

데이터 품질관리 기준

◆ 데이터 스테이션 등록 시스템 내 데이터 구성

- ☑ 데이터 품질관리 기준은 시스템에서 자동으로 검증되는 항목을 제외하고, 품질관리자가 메타데이터를 직접 검수 시 기준으로 활용할 수 있는 부분으로 구성함
- ☑ '24년 표준등록양식을 기준으로 BioProject, BioSample 및 표준등록양식 전체에 해당하는 데이터 종류의 메타데이터 품질관리 기준에 대해 작성함
- ☑ '24년 K-BDS 고도화 이후 단백질체·화합물·대사체·바이오이미징 품질선도센터 데이터를 포함한 전 분야의 메타데이터 및 데이터 파일에 대한 품질관리 기준을 작성하고자 함

Checklist Fields					
Filter fields...					
Filter by type:					
non-sample terms		Field Name	Field Format	(Field Restriction)	Requirement (Units)
Collection event information		project name	free text		mandatory
internal environment		experimental factor	free text		optional
sample collection		ploidy	free text		optional
host disorder		number of replicons	restricted text	regular expression	optional
		extrachromosomal elements	restricted text	regular expression	optional

데이터 품질관리 매뉴얼

목차

CONTENT

별첨 1. 데이터 품질관리 기준

01

1-1. BioProject 정보	04
1-1-1. 개요	04
1-1-2. 품질관리 체크리스트	05
파트1. Submitter information (등록자 인적사항)	05
파트2. General information of the project (프로젝트의 개괄적 정보)	06
파트3. Grant information (연구 과제 정보)	07
파트4. Publication and patent information (논문 및 특허 성과 정보)	08
1-2. BioSample 정보	09
1-2-1. 개요	09
1-2-2. 품질관리 기준	10
파트1. Human	10
파트2. Model organism or animal	14
파트3. Invertebrate	19
파트4. Plant	23
파트5. Microbe	27
파트6. Virus	31
파트7. Clinical or host-associated pathogen	35
파트8. Environmental, food, or other pathogen	39
파트9. Metagenome or environment	42
1-3. NGS(차세대시퀀싱) 데이터	44
1-3-1. 개요	44
1-3-2. 품질관리 기준	45
파트1. NGS metadata (차세대시퀀싱 속성정보)	45
파트2. Files	47

데이터 품질관리 매뉴얼

1-4. Microarray(마이크로어레이) 데이터	49
1-4-1. 개요	49
1-4-2. 품질관리 기준	50
파트1. Series (시리즈)	50
파트2. Sample detail (샘플 상세정보)	50
파트3. Protocol (실험 프로토콜)	51
파트4. Result files (결과파일)	51
1-5. Nucleotide sequence(염기서열) 데이터	53
1-5-1. 개요	53
1-5-2. 품질관리 기준	54
파트1. Nucleotide type	54
파트2. Sequencing technology 정보	54
파트3. Sequences/Nucleotide 정보	56
파트4. Feature 정보	57
파트5-1. File 정보 (Case 1)	63
파트5-2. File 정보 (Case 2)	63

1-1. BioProject 정보

1-1-1. 개요

◆ 목적

- ☑ 이 지침은 한국생명공학연구원 국가생명연구자원정보센터 바이오 데이터 스테이션의 표준등록 양식 중 BioProject 정보 품질관리에 대한 기준을 규정하여 업무의 표준화를 기하기 위함임

◆ 적용범위

- ☑ 모든 바이오 연구 데이터 등록의 필수 선행사항으로서 등록자 및 연구 프로젝트에 대한 정보를 품질 관리하기 위한 기준임
- ☑ BioProject 품질관리는 '23년도 표준등록양식 기준 '2. BioProject 정보' 를 기반으로 항목별 품질관리 기준을 규정함
- ☑ '24년도 양식 고도화에 따른 변경 부분은 붉은색으로 기재함

◆ 개정이력

개정번호	개정일	개정사유	작성자	관리기관
1	2022. 11. 01	신규제정	KOBIC	KOBIC
2	2024. 01. 31	'23년 품질관리 시스템 고도화	KOBIC	KOBIC

1-1-2. 품질관리 체크리스트

파트 1. Submitter information (등록자 인적사항)			
No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Name of submitter (등록자의 영문 이름)	M	① 영문 기재 여부
2	Name of submitter in Korean (등록자의 국문 이름)	M	① 한글 기재 여부
3	E-mail of submitter (등록자의 이메일 주소)	M	① 이메일 형태 작성 여부 (abc@kribb.re.kr)
4	Name of submitter's organization (등록자 소속 기관의 영문명)	M	① 기관 홈페이지에 게시된 공식 명칭 사용 [예시] Seoul National University
5	Name of submitter's organization in Korean (등록자 소속 기관의 국문명)	M	① 공공 기관의 경우 공공기관 경영정보 공개시 스템 (ALIO, https://www.alio.go.kr/guide/publicAgencyList.do) 항목과 일치 여부 ② ①에 해당하지 않는 기관의 경우 기관 홈페이지에 게시된 공식 명칭 사용 [예시] 서울대학교
6	Department of submitter (등록자 소속 학과/부서의 영문명)	O	① 영문 기재 여부 [예시] Department of Chemical Engineering
7	Department of submitter in Korean (등록자 소속 학과/부서의 국문명)	O	① 한글 기재 여부 ② 6번 항목과 일치 여부 [예시] 화학공학과
8	Phone number of submitter (등록자의 전화번호)	O	① 국제 전화 번호 형식 기재 여부 [예시] 010-1111-2222
9	Country of submitter (등록자의 국가)	M	① 영문 기재 여부 [예시] South Korea
10	Researcher ID of submitter (등록자의 연구자 고유번호)	M	① 숫자 기재 여부 (8자리) [예시] 10155792

파트 2. General information of the project (프로젝트의 개괄적 정보)

No	항목	필수(M)/ 선택(O)	품질관리 기준				
1	Project title (프로젝트의 영문 제목)	M	① 영문 기재 여부 [예시] <table><tr><th>Examples</th></tr><tr><td>Chromosome Y sequencing</td></tr><tr><td>Opportunistic pathogen that causes important food-borne disease</td></tr><tr><td>Global studies of microbial diversity on human skin</td></tr></table>	Examples	Chromosome Y sequencing	Opportunistic pathogen that causes important food-borne disease	Global studies of microbial diversity on human skin
Examples							
Chromosome Y sequencing							
Opportunistic pathogen that causes important food-borne disease							
Global studies of microbial diversity on human skin							
2	Project title in Korean (프로젝트의 국문 제목)	M	① 한글 기재 여부 [예시] <table><tr><th>예시</th></tr><tr><td>Y 염색체 시퀀싱</td></tr><tr><td>주요 식품매개질환을 일으키는 기회감염병원체</td></tr><tr><td>인간 피부에서의 미생물 다양성에 대한 포괄적 연구</td></tr></table>	예시	Y 염색체 시퀀싱	주요 식품매개질환을 일으키는 기회감염병원체	인간 피부에서의 미생물 다양성에 대한 포괄적 연구
예시							
Y 염색체 시퀀싱							
주요 식품매개질환을 일으키는 기회감염병원체							
인간 피부에서의 미생물 다양성에 대한 포괄적 연구							
3	Project description (프로젝트의 영문 설명)	M	① 영문 기재 여부				
4	Project description in Korean (프로젝트의 국문 설명)	M	① 한글 기재 여부 ② 3번 항목과 일치 여부				
5	Private message to K-BDS admin (관리자에 요청하는 사항)	O					
6	Existence of umbrella BioProject (상위 umbrella BioProject의 유무)	M	① 가능한 값 중 선택 여부				
7	Accession ID of umbrella BioProject (Umbrella BioProject의 accession ID)	M	① 이미 등록된 상위 umbrella BioProject의 accession ID 기재 여부				
8	Title of umbrella BioProject (Umbrella BioProject의 영문 제목)	O	① 영문 기재 여부				

파트 3. Grant information (연구 과제 정보)

No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Major grant ID (주요 연구과제 번호)	M	외부 DB 연동으로 품질관리 기준 없음
2	Major grant NTIS ID (주요 연구과제의 NTIS 번호)	M	
3	Major grant title (주요 연구과제의 영문 제목)	M	
4	Major grant title in Korean (주요 연구과제의 국문 제목)	M	
5	Funding agency (연구비 지원기관)	M	
6	Funding government department (연구비 지원부처)	M	
7	Additional grant ID (추가적인 연구과제 번호)	O	

파트 4. Publication and patent information (논문 및 특허 성과 정보)

No	항목	필수(M)/ 선택(O)	품질관리 기준
1	PubMed ID of publication (논문의 PubMed ID)	O	① PubMed의 기재 여부 [예시] 31974418
2	DOI of publication (논문의 DOI)	O	① 해당 논문의 PubMed ID와 DOI 일치 여부 [예시] 10.1038/s41598-020-57919-6
3	Domestic patent number (국내 특허번호)	O	[예시] 10-2017-0012076
4	International patent number (해외 특허번호)	O	[예시] US20170233479A1

1-2. BioSample 정보

1-2-1. 개요

◆ 목적

- ☑ 이 지침은 한국생명공학연구원 국가생명연구자원정보센터 바이오 데이터 스테이션의 표준등록 양식 중 BioSample 정보 품질관리에 대한 기준을 규정하여 업무의 표준화를 기하기 위함임

◆ 적용범위

- ☑ 유전체, 단백질체 및 대사체 바이오 연구 데이터 등록의 필수 선행사항으로서 등록자 및 연구 프로젝트에 대한 정보를 품질 관리하기 위한 기준
- ☑ BioSample 품질관리는 '23년도 표준등록양식 기준 '3. BioSample 정보'를 기반으로 항목별 품질관리 기준을 규정함
- ☑ 총 9가지 생물군에 따라 입력해야 하는 항목들의 차이가 있어 각 생물군에 따른 양식으로 파트를 나누어 작성함

◆ 개정이력

개정번호	개정일	개정사유	작성자	관리기관
1	2022. 11. 01	신규제정	KOBIC	KOBIC
2	2024. 01. 31	'23년 품질관리 시스템 고도화	KOBIC	KOBIC

데이터 품질관리 기준

1-2-2. 품질관리 기준

파트 1. Human			
No	항목	M/O	품질관리 기준
1	age (나이)	M	① 나이 + 단위 기재 여부 ② 세포주 등 나이가 성립되지 않는 샘플인 경우 “not applicable” ③ 샘플 나이의 범위만 알고 있는 경우 등 정확한 나이를 모를 경우, 이 항목에는 “missing: not collected”를 입력한 후, 기타 특성 (other characteristics) 항목에 알고 있는 나이 정보를 입력 [예시] - 28 years - 2.5 months - 13 weeks - 3 days - 10 embryo age
2	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr. Gildong Hong from Seoul National University - not applicable
3	cell line (세포주명)	O	① 적절한 (검색 가능한) 세포주의 이름인지 확인 [예시] - A549 - HEK293 - HeLa
4	cell subtype (세포 서브타입)	O	[예시] CD4+ T cell
5	cell type (세포 타입)	O	[예시] - T cell - NK cell - epithelial cell
6	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
7	culture collection (배양 컬렉션)	O	① <institution-code>[:<collection-code>][:<culture_id>] 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370

No	항목	M/O	품질관리 기준
8	description (샘플 설명)	O	
9	development stage (발생 상태)	O	[예시] - adult - infant - newborn - fetal - pluripotency
10	disease (질병)	O	① 진단된 질병 목록 기재 여부. 다수 입력 가능 ② 다음 두 종류의 controlled vocabulary 중 하나를 사용: (1) Human Disease Ontology (http://bioportal.bioontology.org/ontologies/1009/) (2) MeSH (Medical Subject Headings) (http://www.ncbi.nlm.nih.gov/mesh/) [예시] Lung adenocarcinoma
11	disease stage (질병단계)	O	[예시] - Stage 2 - NASH1
12	ethnicity (민족)	O	[예시] European-American
13	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
14	growth protocol (배양 프로토콜)	O	[예시] ES cell-derived NS cells were routinely generated by re-plating 7 adherent neural differentiation cultures (typically 2-3 × 10 ⁶ 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).
15	health state (건강 상태)	O	① '질병단계' 항목과 구분되어 샘플 수집 당시 질병 진행 추이로 인한 건강 상태를 기술 [예시] - death - chronic - recovery
16	isolate	M	[예시] - Stage 3 lung cancer patient #17
17	karyotype (핵형)	O	[예시] - copy neutral loss of heterozygosity - 47XY+21_1 - haploid
18	NCBI taxonomy ID (NCBI 분류 ID)	M	① 9606 (고정값)
19	organism (생명체 명)	M	① Homo Sapiens (고정값)

데이터 품질관리 기준

No	항목	M/O	품질관리 기준
20	phenotype (표현형)	O	<p>① Phenotypic Quality Ontology (PATO) 에서 제공하는 controlled vocabulary 입력 여부 (http://bioportal.bioontology.org/visualize/44601/)</p> <p>[예시] antimicrobial humoral immune response mediated by antimicrobial peptide</p>
21	population (집단)	O	<p>① 특정 서식 지역을 기반으로 동일한 그룹 혹은 종의 개체들을 묶어 표현할 수 있는 이름을 작성 ② '생물군 종류' 에서 'Human' 선택 시 코호트 이름 등을 작성</p> <p>[예시] Icelandic</p>
22	race (인종)	O	<p>① 특성명:특성값 포맷 사용 여부 ② '민족' 항목과 구분되어 유전학적으로 나뉠 수 있는 물리적인 특징을 기술하는 항목</p> <p>[예시] - Hair color:black - Eye color:black</p>
23	sample name (샘플명)	M	<p>① 단일 BioProject 내에서 각각 고유한지 확인</p> <p>[예시] - A549 lung adenocarcinoma smRNA-seq 1 - A549 lung adenocarcinoma smRNA-seq 2</p>
24	sample type (샘플 타입)	O	<p>① cell culture, mixed culture, tissue sample, whole organism, single cell 등 이질성(heterogeneity) 관련 사항을 입력하는 항목</p> <p>[예시] - cell culture - mixed culture - tissue sample - whole organism - single cell</p>
25	sex (성별)	M	<p>① 가능한 값 중 선택 여부</p>
26	specimen voucher (표본 식별자)	M	<p>① [<Institution-code>:<collection-code>:]<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 "not applicable" 입력</p> <p>[예시] - KCLB:30082 - not applicable</p>
27	tissue (조직)	M	<p>[예시] - lung - liver</p>
28	treatment (처리 조건)	O	<p>[예시] Embryos were dechorionated with 50% bleach, put on a cover slip and covered with Halocarbon oil 27 (Sigma). Embryos of the appropriate stage were manually selected under the dissecting scope. Selected embryos were transferred to a basket, rinsed with PBS with 0.7% NaCl, 0.04% triton-X100 and placed on ice in the Trizol solution (GibcoBRL).</p>

No	항목	M/O	품질관리 기준
29	other characteristics (기타 특성)	O	<p>① “특성명:특성값” 형식 사용 여부 ② 구분자로 “//”를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음</p> <p>[예시] - BRCA1 mutation status:Yes - KRAS mutation status:Yes//Chemotherapy dosage:High - organ:brain//subregion:hipocampus//neuronal type:GABAergic</p>

파트 2. Model organism or animal			
No	항목	M/O	품질관리 기준
1	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
2	birth date (출생일자)	O	① 샘플링 시점의 나이. 생물종 및 연구에 따라 적절한 단위와 함께 기재해야 함 ② 세포주 등 나이가 성립되지 않는 경우 not applicable을 입력 ③ 정확한 샘플 나이를 모를 경우 missing: not collected를 입력
3	birth location (출생장소)	O	[예시] Seoul, Korea
4	breeding history (품종개량 이력)	O	[예시] Inbred stock Daejeon first season 2021 and Stop F13
5	breeding method (품종개량 방법)	O	[예시] - Inbreeding - Our-crossing - Cross-breeding - Interspecific Hybridization
6	cell line (세포주명)	O	① 적절한 (검색 가능한) 세포주의 이름인지 확인 [예시] - C2C12 - HeLa - NIH3T3
7	cell subtype (세포 서브타입)	O	[예시] CD4+ T cell
8	cell type (세포 타입)	O	[예시] - T cell - NK cell - epithelial cell
9	collected by (수집자)	O	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
10	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z

No	항목	M/O	품질관리 기준
11	culture collection (배양 컬렉션)	O	① <institution-code>:[<collection-code>:]<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370
12	death date (사망일자)	O	① YYYY-mm-dd 또는 YYYY-mm-ddThh:mm:ssZ 포맷 사용 여부 [예시] - 2021-09-10 - 2021-09-10T12:24:10Z
13	description (샘플 설명)	O	
14	disease (질병)	O	① 진단된 질병 목록 기재 여부. 다수 입력 가능 ② 다음 두 종류의 controlled vocabulary 중 하나를 사용: (1) Human Disease Ontology (http://bioportal.bioontology.org/ontologies/1009/) (2) MeSH (Medical Subject Headings) (http://www.ncbi.nlm.nih.gov/mesh/) [예시] - cervical adenocarcinoma - hepatocarcinoma
15	disease stage (질병단계)	O	[예시] - Stage 2 - NASH1
16	genotype (유전형)	O	[예시] - Bhlhe40-/- - IFNAR1 KO - Eed null - Wild type
17	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
18	growth protocol (배양 프로토콜)	O	[예시] ES cell-derived NS cells were routinely generated by re-plating adherent neural differentiation cultures (typically 2-3 × 10 ⁶ 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).
19	health state (건강 상태)	O	① '질병단계' 항목과 구분되어 샘플 수집 당시 질병 진행 추이로 인한 건강 상태를 기술 [예시] - death - chronic - recovery
20	isolation source (분리 소스)	O	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] Oral Secretion

No	항목	M/O	품질관리 기준
21	latitude and longitude (위도 및 경도)	O	① “d[d.ddd] N S d[dd.ddd] W E” 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W
22	NCBI taxonomy ID (NCBI 분류 ID)	M	① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부 [예시] - 10090 - 158481 - NA
23	organism (생명체 명)	M	① 확인종 ② 미확인종: 적절한 속(genus).sp ③ 새로운 종: 원하는 생물명 ④ Virus 분류군 미확인 종: Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱: 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기 [예시] - Mus musculus - Mus sp. Kiwit - Pseudomonas sp. UK4 - (원하는 생물명)
24	phenotype (표현형)	O	① Phenotypic Quality Ontology (PATO) 에서 제공하는 controlled vocabulary 입력 여부 (http://bioportal.bioontology.org/visualize/44601/) [예시] antimicrobial humoral immune response mediated by antimicrobial peptide
25	sample name (샘플명)	M	① 단일 BioProject 내에서 각각 고유한지 확인 [예시] - Mus musculus C2C12 cells H3K4me3 control 1 - Mus musculus C2C12 cells H3K4me3 control 2
26	sample type (샘플 타입)	O	① cell culture, mixed culture, tissue sample, whole organism, single cell 등 이질성(heterogeneity) 관련 사항을 입력하는 항목 [예시] - cell culture - mixed culture - tissue sample - whole organism - single cell
27	sex (성별)	M	① 가능한 값 중 선택 여부
28	specimen voucher (표본 식별자)	M	① [Institution-code]:[collection-code]:[specimen_id] 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 “not applicable” 입력 [예시] - KCLB:30082 - not applicable

No	항목	M/O	품질관리 기준
29	storage conditions (보관 조건)	O	[예시] -80 degree Celsius, 96 hours
30	stud book number (혈통 번호)	O	① 적절한 (검색 가능한) 동물의 혈통 번호 여부 [예시] - 3220DA - F17-D-F03
31	tissue (조직)	M	[예시] - lung - liver
32	treatment (처리 조건)	O	[예시] Embryos were dechorionated with 50% bleach, put on a cover slip and covered with Halocarbon oil 27 (Sigma). Embryos of the appropriate stage were manually selected under the dissecting scope. Selected embryos were transferred to a basket, rinsed with PBS with 0.7% NaCl, 0.04% triton-X100 and placed on ice in the Trizol solution (GibcoBRL).
33	other characteristics (기타 특성)	O	① “특성명:특성값” 형식 사용 여부 ② 구분자로 “//”를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음 [예시] - BRCA1 mutation status:Yes - KRAS mutation status:Yes//Chemotherapy dosage:High - organ:brain//subregion:hipocampus//neuronal type:GABAergic

No	그룹항목(G)	품질관리 기준
Age/Stage Group		
1	age (나이)	<p>① 나이 + 단위 기재 여부 ② 세포주 등 나이가 성립되지 않는 샘플인 경우 “not applicable” ③ 샘플 나이의 범위만 알고 있는 경우 등 정확한 나이를 모를 경우, 이 항목에는 “missing: not collected”를 입력한 후, 기타 특성(other characteristics) 항목에 알고 있는 나이 정보를 입력</p> <p>[예시] - 28 years - 2.5 months - 13 weeks - 3 days - 10 embryo age</p>
2	development stage (발생 상태)	<p>[예시] - adult - infant - newborn - fetal - pluripotency</p>
Organism Group		
1	breed (품종명)	<p>[예시] - Beagle - Angean - Angora - mixed</p>
2	cultivar (재배품종명)	<p>[예시] CS69111</p>
3	ecotype (생태형)	<p>[예시] Columbia</p>
4	isolate	<p>[예시] Stage 3 lung cancer #17</p>
5	strain (균주, 계통)	<p>① 적절한 (검색 가능한) 균주, 계통 여부</p> <p>[예시] - BALB/c - C57BL/6</p>

파트 3. Invertebrate			
No	항목	M/O	품질관리 기준
1	age (나이)	O	① 나이 + 단위 기재 여부 ② 세포주 등 나이가 성립되지 않는 샘플인 경우 “not applicable” ③ 샘플 나이의 범위만 알고 있는 경우 등 정확한 나이를 모를 경우, 이 항목에는 “missing: not collected”를 입력한 후, 기타 특성 (other characteristics) 항목에 알고 있는 나이 정보를 입력 [예시] - 28 years - 2.5 months - 13 weeks - 3 days - 10 embryo age
2	altitude (고도)	O	① 고도값 + 단위 기재 여부 [예시] 15.2 m
3	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr. Gildong Hong from Seoul National University - not applicable
4	broad-scale environmental context (주요 환경 유형)	O	① 환경 유형 [ENVO:환경 유형 ID] 포맷 사용 여부 (Environment Ontology(ENVO)에서 biome(생물군계)의 하위 클래스에 있는 용어 및 그 용어의 ID) ② 다수 기재 시 파이프() 구분자 사용 환경 유형1 [ENVO:환경 유형1 ID] 환경 유형2 [ENVO:환경 유형2 ID] [예시] - mangrove biome [ENVO:01000181] - mangrove biome [ENVO:01000181] estuarine biome [ENVO:01000020]
5	collected by (수집자)	O	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
6	collection date (수집일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z

No	항목	M/O	품질관리 기준
7	depth (깊이)	O	① 깊이 값 + 단위 기재 여부 [예시] 2.5 m
8	description (샘플 설명)	O	
9	development stage (발생 상태)	O	[예시] - embryo - adult
10	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
11	growth protocol (배양 프로토콜)	O	[예시] 30 min egg collections of OreR and yw flies at 25 degree Celsius were aged at room temperature (RT) according to the different temporal classes T0-T4.
12	host tissue sampled (숙주 조직)	O	① BRENDA Tissue and Enzyme Source Ontology 제공 용어 사용 여부 (http://bioportal.bioontology.org/ontologies/1005/) [예시] - ovarian fluid - dental plaque
13	identified by (식별한 자)	O	① 제출한 샘플을 찾은 실험실의 연구자 이름을 기재하는 것이 아니라 샘플의 Taxonomy를 처음 발견한 분류학자의 이름을 기재 [예시] Dan Janzen
14	latitude and longitude (위도 및 경도)	O	① “d[d.ddd] N S d[dd.ddd] W E” 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W
15	NCBI taxonomy ID (NCBI 분류 ID)	M	① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부 [예시] - 78423 - 2790805 - NA
16	organism (생명체 명)	M	① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기 [예시] - Sepioteuthis sepioidea - Sepioteuthis sp. HAW01 - (원하는 생물명)

No	항목	M/O	품질관리 기준
17	sample name (샘플명)	M	<p>① 단일 BioProject 내에서 각각 고유한지 확인</p> <p>[예시]</p> <ul style="list-style-type: none"> - <i>Sepioteuthis sepioidea</i> WGS sample 1 - <i>Sepioteuthis sepioidea</i> WGS sample 2
18	sex (성별)	O	<p>① 가능한 값 중 선택 여부</p>
19	specimen voucher (표본 식별자)	M	<p>① [(Institution-code):[collection-code]:]<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용)</p> <p>② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력</p> <p>③ 해당 없는 경우 "not applicable" 입력</p> <p>[예시]</p> <ul style="list-style-type: none"> - KCLB:30082 - not applicable
20	temperature (온도)	O	<p>① 섭씨: degree Celsius, 화씨: degree Fahrenheit 단위 기재 여부</p> <p>[예시]</p> <ul style="list-style-type: none"> - 24 degree Celsius - 75.2 degree Fahrenheit
21	tissue (조직)	M	<p>[예시]</p> <p>whole organism</p>
22	treatment (처리 조건)	O	<p>[예시]</p> <p>Embryos were dechorionated with 50% bleach, put on a cover slip and covered with Halocarbon oil 27 (Sigma). Embryos of the appropriate stage were manually selected under the dissecting scope. Selected embryos were transferred to a basket, rinsed with PBS with 0.7% NaCl, 0.04% triton-X100 and placed on ice in the Trizol solution (GibcoBRL).</p>
23	other characteristics (기타 특성)	O	<p>① "특성명:특성값" 형식 사용 여부</p> <p>② 구분자로 "/"를 사용했는지 여부</p> <p>③ 작성이 불필요한 경우 이 항목을 입력하지 않음</p> <p>[예시]</p> <ul style="list-style-type: none"> - BRCA1 mutation status:Yes - KRAS mutation status:Yes//Chemotherapy dosage:High - organ:brain//subregion:hipocampus//neuronal type:GABAergic

No	그룹항목(G)	품질관리 기준
Organism group		
1	breed (품종명)	[예시] - Koi - Comet - Ranchu - mixed
2	isolate	[예시] Stage 3 lung cancer sample #17
Source group		
1	host (숙주)	① 완전한 분류명 기재 여부 ② '실험실 숙주' 항목과 구분되어 실험을 목적으로 이용한 숙주가 아닌 자연감염 숙주를 기입 [예시] - Homo sapiens - Mus musculus
2	isolation source (분리 소스)	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] surface seawater

파트 4. Plant			
No	항목	M/O	품질관리 기준
1	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
2	cell line (세포주명)	O	① 적절한 (검색 가능한) 세포주의 이름인지 확인 [예시] PSB-D
3	cell type (세포 타입)	O	[예시] - parenchyma cell - xylem cell
4	collected by (수집자)	O	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
5	collection date (수집일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
6	culture collection (배양 콜렉션)	O	① <institution-code>:[<collection-code>:]<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 콜렉션코드 사용) [예시] ATCC:26370
7	description (샘플 설명)	O	
8	disease (질병)	O	① 진단된 질병 목록 기재 여부. 다수 입력 가능 ② 다음 두 종류의 controlled vocabulary 중 하나를 사용: (1) Human Disease Ontology (http://bioportal.bioontology.org/ontologies/1009/) (2) MeSH (Medical Subject Headings) (http://www.ncbi.nlm.nih.gov/mesh/) [예시] - rice stripe virus - clubroot
9	disease stage (질병 단계)	O	[예시] - stage 3 - initial infection stage

No	항목	M/O	품질관리 기준
10	genotype (유전형)	O	[예시] - asil1-1 asia2-2 - djc31 djc62 double mutant
11	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
12	growth protocol (배양 프로토콜)	O	[예시] Col-0 cell culture grown in Gamborgs B5 basal medium and on a rotary shaker at 160 rpm under constant light at 23 degree Celsius. The cells were subcultured every 7 days with a 1:10 (inoculum: fresh medium) dilution, as described in Lee et al, 2010.
13	height or length (높이 또는 길이)	O	① 높이 값 (길이 값) + 단위 기재 여부 [예시] 1.2 m
14	isolation source (분리 소스)	O	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] Potato Leaf
15	latitude and longitude (위도 및 경도)	O	① "d[d.ddd] N S d[dd.ddd] W E" 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W
16	NCBI taxonomy ID (NCBI 분류 ID)	M	① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부 [예시] - 3711 - 2695493 - NA
17	organism (생명체 명)	M	① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기 [예시] - Brassica rapa - Brassica sp. CP351 - (원하는 생물명)
18	phenotype (표현형)	O	① Phenotypic Quality Ontology (PATO) 에서 제공하는 controlled vocabulary 입력 여부 (http://bioportal.bioontology.org/visualize/44601/) [예시] antimicrobial humoral immune response mediated by antimicrobial peptide

No	항목	M/O	품질관리 기준
19	population (집단)	O	① 특정 서식 지역을 기반으로 동일한 그룹 혹은 종의 개체들을 묶어 표현할 수 있는 이름을 작성 ② '생물군 종류' 에서 'Plant' 선택 시 세대, 자손 수, 유전적 구조 등을 가리키는 이름 작성 [예시] F2 population
20	sample name (샘플명)	M	[예시] Late heart stage Brassica embryos RNA-seq 1
21	sample type (샘플 타입)	O	① cell culture, mixed culture, tissue sample, whole organism, single cell 등 이질성(heterogeneity) 관련 사항을 입력하는 항목 [예시] - cell culture - mixed culture - tissue sample - whole organism - single cell
22	sex (성별)	O	① 가능한 값 중 선택 여부
23	specimen voucher (표본 식별자)	M	① [<Institution-code>:<collection-code>:<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 콜렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 "not applicable" 입력 [예시] - KCLB:30082 - not applicable
24	temperature (온도)	O	① 섭씨: degree Celsius, 화씨: degree Fahrenheit 단위 기재 여부 [예시] - 24 degree Celsius - 75.2 degree Fahrenheit
25	tissue (조직)	M	[예시] - plantlet - mesophyll
26	treatment (처리 조건)	O	[예시] - wounding-treated post 2 hours - fig22 treatment
27	other characteristics (기타 특성)	O	① "특성명:특성값" 형식 사용 여부 ② 구분자로 "/"를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음 [예시] - ABA-deficient mutation status:Yes - XTC2 mutation status:Yes//transformed leaf rate:High - organ:leaves//subregion:mesophyll//type:spongy mesophyll

No	그룹항목(G)	품질관리 기준
Age/Stage Group		
1	age (나이)	① 나이 + 단위 기재 여부 ② 세포주 등 나이가 성립되지 않는 샘플인 경우 “not applicable” ③ 샘플 나이의 범위만 알고 있는 경우 등 정확한 나이를 모를 경우, 이 항목에는 “missing: not collected”를 입력한 후, 기타 특성(other characteristics) 항목에 알고 있는 나이 정보를 입력 [예시] - 28 years - 2.5 months - 13 weeks - 3 days
2	development stage (발생 상태)	[예시] - late heart stage - four-leaves stage - germination stage
Organism Group		
1	cultivar (재배품종명)	[예시] CS69111
2	ecotype (생태형)	[예시] Columbia
3	isolate	[예시] - Stage 3 lung cancer patient #17 - 2-year-old Panax ginseng

파트 5. Microbe			
No	항목	M/O	품질관리 기준
1	altitude (고도)	O	① 고도값 + 단위 기재 여부 [예시] 15.2 m
2	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 "not applicable" 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
3	broad-scale environmental context (주요 환경 유형)	O	① 환경 유형 [ENVO:환경 유형 ID] 포맷 사용 여부 (Environment Ontology(ENVO)에서 biome(생물군계)의 하위 클래스에 있는 용어 및 그 용어의 ID) ② 다수 기재 시 파이프(I) 구분자 사용 환경 유형1 [ENVO:환경 유형1 ID] 환경 유형2 [ENVO:환경 유형2 ID] [예시] - mangrove biome [ENVO:01000181] - mangrove biome [ENVO:01000181] estuarine biome [ENVO:01000020]
4	collected by (수집자)	O	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
5	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 "/" 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
6	culture collection (배양 컬렉션)	O	① <institution-code>:[<collection-code>]:<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370
7	depth (깊이)	O	① 깊이 값 + 단위 기재 여부 [예시] 2.5 m
8	description (샘플 설명)	O	
9	ecotype (생태형)	O	[예시] yellow mycelium

No	항목	M/O	품질관리 기준
10	geographic location (지리적 장소)	M	<p>① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/</p> <p>② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함</p> <p>[예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps</p>
11	growth protocol (배양 프로토콜)	O	<p>[예시] 30 min egg collections of OreR and yw flies at 25 degree Celsius were aged at room temperature (RT) according to the different temporal classes T0-T4.</p>
12	host tissue sampled (숙주 조직)	O	<p>① BRENDA Tissue and Enzyme Source Ontology 제공 용어 사용 여부 (http://bioportal.bioontology.org/ontologies/1005/)</p> <p>[예시] - ovarian fluid - dental plaque</p>
13	identified by (식별한 자)	O	<p>① 제출한 샘플을 찾은 실험실의 연구자 이름을 기재하는 것이 아니라 샘플의 Taxonomy를 처음 발견한 분류학자의 이름을 기재</p> <p>[예시] Dan Janzen</p>
14	lab host (실험실 숙주)	O	<p>① 실험실 숙주의 적절한 학명 및 설명 기재 여부</p> <p>[예시] - Canis familiaris - Escherichia coli DH5a</p>
15	latitude and longitude (위도 및 경도)	O	<p>① “d[d.ddd] N S d[dd.ddd] W E” 포맷 사용 여부</p> <p>[예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W</p>
16	mating type (교배 형태)	O	<p>[예시] MAT alpha</p>
17	NCBI taxonomy ID (NCBI 분류 ID)	M	<p>① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부</p> <p>[예시] - 239935 - 1945963 - NA</p>
18	organism (생명체 명)	M	<p>① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기</p> <p>[예시] - Akkermansia muciniphila - Akkermansia sp. UBA3271 - (원하는 생물명)</p>
19	passage history (계대 배양 이력)	O	<p>[예시] - Brain heart infusion Broth, 5 - Gentle cell dissociation reagent (STENCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seed at a 1:6 split ratio</p>

No	항목	M/O	품질관리 기준
20	sample name (샘플명)	M	<p>① 단일 BioProject 내에서 각각 고유한지 확인</p> <p>[예시] - Akkemansia muciniphila WGS sample 1 - Akkemansia muciniphila WGS sample 2</p>
21	sample size (샘플 사이즈)	O	<p>① 사이즈 값 + 단위 기재 여부</p> <p>[예시] - 1 L - 0.3 kg - 0.2 m²</p>
22	sample type (샘플 타입)	M	<p>① cell culture, mixed culture, tissue sample, whole organism, single cell 등 이질성(heterogeneity) 관련 사항을 입력하는 항목</p> <p>[예시] - cell culture - mixed culture - tissue sample - whole organism - single cell</p>
23	serotype (항원형)	O	<p>[예시] serotype="H1N1" in Influenza A virus CY098518</p>
24	serovar (혈청형)	O	<p>[예시] Leptospira, serovar saopaulo S76607 (65357 in Entrez)</p>
25	specimen voucher (표본 식별자)	M	<p>① [<Institution-code>:<collection-code>:<specimen_id>] 포맷 사용 여부 (INSDC 제공 기관코드 및 콜렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 "not applicable" 입력</p> <p>[예시] - KCLB:30082 - not applicable</p>
26	temperature (온도)	O	<p>① 섭씨: degree Celsius, 화씨: degree Fahrenheit 단위 기재 여부</p> <p>[예시] - 24 degree Celsius - 75.2 degree Fahrenheit</p>
27	treatment (처리 조건)	O	<p>[예시] - diamide treatment - 0.05% SDS treatment</p>
28	other characteristics (기타 특성)	O	<p>① "특성명:특성값" 형식 사용 여부 ② 구분자로 "/"를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음</p> <p>[예시] - relationship to oxygen:facultative - relationship to oxygen:facultative//subspecies:enterica - relationship to oxygen:facultative//subspecies:enterica//environmental medium:soil</p>

No	그룹항목(G)	품질관리 기준
Organism group		
1	isolate	[예시] Stage 3 lung cancer sample #17
2	strain (균주, 계통)	① 적절한 (검색 가능한) 균주, 계통 여부 [예시] MG1234
Source group		
1	host (숙주)	① 완전한 분류명 기재 여부 ② '실험실 숙주' 항목과 구분되어 실험을 목적으로 이용한 숙주가 아닌 자연감염 숙주를 기입 [예시] - Homo sapiens - Mus musculus
3	isolation source (분리 소스)	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] Oral Secretion

파트 6. Virus			
No	항목	M/O	품질관리 기준
1	altitude (고도)	O	① 고도값 + 단위 기재 여부 [예시] 15.2 m
2	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 "not applicable" 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
3	broad-scale environmental context (주요 환경 유형)	O	① 환경 유형 [ENVO:환경 유형 ID] 포맷 사용 여부 (Environment Ontology(ENVO)에서 biome(생물군계)의 하위 클래스에 있는 용어 및 그 용어의 ID) ② 다수 기재 시 파이프(I) 구분자 사용 환경 유형1 [ENVO:환경 유형1 ID] 환경 유형2 [ENVO:환경 유형2 ID] [예시] - mangrove biome [ENVO:01000181] - mangrove biome [ENVO:01000181] estuarine biome [ENVO:01000020]
4	collected by (수집자)	O	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
5	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 "/" 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
6	culture collection (배양 컬렉션)	O	① <institution-code>:[<collection-code>]:<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370
7	depth (깊이)	O	① 깊이 값 + 단위 기재 여부 [예시] 2.5 m
8	description (샘플 설명)	O	

No	항목	M/O	품질관리 기준
9	genotype (유전형)	O	[예시] - 27 G - 68/73
10	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
11	growth protocol (배양 프로토콜)	O	[예시] 30 min egg collections of OreR and yw flies at 25 degree Celsius were aged at room temperature (RT) according to the different temporal classes T0-T4.
12	host tissue sampled (숙주 조직)	O	① BRENDA Tissue and Enzyme Source Ontology 제공 용어 사용 여부 (http://bioportal.bioontology.org/ontologies/1005/) [예시] - ovarian fluid - dental plaque
13	identified by (식별한 자)	O	① 제출한 샘플을 찾은 실험실의 연구자 이름을 기재하는 것이 아니라 샘플의 Taxonomy를 처음 발견한 분류학자의 이름을 기재 [예시] Dan Janzen
14	isolate	M	[예시] Tomato Leaf
15	isolation source (분리 소스)	M	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] Oral Secretion
16	latitude and longitude (위도 및 경도)	O	① "[d.dddd] N[S d[dd.dddd] W]E" 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - 23.456 S -10.234 W
17	NCBI taxonomy ID (NCBI 분류 ID)	M	① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부 [예시] - 507474 - 1498504 - NA
18	organism (생명체 명)	M	① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기 [예시] - Tomato leaf curl Cebu virus - Tomato leaf curl Barka virus - (원하는 생물명)

No	항목	M/O	품질관리 기준
19	passage history (계대 배양 이력)	M	<p>[예시]</p> <ul style="list-style-type: none"> - Brain heart infusion Broth, 5 - Gentle cell dissociation reagent(STENCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seed at a 1:6 split ratio
20	sample name (샘플명)	O	<p>① 단일 BioProject 내에서 각각 고유한지 확인</p> <p>[예시]</p> <p>Leguminosae-associated picorna-like virus 1 nonstructural protein genes, partial CDS</p>
21	sample size (샘플 사이즈)	M	<p>① 사이즈 값 + 단위 기재 여부</p> <p>[예시]</p> <ul style="list-style-type: none"> - 1 L - 0.3 kg - 0.2 m²
22	serotype (항원형)	O	<p>[예시]</p> <p>serotype="H1N1" in Influenza A virus CY098518</p>
23	specimen voucher (표본 식별자)	M	<p>① [<Institution-code>:<collection-code>:]<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용)</p> <p>② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력</p> <p>③ 해당 없는 경우 "not applicable" 입력</p> <p>[예시]</p> <ul style="list-style-type: none"> - KCLB:30082 - not applicable
24	strain (균주, 계통)	O	<p>① 적절한 (검색 가능한) 균주, 계통 여부</p> <p>[예시]</p> <p>Influenza A/Aichi/1968(H3N2)</p>
25	temperature (온도)	O	<p>① 섭씨: degree Celsius, 화씨: degree Fahrenheit 단위 기재 여부</p> <p>[예시]</p> <ul style="list-style-type: none"> - 24 degree Celsius - 75.2 degree Fahrenheit
26	treatment (처리 조건)	O	<p>[예시]</p> <ul style="list-style-type: none"> - 48h infection - 100nM 17AAG in 0.1% DMSO for 48h
27	other characteristics (기타 특성)	O	<p>① "특성명:특성값" 형식 사용 여부</p> <p>② 구분자로 "/"를 사용했는지 여부</p> <p>③ 작성이 불필요한 경우 이 항목을 입력하지 않음</p> <p>[예시]</p> <ul style="list-style-type: none"> - replicate:biological replicate - replicate:biological replicate 3//subspecies:carotovorum

No	그룹항목(G)	품질관리 기준
Host group		
1	host (숙주)	① 완전한 분류명 기재 여부 ② '실험실 숙주' 항목과 구분되어 실험을 목적으로 이용한 숙주가 아닌 자연감염 숙주를 기입 [예시] - Escherichia coli - Phaseolus vulgaris
2	lab host (실험실 숙주)	① 실험실 숙주의 적절한 학명 및 설명 기재 여부 [예시] Epithelium papillosum cyprini (EPC) cell line

파트 7. Clinical or host-associated pathogen			
No	항목	M/O	품질관리 기준
1	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
2	collected by (수집자)	M	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
3	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
4	culture collection (배양 컬렉션)	O	① <institution-code>:{<collection-code>}<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370
5	description (샘플 설명)	O	
6	genotype (유전형)	O	[예시] - 27G - 68/73
7	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
8	growth protocol (배양 프로토콜)	O	[예시] 30 min egg collections of OreR and yw flies at 25 degree Celsius were aged at room temperature (RT) according to the different temporal classes T0-T4.

No	항목	M/O	품질관리 기준
9	host (숙주)	M	<p>① 완전한 분류명 기재 여부 ② '실험실 숙주' 항목과 구분되어 실험을 목적으로 이용한 숙주가 아닌 자연감염 숙주를 기입</p> <p>[예시] - Homo sapiens - Mus musculus</p>
10	host age (숙주 나이)	O	<p>① 나이 + 단위 기재 여부 ② 세포주 등 나이가 성립되지 않는 샘플인 경우 "not applicable" ③ 샘플 나이의 범위만 알고 있는 경우 등 정확한 나이를 모를 경우, 이 항목에는 "missing: not collected"를 입력한 후, 기타 특성(other characteristics) 항목에 알고 있는 나이 정보를 입력</p> <p>[예시] - 28 years - 2.5 months - 13 weeks - 3 days</p>
11	host description (숙주 추가 정보)	O	<p>[예시] ethnicity: European-American, genetic disorder: Yes European-American, genetic</p>
12	host disease (숙주 질병)	M	<p>① 다음 두 종류의 controlled vocabulary 중 하나를 사용: (1) Human Disease Ontology (http://bioportal.bioontology.org/ontologies/1009/) (2) MeSH (Medical Subject Headings) (http://www.ncbi.nlm.nih.gov/mesh/)</p> <p>[예시] - vibrio cholerae - vibrio infections - Salmonella gastroenteritis</p>
13	host disease outcome (숙주 질병 결과)	O	<p>[예시] - death - chronic disease - recovery</p>
14	host disease stage (숙주 질병 단계)	O	<p>[예시] - Stage 2 - NASH1</p>
15	host health state (숙주 건강 상태)	O	<p>① '질병단계' 항목과 구분되어 샘플 수집 당시 질병 진행 추이로 인한 건강 상태를 기입 여부</p> <p>[예시] - death - chronic disease - recovery</p>
16	host sex (숙주 성별)	O	<p>① 가능한 값 중 선택 여부</p>
17	host subject id (숙주 개체 ID)	O	<p>[예시] #131</p>

No	항목	M/O	품질관리 기준
18	host tissue sampled (숙주 조직)	O	<p>① BRENDA Tissue and Enzyme Source Ontology 제공 용어 사용 여부 (http://bioportal.bioontology.org/ontologies/1005/)</p> <p>[예시] - ovarian fluid - dental plaque</p>
19	isolation source (분리 소스)	M	<p>① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부</p> <p>[예시] Oral Secretion</p>
20	latitude and longitude (위도 및 경도)	M	<p>① “d[d.ddd] N S d[dd.ddd] W E” 포맷 사용 여부</p> <p>[예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W</p>
21	NCBI taxonomy ID (NCBI 분류 ID)	M	<p>① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부</p> <p>[예시] - 666 - 161723 - NA</p>
22	organism (생명체 명)	M	<p>① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기</p> <p>[예시] - Vibrio cholerae - Vibrio sp. 01 - (원하는 생물명)</p>
23	passage history (계대 배양 이력)	O	<p>[예시] - Brain heart infusion Broth, 5 - Gentle cell dissociation reagent (STENCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seed at a 1:6 split ratio</p>
24	pathotype (병원성)	O	<p>[예시] - Escherichia coli - STEC, UPEC - Extended-spectrum beta-lactamase (ESBL)</p>
25	sample name (샘플명)	M	<p>① 단일 BioProject 내에서 각각 고유한지 확인</p> <p>[예시] ChiP-seq of the Vibrio cholerae AphA protein3</p>
26	serotype (항원형)	O	<p>[예시] serotype=“H1N1” in Influenza A virus CY098518</p>
27	serovar (혈청형)	O	<p>[예시] Leptospira, serovar saopaulo S76607 (65357 in Entrez)</p>

No	항목	M/O	품질관리 기준
28	specimen voucher (표본 식별자)	M	① [<Institution-code>:<collection-code>:]<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 “not applicable” 입력 [예시] - KCLB:30082 - not applicable
29	subgroup (서브그룹)	O	[예시] Clostridium botulinum Group I
30	subtype (서브타입)	O	[예시] HIV type 1, Group M, Subtype A
31	treatment (처리 조건)	O	[예시] - 0.05% SDS treatment - diamide treatment
32	other characteristics (기타 특성)	O	① “특성명:특성값” 형식 사용 여부 ② 구분자로 “//”를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음 [예시] - isolate name alias:ROSS 308 broiler - isolate name alias:ROSS 308 broiler// purpose of sampling:Diagnostic testing - anatomical part:Anterior Nares//purpose of sequencing:baseline surveillance//body product:Mucus

No	그룹항목(G)	품질관리 기준
Organism group		
1	isolate	[예시] Stage 3 lung cancer sample #17
2	strain (균주, 계통)	① 적절한 (검색 가능한) 균주, 계통 여부 [예시] MG1234

파트 8. Environmental, food, or other pathogen			
No	항목	M/O	품질관리 기준
1	biomaterial provider (생물질 제공자)	M	① 소재은행명 또는 기관명 또는 개인 연구자명 또는 해당 사항이 없다면 “not applicable” 기재 여부 ② 등록자가 biomaterial을 제공받은 가장 최근의 제공자(Provider)를 입력해야만 함 [예시] - Korean Cell Line Bank (KCBL) - Korean Collection for Type Cultures (KCTC) - Dr.Gildong Hong from Seoul National University - not applicable
2	collected by (수집자)	M	① 개인 연구자명 또는 기관명 기재 여부 [예시] - Hong Gildong - Korea Bioinformation Center (KOBIC)
3	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 “/” 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
4	culture collection (배양 컬렉션)	O	① <institution-code>:{<collection-code>}<culture_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) [예시] ATCC:26370
5	description (샘플 설명)	O	
6	genotype (유전형)	O	[예시] SB0140
7	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
8	growth protocol (배양 프로토콜)	O	[예시] 30 min egg collections of OreR and yw flies at 25 degree Celsius were aged at room temperature (RT) according to the different temporal classes T0-T4.
9	isolation source (분리 소스)	M	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] permanent Antarctica sea ice

No	항목	M/O	품질관리 기준
10	latitude and longitude (위도 및 경도)	M	① “d[d.ddd] N S d[dd.ddd] W E” 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W
11	NCBI taxonomy ID (NCBI 분류 ID)	M	① NCBI가 부여하는 분류 ID 또는 해당 사항이 없다면 NA 기재 여부 [예시] - 446 - 907781 - NA
12	organism (생명체 명)	M	① 확인종 ② 미확인종 : 적절한 속(genus).sp ③ 새로운 종 : 원하는 생물명 ④ Virus 분류군 미확인 종 : Taxonomy ID에 적합한 생명체 명 ⑤ 비-메타게놈 시퀀싱 : 균주 또는 분리주의 이름도 입력 (③을 제외하고 NCBI taxonomy ID에 해당하는 학명 기재) ⑥ 식물의 경우, 국가표준식물목록에 등재된 이름 표기 [예시] - Legionella pneumophila - Legionella sp.11127 - (원하는 생물명)
13	passage history (계대 배양 이력)	O	[예시] - Brain Heart Infusion Broth, 5 - Gentle cell dissociation reagent(STENCELL Technologies, 07174) was used to passage cPP cells as aggregates which were then seed at a 1:6 split ratio
14	pathotype (병원성)	O	[예시] - Escherichia coli - STEC, UPEC - Extended-spectrum beta-lactamase (ESBL)
15	sample name (샘플명)	M	① 단일 BioProject 내에서 각각 고유한지 확인 [예시] Water-related pathogen sample WGS 2
16	serotype (항원형)	O	[예시] serotype=“H1N1” in Influenza A virus CY098518
17	serovar (혈청형)	O	[예시] Leptospira, serovar saopaulo S76607 (65357 in Entrez)
18	specimen voucher (표본 식별자)	M	① [(Institution-code):[collection-code]:]<specimen_id> 포맷 사용 여부 (INSDC 제공 기관코드 및 컬렉션코드 사용) ② 등록되지 않은 기관의 경우 영문기관명(영문기관약자) 형태로 institution code 입력 ③ 해당 없는 경우 “not applicable” 입력 [예시] - KCLB:30082 - not applicable
19	subgroup (서브그룹)	O	[예시] Clostridium botulinum Group I
20	subtype (서브타입)	O	[예시] HIV type 1, Group M, Subtype A

No	항목	M/O	품질관리 기준
21	treatment (처리 조건)	O	[예시] - 0.05% SDS treatment - diamide treatment
22	other characteristics (기타 특성)	O	① “특성명:특성값” 형식 사용 여부 ② 구분자로 “//”를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음 [예시] - host:Gallus gallus domesticus - host:Gallus gallus domesticus //sequence_type:315 - subspecies: enterica //sequence_type:315 //potential_contamination:none

No	그룹항목(G)	품질관리 기준
Organism group		
1	isolate	[예시] Stage 3 lung cancer sample #17
2	strain (균주, 계통)	① 적절한 (검색 가능한) 균주, 계통 여부 [예시] MG1234

파트 9. Metagenome or environment			
No	항목	M/O	품질관리 기준
1	collection date (수집 일자)	M	① 형식을 지켰는지 확인 (ISO 8601 국제 표준 규격) ② 두 개의 날짜 및 시간의 범위 입력 시 "/" 구분자 사용 여부 [예시] - 1990-10-30 - 1990-10 - 1990 - 1952-10-21/1953-02-15 - 2015-10-11T17:53:03Z
2	description (샘플 설명)	O	
3	geographic location (지리적 장소)	M	① 적절한 용어 사용 여부. 적절한 이름(국가, 해양 등)은 다음 리스트에 나와 있음 : http://www.insdc.org/documents/country-qualifier-vocabulary/ ② 적절한 양식 사용 여부. 국가/해양 등의 이름 보다 더 자세한 지리적 위치는 국가/해양 등의 이름 다음에 콜론(:)을 붙인 후 입력함 [예시] - Canada: Vancouver - Germany: halfway down Zugspitze, Alps
4	latitude and longitude (위도 및 경도)	M	① "d[d.ddd] N S d[dd.ddd] W E" 포맷 사용 여부 [예시] - 47.9412 N 28.1201 E - -23.456 S -10.234 W
5	NCBI taxonomy ID (NCBI 분류 ID)	M	① 256318 (고정값)
6	organism (생명체 명)	M	① metagenome (고정값)
7	reference for biomaterial (생물질 참고문헌)	O	① DOI 또는 PubMed ID 또는 URL 포맷 사용 여부 [예시] - 10.1038/s41586-032-03689-8 - 32234345 - https://pubmed.ncbi.nlm.nih.gov/34224345
8	relationship to oxygen (산소 관련성)	O	① 가능한 값 중 선택 여부
9	sample collection device or method (샘플 수집 장치 또는 방법)	O	[예시] - Seabird 9/11 plus CTD - centrifuge - bronchoscopy
10	sample material processing (샘플 물질 처리 방법)	O	[예시] - DNA extraction - RNAlater - mild sonication, wet sieving, centrifugation

No	항목	M/O	품질관리 기준
11	sample name (샘플명)	M	① 단일 BioProject 내에서 각각 고유한지 확인 [예시] Metagenome sample from Natural attenuation in co-immobilized experiment
12	sample size (샘플 사이즈)	O	① 사이즈 값 + 단위 기재 여부 [예시] - 1 L - 0.3 kg - 0.2 m ²
13	sample type (샘플 타입)	O	① cell culture, mixed culture, tissue sample, whole organism, single cell 등 이질성(heterogeneity) 관련 사항을 입력하는 항목 [예시] - cell culture - mixed culture - tissue sample - whole organism - single cell
14	source material identifiers (원천 물질 ID)	O	[예시] - AG0246-39-058 - HM415
15	other characteristics (기타 특성)	O	① “특성명:특성값” 형식 사용 여부 ② 구분자로 “//”를 사용했는지 여부 ③ 작성이 불필요한 경우 이 항목을 입력하지 않음 [예시] - Perturbation:salt - Tissue: pancreatic tissue//Broad-scale environmental context:00009003 - Host age:5 weeks//Plant body site:root//Air temperature regiment:22 degree Celsius

No	그룹항목(G)	품질관리 기준
Source group		
1	host (숙주)	① 완전한 분류명 기재 여부 ② ‘실험실 숙주’ 항목과 구분되어 실험을 목적으로 이용한 숙주가 아닌 자연감염 숙주를 기입 [예시] - Homo sapiens - Mus musculus
2	isolation source (분리 소스)	① 샘플이 유래된 물리적, 환경적 및/또는 지역적 지리 출처 여부 [예시] surface seawater

1-3. NGS(차세대시퀀싱) 데이터

1-3-1. 개요

◆ 목적

- ☑ 이 지침은 한국생명공학연구원 국가생명연구자원정보센터 바이오 데이터 스테이션의 표준등록 양식 중 NGS (차세대시퀀싱) 데이터 품질관리에 대한 기준을 규정하여 업무의 표준화를 기하기 위함임

◆ 적용범위

- ☑ DNA 및 RNA 서열의 NGS (Next-Generation Sequencing) 기반 high-throughput 시퀀싱 데이터를 품질 관리하기 위한 기준
- ☑ NGS (차세대시퀀싱) 데이터는 시퀀싱 데이터를 생산할 때 선택한 분석기법 및 결과에 대한 정보로, '23년도 표준등록양식 기준 '4.1. NGS (차세대시퀀싱) 데이터'를 기반으로 항목별 품질관리 기준을 규정함

◆ 개정이력

개정번호	개정일	개정사유	작성자	관리기관
1	2022. 11. 01	신규제정	KOBIC	KOBIC
2	2024. 01. 31	'23년 품질관리 시스템 고도화	KOBIC	KOBIC

1-3-2. 품질관리 기준

파트 1. NGS metadata (차세대시퀀싱 속성정보)			
No	항목	필수(M)/선택(O)	품질관리 기준
1	Title (제목)	M	① {서열 생산 방법} of {생물종} : {샘플 정보} 포맷 사용 권장 [예시] RNA-Seq of Mus musculus: adult female spleen.
2	Data identifier (데이터의 식별자)	M	[예시] ① 한 샘플에 대하여 두 개의 paired-end fastq 파일이 생산된 경우: 샘플에 해당하는 식별자를 입력 - ID0001_R1.fastq 및 ID0001_R2.fastq 파일이면 ID0001 입력 ② 동일 샘플에 대하여 여러 번의 technical replication이 이루어진 후 각 replicate별로 fastq 파일이 생산된 경우: 샘플 및 replicate 조합에 해당하는 식별자를 입력 - ID0002_rep1.fastq, ID0002_rep2.fastq 등의 파일이면 ID0002_rep1 또는 ID0002_rep2 입력 ③ 동일 샘플에 대하여 여러 번의 library prep이 이루어진 후 각 prep별로 fastq 파일이 생산된 경우: 샘플 및 prep 조합에 해당하는 식별자를 입력 - ID0002_prep1.fastq, ID0002_prep2.fastq 등의 파일이면 ID0002_prep1 또는 ID0002_prep2 입력
3	Library strategy (시퀀싱 기법)	M	① 가능한 값 중 선택 여부
4	Library source (라이브러리 출처)	M	① 가능한 값 중 선택 여부
5	Library selection (라이브러리 선택 방법)	M	① 가능한 값 중 선택 여부
6	Library construction protocol (라이브러리 제작 프로토콜)	M	① DNA fragmentation, ligation, enrichment 에 대한 프로토콜 기술 여부 ② library preparation kit 를 사용한 경우, 키트의 이름과 버전 기술 권장 (예: Illumina Nextera DNA Library Preparation Kit)
7	Library strandedness (라이브러리 방향성)	O	① 가능한 값 중 선택 여부
8	Library layout (라이브러리 레이아웃)	M	① 가능한 값 중 선택 여부 [비고] Long-read sequencing인 경우 옵션 1을 선택
9	Insert length (Insert 길이)	O	① 단위: bp ② 숫자 기재 여부 [예시] 100 bp [비고] Paired-end read 데이터인 경우에만 해당

No	항목	필수(M)/ 선택(O)	품질관리 기준
10	Platform (플랫폼)	M	① 가능한 값 중 선택 여부
11	Instrument model (기기모델)	M	① 가능한 값 중 선택 여부
12	Design description (디자인 설명)	M	

파트 2. Files			
No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Type of primary data file (주요 데이터 파일의 타입)	M	① 가능한 값 중 선택 여부 [비고] Long-read 데이터는 fastq 또는 fasta 형식 업로드 가능. Short-read 데이터는 fastq 형식으로 업로드
2	Name of primary data file (주요 데이터 파일의 이름)	M	① 확장자를 포함한 파일 이름의 파일 업로드 여부 ② md5 파일 업로드 여부 [비고] “Library layout” 항목의 값이 “paired” 인 경우, 쌍방향 서열 분석의 경우 두 개의 파일을 등록하게 되어 있으므로 파일 이름도 두 개를 입력. 압축형식은 gzip만 허용
3	Type of secondary data file (부가 데이터 파일의 타입)	O	① 가능한 값 중 선택 여부 [비고] 여러 부가 데이터 파일 업로드 가능
4	Name of secondary data file (부가 데이터 파일의 이름)	O	① 확장자를 포함한 파일 이름의 파일 업로드 여부 ② md5 파일 업로드 여부
5	Description of secondary data file (부가 데이터 파일 설명)	O	[비고] - 부가 데이터 파일을 업로드한 경우만 해당 - Reference 사용 시 chrX, chrY, chrM 제거 등 특이사항 명시 여부
6	Type of reference data (참조 서열의 타입)	O	① 가능한 값 중 선택 여부 [비고] 주요 데이터 파일에 BAM 파일을 업로드 한 경우 및 부가 데이터 파일을 업로드한 경우 해당. 단, PacBio나 IonTorrent로 생산된 BAM을 제출한 경우, 참조 서열과 매핑 테이블 제출 불필요
7	RefSeq accession ID (표준 참조 서열의 accession ID)	O	① 6번 항목에서 ‘RefSeq’ 선택 여부 ② ‘accession.version’ 포맷 사용 여부 [예시] NC_000001.11
8	Name of reference data file (참조 서열 데이터 파일 이름)	O	① 6번 항목에서 ‘Assembly’ 선택 여부 ② multi-fasta 포맷의 파일 업로드 여부

No	항목	필수(M)/ 선택(O)	품질관리 기준
9	SN-reference mapping table (SN-참조 서열 간 연결 테이블)	O	① 탭으로 구분한 텍스트 파일 업로드 여부
10	Release date (공개날짜)	M	① YYYY-MM-DD 포맷 사용 여부 [예시] 2018-12-22

1-4. Microarray (마이크로어레이) 데이터

1-4-1. 개요

◆ 목적

- ☑ 이 지침은 한국생명공학연구원 국가생명연구자원정보센터 바이오 데이터 스테이션의 표준등록 양식 중 Microarray (마이크로어레이) 데이터 품질관리에 대한 기준을 규정하여 업무의 표준화를 기하기 위함임

◆ 적용범위

- ☑ DNA, RNA 및 protein을 microarray 기법으로 profiling 한 실험 데이터
- ☑ Microarray (마이크로어레이) 데이터는 Gene expression, Methylation, SNP, Protein 등 모든 종류의 array를 포함하며, '23년도 표준등록양식 기준 '5. Microarray (마이크로어레이) 데이터'를 기반으로 항목별 품질관리 기준을 규정함

◆ 개정이력

개정번호	개정일	개정사유	작성자	관리기관
1	2022. 11. 01	신규제정	KOBIC	KOBIC
2	2024. 01. 31	'23년 품질관리 시스템 고도화	KOBIC	KOBIC

1-4-2. 품질관리 기준

파트 1. Series (시리즈)			
No	항목	필수(M)/선택(O)	품질관리 기준
1	Title (제목)	M	① 120자 이내 여부 [예시] Expression data from early Drosophila embryo
2	Overall design (전반적 디자인)	M	[예시] Drosophila embryos were selected at successive stages of early development for RNA extraction and hybridization on Affymetrix microarrays. We sought to obtain homogeneous populations of embryos at each developmental stage in order to increase the temporal resolution of expression profiles. To that end, we hand-selected embryos according to morphological criteria at five time-points: before pole cell formation, i.e. before zygotic transcription (T0), during the slow phase (T1) and the fast phase (T2) of cellularisation and at the beginning (T3) and the end (T4) of gastrulation.

파트 2. Sample detail (샘플 상세정보)			
No	항목	필수(M)/선택(O)	품질관리 기준
1	detailed sample title (상세 샘플 제목)	M	① [생물질]-[실험조건]-[반복회차] 포맷 사용 권장 [예시] embryo at T0, biological rep1, Muscle_exercised_60min_rep2
2	source name (소스 명)	M	[예시] Drosophila embryos before nuclear cycle 9 (maternal transcripts), vastus lateralis muscle, exercised, 60 min.
3	molecule (분자)	M	① 가능한 값 중 선택 여부
4	label (라벨)	M	[예시] biotin, Cy3, Cy5, ³³ P
5	description (설명)	M	[예시] Gene expression data from embryos younger than nuclear cycle 9, i.e. before zygotic genome activation.
6	platform (플랫폼)	M	① GPLxxx 포맷 사용 여부 (GEO 플랫폼(GPL) 식별번호 링크: https://ncbi.nlm.nih.gov/geo/browse/?view=platforms.) [예시] GPL6805 [비고] GPL 번호가 없는 경우(custom array 등) 파트 4의 processed data file(가공 데이터 파일) 항목에 annotation data table 업로드 여부

파트 3. Protocol (실험 프로토콜)			
No	항목	필수(M)/ 선택(O)	품질관리 기준
1	extract protocol (추출 프로토콜)	M	[예시] Trizol extraction of total RNA was performed according to the manufacturer's instructions.
2	label protocol (라벨링 프로토콜)	M	[예시] Biotinylated cRNA were prepared according to the standard Affymetrix protocol from 6 ug total RNA (Expression Analysis Technical Manual, 2001, Affymetrix).
3	hybridization protocol (혼성화 프로토콜)	M	[예시] Following fragmentation, 10 ug of cRNA were hybridized for 16 hr at 45°C on GeneChip Drosophila Genome Array. GeneChips were washed and stained in the Affymetrix Fluidics Station 400.
4	scan protocol (스캔 프로토콜)	M	[예시] GeneChips were scanned using the Hewlett-Packard GeneArray Scanner G2500A.
5	data processing (데이터 프로세싱)	M	[예시] The data were analyzed with Microarray Suite version 5.0 (MAS 5.0) using Affymetrix default analysis settings and global scaling as normalization method. The trimmed mean target intensity of each array was arbitrarily set to 100.

파트 4. Result files (결과파일)

No	항목	필수(M)/ 선택(O)	품질관리 기준										
1	raw data file (원시 데이터 파일)	M	<div> <div> <div>① 원시 데이터 파일 업로드 여부</div> <div>② 파트 2의 “detailed sample title(상세 샘플 제목)”과 동일하게 작성하되 띄어쓰기 대신 ‘_’로 분리 표기하여 등록 여부</div> </div> <div> <div>[예시]</div> <div>Embryo_at_T0_biological_rep1, Muscle_exercised_60min_rep2</div> </div> </div> <table> <tr> <th>Vendor</th> <th>Format</th> </tr> <tr> <td>Affymetrix</td> <td>CEL</td> </tr> <tr> <td>Agilent</td> <td>Agilent Feature Extraction result file. GPR files are also acceptable if they include the Agilent Probe Names. Depending on the platform, tab-delimited text files are also acceptable.</td> </tr> <tr> <td>Nimblegen</td> <td>PAIR</td> </tr> <tr> <td>Illumina</td> <td>IDAT and associated files or a Matrix worksheet containing non-normalized data</td> </tr> </table>	Vendor	Format	Affymetrix	CEL	Agilent	Agilent Feature Extraction result file. GPR files are also acceptable if they include the Agilent Probe Names. Depending on the platform, tab-delimited text files are also acceptable.	Nimblegen	PAIR	Illumina	IDAT and associated files or a Matrix worksheet containing non-normalized data
Vendor	Format												
Affymetrix	CEL												
Agilent	Agilent Feature Extraction result file. GPR files are also acceptable if they include the Agilent Probe Names. Depending on the platform, tab-delimited text files are also acceptable.												
Nimblegen	PAIR												
Illumina	IDAT and associated files or a Matrix worksheet containing non-normalized data												
2	processed data file (가공 데이터 파일)	M	<div> <div>① 가공된 데이터 파일 업로드 여부</div> <div>[예시]</div> <div>log2 transformation, Quantile RMA normalized</div> </div>										

No	항목	필수(M)/ 선택(O)	품질관리 기준										
			<div>[비고]</div> <div><div>- raw data file (원시 데이터 파일)을 분석한 결과 데이터 테이블을 업로드하며, 결과 값의 정량 지표를 표기</div><div>- 파트 2 “platform” 항목에 GPL 번호가 없는 경우(custom array 등) 파트 4의 processed data file(가공 데이터 파일) 항목의 annotation data table 업로드</div></div> <div><table><tr><th>Vendor</th><th>Description</th></tr><tr><td>Affymetrix</td><td>Usually probe set summary data generated by the primary analysis software (e.g., Expression Console, Microarray Suite 5.0, Genotyping Console, GTYPE/CNAT, GTGS, Tiling Array Software, or GeneChip-compatible/other 3rd-party software). These data may be submitted either as CHP files or a tab-delimited text file (see examples in templates below). Please submit the data used to draw the conclusions of your study. For instance, do not submit CHP files analyzed with MAS5.0 if your submission is related to a publication based on GC-RMA data. In this case, you should submit the GC-RMA probe set summary data instead of MAS5.0 CHP files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_affy.html)</td></tr><tr><td>Agilent</td><td>Processed/normalized data used to draw the conclusions from your study. Two-color experiments (e.g., Cy3/Cy5) will typically have log (test/control) ratios; one-color experiments will have normalized signal intensities. Processed data should be provided as a value matrix containing data for all Samples. Data can be provided in an Excel worksheet or a tab-delimited text file. The table should include the Agilent Probe Names (e.g., A_23_P158231). (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_agil.html)</td></tr><tr><td>Nimblegen</td><td>A Matrix worksheet containing the processed/normalized data used to draw conclusions from your study (e.g., RMA-normalized signals, scaled log2 ChIP/Input ratios, etc.). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). Alternatively, if your data were processed with Nimblescan software, the processed data may be supplied as native .gff files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_nimb.html)</td></tr><tr><td>Illumina</td><td>A Matrix worksheet containing the final processed/normalized data used to draw the conclusions from your study (e.g., cubic spline). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). If submitting a new Illumina Platform, also include Platform annotation columns. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_illu.html)</td></tr></table></div>	Vendor	Description	Affymetrix	Usually probe set summary data generated by the primary analysis software (e.g., Expression Console, Microarray Suite 5.0, Genotyping Console, GTYPE/CNAT, GTGS, Tiling Array Software, or GeneChip-compatible/other 3rd-party software). These data may be submitted either as CHP files or a tab-delimited text file (see examples in templates below). Please submit the data used to draw the conclusions of your study. For instance, do not submit CHP files analyzed with MAS5.0 if your submission is related to a publication based on GC-RMA data. In this case, you should submit the GC-RMA probe set summary data instead of MAS5.0 CHP files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_affy.html)	Agilent	Processed/normalized data used to draw the conclusions from your study. Two-color experiments (e.g., Cy3/Cy5) will typically have log (test/control) ratios; one-color experiments will have normalized signal intensities. Processed data should be provided as a value matrix containing data for all Samples. Data can be provided in an Excel worksheet or a tab-delimited text file. The table should include the Agilent Probe Names (e.g., A_23_P158231). (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_agil.html)	Nimblegen	A Matrix worksheet containing the processed/normalized data used to draw conclusions from your study (e.g., RMA-normalized signals, scaled log2 ChIP/Input ratios, etc.). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). Alternatively, if your data were processed with Nimblescan software, the processed data may be supplied as native .gff files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_nimb.html)	Illumina	A Matrix worksheet containing the final processed/normalized data used to draw the conclusions from your study (e.g., cubic spline). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). If submitting a new Illumina Platform, also include Platform annotation columns. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_illu.html)
Vendor	Description												
Affymetrix	Usually probe set summary data generated by the primary analysis software (e.g., Expression Console, Microarray Suite 5.0, Genotyping Console, GTYPE/CNAT, GTGS, Tiling Array Software, or GeneChip-compatible/other 3rd-party software). These data may be submitted either as CHP files or a tab-delimited text file (see examples in templates below). Please submit the data used to draw the conclusions of your study. For instance, do not submit CHP files analyzed with MAS5.0 if your submission is related to a publication based on GC-RMA data. In this case, you should submit the GC-RMA probe set summary data instead of MAS5.0 CHP files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_affy.html)												
Agilent	Processed/normalized data used to draw the conclusions from your study. Two-color experiments (e.g., Cy3/Cy5) will typically have log (test/control) ratios; one-color experiments will have normalized signal intensities. Processed data should be provided as a value matrix containing data for all Samples. Data can be provided in an Excel worksheet or a tab-delimited text file. The table should include the Agilent Probe Names (e.g., A_23_P158231). (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_agil.html)												
Nimblegen	A Matrix worksheet containing the processed/normalized data used to draw conclusions from your study (e.g., RMA-normalized signals, scaled log2 ChIP/Input ratios, etc.). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). Alternatively, if your data were processed with Nimblescan software, the processed data may be supplied as native .gff files. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_nimb.html)												
Illumina	A Matrix worksheet containing the final processed/normalized data used to draw the conclusions from your study (e.g., cubic spline). If the processed data row count exceeds Excel's limit, please supply your Matrix as a tab-delimited, plain-text file (.txt). If submitting a new Illumina Platform, also include Platform annotation columns. (Taken from https://www.ncbi.nlm.nih.gov/geo/info/geo_illu.html)												
3	Release date (공개날짜)	M	<div>① YYYY-MM-DD 포맷 사용 여부</div> <div>[예시]</div> <div>2018-12-22</div>										

1-5. Nucleotide sequence (염기서열) 데이터

1-5-1. 개요

◆ 목적

- ☑ 이 지침은 한국생명공학연구원 국가생명연구자원정보센터 바이오 데이터 스테이션의 표준등록 양식 중 Nucleotide sequence (염기서열) 데이터 품질관리에 대한 기준을 규정하여 업무의 표준화를 기하기 위함임

◆ 적용범위

- ☑ Nucleotide sequence (염기서열)와 관련한 정보를 입력하기 위한 양식으로, NGS raw read sequence와 PCR primer sequence의 경우는 본 양식이 아니라 NGS 데이터의 양식과 PCR 프라이머 품질관리 기준을 각각 적용함

◆ 개정이력

개정번호	개정일	개정사유	작성자	관리기관
1	2022. 11. 01	신규제정	KOBIC	KOBIC
2	2024. 01. 31	'23년 품질관리 시스템 고도화	KOBIC	KOBIC

1-5-2. 품질관리 기준

파트 1. Nucleotide type			
No	항목	필수(M)/선택(O)	품질관리 기준
1	Assemble type (염기서열 형식)	M	① 가능한 값 중 선택 여부
2	Submission type (데이터 등록 형식)	M	① 가능한 값 중 선택 여부 [비고] - Case 1: 1. Single sequence, 6. Targeted locus study, 7. Other (Please specify) - Case 2: 2. Whole genome sequence, 3. Transcriptome shotgun assembly (TSA) sequences, 4. High-throughput genome (HTG) sequences, 5. Third party annotation (해당 genome이 있는 경우) • 이후 작성 순서 - Case 1: 파트 3 ⇨ 파트 4 ⇨ 파트 5-1 - Case 2: 파트 2 ⇨ 파트 4 ⇨ 파트 5-2

파트 2. Sequencing technology 정보			
No	항목	필수(M)/선택(O)	품질관리 기준
1	Assembly date (어셈블리 일자)	O	① YYYY-MM-DD 포맷 사용 여부
2	Assembly method (어셈블리 방식)	M	① 가능한 값 중 선택 여부 [비고] Other를 선택한 경우 정확한 알고리즘 기입
3	Program version (프로그램 버전)	M	① 우선적으로 어셈블리에 사용한 프로그램 버전 기입. 없는 경우 실행 시 날짜 기입 여부 [예시] 2.3 or MAY-2011
4	Assembly name (어셈블리명)	O	[비고] - 일반적으로 기관에 대한 짧은 3-4 자 약어, 생명체 이름의 4-6 자 약어 및 버전 권장 - 생명체 이름 약어의 경우 “속 + 종” 형식(예 : Pan troglodytes의 경우 “PTro” 또는 “PaTrog” 또는 “Patro”) 을 지킬 것. 예: UCLA_Agam_2.1 (이름 지정 권장 사항 참조) [예시] UCLA_Agam_2.1 for version 2.1 of the <i>Anopheles gambiae</i> genome from a lab in UCLA.

No	항목	M/O	품질관리 기준
5	Genome coverage (게놈 커버리지)	M	① 숫자 기재 여부 [예시] 12x [비고] - 유전체 전반에 걸친 추정된 염기 범위로 반드시 최종 어셈블리를 바탕으로 이를 계산해서 기입 - (염기서열 수 / 예상 게놈 크기) × 최종 어셈블리에 배열된 염기의 백분율 = 어셈블리에 배열된 염기서열 수 / 예상 게놈 크기 - 염기서열 수 / 예상 게놈 크기로 간단히 나타낼 수도 있음
6	Total raw read length (전체 원시데이터 길이)	M	① 단위 : bp ② 숫자 기재 여부
7	Read throughput (총 염기서열 read 수)	M	① 숫자 기재 여부
8	The number of contigs (전체 contig 수)	M	① 숫자 기재 여부 [비고] 전체 Contig 수를 적되, 필요한 경우 Scaffold나 Chromosome을 적을 수 있음
9	Contig length (전체 contig 길이)	M	① 단위 : bp ② 숫자 기재 여부
10	N50	M	① 단위 : bp ② 숫자 기재 여부
11	Sequencing technology (시퀀싱 기술)	M	① 가능한 값 중 선택 여부 [비고] Other를 선택한 경우 정확한 시퀀싱 명칭 기재 여부
12	Confirm full genome (전체 게놈 포함 여부 확인)	M	① 가능한 값 중 선택 여부
13	Description for subset of the genome (게놈 영역 설명)	M	[비고] “Confirm full genome” 항목의 값이 “No”인 경우에만 해당
14	Confirm final version (최종버전 여부)	M	① 가능한 값 중 선택 여부 [비고] 해당 게놈의 더 많은 시퀀싱 또는 어셈블리를 기대하지 않는 경우 “Yes”로 대답
15	Confirm de novo assembly (드 노보 어셈블리 여부)	M	① 가능한 값 중 선택 여부

No	항목	필수(M)/ 선택(O)	품질관리 기준
16	Reference assembly name or accession (참조 어셈블리 이름)	M	[비고] - “Confirm de novo assembly” 항목의 값이 “No”인 경우에만 해당 - NCBI의 Accession Number 등 Public Repository에서 확인할 수 있는 어셈블리여야 하며, 그렇지 않은 경우 참조 어셈블리를 먼저 등록하고 이후에 어셈블리를 등록할 것
17	Confirm update of existing submission (업데이트 여부)	M	① 가능한 값 중 선택 여부 [비고] 기존 존재하는 제출물의 업데이트 데이터인지에 대한 확인
18	Existing genome accession (기존 게놈 등록번호)	M	[비고] - “Confirm update of existing submission” 항목의 값이 Yes인 경우에만 해당 - 기존 존재하는 제출물과 연동하여야 하기에 확인 등록번호 확인필요

파트 3. Sequences/Nucleotide 정보

No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Molecule Type (분자 유형)	M	① 가능한 값 중 선택 여부
2	Topology (위상)	M	① 가능한 값 중 선택 여부
3	Source Organelle/ Location information (원천 소기관 /위치 정보)	M	① 가능한 값 중 선택 여부
4	Chimera check (키메라 서열 확인)	M	① 가능한 값 중 선택 여부
5	Chimera check program name (키메라 확인 프로그램명)	O	[비고] “Chimera check” 항목에서 Yes일 경우 해당되는 항목
6	Chimera check program version (키메라 확인 프로그램 버전)	O	[비고] “Chimera check” 항목에서 Yes일 경우 해당되는 항목

No	항목	필수(M)/선택(O)	품질관리 기준
7	Cultured or Uncultured (배양 여부)	M	① 가능한 값 중 선택 여부 [비고] Bacteria 및 Archaea 시퀀스인 경우, 시퀀스를 배양된 샘플로부터 얻었는지 배양되지 않은 샘플로부터 얻었는지 선택
8	Primer Type (프라이머 유형)	M	① 가능한 값 중 선택 여부 [비고] “Cultured or uncultured” 항목의 값이 “Uncultured, bulk environmental DNA”인 경우 해당되는 항목
9	Forward primer name (포워드 프라이머명)	O	
10	Forward primer sequence (포워드 프라이머 시퀀스)	O	① Case 1 에 해당하는 경우 IUPAC에서 제공하는 염기서열 표기방식 이용한 기입 여부
11	Reverse primer name (리버스 프라이머명)	O	
12	Reverse primer sequence (리버스 프라이머 시퀀스)	O	① Case 1 에 해당하는 경우 IUPAC에서 제공하는 염기서열 표기방식 이용한 기입 여부

파트 4. Feature 정보

※ 하나 이상의 Feature 정보 기입 여부 및 각 엑셀의 기입양식 준용여부
 ※ Case 2의 경우 Feature가 없더라도 받을 수 있음

No	항목	필수(M)/선택(O)	품질관리 기준
Coding region (CDS) / Gene / mRNA			
1	Providing intervals	O	① 인트론 서열 포함 CDS, 유전자, mRNA 서열 - 작성하여 업로드할 엑셀파일 양식: PI.file.xlsx
2	Providing protein sequence data	O	① 단백질 서열 포함 CDS, 유전자, mRNA 서열 - 작성하여 업로드할 엑셀파일 양식: PP.file.xlsx
RNA			
3	precursor RNA	O	① 번역이 일어나지 않는 서열을 포함한 premature RNA - 작성하여 업로드할 엑셀파일 양식: PreRNA.file.xlsx

No	항목	필수(M)/ 선택(O)	품질관리 기준
4	messenger RNA	O	① Splicing 이후 비번역부위를 포함한 mRNA (cDNA) - 작성하여 업로드할 엑셀파일 양식: mRNA.file.xlsx
5	transfer RNA	O	① 핵산 서열을 아미노산 서열로 번역하는 역할을 하는 tRNA - 작성하여 업로드할 엑셀파일 양식: tRNA.file.xlsx
6	ribosomal RNA	O	① Ribonucleoprotein 구조의 RNA 서열 - 작성하여 업로드할 엑셀파일 양식: rRNA.file.xlsx
7	non-coding RNA	O	① 비암호화 유전자 - 작성하여 업로드할 엑셀파일 양식: ncRNA.file.xlsx
8	transfer-messenger RNA	O	① tRNA와 rRNA의 특성을 모두 갖는 RNA 분자 - 작성하여 업로드할 엑셀파일 양식: tmRNA.file.xlsx
9	miscellaneous RNA	O	① 다른 RNA 항목로 정의할 수 없는 모든 전사체 또는 RNA - 작성하여 업로드할 엑셀파일 양식: miscRNA.file.xlsx
Repeat sequence / mobile element / satellites			
10	repeat region	O	① 반복 서열 단위를 포함하는 유전체 영역 - 작성하여 업로드할 엑셀파일 양식: repeat.file.xlsx
Regulatory feature			
11	-35 Signal	O	① 원핵생물 프로모터의 -35 box - 작성하여 업로드할 엑셀파일 양식: 35Sig.file.xlsx
12	-10 Signal	O	① 원핵생물 프로모터의 Pribnow 상자 - 작성하여 업로드할 엑셀파일 양식: 10Sig.file.xlsx
13	Attenuator	O	① 전사 종결에 관여된 서열 - 작성하여 업로드할 엑셀파일 양식: Attenuator.file.xlsx
14	CAAT_signal	O	① 진핵생물의 CAAT box - 작성하여 업로드할 엑셀파일 양식: CAAT.file.xlsx
15	DNase I Hypersensitive Site	O	① DNase I 효소에 의한 절단이 빈번히 일어나는 염색질 영역 - 작성하여 업로드할 엑셀파일 양식: DNaseHyp.file.xlsx
16	Enhancer	O	① 시스 위치에서 프로모터 기능을 하는 인핸서 - 작성하여 업로드할 엑셀파일 양식: Enhancer.file.xlsx

No	항목	필수(M)/ 선택(O)	품질관리 기준
17	Enhancer Blocking Element	O	① 인핸서 억제 요소 - 작성하여 업로드할 엑셀파일 양식: EBE.file.xlsx
18	GC_signal	O	① 진핵생물 프로모터의 GC box - 작성하여 업로드할 엑셀파일 양식: GC.file.xlsx
19	Imprinting Control Region	O	① Need to Update / 부모 imprinting과 관련 - 작성하여 업로드할 엑셀파일 양식: ICR.file.xlsx
20	Insulator	O	① 먼 영역에서 조절하는 전사조절요소 - 작성하여 업로드할 엑셀파일 양식: Insulator.file.xlsx
21	Locus Control Region	O	① 연관 유전자의 발현을 증가시키는 먼 영역의 전사조절 요소 - 작성하여 업로드할 엑셀파일 양식: LCR.file.xlsx
22	Matrix Attachment Region	O	① 핵기질이 부착되는 진핵생물 염색체의 DNA 서열 - 작성하여 업로드할 엑셀파일 양식: MAR.file.xlsx
23	PolyA Signal	O	① 폴리아데닐화를 위한 절단 신호 - 작성하여 업로드할 엑셀파일 양식: PolyA.file.xlsx
24	Promoter	O	① 전사 개시에 관여하는 영역 - 작성하여 업로드할 엑셀파일 양식: Promoter.file.xlsx
25	Recoding Stimulatory Region	O	① 코드 재작성 촉진 영역 - 작성하여 업로드할 엑셀파일 양식: RSR.file.xlsx
26	Replication Regulatory Region	O	① 복제 조절 영역 - 작성하여 업로드할 엑셀파일 양식: RRR.file.xlsx
27	Response Element	O	① 특정 전사인자에 결합할 수 있는 프로모터 또는 인핸서 영역 내의 짧은 DNA 서열 - 작성하여 업로드할 엑셀파일 양식: RE.file.xlsx
28	Ribosome Binding Site	O	① mRNA 전사체의 시작 코돈 upstream의 염기서열 - 작성하여 업로드할 엑셀파일 양식: RBS.file.xlsx
29	Riboswitch	O	① 작은 분자구조에 결합하는 mRNA 분자의 조절 세그먼트 - 작성하여 업로드할 엑셀파일 양식: RiboSwi.file.xlsx

No	항목	필수(M)/ 선택(O)	품질관리 기준
30	Silencer	O	① 시스(cis-) 위치에서 발현 억제를 하는 서열 - 작성하여 업로드할 엑셀파일 양식: Silencer.file.xlsx
31	TATA signal	O	① 진핵생물의 TATA box - 작성하여 업로드할 엑셀파일 양식: TATA.file.xlsx
32	Terminator	O	① 전사 종결을 조절하는 서열 - 작성하여 업로드할 엑셀파일 양식: Terminator.file.xlsx
33	Transcriptional Cis Regulatory Region	O	① 전사 조절을 하는 Cis 영역 - 작성하여 업로드할 엑셀파일 양식: TCRR.file.xlsx
34	Other Regulatory Region	O	① 분류되지 않는 조절 영역 - 작성하여 업로드할 엑셀파일 양식: OtherRegul.file.xlsx
Others			
35	C_region	O	① 면역글로불린 경쇄 및 중쇄 그리고 T 세포 수용체 알파, 베타 및 감마 사슬의 보존 영역 - 작성하여 업로드할 엑셀파일 양식: Cregion.file.xlsx
36	D-loop	O	① 분리된 이종가닥 DNA에서 세 번째 DNA 가닥에 의해 분리되어 있는 DNA 구조 - 작성하여 업로드할 엑셀파일 양식: Dloop.file.xlsx
37	D_segment	O	① 면역글로불린 경쇄 및 중쇄 그리고 T 세포 수용체 알파, 베타 및 감마 사슬의 다양한 영역 - 작성하여 업로드할 엑셀파일 양식: Dseg.file.xlsx
38	Exon	O	① 엑손 서열만 있는 경우 - 작성하여 업로드할 엑셀파일 양식: Exon.file.xlsx
39	Gap	O	① 염기서열 간의 비결정 서열영역 - 작성하여 업로드할 엑셀파일 양식: Gap.file.xlsx
40	Gene	O	① 유전자 서열만 있는 경우 - 작성하여 업로드할 엑셀파일 양식: Gene.file.xlsx
41	Intron	O	① 인트론 서열만 있는 경우 - 작성하여 업로드할 엑셀파일 양식: Intron.file.xlsx
42	J_segment	O	① 면역글로불린 경쇄 및 중쇄 그리고 T 세포 수용체 알파, 베타 및 감마 사슬의 연결 영역 - 작성하여 업로드할 엑셀파일 양식: Jseg.file.xlsx

No	항목	필수(M)/ 선택(O)	품질관리 기준
43	Mature Peptide	O	① 펩타이드 또는 단백질 코딩 서열 - 작성하여 업로드할 엑셀파일 양식: MatPep.file.xlsx
44	Misc.Binding	O	① 다른 바인딩 항목(primer_bind 또는 protein_bind)로 설명할 수 없는 결합 영역 - 작성하여 업로드할 엑셀파일 양식: MiscBind.file.xlsx
45	Misc.Difference	O	① 다른 서열 항목(variation, 또는 modified base)로 설명할 수 없는 영역 - 작성하여 업로드할 엑셀파일 양식: MiscDiff.file.xlsx
46	Misc.Feature	O	① 다른 항목으로 분류할 수 없는 서열 영역 - 작성하여 업로드할 엑셀파일 양식: MiscFeat.file.xlsx
47	Misc. Recombination	O	① 모든 일반화, 특이적 또는 복제 재조합이 일어나는 부위 - 작성하여 업로드할 엑셀파일 양식: MiscRecomb.file.xlsx
48	Misc.Signal	O	① 다른 신호 항목으로 분류할 수 없는 서열 영역 - 작성하여 업로드할 엑셀파일 양식: MiscSig.file.xlsx
49	Misc.Structure	O	① 다른 항목(stem_loop 및 D-loop)으로 설명할 수 없는 2차 또는 3차 염기서열 구조 - 작성하여 업로드할 엑셀파일 양식: MiscStruc.file.xlsx
50	Modified Base	O	① 표시된 분자로 치환된 변형된 염기서열 - 작성하여 업로드할 엑셀파일 양식: ModBase.file.xlsx
51	N_Region	O	① 재배열된 면역글로불린 세그먼트 사이에 삽입된 추가 염기서열 - 작성하여 업로드할 엑셀파일 양식: Nregion.file.xlsx
52	Operon	O	① 유전자 클러스터를 포함하는 폴리시스트론 전사서열 - 작성하여 업로드할 엑셀파일 양식: Operon.file.xlsx
53	Origin of Transfer(oriT)	O	① 원핵생물 접합시 transfer가 시작되는 염기서열 - 작성하여 업로드할 엑셀파일 양식: OriT.file.xlsx
54	PolyA Site	O	① 전사 후 폴리아데닐화시 아데닌 잔기가 추가되는 RNA 전사체의 염기서열 - 작성하여 업로드할 엑셀파일 양식: PolyAsite.file.xlsx

No	항목	필수(M)/ 선택(O)	품질관리 기준
55	Primer Bind	O	① 복제, 전사 또는 역전사 개시를 위한 프라이머 결합 부위 - 작성하여 업로드할 엑셀파일 양식: PrimBind.file.xlsx
56	Propeptide	O	① 프로펩타이드 코딩 서열 - 작성하여 업로드할 엑셀파일 양식: Propep.file.xlsx
57	Protein Bind	O	① 핵산의 단백질 결합 부위 - 작성하여 업로드할 엑셀파일 양식: ProtBind.file.xlsx
58	Replication Origin	O	① 염기서열 복제를 위한 시작지점 - 작성하여 업로드할 엑셀파일 양식: OriR.file.xlsx
59	S_Region	O	① 면역글로불린 중쇄의 스위치 영역 - 작성하여 업로드할 엑셀파일 양식: Sregion.file.xlsx
60	Signal peptide	O	① 분비신호 펩타이드 코딩 염기서열 - 작성하여 업로드할 엑셀파일 양식: Sigpep.file.xlsx
61	Stem Loop	O	① RNA 또는 DNA의 단일 가닥에 있는 헤어핀 구조의 상보적 서열 - 작성하여 업로드할 엑셀파일 양식: StemLoop.file.xlsx
62	Transit Peptide	O	① 세포 소기관으로 향하는 N-말단 서열이 있는 단백질 코딩 서열 - 작성하여 업로드할 엑셀파일 양식: TranPep.file.xlsx
63	V_Region	O	① 면역글로불린 경쇄 및 중쇄 그리고 T 세포 수용체 알파, 베타 및 감마 사슬의 가변 영역 - 작성하여 업로드할 엑셀파일 양식: Vregion.file.xlsx
64	V_Segment	O	① 면역글로불린 경쇄 및 중쇄 그리고 T 세포 수용체 알파, 베타 및 감마 사슬의 가변 세그먼트 - 작성하여 업로드할 엑셀파일 양식: Vseg.file.xlsx
65	Variation	O	① 근연종 동일한 유전자의 지속적 돌연변이(예: RFLP, 다형성 등)를 포함합니다. - 작성하여 업로드할 엑셀파일 양식: Var.file.xlsx
66	3'UTR	O	① 성숙한 전사체의 3' 말단 영역(stop codon 다음영역) ② 엑셀파일 업로드 - 작성하여 업로드할 엑셀파일 양식: 3UTR.file.xlsx
67	5'UTR	O	① 성숙한 전사체의 5' 말단에 있는 영역(Start codon 앞) - 작성하여 업로드할 엑셀파일 양식: 5UTR.file.xlsx

파트 5-1. File 정보 (Case 1)			
No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Reference file	O	① Accession Number 기입 여부
2	FASTA or SQN file (FASTA 또는 SQN 파일)	M	① FASTA 또는 Sqn 파일 업로드 여부 [비고] Sqn file로 등록 시 앞의 다른 feature들이 이미 들어가 있는 경우 feature를 등록하지 않아도 등록 가능(단 어떤 양식이던 하나 이상의 feature가 반드시 포함되어야 함)
3	Release date (공개날짜)	M	① YYYY-MM-DD 포맷 사용 여부 [예시] 2018-12-22

파트 5-2. File 정보 (Case 2)			
No	항목	필수(M)/ 선택(O)	품질관리 기준
1	Assembly state (어셈블리 상태)	O	① 가능한 값 중 선택 여부
2	Confirm AGP file for unplaced scaffolds (비배치 스캐폴드 AGP 파일 유무)	O	① 가능한 값 중 선택 여부 [비고] “Assembly state” 항목의 값이 3번인 경우에만 해당
3	Confirm AGP file for unlocalized scaffolds (비위치화 스캐폴드 AGP 파일 유무)	O	① 가능한 값 중 선택 여부
4	How to assemble using AGP (AGP 이용 조립방법)	O	① 가능한 값 중 선택 여부 [비고] “Confirm AGP file for unlocalized scaffold” 항목의 값이 Yes인 경우에만 해당
5	Confirm annotation in AGP (AGP 주석 유무)	O	① 가능한 값 중 선택 여부

No	항목	필수(M)/ 선택(O)	품질관리 기준
6	AGP file (AGP 파일)	O	① AGP 파일 업로드 여부 [비고] - 어셈블리만 올리는 경우 Feature 항목을 올리지 않고 파트 5.2에서 fasta 파일만 업로드하며, 유전자 주석을 포함한 유전체 정보를 올리는 경우 Feature를 올리거나 sqn을 통해 파일을 업로드해야 함 - Gap이 포함된 어셈블리의 경우 AGP 파일을 등록해 Gap filing에 대한 정보를 기입할 것
7	Annotation file (주석 파일)	O	① gff 파일 업로드 여부 [비고] 유전자 주석(annotation)이 있는 경우 GFF format에 맞추어 feature 정보를 기입할 수 있으며, 앞의 다른 feature들이 이미 들어가 있는 경우 feature를 등록하지 않아도 등록 가능(단 어떤 양식이던 하나 이상의 feature가 반드시 포함되어야 함)
8	Sequence file format (염기서열 파일 포맷)	M	① 가능한 값 중 선택 여부 [비고] - WGS 등의 단일 염기서열 (single feature)의 경우에는 FASTA 및 Sqn 선택 가능 - WGS with third party annotation 등의 다중 염기서열 (multiple features)의 경우에는 Sqn만 선택 가능
9	Sequence file (염기서열 파일)	M	① 염기서열 파일 업로드 여부
10	Release date (공개날짜)	M	① YYYY-MM-DD 포맷 사용 여부 [예시] 2018-12-22