

## 研究论文

# 肝脏糖脂代谢研究的趋势变化分析与展望

石磊<sup>1,2</sup>, 瞿华<sup>3</sup>, 郑宏庭<sup>3,\*</sup>, 王秀杰<sup>1,2,\*</sup>

<sup>1</sup>中国科学院遗传与发育生物学研究所分子系统生物学研究中心, 北京 100101; <sup>2</sup>中国科学院大学, 北京 100049; <sup>3</sup>陆军军医大学第二附属医院内分泌科, 重庆 400037

**摘要:** 糖脂代谢稳态平衡是多因素、多器官协调互作的结果, 是维持人类健康的基本要素之一。肝脏作为最主要的代谢器官, 在机体糖脂代谢稳态调控中发挥至关重要的作用。随着研究的不断深入, 肝脏糖脂代谢相关的论文数量日益增长, 给快速了解肝脏糖脂代谢领域的研究热点和发展趋势带来了挑战。为解决这一问题, 本研究开发了一种信息分析方法, 通过对肝脏糖脂代谢领域研究论文涉及的《医学主题词表》(Medical Subject Headings, MeSH)中的词条和相关高通量研究数据的收集与解析, 对肝脏糖脂代谢领域近30年来的研究现状、研究手段和研究热点变化等因素进行了系统分析。结果显示, 本世纪以来, 肝脏糖脂代谢研究相关的论文数量, 尤其是中国学者发表的论文数量快速提升, 论文的篇均作者数量和篇均作者单位数量也显著增加, 并且这一数量变化与论文的影响力之间存在一定的正相关。各种高通量方法逐渐成为肝脏糖脂代谢遗传学相关研究的主要手段, 过氧化物酶体增殖物激活受体(peroxisome proliferator-activated receptors, PPARs)、固醇调节元件结合蛋白(sterol regulatory element binding proteins, SREBPs)、核因子E2相关因子2 (NF-E2-related factor 2, Nrf2)等转录因子成为肝脏糖脂代谢研究的新兴热点。这些结果系统展示了肝脏糖脂代谢研究的关注重点和变化趋势, 本文所开发的方法也可被应用于对其他领域进行类似的研究热点分析。

**关键词:** 肝脏; 糖脂代谢; 文献分析; MeSH词条; GEO数据

**中图分类号:** R333.4

## Developmental trend analysis and perspectives on researches related to hepatic glucose and lipid metabolism

SHI Lei<sup>1,2</sup>, QU Hua<sup>3</sup>, ZHENG Hong-Ting<sup>3,\*</sup>, WANG Xiu-Jie<sup>1,2,\*</sup>

<sup>1</sup>Center for Molecular Systems Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, China; <sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China; <sup>3</sup>Department of Endocrinology, The Second Affiliated Hospital of Army Medical University, Chongqing 400037, China

**Abstract:** The balance of glucose and lipid metabolism is a coordinated result of multiple factors and organs, and is one of the fundamental requirements for the maintenance of human health. As the most important organ for human metabolism, liver plays a key role in regulating glucose and lipid metabolism. With the advances of researches, the number of publications related to hepatic glucose and lipid metabolism has increased rapidly, which posed a challenge for grasping the hot research topics and developmental trends of hepatic glucose and lipid metabolism in a short time. To solve such problem, we developed an information analysis method, which systematically analyzes the research status, research techniques, and hot research topics of the hepatic glucose and lipid metabolism research field through Medical Subject Headings (MeSH) of related papers and high-throughput experimental data. The results showed that the number of publications related to hepatic glucose and lipid metabolism, especially publications by Chinese scholars, has increased dramatically in this century, along with the remarkable increment of the numbers of authors and affiliations per paper. Such

Received 2021-01-04 Accepted 2021-07-12

This work was supported by grants from the Chinese Academy of Sciences (No. QYZDJ-SSW-SMC015) and the National Science Fund for Distinguished Young Scholars (No. 81925007).

\*Corresponding authors. WANG Xiu-Jie: Tel: +86-10-64806590, E-mail: xjwang@genetics.ac.cn; ZHENG Hong-Ting: Tel: +86-23-68755709, E-mail: fnf7703@hotmail.com

increment is in part positively correlated with the impact of publications. Nowadays, various types of high-throughput experimental techniques have become the main research methods for genetic studies of hepatic glucose and lipid metabolism. Transcription factors, such as peroxisome proliferator-activated receptors (PPARs), sterol regulatory element binding proteins (SREBPs), and NF-E2-related factor 2 (Nrf2), have become the new research hotspots. These results systematically showed the current focuses and developmental trends of hepatic glucose and lipid metabolism research, and the data analysis method developed in this work can also be applied to other research fields.

**Key words:** liver; glucose and lipid metabolism; publication analysis; MeSH term; GEO data

糖类和脂类是生物体维持生命活动所需能量的主要来源,其调控网络的稳态失衡也是多种代谢疾病发生和发展的重要诱导因素。糖类物质按其分子组成,可以分为单糖(如葡萄糖、果糖)、二糖(如麦芽糖、蔗糖、乳糖)和多糖(如淀粉、糖原、纤维素)等<sup>[1]</sup>。食物中的糖类物质(以淀粉为主)被人体消化吸收后,以葡萄糖的形式随血液运输,到达各组织器官进行代谢。临床常用的检测指标血糖,即血浆中的葡萄糖浓度,其相对恒定对于维持组织器官的正常功能和人体健康至关重要<sup>[2]</sup>。脂类是脂肪、油、类脂等不溶于水的生物大分子的总称,主要包括脂肪、磷脂、鞘脂、胆固醇等,其中脂肪的主要成分是甘油三酯<sup>[3]</sup>。脂类、糖类与蛋白质共同组成人体所需的三大基本营养素,是细胞供能和储能的主要物质之一<sup>[4]</sup>。脂类也是组成细胞膜的必要成分,并且在调节细胞代谢、炎症和免疫反应等基本生理活动方面发挥重要作用<sup>[5,6]</sup>。

在人体涉及糖脂代谢的各类器官中,肝脏扮演着重要的角色。人体细胞主要依赖肝脏来维持胞内葡萄糖和脂类物质的稳态。在餐后血糖浓度升高时,葡萄糖随血液运输至肝细胞后,首先被磷酸化为 6-磷酸葡萄糖<sup>[7,8]</sup>。6-磷酸葡萄糖不能再穿透细胞膜回到血液,而是作为肝脏碳水化合物代谢的源头物质,开启糖原合成、糖酵解、戊糖磷酸化和脂肪合成等代谢途径<sup>[9,10]</sup>。在饥饿状态下,肝细胞也可以通过糖异生过程将乳酸、生糖氨基酸等非糖物质转变为 6-磷酸葡萄糖<sup>[11]</sup>。肝脏是人体内糖类物质转变为脂肪的两大主要场所之一<sup>[12]</sup>。肝脏也是脂类物质消化、吸收、合成与运输的调控中枢。肝脏通过分泌胆汁而促进脂类物质的消化吸收,是甘油三酯、脂肪酸、胆固醇、酮体等脂类物质合成的最主要场所<sup>[13]</sup>。在正常情况下,这些脂类代谢产物在肝脏合成后,被血液运输到肝外组织储存或加以利用<sup>[14]</sup>。

肝脏糖脂代谢失衡可导致多种代谢性疾病,最常见的包括糖尿病、肥胖和非酒精性脂肪肝。随着

这些代谢性疾病患病率在全球的快速增长,深入探讨肝脏糖脂代谢调控规律及其在机体糖脂代谢稳态中的功能显得愈发重要。在遗传和(或)环境因素影响下,机体胰岛素分泌减少或信号转导障碍(胰岛素抵抗)会导致葡萄糖来源的供能减少和脂质的分解与释放增强,后者刺激肝脏葡萄糖异生、促进肝糖释放,并参与糖尿病及其并发症的发生和发展<sup>[15,16]</sup>。此外,基因和(或)营养过剩等因素也会促进肝脏合成更多的甘油三酯,多余的葡萄糖也可以通过脂肪生成反应转变为甘油三酯<sup>[17,18]</sup>。这些过量合成的甘油三酯在肝脏中累积,则可能导致非酒精性脂肪肝<sup>[12,19]</sup>。

鉴于糖脂代谢对于人体健康的重要性和肝脏在调控糖脂代谢稳态中的核心功能,解析肝脏糖脂代谢调控的分子机制已成为科研人员持续关注研究方向。为总结近年来有关肝脏糖脂代谢研究的热点变化和发展趋势,我们对肝脏糖脂代谢研究的相关文献和高通量研究数据进行了系统收集,并基于文献的作者组成、《医学主题词表》(Medical Subject Headings, MeSH)的词条分类、高通量研究数据的类别与数量等信息,对肝脏糖脂代谢研究领域近 30 年来的研究模式、研究方法和研究热点的变化趋势进行了分析。

## 1 材料与方 法

**1.1 数据收集** 本研究所使用的文献数据全部来源于 PubMed 文献数据库 (<https://pubmed.ncbi.nlm.nih.gov/>)。文献收集的标准为:在文献的标题或摘要中存在符合如下检索式【("liver" OR "hepatic") AND ("glycolipid" OR "glucose and lipid" OR "glucose-lipid" OR "glucose" OR "lipid" OR "fat")】的词条的文献。应用上述标准,我们基于 Python 程序包 Scrapy (<https://scrapy.org/>) 构建了网页抓取程序,从 PubMed 获取了 147 252 篇肝脏糖脂代谢文献数据,这些数据包括但不限于文章标题、期刊、引用

信息、作者和 doi 编号 (截至 2020 年 11 月 6 日)。为了保证下游分析结果的质量,我们首先删除了在标题、期刊、引用信息、作者或 doi 编号等属性出现缺失的文献,并且剔除了发表在 1930 年之前的文献(年发表文献量持续低于 3 篇)。经过上述数据清理后,共有 121 326 篇文献符合标准,进入下游分析环节。

为了对文献的研究内容进行分类,我们同样基于自建的网页抓取程序,对每篇文献对应的 MeSH 词条和整个 MeSH 词集进行了系统收集。MeSH 词集是由美国国家医学图书馆(The United States National Library of Medicine)编制的主题词表,共包含 30 000 多个医学主题词,MeSH 词条之间通过有向无环图的结构展示其相互关系。我们对上述文献中的 MeSH 词条进行了统计,并剔除了词频数低于 20 的 MeSH 词条和无义词条。最终,本研究所选用的 121 326 篇文献共对应 3 629 个 MeSH 词条。通过这些 MeSH 词条,我们可以从不同层次、不同角度对文献的研究内容及其所属类别进行分析。

本研究所使用的肝脏糖脂代谢相关高通量研究数据来自于 NCBI 的 Gene Expression Omnibus (GEO) 数据库(<https://www.ncbi.nlm.nih.gov/geo/>)。高通量研究数据收集的标准为:在 GEO 数据库中的数据说明满足如下检索式【"liver"[MeSH Terms] AND ("glucose"[MeSH Terms] OR "lipids"[MeSH Terms))]的 Series 类型数据,这些数据包含了高通量研究数据的 GEO ID、标题、研究对象等信息。利用相似的网页抓取程序,我们共收集到 5 176 组来自 GEO 数据库的高通量研究数据(截至 2020 年 11 月 6 日)。

**1.2 肝脏糖脂代谢研究论文的基本情况分析** 本研究根据由科睿唯安(Clarivate Analytics)发布的 2019 年全球 SCI 期刊影响因子(impact factor, IF)来评价期刊的影响力,将发表在  $IF \geq 10$  的期刊上的文献定义为高影响力论文。本研究所使用的 121 326 篇文献中,共有 90 011 篇文献发表的期刊有相应的 SCI 影响因子(期刊名称及其影响因子见附表 1,见补充材料:<https://www.actaps.com.cn/supplement/APS-2021-01-0004table1.pdf>)。每篇文献的作者数目和研究机构数目通过文献的作者和研究机构信息获取。除此之外,文献的研究机构信息还被用于判断文献的产出是否有中国的研究机构参与。

### 1.3 肝脏糖脂代谢研究方法的变化及高通量研究技术在肝脏糖脂代谢研究中的应用分析

我们开发了 Python 程序,对 MeSH 模块和 MeSH 词条之间的关系进行提取和判定。通过上述程序,我们获取了自 1990 年以来具有 MeSH 词条标注的文献共 86 650 篇。肝脏糖脂代谢研究方法的变化情况通过统计上述文献集对应的 MeSH 词集中隶属于“Genetic Techniques”的主要词条的年出现频率而获得。主要词条指隶属于某一词条(如“Genetic Techniques”)并且在本研究使用文献中出现的最高一级子词条。近年来兴起的高通量研究技术在肝脏糖脂代谢研究中的应用情况通过对上述收集的 GEO 数据按照其提交时间和 GEO 数据库中注释的数据类型(Type)进行统计而获得。

**1.4 肝脏糖脂代谢研究热点变化分析** 肝脏糖脂代谢研究热点分析包括总体研究热点分析和研究热点趋势变化分析两部分。肝脏糖脂代谢领域的总体研究热点分析通过汇总上述 86 650 篇文献所涉及的 MeSH 词条的出现频数及其对应的 MeSH 类别的占比情况而获得。不同 MeSH 类别下,高频 MeSH 词条的分布情况通过词频图展示。词频图通过 Python 第三方包 wordcloud (<https://github.com/amueller/wordcloud>) 绘制得到,其中 max\_words 参数的设置为 30。研究热点趋势变化分析首先根据总体研究热点的分析结果,提取感兴趣的 MeSH 类别,再分析该类别下不同层级的主要 MeSH 词条在文献中出现频率的逐年变化趋势。该部分分析使用的文献为发表于 1990 年之后的 86 650 篇文献,主要 MeSH 词条的选取标准与 1.3 部分的标准一致。

**1.5 关联网络构建与统计检验** 与过氧化物酶体增殖物激活受体(peroxisome-proliferator activated receptors, PPARs)、固醇调节元件结合蛋白(sterol regulatory element binding proteins, SREBPs)或核因子 E2 相关因子 2 (NF-E2-related factor 2, Nrf2) 相关的 MeSH 词条的筛选通过分析文献集中与三者具有显著相关性(皮尔逊相关性分析,  $r > 0.85$ )的 MeSH 词条,并去除“Liver”、“Cells, Cultured”等低信息量词条而完成。由于三者均为转录因子,定义其属性的“Transcription Factor”和“DNA-Binding Protein”词条也被去除(详细的相关 MeSH 词条见附表 2,见补充材料:<https://www.actaps.com.cn/supplement/APS-2021-01-0004table2.pdf>)。PPARs 和 SREBPs 的重叠关联词中综合相关性最高的 Top 5 MeSH 词条和

与 Nrf2 关联性最高的 Top 5 MeSH 词条被分别用于词频年度变化分析。综合相关性定义如下：

$$score_{MeSH} = \frac{\sum_{i=1}^2 index_{MeSH,i} / length_i}{2}$$

其中 MeSH 为特定的 MeSH 词条, *index* 为其在不同组别 (PPARs 和 SREBPs) 的排序索引, *length* 为不同组别的长度。

Nrf2 或其他肝脏糖脂代谢研究的主要目标蛋白与不同类别 MeSH 词条的关联网络通过目标蛋白和与其在文献中高频共现的 MeSH 词条构建。与目标蛋白共出现的高频 MeSH 词条通过卡方检验筛选, 根据不同蛋白的出现频次, 以错误发现率 (False Discovery Rate, FDR)  $< 1 \times 10^{-5}$  或  $1 \times 10^{-2}$  作为 MeSH 词条在网络中展示的入选标准。网络中 MeSH 词条间的关系依据 MeSH 词库中词条间的隶属层级关系构建。为避免网络过于混杂, 目标蛋白仅与不同 MeSH 类别的最底层词条间建立连线。

## 2 结果

### 2.1 肝脏糖脂代谢研究论文的总体情况

为了解肝脏糖脂代谢研究领域的总体情况, 我们首先对本研究使用的 121 326 篇文献进行了系统分析。结果表明, 肝脏糖脂代谢研究相关文献的年发表数量自上世纪 60 年代起呈现稳步增加的趋势, 并且文献年发表数量在 21 世纪后的提升速度进一步加快 (图 1A)。值得关注的是, 2004 年和 2010 年发表的肝脏糖脂代谢相关研究文献的数量与前一年相比有较大幅度的跃升, 并且在之后的年份也保持更大的增长比例, 提示这两个时间点可能出现了新的研究热点或技术理论突破 (图 1A)。相应地, 这两个时间点发表的高影响力文献数量也较之前的时间点有较大增长, 并且在之后的一段时间内保持小幅增加 (图 1A)。其中, 2004 年发表在 IF  $\geq 10$  的期刊上的论文为 159 篇, 2010 年发表在 IF  $\geq 10$  的期刊上的论文为 204 篇 (图 1A)。值得一提的是, 我国肝脏糖脂代谢研究虽然起步较晚 (约起始于 20 世纪 90 年代), 但发展迅速。中国学者发表的肝脏糖脂代谢研究论文在整个领域论文中的占比已由 2000 年的 1.1% 攀升至 2019 年的 29.2% (图 1A)。在 121 326 篇文献中, 绝大多数为基础研究相关的论文 (General Research), 其次为综述类文章 (Review), 临床试验类论文 (Randomized Controlled Trial 和 Clinical Trial) 与病例报道类论文 (Case Reports) 最少 (图 1B)。

我们对历年来发表在 IF  $\geq 10$  的期刊上的文献分布情况进行了统计。其中, 发文数占比最高的杂志为 *Hepatology*, 其次为 *Journal of Clinical Investigation* 和 *Journal of Hepatology*。顶级综合或医学期刊 *Cell*、*Science*、*Lancet*、*New England Journal of Medicine* 的发文占比均在 1%~3%, *Nature* 的发文占比稍高, 为 4.1% (图 1C)。

进一步地, 我们对上述所有 121 326 篇文献的作者数量和单位组成进行了分析, 以期揭示研究团队组成模式的变化情况。结果显示, 自 1930 年以来, 肝脏糖脂代谢相关研究论文的作者数量呈现稳步上升的趋势, 这种篇均作者数量增多的趋势在本世纪尤其明显 (图 2A)。与篇均作者数量的变化趋势不完全一致, 篇均作者单位数量长期以单一单位为主, 自 2010 年后才有较明显的整体上升, 并且在 2015 年之后有较大的跃升 (图 2B)。多作者与多单位协作的一个成效是产出的高影响力期刊 (IF  $\geq 10$ ) 论文的占比显著提升, 这种趋势在篇均作者数大于 8 或篇均作者单位数大于 5 时呈快速线性递增 (图 2C、D)。

### 2.2 肝脏糖脂代谢研究方法变化分析

新的研究技术与方法在催生新的科学发现方面经常发挥至关重要的作用。通过对文献对应的 “Genetic Techniques” 类别 MeSH 词条的出现频率进行分析, 我们发现在 2000 年前常用的 “Molecular Cloning” 和 “Chromosome Mapping” 等方法, 在 2000 年后随着高通量测序等新研究方法的出现, 所占的比例越来越少 (图 3A)。基于高通量测序的 “Gene Expression Profiling” 方法在 2000 年开始出现, 并快速发展为肝脏糖脂代谢相关遗传学研究最常用的方法 (图 3A)。“Gene Targeting” 词条的出现频率在 2010 年及之后有较大幅度的增加, 这可能与 CRISPR/Cas9 等基因编辑技术的开发和普及有关 (图 3A)。

为了进一步分析肝脏糖脂代谢研究中高通量研究技术的应用情况, 我们提取了收录在 NCBI GEO 数据库中肝脏糖脂代谢相关的高通量研究数据信息。自 21 世纪以来, 每年提交至 GEO 数据库中的肝脏糖脂代谢相关高通量数据呈快速增长趋势, 至 2014 年起达到平台期, 每年新增数据约 400 套 (图 3B)。2020 年提交的数据量有较明显下降, 一方面由于数据统计时间仅截至 2020 年 11 月 6 日, 另一方面也可能与当年全球新型冠状病毒肺炎疫情对实验活动的影响有关 (图 3B)。在提交至 GEO 数据

库的肝脏糖脂代谢相关的高通量数据中，基于基因芯片的表达谱检测数据“Expression profiling by microarray”为当前最主要的数据类型，占比为57.44%，其次为依赖于高通量测序的“Expression profiling by high throughput sequencing”，占比为26.80%，而与转录因子结合或染色质高级结构研究相关的“Genome binding/occupancy profiling”和与非编码

RNA 研究相关的“Non-coding RNA profiling”占比还相对较少(图3C)。

### 2.3 肝脏糖脂代谢研究内容的变化趋势分析

为了探究肝脏糖脂代谢领域的研究重点，我们首先对文献涉及的MeSH词条进行了系统分析。其中，“Chemicals and Drugs Category”的Mesh词条出现次数最多，占比47.9%，之后依次为“Anatomy

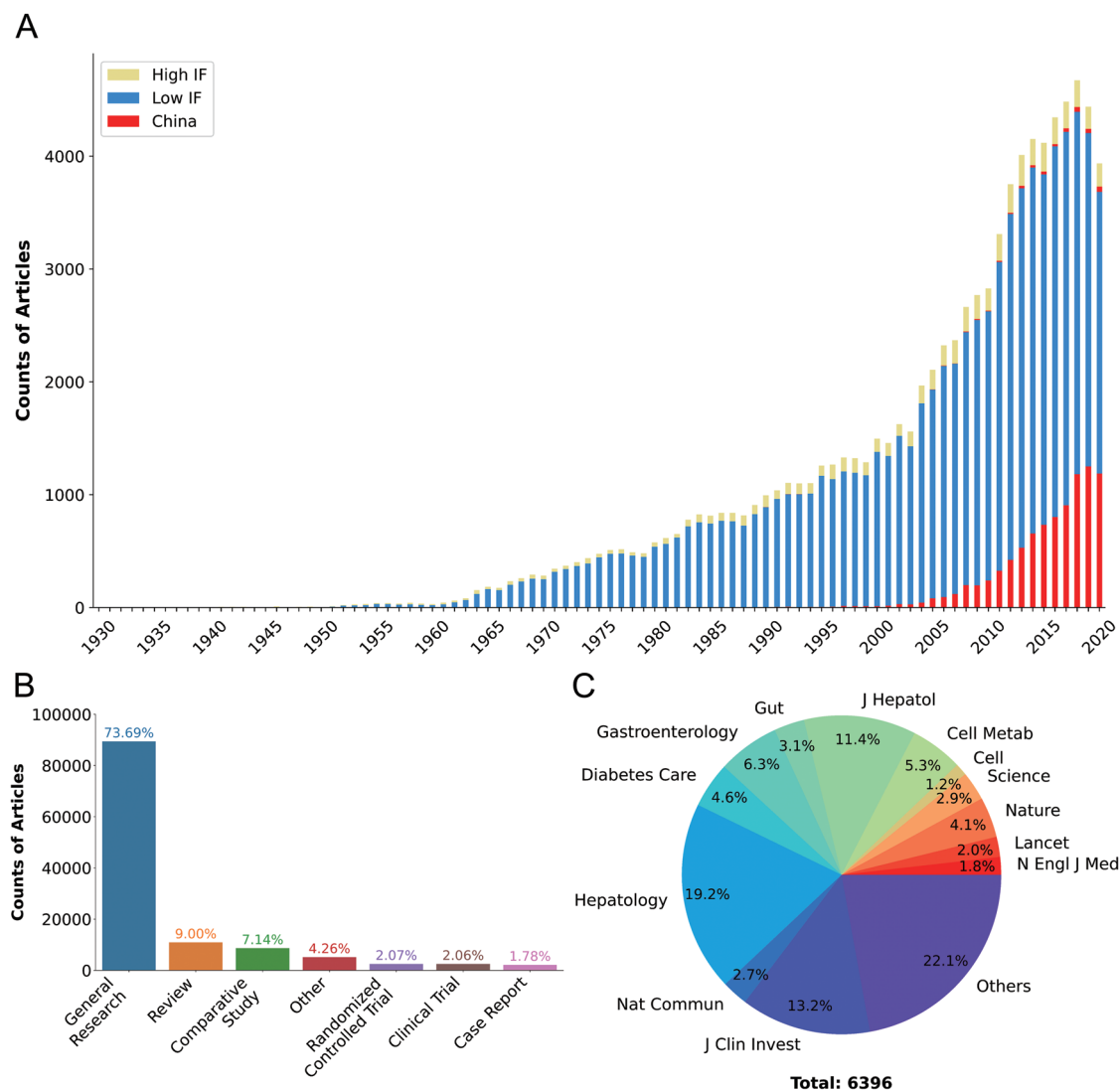


图 1. 肝脏糖脂代谢领域文献总体情况分析。A: PubMed数据库中肝脏糖脂代谢领域年文献数量统计。以文献发表期刊影响因子(impact factor, IF)为依据, 将文献分为高影响因子文献(High IF,  $IF \geq 10$ )和低影响因子文献(Low IF,  $IF < 10$ )两类。图中红色柱状图部分表示由中国机构参与发表的High IF或Low IF文献。B: 领域文献类型分布。C: High IF文献分布统计。期刊按照其IF由大到小的顺序由红色至紫色表示。

Fig. 1. Systematic analysis of articles related to hepatic glucose and lipid metabolism. A: Annual counts of PubMed articles related to hepatic glucose and lipid metabolism. High IF and low IF represent articles published on journals with impact factor (IF)  $\geq 10$  and  $< 10$ , respectively. Articles contributed by Chinese research institutions are shown as red bars in both the high IF and low IF categories. B: Bar plot of article type distributions. C: Statistics of articles published on high IF journals. Journals are ordered by decreasing IF factors with corresponding colors (from red to purple).

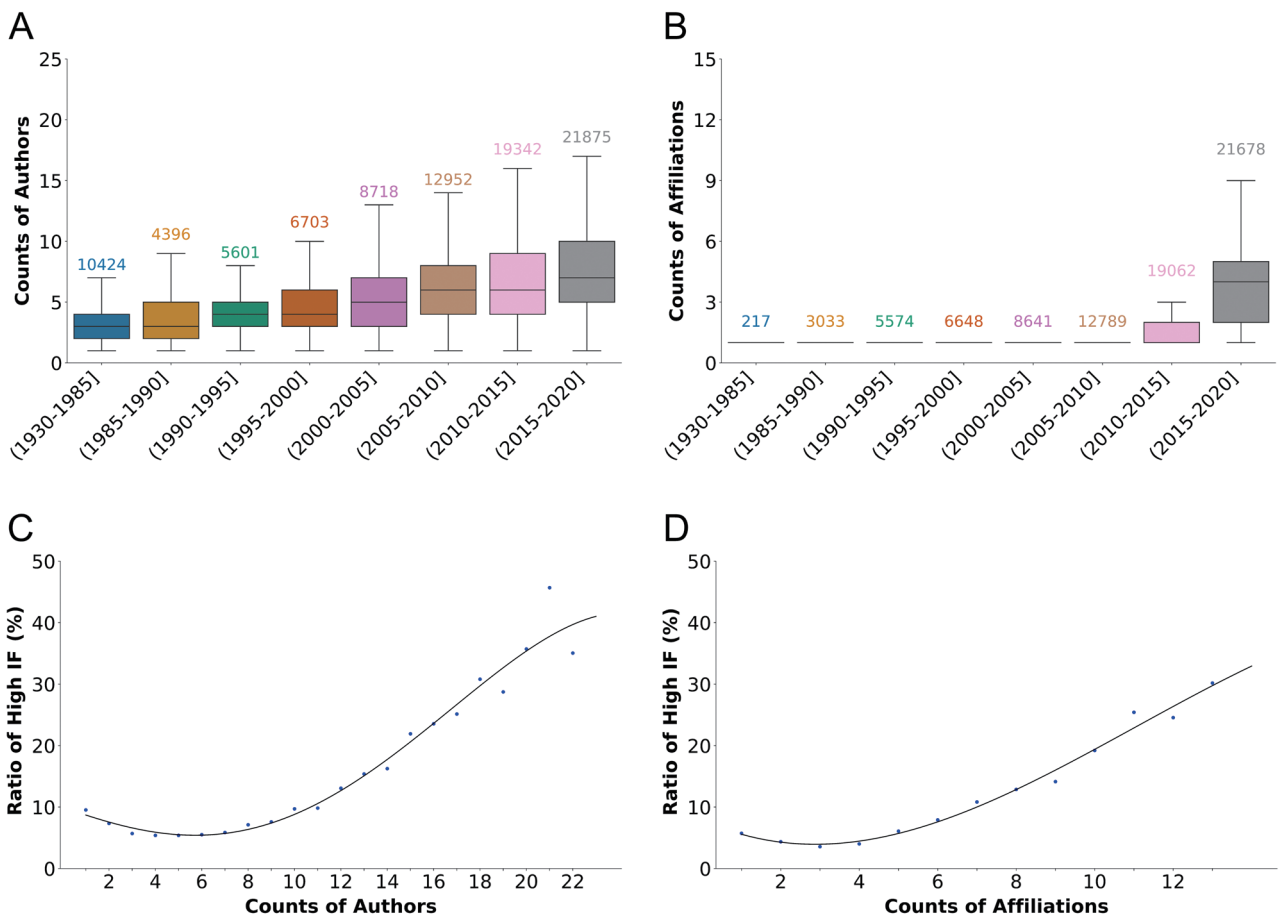


图 2. 肝脏糖脂代谢领域文献篇均作者数量与篇均作者单位数量变化分析。A~B: 文献篇均作者数量(A)或篇均作者单位数量(B)随时间变化情况。箱图上的数字表示在相应时间段内分析的含有作者(A)或作者单位(B)信息的文献总数。C~D: 文献篇均作者数量(C)或篇均作者单位数量(D)与文献发表期刊IF间关系的回归分析。Y轴表示相应文献集中High IF文献的比例。

Fig. 2. Analysis of the changes on author number or affiliation number per article related to hepatic glucose and lipid metabolism. A~B: Box plot showing the distribution of author number (A) or affiliation number (B) per article within different time intervals. Numbers at the top of variation bars represent the total number of articles with author (A) or affiliation (B) information published within the corresponding time interval. C~D: Correlations between the number of authors (C) or affiliations (D) per article and the impact of publications. Y-axis represents the proportion of high IF papers among all papers published by the number of authors (C) or affiliations (D) indicated by the X-axis.

Category”、“Analytical, Diagnostic and Therapeutic Techniques and Equipment Category”、“Diseases Category”等(图4A)。我们进一步对“Chemicals and Drugs Category”类别的MeSH子词条出现频次进行了分析。结果显示,约1/4(24.7%)隶属于“Chemicals and Drugs Category”的子词条属于“Amino Acids, Peptides and Proteins”类别(图4B)。在“Chemicals and Drugs Category”类别出现最频繁的前30个MeSH词条中,除了糖脂代谢直接涉及的血糖(Blood Glucose)、葡萄糖(Glucose)、胰岛素(Insulin)、胆固醇(Cholesterol)等,抗氧化相关的词条也占较大

比例,包括“Antioxidants”、“Catalase”、“Superoxide Dismutase”等(图4D)。类似地,“Diseases Category”类别MeSH词条的具体分析显示,营养与代谢疾病(Nutritional and Metabolic Diseases)和消化系统疾病(Digestive System Diseases)是被关注最多的疾病(图4C)。隶属于“Diseases Category”类别的子词条中,出现频率最高的包括肥胖(Obesity)、胰岛素抵抗(Insulin Resistance)、脂肪肝(Fatty Liver)、2型糖尿病(Diabetes Mellitus, Type 2)、肝脏肿瘤(Liver Neoplasms)等(图4E)。

为了发现肝脏糖脂代谢领域重点关注的蛋白

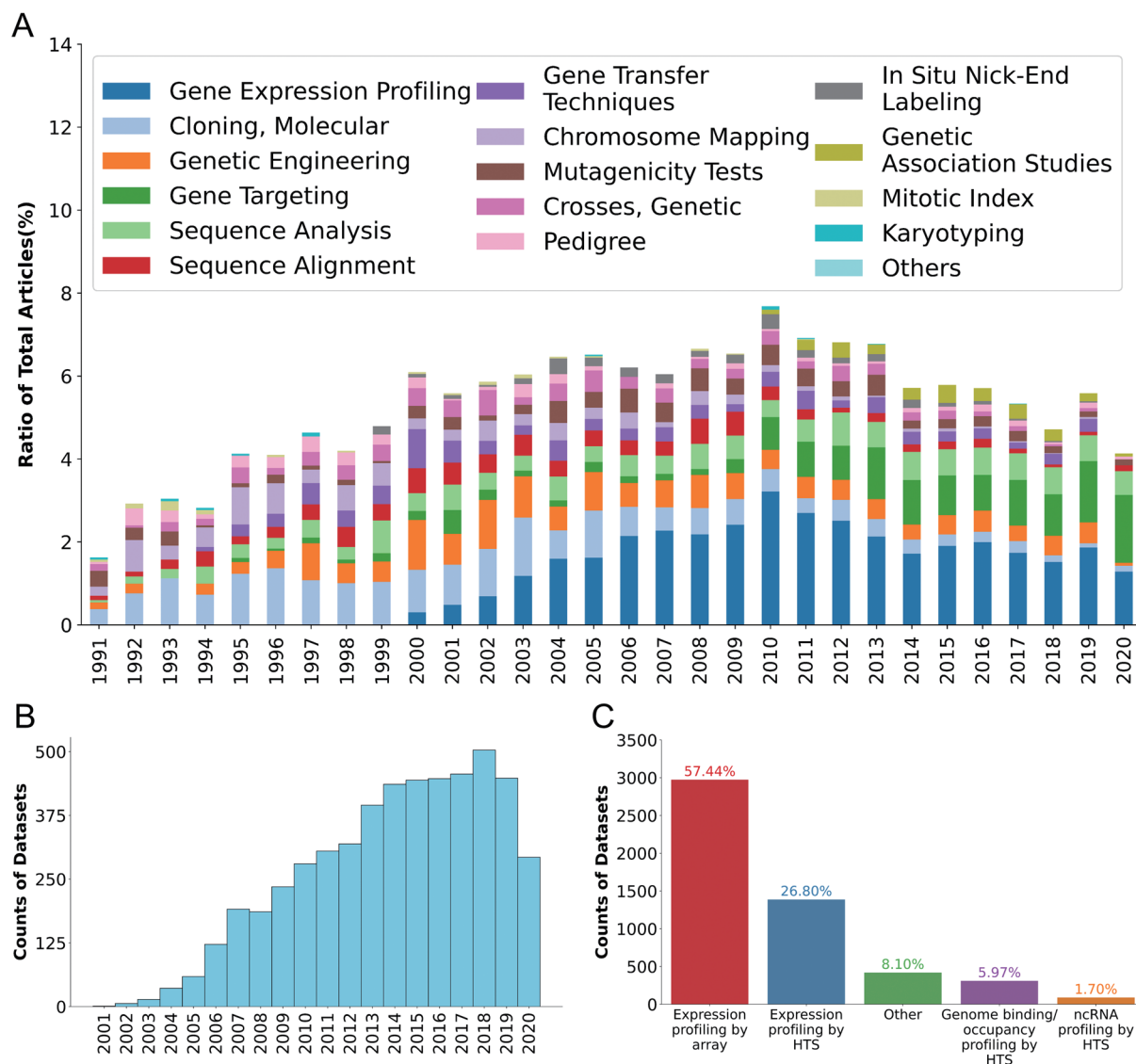


图 3. 肝脏糖脂代谢常用研究方法的变化分析。A: 包含隶属于“Genetic Techniques”的MeSH词条的文献分布统计。只展示了词频比例 $\geq 0.03$ 的MeSH 词条。B: GEO数据库中肝脏糖脂代谢相关高通量实验数据集的年提交数目统计。C: 高通量数据集的类型分布。只展示了Top 5的高通量数据集类型，其中“Other”类型高通量数据集由GEO数据定义。

Fig. 3. Analysis of common research methods used in hepatic glucose and lipid metabolism studies. A: Stacked bar plot showing the proportion of research articles using different experimental techniques under the “Genetic Techniques” MeSH category. MeSH terms with frequencies no less than 0.03 are included. B: Annual counts of hepatic glucose and lipid metabolism study-related high-throughput experimental datasets in the GEO database. C: Classification of high-throughput datasets by experimental types. Shown are the top 5 most abundant experimental types among the GEO datasets. HTS means high throughput sequencing and ncRNA means non-coding RNA. “Other” is defined by the GEO database representing datasets with uncharacterized experimental types.

质，我们对“Amino Acids, Peptides, and Proteins”分类下的MeSH词条自1990年以来在文献中的出现频率进行了详细分析。结果显示，对上世纪90年代早期研究较多的“Protein Precursors”和“Blood Proteins”的关注度在1997年之后逐年下降，其中对“Protein Precursors”的关注度下降最明显(图

5A)。几乎在同时期，涉及“Transcription Factors”和“DNA-Binding Proteins”的文献数目逐年增加(图5A)。2000年之后，涉及“Receptors, Cytoplasmic and Nuclear”的研究也有明显增加(图5A)。此外，“Membrane Proteins”、“Carrier Proteins”和“Glycoproteins”的词频占比相对比较稳定，提示其一直为

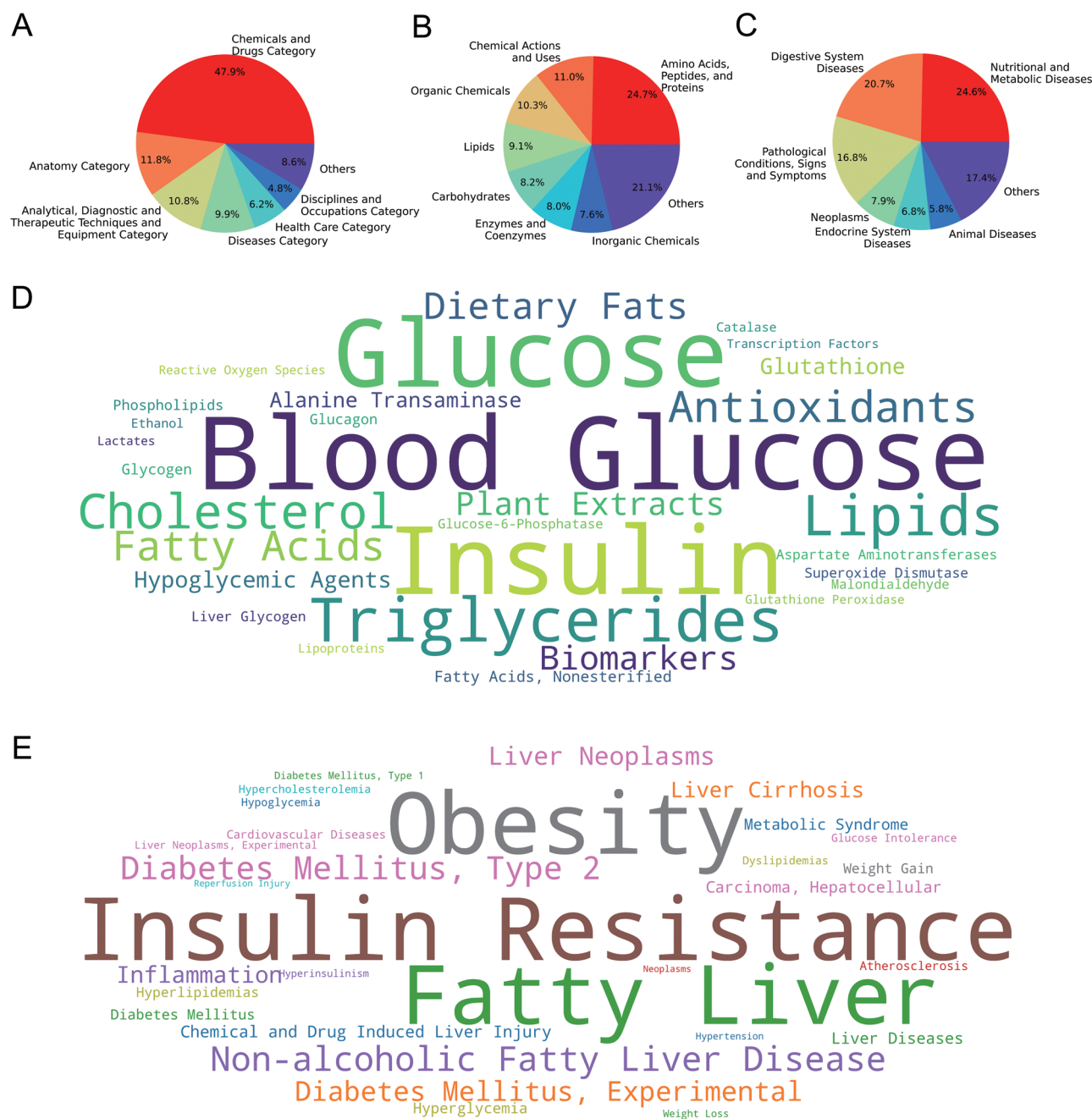


图 4. 肝脏糖脂代谢相关文献涉及的主要MeSH词条分类。A~C: MeSH词条“*All MeSH Categories*”(A)及其子代词条“*Chemicals and Drugs Category*”(B)和“*Diseases Category*”(C)的词频扇形分布图。图A~C中仅展示词频不小于4%的MeSH词条。D~E: “*Chemicals and Drugs Category*”(D)和“*Diseases Category*”(E) MeSH词条的子代词频排名前30的词云图。  
Fig. 4. Classification of key MeSH terms from articles related to hepatic glucose and lipid metabolism studies. A–C: Frequency pie charts of the first-level child MeSH terms under the “*All MeSH Categories*” (A), “*Chemicals and Drugs Category*” (B), and “*Diseases Category*” (C). Only MeSH terms with appearance frequency no less than 4% are shown in A–C. D–E: Word clouds of the top 30 MeSH terms under the “*Chemicals and Drugs Category*” (D) and “*Diseases Category*” (E).

肝脏糖脂代谢领域的重要研究对象(图5A)。

我们随后对“*Transcription Factors*”和“*Receptors, Cytoplasmic and Nuclear*”类别的子代MeSH词条进行了进一步分析。结果显示,“*Transcription Factors*”

相关研究出现频次最高的子词条中,除“*Basic-Leucine Zipper Transcription Factors*”和“*Basic Helix-Loop-Helix Transcription Factors*”外,其余高频词条多数与受体相关,并且与“*Receptors, Cytoplasmic and*



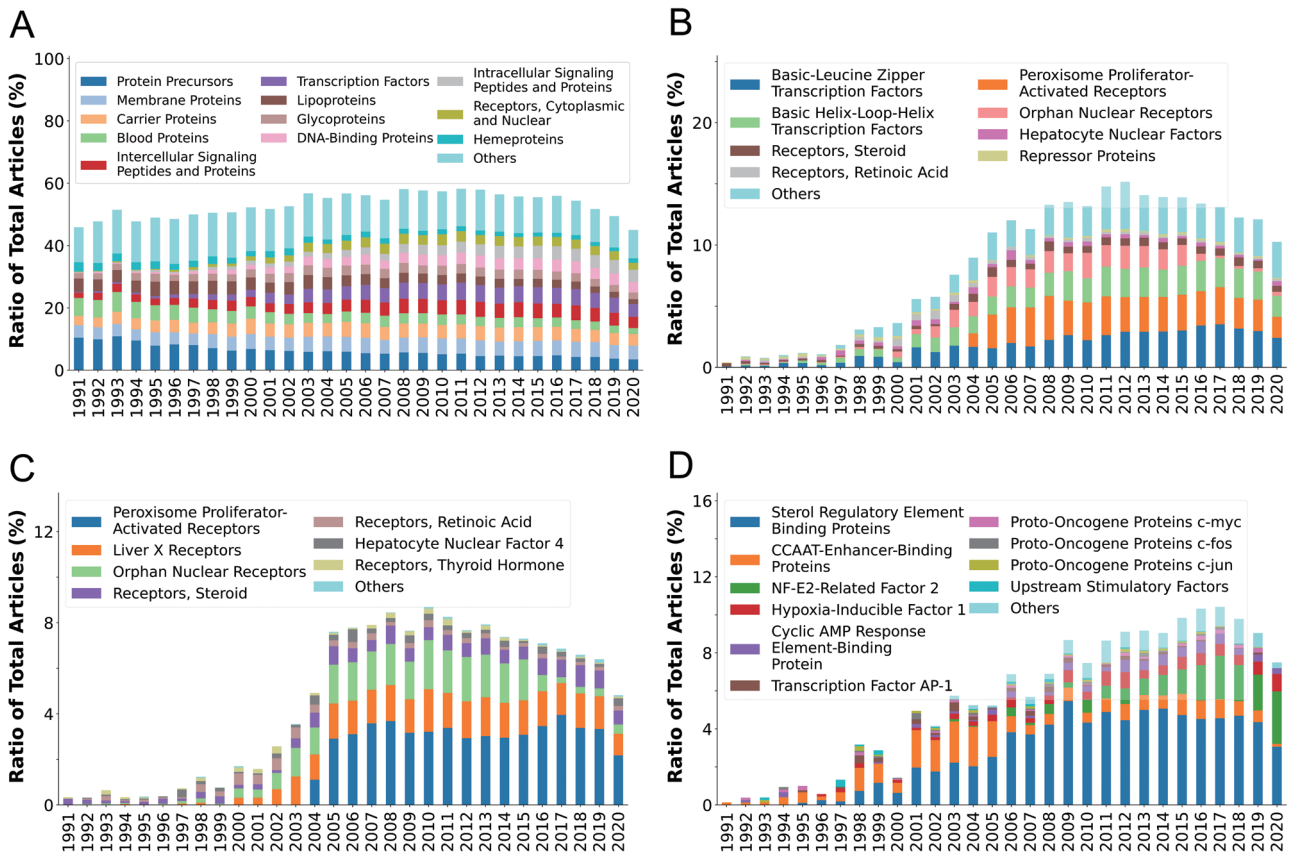


图 5. 肝脏糖脂代谢研究关注的热点蛋白分析。A: “Proteins” MeSH 词条的第一层级子代词条的相关文献年度分布分析。B, C: “Transcription Factors” (B) 或 “Receptors, Cytoplasmic and Nuclear” (D) MeSH 词条下的主要词条的相关文献年度分布分析。D: “Basic-Leucine Zipper Transcription Factors” 或 “Basic Helix-Loop-Helix Transcription Factors” MeSH 词条的主要子代词条的相关文献分布分析。只展示了词频比例 $\geq 0.05$  (A) 或 0.1 (B, D) 的 MeSH 词条。

Fig. 5. Analysis of major proteins focused by hepatic glucose and lipid metabolism researches. A: Stacked bar plot representing the annual frequency distribution of articles containing the first-level child MeSH terms under the “Proteins” category. MeSH terms with frequencies no less than 0.05 are included. B–C: Stacked bar plots representing the annual frequency distribution of articles containing the major child MeSH terms under the “Transcription Factors” (B) or “Receptors, Cytoplasmic and Nuclear” (C) category. D: Stacked bar plots representing the annual frequency distribution of articles containing the major child MeSH terms under the “Basic-Leucine Zipper Transcription Factors” or “Basic Helix-Loop-Helix Transcription Factors” category. In B–D, MeSH terms with frequencies no less than 0.1 are included.

Nuclear” 类别的子代高频词条重叠 (图 5B, 5C)。其中, 受关注最多的受体为 “Peroxisome Proliferator-Activated Receptors (PPARs)”。关于 PPARs 的肝脏糖脂代谢研究论文自 2004 年才开始出现, 在 2005 年相关文献数目就翻倍, 并在之后的年份持续保持高占比, 显示出 PPARs 在肝脏糖脂代谢调控中的重要作用 (图 5B, 5C)。我们还对 “Basic-Leucine Zipper Transcription Factors” 和 “Basic Helix-Loop-Helix Transcription Factors” 词条的子词频进行了分析。结果显示, “Sterol Regulatory Element Binding Proteins (SREBPs)” 自 2006 年起成为最主要的研究对象 (图

5D)。关于 “CCAAT-Enhancer-Binding Proteins (C/EBPs)” 的研究在 2005 年之前较多, 之后被与 C/EBPs 同为 PI3K 下游的 “NF-E2-Related Factor 2 (Nrf2)” 所取代, 针对 Nrf2 的研究自 2010 年之后显著增多 (图 5D)。Nrf2 在抗氧化方面具有重要作用, 已被证明可改善胰岛素抵抗和非酒精性脂肪肝等, 并可促进糖尿病溃疡创面愈合 [20, 21]。

进一步地, 我们分析了 2010 年之后与 PPARs、SREBPs 或 Nrf2 一同出现的 MeSH 主要词条, 以期揭示肝脏糖脂代谢热点领域相关研究的侧重点。统计结果显示, PPARs、SREBPs 和 Nrf2 分别拥有 11、

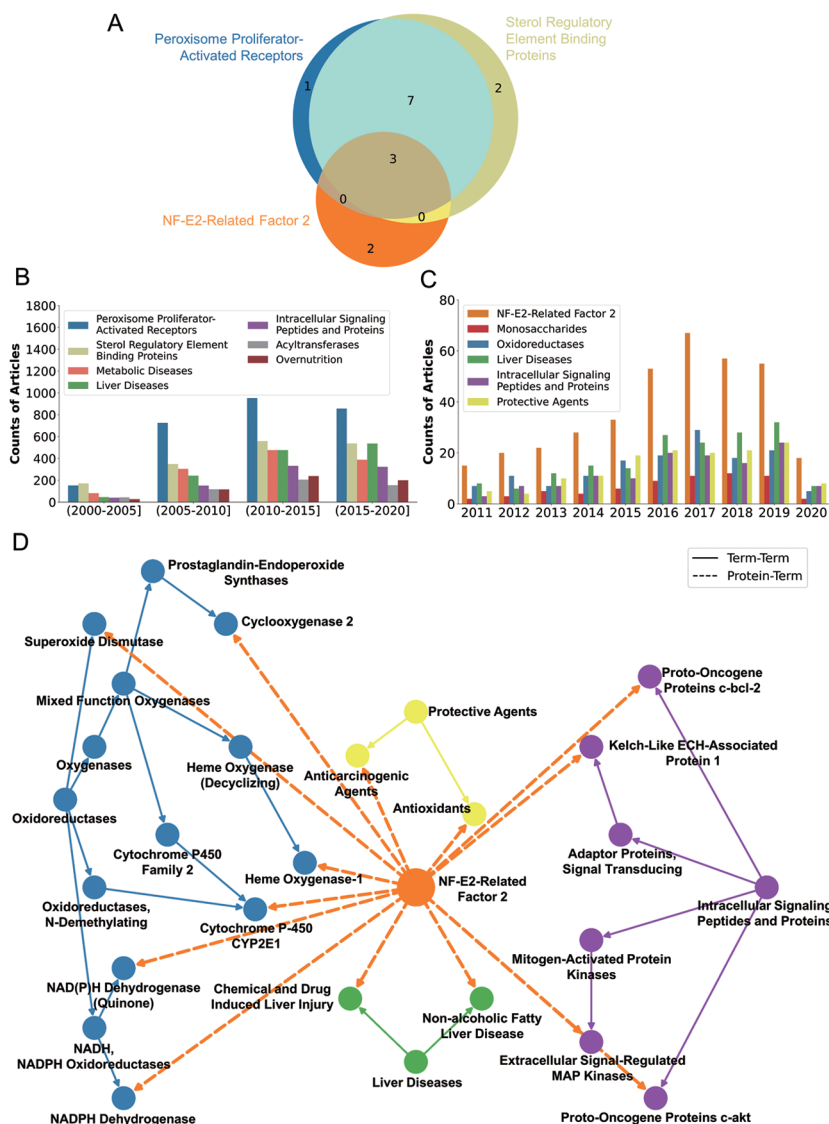


图 6. 肝脏糖脂代谢研究热点蛋白及其相关MeSH词条的关联分析。A: 与“Peroxisome Proliferator-Activated Receptors”、“Sterol Regulatory Element Binding Proteins”和“NF-E2-Related Factor 2”高度相关(皮尔逊相关系数 $r > 0.85$ )的最高层级MeSH词条的重叠分析。B: 包含“Peroxisome Proliferator-Activated Receptors”、“Sterol Regulatory Element Binding Proteins”或与二者之一相关度最高的前5个MeSH词条(皮尔逊相关系数 $r > 0.85$ )的年文献数量统计。C: 包含“NF-E2-Related Factor 2”或与其相关度最高的前5个MeSH词条(皮尔逊相关系数 $r > 0.85$ )的年文献数量统计。D: “NF-E2-Related Factor 2”及与其高度相关并且隶属于“Protective Agents”(黄色)、“Oxidoreductases”(蓝色)、“Liver Diseases”(绿色)、“Intracellular Signaling Peptides and Proteins”(紫色)的子代MeSH词条的关联网络图。图中只展示了与“NF-E2-Related Factor 2”高度相关的MeSH词条(卡方检验和多重假设检验 $FDR < 1 \times 10^{-5}$ )。为了减少网络的复杂度, 连线只展示了“NF-E2-Related Factor 2”与每个MeSH分类最底层节点之间的连接。

Fig. 6. Association analysis of hot proteins and their related MeSH terms in hepatic glucose and lipid metabolism researches. A: Overlap of top-level MeSH terms highly correlated (Pearson correlation coefficient  $r > 0.85$ ) with “Peroxisome Proliferator-Activated Receptors”, “Sterol Regulatory Element Binding Proteins” or “NF-E2-Related Factor 2” in articles. B: Counts of articles related to “Peroxisome Proliferator-Activated Receptors” and “Sterol Regulatory Element Binding Proteins” and their highly correlated (Pearson correlation coefficient  $r > 0.85$ ) top 5 MeSH terms. C: Annual counts of articles related to “NF-E2-Related Factor 2” and its highly correlated (Pearson correlation coefficient  $r > 0.85$ ) top 5 MeSH terms. D: Network of “NF-E2-Related Factor 2” and its highly correlated MeSH terms under the “Protective Agents” (yellow), “Oxidoreductases” (blue), “Liver Diseases” (green) and “Intracellular Signaling Peptides and Proteins” (purple) categories. Only terms highly correlated (Chi-square test and  $FDR < 1 \times 10^{-5}$ ) with “NF-E2-Related Factor 2” are shown. To reduce network complexity, only links between “NF-E2-Related Factor 2” and the bottom layer MeSH terms of each category are drawn.

12 和 5 个共现热点词条，其中 PPARs 和 SREBPs 的共现热点词条具有明显的重复性(图 6A)。“Metabolic Diseases”、“Liver Diseases”、“Intracellular Signaling Peptides and Proteins”、“Acyltransferases”和“Overnutrition”是与 PPARs 和 SREBPs 具有最高综合相关性的 MeSH 主要词条，而与 Nrf2 相关性最高的 MeSH 主要词条分别是“Monosaccharides”、“Oxidoreductases”、“Liver Diseases”、“Intracellular Signaling Peptides and Proteins”和“Protective Agents”(图 6B, 6C)。进一步地，我们选取了与 Nrf2 共出现的高频词条及其与 Nrf2 共同富集的子词条，构建了以 Nrf2 为中心的 MeSH 词条网络，MAP Kinase、c-Akt、NADPH Dehydrogenase 等重要调控蛋白均在此网络中(图 6D)。在疾病和药物方面，与 Nrf2 关联性最强的是肝损伤、非酒精性脂肪肝和抗氧化与抗癌药物(图 6D)。

#### 2.4 肝脏糖脂代谢研究主要目标蛋白相关的疾病与药物分析

根据 MeSH 词条出现的频率，我们选择了蛋白类别词频最高的 5 个词条，包括“Liver X Receptors”、“PPAR alpha”、“Sterol Regulatory Element Binding Protein”、“PPAR gamma”和“NF-kappa B”，分析其与“Nutrition Disorders”、“Liver Diseases”和“Metabolic Diseases”类别下疾病相关 MeSH 词条的关系。我们首先收集了“Nutrition Disorders”、“Liver Diseases”和“Metabolic Diseases”类别下与上述 5 个蛋白词条的共出现频率有显著相关性(卡方检验， $FDR < 1 \times 10^{-2}$ )的所有子词条，根据子词条间的层级关系及其与蛋白的共出现关系，构建了上述 5 个蛋白词条和疾病的关联网络(图 7A)。可以看出，上述五种蛋白作为肝脏糖脂代谢研究的最热门对象，均广泛参与多种糖脂代谢相关疾病的调控。值得指出的是，NF-kappa B 虽然也在糖尿病等代谢类疾病中发挥重要的调控作用，但由于其与代谢性疾病的共现频率未达到统计阈值，因此在网络中并未展示出直接关联。

通过检索 Clinical Trials 网站(<https://clinicaltrials.gov/>)，我们发现多个针对上述靶点的临床试验与肝脏糖脂代谢相关疾病有关，如 PPAR alpha 激动剂非诺贝特是临床常用的血脂调节药，传统用于高甘油三酯血症的治疗，但目前已有 4 项临床试验研究非诺贝特在治疗非酒精脂肪肝和肝脏胰岛素抵抗中的作用。PPAR gamma 激动剂——罗格列酮和吡格

列酮是已上市多年的噻唑烷二酮类抗糖尿病药物，虽然这些药物在心血管方面的副作用还存在争议，但其改善肝脏胰岛素抵抗、减少肝糖输出的功效已得到公认。目前共有 28 项临床试验研究罗格列酮和吡格列酮在降糖以外对非酒精性脂肪肝的作用。直接靶向“Liver X Receptors”、“Sterol Regulatory Element Binding Proteins”的药物目前多处于临床前期阶段，在 Clinical Trials 网站注册的相关研究仅各有一项临床试验，其针对的疾病种类为糖尿病视网膜病变及甲状腺癌。

我们进一步分析了上述 5 种蛋白与“Therapeutic Uses”类别下高频词的共现关系(图 7B)。与图 6A 中展示的蛋白在多种疾病中的广泛调控作用不同，在已上市的临床药物中，只有针对 PPAR alpha 的药物功效较广泛，针对其他几种蛋白的药物在功效方面还相对比较集中。现有文献中，针对 NF-kappa B 的临床药物相关研究最多，主要集中在抗炎相关的功能(图 7B)。鉴于这些蛋白在糖脂代谢方面的功能多样性，预期以这些蛋白为靶点的其他功效的药物在未来也将相继问世。

### 3 讨论

随着人们生活条件的改善和生活方式的改变，肥胖和糖尿病等糖脂代谢相关疾病的患病率持续攀升，针对糖脂代谢调节的分子机制和新药开发研究变得愈加重要。为帮助研究人员快速了解肝脏糖脂代谢领域的研究状况，本文利用 MeSH 词条和 GEO 数据库中的高通量数据，对肝脏糖脂代谢领域的研究现状、研究方法和研究热点等因素的变化趋势进行了系统分析。我们发现：(1) 肝脏糖脂代谢研究自本世纪以来进入快速发展时期，领域内研究人员和研究机构间的合作也明显增多，这种多人员、多单位的合作更有利于高质量研究成果的产出；(2) 中国的肝脏糖脂代谢研究虽然起步较晚，但发展迅速，现已成为推动肝脏糖脂代谢研究发展的重要动力；(3) 基因芯片、高通量测序等新兴高通量研究手段已逐渐成为肝脏糖脂代谢领域遗传学相关研究常用的方法；(4) 针对肝脏糖脂代谢的研究主要围绕肝脏糖脂代谢失衡所引起的各类代谢疾病和肝脏疾病及其分子机制开展，除糖脂代谢本身所涉及的反应底物和反应产物之外，对糖脂代谢有重要调控作用的蛋白质、多肽等物质也是被重点关注的研究对象；(5) 以 PPARs、SREBPs 和 Liver X Receptors

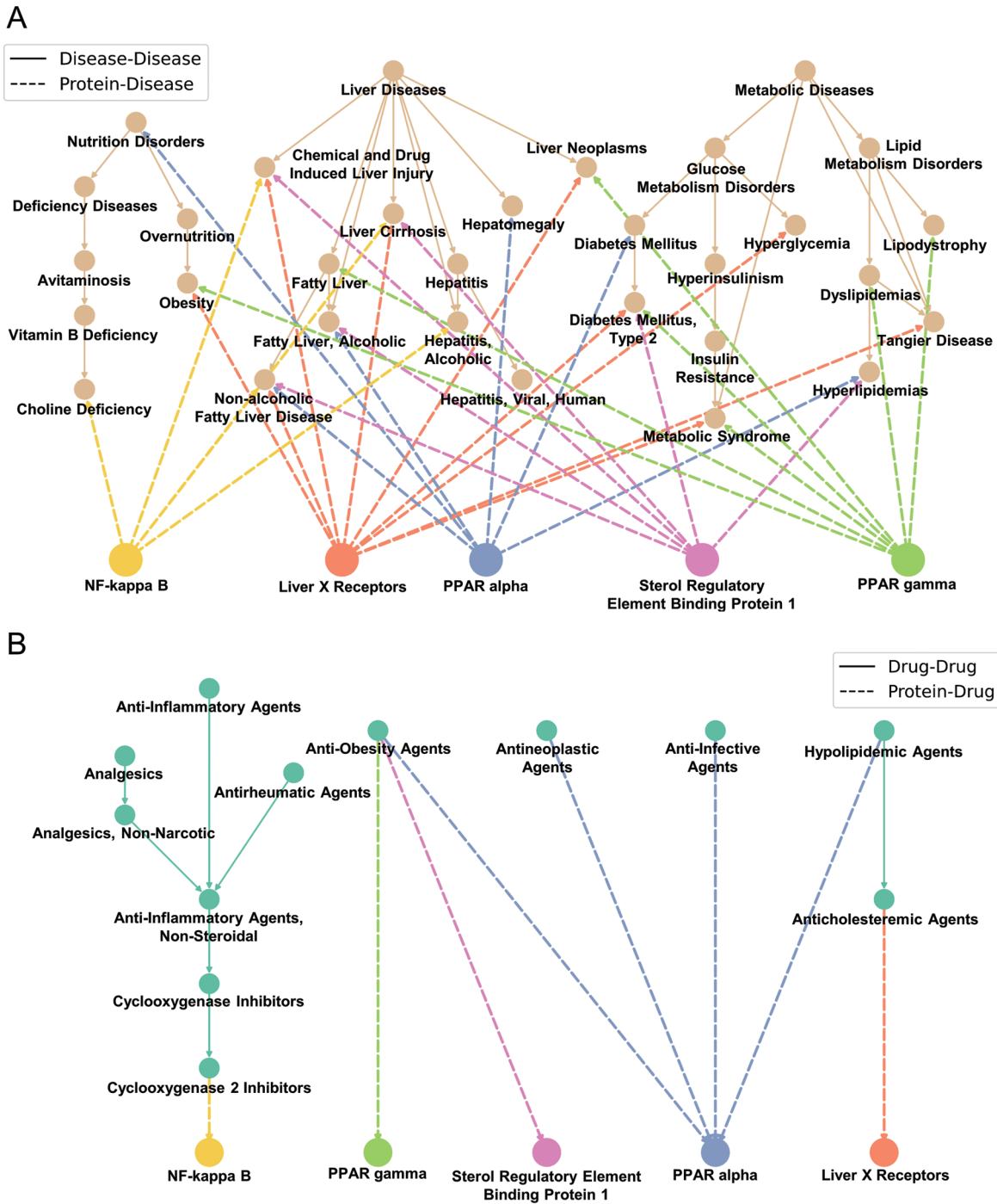


图 7. 肝脏糖脂代谢研究热点蛋白相关的疾病与药物功效。A: 研究最多的5种蛋白及与其高度相关的“Nutrition Disorders Disease”、“Liver Diseases”和“Metabolic Diseases” MeSH词条的子代词条关联网络图。B: 研究最多的5种蛋白及其与“Therapeutic Uses” MeSH词条的子代词条的关联网络图。在A和B中, 只展示了与五种蛋白中的任一种高度相关的MeSH词条(卡方检验和多重假设检验 $FDR < 1 \times 10^{-2}$ )。为了减少网络的复杂度, 如果一个蛋白与一个MeSH子树的多个节点均相关, 只展示了该蛋白与MeSH子树最底层节点之间的连接。

Fig. 7. Diseases and drug therapeutic effects related to the hot proteins of hepatic glucose and lipid metabolism researches. A: Network of the top 5 most studied proteins and their highly correlated Mesh terms under the “Nutrition Disorders Disease”, “Liver Diseases” and “Metabolic Diseases” categories. B: Network of the top 5 most studied proteins and their highly correlated Mesh terms under the “Therapeutic Uses” category. In both A and B, only terms highly correlated ( $\chi^2$  test and  $FDR < 1 \times 10^{-2}$ ) with any of the 5 proteins are shown. To reduce network complexity, if a protein has correlations with multiple components of a MeSH term subtree, only links between the protein and the bottom layer MeSH terms of the subtree are drawn.

为代表的转录因子、核受体等在 2001 年之后成为受关注较多的研究对象。过去 10 年左右，针对抗氧化应激转录因子 Nrf2 的研究快速增多，成为新兴研究热点。

从我们的分析结果可以看出，开展多人员、跨单位的合作研究是本领域的发展趋势，并且也有助于科研产出质量的提升。本研究的统计结果显示，论文作者在 3~7 人或研究机构在 2~4 个时，其科研成果中高影响力研究论文的比例相对较少。这种情况有可能是因为处于这一范围的研究论文数目偏多而导致的，并不能代表在这一范围的研究组合的相对科研能力。此外，本研究仅仅依据论文发表期刊的影响因子来评价论文的影响力，也存在一定的偏颇。

二十一世纪以来相继出现的多种高通量研究方法在很大程度上改变了生命科学的研究范式，其在肝脏糖脂代谢研究领域的应用越来越广。目前统计的肝脏糖脂代谢相关高通量数据中，基于基因芯片的数据还占大多数，这主要是由于目前 GEO 数据库收录的基因芯片数据总量多于高通量测序数据。随着基因芯片应用的减少和高通量测序方法的进一步普及，预期未来基于高通量测序方法产生的数据量会有较快速提升。

PPARs 作为一类依赖配体激活的核激素受体，在脂肪酸氧化和胰岛素信号传递等糖脂代谢的重要途径中都发挥关键调控作用。如前所述，目前已有多种 PPAR 家族蛋白激动剂类药物被用于糖尿病、血脂异常和非酒精性脂肪肝等糖脂代谢相关疾病的治疗<sup>[22]</sup>。肝脏糖脂代谢研究关注的热点蛋白，如 SREBPs、Liver X Receptors、Nrf2 等也已成为药物研发的热门靶点，并且已有多项相关药物进入各期临床试验，预计未来会有更多通过调控这些靶点治疗糖脂代谢相关疾病的新药问世。此外，还有一大批新兴的肝糖脂代谢疾病干预靶点陆续被发现，如线粒体内膜蛋白 mGPDH、凋亡调节子 CFLAR、胞质多聚腺苷酸化元件结合蛋白 CPEB4 等，这些潜在的干预靶点将为肝糖脂代谢相关疾病的防治带来新的希望<sup>[23-25]</sup>。

随着各领域研究的不断深入和相关科研论文的快速生长，科研人员对于快速获取某一研究领域的总体情况和研究热点变化等宏观信息的需求日益增大。本研究以肝脏糖脂代谢为范例，展示了利用基于文献和公共数据的信息挖掘整理技术解决上述问

题的可行性。本研究开发的方法也可被应用于其他研究领域，从而为科研界、企业界和政府部门相关人员快速了解领域现状和发展趋势提供便利。

### 参考文献

- 1 Nils-Georg A, John C, Erik ODB, Mark D, Mukund G, Seyed MK, Shuhei K, David L, MANN JI, Ruth OO, James H, Kraissid T, Hester HV, Alison S, Mark W, Thomas W. The role of carbohydrates in nutrition. In: Carbohydrates in Human Nutrition. Report of the joint FAO/WHO Expert Consultation, Rome, 14-18 April, 1997.
- 2 Wasserman DH. Four grams of glucose. *Am J Physiol Endocrinol Metab* 2009; 296(1): E11-E21.
- 3 Fahy E, Subramaniam S, Murphy RC, Nishijima M, Raetz CR, Shimizu T, Spener F, van Meer G, Wakelam MJ, Dennis EA. Update of the LIPID MAPS comprehensive classification system for lipids. *J Lipid Res* 2009; 50 Suppl(Suppl): S9-S14.
- 4 Chen J, Chevassus-Agnès S, Gilani GS, Harnly J, Livesey G, William C. MacLean J, Mathioudakis B, Chavez MMd, Torun B, Vasconcellos MTLd, Warwick P, Sizaret F, Kennedy G, Prinzo ZW. Calculation of the energy content of foods-energy conversion factors. In: FAO (ed.) FAO Food Nutrition Paper, Food energy-methods of analysis and conversion factors. Food and Agriculture Organisation of the United Nations, Rome, Italy, 2003.
- 5 Boyce JA. Eicosanoids in asthma, allergic inflammation, and host defense. *Curr Mol Med* 2008; 8(5): 335-349.
- 6 Beltowski J. Liver X receptors (LXR) as therapeutic targets in dyslipidemia. *Cardiovasc Ther* 2008; 26(4): 297-316.
- 7 Rajas F, Gautier-Stein A, Mithieux G. Glucose-6 phosphate, a central hub for liver carbohydrate metabolism. *Metabolites* 2019; 9(12): 282.
- 8 Adeva-Andany MM, Pérez-Felpete N, Fernández-Fernández C, Donapetry-García C, Pazos-García C. Liver glucose metabolism in humans. *Biosci Rep* 2016; 36(6): e00416.
- 9 Lubert Stryer, Berg JM, Tymoczko JL. *Biochemistry*. New York: W.H. Freeman and Company, 2002.
- 10 McDevitt RM, Bott SJ, Harding M, Coward WA, Bluck LJ, Prentice AM. *De novo* lipogenesis during controlled overfeeding with sucrose or glucose in lean and obese women. *Am J Clin Nutr* 2001; 74(6): 737-746.
- 11 Mithieux G, Rajas F, Gautier-Stein A. A novel role for glucose 6-phosphatase in the small intestine in the control of glucose homeostasis. *J Biol Chem* 2004; 279(43): 44231-44234.
- 12 Nguyen P, Leray V, Diez M, Serisier S, Le Bloc'h J, Siliart B, Dumon H. Liver lipid metabolism. *J Anim Physiol Anim*

- Nutr (Berl) 2008; 92(3): 272–283.
- 13 Alves-Bezerra M, Cohen DE. Triglyceride metabolism in the liver. *Compr Physiol* 2017; 8(1): 1–8.
  - 14 Choi SH, Ginsberg HN. Increased very low density lipoprotein (VLDL) secretion, hepatic steatosis, and insulin resistance. *Trends Endocrinol Metab* 2011; 22(9): 353–363.
  - 15 Lin Hua V, Accili D. Hormonal regulation of hepatic glucose production in health and disease. *Cell Metab* 2011; 14(1): 9–19.
  - 16 Hatting M, Tavares CDJ, Sharabi K, Rines AK, Puigserver P. Insulin regulation of gluconeogenesis. *Ann N Y Acad Sci* 2018; 1411(1): 21–35.
  - 17 Smith BW, Adams LA. Non-alcoholic fatty liver disease. *Crit Rev Clin Lab Sci* 2011; 48(3): 97–113.
  - 18 Hazlehurst JM, Woods C, Marjot T, Cobbold JF, Tomlinson JW. Non-alcoholic fatty liver disease and diabetes. *Metabolism* 2016; 65(8): 1096–1108.
  - 19 Bechmann LP, Hannivoort RA, Gerken G, Hotamisligil GS, Trauner M, Canbay A. The interaction of hepatic lipid and glucose metabolism in liver diseases. *J Hepatol* 2012; 56(4): 952–964.
  - 20 Xu L, Nagata N, Ota T. Impact of glucoraphanin-mediated activation of Nrf2 on non-alcoholic fatty liver disease with a focus on mitochondrial dysfunction. *Int J Mol Sci* 2019; 20(23): 5920.
  - 21 Long M, Rojo de la Vega M, Wen Q, Bharara M, Jiang T, Zhang R, Zhou S, Wong PK, Wondrak GT, Zheng H, Zhang DD. An essential role of NRF2 in diabetic wound healing. *Diabetes* 2016; 65(3): 780–793.
  - 22 Pawlak M, Lefebvre P, Staels B. Molecular mechanism of PPAR $\alpha$  action and its impact on lipid metabolism, inflammation and fibrosis in non-alcoholic fatty liver disease. *J Hepatol* 2015; 62(3): 720–733.
  - 23 Zheng Y, Qu H, Xiong X, Wang Y, Liu X, Zhang L, Liao X, Liao Q, Sun Z, Ouyang Q, Yang G, Zhu Z, Xu J, Zheng H. Deficiency of mitochondrial glycerol 3-phosphate dehydrogenase contributes to hepatic steatosis. *Hepatology* 2019; 70(1): 84–97.
  - 24 Wang PX, Ji YX, Zhang XJ, Zhao LP, Yan ZZ, Zhang P, Shen LJ, Yang X, Fang J, Tian S, Zhu XY, Gong J, Zhang X, Wei QF, Wang Y, Li J, Wan L, Xie Q, She ZG, Wang Z, Huang Z, Li H. Targeting CASP8 and FADD-like apoptosis regulator ameliorates nonalcoholic steatohepatitis in mice and nonhuman primates. *Nat Med* 2017; 23(4): 439–449.
  - 25 Maillo C, Martín J, Sebastián D, Hernández-Alvarez M, García-Rocha M, Reina O, Zorzano A, Fernandez M, Méndez R. Circadian- and UPR-dependent control of CPEB4 mediates a translational response to counteract hepatic steatosis under ER stress. *Nat Cell Biol* 2017; 19(2): 94–105.