

• 专题研究:肿瘤 •

## 基于 cfDNA 多组学模型进行头颈部鳞状细胞癌早期诊断的研究进展

杨明哲, 陈仁杰\*

南京医科大学第二附属医院耳鼻咽喉头颈外科, 江苏 南京 210003

**[摘要]** 头颈部鳞状细胞癌(head and neck squamous cell carcinomas, HNSCC)是全球第六大常见癌症,其早期诊断因症状隐匿和特异性标志物缺乏而存在困难。随着液体活检技术的发展,循环游离DNA(circulating free DNA, cfDNA)作为其核心检测对象,具有非侵入性、动态监测和反映肿瘤特征的优势,在HNSCC早期筛查、预后评估和疗效监测中展现出应用前景。但现有的单组学研究仍受限于低灵敏度和肿瘤异质性的影响。通过结合基因组、转录组、表观组、蛋白质组信息的多组学整合策略,可显著提高诊断准确率,并揭示疾病分子机制。近年来,随着深度学习和机器学习技术被广泛运用于cfDNA的多组学数据分析,推动了相关标志物筛选和模型构建。现有研究表明,基于cfDNA的多组学模型有望提高HNSCC的早期诊断水平,并为精准医学的发展提供新方向。

**[关键词]** 头颈部鳞状细胞癌;循环游离DNA;液体活检;多组学整合;深度学习;早期诊断;生物标志物;精准医学

**[中图分类号]** R739.91

**[文献标志码]** A

**[文章编号]** 1007-4368(2026)01-21-10

**doi:** 10.7655/NYDXBNSN251049

## Research progress on early diagnosis of head and neck squamous cell carcinoma based on cfDNA multi-omics models

YANG Mingzhe, CHEN Renjie\*

Department of Otolaryngology and Head and Neck Surgery, the Second Affiliated Hospital of Nanjing Medical University, Nanjing 210003, China

**[Abstract]** Head and neck squamous cell carcinomas (HNSCC) rank as the sixth most common cancer globally, with early diagnosis challenging due to subtle symptoms and the absence of specific biomarkers. With the development of liquid biopsy technology, circulating free DNA (cfDNA), as its core test object, has the advantages of non-invasive, dynamic monitoring and reflecting tumor characteristics, and shows promising applications in early screening, prognostic assessment and efficacy monitoring of HNSCC. However, existing single-omics studies are still limited by low sensitivity and tumor heterogeneity. Multi-omics integration strategies combining genomic, transcriptomic, epigenomic, and proteomic information can significantly enhance diagnostic accuracy and reveal disease molecular mechanisms. In recent years, as deep learning and machine learning technologies have been widely applied to analyze multi-omics data of cfDNA, promoting relevant marker screening and model construction. Existing research indicates that multi-omics models based on cfDNA hold promise for improving early diagnosis of HNSCC and offer new directions for precision medicine development.

**[Key words]** head and neck squamous cell carcinomas; cell-free DNA; liquid biopsy; multi-omics integration; deep learning; early diagnosis; biomarkers; precision medicine

[J Nanjing Med Univ, 2026, 46(01): 21-30]

**[基金项目]** 江苏省科技计划项目(BL2013038)

\*通信作者(Corresponding author), E-mail: crjent@njmu.edu.cn (ORCID: 0000-0002-4768-7650)

头颈部鳞状细胞癌(head and neck squamous cell carcinomas, HNSCC)是头颈部最常见的恶性肿瘤之一。在世界范围内,每年HNSCC的发病人数超过89万,且发病率仍在持续上升<sup>[1]</sup>。HNSCC起源于鼻窦、口腔、鼻咽、口咽、喉咽、喉的黏膜上皮细胞,并遵循从黏膜上皮细胞增生—异型增生—原位癌—浸润性癌的进展顺序。HNSCC的主要危险因素包括吸烟、饮酒、人乳头瘤病毒(human papilloma virus, HPV)感染等。HNSCC早期诊断的主要方法包括视诊、触诊、内镜检查、影像学检查等,但由于HNSCC具有早期症状隐匿和缺乏特异性的特点,其早期诊断十分困难。HNSCC必须通过原发肿瘤和/或颈部肿块活检来确诊<sup>[2]</sup>,但是组织活检技术具有侵袭性、取样偏差和无法实时动态监测等不可避免的缺陷。早期诊断是提高HNSCC患者生存率的关键,然而,由于缺乏特定症状和有效的早期筛查方法,许多肿瘤未能在早期发现而延误诊治,开展HNSCC早期筛查是目前的一项重要研究课题。

液体活检(liquid biopsy)技术凭借着非侵入性、高安全性、高效率性、重复性强和动态反映病情等优势在肿瘤的早期诊断、寻找相关生物标志物和预后评估等方面都具有显著优势。循环游离DNA(cell-free DNA, cfDNA)是指未被细胞膜包裹的DNA片段,通常存在于外周血浆、唾液或尿液中由机体组织细胞在发生凋亡、坏死或主动分泌时释放入循环。cfDNA中存在一种来自肿瘤细胞的凋亡、坏死或主动释放的循环肿瘤DNA(circulating tumor DNA, ctDNA),其反映了肿瘤的基因组特征和动态演变,通俗来讲,可称之为“肿瘤指纹”。ctDNA已在多个癌种的临床检测中得到应用,于2020年首次获得美国食品和药品管理局批准,成为泛肿瘤液体活检产品<sup>[3]</sup>,使得HNSCC的早期诊断和精准治疗成为可能。

目前仍有许多依赖cfDNA单组学分析方法的研究和临床试验正在进行。单组学分析方法虽然在某种癌症或某个研究阶段存在一定实用价值,但由于其只研究了cfDNA中的一类特征信息,必然会存在某种限制,如检测灵敏度不足、受肿瘤异质性影响、背景噪音干扰以及生物信息维度单一等一系列问题,这些问题与肿瘤早期诊断所需要的高敏感度背道而驰。所以有关cfDNA的研究正由单组学向多组学转变,这不仅需要整合cfDNA的突变、甲基化、片段长度特征,甚至还需要融合蛋白组学以及转录组信息,以提高检测的灵敏度、特异度和肿瘤全景描述能力,从而更有效服务于肿瘤早筛和精

准治疗。

在近年来的研究中,深度学习已成为多组学模型分析的重要工具。在处理存在灵敏度不足、背景噪音干扰等问题的cfDNA数据时,可以借助深度学习强大的非线性特征学习能力和跨维度建模能力,从中提取出癌症的相关信号,以实现其早期诊断。

文章旨在建立一个基于深度学习和多组学数据分析的HNSCC cfDNA液体活检早期筛查模型,为HNSCC的早期诊断和精准治疗提供新的科学依据和技术路径,进而提高治疗的成功率并减轻患者的负担。

## 1 cfDNA的生物学特性、功能与相关检测技术

cfDNA最早由Mandel和Metais<sup>[4]</sup>于1948年在血液中发现,主要来源于细胞的程序性死亡、坏死或主动释放。随着近年来高通量测序和数字PCR等技术的发展,cfDNA在液体活检领域的应用逐渐增多,并在肿瘤早期诊断、治疗效果检测、预后评估等方面显示出巨大的发展潜力<sup>[5]</sup>。

### 1.1 cfDNA的生物学特性

#### 1.1.1 片段化

cfDNA主要以双链的形式存在,片段长度集中在160~180 bp(单核小体)和320~360 bp(双核小体)<sup>[6]</sup>。肿瘤组织及细胞所释放的ctDNA则可能表现为更短的片段。Mouliere等<sup>[7]</sup>开发了基于片段大小的分析方法,发现HNSCC患者cfDNA中低于150 bp的片段比例显著增加,这一特征可用于区分肿瘤患者和健康个体。

#### 1.1.2 浓度的动态变化

一般来说,健康个体的cfDNA浓度通常较低,但在病理条件下,如肿瘤、创伤或器官移植后的排异反应等情况,cfDNA的浓度会显著提高<sup>[8]</sup>。但是肿瘤来源的ctDNA通常只占据cfDNA总量的很小一部分,故而需要提高检测技术的灵敏度。

#### 1.1.3 甲基化特征

cfDNA的甲基化模式和突变特征可以反映来源组织。例如,来源于肿瘤的cfDNA,即ctDNA通常会携带相关突变基因和特异性甲基化模式<sup>[9]</sup>。主要包括TP53、CDKN2A、PIK3CA、NOTCH1、FAT1等基因。

#### 1.1.4 快速清除

cfDNA在血液中的半衰期较短,一般认为在16~150 min。这种特性使其能够成为实时监测疾病动态、疾病的早期诊断和干预、疾病负荷评估的理想标志物(表1)。

表1 cfDNA 的生物学特性  
Table 1 Biological characteristics of cell-free DNA

Characteristic	Description	Significance
Fragmentation	Predominantly exists as fragments of specific sizes. Shorter fragments increase in certain diseases	Fragment size pattern can differentiate healthy and diseased states
Variable concentration	Low baseline in health; significantly elevated in pathology (e.g., cancer, trauma)	Serves as a sensitive indicator of active disease or tissue injury
Tissue-specific marks	Carries DNA methylation and genetic signatures of its cell of origin	Enables non-invasive identification of the contributing tissue or disease type
Rapid clearance	Rapidly cleared from blood, with a half-life ranging from minutes to a few hours	Provides a real-time snapshot of disease dynamics for monitoring

## 1.2 cfDNA 的生物学功能

cfDNA 不仅是细胞程序性死亡的产物, 还包含多种生物学功能。

### 1.2.1 细胞间通讯

cfDNA 可以参与细胞间通讯, 主要以微囊泡或外泌体的形式进行。相关研究表明, 由肿瘤细胞所释放的 cfDNA 可以起到调节肿瘤微环境的作用, 从而促进肿瘤的生长和转移<sup>[10]</sup>。

### 1.2.2 免疫调控

cfDNA 是免疫系统中的调节因子, 主要通过病原相关分子模式 (pathogen-associated molecular pattern, PAMP)、损伤相关分子模式 (damage-associated molecular pattern, DAMP) 和参与自身免疫反应这 3 种形式来参与。

### 1.2.3 传递遗传信息

cfDNA 携带了细胞的全部或部分基因组信息, 可以反映原细胞的基因突变、甲基化等表观遗传学改变<sup>[11]</sup>。例如来自肿瘤细胞的 cfDNA, 其中就包含着肿瘤相关的特异性突变基因, 可用于肿瘤相关的基因组分析<sup>[12]</sup>。这让 cfDNA 成为液体活检的重要生物标志物, 可应用于肿瘤的早期诊断、治疗检测和预后评估。

## 1.3 cfDNA 的相关检测技术

cfDNA 的检测流程主要包括样本采集与处理、提取 cfDNA、定量、质控和下游分析<sup>[13]</sup>。cfDNA 的样本类型包括血液、尿液、唾液等。血浆是目前最常用的检测标本类型, 因为血浆中的 cfDNA 浓度相对稳定, 且受细胞的基因组 DNA (genomic DNA, gDNA) 污染的程度较小。血液样本采集通常使用 Streck 管, 使用该管无需立即离心, 可室温保存 7~14 d, 也可防止 gDNA 污染及 cfDNA 的自行降解。之后检测时通过双重离心法分离血浆中的 cfDNA, 可有效避免 gDNA 的污染。cfDNA 常见的提取方法包括

酚-氯仿法、磁珠法和硅胶膜柱法。硅胶膜柱法是目前最广泛使用的提取方法, 主要依赖于核酸与硅胶膜的选择性结合, 缺点则在于该方法的回收效率有限<sup>[10]</sup>。磁珠法则是利用磁性颗粒吸附 cfDNA, 并可用于自动化提取, 达成更高的回收效率, 但成本较为高昂<sup>[14]</sup>。这两种方法各有优劣。定量和质控方法包括荧光染料法、紫外分光光度法、实时荧光定量 PCR (real-time PCR, qPCR)、数字 PCR 等。下游分析方法则包括检测基因突变的扩增难变性突变系统 PCR (amplification refractory mutation system PCR, ARMS-PCR)、BEAMing (beads, emulsion, amplification, and magnetics) 和高通量测序 (next-generation sequencing, NGS); 分析 cfDNA 甲基化情况的亚硫酸氢盐转化法、甲基化特异性 PCR (methylation-specific PCR, MSP) 及甲基化测序; 分析 cfDNA 片段化模式的 NGS 以及针对拷贝数变异检测的 NGS 或微阵列技术 (microarray technology)。目前已有相关研究开发出利用 TAPS (TET-assisted pyridine borane sequencing), 即一种替代亚硫酸氢盐测序的无损甲基化测序方法, 能够在同一数据中保留基因组和甲基化信息<sup>[15]</sup>, 使用液滴数字 PCR (droplet digital PCR, ddPCR) 和 NGS 技术动态监测 cfDNA。在实际实验或应用中, 需要根据具体需求来选择合适的方法, 以达成预期的目标效果。

## 2 cfDNA 在 HNSCC 中的研究现状

cfDNA 在 HNSCC 中的研究和临床应用主要集中于疾病的早期诊断、预后评估和治疗监测。

### 2.1 早期诊断

在 HNSCC 的早期阶段, cfDNA 检测可作为一个非侵入性的工具。在 HNSCC 患者血液 cfDNA 中可以检测到肿瘤相关的特异性突变和甲基化标志物。一项包含 200 例 HNSCC 患者的研究显示, 基于

cfDNA 中 TP53 突变和 CDKN2A 甲基化的联合检测, 诊断灵敏度达 78%, 特异度为 92%<sup>[16]</sup>。

## 2.2 预后评估

多项研究表明, 治疗前 cfDNA 浓度和其中的 ctDNA 水平往往与 HNSCC 患者的预后密切相关。血浆中 cfDNA 浓度越高, cfDNA 中的 ctDNA 水平越高, 患者预后往往越差。Chen 等<sup>[17]</sup>的相关研究报道, cfDNA >30 ng/mL 的患者 3 年生存率仅为 25%, 而 <30 ng/mL 组为 65%。并且有相关研究报道, cfDNA 中的特定基因突变, 例如 NOTCH1 突变的存在也与肿瘤分化障碍和肿瘤侵袭性增强相关, 可造成患者预后不良<sup>[18]</sup>。所以, 通过对 cfDNA 分子层面的分析, 可以揭示 HNSCC 患者的分子特征, 从而预测患者的疾病复发、转移风险和生存时间, 较传统的 TNM 分期可以提供更加个体化、精细化的预后评估。

## 2.3 治疗检测

cfDNA 的动态变化可以反映 HNSCC 的治疗效果。相关研究表明, 治疗 2 周后 cfDNA 浓度下降 >50% 的患者, 其客观缓解率显著高于下降幅度小的患者 (78% vs. 32%)<sup>[19]</sup>。另外一项 II 期临床试验显示, 基于 cfDNA 监测调整治疗方案可使 HNSCC 患者中位无进展生存期延长 3.2 个月<sup>[20]</sup>。

以上相关研究虽在 HNSCC 的早期诊断、预后评估及治疗检测中展现出较大潜力, 但其单组学研究存在的固有限制使如今的研究人员不得不转向多组学整合。

## 3 用于 HNSCC 早期诊断的 cfDNA 多组学整合策略与标志物

### 3.1 多组学整合的概念及其核心理念

多组学整合, 即同时采集和分析多种生物学层面的数据, 如通过整合基因组、转录组、蛋白组和代谢组等多维度数据, 并通过生物信息学方法融合分析, 以揭示疾病的分子机制<sup>[21-22]</sup>, 是后基因组时代的重要方法论。多组学整合有三大核心理念。

**互补性:** 多组学整合通过整合基因组学、表观组学、片段组学和蛋白质组与代谢组, 将上述生物信息组别协同解码, 揭示了生物系统的复杂性与互补性。基因组学通过识别遗传变异揭示疾病发生的基础, 例如 TP53 突变需结合染色质开放区域 (片段组学) 和启动子超甲基化 (表观组) 验证其功能性影响。而表观组学通过 DNA 甲基化和组蛋白修饰调控基因表达, 表观沉默基因的蛋白质表达缺失可能早于转录水平变化, 提示跨组学验证的必要

性<sup>[21]</sup>。片段组学绘制染色质空间构象图谱, 如拓扑关联域重构与增强子-启动子环化共同解释 MYC 基因过度表达<sup>[23]</sup>。蛋白质组与代谢组作为功能输出的终端验证, 例如表皮生长因子受体 (epidermal growth factor receptor, EGFR) 突变激活丝裂原活化蛋白激酶 (mitogen activated protein kinases, MAPK) 通路促进乳酸生成, 代谢通量分析甚至可推断未检测的酶活性变化<sup>[22]</sup>。

**协同增效:** 多组学的协同增效将突破单组学的固有限制, 在敏感性、特异性和系统性 3 个方面实现全面提升。单组学的敏感性较低, 这源于生物学的复杂性和相关检测技术的局限性。相关研究表明, 将 ctDNA、外泌体 miRNA 和相关蛋白质标志物整合, 可将早期癌症的检出率提升至 92%<sup>[24]</sup>。单组学研究中数据的高噪声导致其数据的假阳性率较多组学明显偏高。研究表明, 通过整合表达数量性状位点 (expression quantitative trait loci, eQTL, 转录组)、甲基化数量性状位点 (methylation quantitative trait loci, meQTL, 表观组) 和蛋白质数量性状位点 (protein quantitative trait loci, pQTL, 蛋白质组) 数据, 可显著提高功能性单核苷酸多态性 (single nucleotide polymorphism, SNP) 的识别率和特异性<sup>[25]</sup>。基于多组学的整合优势, 可构建动态调控网络, 较单组学的线性关系有着巨大的系统性的解析潜力。

**数据融合策略:** 通过整合基因组、表观组、蛋白质组和代谢组等多组学数据, 并将其融合, 可以进一步展示生物系统间的复杂关系, 并为精准医学提供基础和技术支持。数据融合策略主要包括特征级、决策级、模型级。特征级融合, 即通过标准化和降维技术将多组学中的特征投射到同一个空间。例如, “泛癌图谱” (Pan-Cancer Atlas) 项目整合了 33 种癌症的基因组、表观组数据, 并结合临床数据, 探索出了跨癌种的共同驱动通路<sup>[26]</sup>。决策级融合是一种通过整合多个独立模型的预测结果从而提高整体分析准确性的方法, 主要通过加权投票和 Stacking 算法实现。目前已有研究基于 Stacking 算法, 利用机器学习开发出了预测准确率和预测性能较传统方法显著提高的新算法<sup>[27]</sup>。模型级融合则主要通过结合多个模型的预测结果来提高整体性能。这 3 种策略各有优势, 共同推动多组学整合在精准医学中的发展。

### 3.2 HNSCC 中各关键组学的标志物

**基因组学:** 可揭示 HNSCC 中的基因突变、染色体重排和拷贝数变异等信息。TP53 突变是 HNSCC

中最常见的基因改变之一,其突变频率较高,常与肿瘤侵袭性和预后不良相关<sup>[28]</sup>。此外,有相关研究揭示 PIK3CA、CDKN2A、NOTCH1 等基因也与 HNSCC 的发生、发展密切相关<sup>[18]</sup>。并且,源于 HNSCC 的病毒相关性,在挖掘生物标志物时需基于病毒状态进行分子分型。Seiwert 等<sup>[29]</sup>的研究显示,HPV 阴性与阳性的突变谱截然不同:阴性富集 TP53、CDKN2A、MLL2、CUL3、NSD1、PIK3CA 等相关的基因突变;阳性则在 DDX3X、FGFR2/3 中显示出独特的突变。而 Lechner 等<sup>[30]</sup>的研究则表明 SSTR2 在 EB 病毒相关的 HNSCC 中呈现出高频且特异的表达。

**转录组学:**可揭示 HNSCC 中基因表达的异常模式,主要包括 mRNA、非编码 RNA 两个方面。例如,EGFR 是许多癌症中常见的驱动基因,其过表达与肿瘤的增殖、侵袭和治疗耐药密切相关<sup>[31]</sup>。非编码 RNA 方面,有相关研究表明,HPV E6/E7 基因表达是 HPV 阳性 HNSCC 的重要标志物<sup>[32]</sup>。

**表观组学:**主要涉及 HNSCC 中 DNA 甲基化、组蛋白修饰、染色质重塑的相关作用。在 HPV 阴性 HNSCC 中,p16 基因的启动子甲基化较为常见,并且这种甲基化与肿瘤抑制基因失活相关<sup>[31]</sup>。

**蛋白质组学:**主要表现为 HNSCC 中关键蛋白表达和功能方面的变化。在 HNSCC 的复发和转移过程中,免疫逃逸是一个重要因素,而 PD-L1 的表达被认为是其中的一部分机制<sup>[33]</sup>。

### 3.3 HNSCC 相关的 cfDNA 多组学标志物研究

在 HNSCC 的相关研究中,cfDNA 的遗传特征和表观遗传特征可以反映出肿瘤分子的变异。基于基因组学、转录组学、表观组学等多组学整合,在标志物的诊断、预测准确率和预测性能等方面可获得显著提高。为了实现多组学数据的整合,使 cfDNA 中的遗传信息和表观遗传信息获得充分的发掘,机器学习算法得到了广泛应用,旨在构建更加精准的标志物组合。随机森林(random forest)、支持向量机(support vector machine, SVM)等机器学习算法能够高效率处理高维度、多变量的组学数据。而深度学习模型,特别是自编码器(autoencoder),则能通过非线性降维,从海量数据中自动学习深层次特征,更有效地解决多组学数据整合中的高维度和噪声问题<sup>[34]</sup>。生物网络分析也是一种重要的整合方法。利用该方法,将多组学数据整合后,可以更加全面地理解疾病发生发展的分子机制,从而发现新的疾病基因、解释基因突变的意义,并识别相关生物标志物<sup>[35]</sup>。多组学通过整合 HNSCC 患者的免疫微环

境、基因突变、HPV 感染及解剖部位等信息,结合预测模型和临床前研究模型,强化了免疫治疗的作用,为开发更精准的个性化治疗方案提供了科学基础,并能够发现新治疗靶点以提高免疫疗法的疗效<sup>[36]</sup>。随着数据整合策略的进一步优化和标志物组合的不断改进,cfDNA 多组学分析的临床应用价值将逐步提高。

## 4 多组学模型的构建、算法和验证

### 4.1 多组学模型构建流程

构建多组学模型的前提是建立起一个针对 HNSCC 的特异性标志物的筛选标准,旨在从基因组、表观组学等多组学维度中筛选出与病因驱动、HPV 状态及临床结局密切相关的高价值特征。在此基础上,模型的构建流程通常包含以下环节:首先,要对数据进行预处理,包括对原始数据的清洗和质控,且正确处理数据的配对关系<sup>[37]</sup>。预处理的方法也应适应不同测序方式,以便于下游应用,同时不能过度增加计算负担。其次,在多组学数据中提取各组的关键特征,例如基因组中的基因突变频率、表观组中的 DNA 甲基化、分子组中的 cfDNA 片段特征等。为了简化分析流程,减少计算的复杂度,应使用特征选择或相关降维技术,例如最小绝对值收敛与选择算子(least absolute shrinkage and selection operator, LASSO)回归、随机森林重要性分析、最大相关最小冗余(maximum relevance minimum redundancy, mRMR)等,用于筛选最能区分不同类别的特征组合。最后,根据筛选出的数据特征选择合适的机器学习和深度学习算法,如随机森林、SVM、梯度提升树(gradient boosting decision trees, GBDT)、神经网络(neural network)等。此外还可以使用超参数优化和早停策略来训练和优化模型。

### 4.2 关键算法和技术

在多组学模型的构建过程中,机器学习和深度学习算法在其中发挥了重要作用。在机器学习算法方面,随机森林算法适用于处理高维数据和非线性关系<sup>[38]</sup>。SVM 算法可通过最大间隔超平面和支持向量选择,在小样本数据中表现优异,并可通过调整惩罚参数部分缓解类别不平衡问题<sup>[39]</sup>。梯度提升树和在其基础上改善的 XGBoost 和 LightGBM 算法,在处理大规模数据时可使处理效率获得显著提升<sup>[40-41]</sup>。而在深度学习方面,其具有的自动特征学习能力和处理非线性关系的能力,能够发现数据中的复杂结构和关联,适用于多组学数据的融合分

析<sup>[42]</sup>。此外,使用合成少数类过采样技术(synthetic minority over-sampling technique, SMOTE),可以解决数据不平衡问题<sup>[43]</sup>。对于批次效应和技术噪音等无法避免的问题,也可通过标准化技术和主成分分析(principal component analysis, PCA)来解决。

#### 4.3 模型验证及其性能评估

模型验证是研究型论文的关键环节,它直接决定了研究结果的可靠性和科学性。通过对构建的模型进行验证,可以评估其在实际应用中的表现,并确保研究结论具有严谨性和可信度。模型验证主要通过10折交叉验证(10-fold cross-validation)、留一法(leave-one-out cross-validation, LOOCV)、自助法(bootstrap)等方法,在选择最佳模型以及检测模型的泛化能力方面具有各自的优势和适用场景<sup>[44]</sup>。而性能评估指标则包括敏感性、特异性、阳性预测值、阴性预测值、准确率、F1分数(F1 score)、受试者工作特征曲线下面积(area under curve, AUC)和PR曲线(关注高召回率下的表现)等。也可以将模型性能与传统的检测方式进行比较,以体现多组学模型的优势<sup>[45]</sup>。

#### 4.4 代表性研究案例

目前已有相关已发表的研究案例,可体现出多组学模型的应用价值。以CancerSEEK为例,其整合了cfDNA的突变程度和蛋白质组学的数据,在检测卵巢、肝、胃、胰腺、食管、结直肠、肺和乳腺癌的早期阶段表现出较高的敏感度和99%的特异度,并且可以通过机器学习算法来预测癌症的组织来源<sup>[46]</sup>。另有相关研究利用cfDNA的片段长度、末端基序、核小体足迹、拷贝数变异这4类不同组学的特征,构建了可以运用于胰腺癌早期诊断和预后检测的模型<sup>[47]</sup>。此外还有基于cfDNA的DELFI(DNA evaluation of fragments for early interception)技术和ichorCNA算法。但上述模型和算法并未基于HNSCC进行开发,故其在HNSCC研究中的适用性尚未得到证实,仍需要针对HNSCC的大规模队列来验证模型的适用性和准确性。

### 5 临床应用场景、挑战和未来方向

#### 5.1 临床应用场景

基于cfDNA多组学模型的早期诊断技术在HNSCC的临床应用中有着广泛的应用场景。在HNSCC发病的高危人群中,吸烟、饮酒和HPV感染者均可通过cfDNA检测来捕捉早期癌症发生发展的信号。cfDNA提供了一种非侵入性、精准度高的筛查方法,

通过检测肿瘤相关的基因突变、DNA甲基化和cfDNA的片段特征来发现早期病变。对于吸烟、饮酒和HPV感染者,定期检测cfDNA中的特异性分子标志物,可早期发现病变信号,从而降低疾病进展风险<sup>[48]</sup>。对有着癌症家族遗传史的个体,cfDNA可以实现动态监测,早期发现疾病风险,有助于在癌症的早期阶段进行干预。在临床诊断方面,cfDNA可以有效辅助影像学不确定的病例进行临床决策。相关报道显示,1例正电子发射计算机断层显像(positron emission tomography-computed tomography, PET-CT)假阳性患者,其cfDNA检测为阴性,避免了本不需要的手术<sup>[49]</sup>。通过分析cfDNA中的肿瘤相关突变及甲基化特征,能够提供比传统组织活检更全面的分子信息,同时避免传统活检的侵入性操作和潜在并发症;此外,cfDNA中的肿瘤相关突变及甲基化特征也可以用于检测和定位肿瘤来源,对于原发灶不明的癌症(如颈部淋巴结转移癌)具有潜在的应用价值,可以指导临床治疗决策<sup>[46]</sup>。cfDNA检测在治疗后的微小残留病灶(minimal residual disease, MRD)检测和治疗效果的动态评估中也有着重要价值。相关文献指出,cfDNA水平的动态变化可以作为评估治疗效果和预测复发风险的重要指标<sup>[50]</sup>。通过动态监测cfDNA的变化,可以为临床医生调整治疗策略提供相关依据,例如增加辅助化疗或免疫治疗。

#### 5.2 目前面临的主要挑战

尽管目前cfDNA技术在HNSCC的早期诊断中有着巨大的潜力,但在临床实际应用上却面临着多方面的挑战。技术层面上,在癌症的早期阶段,cfDNA的丰度极低,常低于0.1%,这对目前检验技术的灵敏度提出了极高的要求<sup>[51]</sup>。另外,正常组织释放的DNA和克隆性造血突变造成的高背景噪音可能干扰分析结果,增加了假阳性风险<sup>[52]</sup>。构建模型所需的样本量一定是巨大的,达到标准的早期癌症患者的血液样本量可能不足以满足测序要求,尤其是在全基因组测序(whole genome sequencing, WGS)或全基因组甲基化测序(whole genome bisulfite sequencing, WGBS)的情况下。并且,目前cfDNA检测技术的成本仍居高不下,限制了该技术在大规模筛查中的应用<sup>[53]</sup>。在cfDNA技术进一步推广中,标准化是其中一个重要且无法回避的问题。cfDNA检测中的变量,如样本采集、处理、建库、测序和生信分析流程均缺乏统一标准,妨碍了跨机构之间的对比,并影响结果的可重复性<sup>[54]</sup>。HNSCC发生的复杂解剖结

构、各式各样的病因和驱动癌症发生的分子变化多样性,导致了 HNSCC 的显著异质性,使得模型开发的复杂性显著增高<sup>[30]</sup>。所以在开发 HNSCC 的模型与其中亚种的特异性模型之间需要进行权衡。同时区分惰性病变(如低度异型增生、癌前病变)和具有侵袭性的癌症也是临床诊断中的一个难题<sup>[55]</sup>。在数据分析层面,如何整合分析 cfDNA 结合突变、甲基化、片段特征等多组学数据,并实现临床解读是一个更加复杂的问题<sup>[56]</sup>。开发高质量的模型则需要大量高质量的数据用于训练,与此同时,构建出的模型也需要通过多中心、前瞻性队列研究进行验证。

### 5.3 发展方向与展望

未来,cfDNA 技术的发展方向将逐渐趋于提高检测的特异度和灵敏度、降低检测成本以及优化标志物组合。如单分子测序、纳米孔测序和新型富集技术等技术创新的应用,将进一步提高检测的特异度和灵敏度。随着人工智能技术的发展,可以不断优化模型的性能,提高诊断的准确率。例如通过结合 Transformer 架构,可轻松实现癌症患者的动态风险预测<sup>[57]</sup>。而通过检测流程的高标准化、真实数据

和证据的累积以及卫生经济学的分析和评价,可以使 cfDNA 的检测成本得到控制,在此基础上使该技术实现普及,从而提高其在基层医疗中的应用率和可操作性。相关研究表明,利用 DELFI 可获得高质量的片段组学信息进行癌症筛查,在保证性能的同时可以显著降低相关检测成本<sup>[58]</sup>。未来研究还需要关注 cfDNA 模型在 HNSCC 不同亚型中的诊断性能差异。目前已有初步数据研究显示 cfDNA 的动态变化在 HPV 阴性和阳性患者中均存在着预测价值,但由于样本量受限,尚未进行亚组分析<sup>[59]</sup>。短期内,需要构建大规模 HNSCC 特异性多组学数据库,并借鉴相关多维特征提取方法,开发兼顾成本和性能的靶向检测工具<sup>[60]</sup>。长期目标则是推动前瞻性多中心临床验证,并整合模型,推广至基层医疗体系。随着 cfDNA 技术在基层医疗体系内的铺开,累积了现实世界中的大量数据,可以用于 cfDNA 检测的公共卫生价值和经济效益方面的评估,给相关政策的制定提供证据。最终将实现从早期筛查到治疗全流程的 cfDNA 检测,以全面提高 HNSCC 患者的诊疗水平。

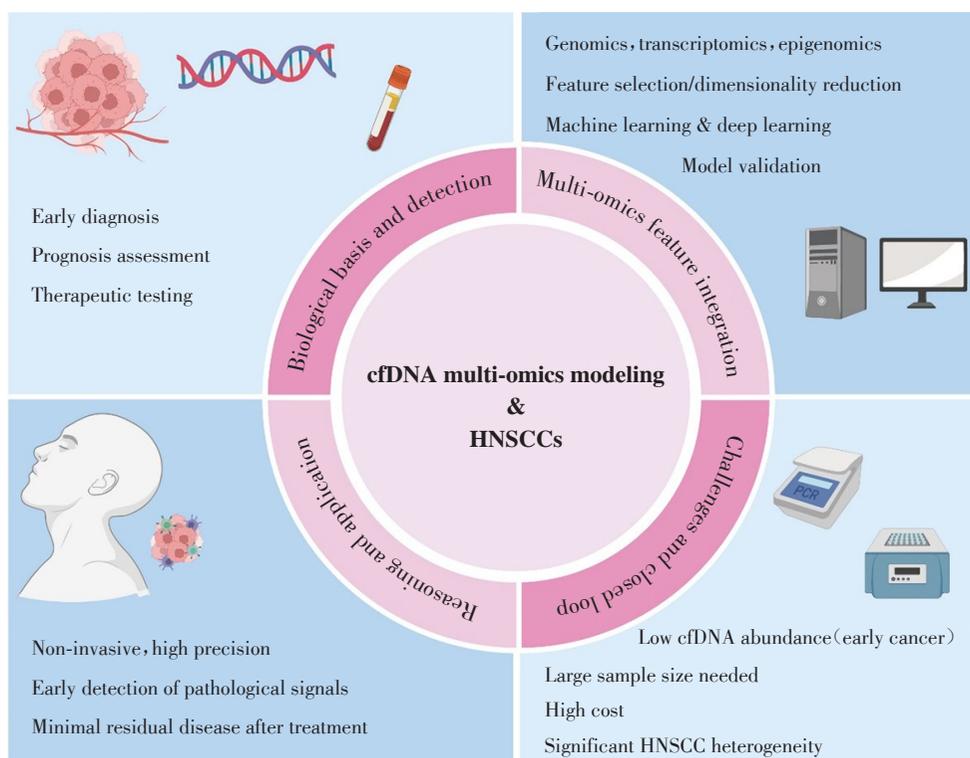


图1 核心概念框架

Figure 1 Core concept framework diagram

## 6 结论

cfDNA 多组学整合技术是解决 HNSCC 早期诊

断瓶颈的一种极具前景的策略,其非侵入性,高灵敏度和特异度等使其可以在一定程度上克服肿瘤异质性的问题。cfDNA 可以反映患者全身的肿瘤

信息,并可以动态监测肿瘤变化,进而实时追踪患者的病情变化,从而实现精准医疗。相比传统检测方法和单组学分析,cfDNA多组学模型通过整合突变基因、cfDNA甲基化和cfDNA片段特征等多维数据,显著提高了早期诊断的准确性。目前,基于cfDNA多组学技术虽在实验室中得到大量验证,构建了可以实际应用的模型,但在向临床转化过程中仍存在不少挑战。在达成样本处理的标准化、提高检测技术的灵敏度和完成基层医疗市场普及之后,将实现从早期筛查到治疗的全过程闭环管理,使HNSCC患者的诊疗水平获得革命性提升。cfDNA技术的发展还需要加强跨学科合作和加大资源投入,需要肿瘤学、生物信息学、分子生物学、检验医学、流行病学等多个学科的专家紧密协作,共同致力于技术优化、数据整合以及临床验证。同时,相关政策的制定和公共卫生部门的加入对于推动cfDNA检测的普及至关重要。通过多方资源的整合与协同努力,cfDNA多组学整合技术有望成为HNSCC早期诊断的重要工具,可以显著提高患者的生存率,为癌症早期诊断领域提供重要的参考和借鉴;也将推动液体活检领域的整体进步,以实现液体活检技术应用于癌症从早期筛查到精准治疗的全过程。

#### 利益冲突声明:

所有作者声明无利益冲突。

#### Conflict of Interests:

All authors declare that there is no conflict of interests.

#### 作者贡献声明:

杨明哲负责文献检索、资料整理以及初稿撰写;陈仁杰负责论文审阅和指导。

#### Author's Contributions:

YANG Mingzhe was responsible for literature retrieval, data organization and first draft writing; CHEN Renjie was responsible for paper review and guidance.

#### [参考文献]

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. *CA Cancer J Clin*, 2021, 71(3):209-249
- [2] PYNNONEN M A, GILLESPIE M B, ROMAN B, et al. Clinical practice guideline: evaluation of the neck mass in adults [J]. *Otolaryngol Head Neck Surg*, 2017, 157(2\_suppl):S1-S30
- [3] WOODHOUSE R, LI M J, HUGHES J, et al. Clinical and analytical validation of FoundationOne Liquid CDx, a novel 324-gene cfDNA-based comprehensive genomic profiling assay for cancers of solid tumor origin [J]. *PLoS One*, 2020, 15(9):e0237802
- [4] MANDEL P, METAIS P. Les acides nucléiques du plasma sanguin chez l'homme [J]. *Nature*, 1948, 162(4121):293-294
- [5] WAN J C M, MASSIE C, GARCIA-CORBACHO J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA [J]. *Nat Rev Cancer*, 2017, 17(4):223-238
- [6] SNYDER M W, KIRCHER M, HILL A J, et al. Cell-free DNA comprises an *in vivo* nucleosome footprint that informs its tissues-of-origin [J]. *Cell*, 2016, 164(1/2):57-68
- [7] MOULIERE F, CHANDRANANDA D, PISKORZ A M, et al. Enhanced detection of circulating tumor DNA by fragment size analysis [J]. *Sci Transl Med*, 2018, 10(466):eaat4921
- [8] LEON S A, SHAPIRO B, SKLAROFF D M, et al. Free DNA in the serum of cancer patients and the effect of therapy [J]. *Cancer Res*, 1977, 37(3):646-650
- [9] CHAN K C, JIANG P Y, CHAN C W M, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing [J]. *Proc Natl Acad Sci USA*, 2013, 110(47):18761-18768
- [10] FAN H C, BLUMENFELD Y J, CHITKARA U, et al. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood [J]. *Proc Natl Acad Sci USA*, 2008, 105(42):16266-16271
- [11] DAWSON S J, TSUI D W Y, MURTAZA M, et al. Analysis of circulating tumor DNA to monitor metastatic breast cancer [J]. *N Engl J Med*, 2013, 368(13):1199-1209
- [12] SCHWARZENBACH H, HOON D S B, PANTEL K. Cell-free nucleic acids as biomarkers in cancer patients [J]. *Nat Rev Cancer*, 2011, 11(6):426-437
- [13] VOLIK S, ALCAIDE M, MORIN R D, et al. Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies [J]. *Mol Cancer Res*, 2016, 14(10):898-908
- [14] CHENG T H T, JIANG P Y, TEOH J Y C, et al. Noninvasive detection of bladder cancer by shallow-depth genome-wide bisulfite sequencing of urinary cell-free DNA for methylation and copy number profiling [J]. *Clin Chem*, 2019, 65(7):927-936
- [15] VAVOULIS D V, CUTTS A, THOTA N, et al. Multimodal cell-free DNA whole-genome TAPS is sensitive and reveals specific cancer signals [J]. *Nat Commun*, 2025, 16

- (1): 430
- [16] CARVALHO A L, HENRIQUE R, JERONIMO C, et al. Detection of promoter hypermethylation in salivary rinses as a biomarker for head and neck squamous cell carcinoma surveillance[J]. *Clin Cancer Res*, 2011, 17(14): 4782-4789
- [17] CHEN K Z, ZHANG J B, GUAN T, et al. Comparison of plasma to tissue DNA mutations in surgical patients with non-small cell lung cancer[J]. *J Thorac Cardiovasc Surg*, 2017, 154(3): 1123-1131
- [18] AGRAWAL N, FREDERICK M J, PICKERING C R, et al. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1[J]. *Science*, 2011, 333(6046): 1154-1157
- [19] RAJA R, KUZIORA M, BROHAWN P Z, et al. Early reduction in ctDNA predicts survival in patients with lung and bladder cancer treated with durvalumab [J]. *Clin Cancer Res*, 2018, 24(24): 6212-6222
- [20] CABEL L, RIVA F, SERVOIS V, et al. Circulating tumor DNA changes for early monitoring of anti-PD1 immunotherapy: a proof-of-concept study[J]. *Ann Oncol*, 2017, 28(8): 1996-2001
- [21] HASIN Y, SELDIN M, LUSIS A. Multi-omics approaches to disease[J]. *Genome Biol*, 2017, 18(1): 83
- [22] HUANG S J, CHAUDHARY K, GARMIRE L X. More is better: recent progress in multi-omics data integration methods[J]. *Front Genet*, 2017, 8: 84
- [23] KARCZEWSKI K J, SNYDER M P. Integrative omics for health and disease[J]. *Nat Rev Genet*, 2018, 19(5): 299-310
- [24] RITCHIE M D, HOLZINGER E R, LI R W, et al. Methods of integrating data to uncover genotype-phenotype interactions[J]. *Nat Rev Genet*, 2015, 16(2): 85-97
- [25] BOYLE A P, HONG E L, HARIHARAN M, et al. Annotation of functional variation in personal genomes using RegulomeDB[J]. *Genome Res*, 2012, 22(9): 1790-1797
- [26] HUTTER C, ZENKLUSEN J C. The cancer genome atlas: creating lasting value beyond its data[J]. *Cell*, 2018, 173(2): 283-285
- [27] RAIHAN F R, ABDULRAZAK L F, RAHMAN M A, et al. StackTHP: a stacking ensemble model for accurate prediction of tumor-homing peptides in cancer therapy[J]. *Comput Biol Med*, 2025, 189: 109958
- [28] STRANSKY N, EGLOFF A M, TWARD A D, et al. The mutational landscape of head and neck squamous cell carcinoma[J]. *Science*, 2011, 333(6046): 1157-1160
- [29] SEIWERT T Y, ZUO Z X, KECK M K, et al. Integrative and comparative genomic analysis of HPV-positive and HPV-negative head and neck squamous cell carcinomas [J]. *Clin Cancer Res*, 2015, 21(3): 632-641
- [30] LECHNER M, SCHARTINGER V H, STEELE C D, et al. Somatostatin receptor 2 expression in nasopharyngeal cancer is induced by Epstein Barr virus infection: impact on prognosis, imaging and therapy [J]. *Nat Commun*, 2021, 12(1): 117
- [31] LEEMANS C R, SNIJDERS P J F, BRAKENHOFF R H. The molecular landscape of head and neck cancer [J]. *Nat Rev Cancer*, 2018, 18(5): 269-282
- [32] GILLISON M L, KOCH W M, CAPONE R B, et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers [J]. *J Natl Cancer Inst*, 2000, 92(9): 709-720
- [33] FERRIS R L, BLUMENSCHIEIN G Jr, FAYETTE J, et al. Nivolumab for recurrent squamous-cell carcinoma of the head and neck[J]. *N Engl J Med*, 2016, 375(19): 1856-1867
- [34] CHAUDHARY K, POIRION O B, LU L Q, et al. Deep learning-based multi-omics integration robustly predicts survival in liver cancer [J]. *Clin Cancer Res*, 2018, 24(6): 1248-1259
- [35] BARABÁSI A L, GULBAHCE N, LOSCALZO J. Network medicine: a network-based approach to human disease[J]. *Nat Rev Genet*, 2010, 12(1): 56-68
- [36] LIU X H, WANG G R, ZHONG N N, et al. Multi-omics in immunotherapy research for HNSCC: present situation and future perspectives [J]. *NPJ Precis Oncol*, 2025, 9(1): 93
- [37] BOLGER A M, LOHSE M, USADEL B. Trimmomatic: a flexible trimmer for illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120
- [38] HO T K. Random decision forests[C]//Proceedings of 3rd international conference on document analysis and recognition. Washinton DC: IEEE, 2002: 278-282
- [39] CORTES C, VAPNIK V. Support - vector networks [J]. *Mach Learn*, 1995, 20(3): 273-297
- [40] CHEN T Q, GUESTIN C. XGBoost: a scalable tree boosting system [C]//Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. New York: ACM, 2016: 785-794
- [41] KE G L, MENG Q, FINLEY T, et al. LightGBM: a highly efficient gradient boosting decision tree [C]//Proceedings of the 31st international conference on neural information processing systems. New York: Curran Associates Inc., 2017: 6000-6010
- [42] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. *Nature*, 2015, 521(7553): 436-444

- [43] CHAWLA N V, BOWYER K W, HALL L O, et al. SMOTE: synthetic minority over-sampling technique[J]. *Jair*, 2002, 16: 321–357
- [44] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection[C]//Proceedings of the 14th international joint conference on artificial intelligence. San Francisco: Morgan Kaufmann Publishers Inc., 1995: 1137–1143
- [45] CHEN G, ZHANG J, FU Q T, et al. Integrative analysis of multi-omics data for liquid biopsy[J]. *Br J Cancer*, 2023, 128(4): 505–518
- [46] COHEN J D, LI L, WANG Y X, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test[J]. *Science*, 2018, 359(6378): 926–930
- [47] WU J M, XU X F, ZHANG Q Z, et al. Cell-free DNA testing for the detection and prognosis prediction of pancreatic cancer[J]. *Nat Commun*, 2025, 16(1): 6645
- [48] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics, 2020[J]. *CA A Cancer J Clinicians*, 2020, 70(1): 7–30
- [49] BOLA S, CUTTS A, VAVOULIS D, et al. Circulating tumour DNA to augment PET-CT in determining clinical outcome after head and neck cancer treatment[J]. *Eur J Cancer*, 2025, 227: 115626
- [50] ABBOSH C, BIRKBAK N J, WILSON G A, et al. Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution[J]. *Nature*, 2017, 545(7655): 446–451
- [51] NEWMAN A M, BRATMAN S V, TO J, et al. An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage[J]. *Nat Med*, 2014, 20(5): 548–554
- [52] JAISWAL S, FONTANILLAS P, FLANNICK J, et al. Age-related clonal hematopoiesis associated with adverse outcomes[J]. *N Engl J Med*, 2014, 371(26): 2488–2498
- [53] HEITZER E, ULZ P, GEIGL J B. Circulating tumor DNA as a liquid biopsy for cancer[J]. *Clin Chem*, 2015, 61(1): 112–123
- [54] MERKER J D, OXNARD G R, COMPTON C, et al. Circulating tumor DNA analysis in patients with cancer: American society of clinical oncology and college of American pathologists joint review[J]. *J Clin Oncol*, 2018, 36(16): 1631–1641
- [55] CALIFANO J, VAN DER RIET P, WESTRA W, et al. Genetic progression model for head and neck cancer: implications for field cancerization [J]. *Cancer Res*, 1996, 56(11): 2488–2492
- [56] SHEN S Y, SINGHANIA R, FEHRINGER G, et al. Sensitive tumour detection and classification using plasma cell-free DNA methylomes[J]. *Nature*, 2018, 563(7732): 579–583
- [57] ADRIAN TSUI W H, JIANG P Y, DENNIS LO Y M. Cell-free DNA fragmentomics in cancer [J]. *Cancer Cell*, 2025, 43(10): 1792–1814
- [58] CRISTIANO S, LEAL A, PHALLEN J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer[J]. *Nature*, 2019, 570(7761): 385–389
- [59] RUIZ-TORRES D A, MERKIN R D, BRYAN M E, et al. Personalized circulating tumor DNA dynamics inform survival and response to immune checkpoint blockade in recurrent/metastatic head and neck cancer[J]. *NPJ Precis Oncol*, 2025, 9(1): 298
- [60] BAO H, YANG S S, CHEN X X, et al. Early detection of multiple cancer types using multidimensional cell-free DNA fragmentomics [J]. *Nat Med*, 2025, 31(8): 2737–2745

(收稿:2025-09-22;修回:2025-10-22;录用:2025-11-10)

(本文编辑:蒋 莉)