

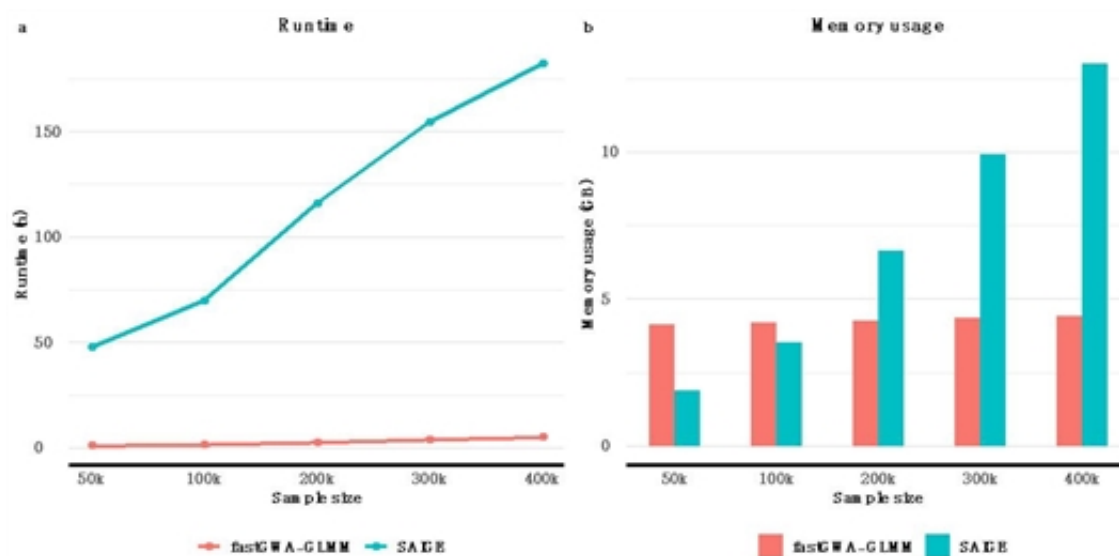
百万级生物样本库混合模型关联分析工具获开发

作者：writer 来源：爱科学

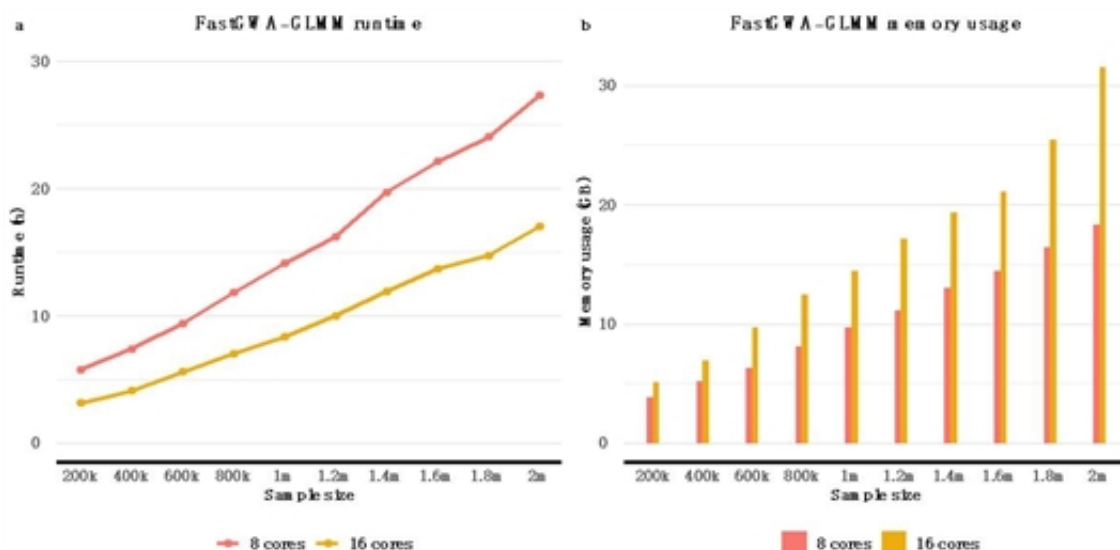
本文原地址：<https://www.iikx.com/news/progress/16450.html>

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

百万级生物样本库混合模型关联分析工具获开发。



fastGWA-GLMM和SAIGE的运算速度和内存使用量的比较。西湖大学供图



fastGWA-GLMM在2百万样本中的运算速度和内存使用量。 西湖大学供图

11月4日，西湖大学生命科学学院教授杨剑团队在《自然—遗传学》发表题为《一款适用于生物样本库规模数据的广义线性混合模型关联分析工具》的论文。该研究团队开发出了一款极为高效的广义线性混合模型（Generalized Linear Mixed Model, GLMM）关联分析工具，将其命名为fastGWA—GLMM，专门针对二元性状的全基因组关联研究（Genome—Wide Association Study，简称GWAS）。fastGWA—GLMM以及该团队之前开发的fastGWA（主要针对连续型性状）是目前唯一可应用于百万级生物样本库的混合模型GWAS分析工具。

GWAS是一种被广泛用于检测遗传变异与表型之间关联的实验设计。但它受制于数据分析能力。尤其是近年来，随着十万级、甚至百万级大型生物样本库的出现，例如英国生物样本库（UK Biobank，简称UKB）、美国的基因检测公司23andMe等，原有的GWAS分析工具捉襟见肘。

人体的性状可以大致分为两类，一种是连续型数量性状，指的是个体间的差异可以用数量区别，例如身高、体重等；另一种是二元性状，表现为对立面的两种状态，例如患病还是不患病。UKB中有3000—4000个性状，其中一半以上是非此即彼的二元性状。

过去常用的GWAS分析工具主要是基于线性回归模型（Linear Regression, LR），但LR的缺点是忽略了群体结构以及人与人之间的亲缘关系，从而会影响结果的准确性。随后出现的基于线性混合模型（Linear Mixed Model, LMM）的方法，可以有效控制样本中的群体结构和亲缘相关，无需移除样本中大量的亲缘相关个体，从而在避免虚假关联的条件下达到更高的统计功效。但线性混合模型原则上只适用于连续型性状，当它被应用于二元性状时，得到的只是近似解，结果并不精确。广义线性模型正好能弥补上述二者的缺陷，但广义线性模型有着比一般线性模型更高的计算复杂度。于是，科学家们开始思考如何开发一种可以高效地分析大型生物样本库数据的广义线性模型方法。

杨剑团队提出了一系列基于稀疏矩阵的算法，突破了传统广义线性模型和线性混合模型耗时、耗计算资源的瓶颈，开发出了一款极为高效的基于广义线性混合模型的二元性状全基因组关联分析工具：fastGWA—GLMM。

该团队用真实数据展示了fastGWA—GLMM极高的运算效率，远超同类二元性状关联分析方法，运算效率最高可达到传统方法的36倍。而在一个模拟的两百万人的群体中（每个人有约1200万个变异位点），fastGWA—GLMM在使用16个CPU核和32GB内存的情况下只需要17小时就可以完成一个二元性状的全基因关联分析，而这对于原有的工具是几乎不可能完成的任务。fastGWA—GLMM对大量数据的快速处理能力，对即将到来的百万级生物样本库具有重大意义。

作为一款稳健、强大且高效的关联分析工具，fastGWA—GLMM可以应用于几乎所有的大型生物样本库的二元性状关联分析。杨剑团队已经用fastGWA—GLMM分析了英国生物样本库中的2989个二元性状，并将所有的关联分析结果共享在他们的在线数据平台上（<http://fastgwa.info/ukbimph>in）。用户可以在这个平台上无限制地浏览、检索、查询、所有的结果数据。并且，该方法已被整合至该团队开发的开源软件包GCTA中（<https://yanglab.westlake.edu.cn/software/gcta>）。fastGWA和fastGWA—GLMM或将成为未来超大型生物样本库关联分析研究不可或缺的工具之一，其破解人类复杂疾病遗传奥秘的应用潜力不可估量。（来源：中国科学报温才妃）

相关论文信息：<https://doi.org/10.1038/s41588-021-00954-4>

版权声明：凡本网注明来源：中国科学报、科学网、科学新闻杂志的所有作品，网站转载，请在正文上方注明来源和作者，且不得对内容作实质性改动；微信公众号、头条号等新媒体平台，转载请联系授权。邮箱：shouquan@stimes.cn。

作者：杨剑等 来源：《自然—遗传学》

更多 科学进展 请访问 <https://www.iikx.com/news/progress/>

本文版权归原作者所有，请勿用于商业用途，[爱科学iikx.com](https://www.iikx.com)转发