

基于基因表达谱的黑腹果蝇组织标志基因的鉴定*

李磊¹, 朵泓睿¹, 李勃¹, 李映红², 郝友进¹

(1. 重庆师范大学 生命科学学院, 重庆 401331; 2. 重庆邮电大学 生物信息学院, 重庆 400065)

摘要:为解决目前黑腹果蝇(*Drosophila melanogaster*)组织标志基因鉴定结果不统一、不准确的问题,分别利用 Fly Cell Atlas 和 FlyAtlas2 数据库中 7 个共有组织的转录组学数据,通过 Seurat 以及计算差异表达倍数的方法鉴定果蝇组织标志基因;经过跨数据集交叉验证、文献比对和基因功能检索,最终确定准确的组织标志基因。从上述 2 个数据集中分别鉴定出 575 和 5 202 个组织标志基因,经过跨数据集交叉验证后得到 366 和 2 671 个组织标志基因,通过文献比对和 FlyBase 数据库基因功能检索,共得到 55 个核心组织标志基因和 22 个候选组织标志基因。建立了准确的果蝇组织标志基因集,为将来果蝇组织类型区分和细胞类型标志基因的鉴定提供了重要参考。

关键词:黑腹果蝇;组织标志基因;转录组;单细胞转录组

中图分类号:Q966

文献标志码:A

文章编号:1672-6693(2024)06-0069-07

组织标志基因是在某一种组织中高表达而在其他组织中低表达或不表达的一类基因。准确的组织标志基因鉴定不仅有利于组织划分,还有助于了解组织的发育和分化,例如有研究者通过鉴定胚胎不同发育时期的标志基因定义了囊胚的 3 种细胞谱系,为研究胚胎的发生及与干细胞的关系提供了分子蓝图^[1]。此外,还有研究者利用组织标志基因可对肿瘤和非肿瘤组织进行区分^[2],以利于后续疾病诊断和药物开发。由此可见,使用准确的组织标志基因对发育生物学和医学领域的研究来说非常重要,而使用不准确的组织标志基因将导致相关研究走向错误的方向并严重影响研究结果的准确度。

黑腹果蝇(*Drosophila melanogaster*)作为模式生物之一,已被广泛应用于遗传学、发育生物学、医学等领域的研究中^[3-5]。果蝇组织标志基因推动了有关发育过程中不同组织和器官形成和分化机制的研究^[6]。作为个别疾病的研究模型,研究人员可通过组织标志基因的表达水平评估药物对组织的影响^[7]。因此,对果蝇组织标志基因的鉴定至关重要。在高通量测序技术成熟前,研究者主要通过研究基因功能与组织的关系^[8]、原位杂交^[9]等技术进行组织标志基因鉴定。但这些研究有的是基于单个组织进行实验,忽视了基因可能同时在不同组织中高表达的情况,因此鉴定出的组织标志基因准确性不高。此外,目前仍然没有研究系统全面地收集和整理果蝇的组织标志基因。

鉴于目前果蝇组织标志基因鉴定结果不统一、不准确的现状,系统全面地收集整理和鉴定果蝇组织标志基因十分必要。本研究利用 2 类转录组测序(RNA-Seq)数据即单细胞转录组测序(scRNA-Seq)和传统转录组测序(bulk RNA-Seq)的数据进行组织标志基因鉴定,然后通过跨数据集验证、文献比对及基因功能检索,共筛选出 77 个最适合的组织标志基因。本研究结果有望为将来果蝇组织区分或细胞类型标志基因鉴定提供参考,让果蝇在各个生物学及医学研究领域中得到更好的应用。

1 数据来源与基因鉴定

1.1 分析设计

本研究过程在 R 语言环境下实现。首先利用 Fly Cell Atlas 数据库^[10]中的 scRNA-Seq 数据和 FlyAtlas2 数据库^[11]中的 RNA-Seq 数据进行组织标志基因鉴定并进行跨数据集交叉验证。将验证后的组织标志基因再经

* 收稿日期:2023-11-11 修回日期:2024-02-02 网络出版时间:2024-12-25T14:04

资助项目:国家自然科学基金青年科学基金项目(No. 31871274);重庆市教育委员会科学技术研究计划项目(No. KJQN202100538)

第一作者简介:李磊,男,研究方向为生态学与生物信息学,E-mail:1393348077@qq.com;通信作者:郝友进,男,教授,博士,E-mail:1390593613@qq.com

网络出版地址:https://link.cnki.net/urlid/50.1165.N.20241225.1007.002

过文献比对和基因功能检索,最终确定核心和候选组织标志基因(图 1)。

1.2 数据的获取

检索 Fly Cell Atlas 和 FlyAtlas2 数据库得到 7 个共有组织:卵巢(ovary)、头部(head)、马氏管(malpighian tubule)、脂肪体(fat body)、气管(trachea)、精巢(testis)和肠道(gut)。分别从以上 2 个数据库下载这 7 个组织的 scRNA-Seq 数据和 RNA-Seq 数据。

1.3 组织标志基因鉴定

1.3.1 Fly Cell Atlas 数据组织标志基因的鉴定

以 scRNA-Seq 数据作为输入文件,使用 Seurat R 包^[12]进行组织标志基因鉴定,具体步骤有以下 9 步。

- 1) 表达矩阵构建:使用 Connect 函数提取单细胞数据中的基因名、细胞名及相应表达值,并构建表达矩阵。
- 2) Seurat 对象构建:删除表达矩阵中基因表达值小于或等于 300 的细胞和在小于或等于 3 个细胞中表达的基因,根据表达矩阵和细胞所属组织信息,利用 CreateSeuratObject 函数构建 Seurat 对象。
- 3) 细胞过滤:使用 Subset 函数选择细胞内总表达值在 300~6 000 的细胞,并删除表达线粒体基因的细胞(表达值大于或等于 5%)。
- 4) 数据标准化:使用 NormalizeData 函数中的 LogNormalize 方法对数据进行归一化,然后进行对数转换。
- 5) 高变基因筛选:使用 FindVariableFeature 函数中的 vst 方法选出前 2 000 个高变基因。
- 6) 数据降维:使用 ScaleData 函数将数据归一化处理后,通过主成分分析(principal components analysis, PCA)进行数据降维。
- 7) 细胞聚类:使用 FindNeighbors 函数在 PCA 空间内构建基于欧氏距离的 K 临近(K -nearest neighbors)图,聚类数(组织数) $n=7$ 。
- 8) 低维可视化:使用 RunUMAP 函数进行细胞聚类和可视化。
- 9) 组织标志基因筛选:使用 FindAllMarkers 函数进行基因差异表达分析,以差异表达倍数大于或等于 2 为标准筛选组织标志基因。

1.3.2 FlyAtlas2 数据组织标志基因鉴定

使用 utils 包中 download.file 函数下载 FlyAtlas2 数据库中每个基因在 7 个组织中的表达矩阵,并将基因的 FlyBase gene ID 转换为 Gene Symbol。为了避免 0 值影响,对表达矩阵中所有数值加上 0.01 后,进行底数为 2 的对数转换,即:

$$F = \log_2(a_x/b_x),$$

其中: a_x 和 b_x 分别代表基因 x 在组织 a 和 b 中的表达值,而 F 代表基因 x 在组织 a 和 b 中的表达值之比经过底数为 2 的对数转化后得到的差异表达倍数,以 $F \geq 2$ 为标准筛选组织标志基因。

1.4 组织标志基因验证

为保证组织标志基因的准确性和可靠性,首先对组织标志基因进行跨数据集交叉验证。将由 Fly Cell Atlas 数据鉴定的组织标志基因代入 FlyAtlas2 数据中,以前文 1.3.2 节中的计算方法进行差异表达倍数计算,以 F 大于或等于 2 为标准筛选组织标志基因。FlyAtlas2 数据鉴定的组织标志基因采取同样的验证方法。另外,以“*Drosophila melanogaster*”“marker gene”和组织名称“fat body”“gut”“head”“malpighian tubule”“ovary”“testis”“trachea”为关键词检索 PubMed 文献数据库(<https://pubmed.ncbi.nlm.nih.gov/>),提取整合文献中有实验证据的组织标志基因。通过数据集交叉验证、文献比对以及 FlyBase 数据库^[13]基因功能查询,最终筛选出核心和候选组织标志基因。

2 结果与分析

2.1 Fly Cell Atlas 数据中的组织标志基因

整合 Fly Cell Atlas 数据库中 7 个组织的 scRNA-Seq 数据后,得到的基因表达谱共含有 10 580 个基因和 255 943 个细胞,后者经进一步过滤后得到 255 938 个细胞。数据经 Seurat 分析和降维后的 UMAP(uniform manifold approximation and projection)可视化结果如图 2 所示。从图中可以看出:气管与肠道、卵巢的细胞在二维空间有一定交集但边界清晰,因而不影响组织标志基因鉴定。除精巢外,其它组织被分为多个细胞团,说明它们具有区分明显的多种细胞类型。通过差异基因表达分析共鉴定出 575 个组织标志基因(存在 1 个基因属于多个组织的标志基因的情况),脂肪体、肠道、头部、马氏管、卵巢、精巢和气管的组织标志基因分别有 66、51、177、

73、24、173 和 16 个。

2.2 FlyAtlas2 数据中的组织标志基因

通过整合 FlyAtlas2 数据,获得 13 959 个基因在 7 个不同组织的表达矩阵。根据差异表达倍数大于或等于 2 的条件,共鉴定出 5 202 个组织标志基因(存在 1 个基因属于多个组织的标志基因的情况)。脂肪体、肠道、头部、马氏管、卵巢和精巢中的组织标志基因个数分别为 219、644、1 278、402、216 和 2 510。主要组织标志基因在不同组织的表达情况如图 3 所示。

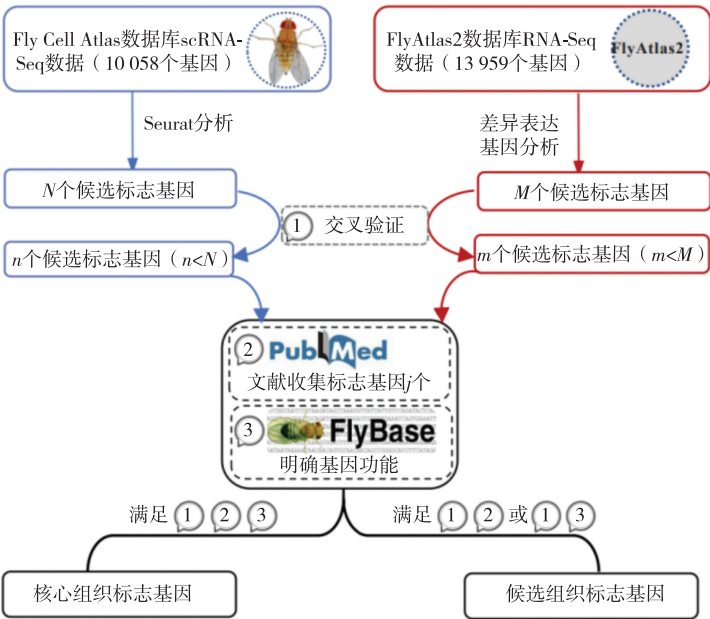


图 1 黑腹果蝇组织标志基因鉴定流程
Fig. 1 Process for identification of tissue marker genes in *D. melanogaster*

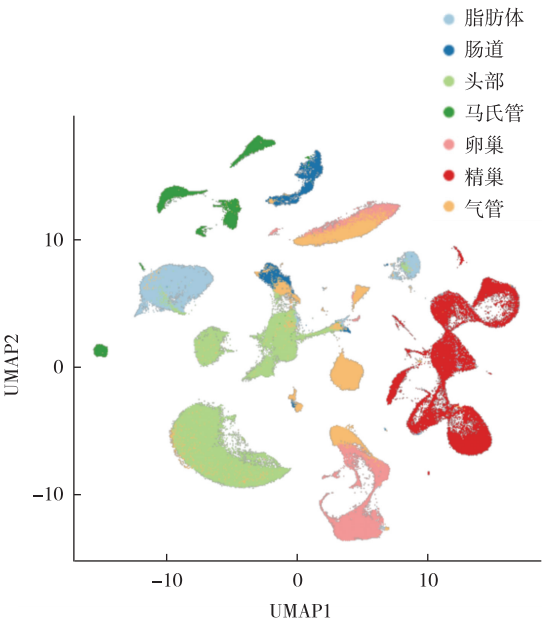
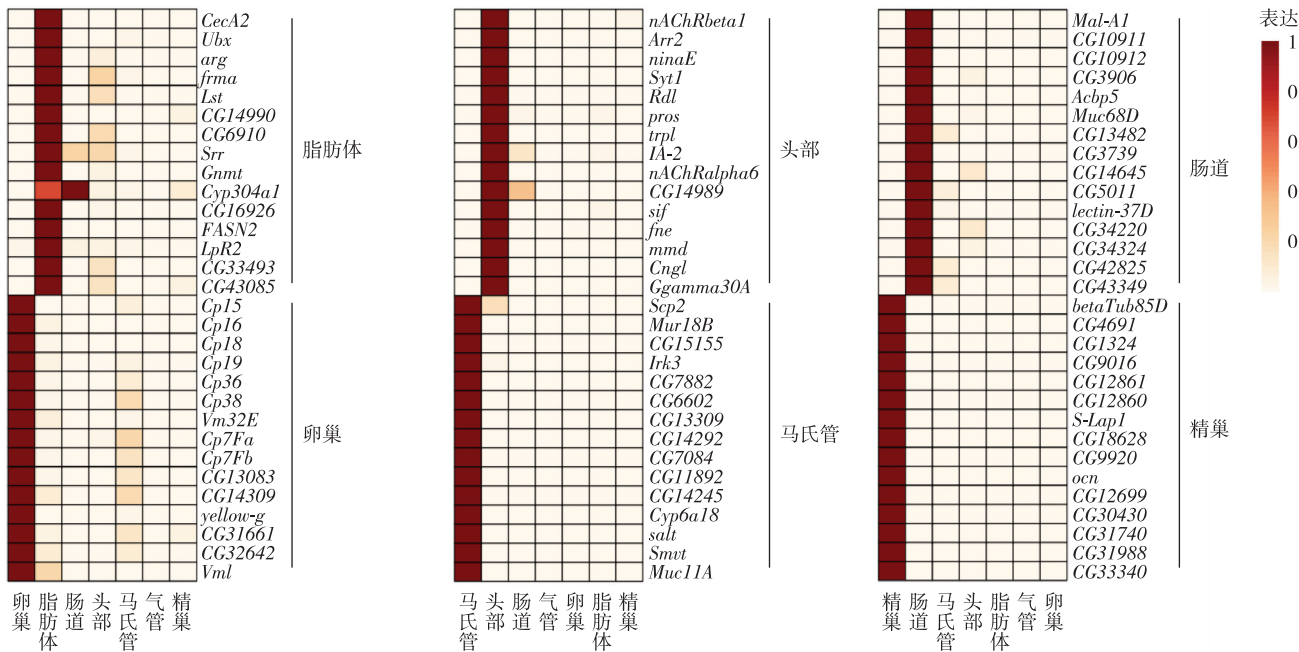


图 2 Fly Cell Atlas 数据低维可视化的 UMAP 图
Fig. 2 Low-dimensional visualization of UMAP plot for Fly Cell Atlas data



注:原始数据已进行归一化处理,将基因在不同组织的表达值等比例放缩到 0~1。

图 3 FlyAtlas2 数据中主要组织标志基因的热图
Fig. 3 Heatmap of major marker genes in FlyAtlas2 data

2.3 组织标志基因的交叉验证的结果

通过跨数据集交叉验证,分别从 Fly Cell Atlas 和 FlyAtlas2 数据中筛选出 366 和 2 670 个适用于外部数据的组织标志基因;整合 271 篇文献的结果,共有 393 个组织标志基因(存在 1 个基因属于多个组织的标志基因的情况)。如图 4 所示,本研究所鉴定的基因当中仅有 36 个通过数据集交叉验证并被以往文献报道,各个组织标志基因集均有大量组织标志基因不能在其他组织标志基因集中复现。

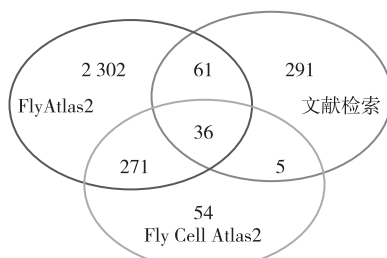


图 4 数据集交叉验证和文献收集的组织标志基因比较

Fig. 4 Comparison of marker genes in cross-data validation and literature searching

基于以上情况,本研究按照如下条件进行最终筛选:1) 通过跨数据集验证;2) 被以往文献报道;3) 有明确的和组织相关的基因功能。满足条件 1)、2)、3) 的基因为核心组织标志基因,满足条件 1)、2) 或条件 1)、3) 的基因为候选组织标志基因。最终在 7 个组织中鉴定出 55 个核心组织标志基因和 22 个候选组织标志基因(表 1)。

表 1 黑腹果蝇组织标志基因

Tab. 1 Tissue marker genes in *D. melanogaster*

组织	组织标志基因
肠道	<i>Amy-p</i> 、 <i>Amy-d</i> 、 <i>betaTry</i> 、 <i>alphaTry</i> 、 <i>thetaTry</i> 、 <i>Mal-A1</i> 、 <i>Mal-A6</i> 、 <i>mesh</i> 、 <i>LManVI</i> 、 <i>Pgant4</i> 、 <i>CG43774</i> 、 <i>PGRP-SC1a</i> 、 <i>CG5065</i>
头部	<i>Rh2</i> 、 <i>Rh3</i> 、 <i>Rh6</i> 、 <i>Rh5</i> 、 <i>sNPF</i> 、 <i>ninaC</i> 、 <i>ninaE</i> 、 <i>pros</i> 、 <i>Pdh</i> 、 <i>SoxN</i> 、 <i>retinin</i> 、 <i>Cpr72Ec</i> 、 <i>Drgx</i> 、 <i>bi</i> 、 <i>Sh</i>
马氏管	<i>Irk1</i> 、 <i>Irk3</i> 、 <i>Dh31-R</i> 、 <i>Eglp2</i> 、 <i>CapaR</i> 、 <i>salt</i> 、 <i>Uro</i> 、 <i>Tsh</i> 、 <i>CG16762</i> 、 <i>Lkr</i> 、 <i>SecCl</i> 、 <i>Alp4</i>
卵巢	<i>CP19</i> 、 <i>CP15</i> 、 <i>Cp18</i> 、 <i>CP36</i> 、 <i>Vm32E</i> 、 <i>Cp38</i> 、 <i>CG31926</i> 、 <i>Vml</i> 、 <i>Femcoat</i> 、 <i>CP7Fb</i> 、 <i>yellow-g</i> 、 <i>yellow-g2</i>
精巢	<i>kl-2</i> 、 <i>kl-3</i> 、 <i>kl-5</i> 、 <i>w-cup</i> 、 <i>c-cup</i> 、 <i>p-cup</i> 、 <i>CG18628</i> 、 <i>soti</i> 、 <i>whip</i> 、 <i>sunz</i> 、 <i>hale</i> 、 <i>aly</i> 、 <i>Rbp4</i> 、 <i>kmg</i> 、 <i>dany</i> 、 <i>fest</i> 、 <i>CG3927</i> 、 <i>schuy</i> 、 <i>Nep5</i>
气管	<i>trh</i> 、 <i>wat</i>
脂肪体	<i>ACC</i> 、 <i>Gnmt</i> 、 <i>CecA2</i> 、 <i>LpR2</i>

注:加粗部分为核心组织标志基因,不加粗部分为候选组织标志基因。

3 讨论

在组织标志基因鉴定过程中,本研究发现仅有约占文献总数 1/4 的文献中的组织标志基因与转录组学数据鉴定结果相吻合。深入分析后发现,有关文献中的部分组织标志基因组织表达的特异性不高,例如基因 *Mhc* 编码的运动蛋白与肌肉运动密切相关且在肌肉细胞中高表达,然而肌肉细胞广泛分布在肠道、头部、马氏管等多个组织中,因此该基因不适合成为单一组织的标志基因。另外有研究通过单细胞核转录组测序(snRNA-Seq)发现基因 *FASN2* 在绛色细胞中特异性表达^[14],但本研究在 FlyAtlas2 数据库中发现该基因在心脏中表达量最高,在脂肪体中的表达量次之,因此 *FASN2* 的多组织表达特点使之不能作为脂肪体的组织标志基因。此外,又如基因 *elav* 在神经细胞的发育过程中调节神经细胞的分型^[15]且已被作为果蝇神经细胞的组织标志基因^[16],但由于果蝇的神经系统存在于大脑和胸腹神经节中,因此该基因可以作为神经细胞的标志基因而不能作为头部的组织标志基因。

基因 *Amy* 编码的淀粉酶在昆虫碳水化合物代谢中起到了重要作用;在碳水化合物存在的条件下,*Amy-p*、*Amy-d* 可在果蝇肠道中高表达^[17]。Li 等人^[18]利用 scRNA-Seq 数据进行细胞分型时已将 *Amy-p* 作为果蝇前肠细胞的组织标志基因。基于 scRNA-Seq 分析,有研究发现 *alphaTry* 和 *betaTry* 这 2 个胰蛋白酶基因在前肠

中特异性高表达,并将它们定为肠道细胞的组织标志基因^[19]。编码溶酶体 α -甘露糖苷酶的 *LManVI* 虽未被作为肠道的组织标志基因,但通过凝胶电泳实验在果蝇肠道中却能检测到该酶的存在^[20]。在分析过程中,本研究发现 *LManVI* 在肠道中的表达值是在其他组织中表达值的 18 倍。因此,*LManVI* 可以作为肠道的候选组织标志基因。

有研究通过原位杂交技术发现,*Rh2* 和 *Rh3* 分别在果蝇单眼和复眼的感光细胞中特异性表达^[9],*Rh5* 和 *Rh6* 在果蝇复眼的感光细胞中特异性表达^[8,21]。上述这些基因编码的光敏感蛋白对果蝇的视觉至关重要,因此可作为头部或眼部的核心组织标志基因。*sNPF* 编码的短神经肽 F 参与调节食物摄入量和体型大小,Johard 等人^[22]通过原位杂交实验和免疫血清分析发现该基因在果蝇头部的凯尼恩细胞(Kenyon cell)中特异性表达。此外,*bi* 编码的 T-box 转录因子能促进大脑和眼球的发育,而 Michki 等人^[23]通过 scRNA-Seq 分析发现 *bi* 在果蝇头部视叶(optic lobe)区域中高表达。

在马氏管的 8 个核心组织标志基因和 4 候选组织标志基因中,Evans 等人^[24]通过原位杂交将钾离子通道蛋白编码基因 *Irk1*、*Irk2* 和 *Irk3* 定位到马氏管的主节段,进一步的定量 RT-PCR 实验证明了它们在马氏管中高表达。但本研究发现 *Irk2* 也在果蝇后肠中高表达,且目前尚无更多实验证明 *Irk2* 是马氏管的组织标志基因,因此本研究只将 *Irk1* 和 *Irk3* 作为马氏管的核心组织标志基因。另外,基因 *Dh31-R* 和 *CapaR* 在马氏管中均有明显的高表达,两者分别编码利尿激素 Dh31 受体和 Capa 神经肽受体,而这 2 种受体都有利尿功能,可控制果蝇排泄并保持体液平衡。虽然 *CG16762* 目前没有明确的功能,但本研究发现它在马氏管中的表达量是在它在其他组织中平均表达量的 64 倍。同时有研究表明 *CG16762* 在马氏管中特异性表达^[25]。因此,本研究认为 *CG16762* 可以作为马氏管的候选组织标志基因。

编码卵黄膜蛋白的 *Vm32E* 以及编码绒毛膜蛋白的 *Cp19*、*Cp15*、*Cp18*、*Cp36*、*Cp38* 等在卵形成过程中的不同时期具有选择性表达特点^[26-27]。*Vm32E* 在卵形成阶段 10 特异性表达,*Cp19*、*Cp36* 和 *Cp38* 在卵形成阶段 10~14 特异性表达,而 *Cp18* 在卵形成阶段 11~14 特异性表达。Gigliotti 等人^[28]利用 Northern blot 实验比较了不同年龄雄性和雌性果蝇中 *Vm32E* 的表达模式,发现 *Vm32E* 的表达与卵巢发育高度相关,并进一步分析发现该基因在卵形成阶段 10 有明显的高表达。虽然 *yellow-g* 和 *yellow-g2* 在 FlyBase 数据库中没有明确的功能注释,但有研究表明两者可能编码多巴胺转化酶。它们与早期绒毛膜基因的表达一致,且 *yellow-g* 已被证明对卵黄膜的完整性至关重要^[29]。因此,本研究认为 *yellow-g* 和 *yellow-g2* 可以作为卵巢的候选组织标志基因。

定位于果蝇 Y 染色体长臂上的 *kl-2*、*kl-3* 和 *kl-5* 编码的微观动力蛋白是轴丝的关键组成部分,轴丝是形成精子鞭毛核心的基于微管的细胞骨架结构^[30]。有研究发现上述 3 个基因只有在精子伸长过程中的轴丝伸长时期特异性高表达^[31]。此外,利用 CRISPR/Cas9 技术使 *kl-2* 和 *kl-5* 产生突变会导致雄性果蝇不育^[32]。Barreau 等人^[33]发现 *p-cup*、*c-cup* 和 *w-cup* 在精子发育的后期高表达,而最近一项单细胞转录组学研究也将上述 3 个基因作为精子标志基因用于区分精子不同发育时期^[34]。总之,以上基因均可以作为精巢的核心组织标志基因。

本研究发现气管中的组织标志基因较少,原因在于果蝇气管系统由延伸到整个体腔的分支小管网络组成且具有分布广特异性差的特点,而 FlyAtlas2 数据库中也并没有成虫的气管组织数据用于鉴定。结合 Fly Cell Atlas 数据鉴定结果、FlyBase 数据库基因功能注释以及其他文献发现 *trh* 编码气管形成所需的 bHLH-PAS 核转录因子,而所有气管基因的表达都需要该因子,*trh* 基因的缺失将导致气管细胞无法内陷形成管^[35]。*wat* 编码的脂肪酰辅酶 A 还原酶参与 C_{24} 和 C_{26} 长链脂肪酸的还原。还原后的脂肪醇作为疏水性物质覆盖气管内壁,促进气管由充满水的状态过渡到充气状态^[36]。*trh* 和 *wat* 可作为目前最佳的气管组织标志基因。

脂肪体分布在昆虫全身各处,多位于体皮下方并围绕肠道和生殖器官,有储存和利用营养物质、分泌抗菌肽、参与解毒等多种功能^[37],分布广泛和功能多样的特点使得该组织的特异性标志基因较少。*ACC* 编码乙酰辅酶 A 羧化酶,后者催化乙酰辅酶 A 羧化为丙二酰辅酶 A,而丙二酰辅酶 A 是脂肪酸合成的限速底物。有相关研究表明该基因高表达于脂肪体细胞,利用 RNAi 敲除 *ACC* 后将导致甘油三酯存储量急剧下降^[38]。到目前为止该基因虽未被列为果蝇组织标志基因,但它的功能与脂肪体高度相关,且 Fly Cell Atlas 数据显示该基因高表达于脂肪体。因此,本研究认为 *ACC* 可作为脂肪体的候选组织标志基因。

总而言之,本研究基于 scRNA-Seq 和 RNA-Seq 数据,对果蝇 7 个主要组织进行了标志基因的鉴定。通过跨数据集交叉验证、文献比对以及基因功能查询,最终鉴定出 55 个核心组织标志基因和 22 个候选组织标志基因。研究结果部分解决了目前果蝇组织标志基因不统一和不准确的问题,也为将来果蝇组织分型和细胞类型标志基因鉴定提供了重要参考。

参考文献:

- [1] BLAKELEY P, FOGARTY N M E, Del VALLE I, et al. Defining the three cell lineages of the human blastocyst by single-cell RNA-seq[J]. *Development*, 2015, 142(20): 3151-3165.
- [2] Van't VEER L J, PAIK S, HAYES D F. Gene expression profiling of breast cancer: a new tumor marker[J]. *Journal of Clinical Oncology*, 2005, 23(8): 1631-1635.
- [3] McGURK L, BERSON A, BONINI N M. *Drosophila* as an *in vivo* model for human neurodegenerative disease[J]. *Genetics*, 2015, 201(2): 377-402.
- [4] BIER E. *Drosophila*, the golden bug, emerges as a tool for human genetics[J]. *Nature Reviews. Genetics*, 2005, 6(1): 9-23.
- [5] TOLWINSKI N S. Introduction: *Drosophila*-a model system for developmental biology[J]. *Journal of Developmental Biology*, 2017, 5(3): 9.
- [6] WAWERSIK S, MAAS R L. Vertebrate eye development as modeled in *Drosophila*[J]. *Human Molecular Genetics*, 2000, 9(6): 917-925.
- [7] PANDEY U B, NICHOLS C D. Human disease models in *Drosophila melanogaster* and the role of the fly in therapeutic drug discovery[J]. *Pharmacological Reviews*, 2011, 63(2): 411-436.
- [8] CHOU W H, HALL K J, WILSON D B, et al. Identification of a novel *Drosophila* opsin reveals specific patterning of the R7 and R8 photoreceptor cells[J]. *Neuron*, 1996, 17(6): 1101-1115.
- [9] POLLOCK J A, BENZER S J N. Transcript localization of four opsin genes in the three visual organs of *Drosophila*; *RH2* is ocellus specific[J]. *Nature*, 1988, 333(6175): 779-782.
- [10] LI H J, JANSSENS J, de WAEGENEER M, et al. Fly Cell Atlas: a single-nucleus transcriptomic atlas of the adult fruit fly[J]. *Science*, 2022, 375(6584): eabk2432.
- [11] LEADER D P, KRAUSE S A, PANDIT A, et al. FlyAtlas 2: a new version of the *Drosophila melanogaster* expression atlas with RNA-Seq, miRNA-Seq and sex-specific data[J]. *Nucleic Acids Research*, 2018, 46(D1): D809-D815.
- [12] HAO Y H, HAO S, ANDERSEN-NISSEN E, et al. Integrated analysis of multimodal single-cell data[J]. *Cell*, 2021, 184(13): 3573-3587.
- [13] THURMOND J, GOODMAN J L, STRELETS V B, et al. FlyBase 2. 0: the next generation[J]. *Nucleic Acids Research*, 2019, 47(D1): D759-D765.
- [14] GHOSH A C, TATTIKOTA S G, LIU Y F, et al. *Drosophila* PDGF/VEGF signaling from muscles to hepatocyte-like cells protects against obesity[J]. *Elife*, 2020, 9: e56969.
- [15] ROBINOW S, WHITE K J D B. The locus *elav* of *Drosophila melanogaster* is expressed in neurons at all developmental stages[J]. *Developmental Biology*, 1988, 126(2): 294-303.
- [16] RAVENSCROFT T A, JANSSENS J, LEE P T, et al. *Drosophila* voltage-gated sodium channels are only expressed in active neurons and are localized to distal axonal initial segment-like domains[J]. *Journal of Neuroscience: the official Journal of the Society for Neuroscience*, 2020, 40(42): 7999-8024.
- [17] CHNG W-B A, SLEIMAN M S B, SCHÜPFER F, et al. Transforming growth factor beta/activin signaling functions as a sugar-sensing feedback loop to regulate digestive enzyme expression[J]. *Cell Reports*, 2014, 9(1): 336-348.
- [18] LI J Q, WANG J J, ZHANG P J, et al. Inferring predictive genetic models and regulatory elements by deep learning of cross-species single-cell gene expression landscapes[EB/OL]. (2022-04-11)[2023-11-11]. <https://www.researchsquare.com/article/rs-1544073/v1>.
- [19] ALLEN A M, SOKOLOWSKI M B. Expression of the foraging gene in adult *Drosophila melanogaster* [J]. *Journal of Neurogenetics*, 2021, 35(3): 192-212.
- [20] FOSTER J M, ROBERTS D B. The soluble α -mannosidases of *Drosophila melanogaster* [J]. *Insect Biochemistry and Molecular Biology*, 1997, 27(7): 657-661.
- [21] HUBER A, SCHULZ S, BENTROP J, et al. Molecular cloning of *Drosophila* Rh6 rhodopsin: the visual pigment of a subset of R8 photoreceptor cells[J]. *FEBS Letters*, 1997, 406(1/2): 6-10.
- [22] JOHARD H A, ENELL L E, GUSTAFSSON E, et al. Intrinsic neurons of *Drosophila* mushroom bodies express short neuropeptide F; relations to extrinsic neurons expressing different neurotransmitters[J]. *Journal of Comparative Neurology*, 2008, 507(4): 1479-1496.
- [23] MICHKI N S, LI Y, SANJASAZ K, et al. The molecular landscape of neural differentiation in the developing *Drosophila* brain revealed by targeted scRNA-seq and multi-informatic analysis[J]. *Cell Reports*, 2021, 35(4): 109039.
- [24] EVANS J M, ALLAN A K, DAVIES S A, et al. Sulphonylurea sensitivity and enriched expression implicate inward rectifier K^+

- channels in *Drosophila melanogaster* renal function[J]. Journal of Experimental Biology, 2005, 208(19): 3771-3783.
- [25] CHINTAPALLI V R, TERHSAZ S, WANG J, et al. Functional correlates of positional and gender-specific renal asymmetry in *Drosophila* [J]. PLoS One, 2012, 7(4): e32577.
- [26] CAVALIERE V, BERNARDI F, ROMANI P, et al. Building up the *Drosophila* eggshell: first of all the eggshell genes must be transcribed[J]. Developmental Dynamics, 2008, 237(8): 2061-2072.
- [27] FAKHOURI M, ELALAYLI M, SHERLING D, et al. Minor proteins and enzymes of the *Drosophila* eggshell matrix[J]. Developmental Biology, 2006, 293(1): 127-141.
- [28] GIGLIOTTI S, GRAZIANI F, de PONTI L, et al. Sex-, tissue-, and stage-specific expression of a vitelline membrane protein gene from region 32 of the second chromosome of *Drosophila melanogaster* [J]. Developmental Genetics, 1989, 10(1): 33-41.
- [29] CLAYCOMB J M, BENASUTTI M, BOSCO G, et al. Gene amplification as a developmental strategy: isolation of two developmental amplicons in *Drosophila* [J]. Developmental Cell, 2004, 6(1): 145-155.
- [30] ZHU L, FUKUNAGA R. RNA-binding protein Maca is crucial for gigantic male fertility factor gene expression, spermatogenesis, and male fertility, in *Drosophila* [J]. PLoS Genetics, 2021, 17(6): e1009655.
- [31] FINGERHUT J M, YAMASHITA Y M. mRNA localization mediates maturation of cytoplasmic cilia in *Drosophila* spermatogenesis [J]. Journal of Cell Biology, 2020, 219(9): e202003084.
- [32] ZHANG J Y, LUO J J, CHEN J Y, et al. The role of Y chromosome genes in male fertility in *Drosophila melanogaster* [J]. Genetics, 2020, 215(3): 623-633.
- [33] BARREAU C, BENSON E, GUDMANSDOTTIR E, et al. Post-meiotic transcription in *Drosophila* testes [J]. Development, 2008, 135(11): 1897-1902.
- [34] WITT E, BENJAMIN S, SVETEC N, et al. Testis single-cell RNA-seq reveals the dynamics of de novo gene transcription and germline mutational bias in *Drosophila* [J]. Elife, 2019, 8: e47138.
- [35] WILK R, WEIZMAN I, SHILO B, et al. Trachealess encodes a bHLH-PAS protein that is an inducer of tracheal cell fates in *Drosophila* [J]. Genes & Development, 1996, 10(1): 93-102.
- [36] JASPERS M H, PFLANZ R, RIEDEL D, et al. The fatty acyl-CoA reductase Waterproof mediates airway clearance in *Drosophila* [J]. Developmental Biology, 2014, 385(1): 23-31.
- [37] ARRESE E L, SOULAGES J L. Insect fat body: energy, metabolism, and regulation [J]. Annual Review of Entomology, 2010, 55: 207-225.
- [38] PARVY J-P, NAPAL L, RUBIN T, et al. *Drosophila melanogaster* Acetyl-CoA-carboxylase sustains a fatty acid-dependent remote signal to waterproof the respiratory system [J]. PLoS Genetics, 2012, 8(8): e1002925.

Molecular Biology and Bioinformatics

Identification of Tissue Marker Genes in *Drosophila melanogaster* Based on Gene Expression Profiling

LI Lei¹, DUO Hongrui¹, LI Bo¹, LI Yinghong², HAO Youjin¹

(1. College of Life Sciences, Chongqing Normal University, Chongqing 401331; 2. School of Bioinformatics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

Abstract: To solve the current problem of inconsistent and inaccurate results of tissue marker gene identification in *Drosophila melanogaster*, *Drosophila* tissue marker genes were identified by Seurat as well as calculation of differential expression folds using transcriptomics data of seven shared tissues from Fly Cell Atlas and FlyAtlas2 databases, respectively. After cross-validation across datasets, literature comparison and gene function searches, accurate tissue marker genes were finally identified. From the two datasets, 575 and 5 202 tissue marker genes were identified, respectively, and 366 and 2 671 tissue marker genes were obtained after cross-validation across the datasets, and a total of 55 core tissue marker genes and 23 candidate tissue marker genes were obtained through literature comparison and gene function search in the FlyBase database. An accurate set of *Drosophila* tissue marker genes was established, providing an important reference for future *Drosophila* tissue type differentiation and cell type marker gene identification.

Keywords: *Drosophila melanogaster*; marker gene; transcriptome; single-cell transcriptome

(责任编辑 方 兴)