

## شناسایی بیوانفورماتیکی تنوع ژنومی تعداد کپی‌ها (Copy number Variation) در مرغان تخمگذار و گوشتی

هما سلیمانی<sup>۱</sup>، جلیل شجاع غیاث<sup>۲</sup>، سید عباس رافت<sup>۳</sup> و صابر قنبری<sup>۳</sup>

تاریخ دریافت: ۱۴۰۰/۸/۲۴ تاریخ پذیرش: ۱۴۰۰/۱۲/۱۵

<sup>۱</sup>دانشجو دکتری گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

<sup>۲</sup>استاد گروه علوم دامی، دانشکده کشاورزی، دانشگاه تبریز

<sup>۳</sup>موسسه ژنتیک و بیومتریک، موسسه زیست‌شناسی حیوانات مزرعه لاینیز، دامستورف، آلمان

\* مسئول مکاتبه: Email: soleimani\_homa@yahoo.com

### چکیده

زمینه مطالعاتی: تنوع تعداد کپی (CNV)، یکی از تغییرات ساختاری نامتعادل در ژنوم است که شامل جهش‌هایی از نوع حذف، اضافه شدن و تکرار بخش‌هایی از DNA در اندازه‌های مختلف از چند ده جفت باز تا چند مگا جفت باز است. بنابراین، این منبع مهم تنوع ژنتیکی، بر الگوهای بیان ژن و سپس، بر تنوع مشاهده شده در سطح فنوتیپی اثرگذار است. در این راستا، یک مطالعه‌ی جامع در مورد شناسایی تنوع تعداد کپی (CNV) در سطح ژنوم مرغ اهلی، می‌تواند اطلاعات ارزشمندی در مورد تنوع ژنتیکی بین نژادها و ارتباط بین این تغییرات ساختاری و صفات مهم اقتصادی در طیور را ارائه دهد. هدف: هدف از انجام پژوهش حاضر، شناسایی انواع تنوع در تعداد کپی (CNV) در سرتاسر ژنوم مرغ‌های گوشتی و تخم‌گذار بود. روش کار: در این مطالعه، یک مقایسه کلی بین مرغان تخمگذار و گوشتی انجام شد. بدین منظور، از داده‌های خام گزارش شده در مطالعه قنبری و همکاران (۲۰۱۹) که در مجموع شامل تعداد ۹۰ نمونه DNA با محتوای اطلاعاتی ۵۰ نمونه‌ی مرغ تخمگذار و ۴۰ نمونه‌ی مرغ گوشتی برای تعیین توالی یابی کل ژنوم استفاده شد. پس از هم‌ردیفی خوانش‌های خام فیلتر شده در ژنوم مرجع (شماره‌ی دسترسی در NCBI: GRCg6a)، از الگوریتم مبتنی بر عمق خوانش، برای شناسایی تنوع تعداد کپی‌ها استفاده شد. نتایج: نتایج بدست آمده از تجزیه و تحلیل‌های بیوانفورماتیکی بین ژنوم مرغان تپ گوشتی و تخمگذار، نشان داد که ۱۳ ناحیه از ۲۹ ناحیه بررسی شده فاقد هر نوع ژن و ناحیه کد شونده بوده و از طرفی ۱۶ ناحیه شناسایی شده دیگر حاوی ۲۸ ژن بود. از این میان، ۱۶ ژن شناسایی شده مربوط به RNAهای بلند غیر کدکننده بود ۱۰ ژن شناسایی شده مربوط به RNA ریبوزومی و ۱۲ ژن هم ژن‌های کدکننده پروتئین بودند. نتیجه گیری نهایی: به طور خلاصه، نتایج به دست آمده نشان دادند که ژن‌های مهمی از جمله ژن‌های DEDs و TNFAIP8 دخیل در مرگ برنامه ریزی شده سلول، دارای تنوع تعداد کپی هستند. همچنین دو ژن NPAL3 و RCAN که در سیستم ایمنی نقش دارند، در نمونه‌های مورد مطالعه، دارای تنوع تعداد کپی بودند. بعلاوه بسیاری از نقاط شناسایی شده حاوی lncRNA بودند که می‌تواند نشان دهنده اهمیت و تاثیر این نواحی بر افتراق دو نژاد متمایز گوشتی و تخمگذار باشد. لذا به نظر می‌رسد از شناسایی تنوع تعداد کپی و بررسی نواحی تنظیمی می‌توان در پژوهش‌های آینده برای اصلاح نژاد کمک گرفت.

واژگان کلیدی: تغییرات ساختاری، تنوع تعداد کپی، توالی یابی کل ژنوم، مرغان تخم‌گذار و گوشتی

## مقدمه

اخیراً، مطالعات ژنومی در حیوانات اهلی، بر شناسایی تنوع‌های جدید در ژنوم و اثر آن در عملکرد فنوتیپی صفات اقتصادی متمرکز شده‌اند. انواع تنوع ژنومی شامل جهش‌های از نوع حذف، اضافه، جای‌گزینی، بازترتیبی، معکوس شدن و تنوع در تعداد کپی قطعه‌ی ژنومی است. تنوع در تعداد کپی قطعه‌ی ژنومی شامل حذف، اضافه یا مضاعف شدن قطعه‌ای از DNA به طول یک کیلو جفت باز تا چندین مگاجفت باز است که شباهت میان این نواحی بیشتر از ۹۰ درصد است (بانکارپاش و همکاران ۲۰۱۸، پیروزنیا و همکاران ۲۰۱۵). مطالعات اخیر، نشان می‌دهد که این تنوع می‌تواند منجر به تغییر در دُز ژن، توالی‌های کُدکننده و تنظیم بیان ژن شود، لذا، اثر آن بر تنوع فنوتیپی باید بیش از جهش‌های تک نوکلئوتیدی باشد (آلکان و همکاران ۲۰۱۱). رودن و همکاران (۲۰۰۶) اولین نقشه CNV انسان را با استفاده از ۲۷۰ فرد حاضر در پروژه Hapmap ایجاد و مشاهده کردند. تقریباً، به طور متوسط ۱۲ درصد از کل ژنوم هر فرد را CNVها تشکیل می‌دهد که این حدوداً ۳۶۰ میلیون جفت باز را شامل می‌شد. بیشتر این CNVها اندازه نسبتاً کوچکی (تقریباً، ۷ درصد از ژنوم موش کمتر از ۲۰ کیلو باز) داشتند. (لاک و همکاران ۲۰۱۵)، ۱۲ درصد از ژنوم انسان (ردون و همکاران ۲۰۰۶)، ۷ درصد از ژنوم گاو (کیل و همکاران ۲۰۱۶) و ۵.۱۲ درصد از ژنوم مرغ (گورلا و همکاران ۲۰۱۷) حاوی CNV است و تقریباً، ۴۰ درصد این ساختارها، حداقل با یک ژن هم پوشانی دارد (کنراد و همکاران ۲۰۱۰). این نوع تنوع با بیماری‌هایی از قبیل شیذوفرنی، دیابت نوع اول و سرطان و غیره در انسان، مرتبط است (پیروزنیا و همکاران ۲۰۱۵).

مطالعات اخیر همچنین، نشان می‌دهد تنوع در تعداد کپی قطعه‌ی ژنومی در حیوانات مزرعه‌ای با تنوع فنوتیپی صفات مورفولوژیکی مرتبط است. مثلاً، تاج

نخودی در طیور، ناشی از تنوع در تعداد کپی قطعه‌ی ژنومی در اینترون یک ژن SOX5 می‌باشد (رایت و همکاران ۲۰۰۹). در گوسفند، جهش‌های از نوع مضاعف شدن نوکلئوتیدها در ژن ASIP منجر به تنظیم رنگ دانه در پوشش بدن می‌شود (نوریس و همکاران ۲۰۰۸). تأخیر در پردرآوری جوجه‌ها، به دلیل دوبله ناقص در ژن‌های PRLR و SPEF2 می‌باشد (الفرینگ و همکاران ۲۰۰۸). دو برابر شدن ۴/۶ کیلو جفت باز در اینترون شش ژن stx17 منجر به خاکستری شدن رنگ بدن در اسب، با افزایش سن می‌شود. حذف ناحیه‌ی بین ژنی به طول ۱۱/۷ کیلو جفت باز در ژنوم بز، منجر به حذف شاخ می‌شود (کلاپ و همکاران ۲۰۱۲). با توجه به پوشش قابل ملاحظه این تنوع در ژنوم و اثر آن بر تنوع فنوتیپی صفات، اهمیت مطالعه‌ی بیشتر روی این شکل تنوع، آشکارتر می‌شود. این پژوهش، با هدف بررسی نحوه‌ی پراکندگی CNVها در سراسر ژنوم و مقایسه‌ی تفاوت‌های CNV در نژادهای مختلف مرغ اهلی اجرا گردید.

## مواد و روش‌ها

## نمونه‌گیری و توالی‌یابی

در این مطالعه، از داده‌های مطالعه انجام شده توسط قنبری و همکاران (۲۰۱۹) شامل مجموع ۹۰ نمونه DNA با جزییات و محتوای اطلاعاتی تعداد ۵۰ نمونه‌ی مرغ تخمگذار و تعداد ۴۰ نمونه‌ی مرغ گوشتی برای توالی‌یابی کل ژنوم استفاده شد. پرندگان از نژادهای تجاری به ترتیب شامل دو گروه گوشتی و دو گروه تخم‌گذار بودند. جوجه‌های گوشتی شامل تعداد ۲۰ نمونه DNA که قبلاً به طور مستقل، به عنوان بخشی از پروژه AVIANDIV نمونه‌گیری شدند.

در گروه تخمگذار ۵۰ پرند ماده از دو جمعیت تجاری مرغان تخم‌گذار (LOHMANN Tierzucht GmbH) با پوشش سفید پوستی (n = 25, WL) و قهوه‌ای (n = 25, BL) در طرح پروژه SYNBREED توالی

### محاسبه‌ی عمق خوانش‌ها

در این مرحله، هدف، بررسی تعداد خوانش‌هایی است که به هر نقطه از ژنوم متصل شده. از آنجاییکه ژنوم به صورت تصادفی قطعه قطعه و توالی‌یابی شده است، انتظار می‌رود خوانش‌های تولید شده هم تقریباً به صورت یکسان در سطح ژنوم رفرنس نقشه‌یابی شود. بنابراین اگر در یک مکان تعداد زیادی خوانش (مثلاً دو یا چند برابر میانگین سایر مکان‌ها) به یک محل متصل شوند احتمالاً در آن نقطه، نسبت به ژنوم رفرنس قطعه‌ی ای DNA اضافه شده و یا اگر به قسمتی از ژنوم رفرنس هیچ خوانشی نقشه‌یابی نشود احتمالاً آن قطعه در نمونه مورد بررسی حذف شده است.

با استفاده از بسته‌ی نرم‌افزاری GATK و تابع Depth Of Coverage، عمق خوانش‌ها برای هر نمونه و در هر نمونه برای تمام قسمت‌های ژنوم محاسبه شد. آنجاییکه مقایسه ژنوم افراد در گروه‌های مختلف زمان بر و از نظر محاسباتی برای تمام ژنوم دشوار بود، ژنوم هر پرنده به پنجره‌هایی بدون هم‌پوشانی به اندازه  $1000 \text{ bp}$  تقسیم شد و متوسط عمق خوانش در هر پنجره محاسبه شد. سپس، یک ضریب تصحیح برای جمعیت ایجاد و آن را بر روی میانگین عمق خوانش برای هر پنجره اعمال شد. نتایج با استفاده از نرم‌افزار R و آنالیز واریانس بررسی شدند (کارنیرو و همکاران ۲۰۱۴) و مناطقی که بین نژادها تفاوت معنی‌دار  $P \leq 0.001$  داشتند مشخص شدند. در پایان آنالیز، هستی‌شناسی برای ژن‌های حاوی تنوع تعداد کپی با استفاده از نسخه‌ی ۶.۷ برنامه تحت وب DAVID (هانگ و همکاران ۲۰۰۹) انجام شد.

### نتایج و بحث

میانگین هم‌ردیفی خوانش‌ها با ژنوم رفرنس (Galgal6)  $98.61\%$  و میانگین عمق خوانش‌ها  $11.51$  بود. پس از محاسبه عمق خوانش‌ها در تمامی افراد، یک مقایسه کلی بین مرغان تخمگذار و مرغان گوشتی به کمک آنالیز واریانس صورت گرفت تا بتوان ژنوم

یابی شدند. پرندگان توالی‌یابی شده از لاین والدین خالص هیبریدهای تجاری لوهمن بودند که در ۶۸ هفته به ترتیب به ۲۹۰ و ۲۷۰ تخم می‌رسند. نژاد لوهمن از ترکیب دو نژاد لگهورن سفید و رود آیلندرد ایجاد شده است. نمونه‌های DNA توسط فن‌آوری Illumina HiSeq 2500 توالی‌یابی شد. توالی‌یابی کامل از دو انتها، طول قطعات  $101 \text{ bp}$  با میانگین عمق خوانش بیشتر از  $10 \times$  صورت گرفت.

کنترل کیفی، پالایش و هم‌ردیفی خوانش‌ها با ژنوم مرجع

در ابتدا، با استفاده از FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) سنجش کیفیت خوانش‌ها انجام شد. داده‌های خام حاصل از توالی‌یابی کل ژنوم با فرمت fastq و با حجم  $15 \text{ Gb}$  برای هر فرد بود. توالی ژنوم مرجع GRCg6a از سایت NCBI دانلود شده (شماره دسترسی در NCBI: GCF\_000002315.5\_GRCg6a) و هم‌ردیفی خوانش‌های با کیفیت با ژنوم مرجع صورت گرفت. برای شناسایی واریانت‌های جدید از تابع Add or Replace Read Group بسته‌ی نرم‌افزاری Picard-Tools استفاده شده و خوانش‌های تکراری توسط تابع Mark Duplicates شناسایی و حذف شدند، نهایتاً، فایل خروجی با نرم‌افزار SAMtools ایندکس شد تا داده‌ها برای هم‌ردیفی مجدد و تصحیح خطاهای احتمالی نرم‌افزار آماده شوند. سپس، هم‌ردیفی مجدد اطراف ایندل‌ها (Indel) صورت گرفت. با استفاده از بسته‌ی نرم‌افزاری GATK و تابع Realigner Target Creator ابتدا، محل نواحی متغیر روی ژنوم مشخص شده سپس با استفاده از تابع Indel Realigner تمامی خوانش‌ها در این نواحی به ژنوم رفرنس هم‌ردیف شدند.

بررسی قرار گرفتند که از این میان، ۶۹۵۲۲ پنجره با  $P < 0.001$  شناسایی شد. برای مناطقی از ژنوم که در دوگروه اختلاف معنی دار داشتند ( $FDR=0.001$ ) منهن پلات رسم شد.

مرغانی که با اهداف اصلاحی متفاوت در طول سالیان متمادی انتخاب شده‌اند را مقایسه و نواحی درگیر در انتخاب که حاوی تنوع تعداد کپی هستند را شناسایی کرد. در این مطالعه، در مجموع ۹۳۵۲۴۷ پنجره مورد

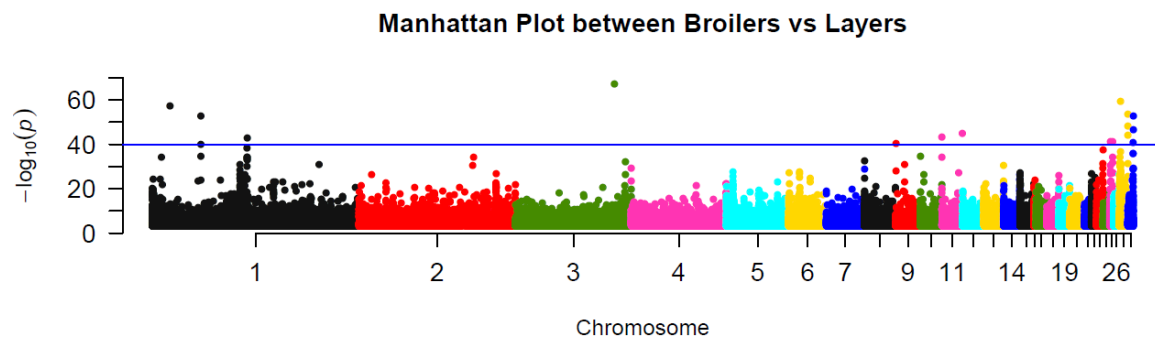


Figure 1. Manhattan plot between Broiler vs Layers ( $FDR = 0.001$ ). The blue line represents  $-\log_{10}(1e-40)$ .

نواحی ژنومی را که بین مرغ‌های تخم‌گذار و جوجه‌های گوشتی تفاوت معنی‌داری دارند، به ترتیب، ژن‌ها و نوع ژن‌های شناسایی شده را فهرست می‌کند.

نقاطی که بالای خط فرضی قرار داشتند شناسایی و در یک دامنه ۲۵± Kbp بررسی شدند تا ژن‌های احتمالی در آن ناحیه شناسایی شوند. جدول ذیل

**Table 1: Genomic regions that differed significantly between Broiler vs Layers. chr = chromosome number, end pos & start pos = start and end position of the interval studied, p-value = significant level of difference observed in the two groups, Gene = gene name in the Ensemble database, Gene Type = type of genes identified**

Chr	start pos	end pos	P_Value	Gene	Gene Type
1	8615849	8616849	8.88E-35	<u>ENSGALG00000047509</u>	LncRNA
1	16743849	16744849	6.39E-58	<u>ENSGALG00000051900</u>	LncRNA
2	108739997	108740997	4.06E-31	<u>ENSGALG00000035863</u>	Protein coding
4	6000	7000	5.23E-30	<u>ENSGALG00000051955</u>	Protein coding
4				<u>ENSGALG00000046911</u>	Protein coding
4				<u>ENSGALG00000046912</u>	-
8	165493	166493	4.98E-33	<u>ENSGALG00000047113</u>	LncRNA
9	1059400	1060400	5.37E-41	<u>ENSGALG00000053235</u>	RRNA
9				<u>ENSGALG00000049906</u>	RRNA
9				<u>ENSGALG00000048300</u>	RRNA
9				<u>ENSGALG00000048577</u>	RRNA
9				<u>ENSGALG00000042075</u>	RRNA
9				<u>ENSGALG00000051124</u>	RRNA
9				<u>ENSGALG00000053692</u>	RRNA
9				<u>ENSGALG00000050473</u>	RRNA
9				<u>ENSGALG00000050500</u>	RRNA
9				<u>ENSGALG00000047553</u>	RRNA
9	9122510	9123510	1.89E-31	<u>ENSGALG00000055140</u>	LncRNA
9				<u>ENSGALG00000051658</u>	LncRNA

9				<u>ENSGALG00000035425</u>	LncRNA
10	622360	623360	1.93E-35	<u>ENSGALG00000047538</u>	LncRNA
11	19984944	19985944	1.52E-45	<u>ENSGALG00000049187</u>	LncRNA
13	19039222	19040222	4.13E-31	<u>ENSGALG00000049252</u>	LncRNA
23	5916500	5917500	5.61E-32	<u>ENSGALG00000004268</u>	Protein coding
				NIPA like domain containing 3-	
				<u>ENSGALG00000004282</u>	Protein coding
				RCAN family member3	
23	5991000	5992000	8.80E-30	<u>ENSGALG00000047337</u>	LncRNA
23	6014000	6015000	4.43E-38		
25	490210	491210	2.89E-31	death effector domain containing- <u>ENSGALG00000043129</u>	Protein coding
25	532210	533210	1.04E-41	prefoldin subunit 2- <u>ENSGALG00000026374</u>	Protein coding
				ATPase H <sup>+</sup> transporting V1 subunit A-like - <u>ENSGALG00000000048</u>	Protein coding
25	3063402	3064402	7.99E-42	<u>ENSGALG00000047551</u>	Protein coding
25	3064402	3065402	8.77E-35	<u>ENSGALG00000054815</u>	LncRNA
25	3068502	3069502	3.21E-32	<u>ENSGALG00000046959</u>	LncRNA
				<u>ENSGALG00000051668</u>	LncRNA
27	8067131	8068131	5.61E-32	<u>ENSGALG00000054749</u>	LncRNA
27	8072131	8073131	2.95E-54	<u>ENSGALG00000048092</u>	LncRNA
27	8074131	8075131	8.16E-45		
27	8075131	8076131	9.75E-49		
28	5061100	5062100	1.33E-36	myeloid derived growth factor- <u>ENSGALG00000004263</u>	Protein coding
28	5063100	5064100	1.25E-36	<u>ENSGALG00000053954</u>	Protein coding
28	5071100	5072100	2.90E-41	TNF alpha induced protein 8 like 1- <u>ENSGALG00000004266</u>	Protein coding
28	5093100	5094100	4.12E-47	<u>ENSGALG00000038201</u>	LncRNA
28	5094100	5095100	1.78E-53	-	-
28	5100200	5101200	7.71E-30		

ها در مرغان اهلی شده جمله ژن‌های RCAN و NIPA را تحت تأثیر قرار داد. RCAN Family Member 3 یک ژن کُدکننده پروتئین است و در جد مشترک همه حیوانات وجود دارد (مولرو و همکاران ۲۰۰۷). سندرم داون از جمله بیماری‌های مرتبط با این ژن است. از جمله مسیرهای مرتبط با آن GPCR<sup>1</sup> Pathway و NFAT<sup>2</sup> و هایپرتروفی قلبی است (مولرو و همکاران ۲۰۰۷).

۱۳ ناحیه از ۲۹ منطقه بررسی شده فاقد هر نوع ژن بوده و تعداد ۱۶ ناحیه دیگر، حاوی ۲۸ ژن بود. از این میان تعداد ۱۶ ژن مربوط به RNAهای بلند غیر کُدکننده بود تعداد ۱۰ ژن مربوط به RNA ریبوزومی و تعداد ۱۲ ژن هم ژن‌های کُدکننده پروتئین بودند. هرچند انجام آنالیز هستی‌شناسی بر روی ژن‌های مناطق انتخاب شده، هیچ عبارت معنی داری را نشان نداد اما نتایج حاشیه نویسی در ناحیه ۵۹۱۷۵۰۰-۵۹۱۶۵۰۰ کروموزوم ۲۳ نشان داد که تنوع تعداد کپی

<sup>1</sup> G-protein coupled receptors

<sup>2</sup> Nuclear factor of activated T-cells (NFAT)

این ایزوفرم‌ها عمدتاً در سیتوپلاسم یافت می‌شوند، اما در هسته‌ها و میتوکندری‌ها نیز قرار دارند. یافته‌ها نشان می‌دهد که عملکرد RCAN شامل اتصال به RNA، تنظیم میتوکندری، تعدیل ساعت زیستی، ترموژن (تولید گرما در بدن) و سوخت و ساز است و ظرفیت مطلوبی برای RCAN به عنوان یک هدف درمانی ارائه می‌دهد (لی و آن ۲۰۲۰). NIPA like domain containing انتقال یون‌های منیزیم (Mg) را از یک طرف غشاء به طرف دیگر امکان پذیر می‌کند (گرمیل و همکاران ۲۰۰۹).

ژن NPAL3 پروتئین‌های بسیار محافظت شده را رمزگذاری می‌کند. اختلال در این ژن، منجر به اختلال در عملکرد سیستم عصبی شده و تعداد سلول‌های NK کاهش می‌یابد و سبب تغییرات IgM، IgG2 و IgA می‌شود، که نشان می‌دهد سیستم ایمنی نیز تحت تأثیر قرار می‌گیرد.

سلول‌های NK<sup>۲</sup> لنفوسیت‌هایی از دستگاه ایمنی بدن هستند که در ایمنی سلولی و ایمنی ذاتی<sup>۳</sup> دخیل هستند. سلول‌های کشنده ی طبیعی (NK) اولین خط دفاعی در از بین بردن سلول‌های سرطانی و سلول‌های آلوده به ویروس می‌باشند. کاهش فعالیت این سلول‌ها می‌تواند با بروز، استقرار و گسترش بسیاری از سرطان‌ها، عفونت‌های ویروسی، سندرم‌های نقص ایمنی و بیماری‌های خودایمنی همراه باشد (کومریچ و همکاران ۲۰۲۰).

ژن‌های DEDs و Atp6v1a و PFDN2 در کروموزوم ۲۵ حاوی مناطقی با تنوع در تعداد کپی بودند. ژن‌های (DEDs) در مرگ سلولی برنامه ریزی شده (آپوپتوز) نقش دارند. پروتئین‌های حاوی DED با تنظیم آپوپتوز و تکثیر گیرنده‌های فاکتور نکروز تومور (TNFalpha) مرتبط هستند. این پروتئین‌ها علاوه بر عملکردهای طرفدار و ضد آپوپتوز، رونویسی، مهاجرت و تکثیر را

گیرنده‌های جفت شده با پروتئین G (GPCRs) بزرگترین و متنوع‌ترین گروه گیرنده‌های غشایی در یوکاریوت‌ها هستند. پروتئین‌های G پروتئین‌های تخصصی با قابلیت اتصال نوکلئوتیدهای گوانوزین تری فسفات (GTP) و گوانوزین دی فسفات (GDP) هستند. پروتئین‌های G که با GPCRs ارتباط دارند هتروتیمر هستند، به این معنی که، آن‌ها دارای سه زیر واحد متفاوت هستند: زیر واحد آلفا، زیر واحد بتا ( $\beta$ ) و زیر واحد گاما ( $\gamma$ ). اتصال لیگاند به GPCR سبب تغییر در ترکیب گیرنده می‌شود که به نوبه خود پروتئین G را متصل کرده و فعال می‌کند. سپس، شکل فعال پروتئین G از سطح گیرنده آزاد می‌شود و به زیر واحدهای  $\alpha$ - و  $\beta/\gamma$  تجزیه می‌شود. سپس هر دو زیر واحد عوامل مؤثر خود را فعال کرده و موجب انتشار پیام رسان‌های دوم می‌شوند. این پیام رسان‌ها توسط پروتئین کینازها شناخته شده و منجر به فعال شدن آن‌ها شده و آبشار سیگنال را به سمت یک رویداد سلولی تحریک می‌کند (گورویچ ۲۰۱۹). عامل هسته‌ای سلول‌های T فعال شده (NFAT) خانواده‌ای از عوامل رونویسی که در پاسخ ایمنی مهم هستند. یک یا چند عضو از این خانواده در بیشتر سلول‌های سیستم ایمنی بیان می‌شود. افزون بر آن، NFAT در توسعه سیستم قلبی، ماهیچه‌های اسکلتی و سیستم عصبی نقش دارد (ویت و بردک ۲۰۰۵).

RCAN یک درون ریز مهم است و برای فرآیندهای سلولی متعدد حیاتی است. به دلیل پیشرفت‌های گسترده در زیست فن آوری و بیوانفورماتیک عملکردهای بی سابقه و غیر منتظره‌ای برای پروتئین‌های خانواده RCAN کشف شد (لی و آن ۲۰۲۰). این ژن در لکوسیت‌های قلب، ماهیچه‌های اسکلتی، کبد، کلیه و خون محیطی بیان می‌شود. سه رونوشت RCAN3 با آلترناتیو اسپلایسینگ تولید می‌شود که سبب ایجاد سه ایزوفرم پروتئین RCAN3 می‌شود (مولرو و همکاران ۲۰۰۷).

<sup>3</sup> Natural killer cell : NK cells

<sup>4</sup> Innate immune system



گیرنده و قاچاق درون سلولی، پردازش و تخریب پروتئین، انتقال مولکول‌ها و یون‌های کوچک و ورود عوامل بیماری‌زا به سلول‌ها از جمله ویروس‌هایی مانند ویروس آنفلوانزا و سموم مانند سم سیاه زخم. غشای پلاسما V-ATPases نقش مهمی در ترشح اسید در کلیه، تحلیل استخوان توسط استئوکلاست‌ها، اسیدی شدن دستگاه تناسلی مردان و تهاجمی بودن سلول‌های تومور و سلول‌های اندوتلیال دارد (کی و همکاران ۲۰۰۷).

تا شدن مناسب پروتئین‌ها یک فرآیند کلیدی برای به دست آوردن ساختار و ترکیب پروتئین است. Prefoldin یک پروتئین چاپرون معروف است که تا شدن صحیح پروتئین‌ها، به ویژه تا شدن صحیح توبولین و اکتین را تنظیم می‌کند توسط ژن PFDN2 کُد می‌شود. این پروتئین در پاتوژنز بیماری‌های شایع عصبی نظیر بیماری آلزایمر، بیماری پارکینسون و بیماری هانتینگتون نقش مهمی دارد. همچنین، هر یک از زیر واحدهای prefoldin عملکردهای متفاوت و مستقل از مجموعه prefoldin دارند. علاوه بر این، بیان غیرطبیعی زیر واحدهای prefoldin در تومورهای مختلف رخ می‌دهد و نقش مهمی در بروز و توسعه‌ی تومورها ایفا می‌کند (لیانگ و همکاران ۲۰۲۰).

در سلول تنظیم می‌کنند، علاوه بر این، در رشد جنین و هموستاز سیستم ایمنی ضروری هستند. آپوپتوز یا مرگ سلولی برنامه ریزی شده، زمانی که سلول‌ها به دلایل مختلفی دچار نقص می‌شوند بسیار حیاتی می‌باشد چرا که عدم حذف سلول‌های آسیب دیده می‌تواند سلول‌های سالم را درگیر کند (والمیکی و راموس ۲۰۰۹). لازم به ذکر است ژن TNFAIP8 که در مهار آپوپتوز و اتوفاژی در انواع سلول‌ها نقش دارد نیز در مناطقی ژنومی که بین دو نژاد تفاوت داشتند شناسایی شد.

V-ATPases مسؤل اسیدی شدن بخش‌های درون سلولی و انتقال پروتون در غشای پلاسمایی است. آن‌ها نقش مهمی در هر دو فرایند طبیعی مانند ترافیک غشایی، تخریب پروتئین، اسیدی شدن ادرار و تحلیل استخوان و همچنین فرآیندهای مختلف بیماری مانند عفونت ویروسی، کشتن سموم، پوکی استخوان و متاستاز تومور دارند (کی و همکاران ۲۰۰۷).

واکوئولار (H<sup>+</sup>)-ATPases یا (V-ATPases) پمپ‌های پروتون وابسته به ATP هستند که هم در محفظه داخل سلولی و هم در غشای پلاسمایی وجود دارند و در انواع فرآیندهای عادی و بیماری عمل می‌کنند. V-ATPases درون سلولی برای اندوسیتوز با واسطه‌ی

## References

- Alkan C, Coe BP and Eichler EE, 2011. Genome structural variation discovery and genotyping. *Nature reviews Genetics* 12: 363-376.
- Batista PG and Chang HY, 2013. Long Noncoding RNAs: Cellular Address codes in development and disease. *Cell* 152:1298-1307.
- Benjamini Y and Hochberg Y, 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1): 289-300.
- Bhanuprakash V, Chhotaray S, Pruthviraj DR, Rawat C, Karthikeyan A and Panigrahi M, 2018. Copy number variation in livestock. A mini review, *Veterinary World* 11(4): 535-541.
- Burt DW. 2005. Chicken genome: current status and future opportunities. *Genome Research* 15(12): 1692-1698.
- Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J, Younis S, Afonso S, Aken B, Alves JM, Barrell D, Bolet G, Boucher S, Burbano HA, Campos R, Chang JL, Duranthon V, Fontanesi L, Garreau H, Heiman D, Johnson J, Mage RG, Peng Z, Queney G, Rogel-Gaillard C, Ruffier M, Searle S, Villafuerte R, Xiong A, Young S, Forsberg-Nilsson K,

- Good JM, Lander ES, Ferrand N, Lindblad-Toh K, Andersson L. 2014. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 345(6200):1074-1079.
- Clop A, Vidal O and Amills M, 2012. Copy number variation in the genomes of domestic animals. *Animal Genetics* 43(5): 503-517.
- Conrad D F, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, 2010. Origins and functional impact of copy number variation in the human genome. *Nature* 464 (7289): 704–712.
- Elferink MG, Vallee AA, Jungerius AP, Crooijmans RP and Groenen MA, 2008. Partial duplication of the PRLR and SPEF2 genes at the late feathering locus in chicken. *BMC Genomics* 9: 391.
- Giuffra E, Evans G, Törnsten A, Wales R, Day A, Looft H, Plastow G and Andersson L, 1999. The Belt mutation in pigs is an allele at the Dominant white (I/KIT) locus. *Mammalian Genome* 10:1132–1136.
- Gorla E, Cozzi MC, Roman-Ponc SI, Ruiz Lopez FJ, Vega-Murillo VE, Cerolini S, Bagnato A and Strillacci MG, 2017. Genomic variability in Mexican chicken population using copy number variants. *BMC Genetics* 18:61.
- Grzmil P, Konietzko J, Boehm D, Hoelter SM, Aguilar A, Javaheri A, Kalaydjiev S, Adler T, Bolle I, Adham I, Dixkens C, Wolf S, Fuchs H, Gailus-Durne V, Wurst W, Ollert M, Busch D, Schulz H, Hrabe de Angelis M and Burfeind P, 2009. Targeted disruption of the mouse Npal3 gene leads to deficits in behavior, increased IgE levels, and impaired lung function. *Cytogenet Genome Research* 125:186-200.
- Gurevich VV and Gurevich EV, 2019. GPCR Signaling Regulation: The Role of GRKs and Arrestins. *Frontiers in Pharmacology* 10:125.
- Keel BN, Keele JW, and Snelling WM, 2016. Genome-wide copy number variation in the bovine genome detected using low coverage sequence of popular beef breeds. *Animal Genetics* 48 (2): 141–150.
- Kumric M, Ticinovic Kurir T, Borovac JA and Bozic J, 2020. The role of Natural Killer (NK) cells in acute coronary syndrome: A Comprehensive Review. *Biomolecules* 10: 1514.
- Lee SK and Ahnn J, 2020. Regulator of calcineurin (RCAN): Beyond down syndrome critical region. *Molecules and Cells* 43(8): 671-685.
- Liang J, Xia L, Oyang L, Lin J, Tan Sh, Yi P, Han Y, Luo X, Wang H, Tang L, Pan Q, Tian Y, Rao S, Su M, Shi Y, Cao D, Zhou Y and Liao Q, 2020. The functions and mechanisms of prefoldin complex and prefoldin-subunits. *Cell Bioscience* 10:87.
- Locke ME, Milojevic M, Eitutus ST, Patel N, Wishart AE, Daley M, Hill KA, 2015. Genomic copy number variation in *Mus musculus*. *BMC Genomics* 16(1): 497.
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA, 2010. The genome analysis toolkit: A mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.
- Mulero MC, Aubareda A, Schlüter A and Perez-Riba M, 2007. RCAN3 , a novel calcineurin inhibitor that down-regulates NFAT-dependent cytokine gene expression. *Biochimica et Biophysica Acta* 1773: 330-341.
- Norris BJ and Whan VA, 2008. A gene duplication affecting expression of the ovine ASIP gene is responsible for white and black sheep. *Genome Research* 18: 1282-1293.
- Pirooznia M, Goes FS and Zandi PP, 2015. Whole-genome CNV analysis: advances in computational approaches. *Frontiers in Genetics* 6:138.
- Qi J, Wang Y and Forgac M, 2007. The vacuolar (H<sup>+</sup>)-ATPase: subunit arrangement and in vivo regulation. *Journal of Bioenergetics and Biomembranes* 39(5-6): 423-426.
- Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME 2006. Global variation in copy number in the human genome. *Nature* 444(7118):444–454.



- Valmiki MG and Ramos JW, 2009. Death Effector Domain-Containing Proteins. *Cellular and Molecular Life Sciences* 66: 814-830.
- Witte H and Bradke F, 2005. Chapter 21 - Guidance of Axons to Targets in Development and in Disease. *Peripheral Neuropathy (Fourth Edition)* 1: 447-481.
- Wright D, Boije H, Meadows JR, Bed'Hom B, Gourichon D, Vieaud A, Tixier-Boichard M, Rubin CJ, Imsland F, Hallbook F and Andersson L, 2009. Copy number variation in intron 1 of SOX5 causes the pea-comb phenotype in chickens. *PLOS Genetics* 5:e1000512.
- Zhang L, Liu R, Luan Y and Yao Y, 2018. Tumor Necrosis Factor- $\alpha$  Induced Protein 8: Pathophysiology, Clinical Significance, and Regulatory Mechanism. *International Journal of Biological Sciences* 14(4):398-405.
- Zhang S, Wang Y, Jia L, Wen X, Du Z, Wang C, Hao Y, Yu D, Zhou L, Chen N, Chen J, Chen H, Zhang H, Celik I, Gulsoy G, Luo J, Qin B, Cui X, Liu Z, Zhang S, Esteban MA, Ay F, Xu W, Chen R, Li W, Hoffman AR, Hu JF and Cui J, 2019. Profiling the long noncoding RNA interaction network in the regulatory elements of target genes by chromatin in situ reverse transcription sequencing. *Genome Research* 29(9):gr.244996.118.

## Bioinformatics identification of genomic diversity of Copy Number Variations in laying and broiler hens

H Soleimani<sup>1</sup>, J Shoja<sup>2</sup>, S A Raafat<sup>2</sup> and S Qanbari<sup>3</sup>

Received: November 15, 2021

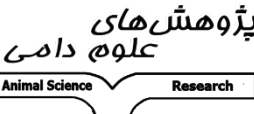

Accepted: March 6, 2022

<sup>1</sup>Phd student, Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

<sup>2</sup>Professor, Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

<sup>3</sup>Institute of Genetics and Biometrics, Institute of Leibniz for Farm Animal Biology (FBN), Dammerstorf, Germany

\*Corresponding author: Email: soleimani\_homa@yahoo.com

 <p>پژوهش‌های علوم دامی</p>	<p>Journal of Animal Science/vol.34 No.1/ 2024/pp 93-103  <a href="https://animalscience.tabrizu.ac.ir">https://animalscience.tabrizu.ac.ir</a></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 (<a href="https://creativecommons.org/licenses/by-nc/2.0/">https://creativecommons.org/licenses/by-nc/2.0/</a>)          DOI: 10.22034/as.2022.48893.1637</p>		

**Introduction:** Copy number variation (CNV) consist of deletion, insertion, and duplications. It is an important source of genetic variation in organisms and thus influences on the gene expression and phenotypic variation. Copy number variation (CNV) is one of the structural variant with an intermediate size class larger than 50bp which involves unbalanced rearrangements that increase or decrease the amount of DNA (Pirooznia et al 2015, Alkan et al 2011). The size of CNVs is larger than 50bp, while smaller segments are known as insertions or deletions (indels). Thereupon these structural variations comprise more polymorphic than SNPs because of enormity, detection of them and their effect on phenotype has caught the attention of many researchers recently. It has been reported that CNVs changes in gene dosage and regulation as well as in transcript structure, and thus contribute to phenotypic variability (Pirooznia et al 2015, Alkan et al 2011).

The pea-comb phenotype is caused by a CNV mapping to intron 1 of the SRY (sex determining region Y)-box 5 (SOX5) gene (Wright et al 2009). Late feathering in chickens is due to incomplete duplication in PRLR and SPEF2 genes (Elfrink et al 2008). In swine, dominant white colour has been related with a duplication of a 450-kb fragment of the KIT gene (Giuffra et al 1999) and a splice mutation causing the skipping of exon 17 (Giuffra et al 1999). In sheep, doubling in the ASIP gene results in the regulation of pigment in body coat (Norris et al 2008). Doubling the 4.6 k base pair into the six introns of the STX17 gene results in a gray body color in the horse with age. Deletion of the intergenic region with a length of 11.7 kbp in the goat genome leads to the removal of horns (Clou et al 2012).

Chicken is the most intensively farmed animal on earth and is a major food source with billions of birds used in meat and egg production each year. A big share of chicken CNVs involves protein coding or regulatory sequences. A comprehensive study of chicken CNV can provide valuable information on genetic diversity and assist future analyses of associations between CNV and economically important traits in chickens. Unique chicken genome with macro and micro chromosomes and its biology make it an ideal organism for studies in development and evolution, as well as applications in agriculture and medicine (Burt 2005). In the last several years, there has been an increasing interest in the study of CNVs in the chicken. This study focuses on comparison of CNV between the broilers and layers chicken to find evidence of domestication on the genome using whole genome sequencing.

**Material and methods:** we used n=90 female birds of two commercial broiler (n=40) and layer (n=50) chicken. The broilers (BRs) were represented by 20 DNA samples of each of two lines (BRA and BRB) established independently and previously collected as part of the AVIANDIV project. In the layer group (LRs), data from 25 birds each from purebred white (WL) and brown (BL) egg laying populations, sequenced in the frame of the SYNBREED project (<http://www.synbreed.tum.de/index.php?id=2>), were included. The paired-end reads with a read length of 101bp were mapped against the current reference genome assembly GalGal6 using the Burrows-Wheeler aligner (bwa, 0.6.2-r126 Version, with default parameters. Duplicate reads were masked during post-processing using the Picard tool set (version 2.9.2, <http://picard.sourceforge.net>). Finally, Genome Analysis Toolkit-3.3.0 was used to realign reads for correcting errors caused by indels. Using GATK software package and Depth of Coverage function (McKenna et al 2010), the depth of readings was calculated for each sample. Then filter out reads with mapping quality below 20. Because comparing the genomes of individuals in different groups was time consuming and computationally difficult for all parts of the genome, the genomes of each individual were divided into 1000 bp non-overlapping windows and the average reading depth per window was calculated. Then the results were normalized against the BL sample that showed highest average depth. In short, we created a correction factor per population and applied it on the depth of coverage value for each window. For all the contrasts, we performed an analysis of variance (ANOVA) as described (Carneiro et al 2014). For the Broilers-Layers contrast we scanned 935247 windows. ۶۹۵۲۲ windows showed significant by FDR with  $P < 0.001$ , with ANOVA using the Benjamini-Hochberg FDR method for multiple corrections (Benjamini and Hochberg 1995).

**Results and discussion:** Mapping sequencing data to galGal6 assembly showed an average 98.61% mapping rate and 11.51 depth. Manhattan plot was plotted for regions of the genome that differed significantly between the two groups (FDR = 0.001). The points above the hypothetical line were identified and examined in a 25 Kbp confidence interval to identify possible genes. 39 regions were identified that half of them does not contain any genes. Although Long noncoding RNAs are under lower selective pressure than protein-coding genes (Batista and Chang 2013), the other 11 regions contained 16 genes related to long non-coding RNAs. Long noncoding RNAs (lncRNAs) play a critical role in organizing the 3-dimensional genome architecture and regulating gene activity in cis or in trans through multiple mechanisms (Zhang et al. 2019, Batista and Chang 2013). 6 Other regions also contained 12 coding genes. Most of the identified genes were somehow linked to the immune system disease or cancer. Genes such as DEDs and TNFAIP8 are involved in programmed cell death (apoptosis) and two genes NPAL3 and RCAN, which are involved in the immune system, had a copy number variation in the studied samples. In addition, RCAN is involved in Down syndrome. The PFDN gene, located on chromosome 25, is also involved in Alzheimer's and Parkinson's disease.

**Key words:** broilers, copy number variation, layers, structural variant, whole genome analysis