

# HIPC Dashboard curation template fields description

Last updated December 17, 2021

## Overview

This document describes the format of the data templates currently in use for the HIPC Dashboard immune response curation projects, including immune exposures of type vaccination and infection, and measured response types of gene expression and cell-type frequency. Legacy curations may not have all columns.

## Capitalization

Values should be entered in lower case except for the following:

| Column name                                       | Example                |
|---|------------------------|
| response_component (for case of gene symbols)     | STAT1                  |
| target_pathogen (vaccines) - first letter of name | Neisseria meningitidis |
| exposure_material - first letter of name          | Menactra               |

# Curation template columns

The table below is a transposed version of the annotation sheets. Each row of the table represents a column of the template. There are currently two main template types: **vaccine** and **infection**. Columns specific to a given template type are color-coded in the table below with blue for **vaccine** and green for **infection**.

In the annotation sheet, the person curating a publication can add as many signature rows as needed to capture all pertinent immune signatures found, one row per signature.

The columns are described in the order they appear in the curation templates. The terms are ordered to provide a logical progression to the curation process.

| <u>column name</u>   | <u>Descriptive text on Dashboard</u>   | <u>Vocabulary</u> | <u>Additional information for curators</u>   |
|----------------------|--|-------------------|--|
| <b>curation_date</b> | curation date (YYYY-MM-DD)   | YYYY-MM-DD        |  |
| <b>cohort</b>        | cohort - any characteristics of the population(s) studied, plus whether the result was taken from a subgroup of the broader cohort tested. | free text         | <p>Example: Hong Kong, Atlanta</p> <p>Also report if the result was limited to a subgroup of the tested cohort, e.g.</p> <ul style="list-style-type: none"><li>• subjects suffering adverse events</li><li>• particular threshold levels of antibody titer level</li><li>• based on a receiving a particular treatment</li></ul> |

|                            |   |   |   |
|----------------------------|---|---|---|
| <b>age_min</b>             | age_min - age of youngest subject including both cases and controls               | number  | Include both case/affected and control subjects   |
| <b>age_max</b>             | age_max - age of oldest subject including both cases and controls                 | number  | include both case/affected and control subjects   |
| <b>age_units</b>           | age units   | choose from list  | hours, days, months, years  |
| <b>number_subjects</b>     | number of subjects - count of case plus control subjects used in the measurement. | number  | Often differs by signature within a publication. If number of subjects for a particular signature is not clear in text, use total for cohort.   |
| <b>tissue_type</b>         | tissue type   | As reported. Only one entry allowed.  | The parent tissue of the response components. Report multiple tissues in different rows. For cell-type results, the tissue is often PBMCs. For gene expression, the tissue might be a particular cell-type. |
| <b>tissue_type_term_id</b> | Cell Ontology ID of tissue  | Choose from list or <a href="#">lookup Cell Ontology code</a> . Use format type:code e.g. | If there is no matching cell type in the pulldown list, the curator can try to look up a matching term in the <a href="#">Cell Ontology</a> . If there is no appropriate Cell                               |

|                           |   |   |   |
|---------------------------|---|---|---|
|                           |   | CL:0000236,<br>UBERON:0000178.  | Ontology term, UBERON codes can also be used.   |
| <b>method</b>             | method - primary experimental method used to measure the response | Choose from list. Only one entry expected. You can add new methods.   | The primary experimental method used to measure the response, e.g. RNA-seq, CyTOF, CITE-seq.  |
| <b>response_component</b> | response component  | Gene or protein <b>symbols</b> can be separated with commas, spaces, or semicolons. <b>Cell types or other names must be separated using semicolons</b> | The entities whose response is being measured. Please copy symbols and names exactly as reported in the publication - <b>except spell out greek letters or other special characters</b> . For cell types, this includes all markers. Examples for a signature with three cell types: T cells CD3+/CD4+/Ki67+; T cells CD3+/CD8+/Ki67+ CD86+ myeloid dendritic cell (DC); CD86+ monocyte |
| <b>is_model</b>           | signature was derived from a computational model                  | Y/N   | Were the response components chosen using a classification or other model-building strategy?  |

|                               |  |  |   |
|-------------------------------|--|--|---|
| <b>response_behavior_type</b> | response behavior type                               | Choose from list. Only one entry allowed.                                | The type of change being measured, e.g. gene expression, cell-type frequency.   |
| <b>response_behavior</b>      | response behavior (direction, correlation type etc.) | Choose from list. Only one entry allowed. Add new behaviors if required. | <p>Common values in the pulldown list:</p> <ul style="list-style-type: none"> <li>• up, down</li> <li>• positively correlated, negatively correlated, correlated</li> <li>• positively predictive, negatively predictive, predictive</li> </ul> |

## comparison

comparison (affected vs control, correlated variable, time vs baseline event)

Free text. **Use "vs" rather than a dash to separate comparison terms (A vs B)**. Separate multiple comparison entries with semicolons (A vs B; C vs D).

Comparisons are typically between two groups, or may reflect a correlation of the response component with some other measured variable. **Only report significant results.**

Examples:

- severe COVID-19 cases vs healthy
- moderate COVID-19 cases vs healthy, severe COVID-19 cases vs healthy
- interferon-stimulated genes in COVID-19 vs healthy
- bacterial DNA levels across COVID-19 and healthy subjects

Include time comparison if relevant, e.g. 7d vs 0d, where the times are relative to the baseline reference event time (0d). Times before baseline event can be entered as negative numbers, e.g. days before vaccination (7d vs -1d).

Please be concise\*

|  |  |  |  |
|--|--|--|--|
| <b>baseline_time_event</b>   | baseline time event  | Free text. Use "none" if no time was reported.   | The reference event from which the time of the experimental response is measured, e.g. hospital admission, onset of symptoms   |
| <b>time_point</b>  | time point relative to baseline event at which response was measured   | Number. Use "none" if no time was reported.  | Time point when response was measured, if available.   |
| <b>time_point_units</b>  | time point units   | Choose from list.<br>Lowercase only  | days, months etc.  |
| <b>exposure_material</b>   | <b>infection:</b> exposure material (pathogen name);<br><b>vaccine:</b> exposure material (vaccine name)             | free text  | Enter the pathogen or vaccine underlying the immune exposure as reported in the publication or use the NCBI Taxonomy term name |
| <b>exposure_material_id</b><br><b>(vaccine and infection templates use different ontologies)</b> | <b>infection:</b> exposure material NCBI taxonomy ID;<br><b>vaccine:</b> exposure material vaccine ontology ID (VO). | <b>infection:</b> Choose from list or use format ncbi_taxid:2697049.<br><b>vaccine:</b> use format VO:0000045. | For infection template, use NCBI Taxonomy ID of pathogen causing disease.  |
| <b>exposure_process</b><br><b>(infection template)</b>   | exposure process - method by which immune exposure occurred  | Choose from list. Enter new process if needed.   | Method by which exposure to pathogen occurred  |

|   |  |                     |  |
|---|--|---------------------|--|
| <b>disease_name (infection template)</b>                | disease name   | free text           |  |
| <b>disease_stage (infection template)</b>               | disease stage - reported disease stage(s) of affected subjects | free text           | Reported disease stage(s) of affected subjects in all comparisons entered in row (not including control subjects). We will not attempt to match these directly with the comparisons. |
| <b>additional_exposure_material (vaccine templates)</b> | exposure material - additional                                 | free text           | Any additional exposure material, e.g. "Live attenuated vaccine TC-83 challenge", "ex-vivo restimulation with live VZV"  |
| <b>target_pathogen (vaccine templates)</b>              | target pathogen  | text                | NCBI taxonomy name (non-influenza pathogens). For influenza, can just note e.g. "influenza A virus; influenza B virus". Values will be filled in by script based on vaccine year.    |
| <b>target_pathogen_taxonid (vaccine templates)</b>      | NCBI TaxonID (non-influenza)                                   | ncbi_taxid:nnnnnnnn | Currently optional but plan to switch to this.   |
| <b>vaccine_year (vaccine templates)</b>                 | vaccine year (influenza only)                                  | YYYY                | The official year of the vaccine, as a key to its composition.   |
| <b>adjuvant (vaccine templates)</b>                     | adjuvant   | free text           |  |



|   |  |   |  |
|---|--|---|--|
| <b>route (vaccine templates)</b>                | route  | free text   | i.m, i.n, i.d., po, subcutaneous etc.  |
| <b>scheduling (vaccine templates)</b>           | scheduling   | free text   | Has been used to record number of doses, e.g. 1 dose, 2 doses.   |
| <b>publication_reference_id</b>                 | publication reference (PMID, bioRxiv etc)          | type:id   | Enter an official identifier for the article curated, using format type:id,such as pmid:32788292                           |
| <b>publication_date</b>                         | print publication date or posted date (YYYY-MM-DD) | YYYY-MM-DD  | Partial dates are OK, e.g. YYYY-MM   |
| <b>publication_reference_url</b>                | publication URL                                    | The URL only, not any associated display text   | URL of the article curated. Please use PubMed if available   |
| <b>signature_source</b>                         | signature source - figure, table or text section   | Free text. Up to two entries, separated by semicolons                                     | Figure, table number etc. where the signature was found, usually the original primary publication rather than PMC          |
| <b>signature_source_url (vaccine templates)</b> | signature source url                               | The URL only, not any associated display text. Up to two entries, separated by semicolons | Link to figure, table number etc. where the signature was found. Please enter the URL directly, not as a hyperlinked name. |

|                         |                                 |                                    |   |
|-------------------------|---------------------------------|------------------------------------|---|
| <b>comments</b>         | comments and additional details | Free text, limit of 255 characters | Details clarifying any aspect of the signature not captured in other fields           |
| <b>curator_comments</b> | curator comments                | Free text, no limit                | Questions or notes for further examination from curator. Will not appear in Dashboard |