CAIRN: Copy Alterations Intuitive-Rendering Navigator

Purpose: To enable the user to quickly and easily graph all copy-number alterations (CNAs) present in a segment file. This could include deletions and amplifications in tumors, for example. Custom data is permitted, so users can be creative in their use of this tool.

Background: Cancer is not genetically different from normal cells in only one way. Single nucleotide changes, which can cause mutations to coding regions for proteins, are a widely researched source of somatic mutation which gives rise to cancer cells. However, an equally important mutation type is changes in copy numbers of genes. This occurs through deletions or amplifications of regions on a chromosome, via complex mechanisms. These CNAs can be very small – less than the length of a single gene – or entire chromosomes. One prevailing theory for why they exist in cancer is that deletions encompass many tumor suppressor genes and amplifications encompass many oncogenes. Hence, being able to view how many deletions or amplifications occur over a given region of DNA may be informative for how a particular cancer type manipulates its genome to drive tumor progression. If deletions and amplifications are roughly equal, that region is unlikely to affect tumor formation or progression. If deletions dominate, this cancer likely requires removal of tumor suppressors within that region of DNA. Similarly, if amplifications dominate the region, a tumor likely requires more copies of one or more oncogenes in the area for maximal proliferation.

With this tool, you can:

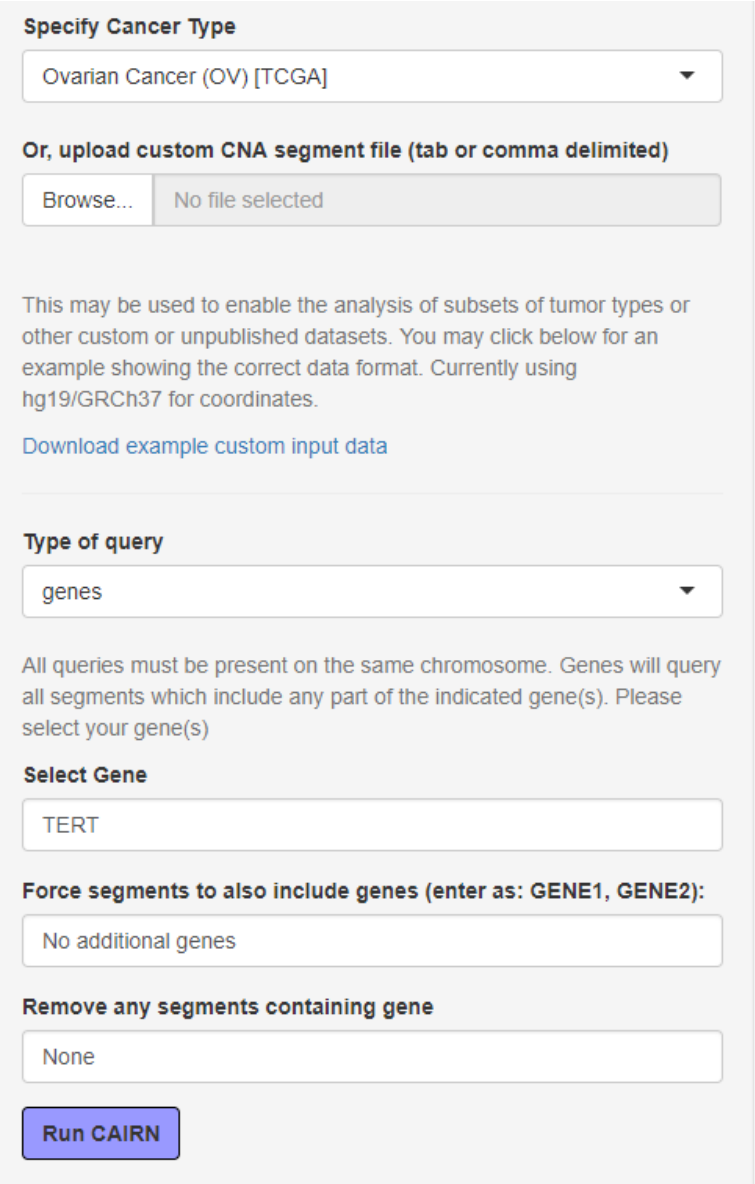
-Display copy-number alterations (CNAs) which overlap a user specified region
-Quantify the number of amplified CNAs and deleted CNAs

Options to modify your output include:

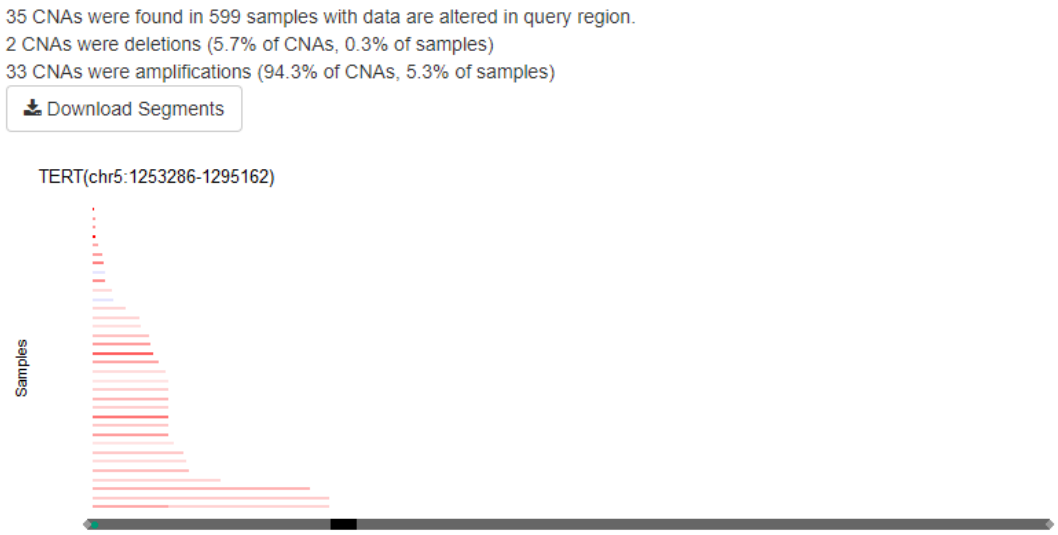
-Use over 30 cancer types (online)
-Use your own custom or downloaded segment dataset (an example is given)
-Choose to display only deletions or only amplifications if desired
-Query CNAs which end on a query region (for break points)
-Query CNAs which overlap a query region (for gene-level CNAs)
-Omit CNAs which also overlap another query region
-Customize the definition of CNAs by strength and length
-Display known oncogene and/or tumor suppressor locations
-Overlay single nucleotide and short indel mutation oncogenic data when available
-Customize the image size
-Download the segments corresponding to your query

An example for using CAIRN:

Say you are interested in telomerase in ovarian cancer: is it deleted or amplified by CNAs? To test this, you need to know the gene symbol for telomerase. For the catalytic subunit, this is TERT. If using the online tool, select or type "ovarian" into the "Specify Cancer Type" pull-down menu. If using the desktop app, download segment data from the UCSC Cancer Genome Browser or other resources and upload into the custom upload button. Ensure "Type of query" is set to "genes". Type "TERT" into the "Select Gene" box. Your query should look something like this:

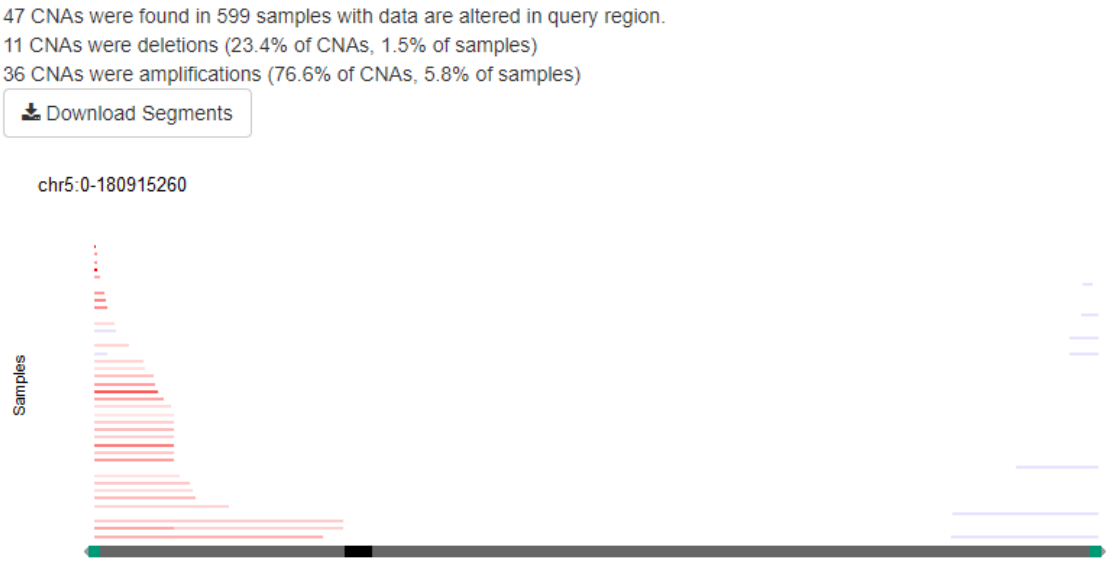


Clicking “Run CAIRN” will produce a graph like this:



Each unit of the Y-axis is a single tumor sample. The darker the red, the more highly amplified the segment is relative to normal 2N-copy DNA. The green dot shows where TERT is located – right next to the telomere (the ends of the grey segment represent telomeres). The black segment in the middle of the grey segment indicates the centromere of the chromosome.

Now you might be curious if all segments which end on the edge of this chromosome are amplifications. To test this, you may change the query to “ends”, type “chr5” as the chromosome, and check the box for “Display segments which start/end near telomeres”. After hitting Run CAIRN, you get a graph like this:



It may be somewhat difficult to see, but there are shallow deletions on the q arm telomere. Thus, not both telomeres of chromosome 5 are amplified, but rather only the telomere containing TERT.

Explanation of inputs:

Specify Cancer Type

In the online version, mutltiple tumor types are preloaded into CAIRN. For the desktop app, segment files must be loaded into the "Segments" folder or uploaded into the custom file input section.

Or, upload custom CNA segment file (tab or comma delimited)

Click to upload a custom CNA file for analysis. This overrides any "Specify Cancer Type" selection. Must be in hg19 coordinates for the original App. An example dataset can be found by clicking "Download example custom input data" or by finding the "Example_segments" file in the downloaded App.

Display segments which start/end near telomeres

Will produce a graph of all segments which have at least one end ending within 2 megabases of the telomeres end points, as defined by hg19 coordinates.

Display only segments which start/end near centromeres

Will produce a graph of all segments which have at least one end ending within 2 megabases of the centromere end points, as defined by hg19 coordinates.

Type of query

Selects between a gene-based analysis or a coordinate-based analysis (ends or overlaps). Ends query regions where CNA break points overlap the query. Overlap queries CNAs which extend beyond both ends of the query region.

Genes

If genes are selected, type in the official gene symbol of your gene of interest. Note that many genes have many synonyms, so search for the official gene symbol if unknown (for example, on uniprot.org or Wikipedia)

Force segments to also include genes (enter as: GENE1, GENE2)

This enables you to search for co-deleted or co-amplified genes. Type in one or more extra official gene symbols for the query if desired

Remove any segments containing gene

This enables a search for segments which only include one gene but not another gene. This may help quantify if two genes are both tumor suppressors or if one may be an oncogene and one a tumor suppressor, when combined with other search data.

Ends

This is used to monitor where break sites may occur in a particular cancer type.

Overlaps

This can be used if a non-gene query is desired. Some examples may include chromosome arm coordinates, entire chromosome coordinates, microRNA locations, or repetitive DNA coordinates.

Run CAIRN
This button will initiate a new calculation based on current input settings. It can also enable other graph labels to appear, but is not needed to resize the graph.

Minimum CNA amplitude
Set the magnitude of segment data values which correspond to a deletion or amplification change. Typically, this is 0.2 for shallow deletions which may occur in only a subset of tumor cells. Higher values may enable searches for more stringent data, or for custom data segments with copy number calls (use 1).

Minimum CNA length (bp)
The number of base-pairs in length each segment must be larger than to be displayed and counted.

Extend query region endpoints (bp)
Enables searches which include some DNA upstream and downstream of a gene or region. Type in zero for a query which does not include any extension beyond the query region.

Label COSMIC tumor suppressor genes
Labels genes which are annotated as tumor suppressors in COSMIC along the chromosome.

Label COSMIC oncogenes
Labels genes which are annotated as oncogenes in COSMIC along the chromosome.

Mark mutant tumor suppressor genes on segments
Additionally graphs triangle on any segment which contains a single nucleotide mutation or short indel of a COSMIC tumor suppressor gene on the query segments. Must click "Run CAIRN" to update the graph.

Mark mutant oncogenes on segments
Additionally graphs a green triangle on any segment which contains a single nucleotide mutation or short indel of a COSMIC tumor suppressor gene on the query segments. Must click "Run CAIRN" to update the graph.

Plot only deletions
Omits graphing of amplified regions.

Plot only amplifications
Omits graphing of deleted regions.

Specify exact plot output size
Check this box to specify the exact number of pixels you wish the graph to be.

Installation of CAIRN to your local computer

If you have never used a programming language, that's ok! This installation guide will still enable you to install and use CAIRN with minimal problems.

Requirements:
You must have R and RStudio installed, with the packages: shiny, data.table, dplyr, tidyr, ggplot2. All installations are freeware.

To install R: https://cran.r-project.org/
To install RStudio: https://www.rstudio.com/

Once these are both installed, restart your computer and open RStudio. In the lower left-hand window labeled "Console", paste in the following command:

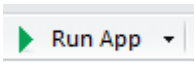
```
install.packages(c("shiny", "data.table", "dplyr", "tidyr", "ggplot2", "stringr"))
```

Hit enter and allow the packages some time to install. If asked to use a local library, click yes.

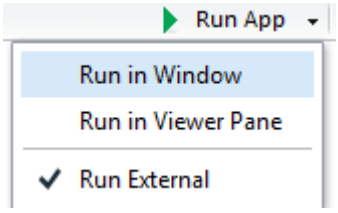
Download the zip file from the publication. Unzip the file onto your hard drive somewhere. Unzipping the file should enable all files to be in the correct location for immediate use on your computer. If you need software to unzip the file, you may Google "7zip" and use their free software.

Locate the "ui.R" file. Double click to open, or open this file from RStudio.

If Shiny Apps installed correctly into your local RStudio, you should see a small button which looks like this :



Click the black arrow on the right side of this button, and select "Run External"



Then click the green arrow, and CAIRN should load into your default browser.

License
Any usage or distribution of the underlying code is subject to the 2-Clause BSD License
https://opensource.org/licenses/BSD-2-Clause