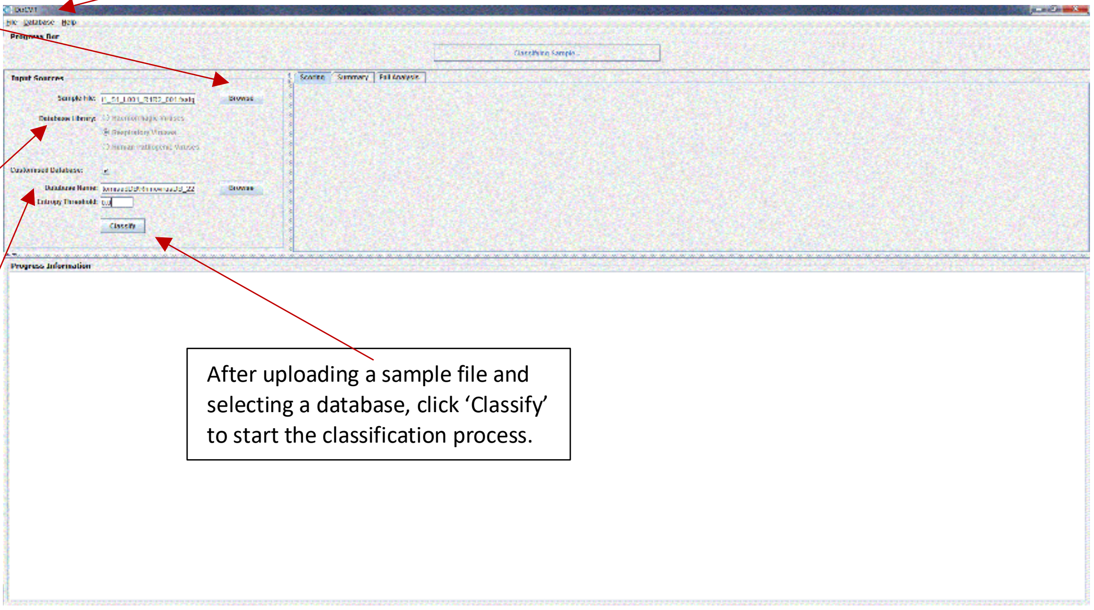
This guide provides an overview demonstration of DisCVR GUI and how to use it to classify a clinical sample from HTS data and validate the results using reference genome alignment. For more details about installation and system requirements, please refer to the Operating Manual.

Sample Classification



Database icon shows a list of all viruses in the Database Library.

Upload a sample file to classify. The file must be .fastq or .fastq.gz but not .fasta

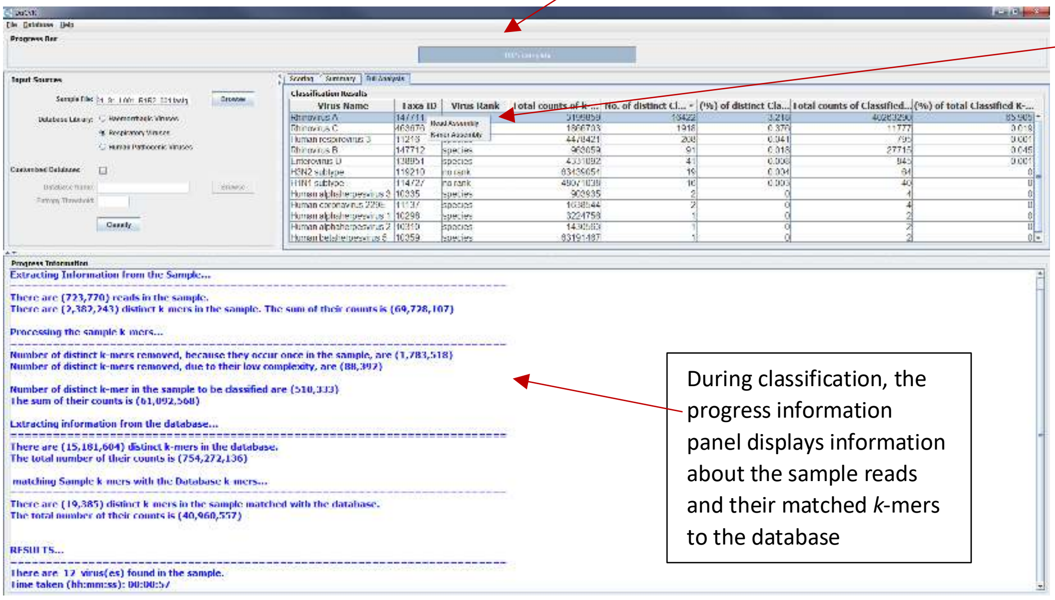
To classify the sample using DisCVR's built-in database, select one from the Database Library.

To classify the sample using your own database, check Customised Database, upload the database file and enter a value for the entropy.

Consult the Operating Manual about building a customised database.

After uploading a sample file and selecting a database, click 'Classify' to start the classification process.

Results Validation



Classification Results

Virus Name	Taxa ID	Virus Rank	Total counts of k-...	No. of distinct Cl...	(%) of distinct Cla...	Total counts of Classified...	(%) of total Classified K-...
Rhinovirus A	147711		5199859	18422	3.210	40283290	85.905
Rhinovirus C	463676		1896703	1918	0.376	11777	0.018
Human respirovirus 3	11216		4478421	208	0.041	795	0.001
Rhinovirus B	147712	species	963059	91	0.018	27715	0.045
Enterovirus D	138951	species	4331092	41	0.008	945	0.001
H3N2 subtype	119210	no rank	83439054	19	0.004	64	0
H1N1 subtype	114727	no rank	49071038	16	0.003	40	0
Human alphaherpesvirus 3	10335	species	903935	2	0	4	0
Human coronavirus 229E	11137	species	1638544	2	0	4	0
Human alphaherpesvirus 1	10298	species	3224758	1	0	2	0
Human alphaherpesvirus 2	10310	species	1430563	1	0	2	0
Human betaherpesvirus 5	10359	species	83191487	1	0	2	0

```
Extracting Information from the Sample...
------------------------------------------------------------------------
There are (723,770) reads in the sample.
There are (2,382,243) distinct k mers in the sample. The sum of their counts is (69,728,107)

Processing the sample k mers...
------------------------------------------------------------------------
Number of distinct k-mers removed, because they occur once in the sample, are (1,783,518)
Number of distinct k-mers removed, due to their low complexity, are (88,392)

Number of distinct k-mer in the sample to be classified are (510,333)
The sum of their counts is (61,092,568)

Extracting information from the database...
========================================================================
There are (15,181,604) distinct k-mers in the database.
The total number of their counts is (754,272,136)

 matching Sample k mers with the Database k mers...
------------------------------------------------------------------------
There are (19,385) distinct k mers in the sample matched with the database.
The total number of their counts is (40,960,557)


RESULTS...
------------------------------------------------------------------------
There are  12  virus(es) found in the sample.
Time taken (hh:mm:ss): 00:00:57
```

The classification contains four stages and the progress bar is updated after each stage.

To validate the significance of the results, right click on a row and then choose either to align the matched *k*-mers, or the sample reads to the reference genome of the selected virus.

Consult the Operating Manual for information about the Scoring and Summary panels.

During classification, the progress information panel displays information about the sample reads and their matched *k*-mers to the database