



Redes de interação proteica revelam fatores de risco associados à perturbação do espectro do autismo

Catarina Correia^{1,2,3}, Guiomar Oliveira^{4,5,6}, Astrid Moura Vicente^{1,2}

astrid.vicente@insa.min-saude.pt

(1) Departamento de Promoção da Saúde e Doenças não Transmissíveis, INSA.

(2) Center for Biodiversity, Functional & Integrative Genomics.

(3) Instituto Gulbenkian de Ciência.

(4) Unidade Neurodesenvolvimento e Autismo. Centro de Desenvolvimento. Hospital Pediátrico, Centro Hospitalar e Universitário de Coimbra.

(5) Centro de Investigação e Formação Clínica. Hospital Pediátrico, Centro Hospitalar e Universitário de Coimbra.

(6) Faculdade de Medicina, Universidade de Coimbra.

Introdução

A Perturbação do Espectro do Autismo (PEA) é uma patologia do neurodesenvolvimento caracterizada por problemas de socialização e por padrões de comportamento repetitivos e estereotipados (1). A sua apresentação clínica é muito heterogénea, em termos de gravidade, nível cognitivo e co-morbilidades presentes, tais como epilepsia, défice intelectual e défice de atenção e hiperatividade (2). Fatores genéticos contribuem substancialmente para o risco de PEA, sendo possível definir uma etiologia genética em cerca de 20% dos casos (3). No entanto, as causas genéticas conhecidas são também muito variáveis, incluindo alterações estruturais nos cromossomas (deleções, amplificações ou translocações) ou variantes na sequência de múltiplos genes (4). No geral, cada uma destas alterações genéticas é rara (presente em menos de 1% da população), embora em conjunto expliquem uma proporção significativa da variância no risco genético. Presume-se que, para os cerca de 80% dos doentes que são idiopáticos, existam outros fatores de risco, incluindo variantes genéticas comuns, fatores epigenéticos e ambiente. As bases neurobiológicas deste cenário complexo estão longe de completamente explicadas, impedindo até agora o desenvolvimento de terapia farmacológica que permita um tratamento eficaz da doença (4).

Em oposição às variantes raras, variantes genéticas comuns são alterações na sequência de genes que são frequentes na população geral e que contribuem individualmente com um pequeno efeito sobre o risco total de PEA (5). O seu efeito sobre o risco para esta

patologia é cumulativo, isto é, múltiplas variantes comuns têm que estar presentes, num dado paciente, para se atingir um limite patofisiológico a partir do qual surgem sintomas. Rastreios de associação genómica (*Genome Wide Association Studies*, GWAS), que têm como objetivo a identificação de variantes genéticas comuns associadas à PEA (4), têm tido um sucesso muito limitado. A dificuldade em obter resultados de associação coerentes em GWAS é consistente com a noção de que variantes genéticas comuns dificilmente podem ser detetadas em rastreios genómicos de associação, porque a dimensão das amostras populacionais até agora utilizadas não tem poder estatístico para tal (4,5). Estratégias alternativas são por isso necessárias para a identificação deste tipo de variantes, que permitam ultrapassar as limitações dos desenhos de estudo atuais.

Objetivo

No presente estudo foi colocada a hipótese de que fatores de risco comuns para a PEA convergem em vias fisiológicas específicas, e cumulativamente levam ao aparecimento de sintomas (6). Para identificar estas vias fisiológicas foi desenvolvido um método de análise de redes de interação proteína-proteína (*protein-protein interaction*, PPI), o qual envolve a sobreposição dos resultados de associação genómica obtidos em GWAS com redes de interação proteica previamente definidas. Esta abordagem pretende assim capturar informação de relevância biológica mesmo nos resultados negativos dos estudos de associação genómicos, e definir uma rede de interações proteicas (PPI) específica para a PEA.

Abordagem e resultados

A análise de redes de interação proteína-proteína foi aplicada ao conjunto de dados gerados num rastreio genómico para a PEA em que participámos (incluindo 2818 pacientes), no contexto do estudo internacional *Autism Genome Project* (AGP) (7). Os resultados foram confirmados com dados de um outro estudo levado a cabo pelo *Autism Genetic Resource Exchange* (AGRE) (8) (com 943 pacientes), disponíveis em base de dados para análise da comunidade científica. Os rastreios de associação genómica envolvem a genotipagem de marcadores polimórficos, designados *Single Nucleotide Polymorphisms* (SNPs), que constituem alterações de uma base dispersas por todo o genoma, intra- e intergenes, e cuja localização é

artigos breves_ n. 1

exatamente conhecida. Estas alterações de uma base são polimórficas na população humana, e constituem marcadores de posição no genoma. Tipicamente, em rastreios mais recentes, são genotipados cerca de 1 milhão de SNPs. É depois efetuado um teste de associação das frequências alélicas de cada SNP com a patologia em questão, utilizando grupos de doentes e de controlos, ou um formato de estudo baseado na frequência de transmissão alélica em famílias constituídas por paciente e ambos os pais (formato utilizado no AGP e AGRE).

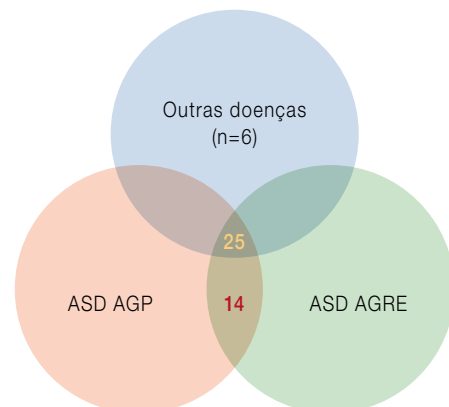
A primeira fase do presente estudo envolveu a ordenação crescente dos resultados dos testes de associação para cada SNP, obtidos no trabalho do AGP, até um limite (neste caso definimos $P < 0,1$) acima da significância estatística definida para rastreios genómicos ($P < 10^{-8}$). Todos os SNPs acima deste valor foram mapeados nos respetivos genes, e os genes convertidos nas proteínas que codificam. Este grupo de proteínas foi usado para a construção de uma rede de interações proteicas baseada na sobreposição dos resultados do rastreio genómico para a PEA com uma rede PPI humana pré-definida, incluindo 12372 proteínas e 58365 interações proteicas. Foram depois examinadas as propriedades topológicas da rede de interação proteica assim construída, indicativas de conectividade de vários níveis de associação genética, para confirmar a hipótese de que genes associados à PEA, a um nível normalmente considerado de "ruído estatístico", estão funcionalmente ligados para além da expectativa aleatória. Para tal, a rede PPI gerada para a PEA, com 416 proteínas, foi comparada com uma rede PPI de igual dimensão, construída a partir de grupos de proteínas selecionadas aleatoriamente da rede PPI humana pré-definida, em 1000 iterações.

As propriedades topológicas examinadas foram a percentagem de interações proteicas diretas, a percentagem de *nodes* isolados, e a dimensão da rede PPI formada. A análise topológica da rede proteica mostrou que as proteínas associadas à PEA, a níveis ($P < 0,1$) acima dos convencionalmente definidos como com significância estatística em GWAS, interagem diretamente mais do que a expectativa aleatória e formam redes de maiores dimensões e com menor número de *nodes* isolados. Estes resultados indicam que as proteínas definidas pelos rastreios genómicos, acima da significância estatística, estão envolvidas num número limitado de processos biológicos interligados, e estão relacionadas funcionalmente.

Verificou-se ainda que a precisão e especificidade com que os genes que codificam este grupo de proteínas são identificados numa lista de genes reconhecidamente envolvidos na PEA são significativamente superiores à do grupo de genes com maior evidência de associação (até $P < 0,1$) no GWAS, indicando que a rede proteica funcionalmente coerente gerada por esta abordagem contém mais informação biológica relevante para a PEA do que os resultados dos testes de associação.

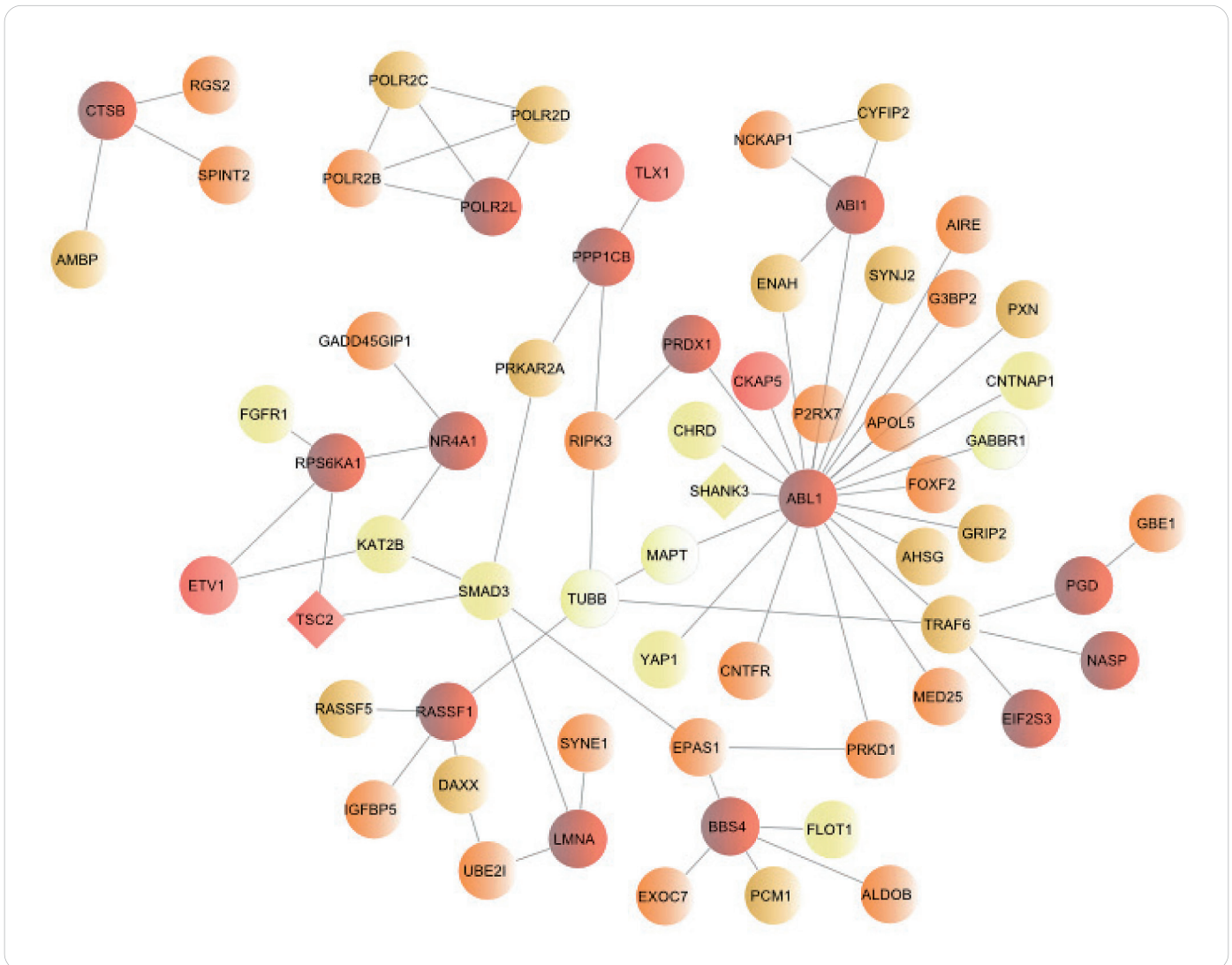
Finalmente, efetuou-se uma análise da intersecção entre as redes PPI obtidas a partir dos dois GWAS para a PEA analisados (AGP e AGRE), e seis conjuntos de dados de GWAS para doenças não relacionadas com a PEA (*Lupus Erythematosus*, Diabetes tipo 1, Neuroblastoma, Esclerose múltipla, Cancro da mama e Doença de Parkinson) (6), representada na **figura 1**. Esta análise identificou 14 proteínas exclusivamente presentes nas redes PPI para a PEA (**figura 2**). Os genes que codificam estas 14 proteínas são expressos no cérebro, e estão envolvidos em processos biológicos fundamentais anteriormente implicados na PEA, como a adesão celular, a organização do citoesqueleto e a migração dos neurónios durante a formação do cérebro.

Figura 1: Intersecção entre as redes obtidas a partir dos dois GWAS para PEA analisados (AGP e AGRE), e seis conjuntos de dados de GWAS para doenças não relacionadas com a PEA (*Lupus Erythematosus*, Diabetes tipo 1, Neuroblastoma, Esclerose múltipla, Cancro da mama e Doença de Parkinson).



artigos breves_ n. 1

Figura 2: Network de interações proteicas na Perturbação do Espectro do Autismo (PEA), com base nas 14 proteínas identificadas em comum entre os estudos AGP e AGRE, e as suas interações direta. Nodes são coloridos de acordo com um esquema de cores baseado num sistema de pontuação que reflete a sua presença nos dois conjuntos de dados de PEA (AGP e AGRE) e ausência nas 6 patologias não relacionadas com PEA; uma coloração mais forte é indicativa de uma pontuação mais elevada, isto é, maior especificidade.



Para vários dos genes que codificam estas proteínas foram anteriormente reportadas alterações em pacientes com PEA: mutações identificadas por sequenciação exómica, incluindo 50 pacientes da nossa amostra populacional (genes *PGD*, *NASP*, *LMNA*, *PPP1CB*, *RASSF1*, *ABL1*, *ABI1*, *BBS4*), genes deletados ou amplificados em

Copy Number Variants (genes *ABL1*, *RPS6KA1*, *PPP1CB*, *NASP*), alterações de expressão em tecido *post-mortem* ou linfoblastos (genes *NASP*, *NR4A1*, *ABI1*, *BBS4*, *LMNA*, *ABL1*) e fenótipos anómalos no sistema nervoso central em modelos animais (genes *CTSB*, *BBS4*, *LMNA* e *ABL1*).

_Conclusão

Em conclusão, a análise de redes de interações proteína-proteína aplicada a dados de rastreios genómicos para a PEA confirmou resultados de estudos por outras abordagens, e destacou novos fatores de risco para esta patologia, envolvidos no funcionamento do sistema nervoso central e com significado biológico para esta patologia. Nos rastreios genómicos analisados, estes genes encontravam-se ao nível do “ruído estatístico”, e não foram valorizados. Sequenciação dos genes agora identificados está em curso num grupo de pacientes com PEA, como *proof-of-concept* experimental desta nova metodologia de análise. Este trabalho releva o valor de estratégias alternativas que permitam uma análise mais profunda da informação gerada em rastreios genómicos de associação, nomeadamente através da exploração de redes de interações proteicas.

Referências bibliográficas:

- (1) Lord C, Jones RM. Annual research review: re-thinking the classification of autism spectrum disorders. *J Child Psychol Psychiatry*. 2012;53(5):490-509. [LINK](#)
- (2) Fakhoury M. Autistic spectrum disorders: a review of clinical features, theories and diagnosis. *Int J Dev Neurosci*. 2015;43:70-7
- (3) Betancur C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res*. 2011;1380:42-77. Epub 2010 Dec 1.
- (4) Geschwind DH. Genetics of autism spectrum disorders. *Trends Cogn Sci*. 2011;15(9):409-16. [LINK](#)
- (5) Anney R, Klei L, Pinto D, et al. Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum Mol Genet*. 2012;21(21):4781-92. [LINK](#)
- (6) Correia C, Oliveira G, Vicente AM. Protein interaction networks reveal novel autism risk genes within GWAS statistical noise. *PLoS One*. 2014;9(11):e112399. [LINK](#)
- (7) Anney R, Klei L, Pinto D, et al. A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet*. 2010;19(20):4072-82. [LINK](#)
- (8) Wang K, Zhang H, Ma D, et al. Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature*. 2009;459(7246):528-33. [LINK](#)