

Genômica comparada de genes candidatos do transtorno depressivo maior em humanos e em modelos animais

Bolsista de Iniciação Científica: Álvaro de Oliveira Franco¹
Orientadora: Maria Cátira Bortolini¹

¹ Laboratório de Evolução Humana e Molecular (LEHM) – Universidade Federal do Rio Grande do Sul (UFRGS)

INTRODUÇÃO

Atualmente, existem hipóteses a respeito da fisiopatologia do transtorno depressivo maior (TDM) e dos mecanismos subjacentes à sua emergência, mas a complexa natureza do TDM traz muitos desafios a serem superados antes do estabelecimento de um panorama neuromolecular mais geral. Isso é refletido nos números de tratamentos refratários relacionados à essa condição e no problemático manejo dos pacientes.

Fatores genéticos representam um importante papel na patogênese do TDM, e alguns genes já foram fortemente associados com a doença. Entretanto, evidências mostram que genes coexpressos, coregulados e cofuncionais podem ser mantidos topologicamente próximos entre si devido às pressões seletivas, ocupando uma mesma vizinhança gênica, como um bloco [1; 2]. Por conseguinte, é possível inferir efeitos modulatórios entre genes de um bloco conservado, com consequências sobre os respectivos fenótipos aos quais são relacionados. Nós hipotetizamos, portanto, que a comparação de algum fenótipo comportamental entre espécies pode ser feita de forma mais acurada quando são levados em consideração os aspectos sintênicos e de identidade proteica dos genes analisados.

Neste estudo, nós investigamos, através do uso de bioinformática aplicada em genômica comparativa, qual dos sete animais usados como modelos em pesquisa experimental assemelham-se mais aos humanos com relação às identidades proteicas de seis genes associados ao TDM e às suas respectivas vizinhanças. Deste modo, desejamos verificar se os murinos, estabelecidos por tradição como animais modelos, são os mais adequados para mimetizar o fenótipo de TDM em estudos experimentais.

OBJETIVO

Verificar qual espécie é mais similar ao perfil genético humano de seis genes associados com o transtorno depressivo maior usando como parâmetros a informação de suas sequências proteicas, bem como a informação sobre a vizinhança desses respectivos genes.

MÉTODOS

Identificação e mapeamento dos genes e vizinhanças: nós utilizamos a base de dados *Mouse Genome Informatics (MGI)* [3] e informações de *Genome-Wide Analysis (GWAS)* [4] para selecionar genes candidatos fortemente associados com o Transtorno Depressivo Maior em humanos e que já estão sendo usados em estudos experimentais com modelos murinos – *DISC1*, *FKBP5*, *HAP1*, *HTR2A*, *SLC6A15* e *TPH2*. Desenvolvemos um modelo matemático para quantificar a semelhança entre humanos e sete animais modelos de pesquisa mais três grandes símios (para propósito de controle positivo), levando em consideração tanto a identidade de aminoácidos quanto os aspectos sintênicos dos seis genes candidatos para TDM. Nós mapeamos a vizinhança desses genes (definidos como 15 genes *down* e *upstream* dos genes candidatos) em humanos e nas seguintes espécies: chimpanzé (*Pan troglodytes*), gorila (*Gorilla gorilla*), orangotango (*Pongo abelii*), macaco rhesus (*Macaca mulatta*), sagui (*Callithrix jacchus*), camundongo (*Mus musculus*), rato (*Rattus norvegicus*), coelho (*Oryctolagus cuniculus*), cachorro (*Canis lupus familiaris*) e galo (*Gallus gallus*). O mapeamento da vizinhança foi realizado através do uso do Genomicus v88.01 [5] e Ensembl [6]. Sequências genéticas foram obtidas através das bases de dados do NCBI [7] e do Ensembl. Alinhamentos foram realizados através do uso do MEGA7 [8] e GUIDANCE [9].

Escore identidade-vizinhança e análise K-means: mapas de vizinhança foram usados para gerar um *escore de vizinhança (EV)* para cada gene respectivo a cada espécie, baseado na sua similaridade com o bloco gênico humano de referência. A presença e a ausência dos ortólogos humanos nos blocos gênicos das espécies animais foram os componentes para gerar tais EVs, bem como a ocorrência de duplicações, inserções e realocações. A partir do alinhamento pareado das sequências de aminoácidos obtidas em bancos de dados, nós obtivemos, pelo uso do LALIGN [8], as *identidades proteicas (IPs)* do conjunto de seis genes candidatos nas espécies estudadas. As IPs foram multiplicadas pelos respectivos EVs objetivando normalizá-los pela filogenia e também levar em conta as diferenças na sequência de aminoácidos, gerando um *escore identidade-vizinhança (EIV)*. EIV foi então utilizado como parâmetro para uma análise de *cluster K-means* ao aplicarmos uma matriz de distância euclidiana, implementado em pacote *pvclust* em linguagem R, que nos permitiu gerar uma árvore baseada na similaridade dos animais selecionados quando ambos aspectos sintênicos e sequências proteicas são analisadas.

REFERÊNCIAS

- [1] Hurst LD, Pál C and Lercher MJ. The evolutionary dynamics of eukaryotic gene order. *Nat Rev Genet* 2004; 5(4):299-310. (doi: 10.1038/nrg1319)
- [2] Michalak P. Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. *Genomics* 2008; 91(3):243-8. (doi: 10.1016/j.ygeno.2007.11.002)
- [3] <http://www.informatics.jax.org>
- [4] Wray NR, Pergadia ML, Blackwood DHR et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol Psychiatry* [Internet] 2012; 17(1):36-48. (doi:10.1038/mp.2010.109)

RESULTADOS

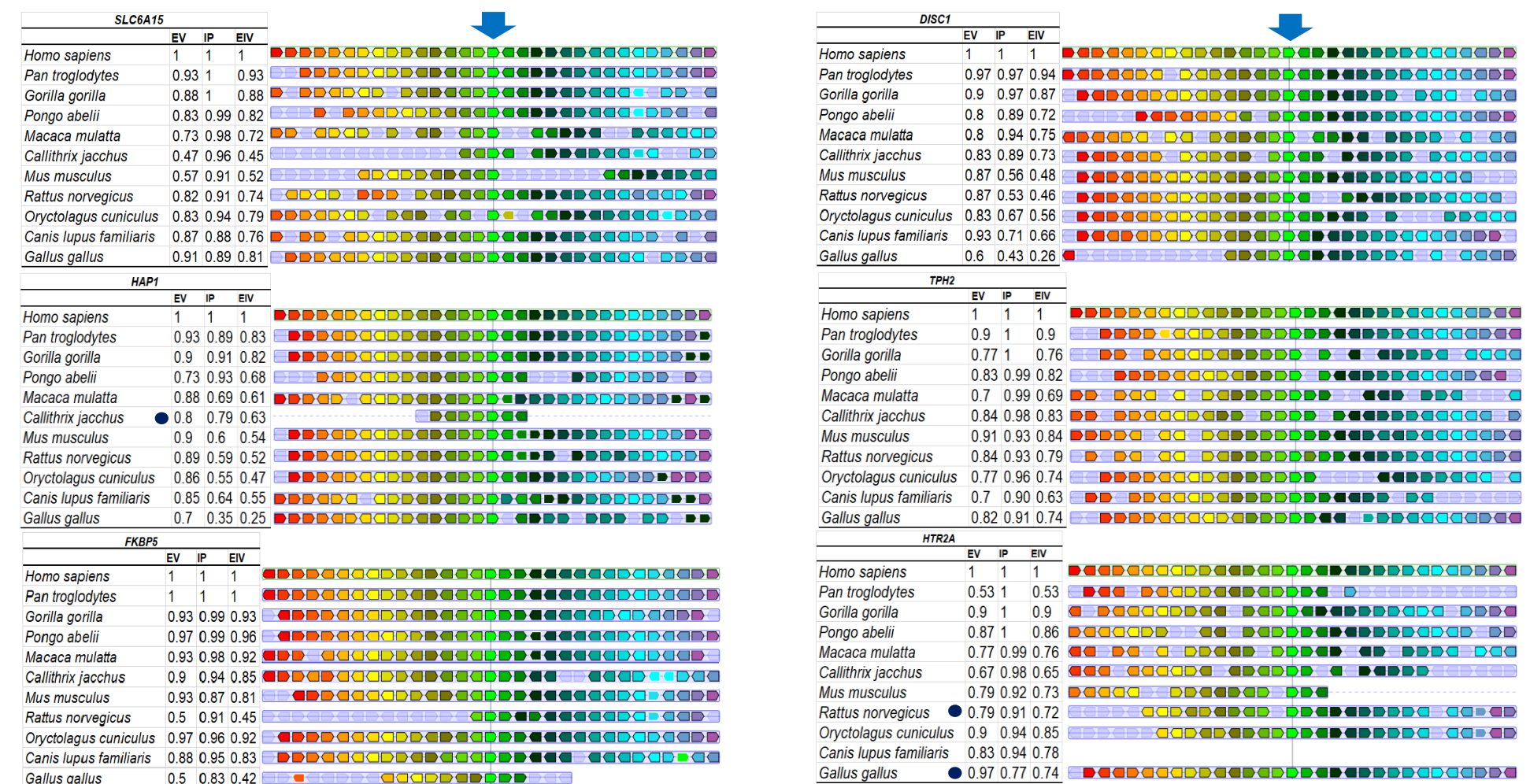


Figura 1: tabelas dos escores mencionados para cada espécie. Cada gene é representado por um polígono. Polígonos acromáticos representam genes não ortólogos daqueles encontrados na vizinhança dos genes candidatos humanos. O diagrama da imagem foi retirado do Genomicus v88.01 e representa os genes candidatos (polígonos verdes indicados pela seta azul) e sua vizinhança (15 genes *down* e *upstream*). As espécies marcadas com o círculo azul tiveram seu mapa de vizinhança completado com dados do Ensembl. EV = Escore de vizinhança. IP = Identidade proteica. EIV = Escore identidade-vizinhança.

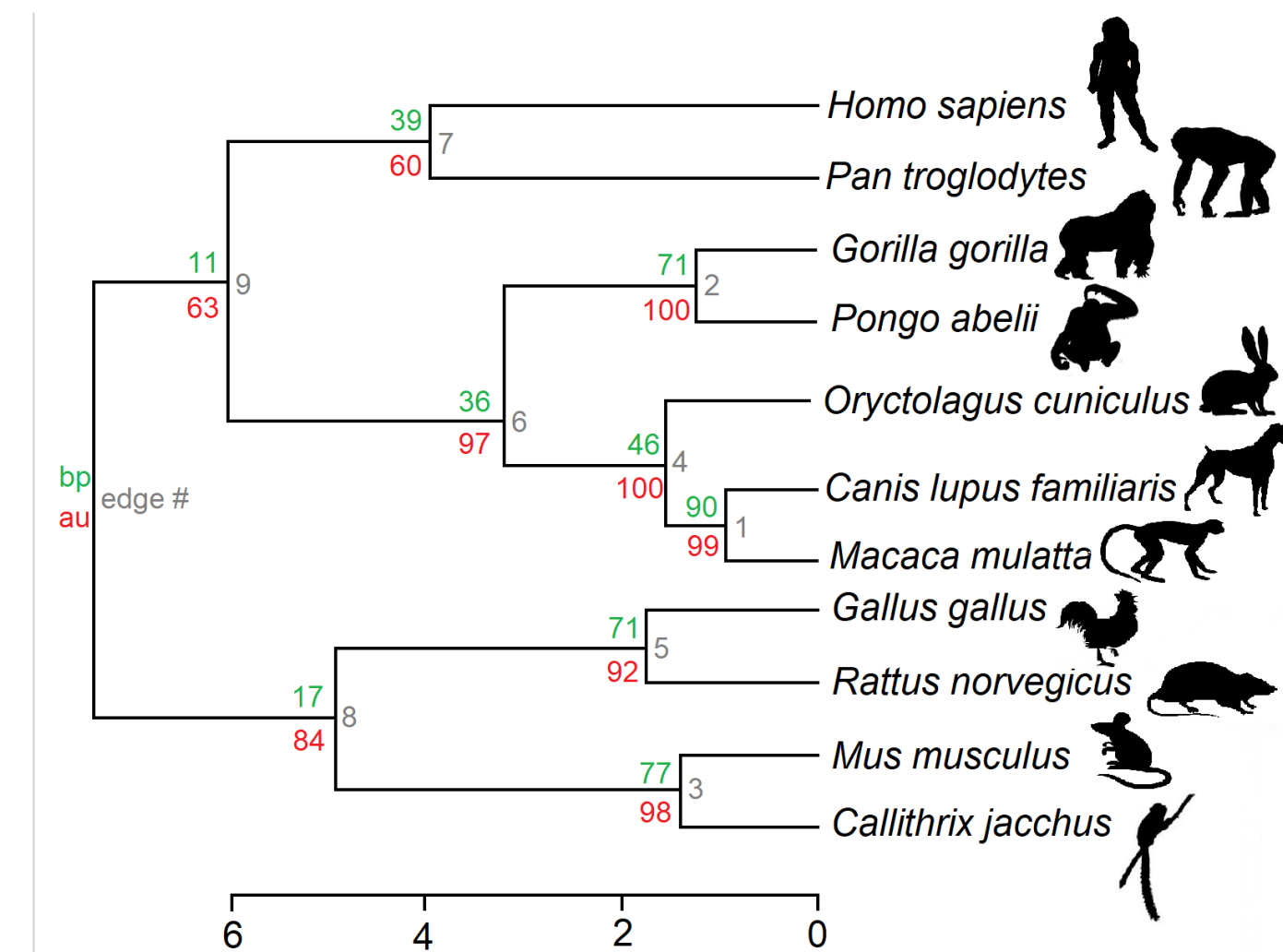


Figura 2: pelo uso da análise K-means em linguagem R, os dados mostrados nas tabelas da figura 1 foram usados como base para a criação da árvore exibida acima. Os números verdes e vermelhos indicam os valores de *bootstrap* (bp) e de *approximately unbiased* (au), respectivamente.

CONCLUSÃO

- Os grandes símios usados como controle são as espécies mais similares aos humanos em relação à sequência de aminoácidos e à vizinhança dos seis genes candidatos para TDM selecionados, sugerindo que o método por nós utilizado segue corretamente a filogenia consenso. Entretanto, eles cada vez menos estão sendo usados como modelos de experimentação animal devido a questões éticas.
- O coelho é a espécie mais similar aos humanos em relação aos parâmetros estabelecidos, quando excluídos os grandes símios. Seguindo a premissa utilizada – de que vizinhanças genéticas possuem impacto na expressão e regulação de seus genes abrangidos e, portanto, no fenótipo relacionado aos respectivos genes –, o coelho parece ser mais adequado para modelos de TDM do que modelos murinos tradicionais.
- Mais estudos são necessários para elucidar a importância dos blocos de vizinhanças gênicas nos fenótipos dos genes que as compõe e para comprovar experimentalmente a validade preditiva deste estudo.

- [5] <http://www.genomicus.biologie.ens.fr>
- [6] <http://www.ensembl.org/index.html>
- [7] <http://www.ncbi.nlm.nih.gov>
- [8] Kumar S, Stecher G, and Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016; 33(7):1870-4. (doi: 10.1093/molbev/msw054)
- [9] <http://guidance.tau.ac.il/ver2/>
- [10] <http://embnet.vital-it.ch>