

北京大学多组学信息平台服务指南

一、平台简介

多组学信息平台目前配备高性能计算平台及大容量数据存储系统,致力于提供高效的(单细胞)转录组、表观组、基因组、蛋白组、代谢组、微生物组等组学分析及多组学整合分析服务。除常规的流程性分析之外,平台可根据用户的具体需求,开展定制化数据分析和生物信息学算法开发服务。

二、平台组织架构

平台负责人: 吴华君

数据分析专家库: 郑小琪(上海交通大学单细胞信息平台主任)、石建涛(中科院生化细胞所信息平台主任)、刘林(上海交通大学)等。

平台单位: 北京大学医学部多组学研究中心、北京大学肿瘤医院科研处

平台联系方式: multiomics@126.com

三、平台管理办法

平台负责人确定对外合作课题内容、方向及合作细节,并指导技术人员进行设备管理、设备使用、技术开发、数据管理和分析等。技术人员和院内外合作者在平台负责人和专家的指导下进行课题的调研和实施,并根据用户需求进行调整和反馈。

四、服务器用户须知与使用指南

用户需求申请表

申请人姓名：_____

身份证号：_____

单位：_____

需求：_____

用户须知

1. 账号负责人必须为北京大学医学部及附属医院在职教师或博士后、学生，北大其他学院及校外机构的人员需要与北京大学医学部开展相关合作事务方可申请账号。账号负责人须通过北京大学网络信息中心申请开通 VPN（首次登录 VPN 需按要求强制修改密码）；
2. 请妥善保管账号，一人一号，不得外借；
3. 本计算平台仅提供教学和科研相关的服务，严禁用于军工等涉密课题；
4. 用户在申请账号前请仔细阅读本须知，确保在使用过程中遵守相关规章制度，维护平台环境，用户不得将平台资源用于与教学或科研无关的事宜。禁止主观破坏集群环境的行为，一经发现，将立即终止服务并追究责任并要求赔偿平台损失；
5. 用户须自行备份重要的数据及相关文件；
6. 其它需注意的问题，请查看用户手册。

申请人：（签字）

二〇 年 月 日

五、用户手册

1. 服务器硬件资源

本计算平台采用高性能高密度存储系统作为基础存储平台，采用模块化刀片服务器构建高密度、高性能的计算集群，同时采用高速交换系统，存储系统直接接入存储光纤交换系统进行本地高速交换。计算集群采用 CentOS 的操作系统，配以 Slurm 作业调度系统统一调度计算节点资源和存储资源。集群包括登录节点和两个计算节点，登录节点包含 32 个 Intel(R) Xeon(R) Silver 4215R @ 3.20GHz CPU 双线程核心，计算节点 1 含有 112 个 Intel(R) Xeon(R) Gold 6330N @ 2.20GHz CPU 双线程核心、2 张 NVIDIA A100 80G 显卡和 2T 内存，计算节点 2 含有 112 个 Intel(R) Xeon(R) Gold 6330N CPU @ 2.20GHz CPU 双线程核心、8 张 NVIDIA GeForce 3090 24G 显卡和 512G 内存，存储节点包含 300T 硬盘空间，平台网络采用高速 56Gb FDR Infiniband 连接。

2. 服务器的登录

服务器的登录节点 IP 地址为 `multiomics.pku.edu.cn`，端口为 22，登录节点为 `head01`。

用户通过 terminal 或 shell 软件登录服务器，方法如下：

```
$ ssh -p 22 username@multiomics.pku.edu.cn
```

输入密码即可登录到用户的 home 路径。

用户在校外登录，需安装北大 VPN 软件，方法详见北京大学 `vpn` 网站中的 VPN 客户端安装说明（https://its.pku.edu.cn/service_1_vpn_client.jsp）。

用户首次登录后需修改初始密码，方法如下：

```
$ passwd
```

输入当前密码、新密码并确认即可修改完成。

shell 软件的选择：Windows 可使用 Xshell 或 MobaXterm，OS 系统可使用 terminal 或 iterm2，同时也推荐有条件的同事、同学安装使用 warp。

3. 办公电脑与服务器之间的文件传输

用户通过 FTP 软件登录服务器，方法如下：

- 1) 新建会话
- 2) 输入主机（即服务器 `head01` 节点 ip 地址，`multiomics.pku.edu.cn`）、端口号（22）、协议

(SFTP)、用户名、密码、远程路径（可选）

即可建立与服务器之间的连接，用于文件的上传下载。

FTP 软件的选择：Windows 可使用 Xftp 或 FileZilla，OS 系统亦可使用 FileZilla。

4. 使用服务器的计算资源

用户在计算节点使用 Slurm 提交任务，用户根据实际需求申请计算资源。

5. 需求申请

新用户申请账号时，需填写《用户需求申请表》提交给管理员，管理员确认开通账号后新用户即可登录和使用服务器。用户如有安装软件等其他需求，也需填写《用户需求申请表》并提交给管理员，等待管理员反馈。

6. 注意事项

- (1) 建议用户使用 nfs 挂载盘的指定路径安装软件、存储数据。
- (2) 不得在非管理员指定的节点提交任务。
- (3) 不得将服务器上的文件和数据等资料外传，不可使用有数据泄露风险的软件工具等。

7. 信息反馈

用户对服务器有任何疑问和需求，请和管理员沟通。

8. 附加材料

- (1) 参考基因组数据路径：/Software/Refs

1) 常用软件的预装路径：

Software	Version	Location
Python	3.8.5	/Software/Python/anaconda3/bin
R	4.0.3	/Software/R/current/bin
Cellranger	5.0.0	/Software/TenXGenomics/Cellranger
Cellranger-arc	1.0.0	/Software/TenXGenomics/Cellranger-arc
Cellranger-atac	1.2.0	/Software/TenXGenomics/Cellranger-atac
nf-core/sarek	2.7	/Software/nf_core/nf-core-sarek-2.7
nf-core/rnaseq	3.0	/Software/nf_core/nf-core-rnaseq-3.0
fastQC	0.11.9	/Software/Packages/FastQC
Trim_galore	0.6.5	/Software/Packages/TrimGalore-master

fastp	0.21.0	/Software/Packages/fastp
picard	2.22.9	/Software/Packages/picard.jar
bwa	0.7.17	/Software/Packages/bwa
STAR	2.7.3	/Software/Packages/STAR
HASAT2	2.2.0	/Software/Packages/hisat2-2.2.0
RSEM	1.3.3	/Software/Python/anaconda3/bin
GATK	4.1.6	/Software/Packages/gatk-4.1.6.0
samtools	1.10	/Software/Packages/samtools-1.10
Scaller	2.0.0	/Software/Packages/SCcaller-master
VarScan	2.4.0	/Software/Packages/varscan
vep	101	/Software/Packages/vep/ensembl-vep-release-101
SnEff	4.3	/Software/Packages/snpEff
bcl2fastq	2.20	/Software/Packages/bcl2fastq2-v2.20/installed-dir/bin
vcftools	0.1.17	/Software/Packages/vcftools
bcftools	1.10.2	/Software/Packages/bcftools-1.10.2
sratoolkit	2.10.5	/Software/Packages/sratoolkit.2.10.5-centos_linux64

(2) 常见问题

Q: 如何使用 slurm 提交计算任务?

脚本示例 (test.slurm):

```
#!/bin/bash
#SBATCH --job-name=test      # 作业名
#SBATCH --partition=cpu     # cpu 队列
#SBATCH -n 20               # 使用核数
#SBATCH --ntasks-per-node=4 # 每节点核数
#SBATCH --output=%j.out
#SBATCH --error=%j.err
Rscript run.R
```

后台提交任务:

```
$ sbatch test.slurm
```

交互式任务:

```
$ srun R --save
```

查看总体资源信息:

```
$ sinfo
```

六、收费标准（暂行）

1. 数据分析服务收费标准

项目	类型	流程性分析			定制化分析（45 个工作日）	
		包含内容	数据类型	价格（元/样本）	包含内容	价格（元/样本）
基因组测序	Bulk	质控、序列比对、SNV/indel 分析、CNV 分析、突变注释	WGS WES Panel	2000 1000 200	驱动基因分析 GISTIC 分析 突变指纹分析 TCR/BCR 组装 HLA 分型 新抗原分析 非编码突变分析 结构变异分析	5000 3000 -
	Hi-C /ChIA-PET	质控、序列比对、互作矩阵矫正、Compartment、TAD、loop 分析	Hi-C	3000	调控元件分析 调控网络分析 整合分析	7000
	单细胞	质控、序列比对、SNV/indel 分析、CNV 分析、突变注释	scWGS scWES	3000	系统进化树分析 整合分析	10000
转录组测序	Bulk	质控、序列比对、定量、差异分析、功能富集、基因/样本聚类	RNA-seq	500	生存分析 免疫细胞浸润分析 TCR/BCR 组装 共表达网络分析 转录调控网络分析	3000
	单细胞	质控、比对、定量、细胞聚类、细胞注释、差异分析、功能富集	scRNA-seq /scATAC-seq	2000	轨迹推断 细胞速率 拷贝数分析 TCR/BCR 组装 受体/配体分析 转录调控网络分析	10000
	空间转录组	质控、比对、定量、细胞聚类、细胞注释、差异分析、功能富集	spatial transcriptomics	2000	空间聚类 空间差异基因分析 空间数据解卷积 单细胞的所有分析	10000
DNA 甲基化	Bulk/单细胞	质控、比对、甲基化定量、差异甲基化分析、功能富集	WGBS RRBS 450k	1000 500 200	样本聚类 ASM 分析 免疫细胞浸润分析	2500 1500 1200

					表达谱整合	
非编码RNA		质控、比对、定量、差异分析、功能富集、分类	miRNA lncRNA circRNA	500	靶基因预测	1500
表观遗传数据		质控、比对、Peak calling、功能富集	ChIP-seq Cut&Tag ATAC-seq	500 500 500	motif 分析 增强子分析 转录因子活性 表达谱整合	3000 3000 3000
微生物组	扩增子	比对、去噪、构建进化树、alpha 多样性分析、beta 多样性分析、物种注释、组成分析	16S/18S	500	物种组成分析 网络分析 代谢通路分析	4000
	宏基因组	质控、比对、物种多样性分析、拼接、分箱	DNA-seq	3000	物种组成分析 网络分析 代谢通路分析	7000
质谱数据	蛋白质组	差异表达、功能富集	蛋白质组	600	样本聚类 标志物筛选 免疫分析	3000
	代谢组	差异分析、功能富集	代谢组	600	样本聚类 标志物筛选	2000
多组学整合	面议	包括已发表的 bulk 组学数据和单细胞组学数据的整合分析	-	-	面议	面议
人工智能算法	机器学习、深度学习	经典机器学习方法、DNN/CNN/RNN/LSTM/ResNet/Transformer/VAE/GNN/GCN 等经典深度学习框架	特征数据、序列数据、影像数据	10000	多模态模型、个性化框架搭建、软件开发	> 20000
新算法和分析流程开发	-	根据用户需求，开发测试新的数据分析算法和流程。按照复杂程度和开发周期收费。	-	-	-	面议

注：北京大学及附属医院课题组可享受[八折优惠](#)（挂名），[五折优惠](#)（共一作第二+共责任第二）。

注：大样本合作方式和价格面议。

注：非紧密合作的定制化数据分析项目一般在 45 个工作日内结题，不负责无限期分析服务。

2. 生信平台使用收费标准（以账户为单位）

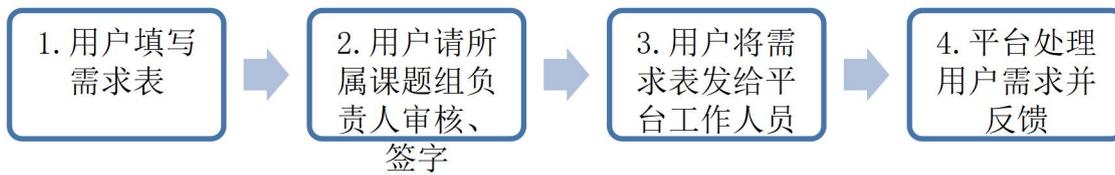
类型	类别	价格
账户	基础服务费	2000/年
计算设备	CPU	0.08 元/核时
	GPU	3 元/核时

存储设备	硬盘	免费存储 1TB, 超出部分 300 元/TB/年(按天扣费)
包年账户	所有服务面议	50000/年

注：仅提供给北京大学及附属医院课题组使用。

七、用户需求申请与定制化服务申请流程

1. 用户需求申请流程



2. 用户定制化服务申请流程

