

knotAnnotSV Manual

Version 1.0

<https://github.com/mobidic/knotAnnotSV>

Copyright (C) 2020-present GUIGNARD Thomas

Please feel free to contact me for any suggestions or bug reports
email: t-guignard@chu-montpellier.fr

TABLE OF CONTENTS

1.	INTRODUCTION	3
2.	REQUIREMENTS / INSTALLATION	3
a)	Requirements	3
b)	Installation	3
3.	INPUT	3
a)	Input files	3
b)	Arguments	4
4.	OUTPUT	4
a)	Output html format	4
b)	Output file path(s) and name(s)	4
c)	Compact and expanded visualization modes	4
d)	Default order of the annotation lines	5
e)	Available functions	5
f)	“Gene name” box color	6
5.	USAGE / OPTIONS	6
6.	Test	6
7.	FAQ	7

1. INTRODUCTION

knotAnnotSV is a program designed to create a customizable html file from an [AnnotSV](#) output file. This will allow to easily filter and explore each SV. Annotations can be directly displayed in the web browser or by mouse hovering (tooltip).

By default, the html generation is configured to optimize the analysis by clinicians and biologists. However, thanks to a configuration file, the user can:

- Customize the order, the choice, the name and the description of the annotation columns
- Choose to display the annotation directly or in a tooltip

Moreover, several functionalities (search, highlight, links to public databases, color coded information...) are available to help the user.

2. REQUIREMENTS / INSTALLATION

a) Requirements

The knotAnnotSV program is written in the Perl language. Modern Unix systems have this scripting language already installed. knotAnnotSV requires the following 2 Perl libraries (available via [CPAN](#)):

- YAML::XS
- Sort::Key::Natural

b) Installation

Please use git to download the most recent development tree.

The sources can be cloned to any directory:

```
cd /path/to/install/  
git clone https://github.com/mobidic/knotAnnotSV.git
```

3. INPUT

a) Input files

knotAnnotSV takes an AnnotSV output file (*'file'.annotated.tsv*) and a configuration file as inputs.

Configuration file:

knotAnnotSV uses an indented yaml file (e.g. config_cyto.yaml) to configure the AnnotSV output file. Several default configuration files are already distributed with the installation:

```
config_cyto.yaml  
config_AnnotSV.yaml
```

These configuration files are used by knotAnnotSV to:

- Define the POSITION (column ordering) of each field you want to display
- Define in COMMENTLIST which associated fields to display in tooltips (by mouse hovering)
- Define in RENAME which column name to display in the output
- Define in HEADERTIPS some information about the column you want to display in the header tooltips (by mouse hovering)

- Inactivate fields that you don't care about either with a starting '#' or 'POSITION: 0' or by deleting the corresponding line

NOTE : use space instead of tab for indentation, tabs are not allowed

b) Arguments

knotAnnotSV takes several arguments as input including options that are detailed in “USAGE / OPTIONS” section. The different arguments can be passed to the program using the command line.

4. OUTPUT

a) Output html format

Giving an AnnotSV tab-separated values file, an html file is created that can be displayed in a web browser.

At each stage of the analysis process, all the set-up filters are locally stored. The html file can thus be closed at any time by biologists and then re-opened to continue the analysis.

Browser compatibility:

Firefox 81.0
 Chrome 86.0.4240.75
 Edge 83.0.478.54
 IE 11

b) Output file path(s) and name(s)

Two options (--outDir and --outPrefix) can be used to specify the output directory and/or file name. By default, the html file is written in the current directory and named by replacing the ".tsv" extension of the AnnotSV file by ".html".

c) Compact and expanded visualization modes

AnnotSV context:

A typical AnnotSV use would be to first look at the annotation and ranking of each SV as a whole (i.e. “full”) and then focus on the content of that SV by genes (i.e. “split”). This is possible thanks to the 2 types of annotation lines provided by AnnotSV:

- An annotation on the “full” length of the SV. Every SV are reported, even those not covering any gene. This type of annotation gives an estimate of the SV itself.
- An annotation of the SV “split” by gene. This type of annotation gives an opportunity to focus on each gene overlapped by the SV. Thus, when a SV spans over several genes, the output will contain as many annotations lines as genes covered.

Visualization modes in knotAnnotSV:

Two visualization modes are available in the web browser:

- “Compact”: only the “full” AnnotSV lines are displayed
- “Expanded”: the “full” and “split” AnnotSV lines are displayed

Useful:

- In the compact mode, a double-click on a « full » line focuses on the relatives « split » lines. Double-click on the same « full » line to come back to the COMPACT mode
- The full lines are highlighted depending on their SV type (LOSS: red; GAIN: blue; INS: pink; INV: orange)

d) Default order of the annotation lines

By default, the annotation lines are sorted from the most (e.g. "1") to the least (e.g. "-1") pathogenic SV using the « AnnotSV_ranking_score » values.

e) Available functions

Top of the HTML page:

- Click to choose the desired visualization mode. For more explanation, see the « Compact and expanded visualization modes » section.
- Choose the number of lines to be displayed (50, 100 or all) in the upper left corner
- To search for a topic (gene name, phenotype...), just type a word or phrase in the Search box in the upper right corner

Header:

- Display the definition of the column name by mouse hovering (tooltip)
- Click small triangle in each column to sort the data
- Below each column name, a search box is available for:
 - Searching words
 - Filtering a range of numerical values (with the ">", ">=", "<" and "<=" symbols) and extracts matching records
e.g. to select frequencies smaller than 1%, type "<0.01"
 - Filtering out some words ("<>word" or "!=word")

Annotation lines:

- The « AnnotSV ID » column is freezed
- Mouse hover each annotation to show complementary information (tooltip)
- Double-click on a « full » line to focus on the relatives « split » lines (only for COMPACT mode). Double-click on the same « full » line to come back to the COMPACT mode
- Click on the « AnnotSV ID » to open the SV coordinates in the UCSC Genome browser (the SV region is automatically highlighted in blue)

Footer:

- Show the count of filtered lines (as well as the total entries count)
- Navigate between the different pages

NOTE:

- At each stage of the analysis process, all the set-up filters are locally stored. The html file can thus be closed at any time by biologists and then re-opened on the same computer to continue the analysis.

- To come back to the default annotation lines order, the user can click on the header of the first column (AnnotSV ID).

f) “Gene name” box color

Two different palettes of colors can be used to visualize for each gene its corresponding LOEUF_bin value (user defined, see the USAGE/OPTIONS section):

- A red-to-green palette (default):

The red color means the lower LOEUF_bin value (0). The green color means the higher LOEUF_bin value (9).

- A sequential palette of red:

Darker colors mean lower LOEUF_bin values (e.g 0).

NOTE : Depending on the gene overlap, the « Gene name » box is fully colored or not:

- Gene totally overlapped : fully colored box
- Gene partially overlapped : half colored box

5. USAGE / OPTIONS

To run knotAnnotSV, the default command line is the following:

```
cd /path/to/install/knotAnnotSV
perl ./knotAnnotSV.pl --configFile config_AnnotSV.yaml --annotSVfile
file.annotated.tsv
```

The command line can be completed by some options.

To show the options simply type:

```
perl ./knotAnnotSV.pl
```

```
--configFile    <YAML config file for customizing output>
--annotSVfile    <AnnotSV annotated file>
--outDir         <output directory (default = current dir)>
--outPrefix      <output file prefix (default = "")>
--genomeBuild    <Genome Assembly reference (default = hg19)>
--LOEUFcolorRange <Number to define which color to use for LOEUF bin: 1 (red-to-green), 2 (red-
shades-only) (default = 1)>
--datatableDir   <Local Path to DataTables directory containing css and js files (default = "", requires
web connection)>
```

6. Test

In order to validate the knotAnnotSV installation and to facilitate its use we have provided with the installation an input/output example in the « example » folder.

1. Change to the repo directory, and run the example

```
cd /path/to/install/knotAnnotSV/  
perl ./knotAnnotSV.pl --annotSVfile ./example/example.annotated.tsv  
--configFile ./config_AnnotSV.yaml
```

2. Display the html output on a web browser
3. Have fun with the exploring!

7. [FAQ](#)

Q: What does knotAnnotSV means ?

If pronounced like « knot a knot SV », knotAnnotSV is a word game meaning that this tool creates a link between the SV data to be analysed :o)