Biotechnology Frontier
Volume 11 Issue 2, December 2024, PP. 20-24

Dynamic Network Biomarkers and Early Warning of Type 2 Diabetes

Hongling Liu, Yiping Lao†, Wenqi Liu

Ludong University, School of Mathematics and Statistics, Yantai, Shandong, 264025, China

Abstract

As a global chronic disease, type 2 diabetes has a serious impact on people's health. Its deterioration mechanism, critical early warning, early diagnosis and treatment strategies have been the focus of biomedical research. The emergence of this disease is usually due to an imbalance in the interaction network of biomolecules. Therefore, studying the dynamic changes of individual biomolecular networks is critical to revealing the pathogenesis of diabetes, in which the network differences of a single sample are particularly important, which helps to obtain early warning signals and key marker genes in the disease process. Based on high-dimensional omics data, using statistics, mathematics and computational biology knowledge, the early warning indicators and key biomarkers of type 2 diabetes were explored in depth. Through the combination of bulk 2 space and DNB algorithm, the identification process of critical points was optimized.

Keywords: Dynamic Network Biomarkers; Type 2 Diabetes; Critical State; Early Warning

2 型糖尿病动态网络生物标志物与早期预警

刘洪领，劳怡萍†，刘文琦

鲁东大学数学与统计科学学院，山东烟台 264025

摘 要：2 型糖尿病作为一种全球性的慢性疾病，严重影响人们健康，其病情恶化机制、临界预警、早期诊断与治疗策略等一直是生物医学研究的热点。这种疾病的产生通常是由于生物分子相互作用网络的失衡。因此，研究个体生物分子网络动态变化对于揭示糖尿病的发病过程十分关键，其中单个样本的网络差异尤为重要，这有助于获得疾病进程中的预警信号和关键标志物基因。基于高维组学数据，运用统计学、数学和计算生物学知识，深入探究 2 型糖尿病的早期预警指标和关键生物标志物，通过 Bulk2space 和动态网络生物标志物算法相结合的方式，优化了临界点的识别过程。

关键词：动态网络生物标志物；2 型糖尿病；临界状态；早期预警

1 引言

1.1 研究背景

二型糖尿病是一种全球性公共卫生问题，其发病率在全球范围内持续上升，因此 2 型糖尿病的有效预防与早期诊断对于降低其发生率至关重要。在疾病发展的早期阶段，由于缺乏显著的体征和症状，不易及时发现。不过，从健康状态到明显疾病状态之间，会存在一个过渡的“临界状态”阶段，在这个阶段采取适当的干预措施，可能会有效遏制糖尿病的发展，防止病情进一步恶化。这一临界状态的认识和干预，对于防治 2 型糖尿病具有重要的临床意义。

复杂疾病的演变过程可以分为正常状态或者称健康状态、发病前状态或者称临界状态，这是一个关键的分水岭、以及发病状态，此时疾病已经完全形成。如果能够在疾病发展的临界状态阶段及时进行干预和治疗，系统的状态有可能逆转回到健康的正常状态。否则，病情将继续发展，超过临界点进入发病状态，在疾病发展这三个阶段当中，动态网络标志物表征的分子网络在临界状态时，事实上是健康状态的极限，

此时这种状态的特点是稳定性和鲁棒性较低[1]。

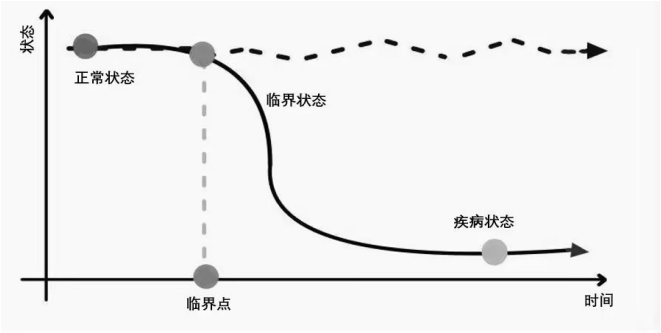


图 1 复杂疾病的动态演化

1.2 生物标志物的研究意义

动态网络生物标志物可以识别疾病的临界状态或临界点，从而预测而不是诊断疾病。但是事实上将动态网络生物标志物理论应用于临床实践较为困难，因为在临界状态下评估动态网络生物标志物需要结合每个个体的多个样本的数据，而这些数据通常不可用，从而限制了动态网络生物标志物的适用性。因此本文结合 Bulk2Space 算法将 RNA-seq 数据转化成单细胞水平的基因表达数据，利用 β-VAE 在聚类空间内将异质转录组解卷积为单细胞转录组数据，然后使用空间转录组参考将生成的单个细胞分配到最佳空间位置，最后进行动态网络生物标志物分析。

根据美国国立卫生研究院的描述[2]，生物标志物是一种可以通过客观手段进行量化和评价的特征，它反映了正常的生理活动、疾病进展或对治疗的反应。生物标志物揭示了遗传和环境因素相互作用的后果，它在疾病的早期发现中起着关键作用。在疾病尚未表现出明显症状的隐匿阶段，生物标志物的检测对于疾病的早期诊断提供了极大帮助。

1.3 基因表达数据分析意义

随着人类基因组计划的完成和生命科学领域的不断进展，大量的动植物及微生物基因组数据得以积累，并且这些数据的总量还在不断上升。得益于基因芯片技术的迅猛发展，与疾病相关的基因表达数据开始被广泛地存储在多个公开的生物信息资源库中。基因芯片技术将特定的 DNA 寡核苷酸或 DNA 片段作为探针，定位固定在芯片的特定区域，并利用碱基互补配对原则进行杂交实验，以此来识别目标序列的存在和其相对的表达水平。在基因芯片技术的促进下，目前存在多个公共数据库收录了与 2 型糖尿病相关的基因表达数据。在本项研究中，所采用的数据是从基因表达综合数据库 GEO 中获取的。

2 数据收集与整理

动物模型是研究糖尿病进展和鉴定 2 型糖尿病相关生物标志物的重要资源。在众多用于糖尿病研究的近交啮齿类动物模型中，GK 大鼠是一个突出的选择。这种模型的独特之处在于其糖尿病的遗传背景是多基因的，这与大多数其他具有单基因缺陷的近交啮齿类动物模型不同。GK 大鼠的多种糖尿病特征，如高血糖、葡萄糖耐量异常和胰岛素抵抗，与人类糖尿病相似，尽管它们高脂血症通常不显著的表现[3]。GK 大鼠的多基因遗传特性以及其 100%的外显率，使其在研究 2 型糖尿病的发生和发展方面具有特别的价值。研究从 GEO 数据库（GSE13268）中获取了 50 份 GK 大鼠的脂肪组织样本，这些大鼠分别接受了正常饮食和高脂肪饮食，并在 5 个不同的年龄阶段（4、8、12、16 和 20 周）进行了宰杀。每个时间点包括 5 个正常饮食喂养的 GK 样本和 5 个高脂肪饮食喂养的 GK 样本。

获取 2 型糖尿病数据集之后，根据数据采样情况，排除异常样本，利用 GEO2R 软件对剩余样本的基因芯片数据进行差异表达分析（见图 2）。

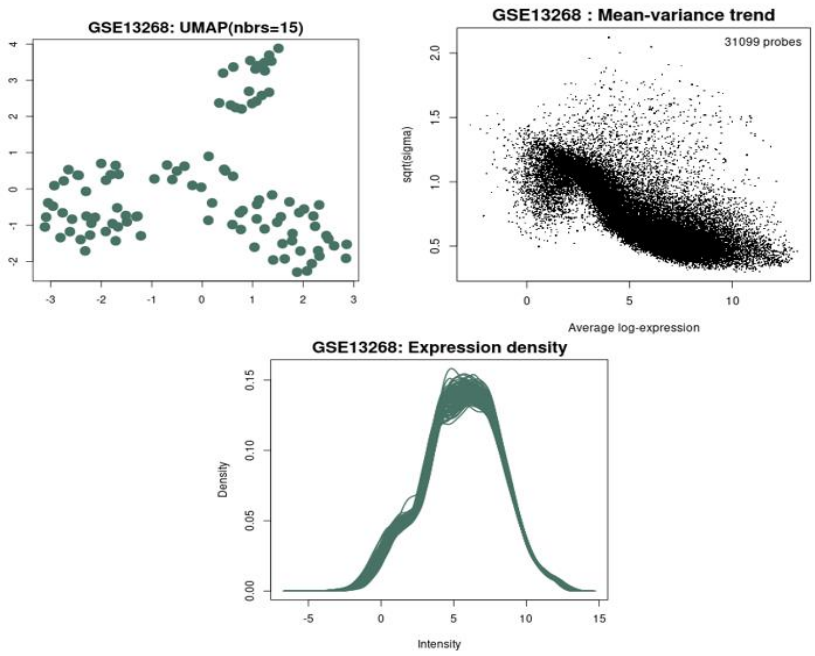


图 2 数据分析图

3 研究方法

3.1 利用动态网络生物标志物探究 2 型糖尿病的预警信号

3.1.1 动态网络生物标志物识别

传统的分子生物标志物通常是通过检测单个代谢物或蛋白质来推断器官的疾病状况，这种方法往往是在器官已经显示出疾病迹象时才能检测到疾病的状态。这种做法虽然能够区分正常和疾病状态，但不适用于早期诊断。动态网络生物标志物多个生物学指标相互作用形成的复杂网络，这些指标的变化可以反映疾病的进展和状态。与传统的单一生物标志物相比，动态网络生物标志物可以更全面、更准确地反映疾病的生物学特征。

动态网络生物标志物作为一种医学上常用的分子生物标志物，其重要性在于它能够揭示疾病状态的存在及其严重程度。在不同的生理状态下，动态网络生物标志物的数值表现出显著的一致性，而在疾病前期，动态网络生物标志物内的分子浓度则会呈现出波动性。这种波动性是动态网络生物标志物区别于传统生物标志物的关键特征，动态网络生物标志物还能实时监测疾病进展和治疗效果，其波动性不仅为早期诊断提供线索，还为个性化医疗和精准治疗开辟新路径[4]。

3.1.2 动态网络生物标志物方法的具体步骤

为了阐述动态网络生物标志物方法的实施步骤，我们提供了详细的流程图，如图 3 所示：

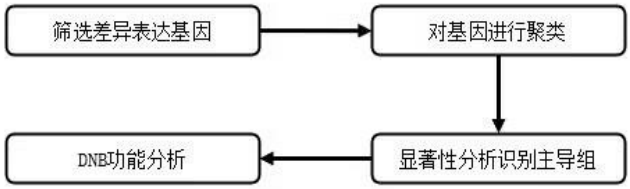


图 3 动态网络生物标志物方法实施流程图

（1）筛选数据中的差异表达基因。在各个时间点，利用多重测试和 T 检验方法，识别出病例组与对照

组之间表达水平存在显著差异的基因，设定的差异变化倍数为 1.5，显著性水平 p 值为 0.05。

（2）对差异表达基因进行时间点特异的相关性聚类。在每一个时间点，对筛选出的差异表达基因进行基于相关性的聚类分析，形成多个基因组，每组中的基因在特定时间点的表达相关性较高。

（3）通过显著性检验确定关键的主导组。根据动态网络生物标志物的三个评判标准，对各个基因组进行显著性检验，选取两个连续时间点间临界指数变化最大的组作为潜在的主导组，同时将对应的时期视为疾病前期阶段的潜在时间点。

（4）执行动态网络生物标志物的功能分析。为了验证所识别的用于监测特定疾病早期预警信号的动态网络生物标志物是否准确预测了即将到来的关键变化，对特定疾病进行了功能性的分析。

3.1.3 应用价值

成纤维细胞脂质运载蛋白、生长因子家族和脂肪细胞型脂肪酸结合蛋白）等血清标志物在 2 型糖尿病的早期预警中展现出至关重要的作用。通过识别 2 型糖尿病的早期预警信号，提高了 2 型糖尿病早期的发现率，从而能够尽早进行干预，有效提升了 2 型糖尿病的防控水平。

3.2 基于 Bulk2Space 算法并结合动态网络生物标志物探究 2 型糖尿病的预警信号

3.2.1 Bulk2Space 算法介绍

Bulk2Space 算法包含两个主要过程：反卷积和空间定位。

该算法处理的是从细胞群体中获得的批量 RNA-seq 数据，而单细胞 RNA-seq 技术则是针对单个细胞进行标记和测序，它们研究的细胞类型和状态是相对应的。在这一前提下，批量转录组数据可以通过精细的反卷积过程被解析成单细胞水平的数据：利用单细胞参考数据对 Bulk 数据进行反卷积从而构建一个模拟的单细胞矩阵。注意这里的反卷积不同于 CIBERSORTx 算法、MuSiC 算法和贝叶斯棱镜算法，这些算法结合 bulk 数据和单细胞数据，给出 Bulk 数据里的细胞类型的比例情况。而 Bulk2space 是直接输出一个基于 Bulk 数据推断的“单细胞表达矩阵”。

第二步空间映射：在假定的“单细胞表达矩”的基础上，将每个细胞映射到空间位置上。本质上，这和传统的单细胞数据映射到空转数据是一个逻辑。在算法第二阶段，尽管现有的空间转录组学技术无法达到单细胞级别的分辨率或全面捕捉转录组信息，却能够基于表达谱的相似性，为算法第一步中生成的单细胞提供空间位置信息。因此，将单细胞数据与空间转录组学技术相结合，能有效克服技术瓶颈，实现批量转录组数据在单细胞分辨率下的空间反卷积（见图 4[5]）。

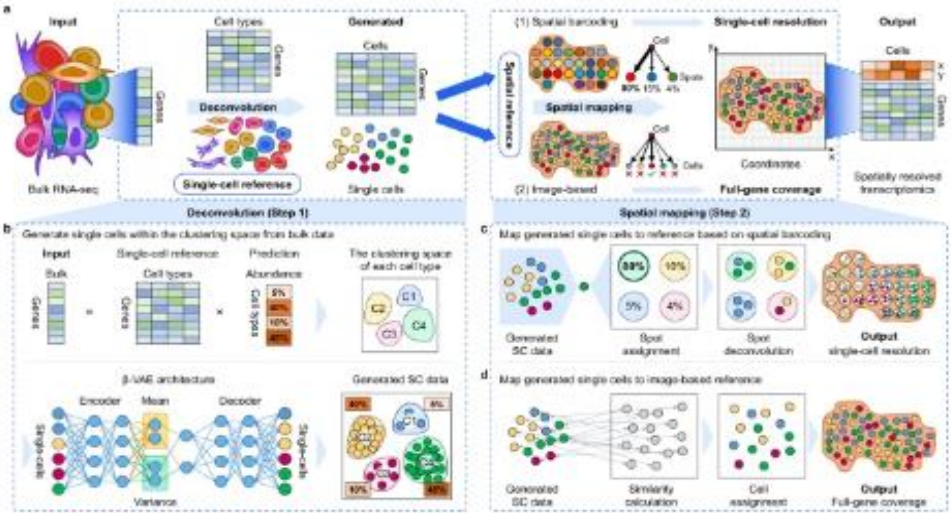


图 4 Bulk2Space 算法原理图

3.2.2 Bulk2Space 算法的应用价值

在深入分析了两种空间分辨转录组学技术的利弊之后，研究员为每种技术制定了特定的空间映射方案。针对基于图像的原位杂交技术，Bulk2Space 算法不仅能够预测目标基因 mRNA 的空间分布，还能预测其他 mRNA 物种的空间表达。同时，该算法有助于对那些通过靶向方法难以进一步分类的细胞进行注释。对于采用空间条形码的 RNA 测序技术，由于每个测序的组织区域包含了来自多个细胞的转录本，因此这些区域被视为细胞混合物，无法达到单细胞分辨率。Bulk2Space 算法能够将单个细胞定位到最合适的空间点，并将这些混合物分解为单个细胞的集合，从而为空间数据提供了一种空间分辨的单细胞去卷积方法。

研究结果显示，Bulk2Space 算法在生物学和临床应用领域有巨大的潜能，包括将组织的分子特征与组织学表型相关联，实现更细致的细胞分群，及预测疾病发展的分子机制。

4 结论与展望

4.1 结论

相比于传统的基于动态网络生物标志物理论的临界点探测方法，结合机器学习算法 Bulk2space 算法将 RNA-seq 数据转化成单细胞水平的基因表达数据进行动态网络生物标志物分析识别出 2 型糖尿病的潜在生物标志物。在动态网络生物标志物的识别能力上，结合 Bulk2Space 算法通过将 RNA-seq 数据转化可以快速地收缩网络，在提取动态网络生物标志物时能够更有效地识别出关键节点的基因，克服了传统方法中由于网络规模过大而导致的识别困难问题，从而更有利于识别出关键节点的基因。此外，该算法还可以对单个细胞的空间表达数据进行空间解剖学分析，以解释单细胞和基因表达的空间分布，相比于传统中每个时刻的数据是独立参与运算的加快了各个基因之间的联系。

我们将 Bulk2Space 改进算法应用到数据集中，复现了文献中报道的方法给出的结果[6]，基于脂肪组织数据的临界预警信号首先出现在 GK 大鼠 8 周龄，这与大鼠血浆葡萄糖和胰岛素浓度的表型变化趋势一致。

在实验过程中，GK 大鼠的血浆葡萄糖水平在 4 至 12 周期间逐渐上升，并在第 12 周时达到最高点，此后整个实验期间维持在峰值水平。与此同时，GK 大鼠的血浆胰岛素浓度在 4 至 8 周期间迅速上升，但在 8 周之后开始降低。根据动态网络生物标志物的三个标准，我们在这一时间点形成了一个特定的模块，这标志着疾病的关键转变和突发变化。通过广泛的文献回顾和直接的动态网络生物标志物分析，我们进一步证实了这一临界点的有效性，并揭示了 2 型糖尿病疾病恶化过程中的临界状态。采用 Bulk2Space 结合动态网络生物标志物技术来检测 2 型糖尿病的早期预警信号，并在疾病恶化前预测疾病的发生，为 2 型糖尿病的预防和控制提供了新的策略。

4.2 展望

动态网络生物标志物为 2 型糖尿病的早期预警提供了新的思路。未来，随着高通量组学技术和网络生物学的发展，研究人员可以更深入地探索 2 型糖尿病的动态网络生物标志物，并开发更精确、更有效的 2 型糖尿病早期预警模型，为 2 型糖尿病的预防和治疗提供重要的科学依据。

参考文献

[1] 刘锐,杨茜然.复杂生物系统临界状态的识别与预警[J].中山大学学报(自然科学版)(中英文),2024-09-27.

[2] 徐俊花,基于动态网络生物标志物理论识别 CML 发展及疗效的生物标志物[D].江南大学,2020.

[3] 成婷婷,张婷倩,黄朝霞,刘晓清,李军.II 型糖尿病动物模型的研究进展[J].生物医学,2023.

[4] 蒋罚,乳腺癌发展演进的动态网络生物标志物研究[D].太原理工大学.

[5] Liao,J.et al. De novo analysis of bulk RNA-seq data at spatially resolved single-cell resolution. Nat Commun 13,6498 (2022).

[6] 闫金铃,基于单样本差异网络的复杂疾病早期预警和关键标志物识别[D].河南科技大学,2022.