

从孟德尔遗传到整合组学: 鸡重要性状遗传机制解析及育种应用

王宇哲^{1,2}, 吴寒宇^{1,2}, 胡晓湘^{1,2*}

1. 中国农业大学生物学院, 畜禽生物育种全国重点实验室, 北京 100193

2. 中国农业大学国家模式动物科学中心, 北京 100193

* 联系人, E-mail: huxx@cau.edu.cn

2024-12-31 收稿, 2025-02-22 修回, 2025-04-09 接受, 2025-04-10 网络版发表

广东省重点领域研发计划(2022B0202100002)、农业生物育种国家科技重大专项(2023ZD0406805)和国家自然科学基金(32272862)资助

摘要 鸡是世界上驯化最成功的陆生脊椎动物之一。它不仅是全球主要肉类来源, 也是发育生物学、免疫学、生理学等研究重要的模式动物。随着2004年基因组组装的完成, 鸡各类遗传机制的解析得到了迅猛发展。本文首先概述了鸡孟德尔质量性状研究进展, 然后总结了鸡基因组组装的研究进程, 随后介绍鸡复杂性状遗传机制解析的方法和成果。接着讨论了多组学和大模型时代基因组选择和分子设计育种的发展方向, 最后本文总结了当前鸡遗传研究的现状, 并展望了未来与人工智能交叉的前景。本文为鸡各类遗传解析工作提供了系统的梳理, 为本领域的进一步发展方向提供参考。

关键词 鸡, 基因组组装, 基因定位, 多组学整合, 基因组选择, 基因编辑

从一种森林栖息的鸟类, 到世界上数量最多的陆生脊椎动物, 没有任何驯化动物像鸡一样, 被人类如此广泛地改造。鸡不仅是全球肉类和蛋白质的重要来源, 也是脊椎动物发育生物学、免疫学、生理学等研究重要的模式动物。中国是鸡资源大国, 根据最新公布的《国家畜禽遗传资源品种名录(2021年版)》, 我国拥有240个鸡品种, 它们在外貌、生长、繁殖等性状上各具特色, 是科学研究和品种改良的宝贵资源库。随着整合组学、基因组选择和基因编辑等技术的发展, 鸡各类性状遗传机制的解析也逐渐深入。本文综述了鸡遗传研究相关进展, 将从孟德尔质量性状研究、鸡基因组学研究、复杂性状多组学解析、基因组选择和分子设计育种研究进展等方面依次阐述。

1 经典孟德尔遗传定律在鸡中的应用已挖掘众多质量性状致因基因

鸡是最早证明孟德尔遗传定律的动物之一。威廉

贝特森在1897年对家鸡冠形等性状进行杂交实验, 不仅发现了与孟德尔类似的3:1分离率, 还强调了对杂种后代进行统计分析的重要性, 这是遗传学领域的一个里程碑, 标志着遗传学研究的一个重要起点。一个世纪以来, 科学家在鸡上开展了大量的研究工作, 相关质量性状的分子机制挖掘取得了巨大进步^[1]。质量性状多数由一对或少数几对主效基因决定, 遗传关系简单, 一般服从孟德尔遗传定律。目前鸡中已经精确定位到致病基因的质量性状有59个, 包括白羽、多趾、裸颈、豆冠、无尾椎骨、慢羽、毛脚等^{[2](表1)}。

鸡的冠型作为鸡最多元化的表型之一, 在对孟德尔遗传学的发现和遗传学基本概念的建立中发挥了重要作用。在鸡冠的科学问题被提出一个世纪以后, 科学家最终解析了鸡冠形成的复杂机制。玫瑰冠^[15]、豆冠^[14]和复冠^[16]的分子遗传学研究结果显示, 这三种冠型的形成均与相关基因非编码区的结构变异有关: 玫

引用格式: 王宇哲, 吴寒宇, 胡晓湘. 从孟德尔遗传到整合组学: 鸡重要性状遗传机制解析及育种应用. 科学通报, 2025, 70: 4360–4369

Wang Y, Wu H, Hu X. From Mendelian inheritance to multi-omics integration: deciphering the genetic mechanism of important traits and breeding applications in chickens (in Chinese). Chin Sci Bull, 2025, 70: 4360–4369, doi: [10.1360/TB-2024-1399](https://doi.org/10.1360/TB-2024-1399)

表 1 部分鸡外貌性状基因定位结果

Table 1 Gene mapping results of chicken feature trait

表型	基因	突变类型	染色体	参考文献
巧克力羽	<i>TYRP1</i>	错义突变	Z	[3]
淡紫色羽	<i>MLPH</i>	错义突变	7	[4]
卷羽	<i>KRT75L4</i>	小片段删除(≤20)	33	[5]
无尾	<i>RUM</i>	大片段删除(>20)	2	[6]
胡须	<i>HOXB</i>	复杂重排	27	[7]
侏儒	<i>TMEM263</i>	无义突变	1	[8]
显性白羽	<i>PME</i>	小片段插入(≤20)	33	[9]
烟灰色羽	<i>PMEL</i>	小片段删除(≤20)	33	[9]
绿壳蛋	<i>SLCO1B3</i>	大片段插入(>20)	1	[10]
哥伦比亚羽	<i>MC1R</i>	错义突变	11	[11]
多指	<i>SHH</i>	调控突变	2	[12]
黑棕色/黄色羽	<i>SOX10</i>	大片段删除(>20)	1	[13]
豆冠	<i>SOX5</i>	重复突变	1	[14]
玫瑰冠	<i>MNR2</i>	复杂重排	7	[15]
复冠	<i>EOMES</i>	重复	2	[16]
凤头	<i>HOXC10</i>	重复	33	[17]
凤头	<i>HOXC8</i>	调控突变	33	[18]
毛脚	<i>TBX5</i>	调控突变	15	[19]
无翼	<i>RAFI</i>	无义突变	12	[20]
丝羽	<i>PDSS2</i>	调控突变	3	[21]
性连锁羽色淡化	<i>CDKN2</i>	复杂重排	Z	[22]

瑰冠的形成与7号染色体上一个7.4 Mb片段的倒位有关, *SOX5*基因内含子上的拷贝数变异与豆冠的形成有关, 而复冠表型的形成则与*EOMES*基因上游一个20 kb的序列重复有关. 鸡的冠型与鸡的繁殖性能还存在关联, 玫瑰冠纯合子个体的倒位突变导致*MNR2*的异位表达, 引起了公鸡生育力的低下^[23]. 鸡冠型的多样性为遗传学家揭示动物遗传规律提供了丰富的研究材料.

鸡的羽毛颜色呈现出惊人的色彩和图案的多样性, 也是品种重要的外貌特征. 羽色的差异主要是由羽毛中色素的含量、比例以及在羽毛中分布的不同引起的, 目前有多个与羽色相关的基因被发现. 位于鸡11号染色体的黑色素皮质素受体1(*MC1R*)控制着鸡的黑羽和麻羽性状^[11], *MC1R*通过调节酪氨酸酶的活性来影响真黑色素和褐黑色素的分布. 家鸡的显性白羽和隐性白羽两种表型分别受33号染色体早期黑素体蛋白17(*PMEL17*)^[9]和酪氨酸酶(*TYR*)控制^[24], 这些基因的突变抑制了黑色素的沉积. 此外, 鸡的银羽和金羽、深棕色羽、横斑羽、淡紫色羽、斑点羽等羽色性状的致因基

因也被报道, 它们通过决定真黑素和褐黑色素分布、抑制色素沉着、降低色素沉着强度等方式塑造着各式各样的羽色及图案.

除此之外, 还有一些我国地方品种独特外貌表型的功能基因被鉴定. 以中国农业大学胡晓湘教授团队工作为例: 惠阳胡须鸡的胡须性状是由27号染色体上3个拷贝数变异组成的复杂结构变异引起, 并导致了*HOXB8*基因的异位表达^[7]. 丝羽乌骨鸡的丝羽表型是由*PDSS2*上游一个顺式调控单核苷酸多态位点(Single nucleotide polymorphism, SNP)的突变引起^[21]. 瓢鸡的无尾性状是由一个新基因*Rum*的表达缺失引起, 进一步导致尾部骨骼发育出现异常^[6]. 此外, 鸡的缨头^[18]、毛脚^[19]、绿蛋壳^[10]、匍匐性状^[25]等特色性状也已完成基因定位. 这些鸡特有性状的研究为理解鸡的遗传发育提供了重要信息, 强调了遗传演化和适应性的动态本质, 是鸟类遗传和进化研究领域突破性的发现. 随着研究工作的不断持续, 机制未知的鸡质量性状在逐步减少, 对应的研究策略也正在向整合更多组学数据的逻辑上演变.

2 高质量基因组组装为鸡新一代遗传学研究奠定基础

鸡是最早进行基因组测序的农业动物. 2004年, 国际鸡基因组测序联盟发布了首个鸡基因组(红色原鸡, *Gallus_gallus*-1.0), 这也是第一个被测序的鸟类基因组. 第一版鸡基因组大小约为1.05 Gb, 初步注释了20000多个基因, 为后续研究提供了基因组框架^[26]. 近年来, 得益于长读长序列的加入、改进的*de novo*组装算法以及BAC克隆序列的整合, 基因组组装的准确性取得了巨大进步. 2004年随后的十余年间, 基因组版本在不断更迭, 其中2018年发布了GRCg6a版, 2021年发布了GRCg7版, 是目前被广泛使用的两个参考基因组版本. GRCg6a版作为目前红色原鸡最好的参考基因组, Scaffold N50值约为20 Mb. GRCg7版本基因组则主要体现在测序技术和组装方法的革新上. 其采用了二代测序、Hi-C、长读长测序等先进技术, 结合TrioCanu, purge_dups, Salsa2 HiC等多种组装工具, 显著提升了基因组拼接的精度和完整性. GRCg7在参考样本选择上采用了一只由白羽肉鸡母鸡和白来航蛋鸡公鸡杂交产生的雌性后代. 这一改变的理由是, 从红色原鸡到现代商业化肉、蛋鸡, 人工选择对基因组变异施加了巨大影响, 以红色原鸡作为参考基因组对商业鸡变异检测将可能导致假阳性结果. GRCg7中也包含了两个版本, GRCg7b用于代表肉鸡品种, 并被作为新的参考基因组, Scaffold N50值约为90 Mb; GRCg7w用于代表蛋鸡品种, 作为备选参考基因组.

当前, 可检索的鸡基因组组装体涵盖了原鸡以及全球各个地方品种, 例如胡须鸡、溧阳鸡、固始鸡、白来航肉鸡等在内约44个版本. 随着越来越多不同品种鸡的基因组序列得以组装, 研究人员发现独立且单一的参考基因组无法完整涵盖物种所有的遗传信息, 可能会阻碍物种基因组变异的精准鉴定, 这促使鸡基因组组装进入了泛基因组时代. 泛基因组可以纳入更多的序列多样性, 2021年, 使用二代测序数据构建了首个鸡的泛基因组, 发现了鸡参考基因组(GRCg6a)中未被组装的66.5Mb序列^[27]. 2022年, 中国农业大学和西北农林科技大学联合使用鸡三代基因组测序数据构建了鸟类中的第一个基于*de novo*组装的高质量泛基因组, 鉴定到相当于参考基因组15%(159 Mb)的泛序列. 该研究发现泛序列中的新基因大多位于染色体亚端粒区和小染色体, 具有较多的串联重复序列. 非经典的

DNA二级结构导致了这些缺失序列难以被捕获, 阻碍了鸡基因组的深入研究. 此外, 研究发现新基因的替代率比已知基因高3倍, 这打破了已有的研究结论, 对鸟类比较基因组学和功能基因组学的研究具有重要意义^[28]. 2023年, 密苏里大学(University of Missouri)利用30个不同品种鸡基因组首次构建了全基因组图谱参考序列, 以解决现有参考序列无法识别复杂结构变异的问题^[29]. 使用这种全基因组图谱参考序列可以更准确地检测现代品种中存在的结构变异, 同时能够减少参考偏差, 提高基因型检测的准确性.

在泛基因组纳入更多序列多样性的同时, 对基因组完整性的要求也在逐步提高. 2021年人类基因组端粒到端粒的完整基因组(telomere-to-telomere, T2T)版正式发布, 完美重建了包括端粒和着丝粒区域的人类基因组所有序列, 标志着基因组T2T时代的来临. 鸡基因组质量也在大幅提升, 2023年研究人员构建了我国特色地方品种惠阳胡须鸡的高质量参考基因组GGswu1^[30]. 该基因组新鉴定了6条鸡中最小的染色体(chr29, chr34~chr38)^[31], gap数量和长度显著减少, 仅剩下约5572 bp的空白序列(由W染色体上的14个gap组成). 目前, 更多以T2T为目标的基因组组装工作正在进行中, 我国农业农村部也在主导构建我国多个特色地方品种的高质量参考基因组. 以T2T打底的多品种图形化基因组构建, 则完美融合了鸡基因组对“完整性”和“特异性”的高要求, 是未来几年内基因组组装的研究方向. 2024年, 基于鸡GRCg7b和GGswu1两个参考基因组构建泛基因组, 注释到1242个和880个新基因, 为功能基因组学研究提供了更加完整的基因集合^[32]. 然而, 新型图形化基因组仍需解决基因组高质量基因注释、开发图形化基因组比对算法、适配下游群体遗传分析等技术难点.

3 多组学整合研究为数量性状遗传机制研究带来新的突破

与外貌性状不同, 鸡的重要经济性状多为数量性状, 其遗传受多个基因座共同影响, 一般通过连锁分析或全基因组关联分析(genome-wide association study, GWAS)等方法来定位关键基因. 目前, 动物数量性状基因座(quantitative trait locus, QTL)数据库(animal QTLdb)收录了18602个与鸡的复杂性状相关的QTL^[33], 覆盖生产、外貌、健康、繁殖和生理五大类性状, 这些QTL资源为鸡重要性状遗传机制的解析提供了重要

参考。然而,由于数量性状遗传结构本身的复杂性,定位结果通常会面临QTL区间较大、关联位点多分布于非编码区、注释基因难以与目标表型直接建立联系等困境,使得揭示GWAS相关候选位点的生物学机制极具挑战性^[34]。传统的孟德尔质量性状Case-Control的研究方法已不适用,需要通过对中心法则链条上基因组、表观组、转录组、表型组等多维度大规模数据的搜集整理,帮助我们建立各组学间的级联反馈,提高复杂性状基因定位的精细程度(图1)。

在2003年启动的DNA元件百科全书(Encyclopedia of DNA Elements, ENCODE)项目,通过建立全球范围的合作,完成了人类和小鼠基因组中调控元件的注释工作。在农业动物中,功能基因组注释计划(Functional Annotation of Animal Genomes, FAANG)于2015年启动,为研究者进一步解析复杂性状提供了重要的数据资源。此外,基因型-组织表达(genotype-tissue expression, GTEx)项目从转录水平上为研究者开辟了解析复杂性状的新思路,其主要思想是以分子表型作为桥梁,填补遗传变异和复杂性状之间的空白,进而精确解析性状背后的机制。从2015年人类GTEx试点项目的开展到2020年项目完成,GTEx项目已经搜集涵盖了838个捐赠者、49个人体组织、超过1.5万个转录组测序样本。基于这些数据,在超过2万个基因中定义到了基因表达QTL,为解析复杂性状遗传调控的分子机制提供了有力支撑。借鉴人类中成功的研究经验,家养动物基因型-组织表达(Farm animal Genotype-Tissue Expression,

FarmGTEx)项目于2018年正式启动,截至目前,牛^[35]、猪^[36]和鸡^[37]三个主要物种试点阶段的GTEx资源已经发表,为农业动物复杂性状的解析提供了重要资源。

在多组学分析方面,常用的整合方法包括全转录组关联分析(transcriptome whole association study, TWAS)、孟德尔随机化(Mendelian randomization, MR)和共定位(colocalization)分析等。TWAS的主要思想是直接将军表达与表型建立联系,进而可以将基因与表型进行关联^[38]。MR则是一种通常在流行病学中用于推断疾病病因的方法,其主要原理是通过使用基因组遗传变异为辅助变量,推断暴露因素是否与目标疾病存在因果关系。而在多组学整合中,SNP通常作为“辅助变量”,基因表达为“暴露因素”,复杂性状表型当作“结果变量”^[39]。共定位分析则是直接对分子QTL和终端复杂性状QTL的关联结果进行整合,以判断两者之间是否受到相同遗传因素的影响^[40]。

在家禽研究方面,2015年启动的FAANG联盟^[41]在表观注释层面,已注释了1573399个调控元件,包含102907个启动子、146045个TSS-近端转录区、765400个增强子、351928个ATAC岛和201377个抑制区^[42]。这一高分辨率的调控元件图谱也被称为基因组序列的密码本,为鸡遗传学研究和基因组学分析提供了宝贵注释资源,帮助预测功能性非编码序列,为表观基因组编辑提供潜在靶点,也为非编码区比较基因组学的发展提供了新的可能^[43]。在转录表达层面,ChickenGTEx (genotype-tissue expression)整合了来自28个组织的

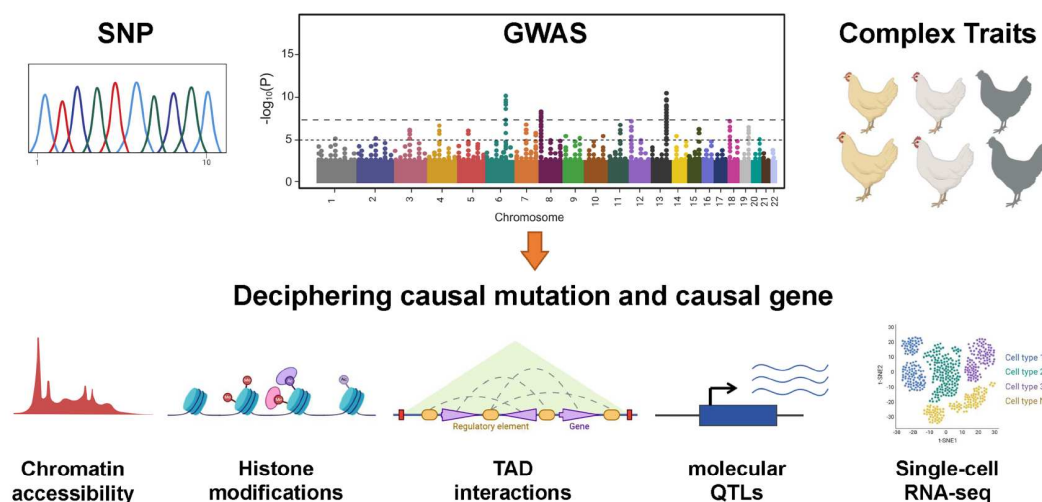


图1 多组学整合解析数量性状示意图(部分素材经BioRender授权使用)

Figure 1 Schematic diagram of multi-omics integrated analysis of complex quantitative traits (Some figure components were created using BioRender software with a license)

7015个RNA-Seq样本和2869个全基因组测序样本,构建了包括eQTL, exQTL, lncQTL, sQTL和3a'QTL在内的五种分子QTL图谱^[37]. 我国科学家构建了Galbase数据库^[44],集成了来自928个基因组、429个转录组、379个表观基因组、15275个QTL条目和7526个关联的鸡多组学数据,也标志着组学数据库的建设模式从单一的遗传定位结果收录,逐步走向多组学的交叉融合. 利用鸡多组学数据,通过孟德尔随机化(summary-data-based Mendelian randomization, SMR)、全转录组关联研究(TWAS)、共定位分析等方法鉴定到210个影响鸡生长性状的基因座,并确定了*KPNA3*基因上影响鸡多个生长性状的功能位点^[37]. 在生长性状遗传解析上,发现鸡1号染色体末端存在影响生长的主效QTL,可解释超过12%的表型变异^[45],但在惠阳胡须鸡-岭南黄鸡深度杂交系中,通过累积重组的方法却不能有效缩短该区间,且该区域还存在差异单倍型效应渐变化、正选择信号马赛克化等独特遗传结构^[46]. 研究人员认为,该区间的显著信号是由多个物理位置上相邻的致因基因联合累加导致,在调控层面则可解释为:多个基因在不同的组织中,通过不同的调控途径对生长性状协同产生影响. 在脂肪沉积研究上,高分辨率Hi-C图谱可以揭示启动子-增强子的远源互作,通过整合三维基因组学和多组学分析,研究发现rs734209466突变作为一个等位基因特异性增强子,通过结合转录因子*IRF4*,远程增强*IGFBP2*和*IGFBP5*的转录,这种调控作用影响前脂肪细胞的分化和增殖,最终影响表型^[47]. 在繁殖性状上,产蛋能力是家禽重要经济性性状,结合多组学数据和实验验证,发现肝脏因子*APOA4*和脂肪因子*ANGPTL2*可通过介导下丘脑-垂体-卵巢轴的组织间通讯,来有效提升产蛋量,揭示了多组织协同调控鸡产蛋性能的分子机制,同时为鸟类生殖发育的调控提供了重要理论依据^[48].

多组学研究目前还在向单细胞层面深入,单细胞RNA测序(scRNA-seq)能够以单细胞级别的分辨率解析基因表达模式^[49]. 利用单细胞转录组绘制鸡肢体发育过程中的细胞图谱,揭示了在不同发育阶段细胞群体的异质性和动态变化. 它不仅解析了主要组织类型(如间充质、骨骼前体、肌肉细胞、皮肤和血细胞),还发现了关键亚群(如顶端外胚层脊细胞),并识别出标志性基因*FGF8*,为理解四肢形态发生的细胞和分子机制提供了新见解^[50]. 通过单细胞转录组识别了在禽流感感染过程中受到病毒直接影响的细胞类型,并分析了

病毒RNA的分布,构建了免疫细胞间的配体-受体互作网络,揭示了不同免疫细胞在病毒感染后的通讯机制,为了解禽类病毒感染的病理生理机制以及相关基因调控网络提供了重要数据支持^[51]. 未来,利用单细胞技术构建细胞类型特异性的QTL图谱,将为复杂性状研究带来更高的精度. 随着计算生物学的发展,单细胞多组学整合工具(如Seurat和Signac)的不断优化,将进一步提升数据分析的能力与可靠性.

4 多组学和大模型时代的鸡基因组选择与分子设计育种

基因组选择(genomic selection, GS)是一种基于全基因组遗传标记预测动物或植物育种价值,从而优化育种决策的现代育种方法. 自21世纪初首次提出以来^[52],GS技术已在农业动植物育种^[53]及人类疾病风险预测^[54]等领域取得广泛应用. 作为重要的模式动物,早在2008年,美国农业部主导设计生产了鸡中首个包含3k标记的SNP芯片,随后推出的60k芯片及600k芯片,进一步推动了GS技术在家禽育种中的应用. 尽管基因组选择技术在动物育种中已取得一定进展,但目前其所能达到的选择准确性仍未能满足实际需求. 根本原因在于我们对复杂性状的调控机制了解有限,只能依赖较为简单、且缺乏生物学背景模型来描述遗传变异与表型之间的关系,而这些模型与真实遗传架构之间仍存在较大差距. 一些研究将已知的先验信息加入模型取得了显著的准确性提升,例如利用GFBLUP (genomic feature best linear unbiased prediction)模型提升已知关联位点的权重,在黄羽肉鸡群体中将预测准确性提升了13.8%~15.2%;另一项研究通过SLDP(selective linkage disequilibrium pruning)的方法富集性状重要QTL区域内的位点,在不改变模型计算复杂度的情况下增加区域权重,实现提升预测准确性的目的. 这些研究提示我们,将已知先验信息加入预测模型,使其具备更多生物学特征是现阶段提升预测准确性的主要思路.

此外,现有的基因组选择方法仅利用基因组这一单一维度的信息,忽略了表观组、转录组、蛋白组和代谢组等多个生物学层面之间错综复杂的作用机制. 已有研究提出了NetGP多组学预测模型,通过结合转录组和基因组数据,有效地提高了预测性能^[55]. 然而,由于中间组学的检测成本仍然较高,在大规模群体中开展多组学数据分析仍面临挑战. 但随着组学技术的不断

断进步,未来动物育种有望从单一的基因组选择转向多组学选择,进一步提升育种效率和准确性^[56]。另一方面,简单的线性关系已难以准确地度量遗传变异与复杂性状之间的联系。机器学习、深度学习等人工智能模型已经开始在育种工作中崭露头角^[57,58]。这些先进的算法能够处理大量高维数据,挖掘各个组学与表型之间潜在的非线性关系。最近DNA大语言模型理念引领了学科交叉的浪潮,基因组学、计算机科学与统计学的融合与交叉,极大拓宽了基因组选择的研究视野,为精准育种提供了新的思路和方法,也为复杂性状的遗传机制解析开辟了更加高效和可行的路径。

基因编辑技术是GS技术之外另一种现代化遗传改良方法,通过精确高效的遗传修饰,能够在基因功能研究、疾病模型创建以及分子设计育种中发挥关键作用。自20世纪80年代,DNA显微注射技术成功制备转基因小鼠以来,哺乳动物基因编辑技术已日趋成熟。然而,鸡的特殊发育模式限制了这些技术的应用。鸡胚发育分为体内子宫期和体外孵化期,新生种蛋胚盘中已有数万个细胞,其中包括少量原始生殖细胞前体。通过胚盘下腔注射外源基因载体,可靶向生殖细胞,获得能传递外源基因的性腺嵌合体后代。1987年,Salter等人^[59]首次利用禽白血病病毒载体制备出转基因鸡。2006年,研究人员首次从鸡胚血液中分离出鸡原始生殖细胞(primitive germ cells, PGCs)并实现体外培养建系,通过遗传改造这些细胞,可高效获得转基因后代^[60]。2013年,Schusser等人^[61]通过同源重组技术在鸡中首次实现基因敲除。随后,TALENs和CRISPR/Cas9等靶向编辑技术被应用于PGCs,成功实现特定基因的靶向编辑。在利用PGCs进行基因编辑个体制备时,为降低受体胚胎中内源PGCs的竞争性挑战,可通过物理(γ 射线辐照或移除HH 14-15期胚胎血液)、化学(白消安处理)等方法消除内源PGCs干扰,提升阳性PGCs的种系传递效率;此外,利用TALENs技术敲除DDX4,可制备雌性不育鸡^[62],通过CRISPR/Cas9介导的同源定向修复(HDR)将iCaspase9插入DAZL基因,还开发了新型可诱导不育宿主鸡^[63]。这些研究为基因编辑鸡的制备提供了有效手段。

从技术角度分析,基因编辑是比基因组选择更为高效的育种策略。基因编辑能够直接对目标基因进行定点修改,无需依赖自然突变的积累。因此基因编辑可以被视为育种工作的终极路径,而未来的育种家们将会化身为基因组的设计师,设计出更加符合人类需求

的动物基因组,从而灵活、高效、标准化地为人类提供所需的生物资源。随着基因编辑技术的不断发展,尤其是单碱基编辑等技术的成熟,基因编辑技术面临的主要挑战已经从“如何编辑”转变为“编辑什么”的问题,寻找重要功能基因或突变作为编辑靶点将是未来的重要工作。多组学时代带来的数据整合以及大数据、大模型带来的跨学科领域融合,为基因编辑提供了强大的支持和新的视角。通过多组学数据,研究人员能够全面了解畜禽的遗传特征和表现,从而识别出大量影响性状的候选突变。此外,深度学习模型的应用也同样为基因编辑技术提供了新思路。通过深度学习和机器学习等先进算法的引入,能够更有效地从庞大而复杂的生物学数据中提取有用信息,构建基因组的特征模型,通过序列扰动,从计算机层面预估编辑前后序列特征和功能的变化,进而为基因编辑寻找合适靶点,这种电子基因编辑模式能够提高功能位点的筛选效率。例如Kathleen等人^[64]利用人类基因组注释信息在超过1300个组织和细胞系中建立了Sei深度学习模型,通过学习不同调控元件内的序列特征建立突变扰动模型,从而实现了对突变调控功能的预测。针对非编码区域训练的DeepSEA模型则能够直接从大规模注释数据中学习序列特征,以单核苷酸敏感性预测序列改变的染色质效应,为基因编辑靶点的预测提供了新的思路。

5 总结与展望

鸡在人类世界中扮演着肉用蛋白供给者和鸟类模式动物的双重身份,意味着对鸡的遗传研究将一直是科学界的重点方向。如果把中心法则比作一座桥梁,基因型和表型就是桥的两岸,随着组学技术的蓬勃发展,这座桥正从两岸向中心逐渐合拢,让我们更清晰地理解基因组上每一个序列对表型各司其职又珠联璧合的完美遗传机制。而基因编辑就像是桥墩,通过反向遗传学的验证拱卫着组学分析的结果。2024年中央一号文件再次强调了农业科技创新和种业振兴,目前多组学数据的应用正在推动鸡育种进入大数据时代,紧随而来的智能化基因育种时代也将到来。信息科学与生命科学的交叉融合为遗传解析实现质的飞跃提供了新的契机,例如基因组大模型Evo^[65],可以准确解码和设计从分子到基因组规模的DNA、RNA和蛋白质序列,打通中心法则。实验室中的基因必需性实验可能需要半年以上的努力,而用神经网络进行几次正向传递就可替代这一过程。在基因组预测方面,已有尝试将多组学数

据与各类机器学习或深度学习算法相结合进行遗传解析和育种评价, 通过无先验假设的形式, 更客观地评估基因组各类非加性效应, 达到更好、更快的预测效果。

目前, 人工智能算法如何与各维度数据有效接驳, 建立基于多模态的评价预测模型, 并实现预测结果的可解释性, 将决定人工智能在本领域的未来前景。

参考文献

- Fuentes-Pardo A P, Ruzzante D E. Whole-genome sequencing approaches for conservation biology: advantages, limitations and practical recommendations. *Mol Ecol*, 2017, 26: 5369–5406
- Nicholas F W. Online mendelian inheritance in animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Res*, 2003, 31: 275–277
- Li J, Bed'hom B, Marthey S, et al. A missense mutation in *TYRP1* causes the chocolate plumage color in chicken and alters melanosome structure. *Pigment Cell Melanoma Res*, 2019, 32: 381–390
- Vaez M, Follett S A, Bed'hom B, et al. A single point-mutation within the melanophilin gene causes the lavender plumage colour dilution phenotype in the chicken. *BMC Genet*, 2008, 9: 7
- Dong J, He C, Wang Z, et al. A novel deletion in KRT75L4 mediates the frizzle trait in a Chinese indigenous chicken. *Genet Sel Evol*, 2018, 50: 68
- Guo Y, Tian J, Song C, et al. Mapping and functional dissection of the rumple trait in piao chicken identifies a causal loss of function mutation in the novel gene *Rum*. *Mol Biol Evol*, 2023, 40: msad273
- Guo Y, Gu X, Sheng Z, et al. A complex structural variation on chromosome 27 leads to the ectopic expression of HOXB8 and the muffs and beard phenotype in chickens. *PLoS Genet*, 2016, 12: e1006071
- Wu Z, Derks M F L, Dibbitts B, et al. A novel loss-of-function variant in transmembrane protein 263 (TMEM263) of autosomal dwarfism in chicken. *Front Genet*, 2018, 9: 193
- Kerje S, Sharma P, Gunnarsson U, et al. The dominant white, dun and smoky color variants in chicken are associated with insertion/deletion polymorphisms in the PMEL17 GeneSequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AY636124, AY636125, AY636126, AY636127, AY636128, AY636129. *Genetics*, 2004, 168: 1507–1518
- Wang Z, Qu L, Yao J, et al. An EAV-HP insertion in 5' flanking region of *SLCO1B3* causes blue eggshell in the chicken. *PLoS Genet*, 2013, 9: e1003183
- Kerje S, Lind J, Schütz K, et al. Melanocortin 1-receptor (*MC1R*) mutations are associated with plumage colour in chicken. *Anim Genet*, 2003, 34: 241–248
- Dorshorst B, Okimoto R, Ashwell C. Genomic regions associated with dermal hyperpigmentation, polydactyly and other morphological traits in the silk chicken. *J Heredity*, 2010, 101: 339–350
- Zhu T, Liu M, Peng S, et al. A deletion upstream of *SOX10* causes light yellow plumage colour in chicken. *Genes*, 2022, 13: 327
- Wright D, Boije H, Meadows J R S, et al. Copy number variation in intron 1 of *SOX5* causes the pea-comb phenotype in chickens. *PLoS Genet*, 2009, 5: e1000512
- Imsland F, Feng C, Boije H, et al. The rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet*, 2012, 8: e1002775
- Dorshorst B, Harun-Or-Rashid M, Bagherpoor A J, et al. A genomic duplication is associated with ectopic eomesodermin expression in the embryonic chicken comb and two duplex-comb phenotypes. *PLoS Genet*, 2015, 11: e1004947
- Li J, Lee M O, Davis B W, et al. The crest phenotype in domestic chicken is caused by a 195 bp duplication in the intron of *HOXC10*. *G3 Genes Genomes Genet*, 2021, 11: jkaa048
- Wang Y, Gao Y, Imsland F, et al. The crest phenotype in chicken is associated with ectopic expression of *HOXC8* in cranial skin. *PLoS One*, 2012, 7: e34012
- Li J, Lee M O, Davis B W, et al. Mutations upstream of the *TBX5* and *PITX1* transcription factor genes are associated with feathered legs in the domestic chicken. *Mol Biol Evol*, 2020, 37: 2477–2486
- Youngworth I, Delany M E. A premature stop codon in *RAF1* is the priority candidate causative mutation of the inherited chicken wingless-2 developmental syndrome. *Genes*, 2019, 10: 353
- Feng C, Gao Y, Dorshorst B, et al. A *cis*-regulatory mutation of *PDSS2* causes silky-feather in chickens. *PLoS Genet*, 2014, 10: e1004576
- Hellström A R, Sundström E, Gunnarsson U, et al. *Sex-linked barring* in chickens is controlled by the *CDKN2A/B* tumour suppressor locus. *Pigment Cell Melanoma Res*, 2010, 23: 521–530
- Wang Y, Li J, Feng C, et al. Transcriptome analysis of comb and testis from Rose-comb Silky chicken (R1/R1) and Beijing Fatty wild type chicken (r/r). *Poultry Sci*, 2017, 96: 1866–1873

- 24 Chang C M, Coville J L, Coquerelle G, et al. Complete association between a retroviral insertion in the tyrosinase gene and the recessive white mutation in chickens. *BMC Genomics*, 2006, 7: 19
- 25 Jin S, Zhu F, Wang Y, et al. Deletion of Indian hedgehog gene causes dominant semi-lethal Creeper trait in chicken. *Sci Rep*, 2016, 6: 30172
- 26 International Chicken Genome Sequencing C. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 2004, 432: 695–716
- 27 Wang K, Hu H, Tian Y, et al. The chicken pan-genome reveals gene content variation and a promoter region deletion in *IGF2BP1* affecting body size. *Mol Biol Evol*, 2021, 38: 5066–5081
- 28 Li M, Sun C, Xu N, et al. *De novo* assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Mol Biol Evol*, 2022, 39: msac066
- 29 Rice E S, Alberdi A, Alfieri J, et al. A pangenome graph reference of 30 chicken genomes allows genotyping of large and complex structural variants. *BMC Biol*, 2023, 21: 267
- 30 Huang Z, Xu Z, Bai H, et al. Evolutionary analysis of a complete chicken genome. *Proc Natl Acad Sci USA*, 2023, 120: e2216641120
- 31 Riddle N C, Elgin S C R. The *Drosophila* dot chromosome: where genes flourish amidst repeats. *Genetics*, 2018, 210: 757–772
- 32 Ren J, Kou W, Xu Y, et al. Pan-genome analyses add ~1000 genes to the “complete” genome assembly of chicken. *J Genet Genomics*, 2025, 52: 116–119
- 33 Hu Z L, Park C A, Reecy J M. Bringing the animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. *Nucleic Acids Res*, 2022, 50: D956–D961
- 34 Tam V, Patel N, Turcotte M, et al. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*, 2019, 20: 467–484
- 35 Liu S, Gao Y, Canela-Xandri O, et al. A multi-tissue atlas of regulatory variants in cattle. *Nat Genet*, 2022, 54: 1438–1447
- 36 Teng J, Gao Y, Yin H, et al. A compendium of genetic regulatory effects across pig tissues. *Nat Genet*, 2024, 56: 112–123
- 37 Guan D L, Bai Z H, Zhu X M, et al. Genetic regulation of gene expression across multiple tissues in chickens. *Nat Genet*, 2025. doi: 10.1038/s41588-025-02155-9
- 38 Mai J, Lu M, Gao Q, et al. Transcriptome-wide association studies: recent advances in methods, applications and available databases. *Commun Biol*, 2023, 6: 899
- 39 Smith GD, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *Intl J Epidemiol*, 2003, 32: 1–22
- 40 Hukku A, Sampson M G, Luca F, et al. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *Am J Hum Genet*, 2022, 109: 825–837
- 41 Andersson L, Archibald A L, Bottema C D, et al. Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. *Genome Biol*, 2015, 16: 57
- 42 Pan Z, Wang Y, Wang M, et al. An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. *Sci Adv*, 2023, 9: eade1204
- 43 Zhu X N, Wang Y Z, Li C, et al. Chicken chromatin accessibility atlas accelerates epigenetic annotation of birds and gene fine-mapping associated with growth traits. *Zool Res*, 2023, 44: 53–62
- 44 Fu W, Wang R, Xu N, et al. Galbase: a comprehensive repository for integrating chicken multi-omics data. *BMC Genomics*, 2022, 23: 364
- 45 Cao X, Wang Y, Shu D, et al. Food intake-related genes in chicken determined through combinatorial genome-wide association study and transcriptome analysis. *anim Genet*, 2020, 51: 741–751
- 46 Wang Y, Cao X, Luo C, et al. Multiple ancestral haplotypes harboring regulatory mutations cumulatively contribute to a QTL affecting chicken growth traits. *Commun Biol*, 2020, 3: 472
- 47 Shen L, Bai X, Zhao L, et al. Integrative 3D genomics with multi-omics analysis and functional validation of genetic regulatory mechanisms of abdominal fat deposition in chickens. *Nat Commun*, 2024, 15: 9274
- 48 Wang D, Tan L, Zhi Y, et al. Genome-wide variation study and inter-tissue communication analysis unveil regulatory mechanisms of egg-laying performance in chickens. *Nat Commun*, 2024, 15: 7069
- 49 Stuart T, Satija R. Integrative single-cell analysis. *Nat Rev Genet*, 2019, 20: 257–272
- 50 Feregrino C, Sacher F, Parnas O, et al. A single-cell transcriptomic atlas of the developing chicken limb. *BMC Genomics*, 2019, 20: 401
- 51 Liukang C, Zhao J, Tian J, et al. Deciphering infected cell types, hub gene networks and cell-cell communication in infectious bronchitis virus via single-cell RNA sequencing. *PLoS Pathog*, 2024, 20: e1012232
- 52 Meuwissen T H E, Hayes B J, Goddard M E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157: 1819–1829
- 53 Wiggans G R, VanRaden P M, Cooper T A. The genomic evaluation system in the United States: past, present, future. *J Dairy Sci*, 2011, 94: 3202–3211

- 54 Choi S W, Mak T S H, O'Reilly P F. Tutorial: a guide to performing polygenic risk score analyses. [Nat Protoc](#), 2020, 15: 2759–2772
- 55 Zhao L, Tang P, Luo J, et al. Genomic prediction with NetGP based on gene network and multi-omics data in plants. [Plant Biotechnol J](#), 2025, 23: 1190–1201
- 56 Zhao T, Zeng J, Cheng H, et al. Extend mixed models to multilayer neural networks for genomic prediction including intermediate omics data. [Genetics](#), 2022, 221: iyac034
- 57 Novakovsky G, Dexter N, Libbrecht M W, et al. Obtaining genetics insights from deep learning via explainable artificial intelligence. [Nat Rev Genet](#), 2023, 24: 125–137
- 58 Mota L F M, Arikawa L M, Santos S W B, et al. Benchmarking machine learning and parametric methods for genomic prediction of feed efficiency-related traits in Nellore cattle. [Sci Rep](#), 2024, 14: 6404
- 59 Salter D W, Smith E J, Hughes S H, et al. Transgenic chickens: insertion of retroviral genes into the chicken germ line. [Virology](#), 1987, 157: 236–240
- 60 van de Lavoie M C, Diamond J H, Leighton P A, et al. Germline transmission of genetically modified primordial germ cells. [Nature](#), 2006, 441: 766–769
- 61 Schusser B, Collarini E J, Yi H, et al. Immunoglobulin knockout chickens via efficient homologous recombination in primordial germ cells. [Proc Natl Acad Sci USA](#), 2013, 110: 20170–20175
- 62 Taylor L, Carlson D F, Nandi S, et al. Efficient TALEN-mediated gene targeting of chicken primordial germ cells. [Development](#), 2017, 144: dev.145367
- 63 Ballantyne M, Woodcock M, Doddamani D, et al. Direct allele introgression into pure chicken breeds using Sire Dam Surrogate (SDS) mating. [Nat Commun](#), 2021, 12: 659
- 64 Chen K M, Wong A K, Troyanskaya O G, et al. A sequence-based global map of regulatory activity for deciphering human genetics. [Nat Genet](#), 2022, 54: 940–949
- 65 Nguyen E, Poli M, Durrant M G, et al. Sequence modeling and design from molecular to genome scale with evo. [Science](#), 2024, 386: eado9336

Summary for “从孟德尔遗传到整合组学: 鸡重要性状遗传机制解析及育种应用”

From Mendelian inheritance to multi-omics integration: deciphering the genetic mechanism of important traits and breeding applications in chickens

Yuzhe Wang^{1,2}, Hanyu Wu^{1,2} & Xiaoxiang Hu^{1,2*}

¹ State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing 100193, China

² National Science Center for Model Animals, China Agricultural University, Beijing 100193, China

* Corresponding author, E-mail: huxx@cau.edu.cn

Chickens serve as the predominant source of meat globally and represent a critical model organism in developmental biology, immunology, and physiology research. Chickens were among the first animals to demonstrate Mendelian inheritance laws. As early as 1897, William Bateson conducted hybridization experiments on chicken comb morphology. Over the past century, 59 quantitative traits linked to causal genes have been identified in chickens, primarily related to phenotypic traits. Since the release of the initial chicken genome in 2004, seven updated versions have been published. However, a single reference genome cannot fully capture the species' complete genetic diversity, driving the advancement of chicken genome research into the pan-genome era. Current efforts focus on constructing a multi-variety graphical genome using telomere-to-telomere (T2T) assembly approaches. To date, the Global Chicken Reference Panel (GCRP) consortium has generated a genome panel of more than 10,000 samples. The Functional Annotation of Animal Genomes (FAANG) consortium has annotated over 1.5 million regulatory elements in chickens. At the transcriptional level, the Chicken Genotype-Tissue Expression (GTEx) Project has mapped diverse molecular quantitative trait loci (QTLs) and is expanding its scope to incorporate single-cell omics. These multi-omics data have facilitated the creation of multiple associated databases and have also transformed the research methodologies for economic traits such as growth and reproduction in chickens. The shift from genome-wide association studies (GWAS) to multi-omics colocalization and integrative analyses has deepened our understanding of how multi-dimensional large-scale data interact with the central dogma, thereby improving the precision of complex trait gene mapping. Downstream applications leveraging these foundational data include two prominent breeding strategies: genomic selection and gene-editing-based design breeding. In chickens, genomic selection primarily addresses the issue of prediction accuracy, propelling the development and application of weighted genomic best linear unbiased prediction (GBLUP) and diverse Bayesian methodologies. The distinct developmental biology of chickens imposes certain limitations on the application of gene-editing technology; however, more stable and efficient editing systems are continually emerging to provide the biological resources required by humans in a flexible, efficient, and standardized manner. Meanwhile, artificial intelligence (AI) models supported by big data are increasingly integrated into animal genetics and breeding. In the future, multimodal prediction and evaluation frameworks may be established to enhance the interpretability of biological outcomes. This review first outlines research progress on Mendelian traits in chickens, followed by a summary of advancements in chicken genome assembly. Next, we introduce methodologies and breakthroughs in deciphering the genetic mechanisms underlying complex traits. Subsequently, we discuss the evolving directions of genomic selection and molecular design breeding within the context of multi-omics integration and large language models (LLMs). Finally, we synthesize the current landscape of chicken genetic research and anticipate future prospects at the intersection of poultry genomics and artificial intelligence.

chicken, genome assembly, gene mapping, multi-omics integration, genomic selection, gene-editing

doi: [10.1360/TB-2024-1399](https://doi.org/10.1360/TB-2024-1399)