SN	Table names	Columns/Attributes	Indexes/Unique	Description
1.	cell_type_enrichment_level1	• level1_id (PK) BIGINT • cell_type_name VARCHAR(128) NOT NULL • lineage VARCHAR (8) NOT NULL • enrichment_score DECIMAL(6,4)	• PK: level1_id • UNIQUE: (cell_type_name)	
2.	Cell_type_enrischment_level2	• level2_id (PK) BIGINT • level1_id (FK) BIGINT NOT NULL • cell_type_name VARCHAR(128) NOT NULL • enrichment_score DECIMAL(6,4)	PK: level2_id UNIQUE: (level1_id, cell_type_name)	
2.	disease	• disease_id (PK) BIGINT • disease_name VARCHAR (128) NULL • category VARCHAR (32) NULL (e.g. “autoimmune”, “CH”, etc)	• PK: disease_id • UNIQUE: (disease_name)	Catalogue of autoimmune / autoinflammatory diseases and related diagnostic labels. Provides a controlled list of disease names and broad categories (e.g. “autoimmune”, “autoinflammatory”) used to annotate patients, studies, and variant–disease associations.
3.	genes	• gene_id (PK) BIGINT • ref_genome_id, FK, NOT NULL • gene_symbol VARCHAR (32) NOT NULL	• PK: gene_id • UNIQUE: (gene_symbol, ref_genome_id) – same symbol can exist on	Registry of genes (one row per gene symbol) that variants and annotations reference. Ensures standardized gene naming across studies and annotation sources (HGNC-style symbols like STAT3, DNMT3A, TET2).

			different genome builds • INDEX: idx_genes_ref_genome(ref_genome_id)	
4.	sequencing_experiment	• seq_exp_id (PK) BIGINT • sample_id FK NOT NULL • study_id (FK) BIGINT • technology ENUM('GWAS','Tapestri','GT_seq','WES','WGS','Other') NOT NULL • coverage DECIMAL	• PK: seq_exp_id • INDEX: idx_seqexp_sample(sample_id) • INDEX: idx_seqexp_study(study_id) • INDEX: idx_seqexp_technology (technology)	Each sequencing run / library prepared from a sample under a given study (paper/cohort). This is the anchor for downstream per-experiment cell types, variant calls, clonality and summaries.
5.	sample_variant_call	• call_id (PK) BIGINT • cell_type_id (FK) BIGINT NOT NULL • seq_exp_id (FK) BIGINT NOT NULL • level1_id BIGINT • level2_id BIGINT • variant_id (FK) BIGINT NOT NULL • vaf DECIMAL(4,3) NULL • variant_origin VARCHAR(32) NULL	• PK: call_id • UNIQUE: (seq_exp_id, variant_id, cell_type_id) – one call per variant/celltype/experiment • INDEX: idx_call_cell_type(cell_type_id) • INDEX: idx_call_variant(variant_id)	Stores individual variant calls observed in a specific sequencing experiment, within a specific cell-type compartment. Each row links a variant to an experiment and cell type, with an associated VAF and call-level origin (somatic/germline). The unique constraint ensures you don't duplicate the same variant–cell-type call within the same experiment.

7.	study	• study_id BIGINT (PK) • study_name VARCHAR (128) name • pmid VARCHAR(16) NULL • year INT NULL • notes TEXT NULL	• PK: study_id • UNIQUE: (pmid) • INDEX: idx_study_year(year)	Projects / publications / cohorts under which experiments are organized. Stores a human-readable name, PubMed ID, year and any notes so that sequencing experiments, patients and downstream analyses can be grouped and reported at the study level.
9.	variant_annotation	• ann_id (PK) BIGINT NOT NULL • variant_id (FK) BIGINT • hgvsp VARCHAR (255) NOT NULL • consequence VARCHAR (32) (e.g., missense_variant, splice_donor_variant) NOT NULL	• PK: ann_id • INDEX: idx_varann_variant(variant_id)	Stores per-variant functional annotations: gene assignment, HGVS protein change and predicted consequence, plus optional mapping to a protein domain. The UNIQUE (variant_id) enforces a single "canonical" annotation row per variant (e.g. one main gene / consequence / protein HGVS per allele).
10.	variants	• variant_id [PK] BIGINT NOT NULL AUTO_INCREMENT • protein_domain_id FK BIGINT NOT NULL • gene_id FK BIGINT NOT NULL • ref_genome_id [FK] INT NOT NULL • chrom VARCHAR (8) NOT NULL • pos INT NOT NULL	• PK: variant_id • UNIQUE: (ref_genome_id, chrom, pos, ref, alt) – ensures no duplicate variants per genome build • INDEX: idx_variant_chrom_pos (chrom, pos)	Central hub of genomic variants: one row per allele defined by genome build, chrom, position, ref and alt, with links to a reference-genome table and a variant-type table (SNV, indel, etc.). All downstream objects (calls, annotations, disease associations, functional validation) reference this table.

		• ref VARCHAR (255) NOT NULL • alt VARCHAR (255) NOT NULL • variant_origin VARCHAR(32) NOT NULL • vaf DECIMAL(4,3)		
11.	protein_domain	• protein_domain_id PK BIGINT • domain_name VARCHAR(16) NULL • domain_type VARCHAR(25) NULL	• PK: protein_domain_id • UNIQUE: (domain_name)	Lookup table for protein domains affected by variants, such as SH2 domains or Pfam/InterPro regions. Enables annotation of which structural/functional region of the protein is altered by a given variant (via variant_annotation.protein_domain_id).
12.	sample	• sample_id PK BIGINT • disease_id FK BIGINT NOT NULL • patient_code VARCHAR(8) NULL • tissue_source VARCHAR(16) NOT NULL • sample_type VARCHAR(64) NULL	• PK: sample_id • INDEX: idx_sample_patient(patient_id)	Biological samples collected from patients (e.g. PBMC, bone marrow, intestinal biopsy). Each sample belongs to one patient and can be sequenced multiple times. sample_type differentiates bulk DNA, single-cell DNA, TCR sequencing, etc.

14.	Functional_validation	• validation_id PK BIGINT • variant_id FK BIGINT NOT NULL • assay_type VARCHAR(64) NULL • result VARCHAR(32) NULL	• PK: validation_id • INDEX: idx_fv_variant(variant_id)	Captures experimental evidence from functional assays for specific variants (e.g. reporter assays, phosphorylation readouts, CRISPR screens). Multiple assays can exist per variant, with the result summarised as gain-of-function, loss-of-function, neutral, etc.
15.	reference_genome	• ref_genome_id BIGINT PK, auto-inc, NOT NULL • ref_genome_name VARCHAR(32) NOT NULL • source VARCHAR(64) NULL • version VARCHAR(16) NULL • notes TEXT NULL	• PK: ref_genome_id • UNIQUE: (ref_genome_name)	