

## Pathway Query: Searching for Unique Identifiers

Unique identifiers can be accession numbers, which apply to a complete sequence record, or sequence identification numbers, which apply to the individual sequences within a record.

The format of accession numbers varies, depending upon the source database. Some descriptions and examples of typical unique identifiers are given below (adapted from NCBI).

Type of Record	Sample Accession Format
Gene ID	This refers to the NCBI Entrez Gene ID. The identifier that is assigned is an integer and is species specific. For genomes that had been represented in NCBI LocusLink, the Gene ID is the same as the Locus ID. The Gene ID is reported in RefSeq records as a ?db_xref? (e.g., /db_xref=?Gene ID:856646,? in GenBank format).

Type of Record	Sample Accession Format
UniProt Accession Number	There are a total of six characters. Use capital letters. Character/Format 1[O,P,Q] 2[0-9] 3[A-Z,0-9] 4[A-Z,0-9] 5[A-Z,0-9] 6[0-9]e.g.: P12345 and Q9JJS7

After February 7, 2006:

UniProt wants to make sure that the space of available accession numbers is adequate to deal with the large expansion in the number of protein sequences and is therefore going to extend the current scheme by allowing the first character of the accession number to be any of the 26 letters (instead of only O, P and Q).



# ImmPort Pathway Queries

## Pathway Query: Searching for Unique Identifiers

Type of Record	Sample Accession Format
UniProt Accession Number	To avoid assigning accession numbers identical to those that have been used by the international nucleotide sequence database, the extension of the first position goes along with a restriction in the third position, which can only be a letter. The new format for UniProtKB accession numbers will therefore be: Character/Format 1[A-N,R-Z] 2[0-9] 3[A-Z] 4[A-Z,0-9] 5[A-Z,0-9] 6[0-9] e.g.: B1J345 and R9JJS7
Type of Record	Sample Accession Format
RefSeq Protein Accession	Two capital letters (NP), an underscore bar, and six digits, e.g.: NP_000483.



From the ImmPort home page, click on Access data, select Reference Data

The screenshot shows the ImmPort website interface. At the top, there is a navigation bar with the ImmPort logo and the text "BIOINFORMATICS FOR THE FUTURE OF IMMUNOLOGY". To the right of the logo are links for "Edit Profile" and "Sign Out". Below the logo is a navigation menu with items: "About Immport", "Admin", "Access Data", "Tools", "Resources", and "News & Events". The "Access Data" menu is open, showing options: "Submit Data", "Submission History", "Submission Resources", "Research Data", "Research Advanced Search", "Reference Data", and "Reference Advanced Search". The "Reference Data" option is selected, opening a sub-menu with options: "Genes", "Proteins", "MHC Alleles", "Pathways", "Protein Networks", "SNPs", "Immunologically", and "Related Genes". The "Pathways" option is selected, opening a sub-sub-menu with options: "ImmPort Gene Lists", "Download All Immunologically", "Related Genes", and "Reference Data History".

Below the navigation menu is a breadcrumb trail: "Home | Genes | Proteins | MHC Alleles | Pathways | Protein Networks | SNPs | ImmPort Reference Data History".

Below the breadcrumb trail is a section titled "Search Options" with the following fields:

- Search Type\*: Gene Symbol
- Search Option\*: Like
- Search Text\*: il20 (Comma delimited, 256 max chars)
- Search Species\*: Homo sapiens (Multiple)

Below the search options are buttons for "Submit" and "Cancel", and a "Results Per Page" dropdown set to 25. A "Select All" link is also present.

A speech bubble points to the search form with the text: "Enter the pathways query parameters and click submit to view the details".

## Pathway Query

Fields marked with an asterisk \* are required.

**Details of the filters on which to do a pathways query**

Search Type\*

Gene Symbol  
Gene Symbol  
Gene Name  
Gene ID  
Chromosome Region  
NCBI mRNA Accession  
NCBI Protein Accession

Search type used in the query

Search Option\*

Like  
Like  
Exact Match  
Start With  
Like

Search option operator used in the query

Search Text\*  
(Comma delimited, 256 max chars)

il20

Search text used in the query like a gene of interest

Search Species\*  
(Multiple)

Homo sapiens  
Mus musculus  
Rattus rattus  
Rattus norvegicus  
Gallus gallus  
Drosophila melanogaster  
Macaca mulatta  
Select All

Species studied can be selected individually or in multiples using the Ctrl Key. OR click 'Select All'

Search Human Build\*  
(for Homosapiens)

Build 36.1, hg18  
Build 36.1, hg18

Human build

Results Per Page

25

Results per page up to a maximum of 100 and click 'Submit'