Naming Conventions

The NCBI RefSeq human mRNA database represents the best effort in defining the most complete and authentic mRNA sequences encoded by the human genome. It currently contains approximately 17,500 "NM" sequences, which have at least some cDNA sequence support, and about 10,000 "XM" sequences, the majority of which are generated by computational prediction.

NP: is for protein, Natural Protein
NM: is for mRNA, Natural Mrna
NR: is for RNA not codifing
NT: contigs (DNA)
XP or XM: these are referenced protein and mRNA seq, generated by insilico approach.
CDs: coding sequence
CON: Constructed
EST: Expressed Sequence Tag from cDNA
GSS: Genome Sequence Scan
STD: Standard
STS: Sequence Tagged Site (piccolo e unico locus genomico di 500 basi per il quale si è ottenuto un prodotto di PCR)
ENV: Environmental Samples
FUN: Fungi
HUM: Human
INV: Invertebrates
MAM: Other Mammals
MUS: Mus musculus
PHG: Bacteriophage
PLN: Plants
PRO: Prokaryotes
ROD: Rodents
SYN: Synthetic
UNC: Unclassified
VRL: Viruses
VRT: Other Vertebrates
NIH: National Institute of Health
NGI: National Institute of Genetic (Japan)
EBI: European Bioinformatics Institute

The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.
There are five different blast programs, that perform the following searches:

BLASTP compares an amino acid query sequence against a protein sequence database;

BLASTN compares a nucleotide query sequence against a nucleotide sequence database;

BLASTX compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database;

TBLASTN compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

TBLASTX compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

Interpreting Sequence Identifiers

The syntax of sequence header lines used by the NCBI BLAST server depends on the database from which each sequence was obtained. The table below lists the identifiers for the databases from which the sequences were derived.

Database Name	Identifier Syntax
GenBank	gb|accession|locus
EMBL Data Library	emb|accession|locus
DDBJ, DNA Database of Japan	dbj|accession|locus
NBRF PIR	pir||entry
Protein Research Foundation	prf||name
SWISS-PROT	sp|accession|entry name
Brookhaven Protein Data Bank	pdb|entry|chain
Kabat's Sequences of Immuno...	gnl|kabat|identifier
Patents	pat|country|number
GenInfo Backbone Id	bbs|number

BLAT on DNA is designed to quickly find sequences of 95% and greater similarity of length 40 bases or more. It may miss more divergent or shorter sequence alignments. It will find perfect sequence matches of 33 bases, and sometimes find them down to 20 bases. BLAT on proteins finds sequences of 80% and greater similarity of length 20 amino acids or more. In practice DNA BLAT works well on primates, and protein blat on land vertebrates.