

BGGN 213

Foundations of Bioinformatics

Barry Grant
UC San Diego

<http://thegrantlab.org/bggn213>

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...

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**

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, Genlilesne, GenLink, GENOTK, GenProtEC, GIFTS, GPCRDB, GRAP, GRBase, gRNAsdb, GRR, GSDB, HAEMB, HAMSTERS, HEART-2DPAGE, HEXAdb, HGMD, HIDB, HIDC, HIVdb, HotMolecBase, HOVERGEN, HPDB, HSC-2DPAGE, ICN, ICTVDB, IL2RGbase, IMGT, Kabat, KDNA, KEGG, KloTho, LGIC, MAD, MaizeDb, MDB, Medline, Mendel, MEROPS, MGDB, MGI, MHCPEP5, Micado, MitoDat, MITOMAP, MJDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSub, 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, TeIDB, 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, BCGB, Beanref, CANSITE, CarbBank, CARBHYD, CATH, CAZy, ChickGBASE, Colibri, COPE, CottonDB, dbSTS, DDBJ, DGP, DictyDb, ECGC, EC02DBASE, FlyBase, GDB, HOMER, KEGG, MHCDB, MycoDB, PDB, PDBe, Pfam, PhosphoBase, PigBASE, PIR, PKR, PMD, PRODOM, Prolysis, PROSITE, PROTOMAP, RatMAP, RDP, REBASE, RGP, SBASE, 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 ..!!!!

There are lots of Bioinformatics Databases

For a annotated listing of major bioinformatics databases please see the online handout
[**< Major Databases.pdf >**](#)

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 in the address bar. The page is titled "NAR Database Summary Paper Category List". The left sidebar lists various database categories, many of which are hyperlinks. To the right of the sidebar is a vertical menu with several options. A large button at the bottom contains a URL.

You are here: [NAR Journal Home](#) » Database Summary Paper Categories

NAR Database Summary Paper Category List

- [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](#)
- [Cell biology](#)

- ▶ [Compilation Paper](#)
- ▶ [Category List](#)
- ▶ [Alphabetical List](#)
- ▶ [Category/Paper List](#)
- ▶ [Search Summary Papers](#)

<http://www.oxfordjournals.org/nar/database/cat/8>

GENBANK & REFSEQ:

NCBI'S NUCLEOTIDE SEQUENCE DATABASES

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



GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI
www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

Nucleotide (KIF5A) AND "Homo sapiens"

Display Settings: GenBank

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

NCBI Reference Sequence: NM_004984.2

FASTA Graphics

Go to:

ACCESSION NM_004984

LOCUS NM_004984 3897 bp mRNA linear PRI 10-JAN-2014

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

VERSION NM_004984.2 GI:45446748

KEYWORDS RefSeq.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3897)

AUTHORS Kawaguchi, K.

TITLE Role of kinesin-1 in the pathogenesis of SPG10, a rare form of her

JOURNAL Neu

PUBMED 227

REMARK Gen spa Rev

REFERENCE 2 Pro

AUTHORS Boh

TITLE alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay

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]

GenBank flat file format has defined fields including unique identifiers such as the **ACCESSION** number.

This same general format is used for other sequence database records too.

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	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/nuccore/NM_004984.2

Nucleotide (KIF5A) AND "Homo sapiens" Search Help

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.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; 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 paraplegia

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

PUBMED 22785106

REMARK GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia 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)

See all...
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 ubiquitylation dynamics reveals a key role for KIF5A in [Nat Cell Biol. 2012]
Pathways for the KIF5A gene
Peptide hormone metabolism
MHC class II antigen presentation

GenBank sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI
www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

Nucleotide (KIF5A) AND "Homo sapiens" Search Help

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 Can set different display formats here

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; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; 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 paraplegia

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

PUBMED 22785106

REMARK GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia 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 ubiquitylation dynamics reveals a key role for KIF5A in [Nat Cell Biol. 2012]

See all...

Pathways for the KIF5A gene

Peptide hormone metabolism

MHC class II antigen presentation

FASTA sequence record

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI
www.ncbi.nlm.nih.gov/nuccore/45446748?report=fasta

Nucleotide Nucleotide Search Sign in to NCBI

Display Settings: FASTA

Send: Change region shown

Customize view

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

NCBI Reference Sequence: NM_004984.2

GenBank **Graphics**

>gi|45446748|ref|NM_004984.2| Homo sapiens kinesin family member 5A (KIF5A), mRNA

ACGCCAGTCGCCGCATCCCGTCCGCAAGAGACAGCGCAGCCCTGCTCCCCAGGCTT CGCCGGGGCGCCCTCAACTCTGCCCCAGAGACTGACCACCTGTCCCTCCGCTCGCCCTGCTGAGAGC CCTCTCTCTGGAGCACACACCACCCCTGCAGCCAAGAAAGAGTCCCAGCCCCACGCCGGCTACCACCAT GGCGGAGACCAACAACGAATGTAGCATCAAGGTGCTCTGCCATTCCGGCCCTGAACCAGGCTGAGATT CTGCGGGAGACAAGTTCATCCCCATTTCAGGGACGACAGCGTCTTATTGGGGGAAGCCATATG TTTTGACCGTGTATTCCCCCAAACACGACTCAAGAGCAAGTTTATCATGCCATGTGCCATGCAGATTGT CAAAGATGTCCTTGTGGCTACATGGCACCATTTGGCTTATGGACAGACATCCTCAGGAAAACACAT ACCATGGAGGGAAAGCTGCACGCCCTCAGCTGATGCGAACATCATTCTCGAATTGCGGGAGACATCTCA ACCACATCTACTCCATGGATGAGAACCTTGAGTTCCACATCAAGGTTCTACTTGAATAATTACCTGGA CAAAATCGTGAACCTCTGGATCTGACCAAGAACATCTGTCCTGCAAGGAGACAAGAACCGGGTCCA TTTGTCAGGGTTGTACTGAAAGCTTGTGTCAGCCGGAGGAGATTCTGGATGTGATTGAGGA AACAAATCGTCATGGCTGACCAACATGAAACACAGCTCAGCTGGAGCTGTATCTGGACCTGGCA CAACATCAAGCAGGAGAACATGGAAACGGAGCAAGCTGTCCTGGACAGGCAAAAGAATATCAACAGT CGGACTGAGAACGTCAGCAAGACTGGAGCAGAGGGACCCCTGCTGGACAGGCAAAAGAATATCAACAGT CACTGTCAGCTCTGGCAATGTACTCCGACTGGCTGAGGGCACTAAAAAGCTATGTTCCATATCGTGA CAGCAAATGACAAGGATTCTCAGGACTCTCTGGGGAAACTGCGGAGCAGTATGTTCATCTGTGC TCACCATCCAGTTAAATGATCCAGAGACCAAGTCCACCCCTGATGTTGGCAGCGGGCAAAGACATTA AGAACACTGCTCAGTAAATTGGAGTTGACTGCTGACGAGTGGAAAGAAGAAATATGAGAAGGAGAAGCA GAAGACAAAGGCCAGAAGGAGCAGATTGCGAAGGCTGAGCTGAGCCGGTGGCGCAATGGAGAG AATGTCCTGAGACAGAGGCCCTGGCTGGGAGGAGCCAGCCCTGGAGGCCAGCTCTGTGAGGAGACCC CTGTCAGTACAACATCCATCTGGCTGCCATCCGGGGAGGACCCAGAAATACGACGACAGAT CGGGCGCTCTATAAGCAGCTGACGACAAGGATGATGAAATCAACCAACAAAGCCAACATAGACAAAG CTCAACCGCAAATGCTGGACCAAGGAAGAGCTGCTGGTCCACCCGAGGAGACAACGAGAAGGTCAGC GGGACCTGAGCCACCTGCAATCAGAGAACGATGCGCTAAGGATGAGCTGAAGGAAGTGCTGCAGGGCCT GGAGGAGCTGGCTGTGAACTATGACCAAGAAGTCCAGGAGGAGGAGAAGAGCCAGCAGAACAGCTT CTGGTGGATGAGCTGTCTCAGAAGGTGGCCACCATGCTGTCCTGGAGTCTGAGTTGCGAGCGGCTACAGG AGGTCAGTGGACACCAGCAGAAACGAATTGCTGAGGTGCTGAACGGGCTGATGAGGATCTGAGCGAGTT

FASTA sequence files consist of records where each record begins with a ">" and header information on that same line. Each subsequent line of the record is sequence information.

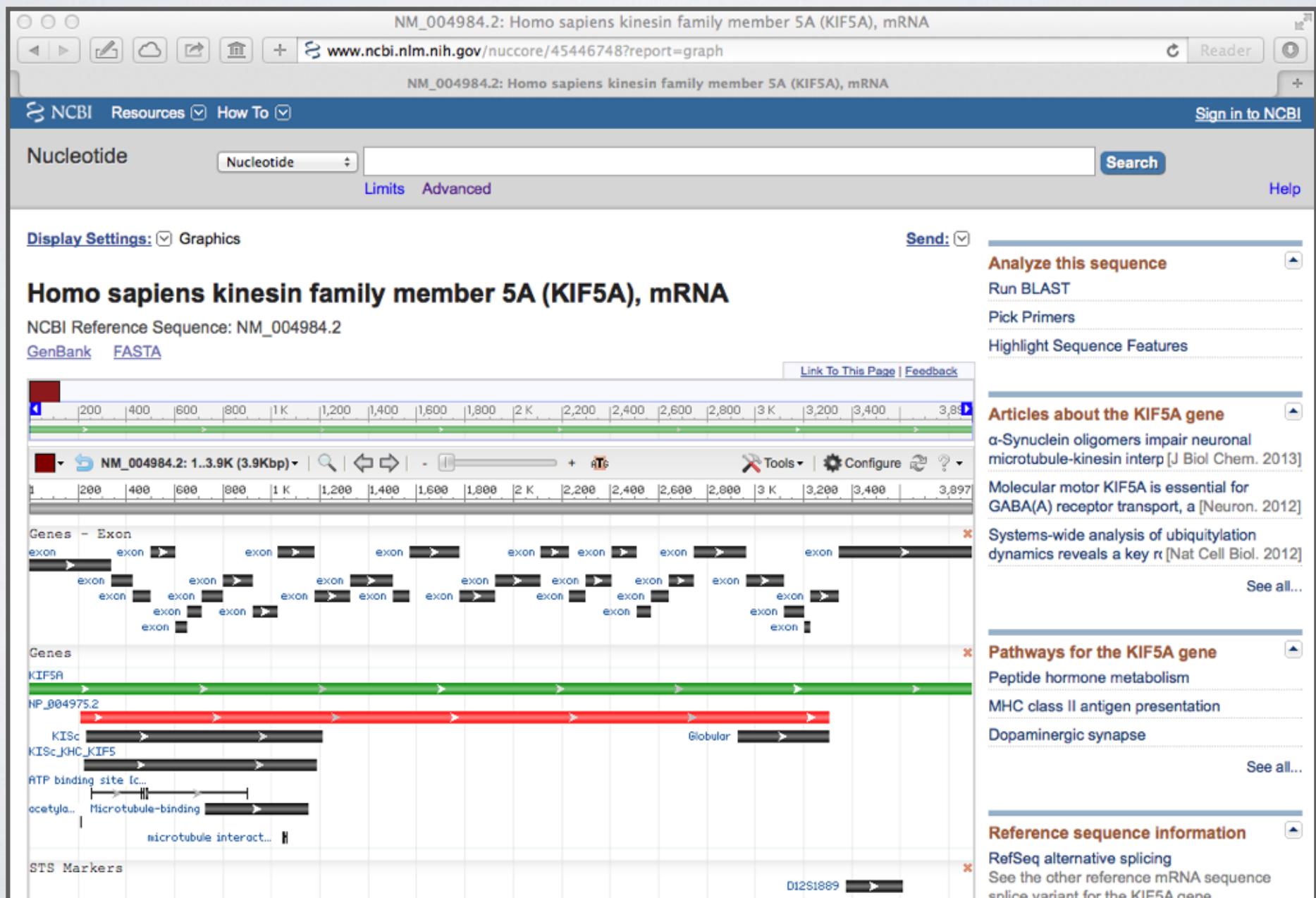
This format is commonly used by sequence analysis programs.

Pathways for the KIF5A gene

Peptide hormone metabolism

MHC class II antigen presentation

GenBank ‘graphics’ sequence record



GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI
www.ncbi.nlm.nih.gov/nuccore/NM_004984.2

Nucleotide (KIF5A) AND "Homo sapiens" Search Help

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.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; 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 paraplegia

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

PUBMED 22785106

REMARK GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplegia 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)

See all...
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 ubiquitylation dynamics reveals a key role for KIF5A in [Nat Cell Biol. 2012]
Pathways for the KIF5A gene
Peptide hormone metabolism
MHC class II antigen presentation

GenBank sequence record, cont.

Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI
www.ncbi.nlm.nih.gov/nuccore/45446748?report=genbank&to=3897#feature_45446748

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

FEATURES Location/Qualifiers OMIM

source 1..3897 Probe

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:[9606](#)" Protein

/chromosome="12"

/map="12q13.13"

gene 1..3897 PubMed

/gene="KIF5A"

/gene_synonym="D12S1889; MY050; NKHC; SPG10"

/note="kinesin family member 5A"

/db_xref="GeneID:[3798](#)"

/db_xref="HGNC:[6323](#)"

/db_xref="HPRD:[09108](#)"

/db_xref="MIM:[602821](#)"

exon 1..337 PubMed (RefSeq)

/gene="KIF5A"

/gene_synonym="D12S1889; MY050; NKHC; SPG10"

/inference="alignment:Splign:1.39.8"

misc_feature 134..136

/gene="KIF5A"

/gene_synonym="D12S1889; MY050; NKHC; SPG10"

/note="upstream in-frame stop codon"

CDS 209..3307

/gene="KIF5A"

/gene_synonym="D12S1889; MY050; NKHC; SPG10"

/note="kinesin, heavy chain, neuron-specific; KIF5A variant protein; neuronal kinesin heavy chain; kinesin heavy chain neuron-specific 1"

/codon_start=1

/product="kinesin heavy chain isoform 5A"

/protein_id="NP_004975.2"

/db_xref="GI:45446749"

/db_xref="CCDS:CCDS8945.1"

/db_xref="GeneID:[3798](#)"

/db_xref="HGNC:[6323](#)"

/db_xref="HPRD:[09108](#)"

/db_xref="MIM:[602821](#)"

translation="MAETNNNECSIKVLCRFRPLNQAEILRGDKFIPIFQGDDSVVIGG KPYVFDRVFPPTTQEQQVYHACAMQIVKDVLAGYNGTIFAYGQTSSGKTHTMEGKLHD PQLMGIIPRIARDIFNHIYSMDENLEPHIKVSYFEIYLKDRLLDVTKTNLNVHEDK NRPVFKGCTERFVSSPEEILDVIDECKSNRHVAVTNMNEHSSRSHSIFLINIKQENM ETEQKLSGKLYLVDLAGSEKVKSTGAEGAVLDEAKNINKSLALGNVISALAEGTKS VPYRDSKMTRILQDSLGGNCRTTMFICCSPPSYNDAETKSTLMFGQRAKTIKNTASVN

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] (1351) Nucleotide

kinesin (37064) Nucleotide

See more...

GenBank sequence record, cont.

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

www.ncbi.nlm.nih.gov/nuccore/45446748?report=genbank&to=3897#sequence_45446748

Reader

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

```
/gene="KIF5A"
/gene_synonym="D12S1889; MY050; NKHC; SPG10"
/standard_name="D12S1889"
/db_xref="UniSTS:48006"

ORIGIN
 1 acggccaggc cgccccgatc ccgtcgccgc aggagagaga cagcgcgcgc cggccctgtct
 61 ccccaggctt cgccccgggg ccctcaactc tggcccccaga gactgagcac ctgtccctcg
121 cctcggcctc tgctgagacg cctcttcctct ggacacaca ccaccctgc agcccaagaa
181 gagtcggcgc cccacgcggc ctaccatcc ggccggagacc aacaacgaat gttagcatcaa
241 ggtgtctgc cgattccggc ccctgaacca ggctgagatt ctgcggggag acaagttcat
301 ccccattttc caaggggacg acagcgtctg tattgggggg aagccatatg tttttgaccg
361 tggatcccccc ccaaacaacgca ctcaagagca agtttatcat gcatgtgccca tgcagattgt
421 caaaatgtc cttgctggct acaatggcac cattttgtct tatggacaga catcctcagg
481 gaaaacacat accatggagg gaaagtgcga cgaccctcg ctgtatggaa tcattctcg
541 aattgcccga gacatcttca accacatcta ctccatggat gagaaccttg agttccacat
601 caagggtttct tactttgaaa ttacatggaa caaaattctgt gacccatctgg atgtgaccaa
661 gacaaatctg tccgtgcacg aggacaagaa ccgggtgcctt tttgtcaagg gttgtactga
721 acgttttgtt tccagccgg aggagattctt ggatgtgattt gatgaaggaa aatcaaatcg
781 tcatgtggct gtcaccaaca tgaatgaaca cagctctcg agccacacgca tccatctcg
841 caacatcaag caggagaaca tggaaacggg gcagaagctc agtggaaagc tggatctgg
901 ggacctggca gggagtggaa aggtcagca gactggagca gagggagccg tggatggacga
961 gccaagaat atcaacaatgt cactgtcgc tctggcaat gtatctccg cactggctga
1021 gggactaaaa agctatgttccatatgtca cagcaaaatg acaaggatttccaggactc
1081 tctcggggga aactggcgga cgactatgtt catatgttgc tcaccatcca gttataatga
1141 tgcagagacc aagtccaccc tggatgttgg gcaacggggca aagaccat aagacactgc
1201 ctcagtaaat ttggagttga ctgctgagca gtggaaagaag aaatatgaga aggagaagga
1261 gaagacaaag gcccagaagg agacgattgc gaagctggag gctgactgaa gcccgtggcg
1321 caatggagag aatgtgcctg agacagagcg cctggctggg gaggaggcag ccctggggagc
1381 cgactctgtt gaggagaccc ctgtgaatga caactcatcc atcgtggcgc gcatcgcc
1441 cgaggagccg cagaaatacg aggaggatcc cccgcgtctc tataaggcgc ttgacgacaa
1501 ggtatgtgaa atcaaccaac aaaggccaact catagagaag ctcaaggc aatgtctgg
1561 ccacccgggg agtgcgttgc ccacccgggg agacaacccgg aagggtccgc gggagcttag
1621 ccacctgcac tcaacccgggg agtgcgttgc ccacccgggg agacaacccgg aagggtccgc
1681 ggaggagctg gttgtgaacttggatggggatgg gttggaggaga agaggccac
1741 gaacccgggg tggatgttgc tggatgttgc gttggaggaga agaggccac
1801 tggatgttgc tggatgttgc tggatgttgc gttggaggaga agaggccac
1861 gaacccgggg tggatgttgc tggatgttgc gttggaggaga agaggccac
1921 gttgttgc tggatgttgc tggatgttgc gttggaggaga agaggccac
1981 cggatgttgc tggatgttgc tggatgttgc gttggaggaga agaggccac
2041 ccacccgggg tggatgttgc tggatgttgc gttggaggaga agaggccac
2101 cccatctctt cggatgttgc tggatgttgc gttggaggaga agaggccac
2161 gttgttgc tggatgttgc tggatgttgc gttggaggaga agaggccac
2221 ccacccgggg tggatgttgc tggatgttgc gttggaggaga agaggccac
2281 tggatgttgc tggatgttgc tggatgttgc gttggaggaga agaggccac
2341 ccacccgggg tggatgttgc tggatgttgc gttggaggaga agaggccac
```

The actual sequence entry starts after the word **ORIGIN**

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

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 two sides of UniProtKB



UniProtKB/**TrEMBL**

Redundant, automatically
annotated - unreviewed

UniProtKB/**Swiss-Prot**

Non-redundant, high-quality
manual annotation - reviewed

★ Unreviewed, UniProtKB/TrEMBL **Q9N0H9** (Q9N0H9_EQUAS)

★ Reviewed, UniProtKB/Swiss-Prot **P38398** (BRCA1_HUMAN)

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

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

[1] "The quaking gene product necessary in embryogenesis and myelination combines features of RNA binding and signal transduction proteins."

Ebersole T.A., Chen Q., Justice M.J., Artzt K.

Nat. Genet. 12:260-265(1996) [PubMed: 8589716] [Abstract]

Cited for: NUCLEOTIDE SEQUENCE [mRNA] (ISOFORM 3), INVOLVEMENT IN QKV, TISSUE SPECIFICITY, MUTAGENESIS OF QKV

[2] "Genomic organization of the mouse *Qkv* gene and expression of its four alternative splicing isoforms during development"

Kondo T., Furuta I., Mitsuhashi K., Ebersole T.A., Shiochi M., Wu J., Artzt K., Yamamoto K., Abe K.

Mamm. Genome 10:662-669(1999) [PubMed: 10384037] [Abstract]

Cited for: NUCLEOTIDE SEQUENCE [GENOMIC DNA / mRNA] (ISOFORMS 2; 3; 4 AND 7), ALTERNATIVE SPlicing (ISOFORM 1).

Strain: 129/J.

References

General annotation (Comments)

Function: RNA-binding protein that plays a central role in myelination. Also required for visceral endoderm function and blood vessel development. Binds to the 5'-NAC(UAU)N-3' UAU(A-Y) RNA core sequence. Acts by regulating pre-mRNA splicing, mRNA export, mRNA stability and mRNA processing. Involved in regulation of mRNA splicing of myelin basic protein (MBP) mRNA. Required for MBP mRNA to promote stability of mRNAs such as MBP and CNH1B to promote oligodendrocyte differentiation. Participates in mRNA transport by regulating the nuclear export of MBP mRNA. Isoform 1 is involved in regulation of mRNA splicing of MAG pre-mRNA by acting as a component of the complex of alternative splicing factors. Isoform 1 also acts as a splicing factor, while neurodifferentiation web other isoforms result in nuclear translocation of isoform 1 and suppression of splicing. Isoform 4 acts as a transcriptional repressor for GKI. May also play a role in smooth muscle development.

Subunit structure: Homodimer. Does not require homodimerize. Able to heterodimerize with B1C1.

Subcellular location: Cytosol. Nucleus. Note-Isoforms 1 localizes predominantly in the nucleus, and at lower level in cytosol. It shuttles between the cytosol and nucleus. Isoform 2 localizes predominantly in the cytosol. Isoform 3 localizes predominantly in the cytosol and at much lower level in nucleus. Isoform 4 localizes both in the cytosol and nucleus.

Tissue specificity: Highly expressed in the developing nervous system. Isoforms 1 and 2 are present in the nervous system, but absent in developing blood cells. Isoform 3 is present in the nervous system, but expression is down-regulated during neuronal differentiation. By contrast, neural progenitors located in specific subcompartments of the vZ maintain expression as they differentiate and migrate away into the emerging nervous system. These have characteristics consistent with the acquisition of a glial rather than neuronal fate. Isoform 4 is present in the nervous system, but absent in early neurons, while isoforms 3 and 4 are present in the nervous system of adult mice. Isoform 4 is present in the peripheral nervous system, but absent in the central nervous system.

Developmental stage: High expression in the developing nervous system, particularly in the midbrain, forebrain, and spinal cord, affecting transport and/or distribution of MBP mRNA. The level of γ-phosphorylation in the developing nervous system is highest in the first postnatal week (P7). During the vigorous accumulation of MBP mRNA between P7 and P20, phosphorylation in the developing myelin drastically declined. By the end of the fourth postnatal week (P28), phosphorylation is reduced approximately 90%.

Involvement in disease: Defects in Qkv cause of quaking-like (qkv). Qkv is a spontaneous mutation resulting in hypomyelination of the central and peripheral nervous systems. Mice with qkv develop normally until postnatal day 10 when they display reduced tremor or "quaking" that is especially pronounced in hindlimbs and expresses progressive tonic-clonic seizures as they mature. Mice with qkv specifically lack isoform 1.

Cell cycle: Regulation of cell proliferation

DNA damage: Traceable author statement. Source: UniProtKB

DNA repair: Regulation of transcription from RNA polymerase II promoter

Fatty acid biosynthesis: Traceable author statement. Source: Protic

Lipid synthesis: Regulation of transcription from RNA polymerase III promoter

Nucleus: Traceable author statement. Source: UniProtKB

Polymorphism: Response to estrogen stimulus

Disease mutation: BRCA1-BARD1 complex

Repeat: Inferred from direct assay. Source: UniProtKB

Zinc-finger: Gamma-tubulin ring complex

DNA-binding: Non-traceable author statement. Source: UniProtKB

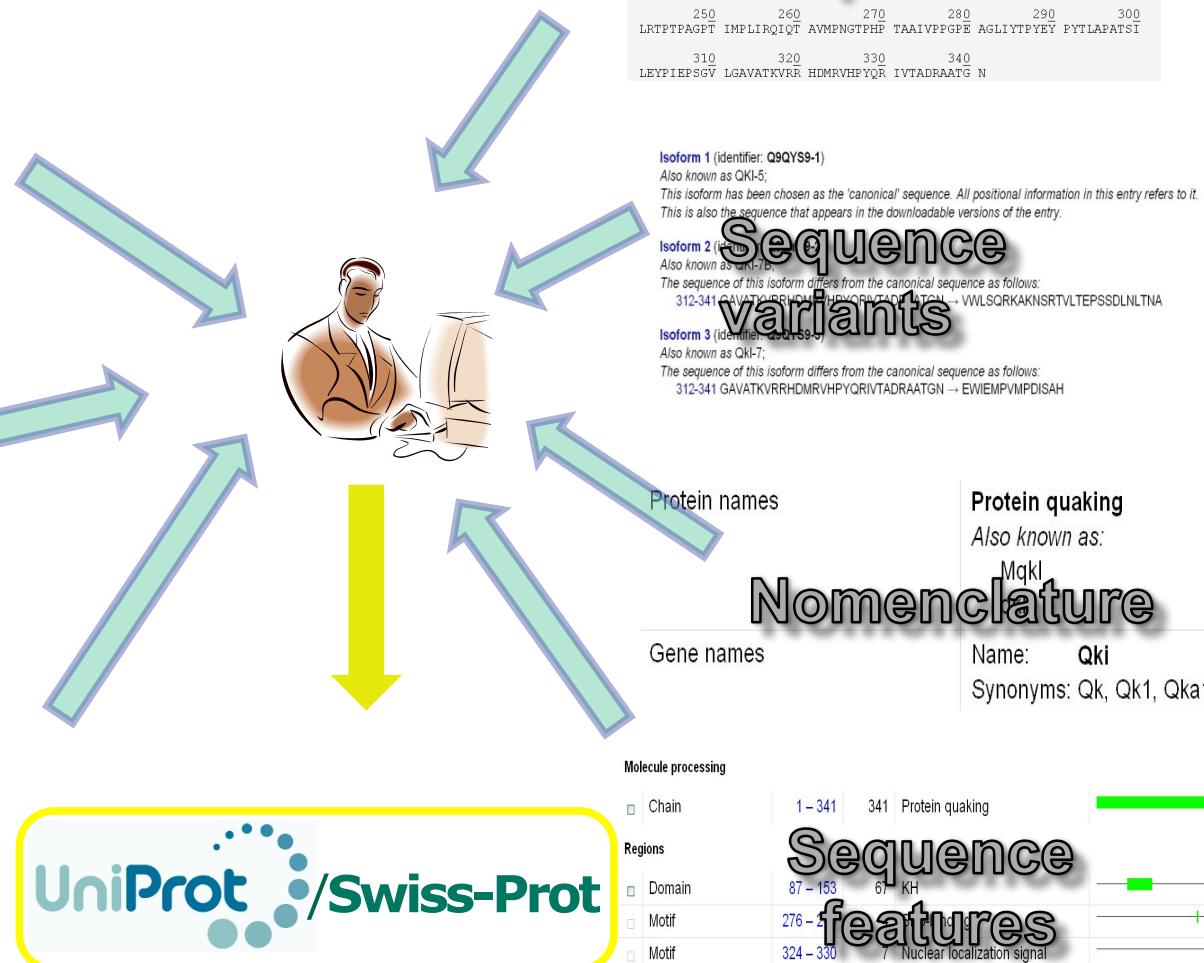
Androgen receptor binding: Traceable author statement. Source: UniProtKB

Enzyme binding: Non-traceable author statement. Source: UniProtKB

Phosphorylation: Inferred from physical interaction. Source: UniProtKB

3D-structure: Inferred from physical interaction. Source: UniProtKB

Ontologies



10	20	30	40	50	60
MVGEMETKEK	PKPTPDYLMQ	LMNDKKLMS	LPNFCGIFNH	LERLLDEEIS	RVRKDNYNTD
70	80	90	100	110	120
LNGSTEKRSA	ELPDAVGPIV	QLQEBKLYPV	KEYPDFNFVG	RILGPRGLTA	KQLEAETGCK
130	140	150	160	170	180
IMVRGKGSMR	DVQKQDQG	KRMKPEHLNEF	IHLVLLVEDA	ONBAEATKLKR	AVEEVKKLV
190	200	210	220	230	240
PAAEGEDSLK	KMQLMELAIL	NGTARDANIK	SPALAFSIAA	TAQAAFRIT	GPAFPVLPAA
250	260	270	280	290	300
LRTPTPAGPT	IMPLIPIQQT	AVMPNGTFHP	TAAIVPPGPE	AGLIYTPKEY	PTYLAPATS
310	320	330	340		
LEYPIEPSGV	LGAVATKVR	HDMRVHPYQR	IVTADRAATG		

Isoform 1 (Identifier: Q9QYS9-1)

Also known as QKI-5;

This isoform has been chosen as the 'canonical' sequence. All positional information in this entry refers to it. This is also the sequence that appears in the downloadable versions of the entry.

Isoform 2 (Identifier: Q9QYS9-2)

Also known as QKI-7;

The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRHHDMRVHPYQRIVTADRAATGN → VVLSQRKAKNSRTVLTEPSDSLNLNTA

Isoform 3 (Identifier: Q9QYS9-3)

Also known as QKI-7;

The sequence of this isoform differs from the canonical sequence as follows:
312-341 GAVATKVRHHDMRVHPYQRIVTADRAATGN → EWIEMPVMPDISAH

Gene names

Name: Qki

Synonyms: Qk, Qk1, Qka1

Molecule processing

Chain

1 - 341 341 Protein quaking

Regions

87 - 153 67 KH

Domain

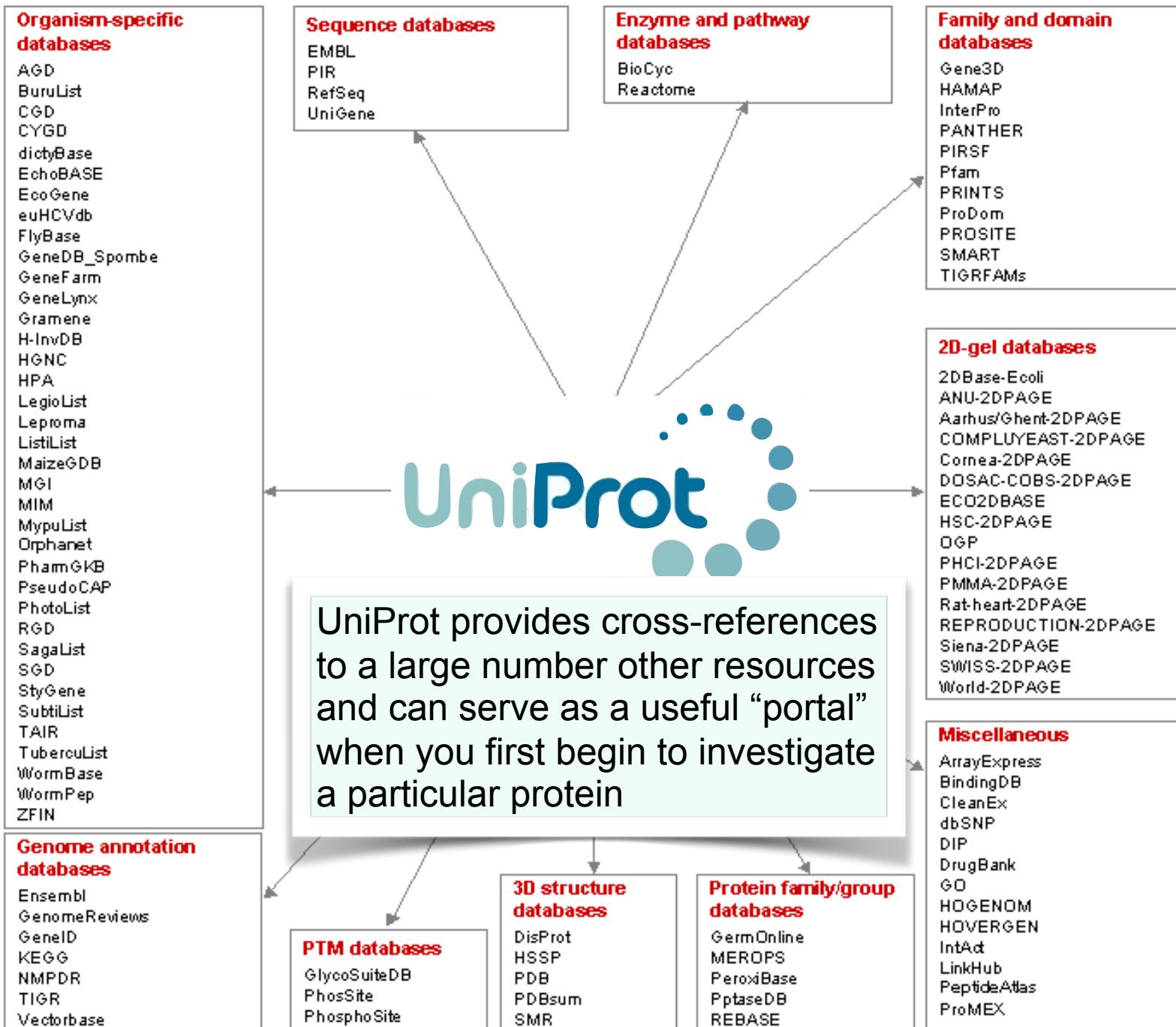
276 - 300 10 KH

Motif

324 - 330 7 Nuclear localization signal

276 - 300 10 KH

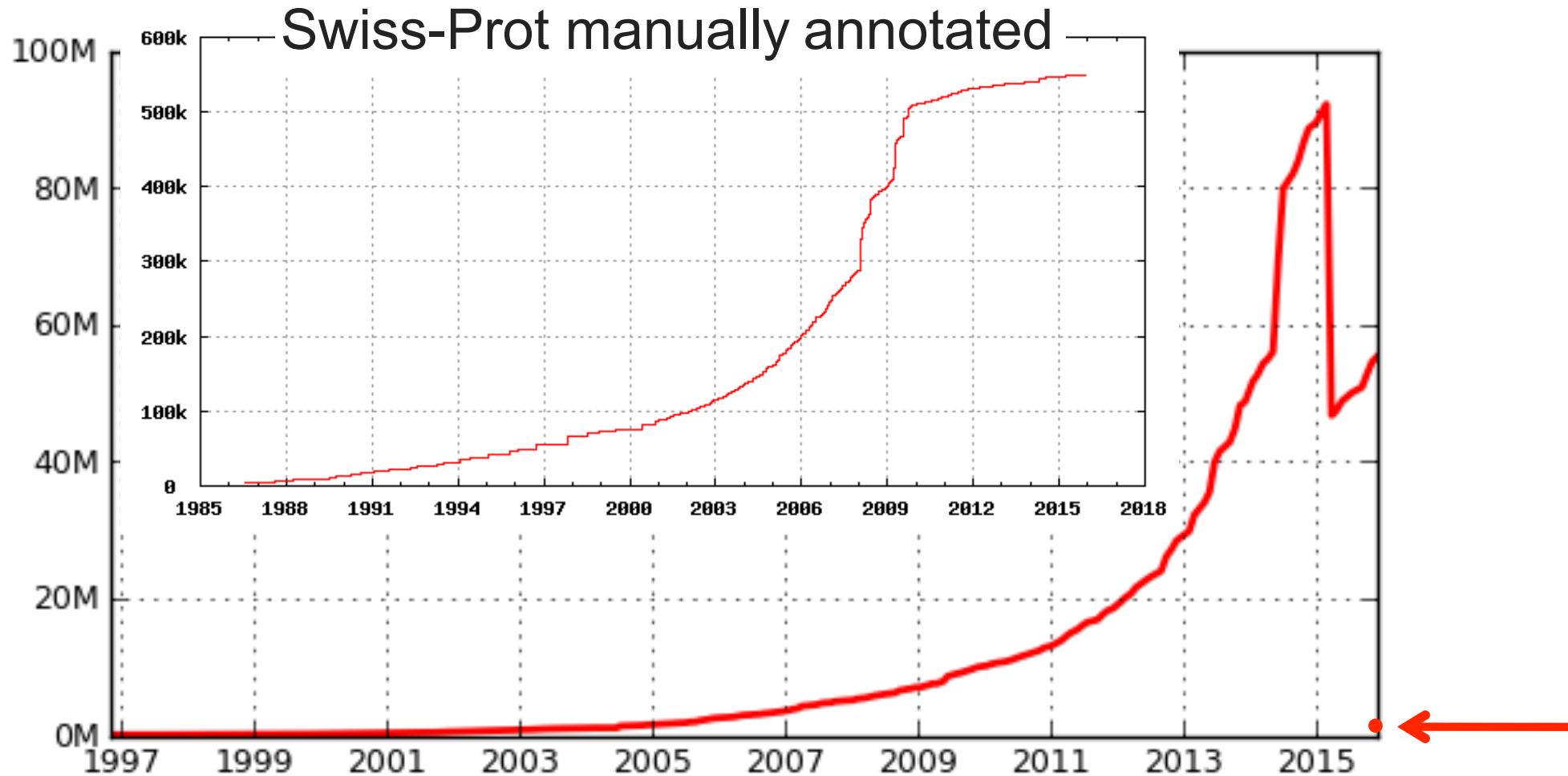
324 - 330 7 Nuclear localization signal



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)



Your Turn!

https://bioboot.github.io/bggn213_f17/lectures/#2

https://bioboot.github.io/bggn213_f17/lectures/#2

Home Gmail Gcal Bitbucket GitHub News Disqus BGGN-213

UC San Diego

BGGN 213

A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the Division of Biological Sciences, UCSD.

Overview

Lectures

Computer Setup

Learning Goals

Assignments & Grading

- 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/bggn213_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
ATGGTGCATCTGACTCCTGTGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG
TTGGTGGTGAGGCCCTGGCAGGCTGCTGGTGGTCTACCCCTGGACCCAGAGGTTCTTGAGTCCTTGG
GGATCTGTCCACTCCTGATGCAGTTATGGCAACCCCTAACGGTGAAGGCTCATGGCAAGAAAGTGCTCGGT
GCCCTTAGTGATGCCCTGGCTCACCTGGACAACCTCAAGGGCACCTTGCCACACTGAGTGAGCTGCACT
GTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGCTCCTGGCAACCTGCTGGTCTGTGTGCTGGCCA
TCACTTGGCAAAGAATTCAACCCCCACCAGTGCAGGCTGCCATCAGAAAGTGGTGGCTGGTGTGGCTAAT
GCCCTGGCCCACAAGTATCAACTAAGCTGCTTCTTGCTGTCCAATTT
```

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:

End times:

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

[9:45 am]

2. GENE database @ **NCBI**

[10:00 am]

— BREAK —

— 10:10 am —

3. UniProt & Muscle @ **EBI**

[10:35 am]

4. PFAM, PDB & NGL

[11:05 am]

— BREAK —

— 11:15 am —

5. Extension exercises

[11:45 am]

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

HOMEWORK

https://bioboot.github.io/bggn213_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.

THANK YOU