NABIC
데이터 구조 상세

STAGE 01	메타데이터 등록 — 프로젝트 및 샘플 정보
STAGE 02	오믹스 및 표현체 분야별 특화 데이터 등록
INDEX	분야별 생명정보 분류 및 업로드 파일

01 메타데이터 등록 - 프로젝트 및 샘플 정보

STAGE 01

PROJECT

BioProject

파트 1 **Submitter information (등록자 인적사항)**
등록자 영문·국문 이름 / 이메일(기관 도메인) / 소속 기관·학과의 영문·국문명 / 국가 / 등록자 그룹 / 실무자 정보

파트 2 **General information (프로젝트의 개괄적 정보)**
프로젝트 유형(Primary · Umbrella) / 상위 Umbrella BioProject 유무 · accession ID / 하위 BioProject accession / 사업단 이름 / 공개날짜(YYYY-MM-DD)

파트 3 **Grant information (연구 과제 정보)**
연구과제 NTIS 번호 / 영문·국문 과제명 / 연구비 지원 기관 / 기준 연도 — Primary Project에서만 작성

파트 4 **Publication and patent information (논문 및 특허 성과 정보)**
논문 PubMed ID(없을 시 DOI) / 국내 특허 번호 / 해외 특허 번호 — "//" 구분자로 복수 입력

SAMPLE

BioSample

파트 1 **BioSample package (BioSample 유형)**
Standard package — Human · Model organism or animal · Invertebrate · Plant or fungi · Microbe · Virus · Clinical or host-associated pathogen · MIMS
Environmental/Metagenome package — 환경 · 메타게놈 서열

파트 2 **Sample attribute — Standard package (일반 패키지 샘플 특성)**
선택한 생물군 패키지별 필수(M) · 선택(O) 속성 항목 기재

파트 3 **Sample attribute — MIMS package (MIMS 패키지 샘플 특성)**
MIMS 표준 속성 항목 — 환경 유형 · 채취 정보 등

02 유전체

GENOMICS – A

NGS (차세대 시퀀싱)

파트 1. Experiment (실험 정보)
시퀀싱 기법 — Bulk sequencing (WGS · RNA-seq · ChIP-seq · ATAC-seq 등) · Single-cell sequencing / 라이브러리 ID · 제목 · Library strategy(AMPLICON · ATAC-seq 등) · 기기 모델 · 공개날짜
※ Spatial transcriptomics sequencing은 전사체 분야로 분류

파트 1-1. Single-cell sequencing experiment
단일세포 시퀀싱 선택 시 추가 작성 — 세포 분리 · 바코드 방식 등

파트 2. Run (파일)
염기서열 원시 파일 업로드

파트 3. Analysis (분석 데이터)
분석 파일명 · 분석 방법 · 정렬 참조 서열

GENOMICS – B

염기서열 (Nucleotide sequence)

파트 1–3. Affiliation · File · Reference
등록자 인적사항 / 서열 파일(FASTA) · 주석 파일(gff) / 논문 정보

파트 4. Assembly Status information
어셈블리 상태 · de novo 여부 · 플랫폼 · 게놈 커버리지 · 프로그램명

파트 5–6. Nucleotide · Gaps information
분자 유형(게놈 DNA/RNA · mRNA · tRNA 등) · 위상 · 완성도 / 갭 정보

파트 7–8. MAG mapping · Protocol
빈(Bin) 번호 · 완전도 · 오염도 / PCR 프라이머 등 프로토콜

GENOMICS – C

유전체 변이 (Genomic variation)

파트 1. Study (연구 내용)
변이 유형(SNP · SV) · 연구 설명 · 샘플 집단

파트 2. SampleSet (샘플셋)
그룹 식별자 · 실험군(Case) · 대조군(Control) · 그룹 사이즈

파트 3. Sample (샘플)
상세 샘플명 · 생물 분류 ID

파트 4–5. Experiment · Dataset
실험 단계 · 검출 방법 · 분석 플랫폼 / 데이터셋 · 서열 파일명 연결

02 전사체

TRANSCRIPTOMICS — A

공간전사체 시퀀싱 (Spatial transcriptomics sequencing)

파트 1. Experiment (실험 정보)
시퀀싱 기법 — Spatial transcriptomics sequencing (샘플 내 분자의 공간적 위치 정보를 보존한 시퀀싱) / 라이브러리 ID · 제목 · Library strategy · 기기 모델 · 공개날짜

파트 1-1. Single-cell and spatial transcriptomic sequencing experiment
공간전사체 선택 시 추가 작성 — 조직 절편 · 스팟 좌표 체계 · 바코드 방식 등 세부 메타데이터

파트 2. Run (파일)
염기서열 원시 파일 업로드

파트 3. Analysis (분석 데이터)
분석 파일명 · 분석 방법 · 정렬 참조 서열 · 공간 좌표 및 이미지

TRANSCRIPTOMICS — B

기능유전체 (Functional genomics)

파트 1. Submission type (데이터 등록 형식)
High-throughput sequencing(HTS) · Microarray / 데이터 유형 — Bulk · Single-cell · Spatial transcriptomic sequencing data / Microarray data

파트 2. Study (연구 내용)
제목(120자 이내) · 기여자 · 실험 유형 · 공개날짜

파트 3. Protocols (프로토콜)
성장 · 처리 · 추출 · 라벨링 · 정규화 변환 프로토콜, 분석 장비 · 소프트웨어

파트 4-1 / 4-2. Sample & Experimental detail
HTS용 · Microarray용 샘플 및 실험 상세 정보 (플랫폼 ID · 샘플 설명)

파트 5-1 / 5-2. Result files (결과 파일)
HTS — expression matrix · peak calls 등 분석 데이터 / Microarray — 원시 · 분석 파일

02 단백체 · 대사체 · 표현체

유전체 | 전사체 | 단백체 | 대사체 | 표현체

03 — PROTEOMICS

5개 파트

단백체

파트 1. Descriptors (설명자)
연구 정의 키워드 · 데이터셋 제목 · 제출 상태 · 공개날짜

파트 2. Sample details (샘플 상세정보)
샘플 종류(체액 · 혈액 · 혈청 · 조직 · 세포 등) · 질병 정보

파트 3. Experimental details (실험 상세정보)
분획(SCX · RP · HILIC) / 분해 효소(Trypsin · Lys-C) / 정량(TMT · iTRAQ · SILAC) / 획득 방법(DDA · DIA · SRM) · 장비

파트 4. Data analysis (데이터 분석)
전처리 · 처리 프로토콜 · 적용 스크립트

파트 5. Files (결과 파일)
원시데이터(raw · wiff · d) / 피크 목록(mzML · mzXML) / 검색 결과(mzid · pepXML · protXML)

04 — METABOLOMICS

13개 파트

대사체

파트 1–2. Descriptors · Extraction
연구 키워드 · 실험 인자 · 공개날짜 / 추출 방법

파트 3. Platform (플랫폼)
분석 기기 대분류 — MS-Chromatography · MS-DI · NMR · Others

파트 4–8. 기기 세팅
Chromatography · Mass spectrometry · NMR sample · NMR spectroscopy · NMR assay

파트 9–11. 변환 · 동정 · 품질관리
Data transformation / Metabolite identification and quantification / 데이터 품질관리 정보

파트 12–13. Files · Metabolite data
정량 결과 템플릿(csv) / 대사물질 식별자(PubChem · HMDB · KEGG · CAS) · 주석 출처

05 — PHENOME

NABIC 자체 분야

표현체

파트 1. 등록 정보
오믹스 정보명(한글 · 영문) / 생물정보 대분류·소분류 — RGB · 초분광(HSI) · 적외선(IR) · X-ray · 기타 / 생물종 · 품종 · 자원 번호(IT · ET) / 공개 시점

파트 2. 메타 정보
실험 명칭 · 측정 목적 / 측정 기간 · 주기 / 촬영 부위 / 장비 스펙 · 세부 설정 / 스케일 기준

파트 3. 파일 정보
표현형 정보 · 지표 정보 · 환경 정보 파일 (드래그 앤 드롭 업로드)

분야별 생명정보 분류 및 업로드 파일

바이오 연구데이터 표준등록양식(2025.12.26)의 양식 체계를 준수함

* 필수항목 · 필수 확장자는 **진하게** 표시

대분류	중분류	데이터 유형 구분	파트 구성	메타데이터 *	원시(raw) 데이터 *	분석·결과 데이터 *
유전체	**NGS** (K-BDS KRA)	Bulk sequencing · Single-cell sequencing	Experiment / Single-cell 특화 / Run / Analysis	등록양식 메타데이터 (BioProject·BioSample·Experiment)	**Sequence file (fastq / bam / sff / hdf5 / fast5)** + **File name · MD5 checksum**	**ABUNDANCE_MEASUREMENT** (tsv · csv · txt) · **REFERENCE_ALIGNMENT** (bam · sam · cram) · **SEQUENCE_VARIATION** (bcf · gvcf · bed · wig · bedgraph · maf) 복수 파일은 tar 또는 tar.gz로 압축해 하나로 제출 가능
	염기서열	Genome · Nucleotide · Gene · MAG (PCR 프라이머 통합)	Affiliation / File / Reference / Assembly / Nucleotide / Gaps / MAG / Protocol	등록양식 메타데이터	**Sequence file (FASTA)**	**GFF · GTF · TBI**
	유전체 변이	SNP · SV	Study / SampleSet / Sample / Experiment / Dataset	등록양식 메타데이터	**vcf · vcf.gz** 변이 파일 — SNP은 VCF 필수, SV는 VCF 또는 xlsx	–
전사체	**공간전사체 시퀀싱** (NGS 양식)	Spatial transcriptomics sequencing	Experiment / Spatial 특화 / Run / Analysis	등록양식 메타데이터 (파트1, 파트1-1 둘 다 작성)	**Sequence file (fastq / bam / sff / hdf5 / fast5 / spatial)** + **File name · MD5 checksum** spatial — 공간 바코드 · 현미경 이미지 파일이 연계된 시퀀싱 데이터 파일	**ABUNDANCE_MEASUREMENT** (tsv · csv · txt) · **REFERENCE_ALIGNMENT** (bam · sam · cram) · **SEQUENCE_VARIATION** (bcf · gvcf · bed · wig · bedgraph · maf)
	기능유전체 — HTS (K-BDS KEA)	HTS (Bulk · Single-cell · Spatial data)	Submission type / Study / Protocols / Sample & Experimental detail / Result files	등록양식 메타데이터	**Cell barcode matrix (barcodes.tsv) · Feature information matrix (features.tsv) · Spatial information (tissue_hires_image.png · tissue_positions.parquet)** Single-cell·Spatial sequencing 선택 시 필수 · Spatial information은 세미콜론(;)으로 복수 입력	**Count matrix – Raw counts (csv · txt)** · Browser extensible data (bed) · narrowPeak · broadPeak · BedGraph (bedgraph) · WIG (wig) · BIGWIG (bw) 선택: Count matrix – Normalize (TPM·FPKM·CPM) · Alignment files (bam·sam) · Image (jpg·tiff)
	기능유전체 — Microarray (K-BDS KEA)	Microarray	Submission type / Study / Protocols / Sample & Experimental detail / Result files	등록양식 메타데이터	**Raw data file** — 벤더별 형식 Affymetrix: **CEL** / Agilent: **Agilent FE 결과 파일** (Agilent Probe Names 포함 시 GPR·탭 구분 텍스트 가능) / Nimblegen: **PAIR** / Illumina: **IDAT** 및 연관 파일 또는 비정규화 Matrix worksheet · 세미콜론(;)으로 구분해 복수 입력	**1차 분석 파일** (발현량 테이블 파일 · 후성 유전체 분석 파일 · 변이 분석 파일 중 해당)
단백체	**단백체 데이터**	DDA · DIA · SRM	Descriptors / Sample details / Experimental details / Data analysis / Files	등록양식 메타데이터	**Raw data (기기 출력: Thermo RAW · Sciex WIFF · Bruker .d)**	**Search result file (mzIdentML · Mascot .dat 등)**
대사체	**대사체 데이터**	MS-Chromatography · MS-DI · NMR · Others	Descriptors / Extraction / Platform / 기기 세팅(4–8) / 변환·동정·QC(9–11) / Files·Metabolite data(12–13)	등록양식 메타데이터	**Acquired raw data (MS RAW · NMR FID, 폴더형은 zip 압축)**	–
표현체	**표현체 데이터** (NABIC)	RGB · 초분광(HSI) · 적외선(IR) · X-ray · 기타	등록 정보 / 메타 정보 / 파일 정보	등록양식 메타데이터	**jpg · png · tiff · GeoTIFF** / ply · obj · las / mp4 · avi / hdr + raw · dat	**표현형 · 지표 · 환경 정보 파일** (xlsx·csv·tsv)