Biocat Database - Entity Relationship Diagram (Chen's Notation)

Nikhil R - PES1UG23CS392
Prakyath P. Nayak - PES1UG23CS431

genetic_variant
- variant_id : INT(10) «PK»
- chromosome_id : INT(10) «FK»
- variant_name : VARCHAR(255)
- variant_type : ENUM
- start_position : BIGINT
- end_position : BIGINT
- reference_allele : TEXT
- alternate_allele : TEXT
- variant_class : VARCHAR(100)
- minor_allele_frequency : DECIMAL(6,5)
- validation_status : VARCHAR(50)
- clinical_significance : VARCHAR(100)

user_account
- user_id : INT(10) «PK»
- username : VARCHAR(100)
- email : VARCHAR(255)
- password_hash : VARCHAR(255)
- full_name : VARCHAR(255)
- institution : VARCHAR(255)
- user_role : ENUM
- is_active : TINYINT(1)
- created_at : TIMESTAMP
- last_login : TIMESTAMP

data_source
- source_id : INT(10) «PK»
- source_name : VARCHAR(100)
- source_url : VARCHAR(500)
- source_version : VARCHAR(50)
- release_date : DATE
- description : TEXT
- contact_email : VARCHAR(255)
- is_active : TINYINT(1)

gene
- gene_id : INT(10) «PK»
- species_id : INT(10) «FK»
- chromosome_id : INT(10) «FK»
- gene_stable_id : VARCHAR(100)
- gene_symbol : VARCHAR(100)
- gene_name : TEXT
- gene_description : TEXT
- gene_biotype : VARCHAR(100)
- start_position : BIGINT
- end_position : BIGINT
- strand : TINYINT
- is_canonical : TINYINT(1)
- created_at : TIMESTAMP
- updated_at : TIMESTAMP

variant_gene_annotation
- annotation_id : INT(10) «PK»
- variant_id : INT(10) «FK»
- gene_id : INT(10) «FK»
- variant_effect : VARCHAR(100)
- impact_severity : ENUM
- annotation_source : VARCHAR(100)

query_history
- query_id : INT(10) «PK»
- user_id : INT(10) «FK»
- query_text : TEXT
- query_type : VARCHAR(50)
- execution_time : DECIMAL(8,4)
- result_count : INT
- query_timestamp : TIMESTAMP
- ip_address : VARCHAR(45)

data_import_log
- import_id : INT(10) «PK»
- source_id : INT(10) «FK»
- import_type : VARCHAR(100)
- records_imported : INT
- records_failed : INT
- import_status : ENUM
- start_time : TIMESTAMP
- end_time : TIMESTAMP
- error_message : TEXT
- imported_by : VARCHAR(100)

gene_go_annotation
- annotation_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- go_id : VARCHAR(20) «FK»
- evidence_code : VARCHAR(10)
- annotation_source : VARCHAR(100)
- annotation_date : DATE

gene_ontology
- go_id : VARCHAR(20) «PK»
- go_name : VARCHAR(255)
- go_namespace : ENUM
- go_definition : TEXT
- is_obsolete : TINYINT(1)

gene_pathway_annotation
- annotation_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- pathway_id : INT(10) «FK»
- annotation_source : VARCHAR(100)
- annotation_date : DATE

pathway
- pathway_id : INT(10) «PK»
- pathway_accession : VARCHAR(50)
- pathway_name : VARCHAR(255)
- pathway_description : TEXT
- pathway_category : VARCHAR(100)
- database_source : VARCHAR(50)

gene_expression
- expression_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- tissue_id : INT(10) «FK»
- expression_level : DECIMAL(10,4)
- expression_unit : VARCHAR(50)
- developmental_stage : VARCHAR(100)
- experimental_condition : VARCHAR(255)
- data_source : VARCHAR(100)
- pubmed_id : INT
- created_at : TIMESTAMP

tissue
- tissue_id : INT(10) «PK»
- tissue_name : VARCHAR(255)
- tissue_description : TEXT
- tissue_category : VARCHAR(100)
- parent_tissue_id : INT(10) «FK»

gene_phenotype_annotation
- annotation_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- phenotype_id : INT(10) «FK»
- annotation_source : VARCHAR(100)
- evidence_code : VARCHAR(10)
- pubmed_id : INT
- annotation_date : DATE

phenotype
- phenotype_id : INT(10) «PK»
- phenotype_accession : VARCHAR(50)
- phenotype_name : VARCHAR(255)
- phenotype_description : TEXT
- phenotype_category : VARCHAR(100)

ortholog
- ortholog_id : INT(10) «PK»
- gene_a_id : INT(10) «FK»
- gene_b_id : INT(10) «FK»
- ortholog_type : ENUM
- orthology_confidence : DECIMAL(5,4)
- dn_ds_ratio : DECIMAL(8,6)
- percentage_identity : DECIMAL(5,2)
- alignment_coverage : DECIMAL(5,2)

gene_publication
- association_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- publication_id : INT(10) «FK»
- association_type : VARCHAR(100)

publication
- publication_id : INT(10) «PK»
- pubmed_id : INT
- doi : VARCHAR(100)
- title : TEXT
- abstract : TEXT
- journal_name : VARCHAR(255)
- publication_date : DATE
- volume : VARCHAR(50)
- issue : VARCHAR(50)
- pages : VARCHAR(50)
- first_author : VARCHAR(255)
- last_author : VARCHAR(255)
- citation_count : INT

species
- species_id : INT(10) «PK»
- taxon_id : INT(10) «FK»
- species_name : VARCHAR(255)
- binomial_name : VARCHAR(255)
- common_name : VARCHAR(255)
- web_name : VARCHAR(255)
- is_... : TINYINT(1)

transcript
- transcript_id : INT(10) «PK»
- gene_id : INT(10) «FK»
- transcript_stable_id : VARCHAR(100)
- transcript_name : VARCHAR(100)
- transcript_biotype : VARCHAR(100)
- start_position : BIGINT
- end_position : BIGINT
- strand : TINYINT
- transcript_length : INT
- is_canonical : TINYINT(1)
- created_at : TIMESTAMP

genome_assembly
- assembly_id : INT(10) «PK»
- species_id : INT(10) «FK»
- assembly_name : VARCHAR(255)
- assembly_accession : VARCHAR(100)
- assembly_level : ENUM
- genome_size : BIGINT
- chromosome_count : INT
- scaffold_count : INT
- contig_count : INT
- n50_value : INT
- gc_content : DECIMAL(5,2)
- assembly_date : DATE
- submitter : VARCHAR(255)
- is_reference : TINYINT(1)
- is_representative : TINYINT(1)
- created_at : TIMESTAMP
- updated_at : TIMESTAMP

ncbi_taxa_node
- taxon_id : INT(10) «PK»
- parent_id : INT(10)
- rank : VARCHAR(32)
- genbank_hidden_flag : TINYINT(1)
- left_index : INT(10)
- right_index : INT(10)
- root_id : INT(10)

protein
- protein_id : INT(10) «PK»
- transcript_id : INT(10) «FK»
- protein_stable_id : VARCHAR(100)
- protein_sequence : LONGTEXT
- protein_length : INT
- molecular_weight : DECIMAL(10,2)
- isoelectric_point : DECIMAL(4,2)
- created_at : TIMESTAMP

chromosome
- chromosome_id : INT(10) «PK»
- assembly_id : INT(10) «FK»
- chromosome_name : VARCHAR(100)
- chromosome_type : ENUM
- sequence_length : BIGINT
- sequence_accession : VARCHAR(100)
- is_circular : TINYINT(1)

ncbi_taxa_name
- taxon_id : INT(10) «FK»
- name : VARCHAR(255) «PK»
- name_class : VARCHAR(50) «PK»

protein_interaction
- interaction_id : INT(10) «PK»
- protein_a_id : INT(10) «FK»
- protein_b_id : INT(10) «FK»
- interaction_type : VARCHAR(100)
- confidence_score : DECIMAL(5,4)
- detection_method : VARCHAR(100)
- interaction_source : VARCHAR(100)
- pubmed_id : INT

protein_domain
- domain_id : INT(10) «PK»
- domain_accession : VARCHAR(50)
- domain_name : VARCHAR(255)
- domain_description : TEXT
- domain_type : VARCHAR(100)
- database_source : VARCHAR(50)

protein_domain_annotation
- annotation_id : INT(10) «PK»
- protein_id : INT(10) «FK»
- domain_id : INT(10) «FK»
- start_position : INT
- end_position : INT
- e_value : DECIMAL(10,3)
- score : DECIMAL(8,2)