



中国科学院北京基因组研究所（国家生物信息中心）

BEIJING INSTITUTE OF GENOMICS CHINESE ACADEMY OF SCIENCES / CHINA NATIONAL CENTER FOR BIOINFORMATION

# 组学原始数据汇交与共享

王彦青

2024年10月26日

# 提纲



**GSA Family介绍**



**GSA数据提交及共享**



**GSA-Human数据提交及共享**

# 组学原始数据管理体系 (GSA Family)



为组学原始数据统一汇交、存储、管理、共享的公共平台，保障国家数据安全



# 组学原始数据归档库GSA



组学原始数据汇交、存储、管理、发布与归档系统

equivalent

SRA  
Sequence  
Read Archive  
@NCBI

DRA  
DDBJ Read  
Archive  
@DDBJ

ENA  
European  
Nucleotide  
Archive  
@EBI

International Nucleotide Sequence Database  
Collaboration (INSDC)

The International Genome Sequence archive databases

Wang, Yanqing et al. *Genomics Proteomics Bioinformatics*, 2017  
Chen, Tingting et al. *Genomics Proteomics Bioinformatics*, 2021

# 国际数据整合

整合国际组学数据，实现SRA数据的本地镜像、同步和共享



# GSA-Human: 人类遗传资源组学原始数据管理

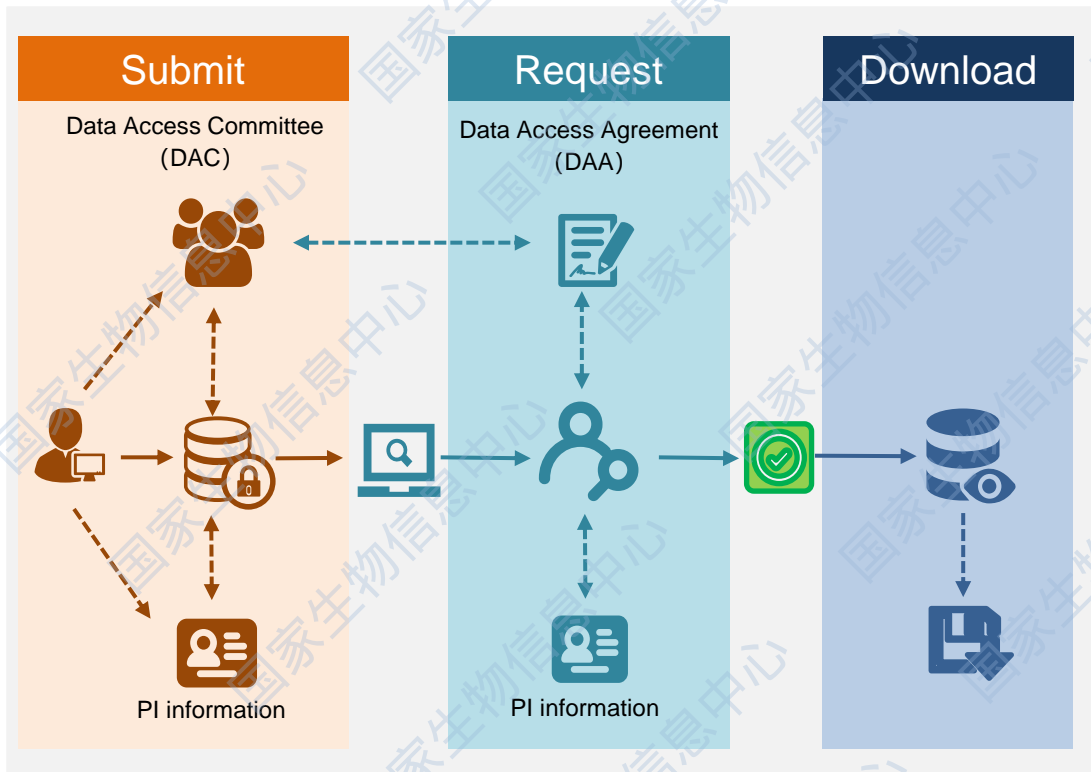


Chen, Tingting et al. *Genomics Proteomics Bioinformatics*, 2021  
Zhang, Sisi et al. *Yi Chuan*, 2021

- 聚焦人类遗传资源组学原始数据汇交、归档、管理、可控共享

- 对标NCBI的dbGap和EBI的EGA
- 对接国家人遗数据管理
- 管理>58万人数据
- 归档数据量**38.13PB**
- 支持**883篇**文章发表

# GSA-Human系统核心功能要素

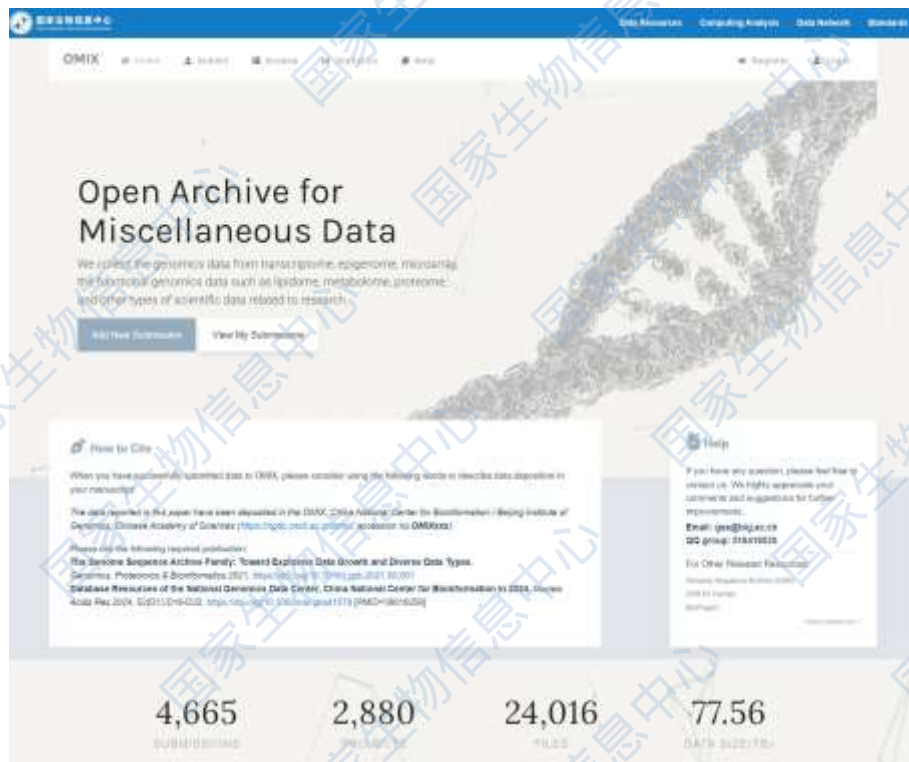


- 支持**5种**研究类型数据汇交
- 支持**2种**文件上传方式
  - Aspera和FTP
- 支持**2种**访问方式
  - 受控访问 (Controlled-access)
  - 公开访问 (Open-access)
- 受控访问采用申请审核制
  - 数据管理委员会 (Data Access Committee, DAC)
  - 审核数据访问申请
  - 授予访问权限
- PI账号提交和申请数据：数据责任人



# 多元数据归档库 OMIX

汇交、存储、共享 来自生物与医学  
领域研究的多元数据



Chen, Tingting et al. *Genomics Proteomics Bioinformatics*, 2021

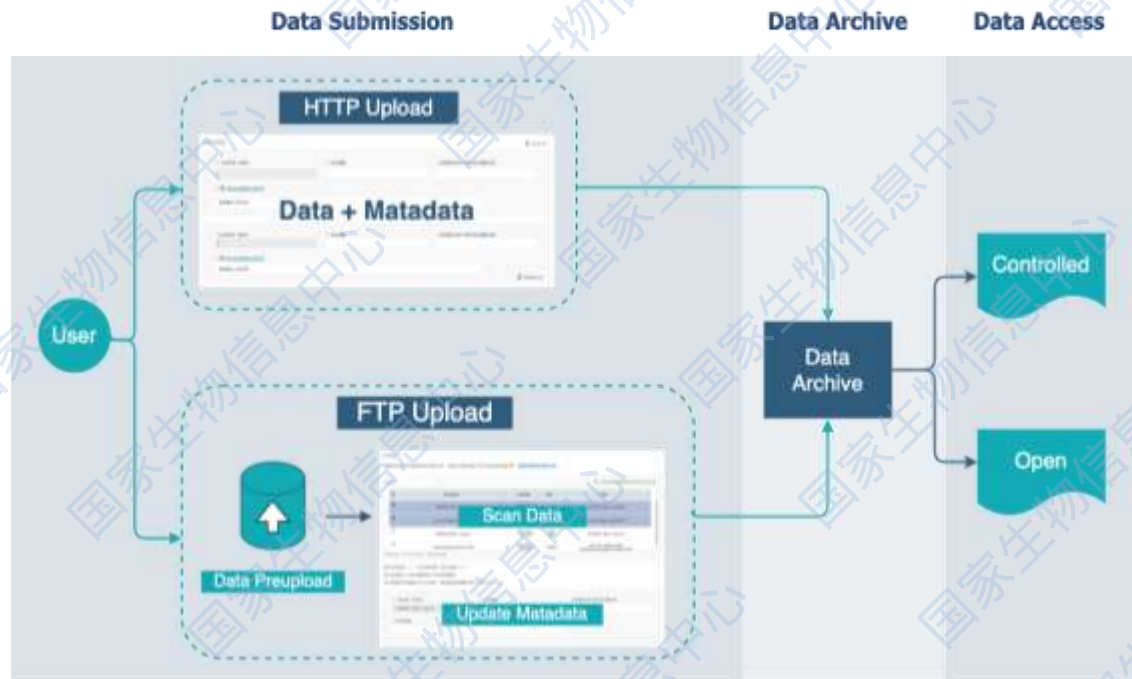


- 已通过 **FAIRSharing** 和 **re3data.org** 认证
- 获得 **40** 个出版集团的 **250** 个期刊认可，支撑文章 **604** 篇



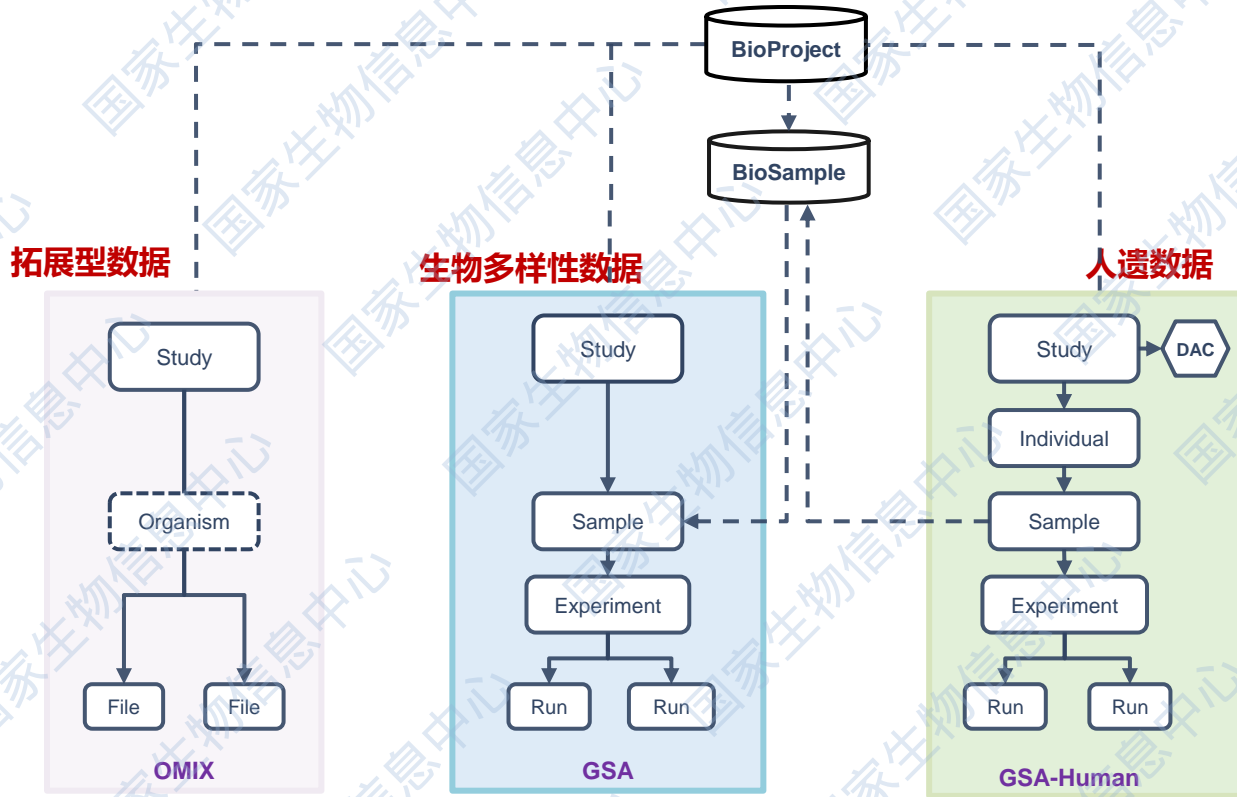


# 应用灵活可扩展的管理模式，支撑多模态数据汇交共享



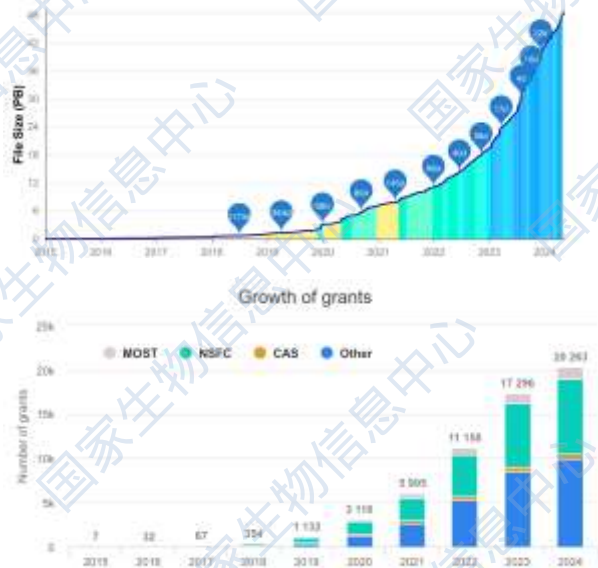
- 支持**10**个大类，**31**个小类，超过**70**种格式数据汇交
- 支持**2**种数据访问策略
  - 开放访问
  - **受控访问** (仅对人遗数据)
- 支持**2**种文件上传方式
  - HTTP在线上传
  - FTP预上传 (>1 TB 可协助上传)

# GSA Family数据模型



# 运行效果

- ❑ GSA Family得到生命领域几乎所有国际主流期刊认可
- ❑ 汇交数据量超 **56 PB**，支撑发表文章**3900** 篇
- ❑ 汇交数据量**每12个月翻一番**，超过“摩尔定律”增速，**增长1PB**数据用时最短**4天**



**>120万** 独立IP  
**>17亿** 次下载

**> 190** 国家/地区  
**> 9PB** 下载量

# 获得国际著名出版商认可

**SPRINGER NATURE**

Research data policy

Biological sciences repository examples

- Imaging
- Nucleic acid sequence and omics
  - Nucleic acid sequence data and metadata should follow the Genome Standards Consortium (GSC) guidance, which can be browsed at FAIRsharing GSC collection.
  - Data types
    - DNA sequence data\*
    - RNA sequence data\*
    - Genome assembly data\*
  - Repositories
    - Any INSOC member repository
    - Genome Sequence Archive (GSA)
  - Genetic variation data
    - dbSNP (human variations less than 50bp)
    - dbVar (human variations greater than 50bp)
    - European Variation Archive (EVA) (all variants)
    - Genome Sequence Archive for Human (human variations)

<https://www.springernature.com/gp/authors/research-data-policy/repositories-bio>

**ELSEVIER**

About Elsevier Products & Solutions Services

Home > Journals > Data in Brief > Policies and Guidelines > Public repositories to store and find data

**Data in Brief**

Submit your Paper View Articles

A Guide for authors Track your paper

Public repositories to store and find data

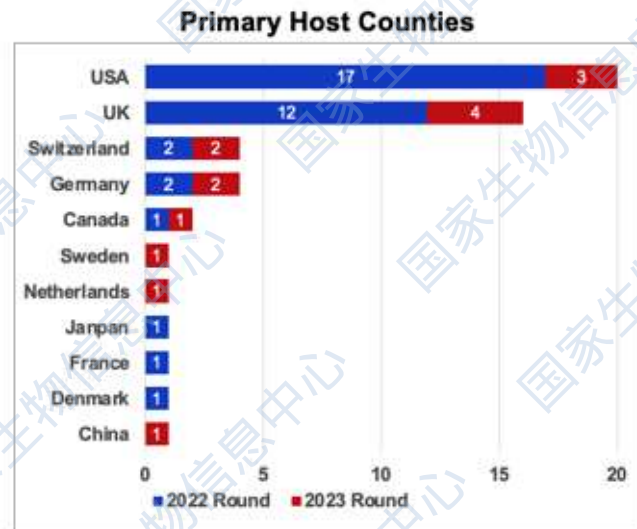
2. Specialized Repositories are recommended. Please note that this list is not exhaustive. We encourage the use of Re3data and FAIRsharing to find additional specialized repositories in a specific discipline.

Discipline	Type of data	Repository name	Cost
Raw sequencing data, genome assemblies, annotated sequences, and sample metadata		INSOC repositories	Free
		Genome Sequence Archive (GSA)	Free
		dbSNP	Free
		dbVar	Free
Genetic variation data		ClinVar	Free
		European Variation Archive (EVA)	Free
		Genome Sequence Archive for Human	Free

<https://www.journals.elsevier.com/data-in-brief/policies-and-guidelines/public-repositories-to-store-and-find-data>

# GSA入选全球核心数据资源 (GCBR)

GSA入选由国际生物数据联盟 (Global Biodata Coalition, GBC) 发起的**全球核心生物数据资源** (GCBR), 是我国目前**唯一入选**的数据库





# 全面保障我国人类遗传资源信息安全管理与共享应用

承担国家人类遗传资源信息汇交备份任务，遵循“人遗条例”，制定专业化数据访问机制“**申请审核制**”，开发人类遗传资源信息备份管理与发布共享的一体化平台

中华人民共和国科学技术部

国科办函社〔2022〕316号

科技部办公厅关于委托中国科学院北京  
基因组研究所（国家生物信息中心）  
承担国家人类遗传资源信息  
管理备份工作的函

中科院办公厅：

为贯彻落实国家人类遗传资源信息备份流程，确保备份业务与发表共享数据的一致性，推动人类遗传资源信息依法依规开展共享与再利用，经科技部研究，决定委托中国科学院北京基因组研究所（国家生物信息中心）承担国家人类遗传资源信息管理备份任务。

委托的主要任务包括整合现有的人类遗传资源信息备份平台（<https://202.108.211.75/>）和国家生物信息中心人类遗传资源管理



人类遗传资源信息  
管理备份平台



<https://hgrip.cncb.ac.cn>  
<https://ngdc.cncb.ac.cn/hgrip>

促进人遗数据资源安全共享



已生成备份号：3335  
已备份数据集：6388  
汇交数据总量：2.29 PB  
已发布受控数据集：1298  
总申请次数：1285  
获得授权次数：477（境外102）  
下载数据量：574.9 TB





# 提纲



GSA Family介绍



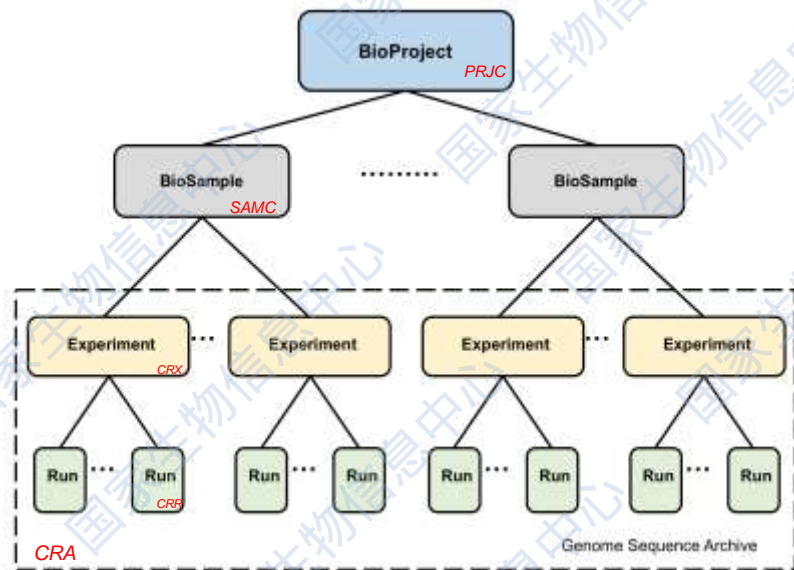
**GSA数据提交及共享**



GSA-Human数据提交及共享



# GSA数据模型及支持的测序平台



数据模型与数据标准：参考INSDC标准

Platform	Data File Type
ILLUMINA	FASTQ, BAM
PACBIO_SMRT	HDF5, BAM, FASTQ
Oxford Nanopore	FASTQ, BAM, Fast5
COMPLETE_GENOMICS	Complete_Genomics_native, FASTQ, BAM
BGISEQ	FASTQ, BAM
BIONANO	BNX
ABI_SOLiD	FASTA/QUAL, FASTQ, BAM
...	...

# 支持的样本类型

## ☐ Pathogen

用于与公共卫生相关的病原体样本

### ☐ Clinical or host-associated

### ☐ Environmental, food or other

## ☐ Microbe

用于能具体写出物种名称的微生物样本，但不包括病毒或病毒样本。宏基因组数据提交选择的Metagenome/Environmental Sample (GSC MIMS unsupported)或Metagenome/Environmental Sample (GSC MIMS compliant)

## ☐ Animal

用于模式生物或动物的多细胞样本或细胞系样本，如大鼠、小鼠、果蝇、牲畜、鱼类、两栖类或其他哺乳动物数据

## ☐ Human

人类遗传资源相关组学数据提交到GSA for Human基因组库。

## ☐ Plant

用于植物样本或植物细胞系样本

## ☐ Virus

用于所有与疾病无关的病毒样本。病毒体应归类为Clinical or host-associated pathogen

### Metagenome/Environmental Sample (GSC MIMS unsupported)

目前用于不满足Metagenome/Environmental Sample (GSC MIMS compliant)的宏基因组生物样本数据

### Metagenome/Environmental Sample (GSC MIMS compliant)

用于宏基因组生物样本数据。下设三个小类分别接收来自人类肠道、土壤和水相关宏基因组数据，即human-gut, soil, water

#### ☐ human-gut

#### ☐ soil

#### ☐ water

新冠病毒相关样本，首选此类型。

\*注意：如果想要受控处理(Controlled-access)，可以提交到GSA-Human

人类型相关样本，请提交到GSA-Human。

不适用于Metagenome/Environmental GSC MIMS类型的宏基因组样本类型，请选此类型

# 数据递交入口: BIG-Sub



<https://ngdc.cncb.ac.cn/gsub>

# 注册并激活账号

Welcome to register for an account of BIGD

Register information:

Account Login Information

Email \*

Password \*

Confirm Password \*

Personal Information

First Name \*

Middle Name \*

Last Name \*

Email Address \*

City \*

State / Province \*

Postal Code \*

Country / Region \*

Institutional Information

Institute / Organization \*

Department \*

Laboratory \*

Title / Position \*

Research Area \*

Check Code \*

Submit

注意：由于需要通过邮件接收激活链接，请确保该邮箱属于自己并能够登录进去点击激活链接！否则账号无法激活和登录！同时请合理设置垃圾邮件判断策略，检查激活邮件是否被邮件系统转移到垃圾邮箱中，或者被邮件系统拒收！

注意：密码必须同时包含大写字母、小写字母、数字、特殊字符，且长度在8~30位之间！

## 激活邮件

### SSO Account Activation

Sender: [bigd-admin <bigd-admin@bigd.ac.cn>](mailto:bigd-admin@bigd.ac.cn)

Time: 2017年10月10日(星期二) 下午2:50

收件人: 您

Dear user,

Please click the URL below to activate your account within 48 hours or you will need to register again.

<http://ss0.bigd.ac.cn/register/active.action?mailAddr=&token=7f6934f45681c87a93757662c8784c22>

Thanks.


BIGD SSO ADMIN

2017-10-10 15:05:03


Note: This email is sent by the system automatically, please do not reply this directly.

点击链接，激活账号

# 登录系统

 CNCB-NGDC

DatabasesTools

 **NGDC** Central Authentication Service

Enter your Username and Password

Email

gsatest1@big.ac.cn

Password

Forgot password?

\*\*\*\*\*

Check code

36bb

36bb

☐ Keep me signed in

LoginResetRegister

Central Authentication Service

[BioProject](#)

[BioSample](#)

[BioCode](#)

[GSA for Human](#)

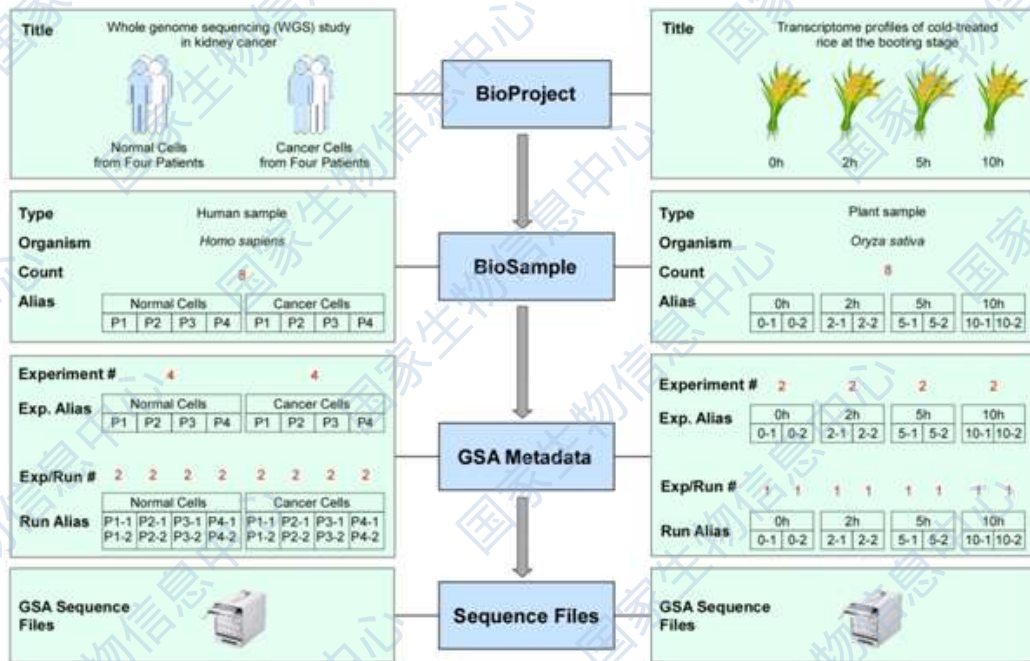
[Genome Sequence Archive \(GSA\)](#)

[Genome WareHouse \(GWH\)](#)

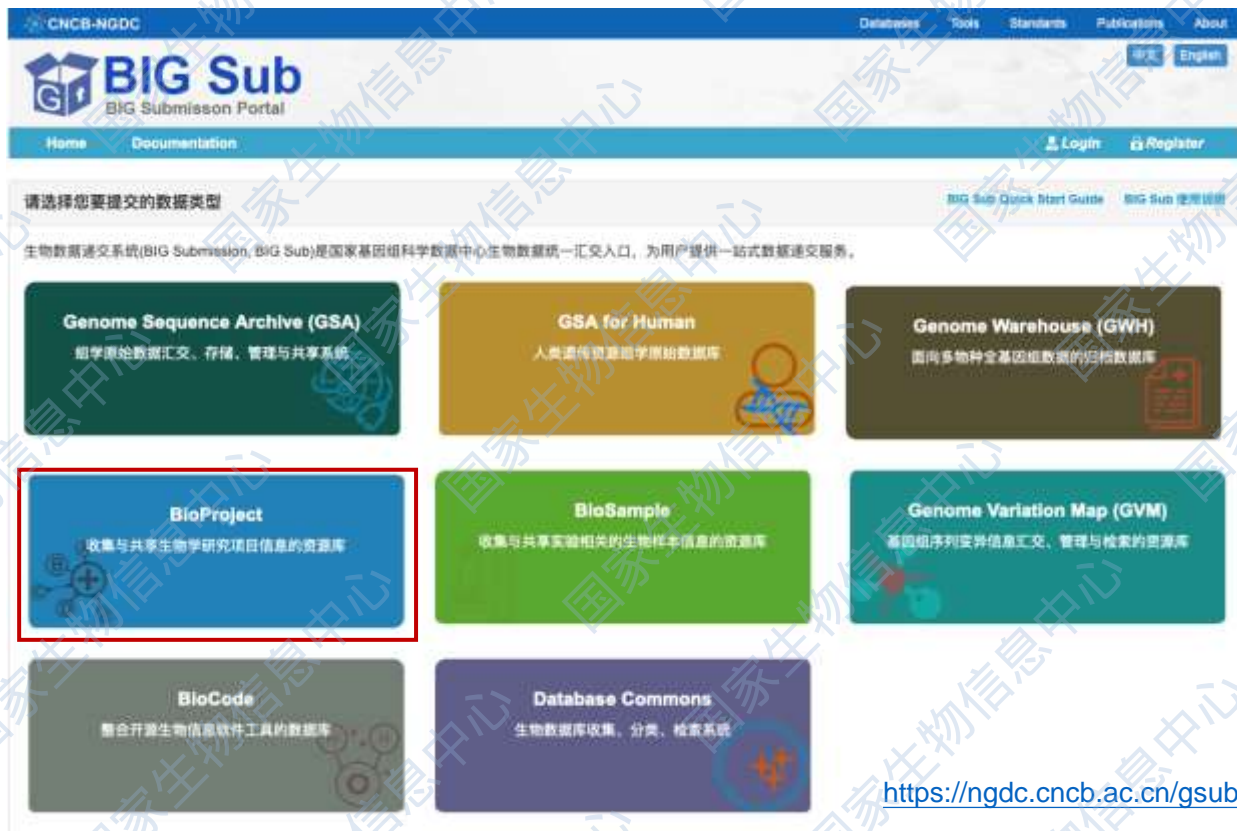
[Genome Variation Map \(GVM\)](#)



# GSA数据递交流程



# 数据递交 - 创建BioProject



CNCB-NGDC

Databases Tools Standards Publications About

**BIG Sub**  
BIG Submission Portal

中文 English

Home Documentation Login Register

请选择您要提交的数据类型

BIG Sub Quick Start Guide BIG Sub 使用指南

生物数据递交系统(BIG Submission, BIG Sub)是国家基因组科学数据中心生物数据统一汇交入口, 为用户提供一站式数据递交服务。

- Genome Sequence Archive (GSA)**  
组学测序数据汇交、存储、管理与共享系统
- GSA for Human**  
人类遗传资源组学原始数据库
- Genome Warehouse (GWH)**  
面向多物种全基因组数据的归档数据库
- BioProject**  
收集与共享生物学研究项目信息的资源库
- BioSample**  
收集与共享实验相关的生物样本信息的资源库
- Genome Variation Map (GVM)**  
基因组序列变异信息汇交、管理与检索的资源库
- BioCode**  
整合开源生物信息软件工具的数据库
- Database Commons**  
生物数据库收集、分类、检索系统

<https://ngdc.cncb.ac.cn/gsub>

# 创建BioProject – 提交列表



中文 English

Home Documentation

Welcome, GSA

BIG Sub / BioProject

**BioProject** 是对研究项目的总体描述。一个BioProject可包含多个生物学样本（BioSample）。

新建BioProject

BioProject Quick Start Guide BioProject使用说明

项目编号	提交编号	项目标题	发布日期	提交状态		操作
PRJCA003418	subPRO004980	GSA_TEST_20200909	2021-09-09	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA002492	subPRO003649	TEST_1_GSATEST1	2020-04-30	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA001010	subPRO001500	project2	2019-12-31	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA001009	subPRO001499	project-test1	2019-12-31	checked OK	Public	

# 创建BioProject – 提交者信息

BiG Sub / BioProject / New BioProject

01 提交者信息 02 基本信息 03 项目类型 04 关联信息 05 审查 & 提交

### 提交者信息

• 名	中间名	• 姓氏
<input type="text" value="GSA"/>	<input type="text" value="middle name"/>	<input type="text" value="BIGD"/>
• 邮箱	备用邮箱	
<input type="text" value="gsatest1@big.ac.cn"/>	<input type="text" value="secondary email"/>	
• 单位	单位网址	• 部门
<input type="text" value="Beijing Institute of Genomics, CN"/>	<input type="text" value="http://www.big.ac.cn"/>	<input type="text" value="BIGD"/>
手机	传真	
<input type="text"/>	<input type="text"/>	
• 街道	• 城市	州/省
<input type="text" value="No.1 Beichen West Road, Chaoyi"/>	<input type="text" value="Beijing"/>	<input type="text"/>
• 邮编	• 国家/地区	
<input type="text" value="100101"/>	<input type="text" value="China"/>	

保存并进入下一项

# 创建BioProject – 基本信息

HiG Sub BioProject New BioProject

01 设置项目信息 02 基本信息 03 项目类型 04 出版信息 05 提交 & 提交

基本信息

发布日期

☐ 审核通过后即可发布 (推荐)

☒ 指定日期发布

2021-10-31

(yyyy-mm-dd)

选择项目

项目

项目标题

涉及领域

项目说明

# 创建BioProject – 基本信息

项目资金来源

机构	项目类别	项目批准号	项目名称
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

[Add another project](#)

项目相关网站链接

网站信息描述	网站地址
<input type="text"/>	<input type="text"/>

[Add another link](#)

关联项目

项目编号	相关描述信息
<input type="text"/>	<input type="text"/>

[Add another project](#)

外部数据库编号

项目编号	数据库名称
<input type="text"/>	<input type="text"/>

[Add another database](#)

[保存并进入下一项](#)



# 创建BioProject – 项目类型

BIG Sub BioProject New BioProject

01 提交表信息 02 样本信息 03 项目类型 04 出版信息 05 概览 & 提交

## 项目类型

### 项目数据类型

- ☐ Whole genome sequencing
- ☐ Clone ends
- ☐ Epigenomics
- ☐ Exome
- ☐ Map
- ☐ Metagenome
- ☐ Phenotype or Genotype
- ☐ Random survey
- ☐ Targeted Locus (Loci)
- ☐ Transcriptome or Gene expression
- ☐ Variation
- ☐ Genome sequencing and assembly
- ☐ Raw sequence reads
- ☐ Genome sequencing
- ☐ Assembly
- ☐ Metagenomic assembly
- ☐ Targeted loci cultured
- ☐ Targeted loci environmental
- ☐ Other

### 样本范围

Monoisolate

### 样本范围选择提示

**Monoisolate:** 单个动物、培养细胞系、近交群体（或可能是从集合样本中产生单个基因型群体）。注意此项适用范围较窄，不为优选项，请谨慎选择。

**Multisolate:** 同一物种的多基因型的个体样本集；

**Multi-species:** 多物种样本集；

**Environment:** 环境样本集，多用于宏基因组研究；

**Synthetic:** 在实验室里制造/合成的样本集；

**Single cell:** 单个细胞测序样本集；

**Other:** 其他样本集/无法归入以上类别的特殊样本集。

保存并进入下一项

# 创建BioProject – 出版信息

BIG Sub / BioProject / New BioProject

01 提交者信息    02 基本信息    03 项目类型    04 出版信息    05 概览 & 提交

出版信息

PubMed ID    OR    DOI

[+ Add another publication](#)

[保存并进入下一项](#)

# 创建BioProject – 概览 & 提交

BIG Sub / BioProject / New BioProject

01  
提交者信息

02  
基本信息

03  
项目类型

04  
出版信息

05  
概览 & 提交

## 概况信息

提交者信息	
提交者	GSA BIGD <a href="mailto:gsatest1@big.ac.cn">gsatest1@big.ac.cn</a>
单位	Beijing Institute of Genomics, Chinese Academy of Sciences
部门	BIGD
国家/地区	China
地址	No.1 Beichen West Road,Chaoyang District Beijing
邮编	100101

项目类型	
项目数据类型	Whole genome sequencing
样本范围	Multisolate
基本信息	
项目标题	TEST_20201015
涉及领域	Medical
项目说明	TEST_20201015
发布日期	2021-10-31

提交

# 创建BioProject – 完成提交

BIG Sub / BioProject

**BioProject** 是对研究项目的总体描述。一个BioProject可包含多个生物学样本 (BioSample) 。

[BioProject Quick Start Guide](#) [BioProject使用说明](#)

新建BioProject

生成的BioProject编号，  
记住这个编号，后边提交GSA信息的时候会用到

项目编号	提交编号	项目标题	发布日期	提交状态		操作
PRJCA003665	subPRO005344	TEST_20201015	2021-10-31	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA003418	subPRO004980	GSA_TEST_20200909	2021-09-09	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA002492	subPRO003649	TEST_1_GSATEST1	2020-04-30	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA001010	subPRO001500	project2	2019-12-31	finished	confidential	<a href="#">修改</a> <a href="#">删除</a>
PRJCA001009	subPRO001499	project-test1	2019-12-31	checked OK	Public	

# 数据递交 – 创建GSA

GSA递交入口



<https://ngdc.cncb.ac.cn/gsub>

# 创建GSA – 提交列表

Home Documentation Welcome, GSA-

BIG Sub / GSA

GSA 是组学原始数据汇交、存储、管理与共享系统。在开始创建GSA数据集前，您需要进入BioProject提交入口完成BioProject（对研究项目的总体描述）的创建。

提示：

- 人类遗传资源相关组学原始数据请提交到GSA for human数据库。
- 同一个项目（BioProject）的数据应归档到同一个GSA数据集下；若BioProject包含多种样本类型（Sample type），如 PRJCA001357，您需要按样本类型创建多个GSA数据集。并将相关类型数据归档到对应的GSA数据集下。
- 如果您在数据上传过程中遇到问题或发现任何系统报错，请通过 [gsa@big.ac.cn](mailto:gsa@big.ac.cn) 邮箱联系我们。

[如何共享和发布GSA数据集](#)

[新建 GSA](#) 点击新建GSA提交

GSA Quick Start Guide GSA使用说明

GSA编号	提交编号	GSA标题	发布日期	提交状态	操作
Unassigned	subCRA002947		2020-04-06	Unfinished at the General Info step Confidential	<a href="#">删除</a>
CRA001076	subCRA001069	GSA-test	2019-12-31	Checked OK Public	



# 创建GSA - 提交者信息

Home Documentation Welcome, GSA

BIG Sub GSA New GSA

01 提交者信息 02 基本信息 03 元数据信息 04 文件 05 审查 & 提交

提交者信息

• 名	GSA	中间名	middle name	• 姓氏	BIGD
• 邮箱	gsatest1@big.ac.cn	备用邮箱	secondary email		
• 单位	Beijing Institute of Genomics, CH	单位网址	http://www.big.ac.cn	• 部门	BIGD
手机		传真			
• 街道	No. 1 Beichen West Road, Chaoyi	• 城市	Beijing	• 州省	
• 邮编	100101	• 国家/地区	China		

保存并进入下一步

保存并进入下一步

# 创建GSA - 基本信息

BIG Sub / GSA / Submission: subCRA010552

01  
提交者信息

02  
基本信息

03  
元数据信息

04  
文件

05  
概览 & 提交

基本信息

发布日期

- ☐ 审核通过后即可发布
- ☐ 指定日期发布

2024-05-26

(yyyy-mm-dd)

选择审核通过后立即发布，或者指定日期发布。  
发布日期必须是从提交起两年之内的日期

## \* 发布策略和免责声明

1. 您可以根据需求设定“发布日期”，在该日期之前，GSA保证数据不公开；
2. 如果引用这些数据与该accession号的文章先于您设定的发布时间而发表，我们将根据文章的发表时间来发布该数据；否则GSA将根据您设定的发布日期而发布该数据；
3. 一旦文章发表，数据可以发布，请把已发表文章的全部信息—作者、题目、期刊、刊号、页数、日期信息发送到该邮箱：[gsa@big.ac.cn](mailto:gsa@big.ac.cn)
4. “发布日期”可以在GSA提交系统内进行修改 ([https://ngdc.cncb.ac.cn/gsub/submit/gsa/\[substitute your GSA submission number\]/finishedOverview](https://ngdc.cncb.ac.cn/gsub/submit/gsa/[substitute your GSA submission number]/finishedOverview))

☐ I accept it. ☐ I don't accept it.

# 创建GSA - 基本信息

### 标题和描述信息

标题

描述信息

### 项目信息

请选择项目编号

OR Go to create [BioProject](#)

选择已有的BioProject编号，  
或者先创建BioProject

### 样本信息

☒ 未创建GSA相关的BioSample信息

☐ 已经创建好GSA相关的BioSample信息

如果您还未创建GSA相关的BioSample，请选择“未创建GSA相关的BioSample信息”，在本次提交流程中创建BioSample(s)

如果您已创建好GSA相关的BioSample，请选择“已经创建GSA相关的BioSample信息”

选择是否已创建BioSample

保存并进入下一项

# 创建GSA - 样本类型

BIQ Sub : GSA : Submission: subCRA004261

01 提交信息 02 样本信息 03 样本类型 04 样本属性 05 元数据信息 06 文件 07 预览 & 提交

样本类型

**Pathogen**  
用于与公共卫生相关的病原体样本

☒ Clinical or host-associated  
Environmental, food or other

**Microbe**  
用于能具体写出物种名称的微生物样本，但不包括致病菌或病毒样本。宏基因组数据建议选择的Metagenome/Environmental Sample (GSC MIMS unsupported)或Metagenome/Environmental Sample (GSC MIMS compliant)

☐ Animal  
用于模式生物或动物的多细胞样本或细胞系样本，如大鼠、小鼠、果蝇、线虫、鱼类、两栖类或其他哺乳动物数据

☐ Plant  
用于植物样本或植物细胞系样本

☐ Virus  
用于所有与疾病无关的病毒样本，病原体应归类为Clinical or host-associated pathogen

**Metagenome/Environmental Sample (GSC MIMS unsupported)**  
目前用于不适于Metagenome/Environmental Sample (GSC MIMS compliant) 的宏基因组生物样本数据

**Metagenome/Environmental Sample (GSC MIMS compliant)**  
用于宏基因组生物样本数据，下设三个小类分别接收来自人类胃肠道、土壤和水相关宏基因组数据，即human-gut, soil, water

☐ human-gut  
☐ soil  
☐ water

保存并进入下一项

## 创建GSA - 样本属性

BIG Sub / GSA / Submission: subCRA004261

01 提交前信息 02 基本信息 03 样本类型 04 样本属性 05 元数据信息 06 文件 07 提交 & 提交

临床或病原微生物样本属性信息

上传BioSample批量提交文件

请选择文件 上传

点击下载Sample模板文件，填写完成后上传

BioSample批量提交模板文件 [Pathogen\\_cl.cn.xlsx](#)，完成填写并检查无误后上传。

BioSample批量提交示例，详见 [e.g Pathogen\\_cl.cn.xlsx](#)。

更多帮助，请查看 [Help](#)。

示例文件。注意：不能编辑、上传此文件

[illegible]

pathogen\_cl.cn.xlsx

# 创建GSA - 样本属性

临床病原微生物样本属性信息

已上传的BioSample批量提交文件

Pathogen\_cl.cn-12.xlsx 33KB

删除 校验

请点击“校验”按钮对批量提交文件进行在线审核。若文件审核通过，请点击“保存并进入下一项”完成BioSample批量提交；若文件审核未通过，请删除已上传文件并修改后再重新上传。

临床病原微生物样本属性信息

已上传的BioSample批量提交文件

Pathogen\_cl.cn-12.xlsx 33KB

删除 校验

下载错误文件: error.txt

Error in Sample sheet  
row 11, column 10: "geographic\_location" ("China: Hubei: Wuhan") Format error. You can use a colon to separate the country or ocean from more detailed information about the location, eg "Canada: Vancouver" or "Germany: halfway down Zugspitze, Alps"

临床病原微生物样本属性信息

已上传的BioSample批量提交文件

Pathogen\_cl.cn-12.xlsx 33KB

删除 校验

Checked OK.

请点击“校验”按钮对批量提交文件进行在线审核。若文件审核通过，请点击“保存并进入下一项”完成BioSample批量提交；若文件审核未通过，请删除已上传文件并修改后再重新上传。

保存并进入下一项

校验错误，删除已上传的表格，按提示修改后重新上传并校验



# 创建GSA - 序列元数据信息

# 元数据表格: Experiment/Run

GSA批量表格由两部分组成Experiment和Run表, 本表用于收集Experiment信息

“绿色标题列”为必填项

“灰色标题列”为选填项, 如果无法提供, 可以为空

“蓝色标题列”为Paired数据必填

“黄色区域”有下拉菜单

当\*Platform (第6列) 为Helicos HeliScopeII时, Planned number of cycles必填

注意: 1. 不要随意删除或插入词条列!

2. 在您使用Excel“自动填充功能”时需谨慎, 以防出

3. 为节省您批量上传表格的审核时间, 请您在提交

1	2	3	4	5	6	7
*ID	*Experiment title	*BioProject accession	*BioSample name	BioSample accession	*Platform	*Library Construction / Experimental Design
E1						
E2						

\*注意: BioSample name需要跟sample表格里  
的sample\_name保持一致

GSA批量表格由两部分组成Experiment和Run表, 本表用于收集Run信息

“绿色标题列”为必填项

“灰色标题列”为选填项, 如果无法提供, 可以为空

“蓝色标题列”为FASTQ格式Paired数据必填项, FASTQ格式只接收.bz2或.gz压缩, 不接收其它格式的压缩文件

“紫色标题列”为BAM格式, 且不是来自PacBio sequel或Ion Torrent系列测序平台需填项: 如果您想把参考序列传到GSA数据库, 请您填写Reference file name和MD5 for reference file (第10和11

“黄色区域”有下拉菜单

注意: 1. 不要随意删除或插入词条列!

2. 在您使用Excel“自动填充功能”时需谨慎,

3. 为节省您批量上传表格的审核时间, 请您

1	2	3	4	5	6	7
*ID	*Run title	*BioProject accession	*Experiment ID	*Run data file type	*File name 1	*MD5 checksum 1
R1						
R2						

GSA\_Template.cn.xlsx

# 创建GSA - 序列元数据信息

The screenshot displays the GSA submission interface. At the top, there is a navigation bar with buttons for steps 01 to 07. Step 05, '元数据信息' (Metadata Information), is currently selected and highlighted in blue. Below the navigation bar, the main content area is titled 'GSA元数据信息'. It shows a list of uploaded files, including 'GSA\_Template.cn\_gztest.xlsx' (31KB), with buttons for '删除' (Delete) and '校验' (Check). A green checkmark and the text 'Checked OK' indicate successful validation. Below the file list, there is a red button labeled '保存并进入下一步' (Save and Enter Next Step).

# 创建GSA - 序列文件上传方式

BiG Sub - GSA Submission: subORA002988

01 提交新数据 02 数据列表 03 提交数据 04 提交数据 05 数据列表 06 文件 07 提交 & 提交

### 文件上传方式

提示:  
上传的文件名和MD5码必须跟您在GSA Metadata表格里面填写的一致。  
上传的文件名必须跟表格一致。  
FASTQ文件必须压缩后上传。目前我们只接受gzip 或 bzip2格式压缩文件。

• 请选择文件上传方式

☒ FTP

使用FTP客户端软件上传文件。FTP地址跟您在BiG Sub的账号一样。  
您可以通过客户端软件FTP上传文件。  
Address: ftp://submit.bi-g.cn  
Username: Submit as you login the BiG Sub  
Password: Same as you login the BiG Sub  
进入FTP后，请进入GSA目录并把文件上传到相应目录下。请不要把文件上传到根目录下。该目录此处填写账号ID和邮箱的文件。  
注意：请使用二进制模式上传文件。如果仍使用FTP客户端软件，请参考软件的说明文档来设置传输模式。如果直接使用FTP命令上传文件，请在输入login命令之前，先输入binary命令，来设置二进制传输模式。

☐ Aspera Command Line

使用Aspera命令上传文件。

建议使用客户端软件上传文件  
注意：请使用二进制模式上传

保存并进入下一项

文件名、MD5码需要跟  
metadata表格里填写的一致

文件名	大小 (字节)	MD5
v1_1.fastq.gz	200,000,000	999,800,000
v1_2.fastq.gz	999,800,000	999,800,000

[illegible]

秘钥文件地址，点击下载秘钥文件  
并将命令行中的[path/to/keyfile]替换为下载的本地路径

注意：本页面不是提交的最后一步。  
请选择文件上传方式后，点击进入下一页，浏览并保存提交

# 创建GSA - 概览 & 提交

创建GSA - 概览 & 提交

提交编号: subGSA04261

标题: TEST\_20211015

创建日期: TEST\_20211015

发布日期: 2021-10-31

提交者: GSA BIGD

单位: Institute of Genomics, Chinese Academy of Science

部门: BIGD

国家/地区: China

地址: No.1 Beichen West Road, Chaoyang District, Beijing

邮编: 100101

样本类型: Clinical or host-associated pathogen

核酸类型: Pathogen, s1-s12, s13

二进制文件: GSA\_Template.on\_gsatoolkit

实验名称	测序平台	样本名称	生物名称	实验选择
experiment1	Illumina HiSeq 2000	Sample1	Severe/acute respiratory syndrome coronavirus 2	unpairedRead
Run名称		Run序列文件处理结果		文件类型
Run1	File: s1_1.fastq.gz File: s1_2.fastq.gz			fastq

Page 1 of 1

提交

检查无误后，点击“提交”，完成本次提交



# 创建GSA – 元数据提交完成

GSA提交编号, 注意不能把这个号写到文章里

点击查看数据处理情况

GSA编号	提交编号	GSA标题	创建日期	发布日期	提交状态	操作
Unassigned	subCRA004261	TEST_20201015	2020-10-15	2021-10-31	Checking detail Confidential	删除
Unassigned	subCRA004036	TEST_20200915	2020-09-15	2020-09-30	Checking detail Confidential	删除
Unassigned	subCRA003988	GSA_TEST_20200910	2020-09-10	2022-11-08	⚠️ Unfinished at the OverView step Confidential	删除
Unassigned	subCRA002947		2020-04-06	2020-04-06	⚠️ Unfinished at the General Info step Confidential	删除
CRA001076	subCRA001069	GSA-test	2018-08-29	2025-02-24	Checked OK Confidential	立即发布 分享

一个审核成功的提交, 分配的GSA编号, 文章里使用此编号

# 创建GSA – 数据审核状态

BIG Sub / GSA / Submission: subCRA004261

### GSA基本信息

GSA 提交信息: subCRA004261 / Release Date : 2021-10-31 / Project : [PRJCA003665](#)

状态 1 Runs are checked failed [detail](#)

标题 TEST\_20201015

发布日期 2021-10-31

更新

**File not found:** 如果元信息刚审核通过，需要耐心等待后台关联审核数据；如果元信息审核通过很久，需自行检查数据是否上传到GSA目录下，文件名是否一致。如果不一致，可点run编号后的edit修改一致。Filezilla上传数据的用户也可以自行修改数据文件名

**md5 is inconsistent:** 自行核对本地文件的md5与元信息填写中是否一致，如果填写有误，点run编号后面的edit修改。如果确认填写无误，则是文件上传过程有问题导致，选择二进制重新上传文件

**processed error:** 数据审核报错，务必联系管理员反馈报错信息

**processed succeed:** 数据审核没有问题，全部成功后等归档就可以收到编号

# 数据发布

Public

Data are fully available by anyone

Confidential

Data are not available publicly through any means, data owners should provide the date release date

数据发布条件:

- 到达用户设定的发布日期
- 文章见刊

• 发布日期

2018-05-24

(yyyy-mm-dd)

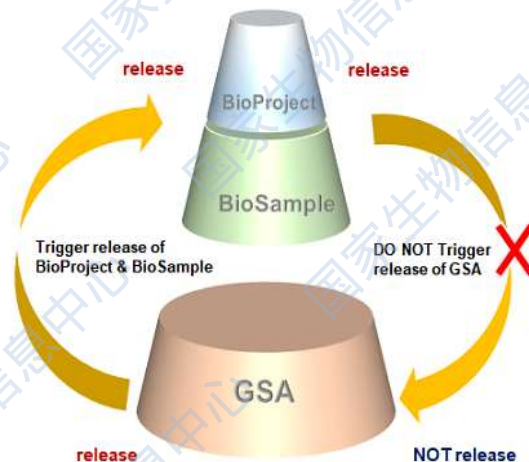
• 发布策略和免责声明

1. 你提交的数据可以保留在一定时间内不发布;
2. 发布时可以在GSA提交系统内进行修改: ([http://bigd.big.ac.cn/gsub/submit/gsa/substitute your GSA accession number/contents](http://bigd.big.ac.cn/gsub/submit/gsa/substitute%20your%20GSA%20accession%20number/contents))
3. 如果引用这些数据或该accession号的文章先于您设定的发布时间而发表,我们将根据文章的发表时间来发布该数据;否则GSA将根据您设定的发布日期发布该数据;
4. 一旦文章发表,数据可以发布,请把已发表文章的全部信息--作者,题目,期刊,刊号,页数,日期信息发送到该邮箱: [GSA@big.ac.cn](mailto:GSA@big.ac.cn)

☐ I accept it. ☒ I don't accept it.

新建 GSA

GSA编号	提交编号	GSA标题	发布日期	提交状态	操作
CRA000532	subCRA000595	Human dataset test	2020-12-31	Checked OK Confidential	<div>立即发布</div> <div>分享</div>



# 设置共享链接

Create GSA		Batch Submission		GSA Quick Start Guide (US) GSA Quick Start Guide (CN)		
Accession	Submission ID	Title	Release date	Status	Operation	
CRA000620	subCRA000620	GSA-test	2019-12-31	Checked OK Confidential	ReleaseNow Share	



Create GSA		Batch Submission		GSA Quick Start Guide (US) GSA Quick Start Guide (CN)		
Accession	Submission ID	Title	Release date	Status	Operation	
CRA000620	subCRA000620	GSA-test	2019-12-31	Checked OK Confidential	ReleaseNow Shared URL: http://bigd.bj.ac.cn/gsa/s/VmVW41CA Cancel Share	

## Review Link Opinion

Accession: CRA000620

Set review link expire time

1 month

勾选，生成数据  
下载链接

☒ Create data access link

NOTE: Reviewers can access the sequence files through this link, please tick the checkbox to generate this link as required.

Cancel

Submit

注：此链接为临时链接，可以将该链接分享给编辑和审稿人，方便其查看数据，但为了您的数据安全请不要将此链接对外公布。数据共享结束后，请点击“Cancel share”按钮，取消数据共享。

Home	Preview
CRA001076 基本信息	
CRA 提交信息: CRA001076 / GSA-test / release time: 2025-02-24	
数据下载	
文件	HTTPS: <a href="https://share.cnbi.ac.cn/7HeHqPAJ">https://share.cnbi.ac.cn/7HeHqPAJ</a>

# 数据引用

## 如何引用GSA?

当您成功提交数据到GSA并通过审核后，请在您要发表的论文中添加如下语句：

*The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRAxxxxxx) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa>.*

请按照以下格式引用我们的文章：

● **The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types.** *Genomics, Proteomics & Bioinformatics* 2021, 19(4):578-583. <https://doi.org/10.1016/j.gpb.2021.08.001> [PMID=34400360] 

[Endnote文件下载](#)

● **Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024.** *Nucleic Acids Res* 2024, Jan 5;52(D1):D18-D32. <https://doi.org/10.1093/nar/gkad1078> [PMID=38018256]  [Endnote文件下载](#)





# 数据访问

国家生物信息中心

Data Resources Computing Analysis Data Network Standards

**GSA**  
Genome Sequence Archive

GSA CRA001962 搜索 高级检索

Home 数据提交 数据浏览 信息检索 数据统计 帮助和支持

Center

- NGDC (40)

Downloadable

- Downloadable (40)

File type

- fastq (40)

Platform

- ILLUMINA (40)

Source

- TRANSCRIPTOMIC (40)

Strategy

- RNASeq (40)

Selection

- PCR (40)

Clear all

Search result:

Total items: 40 | Items 1 of 40 | Page: 1 / 3 | First Prev Last | Jump To: 1 | Go

<input type="checkbox"/>	Title: Spleen-2-5-IP Accession: CRX000013 Platform name: ILLUMINA Strategy name: RNASeq	Source name: TRANSCRIPTOMIC Selection name: PCR
<input type="checkbox"/>	Title: Heart-1-5-IP Accession: CRX000004 Platform name: ILLUMINA Strategy name: RNASeq	Source name: TRANSCRIPTOMIC Selection name: PCR
<input type="checkbox"/>	Title: Lung-2-5-Input Accession: CRX000013 Platform name: ILLUMINA Strategy name: RNASeq	Source name: TRANSCRIPTOMIC Selection name: PCR
<input type="checkbox"/>	Title: Lung-1-4-IP Accession: CRX000011 Platform name: ILLUMINA Strategy name: RNASeq	Source name: TRANSCRIPTOMIC Selection name: PCR
<input type="checkbox"/>	Title: Liver-1-5-Input Accession: CRX000009 Platform name: ILLUMINA Strategy name: RNASeq	Source name: TRANSCRIPTOMIC Selection name: PCR

Send to : ▼

Send to : ▼

## Choose Destination

☒ File

Download 40 items.

Format

RunInfo ▼

Create Files





Center

- INSDC (5715855)
- NGDC (88449)

Downloadable

- Non-downloadable (4282444)
- Downloadable (1521725)

File type

- ab1 (20)
- sff (299)
- sra (5715855)
- oxfordnanopore\_native (5)
- pacbio\_sequel\_native (5)
- fastq (85969)
- fasta (334)
- bam (1814)

Platform

- unspecified (36681)
- ILLUMINA (5270548)

Search result:

Total items: 5804304 | Items 1 of 20 | Page: 1 / 290215 | First Next Last | Jump To 1 Go

1.	Title:	Illumina HiSeq 1000 paired end sequencing of SAMD00082794		
	Accession:	DRX088646	Source name:	METAGENOMIC
	Platform name:	ILLUMINA	Selection name:	RANDOM
	Strategy name:	Genome		
2.	Title:	Illumina HiSeq 1000 paired end sequencing of SAMD00082790		
	Accession:	DRX088642	Source name:	METAGENOMIC
	Platform name:	ILLUMINA	Selection name:	RANDOM
	Strategy name:	Genome		
3.	Title:	Illumina HiSeq 1000 paired end sequencing of SAMD00082792		
	Accession:	DRX088644	Source name:	METAGENOMIC
	Platform name:	ILLUMINA	Selection name:	RANDOM
	Strategy name:	Genome		
4.	Title:	Illumina HiSeq 1000 paired end sequencing of SAMD00082791		
	Accession:	DRX088643	Source name:	METAGENOMIC
	Platform name:	ILLUMINA	Selection name:	RANDOM
	Strategy name:	Genome		

Send to : ▼

Send to : ▼

Choose Destination

☐ File

Download 5804385 items.

Format

☒ Accession List

☐ RunInfo

☐ DataSet Accession List

# 数据访问

Home / GSA / CRA003018

### CRA003018 基本信息

标题: ZEA affects porcine ovarian granulosa cells      项目编号: PRJCA003148 /      发布日期: 2020-09-08  
文件个数: 27      文件大小: 69.51 GB

### 数据下载

元数据信息: [CRA003018.xlsx](#)

文件: [HTTPS: https://download.cncb.ac.cn/gsa/CRA003018](https://download.cncb.ac.cn/gsa/CRA003018) [推荐](#) [EdgeTurbo \(测试\)](#) : [文件下载](#) [Aspera命令行](#) [帮助](#)  
FTP: <ftp://download.big.ac.cn/gsa/CRA003018>

提示: HTTP下载速度有限, 推荐使用Edge Turbo或FTP客户端 (比如 FileZilla Client) 下载数据。  
EdgeTurbo支持Linux命令行, Windows/Mac平台的Chrome, Edge和Firefox浏览器。

### Experiments和Runs信息

实验编号	实验名称	生物名称	测序平台	样品编号
CRX130439	S3_3_miRNA	Sus scrofa	Illumina NovaSeq 6000	SAMC206797
Run编号	Run别名		Run序列文件信息	
CRR156624	S3_3_miRNA		File: CRR156624.fq.gz	
CRX130438	S3_2_miRNA	Sus scrofa	Illumina NovaSeq 6000	SAMC206796
Run编号	Run别名		Run序列文件信息	
CRR156623	S3_2_miRNA		File: CRR156623.fq.gz	

点击下载元数据

多种文件下载方式



# 提纲



GSA Family介绍



GSA数据提交及共享

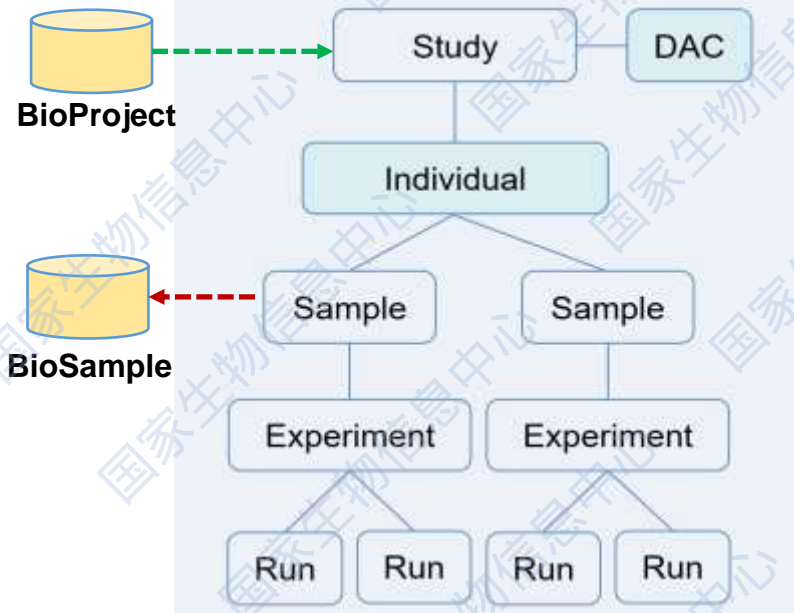


**GSA-Human数据提交及共享**



# 数据模型和支持的数据类型

数据组织模型



以individual为核心，组织元信息与数据文件

研究类型		适用元数据模板
疾病研究 (Disease study)	肿瘤研究 (Cancer)	Tumor_1.0_us.xlsx
	自身免疫研究 (Autoimmune diseases)	Autoimmune_1.0_us.xlsx
	其他疾病研究 (Other diseases)	General_1.0.xlsx
队列研究 (Cohort study)		General_1.0.xlsx
细胞系相关研究 (Cell line related study)		Cellline_1.0.xlsx
临床病原体研究 (Clinical pathogen)		Clinical_pathogen_1.0.xlsx
人体相关宏基因组研究 (Human associated metagenome)		Human_associated_metagenome_1.0.xlsx

测序平台	文件格式
ILLUMINA	FASTQ, BAM
PACBIO_SMRT	HDF5, BAM, FASTQ
Oxford Nanopore	FASTQ, BAM, Fast5
COMPLETE_GENOMICS	Complete_Genomics_native, FASTQ, BAM
BGISEQ	FASTQ, BAM
BIONANO	BNX
ABI_SOLiD	FASTA/QUAL, FASTQ, BAM
...	...

更多数据格式信息: [https://ngdc.cnbc.ac.cn/gsa/support/standardsGsa#file\\_1](https://ngdc.cnbc.ac.cn/gsa/support/standardsGsa#file_1)

# 数据提交须知

- 数据提交必须使用PI账号 – 数据责任人

- 科研机构独立研究组的课题组长 (PI) 或具有高级职称的人员
- 高校教授/副教授
- 医院主任/副主任医师
- 企业部门负责人

- 元数据信息必须脱敏，即不能包含受试者隐私信息

- 提交时设置数据访问方式 (用户自行设定)

- 受控访问 (默认)
- 公开访问

- 受控访问数据，须提供DAC (数据管理委员会) 及成员信息，并指定DAC联系人

- 对外共享数据集前，应根据《中华人民共和国人类遗传资源管理条例》的规定，获得相关数据集在“科学技术部政务服务平台”中“开放使用”类别的备案号

[https://ngdc.cncb.ac.cn/gsa-human/document/Principle\\_of\\_Accessing\\_Human\\_Genetic\\_Resource\\_Data\\_in\\_NGDC\\_V1.pdf](https://ngdc.cncb.ac.cn/gsa-human/document/Principle_of_Accessing_Human_Genetic_Resource_Data_in_NGDC_V1.pdf)

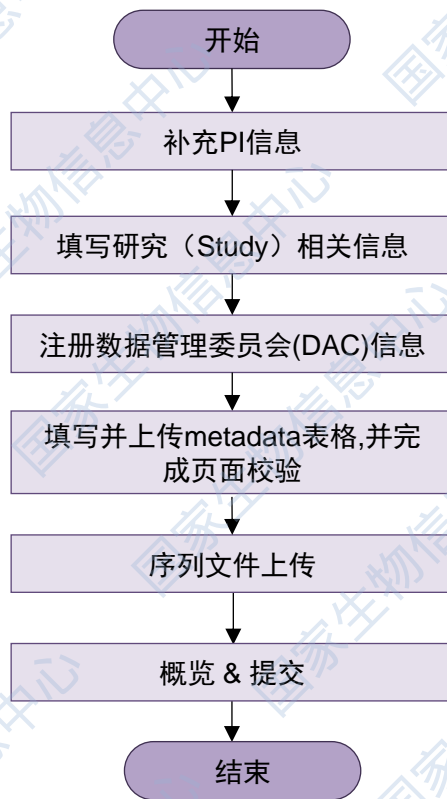
# 受控访问数据的“申请-审核”模式



遗传, 2021



# 数据提交流程



# 个人信息补充



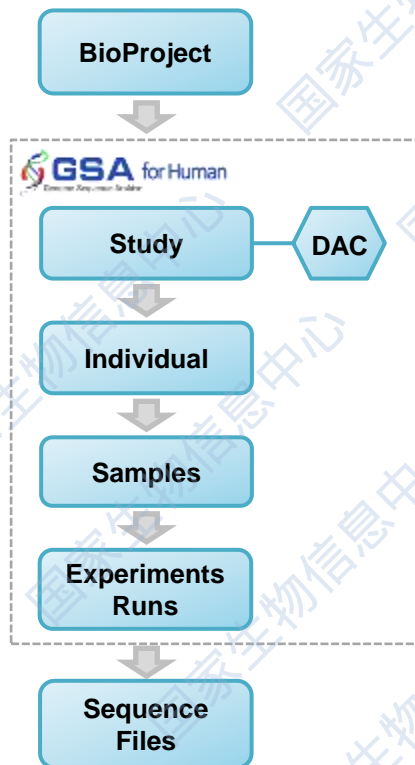
<https://ngdc.cncb.ac.cn/gsub/>

登录后，需要完善个人信息：

补充职称、个人主页/通讯文章链接、单位  
主页信息

个人主页上的邮箱信息最好与账号邮箱一致  
不符合条件的账号进行数据提交，将会审批  
不通过

# 新建GSA-Human提交



**GSA-Human数据提交**

GSA-Human是一个专注于生物医学研究相关的人类遗传资源信息档案。在开始创建GSA-human数据前，您需要通过BioProject提交入口完成BioProject（对研究项目的总体描述）的创建。

- GSA-Human的数据提交者需满足下列条件之一：
  - 从事基础或应用研究的课题组长（PI）或具有高级职称的人员
  - 高等院校副教授
  - 医院主任/副主任医师
  - 企业部门负责人

如果您不属于以上人员，请邀请或聘用满足条件的人员负责注册账号，并进行数据提交。

- 关于个人主页/通讯作者文章链接信息，建议提供单位网页，姓名等信息的企业个人主页，或链接一至五年内作为通讯作者的文章链接，或PubMed文章链接。如单位需要被引用的个人主页请提供作者文章链接信息，请点击此链接修改。

**提交文档**

人类遗传资源科学数据信息（HGRIS）使用指南，包括元数据信息提交、元数据修改、数据文件上传数据提交等过程的详细操作描述。 [CHN](#)

人类遗传资源信息数据流程。 [CHN](#)

国家基因组科学数据中心人类遗传资源数据共享指南。 [LNN](#)

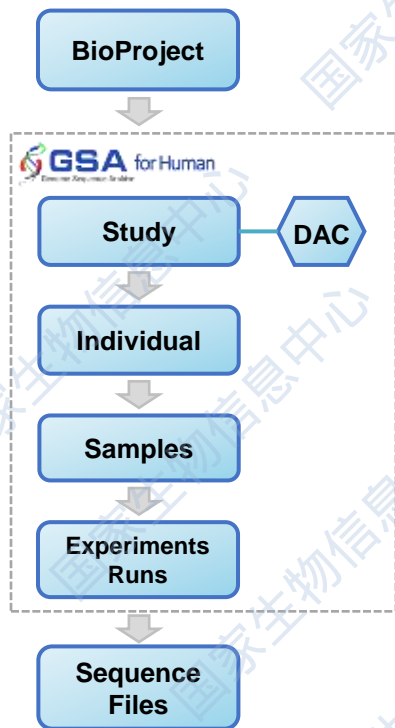
如果您在使用系统过程中遇到任何问题，请通过 [gsa@big.ac.cn](mailto:gsa@big.ac.cn) 联系我们。

如果您需要上传超过30TB的数据，请与我们联系，邮件 [gsa@big.ac.cn](mailto:gsa@big.ac.cn)。

**新建GSA-human**

编号	提交状态	访问类型 (申请次数)	操作

# GSA-Human提交：数据提交者信息

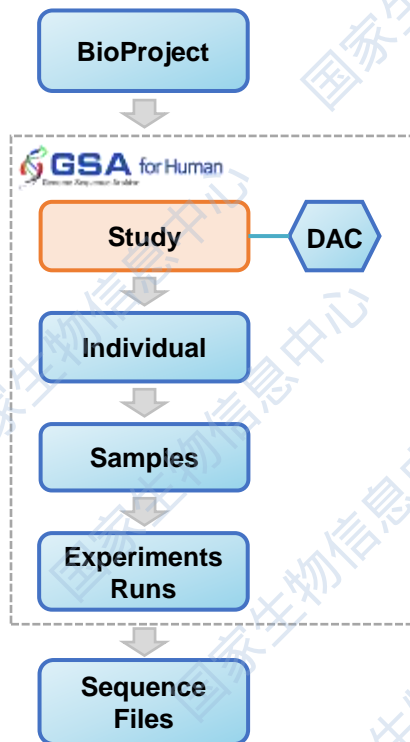


The screenshot shows the 'New Submission' form on the GSA-Human website. The form is divided into several sections: 'Subscriber Information', 'Administrative Contact Information', and 'Privacy Policy for the Release of Human Genetic Resources Data in HRCDB'. The 'Subscriber Information' section includes a checkbox for 'I want to provide new contact information'. The 'Administrative Contact Information' section contains fields for 'First Name', 'Middle Name', 'Last Name', 'Email', 'Institution/Organization', 'Department', 'City', 'State/Province', 'Country/Region', 'Postal Code', and 'Phone'. The 'Privacy Policy' section contains a checkbox for 'I have read and accepted the terms above'. A red box highlights the 'Administrative Contact Information' section, and another red box highlights the 'Privacy Policy' section. A red arrow points from the 'I want to provide new contact information' checkbox to the 'Administrative Contact Information' section.

请注意，建议添加提交联系人信息。如果不添加，数据集问题的反馈默认仅通知注册账号，而往往数据提交和问题处理不一定由PI本人负责。添加后两个人都会收到通知信息，多一个人有备而无患

进入下一步前，务必请先阅读并接受我们的数据管理规则

# GSA-Human提交：研究信息



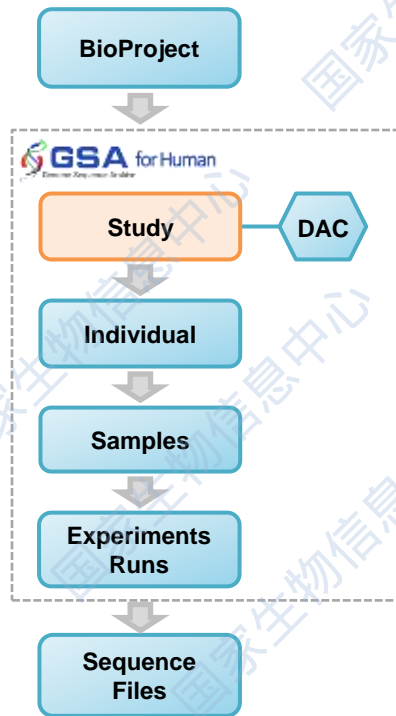
The screenshot shows the GSA-Human submission form. The form is divided into several sections, each highlighted with a red box and an arrow pointing to a text box on the right. The sections are: 'Research data', 'Study information', 'Biological information', and 'Data availability'. The 'Research data' section includes a 'Release date' field. The 'Study information' section includes a 'Study title' field, a 'Description' field, and a 'Study type' field. The 'Biological information' section includes a 'Please select biological system' field. The 'Data availability' section includes a 'Data availability' field.

**发布时间：**可选**两年之内**的任意时间作为发布时间；  
只要数据未公开，可随时修改该时间

**基本信息：**按照研究情况，选择研究类型及其相关信息

**BioProject信息：**填写已有BioProject编号，或者先  
跳转至BioProject新建

# GSA-Human提交：研究信息



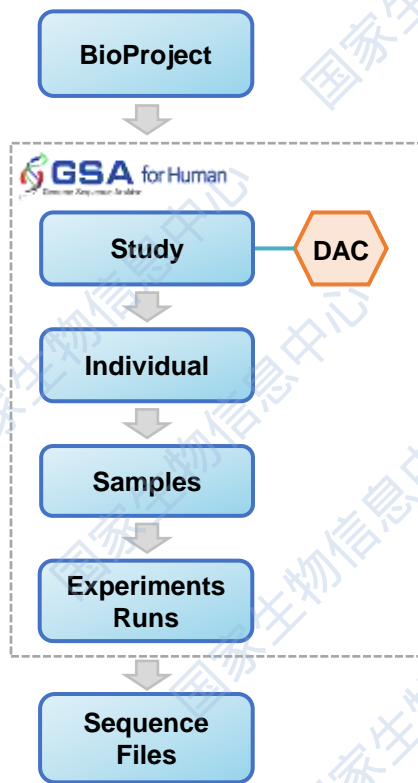
The screenshot shows the 'GSA for Human' submission form. The 'Study' section is highlighted with a red box. The 'Data Accessibility' section is also highlighted with a red box. The 'Data release policy' section is highlighted with a red box. The 'Data sharing limitation' section is highlighted with a red box. The 'Whether or not the confirmation of the data user's institution is required when requesting for this data?' section is highlighted with a red box. The 'I have read and accepted the DATA ACCESS AGREEMENT' section is highlighted with a red box.

## 数据访问方式

- 选择数据访问方式（受控访问或开放访问）与使用限制
- 如果选择受控访问，请确定未来使用者申请数据时，是否需要在双方签订的数据访问协议（DAA）上加盖单位公章



# GSA-Human提交：数据管理委员会（DAC）信息



BIG Data > GSA-Human | Submission: subHRA000497

Submission ID: subHRA000497

Submitter Study Metadata Data Access Committee Research Files Overview & Submit

Data Access Committee (DAC) is a body of one or more named individuals who are responsible for data release to external requestors based on current applicable National Research Ethics laws.  
A DAC is typically formed from the same or multiple organisations that collected the samples and generated any associated analyses.  
2 Multiple datasets may be affiliated to a single DAC.

DAC Information

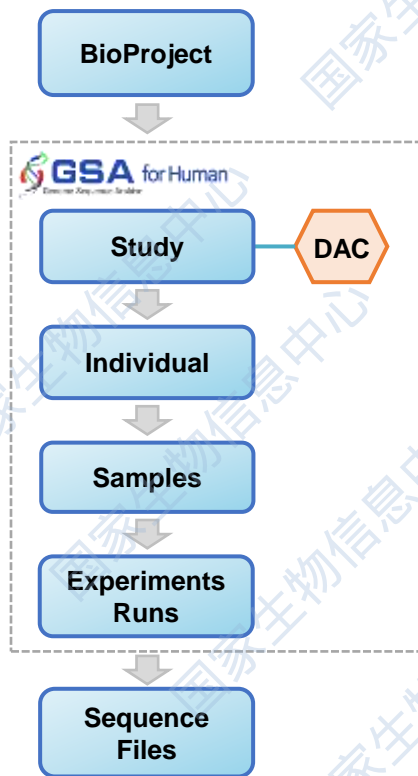
\* Did you already register a DAC for this study?  
☐ No ☐ Yes

Data Access Committee

DAC accession	DAC name	DAC Members	Email	Organization	Department	Is Contact Person	Contact Phone
HDAC000186				Beijing Institute of Genomics (BIG)	BIG Data Center	Yes	

如果数据访问方式使用“受控访问 (Controlled-access)”，需建立数据管理委员会 (DAC) 并指定DAC联系人，负责数据发布后的申请审批与授权

# GSA-Human提交：数据管理委员会（DAC）信息



The screenshot shows the 'DAC Information' section of the GSA-Human submission form. The 'Did you already register a DAC for this study?' question is highlighted with a red box, and the 'No' radio button is selected. A red arrow points from this selection to the 'Add DAC members' section. The 'Add DAC members' section contains a table for existing DACs and a form for adding new members.

**DAC Information**

Did you already register a DAC for this study?  
☒ No ☐ Yes

**Data Access Committee**

DAC accession	DAC name	DAC Members
HDAC000186		

**Add DAC members**

Did you already register a DAC for this study?  
☒ No ☐ Yes

DAC name:

Description:

**DAC Members**

Email	First Name	Last Name	Organization	Department	Contact
gustad@ngdc.ac.cn	Gust	ADG	Beijing Institute of	ADG	

**Add DAC members**

☒ Contact Person (Selected)

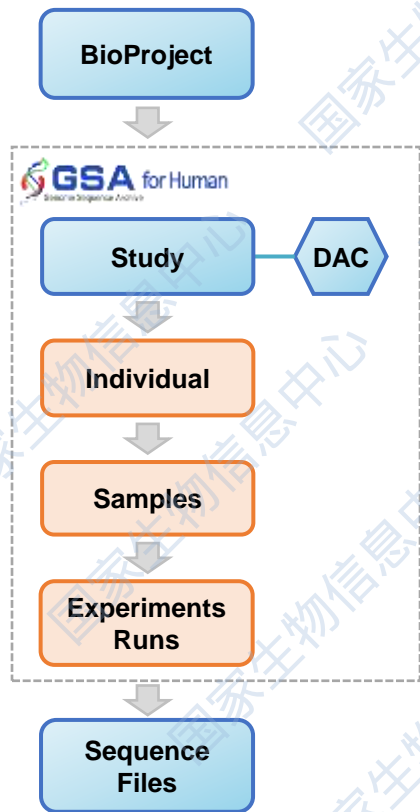
Please follow the guidelines to complete the required information.  
ONLY ONE member can be selected as the contact person to receive data access requests.  
Please choose carefully and complete the contact phone number.

First Name:  Email:  Phone Number:

**Save and forward**

选择“No”，创建新DAC

# GSA-Human提交：元数据信息



GSA-Human Submission: subHRA000650

Submission ID: subHRA000650

Submitter Study Information Data Accession Committee Metadata Files Overview & Submit

General Package

Upload the metadata file using Excel format.

Please selected file

**Download Excel Template General 1.0 spreadsheet** edit, save and then upload the modified template.

If you have any question or would like to give us any suggestion/comment or report a bug, please feel free to contact us at [gaa@big.ac.cn](mailto:gaa@big.ac.cn)

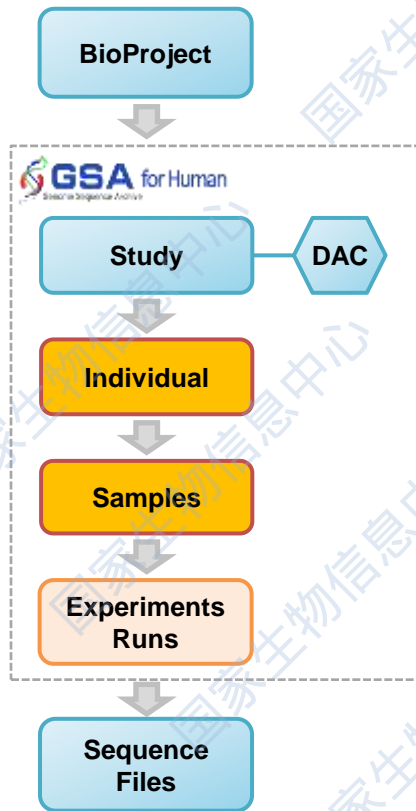
## 研究信息中的研究类型 (Study Type)

研究类型	适用元数据模板
疾病研究 (Disease study)	General_1.0.xlsx
队列研究 (Cohort study)	General_1.0.xlsx
细胞系相关研究 (Cell line related study)	Cellline_1.0.xlsx
临床病原体研究 (Clinical pathogen)	Clinical_pathogen_1.0.xlsx
人体相关宏基因组研究 (Human associated metagenome)	Human_associated_metagenome_1.0.xlsx

元数据的模板文件根据研究类型 (Study Type) 进行**自动匹配**, 以适应不同的研究方向需求

# GSA-Human提交：元数据信息

以General\_1.0.xlsx为例：主要用于疾病研究和队列研究



This sheet is used for individual information.  
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.  
GREY fields are optional. Leave optional fields empty if no information is available.  
ORANGE fields require that the individual accession number, like HR000001, already exists in GSA for Human database.  
YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

**CAUTION:** DO NOT delete any columns! But you can insert the custom attributes at the end of all column as needed!  
Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before sending to us.

1	2	3	4	5	6	7	8	9	10	11
ID	individual name	gender	existing individual accession	experiment group	disease name	disease stage	treatment	survival time	is smoker	is drinker
D1										

12	13	14	15	16	17	18	19	20
height (m)	body weight (kg)	body mass index(BMI)	ethnicity	province	occupation	race	birthplace	custom attributes 1

自定义字段

Individual 表格

1:N

常见问题1：对应关系

单个个体可以多对应多个样本，请务必注意每个样本所属关系。这是最容易出现问题的地方。因为个体为核心，一旦出错，所有元信息都会受到影响。

This sheet is used for Sample information.  
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.  
GREY fields are optional. Leave optional fields empty if no information is available.  
YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

**CAUTION:** DO NOT delete any columns! But you can insert the custom attributes at the end of all column as needed!  
Be aware that Excel may automatically apply formatting to your data.  
In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate before sending to us.

1	2	3	4	5	6	7	8	9	10	11
ID	sample name	tissue	individual ID	age	age unit	public description	sample title	collection date	biomaterial provider	culture collection
S1										

12	13	14	15
karyotype	phenotype	population	custom attributes 1

自定义字段

Sample 表格

常见问题2：敏感信息问题

individual和sample的名称中不可包括受试个体名字拼音全称或简称

提示：请确保信息的完整性与准确性。完整的信息可提供全面的背景和上下文，对于管理员审核、审稿人审阅及未来数据使用有重要帮助。如果已有列无法详细描述，可自定义添加相关信息。



# GSA-Human提交：元数据信息

以General\_1.0.xlsx为例：主要用于 **疾病研究** 和 **队列研究**

DAC

This sheet is used for Experiment information.

GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.

GREY fields are optional. Leave optional fields empty if no information is available.

BLUE fields require for paired-end data only

YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

When the Platform (column 4) is Helicos HelScope, the Planned number of cycles (column 16) is required.

CAUTION: DO NOT delete or insert column!

Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate.

1	2	3	4	5	6	7
ID	Experiment title	sample_ID	Platform	Library Construction / Experimental Design	Library name	Strategy
E1						

11	12	13	14	15	16
Read length for read 1 (bp)	Read length for mate 2 (bp)	Insert size (bp)	Nominal size (bp)	Nominal standard deviation (bp)	Planned number of cycles

**常见问题1：对应关系**

如果单个样本采用多种测序策略，则样本对应多个Experiment；Exp  
应关系也为一对多。

This sheet is used for Run information and needs to be created after Experiment sheet.

GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed.

BLUE fields require for paired-end data only

PURPLE When your Run data file type (column 4) select bam. If you want to submit your reference file to our FTP Site, you need to fill in the Reference file name and MD5 file

YELLOW columns have drop-down menus that allow you to select from a controlled vocabulary. Once specified for one row, these values can be copied-and-pasted down.

CAUTION: DO NOT delete or insert column!

Be aware that Excel may automatically apply formatting to your data. In particular, take care with dates, incrementing autofills and special characters like / or -. Doublecheck that your text file is accurate.

1	2	3	4	5	6	7
ID	Run title	Experiment ID	Run data file type	File name 1	MD5 checksum 1	File name 2
R1						

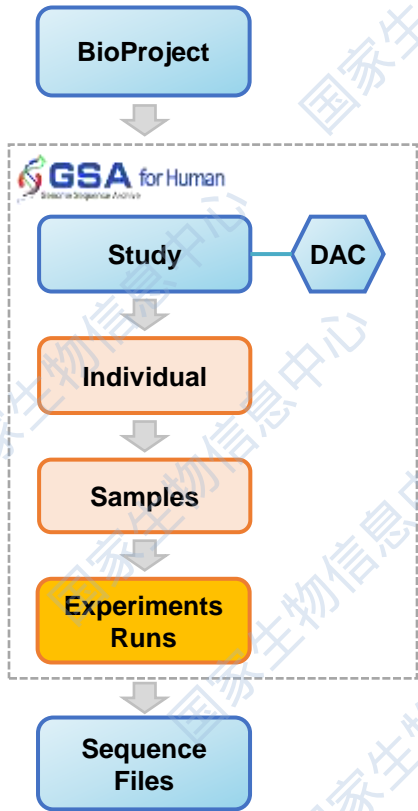
9	10	11	12
Reference file name	MD5 for reference file	Assembly Name or Accession	Assembly Accession URL

**常见问题2：敏感信息问题**

Experiment和Run title不可包括患者名字拼音  
全称或简称；文件的名称也不可包含此类信息

**常见问题3：文件名称一致问题**

文件名称必须与上传文件完全一致  
质控时将系统无法定位，从而影响



**以General\_1.0.xlsx为例：主要用于疾病研究和队列研究**

[illegible]

## Experiment 表格

### 常见问题1：对应关系

如果单个样本采用多种测序策略，则样本对应多个Experiment；Experiment与Run之间对应关系也为一对多。

1:N ↓

[illegible]

## Run 表格

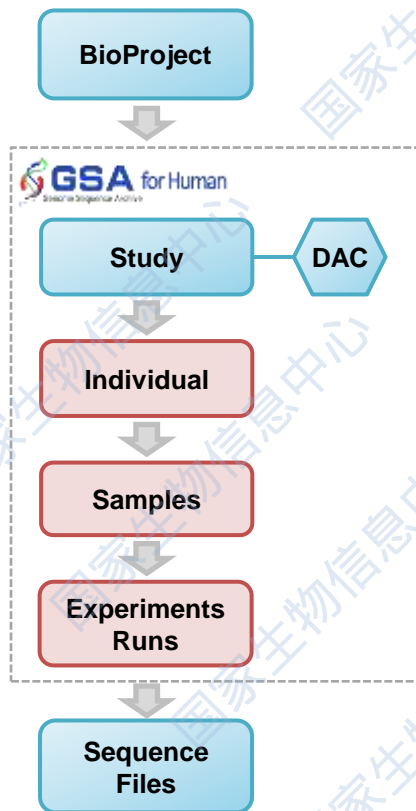
## 常见问题2：敏感信息问题

Experiment和Run title不可包括患者名字拼音  
全称或简称；文件的名称也不可包含此类信息

### 常见问题3：文件名称一致问题

文件名称必须与上传文件**完全一致**。如果不一致，后期  
质控时将系统无法定位，从而**影响文件处理与归档效率**

# GSA-Human提交：元数据信息



BIG Sub : GSA-Human : Submission: subHRA000660

Submission ID: subHRA000660

Submitter Study Information Data Accession Overview Metadata Files Overview & Submit

General Package

Upload the metadata file using Excel format

Please select file **Upload** 上传填写完成的表格

Download Excel Template General 1.0 spreadsheet, edit, save and then upload the modified template.

If you have any question or would like to give us any suggestion/comment or report a bug, please feel free to contact us at [gsa@big.ac.cn](mailto:gsa@big.ac.cn)

Upload the metadata file using Excel format

General\_1.0\_us-GSATEST.xlsx 10KB **Check** 点击进行“格式”校验

Download Excel Template General 1.0 spreadsheet, edit, save and then upload the modified template. For more help, please see the [Example File](#).

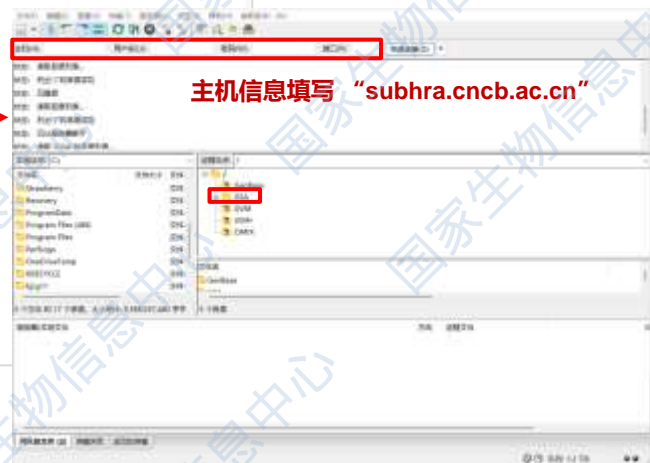
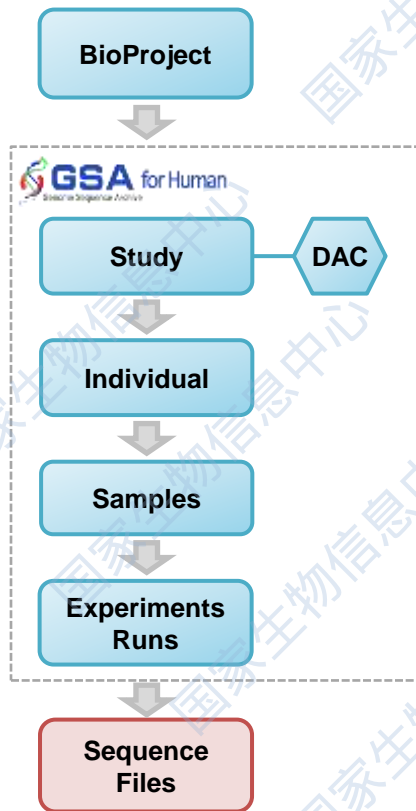
If you have any question or would like to give us any suggestion/comment or report a bug, please feel free to contact us at [gsa@big.ac.cn](mailto:gsa@big.ac.cn)

Save and Forward 校验通过后，点击“Save and forward”进入下一步

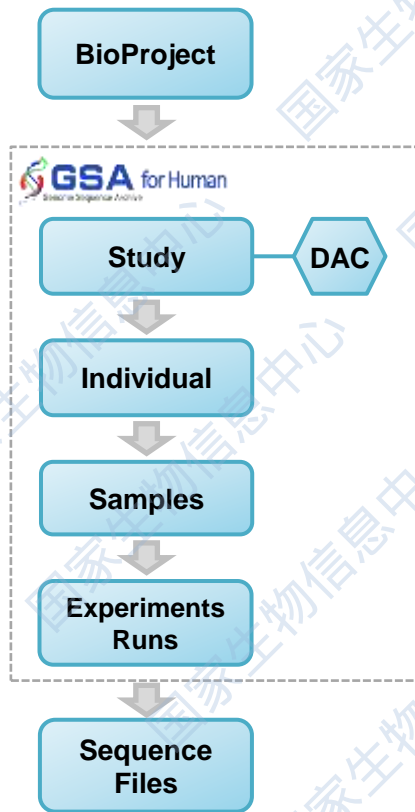
- 若文件审核未通过，请按系统提示修改信息后重新上传校验
- 如遇到任何问题，欢迎加入QQ群（548170081）交流，或通过电子邮件 [gsa@big.ac.cn](mailto:gsa@big.ac.cn) 联系我们。



# GSA-Human提交：序列文件上传方式



# GSA-Human提交: 概览 & 提交



The screenshot shows the GSA-Human submission form. The form is divided into several sections: 'Study Information', 'Study Accession', 'Study Type', 'Study Name', 'Study Date', 'Study Description', 'Study Accession', 'Study Type', 'Study Name', 'Study Date', 'Study Description', 'Study Accession', 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Information' section includes fields for 'Study Name', 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Accession' section includes fields for 'Study Accession', 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Type' section includes fields for 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Name' section includes fields for 'Study Name', 'Study Date', 'Study Description'. The 'Study Date' section includes fields for 'Study Date', 'Study Description'. The 'Study Description' section includes fields for 'Study Description'. The 'Study Accession' section includes fields for 'Study Accession', 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Type' section includes fields for 'Study Type', 'Study Name', 'Study Date', 'Study Description'. The 'Study Name' section includes fields for 'Study Name', 'Study Date', 'Study Description'. The 'Study Date' section includes fields for 'Study Date', 'Study Description'. The 'Study Description' section includes fields for 'Study Description'.



请务必点击“提交”，提交后进入数据审核流程，请耐心等待反馈信息

# GSA-Human提交：等待审核

## 数据审核流程

### 第一阶段：人工审核

账号的身份认证

是否包含隐私信息

关联关系是否合理

信息的合理与完整

元数据信息审核周期：  
**1-2**个工作日

### 第二阶段：数据关联与质控

数据文件关联

文件格式检查

质量信息分析

文件处理周期与文件  
**大小与数量** 有关

Accession	Title	Submission ID	Release Date	Package	Status	Access(Request)	Operation
Unassigned	aaaaaa	subHRA000350	2021-09-30	General	Checking detail Confidential	Controlled(0)	<button>Delete</button> <button>Update</button>
HRA000165	test	subHRA000201	2024-03-22	Cell Line	Checked OK Confidential	Open	<button>ReleaseNow</button> <button>Share</button> <button>Update</button>

- 数据审核与文件处理的状态可通过网页进行查看。如遇到任何问题，欢迎加入我们的QQ群 ([548170081](#)) 交流，或通过电子邮件[gsa@big.ac.cn](mailto:gsa@big.ac.cn)联系我们
- 数据归档后，将分配唯一的**序列号 (Accession)**。

# 设置共享链接与数据引用

分配的  
编号

New Submission							
Accession	Submission ID	Release Date	Package	Status	Access	Request	Operation
HRA000049	subHRA000005	2022-04-17	General	Checked OK Confidential	Controlled	0	<a href="#">Release Data</a> <a href="#">Share</a>



New Submission							
Accession	Submission ID	Release Date	Package	Status	Access	Request	Operation
HRA000049	subHRA000005	2022-04-17	General	Checked OK Confidential	Controlled	0	<a href="#">Release Data</a> Shared URL: <a href="https://gsa.tig.ac.cn/gsa-human/s-xxxxxx">https://gsa.tig.ac.cn/gsa-human/s-xxxxxx</a> <a href="#">Cancel Share</a>

- 用户可**自行设定共享链接**，决定其有效期以及是否允许获得者查看数据
- 注意，此**链接不应对外公布**，亦**不应在任何文章中引用**

当您成功提交数据到GSA-Human并通过审核后，请在您要发表的论文中添加如下语句：

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) in National Genomics Data Center (Nucleic Acids Res 2024), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences (GSA-Human: **HRAxxxxxx**) that are publicly accessible at <https://ngdc.cncb.ac.cn/gsa-human>.

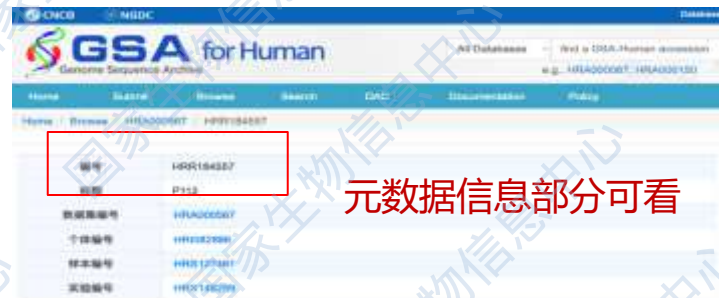
请按照以下格式引用我们的文章：

- **The Genome Sequence Archive Family: Toward Explosive Data Growth and Diverse Data Types.** *Genomics, Proteomics & Bioinformatics* 2021, 19(4):578-583  
<https://doi.org/10.1016/j.gpb.2021.08.001> [PMID=34400360]
- **Database Resources of the National Genomics Data Center, China National Center for Bioinformation in 2024.** *Nucleic Acids Res* 2024, Jan 5;52(D1):D18-D32. <https://doi.org/10.1093/nar/gkad1078> [PMID=38018256]

# GSA-Human数据访问权限

访问类型	数据情况	访问权限分类	元信息查看	元信息下载	序列文件下载
公开访问数据	<ul style="list-style-type: none"> <li>符合开放共享条件*</li> </ul>	可获取	全部	全部	可下载
	<ul style="list-style-type: none"> <li>不符合开放共享条件</li> </ul>	不可获取	部分	部分	不能
受控访问数据	<ul style="list-style-type: none"> <li>符合开放共享条件*</li> <li>已进行数据申请</li> <li>DAC审核通过</li> </ul>	可获取	全部	全部	可下载
	<ul style="list-style-type: none"> <li>符合开放共享条件*</li> <li>未进行数据申请或DAC审核未通过</li> </ul>	可获取	部分	部分	不能
	<ul style="list-style-type: none"> <li>不符合开放共享条件</li> </ul>	不可获取	部分	部分	不能

\*注：符合开放共享条件，是指数据已经获得科技部备案，或者数据为国外数据，或者数据为临床病原体、古人类或化石、商用细胞系数据



# GSA-Human数据浏览与获取

The screenshot shows the GSA for Human website interface. The top navigation bar includes links for Home, About, Datasets, Methods, and Downloads. The main content area displays a table of datasets with columns for Dataset ID, Title, Institution, Accession, and Status. A red box highlights the 'Download' button in the top left corner. Another red box highlights a specific dataset entry at the bottom, which is linked to a detailed view page.

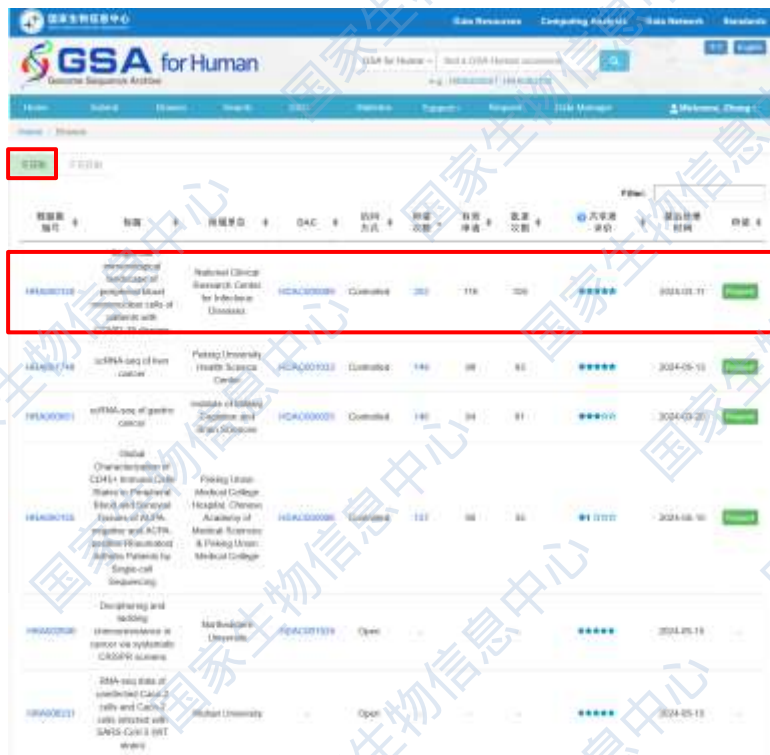
The screenshot shows the download details for a specific dataset. The top section displays the dataset ID and the download link. A red box highlights the download link and the 'Download' button. Another red box highlights the 'Download' button in the top right corner.

可获取栏中访问方式为“Open”的数据集，任何人都可在线查看完整信息并下载数据文件

<https://ngdc.cnbc.ac.cn/gsa-human/browse/>



# GSA-Human数据浏览与获取



数据集编号	标题	所属单位	DAC	访问方式	数据格式	数据量	数据更新	申请状态	申请日期
HS000001	Genomic and transcriptomic data of patients with COVID-19	National Clinical Research Center for Infectious Diseases	Controlled	Control	FASTQ	118	2020-03-11	待审核	2020-03-11
HS000002	scRNA-seq of liver cancer	Peking University Health Science Center	Controlled	Control	FASTQ	98	2020-05-18	待审核	2020-05-18
HS000003	scRNA-seq of gastric cancer	Institute of Basic Medicine and Clinical Research	Controlled	Control	FASTQ	94	2020-03-20	待审核	2020-03-20
HS000004	Whole genome sequencing of COVID-19 patients	Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College	Controlled	Control	FASTQ	86	2020-03-10	待审核	2020-03-10
HS000005	Single-cell sequencing of COVID-19 patients	Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College	Controlled	Control	FASTQ	86	2020-03-10	待审核	2020-03-10
HS000006	Genomic and transcriptomic data of patients with COVID-19	National Clinical Research Center for Infectious Diseases	Controlled	Control	FASTQ	118	2020-03-11	待审核	2020-03-11
HS000007	scRNA-seq of liver cancer	Peking University Health Science Center	Controlled	Control	FASTQ	98	2020-05-18	待审核	2020-05-18
HS000008	scRNA-seq of gastric cancer	Institute of Basic Medicine and Clinical Research	Controlled	Control	FASTQ	94	2020-03-20	待审核	2020-03-20
HS000009	Whole genome sequencing of COVID-19 patients	Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College	Controlled	Control	FASTQ	86	2020-03-10	待审核	2020-03-10
HS000010	Single-cell sequencing of COVID-19 patients	Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College	Controlled	Control	FASTQ	86	2020-03-10	待审核	2020-03-10



应用方式: Control

数据类型: Control

申请状态: 待审核

申请日期: 2020-03-11

申请说明: 本数据集为新冠病毒感染患者的基因组和转录组数据，用于研究新冠病毒的基因组变异和转录组表达。数据集包含118个样本的FASTQ文件，数据量为118GB。数据集的访问方式为Control，数据类型为FASTQ。数据集的申请状态为待审核，申请日期为2020-03-11。

- 可获取栏中访问方式为“Control”的数据集，申请审核制访问
- 数据申请：仅允许PI申请下载（与数据提交者相同的条件）
- 申请审核
  - 系统管理员将对申请者身份进行形式审查
  - DAC对数据申请做最终审核，并决定是否授权

<https://ngdc.cnbc.ac.cn/gsa-human/browse/>

# 受控数据申请

Basic Information   Research Information   Study Information   Overview   Agreement

### Basic Information

Request title

input short title

填写本次申请的标题

### PI Information

Email	chentl@big.ac.cn
First Name*	Tingting
Middle Name	
Last Name	Chen
Country/Region	
Institute/Organization	Beijing Institute of Genomics, Chinese Academy of Sciences
Department	BIG Data Center
Title/Position	

Save and forward

# 受控数据申请

## New Request

Request Study: HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy

Basic information

Research information

Study information

Overview

Agreement

### Research Information

#### Research Activity Title

a short title for the research activity

研究活动名称

#### Choose research period

2 years

研究周期

#### Research purpose

研究的目的

#### Methods & techniques

会使用到的方法和技术

Save and forward

收集数据申请者利用该数据集拟开展的研究的相关信息

# 受控数据申请

## New Request

Request Study: HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy

Basic Information Research Information Study Information Overview Agreement

### Requested study

#### Study Accession

HRA000067

#### Study Title

Multi-omics Characterization of Gastric Cancer Evolution unde

#### Uses and Limitations

General purpose use

Whether or not the confirmation of the data user's institution is required?

NO

#### Data Access Agreement

☐ I have read the data access agreement, and I agree it

Save and forward

确认申请的数据集的信息和数据提交者的要求

请仔细阅读数据访问协议 (Data Access Agreement)

# 受控数据申请

## New Request

Request Study: HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy

提供对本次数据申请信息的整体概览

Basic Information

Research Information

Study Information

Overview

Agreement

### Overview of the request

#### Basic Information

##### Request title

Test study

#### PI Information

##### Email

chenlt@big.ac.cn

##### First Name

Tingting

##### Middle Name

##### Last Name

Chen

##### Country/Region

##### Institute/Organization

Beijing Institute of Genomics, Chinese Academy of Sciences

##### Department

BIG Data Center

##### Title/Position



# 受控数据申请

## New Request

Request Study: HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy

提供对本次数据申请信息的整体概览

Basic Information

Research Information

Study Information

Overview

Agreement

### Overview of the request

#### Basic Information

##### Request title

Test study

#### PI Information

##### Email

chenlt@big.ac.cn

##### First Name

Tingting

##### Middle Name

##### Last Name

Chen

##### Country/Region

##### Institute/Organization

Beijing Institute of Genomics, Chinese Academy of Sciences

##### Department

BIG Data Center

##### Title/Position



# 受控数据申请

New Request  
Request Study: HRA000067  
Title: Multi-Omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy

Basic information | Research information | Study information | Overview | Agreement

Agreement

☐ I have read the following GSA-Human data policy, and I agree it.

1. Rights of data user

- 1) The data user can make applications for the data needed through the GSA-Human system.
- 2) The data user can download the data in certain time after obtaining the approval of accessing the dataset.
- 3) The data user has the intellectual property of the subsequent research result based on the dataset.
- 4) The data user can download the approved data by tools like Iip or Aspera, which is designated by GSA-Human.

2. Responsibilities of data user

- 1) The data user should use the downloaded dataset in the promised research scope of application. The dataset is restricted to research purpose, and can only be used for the research group and its research collaborators. Dataset distribution to other person is not allowed and cannot be used to identify individuals.
- 2) The data user should verify the quality, content and scientific validity of the downloaded dataset by his/her own.
- 3) If the data user publish a paper using the downloaded dataset from GSA-Human, the data accession number must be indicated in the paper with the note that the data can be acquired from GSA-Human.

Choose signature type

☐ Online signature ☒ Offline signature

☐ Upload a signature picture (recommend size: 120px \* 50px)

☐ Generate signature by using mouse

☐ Create a signature automatically

Save and forward

确认协议内容

签署数据访问协议：  
数据申请者可通过在线生成签名或离线手写签名两种方式签署数据访问协议

注意：如果数据提交者要求数据申请者提供单位盖章，只能通过离线方式签署协议

上传电子签名图片  
使用鼠标签名  
自动生成签名

# 受控数据申请

Home / My Request

**My Request List**

[New Request](#)

Request Number	Request ID	Request Title	Request Date	Study Accession	Status	Expire Date	Operation
HREQ000133	req000000306	Request download HRA000067 dataset	2020-07-09	HRA000067	Finished   Waiting for Curation	Use expire 2022-07-03	<a href="#">Cancel</a> <a href="#">Modify</a> <a href="#">Review History</a>

**My Request List**

[New Request](#)

Search:

Request Number	Request ID	Request Title	Request Date	Study Accession	Status	Expire Date	Operation
HREQ000133	req000000470	test2	2020-07-02	HRA000219	Approved download	Use expire 2022-07-02 Download expire 2020-08-01	<a href="#">Review History</a>

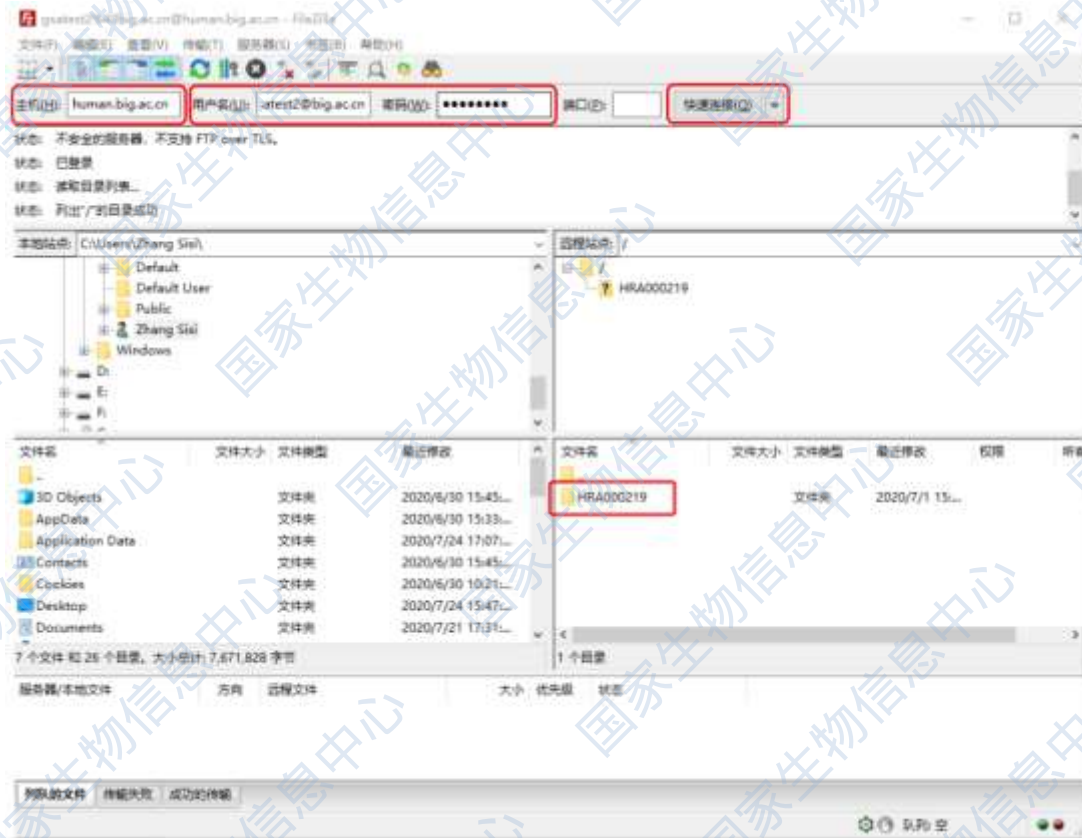
Showing 1 to 2 of 2 entries

Previous 1 Next

完成申请提交后，列表中状态（Status）变为 Finished | Waiting for Curation。

- DAC 审核通过，申请的状态变为 “Approved Download”。
- DAC 审核不通过，状态变为 “Rejected”。可点击 “Review History” 在线查看反馈意见。

# 受控数据申请-数据下载



# 用户服务



Email: [gsa@big.ac.cn](mailto:gsa@big.ac.cn)

QQ Group: 548170081