

SARS-CoV-2 Drug Resistance Database

April 8, 2021

Version 1.0

Descriptions of tables

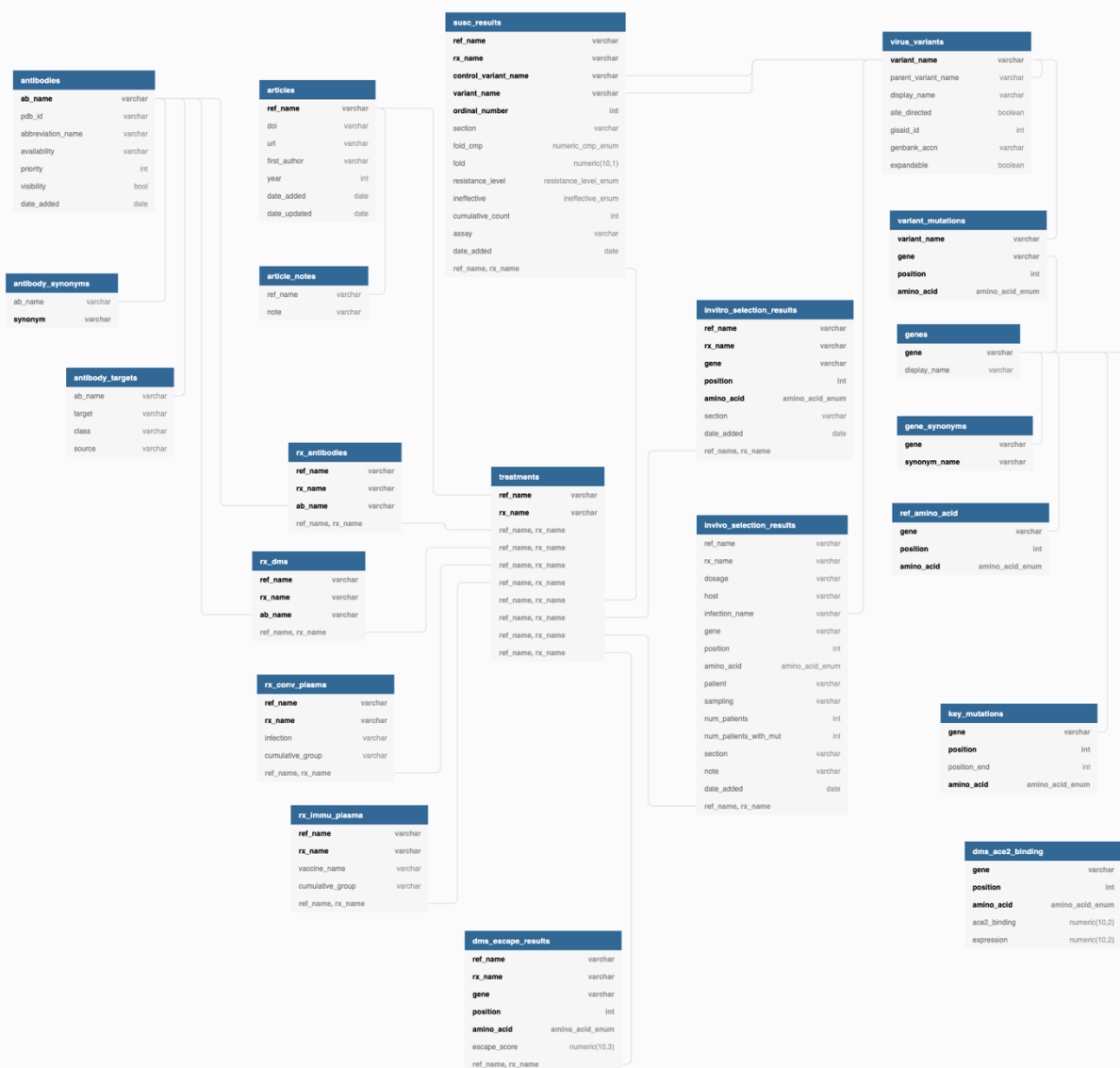
The main goals of this database are to capture the following data:

- I. The neutralizing susceptibility of variants to (1) monoclonal antibodies (mAbs); (2) convalescent plasma (CP); and (3) plasma from vaccinated persons (VP).
- II. The spike mutations selected in vitro by mAbs and convalescent plasma, and in vivo by mAbs.

The main tables include:

- I. The **articles** table contains data about each reference.
- II. The **virus_variants** table contains a list of variants, including individual mutations, combinations of mutations, and named variants (i.e., a variant of concern or variant of interest).
- III. The **variant_mutations** table links mutations to variants.
- IV. The **treatments** table contains unique pairs of studies and treatments and which functions as a parent table to data about mAbs (**rx_antibodies**), convalescent plasma (**rx_conv_plasma**), and plasma from vaccinated persons (**rx_immu_plasma**). The **rx_** tables contain information about mAbs, convalescent plasma, and plasma from vaccinated persons.
- V. The tables **antibodies**, **antibody_synonyms**, and **antibody_targets** contain data about each mAb.
- VI. The **susc_results** table contains neutralization data linking a virus variant and treatment.
- VII. The **invitro_selection_results** and **invivo_selection_results** tables contain mutations associated with a treatment.

Database schema



Footnote: For each rectangle, text on darkblue background is the table name, texts in bold style are primary keys. Table relationships are represented by texts on lightblue background and blue linkage lines. If the foreign key contains one field, the linkage line is linked between corresponding fields on each table. If the foreign key contains multiple lines, combination of fields is shown at the bottom of the table (fields are separated by comma) and linked to corresponding combination of fields of another table. “*” means many. You can see high resolution [here](#).

Entity tables

Table name	Table definition	Field name	Field definition
articles	Refererence meta info.	ref_name	Assigned name (RefID, RefName) for a reference, usually formed by the surname of first author, two-digit year and an optional alphabetical letter for distinguishing studies published in the same year.
		doi	Digital Object Identifier (DOI) of a reference.
		url	Web page of a reference, should be specified when DOI is not available.
		first_author	First author name, surname + first letters of other names.
		year	Year the reference was published.
		date_added	Date the reference added in database.
		date_updated	Date the reference updated in database.
article_notes	Notes and comments of references.	ref_name	Foreign key for articles.
		note	Additional notes and comments of the reference, free text.
antibodies	mAb meta info.	ab_name	Assigned mAb name. This is the most commonly recognized name.
		pdb_id	PDB ID when available.
		abbreviation_name	mAb abbreviation.
		availabiliy	Availability of mAb, for example, EUA, Phase 3, etc.
		priority	Priority when shown in susceptibility reports. Lower number means higher priority.
		visibility	Shown in report or not.
		date_added	Date the mAb added in database.
antibody_targets	mAb structure and classification meta info.	ab_name	Assigned mAb name. Foreign key for antibodies.
		target	Target sub-domain in SARS-CoV-2 spike (i.e., NTD, RBD, S1-CTD, or S2)
		class	Epitope classification of RBD-targeting mAbs (i.e., RBM class I, RBM class II, RBD-Core I, or RBD-Core II).
		source	Source of the classification. If the classification is from PDB file, source is "structure". If the classification is from the reference, source is "author".
antibody_synonyms	Synonyms of mAb	ab_name	Assigned mAb name. Foreign key for antibodies.
		synonyms	mAb synonym.
virus_variants	Virus variants, mutations or combination of mutations.	variant_name	Unique variant name. Allowable formats include: (i) Pango lineage plus any additional required descriptors such as a mutation; (ii) Commonly used name for a variant, such as a name assigned by the WHO or CDC; VOC name; (iii) one or more mutations

			(<gene_name>:<position><mutation>) separated by "+". This format is used for site-directed mutants or for viruses defined by the presence of specific mutations; (iv) another trackable name (GISAID ID, NextStrain lineage name, name used in a published reference).
		parent_variant_name	Parent variant name. Foreign key for virus_variants.
		display_name	Display name of variant.
		site_directed	Variant created by site-directed mutagenesis or not.
		gisaid_id	GISAID isolate accession ID without the common prefix "EPI_ISL_" (optional).
		genbank_accn	Genbank accession number (optional).
		expandable	For SARS-CoV-2 related variants, the program using this database can expand all the mutations, for non-SARS-CoV-2 coronaviruses the number of mutations is huge, the program should not expand all the mutations. (Yes/No)
variant_mutations	Mutation list of a variant. One row represents one mutation of variant.	variant_name	Foreign key for virus_variants.
		gene	Foreign key for gene.
		position	Mutation position.
		amino_acid	Mutation amino acid.
genes		gene	SARS-CoV-2 gene name.
		display_name	SARS-CoV-2 gene display name.
gene_synonyms		gene	Foreign key for gene.
		synonym_name	Synonym of gene.
ref_amino_acid		gene	Foreign key for gene.
		position	Gene position.
		amino_acid	Reference amino acid of reference SARS-CoV-2 genome.
key_mutations	Key mutations for generating reports	gene	Foreign key for gene.
		position	Gene position.
		position_end	End of a range of positions. For insertion or deletion.
		amino_acid	Amino acid mutation.
dms_ace2_binding	Deep mutation sequencing results for all amino acid mutations	gene	Foreign key for gene.
		position	Gene position.
		amino_acid	Amino acid mutation.
		ace2_binding	ACE2 binding score. Minus sign means reduction of binding.
		expression	Spike protein expression score, minus means reduction of expression.

Treatment tables

Table name	Table definition	Field name	Field definition
treatments	Contains the list of distinct treatments for each reference.	ref_name	Foreign key for articles.
		rx_name	The “name” of the mAb, CP, or VP used in the reference.
rx_antibodies	mAb treatment meta info	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	mAb name used in the reference. Can be a single mAb or mAb combination.
		ab_name	Assigned mAb name. Foreign key for antibodies.
rx_conv_plasma	CP treatment meta info	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	Description of CP used in the reference. Most commonly this is represented by “CP”. But it also may be followed by additional descriptors such as the likely variant responsible for infection, the severity of infection, the time since infection, and the neutralizing antibody titer.
		infection	The variant responsible for infection. When left blank it means “Unknown”.
		cumulative_group	Cumulative group name for aggregated samples. A group of samples from same treatment can be combined together for additional analysis.
rx_immuplasma	VP treatment meta info	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	Description of VP including the vaccine used and any additional information such as the time since vaccination.
		vaccine_name	Enumerated list of vaccines [BNT162b2, mRNA-1273, AZD1222, NVX-CoV2373, BBIBP-CorV, CoronaVac]
		cumulative_group	Cumulative group name for aggregated samples. A group of samples from same treatment can be combined together for additional analysis.
rx_dms	DMS treatment meta info	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	mAb name used in the DMS reference. Can be a single mAb or mAb combination.
		ab_name	Assigned mAb name. Foreign key for antibodies.

Result tables

Table name	Table definition	Field name	Field definition
susc_results	Susceptibility/Resistance results.	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	The “name” of the mAb, CP, or VP used in the reference.
		control_variant_name	Variant used for control (i.e fold reduced susceptibility is the IC ₅₀ of the experimental variant / IC ₅₀ of the control variant). Commonly used controls including Wuhan variant, D614G, and one of the VOCs. Foreign key for virus_variants
		variant_name	Variant undergoing susceptibility testing. Foreign key for virus_variants.
		ordinal_number	For CP or VP samples tested against the same variant, plasmas from the same group should have an ordinal number to distinguish the results and record in different rows. For mAb samples, the value is always 1.
		section	The location in the reference containing the result (e.g. Figure 1, table 1 or page 1, para 4).
		fold_cmp	"=", "~", ">", or "<", default "=", modifier of fold.
		fold	Fold change compared to the control variant. This is usually based on the IC ₅₀ . It is occasionally based on a binding assay if neutralization data are not available.
		resistance_level	This is for studies that do not report qualitative data.
		ineffective	If the neutralizing result on the control variant or the experimental variant is below detectable threshold, this field should be used. “control”: control variant neutralizing is undetectable, “experimental”: experimental variant neutralizing is undetectable, “both”: both of them are undetectable.
		cumulative_count	For some studies, it’s not possible to extract individual results but a mean or median value is available. In this case the fold field will contain the mean or median and the number of aggregated samples

			is recorded in this field. If the result is not for aggregated data, this field will contain 1.
		assay	Pseudovirus, live virus, or name of binding assay.
		date_added	Date the result added in the database.
invitro_selection_results	<i>in vitro</i> selected mutations, one row represents one selected mutation.	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	mAb name used in the reference. Can be a single mAb or mAb combination.
		gene	Mutation gene.
		position	Mutation position.
		amino_acid	Mutation amino acid.
		section	The location in the reference containing the result (e.g. Figure 1, table 1 or page 1, para 4).
		date_added	Date the result added in the database.
invivo_selection_results	<i>in vivo</i> selected mutations, one row represents one selected mutation.	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	mAb name used in the reference. Can be a single mAb or mAb combination.
		dosage	Dosage used in the treatment.
		host	Host of the in vivo selection experiment.
		infection_name	Virus variant responsible for infection. Foreign key for virus_variants.
		patient	Patient description in the reference, for example, hospitalized or nonhospitalized.
		sampling	Sampling date after treatment or infection. "D" means days, "M" means months.
		section	The location in the reference containing the result (e.g. Figure 1, table 1 or page 1, para 4).
		note	Additional information from the reference, free text.
		date_added	Date the result added in the database.
dms_escape_results	DMS results for mAbs	ref_name	Together with the rx_name this is a foreign key for treatments.
		rx_name	mAb name used in the reference. Can be a single mAb or mAb combination.
		gene	Spike.

		position	Mutation position.
		amino_acid	Mutation amino acid.
		escape_score	Escape score of the DMS result. 0 means no escape, 1 means full escape.