

# Representation of sequence – sequence file format

1.) FASTA – simple and clean

> gene\_name, (other info)

MASASASKJHKLJLKJLDSDFSF

SSDSASFSD...

Practice / DIY: retrieve sequence in Fasta format and save the file in the local computer.

# Public Resources for Bioinformatics

- ❖ Databases - how to find relevant information.
- ❖ Analysis Tools

**Observe: List of databases and service at NCBI, EBI, KEGG, and Ensembl.**

# Pet Project: Identify an ortholog of..

**IL6, or your favorite gene**

**Make a folder for the project in /GMS6014/**

- **Generate subfolders, e.g. “info”, “seqs”.**
  - **/seqs/ for raw sequence files in fasta format.**
  - **/info/ for saving webpages of curated entry.**

# Observe/Practice

Search for IL6 (or your favorite gene) in the Gene database and the Proteins databases.

- why do we get so many hits?

Search for IL6 in the default “All Text” vs. search in the [Gene Name] field only in the Gene database.

Compare results.

# Database concept – tables in relational databases

“IL6”=IL6[All Fields]

IL6[Name]

Accession	Organ.	Ref.	Name	Key words	Features	
....	.....	medline1	TNF	.....	.....	.....
....	....	medline2	P53	....	.....	.....

**Gene table**

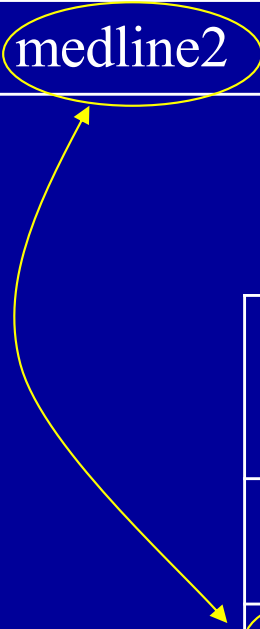
# Database concept – relationship between tables allows linkage

Accession	Organ.	Ref.	Name	Key words	Features	
....	....	medline1	P27	....	.....	....
....	....	medline2	P53	....	.....	....

**Protein table**

ID	title	year	author	abstract	
medline1	....	1970	....	.....	....
medline2	....	1980	....	.....	....

**Reference table**



# Observe/Practice

Observe the links from the IL6 **Gene** page:

- RefSeq
- OMIM
- SNP
- GEO
- Etc.

Gene Entry is the pivotal point for many NCBI resources.

# Representation of genes and related information

The need to represent associated info with sequence

- Different aspects of the gene (such as protein, nucleotide, structure (PDB), OMIM etc.)
- Specialized databases (such GEO, SNP)
- Complex / customized data structure
  - Object-oriented data representation



# Observe

Observe entries involving IL6 (or your gene)  
in Reactome.

# Pet Project: Identify an ortholog of..

**IL6, or your favorite gene**

**What can we know about this gene?**

- **Search for “curated” databases.**
- **To prepare for future analysis, save annotated sequence files as `genename.html` (in a target folder).**
- **For downstream sequence analysis, save pure sequence as FASTA format file.**

# Where and how much information are available for my gene?

Observe: The information contents and presentation format for the same gene in NCBI Genes, NCBI protein, and UniProt, etc..