35344659 35344424 35344176 35343807 35343785 35343771 35343620 35342821 35342158 Retrieve abstracts from articles found efetch: returns abstracts of articles in a simple text format Here we return abstracts from first two articles In [4]: | mshandle = Entrez.efetch(db="pubmed", id=record["IdList"][0:2], rettype="abstract", retmode="text") print(mshandle.read()) 1. Antiviral Res. 2022 Mar 27:105299. doi: 10.1016/j.antiviral.2022.105299. [Epub ahead of print] Reliable quantification of Cytomegalovirus DNAemia in Letermovir treated Weinberger S(1), Steininger C(2). Author information: (1)Division of Infectious Diseases and Tropical Medicine, Department of Medicine I, Medical University of Vienna, Vienna, Austria. (2)Division of Infectious Diseases and Tropical Medicine, Department of Medicine I, Medical University of Vienna, Vienna, Austria. Electronic address: christoph.steininger@meduniwien.ac.at. Polymerase chain reaction (PCR) based methods are a fast and sensitive approach to detect and monitor viral load in Cytomegalovirus (CMV) patients. Letermovir (LMV) acts at a late stage during the CMV replication cycle and does not inhibit CMV DNA replication per se. Therefore, quantitative nucleic acid amplification testing might lead to the overestimation of viral load in patients treated with LMV and underestimate treatment success. To study this discrepancy, we treated infected cells with LMV or Ganciclovir (GCV) and compared viral progeny DNA levels. Prior to nucleic acid extraction and qPCR measurements we pretreated cell lysates and cell culture supernatants from infected cells with DNase I. This step assumes the degradation of DNA which is not protected from a viral capsid. LMV treatment did not reduce genomic copies (GC) in samples from whole cell lysates compared to samples treated with GCV. DNase treatment prior to DNA extraction, decreased GC in the LMV treated group to comparable levels as seen in the GCV group. In cell culture supernatants, LMV or GCV treatment led to an equivalent reduction of CMV GC. In this case, DNase treatment exerted a negligible effect on both groups. We conclude that the accumulation of concatemeric DNA within cells seems to be a confounding variable when monitoring LMV efficacy via qPCR. However, qPCR shows to be a reliable method to evaluate antiviral efficacy of LMV in cell free specimens. These results have strong clinical implications for the monitoring of CMV therapy during LMV treatment. Copyright © 2022. Published by Elsevier B.V. DOI: 10.1016/j.antiviral.2022.105299 PMID: 35354065 Conflict of interest statement: Declaration of competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. 2. Hum Genet. 2022 Mar 30. doi: 10.1007/s00439-021-02333-9. [Epub ahead of print] Genetic testing for pediatric hearing loss: no time to waste. Kenna MA(1). Author information: (1)Sarah Fuller Chair for Hearing Loss and Hearing Restoration, Department of Otolaryngology and Communication Enhancement, Boston Children's Hospital, Harvard Medical School, 300 Longwood Ave. BCH 3129, Boston, MA, 02115, USA. Margaret.Kenna@childrens.harvard.edu. Since the discovery of the first human deafness gene a quarter of a century ago, our approach to clinical evaluation of children with hearing loss has changed dramatically. What was once a low-yield scattershot approach has changed to a clearly definable pathway involving genetic testing, imaging, and congenital cytomegalovirus testing. There still however is a great deal of work to be done to expand the correct use of this testing, particularly genetic testing. © 2022. The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature. DOI: 10.1007/s00439-021-02333-9 PMID: 35353226 Retrieve nucleotide sequences for rhodopsin from Homo Sapiens esearch returns results in XML format In [5]: handle = Entrez.esearch(db = "nucleotide", term = "rhodopsin AND Homo sapiens[orgn]") record = Entrez.read(handle) print(record['IdList']) Entrez.efetch(db = "nucleotide", id = record["IdList"]) ['1675034719', '1653962141', '1523964325', '1520687782', '1519473570', '1519315165', '1519313919', '1 519244045', '1062594141', '1236508655', '219281886', '1783531726', '1677539460', '1677530135', '20759 19630', '2068205929', '1519312296', '1918095546', '1654220746', '1654220743'] Out[5]: <\_io.TextIOWrapper encoding='UTF-8'> We point [org] after species name to specify that we want to search for requested protein directly in *H.sapiens* (and not in other organisms connected with it, e.g microorganims inhabiting it). Find ID of an organism (Mus Musculus) in database In [6]: | handle = Entrez.esearch(db = "taxonomy", term = "Mus musculus") record = Entrez.read(handle) print(record) {'Count': '1', 'RetMax': '1', 'RetStart': '0', 'IdList': ['10090'], 'TranslationSet': [], 'TranslationStack': [{'Term': 'mus musculus[All Names]', 'Field': 'All Names', 'Count': '1', 'Explode': 'N'}, 'G ROUP'], 'QueryTranslation': 'mus musculus[All Names]'} In [7]: print('Mus Musculus ID:', record['IdList'][0]) Mus Musculus ID: 10090 Retrieve nucleotide sequences for rhodopsin from H. Sapiens and return results in a table (not XML) esearch + esummary : makes a request in a database by gene name and then returns a table with UID (in XML it corresponds to Id), accession number (in XML it corresponds to Caption) and sequence length (Slen). We retrieve here all PMIDs and then iterate through them to obtain some basic description In [8]: handle = Entrez.esearch(db="protein", term="rhodopsin AND Homo sapiens[orgn]") record = Entrez.read(handle) with open("results/id\_capt\_len\_py.txt", "w") as ouf: for rec in record["IdList"]: temphandle = Entrez.read(Entrez.esummary(db="protein", id=rec, retmode="text")) ouf.write(f"{temphandle[0]['Id']}\t{temphandle[0]['Caption']}\t{temphandle[0]['Length']}\n") In [9]: | temphandle Out[9]: [{'Item': [], 'Id': '166362740', 'Caption': 'NP\_001983', 'Title': 'proteinase-activated receptor 1 is oform 1 precursor [Homo sapiens]', 'Extra': 'qi|166362740|ref|NP\_001983.2|[166362740]', 'Gi': Integer Element(166362740, attributes={}), 'CreateDate': '1999/03/19', 'UpdateDate': '2022/03/12', 'Flags': I ntegerElement(512, attributes={}), 'TaxId': IntegerElement(9606, attributes={}), 'Length': IntegerEle ment(425, attributes={}), 'Status': 'live', 'ReplacedBy': '', 'Comment': ' ', 'AccessionVersion': 'N P\_001983.2'}] In [10]: ! cat results/id\_capt\_len\_py.txt 523704349 NP\_001265723 375 1043261393 NP\_001316355 74 439 1002819403 NP\_001307687 539848521 NP\_064445 364 523704347 NP\_005963 375 222080049 NP\_000858 481 88758606 NP\_001034674 297 52317263 NP\_001004723 307 11386179 NP\_068774 74 485837026 NP\_057726 3674 NP\_997253 609 282403488 153945858 NP\_002089 200 116008178 NP\_000532 405 NP\_775904 110556629 493 NP\_000946 38505194 402 31083344 NP\_064707 362 23312386 NP\_006230 753 7662494 NP\_054798 416 1062594142 NP\_001318005 374 166362740 NP\_001983 425 Retrieve nucleotide sequences in fasta format for requested gene Return sequences in a text format handle = Entrez.esearch(db="protein", term="rhodopsin AND Homo sapiens[orgn]") In [11]: record = Entrez.read(handle) print(Entrez.efetch(db="protein", id=record["IdList"], retmode="text", rettype="fasta").read()) >NP\_001265723.1 neuropeptide Y receptor type 4 [Homo sapiens] MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVMVFIVTSYSIETVVGVLGNLCLMCVTVRQK EKANVTNLLIANLAFSDFLMCLLCQPLTAVYTIMDYWIFGETLCKMSAFIQCMSVTVSILSLVLVALERH QLIINPTGWKPSISQAYLGIVLIWVIACVLSLPFLANSILENVFHKNHSKALEFLADKVVCTESWPLAHH RTIYTTFLLLFQYCLPLGFILVCYARIYRCLQRQGRVFHKGTYSLRAGHMKQVNVVLVVMVVAFAVLWLP LHVFNSLEDWHHEAIPICHGNLIFLVCHLLAMASTCVNPFIYGFLNTNFKKEIKALVLTCQQSAPLEESE HLPLSTVHTEVSKGSLRLSGRSNPI >NP\_001316355.1 guanine nucleotide-binding protein G(T) subunit gamma-T1 precursor [Homo sapiens] MPVINIEDLTEKDKLKMEVDQLKKEVTLERMLVSKCCEEVRDYVEERSGEDPLVKGIPEDKNPFKELKGG CVIS >NP\_001307687.1 5-hydroxytryptamine receptor 2B isoform 2 [Homo sapiens] MKQIVEEQGNKLHWAALLILMVIIPTIGGNTLVILAVSLEKKLQYATNYFLMSLAVADLLVGLFVMPIAL LTIMFEAMWPLPLVLCPAWLFLDVLFSTASIMHLCAISVDRYIAIKKPIQANQYNSRATAFIKITVVWLI SIGIAIPVPIKGIETDVDNPNNITCVLTKERFGDFMLFGSLAAFFTPLAIMIVTYFLTIHALQKKAYLVK NKPPQRLTWLTVSTVFQRDETPCSSPEKVAMLDGSRKDKALPNSGDETLMRRTSTIGKKSVQTISNEQRA SKVLGIVFFLFLLMWCPFFITNITLVLCDSCNQTTLQMLLEIFVWIGYVSSGVNPLVYTLFNKTFRDAFG RYITCNYRATKSVKTLRKRSSKIYFRNPMAENSKFFKKHGIRNGINPAMYQSPMRLRSSTIQSSSIILLD TLLLTENEGDKTEEQVSYV >NP\_064445.2 long-wave-sensitive opsin 1 [Homo sapiens] MAQQWSLQRLAGRHPQDSYEDSTQSSIFTYTNSNSTRGPFEGPNYHIAPRWVYHLTSVWMIFVVTASVFT NGLVLAATMKFKKLRHPLNWILVNLAVADLAETVIASTISIVNQVSGYFVLGHPMCVLEGYTVSLCGITG LWSLAIISWERWMVVCKPFGNVRFDAKLAIVGIAFSWIWAAVWTAPPIFGWSRYWPHGLKTSCGPDVFSG SSYPGVQSYMIVLMVTCCIIPLAIIMLCYLQVWLAIRAVAKQQKESESTQKAEKEVTRMVVVMIFAYCVC WGPYTFFACFAAANPGYAFHPLMAALPAYFAKSATIYNPVIYVFMNRQFRNCILQLFGKKVDDGSELSSA SKTEVSSVSSVSPA >NP\_005963.4 neuropeptide Y receptor type 4 [Homo sapiens] MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVMVFIVTSYSIETVVGVLGNLCLMCVTVRQK EKANVTNLLIANLAFSDFLMCLLCQPLTAVYTIMDYWIFGETLCKMSAFIQCMSVTVSILSLVLVALERH QLIINPTGWKPSISQAYLGIVLIWVIACVLSLPFLANSILENVFHKNHSKALEFLADKVVCTESWPLAHH RTIYTTFLLLFQYCLPLGFILVCYARIYRCLQRQGRVFHKGTYSLRAGHMKQVNVVLVVMVVAFAVLWLP LHVFNSLEDWHHEAIPICHGNLIFLVCHLLAMASTCVNPFIYGFLNTNFKKEIKALVLTCQQSAPLEESE HLPLSTVHTEVSKGSLRLSGRSNPI >NP\_000858.3 5-hydroxytryptamine receptor 2B isoform 1 [Homo sapiens] MALSYRVSELQSTIPEHILQSTFVHVISSNWSGLQTESIPEEMKQIVEEQGNKLHWAALLILMVIIPTIG GNTLVILAVSLEKKLQYATNYFLMSLAVADLLVGLFVMPIALLTIMFEAMWPLPLVLCPAWLFLDVLFST ASIMHLCAISVDRYIAIKKPIQANQYNSRATAFIKITVVWLISIGIAIPVPIKGIETDVDNPNNITCVLT KERFGDFMLFGSLAAFFTPLAIMIVTYFLTIHALQKKAYLVKNKPPQRLTWLTVSTVFQRDETPCSSPEK VAMLDGSRKDKALPNSGDETLMRRTSTIGKKSVQTISNEQRASKVLGIVFFLFLLMWCPFFITNITLVLC DSCNQTTLQMLLEIFVWIGYVSSGVNPLVYTLFNKTFRDAFGRYITCNYRATKSVKTLRKRSSKIYFRNP MAENSKFFKKHGIRNGINPAMYQSPMRLRSSTIQSSSIILLDTLLLTENEGDKTEEQVSYV >NP\_001034674.1 prostaglandin F2-alpha receptor isoform b precursor [Homo sapiens] MSMNNSKQLVSPAAALLSNTTCQTENRLSVFFSVIFMTVGILSNSLAIAILMKAYQRFRQKSKASFLLLA SGLVITDFFGHLINGAIAVFVYASDKEWIRFDQSNVLCSIFGICMVFSGLCPLLLGSVMAIERCIGVTKP IFHSTKITSKHVKMMLSGVCLFAVFIALLPILGHRDYKIQASRTWCFYNTEDIKDWEDRFYLLLFSFLGL LALGVSLLCNAITGITLLRVKFKSQQHRQGRSHHLEMVIQLLAIMCVSCICWSPFLGYRIILNGKEKYKV YEEQSDFLHRLQWPTLE >NP\_001004723.1 olfactory receptor 4N2 [Homo sapiens] MESENRTVIREFILLGLTQSQDIQLLVFVLVLIFYFIILPGNFLIIFTIKSDPGLTAPLYFFLGNLAFLD ASYSFIVAPRMLVDFLSAKKIISYRGCITQLFFLHFLGGGEGLLLVVMAFDRYIAICRPLHYPTVMNPRT CYAMMLALWLGGFVHSIIQVVLILRLPFCGPNQLDNFFCDVPQVIKLACTDTFVVELLMVFNSGLMTLLC FLGLLASYAVILCRIRGSSSEAKNKAMSTCITHIIVIFFMFGPGIFIYTRPFRAFPADKVVSLFHTVIFP LLNPVIYTLRNQEVKASMKKVFNKHIA >NP\_068774.1 guanine nucleotide-binding protein G(T) subunit gamma-T1 precursor [Homo sapiens] MPVINIEDLTEKDKLKMEVDQLKKEVTLERMLVSKCCEEVRDYVEERSGEDPLVKGIPEDKNPFKELKGG CVIS >NP\_057726.4 spectrin beta chain, non-erythrocytic 5 [Homo sapiens] MAGQPHSPRELLGAAGHRSRRPSTELRVPPSPSLTMDSQYETGHIRKLQARHMQMQEKTFTKWINNVFQC GQAGIKIRNLYTELADGIHLLRLLELISGEALPPPSRGRLRVHFLENSSRALAFLRAKVPVPLIGPENIV DGDQTLILGLIWVIILRFQISHISLDKEEFGASAALLSTKEALLVWCQRKTASYTNVNITDFSRSWSDGL GFNALIHAHRPDLLDYGSLRPDRPLHNLAFAFLVAEQELGIAQLLDPEDVAAAQPDERSIMTYVSLYYHY CSRLHQGQTVQRRLTKILLQLQETELLQTQYEQLVADLLRWIAEKQMQLEARDFPDSLPAMRQLLAAFTI FRTQEKPPRLQQRGAAEALLFRLQTALQAQNRRPFLPHEGLGLAELSQCWAGLEWAEAARSQALQQRLLQ LQRLETLARRFQHKAALRESFLKDAEQVLDQARAPPASLATVEAAVQRLGMLEAGILPQEGRFQALAEIA DILRQEQYHSWADVARRQEEVTVRWQRLLQHLQGQRKQVADMQAVLSLLQEVEAASHQLEELQEPARSTA CGQQLAEVVELLQRHDLLEAQVSAHGAHVSHLAQQTAELDSSLGTSVEVLQAKARTLAQLQQSLVALVRA RRALLEQTLQRAEFLRNCEEEEAWLKECGQRVGNAALGRDLSQIAGALQKHKALEAEVHRHQAVCVDLVR RGRDLSARRPPTQPDPGERAEAVQGGWQLLQTRVVGRGARLQTALLVLQYFADAAEAASWLRERRSSLER ASCGQDQAAAETLLRRHVRLERVLRAFAAELRRLEEQGRAASARASLFTVNSALSPPGESLRNPGPWSEA SCHPGPGDAWKMALPAEPDPDFDPNTILQTQDHLSQDYESLRALAQLRRARLEEAMALFGFCSSCGELQL WLEKQTVLLQRVQPQADTLEVMQLKYENFLTALAVGKGLWAEVSSSAEQLRQRYPGNSTQIQRQQEELSQ RWGQLEALKREKAVQLAHSVEVCSFLQECGPTQVQLRDVLLQLEALQPGSSEDTCHALQLAQKKTLVLER RVHFLQSVVVKVEEPGYAESQPLQGQVETLQGLLKQVQEQVAQRARRQAETQARQSFLQESQQLLLWAES VQAQLRSKEVSVDVASAQRLLREHQDLLEEIHLWQERLQQLDAQSQPMAALDCPDSQEVPNTLRVLGQQG QELKVLWEQRQQWLQEGLELQKFGREVDGFTATCANHQAWLHLDNLGEDVREALSLLQQHREFGRLLSTL GPRAEALRAHGEKLVOSOHPAAHTVREQLOSIQAQWTRLOGRSEQRRRQLLASLQLQEWKQDVAELMQWM

**Working with NCBI** 

{'Count': '53107', 'RetMax': '20', 'RetStart': '0', 'IdList': ['35354065', '35353226', '35352632', '35352508', '35351972', '35349757', '35348580', '35348198', '35347236', '35346077', '35345706', '353446 59', '35344424', '35344176', '35343807', '35343785', '35343771', '35343620', '35342821', '35342158'], 'TranslationSet': [{'From': 'cytomegalovirus', 'To': '"cytomegalovirus"[MeSH Terms] OR "cytomegalovir

us"[All Fields]'}], 'TranslationStack': [{'Term': '"cytomegalovirus"[MeSH Terms]', 'Field': 'MeSH Terms', 'Count': '22284', 'Explode': 'Y'}, {'Term': '"cytomegalovirus"[All Fields]', 'Field': 'All Fields', 'Count': '53110', 'Explode': 'N'}, 'OR', 'GROUP'], 'QueryTranslation': '"cytomegalovirus"[MeSH Te

Libraries

from Bio import Entrez

Some basic commands

record = Entrez.read(handle)

print(record)

esearch: finds an article in NCBI by our request

rms] OR "cytomegalovirus"[All Fields]'}

Pubmed IDs with articles on cytomegalovirus:

print(\*record['IdList'], sep='\n')

Entrez.email = "isemenov.bioinfo1998@gmail.com"

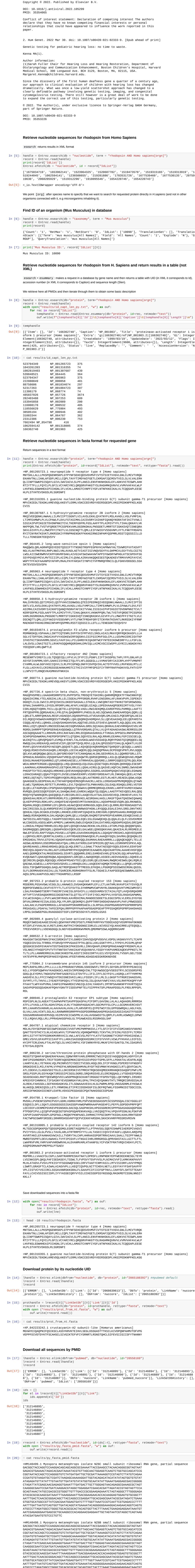
Find an article in NCBI connected with cytomegalovirus

By default RetMax = 20 => we retrieve only the first 20 articles from 52959 articles found

In [2]: | handle = Entrez.esearch(db='pubmed', term='cytomegalovirus')

In [1]:

In [3]:



EEKGLMAAHEPSGARRNILQTLKRHEAAESELLATRRHVEALQQVGRELLSRRPCGQEDIQTRLQGLRSK WEALNRKMTERGDELQQAGQQEQLLRQLQDAKEQLEQLEGALQSSETGQDLRSSQRLQKRHQQLESESRTLAAKMAALASMAHGMAASPAILEETQKHLRRLELLQGHLAIRGLQLQASVELHQFCHLSNMELSWVAEHM PHGSPTSYTECLNGAQSLHRKHKELQVEVKAHQGQVQRVLSSGRSLAASGHPQAQHIVEQCQELEGHWAE LERACEARAQCLQQAVTFQQYFLDVSELEGWVEEKRPLVSSRDYGRDEAATLRLINKHQALQEELAIYWS SMEELDQTAQTLTGPEVPEQQRVVQERLREQLRALQELAATRDRELEGTLRLHEFLREAEDLQGWLASQK QAAKGGESLGEDPEHALHLCTKFAKFQHQVEMGSQRVAACRLLAESLLERGHSAGPMVRQRQQDLQTAWS ELWELTQARGHALRDTETTLRVHRDLLEVLTQVQEKATSLPNNVARDLCGLEAQLRSHQGLERELVGTER QLQELLETAGRVQKLCPGPQAHAVQQRQQAVTQAWAVLQRRMEQRRAQLERARLLARFRTAVRDYASWAA RVRQDLQVEESSQEPSSGPLKLSAHQWLRAELEAREKLWQQATQLGQQALLAAGTPTKEVQEELRALQDQ RDQVYQTWARKQERLQAEQQEQLFLRECGRLEEILAAQEVSLKTSALGSSVEEVEQLIRKHEVFLKVLTA QDKKEAALRERLKTLRRPRVRDRLPILLQRRMRVKELAESRGHALHASLLMASFTQAATQAEDWIQAWAQ QLKEPVPPGDLRDKLKPLLKHQAFEAEVQAHEEVMTSVAKKGEALLAQSHPRAGEVSQRLQGLRKHWEDL RQAMALRGQELEDRRNFLEFLQRVDLAEAWIQEKEVKMNVGDLGQDLEHCLQLRRRLREFRGNSAGDTVG DACIRSISDLSLQLKNRDPEEVKIICQRRSQLNNRWASFHGNLLRYQQQLEGALEIHVLSRELDNVTKRI QEKEALIQALDCGKDLESVQRLLRKHEELEREVHPIQAQVESLEREVGRLCQRSPEAAHGLRHRQQEVAE SWWQLRSRAQKRREALDALHQAQKLQAMLQELLVSAQRLRAQMDTSPAPRSPVEARRMLEEHQECKAELD SWTDSISLARSTGQQLLTAGHPFSSDIRQVLAGLEQELSSLEGAWQEHQLQLQQALELQLFLSSVEKMER WLCSKEDSLASEGLWDPLAPMEPLLWKHKMLEWDLEVQAGKISALEATARGLHQGGHPEAQSALGRCQAM LLRKEALFRQAGTRRHRLEELRQLQAFLQDSQEVAAWLREKNLVALEEGLLDTAMLPAQLQKQQNFQAEL DASMHQQQELQREGQRLLQGGHPASEAIQERLEELGALWGELQDNSQKKVAKLQKACEALRLRRSMEELE NWLEPIEVELRAPTVGQALPGVGELLGTQRELEAAVDKKARQAEALLGQAQAFVREGHCLAQDVEEQARR LLQRFKSLREPLQERRTALEARSLLLKFFRDADEEMAWVQEKLPLAAAQDYGQSLSAVRHLQEQHQNLES EMSSHEALTRVVLGTGYKLVQAGHFAAHEVAARVQQLEKAMAHLRAEAARRRLLLQQAQEAQQFLTELLE AGSWLAERGHVLDSEDMGHSAEATQALLRRLEATKRDLEAFSPRIERLQQTAALLESRKNPESPKVLAQL QAVREAHAELLRRAEARGHGLQEQLQLHQLERETLLLDAWLTTKAATAESQDYGQDLEGVKVLEEKFDAF RKEVQSLGQAKVYALRKLAGTLERGAPRRYPHIQAQRSRIEAAWERLDQAIKARTENLAAAHEVHSFQQA AAELQGRMQEKTALMKGEDGGHSLSSVRTLQQQHRRLERELEAMEKEVARLQTEACRLGQLHPAAPGGLA KVQEAWATLQAKAQERGQWLAQAAQGHAFLGRCQELLAWAQERQELASSEELAEDVAGAEQLLGQHEELG QEIRECRLQAQDLRQEGQQLVDNSHFMSAEVTECLQELEGRLQELEEAWALRWQRCAESWGLQKLRQRLE QAEAWLACWEGLLLKPDYGHSVSDVELLLHRHQDLEKLLAAQEEKFAQMQKTEMEQELLLQPQELKPGRA GSSLTSFQWRPSGHQGLGAQLAETRDPQDAKGTPTMEGSLEFKQHLLPGGRQPSSSSWDSCRGNLQGSSL SLFLDERMAAEKVASIALLDLTGARCERLRGRHGRKHTFSLRLTSGAEILFAAPSEEQAESWWRALGSTA

>NP\_997253.2 probable G-protein coupled receptor 153 [Homo sapiens] MSDERRLPGSAVGWLVCGGLSLLANAWGILSVGAKQKKWKPLEFLLCTLAATHMLNVAVPIATYSVVQLR RQRPDFEWNEGLCKVFVSTFYTLTLATCFSVTSLSYHRMWMVCWPVNYRLSNAKKQAVHTVMGIWMVSFI LSALPAVGWHDTSERFYTHGCRFIVAEIGLGFGVCFLLLVGGSVAMGVICTAIALFQTLAVQVGRQADRR AFTVPTIVVEDAQGKRRSSIDGSEPAKTSLQTTGLVTTIVFIYDCLMGFPVLVVSFSSLRADASAPWMAL CVLWCSVAQALLLPVFLWACDRYRADLKAVREKCMALMANDEESDDETSLEGGISPDLVLERSLDYGYGG DFVALDRMAKYEISALEGGLPQLYPLRPLQEDKMQYLQVPPTRRFSHDDADVWAAVPLPAFLPRWGSGED LAALAHLVLPAGPERRRASLLAFAEDAPPSRARRRSAESLLSLRPSALDSGPRGARDSPPGSPRRRPGPG PRSASASLLPDAFALTAFECEPQALRRPPGPFPAAPAAPDGADPGEAPTPPSSAQRSPGPRPSAHSHAGS

>NP\_002089.4 guanylyl cyclase-activating protein 2 [Homo sapiens]

TPEEVVDRIFLLVDENGDGQLSLNEFVEGARRDKWVMKMLQMDMNPSSWLAQQRRKSAMF

VATEVPFRLMHPQPEDPAKESYQDANLVFEEFARHNLKDAGEAEEGKRDKNDVDE

MGQEFSWEEAEAAGEIDVAELQEWYKKFVMECPSGTLFMHEFKRFFKVTDDEEASQYVEGMFRAFDKNGD NTIDFLEYVAALNLVLRGTLEHKLKWTFKIYDKDGNGCIDRLELLNIVEGIYQLKKACRRELQTEQGQLL

MAASGKTSKSEPNHVIFKKISRDKSVTIYLGNRDYIDHVSQVQPVDGVVLVDPDLVKGKKVYVTLTCAFR YGQEDIDVIGLTFRRDLYFSRVQVYPPVGAASTPTKLQESLLKKLGSNTYPFLLTFPDYLPCSVMLQPAP QDSGKSCGVDFEVKAFATDSTDAEEDKIPKKSSVRLLIRKVQHAPLEMGPQPRAEAAWQFFMSDKPLHLA VSLNKEIYFHGEPIPVTVTVTNNTEKTVKKIKAFVEQVANVVLYSSDYYVKPVAMEEAQEKVPPNSTLTK TLTLLPLLANNRERRGIALDGKIKHEDTNLASSTIIKEGIDRTVLGILVSYQIKVKLTVSGFLGELTSSE

>NP\_775904.2 transmembrane protein 145 isoform 2 precursor [Homo sapiens] MEPLRAPALRRLLPPLLLLLLSLPPRARAKYVRGNLSSKEDWVFLTRFCFLSDYGRLDFRFRYPEAKCCQ NILLYFDDPSQWPAVYKAGDKDCLAKESVIRPENNQVINLTTQYAWSGCQVVSEEGTRYLSCSSGRSFRS GDGLQLEYEMVLTNGKSFWTRHFSADEFGILETDVTFLLIFILIFFLSCYFGYLLKGRQLLHTTYKMFMA AAGVEVLSLLFFCIYWGQYATDGIGNESVKILAKLLFSSSFLIFLLMLILLGKGFTVTRGRISHAGSVKL SVYMTLYTLTHVVLLIYEAEFFDPGQVLYTYESPAGYGLIGLQVAAYVWFCYAVLVSLRHFPEKQPFYVP FFAAYTLWFFAVPVMALIANFGIPKWAREKIVNGIQLGIHLYAHGVFLIMTRPSAANKNFPYHVRTSQIA SAGVPGPGGSQSADKAFPQHVYGNVTFISDSVPNFTELFSIPPPATSPLPRAAPDSGLPLFRDLRPPGPL

>NP\_000946.2 prostaglandin E2 receptor EP1 subtype [Homo sapiens]

YILLRQAVLRQLLRLLPPRAGAKGGPAGLGLTPSAWEASSLRSSRHSGLSHF

ARSIDFNKDGHIDINEFLEAFRLVEKSCPEGDASECPQATNAKDSGCSSPGAH

>NP\_054798.1 Krueppel-like factor 15 [Homo sapiens]

>NP\_064707.1 atypical chemokine receptor 3 [Homo sapiens]

MSPCGPLNLSLAGEATTCAAPWVPNTSAVPPSGASPALPIFSMTLGAVSNLLALALLAQAAGRLRRRRSA ATFLLFVASLLATDLAGHVIPGALVLRLYTAGRAPAGGACHFLGGCMVFFGLCPLLLGCGMAVERCVGVT RPLLHAARVSVARARLALAAVAAVALAVALLPLARVGRYELQYPGTWCFIGLGPPGGWRQALLAGLFASL GLVALLAALVCNTLSGLALLRARWRRRSRRPPPASGPDSRRRWGAHGPRSASASSASSIASASTFFGGSR SSGSARRARAHDVEMVGQLVGIMVVSCICWSPMLVLVALAVGGWSSTSLQRPLFLAVRLASWNQILDPWV

MDLHLFDYSEPGNFSDISWPCNSSDCIVVDTVMCPNMPNKSVLLYTLSFIYIFIFVIGMIANSVVVWVNI QAKTTGYDTHCYILNLAIADLWVVLTIPVWVVSLVQHNQWPMGELTCKVTHLIFSINLFGSIFFLTCMSV DRYLSITYFTNTPSSRKKMVRRVVCILVWLLAFCVSLPDTYYLKTVTSASNNETYCRSFYPEHSIKEWLI GMELVSVVLGFAVPFSIIAVFYFLLARAISASSDQEKHSSRKIIFSYVVVFLVCWLPYHVAVLLDIFSIL HYIPFTCRLEHALFTALHVTQCLSLVHCCVNPVLYSFINRNYRYELMKAFIFKYSAKTGLTKLIDASRVS

MGSGTSTQHHFAFQNAERAFKAAALIQRWYRRYVARLEMRRRCTWSIFQSIEYAGQQDQVKLHDFFSYLM DHFIPSSHNDRDFLTRIFTEDRFAQDSEMKKCSDYESIEVPDSYTGPRLSFPLLPDHATALVEAFRLKQQ LHARYVLNLLYETKKHLVQLPNINRVSTCYSEEITVCGDLHGQLDDLIFIFYKNGLPSPERSYVFNGDFV DRGKDSVEILMILFAFMLVYPKEFHLNRGNHEDHMVNLRYGFTKEVMNKYKVHGKEILRTLQDVFCWLPL ATLIDEKVLILHGGVSDITDLELLDKIERSKIVSTMRCKTRQKSEKQMEEKRRANQKSSAQGPIPWFLPE SRSLPSSPLRLGSYKAQKTSRSSSIPCSGSLDGRELSRQVRSSVELELERCRQQAGLLVTGEKEEPSRSA SEADSEAGELRKPTQEEWRQVVDILWSDPMAQEGCKANTIRGGGCYFGPDVTQQLLQKYNMQFLIRSHEC KPEGYEFCHNRKVLTIFSASNYYEVGSNRGAYVKLGPALTPHIVQYQANKVTHTLTMRQRISRVEESALR ALREKLFAHSSDLLSEFKKHDADKVGLITLSDWAAAVESVLHLGLPWRMLRPQLVNSSADNMLEYKSWLK NLAKEQLSRENIQSSLLETLYRNRSNLETIFRIIDSDHSGFISLDEFRQTWKLFSSHMNIDITDDCICDL

MVDHLLPVDENFSSPKCPVGYLGDRLVGRRAYHMLPSPVSEDDSDASSPCSCSSPDSQALCSCYGGGLGT ESQDSILDFLLSQATLGSGGGSGSSIGASSGPVAWGPWRRAAAPVKGEHFCLPEFPLGDPDDVPRPFQPT LEEIEEFLEENMEPGVKEVPEGNSKDLDACSQLSAGPHKSHLHPGSSGRERCSPPPGGASAGGAQGPGGG PTPDGPIPVLLQIQPVPVKQESGTGPASPGQAPENVKVAQLLVNIQGQTFALVPQVVPSSNLNLPSKFVR IAPVPIAAKPVGSGPLGPGPAGLLMGQKFPKNPAAELIKMHKCTFPGCSKMYTKSSHLKAHLRRHTGEKP FACTWPGCGWRFSRSDELSRHRRSHSGVKPYQCPVCEKKFARSDHLSKHIKVHRFPRSSRSVRSVN

MLTGSCGDPQKKPQVTQDSGPQSMGLEGRETAGQPRVTLLPTPHVSGLSQEFESHWPEIAERSPCVAGVI PVIYYSVLLGLGLPVSLLTAVALARLATRTRRPSYYYLLALTASDIIIQVVIVFAGFLLQGAVLARQVPQ AVVRTANILEFAANHASVWIAILLTVDRYTALCHPLHHRAASSPGRTRRAIAAVLSAALLTGIPFYWWLD MWRDTDSPRTLDEVLKWAHCLTVYFIPCGVFLVTNSAIIHRLRRRGRSGLQPRVGKSTAILLGITTLFTL LWAPRVFVMLYHMYVAPVHRDWRVHLALDVANMVAMLHTAANFGLYCFVSKTFRATVRQVIHDAYLPCTL

MGPRRLLLVAACFSLCGPLLSARTRARRPESKATNATLDPRSFLLRNPNDKYEPFWEDEEKNESGLTEYR LVSINKSSPLQKQLPAFISEDASGYLTSSWLTLFVPSVYTGVFVVSLPLNIMAIVVFILKMKVKKPAVVY MLHLATADVLFVSVLPFKISYYFSGSDWQFGSELCRFVTAAFYCNMYASILLMTVISIDRFLAVVYPMQS LSWRTLGRASFTCLAIWALAIAGVVPLLLKEQTIQVPGLNITTCHDVLNETLLEGYYAYYFSAFSAVFFF VPLIISTVCYVSIIRCLSSSAVANRSKKSRALFLSAAVFCIFIICFGPTNVLLIAHYSFLSHTSTTEAAY FAYLLCVCVSSISCCIDPLIYYYASSECQRYVYSILCCKESSDPSSYNSSGQLMASKMDTCSSNLNNSIY

>NP\_001318005.1 probable G-protein coupled receptor 142 isoform b [Homo sapiens]

>NP\_001983.2 proteinase-activated receptor 1 isoform 1 precursor [Homo sapiens]

lne = Entrez.efetch(db="protein", id=rec, retmode="text", rettype="fasta").read()

>NP\_001316355.1 guanine nucleotide-binding protein G(T) subunit gamma-T1 precursor [Homo sapiens]

lhandle = Entrez.elink(dbfrom="nucleotide", db="protein", id="2065188392") ##pubmed default

rrecord = Entrez.efetch(db="protein", id=prothandle, rettype="fasta", retmode="text")

{'Id': '312146892'}, {'Id': '312146891'}, {'Id': '312146890'}, {'Id': '312146889'}, {'Id': '3

rrecord = Entrez.efetch(db="nucleotide", id=ids[:4], rettype="fasta", retmode="text")

>HM140499.1 Myospora metanephrops isolate NZ6C small subunit ribosomal RNA gene, partial sequence

>HM140498.1 Myospora metanephrops isolate NZ6B small subunit ribosomal RNA gene, partial sequence

>HM140497.1 Myospora metanephrops isolate NZ6A small subunit ribosomal RNA gene, partial sequence

>HM140496.1 Myospora metanephrops isolate NZ5D small subunit ribosomal RNA gene, partial sequence

>NP\_006230.2 serine/threonine-protein phosphatase with EF-hands 2 [Homo sapiens]

AQSLSPKLKAKPVSSLNECTTKDARPGCLLRSDP

LRPGLSASWGEPGGLRAAGGGGSTSSFLSSPSESSGYATLHSDSLGSAS

>NP\_000532.2 S-arrestin [Homo sapiens]

RDL

**ETEYSALEQSTK** 

ASQPEGMAAKPVMEPPGLPTGAEV

Save downloaded sequences into a fasta file

ouf.write(lne)

HLPLSTVHTEVSKGSLRLSGRSNPI

for rec in record["IdList"]:

! head -10 results/rhodopsin.fasta

Download protein by its nucleotide UID

lrecord = Entrez.read(lhandle)

ouf.write(rrecord.read())

Download all sequences by PMID

'DbFrom': 'pubmed', 'IdList': ['20558169']}]

for el in lrecord[0]["LinkSetDb"][0]["Link"]:

with open ("results/py\_fasta\_pmid.fasta", "w") as ouf:

GACGGCTACCAGGTCCAAGGACAGCAGCAGCGCGCAAAATTACCGAAGCCTACAACAGGGCGGTAGTAAT GAGACGTGAAAACTAGACACGAATAAAATACGTGTTAGCAACTGGAGGTCAAGTCTGGTGCCAGCATCCG CGGTAATACCAGCTCCAGGGGTGTCTATGATGATTGCTGCGATTAAAAGGTCCGTAGTCTTATGTCAGAA ATAGAGATGTTATATTTGTAATGTTGATATGTATATGGTGCAATATATTGAAATGAGGAGCGACCGGGGG CAAGGGCGAATCCGATGATCAAGGACGTAGGCTGGAGGATCGAACACGATTAGATACCGTAGTAGTTCCA GCAGTAAACTATGCCGACGCCGTGGGTTGTTTGACCCGCGGAAGAGAAATCTAGTAGGGCTTTGGGGAGA GTACGCGCGCAAGCGATAAATTTAAAGGAAATTGACGGAAGAACACCACAAGGAGTGGAGTGTGCGGCTT AATTTGACTCAACGCGGGACAGCTTACCAGGCCCGAGGATTGCACGAGCGAATACGCGATAGATCTGAAA GTGGTGCATGGCCGTTATCGACGAATGGAGTGATCTTTTGGTTAAATCCGTCAATTCGTGAGACCCTTTT AATTTGATTAATGTCAGTGGTTGATACAGGTATGAAAATACAGGGGGGAAAGGACAAGAACAGGTCAGTG ATGCCCTTAGATGGCCTGGGCTGCACGCGCACTACAGTGGTTACTTTAAACTGAAGAGAGGAATAAATGT AATCGAGAGGGAATGAGCCCTGCAAGGCCCACAGGAACGAGGAATTGCTAGTAATCGTAGGCTCAGTAAG

GACGGCTACCAGGTCCAAGGACAGCAGCAGGCGCGAAAATTACCGAAGCCTACAACAGGGCGGTAGTAAT GAGACGTGAAAACTAGACACGAATAAAATACGTGTTAGCAACTGGAGGTCAAGTCTGGTGCCAGCATCCG CGGTAATACCAGCTCCAGGGGTGTCTATGATGATTGCTGCGATTAAAAGGTCCGTAGTCTTATGTCAGAA ATAGAGATGTTATATTTGTAATGTTGATATGTATATGGTGCAATATATTGAAATGAGGAGCGACCGGGGG CAAGGGCGAATCCGATGATCAAGGACGTAGGCTGGAGGATCGAACACGATTAGATACCGTAGTAGTTCCA GCAGTAAACTATGCCGACGCCGTGGGTTGTTTGACCCGCGGAAGAGAAATCTAGTAGGGCTTTGGGGAGA GTACGCGCGCAAGCGATAAATTTAAAGGAAATTGACGGAAGAACACCACAAGGAGTGGAGTGCGGCTT AATTTGACTCAACGCGGGACAGCTTACCAGGCCCGAGGATTGCACGAGCGAATACGCGATAGATCTGAAA GTGGTGCATGGCCGTTATCGACGAATGGAGTGATCTTTTGGTTAAATCCGTCAATTCGTGAGACCCTTTT AATTTGATTAATGTCAGTGGTTGATACAGGTATGAAAATACAGGGGGGAAAGGACAAGAACAGGTCAGTG ATGCCCTTAGATGGCCTGGGCTGCACGCGCACTACAGTGGTTACTTTAAACTGAAGAGAGGAATAAATGT AATCGAGAGGGAATGAGCGCTGCAAGGCGCACAGGAACGAGGAATTGCTAGTAATCGTAGGCTCAGTAAG

GACGGCTACCAGGTCCAAGGACAGCAGCAGCGCGCAAAATTACCGAAGCCTACAACAGGGCGGTAGTAAT GAGACGTGAAAACTAGACACGAATAAAATACGTGTTAGCAACTGGAGGTCAAGTCTGGTGCCAGCATCCG CGGTAATACCAGCTCCAGGGGTGTCTATGATGATTGCTGCGATTAAAAGGTCCGTAGTCTTATGTCAGAA CAAGGGCGAATCCGATGATCAAGGACGTAGGCTGGAGGATCGAACACGATTAGATACCGTAGTAGTTCCA GCAGTAAACTATGCCGACGCCGTGGGTTGTTTGACTCGCGGAAGAGAAATCTAGTAGGGCTTTGGGGAGA GTACGCGCGCAAGCGATAAATTTAAAGGAAATTGACGGAAGAACACCACAAGGAGTGGAGTGCGGCTT AATTTGACTCAACGCGGGACAGCTTACCAGGCCCGAGGATTGCACGAGCGAATACGCGATAGATCTGAAA GTGGTGCATGGCCGTTATCGACGAATGGAGTGATCTTTTGGTTAAATCCGTCAATTCGTGAGACCCTTTT AATTTGATTAATGTCAGTGGTTGATACAGGTATGAAAATACAGGGGGGAAAGGACAAGAACAGGTCAGTG ATGCCCTTAGATGGCCTGGGCTGCACGCGCACTACAGTGGTTACTTTAAACTGAAGAGAGGAATAAATGT AATCGAGAGGGAATGAGCGCTGCAAGGCGCACAGGAACGAGGAATTGCTAGTAATCGTAGGCTCAGTAAG

GACGGCTACCAGGTCCAAGGACAGCAGCAGCGCGCAAAATTACCGAAGCCTACAACAGGGCGGTAGTAAT GAGACGTGAAAACTAGACACGAATAAAATACGTGTTAGCAACTGGAGGTCAAGTCTGGTGTCAGCATCCG CGGTAATACCAGCTCCAGGGGTGTCTATGATGATTGCTGCGATTAAAAGGTCCGTAGTCTTATGTCAGAA ATAGAGATGTTATATTTGTAATGTTGATATGTATATGGTGCAATATATTGAAATGAGGAGCGACCGGGGG CAAGGGCGAATCCGATGATCAAGGACGTAGGCTGGAGGATCGAACACGATTAGATACCGTAGTAGTTCCA GCAGTAAACTATGCCGACGCCGTGGGTTGTTTGACCCGCGGAAGAGAAATCTAGTAGGGCTTTGGGGAGA GTACGCGCGCAAGCGATAAATTTAAAGGAAATTGACGGAAGAACACCACAAGGAGTGGAGTGCGGCTT AATTTGACTCAACGCGGGACAGCTTACCAGGCCCGAGGATTGTACGAGCGAATACGCGATAGATCTGAAA GTGGTGCATGGCCGTTATCGACGAATGGAGTGATCTTTTGGTTAAATCCGTCAATTCGTGAGACCCTTTT AATTTGATTAATGTCAGAGGTTGATACAGGTATGAAAATACAGGGGGGAAAGGACAAGAACAGGTCAGTG ATGCCCTTAGATGGCCTGGGCTGCACGCGCACTACAGTGGTTACTTTAAACTGAAGAGACGAATAAATGT AATCGAGAGGGAATGAGCGCTGCAAGGCGCACAGGAACGAGGAATTGCTAGTAATCGTAGGCTCAGTAAG

lrecord = Entrez.read(lhandle)

ids.append(el['Id'])

ouf.write(rrecord.read())

ATACGATGAATGTGTCCCTGTTC

ATACGATGAATGTGTCCCTGTTC

ATACGATGAATGTGTCCCTGTTC

ATACGATGAATGTGTCCCTGTTC

lrecord

lrecord

'312146894', '312146893', '312146892' '312146891' '312146890' '312146889' '312146888' '312146887']

with open("results/rhodopsin.fasta", "w") as ouf:

>NP\_001265723.1 neuropeptide Y receptor type 4 [Homo sapiens]

prothandle = lrecord[0]["LinkSetDb"][0]['Link'][0]['Id']

>XP\_042223242.1 crustacyanin-A2 subunit-like [Homarus americanus]

MGVWYEIQAQPNIFQSIKSCLASSYKRVKTEIHVLSEGLDSSGASTTTKSILKIVDPQNPAHMVTDFVPG VEPPFDIVDTDYKTFSCAHSCLSIVGIKTEFVFIYSRNRTLRSNSTQHCLSIFEVSIIGIISFYTNANNY

lhandle = Entrez.elink(dbfrom="pubmed", db="nucleotide", id="20558169")

with open ("results/prot\_from\_nt.fasta", "w") as ouf:

MNTSHLLALLLPKSPQGENRSKPLGTPYNFSEHCQDSVDVMVFIVTSYSIETVVGVLGNLCLMCVTVRQK EKANVTNLLIANLAFSDFLMCLLCQPLTAVYTIMDYWIFGETLCKMSAFIQCMSVTVSILSLVLVALERH QLIINPTGWKPSISQAYLGIVLIWVIACVLSLPFLANSILENVFHKNHSKALEFLADKVVCTESWPLAHH RTIYTTFLLLFQYCLPLGFILVCYARIYRCLQRQGRVFHKGTYSLRAGHMKQVNVVLVVMVVAFAVLWLP LHVFNSLEDWHHEAIPICHGNLIFLVCHLLAMASTCVNPFIYGFLNTNFKKEIKALVLTCQQSAPLEESE

MPVINIEDLTEKDKLKMEVDQLKKEVTLERMLVSKCCEEVRDYVEERSGEDPLVKGIPEDKNPFKELKGG

**KKLLT**