



## Example Vignette Questions:

- What chromosome location and what genes are in the vicinity of a given query gene? **NCBI GENE**
- What can you find out about molecular functions, biological processes, and prominent cellular locations? **EBI GO**
- What amino acid positions in the protein are responsible for ligand binding? **EBI UniProt**
- What variants of this gene are associated with gastric cancer and other human diseases? **NCBI OMIM**
- Are high resolution protein structures available to examine the details of these mutations? How might we explain their potential molecular effects? **RCSB PDB**
- What is known about the protein family, its species distribution, number in humans and residue-wise conservation? **EBI PFAM**

## Recap From Last Time:

- Bioinformatics is computer aided biology.
  - Deals with the collection, archiving, organization, and interpretation of a wide range of biological data.
- There are a large number of primary, secondary and tertiary bioinformatics databases (see [handout!](#)).
- The NCBI and EBI are major online bioinformatics service providers.
- Introduced Gene, UniProt and PDB databases as well as a number of 'boutique' databases including PFAM and OMIM.
- Introduced the notion of *controlled vocabularies* and *ontologies* via exploring GO annotations.
- Also covered: Course structure; Introductions, Software setup and Database Vignette...

## TODAYS MENU

- More hands-on exploration of these databases and their associated tools (searching with a propose!)
- Major hands-on sections include:
  1. BLAST, GenBank and OMIM @ **NCBI** [~35 mins]
  2. GENE database @ **NCBI** [~15 mins]  
— BREAK —
  3. UniProt & Muscle @ **EBI** [~25 mins]
  4. PFAM, PDB & NGL [~30 mins]  
— BREAK —
  5. Optional extension exercises [~20 mins]

- Please do answer the last review question (**Q19**).
- We encourage discussion and exploration!

## SideNote: Bioinformatics Databases

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolmage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTKbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFC, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty\_cDB, DIP, DOGS, DOMO, DPD, DPInteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDB, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genllesne, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, Klobo, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5, Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSSub, O-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene, Sub2D, SubtiList, SWISS-2DPAGE, SWISS-3DIMAGE, SWISS- MODEL Repository, SWISS-PROT, TelDB, TGN, tmRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT, WormPep, etc ..... !!!!

## SideNote: Bioinformatics Databases

AATDB, AceDb, ACUTS, ADB, AFDB, AGIS, AMSdb, ARR, AsDb, BBDB, BCGD, Beanref, Biolmage, BioMagResBank, BIOMDB, BLOCKS, BovGBASE, BOVMAP, BSORF, BTKbase, CANSITE, CarbBank, CARBHYD, CATH, CAZY, CCDC, CD4OLbase, CGAP, ChickGBASE, Colibri, COPE, CottonDB, CSNDB, CUTG, CyanoBase, dbCFC, dbEST, dbSTS, DDBJ, DGP, DictyDb, Picty\_cDB, DIP, DOGS, DOMO, DPD, DPInteract, ECDC, ECGC, EC02DBASE, EcoCyc, EcoGene, EMBL, EMD db, ENZYME, EPD, EpoDB, ESTHER, FlyBase, FlyView, GCRDB, GDB, GENATLAS, Genbank, GeneCards, Genllesne, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, Klobo, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5, Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSSub, O-lycBase, OMIA, OMIM, OPD, ORDB, OWL, PAHdb, PatBase, PDB, PDD, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PPDB, PRESAGE, PRINTS, ProDom, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, SCOP, SeqAnaiRef, SGD, SGP, SheepMap, Soybase, SPAD, SRNA db, SRPDB, STACK, StyGene, Sub2D, SubtiList, SWISS-2DPAGE, SWISS-3DIMAGE, SWISS- MODEL Repository, SWISS-PROT, TelDB, TGN, tmRDB, TOPS, TRANSFAC, TRR, UniGene, URNADB, V BASE, VDRR, VectorDB, WDCM, WIT, WormPep, etc ..... !!!

## Side-note: Databases come in all shapes and sizes



Databases can be of variable quality and often there are multiple databases with overlapping content.

## Finding Bioinformatics Databases

The screenshot shows a web browser window with the URL [www.oxfordjournals.org/nar/database/cat/8](http://www.oxfordjournals.org/nar/database/cat/8). The page is titled "OXFORD JOURNALS". The main content area is titled "NAR Database Summary Paper Category List". It lists several categories of databases, including Nucleotide Sequence Databases, RNA sequence databases, Protein sequence databases, Structure Databases, Genomics Databases (non-vertebrate), Metabolic and Signaling Pathways, Human and other Vertebrate Genomes, Human Genes and Diseases, CancerResource, Protein Mutant Database, General human genetics databases, General polymorphism databases, Cancer gene databases, Gene-, system- or disease-specific databases, Microarray Data and other Gene Expression Databases, Proteomics Resources, Other Molecular Biology Databases, Organelle databases, Plant databases, Immunological databases, and Cell biology. On the right side, there is a sidebar with links: "Compilation Paper", "Category List", "Alphabetical List", "Category/Paper List", and "Search Summary Papers". At the bottom right, there is a link to the URL <http://www.oxfordjournals.org/nar/database/cat/8>.

# GENBANK & REFSEQ: NCBI'S NUCLEOTIDE SEQUENCE DATABASES

## GenBank sequence record

The screenshot shows a detailed view of a GenBank sequence record for Homo sapiens kinesin family member 5A (KIF5A), mRNA. The page includes a search bar, navigation links, and various analysis tools like BLAST and primer pickers. Key data points include:

- LOCUS**: NM\_004984
- DEFINITION**: Homo sapiens kinesin family member 5A (KIF5A), mRNA
- ACCESSION**: NM\_004984
- VERSION**: NM\_004984.2
- KEYWORDS**: RefSeq, Homo sapiens (human)
- ORGANISM**: Homo sapiens
- REFERENCE**: Kawaguchi, K. et al. (2013) *J. Biol. Chem.* 288: 21742-21754.
- TITLE**: Role of kinesin-1 in the pathogenesis of SPC10, a rare form of hereditary spastic paraparesis.
- JOURNAL**: Neuropediatrics 2012; 43(2): 111-116.
- PURMED**: 227
- REMARK**: This same general format is used for other sequence database records too.
- REFERENCE**: Prokopenko, D. et al. (2013) *J. Biol. Chem.* 288: 21742-21754.
- Peptide hormone metabolism**
- MHC class II antigen presentation**

A callout box highlights the **ACCESSION** number (NM\_004984) and notes that the GenBank flat file format defines unique identifiers such as the ACCESSION number.

## What is GenBank?

- GenBank is NCBI's primary nucleotide only sequence database
  - Archival in nature - reflects the state of knowledge at time of submission
  - Subjective - reflects the submitter point of view
  - Redundant - can have many copies of the same nucleotide sequence
  - GenBank is actually three collaborating international databases from Europe, US and Japan



## Side node: Database accession numbers

Database **accession numbers** are strings of letters and numbers used as **identifying labels** for sequences and other data within databases

- Examples (all for retinol-binding protein, RBP4):

X02775	GenBank genomic DNA sequence	DNA
NT_030059	Genomic contig	
N91759.1	An expressed sequence tag (1 of 170)	
NM_006744	RefSeq DNA sequence (from a transcript)	RNA
NP_007635	RefSeq protein	
AAC02945	GenBank protein	
Q28369	UniProtKB/SwissProt protein	
1KT7	Protein Data Bank structure record	
PMID: 12205585	PubMed IDs identify articles at NCBI/NIH	Literature

## GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI

www.ncbi.nlm.nih.gov/nucleotide/NM\_004984.2

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI

NCBI Resources How To Sign In to NCBI

Nucleotide (KIF5A) AND "Homo sapiens" Search

Limits Advanced

Display Settings: GenBank

Send: Change region shown

Customize view

**Homo sapiens kinesin family member 5A (KIF5A), mRNA**

NCBI Reference Sequence: NM\_004984.2

FASTA Graphics

Go to:

LOCUS NM\_004984 3897 bp mRNA linear PRI 10-JAN-2014

DEFINITION Homo sapiens kinesin family member 5A (KIF5A), mRNA.

ACCESSION NM\_004984

VERSION NM\_004984.2 GI:45446748

KEYWORDS RefSeq, Human, Homo sapiens

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euchortoglires; Primates; Homoplorthini; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3897)

AUTHORS Kawaguchi, K.

TITLE Role of kinesin-1 in the pathogenesis of SPC10, a rare form of hereditary spastic paraparesis

JOURNAL Neuroscientist 19 (4), 336-344 (2013)

PUBMED 22785106

REMARK GenRef: A review of the mechanism of pathogenesis involved in spastic paraparesia type 10 when KIF5A is inactivated by mutations. Review article

REFERENCE 2 (bases 1 to 3897)

AUTHORS Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R., Bohm, K.J. and Winner, B.

TITLE alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay

JOURNAL J. Biol. Chem. 288 (30), 21742-21754 (2013)

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the KIF5A gene

*$\alpha$ -Synuclein oligomers impair neuronal microtubule-kinesin interplay* [J. Biol. Chem. 2013]

Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012]

Systems-wide analysis of ubiquitination dynamics reveals a key role in [Nat. Cell Biol. 2012]

Pathways for the KIF5A gene

Peptide hormone metabolism

MHC class II antigen presentation

## GenBank sequence record

Home sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI  
www.ncbi.nlm.nih.gov/nucleotide/NM\_004984.2

Home sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide : (KIF5A) AND "Homo sapiens" Search

Limits Advanced

Display Settings: GenBank

Send: Change region shown

Display Settings: GenBank

Can set different display formats here

FASTA Graphics

LOCUS NM\_004984 3897 bp mRNA linear PRI 10-JAN-2014

DEFINITION Homo sapiens kinesin family member 5A (KIF5A), mRNA.

ACCESSION NM\_004984

VERSION NM\_004984.2 GI:45446748

KEYWORDS RefSeq, Homo sapiens (human)

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarhontognathes; Primates; Haplorrhini; Catarrhini; Hominoidea; Homo.

REFERENCE 1 (bases 1 to 3897)

AUTHORS Kawaguchi, K.

TITLE Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraparesis

JOURNAL Neuroscientist 19 (4), 336-344 (2013)

PUBMED 22785106

REMARK Gene review: A review of the mechanism of pathogenesis involved in spastic paraparesis type 10 when KIF5A is inactivated by mutations. Review article.

REFERENCE 2 (bases 1 to 3897)

AUTHORS Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R., Boehm, K.J. and Winnier, B.

TITLE alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay

JOURNAL J. Biol. Chem. 288 (30), 21742-21754 (2013)

Analyze this sequence

Run BLAST

Pick Primers

Highlight Sequence Features

Find in this Sequence

Articles about the KIF5A gene

alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay [J. Biol. Chem. 2013]

Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012]

Systems-wide analysis of ubiquitination dynamics reveals a key [Nat Cell Biol. 2012]

See all...

Pathways for the KIF5A gene

Peptide hormone metabolism

MHC class II antigen presentation

# FASTA sequence record

# GenBank ‘graphics’ sequence record

The screenshot shows the NCBI Nucleotide search results for NM\_004984.2. The top navigation bar includes links for NCBI Resources, How To, and Sign in to NCBI. The main content area displays the genomic structure of KIF5A, showing exons as black boxes and introns as white spaces. Annotations include protein domains (KIF5c, KIF5c\_KIF5c, Globular), post-translational modifications (RTP binding site, acetylation, phosphorylation), and interactions (microtubule interaction). The right sidebar provides options to Analyze this sequence, Run BLAST, Pick Primers, and Highlight Sequence Features. It also lists articles about the KIF5A gene and pathways it is involved in, such as Peptide hormone metabolism, MHC class II antigen presentation, and Dopaminergic synapse. A reference sequence information section is also present.

## GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

www.ncbi.nlm.nih.gov/nucleotide/NM\_004984.2

Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide - NCBI

NCBI Resources How To Sign in to NCBI

Nucleotide Nucleotide (KIF5A) AND "Homo sapiens" Search Limits Advanced

Display Settings: GenBank Send: Change region shown

Help

## Homo sapiens kinesin family member 5A (KIF5A), mRNA

NCBI Reference Sequence: NM\_004984.2

FASTA Graphics

Go to:

LOCUS NM\_004984 3897 bp mRNA linear PRI 10-JAN-2014  
DEFINITION Homo sapiens kinesin family member 5A (KIF5A), mRNA.  
ACCESSION NM\_004984  
VERSION NM\_004984.2 GI:45446748  
KEYWORDS RefSeq  
SOURCE Homo sapiens (Human)  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Butheria; Euchontoglires; Primates; Haplorrhini; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 3897)  
AUTHORS Kawaguchi, K.  
TITLE Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraparesis  
JOURNAL Neuropathol Appl Neurobiol 19 (4), 336-344 (2013)  
PUBLISHED 22/05/16  
REMARK SevenTRP: A review of the mechanism of pathogenesis involved in spastic paraparesia type 10 when KIF5A is inactivated by mutations. Review article  
REFERENCE 2 (bases 1 to 3897)  
AUTHORS Protz, I., Veber, V., Breyer, S., Campioni, S., Budner, K., Riek, R., Bohm, K.J. and Winer, B.  
TITLE alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay  
JOURNAL J. Biol. Chem. 288 (30), 21742-21754 (2013)  
  
Send: Change region shown  
Customize view  
  
Analyze this sequence  
Run BLAST  
Pick Primers  
Highlight Sequence Features  
Find in this Sequence  
  
Articles about the KIF5A gene  
alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay [J. Biol. Chem. 2013]  
Molecular motor KIF5A is essential for GABA(A) receptor transport, [Neuron. 2012]  
Systems-wide analysis of ubiquitylation dynamics reveals a key [Nat Cell Biol. 2012]  
See all...  
  
Pathways for the KIF5A gene  
Peptide hormone metabolism  
MHC class II antigen presentation

## GenBank sequence record, cont.

## GenBank sequence record, cont.

The **FEATURES** section contains annotations including a conceptual translation of the nucleotide sequence.

**Recent activity**

- Turn Off Clear
- Homo sapiens kinesin family member 5A (KIF5A), mRNA Nucleotide
- (kinesin) AND "Homo sapiens"[orgn] Nucleotide
- kinesin (37064) Nucleotide

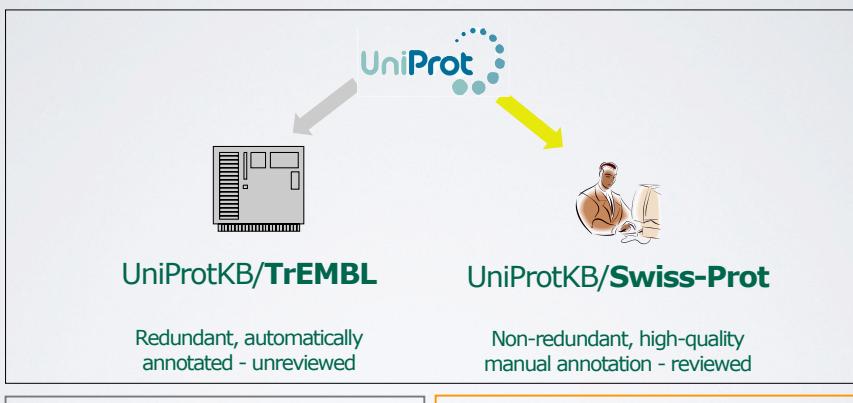
See more...

# RefSeq: NCBI's Derivative Sequence Database

- RefSeq entries are hand curated best representation of a transcript or protein (in their judgement)
  - Non-redundant for a given species although alternate transcript forms will be included if there is good evidence
    - Experimentally verified transcripts and proteins accession numbers begin with “NM\_” or “NP\_”
    - Model transcripts and proteins based on bioinformatics predictions with little experimental support accession numbers begin with “XM\_” or “XP\_”
    - RefSeq also contains contigs and chromosome records

# **UNIPROT:** THE PREMIER PROTEIN SEQUENCE DATABASE

# The two sides of UniProtKB



Indicators of which part of UniProt an entry belongs to include the color of the stars and the ID

# UniProt: Protein sequence database

UniProt is a comprehensive, high-quality resource of protein sequence and functional information

- UniProt comprises four databases:

- ## 1. UniProtKB (Knowledgebase)

Containing **Swiss-Prot** and **TrEMBL** components  
(these correspond to hand curated and automatically annotated entries respectively)

- ## 2. UniRef (Reference Clusters)

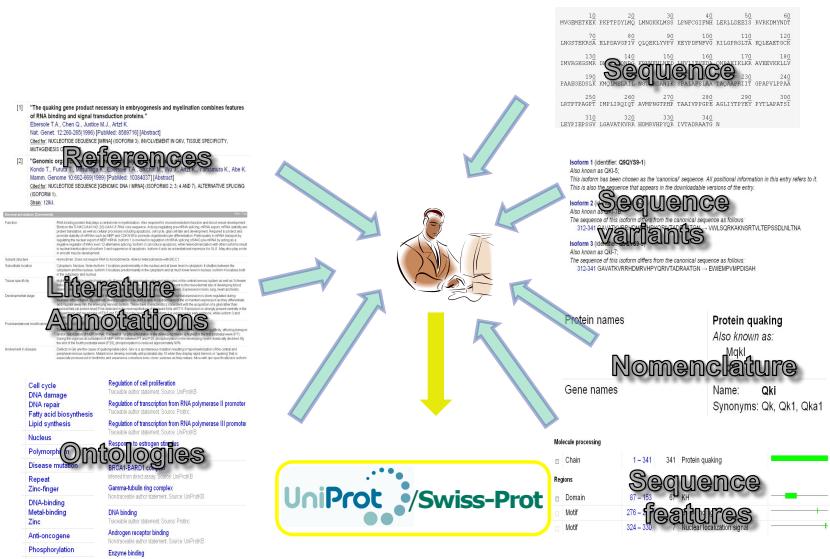
## Filtered version of UniProtKB at various levels of sequence identity

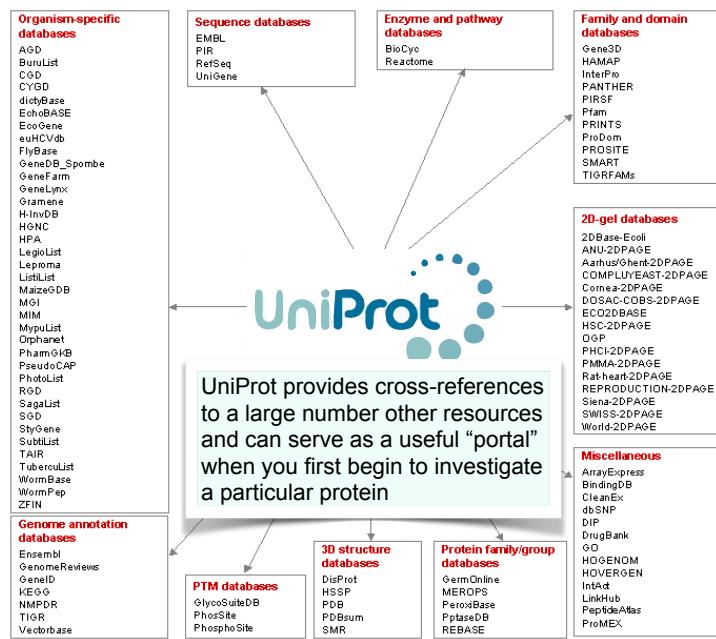
e.g. UniRef90 contains sequences with a maximum of 90% sequence identity to each other

- 3. UniParc** (Archive) with database cross-references to source.

- #### **4. UniMES (Metagenomic and Environmental Sequences)**

# The main information added to a UniProt/Swiss-Prot entry



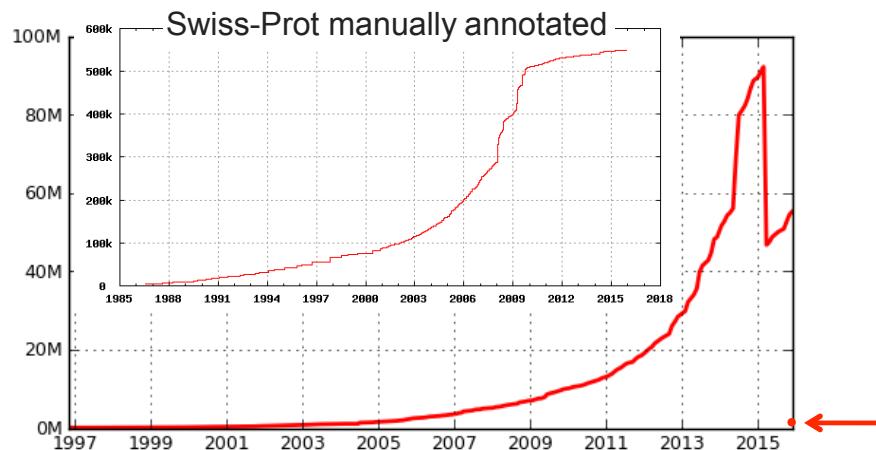


25

## UniProt/Swiss-Prot vs UniProt/TrEMBL

- UniProtKB/Swiss-Prot is a **non-redundant** database with one entry per protein
- UniProtKB/TrEMBL is a **redundant** database with one entry per translated ENA entry (ENA is the EBI's equivalent of GenBank)
  - Therefore TrEMBL can contain multiple entries for the same protein
  - Multiple UniProtKB/TrEMBL entries for the same protein can arise due to:
    - Erroneous gene model predictions
    - Sequence errors (Frame shifts)
    - Polymorphisms
    - Alternative start sites
    - Isoforms
    - OR because the same sequence was submitted by different people

## Side note: Automatic Annotation (a.k.a. sharing the wealth)



27

## Your Turn!

[https://bioboot.github.io/bggm213\\_f17/lectures/#2](https://bioboot.github.io/bggm213_f17/lectures/#2)

The screenshot shows a web browser displaying the BGGN 213 course page. The page features the UC San Diego logo and the text "BGGN 213". It includes a brief description of the course: "A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the Division of Biological Sciences, UCSD". Below this, there are navigation links for "Overview", "Lectures" (which is highlighted in red), "Computer Setup", "Learning Goals", and "Assignments & Grading". To the right of the page, there is a list of course material and a "Homework" section.

- Be able to describe how nucleotide and protein sequence and structure data are represented (FASTA, FASTQ, GenBank, UniProt, PDB).
- Be familiar with online tools at the EBI and NCBI including Muscle and BLAST.
- The goals of the hands-on session is to introduce a range of core bioinformatics databases and associated online services whilst actively investigating the molecular basis of several common human disease.

**Material:**

- Lecture Slides: Large PDF, Small PDF,
- Handout: Major Bioinformatics Databases
- Hands-on section worksheet
- Muddy point assessment

**Homework:**

## BGGN-213: FOUNDATIONS OF BIOINFORMATICS (Lecture 2)

### Bioinformatics Databases and Key Online Resources

[https://bioboot.github.io/bggm213\\_f17/lectures/#2](https://bioboot.github.io/bggm213_f17/lectures/#2)

Dr. Barry Grant  
Oct 2017

**Overview:** The purpose of this lab session is to introduce a range of bioinformatics databases and associated services available on the Web whilst investigating the molecular basis of a common human disease.

Sections 1 and 2 deal with querying and searching GenBank, GENE and OMIM databases at NCBI. Sections 3 and 4 provide exposure to EBI resources for comparing proteins and visualizing protein structures. Finally, section 5 provides an opportunity to explore these and other databases further with additional examples.

**Side-note:** The Web is a dynamic environment, where information is constantly added and removed. Servers "go down", links change without warning, etc. This can lead to "broken" links and results not being returned from services. Don't give up - give it a second go and try a search engine using terms related to the page you are trying to access.

#### Section 1

The following transcript was found to be abundant in a human patient's blood sample.

```
>example1
ATGGTGTGAACTGACCTCCCTGGAGAAGTCTGCCGTACTGCCCTGGGGCAAGGTGAACGTGGATGAA
TTGGTGTGAAGCCCTGGCAGGGCTGCTTGACCCCAGGGTCTTGGTGTGAAAGTGTCTTGGTGTG
GGATCTGCACCTGGATGAGTTATGGGCAAACTTAAGGTGAAGGCTCATGGCAAAAGTGCTCGGT
GCCCTTATGATGATGCCCTGGCTCACCTGACAACCTAACGGGCACTTGGTCCACACTGAGTGAGCTGCACT
GTGACAACTGTCACCTGATCTGAGACTTCAAGGCTCTGGCCACACTGCTGCTGCTGCTGCTGCTG
TCACTTTGGCAAAAGATTACCCACCAAGTGCAAGGCTGGCTATCAGAAAATGGTGGCTGGTGTGGCTAAT
GCCCTGGCCCAAGACTACATAAGCTGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTAAT
```

The only information you are given is the above sequence so you must begin your investigation with a sequence search - for this example we will use NCBI's **BLAST** service at: <http://blast.ncbi.nlm.nih.gov/>

Note that there are several different "basic BLAST" programs available at NCBI (including nucleotide BLAST, protein BLAST, and BLASTx).

## YOUR TURN!

- There are five major hands-on sections including:

### 1. BLAST, GenBank and OMIM @ NCBI

[~35 mins]

### 2. GENE database @ NCBI

[~15 mins]

— BREAK —

### 3. UniProt & Muscle @ EBI

[~25 mins]

### 4. PFAM, PDB & NGL

[~30 mins]

— BREAK —

### 5. Extension exercises

[~30 mins]

- Please do answer the last review question (**Q19**).
- We encourage discussion and exploration!

## YOUR TURN!

- There are five major hands-on sections including:

### 1. BLAST, GenBank and OMIM @ NCBI

### 2. GENE database @ NCBI

— BREAK —

### 3. UniProt & Muscle @ EBI

### 4. PFAM, PDB & NGL

— BREAK —

### 5. Extension exercises

End times:

[ 9:45 am ]

[ 10:00 am ]

— 10:10 am —

[ 10:35 am ]

[ 11:05 am ]

— 11:15 am —

[ 11:45 am ]

- Please do answer the last review question (**Q19**).
- We encourage discussion and exploration!

## HOMEWORK

[https://bioboot.github.io/bggm213\\_f17/lectures/#2](https://bioboot.github.io/bggm213_f17/lectures/#2)

Please do answer the last review question from today (**Q19**)

Complete the **lecture 1 homework questions** for Thur.

Check out the "**Background Reading**" material online.

