SynMICdb

Synonymous Mutations In Cancer database

The Synonymous Mutations In Cancer database (SynMICdb) is a curated database of synonymous mutations in cancer. SynMICdb allows biologists to easily extract and download synonymous mutations in cancer as well as orthogonal data using multiple search options. It also integrates the predicted impact of synonymous mutations on structural changes in RNA using structural prediction algorithms.

Several independent search criteria are available in SynMICdb such as the gene name, the genomic coordinates, the position of the mutations within the coding sequence (CDS), their evolutionary conservation, the organ system, organ and tumor type, their link to cancer (Cancer Gene Census) or the SynMICdb score. Each search option is described in detail below.

Search by Gene

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

1. HGNC gene symbol
2. Gene name
3. ENSEMBL ID

Alias names for genes (P53 for TP53) are allowed and the search is case-insensitive.

Figure 1. Search option "Search by Gene".

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

---

[1] http://cancer.sanger.ac.uk/census/

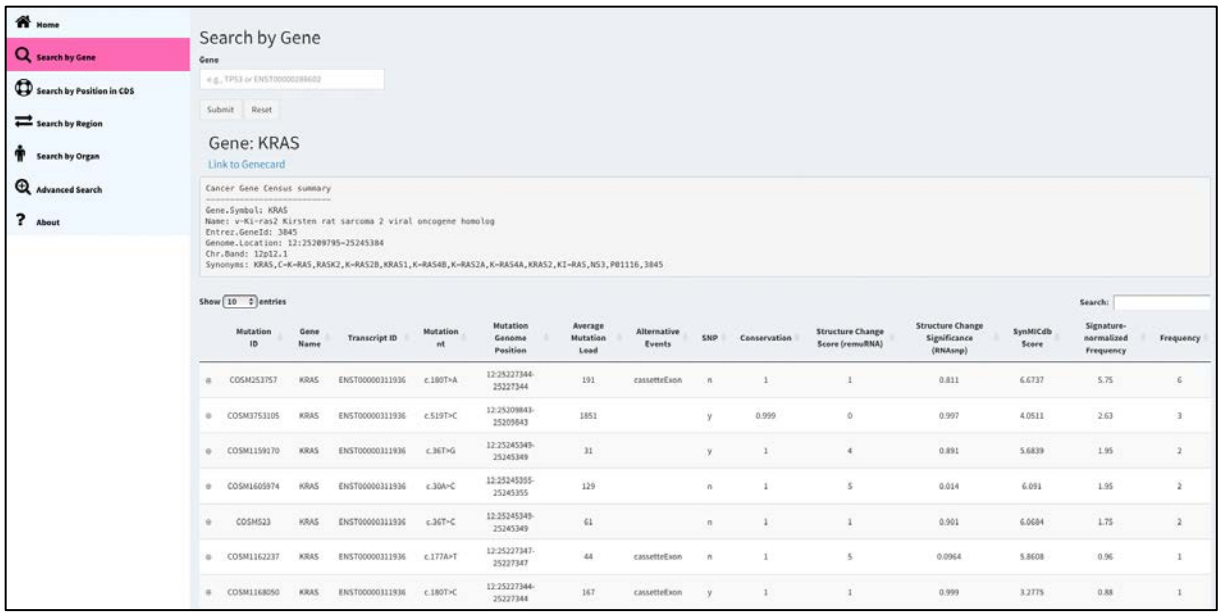
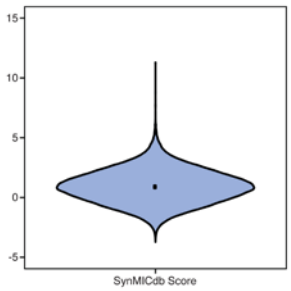


Figure 2. Example of result page for "Search by Gene" in SynMICdb.

The result columns provide the following information:

- **Mutation ID**: Unique identifier of each mutation (as present in COSMIC database).
- **Gene Name**: Abbreviated name of the gene.
- **Transcript ID**: ENSEMBL transcript ID for the corresponding mutation.
- **Mutation nt**: Number and nucleotide change of mutation: e.g. c.36T>G indicates a change of coding nucleotide number 36 from T to G.
- **Mutation genome position**: Genomic coordinates of each respective mutation in human genome assembly GRCh38 (chromosome:start-end).
- **SynMICdb Score**: The SynMICdb score shall reflect the probable impact of the synonymous mutation and is based on the mutation frequency, the probability due to mutational bias by mutation signatures, the average mutation load of the tumors with this mutation, the evolutionary conservation, the listing of the affected gene as cancer gene in the Cancer Gene Census, the listing of the mutation in the SNPdb, the FATHMM-MKL score, the CADD score and the predicted impact on RNA secondary structure. The score ranges from -4 to +12 and high numbers indicate a higher likelihood of a functional impact of the synonymous mutation. The distribution of the SynMICdb score is illustrated by the following table and violin plot:

Quantile	SynMICdb Score
top 50%	0.89
top 25%	1.83
top 10%	2.70
top 1%	4.38
top 0.1%	5.83
top 0.01%	8.08



Thus, a SynMICdb score of above 4.38 indicates that the synonymous mutation is among the top 1% of synonymous mutations in this study.

[2] http://www.genecards.org/

- **Average Mutation Load**: This column indicates the average number of mutations found in the genome-wide analysis of the tumor samples harboring this specific mutation.
- **Alternative Events**: This column provides information about alternative events as indicated by GENCODE like alternative splicing and other events that result in more than a single transcript from the same gene characterized by the UCSC genome browser[3].
- **SNP**: This column provides information whether this mutation has been listed as a Single Nucleotide Polymorphism (SNP) in the SNP database. y = yes, n = no.
- **Conservation**: This column lists the conservation scores of human vs. 99 vertebrate genomes (PhastCons100). The score ranges between 0 to 1 with 1 indicating the highest conservation levels among the 100 species.
- **Structure Change Score (remuRNA)**: This column depicts scores for structural change predictions for the respective mutation calculated by remuRNA. The score ranges from -5 to +20 and high numbers indicate a higher likelihood of a structural change caused by the mutation.
- **Structure Change Significance (RNAsnp)**: This column has *p-values* for significant structural change predictions for the respective mutation calculated by RNAsnp p0. The p-value ranges from 0 to 1 and low numbers indicate a higher likelihood of a structural change caused by the mutation.
- **Exon Type**: This column displays information about the exon type (1 = first exon, 2 = internal exon, 3 = last exon, 4 = monoexonic transcript).
- **Distance to Closest Exon Bundary**: This column indicates the distance to the closest exon boundary for each synonymous mutation in nucleotides.
- **Any ESE/ESS Change**: This column lists the gains and losses of exonic splicing enhancer (ESE) or exonic splicing silencer (ESS) motifs according to RegRNA 2.0 or SpliceAidF. Details for this analysis for ESEs and ESSs separately for the two prediction algorithms are provided in the full data table upon "Download Full Results". Please note that 23 motifs were assigned "ESE" as well as "ESS" properties in SpliceAidF and hence are listed separately as "ESE & ESS".
- **Signature-normalized Frequency**: In this column, the Frequency of the mutation has been corrected for the mutation bias due to mutational signatures frequently observed in cancer - thus, the Frequency has been multiplied with (1 - p) with p indicating the probability of the nucleotide change according to the most prevalent mutational signature in cancer.
- **Frequency**: This column shows the recurrence level of each mutation. The number in this column represents the total number of tumor samples in which the respective mutation was found.

By default, the results are grouped by Mutation ID and sorted by their frequency. For each Mutation ID, only one line is given in this view.
Detailed information for each sample can be viewed by clicking on the ⊕ icon.
Figure 3 shows an example of sample information for mutation ID COSM253757.

[3] For more details, please visit: https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=603403771_9k9O4FUq13hjAk0gvJCQPmO4vctG&c=chr12&g=knownAlt

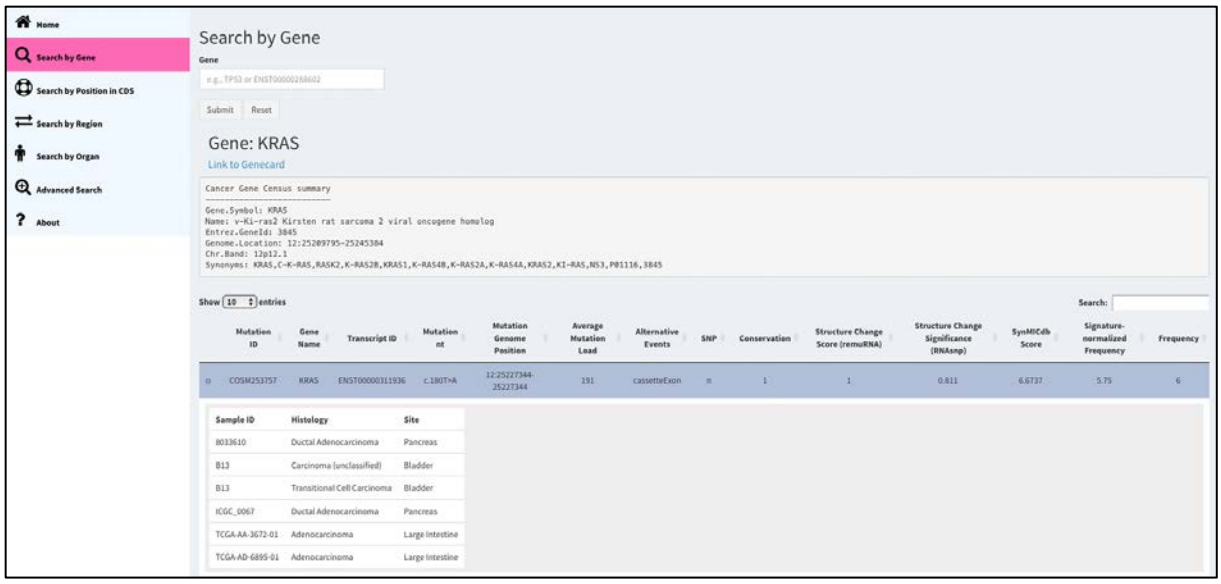


Figure 3. Detailed sample information for all samples having the Mutation ID COSM253757 in the table of results.

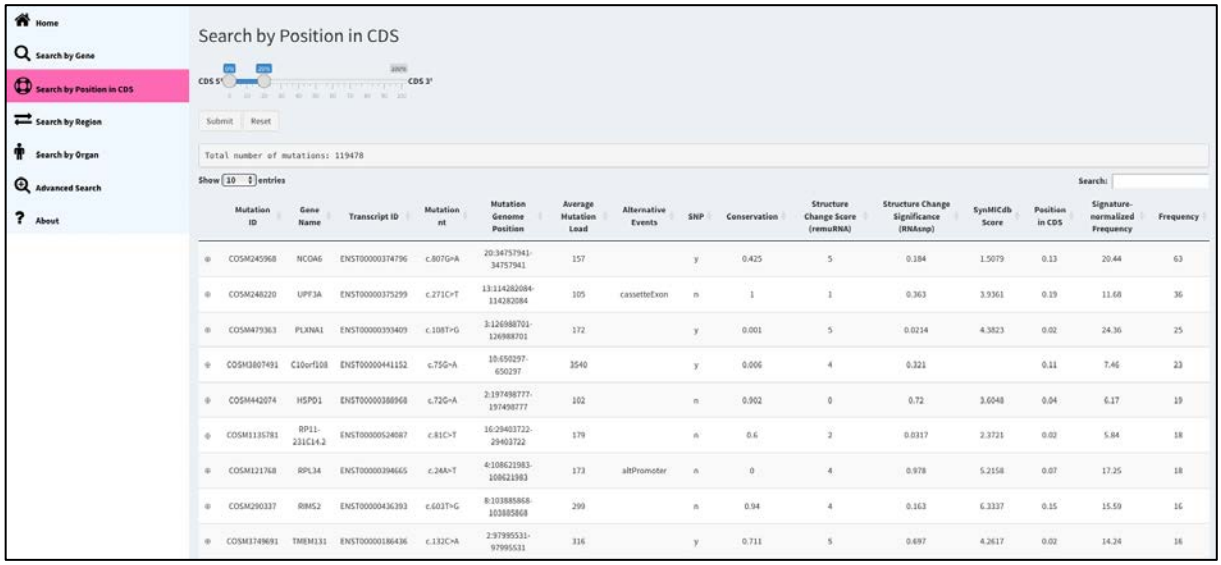
Download Options

The user can download the results using one of the following two options:

- **Download Table**: This button allows the user to download the displayed results as a csv file.
- **Download Full Results**: This button allows user to download the displayed results plus additional information like affected codon and amino acid, the mutation load of each affected sample, the position of the mutation within the CDS as well as the classification by the Cancer Gene Census (CGC).

Search by Position in CDS

This option allows the user to search for mutations on the basis of their location within the coding sequence (CDS) of genes (e.g. Figure 4 shows mutations present within the first 20% of the CDS). This facilitates the user to study synonymous mutation within a specific region of interest, for example towards the 5'-end of the coding region within the translation initiation and ramping region.

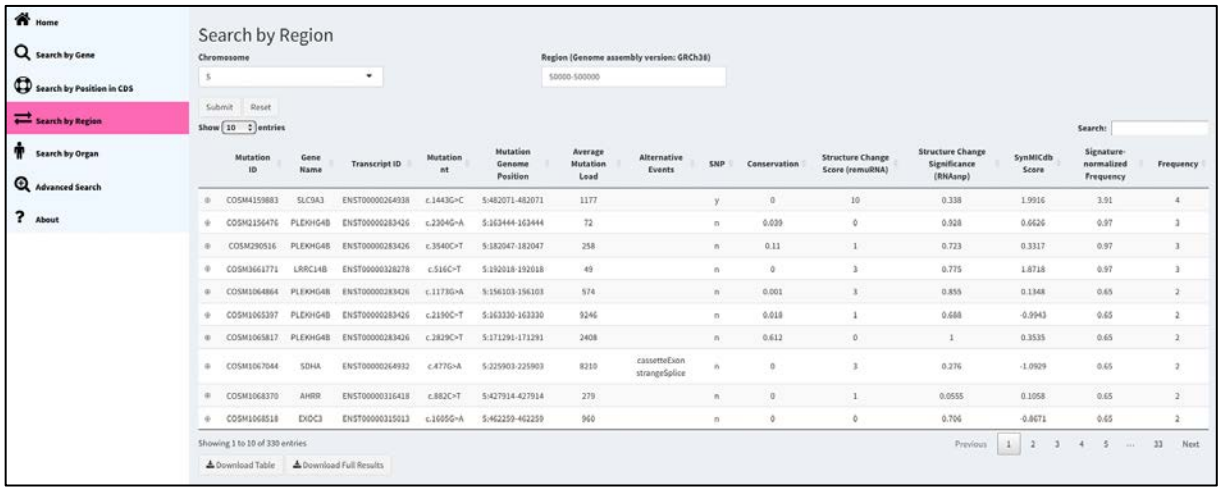


Mutation ID	Gene Name	Transcript ID	Mutation nt	Mutation Genome Position	Average Mutation Load	Alternative Events	SNP	Conservation	Structure Change Score (remuRNA)	Structure Change Significance (RNAsnp)	SynMICdb Score	Position in CDS	Signature-normalized Frequency	Frequency
COSM245968	NCOA6	ENST00000374796	c.807G>A	20:34757941-34757941	157		y	0.425	5	0.184	1.5079	0.13	20.44	63
COSM248220	UPF3A	ENST00000375299	c.271C>T	13:114282084-114282084	105	cassetteExon	n	1	1	0.363	3.9361	0.19	11.68	36
COSM479363	PLXNA1	ENST00000393409	c.108T>G	3:126988701-126988701	172		y	0.001	5	0.0214	4.3823	0.02	24.36	25
COSM3807491	C10orf108	ENST00000441152	c.75G>A	10:650297-650297	3540		y	0.006	4	0.321		0.11	7.46	23
COSM442074	HSPD1	ENST00000380968	c.72G>A	2:197498777-197498777	102		n	0.902	0	0.72	3.6048	0.04	6.17	19
COSM1135781	RP11-231C14.2	ENST00000524087	c.81C>T	16:29403722-29403722	179		n	0.6	2	0.0317	2.3721	0.02	5.84	18
COSM121768	RPL34	ENST00000394665	c.24A>T	4:108621983-108621983	173	altPromoter	n	0	4	0.978	5.2158	0.07	17.25	18
COSM290337	RIMS2	ENST00000436393	c.603T>G	8:103885868-103885868	299		n	0.94	4	0.163	6.3337	0.15	15.59	16
COSM1749691	TMEM131	ENST00000186436	c.132C>A	2:97995531-97995531	316		y	0.711	5	0.497	4.2617	0.02	14.24	16

Figure 4. Example of results page for "Search by Position in CDS".

Search by Region

This option allows the user to search for mutations present within a region defined by genomic coordinates of human genome assembly GRChg38 (note: chromosome 23 = X, 24 = Y and 25 = M). For example, Figure 5 shows the list of mutations present in chromosome 5 region 50000-500000.



Mutation ID	Gene Name	Transcript ID	Mutation nt	Mutation Genome Position	Average Mutation Load	Alternative Events	SNP	Conservation	Structure Change Score (remuRNA)	Structure Change Significance (RNAsnp)	SynMICdb Score	Signature-normalized Frequency	Frequency
COSM4159883	SLC9A3	ENST00000264938	c.1443G>C	5:482071-482071	1177		y	0	10	0.338	1.9916	3.91	4
COSM2156476	PLEKHG4B	ENST00000283426	c.2304G>A	5:163444-163444	72		n	0.039	0	0.928	0.6626	0.97	3
COSM290516	PLEKHG4B	ENST00000283426	c.3540C>T	5:182047-182047	258		n	0.11	1	0.723	0.3317	0.97	3
COSM3661771	LRRC14B	ENST00000328278	c.516C>T	5:192018-192018	49		n	0	3	0.775	1.8718	0.97	3
COSM1064864	PLEKHG4B	ENST00000283426	c.1173G>A	5:156103-156103	574		n	0.001	3	0.855	0.1348	0.65	2
COSM1065397	PLEKHG4B	ENST00000283426	c.2190C>T	5:163330-163330	9246		n	0.018	1	0.688	-0.9943	0.65	2
COSM1065817	PLEKHG4B	ENST00000283426	c.2829C>T	5:171291-171291	2408		n	0.612	0	1	0.3535	0.65	2
COSM1067044	SDHA	ENST00000264932	c.477G>A	5:225903-225903	8210	cassetteExon strangeSplice	n	0	3	0.276	-1.0929	0.65	2
COSM1068370	AHRR	ENST00000316418	c.882C>T	5:427914-427914	279		n	0	1	0.0555	0.1058	0.65	2
COSM1068518	DXDC3	ENST00000315013	c.1605G>A	5:462259-462259	960		n	0	0	0.706	-0.8671	0.65	2

Figure 5. "Search by Region" using genome coordinates.

Search by Organ

This option allows the user to search for synonymous mutations in cancer on the basis of their site of origin in a hierarchical manner. The user first selects an organ system and then a site and histology of interest. Nine organ systems are listed (as depicted in Figure 6): Cardiovascular System, Digestive System, Endocrine System, Genitourinary System, Integumentary System, Lymphatic System, Musculoskeletal System, Nervous System and Respiratory System.

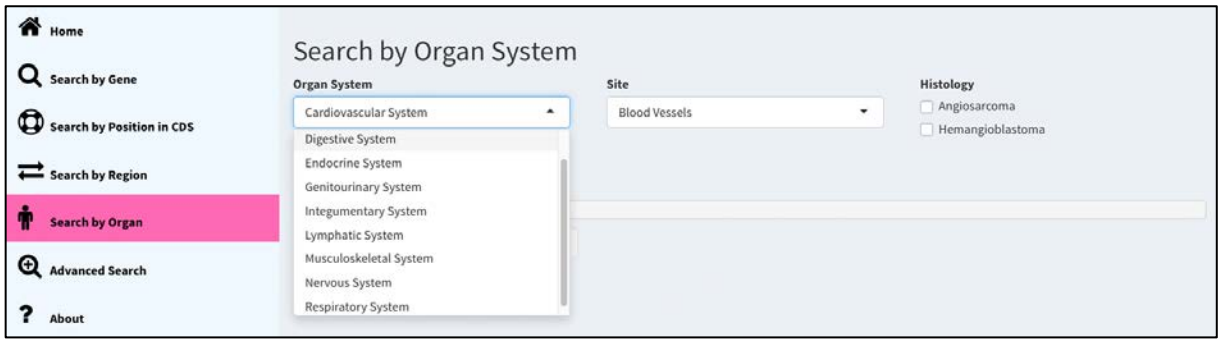


Figure 6. "Search by organ" - selection of the organ system of interest.

After selecting the organ system, the user selects first the primary site and optionally the histology of interest. The following example depicts a search and result of synonymous mutations present in the "Digestive System" as organ system following the selection of the "Large Intestine" as primary site (Figure 7) and "Adenocarcinoma" as histology (Figure 8).

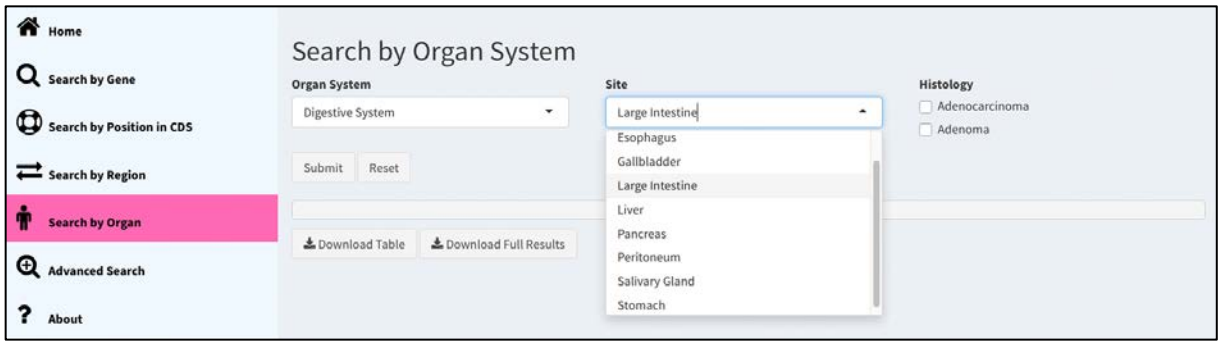


Figure 7. "Search by Organ" - selection of primary site.

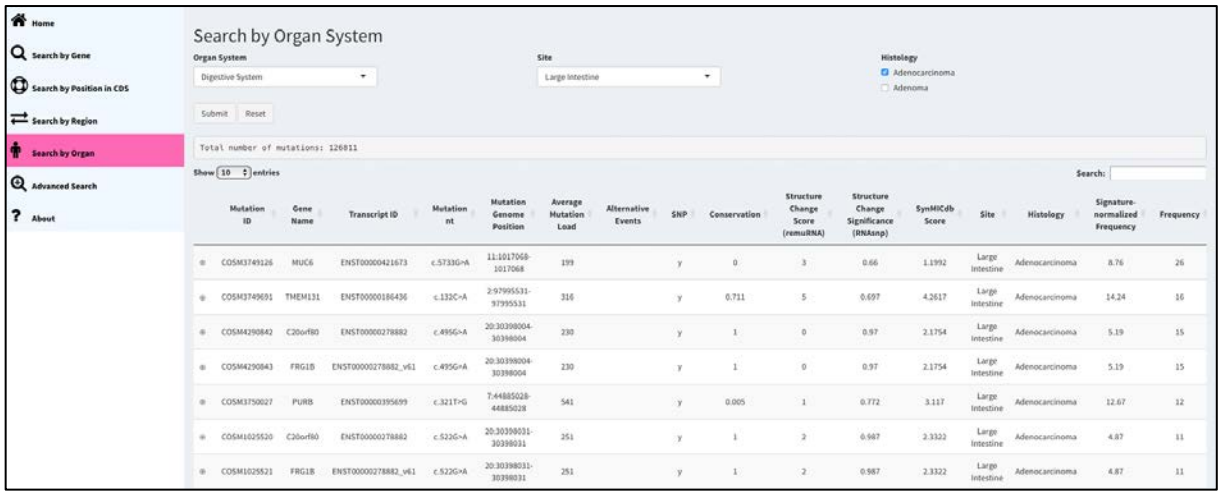


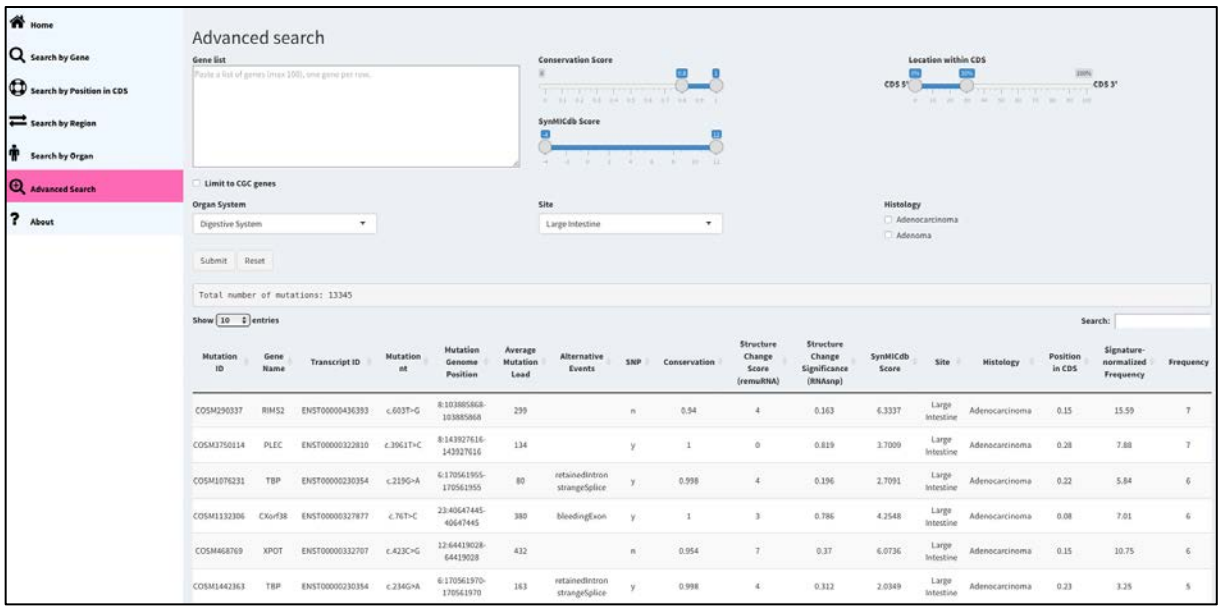
Figure 8. "Search by Organ" - selection of primary histology and results.

Advanced search

This search option allows the combination of multiple search parameters and offers additional search criteria including Gene Names, Cancer Gene Census genes, Conservation, Location within CDS, SynMICdb Score, Organ System, Site, and Histology of synonymous mutations. Here, users can also perform batch searches by providing a list of up to 100 genes (Figure 9).

Figure 9. Panel of the "Advanced Search".

Below is an example of search for synonymous mutations that are >80% conserved and only present in the first 30% of the CDS (Figure 10). The user can limit the output to genes listed as cancer genes in the Cancer Gene Census (CGC) database by clicking the "Limit to CGC genes" option.



Mutation ID	Gene Name	Transcript ID	Mutation nt	Mutation Genome Position	Average Mutation Load	Alternative Events	SNP	Conservation	Structure Change Score (remuRNA)	Structure Change Significance (RNAsnp)	SynMICdb Score	Site	Histology	Position in CDS	Signature-normalized Frequency	Frequency
COSM290337	RIMS2	ENST00000436393	c.603T>G	8:103885868-103885868	299		n	0.94	4	0.163	6.3337	Large Intestine	Adenocarcinoma	0.15	15.59	7
COSM3750114	PLEC	ENST00000322810	c.3961T>C	8:143927616-143927616	134		y	1	0	0.819	3.7009	Large Intestine	Adenocarcinoma	0.28	7.88	7
COSM1076231	TBP	ENST00000230354	c.219G>A	6:170561955-170561955	80	retainedIntron strangeSplice	y	0.998	4	0.196	2.7091	Large Intestine	Adenocarcinoma	0.22	5.84	6
COSM1132306	CXorf38	ENST00000327877	c.76T>C	23:40647445-40647445	380	bleedingExon	y	1	3	0.786	4.2548	Large Intestine	Adenocarcinoma	0.08	7.01	6
COSM468769	XPOT	ENST00000332707	c.423C>G	12:64419028-64419028	432		n	0.954	7	0.37	6.0736	Large Intestine	Adenocarcinoma	0.15	10.75	6
COSM1442363	TBP	ENST00000230354	c.234G>A	6:170561970-170561970	163	retainedIntron strangeSplice	y	0.998	4	0.312	2.0349	Large Intestine	Adenocarcinoma	0.23	3.25	5

Figure 10. "Advanced Search" - results listing synonymous mutations in the large intestinal tumors with a conservation score >= 0.8 and mutation position within the first 30% from the 5' end of the CDS.