

• 转录组测序结果

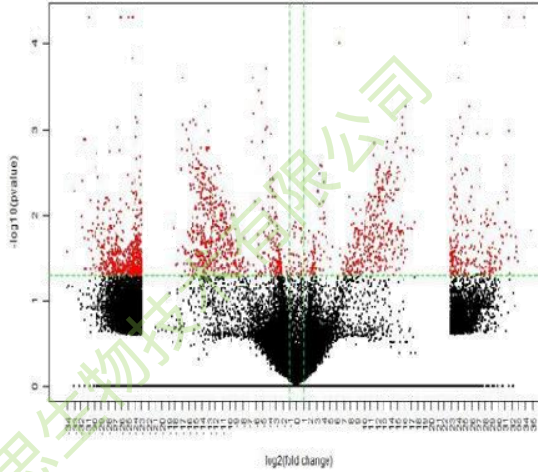


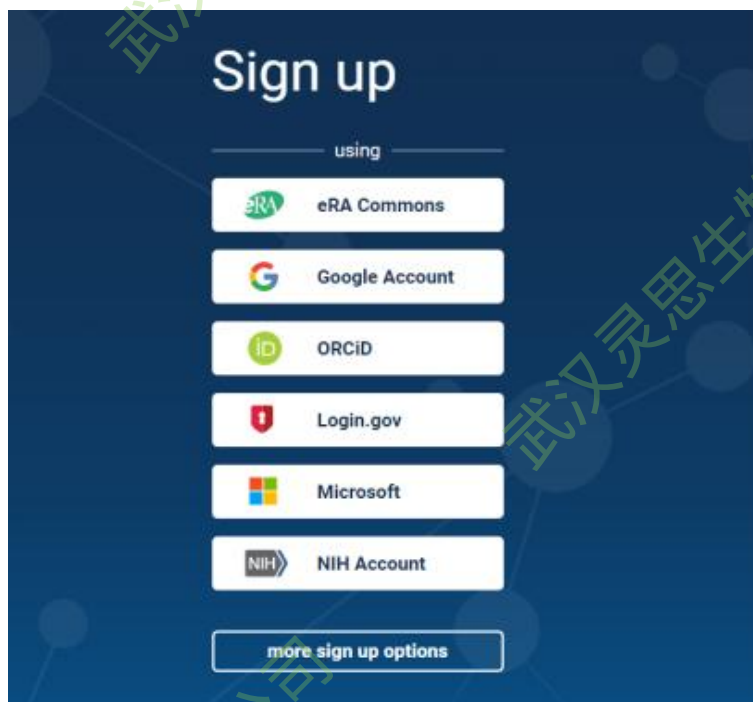
表 5.1 差异表达转录本筛选结果文件。

ter_id	gene_id	gene	locus	chr	strand	chr	start	end	diffho	Fold	log2(fold change)	test_stat	p_value	q_value	significant
ENS0000028200001440100			7:32727	+	strand	OE	41.3147	4.01020	-11.8088	-0.20274	0.06816	4.89327	0.0001	0.0001	1
ENS000002820000082571			2:16500	+	strand	OE	0.900281	41.1075	17.2626	0.01249	0.01476	4.89327	0.0001	0.0001	1
ENS000002820000082574			2:69601	+	strand	OE	1.14.842	23.3014	-2.70117	-2.22158	0.02205	4.89327	0.0001	0.0001	1
ENS000002820000081922			1:143047	+	strand	OE	2.51487	0.282276	-2.45962	-2.23269	0.0481	4.89327	0.0001	0.0001	1
ENS000002820000082570			6:117094	+	strand	OE	3.92043	0.00011	15.6225	0.07139	0.04255	4.89327	0.0001	0.0001	1
ENS000002820000081994			1:98041	+	strand	OE	9.84843	0.00787	-12.6669	-0.20287	0.04876	4.89327	0.0001	0.0001	1
ENS000002820000081977			2:48926	+	strand	OE	4.28293	0.00145	-0.71643	-0.13861	0.0282	4.89327	0.0001	0.0001	1
ENS000002820000082573			2:17134	+	strand	OE	0.90389	25.4662	12.8772	0.13439	0.0051	4.89327	0.0001	0.0001	1
ENS000002820000081921			9:42411	+	strand	OE	0.902575	6.32122	-11.6205	0.131289	0.0187	4.89327	0.0001	0.0001	1
ENS000002820000081976			9:45441	+	strand	OE	3.55724	0.00175	-10.3821	-0.1821	0.02725	4.89327	0.0001	0.0001	1
ENS000002820000081979			7:109461	+	strand	OE	7.7723	0.00687	-15.4662	-0.0138	0.00313	4.89327	0.0001	0.0001	1
ENS000002820000081978			9:14075	+	strand	OE	118.147	723.618	2.41464	2.3364	0.0176	4.89327	0.0001	0.0001	1
ENS000002820000081920			2:37465	+	strand	OE	185.837	12.9501	-3.78262	-3.29452	0.0447	4.89327	0.0001	0.0001	1
ENS000002820000081923			2:28864	+	strand	OE	15.3111	0.941094	-8.83276	-0.37447	0.0245	4.89327	0.0001	0.0001	1
ENS000002820000081924			2:28524	+	strand	OE	4.72633	0.00229	-11.8326	-0.27508	0.02965	4.89327	0.0001	0.0001	1
ENS000002820000081925			2:43061	+	strand	OE	3.14443		-23.7165	0.04585	4.89327	0.0001	0.0001	1	
ENS000002820000081926			2:28677	+	strand	OE	194.061	22.6422	-2.36579	-2.38963	0.02905	4.89327	0.0001	0.0001	1
ENS000002820000081927			2:41011	+	strand	OE	4.28293	0.00505	-5.7186	-0.20885	0.0172	4.89327	0.0001	0.0001	1
ENS000002820000081928			4:28427	+	strand	OE	15.2313	103.261	2.78189	2.50211	0.0281	4.89327	0.0001	0.0001	1
ENS000002820000081929			30:54974	+	strand	OE	6.16472	0.000155	-15.2759	-0.03994	0.03345	4.89327	0.0001	0.0001	1
ENS000002820000081930			30:28972	+	strand	OE	2.34912	0.000292	-12.9739	-0.06482	0.0074	4.89327	0.0001	0.0001	1
ENS000002820000081931			30:28972	+	strand	OE	0.900863	3.80638	12.1185	0.09494	0.003	4.89327	0.0001	0.0001	1
ENS000002820000081932			30:33.48	+	strand	OE	4.92633	0.001247	-11.6401	-0.06862	0.0027	4.89327	0.0001	0.0001	1
ENS000002820000081933			8:27192	+	strand	OE	198.231	2.39416	-0.71711	-5.99977	0.00025	4.89327	0.0001	0.0001	1
ENS000002820000081934			6:87178	+	strand	OE	0.90449	11.311	14.2644	0.00164	0.04875	4.89327	0.0001	0.0001	1

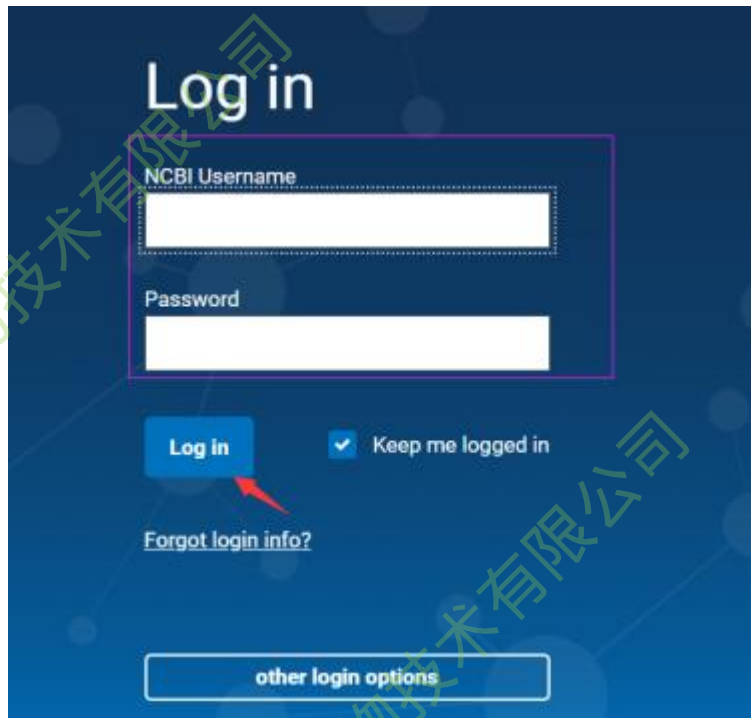
NCBI 数据上传：转录组测序 (RNA-Seq) 数据

一、注册及登录账号

(1) 注册账号：进入 <https://www.ncbi.nlm.nih.gov/> 网页，找到页面右上角 Sign in，进入页面后，点击页面中的 Sign up，根据自己喜好，选择对应的账号进行注册，按照提示填写账号、密码、邮箱等信息。



(2) 登录账号：输入账号及密码（注册时使用哪种账号注册，登录时选择对应的登录通道登录），点击 Log in，然后点击左上角的 NCBI 大图标回到 NCBI 的主页，点击图中 Submit 按钮进入提交数据页面。



二、生成 Biosample 编号

(1) 进入 Submit 界面之后，选择 My submissions。（提交数据前必须先把样品信息录入生成 Biosample 编号）



Type a few words about the sequence data you are submitting and select an option to learn more. You can also browse submission information below.

What do you want to submit?

Enter a few words about your sequence data.



Suggest tool

SARS-CoV-2

16S rRNA

genome

ITS

SRA

(2) 在下一个界面中选择 BioSample，进入新的页面，点击蓝色按钮 New submission。

Submission Portal

Your submissions

Start a new submission

- GenBank
- Sequence Read Archive
- Genome
- TSA
- BioProject
- BioSample
- Supplementary Files
- API

Submission Portal

BioSample

New submission

Download batch submission template



Note: to update an existing record or recent submission, please email your request.

Short description and brief instructions

(3) 进入如下页面，根据实际情况填写提交者的信息：包括姓名，邮箱（需要填写一个以单位后缀邮箱），学校学院，单位地址，填写完成点击 Continue。

Submission Portal

BioSample submission: SUB11603082

New

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 DESCRIPTION 6 REVIEW & SUBMIT

Submitter

* First (given) name Middle name * Last (family) name

* Email (primary) * Email (secondary)

1040531904@qq.com

At least one email should be from NCSL. Submitters with a qq.com, 163.com, or foxmail.com email address do not always receive mail from NCSL. Please provide an alternative email address from a different domain to ensure that we are able to communicate with you.

Group for this submission

No group (affiliation from my personal profile)

You can create a group for shared submissions in your profile.

* Submitting organization Submitting organization URL * Department

Phone Fax

(4) 接下来填写数据释放的时间，可选择立即释放，也可选择指定日期；下一个选项选择 batch/Multiple Biosamples，然后点击 Continue 即可。

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 DESCRIPTION 6 REVIEW & SUBMIT

General Information

Release date

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

* When should this submission be released to the public?

Release immediately following processing

Release on specified date or upon publication, whichever is first

* Projected release date

YYYY-MM-DD

* Specify if you are submitting a single sample or a file containing multiple samples

Batch/Multiple BioSamples

You will be asked to upload a tab-delimited text file that describes each of your samples and their attributes. Submission template files can be downloaded from the Attributes tab or the templates tab.

Single BioSample

You will be asked to manually complete a web form to describe one sample and its attributes.

(5) 接下来选择样品类型 Sample Type, 通常的转录组样品选择比较多: 原核转录组
可选前两个, 真核的根据样品可以选择: Pathogen affecting public health, Model orga
nism or animal sample, Plant sample 等等。然后点击最下方的 Continue。

laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.

- Metagenome or environmental**
Use for metagenomic and environmental samples when it is not appropriate or advantageous to use [MixS](#) packages.
- Invertebrate**
Use for any invertebrate sample.
- Human**
WARNING: Only use for human samples or cell lines that have no privacy concerns. For all studies involving human subjects, it is the submitter's responsibility to ensure that the information supplied protects participant privacy in accordance with all applicable laws, regulations and institutional policies. Make sure to remove any direct personal identifiers from your submission. If there are patient privacy concerns regarding making data fully public, please submit samples and data to NCBI's [dbGaP](#) database. [dbGaP](#) has controlled access mechanisms and is an appropriate resource for hosting sensitive patient data. For samples isolated from humans use the Pathogen, Microbe or appropriate [MixS](#) package.
- Plant**
Use for any plant sample or cell line.
- Viral**
Use for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.
- Beta-lactamase**
Use for beta-lactamase gene transformants that have sequence and antibiotic resistance data. Please use the 'Supplementary Files' wizard to submit corresponding Sequin and Antibigram files.

28S rRNA or COI obtained from cultured or voucher-identifiable specimens. Organism cannot contain the term 'metagenome'.

- MIMARKS Survey related**
Use for any type of marker gene sequences, eg, 16S, 18S, 23S, 28S rRNA or COI obtained directly from the environment, without culturing or identification of the organisms. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.
- MIMS Environmental/Metagenome**
Use for environmental and metagenome sequences. Organism must be a metagenome, where lineage starts with [unclassified sequences](#) and scientific name ends with 'metagenome'.
- MISAG Single Amplified Genome**
Use for single amplified genome sequences produced by isolating individual cells, amplifying the genome of each cell using whole genome amplification, and then sequencing the amplified DNA. Organism cannot contain the term 'metagenome'.
- MIUVIG Uncultivated Virus Genome**
Use for uncultivated virus genome identified in metagenome and metatranscriptome datasets. Organism must have lineage [Viruses](#).

Continue

(6) 接下来填写样品信息 Attributes, 在线填写和上传 excel 文件都可, 小编习惯下载他们的模板填写后上传的方式: 点击第二个选项, 然后 Download Excel。

BioSample submission: SUB11603082
Plant sample

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 REVIEW & SUBMIT

Attributes
Package Plant, version 1.0

★ How do you want to provide your BioSample attributes?

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited) that includes the attributes for each of your BioSamples

★ Attributes file

or drag and drop it here

- Template for BioSample package Plant, version 1.0
Download Excel or download TSV
For column explanations and examples, please see the sample attributes page.
For more information, please see creating sample attribute file.

Continue

(7) 打开该 Excel，文件如下，以本次提交的 3 个样品为例，其中绿色的栏目为必填项，如果不知道具体信息就填写“not collected”，填的越详细 NCBI 审核的速度会越快，如果都填“not collected”也可以，不过 NCBI 反馈给你 Biosample 编号会慢一些，大约需要 2-3 天，样品名称建议使用唯一的名称。上传之后就可以点击 Continue 进入下一个关卡了。



This is a submission template for batch deposit of "Plant" version 1.0 samples to the NCBI Biosample database (<https://www.ncbi.nlm.nih.gov/biosample/>).
Some fields are required, but optional fields are not required. If information is available for any required field, please fill it in. If you are unsure of the correct value, please contact us at submit@ncbi.nlm.nih.gov.
All fields are optional, leave optional fields empty or delete them if no information is available.
You can add any number of custom fields to fully describe your Biosamples, simply include them in the table.
More user fields are to view definition, as see <https://www.ncbi.nlm.nih.gov/biosample/docs/attributes/>.
CAUTION: In some tabs Excel may automatically apply formatting to your data. In particular, take care with dates, increasing serials and special characters.
TO MAKE A SUBMISSION
1. Complete this template table.
2. Upload the file on the "Attributes" tab of the Biosample Submission Portal at <https://submit.ncbi.nlm.nih.gov/biosample/>.
If you have any questions, please contact us at submit@ncbi.nlm.nih.gov.

sample_name	sample_title	bioproject_accession	Purpose	isolate	multitier	sextype	age	dev_stage	sex	loc_name	strains	hit
ABC1			BACE	not collect	not collect	not collect	not collect	not collect	not collect	not collect	root	
ABC2			BACE	not collect	not collect	not collect	not collect	not collect	not collect	not collect	root	
ABC3			BACE	not collect	not collect	not collect	not collect	not collect	not collect	not collect	root	

(8) 接下来就是生成 BioSample 的最后一步，确认信息后点 Submit 即可。

BioSample submission: SUB11603082

Plant sample

1 SUBMITTER 2 GENERAL INFO 3 SAMPLE TYPE 4 ATTRIBUTES 5 REVIEW & SUBMIT

Review & Submit

This BioSample submission will be released on **2023-07-20** or upon publication, whichever is first.

Note: Release of BioProject or BioSample is also triggered by the release of linked data.

Submitter Information

Submitter

General Information

Package

Attribute file

(9) 最后就是坐等 NCBI 的邮件，第一封邮件是点击 Submit 后自动发送的，可以忽略。我们需要等的邮件是包含 Biosample 编号的邮件，如下图所示。这封邮件里会详细说明您的每个样品跟 Biosample accession 号的对应关系，后续上传 raw data 需要用到。

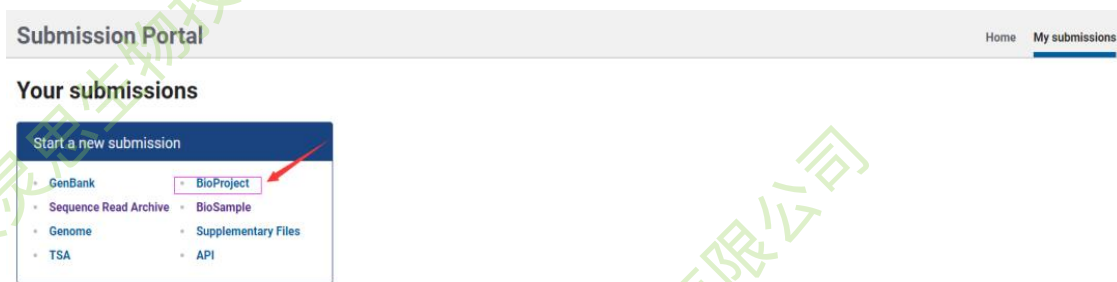
Dear [REDACTED]

This is an automatic acknowledgment that your recent submission to the BioSample database has been successfully processed.

BioSample accessions: [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED], [REDACTED]
Temporary SubmissionID: SUB06-[REDACTED]
Release date: [REDACTED] as soon as processing is complete.

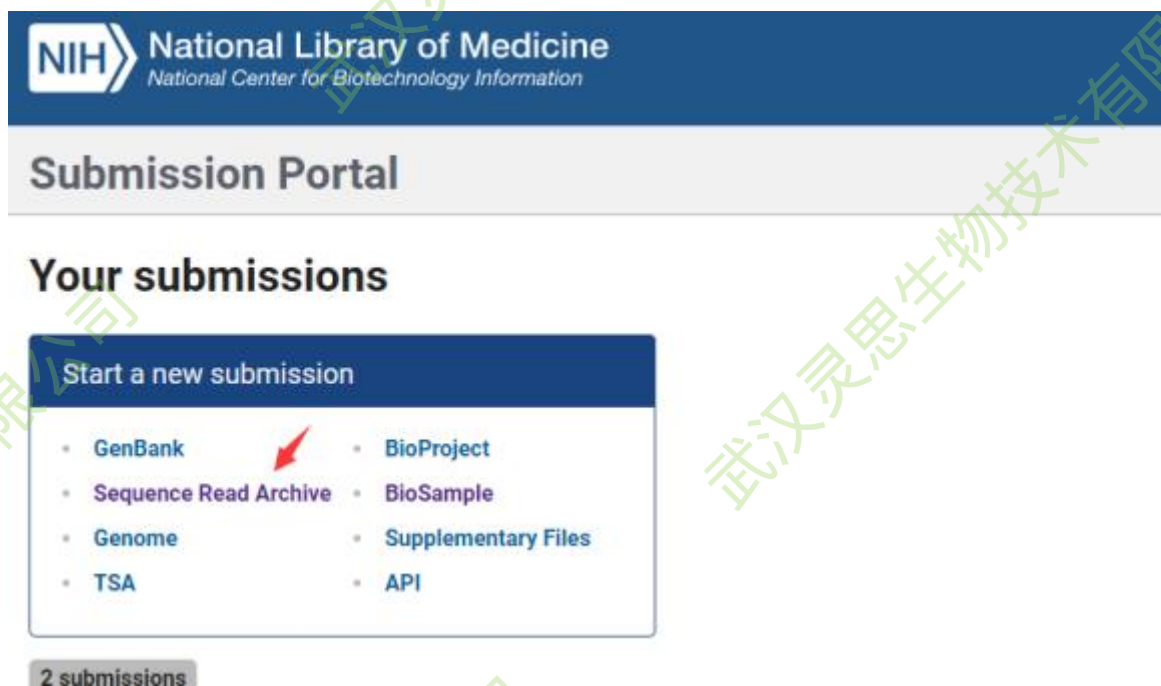
A submission summary and the links by which your BioSample records will be accessible are appended [REDACTED]

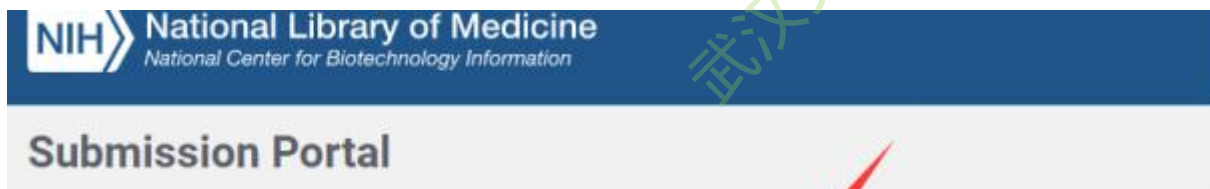
(10) Bioproject 编号申请与 BioSample 编号申请步骤一致，在以下页面点击 Bioproject，按照提示填写相关信息即可。



三、raw data 上传

(1) 待收到 BioSample 编号之后，我们要再次登录 NCBI，进入 Submit 界面，这次我们选择 Sequence Read Archive(SRA)，然后点击 New submission 进入最后的数据上传关卡。





Sequence Read Archive (SRA)

New submission

The SRA accepts genetic data and the associated quality scores produced by next generation sequencing technologies. Please refer to the [File Format Guide](#).

- Files can be compressed using **gzip** or **bzip2**, and may be submitted in a tar archive but archiving

(2) 填写个人信息。

(3) 第二步填写通用信息，如果提前注册好了 Bioproject 号。Biosample 这里选择 yes，Release date 根据需要选择即可。点击 Continue。

(4) 接下来，根据实际项目内容填写项目的摘要，简单的几个词描述一下就可以了。后边的几个选项没有特殊情况的话第一个选择 No，其他的不填即可。然后 Continue。

Sequence Read Archive (SRA) submission: SUB11603118

New

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 SRA METADATA 5 FILES 6 REVIEW & SUBMIT

Project Info

★ Project title

★ Public description

Relevance

★ Is your project part of a larger initiative which is already registered with NCBI?
 No Yes (not very common)

(5) 跟 Biosample 差不多，也需要填写一个表格，在线填写和下载模板填写再上传一样的，小编以
下载 Excel 表格填写为例。

Sequence Read Archive (SRA) submission: SUB11603118

123, Jun 13 '22

1 SUBMITTER 2 GENERAL INFO 3 PROJECT INFO 4 SRA METADATA 5 FILES 6 REVIEW & SUBMIT

SRA Metadata

For more detailed help with SRA submission please read the SRA Submission Wizard Help.

★ How do you want to provide your metadata?

- Use built-in table editor
- Upload a file using Excel or text format (tab-delimited)

★ Metadata file

or drag and drop it here

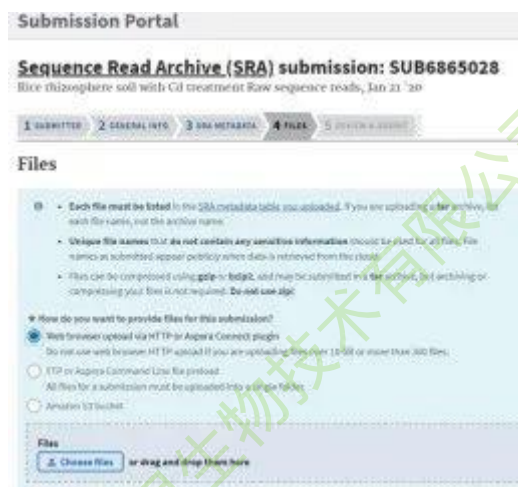
Download Excel spreadsheet (designed to make it easier to select the correct metadata values), edit, save and then upload the modified Excel file.

(6) 这个表格是这个样子的：biosample accession 编号就是邮件里的编号，library_ID 这里可以自行命名只要不重复即可。剩下的选项如图所示，通常情况是这样填写，在 filename 那里，请各位同学一定要保证跟后续要上传的文件名完全一致，并且不能有“.”、“-”等奇奇怪怪的符号，如果有可以改成下

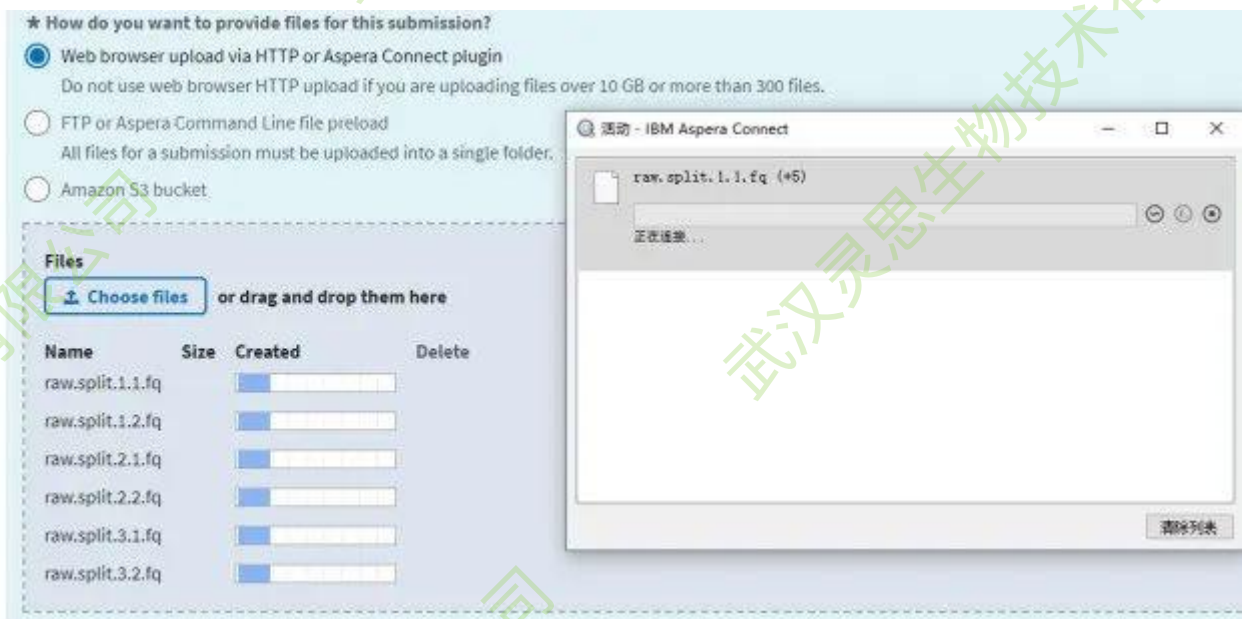
划线“_”或者删掉。转录组的数据选项依次是：library strategy: RNA-seq; library source: Transcriptome; library selection; layout: paired。（至于后边的测序仪型号什么的可以咨询给您做项目的销售）。

sample_access	library_ID	title	library_strategy	library_source	library_selection	library_layout	platform	instrument_model	design_descipt	filetype
D01	abc	abc	RNA-Seq	TRANSCRIPTOMIC	Paired	paired	IONSEQ	IONSEQ-T7		fastq
D02	abc	abc	RNA-Seq	TRANSCRIPTOMIC	Paired	paired	IONSEQ	IONSEQ-T7		fastq
D03	abc	abc	RNA-Seq	TRANSCRIPTOMIC	Paired	paired	IONSEQ	IONSEQ-T7		fastq

(7) 接下来就是最后一步了，上传数据。在这里，我们可以安装一个 IBM 的软件——Aspera Connect 用于大数据传输（下载地址：<https://downloads.asperasoft.com/en/downloads/8?list>，下载安装即可，推荐使用 360 浏览器，与 Aspera 关联比较方便。



(8) 选择图中第一个选项，在 choose files 选择你的所有文件，网站会自动使用 Aspera 帮您上传文件，速度很快。出现下图的界面就说明正在上传中了。



(9) 数据上传完后，点击 Continue 即可。最后一步，核查相关信息，无问题后，点击提交，等待 N CBI 审核。审核通过后会收到相关邮件。

友情提示

转录组测序的 raw data 通常较大，如果样品个数少的话可以通过第一个选项 Web browser upload via HTTP or Aspera Connect plugin。如果是 10 个样品以上，推荐用第二个选项 Aspera 命令行上传（当然用第一个选项上传也是可以的，不过需要一定的耐心）。

咨询电话：18171096640

邮箱：market@lingsibio.cn

网址：www.lingsibio.cn

公司地址：武汉市东湖新技术开发区花城大道 8 号光谷智慧健康园 D2 栋 4 层