·肿瘤研究与诊疗前沿交叉技术·
DOI:10.16262/j.cnki.1000-8217.20250227.002

从数据到机制:医数交叉驱动肿瘤精准诊疗的研究现状与展望*

孙端辰[1,4] 韩仁敏[2] 马成林[2] 段贸腾[1] 王 斐[3,4**] 刘丙强[1**]

1. 山东大学 数学学院,济南 250100
2. 山东大学 数学与交叉科学研究中心,青岛 266237
3. 山东大学 第二医院,济南 250033
4. 山东省癌症数字医疗重点实验室,济南 250033

[摘 要] 测序组学、电子健康记录和医学影像等多维度异质医学数据的迅速积累,不仅凸显了传统还原论研究范式的局限性,同时也为医学研究革新带来机遇。近年来,生物医学与智能信息科学的深度交叉融合取得显著进展,推动了疾病预测与精准医疗的发展进程,数学已逐步成为其底层核心驱动。通过深化医学—数学交叉研究实现对生命系统本质规律的定量解析将成为本领域取得变革性突破的关键路径。本文系统综述了医数交叉领域的研究进展,重点探讨数学模型在肿瘤诊断、治疗及肿瘤发生发展机制解析等方面的关键作用,深入展望医数交叉在机制导向的数学模型构建、数字生命和虚拟健康等领域的创新潜力与应用前景。通过数学模型的精准构建与应用,实现从"数据关联"向"机制解析"的迈进,医数交叉将为肿瘤预防和诊疗提供突破性解决方案,推动医学的高效化、精准化、智能化变革。

[关键词] 医数交叉;数据分析;数学模型;生物信息学;生物数学;肿瘤诊疗

医学直接关乎人类的生命健康,是充满人文关怀的"前沿阵地"。其进步历程充分体现了多学科交叉融合的深远影响。例如,18 世纪化学的突破为医学注入新动力,约瑟夫·普里斯特利发现氧气,为现代呼吸学奠定了科学基础;19 世纪中期,物理学与医学结合催生了现代影像学,彻底革新了医学诊断技术;19 世纪末,心理学的兴起为精神病学提供理论支撑,西格蒙德·弗洛伊德的精神分析学说更是深刻影响了现代心理治疗的发展。这些跨学科的融合不仅推动了医学理论与技术的不断演变,更彰显了学科交叉在医学发展中的核心作用。

基础研究是科学之本和创新之源,是所有技术问题的总机关。强大的基础科学研究是建设科技强国的基石,更是提升原始创新能力的根本途径。数学是所有自然科学的基础,是科技创新的底层驱动力。数学与医学研究和实践的紧密结合,展现了其严谨性和解析性在解决生物医学复杂问题中的独特优势。通过时空动态、因果性、临界性的定性和定量分析方法,数学为理解生物医学现象提供了强有力的工具[1]。例如,研究人员通过建立数学模型模拟生物系统的动态行为,深入揭示了肿瘤发生与发展的内在机制[2]。

收稿日期:2024-11-30;修回日期:2025-02-21
* 本文根据国家自然科学基金委员会第 373 期"双清论坛"讨论的内容整理。
** 通信作者,Email: fei.wang@sdu.edu.cn; bingqiang@sdu.edu.cn
本文受到国家重点研发计划(2020YFA0712400,2022YFA1004801)和国家自然科学基金项目(62272270,62202269)的资助。

引用格式:孙端辰,韩仁敏,马成林. 从数据到机制:医数交叉驱动肿瘤精准诊疗的研究现状与展望. 中国科学基金,2025,39(1):174—184.
Sun DC, Han RM, Ma CL. Current status and prospects of digital-driven interdisciplinary research in biomedicine and mathematics. Bulletin of National Natural Science Foundation of China, 2025, 39(1): 174-184. (in Chinese)

当前，生物医学技术的飞速革新正加速医学的数字化和精准化进程，催生了多层次、多尺度生物医学数据的爆炸式增长[3]。从人工智能到大数据分析，再到生物医药的数字化转型，数据驱动的医疗体系正在逐步取代传统的经验性诊疗模式。然而，尽管海量数据为肿瘤研究和个性化医疗提供了前所未有的资源，如何从中高效、准确地挖掘有效信息仍面临诸多挑战[4]。这些挑战不仅源于测量噪声以及数据在维度、尺度和结构上的差异[5-7]，更与生命系统本身的系统性、高度组织性、动态性和随机性密切相关。许多生物学现象和机制尚缺乏统一的概念模型来描述其原理，这使得目标导向的建模方法难以揭示生命系统紊乱的底层机理，所构建的因果链条往往模糊且不完整。在这样的时代背景下，医学研究亟需新的研究范式。借助数学模型、算法和理论应用于生物医学领域（医数交叉），有望在癌症基因组学和数据驱动的精准治疗等场景实现突破性进展。与此同时，医学实践也为数学学科提供丰富的应用场景与试验平台，成为推动数学理论创新的“催化剂”。两者相辅相成、共同进步。肿瘤作为复杂疾病的典型代表，其发生与发展充分体现了生命内在机制的系统性、动态性和随机性。因此，肿瘤的诊疗与机制研究将成为医数交叉刻画复杂疾病的理想试验田，并成为推动医学变革的重要阵地。

肿瘤学中医数交叉研究的核心在于综合利用多元数学理论，深入刻画疾病发生发展内在生命原理所蕴含的系统性、高度组织性、动态性和随机性（图 1）。优化理论、组合数学、图与网络、微分方程和概率统计等数学工具，为建模和分析复杂生命机制提供了形式化手段、模型化工具和科学化语言。只有当这些数学理论模型与生命系统运作高度吻合时，才能有效捕捉和描述生命过程中的系统性和复杂性。例如，生命体通过亿万年的进化，自然地遵循着一套内在优化原则，如“最大化利用”“最小能量消耗”和“最稳定状态”等。这些原则贯穿于能量代谢、酶催化反应、细胞分裂和蛋白质折叠等生物过程的各个层面。此外，药物代谢、肿瘤生长和免疫反应等生物医学现象均具有时变特征，涉及离散对象之间的连续动态非线性关系，而微分方程则为描述这种关系提供了精确的数学框架，能够准确反映生物系统的演化规律。与新兴的大语言模型或人工智能（Artificial Intelligence，AI）技术相比，数学方法在揭示复杂生物医学问题的内在逻辑和底层机理方面独具优势。AI 技术虽在特定任务中表现出色，但其基于概率统计的框架在系统深层次机理的解释性方面存在局限。实际上，数学方法与 AI 技术并非对立，而是相辅相成。AI 技术优势在于处理高维、非结构化数据（如医学影像），可以作为数学建模的有力补充，而数学理论则为 AI 模型提供了可解释性的理论基础。两者的有机结合将推动医数交叉研究向更深层次发展，实现从生命现象描述到机制解析的跨越。

本文以肿瘤精准诊疗和机制研究为例，全面综述了医数交叉领域的研究进展，系统阐述了数学理论、模型和算法在其中的关键作用。同时，本文展望了医数交叉在机制导向的数学模型构建、数字生命和虚拟健康等领域的创新潜力与应用前景。我们期待，通过数学理论的精准运用、数学模型的系统构建以及数学算法的开发应用，医数交叉能够为以肿瘤为代表的复杂疾病精准预防、早期诊断和高效治疗提供切实可行的解决方案，进而推动现代医学向高效化、精准化与智能化的方向实现变革性发展。

1　医数交叉研究现状

在健康管理和疾病防控的全生命周期中，数学理论与方法作为核心支撑工具，贯穿于未病先防、早期检测和诊断治疗等各个环节，发挥着不可替代的作用。在疾病预防阶段，基于系统动力学的临界理论分析为个体患病风险预测提供了科学依据，使早期预防干预更具针对性和时效性；在诊断治疗环节，数学方法通过对高维复杂数据的深度挖掘，有效提取肿瘤精准特征，显著提升诊断准确性，同时基于患者个体化特征提供最优治疗方案；在肿瘤发生发展机制研究方面，数学模型为驱动突变基因和致病基因识别提供了可靠的技术支撑，不仅有助于解析复杂疾病的转录调控机理，更能深入揭示细胞间通讯的关键特征。

1.1　肿瘤预防与早期诊断

数学理论与方法在肿瘤诊断的各环节中均发挥着重要作用，尤其在数据驱动的疾病临界点预测、生物标志物识别以及医学影像数据重建等领域展现出显著优势。

1.1.1　肿瘤演进临界点预测

肿瘤的发生是一个高度复杂的生物学过程，其诱因和机制因肿瘤类型而异，导致预防措施难以实现普适性[8]。目前，尽管多种肿瘤筛查技术已在临床实践中广泛应用，但仍存在显著的局限性。以肿瘤标志物检测为例，尽管其作为一种常用的筛查手

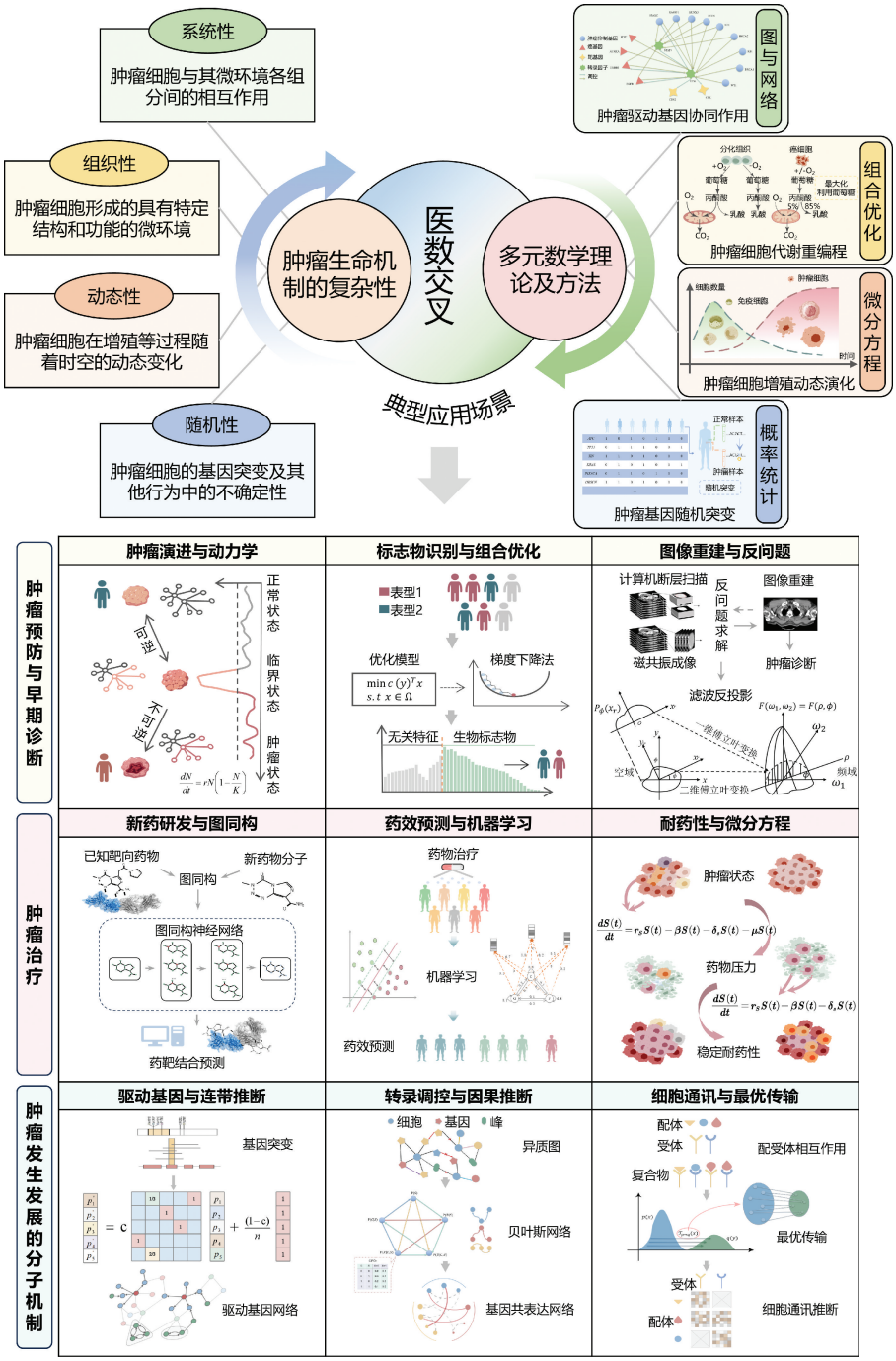

图 1　肿瘤生命机制的复杂性、潜在的数学交叉领域以及医数交叉在肿瘤精准诊疗中的典型应用场景

Fig. 1　The Complexity of Tumor Biology, Potential Interdisciplinary Fields in Mathematics, and Typical Application Scenarios of Medical-Mathematical Integration in Precision Diagnosis and Treatment

段，但部分标志物的特异性和敏感性不足，限制了其在早期诊断中的准确性；影像学检查虽能有效识别部分早期肿瘤，但对微小病灶的检测能力仍面临挑战，同时频繁的检查不仅增加了医疗成本，还可能带来潜在的辐射暴露风险。

生命系统具有高度复杂性和时空动态变化特征，复杂疾病的发生发展过程普遍呈现出非线性临界现象。因此，基于动力学的数据科学作为一门新兴的数学交叉理论应运而生，其核心在于将生命系统转化为动力系统，从而精准捕捉不同疾病状态间的临界转变过程。在肿瘤研究领域，通过刻画系统的动态行为及其演化规律，能够为复杂的肿瘤过程构建临界理论框架，赋予肿瘤发生发展过程以“可解释性”“可预测性”和“可拓展性”。这一理论不仅为

深入理解肿瘤的内在机制和行为特征提供了科学依据，更实现了肿瘤演进临界点的科学量化与精准预测。

几个典型例子包括：Liu 等人利用非线性动力学理论识别与疾病进展临界点相关的动态特征，成功检测到疾病早期的若干预警信号[9]。该方法在肺腺癌研究中取得了重要突破，通过整合两个临床队列数据，研究团队发现肺腺癌的转移临界点位于 IIB 期，且在该临界点之后患者的肿瘤转移风险显著增高。基于这一理论框架，Hong 等人进一步提出了相对熵模型，用于检测疾病进展过程中状态剧烈转变的预警信号[10]。该模型在结直肠癌、肺腺癌、甲状腺癌和肾细胞癌等研究中展现出强大的预测能力，其识别的信号网络在特定阶段表现出明显变化，为多种疾病的演化进程提供了可靠的预测依据。此外，Chen 等人开发了一种基于无监督隐马尔可夫模型的计算方法，通过提取可判别、可解释的特征，揭示了肿瘤演进接近临界点时动力学过程的潜在机制[11]。这一方法在乳腺癌研究中得到验证，成功识别了 *CEBPA*、*SMAD3* 和 *GSK3B* 等基因在肿瘤演进临界点时的动态变化，为疾病的早期预防和干预提供了重要的分子依据。

1.1.2　生物标志物识别

识别具有高特异性和高精度的生物标志物在疾病精准诊断、治疗方案优化、预后评估、药物研发和靶点预测等领域具有重要的科学价值和应用意义[12, 13]，已成为生物信息学研究的核心热点之一[14, 15]。

从数学视角来看，生物标志物识别问题本质上可归结为从特征集合中筛选最优特征子集的组合优化问题。这一问题的核心在于在离散的组合空间中寻找满足特定约束条件的最优解。然而，其庞大的解空间规模使得传统的穷举搜索方法不切实际。因此，研究者通常采用启发式算法或近似算法，如贪婪算法、模拟退火算法和遗传算法等。这些数学方法不仅为生物标志物识别提供了高效的计算工具，还显著提升了精准医疗中特征筛选的效率与模型构建的准确性，为疾病诊断和治疗策略的优化奠定了重要基础[16]。

在实际建模研究中，Chuang 等人创新性地将蛋白质相互作用网络应用于乳腺癌远处转移预测，将生物标志物识别问题转化为最优子图的识别问题，并利用贪婪算法成功筛选出具有高分值和高统计显著性的连通子网络作为潜在生物标志物[17]。该方法在乳腺癌数据中表现优异，分别鉴定出 149 个和 243 个子网络。与传统的基于单个基因表达分析的模式相比，该方法从系统生物学角度出发，充分考虑了功能网络中基因间相互作用的整体状态，从而更全面地揭示了所识别生物标志物的功能特性。Liu 等人则提出了针对每个疾病样本构建特异性网络的方法，通过差异网络分析成功识别了不同癌症个体间的差异基因和网络模式[18, 19]。此外，基于机器学习的标志物识别方法在肿瘤诊断中发展迅速。例如，Steyaert 等人开发了一个整合多组学数据、临床数据和医学影像的深度学习框架[20]，通过监督学习有效识别出与癌症发展或预后相关的生物标志物，为肿瘤精准诊断提供了新的技术手段。

1.1.3　医学图像影像辅助诊断

医学影像技术是肿瘤辅助诊断的重要工具，其典型代表包括计算机断层扫描(Computed Tomography，CT)、磁共振成像(Magnetic Resonance Imaging，MRI)以及病理图像等。CT 和 MRI 均利用特定的物理原理对人体内部进行成像，而病理图像则是通过高精度扫描仪将传统的玻片病理切片数字化，实现病理信息的可视化与定量分析。

在基于医学图像影像的辅助诊断中，肿瘤的自动分割与标注是实现精准诊断的关键任务。尽管手动分割长期以来被视为描绘解剖结构和病理区域的黄金标准，但其过程不仅耗时耗力，还高度依赖于专业人员的经验。相比之下，半自动或全自动分割方法能够显著提高效率，降低人力成本，并支持大规模数据集的分析。近年来，随着深度学习技术的快速发展，大规模预训练模型(大模型)的出现为全自动分割方法的实现提供了新的可能。这些模型的成功在很大程度上得益于其蕴含的数学基础：线性代数为模型的表示和运算提供了框架，优化理论指导了模型的训练过程，而信息论则为模型性能的评估提供了理论依据。

在影像诊断中，一个关键问题是如何通过优化图像重构算法，在尽可能减少辐射对人体伤害的同时获取高质量的图像数据。CT 和 MRI 成像本质上属于反问题求解，即从投影测量数据中反向推断待测物体的衰减系数分布函数，从而生成人体截面图像。为了降低辐射暴露风险，CT 扫描通常采用低剂量和减少测量视角的策略，而 MRI 则通过减少空间采样量来缩短成像时间。然而，这些策略不可避免地会降低成像的分辨率和信噪比，进而影响图像质量。因此，从数学模型与计算精度的角度出发，开发

高效的图像重建方法成为提升成像效果的核心任务。滤波反投影(Filtered Back Projection, FBP)算法是图像重建中反问题求解的经典算法之一,其基于投影定理和频域滤波原理,通过对采集的投影数据进行高通滤波,再将其反投影至图像空间以生成重建图像。FDK(Feldkamp Davis Kress)算法[21]作为 FBP 算法的扩展,通过处理斜投影数据,能够有效应对非平行投影数据的重建问题。在 MRI 成像中,由于采样点通常呈非均匀分布,传统的快速傅里叶变换(Fast Fourier Transform, FFT)难以直接应用。NUFFT(Non-uniform FFT)算法[22]通过将非均匀采样数据插值转换为均匀数据,从而利用 FFT 实现高效计算。这些算法显著提升了医学成像中数据处理和图像重建的精确性,能够在更少的扫描数据基础上重建出清晰的人体截面图像,不仅降低了辐射对人体的伤害,还为精准诊疗提供了高质量的影像支持。

1.2 肿瘤治疗

新药研发、药效预测和耐药性研究等领域不仅是生物学与药理学深度融合的体现,更是数学建模与大数据分析等前沿技术的重要应用场景。通过医数交叉研究,数学模型能够为药物设计提供精准的理论依据,优化靶点筛选过程,并预测药物在不同患者群体中的疗效和副作用,从而为个性化治疗的实现提供科学支持。

1.2.1 新药研发与图同构

新药研发在肿瘤治疗中具有重要的科学价值和临床意义[23]。抗肿瘤药物的研发是一个复杂且多维的过程,涵盖了分子结构设计与解析、候选药物筛选以及药物与靶点相互作用等多个关键环节。面对研发过程中诸如靶点识别、药物优化和疗效预测等挑战,亟须借助数学工具与方法提供理论支撑和技术突破。

在新药研发领域,图同构问题发挥着至关重要的作用,其核心在于通过判断两种化合物分子结构或生物活性网络的相似性,捕捉化学分子图中原子与键的拓扑结构一致性[24]。药物分子可以通过分子图进行表示,基于图同构分析,研究人员能够设计新颖的分子结构,确保其具备尚未被充分探索的化学特性,从而拓宽新药研发的可能性。近年来,图同构与深度学习的结合,尤其是图神经网络,已成为药物开发领域的研究热点[25]。通过图同构的应用,分子表征的精准性得到显著提升,不仅有助于准确地预测化合物的药效、毒性和药代动力学性质,还为药物设计与优化提供了强有力的理论支持和技术手段。

举例来说,Bao 等人开发了基于多任务图同构网络的小分子抑制活性预测模型[26]。该模型利用图同构神经网络作用于药物分子图,通过加和聚合函数生成全图表示,实现了分子特征的高效提取。Du 等人提出的 CT-GINDTI 模型采用全局最大池化的图同构网络对药物分子图进行表征,并通过循环训练方法有效解决了样本不平衡问题[27]。此外,Wang 等人基于跨模态相似性,设计了图同构网络的图自动编码器,用于预测药物与特定靶标的相互作用[28]。这些研究表明,图同构网络凭借强大的结构捕捉能力和灵活的任务适配性,在药物分子设计和药物—靶点相互作用预测中展现出独特优势。与传统高成本实验方法相比,图同构网络能够通过精确的分子拓扑分析快速筛选候选分子,显著降低研发成本;同时,其自动学习能力超越了传统统计模型对手工特征的依赖,能够更全面地表征分子间的复杂关系。这些数学方法为加速新药研发、优化候选分子及开发靶向性更强的抗肿瘤药物提供了重要的理论支撑和技术保障。

1.2.2 药效预测

不同患者对相同药物的反应存在显著差异,而肿瘤类型及其分子特征的多样性进一步增加了制定有效治疗方案的难度,因此药效预测在肿瘤治疗中尤为关键[29]。药效预测的核心在于将多源数据中的复杂关联关系转化为可量化的数学模型,从而有效揭示患者分子特征与药物反应之间的潜在规律。

利用机器学习技术构建药效预测模型,已成为应对肿瘤治疗中个体差异性挑战的有效方法。生物医学数据通常具有高维度、高异质性和高噪声等特性,这使得传统方法难以有效建模其中的非线性关系和复杂交互作用。在此背景下,机器学习通过最小化损失函数或最大化似然函数,能够自动捕捉数据中的复杂结构,学习变量间的潜在关系,提取关键特征,并减少冗余噪声的干扰,从而显著提升药效预测的准确性和模型鲁棒性。特别是结合多模态数据时,机器学习模型可跨越单一数据源的限制,综合利用分子特征、临床数据和病理信息,构建更精确的药效预测模型。

在药效预测领域,数据驱动的机器学习方法展现出强大的应用潜力。例如,Sugasawa 等人提出了基于梯度提升树的个体治疗效应估计方法——

SGBT[30]。该方法通过计算每位患者在接受和未接受药物治疗时的预期反应值差异，精准估计药物对每位患者的具体效应。Brindha 等人开发了一种基于支持向量机的抗癌药物效力预测方法——ELM_SVR[31]。该方法整合口腔鳞状细胞癌患者的临床和分子数据，成功预测了五种抗癌药物的效力。此外，在某些罕见亚型肿瘤的研究中，由于样本量有限，模型的训练效果可能受到影响，从而限制了预测的准确性和普适性。未来，若能结合生成对抗网络和对比学习等新兴技术，将有望增强对数据稀缺领域的建模能力，进一步提高模型的可解释性和推广性。

1.2.3　肿瘤耐药动力学模型

耐药性是导致肿瘤治疗失败的主要原因之一。在临床实践中，通常通过监测肿瘤标志物水平升高或依赖 CT、MRI 等影像学数据判断患者是否产生耐药性。然而，耐药性可在肿瘤增殖的任何阶段发生，甚至在诊断和药物治疗前就已存在[32]，这些早期耐药现象往往无法通过传统的影像学检测手段进行量化评价。从生物学机制来看，肿瘤细胞种群的死亡率低于其增长率，以及耐药细胞逐步成为主导细胞群，是耐药性发生的重要基础。因此，构建肿瘤耐药动力学模型不仅有助于深入理解耐药性发展的内在机制，还能为肿瘤演化规律的解析和治疗策略的优化提供新的科学见解。

概率模型是研究肿瘤耐药动力学的常用方法之一。其理论基础在于，细胞生命周期的终结阶段通常有两种可能的结果：细胞死亡或分裂成两个新的细胞。在增殖过程中，敏感细胞会以一定的概率发生突变，产生耐药细胞。敏感细胞和耐药细胞的增殖均符合连续时间分支（生灭）过程。因此，细胞增殖过程常通过分支（生灭）过程来建模。借助概率母函数或特征函数等数学工具，可以求解耐药细胞数量的均值及其概率分布[33]。肿瘤细胞种群规模可以通过影像学等医学手段进行测量，进而通过模型对耐药细胞进行量化评估。例如，Komarova 等人利用概率模型，揭示了肿瘤细胞与耐药细胞之间的数量关系。这一研究为癌症确诊时分析患者的肿瘤细胞构成提供了重要的理论依据，对治疗方案的制定具有重要的指导意义[34]。

除了概率模型，确定性微分方程也是研究肿瘤耐药动力学的重要方法之一。在肿瘤诊断时，细胞数量通常已达到庞大规模，此时细胞种群的演化几乎呈现出确定性特征。微分方程能够从宏观角度描述细胞种群数量随时间的变化趋势，综合考虑环境容许量、药物影响以及不同细胞之间的相互作用等因素，从而刻画敏感细胞与耐药细胞种群的动态演化过程。例如，Liu 等人构建了敏感细胞与耐药细胞之间的微分博弈模型，并提出了带约束的动态最优问题。通过仿真模拟实验和量化分析，验证了其所提出的最优适应性治疗方案在疗效和耐药性控制方面优于现有治疗方案[35]。

1.3　肿瘤发生发展的分子机制

肿瘤的发生与发展是一个多层次、多维度的复杂生物学过程，涉及驱动基因突变、异常基因表达、转录调控异常以及细胞间通讯等多种生物机制的动态相互作用。为全面解析这一复杂过程，数学的定性与定量分析方法成为不可或缺的工具。通过构建数学模型，可以系统探索肿瘤的分子特征，揭示其内在机制的相互作用规律，从而为绘制肿瘤发生发展的“全景图”提供关键理论支撑。

1.3.1　驱动基因识别

癌症的发生通常是体细胞突变逐渐积累的结果。在这一过程中，少数赋予肿瘤细胞显著增殖优势的“驱动基因”突变成为肿瘤发展和扩散的主要推动力。精准识别这些驱动基因不仅对于深入解析肿瘤的生物学机制具有重要科学意义，还为开发针对性的靶向治疗方法提供了关键依据。

在生物信息学和系统生物学领域，“连带推断”（guilt-by-association）方法[36]通过分析生物网络中已知节点的特征来推测邻近未知节点的功能。这一策略在基因功能注释、疾病基因识别和蛋白质功能预测等方面得到广泛应用，为揭示复杂疾病的分子机制和发现新的治疗靶点提供了重要工具。该方法基于一个核心假设：与已知致病基因在表达模式、蛋白质相互作用或遗传连锁等方面密切相关的基因或蛋白，可能参与类似的病理过程。通过应用如 PageRank 等网络扩散算法[37,38]，可以从已验证的致病基因出发，在生物网络中传播信息，从而对与疾病相关的未知基因进行排序和优先级评估。

为了精准识别驱动基因，研究人员基于“连带推断”方法开发了多种先进算法。其中，Leiserson 等人提出的 HotNet2 算法[39]利用热扩散理论，在蛋白质互作网络及生物通路中高效识别罕见突变子网络。通过对癌症基因组图谱（The Cancer Genome Atlas Program，TCGA）进行泛癌分析，HotNet2 成功鉴定了 16 个显著突变的子网络，为癌症类型的诊断和治疗提供了新的科学见解。Bashashati 等人开

发的 DriverNet 算法[40]则通过整合生物网络和基因表达数据，揭示了潜在的驱动基因，增强了对复杂疾病机制的理解。该算法在胶质母细胞瘤数据分析中，成功识别了 *KRAS* 和 *AKT1* 这两个罕见驱动基因，其中 *AKT1* 的激活已被证实与许多恶性肿瘤的发生密切相关。然而，由于癌症样本中存在大量乘客基因突变，准确区分驱动基因和乘客基因仍面临巨大挑战。针对这一问题，Hou 等人提出了一种基于最大突变影响函数的新方法[41]，通过量化突变对基因功能的影响，有效区分了驱动基因与乘客基因，为驱动基因的精准识别提供了新的技术手段。

1.3.2 转录调控与因果推断

转录调控是决定细胞内基因表达模式的核心生物过程，通过激活或抑制特定基因的表达，调控细胞的行为与功能。转录调控的异常可导致致癌基因的激活或抑癌基因的失活，这些变化是促进肿瘤发展的重要驱动因素。因此，准确推断转录调控关系对于理解细胞内信息处理机制、基因表达调控网络以及细胞如何响应环境变化具有关键意义[42]。然而，在转录调控网络的研究中，一个核心挑战是如何建立网络中的因果关系，即深入理解节点间的有向相互作用及其调控机制。

概率模型能够有效捕捉转录调控网络中潜在的随机性与不确定性，而因果推断则能够识别调控关系中的因果效应。贝叶斯网络作为一种经典的概率图模型，通过有向无环图（Directed Acyclic Graph，DAG）刻画变量间的依赖关系，能够对基因调控路径进行过滤和优先级排序，从而帮助识别关键的生物学过程。动态贝叶斯网络（Dynamic Bayesian Network，DBN）进一步扩展了 DAG 的应用范围，尤其适用于处理时间序列数据，能够捕捉基因表达随时间的动态变化，从而更准确地描述基因间的动态调控关系。此外，回归分析和结构方程模型也是因果推断的重要数学工具。回归分析通过探讨因变量与自变量之间的关系，能够推断调控网络中的作用强度和方向；而结构方程模型则通过分析复杂的变量间相互作用及潜在变量关系，为转录调控网络研究提供了更强的理论支持。

这些方法的实际应用充分展示了其在转录调控研究中的巨大潜力。例如，Feng 等人提出了一种基于因果推断与图神经网络相结合的基因调控网络推断方法 GRINCD[43]。该方法通过整合图表示学习与因果非对称学习技术，实现了对转录因子—靶基因调控关系的精准预测。在临床研究中，GRINCD 成功揭示了炎症性肠病向结直肠癌转变过程中的关键调控因子，并发现了潜在驱动因子 *WWTR1* 和 *E2F6*。在临床决策支持方面，Jonathan 等人创新性地将因果推断理论应用于诊断模型构建，开发了一种基于反事实推理的智能诊断算法[44]。通过算法诊断结果与 44 名临床医生的判断进行系统性比较，反事实诊断算法在诊断准确率上展现出显著优势（77.26% vs. 71.4%）。

1.3.3 细胞通讯与最优传输

细胞通讯，又称细胞信号传导，是指细胞间通过信号分子进行信息传递的复杂机制，对细胞响应环境变化和调节其行为具有至关重要的作用。在细胞通讯研究中，针对细胞个体之间的复杂关联关系，需要构建图与网络的数学模型，从而以数量化的形式刻画细胞之间的分子互作和信息传递过程。这种基于数学模型的定量分析方法能够深入揭示调控组织生理、疾病和发育过程的分子机制[45]。

结合最优传输理论方法和空间位置网络构建数学模型，已成为研究细胞通讯的一种可行途径。最优传输作为一种描述物质传递过程的优化工具，能够量化离散对象在不同状态间转移的“成本”，并确定不同概率分布之间转移物质所需的最小成本传输方式[46, 47]。由于细胞通讯通常发生在有限的空间距离内，引入空间信息至关重要，这不仅能够显著降低非空间数据引发的假阳性推断，还能更准确地识别特定微环境中的细胞通讯机制。最优传输方法通过将空间位置信息与基因表达信息结合，构建“传输矩阵”，求解最小化转移成本的连续优化问题，从而推断细胞间最可能的配体—受体信号分子流动情况。这种方法为解析细胞间信息的传递与处理机制提供了相对精准且强有力的数学支持[48]，不仅深化了对细胞通讯网络的理解，还为揭示组织微环境中的信号调控机制和疾病发生发展的分子基础提供了新的研究范式。

其中的成功案例包括：Cang 等人基于单细胞与空间转录组数据，提出了一种结构化最优传输方法 SpaOTsc[49]。通过为非空间细胞建立空间度量，SpaOTsc 计算细胞间的最优传输距离，从而推断配体—受体互动介导的细胞通讯，揭示细胞间的空间调控关系。而 COMMOT 则采用集体最优传输方法，考虑多种配体—受体对之间的竞争关系，并优化不同位置间资源传输的分配策略，以推断空间信号

传导方向[50]。这些创新方法为精准量化组织内细胞通讯细节提供了有力工具，不仅能够深入揭示癌细胞通过信号传递适应高异质性肿瘤微环境的机制，还为解析复杂生物学过程开辟了新的路径。

2　医数交叉研究展望

2.1　机制导向的数学模型构建

机制导向的数学模型是一种基于系统内在机制或原理来描述和预测系统行为的数学工具。它以物理规律、生物原理或社会行为等为理论基础，通过构建明确的因果关系链来解释现象并预测未来行为。例如，流体力学中的 Navier-Stokes 方程、化学反应中的 Michaelis-Menten 方程，都是机制导向模型的经典代表。与数据驱动模型相比，机制导向模型的核心优势在于其因果性和解释性。它不仅依赖观察数据的统计模式，更注重揭示变量之间的本质联系，从而深入刻画系统的内在规律，有效弥补传统数据驱动方法的不足。

随着多组学数据的快速积累和人工智能技术的迅猛发展，机制导向的数学模型在生物医学领域的重要性日益凸显。以新药研发中的医数交叉研究为例，尽管图同构算法能够有效捕捉分子结构相似性并预测生物活性，但在处理复杂分子系统的多尺度、多层次特性时仍存在局限性。药物分子在体内的行为不仅取决于其分子结构，还受到代谢动力学、药物分布和靶点结合动力学等多重因素的共同影响。因此，通过深入融合微分方程、概率模型与图深度学习框架，构建一个涵盖分子结构、药代动力学和靶点结合动态过程的统一理论体系，将成为药物设计的关键。这种机制导向的数学模型不仅能够真实地模拟生物系统，还有望显著提升新药研发的效率和成功率，为开发更安全、更高效的药物带来革命性突破。

机制导向的数学模型的另一典型应用场景是刻画肿瘤发生发展的分子机制。例如，现有的调控网络推断模型往往过于简化转录因子的调控作用，且未能充分考虑基因表达的时空动态性。实际上，转录调控过程是一个复杂的动态非线性系统，涉及多层次的相互作用和反馈循环。通过构建基于微分方程的基因表达动态模型，并结合概率模型、因果推断方法以及深度学习、大语言模型等前沿技术，可以更精细地模拟和预测转录调控网络的动态行为，从而精确识别与肿瘤进展和转移相关的关键分子。此外，机制导向的数学模型在罕见病研究和传染病防控等领域也具有广泛的应用潜力。这些应用不仅展示了机制导向数学模型在解决复杂生物医学问题中的独特优势，也为精准医疗和公共卫生体系的建设提供了强有力的科学工具。

2.2　数字生命和虚拟健康

《“健康中国 2030”规划纲要》明确提出，到 2030 年中国健康服务业总规模将达到 16 万亿元。在这一背景下，随着数字化技术对医疗健康领域的全面渗透、深度融合和持续赋能，数字健康产业无疑将成为未来医疗健康体系的主导力量。在 5G 网络、虚拟现实、人工智能和量子计算等前沿技术的飞速发展驱动下，医数交叉在数字生命和虚拟健康领域展现出广阔的应用前景。数字生命通过整合个体的生理、基因、行为和环境等多维度数据，构建物理实体的动态虚拟化身，为个性化健康管理和疾病预防提供了全新的技术范式；而虚拟健康则通过远程监控、数字临床试验、虚拟诊疗和健康教育等创新手段，将传统的面对面健康服务扩展至数字化、虚拟化的新维度，极大地提升了健康管理的可及性和效率。

数字生命和虚拟健康的结合正在深刻变革个体的健康管理方式与科学研究范式。以数字孪生为代表的新兴技术，依托数学模型重建人体的微观、介观和宏观网络化动态生命系统，能够精准模拟人体的生理和病理过程，进而定性、定量地描述生命活动的状态，构建数字生命与全息人体，为探索生命本质提供了更加系统化和全面化的认知框架。美国国家科学院、工程院和医学院已将计算机模拟药物发现、临床试验、预防性医疗和行为干预计划、临床团队协调以及大流行病应急准备等列为数字孪生在医学领域的初始应用场景。然而，数字孪生在生物医学领域仍面临重大挑战，其根本原因在于人体（包括肿瘤）具有多层次、多尺度的复杂性。例如，虚拟表示作为数字孪生的核心要素之一，可采用包括动态系统、微分方程和统计模型等多种数学形式，但无论选择哪种形式，都必须结合特定场景优化模型类型、保真度、分辨率和参数等。人体的多层次、多尺度特性使得现有数学模型难以全面适应各个层次的复杂性；同时，当多个模型在系统中耦合时，还面临稳定性和精度的双重挑战；此外，数据隐私保护、算法可靠性检验以及技术标准化等关键问题也亟待解决。因此，推动数学与生物医学、计算机科学等多领域学科

的深度交叉协作，有望突破上述瓶颈，从而推动医数交叉研究向更加综合化、动态化的方向发展，为精准医疗和个性化健康管理提供更强大的理论支持和技术保障。

3 结 语

总之，医学与数学的深度融合正在开辟学科交叉的新前沿，推动生物医学研究范式的根本性变革。通过将数学与现代生物医学技术紧密结合，人们能够更精准地解析疾病发生与发展的复杂分子机制，推动肿瘤治疗向个性化与精细化方向发展，并加速研究成果的临床转化。未来，随着跨学科合作的不断深入，医数交叉有望成为全球健康领域创新的核心引擎，推动医疗服务的创新与普及，助力人们迈向一个更加健康的未来。与此同时，医数交叉不仅为医学带来创新突破，也为数学学科开辟新的应用场景和理论挑战，促进两者在相互推动中实现共同进步。

参 考 文 献

[1] Sun DC, Guan XN, Moran AE, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. Nature Biotechnology, 2022, 40(4): 527—538.

[2] Shi JF, Aihara K, Chen LN. Dynamics-based data science in biology. National Science Review, 2021, 8(5): nwab029.

[3] Acosta JN, Falcone GJ, Rajpurkar P, et al. Multimodal biomedical AI. Nature Medicine, 2022, 28(9): 1773—1784.

[4] Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, 2011, 13(1): 36—46.

[5] Eraslan G, Simon LM, Mircea M, et al. Single-cell RNA-seq denoising using a deep count autoencoder. Nature Communications, 2019, 10(1): 390.

[6] Ma AJ, Wang XY, Li JX, et al. Single-cell biological network inference using a heterogeneous graph transformer. Nature Communications, 2023, 14(1): 964.

[7] Cao ZJ, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nature Biotechnology, 2022, 40(10): 1458—1466.

[8] Umar A, Dunn BK, Greenwald P. Future directions in cancer prevention. Nature Reviews Cancer, 2012, 12(12): 835—848.

[9] Liu XP, Chang X, Leng SY, et al. Detection for disease tipping points by landscape dynamic network biomarkers. National Science Review, 2019, 6(4): 775—785.

[10] Hong RH, Tong YY, Liu HS, et al. Edge-based relative entropy as a sensitive indicator of critical transitions in biological systems. Journal of Translational Medicine, 2024, 22(1): 333.

[11] Chen P, Li YJ, Liu XP, et al. Detecting the tipping points in a three-state model of complex diseases by temporal differential networks. Journal of Translational Medicine, 2017, 15(1): 217.

[12] Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nature Reviews Cancer, 2005, 5(11): 845—856.

[13] Zhou Y, Tao L, Qiu JH, et al. Tumor biomarkers for diagnosis, prognosis and targeted therapy. Signal Transduction and Targeted Therapy, 2024, 9(1): 132.

[14] Sawyers CL. The cancer biomarker problem. Nature, 2008, 452(7187): 548—552.

[15] Liu ZP. Identifying network-based biomarkers of complex diseases from high-throughput data. Biomarkers in Medicine, 2016, 10(6): 633—650.

[16] Sun DC, Ren XW, Ari E, et al. Discovering cooperative biomarkers for heterogeneous complex disease diagnoses. Briefings in Bioinformatics, 2019, 20(1): 89—101.

[17] Chuang HY, Lee E, Liu YT, et al. Network-based classification of breast cancer metastasis. Molecular Systems Biology, 2007, 3: 140.

[18] Liu XP, Wang YT, Ji HB, et al. Personalized characterization of diseases using sample-specific networks. Nucleic Acids Research, 2016, 44(22): e164.

[19] Liu XP, Liu ZP, Zhao XM, et al. Identifying disease genes and module biomarkers by differential interactions. Journal of the American Medical Informatics Association, 2012, 19(2): 241—248.

[20] Steyaert S, Pizurica M, Nagaraj D, et al. Multimodal data fusion for cancer biomarker discovery with deep learning. Nature Machine Intelligence, 2023, 5(4): 351—362.

[21] Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. Journal of the Optical Society of America A, 1984, 1(6): 612—619.

[22] Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. IEEE Transactions on Signal Processing, 2003, 51(2): 560—574.

[23] Savage SR, Yi XP, Lei JT, et al. Pan-cancer proteogenomics expands the landscape of therapeutic targets. Cell, 2024, 187(16): 4389—4407.e15.

[24] Bongini P, Bianchini M, Scarselli F. Molecular generative graph neural networks for drug discovery. Neurocomputing, 2021, 450: 242—252.

[25] Mullowney MW, Duncan KR, Elsayed SS, et al. Artificial intelligence for natural product drug discovery. Nature Reviews Drug Discovery, 2023, 22(11): 895—916.

[26] Bao LJ, Wang Z, Wu ZX, et al. Kinome-wide polypharmacology profiling of small molecules by multi-task graph isomorphism network approach. Acta Pharmaceutica Sinica B, 2023, 13(1): 54—67.

[27] Du YH, Yao YB, Tang JX, et al. Drug-target interactions prediction *via* graph isomorphic network and cyclic training method. Expert Systems with Applications, 2024, 249: 123730.

[28] Wang MD, Lei XJ, Liu L, et al. GIAE-DTI: predicting drug-target interactions based on heterogeneous network and GIN-based graph autoencoder. IEEE Journal of Biomedical and Health Informatics, 2024, DOI: 10. 1109/JBHI. 2024. 3458794.

[29] Tirosh I, Izar B, Prakadan SM, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science, 2016, 352(6282): 189—196.

[30] Sugasawa S, Noma H. Estimating individual treatment effects by gradient boosting trees. Statistics in Medicine, 2019, 38(26): 5146—5159.

[31] Brindha GR, Rishiikeshwer BS, Santhi B, et al. Precise prediction of multiple anticancer drug efficacy using multi target regression and support vector regression analysis. Computer Methods and Programs in Biomedicine, 2022, 224: 107027.

[32] Luria SE, Delbrück M. Mutations of bacteria from virus sensitivity to virus resistance. Genetics, 1943, 28(6): 491—511.

[33] Dewanji A, Luebeck EG, Moolgavkar SH. A generalized *Luria*-Delbrück model. Mathematical Biosciences, 2005, 197(2): 140—152.

[34] Komarova NL, Wu L, Baldi P. The fixed-size *Luria*-Delbruck model with a nonzero death rate. Mathematical Biosciences, 2007, 210(1): 253—290.

[35] Liu RY, Wang S, Tan XW, et al. Identifying optimal adaptive therapeutic schedules for prostate cancer through combining mathematical modeling and dynamic optimization. Applied Mathematical Modelling, 2022, 107: 688—700.

[36] Bowcock AM. Genomics: guilt by association. Nature, 2007, 447(7145): 645—646.

[37] Page L, Brin S, Motwani R, et al. The pagerank citation ranking: Bringing order to the web. (2001-10-30)/[2024-11-20\] http://ilpubs. stanford. edu: 8090/422/? utm _ campaign=Technical%20SEO%20Weekly&utm_medium=email&utm_source=Revue%20newsletter.

[38] Yu JT, Leng JC, Sun DC, et al. Network Refinement: Denoising complex networks for better community detection. Physica A: Statistical Mechanics and Its Applications, 2023, 617: 128681.

[39] Leiserson MDM, Vandin F, Wu HT, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. Nature Genetics, 2015, 47(2): 106—114.

[40] Bashashati A, Haffari G, Ding JR, et al. DriverNet: uncovering the impact of somatic driver mutations on transcriptional networks in cancer. Genome Biology, 2012, 13(12): R124.

[41] Hou YN, Gao B, Li GJ, et al. MaxMIF: a new method for identifying cancer driver genes through effective data integration. Advanced Science, 2018, 5(9): 1800640.

[42] Fan JL, Nazaret A, Azizi E. A thousand and one tumors: the promise of AI for cancer biology. Nature Methods, 2024, 21(8): 1403—1406.

[43] Feng K, Jiang HY, Yin CY, et al. Gene regulatory network inference based on causal discovery integrating with graph neural network. Quantitative Biology, 2023, 11(4): 434—450.

[44] Richens JG, Lee CM, Johri S. Improving the accuracy of medical diagnosis with causal machine learning. Nature Communications, 2020, 11(1): 3923.

[45] Armingol E, Baghdassarian HM, Lewis NE. The diversification of methods for studying cell-cell interactions and communication. Nature Reviews Genetics, 2024, 25(6): 381—400.

[46] Villani C. Optimal transport: old and new. Berlin: Springer, 2009.

[47] Cao K, Gong QY, Hong YG, et al. A unified computational framework for single-cell data integration with optimal transport. Nature Communications, 2022, 13(1): 7419.

[48] Zuo CM, Xia JJ, Chen LN. Dissecting tumor microenvironment from spatially resolved transcriptomics data by heterogeneous graph learning. Nature Communications, 2024, 15(1): 5057.

[49] Cang ZX, Nie Q. Inferring spatial and signaling relationships between cells from single cell transcriptomic data. Nature Communications, 2020, 11(1): 2084.

[50] Cang ZX, Zhao YX, Almet AA, et al. Screening cell-cell communication in spatial transcriptomics *via* collective optimal transport. Nature Methods, 2023, 20(2): 218—228.

From Data to Mechanisms: Medical-Mathematical Integration Driven Advances in Precision Oncology-Current Research Status and Future Perspectives

Duanchen Sun[1,4] Renmin Han[2] Chenglin Ma[2] Maoteng Duan[1] Fei Wang[3,4*] Bingqiang Liu[1*]

1. *School of Mathematics, Shandong University, Jinan 250100, China*
2. *Research Center for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, China*
3. *The Second Hospital, Shandong University, Jinan 250033, China*
4. *Shandong Key Laboratory of Cancer Digital Medicine, Jinan 250033, China*

Abstract The rapid accumulation of complex, multi-level, and heterogeneous medical data makes traditional research paradigms less effective. However, it also presents an opportunity to reform current medical research. In recent years, interdisciplinary research in biomedicine and intelligence science has significantly progressed, driving the rapid development of disease prediction and precision medicine, in which mathematics has played a fundamental role. Deep interdisciplinary research in biomedicine and mathematics makes achieving a fundamental characterization of life mechanisms possible. This article reviews the current status of interdisciplinary research in biomedicine and mathematics, exploring the critical role of mathematics in tumor diagnosis, treatment, and the mechanisms underlying tumorigenesis. Next, we discuss interdisciplinary research's prospects and potential in mathematical model design, digital life, and virtual health. By constructing and applying mathematical models, interdisciplinary research in biomedicine and mathematics is expected to provide powerful cancer prevention and treatment solutions, guiding biomedicine research more efficiently, precisely, and intelligently.

Keywords interdisciplinary research; data analysis; mathematical models; bioinformatics; biomathematics; oncological diagnosis and treatment

王 斐 山东大学第二医院副主任医师，研究员，临床教授，博士生导师。从事乳腺癌临床诊疗与科学研究，主要围绕乳腺癌风险干预及诊疗策略优化开展流行病学与基础研究。主持国家自然科学基金等科研项目，获评山东省“泰山学者”青年专家。

刘丙强 山东大学数学学院教授、博士生导师。从事数学与生物医学交叉研究，面向以肿瘤为代表的复杂疾病，研究生物医学大数据处理与分析中的数学模型与算法挑战问题。主持国家重点研发计划、国家自然科学基金、山东省杰出青年基金等项目。获评教育部“长江学者”青年学者、山东省“泰山学者”青年专家。获教育部国家教学成果二等奖、山东省教学成果特等奖。

孙端辰 山东大学数学学院教授、博士生导师。从事数学、计算机与前沿生物医学问题的交叉研究，围绕肿瘤等复杂疾病建模中存在的若干计算问题开展数学模型和算法设计。主持国家自然科学基金项目 1 项，参与国家重点研发计划 1 项。

（责任编辑 陈 鹤 张 强）

* Corresponding Authors, Email: fei.wang@sdu.edu.cn; bingqiang@sdu.edu.cn