

○川添将仁、青野英雄、Andrea Ghelfi、李慶範、飯塚朋代、大城戸利久、児玉悠一、小菅武英、杉田里江、鈴木紀美子、高木佳苗、時松敏明、秦千比呂、福田亜沙美、真島淳、横山会美、藤澤貴智

情報・システム研究機構 国立遺伝学研究所生命情報・DDBJ センター

要旨

DDBJセンターは、米国のNCBI、欧州のEBIと共同運営する国際塩基配列データベースINSDCの構成メンバーであり、国立遺伝学研究所スーパーコンピュータを基盤として国内外の生命科学研究から産出されるデータの収集・共有を行うと共に、解析プラットフォームを提供している。DDBJセンターでは、アノテーションを付与/アセンブルした塩基配列データ:DDBJ、次世代シーケンスデータ:DRA、機能ゲノミクスデータ:GEA、アクセス制限のある個人ゲノムデータ:JGA、研究プロジェクト・サンプルのデータ:BioProject・BioSampleの各データベースを運営しており、2020年10月よりメタボロームレポジトリ:MetaboBankの運用を開始した。また、2021年3月に、3PBのストレージをユーザホーム領域に追加し、解析環境の拡充を行った。今後はヒトバリエーションデータベース:JVarの新設を予定しており、時代と共に多様化、大容量化する生命科学研究データに対応したデータベースの拡充および解析環境を構築することで、生命科学の情報基盤の整備および発展に寄与している。



大学共同利用機関法人情報・システム研究機構国立遺伝学研究所生命情報・DDBJセンター
〒411-8540 静岡県三島市谷田1111 問い合わせ：<https://www.ddbj.nig.ac.jp/contact.html>

オープンアクセスデータベース

塩基配列

DDBJ *1

アノテーション付・アセンブル済塩基配列

DRA *2

DDBJ Sequence Read Archive
NGS 生データとアライメント情報

GEA *3

Genomic Expression Archive
遺伝子発現・エピジェネティクス解析・
マイクロアレイなど

JVar (準備中) *4

Japan Variation Database
バリエーション: SNP, INDEL, CNV, 構造変異等

メタデータ

BioProject

研究プロジェクト情報

BioSample

サンプル情報

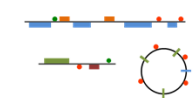
MetaboBank

メタボロームデータレポジトリ
(2020 年 10 月サービス開始)

収集・公開しているデータベース例 (一部抜粋)

DDBJ *1

アノテーション



アセンブリ



アセンブリ・アノテーション

GenBank
ENA

DRA *2

アライメント



リード

Quality value

次世代シーケンス生データとアライメント情報

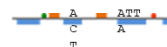
GenBank
ENA

GEA *3

機能ゲノミクス
データタグカウント
ピークコール遺伝子発現・エピジェネティクス解析・
マイクロアレイなどGEO
ArrayExpress

JVar *4

短いバリエーション



構造変異



SNP, INDEL, CNV, 構造変異等のバリエーション

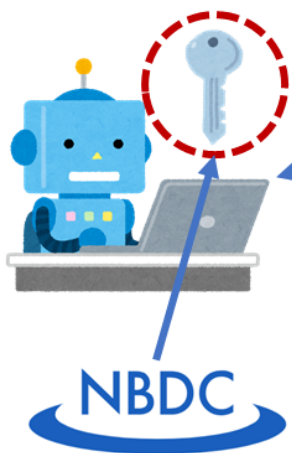
dbSNP/dbVar
EVA/DGVa

研究プロジェクトと用いられた生物試料の情報を BioProject、BioSample に格納し、得られたデータにその ID を引用することで、プロジェクト、試料、データを総合的に管理。



アクセス制限データベース

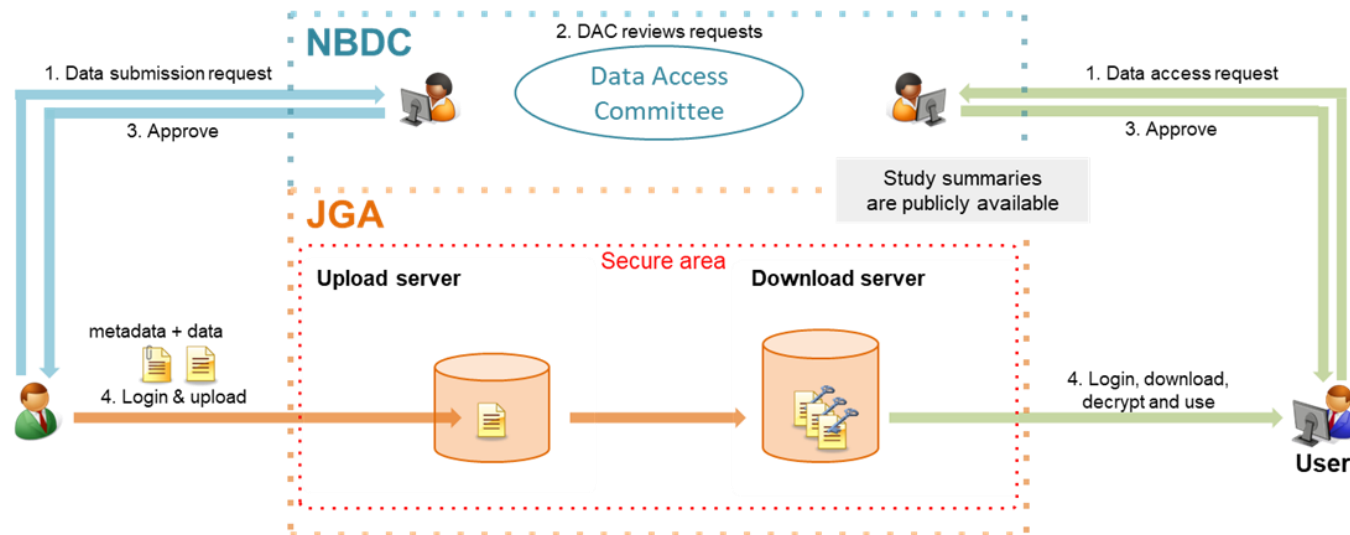
制限ユーザによる登録・利用



JGA

Japanese Genotype-phenotype Archive
匿名化した個人レベルの遺伝型と表現型ヒトデータ審査委員会による
登録・利用の審査

登録・利用のフロー



登録と利用でそれぞれ個別の審査が必要となる。



NIGスーパーコンピュータ

一般解析区画

- 国内の研究者は無料(*)で利用可能(* 一般解析区画のみ)
- BioContainers プロジェクト が公開している Singularity コンテナイメージをスパコン内で提供
- 様々な DB のデータに直接アクセス可能

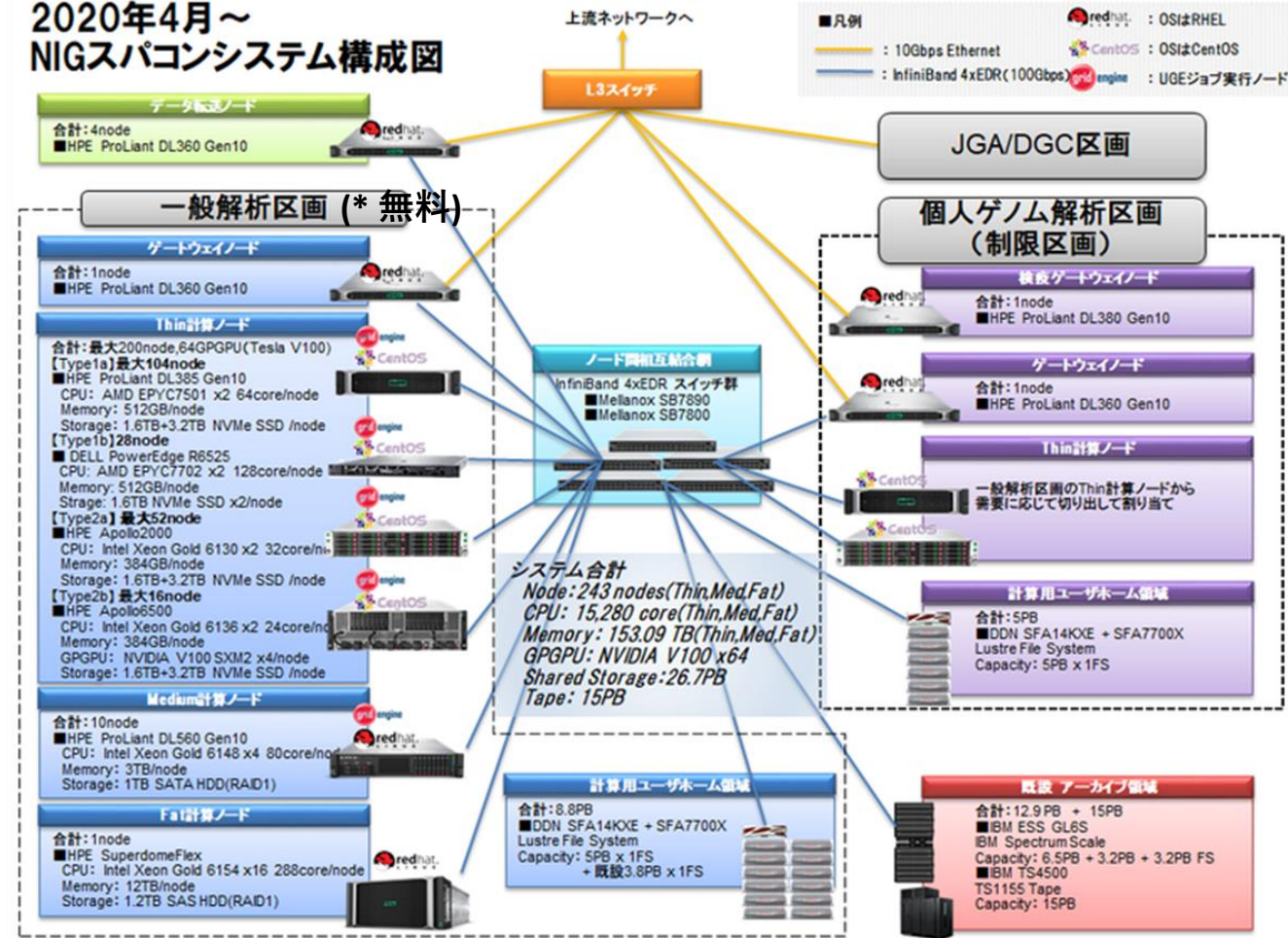
個人ゲノム解析区画

- アクセス制限区画
- JGA データ等個人ゲノムデータの解析

利用可能 DB 名

ChEMBL
DDBJ-unified-all
DDBJ-unified-new
EMBL
EMBL-daily
GenBank
GenBank-daily
illumina iGenomes
NCBI-dbEST
NCBI-dbGSS
NCBI-HTGS
NCBI-nr
NCBI-nt
NCBI-patnt
NCBI-STs
NCBI-taxonomy
NCBI-v5
PDB
Pfam
PubChem
RefSeq-daily
RefSeq-Genomic
UniProt

2020年4月～ NIGスパコンシステム構成図



2021年3月に3PBのストレージをユーザホーム領域に追加済