NCBI Molecular Biology Resources

Using Entrez

March 2007

1988 as a national
olecular biology
CBI creates public
nducts research in
biology, develops
for analyzing genome
eminates biomedical
l for the better
of molecular processes
n health and disease.

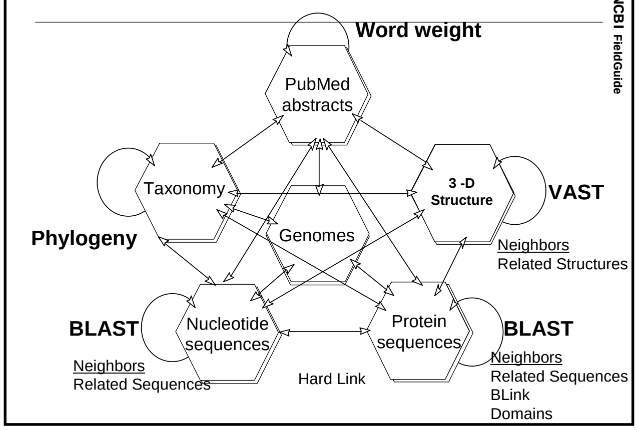
ome Association (WGA) resource
rs with access to genotype and
ype information that will help
etween genes and disease. For
click here to see the the WGA
resource page and click here to read the press
release.

Two papers authored by researchers at NCBI are
among the top 40 most cited articles for 2005
according to ScienceWatch's "Hottest Research

WWW
Access

Entrez
&
BLAST

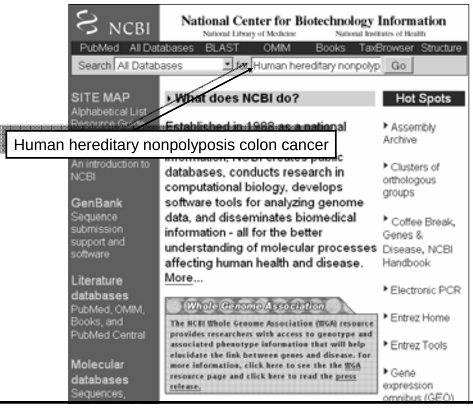
Entrez: Database Integration



Database Searching with Entrez

◆Using limits and field restriction to find human MutL homolog
◆Linking and neighboring with MutL
◆Mapping SNPs onto structure and the genome

Global NCBI (Entrez) Search



Human hereditary nonpolyposis colon cancer

What does NCBI do?

Established in 1988 as a national … information, NCBI creates public databases, conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information - all for the better understanding of molecular processes affecting human health and disease. More...

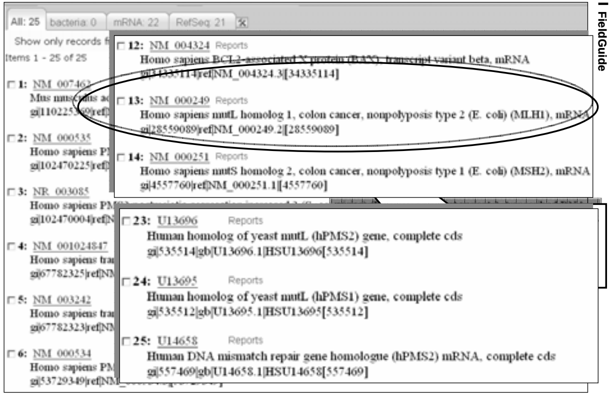
The NCBI Whole Genome Association (WGA) resource provides researchers with access to genotype and associated phenotype information that will help elucidate the link between genes and disease. For more information, click here to see the the WGA resource page and click here to read the press release.

Global Entrez Search Results

Search across databases: Human hereditary nonpolyposis colon cancer

1981	**PubMed:** biomedical literature citations and abstracts	13	**Books:** online books
178	**PubMed Central:** free, full text journal articles	30	**OMIM:** online Mendelian Inheritance in Man
none	**Site Search:** NCBI web and FTP sites	none	**OMIA:** Online Mendelian Inheritance in Animals
25	**Nucleotide:** sequence database (includes GenBank)	1	**UniGene:** gene-oriented clusters of transcript sequences
33	**Protein:** sequence database	1	**CDD:** conserved protein domain database
none	**Genome:** whole genome sequences	none	**3D Domains:** domains from Entrez Structure
none	**Structure:** three-dimensional macromolecular structures	none	**UniSTS:** markers and mapping data
none	**Taxonomy:** organisms in GenBank	none	**PopSet:** population study data sets
none	**SNP:** single nucleotide polymorphism	none	**GEO Profiles:** expression and molecular abundance profiles
6	**Gene:** gene-centered information	none	**GEO DataSets:** experimental sets of GEO data
6	**HomoloGene:** eukaryotic homology groups	1	**Cancer Chromosomes:** cytogenetic databases
none	**PubChem Compound:** unique small molecule chemical structures	none	**PubChem BioAssay:** bioactivity screens of chemical substances
none	**PubChem Substance:** deposited chemical substance records	none	**GENSAT:** gene expression atlas of mouse central nervous system
none	**Genome Project:** genome project information	none	**Probe:** sequence-specific reagents

Nucleotide Sequences



☐ 1: NM_007462
Mus musculus ad
gi|110225369|ref|

☐ 2: NM_000535
Homo sapiens P
gi|102470225|ref|

☐ 3: NR_003085
Homo sapiens PM
gi|102470004|ref|N

☐ 4: NM_001024847
Homo sapiens tra
gi|67782325|ref|NM

☐ 5: NM_003242
Homo sapiens tra
gi|67782323|ref|NM

☐ 6: NM_000534
Homo sapiens PM
gi|53729349|ref|NM

☐ 12: NM_004324 Reports
Homo sapiens BCL2-associated X protein (BAX), transcript variant beta, mRNA
gi|34335114|ref|NM_004324.3|[34335114]

☐ 13: NM_000249 Reports
Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA
gi|28559089|ref|NM_000249.2|[28559089]

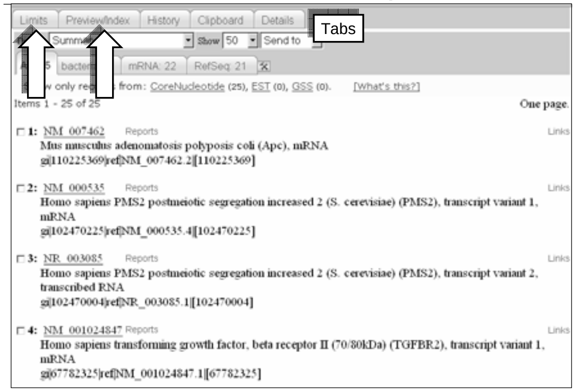
☐ 14: NM_000251 Reports
Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA
gi|4557760|ref|NM_000251.1|[4557760]

☐ 23: U13696 Reports
Human homolog of yeast mutL (hPMS2) gene, complete cds
gi|535514|gb|U13696.1|HSU13696[535514]

☐ 24: U13695 Reports
Human homolog of yeast mutL (hPMS1) gene, complete cds
gi|535512|gb|U13695.1|HSU13695[535512]

☐ 25: U14658 Reports
Human DNA mismatch repair gene homologue (hPMS2) mRNA, complete cds
gi|557469|gb|U14658.1|HSU14658[557469]

Advanced Search Options



Tabs

…w only re… from: CoreNucleotide (25), EST (0), GSS (0). [What's this?]

Items 1 - 25 of 25 One page.

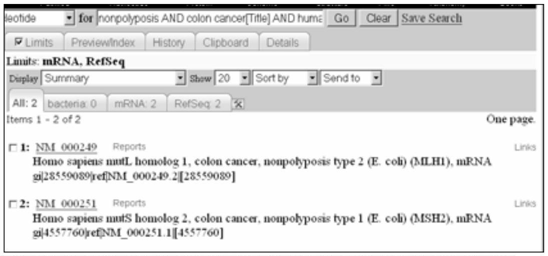
☐ 1: NM_007462 Reports Links
Mus musculus adenomatosis polyposis coli (Apc), mRNA
gi|110225369|ref|NM_007462.2|[110225369]

☐ 2: NM_000535 Reports Links
Homo sapiens PMS2 postmeiotic segregation increased 2 (S. cerevisiae) (PMS2), transcript variant 1, mRNA
gi|102470225|ref|NM_000535.4|[102470225]

☐ 3: NR_003085 Reports Links
Homo sapiens PMS2 postmeiotic segregation increased 2 (S. cerevisiae) (PMS2), transcript variant 2, transcribed RNA
gi|102470004|ref|NR_003085.1|[102470004]

☐ 4: NM_001024847 Reports Links
Homo sapiens transforming growth factor, beta receptor II (70/80kDa) (TGFBR2), transcript variant 1, mRNA
gi|67782325|ref|NM_001024847.1|[67782325]

More Precise Nucleotides Search



nonpolyposis AND colon cancer[Title] AND huma

Limits: mRNA, RefSeq

Display: Summary Show: 20

Items 1 - 2 of 2 One page.

☐ 1: NM_000249

Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) (MLH1), mRNA
gi|28559089|ref|NM_000249.2|[28559089]

☐ 2: NM_000251

Homo sapiens mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli) (MSH2), mRNA
gi|4557760|ref|NM_000251.1|[4557760]

nonpolyposis[All Fields] AND colon cancer[Title] AND human[Organism] AND biomol_mrna[Properties] AND srcdb_refseq[Properties]

Useful Field Restrictions

[Title]: Definition line in GenBank / GenPept format shown in Summary format

glyceraldehyde 3 phosphate dehydrogenase[Title]

[Organism]: NCBI's taxonomy. Organizing system for molecular databases

mouse[organism]; green plants[organism]; Streptomyces coelicolor[organism]

[Properties]: molecule type, location, database source

biomol_mrna[properties]; biomol_genomic[properties];
gene_in_mitochondrion[properties]; srcdb pdb[properties]

[Filter]: subsets of data, Entrez links

all[filter]; nucleotide mapview[filter]; nucleotide omim[filter]

Organism Field: NCBI's Taxonomy

Homo sapiens

Taxonomy ID: 9606
Genbank common name: **human**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 2 (Vertebrate Mitochondrial)
Other names:
common name: man

Lineage(full)
cellular organisms; Eukaryota; Fungi/Metazoa group; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homo/Pan/Gorilla group; Homo

Entrez records		
Database name	Subtree links	Direct links
Nucleotide	11,501,520	11,501,512
Protein	315,017	315,016
Structure	8,347	8,347
Genome Sequences	51	51
Genome Projects	1	1
Popset	20,788	20,788
SNP	12,702,095	12,702,095
3D Domains	31,280	31,280
Domains	24	24
GEO Datasets	2,819	2,819
GEO Expressions	8,804,567	8,804,567
UniGene	86,810	86,810
UniSTS	322,092	322,092
PubMed Central	2,983	2,983
Gene	39,092	39,092
HomoloGene	19,250	19,250
Taxonomy	2	1

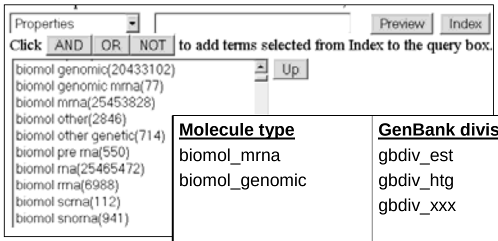
Genome Information

See the NCBI Genome homepage
Go to NCBI genomic BLAST page for Homo sapiens

Genome view: 24 chromosomes

Names 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y

Useful Properties Field Terms

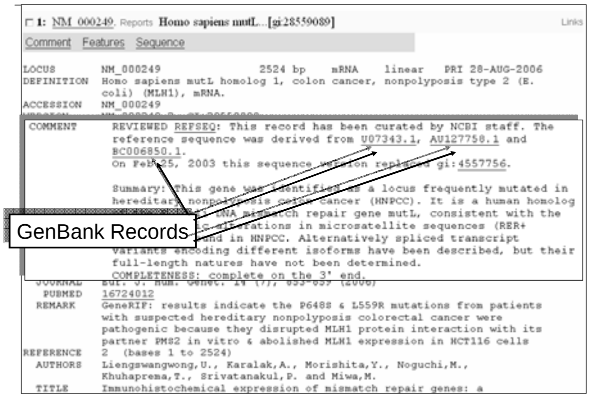


Click AND OR NOT to add terms selected from Index to the query box.

biomol genomic(20433102)
biomol genomic mrna(77)
biomol mrna(25453828)
biomol other(2846)
biomol other genetic(714)
biomol pre rna(550)
biomol rna(25465472)
biomol rrna(6988)
biomol scrna(112)
biomol snorna(941)

Molecule type	GenBank division
biomol_mrna biomol_genomic	gbdiv_est gbdiv_htg gbdiv_xxx
Gene location	**Source Database**
gene_in_mitochondrion gene_in_chloroplast gene_in_genomic	srcdb_refseq srcdb_pdb srcdb_swiss_prot

Human MutL RefSeq

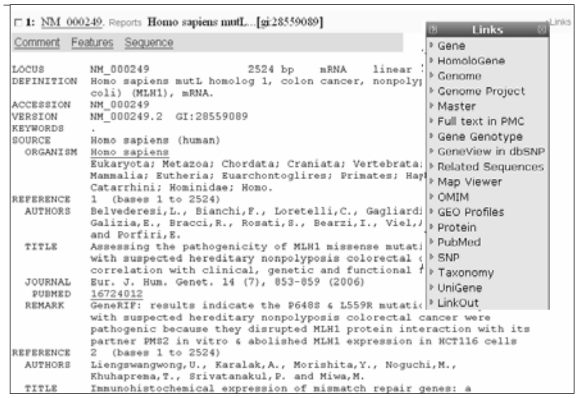


```
LOCUS       NM_000249                2524 bp    mRNA    linear   PRI 28-AUG-2006
DEFINITION  Homo sapiens mutL homolog 1, colon cancer, nonpolyposis type 2 (E.
            coli) (MLH1), mRNA.
ACCESSION   NM_000249
COMMENT     REVIEWED REFSEQ: This record has been curated by NCBI staff. The
            reference sequence was derived from U07343.1, AU127758.1 and
            BC006850.1.
            On Feb 25, 2003 this sequence version replaced gi:4557756.

            Summary: This gene was identified as a locus frequently mutated in
            hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog
                            DNA mismatch repair gene mutL, consistent with the
                            alterations in microsatellite sequences (RER+
                            and in HNPCC. Alternatively spliced transcript
            variants encoding different isoforms have been described, but their
            full-length natures have not been determined.
            COMPLETENESS: complete on the 3' end.
  PUBMED    16724012
  REMARK    GeneRIF: results indicate the P648S & L559R mutations from patients
            with suspected hereditary nonpolyposis colorectal cancer were
            pathogenic because they disrupted MLH1 protein interaction with its
            partner PMS2 in vitro & abolished MLH1 expression in HCT116 cells
REFERENCE   2  (bases 1 to 2524)
  AUTHORS   Liengswangwong,U., Karalak,A., Morishita,Y., Noguchi,M.,
            Khuhaprema,T., Srivatanakul,P. and Miwa,M.
  TITLE     Immunohistochemical expression of mismatch repair genes: a
```

GenBank Records

NM_000249: Links

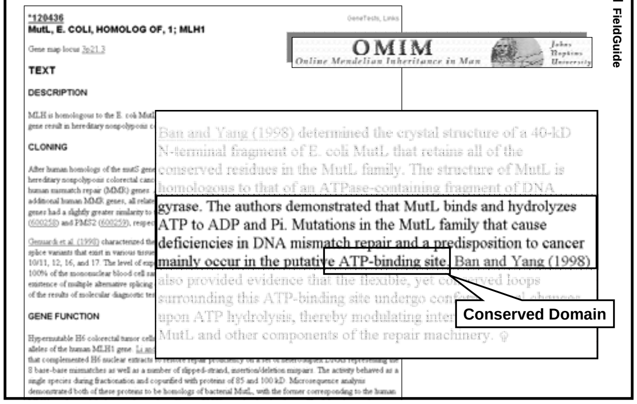


```
LOCUS       NM_000249                2524 bp    mRNA    linear
DEFINITION  Homo sapiens mutL homolog 1, colon cancer, nonpoly
            coli) (MLH1), mRNA.
ACCESSION   NM_000249
VERSION     NM_000249.2  GI:28559089
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata;
            Mammalia; Eutheria; Euarchontoglires; Primates; Hap
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2524)
  AUTHORS   Belvederesi,L., Bianchi,F., Loretelli,C., Gagliardi
            Galizia,E., Bracci,R., Rosati,S., Bearzi,I., Viel,A
            and Porfiri,E.
  TITLE     Assessing the pathogenicity of MLH1 missense mutati
            with suspected hereditary nonpolyposis colorectal c
            correlation with clinical, genetic and functional f
  JOURNAL   Eur. J. Hum. Genet. 14 (7), 853-859 (2006)
   PUBMED   16724012
  REMARK    GeneRIF: results indicate the P648S & L559R mutatio
            with suspected hereditary nonpolyposis colorectal cancer were
            pathogenic because they disrupted MLH1 protein interaction with its
            partner PMS2 in vitro & abolished MLH1 expression in HCT116 cells
REFERENCE   2  (bases 1 to 2524)
  AUTHORS   Liengswangwong,U., Karalak,A., Morishita,Y., Noguchi,M.,
            Khuhaprema,T., Srivatanakul,P. and Miwa,M.
  TITLE     Immunohistochemical expression of mismatch repair genes: a
```

Literature Links

OMIM

OMIM: Human Disease Genes



*120436

MutL, E. COLI, HOMOLOG OF, 1; MLH1

Gene map locus 3p21.3

TEXT

DESCRIPTION

CLONING

GENE FUNCTION

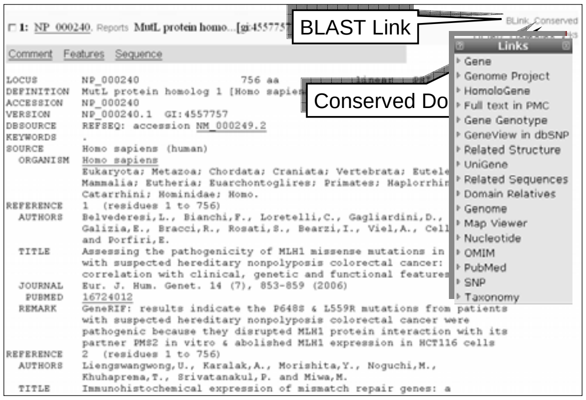
Ban and Yang (1998) determined the crystal structure of a 40-kD N-terminal fragment of E. coli MutL that retains all of the conserved residues in the MutL family. The structure of MutL is homologous to that of an ATPase-containing fragment of DNA gyrase. The authors demonstrated that MutL binds and hydrolyzes ATP to ADP and Pi. Mutations in the MutL family that cause deficiencies in DNA mismatch repair and a predisposition to cancer mainly occur in the putative ATP-binding site. Ban and Yang (1998) also provided evidence that the flexible, yet conserved loops surrounding this ATP-binding site undergo conf changes upon ATP hydrolysis, thereby modulating inter MutL and other components of the repair machinery.

Conserved Domain

Sequence Links

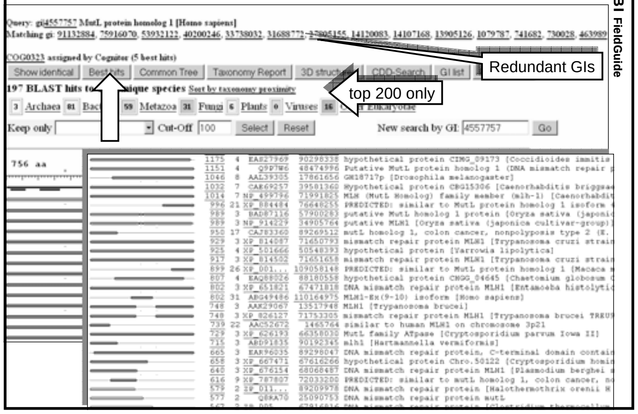
Finding Homologs and Structures

Protein Link



```
LOCUS       NP_000240                 756 aa
DEFINITION  MutL protein homolog 1 [Homo sapie
ACCESSION   NP_000240
VERSION     NP_000240.1  GI:4557757
DBSOURCE    REFSEQ: accession NM_000249.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Eutele
            Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhin
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 756)
  AUTHORS   Belvederesi,L., Bianchi,F., Loretelli,C., Gagliardini,D.,
            Galizia,E., Bracci,R., Rosati,S., Bearzi,I., Viel,A., Cell
            and Porfiri,E.
  TITLE     Assessing the pathogenicity of MLH1 missense mutations in
            with suspected hereditary nonpolyposis colorectal cancer:
            correlation with clinical, genetic and functional features
  JOURNAL   Eur. J. Hum. Genet. 14 (7), 853-859 (2006)
   PUBMED   16724012
  REMARK    GeneRIF: results indicate the P648S & L559R mutations from patients
            with suspected hereditary nonpolyposis colorectal cancer were
            pathogenic because they disrupted MLH1 protein interaction with its
            partner PMS2 in vitro & abolished MLH1 expression in HCT116 cells
REFERENCE   2  (residues 1 to 756)
  AUTHORS   Liengswangwong,U., Karalak,A., Morishita,Y., Noguchi,M.,
            Khuhaprema,T., Srivatanakul,P. and Miwa,M.
  TITLE     Immunohistochemical expression of mismatch repair genes: a
```

BLink: BLAST Link



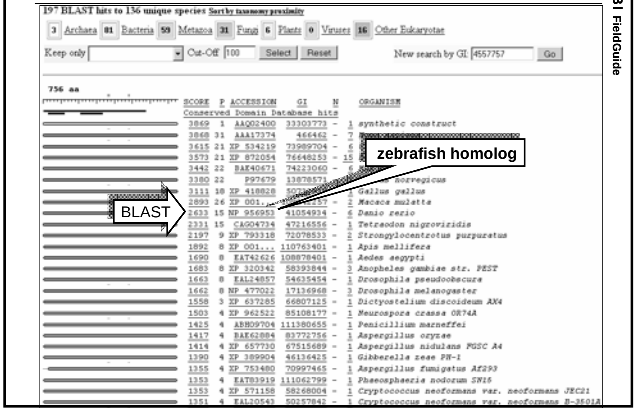
Query: gi4557757 MutL protein homolog 1 [Homo sapiens]
Matching gi: 91132884, 75916070, 53932122, 40200246, 33738032, 31688772, 37805155, 14120083, 14107168, 13905126, 1079787, 741682, 730028, 463989

COG0323 assigned by Cognitor (5 best hits)

197 BLAST hits to unique species Sort by taxonomy proximity

```
1175  4  EAS27969   90298338 hypothetical protein CIMG_09173 [Coccidioides immitis
1151  4    Q9P7W6   48474996 Putative MutL protein homolog 1 (DNA mismatch repair p
1046  8  AAL39305   17861656 GM18717p [Drosophila melanogaster]
1032  7  CAE69257   39581360 Hypothetical protein CBG15306 [Caenorhabditis briggsae
1014  7 NP_499796   71991825 MLH (MutL Homolog) family member (mlh-1) [Caenorhabdit
 996 21 XP_884484   76648255 PREDICTED: similar to MutL protein homolog 1 isoform
 989  3  BAD87116   57900283 putative MutL homolog 1 protein [Oryza sativa (japonic
 989  3 NP_914229   34905764 putative MLH1 [Oryza sativa (japonica cultivar-group)]
 950 17  CAJ83360   89269512 mutL homolog 1, colon cancer, nonpolyposis type 2 (E.
 929  3 XP_814087   71650793 mismatch repair protein MLH1 [Trypanosoma cruzi strain
 925  4 XP_501666   50548393 hypothetical protein [Yarrowia lipolytica]
 917  3 XP_814502   71651658 mismatch repair protein MLH1 [Trypanosoma cruzi strain
 899 26 XP_001...  109058148 PREDICTED: similar to MutL protein homolog 1 [Macaca m
 807  4  EAQ88026   88180558 hypothetical protein CHGG_04645 [Chaetomium globosum C
 802  3 XP_651821   67471818 DNA mismatch repair protein MLH1 [Entamoeba histolytic
 802 31  ABG49486  110164975 MLH1-Ex(9-10) isoform [Homo sapiens]
 748  3  AAK29067   13517948 MLH1 [Trypanosoma brucei]
 748  3 XP_826127   71753305 mismatch repair protein MLH1 [Trypanosoma brucei TREU9
 739 22  AAC52672    1465764 similar to human MLH1 on chromosome 3p21
 729  3 XP_626193   66358030 MutL family ATpase [Cryptosporidium parvum Iowa II]
 715  3  ABD91835   90192345 mlh1 [Hartmannella vermiformis]
 665  3  EAR96035   89298047 DNA mismatch repair protein, C-terminal domain contain
 658  3 XP_667471   67616266 hypothetical protein Chro.50122 [Cryptosporidium homin
 640  3 XP_676154   68068487 DNA mismatch repair protein MLH1 [Plasmodium berghei s
 616  9 XP_787807   72033200 PREDICTED: similar to mutL homolog 1, colon cancer, no
 579  2 ZP_011...   89209978 DNA mismatch repair protein [Halothermothrix orenii H
 577  2    Q8RA70   25090753 DNA mismatch repair protein mutL
```

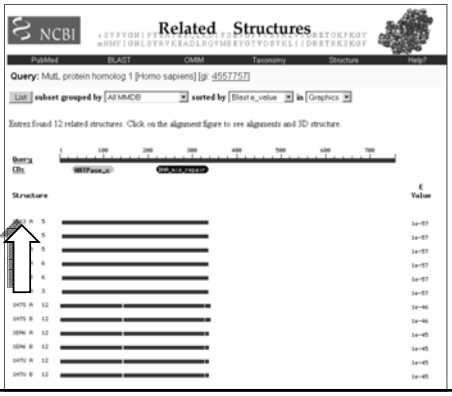
BLink: non-redundant relatives



197 BLAST hits to 136 unique species Sort by taxonomy proximity

```
SCORE  P ACCESSION    GI       N    ORGANISM
Conserved Domain Database hits
3869   1  AAQ02400   33303773 -  1 synthetic construct
3868  31  AAA17374     466462 -  7
3615  21 XP 534219   73989704 -  6
3573  21 XP 872054   76648253 - 15
3442  22  BAK40671   74223060 -  6
3380  22    P97679   13878571        norvegicus
3111  18 XP 418828   507         1 Gallus gallus
2893  26 XP 001...      42257 -  2 Macaca mulatta
2633  15 NP 956953   41054934 -  6 Danio rerio
2331  15  CAG04734   47216556 -  1 Tetraodon nigroviridis
2197   9 XP 793318   72078533 -  2 Strongylocentrotus purpuratus
1892   8 XP 001...  110763401 -  1 Apis mellifera
1690   8  EAT42626  108878401 -  1 Aedes aegypti
1683   8 XP 320342   58393844 -  3 Anopheles gambiae str. PEST
1663   8  EAL24857   54635454 -  1 Drosophila pseudoobscura
1662   8 NP 477022   17136968 -  3 Drosophila melanogaster
1558   3 XP 637285   66807125 -  1 Dictyostelium discoideum AX4
1503   4 XP 962522   85108177 -  1 Neurospora crassa OR74A
1425   4  ABH09704  111380655 -  1 Penicillium marneffei
1417   4  BAE62884   83772756 -  1 Aspergillus oryzae
1414   4 XP 657730   67515689 -  1 Aspergillus nidulans FGSC A4
1390   4 XP 389904   46136425 -  1 Gibberella zeae PH-1
1355   4 XP 753480   70997465 -  1 Aspergillus fumigatus Af293
1353   4  EAT83919  111062799 -  1 Phaeosphaeria nodorum SN15
1353   4 XP 571158   58268004 -  1 Cryptococcus neoformans var. neoformans JEC21
1351   4  EAL20543   50257842 -  1 Cryptococcus neoformans var. neoformans B-3501A
```

Short Cut: Related Structures



Related Structures

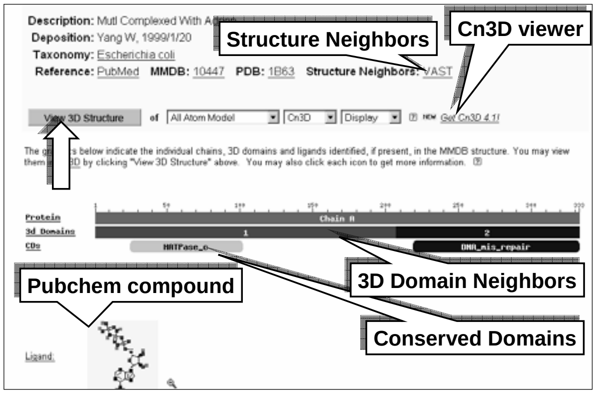
PubMed BLAST OMIM Taxonomy Structure Help?

Query: MutL protein homolog 1 [Homo sapiens] [gi: 4557757]

List **subset grouped by** All MMDB **sorted by** Blast e_value **in** Graphics

Entrez found 12 related structures. Click on the alignment figure to see alignments and 3D structure.

E. coli MutL Structure



Description: Mutl Complexed With A

Deposition: Yang W, 1999/1/20

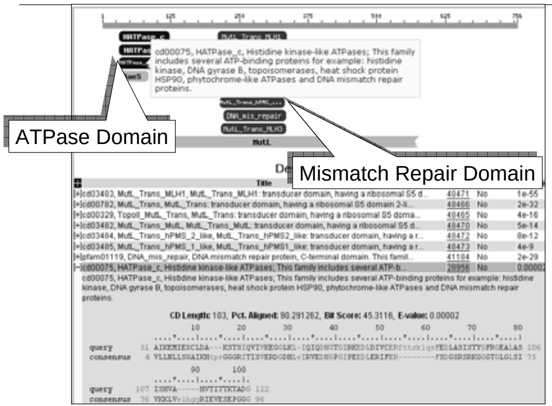
Taxonomy: Escherichia coli

Reference: PubMed **MMDB:** 10447 **PDB:** 1B63 **Structure Neighbors:** VAST

View 3D Structure of All Atom Model Cn3D Display NEW Get Cn3D 4.1!

The g s below indicate the individual chains, 3D domains and ligands identified, if present, in the MMDB structure. You may view them 3D by clicking "View 3D Structure" above. You may also click each icon to get more information.

MLH1 Domain Structure: CDD



cd00075, HATPase_c, Histidine kinase-like ATPases; This family includes several ATP-binding proteins for example: histidine kinase, DNA gyrase B, topoisomerases, heat shock protein HSP90, phytochrome-like ATPases and DNA mismatch repair proteins.

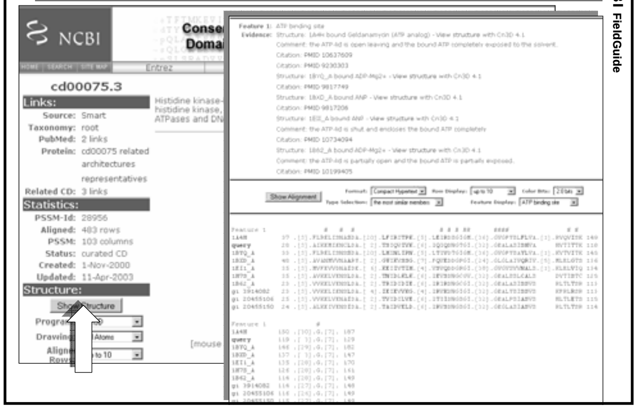
Title			
[+]cd03483, MutL_Trans_MLH1, MutL_Trans_MLH1: transducer domain, having a ribosomal S5 d...	48471	No	1e-55
[+]cd00782, MutL_Trans, MutL_Trans: transducer domain, having a ribosomal S5 domain 2-li...	48466	No	2e-32
[+]cd00329, TopoII_MutL_Trans, MutL_Trans: transducer domain, having a ribosomal S5 doma...	48465	No	4e-16
[+]cd03482, MutL_Trans_MutL, MutL_Trans_MutL: transducer domain, having a ribosomal S5 d...	48470	No	5e-14
[+]cd03484, MutL_Trans_hPMS_2_like, MutL_Trans_hPMS2_like: transducer domain, having a r...	48472	No	8e-12
[+]cd03485, MutL_Trans_hPMS_1_like, MutL_Trans_hPMS1_like: transducer domain, having a r...	48473	No	4e-9
[+]pfam01119, DNA_mis_repair, DNA mismatch repair protein, C-terminal domain. This famil...	41184	No	2e-29
[-]cd00075, HATPase_c, Histidine kinase-like ATPases; This family includes several ATP-b...	28956	No	0.00002

cd00075, HATPase_c, Histidine kinase-like ATPases; This family includes several ATP-binding proteins for example: histidine kinase, DNA gyrase B, topoisomerases, heat shock protein HSP90, phytochrome-like ATPases and DNA mismatch repair proteins.

CD Length: 103, **Pct. Aligned:** 90.291262, **Bit Score:** 45.3116, **E-value:** 0.00002

```
                 10        20        30        40        50        60        70        80
          ....*....|....*....|....*....|....*....|....*....|....*....|....*....|....*....|
query      31 AIKEMIENCLDA---KSTSIQVIVKEGGLKL-IQIQDNGTGIRKEDLDIVCERftsklqsFEDLASISTYGFRGEALAS 106
consensus   4 VLLNLLSNAIKHtpeGGGRITISVERDGDHLeIRVEDNGPGIPEEDLERIFER--------FSDGSRSRKGGGTGLGLSI 75
                 90       100
          ....*....|....*....|.
query     107 ISHVA-----HVTITTKTADG 122
consensus  76 VKKLVelhggRIEVESEPGGG 96
```

MLH1: ATPase Domain



Conse Domai

HOME SEARCH SITE MAP Entrez

cd00075.3

Links:

Source: Smart

Taxonomy: root

PubMed: 2 links

Protein: cd00075 related architectures representatives

Related CD: 3 links

Statistics:

PSSM-Id: 28956

Aligned: 483 rows

PSSM: 103 columns

Status: curated CD

Created: 1-Nov-2000

Updated: 11-Apr-2003

Structure:

Show Structure

Program Drawing Aligned Rows

Histidine kinase- histidine kinase, ATPases and DN

[mouse

Feature 1: ATP binding site

Evidence: Structure: 1A4H bound Geldanamycin (ATP analog) - View structure with Cn3D 4.1

Comment: the ATP-lid is open leaving and the bound ATP completely exposed to the solvent.

Citation: PMID 10637609

Citation: PMID 9230303

Structure: 1BYQ_A bound ADP-Mg2+ - View structure with Cn3D 4.1

Citation: PMID 9817749

Structure: 1BXD_A bound ANP - View structure with Cn3D 4.1

Citation: PMID 9817206

Structure: 1EI1_A bound ANP - View structure with Cn3D 4.1

Comment: the ATP-lid is shut and encloses the bound ATP completely

Citation: PMID 10734094

Structure: 1B62_A bound ADP-Mg2+ - View structure with Cn3D 4.1

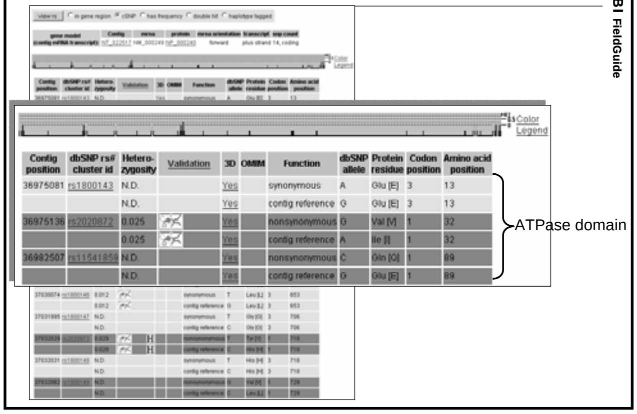
Comment: the ATP-lid is partially open and the bound ATP is partially exposed.

Citation: PMID 10199405

Show Alignment Format: Compact Hypertext Row Display: up to 10 Color Bits: 2.0 bits Type Selection: the most similar members Feature Display: ATP binding site

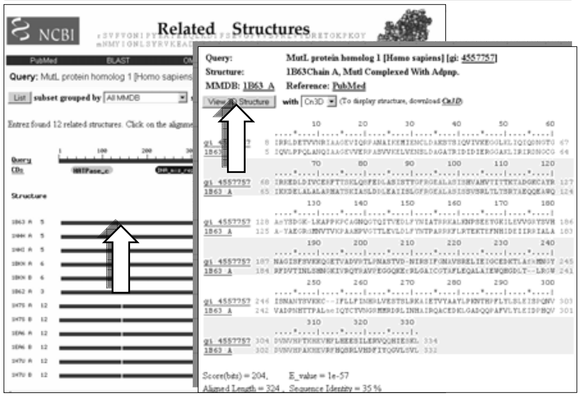
Mapping Polymorphisms onto Structure

GeneView: Variations Human MLH1



Contig position	dbSNP rs# cluster id	Hetero-zygosity	Validation	3D	OMIM	Function	dbSNP allele	Protein residue	Codon position	Amino acid position
36975081	rs1800143	N.D.		Yes		synonymous	A	Glu [E]	3	13
		N.D.		Yes		contig reference	G	Glu [E]	3	13
36975136	rs2020872	0.025		Yes		nonsynonymous	G	Val [V]	1	32
		0.025		Yes		contig reference	A	Ile [I]	1	32
36982507	rs11541859	N.D.		Yes		nonsynonymous	C	Gln [Q]	1	89
		N.D.		Yes		contig reference	G	Glu [E]	1	89

Related Structures



Query: MutL protein homolog 1 [Homo sapiens]

List subset grouped by All MMDB

Entrez found 12 related structures. Click on the alignme

Query: MutL protein homolog 1 [Homo sapiens] [gi: 4557757]

Structure: 1B63Chain A, Mutl Complexed With Adpnp.

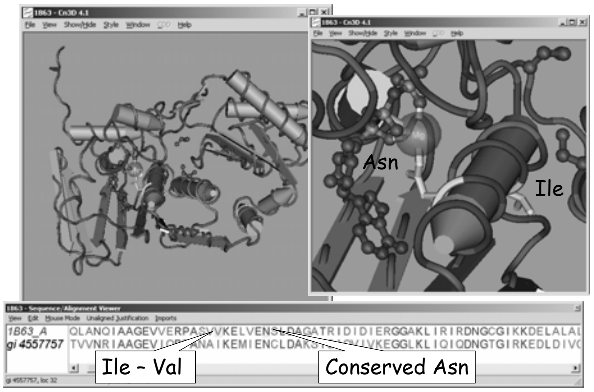
MMDB: 1B63_A Reference: PubMed

View 3D Structure with Cn3D (To display structure, download Cn3D)

Score(bits) = 204, E_value = 1e-57

Aligned Length = 324 , Sequence Identity = 35 %

Mapping Variation Onto Structure

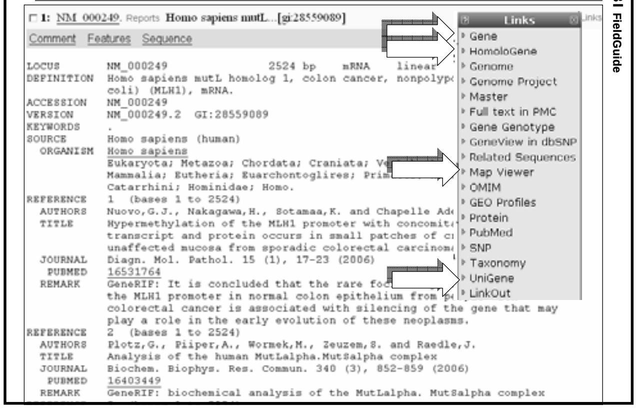


```
1B63_A      QLANQIAAGEVVERPASVVKELVENSLDAGATRIDIDIERGGAKLIRIRDNGCGIKKDELALAL
gi 4557757  TVVNRIAAGEVIQRPANAIKEMIENCLDAKSTSIQVIVKEGGLKLIQIQDNGTGIRKEDLDIVC
```

gi 4557757, loc 32

Genome Resources

NM_000249: Genome Links

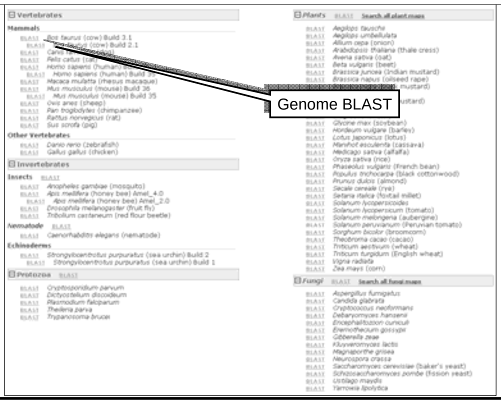


☐ 1: NM_000249. Reports **Homo sapiens mutL...[gi:28559089]**

Comment Features Sequence

```
LOCUS       NM_000249                2524 bp    mRNA    linear
DEFINITION  Homo sapiens mutL homolog 1, colon cancer, nonpolyp
            coli) (MLH1), mRNA.
ACCESSION   NM_000249
VERSION     NM_000249.2  GI:28559089
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Ve
            Mammalia; Eutheria; Euarchontoglires; Prim
            Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2524)
  AUTHORS   Nuovo,G.J., Nakagawa,H., Sotamaa,K. and Chapelle Ade
  TITLE     Hypermethylation of the MLH1 promoter with concomit
            transcript and protein occurs in small patches of c
            unaffected mucosa from sporadic colorectal carcinoma
  JOURNAL   Diagn. Mol. Pathol. 15 (1), 17-23 (2006)
   PUBMED   16531764
  REMARK    GeneRIF: It is concluded that the rare foc
            the MLH1 promoter in normal colon epithelium from p
            colorectal cancer is associated with silencing of the gene that may
            play a role in the early evolution of these neoplasms.
REFERENCE   2  (bases 1 to 2524)
  AUTHORS   Plotz,G., Piiper,A., Wormek,M., Zeuzem,S. and Raedle,J.
  TITLE     Analysis of the human MutLalpha.MutSalpha complex
  JOURNAL   Biochem. Biophys. Res. Commun. 340 (3), 852-859 (2006)
   PUBMED   16403449
  REMARK    GeneRIF: biochemical analysis of the MutLalpha. MutSalpha complex
```

The Map Viewer



Genome BLAST

⊟ Vertebrates

Mammals
- BLAST Bos taurus (cow) Build 3.1
  - BLAST (cow) Build 2.1
- BLAST Canis
- BLAST Felis catus (cat)
- BLAST Homo sapiens (human)
  - BLAST Homo sapiens (human) Build 35
- BLAST Macaca mulatta (rhesus macaque)
- BLAST Mus musculus (mouse) Build 36
  - BLAST Mus musculus (mouse) Build 35
- BLAST Ovis aries (sheep)
- BLAST Pan troglodytes (chimpanzee)
- BLAST Rattus norvegicus (rat)
- BLAST Sus scrofa (pig)

Other Vertebrates
- BLAST Danio rerio (zebrafish)
- BLAST Gallus gallus (chicken)

⊟ Invertebrates

Insects BLAST
- BLAST Anopheles gambiae (mosquito)
- BLAST Apis mellifera (honey bee) Amel_4.0
  - BLAST Apis mellifera (honey bee) Amel_2.0
- BLAST Drosophila melanogaster (fruit fly)
- BLAST Tribolium castaneum (red flour beetle)

Nematode BLAST
- BLAST Caenorhabditis elegans (nematode)

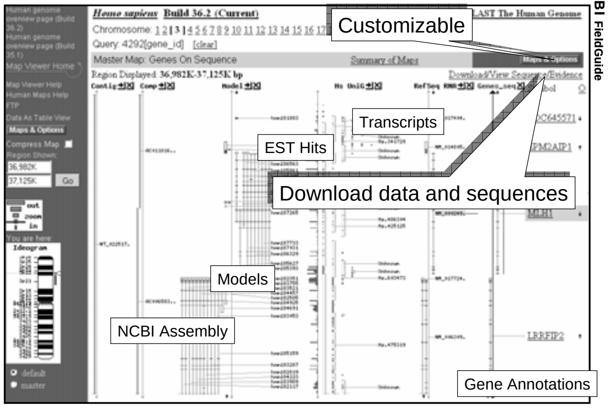
Echinoderms
- BLAST Strongylocentrotus purpuratus (sea urchin) Build 2
  - BLAST Strongylocentrotus purpuratus (sea urchin) Build 1

⊟ Protozoa BLAST
- BLAST Cryptosporidium parvum
- BLAST Dictyostelium discoideum
- BLAST Plasmodium falciparum
- BLAST Theileria parva
- BLAST Trypanosoma brucei

⊟ Plants BLAST Search all plant maps
- BLAST Aegilops tauschii
- BLAST Aegilops umbellulata
- BLAST Allium cepa (onion)
- BLAST Arabidopsis thaliana (thale cress)
- BLAST Avena sativa (oat)
- BLAST Beta vulgaris (beet)
- BLAST Brassica juncea (Indian mustard)
- BLAST Brassica napus (oilseed rape)
- BLAST mustard)
- ustard)
- BLAST Glycine max (soybean)
- BLAST Hordeum vulgare (barley)
- BLAST Lotus japonicus (lotus)
- BLAST Manihot esculenta (cassava)
- BLAST Medicago sativa (alfalfa)
- BLAST Oryza sativa (rice)
- BLAST Phaseolus vulgaris (French bean)
- BLAST Populus trichocarpa (black cottonwood)
- BLAST Prunus dulcis (almond)
- BLAST Secale cereale (rye)
- BLAST Setaria italica (foxtail millet)
- BLAST Solanum lycopersicoides
- BLAST Solanum lycopersicum (tomato)
- BLAST Solanum melongena (aubergine)
- BLAST Solanum peruvianum (Peruvian tomato)
- BLAST Sorghum bicolor (broomcorn)
- BLAST Theobroma cacao (cacao)
- BLAST Triticum aestivum (wheat)
- BLAST Triticum turgidum (English wheat)
- BLAST Vigna radiata
- BLAST Zea mays (corn)

⊟ Fungi BLAST Search all fungi maps
- BLAST Aspergillus fumigatus
- BLAST Candida glabrata
- BLAST Cryptococcus neoformans
- BLAST Debaryomyces hansenii
- BLAST Encephalitozoon cuniculi
- BLAST Eremothecium gossypii
- BLAST Gibberella zeae
- BLAST Kluyveromyces lactis
- BLAST Magnaporthe grisea
- BLAST Neurospora crassa
- BLAST Saccharomyces cerevisiae (baker's yeast)
- BLAST Schizosaccharomyces pombe (fission yeast)
- BLAST Ustilago maydis
- BLAST Yarrowia lipolytica

Map Viewer: Human MLH1



Customizable

Transcripts

EST Hits

Download data and sequences

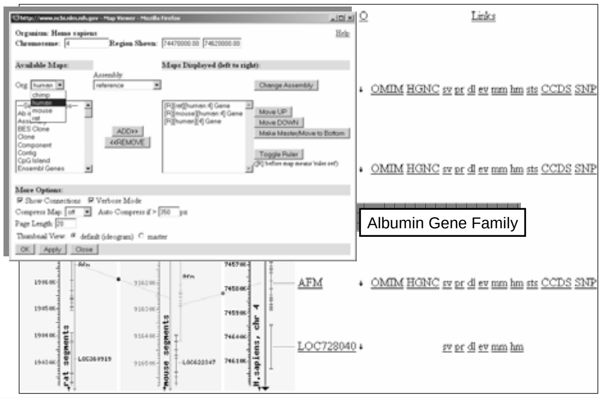
Models

NCBI Assembly

Gene Annotations

Synteny: Mammalian Genomes

NCBI FieldGuide

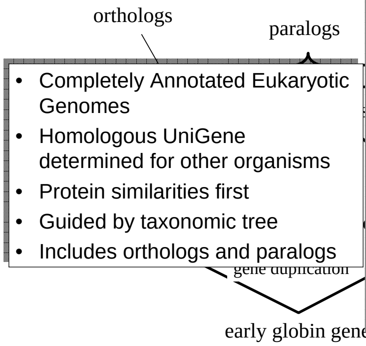


Organism: Homo sapiens
Chromosome: 4 Region Shown: 74470000.00 74620000.00
Org: human
Assembly: reference
Maps Displayed (left to right):
[R][rat][human:4] Gene
[R][mouse][human:4] Gene
[R][human][4] Gene
☑ Show Connections ☑ Verbose Mode
Compress Map: off Auto Compress if > 350 px
Page Length: 20
Thumbnail View: ● default (ideogram) ○ master

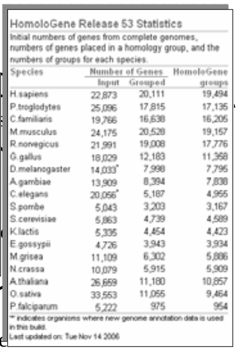
Albumin Gene Family

Homologene

NCBI FieldGuide



- Completely Annotated Eukaryotic Genomes
- Homologous UniGene determined for other organisms
- Protein similarities first
- Guided by taxonomic tree
- Includes orthologs and paralogs



HomoloGene Release 53 Statistics

Initial numbers of genes from complete genomes, numbers of genes placed in a homology group, and the numbers of groups for each species.

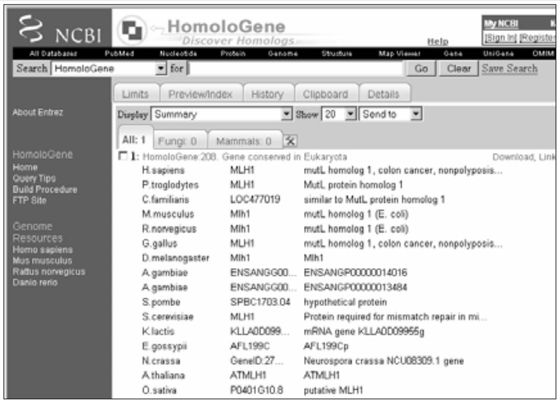
Species	Number of Genes		HomoloGene
	Input	Grouped	groups
H.sapiens	22,873	20,111	19,494
P.troglodytes	25,096	17,815	17,135
C.familiaris	19,766	16,638	16,205
M.musculus	24,175	20,528	19,157
R.norvegicus	21,991	19,008	17,776
G.gallus	18,029	12,183	11,358
D.melanogaster	14,033*	7,998	7,796
A.gambiae	13,909	8,394	7,838
C.elegans	20,056*	5,187	4,955
S.pombe	5,043	3,203	3,167
S.cerevisiae	5,863	4,739	4,589
K.lactis	5,335	4,454	4,423
E.gossypii	4,726	3,943	3,934
M.grisea	11,109	6,302	5,886
N.crassa	10,079	5,915	5,909
A.thaliana	26,659	11,180	10,857
O.sativa	33,553	11,055	9,464
P.falciparum	5,222	975	954

"*" indicates organisms where new genome annotation data is used in this build.

Last updated on: Tue Nov 14 2006

Homologene Cluster

NCBI FieldGuide



Search HomoloGene for

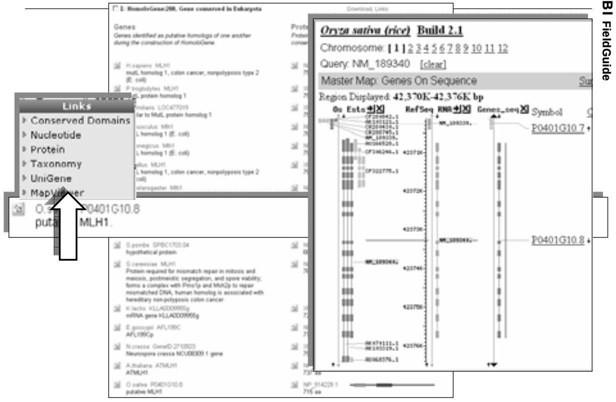
Display Summary Show 20 Send to

☐ 1: HomoloGene:208. Gene conserved in Eukaryota

H.sapiens	MLH1	mutL homolog 1, colon cancer, nonpolyposis...
P.troglodytes	MLH1	MutL protein homolog 1
C.familiaris	LOC477019	similar to MutL protein homolog 1
M.musculus	Mlh1	mutL homolog 1 (E. coli)
R.norvegicus	Mlh1	mutL homolog 1 (E. coli)
G.gallus	MLH1	mutL homolog 1, colon cancer, nonpolyposis...
D.melanogaster	Mlh1	Mlh1
A.gambiae	ENSANGG00...	ENSANGP00000014016
A.gambiae	ENSANGG00...	ENSANGP00000013484
S.pombe	SPBC1703.04	hypothetical protein
S.cerevisiae	MLH1	Protein required for mismatch repair in mi...
K.lactis	KLLA0D099...	mRNA gene KLLA0D09955g
E.gossypii	AFL199C	AFL199Cp
N.crassa	GeneID:27...	Neurospora crassa NCU08309.1 gene
A.thaliana	ATMLH1	ATMLH1
O.sativa	P0401G10.8	putative MLH1

Rice Homolog

NCBI FieldGuide



S.cerevisiae MLH1
Protein required for mismatch repair in mitosis and meiosis, postmeiotic segregation, and spore viability; forms a complex with Pms1p and Mlh2p to repair mismatched DNA; human homolog is associated with hereditary non-polyposis colon cancer

O.sativa P0401G10.8
putative MLH1

The Gene Database

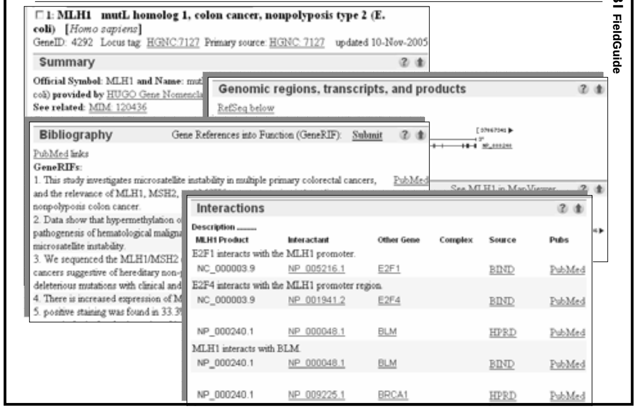
NCBI FieldGuide

- Gene Centered Information
- Unifies LocusLink and microbial Genomes
- 2.4 million records for 3,822 taxa

Human	38,603	Sea Urchin	30,603
Chimpanzee	31,502	Mosquito	13,763
Mouse	60,746	Fruit Fly	21,116
Rat	38,117	*C. elegans*	20,935
Dog	20,154	Fungi	168,802
Cow	23, 677	Green Plants	76,847
Chicken	18, 469	Archea	74,627
Zebrafish	38, 594	Bacteria	1,361,390

Genes MLH1: One Stop Shopping

NCBI FieldGuide



☐ 1: **MLH1 mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)** [*Homo sapiens*]
GeneID: 4292 Locus tag: HGNC:7127 Primary source: HGNC: 7127 updated 10-Nov-2005

Summary

Official Symbol: MLH1 **and Name**: mut
coli) **provided by** HUGO Gene Nomencla
See related: MIM: 120436

Genomic regions, transcripts, and products

RefSeq below

Bibliography Gene References into Function (GeneRIF): Submit

PubMed links
GeneRIFs:
1. This study investigates microsatellite instability in multiple primary colorectal cancers, and the relevance of MLH1, MSH2, nonpolyposis colon cancer. PubMed
2. Data show that hypermethylation o pathogenesis of hematological maligna microsatellite instability.
3. We sequenced the MLH1/MSH2 cancers suggestive of hereditary non- deleterious mutations with clinical and
4. There is increased expression of M
5. positive staining was found in 33.3

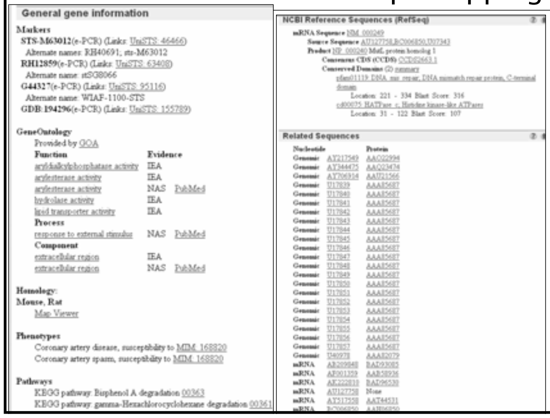
Interactions

Description

MLH1 Product	Interactant	Other Gene	Complex	Source	Pubs
E2F1 interacts with the MLH1 promoter.					
NC_000003.9	NP_005216.1	E2F1		BIND	PubMed
E2F4 interacts with the MLH1 promoter region.					
NC_000003.9	NP_001941.2	E2F4		BIND	PubMed
NP_000240.1	NP_000048.1	BLM		HPRD	PubMed
MLH1 interacts with BLM.					
NP_000240.1	NP_000048.1	BLM		BIND	PubMed
NP_000240.1	NP_009225.1	BRCA1		HPRD	PubMed

Genes MLH1: One Stop Shopping (cont.)

NCBI FieldGuide



General gene information

Markers
STS-M63012(e-PCR) (Links: UniSTS: 46466)
Alternate names: RH40691; sts-M63012
RH12859(e-PCR) (Links: UniSTS: 63408)
Alternate name: stSG8066
G44327(e-PCR) (Links: UniSTS: 95116)
Alternate name: WIAF-1100-STS
GDB:194296(e-PCR) (Links: UniSTS: 155789)

GeneOntology
Provided by GOA

Function	**Evidence**	
aryldialkylphosphatase activity	IEA	
arylesterase activity	IEA	
arylesterase activity	NAS	PubMed
hydrolase activity	IEA	
lipid transporter activity	IEA	
Process		
response to external stimulus	NAS	PubMed
Component		
extracellular region	IEA	
extracellular region	NAS	PubMed

Homology:
Mouse, Rat
Map Viewer

Phenotypes
Coronary artery disease, susceptibility to MIM: 168820
Coronary artery spasm, susceptibility to MIM: 168820

Pathways
KEGG pathway: Bisphenol A degradation 00363
KEGG pathway: gamma-Hexachlorocyclohexane degradation 00361

NCBI Reference Sequences (RefSeq)

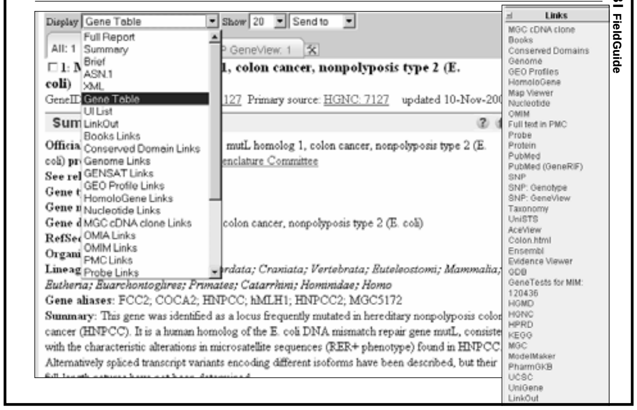
mRNA Sequence NM_000249
Source Sequence AU127758,BC006850,U07343
Product NP_000240 MutL protein homolog 1
Consensus CDS (CCDS) CCDS2663.1
Conserved Domains (2) summary
pfam01119: DNA_mis_repair; DNA mismatch repair protein, C-terminal domain
Location: 221 - 334 Blast Score: 316
cd00075: HATPase_c; Histidine kinase-like ATPases
Location: 31 - 122 Blast Score: 107

Related Sequences

Nucleotide **Protein**

Genes: Display Options and Links

NCBI FieldGuide



Display: Gene Table; Show: 20; Send to

☐ 1: M... 1, colon cancer, nonpolyposis type 2 (E. coli)
GeneID ...127 Primary source: HGNC: 7127 updated 10-Nov-200

Official ... mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) pr... enclature Committee
Gene d... colon cancer, nonpolyposis type 2 (E. coli)
Lineag... rdata; Craniata; Vertebrata; Euteleostomi; Mammalia; *Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo*
Gene aliases: FCC2; COCA2; HNPCC; hMLH1; HNPCC2; MGC5172
Summary: This gene was identified as a locus frequently mutated in hereditary nonpolyposis colon cancer (HNPCC). It is a human homolog of the E. coli DNA mismatch repair gene mutL, consiste with the characteristic alterations in microsatellite sequences (RER+ phenotype) found in HNPCC. Alternatively spliced transcript variants encoding different isoforms have been described, but their