

ILLUMINA HiSeq 2000/2500/3000/4000 高通量测序数据上传GEO方法

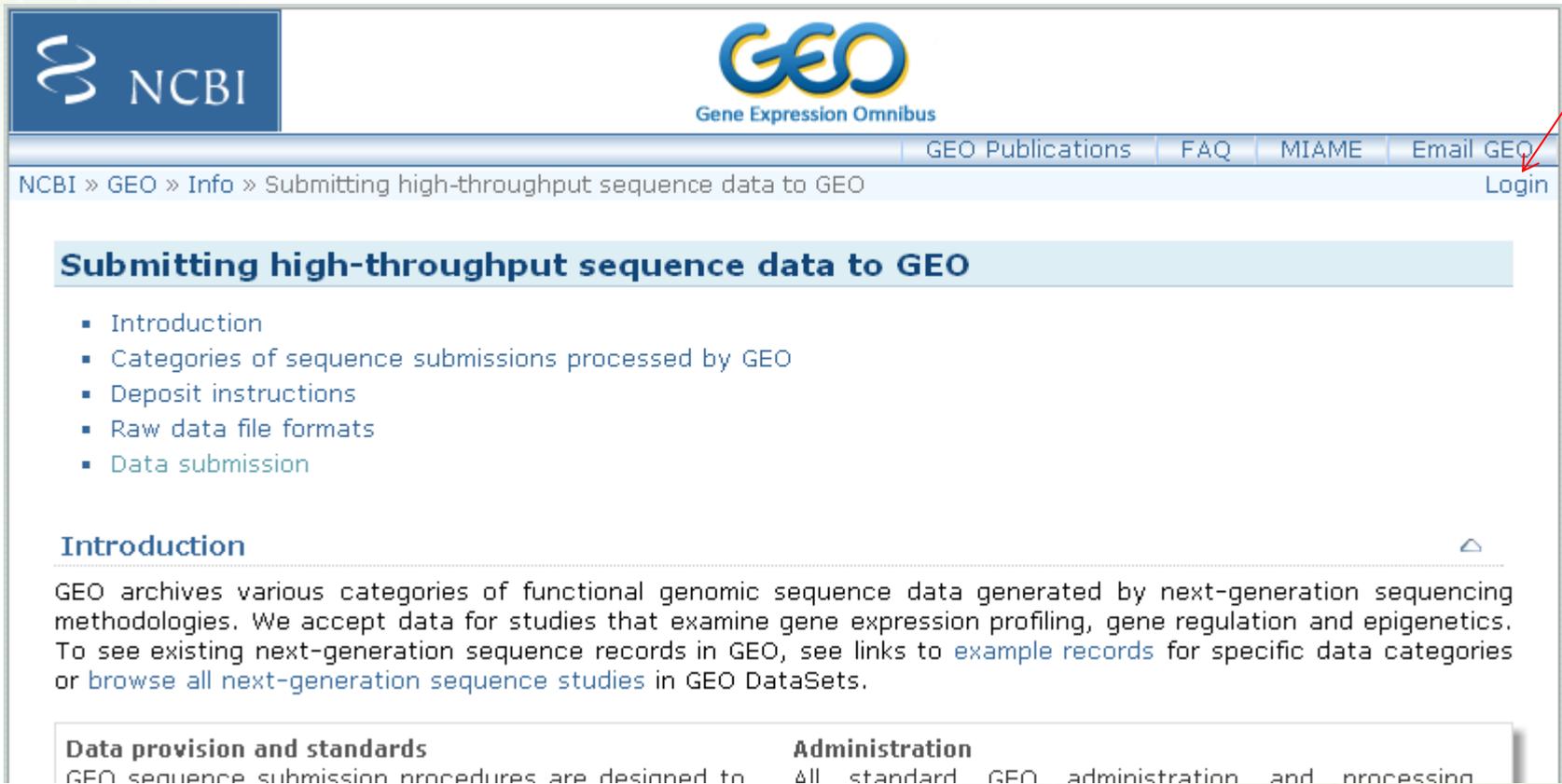
2018/08/08

高通量测序数据GEO上传步骤说明

1.1 上传网页:

<http://www.ncbi.nlm.nih.gov/geo/info/seq.html>

点击 **Login**



NCBI » GEO » Info » Submitting high-throughput sequence data to GEO

[GEO Publications](#) | [FAQ](#) | [MIAME](#) | [Email GEO](#) | [Login](#)

Submitting high-throughput sequence data to GEO

- [Introduction](#)
- [Categories of sequence submissions processed by GEO](#)
- [Deposit instructions](#)
- [Raw data file formats](#)
- [Data submission](#)

Introduction

GEO archives various categories of functional genomic sequence data generated by next-generation sequencing methodologies. We accept data for studies that examine gene expression profiling, gene regulation and epigenetics. To see existing next-generation sequence records in GEO, see links to [example records](#) for specific data categories or [browse all next-generation sequence studies](#) in GEO DataSets.

Data provision and standards
GEO sequence submission procedures are designed to

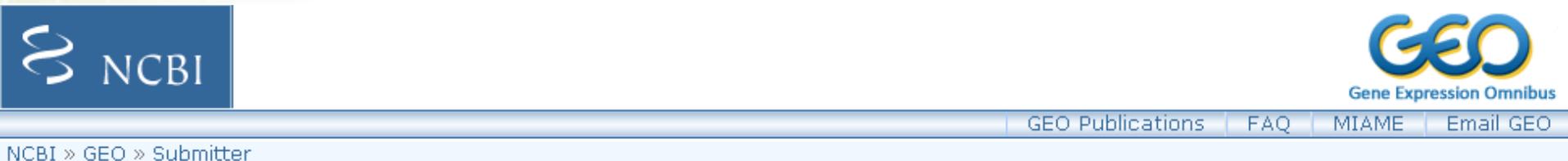
Administration
All standard GEO administration and processing

高通量测序数据GEO上传步骤说明

1.2 进入 <http://www.ncbi.nlm.nih.gov/geo/submitter/>

所有的GEO提交都是通过NCBI来实现的。

如果没有NCBI账户，请点击左边的箭头创建NCBI账户；否则就点击右边的箭头进入GEO登录页面。之后按照最下面的图进入GEO的主页。



Submitter login

You only need to log in if you want to submit data or update existing submissions. Browsing or downloading GEO's public holdings is unrestricted and does not require an account.

Effective May 1st, 2013, all GEO submitters must log in through My NCBI. If you already have a My NCBI account, please [login](#). Otherwise, please create a My NCBI account [here](#).

IMPORTANT: After logging in to My NCBI you will be given the option to link that account with your legacy GEO account and thereby gain access to existing submissions.



Submitter login

Investigator

Use this section to provide details about the primary investigator. This information will be

Organization name*

Submitter (Account manager)

If the person responsible for submitting the data to GEO is different from the

1.3 进入GEO主页 <http://www.ncbi.nlm.nih.gov/geo/> 点击: [Submission Guidelines](#)

NCBI Resources How To kangcheng My NCBI Sign Out
GEO Home Documentation Query & Browse Email GEO My GEO Submissions

Gene Expression Omnibus

GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. Array- and sequence-based data are accepted. Tools are provided to help users query and download experiments and curated gene expression profiles.

Keyword or GEO Accession **Search**

Getting Started

- Overview
- FAQ
- About GEO DataSets
- About GEO Profiles
- About GEO2R Analysis
- How to Construct a Query
- How to Download Data

Tools

- Search for Studies at GEO DataSets
- Search for Gene Expression at GEO Profiles
- Search GEO Documentation
- Analyze a Study with GEO2R
- GEO BLAST
- Programmatic Access
- FTP Site

Browse Content

Repository Browser

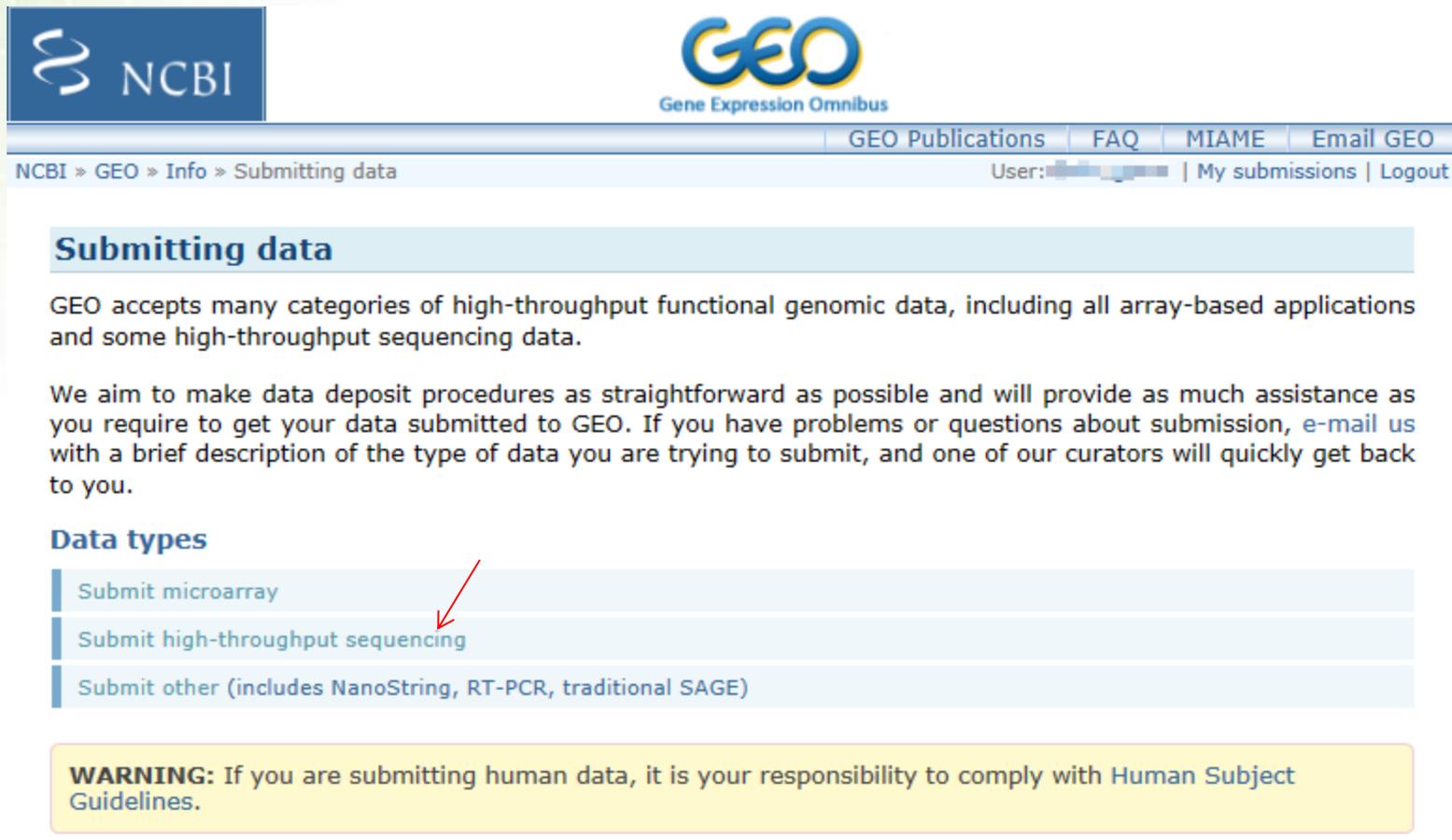
| | |
|------------|---------|
| DataSets: | 3413 |
| Series: | 47499 |
| Platforms: | 12934 |
| Samples: | 1134860 |

Information for Submitters

| | | |
|--------------------|---------------------------------------|---------------------------|
| My GEO Submissions | Submission Guidelines | MIAME Standards |
| My GEO Profile | Update Guidelines | Citing and Linking to GEO |

高通量测序数据GEO上传步骤说明

1.4 点击[High-throughput sequence submissions](#)



NCBI > GEO > Info > Submitting data

User: [redacted] | My submissions | Logout

Submitting data

GEO accepts many categories of high-throughput functional genomic data, including all array-based applications and some high-throughput sequencing data.

We aim to make data deposit procedures as straightforward as possible and will provide as much assistance as you require to get your data submitted to GEO. If you have problems or questions about submission, [e-mail us](#) with a brief description of the type of data you are trying to submit, and one of our curators will quickly get back to you.

Data types

- Submit microarray
- Submit high-throughput sequencing
- Submit other (includes NanoString, RT-PCR, traditional SAGE)

WARNING: If you are submitting human data, it is your responsibility to comply with Human Subject Guidelines.

高通量测序数据GEO上传步骤说明

1.5 进入GEO提交网页 <http://www.ncbi.nlm.nih.gov/geo/info/seq.html>



The screenshot shows the NCBI GEO website. The top navigation bar includes links for 'GEO Publications', 'FAQ', 'MIAME', and 'Email GEO'. The breadcrumb trail reads 'NCBI > GEO > Info > Submitting high-throughput sequence data to GEO'. The user is logged in as 'User: [redacted] | My submissions | Logout'. The main content area is titled 'Submitting high-throughput sequence data to GEO' and contains a list of links:

- Assembling your submission
 - Metadata spreadsheet
 - Processed data files
 - Raw data files
- Uploading your submission
- General Information
 - Data provisions, standards and administration
 - Categories of sequence submissions accepted by GEO

A red arrow points to the 'Raw data files' link in the 'Assembling your submission' section.

Uploading your submission



There are two steps for submission:

1. Transfer your files to the GEO FTP server. On your computer, create a folder named using your GEO username (**/kckangchen**) which includes all required submission files. Transfer the folder using instructions at [GEO File Transfer Protocol \(FTP\)](#). Do not transfer files unless you are confident that you have a submission that includes all required components (raw data files, processed data files and metadata spreadsheet).
2. Notify us **after** the FTP upload is complete using the [Submit to GEO](#) web form.

Submit

高通量测序数据GEO上传所需内容

| | |
|--|--|
| Metadata spreadsheet(元数据表单) | 实验设计和各个样本的描述信息,以及各步骤的 protocols 内容, processed 和 raw data file names 信息 |
| Processed data files (处理过程中的数据) | 与论文中的结论相关的数据, 表达谱数据如 mRNA和miRNA放置raw和normalized数据, 比如miRNA-seq得到的TPM, mRNA-seq中的 Cufflinks和Cuffdiff处理得到的结果。ChIP-seq和MEDIP-seq放置tag density 文件和peak文件, 比如bed, wig, tab分隔的txt文档等等。 |
| Raw data files (原始数据) | 测序平台产生的序列和质量值数据(*.gz) |

高通量测序数据GEO上传步骤说明

2. 上传准备

2.1 建立以你的用户名为文件名的文件夹：比如
kangchen

2.2 把做好的Metadata spreadsheet, Processed data files, Raw data files放进去。其中Raw data files放置原始的以gz压缩好的fastq文件。

高通量测序数据GEO上传步骤说明

元数据表单模板

Metadata spreadsheet

Download metadata spreadsheet (template and examples)

Metadata refers to descriptive information about the overall study, individual samples, all protocols, and references to processed and raw data file names. Information is supplied by completing all fields of a metadata template spreadsheet. Guidelines on the content of each field are provided within the spreadsheet.

Metadata spreadsheet.xls [兼容模式] - Microsoft Excel

开始 插入 页面布局 公式 数据 审阅 视图 帮助

剪贴板 字体 对齐方式 数字 样式 单元格 编辑

A19

| | | | | | | |
|----|--|---|--------------------|-----------------|--------------------------------|--------------------------------|
| 1 | # High-throughput sequencing metadata template (version 2.1). | | | | | |
| 2 | # All fields in this template must be completed. | | | | | |
| 3 | # Templates containing example data are found in the METADATA EXAMPLES spreadsheet tabs at the foot of this page. | | | | | |
| 4 | # Field names (in blue on this page) should not be edited. Hover over cells containing field names to view field content guidelines. | | | | | |
| 5 | # Human data. If there are patient privacy concerns regarding making data fully public through GEO, please submit to NCBI's dbGaP (http://www.ncbi.nlm.nih.gov/gap/) database. dbGaP has cont | | | | | |
| 6 | | | | | | |
| 7 | SERIES | | | | | |
| 8 | # This section describes the overall experiment. | | | | | |
| 9 | title | Next Generation Sequencing Facilitates Quantitative Analysis of Wild Type and Nrl-/- Retinal Transcriptomes | | | | |
| 10 | summary | Purpose: Next-generation sequencing (NGS) has revolutionized systems-based analysis of cellular pathways. The goals of this study are to compare NGS-derived | | | | |
| 11 | summary | Methods: Retinal mRNA profiles of 21-day-old wild-type (WT) and neural retina leucine zipper knockout (Nrl-/-) mice were generated by deep sequencing, in triplic | | | | |
| 12 | summary | Results: Using an optimized data analysis workflow, we mapped about 30 million sequence reads per sample to the mouse genome (build mm9) and identified 16,0 | | | | |
| 13 | summary | Conclusions: Our study represents the first detailed analysis of retinal transcriptomes, with biologic replicates, generated by RNA-seq technology. The optimized d | | | | |
| 14 | overall design | Retinal mRNA profiles of 21-day old wild type (WT) and Nrl-/- mice were generated by deep sequencing, in triplicate, using Illumina GAIIx. | | | | |
| 15 | contributor | Rebecca A, Smith | | | | |
| 16 | contributor | David, Doe | | | | |
| 17 | supplementary file | | | | | |
| 18 | SRA center name code | | | | | |
| 19 | | | | | | |
| 20 | SAMPLES | | | | | |
| 21 | # This section lists and describes each of the biological Samples under investigation, as well as any protocols that are specific to individual Samples. | | | | | |
| 22 | # Additional "processed data file" or "raw file" columns may be included. | | | | | |
| 23 | Sample name | title | source name | organism | characteristics: strain | characteristics: tissue |
| 24 | Sample 1 | WT rep1 | Retina | Mus musculus | C57BL/6 | retina |
| 25 | Sample 2 | WT rep2 | Retina | Mus musculus | C57BL/6 | retina |
| 26 | Sample 3 | Nrl-KO rep1 | Retina | Mus musculus | C57BL/6 | retina |
| 27 | Sample 4 | Nrl-KO rep2 | Retina | Mus musculus | C57BL/6 | retina |
| 28 | | | | | | |
| 29 | PROTOCOLS | | | | | |

高通量测序数据GEO上传步骤说明

3. 通过FTP上传

3.1 下载FileZila

- **FTP instructions:**

- Files may be transferred by many methods. Here are the ones we recommend:

- **Windows and Mac OS X:** we recommend the free client software, [FileZilla](#).

1. Connect using the following FTP login information:

host ftp-private.ncbi.nlm.nih.gov
Please use the 'fasp' directory.

username D0gDAzr0va
password D0gDAzr0va

2. Drag-n-drop directory or file(s) into the /fasp directory on the FTP server. When transferring multiple files please drop the files into a directory that includes your GEO username.

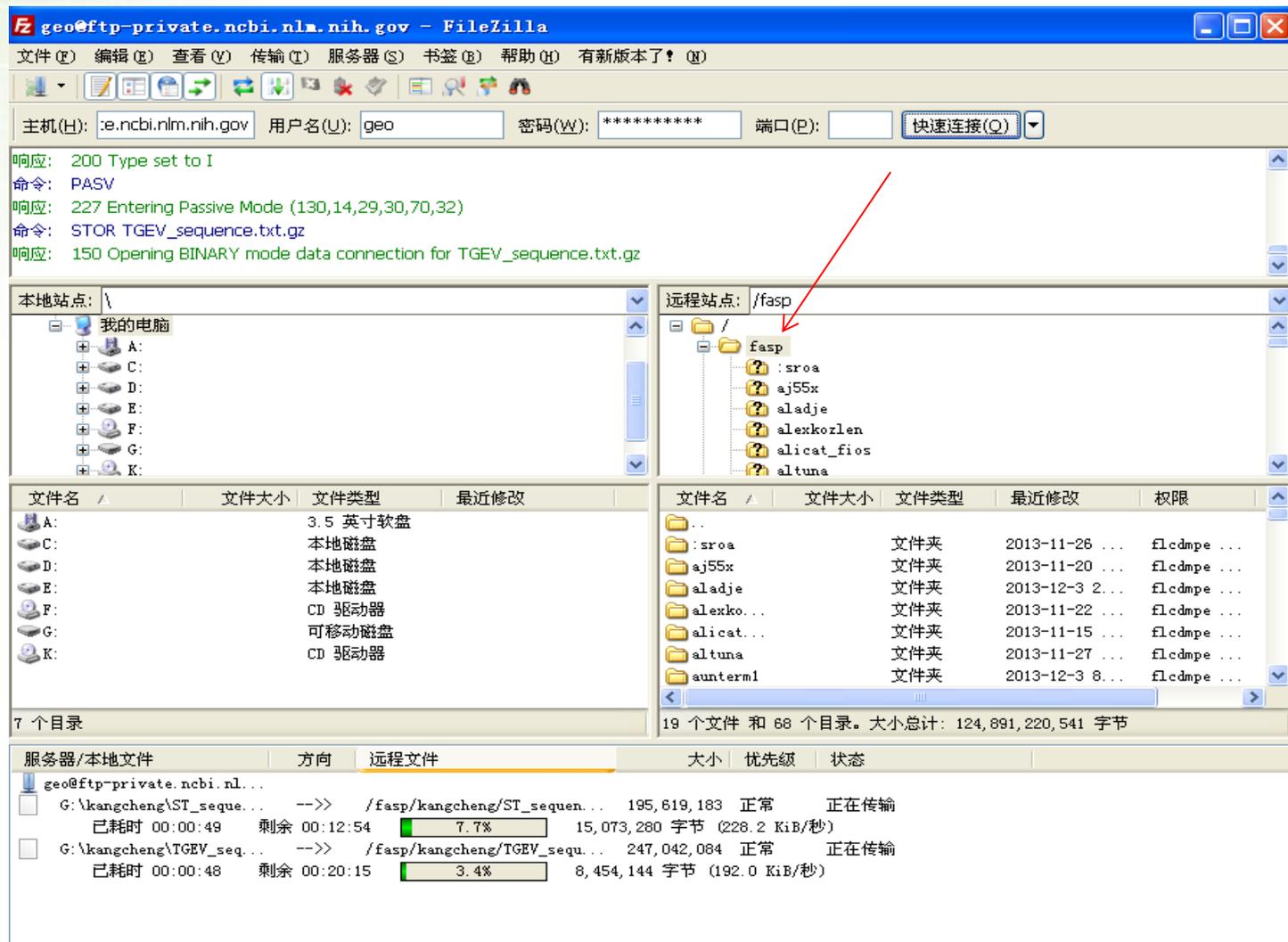
- **Linux/Unix:** we recommend that you try 'ncftp'. Optimized settings are detailed in this [README](#) file.

Here is a typical 'ncftp' session:

1. Connect to the server:
ncftp ftp://geo:D0gDAzr0va@ftp-private.ncbi.nih.gov/fasp
2. Set buffer size (recommended for faster transfer):
set so-bufsize 33554432
3. Transfer an entire directory (named using your GEO username) plus content using:
put -R GEOusername_directory

高通量测序数据GEO上传步骤说明

3.2 根据上页信息登录FTP，然后将以你的用户名为文件夹名的文件夹拖入fasp，等待上传完毕



GEO上传数据验证流程

数据上传成功后，**一定要给GEO发一封邮件**，写明下图中的三点内容



GEO curator 开始处理递交的数据

Contact you



解释并修正数据中他们认为有问题的地方
5个工作日

• Email notification:

• After file transfer is complete, please **e-mail GEO** with the following information:

1. GEO account username (**kangcheng**);
2. Names of the directory and files deposited;
3. Public release date (required - up to 3 years from now - **FAQ**).

• **It is important to send us this e-mail notification because unannounced files will be removed from our FTP site without being processed. We do not send automated confirmation that files have been received. You should expect to receive an e-mail from a curator within 5 business days after you send us the notification (FAQ).**