

《中国细胞生物学学报》数据论文写作模板 (ScienceDB)

题目

要求：客观、准确、只体现数据集，不体现科研结论。格式统一为：

【研究对象+实验条件+数据类型+数据集】。

作者及单位（中英文）

列出所有数据采集、整理、分析人员，标注单位、邮编。

作者贡献

需要说明每位作者在数据采集、整理、分析、论文撰写等环节的具体分工，推荐使用 CRediT (Contributor Roles Taxonomy) 分类法，如“数据汇总、论文撰写、数据分析、样品采集、数据审核”等。

摘要（中英文）

数据论文摘要四段式写作：研究背景与数据构建目的、数据采集与处理方法、数据集构成与质控结果、数据共享价值与复用前景。禁止出现机制结论、差异结果、创新性推论。末尾明确标注数据存储平台、DOI 与共享协议。（300–500 字）。

基础参数	填写内容说明
数据集名称	与论文题目对应，精准匹配研究对象与数据类型
数据集版本	默认 V1.0，迭代更新可标注后续版本
数据产出时间	实验数据采集、整理完成的具体时间段
数据学科领域	细胞生物学、分子生物学、肿瘤生物学等精准细分领域
数据类型	测序数据、细胞成像数据、蛋白检测数据、细胞功能实验数据等
数据总容量	完整数据集文件总大小（精确到 MB/GB）
数据文件总数	包含原始数据、处理数据、质控文件、说明文档的全部文件数量
共享协议	本数据论文的文章版权归属《中国细胞生物学学报》；存储于 ScienceDB 的数据集采用 CC0 1.0 协议，二者相互独立
存储平台	科学数据银行（ScienceDB）
数据集 DOI	ScienceDB 审核通过后分配的唯一永久 DOI
访问链接	数据集官方永久公开访问链接

关键词（中英文） 3~8 个

数据概括

这是数据论文特有的前置要素，用于简要描述数据集的核心特征（数据集名称、所属学科、研究主题、时间范围、数据格式、数据量等）。

脚注（中英文） 基金及课题支持、通信作者及联系方式。

引言

简要阐述该细胞生物学研究方向的研究现状、现有公开数据集的不足，说明本数据集的构建必要性。介绍本数据的实验体系、数据覆盖范围，说明数据公开共享对领域可重复性研究、二次挖掘、方法学优化的意

义。无需大量文献综述，不展开科学假说与机制探讨。

1 材料与方法

本章需写至可完全重复实验程度，是数据论文核心章节。

1.1 实验材料与细胞模型

详细说明细胞系来源、培养体系、处理条件、分组方式、生物学重复数量、传代与培养参数；所用试剂、仪器、耗材型号与来源；涉及动物/人体样本需说明伦理审批情况。

1.2 数据采集方法

根据实验类型完整描述：显微成像、细胞功能检测、蛋白核酸实验、高通量测序等全部操作流程，写明仪器参数、采集模式、实验时间、操作标准。

1.3 数据预处理方法

说明原始数据清洗、格式统一、降噪、校正、归一化、序列比对、图像分割等全部处理步骤；写明所用软件、版本、关键参数与筛选阈值。

1.4 数据质量控制标准

明确本数据集质控指标，如测序 Q30、样本相关性、重复一致性、图像清晰度、数据缺失率、异常值剔除标准等。

1.5 数据显著性分析

说明使用的统计软件、版本、检验方法、显著性水平。

2 数据样本描述

2.1 数据信息及结构

详细阐述数据集整体架构、文件分类、文件夹层级结构、统一命名规则，完整说明原始数据、预处理数据、质控文件、元数据文件、分析脚本、README 说明文档等配套资料的格式特征、存储形式与组织逻辑。清晰区分各类数据文件的生成阶段、数据格式、适配软件，保障数据集结构规范、条理清晰、可直接复用。

2.2 数据样本

系统、具体描述本数据集对应的全部实验样本信息，涵盖细胞样本来源、品系、培养条件、分组设置、生物学重复与技术重复数量、样本处理梯度、实验批次等核心样本基础信息。明确样本筛选标准、剔除规则，完整呈现样本整体特征与实验排布体系，保证样本信息可溯源、实验可重复。

其中展示的图表及数据的要求如下

图表及图注/表注

数据集整体架构图：展示文件夹层级、文件类型、数据流向；

样本信息汇总表：列出全部实验样本的分组、重复数、处理条件等；

质控指标汇总表：Q30、样本相关性、缺失率等关键质控数据；

图表需具备自明性，图题在图下方、表题在表上方。

数据处理工具或代码可用性（可选但有则必备）

如果数据预处理使用了自定义脚本、算法或分析流程，需要提供代码存储位置（如 GitHub、Zenodo）及获取方式。

3 数据质量评价

3.1 数据完整性

说明数据集是否完整覆盖实验全部样本、全部实验组别、全部检测指标，无关键数据缺失，配套元数据（元数据文件必填字段清单样本 ID、实验批次、仪器编号、试剂批号、实验日期等）、实验参数、样本信息齐全，可完整还原全部实验流程。

3.2 数据准确性与可靠性

结合具体质控数据、实验重复结果、仪器检测标准，佐证数据真实准确，无人工篡改、系统误差、异常干扰，生物学重复、技术重复一致性良好，符合细胞生物学实验数据通用质控标准。

3.3 数据集局限性与适用边界

客观、严谨说明数据集的适用范围、存在的客观局限，明确数据复用边界，避免不当误用，保证数据二次挖掘的规范性与科学性。

4 数据应用价值

结合细胞生物学领域特点，说明本数据集可支撑的后续研究方向，包括机制研究、方法验证、数据建模、算法训练、标志物筛选、同类实验对照等。说明数据开放对提升领域研究可重复性、推动学科二次挖掘、优化实验方法体系的积极作用。

数据论文以客观展示数据集为核心，无需撰写研究结论、机制总结及创新性推论，本章节默认省略。

后置部分	位置	说明
数据可用性声明	正文后、参考文献前	本研究构建的全部原始数据、处理数据及配套说明文件已公开存储于科学数据银行（ScienceDB）。数据集 DOI：_____，永久访问链接：_____。数据集采用 CC0 公共领域协议开放共享，全球用户可免费获取、复用与二次分析。
利益冲突声明	数据可用性声明后	声明经济/个人/职业关系
共享约定	利益冲突声明后	本数据论文的文章版权归属《中国细胞生物学学报》；存储于 ScienceDB 的数据集采用 CC0 1.0 协议，二者相互独立
伦理声明	共享约定后	涉及人体/动物实验时标注审批编号， 动物实验：缺少实验动物许可证号、3R 原则说明要求； 人体样本：缺少知情同意 / 豁免知情同意说明、样本去标识化 / 数据脱敏规则； 细胞样本：如果使用商品化细胞系，1.1 仅提细胞来源，投稿须知未强制要求配套细胞 STR 鉴定报告作为数据集附件（细胞生物学稿件硬性要求）；人源细胞系还需说明细胞来源伦理
致谢	伦理声明后	平台支持等（有则必备）
参考文献	全文末尾	遵照《中国细胞生物学学报》GB/T 7714-2025 格式规范，引用方法学文献、领域标准、同类数据集文献