A tutorial of data submission to GEO

NTU Center of Genomic and Precision Medicine

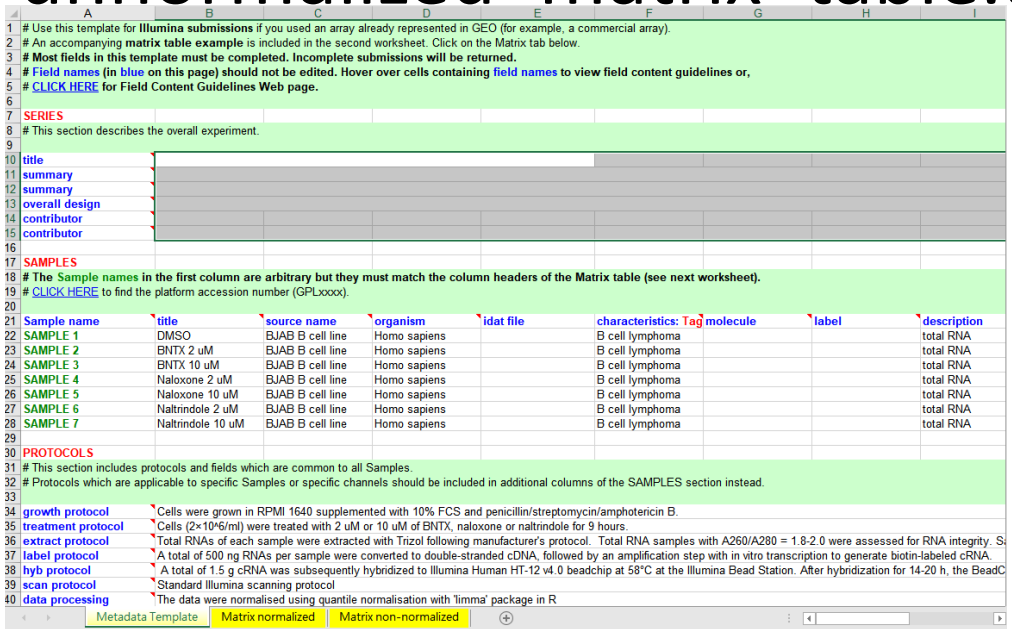
Yi-Wen Hsiao

Date: 2017.06.16

For illumina array data

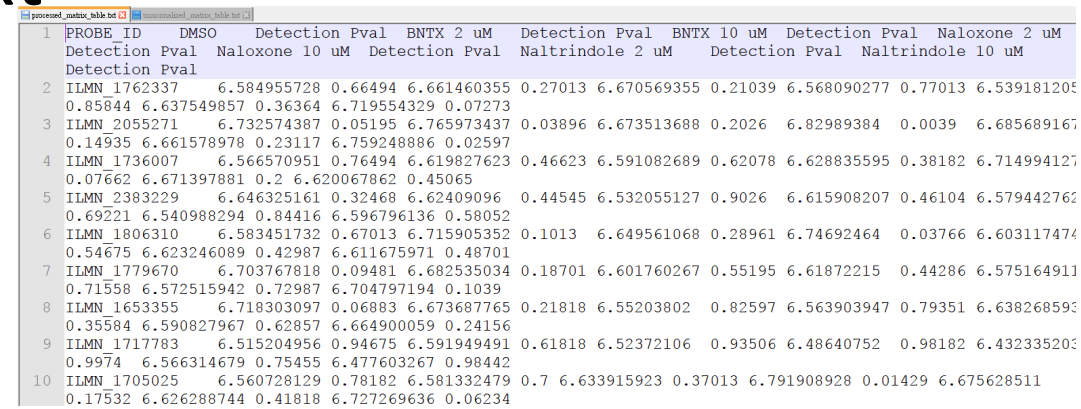
A XXXX.rar

- Includes:
  - GA_illumina_expression_metasheet.xls
  - processed_matrix_table.txt
  - unnormalized_matrix_table.txt



	A	B	C	D	E	F	G	H	I
1	# Use this template for **Illumina submissions** if you used an array already represented in GEO (for example, a commercial array).								
2	# An accompanying **matrix table example** is included in the second worksheet. Click on the Matrix tab below.								
3	# **Most fields in this template must be completed. Incomplete submissions will be returned.**								
4	# **Field names (in blue on this page) should not be edited. Hover over cells containing field names to view field content guidelines or,**								
5	# CLICK HERE **for Field Content Guidelines Web page.**								
6									
7	SERIES								
8	# This section describes the overall experiment.								
9									
10	title								
11	summary								
12	summary								
13	overall design								
14	contributor								
15	contributor								
16									
17	SAMPLES								
18	**# The Sample names in the first column are arbitrary but they must match the column headers of the Matrix table (see next worksheet).**								
19	# CLICK HERE to find the platform accession number (GPLxxxx).								
20									
21	Sample name	title	source name	organism	idat file	characteristics: Tag	molecule	label	description
22	SAMPLE 1	DMSO	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
23	SAMPLE 2	BNTX 2 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
24	SAMPLE 3	BNTX 10 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
25	SAMPLE 4	Naloxone 2 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
26	SAMPLE 5	Naloxone 10 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
27	SAMPLE 6	Naltrindole 2 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
28	SAMPLE 7	Naltrindole 10 uM	BJAB B cell line	Homo sapiens		B cell lymphoma			total RNA
29									
30	PROTOCOLS								
31	# This section includes protocols and fields which are common to all Samples.								
32	# Protocols which are applicable to specific Samples or specific channels should be included in additional columns of the SAMPLES section instead.								
33									
34	growth protocol	Cells were grown in RPMI 1640 supplemented with 10% FCS and penicillin/streptomycin/amphotericin B.							
35	treatment protocol	Cells (2×10^6/ml) were treated with 2 uM or 10 uM of BNTX, naloxone or naltrindole for 9 hours.							
36	extract protocol	Total RNAs of each sample were extracted with Trizol following manufacturer's protocol. Total RNA samples with A260/A280 = 1.8-2.0 were assessed for RNA integrity. S							
37	label protocol	A total of 500 ng RNAs per sample were converted to double-stranded cDNA, followed by an amplification step with in vitro transcription to generate biotin-labeled cRNA.							
38	hyb protocol	A total of 1.5 g cRNA was subsequently hybridized to Illumina Human HT-12 v4.0 beadchip at 58°C at the Illumina Bead Station. After hybridization for 14-20 h, the BeadC							
39	scan protocol	Standard Illumina scanning protocol							
40	data processing	The data were normalised using quantile normalisation with 'limma' package in R							

Metadata Template | Matrix normalized | Matrix non-normalized

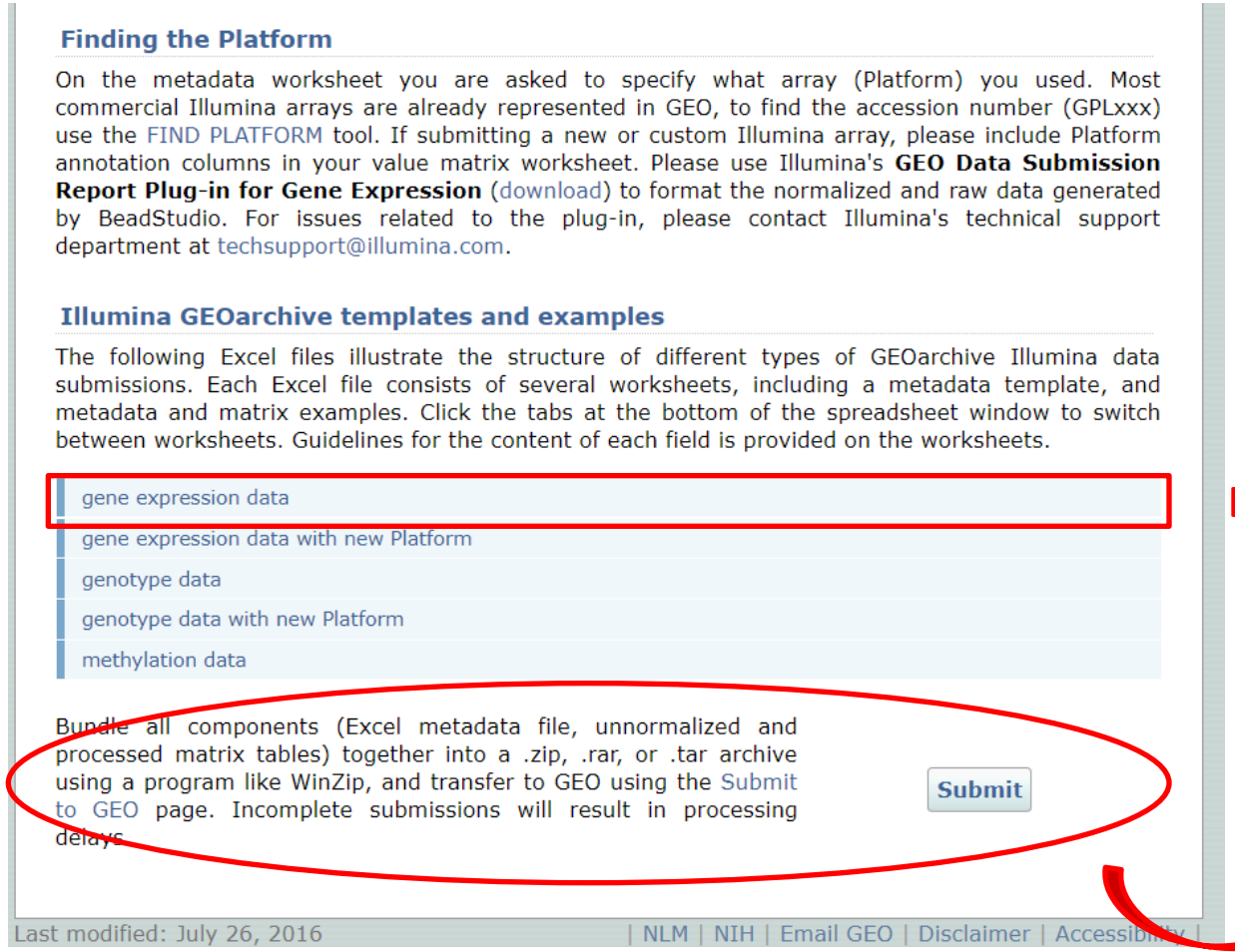


processed_matrix_table.txt | unnormalized_matrix_table.txt

```
1  PROBE_ID    DMSO    Detection Pval  BNTX 2 uM   Detection Pval  BNTX 10 uM  Detection Pval  Naloxone 2 uM
   Detection Pval  Naloxone 10 uM  Detection Pval  Naltrindole 2 uM    Detection Pval  Naltrindole 10 uM
   Detection Pval
2  ILMN_1762337    6.584955728 0.66494 6.661460355 0.27013 6.670569355 0.21039 6.568090277 0.77013 6.539181205
   0.85844 6.637549857 0.36364 6.719554329 0.07273
3  ILMN_2055271    6.732574387 0.05195 6.765973437 0.03896 6.673513688 0.2026  6.82989384  0.0039  6.685689167
   0.14935 6.661578978 0.23117 6.759248886 0.02597
4  ILMN_1736007    6.566570951 0.76494 6.619827623 0.46623 6.591082689 0.62078 6.628835595 0.38182 6.714994127
   0.07662 6.671397881 0.2 6.620067862 0.45065
5  ILMN_2383229    6.646325161 0.32468 6.62409096  0.44545 6.532055127 0.9026  6.615908207 0.46104 6.579442762
   0.69221 6.540988294 0.84416 6.596796136 0.58052
6  ILMN_1806310    6.583451732 0.67013 6.715905352 0.1013  6.649561068 0.28961 6.74692464  0.03766 6.603117474
   0.54675 6.623246089 0.42987 6.611675971 0.48701
7  ILMN_1779670    6.703767818 0.09481 6.682535034 0.18701 6.601760267 0.55195 6.61872215  0.44286 6.575164911
   0.71558 6.572515942 0.72987 6.704797194 0.1039
8  ILMN_1653355    6.718303097 0.06883 6.673687765 0.21818 6.55203802  0.82597 6.563903947 0.79351 6.638268593
   0.35584 6.590827967 0.62857 6.664900059 0.24156
9  ILMN_1717783    6.515204956 0.94675 6.591949491 0.61818 6.52372106  0.93506 6.48640752  0.98182 6.432335203
   0.9974  6.566314679 0.75455 6.477603267 0.98442
10 ILMN_1705025    6.560728129 0.78182 6.581332479 0.7 6.633915923 0.37013 6.791908928 0.01429 6.675628511
   0.17532 6.626288744 0.41818 6.727269636 0.06234
```

Submission page

- https://www.ncbi.nlm.nih.gov/geo/info/geo_illu.html



Finding the Platform

On the metadata worksheet you are asked to specify what array (Platform) you used. Most commercial Illumina arrays are already represented in GEO, to find the accession number (GPLxxx) use the FIND PLATFORM tool. If submitting a new or custom Illumina array, please include Platform annotation columns in your value matrix worksheet. Please use Illumina's **GEO Data Submission Report Plug-in for Gene Expression** (download) to format the normalized and raw data generated by BeadStudio. For issues related to the plug-in, please contact Illumina's technical support department at techsupport@illumina.com.

Illumina GEOarchive templates and examples

The following Excel files illustrate the structure of different types of GEOarchive Illumina data submissions. Each Excel file consists of several worksheets, including a metadata template, and metadata and matrix examples. Click the tabs at the bottom of the spreadsheet window to switch between worksheets. Guidelines for the content of each field is provided on the worksheets.

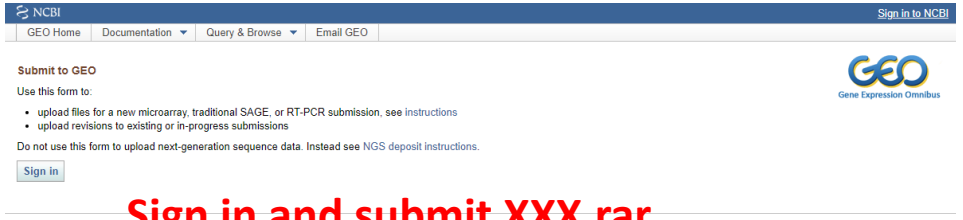
- gene expression data
- gene expression data with new Platform
- genotype data
- genotype data with new Platform
- methylation data

Bundle all components (Excel metadata file, unnormalized and processed matrix tables) together into a .zip, .rar, or .tar archive using a program like WinZip, and transfer to GEO using the Submit to GEO page. Incomplete submissions will result in processing delays.

Submit

Last modified: July 26, 2016 | NLM | NIH | Email GEO | Disclaimer | Accessibility |

Download metadata file format



Sign in and submit XXX.rar

Email Template
(To geo@ncbi.nlm.nih.gov)

Dear Sir,

We'd like to submit our Illumina HumanHT-12 V4.0 expression beadchip array result to your GEO website. We have uploaded our data file (XXXX.rar) to the GEO account name OOO. And we want to release our data on XX/XX/20XX.

XXXX.rar includes:

GA_illumina_expression_metasheet.xls

processed_matrix_table.txt

unnormalized_matrix_table.txt

Thank you for your assistance.

Best regards,

AAA

For NGS data (except metagenomics data)

Two main portals

- **SRA submission** (raw data)
  1. BioProject: to describe the goal of the sequencing study
  2. BioSample: to provide sample attributes
  3. SAR: to up-load raw data file (e.g. fastq, bam …)

- **GEO submission** (experimental description and processed data)

Three Main Materials (except for raw data)

- **Submitter information**: word file (for first time)
- **BioProject**: word file
- **BioSample**: excel file; provided by NCBI
- **GEO Metasheet**: excel file; provided by NCBI

[Submitter information]

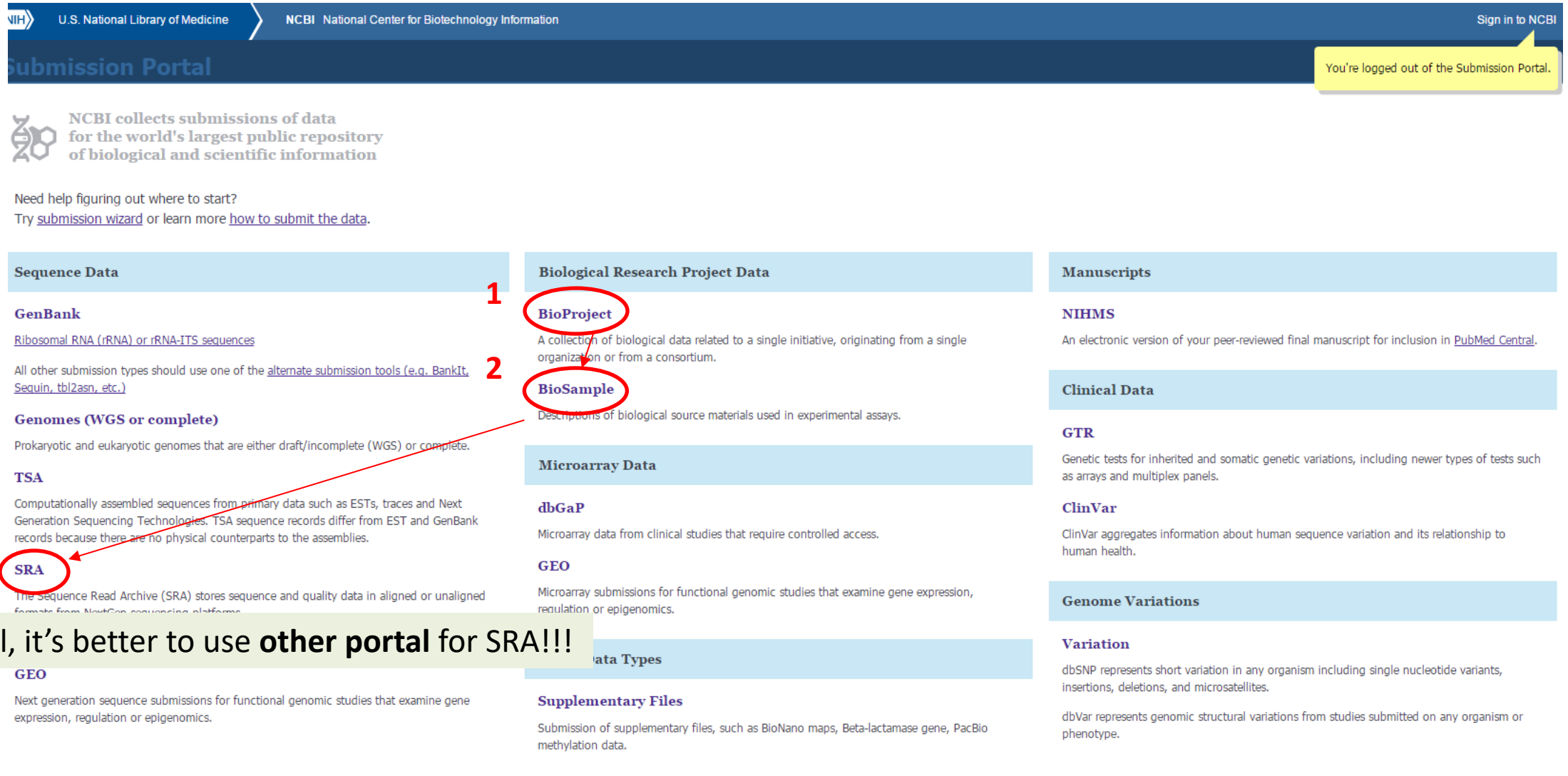
SUBMITTER INFO.
- First (given) name:
- Last (family) name:
- E-mail(primary):
- Submitting organization:
- Department:
- Street, City, Postal code & Country:

[BioProject]

- Project Type
  - Project data type:
  - Sample scope:
- Target
  - Organism name:
- General information (Project detail)
  - project title:
  - project description:
  - Relevance:
  - Release Date:

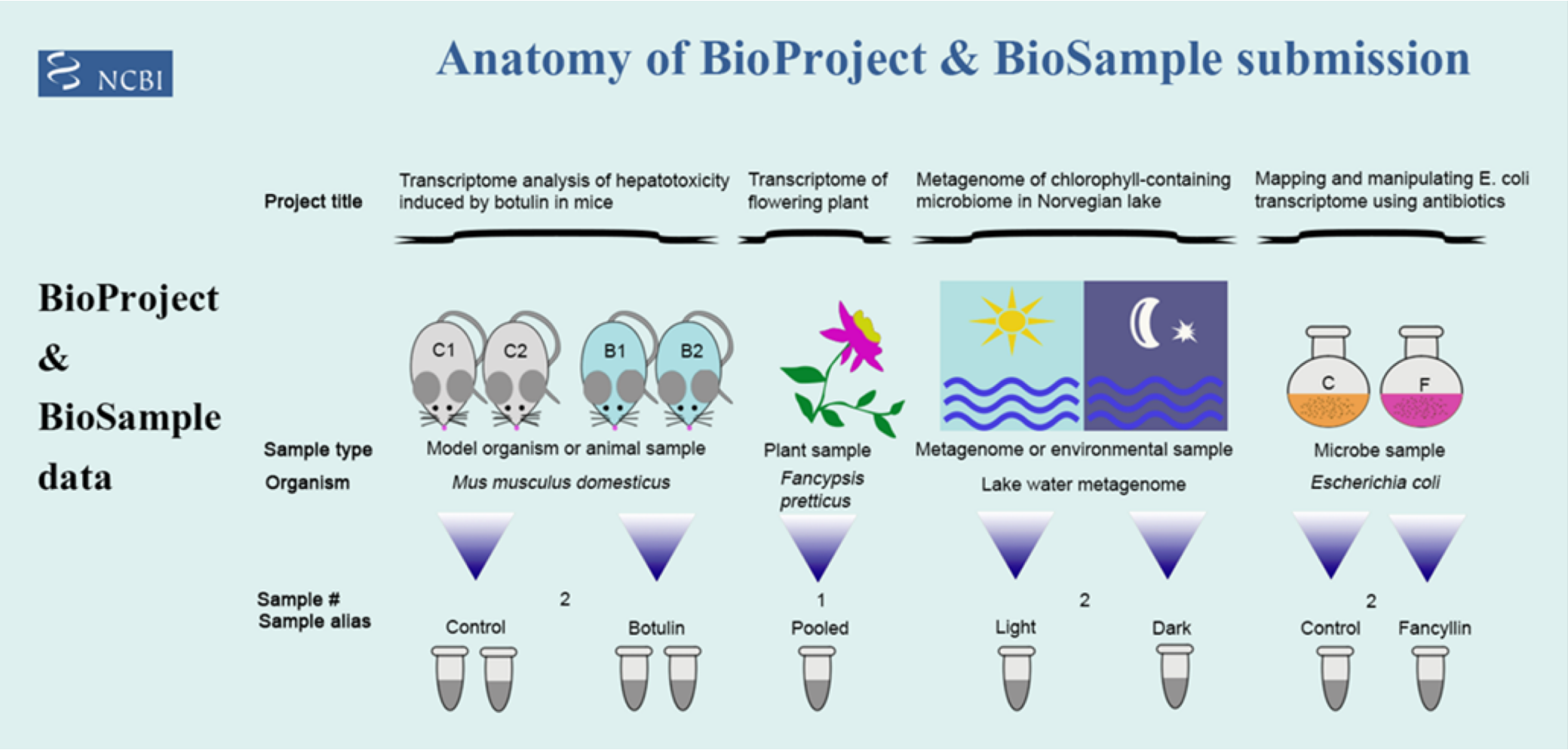
I. SRA submission

Submission Portal: https://submit.ncbi.nlm.nih.gov/



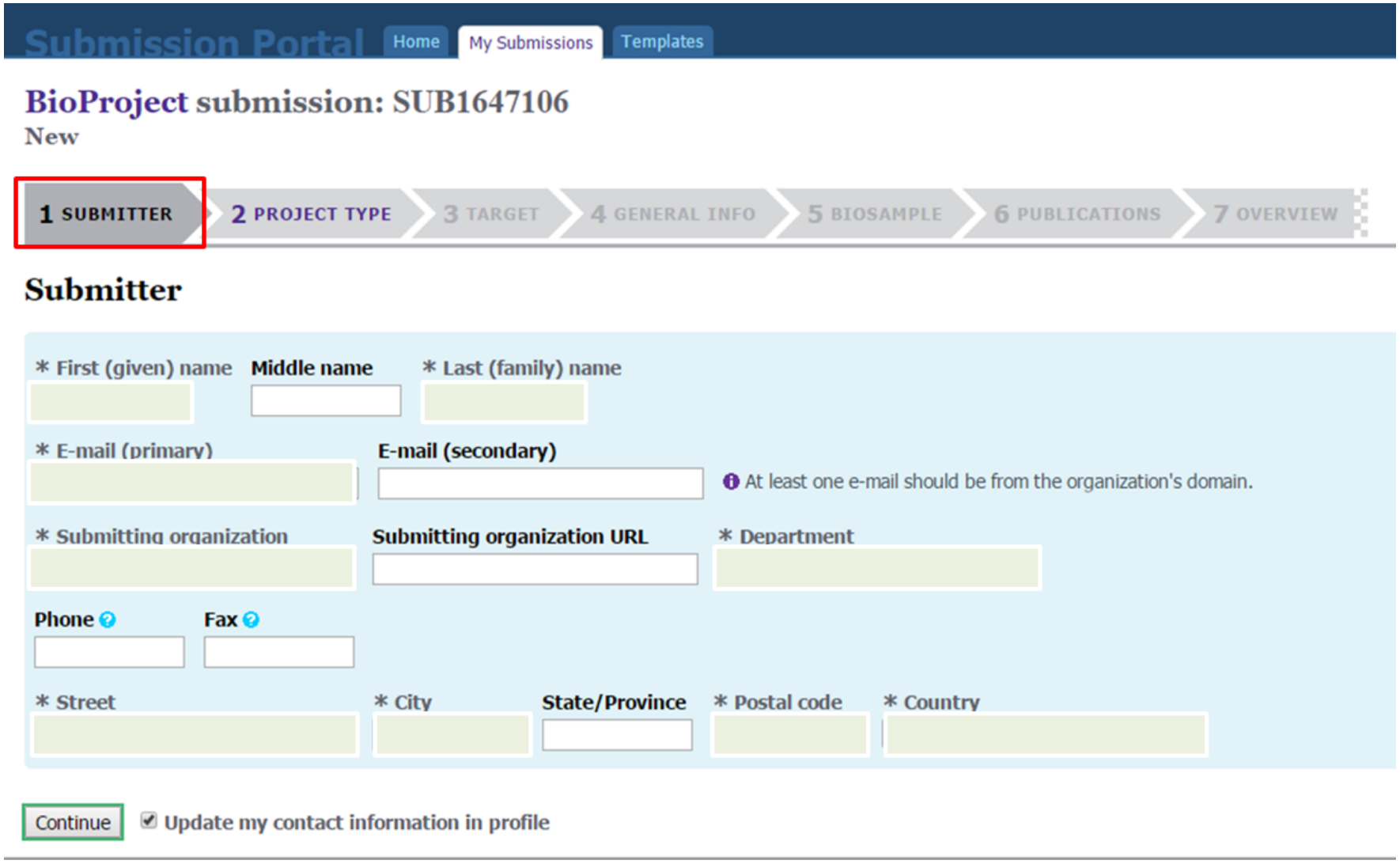
Well, it's better to use **other portal** for SRA!!!

Bioproject & Biosample



I. SRA submission_1. BioProject

BioProject: to describe the goal of the sequencing study



1 SUBMITTER | 2 PROJECT TYPE | 3 TARGET | 4 GENER

Project Type

* Project data type

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

Indicates the scope and purity of the biological sample used for the study. (Click to see more)

* Sample scope

Monoisolate
Multiisolate
Multispecies
Environment
Synthetic
Other

Target

A descriptive label for the sample scope selections multispecies, environment, or synthetic. For example: 1) primates; 2) acidic hot springs; 3) synthetic bacterium.

* Synthetic organism name

Strain | Breed | Cultivar | Isolate name | Label

* Synthetic organism description

1 SUBMITTER | 2 PROJECT TYPE | 3 TARGET | 4 GENERAL INFO | 5 BIOSAMPLE | 6 PUBLICATIONS | 7 OVERVIEW

General Info

* When this submission should be released to the public:

- Release immediately following curation
- Release on specified date (not viewable until this date or the release of linked data, whichever is first)

Release Date

* Project title

ACV Genome sequencing and assembly

Project title

* Public description

Private comments to NCBI staff

Project description

Relevance

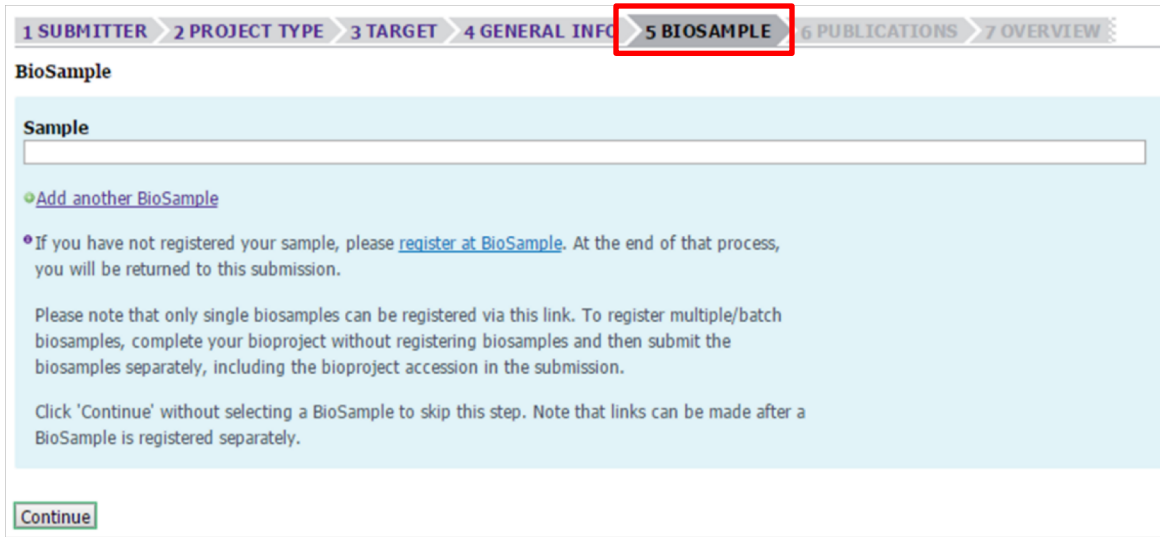
Relevance

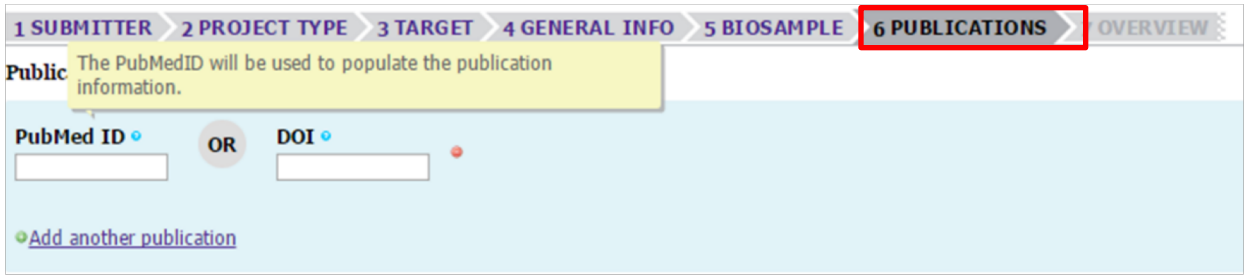
- Agricultural
- Medical
- Industrial
- Environmental
- Evolution
- Model organism
- Other

* Is your project part of a larger initiative which is already registered?

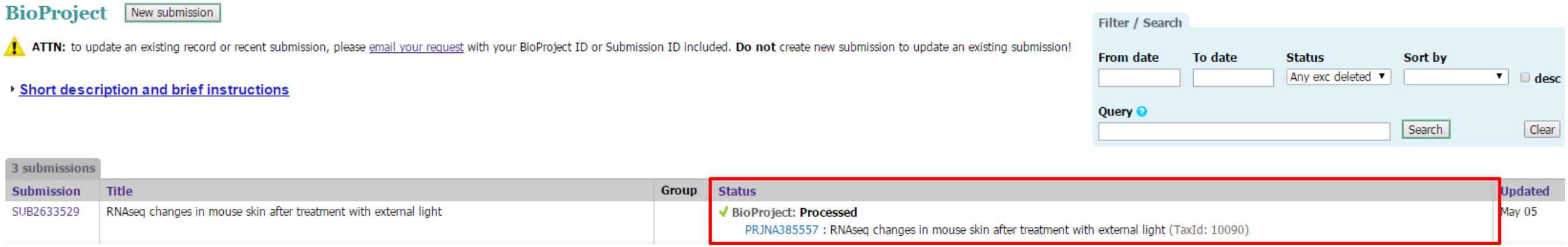
- No
- Yes

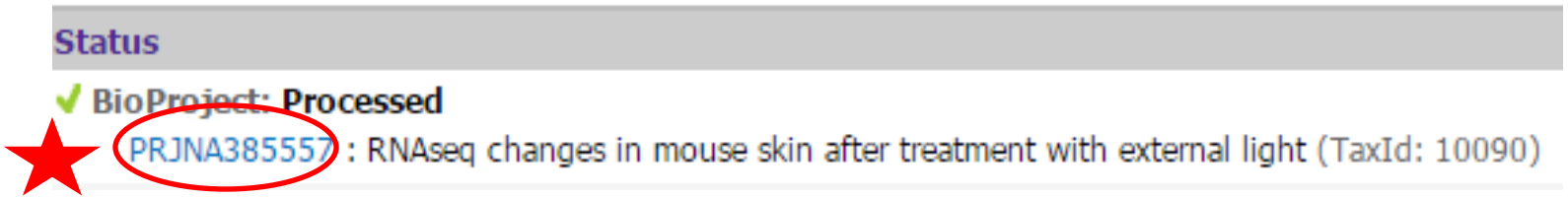
Provide a brief title, as a phrase or short sentence for public display. Examples: 1) Chromosome Y sequencing; 2) Opportunistic pathogen that causes important food-born disease; 3) Global studies of microbial diversity on human skin.





After submission…





After that, you will receive **PRJNA#** for GEO submission.

An example (from Dr.Lai's case)

Project Type

Project data type: Genome sequence and assembly

Sample scope: Monoisolate

Target

Organism name: MCF-7 cells

General information (Project detail)

project title: Next-generation sequencing of lncRNAs regulated by oxygen in MCF-7 cells.

project description: To identify the oxygen-responsive lncRNAs, next-generation sequencing was performed. Cells were harvested under normoxia (O2), hypoxia (N2) and re-oxygenation (Re-O2) conditions. Each condition was done in triplicate.

Relevance: Medical

Release Date: Release immediately after curation.

I. SRA submission_2. BioSample

BioSample submission: SUB1733038
New

Delete submission

1 SUBMITTER | 2 GENERAL INFO | 3 SAMPLE TYPE | 4 ATTRIBUTES | 5 COMMENTS | 6 OVERVIEW

Submitter

Required fields are marked with asterisk *

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

* E-mail (primary) | E-mail (secondary)

At least one e-mail should be from the organization's domain.

* Submitting organization | Submitting organization URL | * Department

Phone | Fax

* Street | * City | State/Province | * Postal code | * Country

Continue

1 SUBMITTER | 2 GENERAL INFO | 3 SAMPLE TYPE | 4 ATTRIBUTES | 5 COMMENTS

General Information

Release Date

Release date

* **When this submission should be released to the public:**

- (•) Release immediately following processing (**recommended**)
- ( ) Release on specified date or upon publication, whichever is first

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

* **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 page.
- ( ) Single BioSample
  You will be asked to manually complete a web form to describe one sample and its attributes.

Click this for multiple samples

Continue

1 SUBMITTER | 2 GENERAL INFO | 3 SAMPLE TYPE | 4 ATTRIBUTES | 5 COMM

Sample Type

* **Select the package that best describes your samples:**

- ( ) **Pathogen affecting public health**
  Use for pathogen samples that are relevant to public health. Required attributes include those considered useful for the rapid analysis and trace back of pathogens.
- ( ) **Microbe**
  Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MIxS, Pathogen or Virus packages.
- ( ) **Model organism or animal sample**
  Use for multicellular samples or cell lines derived from common laboratory model organisms, e.g., mouse, rat, Drosophila, worm, fish, frog, or large mammals including zoo and farm animals.
- ( ) **Metagenome or environmental sample**
  Use for metagenomic and environmental samples when it is not appropriate or advantageous to use MIxS packages.
- ( ) **Invertebrate**
  Use for any invertebrate sample.
- ( ) **Human sample**
  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

1 SUBMITTER | 2 GENERAL INFO | 3 SAMPLE TYPE | 4 ATTRIBUTES | 5 COMMENTS | 6 OVERVIEW

Attributes

選擇檔案 未選擇任何檔案

Convert xcl to tab-delimited text file for upload

Template for BioSample package **Model organism or animal; version 1.0**
Download Excel Download TSV
For more information, please see creating sample attribute file.

Continue

This is a submission template for batch deposit of 'Human; version 1.0' samples to the NCBI BioSample database (http://www.ncbi.nlm.nih.gov/biosample/).
GREEN fields are mandatory. Your submission will fail if any mandatory fields are not completed. If information is unavailable for any mandatory field, please enter 'not collected', 'not applicable' or 'missing' as appropriate.
YELLOW 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.
Hover over field name to view definition, or see http://www.ncbi.nlm.nih.gov/biosample/docs/attributes/.
CAUTION: 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 uploading to BioSample.

TO MAKE A SUBMISSION:

1. Complete this template table
2. Save the worksheet as a Text (Tab-delimited) file -- (use 'File, Save as, Save as type: Text (Tab-delimited)')
3. Upload the text file on the 'Attributes' tab of the BioSample Submission Portal at https://submit.ncbi.nlm.nih.gov/subs/biosample/.

If you have any questions, please contact us at biosamplehelp@ncbi.nlm.nih.gov.

*Organism: Mus musculus

*sample_name	sample_title	bioproject_accession	*organism	*isolate	*age	*biomaterial_provider	*sex	*tissue	cell_line	cell_subtype	cell_type	culture_collection	dev_stage	disease

treatment	description

If there is no information for GREEN column, please type “not collected”

**Replicate/ different treatment

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

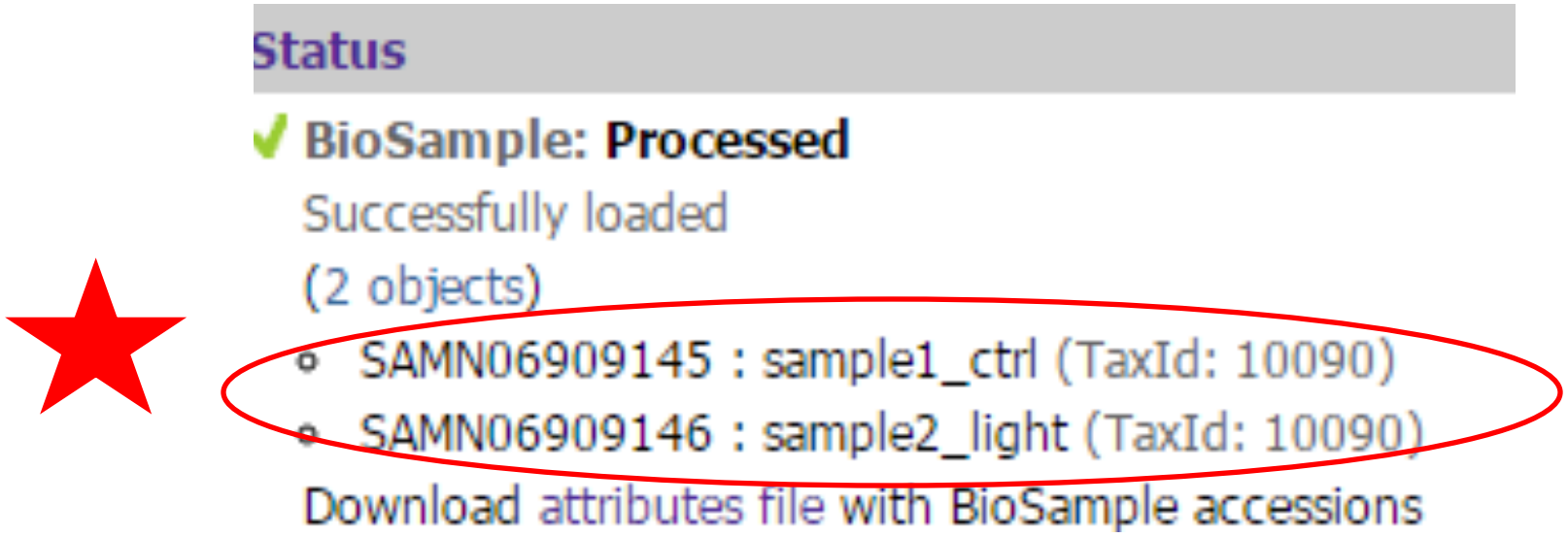
Filter / Search

From date | To date | Status: Any exc deleted | Sort by

Query | Search

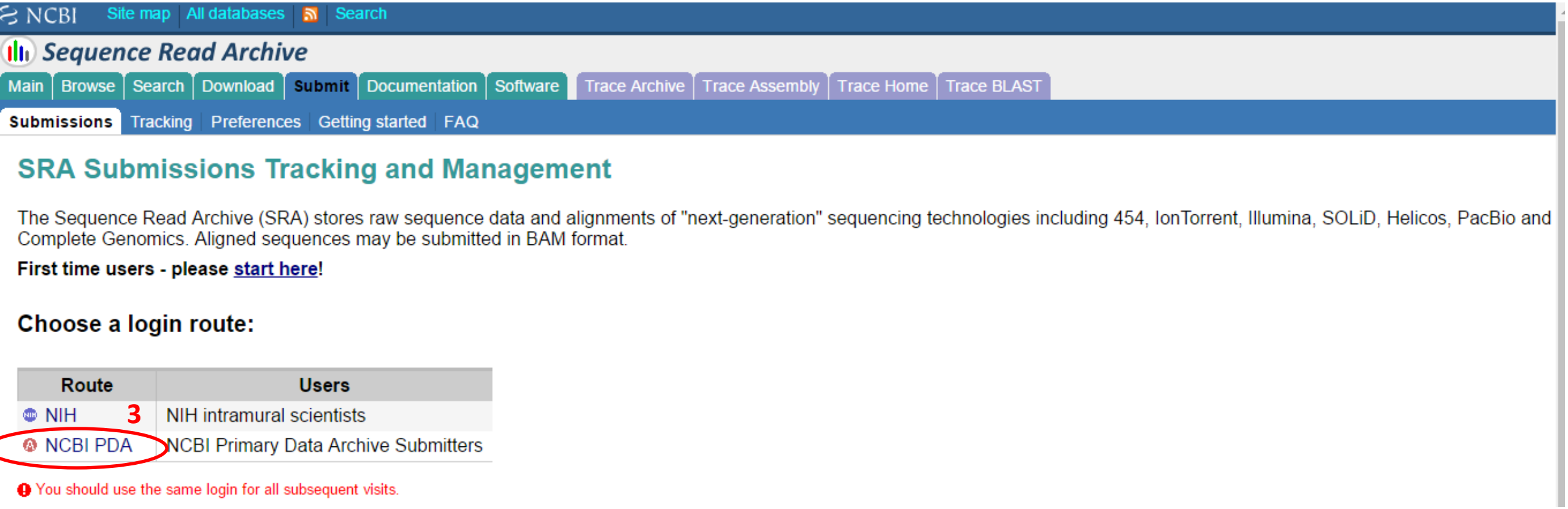
3 submissions

Submission	Title	Group	Status	Updated
SUB2633554	Model organism or animal sample		✓ **BioSample: Processed** Successfully loaded (2 objects) ∘ SAMN06909145 : sample1_ctrl (TaxId: 10090) ∘ SAMN06909146 : sample2_light (TaxId: 10090) Download attributes file with BioSample accessions	May 07



I. SRA submission_3. SRA

SRA submission portal: http://www.ncbi.nlm.nih.gov/Traces/sra_sub/sub.cgi



Layout

Main | Browse | Search | Download | **Submit** | Software | Trace Archive | Trace Assembly | Trace BLAST

Submissions | Tracking | Preferences | Getting started | FAQ

Submission: RNA-Seq_Mu_skin

Submission Id	Submitter	Updated	State	Status	Comments
National Taiwan University : RNA-Seq_Mu_skin		2017-05-08 00:03	completed	7	• SRP106582 : PRJNA385557 • 2 samples • 2 experiments • 2 runs

Files

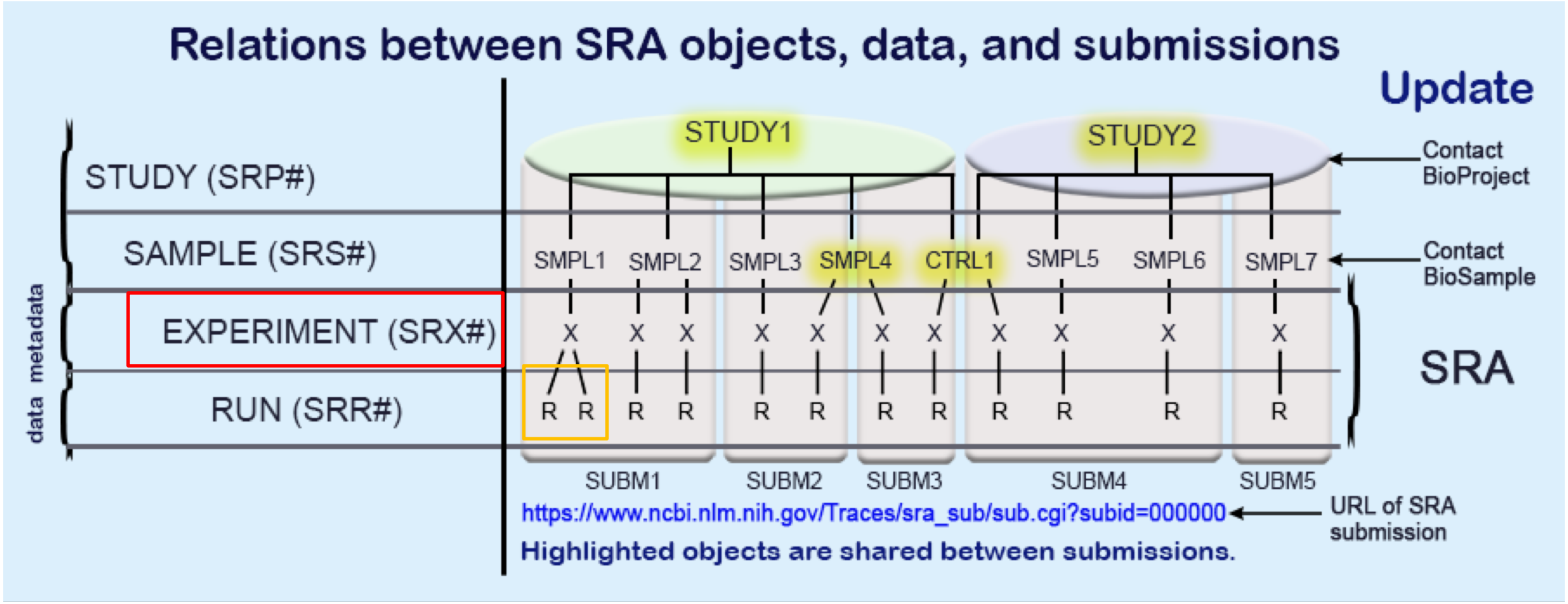
Type	Alias	Accession	Uploaded	Links	Files	Released
STUDY	PRJNA385557	SRP106582	32 h	ok	done	**2017-05-10 00:00:00**
SAMPLE	sample1_ctrl	SRS2169853	32 h	ok	done	2017-05-07 21:54:15
EXPERIMENT New Run	sample1_ctrl	SRX2788265	30 h	ok	done	**2017-05-10 00:00:00**
RUN	sample1_ctrl	SRR5514884	30 h	ok	done	**2017-05-10 00:00:00**
SAMPLE	sample2_light	SRS2169863	31 h	ok	done	2017-05-07 22:39:14
EXPERIMENT New Run	sample2_light	SRX2788275	30 h	ok	done	**2017-05-10 00:00:00**
RUN	sample2_light	SRR5514885	30 h	ok	done	**2017-05-10 00:00:00**

New Experiment

The SRA web submission interface for Sample creation has been replaced by the BioSample Submission Portal. Please make all sample submissions through the portal. SRA XML submissions are unchanged.

Set release date to: 2017-05-10 (YYYY-MM-DD)

Study, Sample, Experiment and Run



New Submission

*Alias

Submission Comment

*Release date
2017-07-31
(YYYY-MM-DD)

Save Cancel

Experiment

Meta information

*Platform

*Alias *Title

*BioProject accession
Look at Entrez BioProject or Submit new BioProject

*BioSample accession
Look at Entrez BioSample or Submit new BioSample

Library Construction / Experimental Design

Library

Library name *Strategy *Source *Selection

*Layout

*Strategy
- WGA
- WGS
- WXS
- RNA-Seq
- miRNA-Seq
- Tn-Seq
- WCS
- CLONE
- POOLCLONE
- AMPLICON
- CLONEEND
- FINISHING
- ChIP-Seq
- MNase-Seq
- DNase-Hypersensitivity
- Bisulfite-Seq
- EST
- FL-cDNA
- CTS

*Source
TRANSCRIPTOMIC
- GENOMIC
- TRANSCRIPTOMIC
- METATRANSCRIPTOMIC
- METAGENOMIC
- SYNTHETIC
- VIRAL RNA
- OTHER

*Selection
- unspecified
- RANDOM
- PCR
- RANDOM PCR
- RT-PCR
- HMPR
- MF
- CF-S
- CF-M
- CF-H
- CF-T
- MDA
- MSLL
- cDNA
- ChIP
- MNase
- DNAse
- Hybrid Selection
- Reduced Representation

*Layout
- FRAGMENT
- PAIRED

Experiment: SRX1177891/MCF-7 RNA seq samples data

Run: SRR2235351

General info

*Alias: Sample_H1

*Run data file type: fastq

bam
fastq
qseq/seq_prb_int
srf

Data files

*File name	*MD5 checksum		File status
H1_GTGGCCT_L007_R1_	aa3043a90b98c7f13e6b0d2e14bf729d	Delete	loaded
H1_GTGGCCT_L007_R2_	4d274380c4f91605e034d5afaddb8d87	Delete	loaded

Add

File name: The full file name, with extension(s) (e.g., sequence_data.fastq, NOT only sequence_data). Please DO NOT include directory or path information like "C:\Documents\" or "/Users/Documents/". We DO NOT accept .zip or .rar compressed files; .gz and .bz2 compressed files ARE accepted.

MD5 checksum: 32-character alphanumeric string that can be computed with native command line tools "md5" (Mac OS X) or "md5sum" (Linux). Windows users will need to download a 3rd party utility.

Transmitting your data files to the SRA:

Please enter the above information and then deposit the files by FTP using the following credentials:
Address: **ftp-private.ncbi.nlm.nih.gov**
Login: **sra**
Password: **Qrjo6iJ4**

Note: Files are regularly moved from FTP to archive. If your files have been removed from FTP but have not yet linked to your SRA submission, please wait 6 hours before contacting sra@ncbi.nlm.nih.gov for assistance. Alternatively, download the Aspera Connect plugin and contact the SRA (sra@ncbi.nlm.nih.gov) for a key file and 'ascp' command line tool instructions.

```
$ md5sum <file>
$ md5sum *.fastq > sra.md5
```

```
$ cd to directory with your data files
$ lftp -u sra ftp-private.ncbi.nlm.nih.gov
Key in Password: Qrjo6iJ4
> mkdir PRJNA#
> cd PRJNA#
> mput *.fastq
> exit
```

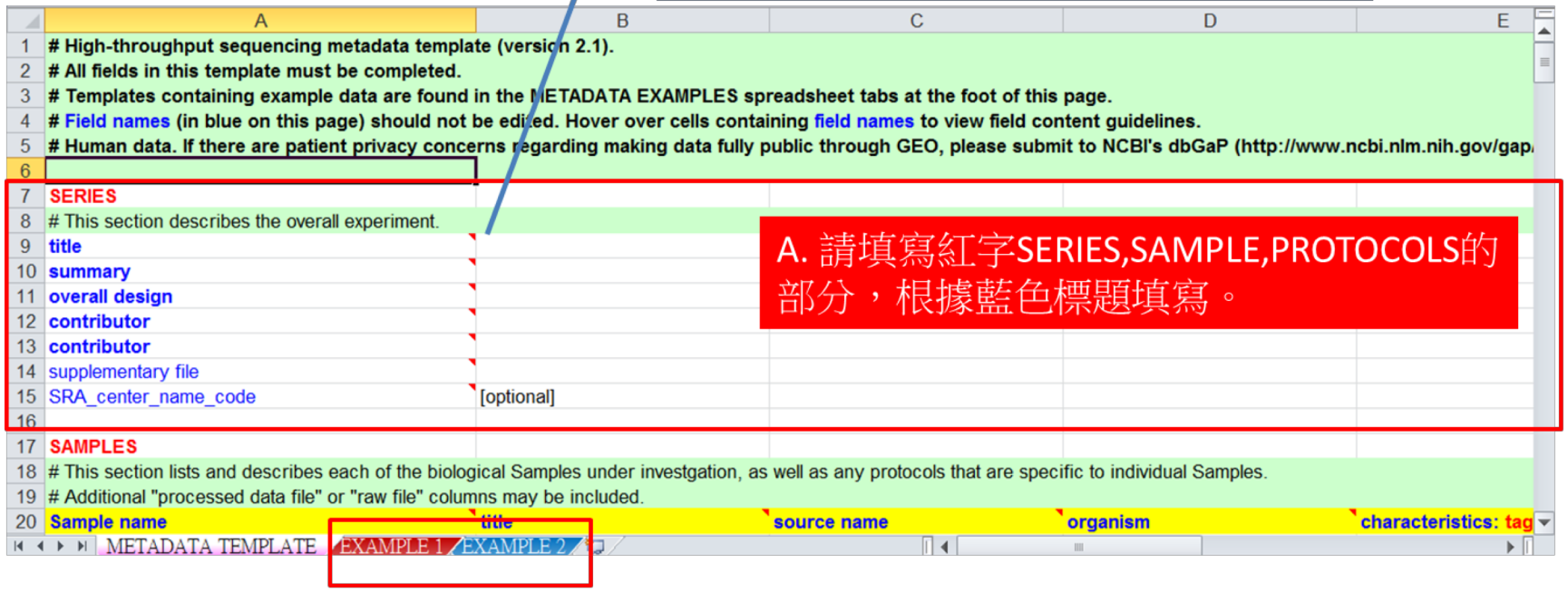
SRA search


SRA
SRA
SRX111436
Create alert
Advanced
Full
Send to:
SRX111436: Whole Exome sequencing for the 1000 Genomes Project
8 ILLUMINA (Illumina HiSeq 2000) runs: 17.8M spots, 2.7G bases, 1.3Gb downloads
Design: Whole Exome sequencing for the 1000 Genomes Project via in-solution hybrid selection
Submitted by: Broad Institute (BI)
Study: Exome sequencing of (KHV) Kinh in Ho Chi minh City, Vietnam HapMap population
PRJNA59815 • SRP004063 • All experiments • All runs
show Abstract
Sample: Coriell HG02047
SAMN00630256 • SRS212513 • All experiments • All runs
Organism: Homo sapiens
Library:
Name: Catch-111931
Instrument: Illumina HiSeq 2000
Strategy: WXS
Source: GENOMIC
Selection: Hybrid Selection
Layout: PAIRED
Spot descriptor:
1
forward
77
reverse
Experiment attributes: (show all 4 attributes...)
Pipeline: show...
Runs: 8 runs, 17.8M spots, 2.7G bases, 1.3Gb
Run # of Spots # of Bases Size Published
SRR389621 2,257,646 343.2M 172.1Mb 2011-12-14
SRR389628 2,222,999 337.9M 169.9Mb 2011-12-14
SRR389633 2,221,570 337.7M 172.2Mb 2011-12-14

II. GEO submission_1. fill-in metasheet

- Materials: metasheet and processed data

B. 若藍色字不知道要填寫甚麼了話，框框內有個紅色三角物件，點選他會有說明。



	A	B	C	D	E
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				
6					
7	SERIES				
8	# This section describes the overall experiment.				
9	title				
10	summary				
11	overall design				
12	contributor				
13	contributor				
14	supplementary file				
15	SRA_center_name_code	[optional]			
16					
17	SAMPLES				
18	# This section lists and describes each of the biological Samples under investgation, as well as any protocols that are specific to individual Samples.				
19	# Additional "processed data file" or "raw file" columns may be included.				
20	Sample name	title	source name	organism	characteristics: tag

METADATA TEMPLATE | EXAMPLE 1 | EXAMPLE 2

A. 請填寫紅字SERIES,SAMPLE,PROTOCOLS的部分，根據藍色標題填寫。

C. 此外，這裡也有填寫的範例可以參考。

Excel file : seq_template_v2.1.xls

**Extra Info_GEO metasheet

[1] Reference the SRA Study accession (SRPnnnn) and BioProject accession (PRJNAnnnn) in the **SERIES 'Overall design' field**.

[2] **In the 'raw file' column of the SAMPLES section**, please list the corresponding SRA Experiment accessions (SRXnnnnnn) so that we can create the appropriate links between the SRA Experiments and GEO Samples.

[3] **Add a 'BioSample' column to the SAMPLES section** and include the corresponding BioSample accessions (SAMNnnnnnn).

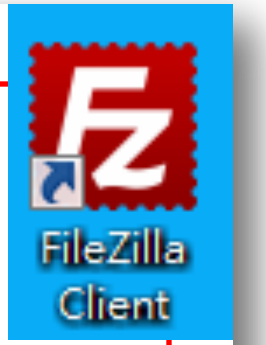
[4] Add an 'instrument model' column to the SAMPLES section and indicate the instrument model that was used for sequencing.

II. GEO submission_2. upload

- **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	geo
password	33%9uyj_fCh?M16H

      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:33%259uyj_fCh%3FM16H@ftp-private.ncbi.nlm.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

    Note: SFTP protocol not supported.

If you have any questions or concerns regarding data transfer, please e-mail us.

II. GEO submission_3. notification

- **Email notification:**
  - After file transfer is complete, please e-mail GEO with the following information:
    1. GEO account username (**llai0619**);
    2. Names of the directory and files deposited;
    3. Public release date (required - up to 3 years from now - see 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 (see FAQ). *****