

gwasfilter: 用于筛选全基因组关联研究的 R 脚本

杨淞淳¹ 李重阳² 胡一祯¹ 孙秋芬¹ 潘建桥² 孙点剑¹ 马宝山² 吕筠^{1,3}
李立明^{1,3}

¹北京大学公共卫生学院流行病与卫生统计学系 100191; ²大连海事大学信息科学技术学院 116026; ³北京大学公众健康与重大疫情防控战略研究中心 100191

杨淞淳和李重阳对本文有同等贡献

通信作者: 吕筠, Email: lvjun@bjmu.edu.cn; 马宝山, Email: mabaoshan@dmlu.edu.cn

【摘要】 目的 开发一套能高效准确地从 GWAS Catalog 公开数据库中筛选全基因组关联研究 (GWAS) 的 R 脚本。方法 参考既往研究制定 GWAS 的筛选原则。将人工在 GWAS Catalog 的筛选过程抽象为标准的算法, 2 名程序员共同撰写 R 脚本 (gwasfilter.R) 后, 由他人多次对脚本进行测试。结果 采用 gwasfilter.R 筛选 GWAS 包含 6 个步骤。该脚本内置 5 个主要函数, 可以同时根据“是否有验证人群”“样本量大小”和“研究人群种族”等条件对 GWAS 进行筛选。筛选单个性状时, 程序用时不超过 1 秒。结论 gwasfilter.R 操作简便, 筛选过程高效而标准化, 可灵活应用于 GWAS 筛选。脚本源代码网址: https://github.com/lab319/gwas_filter。

【关键词】 全基因组关联研究; GWAS Catalog; R 脚本

基金项目: 国家自然科学基金 (91846303, 61471078)

gwasfilter: an R script to filter genome-wide association study

Yang Songchun¹, Li Chongyang², Hu Yizhen¹, Sun Qiufen¹, Pan Jianqiao², Sun Dianjianyi¹, Ma Baoshan², Lyu Jun^{1,3}, Li Liming^{1,3}

¹Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China; ²College of Information Science and Technology, Dalian Maritime University, Dalian 116026, China; ³Peking University Center for Public Health and Epidemic Preparedness and Response, Beijing 100191, China

Yang Songchun and Li Chongyang contributed equally to the article

Corresponding authors: Lyu Jun, Email: lvjun@bjmu.edu.cn; Ma Baoshan, Email: mabaoshan@dmlu.edu.cn

【Abstract】 Objective To develop an R script that can efficiently and accurately filter genome-wide association studies (GWASs) from the GWAS Catalog Website. **Methods** The selection principles of GWASs were established based on previous studies. The process of manual filtering in the GWAS Catalog was abstracted as standard algorithms. The R script (gwasfilter.R) was written by two programmers and tested many times. **Results** It takes six steps for gwasfilter.R to filter GWASs. There are five main self-defined functions among this R script. GWASs can be filtered based on "whether the GWAS has been replicated" "sample size" "ethnicity of the study population" and other conditions. It takes no more than 1 second for this script to filter GWASs of a single trait. **Conclusions** This R script (gwasfilter.R) is user-friendly and provides an efficient and standard process to filter GWASs flexibly. The source code is available at github (https://github.com/lab319/gwas_filter).

【Key words】 Genome-wide association study; GWAS Catalog; R script

Fund programs: National Natural Science Foundation of China (91846303, 61471078)

DOI: 10.3760/cma.j.cn112338-20200731-01003

收稿日期 2020-07-31 本文编辑 李银鸽

引用本文: 杨淞淳, 李重阳, 胡一祯, 等. gwasfilter: 用于筛选全基因组关联研究的 R 脚本[J]. 中华流行病学杂志, 2021, 42(10): 1876-1881. DOI: 10.3760/cma.j.cn112338-20200731-01003.



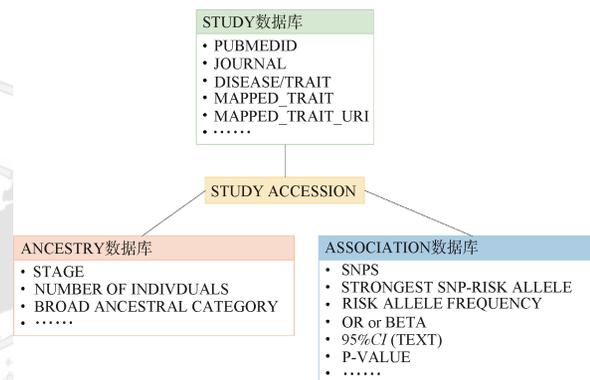
全基因组关联研究 (genome-wide association study, GWAS) 是在整个人类基因组范围内检验遗传变异 (genetic variant) 与疾病或性状之间统计学关联的重要研究方法^[1]。深入分析 GWAS 发现的显著遗传位点对于确定疾病易感位点 (casual variant)、阐明疾病遗传机制、实现精准医疗和精准预防都具有重要意义^[2-5]。NHGRI-EBI GWAS Catalog (<https://www.ebi.ac.uk/gwas/>, GWAS Catalog) 自 2008 年起持续收集所有已发表的符合该网站纳入标准的 GWAS, 是目前信息最为全面、更新相对及时、数据质量较高且能够公开获取的 GWAS 数据库^[6]。截至 2021 年 8 月, GWAS Catalog 已收录超过 5 000 篇 GWAS 文献的研究信息。

在报告新的 GWAS 位点 (new loci)、构建遗传风险评分 (genetic risk score)、开展基因多效性 (pleiotropy) 分析、遗传相关 (genetic correlation) 分析、孟德尔随机化 (mendelian randomization) 或基因-环境交互作用 (gene-environment interaction) 分析以及通路分析 (pathway analysis) 等之前^[7-12], 都需要检索、利用既往 GWAS 的研究结果。对已发表的 GWAS 研究通常要参考以下一些原则进行适当筛选, 以获得可信度更高、与目标人群更相关的研究。例如, 在 GWAS 中, 发现人群中观察到的结果若能在另一个独立人群中得到验证 (replication), 则具有更高的可信度^[13]。由于常见的遗传变异与疾病或性状之间的关联强度往往较弱, 大样本量将更有助于发现潜在的关联^[13]。同时, 由于不同种族人群遗传背景的天然差异, 在某个种族中发现的遗传位点可能无法直接应用于其他种族^[13]。在 GWAS Catalog 中筛选目标 GWAS 研究时, 应用上述这些条件, 需要人工逐一判断和筛选, 特别是同时需要筛选较多性状时, 费时费力, 且容易出错。有研究者最近开发了针对 GWAS Catalog 数据库的 R 程序包 gwasrappid, 提供的函数 get_studies() 只能基于特定的性状检索词查询 GWAS Catalog 中匹配的研究, 但没有应用特定条件筛选研究的功能^[14]。因此, 我们开发了一套基于 R 语言的程序脚本, 用于高效、准确地从 GWAS Catalog 中筛选满足特定条件的 GWAS。

研究方法

1. 筛选原理: GWAS Catalog 网站允许直接下载完整的结构化数据库, 包括研究设计、研究人群、性状和遗传位点等信息 (图 1)。在 GWAS Catalog 中,

性状主要采用 Experimental Factor Ontology (EFO) 分类系统, 标准性状与 EFO 编码一一对应^[6]。由于该分类系统具有跨域 (cross-domain) 分类的特点^[15], 通过某一 EFO 编码检索到的研究并非都能满足目标性状的要求。例如, 若我们只关注在自然人群中开展的某性状的 GWAS 结果, 则需剔除在患者中开展的 GWAS、该性状对药物治疗反应的 GWAS、该性状与其他性状交互作用的 GWAS。因此, 研究者首先需要根据 GWAS Catalog 提供的原始文献中“报告性状” (reported traits) 的名称人工判断哪些是自己期望的性状, 据此获取所有相关的 GWAS 后, 则可以根据“是否有验证人群”“样本量大小”和“研究人群种族”等条件进一步筛选。



注: STUDY 数据库中包含 GWAS 的发表、研究设计、性状等基本信息; ANCESTRY 数据库中包含研究人群 (含发现人群与验证人群) 的种族及样本量信息; ASSOCIATION 数据库中包含遗传位点与表型的关联信息及位点注释信息等。3 个子数据库之间通过变量“STUDY ACCESSION”互相链接。数据库下载地址: <https://www.ebi.ac.uk/gwas/docs/file-downloads>

图 1 GWAS Catalog 数据库基本结构

2. 脚本完成过程: 基于以上原理, 由 2 名程序员采用 R 语言共同撰写了脚本 gwasfilter.R^[16]。撰写思路为: ①将人工筛选研究的过程抽象为标准化的算法; ②明确各环节的输入和输出; ③采用 R 语言中的函数 (function) 对每个环节进行编译。该脚本基于 R 3.5.1 x64 环境开发, 开发完成后, 由 3 名测试员经过多次测试, 在 R 3.5 软件及以上平台、32 位操作系统及 Linux 操作系统均可正常运行。脚本源代码可在 github 平台 (https://github.com/lab319/gwas_filter) 下载获得。

结 果

1. 脚本操作具体步骤: gwasfilter.R 包含 6 个标

准步骤(图 2), 内置 5 个预设函数(表 1), 当前支持 2 种使用模式。模式 1 为从头筛选模式, 包含完整的 6 个步骤, 在未明确性状 EFO 编码和性状筛选词时使用。模式 2 为直接筛选模式, 在已备好性状 EFO 编码文件和性状筛选词文件的前提下, 可跳过步骤 3~5 直接进行步骤 6。以下为各步骤简介:

步骤 1: 准备 R 工作平台。在使用前需运用 source 命令加载 gwasfilter.R 中的预设函数。加载过程中程序会自动检查依赖包 (readr、dplyr、stringr、data.table) 的安装情况。

步骤 2: 下载/更新 GWAS Catalog 数据库。函数 get_gwasdata() 将 GWAS Catalog 网站的数据库下载至本地并读入内存。该函数会自动检查本地已有数据库是否需更新。

步骤 3: 获取目标性状 EFO 编码。函数 get_efo() 可返回查询给定关键词时所有相关的 EFO 编码。

步骤 4: 获取“报告性状”。函数 obtain_trait() 可返回在 GWAS Catalog 中检索某个 EFO 编码(一次只能选一个 EFO 编码)时所有相关研究的“报告性状”。

步骤 5: 明确目标性状筛选词。函数 store_trait() 通过直接记录正确的“报告性状”的序号向量 traitindex, 能方便地完成性状筛选词的存储过程。在执行步骤 6 前, 步骤 3~5 可以循环进行(图 2), 从而实现程序的批量筛选功能。

步骤 6: 执行筛选策略。函数 gwasfilter() 包含多个功能参数(表 1)。用户可以通过设置这些参数, 灵活组合多个筛选条件: ①是否剔除未观察到满足特定统计学显著性水平 ($P < 1.0 \times 10^{-5}$) 的遗传位点的研究 (exclude0); ②是否有验证人群 (replication); ③发现阶段样本量 (sample); ④发现人群的种族 (ancestry.eas 和 eas.only); ⑤期刊特征, 如以 2018 年期刊引用报告 (JCR) 中的杂志分区 (journal) 为例。通过设置参数 association=TRUE 还可以额外导出入选研究对应的遗传位点关联数据库。在不设置任何筛选参数时, gwasfilter() 可返回筛选前的所有记录。

2. 测试情况: 以筛选总胆固醇水平 (total cholesterol) 的 GWAS 为例进行测试。测试所用平台为 R 4.0.1 软件, 所用 GWAS Catalog 数据库的更

表 1 gwasfilter.R 内置函数参数说明

序号	步骤内容	函数名(参数名)	参数简介
1	下载/更新 GWAS Catalog 数据库	get_gwasdata(path, checknew)	path: GWAS Catalog 数据库的下载路径; 默认为当前路径 getwd() checknew: logical 值, 在本地已有 GWAS Catalog 数据库时是否检查更新; 默认为 TRUE
2	获取性状 EFO 编码	get_efo(trait)	trait: 目标性状的关键词, 不区分大小写
3	获取目标性状筛选词	obtain_trait(efolist, efoindex, append)	efolist: 待选择的 EFO 数据框; 默认为最近一次使用 get_efo() 函数时输出的信息 efoindex: 数值, EFO 数据框中目标性状的序号 append: 是否保留之前已查询过的性状, 默认为 TRUE (保留)
4	明确目标性状筛选词	store_trait(aim.efo, traitlist, traitindex)	aim.efo: 目标性状的 EFO 编码; 默认为最近一次使用 obtain_trait() 函数时 efoindex 指定的 EFO 编码 traitlist: obtain_trait() 函数得到的性状筛选词数据框; 默认为最近一次使用 obtain_trait() 函数时输出的信息 traitindex: 人工筛选后保留的性状筛选词的序号向量
5	执行筛选策略	gwasfilter(efofile, traitfile, exclude0, replication, sample, ancestry.eas, eas.only, journal, JCRfile, association)	efofile: 需要筛选研究的目标性状清单文件; 默认为 "TMP_EFO_LIST.csv" traitfile: 目标性状清单对应的性状筛选词的文件; 默认为 "TMP_EFO_TRAITS.csv" exclude0: logical 值, 是否剔除研究库中 ASSOCIATION COUNT 为 0 的研究; 默认为 FALSE (不剔除) replication: logical 值, 是否只保留验证阶段的研究; 默认为 FALSE (均保留) sample: 数值, 只保留发现阶段样本量大于等于该值的研究; 默认为 0 ancestry.eas: logical 值, 是否只保留发现人群包含东亚人的研究; 默认为 FALSE (均保留) eas.only: logical 值, 在 ancestry.eas 设置为 TRUE 时, 是否只保留发现人群仅为东亚人的研究; 默认为 FALSE (均保留) journal: 数值向量, c(1, 2, 3, 4, 9) 的子集, 只保留期刊数据选定 JCR 分区的研究, 9 代表 JCR 分区不确定; 默认为 c(1, 2, 3, 4, 9) JCRfile: JCR 的杂志分区文件; 默认为 "Catalog_JCR_2018.csv" association: logical 值, 是否需要导出对应的遗传位点关联数据库; 默认为 FALSE (不导出)

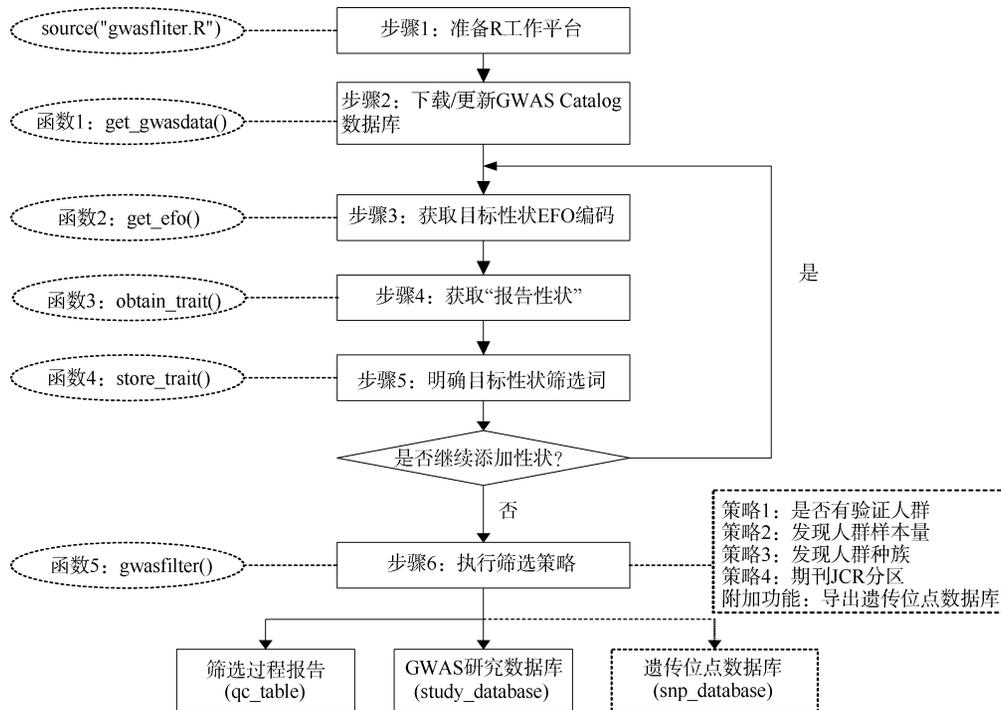


图2 gwasfilter.R脚本原理图

新日期为 2021-08-16, 测试语句如下:

```
source("gwasfilter.R")#步骤 1: 加载脚本函数;
get_gwasdata()#步骤 2: 获取 GWAS Catalog 数据库;
get_efo(trait="Cholesterol")#步骤 3: 查询胆固醇相关的 EFO 编码;
obtain_trait(efoindex=1)#步骤 4: 查询目标性状对应的报告性状;
store_trait[traitindex=c(1, 2, 11, 12)]#步骤 5: 明确目标性状筛选词;
gwasfilter(replication = T, ancestry.eas = T, eas.only = F, journal=c(1), association=T)#步骤 6: 执行筛选策略(有验证阶段, 发现人群包含东亚人, 期刊为 JCR Q1 区, 同时输出对应的遗传位点关联数据库)。
```

上述步骤中, 除步骤 2 下载与预处理数据耗时约 4 min 以外(与具体网速有关), 其余步骤程序用时均 < 1 秒。步骤 6 完成后输出 4 个 csv 文件(表 2), 分别为 qc_table1、qc_table2、study_database 和 snp_database。在上述筛选条件下, 共有 6 项研究符合筛选要求。

讨 论

当前尚无其他的专注于筛选 GWAS 的工具, 仅

有的一个与 GWAS Catalog 相关的 R 程序包 *gwasrapid* 仅能简单查询 GWAS Catalog 中的研究, 没有进一步的筛选功能^[14]。与人工筛选相比, *gwasfilter.R* 具有 4 大优势。首先, 使用过程简便。用户只需要记录 2 个参数(*efoindex* 和 *traitindex*), 并根据需要确定筛选策略, 便可完成筛选, 省去了在 GWAS Catalog 网站检索性状、下载数据然后逐一判断的繁琐过程; 在筛选条件较多、筛选量较大时, 也能避免人工操作导致的失误。第二, 筛选过程高效。除下载数据外, 程序执行过程所需时间几乎可以忽略。在需要筛选的性状较多时, 脚本的批量筛选功能可以大幅度地提升筛选效率。第三, 筛选过程标准化。*gwasfilter.R* 导出的文件 *qc_table1.csv* 记录了筛选过程的关键数据, 每个性状所用的性状筛选词也能即时存储, 便于对筛选过程进行核验; 每个性状存储的性状筛选词也可在之后重复使用。第四, 筛选方式灵活。用户可以根据研究需要, 灵活地组合不同的筛选条件。

当然, *gwasfilter.R* 也存在一定的局限性。第一, 该脚本的运行依赖于 GWAS Catalog 定期发布的数据库, 因此, 筛选的准确度以及导出的遗传位点数据库的数据质量取决于 GWAS Catalog 数据库本身的数据质量。第二, 在明确目标性状筛选词的环节(步骤 5), 不同研究者的意见可能存在分歧。通过多人平行地甄别性状筛选词将有助于提高筛

表 2 gwasfilter.R 输出文件变量注释

变量名	变量标签	变量值标签
文件 1:qc_table1		
MAPPED_TRAIT	性状标准名称	
EFO	性状 EFO 编码	
N_reported.traits	性状筛选词的数量	
status	筛选状态	1=完成,0=未完成
EFO_record	通过 EFO 编码检索到的研究数(N1)	
merged_record	在 N1 的基础上,符合性状筛选词的研究数(N2)	
nonzero_merged_record	在 N2 的基础上,符合性状筛选词且关联数不为 0 的研究数(N3)	
is_replicated	在 N2 或 N3 的基础上(取决于 exclude0 参数的设定),有验证阶段的研究数	
Sample.size.median	在 N2 或 N3 的基础上,发现人群样本量的中位数	
is_EAS	在 N2 或 N3 的基础上,发现人群中包含东亚地区人群的研究数	
JCR.Q1	在 N2 或 N3 的基础上,期刊水平属于 JCR Q1 区的研究数	
total_kept	经过筛选后保留的研究数	
文件 2:qc_table2		
EFO	性状 EFO 编码	
reported.traits	性状筛选词	
related.EFO	性状筛选词中所包含的相关性状的 EFO 编码	
N_merged_record	根据性状筛选词可以查询到的研究	
文件 3:study_database ^a		
EFO	性状 EFO 编码	
STUDY ACCESSION	GWAS Catalog 中的研究编号	
replicated	是否有验证阶段	1=是,0=否
Sample.size	发现人群样本量	
is_EAS	发现人群是否包含东亚人	1=是,0=否
is_TRA	发现人群是否包含多个种族	1=是,0=否
JCRrank	期刊 JCR 分区	1=Q1,2=Q2,3=Q3,4=Q4,9=未知
Chr.missing	遗传位点数据库中染色体编号缺失的位点数	
Pos.missing	遗传位点数据库中位置缺失的位点数	
snpname.invalid	遗传位点数据库中位点名称非标准 rsid 的位点数	
EA.missing	遗传位点数据库中 effect allele 缺失的位点数	
EA.invalid	遗传位点数据库中 effect allele 碱基不明确的位点数	
betaOR.missing	遗传位点数据库中 beta 值和 OR 值缺失的位点数	
P.missing	遗传位点数据库中 P 值缺失的位点数	
文件 4:snp_database		
EFO	性状 EFO 编码	
STUDY ACCESSION	GWAS Catalog 中的研究编号	
Region	所属染色体区域	
Chr	染色体编号	
Pos(hg38)	位置(build 为 hg38)	
Locus	临近基因	
rsid	位点名称	
Effect_allele	效应碱基	
Effect_allele_freq	效应碱基频率	
Function	位点功能注释	
beta	beta 值	
OR	OR 值	
LCI	beta 值或 OR 值的下 95%CI	
UCI	beta 值或 OR 值的上 95%CI	
P-VALUE	P 值	
P-VALUE(TEXT)	对 P 值的注释	
Chr.missing	染色体编号是否缺失	1=是,0=否
Pos.missing	位置是否缺失	1=是,0=否
snpname.invalid	位点名称是否非标准 rsid	1=是,0=否
EA.missing	效应碱基是否缺失	1=是,0=否
EA.invalid	效应碱基是否碱基不明确	1=是,0=否
EAF.missing	效应碱基频率是否缺失	1=是,0=否
betaOR.missing	beta 值和 OR 值是否缺失	1=是,0=否
P.missing	P 值是否缺失	1=是,0=否

注:^a study_database 中部分变量直接来源于 GWAS Catalog study 数据库,其变量标签可在 <https://www.ebi.ac.uk/gwas/docs/fileheaders> 查询

选的准确度。第三,由于一项GWAS从发表到收录至GWAS Catalog有一定的时间间隔,对于3个月内新发表的研究,则需要在PubMed上检索。

综上所述,我们开发了一套基于GWAS Catalog数据库筛选GWAS的R脚本。该脚本操作简便,筛选研究的过程高效而标准化,可灵活应用于实际研究。后续研究可以在当前脚本的基础上进一步与其他公共数据库整合。

利益冲突 所有作者均声明不存在利益冲突

参 考 文 献

[1] 陈峰,柏建岭,赵杨,等.全基因组关联研究中的统计分析方法[J].中华流行病学杂志,2011,32(4):400-404. DOI:10.3760/cma.j.issn.0254-6450.2011.04.018.
 Chen F, Bai JL, Zhao Y, et al. Statistical methodologies used in genome-wide association studies[J]. Chin J Epidemiol, 2011, 32(4):400-404. DOI:10.3760/cma.j.issn.0254-6450.2011.04.018.
 [2] Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping[J]. Nat Rev Genet, 2018, 19(8):491-504. DOI:10.1038/s41576-018-0016-z.
 [3] Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function[J]. Am J Human Genet, 2018, 102(5):717-730. DOI:10.1016/j.ajhg.2018.04.002.
 [4] Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores[J]. Nat Rev Genet, 2018, 19(9):581-590. DOI:10.1038/s41576-018-0018-x.
 [5] Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications[J]. Nat Genet, 2015, 47(8):856-860. DOI:10.1038/ng.3314.
 [6] Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary

statistics 2019[J]. Nucleic Acids Res, 2019, 47(D1):D1005-1012. DOI:10.1093/nar/gky1120.
 [7] 王斌,戴俊程,孙义民,等.遗传风险评估的原理与方法[J].中华流行病学杂志,2015,36(10):1062-1064. DOI:10.3760/cma.j.issn.0254-6450.2015.10.005.
 Wang C, Dai JC, Sun YM, et al. Genetic risk score: principle, methods and application[J]. Chin J Epidemiol, 2015, 36(10):1062-1064. DOI:10.3760/cma.j.issn.0254-6450.2015.10.005.
 [8] Solovieff N, Cotsapas C, Lee PH, et al. Pleiotropy in complex traits: challenges and strategies[J]. Nat Rev Genet, 2013, 14(7):483-495. DOI:10.1038/nrg3461.
 [9] Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits[J]. Nat Genet, 2015, 47(11):1236-1241. DOI:10.1038/ng.3406.
 [10] 林丽娟,魏永越,张汝阳,等.孟德尔随机化方法在观察性研究因果推断中的应用[J].中华预防医学杂志,2019,53(6):619-624. DOI:10.3760/cma.j.issn.0253-9624.2019.06.015.
 Lin LJ, Wei YY, Zhang RY, et al. Application of mendelian randomization methods in causal inference of observational study[J]. Chin J Prev Med, 2019, 53(6):619-624. DOI:10.3760/cma.j.issn.0253-9624.2019.06.015.
 [11] McAllister K, Mechanic LE, Amos C, et al. Current challenges and new opportunities for gene-environment interaction studies of complex diseases[J]. Am J Epidemiol, 2017, 186(7):753-761. DOI:10.1093/aje/kwx227.
 [12] Wang K, Li MY, Hakonarson H. Analysing biological pathways in genome-wide association studies[J]. Nat Rev Genet, 2010, 11(12):843-854. DOI:10.1038/nrg2884.
 [13] McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges[J]. Nat Rev Genet, 2008, 9(5):356-369. DOI:10.1038/nrg2344.
 [14] Magno R, Maia AT. gwasrapidd: an R package to query, download and wrangle GWAS catalog data[J]. Bioinformatics, 2019, 36(2):649-650. DOI:10.1093/bioinformatics/btz605.
 [15] Malone J, Holloway E, Adamusiak T, et al. Modeling sample variables with an experimental factor ontology[J]. Bioinformatics, 2010, 26(8):1112-1118. DOI:10.1093/bioinformatics/btq099.
 [16] R Core Team. R: A language and environment for statistical computing[EB/OL]. R Foundation for Statistical Computing, Vienna, Austria, 2018. https://www.R-project.org/.

读者·作者·编者

本刊常用缩略语

本刊对以下较为熟悉的一些常用医学词汇将允许直接用缩写,即在文章中第一次出现时,可以不标注中文和英文全称。

OR	比值比	HBcAg	乙型肝炎核心抗原
RR	相对危险度	HBeAg	乙型肝炎e抗原
CI	可信区间	HBsAg	乙型肝炎表面抗原
P _n	第n百分位数	抗-HBs	乙型肝炎表面抗体
AIDS	艾滋病	抗-HBc	乙型肝炎核心抗体
HIV	艾滋病病毒	抗-HBe	乙型肝炎e抗体
MSM	男男性行为者	ALT	丙氨酸氨基转移酶
STD	性传播疾病	AST	天冬氨酸氨基转移酶
DNA	脱氧核糖核酸	HPV	人乳头瘤病毒
RNA	核糖核酸	DBP	舒张压
PCR	聚合酶链式反应	SBP	收缩压
RT-PCR	反转录聚合酶链式反应	BMI	体质指数
Ct值	每个反应管内荧光信号达到设定的 阈值时所经历的循环数	MS	代谢综合征
PAGE	聚丙烯酰胺凝胶电泳	FPG	空腹血糖
PFGE	脉冲场凝胶电泳	HDL-C	高密度脂蛋白胆固醇
ELISA	酶联免疫吸附试验	LDL-C	低密度脂蛋白胆固醇
A值	吸光度值	TC	总胆固醇
GMT	几何平均滴度	TG	甘油三酯
HBV	乙型肝炎病毒	CDC	疾病预防控制中心
HCV	丙型肝炎病毒	WHO	世界卫生组织
HEV	戊型肝炎病毒		