



Modelo preditivo de risco genético de fissura labial com ou sem fissura palatina não-sindrômica na população brasileira

Carolina de Oliveira Silva*, Renato Assis Machado, Ricardo D. Coletta.

Resumo

Neste estudo, tivemos como objetivo determinar um modelo preditivo de risco genético para as fissuras labiais com ou sem fissura palatina não-sindrômicas (FL±PNS) na população brasileira, através de métodos machine learning. Inicialmente tivemos as fases de treinamento e validação das amostras, das quais 17 SNPs demonstraram altos scores de importância e 13 deles valores de p permutado significantes. Esse modelo contendo 13 SNPs deu origem a uma precisão maior que 90% de predição na classificação dos indivíduos com FL±PNS. Interações SNP-SNP significantes foram encontradas em IRF6 (rs642961), GRHL3 (rs41268753) e VAX1 (rs7078160). Nossos resultados revelam um modelo capaz prever o risco de desenvolvimento de FL±PNS na população brasileira.

Palavras-chave:

Fissuras labiais com ou sem fissura palatina não-sindrômica, machine learning, risco genético.

Introdução

As fissuras labiais com ou sem fissura palatina não-sindrômica (FL±PNS) são as malformações craniofaciais mais comum em seres humanos. Embora múltiplos genes e loci são associados à sua etiologia, não há modelos que transformem estas informações em modelos de predição de risco para as FL±PNS. O objetivo deste estudo foi estabelecer um modelo preditivo de risco genético para as FL±PNS na população brasileira a partir de métodos de aprendizado de máquina (*machine learning*).

Resultados e Discussão

Após fases de treinamento (2/3 das amostras) e validação (1/3 da amostra), 17 SNPs demonstraram altos escores de importância e uma taxa de erro de aproximadamente 3%, mas apenas 13 deles apresentaram valores significantes de p permutado. O modelo contendo os 13 SNPs resultou em uma acurácia de predição na classificação dos indivíduos com FL±PNS superior a 90%. Interações entre os SNPs localizados em IRF6 (rs642961), GRHL3 (rs41268753) e VAX1 (rs7078160) foram significantemente detectadas.

Tabela 1. Top 17 SNPs usando o escore de importância permutado baseada na análise de Random Forest.

SNP	MeanDecrease Accuracy	MeanDecrease Gini	Valor de P Perm.
rs2274976 (MTHFR)	4.36e-02	90.62	0.0009
rs2235371 (IRF6)	5.10e-02	75.27	0.0009
rs2236225 (MTHFD1)	3.79e-02	62.63	0.0009
rs41268753 (GRHL3)	2.29e-02	40.06	0.0009
rs1875735 (FGF12)	1.98e-02	30.13	0.0009
rs7078160 (VAX1)	1.58e-02	23.91	0.0009
rs10762573 (VCL)	1.33e-02	20.44	0.0009
rs227731 (17q22)	8.70e-03	19.25	0.0009
rs2131960 (VCL)	1.33e-02	18.63	0.0009
rs642961 (IRF6)	7.19e-03	16.33	0.003
rs2013162 (IRF6)	8.85e-03	16.24	0.68
rs17085106 (RHPN2)	1.02e-02	14.97	0.0009
rs4746172 (VCL)	1.19e-02	13.96	1.00
rs28372960 (TCOF1)	8.77e-03	13.31	0.004
rs1443434 (FOXE1)	6.33e-03	11.37	1.00
rs11717284 (FGF12)	1.06e-02	10.71	0.0009
rs7086344 (VAX1)	9.16e-03	10.55	1.00

Figura 1. Modelo Neural Network para discriminação de pacientes com FL±PNS e controles

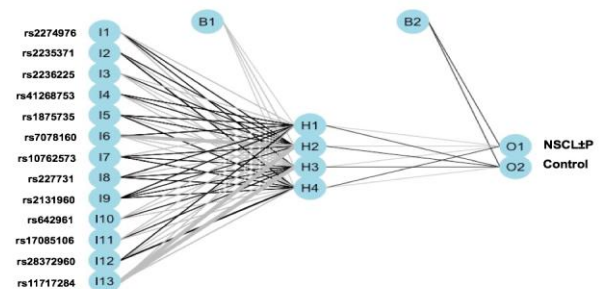


Tabela 2. Interações SNP-SNP baseados no modelo de redução de dimensionalidade multifatorial.

SNP1	SNP2	Valor de P	Valor de P Perm.
rs17085106 (RHPN2)	rs7078160 (VAX1)	2.98e-15	<0.001
rs7078160 (VAX1)	rs227731 (17q22)	3.59e-15	<0.001
rs7078160 (VAX1)	rs28372960 (TCOF1)	7.19e-13	<0.001
rs7078160 (VAX1)	rs642961 (IRF6)	3.40e-12	<0.001
rs17085106 (RHPN2)	rs28372960 (TCOF1)	5.35e-08	<0.001
rs7078160 (VAX1)	rs41268753 (GRHL3)	7.72e-08	<0.001
rs17085106 (RHPN2)	rs642961 (IRF6)	3.59e-06	<0.001
rs227731 (17q22)	rs2013162 (IRF6)	3.64e-06	<0.001
rs7086344 (VAX1)	rs227731 (17q22)	5.24e-06	<0.001
rs227731 (17q22)	rs10762573 (VCL)	3.88e-05	0.001
rs227731 (17q22)	rs2131960 (VCL)	8.72e-05	0.004
rs227731 (17q22)	rs2274976 (MTHFR)	3.00e-04	<0.001
rs227731 (17q22)	rs2235371 (IRF6)	3.25e-04	0.001
rs227731 (17q22)	rs28372960 (TCOF1)	3.80e-04	<0.001

Conclusões

Nossos resultados apontam um modelo capaz prever o risco de desenvolvimento de FL±PNS na população brasileira. Embora validação é garantida, este painel pode gerar um impacto muito grande na prática clínica, com potencial inovador e translacional.

Agradecimentos

SAE/UNICAMP

Li, Q., Kim, Y., Sukitipat, B., Hetmanski, J.B., Marazita, M.L., Duggal, P., Beaty, T.H., e Bailey-Wilson, J.E. *Genet Epidemiol.* 2015, 39, 385-94.
Zhang, S.J., Meng, P., Zhang, J., Jia, P., Lin, J., Wang, X., Chen, F., e Wei, X. *Genomics Proteomics Bioinformatics.* 2018, 16, 354-364.