

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 <u>handout</u>!).
- 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? EB| UniProt
- What variants of this gene are associated with gastric cancer and other human diseases? NCBI OMIN
- 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:
 - BLAST, GenBank and OMIM @ NCBI [~35 mins]
 GENE database @ NCBI [~15 mins] BREAK —
 UniProt & Muscle @ EBI [~25 mins]
 PFAM, PDB & NGL [~30 mins] BREAK —
 Optional extension exercises [~20 mins]
 - Please do answer the last review question (Q19).
 - We encourage <u>discussion</u> and <u>exploration</u>!

BGGN-213

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, MIDB, MmtDB, Mol-R-Us, MPDB, MRR, MutBase, MycDB, NDB, NRSub, 0-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, SegAnaiRef, 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,

SideNote: Bioinformatics Databases



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

	www.oxfordjournals.org/nar/database	/cat/8 Č	t o
Home Gmail	Gcal Bitbucket GitHub News ~	Disqus BGGN-213	+
OXFORD Jour	nals		
You are here: NAR Journal Home » Database Summa	ary Paper Categories		
NAR Database Summary Paper Nucleotide Sequence Databases RNA sequence databases	Category List	Compilation Paper Category List Alphabetical List Category/Paper List Search Summary Papers	
Protein sequence databases Structure Databases Genomics Databases (non-vertebrate) Metabolic and Signaling Pathways Human and other Vertebrate Genomes Human Genes and Diseases	http://www.oxforc	journals.org/nar/c	latabase/c
CancerResource Protein Mutant Database General human genetics databases General polymorphism databases Cancer gene databases			_
Gene-, system- or disease-specific datab Microarray Data and other Gene Expression Da Proteomics Resources			
Other Molecular Biology Databases Organelle databases Plant databases			
Immunological databases			

GENBANK & REFSEQ: NCBI'S NUCLEOTIDE SEQUENCE DATABASES

What is GenBank?

- GenBank is NCBI's <u>primary</u> 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

00	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NC	
	🗅 🖻 🏦 + S www.ncbi.nlm.nih.gov/nuccore/NM_004984.2	C Reader
	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI	+
S NCBI Res	ources 🖸 How To 🗹	Sign in to NCBI
Nucleotide	Nucleotide : (KIF5A) AND "Homo sapiens"	Search
	Limits Advanced	Help
Display Setting	<u>s:</u> ⊙ GenBank Sen	d: 🔍
		Change region shown
Homo sa	piens kinesin family member 5A (KIF5A), mRNA	
	e Sequence: NM_004984.2	Customize view
FASTA Graphi	<u>cs</u>	Customize view
<u>Go to:</u> ⊡		
LOCUS	M 004984 3897 bp mRNA linear PRI 10-JAN-2014	Analyze this sequence
DEFINITION	Jomo sapiens kinesin family member 5A (KIF5A), mRNA.	Run BLAST
	1M_004984 1M_004984.2 GI:45446748	Pick Primers
	RefSeq.	Highlight Sequence Features
	lomo sapiens (human)	Find in this Sequence
	<u>Homo sapiens</u> Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	fammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;	Articles about the KIF5A gene
	Catarrhini; Hominidae; Homo. L (bases 1 to 3897)	a-Synuclein oligomers impair neuronal
	Kawaguchi,K.	microtubule-kinesin interp [J Biol Chem. 2013]
	Role of kinesin-1 in the mathogenesis of SPG10. a rare form of	
	Neu ConDonk flat file format has defined fields inclu	
	GenBank flat file format has defined fields inclu	laing unique identifiers
	such as the ACCESSION number.	·
REFERENCE	This same general format is used for other segue	ence database records to
	5	
	Soh alpha-Synuclein oligomers impair neuronal microtubule-kinesin	- animajo for moren or gone
		Peptide hormone metabolism
1	interplay J. Biol. Chem. 288 (30), 21742-21754 (2013)	MHC class II antigen presentation

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

GenBank sequence record

	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleot	ide – NCBI
< > Z		C Reader 🔘
	Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucleotide -	NCBI +
NCBI Re	sources 🖸 How To 🖸	Sign in to NCBI
Nucleotide	Nucleotide = (KIF5A) AND "Homo sapiens"	Search
	Limits Advanced	Help
Display Settin	ga: ⊡ GenBank	Send: 🖂
		Change region shown
Homo sa	piens kinesin family member 5A (KIF5A), mRNA	
	ce Sequence: NM 004984.2	
FASTA Grad		Customize view
<u>Go to:</u> ⊡		Analyze this sequence
LOCUS	NM 004984 3897 bp mRNA linear FRI 10-JAN-2014	Analyze and bequence
DEFINITION	Homo sapiens kinesin family member 5A (KIF5A), mRNA.	Run BLAST
ACCESSION	NM_004984 NM_004984.2 GI:45446748	Pick Primers
KEYWORDS	RefSeq.	Highlight Sequence Features
SOURCE	Homo sapiens (human)	Find in this Sequence
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;	Articles about the KIE5A gene
	Catarrhini; Hominidae; Homo.	Antioles about the full on gene
AUTHORS	1 (bases 1 to 3897) Kawaguchi.K.	α-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013]
TITLE	Role of kinesin-1 in the pathogenesis of SPG10, a rare form of	Molecular motor KIF5A is essential for
	hereditary spastic paraplegia	GABA(A) receptor transport, a [Neuron. 2012]
JOURNAL	Neuroscientist 19 (4), 336-344 (2013) 22785106	Systems-wide analysis of ubiquitylation
REMARK	GeneRIF: A review of the mechanism of pathogenesis involved in	dynamics reveals a key rr [Nat Cell Biol. 2012
	spastic paraplegia type 10 when KIF5A is inactivated by mutations.	See all.
REFERENCE	Review article 2 (bases 1 to 3897)	566 dil
AUTHORS	Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R.,	
TITLE	Bohm,K.J. and Winner,B. alpha-Synuclein oligomers impair neuronal microtubule-kinesin	Pathways for the KIF5A gene
TITLE	aipna-Synuclein oligomers impair neuronai microtubule-kinesin interplay	Peptide hormone metabolism
	J. Biol. Chem. 288 (30), 21742-21754 (2013)	

GenBank sequence record

	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI	
∢ ▶] 🛃	△ (2) (1) (1) (1) (2) (2) (2) (2) (2) (2) (2) (2) (2) (2	🖒 Reader
	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – NCBI	
S NCBI Re	sources 🖂 How To 🖂	Sign in to NCE
Nucleotide	Nucleotide : KIF5A) AND "Homo sapiens"	Search
	Limits Advanced	Hel
Diaplay Sattin	as: ♡ GenBank Send: ♡	
Jispiay Settin		Change region shown
Homoer	piens kinesin family member 5A (KIF5A), mRNA	Change region shown
	ce Sequence: NM_004984.2	tomize view
FASTA Grap	🗠 🗲 Can set different display formats here	
<u>3o to:</u> ⊘		
Locus	NU 001001	Analyze this sequence
DEFINITION	NM_004984 3897 bp mRNA linear PRI 10-JAN-2014 Homo sapiens kinesin family member 5A (KIF5A), mRNA.	Run BLAST
CCESSION	NM 004984	Pick Primers
ERSION	NM_004984.2 GI:45446748	Usebisti Oserana Frataria
EYWORDS	RefSeq.	Highlight Sequence Features
OURCE	Homo sapiens (human)	Find in this Sequence
ORGANISM	Homo sapiens Eukarvota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia: Eutheria: Euarchontoglires: Primates: Haplorrhini:	
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini; Catarrhini; Hominidae; Homo.	Articles about the KIF5A gene
	Catarrhini; Hominidae; Homo. 1 (bases 1 to 3897)	α-Synuclein oligomers impair neuronal
AUTHORS	Catarrhini; Hominidae; Homo. 1 (bases 1 to 3897) Kawaguchi,K.	Antioles about the full of gene
	Catarrhini Hominidae; Homo. 1 (baeei 1 ol 3987) Xawayohi,X. Role of kinesin-1 in the pathogenesis of SPG10, a rare form of	a-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013 Molecular motor KIF5A is essential for
AUTHORS	Catarrhini; Hominidae; Homo. 1 (bases 1 to 3897) Kawaguchi,K.	α-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013
AUTHORS	Catarrhini Hominidae; Homo. 1 (based 1 to 3997) Kawaguchi,K. Role of kinesin-1 in the pathogenesis of SPC10, a rare form of heredicary passic paraplogia Meuroscientist 19 (4), 336-344 (2013) 22785106	A-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013 Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012 Systems-wide analysis of ubiquitylation
AUTHORS TITLE JOURNAL	Catarrhini, Hominides, Homo. 1 (basea 1 to 3987) Xawagnohi,X. Role of kinesin-1 in the pathogenesis of SPC10, a rare form of hereditary spastic paraplogia Neuroscientics 19 (4), 336-344 (2013) <u>22785106</u> GeneRIF: A review of the mechanism of pathogenesis involved in	a-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013 Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012
AUTHORS TITLE JOURNAL PUBMED	Catarrhini Hominidae; Homo. 1 (basea 1 to 3897) Kawaguchi,K. Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplogia Neuroscientist 19 (4), 336-344 (2013) <u>2785106</u> GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplogia type 10 whom KIF5A is inactivated by mutations.	A-Synuclein oligomers impair neuronal microtubule-kinesin interp [J Biol Chem. 2013 Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012 Systems-wide analysis of ubiquitylation
AUTHORS TITLE JOURNAL PUBMED REMARK	Catarrhini, Hominidae; Homo. 1 (basea 1 to 3987) Xawaguohi,X. Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplogia Neuroscientist 19 (4), 336-344 (2013) <u>22785106</u> GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplogia type 10 when KIP5A is inactivated by mutations. Review article	Autor a Synuclein ofigomers impair neuronal microtubule-kinesin interp [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 rr [Nat Cell Biol. 2012
AUTHORS TITLE JOURNAL PUBMED REMARK	Catarrhini Hominidae; Homo. 1 (basea 1 to 3897) Kawaguchi,K. Role of kinesin-1 in the pathogenesis of SPG10, a rare form of hereditary spastic paraplogia Neuroscientist 19 (4), 336-344 (2013) <u>2785106</u> GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplogia type 10 whom KIF5A is inactivated by mutations.	Autor a Synuclein ofigomers impair neuronal microtubule-kinesin interp [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 rr [Nat Cell Biol. 2012
AUTHORS TITLE JOURNAL PUBMED REMARK	Catarrhini; Hominidae; Homo. 1 (bases 1 to 3987) Kawaguohi,K. Role of kinesin-1 in the pathogenesis of SPC10, a rare form of heredicary spastic paraplegia Meuroscientics 19 (4), 316-3144 (2013) 22725126 A review of the mechanism of pathogenesis involved in spastic paraplegia type 10 when KIP5A is inactivated by mutations. Review article 2 (bases 1 to 3987)	a Synuclein oligomers impair neuronal microtubule-kinesin inter J Biol Chem. 2017 Molecular moto: KIFSA is essential for GABA(A) receptor transport, a [Neuron. 2017 Systems-wide analysis of ubiquitylation dynamics reveals a key rr [Nat Cell Biol. 2017 See all
TITLE JOURNAL PUBMED REMARK REFERENCE	Catarrhini, Hominidae; Homo. 1 (basea 1 to 3987) Kawaguchi,K. Role of Kinesin-1 in the pathogenesis of SPC10, a rare form of hereditary spastic paraplogia Neuroscientical 19 (4), 336-344 (2013) 22785106 GeneRIF: A review of the mechanism of pathogenesis involved in spastic paraplogia type 10 when KIP5A is inactivated by mutations. Review article 2 (bases 1 to 3987) Prots.J., Veber,V., Brey,S., Campioni,S., Buder,K., Riek,R.,	a Synuclein oligomers impair neuronal microtubule-kinesin inter J Biol Chem. 2017 Molecular moto: KIFSA is essential for GABA(A) receptor transport, a [Neuron. 2017 Systems-wide analysis of ubiquitylation dynamics reveals a key rr [Nat Cell Biol. 2017 See all

FASTA sequence record

	-					NA – Nucleoti				
		+ 8 www.no	bi.nlm.nih.gov/n	nuccore/45446748?report	t=fasta				C Reade	er 🚺 🖸
			Homo sapiens k	kinesin family member 5A	(KIF5A), mRNA	- Nucleotide -	NCBI			+
S NCBI Resource	s 🖂 How To								<u>Sign in</u>	to NCBI
Nucleotide										
NUCIEO LIGE	Nucleo		mits Advanced					Sea	arch	Help
		LI	mits Advanced							нөр
Display Settings: 🖂	FASTA						Send: 🖂	a		
								Change region sh	lown	-
Homo sapie	ns kine	sin family	/ member	⁻ 5A (KIF5A), m	RNA			enange region en		
NCBI Reference Sec		-								
GenBank Graphics		001001.2						Customize view		
	NM 004984	.2 Homo sap	iens kinesin	family member 5A (1	KIF5A), mRNA					
			IGACAGCGCGCCCC	CGGCCCTGCTCCCCAGGCT	r					
		CAGAGACTGAGC	ACCTGTCCTCCG	CCTCGGCCTCTGCTGAGAG	c			C 11		- 1
CTCTCCTCTGGAGCA	CACACCACCO	CAGAGACTGAGC	ACCTGTCCTCCGG		c	STA se	quenc	e files con	sist of	
CTCTCCTCTGGAGCA GCGGAGACCAACAAC TGCGGGGGAGACAAGT	CACACCACCO GAATGTAGCA FCATCCCCAT	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA	ACCTGTCCTCCGG AAGAGTCCCAGCG GCCGATTCCGGCG CGACAGCGTCGT	CCTCGGCCTCTGCTGAGAGG CCCACGCCGGCTACCACCA CCCTGAACCAGGCTGAGAT TATTGGGGGGGAAGCCATAT	FAS					
CTCTCCCTCTGGAGCA CCCGCAGACCAACAAC TGCGGGGAGACAAGT TTTTGACCGTGTATT	CACACCACCO GAATGTAGCA FCATCCCCAT CCCCCCAAAC	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACTCAAGAG	ACCTGTCCTCCG AAGAGTCCCAGCO GCCGATTCCGGCO (CGACAGCGTCGT) ICAAGTTTATCATO	CCTCGGCCTCTGCTGAGAGG CCCACGCCGGCTACCACCA CCCTGAACCAGGCTGAGAT TATTGGGGGGGAAGCCATAT GCATGTGCCATGCAGATTG	FAS rec	ords w	here e	each record	d begins	3
CTETCETETGGAGEA GEEGGAGACCAACAAC TGEEGGGGAGACAAGT TTTTGACCGTGTATT CAAAGATGTCCTTGCT	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACTCAAGAG GCACCATTTTTG	ACCTGTCCTCCG AAGAGTCCCAGC GCCGATTCCGGC (CGACAGCGTCGT) CAAGTTTATCAT (CTTATGGACAGA	CCTCGGCCTCTGCTGAGAGG CCCACGCCGGCTACCACCA CCCTGAACCAGGCTGAGAT TATTGGGGGGGAAGCCATAT	FAS F FAS	ords w	here e		d begins	3
CTCTCCTCTGGAGCA GCGGAGACCAACAAC TTCGGGGGAGACAAGT TTTTGACCGTGTATT CAAAGATGTCCTTGCT ACCATGAGGGAAAGC ACCACATCTACTCCAT	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG TGCACGACCC GGATGAGAAC	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACTCAAGGGGA GCACCATTTTTG TCAGCTGATGGG CTTGAGTTCCAC	ACCTGTCCTCCG AAGAGTCCCAGC GCCGATTCCGGC CGACAGCGTCCT CCAAGTTATCAT CTTATGGACAGA AATCATTCCTCG ATCAAGGTTTCT	CCTCGGCCTCTGCTGAGAG CCCAGGCCGGCTACACCAC CCCTGAACCAGGCTGAGAT CCCTGACCAGGGGAACCCATAT GCATGTGCCATGCAGATG CATCTCCCGGAGAAACCAC ANTGCCCCGGAGACATCTCC TACTTGAAATTACCTGG	FAS rec	ords w n a ">"	here e and he	each record eader infor	d begins	5
CTCTCCTCTGGAGCA SCCGGAGACCAACAAC TTCTGGCCGGGAGACAAGT TTTTGACCGTGTATT CAABGATGTCCTTGCT ACCATGGAGGGAAAGC ACCACGATCTACTCCAT CAABATTCGTGACCTT	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG GGCTACAATG GGATGAGAAC CTGGATGTGA	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACTCAAGAG GCACCATTTTG CCACTGATGGG CTTGAGTTCCAC ACCAAGACAAATC	ACCTGTCCTCCGG (AAGAGTCCCAGC) (GCCGATTCCGGC) (CGACGCGCCGT) (CAAGTTTATCATC) (CTTATGGACAGA) (AATCATCGTCGGCAGA) (AATCATCCTCG] (TCCAGGCACGA)	CCTCGCCTCTGCTGAGAG CCCAGCCGCGCATACCACCA CCCTGAACCAGGCTGAGAT TATTGGGGGGAAGCCATAT GCATCTCCAGGGAAACACA AATTGCCGGGGAACACAT TACTTGAAATTACCTGG AGGACAAGACCGGCTGCC	FAS Freck With A on	ords w n a ">"	here e and he	each record	d begins	;
CTCTCCTCTGGAGCA SGCGGAGACCAACAAC TTGCGGGGAGACAAGT TTTTTGACCGTGTATT SAAAGATGTCCTTGCT SCATGGAGGGAAAGC ACCACATCTACTCCAT SAAAATCGTGACCCAT TTGTCAAGGGTTGTA	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG TGCACGACCC GGATGAGAAC CTGGATGTGA CTGGACGCTT	CAGAGACTGAGC CCTGCAGCCCAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACCATTTTG GCACCATTTTTG TCAGCTGATGCGA CCTTGAGTTCCAG CCCAAGACAAATC TGTGTCCAGCCC	ACCTGTCCTCCG (AAGAGTCCCAGC) (GCCGATTCCGGC) (GACAGCGTCGT) (CAAGTTATCAT(CTTATGGACAGAGAGAC) (AATCAAGGTTCT) (ATCAAGGTTCT) (GGCCGTGCACG) (GGAGGAGATTCT)	CCTCGGCCTCTGCTGAGAG CCCAGGCCGGCTACACCAC CCCTGAACCAGGCTGAGAT CCCTGACCAGGGGAACCCATAT GCATGTGCCATGCAGATG CATCTCCCGGAGAAACCAC ANTGCCCCGGAGACATCTCC TACTTGAAATTACCTGG	FAS rec with on	ords w n a ">" that sa	here e and he me lin	each record eader infor ie. Each	l begins mation	5
CTOTCCTCTGGAGCA SGCGGAGACCAACAAC TTGCGGGGAGACAAGT TTTTGACGGTATT SAAAATGTCCTTGCT CCATGGAGGAGAAAGC CCATGGAGGAGAAAGC SCACATCTACTCCAT TTTCCAAGGTTGTA VATCAAACGCAGGAG	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG GGCTACAACG CTGGACGACG CTGGACGCTT GGCTGTCACC AACATGGAAA	CAGAGACTGAGC CTGCAGCCCAAG TCAAGGTGCTCT TTTCCAAGGGGA ACGACTCAAGAG GCACCCATTTTG TCAGCTGATGGG CCTGAGTCCAC CCTGAGTCCAC CCTGAGTCCAGCCC TGTGTCCAGCCC TGTGTCCAGCCC TGTGTCCAGCCC	ACCTGTCCTCCG (ACGAGTCCCAGC) (GCCGATTCCGGC) (GACAGCGTCGT (CAAGTTATCATC (CTTATGGACAGA (AATCATCTTCG) (ATCAGTCTCTCG) (GAGGGGATTCTC) (GCAGGGGGAACC) (TCAGTGGGAACC)	CCTOBCCTTOCTGAGAG CCCAGCCGCGCTACCACCA CCCCGCACCCAGCTOACAT TATTGCGGCGAAGCCATAT GCATGTCCTCAGGGAAACCATAT AATTGCCCGGAAACCAC TACTTCAGAATTTACCTGG AGCCAAGACCCGGCTCCT GGATGTGATTGATGAAGGG ACCACACATCTTCCTCA TGTTCTGCGACTGCCC	F FAS F rec with on sub	ords w n a ">" that sa seque	here e and he me lin nt line	each record eader infor le. Each of the reco	l begins mation	5
CTCTCCTCGGAGA SCCGAGACCAACHAC TTTTGCGGGGAGACAACT TTTTGCGGGGGGAAACC CCATGGAGGAAACC SCACATCTACTCCTACT SCACATCTACTCCTACT TTTCTCAAGGGTTTTA ANTCAAACCACGAGAG SGGACTAAACACGAA	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAATG CCCCCCAATG CGCACGACCC GGCTGCACGAC CTGGATGTGA CTGGATGTCACC AACATGGAA GCAAGACTGG	CAGAGACTGAGC CTGCAGCCCAAG TCCAAGTGCTT TTCCAAGGGGA ACGACTCAAGGG CTTGAGTTCCAG CTTGAGTTCCAG CCTAGATTCCAG CCCAAGACAAATCCA TGTGTCCAGCCC AACATCAATCAA CGGAGCAGAAGC	ACCTGTCCTCCGG (AAGAGTCCAGCO (CGACAGCGTCGT) (CAAGTTATCAT (CTATGGACAGA (AATCATTCCTCG) (ATCAAGTTTCTCG) (CAAGGGAGATTCTC (CACAGCTCCGG) (CGAGGGAGAGCC) (CGTGCTGGACGA)	CCTOBCCTTOCTRAGAGE CCCACGCCGCCTACACCA CCCCGCACCCCGCCACCAC CCCGCACCCCGGAAGCCATAT GCATOTGCCGGGAAGCCATAT GCATOTGCCGGGAAACACAC ANTGCCCGGGAAACACCTC ATGCTCGAAAATTACCTGG GGATOTGATTGATGAAGG ACCCACACATCTCCTCA TOTATCTGGTGGACCTGGC GGCAAAGAATTCCACAC	F FAS F FAS F rec with N on F sub	ords w n a ">" that sa seque	here e and he me lin nt line	each record eader infor ie. Each	l begins mation	- - - 1
CTTTCCTTGGAGCA GCGGGGACCAACAC GCGGGGACCAACA TTTTGGGGGGGACCAACT TTTTGACCTTGTT CAACATCTACCTTGT CACACATCTACCCAT TTTCTCAACGCTTGTA UACATCAACGCGGTGTA UACAATCAACGCACGG GGGACTACAACGCACGG GGGACTACAACGCCCA	CACACCACCO GAATGTAGCA TCATCCCCAA CCCCCCAAACG GGCTACAATG TGCACGACCC GGATGAGAAC CTGGATGTGAACG CTGGATGTGAACG GGCTGTCACC GGCTGTCACC CAATGTGAAC	CAGAGACTGAGC CTGCAGGCGAGC TCCAAGGGGAC TTTCCAAGGGGA ACGACTCAAGAG GGCACCATTTTG CTGGGCATCGAG CCTGGAGTTCCAC CCAAGGAAGCATCCAC CCAAGACAACCAC CCGCACCAGACC TCCGCACGGCC	NCCTGTCCTCCGG NAGAGTCCGGCC (CGACAGCGCGTT (CAAGCGTCGTT (CTATGTGACAGAC INTCAAGGTTTTCT ITCTCGTGCACGG INTCAAGGTTCTCG (CGAGGACATTCT (CACAGCTCTCGG) ITCATGGCAGCG (GAGGCACTAAJ)	CCTOBCCTTOCTGAGAG CCCAGCCGCGCTACCACCA CCCCGCACCCAGCTOACAT TATTGCGGCGAAGCCATAT GCATGTCCTCAGGGAAACCATAT AATTGCCCGGAAACCAC TACTTCAGAATTTACCTGG AGCCAAGACCCGGCTCCT GGATGTGATTGATGAAGGG ACCACACATCTTCCTCA TGTTCTGCGACTGCCC	F FAS F rec with a on sub	ords w n a ">" that sa seque juence	here e and he me lin nt line inforn	each record eader infor e. Each of the reconnation.	d begins mation ord is	
CICTCCCTCGGACCA GCGGGACCAACAC TCCGGGGACCAACT TTTTGCGGGGACCAACT TTTTGCGGGGGAACCA CACACTCTTGCT CACACTCCTTGCT CACACTCCTGCCTTC TTCCCACGGCTCTA AAATTCCAGCGCTCTA CAACTCAAGCACGGC CACACACCACCACGTCTA CACCTCCACCTCGGG CACCAAAATCACACAG CACCACCACGTTATA	CACACCACCO GAATGTACCA TCATCCCCAT CCCCCCAAAC GGCTACAATG GGCTACAATG CTGGATGTGA CTGGATGTGA CTGGATGTGAA GCAAGACTG GAAGACTGGAA GCAAGACTGG CAATGCAGGA	CIGADACTAGE (CTGCAGCCAAG (CTGCAGCCAT (CTGAGCCATGCAT (CTCAAGGCAAGGA (GCACCATTTTG (CCAAGCAAAGA (CCAAGCAAAGCA (CCAAGCAAAGCA (CCGAGCACAAGCA (CCGAGCACAGCA (CCGACGACGAC (CCCCACGGCA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCGGGA (CCCCCCCGGGA (CCCCCCCGGGA (CCCCCCCGGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCGGA (CCCCCCCCGGA (CCCCCCCCGGA (CCCCCCCCGGA (CCCCCCCCCGGA (CCCCCCCCGGA (CCCCCCCCCGGA (CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	INCTOPECTOCS INCOMPTOGED INCOMPTOGED INCOMPTOS INCOMPTO I	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	F FAS rec with on sub sec	ords w n a ">" that sa seque juence	here e and he me lin nt line inforn	each record eader infor e. Each of the reconnation.	d begins mation ord is	
CTCTCCTCTGGAGCA SIGCGGAGCAACAAC TTTTTGACCGTGATT TTTTGACCGTGTTTT LCATGGAGGGAAAGC LCACATCTGACGTATA NARATTCGTGACGTTTA NARATTCGTGACGTTTA NATCAAACGACAGGA GGAGTGAGAAGCAGGA GGCGACAACGACGTGTA NACCAACATGACAAGG CAACATCGACTCAGTTATA	CACACCACCO GAATGTAGCA TCATCCCCAT CCCCCCAAAC GGCTACAATG TGCACGACCC GGATGAGAAC CTGGATGTGAC CTGGATGTCACC CTGGATGTCACC GATGTGGATC ATCATCCAGA ANATTTGGAC	CAGAGACTGAGC CTCAAGCCAAG TCCAAGCCAT TTCCAAGCCAT TTTCCAAGGCA ACGACCATTTG TCCACCTGATGGG CCTAGACTAAGA CCCAAGACAAAGC AACATGAAAGAAAGC AACATGAAGAGACA CCCGACTGCAGG GACCAAGTCCAC	NOCTOTOCOGO ANAGATOCOGOC (CORARGOTOGT) (CANOTTATOTO (CANOTTATOTO (CANOTTATOTO (CANOTTATOTO TOCOGOCOGO (CANOTTTOCOGO (COTOCTOGOCOM (CANOCTANA) (CANOCTANA) (CANOCTANA) (CANOCTANA)	CECCOCCEPTECTCALABOR CECNOCCECCENTRACCACCA CECNOCCECCENTRACCACCA CECTOTACACACACETRAGA CENTERCONCECTATA CENTERCANACCA ANTERCECCALACACETER CENTERCANACCACA CENTERCANACCACA CENTERCANACCACA CENTERCANACCACA CENTERCANACCACA CENTERCANACCACACA CENTERCANACCACACA CENTERCANACCACACA CENTERCANACCACACA CENTERCANACCACACA CENTERCENTRACCACACA CENTERCENTRACCACACA CENTERCENTRACCACACA CENTERCENTRACCACACACACA CENTERCENTRACCACACACACACA CENTERCENTRACCACACACACACACACACACACACACACACACACACA	F FAS F rec with Sub S Sec Thi	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each recorc eader infor le. Each of the reconnation. commonly	d begins mation ord is v used	
ICTETCOTTOGICADA SIGCIOMANCANCANCANC SIGCIOGUALACANOT TITTTOCOCOLACICATO TITTTOCOCOLACICOTTOCT COCONCOCTITOCT NAMITICOTINICOTTO NAMITICOTINICOTTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOTICACTO NACANTCOCICA	CACACCACCC GAATGTACCA TCATCCCCAT CCCCCCAAAC GGCTACAACG GGCAGAACC CTGGACGACG CTGGACGACG CTGGACGACG CTGGACGACGAAA GCAAGACTGG ATCCCCAGG ATCATGCAGA ANATTCGCAG AAAGTGCAGA	CRADACTAGE (CTGCAGCCAAG (CTGCAGCCAAGGAA ACCAACTCAAGGAC ACCAACTCAAGGAC (CTGAGTGCAAGGAC (CTGAGTGCAGCCC AACATCAAGGAAC (CGGAGCACAAGGACC TCCGCACGGACG (TCGCGACGAGTCCAC (TTGCGAACCTGG	ACCTORCECCGG INAGNOTCCCAGC (COALDAGCTCOT) (COALDAGCTCOT) (COALDAGCTCOT) (COALDAGCTCOC) (CAACOTTCOT) (COACOGCAGCACCAC) (CAACOTTCOT) (CACAGCTCACCAC) (COCCAACACCACAC) (COCCAACCACCACAC) (COCCAACCACCACAC) (COCCAACCACAC) (CACAGCCAACCAC)	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	FF FAS FF FEC with on sub FF Sec Thi by	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each record eader infor e. Each of the reconnation.	d begins mation ord is v used	
CECTCOLTOGAGOS CEGGICIACICANONA CEGGICIACICANONA CEGGICIACICANON CEGICIACICANONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIACICIONA CENTOCIONA CENT	CACACCACCO DATOTACA DETACOCCATA DECECCAAAO SCTACACAT DECECCAAAO SCTACACATO DECACACCO GGATGAGAAO TEGAACATO SGCTOTCACCA ACATGGAAA CAAGTGCAGA CAATGTGATCA AAGGACAGA AAGGACAGACA AAGGCACGGC	CIGADACTAGE (CTGCAGCCAGE (CTGCAGCCAGE (CTGAGTCCTTT (CTCAAGTGCAAGGGA (CTGAGTCCACCACATTTTG (CCCAAGACAATGCA (CCCAAGACAATGCAC (CCCAAGCACACCACCACCACCACCACCACCACCACCACCA	IACTORECTEORS IACADATECTORS (COACAGOTOGT) (CANOTTATCATO (CTATGACAGA (CANOTTATCATO (CTATGACAGA) (CANOTTATCATO (CANOTTATCATO (CANOTTATCATO (CANOTACACA) (CANOTACACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COTGATGACAGA) (COCCAGAGACAGA)	CCCCGCCCCCTCTCCCAAAG CCCCCAACCAGCCCAAAC CCCCTAACCAGCCCAAAC CCCCTAACCAGCCCAAAC TATTGGGGGGAACAACAAC CCCCCAACACAAACAAC CACCAACAACAA	FAS Free with Sub Free Sub Free Sub Free Sub Free Sub Free Sub Free Sub Free Free Free Free Free Free Free Fre	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each recorc eader infor le. Each of the reconnation. commonly	d begins mation ord is v used	
CECTCOTTOGAGEM GEORANCACONTOCOTTOGAGEMA TECESOGIALCANOR TECESOGIALCANOR CONTOCASOCIAL AND CONTOCASOCIAL AND CONTOCASOCIAL AND CONTOCASOCIAL AND CONTOCASOCIAL AND CONTOCASOCIAL CANON CONTOCASOCIAL CANON CONT	CACACCACCO CAACOTACCA TRATECCAA SGCTACAATG TCCCCCCCAAAG SGCTACAATG TCCCACGACC SGCTACAATG TCCCAGG TCGAATGTGA TCGAATGGAA AACATGGAAA AACATGGAAA AAGGAATCGAC AAGGATCGAC AACGCTGGC	CLADAGACTAGC CITACACCCAMA CITACACCCAMA CLAAGGCACAAGGA ACCACTCAAGGA CCAAGCACATITTC CCACAGCAATGATG CITGAGTGCAACCC MACATGATGCAA CGACGACACAGAC MCCATGCTGGAC MCCTCCGACGACGAC CCCAGCCAGCCA MCCATGCTGGAC MCCTCCGACGACGAC TTCCGAACGACC TTCCGAACGACC TTCCGAACGACC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCAC TCCGAACGACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCAC TCCCAACCACCACCACCAC TCCCAACCACCACCACCAC TCCCAACCACCACCACCAC TCCCAACCACCACCACCACCAC TCCCAACCACCACCACCACCACCACCACCACCACCACCAC	ACTORECTOCOGO INGENERACIÓN INGENERACIÓN INGENERACIÓN INTENTICATO INTENTICATO INTENTICATO INTENTICATO INTENTICATO INTENTICATO INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INGENERACIÓN INTENTIC	CCTCGCCCTCTCTCALABA CCCACCCTCALACCAGCCTCALADT TATTGGGGGCAACCAGCCTCALADT TATTGGGGGCAACCACAT CACCCTCCTCGGGCAAACCACA AATTGCCCCAGCACACAT AATTGCCCCAGCACACAT AATTGCCCACACACACAT AATTGCCCACACACACACACA AATTGCCCACACACACACACA ACTACACACACACACACACACAC	FAS Free FAS Free with a on sub Free Sec Thi by	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each recorc eader infor le. Each of the reconnation. commonly	d begins mation ord is v used	
CECTCOLTOGAGOS GEGGIANCANONA CEGGIANCANONA CEGGIANCANOT CEGNARATOCITOTO NAMATROCITATO NAMATROCITATO NACANANCOLATO NACA	CACACCACC CAACCACCC CAACCACACCACACCACA		IACTORECTECCS IACADATECTORE (COADATECTORE) (COADATECTORE) (CANOTITATCATE (CITATEGACAGA) (CANOTITATCATE (CITATEGACAGA) (CANOTECCC) (CANOTECCC) (CANOTECCC) (COTORTOGACAC) (COTORTOGACAC) (COTORTOGACAC) (COTORTOGACAC) (COTORTOGACAC) (COTORTOGACAC) (COTORTOGACC) (COTORTOGACC) (COCORGACCTON (COCORGACC) (COCORGACCTON (COCORGACC) (COCORGAC) (COCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCOCO	CCCCGCCCCCTCTCCCAAAG CCCCCAACCAGCCCAAAC CCCCTAACCAGCCCAAAC CCCCTAACCAGCCCAAAC TATTGGGGGGAACAACAAC CCCCCAACACAAACAAC CACCAACAACAA	FAS rec with on sub sec Thi	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each recorc eader infor le. Each of the reconnation. commonly	d begins mation ord is v used ograms	
	CACACCACC CACACOACCA CALCONACA CONTRACCACA CONTRACCACA CONTRACTOR CONTRACTOR CONTRACTOR CONTRACTOR CANONACTOR CANONACTOR CANONACTOR CANONACTOR CANONACTOR CONTRACTON		ACTOPECTOCO INCOMPTOCIDACIÓN (CACADOCTOCITICA CICACADOCTOCITICA CICACADOCTOCITICA CICACADOCTOCITICA INTENTOCICO (CACADOCTOCO CONTINUE INTENTOCICO (CACADOCTOCO (CACADOCADACIÓN (CACADOCTOCO (CACADOCADACIÓN (CACADOCTOCO (CACADOCTOCO (CACADOCTOCO (CACADOCADACIÓN) (CACADOCADACIÓN (CACADOCADACIÓN) (CACADOCADACIÓN (CACADOCADACIÓN) (CACADOCADACIÓN (CACADOCADACIÓN) (CACADOCADACIÓN (CACADOCADACIÓN) (CACADOCADACIÓN (CACADOCADACIÓN) (CACAD	CCCOCCCCTCTCCCAAAAA CCCACCAACCAACCCAACAA CCCCTCAACCAAC	FAS rec with on sub sec by	ords w n a ">" that sa seque juence s form	here e and he me lin nt line inform	each record eader infor le. Each of the reconnation. commonly nalysis pro	d begins mation ord is v used ograms	1 1 1 1

GenBank 'graphics' sequence record

🔺 🕨 🖉 🖄 🏦 🕂 🗧 www.ncbi.nlm.nih.gov/nuccore/45446748?report=graph		C Reader
NM_004984.2: Homo sapiens kinesin family member 5A (KIFSA),	mRNA	
S NCBI Resources 🖸 How To 🗹		Sign in to NCI
Nucleotide :		Search
Limits Advanced		He
Display Settings: Graphics	Send: 🖂	
		Analyze this sequence
Homo sapiens kinesin family member 5A (KIF5A), mRNA		Run BLAST
NCBI Reference Sequence: NM_004984.2		Pick Primers
GenBank FASTA		Highlight Sequence Features
Link To	This Page Feedback	
	3,400 3,8\$	Articles about the KIF5A gene
		α-Synuclein oligomers impair neuronal
	oomiguro ne :	microtubule-kinesin interp [J Biol Chem. 201
1 208 499 688 988 1.K 1.208 1.409 1.609 2.K 2.208 2.409 2.698 2.698 3.K 3.209		Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 201:
		Systems-wide analysis of ubiquitylation dynamics reveals a key rc [Nat Cell Biol. 201
		dynamics reveals a key rt [Nat Cell Biol. 201
		dynamics reveals a key rt [Nat Cell Biol. 201
	*	dynamics reveals a key rt [Nat Cell Biol. 201 See all
	*	dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene
eon eon <td>*</td> <td>dynamics reveals a key rr [Nat Cell Biol. 201 See al Pathways for the KIF5A gene Peptide hormone metabolism</td>	*	dynamics reveals a key rr [Nat Cell Biol. 201 See al Pathways for the KIF5A gene Peptide hormone metabolism
eon eon <td>*</td> <td>dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation</td>	*	dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation
con eon eon <td>*</td> <td>dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation Dopaminergic synapse</td>	*	dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation Dopaminergic synapse
000 000 <td>×</td> <td>dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation Dopaminergic synapse</td>	×	dynamics reveals a key rr [Nat Cell Biol. 201 See all Pathways for the KIF5A gene Peptide hormone metabolism MHC class II antigen presentation Dopaminergic synapse

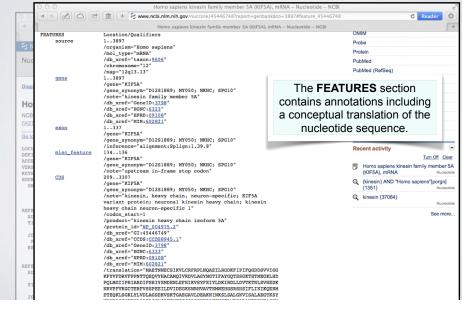
GenBank sequence record, cont.

00	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotid	e – NCBI II
< ► 🛃		🖒 Reader 🔘
	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotide – N	CBI +
NCBI Re	sources 🖸 How To 🖂	Sign in to NCBI
Nucleotide	Nucleotide : (KIF5A) AND "Homo sapiens"	Search
100000000	Limits Advanced	Help
Display Settin	as:⊡ GenBank	Send: 🖂
		Change region shown
Homo sa	piens kinesin family member 5A (KIF5A), mRNA	
NCBI Referen	ce Sequence: NM_004984.2	Customize view
FASTA Grap	hics	Customize view
<u>3o to:</u> ⊡		
Locus	NM_004984 3897 bp mRNA linear PRI 10-JAN-2014	Analyze this sequence
DEFINITION ACCESSION	Homo sapiens kinesin family member 5A (KIF5A), mRNA. NM 004984	Pick Primers
TERSION	NM_004984.2 GI:45446748	Highlight Sequence Features
KEYWORDS SOURCE	RefSeq. Homo sapiens (human)	Find in this Sequence
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;	Articles about the KIF5A gene
REFERENCE	Catarrhini; Hominidae; Homo. 1 (bases 1 to 3897)	α-Synuclein oligomers impair neuronal
AUTHORS	Kawaguchi,K. Role of kinesin-1 in the pathogenesis of SPG10, a rare form of	microtubule-kinesin interp [J Biol Chem. 2013]
	hereditary spastic paraplegia	Molecular motor KIF5A is essential for GABA(A) receptor transport, a [Neuron. 2012]
JOURNAL	Neuroscientist 19 (4), 336-344 (2013) 22785106	Systems-wide analysis of ubiguitylation
REMARK	GeneRIF: A review of the mechanism of pathogenesis involved in	dynamics reveals a key rr [Nat Cell Biol. 2012]
	spastic paraplegia type 10 when KIF5A is inactivated by mutations. Review article	See all
AUTHORS	2 (bases 1 to 3897) Prots, I., Veber, V., Brey, S., Campioni, S., Buder, K., Riek, R.,	
	Bohm,K.J. and Winner,B.	Pathways for the KIF5A gene
TITLE	alpha-Synuclein oligomers impair neuronal microtubule-kinesin interplay	Peptide hormone metabolism
	J. Biol. Chem. 288 (30), 21742-21754 (2013)	

GenBank sequence record, cont.

0.0	Homo sapiens kinesin family member 5A (KIF5A), mRNA - Nucle	eotide - NCBI
0.04	🔺 🕨 🖒 🖄 🏦 🕂 🗧 www.ncbi.nlm.nih.gov/nuccore/45446748?report=genbank&to=3897#sequ	uence_45446748 C Rea
	Homo sapiens kinesin family member 5A (KIF5A), mRNA – Nucleotid	e – NCBI
FEA	/gene synonym="D12S1889: MY050; NKHC; SPG10"	
	/standard_name="D12S1889"	
NCBI	/db_xref="UniSTS: <u>48006</u> "	
	ORIGIN	
ucleotic	1 acgeccaggt egecegeate cegetgeege aggagagaga cagegegeee eggecetget 61 ecceaggett egeceggeeg eccteaaete tgteeeeaga gaetgageae etgteeteeg	
	61 ccccaggett cgcccgggeg ccctcaactc tgtcccccaga gactgagcac ctgtcctccg 121 cctcggectc tgctgagage cctctcctct ggagcacaca ccacccctge ageccaagaa	
	121 ceteggeete tgetgagage eeteteetet ggageacaa eeaeeetge ageecaagaa 181 gagteecage eecaegeegg etaceaceat ggeggagaee aacaaegaat gtageateaa	
	241 ggtgetetge cealgeegg etaleaceat ggeggagate aacaaegaat gtageateaa 241 ggtgetetge egatteegge eeetgaacea ggetgagatt etgeggggag acaagtteat	
play Set	301 coccatttto caaggggacg acagogtogt tattgggggg aagcoatatg tttttgacog	
	361 tgtattcccc ccaaacacga ctcaagagca agtttatcat gcatgtgcca tgcagattgt	The estuel seguence
	421 caaagatgtc cttgctggct acaatggcac cattttgct tatggacaga catcctcagg	The actual sequence
omo :	481 gaaaacacat accatgqagg gaaagctgca cgaccctcag ctgatgggaa tcattcctcg	and a standa after the
B Refer	541 aattgcccga gacatettea accacateta etecatggat gagaacettg agttccacat	entry starts after the
	601 caaggtttct tactttgaaa tttacctgga caaaattcgt gaccttctgg atgtgaccaa	
TA G	661 gacaaatotg toogtgcacg aggacaagaa cogggtgcca tttgtcaagg gttgtactga	word ORIGIN
	721 acgetttgtg tecageeegg aggagattet ggatgtgatt gatgaaggga aateaaateg	
	781 tcatgtggct gtcaccaaca tgaatgaaca cagctctcgg agccacagca tcttcctcat	
	841 caacatcaag caggagaaca tggaaacgga gcagaagotc agtgggaago tgtatotggt	
16	901 ggacctggca gggagtgaga aggtcagcaa gactggagca gagggagccg tgctggacga	
NITIO	961 ggcaaagaat atcaacaagt cactgtcagc tctgggcaat gtgatctccg cactggctga	
SSION	1021 gggcactaaa agctatgttc catatcgtga cagcaaaatg acaaggattc tccaggactc	
SION	1081 tetegggggga aactgeegga egaetatgtt eatetgttge teaceateea gttataatga	
WORDS	1141 tgcagagacc aagtccaccc tgatgtttgg gcagcgggca aagaccatta agaacactgc	
RCE	1201 ctcagtaaat ttggagttga ctgctgagca gtggaagaag aaatatgaga aggagaagga	
JRGANIS	1261 gaagacaaag gcccagaagg agacgattgc gaagctggag gctgagctga	
	1321 caatggagag aatgtgcctg agacagagcg cctggctggg gaggaggcag ccctgggagc	
	1381 cgagetetgt gaggagaece etgtgaatga caacteatee ategtggtge geategegee	
PERENCE	1441 cgaggagogg cagaaatacg aggaggagat cogcogtoto tataagcago ttgacgacaa	
AUCHORS	1501 ggatgatgaa atcaaccaac aaagccaact catagagaag ctcaagcagc aaatgctgga	
FIFLE	1561 ccaggaagag ctgctggtgt ccacccgagg agacaacgag aaggtccagc gggagctgag	
	1621 ccacctgcaa tcagagaacg atgccgctaa ggatgaggtg aaggaagtgc tgcaggccct	
JOURNAL	1681 ggaggagotg gotgtgaact atgaccagaa gtoccaggag gtggaggaga agagccagca	
PUBMED	1741 gaaccagett etggtggatg agetgtetea gaaggtggee accatgetgt eeetggagte	
EMARK	1801 tgagttgcag cggctacagg aggtcagtgg acaccagcga aaacgaattg ctgaggtgct	
	1861 gaacgggctg atgaaggatc tgagcgagtt cagtgtcatt gtgggcaacg gggagattaa	
	1921 gctgccagtg gagatcagtg gggccatcga ggaggagttc actgtggccc gactctacat	
ERENCE	1981 cagcaaaatc aaatcagaag tcaagtotgt ggtcaagcgg tgccggcagc tggagaacct	
AUTHORS	2041 ccaggtggag tgtcaccgca agatggaagt gaccggggcgg gagctctcat cctgccagct 2101 cctcatctct cagcatgagg ccaagatccg ctcgcttacg gaatacatgc agagcgtgga	
	2101 cotcatotot cagcatgagg coaagatoog otogottacg gaatacatgo agagogtgga 2161 gotaaagaag oggcacotgg aagagtoota tgactoottg agogatgago tggocaagot	
TITLE	2221 ccaggcccag gaaactgtgc atgaagtggc cctgaaggac aaggagcctg acactcagga	
	2281 tgcagatgaa gtgaagaagg ctctggagct gcagatggag agtcaccggg aggcccatca	
JOURNAL	2341 cooreagetya goograactoo graecagat caacgatya aytootootoa	

GenBank sequence record, cont.



RefSeq: NCBI's Derivative Sequence Database

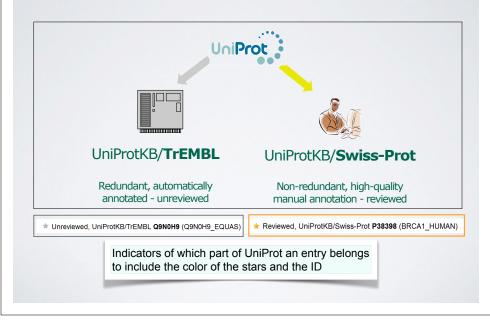
- RefSeq entries are hand curated best representation of a transcript or protein (in their judgement)
- Non-redundant for a given species although alternate transcript forms will be included if there is good evidence
 - Experimentally verified transcripts and proteins accession numbers begin with "NM_" 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



UniProt: Protein sequence database

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

- UniProt comprises four databases:
- UniProtKB (Knowledgebase) Containing <u>Swiss-Prot</u> and <u>TrEMBL</u> components (these correspond to hand curated and automatically annotated entries respectively)
- 2. UniRef (Reference Clusters)

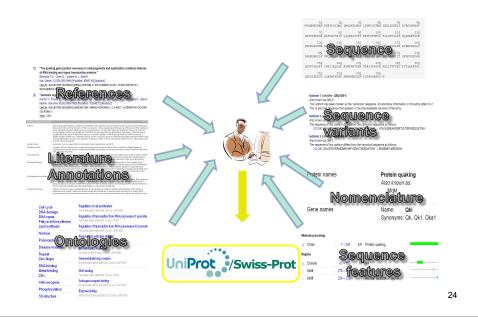
Filtered version of UniProtKB at various levels of sequence identity

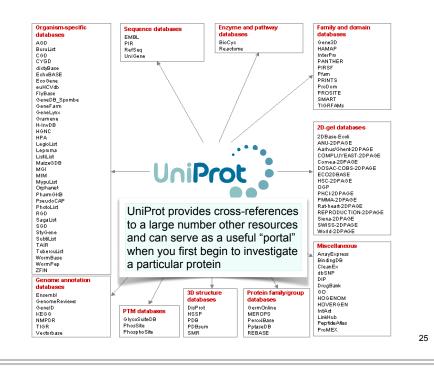
e.g. <u>UniRef90</u> contains sequences with a maximum of 90% sequence identity to each other

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

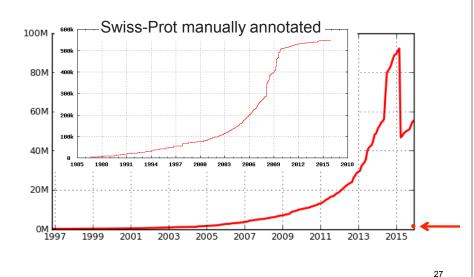
4. UniMES (Metagenomic and Environmental Sequences)

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





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



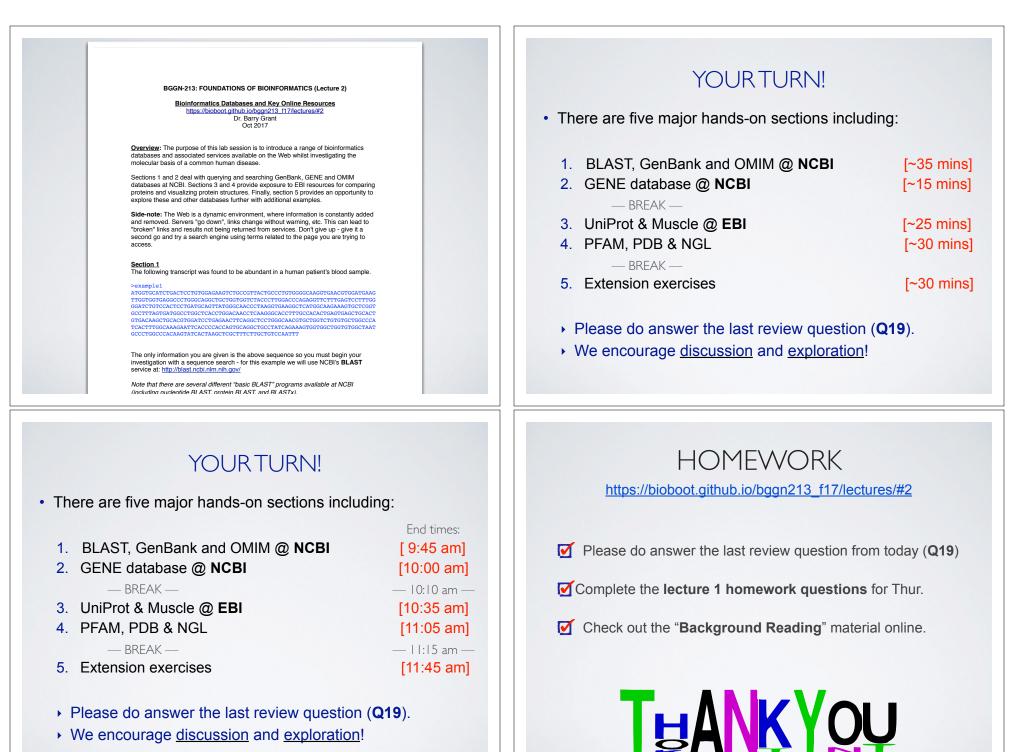
UniProt/Swiss-Prot vs UniProt/TrEMBL

- <u>UniProtKB/Swiss-Prot</u> is a non-redundant database with one entry per protein
- <u>UniProtKB/TrEMBL</u> 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

Your Turn!

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

	https://bioboot.github.io/bggn213_f17/lectures/#2	1 D
Hom	e Gmail Gcal Bitbucket GitHub News V Disqus BGGN-213	+
UC San Diego	 Be able to describe how nucleotide and protein sequence and structure data are represented (FASTA, FASTQ, GenBank, UniProt, PDB). 	
BGGN 213	 Be familiar with online tools at the EBI and NCBI including Muscle and BLAST. 	
A hands-on introduction to the computer-based analysis of genomic and biomolecular data from the Division of	 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. 	
Biological Sciences, UCSD 🗵.	Material:	
Overview	Lecture Slides: Large PDF, Small PDF,	
Lectures	Handout: Major Bioinformatics Databases	
	Hands-on section worksheet	
Computer Setup	Muddy point assessment	
Learning Goals Assignments & Grading	Homework:	
Assignments & Graumt		



- Please do answer the last review question (Q19).
- We encourage <u>discussion</u> and <u>exploration</u>!