Guidance for Making Data Access Request of Genome Sequence Archive for Human

Introduction 2

User Registration 2

Making Data Access Request 5

Waiting for Review 12

Data Download 14

Introduction

The Genome Sequence Archive for Human (GSA-Human), as a part of Genome Sequence Archive (GSA) in the National Genomics Data Center (NGDC), is a data repository specialized for human genetic related data derived from biomedical researches. Two different services of data access are provided by GSA-Human: Open-access and Controlled-access. For Open-access, data are public for global researchers. While for Controlled-access, data can only be downloaded after being authorized by its Data Access Committee (DAC), which means making a successful data access request.

This documentation is aimed to give detailed guidance on how to make a data access request for Controlled-access data via GSA-Human System.

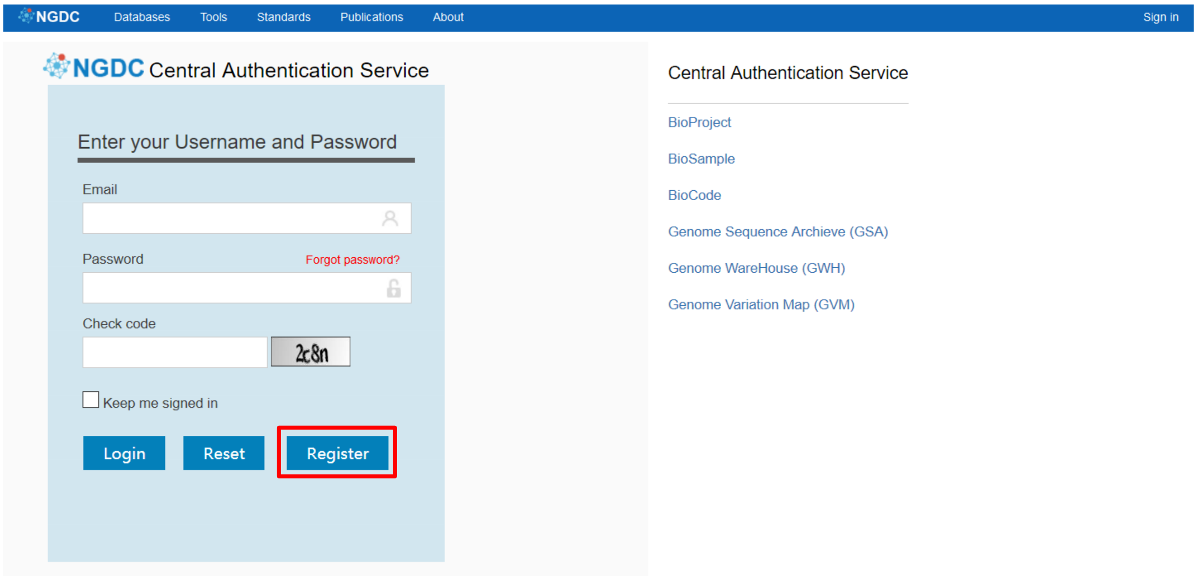
User Registration

Step 1. Register Account in BIG Single Sign-On (BIG SSO) System of NGDC and Login

It is important to note that only the Principle Investigator (PI) of the research group can be allowed to apply for Controlled-access data. Besides, before making a data access request, you are supposed to register an account via BIG Single Sign-On (BIG SSO) and sign in (See Figure below).

If you are the PI of the group and already had an account, just go to Step 2; Or if you are the staff or student of the group, please be sure to invite the PI of your group to register an account via BIG SSO, as well as you can use the PI's information to do it after permission.

If you have any problems during the process mentioned above, please feel free to contact gsa@big.ac.cn. Once finish Step 1, you can go to Step 2.


NGDC
Databases
Tools
Standards
Publications
About
Sign in

NGDC Central Authentication Service

Enter your Username and Password

Email

Password

Forgot password?

Check code

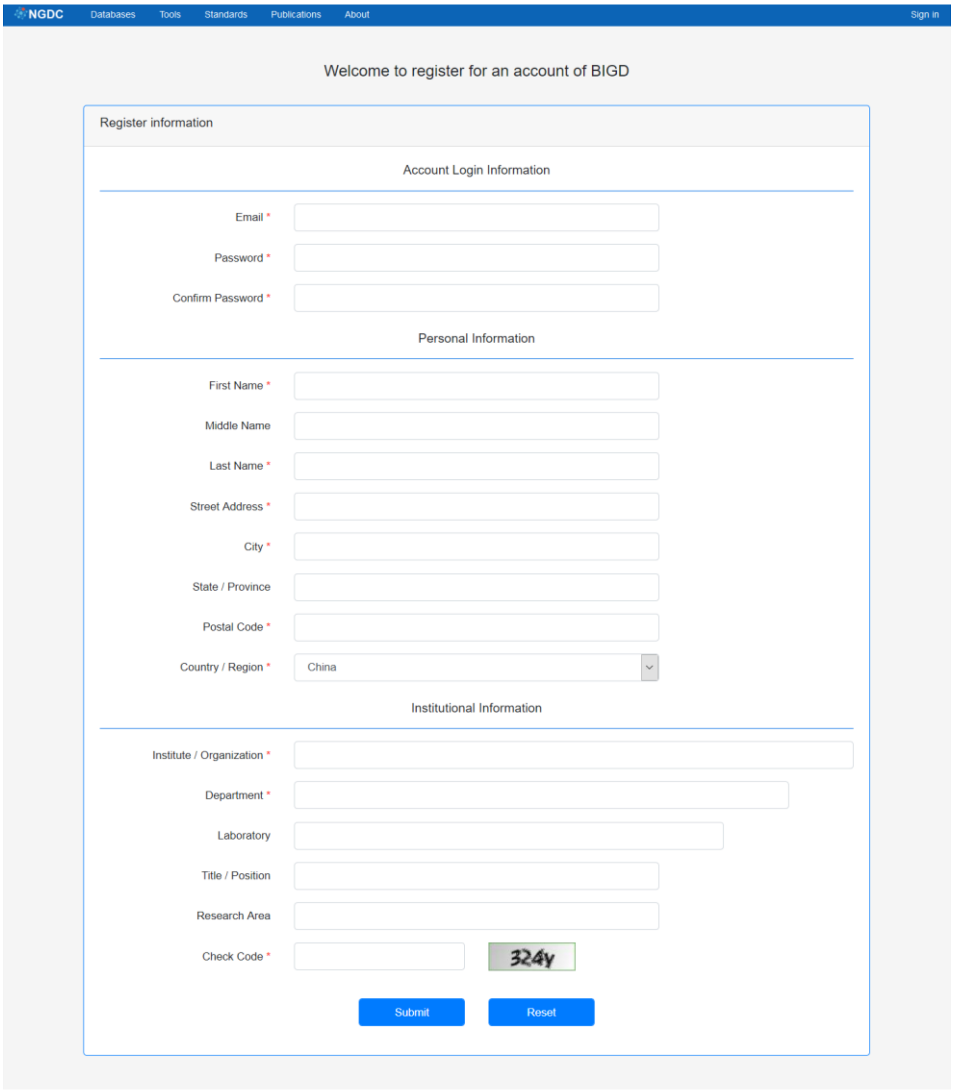
2c8n

☐ Keep me signed in

Login
Reset
Register

Central Authentication Service

- BioProject
- BioSample
- BioCode
- Genome Sequence Archieve (GSA)
- Genome WareHouse (GWH)
- Genome Variation Map (GVM)


NGDC
Databases
Tools
Standards
Publications
About
Sign in

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 *

Street Address *

City *

State / Province

Postal Code *

Country / Region * China

Institutional Information

Institute / Organization *

Department *

Laboratory

Title / Position

Research Area

Check Code *

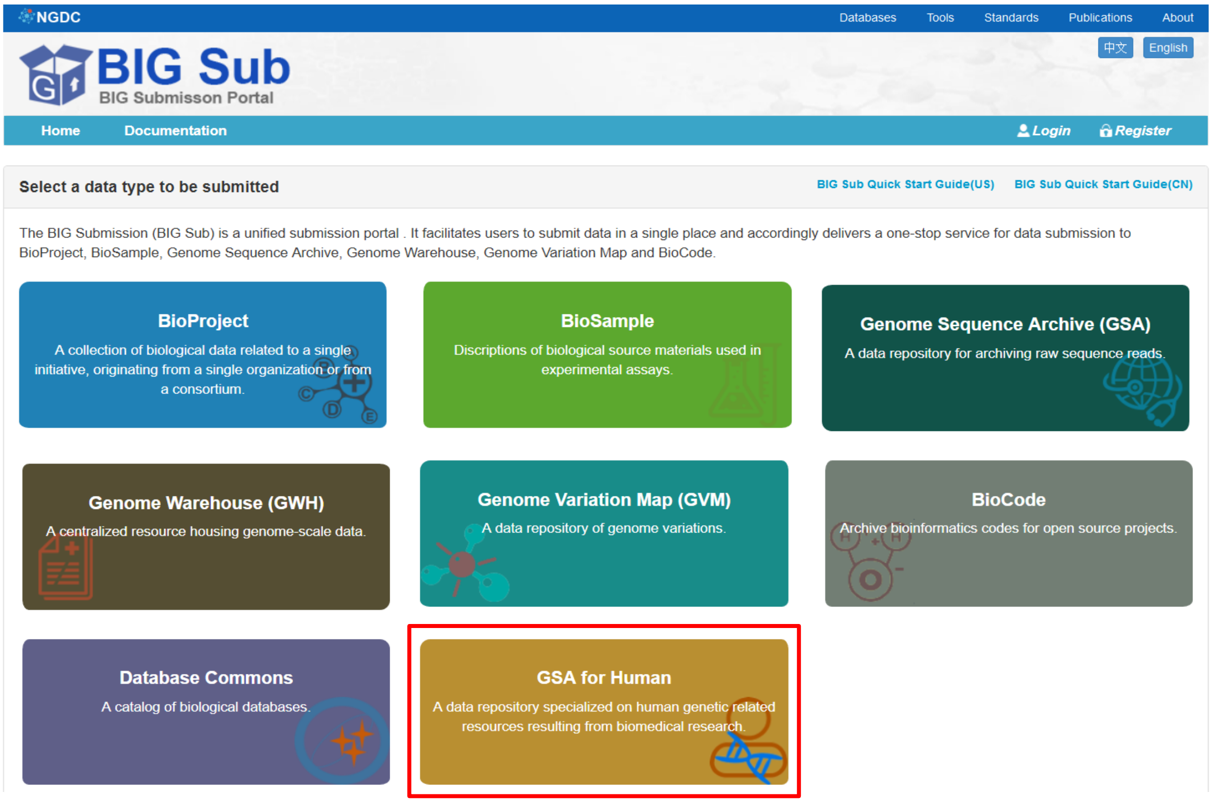
324y
Submit
Reset

Step 2. Complete the PI information in GSA-Human System

With an account of PI, you can go to the GSA-Human pages directly via BIG Submission (BIG Sub). As can be seen in the following figure, however, if it’s the first time for you to use this system, you are supposed to complete the required information of PI.

For Personal Homepage, please provide a web-link to your personal homepage in your research institution or laboratory.

Note: This information is aimed to improve your reliability as well as the success rate of your data access request, since DAC may review your homepage and research areas, and so on. While if you do not have the homepage link, other types of links can be accepted, such as the articles that you published as the corresponding author or the public website which contains your personal information.



Select a data type to be submitted

The BIG Submission (BIG Sub) is a unified submission portal . It facilitates users to submit data in a single place and accordingly delivers a one-stop service for data submission to BioProject, BioSample, Genome Sequence Archive, Genome Warehouse, Genome Variation Map and BioCode.

BioProject
A collection of biological data related to a single initiative, originating from a single organization or from a consortium.

BioSample
Discriptions of biological source materials used in experimental assays.

Genome Sequence Archive (GSA)
A data repository for archiving raw sequence reads.

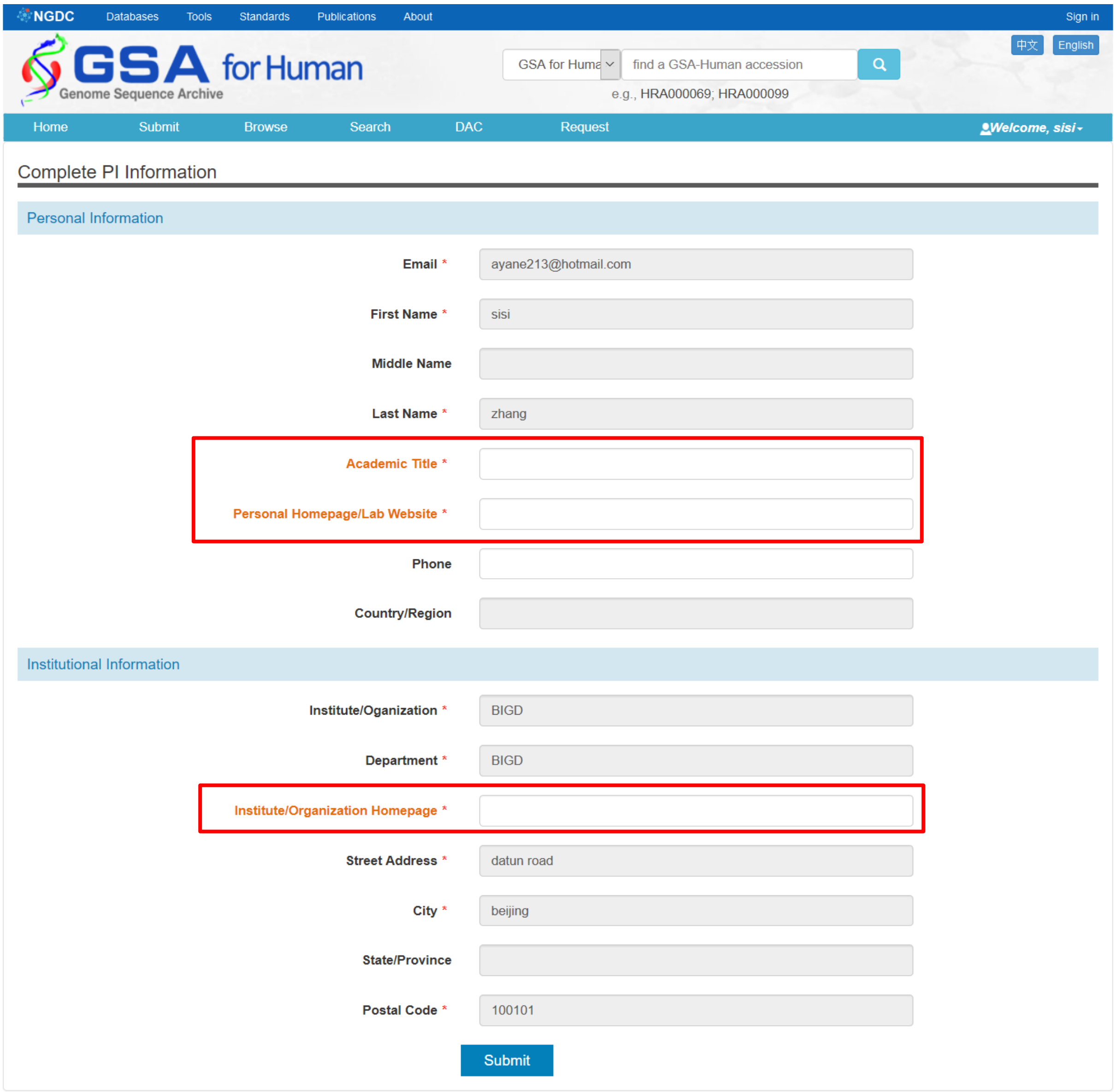
Genome Warehouse (GWH)
A centralized resource housing genome-scale data.

Genome Variation Map (GVM)
A data repository of genome variations.

BioCode
Archive bioinformatics codes for open source projects.

Database Commons
A catalog of biological databases.

GSA for Human
A data repository specialized on human genetic related resources resulting from biomedical research.


NGDC
Databases
Tools
Standards
Publications
About
Sign in
中文
English
GSA for Human
Genome Sequence Archive
GSA for Human
find a GSA-Human accession
e.g., HRA000069; HRA000099
Home
Submit
Browse
Search
DAC
Request
Welcome, sisi
Submit

Complete PI Information

Personal Information

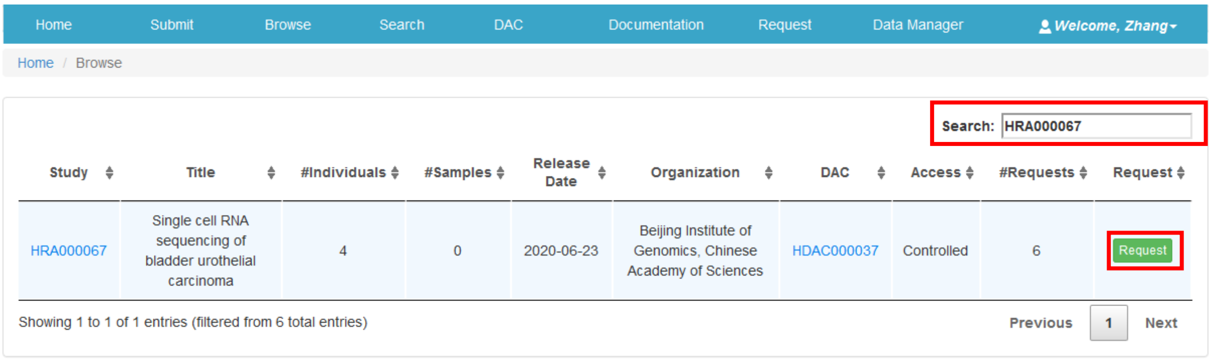
Email *	ayane213@hotmail.com
First Name *	sisi
Middle Name	
Last Name *	zhang
Academic Title *	
Personal Homepage/Lab Website *	
Phone	
Country/Region	

Institutional Information

Institute/Oganization *	BIGD
Department *	BIGD
Institute/Organization Homepage *	
Street Address *	datun road
City *	beijing
State/Province	
Postal Code *	100101

Making Data Access Request

There are two ways to enter the request system to make a new request:
Click the navbar: Browse, find the data you are willing to request via Search Bar, click the “Request” button


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / Browse
Previous
1
Next

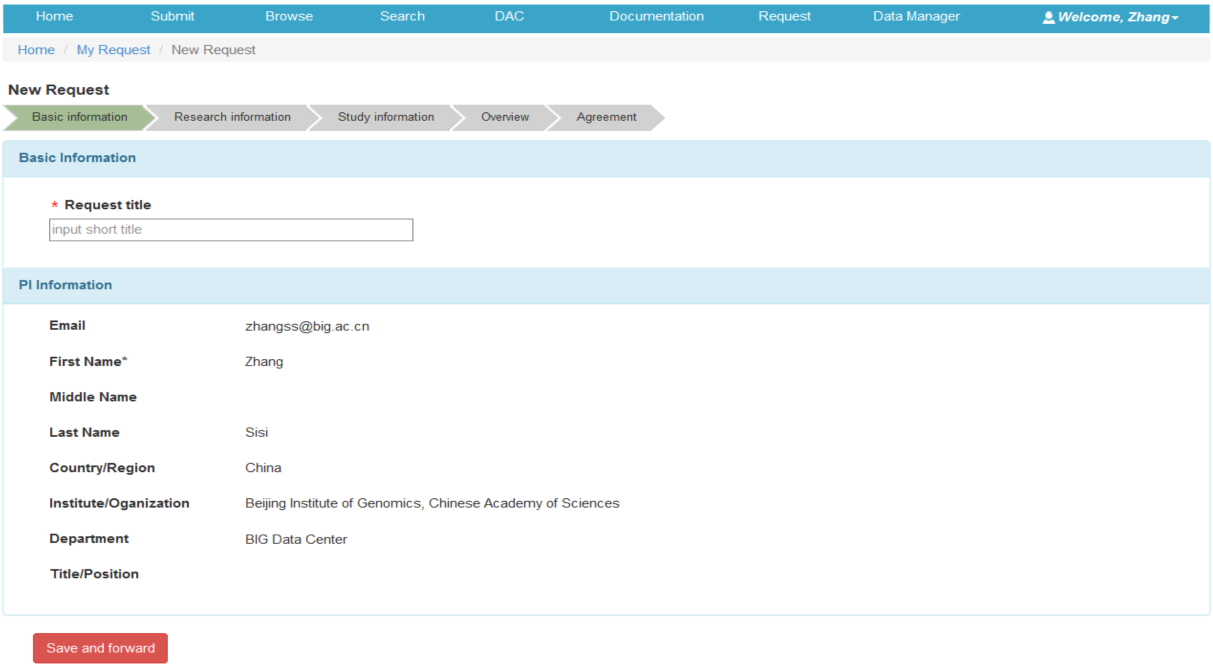
Search: HRA000067

Study	Title	#Individuals	#Samples	Release Date	Organization	DAC	Access	#Requests	Request
HRA000067	Single cell RNA sequencing of bladder urothelial carcinoma	4	0	2020-06-23	Beijing Institute of Genomics, Chinese Academy of Sciences	HDAC000037	Controlled	6	Request

Showing 1 to 1 of 1 entries (filtered from 6 total entries)

Then, please follow the steps listed below to make your Data Access Request:

Step 1. Fill Basic Information – the title of your request


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / My Request / New Request
Basic information
Research information
Study information
Overview
Agreement
Save and forward

New Request

Basic Information

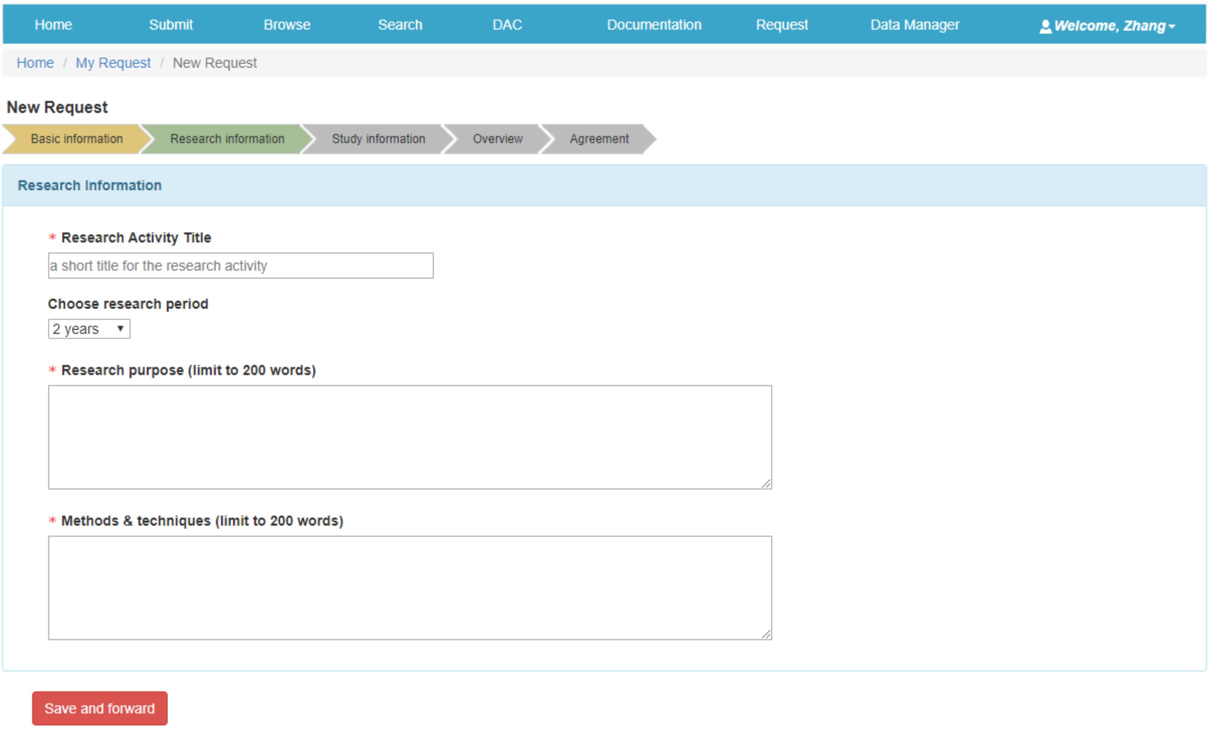
* Request title

input short title

PI Information

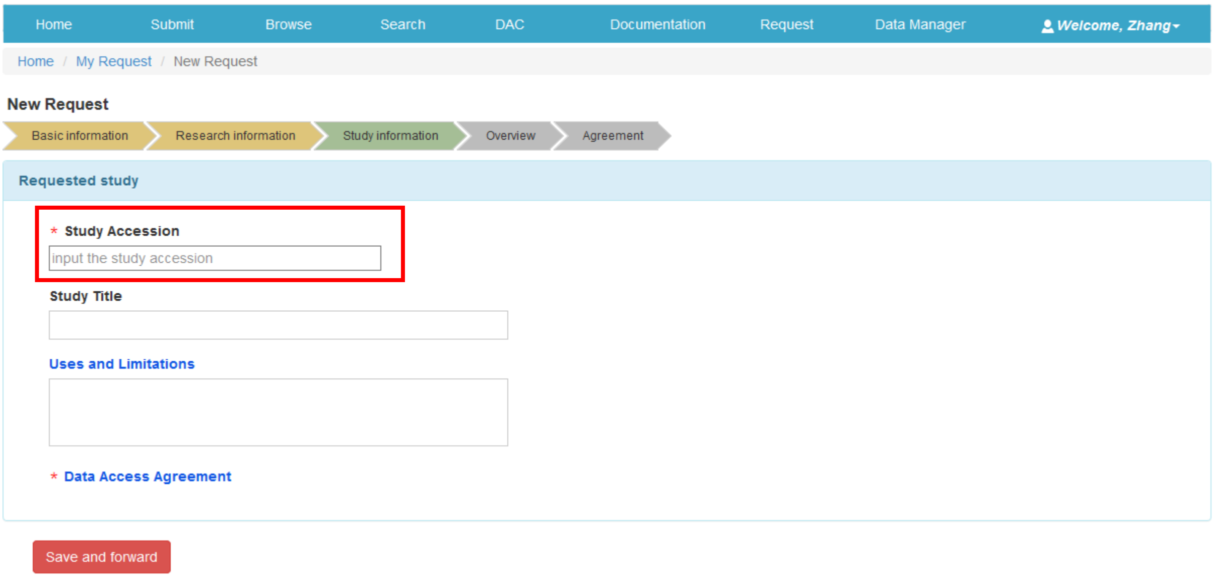
Email	zhangss@big.ac.cn
First Name*	Zhang
Middle Name	
Last Name	Sisi
Country/Region	China
Institute/Oganization	Beijing Institute of Genomics, Chinese Academy of Sciences
Department	BIG Data Center
Title/Position	

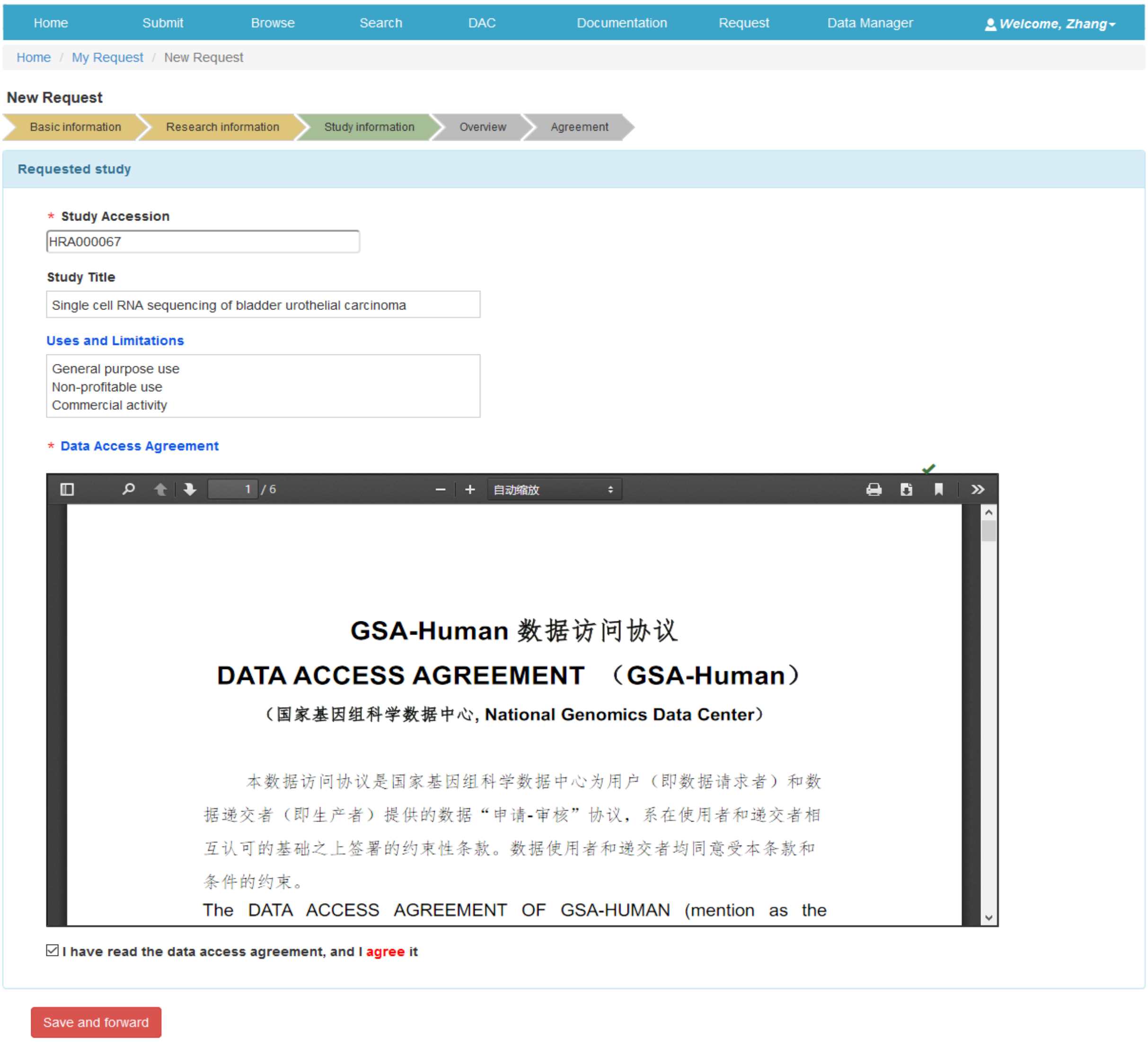
Step 2. Fill Research Information – tell the DAC how you will use this data in your research, including Research Activity Title, Choose research period, Research purpose, Methods & techniques.


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / My Request / New Request
New Request
Basic information
Research information
Study information
Overview
Agreement
Research Information
* Research Activity Title
a short title for the research activity
* Research purpose (limit to 200 words)
* Methods & techniques (limit to 200 words)
Save and forward

Choose research period
2 years

Step 3. Study information – type in the Study Accession of the data you want to apply (such as HRA000067), and you can easily review the basic information of the data, including Study Title, Use and Limitation, Data Access Agreement.


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / My Request / New Request
New Request
Basic information
Research information
Study information
Overview
Agreement
Requested study
* Study Accession
input the study accession
Study Title
Uses and Limitations
* Data Access Agreement
Save and forward


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / My Request / New Request
Basic information
Research information
Study information
Overview
Agreement
1 / 6
自动缩放
Save and forward

New Request

Requested study

* **Study Accession**
HRA000067

Study Title
Single cell RNA sequencing of bladder urothelial carcinoma

Uses and Limitations
General purpose use
Non-profitable use
Commercial activity

* **Data Access Agreement**

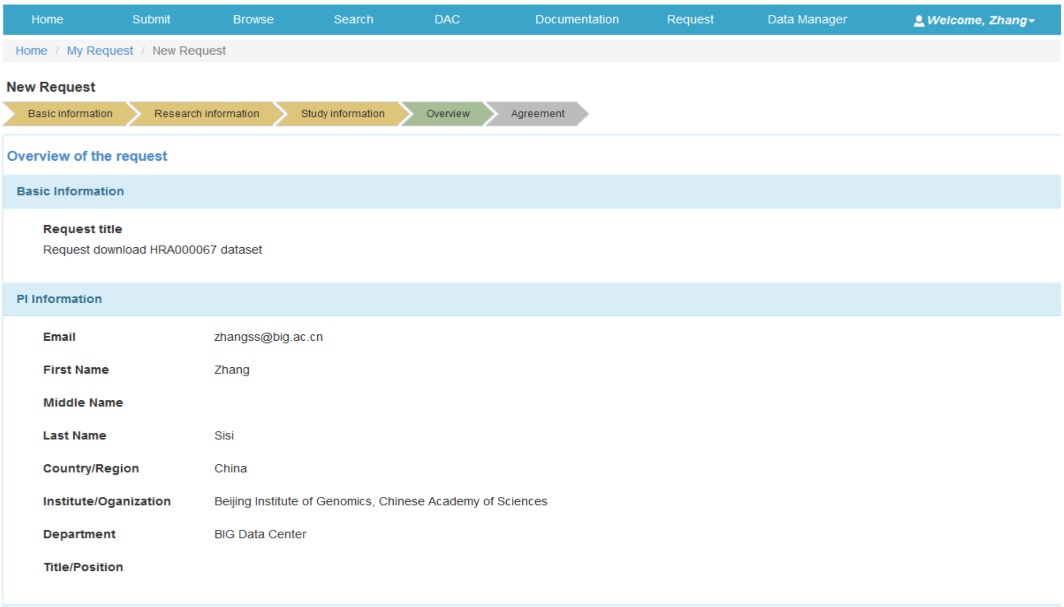
GSA-Human 数据访问协议
DATA ACCESS AGREEMENT （GSA-Human）
（国家基因组科学数据中心, National Genomics Data Center）

本数据访问协议是国家基因组科学数据中心为用户（即数据请求者）和数据递交者（即生产者）提供的数据“申请-审核”协议，系在使用者和递交者相互认可的基础之上签署的约束性条款。数据使用者和递交者均同意受本条款和条件的约束。

The DATA ACCESS AGREEMENT OF GSA-HUMAN (mention as the

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

Step 4. Overview – you can review your request information on this page


Home
Submit
Browse
Search
DAC
Documentation
Request
Data Manager
Welcome, Zhang
Home / My Request / New Request
Basic information
Research information
Study information
Overview
Agreement

New Request

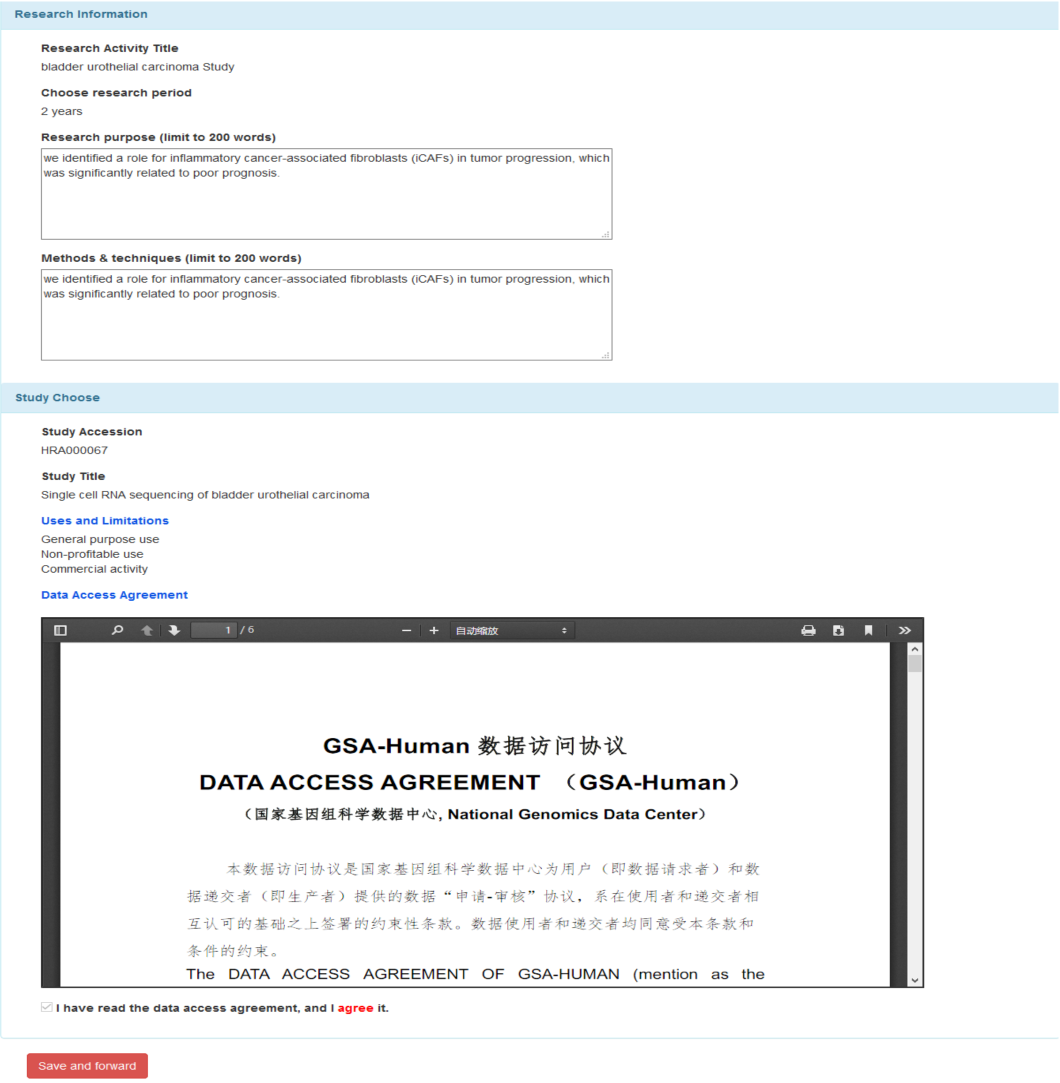
Overview of the request

Basic Information

Request title
Request download HRA000067 dataset

PI Information

Email	zhangss@big.ac.cn
First Name	Zhang
Middle Name	
Last Name	Sisi
Country/Region	China
Institute/Oganization	Beijing Institute of Genomics, Chinese Academy of Sciences
Department	BIG Data Center
Title/Position	



Research Information

Research Activity Title
bladder urothelial carcinoma Study

Choose research period
2 years

Research purpose (limit to 200 words)
we identified a role for inflammatory cancer-associated fibroblasts (iCAFs) in tumor progression, which was significantly related to poor prognosis.

Methods & techniques (limit to 200 words)
we identified a role for inflammatory cancer-associated fibroblasts (iCAFs) in tumor progression, which was significantly related to poor prognosis.

Study Choose

Study Accession
HRA000067

Study Title
Single cell RNA sequencing of bladder urothelial carcinoma

Uses and Limitations
General purpose use
Non-profitable use
Commercial activity

Data Access Agreement

GSA-Human 数据访问协议

DATA ACCESS AGREEMENT （GSA-Human）

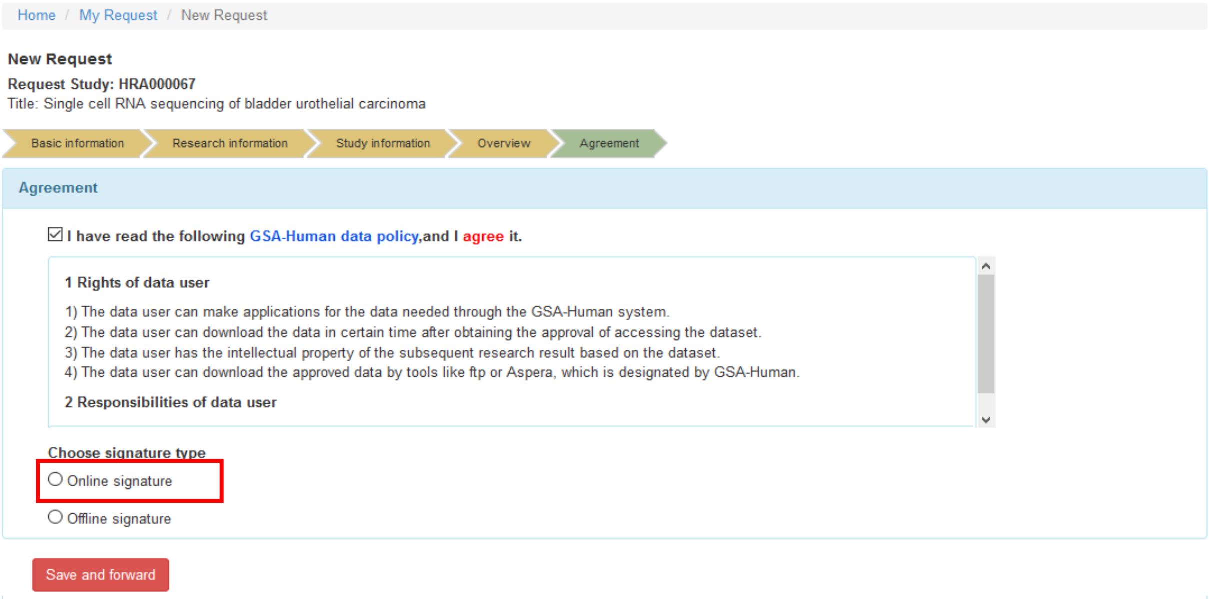
（国家基因组科学数据中心, National Genomics Data Center）

本数据访问协议是国家基因组科学数据中心为用户（即数据请求者）和数据递交者（即生产者）提供的数据“申请-审核”协议，系在使用者和递交者相互认可的基础之上签署的约束性条款。数据使用者和递交者均同意受本条款和条件的约束。

The DATA ACCESS AGREEMENT OF GSA-HUMAN (mention as the

☑ **I have read the data access agreement, and I agree it.**

Step 5. Agreement – you have to sign the Data Access Agreement, via Online signature or Offline signature.



New Request

Request Study: HRA000067

Title: Single cell RNA sequencing of bladder urothelial carcinoma

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 ftp or Aspera, which is designated by GSA-Human.

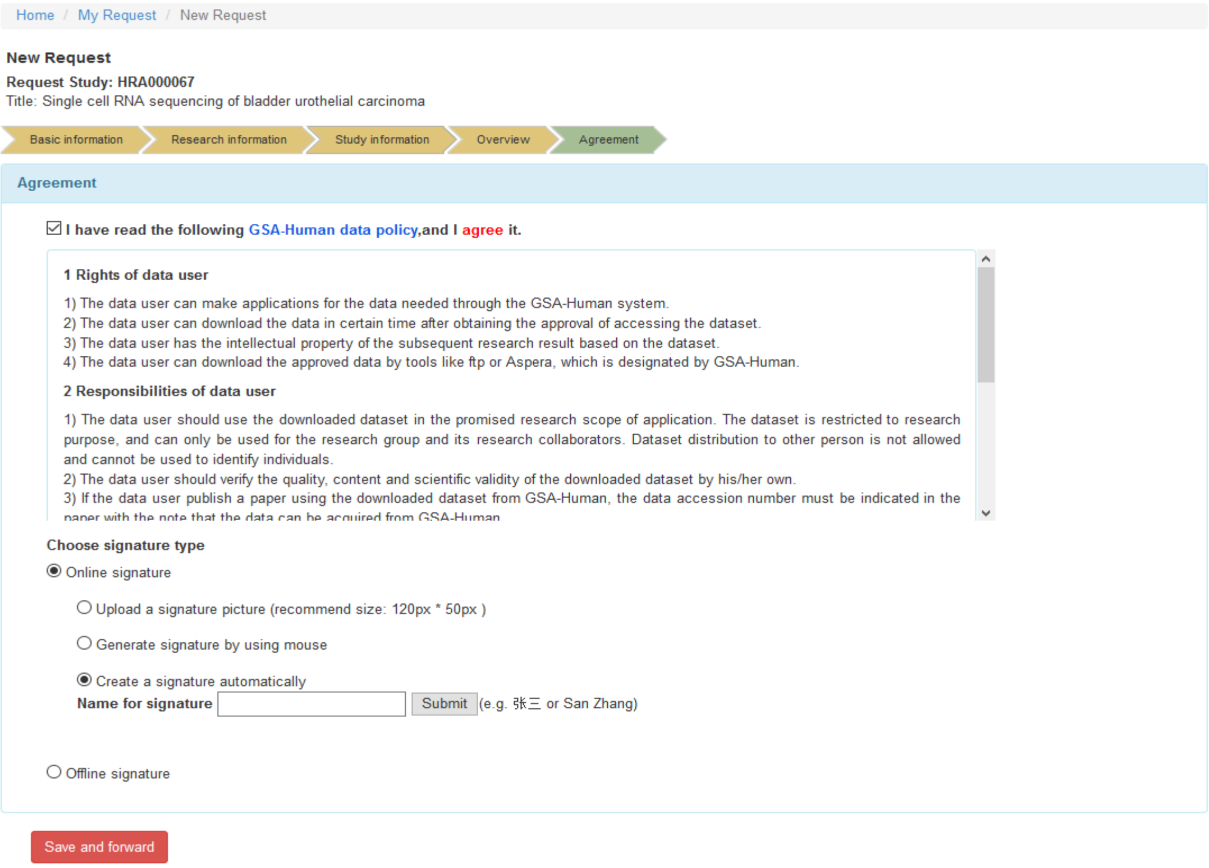
2 Responsibilities of data user

Choose signature type

○ Online signature

○ Offline signature

1. For Online signature, this page provides three methods: Upload a signature picture, generate a signature by using mouse, and Create a signature automatically. Using the third methods as an example, the following process can be seen below:



New Request

Request Study: HRA000067

Title: Single cell RNA sequencing of bladder urothelial carcinoma

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 ftp 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

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

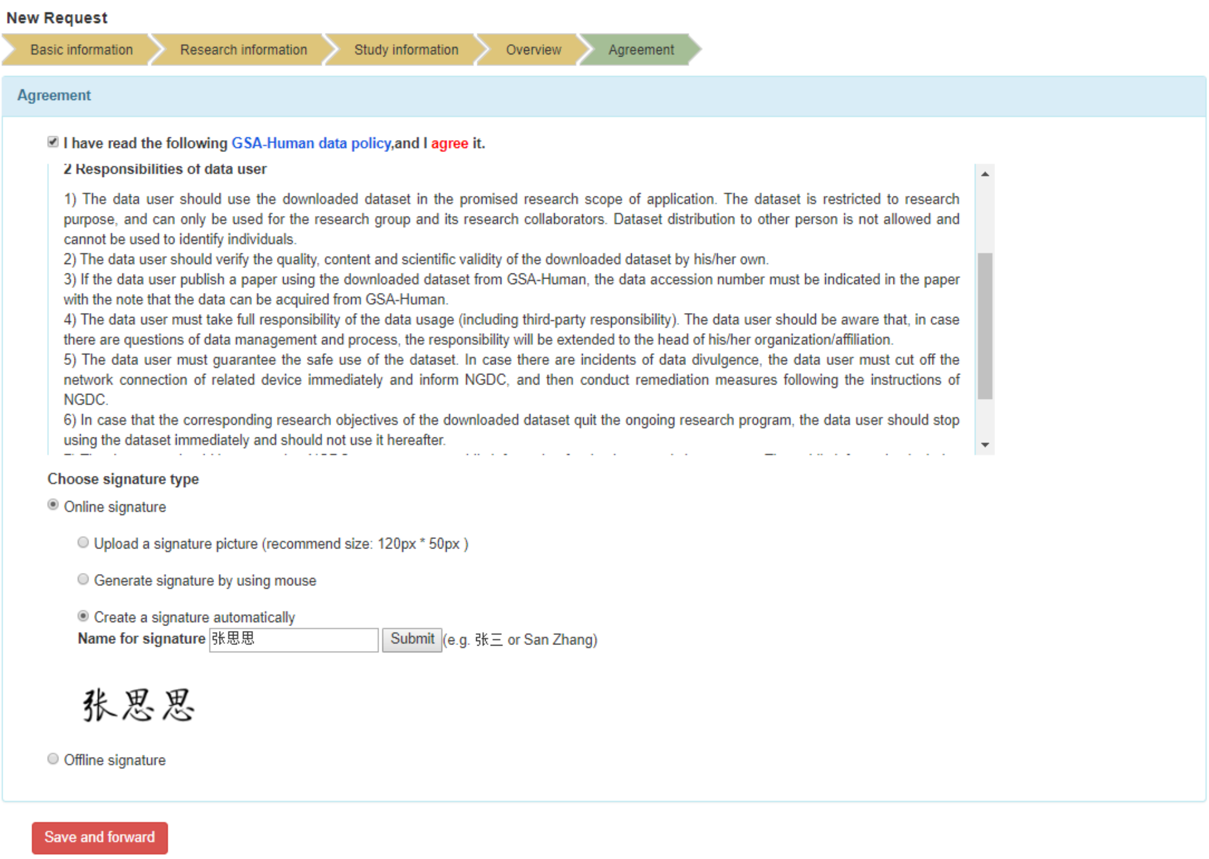
○ Generate signature by using mouse

● Create a signature automatically

Name for signature ＿＿＿＿ Submit (e.g. 张三 or San Zhang)

○ Offline signature

⇩



Agreement

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

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.
4) The data user must take full responsibility of the data usage (including third-party responsibility). The data user should be aware that, in case there are questions of data management and process, the responsibility will be extended to the head of his/her organization/affiliation.
5) The data user must guarantee the safe use of the dataset. In case there are incidents of data divulgence, the data user must cut off the network connection of related device immediately and inform NGDC, and then conduct remediation measures following the instructions of NGDC.
6) In case that the corresponding research objectives of the downloaded dataset quit the ongoing research program, the data user should stop using the dataset immediately and should not use it hereafter.

Choose signature type

● Online signature

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

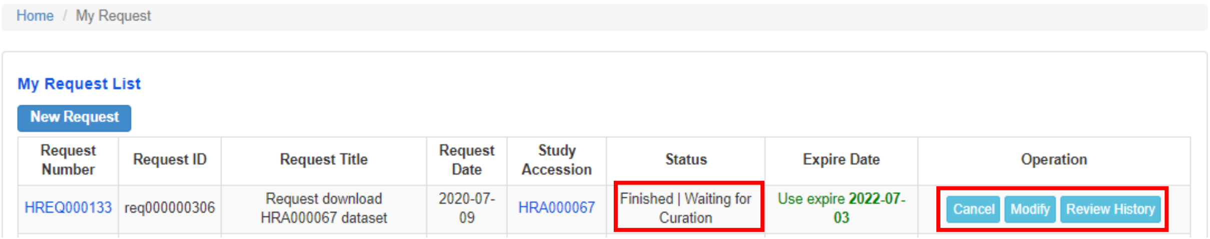
○ Generate signature by using mouse

● Create a signature automatically

Name for signature 张思思 Submit (e.g. 张三 or San Zhang)

张思思

○ Offline signature

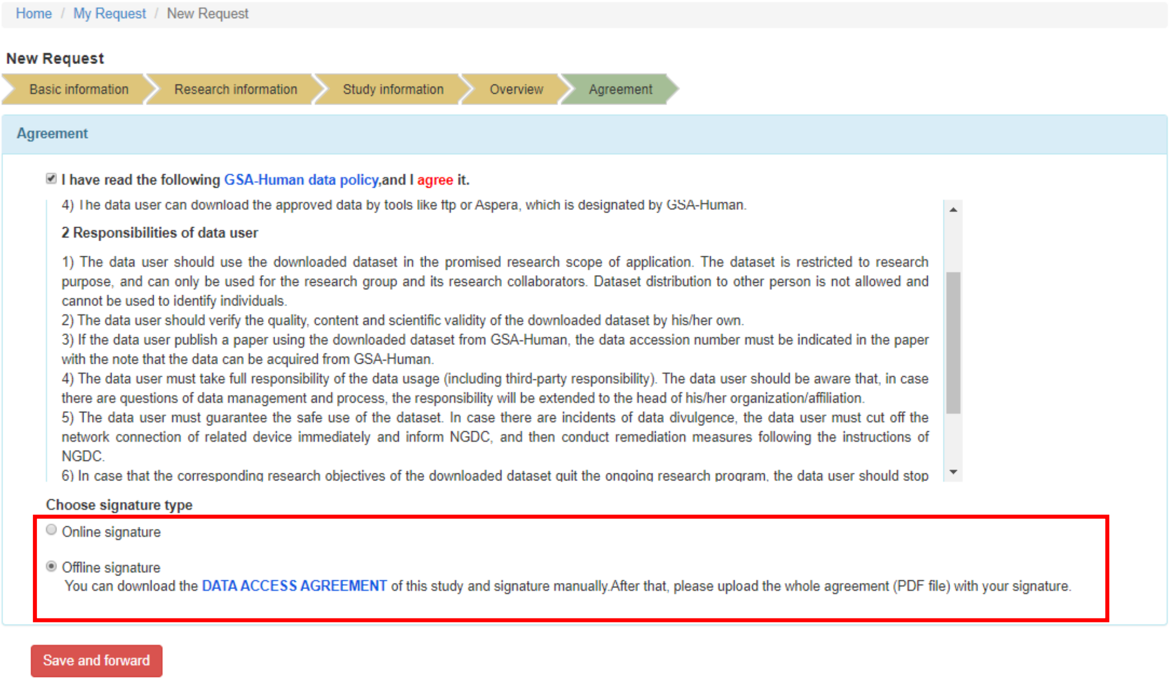


My Request List

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	Cancel Modify Review History

Note: After generating and submitting the signature, your request will be submitted to the system and appear in your request list. In addition, its status will be "Finished | Waiting for Curation". Please wait for the system admin doing the first check for you. Before the request being checked, you can freely cancel, modify, review your request.

2. For Offline signature, the process is as follows:
    - Please click the “DAA Download” button and download the DAA file of this study.
    - After that, please “upload” button to upload the Scanned DAA file. **For the domestic users**, please prove the scanning copy of DAA document with the signature of you and your organization and the stamp of your organization. **For international users**, please prove the scanning copy of signed DAA document from you and your organization.
    - Then please wait for the system admin doing the first review for you. Before the request being reviewed, you can freely cancel, modify, review your request.



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

4) The data user can download the approved data by tools like ftp or Aspera, which is designated by GSA-Human.

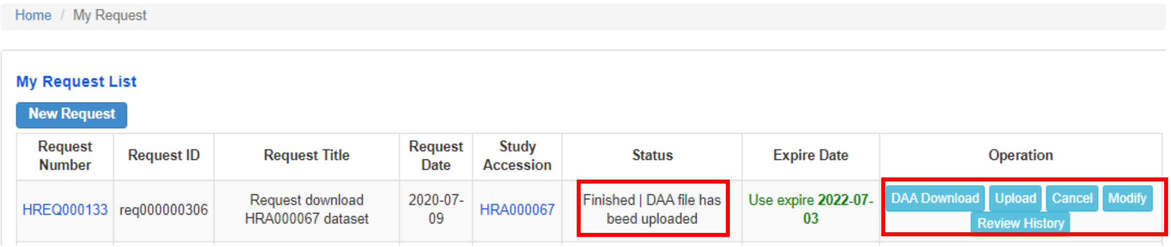
2 Responsibilities of data user

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.
4) The data user must take full responsibility of the data usage (including third-party responsibility). The data user should be aware that, in case there are questions of data management and process, the responsibility will be extended to the head of his/her organization/affiliation.
5) The data user must guarantee the safe use of the dataset. In case there are incidents of data divulgence, the data user must cut off the network connection of related device immediately and inform NGDC, and then conduct remediation measures following the instructions of NGDC.
6) In case that the corresponding research objectives of the downloaded dataset quit the ongoing research program, the data user should stop

Choose signature type

○ Online signature

● Offline signature
You can download the DATA ACCESS AGREEMENT of this study and signature manually.After that, please upload the whole agreement (PDF file) with your signature.



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 | DAA file has beed uploaded	Use expire 2022-07-03	DAA Download Upload Cancel Modify Review History

Waiting for Review

After submitting the request successfully, it may take some time for being reviewed by both system and the DAC, thanks for your patient. Once you have passed the review, congratulation, you will be approved to download the data.

System Review: The system admin will do the first and basic review for you. Please fill in all of the required information carefully. For instance, if your identity is not reliable enough (miss the Personal Homepage information) or lack of required documentation, your request will be rejected. If the request has be approved, its status will be "**Waiting for DAC review**"; otherwise, its status will be "**Rejected**". Please click the "**Review History**" to check the detail of the feedback.

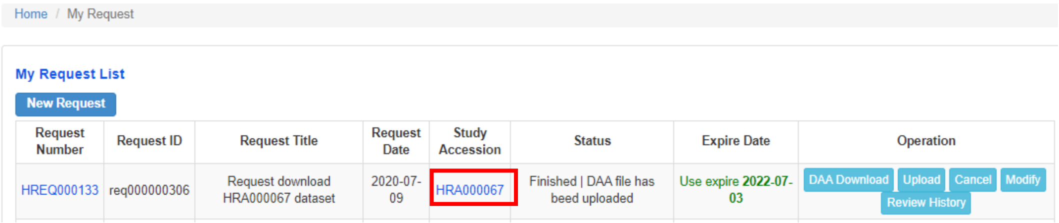
DAC Review: The requests that passed the System Review will be transferred to the DAC Review. While passing the DAC Review, you will receive a notification by email, so please pay attention to your email. If the request has be approved, its status will be "**Approved Download**"; otherwise, its status will be "**Rejected**". Please click the "**Review History**" to check the detail of the feedback.

If the DAC does not give their feedback after two weeks, the system will send an email to notify them automatically. There are differences among the frequency of DAC contacts processes the request. You can check the last processed time or the number of approved requests by searching the accession you requested in the browse page (https://ngdc.cncb.ac.cn/gsa-human/browse/)**.**

- The columns of the "**requests**" show the number of requests that users submitted.
- The columns of the "**approve**" show the number of requests that DAC contact's approved.
- The columns of the "**last processed**" show the last time that DAC contact's login in the GSA-Human database.

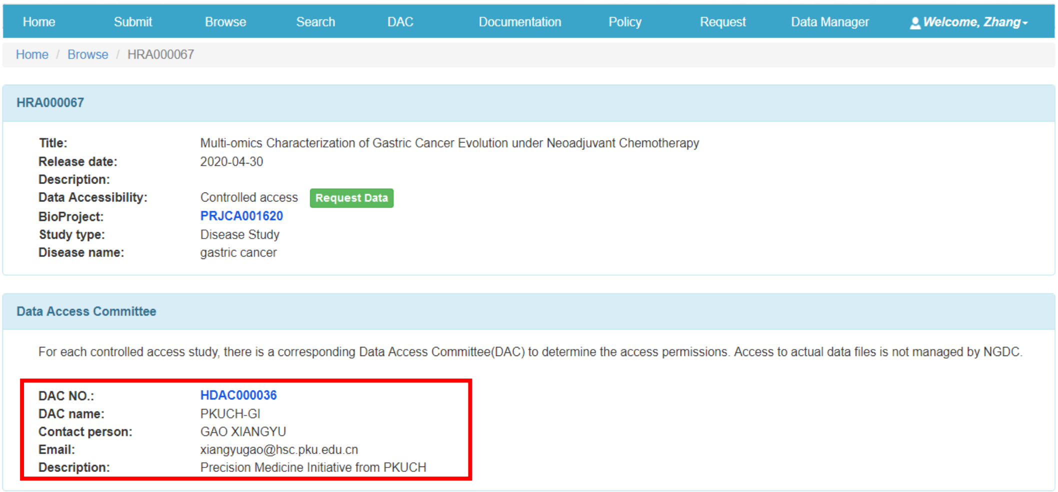
If the DAC has not processed the request for a long time, you can try to email the DAC contacts to get the real reasons. Please follow the steps below to find the DAC contacts' email.

- Open the browse page (https://ngdc.cncb.ac.cn/gsa-human/browse/) and find the dataset that you requested.
- Click the study accession of the dataset to enter the dataset browsing page, and you can find the DAC contact person's email in Data Access Committee part.


Home / My Request
New Request
DAA Download
Upload
Cancel
Modify
Review History

My Request List

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 | DAA file has beed uploaded	Use expire 2022-07-03	DAA Download Upload Cancel Modify Review History


Home
Submit
Browse
Search
DAC
Documentation
Policy
Request
Data Manager
Welcome, Zhang
Home / Browse / HRA000067
Request Data

HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy
Release date: 2020-04-30
Description:
Data Accessibility: Controlled access
BioProject: PRJCA001620
Study type: Disease Study
Disease name: gastric cancer

Data Access Committee

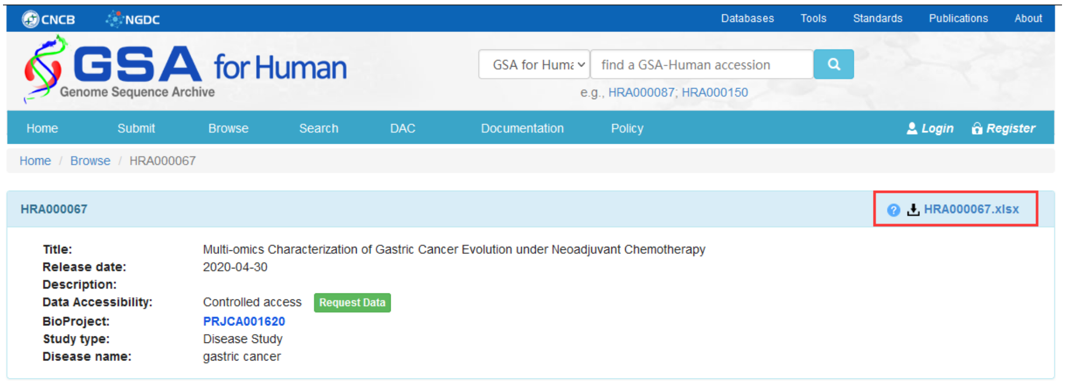
For each controlled access study, there is a corresponding Data Access Committee(DAC) to determine the access permissions. Access to actual data files is not managed by NGDC.

DAC NO.: HDAC000036
DAC name: PKUCH-GI
Contact person: GAO XIANGYU
Email: xiangyugao@hsc.pku.edu.cn
Description: Precision Medicine Initiative from PKUCH

NOTICE: If the dataset data not yet been registered in the Human Genetic Resource Management Platform of Ministry of Science and Technology of the People´s Republic of China, the international users are unavailable to request this dataset currently. Please wait until the owners get the Registration and backup numbers from Human Genetic Resource Management Platform.

Data and Metadata table Download

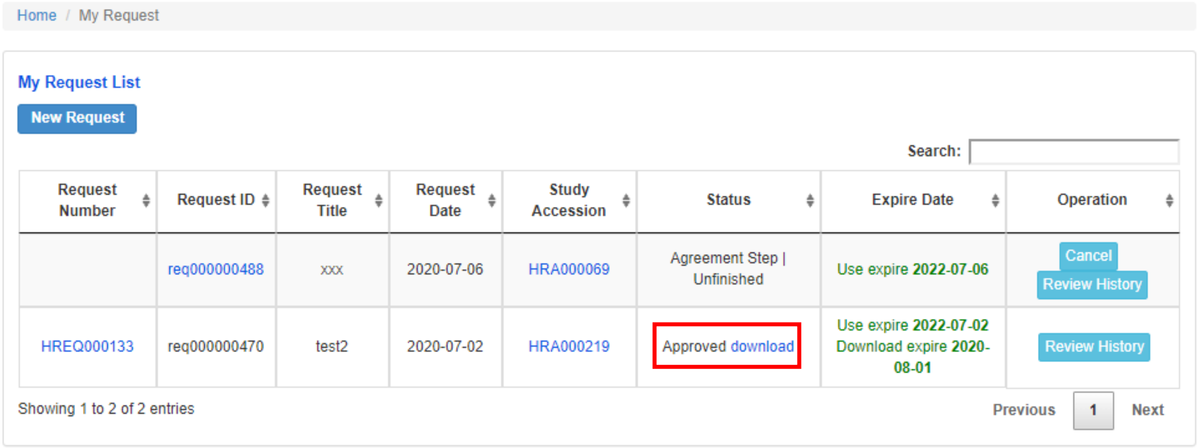
After passing the review, the status of your request in the list will become “**Approved**”. For downloading metadata table, please login and open the dataset you requested in the browsing page, like https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000067. Then click the table download button as following picture below.


CNCB
NGDC
Databases
Tools
Standards
Publications
About
GSA for Human
Genome Sequence Archive
GSA for Huma
find a GSA-Human accession
e.g., HRA000087; HRA000150
Home
Submit
Browse
Search
DAC
Documentation
Policy
Login
Register
Home / Browse / HRA000067
HRA000067.xlsx
Request Data

HRA000067

Title: Multi-omics Characterization of Gastric Cancer Evolution under Neoadjuvant Chemotherapy
Release date: 2020-04-30
Description:
Data Accessibility: Controlled access
BioProject: PRJCA001620
Study type: Disease Study
Disease name: gastric cancer

For downloading data file, you are able to click the “download” link and download the data from the data directory directly via FTP with your account login.


Home / My Request

My Request List

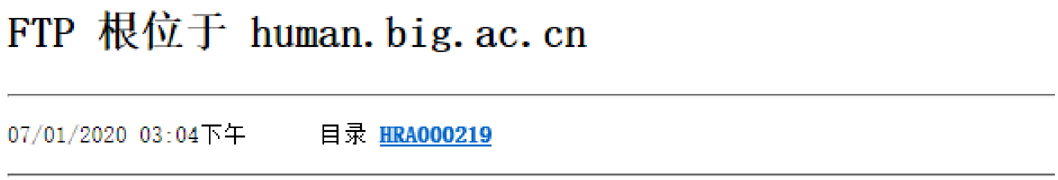
New Request

Search:

Request Number	Request ID	Request Title	Request Date	Study Accession	Status	Expire Date	Operation
	req000000488	xxx	2020-07-06	HRA000069	Agreement Step | Unfinished	Use expire 2022-07-06	Cancel Review History
HREQ000133	req000000470	test2	2020-07-02	HRA000219	Approved download	Use expire 2022-07-02 Download expire 2020-08-01	Review History

Showing 1 to 2 of 2 entries

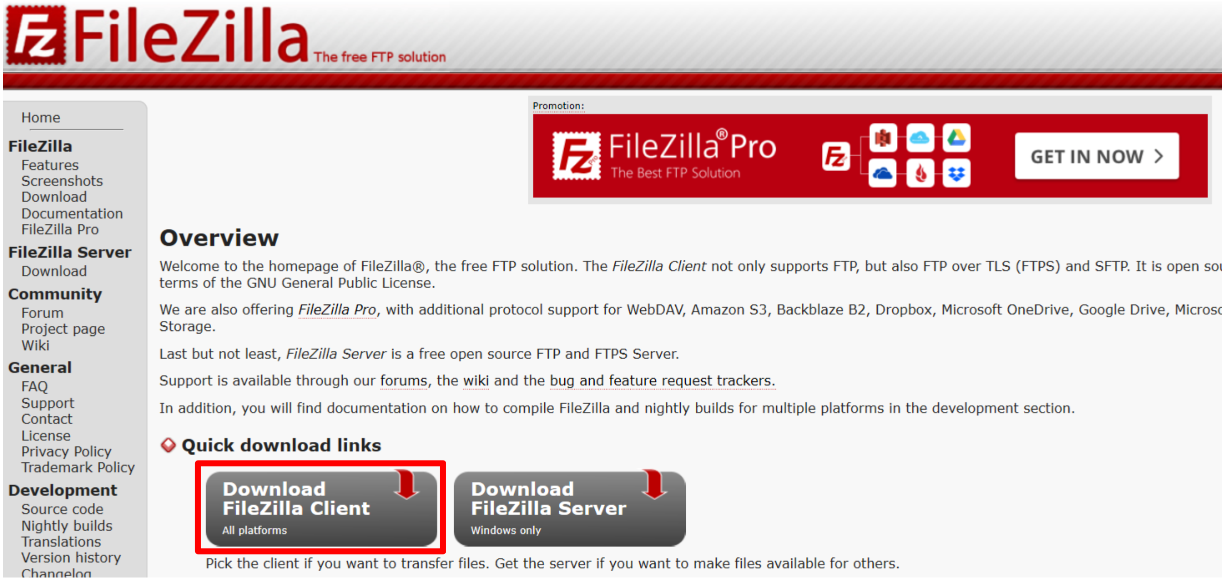
Previous 1 Next


FTP 根位于 human.big.ac.cn

07/01/2020 03:04下午 目录 HRA000219

As downloading data through the website is not stable, we highly recommend you use an FTP client to sign into our FTP server for big data downloading, such as FileZilla. Taking FileZilla as an example, you can download the data following four steps:

Step 1. download and install FileZilla (https://filezilla-project.org/), as shown in the figure below, click “Download FileZilla Client” and install the software following the guidance.


FileZilla The free FTP solution
Home
FileZilla
Features
Screenshots
Download
Documentation
FileZilla Pro
FileZilla Server
Download
Community
Forum
Project page
Wiki
General
FAQ
Support
Contact
License
Privacy Policy
Trademark Policy
Development
Source code
Nightly builds
Translations
Version history
Changelog
Promotion:
FileZilla® Pro The Best FTP Solution
GET IN NOW ›
Download FileZilla Client All platforms
Download FileZilla Server Windows only

Overview

Welcome to the homepage of FileZilla®, the free FTP solution. The *FileZilla Client* not only supports FTP, but also FTP over TLS (FTPS) and SFTP. It is open sou terms of the GNU General Public License.

We are also offering *FileZilla Pro*, with additional protocol support for WebDAV, Amazon S3, Backblaze B2, Dropbox, Microsoft OneDrive, Google Drive, Microsc Storage.

Last but not least, *FileZilla Server* is a free open source FTP and FTPS Server.

Support is available through our forums, the wiki and the bug and feature request trackers.

In addition, you will find documentation on how to compile FileZilla and nightly builds for multiple platforms in the development section.

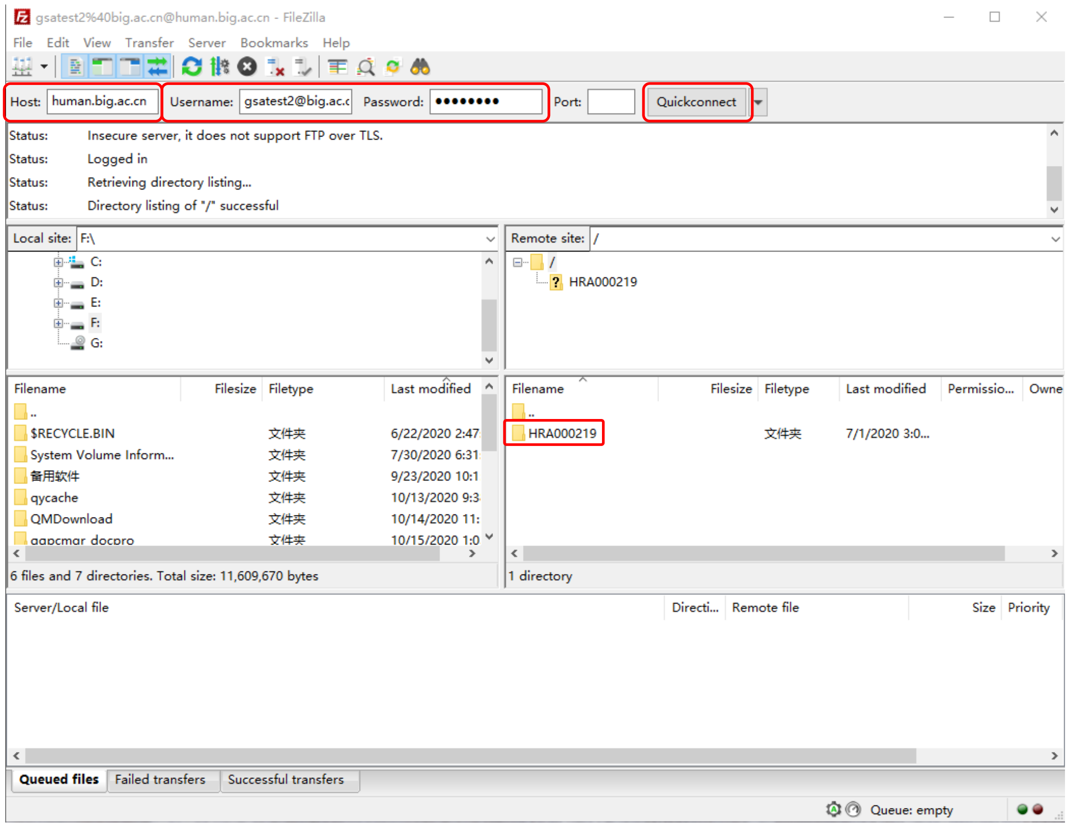
Quick download links

Pick the client if you want to transfer files. Get the server if you want to make files available for others.

Step 2. Open FileZilla (see figure below), input the Host as “human.big.ac.cn”, User and Password are the same as your account (the login email and password), do not need to fill in the port, finally, click “Quickconnect”. The status in the message log will show you have login successfully. If not, please check the issues based on the message shown in the log.

Step 3. After login, in the local pane, set where you want to save the data.

Step 4. In the remote pane, double click the folder of the dataset, such as “**HRA000219**”, and just drag the files/directories from the remote pane to the local pane.


gsatest2%40big.ac.cn@human.big.ac.cn - FileZilla
File Edit View Transfer Server Bookmarks Help
Quickconnect

Host: human.big.ac.cn Username: gsatest2@big.ac.c Password: •••••••• Port:

```
Status:    Insecure server, it does not support FTP over TLS.
Status:    Logged in
Status:    Retrieving directory listing...
Status:    Directory listing of "/" successful
```

Local site: F:\

Remote site: /

Filename	Filesize	Filetype	Last modified
..			
$RECYCLE.BIN		文件夹	6/22/2020 2:47
System Volume Inform...		文件夹	7/30/2020 6:31
备用软件		文件夹	9/23/2020 10:1
qycache		文件夹	10/13/2020 9:3
QMDownload		文件夹	10/14/2020 11:
qqpcmgr_docpro		文件夹	10/15/2020 1:0

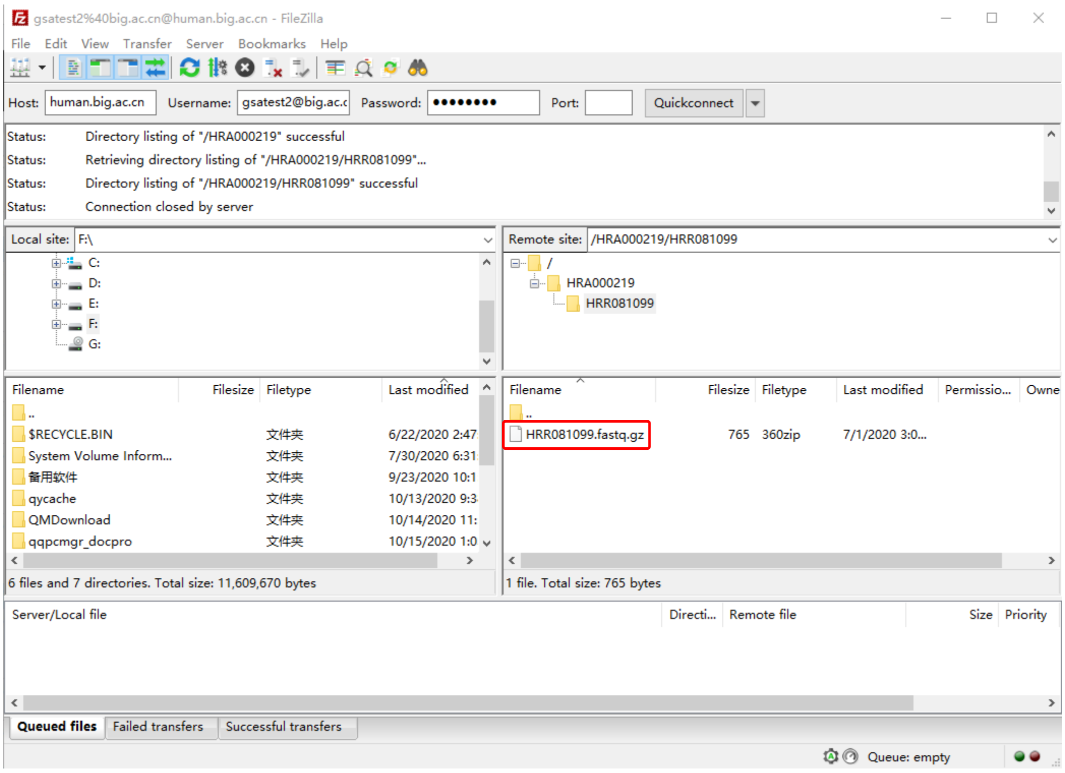
6 files and 7 directories. Total size: 11,609,670 bytes

Filename	Filesize	Filetype	Last modified	Permissio...	Owne
..					
HRA000219		文件夹	7/1/2020 3:0...		

1 directory

Server/Local file
Directi...
Remote file
Size
Priority
Queued files
Failed transfers
Successful transfers
Queue: empty

⇩


gsatest2%40big.ac.cn@human.big.ac.cn - FileZilla
File Edit View Transfer Server Bookmarks Help
Quickconnect

Host: human.big.ac.cn Username: gsatest2@big.ac.c Password: •••••••• Port:

```
Status:    Directory listing of "/HRA000219" successful
Status:    Retrieving directory listing of "/HRA000219/HRR081099"...
Status:    Directory listing of "/HRA000219/HRR081099" successful
Status:    Connection closed by server
```

Local site: F:\

Remote site: /HRA000219/HRR081099

Filename	Filesize	Filetype	Last modified
..			
$RECYCLE.BIN		文件夹	6/22/2020 2:47
System Volume Inform...		文件夹	7/30/2020 6:31
备用软件		文件夹	9/23/2020 10:1
qycache		文件夹	10/13/2020 9:3
QMDownload		文件夹	10/14/2020 11:
qqpcmgr_docpro		文件夹	10/15/2020 1:0

6 files and 7 directories. Total size: 11,609,670 bytes

Filename	Filesize	Filetype	Last modified	Permissio...	Owne
..					
HRR081099.fastq.gz	765	360zip	7/1/2020 3:0...		

1 file. Total size: 765 bytes

Server/Local file
Directi...
Remote file
Size
Priority
Queued files
Failed transfers
Successful transfers
Queue: empty