EVALUATION BETWEEN MTR 2756A>G POLYMORPHISM AND TYPE 2 DIABETES RISK: AN INTEGRATED APPROACH WITH MACHINE LEARNING

João Pedro Costa Roriz [1]; Diolina Gonçalves da Silva [1]; Nayane Soares de Lima [1]; Stela Silva Noleto Leite de Carvalho [1]; Leandro do Prado Assunção [1]; Rodrigo da Silva Santos [1,2] e Angela Adamski da Silva Reis [1,2]

[1] Núcleo de Pesquisa em Neurogenética (NeuroGene), Instituto de Ciências Biológicas (ICB II), Universidade Federal de Goiás (UFG), Avenida Esperança, s/n, Campus Samambaia (Campus II), Zip Code: 74690-900, Goiânia, GO, Brazil.

[2] Departamento de Bioquímica e Biologia Molecular, Instituto de Ciências Biológicas (ICB II), Universidade Federal de Goiás (UFG), Avenida Esperança, s/n, Campus Samambaia (Campus II), Zip Code: 74690-900, Goiânia, GO, Brazil.

Angela Adamski da Silva Reis angela@ufg.br

INTRODUCTION

Type 2 diabetes *mellitus* (T2DM) and the folate cycle are interconnected through essential biochemical processes involved in cellular metabolism. Methionine synthase (MTR) converts homocysteine (Hcy) into methionine (Met) using folate and vitamin B12, preventing its accumulation and reducing metabolic and vascular risks. Furthermore, folate regulates DNA methylation, influencing the expression of genes associated with insulin resistance (IR). Alterations in this cycle can impair metabolic functions, contributing to the development of T2DM. This study investigated the association of the single nucleotide polymorphism (SNP) in the *MTR* gene, rs1805087, with the risk of developing T2DM.

METHODOLOGY

Case group 100 patients with T2DM

Control group 119 patients without T2DM

DNA extraction

Amplification and genotyping with PCR-RFLP SYBR Green assay followed by visualization of the RFLP products in 15% polyacrylamide gel

Allelic and genotypic frequencies compared using chi-square test and logistic regression analyzed diferente inheritance models, odds ratio was calculated to estimate association of genotypes with T2DM risk

Data was submitted to machine learning (ML) models which were used to predict getonypes and phenotype association. These models were avaliated by accuracy, precision and recall first in a training and test session

The results of those session were seen in confusion matrix and ROC curve. Area under the curve was calculated to evaluate the best model

RESULTS AND DISCUSSION

Results indicated that the SNP rs1805087 showed no association with T2DM risk across all inheritance models. Additionally, permutations tests revealed that the models RF, KNN and SVM identified age, gender, SNP, smoking and Alcohol comsumption as major predictors of T2DM. RF model proved to be the best model validated by na ROC curve of 0.99 but it was disconsidered due to suspicion of overfitting.

Table 1. Distribution on *MTR* genotypes in case and control groups under diferente inheritance models

Model	Genotype	Diabetes *mellitus* 2 No	Diabetes *mellitus* 2 Yes	Odds (CI 95%)	p-value	AIC	BIC
Codominant	AA	67(56.303%)	68(68%)	Ref	Ref		
Codominant	AG	47(39.496%)	29(29%)	0.608(0.34-1.073)	0.089	304.782	314.949
Codominant	GG	5(4.202%)	3(3%)	0.591(0.118-2.506)	0.484		
Dominant	AA	67(56.303%)	68(68%)	Ref	Ref		
Dominant	AG + GG	52(43.697%)	32(32%)	0.606(0.346-1.052)	0.077	302.783	309.561
Recessive	AA + AG	114(95.798%)	97(97%)	Ref	Ref		
Recessive	GG	5(4.202%)	3(3%)	0.705(0.142-2.948)	0.638	305.722	312.5
Overdominant	AA + GG	72(60.504%)	71(71%)	Ref	Ref		
Overdominant	AG	47(39.496%)	29(29%)	0.626(0.352-1.099)	0.105	303.288	310.066
Allele	A	181	165	-	0.125	-	-
Allele	G	57	35	-		-	-

CI = Confidence interval AIC = Akaike information criteria; BIC = Bayesian information criteria; Odds ratio with 95% CI; significance level = 0.05; p-value < 0.05

Table 2. Hardy-Weinberg equilibrium analysis for *MTR* genotypes in case and control groups

Group (n)	Genotype	Observed	Expected	Allelic frequency A (%)	Allelic fequency G (%)	x^2 (gl =1)	p-value
Control (119)	A/A	67	68,81	--	--	0,837	0,3602
	A/G	47	43,34	--	--		
	GG	5	6,81	--	--		
	A	181	--	0,7605 (76.05%)	--		
	G	35	--	--	0,2395 (23.95%)		
Case (100)	AA	68	68,06	--	--	0,0018	0,9654
	A/G	29	28,84	--	--		
	G/G	3	3	--	--		
	A	165	--	0,825 (82.5%)	--		
	G	35	--	--	0,175 (17.5%)		

Tests used: x^2 = Chi-square; Chi-square calculated with 1 level of liberty; significance level = 0.05; p-value < 0.05 and Hardy-Weinberg equilibrium caculated using x^2 results.

Table 3. Training and test of ML models

	ML	Accuracy	Precision	Recall
Train	KNN	0,720	0,736	0,728
Train	SVM	0,640	0,655	0,663
Train	RF	0,931	0,954	0,913
Test	KNN	0,591	0,667	0,667
Test	SVM	0,568	0,653	0,629
Test	RF	0,477	0,600	0,444

Training = 80%; Test = 20%; Metrics accessed were accuracy, precision and recall for model fitting.

Figure 1. ROC Curve and AUC

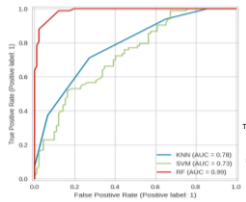

Figure 2 KNN Permutation of variables

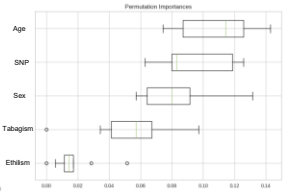

Figure 3 RF Permutation of variables

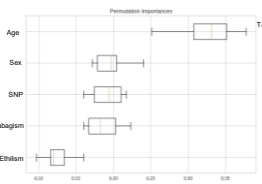

Figure 4 SVM Permutation of variables

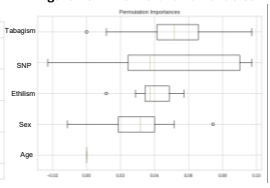

Table 4. Pattern prediction by ML models identifying gene-enviroment

Features	Codominant	Dominant	Recessive	Overdominant
Sex	0,279	0,283	1,000	0,186
Tabagism	**0,009**	0,068	**0,041**	0,386
Ethilism	0,220	0,158	0,916	0,258
HAS	0,575	1,000	0,735	0,898
RD	0,878	1,000	1,000	0,943
ND	0,443	0,387	0,916	0,570
NPP	0,248	0,212	0,892	0,360
PD	0,619	0,809	1,000	0,702
CD	0,684	1,000	0,916	1,000

HAS = Sistemic arterial hypertension; RD = Diabetic retinopathy; ND = Diabetic nephropathy; NPP = Diabetic peripheric neuropathy; PD = Diabetic feet; CD = Diabet cardiopathy; Inheritance models applied to ML with significance level = 0.05; p-value < 0.05.

CONCLUSION

In conclusion, the lack of association between rs1805807 and T2DM suggest that this SNP, alone, was not a determinant factor in the studied population. Conversely, the use of ML proved to be an effective approach in identifying patterns and predictors of T2DM, demonstrating that the combination of genetic and enviromental factos can significantly influence disease predisposition.

REFERENCES

SUPPORT

UFG CAPES CNPq

Espaço livre para inserir QR codes para referências, suplementos (planilhas, figuras, currículo, etc.) de acordo com o trabalho. Fonte Aptos, tamanho 30.