

Biological Databases Quick Reference

A practical map of the major databases used to find sequences, expression studies, proteins and genomic datasets.

Core biological databases

Resource	Best for	Typical identifier
NCBI	Entry point to many biological databases	Gene ID, BioProject, accession
GenBank	Nucleotide sequence records	NM_, NC_, accession
SRA	Raw sequencing reads	SRR / SRX / SRS
GEO	Functional genomics and expression studies	GSE / GSM
Ensembl	Genes, transcripts, variants, genomes	ENSG / ENST
UniProt	Protein sequences and functional annotation	UniProt accession
PDB	Experimentally determined 3D structures	4-character PDB ID
ENCODE	Functional genomics and regulatory datasets	ENCSR / ENCFF

Which database should I start with?

Question	Start here
I need a gene or transcript sequence	NCBI / GenBank / Ensembl
I need raw FASTQ sequencing data	SRA
I need an expression study	GEO
I need protein function and sequence	UniProt
I need a protein structure	PDB
I need regulatory genomics data	ENCODE