

Biomedical Knowledge / Context Graphs Foundation Model

What if Biology were software ?

Andrea R. Beccari, PhD
VP Exscalate Drug Discovery Platform
Dompe Farmaceutici SpA



Dompe

**“Any sufficiently
evolved technology
becomes software”**

Nokia didn't build the iPhone
Hollywood didn't create YouTube > Netflix
Record labels didn't launch iTunes > Spotify
Publishers didn't invent Amazon > Kindle
Yellow Cab didn't build Uber
Uber do not evolve to Waymo
Banks didn't create Bitcoin > Tether

...

**Disruption never comes from evolution.
it comes from revolution**

Is medicine reaching an inflection point ?

Can Became a Software ?

Where we are ?

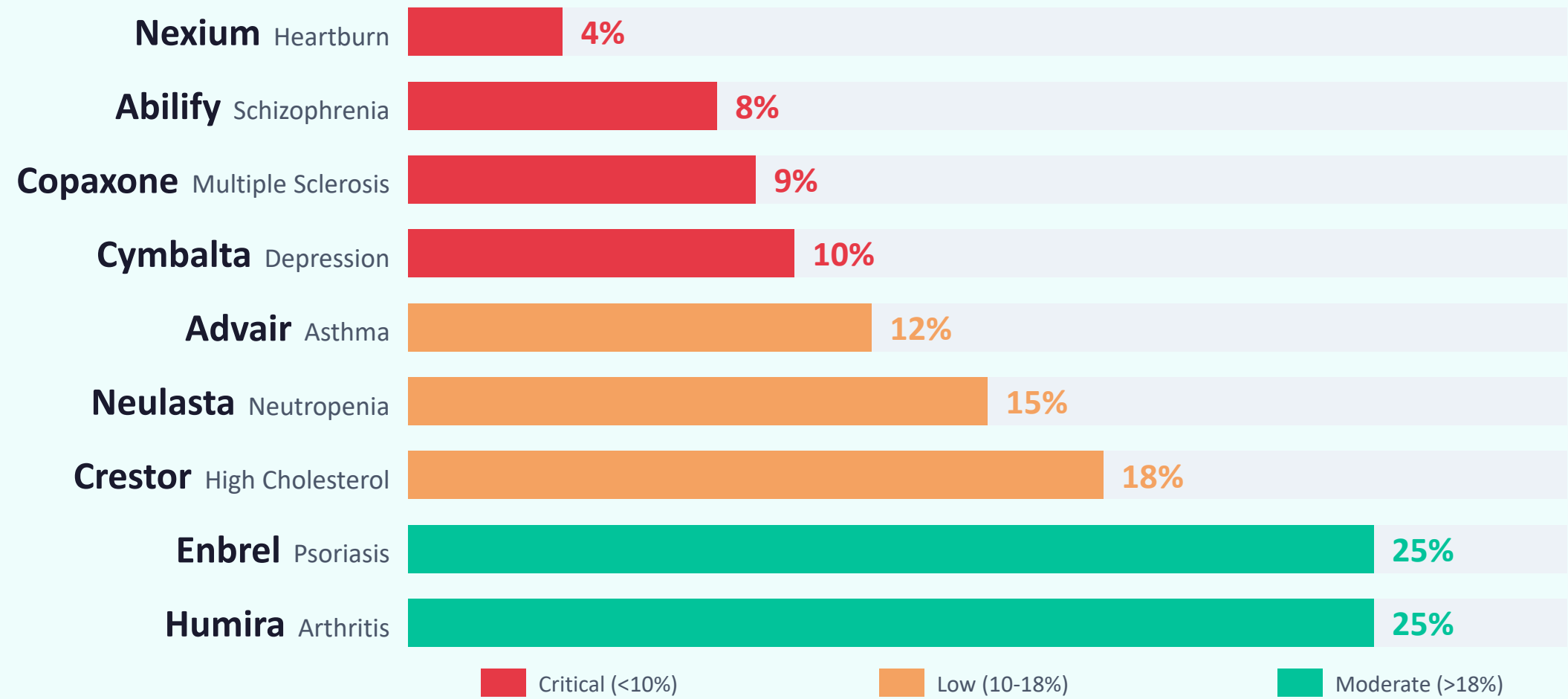
> 90% Failure Rate

> 3B\$ Cost per drug

> 12 years of R&D

The Best-Selling Drugs Work for Shockingly Few

True responder rates of top-selling drugs range from just 4% to 25%

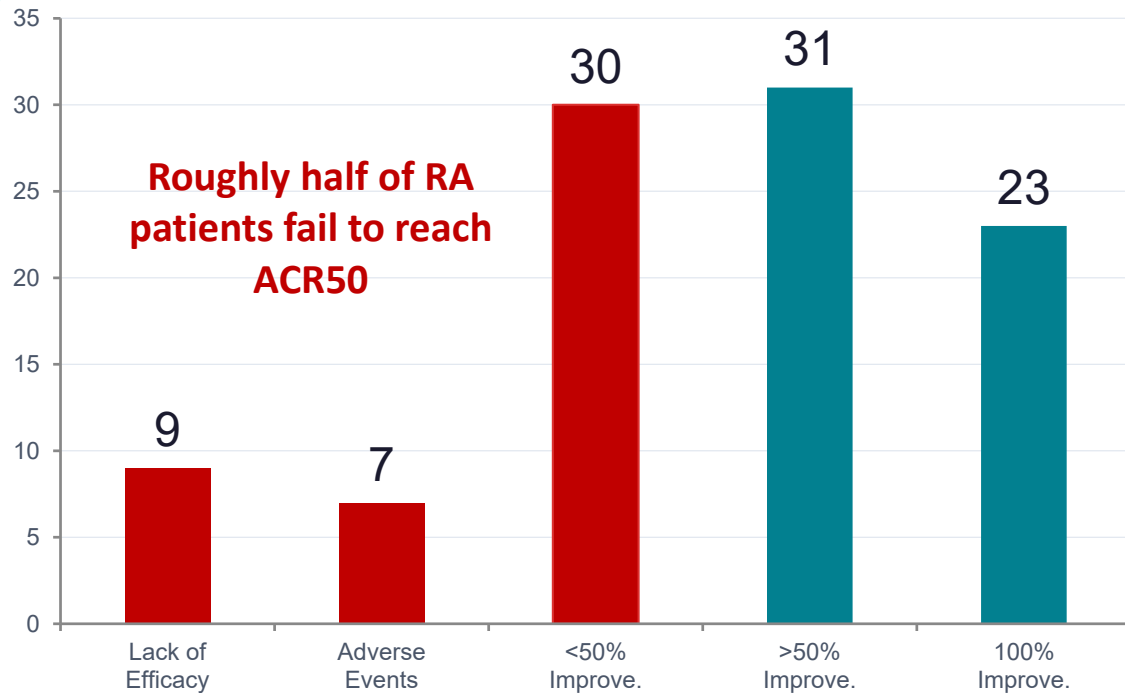


Source: Nature — Personalized Medicine: Time for One-Person Trials (2015)

Does a Single Rheumatoid Arthritis Even Exist?

The root cause: phenotypic diagnosis ignores molecular diversity

Etanercept (Enbrel) Patient Outcomes



% of patients in each outcome category



Phenotypic Blind Spot

Physicians diagnose on observable symptoms, irrespective of molecular mechanisms



Multiple Molecular Causes

A single disease label can mask fundamentally different underlying pathologies



One Target, One Drug: Insufficient

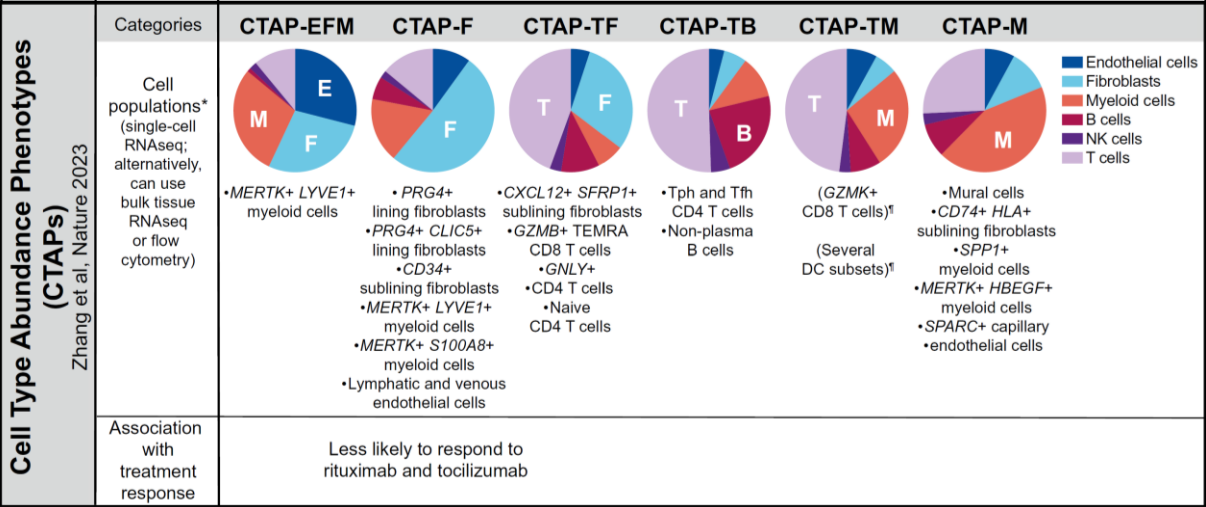
Disease complexity demands multi-target approaches, not single-pathway thinking

Effective treatment requires deep molecular knowledge — not just symptom matching

Actually 7 Rheumatoid Arthritis... so far

Phenotypic diagnosis can be implemented today: **Deconstruction of rheumatoid arthritis synovium defines inflammatory subtypes.** Nature. 2023 [PMID: 37938773]

Six cell-type abundance phenotypes (CTAPs) in a cohort of patients who were already inadequate responders to Embrel demonstrate the diversity of synovial inflammation in rheumatoid arthritis, ranging from samples enriched for T and B cells to those largely lacking lymphocytes.



Disease Phenotyping

RNA-sequencing and molecular profiling to characterize clinical response to anti-TNF



7 Molecular Causes

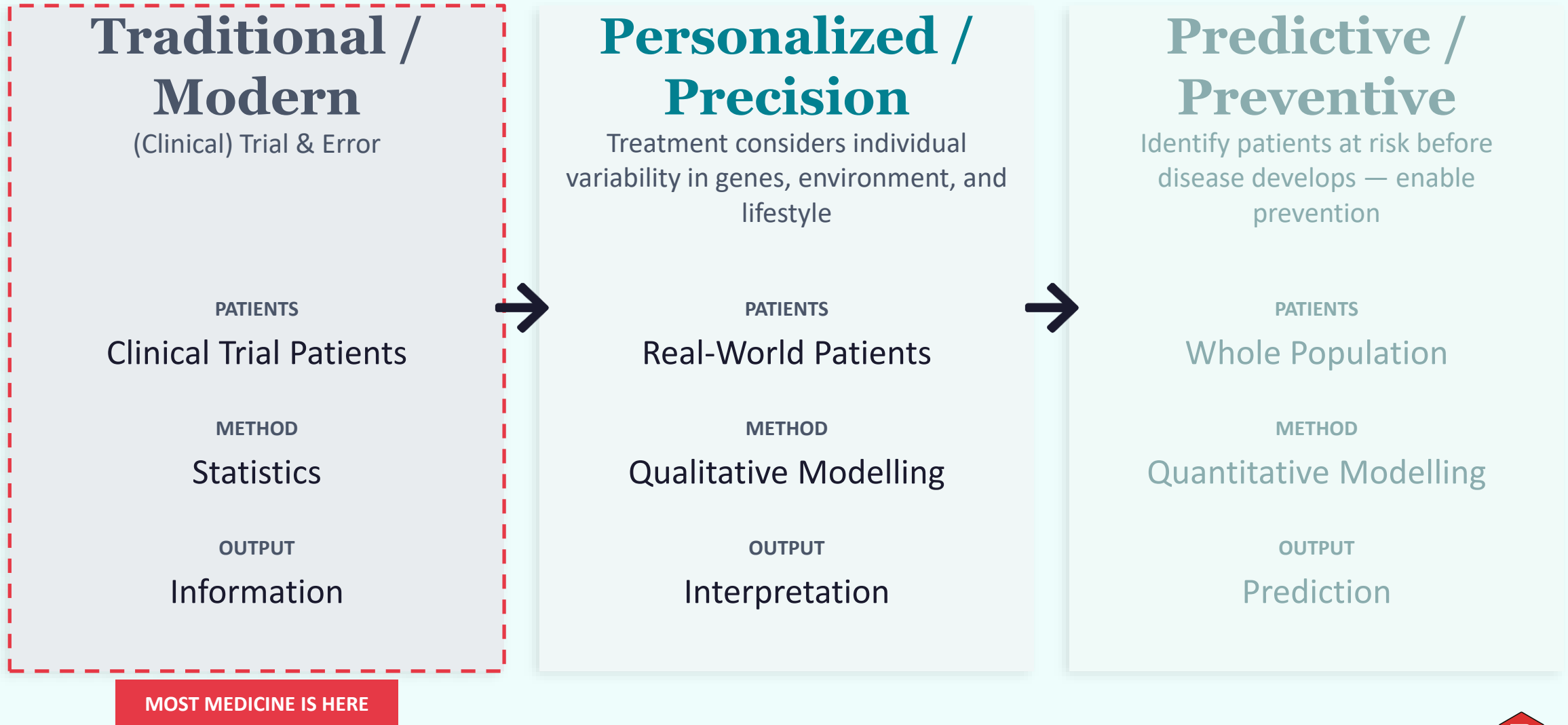
A single disease label can mask fundamentally different underlying pathologies



Predictive medicine

biomarkers differentiating good responders and non-responders in rheumatoid arthritis (RA).

Evolution of Medicine



Biology is too complex to be modelled !

1 Drug > 1 Target > 1 Disease

Biology at its most fundamental level is an information processing system, and there may be a common underlying structure between biology and information science - an isomorphic mapping between the two.



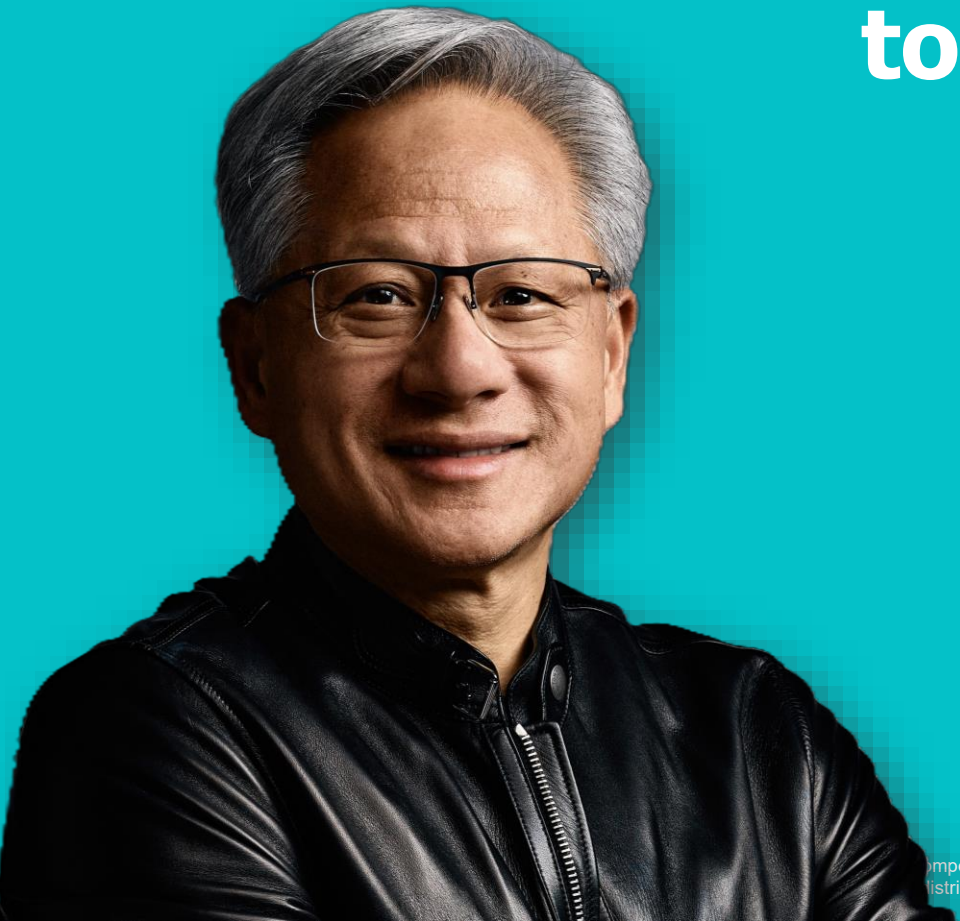
Demis Hassabis

Isomorphic Labs

Google DeepMind spin-off

ential.

“For the very first time in human history, biology has the opportunity to be engineering not science ”



Jensen Huang
founded NVIDIA

Dompe and strictly confidential.
distributed to any third party



"I would estimate we might know 10 to 15% of human biology, so the machine is not going to be good at all until we get way above 50%."

David Ricks
Chair, CEO Eli Lilly and Company

Confidential and strictly confidential.
Not to be distributed to any third party



“Combining our volumes of data and scientific knowledge with NVIDIA’s computational power and model-building expertise could reinvent drug discovery as we know it.”

“LillyPod is a continuous AI-driven discovery infrastructure tightly coupling “agentic wet labs” with computational dry labs for 24/7 experimentation and learning loops “



David Ricks
Chair, CEO Eli Lilly and Company

Confidential and strictly confidential.
Not to be distributed to any third party



If we would model biology what we need ?

Large-Scale,
High-Quality
Experimental
& Synthetic
Data

Hyperscale
HPC &
Quantum
Computing
Infrastructure

L(L,V,Q)Ms

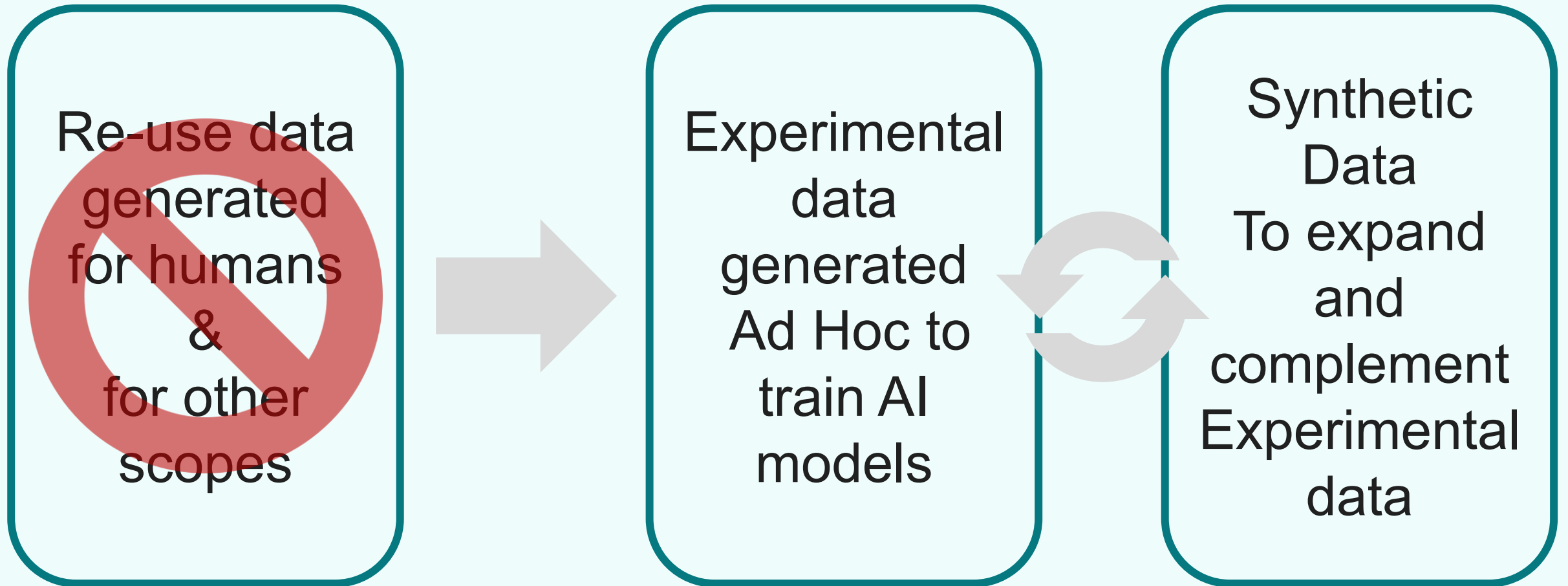
EBRM

Formal
Systems
Validators

Black-box models earn black-box data.

**From human-interpretable hypotheses
to industrial-scale systems biology,
profiled population-wide.**

Data Design for Model Generation



Black box data for black box models

Atlases tell us WHY a perturbation alters a cell. Biobanks tell us WHICH perturbations matter in real patients. Together they form the multi-scale substrate for every virtual-cell claim.

01 · MECHANISTIC LAYER

Cellular Perturbation Atlases

- ILLUMINA BILLION CELL ATLAS**
1B → 5B cells in 3y, ~20 PB/yr, 250 disease lines
- CZI VIRTUAL CELLS**
Petabyte scRNA-seq + spatial + imaging
- ARC INSTITUTE VCC**
Virtual Cell Challenge benchmark
- X-ATLAS / PISCES**
25.6 M perturbed cells across 16 contexts
- TAHOE-100M**
100 M cells × 1,100 compounds × 50 lines

02 · TRANSLATIONAL LAYER

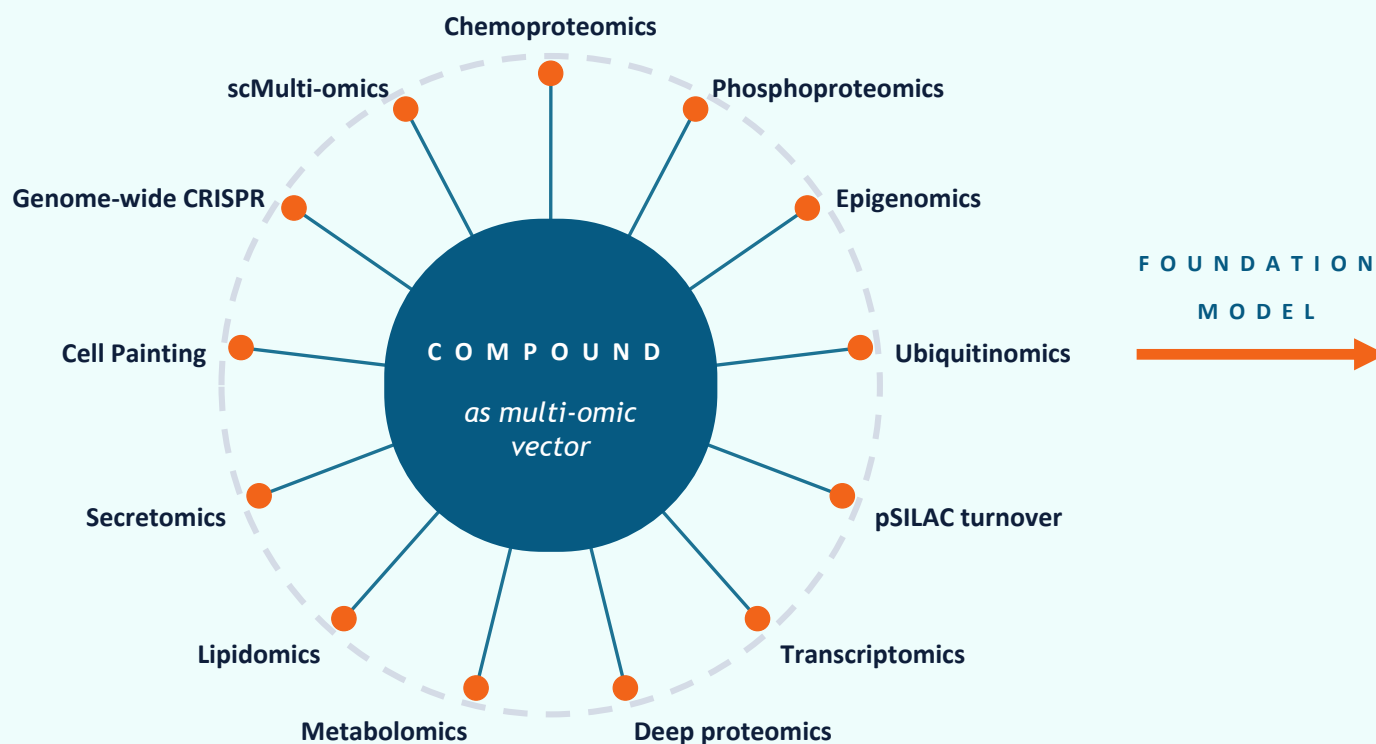
Population Biobanks

- UK BIOBANK**
500K WGS + Olink ~3,000 + imaging + EHR/GP
- NIH ALL OF US**
1 M target, 414K WGS, 1.2 B variants
- VA MILLION VETERAN**
900K enrolled, 2,068 GWAS traits
- TRUVETA GENOME**
10 M exomes on 120 M-patient EHR base
- LONGITUDINAL EHR**
Real-world phenotype at population scale

Two layers, one substrate: from molecule to population

From Magic Bullet to Pleiotropic Vector

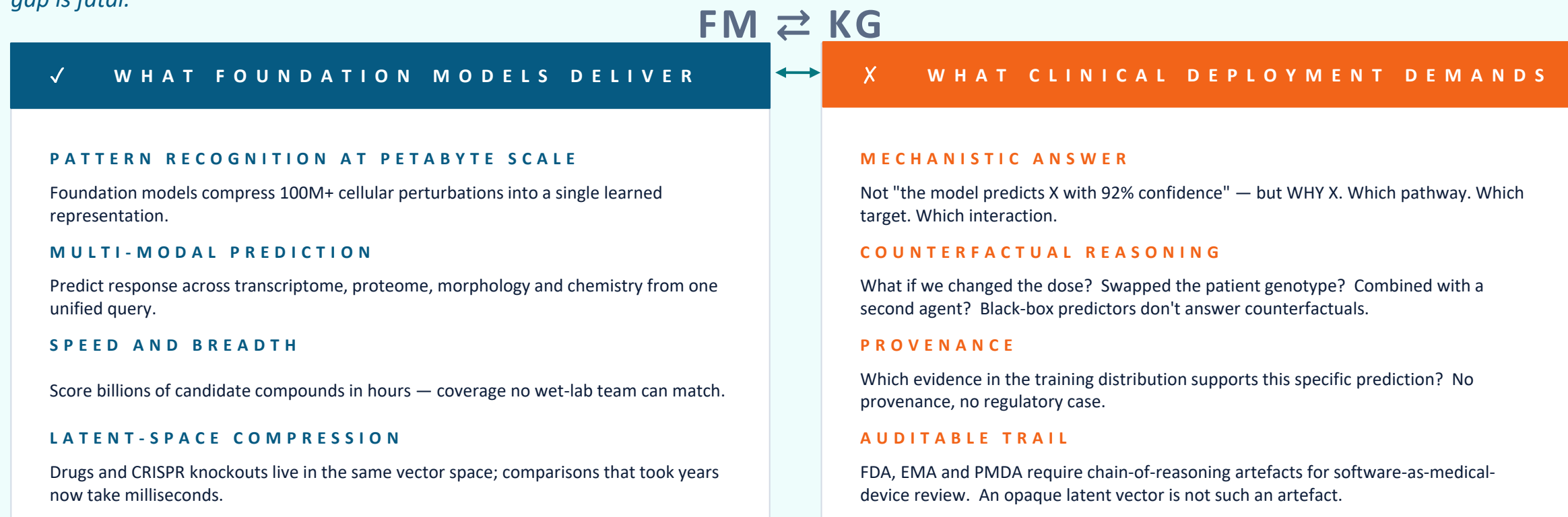
Stop describing a compound by its primary target and its side activities. Describe it by its coordinates in a n-dimensional readout space and let the foundation model do the rest.



- 01 Mechanism of Action**
Pathways perturbed, binding partners, downstream remodelling.
- 02 Polypharmacology**
Off-target activity, secondary engagement, full target portrait.
- 03 Efficacy**
Predicted response across cell lines, tissues, stratified cohorts.
- 04 Safety**
Predicted on-target liabilities and dose-response margins.
- 05 Toxicology**
Organ-system toxicity signatures inferred from profile geometry.

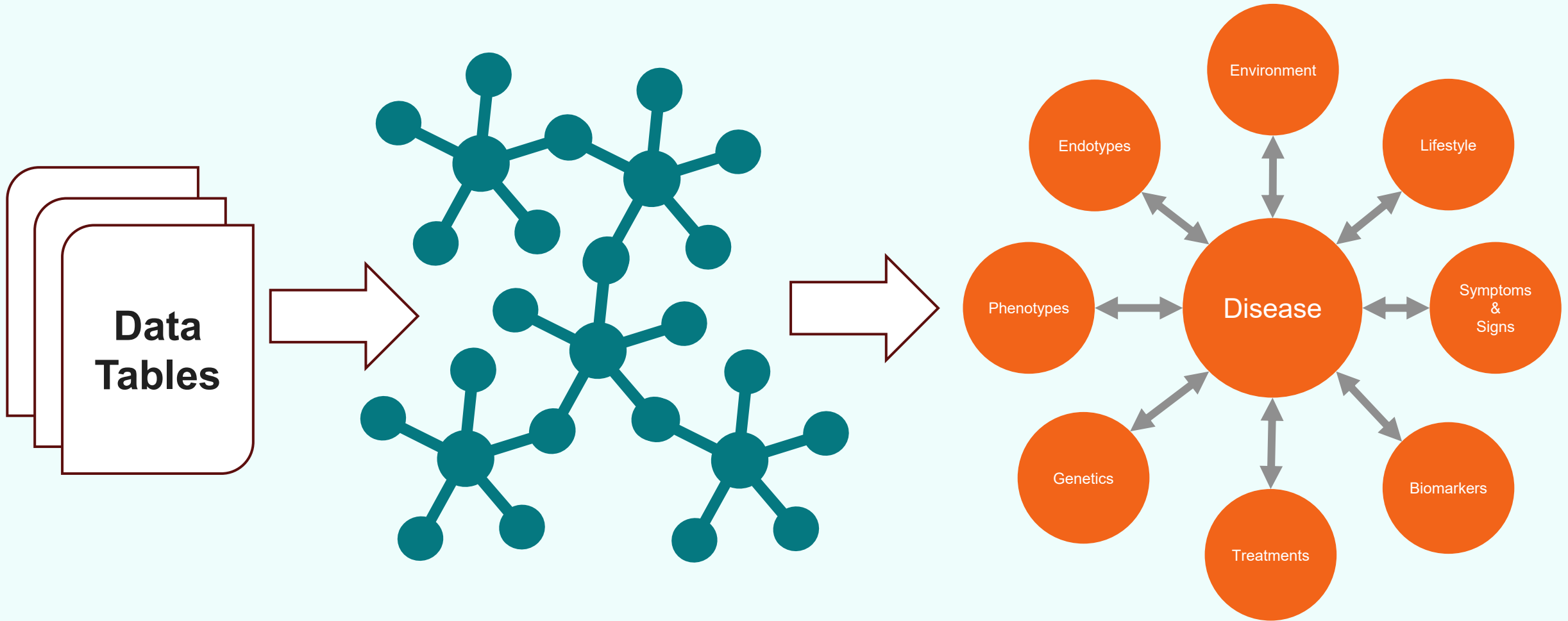
Black-Box Accuracy Without Trust ?

Foundation models reached the ceiling of prediction. They have not reached the floor of explanation — and for clinical, regulatory, and payer use, that gap is fatal.



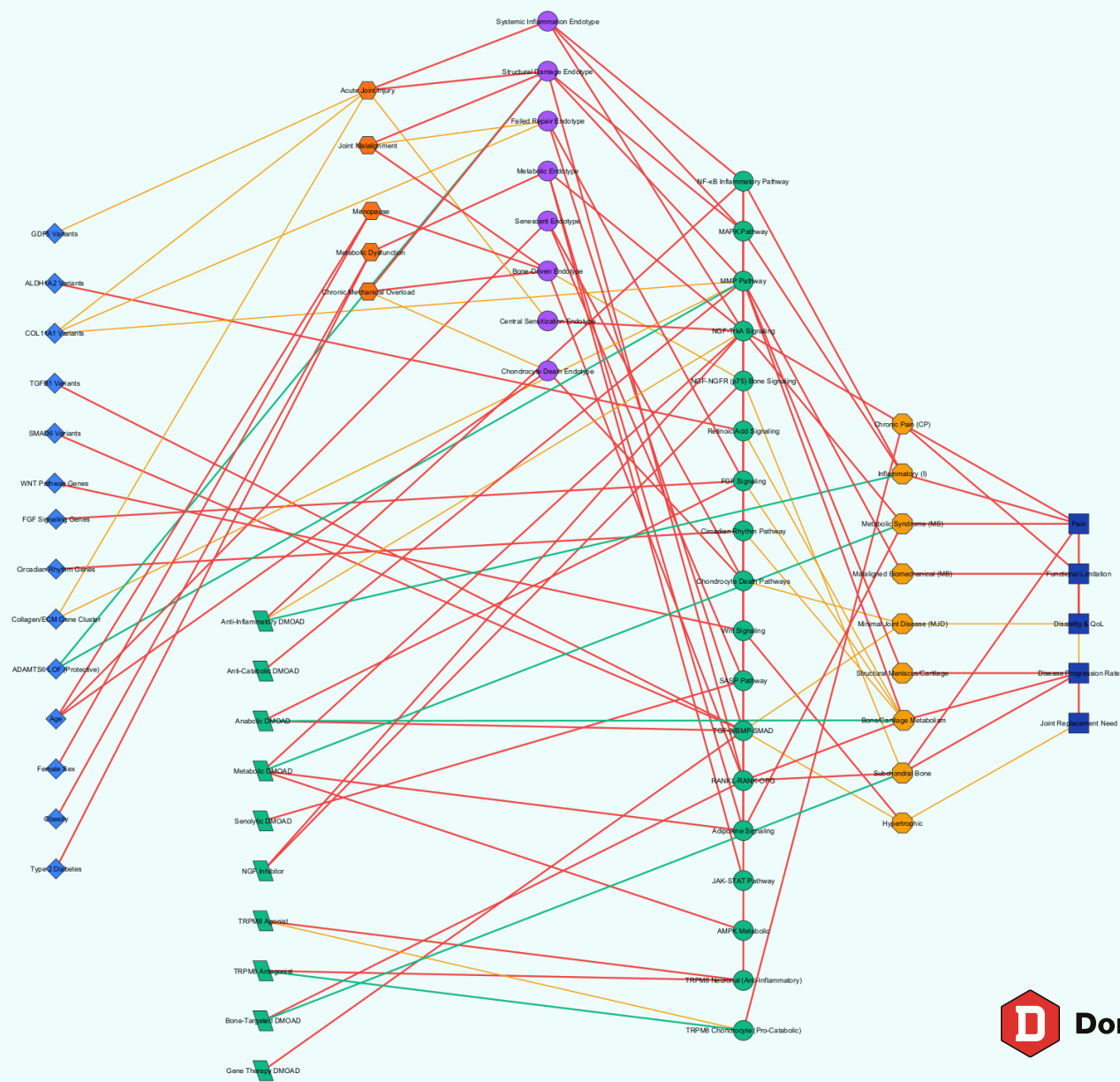
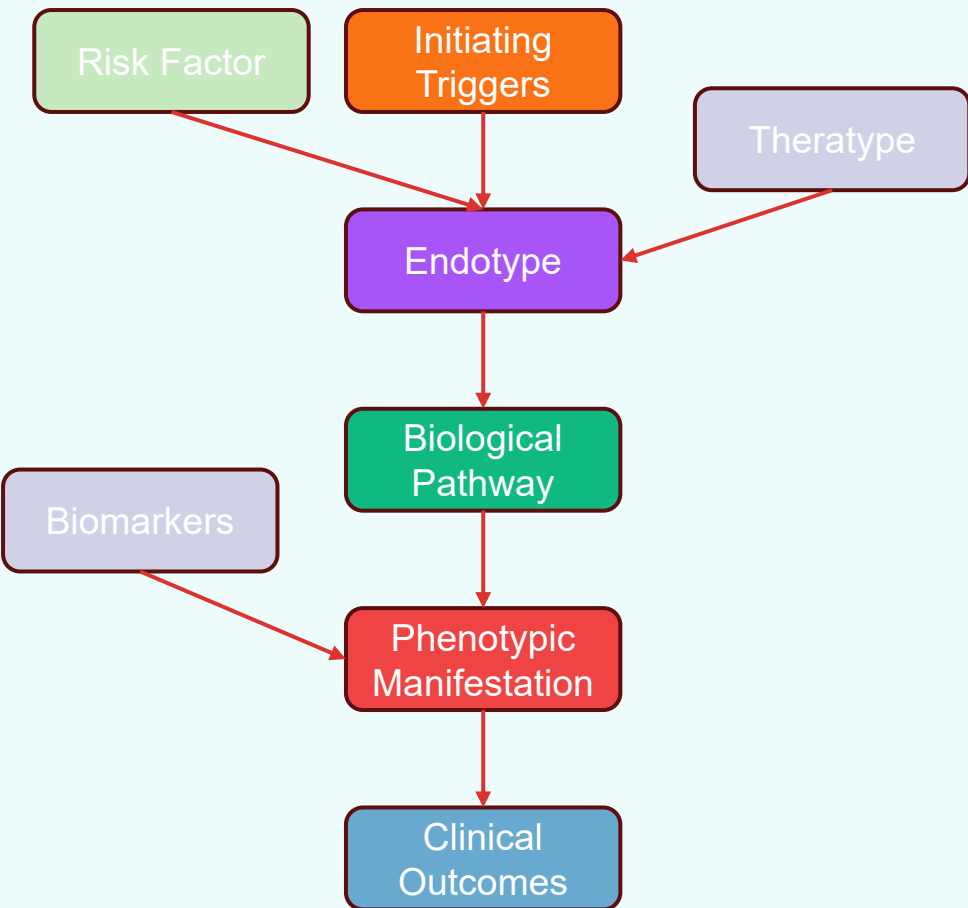
The plateau is no longer accuracy. It is interpretability. And interpretability is not a property a foundation model acquires by training longer — it is a property of the architecture surrounding the model.

From measurements, through the graph, into meaning. Numbers become a biomedical world model.



As of September 2024

Contextualized KGML is where we are going



Biology will be soon a software !

New generation of data,
next gen AI & Hyperscalers
are Rewriting Medicine

Andrea R. Beccari, PhD
VP Exscalate Drug Discovery Platform
Dompe Farmaceutici SpA

