

Perturbação do espectro do autismo: identificação de variantes em genes envolvidos no mecanismo regulador da expressão génica *nonsense-mediated mRNA decay*

Autism spectrum disorder: identification of gene variants involved in the nonsense-mediated mRNA decay

Ana Rita Marques^{1,2}, João Xavier Santos^{1,2}, Hugo Martiniano^{1,2}, Joana Vilela^{1,2}, Célia Rasga^{1,2}, Luísa Romão^{1,3}, Astrid Moura Vicente^{1,2}

astrid.vicente@insa.min-saude.pt

(1) Departamento de Promoção da Saúde e Prevenção de Doenças Não Transmissíveis, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

(2) Instituto de Biosistemas e Ciências Integrativas. Faculdade de Ciência, Universidade de Lisboa, Lisboa, Portugal

(3) Departamento de Genética Humana, Instituto Nacional de Saúde Doutor Ricardo Jorge, Lisboa, Portugal

_Resumo

A Perturbação do Espectro do Autismo (PEA) é uma patologia do neurodesenvolvimento caracterizada por dificuldades de socialização e/ou comunicação e por comportamentos estereotipados e repetitivos. Estudos de heritabilidade indicam que fatores genéticos contribuem para 50-80% do risco de desenvolver PEA, mas os mecanismos genéticos não são bem conhecidos. Neste estudo, exploramos a contribuição para o risco de PEA de genes que codificam proteínas envolvidas num mecanismo regulador da expressão génica, o *Nonsense-Mediated mRNA Decay* (NMD). Para este fim, compilámos uma lista de 46 genes envolvidos no NMD e investigámos a presença de *Single Nucleotide Variants* (SNVs) e *Copy Number Variants* (CNV) nestes genes em duas amostras populacionais de indivíduos diagnosticados com PEA (n=1828 e n=3570, respetivamente). Identificámos variantes raras (MAF<1% em controlos) com efeito deletério de acordo com uma predição *in silico*, e verificámos se afetam domínios proteicos necessários para o NMD. Identificámos 270 SNVs em 524 pacientes com PEA e 38 CNVs em 38 pacientes com PEA. Destas, 136 variantes (122 SNVs e 11 CNVs) estão localizadas em regiões dos genes que codificam domínios proteicos importantes para o funcionamento do mecanismo de NMD. Deste modo, colocamos a hipótese de que estas variantes possam afetar a função normal do NMD e, consequentemente, contribuir para alterações na expressão dos genes-alvo, podendo assim constituir um fator de risco para a PEA.

_Abstract

*Autism Spectrum Disorder (ASD) is a neurodevelopmental condition characterized by impaired social/communication skills and stereotyped/repetitive behaviors. Genetic factors account for 50-80% of the familial risk of ASD, but genetic determinants are not fully understood. In this study, we explored the contribution to ASD etiology of genes involved in an important post-transcriptional regulatory mechanism, the Nonsense-Mediated mRNA Decay (NMD). We first compiled a group of 46 genes encoding NMD factors and regulators. In these genes, we searched for Single Nucleotide Variants (SNVs) and Copy Number Variants (CNVs) in two samples of ASD patients (n=1828 and n=3570, respectively). In genes with rare variants (MAF<1% in controls) predicted *in silico* to be deleterious, we further investigated whether these variants affect protein domains required for NMD. We identified 270 SNVs in 524 ASD patients and 38 CNVs in 38 ASD patients. We found 136 variants (122 SNVs and 11 CNVs)*

located within protein domains that may affect proper NMD function and consequently contribute to changes in the expression of NMD targets. This study suggests that genetic variants that cause NMD impairment may constitute a risk factor to ASD pathophysiology.

_Introdução

A Perturbação do Espectro do Autismo (PEA) é uma patologia do neurodesenvolvimento caracterizada por dificuldades de comunicação e socialização associadas à presença de comportamentos repetitivos e estereotipados (1) e manifestações clínicas heterogéneas. A etiologia da PEA ainda não está bem esclarecida, o que impossibilita o desenvolvimento de terapias eficazes. Muitos estudos têm sido desenvolvidos para determinar as causas da PEA, conduzindo à descoberta de fatores genéticos, epigenéticos e ambientais associados à suscetibilidade para a perturbação. Sendo a origem da PEA tão heterogénea, coloca-se a hipótese de que mecanismos reguladores da expressão génica possam contribuir para o desenvolvimento da patologia quando afetados.

Existem vários mecanismos de controlo de qualidade da expressão génica, entre os quais, o decaimento do mRNA mediado por mutações nonsense (*Nonsense-Mediated mRNA Decay*, NMD). O mecanismo de NMD permite degradar transcritos que contenham codões de terminação de tradução prematura (*premature translation termination codon*, PTC), reduzindo a sua abundância, mas também é responsável pela regulação da expressão de ~10% de transcritos fisiológicos (2). A função do NMD é essencial durante o



neurodesenvolvimento (3-6) e depende de várias proteínas. Alguns genes que codificam fatores do NMD foram anteriormente implicados na PEA. Por exemplo, foram identificadas mutações no gene *UPF3B* em indivíduos que exibem características do autismo (6,7) e foi demonstrado que mutações *missense* neste gene alteram a diferenciação das células progenitoras neuronais devido a uma disfunção do processo de NMD (5,6). Adicionalmente, um estudo recente mostrou que a perda da proteína FMRP, codificada pelo gene *FMR1*, que representa uma das principais causas de deficiência intelectual e autismo, resulta numa hiperativação do mecanismo de NMD (8). Assim, testámos a hipótese de que alterações em genes que codificam para proteínas envolvidas no NMD podem contribuir para a PEA através da análise de variantes em genes envolvidos no mecanismo NMD, numa amostra populacional de grande dimensão com PEA.

_Objetivos

Este estudo teve como objetivo a identificação de variantes em genes que codificam proteínas envolvidas no mecanismo de NMD em indivíduos diagnosticados com perturbação do espectro do autismo (PEA) e a análise do seu contributo para o risco de PEA.

_Material e métodos

Os genes que codificam fatores e reguladores do NMD (do-ravante denominados genes NMD), foram selecionados através de uma pesquisa na literatura e em bases de dados, incluindo a Gene Ontology AmiGO (<http://amigo.geneontology.org>) (9) e a Reactome (www.reactome.org) (10).

Foram analisadas três *datasets* de indivíduos com PEA: *Autism Sequencing Consortium* (ASC, n=1828) (11), *Autism Genome Project* (AGP, n=2446) (12) e *Simons Simplex Collection* (SSC, n=1024) (13) para identificar Variantes Nucleotídicas Pontuais (*Single Nucleotide Variants*, SNVs) e Variantes de Número de Cópias (*Copy Number Variants*, CNVs) nos genes NMD. A sequenciação do exoma completo permite identificar SNVs nas regiões codificantes do genoma (exões), enquanto os CNVs nestes *datasets* foram detetados através da tecnologia de arrays. Estimámos ainda a fre-

quência destas variantes em populações controlo sem historial de doença neurológica, nomeadamente 60146 controlos da GnomAD v2.1.1 (<https://gnomad.broadinstitute.org/>, para SNVs) e 10355 controlos da Database of Genomic Variants (<http://dgv.tcag.ca/dgv/app/home>, para CNVs). Para identificar as variantes potencialmente deletérias, seleccionámos apenas variantes raras em controlos (MAF<1%), nomeadamente os CNVs que continham regiões codificantes dos genes em análise, assim como os SNVs classificados como deletérios *in silico* pelos programas Polyphen (14) e SIFT (15) (*missense predicted damaging and deleterious*, MISPD) ou como perda de função (*loss-of-function*, LoF). Adicionalmente, pesquisamos na literatura quais os domínios proteicos comprovadamente necessários para o correto funcionamento do NMD e identificamos as variantes deletérias localizadas nestas regiões.

_Resultados e discussão

Da literatura e bases de dados foram identificados 46 genes que codificam proteínas envolvidas no processo de NMD (tabela 1). Estes genes foram divididos em grupos de acordo com sua função no NMD: EJC – fatores e reguladores do complexo de junção exão-exão (*Exon Junction Complex*, EJC); SURF-DECID – componentes do complexo SMG1-UPF1-eRF1-eRF3 (SURF) e indutores de decaimento (DECID); mRNA decay – fase de decaimento do mRNA; ER-NMD – resposta NMD no Retículo Endoplasmático (RE); Regulatório – envolvido na regulação do NMD (tabela 1). Destes 46 genes, 7 são genes candidatos para PEA descritos na base de dados SFARI (<https://gene.sfari.org/>). Verificámos ainda na *Human Protein Atlas* (www.proteinatlas.org) (16) que todos os 46 genes são expressos no cérebro humano.

A presença de SNVs nestes genes foi analisada nos exomas do *dataset* ASC (N=1828) e de CNVs nos *datasets* AGP (n=2446) e SSC (n=1024). Identificámos 270 SNVs potencialmente deletérios (por predição *in silico*) em 38 genes NMD em 28,7% (524/1828) dos pacientes com PEA, dos quais 11,5% (31/270) são classificados como LoF e 88,5% (239/270) como MISPD (figura 1).



Tabela 1: ↓ Lista de genes humanos que codificam fatores e reguladores do mecanismo de *nonsense-mediated mRNA decay* (NMD).

Grupo	Símbolo do gene		
EJC	<i>EIF4A3</i>	<i>MAGOHB</i>	<i>PYM1</i>
	<i>RBM8A</i>	<i>CASC3</i>	<i>RNPS1</i>
	<i>MAGOH</i>	<i>ICE1</i>	
SURF-DECID	<i>UPF1</i>	<i>GSPT1</i>	<i>RUVBL2</i>
	<i>UPF2</i>	<i>NCBP1</i>	<i>SMG1</i>
	<i>UPF3B</i>	<i>NCBP2</i>	<i>SMG8</i>
	<i>EIF4E</i>	<i>DHX34</i>	<i>SMG9</i>
	<i>ETF1</i>	<i>RUVBL1</i>	
mRNA decay	<i>SMG5</i>	<i>PNRC2</i>	<i>PARN</i>
	<i>SMG7</i>	<i>DCP2</i>	<i>MOV10</i>
	<i>SMG6</i>	<i>XRN1</i>	<i>PPP2CA</i>
	<i>CNOT8</i>	<i>DIS3L</i>	<i>PPP2R1A</i>
	<i>DCP1A</i>	<i>DIS3L2</i>	<i>EXOSC10</i>
ER-NMD		<i>NBAS</i>	
Regulatório	<i>PABPC1</i>	<i>FMR1</i>	<i>SEC13</i>
	<i>EIF4G1</i>	<i>EIF3E</i>	<i>GNL2</i>
	<i>UPF3A</i>	<i>SRSF1</i>	

A cinza estão marcados os genes candidatos do autismo de acordo com a base de dados SFARI (<https://gene.sfari.org/>).

O *NBAS*, um fator importante para a resposta do NMD no retículo endoplasmático (17), foi o gene com mais SNVs identificados: 25 em 5.4% (98/1828) da população ASC (tabela 2). Adicionalmente, os genes com SNVs presentes em mais de 1% dos indivíduos PEA foram: *DIS3L2* (4%), *DIS3L* (3%), *SMG7* (2%), *CASC3* (1,9%), *DCP1A* (1,7%), *ICE1* (1,6%), *DHX34* (1,5%) e *SMG6* (1,3%) (figura 1). Três destes genes codificam as proteínas *CASC3*, *ICE1* e *DHX34*, que promovem a ativação do NMD, enquanto os outros genes codificam proteínas envolvidas na degradação do mRNA. Analisámos os domínios proteicos afetados pelos SNVs identificados neste estudo, e verificamos que grande parte das variantes se localizam nos domínios proteicos necessários para o NMD. Por exemplo, os SNVs identificados nos genes

que codificam os fatores *up-frameshift* (*UPF1*, *UPF2* e *UPF3*) e supressor *with morphogenesis in genitalia* (*SMG1*, *SMG5*, *SMG6*, *SMG7*, *SMG8* e *SMG9*) localizam-se maioritariamente em regiões que codificam domínios essenciais para a função do NMD. Adicionalmente, encontramos variantes numa região do *UPF3B*, onde já foram descritas mutações que interferem com a função do NMD (6).

A análise conjunta dos *datasets* AGP e SSC, levou à descoberta de 38 CNVs contendo a totalidade ou parte de 18 dos 46 genes NMD, em 1% (38/3570) dos indivíduos com PEA. Destes CNVs, 8 são deleções e 30 são duplicações. Verificámos que os CNVs que incluem genes inteiros ou domínios funcionais importantes para o NMD são raros nos indivíduos com PEA (0,4%), mas extremamente raros em indivíduos controlo (<0,1%).

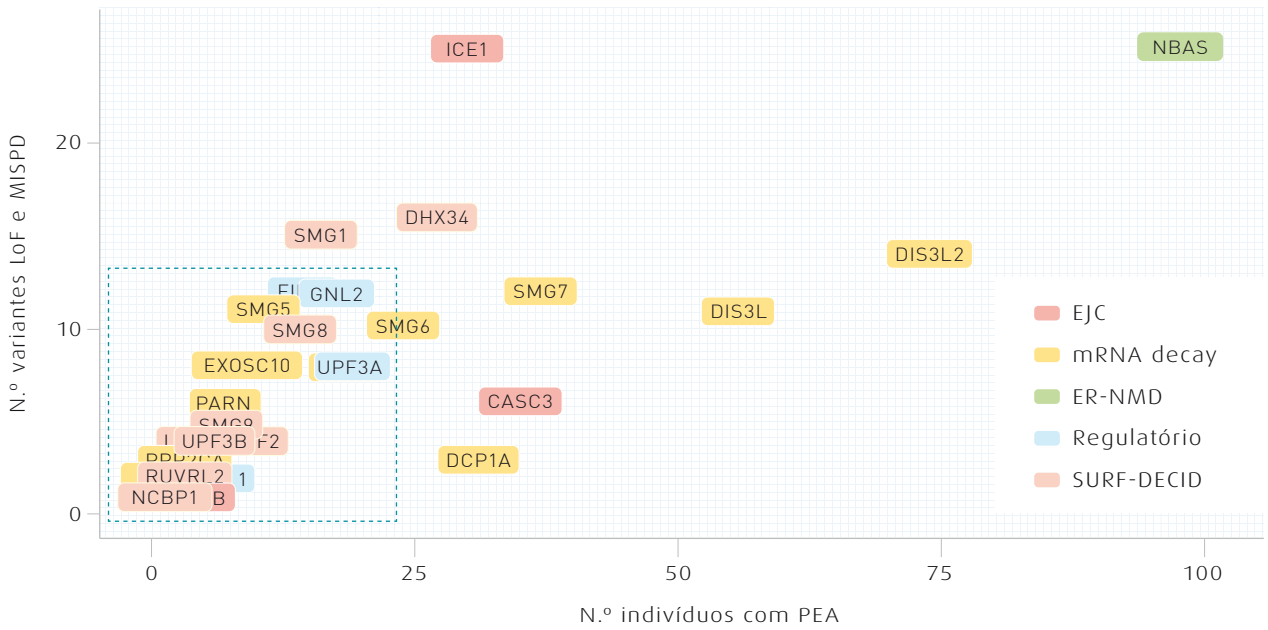
No total, foram identificadas 136 variantes (122 SNVs e 11 CNVs) em 23 genes, localizadas em regiões que codificam domínios proteicos importantes para a sua função no NMD (tabela 2), em 258 indivíduos com PEA. A *classificação in silico* dos SNVs como tendo um efeito deletério, sugere que possam afetar o mecanismo de NMD e contribuir para alterações na expressão dos genes-alvo. É notável que existam pelo menos 1277 genes-alvo de NMD expressos em células neuronais (8). Será agora fundamental a validação funcional destas variantes, através de estudos funcionais em linhas celulares ou da análise de expressão dos seus genes-alvo.

O NMD desempenha um papel importante na diferenciação neuronal (6) e a magnitude da sua resposta pode influenciar a diferenciação celular numa fase inicial do desenvolvimento (3). Muitas das variantes genéticas reportadas em indivíduos diagnosticados com PEA são mutações LoF, que levam à inserção de um PTC. O NMD é essencial na proteção contra estes erros genéticos, sendo, portanto, plausível que a disfunção deste mecanismo afete a expressão destes genes, influenciando o risco de PEA. No contexto da literatura hoje, os presentes resultados sugerem que uma disfunção do mecanismo de NMD pode constituir um fator de risco para a fisiopatologia da PEA.



Figura 1: Representação dos genes com SNVs identificados em indivíduos com perturbação do espectro do autismo (PEA).

a) Genes identificados, marcados com cores de acordo com a sua função no *nonsense-mediated mRNA decay* (NMD)



b) Grupo de genes com variantes identificadas em <1% dos indivíduos com PEA

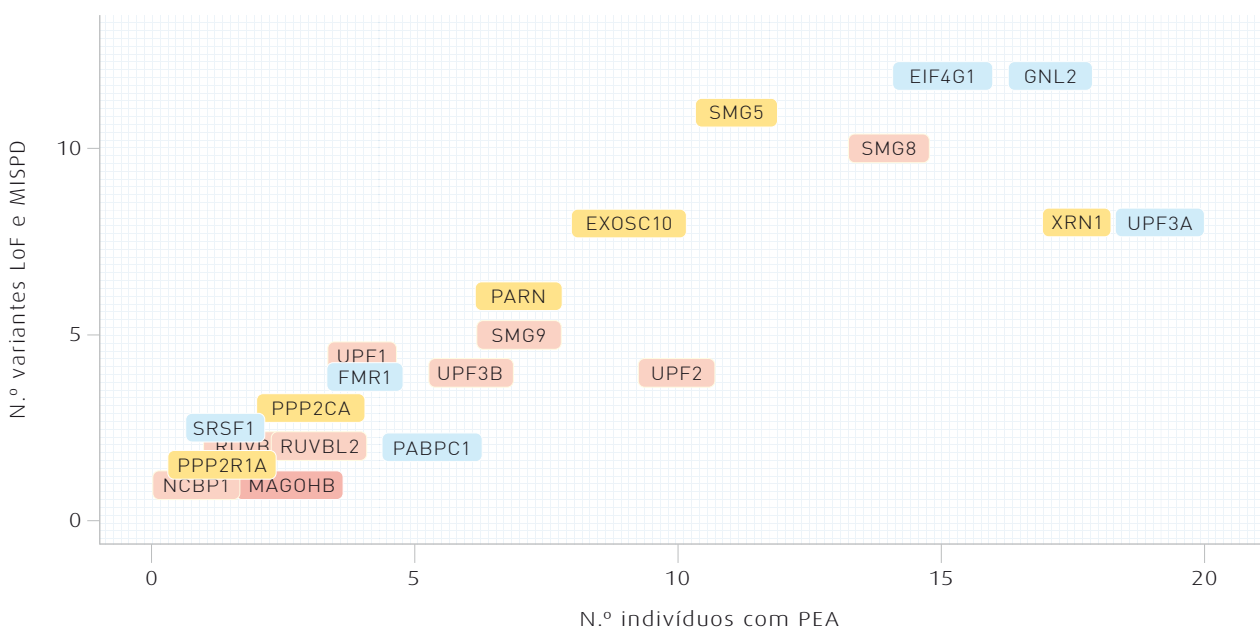




Tabela 2: ↓ Número de SNVs e CNVs localizados em regiões dos genes que codificam domínios importantes para a sua função no *non-sense-mediated mRNA decay* (NMD) em 258 indivíduos com perturbação do espectro do autismo (PEA).

Gene	Localização	SNVs ⁽¹⁾		CNVs ⁽¹⁾	PEA n ⁽²⁾
		LoF	MISPD		
<i>EJNBAS</i> ⁽³⁾	2p24.3	5	20	1	99
<i>DHX34</i>	19q13.32	2	12	1	25
<i>GNL2</i> ⁽³⁾	1p34.3	1	11	–	17
<i>SMG7</i>	1q25.3	–	8	–	14
<i>SMG1</i>	16p12.3	1	12	–	13
<i>UPF2</i>	10p14	1	3	1	11
<i>UPF3A</i>	13q34	1	2	2	11
<i>ICE1</i>	5p15.32	2	7	–	10
<i>SMG8</i>	17q22	–	6	–	8
<i>GSPT1</i>	16p13.13	–	4	1	7
<i>UPF3B</i>	Xq24	1	3	1	7
<i>ETF1</i>	5q31.2	1	3	–	6
<i>SMG6</i>	17p13.3	–	2	1	6
<i>RBM8A</i>	1q21.1	–	–	2	4
<i>SMG9</i>	19q13.31	1	1	–	4
<i>UPF1</i>	19p13.11	–	4	–	4
<i>CASC3</i>	17q21.1	–	3	–	3
<i>RUVBL2</i>	19q13.33	–	1	1	2
<i>SEC13</i> ⁽³⁾	3p25.3	–	2	–	2
<i>SMG5</i>	1q22	–	2	–	2
<i>EIF4A3</i>	17q25.3	–	1	–	1
<i>RUVBL1</i>	3q21.3	–	1	–	1
<i>SRSF1</i>	17q22	–	1	–	1

⁽¹⁾ Número de SNVs e CNVs localizados em regiões do gene que codificam domínios necessários para a função do mecanismo de NMD; ⁽²⁾ Número de indivíduos com PEA para os quais foram identificados SNVs e CNVs nos genes indicados; ⁽³⁾ Estes genes são fatores conservados do mecanismo de NMD mas ainda não se conhecem os domínios essenciais para a sua função.

SNVs, *Single Nucleotide Variants*; CNVs, *Copy Number Variants*.



_Conclusões

Este estudo encontrou evidência de um possível contributo do funcionamento anómalo do *Nonsense-Mediated mRNA Decay* (NMD) na perturbação do espectro do autismo (PEA).

Tendo um papel importante na regulação pós-transcricional, o NMD, se disfuncional, pode contribuir para a PEA através da desregulação da expressão dos seus genes-alvo. Por outro lado, pode ainda contribuir para a PEA, e para a sua heterogeneidade fenotípica, se a capacidade de degradação do mRNA mediada por mutações *nonsense* for afetada.

Uma melhor compreensão do mecanismo de NMD e das suas funções pode constituir uma oportunidade de desenvolvimento de intervenções terapêuticas para o tratamento e prevenção da PEA.

Referências bibliográficas:

- (1) American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th ed. Washington: APA, 2013. <http://dx.doi.org/10.1176/appi.books.9780890425596>
- (2) Wittmann J, Hol EM, Jäck HM. hUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated mRNA decay. *Mol Cell Biol*. 2006 Feb;26(4):1272-87. <https://doi.org/10.1128/MCB.26.4.1272-1287.2006>
- (3) Lou CH, Dumdie J, Goetz A, et al. Nonsense-Mediated RNA Decay Influences Human Embryonic Stem Cell Fate. *Stem Cell Reports*. 2016 Jun 14;6(6):844-57. <https://doi.org/10.1016/j.stemcr.2016.05.008>
- (4) Notaras M, Allen M, Longo F, et al. UPF2 leads to degradation of dendritically targeted mRNAs to regulate synaptic plasticity and cognitive function. *Mol Psychiatry*. 2020 Dec;25(12):3360-79. Epub 2019 Oct 21. <https://doi.org/10.1038/s41380-019-0547-5>
- (5) Jolly LA, Homan CC, Jacob R, et al. The UPF3B gene, implicated in intellectual disability, autism, ADHD and childhood onset schizophrenia regulates neural progenitor cell behaviour and neuronal outgrowth. *Hum Mol Genet*. 2013 Dec 1;22(23):4673-87. <https://doi.org/10.1093/hmg/ddt315>
- (6) Alrahbeni T, Sartor F, Anderson J, et al. Full UPF3B function is critical for neuronal differentiation of neural stem cells. *Mol Brain*. 2015 May 27;8:33. <https://doi.org/10.1186/s13041-015-0122-1>
- (7) Domingo D, Nawaz U, Corbett M, et al. A synonymous UPF3B variant causing a speech disorder implicates NMD as a regulator of neurodevelopmental disorder gene networks. *Hum Mol Genet*. 2020 Aug 29;29(15):2568-78. <https://doi.org/10.1093/hmg/ddaa151>
- (8) Kurosaki T, Imamachi N, Pröschel C, et al. Loss of the fragile X syndrome protein FMRP results in misregulation of nonsense-mediated mRNA decay. *Nat Cell Biol*. 2021 Jan;23(1):40-48. <https://doi.org/10.1038/s41556-020-00618-1>
- (9) Carbon S, Ireland A, Mungall CJ, et al. ; AmiGO Hub; Web Presence Working Group. AmiGO: online access to ontology and annotation data. *Bioinformatics*. 2009 Jan 15;25(2):288-9. Epub 2008 Nov 25. <https://doi.org/10.1093/bioinformatics/btn615>
- (10) Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome diagram viewer: data structures and strategies to boost performance. *Bioinformatics*. 2018 Apr 1;34(7):1208-14. <https://doi.org/10.1093/bioinformatics/btx752>
- (11) Buxbaum JD, Daly MJ, Devlin B, et al; Autism Sequencing Consortium. The autism sequencing consortium: large-scale, high-throughput sequencing in autism spectrum disorders. *Neuron*. 2012 Dec 20;76(6):1052-6. <https://doi.org/10.1016/j.neuron.2012.12.008>
- (12) Pinto D, Pagnamenta AT, Klei L, et al. Functional impact of global rare copy number variation in autism spectrum disorders. *Nature*. 2010 Jul 15;466(7304):368-72. <https://doi.org/10.1038/nature09146>
- (13) Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, et al. Multiple recurrent de novo CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron*. 2011 Jun 9;70(5):863-85. <https://doi.org/10.1016/j.neuron.2011.05.002>
- (14) Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nat Methods*. 2010 Apr;7(4):248-9. <https://doi.org/10.1038/nmeth0410-248>
- (15) Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073-81. <https://doi.org/10.1038/nprot.2009.86>
- (16) Sjöstedt E, Zhong W, Fagerberg L, et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. *Science*. 2020 Mar 6;367(6482):eaay5947. <https://doi.org/10.1126/science.aay5947>
- (17) Longman D, Jackson-Jones KA, Maslon MM, et al. Identification of a localized nonsense-mediated decay pathway at the endoplasmic reticulum. *Genes Dev*. 2020 Aug 1;34(15-16):1075-88. <https://doi.org/10.1101/gad.338061.120>