

Determinantes genéticos de doença arterial coronária em uma amostra da população brasileira

A importância da doença arterial coronariana (DAC) na sociedade "moderna" é atestada pelo grande número de pessoas afetadas por ela. Os dados da Organização Mundial de Saúde estimam que a doença isquêmica do coração será a segunda principal causa de morte em todo o mundo até 2030. Além disso, com o advento das ferramentas de predição de riscos e estudos de associação genômica ampla (GWAS, do inglês Genome-Wide Association Studies), vários estudos tem demonstrado o uso de escores genéticos de risco para prever DAC, porém estes escores apresentam apenas uma discreta melhora na capacidade de avaliação.

O presente estudo desenvolveu um escore genético de risco (GRS, do inglês Genetic risk score) para DAC, utilizando polimorfismos previamente associados a DAC e/ou a fenótipos relacionados à doença. Mil, trezentos e quarenta e nove indivíduos, provenientes do Estudo Longitudinal da Saúde do Adulto ELSA-BRASIL (Aquinio et al. 2012), foram genotipados para nosso estudo. Usando polimorfismos de nucleotídeo único (SNP, do inglês Single Nucleotide Polymorphisms) obtidos a partir da base de dados online de GWAS avaliamos a associação de cada polimorfismo escolhido com os fatores de risco para doença cardiovascular.

Modelos Zero-inflacionados foram utilizados para a construção do escore de risco genético (GRS). Escore de risco de Framingham para DAC em 10 anos (FRS, do inglês Framingham Risk Score) foi calculado de acordo com o descrito por Wilson e D'Agostino. Também foram criados outros dois escores: o primeiro, um modelo com apenas a idade e sexo; o segundo, apenas utilizando a informação genética. Comparamos o modelo Idade+Sexo com o GRS e FRS por meio de curva ROC e avaliamos o desempenho dos modelos em comparação com FRS.

INTRODUÇÃO

A doença arterial coronária (DAC), dentre as doenças cardiovasculares, é a principal causa de mortalidade e a que mais consome recursos na área de saúde nos países industrializados (Murray et al. 2012). No Brasil, segundo o DATASUS, foram cerca de 250 mil internações em 2014, o que resultou num custo direto de mais de um bilhão de reais.

A etiologia da DAC inclui fatores genéticos e ambientais e a interação entre ambos. A detecção precoce e estratificação de indivíduos de alto risco são de grande importância e podem direcionar as decisões sobre a prevenção de novos eventos e tratamento (Chen et al. 2011). O cálcio arterial coronariano (CAC) é um componente da aterosclerose, detectado, quase que exclusivamente, em artérias ateroscleróticas (Youssef et al. 2013), assim em estudos observacionais de grande escala, o escore de cálcio adiciona valor de prognóstico à predição de morte cardíaca e infarto do miocárdio, especialmente em pacientes de risco intermediário para eventos (Labree et al. 2004; Elias-Smale et al. 2010; Polonsky et al. 2010). A detecção do CAC tornou-se um método estabelecido para avaliação de risco cardiovascular, por isso é usada como nossa variável resposta no estudo.

Por se tratar de uma doença complexa a DAC está associada a múltiplos fatores: ambientais, genéticos e a interação entre esses dois (Chen et al. 2011). Os fatores de risco clássicos associados são: dislipidemia, hipertensão arterial, diabetes, tabagismo e obesidade, história familiar precoce para a DAC e sedentarismo (Unal et al. 2004).

Muitos estudos têm demonstrado que os mecanismos complexos, determinados por variantes genéticas, podem influenciar a incidência de eventos cardíacos adversos importantes. Uma das ferramentas mais usadas para encontrar associação de novos polimorfismos (variantes genéticas) às doenças estudadas são os Estudos de associação genômica ampla (GWAS, do inglês Genome-Wide Association Studies). Eles fornecem informações preliminares sobre a associação entre genes e

fenótipos, sem que o pesquisador precise se valer de conhecimentos prévios sobre a fisiopatologia da doença para gerar candidatos a serem testados (Baudhuin 2009).

Escores de risco genéticos permitem fazer uma avaliação composta por dados genéticos em doenças complexas. Estas ferramentas permitem combinar o risco relacionado à saúde modelada pela informação genética em um único número, o que permite uma medida mais direta a ser utilizado para predição de risco (Smith et al. 2015).

No geral, o uso de GRS (do inglês, Genetic Risk Score) tem sido aplicado para melhorar a avaliação de risco com base em modelos de predição já estabelecidos. Partindo deste cenário, este trabalho teve como objetivo criar um GRS para estimar o risco de DAC, a partir de SNPs selecionados por meio de catálogo online de GWAS.

Modelos lineares generalizados foram utilizados de acordo com o comportamento dos dados de cada fator de risco. Variáveis como glicose, colesterol total, HDL-C, LDL-C, triglicerídeos e proteína C-reativa tiveram seus logs obtidos para serem ajustados ao modelo. O modelo foi ajustado por idade, sexo e ancestralidade africana. Depois disso, a análise de associação foi realizada entre CAC e os polimorfismos utilizando o modelo zero-inflacionado.

Para descrever os dados de CAC, foi utilizado o modelo ZIGamma ajustado para idade, sexo, ancestralidade africana e cada um dos SNP selecionados. Polimorfismos com valores de $p < 0.05$ foram selecionadas para o modelo de regressão múltipla. O CAC foi definido de acordo com Agatston et al., e foi usado como um variável contínua (Agatston and Janowitz 1990).

A partir dos resultados deste modelo um escore de risco genético foi formulado e, de forma análoga, outros dois escores também foram criados, sendo um deles apenas para Idade+Sexo e outro com somente a informação genética dos SNPs. Nós utilizamos a análise da curva ROC para estabelecer os melhores pontos de corte dos escores e para classificar os indivíduos como afetados com DAC (CAC ≥ 100) ou não afetados com a doença (CAC < 100).

Nós avaliamos o desempenho dos valores preditivo positivo e negativo do GRS, do modelo Idade+Sexo e FRS, analisamos também a sensibilidade e a especificidade. As análises foram realizadas com R versão 3.1.1 e SPSS versão 20.

SELEÇÃO DOS SNPs CANDIDATOS.

Os polimorfismos elegíveis a marcadores moleculares para DAC foram selecionados a partir de GWAS publicados até novembro de 2012 e organizados no banco de dados do National Human Research Genome Institute, de acordo com o fenótipo da doença ou o traço, a que o SNP está relacionado.

ANALISES DE ASSOCIAÇÃO DE CAC COM SNPs

Inicialmente foi realizada uma análise descritiva da variável CAC, para analisar algumas características de sua distribuição. A Tabela 1, apresenta algumas medidas do CAC. Observa-se que tem uma alta variabilidade e até o valor da mediana se tem zeros, ao observar com mais detalhe se determino que aproximadamente 78% da amostra apresenta valor de **zero** para o escore de cálcio.

Variable	minimo	Q1	median	media	Q3	maximo	desvio padrão
CAC	0	0	0	34.86	0	5363	184.107

Tabela 1. Medidas de resumo para a variável CAC.

O CAC tem uma distribuição contínua com um "ponto de massa em zero"; com isso, se mostra um caso especial de dados zero-inflacionado. Os modelos de regressão linear são inadequados para esse tipo de

dados. Em vez disso, usamos uma distribuição zero-inflacionada gama (ZIGamma) permitindo que os dados possam ser decompostos em duas partes: a primeira, um modelo binário que determina a probabilidade de que o resultado seja 0 ou superior a 0; e a segunda, uma distribuição contínua que modela o valor previsto para um escore de cálcio condicionado ao valor maior de zero. Este tipo de modelo é similar ao aplicado em estudos anteriores; por exemplo, o MESA (Multi-Ethnic Study of atherosclerosis) (Ma et al.2010). O modelo ZIGamma foi ajustado para idade, sexo, ancestralidade africana e cada um dos SNP selecionados.

$$CAC = \text{Modelagem contínua } (\mu) + \text{Modelagem discreta } (\eta)$$

Em que

$$\mu = \text{idade} + \text{sexo} + \text{ancestralidade} + \text{SNP1} + \epsilon$$

e

$$\eta = \text{idade} + \text{sexo} + \text{ancestralidade} + \text{SNP1} + \epsilon.$$

Polimorfismos com valores de $p < 0.05$ foram selecionadas para o modelo de regressão múltipla. O CAC foi definido de acordo com Agatston et al., e foi usado como um variável contínua (Agatston and Janowitz 1990).

Em total foram 265 modelos corridos. Como resultado do ajuste do modelo encontramos 56 SNPs que foram significativamente associados com o CAC; sendo que, seis deles foram significativos em ambas as partes do modelo. Na parte contínua do modelo, 33 SNPs foram estatisticamente associados ao CAC; sendo que, nove destes foram previamente associados diretamente com DAC, e oito foram previamente associados com diabetes mellitus tipo II. Na parte discreta do modelo, 17 SNPs foram significativamente associados com CAC; sendo que, quatro foram previamente associados com doença arterial coronariana, e cinco foram previamente associados com diabetes tipo II.

FORMULAÇÃO ESCORE GENÉTICO

Para a derivação do GRS, foi ajustado um modelo zero-inflacionado múltiplo com 871 indivíduos. Estes indivíduos são o resultado da genotipagem de 47 SNPs escolhidos para obter o escore.

A partir disso, foram calculados os valores ajustados da parte discreta e contínua do modelo. Os valores ajustados da parte contínua fornecem o valor estimado do CAC, dado $CAC > 0$. Os valores da parte discreta do modelo fornecem a probabilidade de o indivíduo ter $CAC = 0$. Depois disso, o valor do GRS é formulado como:

$$GRS = W * (1 - P(\widehat{CAC} = 0)) * (\widehat{CAC} | CAC > 0)$$

W corresponde a um peso específico, calculado como um peso que minimiza o erro quadrado médio dos valores ajustados de CAC. Em outras palavras, W varia de 0,1 a 5, em incrementos de 0,1, o W para este modelo corresponde a 0,76. Desta forma, o GRS é calculado como:

$$GRS = 0.76 * (1 - \hat{P}(CAC = 0)) * (\widehat{CAC} | CAC > 0)$$

O GRS foi derivado a partir do modelo zero-inflacionado, como descrito acima. Nós comparamos seu desempenho com o escore Idade+Sexo e FRS para DAC em 10 anos por meio de análise da área sobre a curva (Figura 10). Os pontos de corte para os escores foram obtidos por meio do ponto máximo da soma entre a sensibilidade e especificidade dos escores obtidos com os modelos ajustados.

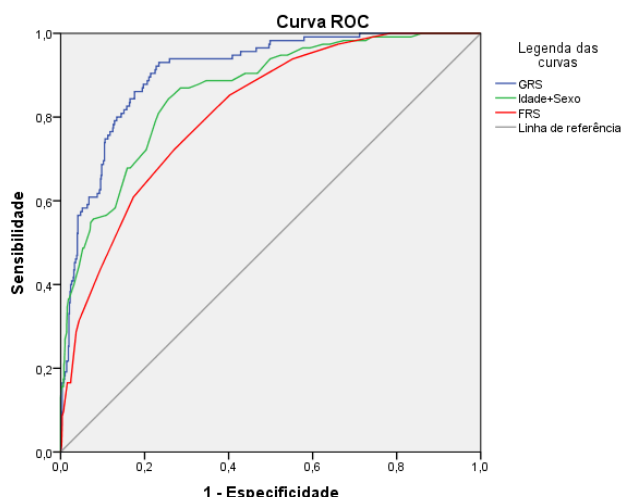


Figura 10. Gráfico das curvas ROC (Receiver operating characteristic). Curvas ROC dos escores GRS, Idade+Sexo e FRS para a predição de CAC. As áreas sob as curvas são 0,908, 0,860 e 0,812, respectivamente.

A Tabela 4 mostra que o GRS alcança uma predição de 91%, que foi maior do que os valores obtidos para o escore Idade+Sexo e FRS (86% e 81%, respectivamente). Além disso, o FRS e escore Idade+Sexo apresentaram menor sensibilidade do que o GRS. Em contrapartida, FRS foi o que apresentou maior valor preditivo positivo (VPP) devido a maior especificidade encontrada.

Tabela 6 – Desempenho do GRS em comparação com o escore Idade+Sexo e FRS

GRS	Idade+Sexo	FRS	
Estatística-C (IC 95%)	0.908	0.86	0.812
...	(0.882 - 0.934)	(0.825 - 0.895)	(0.774 - 0.850)
Sensibilidade (%)	93	84	43
Especificidade (%)	76	74	90
VPP (%)	38	33	41
VPN (%)	99	97	90

REFERÊNCIAS BIBLIOGRÁFICAS

Agatston A, Janowitz W. Quantification of coronary artery calcium using ultrafast computed tomography. J ... [Internet]. 1990 [cited 2015 Jun 11];15(4):827–32. Available from: <http://content.onlinejacc.org/article.aspx?articleid=1115376>

Aquino EM, Barreto SM, Bensenor IM, Carvalho MS, Chor D, Duncan BB, et al. Brazilian Longitudinal Study of Adult Health (ELSA-Brasil): objectives and design. Am J Epidemiol [Internet]. 2012/01/12 ed. 2012;175(4):315–24. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22234482

Baudhuin L. Genetics of coronary artery disease: focus on genome-wide association studies. Am J Transl Res [Internet]. 2009 [cited 2015 Jun 11];1(3):221–34. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2776326/>

Chen Z, Xie F, Ma G, Feng Y, Qian Q, Liu N. Study of the association between growth differentiation factor 15 gene polymorphism and coronary artery disease in a Chinese population. Mol Biol Rep [Internet]. 2011 Nov [cited 2015 Jun 10];38(8):5085–91. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21312063>

Elias-Smale SE, Proença RV, Koller MT, Kavousi M, van Rooij FJ a, Hunink MG, et al. Coronary calcium score improves classification of coronary heart disease risk in the elderly: the Rotterdam study. *J Am Coll Cardiol* [Internet]. Elsevier Inc.; 2010 Oct 19 [cited 2015 Jun 10];56(17):1407–14. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20946998>

Johnson RC, Leopold J a, Loscalzo J. Vascular calcification: pathobiological mechanisms and clinical implications. *Circ Res* [Internet]. 2006 Nov 10 [cited 2015 Jun 10];99(10):1044–59. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17095733>

Labree L, Azen SP, Doherty TM. Coronary Artery Calcium Score Combined With Framingham Score for Risk Prediction in Asymptomatic Individuals. *JAMA*. 2004;291(2):210–5.

Ma S, Liu A, Carr J, Post W, Kronmal R. Statistical modeling of Agatston score in multi-ethnic study of atherosclerosis (MESA). *PLoS One*. 2010 Jan;5(8):e12036.

Morrison AC, Bare L a, Chambless LE, Ellis SG, Malloy M, Kane JP, et al. Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study. *Am J Epidemiol*. 2007 Jul;166(1):28–35.

Murray CJL, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* [Internet]. 2012 Dec 15 [cited 2014 Jul 10];380(9859):2197–223. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23245608>

Polonsky TS, McClelland RL, Jorgensen NW, Bild DE, Burke GL, Guerci AD. Coronary Artery Calcium Score and Risk Classification for Coronary Heart Disease Prediction. *JAMA*. 2010;303(16):1610–6.

Smith JA, Ware EB, Middha P, Beacher L, Kardia SLR. Current Applications of Genetic Risk Scores to Cardiovascular Outcomes and Subclinical Phenotypes. *Curr Epidemiol Rep*. 2015;2:180–90.

Smith SC, Allen J, Blair SN, Bonow RO, Brass LM, Fonarow GC, et al. AHA/ACC guidelines for secondary prevention for patients with coronary and other atherosclerotic vascular disease: 2006 update: endorsed by the National Heart, Lung, and Blood Institute. *Circulation* [Internet]. 2006 May 16 [cited 2015 Mar 30];113(19):2363–72. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16702489>

Unal B, Critchley JA, Capewell S. Explaining the decline in coronary heart disease mortality in England and Wales between 1981 and 2000. *Circulation* [Internet]. 2004 Mar 9 [cited 2015 Jun 11];109(9):1101–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1499313>