BMG 744
Bioinformatics/Genomics/Proteomics

The “ics” revolution
March 4, 2003
Contact Information

Elliot Lefkowitz

✓ Email
  – ElliotL@uab.edu

✓ Web Site
  – http://www.genome.uab.edu

✓ Office
  – BBRB 277A

✓ Phone
  – 934-1946
Molecular and Genetic Bioinformatics Facility

General Information

- UAB Bioinformatics Resources
- Sequence Analysis at UAB
- MIC753 - "Practical Applications of Sequence Analysis"
- CIS 640 - Bioinformatics I "Lectures on Practical Bioinformatics" pdf

Genomic Sequencing

- The Poxvirus Bioinformatics Resource
- The Streptococcus pneumoniae genome diversity project
- The Streptococcus pneumoniae strain SpP6 genome sequencing project
- The Ureaplasma urealyticum genome sequencing project

UAB Only (Password required. Call or Email Elliot for access)

- CCG at UAB
- SeqWeb - Web interface to CCG
- CCG 10 Documentation
- CCG 10 Documentation - Downloadable pdf files

For information contact:

<table>
<thead>
<tr>
<th>Elliot Lefkowitz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone: 205-934-1946</td>
</tr>
<tr>
<td>Email: <a href="mailto:EliotL@uab.edu">EliotL@uab.edu</a></td>
</tr>
<tr>
<td>Office: PBEB 277A</td>
</tr>
</tbody>
</table>

Funding for the Molecular and Genetic Bioinformatics Facility has been provided in part by the UAB Health Services Foundation and the UAB Center for AIDS Research.
 Definitions  (From NIH)

✓ Bioinformatics
  – Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.

✓ Computational Biology
  – The development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems.
Elliot’s Definition of Bioinformatics

✓ Computer-aided analysis of biological information

✓ Caveat:
  – In the end, bioinformatics (a.k.a. computers) can only help in making inferences concerning biological processes.
  – These inferences (or hypotheses) have to be tested in the laboratory
Why Bioinformatics?

✔ Assists in the understanding of the basic genetic components of living systems.

✔ Understanding Genetics
  - Sequence
  - Genetic organization
  - Comparative analysis
  - Variation

✔ Structure Prediction

✔ Functional Inference
  - Enzymatic/Regulatory/Structural
  - Expression
  - Protein interactions
Why Bioinformatics Now?

✓ Data overload
  – Genomics
  – Gene expression arrays
  – Proteomics
  – Structural biology
  – Future biological revolutions
Genomics
The Human Genome Project

- Mapping and Sequencing the Genomes of Model Organisms
- Data Collection and Distribution
- Ethical, Legal, and Social Considerations
- Research Training
- Technology Development
- Technology Transfer
Genomes of Humans and their “cousins”

- Eukaryotic
- Prokaryotic
- Archaea
- Viruses
Human Genome Sequencing as of 1/5/2003
95.8% finished; chromosomes 6, 7, 13, 20, 21, 22 and Y are considered complete

<table>
<thead>
<tr>
<th>Chr</th>
<th>Tiling Path Clones</th>
<th>Sequenced Clones Draft (%)</th>
<th>Sequenced Clones Finished (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2202</td>
<td>71 (3.2)</td>
<td>2089 (94.9)</td>
</tr>
<tr>
<td>2</td>
<td>1966</td>
<td>14 (0.7)</td>
<td>1948 (99.1)</td>
</tr>
<tr>
<td>3</td>
<td>1740</td>
<td>82 (4.7)</td>
<td>1630 (93.7)</td>
</tr>
<tr>
<td>4</td>
<td>1629</td>
<td>47 (2.9)</td>
<td>1572 (96.5)</td>
</tr>
<tr>
<td>5</td>
<td>1775</td>
<td>54 (3.0)</td>
<td>1700 (95.8)</td>
</tr>
<tr>
<td>6</td>
<td>1795</td>
<td>2 (0.1)</td>
<td>1790 (99.7)</td>
</tr>
<tr>
<td>7</td>
<td>1526</td>
<td>1 (0.1)</td>
<td>1514 (99.2)</td>
</tr>
<tr>
<td>8</td>
<td>1239</td>
<td>79 (6.4)</td>
<td>1133 (91.4)</td>
</tr>
<tr>
<td>9</td>
<td>999</td>
<td>22 (2.2)</td>
<td>963 (96.4)</td>
</tr>
<tr>
<td>10</td>
<td>1132</td>
<td>6 (0.5)</td>
<td>1117 (98.7)</td>
</tr>
<tr>
<td>11</td>
<td>1147</td>
<td>43 (3.7)</td>
<td>1088 (94.9)</td>
</tr>
<tr>
<td>12</td>
<td>1140</td>
<td>74 (6.5)</td>
<td>1044 (91.6)</td>
</tr>
<tr>
<td>13</td>
<td>854</td>
<td>0 (0.0)</td>
<td>854 (100.0)</td>
</tr>
<tr>
<td>14</td>
<td>655</td>
<td>1 (0.2)</td>
<td>642 (98.0)</td>
</tr>
<tr>
<td>15</td>
<td>710</td>
<td>68 (9.6)</td>
<td>618 (87.0)</td>
</tr>
<tr>
<td>16</td>
<td>725</td>
<td>44 (6.1)</td>
<td>672 (92.7)</td>
</tr>
<tr>
<td>17</td>
<td>692</td>
<td>132 (19.1)</td>
<td>541 (78.2)</td>
</tr>
<tr>
<td>18</td>
<td>600</td>
<td>6 (1.0)</td>
<td>585 (97.5)</td>
</tr>
<tr>
<td>19</td>
<td>861</td>
<td>9 (1.0)</td>
<td>852 (99.0)</td>
</tr>
<tr>
<td>20</td>
<td>632</td>
<td>0 (0.0)</td>
<td>632 (100.0)</td>
</tr>
<tr>
<td>21</td>
<td>473</td>
<td>0 (0.0)</td>
<td>472 (99.8)</td>
</tr>
<tr>
<td>22</td>
<td>527</td>
<td>0 (0.0)</td>
<td>527 (100.0)</td>
</tr>
<tr>
<td>X</td>
<td>1588</td>
<td>37 (2.3)</td>
<td>1508 (95.0)</td>
</tr>
<tr>
<td>Y</td>
<td>200</td>
<td>0 (0.0)</td>
<td>200 (100.0)</td>
</tr>
<tr>
<td>total</td>
<td>26807</td>
<td>792 (3.0)</td>
<td>25691 (95.8)</td>
</tr>
</tbody>
</table>
Genome Project Organization

- Cloning
- Mapping
- Sequencing
- Annotation
- Analysis
Bioinformatic Information Flow

✓ “Raw” data generation
  – Sequence generation and assembly

✓ Analytical tools
  – Pattern matching

✓ Database generation
  – Construction and data import

✓ Visualization (publication) of results
  – Static: Table or graph
  – Dynamic: Web page/Java applet
Annotation and Analysis

- Gene prediction
  - Identify patterns characteristic of ORFs
- Functional assignment
  - Similarity searching
- Metabolic pathway modeling
- Comparative analysis
  - Identification and comparison with related genes
“Similar” sequences
- Sequences related by primary sequence similarity

Homologs
- Sequences related by evolution
  - Orthologs
    - Related due to speciation
  - Paralogs
    - Related due to gene duplication
Bioinformatic Databases

Something to compare against
Sequence Databases

✔ DNA
  - Genbank (NCBI)
  - EMBL
  - DDBJ

✔ Protein
  - PIR
  - Swiss-Prot
  - Swiss-Prot TrEMBL
  - PDB
Other Databases

✔ Structural
  – Protein Data Bank (PDB): http://www.rcsb.org/pdb/

✔ Expression
  – Microarray Gene Expression Data Society (MGED): http://www.mged.org/
  – Gene Expression Omnibus (GEO – NCBI)

✔ Proteomic
  – Mascot: http://www.matrixscience.com/

✔ Metabolism
  – BioCyc: http://biocyc.org/

✔ Ontology
Genbank

- Primary nucleic acid sequence database
- Maintained by NCBI
  - National Center for Biotechnology Information
- February 15, 2003; Release 134
  - 29,358,082,791 bases
  - 23,035,823 sequences
Growth of GenBank

- Sequences (millions)
- Base Pairs of DNA (millions)


Base Pairs
Sequences
Genbank Divisions

1. PRI - primate sequences
2. ROD - rodent sequences
3. MAM - other mammalian sequences
4. VRT - other vertebrate sequences
5. INV - invertebrate sequences
6. PLN - plant, fungal, and algal sequences
7. BCT - bacterial sequences
8. VRL - viral sequences
9. PHG - bacteriophage sequences
10. SYN - synthetic sequences
11. UNA - unannotated sequences
12. EST - EST sequences (expressed sequence tags)
13. PAT - patent sequences
14. STS - STS sequences (sequence tagged sites)
15. GSS - GSS sequences (genome survey sequences)
16. HTG - HTGS sequences (high throughput genomic sequences)
17. HTC - unfinished high-throughput cDNA sequencing
High Throughput Genomic Sequences

✓ ‘Unfinished' DNA sequences generated by the high-throughput sequencing centers

✓ Phase 0
  - Single-few pass reads of a single clone (not contigs)

✓ Phase 1
  - Unfinished, may be unordered, unoriented contigs, with gaps

✓ Phase 2
  - Unfinished, ordered, oriented contigs, with or without gaps

✓ Phase 3
  - Primary division (Genbank)
  - Finished, no gaps (with or without annotations)
EST

✓ Expressed Sequence Tags
  – “Single-pass" cDNA sequences
  – Generally representative of the 3’ ends of cDNAs
  – More “full-length” ESTs now available

✓ NCBI also has a dbEST database
  – Same content
  – Different format
STS

✓ Sequence Tagged Sites
  – Sequence and mapping data
  – Short genomic landmark sequences

✓ NCBI also has a dbSTS database
  – Same content
  – Different format
GSS

✓ Genome Survey Sequences
✓ Similar to the EST division, except that its sequences are genomic in origin, rather than cDNA
  – Random “single pass read” genome survey sequences.
  – Cosmid/BAC/YAC end sequences
  – Exon trapped genomic sequences
  – alu PCR sequences
✓ NCBI also has a dbGSS database
  – Same content
  – Different format
Other NCBI Databases

- RefSeq
- Unigene
- HomoloGene
- Genomic
- dbSNP
RefSeq

✓ NCBI Reference Sequence project
✓ Provides reference sequence standards for the naturally occurring molecules from chromosomes to mRNAs to proteins
✓ Stable reference point for:
  – mutation analysis
  – gene expression studies
  – polymorphism discovery
✓ Accession numbers have two letters, an underscore, and six numbers
  – NM_123456
RefSeq...

- Curated RefSeq
  - transcripts and proteins
- Genome Annotation
  - contigs, transcripts, and proteins
- Complete Genomes
  - genomes, chromosomes, and proteins
Unigene

- GenBank sequences partitioned into a non-redundant set of gene-oriented clusters
  - Each UniGene cluster contains sequences that represent a unique gene, as well as related information such as the tissue types in which the gene has been expressed and map location.
- Includes EST and complete cDNA sequences
- Provides information on differentially-spliced transcripts
# Unigene Organisms

## Vertebrata

### Mammalia
- **Bos taurus** (cow) 12,808 entries
- **Homo sapiens** (human) 138,826 entries
- **Mus musculus** (mouse) 90,444 entries
- **Rattus norvegicus** (rat) 63,253 entries
- **Sus scrofa** (pig) 14,344 entries

### Aves
- **Gallus gallus** (chicken) 5,068 entries

### Amphibia
- **Xenopus laevis** (frog) 19,512 entries

### Actinopterygii
- **Danio rerio** (zebrafish) 16,355 entries

## Urochordata

### Ascidiae
- **Ciona intestinalis** (sea squirt) 13,674 entries

## Arthropoda

### Insecta
- **Anopheles gambiae** (malaria mosquito) 3,270 entries
- **Drosophila melanogaster** (fruit fly) 14,779 entries

## Nematoda

### Chromadorea
- **Caenorhabditis elegans** 20,137 entries

## Embryophyta

### Eudicotyledons
- **Arabidopsis thaliana** (thale cress) 27,141 entries
- **Glycine max** (soybean) 8,967 entries
- **Lycopersicon esculentum** (tomato) 3,740 entries
- **Medicago truncatula** (barel medic) 5,729 entries

### Liliopsida
- **Hordeum vulgare** (barley) 7,944 entries
- **Oryza sativa** (rice) 19,223 entries
- **Triticum aestivum** (wheat) 20,454 entries
- **Zea mays** (maize) 19,512 entries

## Chlorophyta

### Chlorophyceae
- **Chlamydomonas reinhardtii** 6,448 entries
HomoloGene

Curated and calculated orthologs and homologs for genes represented in UniGene and LocusLink. Organisms include:

- Arabidopsis thaliana
- Bos taurus
- Caenorhabditis elegans
- Danio rerio
- Drosophila melanogaster
- Homo sapiens
- Hordeum vulgare
- Lycopersicon esculentum
- Medicago truncatula
- Mus musculus
- Oryza sativa
- Rattus norvegicus
- Sus scrofa
- Triticum aestivum
- Xenopus laevis
- Zea mays
Genomic DBs

- Human
- Mouse
- Rat
- Zebrafish
- Drosophila
- Nematode
- Plant genomes
- Yeast
- Malaria
- Microbial genomes
- Viruses
- Viroids
- Plasmids
- Eukaryotic organelles
dbSNP

✓ Single Nucleotide Polymorphisms
  - Single base changes
  - Small-scale insertions/deletions
  - Polymorphic repetitive elements
  - Microsatellite variation
LocusLink

✓ Provides a single query interface to curated sequence and descriptive information about genetic loci
  – Nomenclature
  – Aliases
  – Sequence accessions
  – Phenotypes
  – EC numbers
  – MIM numbers
  – UniGene clusters
  – Homology
  – Map locations
  – Web sites
OMIM

✔ Online Mendelian Inheritance in Man
✔ Database of gene-linked genetic disorders
✔ Maintained at Johns Hopkins University
  – Dr. Victor A. McKusick
What human genes are related to hypertension? Which of those genes are on chromosome 17?

List the OMIM entries that describe genes on chromosome 10.

List the OMIM entries that contain information about allelic variants.

Retrieve the OMIM record for the cystic fibrosis transmembrane conductance regulator (CFTR), and link to related protein sequence records via Entrez.

Find the OMIM record for the p53 tumor protein, and link out to related information in LocusLink and the p53 Mutation Database.
EMBL and DDBJ

✔ European Molecular Biology Laboratory
  – Hinxton, UK
  – http://www.ebi.ac.uk/

✔ DNA Data Bank of Japan
  – Mishima, Japan
  – http://www.ddbj.nig.ac.jp/
Coordination with Genbank

- Prevents duplication
- Genbank enters sequences from U.S. journals and researchers
- EMBL handles European data
- DDBJ handles Asian data
- Data exchanged daily
Accession Numbers

✓ Each sequence submitted to a database is assigned a unique primary accession number.
✓ Accession numbers do not change.
✓ If a sequence is merged with another, a new accession number is assigned, and the original number becomes a secondary accession number.
✓ Accession numbers may include version numbers:
  - AO2428.2
The Sequence Record

- Different for each database
- Locus (Name)
- Accession Number
- Keywords
- Description
- Properties
- References
- The Sequence
GenBank Sample Record

LOCUS      HUMCFTTRM    6129 bp    mRNA    PRI    15-DEC-1989
DEFINITION Human cystic fibrosis mRNA, encoding a presumed transmembrane conductance regulator (CFTR).
ACCESSION M28668
NID        g180331
KEYWORDS   cystic fibrosis; transmembrane conductance regulator.
SOURCE     Human, cDNA to mRNA.
ORGANISM  Homo sapiens
           Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 6129)
TITLE      Identification of the cystic fibrosis gene: Cloning and characterization of complementary DNA
JOURNAL    Science 245, 1066-1073 (1989)
MEDLINE    89368940
A three base-pair deletion spanning positions 1654-1656 is observed in cDNAs from cystic fibrosis patients.

<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Location/Qualifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>1. 6129</td>
</tr>
<tr>
<td></td>
<td>/organism=&quot;Homo sapiens&quot;</td>
</tr>
<tr>
<td></td>
<td>/db_xref=&quot;taxon:9606&quot;</td>
</tr>
<tr>
<td>CDS</td>
<td>133. 4575</td>
</tr>
<tr>
<td></td>
<td>/note=&quot;cystic fibrosis transmembrane conductance regulator&quot;</td>
</tr>
<tr>
<td></td>
<td>/codon_start=1</td>
</tr>
<tr>
<td></td>
<td>/db_xref=&quot;PID:g180332&quot;</td>
</tr>
<tr>
<td></td>
<td>/translation=&quot;MQRSPLEKASVVSKLFFSWTRPILRKGYRQRLSDDIYQIPSVD SADNLSEKLEREDRELASKKNPKNLVALRRCFFWRFMYGFILYLGJEVTKAQPLLL LNRFSKDIAILDDLPLTIFDFQLLLIVIGAIAVVAVLQPYIFVATVPVIVAFIMLR AYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLWLTLRAFGRQPYFETLHFHKALNHTAN WFYLSTLRWFQMRQMIEMIFVIFFIAVTFISILTTGEGEGRVGIILNLAMINMSTLQWA VNSSIDVDSLMSVSRVKFIDMTEGKPTKSTKPYKNGQLSKVMIIENSHVKKDDIW PSGGGMTVKDLTAKYTEGNAILENISFSISPQRVGLLLGRTGSKSTLLSAFLRLLN TEGEIQIDGVSWDSITLQQWRKAFGVIPQVFIFSSTFRKNDPYEQWDQEIEKAD EVGLRSVIEQFPGKLDFVLDGCGVSQHGLKQLMCLRSVLSKAKILLLDEPSAHLDP VTYQIIIRRTLKQAFADCTVILCEHRIEAMLECQQFLVIEENKVRQYDISIQKLLNERSL FRQAISPSDRVLFPHRNSSKCKSKQPQIAALKEETEEEVQDTRL&quot;</td>
</tr>
</tbody>
</table>

| BASE COUNT | 1886 a 1181 c 1330 g 1732 t |

ORIGIN
1  AATTGGAAAGC AAATGACATC ACAGCAGGTC AGAGAAAAAG GGTTGAGCGG
51  CAGGCACCCA GAGTAGTAGG TCTTTGGCAT TAGGAGCTTG AGCCCAGACG
101 GCCCTAGCAG GGACCCCCAGC GCCGAGAGA CCATGCAGAG GTCGCCTCTG
151 GAAAAGGCCA GCGTTGTCTC CAAACTTTTT TTCAGCTGGA CCAGACCAAT
201 TTTGAGGAAA GGATACAGAC AGCGCCTGGA ATTGTCAGAC ATATACCAAA
251 TCCCTTCTGT TGATTCTGCT GACAATCTAT CTGAAAAATT GGAAAGAGAA
301 TGGGATAGAG AGCTGGCTTC AAAGAAAAAT CCTAAACTCA TTAATGCCCT
351 TCGGCGATGT TTTTTCTGGA GATTTATGTT CTATGGAATC TTTTTATATT
401 TAGGGGAAGT CACCAAAGCA GTACAGCCTC TCTTACTGGG AAGAATCATA
451 GCTTCCTATG ACCCGGATAA CAAGGAGGAA CGCTCTATCG CGATTTATCT
LOCUS       HUMIFNRF1A   7721 bp   DNA             PRI       10-NOV-1992
DEFINITION  Homo sapiens interferon regulatory factor 1 gene, complete cds.
ACCESSION   L05072
NID         g184648
KEYWORDS    interferon regulatory factor 1.
SOURCE      Homo sapiens Placenta DNA.
            Homo sapiens
            Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
            Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 7721)
            Cha,Y., Sims,S.H., Romine,M.F., Kaufmann,M. and Deisseroth,A.B.
            TITLE     Human interferon regulatory factor 1: intron/exon organization
            JOURNAL   DNA Cell Biol. 11, 605-611 (1992)
            MEDLINE   93000481
FEATURES

source 1. .7721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="Placenta"
/map="5q23-q31"

exon 1. .219
/gene="IRF1"
/number=1
/note="putative"

5'UTR
/join(1. .219,1279. .1287)
/gene="IRF1"

gene
/join(1. .219,1279. .1287)
/gene="IRF1"

intron 220. .1278
/gene="IRF1"
/number=1

exon 1279. .1374
/gene="IRF1"
/number=2

CDS
/join(1288. .1374,2738. .2837,3630. .3806,3916. .3965,4073. .4202,4386. .4508,5040. .5089,6248. .6383,6670. .6794)
/gene="IRF1"
/codon_start=1
/product="interferon regulatory factor 1"
/db_xref="PID:g184649"
/translation="MPITRMRMRPWLEMOINSQIPGLIWINKEEMIFQIPWKHAAKHGWDINKDACLFRSAIHTGKYKAGEKPEPDKTWANKRFAMNSLPDIEEVKQDSRNKGSSAVRVRMLPPLTKNQRKSKSRSRASKAKRSCGDSSPDTFSDGLSSTLPDDHSSYTVPGYMQDLEVEQALTPALSPLCAVSSTLPDWHIPVEVVPDSTSDLYNFQVSPMPSTSEATTREDTECKLPRDINKLFOESEWORTNYECKSKYLLNEDCVOPTSYKREVCKF"
```
intron    1375. .2737  
        /gene="IRF1"
        /number=2
exon     2738. .2837  
        /gene="IRF1"
        /number=3
intron   2838. .3629  
        /gene="IRF1"
        /number=3
exon     3630. .3806  
        /gene="IRF1"
        /number=4
intron   3807. .3915  
        /gene="IRF1"
        /number=4
exon     3916. .3965  
        /gene="IRF1"
        /number=5
intron   3966. .4072  
        /gene="IRF1"
        /number=5

...  
exon     5040. .5089  
        /gene="IRF1"
        /number=8
intron   5090. .6247  
        /gene="IRF1"
        /number=8
exon     6248. .6383  
        /gene="IRF1"
        /number=9
intron   6384. .6669  
        /gene="IRF1"
        /number=9
exon     6670. .7656  
        /gene="IRF1"
        /number=9
exon     6670. .7656  
        /gene="IRF1"
        /number=9
3'UTR    6795. .7656  
BASE COUNT 1750 a  1946 c  2253 g  1772 t
ORIGIN
```
<table>
<thead>
<tr>
<th>LOCUS</th>
<th>HUM091226F</th>
<th>152 bp mRNA</th>
<th>EST</th>
<th>02-APR-1996</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEFINITION</td>
<td>Homo sapiens retinal fovea EST HFV091226 sequence.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACCESSION</td>
<td>L48850</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NID</td>
<td>g1254959</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEYWORDS</td>
<td>EST; expressed sequence tag.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOURCE</td>
<td>Homo sapiens (clone: EST HFV091226) age normalized retinal foveae cDNA to mRNA.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORGANISM</td>
<td>Homo sapiens Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>1 (sites)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUTHORS</td>
<td>Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITLE</td>
<td>3,400 new expressed sequence tags identify diversity of transcripts in human brain</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEDLINE</td>
<td>93364420</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>2 (sites)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITLE</td>
<td>A catalogue of genes in the cardiovascular system as identified by expressed sequence tags</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEDLINE</td>
<td>95024171</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>REFERENCES</td>
<td>3 (bases 1 to 152)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUTHORS</td>
<td>Bernstein,S.L., Borst,D.E., Neuder,M.E. and Wong,P.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TITLE</td>
<td>Characterization of a human fovea cDNA library and regional differential gene expression in the human retina</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JOURNAL</td>
<td>Genomics 32 (3), 301-308 (1996)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEATURES</td>
<td>Location/Qualifiers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>source</td>
<td>1. .152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/organism=&quot;Homo sapiens&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/note=&quot;Expressed sequence tags (first pass sequencing) from randomly selected bacteriophage clones (mRNA-cDNA) from human retinal fovea. The library is age normalized from ten sets of donor foveae 2-79 years old.&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/db_xref=&quot;taxon:9606&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/clone=&quot;EST HFV091226&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/dev_stage=&quot;age normalized&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/tissue_type=&quot;retinal foveae&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mRNA</td>
<td>&lt;1. .&gt;152</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>/standard_name=&quot;EST HFV091226&quot;</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<p>| BASE COUNT | 31 a     42 c     41 g     36 t     2 others |</p>
<table>
<thead>
<tr>
<th><strong>LOCUS</strong></th>
<th>HUMSWX153</th>
<th><strong>bp</strong></th>
<th>DNA</th>
<th><strong>STS</strong></th>
<th><strong>24-MAY-1993</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>DEFINITION</strong></td>
<td>Human chromosome X STS sWXD153; single read.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ACCESSION</strong></td>
<td>L15212</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NID</strong></td>
<td>g292645</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>KEYWORDS</strong></td>
<td>STS; primer; sequence tagged site.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>SOURCE</strong></td>
<td>Homo sapiens DNA.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ORGANISM</strong></td>
<td>Homo sapiens</td>
<td>Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>REFERENCE</strong></td>
<td>1 (bases 1 to 192)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TITLE</strong></td>
<td>Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>JOURNAL</strong></td>
<td>Genomics 14, 241-248 (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>MEDLINE</strong></td>
<td>93052321</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
COMMENT
Submitted by: David Schlessinger,
Center for Genetics in Medicine,
Washington University School of Medicine, Box 8232 4566 Scott
Avenue, St. Louis, MO 63110, USA
e-mail: davids@wugenmail.wustl.edu
Primer A:   TAAAGGGATCGCCAAGGAC
Primer B:   CTTACTCATTTGCTGGATTCTC
STS size:   85bp
Template:   600 ng/100ul
Primer:     40 pmoles/100ul
dNTPs:      100 uM
MgCl2:      1.5 mM
KCl:        100 mM
TrisHCl:    10 mM
Taq Polymerase: 0.125 U
NH4Cl:      5 mM
pH:         8.6
Total Vol:  5 ul
PCR Profile:
Denaturation:  94 degrees C for 1.00 minute(s)
Annealing:    55 degrees C for 2.00 minute(s)
Polymerization: 72 degrees C for 2.00 minute(s)
PCR Cycles:   35
Thermal Cycler:  P-E.
<table>
<thead>
<tr>
<th>FEATURES</th>
<th>Location/Qualifiers</th>
</tr>
</thead>
<tbody>
<tr>
<td>source</td>
<td>1. .192</td>
</tr>
<tr>
<td></td>
<td>/organism=&quot;Homo sapiens&quot;</td>
</tr>
<tr>
<td></td>
<td>/db_xref=&quot;taxon:9606&quot;</td>
</tr>
<tr>
<td></td>
<td>/map=&quot;Xq13-q24&quot;</td>
</tr>
<tr>
<td>STS</td>
<td>60. .144</td>
</tr>
<tr>
<td></td>
<td>/standard_name=&quot;sWXD153&quot;</td>
</tr>
<tr>
<td>primer_bind</td>
<td>60. .78</td>
</tr>
<tr>
<td>primer_bind</td>
<td>complement(123. .144)</td>
</tr>
<tr>
<td>BASE COUNT</td>
<td>72 a 26 c 60 g 29 t 5 others</td>
</tr>
<tr>
<td>ORIGIN</td>
<td>analyze%</td>
</tr>
</tbody>
</table>
Entrez Searching

✔ Search via text patterns
✔ Access to all of NCBI’s databases
  – Sequence
  – PubMed
  – OMIM
  – Linkage information
Swiss-Prot

- Protein Database
- University of Geneva
- Arranged by protein function
- Release 40.44
- February 22, 2003
- 44,864,044 amino acids 122,214 entries
- Provides annotated protein records
Swiss-Prot TrEMBL

- Translation of all EMBL Nucleic Acid coding sequences not yet present in Swiss-Prot
- Allows rapid availability without immediate annotation
- Release 21.13
- February 14, 2003
- 725,373 entries
Database begun over twenty years ago by Margaret O. Dayhoff

Originally published sequences in book form

Started with sequences derived from direct amino acid sequencing
PIR

- http://pir.georgetown.edu/
- Protein Identification Resource
  - PIR-International Protein Sequence Database (PSD)
- National Biomedical Research Foundation
- Georgetown University
- Release 75.04, March 03, 2003
- 283,290 Entries
PIR-NREF

- Non-redundant REFerence protein database
- Current Release 1.17
- March 3, 2003
- 1,159,203 Entries
iProClass Database - PIR

http://pir.georgetown.edu/iproclass/

✔ Comprehensive family relationships and structural/functional classifications and features of proteins
  – Superfamilies
  – Families
  – Domains
So I have a protein. What next?
NCBI Sequence Services

✔ Obtain sequences directly from NCBI using Entrez
  - Sequence Searches
  - Sequence Retrieval

✔ Other services
  - BLAST Searches
  - Sequence Submission
  - PubMed Searches

Sequence Similarities

✓ What other sequences have some primary sequence similarity to my query sequence?
✓ Time and cost of the search is dependent on the size of the database
  – Restrict the size of the database
BLAST

- Search a sequence database for primary sequence similarities to some query sequence
- Provides a measure of the significance of the similarity
- Does not necessarily imply common evolutionary origin
BLAST

- All search combinations possible
  - nt vs. nt database
    - blastn
  - protein vs. protein database
    - blastp
  - translated nt vs. protein database
    - blastx
  - protein vs. translated nt database
    - tblastn
  - translated nt vs. translated nt database
    - tblastx
A Few Genome Resources

✓ NCBI Genome Resources
  – National Center for Biotechnology Information

✓ Ensembl Human Genome Server
  – www.ensembl.org

✓ UCSC Human Genome Browser
  – genome.ucsc.edu
Analysis of PnP

Human Purine nucleoside phosphorylase
P00491. Purine nucleoside...[gi:130377]

LOCUS P00491 289 aa linear FRI 15-JUN-2002
DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNF).
ACCESSION P00491
VERSION P00491 GI:130377
DBSOURCE swissprot: locus PNH_HUMAN, accession P00491;
class: standard.
extras: gi: 35564, gi: 35565, gi: 190150, gi: 387638, gi: 190147,
gi: 190148, gi: 190149, gi: 66583, gi: 230387, gi: 230388
extras (non-sequence databases): Aarhus/Ghent-2DPAE108, MIM
164050, InterPro00139, PfamPF00696, PROSITEPS01240
KEYWORDS Transferase; Glycoyltransferase; Polymorphism; Disease mutation;
3D-structure.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
       Bacteria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 289)
       TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
            clone characterization
       JOURNAL Nucleic Acids Res. 12 (14), 5779-5787 (1984)
       MEDLINE 94272252
       PUBMED 5087295
       REMARK SEQUENCE FROM N.A.
       REFERENCE 2 (residues 1 to 289)
       AUTHORS Williams,S.R., Geikel,Y., Melvor,R.S. and Martin,D.M. Jr.
       TITLE A human purine nucleoside phosphorylase deficiency caused by a
FEATURES
  source 1..289
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  gene 1..289
  /gene="NP"
  /note="synonym: PHP"
  Protein 1..289
  /gene="NP"
  /product="Purine nucleoside phosphorylase"
  /EC_number="2.4.2.1"
  Region 7
  /gene="NP"
  /region_name="Hydrogen bonded turn"
  Region 8..17
  /gene="NP"
  /region_name="Helical region"
  Region 18..20
  /gene="NP"
  /region_name="Hydrogen bonded turn"
  Region 27..31
  /gene="NP"
  /region_name="Beta-strand region"
  Region 36..41
  /gene="NP"
  /region_name="Helical region"
  Region 46..48
  /gene="NP"
  /region_name="Beta-strand region"
  Region 50..52
  /gene="NP"
  /region_name="Helical region"
  Region 51
  /gene="NP"
  /region_name="Variant"
  /note="S -> G (IN DBSNP:1045364), /FTId=VAR_002243."
  Region 54..55
  /gene="NP"
/gene="NP"
/region_name="Helical region"
Region
231..232
  /gene="NP"
  /region_name="Hydrogen bonded turn"
Region
234..244
  /gene="NP"
  /region_name="Beta-strand region"
Region
234
  /gene="NP"
  /region_name="Variant"
  /note="R -> P (IN NP DEFICIENCY). /FTId=VAR_002247."
Region
263..278
  /gene="NP"
  /region_name="Helical region"
Region
279..281
  /gene="NP"
  /region_name="Helical region"

ORIGIN

1 mengtyedy kntaewllsh tkhrpqvaii cgsqglgltd kitgqiqfdy seipnpfrst
  61 vphagrlyf gfingrracvm mqgrfifyeg yplwqtvfvp rvfhlqtvjt lvvtncaggl
 121 npkfevgdmin lirdhinlpq fgqvgplrcp nderfgrfip amdsdayrnm rqralslwkq
 181 mgeqrlqeg tvymvagpsf etvaercrlq klggadagmnvs tpeeviwarh cglrvfgfsl
 241 itnkvimdyse slekanheev laagkqaaqk leqfvsilma sipipdkas

//

Revised: August 5, 2002.
**NCBI Sequence Viewer**

**Search** Protein  

**Display** FASTA

**LOCUS** P00419

**DEFINITION** Nucleoside phosphorylase (Inosine phosphorylase) (PNP).

**FEATURES**
- **Name:** Nucleoside phosphorylase (Inosine phosphorylase) (PNP).
- **Type:** enzyme.
- **Accession:** gi:130377
- **Length:** 289 aa
- **Status:** linear
- **Date:** PRI 15-JUN-2002

**SUMMARY**

- **Source:** NCBI
- **Location:** g310501/2022
- **Report:** P00419

**KEYWORDS**
- Transferase; Glycosyltransferase; Polymorphism; Disease mutation; 3D-structure.

**SOURCE**
- Homo sapiens (human)

**ORGANISM** Homo sapiens

- Bacteria: Meta; Chorist; Craniata; Vertebrata; Heterostomia;
- Mammalia: Eutheria; Primates; Catarrhini; Hominidae; Homo.

**REFERENCE**

1. (residues 1 to 289)

**AUTHORS**
1: P00491. Purine nucleoside...[gi:130377]

>gi|130377|sp|P00491|PHK_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (ENP)
MENGTYEYKYNTAELLLSHTKHEQVAILIGSGLGLTDKLSTQAQQEFYSEILPNFERSIT8GVHAGRLVF
GELNGRACVM4QGRFIMYEGYLKVFTEVVPVPHLLGDLTVTVAAGA6NLNPVEGDI3MLIRDHINLEG
ESGSQNEFIPDGNRFGDFRPRANSAYDRTRMQALAS7WGGQEIQRELTQTYMYAG888SETVAEACRVLQ
KLEGR4AVMGSTYPEVTVARHCGLRVPGF8ILTPNVI9MDYESTE9NH3EV1AAAGAQKQAQKLEQPGVTSILMA
SIIIPDRAS

Revised: August 5, 2002.
Searching Genomic Sequences

✓ Where is my sequence located in the human genome?
  – Chromosome; band; mapping data
  – Genetic linkage relationships

✓ What is the genomic context of my sequence?
  – Alternative splicing
  – Regulation

✓ Are there any paralogs?

✓ Are there any pseudogenes?

✓ Comparative analysis with the same gene in other genomes
BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

Database: genome Program: blastn

Use MegaBLAST

Begin Search

Enter an accession, gi, or a sequence in FASTA format:

Optional parameters
Expect Filter Descriptions Alignments

Advanced options:

Begin Search Clear Input
BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

**Database:** genome  **Program:** blastn

**Optional parameters**
- Expect
- Filter
- Descriptions
- Alignments

**Advanced options:**

**Begin Search**  **Clear Input**
BLAST the Human genome

Compare your query sequence to the working draft sequence of the human genome or its mRNA and protein products.

Database: genome Program: blastn

Use MegaBLAST

Enter an accession, gi, or a sequence in FASTA format:

Optional parameters

Expect Filter Descriptions Alignments

Advanced options:

Begin Search Clear Input
Your request has been successfully submitted and put into the Blast Queue.

Query = PNP [Homo sapiens] gi|35565|emb|CAA25320.1 | (289 letters)

The request ID is 1046634425-024719-9809

The results are estimated to be ready in 6 minutes but may be done sooner.

Please press "FORMAT" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT" again. You may also request results of a different search by entering any valid request ID to see other recent jobs.
**results of BLAST**

**Reference:**

RID: 1044624425-024719-9809

**Query:** PNP [Homo sapiens] gi|35565|emb|CAA25320.1 |
(269 letters)

**Database:** Homo sapiens genomic contig sequences
953 sequences; 2,881,589,868 total letters

If you have any problems or questions with the results of this search please refer to the BLAST FAQs

Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

**Distribution of 7 Blast Hits on the Query Sequence**

Mouse-over to show details and scores. Click to show alignments
Sequences producing significant alignments:

ref|NT_022184.10|Hs2_22340 Homo sapiens chromosome 2 genomic... 327 5e-88
ref|NT_037845.1|Hs14_37849 Homo sapiens chromosome 14 genomic... 141 9e-62
ref|NT_037734.1|Hs9_37738 Homo sapiens chromosome 9 genomic... 47 0.001

Alignments

>ref|NT_022184.10|Hs2_22340 Homo sapiens chromosome 2 genomic contig

Score = 327 bits (330), Expect = 5e-88
Identities = 132/298 (44%), Positives = 211/298 (70%), Gaps = 9/298 (3%)
Frame = -2

Query: 1
MENGYYEDYKNTAEWLL3H7KH65VQVIQC8G8SLGGSLITDLTQ4AQ1FDS1FNFPREST 60
MENGYYEDY++TAEWLL HTKH QV +IC8S LG LTDRKL QAQ1F+ SS+ NF +8T
Sbjct: 6971389 MENGYYEDYQSTAEWLLHTKH*QV6TIGSGLSGLITDLTRQ4AQ1FSD1FNFPREAT 6971210

Query: 61
VEGHAFLRLVGGFLNQKCRVYKMQRHMYNYGMYLPLMKYTVFQRFYFHLLGVDLLTIINVTAAGG 120
VEGHA LVEGFLNQVRVYKMQRHMYNYGMYLPLMKYTVFQRFYFHLLGVDLLTIINVTAAGG
Sbjct: 6971209 VEGHAY*LVFGFLNGTVQKCRVYKMQRHMYNYGMYLPLMKYTVFQRFYFHLLGVDLLTIINVTAAGG 6971030

Query: 121
NKFQVGDIMLRLDHIHNLPGISQCNELRGENDRRGDREPM8DAYTRHQRPL8TWWQ 180
NKFQVGDIMLRLDHIHNLPGISQCNELRGENDRRGDREPM8DAYTRHQRPL8TWWQ
Sbjct: 6971029 NKFQVGDIMLRLDHIHNLPGISQCNELRGENDRRGDREPM8DAYTRHQRPL8TWWQ 6970850

Query: 181
MGEPQGSL3BGTVMVAGASETVAECRVLQKLGADAVGMSITVPEVTVL--ARHGSLEVF-- 237
MGEPQGSL3BGTVMVAGASETVAECRVLQKLGADAVGMSITVPEVTVL--ARHGSLEVF--
Sbjct: 6970649 MGEPQGSL3BGTVMVAGASETVAECRVLQKLGADAVGMSITVPEVTVL--ARHGSLEVF-- 6970600

Query: 230
--------PFLITHKVIDVEWSLEKANB7SEVLAAGKQAQQKLENQFVSLILMASIPLFDKAS 289
PFLITHKVIDVEWSLEKANB7SEVLAAGKQAQQKLENQFVSLILMASIPLFDKAS
Sbjct: 6970699 PFLITHKVIDVEWSLEKANB7SEVLAAGKQAQQKLENQFVSLILMASIPLFDKAS 6970526
**Homo sapiens** genome view   **build 31**

**BLAST search** the human genome

**BLAST search results:** 3 BLAST hits found

- **Query PNP** [Homo sapiens] gi|35565|emb|CAA25320.1|

<table>
<thead>
<tr>
<th>Chr</th>
<th>Hit GI</th>
<th>Hits</th>
<th>Score</th>
<th>E value</th>
<th>Map element</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>27481183</td>
<td>1</td>
<td>327</td>
<td>5e-88</td>
<td>Homo sapiens chromosome 2 genomic contig</td>
</tr>
<tr>
<td>9</td>
<td>27481904</td>
<td>1</td>
<td>47.4</td>
<td>0.001</td>
<td>Homo sapiens chromosome 9 genomic contig</td>
</tr>
<tr>
<td>14</td>
<td>27499846</td>
<td>5</td>
<td>141</td>
<td>4e-32</td>
<td>Homo sapiens chromosome 14 genomic contig</td>
</tr>
</tbody>
</table>
Human BLAT Search

BLAT Search Genome

Genome: Human
Assembly: Human Nov. 2002
Query type: BLAT's guess
Sort output: query, score
Output type: hyperlink

BLAT's guess
DNA
protein
translated RNA
translated DNA

>ENP [Homo sapiens] gi|33565|emb|CAA25320.1
MENGTYEDYKNTAELLSHKTFRPQVAIECSGSGLGSLTDKLTPQAIFPDYSEINPPFRSPTVGHAGLWF
GFIINSGRAVWQQGFHMYEGYTLKVTFRVRFIVLLGVDTLVTNAAAGLNPKFEGVCEMLRDRHINLPG
FSGQMFPLRGPNDRFGDFPAMSMDAYDTMRQALSTMKQMGSRQRLQEGTVYVMVAGPSPETVAECRVLQ
KLGADAVGMSVTPEVIVARGCGLRVPGFSLITKVIMDYESLEKANHEEVLAIAGKQAQKLGQFVSVILMA
SIPLPKAS

Rather than pasting a sequence, you can choose to upload a text file containing the sequence.
Upload sequence: Browse... Submit File

Only DNA sequences of 25,000 or less bases and protein or translated sequence of 5000 or less letters will
**Human BLAT Results**

**BLAT Search Results**

<table>
<thead>
<tr>
<th>ACTIONS</th>
<th>QUERY</th>
<th>SCORE</th>
<th>START</th>
<th>END</th>
<th>QSIZE</th>
<th>IDENTITY</th>
<th>CHRO</th>
<th>STRAND</th>
<th>START</th>
<th>END</th>
</tr>
</thead>
<tbody>
<tr>
<td>browser details PNP</td>
<td>279</td>
<td>4</td>
<td>289</td>
<td>289</td>
<td>99.7%</td>
<td></td>
<td>14</td>
<td>++</td>
<td>14727930</td>
<td>14732220</td>
</tr>
<tr>
<td>browser details PNP</td>
<td>101</td>
<td>0</td>
<td>288</td>
<td>289</td>
<td>70.0%</td>
<td></td>
<td>2</td>
<td>+-</td>
<td>76697947</td>
<td>76698808</td>
</tr>
</tbody>
</table>
Homo sapiens Map View build 31 BLAST the Human Genome

Chromosome: 1 2 3 4 5 6 7 8 9 10 11 12 13 [14] 15 16 17 18 19 20 21 22 XY

Query: BLAST: PNP [Homo sapiens]
gi|35565|emb|CAA25320.1] [clear]

Color Key for Alignment Scores:
<40 40-50 50-80 80-200 >=200

Master Map: Contig Total Contigs On Chromosome: 7 [1 not localized]
Region Displayed: 14,727K-14,732K
bp Download/View Sequence/Evidence
Contigs Labeled: 6 Total Contigs in Region: 6

Blast hit Identity=94% query: 5.62

Blast hit Identity=97% query: 60.96
Blast hit Identity=98% query: 95.154
Blast hit Identity=100% query: 154.217

NT_037845.1

Blast hit Identity=100% query: 218.288
### Human Chromosome 14

Jump to chromosome. Jump to mapview for chromosome statistics.

#### Homology Matches

<table>
<thead>
<tr>
<th>Homo_sapiens Genes</th>
<th>Mus_musculus Homologues</th>
</tr>
</thead>
<tbody>
<tr>
<td>OSGEP</td>
<td>Q99L6e</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.79 Mb)</td>
</tr>
<tr>
<td>APEX</td>
<td>Apex1</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.79 Mb)</td>
</tr>
<tr>
<td>Q8WUC0</td>
<td>ENSMUSG0000036963</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.79 Mb)</td>
</tr>
<tr>
<td>NP</td>
<td>Pnp</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.81 Mb)</td>
</tr>
<tr>
<td>Q06PS6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>ENSG00000165787</td>
<td>4930474F22Rik</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.87 Mb)</td>
</tr>
<tr>
<td>Q8TAA1</td>
<td>ENSMUSG0000036932</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.83 Mb)</td>
</tr>
<tr>
<td>RNASE4</td>
<td>Rnase4</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.96 Mb)</td>
</tr>
<tr>
<td>ANG</td>
<td>Angrp</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 43.06 Mb)</td>
</tr>
<tr>
<td>EP3A_HUMAN</td>
<td>ENSMUSG0000021678</td>
</tr>
<tr>
<td></td>
<td>(chr 14: 42.80 Mb)</td>
</tr>
</tbody>
</table>
Detailed View

Jump to Chromosome: 2   bp: 76573222 to 76673222

Features▼ DAS Sources▼ Repeats▼ Decorations▼ Export▼ Jump to▼ Image size▼ Help▼

Length
BLAST hits
DNA(contigs)
BLAST hits
Genes
Proteins
Pat. matches
SNPs
Tisspath
SNP legend

There are currently 99 tracks switched off, use the mouse above the image to turn these on.
### UCSC Genome Browser v1.7 - Microsoft Internet Explorer

#### Mapping and Sequencing Tracks
- **Base Position**
  - on
- **Chromosome Band**
  - dense
- **STS Markers**
  - hide
- **FISH Clones**
  - hide
- **GenMapDB Clones**
  - hide
- **Recomb Rate**
  - hide
- **Map Contigs**
  - hide
- **Assembly Gap**
  - dense
- **Coverage**
  - hide
- **BAC End Pairs**
  - hide
- **Fosmid End Pairs**
  - hide
- **GC Percent**
  - hide
- **BLAT Sequence**
  - full

#### Genes and Gene Prediction Tracks
- **Known Genes**
  - pack
- **RefSeq Genes**
  - hide
- **Assembly Genes**
  - dense
- **Twisscan**
  - dense
- **SGP Genes**
  - dense
- **Fgenesh++ Genes**
  - hide
- **Gene ID Genes**
  - hide
- **GeneScan Genes**
  - pack

#### mRNA and EST Tracks
- **Human mRNAs**
  - pack
- **Spliced ESTs**
  - dense
- **Human ESTs**
  - hide
- **Nonhuman mRNA**
  - dense
- **Nonhuman EST**
  - hide
- **UniGene**
  - hide
- **Gene Boundaries**
  - hide

#### Expression and Regulation
- **CpG Islands**
  - hide

#### Comparative Genomics
- **Fugu Blat**
  - dense
- **Mouse Cons**
  - full
- **Tight Mouse**
  - hide
- **Best Mouse**
  - dense
- **Blastz Mouse**
  - hide
- **Mouse Synteny**
  - full
- **Rat Synteny**
  - hide

#### Variation and Repeats
- **Overlap SNPs**
  - dense
- **Random SNPs**
  - dense
- **RepeatMasker**
  - dense
- **Simple Repeats**
  - hide
Gene Information

- **BLink**
  - BLAST Hits

- **Domains**
  - Protein domains

- **Links**
  - Varies with available information

- **LinkOut**
  - “Custom” links to other relevant databases
P00491. Purine nucleoside...[gi:130377]

**LOCUS** P00491 269 aa linear FRI 15-JUN-2002
**DEFINITION** Purine nucleoside phosphorylase (Inosine phosphorylase) (PNF).
**ACCESSION** P00491
**VERSION** P00491 GI:130377
**DBSOURCE** swissprot; locus PNPH_HUMAN, accession P00491;
class: standard.
sequence updated: Jul 21, 1986.
xrefs: gi: 35564, gi: 33565, gi: 190149, gi: 357033, gi: 190147,
   gi: 190148, gi: 190149, gi: 65583, gi: 230387, gi: 230388
xrefs (non-sequence databases): Aarhus/Ghent-2DPAGE2108, MIM
   164050, Interpro:IPR001349, Pfam:PF00696, PROSITE:PS001240
**KEYWORDS** Transferase; Glycoehyltransferase; Polymorphism; Disease mutation;
   3D-structure.
**SOURCE** Homo sapiens (human)
**ORGANISM** Homo_sapiens
Bukaryota; Metazoa; Chordata; Cranista; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrrhini; Hominaidae; Homo.
**REFERENCE** 1 (residues 1 to 269)
   **TITLE** Human purine nucleoside phosphorylase cDNA sequence and genomic
   clone characterization
   **JOURNAL** Nucleic Acids Res. 12 (11), 5779-5787 (1984)
   **MEDLINE** 84272252
   **PUBMED** 6087295
   **REMARK** SEQUENCE FROM N.A.
   **REFERENCE** 2 (residues 1 to 269)
   **AUTHORS** Williams, S.R., Gekeler, Y., Melvor, R.S. and Martin, D.W. Jr.
   **TITLE** A human purine nucleoside phosphorylase deficiency caused by a
Query: gi|130377 purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
Matching gi: 355665, 4557801, 66583, 230387, 230388

COG0005 assigned by Cognitor (35 best hits)

<table>
<thead>
<tr>
<th>Best hits</th>
<th>Common Tree</th>
<th>Taxonomy Report</th>
<th>3D structures</th>
<th>CDD-Search</th>
<th>GI list</th>
</tr>
</thead>
</table>

148 BLAST hits to 98 unique species Sort by taxonomy proximity

Keep only Cut-Off 100 Select Reset

<table>
<thead>
<tr>
<th>Score</th>
<th>Accession</th>
<th>Gi</th>
<th>Protein Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1515</td>
<td>AAM38666</td>
<td>387033</td>
<td>purine nucleoside phosphorylase [Homo sapiens]</td>
</tr>
<tr>
<td>1501</td>
<td>BAC05327</td>
<td>217580578</td>
<td>unnamed protein product [Homo sapiens]</td>
</tr>
<tr>
<td>1341</td>
<td>188948</td>
<td>4558112</td>
<td>Chain A, Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1341</td>
<td>P56859</td>
<td>3287982</td>
<td>Purine nucleoside phosphorylase [inosine phosphorylase]</td>
</tr>
<tr>
<td>1339</td>
<td>1FXUA</td>
<td>11514560</td>
<td>Chain A, Purine Nucleoside Phosphorylase From C.</td>
</tr>
<tr>
<td>1334</td>
<td>AAB34886</td>
<td>1042206</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1332</td>
<td>IAG0</td>
<td>3002089</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1337</td>
<td>NFP</td>
<td>2624420</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1337</td>
<td>1A9T</td>
<td>3318947</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1329</td>
<td>1PNB</td>
<td>1311143</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1290</td>
<td>CAA9888</td>
<td>53750</td>
<td>chain , Bovine Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1290</td>
<td>AAO5635</td>
<td>3009261</td>
<td>purine nucleoside phosphorylase</td>
</tr>
<tr>
<td>1297</td>
<td>AAM39835</td>
<td>200028</td>
<td>purine nucleoside phosphorylase</td>
</tr>
<tr>
<td>1267</td>
<td>11425941</td>
<td>12442158</td>
<td>unnamed protein product [Mus musculus]</td>
</tr>
<tr>
<td>1001</td>
<td>XP 214155</td>
<td>27674986</td>
<td>similar to purine-nucleoside phosphorylase [Mus</td>
</tr>
<tr>
<td>814</td>
<td>EAA11700</td>
<td>21299555</td>
<td>agCP6049 [Anopheles gambiae str. PEST]</td>
</tr>
<tr>
<td>780</td>
<td>AAF47854</td>
<td>7292245</td>
<td>CG18758-PB [Drosophila melanogaster]</td>
</tr>
</tbody>
</table>
NCBI Conserved Domain Summary

Query: gi|130377|sp|P00491|ENPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (FNP) (289 letters)

Database: cdd.v.1.60

gi|CDD|4371 pfam00896, Mtap_PNP, Phosphorylase... S= 295 E=3e-:

[Diagram showing domain locations and lengths]
1: P00491. Purine nucleoside...[gi:130377]

LOCUS P00491 289 aa linear PRI 15-d
DEFINITION Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP)
ACCESSION P00491
VERSION P00491 GI:130377
DESOUCE swissprot: locus PMNH_HUMAN, accession P00491;
class: standard.
sequence updated: Jul 21, 1986.
xrefs: gi: 35564, gi: 35565, gi: 190150, gi: 387033, gi: 190147,
  gi: 190148, gi: 190149, gi: 66933, gi: 230387, gi: 230388
xrefs (non-sequence databases): Asrhus/Ghent-2DPA642108, MIM
  164050, InterProIPR001369, PfamPF00896, PROSITE01240
KEYWORDS Transerase; Glycosyltransferase; Polymorphism; Disease mutation;
3D-structure.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 289)
TITLE Human purine nucleoside phosphorylase cDNA sequence and genomic
clone characterization
*164050

NUCLEOSIDE PHOSPHORYLASE; NP

Alternative titles; symbols

PURINE-NUCLEOSIDE:ORTHO PHOSPHATE RIBOSYLTRANSFERASE; PNP
NUCLEOSIDE PHOSPHORYLASE DEFICIENCY, INCLUDED
ATAXIA WITH DEFICIENT CELLULAR IMMUNITY, INCLUDED

Gene map locus 14q13.1

TEXT

Edwards et al. (1971) described electrophoretic variants of nucleoside phosphorylase (EC 2.4.2.1), the enzyme that catalyzes the phosphorolytic cleavage of inosine to hypoxanthine. The enzyme appeared to be a trimer. Family studies indicated autosomal codominant inheritance of the variants. Zannis et al. (1978) and Williams et al. (1984) demonstrated that human PNP is a symmetric trimer composed of 3 identical 32,153-Da subunits, each with a substrate-binding site. PNP reversibly catalyzes the phosphorolysis of the purine nucleosides, (deoxy)inosine and (deoxy)guanosine, to their respective purine bases and the corresponding ribose-1-phosphate. 🤔
Deficiency of nucleoside phosphorylase results in defective T-cell immunity (Giblett et al., 1975). This may not be surprising since deficiency of adenosine deaminase, the next enzyme in the pathway, results in combined immune deficiency disease (102700). Absence of red cell NP was observed in a child with severe T-cell immunodeficiency. The parents were consanguineous and showed less than half the normal activity of the enzyme in their red cells (Berghout et al., 1975). In a patient with deficiency of nucleoside phosphorylase, Cohen et al. (1976) found severe hypouricemia and hypouricosuria, but excessive amounts of purines (mainly inosine and guanosine) in the urine. The immune defect was thought to be related to inhibition of adenosine deaminase by inosine. Mitchell et al. (1978) found that deoxycytidine and deoxyguanosine are particularly toxic to T cells but not to B cells. Addition of deoxycytidine or dipyridamole prevented deoxuribonucleoside toxicity. Stoop et al. (1977) studied a 15-month-old girl, 2 sisters of whom had died of immunodeficiency. NP was lacking from red cells and lymphocytes. The parents and a normal brother had intermediate levels. Both T cells and B cells were normal at birth, but thereafter a gradual decrease in T-cell immunity occurred. The patient showed high inosine and guanosine levels in the blood, as well as hypouricemia and hypouricosuria. Spastic tetraparesis was present. In one patient with severely defective T-cell function and normal B-cell function, Osborne et al. (1977) found no detectable red cell NP and no detectable immunologically reactive material. The parents, second cousins, had less than half the normal enzyme activity. Two patients in a second family had 0.5% residual enzyme activity and about half normal immunologically reactive material. The parents, who were not related, showed electrophoretically different mutant enzymes that were also different from those in the first family. Thus the affected children in the second family were genetic compounds, not true homozygotes. In T cells, the absence of PNP activity is thought to lead to an accumulation of deoxyguanosine triphosphate, which inhibits the enzyme ribonucleotide reductase (Mitchell et al., 1978; Ullman et al., 1979). This inhibition blocks DNA synthesis, thereby preventing the cellular proliferation required for an immune response.

The immune defect from NP deficiency is often accompanied by a neurologic disorder. Watson et al. (1981) reported the case of a 2.5-year-old boy who died of malignant lymphoma of the B-immunoblastic type. He had spastic tetraplegia also. Rijksen et al. (1987) described a case in a 3-year-old boy who was admitted for investigation of a behavior disorder and spastic diplegia. Severe lymphopenia was found; however, clinical symptoms of immune deficiency did not become apparent until the age of 4 years. Stephenson and Tolmie (1990) informed me that the family reported by Graham-Pole et al. (1975) as having 'familial dysequilibrium-diplegia with T-lymphocyte deficiency' (209000) turned out to have PNP deficiency. The condition was diagnosed retrospectively from stored fibroblasts from an affected child and from demonstration that both parents had half-normal activity of PNP. Stephenson and Tolmie (1990) were prompted to restudy this family after diagnosing PNP deficiency in a young girl who presented with dysequilibrium syndrome with pyramidal signs (extensor plantar responses and exaggerated reflexes but not prominent spasticity) very similar to the neurologic picture in the family reported by Graham-Pole et al. (1975). The child had defective cell-mediated immunity and died of lymphoma shortly after her third birthday.

Although early studies suggested that B-cell function is normal or even increased in PNP deficiency, later studies showed that B-cell function can be disrupted as well (Markert, 1991). This was the case in a patient in whom the nature of the molecular defects was demonstrated by Aust et al. (1992): she had normal B-cell counts but significantly depressed immunoglobulin levels.
ALLELIC VARIANTS
(selected examples)

.0001 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, GLU89LYS]

Williams et al. (1987) cloned the mutant gene from an NP-deficient patient who was the offspring of a consanguineous mating. A single base difference was found in the coding region of the mutant gene, a G-to-A transition in the third exon. This single base mutation altered the codon at position 89 from glu-to-lys, a result consistent with previously published peptide mapping data. The patient was demonstrated to be homozygous for the single base mutation on the basis of hybridization of synthetic oligomers to genomic DNA digests.

.0002 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ALA174PRO]

Markert and Barrett (1989) demonstrated a G-to-C change of nucleotide 520, resulting in a substitution of proline as amino acid 174. The other allele carried the mutation described by Williams et al. (1987), namely, a G-to-A change of nucleotide 265, resulting in a glu-to-lys change in amino acid 89 (164050.0001). Markert (1992) indicated that when site-directed mutagenesis was used to create this mutation and the mutant allele was expressed in COS cells, it was found to have normal function. The possibility remains, however, that the mutation was the cause of the patient's clinical disorder, with an abnormality in protein stability or other posttranscriptional stages.

.0003 NUCLEOSIDE PHOSPHORYLASE DEFICIENCY [NP, ASP128GLY]

In a patient with nucleoside phosphorylase deficiency, Aust et al. (1992) found an asp128-to-gly substitution in the maternal allele and an arg234-to-pro mutation (164050.0004) in the paternal allele. In addition, the patient was homozygous for a ser51-to-gly substitution (164050.0005), which is a polymorphism. In order to prove that the 2 mutations were responsible for the disease state, each of the 3 mutations was constructed separately by site-
REFERENCES

1. Aitken, D. A.; Ferguson-Smith, M. A.:
   Regional assignment of nucleoside phosphorylase by exclusion to 14q13.
   PubMed ID: 110525

2. Allderdice, P. W.; Miller, O. J.; Miller, D. A.; Klinger, H. P.:
   PubMed ID: 263441

   PubMed ID: 1384322


5. Carapella De Luca, E.; Stegano, M.; Dionisi Vici, C.; Paesano, R.; Fairbank, M. G. S.; Simmonds, H. A.:
   PubMed ID: 3089796

   PubMed ID: 825775
Query: gi130377 purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human
Matching gi: 35565, 4557801, 66583, 230387, 230388
COG0005 assigned by Cognitor (35 best hits)

289 aa

<table>
<thead>
<tr>
<th>SCORE</th>
<th>P</th>
<th>ACCESSION</th>
<th>GI</th>
<th>PROTEIN DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1515</td>
<td>27</td>
<td>AAA38460</td>
<td>387033</td>
<td>purine nucleoside phosphorylase [Homo sapiens]</td>
</tr>
<tr>
<td>1501</td>
<td>27</td>
<td>DAC33293</td>
<td>23359538</td>
<td>unnamed protein product [Homo sapiens]</td>
</tr>
<tr>
<td>1241</td>
<td>21</td>
<td>1851864</td>
<td>4558113</td>
<td>Chain A, Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1241</td>
<td>21</td>
<td>P30739</td>
<td>2397982</td>
<td>purine nucleoside phosphorylase [Homo sapiens]</td>
</tr>
<tr>
<td>1335</td>
<td>21</td>
<td>1851864</td>
<td>11514560</td>
<td>Chain A, Purine Nucleoside Phosphorylase From Calf Sple</td>
</tr>
<tr>
<td>1334</td>
<td>21</td>
<td>AAB34894</td>
<td>1042226</td>
<td>purine nucleoside phosphorylase [Homo sapiens]</td>
</tr>
<tr>
<td>1332</td>
<td>21</td>
<td>1A9O</td>
<td>3102082</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase Complex</td>
</tr>
<tr>
<td>1331</td>
<td>21</td>
<td>1FEN</td>
<td>2644420</td>
<td>Chain , Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1329</td>
<td>21</td>
<td>1A9T</td>
<td>3318947</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase Complex</td>
</tr>
<tr>
<td>1329</td>
<td>21</td>
<td>1FEN</td>
<td>1311183</td>
<td>Chain , Purine Nucleoside Phosphorylase</td>
</tr>
<tr>
<td>1324</td>
<td>21</td>
<td>1A9O</td>
<td>3402091</td>
<td>Chain , Bovine Purine Nucleoside Phosphorylase Complex</td>
</tr>
<tr>
<td>1290</td>
<td>21</td>
<td>CAA39888</td>
<td>55750</td>
<td>purine-nucleoside phosphorylase [Mus musculus]</td>
</tr>
<tr>
<td>1290</td>
<td>21</td>
<td>AAC37635</td>
<td>388921</td>
<td>purine nucleoside phosphorylase</td>
</tr>
<tr>
<td>1287</td>
<td>21</td>
<td>AAA93835</td>
<td>200098</td>
<td>purine nucleoside phosphorylase</td>
</tr>
<tr>
<td>1282</td>
<td>21</td>
<td>AAC37706</td>
<td>380923</td>
<td>purine nucleoside phosphorylase</td>
</tr>
<tr>
<td>1267</td>
<td>21</td>
<td>BAE255491</td>
<td>12842148</td>
<td>unnamed protein product [Mus musculus]</td>
</tr>
</tbody>
</table>
1: 1B8N_A, Chain A, Purine N...[gi:4558113]

LOCUS 1B8N_A  284 aa  linear  MAM 02
DEFINITION Chain A, Purine Nucleoside Phosphorylase.
ACCESSION 1B8N_A
VERSION 1B8N_A GI:4558113
DBSOURCE pdb: molecule 1B8N, chain A, release Feb 2, 1999;
         deposition: Feb 2, 1999;
         class: Transferase;
         source: Mol_id: 1; Organism_scientific: Bos Taurus;
         Organism_common: Bovine; Organ: Spleen;
         Exp. method: X-Ray Diffraction.

KEYWORDS .
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (residues 1 to 284)
TITLE Calf spleen purine nucleoside phosphorylase complexed with
        substrates and substrate analogues
JOURNAL Biochemistry 37 (20), 7135-7146 (1998)
MEDLINE 93254498
**Description:** Purine Nucleoside Phosphorylase.


**Taxonomy:** Bos taurus

**Reference:** PubMed  MMDB: 13072  PDB: 1B8N

**View 3D Structure** of Best Model with Cn3D Display

**Disclaimer** | Write to the Help Desk
NCBI | NLM | NIH
5 records satisfy the query 4860[loc]

- ESTs, Highly similar to PNPH_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (FNP) [H.sapiens]-Bt.3800
- ESTs, Weakly similar to PHHUPN purine-nucleoside phosphorylase (EC 2.4.2.1) [validated] - human [H.sapiens]-Dr.3216
- purine-nucleoside phosphorylase-Pnp
- nucleoside phosphorylase-NP
- SPARC-like 1-Sparc11
**HOMOLOGENE ENTRY**

<table>
<thead>
<tr>
<th>Organism</th>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mus musculus</td>
<td>Pnp</td>
<td>purine-nucleoside phosphorylase</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>SparcI</td>
<td>SPARC-like 1</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
<td>CG16758</td>
<td>Drosophila melanogaster CG16758 gene</td>
</tr>
<tr>
<td>Danio rerio</td>
<td>Dr.3216</td>
<td>ESTs, Weakly similar to PHHUPN purine-nucleoside phosphorylase (EC 2.4.2.1) [validated]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- human [H.sapiens]</td>
</tr>
<tr>
<td>Xenopus laevis</td>
<td>XI.16206</td>
<td>ESTs, Weakly similar to PNP_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens]</td>
</tr>
<tr>
<td>Homo sapiens</td>
<td>NP</td>
<td>nucleoside phosphorylase</td>
</tr>
<tr>
<td>Bos taurus</td>
<td>Bt.3800</td>
<td>ESTs, Highly similar to PNP_HUMAN Purine nucleoside phosphorylase (Inosine phosphorylase) (PNP) [H.sapiens]</td>
</tr>
</tbody>
</table>
### CALCULATED ORTHOLOGS

Listed below are the nucleotide sequence comparisons used in determining homology. The pairs below represent reciprocal best hits; each alignment is the best one for both organisms. The percent ID below represents identity over an aligned region. The value is a hyperlink to Pairwise Blast, comparing the indicated sequences. When present, red arrows (↑) point out a group of sequence matches which are part of a triplet, being consistent between more than two organisms.

<table>
<thead>
<tr>
<th>Organism-Gene</th>
<th>Sequence</th>
<th>Percent ID</th>
<th>Sequence</th>
<th>Organism-Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>M.musculus-Pnp</td>
<td>NM_013632.1</td>
<td>92.9</td>
<td>BF543947.1</td>
<td>R.norvegicus - Sparc11</td>
</tr>
<tr>
<td>M.musculus-Pnp</td>
<td>NM_013632.1</td>
<td>86.5</td>
<td>BF230403.1</td>
<td>B.taurus - Bt.3800</td>
</tr>
<tr>
<td>M.musculus-Pnp</td>
<td>NM_013632.1</td>
<td>85.0</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>M.musculus-Pnp</td>
<td>NM_013632.1</td>
<td>73.1</td>
<td>BJ060275.1</td>
<td>X.laevis - XL.16206</td>
</tr>
<tr>
<td>M.musculus-Pnp</td>
<td>NM_013632.1</td>
<td>68.7</td>
<td>BF717799.1</td>
<td>D.melanogaster - CG16758</td>
</tr>
</tbody>
</table>

### ADDITIONAL CALCULATED ORTHOLOGS

<table>
<thead>
<tr>
<th>Organism-Gene</th>
<th>Sequence</th>
<th>Percent ID</th>
<th>Sequence</th>
<th>Organism-Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.taurus-Bt.3800</td>
<td>BF230403.1</td>
<td>90.6</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>R.norvegicus-Sparc11</td>
<td>BF543947.1</td>
<td>87.6</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>B.taurus-Bt.3800</td>
<td>AV606522.1</td>
<td>84.9</td>
<td>BF543947.1</td>
<td>R.norvegicus - Sparc11</td>
</tr>
<tr>
<td>H.sapiens-NP</td>
<td>NM_000270.1</td>
<td>84.9</td>
<td>BF543947.1</td>
<td>R.norvegicus - Sparc11</td>
</tr>
<tr>
<td>R.norvegicus-Sparc11</td>
<td>BF543947.1</td>
<td>84.9</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>X.laevis-XL.16206</td>
<td>BJ676077.1</td>
<td>73.3</td>
<td>BF543947.1</td>
<td>R.norvegicus - Sparc11</td>
</tr>
<tr>
<td>B.taurus-Bt.3800</td>
<td>AV603243.1</td>
<td>71.3</td>
<td>BJ17799.1</td>
<td>D.melanogaster - CG16758</td>
</tr>
<tr>
<td>B.taurus-Bt.3800</td>
<td>AV667617.1</td>
<td>71.3</td>
<td>BF717799.1</td>
<td>D.melanogaster - CG16758</td>
</tr>
<tr>
<td>D.rerio-Dr.3216</td>
<td>BF717799.1</td>
<td>71.3</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>X.laevis-XL.16206</td>
<td>BJ060275.1</td>
<td>70.5</td>
<td>NM_000270.1</td>
<td>H.sapiens - NP</td>
</tr>
<tr>
<td>B.taurus-Bt.3800</td>
<td>BE481281.1</td>
<td>70.3</td>
<td>BJ060275.1</td>
<td>X.laevis - XL.16206</td>
</tr>
<tr>
<td>X.laevis-XL.16206</td>
<td>BJ676077.1</td>
<td>70.3</td>
<td>BJ060275.1</td>
<td>X.laevis - XL.16206</td>
</tr>
<tr>
<td>D.melanogaster - CG16758</td>
<td>BF543947.1</td>
<td>70.3</td>
<td>BJ060275.1</td>
<td>X.laevis - XL.16206</td>
</tr>
<tr>
<td>D.melanogaster - CG16758</td>
<td>BF543947.1</td>
<td>70.3</td>
<td>BJ060275.1</td>
<td>X.laevis - XL.16206</td>
</tr>
</tbody>
</table>
Welcome to the WISCONSIN PACKAGE
Version 10.3-UNIX
Installed on solaris

Copyright (c) 1982 - 2001, Accelrys Inc.
A wholly owned subsidiary of Pharmacopela, Inc. All rights reserved.

Published research assisted by this software should cite:
Wisconsin Package Version 10.3, Accelrys Inc., San Diego, CA

Databases available:

<table>
<thead>
<tr>
<th>Database</th>
<th>Release</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>GenBank</td>
<td>133.0</td>
<td>12/2002</td>
</tr>
<tr>
<td>EMBL (Abridged)</td>
<td>73.0</td>
<td>12/2002</td>
</tr>
<tr>
<td>GenPept</td>
<td>133.0</td>
<td>12/2002</td>
</tr>
<tr>
<td>PIR-Protein</td>
<td>74.05</td>
<td>12/2002</td>
</tr>
<tr>
<td>NRL 3D</td>
<td>28.0</td>
<td>01/2001</td>
</tr>
<tr>
<td>SWISS-PROT</td>
<td>40.38</td>
<td>12/2002</td>
</tr>
<tr>
<td>SP-TREMBL</td>
<td>22.0</td>
<td>10/2002</td>
</tr>
<tr>
<td>PROSITE</td>
<td>17.31</td>
<td>12/2002</td>
</tr>
<tr>
<td>Pfam</td>
<td>7.8</td>
<td>12/2002</td>
</tr>
<tr>
<td>Restriction Enzymes (REBASE)</td>
<td></td>
<td>12/2002</td>
</tr>
</tbody>
</table>

Technical support see: http://www.accelrys.com/support/

Or call Elliot at (205) 934-1946 or e-mail him at ElliotL@uab.edu

online help: % genhelp or http://genome.microbio.uab.edu/GCG/

**************
UAB Molecular and Genetic Bioinformatics Facility
For help contact Elliot Lekowitz
(205) 934-1946
ELLIOITL@uab.edu
or visit our web site: http://www.genome.uab.edu/

**************
> blast sw:pnnph_human

BLAST searches one or more nucleic acid or protein databases for sequences similar to one or more query sequences of any type. BLAST can produce gapped alignments for the matches it finds.

```
Begin (* 1 *) ?
End (* 289 *) ?
```

Search for query in what sequence database:

1) pir    p Protein Information Resource
2) nrl_3d p NRL_3D Protein Sequence-Structure Database
3) swplus p SWISS-PROT + SP-TREMBL
4) genembl n GenBank + EMBL (HTGs Removed)
5) htg    n High Throughput Genomes (HTG from GenBank and EMBL)
6) est_human n Human Expressed Sequence Tags (GenBank and EMBL)
7) est_mouse n Mouse Expressed Sequence Tags (GenBank and EMBL)
8) est_other n All Other Expressed Sequence Tags (GenBank and EMBL)
9) htc    n High Throughput cDNA (GenBank and EMBL)
10) gas    n Genome Survey Sequences (GSS from GenBank and EMBL)
11) genpept p GenPept (Translated GenBank)

Please choose one (* 1 *): 3
SW:PNPH_HUMAN  Begin: 1  End: 289
!P00491 homo sapiens (human). purine nucleoside pho... 590 e-168
SW:PNPH_BOVIN  Begin: 1  End: 284
!P55859 bos taurus (bovine). purine nucleoside phos... 521 e-147
SW:PNPH_MOUSE  Begin: 1  End: 289
!P23492 mus musculus (mouse). purine nucleoside pho... 500 e-141
SP_IN:Q9W004  Begin: 66  End: 350
!Q9w004 drosophila melanogaster (fruit fly). cg16758... 297 9e-80
SW:PNPH_YEAST  Begin: 29  End: 311
!Q05788 saccharomyces cerevisiae (baker's yeast). p...

"pnp.list" [New file] 11 lines, 509 characters
> pileup @pnp.list

PileUp creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. It can also plot a tree showing the clustering relationships used to create the alignment.

1  PNPH_HUMAN   289 aa
2  PNPH_BOVIN    284 aa
3  PNPH_MOUSE    289 aa
4  Q9W004       285 aa
5  PNPH_YEAST   283 aa

What is the gap creation penalty (* 8 *)? 5

What is the gap extension penalty (* 2 *)? 1

This program can display the clustering relationships graphically.
Do you want to:

A) Plot to a FIGURE file called "pileup.figure"
B) Plot graphics on HP7550 attached to /dev/tty15
C) Suppress the plot

Please choose one (* A *): C

What should I call the output file name (* pnp.msf *)?

Determining pairwise similarity scores...

1   x   2   4.72
1   x   3   4.45
1   x   4   2.69
1   x   5   2.52
2   x   3   4.45
2   x   4   2.70
<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>MRNSVTYBED KNTABWLLSH TKHRPQVAIL CS3GLGGLTD KLTQ....AQ</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>MRNSVTYBED ODTAVWLLSH TKHRPQAVAI CS3GLGGLVN KLTQ....AQ</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>MRNSVTYBED BNTAVWLLHQ TEYRPQVAIL CS3GLGGLAL HLEK....AQ</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>~SVTPYVSV E11ADIPIFS G5RPDKKII CS3GLGSLAD KIQD....PK</td>
</tr>
<tr>
<td><strong>ENH_YEAST</strong></td>
<td>~~~~~~~~F KHTTTFN.... .EFPRLTII CS3GLGGLST KL3RDNPFPV</td>
</tr>
<tr>
<td>51</td>
<td>100</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>IFDDSYE1NF PBTSTVQHAG KLVFGLNGR ACVM.MQRGF HM7EYFELMK</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>IFDDSYE1NF PBSTSTVQHAG KLVFGLNLR ACVM.MQRGF HM7EYFELMK</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>IFDDSYE1NF PBTSTVQHAG KLVFGLNLR ACVM.MQRGF HM7EYFELMK</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>~SBBYKTPWVF PVSTVQHAG KLVFGLREG. ATVMAMQGRF HM7EYFELAK</td>
</tr>
<tr>
<td><strong>ENH_YEAST</strong></td>
<td>TSQTVQHAG 1574STVQHAG TLQFQ4NS8 PVVL.MQRGAL HG1EGLNEFE</td>
</tr>
<tr>
<td>101</td>
<td>150</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>VTPPVSVPHL LG.VDRLWVT NAAGGLNPKF ESGDMLTRD HINLPFGS6O</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>VTPPVSVPHL LG.VDRLWVT NAAGGLNPPF ESGDMLTRD HINLPFGS6E</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>VTPPVSVPHL LG.VDRLWVT NAAGGLNPPF ESGDMLTRD HINLPFGS6E</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>CSQPV4MLKL CS.VELF1AT NAAGGLNPPF AVEDDMLMTD HVMGLFA4N</td>
</tr>
<tr>
<td><strong>ENH_YEAST</strong></td>
<td>TTPPRLV1NH MSVHNL1VT NAAG3H2AKY QACD1MCTYD H1NLPFG6AQ</td>
</tr>
<tr>
<td>151</td>
<td>200</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>NL8AG2NDER FQGDFRSAMLQ AYDRMR2QRA LSTWKQKMS6Q RLCQG3TYV6</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>NL8AG2NDER FQGDFRSAMLQ AYDRMR2QRA LSTWKQKMS6Q RLCQG3TYV6</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>NL8AG2NDER FQGDFRSAMLQ AYDRMR2QRA LSTWKQKMS6Q RLCQG3TYV6</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>NL8AG2NDER FQGDFRSAMLQ AYDRMR2QRA LSTWKQKMS6Q RLCQG3TYV6</td>
</tr>
<tr>
<td><strong>ENH_YEAST</strong></td>
<td>NL8AG2NDER FQGDFRSAMLQ AYDRMR2QRA LSTWKQKMS6Q RLCQG3TYV6</td>
</tr>
<tr>
<td>201</td>
<td>250</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>VASPTSTVUA BCRVLOILGA DAVSMSTPE VTVARHCGLR VGFS3LTN1K</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>LGGQSTSTVUA BCRVLOILGA DAVSMSTPE VTVARHCGLR VGFS3LTN1K</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>LGGQSTSTVUA BCRVLOILGA DAVSMSTPE VTVARHCGLR VGFS3LTN1K</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>LGGQSTSTVUA BCRVLOILGA DAVSMSTPE VTVARHCGLR VGFS3LTN1K</td>
</tr>
<tr>
<td><strong>ENH_YEAST</strong></td>
<td>V238PTSTVRA BCRVLOILGA DAVSMSTPE VTVARHCGLR VL2SLTNT1K</td>
</tr>
<tr>
<td>251</td>
<td>300</td>
</tr>
<tr>
<td><strong>ENH_HUMAN</strong></td>
<td>VINDY..... .ES....L...E KAHHEEVLAA GQ5AAKQLQDF PSVSLMNSTP</td>
</tr>
<tr>
<td><strong>ENH_BOVIN</strong></td>
<td>VINDY..... .ES....Q...G KAHHEEVLAA GQ5AAKQLQDF PSVSLMNSTP</td>
</tr>
<tr>
<td><strong>ENH_MOUSE</strong></td>
<td>VVMY.... .BN....L...E KAHHEEVLDD GAAACDELER PSVSLMNSTP</td>
</tr>
<tr>
<td>Q6WD4</td>
<td>SAGY..... .BD....KIID BAHDENYAV AS1KQQGACSC LV8BLRBEH</td>
</tr>
</tbody>
</table>
| **ENH_YEAST** | CVGEPDASAL D5EVE3LBS KAHHEYLNN GKLAEV1QN LIAAY1P2L...

| 301 |   |
| **ENH_HUMAN** | LPKKA3 |
| **ENH_BOVIN** | V~~~~|
| **ENH_MOUSE** | LPKKA3 |
| Q6WD4 | ~~~~~|
| **ENH_YEAST** | ~~~~~~|
Only One Final Word of Wisdom...

✓ “...although the computer is a wonderful helpmate for the sequence searcher and comparer, biochemists and molecular biologists must guard against the blind acceptance of any algorithmic output; given the choice, think like a biologist and not a statistician.”

– - Russell F. Doolittle, 1990
Farewell!