

DOI: 10.22034/AS.2021.46637.1621

مطالعه‌ی پویش کل ژنوم برپایه‌ی تجزیه و تحلیل غنی‌سازی مجموعه‌ی ژنی مرتبط با صفات رشد در لاین اینترکراس پیشرفته مرغ

حسین محمدی^{۱*}، امیر حسین خلت آبادی فراهانی^۲، محمد حسین مرادی^۲ و ایمان حاج خدادادی^۲

تاریخ دریافت: ۱۴۰۰/۳/۳۱ تاریخ پذیرش: ۱۴۰۰/۷/۴

^۱استادیار گروه علوم دامی، دانشکده کشاورزی و محیط زیست، دانشگاه اراک

^۲دانشیار گروه علوم دامی، دانشکده کشاورزی و محیط زیست، دانشگاه اراک

*مسئول مکاتبه: Email: H-mohammadi64@araku.ac.ir

چکیده

زمینه مطالعه‌ی: درک نحوه کنترل ژنتیکی صفات رشد برای اصلاح نژاد موثر طیور ضروری است. هدف: پژوهش حاضر، به منظور مطالعه‌ی پویش کل ژنوم بر مبنای تجزیه و تحلیل غنی‌سازی مجموعه‌ی ژنی جهت شناسایی جایگاه‌های ژنی مؤثر بر برخی صفات مرتبط با وزن بدن و طول و قطر شانک در یک لاین اینترکراس پیشرفته (AIL) با استفاده از روش ارزیابی ژنوتیپی با توالی‌یابی اجرا گردید. روش کار: به منظور مطالعه‌ی پویش کل ژنوم از ۵۹۹ قطعه مرغ و رکورد فنوتیپی مرتبط با صفات رشد شامل وزن بدن از هج تا ۱۴ هفتگی و طول و قطر شانک از ۴ تا ۱۲ هفتگی استفاده شد. آنالیز پویش کل ژنوم در برنامه‌ی GCTA نسخه ۱/۹۲ انجام شد. در مرحله‌ی بعد، آنالیز غنی‌سازی ژنی با استفاده از بسته‌ی نرم افزاری goseq برنامه R و هدف شناسایی عملکرد بیولوژیکی ژن‌های نزدیک در مناطق انتخابی کاندیدا از طریق پایگاه‌های برخط GO، Metacyc، KEGG، Reactome و Panther انجام گردید. نتایج: تجزیه و تحلیل بیوانفورماتیکی نشان داد که مناطق ژنومی شناسایی شده به طور مستقیم و غیر مستقیم با ژن‌های مؤثر بر رشد عضلات اسکلتی، وزن بدن، متابولیسم انرژی و رشد و توسعه‌ی استخوان همپوشانی دارند. در این پژوهش نشانگر تک نوکلئوتیدی معنی‌داری واقع روی کروموزوم‌های ۱، ۲، ۴، ۵، ۷، ۸، ۱۰، ۱۱ و ۲۷ شناسایی شدند که با ژن‌های *IGF1 ANXA2 PNPLA3 CAPN3 MSTN EDN1 MC4R TMEM135 FN1 LEPR LDB2 ADAMTS18* و مرتبط بودند. در آنالیز غنی‌سازی مجموعه‌ی ژنی تعداد ۱۹ مسیر هستی‌شناسی ژنی و مسیر KEGG با صفات وزن بدن و طول و قطر شانک مرتبط بودند ($P < 0.05$). این بین، مسیرهای Regulation of muscle organ development، Regulation of cell growth و Anatomical structure و homeostasis نقش مهمی در رشد و توسعه‌ی عضلات اسکلتی داشتند. همچنین در ارتباط با طول و قطر شانک مسیرهای Positive regulation of ossification، Positive regulation of cytosolic calcium ion و Calcium signaling pathway ارتباط معنی‌داری داشتند. نتیجه‌گیری نهایی: با توجه به تأیید مناطق قبلی پویش ژنومی صفات وزن بدن و طول و قطر شانک، همچنین شناسایی مناطق ژنومی جدید، آنالیز غنی‌سازی مجموعه‌ی ژنی درک بهتری از کنترل ژنتیکی صفات رشد را نشان می‌دهد. استفاده از نتایج این تحقیق می‌تواند سبب تسریع در پیشرفت ژنتیکی برنامه‌های اصلاح نژادی مرغ شود.

واژگان کلیدی: آنالیز غنی‌سازی، پویش ژنوم، ژن کاندیدا، مرغ، وزن بدن

مقدمه

گوشتی است که باعث کوتاه‌تر شدن دوره پرورش آنها شده است (محمدی و صادقی ۲۰۱۰). امروزه با توسعه تراشه‌های ژنومی و امکان مطالعات ژنومی می‌تواند باعث

برنامه‌های اصلاحی تجاری برای افزایش سرعت رشد به طور مداوم در حال انتخاب روی دام‌ها و جوجه‌های

افزایش پیشرفت ژنتیکی گردد. هدف از مطالعات پویش ژنومی که به منظور شناسایی ارتباط بین یک نشانگر SNP و یک صفت با استفاده از نشانگرهای با تراکم بالا در سطح ژنوم می‌باشد، پیدا کردن جهش‌های با اثر مستقیم بر فنوتیپ یک صفت تولیدی، تولیدمثلی و یا بیماری می‌باشد. این اطلاعات می‌تواند برای انتخاب به کمک نشانگر مفید بوده و به درک مکانیسم مولکولی صفات مورد مطالعه کمک نماید.

داده‌های مورد استفاده در این تحقیق اولین بار توسط وانگ و همکاران (۲۰۲۰) برای صفات مرتبط با وزن بدن از هفته اول تا ۱۴ هفتگی، طول و قطر شانک از ۴ تا ۱۲ هفتگی مورد آنالیز قرار گرفته بودند که در آن تحقیق با استفاده مدل‌های خطی مختلط تک نشانگری و هاپلوتیپ تجزیه شده و از آزمون بنفرونی برای تعیین آستانه معنی‌داری و جلوگیری از خطای نوع اول استفاده شده بود. ژن‌های *ZNF652*, *JGF2BP1*, *PHOSPHO1* و *GIP* به عنوان ژن‌های مرتبط با صفات رشد گزارش شده بودند.

مطابق یافته‌های مطالعه مدنظر و مقالات مختلف در مطالعات پویش ژنومی مبتنی بر مدل‌های خطی و غیرخطی، مشکل بالا بودن نرخ خطای نوع اول و بیش برآورد اثر نشانگرهای SNPها می‌باشد. به عبارت دیگر یکی از ایرادات تحقیقات مطالعات پویش ژنومی در نظر گرفتن آستانه معنی‌داری برای جلوگیری از بروز خطای نوع اول است. در حالی که پرهیز از خطای نوع اول سبب افزایش خطای نوع دوم یعنی در نظر نگرفتن SNPهای دارای اثر معنی‌دار پایین‌تر از آستانه در نظر گرفته شده می‌شود (عزیز پور و همکاران ۲۰۲۰).

در نتیجه یک جایگزین مناسب برای حل این مشکل، انجام مطالعات پویش کل ژنومی بر مبنای مسیر^۱ و با استفاده از تجزیه و تحلیل مجموعه‌های ژنی است. در واقع در این روش به جای انجام تجزیه برای یک SNP یا یک ژن، ارتباط بین صفت مورد مطالعه و واریانت‌های ژنتیکی را در یک دسته یا گروه ژنی که به طور عملکردی با هم مرتبط هستند بررسی می‌کند. به عبارتی دیگر آنالیز

پیوستگی بین یک مجموعه ژنی معنی‌دار زیستی با فنوتیپ مورد آزمون قرار می‌گیرد (وانگ و همکاران ۲۰۱۱). در حقیقت در این روش به دنبال ژن‌هایی هستیم که به تنهایی اثر آن‌ها بر صفت مورد نظر معنی‌دار نشده و دارای اثرات متوسطی می‌باشند، ولی اثر تجمعی آنها روی صفت دارای اثر معنی‌دار است. برای این که بتوان تفسیر درستی از کنار هم قرار دادن این ژن‌ها حاصل شود، از مسیرهای زیستی به عنوان بستری معنی‌دار که عملکرد مجموع ژن‌ها در آن‌ها یک یا چند هدف واحد را پیگیری می‌کند استفاده می‌شود (مونی و ویلموت ۲۰۱۵). برای اولین بار پناگریکانو و همکاران (۲۰۱۳) نشان دادند که آنالیز پویش ژنومی بر مبنای مسیر دقت شناسایی مناطق ژنومی مؤثر بر صفت نرخ باروری گاوهای نر را بالا برده است، زیرا با استفاده از این روش تمامی نشانگرهای معنی‌دار در سطح ۰/۰۵ آنالیز می‌شوند و در نتیجه میزان خطای نوع اول و بیش برآوردها کاهش پیدا می‌کند. در تحقیق دیگر مطالعه پویش ژنومی بر مبنای مسیر با استفاده از آنالیز غنی شدن مجموعه‌های ژنی روی خصوصیات مرتبط با تولید شیر و امتیاز سلول‌های بدنی گاو هلشتاین انجام شد. نتایج آنالیز براساس مسیر، منجر به شناسایی ۲۵ طبقه مختلف عملکردی هستی شناسی ژن و مسیر بیوشیمیایی در KEGG^۲ شد که شامل مسیرهای sensory perception of chemical، positive regulation of inflammatory stimulus، defense response و نقش مهمی در سیستم ایمنی و مرتبط با امتیاز سلول‌های بدنی داشتند (عزیز پور و همکاران ۲۰۲۰). اخیراً مطالعه پویش کل ژنومی برپایه آنالیز مسیر جهت شناسایی جایگاه‌های ژنی مؤثر بر وزن بدن در چهار نژاد مرغ چوآ، سیلک، لنگشن و بیرد انجام شده است و مسیرهای cytoskeletal anatomical structure development، protein binding و Tricarboxylic acid cycle که شامل ژن‌های کاندیدای *MYO1B*, *MYH11*, *MYH10*, *MYOD1*, *ABCG1*, *SLC2A8*, *MYL3*, *MYL2*, *MYL1*, *MYO1E*, *MYO1C*

² Kyoto Encyclopedia of Genes and Genomes

¹ pathway-based analysis

هتروزیگوت بودند و SNP‌های با موقعیت نامشخص روی کروموزوم، برای آنالیزهای بعدی نگه داشته شدند (وانگ و همکاران ۲۰۲۰). در نهایت بعد از کنترل کیفیت تعداد ۱۶۱۳۷۶ SNP برای آنالیزهای پویش کل ژنومی بر پایه تجزیه و تحلیل غنی‌سازی مجموعه ژنی باقی ماندند. جهت ارتباط فنوتیپ‌ها با ژنوتیپ‌ها از نرم افزار GCTA نسخه ۱/۹۲ (یانگ و همکاران ۲۰۱۱) استفاده شد. مدل مورد استفاده بر پایه مدل خطی مختلط تک صفت به شکل زیر بود:

$$y = a + bx + g + e$$

که y : بردار مشاهدات فنوتیپی، a : اثر میانگین، b : اثر افزایشی SNP کاندیدا، x : بردار ژنوتیپ‌های SNP، g : اثر تصادفی پلی ژنیک و e : اثر تصادفی باقی مانده.

آنالیز پویش کل ژنومی براساس مجموعه‌های ژنی (GSEA-SNP)

اساساً آنالیز پویش ژنومی بر پایه تجزیه و تحلیل مجموعه‌های ژنی در سه مرحله انجام می‌گردد: (۱) تعیین مکان SNP‌های معنی‌دار با ژن (۲) ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی (۳) پویش کل ژنومی بر پایه آنالیز مسیر.

۱- تعیین مکان SNP‌ها با ژن‌ها: SNP‌هایی که مقدار P -value آن‌ها کمتر از ۰/۰۰۵ (اسرای‌کن و همکاران ۲۰۲۰) بود با استفاده از بسته نرم افزاری *biomaRt2* (دیورینسک و همکاران، ۲۰۰۹) در محیط R و با استفاده از ژنوم مرجع مرغ نسخه (*GRCg6a*) به ژن‌هایی که نشانگر SNP مورد نظر در داخل آن ژن و یا ۱۵ kb بالادست یا پایین دست آن ژن قرار داشت، ارتباط داده شدند (تمامی ژن‌ها شامل کدکننده، غیرکدکننده و کاذب). در این مرحله ژنی که حداقل حاوی یک SNP معنی‌دار در سطح ۰/۰۰۵ باشد، ژن معنی‌دار به شمار می‌آید.

۲- ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی: جهت تعیین طبقات عملکردی ژنی و مسیرهای متابولیکی و تنظیمی ژن‌های معنی‌دار از نرم افزار GO2MSIG (پاول ۲۰۱۴) استفاده شد. برنامه GO2MSIG از ۵ پایگاه‌های اطلاعاتی شامل هستی‌شناسی ژن (<http://www.geneontology.org>), (GO), مسیرهای

ACACA, *ACOX1*, *ACOX2* و *PNPLA2* بودند، را گزارش کرده‌اند (خلت‌آبادی فراهانی و همکاران ۲۰۲۰). هدف از انجام پژوهش حاضر، تجزیه و تحلیل مجموعه‌های ژنی و مسیرهای زیستی مرتبط با صفات رشد شامل وزن بدن و طول و قطر شانک در سنین مختلف مرغ براساس پویش کل ژنوم بر مبنای مسیر می‌باشد.

مواد و روش‌ها

در این پژوهش ابتدا فایل‌های تعیین ژنوتیپ با استفاده از تراشه‌های ژنومی همراه با رکوردهای فنوتیپی مرتبط با صفات رشد که مربوط به گونه مرغ بوده و در پایگاه‌های ذخیره‌ای عمومی مختلف داده‌های ژنومی (Dryad, Figshare, Zenodo, Animal Genome, Frontiersin) ذخیره شده بودند، استخراج و بر حسب اطلاعات مفید گروه بندی انجام شد. در نهایت از اطلاعات نسل نهم مرتبط با وزن بدن از هج تا ۱۲ هفتگی، طول شانک از ۴ تا ۱۲ هفتگی و قطر شانک از ۴ تا ۱۲ هفتگی در لاین اینترکراس پیشرفته با هدف پویش کل ژنومی که براساس مدل خطی تک نشانگری و هاپلوتیپ بود، استفاده گردید. برای دستیابی به اطلاعات رکوردهای ژنوتیپی از سایت ([doi: 10.6084/m9.figshare.12581063.v1](https://doi.org/10.6084/m9.figshare.12581063.v1)) و رکوردهای فنوتیپی (<https://www.frontiersin.org/articles/10.3389/fgen>) (e.2020.00894/full#supplementary-material) استفاده گردید. نژاد مورد مطالعه حاصل تلاقی بین نژاد گوشتی تجاری Anak و جوجه‌های نژاد Huiyang Bearded بومی چینی بودند. رکوردهای فنوتیپی شامل وزن بدن جوجه‌ها از زمان هج تا ۱۲ هفتگی که هر دو هفته وزن-کشی انجام شده بود، همچنین طول و قطر شانک از ۴ تا ۱۲ هفتگی هر دو هفته رکوردبرداری شده بود. DNA نمونه‌ها با استفاده از تکنیک توالی‌یابی GBS دستگاه ایلومینا مدل Nexseq500 تعیین ژنوتیپ شده بودند.

نشانگرهای SNP که از تمام مراحل کنترل کیفیت شامل نشانگرهای با حداقل فراوانی آلی بالاتر از ۰/۰۵ و میزان فراوانی آلی بالاتر از ۰/۹۰، کنار گذاشتن SNP‌های کروموزوم Z که ژنوتیپ آنها در افراد ماده به صورت

برای اختلاف طول ژن‌ها نیز تصحیح انجام می‌شود. علاوه بر این جهت خوشه‌بندی ژن‌های معنی‌دار از پایگاه برخط DAVID (<https://david.ncifcrf.gov>) و الگوریتم خوشه‌بندی اکتشافی فازی^۲ براساس ضریب غنی‌سازی^۳ مطابق رابطه ۲ استفاده گردید (فَن و همکاران ۲۰۱۵).

$$ES = \left(\frac{\sum_{G_i^* \in S, i^* \leq i} \frac{|r(i^*)|^p}{N_R}}{\sum_{G_i^* \in S, i^* \leq i} \frac{1}{N - N_S}} \right)^{\frac{1}{p}} \quad (\text{رابطه ۲})$$

برای تفسیر بهتر عملکرد ژن‌های به دست آمده از پایگاه‌های اطلاعاتی آنلین GeneCards (<http://www.genecards.org>) و UniProtKB (<http://www.uniprot.org>) استفاده شد.

نتایج و بحث

آمار توصیفی صفت وزن بدن از هج تا ۱۴ هفتگی، طول و قطر شانک از ۴ تا ۱۲ هفتگی به تفکیک هر دو هفته در جدول ۱ نشان داده شده است.

شناسایی مسیرهای بیوشیمیایی با استفاده از تجزیه GSEA-SNP^۴

تعداد ۱۵۵۶۴ ژن از ۱۶۸۷۸ ژن‌های کدکننده حاشیه نویسی شده در ژنوم مرغ به وسیله نشانگرهای SNP پوشش داده شد که در این میان مجموع ۴۰۷۰ ژن دارای اثر معنی‌دار روی صفات وزن بدن از هج تا ۱۴ هفتگی و طول و قطر شانک از چهار تا ۱۲ هفتگی بودند، یعنی حداقل یک نشانگر با P-value کمتر از ۰/۰۰۵ در داخل و یا در بالا یا پایین دست این ژن تا فاصله ۱۵ kb قرار گرفت. این ژن‌ها به عنوان ژن‌های معنی‌دار مرتبط با صفات وزن بدن برای تجزیه GSEA-SNP انتخاب شدند.

Panther (KEGG, <http://www.genome.jp/kegg>)، Metacyc (<http://www.pantherdb.org>)، Reactome (<http://www.reactome.org>) و جهت تعیین طبقات عملکردی و مسیرهای بیوشیمیایی استفاده می‌نمائید. در این مرحله، فرض بر این است که ژن‌هایی که در یک طبقه عملکردی قرار می‌گیرند می‌توانند به عنوان یک گروه از ژن‌هایی که برخی از ویژگی‌های خاص و مشترک دارند مانند شرکت در ۳ فرآیند هستی شناسی شامل فرآیندهای زیستی، عملکرد مولکولی و اجزای سلولی در نظر گرفته شوند.

۳- پویش کل ژنومی بر پایه آنالیز مسیر: ارتباط‌های معنی‌دار مسیرهای عملکردی با صفات مرتبط با وزن بدن، طول و قطر شانک با استفاده از توزیع فوق هندسی^۱ و آماره Fisher's exact test مورد آزمون قرار گرفت. P-value مسیرهای عملکردی که تعداد k ژن معنی‌دار در آن قرار دارد براساس رابطه ۱ محاسبه شد (پناگریکانو و همکاران ۲۰۱۳):

$$P - \text{value} = 1 - \sum_{i=1}^{k-1} \frac{\binom{S}{i} \binom{N-S}{m-i}}{\binom{N}{m}} \quad (\text{رابطه ۱})$$

در این رابطه، k برابر با تعداد ژن‌های معنی‌دار در طبقه عملکردی، S برابر با تعداد کل ژن‌های معنی‌دار مرتبط با صفات مورد بررسی، N برابر با کل تعداد ژن‌هایی که در این مطالعه آنالیز شدند و m برابر با تعداد کل ژن‌های موجود در مسیر عملکردی. تجزیه غنی‌سازی مجموعه ژنی با استفاده از بسته نرم افزاری goseq (یانگ و همکاران ۲۰۱۰) در محیط نرم افزار R انجام گردید. بسته goseq از آزمون فوق هندسی استفاده می‌کند. همچنین

Table 1- Descriptive statistics of the phenotypes

Phenotype	Number	Mean	SD	Min	Max	SE
BW0 (g)	587	30.17	2.85	21.5	39.8	0.12
BW2 (g)	589	166.63	17.98	99.6	219.3	0.74
BW4 (g)	585	433.93	58.79	246.5	630.5	2.43
BW6 (g)	599	775.54	118.59	315	1199	4.85

³ Enrichment Score (ES)

⁴ Gene Set Enrichment Analysis- SNP

¹ Hypergeometric

^۲ Heuristic fuzzy

BW8 (g)	598	1147.7	185.55	662	1855.2	7.59
BW10 (g)	598	1510.34	272.44	760	2419	11.14
BW12 (g)	598	1895.33	365.81	936	3094.5	14.96
BW14 (g)	588	2090.48	398.21	1079	3127.5	16.42
SC4 (cm)	590	2.99	0.2	2.4	3.7	0.01
SC6 (cm)	599	3.52	0.26	2.9	4.4	0.01
SC8 (cm)	590	3.77	0.3	3	4.8	0.01
SC10 (cm)	596	4.02	0.34	3.3	5	0.01
SC12 (cm)	592	4.23	0.36	3.5	5.3	0.01
SL4 (cm)	590	51.21	2.98	42.99	59.24	0.12
SL6 (cm)	599	66.6	4.44	53.79	79.33	0.18
SL8 (cm)	595	76.66	6.11	62.37	96.66	0.25
SL10 (cm)	585	84.79	8.28	66.77	107.41	0.34
SL12 (cm)	576	91.36	10.5	71.28	119.8	0.44

BW0: body weights at hatching; **BW2**: body weights at 2 weeks of age; **BW4**: body weights at 4 weeks of age; **BW6**: body weights at 6 weeks of age; **BW8**: body weights at 8 weeks of age; **BW10**: body weights at 10 weeks of age; **BW12**: body weights at 12 weeks of age; **BW14**: body weights at 14 weeks of age. **SC4**: shank circumference at 4 weeks of age; **SC6**: shank circumference at 6 weeks of age; **SC8**: shank circumference at 8 weeks of age; **SC10**: shank circumference at 10 weeks of age; **SC12**: shank circumference at 12 weeks of age. **SL4**: shank length at 4 weeks of age; **SL6**: shank length at 6 weeks of age; **SL8**: shank length at 8 weeks of age; **SL10**: shank length at 10 weeks of age; **SL12**: shank length at 12 weeks of age.

(genecards). در مطالعات مختلفی در مرغان بومی آذربایجان غربی و مازندران با بررسی چند شکلی ژن کاندیدای میوستاتین با صفت وزن بدن در ۱۲ هفته‌ی ارتباط معنی‌داری گزارش شده است (زندگی و همکاران ۲۰۱۳ و عطارچی و همکاران ۲۰۱۷).

کالپاین در ارتباط با کالپاستاتین سیستم پروتئازی وابسته به یون کلسیم به نام سیستم کالپاین-کالپستاتین تشکیل می‌دهند (genecards). با بررسی چند شکلی ژن کاندیدای *CAPN3* با صفات مرتبط با لاشه در جوجه‌های لاین خالص تجاری، ارتباط معنی‌داری گزارش شده است (ژانگ و همکاران ۲۰۰۹). علاوه بر این ژن کاندیدای گیرنده لپتین در تنظیم مصرف غذا، متابولیسم انرژی و تولید مثل نقش اساسی دارد. با مطالعه پویش کل ژنومی در جوجه‌های گوشتی، ژن *LEPR* مرتبط با وزن بدن گزارش شده است (زو و همکاران ۲۰۱۳).

تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌های مختلف شامل ۱۰۵۶ طبقه هستی‌شناسی، ۸۴ مسیر بیوشیمیایی PANTHER و مسیر Reactome بود. همان طور که در جداول ۲ و ۳ مشاهده می‌شود تعداد ۱۹ طبقه عملکردی در هستی‌شناسی فرآیندهای زیستی، عملکرد مولکولی، اجزای سلولی و مسیرهای بیوشیمیایی با صفات وزن بدن در هفته‌های مختلف از ۱ تا ۱۴ هفته‌ی طول و قطر شانک دارای ارتباط هستند ($P < 0.05$). مسیرهای که بیشتر از ۳ ژن و کمتر از ۳۰۰ ژن داشتند گزارش شده‌اند.

از مسیرهای بیولوژیکی مهم و معنی‌دار مرتبط با وزن بدن می‌توان به مسیرهای Regulation of muscle organ و Bone growth و development اشاره نمود که از بین ژن‌های معنی‌دار در این مسیرها، ژن کاندیدای *MSTN*، *CAPN3* و *LEPR* مرتبط با رشد عضله اسکلتی می‌باشند

Table 2- Gene set enrichment analysis significantly ($P < 0.05$) associated with body weight

Category	Term	No. of Genes	Genes	P value	FDR
GO_BP	GO:0048634: Regulation of muscle organ development	8	<i>TBX20, CAPN3, HEY2, LEF1, MAPK14, MEGF10, MSTN, ABCC4</i>	5.36E-09	1.95E-02
	GO:0006641:Triglyceride metabolic process	6	<i>ABHD4, PNPLA3, PNPLA2, LPIN1, LIPC, GK5</i>	2.54E-05	3.29E-02
	GO:0055088: Lipid homeostasis	6	<i>ABHD4, PNPLA3, PNPLA2, TMEM56, MTPP, TLCD2</i>	6.08E-06	4.05E-02
	GO:1900046:Regulation of hemostasis	9	<i>ADAMTS18, ANXA2, CPB2, PRKCD, SERPINC1, STX2, TEC, TLR4, VAV2</i>	7.20E-08	4.29E-03

	GO:0001558: Regulation of cell growth	6	<i>HOMER3, IGF1, BCL2, ADAM10, MYOCD, ABL1</i>	9.24E-07	5.76E-02
	GO:0060249: Anatomical structure homeostasis	14	<i>ABCA4, BCL2, LDB2, PIF1, SOX9, TYRO3, EPAS1, HMBOX1, ILDR2, LPCAT1, MC4R, PIGR, TLR4, TNFSF11</i>	4.71E-05	2.16E-02
	GO:0098868: Bone growth	4	<i>BNC2, COL27A1, LEPR, MMP13</i>	7.28E-10	3.51E-02
GO_CC					
	GO:0005925: Focal adhesion	5	<i>ADAM10, NCKAP1, AFAP1, CYFIP1, CDH2,</i>	1.78E-04	3.11E-02
GO_MF					
	GO:0015485: Cholesterol binding	5	<i>OSBPL6, OSBPL10, OSBPL1A, TSP0, OSBP2</i>	3.72E-07	6.93E-03
KEGG Pathway					
	bta04921: Oxytocin signaling pathway	20	<i>ADCY3, ADCY5, ADCY8, ADCY9, CACNA2D2, CACNA2D3, CACNG1, CACNA1D, CAMK2A, CAMK4, EGFR, GUCY1A2, ITPR1, MAP2K5, PLA2G4D, PRKAA1, PRKCA, PRKCB, RYR2, RYR3</i>	3.7E-4	1.93E-02

پویش کل ژنومی با وزن بدن در سنین ۸، ۱۰ و ۱۳ هفتگی و طول و قطر شانک در ۱۳ هفتگی ارتباط معنی‌داری با ژن *IGF1* در جوجه‌های *Silkie* گزارش شده است (تانگ و همکاران ۲۰۱۰). همچنین ناحیه معنی‌دار مؤثر روی وزن بدن در ناحیه ۳/۷ مگابازی کروموزوم ۲۷ به دست آمد که ژن کاندیدای *IGF2BP1* با مطالعه قبلی مشترک بود (وانگ و همکاران، ۲۰۲۰).

از مسیرهای معنی‌دار مرتبط با وزن بدن می‌توان به مسیر Regulation of cell growth اشاره نمود که از بین ژن‌های موجود در این مسیر، ژن *IGF1* در رشد بافت‌های مختلف شامل: بافت عضله (تکثیر و تمایز سلول‌های عضلانی)، بافت غضروفی (تشکیل مجموعه کندروسیت و فعالیت آلکالین فسفاتاز) و بافت استخوانی (تکثیر و تقسیم سلول‌های استخوان) نقش کلیدی دارد. در مطالعه

Table 3- Gene set enrichment analysis significantly ($P < 0.05$) associated with bone traits

Category	Term	No. of genes ¹		P value	FDR
GO_BP					
	GO:0048646: Anatomical structure formation involved in morphogenesis	12	<i>ARL13B, ABCC4, DSCAM, HERC1, FGF18, FNI, IGF1, ITGB1BP1, ITGB1, LEPR, PIBF1, TGFB3</i>	6.21E-10	6.51E-03
	GO:0045778: Positive regulation of ossification	14	<i>SMAD1, SUCO, ACVR2A, ACVR2B, ANO6, TMEM135, BMPR1B, CYR61, GTF2H4, IGF1, JAG1, TOB2, TGFB3, TP63</i>	8.20E-06	4.59E-02
	GO:0060560: Developmental growth involved in morphogenesis	6	<i>FNI, ITGB1, IFRD1, OGN, PLXNA1, PLXNA2</i>	9.32E-07	3.41E-02
	GO:0045669: Positive regulation of osteoblast differentiation	9	<i>SMAD1, SUCO, ACVR2B, ACVR2B, BMPR1B, CYR61, GTF2H4, IGF1, TP63</i>	1.48E-06	4.15E-02
	GO:0006816: Calcium ion transport	7	<i>HTR1D, HTR1E, ABL1, CACNA1G, CAPN3, FFAR4, TRPC1</i>	4.02E-05	4.10E-02

	GO:0007204: Positive regulation of cytosolic calcium ion concentration	6	<i>CAPN3</i> , <i>FFAR4</i> , <i>PTGFR</i> , <i>PTPN6</i> , <i>CASQ2</i> , <i>EDNI</i>	7.28E-09	2.03E-03
GO_CC	GO:0034704: Calcium channel complex	11	<i>AKAP6</i> , <i>SESTD1</i> , <i>ANXA2</i> , <i>ANXA5</i> , <i>CACNA1C</i> , <i>CACNA1B</i> , <i>CACNA1G</i> , <i>CACNA1H</i> , <i>CCDC109B</i> , <i>SMDT1</i> , <i>TRPC4</i>	1.78E-05	2.10E-02
GO_MF	GO:0005262: Calcium channel activity	6	<i>DENND5A</i> , <i>DENND5B</i> , <i>TRPC1</i> , <i>TRPC3</i> , <i>TRPC4</i> , <i>TPCN2</i>	3.42E-07	1.98E-02
KEGG Pathway	bta04020: Calcium signaling pathway	12	<i>ADCY5</i> , <i>PTGFR</i> , <i>ADCY2</i> , <i>CHRM3</i> , <i>CHRM5</i> , <i>ADAMTS18</i> , <i>ITPR2</i> , <i>MYLK2</i> , <i>MYLK</i> , <i>PLCZ1</i> , <i>PLCG1</i> , <i>ERBB4</i>	3.7E-4	2.17E-02

از مسیرهای بیولوژیکی معنی‌دار مرتبط با طول و قطر شانک می‌توان به مسیر *Anatomical structure formation involved in morphogenesis* اشاره نمود که جزء مسیرهای فرآیندهای زیستی مرتبط با رشد می‌باشد که ژن‌های کاندیدای *FNI* و *IGF1* دارای نقش بیولوژیکی مستقیمی با صفات مرتبط با تولید داشت که بیشتر بحث خواهد شد. ژن کاندیدای *FNI* دارای نقش اساسی در انتقال گلوکز به داخل سلول‌ها و تنظیم هموستازی گلوکز دارد (genecards). با بررسی پروفایل ترانسکرپتوم کل ژنوم در جوجه‌ها در ارتباط با تفرق بین ژن‌های مرتبط با رشد اولیه، ژن فیبرونکتین ۱ (*FNI*) بیان بالایی داشته و نقش کلیدی در تشکیل کلاژن داشته است (زو و همکاران ۲۰۱۷). همچنین در بررسی چند شکلی بین ژن کاندیدای *IGF1* با صفات مرتبط با حجم و تراکم استخوان، طول و وزن شانگ ارتباط معنی‌داری در جمعیت F2 نژاد لگهورن و Fayoumi گزارش شده است (ژاو و همکاران ۲۰۰۵).

مسیرهای بیولوژیکی *Positive regulation of ossification* و *Positive regulation of cytosolic calcium ion concentration* از دیگر مسیرهایی بودند که در ارتباط با طول و قطر شانک شناسایی شدند. مطالعه پویش ژنومی با هدف شناسایی جایگاه‌های ژنی مرتبط با صفات توسعه سیستم اسکلتی در استخوان

شاید بتوان مسیر *Anatomical structure homeostasis* که جزء هستی‌شناسی فرآیندهای بیولوژیکی است را یکی از مهمترین مسیرهای مؤثر بر فرآیند وزن بدن در ارتباط با صفات رشد دانست. از بین ژن‌های معنی‌دار در این مسیر، می‌توان به ژن‌های کاندیدای *MC4R* و *LDB2* اشاره نمود. در مطالعه‌ای با بررسی ارتباط چندشکلی در ژن کاندیدای *MC4R* با وزن بدن در سنین ۲، ۴ و ۱۰ هفتگی در جوجه‌های *Korat* ارتباط معنی‌داری گزارش شده است (کیوبت و همکاران ۲۰۱۹). در مطالعه پویش کل ژنومی با هدف شناسایی مناطق ژنومی مرتبط با وزن بدن در سنین ۷ تا ۱۲ هفتگی و افزایش وزن روزانه ۷ تا ۱۲ هفتگی در جمعیت جوجه‌های F2 حاصل از پلیموت راک و سیلکی ژن *LDB2* گزارش شده است (گو و همکاران ۲۰۱۱).

مسیر دیگر معنی‌دار مرتبط با وزن بدن را می‌توان *Regulation of hemostasis* نام برد که از بین ژن‌ها، ژن کاندیدای *ANXA2* در متابولسم لیپید و اسیدهای چرب در بدن مرغ نقش دارند (genecards). در مطالعه پویش ژنومی بر پایه پروفایل *MicroRNA* مرتبط با رشد عضله در جوجه‌های گوشتی تجاری در مقابل جوجه‌های نژاد بومی، ژن آنکسین A2 (*ANXA2*) دارای نقش تنظیمی با بیان ژن بالا بود (کتری و همکاران ۲۰۱۸).

حاضر استفاده کرد. به طور کلی مسیرهای شناسایی شده در این تحقیق می‌تواند ما را به سمت ژن‌های مؤثر بر صفات هدایت کند و بررسی بیشتر نواحی مهم ژنومی شناسایی شده، به خصوص مناطقی که به طور همزمان با چند صفت وزن بدن بصورت مشترک معنی‌دار هستند، با استفاده از آزمون‌های آزمایشگاهی مختلف می‌تواند در تأیید نتایج بدست آمده در این پژوهش مؤثر باشند. از طرفی در این تحقیق به دلیل عدم دسترسی به رکوردهای فنوتیپی و اطلاعات ژنوتیپی مرتبط با صفات رشد در مرغان بومی کشور، از اطلاعات فنوتیپی و ژنوتیپی نژادهای خالص و بومی کشور چین استفاده شد. لذا استفاده از نتایج این تحقیق در مرغان بومی کشور نیاز به مطالعات بیشتر دارد تا در این جمعیت‌ها نیز تأیید شوند.

نتیجه‌گیری

با توجه به عملکرد بیولوژیکی مسیرهای شناسایی شده در این پژوهش، به نظر می‌رسد این ژن‌ها در بروز فنوتیپی صفات مرتبط با وزن بدن نقش ایفا می‌کنند، در نتیجه می‌توان کارآیی روش تجزیه و تحلیل غنی‌سازی مجموعه‌ی ژنی برای پویش ژنومی صفات تولیدی اقتصادی را نیز مورد تأیید قرار داد. همچنین با بررسی چند شکلی موجود در ژن‌های کاندیدای شناسایی شده از تجزیه و تحلیل غنی‌سازی مجموعه ژنی مرتبط با وزن بدن از طریق مطالعات آزمایشگاهی در نژادهای مرغ بومی و نتایج به‌دست آمده را برای مطالعات اصلاحی با هدف افزایش وزن بدن از طریق شناسایی ژن‌های کاندیدای عمده اثر و استفاده از آنها در برنامه انتخاب به کمک نشانگر به کار برد.

درشت نی در تلاقی بین نژادهای گوشتی و تخم‌گذار ژن *EDNI* گزارش شده است (فوفری و همکاران ۲۰۱۹). همچنین با مطالعه پویش کل ژنومی در جمعیت F2 حاصل از تلاقی بین لاین آراین با مرغ بومی آذربایجان ژن *TMEM135* را به عنوان ژن مرتبط با طول شانک گزارش کردند (عمرانی و همکاران ۲۰۲۰).

تحلیل مسیرهای زیستی KEGG نشان‌دهنده این است که ژن‌های *ADAMTS18* و *ERBB4* به طور معنی‌داری با مسیر Calcium signaling pathway در ارتباط می‌باشند. این مسیر زیستی نقش مهمی در رشد و توسعه استخوان دارد. در شکل ۱ مسیر سیگنالی کلسیم و ژن‌های مربوطه شناسایی شده از پایگاه برخط KEGG ارائه شده است. با مطالعه پویش کل ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با تراکم استخوان، ژن *ADAMTS18* گزارش شده است (زیانگ و همکاران ۲۰۰۹).

در تجزیه و تحلیل‌های غنی‌سازی مجموعه ژنی تلاش گردید از آزمون‌های پر کاربرد و سختگیرانه (مانند برنامه GO2MSIG) برای شناسایی مسیرهای زیستی معنی‌دار استفاده گردد. به هر حال باید توجه کرد که یکی از محدودیت‌های اجرای مطالعات پویش ژنومی، کافی نبودن تعداد رکوردهای فنوتیپی و جمعیت غیر خویشاوند برای آنالیز پیوستگی می‌باشد که منجر به افزایش خطای نوع اول یا به عبارتی شناسایی نشانگرهای معنی‌دار مثبت کاذب می‌شود (سی‌بری و همکاران ۲۰۱۷). با توجه به تعداد کم نمونه مورد استفاده در پژوهش حاضر (۵۹۹ مرغ) نیز باید در استفاده از ژن‌های کاندیدای شناسایی شده مرتبط با صفات رشد در برنامه‌های اصلاحی با احتیاط عمل کرد. البته می‌توان با ادغام داده‌های تحقیقات مشابه جدید و استفاده از آنالیزهای آماری جامع‌تر برای تأیید نتایج پژوهش

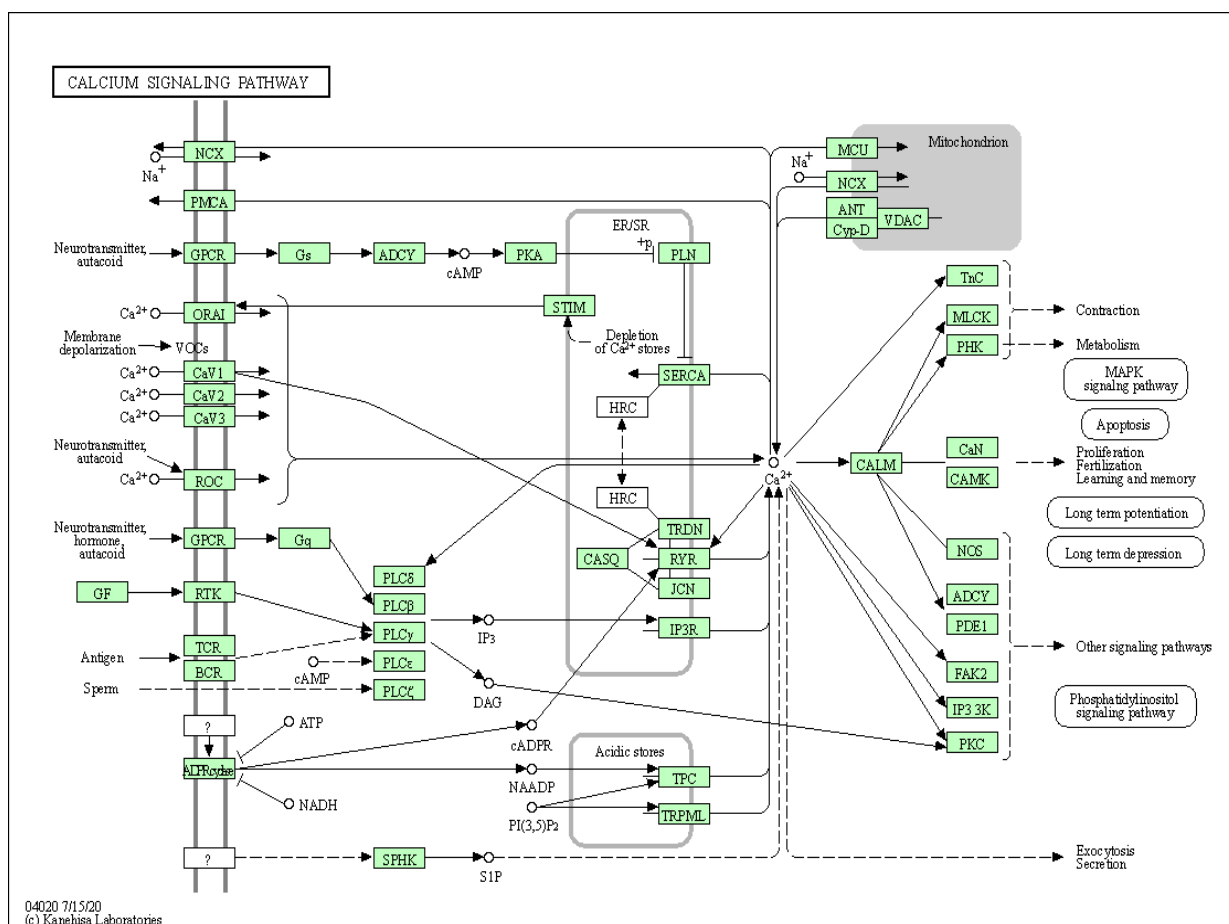


Figure 1- Calcium signaling pathway and candidate genes related to bone traits highlighted (KEGG database)

سیاسگزاری

بدین وسیله نگارندگان مقاله از پروفیسور *Hu* از دانشگاه کشاورزی چین برای تعیین ژنوتیپ‌ها نمونه‌ها و در اختیار قرار دادن رکوردهای فنوتیپی قدردانی می‌گردد.

منابع مورد استفاده

- Attarchi H, Tahmoorespur M, Ahani azari M, Sekhavati M and Mohajer M, 2017. Evaluation of allelic polymorphism of myostatin gene and its association with growth and carcass traits in Mazandaran native chicken. *Journal of Animal Environment* 9(4): 109-112.
- Azizpour N, Khaltabadi Farahani AH, Moradi MH and Mohammadi H, 2020. Genome-wide association study based on Gene-set enrichment analysis associated with milk yield in Holstein cattle. *Journal of Animal Science Researches* 30(1): 79-92.
- Durinck S, Spellman PT, Birney E and Huber W, 2009. Mapping identifiers for the integration of genomic datasets with the R/bioconductor package biomaRt. *Nature Protocols* 4: 1184–1191.
- Emrani H, Masoudi AA, Vaez Torshizi R and Ehsani A, 2020. Genome-wide association study of shank length and diameter at different developmental stages in chicken F2 resource population. *Animal Genetics* 51(5):722-730.
- Fan H, Wu Y, Zhou X, Xia J, Zhang W, Song Y, Liu F, Chen Y, Zhang L, Gao X, Gao H and Li J, 2015. Pathway-Based Genome-Wide Association Studies for Two Meat Production Traits in Simmental Cattle. *Scientific Reports* 5:18389.

- Faveri JC, Pinto LFB, de Camargo GMF, Pedrosa VB, Peixoto JO, Marchesi JAP, Kawski VL, Coutinho LL and Ledur MC, 2019. Quantitative trait loci for morphometric and mineral composition traits of the tibia bone in a broiler × layer cross. *Animal* 13(8): 1563-1569.
- Guo J, Sun C, Qu L, Shen M, Dou T, Ma M, Wang K and Yang N, 2017. Genetic architecture of bone quality variation in layer chickens revealed by a genome-wide association study. *Scientific Reports* 6(7): 45317.
- Khaltabadi Farahani AH, Mohammadi H, Moradi MH, Ghasemi HA and Hajkhodadadi I, 2020. Gene-set enrichment analysis to identify genes and biological pathways associated with body weight in Chicken. *Animal Production Research* 9(3): 47-57.
- Kubota S, Vandee A, Keawnakient P, Molee W, Yongsawatdikul J and Molee A, 2019. Effects of the MC4R, CAPN1, and ADSL genes on body weight and purine content in slow-growing chickens. *Poultry Science* 1;98(10): 4327-4337.
- Khatri B, Seo D, Shouse S, Pan JH, Hudson NJ, Kim JK, Bottje W and Kong BC, 2018. MicroRNA profiling associated with muscle growth in modern broilers compared to an unselected chicken breed. *BMC Genomics* 17;19(1):683.
- Mooney MA and Wilmot B, 2015. Gene Set Analysis: A Step-By-Step Guide. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 168: 517-527.
- Mohammadi H and Sadeghi M, 2010. Estimation of Genetic Parameters for Growth and Reproduction Traits and Genetic Trends of Growth Traits in Zel Sheep Breed under Rural Production System. *Iranian Journal of Animal Science* 41(3), 231-241.
- Peñagaricano F, Weigel KA, Rosa GJ and Khatib H, 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Frontiers in Genetics* 3: 307-314.
- Powell JA, 2014. GO2MSIG, an automated GO based multi-species gene set generator for gene set enrichment analysis. *BMC Bioinformatics* 15: 146-149.
- Seabury CM, Oldeschulte DL, Saatchi M, Beaver JE, Decker JE and Taylor JF, 2017. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics* 18(1): 386-396.
- Srikanth K, Lee SH, Chung KY, Park JE, Jang GW, Park MR, Kim NY, Kim TH, Chai HH, Park WC and Lim D, 2020. A Gene-Set Enrichment and Protein-Protein Interaction Network-Based GWAS with Regulatory SNPs Identifies Candidate Genes and Pathways Associated with Carcass Traits in Hanwoo Cattle. *Genes (Basel)*.11(3):316.
- Tang S, Sun D, Ou J, Zhang Y, Xu G and Zhang Y, 2010. Evaluation of the IGFs (IGF1 and IGF2) genes as candidates for growth, body measurement, carcass, and reproduction traits in Beijing You and Silkie chickens. *Animal Biotechnology* 21(2): 104-13.
- Wang Y, Bu L, Cao X, Qu H, Zhang C, Ren J, Huang Z, Zhao Y, Luo C, Hu X, Shu D and Li N, 2020. Genetic Dissection of Growth Traits in a Unique Chicken Advanced Intercross Line. *Frontiers in Genetics* 11: 894.
- Wang L, Jia P, Wolfinger RD, Chen X and Zhao Z, 2011. Gene set analysis of genome-wide association studies: Methodological issues and perspectives. *Genomics* 98: 1–8.
- Xu Z, Nie Q and Zhang X, 2013. Overview of Genomic Insights into Chicken Growth Traits Based on Genome-Wide Association Study and microRNA Regulation. *Current Genomics* 14(2): 137-46.
- Xue Q, Zhang G, Li T, Ling J, Zhang X and Wang J, 2017. Transcriptomic profile of leg muscle during early growth in chicken. *PLoS One* 14;12(3):e0173824.
- Xiong DH, Liu XG, Guo YF, Tan LJ, Wang L, Sha BY, Tang ZH, Pan F, Yang TL, Liu YJ, Zmuda JM and Deng HW, 2009. Genome-wide association and follow-up replication studies identified ADAMTS18 and TGFBR3 as bone mass candidate genes in different ethnic groups. *American Journal of Human Genetics* 84(3): 388-98.
- Yang J, Lee SH, Goddard ME and Visscher PM, 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics* 88: 76–82.
- Young MD, Wakefield MJ, Smyth GK and Oshlack A, 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology* 11: 14-23.

- Zandi S, Zamani P and Mardani K, 2013. Myostatin Gene Polymorphism and Its Association with Production Traits in Western Azerbaijan Native Chickens. *Iranian Journal of Applied Animal Science* 3(3): 611-615.
- Zhang XX, Ran JS, Lian T, Li ZQ, Yang CW, Jiang XS, Du HR, Cui ZF and Liu YP, 2019. The Single Nucleotide Polymorphisms of Myostatin Gene and Their Associations with Growth and Carcass Traits in Daheng Broiler. *Brazilian Journal of Poultry Science* 21(3): 1-8.
- Zhang ZR, Liu YP, Yao YG, Jiang XS, Du HR and Zhu Q, 2009. Identification and association of the single nucleotide polymorphisms in calpain3 (CAPN3) gene with carcass traits in chickens. *BMC Genetics* 5; 10:10.
- Zhou H, Mitchell AD, McMurtry JP, Ashwell CM and Lamont SJ, 2005. Insulin-like growth factor-I gene polymorphism associations with growth, body composition, skeleton integrity, and metabolic traits in chickens. *Poultry Science* 84(2):212-219

Genome wide association study based on gene-set enrichment analysis of growth traits in a Chicken advanced intercross line

H Mohammadi^{1*}, A H Khaltabadi Farahani², M H Moradi² and I Hajkhodadadi²



Received: June 17, 2021

Accepted: September 26, 2021

¹Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Environmental Science, Arak University.

²Associate Professor, Department of Animal Sciences, Faculty of Agriculture and Environmental Science, Arak University.

*Corresponding author: E-mail: H-mohammadi64@araku.ac.ir

 <p>پژوهش‌های علوم دامی Animal Science Research</p>	<p>Journal of Animal Science/vol.31 No.3/ 2021/pp 99-111 https://animalscience.tabrizu.ac.ir</p>	
<p>© 2009 Copyright by Faculty of Agriculture, University of Tabriz, Tabriz, Iran This is an open access article under the CC BY NC license (https://creativecommons.org/licenses/by-nc/2.0/) DOI: 10.22034/AS.2021.46637.1621</p>		

Introduction: Understanding the genetic control of growth traits is essential for effective poultry breeding poultry. One way to identify new loci and confirm existing QTL is through genome-wide association analysis (GWAA) (Wang et al., 2011). In addition, identifying loci with large effects on economically important traits, has been one of the important goal to poultry breeding. QTL assisted selection and genomic regions affecting the production traits have been considered to increase the efficiency of selection and improve production performance (Seabury et al., 2017). Genome wide association studies typically focus on genetic markers with the strongest evidence of association. However, single markers often explain only a small component of the genetic variance and hence offer a limited understanding of the trait under study. A solution to tackle the aforementioned problems, and expand understanding of the genetic background of complex traits, is to move up the analysis from the SNP to the gene and gene-set levels (Peñagaricano et al., 2013). In a gene-set analysis, a group of related genes that harbor significant SNP previously identified in GWAS, is tested for over-representation in a specific pathway.

Material and methods: The aim of the this study was to genome wide association studies (GWAS) based on Gene set enrichment analysis for identifying the loci associated with related to body weight and shank length and diameter traits in advanced intercross line (AIL) using the high-confidence SNPs that enable us to study 161376 SNP markers simultaneously. For this purpose, the 599 advanced intercross line and 161376 markers were performed with mixed linear model (MLM) approach was used for the GWAS of the F9 generation, as implemented in the GCTA package (v1.92) (Yang et al., 2011) and no any correction to adjust the error rate. The gene set analysis consisted of three different steps: (1) the assignment of SNPs to genes, (2) the assignment of genes to functional categories, and (3) the association analysis between each functional category and the phenotype of interest.

In brief, for each trait, nominal P-values < 0.005 from the GWAS analyses were used to identify significant SNP. Using the *biomaRt* R package, the SNP were assigned to genes if they were within the genomic sequence of the gene or within a flanking region of 15 kb up- and downstream of the gene, to include SNP located in regulatory regions. For the assignment of the genes to functional categories, the Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathway databases were used. The GO database designates biological descriptors to genes based on attributes of their encoded products and it is further partitioned into 3 components: biological process, molecular

function, and cellular component. The KEGG pathway database contains metabolic and regulatory pathways, representing the actual knowledge on molecular interactions and reaction networks. Finally, a Fisher's exact test was performed to test for overrepresentation of the significant genes for each gene-set. The gene enrichment analysis was performed with the *goseq* R package. In the next step, a bioinformatics analysis was implemented to identify the biological pathways performed in BioMart, Panther, DAVID and GeneCards databases.

Results and discussion:

Gene set enrichment analysis has proven to be a great complement of genome-wide association analysis (Abdalla et al., 2016). Among available gene set databases, GO is probably the most popular, whereas KEGG is a relatively new tool that is gaining ground in livestock genomics (Morota et al., 2015, 2016). We had hypothesized that the use of gene set information could improve prediction. However, neither of the gene set SNP classes outperformed the standard whole-genome approach. Gene sets have been primarily developed using data from model organisms, such as mice and flies, so it is possible that some of the genes included in these terms are irrelevant for meat production. It is likely that a better understanding of the biology underlying meat production specifically, plus an advance in the annotation of the chicken genome, can provide new opportunities for predicting production using gene set information. Eleven SNP markers on chromosomes 1, 2, 4, 5, 7, 8, 10, 11, and 27 located in *MSTN*, *CAPN3*, *PNPLA3*, *ANXA2*, *IGF1*, *LDB2*, *LEPR*, *FN1*, *TMEM135*, *MC4R*, *EDNI*, and *ADAMTS18* genes were identified. Some of the genes found to be consistent with some previous studies. Those seem to be involved in biological pathways related to muscle skeletal growth, energy metabolism and bone growth and development. According to pathway analysis, 19 pathways from gene ontology and KEGG pathway were associated with the body weight, shank length and diameter trait ($P < 0.05$). Among those pathways, the regulation of muscle organ development, regulation of cell growth and anatomical structure homeostasis biological pathway have important roles in the growth and skeletal muscle development. Also, the anatomical structure formation involved in morphogenesis, positive regulation of ossification and calcium signaling pathway presumably has significant association with body weight and shank length as well as diameter traits. Some of these regulatory regions, such as enhancers, are located far from the genes. Therefore, although the gene might be part of the analysis, the relevant variant would probably not be included in the gene set SNP class. Finally, linkage disequilibrium interferes with the use of biological information in prediction because irrelevant regions (regions without any biological role) capture part of the information encoded in relevant regions, causing both regions to exhibit similar predictive abilities. The use of very high density SNP data or even whole genome sequence data could alleviate some of these issues.

Finally, it is worth noting that our gene-set enrichment analysis was conducted using a panel of SNP obtained from a single marker regression GWAS, which relies on a simplified theory of the genomic background of traits, without considering for instance the joint effect of SNP. Hence, other approaches (e.g., GWAS exploring SNP by SNP interactions) might provide a better basis for biological pathway analysis.

Conclusion: Our observations agreed with the previous results from GWAS of body weight, shank length and diameter traits. Moreover, additional regions in the chicken genome associated with economically important traits were revealed. Our findings would contribute to a better understanding of the genetic control of growth traits in broiler chicken accelerating the genetic progress in poultry breeding programs.

Keywords: Body weight, Candidate gene, Chicken, Genome scan, Gene set enrichment analysis