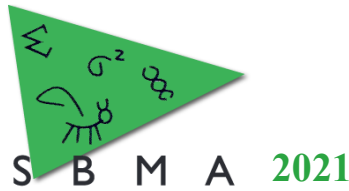


XIV Simpósio Brasileiro de Melhoramento Animal

18 a 19 de Outubro de 2021
On-line





Identificação e caracterização de variantes funcionais no transcriptoma de duodeno em poedeiras

Adriana Mércia Guaratini Ibelli^{1,2*}; Letícia Alves Salmória²; Fernando de Castro Tavernari^{2,3}; Jane de Oliveira Peixoto^{1,2}; Débora Ester Petry Marcelino⁴; Mariane Spudeit Dal Pizzol³; Maurício Egídio Cantão²; Mônica Corrêa Ledur^{2,3}.

¹ Embrapa Suínos e Aves, Concórdia, SC, Brasil.

² Programa de Pós-Graduação em Ciências Veterinárias, Universidade Estadual do Centro Oeste do Paraná, Guarapuava, PR, Brasil.

³ Programa de Pós-Graduação em Zootecnia, UDESC-Oeste, Chapecó, SC, Brasil.

⁴ Faculdade de Concórdia FACC, Concórdia, SC, Brasil.

*Autor correspondente: adriana.ibelli@embrapa.br

Resumo: O aumento no número de trabalhos de sequenciamento de RNA (RNA-Seq) tem permitido que estes dados sejam utilizados em abordagens além da análise de expressão gênica diferencial. Entre elas, pode-se citar a identificação de variantes funcionais baseadas em dados de RNA-Seq que tem se tornado uma alternativa para prospectar polimorfismos em diversas espécies. Sabendo que esta estratégia é pouco utilizada nos estudos com galinhas, o objetivo deste trabalho foi identificar e caracterizar variantes funcionais no transcriptoma de duodeno de galinhas poedeiras. Para isso, dados de transcriptoma de duodeno de galinhas foram utilizados para a detecção de variantes utilizando o software GATK. Em seguida, as variantes identificadas foram anotadas com a ferramenta *Variant Effect Predictor* (VEP) do Ensembl. Foram identificadas 119.046 variantes nos transcriptomas avaliados, sendo que 106.644 já haviam sido descritas (89,6%) e 12.402 (10,4%) foram descritas neste trabalho. Entre as novas variantes encontradas, podem-se citar mutações *missense* com efeitos deletérios nos genes *TNFRSF10B*, *DHX8*, *NRDC* e *SSPN* que estão em regiões de QTL para diversas características de interesse para a avicultura. Estes dados constituem uma fonte de conhecimento da variabilidade genética desta espécie, contribuindo para um melhor entendimento da genômica funcional de aves.

Palavras-chave: polimorfismos, galinhas, SNPs, InDels

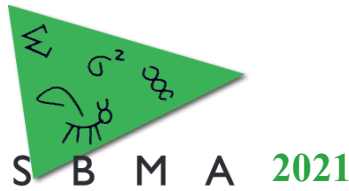
Identification and characterization of functional variants in laying hens' duodenum transcriptome

Abstract: The increase in the RNA sequencing (RNA-Seq) studies has allowed the use of these datasets in new approaches other than differentially expression analysis. One of them is the functional variants identification that has become an alternative way to prospect polymorphisms in several species. Since there are few studies using RNA-seq to find variants in chickens, this study aimed to identify and to characterize the presence of functional variants in the laying hens' duodenum transcriptome. To this, chicken duodenal transcriptome data were used for variant detection using GATK software. The identified variants were annotated using Ensembl's Variant Effect Predictor (VEP) tool. A total of 119,046 variants were found in the evaluated transcriptomes, where 106,644 were existing (89.6%) and 12,402 (10.4%) were novel variants. Among them, it is possible to highlight some missense mutations with deleterious effects predicted in the *TNFRSF10B*, *DHX8*, *NRDC* and *SSPN* genes that are in QTL regions for several traits of interest to poultry industry. Our data constitute a source of knowledge in the chicken genetic variability, contributing to a better understanding on its functional genomics.

Keywords: polymorphisms, chickens, SNPs, InDels

Introdução

O aumento no número de trabalhos de sequenciamento de RNA (RNA-Seq) tem permitido que estes dados sejam utilizados em abordagens além da análise de expressão gênica diferencial. Entre elas, pode-se citar a identificação de variantes funcionais baseadas em dados de RNA-Seq que tem se tornado uma alternativa em diversas espécies, como humanos, animais modelo (Piskol et al., 2013) e também em animais de produção (Bakhtiarizadeh et al., 2020). É importante ressaltar que os dados de RNA-Seq podem apresentar maior variabilidade devido a natureza do RNA, de modo que há a necessidade de aplicar filtros mais conservadores nas análises destes dados para evitar a identificação de variantes falso-positivas (Jehl et al., 2021). No entanto, tem se observado uma concordância de acima de 91% na detecção de polimorfismos quando se compara sequenciamento de RNA e DNA (Piskol et al., 2013; Jehl



et al., 2021). Em aves, há poucos estudos utilizando este tipo de abordagem, que pode ser útil tanto para a detecção de variantes funcionais e causais associadas a fenótipos de interesse, como na identificação de novos polimorfismos, o que permite um avanço no conhecimento do genoma da galinha. Desta forma, o objetivo deste trabalho foi identificar e caracterizar a presença de variantes funcionais no transcriptoma de duodeno de galinhas poedeiras.

Material e Métodos

Para este trabalho foram utilizadas 27 amostras de duodeno de galinhas poedeiras comerciais com aproximadamente 70 semanas de idade alojadas na Embrapa Suínos e Aves. As amostras de duodeno foram coletadas, submetidas a extração de RNA total utilizando o reagente Trizol (Invitrogen) e por purificação com coluna de sílica seguindo o protocolo do kit Qiagen RNeasy (Qiagen). O RNA total foi quantificado em equipamento Biodrop e a integridade foi avaliada em Bioanalyzer (Agilent 2012), sendo consideradas para as análises posteriores aquelas com RIN ≥ 8 . As bibliotecas de RNA-Seq foram preparadas com o kit Illumina Stranded (Illumina) e enviadas para sequenciamento em equipamento HiSeq 2500 (Illumina), seguindo o protocolo paired-end (2x100 pb). Após o recebimento das sequências, estas foram submetidas ao controle de qualidade usando o Trimmomatic e mapeadas contra o genoma referência da galinha (GRCg6a) com o software STAR no protocolo *two-pass mode*. A identificação dos SNPs foi realizada no programa GATK v. 3.8 e os filtros SNPcluster considerando 3 variantes em uma janela de 35bp, FS > 30.0, QD < 5.0, MQ < 50.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, GQ < 5.0, QUAL ≥ 30.0 e DP ≥ 100.0 foram utilizados para minimizar a identificação de variantes falso-positivas. Estes dados foram submetidos a ferramenta *Variant Effect Predictor* (VEP) do Ensembl na versão 104 para caracterização das variantes. A anotação funcional das variantes foi realizada na base de dados do Panther (<http://www.pantherdb.org/>) e no MSigDB (Subramanian et al., 2005). A ferramenta Revigo (Supek et al., 2005) foi utilizada para reduzir as ontologias gênicas.

Resultados e Discussão

A identificação de polimorfismos a partir de RNA-seq tornou-se uma alternativa para a descoberta de variantes associadas a fenótipos de interesse (Piskol et al., 2013). Foram identificadas 119.046 variantes nos transcriptomas avaliados, sendo que 106.644 já haviam sido descritas (89,6%) e 12.402 (10,4%) foram identificadas neste trabalho. Considerando o efeito de todas as variantes, 43% localizaram-se em éxons, 21% em região 3'UTR, 17% em regiões intergênicas, 15% em introns, 2% em regiões 5'UTR e 2% em transcritos não codificantes. daquelas encontradas em regiões codificantes, 79% foram sinônimas, 20% *missense* e 1% foram deleção, inserção, *frameshift* entre outras. De acordo com o VEP, 9750 variantes foram caracterizadas como de alto e moderado impacto, ou seja, que podem alterar a produção de proteínas, estando presentes em 3626 genes, envolvidos em 135 processos biológicos como adesão celular, regulação celular ao estresse, de atividade enzimática entre outros. No MSigDB, que utiliza informações de humanos, foi possível observar que, em sua maioria, estes genes encontram-se em famílias gênicas de fatores de transcrição, de marcação de diferenciação celular, oncogenes, citocinas e fatores de transcrição. Dentre as novas variantes encontradas, podem-se citar algumas classificadas como *missense*: 1) 1 SNP deletério (SIFT = 0) no gene *TNFRSF10B*, envolvido com resposta imune e que está localizado em uma região de QTL para resistência a vírus em galinhas; 2) 1 SNP deletério (SIFT = 0,04) no gene *DHX8*, que regula a liberação de mRNAs no spliceossomo, localizado em uma região de QTL para conteúdo mineral, peso e tamanho do fêmur, complemento do intestino, gordura abdominal, entre outros; 3) 1 SNP tolerante (SIFT = 1) no gene *NRDC* localizado em uma região de QTLs para peso de ovário e comprimento do duodeno; 4) 1 SNP deletério no gene *SSPN* (SIFT = 0.03), que faz a ligação entre citoesqueleto e matriz extracelular de células de matriz celular e também está em uma região de QTLs para gordura intramuscular, peso corporal, peso do ovo e ovário entre outros. Desta forma, neste trabalho, foi possível identificar novos polimorfismos em regiões de QTL e que poderão ser validados como marcadores genéticos associados a características de interesse. Estes dados constituem uma fonte de conhecimento da variabilidade genética desta espécie, contribuindo para um melhor entendimento da genômica funcional de aves.



Conclusão

Foram identificadas variantes funcionais no transcriptoma no duodeno de galinhas poedeiras, sendo aproximadamente 10% delas descritas pela primeira vez neste trabalho. Entre elas, destacam-se mutações *missense* com efeitos deletérios nos genes *TNFRSF10B*, *DHX8*, *NRDC* e *SSPN* que estão em regiões de QTL para diversas características de interesse para a avicultura.

Agradecimentos

Este estudo foi financiado pelo projeto n° 13.16.04.005.00.00 da EMBRAPA. Os autores LAS e DEPM agradecem a Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e CNPq/PIBIC na Embrapa Suínos e Aves, respectivamente, pela concessão de bolsa de estudos. MCL e FCT são bolsistas de produtividade de pesquisa do CNPq. O presente trabalho foi realizado com apoio da CAPES, Código de Financiamento 001.

Literatura citada

- Piskol R, Ramaswami G, Li JB. 2013. Reliable identification of genomic variants from RNA-seq data. **Am J Hum Genet.**,93, 641-51.
- Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., Paulovich A., Pomeroy S.L., Golub T.R., Lander E.S., Mesirov J.P. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. **Proc Natl Acad Sci USA.** 102, 15545-50.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. **PLoS One.**;6:e21800.
- Bakhtiarzadeh MR, Alamouti AA. 2020. RNA-Seq based genetic variant discovery provides new insights into controlling fat deposition in the tail of sheep. **Sci Rep.**10:13525.
- Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Désert C, Coulée M, Bouchez O, Leroux S, Abasht B, Tixier-Boichard M, Bed'hom B, Burlot T, Gourichon D, Bardou P, Acloque H, Foissac S, Djebali S, Giuffra E, Zerjal T, Pitel F, Klopp C, Lagarrigue S. 2021. RNA-Seq Data for Reliable SNP Detection and Genotype Calling: Interest for Coding Variant Characterization and Cis-Regulation Analysis by Allele-Specific Expression in Livestock Species. **Front Genet.** 12:655707.