COSMIC - Catalogue of Somatic Mutations in Cancer

http://cancer.sanger.ac.uk/cosmic

The COSMIC Project

Expert knowledge curation

Large-scale cancer genomics

Curate, Integrate, Combine

http://cancer.sanger.ac.uk/
Large-scale systematic screens

ICGC/TCGA/CGP

Publication datasheets

Data Extraction

GRCh37 co-ords & allele seqs

Sample Details

QC (data input format, details & noise)

Detailed genic analyses

Deep genic data

Standardisation

COSMIC

Data In

QC (reject ~30%)

Broad genomic data

Genic annotation (HGVS syntax)

QC (sequence and annotation checks)

ATRAC

Data Extraction

Sample Details

GRCh37 co-ords & allele seqs

Publication datasheets

ICGC/TCGA/CGP

QC (data input format, details & noise)

Detailed genic analyses

Deep genic data

Standardisation

COSMIC

Data In

QC (reject ~30%)

Broad genomic data

Genic annotation (HGVS syntax)

QC (sequence and annotation checks)

ATRAC

Educating for best practices in clinical cancer genomics
Search for Disease, Gene, Variant, Sample, Study in either search bar.

- Much information is open to public, but some features require a subscription/login.
- News window in upper right.
- Tools – Cancer Browser
- Data Curation information and details
- Release Notes – inform of new additions/advancements to COSMIC

***This is no longer the “Beta” version as of 8/1/17.
Select Data Source - COSMIC vs. Cell Lines

- Cell Lines Project and COSMIC house different data.
- To select a cell line to look into, look at the Cell Line Browser, or NCI-60 Browser if applicable.
Cell Lines Data

Once you select a cell line to view, sample page opens.

Menu on left – click on any item in menu to navigate to that item or scroll to desired information.

Overview – cell line information: sample type, demographics, institute, microsatellite stability, STR Profile, Sequence stats, and Downloadable files

Mutation Spectrum tab – information can be matched to mutation profiles through COSMIC Home Page

Sequence Context tab – bases flanking mutations

Heatmap tab – shows frequent mutations
Cell Lines Variants tab

- Can view by variant type – tabs in window
  - Breakpoints: search by hg38 coordinates

- Exportable in .TSV and .CSV

- Click on Gene name to go to gene page

- Click on mutation nomenclature to go to mutation page – can determine mutation recurrence in cell lines dataset and pathway involvement
Mutation Filters

• Scroll right to see expandable mutation filters.

COSMIC
Catalogue Of Somatic Mutations In Cancer

Educating for best practices in clinical cancer genomics
CNV and Expression Data

- Under/over expression
  - Z-Score
- CN Type
  - Minor allele
  - Copy number observed
  - Hg38 coordinates
- Average Ploidy
Searching in COSMIC

- Search by disease type – brings you to Cancer Browser – see slides 24-30
- Search by gene – brings you to search results page that can link to gene/mutation overview (same format as Cell Lines Project)
- Search by mutation – bring you to mutation overview page (same format as Cell Lines Project)
- All data are derived directly from a cited data source
  - Usually a PubMed ID
Gene Overview Page

- Several Items on Menu on left panel
  - Gene view - view domains within gene and where mutations reside
  - Genome browser - view genomic context
  - Overview - general information
  - Tissue - expression/mutations in each tissue type
  - Distribution - what kind of mutations observed?
  - Drug Resistance - do mutations confer drug resistance
  - Variants - table of observed variants in gene
  - References - curated references for gene

Educating for best practices in clinical cancer genomics
Gene Overview Page

Navigate the gene page by selecting an item from the left menu or scrolling through page.

Educating for best practices in clinical cancer genomics
p53 – an inactivating mutation profile

p53 is oncogenic when its activity is destroyed by inactivating mutations: **Loss-of-function.**

including substitutions, insertions and deletions which cause an early STOP codon
p53 – dominant missense mutations

However, specific p53 mutations also show **Gain – of – function**

- key missense substitutions cause an increase in malignancy
- and cause resistance to chemotherapies (eg cisplatin)

**p.R175H reduces apoptosis**

**p.R248Q & p.R273H/C increase proliferation & invasiveness**

**Missense (R273H)**

GTG CGT GTT
Val Arg Val

GTG CAT GTT
Val His Val

p53 is a DNA binding protein, affecting the work of many other genes – GOF mutations alter how p53 binds to DNA, changing its checkpoint function
Mutation Overview

• Overview – general mutation information
• Tissue distribution – mutation frequency in tissue types
• Samples – table of affected samples
• Pathways Affected – table of affected pathways
  – Click link to go to WikiPathways
• References – links to relevant publications
Mutation Overview Page

**Mutation**

- COSM474

**Overview**

This section shows a general overview of the selected mutation. It describes the source of the mutation i.e gene name sequence level. You can see more information on our help pages.

- **Mutation ID**: COSM474
- **Gene name**: BRAF
- **AA mutation**: p.V600R (Substitution - Missense, position 600, V→R)
- **CDS mutation**: c.1798_1799GT>AG (Complex)
- **Nucleotides inserted**: ag
- **Genomic coordinates**: GRCh38, 7:140753336..140753337, view Ensembl contig
- **CDD**: NP_004324.2
- **HomoloGene**: 3197, view the multiple sequence alignment
- **Ever confirmed somatic?**: Yes
- **FATHMM prediction**: n/a (score 0.00)
- **Remark**: n/a
- **Recurrent**: n/a
- **Drug resistance**: n/a
Cancer Gene Census

- **Main Links** under “Projects” Section on COSMIC home page.
- **List** is in table format and (with license) is exportable in .CSV or .TSV format.
- **Genes** marked as Hallmarks of Cancer have more functional descriptions available
  - Click on

---

Cancer Gene Census

The Cancer Gene Census (CGC) is an ongoing effort to catalogue those genes for which mutations have been causally implicated in cancer. The original census and analysis was published in *Nature Reviews Cancer*.

The census is not static but rather is updated regularly as needed. In particular, we are grateful to Felix Pittelkow and his colleagues in providing information on more genes involved in uncommon translocations in leukemias and lymphomas. Currently, more than 1% of all human genes are implicated via mutation in cancer. Of these, approximately 90% have somatic mutations in cancer; 20% bear germline mutations that predispose to cancer and 10% show both somatic and germline mutations.

---

<table>
<thead>
<tr>
<th>Gene Symbol</th>
<th>Name</th>
<th>Entrez Genoid</th>
<th>Genome Location</th>
<th>Hallmark</th>
<th>Chr Band</th>
<th>Somatic</th>
<th>Germline</th>
<th>Tumour Types (Somatic)</th>
<th>Tumour Types (Germline)</th>
<th>Cancer Syndrome</th>
<th>Tissue Type</th>
<th>Molecular Genetics</th>
<th>Role in Cancer</th>
<th>Mut Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABL1</td>
<td>abl-interactor 1</td>
<td>10006</td>
<td>10:26748570-268460863</td>
<td>10p11.2</td>
<td>yes</td>
<td>AML</td>
<td>L</td>
<td>Dom</td>
<td>TSG; fusion</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABL1</td>
<td>v-abl Abelson murine leukemia viral oncogene homolog 1</td>
<td>252</td>
<td>9:1308354647-132085663</td>
<td>9q34.1</td>
<td>yes</td>
<td>CML; ALL; T-ALL</td>
<td>L</td>
<td>Dom</td>
<td>oncogene; fusion</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABL2</td>
<td>c-abl oncogene 2; non-receptor tyrosine kinase</td>
<td>272</td>
<td>1:179107718-17913044</td>
<td>1q24-25</td>
<td>yes</td>
<td>AML</td>
<td>L</td>
<td>Dom</td>
<td>oncogene; fusion</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACKR3</td>
<td>atypical chemokine receptor 3</td>
<td>57007</td>
<td>5p</td>
<td>2q37.3</td>
<td>yes</td>
<td>Isopna</td>
<td>M</td>
<td>Dom</td>
<td>oncogene; fusion</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACSL3</td>
<td>acyl-CoA synthetase long-chain family</td>
<td>2181</td>
<td>2q36</td>
<td>prostate</td>
<td>E</td>
<td>Dom</td>
<td>fusion</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Cancer Gene Census

The Cancer Gene Census (CGC) is an ongoing effort to catalogue those genes implicated in cancer. The census is not static but rather is updated regularly/as needed. In particular, over 1% of all human genes have been implicated via somatic and germline mutations.

The table below shows the following columns:
- **Gene Symbol**: The symbol for the gene.
- **Name**: The name of the gene.
- **Entrez GeneId**: The Entrez Gene ID.
- **Genome Location**: The genome location of the gene.
- **Hallmark**: Any hallmark annotations associated with the gene.

Click on buttons near the top of the browser or scroll down to see the “Breakdown” of the types of gene abnormalities in the Census or abbreviations used in the table.
Hallmarks of Cancer

ACKR3
atypical chemokine receptor 3

Function summary
scavenger receptor for chemokine CXCL12 in vascular endothelium [PubMed]

Role in cancer
oncogene

Types of alteration in cancer
overexpressed in breast cancer [PubMed]

Promotes
proliferative signalling
promotes tumour growth in breast cancer [PubMed]

invasion and metastasis
limits metastasis in breast cancer [PubMed]

Suppresses
Drug Resistance

• Main Link under “Data Curation” section on COSMIC home page.

• Resistance Data are available for 12 genes:
  – ABL1, ALK, BRAF, EGFR, ESR1, FLT3, KIT, MAP2K1, MAP2K2, PDGFRA, SMO, MET

• Table to view drug and resistance mutation frequency
Drug Resistance

- Manually curated list of genes that can carry resistance mutations to particular drugs.
- Click on gene name in table to see Gene Page. Scroll down to the drug resistance section.
Mutational Signatures

- Selection under “Data” in top banner.

- Describes mutational signatures in detail and specifies diseases that carry each mutational signature.
GRCh37 Cancer Archive

• If your lab has not converted to hg38, there is an option to view GRCh37/hg19 information on COSMIC.
• Under “Projects” in top banner, select GRCh37 archive to see legacy site.
Data Curation

• Other links under Expert Curation section on home page
  – **Gene Curation:** list of all genes with deep curation completed. Curation details included on Curated Genes page. Click on gene to go to gene page.
  – **Gene Fusion Curation:** list of all gene fusions with deep curation completed for solid tumors. Click on pair of gene to go to fusion overview page.
    • Fusion Overview page – information about in what kind of tissue fusion was observed and PMID.
  – **Genome Annotation:** Information and parameters used to analyze data.
  – **Drug Resistance:** See slides 19-20.
Cancer Browser

- Main link under “Tools” on home page
- Select tumor of interest by tissue and histology
- Press ‘Go’ in lower right corner of window
Cancer Browser

- Opens to bar graph of top 20 mutated genes in your selected type of cancer.
- In table format you can view and filter genes with and without mutations by mutation frequency or number of samples tested.
  - Exportable in .CSV or .TSV format (with license)
Cancer Browser

- Can view genome browser in cancer browser window.
- Mutation matrix – top 20 mutations by sample
Cancer Browser: Mutation Matrix

- Can redraw image using drop-down menu on top of window to view data based on a certain mutation type.
Cancer Browser: Distribution

- Distribution of mutations within your selected cancer type.
- Substitutions are described similarly base>base (not shown).
- Indels described by size in bps (Insertions displayed similar to deletions).

---

**Summary**

An overview of the types of mutation observed.

<table>
<thead>
<tr>
<th>Colour</th>
<th>Mutation type</th>
<th>Number of samples (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blue</td>
<td>Nonsense substitution</td>
<td>33 (3.00%)</td>
</tr>
<tr>
<td>Green</td>
<td>Missense substitution</td>
<td>952 (86.55%)</td>
</tr>
<tr>
<td>Red</td>
<td>Synonymous substitution</td>
<td>3 (0.27%)</td>
</tr>
<tr>
<td>Black</td>
<td>Inframe insertion</td>
<td>4 (0.36%)</td>
</tr>
<tr>
<td>Orange</td>
<td>Frameshift insertion</td>
<td>18 (1.64%)</td>
</tr>
<tr>
<td>Pink</td>
<td>Inframe deletion</td>
<td>21 (1.91%)</td>
</tr>
<tr>
<td>Red</td>
<td>Frameshift deletion</td>
<td>64 (5.82%)</td>
</tr>
<tr>
<td>Red</td>
<td>Complex mutation</td>
<td>7 (0.64%)</td>
</tr>
<tr>
<td>Other</td>
<td></td>
<td>34 (3.09%)</td>
</tr>
<tr>
<td><strong>Total unique samples</strong></td>
<td>1100</td>
<td></td>
</tr>
</tbody>
</table>

---

**Deletions**

This histogram shows the distribution of deletion size across samples. You can see all samples with deletions.
Cancer Browser: Variants

- Observed Fusions, Mutations, Methylation status, and CNV/Expression data given in table format.
  - Exportable in .CSV and .TSV format (with license)
  - Click on mutations for mutation overview
  - Click on fusion pairing for Fusion overview
  - Click on gene for gene overview.
Cancer Browser: Samples

• Table of mutated and non-mutated samples
  – Exportable in .TSV or .CSV (with license)
  – Click on sample ID for sample overview page
Sample Overview Page

<table>
<thead>
<tr>
<th>Sample information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample name</td>
</tr>
<tr>
<td>COSMIC sample ID</td>
</tr>
<tr>
<td>Tumour location</td>
</tr>
<tr>
<td>Screening method</td>
</tr>
<tr>
<td>Source</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Microsatellite instability (MSI)</td>
</tr>
<tr>
<td>Curated features</td>
</tr>
<tr>
<td>Tumour details</td>
</tr>
<tr>
<td>Individual details</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Other samples linked to the same individual</td>
</tr>
<tr>
<td>Supplier</td>
</tr>
<tr>
<td>External links</td>
</tr>
<tr>
<td>STR profile data</td>
</tr>
<tr>
<td>Sequence statistics</td>
</tr>
</tbody>
</table>

- Click on menu on left side of page or scroll to desired information
- Primary reference(s) and/or large study name are available in the reference section
Educating for best practices in clinical cancer genomics

GA4GH Beacon

- Main link under “Tools” on COSMIC home page
- GA4GH = Global Alliance for Genomics and Health – see Information section
- Shared URL format for query
  - http://cancer.sanger.ac.uk/api/ga4gh/beacon/query?chrom=\?;pos=\?;allele=\?
- Can query if COSMIC or Cell Lines Project has observed a specific mutation at a designated genomic locus.
SNV frequency depicted by yellow/orange segments
Predicted small molecule binding sites in blue.
Can take screenshot, pop screen out, drag mouse or click button to rotate protein structure for better view.
Several viewing and toggle options
Protein property information
UniProt sequence information and mutation information integrated.
There are so many features within this application. Check them out!
COSMIC BigQuery

• If you have a COSMIC account, register at no charge for access to BigQuery.
• Research based queries of large datasets at minimal cost per query.
• Supports Standard SQL programming language and can be accessed by R

Educating for best practices in clinical cancer genomics
• CONAN = COpy Number ANalysis – under “Tools” on Home page.
• Search by gene (HGNC or Ensembl) or genomic region (hg38)
### Cosmic Copy Number Analysis (CONAN) » BAP1

#### Results for BAP1

<table>
<thead>
<tr>
<th>Tissues</th>
<th>Amplification</th>
<th>Homozygous Deletion</th>
<th>LOH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Breast (1749)</td>
<td>0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Central nervous system (1093)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Carcin (313)</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Eye (80)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Kidney (1027)</td>
<td>0</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Large intestine (771)</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Lung (1185)</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Oesophagus (220)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pleura (108)</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Skin (630)</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Stomach (501)</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Urinary tract (419)</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Showing 1 to 12 of 12 entries

#### Cosmic Copy Number Analysis (CONAN) » BAP1

#### Results for BAP1

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample ID</th>
<th>CNV</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCGA-35-5375-01</td>
<td>1780088</td>
<td>LOH</td>
</tr>
<tr>
<td>TCGA-95-7043-01</td>
<td>1514101</td>
<td>LOH</td>
</tr>
<tr>
<td>TCGA-EP-947F-01</td>
<td>2194749</td>
<td>LOH</td>
</tr>
</tbody>
</table>
Data Downloads

• Main Link under “Data” on Home Page
• Need license to download data
• Depending on the data you would like to download, you may need to download off of a SFTP site using a FTP tool like WinSCP or Filezilla
Data Submission

- Under Help → Data Submission
- Using COSMIC data or describing mutation data in results

- Submitting Data to COSMIC

Educating for best practices in clinical cancer genomics
Other Links Under Data

- Genome Annotation
- Datasheets – details about COSMIC versions
- Help – links to helpful overview resources
- FAQ – Frequently Asked Questions
Scenario #1

• You are a laboratory professional trying to identify genes to put on your lab’s new breast carcinoma NGS panel.
  – Cancer Browser
  – Search Cancer Gene Census
Educating for best practices in clinical cancer genomics
## Results from Queries

### Cancer Genome Browser Query (Criteria ≥ 3% mutated) = 16 genes
- PIK3CA
- TP53
- CDH1
- GATA3
- KMT2C
- ESR1
- PTEN
- RB1
- SPEN
- ARID1A
- NCOR1
- KMT2D
- MAP2K4
- NF1
- TBX3
- LRP1B

### Cancer Gene Census Query = 32 genes
- CCND1
- KEAP1
- NOTCH1
- GATA3
- SMARCAD1
- ESR1
- MAP3K13
- NCOR1
- TP53
- AKT1
- ARID1B
- TBX3
- ERBB2
- BRCA1
- BRCA2
- FOXA1

9 genes overlap between two lists = 39 total genes
Scenario #2

• A whole genome sequencing (WGS) case in your lab has a mutation that you haven’t observed before. Does this mutant allele exist in COSMIC?
  – GA4GH Beacon
  – Chr7:140753333 (BRAF)
  – Mutated Allele = A
Yes - this is in COSMIC

- Mutation can be found in Genome browser on BRAF gene page or by entering the AA mutation or CDS Mutation Nomenclature.
Scenario #3

• Your lab is building a Pan Cancer List based on multiple resources.
  – Cancer Gene Census
Using Data from Cancer Gene Census

• If you have a license, download as .CSV or .TSV
  – This is useful as Excel if you’d like to sort data in different ways to stratify genes based on other criteria in COSMIC’s table.

• Cancer Gene Census is updated with most version releases of COSMIC.
Scenario #4

- Mate Pair sequencing has picked up a translocation your lab has not yet reported.
  - NUP98/KDM5A
  - Search in “Gene fusion curation” under Data Curation
Curated Fusions

Gene fusions, or translocations, resulting from chromosomal rearrangements are the most common mutation class. They lead to chimeric transcripts or to deregulation of genes through juxtapositioning of novel promoter or enhancer regions.

Gene fusions are manually curated from peer reviewed publications by expert COSMIC curators. A comprehensive literature curation is completed for each fusion pair when it is released in the database. Currently COSMIC includes information on fusions involved in solid tumours. Annotation of fusions associated with leukaemias and lymphomas will be added.

Select a gene pair from the list to go to the Overview page for that fusion and see all curated fusions for that pair, as well as the tissue types in which they were found and the associated publications. Many additional data points are curated for individuals (e.g. age, gender), tumour (e.g. stage, drug response) and samples (e.g. histology, sample source). This information can be found on the Sample Overview page.

- Click on fusion gene pairing to open fusion overview.
- Link to PMID and/or large scale studies on Reference tab.
Scenario #5

• Your lab has a significant pool of data that you would like to contribute to COSMIC to enhance the knowledgebase.
  – COSMIC is exploring collaborative relationships with institutions aiming to release anonymized somatic mutation calls in COSMIC
• COSMIC will highlight these laboratories in their webpages and strongly cite them as sources.
• Contact COSMIC directly
  – Simon Forbes (Head of COSMIC): saf@sanger.ac.uk
  – COSMIC Helpdesk: cosmic@sanger.ac.uk
Future Features: Coming Soon

- Guide to Actionability
- Organoids
  - Data release for the Human Cancer Model Initiative
  - Better laboratory models – in preparation
- Cancer Mutation Census
  - Identify driver mutations across all diseases
Online Tutorials

- COSMIC Site Overview: https://www.youtube.com/watch?v=whxIL86gnKA
- Sample Data Tutorial: https://www.youtube.com/watch?v=5dqInH8_LAo
- Cancer Browser Tutorial: https://www.youtube.com/watch?v=k477uAiKx74
- Gene Pages Tutorial: https://www.youtube.com/watch?v=2FD5RabgK6o
- Fusions Tutorial: https://www.youtube.com/watch?v=M9ILszjsuAw
- COSMIC Search Tutorial: https://www.youtube.com/watch?v=SVfloi4hfNM
- DISCLAIMER: These tutorials all provide instruction on how to use older versions of COSMIC. The latest version of COSMIC looks different and has different features.
Citation

Contacts

- Twitter Handle: @cosmic_sanger
- Facebook: cosmic.sanger
- Email: cosmic@sanger.ac.uk
  - Simon Forbes (Head of COSMIC): saf@sanger.ac.uk