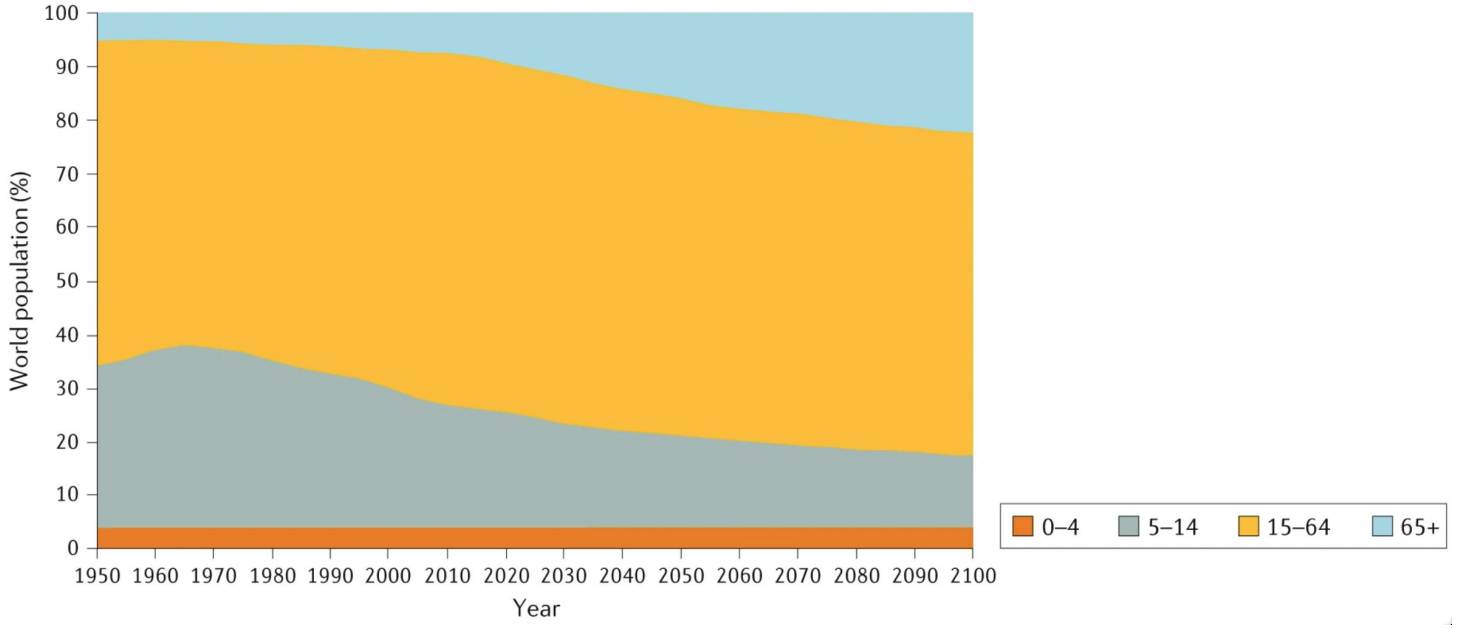
MAPLE:
一种稳定且精准的甲基化年龄预测与疾病风险评估框架

2024 年 10 月 29 日

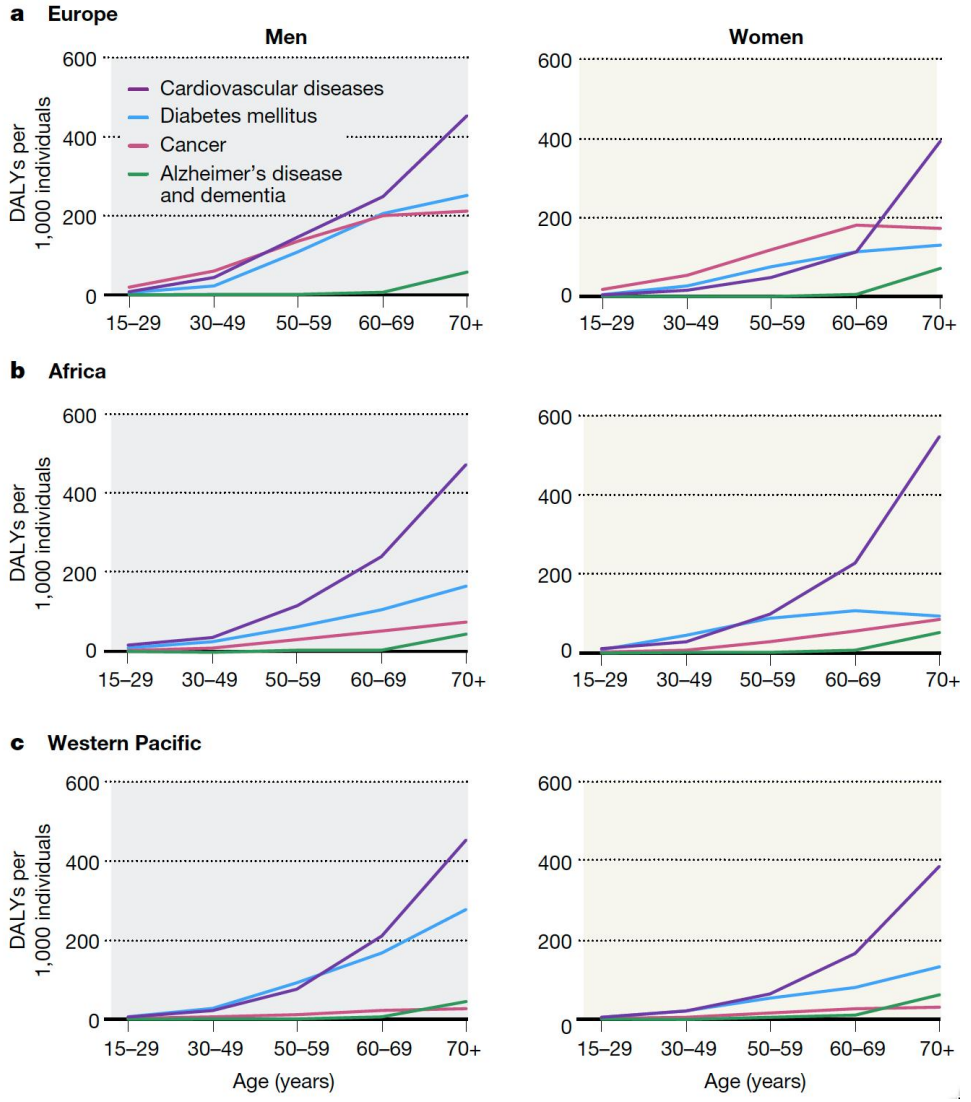
介绍内容

老龄化社会的挑战与抗衰老产业的机遇

全球老龄化趋势加速



衰老导致各种疾病的发生，降低居民生活质量

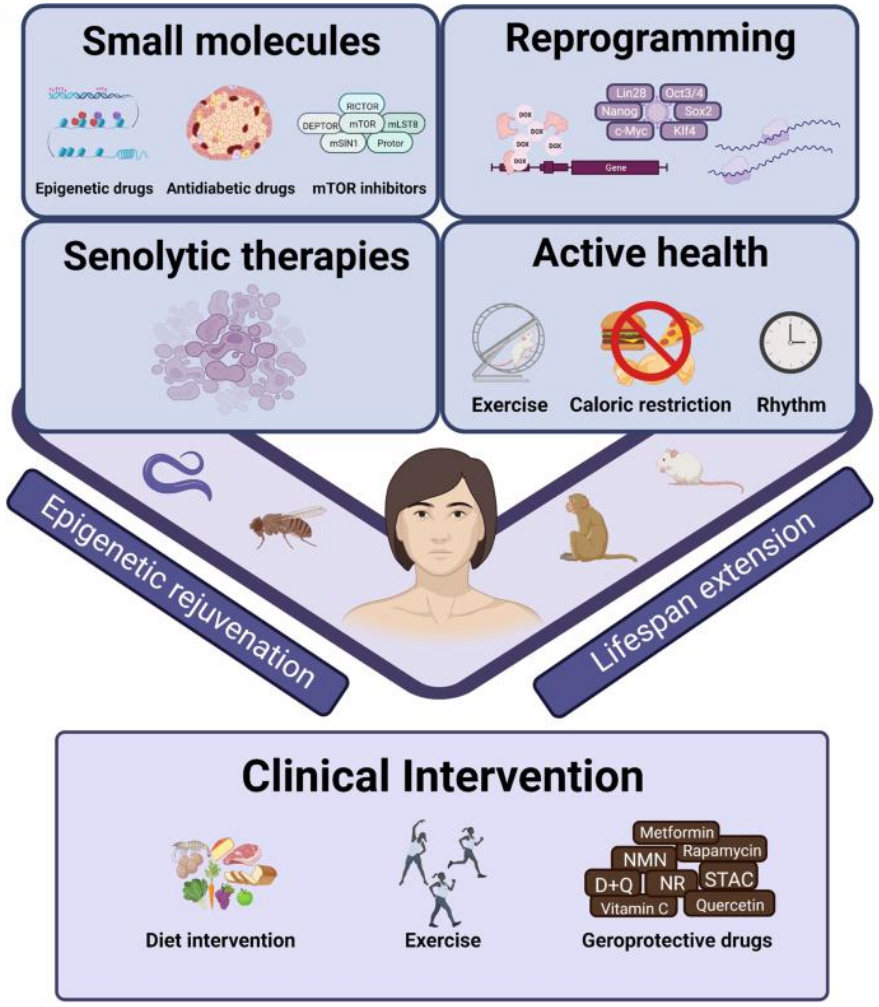


Partridge, L., Fuentealba, M. & Kennedy, B.K. The quest to slow ageing through drug discovery. *Nat Rev Drug Discov* **19**, 513–532 (2020).

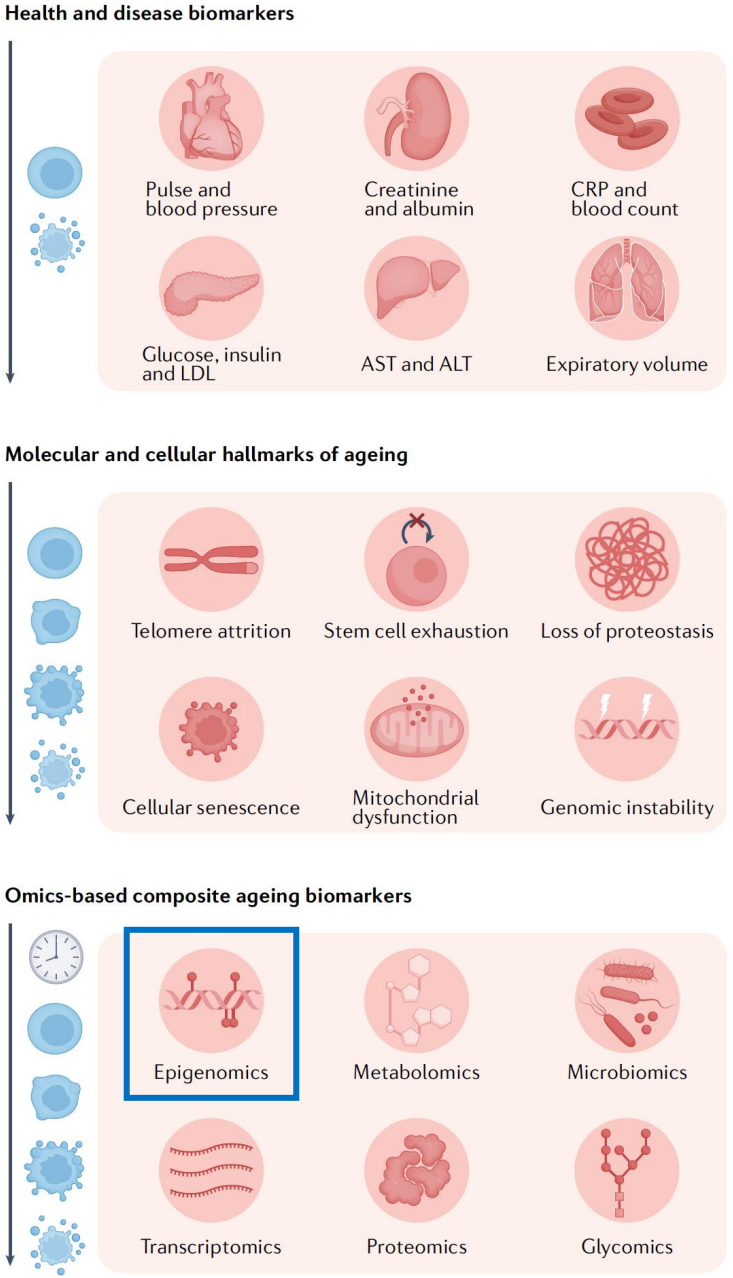
Partridge, L., Deelen, J. & Slagboom, P.E. Facing up to the global challenges of ageing. *Nature* **561**, 45–56 (2018).

衰老干预及生物学年龄度量

衰老干预



生物学年龄度量指标



临床上对疾病的诊断指标：
- 如果指标异常，则表明身体进入疾病状态；
- 不能对人体衰老进程进行度量

衰老在分子与细胞层面的表现：
- 无法进行度量，或对衰老的度量不够精准

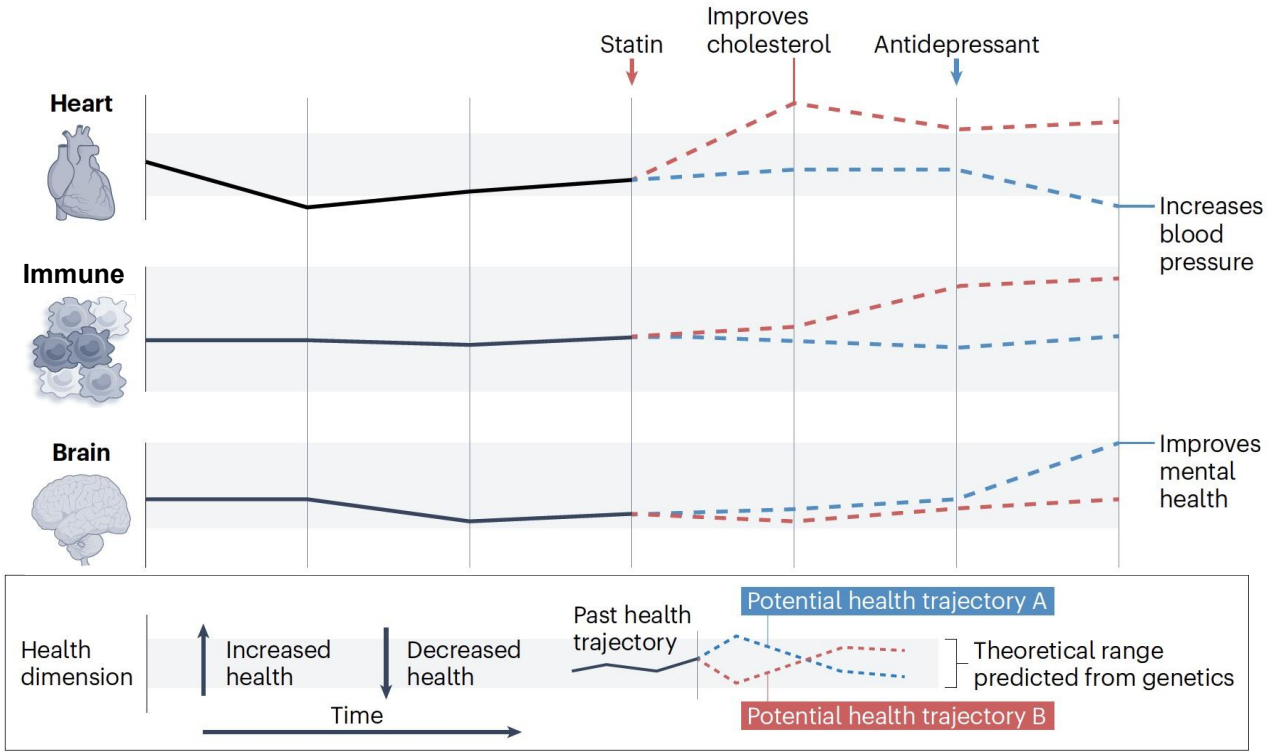
通过组学数据度量衰老：
- 组学数据反映的是衰老过程在原因层面的信息
- 测量的特征数量多，方便建模；
- DNA 甲基化最为稳定

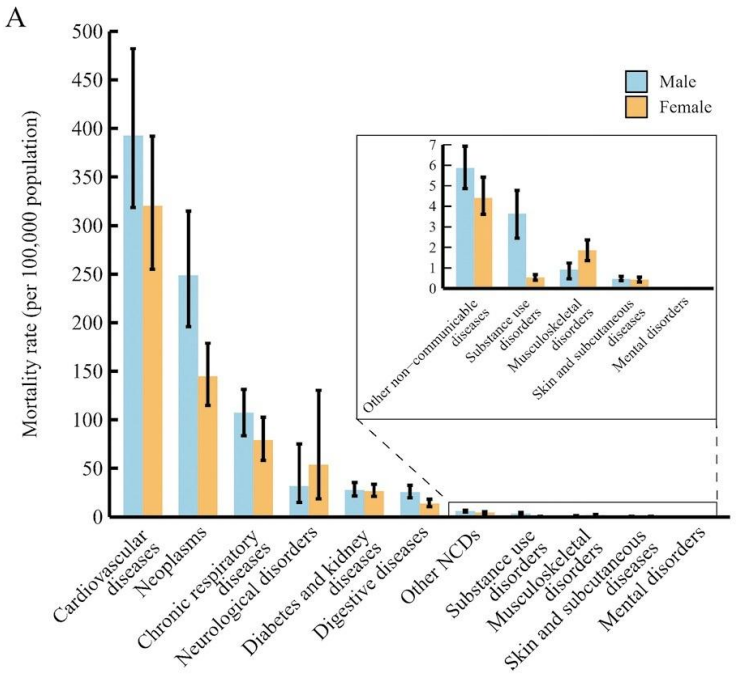
Wang, K., Liu, H., Hu, Q. *et al.* Epigenetic regulation of aging: implications for interventions of aging and diseases. *Sig Transduct Target Ther* 7, 374 (2022).
Rutledge, J., Oh, H. & Wyss-Coray, T. Measuring biological age using omics data. *Nat Rev*

对衰老精细预测，精准干预

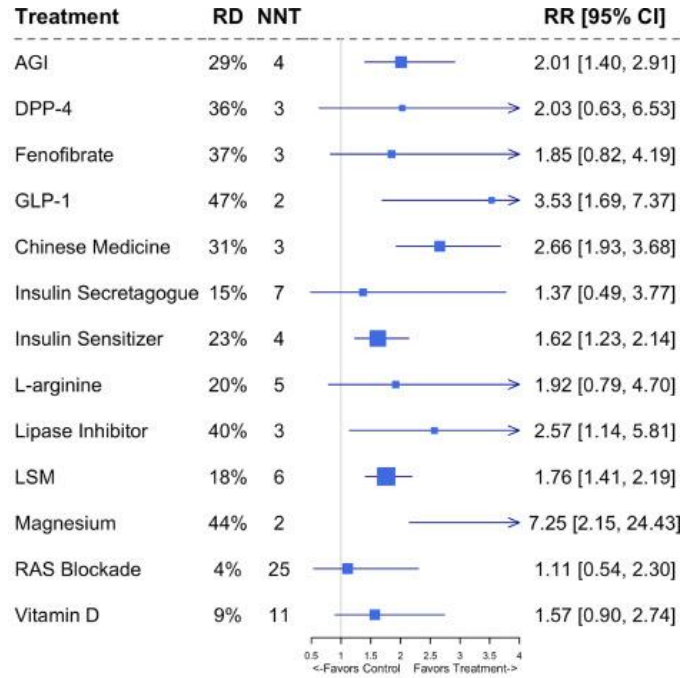
按器官或功能对人体衰老进行更加精细的预测

- **对人体衰老进行更加精细的划分与描述，如心脑血管衰老、糖代谢衰老、免疫衰老、皮肤衰老等**
- **根据个性化衰老检测结果，进行精准的衰老干预**





心血管疾病和糖尿病分别是中国人的第一和第五大致死疾病



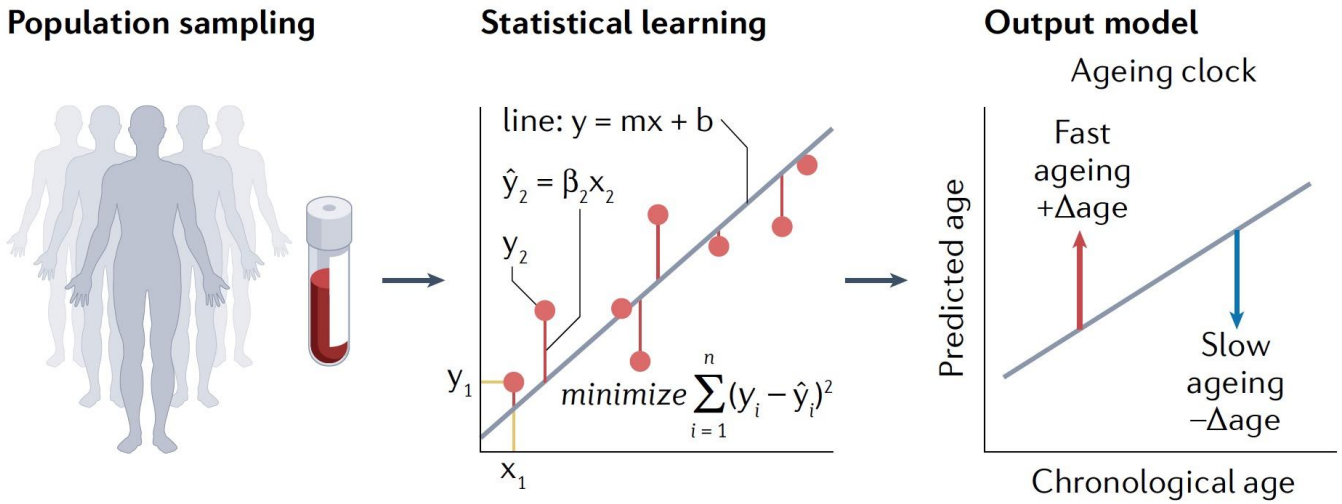
在疾病前驱状态进行干预能够有效防止疾病发生

Liu, Hanxiao; Yin, Peng; Qi, Jinlei *et al.* Burden of non-communicable diseases in China and its provinces, 1990–2021: Results from the Global Burden of Disease Study 2021. Chinese Medical Journal 137(19):p 2325-2333, October 05, 2024.

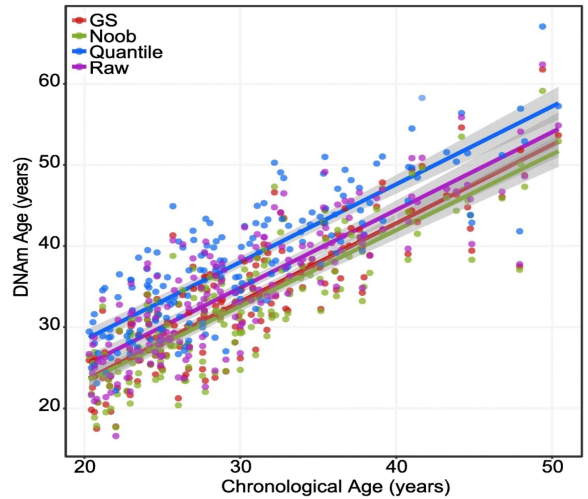
Galaviz KI, Weber MB, Suvada K BS et al. Interventions for Reversing Prediabetes: A Systematic Review and Meta-Analysis. Am J Prev Med. 2022 Apr;62(4):614-625.

现有甲基化时钟的局限性

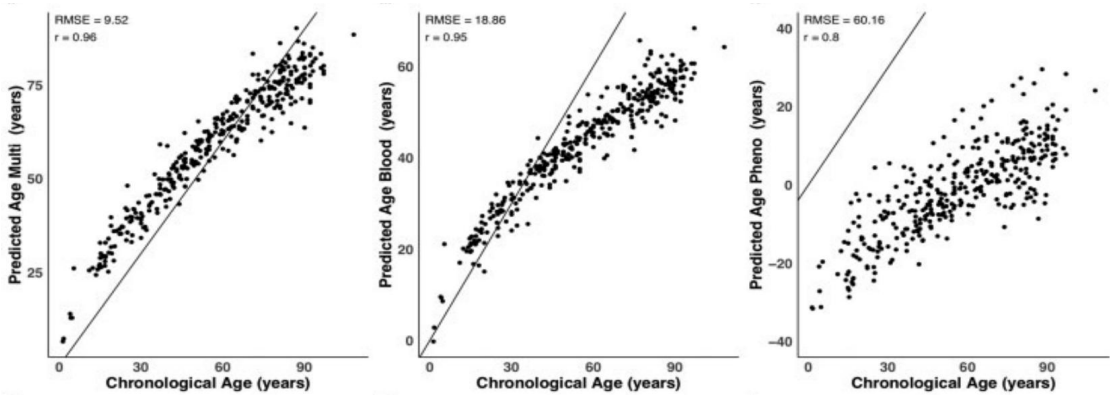
甲基化时钟的构建流程



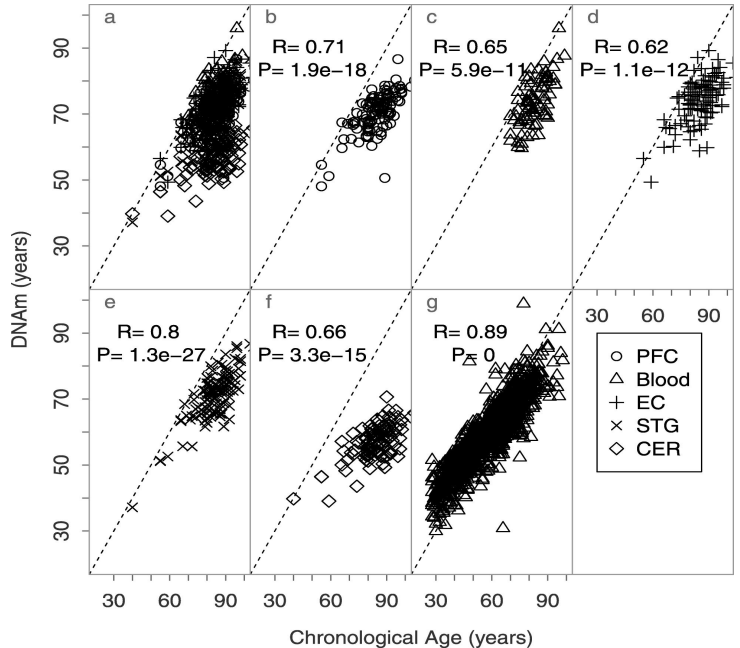
预处理流程的影响



人群差异的影响（老年人预测效果不好）



组织类型的影响



McEwen, L.M. et al. Systematic evaluation of DNA methylation age estimation with common preprocessing methods and the Infinium MethylationEPIC BeadChip array. Clin Epigenetics 10, 123 (2018)

El Khoury, L.Y. et al. Systematic underestimation of the epigenetic clock and age acceleration in older subjects. Genome Biol 20, 283 (2019).

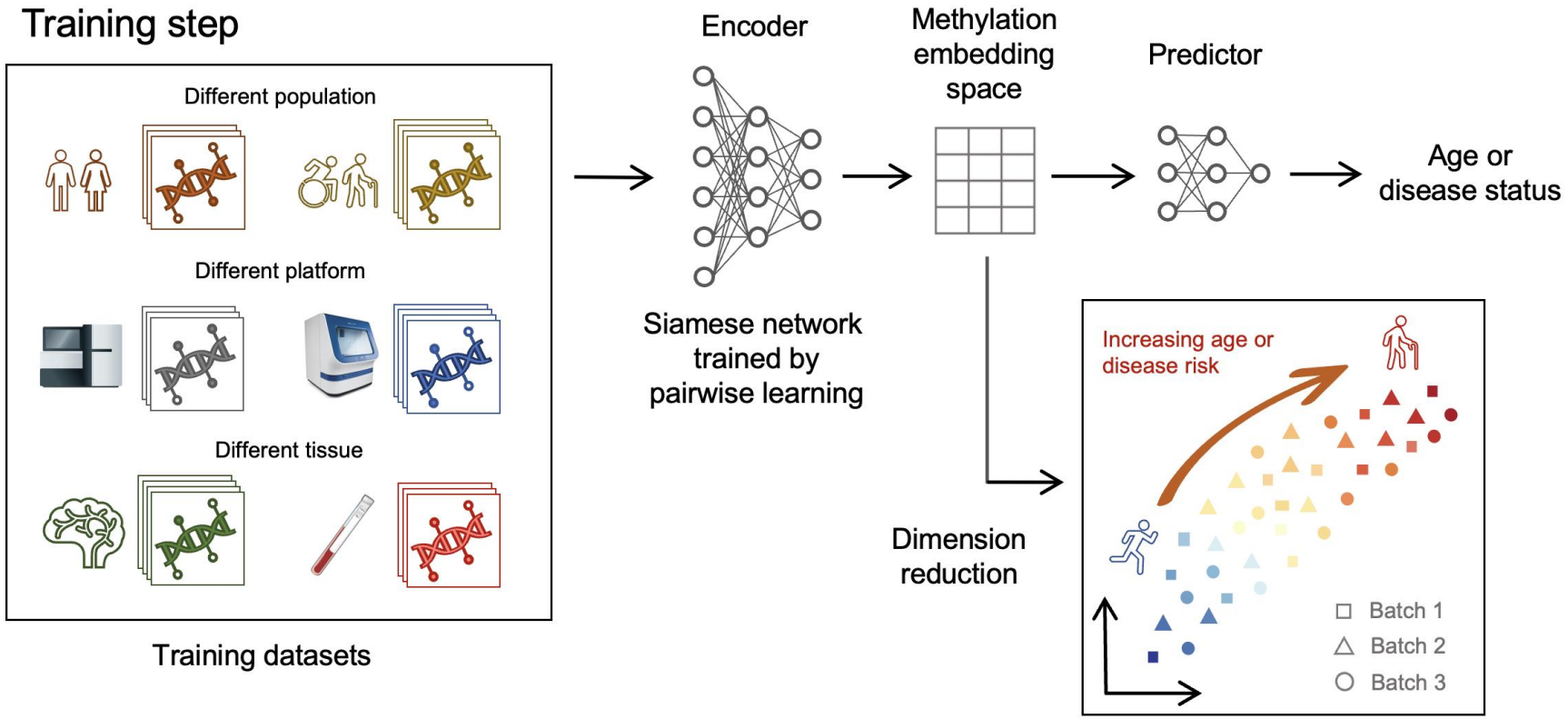
Shireby, G.L. et al. Recalibrating the epigenetic clock: implications for assessing biological age in the human cortex. Brain 143, 3763-3775 (2020).

介绍内容

使用对比学习实现生物学年龄与疾病风险的预测

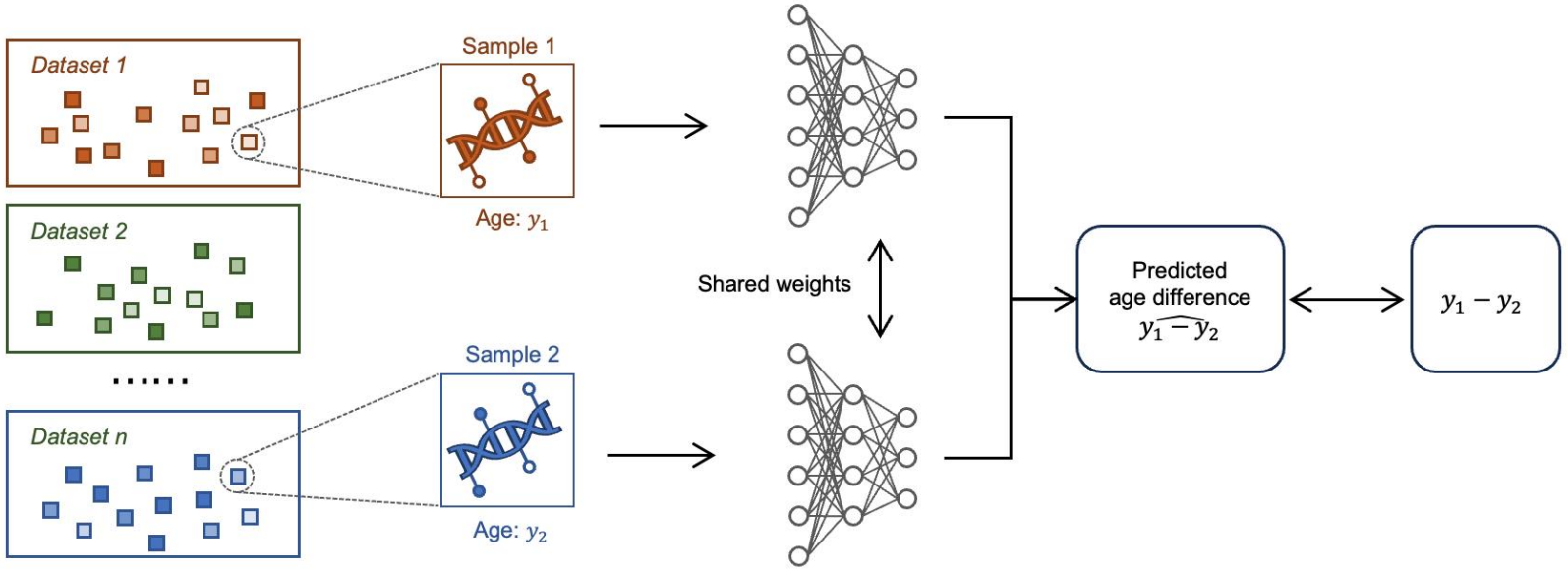
INF无限光年

MAPLE: a framework for **M**ethylation **A**ge and disease-risk **P**rediction based on pairwise **LE**arning

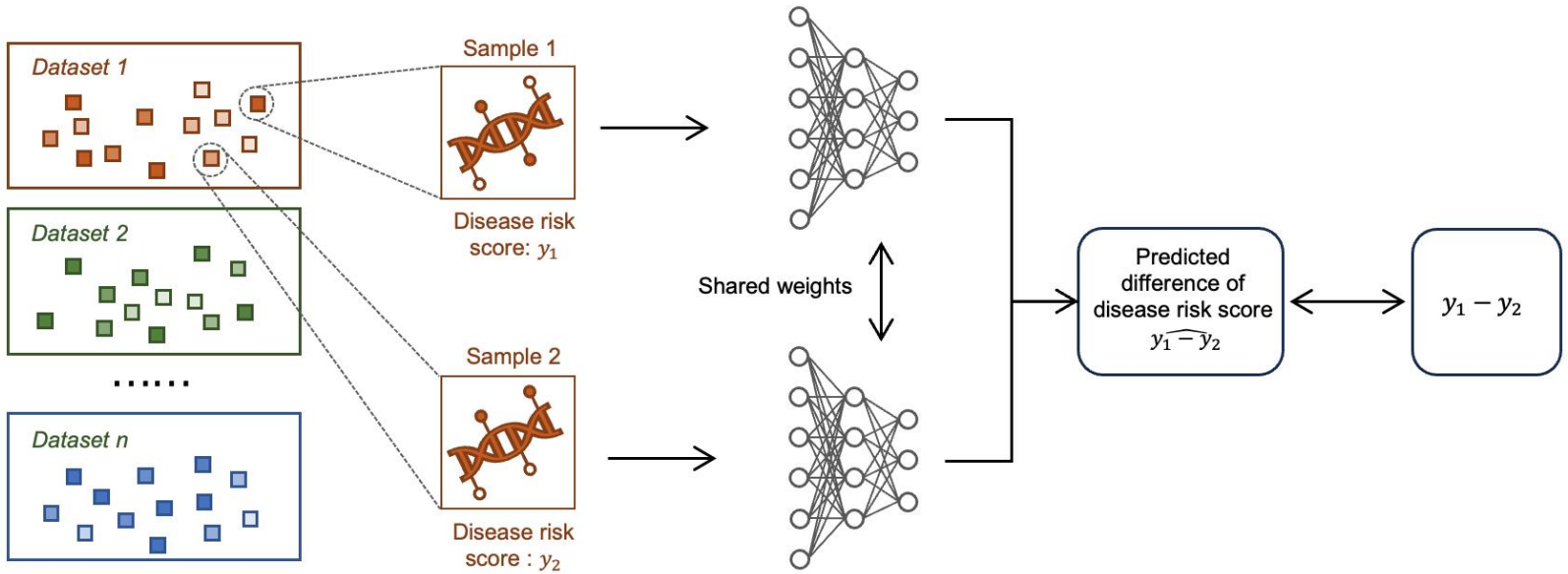


MAPLE 通过对比学习消除不同来源的甲基化数据之间的批次效应

在年龄预测任务和疾病风险预测任务中使用不同的对比策略

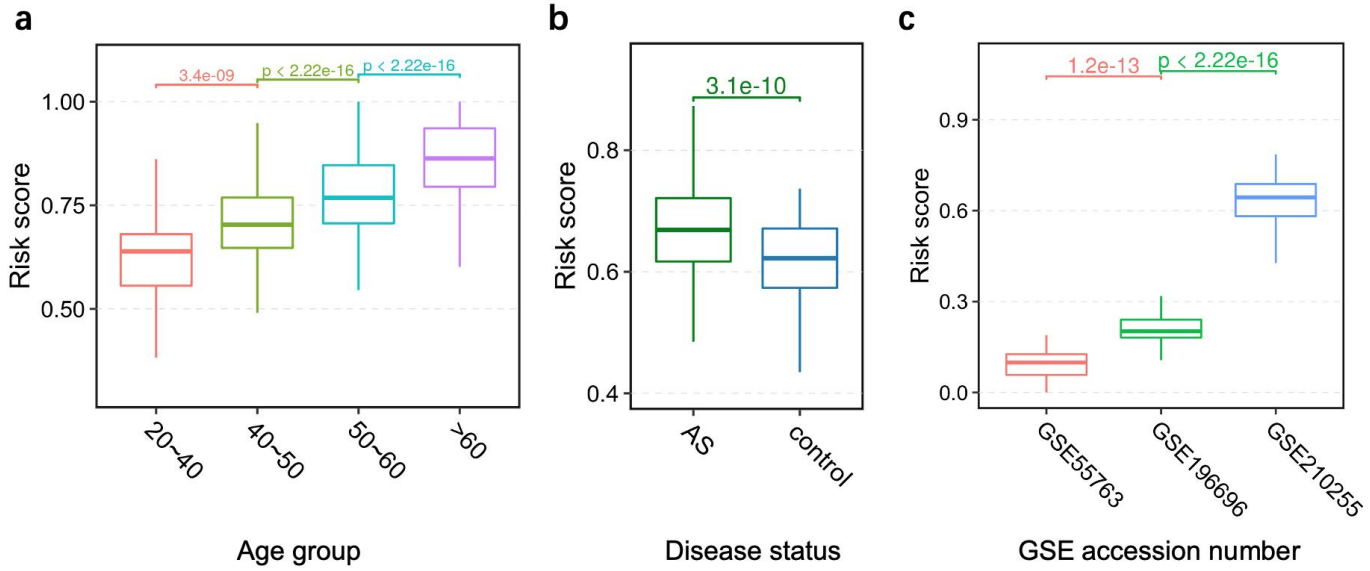


年龄预测模型：对组间与组内样本同时进行对比

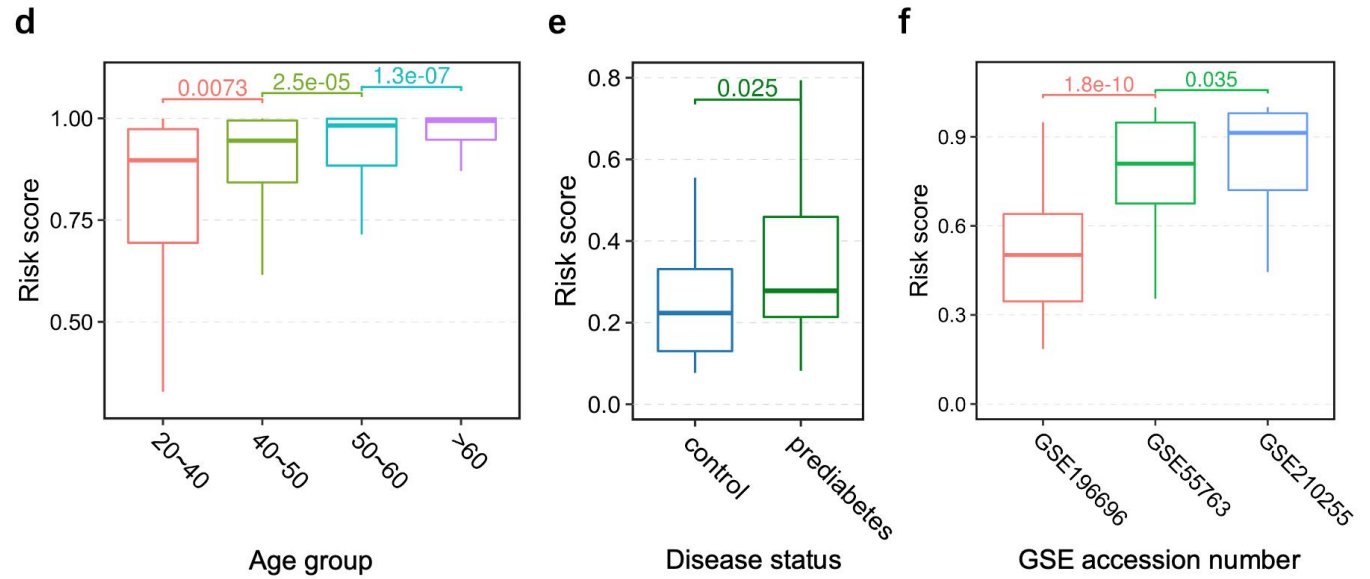


疾病风险预测模型：仅对组内样本之间进行对比

疾病风险分数的标注仅在同一数据集中可比



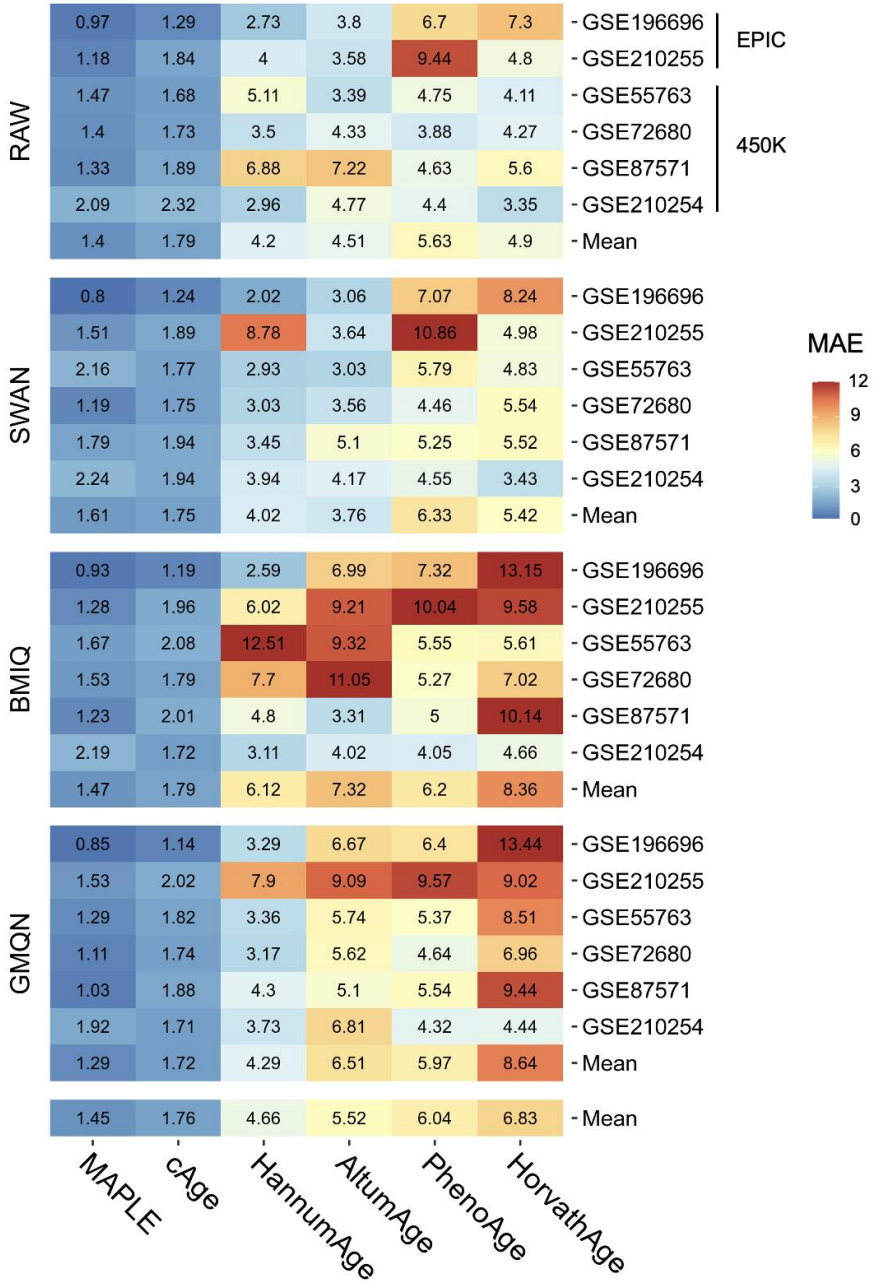
年龄预测模型：对组间与组内样本同时进行对比



疾病风险预测模型：仅对组内样本之间进行对比

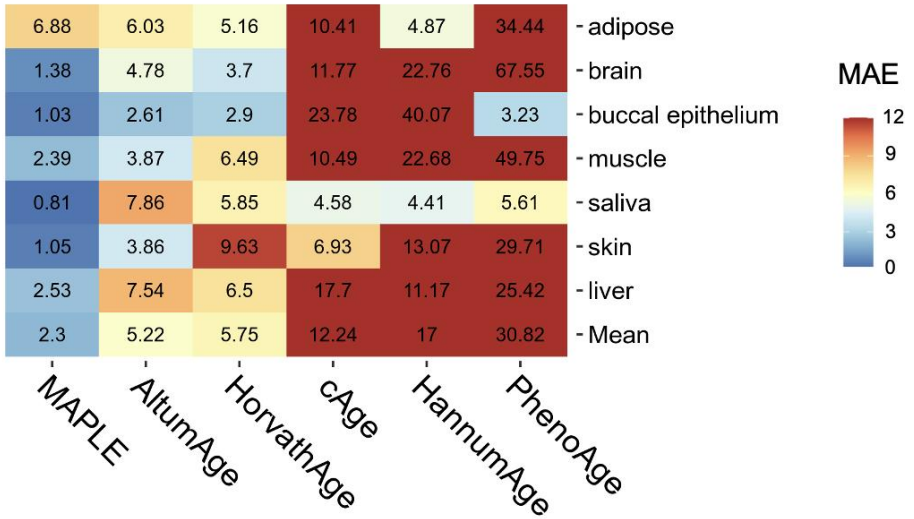
介绍内容

MAPLE 在年龄预测任务上达到国际领先水平



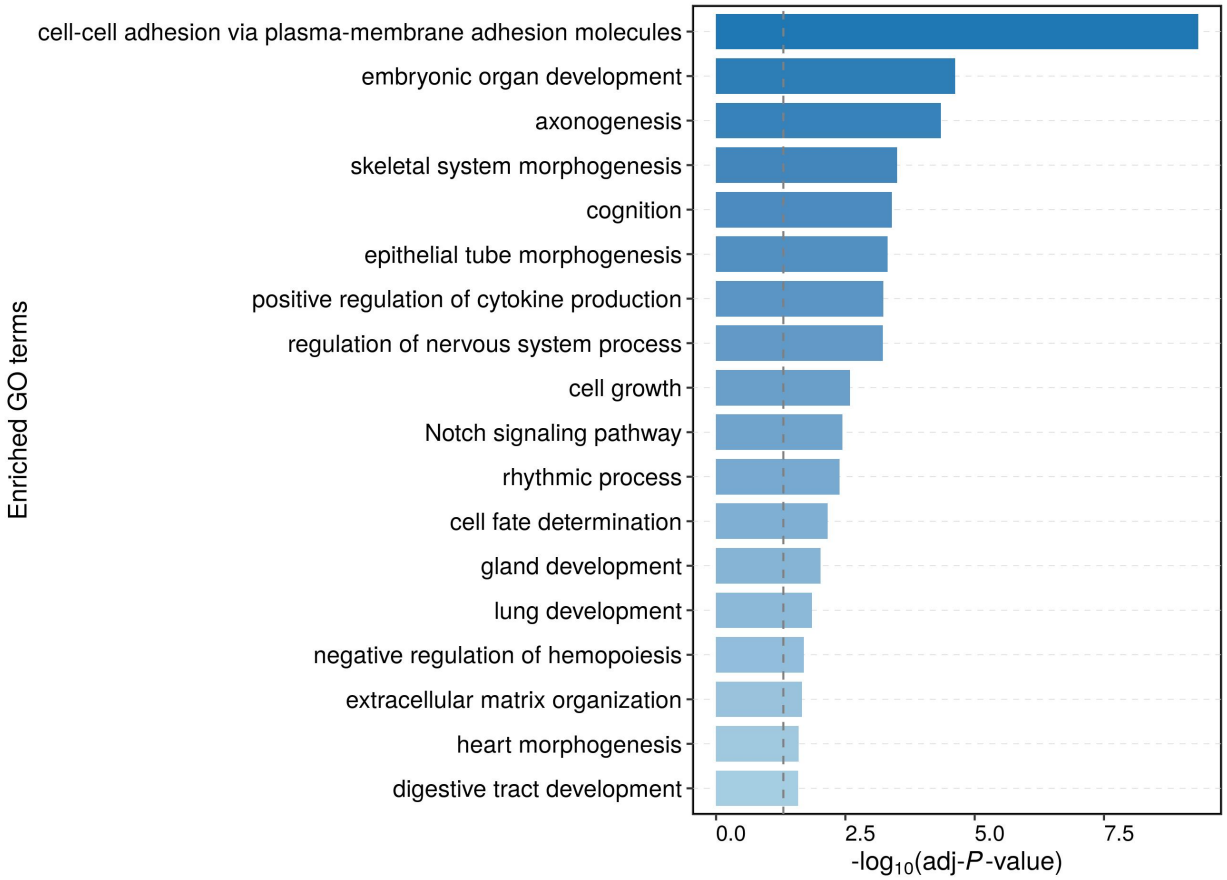
在血液样本上测试多中心、多平台、多数据预处理方法的预测结果

非血液组织的预测结果

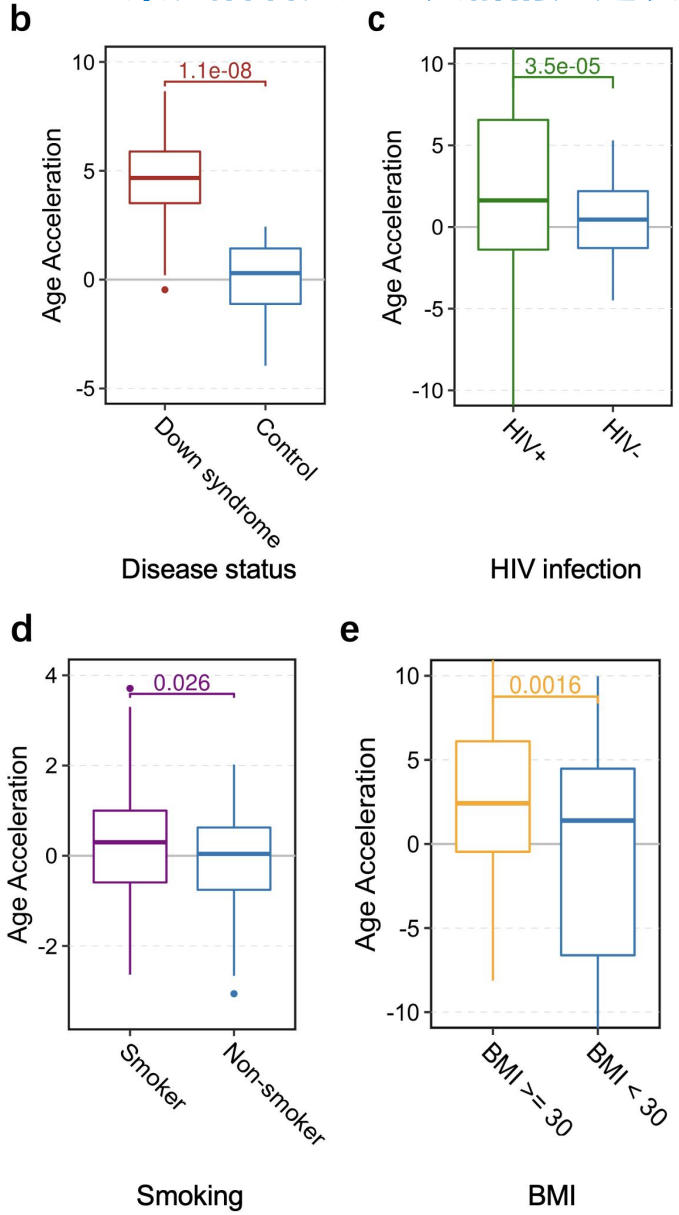


在衰老生物学框架下分析 MAPLE 的结果

MAPLE 能够捕捉到衰老相关的生物学功能



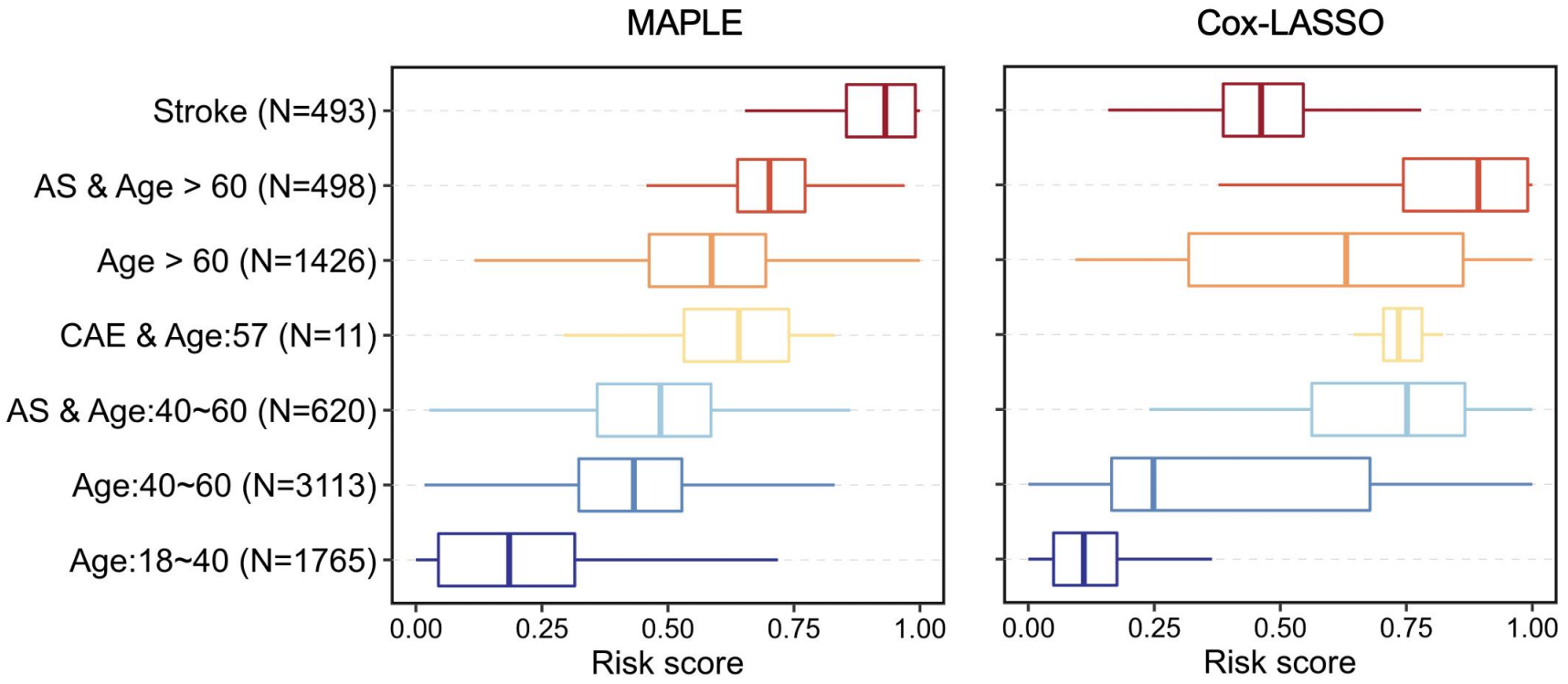
MAPLE 能够识别唐氏综合征、HIV 感染、抽烟、肥胖人群的加速衰老

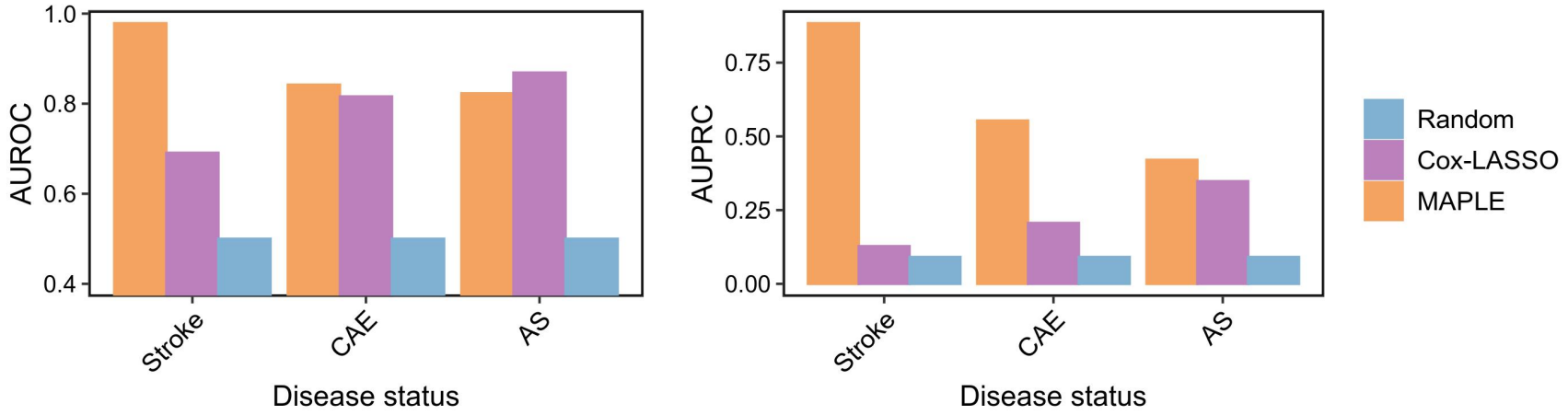


MAPLE 预测心脑血管衰老风险

MAPLE 的预测性能大幅优于传统模型：

- **各人群分组的预测风险符合临床经验**
- **预测中风 AUROC 0.98，AUPRC 0.88**
- **预测冠状动脉扩张 AUROC 0.82，AUPRC 0.55**

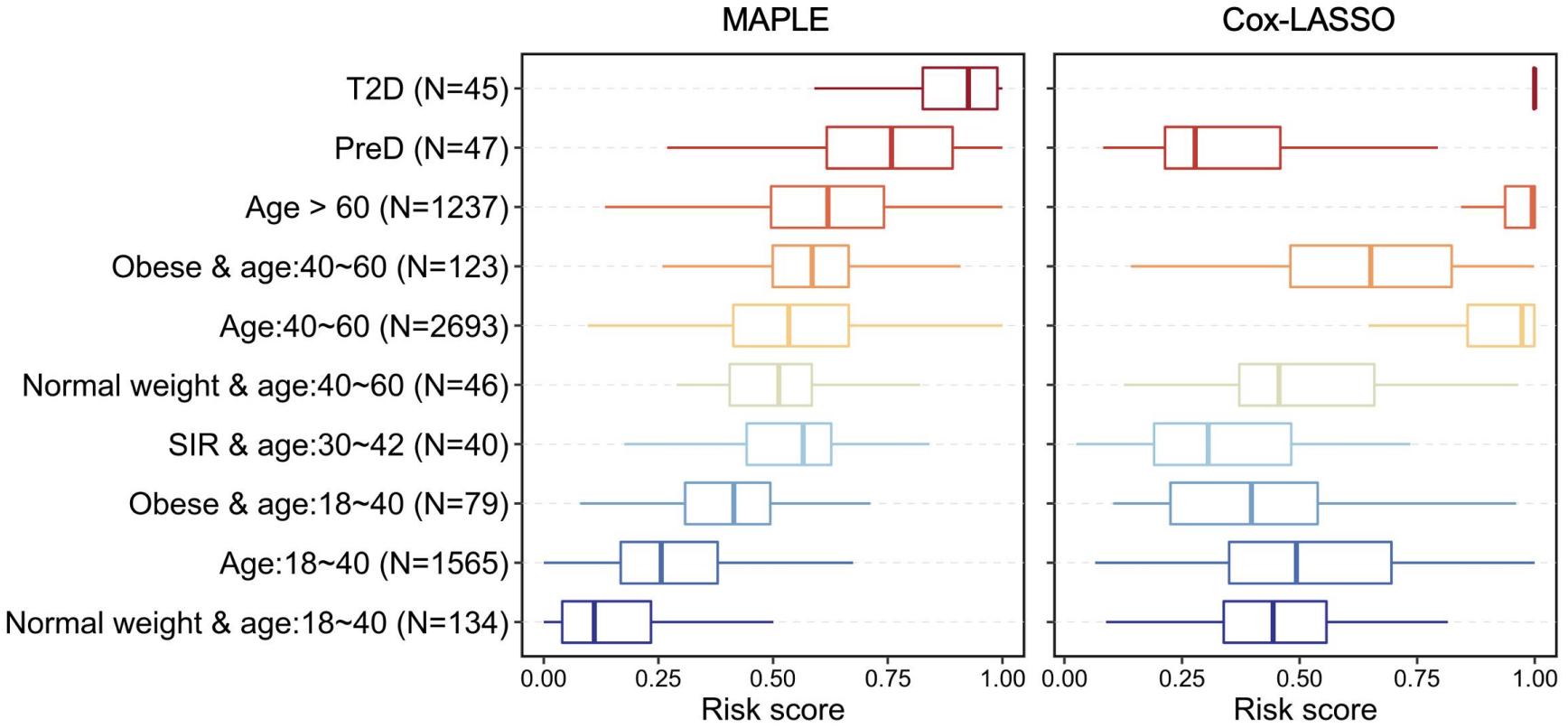


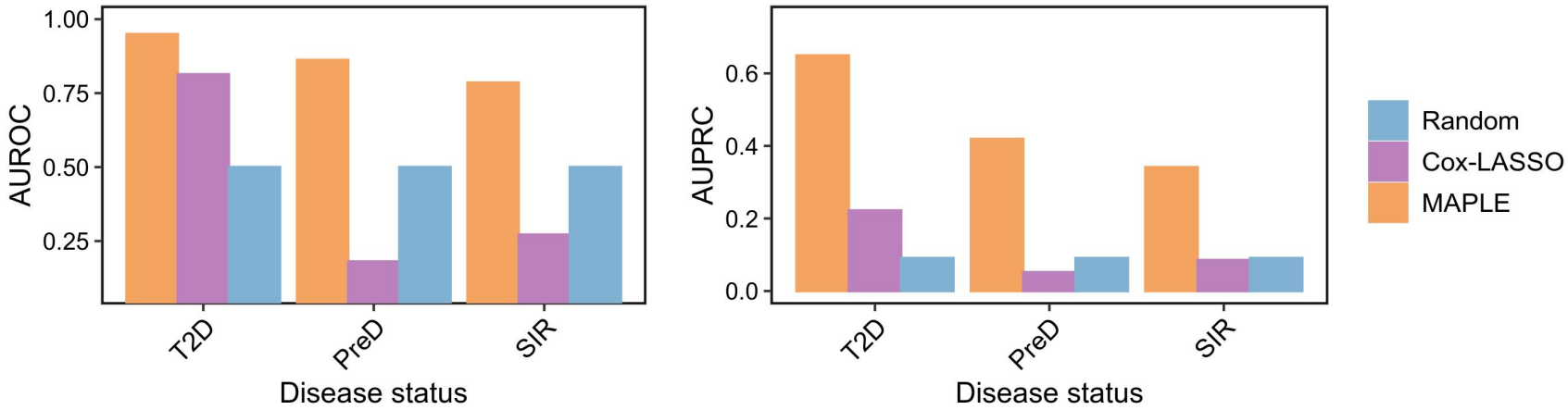


MAPLE 预测糖代谢衰老风险

MAPLE 的预测性能大幅优于传统模型：

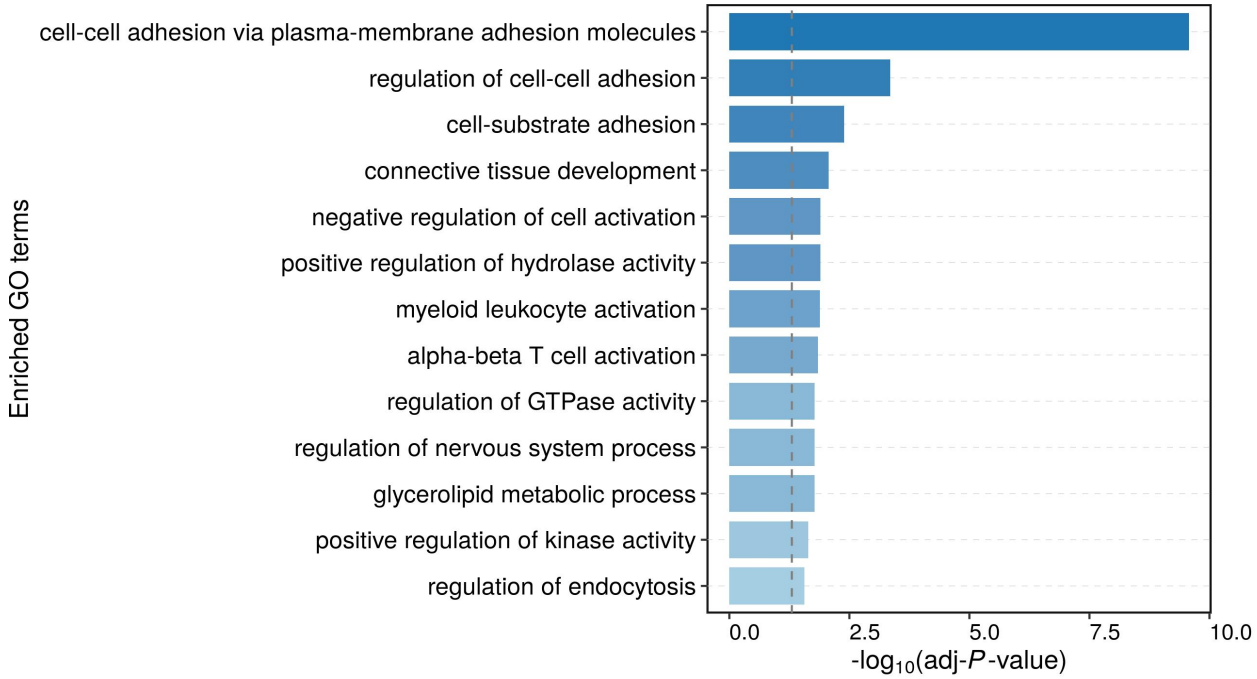
- **各人群分组的预测风险符合临床经验**
- **预测 2 型糖尿病 AUROC 0.95，AUPRC 0.65**
- **预测前驱糖尿病 AUROC 0.86，AUPRC 0.42**



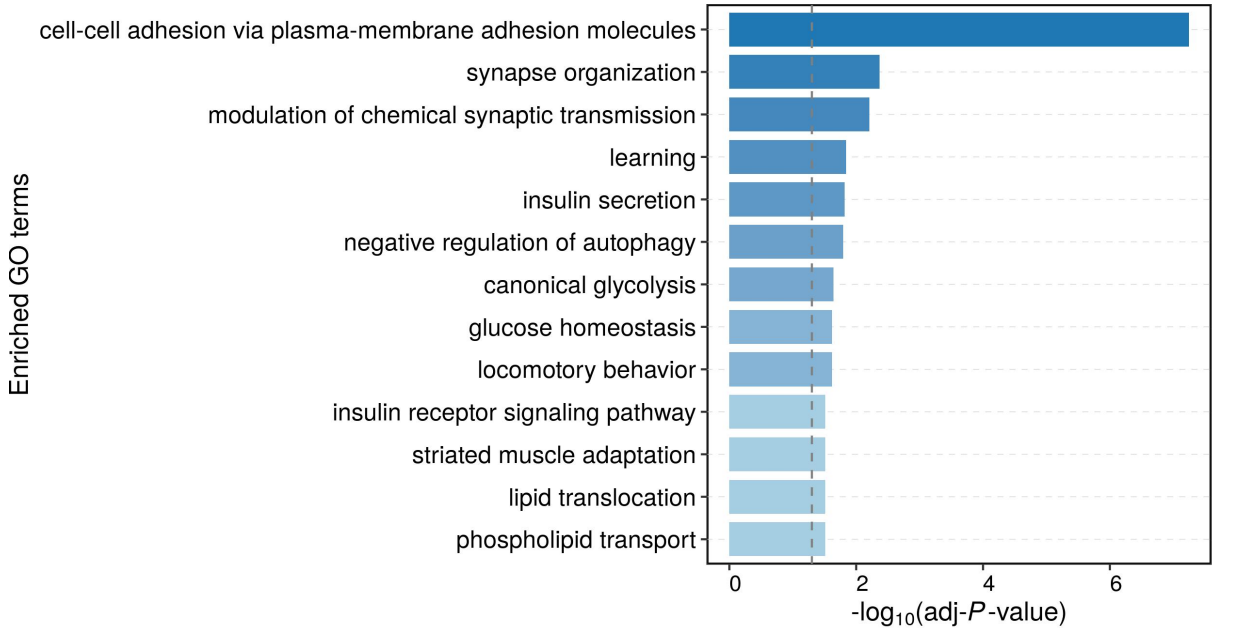


MAPLE 捕获到与疾病风险相关的生物学因素

MAPLE 能够捕捉到导致心脑血管衰老的生物学信号：CpGs 位点所在的基因被报道为 CVD 疾病机制的重要环节或治疗靶点



MAPLE 能够捕捉到导致糖代谢衰老的生物学信号： CpGs 位点所在的基因参与糖代谢相关过程，或与 T2D 并发症相关



介绍内容

总结与讨论

1. MAPLE 基于对比学习算法，能够在消除甲基化数据中批次效应的同时，保留研究者关注的生物学信号
   - 对比学习策略使得训练样本平方级增长，减少过拟合风险
   - 对比学习策略关注样本之间的年龄或风险差，忽略人群差异、平台差异等带来的系统误差
2. MAPLE 在多中心、多平台、多组织、多预处理方法的使用场景下达到 1.6 岁的预测误差
3. MAPLE 能够预测心脑血管疾病与 2 型糖尿病的疾病风险
   - 在疾病分类任务上达到 0.96 的 AUROC
   - 在疾病前驱状态识别任务上达到 0.83 的 AUROC

展望

- MAPLE 作为一个灵活的计算框架，能够扩展到免疫衰老、皮肤衰老、认知衰老、虚弱指数等更多维度的衰老预测任务上
- MAPLE 仅基于队列研究公开的线性模型参数即可对疾病风险预测的性能带来巨大提升，能够在保护受试者隐私的前提下利用队列研究的数据进行疾病风险评估

Q&A