

UKB RAP 平台使用教程

UK Biobank Research Analysis Platform Tutorial

云端生物信息学分析平台完整指南

涵盖项目管理 · 工具使用 · 数据分析 · 任务监控 · JupyterLab 等核心功能

日期: 2026 年 2 月 17 日

作者: Cancan

许可: 仅供个人学习使用

版权所有 © 2025 Cancan

本作品采用[知识共享署名-非商业性使用-禁止演绎 4.0 国际许可协议](#)进行许可

严禁商用、转载、修改或用于任何盈利目的

本教程旨在帮助研究人员快速掌握 UKB RAP 平台的使用方法

适合生物信息学研究者、数据分析师及相关领域从业者

目录

1 简介	1
1.1 平台特点	1
2 创建和管理 Project	1
2.1 创建新 Project	1
2.2 Project 文件管理	2
3 使用 Tools Library	3
3.1 访问工具库	3
3.2 常用工具	3
4 运行分析任务	3
4.1 配置运行参数	4
4.2 选择输出位置	5
4.3 详细的输出位置选择	5
5 使用 Table Exporter 工具	6
5.1 配置 Table Exporter	6
5.2 选择输入数据集	7
5.3 查看已选择的文件	7
6 数据上传	8
6.1 上传本地文件	8
6.2 选择上传文件	9
6.3 Table Exporter 完整配置	10
6.4 审查并启动任务	11
7 监控任务执行	11
7.1 查看任务状态	11
7.2 任务状态说明	12
7.3 查看任务日志	13
8 使用 JupyterLab	13
8.1 JupyterLab 列表	14

8.2 创建新的 JupyterLab 环境	14
8.3 在 JupyterLab 中工作	15
9 会话管理	16
9.1 会话超时处理	16
10 最佳实践建议	16
10.1 项目组织	16
10.2 资源管理	17
10.3 数据安全	17
10.4 分析流程	17
10.5 使用 Table Exporter 的建议	17
10.6 使用 JupyterLab 的建议	17
11 常见问题	18
11.1 任务失败怎么办?	18
11.2 如何优化分析速度?	18
11.3 数据下载	18
11.4 JupyterLab 相关问题	18
11.5 成本控制	19
12 附录	19
12.1 相关资源	19
12.2 联系方式	19
12.3 常用命令	19
12.3.1 dx-toolkit 命令行工具	19
12.3.2 JupyterLab 常用操作	20
13 总结	20

1 简介

UK Biobank Research Analysis Platform (UKB RAP) 是一个基于云计算的研究分析平台，为研究人员提供了强大的数据分析和处理能力。本教程将详细介绍如何使用该平台进行数据分析工作。

1.1 平台特点

- 基于 DNAnexus 平台构建
- 提供海量的 UK Biobank 数据资源
- 支持多种生物信息学分析工具
- 云端计算, 无需本地配置环境
- 灵活的资源管理和费用控制
- 支持 JupyterLab 交互式分析环境

2 创建和管理 Project

2.1 创建新 Project

在 UKB RAP 平台上, 所有的分析工作都需要在 Project 中进行。

1. 登录平台后, 点击右上角的「New Project」按钮创建新项目
2. 或者点击已有的 Project 进入项目管理界面

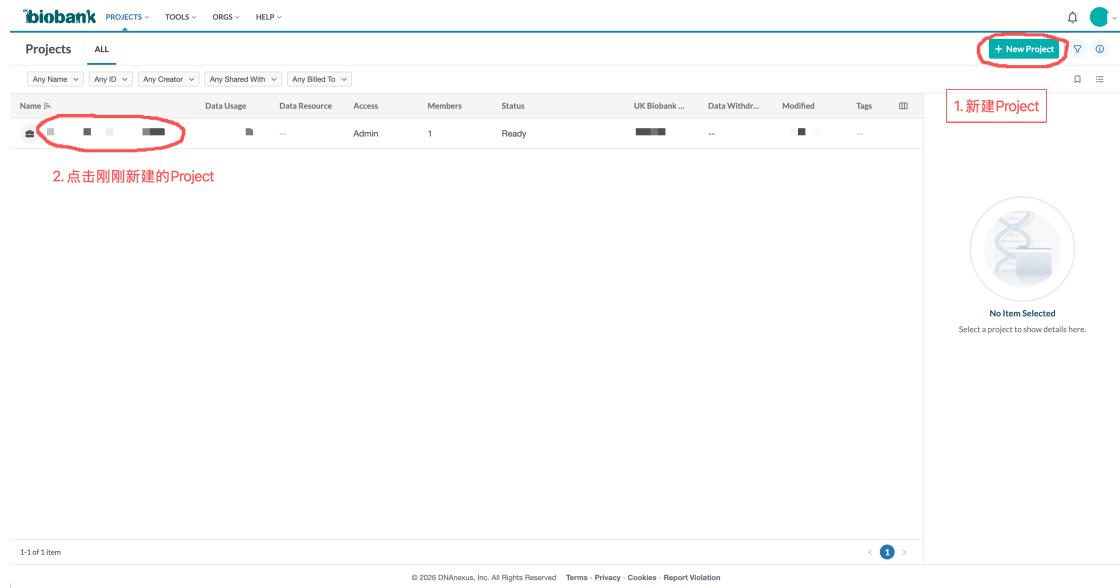


图 1: 创建新 Project 界面

2.2 Project 文件管理

进入 Project 后, 您可以看到文件管理界面:

- 左侧显示文件夹树形结构
- 中间区域显示当前文件夹内容
- 可以查看文件的类型、创建时间、大小等信息

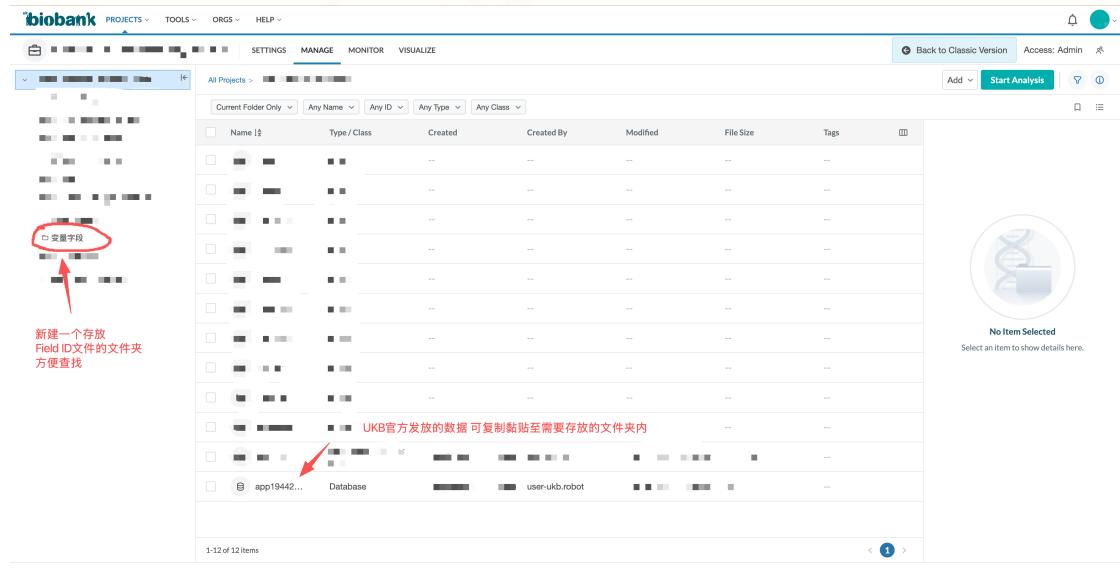


图 2: Project 文件管理界面

3 使用 Tools Library

3.1 访问工具库

UKB RAP 提供了丰富的生物信息学分析工具。

1. 点击顶部菜单栏的「TOOLS」
2. 进入 Tools Library 查看所有可用工具
3. 可以按名称、类别、类型等进行筛选

Name	Category	Type	Provided By	Latest Version	Region
REGENIE Marker Association ...	Translational Informatics	App	org-dnanexus_apps	2.1.0	◇
REGENIE Whole Genome Reg... SAIGE GWAS - Gene and regi...	Translational Informatics	App	org-dnanexus_apps	2.1.0	◇
SAIGE GWAS - Single variant ... SAIGE GWAS - Single variant asso...	...	App	org-dnanexus_apps	3.0.0	◇
SAIGE GWAS - Single variant ... SAIGE GWAS - Single variant asso...	...	App	org-dnanexus_apps	4.0.0	◇
SAIGE GWAS GRM SAIGE - MNLLGLMM to generate	App	org-dnanexus_apps	4.0.0	◇
SAIGE-GWAS Sparse GRM Creates a sparse genetic relation...	...	App	org-dnanexus_apps	2.0.5	◇
SnpEff Annotate Annotates variants in VCF format w...	Annotation	App	org-dnanexus_apps	3.1.0	◇
SnpSift Annotate Runs SnpSift annotation feature on ...	Annotation	App	org-dnanexus_apps	2.1.0	◇
Swiss Army Knife A multi-purpose tool for all your ba...	...	App	org-dnanexus_apps	5.1.0	◇
Table exporter Export a specific entity or the cohort	Translational Informatics	App	org-dnanexus_apps	2.0.84	◇
Tarball App Builder (Focal) Builds an app from a tarball	Debugging	App	org-dnanexus_apps	1.2.2	◇
Tarball App Builder (Noble) Builds an app from a tarball	Debugging	App	org-dnanexus_apps	0.1.0	◇
Tarball App Builder (Xenial v1) Builds an app from a tarball	Debugging	App	org-dnanexus_apps	1.2.2	◇
Tarball App Builder (Xenial) Builds an app from a tarball	Debugging	App	org-dnanexus_apps	1.1.1	◇

图 3: Tools Library 工具库

3.2 常用工具

平台提供的工具包括但不限于：

- **REGENIE:** 全基因组关联分析工具
- **SAIGE:** 基因组关联分析工具
- **Table exporter:** 表格数据导出工具
- **Swiss Army Knife:** 多功能数据处理工具
- 各类注释、调试工具等

4 运行分析任务

4.1 配置运行参数

选择工具后, 需要配置运行参数:

1. 点击工具的「Run App」按钮
2. 设置 Job Name(任务名称)
3. 选择 Output Location(输出位置)
4. 配置其他必要参数

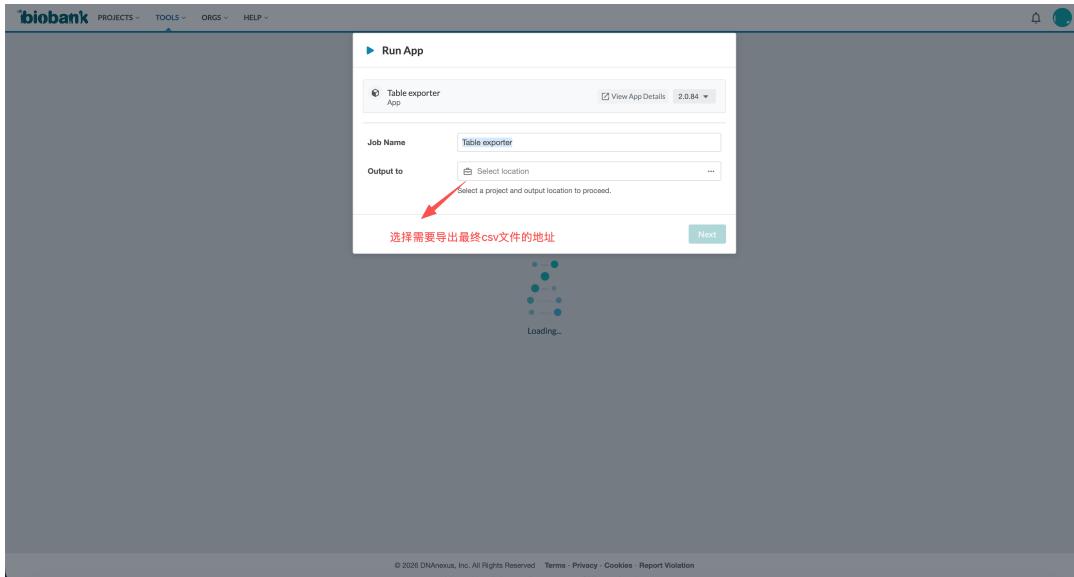


图 4: 配置运行参数

4.2 选择输出位置

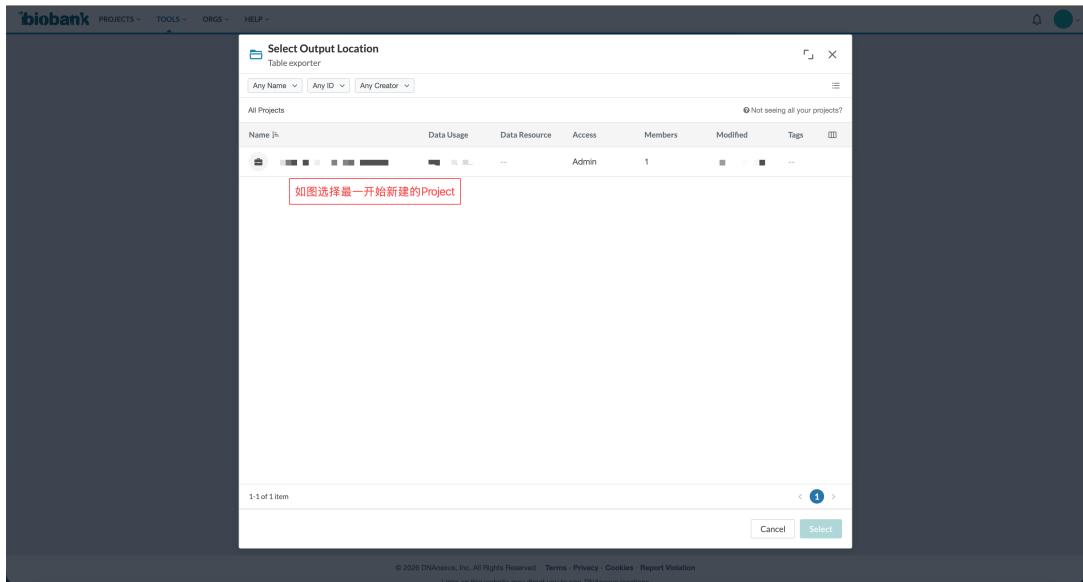


图 5: 选择输出位置

如图所示, 选择第一个开始的 Project 作为输出目录。

4.3 详细的输出位置选择

在选择输出位置时, 系统会弹出文件浏览器:

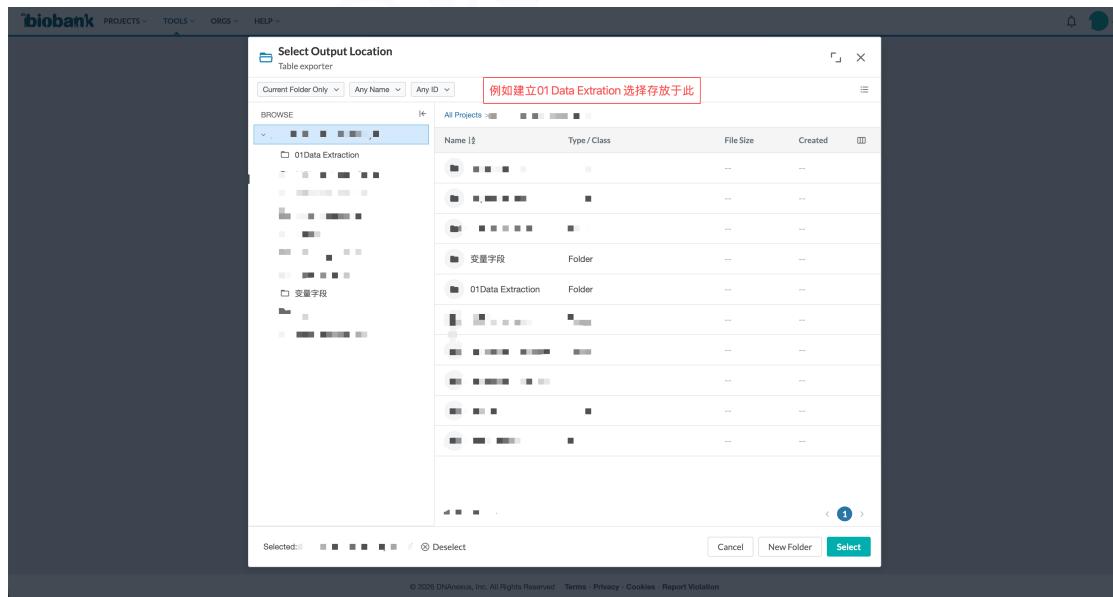


图 6: 输出位置文件浏览器

- 可以浏览 Project 中的所有文件夹
- 选择合适的文件夹作为输出位置

- 建议为不同的分析任务创建不同的输出文件夹

5 使用 Table Exporter 工具

Table Exporter 是 UKB RAP 平台上常用的数据导出工具, 可以将数据库中的表格数据导出为 CSV 等格式。

5.1 配置 Table Exporter

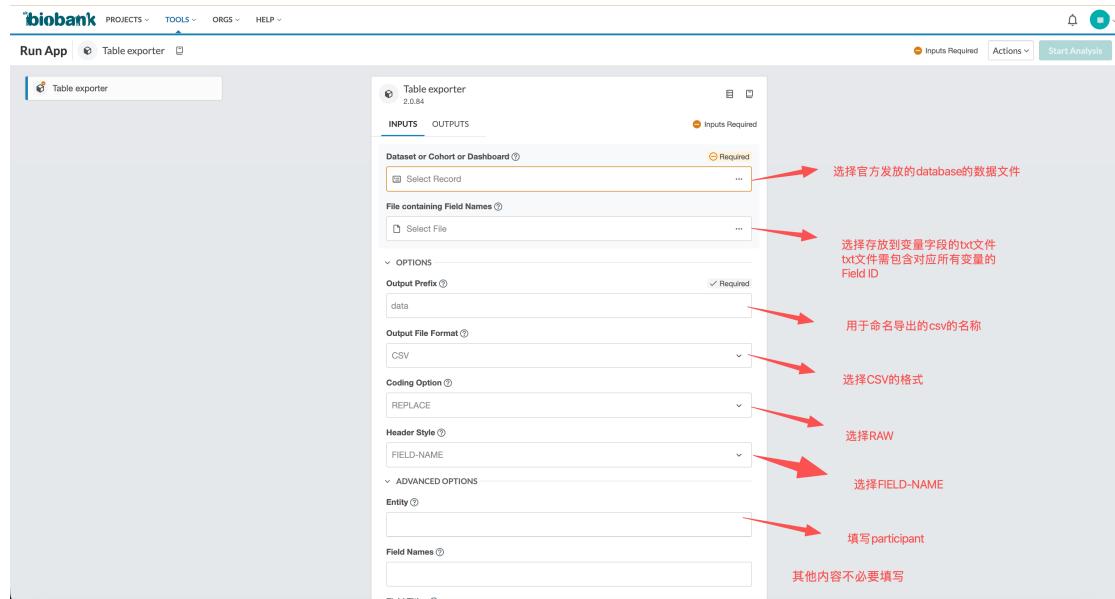


图 7: Table Exporter 配置界面

在配置界面中需要设置:

- Dataset or Cohort or Dashboard**: 选择要导出的数据集
- File containing Field Names**: 选择包含 Field ID 的文件
- Output Prefix**: 设置输出文件的前缀名称
- Output File Format**: 选择输出格式 (CSV/TSV 等)
- Coding Option**: 选择编码方式 (建议选择 RAW)
- Header Style**: 选择表头样式 (如 FIELD-NAME)

5.2 选择输入数据集

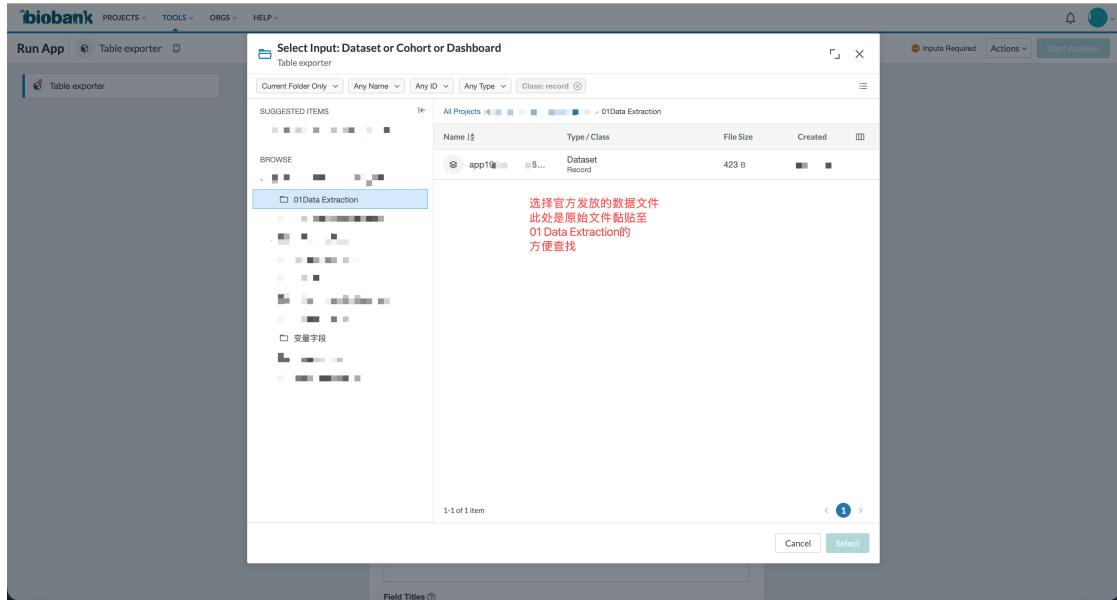


图 8: 选择输入数据集

在弹出的对话框中:

- 浏览并选择 01 Data Extraction 文件夹
- 选择相应的数据集文件

5.3 查看已选择的文件

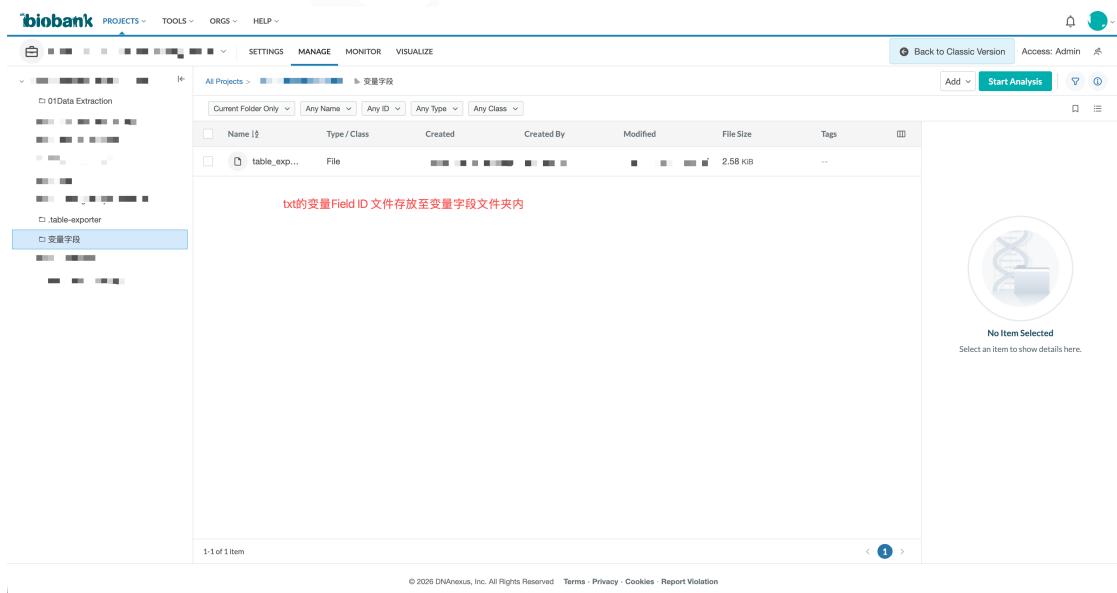


图 9: 已选择的 Field ID 文件

确认已正确选择包含 Field ID 的 txt 文件。

6 数据上传

6.1 上传本地文件

如果需要使用本地数据进行分析：

1. 点击右上角的「Add」按钮
2. 选择「Upload Data」
3. 可以拖拽文件或选择文件上传
4. 从本地上传到 AWS UKB-RNS-Rundev

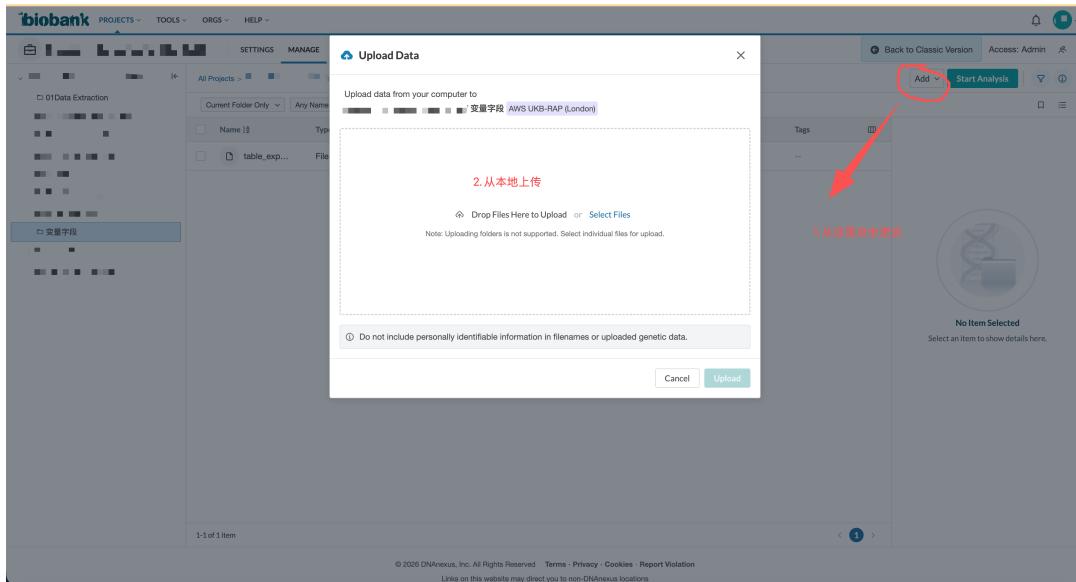


图 10：上传数据界面

6.2 选择上传文件

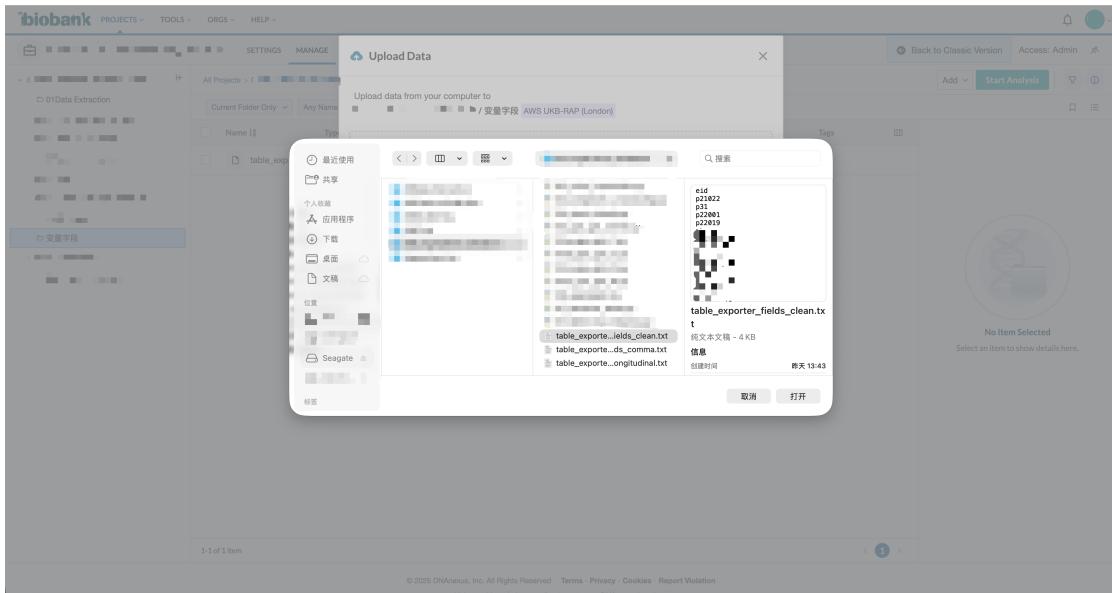


图 11: 本地文件选择界面

在文件选择对话框中:

- 浏览本地文件系统
- 选择需要上传的文件
- 支持多文件选择
- 可以看到文件的预览和大小信息

6.3 Table Exporter 完整配置

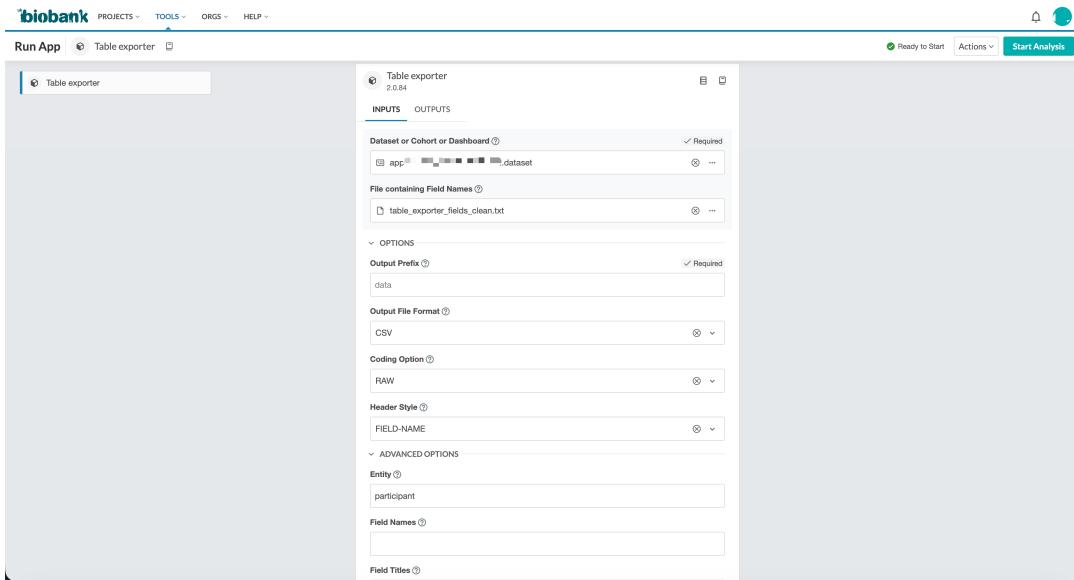


图 12: Table Exporter 完整配置

完成所有参数配置后，确认：

- 输入数据集已选择
- Field Names 文件已指定
- 输出格式和选项已设置
- 高级选项（如 Entity 和 Field Names）已配置

6.4 审查并启动任务

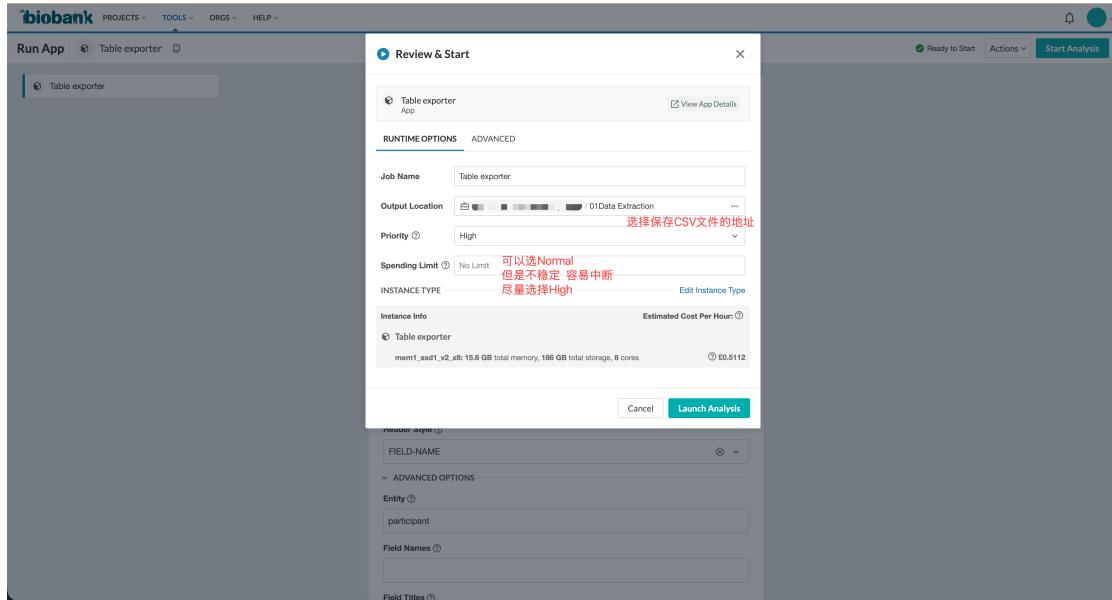


图 13: 审查并启动分析任务

在启动任务前, 需要:

- 设置 Job Name(任务名称)
- 确认 Output Location(输出位置)
- 设置 Priority(优先级, 通常选择 High)
- 设置 Spending Limit(花费限制, 可选择 Normal 或 High)
- 选择 Instance Type(实例类型)
- 查看预估成本
- 点击「Launch Analysis」启动任务

7 监控任务执行

7.1 查看任务状态

提交任务后, 可以在 Monitor 页面查看执行情况:

1. 点击顶部的「MONITOR」标签
2. 查看所有正在运行和已完成的任务
3. 可以看到任务的状态、执行时间、资源消耗等信息

The screenshot shows the UKB RAP Monitor interface. At the top, there are navigation tabs: PROJECTS, TOOLS, ORGS, HELP, SETTINGS, MANAGE, MONITOR (selected), and VISUALIZE. On the right, there are buttons for 'Access: Admin', 'Terminate All', and other monitoring options. The main area is titled 'All Executions /' and contains a table with columns: Name, State, Executable, Tags, Launched By, Launched On, Started Running, Duration, Cost, Priority, Worker, and a detailed icon. There are 8 executions listed:

Name	State	Executable	Tags	Launched By	Launched On	Started Running	Duration	Cost	Priority	Worker
Table exporter	Waiting	Table exporter (v2.0.84)					--	--	high	--
Table exporter	Done	Table exporter (v2.0.84)				6m	£0.0478 Final		high	--
Table exporter	Terminated	Table exporter (v2.0.84)				9h 5m	£1.9477 Final		high	--
Table exporter	Failed	Table exporter (v2.0.84)				5m	£0.0063 Final		normal	--
Table exporter	Failed	Table exporter (v2.0.84)				4m	£0.0053 Final		normal	--
如果没有按照之前步骤来 或者txt文件格式/Field ID格式有问题 会提示Failed										
成功提取时间大概5-6min 如果时间超过6m左右 就是某些格式问题 代码系统卡住 但不会有任何直接提示 会一直扣钱 需要及时停止										

At the bottom left, it says '1-8 of 8 Executions'. On the right, there's a 'No Item Selected' message and a note: 'Select an item to show details here.'

图 14: 任务列表和状态监控

在任务列表中可以看到：

- 任务名称和状态
- 执行时间和持续时间
- 成本信息
- 优先级和实例类型

The screenshot shows the UKB RAP Monitor interface with a task in the 'Running' state highlighted. A context menu is open over this task, with the 'View Log' option circled in red.

The task details in the context menu are:

- View Info
- Launch as New Job
- View Log** (highlighted)
- Terminate

The rest of the interface is similar to Figure 14, showing a list of tasks and their status.

图 15: 任务监控详细界面

7.2 任务状态说明

任务可能处于以下状态：

- **Waiting:** 等待中
- **Running:** 正在运行
- **Done:** 已完成
- **Failed:** 失败
- **Terminated:** 已终止

7.3 查看任务日志

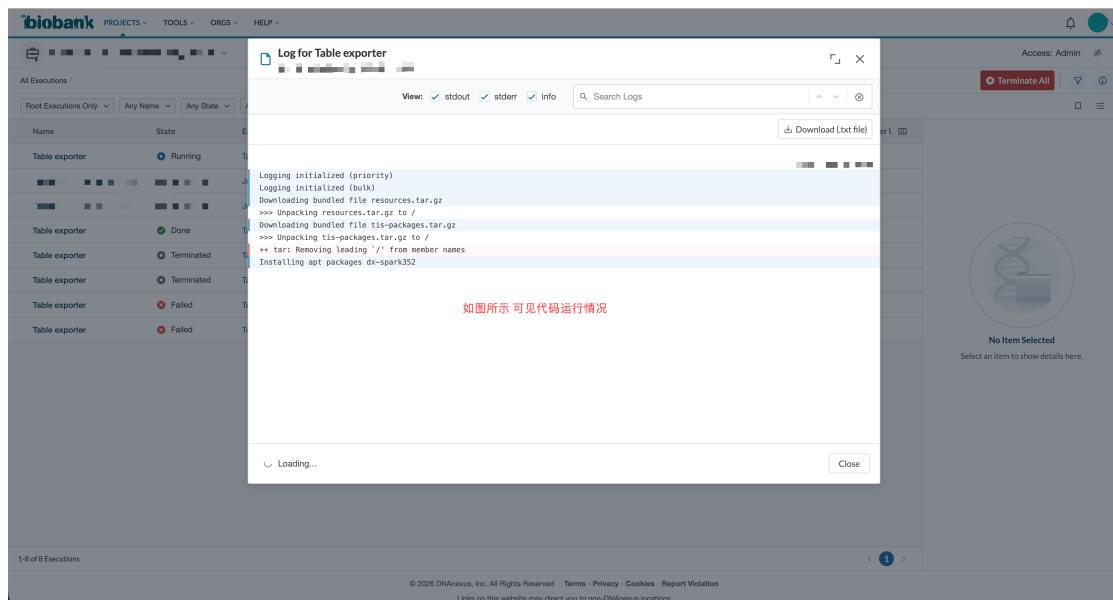


图 16: 查看任务执行日志

点击任务可以查看详细的执行日志:

- 查看任务的实时输出
- 检查是否有错误信息
- 了解任务执行进度
- 可以下载日志文件进行离线分析

8 使用 JupyterLab

UKB RAP 平台支持 JupyterLab 交互式分析环境, 可以进行更灵活的数据分析和可视化。

8.1 JupyterLab 列表

The screenshot shows the 'JupyterLab' section of the UKB RAP platform. At the top, there's a navigation bar with 'PROJECTS ~', 'TOOLS ~', 'ORGs ~', and 'HELP ~'. On the right side of the header, there are icons for user profile, notifications, and search. Below the header, a sub-header 'JupyterLab' is displayed. A table lists the existing JupyterLab environment. The table has columns: 'Name', 'Project', 'Executable Name', 'Version', 'Created At', and 'Status'. One entry is shown: 'JupyterLab -' (Executable Name: dxjupyterlab, Version: 2.10.0). To the right of the table, there's a circular icon with a DNA helix and a beaker, labeled 'No Item Selected' with the sub-instruction 'Select an item to show details here.' Below the table, a message '在 Tool 里面建立一个 JupyterLab' is overlaid. At the bottom of the page, there's a footer with copyright information: '© 2026 DNAnexus, Inc. All Rights Reserved Terms · Privacy · Cookies · Report Violation'. Navigation arrows and a page number '1-2 of 2 items' are also present.

图 17: JupyterLab 环境列表

在 JupyterLab 页面可以:

- 查看已创建的 JupyterLab 环境
- 查看环境的版本和状态
- 点击「New JupyterLab」创建新环境

8.2 创建新的 JupyterLab 环境

The screenshot shows the 'New JupyterLab' dialog box. The dialog has a title '+ New JupyterLab'. It contains several input fields and dropdown menus:

- Environment Name:** JupyterLab -
- Project:** (dropdown menu showing a project name)
- Snapshot (Optional):** (dropdown menu with 'Select Snapshot' option)
- Priority:** High (dropdown menu with '尽量使用High 防止中断' option)
- Cluster Configuration:** Single Node (radio button selected)
- Instance Type:** mem2_ssd2_gput1_v2_x32 (dropdown menu)
- Duration (in hours):** 4 (input field)
- Feature:** PYTHON_R (dropdown menu)

 At the bottom, it says 'Estimated Price: £5.5984' based on the selected parameters. There are 'Cancel' and 'Start Environment' buttons. A note on the right side of the dialog box reads:

Instance Type 按照实际需求选择
如进行深度学习等尽量使用 GPU
提高效率，价钱不会相差太多
建议代码提前本地准备分
因为 RAP 平台按需自动收费

图 18: 创建新的 JupyterLab 环境

创建 JupyterLab 环境时需要配置:

- **Environment Name:** 环境名称
- **Project:** 关联的 Project
- **Snapshot:** 快照设置 (可选)
- **Priority:** 优先级 (建议使用 High)
- **Cluster Configuration:** 集群配置
 - Single Node: 单节点
 - Spark Cluster: Spark 集群
- **Instance Type:** 实例类型 (根据需求选择)
- **Duration:** 运行时长 (小时)
- **Runtime:** 运行时环境 (如 PYTHON R)

8.3 在 JupyterLab 中工作

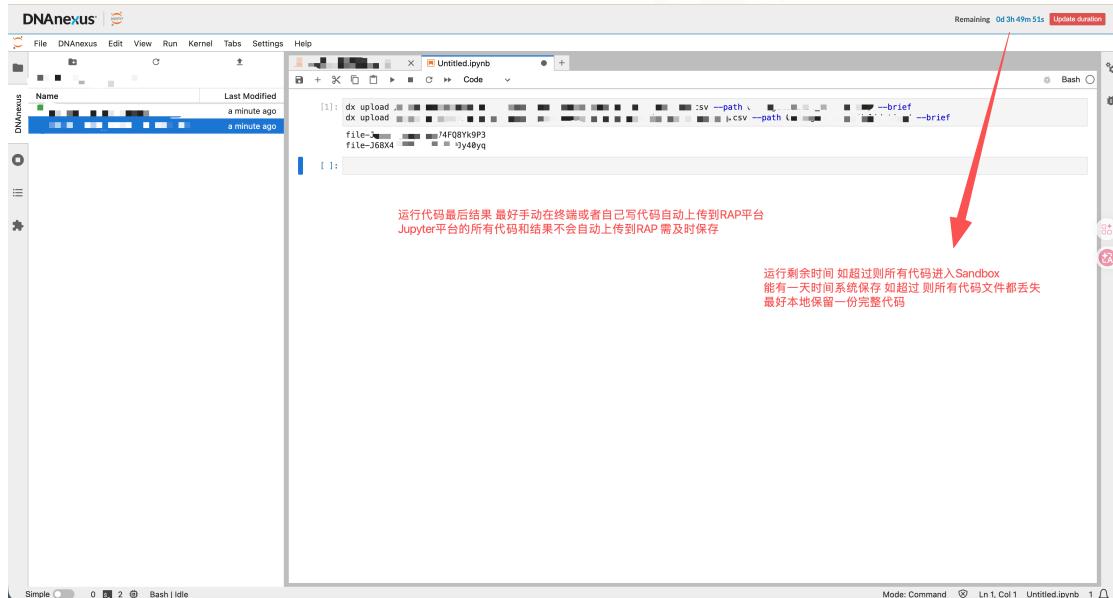


图 19: JupyterLab 工作界面

JupyterLab 提供了完整的交互式开发环境:

- 支持 Python、R 等多种编程语言
- 可以直接访问 Project 中的数据文件
- 支持数据可视化和交互式分析
- 可以安装自定义的 Python 包
- 支持终端命令行操作

- 可以创建和编辑多种类型的文件

9 会话管理

9.1 会话超时处理

UKB RAP 平台会话有时间限制, 当会话即将过期时:

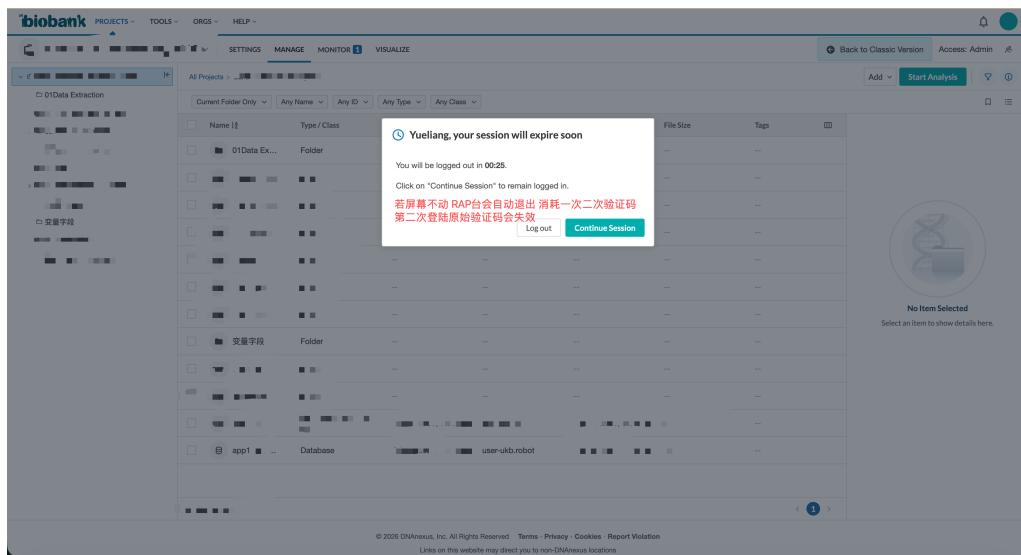


图 20: 会话超时提示

10 最佳实践建议

10.1 项目组织

1. **合理规划文件夹结构:** 为不同类型的数据和分析结果创建独立文件夹
 - 原始数据文件夹
 - 中间结果文件夹
 - 最终结果文件夹
 - Field ID 文件夹
2. **命名规范:** 使用清晰、有意义的文件名和任务名
3. **定期清理:** 删除不需要的临时文件, 节省存储空间
4. **文档记录:** 记录分析流程和参数设置

10.2 资源管理

1. **选择合适的实例类型:** 根据任务需求选择适当的计算资源
2. **监控资源使用:** 定期查看 CPU、内存使用情况
3. **及时终止任务:** 对于不需要的或出错的任务及时终止，避免浪费资源
4. **成本控制:**
 - 设置合理的 Spending Limit
 - 不使用时关闭 JupyterLab 环境
 - 选择合适的优先级

10.3 数据安全

1. **遵守数据使用协议:** 严格按照 UK Biobank 的数据使用规定操作
2. **不上传敏感信息:** 上传文件时注意不要包含个人身份信息
3. **定期备份:** 重要的分析结果应及时下载备份
4. **访问控制:** 合理设置 Project 的访问权限

10.4 分析流程

1. **先小规模测试:** 在大规模分析前，先用小数据集测试流程
2. **记录分析参数:** 详细记录每次分析使用的参数和配置
3. **查看日志:** 任务完成后检查日志，确保分析正确执行
4. **验证结果:** 对分析结果进行合理性检查
5. **版本控制:** 对于复杂的分析脚本，建议使用版本控制

10.5 使用 Table Exporter 的建议

1. **准备 Field ID 列表:** 提前准备好需要导出的 Field ID
2. **选择合适的编码方式:** 根据需求选择 REPLACE 或其他编码选项
3. **注意输出格式:** CSV 格式通用性好，但大文件可能需要压缩
4. **检查导出结果:** 导出后检查数据完整性和正确性

10.6 使用 JupyterLab 的建议

1. **选择合适的实例:** 根据数据量和计算需求选择实例类型
2. **设置合理时长:** 避免环境自动关闭导致工作丢失
3. **定期保存:** 养成定期保存的习惯

4. 下载重要结果: 及时下载分析结果和脚本
5. 使用快照功能: 对于重要的环境配置可以创建快照

11 常见问题

11.1 任务失败怎么办?

- 查看任务的 Log 日志, 找出错误原因
- 检查输入文件格式是否正确
- 确认参数设置是否合理
- 检查是否有足够的存储空间
- 确认 Field ID 文件格式正确
- 如果问题持续, 联系技术支持

11.2 如何优化分析速度?

- 选择更高配置的实例类型
- 合理分割大任务为多个小任务并行执行
- 使用平台推荐的优化工具
- 避免在高峰时段提交大量任务
- 对于重复性任务, 考虑使用 JupyterLab

11.3 数据下载

- 小文件可以直接通过浏览器下载
- 大文件建议使用命令行工具 (如 dx-toolkit)
- 注意下载的数据仍需遵守使用协议
- 可以在 JupyterLab 中直接下载分析结果

11.4 JupyterLab 相关问题

- 环境启动慢: 选择 High 优先级可以减少等待时间
- 会话断开: 检查网络连接, 必要时重新连接
- 包安装失败: 检查包名和版本, 使用 pip 或 conda 安装
- 内存不足: 选择更大内存的实例类型
- 文件访问问题: 确认文件路径正确, 检查权限设置

11.5 成本控制

- 定期检查账户余额和使用情况
- 及时终止不需要的任务和环境
- 选择合适的实例类型, 避免过度配置
- 使用 Spending Limit 控制单个任务的成本
- 对于长时间运行的任务, 评估成本效益

12 附录

12.1 相关资源

- UK Biobank 官网:** <https://www.ukbiobank.ac.uk/>
- DNAexus 文档:** <https://documentation.dnanexus.com/>
- RAP 用户指南:** 平台内置帮助文档
- JupyterLab 文档:** <https://jupyterlab.readthedocs.io/>

12.2 联系方式

如有问题, 可通过以下方式获取帮助:

- UK Biobank 技术支持邮箱
- 平台内的 Help 菜单
- 用户社区论坛
- DNAexus 技术支持

12.3 常用命令

12.3.1 dx-toolkit 命令行工具

```
1 # 登录
2 dx login
3
4 # 列出项目
5 dx ls
6
7 # 下载文件
8 dx download <file-id>
9
```

```
10 # 上传文件  
11 dx upload <local-file>  
12  
13 # 查看任务状态  
14 dx find jobs
```

12.3.2 JupyterLab 常用操作

```
1 # 安装 Python 包  
2 !pip install <package-name>  
3  
4 # 列出文件  
5 !ls -lh  
6  
7 # 读取数据  
8 import pandas as pd  
9 data = pd.read_csv('data.csv')  
10  
11 # 数据可视化  
12 import matplotlib.pyplot as plt  
13 plt.plot(data['x'], data['y'])  
14 plt.show()
```

13 总结

本教程详细介绍了 UKB RAP 平台的完整使用方法, 包括:

- Project 的创建和管理
- 工具库的使用
- Table Exporter 的详细配置流程
- 数据上传和文件管理
- 任务的配置、启动和监控
- JupyterLab 交互式分析环境的使用
- 会话管理和最佳实践建议

希望本教程能帮助您快速上手 UKB RAP 平台, 顺利开展研究工作。随着平台的不断更新, 建议定期查看官方文档获取最新信息。

祝您研究顺利!