User Guide
Database of Nonstop Mutations in Cancer
(http://NonStopDb.dkfz.de)

The database of Nonstop Mutations in Cancer (NonStopDb) is a curated database of nonstop extension and readthrough mutations in human cancer. NonStopDb allows biologists to easily extract and download nonstop mutations in cancer using multiple search options. Search criteria available in NonStopDb include the amino acid extension length, recurrence of mutation, site of mutation and cancer type. We describe each search facet below.

Search by Gene

This feature allows the user to search for nonstop mutations present in a gene of interest (Figure 1) using one of the following nomenclature:

1. Gene Name
2. Ensembl IDs.

Synonymous or alias gene names (e.g. p53 for TP53) are allowed and the search is case-insensitive.
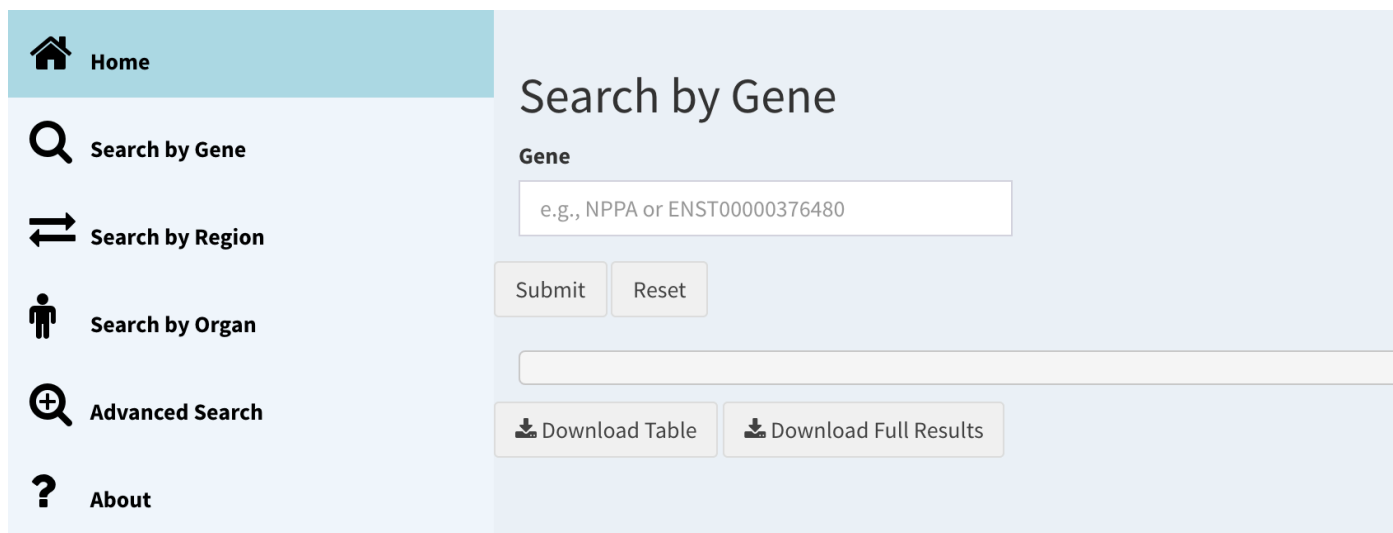

Figure 1: Search by Gene.

For example, Figure 2 shows the results page for the gene *SMAD4*. The summary information in Cancer Gene Census[1] for the gene is shown. The link to Genecards[2] for the gene is also shown.

The user can download the results using one of the two options:
- **Download Table**: This button allows the user to download the displayed results as a csv file.
- **Download Full Results**: This button allows the user to download results for all mutations along with detailed sample information plus additional information not displayed in the NonStopDb table including e.g. FATHMM pathogenicity scores and the mutation burden of the respective tumor sample (the total number of mutations annotated in COSMIC for the tumor if whole genome sequenced).

Figure 2: Results for search by gene in NonStopDb.

The result columns provide following information:
- **Gene name**: Name of the gene.
- **Mutation CDS**: Mutation position in the coding sequence and nucleotide change (For e.g., c.1659A>G indicates a change in the coding sequence at nucleotide 1659 mutated from A to G.).
- **Mutation AA**: Mutation position listed as amino acid from stop codon (*) to an amino acid in the one-letter-code.
- **Mutation ID**: Unique identifier of each mutation (as present in COSMIC database).

[1] http://cancer.sanger.ac.uk/census/
[2] http://www.genecards.org/

- **Mutation ID Count**: This column shows the recurrence level of each mutation. The number in this column represents the number of times that the respective mutation was found in different tumor samples.
- **Mutation Class**: This column shows recurrence level of each mutation. The number in this column represents number of times that the respective mutation was found in different tumor samples.
- **Extension length (AA)**: Number of amino acids appended due to a nonstop mutation including the amino acid encoded by the mutated stop codon.
- **Sample name**: Name of the tumor sample.
- **Organ System:** Organ system for the sample.
- **Site:** Organ or tissue of the sample with the respective mutation.
- **Histology:** Histology of the sample.
- **Mutation genome position**: Genomic coordinates of each respective mutation in human genome assembly GRCh38.
- **SNP COSMIC**: Information whether the mutation is also listed as SNP in the COSMIC database (yes/no).
- **CGC**: Information whether the mutated gene is also listed in the Cancer Gene Census (CGC) (yes/no).

Detailed information for each sample can be viewed by clicking on the ⊕ icon.
Figure 3 shows an example of sample information for mutation ID COSM308304.

- **Chrom**: Chromosome information.
- **Start**: Start position of the mutation.
- **End**: End position of the mutation.
- **Strand**: Strand information.
- **Transcript ID Update**: Update of the transcript ID.
- **Transcript ID updated – Mutation CDS**: Mutation position in the updated Transcript ID coding sequence and nucleotide change.
- **EXT NT Sequence incl. mut STOP**: Nucleotide sequence of the extension including the mutated position in the STOP codon (in the example below, the STOP codon TAG changed to TGG) as well as the second STOP codon (TGA in the example).
- **EXT AA Sequence incl. mut STOP**: Amino acids sequence of the extension including the mutated STOP codon (in the example below, the mutated STOP codon produced a tryptophan – first "W" in the sequence).

Figure 3: Detailed sample information for each sample having the mutation ID COSM308304 in the table of results.

Search by Region

This option allows the user to search for mutations present within the given genomic coordinates of human genome assembly GRCh38 / hg38. For example, Figure 4 shows the list of mutations present in chromosome 2 region 2000000-12000000.

Figure 4: Search by Region (genome coordinates).

Search by Organ

This option allows the search for nonstop mutations in a specific tumor entity on the basis of their site of origin in a hierarchical manner. The user first selects an organ system followed by site (tumor / organ) and histology of interest. The organ systems are divided into nine classes (shown in Figure 5): Cardiovascular System, Digestive System, Endocrine System, Genitourinary System, Integumentary System, Lymphatic System, Musculoskeletal System, Nervous System and Respiratory System.
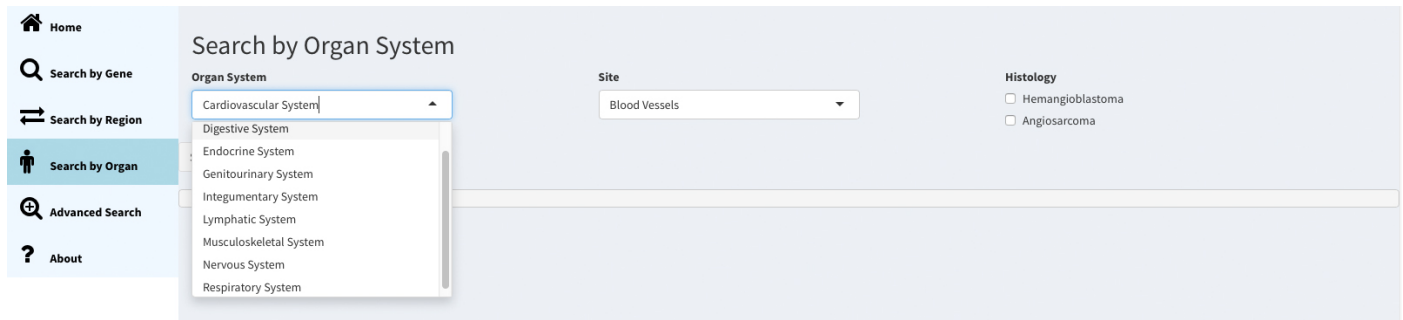
Figure 5: Search by organ - selection of organ system.

After selection of the organ system, the user selects an organ of interest / primary site and histology (Figure 6). The following example shows a list of nonstop mutations present in the digestive system followed by the choice of site and histology (Figure 7).
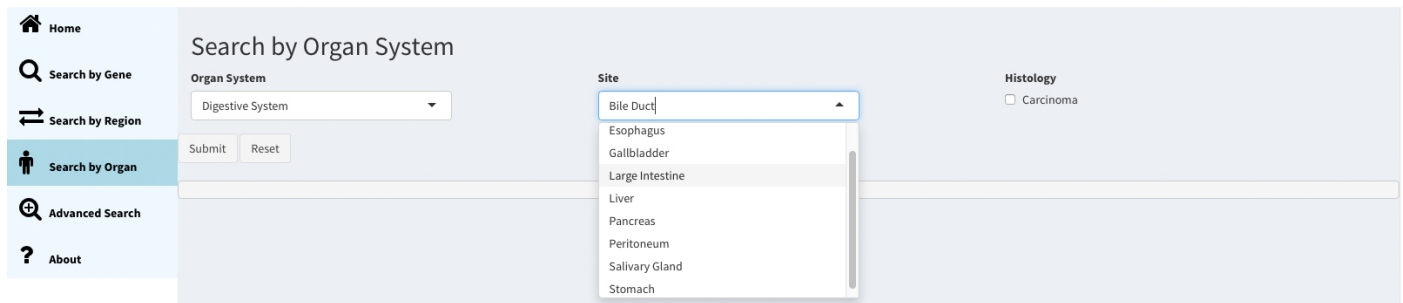

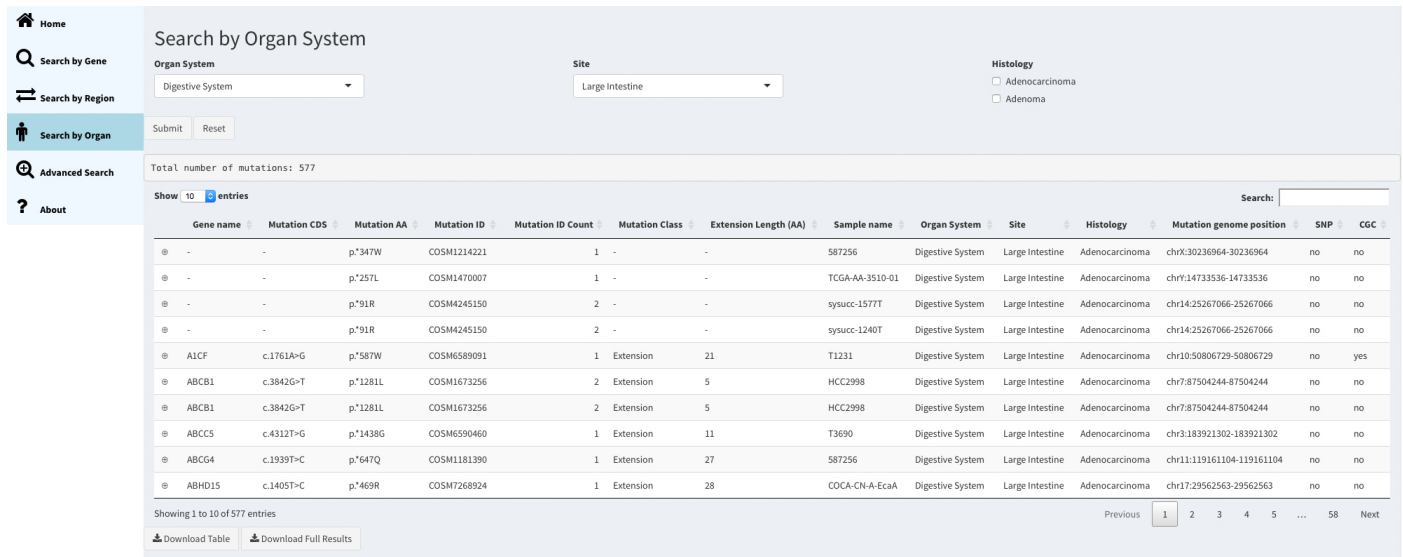*Figure 6: Search by organ - selection of primary site.*

Figure 7: Search by organ - selection of primary histology and results.

Advanced search

This search option has multiple search parameters including extension length, site and histology of nonstop mutations, as well as the mutation type (Figure 8). Here, users can also perform batch searches by providing a list of up to 100 genes. Additionally, all or only selected genes listed as cancer-related genes in the Cancer Gene Census (v89) can be included.

Figure 8: Advanced search.

Below is an example of a search for nonstop mutations that are found in cancer of the large intestine in in all known cancer genes (Figure 9).

Figure 9: Advanced search - results showing only mutations within CGC genes occurring in cancer of the large intestine.