



## ESTUDO DE ASSOCIAÇÃO DE SNPS CANDIDATOS EM PACIENTES COM EPILEPSIA GENERALIZADA GENÉTICA

Felipe S. Kaibara\*, Rodrigo Secolin, Iscia Lopes-Cendes.

### I. INTRODUÇÃO

As epilepsias generalizadas genéticas (EGGs) são um grupo de síndromes caracterizadas pela presença de crises epiléticas associadas a padrões epileptiformes generalizados identificados por meio de eletroencefalograma. Variantes genéticas têm sido identificadas em pacientes com EGG em populações europeias e asiáticas. Entretanto, a ancestralidade e a diversidade genômica pode afetar os estudos de identificação de fatores genéticos de diversas doenças complexas, incluindo EGG. Portanto, é de extrema importância avaliar se as variantes genéticas associadas nestas populações também aparecem associadas na nossa população de pacientes brasileiros miscigenados com EGG.

### II. OBJETIVO PRINCIPAL

Avaliar associação entre variantes genéticas candidatas e pacientes brasileiros miscigenados com EGG.

### III. MATERIAIS E MÉTODOS

A amostra incluiu 67 pacientes com EGG, provenientes do Hospital de Clínicas da UNICAMP, e 340 indivíduos controles provenientes do banco de dados BIPMed (<https://bipmed.org/>) (1). Foram analisados 52 variantes candidatas em termos de SNPs, os quais foram extraídos dos dados de *SNP array* gerados pela plataforma *Genome-Wide Human SNP Array 6.0* (Affymetrix Inc) e já disponíveis no laboratório de Genética Médica e Medicina Genômica da FCM-UNICAMP. Para o processamento e análise estatística dos resultados, foram utilizados os programas PLINK v1.9 (2) e R Studio (<https://rstudio.com/>). Entretanto, somente 17 variantes foram encontradas nos dados de *SNP array*. Neste caso, foi realizada a imputação das variantes não encontradas no BIPMed pelo programa Beagle 5.1 (3). A fim de agilizar o processo computacional, somente regiões dos SNPs candidatos foram imputadas.

A fim de avaliar se a amostra de casos e controles apresenta estratificação devido a miscigenação da população, foi realizado o teste de AMOVA (*Analysis of Molecular Variance*), por meio do pacote *poppr.amova* (4) no programa R (<https://www.r-project.org/>), e a partir de um conjunto de 10.000 SNPs aleatórios do genoma extraídos do *SNP array*. Além disso, a estrutura genômica da nossa amostra foi comparada com outras populações continentais por meio da análise dos componentes principais (*principal component analysis-PCA*).

#### IV. RESULTADOS

Dos 52 SNPs do estudo, 4 foram desconsiderados desde o início por serem *indels* (inserções ou deleções). Os 48 restantes apresentaram uma frequência alélica maior que 0.01 em nossas amostras, porém oito destes (rs12059546, rs12185644, rs1332470, rs1046276, rs1402398, rs2947349, rs4596374 e rs55670112) apresentaram desequilíbrio de *Hardy-Weinberg*. Devido a isso, foram descartados das análises posteriores, pois poderiam oferecer resultados errôneos devido à distribuição genotípica tendenciosa e não de associação.

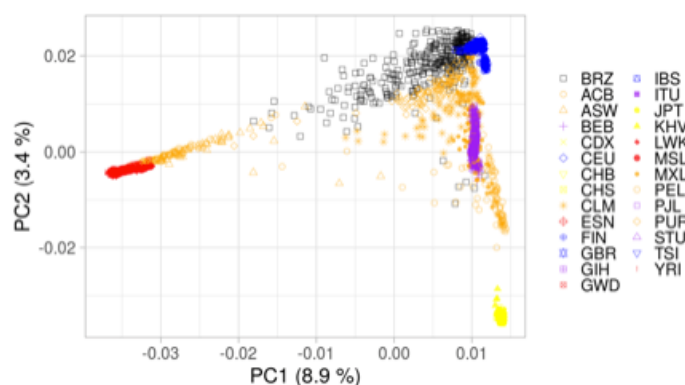
Por meio dos resultados da análise de regressão logística, o SNP rs9788 mostrou associação com EGG, com um valor de p nominal=0.03663 (OR=1.44; IC95%=1.02 – 2.08). No entanto, este SNP não apresentou resultados satisfatórios com a correção para múltiplos testes por Bonferroni (p corrigido=0.6226). Outros dois SNPs, rs1939012 e rs4794333, apresentaram casos similares. O rs1939012 apresentou valor de p nominal=0.01641 (OR=0.6166; IC=95% 0.4154 – 0.9152) e rs4794333 obteve p nominal= 0.01493 (OR=0.6074; IC=95% 0.4066 – 0.9075). Ambos também não passaram pela correção para múltiplos testes de Bonferroni.

**Tabela 1.** Análises descritiva e de regressão logística dos 48 SNPs analisados. Em cinza estão os dados de SNPs imputados, em branco, dados genômicos do SNP-array. Em amarelo, SNPs citados no texto como possível associação.

CHR	SNP	Position (BP)	MAF	HWE p-value	Allele	OR	95% CI	p-value	
								nominal	adjusted
1	rs12136213	10046460	0.2681	0.7882	G	0.792	0.5138 - 1.221	0.2909	1
1	rs12059546	239970097	0.2647	0.005	G	-	-	-	-
2	rs4664630	23898317	0.1824	0.019	G	1.33	0.91 – 1.95	0.1353	1
2	rs13026414	57934055	0.3486	1	T	0.9571	0.6477 - 1.414	0.826	1
2	rs4671319	57950346	0.4222	0.1598	G	0.7494	0.5168 - 1.087	0.1283	1
2	rs1402398	58042241	0.35	0.01469	G	-	-	-	-
2	rs12185644	58051769	0.3309	0.008	C	-	-	-	-
2	rs2947349	58059803	0.3069	0.006447	C	-	-	-	-
2	rs10496964	145359909	0.1324	0.310	A	1.29	0.83 – 2.02	0.2573	1
2	rs13020210	145381225	0.2125	0.4299	G	1.2	0.7805 - 1.845	0.4062	1
2	rs11890028	166943277	0.2569	0.3361	G	1.132	0.7363 - 1.742	0.5713	1
2	rs13406236	167084615	0.2944	1	C	0.9729	0.6462 - 1.465	0.8952	1
2	rs887696	191583507	0.4014	0.8268	C	1.184	0.8145 - 1.72	0.3767	1
3	rs624755	61733962	0.3838	0.758	C	0.86	0.61 – 1.23	0.4158	1
3	rs1374679	63075267	0.3868	0.125	C	0.94	0.66 – 1.35	0.7392	1
3	rs782728	66326302	0.4681	1	A	0.8784	0.6047 - 1.276	0.4962	1
3	rs111577701	167861408	0.07361	0.4246	T	0.8584	0.4143 - 1.778	0.6812	1
4	rs1044352	31147874	0.4485	0.770	T	1.18	0.84 – 1.65	0.3407	1
4	rs28498976	31151357	0.4458	0.2002	A	1.288	0.8738 - 1.899	0.2011	1
4	rs535066	46240287	0.4125	1	G	0.9308	0.637 - 1.36	0.7109	1
4	rs11943905	46397617	0.2931	0.3092	T	1.248	0.8476 - 1.839	0.2613	1
5	rs4596374	114221505	0.475	0.02627	C	-	-	-	-
5	rs55670112	114268470	0.4819	0.04578	C	-	-	-	-
5	rs357608	150840380	0.4806	0.2452	T	0.9894	0.6903 - 1.418	0.9536	1
5	rs2069347	162867195	0.4618	0.697	G	1.14	0.82 – 1.58	0.4487	1

5	rs1025482	166893257	0.4875	0.4619	C	0.9905	0.6881 - 1.426	0.9592	1
5	rs1432881	166932520	0.4221	0.429	A	1.15	0.81 - 1.61	0.4375	1
5	rs244903	167913510	0.3639	0.4278	G	1.118	0.7559 - 1.655	0.5755	1
6	rs68082256	16971575	0.1833	0.2934	A	0.7225	0.4353 - 1.199	0.2085	1
6	rs13200150	128309768	0.2222	0.1711	G	1.019	0.641 - 1.619	0.937	1
11	rs1939012	102595135	0.4458	0.6695	T	0.6166	0.4154 - 0.9152	0.01641	1
13	rs1008812	23966145	0.4653	1	A	0.8587	0.5907 - 1.248	0.425	1
13	rs1332470	91417190	0.3029	3.5e-05	G	-	-	-	-
16	rs1046276	30914626	0.3647	7.7 e-13	A	-	-	-	-
16	rs4638568	50045839	0.0853	0.778	T	1.55	0.92 - 2.61	0.1006	1
17	rs12951323	46027565	0.2056	0.1075	A	0.6177	0.3651 - 1.045	0.07265	1
17	rs4794333	46045495	0.3556	0.05109	C	0.6074	0.4066 - 0.9075	0.01493	1
17	rs72823592	46123004	0.1806	0.7209	A	0.6197	0.3589 - 1.07	0.08595	1
18	rs2665558	48402338	0.4514	0.6707	T	0.9181	0.6284 - 1.341	0.6587	1
18	rs2255610	48404784	0.4736	0.7518	G	1.102	0.757 - 1.603	0.6131	1
18	rs608781	48407326	0.1181	0.4472	C	1.61	0.932 - 2.782	0.08775	1
18	rs2850545	48414235	0.4708	0.5976	A	1.078	0.7391 - 1.572	0.6964	1
18	rs645088	48456903	0.3324	0.380	T	0.85	0.59 - 1.23	0.3895	1
18	rs649224	48458662	0.1309	0.403	T	1.16	0.73 - 1.84	0.5322	1
18	rs654136	48464204	0.4971	0.923	T	0.90	0.65 - 1.26	0.5452	1
19	rs9788	53719250	0.3029	0.656	T	1.46	1.02 - 2.08	0.0366	0.6226
21	rs2833098	32183996	0.3694	0.9099	G	1.081	0.7358 - 1.589	0.6905	1
21	rs2839377	48077812	0.4750	0.384	A	1.19	0.85 - 1.68	0.3091	1

Os resultados do teste de AMOVA mostraram que 99,69% da fonte de variação genética foi encontrada dentro de cada grupo, e somente 0,31% da variação foi encontrada entre os casos e controles (AMOVA  $\phi$ -statistics = 0.0031; p-value = 0.0001), o que indica que nossa amostra de casos e controles não apresenta estratificação por subestrutura genética, a qual não apresenta resultados espúrios de associação genética. Além disso, de acordo com a figura 1, a nossa amostra de casos e controles apresenta uma maior proximidade com as populações europeias, seguido das populações africanas subsaarianas e de nativos americanos, mostrando que nossa amostra não apresenta desvios que poderiam ser considerados erros de genotipagem ou análise.



**Figura 1.** Gráfico de dispersão, mostrando os valores dos componentes principais PC1 e PC2 e suas respectivas porcentagens de variabilidade dos dados, das amostras de pacientes com EGG e controles brasileiros (BRZ, em preto) e as demais populações descritas no 1KGP. As cores representam as populações em nível continental: vermelho=África sub-saariana; azul=Europa; laranja=Americanos miscigenados; roxo=sudoeste Asiático; amarelo=Ásia oriental.

## V. DISCUSSÃO

**Discussão dos resultados da análise descritiva e regressão logística.** Os resultados obtidos da análise de regressão logística levam à conclusão inicial de que nenhum SNP obteve resultado 100% comprobatório de associação com EGG. No entanto, deve-se levar em consideração as variantes rs9788, rs1939012 e rs4794333. No caso do SNP rs9788, o valor de  $p$  nominal=0.03663 é um resultado significativo (menor que 0.05) e, mesmo não tendo apresentado um desempenho satisfatório após a correção de múltiplos testes de Bonferroni, tal restrição não descarta por completo a possível contribuição deste SNP em pacientes com EGG. Também é importante ressaltar que a variante rs9788 obteve resultados similares tanto na análise do SNP-array quanto na análise dos dados imputados, o que reforça a atenção especial a esta variante e atribui maior confiabilidade à imputação.

Os SNPs rs1939012 e rs4794333 apresentaram valor de  $p$  nominal sugestivos de associação na análise dos dados imputados, porém nenhum obteve desempenho satisfatório após a correção de múltiplos testes de Bonferroni. Também devem ser analisados em futuros estudos.

Em suma, especialmente o SNP rs9788, mas também rs1939012 e rs4794333, merecem atenção em futuras pesquisas de maior coorte, visto que apresentaram resultados diferenciados das demais variantes.

**Discussão do perfil populacional da amostra.** O gráfico PCA sugeriu uma maior proximidade da população brasileira à europeia e logo em seguida, à africana. A amostra, de forma contraintuitiva a princípio, não se aproxima tanto de outras populações americanas miscigenadas (Figura 1). De fato, estudos mostram que a população brasileira apresenta maior proporção do componente ancestral europeu, seguida do componente africano subsaariano e por final, nativo americano (6-8). A proporção destes componentes é diferente em comparação com outras populações americanas miscigenadas (9, 10).

Ao comparar frequências alélicas dos SNPs candidatos entre a amostra do estudo e populações de referência, verificamos que 29 de 48 SNPs possuem frequências alélicas diferentes entre a nossa amostra e estas populações. Tal informação favorece muito a hipótese de que os estudos genéticos de EGG realizados nestes locais (principalmente Europa, da qual grande parte dos artigos científicos são oriundos) podem não ser aplicáveis no Brasil, uma vez que as frequências alélicas são diferentes na população brasileira em 60% das variantes candidatas, sendo que grande parte das variantes utilizadas no estudo foram encontradas em estudos com populações europeias. Desta forma, mostra-se a importância de novos estudos na população brasileira no que se refere à EGG, uma vez que os resultados obtidos em populações europeias, por exemplo, nem sempre são replicáveis em outras populações, como a brasileira.

## VI. CONCLUSÃO

Em conclusão, estes resultados reforçam a hipótese de que as variantes que podem estar associadas à EGG na nossa população brasileira miscigenada, em comparação com os resultados encontrados para populações europeias e asiáticas, diferem. Isso indica a importância do avanço de saúde de precisão e de futuros estudos

científicos de associação genética e EGG em amostras brasileiras, a fim de fornecer tratamentos e diagnósticos mais precisos e eficientes para os pacientes da população brasileira.

## VII. REFERÊNCIAS

1. Rocha CS, Secolin R, Rodrigues MR, Carvalho BS, Lopes-Cendes I. The Brazilian Initiative on Precision Medicine (BIPMed): fostering genomic data-sharing of underrepresented populations. *npj Genomic Medicine*. 2020;5(1):42.
2. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*. 2007;81(3):559-75.
3. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am J Hum Genet*. 2018;103(3):338-48.
4. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*. 1992;131(2):479-91.
5. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*. 2020;581(7809):434-43.
6. Rodrigues de Moura R, Coelho AVC, de Queiroz Balbino V, Crovella S, Brandão LAC. Meta-analysis of Brazilian genetic admixture and comparison with other Latin America countries. *American Journal of Human Biology*. 2015;27(5):674-80.
7. Kehdy FSG, Gouveia MH, Machado M, Magalhães WCS, Horimoto AR, Horta BL, et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences*. 2015;112(28):8696-701.
8. Secolin R, Mas-Sandoval A, Arauna LR, Torres FR, de Araujo TK, Santos ML, et al. Distribution of local ancestry and evidence of adaptation in admixed populations. *Sci Rep*. 2019;9(1):13900.
9. Ruiz-Linares A, Adhikari K, Acuña-Alonzo V, Quinto-Sanchez M, Jaramillo C, Arias W, et al. Admixture in Latin America: Geographic Structure, Phenotypic Diversity and Self-Perception of Ancestry Based on 7,342 Individuals. *PLoS Genetics*. 2014;10(9):e1004572-e.
10. Homburger JR, Moreno-Estrada A, Gignoux CR, Nelson D, Sanchez E, Ortiz-Tello P, et al. Genomic Insights into the Ancestry and Demographic History of South America. *PLoS Genetics*. 2015;11(12).