

Korean Nucleotide Archive (KoNA)

사 용 매 뉴 어 리

<b>I INTRODUCTION.....</b>	<b>2</b>
1-1. About KoNA .....	2
1-2. Structure of KoNA.....	3
1-3 How to cite.....	5
1-4 Contact us.....	6
<b>II DATA SUBMISSION.....</b>	<b>7</b>
2-1 Submission overview.....	7
2-2 Create account.....	8
2-3 Submit study information (BioProject).....	11
2-4 Submit sample information (BioSample).....	16
2-5 Submit metadata and raw data (KAD).....	24
2-6 Manage my submission (MyPage).....	29
<b>III DATA SEARCH AND DOWNLOAD.....</b>	<b>31</b>
3-1 Search Umbrella BioProject.....	31
3-2 Search BioProject s.....	32
3-3 Search BioSample.....	36
3-4 Search KRA and download raw data.....	37
3-5 Statistics.....	39
<b>IV FAQ.....</b>	<b>41</b>
<b>VI APPENDIX.....</b>	<b>44</b>
4-1 GBox user guide.....	44
4-2 More information.....	48

# I INTRODUCTION

## 1-1. About KoNA

Korean Nucleotide Archive(KoNA)는 차세대유전체기술(NGS, Next Generation Sequencing) 기반으로 생산된 대용량(high-throughput) 유전체 서열(read)의 저장 및 공유를 목적으로 국가생명연구자원정보센터(KOBIC, Korea Bioinformation Center)에서 개발하여 운영하는 데이터 저장소(repository)입니다.

KoNA의 게놈 데이터의 표준 형식은 Nucleotide Sequence collaboration (INSDC)에 의해 규정된 형식을 따릅니다. 따라서, KoNA에 등록된 데이터는 National Center for Biotechnology Information(NCBI)의 Sequence Read Archive(SRA) 및 European Bioinformatics Institute(EBI)의 European Nucleotide Archive (ENA)의 데이터와 호환됩니다.

또한, KoNA에 저장된 모든 데이터는 큐레이션되고 품질 관리됩니다. 그런 다음, 이 모든 검증된 데이터가 연구원들에게 공개됩니다. KoNA에서 제공하는 모든 공개 데이터는 인체에서 생성된 데이터를 제외하고 제한 없이 사용할 수 있습니다.

2020년 9월 보건복지부 개인정보보호위원회가 마련한 '생명윤리 및 안전에 관한 법률 시행규칙(생명윤리법)'과 '국내 보건의료데이터 활용지침'에 따라 인체소재의 게놈 데이터를 공유·활용해야 합니다. 본 법률 및 지침에 따라 모든 데이터를 수집하고 공유할 수 있습니다. 향후, KoNA는 최첨단 single-cell 기술과 같은 여러 게놈 플랫폼에서 생성된 모든 데이터의 저장, 관리 및 공유를 위해 확장해 나갈 것입니다.

## 1-2. Structure of KoNA

### ◎ KoNA data format

KoNA 는 INSDC 의 데이터 형식 및 구조 모델을 기반으로 구축되었습니다. 주요 국가와 유사한 genome 데이터 저장 서비스를 국내 및 전 세계 연구자에게 제공합니다. 미국 SRA, 유럽 ENA 등 INSDC 데이터베이스에 데이터를 제출했거나 검색 서비스를 이용한 경험이 있는 연구자는 KoNA 의 데이터 구조를 빠르게 이해할 수 있습니다

KoNA 에 저장된 게놈 데이터는 주로 raw 데이터와 메타데이터로 나뉩니다. raw 데이터는 FASTQ 파일을 포함한 시퀀싱 장치를 통해 생성된 genome read 입니다. 메타데이터는 raw 데이터를 설명하는 정보이며 주로 프로젝트, 샘플, 실험 및 Run 으로 구성됩니다. 이 형식은 INSDC 데이터베이스의 형식과 동일합니다.

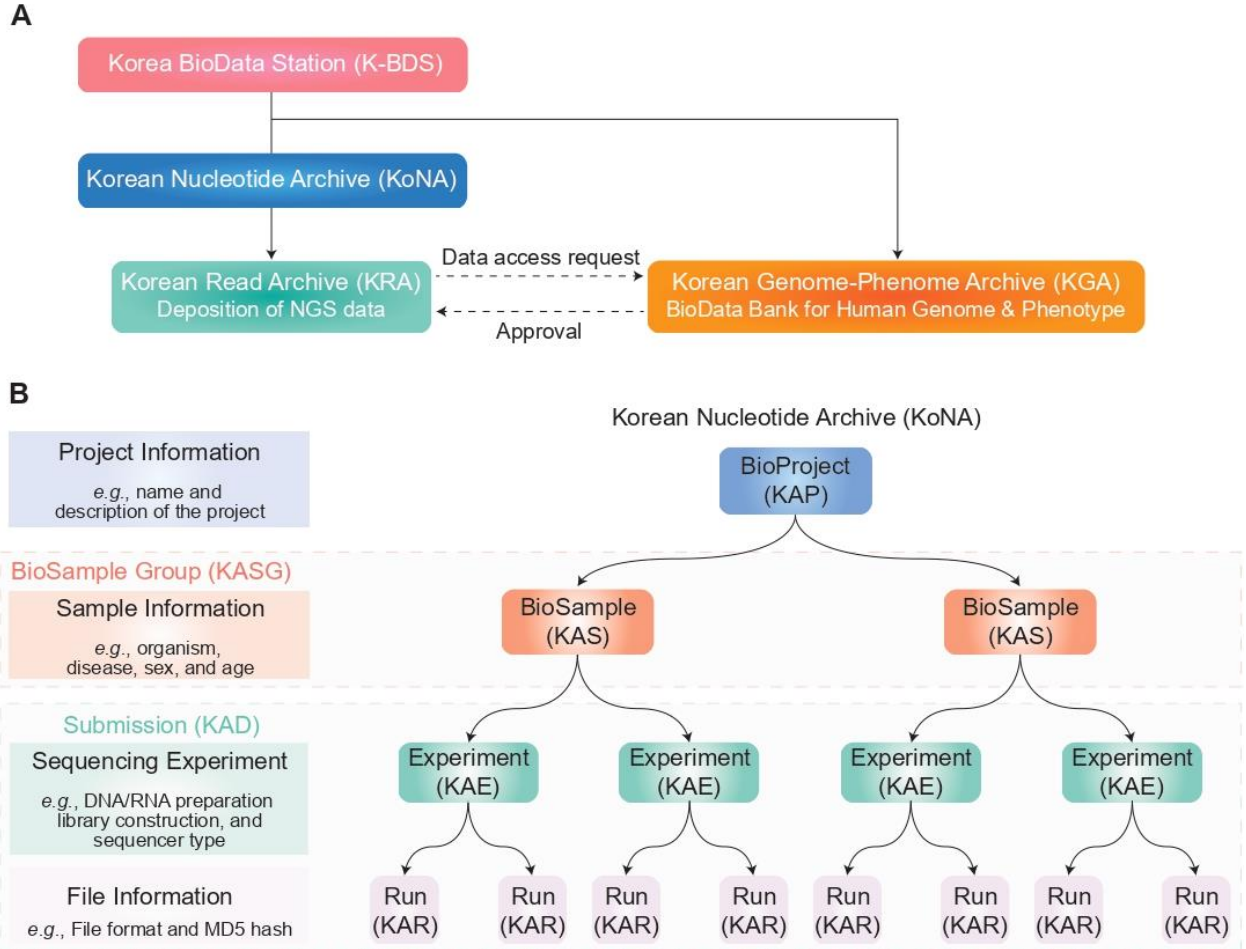
### ◎ Metadata format

메타데이터는 주로 관련된 프로젝트, 샘플, 실험 및 Run 으로 구성됩니다. 즉, raw 데이터를 제출하려면 프로젝트-표본-실험-Run 정보가 존재하고 서로 연결되어야 합니다. 프로젝트 정보는 프로젝트에 대한 설명과 genome 생성을 위한 연구 정보를 포함한 바이오 프로젝트 데이터베이스로 구현됩니다. BioProject 에는 제출자, 날짜, 프로젝트 설계 및 발행과 같은 정보가 포함됩니다. 바이오 프로젝트는 독립적으로 운용되며 다른 데이터와 연결되지 않지만 포괄적인 프로젝트와 연결될 수 있습니다. 또한 대규모 프로젝트 또는 컨소시엄(예: 한국 포스트 게놈 프로젝트)에 포함될 수 있는 바이오 프로젝트가 있는 경우, 바이오 프로젝트에는 그들의 포괄적인 프로젝트에 대한 정보가 포함됩니다. 바이오 프로젝트 가입 ID 는 "KAP"로 시작합니다

샘플 정보는 BioSample 데이터베이스로 구현되며, 데이터 생성에 사용되는 생물학적 소스 재료 또는 샘플에 대한 정보를 포함합니다. BioSample 은 BioProject 와 구조적으로 연결되어야 합니다. 따라서 BioSample 에 샘플 정보를 제출하기 전에 BioProject 에서 프로젝트 정보를 제출하고 KoNA 관리자의 승인을 받아야 합니다(즉, BioSample 을 제출하려면 '상위' BioProject 가입 ID 가 필요함). 동일한 유형의 여러 바이오 검체(예: 여러 폐암 환자의 검체)를 제출하려면 바이오 검체 그룹이 필요합니다. 제출자는 스프레드시트를 사용하여 바이오샘플에 대한 정보를 설명할 수 있습니다. 스프레드시트의 각 행은 각 바이오샘플을 나타내며 스프레드시트 자체는 바이오샘플 그룹을 나타냅니다. BioSample Group 의 등록 ID 는 "KASG"로 시작하며, 각 BioSample 의 등록 ID 는 "KAS"로 시작합니다

각 실험은 특정 샘플에 대한 고유한 시퀀싱 라이브러리를 나타내며, 실험 및 Run 은 KAD(Korea Read Archive)로 그룹화됩니다.

▶ 4 가지 유형의 메타데이터 객체



< BioProject, BioSample 및 KAR 의 관계 >

BioProject, BioSample Group 및 KAD 데이터는 위와 같이 연결되어 있습니다. BioProject 의 메타데이터 개체에는 umbrella 프로젝트를 제외한 다른 데이터와의 연결 정보가 포함되어 있지 않습니다. 따라서 바이오 프로젝트는 독립적입니다. 바이오샘플 데이터에는 바이오프로젝트와의 연결 정보가 있어야 하며 샘플이 생성된 프로젝트 정보를 얻을 수 있습니다. KAD 에는 연결된 바이오샘플 그룹에 대한 정보가 있어야 합니다. 이와 같이 BioProject, BioSample Group, KAD 는 서로 연계되어 있습니다.

## 1-3. How to cite

귀하의 논문에서 바이오 프로젝트, 바이오 샘플 그룹 또는 KAD 제출을 인용하려면:

"본 출판물에서 논의된 raw 염기서열 데이터는 KAP20xxxxx의 등록 번호로 KoNA(Korean Nucleotide Archive)에 보관되었습니다." 제출자는 KAP20xxxx(바이오 프로젝트) 대신 KASG20xxxx(바이오 샘플 그룹) 또는 KAD20xxxx(KAD 제출)를 사용할 수도 있습니다."

KoNA 데이터베이스 인용하기:

"우리는 이제 peer reviewed journal 에 대한 기사를 준비합니다. 출판 절차가 완료될 때까지 URL 주소(<https://kobic.re.kr/kona>) 를 인용해 주시기 바랍니다.

### ◎ KoNA 구조

KoNA 의 모든 표준 및 구조가 INSDC 표준을 따르므로 KoNA 가 발행하는 등록 ID 시스템은 INSDC 에서 일반적으로 사용되는 형식을 따릅니다. 앞서 설명한 바와 같이, KoNA 의 모든 메타데이터는 다음과 같은 네 가지 객체로 구성됩니다: 바이오 프로젝트, 바이오 샘플, 실험 및 Run. 바이오샘플 그룹과 KAD 는 각각 여러 바이오샘플과 실험 및 Run 을 그룹화하는 데 사용됩니다. 각 개체에는 고유한 등록 ID 가 있습니다. 모든 등록 ID 는 대문자와 숫자로 구성됩니다. 문자는 객체를 나타내고 숫자는 승인 연도와 순서를 나타냅니다. 국가를 나타내는 문자의 경우 "K"가 포함되어 Korea 을 나타냅니다(다른 국가의 경우 "S", "E", "D" 및 "C"는 각각 U.S.A., Europe, Japan, China 를 나타냅니다).

먼저, 바이오 프로젝트의 가입 ID 는 "KAP"로 시작하고 그 뒤에 숫자가 나옵니다. "KA"는 Korean Archive 를 의미하고 "P"는 프로젝트를 의미합니다. umbrella 프로젝트의 가입 ID 는 "KAU"로 시작하며 "U"는 umbrella 를 나타냅니다.

BioSample 의 등록 ID 는 "KAS"로 시작하고 그 뒤에 숫자가 나옵니다. "S"는 Sample, "KA"는 Korean Archive 를 나타냅니다. 동일한 유형의 여러 샘플 그룹의 경우, 등록 ID 는 "KASG"로 시작하고 "G"는 그룹을 나타냅니다.

Experiment 및 Run 의 등록 ID 는 각각 "KAE" 및 "KAR"로 시작합니다. 첫 글자 "K"는 Korea 를 나타냅니다.

Object Name	KoNA Accession
Umbrella Project	KAU + 6 digits of number
BioProject	KAP+ 6 digits of number
BioSample Group	KASG + 6 digits of number
BioSample	KAS + 8 digits of number
KAD	KAD + 7 digits of number
Experiment	KAE + 8 digits of number
Run	KAR + 8 digits of number

## 1-4. Contact Us

데이터 작성/등록 문의 시 작성 중 또는 제출한 데이터의 임시 ID나 문제의 페이지를 캡처하여 보내주시면 더 빠르고 정확한 답변을 받을 수 있음

### ◎ EMAIL 문의

▶ [data@kobic.kr](mailto:data@kobic.kr)

### ◎ 유선 및 팩스 문의

▶ Tel : +82-42-879-8549

▶ Fax : +82-42-879-8519

### ◎ 주소

▶ Address : Korean Bioinformation Center (KOBIC), Korea Research Institute of Bioscience & Biotechnology (KRIBB), Daejeon 34141, Republic of Korea

## II DATA SUBMISSION

### 2-1. Submission overview

#### ◎ 데이터 등록 절차



#### ① 계정 생성 및 로그인

- ▶ KoNA에 기존 계정이 없는 경우 계정 생성 필수
- ▶ 데이터 작성 전 로그인 필수

#### ② 데이터 작성 단계

- ▶ 기본 연구 과제 정보(BioProject)와 샘플(BioSample) 작성\*
  - \* 항목에 대해서 빠짐없이 모두 입력해야 제출 가능
- ▶ 실험데이터(raw data) 업로드 및 메타데이터 작성
- ▶ 필요시 (제출 전) 데이터 추가, 수정, 삭제 가능\*
  - \* 제출 완료 이후 관리자 요청을 통한 수정 가능

#### ③ 데이터 검토 단계

- ▶ 제출된 데이터에 대해 품질관리자가 검토
- ▶ 검토 반려 시 제출자가 데이터 추가/수정하여 재제출

#### ④ 데이터 등록단계

- ▶ 검수 완료된 데이터에 대해 등록 완료 및 공개

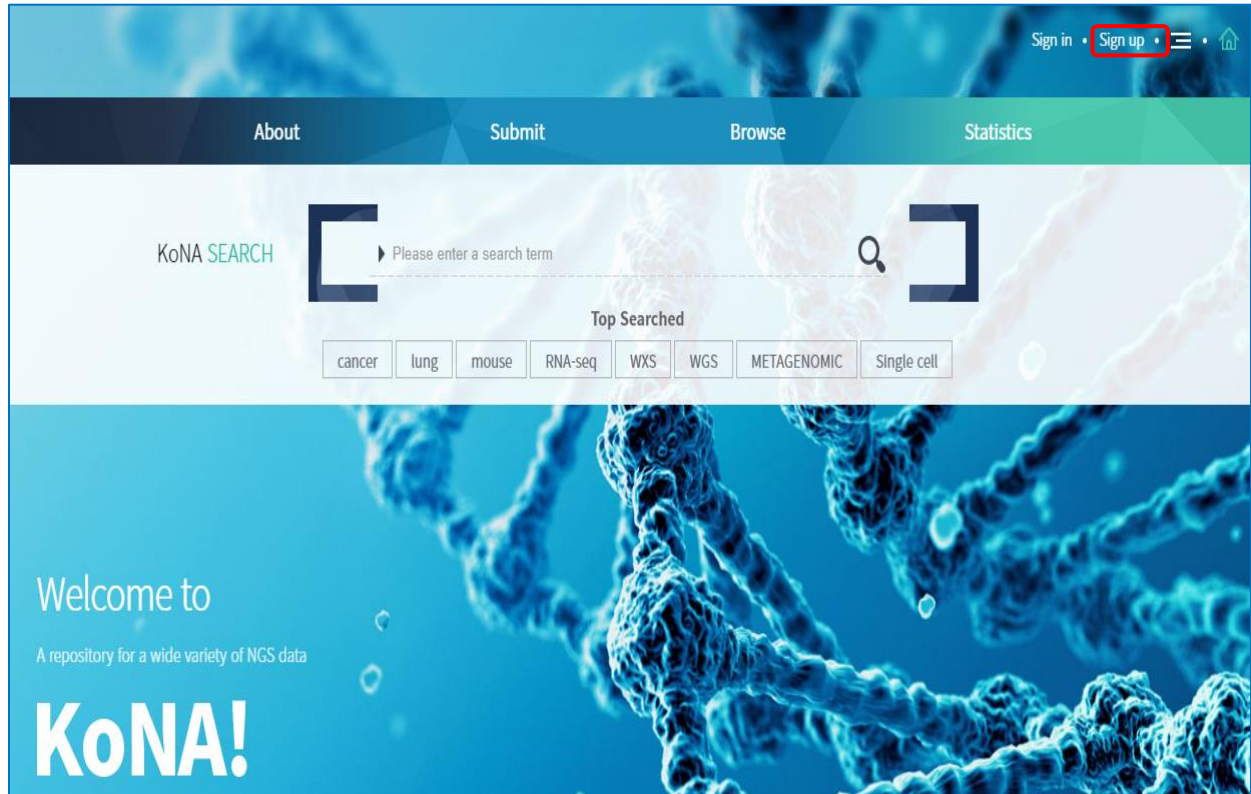


## 2-2. Create account

### ◎ 회원가입 및 로그인

▶ 회원가입은 [개인정보입력 » 가입완료] 절차 순으로 진행

① 메인페이지 우측 상단에서 [Sign up]을 클릭하여 페이지 이동



② “회원유형 선택”에서 [내국인] 또는 [외국인] 선택

③ 회원정보를 빠짐없이 입력 후 [Confirm]를 클릭 하여 제출한다.

\*회원가입에 문제가 발생 했을 경우 , kbds.help@kobic.kr로 문의

## Sign up

01 Agree 02 Info 03 Sign up

(\*Required)

### Enter a personal information

Korean Name

First Name \* Last Name \*

ID \* Duplication Check

Ⓞ Only 4 to 12 digits of English, numbers, and special characters ( \_ ) are allowed.

Password \* Confirm Password \*

Ⓞ It must be 8-20 characters (upper or lower) long and contain at least one number and special character.

E-mail \* Duplication Check

Institution \*

Department


Position

South Korea

Address

Phone Mobile

Ⓞ Enter numbers without '-'.

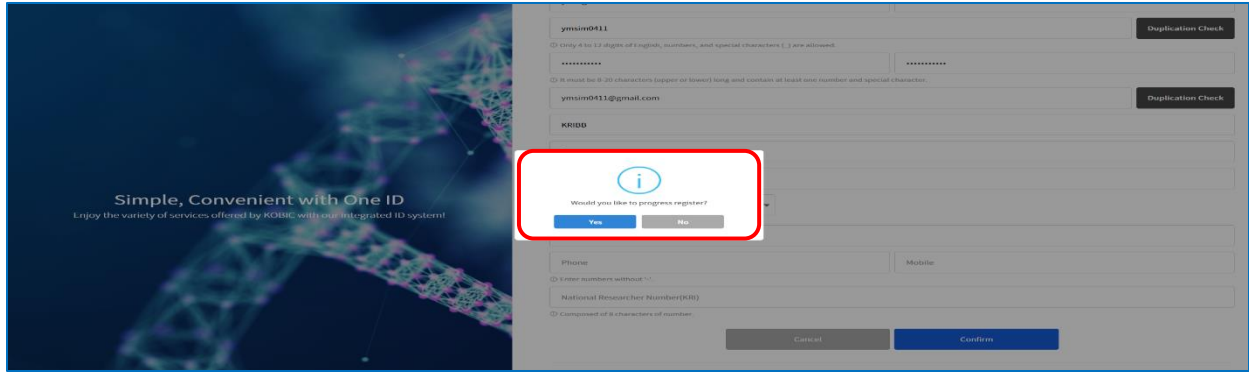


### Select membership type

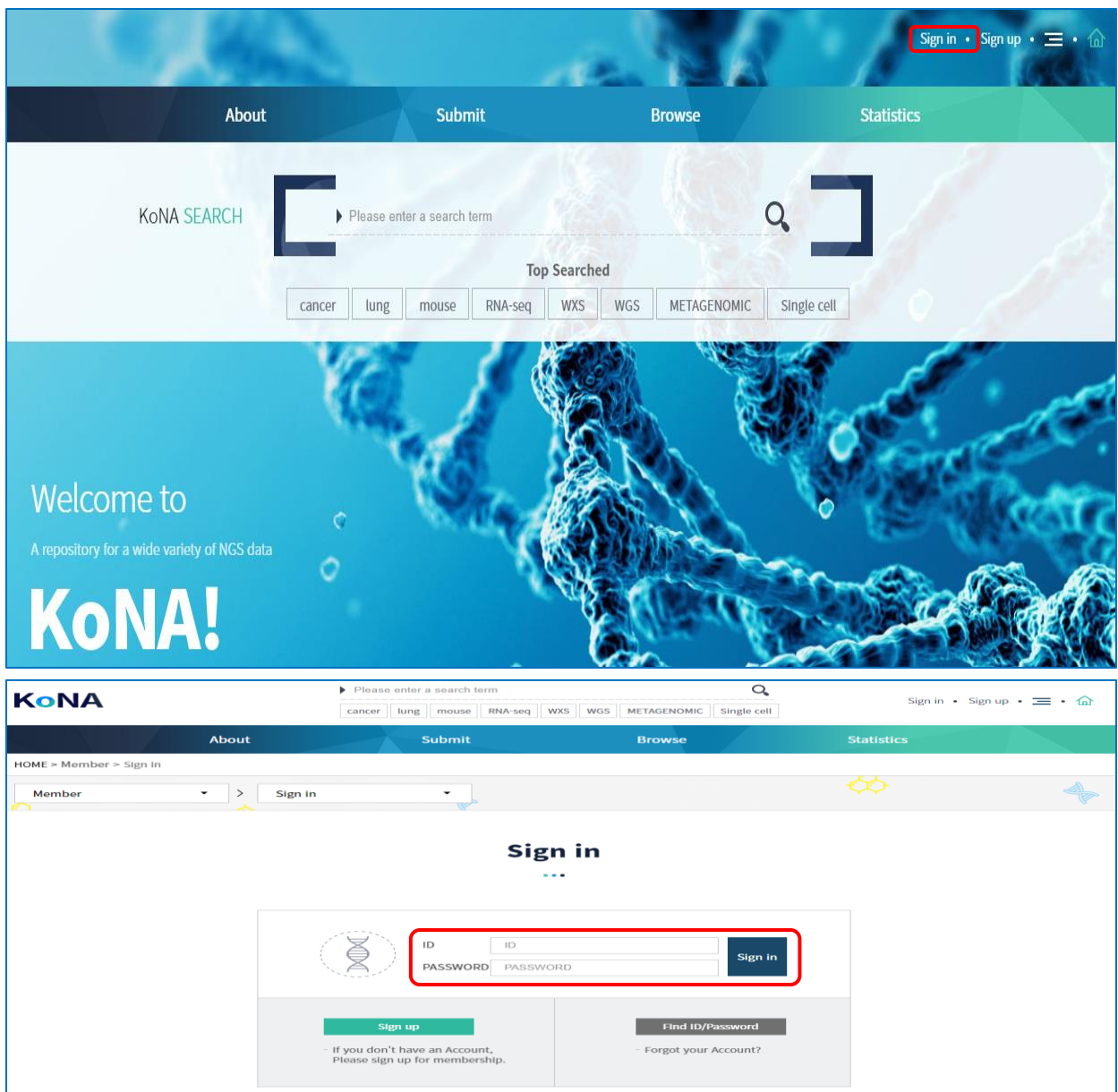
The registration process differs depending on the membership type. Please select the membership type to join.

Korean Foreigner

Privacy Policy  
125, Gwahak-ro, Yuseong-gu, Daejeon, Republic of Korea  
Phone: 042-879-8042 / Fax: 042-879-8039 / E-mail: info@kobic.kr  
Copyright © 2018 BY KOREA INFORMATION CENTER(KOBIC). ALL RIGHTS RESERVED.



④ 계정 생성 후 우측 상단의 [Sign in]을 클릭하여 로그인



## 2-3. Submit study information (BioProject)

### ◎ 기본 연구 정보 작성 및 등록

기본 연구 정보 작성 및 등록 기본 연구 정보 등록을 위해 BioProject를 작성 및 제출

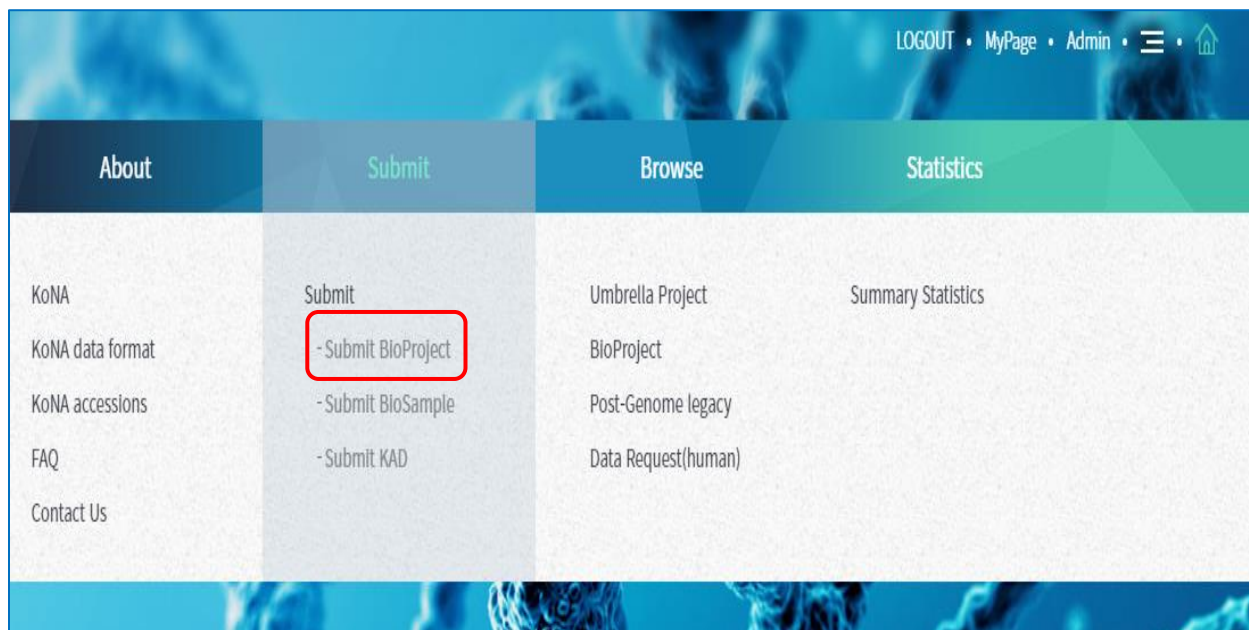
#### • BioProject

- ▶ 연구 프로젝트에 대한 개괄적인 정보를 작성
- ▶ KoNA내 등록되는 모든 데이터 타입에 대해 필수적으로 작성되어야 함
- ▶ 제출자의 정보, 과제정보, 연구성과 등의 정보를 작성
- ▶ 제출된 BioProject는 관리자 검토를 통해 승인 후 등록 완료

#### • BioProject 작성단계



1. 데이터 등록을 위하여 koNA 메인페이지(<https://www.kobic.re.kr/kona/>) 상단 메뉴의 [Submit » Submit BioProject] 페이지로 이동



2. Create 메뉴를 눌러 새로운 BioProject 페이지를 활성화 한다.

## Submit BioProject

...

A BioProject is a collection of biological data related to a single initiative originating from a single organization or a consortium.  
 A BioProject record provides users a single place to find links to the diverse data types generated for that project.  
 The KoNA BioProject issues accession numbers with the prefix 'KAP' to the submitted projects.

Total : 1 / Page 1  
 Submit 0 | Return 0 | Approval 0

Create

NO	BioProject Accession ID	Project Title	Registration Date	Status	Operation
1	KAP230591	-	-	Writing	<div style="display: flex; gap: 5px;"> <span>Edit</span> <span>Delete</span> </div>

### 3. 등록자의 인적사항

① 등록자 인적사항을 빠짐없이 입력

\* BioProject 등록 시 PI (Principal Investigator; 연구책임자) 인적사항 기입을 권장

② 등록자 인적사항은 회원가입 시의 정보가 자동 입력되며, 등록자 인적사항에 변동이 있는 경우, 웹페이지 내에서 수정 가능

## Submit BioProject

...

A BioProject describes information for a collection of biological data provided by a single laboratory/organization or initiative/consortium.

Required \* / Conditionally required \*

**Submitter**  Import the user profile

<b>Name *</b>	<input type="text" value="young mi"/> <input type="text" value="Sim"/> <small>Last name, First name (ex, Hong, Gil Dong)</small>	
<b>Email *</b>	<input type="text" value="ymsim"/> @ <input type="text" value="kribb.re.kr"/>	
<b>Organization *</b>	<input type="text" value="test"/>	
<b>Department *</b>	<input type="text" value="test"/>	
<b>Address *</b>	<input type="text" value="125, Gwahak-ro, Yuseong-gu, Daejeon, Korea 34141"/>	<input type="button" value="Search Postal code"/>
<b>Country *</b>	<input type="text" value="South Korea"/>	

#### 4. 연구 과제 정보 기입 및 프로젝트의 제목 및 설명

- ① NTIS 과제 정보 입력 (해당 사항 없으면 공란 처리)
- ② 프로젝트의 영문 제목 : 연구 과제 정보의 '주요 연구과제의 제목'이 아닌 등록자가 BioProject에 포함할 시료/기기 정보와 실험 데이터의 특성을 고려한 연구 프로젝트의 제목을 명명하여 입력
- ③ 프로젝트의 영문 설명 : 프로젝트에 대한 상세 설명을 영문으로 입력
- ④ Project Data type 및 sample Scope 입력

Project Design	
NTIS Number	<input type="text"/>
Project title *	<input type="text"/>
Relevance *	<div style="border: 1px solid red; padding: 2px;"> <input type="text" value="Select"/> </div> <p>Select or provide the primary general relevance of the project</p>
Description *	<input type="text"/>
Project Data Type *	<p>Indicate the general label of the primary study goal.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Whole Genome sequencing : Whole or partial, genome sequencing project (with or without transcriptome)</li> <li><input type="checkbox"/> Clone ends Clone-end sequencing project</li> <li><input type="checkbox"/> Epigenomics : DNA methylation, histone modification, chromatin accessibility data</li> <li><input type="checkbox"/> Exome : Exome resequencing project</li> <li><input type="checkbox"/> Map : Project that results in non-sequence map data such as genetic map, radiation hybrid map, cytogenetic map, optical map, and etc.</li> <li><input type="checkbox"/> Metagenome : Sequence analysis of environmental samples</li> <li><input type="checkbox"/> Phenotype/Genotype : Project correlating phenotype and genotype</li> <li><input type="checkbox"/> Random Survey : Sequence generated from a random sampling of the collected sample.</li> <li><input type="checkbox"/> Targeted Locus (Loci) : Project to sequence specific loci, such as 16S rRNA sequencing</li> <li><input type="checkbox"/> Transcriptome or Gene Expression : Large scale RNA sequencing or expression analysis. Includes cDNA, EST, RNA-Seq, and microarray.</li> <li><input type="checkbox"/> Variation : Project with a primary goal of identifying large or small sequence variation across populations.</li> <li><input type="checkbox"/> Other <input type="text"/> : A free text description is provided to indicate Other data type</li> </ul>
Sample Scope *	<p>The scope and purity of the biological sample used for the study.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Monoisolate : A single animal, cultured cell-line, inbred population or possibly a heterogeneous population</li> <li><input type="checkbox"/> Multiisolate : Multiple individuals, a population (representation of a species).</li> <li><input type="checkbox"/> Multi-species : Sample represents multiple species.</li> <li><input type="checkbox"/> Environment : Species content of the sample is not known. Nucleic acid is directly isolated from an environmental sample for analysis.</li> <li><input type="checkbox"/> Synthetic : Sample is synthetically created in a laboratory.</li> <li><input type="checkbox"/> Single cell : Single cell sequencing examines the sequence information from individual cells.</li> <li><input type="checkbox"/> Other <input type="text"/> : A free text description is provided to indicate Other data type</li> </ul>

## 5. Umbrella Project 정보[선택사항]

① 연구 과제 정보는 오른쪽 상단의 [Search]을 클릭하여 검색조건으로 해당과제를 찾아 선택하면 자동으로 입력됨

## 6. 논문 및 성과 정보[선택사항]

① 논문의 PubMed ID: 해당 프로젝트의 수행 결과로 산출된 논문의 PubMed ID 입력

② 논문의 DOI: 논문의 PubMed ID가 없을 경우 DOI (Digital Object Identifier) 입력

## 7. BioProject 제출

① 해당 항목을 모두 기입 후 [Save] 버튼을 선택

② 데이터가 저장되었다는 [알림] 팝업창 확인

The image shows a web form for submitting a BioProject. It is divided into two main sections: 'Hierarchy' and 'Publications'.

**Hierarchy Section:** Contains a search bar with a 'Search' button. Below it is a text box for 'Umbrella Project' with a 'Search' button. The text below the search bar reads: "If your project is belonged to some umbrella project, choose correct one. If you intend to submit an umbrella project, please inform us that "this is umbrella project". ###Umbrella projects may be created automatically using a rule-based logic or may be created by database staff upon request or upon identification of a needed grouping. Umbrella projects cannot be kept private. (https://www.ddbj.nig.ac.jp/bioproject/submission-e.html)".

**Publications Section:** Contains two input fields. The first is for 'PubMed IDd' with the label 'Provide a PubMed ID'. The second is for 'DOI' with the label 'Provide a DOI if a PubMed ID is not available'. There are '+ Add' and '- Remove' buttons to the right of the input fields.

**Buttons:** At the bottom of the form are 'Save' and 'Submit' buttons. A red box highlights these buttons, with a blue circle '1' pointing to the 'Save' button.

**Confirmation Dialog:** A red box highlights a confirmation dialog that appears after clicking 'Save'. The dialog contains a green checkmark icon, the text 'Save It's saved.', and an 'OK' button. A blue circle '2' points to the dialog, and a blue circle '3' points to the 'Submit' button.

③ [Submit] 버튼을 클릭하여 BioProject 제출 완료

**Preview of BioProject**

**Submitter**

Name *	Sim young mi
Email *	ymsim@kribb.re.kr
Organization *	test
Department *	test
Address *	125, Gwahak-ro, Yuseong-gu, Daejeon, Korea 34141
Country *	South Korea

---

**Project Design**

NTIS Number	1711044780
Project title *	Single-cell multiomics analysis based biomarker and new target development for immune cell therapy
Relevance *	agricultural
Description *	test
Project Data Type *	Indicate the general label of the primary study goal. <input checked="" type="checkbox"/> Whole Genome sequencing <input type="checkbox"/> Clone ends <input type="checkbox"/> Epigenomics <input type="checkbox"/> Exome <input type="checkbox"/> Map <input type="checkbox"/> Metagenome <input type="checkbox"/> Phenotype/Genotype <input type="checkbox"/> Random Survey <input type="checkbox"/> Targeted Locus (Loci) <input type="checkbox"/> Transcriptome or Gene Expression <input type="checkbox"/> Variation <input type="checkbox"/> Other <input type="text"/>
Sample Scope *	The scope and purity of the biological sample used for the study. <input checked="" type="checkbox"/> Monoisolate <input type="checkbox"/> Multisolate <input type="checkbox"/> Multi-species <input type="checkbox"/> Environment <input type="checkbox"/> Synthetic <input type="checkbox"/> Single cell <input type="checkbox"/> Other <input type="text"/>

**Hierarchy**

Umbrella project	
------------------	--

**Publications**

PubMed IDd	<input type="text"/> Provide a PubMed ID
DOI	<input type="text"/> Provide a DOI if a PubMed ID is not available

**Save**  
It's been submitted.

1



④ 제출된 BioProject는 MyPage에서 확인 가능

\* BioProject의 승인 진행 사항을 확인 할 수 있음

The screenshot shows a user's MyPage dashboard. At the top, it says "MyPage" and "Hello, ymsim". Below this, there's a section for "Registered Assignments" with a table:

Assignment	All	Writing	Submit	Return	Approval
BioProject	All 1	Writing 0	Submit 1	Return 0	Approval 0
BioSample	All 1	Writing 1	Submit 0	Return 0	Approval 0
KAD	All 1	Writing 1	Submit 0	Return 0	Approval 0

Below this is a "Data Request" section with another table:

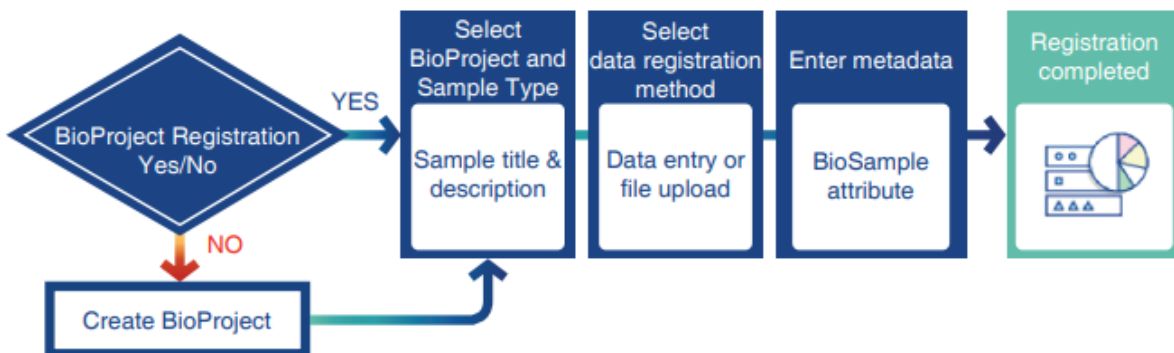
Request	All	Submit	Return	Approval	Destruct
KAD	All 0	Submit 0	Return 0	Approval 0	
Reviewer	All 0	Approval 0	Destruct 0		

## 2-4. Submit sample information (BioSample)

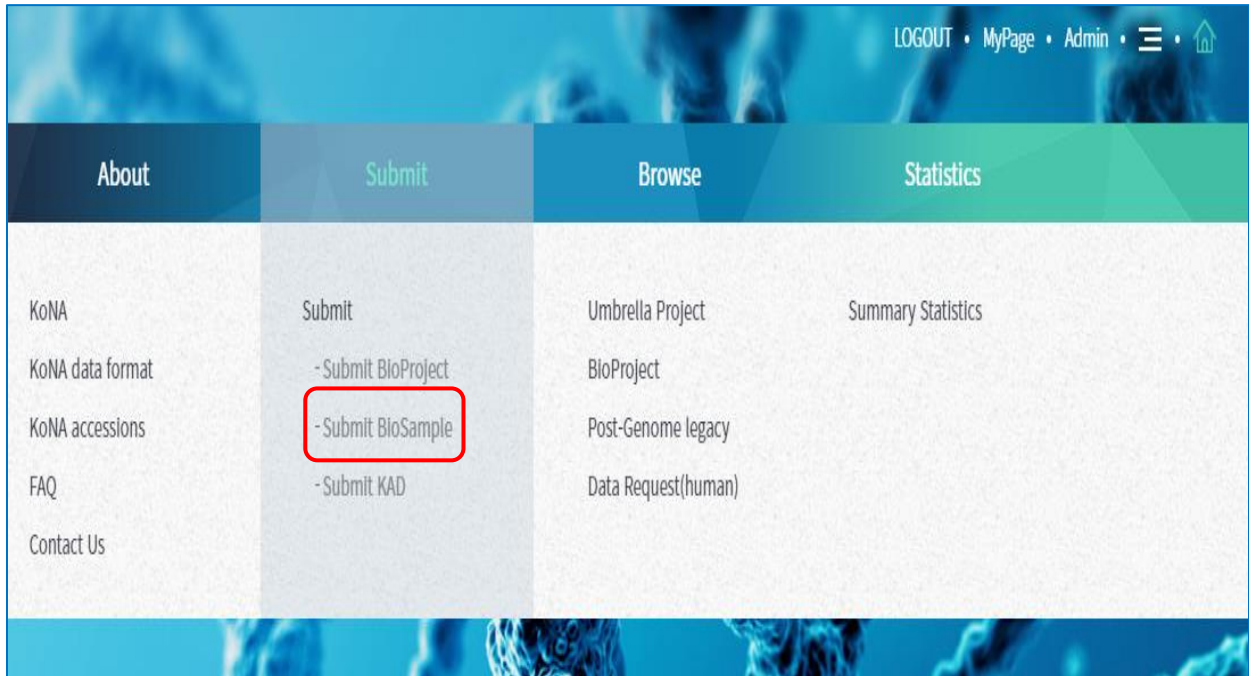
### ◎ 기본 샘플 정보 (BioSample)

생물 유래 샘플 대상으로 실험하여 얻은 실험 데이터를 등록하기 전에 필수 등록해야 하는 정보로서 그 샘플에 대한 개괄적인 정보를 등록하기 위해 사용, 샘플 (BioSample) 정보 작성 및 제출

#### • BioSample 작성단계



1. 데이터 등록을 위하여 KoNA 메인페이지(<https://www.kobic.re.kr/kona/>) 상단 메뉴의 [Submit » Submit



2. Create 메뉴를 눌러 새로운 BioSample 페이지를 활성화 한다.

## Submit BioSample

...

BioSample is a sample repository where you can search, submit and curate sample metadata used in various projects.  
 The BioSample is also being developed to capture descriptive information about the biological samples investigated in projects.  
 BioProject and BioSample records link to corresponding data stored in KoNA.  
 The KoNA BioSample issues accession numbers with the prefix 'KAS' to the submitted sample(s), and with the prefix "KASG" to the submitted sample group.

Total : 1 / Page 1  
 Submit 0 | Return 0 | Approval 0

Create

NO	Sample Group Accession ID	BioProject Accession ID	Sample Title	Organism	Release Date	Status	Operation
1	KASG231005	KAP230591	test	Clinical or host-associate...	-	Writing	<div style="display: flex; gap: 5px;"> <div style="background-color: #333; color: white; padding: 2px 5px;">Edit</div> <div style="background-color: #333; color: white; padding: 2px 5px;">Delete</div> </div>

### 3. 프로젝트 및 샘플설명 작성

※ BioProject 제출을 선행하지 않은 경우, BioProject 등록 절차에 따라 작성 및 등록 진행

① BioProject 제출 후, [Search] 버튼을 클릭하고 창이 나오면 앞서 제출 또는 등록된 BioProject 중, 샘플 등록을 위한 BioProject 를 선택

② 등록하고자 하는 샘플의 공개 날짜를 선택, 즉시 공개와 공개 일자 지정을 할 수 있음

③ Sample Design에 Sample에 대한 Title과 설명을 작성

④ 등록하고자 하는 Sample type을 선택

▶ Sample type (Human Sample 일 경우 신청 한 IRB를 함께 등록 해야 함)

Type	Category
Sample type	Clinical or host-associated pathogen
	Environmental, food or other pathogen
	Microbe
	Model organism or animal sample
	Human sample
	Plant sample
	Virus sample

## Submit BioSample

...

Please fill the below fields to describe your BioSamples.

Required \* / Conditionally required \*

**Hierarchy**

BioProject Accession ID \*  Search

**Date**

Release Date Selection \*  Release immediately following curation(recommended)  Release on specified date

**Sample Design**

Sample Title \*

Description \*

Sample Type \*

Select appropriate types that best describes your samples :

- Clinical or host-associated, pathogen
- Environmental, food or other pathogen
- Microbe : includes bacteria or other unicellular microbes that are not appropriate to to Pathogen or Virus types.
- Model organism or animal sample : includes multicellular samples or cell lines derived from common laboratory model organisms.
- Human sample
- Plant sample : includes any plant sample or cell line.
- Virus sample

Upload IRB \*

Please upload the IRB provision approval (institutional free form) as a PDF file.

Open

Applicable  Not Applicable

Sample Type *	<p>Select appropriate types that best describes your samples :</p> <p><input type="radio"/> Clinical or host-associated, pathogen</p> <p><input type="radio"/> Environmental, food or other pathogen</p> <p><input type="radio"/> Microbe : includes bacteria or other unicellular microbes when it is not appropriate or advantageous to Pathogen or Virus types.</p> <p><input type="radio"/> Model organism or animal sample : includes multicellular samples or cell lines derived from common laboratory model organisms.</p> <p><input checked="" type="radio"/> Human sample</p> <p><input type="radio"/> Plant sample : includes any plant sample or cell line.</p> <p><input type="radio"/> Virus sample</p>
Upload IRB *	<p>Please upload the IRB provision approval (institutional free form) as a PDF file.</p> <div style="border: 1px solid red; padding: 5px; display: inline-block;"> <input type="text"/> <input type="button" value="Open"/> </div> <p><input checked="" type="radio"/> Applicable <input type="radio"/> Not Applicable</p>

⑤ 샘플 등록에 대한 약관을 꼼꼼히 읽은 후 동의란에 체크

	<p style="border: 1px solid red; padding: 2px;"><b>Agreement on data deposition</b></p> <p>1. The principal investigator (a person who submit the data to KoNA will be hereinafter referred to as "principal investigator") agrees to provide genome research data free of charge to Korean Nucleotide Archive (hereinafter abbreviated as "KoNA") of Korea Bioinformatics Center (hereinafter abbreviated as "KOBIC"). This does not imply a transfer of patents or other intellectual property rights of the deposited data.</p> <p>2. Pursuant to Articles 16, 18, 37, and 38 of the "Bioethics and Safety Act", when a human subject of research or donor of a human material gives written consents to providing his/her personal information to a third party (a person who browses, downloads, and uses the data deposited in KoNA will be hereinafter referred to as "third party"), the relevant principal investigator (relevant human subjects researcher or human materials researcher in the Act) can submit the personal genome data to KoNA, subject to examination thereof by the competent institutional review board (hereinafter abbreviated as "IRB").</p> <div style="border: 1px solid gray; padding: 5px; margin: 5px 0;"> <p style="text-align: center;"><b>Bioethics and Safety Act</b></p> <p><b>Article 2 (Definitions)</b> The terms used in this Act shall be defined as follows:</p> <p>18. The term "personal information" means information about an individual, such as personally identifiable information, genetic information, or information about health.</p> <p><b>Article 16 (Consent to Research on Human Subjects)</b> (1) A human subjects researcher shall obtain written consent (including consent by an electronic document; hereinafter the same shall apply) regarding the following matters from human subjects of research before commencing a human subjects research project:</p> <ol style="list-style-type: none"> <li>2. Duration, procedure for, and methods of participation of human subjects of research;</li> <li>3. Foreseen risks and benefits to human subjects of research;</li> <li>4. Protection of personal information;</li> <li>5. Compensation for losses incurred through participation in the research project;</li> <li>6. Provision of personal information;</li> <li>7. Withdrawal of consent;</li> <li>8. Other matters the competent institutional committee deems necessary.</li> </ol> <p><b>Article 18 (Provision of Personal Information)</b> (1) When a human subject of research gives written consents to providing his/her personal information to a third party pursuant to Article 16 (1), the relevant human subjects researcher may provide his/her personal information to a third party, subject to examination thereof by the competent institutional committee.</p> </div> <p>principal investigator by registered e-mail: the e-mail address shall be registered according to the relevant procedure on the KoNA website (<a href="https://www.kobic.re.kr/kona">https://www.kobic.re.kr/kona</a>).</p> <p>4. The principal investigator shall designate the release date of the submitted data in KoNA during the submission on the website. BioProject shall be released immediately, while the other metadata shall be released after the designated release date. However, when the principal investigator publishes a research article relevant to the submitted genome data, the data shall be released immediately in KoNA regardless of the designated release date.</p> <p>5. The principal investigator has the duties and responsibilities for the submitted data, as described below:</p> <p>(A) The principal investigator shall guarantee that the submitted data is actual information generated through experiment and data processing and is not arbitrarily manipulated or incorrect information.</p> <p>(B) The principal investigator has a competent authority to modify and change the submitted genome data and relevant metadata. However, KOBIC may modify and change the data and metadata based on the request of the principal investigator.</p> <p>6. A third party can produce the secondary data from the submitted genome with metadata, however, the third party shall not claim ownership or demand fees for using the secondary data.</p> <p>7. KoNA is not responsible for any direct or indirect losses that may arise from the submitted data.</p> <p>8. Data submission may be suspended or canceled in the following cases:</p> <ul style="list-style-type: none"> <li>- Insufficient or incorrect contents in the submitted data and metadata</li> <li>- Submission of a damaged file or a file which are not matched to the submitted metadata on the website</li> <li>- Insufficient documents required for the data submission</li> <li>- Submission of the data which has been previously submitted to two or more domestic Information Centers of Bio-Resources for Research in Korea.</li> <li>- Other cases when KoNA curators found a significant problem in the submitted data</li> </ul> <p style="border: 1px solid red; padding: 2px;"><input type="checkbox"/> I have thoroughly understand the contents of the consent form above and agree to submit the genome data with the relevant metadata in accordance with the above conditions.</p>
--	---

#### 4. 데이터 등록방법

① Sample Data Input Type 의 [직접 입력] 또는 [파일 업로드] 선택

1) [직접 입력] 다음 단계인 메타데이터 입력 시 하단 Spread sheet 에 직접 입력하는 방식으로 처음 작성하는 경우나 데이터가 많지 않은 경우에 적합

2) [파일 업로드 방식] 다음 단계 메타데이터 입력 시 Spread sheet 파일 형태의 양식 파일을 다운로드하여 작성 후 업로드하는 방식으로 많은 데이터를 입력하는 경우에 적합

#### 5. 메타데이터

① 각 항목에 대한 설명과 예시에 따라 작성하고 \*필수 항목은 반드시 작성

\*M : 필수항목, O : 선택 항목

② [직접 입력]의 경우 하단 Spread sheet 에 작성해야 할 정보를 항목별로 확인 후, 빠짐없이 작성하고 아래의 [Save] 클릭하여 저장됨

③ 오른쪽의 [Submit] 버튼을 클릭하여 BioSample 제출 완료

※ 제출을 진행하면 품질관리자 반려 전에는 수정 및 삭제가 불가능

Sample Data Input Type \*

Direct Input

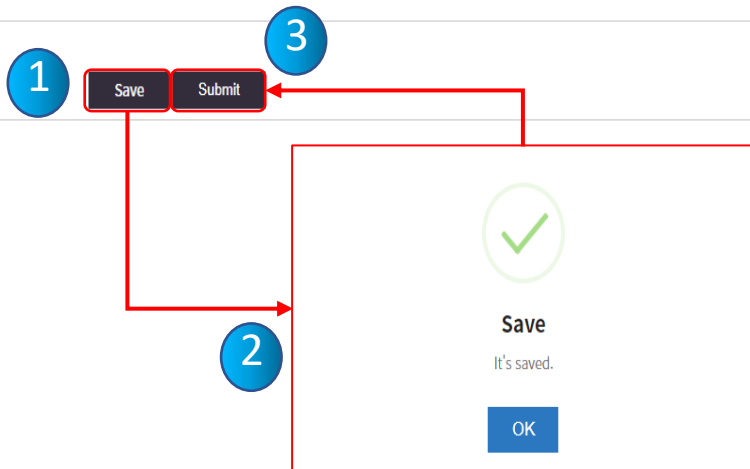
Direct Input

File Upload

Upload Results \*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	1.Name / Designation							2.Sample history						
2	Sample name	Organism	Strain	Subgroup	Subtype	Culture collection	Specimen voucher	Biological replicate	Collected by	Collection date	Identified by	Passage history	Isolation source	Isolate Geo
3	M	M	O	O	O	O	O	O	M	M	O	O	M	O
4	test	test							test	test			test	test
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														

Add Column Save Input Results



④ [파일 업로드]의 경우 [Download form]을 클릭하여 양식을 다운받아 작성하고 [Open]을 클릭하여 작성한 양식을 업로드한 후 [Submit]을 클릭

Sample Data Input Type \*

File Upload ▾

	Entity attribute	Development stage	Developmental stage at the time of sampling	<input type="checkbox"/>
		Disease	List of diseases diagnosed; can include multiple diagnoses	<input type="checkbox"/>
		Disease stage	Stage of disease at the time of sampling	<input type="checkbox"/>
		Health state	Health or disease status of sample at time of collection	<input type="checkbox"/>
		Phenotype	Phenotype of sampled organism	<input type="checkbox"/>
		Population	Population is a summation of all the organisms of the same group or species, which live in a particular geographical area. For human: ; for plants: filial generation, number of progeny, genetic structure	<input type="checkbox"/>
		Race	Race of sample. Race refers to a person's physical characteristics, such as bone structure and skin, hair, or eye color.	<input type="checkbox"/>
		Ethnicity	Ethnicity of the subject. Ethnicity refers to cultural factors, including nationality, regional culture, ancestry, and language.	<input type="checkbox"/>

Save

Reset

---

**Step 2 : Download metadata template.**

1

Download form

Download Input sample

**Step 3 : Upload your metadata.**

Open

**Step 4 : Please submit the attached form finally.**

Submit

⑤ Submit 클릭 후 성공적으로 업로드가 되면, 하단 Spread sheet에 업로드한 파일의 정보가 입력됨. 내용 확인 후 Save 및 Submit를 클릭하여 BioSample 제출을 완료

Step 2 : Download metadata template.

Download form   Download Input sample

Step 3 : Upload your metadata.

C:\fakepath\2\_BioSample\_7\_AF   Open

2\_BioSample\_7\_ARD.xlsx   Remove

Step 4 : Please submit the attached form finally.

Submit 1

Upload Results \* 2

1	BioSample 정보							
2	생물군 종류	샘플명	나이		생물질 제공자	세포주명	세포 타입	수집가
3	taxonomicType	sampleName	age	age_unit	biomaterialProvider	cellLine	cellType	collector
7	Human	ARD-GARDWGSN00004	NA	NA	NA	NA	NA	
8	Human	ARD-GARDWGSN00005	NA	NA	NA	NA	NA	
9	Human	ARD-GARDWGSN00007	NA	NA	NA	NA	NA	
10	Human	ARD-GARDWGSN00008	NA	NA	NA	NA	NA	
11	Human	ARD-GARDWGSN00009	NA					
12	Human	ARD-GARDWGSN00011	NA					
13	Human	ARD-GARDWGSN00012	NA					
14	Human	ARD-GARDWGSN00013	NA					
15	Human	ARD-GARDWGSN00016	NA					
16	Human	ARD-GARDWGSN00019	NA					
17	Human	ARD-GARDWGSN00020	NA					
18	Human	ARD-GARDWGSN00021	NA					
19	Human	ARD-GARDWGSN00022	NA					
20	Human	ARD-GARDWGSN00023	NA					
21	Human	ARD-GARDWGSN00024	NA					

Save It's saved. OK 3

Save It's been submitted. OK 4

save Submit

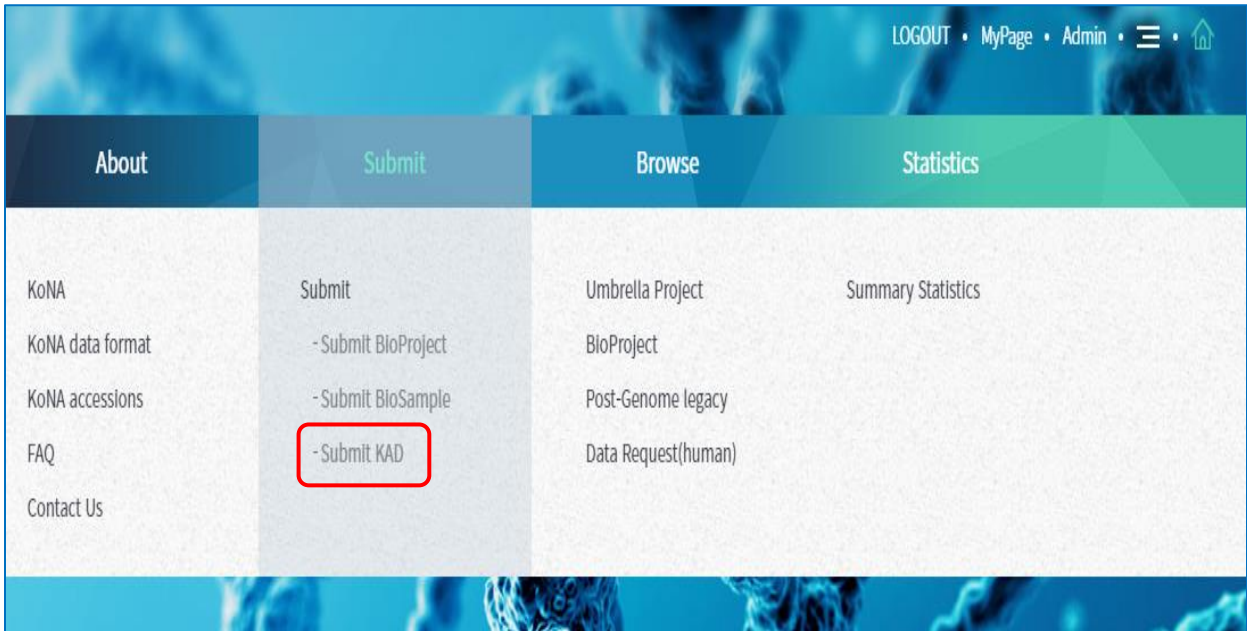
Add Column   Save Input Results



## 2-5. Submit NGS data (KRA)

### ◎ 실험 데이터 작성 및 등록

1. 데이터 등록을 위하여 KoNA 메인페이지(<https://www.kobic.re.kr/kona/>) 상단 메뉴의 [Submit » Submit KAD] 페이지로 이동



2. Create 메뉴를 눌러 새로운 KAD 페이지를 활성화 한다.

### Submit KAD

...

The Korean Read Archive (KAD) is a data repository for collecting, archiving, managing, and sharing raw NGS sequence data.  
The KAD is the first repository of genome sequence data with international journal recognition in Korea.  
Before creating a new KAD submission, you must create a BioProject and BioSample(s).  
If you have any questions or would like to give us any suggestions/comments or report a bug, please feel free to contact us: [data@kobic.re.kr](mailto:data@kobic.re.kr)

Total : 1 / Page 1  
Submit 0 | Return 0 | Approval 0 Create

NO	BioProject Accession ID	Sample Group Accession ID	KAD Accession ID	Submission Date	Registration Date	Reviewer	Status	Operation
1	KAP230591	KASG231005	KAD2301064	-	-	-	Writing	<span>Edit</span> <span>Delete</span>

3. 실험 데이터 입력에 앞서 해당 BioProject와 BioSample 선택

① [Open] 버튼을 클릭하고 창이 나오면 앞서 제출 또는 등록된 BioProject와 BioSample 선택

② 등록하고자 하는 실험 데이터의 공개 날짜를 선택, 즉시 공개와 공개 일자 지정을 할 수 있음

③ [데이터 등록방법]을 선택 [데이터 등록 방법]에는 다음과 아래와 같은 방법이 있음

- 1) 직접 입력 : 실험데이터 경로와 메타데이터를 웹 양식에 직접 입력하는 방식
- 2) 파일 업로드 : 여러 개의 실험데이터경로와 메타데이터를 한 번에 파일로 업로드하는 방식

## Submit KAD

Required \* / Conditionally required \*

**Hierarchy**

BioProject Accession ID \*

Open

Select the Project ID related to the sample. Link data to BioProject that describes the research

BioSample Group Accession \*

Open

Enter a BioSample or KAD Sample Accession. BioSample accessions have 'KAS' prefix. KAD Sample Accessions have 'KDS' prefix. A BioSample describes the biological source material.

**Submit KAD**

Formatting cannot be used in the spreadsheet. Please re-submit.

input type \*

Direct Input

Project accession

Select	Project title (English)	Registration date
<input type="radio"/>	Single-cell multiomics analysis based biomarker and new target development for immune cell the...	-

Select

<< < 1 > >>

Upload Results \*

	A	B	C	D	E	F	G	H	I	J	K	L
1	Experiment											
2	Experiment Design											
3	Sample name	Experiment title	Library name	Instrument model	Platform	Library Construction/Experiment Design	Library strategy	Source	Selection	General description	Fragment/ Paired read	Inte
4	M	M	M	M	M	M	O	M	M	M	O	O
5												
6												
7												
8												
9												
10												
11												
12												
13												
14												
15												
16												
17												
18												

Add Column

Save

Release date selection \*

Release immediately following curation(recommended)
  Release on specified date

#### 4. 메타데이터 작성 및 제출

##### 1) 데이터 등록 방법에서 “직접 입력”을 선택한 경우

The screenshot shows the 'Submit KAD' interface. At the top, there is a message: 'Formatting cannot be used in the spreadsheet. Please re-submit.' Below this, there is a dropdown menu for 'input type' set to 'Direct Input'. The main area is a spreadsheet titled 'Upload Results \*' with columns A through L. The spreadsheet has a header row for 'Experiment Design' and a data row with 'M' and 'O' values. Below the spreadsheet, there are two 'Save' buttons: one labeled 'Save' and 'Submit' at the bottom, and another labeled 'Add Column' and 'Save' on the right. Two red boxes highlight the 'Save' buttons, with arrows pointing to them from the numbered circles 1 and 2. A 'Release date selection' section is at the bottom left, with radio buttons for 'Release immediately following curation(recommended)' (selected) and 'Release on specified date'. A 'Save' dialog box is also visible, showing a green checkmark and the text 'Save It's been submitted.' with an 'OK' button.

① 빅데이터 고속 전송 시스템(\*GBox)를 통해 KoNA스토리지로 실험데이터 파일을 업로드

\* GBox 다운로드, 설치 및 데이터 업로드에 대한 상세한 설명은 Appendix의 "GBox user guide" 참고

② 메타데이터 작성 항목은 \*필수항목과 선택항목으로 이루어져 있으며, 필수항목에 빠짐없이 실험 정보를 입력 후 저장하면 하단 Spread sheet 창에 적용

\*M : 필수항목, O : 선택 항목

③ Gbox를 다운로드 받아 해당 폴더에 옮긴 실험데이터의 경로를 하단 Spread sheet의 path 부분에 작성

- ④ Run 부분에 추가 되는 파일이 있을 경우, [Add Column]을 클릭하여 OTHER column 옆에 새로운 column이 생성 추가 파일을 입력 할 수 있음.
- ⑤ 스프레드시트 밑에 Save 버튼을 눌러 작성한 내용을 저장
- ⑥ 공개일자 지정 후 창 하단의 [Save]을 클릭하여 전체적인 내용을 저장 후 [Submit] 클릭하여 제출 완료

2)데이터 등록 방법에서 “파일 업로드”를 선택한 경우

**Submit KAD**

Input Type \* File Upload

Proceed with Sample typeUpload는 in thr following order.

**Step 1 : Select data entry form.**

- To enter data, check 'Enter or not'.
- If you have a field that you want to keep data private, please select 'Private status'.

After selecting all, select 'Save'. If not saved, the selection information will not reflected.

카테고리	필드	M/O	설명	입력여부
sample	Sample name	M	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.	
	Experiment title (English)	M	"Experiment title. Short description that will identify the dataset on public pages. A clear and concise formula for the title would be like:{methodology} of {organism}; {sample info} (e.g. ""RNA-Seq of Mus musculus: adult female spleen"")"	
Experiment Design	Experiment title (Korean)	M	데이터베이스 공개 페이지에서 나타낼 실험 제목. 제목에 종을 포함한 샘플의 실험 정보와 사용한 시퀀싱 타임을 간략히 표현하는 것을 추천함. (예시: "성체 쥐의 비장에 대한 RNA-Seq")	
	Library name	M	Short unique identifier for the sequencing library. Each library name MUST be unique! (Exception: libraries of the technical replicates are allowed to have the same library name)	
	Platform	M	Sequencing platform	
	Instrument model	M	Sequencing platform 중 Instrument model 부분	
	Library Construction/Experiment Design	M	Enter the details about your experimental design and molecular strategies including hybrid selection and affinity capture reagents; any detail that distinguishes your experiment from other similar experiments. This field should describe: - the protocols used to extract and prepare the material to be sequenced - the library construction protocol - name of the library preparation kit	

- ① 빅데이터 고속 전송 시스템(\*GBox)를 통해 KoNA스토리지로 실험데이터 파일을 업로드  
 \* GBox 다운로드, 설치 및 데이터 업로드에 대한 상세한 설명은 Appendix의 " GBox user guide " 참고
- ② [파일 업로드]의 경우 [Download form]을 클릭하여 양식을 다운로드 받아 실험 정보를 작성하고 [Open]을 클릭하여 작성한 파일을 업로드한 후 [Submit]을 클릭  
 성공적으로 업로드가 되면, 하단 Spread sheet에 업로드한 파일의 정보가 입력됨.

Other  +

Save Reset

---

**Step 2 : Download metadata template.**

Download form Download input sample

---

**Step 3 : Upload your metadata.**

Open

KRA\_test\_3.xlsx Remove

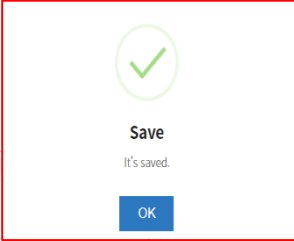
---

**Step 4 : Please submit the attached form finally.**

Submit

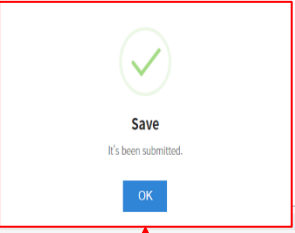
Upload Results \*

Experiment												
Experiment Design												
Sample name	Experiment title	Library name	Instrument model	Platform	Library Construction	Experiment Design	Library	Strategy	Source	Selection	General description	Fragment/ Paired read
M	M	M	M	M			O	M	M	M	O	O
OB_S1_B1	scRNA-seq of IOB_S1_B1_DNBSEQ-G400FMGI	IOB_S1_B1_DNBSEQ-G400FMGI	10X Chromium_5p	10X Chromium_5p				RNA-Seq	TRAN	Poly-A		Paired-end
OB_S1_B1	scRNA-seq of IOB_S1_B1_DNBSEQ-G400FMGI	IOB_S1_B1_DNBSEQ-G400FMGI	10X Chromium_5p	10X Chromium_5p				RNA-Seq	TRAN	PCR	TCR region	Paired-end
OB_S1_B2	scRNA-seq of IOB_S1_B2_DNBSEQ-G400FMGI	IOB_S1_B2_DNBSEQ-G400FMGI	10X Chromium_5p	10X Chromium_5p				RNA-Seq	TRAN	Poly-A		Paired-end
OB_S1_B2	scRNA-seq of IOB_S1_B2_DNBSEQ-G400FMGI	IOB_S1_B2_DNBSEQ-G400FMGI	10X Chromium_5p	10X Chromium_5p				RNA-Seq	TRAN	PCR	TCR region	Paired-end



**Save**  
It's saved.

OK



**Save**  
It's been submitted.

OK

1 Save

2 Submit

**Release date selection \***

Release immediately following curation(recommended)  Release on specified date

"Release immediately following curation (recommended)" OR "Release on specified date"

③ 이때도 직접입력 방식과 마찬가지로 Gbox 폴더에 옮긴 실험데이터의 경로를 다운받은 엑셀 파일의 path 부분에 작성

④ 제출된 실험 정보를 Spread sheet 창에서 다시 한번 검토하고 공개 일자 지정 후 창 하단의 [Save]을 클릭하여 전체적인 내용을 저장 후 [Submit] 클릭하여 제출 완료

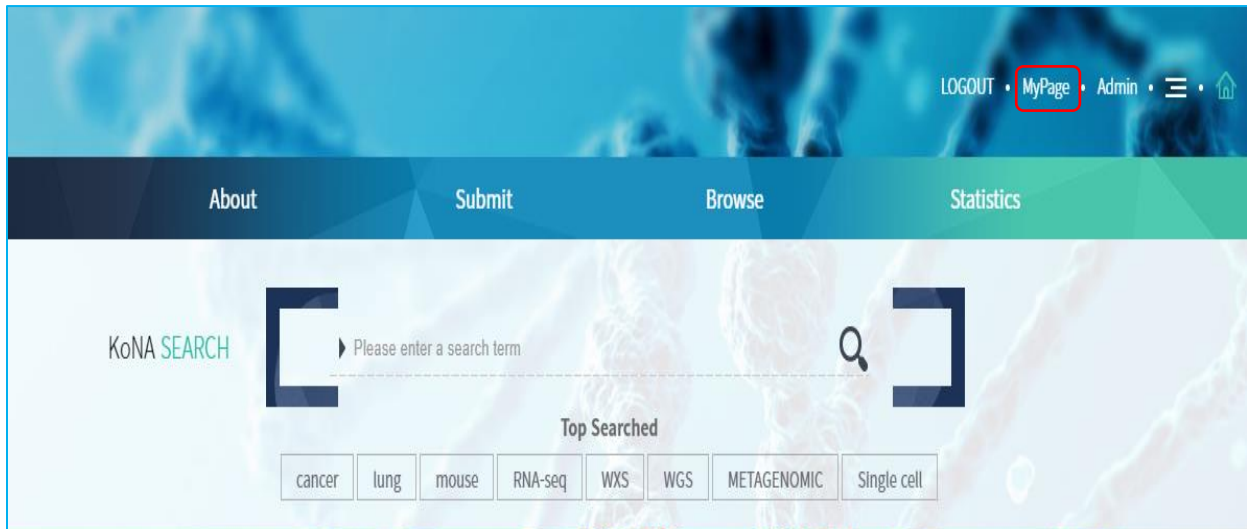
## 2-6 Manage my submission (MyPage)

### ◎ 제출 이후 절차 안내



### ◎ 제출 후 검토 단계

▶ 제출 이후 진행 과정은 [MYPAGE]에서 확인 가능




▶ 검토 단계는 품질관리자가 제출된 데이터를 검수

▶ 검토 단계를 통과하면 '승인'되어 등록이 가능한 상태로 변경되고, 통과하지 못한 데이터에 대해서는 '반려' 처리가 되어 제출자가 수정하여 재제출 필요

## MyPage

...



Hello, ymsim

Thank you for visiting KoNA. You can move directly through the main menu.

[My Info](#)

---

### Registered Assignments

BioProject	All 0	Writing 0	Submit 0	Return 0	Approval 0
BioSample	All 0	Writing 0	Submit 0	Return 0	Approval 0
KRA	All 0	Writing 0	Submit 0	Return 0	Approval 0

### Data Request

KRA	All 0	Submit 0	Return 0	Approval 0
Reviewer	All 0	Approval 0	Destruct 0	

## ◎ 등록 이후

검토 이후 최종 등록이 완료되면, 등록번호가 부여되고, 입력한 공개 일자에 데이터는 공개됨

# III DATA SEARCH AND DOWNLOAD

## 3-1. Search Umbrella BioProject

1. KoNA 메인페이지(<https://www.kobic.re.kr/kona/>) 상단 메뉴의 [Browse » Umbrella Project] 페이지로 이동, 등록된 Umbrella Project를 확인 할 수 있음

The screenshot shows the KoNA website interface. At the top, there is a search bar with the text "Please enter a search term" and a search icon. Below the search bar, there are several tabs: "cancer", "lung", "mouse", "RNA-seq", "WXS", "WGS", "METAGENOMIC", and "Single cell". To the right of the search bar, there are links for "LOGOUT", "MyPage", "Admin", and a home icon. Below the search bar, there are four main navigation tabs: "About", "Submit", "Browse", and "Statistics". The "Browse" tab is selected and highlighted in green. Under the "Browse" tab, there is a list of options: "Umbrella Project" (highlighted with a red box), "BioProject", "Post-Genome legacy", and "Data Request(human)".

Below the navigation tabs, there is a section titled "Umbrella Project" with a description: "An umbrella project provides an organizational structure to a large collaborative project and group projects related via funding or submitting sources or collaboration. The umbrella project is created upon the request of the submitter, a funding agency, or by KOBIC staff to group multiple projects that are part of a large initiative or collaboration or funding source. Umbrella projects is indirectly connected to data through the linked primary submission projects. For example, Umbrella projects reflect the general organizational structure of the Korea Post-Genome Project. If you have any questions about umbrella project(s), please contact us: [data@kobic.re.kr](mailto:data@kobic.re.kr)."

Below the description, there is a message: "There are a total of 1 data registered in the "Umbrella project"".

At the bottom of the page, there is a table with the following information:

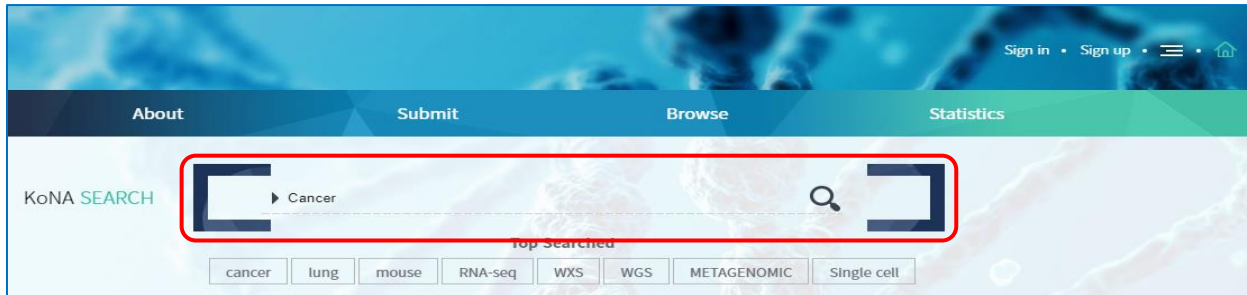
<b>Umbrella Project ID :</b> <a href="#">KAU200000</a>	<b>Registration Date :</b> 2020-10-16
<b>Umbrella Title :</b> Korea Post-Genome Project	
<b>Description :</b> Korea Post-Genome Project has launched to achieve the following aims: 1. Development of diagnostic methods enabling precision medicine; 2. Development of bioresources facilitating genome information of animal, plant, marine species; 3. Establishment of the standard protocol for genome analysis technology; and 4. Collaboration between government ministries to study disease mechanism, host-microbe interaction, and human reference genome as well as cwork with global consortium and train experts in genomics and bioinformatics	



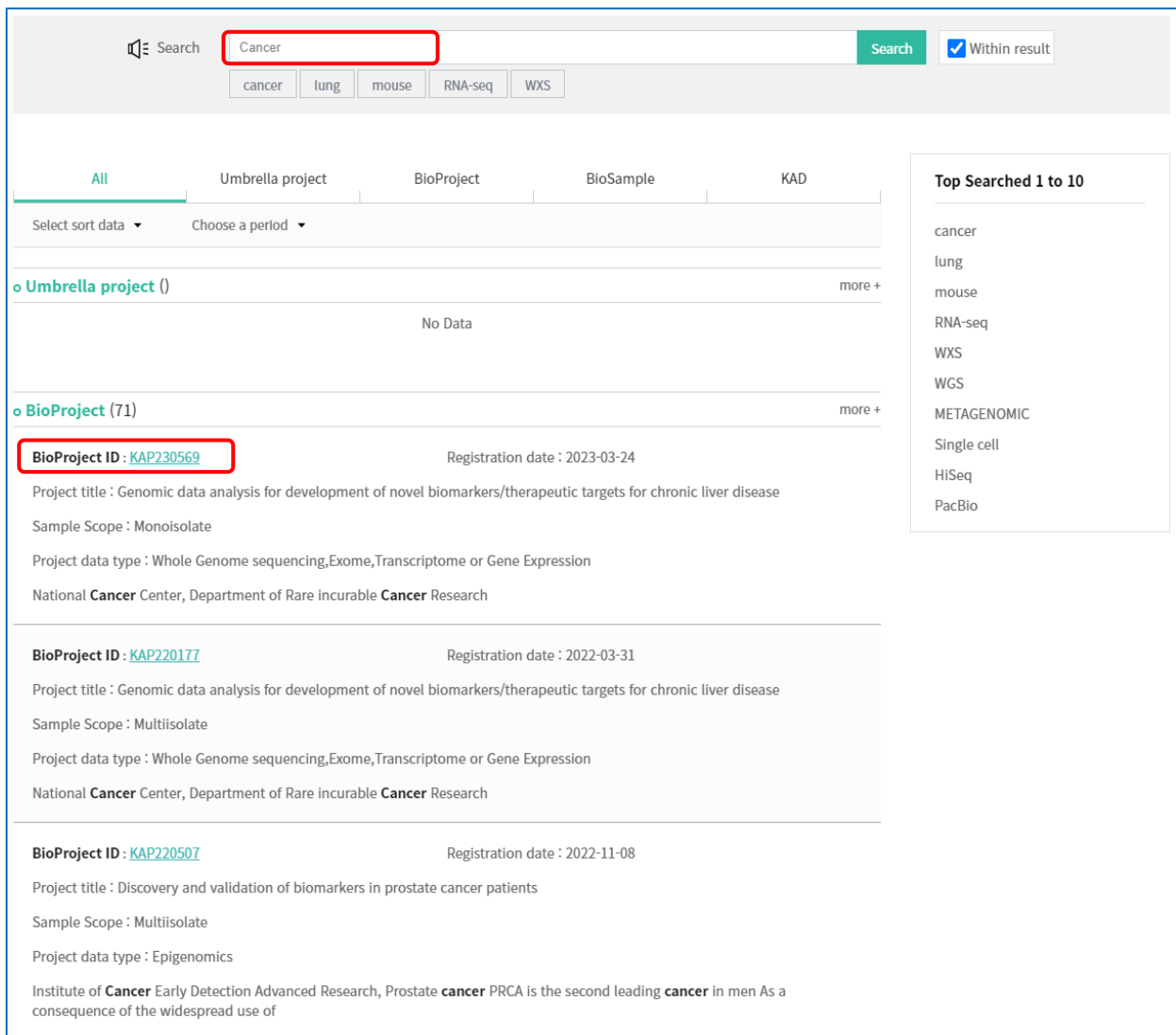
## 3-2. Search BioProject

### ◎ 메인페이지 통합 검색

1. 메인페이지의 검색 바에서 원하는 검색어(예: cancer)를 입력 후, 우측 아이콘을 클릭



2. 입력한 검색어에 대한 BioProject 와 BioSample Group, KAD 에서 원하는 데이터를 선택



o BioSample Group (136)

more +

Sample Group accession ID : [KASG210092](#)

Registration date : 2021-09-08

Sample title : microRNA sequencing data generated from ovarian cancer stem cells and ovarian non-cancer stem cells

Sample type : Human sample

microRNA sequencing data generated from ovarian **cancer** stem cells and ovarian non-**cancer** stem cells, 난소암 **cancer** stem cells과 non-**cancer** stem cells에서 생산한 마이크로알엔에이 시퀀싱 데이터, To detect differentially expressed microRNAs between ovarian **cancer** stem cells and ovarian non-**cancer**, 마이크로알엔에이 시퀀싱 데이터 분석을 통하여 난소암 **cancer** stem cell과 non-**cancer** stem cell에서 차별발현되는 마이크로알엔에이를 발굴함

Sample Group accession ID : [KASG220245](#)

Registration date : 2022-04-12

Sample title : Whole genome sequencing of lung cancer

Sample type : Human sample

Whole genome sequencing of lung **cancer**, Whole genome sequence of **cancer** tissues and surrounding normal tissues of lung **cancer** patients

Sample Group accession ID : [KASG220268](#)

Registration date : 2022-04-12

Sample title : Whole genome sequencing on lung cancer

Sample type : Human sample

Whole genome sequencing on lung **cancer**, Whole genome seq analysis of **cancer** tissues and surrounding normal tissues of lung **cancer** patients

o KAD (95)

more +

KAD accession ID : [KAD2000013](#)

Registration date : 2020-12-16

Platform : BGISEQ

Strategy : WGS

Selection : other (not size selecti...

Extraction of ctDNA from blood samples of patients with triple negative breast **cancer** 2.

KAD accession ID : [KAD2100122](#)

Registration date : 2021-09-08

Platform : ILLUMINA

Strategy : smRNA-Seq

Selection : size fractionation

curcumin-untreated SK-OV-3 **cancer** stem cells\_miRNA-Seq,curcumin-treated SK-OV-3 **cancer** stem cells\_miRNA-Seq, curcumin을 처리하지 않은 SK-OV-3 **cancer** stem cell의 miRNA,curcumin을 처리한 SK-OV-3 **cancer** stem cell의 miRNA,curcumin을

KAD accession ID : [KAD2100074](#)

Registration date : 2021-03-22

Platform : ION\_TORRENT,Illumina

Strategy : Targeted-Capture

Selection : PCR,Hybrid Selection

The development of kit for screening of **cancer** related target genes with high sensitivity in the liquid, Lung **Cancer** Panel-Targeted Sequencing (LCPT-Seq), Lung **Cancer** Panel-Targeted Sequencing (LCPT-Seq)

3. 선택한 데이터에 대한(공개일자가 지난데이터) 메타데이터(웹상 엑셀파일로 바로 다운로드) 및 실험데이터의(Gbox 사용) 다운로드가 가능

### Search KAD

...

Submission : 2020-12-16 | BioSample Group ID : KASG200023

#### KAD

BioProject ID	KAP200001
Sample Group Accession ID	KASG200023
KAD Accession ID	KAD2000013
Registration Date	2020-12-16
Platform	DNBSEQ-G400
Strategy	NA
Selection	GENOMIC
Release Date Selection	specified
Release Date	2020-12-31

#### Submit KAD

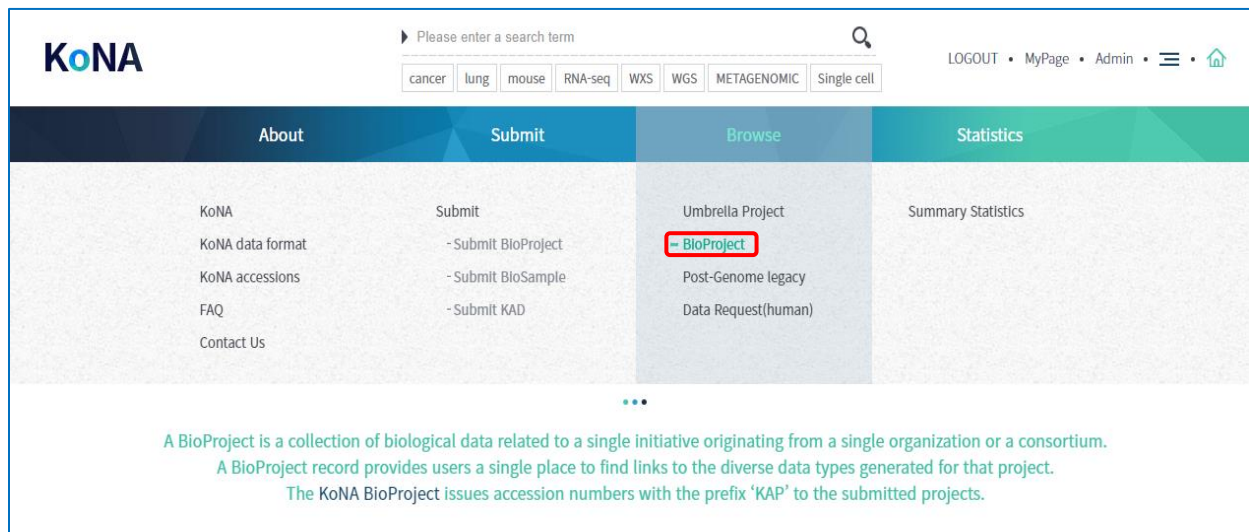
Full Screen

	A	B	C	D	E	F	G	H	I	J	K	
1	Experiment											
2	Experiment Design											
3	Sample name	Experiment title (영문)	Experiment title (한글)	Library name	Instrument mode	Platform	Library Construction/Experiment Design	Library	Strategy	Source	Selection	General de
4	M	M	M	M	M	M	M	O	M	M	M	
5	TNBC001	NA from blood of TNBC pat	유방암 환자 혈액	TNBC001_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
6	TNBC003	NA from blood of TNBC pat	유방암 환자 혈액	TNBC003_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
7	TNBC005	NA from blood of TNBC pat	유방암 환자 혈액	TNBC005_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
8	TNBC006	NA from blood of TNBC pat	유방암 환자 혈액	TNBC006_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
9	TNBC007	NA from blood of TNBC pat	유방암 환자 혈액	TNBC007_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
10	TNBC008	NA from blood of TNBC pat	유방암 환자 혈액	TNBC008_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
11	TNBC009	NA from blood of TNBC pat	유방암 환자 혈액	TNBC009_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
12	TNBC010	NA from blood of TNBC pat	유방암 환자 혈액	TNBC010_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
13	TNBC011	NA from blood of TNBC pat	유방암 환자 혈액	TNBC011_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
14	TNBC014	NA from blood of TNBC pat	유방암 환자 혈액	TNBC014_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
15	TNBC015	NA from blood of TNBC pat	유방암 환자 혈액	TNBC015_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
16	TNBC016	NA from blood of TNBC pat	유방암 환자 혈액	TNBC016_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
17	TNBC018	NA from blood of TNBC pat	유방암 환자 혈액	TNBC018_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	
18	TNBC019	NA from blood of TNBC pat	유방암 환자 혈액	TNBC019_GENO	DNBSEQ-G400	BGISEQ	duction of sequencing library using MGIEasy cell	NA	WGS	GENOMIC	selection, r	

Download Excel

## ◎ BioProject 데이터 활용 검색

1. [메인페이지 » Browse» BioProject] 페이지로 이동
2. 주요 과제 중 관심 BioProject 클릭하여 이동
3. 해당 BioProject의 데이터에 대한(공개일자가 지난데이터) 메타데이터(웹상 엑셀 파일로 바로 다운로드) 및 실데이터의(Gbox 사용) 다운로드가 가능



**KoNA** Please enter a search term

LOGOUT • MyPage • Admin •

cancer lung mouse RNA-seq WXS WGS METAGENOMIC Single cell

About Submit Browse Statistics

KoNA  
KoNA data format  
KoNA accessions  
FAQ  
Contact Us

Submit  
- Submit BioProject  
- Submit BioSample  
- Submit KAD

Umbrella Project  
**BioProject**  
Post-Genome legacy  
Data Request(human)

Summary Statistics

...

A BioProject is a collection of biological data related to a single initiative originating from a single organization or a consortium.  
A BioProject record provides users a single place to find links to the diverse data types generated for that project.  
The KoNA BioProject issues accession numbers with the prefix 'KAP' to the submitted projects.

### Browse BioProject

...

A BioProject is a collection of biological data related to a single initiative originating from a single organization or a consortium.  
A BioProject record provides users a single place to find links to the diverse data types generated for that project.  
The KoNA BioProject issues accession numbers with the prefix 'PRJK' to the submitted projects.

There are a total of 457 data registered in the " BioProject"

NO	BioProject Accession ID	Project Title	Sample Scope	Project Data Type	Registration Date
1	KAP210106	Construction of next generation sequencing system center	Monoisolate	Transcriptome or Gene Expression	2021-11-25
2	KAP230581	Multifaceted roles of retrotransposon-fusion RNAs	Multiisolate	Whole Genome sequencing,Epigenomics,Transcriptome or Gene Expression	2023-05-09
3	KAP220287	Development of molecular markers using comparative genomics	Monoisolate	Whole Genome sequencing,Transcriptome or Gene Expression	2022-05-02
4	KAP220455	Marine animal genome analysis research	Multiisolate	Whole Genome sequencing,Transcriptome or Gene Expression	2022-07-27
5	KAP230562	Targeted sequencing data to discover effective antibodies	Synthetic	Targeted Locus (Loci)	2023-03-09
6	KAP220451	Pipeline discovery of useful genetic resources from large-capacity genetic information and development of useful enzymes	Environment	Metagenome	2022-07-27
7	KAP220472	Multi-genomic analysis for biomarker development in colon cancer	Multiisolate	Whole Genome sequencing	2022-10-17
8	KAP220449	Culture Collection of Multifunctional Novel Bacteria with Biocontrol Effects using Microbial Culturomics	Environment	Whole Genome sequencing,Metagenome	2022-07-27
9	KAP230587	Noggin contributes to brain metastatic colonization of lung cancer cells	Monoisolate	Transcriptome or Gene Expression	2023-06-12
10	KAP220214	Genome Analysis of Marine and Fisheries Organisms and Development of Functional Application	Monoisolate	Whole Genome sequencing,Transcriptome or Gene Expression	2022-04-27

« < 1 2 3 4 5 6 7 8 9 10 > »

### 3-3. Search BioSample

#### ◎ BioSample 데이터 활용 검색

1. [메인페이지 » Browse » BioProject] 페이지로 이동
2. 선택한 BioProject 내 사용하고자 하는 Sample 데이터를 클릭하여 이동
3. 해당 Sample Group ID 의 데이터에 대한(공개일자가 지난데이터) 메타데이터(웹상 엑셀파일로 바로 다운로드) 및 실험데이터의(Gbox 사용) 다운로드 가능

Sample Group			
BioSample Accession ID	Project Title	Sample Type	Registration Date
KASG230995	Cancer cell line	Human sample	2023-05-24
KASG230983	ONT whole genome sequencing for colorectal cancer patients	Human sample	2023-05-09

Submitters's Submissions				
BioProject Accession ID	Project Title	Sample Scope	Project Data Type	Registration Date
KAP220480	Single cell transcriptome based biomarker development in colorectal cancer	Multiisolate	Whole Genome sequencing, Transcriptome or Gene Expression	2022-10-17
KAP220465	Single-cell multiomics analysis based biomarker and new target development for immune cell therapy	Single cell	Transcriptome or Gene Expression	2022-09-14
KAP220466	Development of prognosis-treatment prediction biomarkers based on single-cell transcripts for colorectal cancer	Single cell	Transcriptome or Gene Expression	2022-09-14
KAP220477	Single cell transcriptome based biomarker development in colorectal cancer	Multiisolate	Whole Genome sequencing	2022-10-17
KAP220479	Single cell transcriptome based biomarker development in colorectal cancer	Multiisolate	Whole Genome sequencing, Transcriptome or Gene Expression	2022-10-17
KAP220355	Single-cell multiomics analysis based biomarker and new target development for immune cell therapy	Single cell	Transcriptome or Gene Expression	2022-07-05
KAP230581	Multifaceted roles of retrotransposon-fusion RNAs	Multiisolate	Whole Genome sequencing, Epigenomics, Transcriptome or Gene Expression	2023-05-09

### Sample Group

BioProject ID	KAP230581
Sample Group Accession ID	KASG230995
Sample Title	Cancer cell line
Description	Lung cancer cell line samples (H1299, HCC827)
Sample Type	Human sample
Registration Date	2023-05-24
Number of BioSamples	2
Number of Experiments	2

### Sample Group Information

[화면 확장](#)

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N
2	Sample name	Organism	Type	Cell line	Tissue	Cell type	Cell subtype	Culture collection	Biomaterial provider	Biological replicate	Treatment	Isolate	Karyotype	Ag
3	M	M	O	O	M	O	O	O	M	O	O	M	O	O
4	NCI-H1299	Homo sapiens	Carcinoma	Cell line	Lung	Epithelial cell			ATCC			NCI-H1299		43
5	HCC827	Homo sapiens	Carcinoma	Cell line	Lung	Epithelial cell			ATCC			HCC827		38
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														

[Download Excel](#)

### KRA

BioProject ID	Sample Group Accession ID	KRA Accession ID	Registration Date	Platform	Strategy	Selection
KAP230581	KASG230995	KAD02301058	2023-05-24	ILLUMINA	RNA-Seq	PolyA

### 3-4. Search KAD and download raw data

#### ☉ KAD 데이터 활용 검색

1. [메인페이지 » Browse » BioProject] 페이지로 이동
2. 선택한 BioPreoject 내 사용하고자 하는 Sample 데이터를 클릭하여 이동
3. 해당 Sample Group ID 에 링크된 KAD 데이터에 대한(공개일자가 지난데이터) 메타데이터(웹상 엑셀 파일로 바로 다운로드) 및 실험데이터의(Gbox 사용) 다운로드 가능
  - ① 본 사용자 가이드 통하여 원하는 데이터를 검색 및 선택하여 페이지로 이동
  - ② [Download Excel] 버튼을 클릭하여 데이터 다운로드 ※ 공개 전 데이터는 다운로드 가능

③ 실험데이터는 하단의 [Request Data Access]를 클릭하여 요청 한 뒤 본인 계정의 Gbox 로 들어가면 요청된 실험데이터 다운로드가 가능

### Browse KAD

...

Submission : [KAD2200848](#) | BioSample Group ID : KASG220783

---

**KAD**

BioProject ID	KAP220344
Sample Group Accession ID	KASG220783
KAD Accession ID	KAD2200848
Registration Date	2022-08-02
Platform	
Strategy	
Selection	
Release Date Selection	specified
Release Date	2029-12-31

---

**Submit KAD**

Full Screen

	A	B	C	D	E	F	G	H	I	J	K	L
1	Experiment Design						Experiment					
2	Sample name	Experiment title	Library name	Instrument model	Platform	Library Construction/Experiment Design	Library	Strategy	Source	Selection	General description	Fragment/Paired read
3	M	M	M	M	M	M	O	M	M	M	O	O
4	4-N	Thyroid cancer	4-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
5	5-N	Thyroid cancer	5-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
6	6-N	Thyroid cancer	6-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
7	8-N	Thyroid cancer	8-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
8	9-N	Thyroid cancer	9-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
9	10-N	Thyroid cancer	10-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
10	11-N	Thyroid cancer	11-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
11	12-N	Thyroid cancer	12-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
12	13-N	Thyroid cancer	13-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
13	14-N	Thyroid cancer	14-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
14	15-N	Thyroid cancer	15-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
15	16-N	Thyroid cancer	16-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
16	17-N	Thyroid cancer	17-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end
17	18-N	Thyroid cancer	18-N_RNA-se	illumina NovaSe	ILLUMIN	illumina sequencing library kit	NA	RNA-Seq	TRAN	PCR	NGS sequencing	Paired-end

**Download Excel**

**Submit KAD**

Full Screen

	A	B	C	D	E	F	G	H	I	J	K	L
1	Experiment Design						Experiment					
2	Sample name	Experiment title	Library name	Instrument model	Platform	Library Construction/Experiment Design	Library	Strategy	Source	Selection	General description	Fragment/Paired read
3	M	M	M	M	M	M	O	M	M	M	O	O
4	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
5	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
6	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
7	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
8	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
9	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
10	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
11	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
12	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
13	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
14	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
15	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
16	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
17	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end
18	2017050800	Whole genom	NA	X Ten	ILLUMIN	not provided		Whole	GENO	NA		Paired-end

**Download Excel**

---

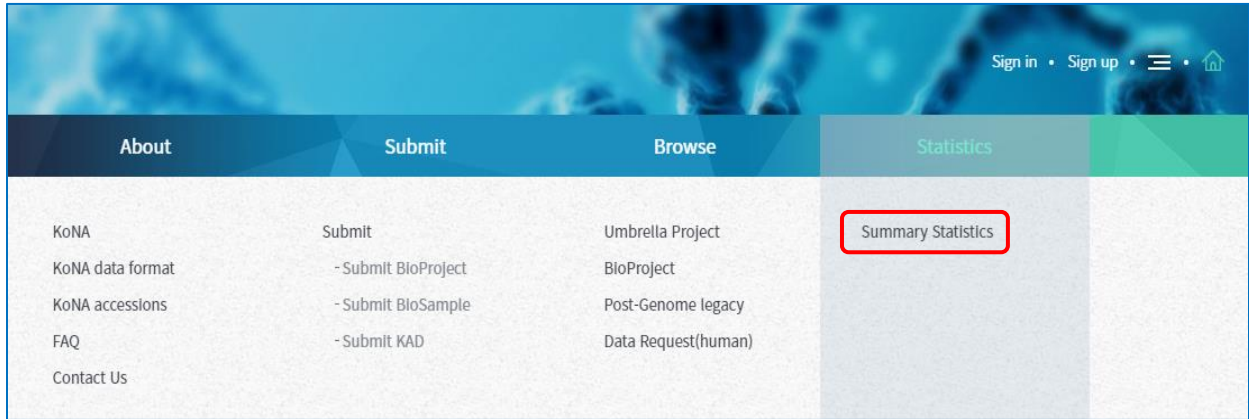
**Is GBOX not installed?**  
 Gbox is a fast file-transfer system used for uploading raw sequencing data. Click the right button to go to the download page.

[Request Data Access](#)

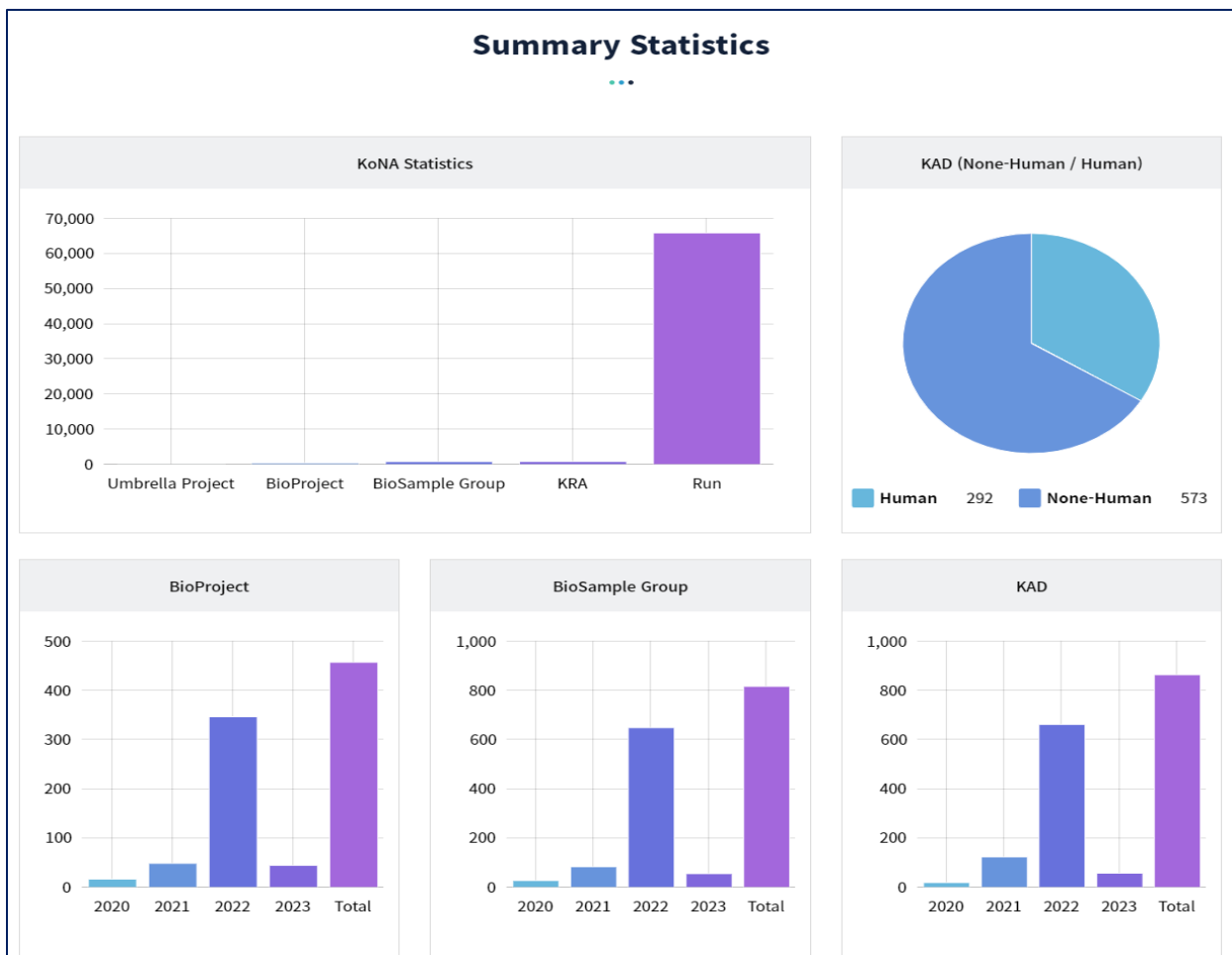
### 3-4. Statistics

#### ☉ 데이터 타입별 통계

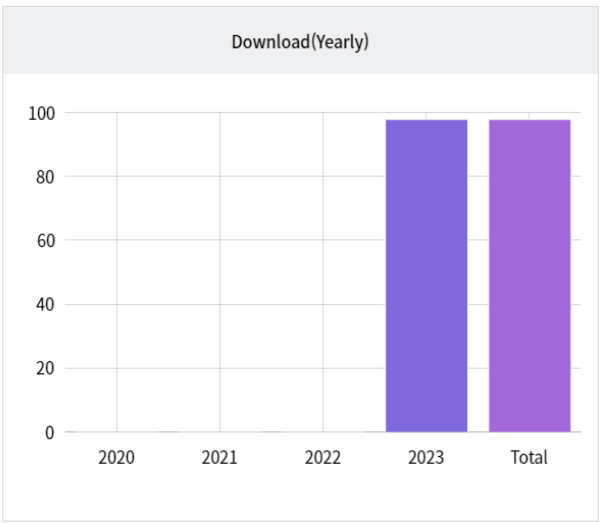
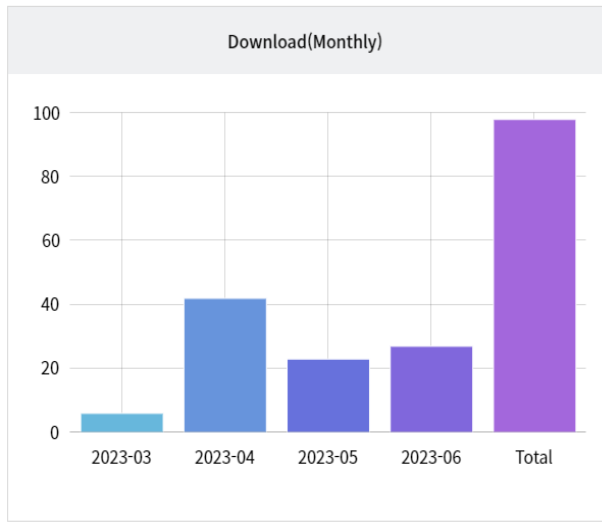
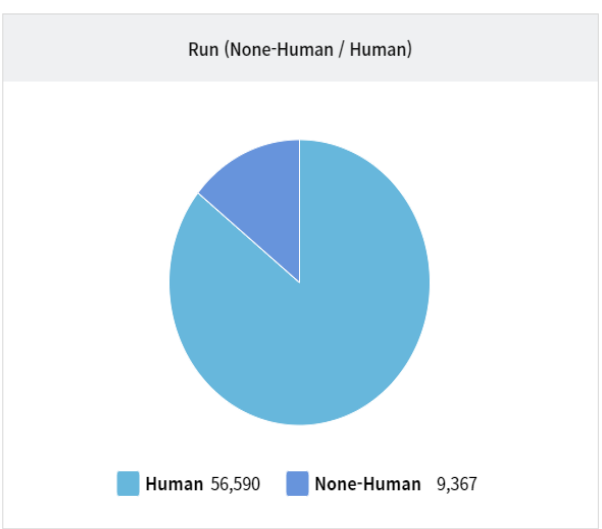
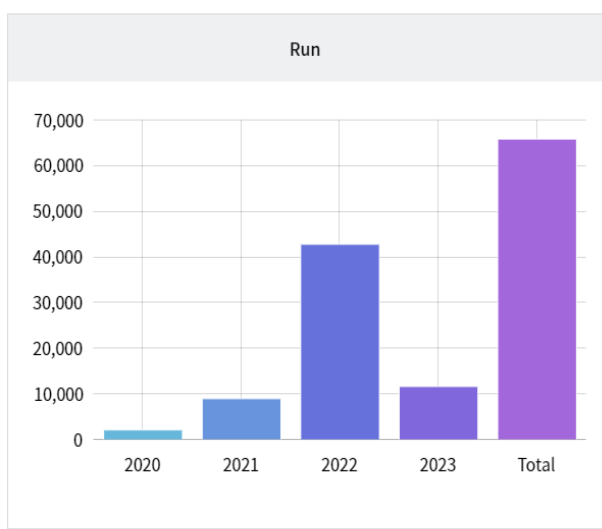
1. [메인페이지 » Statistics » Summary Statistics] 페이지로 이동



2. "데이터 통계" 페이지에서 데이터별 등록 현황, 등록 추이 및 년도별 등록 현황 확인







## IV FAQ

F : How many samples and experiments do I need?

A : A1. You may have more than one experiment per samples.

SAMPLE is a record of biological isolate with unique physical properties.

Please refer to the KoNA Guide for detailed descriptions and guideline.

EXPERIMENT is a unique sequencing result for a specific sample.

A2. In the KoNA Metatdata template, EXPERIMENT is represented by a combination of library + strategy + layout + instrument model and corresponds to one row in the table.

Please refer to the KoNA Metadata Overview for better understanding of KoNA data organization.

F : What KoNA accession do I use for my publication and where can I find it?

A : A1. We recommend using the BioProject accession (KAP#) in publications.

A BioProject is an umbrella for all data submitted to the NCBI for a given study.

Each BioProject can be associated with more than one submission in KoNA. After you have

successfully completed your submission, the KoNA will send an email notification with the BioProject accession.

F : My manuscript reviewer is requesting a link to my submission, how do I get one?

A : Log into the KoNA's MyPage Interface

Find BioProject of interest by browsing, searching, or filtering all your data

Press the button "Reviewer link"

F : Why doesn't my GBox upload work?

A : For troubleshooting GBox service please refer to Bio-Express Support

F : How do I create more than one EXPERIMENT (library/replicate) per SAMPLE for the KAD submission in Bio-Sub menu?

A : Each row in the KAD metadata template represents one EXPERIMENT. Simply use the same BioSample accession or name (depending on the template) in rows (EXPERIMENTs) that you want to associate with this sample.

F : How do I create more than one RUN per EXPERIMENT in Bio-Sub menu?

A : Only one RUN per EXPERIMENT is allowed for KRA submission in the Bio-Sub Portal. When libraries are indeed identical (same combination of library + strategy + layout + instrument model), all files should be placed in the same RUN. To do this simply enter the file names consecutively in the same row.

F : How do I delete my submission (SUB#: e.g., BIOPROJECT#, BIOSAMPLE GROUP#, and KAD#) in Bio-Sub menu?

A : To delete KoNA submissions please contact KoNA staff for assistance at data@kobic.kr. Provide SUB# and reason.

F : How do I add more samples to my KAD submission in Bio-Sub menu?

A : KAD Submission is a discrete act of depositing data (transaction). You cannot add data to a KAD after corresponding KAD accession ID has been issued. Instead, you need to add new samples and new KoNA data and/or new KoNA data to existing samples to a BioProject. BioProject serves as an umbrella for all data submitted to the NCBI for a given research project. In order to update your BioProject with new samples and/or KoNA data you should create a new submission where you provide the BioProject accession (KAP#) of the project you want to update. This will ensure that all pertinent data are linked to the same accession in Entrez and become searchable across databases.

F : My manuscript has been published. How do I update my KoNA submission with the publication?

A : Please refer to the How to cite in KoNA

F : How do I change the release date of my KoNA submission?

A : Please contact KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr)

F : How do I improve/correct my metadata after submitting?

A : Only KoNA staff can correct the metadata after the accession ID is issued. Please contact KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr).

F : How do I withdraw my KoNA BIOPRJECT, BIOSAMPLE GROUP, KAD in KoNA (KAP#/KASG#/KAD#)?

A : Contact the KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr). Provide KAP#, KASG#, KAD#, etc.

F : How do I re-name my EXPERIMENT/RUN?

A : Aliases of EXPERIMENTs and RUNs cannot be changed.

F : How do I change the PLATFORM in my EXPERIMENTs?

A : Contact KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr).

F : How do I link an EXPERIMENT to a different SAMPLE?

A : Only KoNA staff can correct the metadata after the accession ID is issued. Please contact KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr).

F : How do I link a RUN to a different EXPERIMENT?

A : Only KoNA staff can correct the metadata after the accession ID is issued. Please contact KoNA staff for assistance at [data@kobic.kr](mailto:data@kobic.kr).

F : Contact KoNA

A : If you were unable to resolve your problem after reading this F&Q, contact KoNA staff at [data@kobic.kr](mailto:data@kobic.kr).

# VI APPENDIX

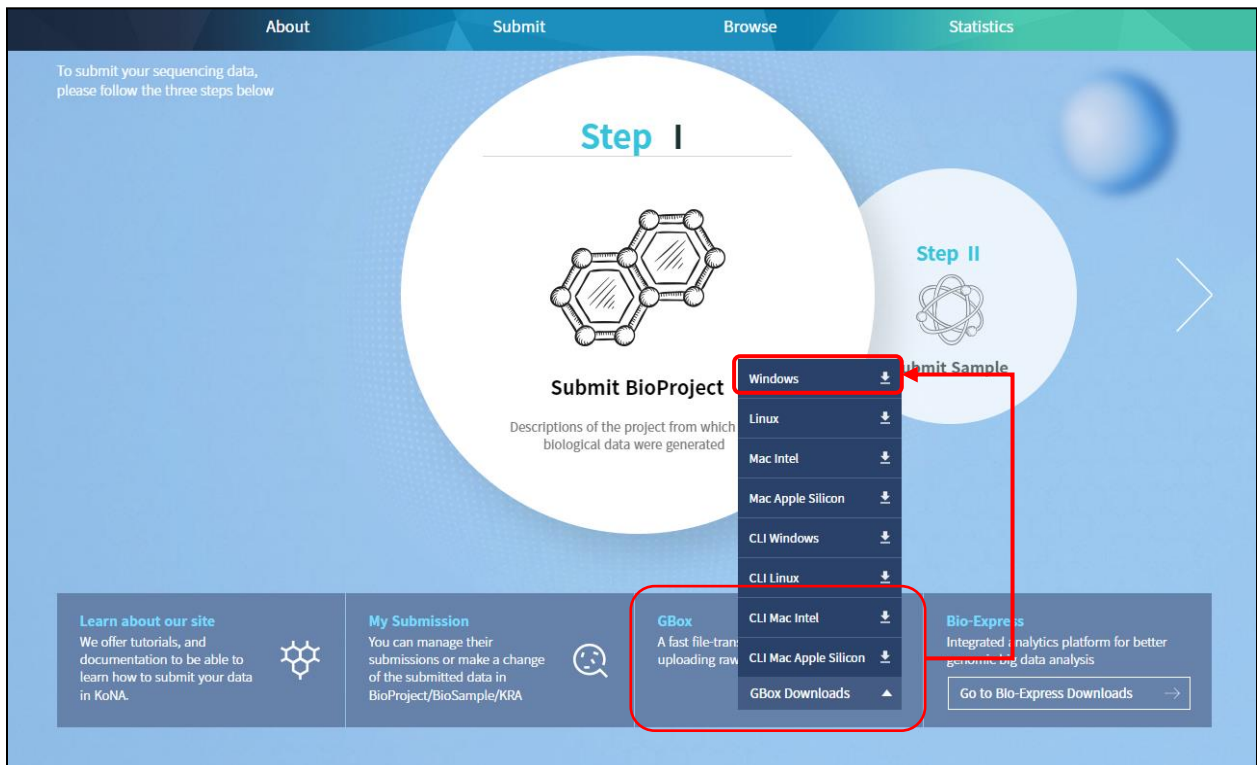
## 4-1. 빅데이터 고속 전송 시스템 (GBox)

### ◎ 빅데이터 고속 전송 시스템 소개

GBox 고속 전송 시스템은 가속기 등의 추가적인 하드웨어 구축이나 ActiveX와 같은 별도의 플러그인 설치 없이 소프트웨어 기술만으로 사용자의 대용량 데이터를 고속으로 전송할 수 있는 시스템

### • 프로그램 다운로드 및 설치

[메인페이지 ⇨ 하단 G-Box] 클릭, 본인의 컴퓨터 환경에 맞는 것을 선택하여 다운로드



### ※ KRA upload raw data

- ▶ 홈페이지상의 Gbox를 각 연구자의 컴퓨터 환경 설정에 맞게 다운로드 하여 설치
- ▶ 설치된 Gbox를 실행 후 로그인 하여 업로드 하고자 하는 raw data를 이동 복사

▶ 데이터가 이동된 디렉토리의 path를 메타데이터 작성시 Run column에 입력

## • 데이터 업로드 및 다운로드

1. 아래 그림의 GBox 프로그램을 통해 사용자는 KoNA 스토리지에 데이터를 업로드하거나 사용자의 PC로 다운로드할 수 있음

Single sign-on for  
**GBOX**

Starts the GBox Workbench.

|

Password

Forgotten ID/Password?

Login

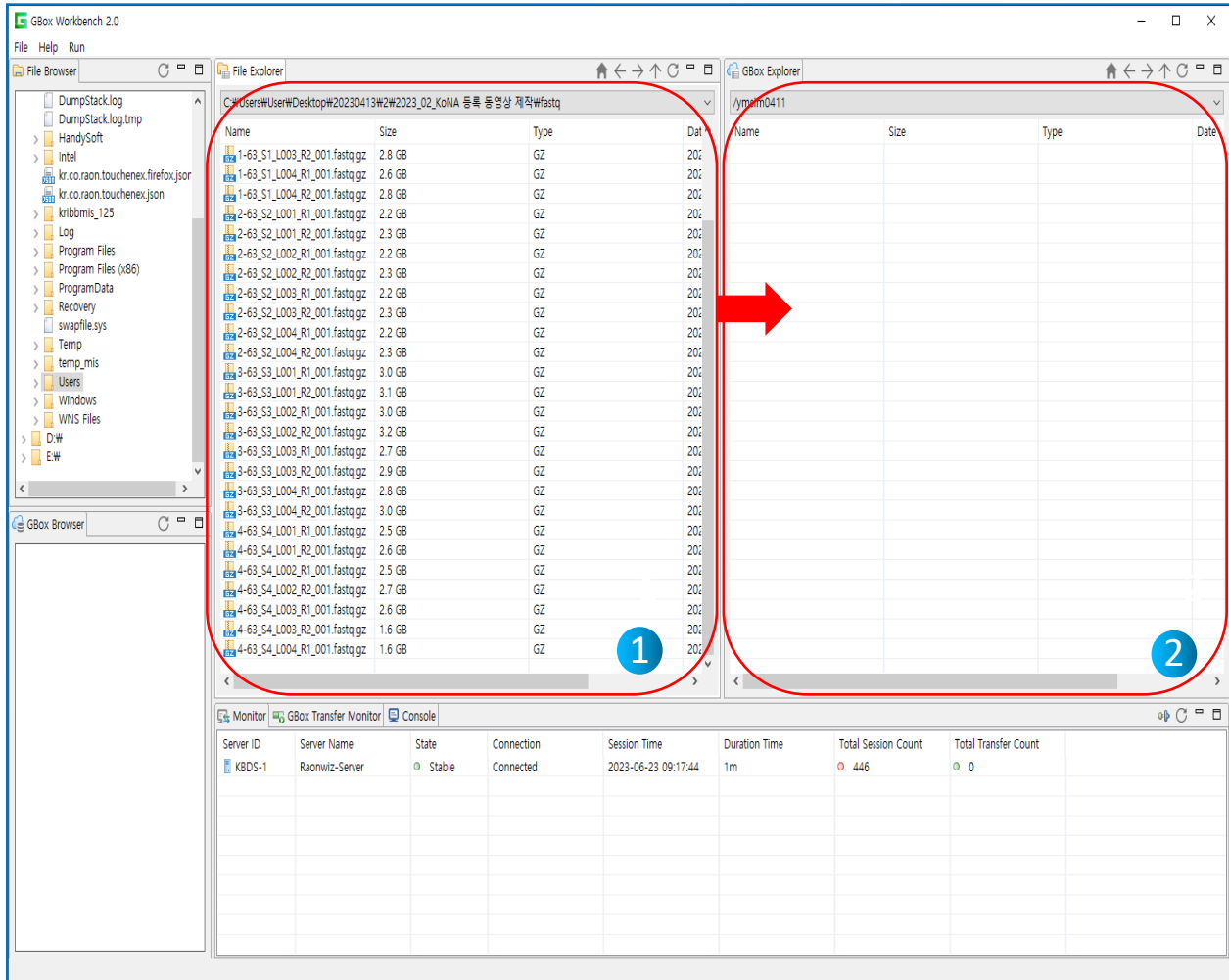
Cancel

Not register yet? [Create Account](#)

Go Bio-Express Service [↗](#)

① [File Browser]나 [File Explorer]를 통해 사용자 PC의 업로드 할 데이터가 있는 폴더로 이동

② 업로드 할 폴더 또는 파일을 선택한 후 [GBox Explorer]로 드래그하거나 [마우스 우클릭] 후 팝업되는 메뉴 창에서 Upload 하면 [KoNA스토리지] 경로로 선택한 폴더 및 파일이 업로드됨



③ [File Explorer] 창에서 [마우스 우클릭]을 통해 데이터 업로드, 폴더 및 파일 생성, rename, 복사/붙여넣기, 삭제, refresh 기능을 지원

④ KoNA 스토리지 내 사용자의 경로는 "/[사용자아이디]"로 시작. 업로드한 폴더 및 파일의 경로는 /[사용자아이디]/업로드한 폴더/업로드한 파일"로 사용

예1) 사용자 아이디 : test, 업로드한 폴더 : RNA\_seq, 해당폴더 내 파일명 : sample.fq.gz

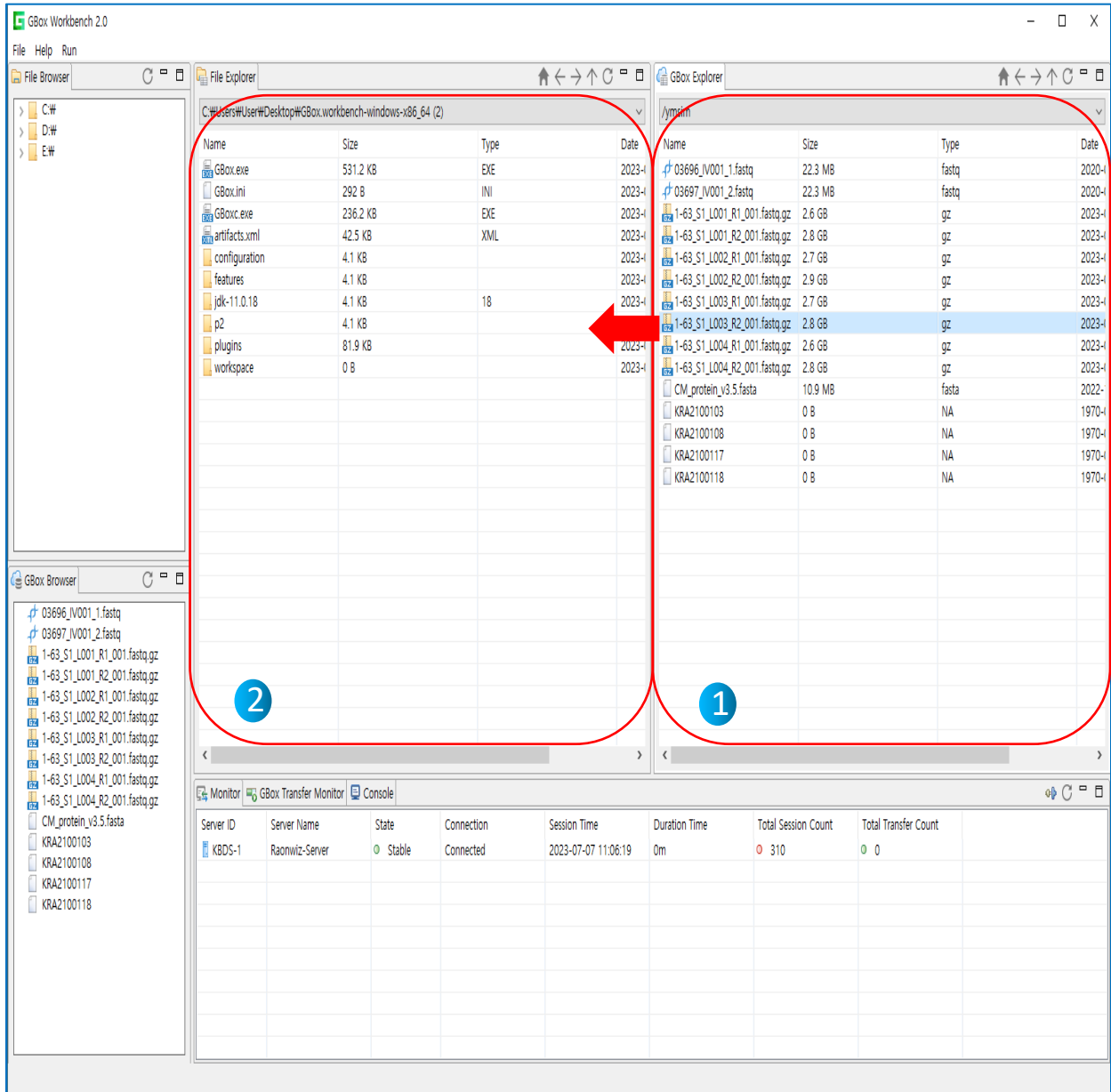
⇒ 업로드한 폴더 및 파일의 경로 : /test/RNA\_seq/sample.fq.gz

예2) 사용자 아이디 : test, 업로드한 파일명 : sample.fq.gz

⇒ 업로드한 폴더 및 파일의 경로 : /test/sample.fq.gz

⑤ [현재 KoNA 스토리지] 서버와 연결 상태를 확인 가능. Connection 창의 상태가 “Connected”가 일 때만 데이터 업로드/다운로드가 가능

⑥ 데이터를 사용자 PC로 다운로드할 때는 [GBox Explorer]창의 폴더 또는 파일을 [File Explorer]로 드래그하거나 GBox Explorer의 파일 및 폴더를 우클릭해 다운로드



## • 업로드한 데이터 경로 확인

① [데이터 업로드]가 시작되면 하단의 progress 창에서 업로드 현황 확인이 가능. 폴더



를 업로드하는 경우 폴더 내의 모든 파일에 대한 파일 크기 측정이 완료된 후 업로드가 시작되며 파일 개수에 따라 측정 시간이 추가로 소요

- ② [업로드]가 완료되면 GBox Explorer창에 업로드된 파일 또는 폴더가 나타남
- ③ [이 파일 및 폴더의 위치]는 표시된 [현재 KoNA스토리지 내 경로] 하위에 파일명 및 폴더명으로 저장
- ④ 업로드한 파일 및 폴더의 경로를 확인하는 다른 방법은 GBox Explorer 창에서 파일 또는 폴더를 선택하고 [마우스 우클릭]을 하여 팝업되는 메뉴에서 Path Copy를 선택하면 해당 파일 또는 폴더의 전체 경로가 복사
- ⑤ 이후 KoNA 웹 페이지에서 직접 입력 등록 또는 파일 업로드 등록 시 입력할 파일명은 파일명이 아닌 위의 방식으로 파악한 파일의 "전체 경로명(절대경로)"을 기입

## 4-2. 상세 설명 가이드

### 1. Sample Design - Sample Type

#### 1.1 Clinical or host-associated pathogen

Description : Clinical or host-associated pathogen

#### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Collected by	Name(s) of person(s) or institute who collected the sample (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")
Collection date	Date of sampling (YYYY-MM-DD) (e.g. "1989-09-20")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")

Geographic location	Geographical origin of the sample; Use a colon to separate the country or ocean from more detailed information about the location (e.g "Korea: Seoul"; "Korea")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.ddd] N S d[dd.ddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")
Host	Name of the natural (as opposed to laboratory) host species to the organism from which the sample was obtained (e.g. "Homo sapiens"; "Gallus gallus domesticus")
Host disease	Name of relevant disease, e.g. Salmonella gastroenteritis (e.g. "Salmonella gastroenteritis"; "Hepatitis C")

### Optional Attributes

Strain	microbial or eukaryotic strain name, number or designation (e.g. "MG1234"; "K12"; "BALB/c")
Subgroup	Taxonomy below subspecies; sometimes used in viruses to denote subgroups taken from a single isolate (e.g. "Clostridium botulinum Group I")
Subtype	Used as classifier in viruses (e.g. HIV type 1, Group M, Subtype A) (e.g. HIV type 1, Group M, Subtype A)
Culture collection	<ul style="list-style-type: none"> <li>* Name of source institute and unique culture identifier.</li> <li>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes (<a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a>) (e.g. "ATCC:26370")</li> </ul>
Specimen voucher	<ul style="list-style-type: none"> <li>* Identifier for the physical specimen that remains after the sample has been obtained.</li> <li>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated museum/collection, or the sequenced sample was deposited in the indicated museum/collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</li> </ul>
Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")
Identified by	The name of the taxonomist who identified the specimen. This field reports the name(s) of the specific person(s) who identified the TAXONOMY of the sample. This does not mean the person(s) in the laboratory who identified the submitted sample. (e.g. "Dan Janzen")

Passage history	Number of passages and passage method (e.g. "13"; "Gentle cell dissociation reagent (STEMCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seeded at a 1:6 split ratio.")
Isolate	Identification or description of the specific individual from which this sample was obtained (e.g. "Patient #152"; "DGGE band PSBAC-13"; "MattSeq37C_S93")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Host subject id	a unique identifier by which each subject can be referred to, de-identified, e.g. #131 (e.g. "#131")
Host tissue sampled	Type of tissue the initial sample was taken from (e.g. "leaves"; "liver")
Host age	Age of host at the time of sampling (e.g. "12 years old"; "2.5 month")
Host sex	Gender or physical sex of the host (e.g. "male"; "female"; "mixed"; "hermaphrodite"; "not determined"; "missing"; "not applicable"; "not collected")
Host disease stage	Stage of disease at the time of sampling (e.g. "Stage 2"; "Illa"; "de novo AML")
Host health state	Information regarding health state of the individual sampled at the time of sampling (e.g. "death"; "chronic disease"; "recovery")
Host disease outcome	Final outcome of disease, e.g., death, chronic disease, recovery (e.g. "death"; "chronic disease"; "recovery")
Host description	Additional information not included in other defined vocabulary fields (e.g. "Patient received prior antiviral therapy but did not achieve sustained virological response.")
Pathotype	Some bacterial specific pathotypes (e.g. "Escherichia coli -STEC, UPEC"; "Extended-spectrum beta-lactamase (ESBL)")
Serotype	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. e.g. serotype="H1N1" in Influenza A virus CY098518 (e.g. "H1N1"; "B1")
Serovar	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy (e.g. "Leptospira, serovar saopaulo S76607 (65357 in Entrez)"; "Salmonella enterica subsp. enterica serovar Braenderup"; "O157")

## 1.2 Environmental, food or other pathogen

Description : Environmental, food or other pathogen

### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Collected by	Name(s) of person(s) or institute who collected the sample (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")

Collection date	Date of sampling (YYYY-MM-DD) (e.g. "1989-09-20")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")
Geographic location	Geographical origin of the sample; Use a colon to separate the country or ocean from more detailed information about the location (e.g "Korea: Seoul"; "Korea")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N S d[dd.dddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")

### Optional Attributes

Strain	microbial or eukaryotic strain name, number or designation (e.g. "MG1234"; "K12"; "BALB/c")
Subgroup	Taxonomy below subspecies; sometimes used in viruses to denote subgroups taken from a single isolate (e.g. "Clostridium botulinum Group I")
Subtype	Used as classifier in viruses (e.g. HIV type 1, Group M, Subtype A) (e.g. HIV type 1, Group M, Subtype A)
Culture collection	<p>* Name of source institute and unique culture identifier.</p> <p>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes (<a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a>) (e.g. "ATCC:26370")</p>
Specimen voucher	<p>* Identifier for the physical specimen that remains after the sample has been obtained.</p> <p>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated museum/collection, or the sequenced sample was deposited in the indicated museum/collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</p>
Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier (e.g. "Sample name_1")
Identified by	The name of the taxonomist who identified the specimen. This field reports the name(s) of the specific person(s) who identified the TAXONOMY of the

	sample. This does not mean the person(s) in the laboratory who identified the submitted sample (e.g. "Dan Janzen").
Passage history	Number of passages and passage method (e.g. "13"; "Gentle cell dissociation reagent (STEMCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seeded at a 1:6 split ratio.")
Isolate	Identification or description of the specific individual from which this sample was obtained (e.g. "Patient #152"; "DGGE band PSBAC-13"; "MattSeq37C_S93")
Temperature	temperature of the sample at time of sampling (e.g. "27 °C")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Pathotype	Some bacterial specific pathotypes (example Escherichia coli -STEC, UPEC) ( e.g. "Escherichia coli -STEC, UPEC"; "Extended-spectrum beta-lactamase (ESBL)")
Serotype	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. e.g. serotype="H1N1" in Influenza A virus CY098518 (e.g. "H1N1"; "B1")
Serovar	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy (e.g. "Leptospira, serovar saopaulo S76607 (65357 in Entrez)"; "Salmonella enterica subsp. enterica serovar Braenderup"; "O157

### 1.3 Microbe

Description : Use for bacteria or other unicellular microbes when it is not appropriate or advantageous to use MlxS, Pathogen or Virus packages.

#### **Mandatory Attributes**

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Strain	microbial or eukaryotic strain name, number or designation (e.g. "MG1234"; "K12"; "BALB/c")
Collection date	Date of sampling (YYYY-MM-DD) (e.g "1989-09-20")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")
Geographic location	Geographical origin of the sample; Use a colon to separate the country or ocean from more detailed information about the location (e.g "Korea: Seoul"; "Korea")

## Optional Attributes

Culture collection	<ul style="list-style-type: none"> <li>* Name of source institute and unique culture identifier.</li> <li>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes (<a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a>) (e.g. "ATCC:26370")</li> </ul>
Specimen voucher	<ul style="list-style-type: none"> <li>* Identifier for the physical specimen that remains after the sample has been obtained.</li> <li>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated museum/collection, or the sequenced sample was deposited in the indicated museum/collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</li> </ul>
Biomaterial provider	<ul style="list-style-type: none"> <li>* The name and address of the lab or PI, or a culture collection identifier who provided the sample to the submitter.</li> <li>* This field is used to annotate source material in biological collections that do not fit into either the 'Culture collection' or the 'Specimen voucher' field categories: <ul style="list-style-type: none"> <li>* Physical specimens from zoos, aquaria, stock centers, seed banks, germplasm repositories, or DNA banks.</li> </ul> </li> <li>* Provide the following information only if the sequenced sample was retrieved directly from the indicated collection, or the sequenced sample was deposited in the indicated collection.</li> <li>* If the value of the field modifier is the name and address of the lab or PI, "Biomaterial provider" and "Collected by" fields can have the same value.</li> <li>* If the value of the field modifier is a culture collection, the identifier should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "Korea Research Institute of Bioscience &amp; Biotechnology(KRIBB)"; "Dan Janzen"; "CGC:CB3912")</li> </ul>
Biological replicate	<p>Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")</p>
Collected by	Name(s) of person(s) or institute who collected the sample (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")
Identified by	The name of the taxonomist who identified the specimen. This field reports the name(s) of the specific person(s) who identified the TAXONOMY of the

	sample. This does not mean the person(s) in the laboratory who identified the submitted sample. (e.g. "Dan Janzen")
Passage history	Number of passages and passage method (e.g. "13"; "Gentle cell dissociation reagent (STEMCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seeded at a 1:6 split ratio.")
Sample size	Amount or size of sample (volume, mass or area) that was collected (e.g. "1 L"; "0.3 kg"; "0.1 m <sup>2</sup> ")
Environment biome	descriptor of the broad ecological context of a sample. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO terms can be found via the link ( <a href="https://www.ebi.ac.uk/ols/ontologies/envo">https://www.ebi.ac.uk/ols/ontologies/envo</a> ) (e.g. "desert"; "taiga"; "deciduous woodland"; "coral reef")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.dddd] N S d[dd.dddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")
Altitude	The altitude of the sample is the vertical distance between Earth's surface above Sea Level and the sampled position in the air (e.g. "-256 m"; "330.12 m")
Depth	Depth is defined as the vertical distance below surface, e.g. for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples (e.g. "15m depth")
Temperature	temperature of the sample at time of sampling (e.g. "27 °C")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Host	Name of the natural (as opposed to laboratory) host species to the organism from which the sample was obtained (e.g. "Homo sapiens"; "Gallus gallus domesticus")
Host tissue sampled	Type of tissue the initial sample was taken from (e.g. "leaves"; "liver")
Lab host	Scientific name and description of the laboratory host used to propagate the source organism or material from which the sample was obtained, e.g., Escherichia coli DH5a, or Homo sapiens HeLa cells (e.g. "Escherichia coli DH5a"; "Homo sapiens HeLa cell")
Serotype	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. e.g. serotype="H1N1" in Influenza A virus CY098518 (e.g. "H1N1"; "B1")
Serovar	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. Same as serovar and serotype. Sometimes used as species identifier in bacteria with shaky taxonomy (e.g. "Leptospira, serovar saopaulo S76607 (65357 in Entrez)"; "Salmonella enterica subsp. enterica serovar Braenderup"; "O157")

#### 1.4 Model organism or animal sample

Description : 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.

##### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Tissue	Type of tissue the sample was taken from (e.g "leaves"; "liver")
Biomaterial provider	<p>* The name and address of the lab or PI, or a culture collection identifier who provided the sample to the submitter.</p> <p>* This field is used to annotate source material in biological collections that do not fit into either the 'Culture collection' or the 'Specimen voucher' field categories:</p> <p>* Physical specimens from zoos, aquaria, stock centers, seed banks, germplasm repositories, or DNA banks.</p> <p>* Provide the following information only if the sequenced sample was retrieved directly from the indicated collection, or the sequenced sample was deposited in the indicated collection.</p> <p>* If the value of the field modifier is the name and address of the lab or PI, "Biomaterial provider" and "Collected by" fields can have the same value.</p> <p>* If the value of the field modifier is a culture collection, the identifier should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "Korea Research Institute of Bioscience &amp; Biotechnology(KRIBB)"; "Dan Janzen"; "CGC:CB3912")</p>
Sex	physical sex of sampled organism (e.g. "male"; "female"; "mixed"; "hermaphrodite"; "not determined"; "missing"; "not applicable"; "not collected")
<b>Optional Attributes</b>	
Strain	microbial or eukaryotic strain name, number or designation (e.g. "MG1234"; "K12"; "BALB/c")
Breed	breed name -chiefly used in domesticated animals or plants (e.g "mixed"; "Beagle")
Cell line	Name of the cell line (e.g "HepG2 cell")
Cell type	Type of cell of the sample or from which the sample was obtained (e.g "T cell")
Cell subtype	The subtype of cell (e.g "CD4+ T cell")
Culture collection	<p>* Name of source institute and unique culture identifier.</p> <p>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes</p>



(ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll\_dump.txt) (e.g. "ATCC:26370")

Specimen voucher	<p>* Identifier for the physical specimen that remains after the sample has been obtained.</p> <p>* Provide the following information only if the sequence you are submitting was obtained from a sample you retrieved directly from the indicated museum/collection, or the sequence was obtained from a sample that you deposited in the indicated museum/collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</p>
Study book number	Study book number of sample
Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")
Collected by	Name(s) of person(s) or institute who collected the sample (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")
Growth protocol	The protocol of growth (e.g. "ES cell-derived NS cells were routinely generated by re-plating d 7 adherent neural differentiation cultures (typically 2–3 × 10 <sup>6</sup> cells into a T75 flask) on uncoated plastic in NS-A medium (Euroclone, Milan, Italy) supplemented with modified N2 and 10 ng/ml of both EGF and FGF-2 (NS expansion medium).")
Storage conditions	Explain how and for how long the sample was stored before DNA extraction
Treatment	Treatment of sample (e.g. "Adalimumab treatment")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")
Geographic location	Geographical origin of the sample; Use a colon to separate the country or ocean from more detailed information about the location (e.g "Korea: Seoul"; "Korea")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[d.ddd] N S d[dd.ddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Age	age at the time of sampling; relevant scale depends on species and study (e.g. "12 years old"; "2.5 month")
Birth date	The date of birth (e.g. "1989-09-20")
Birth location	The location of birth (e.g. "Seoul, Korea")
Breeding history	The history of breeding

Breeding method	The method of breeding (e.g. "Clinical veterinary/Client-owned dog")
Death date	The date of death (e.g. "1989-09-20")
Development stage	Developmental stage at the time of sampling (e.g. "adult"; "Dairy cows second seasonal lactation")
Disease	List of diseases diagnosed; can include multiple diagnoses (e.g. "Rheumatoid arthritis"; "Lung cancer")
Disease stage	Stage of disease at the time of sampling (e.g. "Stage 2"; "IIIa"; "de novo AML")
Health state	Health or disease status of sample at time of collection (e.g. "death"; "chronic disease"; "recovery")
Phenotype	Phenotype of sampled organism (e.g. "Leaf senescence"; "decreased tolerance to AI stress")

### 1.5 Human sample

Description : **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.

#### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Tissue	Type of tissue the sample was taken from (e.g "leaves"; "liver")
Biomaterial provider	<ul style="list-style-type: none"> <li>* The name and address of the lab or PI, or a culture collection identifier who provided the sample to the submitter.</li> <li>* This field is used to annotate source material in biological collections that do not fit into either the 'Culture collection' or the 'Specimen voucher' field categories:</li> <li>* Physical specimens from zoos, aquaria, stock centers, seed banks, germplasm repositories, or DNA banks.</li> <li>* Provide the following information only if the sequenced sample was retrieved directly from the indicated collection, or the sequenced sample was deposited in the indicated collection.</li> <li>* If the value of the field modifier is the name and address of the lab or PI, "Biomaterial provider" and "Collected by" fields can have the same value.</li> <li>* If the value of the field modifier is a culture collection, the identifier should be provided using the following format 'institution-code:(optional</li> </ul>

collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen"; "CGC:CB3912")

Isolate	Identification or description of the specific individual from which this sample was obtained (e.g. "Patient #152"; "DGGE band PSBAC-13"; "MattSeq37C_S93")
Sex	physical sex of sampled organism (e.g. "male"; "female"; "mixed"; "hermaphrodite"; "not determined"; "missing"; "not applicable"; "not collected")

### Optional Attributes

Type	Sample type, such as cell culture, mixed culture, tissue sample, whole organism, single cell, and metagenomics assembly
Cell line	Name of the cell line (e.g "HepG2 cell")
Cell type	Type of cell of the sample or from which the sample was obtained (e.g "T cell")
Cell subtype	The subtype of cell (e.g "CD4+ T cell")
Culture collection	* Name of source institute and unique culture identifier. * Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection. * This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes ( <a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a> ) (e.g. "ATCC:26370")
Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")
Treatment	Treatment of sample (e.g. "Adalimumab treatment")
Karyotype	Karyotype of sampled organism. karyotype is the number and appearance of chromosomes in the nucleus of a eukaryotic cell. ( e.g. "copy neutral loss of heterozygosity"; "47XY+21_1"; "haploid")
Age	age at the time of sampling; relevant scale depends on species and study (e.g. "12 years old"; "2.5 month")
Development stage	Developmental stage at the time of sampling (e.g. "adult"; "Dairy cows second seasonal lactation")
Disease	List of diseases diagnosed; can include multiple diagnoses (e.g. "Rheumatoid arthritis"; "Lung cancer")
Disease stage	Stage of disease at the time of sampling (e.g. "Stage 2"; "IIIa"; "de novo AML")
Health state	Health or disease status of sample at time of collection (e.g. "death"; "chronic disease"; "recovery")
Phenotype	Phenotype of sampled organism (e.g. "Leaf senescence"; "decreased tolerance to AI stress")

Population	Population is a summation of all the organisms of the same group or species, which live in a particular geographical area. For human: ; for plants: filial generation, number of progeny, genetic structure (e.g. "F2 population")
Race	Race of sample. Race refers to a person's physical characteristics, such as bone structure and skin, hair, or eye color. (e.g. "Hair color: black"; "Eye color: black")
Ethnicity	Ethnicity of the subject. Ethnicity refers to cultural factors, including nationality, regional culture, ancestry, and language. (e.g. "Korean"; "Chinese Han"; "Asian"; "Caucasian")

## 1.6 Plant sample

Description : Use for any plant sample or cell line.

### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique. (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Cultivar	Cultivar name: cultivated variety of plant from which the sample was obtained
Tissue	Type of tissue the sample was taken from (e.g "leaves"; "liver")
Biomaterial provider	<ul style="list-style-type: none"> <li>* The name and address of the lab or PI, or a culture collection identifier who provided the sample to the submitter.</li> <li>* This field is used to annotate source material in biological collections that do not fit into either the 'Culture collection' or the 'Specimen voucher' field categories:</li> <li>* Physical specimens from zoos, aquaria, stock centers, seed banks, germplasm repositories, or DNA banks.</li> <li>* Provide the following information only if the sequenced sample was retrieved directly from the indicated collection, or the sequenced sample was deposited in the indicated collection.</li> <li>* If the value of the field modifier is the name and address of the lab or PI, "Biomaterial provider" and "Collected by" fields can have the same value.</li> <li>* If the value of the field modifier is a culture collection, the identifier should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "Korea Research Institute of Bioscience &amp; Biotechnology(KRIBB)"; "Dan Janzen"; "CGC:CB3912")</li> </ul>

Sex	physical sex of sampled organism (e.g. "male"; "female"; "mixed"; "hermaphrodite"; "not determined"; "missing"; "not applicable"; "not collected")
-----	--

### Optional Attributes

Type	Sample type, such as cell culture, mixed culture, tissue sample, whole organism, single cell, and metagenomics assembly
Cell line	Name of the cell line (e.g. "HepG2 cell")
Cell type	Type of cell of the sample or from which the sample was obtained (e.g. "T cell")
Culture collection	<p>* Name of source institute and unique culture identifier.</p> <p>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes (<a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a>) (e.g. "ATCC:26370")</p>
Specimen voucher	<p>* Identifier for the physical specimen that remains after the sample has been obtained.</p> <p>* Provide the following information only if the sequence you are submitting was obtained from a sample you retrieved directly from the indicated museum/collection, or the sequence was obtained from a sample that you deposited in the indicated museum/collection.</p> <p>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</p>
Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")
Collected by	Name(s) of person(s) or institute who collected the sample ("Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")
Collection date	Date of sampling (YYYY-MM-DD) (e.g. "1989-09-20")
Growth protocol	The protocol of growth (e.g. "ES cell-derived NS cells were routinely generated by re-plating d 7 adherent neural differentiation cultures (typically 2–3 × 10 <sup>6</sup> cells into a T75 flask) on uncoated plastic in NS-A medium (Euroclone, Milan, Italy) supplemented with modified N2 and 10 ng/ml of both EGF and FGF-2 (NS expansion medium).")
Treatment	Treatment of sample (e.g. "Adalimumab treatment")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent

	Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[.dddd] N S d[.dddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")
Temperature	temperature of the sample at time of sampling (e.g. "27 °C")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Age	age at the time of sampling; relevant scale depends on species and study (e.g. "12 years old"; "2.5 month")
Development stage	Developmental stage at the time of sampling (e.g. "adult"; "Dairy cows second seasonal lactation")
Disease	List of diseases diagnosed; can include multiple diagnoses (e.g. "Rheumatoid arthritis"; "Lung cancer")
Disease stage	Stage of disease at the time of sampling (e.g. "Stage 2"; "IIIa"; "de novo AML")
Phenotype	Phenotype of sampled organism (e.g. "Leaf senescence"; "decreased tolerance to AI stress")
Height or length	Measurement of height or length (e.g. "Height: 135 cm" )
Population	Population is a summation of all the organisms of the same group or species, which live in a particular geographical area. For human: ; for plants: filial generation, number of progeny, genetic structure (e.g. "F2 population")

### 1.7 Virus sample

Description : Use for all virus samples not directly associated with disease. Viral pathogens should be submitted using the Pathogen: Clinical or host-associated pathogen package.

#### **Mandatory Attributes**

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Organism	The most descriptive organism name for this sample (to the species, if relevant) (e.g. "Candida metapsilosis"; "Hepacivirus C"; "Homo sapiens"; "metagenome")
Collection date	Date of sampling (YYYY-MM-DD) (e.g. "1989-09-20")
Isolation source	Describes the physical, environmental and/or local geographical source of the biological sample from which the sample was derived (e.g. "rumen isolates from standard Pelleted ration-fed steer #67"; "permanent Antarctic sea ice"; "denitrifying activated sludge from carbon_limited continuous reactor")
Isolate	Identification or description of the specific individual from which this sample was obtained (e.g. "Patient #152"; "DGGE band PSBAC-13"; "MattSeq37C_S93")

Geographic location	Geographical origin of the sample; Use a colon to separate the country or ocean from more detailed information about the location (e.g "Korea: Seoul"; "Korea")
Host	Name of the natural (as opposed to laboratory) host species to the organism from which the sample was obtained (e.g. "Homo sapiens"; "Gallus gallus domesticus")

### Optional Attributes

Strain	microbial or eukaryotic strain name, number or designation 칭 (e.g. "MG1234"; "K12"; "BALB/c")
Subgroup	Taxonomy below subspecies; sometimes used in viruses to denote subgroups taken from a single isolate (e.g. "Clostridium botulinum Group I")
Subtype	Used as classifier in viruses (e.g. HIV type 1, Group M, Subtype A)
Culture collection	<ul style="list-style-type: none"> <li>* Name of source institute and unique culture identifier.</li> <li>* Provide the following information only if the sequenced sample you are submitting was retrieved directly from the indicated culture collection, or the sequenced sample was deposited in the indicated culture collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):culture-id'. culture-id and institution-code are mandatory. See the list of allowed institutes (<a href="ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt">ftp://ftp.ncbi.nlm.nih.gov/pub/taxonomy/coll_dump.txt</a>) (e.g. "ATCC:26370")</li> </ul>
Specimen voucher	<ul style="list-style-type: none"> <li>* Identifier for the physical specimen that remains after the sample has been obtained.</li> <li>* Provide the following information only if the sequence you are submitting was obtained from a sample you retrieved directly from the indicated museum/collection, or the sequence was obtained from a sample that you deposited in the indicated museum/collection.</li> <li>* This should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "UAM:Mamm:52179"; "AMCC:101706"; "USNM:field series 8798"; "personal:Dan Janzen:99-SRNP-2003"; "99-SRNP-2003")</li> </ul>
Biomaterial provider	<ul style="list-style-type: none"> <li>* The name and address of the lab or PI, or a culture collection identifier who provided the sample to the submitter.</li> <li>* This field is used to annotate source material in biological collections that do not fit into either the 'Culture collection' or the 'Specimen voucher' field categories:</li> <li>* Physical specimens from zoos, aquaria, stock centers, seed banks, germplasm repositories, or DNA banks.</li> <li>* Provide the following information only if the sequenced sample was retrieved directly from the indicated collection, or the sequenced sample was deposited in the indicated collection.</li> <li>* If the value of the field modifier is the name and address of the lab or PI, "Biomaterial provider" and "Collected by" fields can have the same value.</li> </ul>

\* If the value of the field modifier is a culture collection, the identifier should be provided using the following format 'institution-code:(optional collection-code):specimen-id'. specimen-id is mandatory, collection-code is optional; institution-code is mandatory when collection code is provided. (e.g. "Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen"; "CGC:CB3912")

Biological replicate	Please clarify BioSamples in a same group (e.g. same species, type of cells, genotype or treatment condition) by using Biological replicate identifier. (e.g. "Sample name_1")
Collected by	Name(s) of person(s) or institute who collected the sample ("Korea Research Institute of Bioscience & Biotechnology(KRIBB)"; "Dan Janzen")
Identified by	The name of the taxonomist who identified the specimen. This field reports the name(s) of the specific person(s) who identified the TAXONOMY of the sample. This does not mean the person(s) in the laboratory who identified the submitted sample. (e.g. "Dan Janzen")
Passage history	Number of passages and passage method (e.g. "13"; "Gentle cell dissociation reagent (STEMCELL Technologies, 07174) was used to passage CPP cells as aggregates which were then seeded at a 1:6 split ratio.")
Sample size	Amount or size of sample (volume, mass or area) that was collected <sup>7</sup> (e.g. "1 L"; "0.3 kg"; "0.1 m <sup>2</sup> ")
Environment biome	descriptor of the broad ecological context of a sample. Examples include: desert, taiga, deciduous woodland, or coral reef. EnvO terms can be found via the link ( <a href="https://www.ebi.ac.uk/ols/ontologies/envo">https://www.ebi.ac.uk/ols/ontologies/envo</a> ) (e.g. "desert"; "taiga"; "deciduous woodland"; "coral reef")
Latitude and longitude	The geographical coordinates of the location where the sample was collected. Specify as degrees latitude and longitude in format "d[.dddd] N S d[.dddd] W E", e.g., 38.98 N 77.11 W (e.g. "47.9412 N 28.1201 E")
Altitude	The altitude of the sample is the vertical distance between Earth's surface above Sea Level and the sampled position in the air (e.g. "-256 m"; "330.12 m")
Depth	Depth is defined as the vertical distance below surface, e.g. for sediment or soil samples depth is measured from sediment or soil surface, respectively. Depth can be reported as an interval for subsurface samples (e.g. "15m depth")
Temperature	temperature of the sample at time of sampling (e.g. "27 °C")
Genotype	observed genotype (e.g. "SB0140"; "Wild Type")
Host tissue sampled	Type of tissue the initial sample was taken from ("leaves"; "liver")
Disease	List of diseases diagnosed; can include multiple diagnoses (e.g. "Rheumatoid arthritis"; "Lung cancer")
Lab host	Scientific name and description of the laboratory host used to propagate the source organism or material from which the sample was obtained, e.g., Escherichia coli DH5a, or Homo sapiens HeLa cells (e.g. "Escherichia coli DH5a"; "Homo sapiens HeLa cell")
Serotype	Taxonomy below subspecies; a variety (in bacteria, fungi or virus) usually based on its antigenic properties. e.g. serotype="H1N1" in Influenza A virus CY098518 (e.g. "H1N1"; "B1")



## 2. Submit KAD (Experiment)

Description : Write experiment data related to BioSample according to the format below.

### Mandatory Attributes

Sample name	A name that you choose for the sample. It can have any format, but we suggest that you make it concise, unique and consistent within your lab, and as informative as possible. Every Sample Name from a single Submitter must be unique.
Experiment title (English)	"Experiment title. Short description that will identify the dataset on public pages. A clear and concise formula for the title would be like:{methodology} of {organism}: {sample info} (e.g. ""RNA-Seq of Mus musculus: adult female spleen"")
Library name	Short unique identifier for the sequencing library. Each library name MUST be unique! (Exception: libraries of the technical replicates are allowed to have the same library name) (e.g "Library2_AHX1609"; "sample name_Strategy" or "sample name_Source")
Platform	part of Instrument model of Sequencing Platform <sup>*table(1)</sup>
Instrument model	Sequencing platform -> Instrument model select <sup>*table(1)</sup>
Library Construction/Experiment Design	Enter the details about your experimental design and molecular strategies including hybrid selection and affinity capture reagents; any detail that distinguishes your experiment from other similar experiments. This field should describe: - the protocols used to extract and prepare the material to be sequenced - the library construction protocol - name of the library preparation kit
Strategy	Sequencing technique intended for this library. <sup>*table(2)</sup>
Source	The library source specifies the type of source material that is being sequenced. <sup>*table (3)</sup>
Selection	Whether any method was used to select and/or enrich the material being sequenced. <sup>*table (4)</sup>
Release date selection	Select "Release immediately following curation (recommended)" or "Release on specified date"
Release date	If you select "Release date selection" in "Release on specified date", must enter a public date. (form: YYYY-MM-DD, ex, 2018-12-22)
Read 1	File name
Read 2	File name

### Optional Attributes

General description (영문)	Free-form description of the methods, including a brief 'Materials and Methods' section, that were not described in the other fields.
Library	The library descriptor specifies the origin of the material being sequenced and any treatments that the material might have

undergone that affect the sequencing result. This specification is needed even if the platform does not require a library construction step per se. This field may include:

- adapter sequence(s) that can be contained in the sequencing results
- RNA integrity number (RIN) of the sample, if the source of the sample is transcription product

Fragment/Paired read	Library Layout specifies whether to expect Single, Paired-end, or Other configuration of reads. In the case of paired reads, information about the relative distance and orientation is specified. (e.g "Single-end"; "Paired-end"; "Other configuration of reads")
Nominal standard deviation	Standard deviation of insert size (typically ~10% of Nominal Size)(Only Paired read) (e.g "25")
Technical replicate	When libraries are indeed identical (same combination of library + strategy + layout + instrument model), please clarify each library (Run) by using technical replicate identifier. (e.g., "technical_rep_2")
Insert size	[For paired-end (Paired read 만 해당)] Average fragment size for Paired reads. Insert size can be a distance between a pair of adapters and estimated by using bioinformatics tools such as SAMtools after read alignment. You need to fill in either 'Insert size' or 'Nominal size' identifier. (e.g "250")
Nominal size	[For paired-end (Paired read 만 해당)] Size of the insert for Paired reads. The nominal size is the expected size of the insert. The insert being the fragment sequenced, as chosen from size fractionation (e.g., cutting out a band from an agarose gel). If nominal length is 500bp, put down "500" as the value. No decimals or ranges (e.g., 100-200) allowed, and it cannot be zero. This information guides sequence aligner algorithms to place the two mates of a pair of reads on the genome sequences separated by a reasonable distance. Therefore, the nominal length value does not need to be absolutely precise, but also should not be an order of magnitude off. You need to fill in either 'Insert size' or 'Nominal size' identifier. (e.g "250")

### 3. table 1

• Table 1 : Sequencing platform & Instrument model

<i>Platforms</i>		<i>Instrument model for each Platform</i>								
<i>_LS454</i>	<i>_LS454</i>	<i>ILLUMINA</i>	<i>HELICOS</i>	<i>ABI_SOLID</i>	<i>COMPLETE_GENOMICS</i>	<i>PACBIO</i>	<i>ION_TORRENT</i>	<i>CAPILLARY</i>	<i>OXFORD_NANOPORE</i>	<i>BGI_SEQ</i>
<b>ABI_SOLID</b>	454 GS	HiSeq X Five	Helicos HeliScope	AB 5500 Geneti	Comple	PacBio RS	Ion Tor	AB 310 Genet	GridION	BGIS EQ-500

				Genomic Analyzer		nt PGM	ic Analyzer		
<b>BGISEQ</b>	454 GS 20	HiSeq X Ten		AB 5500xl Genetic Analyzer		PacBio RS II	Ion Torrent Proton	AB 3130 Genetic Analyzer	MinION
<b>CAPILLARY</b>	454 GS FLX	Illumina Genome Analyzer		AB 5500x- WL Genetic Analyzer		PacBio Sequel	Ion Torrent S5 XL	AB 3130x L Genetic Analyzer	PromethION
<b>COMPLETE_GENOMICS</b>	454 GS FLX+	Illumina Genome Analyzer II		AB SOLiD 3 Plus System		PacBio Sequel II	Ion Torrent S5	AB 3500 Genetic Analyzer	
<b>HELICOS</b>	454 GS FLX Titanium	Illumina Genome Analyzer IIx		AB SOLiD 4 System				AB 3500x L Genetic Analyzer	
<b>ILLUMINA</b>	454 GS Junior	Illumina HiScan SQ		AB SOLiD 4hq System				AB 3730 Genetic Analyzer	
<b>ION_TORRENT</b>		Illumina HiSeq 1000		AB SOLiD PI System				AB 3730x L Genetic Analyzer	
<b>OXFORD_NANOPORE</b>		Illumina HiSeq 1500		AB SOLiD System					

<b>PACBI O _SMRT</b>	Illumina HiSeq 2000	AB SOLiD System 2.0
	Illumina HiSeq 2500	AB SOLiD System 3.0
	Illumina HiSeq 3000	
	Illumina HiSeq 4000	
	Illumina iSeq 100	
	Illumina NovaSeq 6000	
	Illumina MiniSeq	
	Illumina MiSeq	
	NextSeq q 500	
	NextSeq q 550	

#### **4. table 2**

- Table 2 : Strategy

<b><i>type</i></b>	<b><i>Description</i></b>
<b>WGA</b>	Whole genome amplification. Random sequencing of the whole genome following non-PCR amplification
<b>WGS</b>	Whole genome sequencing. Random sequencing of the whole genome
<b>WXS</b>	Whole exome sequencing. Random sequencing of exonic regions selected from the genome
<b>RNA-Seq</b>	Random sequencing of whole transcriptome

<b>smRNA-Seq</b>	MicroRNA and other small non-coding RNA sequencing
<b>WCS</b>	Random sequencing of a whole chromosome or other replicon isolated from a genome
<b>CLONE</b>	Genomic clone based (hierarchical) sequencing
<b>POOLCLONE</b>	Shotgun of pooled clones (usually BACs and Fosmids)
<b>AMPLICON</b>	Sequencing of overlapping or distinct PCR or RT-PCR products
<b>CLONEEND</b>	Clone end (5', 3', or both) sequencing
<b>FINISHING</b>	Sequencing intended to finish (close) gaps in existing coverage
<b>ChIP-Seq</b>	Direct sequencing of chromatin immunoprecipitates
<b>MNase-Seq</b>	Direct sequencing following MNase digestion
<b>DNase-Hypersensitivity</b>	Sequencing of hypersensitive sites, or segments of open chromatin that are more readily cleaved by DNaseI
<b>Bisulfite-Seq</b>	Sequencing following treatment of DNA with bisulfite to convert cytosine residues to uracil depending on methylation status
<b>Tn-Seq</b>	Gene fitness determination through transposon seeding. Sequencing from transposon insertion sites
<b>EST</b>	Single pass sequencing of cDNA templates
<b>FL-cDNA</b>	Full-length sequencing of cDNA templates
<b>CTS</b>	Concatenated Tag Sequencing
<b>MRE-Seq</b>	Methylation-Sensitive Restriction Enzyme Sequencing strategy
<b>MeDIP-Seq</b>	Methylated DNA Immunoprecipitation Sequencing strategy
<b>MBD-Seq</b>	Direct sequencing of methylated fractions sequencing strategy
<b>Synthetic-Long-Read</b>	Binning and barcoding of large DNA fragments to facilitate assembly of the fragment
<b>ATAC-Seq</b>	Assay for Transposase-Accessible Chromatin (ATAC) strategy is used to study genome-wide chromatin accessibility. Alternative method to DNase-Seq that uses an engineered Tn5 transposase to cleave DNA and to integrate primer DNA sequences into the cleaved genomic DNA
<b>ChIA-PET</b>	Direct sequencing of proximity-ligated chromatin immunoprecipitates
<b>FAIRE-Seq</b>	Formaldehyde Assisted Isolation of Regulatory Elements
<b>Hi-C</b>	Chromosome Conformation Capture technique where a biotin-labeled nucleotide is incorporated at the ligation junction, enabling selective purification of chimeric DNA ligation junctions followed by deep sequencing
<b>ncRNA-Seq</b>	Capture of other non-coding RNA types, including post-translation modification types such as snRNA (small nuclear RNA) or snoRNA (small nucleolar RNA), or expression regulation types such as siRNA (small interfering RNA) or piRNA/piwiRNA (piwi-interacting RNA).
<b>RAD-Seq</b>	Restriction Site Associated DNA Sequence

<b>RIP-Seq</b>	Direct sequencing of RNA immunoprecipitates (includes CLIP-Seq, HITS-CLIP and PAR-CLIP)
<b>SELEX</b>	Systematic Evolution of Ligands by EXponential enrichment
<b>ssRNA-Seq</b>	strand-specific RNA sequencing
<b>Targeted-Capture</b>	Targeted-Capture sequencing
<b>Tethered Chromatin Conformation Capture</b>	Tethered Chromatin Conformation Capture sequencing
<b>OTHER</b>	Library strategy not listed (please include additional info in the "Description" field of Experiment section)

### 5. table 3

- Table 3 : Source

<i>type</i>	<i>Description</i>
<b>GENOMIC</b>	Genomic DNA (includes PCR products from genomic DNA)
<b>EPIGENOMIC</b>	Genomic DNA used for studying epigenetic changes in a cell that affect gene expression without altering the DNA sequence
<b>TRANSCRIPTOMIC</b>	Transcription products or non-genomic DNA (EST, cDNA, RT-PCR, screened libraries)
<b>METATRANSCRIPTOMIC</b>	Transcription products from community targets
<b>METAGENOMIC</b>	Mixed material from metagenome
<b>SYNTHETIC</b>	Synthetic DNA
<b>VIRAL RNA</b>	Viral RNA
<b>OTHER</b>	Other, unspecified, or unknown library source material (please include additional info in the "Description" of Experiment design)

### 6. table 4

- Table 4 : Selection

<i>type</i>	<i>Description</i>
<b>unspecified</b>	Library enrichment, screening, or selection is not specified (please include additional info in the "Description" of Experiment design)
<b>RANDOM</b>	Random selection by shearing or other method
<b>PCR</b>	Source material was selected by designed primers
<b>RANDOM PCR</b>	Source material was selected by randomly generated primers
<b>RT-PCR</b>	Source material was selected by reverse transcription PCR
<b>HMPR</b>	Hypo-methylated partial restriction digest
<b>MF</b>	Methyl Filtrated
<b>CF-S</b>	Cot-filtered single/low-copy genomic DNA
<b>CF-M</b>	Cot-filtered moderately repetitive genomic DNA

<b>CF-H</b>	Cot-filtered highly repetitive genomic DNA
<b>CF-T</b>	Cot-filtered theoretical single-copy genomic DNA
<b>MDA</b>	Multiple displacement amplification
<b>MSLL</b>	Methylation Spanning Linking Library
<b>cDNA</b>	complementary DNA
<b>ChIP</b>	Chromatin immunoprecipitation
<b>MNase</b>	Micrococcal Nuclease (MNase) digestion
<b>DNase</b>	Deoxyribonuclease (DNase) digestion
<b>Hybrid Selection</b>	Selection by hybridization in array or solution
<b>Reduced Representation</b>	Reproducible genomic subsets, often generated by restriction fragment size selection, containing a manageable number of loci to facilitate re-sampling
<b>Restriction Digest</b>	DNA fractionation using restriction enzymes
<b>5-methylcytidine antibody</b>	Selection of methylated DNA fragments using an antibody raised against 5-methylcytosine or 5-methylcytidine (m5C)
<b>MBD2 protein methyl-CpG binding domain</b>	Enrichment by methyl-CpG binding domain
<b>CAGE</b>	Cap-analysis gene expression
<b>RACE</b>	Rapid Amplification of cDNA Ends
<b>size fractionation</b>	Physical selection of size appropriate targets
<b>Padlock probes capture method</b>	Circularized oligonucleotide probes
<b>Poly-A</b>	polyA enriched RNA-Seq
<b>other</b>	Other library enrichment, screening, or selection process (please include additional info in the "Description" of Experiment design)