HIPC Dashboard curation template fields description

Last updated 8/23/2021

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.

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.

column name	Descriptive text on Dashboard	Vocabulary	Additional information for curators
curation_date	Date curation performed (YYYY-MM-DD)	YYYY-MM-DD	
cohort	Any characteristics of the population(s) studied, plus whether the result was taken from a subgroup of the cohort tested.	free text	Example: Hong Kong, Atlanta Also report if the result was limited to a subgroup of the tested cohort, e.g. - subjects suffering adverse events - particular threshold levels of antibody titer level - based on a receiving a particular treatment

age_min	age of youngest subject including both cases and controls	number	Include both case/affected and control subjects
age_max	age of oldest subject including both cases and controls	number	include both case/affected and control subjects
age_units	age units	choose from list	Hours, days, months, years
number_subjects	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.
tissue_type	tissue type (as reported)	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.
tissue_type_term_id	Cell Ontology ID of tissue	Choose from list or lookup Cell Ontology code	If there is no matching cell type in the pulldown list, the curator can try to look up a matching term in the Cell Ontology.

method **(new column)**	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.
response_component	response component - genes, proteins, cell types etc. are the entities whose change with infection is being measured	Gene or protein **symbols** can be separated with commas, spaces, or semicolons. **Cell types or other names must be separated using semicolons**	Please copy symbols and names exactly as reported in the publication - **except spell out greek letters or other special characters**. 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
is_model (gene expression template)	signature is derived from a computational model	Y if yes, N or blank otherwise	Were the response components chosen using a classification or other model-building strategy?
response_behavior_type	response behavior type	Choose from list. Only one entry allowed.	The type of change being measured, e.g. gene expression, cell-type frequency.

response_behavior	response behavior (direction, correlation type etc.)	Choose from list. Only one entry allowed. Add new behaviors if required.	Common values in the pulldown list: • up, down • positively correlated, negatively correlated, correlated • positively predictive, negatively predictive, predictive

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*

baseline_time_event	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
time_point	time 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.
time_point_units	time point units	Choose from list	Days, months etc.
exposure_material	infection: exposure material (pathogen name); vaccine: exposure material (vaccine name)	free text	Enter the pathogen or vaccine underlying the immune exposure as reported in the publication
exposure_material_id **(vaccine and infection templates use different ontologies)**	infection:exposure material NCBI taxonomy ID; vaccine: exposure material vaccine ontology ID (VO).	infection: Choose from list or use format ncbi_taxid:2697049. vaccine: use format VO:0000045.	For infection template, use NCBI Taxonomy ID of pathogen causing disease.
exposure_process (infection template)	exposure process - method by which immune exposure occurred	Choose from list. Enter new process if needed.	Method by which exposure to pathogen occurred

disease_name (infection template)	disease name	free text	
disease_stage (infection template)	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.
additional_exposure_material (vaccine templates)	exposure material - additional	free text	Any additional exposure material, e.g. "Live attenuated vaccine TC-83 challenge", "ex-vivo restimulation with live VZV"
target_pathogen (vaccine templates)	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.
target_pathogen_taxonid (vaccine templates)	NCBI TaxonID (non-influenza)	ncbi_taxid:nnnnnnnn	Currently optional but plan to switch to this.
vaccine_year (vaccine templates)	vaccine year (influenza only)	YYYY	The official year of the vaccine, as a key to its composition.
adjuvant (vaccine templates)	adjuvant	free text	

route (vaccine templates)	route	free text	i.m, i.n, i.d., po, subcutaneous etc.
scheduling (vaccine templates)	scheduling	free text	Has been used to record number of doses, e.g. 1 dose, 2 doses.
publication_reference_id	publication reference (PMID, bioRxiv etc)	type:id	Enter an official identifier for the article curated, using format type:id,such as pmid:32788292 (default pmid)
publication_year	print publication year	YYYY	The year in which the official printed article appeared.
publication_reference_url	publication URL	URL	URL of the article curated. Primary source publication (e.g. Cell, Nature) preferred over PMC
signature_source	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
signature_source_url	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, usually the original primary publication rather than PMC. Please enter the URL directly, not as a hyperlinked name.

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