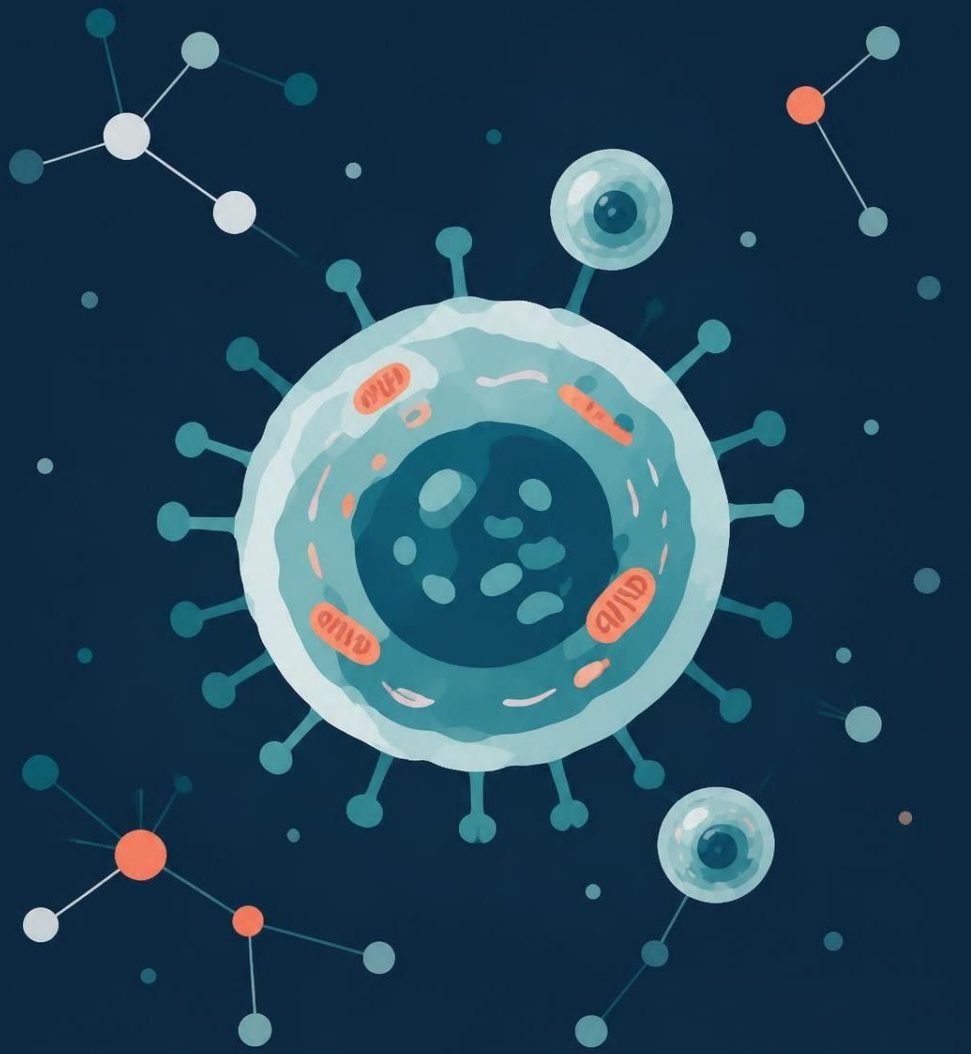
LONG COVID INTELLIGENCE

A Biomedical AI Research Platform

An executive research strategy from The COVID-19 Long-haul Foundation

John Murphy, President & CEO,
The COVID-19 Long-haul Foundation

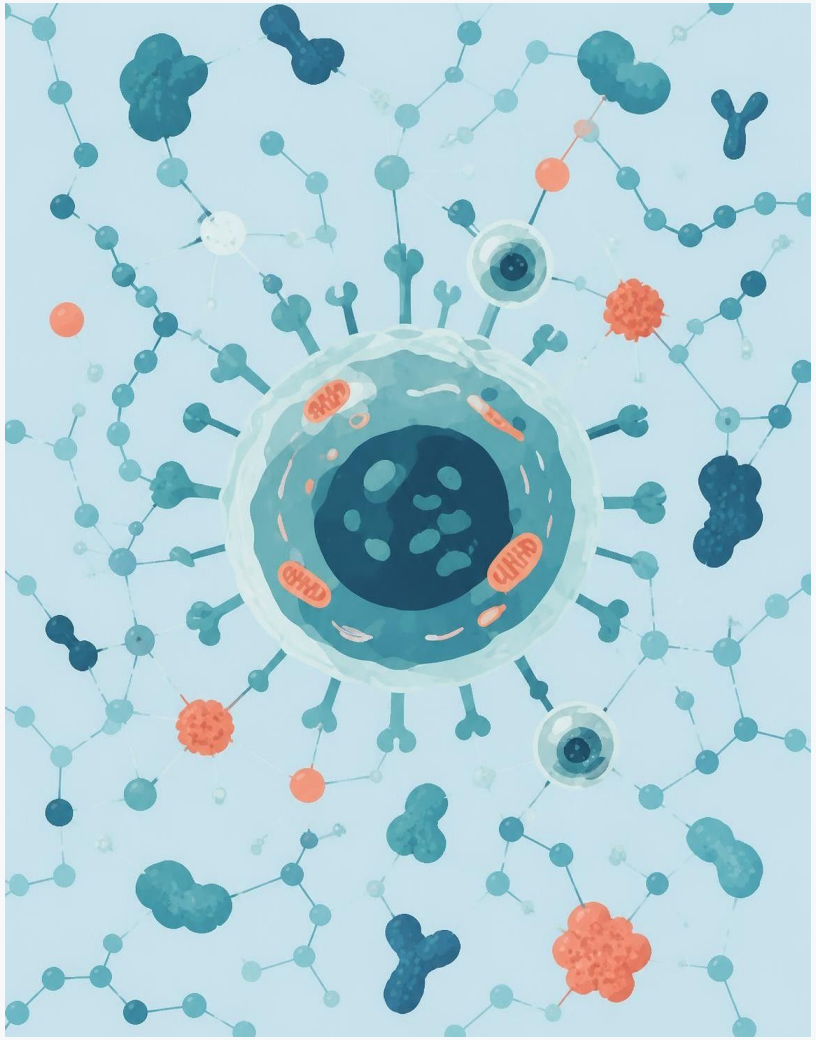


A research intelligence platform for the whole Long COVID continuum

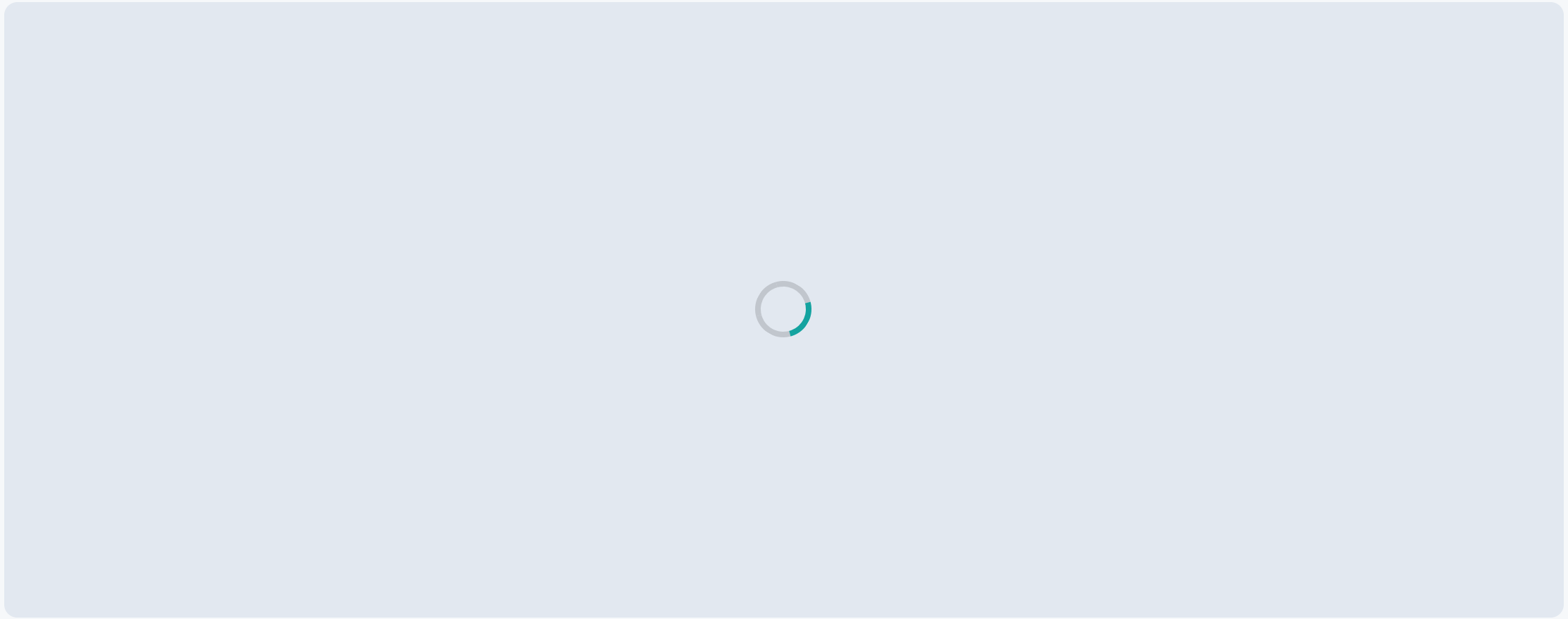
Why does SARS-CoV-2 lead to complete recovery in some people, prolonged multisystem illness in others, and persistent organ dysfunction in still others?

The proposed LCI Research Platform investigates genetic susceptibility before infection, acute exposure, molecular injury, clinical manifestations, progression, recovery and therapeutic response.

It combines a domain-adapted language model with continuously updated literature, longitudinal clinical databases, genomic and molecular datasets, statistical modeling and machine learning.

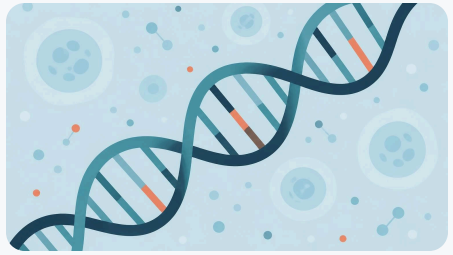


Nine interconnected biological domains



Domains 1–3: Susceptibility, baseline and acute infection

1

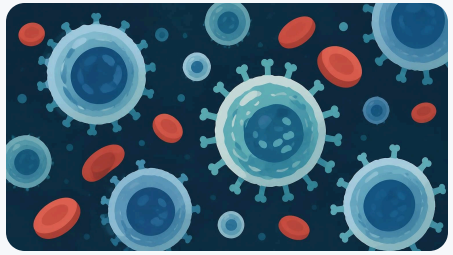


Pre-infection genetic susceptibility

Inherited variants tied to infection, severe disease, immune dysregulation and persistent complications.

GWAS, HLA variation, interferon

2

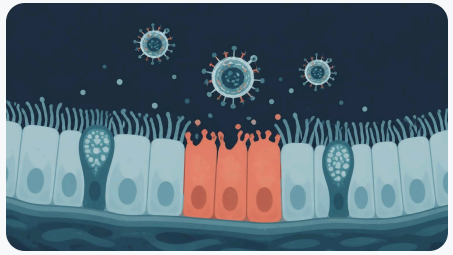


Baseline immune and metabolic status

Inflammatory profile, insulin resistance, obesity, cardiovascular risk, renal function.

Cytokines, T- and B-cell

3



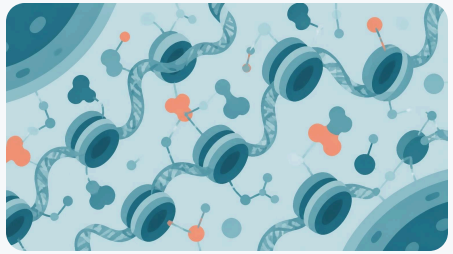
Acute infection and viral exposure

Severity, viral load, replication duration, reinfection and vaccination history.

Research evidence: sequencing records, persistence, antigen

Domains 4–6: Molecular, mitochondrial and vascular injury

4



Epigenetic and transcriptional change

DNA methylation, chromatin accessibility, histone modifications, RNA sequencing, microRNAs and single-cell transcriptomics.

5

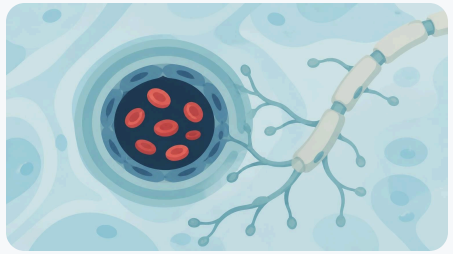


Mitochondrial dysfunction

Impaired energy production, oxidative stress and altered dynamics.

Oxidative phosphorylation, mitochondrial DNA, ATP,

6



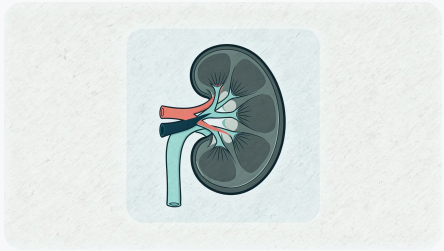
Endothelial and autonomic injury

Vascular dysfunction, dysautonomia, POTS, orthostatic hypotension and impaired cerebral perfusion.

Endothelial biomarkers, vascular

Domains 7–9: Organ complications, trajectory and treatment

7

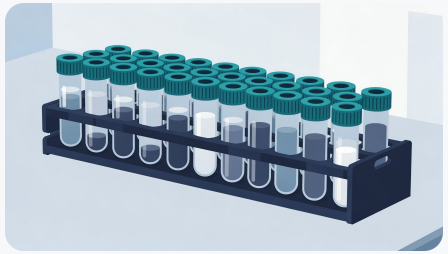


Renal and neurological complications

Acute kidney injury, chronic kidney disease, neuroinflammation, neuropathy and cognitive impairment.

eGFR trajectories, albuminuria,

8



Longitudinal recovery and progression

How abnormalities evolve over months and years.

Repeated labs, symptom trajectories, functional outcomes and organ recovery.

9

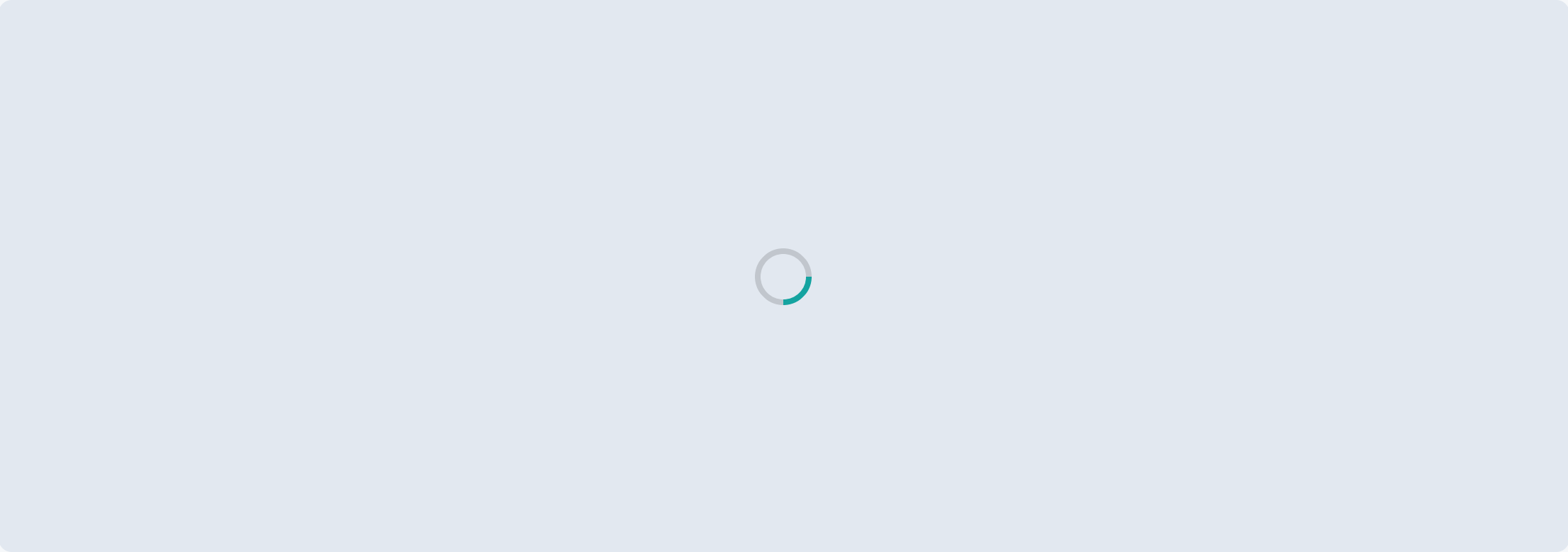


Therapeutic response

Which treatments help which biological subgroups, which fail, and which biomarkers predict response?

Comparative effectiveness, trials, safety,

A hybrid architecture in six components



LCI adapts an existing language model. Training a new one from scratch would be extraordinarily expensive and duplicate existing language capabilities.



Components A and B: The interface and the evidence library

A · Language model adapts an existing capable model through biomedical instructions, evaluation and possible fine-tuning. It reads publications, explains mechanisms, compares hypotheses, drafts protocols, synthesizes across domains and produces reports.

3,000+

Resources already in the Foundation's library

B · Literature repository

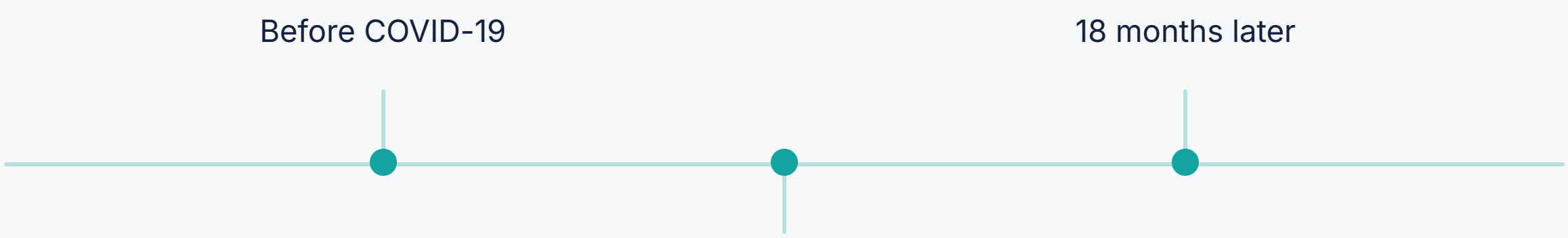
PubMed/MEDLINE, PubMed Central, ClinicalTrials.gov, NIH RePORTER, WHO, CDC, FDA, GEO, SRA, ENCODE, GWAS Catalog, dbGaP.

Index each publication by domain, phenotype, population, design, sample size, biomarkers, genes and pathways, outcomes, interventions and strength of evidence.

Component C: A longitudinal database built on time

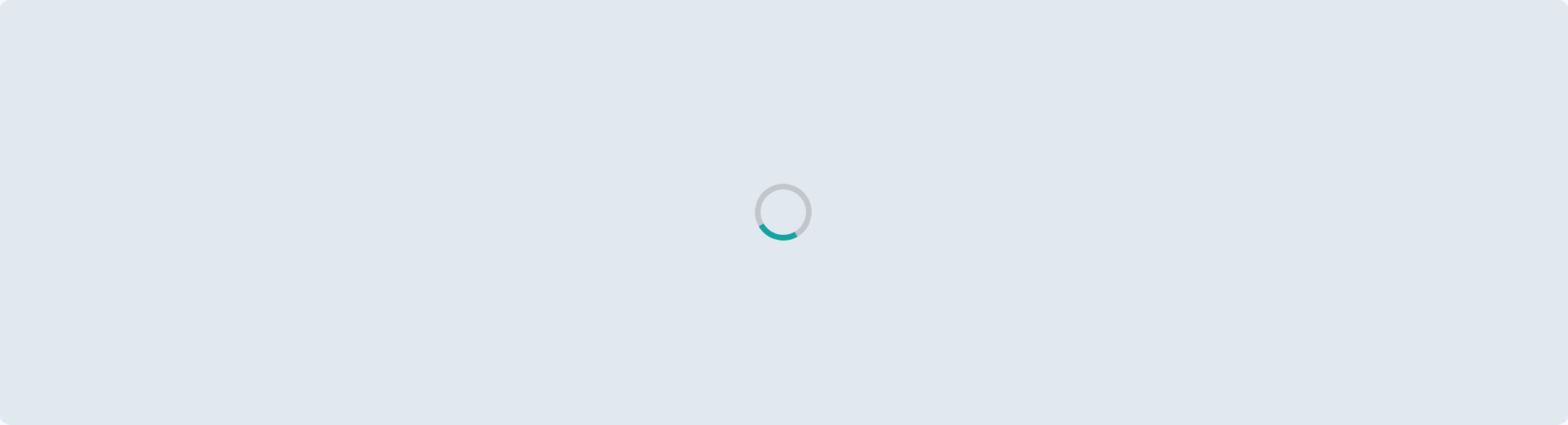
One common patient-level research identifier links nine databases that mirror the nine domains.

LCI-01 · Genomics	**LCI-02** · Baseline immune and metabolic	**LCI-03** · Acute infection
LCI-04 · Epigenomics and transcriptomics	**LCI-05** · Mitochondrial and metabolic	**LCI-06** · Endothelial and autonomic
LCI-07 · Renal and neurological	**LCI-08** · Longitudinal trajectories	**LCI-09** · Therapeutic exposure and outcomes



Component D: A knowledge graph that connects, not just stores

ILLUSTRATIVE HYPOTHESIS MAP · NOT AN ESTABLISHED UNIVERSAL CAUSAL PATHWAY



Links genes to proteins, pathways, cellular abnormalities, organ dysfunction, clinical manifestations and treatment responses.

Each relationship carries an evidence record: experimentally established, observational, hypothesized or contradicted.

Components E and F: Computation and scientific quality control

COMPONENT E

Statistical and ML engine

Dedicated models for GWAS, differential