



Report

 汇报人: Lilian

2022/3/11



文献来源

[nature](#) > [nature communications](#) > [articles](#) > article

Article | [Open Access](#) | [Published: 05 January 2021](#)

Benchmarking joint multi-omics dimensionality reduction approaches for the study of cancer

[Laura Cantini](#) , [Pooya Zakeri](#), [Celine Hernandez](#), [Aurelien Naldi](#), [Denis Thieffry](#), [Elisabeth Remy](#) & [Anaïs Baudot](#) 

[Nature Communications](#) **12**, Article number: 124 (2021) | [Cite this article](#)

13k Accesses | **14** Citations | **61** Altmetric | [Metrics](#)

用于癌症研究的联合多组学降维方法的基准



Abstract

- ✓ 有效整合高维多组学数据可以极大地增强我们对生物系统的理解
- ✓ 联合降维 (joint Dimensionality Reduction, jDR) 方法是最有效的方法之一;
但当前jDR方法较多, 原理各异, 需要一个具有指导意义的基准来方便研究人员选择合适的方法;

文章主要工作——使用了三个互补基准对九种jDR方法进行了评估

评估它们在从模拟的多组学数据集中检索真实样本聚类的性能

使用 TCGA 癌症数据来评估它们在预测生存、临床注释和已知途径/生物过程方面的优势

评估了他们对多组学单细胞数据的分类



Introduction

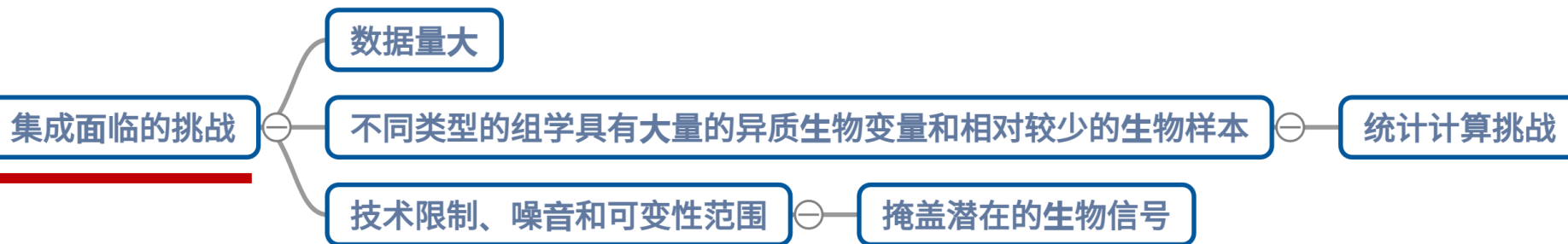
- ✓ 多组学数据研究面临的一大问题就是如何将数据进行有效的整合，目前主要通过顺序组合在单个组学上获得的结果，即**早期集成** (early integration) 和**晚期集成** (late integration)，但真正的多组学数据联合分析（又名**中间集成**）仍然非常罕见

多组学主要集成方法

- ✓ 通过整合多源组学数据，我们可以减少实验和生物噪声的影响、不同的组学技术有望捕捉细胞在不同生物过程的不同功能。不同的组学是互补的，每个组学都包含其他组学中不存在的信息，因此**多组学集成有望提供更全面的生物系统概述**。在癌症研究中，组学已经在不同的分子层进行了分析，例如基因组、转录组、表观基因组和蛋白质组。**集成的有利之处**



Introduction



已有集成方法



Introduction

从这些初步审查中，DR 方法表现得特别好。值得注意的是，DR 用于计算生物学的不同环境中，例如数据可视化或矩阵完成。

DR 方法非常适合解决高维数学问题。此外，其输出中包含的丰富信息增强了它们与多组学集成的相关性。

DR 方法能够对样本进行分类（聚类/亚型）、确定的聚类/亚型的临床表征以及各种其他下游分析，包括细胞过程和/或途径的分析。

因此，DR 与专门的下游分析相结合，提供了有关上述所有关键目标的信息，即将样本分类为亚型、它们与结果/生存的关联，以及重建其潜在的分子机制。

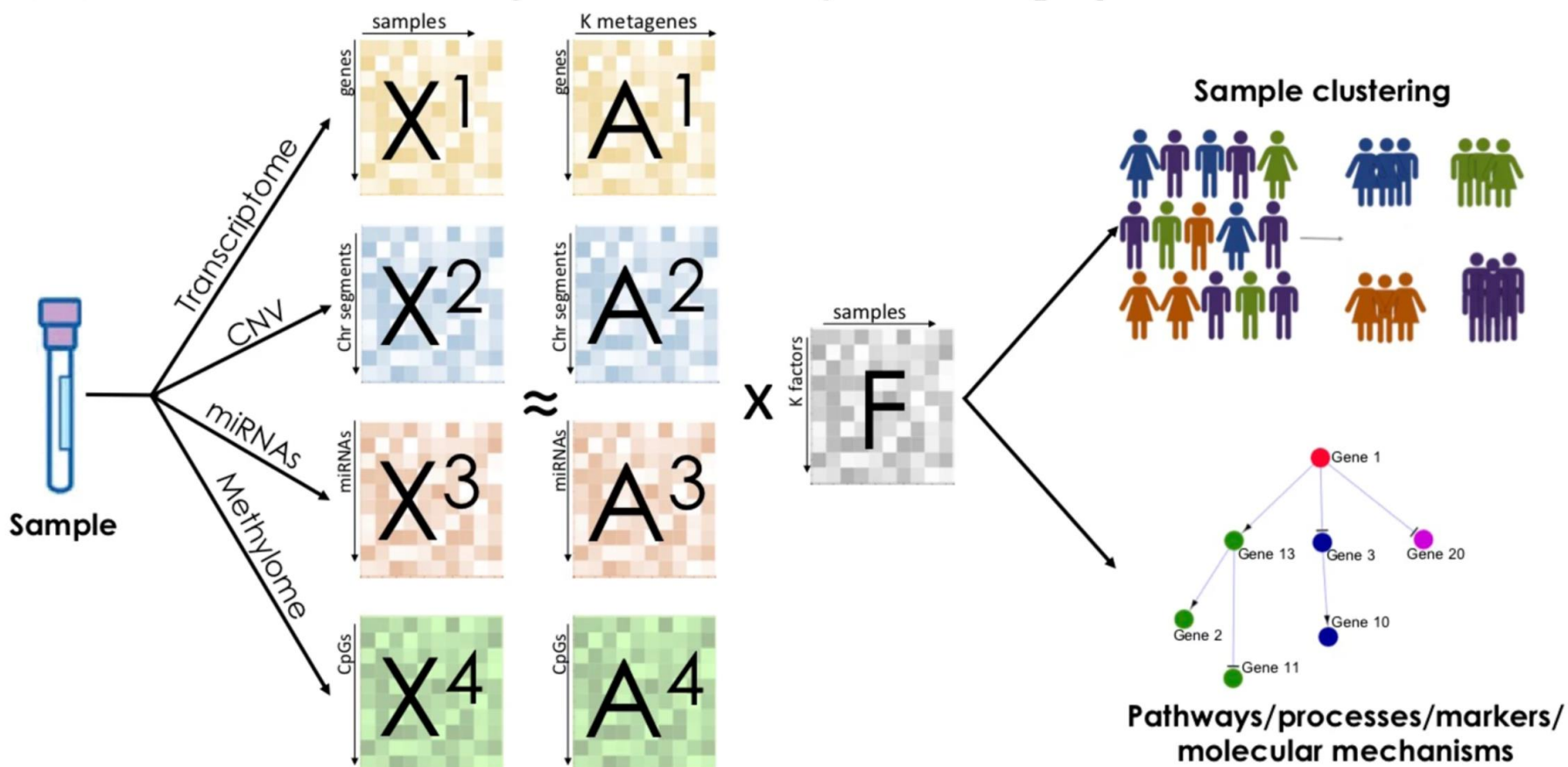
DR方法的优点



Introduction

多组学是从同一个样本中分析的。每个组学对应一个不同的矩阵 X^i 。jDR 方法将矩阵 X^i 分解为因子矩阵 F 和权重矩阵 A^i 的乘积。然后可以使用这些矩阵对样本进行聚类并识别分子过程。

A Multi-omics joint Dimensionality Reduction (jDR)

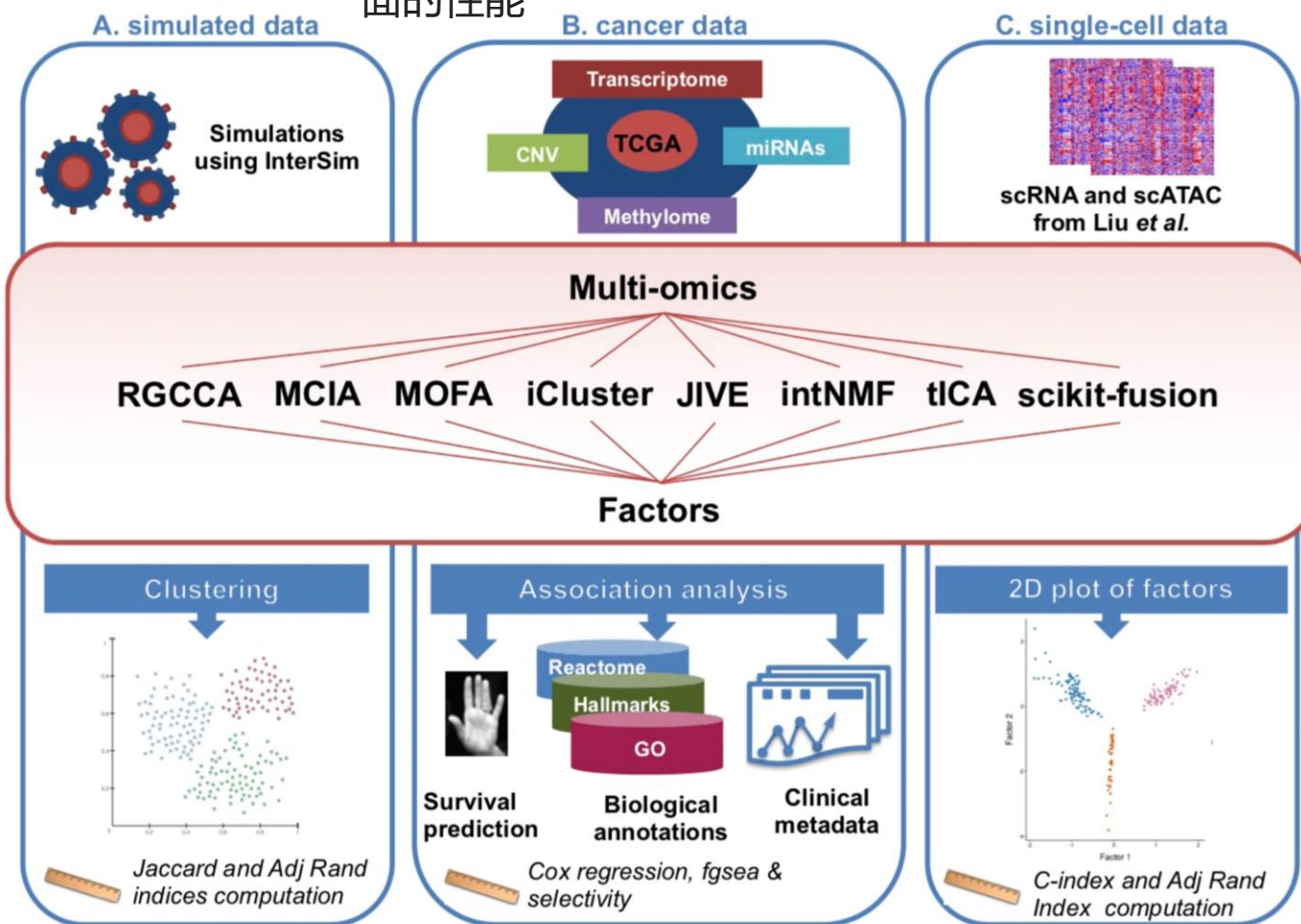




Introduction

B

- 1、模拟了多组学数据集并评估了九种 jDR 方法在检索真实样本聚类方面的性能;
- 2、使用 TCGA 多组学癌症数据来评估 jDR 方法在预测生存、临床注释和已知途径/生物过程方面的优势
- 3、估了这些方法在对来自癌细胞系的多组学单细胞数据进行分类方面的性能



Results & Methods

Joint Dimensionality Reduction approaches and principles

联合降维方法和原则

联合降维 (jDR) 方法旨在将高维组学数据减少到低维空间。在生物学中使用 jDR 的基本原理是，生物样品的状态由多个同时存在的生物因素决定，从一般过程（例如增殖和炎症）到细胞特异性过程。

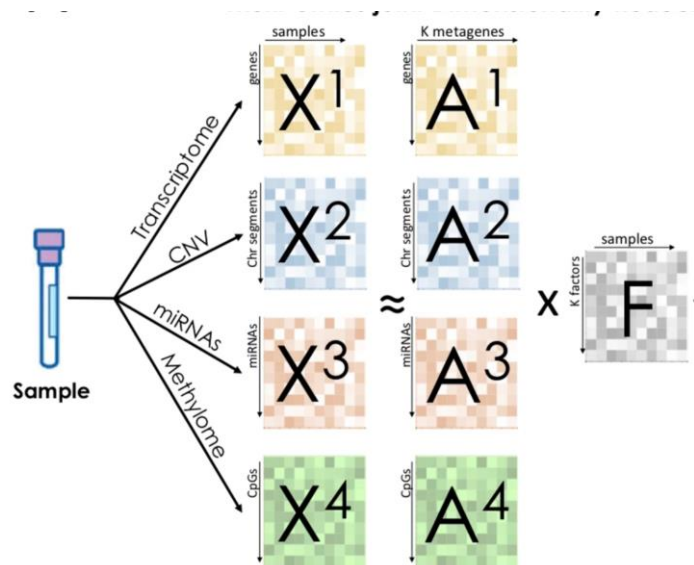
在测量组学数据时，对生物样本的状态进行快照，从而检测样本中活跃的各种生物信号的复杂混合物。jDR 的目标是对这种混合物进行反卷积，并暴露有助于生物样品状态的不同生物信号。

Results & Methods

Joint Dimensionality Reduction approaches and principles

联合降维方法和原则

We consider P omics matrices $\mathbf{X}^i, i = 1, \dots, P$ is thus the number of factors of the decomposition. Here and in the following, we will denote as factors the rows of the factor matrix and as *metagenes* the columns of the weight/projection matrix corresponding to transcriptomic data.



$$\mathbf{X}^i = \mathbf{A}^i * \mathbf{F}$$

\mathbf{A}^i 为某特定组学的权重矩阵

\mathbf{F} 为因子矩阵，其中行代表，列为 metagenes

方法1: *intNMF*

综合非负矩阵分解

*intNMF*¹⁷ is one of the numerous generalizations of NMF to multiomics data. The method decomposes each omics matrix \mathbf{X}^i into a product of non-negative matrices: the factor matrix \mathbf{W} , and an omics-specific matrix \mathbf{H}^i :

$$\mathbf{X}^i = \mathbf{W}\mathbf{H}^i, \text{ for } i = 1, \dots, P \text{ with } \mathbf{W} \text{ and } \mathbf{H}^i \text{ positive matrices} \\ \text{for } i = 1, \dots, P. \quad (1)$$

The algorithm minimizes the objective function $\mathbf{Q} = \min_{\mathbf{W}, \mathbf{H}} \sum_{i=1}^P \theta^i \|\mathbf{X}^i - \mathbf{W}\mathbf{H}^i\|$.

Once the matrices \mathbf{W} and \mathbf{H}^i $i = 1, \dots, P$

1、基于非负矩阵分解;

2、将每个组学矩阵 \mathbf{X}^i 分解为非负矩阵的乘积: 因子矩阵 \mathbf{W} 和组学特定矩阵 \mathbf{H}^i

所有组学特征具有统一的因子矩阵, **考虑了所有组学数据中的共享因素**

方法2: *Joint and individual variation explained (JIVE)*

JIVE¹⁸ is an extension of PCA to multiomics data. JIVE decomposes each omics matrix into three structures: a joint factor matrix (**J**), a omics-specific factor matrix (**A**) and a residual noise (**E**):

$$\mathbf{X}^i = \mathbf{J} + \mathbf{A}^i + \mathbf{E}^i = \mathbf{U}^i \mathbf{S} + \mathbf{A}^i + \mathbf{E}^i, \text{ for } i = 1, \dots, P \quad (2)$$

with $\mathbf{E}^i, \mathbf{A}^i$ and \mathbf{U}^i are $(n_i \times k)$

The algorithm minimizes $\|\mathbf{E}\|^2$ $\mathbf{E}^i = \mathbf{X}^i - \mathbf{U}^i \mathbf{S} - \mathbf{A}^i$ $\mathbf{E} = [\mathbf{E}^1 \dots \mathbf{E}^P]^T$

- 1、基于主成分分析 (PCA) ;
- 2、 将每个组学矩阵分解为三个结构：联合因子矩阵 (**J**)、 组学特定因子矩阵 (**A**) 和残余噪声 (**E**)

既考虑了**包含在所有组学中共享的唯一因子矩阵**，也描述了**包含不同组学特定因子矩阵**。

方法3: *Multiple co-inertia analysis (MCIA)*

多重协惯性分析

MCIA¹⁹, is an extension of co-inertia analysis (CIA) to more than two omics datasets. MCIA factorizes each omics into omics-specific factors

$$\mathbf{X}^i = \mathbf{A}^i \mathbf{F}^i + \mathbf{E}^i, \text{ for } i = 1, \dots, P, \quad (3)$$

by first applying a dimensionality reduction approach, such as PCA, to each omics matrix X^i separately and then maximizing their co-inertia, i.e., the sum of the squared covariance between scores of each factor:

$$\operatorname{argmax}_{\mathbf{q}_1^1 \dots \mathbf{q}_1^P} \sum_{k=1}^P \operatorname{cov}^2 (X_k^i q_k^i, \mathbf{X}^i \mathbf{q}^i) \quad (4)$$

- 1、基于协惯性分析（co-inertia analysis, CIA）；
- 2、首先对每个组学矩阵 X^i 分别应用降维方法（例如 PCA），然后最大化它们的协惯性，即每个因子得分之间的平方协方差之和
- 3、只考虑了特定组学的因子，没有考虑联合因素

方法4: *Regularized generalized canonical correlation analysis (RGCCA)*

正则化广义典型相关分析

方法5: *iCluster*

方法6: *Multimics factor analysis (MOFA)*

多组学因子分析

$$\mathbf{X}^i = \mathbf{A}^i \mathbf{F} + \mathbf{E}^i, \text{ for } i = 1, \dots, P.$$

1、RGCCA 通过找到投影向量来最大化组学特定因素之间的相关性，使得投影数据具有最大相关性

$$\operatorname{argmax}_{i,j} \operatorname{Corr}(\mathbf{X}^i \mathbf{u}^i, \mathbf{X}^j \mathbf{u}^j) \text{ for all possible couples } i, j = 1, \dots, P.$$

考虑各组学的特定因子

2、icluster ,假设误差 E^i 和因子矩阵 F 都是正态分布的，先导出相同方程的基于似然的公式，然后应用期望最大化 (EM) 来求解该方程。最后，通过应用 K-means 从因子矩阵中获得聚类。

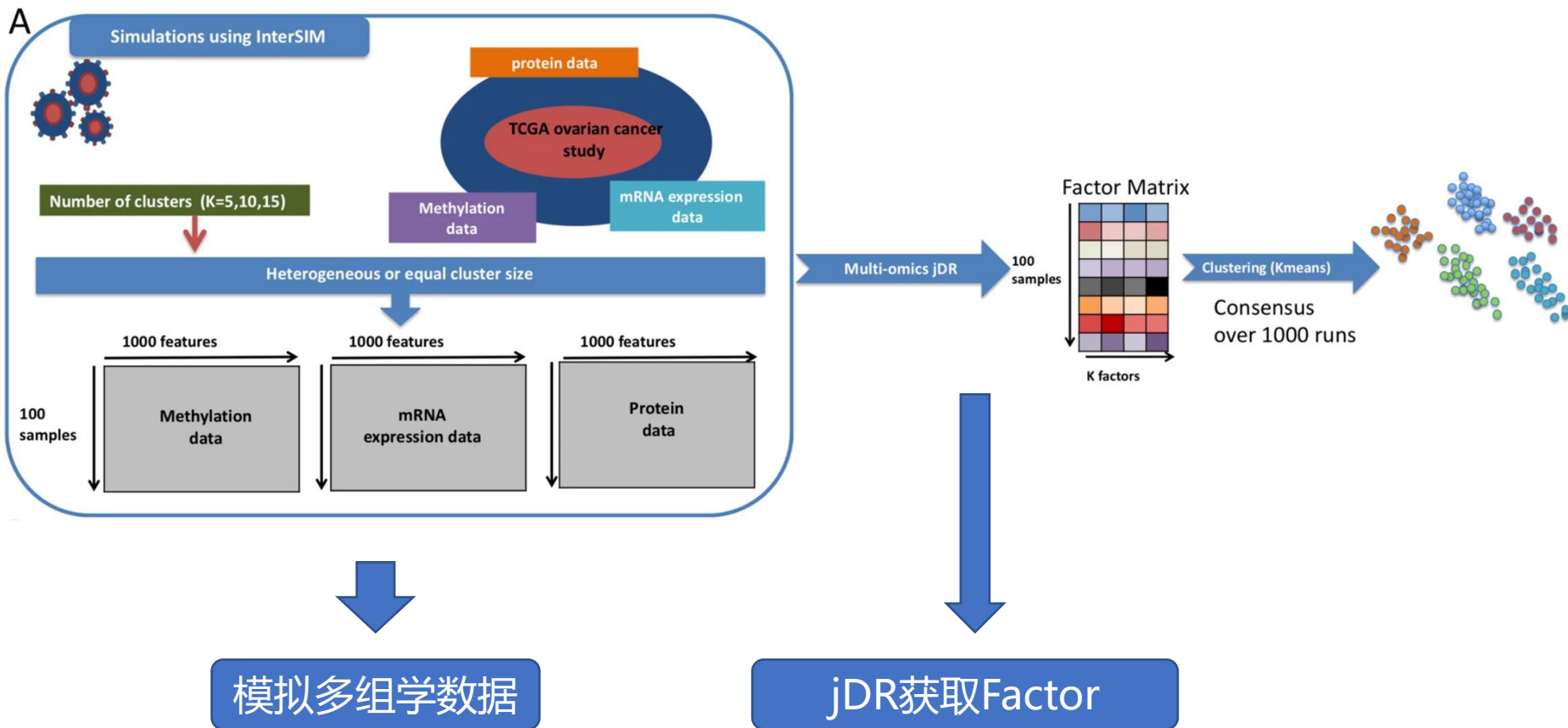
3、MOFA 首先在概率贝叶斯模型中制定上述方程，对所有未观察到的变量 A_i 、 F 、 E_i 进行先验分布。



Joint Dimensionality Reduction approaches and principles

联合降维方法和原则

jDR approach	Underlying approach	Constraints on the factors	Features or samples matching requirements
RGCCA	Canonical Correlation Analysis (CCA)	omics-specific factors	matching of samples
MCIA	Co-Inertia Analysis (CIA)	omics-specific factors	matching of samples
MOFA	Factor Analysis (FA) (Bayesian)	mixed factors	matching of samples (partial matching allowed)
MSFA	Factor Analysis (FA) (Bayesian)	mixed factors	matching of samples
intNMF	Non-Negative Matrix Factorization (NMF)	shared factors	matching of samples
iCluster	Gaussian latent variable model	shared factors	matching of samples
JIVE	Principal Component Analysis (PCA)	mixed factors	matching of samples (partial matching allowed)
tICA	Independent Component Analysis (ICA)	shared factors	matching of both samples and features (tensor)
Scikit-fusion	Matrix tri-factorization	shared factors	matching of samples

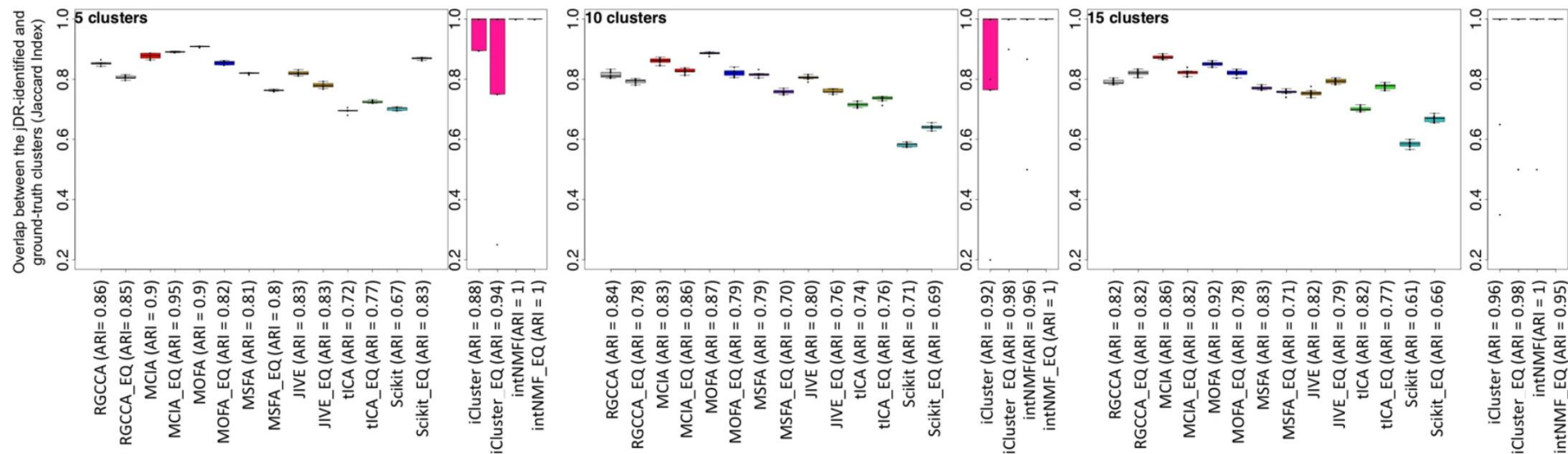


InterSIM基于来自 TCGA 卵巢癌数据的 DNA 甲基化、mRNA 基因表达和蛋白质表达，模拟具有现实内部和相互关系的多种相互关联的数据类型。

1、考虑了五个、十个和十五个集群

2、考虑了具有相同大小的集群和可变随机大小的两个条件

b



$$J(A,B) = \frac{A \cap B}{A \cup B}$$

$$ARI = \frac{RI - \text{expected value } RI}{\max RI - \text{expected value } RI}$$

- 1、所有方法在不同的模拟场景中都表现得相当好 ($JIs \geq 0.6$, $ARI \geq 0.6$)
- 2、本质上为聚类设计的两种算法, 即 intNMF 和 iCluster, 显示出最佳性能。特别是, intNMF 可以完美地检索到真实聚类 ($JI \sim 1$, $0.9 < ARI \leq 1$)
- 3、tICA 和 scikit-fusion 是该基准测试中效果较差的方法。

考虑到, tICA 将多组学数据结构化为张量。获得这些张量, 要将组学数据转换为相关矩阵, 这可能会导致信息丢失。

scikit-fusion 旨在处理辅助信息, 用于构建连接各种实体 (例如, 样本、基因、蛋白质) 的关系网络。但是, 为了与其他 jDR 方法进行比较, 没有考虑边信息, 这可能会影响算法的结果。



Benchmarking joint Dimensionality Reduction approaches on cancer datasets

对癌症数据集的联合降维方法进行基准测试

下载了十种不同癌症类型的 TCGA 多组学数据，由三个组学层组成：基因表达、DNA 甲基化和 miRNA 表达。样本数量从急性髓性白血病 (AML) 的 170 个到乳腺癌的 621 个不等。而且，没有有关癌症亚型的 ground-truth 来评估 jDR 方法的性能。

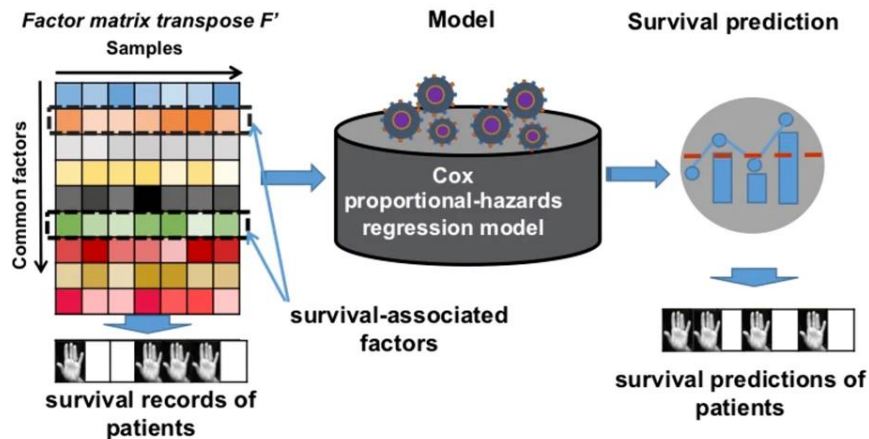
为了评估全套癌症多组学的方法，测试了这些因素与生存、临床注释和生物学注释的关联。由于多因子分析方法 MSFA 没有收敛到任何解决方案，因此没有被考虑。**将剩余的八种 jDR 方法应用于十个癌症多组学数据集中的每一个，将它们共同分解为十个因素。**

比较了八种 jDR 算法在识别预测生存因素的能力以及与临床注释相关的因素方面的性能。还通过评估它们在已知生物途径和过程中的富集来评估由 jDR 方法产生的权重矩阵。



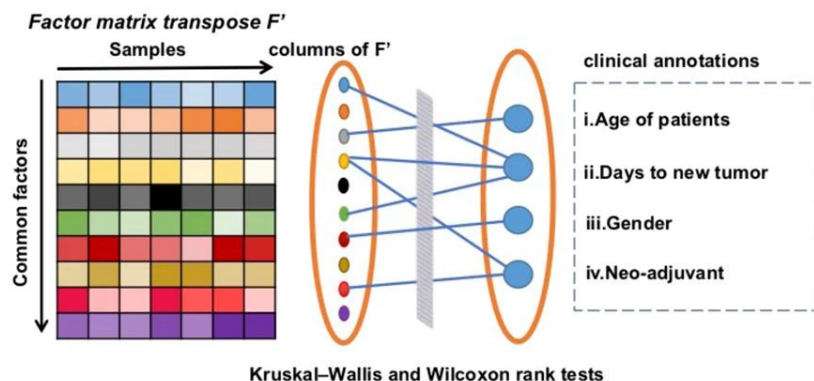
a 测试 jDR 因素与生存的关联

A. Assessing the association of the jDR factors with survival

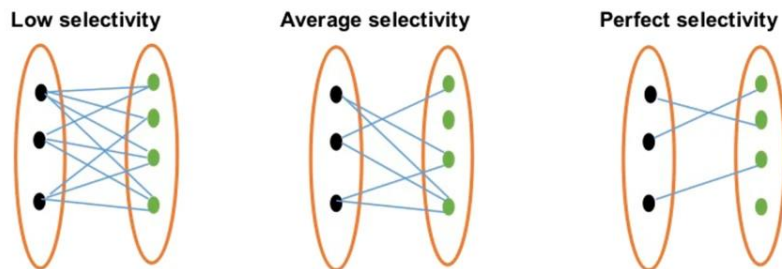


b 测试 jDR 因素与临床注释的关联

B. Assessing the association of the jDR factors with clinical annotations

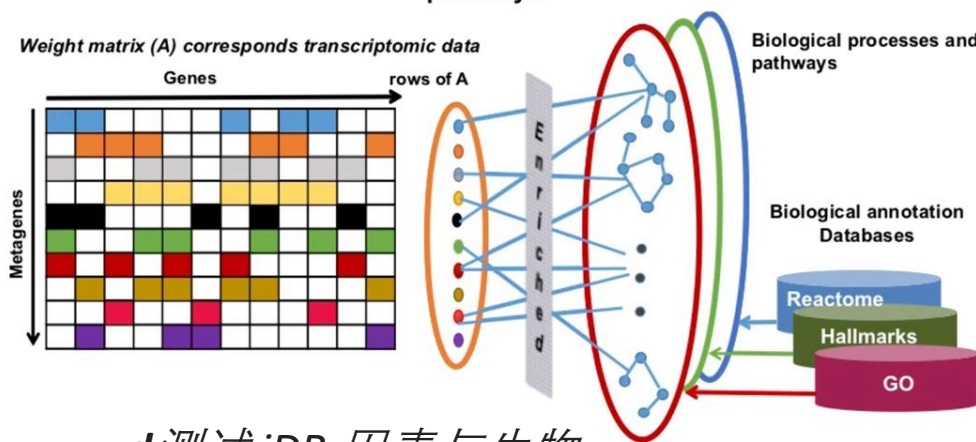


C. Selectivity: Measuring a one-to-one mapping between a factor/metagene and a clinical/biological annotation



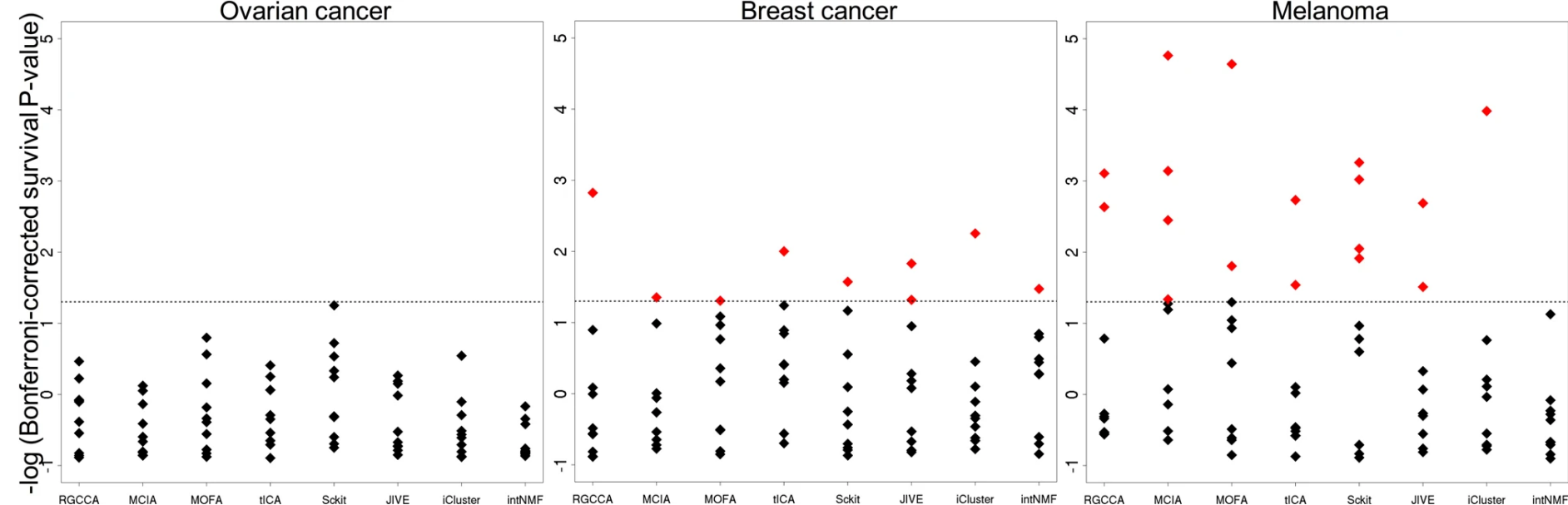
c 选择性评分的图形解释：测量因素和临床/生物学注释之间的一对一映射

D. Assessing the association of the jDR factors with biological processes and pathways



d 测试 jDR 因素与生物过程和途径的关联

通过 jDR 方法鉴定卵巢癌、乳腺癌和黑色素瘤癌症样本中的存活预测因子



使用了 Cox 比例风险回归模型

- 1、与生存相关的因素的数量更多地取决于癌症类型而不是 jDR 算法
- 2、对于三种癌症类型（结肠癌、肺癌和卵巢癌），没有一种 jDR 方法能够识别与生存相关的因素；在其他四种癌症类型（肉瘤、肝癌、肾癌和乳腺癌）中，所有 jDR 算法都确定了一个或两个与生存相关的因素；在黑色素瘤、GBM 和 AML 中，大多数 jDR 方法确定了三个或四个与生存相关的因素。
- 3、总体而言，MCIA、RGCCA 和 JIVE 表现最好，在十种癌症类型中的七种中发现了与生存率显著相关的因素。



Briefings in Bioinformatics

Issues

Advance articles

Submit ▼

Purchase

Alerts

About ▼

All Briefings in B



Volume 23, Issue 1

January 2022

(In Progress)

BioNet: a large-scale and heterogeneous biological network model for interaction prediction with graph convolution

Xi Yang, Wei Wang, Jing-Lun Ma, Yan-Long Qiu, Kai Lu, Dong-Sheng Cao ✉, Cheng-Kun Wu ✉

Briefings in Bioinformatics, Volume 23, Issue 1, January 2022, bbab491,

<https://doi.org/10.1093/bib/bbab491>

Published: 25 November 2021 **Article history** ▼

BioNet: 一种大规模的异构生物网络模型——使用图卷积进行交互预测

本文概述

作者开发了BioNet，通过整合与化合物、基因、生物通路和疾病相关的数据集，构建了一个大规模的异质生物相互作用网络，并**基于编码器-解码器架构提出了一种深度图卷积网络模型**。

该模型利用图卷积编码器从子图中学习网络中节点的嵌入表示，并采用张量分解解码器计算化合物-基因相互作用的概率；

为了解决大规模图模型训练的效率问题，开发对应的**并行策略**，有效加速了模型的计算过程并提升了模型的可扩展性。

最后，利用BioNet预测与**中风和癌症相关的化合物-基因相互作用**，并通过查找文献和对比相关实验数据的方式证明了结果的可靠性。BioNet 表现出出色的预测性能，接收器操作特征 (ROC) 曲线下的最佳面积为 0.952，显著超过了已有方法。

研究人员利用BioNet进行了针对新冠关键靶标的智能药物推荐，得到的结果与基于分子动力学模拟的方法得到的结果在较大程度上可以互相印证。



nature medicine

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature medicine](#) > [articles](#) > [article](#)

[Article](#) | [Open Access](#) | [Published: 07 February 2022](#)

Long-term cardiovascular outcomes of COVID-19

[Yan Xie](#), [Evan Xu](#), [Benjamin Bowe](#) & [Ziyad Al-Aly](#)

[Nature Medicine](#) (2022) | [Cite this article](#)

415k Accesses | **5** Citations | **15275** Altmetric | [Metrics](#)

长期 COVID-19 的下对心血管的影响



本文概述

使用美国退伍军人事务部的国家医疗保健数据库建立了一个由 153,760 名 COVID-19 患者组成的队列，以及两组分别有 5,637,647 名（同时期对照）和 5,859,411 名（历史对照）个体的对照队列，以估计风险和一组预先指定的心血管事件的 1 年负担。

研究表明，在感染后的前 30 天后，**COVID-19 患者发生心血管疾病的风险增加，包括脑血管疾病、心律失常、缺血性和非缺血性心脏病、心包炎、心肌炎、心力衰竭和血栓栓塞性疾病。**即使在感染急性期末住院的个体中，这些风险和负担也很明显，并且根据急性期的护理环境（未住院、住院和重症监护）以分级方式增加。

建议COVID-19 急性发作后幸存者的护理途径应包括**关注心血管健康和疾病**（住院并进入重症监护室）



nature communications

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature communications](#) > [articles](#) > [article](#)

Article | [Open Access](#) | [Published: 14 February 2022](#)

A deep-learning system bridging molecule structure and biomedical text with comprehension comparable to human professionals

[Zheni Zeng](#), [Yuan Yao](#), [Zhiyuan Liu](#) ✉ & [Maosong Sun](#) ✉

[Nature Communications](#) **13**, Article number: 862 (2022) | [Cite this article](#)

4538 Accesses | **13** Altmetric | [Metrics](#)

一个连接分子结构和生物医学文本的深度学习系统，其理解力可与人类专业人士媲美

本文概述

为了加速生物医学研究过程，开发了深度学习系统，通过读取大规模生物医学数据来自自动获取有关分子实体的知识。

受人类从分子结构和生物医学文本信息的多功能阅读中学习深层分子知识的启发，本文提出了一个机器阅读系统，该系统在**统一的深度学习框架中连接这两种类型的信息**，以提供全面的生物医学研究援助。

研究人员解决了现有机器阅读模型只能分别处理不同类型数据的问题，从而实现对分子实体的全面透彻的理解。

通过在不同信息源内和跨不同信息源以无人监督的方式掌握元知识，该系统可以促进各种现实世界的生物医学应用，包括分子性质预测、生物医学关系提取等。实验结果表明，我们的系统在分子特性理解能力方面甚至超过了人类专业人员，也揭示了其在促进未来自动药物发现和记录方面的巨大潜力。



nature metabolism

Explore content ▾

About the journal ▾





Publish with us ▾

Subscribe

[nature](#) > [nature metabolism](#) > [articles](#) > [article](#)

Article | [Published: 21 October 2019](#)

A nutritional memory effect counteracts the benefits of dietary restriction in old mice

[Oliver Hahn](#), [Lisa F. Drews](#), [An Nguyen](#), [Takashi Tatsuta](#), [Lisonia Gkioni](#), [Oliver Hendrich](#), [Qifeng Zhang](#), [Thomas Langer](#), [Scott Pletcher](#), [Michael J. O. Wakelam](#) , [Andreas Beyer](#) , [Sebastian Grönke](#)  & [Linda Partridge](#) 

[Nature Metabolism](#) **1**, 1059–1073 (2019) | [Cite this article](#)

12k Accesses | **31** Citations | **402** Altmetric | [Metrics](#)

营养记忆效应抵消了老年小鼠饮食限制的好处



本文概述

成年期的饮食限制 (DR) 可以极大地延长寿命并改善不同物种的代谢健康。然而，在哺乳动物中首次应用 DR 是否仍然有效仍是未知数。

在该研究中，研究人员使用了800只小鼠进行了晚期饮食限制转换的实验。24个月大的高龄雌鼠从随意进食改为进食限制，反之亦然。

如果小鼠在成年后的进食量比随意进食组减少40%，那么它们在老年时活得更长，也更健康。进食限制组小鼠被喂以富含维生素和矿物质的食物，以防止营养不良；令人惊讶的是，从进食限制组改为随意进食会极大地增加死亡率；另一方面，在减少食物摄入一段时间后，如果允许小鼠改为随意进食，则对它们没有长期的保护作用，所以必须持续减少食物摄入才能使小鼠获得长期益处。

因此，减少食物摄入量必须尽早实施，并持续到他们的生命终结，才能对老年健康产生终极的有益影响。



nature communications

[Explore content](#) ▾ [About the journal](#) ▾ [Publish with us](#) ▾

[nature](#) > [nature communications](#) > [articles](#) > article

Article | [Open Access](#) | [Published: 04 March 2022](#)

Associations between alcohol consumption and gray and white matter volumes in the UK Biobank

[Remi Daviet](#) ✉, [Gökhan Aydogan](#), [Kanchana Jagannathan](#), [Nathaniel Spilka](#), [Philipp D. Koellinger](#), [Henry R. Kranzler](#), [Gideon Nave](#) ✉ & [Reagan R. Wetherill](#) ✉

酒精消费与灰质和白质体积之间的关联



本文概述

研究团队分析了英国生物样本库（UK Biobank）中超过36000名成年人的数据后发现，饮酒和脑容量减少有关，即使是轻度到中度的饮酒也会对大脑造成伤害，平均每天1瓶啤酒的饮酒量，对大脑的伤害就相当于衰老两年。

研究团队对英国生物样本库的数据进行了分析，该样本库包含了50万英国成年人的遗传和医学信息，并特别观察了其中36000多人的大脑磁共振成像（MRI）数据。这些数据可以用来计算大脑不同区域的白质和灰质体积。

每天0到1个酒精单位对脑容量没有明显影响，但每天1到2个酒精单位或2到3个酒精单位与大脑灰质和白质减少有关。而且这种影响不是线性的，而是指数级的，喝酒越多，情况越糟。

 1 UNIT Single shot of spirits (25ml, ABV 40%)	 1.5 UNITS Alcopop (275ml, ABV 5.5%)	 1.5 UNITS Small glass of red / white / rosé / sparkling wine (125ml, ABV 12%)	 2 UNITS Can of beer, ale, lager or cider (440ml, ABV 5.5%)
 2.1 UNITS Standard glass of red / white / rosé / wine (175ml, ABV 12%)	 3 UNITS Pint of beer, ale, lager or cider (568ml, ABV 5.2%)	 3 UNITS Large glass of red / white / rosé / wine (250ml, ABV 12%)	 9 UNITS Bottle of red / white / rosé / sparkling wine (750ml, ABV 12%)

美国国家酒精滥用和酒精中毒研究所建议女性每天饮酒量不超过1个酒精单位，男性不超过2个酒精单位。