



تجزیه و تحلیل مجموعه‌های ژنی جهت شناسایی ژن‌ها و مسیرهای زیستی مرتبط با صفات وزن بدن در مرغ

امیر حسین خلت آبادی فراهانی^{۱*}، حسین محمدی^۱، محمد حسین مرادی^۱، حسینعلی قاسمی^۲، ایمان حاج خدادی^۱

۱- استادیار گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک

۲- دانشیار گروه علوم دامی، دانشکده کشاورزی و منابع طبیعی، دانشگاه اراک

(تاریخ دریافت: ۹۸/۰۸/۱۳ - تاریخ پذیرش: ۹۸/۱۲/۱۳)

چکیده

این پژوهش به منظور مطالعه پویای کل ژنوم بر پایه تجزیه و تحلیل غنی‌سازی مجموعه ژنی جهت شناسایی جایگاه‌های ژنی مؤثر بر وزن بدن در سنین مختلف چهار نژاد مرغ چوآ، سیلک، لنگشن و بیرد با استفاده از آرایه‌های ژنومی با تراکم بالا انجام شد. اطلاعات رکوردهای فنوتیپی و ژنوتیپی نمونه‌ها از پایگاه ذخیره ژنومی برخط Frontiersin استفاده شد. مطالعه پویای کل ژنومی از ۴۰۲ قطعه مرغ و خروس با صفات وزن بدن از هفته اول تا ۱۵ هفته در برنامه GenABEL ارزیابی شد. در مرحله بعد، تجزیه غنی‌سازی مجموعه ژنی با بسته نرم افزاری gseq برنامه R با هدف شناسایی عملکرد زیستی ژن‌های نزدیک در مناطق انتخابی کاندیدا با پایگاه‌های GO، KEGG، DAVID و PANTHER انجام شد. در این پژوهش، تعداد ۱۰ نشانگر تک نوکلئوتیدی واقع روی کروموزوم‌های ۱، ۲، ۵، ۷، ۱۰، ۱۴، ۱۸، ۱۹، ۲۰ و ۲۷ شناسایی شدند که با ژن‌های *ABCG1*، *ACOX1*، *ACACA*، *SLC2A8*، *MYL3*، *MYL2*، *MYL1*، *MYO1E*، *MYO1C*، *MYO1B*، *MYH11*، *MYH10*، *MYOD1* و *PNPLA2* مرتبط بودند. برخی از این ژن‌ها در مناطق معنی‌دار با مطالعات قبلی هم‌خوانی داشتند. در تجزیه غنی‌سازی مجموعه ژنی، تعداد ۱۷ مسیر هستی‌شناسی ژنی و بیوشیمیایی با صفات وزن بدن شناسایی شدند ($P < 0.01$). از این بین، مسیرهای *anatomical structure development*، *cytoskeletal protein binding* و *Tricarboxylic acid cycle* نقش مهمی در توسعه الیاف عضلانی اسکلتی و سوخت و ساز چربی داشتند. با توجه به تأیید مناطق قبلی پویای ژنومی صفات وزن بدن و شناسایی مناطق ژنومی جدید، استفاده از یافته‌های این تحقیق می‌تواند سبب تسریع در پیشرفت ژنتیکی برنامه‌های اصلاح نژادی مرغ شود.

واژه‌های کلیدی: تجزیه بر پایه مسیر، ژن کاندیدا، مرغ، وزن بدن

* نویسنده مسئول: amfarahanikh@gmail.com

مقدمه

وزن بدن یکی از مهمترین صفات اقتصادی در صنعت پرورش مرغ است (Yousefi Zonuz *et al.*, 2013). هدف از مطالعات پویش ژنومی که به منظور شناسایی وابستگی بین یک نشانگر چندشکل تک‌نوکلئوتیدی یا SNP و یک صفت با استفاده از نشانگرهای با تراکم بالا در سطح ژنوم است، پیدا کردن ژنی حامل جهش‌های با اثر مستقیم بر فنوتیپ یک صفت تولیدی، تولیدمثلی و یا بیماری است. این اطلاعات می‌تواند برای انتخاب به کمک نشانگر مفید بوده و به درک ساز و کار مولکولی صفات مورد مطالعه کمک نماید.

داده‌های مورد استفاده در این تحقیق اولین بار به وسیله Yuan *et al.* (2018) برای صفات مرتبط با وزن از هفته اول تا هفته ۱۵ پرورش، مورد تجزیه قرار گرفتند، که در آن تحقیق با استفاده از مدل‌های خطی مختلط تک نشانگری و هاپلویتپ تجزیه شده و از آزمون permutation برای تعیین آستانه معنی‌داری و جلوگیری از خطای نوع اول استفاده شد. در مجموع هشت نشانگر تک‌نوکلئوتیدی و شش بلوک هاپلویتپی معنی‌دار مرتبط با وزن بدن در سنین ۴، ۶، ۷، ۸، ۹، ۱۰، ۱۱، ۱۲، ۱۳، ۱۴ و ۱۵ هفتگی شناسایی شد. مطابق یافته‌های مطالعه مدنظر و مقالات مختلف در مطالعات پویش ژنومی مبتنی بر مدل‌های خطی و غیر خطی، مشکل مربوطه بالا بودن نرخ خطای نوع اول و بیش برآورد اثر نشانگرهای SNP است. 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 (Peng *et al.*, 2010).

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 در یک دسته یا گروه ژنی که به طور عملکردی با هم مرتبط هستند بررسی می‌شود. به عبارت دیگر تجزیه

پوستگی بین یک مجموعه ژنی معنی‌دار زیستی با فنوتیپ مورد آزمون قرار می‌گیرد (Wang *et al.*, 2011). در حقیقت در این روش به دنبال ژن‌هایی هستیم که به تنهایی اثر آنها بر صفت مورد نظر معنی‌دار نشده و دارای آثار متوسطی هستند، ولی اثر تجمعی آنها روی صفت دارای اثر معنی‌دار است. برای اینکه بتوان تفسیر درستی از کنار هم قرار دادن این ژن‌ها حاصل شود، از مسیرهای زیستی به عنوان بستری معنی‌دار که عملکرد مجموع ژن-ها در آنها یک یا چند هدف واحد را پیگیری می‌کند استفاده می‌شود (Mooney and Wilmot, 2015).

برای اولین بار (Peñagaricano *et al.*, 2013) نشان دادند که تجزیه 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 تجزیه 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 تجزیه غنی‌سازی مجموعه‌های ژنی روی خصوصیات لخته‌شدگی شیر، تولید پنیر و استحکام دلمه انجام شد. بر اساس این نتایج، تجزیه بر اساس مسیر منجر به شناسایی ۲۱ طبقه مختلف عملکردی هستی‌شناسی ژن و ۱۷ مسیر بیوشیمیایی KEGG معنی‌دار مرتبط با این صفات شد (Dadousis *et al.*, 2017). در پژوهشی دیگر که با استفاده از داده‌های حساسیت به ویروس لکومی از یک جمعیت گاو شیری هلشتاین انجام شده بود روش پویش ژنومی بر مبنای مسیر، کارایی بهتری برای یافتن مناطق ژنومی، درک بهتر ساز و کار بیماری و معماری ژنتیکی بیماری نسبت به تجزیه بهترین پیش-بینی ناریب خطی ژنومی داشت (Abdalla *et al.*, 2016). هدف از انجام پژوهش حاضر، تجزیه و تحلیل مجموعه-های ژنی و مسیرهای زیستی مرتبط با صفات وزن بدن در سنین مختلف چهار نژاد مرغ بر اساس پویش کل ژنومی بر مبنای مسیر است.

مواد و روش‌ها

در این پژوهش ابتدا فایل‌های تعیین ژنوتیپ با استفاده از تراشه‌های ژنومی همراه با رکوردهای فنوتیپی مرتبط با صفات وزن بدن که مربوط به گونه مرغ بوده و در پایگاه-های ذخیره‌ای عمومی مختلف داده‌های ژنومی (Dryad،

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

۱- تعیین مکان SNP‌ها با ژن‌ها: SNP‌هایی که مقدار P -value آنها کمتر از ۰/۰۵ بود با استفاده از بسته نرم افزاری *biomaR2* (Durinck et al., 2009) 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 ۲۰kb 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 ۰/۰۵ باشد، ژن معنی‌دار به شمار می‌آید.

۲- ارتباط ژن‌ها به طبقات عملکردی و مسیرهای بیوشیمیایی: جهت تعیین طبقات عملکردی ژنی و مسیرهای سوخت و سازی و تنظیمی ژن‌های معنی‌دار از نرم افزار *GO2MSIG* (Powell, 2014) استفاده شد. برنامه *GO2MSIG* از پنج پایگاه‌های اطلاعاتی شامل هستی‌شناسی ژن (<http://www.geneontology.org>), مسیره‌های (<http://www.genome.jp/kegg/>), مسیره‌های (<http://www.pantherdb.org>), مسیره‌های (<http://www.metacyc.org>) و (<http://www.reactome.org>) جهت تعیین طبقات عملکردی و مسیرهای بیوشیمیایی استفاده می‌نماید. 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 سه فرآیند هستی‌شناسی شامل فرآیندهای زیستی، عملکرد مولکولی و اجزای سلولی در نظر گرفته شوند.

۳- پویش کل ژنومی بر پایه تجزیه مسیر: ارتباط‌های معنی‌دار مسیرهای عملکردی با صفات مرتبط با وزن بدن با استفاده از توزیع فوق هندسی^۵ و آماره Fisher's exact

(Frontiersin, Animal Genome, Zenodo, NABIC ذخیره شده بودند، استخراج و بر حسب اطلاعات مفید گروه‌بندی انجام شد. در نهایت از اطلاعات مرتبط با وزن بدن از هج تا ۱۵ هفتگی در چهار نژاد مرغ چوآ^۱، سیلک^۲، لنگشن^۳ و بیرد^۴ با هدف پویش کل ژنومی که بر اساس مدل خطی تک نشانگری و هاپلوتیپ بود، استفاده شد. برای دستیابی به اطلاعات رکوردهای فنوتیپی از سایت <https://www.frontiersin.org/articles/10.3389/fgene.2018.00229/full#supplementary-material> استفاده شد. دو نژاد چوآ و سیلک با وزن کم بدن و نژادهای لنگشن و بیرد برای وزن بدن بالا انتخاب شده بودند. اطلاعات شامل تعداد ۴۰۲ قطعه مرغ از هر نژاد و ۱۰۰ نمونه تقریبی خروس و مرغ از هر نژاد انتخاب شده بودند (Yuan et al., 2018). رکوردهای فنوتیپی شامل وزن بدن جوجه‌ها از زمان هج تا ۱۵ هفتگی بود که وزن‌کشی به صورت هفتگی انجام شده بود. DNA نمونه‌ها با استفاده از آرایه‌های Illumina 60K chicken بر اساس پروتکل استاندارد Illumina تعیین ژنوتیپ شد. آرایه‌های طراحی شده امکان تعیین ژنوتیپ ۵۷۶۳۶ جایگاه نشانگری SNP در مرغ را فراهم می‌آورد.

نشانگرهای SNP که از تمام مراحل کنترل کیفیت شامل نشانگرهای با حداقل فراوانی آلی بالاتر از ۰/۰۵ و میزان فراخوانی آلی بالاتر از ۰/۹۰، کنار گذاشتن SNP‌های کروموزوم Z که ژنوتیپ آنها در افراد ماده به صورت هتروزیگوت بودند و SNP‌های با موقعیت نامشخص روی کروموزوم گذر کردند برای تجزیه‌های بعدی نگه داشته شدند. جهت ارتباط فنوتیپ‌ها با ژنوتیپ‌ها از نرم افزار GenABEL (Aulchenko et al., 2007) در برنامه R استفاده شد. مدل مورد استفاده بر پایه مدل خطی مختلط تک‌صفتی به شکل زیر بود:

$$y = \mu + F\beta + X_S a_S + G + e$$

که y : بردار مشاهدات فنوتیپی، μ : اثر میانگین، β : آثار ثابت شامل وزن هج، جنس و نژاد، F : ماتریس طرح که مشاهدات را به آثار ثابت ربط می‌دهد، a_S : اثر ژنوتیپ نشانگر، X_S : بردار ژنوتیپ‌های SNP، G : اثر تصادفی پلی-ژنیک و e : اثر تصادفی باقیمانده.

1. Chahua
2. Silkie
3. Langshan
4. Beard

بود ($P < 0.01$)، علیرغم اینکه وزن بدن در تمامی هفته‌ها بین نژادهای چوآ و سیلک (نژادهای با وزن بدن کم) تفاوت معنی‌داری نداشت ($P > 0.05$). در حالی که وزن بدن از ۳ تا ۱۴ هفتگی بین نژاد لنگشن و بیرد تفاوت معنی‌داری داشت ($P < 0.05$)، که نشان‌دهنده تنوع بالایی در وزن بدن برای تجزیه‌های مطالعه پویش کل ژنومی بود. همچنین در نهایت بعد از کنترل کیفیت، تعداد ۳۸۸ فرد و تعداد ۴۶۲۱۱ SNP برای تجزیه‌های پویش کل ژنومی بر پایه تجزیه و تحلیل غنی‌سازی مجموعه ژنی باقی ماندند.

شناسایی مسیرهای بیوشیمیایی با استفاده از تجزیه $GSEA-SNP^R$. تعداد ۱۵۷۲۳ ژن از ۱۶۸۷۸ ژن کدکننده حاشیه‌نویسی شده در ژنوم مرغ به وسیله نشانگرهای SNP پوشش داده شد که در این میان مجموع ۲۸۵۰ ژن دارای اثر معنی‌دار روی صفات وزن بدن از هیچ تا ۱۵ هفتگی بودند، یعنی حداقل یک نشانگر با P -value کمتر از ۰/۰۵ در داخل و یا در بالا یا پایین دست این ژن تا فاصله ۲۰ kb قرار گرفت. این ژن‌ها به عنوان ژن‌های معنی‌دار مرتبط با صفات وزن بدن برای تجزیه $GSEA-SNP$ انتخاب شدند.

تعداد مجموعه‌های ژنی حاصل از پایگاه‌های داده‌ای مختلف شامل ۱۴۱۷ طبقات هستی‌شناسی، ۱۲۴ مسیر بیوشیمیایی PANTHER و مسیر Reactome بود. همان‌طور که در جدول ۲ مشاهده می‌شود تعداد ۱۷ طبقه عملکردی در هستی‌شناسی فرآیندهای زیستی، عملکرد مولکولی، اجزای سلولی و مسیرهای بیوشیمیایی با صفات وزن بدن در هفته‌های مختلف از ۱ تا ۱۵ هفتگی دارای ارتباط هستند ($P < 0.01$). مسیرهای که بیشتر از سه ژن و کمتر از ۳۰۰ ژن داشتند گزارش شده‌اند.

test مورد آزمون قرار گرفت. P -value مسیرهای عملکردی که تعداد k ژن معنی‌دار در آن قرار دارد بر اساس رابطه زیر محاسبه شد:

$$P - \text{value} = 1 - \sum_{i=1}^{k-1} \frac{\binom{S}{i} \binom{N-S}{m-i}}{\binom{N}{m}}$$

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 تجزیه شدند و m برابر با تعداد کل ژن‌های موجود در مسیر عملکردی هستند (Dadousis et al., 2017). تجزیه غنی‌سازی مجموعه ژنی با استفاده از بسته نرم افزاری goseq (Young et al., 2010) در نرم افزار R انجام شد. بسته goseq از آزمون فوق هندسی استفاده می‌کند. همچنین 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^۱ براساس ضریب غنی‌سازی^۲ مطابق رابطه زیر استفاده شد:

$$ES = \max_{1 \leq j \leq N} \left\{ \frac{\sum_{G_{i^*} \in S, i^* \leq i} \frac{|r(i^*)|^p}{N_R} - \sum_{G_{i^*} \in S, i^* \leq i} \frac{1}{N - N_S}}{N - N_S} \right\}$$

مراحل کاری تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی در شکل ۱ ارائه شده است. برای تفسیر بهتر عملکرد ژن‌های بدست آمده از پایگاه‌های اطلاعاتی آنلاین GeneCards (<http://www.genecards.org>) و UniProtKB (<http://www.uniprot.org>) استفاده شد.

نتایج و بحث

آمار توصیفی صفت وزن بدن از هیچ تا ۱۵ هفتگی به تفکیک هر هفته در جدول ۱ نشان داده شده است. ۸ فرد (۲ سیلک، ۴ لنگشن و ۲ بیرد) به علت داشتن داده‌های پرت در رکورد‌های فنوتیپی از تجزیه‌های بعدی حذف شدند. همان‌طور که از اعداد جدول ۱ مشخص است، وزن بدن ۱۵ هفتگی در نژادهای لنگشن و بیرد به طور معنی‌داری نسبت به وزن بدن نژادهای چوآ و سیلک سنگین‌تر

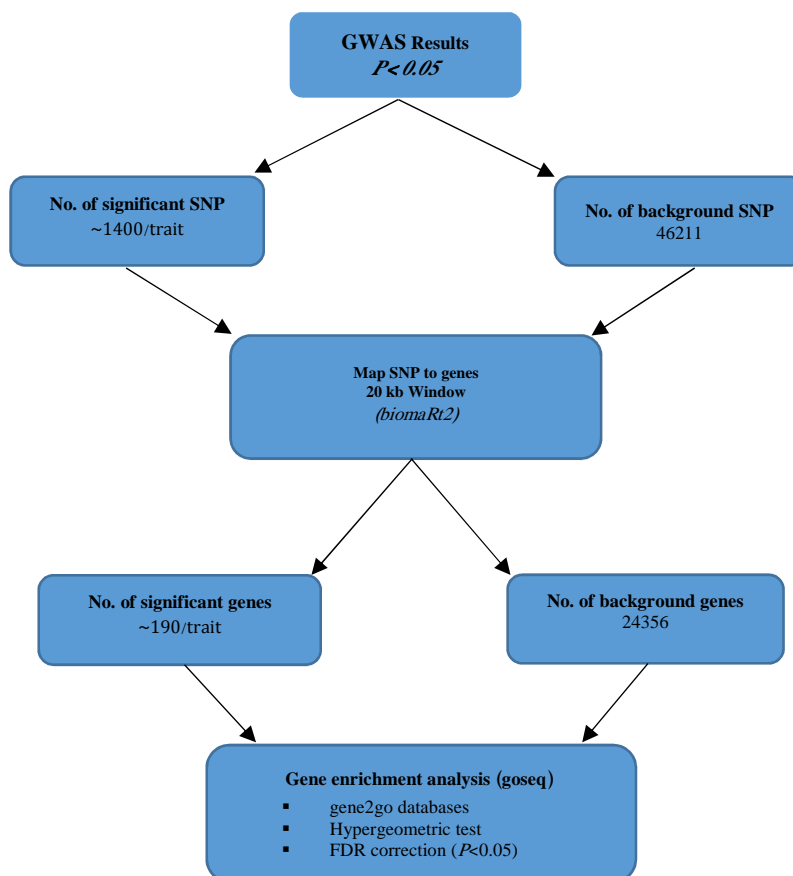


Fig. 1. Flowchart for the gene set enrichment analysis
شکل ۱- مراحل کاری تجزیه و تحلیل غنی‌سازی مجموعه ژنی

جدول ۱- آمار توصیفی صفات وزن بدن در چهار نژاد مورد مطالعه

Table 1. Descriptive statistics of the body weight traits in four under study breeds

Trait	Breed			
	Chahua (g) Mean (g) ±SD	Silkie (g) Mean (g) ±SD	Langshan (g) Mean (g) ±SD	Beard (g) Mean (g) ±SD
Birth -weight	21.9 ± 1.5	24.4 ± 1.9	35.7 ± 2.3	35.8 ± 3.5
BW1	46.7 ± 4.2	49.5 ± 5.3	66.8 ± 5.3	66.6 ± 6.9
BW2	73.4 ± 8.7	76.8 ± 8.3	126.2 ± 11.4	127.6 ± 14.3
BW3	113.9 ± 15.6	119.2 ± 17.4	201.9 ± 18.4	209.1 ± 27.8
BW4	159.1 ± 24.6	153.4 ± 24.8	299.7 ± 30.3	314.5 ± 43.7
BW5	211.4 ± 33.3	221.7 ± 39.1	391.5 ± 44.0	427.3 ± 64.5
BW6	277.9 ± 44.2	283.2 ± 49.8	482.9 ± 58.0	538.4 ± 90.0
BW7	349.4 ± 55.6	341.6 ± 64.5	582.7 ± 82.5	680.5 ± 124.4
BW8	396.3 ± 67.8	414.9 ± 84.4	701.3 ± 99.1	803.6 ± 141.6
BW9	471.4 ± 82.6	485.6 ± 97.6	859.6 ± 119.0	965.3 ± 167.6
BW10	547.7 ± 96.3	569.5 ± 111.4	1005.7 ± 140.4	1091.4 ± 197.0
BW11	613.7 ± 112.0	628.4 ± 126.9	1125.2 ± 172.7	1207.9 ± 214.5
BW12	689.9 ± 127.0	700.5 ± 143.7	1262.5 ± 204.8	1365.4 ± 239.2
BW13	740.9 ± 131.1	768 ± 156.3	1334.7 ± 227.1	1436.3 ± 242.8
BW14	791.9 ± 139.6	828.6 ± 164.4	1471.6 ± 250.5	1542.1 ± 251.8
BW15	817.2 ± 143.9	844.8 ± 169.7	1581.9 ± 277.5	1617.1 ± 240.7

جدول ۲- تجزیه و تحلیل غنی‌سازی مجموعه‌های ژنی معنی دار ($P < 0.01$) مرتبط با صفات وزن بدن

Table 2. Gene set enrichment analysis significantly ($P < 0.01$) associated with body weight

GO ID	Term (GO hierarchy level)	No. genes in the GO term	No. significant genes	FDR
Biological process				
GO:0006099	Tricarboxylic acid cycle	27	19	4.33E-02
GO:0010720	Positive regulation of cell development	42	26	4.32E-02
GO:0050793	Regulation of developmental process	125	62	3.12E-02
GO:0045597	Positive regulation of cell differentiation	49	30	3.05E-02
GO:0051130	Positive regulation of cellular component organization	54	32	3.42E-02
GO:1901135	Carbohydrate derivative metabolic process	262	115	3.38E-02
GO:0048856	Anatomical structure development	320	139	2.34E-02
GO:0006464	Cellular protein modification process	334	242	1.49E-03
Molecular Function				
GO:0008092	Cytoskeletal protein binding	377	159	2.16E-02
GO:0004674	Protein serine/threonine kinase activity	343	160	7.11E-04
GO:0003690	Double-stranded DNA binding	98	13	3.89E-03
Cellular component				
GO:0030054	Cell junction	130	65	1.30E-02
GO:0034703	Cation channel complex	91	46	3.80E-02
PANTHER Pathways				
P00018	EGF receptor signaling pathway	133	67	1.22E-02
P00052	TGF-beta signaling pathway	104	52	2.74E-02
P00021	FGF signaling pathway	119	58	2.60E-02
Reactome pathways				
R-CEL-446203	Asparagine N-linked glycosylation	259	116	3.10E-02

از مسیرهای زیستی مهم و معنی‌دار مرتبط با وزن بدن می‌توان به مسیرهای cytoskeletal protein binding و positive regulation of cell development اشاره نمود، که از بین ژن‌های معنی‌دار در این مسیرها، ژن کاندیدای *MYOD1* جزء ابر خانواده ژنی میوزین در سلول‌های عضله اسکلتی است. *MYOD1* پروتئینی است که نقش عمده‌ای در تنظیم عضلانی دارد و به عنوان یک عامل رونویسی به طور مستقیم بیان و تفرق سلول‌های عضلانی را تنظیم می‌کند. این ژن به عنوان یک عامل نظارتی در فرآیند میوزنیزاسیون^۱ در سطوح اولیه بیان می‌شود. در مطالعه قبلی پویش ژنومی در سطح mRNA با استفاده از تکنیک RNA-Seq با هدف شناسایی ژن‌های کاندیدای مرتبط با رشد در مرغ، مشخص شد ژن *MYOD1* نقش کلیدی در توسعه عضله مرغ دارد (Chen et al., 2015). در مطالعه قبلی پویش ژنومی صفات مرتبط با وزن بدن که بر اساس مدل خطی مختلط تک متغیره و به وسیله نرم افزار

GenABEL انجام شد، ارتباط وزن بدن با سنین ۸، ۱۰، ۱۱ و ۱۲ هفتگی روی کروموزوم ۵ (*GGA5*) گزارش شده بود که با منطقه شناسایی شده در پژوهش حاضر همخوانی داشت (Yuan et al., 2018). همچنین از مسیرهای معنی‌دار مرتبط با وزن بدن می‌توان به مسیر Tricarboxylic acid cycle اشاره نمود که از بین ژن‌های موجود در این مسیر، ژن کاندیدای *ABCG1* نقش کلیدی در سوخت و ساز و هموستازی چربی در بدن دارد. در مطالعه پویش ژنومی با هدف شناسایی ژن‌های کاندیدای مرتبط با افزایش وزن روزانه شناسایی شد (Seabury et al., 2017). همچنین ژن کاندیدای *PNPLA2* که در سوخت و ساز گلیسرولیپید نقش دارد در مطالعه تفرق بیان ژن در مرغ جزء ژن‌های کاندیدای مرتبط با وزن بدن بود (Chen et al., 2015). همچنین ناحیه معنی‌دار مؤثر روی وزن بدن در ناحیه ۴۹ مگابازی

است (Abdalla *et al.*, 2018). مسیر زیستی anatomical structure development از دیگر مسیرهایی بود که در ارتباط با وزن بدن شناسایی شدند. ژن‌های کاندیدای *MYO1B*، *MYO1C* و *MYO1E* در توسعه بافت عضله در لاین پدری جوجه گوشتی با هدف شناسایی نشانه‌های انتخاب مرتبط با صفات عملکردی شناسایی شدند (Almeida *et al.*, 2019).

از ژن‌های کاندیدای دیگر می‌توان به ژن‌های *MYL2* و *MYL3* اشاره نمود که جزء ژن‌های موثر بر ذخیره چربی در بدن و چربی داخل ماهیچه‌ای در جوجه‌های گوشتی تجاری با تفرق بیان ژنی معنی‌دار بودند (Cui *et al.*, 2012). همچنین در مطالعه‌ای روی بلدرچین ژاپنی مشخص شد ژن‌های کاندیدای *MYL1* و *MYL3* نقش کلیدی در سوخت و ساز کلسیم در تفرق عضله دارند (Park *et al.*, 2018). علاوه بر این، در مطالعه تجزیه پروتئوم با تکنیک وسترن بلات عضلات اسکلتی در مرغ، پروتئین‌های کاندیدای *MYL1* و *MYL3* در رشد عضلات اسکلتی در دوران جنینی نقش کلیدی داشتند (Ouyang *et al.*, 2017). همچنین ناحیه معنی‌دار مؤثر روی وزن بدن در هفته‌های ۴ و ۶ هفتگی در ناحیه ۱۴ مگابازی کروموزوم ۲ بدست آمد که ژن کاندیدای *MYL3* با مطالعه قبلی مشترک بود (Yuan *et al.*, 2018). مسیر معنی‌دار دیگر مرتبط با وزن بدن Regulation of developmental process است که از بین ۶۲ ژن، ژن کاندیدای ژن‌های کاندیدای *ACOX1* و *ACOX2* در سوخت و ساز لیپید و اسیدهای چرب در بدن مرغ نقش دارند (Resnyk *et al.*, 2015).

در جدول ۳ نتایج خوشه‌بندی ژن‌های شناسایی شده در مناطق ژنومی معنی‌دار ارائه شده است. خوشه اول با دارا بودن ژن‌های مؤثر بر وزن بدن (۴۱ ژن) دارای بیشترین مقدار ضریب غنی‌سازی (۲/۹۶) و مقدار *P*-value برابر با $7.9E-5$ را به خود اختصاص داد. ژن‌ها در سایر خوشه‌ها به جز خوشه ۳ نیز به طور مستقیم در افزایش وزن بدن نقش داشتند. از آنجا که ژن‌های شناسایی شده در این خوشه‌ها به طور مستقیم در رشد و توسعه عضلات اسکلتی و سوخت و ساز چربی و رشد و نمو اندام‌های بدن نقش داشتند و همچنین به طور غیر مستقیم در فرایند تکامل و رشد و نمو اندام‌های بدن نقش دارند، باید نقش و

کروموزوم ۵ بدست آمد که ژن کاندیدای *PNPLA2* با مطالعه قبلی مشترک بود (Yuan *et al.*, 2018).

شاید بتوان مسیر cell junction که جزء هستی‌شناسی اجزای سلولی است را یکی از مهمترین مسیرهای مؤثر بر فرآیند ارتباط وزن بدن با صفات رشد دانست. در این مسیر، پروتئین‌های واسطه جهت ارتباط دادن اسکلت سلولی یک سلول به اسکلت سلولی مجاور یا یک پروتئین در ماتریکس خارج سلولی از سطح غشای پلاسمایی گسترش می‌یابند. این مسیر دارای چندین chiled Term است که همه دارای عملکردهای مهمی در فرآیندهای ارتباط و اتصال سلول به هم و یا به بافت همبند پایه هستند. از بین ژن‌های معنی‌دار در این مسیر، می‌توان به ژن‌های کاندیدای *MYH10* و *MYH11* اشاره نمود. عملکرد مولکولی ژن *MYH10* اتصال به فیلامنت-های آکتین، اتصال به مولکول ATP و تولید نیرو در میکرو فلانمت‌ها و میکرو توبول‌ها است. در مطالعه تجزیه ترانسکریپتوم با استفاده از تکنیک RNA-Seq با هدف شناسایی ژن‌های کاندیدای مرتبط با توسعه عضله سینه مرغ، ژن زنجیره سنگین میوزین (*MYH11*) جزء ژن‌های کاندیدای شناسایی شده بود (Li *et al.*, 2019). همچنین در تجزیه ترانسکریپتوم بافت عضله مرغ، ژن کاندیدای *MYH10* به عنوان ژن کلیدی در رشد الیاف عضلانی شناسایی شده بود (Ye *et al.*, 2017).

از دیگر مسیرهای زیستی معنی‌دار مرتبط با وزن بدن می‌توان به مسیر Carbohydrate derivative metabolic process اشاره نمود که جزء مسیرهای فرآیندهای زیستی مرتبط با تولید است، که ژن‌های کاندیدای *SLC2A8* و *ACACA* دارای نقش زیستی مستقیمی با صفات مرتبط با تولید هستند که بیشتر بحث خواهد شد. ژن کاندیدای *SLC2A8* دارای نقش اساسی در انتقال گلوکز به داخل سلول‌ها و تنظیم هموستازی گلوکز است. ارتباط معنی‌داری بین بیان ژن *SLC2A8* در دوران هج و توسعه عضله مرغ گزارش شده است (Coudert *et al.*, 2018). ژن *ACACA* آنزیم کلیدی در تنظیم تولید اسید چرب در بدن است که از راه افزایش در فعالیت استیل کوآ کربوکسیلاز آلفا، موجب افزایش تولید مالونیل کوآ از مسیر کربوکسیله کردن استیل کوآ و در نتیجه افزایش تولید اسیدهای چرب می‌شود. ارتباط معنی‌داری بین ژن *ACACA* با وزن چربی داخل احشایی مرغ گزارش شده

توان با ادغام داده‌های تحقیقات مشابه جدید و استفاده از تجزیه‌های آماری جامع‌تر برای تأیید نتایج پژوهش 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 به صورت 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 وزن بدن 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.

عملکرد آنها از راه مطالعات آزمایشگاهی بیشتر مورد بررسی قرار گیرند. با بررسی منابع انجام شده، پژوهش حاضر اولین مطالعه پویش کل ژنومی بر پایه تجزیه مسیر مرتبط با وزن بدن در مرغ بوده است، لذا در تجزیه و تحلیل‌ها تلاش شد از آزمون‌های پر کاربرد و سختگیرانه (مانند برنامه GO2MSIG) برای شناسایی مسیرهای زیستی معنی‌دار استفاده شود. به هر حال باید توجه کرد که یکی از محدودیت‌های اجرای مطالعات پویش ژنومی، کافی نبودن تعداد رکوردهای فنوتیپی و جمعیت غیر خویشاوند برای تجزیه پیوستگی که منجر به افزایش خطای نوع اول یا به عبارتی شناسایی نشانگرهای معنی‌دار مثبت کاذب می‌شود بود (Seabury *et al.*, 2017). با توجه به تعداد کم نمونه مورد استفاده در پژوهش حاضر (۳۸۸ مرغ) باید در استفاده از ژن‌های کاندیدای شناسایی شده مرتبط با وزن بدن در برنامه‌های اصلاحی با احتیاط عمل کرد. البته می‌

جدول ۳- خوشه‌بندی ژن‌های معنی‌دار شناسایی شده از هستی‌شناسی فرآیندهای زیستی و عملکردهای مولکولی

Table 3. Enriched (DAVID Enrichment score ES>2) functional clusters of Gene Ontology (GO) biological processes (BP) and molecular functions (MF)

Annotation Cluster 1	Enrichment Score: 2.96	Count	P-value	Benjamini
INTERPRO	Myosin tail 2	3	7.9E-5	5.7E-2
GOTERM_CC_DIRECT	myosin complex	10	2.2E-4	3.6E-2
UP_KEYWORDS	Myosin	8	4.3E-4	1.5E-1
INTERPRO	Myosin head, motor domain	6	3.2E-3	1.6E-1
GOTERM_MF_DIRECT	motor activity	5	8.8E-4	7.0E-2
UP_KEYWORDS	Motor protein	9	7.9E-2	6.4E-1
Annotation Cluster 2	Enrichment Score: 2.31			
GOTERM_BP_DIRECT	lipid homeostasis	7	8.4E-4	9.2E-1
GOTERM_MF_DIRECT	fatty-acyl-CoA binding	7	2.2E-3	8.2E-1
KEGG_PATHWAY	PPAR signaling pathway	8	5.0E-4	8.5E-2
GOTERM_MF_DIRECT	acyl-CoA dehydrogenase activity	5	6.6E-3	9.7E-1
Annotation Cluster 3	Enrichment Score: 2.14			
GOTERM_MF_DIRECT	phosphatidic acid binding	6	5.1E-3	8.6E-1
GOTERM_MF_DIRECT	phosphatidylinositol-3,5-bisphosphate binding	7	6.4E-3	8.5E-1
GOTERM_MF_DIRECT	phosphatidylinositol-3,4-bisphosphate binding	6	1.9E-2	9.0E-1
GOTERM_MF_DIRECT	phosphatidylinositol-3,4,5-trisphosphate binding	5	2.5E-4	4.9E-1
Annotation Cluster 4	Enrichment Score: 2.09			
UP_KEYWORDS	Fatty acid biosynthesis	7	2.9E-3	9.8E-1
UP_KEYWORDS	Fatty acid metabolism	7	3.2E-3	9.4E-1
GOTERM_BP_DIRECT	very long-chain fatty acid biosynthetic process	4	5.4E-3	7.3E-1
GOTERM_MF_DIRECT	fatty acid elongase activity	3	2.1E-3	1.0E0
KEGG_PATHWAY	Fatty acid elongation	6	3.4E-3	9.0E-1

تجزیه و تحلیل غنی‌سازی مجموعه ژنی برای پویش ژنومی صفات تولیدی اقتصادی را نیز مورد تأیید قرار داد. همچنین با بررسی چندشکلی موجود در ژن‌های کاندیدای شناسایی شده از تجزیه و تحلیل غنی‌سازی مجموعه ژنی مرتبط با وزن بدن از راه مطالعات آزمایشگاهی در نژادهای مرغ بومی و نتایج بدست آمده را برای مطالعات اصلاح نژادی به کار برد.

تشکر و قدردانی

نویسندگان مقاله از دکتر یوآن و همکاران در گروه علوم دامی دانشگاه کشاورزی هوژن به خاطر فراهم نمودن اطلاعات مورد نیاز این تحقیق صمیمانه تقدیر و تشکر می‌نمایند.

استفاده از ژن‌های عمده اثر شناسایی شده در پژوهش حاضر و سایر پژوهش‌های مرتبط در نژادهای مختلف مرغ (Emrani *et al.*, 2017) می‌تواند در این زمینه کاربرد داشته باشد.

نتیجه‌گیری کلی

بررسی مناطق ژنومی روی کروموزوم‌های ۱، ۲، ۵، ۷، ۱۰، ۱۴، ۱۸، ۱۹، ۲۰ و ۲۷ با استفاده از پایگاه‌های داده مختلف نشان داد که این مناطق با صفات مرتبط با وزن بدن در سنین مختلف مرتبط هستند. با توجه به عملکرد زیستی 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

فهرست منابع

- Abdalla E. A., Peñagaricano F., Byrem T. M., Weigel K. A. and Rosa G. J. 2016. Genome-wide association mapping and pathway analysis of leukosis incidence in a US Holstein cattle population. *Animal Genetics*, 47: 395-407.
- Abdalla B. A., Chen J., Nie Q. and Zhang X. 2018. Genomic insights in to the multiple factors controlling abdominal fat deposition in a chicken model. *Frontiers in Genetics*, 19(9): 262-271.
- Almeida O. A. C., Moreira G. C. M., Rezende F. M., Boschiero C., de Oliveira Peixoto J., Ibelli A. M. G., Ledur M. C., de Novais F. J. and Coutinho L. L. 2019. Identification of selection signatures involved in performance traits in a paternal broiler line. *BMC Genomics*, 20(1): 449-461.
- Aulchenko Y. S., Ripke S., Isaacs A. and van Duijn C. M. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23: 1294-1296.
- Chen B., Xu J., He X., Xu H., Li G., Du H., Nie Q. and Zhang X. 2015. A genome-wide mRNA screen and functional analysis reveal FOXO3 as a candidate gene for chicken growth. *PLoS One*, 10(9): e0137087.
- Coudert E., Praud C., Dupont J., Crochet S., Cailleau-Audouin E., Bordeau T., Godet E., Collin A., Berri C., Tesseraud S. and Métayer-Coustard S. 2018. Expression of glucose transporters SLC2A1, SLC2A8, and SLC2A12 in different chicken muscles during ontogenesis. *Journal of Animal Science*, 96(2): 498-509.
- Cui H. X., Liu R. R., Zhao G. P., Zheng M. Q., Chen J. L. and Wen J. 2012. Identification of differentially expressed genes and pathways for intramuscular fat deposition in pectoralis major tissues of fast-and slow-growing chickens. *BMC Genomics*, 13: 213-220.
- Dadousis C., Pegolo S., Rosa G. J. M., Gianola D., Bittante G. and Cecchinato A. 2017. Pathway-based genome-wide association analysis of milk coagulation properties, curd firmness, cheese yield, and curd nutrient recovery in dairy cattle. *Journal of Dairy Science*, 100: 1223-1231.
- Durinck S., Spellman P. T., 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., Vaez Torshizi R., Masoudi A. A. and Ehsani A. R. 2017. Identification of new loci for body weight traits in F₂ chicken population using genome-wide association study. *Livestock Science*, 206: 125-131.
- Li Y., Chen Y., Jin W., Fu S., Li D., Zhang Y., Sun G., Jiang R., Han R., Li Z., Kang X. and Li G. 2019. Analyses of microRNA and mRNA expression profiles reveal the crucial interaction networks and pathways for regulation of chicken breast muscle development. *Frontiers in Genetics*, 10: 197-205.
- Mooney M. A. and Wilmot B. 2015. Gene Set Analysis: A Step-By-Step Guide. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 168: 517-527.
- Ouyang H., Wang Z., Chen X., Yu J., Li Z. and Nie Q. 2017. Proteomic analysis of chicken skeletal muscle during embryonic development. *Frontiers in Physiology*, 8: 281-289.
- Park J. W., Lee J. H., Kim S. W., Han J. S., Kang K. S., Kim S. J. and Park T. S. 2018. Muscle differentiation induced up-regulation of calcium-related gene expression in quail myoblasts. *Asian-Australasian Journal of Animal Science*, 31(9): 1507-1515.
- Peñagaricano F., Weigel K. A., Rosa G. J. and Khatib H. 2013. Inferring quantitative trait pathways associated with bull fertility from a genome-wide association study. *Frontiers in Genetics*, 3: 307-314.

- Peng G., Luo L., Siu H., Zhu Y., Hu P., Hong S., Zhao J., Zhou X., Reveille J. D., Jin L., Amos C. I. and Xiong M. 2010. Gene and pathway-based second wave analysis of genome-wide association studies. *European Journal of Human Genetics*, 18: 111-117.
- Powell J. A. 2014. GO2MSIG, an automated GO based multi-species gene set generator for gene set enrichment analysis. *BMC Bioinformatics*, 15: 146-149.
- Resnyk C. W., Chen C., Huang H., Wu C. H., Simon J., Le Bihan-Duval E., Duclos M. J. and Cogburn L. A. 2015. RNA-Seq analysis of abdominal fat in genetically fat and lean chickens highlights a divergence in expression of genes controlling adiposity, hemostasis, and lipid metabolism. *PLoS One*, 9(10): e0139549.
- Seabury C. M., Oldeschulte D. L., Saatchi M., Beever J. E., Decker J. E., Halley Y. A., Bhattarai E. K., Molaei M., Freetly H. C., Hansen S. L., Yampara-Iquise H., Johnson K. A., Kerley M. S., Kim J., Loy D. D., Marques E., Neibergs H. L., Schnabel R. D., Shike D. W., Spangler M. L., Weaber R. L., Garrick D. J. and Taylor J. F. 2017. Genome-wide association study for feed efficiency and growth traits in U.S. beef cattle. *BMC Genomics*, 18(1): 386-396.
- Wang L., Jia P., Wolfinger R. D., Chen X. and Zhao Z. 2011. Gene set analysis of genome-wide association studies: Methodological issues and perspectives, *Genomics*, 98: 1-8.
- Young M. D., Wakefield M. J., Smyth G. K. and Oshlack A. 2010. Method gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biology*, 11: 14-23.
- Yousefi Zonuz A., Alijani A., Mohammadi H., Rafat S. A. and Daghigh Kia H. 2013. Estimation of genetic parameters for productive and reproductive traits in Esfahan native chickens. *Journal of Livestock Science and Technologies*, 1: 36-40.
- Ye M., Ye F., He L., Luo B., Yang F., Cui C., Zhao X., Yin H., Li D., Xu H., Wang Y. and Zhu Q. 2017. Transcriptomic analysis of chicken Myozenin 3 regulation reveals its potential role in cell proliferation. *PLoS One*, 13(12): e0189476.
- Yuan Y., Peng D., Gu X., Gong Y., Sheng Z. and Hu X. 2018. Polygenic basis and variable genetic architectures contribute to the complex nature of body weight -A genome-wide study in four Chinese indigenous chicken breeds. *Frontiers in Genetics*, 2(9): 229-238.



Research paper

Gene set enrichment analysis to identify genes and biological pathways associated with body weight in chicken

A. H. Khaltabadi Farahani^{1*}, H. Mohammadi¹, M. H. Moradi¹, H. A .Ghasemi², I. Hajkhodadadi¹

1. Assistant Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran
2. Associate Professor, Department of Animal Sciences, Faculty of Agriculture and Natural Resources, Arak University, Arak, Iran

(Received: 04-11-2019 – Accepted: 03-03-2020)

Abstract

The aim of the present study was to conduct a genome wide association studies (GWAS) based on gene set enrichment analysis for identifying the loci associated with body weight in four chicken breeds including Chahua, Silkie, Langshan, and Beard using the high throughput single nucleotide polymorphisms (SNPs). Phenotypic and genotypic data were obtained from the Frontiersin online public repository. In this study, the 402 cocks and hens were used with body weight from 1 to 15 weeks using GenABEL software. 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 GO, KEGG, DAVID and PANTHER databases. Ten SNP markers on chromosomes 1, 2, 5, 7, 10, 14, 18, 19, 20 and 27 located in ABCG1, MYOD1, MYH10, MYH11, MYO1B, MYO1C, MYO1E, MYL1, MYL2, MYL3, SLC2A8, ACACA, ACOX1, ACOX2, and PNPLA2 genes were identified. Some of the genes were found are consistent with some previous studies and to be involved in biological pathways related to body weight. According to pathway analysis, 17 pathways from gene ontology and KEGG pathway were associated with the body weight ($P<0.01$). Among those pathways, the cytoskeletal protein binding, anatomical structure development and tricarboxylic acid cycle had significant association with skeletal muscle fiber and metabolism lipid traits. In total, this study supported previous results from GWAS of body weight; also revealed additional regions in the chicken genome associated with these economically important traits. The use of these findings can accelerate the genetic progress in the breeding programs.

Keywords: Pathway-based analysis, Candidate gene, Chicken, Body weight

*Corresponding author: amfarahanikh@gmail.com