*1*2
mediaPAS	0.2832*1*2	0.3272*1*2	0.2092*1*2	0.1108	1.0000	0.4938*1*2
mediaPAD	0.1295*1*2	-0.0427	0.1137*1	0.2259*1*2	0.4997*1*2	1.0000

Tabela 10: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

4.1.3 Correlações Totais

4.1.3.1 Etapas 1 e 2

	Circ. abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0761 *1 *2	-0.1096 *1 *2	0.2288 *1 *2	0.1316 *1 *2	0.0906 *1 *2
GLICOSE	0.0108	1.0000	-0.0351	0.1381 *1 *2	0.1154 *1 *2	-0.0213
CHDL	-0.1411	-0.0576	1.0000	-0.0904 *1 *2	-0.0237	0.073 *1 *2
TRIG	0.1474	0.0985	-0.0959	1.0000	0.0186	0.0962 *1 *2
mediaPAS	0.0531	0.0479	0.0030	-0.0071	1.0000	0.7061 *1 *2
mediaPAD	0.1218	0.0195	0.0673	0.0981	0.7052 *1 *2	1.0000

Tabela 11: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior) considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.

Para o cálculo do tamanho amostral total, utilizou-se o tamanho efetivo da amostra e o tamanho efetivo de Yang. O segundo valor, menor, foi utilizado, pois o tamanho efetivo da amostra acabou por tornar a análise extremamente sensível resultando em arestas mesmo quando a correlação parcial apresentou valores baixos, ou seja, tornou a análise muito sensível.

4.1.3.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0128	-0.1409	0.1484	0.0498	0.1216
GLICOSE	0.0114	1.0000	-0.0551	0.0903	0.0550	0.0126
CHDL	-0.1409	-0.0572	1.0000	-0.0895	0.0066	0.0596
TRIG	0.1468	0.0966	-0.0973	1.0000	-0.0048	0.0989
mediaPAS	0.0561	0.0468	0.0028	-0.0081	1.0000	0.7058*1*2
mediaPAD	0.1205	0.0205	0.0668	0.0988	0.7048*1*2	1.0000

Tabela 12: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

A matriz de correlação parcial adicionando, além de sexo e idade, o IMC como covariável também foi avaliada. Porém, observou-se que o uso do IMC não trouxe resultados satisfatórios, além de eliminar arestas que estavam consistentes com o modelo. Dessa maneira, decidiu-se por não mais considerar o IMC como covariável nas demais análises executadas. Esses resultados encontram-se no Apêndice 5.

4.2 RESULTADOS MLMP_UNIVARIADO

4.2.1 Correlações parciais para o componente Poligênico

4.2.1.1 Etapas 1 e 2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.4492 *1	-0.2013	0.4115 *1	0.2507	0.0767
GLICOSE	0.1947	NA	0.1838	0.1694	0.1458	0.0224
CHDL	-0.3010	-0.2012	NA	0.0744	-0.0014	0.1354
TRIG	0.1667	0.0462	0.0702	NA	-0.3280 *1	0.3066
mediaPAS	0.3697 *1	0.0694	0.1426	-0.2338	NA	0.7967 *1 *2
mediaPAD	-0.0372	0.1080	0.0247	0.2987	.6846 *1 *2	NA

Tabela 13: Correlações parciais do componente poligênico entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2, (triângular inferior) considerando α = 5% para os casos sem correção e com correção por FDR.

4.2.1.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.2004	-0.2220	0.2089	0.4269*1	-0.1239
GLICOSE	0.2349	NA	-0.2433	0.0454	-0.0090	0.0640
CHDL	-0.2702	-0.2298	NA	0.0708	0.1588	0.0365
TRIG	0.1382	0.0332	0.0575	NA	-0.2099	0.2497
mediaPAS	0.3778*1	0.0284	0.1143	-0.2436	NA	0.7217*1*2
mediaPAD	-0.0215	0.0806	0.0212	0.2780	0.6981*1*2	NA

Tabela 14: Correlações parciais do componente poligênico entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

4.2.2 Correlações parciais para o componente Ambiental

4.2.2.1 Etapas 1 e 2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.2126 *1 *2	-0.1533 *1	0.3009 *1 *2	0.1366 *1	0.2772 *1 *2
GLICOSE	0.0230	NA	-0.1285 *1	0.0660	0.1054 *1	-0.0060
CHDL	-0.2319 *1 *2	-0.0539 *1	NA	0.0294	-0.0409	0.112 *1
TRIG	0.1428 *1	0.0028	0.0488	NA	0.0788	0.0438
mediaPAS	-0.0511 *1	0.0020	0.0234	-0.0141	NA	0.6471 *1 *2
mediaPAD	0.2447 *1 *2	0.0271	0.0020	0.0861	0.6230 *1 *2	NA

Tabela 15: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis (triangular superior) e com covariáveis sexo, idade, PC1 e PC2 (triangular inferior), considerando α = 5% para os casos sem correção e com correção por FDR.

4.2.2.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0407	-0.2449*1*2	0.1347	-0.0437	0.2079*1*2
GLICOSE	0.0439	NA	-0.0613	0.0378	-0.0001	0.0152
CHDL	-0.2102*1 *2	-0.0722*1	NA	0.0434	0.0793	0.0079
TRIG	0.1430*1	0.0053	0.0540	NA	0.0056	0.0662
mediaPAS	-0.0655*1	0.0099	0.0339	-0.0233	NA	0.6331*1*2
mediaPAD	0.2356*1 *2	0.0140	0.0049	0.0959	0.6198*1 *2	NA

Tabela 16: Correlações parciais do componente ambiental entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

4.2.3 Correlações totais

4.2.3.1 Etapas 1 e 2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0761 *1 *2	-0.1095 *1 *2	0.2289 *1 *2	0.1316 *1 *2	0.0905 *1 *2
GLICOSE	0.0108	NA	-0.0349	0.1382 *1 *2	0.1153 *1 *2	-0.0214
CHDL	-0.1411 *1 *2	-0.0575	NA	-0.0903 *1 *2	-0.0237	0.073 *1 *2
TRIG	0.1473*1 *2	0.0985*1 *2	-0.0959	NA	0.0186	0.0962 *1 *2
mediaPAS	0.0532*1 *2	0.0479	0.0029	-0.0071	NA	0.7061 *1 *2
mediaPAD	0.1218*1 *2	0.0195	0.0673*1 *2	0.0981*1 *2	0.7052*1 *2	NA

Tabela 17: Correlações parciais do componente ambiental entre os fenótipos sem covariáveis (triângular superior) e com covariáveis sexo, idade, PC1 e PC2 (triângular inferior), considerando α = 5% para o caso sem correção e α = 5% para o caso com correção por FDR.

4.2.3.2 Etapa 3

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0109	-0.1410*1*2	0.1478*1*2	0.0491	0.1209*1*2
GLICOSE	0.0103	NA	-0.055	0.0920*1*2	0.0478	0.0145
CHDL	-0.1409*1 *2	-0.0568	NA	-0.0917*1*2	0.0036	0.0624*1
TRIG	0.1474*1 *2	0.0970*1 *2	-0.0962*1 *2	NA	-0.0031	0.0979*1*2
mediaPAS	0.0535	0.0461	0.0027	-0.0068	NA	0.7032*1*2
mediaPAD	0.1218*1 *2	0.0192	0.0667*1 *2	0.0978*1 *2	0.7044*1 *2	NA

Tabela 18: Correlações parciais totais entre os fenótipos com covariáveis sexo, idade, PC1, PC2 e SNPs, considerando α = 5% para os casso sem correção e com correção por FDR. SNPs do banco de Baependi estão representados na triangular superior, e SNPs oriundos da literatura na triangular inferior.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

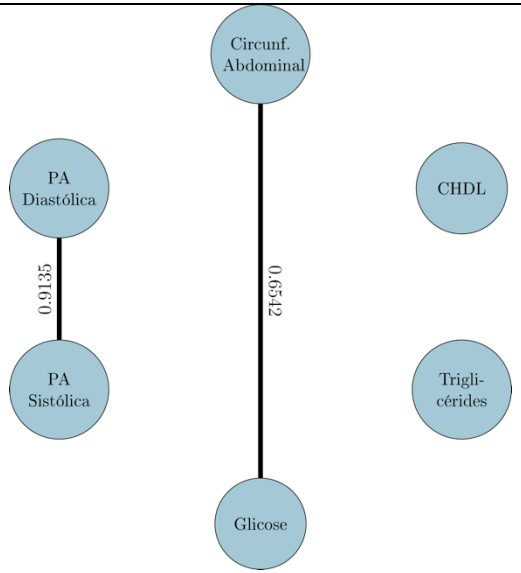
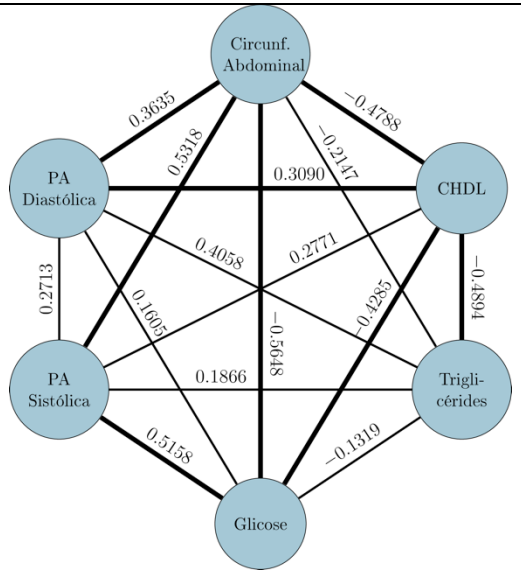
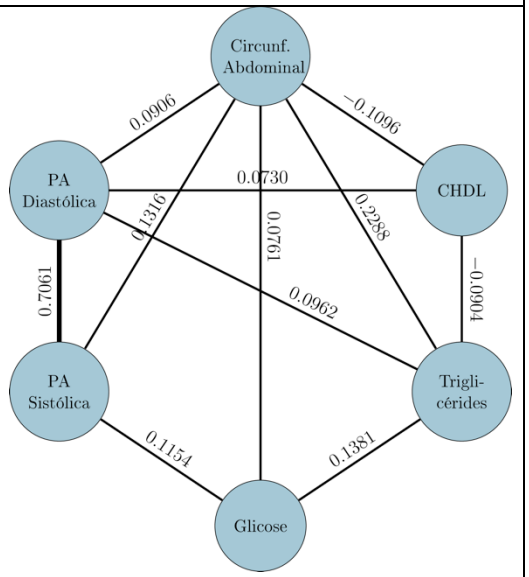
4.3 RESULTADOS PC ALGORITMO

Os resultados das correlações parciais que resultaram em arestas (direcionadas ou não) para o PC Algortimo estão apresentados nos próprios grafos e os demais resultados estão apresentados no Apêndice 4.

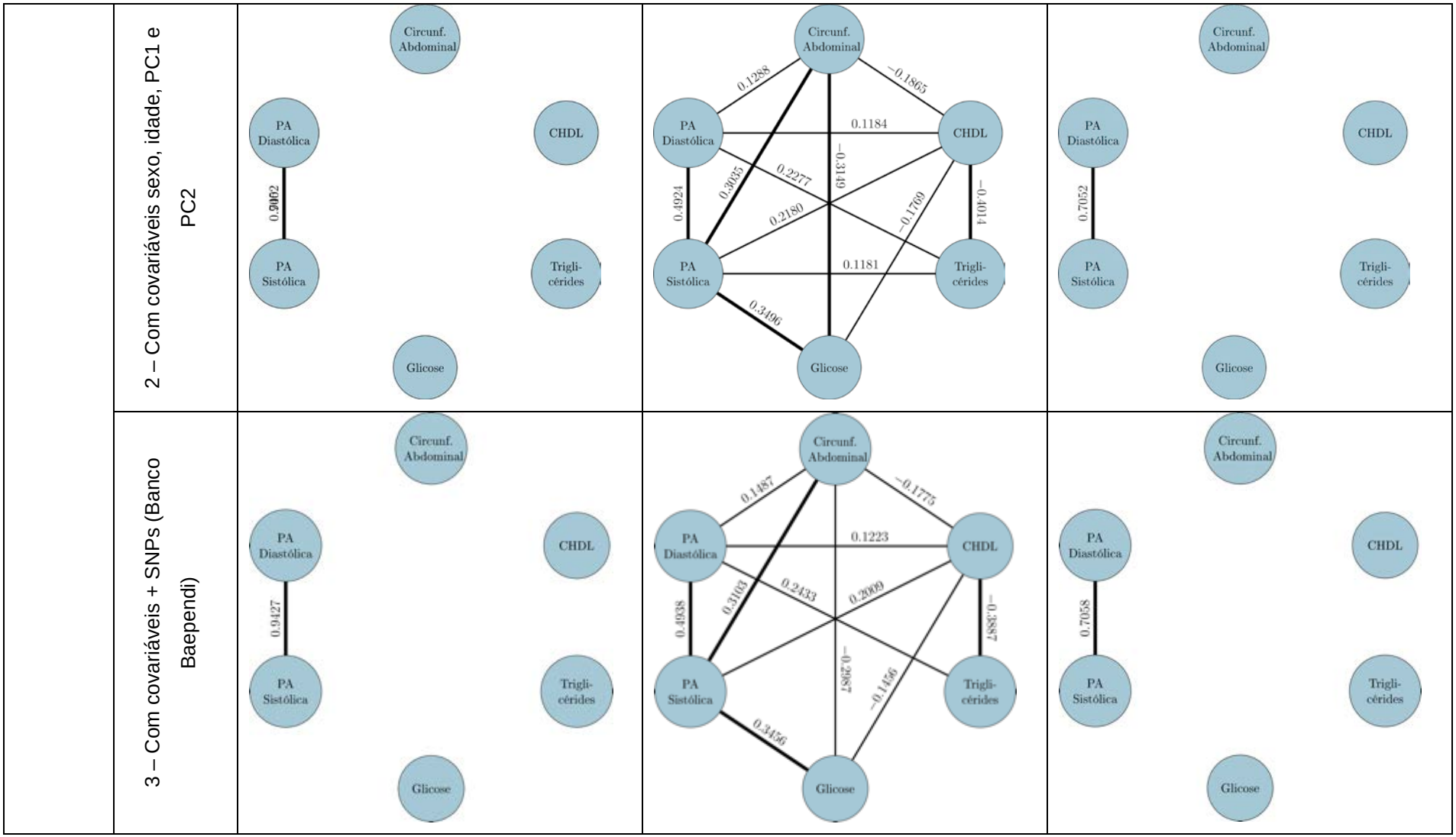
4.4 GRAFOS

Na figura 23, seguem os grafos não-direcionados (UDG) e os grafos direcionados (DAG) para as três metodologias (MLMP_Multivariado, MLMP_Univariado e Algoritmo PC) considerando as três etapas de estudo (Etapa 1: sem covariáveis, Etapa 2: com covariáveis sexo, idade, PC1 e PC2 e Etapa 3: com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi e Literatura).

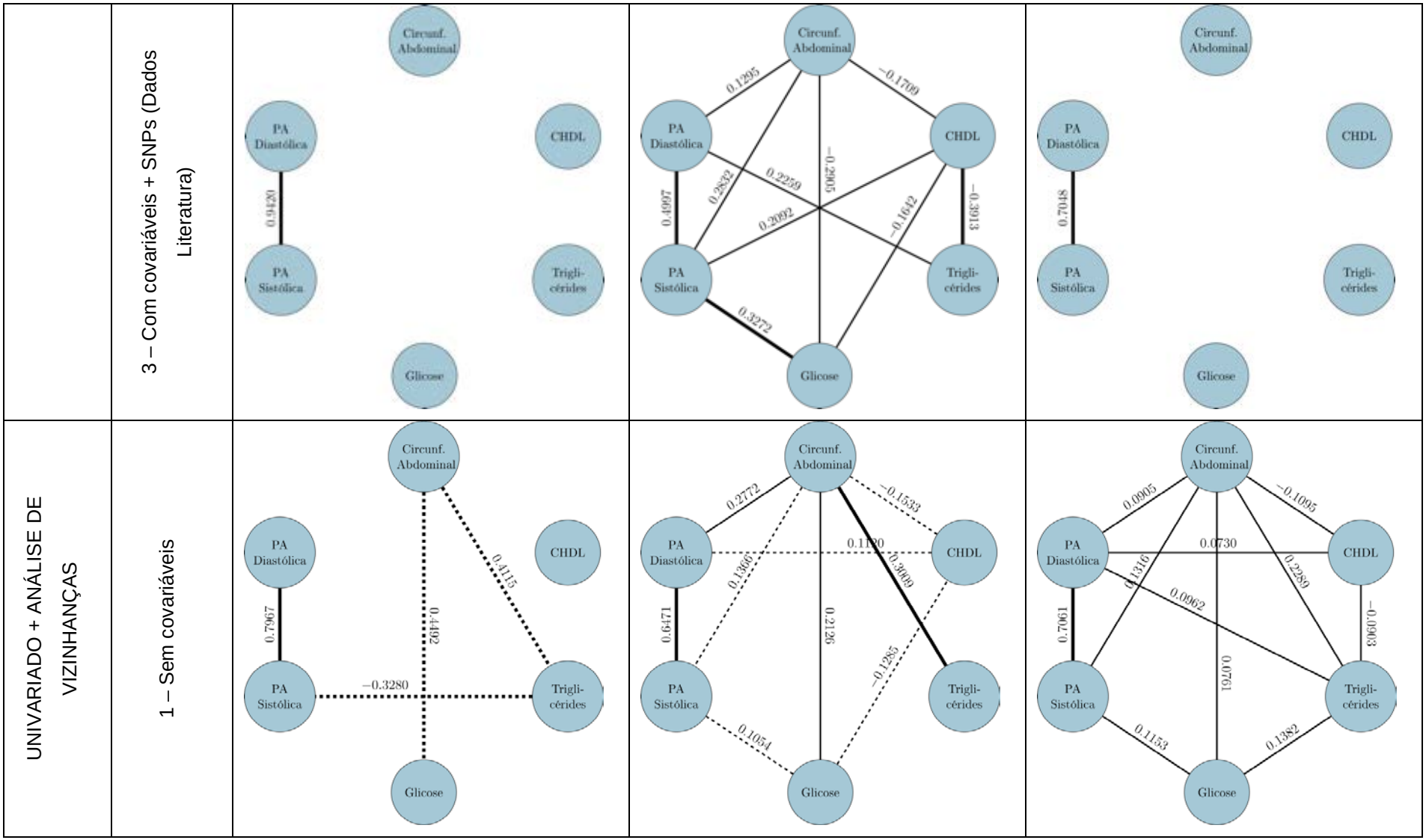
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

		COMPONENTE		
MODELO	ETAPA	POLIGÊNICO	AMBIENTAL	TOTAL
MULTIVARIADO (Oualkacha)	1 – Sem covariáveis			

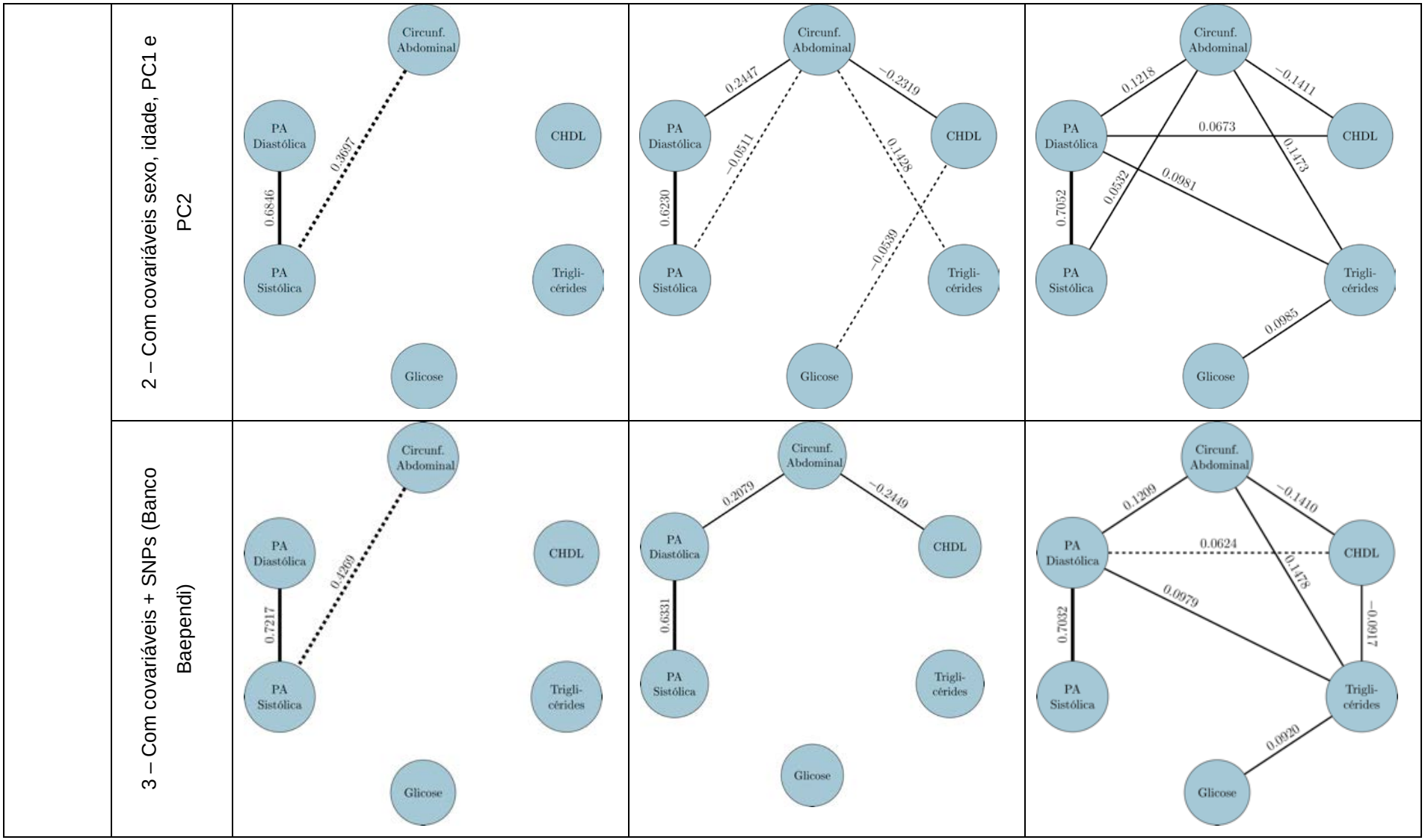
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



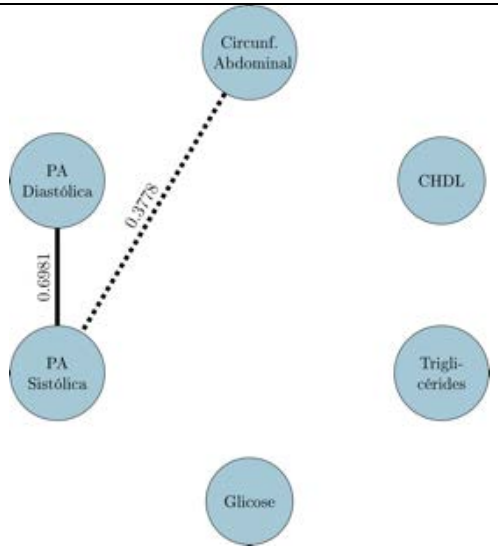
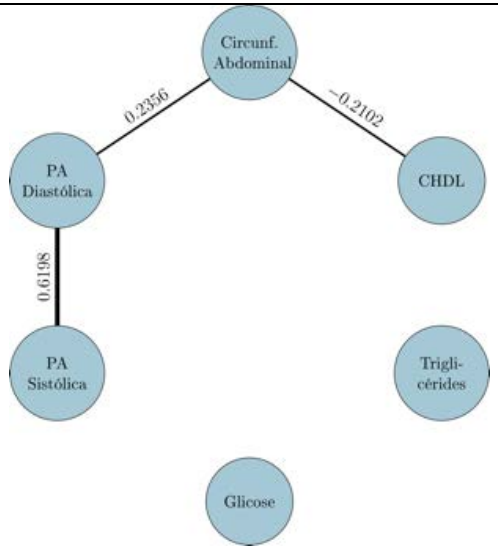
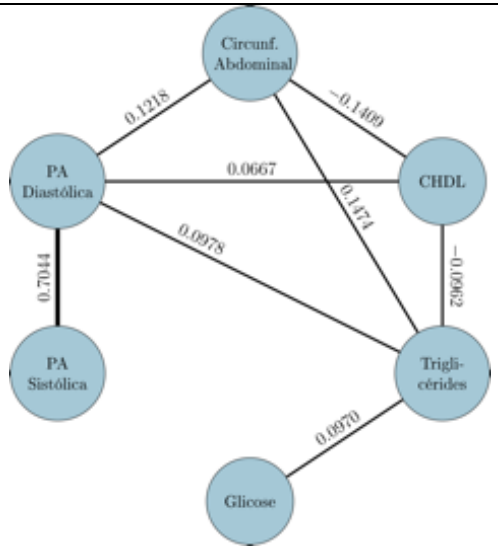
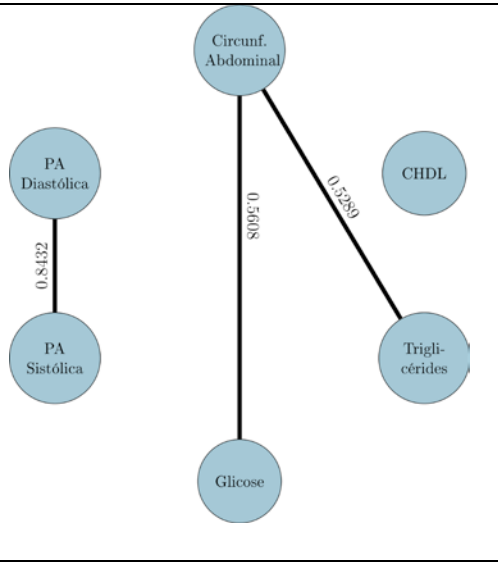
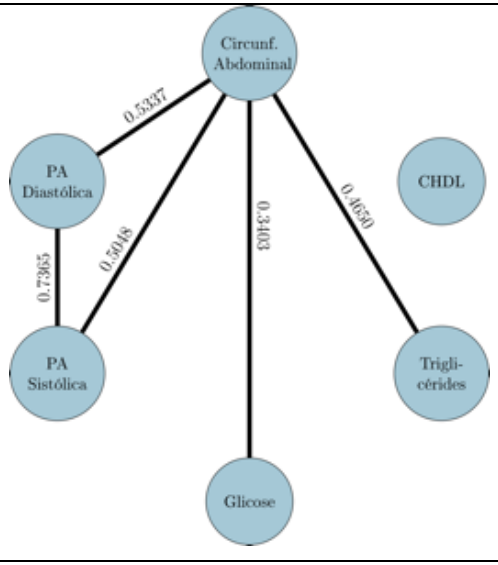
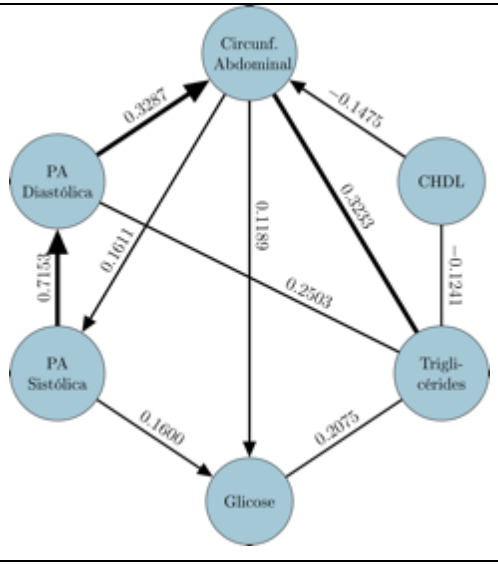
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



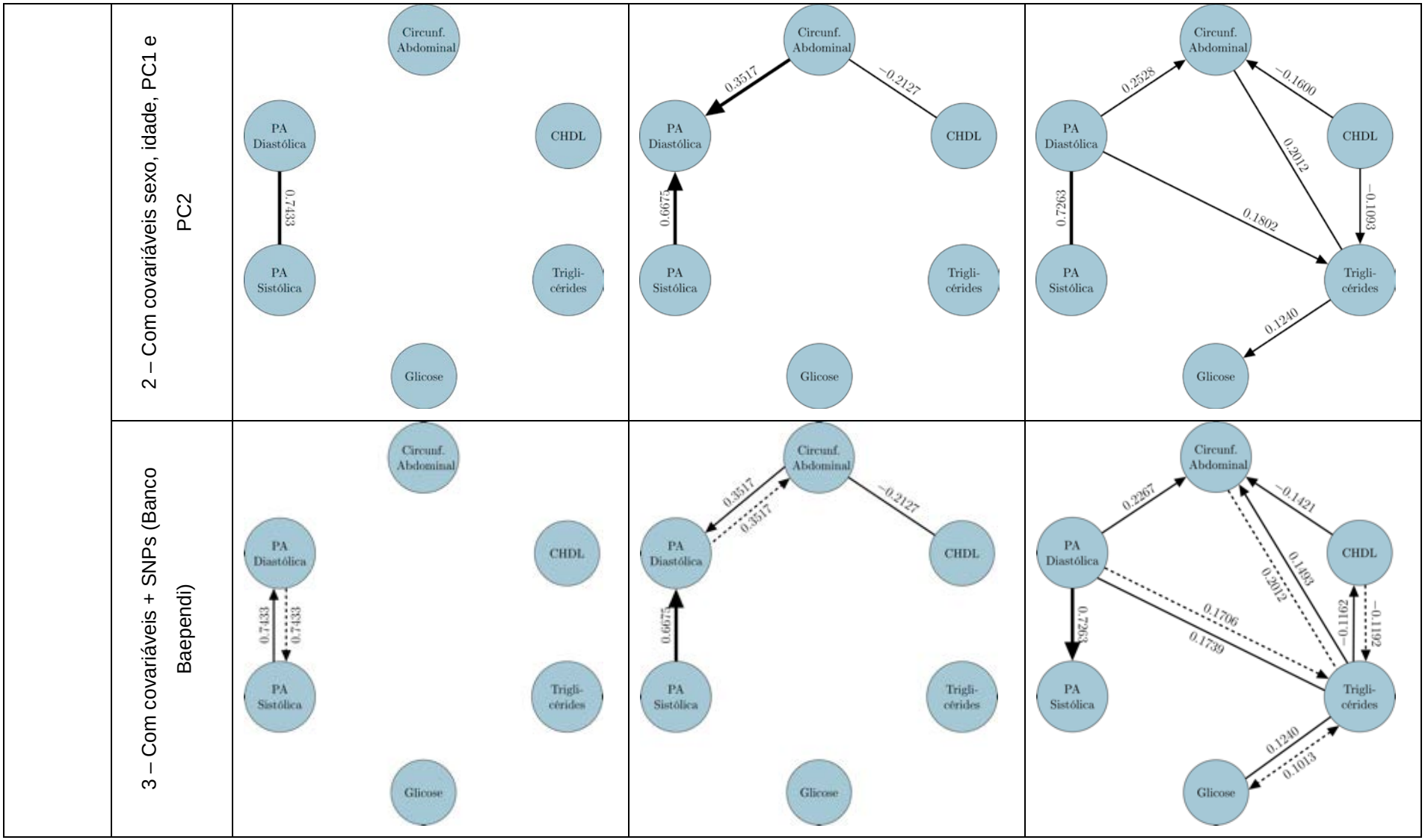
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

	3 – Com covariáveis + SNPs (Dados Literatura)			
ALGORITMO PC	1 – Sem covariáveis			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

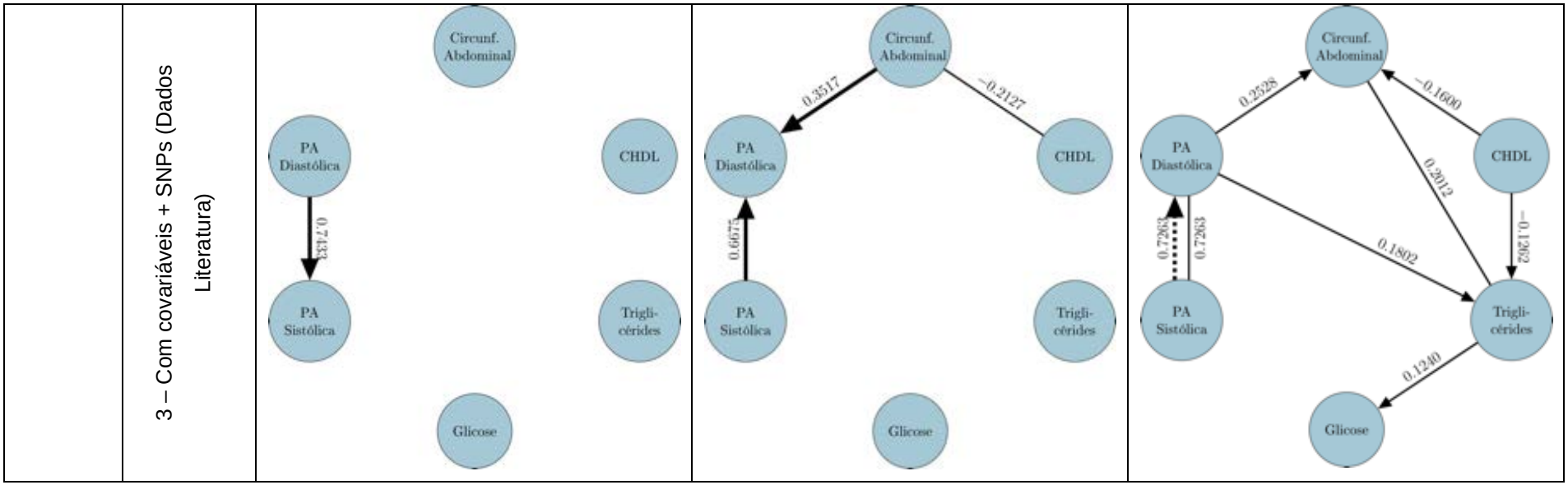


Figura 23: Grafos obtidos para as três abordagens (MLMP_Multivariado, MLMP_Univariado e Algoritmo PC) nas três etapas (sem covariáveis, com covariáveis sexo, idade, PC1 e PC2 e com covariáveis sexo, idade, PC1, PC2 e SNPs)

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

No Apêndice 6 pode-se observar as comparações entre os resultados utilizando o modo *Conservative* e o modo *Major Rule* do Algoritmo PC, e no Apêndice 7 o comparativo entre a utilização de 1 SNP por fenótipo e mais de 1 SNP por fenótipo.

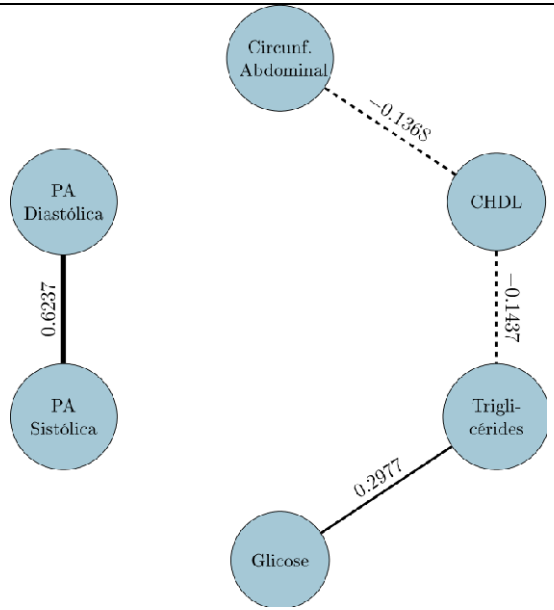
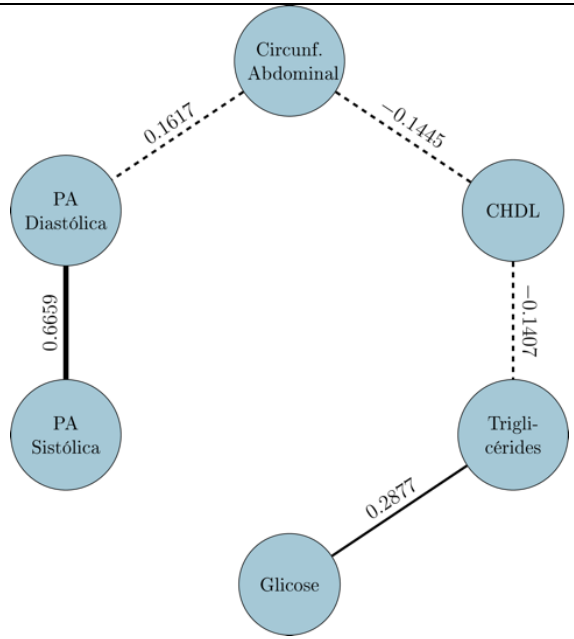
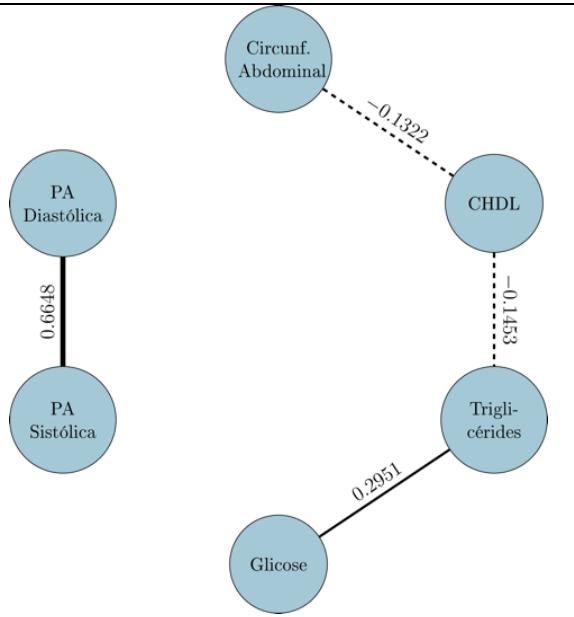
5 VALIDAÇÃO DOS RESULTADOS

Como já mencionado anteriormente, dado o pioneirismo da metodologia adotada neste trabalho, os resultados obtidos são considerados de forma exploratória mais do que conclusivos. A validação dos resultados obtidos é conduzida de duas formas: avaliando a consistência entre as diferentes metodologias empregadas e considerando um subconjunto dos dados consistindo de indivíduos independentes (sem grau de parentesco).

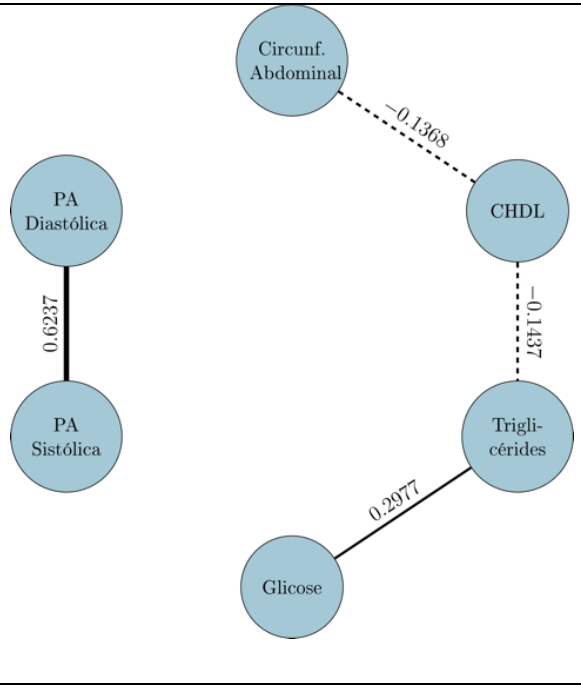
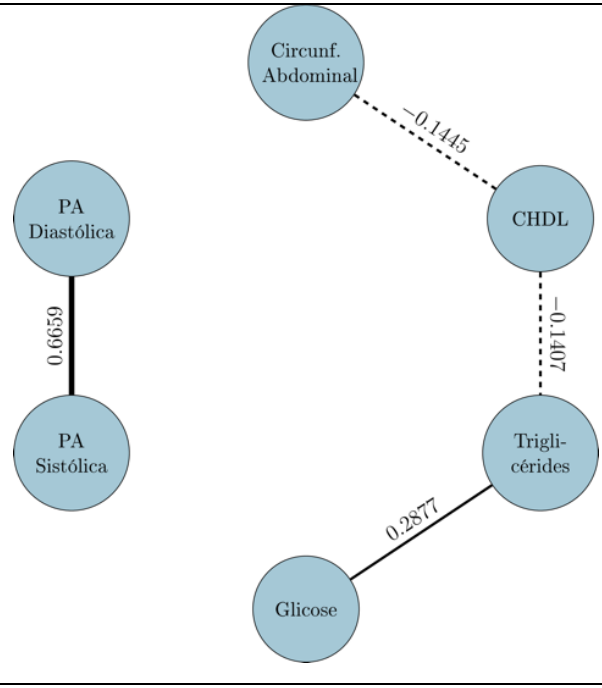
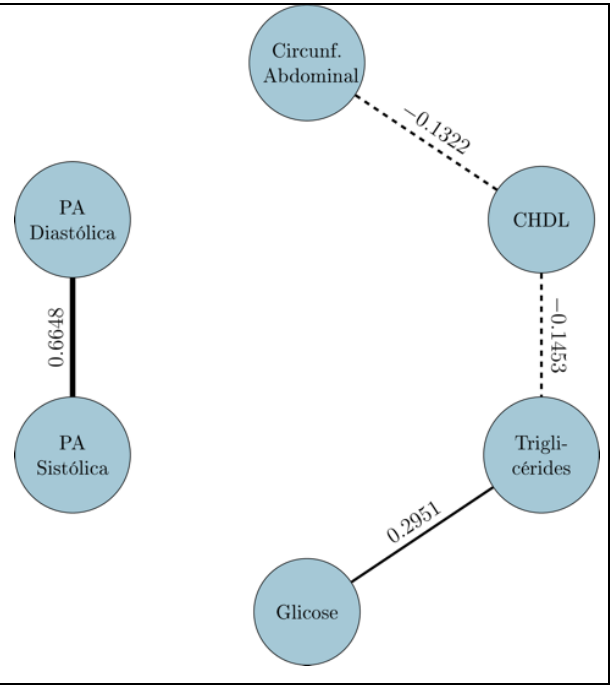
O fato de se estar executando as análises mediante três diferentes metodologias, MLMP_Multivariado, MLMP_Univariado e Algoritmo PC, permite investigar a consistência entre os grafos obtidos em cada etapa, para cada componente, dentro dos diferentes modelos estatísticos e implementações computacionais empregados.

Dessa maneira, como os resultados dos modelos apresentaram-se de maneira consistente, salvo alguns casos específicos, pode-se dizer que eles são robustos desse ponto de vista metodológico.

Adicionalmente, sabe-se que os estudos de aprendizado de redes fenotípicas envolvendo dados de famílias são bastante escassos na literatura (Rosa et al., 2011; Talluri and Shete, 2014), além de envolverem, na maioria dos casos, populações bastante específicas e que, a grande maioria dos estudos dizem respeito a dados de indivíduos independentes. Assim, extraiu-se da base de dados de Baependi os indivíduos fundadores das famílias, obtendo-se assim uma subamostra de indivíduos independentes e, então, aplicou-se a esses dados as metodologias do estudo. Vale dizer que o MLMP_Multivariado foi substituído pela correspondente matriz de covariâncias fenotípicas, neste caso, associada a uma matriz diagonal para a estrutura de indivíduos. Além disso, optou-se por não realizar o ajuste por componentes principais de ancestralidade, visto que, a inclusão destes componentes teve pouco impacto nos resultados para os dados de famílias. Assim, analisou-se os ajustes, sem covariáveis, com as covariáveis sexo, idade e, em seguida, sexo, idade e IMC. Os resultados podem ser vistos na Figura 24.

	Etapa 1 – Sem Covariáveis	Etapa 2 – covariáveis sexo e idade	Etapa 3 – covariáveis sexo, idade e IMC
Matriz de Covariâncias			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

Vizinhança			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

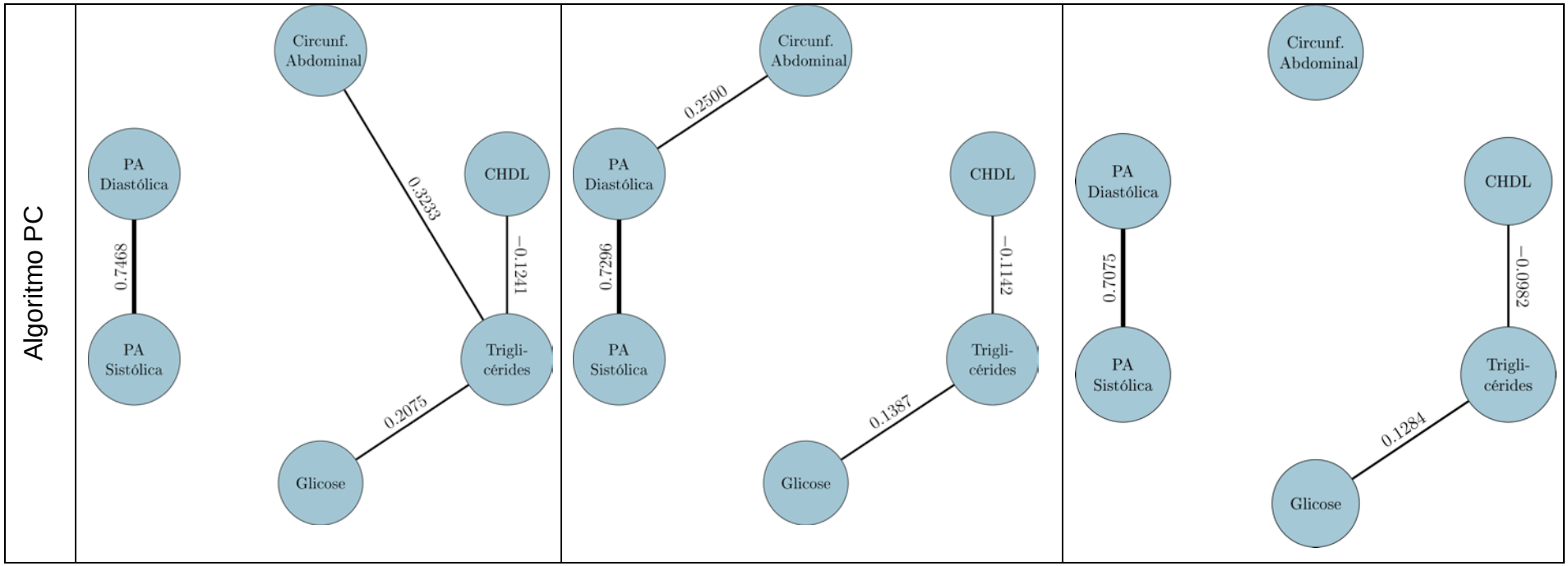


Figura 24: Grafos obtidos a partir de indivíduos considerados fundadores usando-se as três metodologias e as três abordagens.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

Os grafos obtidos são condizentes com aqueles anteriormente obtidos para o modelo envolvendo dados de famílias nas correspondentes etapas consideradas, porém algumas considerações devem ser levantadas. Como esperado, a conexão PAS – PAD manteve-se em todos os grafos, visto que se trata do mesmo fenótipo. Esta relação (conexão) se manteve tanto na análise considerando o componente poligênico (com dados de família) como também com o subconjunto de dados independentes. Contudo, de maneira geral, a comparação dos grafos obtidos sob observações independentes deve ser feita relativamente aos grafos do componente ambiental, obtido da análise de dados de famílias. Os grafos obtidos (Figura 24) se assemelham aos obtidos para o modelo univariado + Análise de Vizinhança (Figura 23), porém, nota-se a presença das conexões entre HDL – Trig no modelos com indivíduos independentes, o que não ocorre para os dados de famílias considerando o componente ambiental. Além disso, a conexão Glicose – Triglicérides também não é observada nos resultados dos componentes ambientais para dados de famílias. Adicionalmente, nota-se a perda da aresta entre CA – PAD para os dados de indivíduos independentes ao passo que esta se mostra uma conexão consolidada nos dados de famílias. Além disso, o fato de os resultados entre as metodologias das matrizes de covariâncias, da análise de vizinhanças e do Algoritmo PC (executado em seu modo clássico), para indivíduos independentes terem apresentado resultados muito semelhantes, já era esperado pois este resultado tem sido exaustivamente demonstrado para o caso de independência. Por outro lado, as semelhanças também encontradas entre os grafos sob independência e para o componente ambiental da análise de famílias, atestam para a validade das metodologias aplicadas neste trabalho. Era esperado que os resultados não fossem exatamente os mesmos, pois a própria estrutura de famílias fornece uma outra forma de tratamento dos dados, bem como várias diferenças são impostas entre as duas análises, por exemplo, o tamanho amostral efetivo que traz grande impacto na significância dos resultados.

6 DISCUSSÃO

O principal objetivo deste trabalho foi o de trazer uma proposta inovadora através da decomposição da rede de associações (causais) entre os fenótipos da Síndrome Metabólica em componente genético e ambiental.

É de fundamental importância ressaltar que, sob esta abordagem, os grafos totais devem ser analisados com cuidado, ou mesmo, sua análise deve ser evitada, visto que não se trata de uma rede combinada dos efeitos de seus componentes, mas sim de uma rede confundida. Apesar disso, os grafos foram elaborados e são apresentados no texto para finalidade de ilustração e discussão desse efeito.

Além disso, torna-se importante mencionar que a matriz de covariância total é a soma das matrizes de covariâncias dos componentes poligênico e ambiental o que significa que, quando os resultados das covariâncias dos componentes apresentam sinais contrários, o resultado da covariância total é diretamente afetado, acarretando em uma redução no valor da correlação parcial e perdendo assim significância. Isso pode ser claramente visto nos grafos da Etapa 3 do Modelo Multivariado (Figura 23) em que o grafo total compreende somente a conexão PAS – PAD.

Deve-se também enfatizar que, antes de se demonstrar que os resultados obtidos são verdadeiros positivos, por exemplo, por meio de análises de poder dos testes, este estudo deve ser considerado ainda exploratório. Deste modo, esforços foram dedicados em explorar diferentes metodologias, diferentes ajustes e critérios de significância, com o objetivo de reduzir tanto resultados falso positivos como falsos negativos.

Neste contexto, a análise proposta neste trabalho permite investigar ou mesmo questionar resultados falsos negativos, ou seja, conexões que poderiam ter ocorrido entre fenótipos, mas que, devido a fatores como o conservadorismo e o tamanho efetivo da amostra adotados nos testes, deixaram de ser encontradas. A busca por falsos negativos ocorreu principalmente no componente poligênico, cujos testes envolvidos no

aprendizado da rede foram muito afetados pelo tamanho efetivo da amostra (que corresponde ao número de graus de liberdade usado na estimação dos componentes de covariâncias) que foi muito baixo, sendo esta uma restrição que demanda muito mais pesquisas, exigindo assim bastante cautela na interpretação dos resultados. O efeito prático disso pode ser ilustrado na construção dos grafos do componente poligênico sob o modelo Multivariado, em que a única aresta que se manteve presente ao longo de todas as etapas foi a de PAS – PAD, o que era não só esperado como desejado, visto que se trata do mesmo fenótipo. Dessa forma, considerando que o tamanho amostral efetivo nestas análises foi cerca de 30 a 40 observações, alguns possíveis resultados falsos negativos são esperados.

Durante a análise, um dos ajustes envolveu somente a inclusão das covariáveis sexo e idade (Apêndice 3 – Grafos), e a aresta Circunferência Abdominal – Glicose ainda estava presente e com coeficiente de correlação parcial de 0,6480, ou seja, uma correlação que pode ser considerada elevada, com p-valor igual a 0,0043. Porém, a partir da inclusão dos componentes principais de ancestralidade PC1 e PC2 houve a perda desta aresta. Uma primeira interpretação deste resultado, é que o efeito anterior ao ajuste pelos PCs estava confundido, possivelmente pela heterogeneidade da população brasileira. Contudo, no caso dos dados de Baependi, esta mudança na significância da aresta pode ser também atribuída à redução no tamanho efetivo da amostra. Na Etapa 1 o tamanho efetivo da amostra que foi calculado em 20,67 foi proveniente de uma amostra total de 1.666 indivíduos. Após a inclusão de PC1 e PC2 passou-se a 1.100 indivíduos (redução de 34%), os quais tiveram dados de SNPs coletados, o que acarretou ainda mais redução do tamanho efetivo da amostra, que passou a 14,24. Dessa maneira, provavelmente não se pode atribuir, sem questionamento, a remoção desta aresta à inclusão de PC1, PC2 e posteriormente dos SNPs, mas argumentar também sobre o baixo tamanho efetivo da amostra que reduz a significância da estatística de teste. Adicionalmente, as correlações parciais entre glicose e as pressões, apesar de não terem fornecido resultados estatisticamente significantes, ocorrendo assim a remoção das arestas, foram altos (para Etapa 2, a correlação parcial entre PAD e Glicose foi de 0,5965 e de PAS e Glicose de -0,6133 e para a Etapa 3 a correlação parcial entre PAD e Glicose foi de

0,6071 e entre PAS e Glicose de -0,6254. Ou seja, tratam-se de prováveis, ou no mínimo questionáveis, falso negativos.

Um outro tópico que deve ser destacado nesta análise é a existência de ambiguidades quando da execução do algoritmo PC. Como parte do próprio algoritmo, existem duas maneiras de se executar a etapa de verificação dos testes. O modo *conservative* verifica todos os testes executados e remove direcionamentos que resultaram em ambiguidades. Já o modo *Major Rule* verifica a maioria dos testes de modo que o resultado final é um grafo com mais arestas direcionadas que o modo anterior. Essas ambiguidades devem ser avaliadas com bastante cautela, pois se trata de uma situação no grafo em que não se pode rejeitar nem a hipótese de cadeia, seguindo qualquer um dos sentidos (A→B→C, ou A←B←C), nem a hipótese de causa comum, formando um garfo (A←B→C).

Assim, um estudo das ambiguidades mostrou que durante a verificação dos testes pelo algoritmo PC, arestas que provavelmente haviam sido anteriormente direcionadas foram removidas em decorrência de ambiguidades entre os fenótipos e SNPs. Isso ocorreu primordialmente na Etapa 3 em que os SNPs foram incluídos.

Outro fator que deve ser levado em conta é o número de SNPs considerados. Quando se adiciona somente um SNP por fenótipo obtém-se um grafo mais direcionado do que quando se adiciona 3 (no caso de Baependi) ou mais (no caso da Literatura). Isso pode ocorrer porque interações entre SNPs, ou seja, epistasias, podem acabar influenciando o direcionamento, gerando um maior número de ambiguidades, bem como a inclusão de mais vértices (de 1 para 3 SNPs) induz a um maior número de testes envolvidos e, portanto, maior dificuldade em encontrar arestas significantes. O trabalho de Kissebah et al. (2000) mostra interações entre locos nos cromossomos 3 e 17 agindo sobre os fenótipos da SM. Apesar disso, o que se observa é a manutenção do esqueleto do grafo entre a Etapa 2 e a Etapa 3, indicando que a influência dos SNPs, sejam 3, seja 1, é muito baixa e não teve “força” para alterar o grafo.

Em termos dos resultados obtidos para o componente ambiental, no modelo Multivariado, permaneceram muitas arestas mesmo que com coeficientes de correlação parcial baixos e isso, provavelmente, está associado ao tamanho efetivo da amostra, neste caso, ser elevado (da ordem da

diferença entre o tamanho amostral total e o número de famílias), o que fez com que correlações parciais relativamente baixas fossem estatisticamente significantes.

Já no modelo Univariado + Análise de Vizinhanças, os coeficientes de correlação parcial do componente poligênico foram baixos, ficando altos somente para o par PAS – PAD, o que, como já mencionado para o caso Multivariado, é esperado e desejado por se tratar de um único fenótipo. Neste modelo, o esqueleto do grafo manteve-se o mesmo com as covariáveis sexo, idade, PC1 e PC2 e após a adição dos SNPs, indicando novamente a baixa influência do tipo de metodologia sobre o grafo encontrado.

No caso dos resultados obtidos por meio do algoritmo PC, é interessante notar que, novamente, assim como no modelo Multivariado, a aresta final no componente poligênico foi entre PAS – PAD. O esqueleto se manteve o mesmo, isto é, sem qualquer aresta, após a adição dos SNPs, no entanto, uma peculiaridade foi observada. O grafo direcionado com os SNPs de Baependi resultou em uma aresta dirigida de PAS → PAD, ao passo que no caso do uso dos SNPs da literatura o direcionamento ocorreu no sentido oposto, de PAD → PAS, sendo que em ambos os casos os coeficientes de correlação parcial foram altos (> 0,74).

Os grafos direcionados do componente ambiental tiveram inicialmente arestas não direcionadas conectando CA – PAS, CA – Glicose e CA – Triglicérides, as quais foram eliminadas com a inclusão de covariáveis. É interessante notar que, não somente os grafos tiveram o mesmo esqueleto, como os mesmos direcionamentos e os mesmos coeficientes de correlação foram obtidos, independentemente do subconjunto de covariáveis. Esse resultado é esperado visto que a atuação dos SNPs ocorre sobre o componente poligênico e não sobre o componente ambiental.

Em termos clínicos, no que diz respeito ao componente ambiental, a relação estabelecida é a da Circunferência Abdominal com a Pressão (Diastólica), porém, pode-se considerar com a pressão de modo geral. Nesse contexto, pode-se dizer que uma intervenção favorecendo a redução da Circunferência Abdominal, teria como consequência direta (causa) a redução da pressão arterial. Apesar de esse fato ser de senso comum e já consolidado,

é interessante ter sido encontrado através da abordagem adotada neste trabalho.

7 CONCLUSÕES

Este foi um estudo inicial e pioneiro na aplicação das metodologias de MLMP_Multivariado, MLMP_Univariado e Algoritmo PC em dados de famílias.

Os resultados obtidos foram bastante interessantes e promissores do ponto de vista metodológico e do uso de covariáveis na eliminação de fatores de confundimento. No entanto, o uso de SNPs não desempenhou o papel esperado na eliminação de fatores de confundimento, possivelmente devido ao reduzido número usado ou ao baixo efeito individual.

Conseguiu-se demonstrar que, analisando-se cada um dos componentes de covariância, pode-se atribuir respostas direcionadas específicas (padrões causais específicos) aos diferentes casos. Pode-se, também, do ponto de vista clínico, fazer intervenções em variáveis precursoras (causais) antes que a variável sucessora (efeitos) seja atingida. Além disso é possível agir de maneira mais contundente sobre as relações relativas ao componente ambiental e saber que essa ação pode ser limitada pelo componente genético.

Apesar disso, este estudo mostrou que o estabelecimento de significâncias nas relações causais entre fenótipos depende dos componentes envolvidos. Na análise destes componentes, do ponto de vista estatístico, é importante adotar estimadores que não estejam confundidos. Neste estudo, para atingir esta finalidade, foi necessário reduzir o tamanho amostral efetivo, principalmente nas estimativas do componente de covariâncias poligênico, o que teve um grande impacto na significância das estatísticas de teste envolvidas. Os resultados também são bastante influenciados pela correção para múltiplos testes adotada. Neste estudo considera-se a correção por FDR bem como opções do algoritmo PC mais conservadoras ou mais moderadas (procedimento “major”), e os resultados dos grafos se mostraram dependentes desta escolha.

Estudos mais detalhados precisam ser desenvolvidos nesta área. Do ponto de vista da Estatística, devem ser estudadas soluções mais calibradas

entre todas essas metodologias e testes envolvidos. Na área Genética, também mais esforços são esperados de modo a tornar possível não só o estabelecimento de uma rede de associações entre os fenótipos com a inclusão de variantes genéticas de maior efeito, mas fazer com que as variantes genéticas se tornem variáveis de maior impacto com o intuito de se construir uma rede de associações entre genótipos e fenótipos. Para o sucesso de qualquer destas iniciativas são necessários desenhos amostrais, métodos estatísticos e recursos computacionais robustos. Este trabalho fez o esforço de explorar os dados de delineamentos familiares na direção da decomposição da estrutura de covariâncias fenotípicas, o que já tem sido bastante explorado em termos de variâncias. Por se tratar de um estudo de caráter pioneiro e exploratório, torna-se importante propor estudos futuros que avaliem os falsos negativos obtidos para que então se torne possível a validação dos verdadeiros positivos através da elaboração de uma estratégia para estabelecer o poder de maneira robusta e contundente, o que, no atual estágio não permite motivação para este desenvolvimento.

Além disso, alguns cuidados devem ser levados em conta no que diz respeito à população alvo deste estudo. Para a realização de inferência causal das variáveis da SM faz-se uso da amostra de famílias de Baependi. Essa amostra tem sido usada em outros estudos para caracterizar a ancestralidade da população brasileira, mas questionamentos devem ser feitos sobre quanto o padrão de associações fenotípicas encontradas nesta amostra são válidas para a população brasileira, bem como se tal padrão pode ser inferido para outras populações mundiais. Além disso, a população de Baependi, pode estar exposta a condições ambientais que podem diferir do restante da população brasileira bem como da de populações alvo de outros estudos.

Adicionalmente, a realização de inferência causal só pode ser garantida pela aceitação de certos pressupostos como suficiência causal, em que todas as variáveis confundidoras foram controladas, aleatorização mendeliana e a seleção de variáveis instrumentais. Os resultados deste estudo deixam questionamentos também sobre o uso destas premissas.

Além disso, o Algoritmo PC apresentou algumas limitações que não permitiram o estabelecimento dos grafos totalmente direcionados, resultando em algumas arestas sem direção.

CRONOGRAMA

2016	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Revisão Bibliográfica			x	x	x	x	x	x	x	x	x	x
Aprimoramento do Projeto			x	x	x							
Tratamento dos Dados											x	x

2017	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez
Revisão Bibliográfica	x	x	x	x								
Exame de Qualificação		x										
Tratamento dos Dados	x	x	x									
Preparação da Dissertação no formato para depósito		x	x	x								
Depósito da Dissertação					x							
Banca						x						

REFERÊNCIAS

AGARWAL, S et al. Metabolic Syndrome derived from principal component analysis and incident cardiovascular events: the multi ethnic study of atherosclerosis (MESA) and Health, Aging and Body composition (Health ABC), Cardiology Research and Practice, **Hindawi Publishing Corporation**, Volume 2012, Article ID 919425, 9 pages, 2012.

ALMASY, L E BLANGERO, J. Multipoint quantitative trait linkage analysis in general pedigrees. **American Journal of Human Genetics,** San Antonio 62:1198-1211, 1998.

ALMASY, L E BLANJERO, J. Variance Component Methods for Analysis of Complex Phenotypes. **Cold Spring Harb Protoc**. 2010(5): PDB.TOP77. DOI:10.1101/PDB.TOP77, 2010.

ALMASY, L. The role of phenotype in gene discovery in the whole genome sequencing era. **Hum. Genet**; 131(10):1533-1540, 2012.

ANDERSON, PJ. Factor analysis of the metabolic syndrome: obesity vs insulin resistance as the central abnormality. **International Journal of Obesity**, 25, 1782-1788, 2001.

AUSTIN, MA et al. Heritability of multivariate factors of the metabolic syndrome in non diabetic Japanese Americans. **Diabetes**, vol 53, 1166-1169, 2004.

BLANGERO, J. et al. Novel family-based approaches to genetic risk in thrombosis. **J Thromb Haemost**: 1: 1391-7, 2003.

DE ANDRADE et al. Extension of Variance Components Approach to Incorporate Temporal trends and longitudinal Pedigree Data Analysis. **Genetic Epidemiology**. 22:221-232, 2002.

DE ANDRADE et al. Global Individual Ancestry Using Principal Components for Family Data. **Human Heredity** 80:1-11, 2015.

DE ANDRADE et al. Methods to Estimate Genetic Components of Variance for Quantitative Traits in Family Studies. **Genetic Epidemiology**. 17:64-76, 1999.

BAE et al. Learning Bayesian Networks from Correlated Data. Scientific Reports. **Nature.** 6:25156, 2016.

BANERJEE, O et al. Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. **Journal of Machine Learning Research**, Berkley. 9:485-516, 2008.

BELIA, A et al. “The Linosa Study”: Epidemiological and heritability data of the Metabolic Syndrome in a Caucasian Genetic Isolate. Nutrition, Metabolism & Cardiovascular Diseases. **Elsevier**. 19: 455-461

BLANN, AD. Angiogenesis, the Metabolic Syndrome and heart disease: Is there a connection? **Journal of Internal Medicine**. 268, 312-315, 2010.

BOLLEN, KA. Structural Equations with Latent Variables. **John Willey & Sons**, 1989.

BOUCHARD, G. Genetics and the metabolic syndrome. **International journal of obesity**. 19: 52–59. Retrieved 2009-04-29, 1995.

BURGES, S et al. Network Mendelian Randomization: using genetic variants as instrumental variables to investigate Mediation in Causal Pathways. **International Journal of Epidemiology**, dyu 176, 2014.

DUARTE, NE et al. On the Equivalence for Methods for Population Stratification and their Application in Genetic Association Studies. **Revista Brasileira de Biometria** 33(4), 2015.

DRTON, M e MAATHUIS, MH. **Structure Learning in Graphical Modeling**. arXiv:1606.02359, Ithaca, 2016.

EGAN, KJ, et al. Cohort profile: the Baependi Heart Study—a family-based, highly admixed cohort study in a rural Brazilian town. **BMJ Open**; 6:e011598. doi:10.1136/bmjopen-2016-011598, São Paulo, 2016.

ECKEL, RH et al. The metabolic syndrome. **Lancet**. Volume 365, No. 9468, p1415–1428, 2005.

FAUCI, AS. Harrison's principles of internal medicine. **McGraw-Hill Medical**. ISBN 0-07-147692-X. New York, 2008.

FOULKES, AS. Applied Statistical Genetics with R – For population-based association studies. Springer, Amherst, MA, 2009.

FRIEDMAN, J et al. Sparse inverse covariance estimation with the graphical lasso. **Biostatistics**. 9, 3, pp. 432–441 doi:10.1093/biostatistics/kxm045, Stanford, CA, 2008.

GIOLO et al. Brazilian Urban Population genetic structure reveals a high degree of admixture. **European Journal of Human Genetic**. 19: 111-116, 2011.

GROOP, L. Genetics of the metabolic syndrome. **British Journal of Nutrition**. 83 (S1): S39–S48. doi:10.1017/S0007114500000945. Retrieved2009-04-29, 2000.

HILLIER, TA et al. Practical way to assess metabolic syndrome using a continuous score obtained from Principal Components Analysis. **Diabetologia**. 49(7):1528-35, 2006.

HORIKOSHI, M et al. Transancestral fine-mapping of four type 2 diabetes susceptibility loci highlights potential causal regulatory mechanisms. **HMG Advance Access**. Oxford, UK, 2016.

KALISCH, M et al. Causal Inference Using Graphical Models with the R Package pcalg. **Journal of Statistical Software**. Issue 11. Volume 27. 2012.

KARN, R et al. Modeling Metabolic Syndrome Through Structural Equations of Metabolic Traits, Comorbid Diseases, and GWAS Variants. **Obesity**. 21, E745-E754, 2013.

KAUR, J. A Comprehensive Review on Metabolic Syndrome. **Cardiology Research and Practice**. Hindawi Publishing Corporation. Volume 2014, Article ID 943162, 21 pages, 2014.

KISSEBAH, AH et al. Quantitative trait loci on chromossomes 3 and 17 influence phenotypes of the metabolic syndrome. **PNAS**, Vol 97, n. 26, 14478-14483, 2000.

LIN, HF et al. Heritabilities of the Metabolic Syndrome and its componets in the Northern Manhattan Family Study. **Diabetologia**. 48(10): 2006-2012, 2005.

LOY, A e HOFMANN, H. Are you normal? The problem of confouded residual structures in hierarchical linear models. **Journal of Computational and Graphical Statistics**, 24:4, 1191-1209, 2015.

MALNIK, SDH e KNOBLER, H. The Medical Complications of Obesity. **QJMed**. 99: 565-579, 2006.

MARKUS, KA. Structural Equation Modeling. In Rogelberg, SG. Encyclopedia of Industrial and Organizational Psychology. **Sage Knowledge**. ISBN 9781412952651, pp. 774–777, 2009.

MEI H, et al. Principal-Component based multivariate regression for genetic association studies of metabolic syndrome components. **BMC Genetic**, 11:100, 2010.

MEIGS, JB et al. Prevalence and Characteristics of the Metabolic Syndrome in the San Antonio Heart and Framingham Offpring Studies. **Diabetes**. Vol 52. 9: 2160-2167. 2003.

MITCHELL, BD et al. Genetic and Environmental Contributions to Cardiovascular Risk Factors in Mexican Americans. **Circulation**. 94:2169-2170, 1996.

MULLER, M. Adipose tissue innate immunity & inflammation – a nutrigenomics perspective of the Metabolic Syndrome. **Health of Medicine, Technology**.

Netherlands Nutrogenomics Centre & Nutrition, Metabolism and Genomic Group, Division of Human Nutrition, Wagenengen University. 2011.

NATIONAL HEART, LUNG AND BLOOD INSTITUTE. **What is Metabolic Syndrome**. 2016. Disponível em: https://www.nhlbi.nih.gov/health/health-topics/topics/ms Acessado em: 10/02/2017

NICOLAI M e BÜHLMANN, P. High-Dimensional Graphs and Variable Selection with the Lasso. **The Annals of Statistics**. Vol. 34, No. 3, 1436–1462 DOI: 10.1214/009053606000000281 Institute of Mathematical Statistics, 2006.

OLIVEIRA, C et al. Heritability of cardiovascular risk factors in a Brazilian population: Baependi Heart Study; **BMC Medical Genetics**. 9:32 doi:10.1186/1471-2350-9-32, 2008.

OLVERA, RL et al. Common genetic influences on depression, alcohol and substances use disorders in Mexican-American families. **Am. J. Med Genet Neuropsychiatr Genet**, 156(5): 561-568, 2011.

OUALKACHA, K et al. Principal Components of Heritability for High Dimension Quantitative Traits and General Pedigrees. **Statistical Applications in Genetics and Molecular Biology**. Vol. 11: Iss. 2, Article 4. DOI: 10.2202/1544-6115.1711, 2012.

PANAZZOLO, DG et al. Obesity, metabolic syndrome, impaired fasting glucose and microvascular dysfunction: a principal component analysis approach. **BMC Cardiovascular Disorders**, 12: 102, 2012.

PEARL J. Causality: models, reasoning, and inference. 2 ed. **Cambridge University Press**. 464p, 2009.

PENA, SDJ et al. The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected. **PLoS ONE** 6(2): e17063. doi:10.1371/journal.pone.0017063, 2011.

PENALVA, DQF. Síndrome metabólica: diagnóstico e tratamento. **Rev Med**. 87(4):245-50. São Paulo, 2008.

PETTERSON, LE e COLEMAN, MA. Machine learning-based receiver operating characteristics (ROC) curves for crisp and fuzzy classification of DNA microaarays in cancer research. **Int. J. Approx. Reason**; 47(1): 17-36, 2008

POLLEX, R.L. e HEGELE, R.A. Genetic determinants of the metabolic syndrome. **Nat. Clin. Pract. Cardiovasc. Med.** 3 (9): 4829. doi: 10.1038/ncpcardio0638. PMID 16932765, 2006.

POULSEN, P et al. Genetic versus environmental etiology of the metabolic syndrome among male and female twins. **Diabetologia**. 44 (5): 537–543. doi:10.1007/s001250051659. PMID 11380071. Retrieved 2009-04-29, 2001.

PRINS, BP et al. Genetics of coronary artery disease: Genome-wide association studies and beyond. **Atherosclerosis** 225, 1-10, 2012.

RAFFA, JD e THOMSON, EA. Power and effective study size in heritability studies. Statistics in Biosciences. **Springer**, 2016.

RIBEIRO, AH et al. Causal Inference and Structure Learning of Genotype-Phenotype Networks using Genetic Variation. In: Wang, KC. Big Data Analytics in Genomics. **Springer International Publishing**, Switzerland. Book Part I, pp 89-143, 2016.

ROSA, GJ et al. Inferring causal phenotype networks using structural equation models. **Genet Sel Evol**, 43(6), 2011.

SHALIZE, CR. **Advanced Data Analysis from na Elementary Point of View**. Carnege Mellon University. **Cambridge University Press**, 2015.

SOUTO, JC et al. Genetic susceptibility to thrombosis and its relationship to physiological risk factors: the GAIT study. **Am. J. Hum. Genetic**, 67: 1452-1459, 2000.

SPIRTES, P et al. Causation, prediction, and search. **MIT press**, 2000.

SUNG, J et al. Heritabilities of the Metabolic Syndrome phenotypes and related factors in Korean Twins. **J Clin, Endocrinol Metab**, 94(12): 4946-4952, 2009.

SUNYER, XP. The Medical Risk of Obesity. **Postgrad Med**. 121(6): 21-33, 2009.

TALLURI, R e SHETE, S. Gaussian graphical models for phenotypes using pedigree data and exploratory analysis using network with genetic and nongenetic factors based on Genetic Analysis Workshop 18 Data. **BMC Proceedings**. 8(Suppl 1): S99, 2014.

TANG, W et al. Familial Clustering for features of the Metabolic Syndrome. **Diabetes Care**. Vol 29. Number 3, p. 631-636, 2006.

TAVARES, HP et al. Metabolic Syndrome: Consensus and Controversy: State of the Art. **Open Journal of Endocrine and Metabolic Diseases**, 5, 124-130. Scientific Research Publishing, 2015.

VALENTE, BD et al. Searching for recursive causal structures in multivariate quantitative genetics mixed models. **Genetic**, 185(2), 633-644, 2010.

VAN DONGEN et al. Heritability of Metabolic Syndrome traits in a large population-based sample. **Journal of Lipid Research.** Vol 54(10). p. 2914-2923, 2013.

VATTIKUTI, S et al. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. **Plos Genetics**,vol 8, Issue 3, e1002637, 2012.

WANG Y et al. A ridge penalized principal-components approach based on heritability for high-dimensional data. **Hum Hered**, 64, 182-191, 2007a.

WANG Y et al. Clustering and principal component analysis for mapping co-regulated genome-wide variation using family data. **BMC Proceedings**, Vol.1. S121.2007b. Supplement 1.

YANG, Y ey al. Effective sample size: quick estimation of the effect of related samples in genetic case-control association analyses. Comput Biol Chem. 35(1): 40-49, 2011.

YUAN M e LIN Y. Model selection and estimation in the Gaussian graphical model. **Biometrika** 94:19-35, 2007.

ZHANG et al. A comprehensive analysis of adiponectin QTLs using SNPs association, SNP cis-effect on peripheral blood gene expression and gene expression correlation identified novel metabolic syndrome (MetS) genes with potential role in carcinogenesis and systemic inflammation. **BMC Medical Genomics**. 6:14, 2013.

ZIMMET, P et al. The Metabolic Syndrome: A Global Public Health Problem and a New Definition. **Journal of Atherosclerosis and Thrombosis**, Vol. 12, No. 6. 295-300, 2005.

ANEXO 1

Teoria dos Grafos (Extraído de Drton, Mathias & Maathuis, Marloes H, 2016 e de RIBEIRO, AH et al, 2016)

Grafos

Os modelos de grafos representam as dependências estocásticas em um conjunto de variáveis aleatória X_v, $v \in V$. Cada modelo está associado com um grafo G=(V,E), onde os vértices V indexam as variáveis e as conexões E impõem um conjunto de dependências condicionais. Especificamente, X_v e X_w devem ser condicionalmente independentes dado $X_c := (X_u: u \in C)$, denotado por $X_v \perp X_w | X_c$, se todos os caminhos entre os nós v e w em G estiverem adequadamente bloqueados pelos nós em C.

Cadeias de Markov compõem um exemplo de modelo de grafos e suas propriedades são utilizadas neste trabalho para mostrar a equivalência entre os modelos de equações estruturais (SEM) e modelos de grafos.

Para isso, serão descritos os grafos não direcionados (UDG) e os grafos acíclicos direcionados (DAG), bem como o aprendizado de estruturas para ambos os tipos, através das propriedades de Markov para pares de Variáveis (M_Pares), Local (M_Local), e Global (M_Global).

Os grafos não direcionados referem-se àqueles em que as conexões são pares não ordenadas (v,w) para v, $w \in V$ distintos. Já os grafos acíclicos direcionados têm as conexões formando pares ordenados (v,w) que representam a conexão apontando de v para w. Também pode-se escrever $v \rightarrow w \in E$.

Os UDGs representam as relações entre as vizinhanças das variáveis e as correlações parciais. Já os DAGs trazem o conceito de ancestrais e descendentes, já que pressupõem relações causais, independências condicionais e d-separação.

Modelos de Grafos Não Direcionados (UDG)

Seja $X = (X_v: v \in V)$ um vetor aleatório indexado pelos vértices de um grafo não direcionado G=(V,E). Então, X satisfaz M_Pares em relação a G se $X_v \perp X_w | X_{V\setminus\{v,w\}}$ sempre que $\{v,w\} \notin E$.

Adicionalmente, X satisfaz M_Local com relação a G se $X_v \perp X_{V\setminus(nbG(v)\cup\{v\}} | X_{nbG(v)}$ para todo $v \in V$, quando se recorda que $nb_G(v) = \{w \in V: \{w,v\} \in E\}$ é um conjunto de vizinhos de v. Finalmente, X satisfaz M_Global em relação a G se $X_A \perp X_B / X_C$ para todos os trios de subconjuntos pareados disjuntos A, B, $C \subset V$ de modo que C separa A e B em G, ou seja, todo caminho entre o nó A e o nó B contém o nó C.

Nota-se que a M_Global engloba M_Local e M_Pares. Também se demonstra que M_Local implica M_Pares.

Apesar de não ser sempre verdadeiro, as três propriedades de Markov são equivalentes quando X satisfaz o axioma da intersecção para independência condicional. Essa equivalência é verdadeira em particular quando X_v são discretas com probabilidade conjunta positiva, ou quando X tem uma densidade positiva e contínua com relação à medida de Lebesgue.

M_Pares para modelos de grafos não direcionados traduz cada aresta ausente em uma independência condicional “cheia”. Por esta razão, o menor grafo G não direcionado quando X é M_Pares também é conhecido como grafo condicional independente de X. Testar todos os $\binom{|V|}{2}$ pareados de independência cheia que podem surgir de $X_v \perp X_w | X_{V\setminus\{v,w\}}$ resulta em um método para estimar esses grafos. Isso permite controlar descobertas de falsas conexões.

Sob M_Local, cada variável X_v pode ser otimamente predita de sua vizinhança $X_{nb}(v)$, o que é usado em um método conhecido como seleção da vizinhança. M_Global é muito útil para resolver questões relativas a independência condicional.

Segundo o teorema de Hammersley-Clifford, se $X = (X_v: v \in V)$ tem uma densidade positiva com relação a uma medida produto, então a distribuição X fatora com respeito a G = (V,E), se e somente se, X satisfizer M_Pares com relação a G.

Um modelo de grafos pode ser definido especificando famílias de funções de probabilidade. Se as funções forem positivas, então M_Pares e M_Global são equivalentes. Adicionalmente, M_Global lista toda independência condicional que é mantida em todas as distribuições com densidades que fatoram.

O modelo Gaussiano associado com um grafo não direcionado G=(V,E) compreende todas as distribuições normais com a inversa da matriz de covariância positiva definida $K=(k_{vw}) \in \mathbb{R}^{VxV}$de modo que $k_{vw} = 0$ quando v-w $\notin$ E.

Seja S uma estimativa de uma matriz de covariância para uma amostra Gaussiana i.i.d. de tamanho *n*. O estimador de máxima verossimilhança (MLE) de $K = \Sigma^{-1}$ no modelo de grafos Gaussianos maximiza a função log-verossimilhança:

$$L(K) = logdet(K)\text{-}tr(SK) \tag{A.1}$$

sujeito a ser matriz positiva definida com zeros correspondendo a arestas de G sem conexão. Se o grafo G for esparso, o MLE de K pode existir unicamente com probabilidade um mesmo se *n* for muito menor que |V|.

Modelos de Grafos Acíclicos Direcionados

DAGs são grafos acíclicos direcionados. Um vetor aleatório $X=(X_v: v \in V)$ satisfaz M_Local em relação a um DAG G se $X_v \perp_X_{ndG}(v)\backslash p_{aG}(v)|Xp_{aG}(v)$ para todo $v \in V$.

Similarmente, X satisfaz M_Global em respeito a G se $X_A \perp X_B|X_C$ para todos os trios de conjuntos pareados disjuntos A, B, C $\subset$ V de modo que C d-separa A e B em G, que será denotado por $A \perp_G B|C$.

Se X satisfizer M_Global em relação a G, então G é chamado um mapa de independência de X. Um DAG G é um mapa perfeito de X se $A \perp_G B|C$, se e somente se, $X_A \perp X_B|X_C$ para todos os conjuntos disjuntos pareados A, B, C $\subset$ V. Um mapa perfeito, portanto, requer M_Global e sua implicação reversa, só ocorre sob Fidelidade de G. A suposição de que G é um mapa perfeito é importante para os métodos de aprendizado de estruturas.

Suponha que X tem uma densidade em respeito a uma medida produto. Então, a distribuição de X fatora de acordo com um DAG G=(V,E) se tiver uma densidade da forma:

$$f(x) = \prod_{v \in V} f(x_v | x(v)_{paG}), \tag{A.2}$$

em que pa_G são os descendentes/pais do nó v em G, de_G são os descendentes de v em G, nd_G são os não descendentes de v em G, $f(x_v|x(v)_{paG})$ são as densidades condicionais com $f(x_v|x_\varnothing)=f(x_v)$.

M_Local e M_Global são equivalentes e sob a suposição de existência de uma densidade positiva eles são equivalentes à propriedade de fatoração. M_Global também é completa, ou seja, estabelece todas as independências condicionais que são implicadas pela factoração da densidade.

Um modelo de grafo dirigido para um DAG G pode ser especificado pelas densidades condicionais $f(x_v|x(v)_{paG})$, v $\in$ V. Se eles não compartilharem parâmetros comuns, a função de verossimilhança do modelo fatora em |V| funções locais de verossimilhança, e o MLE de $f(x_v|x(v)_{paG})$, v $\in$ V pode ser calculado separadamente. Para dados normais multivariados, as médias e variâncias condicionais são encontradas através de regressão linear de cada X_v em $X_{paG}(v)$.

O conceito de d-separação é muito importante em DAGs e envolve a definição de X → Z ← em que dois nós, v e w, em um DAG G=(V,E) são d-

conectados dado $C \subseteq V\backslash\{v,w\}$ se G contém um caminho Π com extremos v e w de modo que

(i) Todos os colisores em Π são em $an_G(C)$, e

(ii) Nenhum não colisor em Π está em C.

Se $\pi = (v_0, v_1, ..., v_k)$ é um caminho, então v_0 e v_k são os extremos do caminho. Um não extremo vi é um colisor em π se $v_{i-1} \rightarrow v_i \leftarrow v_{i+1}$ é um sub-caminho de π. Caso contrário, v_i é um não colisor em π. Se cada conexão em π for da forma $v_{i-1} \rightarrow$ vi, então v_0 é um ancestral de v_k, e v_k é um descendente de v_0.

A notação $adj_G(v)$ se refere a um nó adjacente de v em G, $an_G(v)$ se refere a um ancestral de v em G e $de_G(v)$ se refere um descendente de v em G.

Generalizando para conjuntos, dois subconjuntos disjuntos $A, B \subset V$ são d-conectados dado $C \subseteq V\backslash(A \cup B)$ se existem dois nós $v \in A$ e $w \in B$ que estão d-conectados dado C. Se este não for o caso, então C d-separa A e B.

Dois DAGs G e G' são Markov equivalentes se $A \perp_G B|C$ é equivalente a $A \perp_{G'} B|C$. DAGs Markov equivalentes são caracterizadas por terem o mesmo esqueleto e a mesma estrutura v, denominada colisão não conectada.

Uma estrutura v é um trio de nós $u \rightarrow v \leftarrow w$ com u e w não adjacentes. Cada classe Markov equivalente pode ser representada por um grafo acíclico direcionado parcialmente completo (CPDAG) que pode ter conexões direcionadas ou não. Um CPDAG tem uma conexão $v \rightarrow w$, se e somente se, o conector $v \rightarrow w$ for comum em todos os DAGs em sua classe equivalente. Se a classe contiver um DAG com $v \rightarrow w$ e um com $v \leftarrow w$, então o CPDAG tem um conector não direcionado v – w.

Os modelos de grafos associados com um DAG G=(V,E) também podem ser pensados como Modelos de Equações Estruturais. De fato, se $\varepsilon=(\varepsilon_v: v \in V)$ for um vetor de variáveis (termos de erros) aleatórias independentes e g_v forem funções de densidade, então o vetor aleatório $X=(X_v: v \in V)$ dado por:

$$X_v=g_v(Xpa_G(v), \varepsilon_v), v \in V \tag{A.3}$$

é Markov em respeito a G. De maneira contrária, se X é Markov com respeito a G, então existem variáveis independentes ε_v e funções de densidade g_v suportadas pela equação acima.

Se todas as funções g_v forem lineares e ε_v forem variáveis aleatórias Normais, então se pode assumir que a equação acima toma a forma de $X = BX + \varepsilon$, onde a matriz $B=(\beta_{vw}) \in \mathbb{R}^{V x V}$ tem $\beta_{vw} = 0$ se $w \in pa_G(v)$. A solução $X = (I-B)^{-1}\varepsilon$ segue uma distribuição normal multivariada com matriz de covariância $Cov(X)=(I-B)^{-1}Cov(\varepsilon)(I-B)^{-T}$. Aqui, I-B é inversível dado que det(I-B)=1 por aciclicidade de G.

SEM e, portanto, todos os modelos de grafos direcionados admitem uma interpretação causal natural. Dessa forma, pode-se ver a equação acima como um mecanismo assinalado, que é esclarecido por:

$$X_v \leftarrow g_v(Xpa_G(v), \varepsilon_v), v \in V. \tag{A.4}$$

As variáveis em $Xpa_G(v)$ são tratadas como causas diretas de X_v, o que significa que mudanças em $Xpa_G(v)$ podem levar a mudanças em X_v, mas não o inverso. Essa interpretação permite determinações sobre a distribuição de X sob intervenções experimentais. Em particular, intervenções no sistema podem ser modeladas por alterar as equações estruturais para precisamente aquelas variáveis que são afetadas pelas intervenções.

A interpretação causal de um DAG permite predições em ambientes com intervenção e, portanto, a estimação dos efeitos causais. Essas ideias podem ser combinadas com os métodos de aprendizado de estruturas para estimar efeitos causais a partir de dados observacionais.

Modelos de Aprendizado em Grafos Não Direcionados

Existem diversos modelos de Aprendizado em grafos não direcionados. Para este estudo serão utilizados:

- Modelos Gaussianos, e

- Seleção de Vizinhança com possível Penalização L1.

Modelos gaussianos fornecem um ponto de partida para muitas modelagens de grafos de observações contínuas. Um grafo gaussiano de independências condicionais pode ser estimado pela determinação das entradas de zeros da inversa da matriz de covariância K $\in \mathbb{R}^{VxV}$.

Seja S=(s_{vw}) a amostra da matriz de covariância para uma amostra de *n* observações, com $s_{vv} > 0$ para todo v $\in$ V. Yuan&Lin(2007) e Benerjée et al (2008) propuseram o estimador de grafos lasso (glasso).

$$\widehat{K^{gl}} = arg \min_{K}\{-logdet(K) + tr(SK) + \lambda||K||_1 \tag{A.5}$$

em que a minimização é sobre matrizes positivas definidas K=(Kvw) $\in \mathbb{R}^{VxV}$ e $\lambda \geq 0$ é um parâmetro de penalização no ajuste.

Defazio & Caetano (2012) estimaram a covariância inversa K=(K_{vw}) pela solução ótima de

$$\min_{k}\{-logdet(K) + tr(SK) + \sum_{v \in V}\sum_{j=1}^{|V|} \lambda_j |K_{v,(j)}|\} \tag{A.6}$$

Além do Modelo Gaussiano, pode-se utilizar o Modelo de Seleção de Vizinhança.

Pode-se estimar um grafo de independências condicionais G=(V,E) estimando-se todas as suas vizinhanças $nb_G(v)$. De acordo com a propriedade local de Markov, X_v depende das outras variáveis apenas através de suas vizinhanças X_w, w $\in$ $nb_G(v)$. Portanto, pode-se prosseguir estimando a distribuição condicional de X_v dado $X_{V\setminus\{v\}}$ e determinar $nb_G(v)$ como o conjunto

indexado das variáveis X_w sobre as quais a distribuição condicional estimada depende.

Seja $X=(X_v: v \in V)$ uma normal multivariada com inversa da matriz de covariância $K=(K_{vw})$. A distribuição condicional de X_v dado todas as variáveis restantes é normal com variância $1/K_{vv}$ e esperança:

$$\mathbb{E}[X_v|X_w, w \neq v] = \sum_{w \in V\{v\}} \left(\frac{-k_{vw}}{k_{vv}}\right) X_w = \sum_{w \in nbG(v)} \left(\frac{-k_{vw}}{k_{vv}}\right) X_w \tag{A.7}$$

Pode-se estimar $nb_G(v)$ como o conjunto de covariáveis ativas significantes em uma regressão linear de X_v em todas as outras variáveis X_w, $w \neq v$.

Estimar cada vizinhança isoladamente pode levar a inconsistências que são geralmente resolvidas post-hoc. Sejam $\widehat{\mathrm{nb}(v)}$, $v \in V$, as vizinhanças estimadas. Se $w \in \mathrm{nb}(v)$, mas $v \in \mathrm{nb}(w)$, então a chamada regra AND exclui a conexão v – w do grafo de independência condicional estimada, enquanto a regra OR inclui essas conexões.

Aprendizado de Modelos de Grafos Direcionados

Assume-se que o DAG G=(V,E) é um mapa perfeito de X-(Xv: $v \in V$), e que se observa uma amostra aleatória de tamanho n i.i.d. de X. Os dados observados são denotados por x.

De modo geral, G não é identificável a partir da distribuição de X, mas se pode identificar sua classe Markov equivalente, ou seu CPDAG.

Existem diversos métodos para estimar a direção das conexões dos grafos. Neste trabalho será utilizado o pacote PC Algorithm do software R.

Para fazer uso do pacote PC Algorithm utiliza-se um método baseado em Restrições, que será explicado a seguir.

A entrada é a matriz de covariância (Σ).

O objetivo é encontrar um DAG que seja compatível com as independências condicionais do conjunto de dados. Inicia-se a discussão desses métodos focando na estimação do esqueleto do DAG (ou equivalentemente, seu CPDAG), que é tipicamente a etapa que necessita de recursos computacionais mais intensos.

Se G é um mapa perfeito de X, então v e w são adjacentes em G se, e somente se, X_v e X_w forem condicionalmente independentes dado X_c para todo $C \subseteq V\backslash\{v,w\}$.

O PC Algorithm (Spirtes et al, 2000) limita o número de testes de independência condicional, usando que v e w são não adjacentes em um DAG G se, e somente se, $X_v \perp X_w | X_{pag(v)}$ ou $X_v \perp X_w | X_{pag(w)}$. Dado que não se tem os conjuntos de pais (o que significaria conhecer o DAG), o PC Algorithm assegura que as seguintes independências condicionais sejam testadas: $X_v \perp X_w | X_C$ para todo $C \subseteq adj_G(v)$ e todo $C \subseteq adj_G(w)$, e remove a conexão se uma independência condicional é encontrada. Isso ainda parece inviável, no entanto, dado que não se conhece $adj_G(v)$ e $adj_G(w)$. O PC Algorithm contorna esse problema sempre trabalhando com supergrafos de esqueleto verdadeiro, e testando as independências condicionais dado o subconjunto de conjuntos adjacentes nesses supergrafos.

Concretamente, o PC Algorithm começa com um grafo completo em V. Em seguida, testa independências marginais para todos os pares de nós, e remove uma conexão se uma independência for encontrada. Em seguida, para cada par de nós que ainda são adjacentes, são testadas independências condicionais dados todos os subconjuntos de cardinalidade 1 dos conjuntos adjacentes aos dois nós. O algoritmo remove uma conexão se uma independência condicional for encontrada. O algoritmo continua dessa maneira, cada vez aumentando a cardinalidade do conjunto condicionado em 1, até que a cardinalidade do conjunto condicionante exceda $\max_{v \in V} |adj_{G'}(v)| - 1$, onde G' é o estado atual do esqueleto. Neste ponto, todas as independências requeridas foram consideradas, e o esqueleto foi encontrado.

Na segunda fase, o PC Algorithm pós-processa os resultados dos testes de independência condicional e aprende as estruturas-v. Finalmente, o

algoritmo aplica algumas regras simples de orientação para orientar as conexões faltantes, baseado no fato de que não podem ser criados ciclos direcionados, nem estruturas-v. Se as independências condicionais reais forem usadas como input para o PC Algorithm, o output final é o CPDAG do DAG G. Nota-se que para grafos com grau limitado, isto é, um limite em $\max_{v \in V}|adj_G(v)|$, o PC Algorithm tem um período de processamento que é polinomial no número de variáveis. O tempo de processamento depende exponencialmente do grau.

Exemplo: suponha que se tem três variáveis aleatórias X_u, X_v, X_w com $X_u \perp X_w$ como a única independência condicional. O PC Algorithm inicia com um grafo completamente não-direcionado em {u,v,w}. Quando testa para independências marginais, remove a conexão u-w após encontrar $X_u \perp X_w$. Nenhuma outra independência condicional é encontrada, então u-v-w é o esqueleto final. Em seguida, o algoritmo detecta que v é um colisor no caminho (u,v,w), pois caso contrário $u \perp_G w$ é violado. Portanto, o DAG (e o CPDAG correspondente) é $u \rightarrow v \leftarrow w$.

O mesmo esqueleto é encontrado se a única independência condicional for $X_u \perp X_w / X_v$. Então, no entanto, v deve ser um não colisor no caminho (u,v,w) para assegurar que $u \perp_G w|v$ se mantém. Isso deixa 3 DAGs possíveis: $u \rightarrow v \rightarrow w$, $u \leftarrow v \leftarrow w$ ou $u \leftarrow v \rightarrow w$, que forma uma classe equivalente de Markov representada pelo CPDAG u – v – w.

Na prática, independências condicionais necessitam ser testadas baseadas em dados. Testes padrão estão disponíveis para dados multivariados Gaussianos e dados Multinomiais. Usualmente, todos os testes são executados no mesmo nível de significância α, e uma conexão é removida se a hipótese nula de independência condicional não for rejeitada. Aqui, α não controla o erro Tipo I total. Ele é um parâmetro de ajuste que tipicamente fornece grafos esparsos para valores menores.

ANEXO 2

Correspondência entre Teoria dos Grafos e Modelagem de Equações Estruturais (adaptado de Ribeira, AH et al, 2016)

Modelos de equações estruturais (SEMs) e modelos de Grafos Probabilísticos (PGMs) são metodologias amplamente adotadas para representar possíveis associações causais entre variáveis. Enquanto SEMs fornecem uma representação funcional desses possíveis mecanismos causais pelos quais o valor de uma variável é gerado, PGMs podem, de maneira equivalente, representar a mesma relação causal combinando teoria de grafos e teoria de probabilidade (Ribeiro et al., 2016).

Combinando elementos das duas metodologias é possível usar uma linguagem matemática intuitiva e rigorosa para inferência causal (Ribeiro et al., 2016).

Representação Funcional Causal

De uma forma geral, um modelo causal é um modelo de equações estruturais que representa as possíveis relações causais entre as variáveis aleatórias.

Considerando as variáveis aleatórias V_i, i = 1, ..., n. Seja pa(V_i) o conjunto de pais de V_i, ou seja, o conjunto de variáveis aleatórias que diretamente determina o valor de V_i. Também, seja e_i, i = 1,..., n as variáveis aleatórias representando erros devido a causas desconhecidas.

O mecanismo através do qual o valor de cada variável V_i é selecionado pode ser definido de acordo com alguma função f_i (usualmente, mas não necessariamente, linear) das variáveis pais e das variáveis erro:

$$V_i = f_i(pa(V_i); e_i); \quad i = 1;\ldots; n \qquad (1)$$

No contexto de redes causais de fenótipo-genótipo e Aleatorização Mendeliana, as variáveis aleatórias são fenótipos quantitativos e SNPs (variantes genéticas associadas com esses fenótipos). Adicionalmente, é

comumente assumido que as relações funcionais da SEM mostrados na expressão (1) são lineares, e que as variantes genéticas são variáveis instrumentais de efeito fixo (Ribeiro et al, 2016).

A SEM linear causal representando o padrão esperado de associações entre os p fenótipos observados e os *q* SNPs para *n* indivíduos é:

$$Y = M + PY + QX + E \qquad (2)$$

Em que:

Y = matriz *pxn* em que cada elemento y_{ij} representa o valor observado do *i-ésimo* fenótipo para o *j-ésimo* indivíduo;

M = $1' \otimes \mu$ uma matriz *pxn* resultante do produto de Kronecker entre o vetor unitário n-dimensional transposto e μ o vetor com os valores esperados de cada fenótipo quantitativo;

Q = matriz *pxq* dos SNPs nos fenótipos. O elemento q_{ik} representa o efeito do *k-ésimo* SNP no *i-ésimo* fenótipo;

X = matriz *qxn* onde cada elemento x_{kj} representa o genótipo observado do *k-ésimo* SNP no *j-ésimo* indivíduo.

P = matriz *pxp* conhecida como matriz de adjacência, contendo os efeitos causais diretos dos fenótipos entre si. O elemento p_{ii} corresponde ao efeito do *ji-ésimo* fenótipo no *i'-ésimo* fenótipo;

E = matriz *pxn* em que cada e_{ij} representa medidas de erro do *i-ésimo* fenótipo do *j-ésimo* indivíduo.

Note que a premissa de que variações nos SNPs precedem variações nos fenótipos é expressa pelo fato de que a matriz Q está sempre do lado direito da equação 2, junto com os efeitos fenotípicos pais, representados na matriz P (Ribeiro et al, 2016).

A matriz P, também pode ser chamada de matriz de adjacência, que, no contexto da correspondência entre Grafos e SEM assume um importante papel. A matriz de adjacência é uma matriz de ordem *pxp*, tal que *p* é o número de vértices. Se *P[i,j] = 1*, existe aresta de i a j. Se *P[i,j] = 0*, não existe aresta de

i a *j*. No caso de grafos direcionados, *P[i,j]* deve indicar ou não a presença de uma aresta divergente de *i* e convergente em *j*, ou seja $i \rightarrow j$.

No caso de marcadores moleculares bialélicos, os efeitos genéticos totais (representado pelas matrizes Q e X) podem ser particionados com o intuito de explicitar efeitos particulares, como efeitos aditivos e de dominância de cada marcador e os possíveis efeitos de interação entre eles (efeitos de epistasia). Tipicamente, marcadores bialélicos têm três estados genotípicos, denotados usualmente por AA (homozigoto dominante), Aa (heterozigoto) e aa (homozigoto recessivo). Esses genótipos devem ser codificados de acordo com o tipo de efeito usando até dois graus de liberdade (Ribeiro et al, 2016).

Efeitos de dominância e interação de genes são muito importantes no mapeamento genético. Porém, com o intuito de realizar inferência causal a partir de estudos observacionais, é preferível selecionar as variáveis instrumentais como variantes genéticas (com efeito aditivo) mais relevantes associadas a cada fenótipo, procurando ser o mais parcimonioso possível (Ribeiro et al, 2016).

Representação Causal Gráfica

O modelo causal da expressão (1) pode ser graficamente representado por um grafo direcionado.

Assim, se $V = \{V_1, \ldots, V_n\}$ for um conjunto finito e $E \subseteq \{(V_i,V_j): V_i, V_j \in V\}$ for um conjunto de pares ordenados de vértices, cada elemento do conjunto V é chamado vértice e cada elemento do conjunto E é chamado de aresta direta. A aresta $(V_i;V_j)$ representa uma conexão direta de V_i com V_j. O par ordenado G = (V,E) é chamado grafo direcionado (Ribeiro et al, 2016).

Sempre se pode construir uma representação gráfica de um modelo de equação estrutural usando um grafo direcionado.

Nesta representação, cada vértice do grafo direcionado corresponde a uma variável aleatória distinta. Adicionalmente, cada aresta (V_i, V_j), se existir, representa uma relação funcional direta da variável Vi com a variável V_j. Neste caso, V_i é chamada pai de V_j e V_j é um filho do vértice V_i. A ausência de uma aresta indica que as variáveis não estão diretamente associadas. Portanto, se desenhar, de cada variável V_i, uma aresta apontando para ela de cada um de

seus parentes, constrói-se o grafo direcionado que procura representar o mecanismo causal. Os termos de erro não são representados no grafo. No entanto, quando os termos de erro estão correlacionados, os pares de variáveis correspondentes devem estar conectados por uma seta apontando para ambos. Quando a relação implica uma possível causalidade, a representação gráfica do modelo causal é referida como grafo causal ou diagrama causal do sistema. A descoberta do grafo causal de um sistema a partir de dados observacionais é o objetivo dos métodos de aprendizado de estruturas causais (Ribeiro et al, 2016).

A seguir define-se as duas classes de grafos de acordo com seus padrões de ciclos. Essa distinção é importante pois muitos resultados e procedimentos para inferir relações causais usando dados observacionais são dependentes da estrutura do grafo (Ribeiro et al, 2016).

Um caminho direto entre dois vértices é uma sequência de arestas direcionadas que começam em um vértice e terminam em outro vértice, com a restrição de que todas as arestas estão orientadas na mesma direção. Sempre que existe um caminho que começa e termina no mesmo vértice tem-se um ciclo. Ciclos de comprimento um são chamados de auto-loops e ciclos de comprimento dois correspondem a uma influência bidirecional ou associação recíproca (Ribeiro et al, 2016).

DAG (Grafo Acíclico Direcionado): Um SEM com termos de erros não correlacionados e que não contém relações cíclicas são chamados de SEM recursivos e sua representação gráfica é chamada de grafo acíclico direcionado (do inglês *directed acyclic graph* - DAG).

Sistemas biológicos reais têm comportamento cíclico natural. Assim, DAGs podem ser bastante restritivos para modelar dados biológicos.

DCG (Grafo Cíclico Direcionado): Um SEM que contém pelo menos um ciclo é chamado de SEM não recursivo e sua representação gráfica é chamado de grafo cíclico direcionado (do inglês *directed cyclic graph* - DCG).

Do ponto de vista algébrico, as arestas conectando fenótipos correspondem aos elementos diferentes de zero em P. Se a matriz P puder ser rearranjada como uma matriz triangular inferior, tem-se um modelo recursivo, e consequentemente, pode ser representada como um DAG. Caso contrário, o sistema contém ciclos e um SEM não recursivo e um DCG é usado para

representá-lo (Ribeiro et al, 2016).

Qualquer SEM pode ser representado por grafos direcionados. Para esse estudo específico, assume-se que os fenótipos (e SNPs, quando aplicável) formam SEMs recursivos e, portanto, são representados por DAGs, além disso, assume-se que os termos de erro são não correlacionados.

Conforme mencionado no início desta seção, algumas definições e conceitos são necessários para que seja possível o entendimento sobre a relação de correspondência entre os modelos de equações estruturais e a teoria de grafos.

APÊNDICE 1
GRÁFICOS MANHATTAN PLOT

Os gráficos abaixo apresentados foram construídos utilizando ajuste para Idade, Idade2, Sexo, os componentes principais da análise de ancestralidade (PC1 e PC2) e os SNPs.

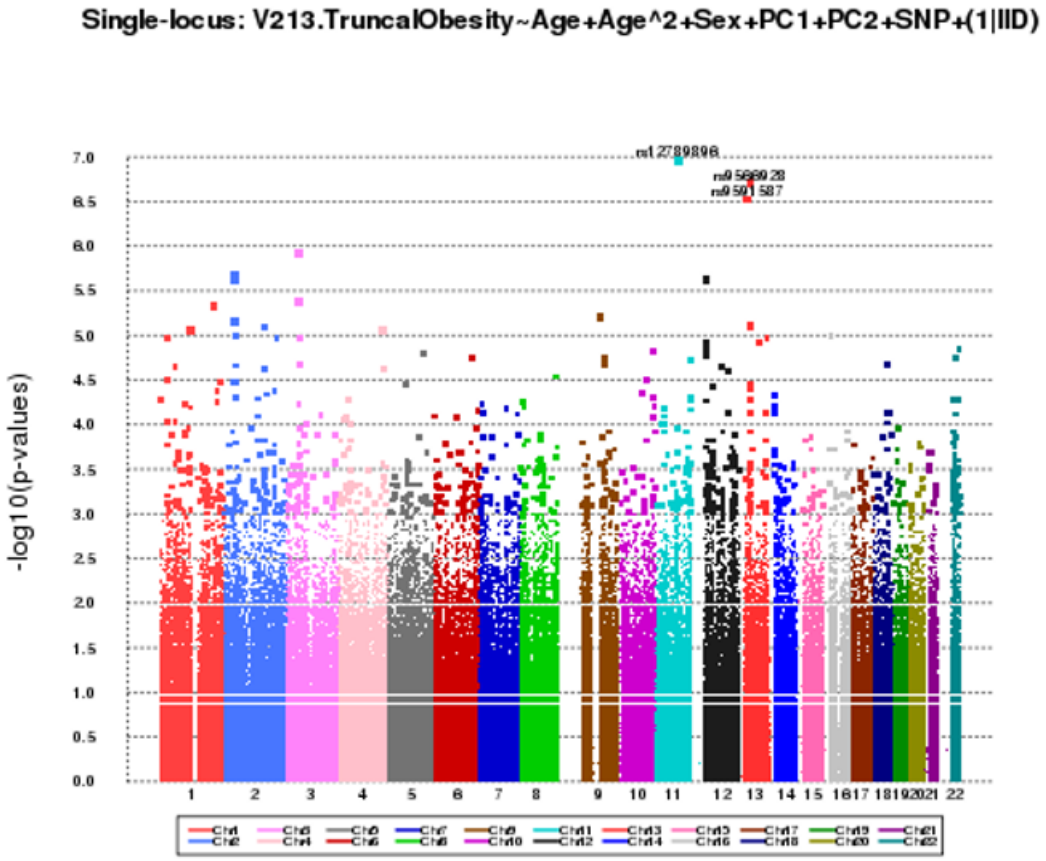


Figura 25: SNPs relativos a Obesidade Truncal.

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	11	rs12789896	83885744	-12,44777	2,342930	-5,31	1,1e-07
2	13	rs9591587	31660717	-12,67000	2,473293	-5,12	3,0e-07
3	13	rs9566928	42708389	-22,34099	4,298361	-5,20	2,0e-07

Tabela 19: Lista de SNPs significantes: Obesidade Truncal

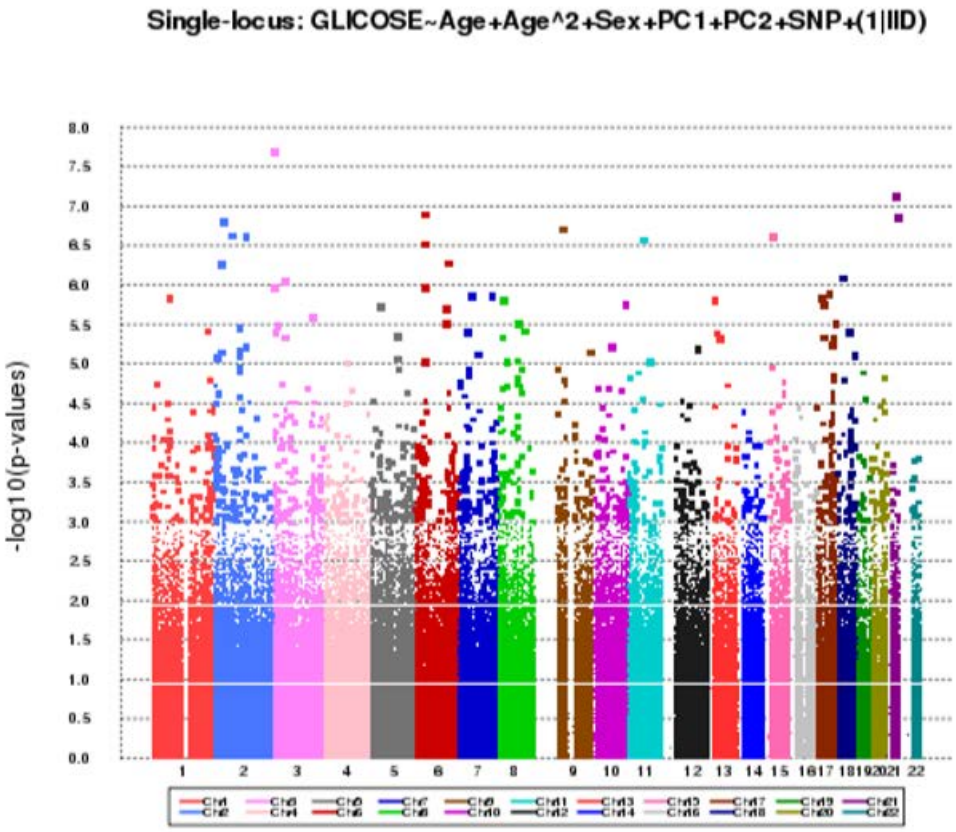


Figura 26: SNPs relativos a Glicose

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	2	Rs625318	33670020	60,94017	12,177312	5,00	5,6e-07
2	2	Rs2430491	37431107	52,86505	10,096781	5,24	1,6e-07
3	2	Rs2303604	71828423	20,12157	3,897021	5,16	2,4e-07
4	2	Rs2897468	129711722	34,03020	6,593211	5,16	2,5e-07
5	3	Rs6778092	4060388	69,52777	12,412269	5,60	2,1e-07
6	3	Rs17076353	44081531	41,55941	8,461720	4,91	9,0e-07
7	6	Rs16885036	34262716	25,69580	5,023639	5,11	3,1e-07
8	6	Rs16889851	34401378	22,83505	4,324490	5,28	1,3e-07
9	6	Rs12529550	130405707	23,27372	4,644620	5,01	5,4e-07
10	9	Rs10967925	27402597	29,86880	5,745375	5,20	2,0e-07
11	11	Rs7116722	58506908	27,94334	5,431932	5,14	2,7e-07
12	11	Rs11826261	58515163	27,94334	5,431932	5,14	2,7e-07
13	15	Rs16961362	35944606	24,09356	4,670927	5,16	2,5e-07
14	18	Rs641876	25629250	35,79785	7,265339	4,93	8,3e-07
15	21	Rs8130136	40314554	68,59901	12,768932	5,37	7,8e-07
16	21	Rs9982923	44811720	47,00565	8,936822	5,26	1,4e-07

Tabela 20: Lista de SNPs significantes: Glicose

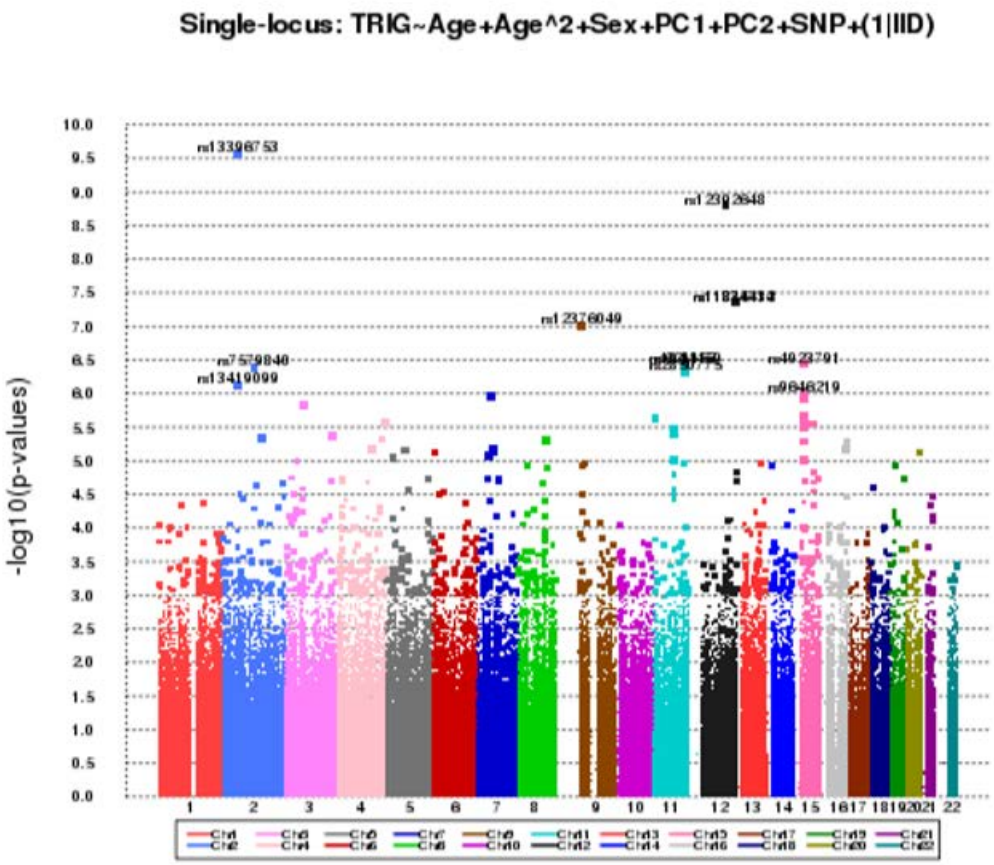


Figura 27: SNPs relativos ao Triglicérides

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	2	Rs13419099	59110526	114,92721	23,217874	4,95	7,4e-07
2	2	Rs13396753	59117009	173,95623	27,535277	6,32	2,7e-07
3	2	Rs7579840	119921881	74,73194	14,750614	5,07	4,1e-07
4	9	Rs12376049	9058319	32,36513	6,070922	5,33	9,8e-07
5	11	Rs2850775	121899825	48,46149	9,621570	5,04	4,7e-07
6	11	Rs1944459	121913807	51,65040	10,162192	5,08	3,7e-07
7	11	Rs621153	121939228	51,60921	10,164735	5,08	3,8e-07
8	12	Rs12302648	91129466	139,59286	23,150682	6,03	1,6e-07
9	12	Rs11834734	126679318	117,16598	21,403283	5,47	4,4e-07
10	12	Rs11835412	126680250	117,16598	21,403283	5,47	4,4e-07
11	15	Rs4923791	38689071	37,70029	7,406419	5,09	3,6e-07
12	15	Rs9646219	38697638	39,54359	8,065988	4,90	9,5e-07

Tabela 21: Lista de SNPs significantes: Triglicérides

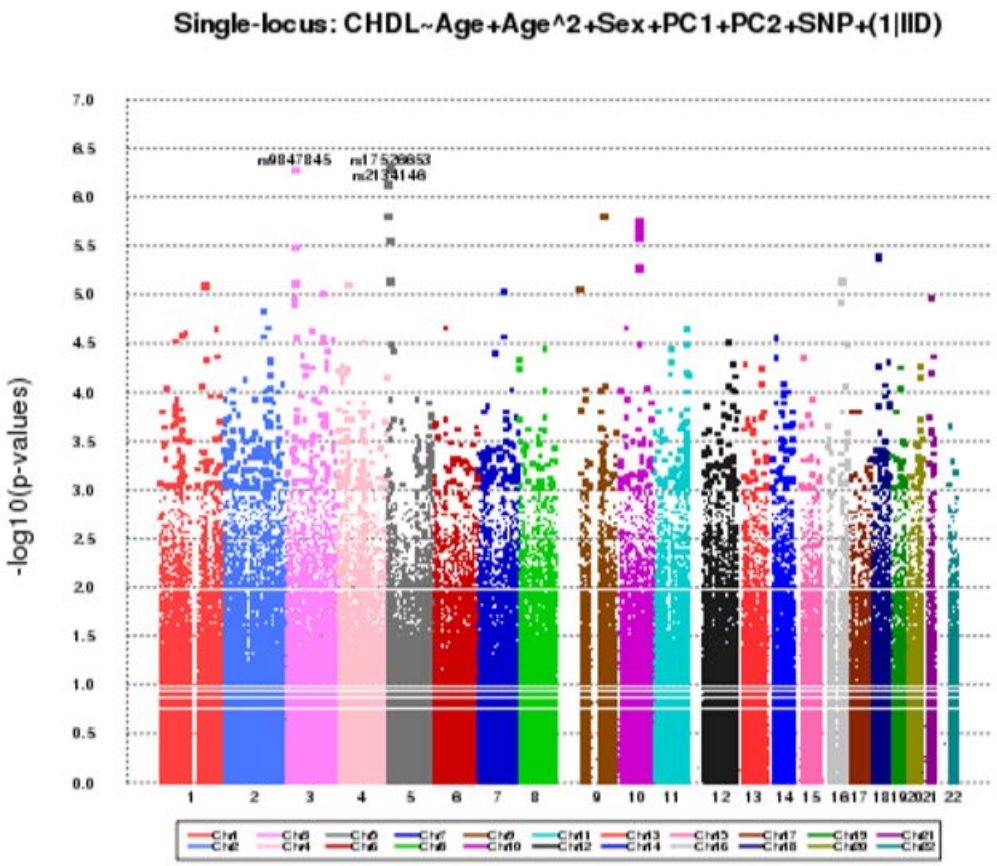


Figura 28: SNPs relativos a cHDL

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	3	Rs9847845	33699640	3,786651	0,7550115	5,02	5,3e-07
2	5	Rs2134146	5062539	-13,487290	2,7272975	-4,95	7,6e-07
3	5	Rs17526653	13905149	-18,607603	3,7059357	-5,02	5,1e-07

Tabela 22: Lista de SNPs significantes: cHDL

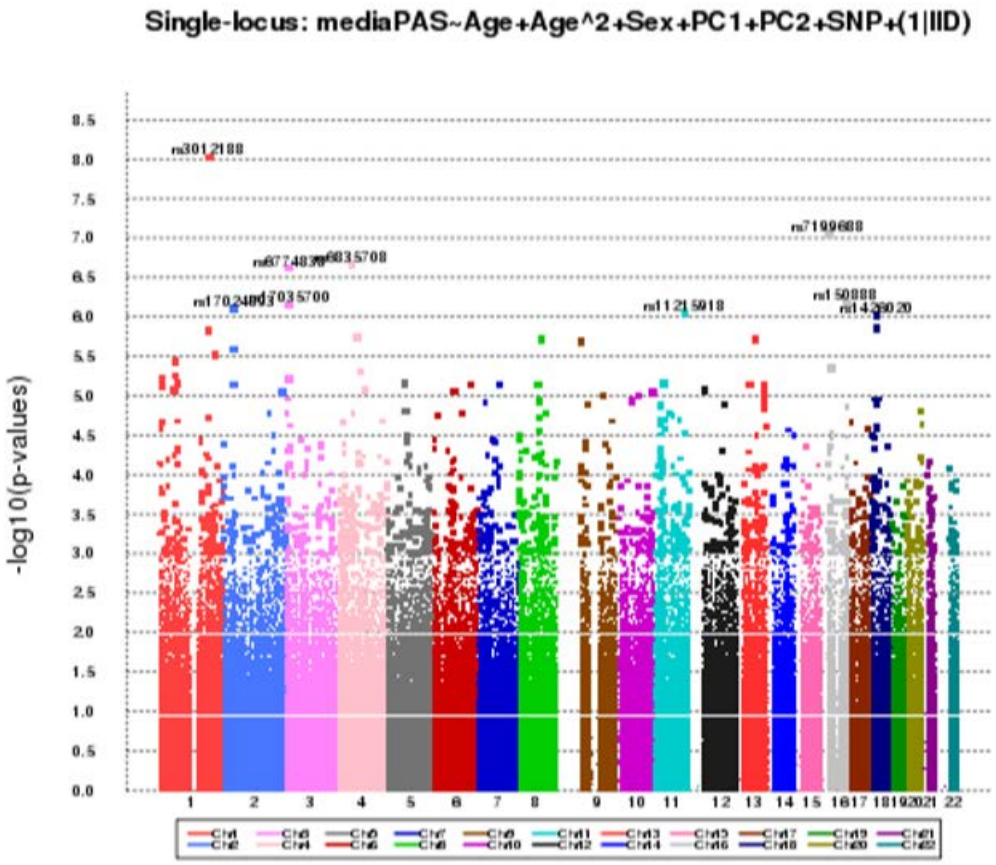


Figura 29: SNPs relativos a média da PAS

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	1	Rs3012188	192036732	14,148012	2,465425	5,74	9,5e-07
2	2	Rs17024893	40260755	15,170536	3,073279	4,94	8,0e-07
3	3	Rs17035700	12068350	-16,597201	3,347782	-4,96	7,1e-07
4	3	Rs6774838	12147403	-12,857824	2,488887	-5,17	2,4e-07
5	4	Rs6835708	54655240	-20,626021	3,980577	-5,18	2,2e-07
6	11	Rs11215918	116273232	-12,434134	2,530270	-4,91	8,9e-07
7	16	Rs7199688	7401712	20,956210	3,915409	5,35	8,7e-07
8	16	Rs150888	76081390	4,230583	0,851109	4,97	6,7e-07
9	18	Rs1426020	19713200	-15,576837	3,179007	-4,90	9,6e-07

Tabela 23: Lista de SNPs significantes: Média PAS

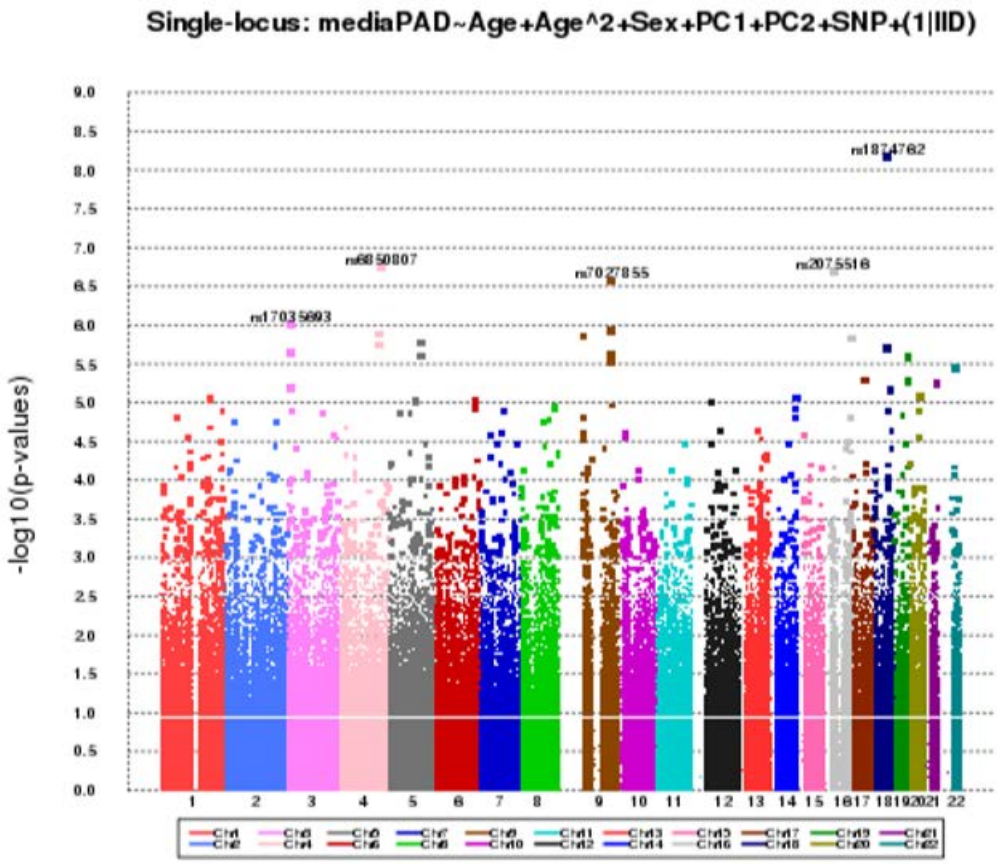


Figura 30: SNPs relativos a média da PAD

	CHR	SNP	BP	Estimate	SE	Teststat	Pvalue
1	3	Rs17035693	12068064	-7,074102	1,4453074	-4,89	9,9e-07
2	4	Rs6850807	164204180	-17,098592	3,2776157	-5,22	1,8e-07
3	9	Rs7027855	111376977	-2,809183	0,5459288	-5,15	2,7e-07
4	16	Rs2075516	15869525	2,654402	0,5106886	5,20	2,0e-07
5	18	Rs1874762	51465914	-18,028866	3,1103564	-5,80	6,8e-07

Tabela 24: Lista de SNPs significantes: Média PAD

APÊNDICE 2

QUADROS COM SNPs RELATIVOS À SM EXTRAÍDOS DA LITERATURA INTERNACIONAL

Gordura abdominal / Visceral				
Fenótipo	*Gene mais próximo*	*Variante*	*p-value*	*Cromossomo*
Circunferência da Cintura - WC (cm) (Zang, 2013)	CDH18	rs10061119	2,18776 e-05	5
		rs133357704	2,18776 e-05	5
		rs10805723	2,69153 e-05	5
		rs16887308	2,69153 e-05	5
		rs4386736	2,69153 e-05	5
		rs6879762	2,69153 e-05	5

Quadro 3: SNPs relacionados a Circunferência da Cintura (Zang, 2013)

Glicemia Alterada					
Fenótipo	*Gene mais próximo*	*Variante*	*p-value*	*OR*	*Cromossomo*
Fasting Glucose - FG (mmol/l) (Zang, 2013)	mRNA TAG1	rs7721328	2,6915 e-04		5
T2D (Horikoshi, 2016)	IGF2BP2	rs11705729	1,3 e-21	1,14	3
	CDKAL1	rs9368222	4,1 e-30	1,17	6
	CDKN2A-B	rs10965246	8,4 e-40)	1,31	9
		rs10757282	2,0 e-12	1,12	9
	KCNQ1	rs231353	1,7 e-11	1,11	11
		rs233448	9,5 e-12	1,12	11
		rs2237897	7,7 e-12	1,19	11
T2D (Prins, 2012)	MTAP	rs7865618		1,18	9
	SMG6, SRR	rs216172		1,07	17
	UBE2Z, GIP, ATP5G1, SNF8	rs46522		1,06	17
	LDLR	rs11226008		1,15	19
	SLC5A3, MRPS6, KCNE2	rs9982601		1,20	21

Quadro 4: SNPs relacionados a Glicemia (Zang, 2013; Horikoshi, 2016; Prins, 2012)

Dislipidemia					
Fenótipo	*Gene mais próximo*	*Variante*	*p-value*	*OR*	*Cromossomo*
HDL Colesterol - HDL-c (mmol/l) (Zang, 2013)	TRIO mRNA	rs6883134	5,6234e-04		5
Triglicérides (Prins, 2012)	HCG27, HLA-C	rs3869109		1,14	6
	PHACTR1	rs12526453		1,12	6
	c6orf105	rs6903956		1,51	6
	TCF21	rs12190287		1,08	6
Triglicérides - TG (mmol/l) (Zang, 2013)	mRNA BC028204	rs16887451	7,9433e-04		5
HDL, hipertrigliceridimia, triglicérides (Prins, 2012)	ZNF259, APOA5-A4-C3-A1	rs964184		1,13	11
	BRAP fs	rs11066001		1,42	12
	ALDH2 fs	rs671		1,43	12
	COL4A1, COL4A2	rs4773144		1,07	13
	HHIPL1	rs2895811		1,07	14
	ADAMTS7, MORF4L1	rs3825807		1,08	15
	ADAMTS7	rs4380028		1,07	15
	RASD1, SMCR3, PEMT	rs12936587		1,07	17

Quadro 5: SNPs relacionados a Dislipidemia (Zang, 2013; Prins, 2012)

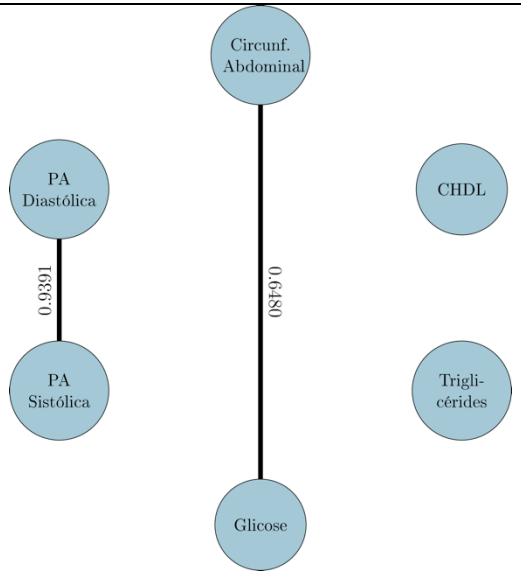
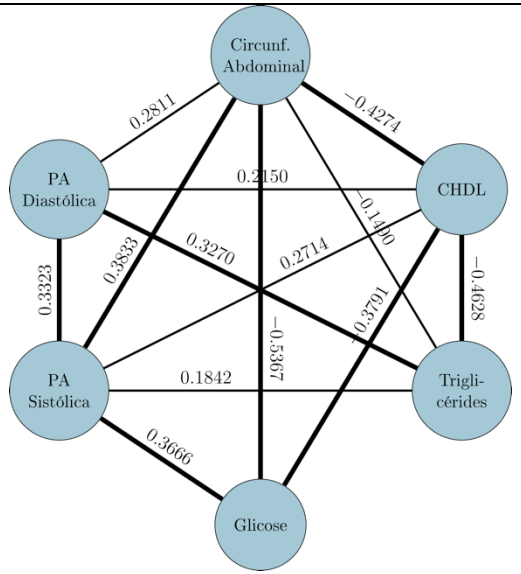
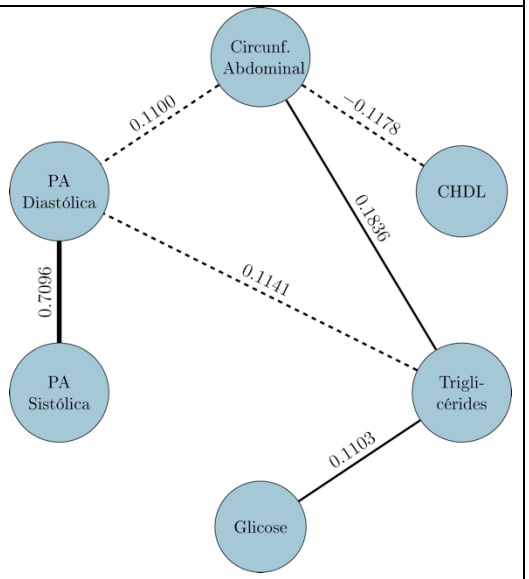
Pressão Arterial Aumentada					
Fenótipo	*Gene mais próximo*	*Variante*	*p-value*	*OR*	*Cromossomo*
Pressao Sistólica - sBP (mmHg) (Zang, 2013)	DAP	rs5745297	2,6303e-04		5
Pressao Diastolica - dBP (mmHg) (Zang, 2013)	mRNA BC028204	rs6450583	5,3703e-04		5
Pressao sanguínea sistólica (Prins, 2012)	LIPA	rs1412444		1,1	10
	CYP17A1, CNNM2, NT5C2	rs12413409		1,12	10
	PDGFD	rs974819		1,07	11

Quadro 6: SNPs relacionados a Pressão Arterial (Zang, 2013; Prins, 2012)

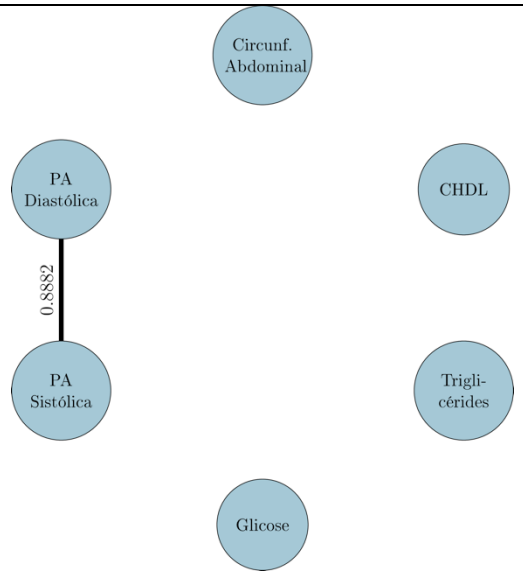
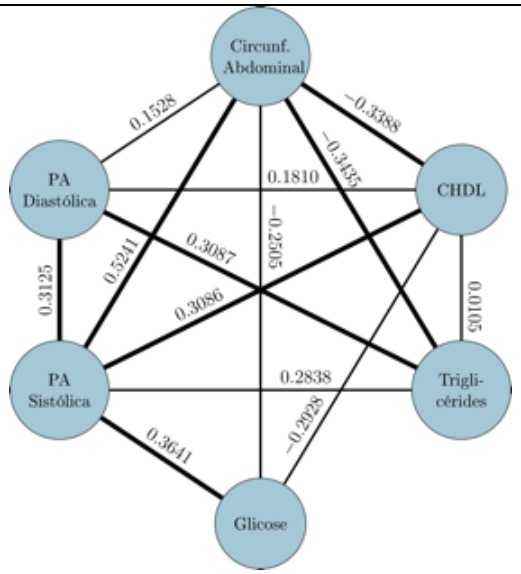
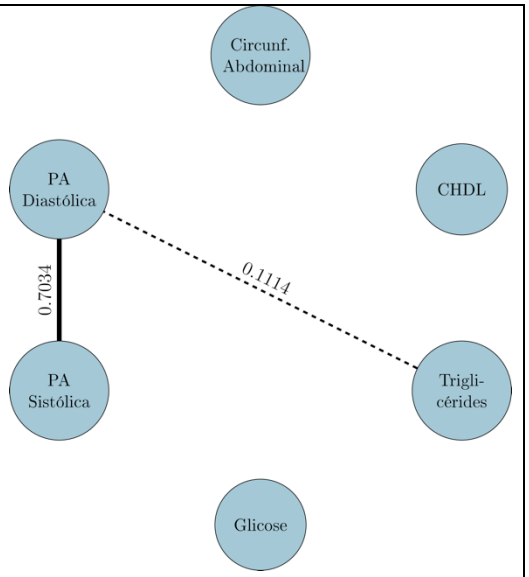
APÊNDICE 3

GRAFOS

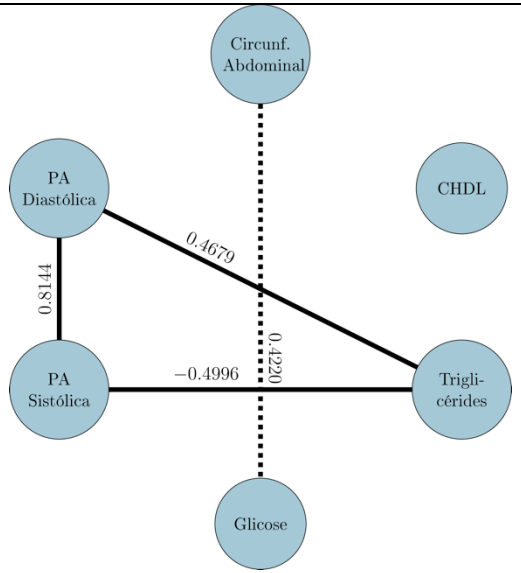
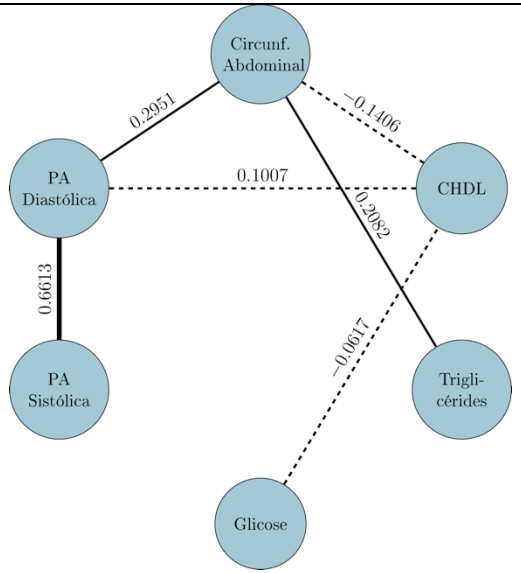
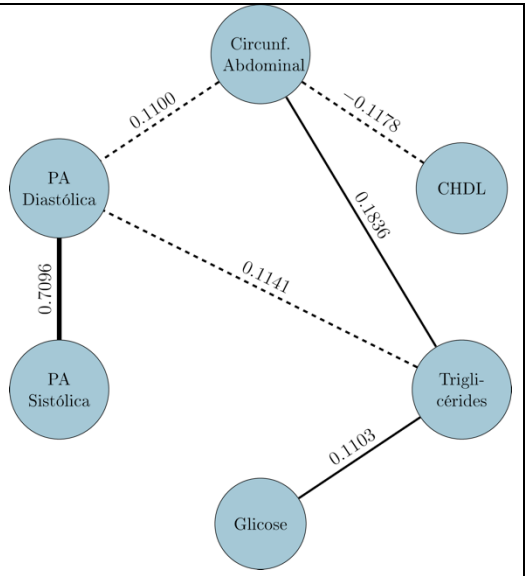
Abaixo seguem os grafos que fizeram parte da construção deste estudo, como a inclusão das covariáveis sexo e idade, bem como das covariáveis sexo, idade e IMC.

		COMPONENTE		
MODELO	ETAPA	POLIGÊNICO	AMBIENTAL	TOTAL
Multivariado	2- Com covariáveis Sexo e Idade			

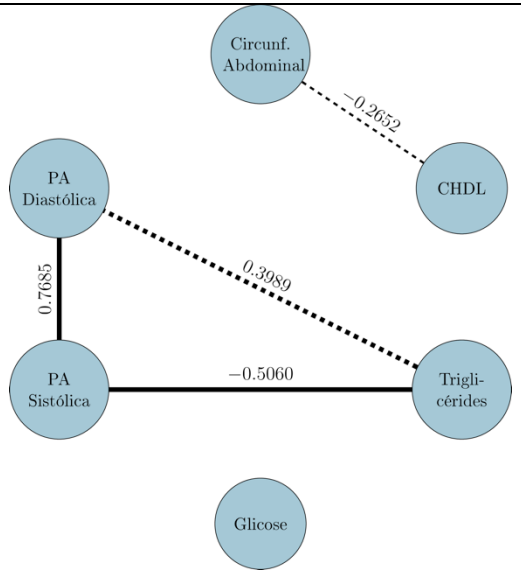
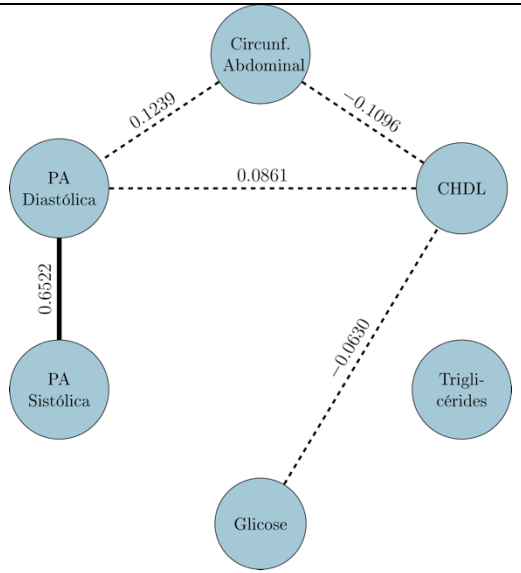
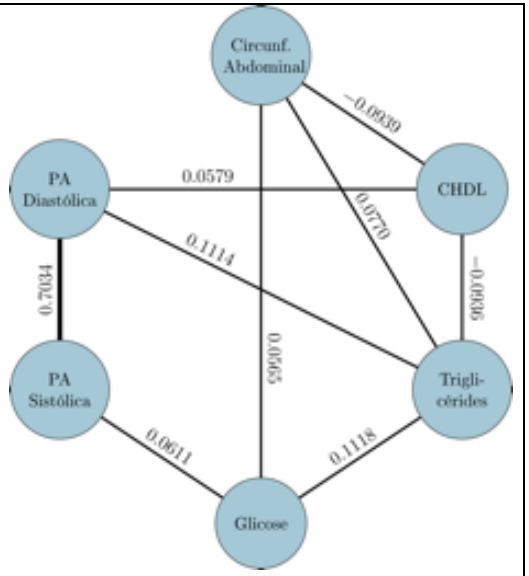
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

	2- Com covariáveis sexo, idade e IMC			

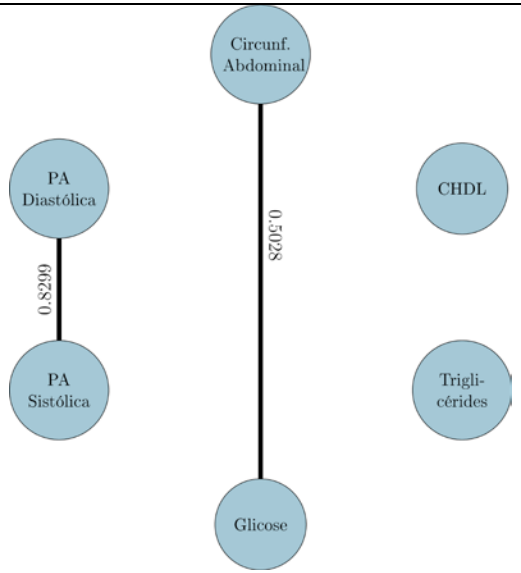
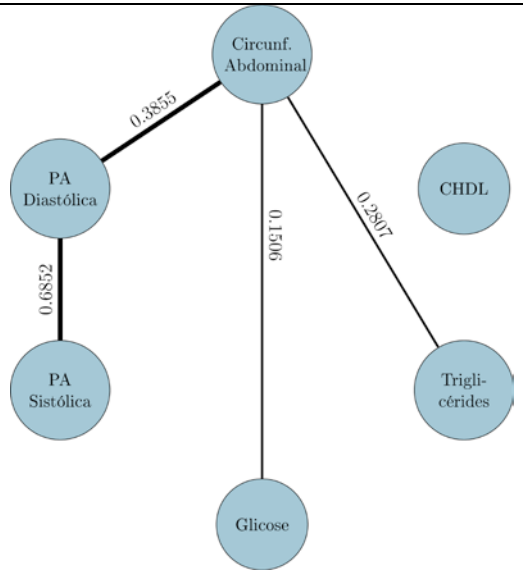
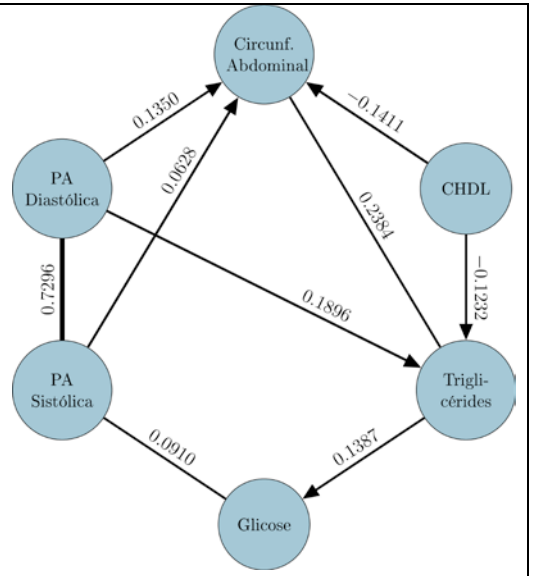
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

Univariado	2- Com covariáveis Sexo e Idade			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

	2- Com covariáveis sexo, idade e IMC			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

Algoritmo PC	2- Com covariáveis Sexo e Idade			

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

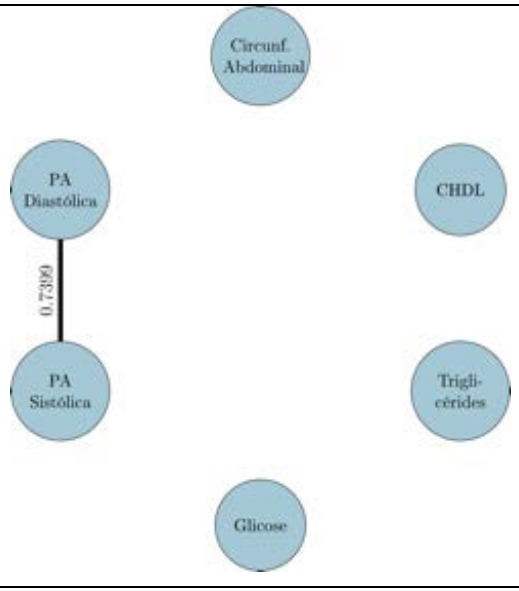
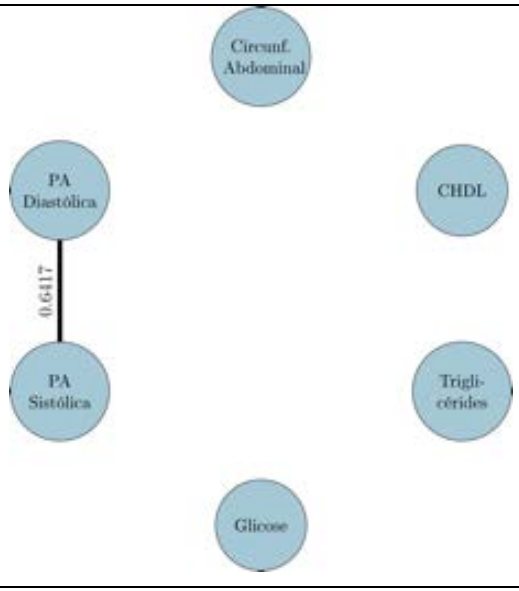
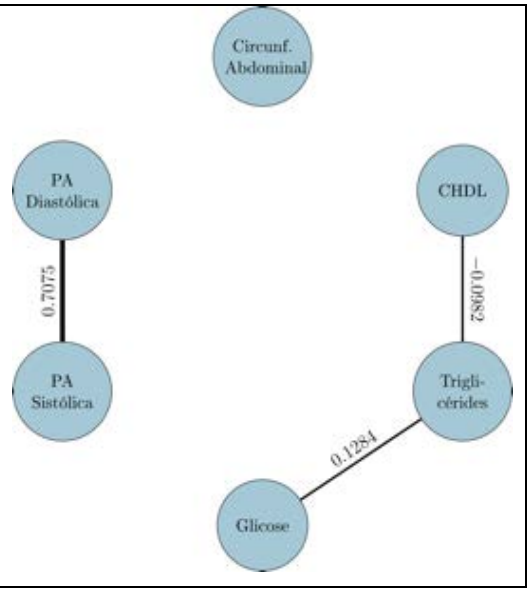
	2- Com covariáveis sexo, idade e IMC			

Figura 31: Grafos relativos às três abordagens (MLMP_Multivariado, MLMP_Univariado e Algortimo PC) para dois estudos feitos durante a Etapa 2 com a inclusão das covariáveis sexo e idade e sexo, idade e IMC.

Justificativa de não inclusão:

1- Com covariáveis Sexo e Idade: A razão pela não inclusão dos resultados obtidos com as covariáveis sexo e idade deve-se ao fato de se ter considerado mais relevante para a análise incluir os Componentes Principais de ancestralidade (PC1 e PC2) nesta análise, tornando a análise mais completa.

2- Com covariáveis Sexo, Idade e IMC: primeiramente foi efetuada a análise conforme acima, considerando as covariáveis sexo e idade, e em seguida adicionou-se a covariável IMC. Os resultados obtidos com a inclusão do IMC mostraram-se insatisfatórios quando comparados aos resultados anteriores, de modo que a covariável IMC acaba por eliminar arestas que

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

permanecem nos grafos das demais análises. Assim, considerou-se que IMC não é uma covariável adequada para este estudo.

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

APÊNDICE 4

p-valores

A) MLMP_MULTIVARIADO

A.1) SEM COVARIÁVEIS

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0038	0.6531	0.4894	0.2306	0.2084
GLICOSE	0.0038	NA	0.6103	0.7968	0.1333	0.1144
CHDL	0.6531	0.6103	NA	0.2080	0.6386	0.5041
TRIG	0.4894	0.7968	0.2080	NA	0.2244	0.1646
mediaPAS	0.2306	0.1333	0.6386	0.2244	NA	0.0000
mediaPAD	0.2084	0.1144	0.5041	0.1646	0.0000	NA

Tabela 25: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo de 20.67 para o MLMP_Multivariado, componente genético, sem covariáveis (Oualkacha, 2012).

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0286	0.6997	0.6874	0.3843	0.3843
GLICOSE	0.0286	1.0000	0.6997	0.7968	0.3843	0.3843
CHDL	0.6997	0.6997	1.0000	0.3843	0.6997	0.6874
TRIG	0.6874	0.7968	0.3843	1.0000	0.3843	0.3843
mediaPAS	0.3843	0.3843	0.6997	0.3843	1.0000	0.0000
mediaPAD	0.3843	0.3843	0.6874	0.3843	0.0000	1.0000

Tabela 26: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, sem covariáveis

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.0000	0.0000	0.0000
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0000
TRIG	0.0000	0.0000	0.0000	NA	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	NA	0.0000
mediaPAD	0.0000	0.0000	0.0000	0.0000	0.0000	NA

Tabela 27: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, sem covariáveis (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.0000	0.0000	0.0000
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
TRIG	0.0000	0.0000	0.0000	1.0000	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
mediaPAD	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

Tabela 28: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.0118	0.0000	0.0021
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0000
TRIG	0.0000	0.0118	0.0000	NA	0.0003	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0003	NA	0.0000
mediaPAD	0.0000	0.0021	0.0000	0.0000	0.0000	NA

Tabela 29: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, sem covariáveis (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.0118	0.0000	0.0023
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
TRIG	0.0000	0.0118	0.0000	1.0000	0.0004	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0004	1.0000	0.0000
mediaPAD	0.0000	0.0023	0.0000	0.0000	0.0000	1.0000

Tabela 30: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, sem covariáveis

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0024	0.0000	0.0000	0.0000	0.0003
GLICOSE	0.0024	NA	0.1624	0.0000	0.0000	0.3963
CHDL	0.0000	0.1624	NA	0.0003	0.3469	0.0037
TRIG	0.0000	0.0000	0.0003	NA	0.4589	0.0001
mediaPAS	0.0000	0.0000	0.3469	0.4589	NA	0.0000
mediaPAD	0.0003	0.3963	0.0037	0.0001	0.0000	NA

Tabela 31: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, sem covariáveis (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0037	0.0000	0.0000	0.0000	0.0005
GLICOSE	0.0037	1.0000	0.2030	0.0000	0.0000	0.4246
CHDL	0.0000	0.2030	1.0000	0.0005	0.4003	0.0050
TRIG	0.0000	0.0000	0.0005	1.0000	0.4589	0.0003
mediaPAS	0.0000	0.0000	0.4003	0.4589	1.0000	0.0000
mediaPAD	0.0005	0.4246	0.0050	0.0003	0.0000	1.0000

Tabela 32: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1480	0.0369	0.0000	0.0120	0.0849
GLICOSE	0.1480	NA	0.5049	0.0084	0.0279	0.6856
CHDL	0.0369	0.5049	NA	0.0857	0.6535	0.1653
TRIG	0.0000	0.0084	0.0857	NA	0.7237	0.0670
mediaPAS	0.0120	0.0279	0.6535	0.7237	NA	0.0000
mediaPAD	0.0849	0.6856	0.1653	0.0670	0.0000	NA

Tabela 33: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, sem covariáveis (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.2220	0.0922	0.0001	0.0451	0.1428
GLICOSE	0.2220	1.0000	0.6311	0.0418	0.0837	0.7237
CHDL	0.0922	0.6311	1.0000	0.1428	0.7237	0.2255
TRIG	0.0001	0.0418	0.1428	1.0000	0.7237	0.1428
mediaPAS	0.0451	0.0837	0.7237	0.7237	1.0000	0.0000
mediaPAD	0.1428	0.7237	0.2255	0.1428	0.0000	1.0000

Tabela 34: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, sem covariáveis

A.2) COM COVARIÁVEIS SEXO E IDADE

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	**0.0043**	0.5104	0.3405	0.2375	0.2692
GLICOSE	**0.0043**	NA	0.6047	0.9624	0.0862	0.0731
CHDL	0.5104	0.6047	NA	0.2272	0.7050	0.6411
TRIG	0.3405	0.9624	0.2272	NA	0.5720	0.4742
mediaPAS	0.2375	0.0862	0.7050	0.5720	NA	**0.0000**
mediaPAD	0.2692	0.0731	0.6411	0.4742	**0.0000**	NA

Tabela 35: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo e idade (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	**0.0324**	0.7397	0.6385	0.5769	0.5769
GLICOSE	**0.0324**	1.0000	0.7397	0.9624	0.3232	0.3232
CHDL	0.7397	0.7397	1.0000	0.5769	0.7553	0.7397
TRIG	0.6385	0.9624	0.5769	1.0000	0.7397	0.7397
mediaPAS	0.5769	0.3232	0.7553	0.7397	1.0000	**0.0000**
mediaPAD	0.5769	0.3232	0.7397	0.7397	**0.0000**	1.0000

Tabela 36: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo e idade

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.0038	0.0000	0.0011
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0000
TRIG	0.0000	0.0038	0.0000	NA	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	NA	0.0000
mediaPAD	0.0000	0.0011	0.0000	0.0000	0.0000	NA

Tabela 37: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.0038	0.0000	0.0012
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
TRIG	0.0000	0.0038	0.0000	1.0000	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
mediaPAD	0.0000	0.0012	0.0000	0.0000	0.0000	1.0000

Tabela 38: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0044	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.1667	0.0000	0.1187
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0000
TRIG	0.0044	0.1667	0.0000	NA	0.0004	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0004	NA	0.0000
mediaPAD	0.0000	0.1187	0.0000	0.0000	0.0000	NA

Tabela 39: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0051	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.1667	0.0000	0.1272
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0001
TRIG	0.0051	0.1667	0.0000	1.0000	0.0005	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0005	1.0000	0.0000
mediaPAD	0.0000	0.1272	0.0001	0.0000	0.0000	1.0000

Tabela 40: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo e idade

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1319	**0.0000**	**0.0000**	**0.0113**	**0.0000**
GLICOSE	0.1319	NA	0.1266	**0.0000**	**0.0190**	0.9861
CHDL	0.0000	0.1266	NA	**0.0002**	0.9621	**0.0170**
TRIG	0.0000	0.0000	0.0002	NA	0.2364	**0.0000**
mediaPAS	0.0113	0.0190	0.9621	0.2364	NA	**0.0000**
mediaPAD	0.0000	0.9861	0.0170	0.0000	0.0000	NA

Tabela 41: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo e idade (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.1648	**0.0000**	**0.0000**	**0.0212**	**0.0000**
GLICOSE	0.1648	1.0000	0.1648	**0.0000**	**0.0285**	0.9861
CHDL	0.0000	0.1648	1.0000	**0.0005**	0.9861	**0.0283**
TRIG	0.0000	0.0000	0.0005	1.0000	0.2728	**0.0000**
mediaPAS	0.0212	0.0285	0.9861	0.2728	1.0000	**0.0000**
mediaPAD	0.0000	0.9861	0.0283	0.0000	0.0000	1.0000

Tabela 42: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.4720	**0.0248**	**0.0004**	0.2265	**0.0362**
GLICOSE	0.4720	NA	0.4659	**0.0357**	0.2629	0.9934
CHDL	0.0248	0.4659	NA	0.0778	0.9819	0.2544
TRIG	0.0004	0.0357	0.0778	NA	0.5720	**0.0297**
mediaPAS	0.2265	0.2629	0.9819	0.5720	NA	**0.0000**
mediaPAD	0.0362	0.9934	0.2544	0.0297	0.0000	NA

Tabela 43: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo e idade (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.5900	0.0904	**0.0032**	0.3943	0.0904
GLICOSE	0.5900	1.0000	0.5900	0.0904	0.3943	0.9934
CHDL	0.0904	0.5900	1.0000	0.1667	0.9934	0.3943
TRIG	0.0032	0.0904	0.1667	1.0000	0.6600	0.0904
mediaPAS	0.3943	0.3943	0.9934	0.6600	1.0000	**0.0000**
mediaPAD	0.0904	0.9934	0.3943	0.0904	0.0000	1.0000

Tabela 44: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo e idade

A.3) COM COVARIÁVEIS SEXO, IDADE E IMC

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.4823	0.2772	0.1657	0.1283	0.4865
GLICOSE	0.4823	NA	0.3942	0.7069	0.6171	0.5872
CHDL	0.2772	0.3942	NA	0.1920	0.5665	0.6052
TRIG	0.1657	0.7069	0.1920	NA	0.4771	0.4760
mediaPAS	0.1283	0.6171	0.5665	0.4771	NA	0.0000
mediaPAD	0.4865	0.5872	0.6052	0.4760	0.0000	NA

Tabela 45: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=20.67 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade e IMC (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.6611	0.6611	0.6611	0.6611	0.6611
GLICOSE	0.6611	1.0000	0.6611	0.7069	0.6611	0.6611
CHDL	0.6611	0.6611	1.0000	0.6611	0.6611	0.6611
TRIG	0.6611	0.7069	0.6611	1.0000	0.6611	0.6611
mediaPAS	0.6611	0.6611	0.6611	0.6611	1.0000	0.0000
mediaPAD	0.6611	0.6611	0.6611	0.6611	0.0000	1.0000

Tabela 46: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade e IMC

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.0581	0.0000	0.6774
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0000
TRIG	0.0000	0.0581	0.0000	NA	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	NA	0.0000
mediaPAD	0.0000	0.6774	0.0000	0.0000	0.0000	NA

Tabela 47: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1587 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.0623	0.0000	0.6774
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
TRIG	0.0000	0.0623	0.0000	1.0000	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
mediaPAD	0.0000	0.6774	0.0000	0.0000	0.0000	1.0000

Tabela 48: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0000	0.0000	0.0035
GLICOSE	0.0000	NA	0.0000	0.3658	0.0000	0.8426
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0005
TRIG	0.0000	0.3658	0.0000	NA	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	NA	0.0000
mediaPAD	0.0035	0.8426	0.0005	0.0000	0.0000	NA

Tabela 49: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0000	0.0000	0.0040
GLICOSE	0.0000	1.0000	0.0000	0.3919	0.0000	0.8426
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0006
TRIG	0.0000	0.3919	0.0000	1.0000	0.0000	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0000	1.0000	0.0000
mediaPAD	0.0040	0.8426	0.0006	0.0000	0.0000	1.0000

Tabela 50: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade e IMC

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0247	0.0002	0.0022	0.7357	0.5993
GLICOSE	0.0247	NA	0.1407	0.0000	0.0149	0.9296
CHDL	0.0002	0.1407	NA	0.0002	0.9824	0.0212
TRIG	0.0022	0.0000	0.0002	NA	0.2057	0.0000
mediaPAS	0.7357	0.0149	0.9824	0.2057	NA	0.0000
mediaPAD	0.5993	0.9296	0.0212	0.0000	0.0000	NA

Tabela 51: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1587 para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC (Oualkacha, 2012).

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0412	0.0006	0.0055	0.8489	0.7492
GLICOSE	0.0412	1.0000	0.2110	0.0000	0.0319	0.9824
CHDL	0.0006	0.2110	1.0000	0.0006	0.9824	0.0398
TRIG	0.0055	0.0000	0.0006	1.0000	0.2805	0.0000
mediaPAS	0.8489	0.0319	0.9824	0.2805	1.0000	0.0000
mediaPAD	0.7492	0.9824	0.0398	0.0000	0.0000	1.0000

Tabela 52: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.2837	0.0736	0.1440	0.8720	0.8020
GLICOSE	0.2837	NA	0.4819	0.0334	0.2451	0.9664
CHDL	0.0736	0.4819	NA	0.0745	0.9916	0.2715
TRIG	0.1440	0.0334	0.0745	NA	0.5458	0.0337
mediaPAS	0.8720	0.2451	0.9916	0.5458	NA	0.0000
mediaPAD	0.8020	0.9664	0.2715	0.0337	0.0000	NA

Tabela 53: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=366.94 para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.4728	0.2234	0.3599	0.9916	0.9916
GLICOSE	0.4728	1.0000	0.7229	0.1687	0.4728	0.9916
CHDL	0.2234	0.7229	1.0000	0.2234	0.9916	0.4728
TRIG	0.3599	0.1687	0.2234	1.0000	0.7443	0.1687
mediaPAS	0.9916	0.4728	0.9916	0.7443	1.0000	0.0000
mediaPAD	0.9916	0.9916	0.4728	0.1687	0.0000	1.0000

Tabela 54: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade e IMC

A.4) COM COVARIÁVEIS SEXO, IDADE, PC1 E PC2

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1722	0.5731	0.7468	0.6596	0.7056
GLICOSE	0.1722	NA	0.8357	0.6027	0.0547	0.0643
CHDL	0.5731	0.8357	NA	0.3353	0.5257	0.5368
TRIG	0.7468	0.6027	0.3353	NA	0.4870	0.4233
mediaPAS	0.6596	0.0547	0.5257	0.4870	NA	0.0000
mediaPAD	0.7056	0.0643	0.5368	0.4233	0.0000	NA

Tabela 55: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=14.24 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.6456	0.8002	0.8002	0.8002	0.8002
GLICOSE	0.6456	1.0000	0.8357	0.8002	0.3215	0.3215
CHDL	0.8002	0.8357	1.0000	0.8002	0.8002	0.8002
TRIG	0.8002	0.8002	0.8002	1.0000	0.8002	0.8002
mediaPAS	0.8002	0.3215	0.8002	0.8002	1.0000	0.0000
mediaPAD	0.8002	0.3215	0.8002	0.8002	0.0000	1.0000

Tabela 56: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1 e PC2

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0877	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.3855	0.0000	0.1360
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0002
TRIG	0.0877	0.3855	0.0000	NA	0.0002	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0002	NA	0.0000
mediaPAD	0.0000	0.1360	0.0002	0.0000	0.0000	NA

Tabela 57: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=42.67 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.1012	0.0000	0.0001
GLICOSE	0.0000	1.0000	0.0000	0.3855	0.0000	0.1457
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0002
TRIG	0.1012	0.3855	0.0000	1.0000	0.0002	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0002	1.0000	0.0000
mediaPAD	0.0001	0.1457	0.0002	0.0000	0.0000	1.0000

Tabela 58: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0007	0.3387	0.0000	0.0206
GLICOSE	0.0000	NA	0.0014	0.6268	0.0000	0.4035
CHDL	0.0007	0.0014	NA	0.0000	0.0001	0.0335
TRIG	0.3387	0.6268	0.0000	NA	0.0340	0.0000
mediaPAS	0.0000	0.0000	0.0001	0.0340	NA	0.0000
mediaPAD	0.0206	0.4035	0.0335	0.0000	0.0000	NA

Tabela 59: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2 (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0014	0.3908	0.0000	0.0309
GLICOSE	0.0000	1.0000	0.0023	0.6268	0.0000	0.4323
CHDL	0.0014	0.0023	1.0000	0.0000	0.0002	0.0425
TRIG	0.3908	0.6268	0.0000	1.0000	0.0425	0.0001
mediaPAS	0.0000	0.0000	0.0002	0.0425	1.0000	0.0000
mediaPAD	0.0309	0.4323	0.0425	0.0001	0.0000	1.0000

Tabela 60: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7304	0.0000	0.0000	0.0898	0.0001
GLICOSE	0.7304	NA	0.0660	0.0016	0.1260	0.5346
CHDL	0.0000	0.0660	NA	0.0022	0.9244	0.0316
TRIG	0.0000	0.0016	0.0022	NA	0.8212	0.0017
mediaPAS	0.0898	0.1260	0.9244	0.8212	NA	0.0000
mediaPAD	0.0001	0.5346	0.0316	0.0017	0.0000	NA

Tabela 61: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=42.67 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2 (Oualkacha, 2012).

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.8427	0.0000	0.0000	0.1347	0.0004
GLICOSE	0.8427	1.0000	0.1099	0.0043	0.1718	0.6683
CHDL	0.0000	0.1099	1.0000	0.0046	0.9244	0.0592
TRIG	0.0000	0.0043	0.0046	1.0000	0.8799	0.0043
mediaPAS	0.1347	0.1718	0.9244	0.8799	1.0000	0.0000
mediaPAD	0.0004	0.6683	0.0592	0.0043	0.0000	1.0000

Tabela 62: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8469	0.0112	0.0080	0.3418	0.0287
GLICOSE	0.8469	NA	0.3029	0.0775	0.3912	0.7279
CHDL	0.0112	0.3029	NA	0.0855	0.9576	0.2283
TRIG	0.0080	0.0775	0.0855	NA	0.8993	0.0787
mediaPAS	0.3418	0.3912	0.9576	0.8993	NA	0.0000
mediaPAD	0.0287	0.7279	0.2283	0.0787	0.0000	NA

Tabela 63: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2 (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.9576	0.0558	0.0558	0.5128	0.1077
GLICOSE	0.9576	1.0000	0.5048	0.1831	0.5334	0.9098
CHDL	0.0558	0.5048	1.0000	0.1831	0.9576	0.4280
TRIG	0.0558	0.1831	0.1831	1.0000	0.9576	0.1831
mediaPAS	0.5128	0.5334	0.9576	0.9576	1.0000	0.0000
mediaPAD	0.1077	0.9098	0.4280	0.1831	0.0000	1.0000

Tabela 64: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1 e PC2

A.5) COM COVARIÁVEIS SEXO, IDADE, PC1, PC2 E SNPs (Baependi)

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1796	0.4961	0.6449	0.6758	0.6614
GLICOSE	0.1796	NA	0.8114	0.5821	0.0550	0.0570
CHDL	0.4961	0.8114	NA	0.2887	0.4935	0.5408
TRIG	0.6449	0.5821	0.2887	NA	0.3743	0.3529
mediaPAS	0.6758	0.0550	0.4935	0.3743	NA	0.0000
mediaPAD	0.6614	0.0570	0.5408	0.3529	0.0000	NA

Tabela 65: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=14.24 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.6734	0.7241	0.7241	0.7241	0.7241
GLICOSE	0.6734	1.0000	0.8114	0.7241	0.2849	0.2849
CHDL	0.7241	0.8114	1.0000	0.7241	0.7241	0.7241
TRIG	0.7241	0.7241	0.7241	1.0000	0.7241	0.7241
mediaPAS	0.7241	0.2849	0.7241	0.7241	1.0000	0.0000
mediaPAD	0.7241	0.2849	0.7241	0.7241	0.0000	1.0000

Tabela 66: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.3246	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.2104	0.0000	0.0765
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0001
TRIG	0.3246	0.2104	0.0000	NA	0.0061	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0061	NA	0.0000
mediaPAD	0.0000	0.0765	0.0001	0.0000	0.0000	NA

Tabela 67: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=42.67 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.3246	0.0000	0.0000
GLICOSE	0.0000	1.0000	0.0000	0.2254	0.0000	0.0882
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0001
TRIG	0.3246	0.2254	0.0000	1.0000	0.0076	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0076	1.0000	0.0000
mediaPAD	0.0000	0.0882	0.0001	0.0000	0.0000	1.0000

Tabela 68: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0014	0.5810	0.0000	0.0074
GLICOSE	0.0000	NA	0.0088	0.4828	0.0000	0.3209
CHDL	0.0014	0.0088	NA	0.0000	0.0003	0.0281
TRIG	0.5810	0.4828	0.0000	NA	0.1242	0.0000
mediaPAS	0.0000	0.0000	0.0003	0.1242	NA	0.0000
mediaPAD	0.0074	0.3209	0.0281	0.0000	0.0000	NA

Tabela 69: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0025	0.5810	0.0000	0.0124
GLICOSE	0.0000	1.0000	0.0132	0.5173	0.0000	0.3702
CHDL	0.0025	0.0132	1.0000	0.0000	0.0006	0.0383
TRIG	0.5810	0.5173	0.0000	1.0000	0.1552	0.0000
mediaPAS	0.0000	0.0000	0.0006	0.1552	1.0000	0.0000
mediaPAD	0.0124	0.3702	0.0383	0.0000	0.0000	1.0000

Tabela 70: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.6820	0.0000	0.0000	0.1117	0.0001
GLICOSE	0.6820	NA	0.0788	0.0039	0.0792	0.6871
CHDL	0.0000	0.0788	NA	0.0042	0.8323	0.0571
TRIG	0.0000	0.0039	0.0042	NA	0.8792	0.0016
mediaPAS	0.1117	0.0792	0.8323	0.8792	NA	0.0000
mediaPAD	0.0001	0.6871	0.0571	0.0016	0.0000	NA

Tabela 71: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=42.67 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012).

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.7928	0.0000	0.0000	0.1523	0.0004
GLICOSE	0.7928	1.0000	0.1188	0.0090	0.1188	0.7928
CHDL	0.0000	0.1188	1.0000	0.0090	0.8792	0.1071
TRIG	0.0000	0.0090	0.0090	1.0000	0.8792	0.0047
mediaPAS	0.1523	0.1188	0.8792	0.8792	1.0000	0.0000
mediaPAD	0.0004	0.7928	0.1071	0.0047	0.0000	1.0000

Tabela 72: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8184	0.0113	0.0075	0.3728	0.0290
GLICOSE	0.8184	NA	0.3247	0.1058	0.3253	0.8214
CHDL	0.0113	0.3247	NA	0.1087	0.9055	0.2864
TRIG	0.0075	0.1058	0.1087	NA	0.9321	0.0763
mediaPAS	0.3728	0.3253	0.9055	0.9321	NA	0.0000
mediaPAD	0.0290	0.8214	0.2864	0.0763	0.0000	NA

Tabela 73: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.9321	0.0563	0.0563	0.5083	0.1089
GLICOSE	0.9321	1.0000	0.4879	0.2330	0.4879	0.9321
CHDL	0.0563	0.4879	1.0000	0.2330	0.9321	0.4879
TRIG	0.0563	0.2330	0.2330	1.0000	0.9321	0.2288
mediaPAS	0.5083	0.4879	0.9321	0.9321	1.0000	0.0000
mediaPAD	0.1089	0.9321	0.4879	0.2288	0.0000	1.0000

Tabela 74: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs

A.6) COM COVARIÁVEIS SEXO, IDADE, PC1, PC2 E SNPs (Literatura)

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1776	0.5808	0.7622	0.6502	0.7043
GLICOSE	0.1776	NA	0.7958	0.5802	0.0483	0.0581
CHDL	0.5808	0.7958	NA	0.3553	0.4907	0.5043
TRIG	0.7622	0.5802	0.3553	NA	0.4750	0.4124
mediaPAS	0.6502	0.0483	0.4907	0.4750	NA	0.0000
mediaPAD	0.7043	0.0581	0.5043	0.4124	0.0000	NA

Tabela 75: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ng=14.24 para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.6661	0.7920	0.7958	0.7958	0.7958
GLICOSE	0.6661	1.0000	0.7958	0.7920	0.2904	0.2904
CHDL	0.7920	0.7958	1.0000	0.7920	0.7920	0.7920
TRIG	0.7958	0.7920	0.7920	1.0000	0.7920	0.7920
mediaPAS	0.7958	0.2904	0.7920	0.7920	1.0000	0.0000
mediaPAD	0.7958	0.2904	0.7920	0.7920	0.0000	1.0000

Tabela 76: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0000	0.0380	0.0000	0.0000
GLICOSE	0.0000	NA	0.0000	0.3028	0.0000	0.1729
CHDL	0.0000	0.0000	NA	0.0000	0.0000	0.0003
TRIG	0.0380	0.3028	0.0000	NA	0.0004	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0004	NA	0.0000
mediaPAD	0.0000	0.1729	0.0003	0.0000	0.0000	NA

Tabela 77: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo ne=1024 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0000	0.0438	0.0000	0.0001
GLICOSE	0.0000	1.0000	0.0000	0.3028	0.0000	0.1852
CHDL	0.0000	0.0000	1.0000	0.0000	0.0000	0.0004
TRIG	0.0438	0.3028	0.0000	1.0000	0.0005	0.0000
mediaPAS	0.0000	0.0000	0.0000	0.0005	1.0000	0.0000
mediaPAD	0.0001	0.1852	0.0004	0.0000	0.0000	1.0000

Tabela 78: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0000	0.0020	0.2450	0.0000	0.0199
GLICOSE	0.0000	NA	0.0031	0.5636	0.0000	0.4450
CHDL	0.0020	0.0031	NA	0.0000	0.0002	0.0413
TRIG	0.2450	0.5636	0.0000	NA	0.0468	0.0000
mediaPAS	0.0000	0.0000	0.0002	0.0468	NA	0.0000
mediaPAD	0.0199	0.4450	0.0413	0.0000	0.0000	NA

Tabela 79: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.0000	0.0038	0.2826	0.0000	0.0299
GLICOSE	0.0000	1.0000	0.0051	0.5636	0.0000	0.4768
CHDL	0.0038	0.0051	1.0000	0.0000	0.0003	0.0563
TRIG	0.2826	0.5636	0.0000	1.0000	0.0585	0.0001
mediaPAS	0.0000	0.0000	0.0003	0.0585	1.0000	0.0000
mediaPAD	0.0299	0.4768	0.0563	0.0001	0.0000	1.0000

Tabela 80: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7160	0.0000	0.0000	0.0734	0.0001
GLICOSE	0.7160	NA	0.0678	0.0020	0.1353	0.5125
CHDL	0.0000	0.0678	NA	0.0019	0.9278	0.0330
TRIG	0.0000	0.0020	0.0019	NA	0.7969	0.0016
mediaPAS	0.0734	0.1353	0.9278	0.7969	NA	0.0000
mediaPAD	0.0001	0.5125	0.0330	0.0016	0.0000	NA

Tabela 81: p-valores do teste de correlação parcial nula usando transformação z de Fisher e tamanho amostral efetivo nt=1024 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Oualkacha, 2012).

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.8261	0.0000	0.0000	0.1101	0.0004
GLICOSE	0.8261	1.0000	0.1101	0.0043	0.1845	0.6406
CHDL	0.0000	0.1101	1.0000	0.0043	0.9278	0.0619
TRIG	0.0000	0.0043	0.0043	1.0000	0.8538	0.0043
mediaPAS	0.1101	0.1845	0.9278	0.8538	1.0000	0.0000
mediaPAD	0.0004	0.6406	0.0619	0.0043	0.0000	1.0000

Tabela 82: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8385	0.0112	0.0082	0.3158	0.0304
GLICOSE	0.8385	NA	0.3062	0.0834	0.4027	0.7136
CHDL	0.0112	0.3062	NA	0.0810	0.9595	0.2322
TRIG	0.0082	0.0834	0.0810	NA	0.8853	0.0765
mediaPAS	0.3158	0.4027	0.9595	0.8853	NA	0.0000
mediaPAD	0.0304	0.7136	0.2322	0.0765	0.0000	NA

Tabela 83: p-valores do teste de correlação parcial usando transformação de Fisher e tamanho amostral efetivo (de Yang) ny=326.32 para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Yang, 2011)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.9485	0.0562	0.0562	0.4737	0.1141
GLICOSE	0.9485	1.0000	0.4737	0.1787	0.5491	0.8920
CHDL	0.0562	0.4737	1.0000	0.1787	0.9595	0.4353
TRIG	0.0562	0.1787	0.1787	1.0000	0.9485	0.1787
mediaPAS	0.4737	0.5491	0.9595	0.9485	1.0000	0.0000
mediaPAD	0.1141	0.8920	0.4353	0.1787	0.0000	1.0000

Tabela 84: p-valores do teste de correlação parcial nula usando correção FDR para o MLMP_Multivariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs

B) MLMP_Univariado

B.1) SEM COVARIÁVEIS

1) Componente genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0145	0.2951	0.0192	0.1962	0.4369
GLICOSE	0.0118	NA	0.2847	0.3727	0.4503	0.9083
CHDL	0.1370	0.3399	NA	0.7014	0.9943	0.4838
TRIG	0.0266	0.3798	0.6353	NA	0.0319	0.1057
mediaPAS	0.1896	0.4208	0.9756	0.0824	NA	0.0000
mediaPAD	0.6926	0.8067	0.6155	0.0845	0.0000	NA

Tabela 85: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1087	0.5902	0.1328	0.4528	0.8093
GLICOSE	0.1087	NA	0.6330	0.6330	0.6433	0.9732
CHDL	0.5902	0.6330	NA	0.8093	0.9943	0.7941
TRIG	0.1328	0.6330	0.8093	NA	0.2818	0.3172
mediaPAS	0.4528	0.6433	0.9943	0.2818	NA	0.0000
mediaPAD	0.8093	0.9732	0.7941	0.3172	0.0000	NA

Tabela 86: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, sem covariáveis

2) Componente ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0007	0.0233	0.0000	0.0434	0.0000
GLICOSE	0.0004	NA	0.0349	0.2797	0.0488	0.3918
CHDL	0.0106	0.3219	NA	0.7129	0.5622	0.1598
TRIG	0.0000	0.2797	0.3210	NA	0.1966	0.4732
mediaPAS	0.0154	0.0683	0.4809	0.1617	NA	0.0000
mediaPAD	0.0000	0.9298	0.0216	0.3820	0.0000	NA

Tabela 87: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0026	0.0582	0.0000	0.0931	0.0002
GLICOSE	0.0026	NA	0.4199	0.3996	0.1281	0.9299
CHDL	0.0582	0.4199	NA	0.7375	0.6023	0.2695
TRIG	0.0000	0.3996	0.7375	NA	0.3105	0.5343
mediaPAS	0.0931	0.1281	0.6023	0.3105	NA	0.0000
mediaPAD	0.0002	0.9299	0.2695	0.5343	0.0000	NA

Tabela 88: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, sem covariáveis

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0019	0.0000	0.0000	0.0000	0.0002
GLICOSE	0.0019	NA	0.1530	0.0000	0.0000	0.3825
CHDL	0.0000	0.1545	NA	0.0002	0.3264	0.0028
TRIG	0.0000	0.0000	0.0002	NA	0.4484	0.0001
mediaPAS	0.0000	0.0000	0.3335	0.4476	NA	0.0000
mediaPAD	0.0002	0.3817	0.0029	0.0001	0.0000	NA

Tabela 89: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0028	0.0000	0.0000	0.0000	0.0004
GLICOSE	0.0028	NA	0.1932	0.0000	0.0000	0.4098
CHDL	0.0000	0.1932	NA	0.0004	0.3848	0.0039
TRIG	0.0000	0.0000	0.0004	NA	0.4484	0.0002
mediaPAS	0.0000	0.0000	0.3848	0.4484	NA	0.0000
mediaPAD	0.0004	0.4098	0.0039	0.0002	0.0000	NA

Tabela 78: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 79: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado sem covariáveis

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	219	219	219	219	219
GLICOSE	270	NA	270	270	270	270
CHDL	159	159	NA	159	159	159
TRIG	270	270	270	NA	270	270
mediaPAS	300	300	300	300	NA	300
mediaPAD	219	219	219	219	219	NA

Tabela 80: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado sem covariáveis

B.2) COM COVARIÁVEIS SEXO E IDADE

1) Componente Genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0226	0.1770	0.0818	0.3700	0.7970
GLICOSE	0.0118	NA	0.1938	0.7682	0.4660	0.6590
CHDL	0.1760	0.1926	NA	0.5395	0.7550	0.8040
TRIG	0.0579	0.7835	0.5499	NA	0.0053	0.0071
mediaPAS	0.2666	0.6543	0.9688	0.0058	NA	0.0000
mediaPAD	0.3606	0.5999	0.7015	0.0105	0.0000	NA

Tabela 81: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0847	0.4154	0.2455	0.6530	0.8317
GLICOSE	0.0847	NA	0.4154	0.8317	0.8317	0.8317
CHDL	0.4154	0.4154	NA	0.8248	0.9688	0.8317
TRIG	0.2455	0.8317	0.8248	NA	0.0427	0.0507
mediaPAS	0.6530	0.8317	0.9688	0.0427	NA	0.0000
mediaPAD	0.8317	0.8317	0.8317	0.0507	0.0000	NA

Tabela 82: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo e idade

2) Componente Ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0681	**0.0376**	**0.0020**	0.3444	**0.0000**
GLICOSE	0.0589	NA	**0.0272**	0.6098	0.5985	0.7220
CHDL	**0.0237**	0.4400	NA	0.3732	0.9004	0.2060
TRIG	**0.0001**	0.5885	0.3770	NA	0.8581	0.3180
mediaPAS	0.3630	0.8290	0.7407	0.8239	NA	**0.0000**
mediaPAD	**0.0000**	0.7732	**0.0259**	0.2231	**0.0000**	NA

Tabela 83: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1703	0.1127	0.0098	0.5953	0.0000
GLICOSE	0.1703	NA	0.6599	0.7954	0.8877	0.8877
CHDL	0.1127	0.6599	NA	0.5953	0.9004	0.4763
TRIG	**0.0098**	0.7954	0.5953	NA	0.8877	0.5953
mediaPAS	0.5953	0.8877	0.9004	0.8877	NA	0.0000
mediaPAD	**0.0000**	0.8877	0.4763	0.5953	**0.0000**	NA

Tabela 84: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo e idade

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1210	0.0000	0.0000	0.0093	0.0000
GLICOSE	0.1220	NA	0.1180	0.0000	0.0161	0.9840
CHDL	0.0000	0.1190	NA	0.0002	0.9686	0.0142
TRIG	0.0000	0.0000	0.0002	NA	0.2234	0.0000
mediaPAS	0.0093	0.0161	0.9620	0.2240	NA	0.0000
mediaPAD	0.0000	0.9830	0.0142	0.0000	0.0000	NA

Tabela 85: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1520	0.0000	0.0000	0.0175	0.0000
GLICOSE	0.1520	NA	0.1520	0.0000	0.0242	0.9836
CHDL	0.0000	0.1520	NA	0.0003	0.9836	0.0237
TRIG	0.0000	0.0000	0.0003	NA	0.2584	0.0000
mediaPAS	0.0175	0.0242	0.9836	0.2584	NA	0.0000
mediaPAD	0.0000	0.9836	0.0237	0.0000	0.0000	NA

Tabela 86: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 87: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo e idade

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	219	219	219	219	219
GLICOSE	270	NA	270	270	270	270
CHDL	159	159	NA	159	159	159
TRIG	270	270	270	NA	270	270
mediaPAS	300	300	300	300	NA	300
mediaPAD	243	243	243	243	243	NA

Tabela 88: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo e idade

B.3) COM COVARIÁVEIS SEXO, IDADE E IMC

1) Componente Genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0549	0.1643	0.2166	0.2180	0.4917
GLICOSE	0.0737	NA	0.0867	0.7200	0.5515	0.9971
CHDL	0.0381	0.1457	NA	0.9472	0.7904	0.7180
TRIG	0.1623	0.6400	0.4519	NA	0.0051	0.0296
mediaPAS	0.2004	0.6010	0.6817	0.0041	NA	0.0000
mediaPAD	0.5011	0.6625	0.8319	0.0321	0.0000	NA

Tabela 89: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.2456	0.3793	0.4087	0.4087	0.7912
GLICOSE	0.2456	NA	0.3793	0.8307	0.8307	0.9971
CHDL	0.3793	0.3793	NA	0.9798	0.8782	0.8914
TRIG	0.4087	0.8307	0.9798	NA	0.0383	0.1603
mediaPAS	0.4087	0.8307	0.8782	0.0383	NA	0.0000
mediaPAD	0.7912	0.9971	0.8914	0.1603	0.0000	NA

Tabela 90: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade e IMC

2) Componente Ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.6108	0.0580	0.2867	0.1374	0.0320
GLICOSE	0.7565	NA	0.0327	0.6430	0.5273	0.6576
CHDL	0.0430	0.4303	NA	0.4269	0.9323	0.2806
TRIG	0.2846	0.6443	0.3866	NA	0.8192	0.2219
mediaPAS	0.1439	0.8769	0.7894	0.7310	NA	0.0000
mediaPAD	0.0286	0.6943	0.0473	0.2257	0.0000	NA

Tabela 91: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8729	0.2176	0.5735	0.4317	0.1962
GLICOSE	0.8729	NA	0.7171	0.8577	0.9072	0.8679
CHDL	0.2176	0.7171	NA	0.7171	0.9323	0.5735
TRIG	0.5735	0.8577	0.7171	NA	0.8777	0.5643
mediaPAS	0.4317	0.9072	0.9323	0.8777	NA	0.0000
mediaPAD	0.1962	0.8679	0.5735	0.5643	0.0000	NA

Tabela 92: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade e IMC

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0212	0.0001	0.0017	0.7254	0.5896
GLICOSE	0.0212	NA	0.1307	0.0000	0.0126	0.9280
CHDL	0.0001	0.1309	NA	0.0001	0.9773	0.0179
TRIG	0.0016	0.0000	0.0001	NA	0.1926	0.0000
mediaPAS	0.7268	0.0125	0.9812	0.1935	NA	0.0000
mediaPAD	0.5898	0.9281	0.0180	0.0000	0.0000	NA

Tabela 93: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0353	0.0004	0.0042	0.8387	0.7372
GLICOSE	0.0353	NA	0.1964	0.0000	0.0271	0.9812
CHDL	0.0004	0.1964	NA	0.0004	0.9812	0.0338
TRIG	0.0042	0.0000	0.0004	NA	0.2639	0.0000
mediaPAS	0.8387	0.0271	0.9812	0.2639	NA	0.0000
mediaPAD	0.7372	0.9812	0.0338	0.0000	0.0000	NA

Tabela 94: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade e IMC

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 95: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade e IMC.

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	300	300	300	300	300
GLICOSE	270	NA	270	270	270	270
CHDL	159	159	NA	159	159	159
TRIG	270	270	270	NA	270	270
mediaPAS	300	300	300	300	NA	300
mediaPAD	243	243	243	243	243	NA

Tabela 96: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade e IMC.

B.4) COM COVARIÁVEIS SEXO, IDADE, PC1 E PC2

1) Componente Genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.3116	0.1126	0.3873	0.0350	0.8480
GLICOSE	0.2291	NA	0.1510	0.5709	0.6086	0.5710
CHDL	0.0783	0.2953	NA	0.5945	0.4606	0.8583
TRIG	0.2528	0.8119	0.7176	NA	0.2223	0.0943
mediaPAS	0.0484	0.7207	0.4110	0.1887	NA	0.0000
mediaPAD	0.8397	0.5772	0.8987	0.1154	0.0000	NA

Tabela 97: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.6232	0.4329	0.7253	0.3628	0.8879
GLICOSE	0.6232	NA	0.6232	0.8879	0.8648	0.7938
CHDL	0.4329	0.6232	NA	0.8648	0.7676	0.8987
TRIG	0.7253	0.8879	0.8648	NA	0.5728	0.4329
mediaPAS	0.3628	0.8648	0.7676	0.5728	NA	0.0006
mediaPAD	0.8879	0.7938	0.8987	0.4329	0.0006	NA

Tabela 98: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1 e PC2

2) Componente Ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7212	0.0003	0.0062	0.4274	0.0001
GLICOSE	0.3521	NA	0.0284	0.9690	0.9773	0.7055
CHDL	0.0011	0.5871	NA	0.5435	0.4320	0.8365
TRIG	0.1482	0.2178	0.6228	NA	0.8469	0.3847
mediaPAS	0.0163	0.5945	0.7922	0.8741	NA	0.0000
mediaPAD	0.0000	0.5907	0.9803	0.2086	0.0000	NA

Tabela 99: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.9408	0.0053	0.4446	0.8100	0.0009
GLICOSE	0.9408	NA	0.8897	0.9803	0.9803	0.9408
CHDL	0.0053	0.8897	NA	0.8897	0.9712	0.9803
TRIG	0.4446	0.9803	0.8897	NA	0.9712	0.8100
mediaPAS	0.8100	0.9803	0.9712	0.9712	NA	0.0000
mediaPAD	0.0009	0.9408	0.9803	0.8100	0.0000	NA

Tabela 100: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1 e PC2

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7209	0.0000	0.0000	0.0781	0.0001
GLICOSE	0.7211	NA	0.0562	0.0011	0.1120	0.5191
CHDL	0.0000	0.0564	NA	0.0014	0.9235	0.0255
TRIG	0.0000	0.0011	0.0014	NA	0.8120	0.0011
mediaPAS	0.0780	0.1121	0.9229	0.8128	NA	0.0000
mediaPAD	0.0001	0.5192	0.0255	0.0011	0.0000	NA

Tabela 101: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8321	0.0000	0.0000	0.1171	0.0002
GLICOSE	0.8321	NA	0.0940	0.0028	0.1528	0.6490
CHDL	0.0000	0.0940	NA	0.0031	0.9235	0.0479
TRIG	0.0000	0.0028	0.0031	NA	0.8708	0.0028
mediaPAS	0.1171	0.1528	0.9235	0.8708	NA	0.0000
mediaPAD	0.0002	0.6490	0.0479	0.0028	0.0000	NA

Tabela 102: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 103: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1 e PC2

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	243	243	243	243	243
GLICOSE	197	NA	197	197	197	197
CHDL	104	104	NA	104	104	104
TRIG	104	104	104	NA	104	104
mediaPAS	129	129	129	129	NA	129
mediaPAD	159	159	159	159	159	NA

Tabela 104: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1 e PC2.

B.5) COM COVARIÁVEIS SEXO, IDADE, PC1, PC2 E SNPs (Baependi)

1) Componente Genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1763	0.2472	0.2767	0.0164	0.5218
GLICOSE	0.2972	NA	0.1694	0.7680	0.9518	0.5606
CHDL	0.0782	0.2035	NA	0.6458	0.4105	0.7417
TRIG	0.2644	0.8152	0.7151	NA	0.2415	0.1859
mediaPAS	0.0209	0.9631	0.3397	0.2746	NA	0.0000
mediaPAD	0.5120	0.6623	0.8510	0.1915	0.0000	NA

Tabela 105: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.5572	0.5535	0.5535	0.1569	0.7827
GLICOSE	0.5572	NA	0.5535	0.9057	0.9631	0.8638
CHDL	0.5535	0.5535	NA	0.8862	0.6842	0.9118
TRIG	0.5535	0.9057	0.8862	NA	0.5535	0.5535
mediaPAS	0.1569	0.9631	0.6842	0.5535	NA	0.0002
mediaPAD	0.7827	0.8638	0.9118	0.5535	0.0002	NA

Tabela 106: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

2) Componente Ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.1503	0.0001	0.0027	0.4976	0.0011
GLICOSE	0.5489	NA	0.0288	0.5775	0.8634	0.8226
CHDL	0.0013	0.5366	NA	0.6618	0.4234	0.9363
TRIG	0.1728	0.1879	0.3873	NA	0.7633	0.5044
mediaPAS	0.0193	0.9988	0.2380	0.9528	NA	0.0000
mediaPAD	0.0000	0.5103	0.7233	0.1200	0.0000	NA

Tabela 107: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7841	0.0066	0.4321	0.7841	0.0066
GLICOSE	0.7841	NA	0.7841	0.7875	0.9988	0.9492
CHDL	0.0066	0.7841	NA	0.8633	0.7841	0.9856
TRIG	0.4321	0.7875	0.8633	NA	0.9856	0.7841
mediaPAS	0.7841	0.9988	0.7841	0.9856	NA	0.0000
mediaPAD	0.0066	0.9492	0.9856	0.7841	0.0000	NA

Tabela 108: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.6712	0.0000	0.0000	0.0812	0.0001
GLICOSE	0.7188	NA	0.0627	0.0018	0.0720	0.6316
CHDL	0.0000	0.0683	NA	0.0021	0.8964	0.0386
TRIG	0.0000	0.0022	0.0023	NA	0.7888	0.0008
mediaPAS	0.1037	0.1133	0.9050	0.9188	NA	0.0000
mediaPAD	0.0001	0.5557	0.0316	0.0012	0.0000	NA

Tabela 109: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8294	0.0000	0.0000	0.1482	0.0002
GLICOSE	0.8294	NA	0.1138	0.0050	0.1545	0.7895
CHDL	0.0000	0.1138	NA	0.0050	0.9188	0.0723
TRIG	0.0000	0.0050	0.0050	NA	0.9188	0.0035
mediaPAS	0.1482	0.1545	0.9188	0.9188	NA	0.0000
mediaPAD	0.0002	0.7895	0.0723	0.0035	0.0000	NA

Tabela 110: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 111: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	243	243	243	243	243
GLICOSE	219	NA	219	219	219	219
CHDL	104	104	NA	104	104	104
TRIG	104	104	104	NA	104	104
mediaPAS	116	116	116	116	NA	116
mediaPAD	177	177	177	177	177	NA

Tabela 112: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Baependi)

B.6) COM COVARIÁVEIS SEXO, IDADE, PC1, PC2 E SNPs (Literatura)

1) Componente Genético

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.2200	0.1563	0.4746	0.0433	0.9119
GLICOSE	0.1922	NA	0.1298	0.8643	0.4772	0.6778
CHDL	0.1108	0.2305	NA	0.7671	0.5550	0.9130
TRIG	0.2283	0.8624	0.6972	NA	0.1971	0.1442
mediaPAS	0.0067	0.8837	0.5403	0.2030	NA	0.0000
mediaPAD	0.9035	0.5024	0.7642	0.1110	0.0000	NA

Tabela 113: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.4609	0.4609	0.8324	0.3248	0.9130
GLICOSE	0.4609	NA	0.4609	0.9130	0.9130	0.9130
CHDL	0.4609	0.4609	NA	0.9130	0.8324	0.9130
TRIG	0.8324	0.9130	0.9130	NA	0.4609	0.4609
mediaPAS	0.3248	0.9130	0.8324	0.4609	NA	0.0004
mediaPAD	0.9130	0.9130	0.9130	0.4609	0.0004	NA

Tabela 114: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente genético, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

2) Componente Ambiental

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.4958	0.0010	0.0055	0.3089	0.0002
GLICOSE	0.3864	NA	0.0348	0.9410	0.8904	0.8449
CHDL	0.0017	0.4662	NA	0.5765	0.4120	0.9607
TRIG	0.1476	0.2440	0.5860	NA	0.7750	0.3860
mediaPAS	0.0147	0.6896	0.7032	0.7929	NA	0.0000
mediaPAD	0.0000	0.4660	0.8532	0.2218	0.0000	NA

Tabela 115: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7828	0.0086	0.4428	0.7128	0.0016
GLICOSE	0.7828	NA	0.7770	0.9607	0.9540	0.9480
CHDL	0.0086	0.7770	NA	0.8371	0.9172	0.9607
TRIG	0.4428	0.9607	0.8371	NA	0.9480	0.7726
mediaPAS	0.7128	0.9540	0.9172	0.9480	NA	0.0000
mediaPAD	0.0016	0.9480	0.9607	0.7726	0.0000	NA

Tabela 116: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, componente ambiental, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

3) Total

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.7333	0.0000	0.0000	0.0763	0.0000
GLICOSE	0.6921	NA	0.0597	0.0013	0.1093	0.5236
CHDL	0.0000	0.0549	NA	0.0014	0.9287	0.0270
TRIG	0.0000	0.0011	0.0013	NA	0.8226	0.0012
mediaPAS	0.0618	0.1268	0.9196	0.7676	NA	0.0000
mediaPAD	0.0001	0.5174	0.0254	0.0011	0.0000	NA

Tabela 117: p-valores do teste de correlação parcial sem correção para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.8461	0.0000	0.0000	0.1145	0.0002
GLICOSE	0.8461	NA	0.0976	0.0030	0.1729	0.6546
CHDL	0.0000	0.0976	NA	0.0030	0.9287	0.0506
TRIG	0.0000	0.0030	0.0030	NA	0.8813	0.0030
mediaPAS	0.1145	0.1729	0.9287	0.8813	NA	0.0000
mediaPAD	0.0002	0.6546	0.0506	0.0030	0.0000	NA

Tabela 118: p-valores do teste de correlação parcial nula usando correção FDR para MLMP_Univariado, total, com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	29	29	29	29	29
GLICOSE	29	NA	29	29	29	29
CHDL	29	29	NA	29	29	29
TRIG	29	29	29	NA	29	29
mediaPAS	29	29	29	29	NA	29
mediaPAD	29	29	29	29	29	NA

Tabela 119: tamanhos amostrais efetivos para o grafo genético do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	243	243	243	243	243
GLICOSE	197	NA	197	197	197	197
CHDL	104	104	NA	104	104	104
TRIG	104	104	104	NA	104	104
mediaPAS	129	129	129	129	NA	129
mediaPAD	159	159	159	159	159	NA

Tabela 120: tamanhos amostrais efetivos para o grafo ambiental do MLMP_Univariado com covariáveis sexo, idade, PC1, PC2 e SNPs (Literatura)

APÊNDICE 5

RESULTADOS IMC

Abaixo seguem os resultados obtidos para as análises incluindo o IMC como covariável:

MLMP_Multivariado:

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	0.1878	-0.2858	0.3583	-0.3897	0.1861
GLICOSE	0.1878	1.0000	0.2265	-0.1014	-0.1344	0.1458
CHDL	-0.2858	0.2265	1.0000	0.3390	-0.1538	0.1389
TRIG	0.3583	-0.1014	0.3390	1.0000	0.1900	-0.1904
mediaPAS	-0.3897	-0.1344	-0.1538	0.1900	1.0000	0.8882
mediaPAD	0.1861	0.1458	0.1389	-0.1904	0.8882	1.0000

Tabela 121: Matriz de Correlações Parciais para o componente poligênico.

	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	1.0000	-0.2505	-0.3388	-0.3435	0.5241	0.1528
GLICOSE	-0.2505	1.0000	-0.2928	-0.0476	0.3641	0.0105
CHDL	-0.3388	-0.2928	1.0000	-0.4862	0.3086	0.1810
TRIG	-0.3435	-0.0476	-0.4862	1.0000	0.2838	0.3087
mediaPAS	0.5241	0.3641	0.3086	0.2838	1.0000	0.3125
mediaPAD	0.1528	0.0105	0.1810	0.3087	0.3125	1.0000

Tabela 122: Matriz de Correlações Parciais para o componente ambiental.

	Circ.abd	GLICOSE	CHDL	TRIG	mediaPAS	mediaPAD
Circ.abd	1.0000	0.0564	-0.0940	0.0769	-0.0085	0.0132
GLICOSE	0.0564	1.0000	-0.0370	0.1117	0.0612	0.0022
CHDL	-0.0940	-0.0370	1.0000	-0.0937	-0.0006	0.0579
TRIG	0.0769	0.1117	-0.0937	1.0000	-0.0318	0.1114
mediaPAS	-0.0085	0.0612	-0.0006	-0.0318	1.0000	0.7034
mediaPAD	0.0132	0.0022	0.0579	0.1114	0.7034	1.0000

Tabela 123: Matriz de Correlações Parciais Total.

MLMP_Univariado:

	Circ.abd	GLICOSE	CHDL	TRIG	mediaPAS	mediaPAD
Circ.abd	NA	0.3603	-0.2652	0.2366	-0.2359	0.1330
GLICOSE	0.3372	NA	0.3237	0.0696	0.1153	-0.0007
CHDL	-0.3869	0.2770	NA	0.0129	0.0516	0.0701
TRIG	0.2665	0.0907	0.1453	NA	-0.5060	0.4042
mediaPAS	-0.2449	0.1013	0.0795	-0.5172	NA	0.7685
mediaPAD	0.1301	0.0846	0.0412	0.3989	0.7709	NA

Tabela 124: Matriz de Correlações Parciais para o componente poligênico.

	Circ.abd	GLICOSE	CHDL	TRIG	mediaPAS	mediaPAD
Circ.abd	NA	0.0295	-0.1096	0.0617	-0.0860	0.1239
GLICOSE	0.0190	NA	-0.1300	0.0283	0.0386	-0.0271
CHDL	-0.1607	-0.0630	NA	0.0634	0.0068	0.0861
TRIG	0.0654	0.0282	-0.0529	NA	-0.0140	0.0746
mediaPAS	-0.0846	0.0090	-0.0155	-0.0199	NA	0.6522
mediaPAD	0.1404	-0.0253	0.1274	0.0780	0.6539	NA

Tabela 125: Matriz de Correlações Parciais para o componente ambiental.

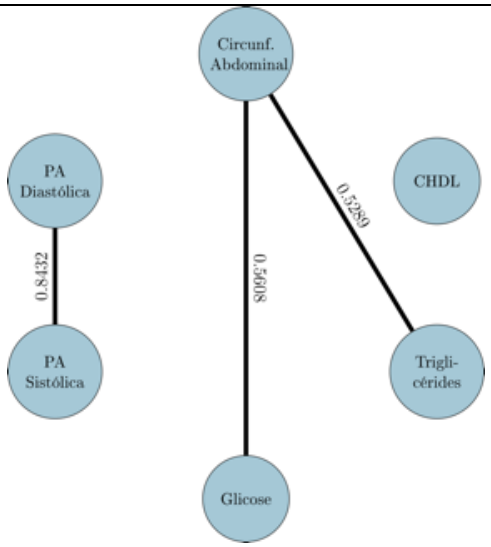
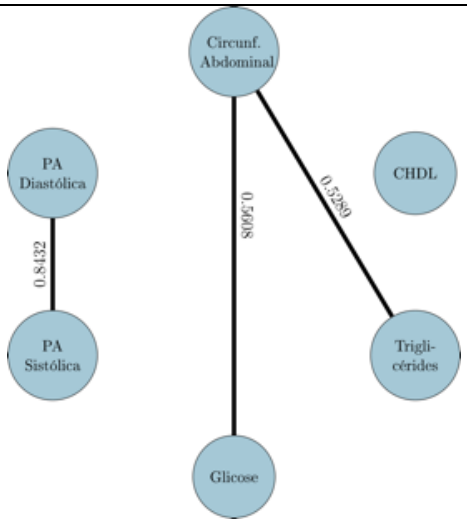
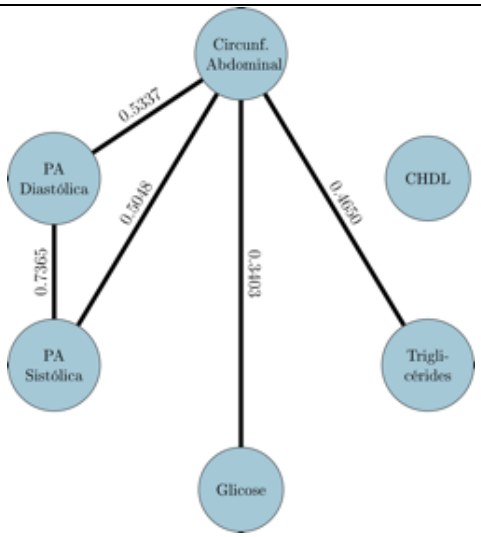
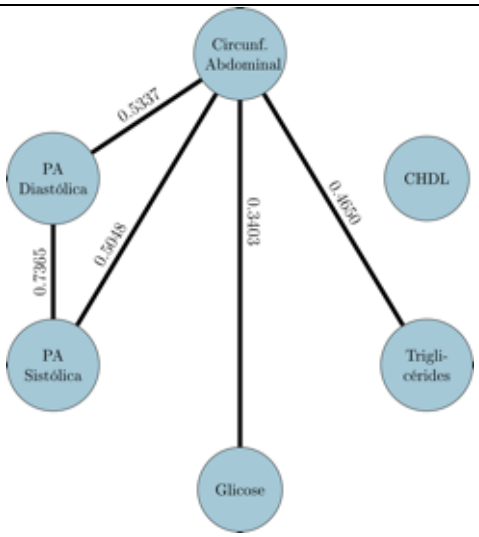
	Circ.abd	*GLICOSE*	*CHDL*	*TRIG*	*mediaPAS*	*mediaPAD*
Circ.abd	NA	0.0565	-0.0940	0.0770	-0.0086	0.0132
GLICOSE	0.0565	NA	-0.0370	0.1118	0.0611	0.0022
CHDL	-0.0939	-0.0370	NA	-0.0936	-0.0007	0.0580
TRIG	0.0771	0.1118	-0.0937	NA	-0.0319	0.1114
mediaPAS	-0.0086	0.0611	-0.0006	-0.0319	NA	0.7034
mediaPAD	0.0132	0.0022	0.0579	0.1114	0.7035	NA

Tabela 126: Matriz de Correlações Parciais Total.

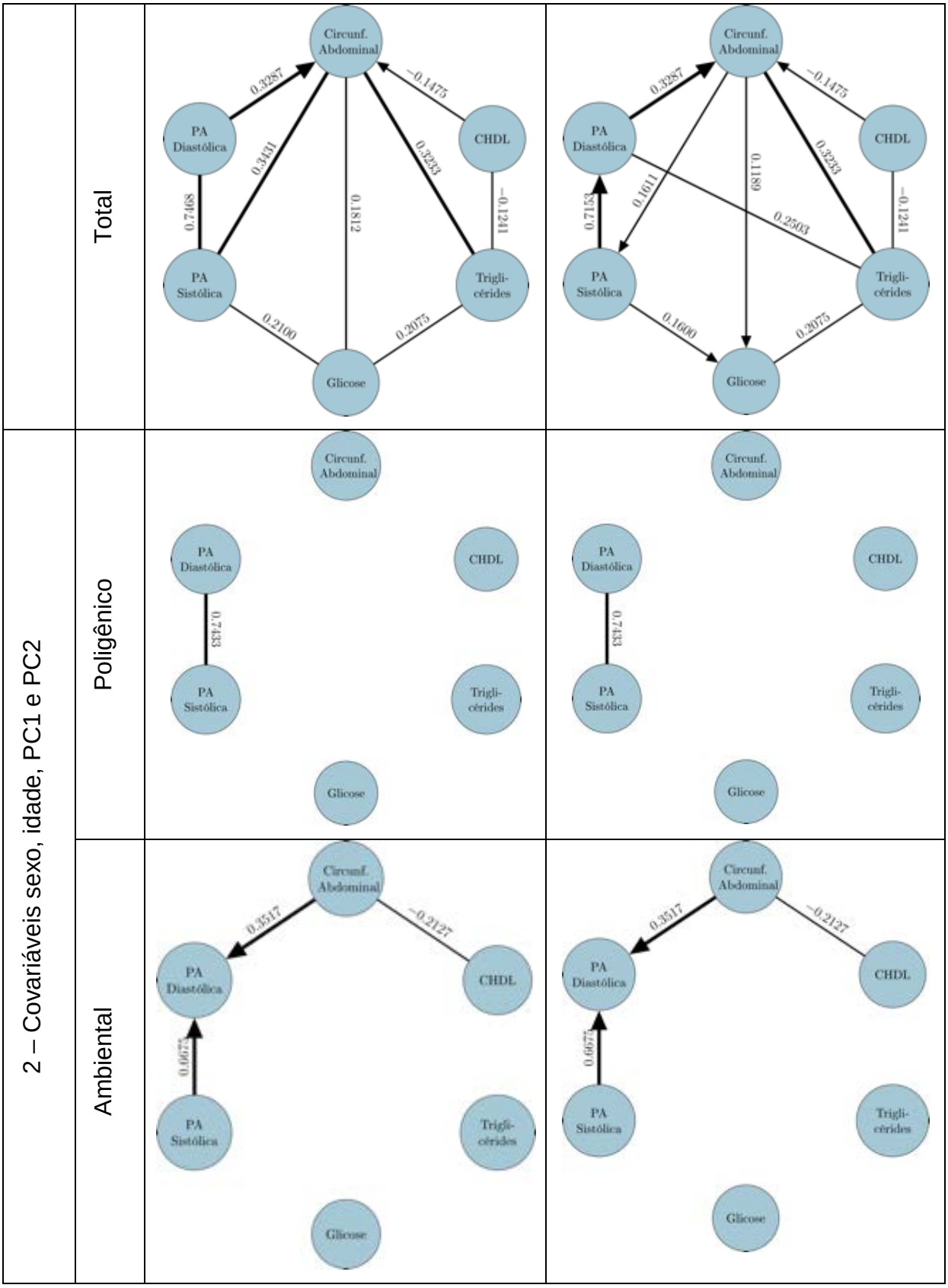
Obs.: Os p-valores encontram-se no **Apêndice 4 – p-valores** e os grafos relativos às três abordagens, Multivariada, Univariada e Algoritmo PC, se encontram no **Apêndice 3 – Grafos**

APÊNDICE 6

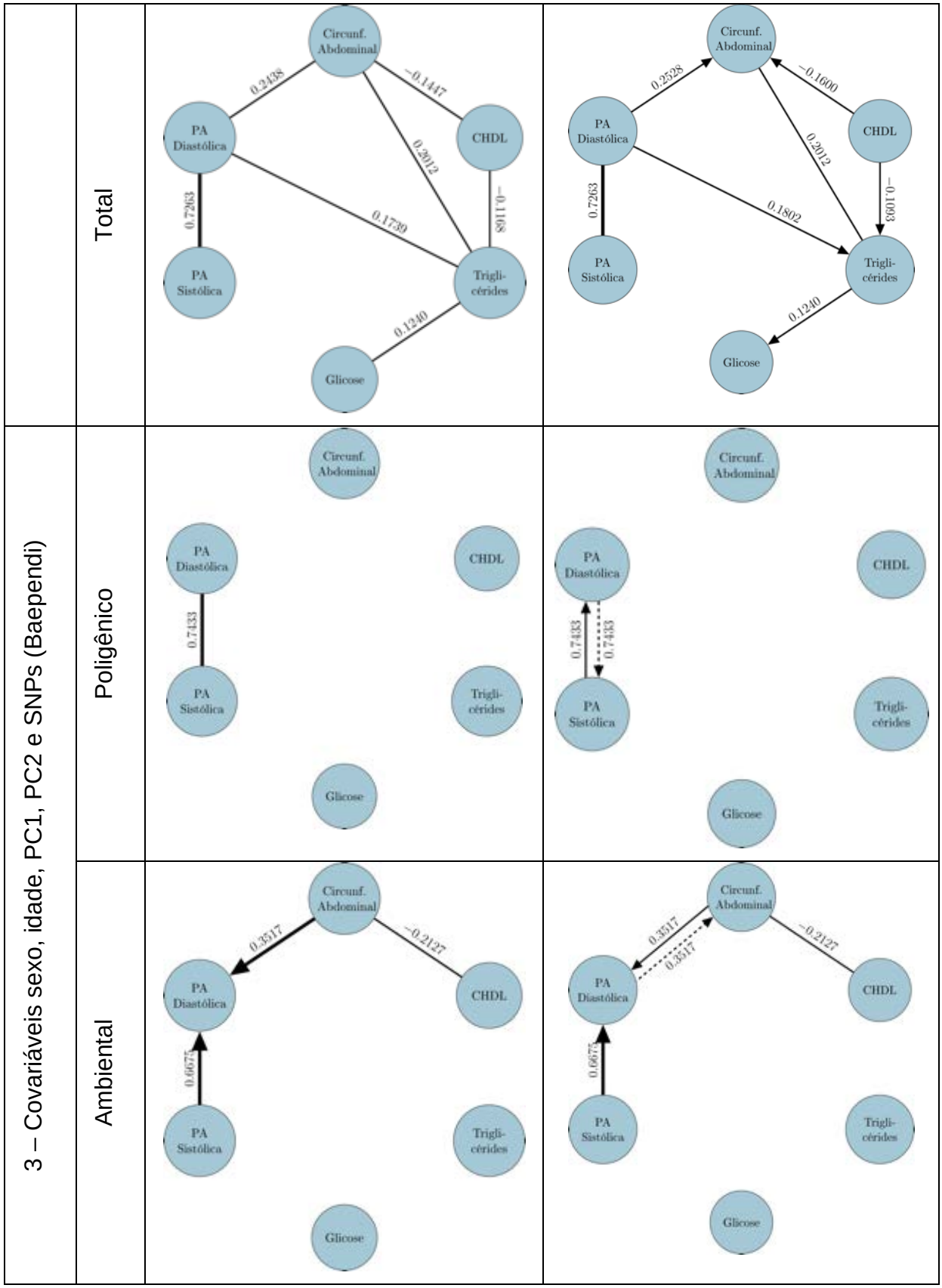
MAJOR RULE VS CONSERVATIVE RULE

Etapa	Comp	Conservative	Major Rule
1 – Sem covariáveis	Poligênico		
	Ambiental		

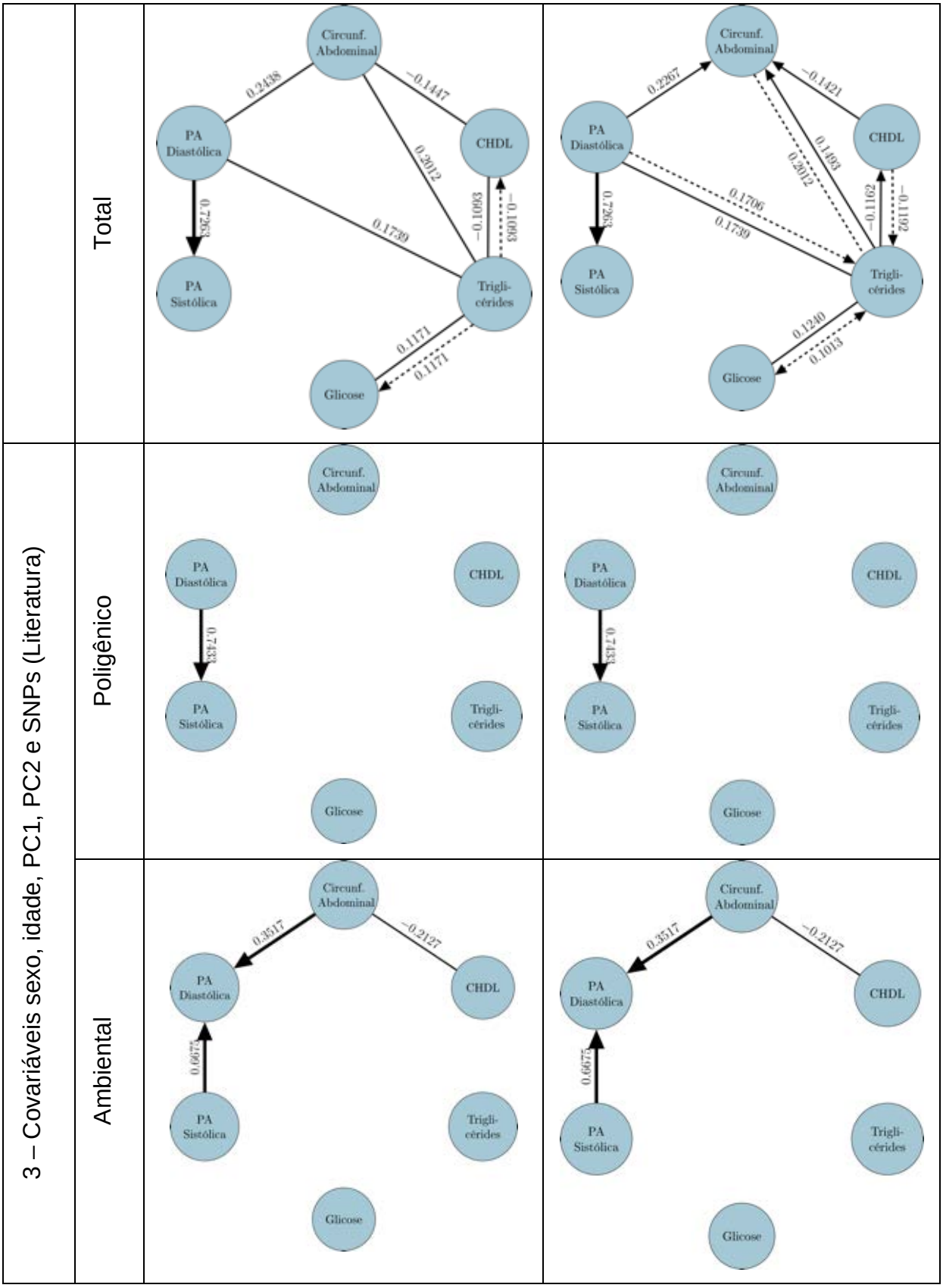
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua



Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

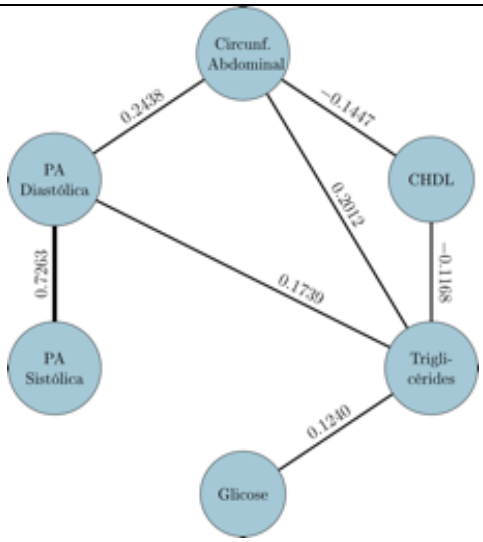
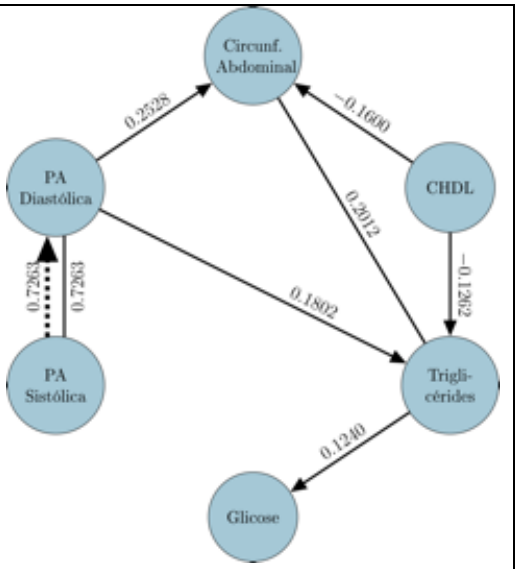
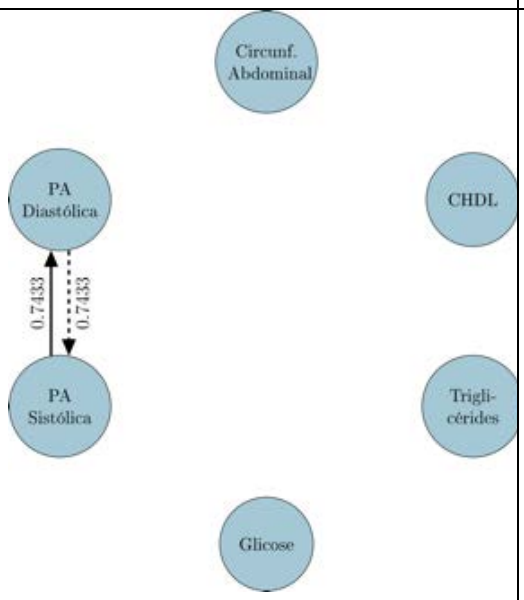
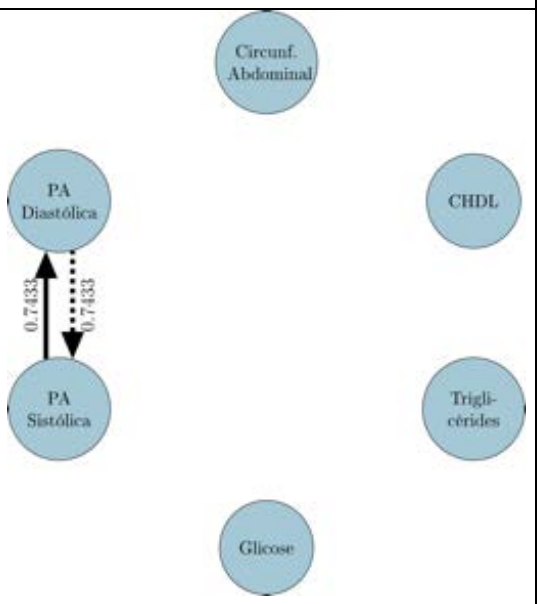
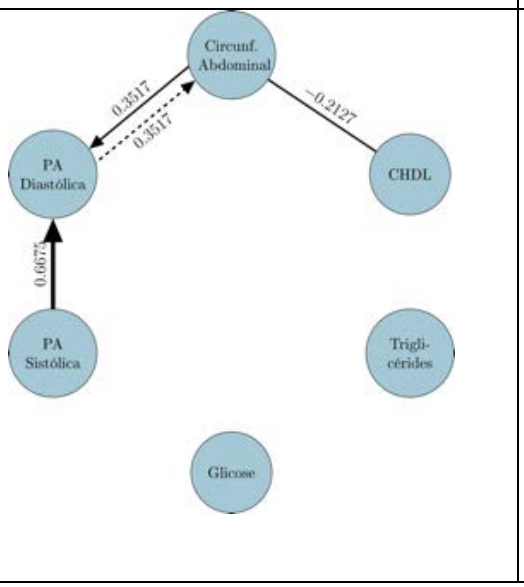
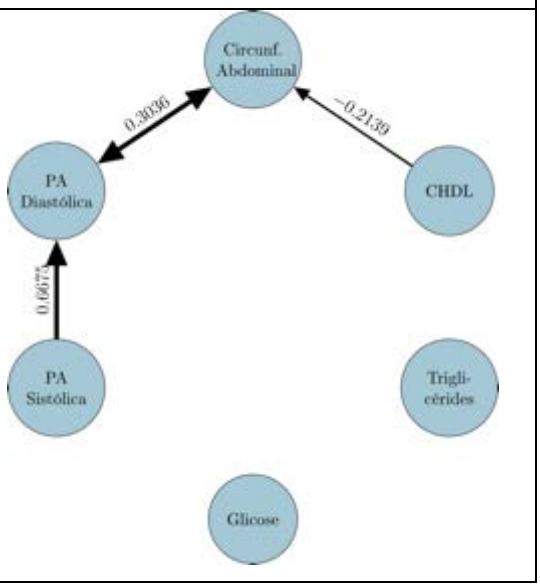
	Total		

Figura 32: Grafos comparativos entre as execuções do Algoritmo PC utilizando modo *Conservative* e modo *Major Rule*

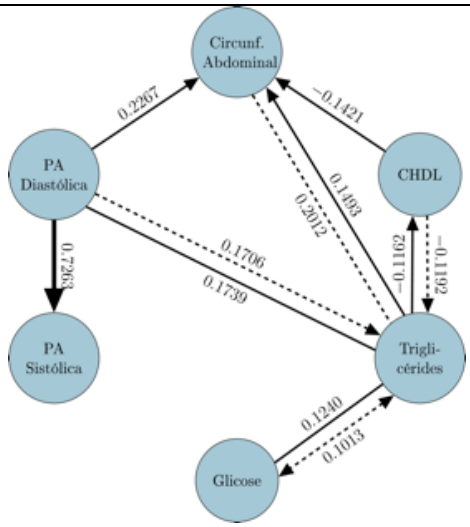
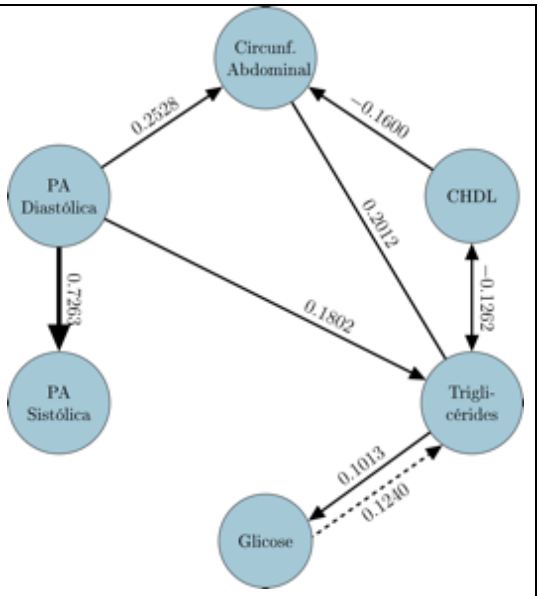
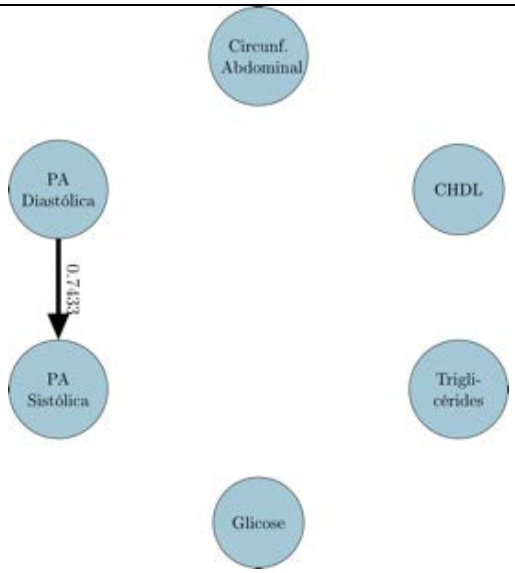
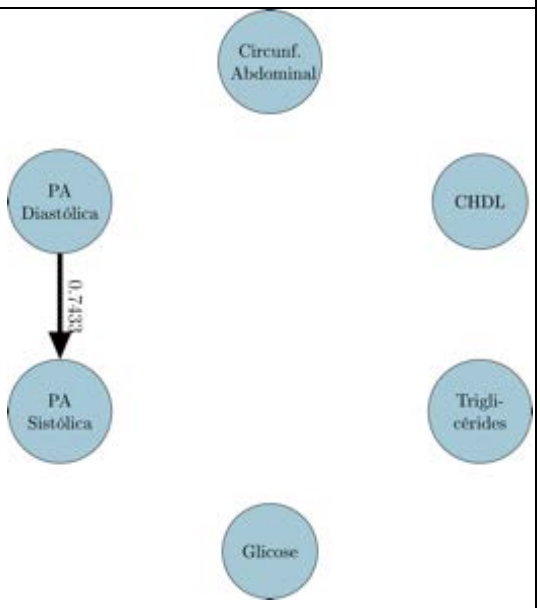
Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

APÊNDICE 7

COMPARAÇÃO ENTRE OS RESULTADOS COM 1 SNP E MAIS DE 1 SNP POR FENÓTIPO

	Comp.	1 SNP	> 1 SNP
Baependi	Poligênico		
	Ambiental		

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

	Total		
Literatura	Poligênico		

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua

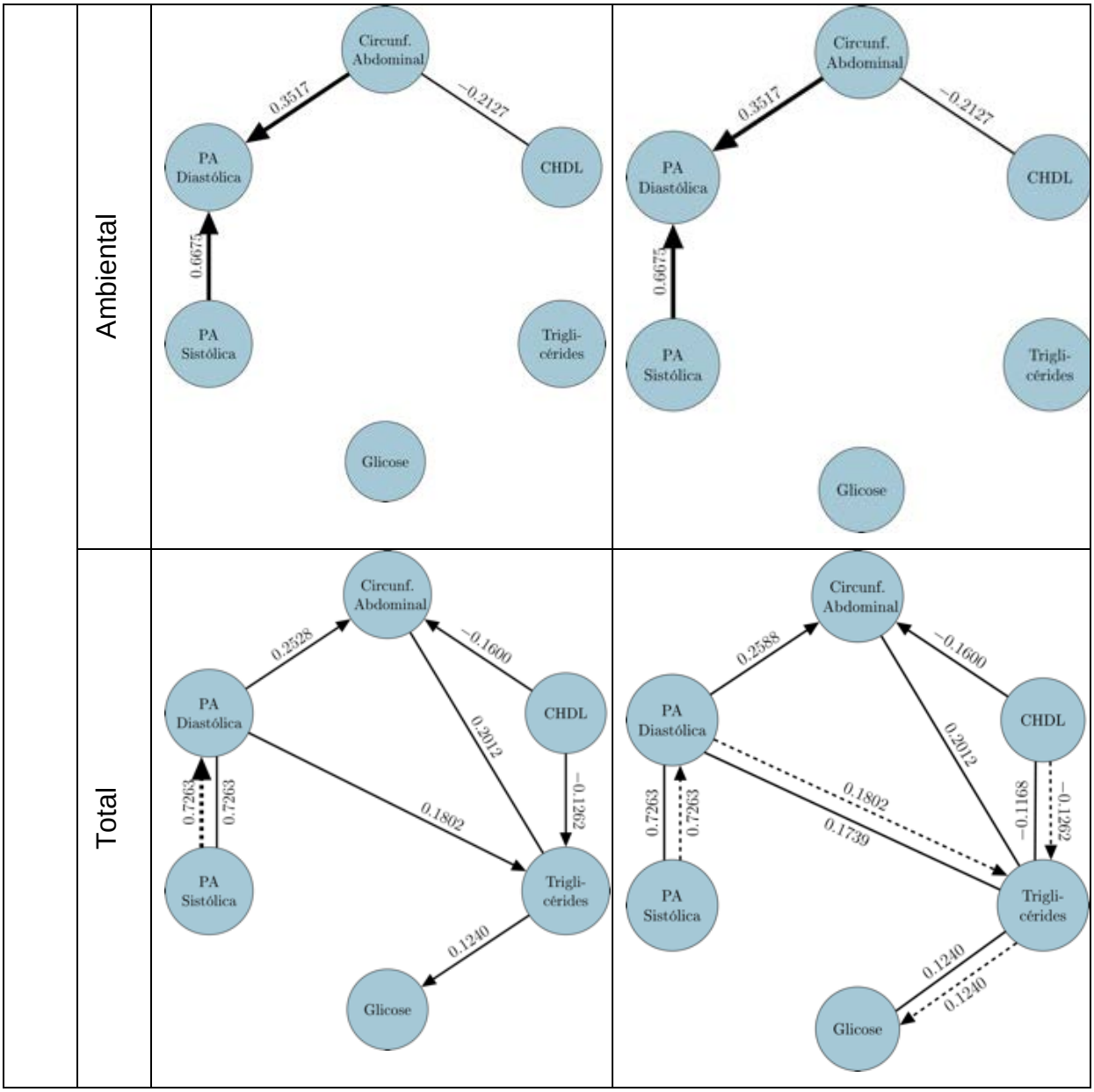


Figura 33: Grafos comparativos utilizando 1 SNP por fenótipo e mais de 1 SNP por fenótipo através da execução do Algoritmo PC utilizando o modo *Major Rule*

Linhas mais grossas nos grafos representam significâncias maiores que 0,30.
*1 – α = 5%, representado no grafo por linha tracejada
*2 – corrigido por FDR, representado no grafo por linha contínua