

## CTTV001 – CTTV database project's data format



### Introduction to the CTTV evidence string format

Samiul Hasan (samiul.x.hasanATgsk.com)

#### Aim

The CTTV “evidence string format” defines a graph-based model for capturing target (i.e. a gene or a protein) to disease evidence from a number of disparate data sources. It is effectively the “currency” for the CTTV database project (CTTV001) to receive and process disease evidence from different CTTV workstreams.

#### Description of the format

The platform is currently accepting data in JSON (javascript object notation) format. Please use the provided JSON template file (evidence\_string\_template.json) as a guide to preparing your data to fit this model.

#### Current status

The JSON template is currently provided in a descriptive format. We are working in the mean time to prepare a JSON schema to ensure validity of the data provided. We will communicate when this is ready. You can use the descriptive template in the mean time to prepare your data. It details the gene/protein identifiers and evidence code/bioontologies and the structure in which to provide this. We may develop formats in addition to JSON in the future for pulling data into the system.

#### Important contacts

- For questions or suggesting changes about the JSON schema: please email Samiul Hasan, cc Michael Maguire (mmaguireATebi.ac.uk).
- For requesting new evidence codes that better “label” the evidence you are providing, please email Samiul Hasan, cc James Malone (maloneATebi.ac.uk).

## CTTV001 – CTTV database project's data format

Figure illustrating the structure of the evidence graph model.

