***v97 (Nov 2022) A focus on blood cancer, 4 census Tier 2 genes, 10 cancer hallmark genes are updated along with resistance data. In this release of COSMIC, we have 44,000 new genomic variants, 127,000 new coding mutations, 27,000 non-coding mutations, 6000 new samples and 1,435 new whole genomes. We have also curated 20 new systematic screen papers.***

# Key Updates

1. New focus curation on Blood cancers
2. Cancer Gene Census: 3 new Tier 2 genes are added ( GOLPH3, FADD, SUB1) and 1 moved from Tier 1 to Tier 2 (SMARCD1)
3. We have created cancer hallmark annotations for each of the 10 Cancer Gene Census Tier 1 genes (NFKBIE, NTRK3, PHF6, POLD1, POLE, PPP2R1A, PRDM1, PTCH1, RPL5, SALL4). By so doing, we are adding functional annotations for 10 genes causally associated with cancer which provide an overview of how the genes contribute to tumour development, in regard to the hallmarks of cancer.
4. 20 new systematic screen papers
5. Data for drug resistance is updated
6. Cancer Mutation Census data is also updated to align with the latest COSMIC data (v97), along with ClinVar and gnomAD datasets
7. COSMIC-3D is also updated to align with the latest COSMIC data (v97), assembly update, new census gene mapped structures and around 1100 more mapped protein structures.
8. Actionability data has been fully updated. Many new trials have been added, the number of trials with results available has substantially increased, several new mutations are represented and a new gene, BTK, has been fully curated.
9. Actionability and CMC downloads are free for non-commercial use, files are available on the download page.
10. New download file to map missing significant variants in the Non-Coding region

# Cancer Focus Curation

## Blood Cancer

As part of release v97 we have focused on updating the expert-curated mutation data for blood tumours. Blood tumours in COSMIC are classified under haematopoietic and lymphoid tissue as haematopoietic neoplasms or lymphoid neoplasms, which include cancer types such as leukaemias, lymphomas and myelomas. 76 additional publications with mutation screening data in these tumour types are included in this release. The types of data includes whole genome studies and studies utilising large next generation sequencing panels to case reports with more unusual clinical details and novel treatments. Over 2,600 samples were curated and 24,356 new variants added from these samples. Release v97 also incorporates 9 new blood tumour types into COSMIC.

# Drug resistance Data

Gene drug pairs are added for website visualisation:

1. NT5C2 - purine Unique samples - 57, Unique Mutation -81
2. FGFR2 - BGJ398 Unique samples - 9, Unique Mutation -29
3. FGFR2 - pemigatinib Unique samples - 8, Unique Mutation -6

All the key statistics have also been updated, for more details please check the Drug resistance page.

# Cancer Mutation Census

Cancer Mutation Census data has been updated for v97 release. These are the key updates:

- Cancer Mutation Census has been updated to align the mutations with COSMIC release 97
- The ClinVar dataset has been updated to release 2022-08
- gnomAD exome frequencies are from release v2.1.1 and contain data from 125,748 exome samples
- gnomAD genome frequencies have been updated to release v3.1 containing 76,156 genome samples. This release also includes a new population - Middle Eastern (MID)
- CMC data are free for non-commercial use, downloads are available on the COSMIC download page.

# COSMIC-3D

COSMIC-3D data has been updated for v97 release. These are the key updates:

- COSMIC-3D has been updated to align the mutations with COSMIC release 97
- Switch from GRCh38 to GRCh37 human genome assemblies in line with the CMC data
- 7 census genes now have mapped structures: (ABI1, ARID2, ATP1A1, FOXA1, FOXL2, SS18, TLR5, TRRAP
- Increased total number of mapped protein structures (pdb ids) from 50,735 to 51, 816

# Actionability and CMC downloads

Actionability and CMC downloads are free for non-commercial use, files are available on the COSMIC download page. Please refer to our licensing page here to understand if you are a Non-Commercial or Commercial user and how to obtain a license.

# New download file NCV CDS syntax mapping

Since the annotation system upgrade in v90, VEP is used to standardise and normalise all variant annotations.https://www.ensembl.org/info/docs/tools/vep/index.html

One unintended consequence of using VEP is that it outputs genomic level (g.) annotations for many non-coding variants in the 5' UTRs of genes. Sometimes these mutations are known or predicted to be functionally significant and have well known CDS annotations in the scientific literature (eg TERT promoter mutations). Previously, these CDS annotations were shown in COSMIC but since the v90 upgrade these are overridden by the standardised VEP genomic annotations.

In order to maintain a standardised dataset, we will continue to show the VEP genomic annotations for all mutations, but we have now produced a mapping file to allow the non-coding variant (NCV) genomic annotations to be linked back to the CDS syntaxes.

The new mapping file NCV_CDS_syntax_mapping.tsv released in v97 can be cross referenced with the CosmicNonCodingVariants.vcf.gz or CosmicNCV.tsv.gz download files to link CDS syntaxes with LEGACY_ID or COSV identifiers.

Generally, on the website we focus on coding mutations, but non-coding variants are displayed on the Genome Browser and can also be viewed directly by searching for the COSN identifier eg: https://cancer.sanger.ac.uk/cosmic/search?q=COSN32285790

In v97, the new mapping file contains only TERT promoter mutations, but we plan to include non-coding mutation mapping for other genes in future releases.
This new file is available on the COSMIC download page.

## Systematic Screen Papers

Follow links below to the 20 papers which are new in v97, or view the full table of papers here.

COSP50547, COSP42471, COSP40854, COSP45619, COSP50467, COSP41068, COSP50471, COSP50438, COSP43701, COSP40675, COSP38483, COSP41453, COSP35787, COSP50319, COSP40730, COSP41722, COSP45615, COSP43650, COSP49463, COSP50070

## COSMIC Statistics:

*Numbers with a '+' at the end of each statistics denotes the increase since the last release.*

| | |
|---|---|
| | **23,443,841** |
| Total genomic variants (COSV) (+44,671) | |
| | **16,015,511** |
| Genomic non-coding variants (+26,839) | |
| | **5,037,981** |
| Genomic mutations within exons (coding variants) (+35,377) | |
| | **8,717,735** |
| Genomic mutations within intronic and other intragenic regions (+13,431) | |
| | **1,515,965** |
| Samples (+6,287) | |
| | **28,880** |
| Papers (+186) | |
| | **41,161** |
| Whole genome screen samples (+1,435) | |
| | **321,804** |
| Genomic Rearrangements (+3,788) | |
| | **19,428** |
| Fusions | |
| | **1,207,190** |
| Copy number variants | |
| | **9,215,470** |
| Gene expression variants | |
| | **7,930,489** |
| Differentially methylated CpGs | |

# Actionability Release v7

COSMIC Actionability v7 includes **11** additional fully-curated genes:
CD274 (PD-L1), HRAS, MAP2K1 (MEK1), AR, GNA11, GNAQ, SMAD4, TSC1, DDR2, ETV6, FOXL2

This means we have a total of 72 fully curated genes:

ABL1, AKT1, AKT2, AKT3, ALK, ASXL1, ATM, BCR ,BRAF, BRCA1, BRCA2, BTK, CDK12, CDK4, CDK6, CEBPA, CTNNB1, DNMT3A, EGFR, ERBB2, ERBB3, EZH2, FGFR1, FGFR2, FGFR3, FGFR4, FLT3, IDH1, IDH2, JAK1, JAK2, JAK3, KIT, KMT2A ,KRAS, MDM2, MDM4, MET, MLH, MPL, MSH2, MSH6, NF1, NF2 ,NPM1, NRAS, PDGFRA, PDGFRB, PIK3CA, PMS2, PTCH1, PTEN, RET, ROS1, RUNX1, SF3B1, SMO, STK11, TET2, TP53, WT1, CD274 (PD-L1), HRAS, MAP2K1 (MEK1), AR, GNA11, GNAQ, SMAD4, TSC1, DDR2, ETV6, FOXL2

To view the full list of curated genes visit the About page on the Actionability website.
All previously-recorded clinical trials have been checked for new or updated results.

### Expressed/not category added to Patient Pre-screening

From v7 onwards the download file contains a new category: 'Expressed/not'
This is used for trials that compare patients that express a protein with those that don't or compare patients with high expression with those with low expression. In practice, there is usually a threshold expression level and the comparison is between patients above/below it. If our curator is able to find out the measure and threshold level that was used, it appears as part of the trial name.

This new value is represented by the term **Patient Pre-Screening**, in the column **mutation_selected_dict**

- positive/above threshold expression - patient number and results value recorded in **treatment values** column

- negative expression/below threshold values - value will appear in the fields used for control values, as is the case for trials that compare a treatment in patients with/without a mutation

There are several trials using this new category in v7.

### Addition of Australian/New Zealand Clinical Trials Registry

Actionability v7 includes the addition of a new datasource: the Australian New Zealand Clinical Trials Registry (ANZCTR). This can be seen in the **Source_Type** column as a value of 9.

| | |
|---|---|
| Genes fully curated (+11) | **72** |
| Genes included (+25) | **311** |
| Drugs (+112) | **1,520** |
| | **3,943** |

| | |
|---|---|
| Drug combinations (+362) | **6,609** |
| Evidence from trials databases (+481) | **2685** |
| Evidence from PubMed and other sources (+447) | **154** |
| Point mutations (+2) | **735** |
| Total mutations (+35) | |