MariaDB

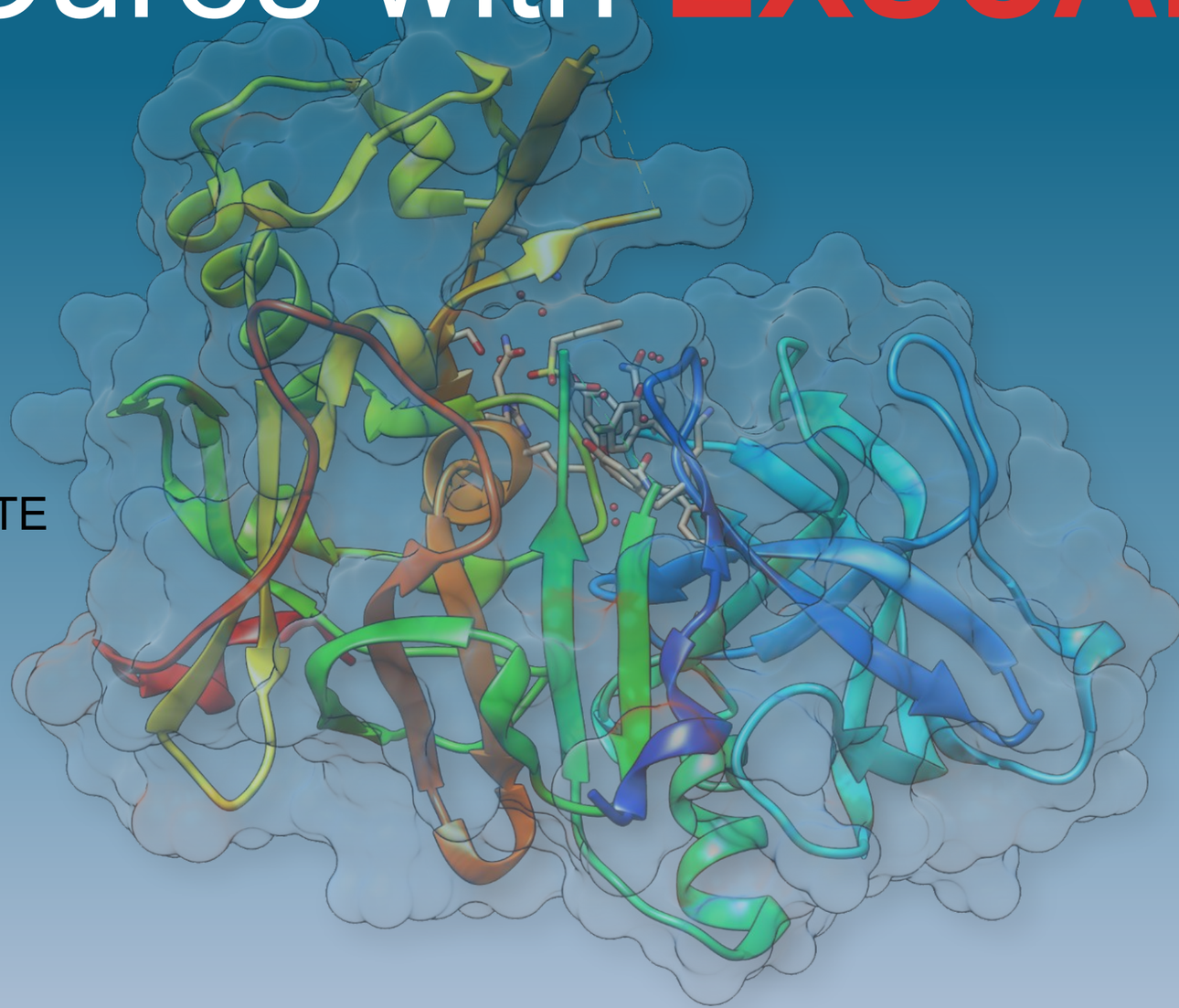
OPENWORKS

# EXSCALATE: Building a Faster, More Efficient & Less Costly Drug Discovery Platform with MariaDB

WILL BEGIN SHORTLY.

# Accelerating the Discovery of Future Cures with EXSCALATE

**Andrea R. Beccari, PhD**
Senior Director EXSCALATE
Dompe' Farmaceutici SpA

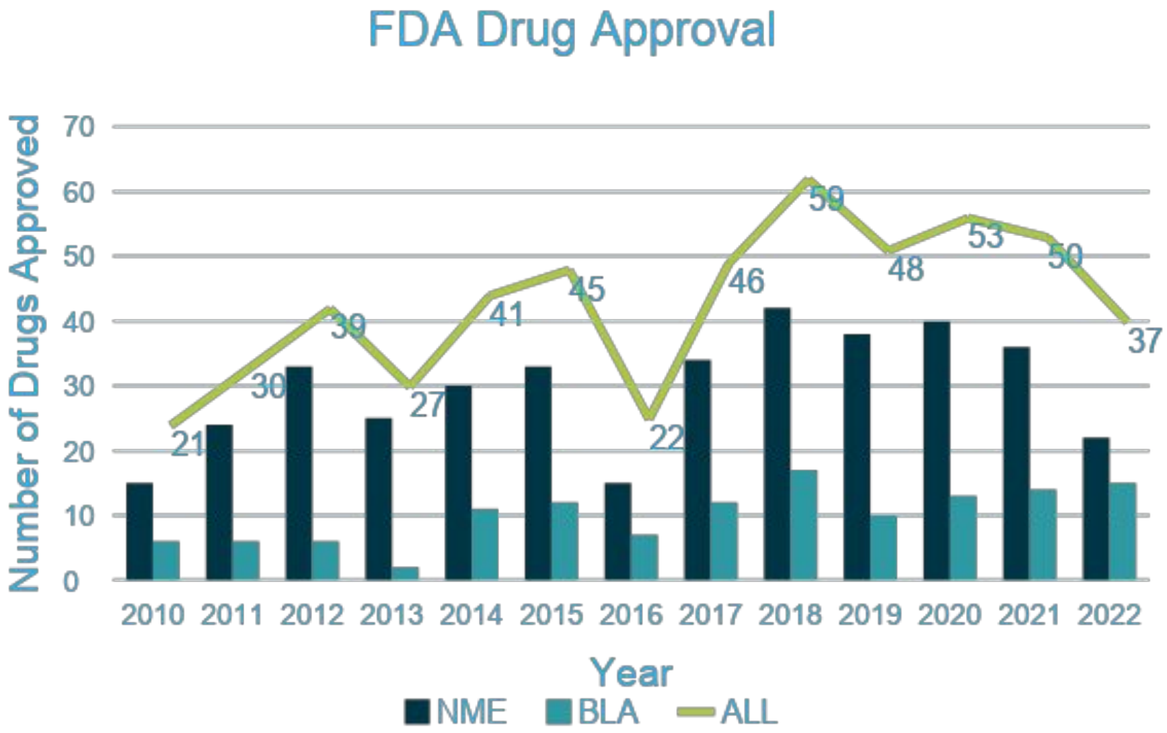


# Unmet Medical Need as a Driver for a Paradigm Shift in Pharmaceutical Sciences…

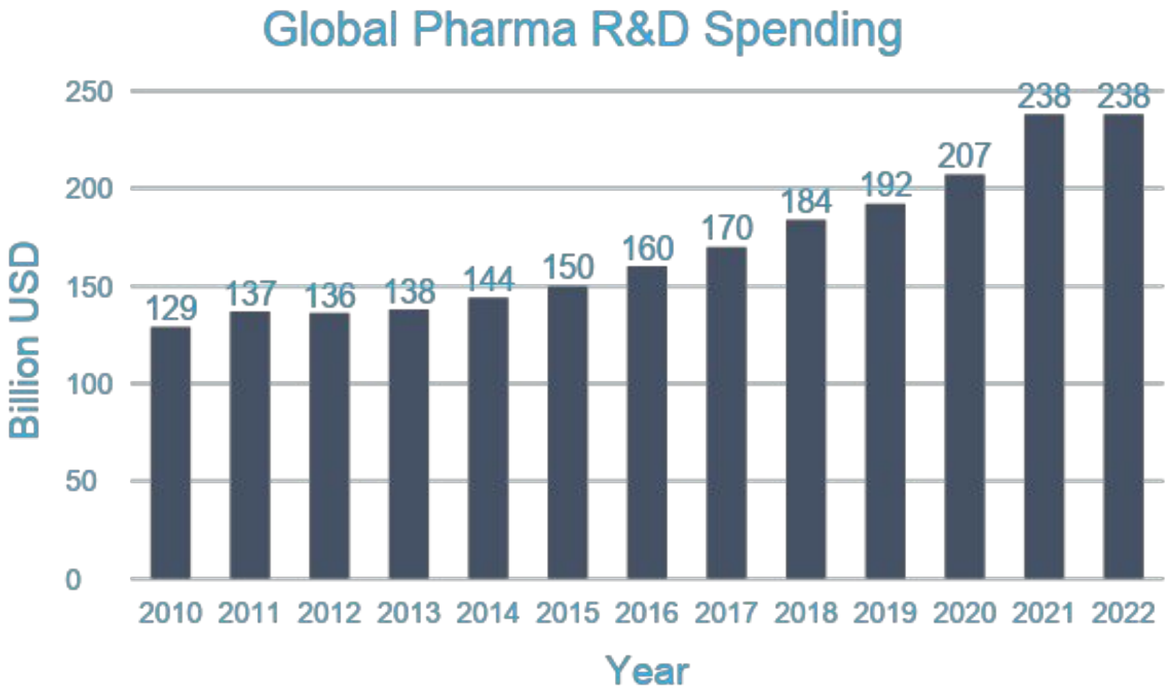
# Only 1/10 of all clinical drugs will arrive to the patient

## Mainly for lack of efficacy (>50%) or safety (> 10%)

### >10 years is the average time to patient




### >2 Billion USD average spending per approved drug



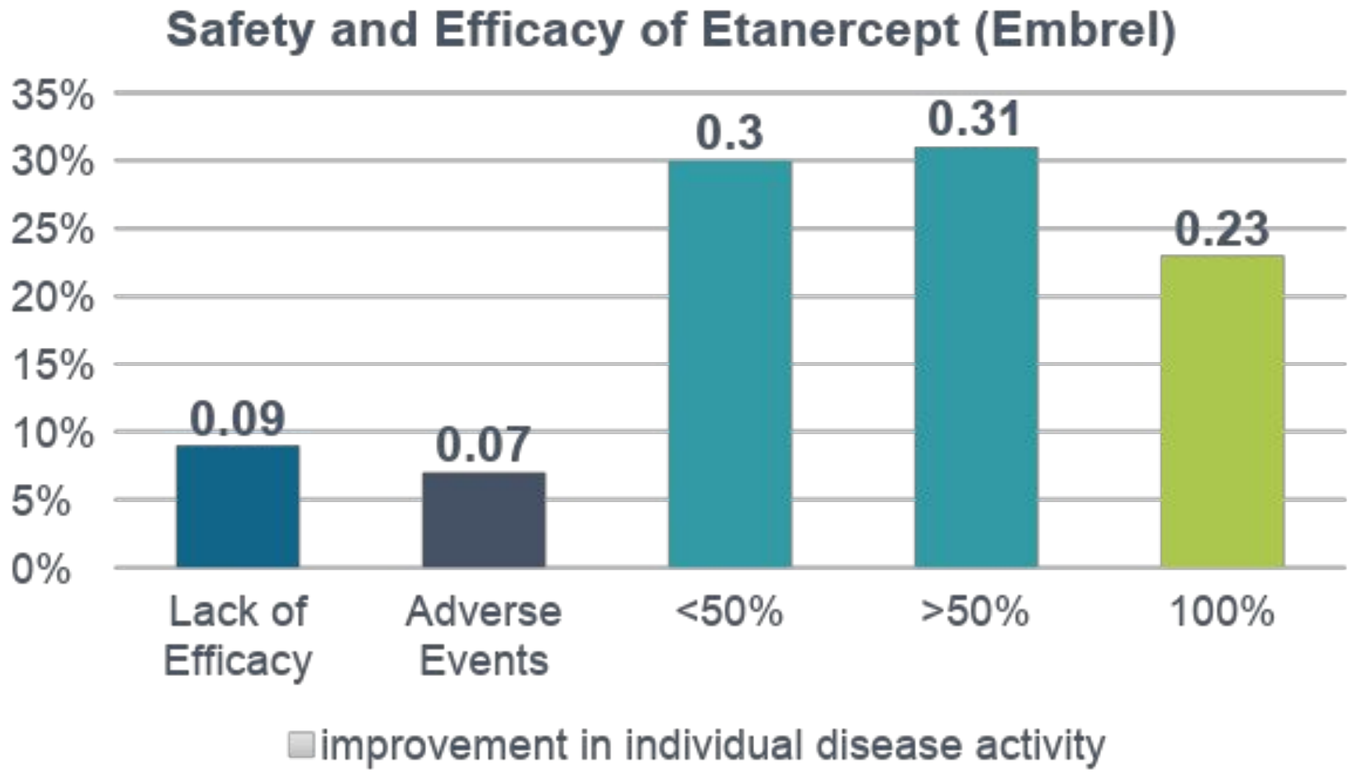

Public

Wouters OJ, McKee M, Luyten J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018.
JAMA. 2020;323(9):844–853. doi:10.1001/jama.2020.1166

# Lack of Knowledge about Disease complexity

## Does it exist a single rheumatoid arthritis ?

**Effective treatment of a disease requires a deep knowledge of the pathological condition and a quantitative understanding of the alterations of molecular mechanism involved.**




The Physician diagnoses a pathology on a phenotypic base irrespective of the molecular bases.

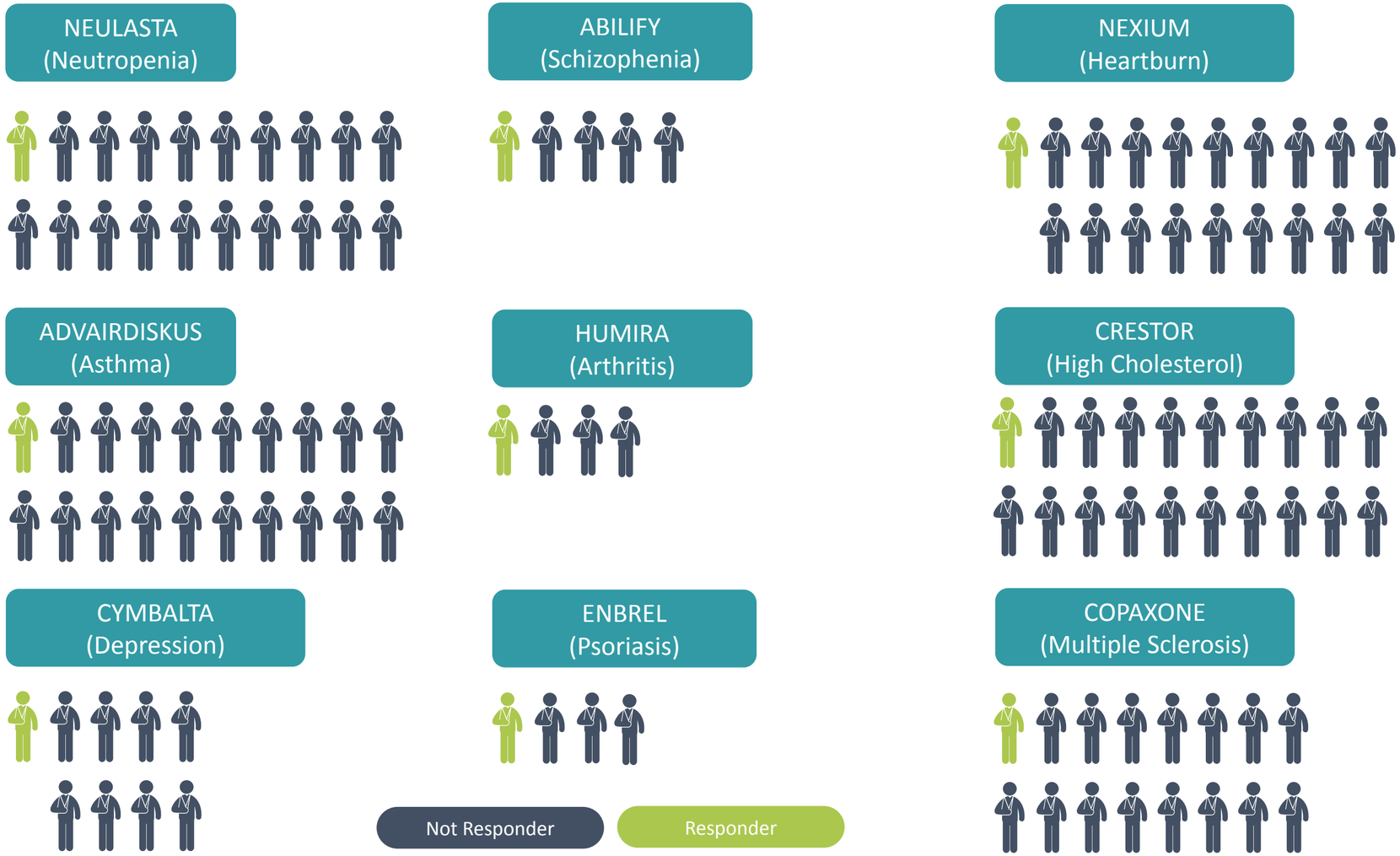
But the same disease may have multiple molecular causes

The model one target –one drug is insufficient to address the disease complexity

Public

# Which is the effect of imprecision medicine ?

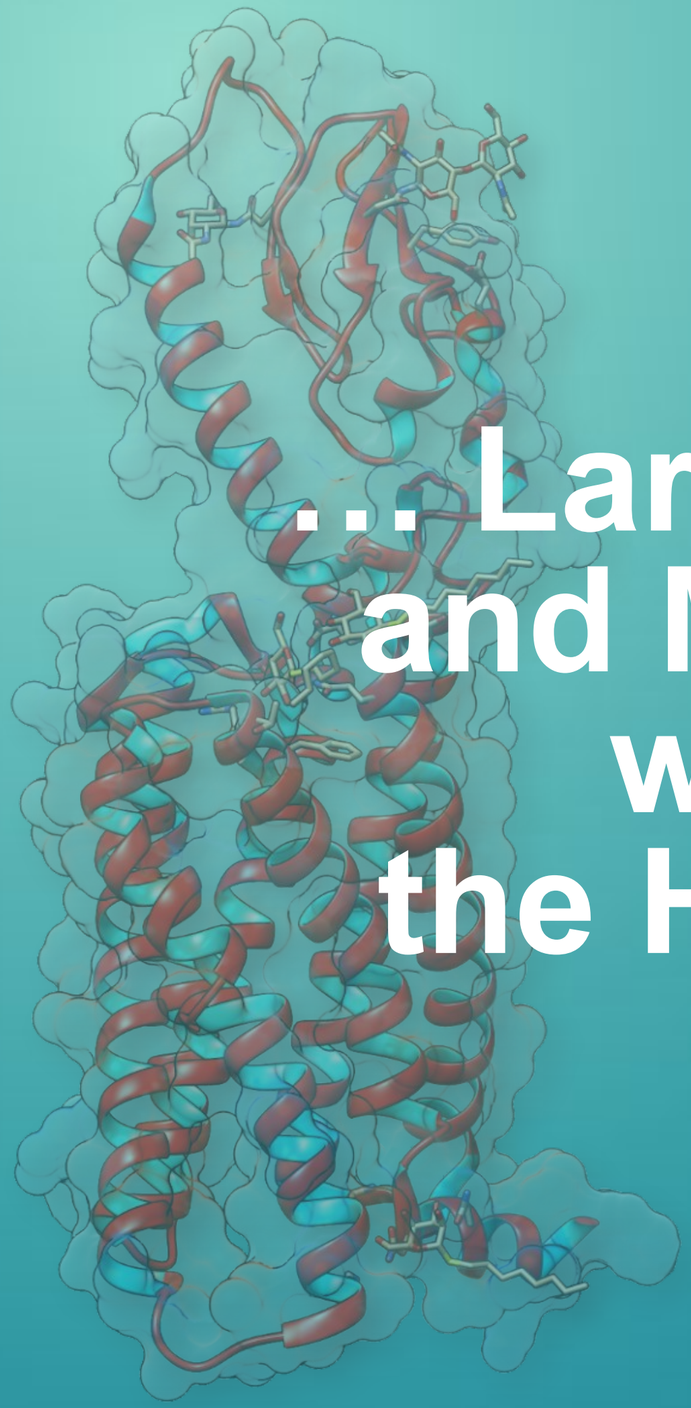
## Is not able to address the medical needs even if succeed in clinics




Some of the top selling drugs are fully active only on a fraction for the patients.
In this selection true responders' spans form the 25% of Humira and Embrel to the 4% of Nexium.

Public

https://www.nature.com/news/personalized-medicine-time-for-one-person-trials-1.17411



# … Large Scale Simulations and Machine Intelligence will Revolutionize the Healthcare Sector …

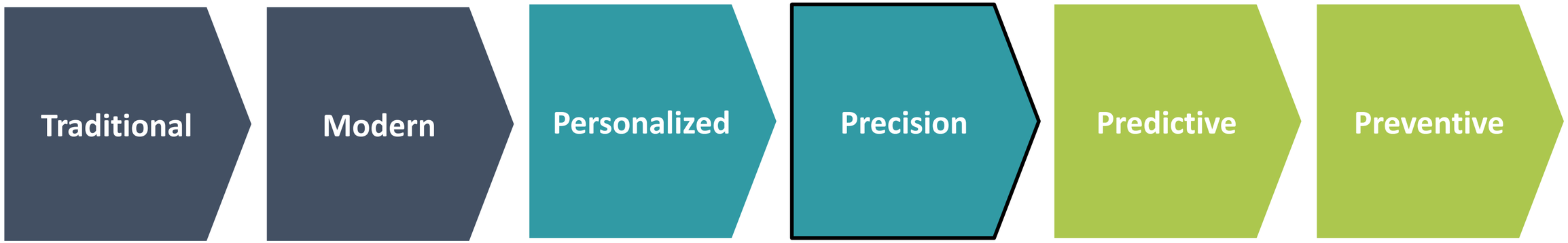
# Evolution of medicine

## Where we really Are ?

**Trial & Error**

**disease treatment and prevention that considers individual variability in genes, environment, and lifestyle for each person**

**Identify patients at risk of developing a disease, thereby enabling either prevention or early treatment of that disease**



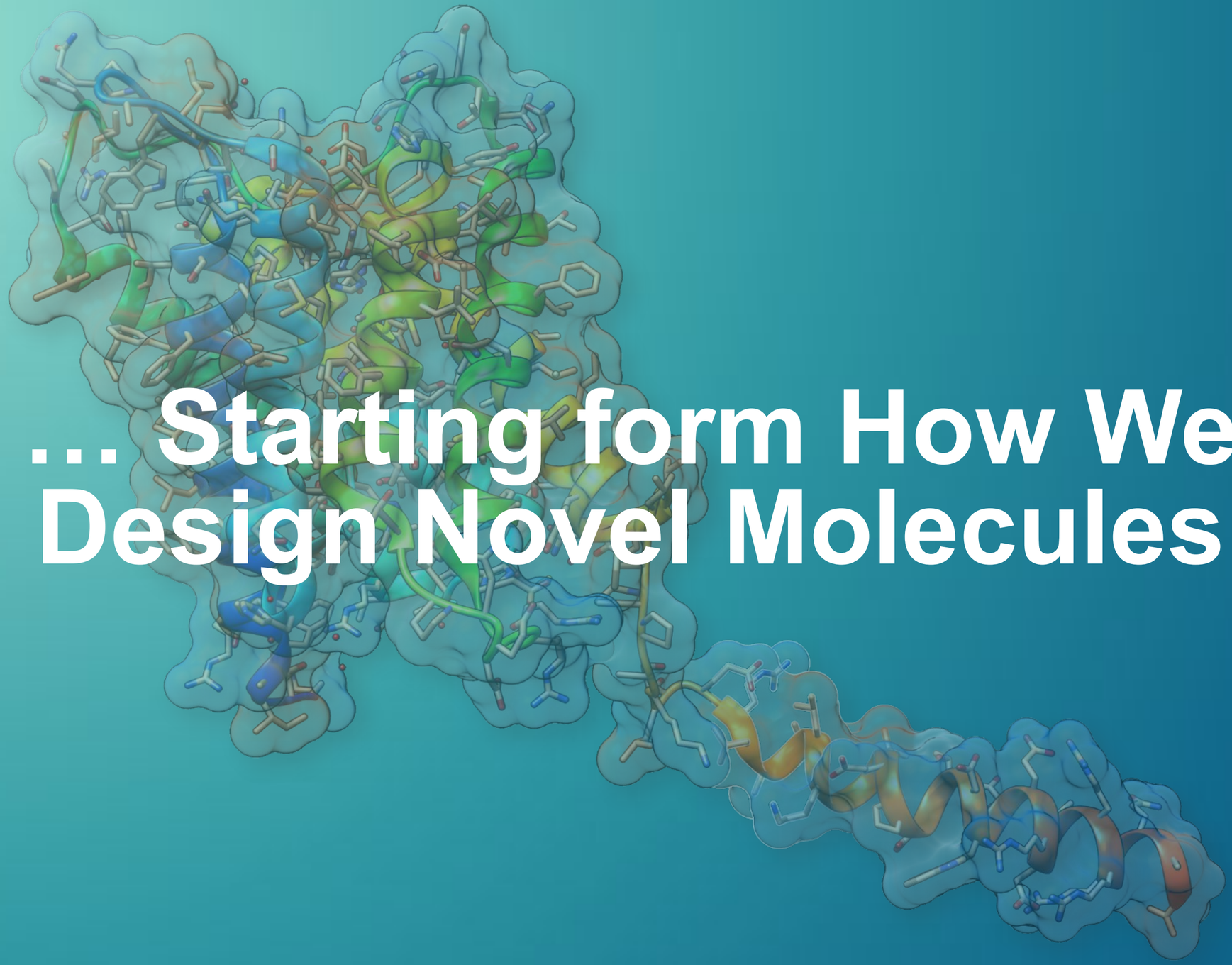

Public

# High Performance Computing in Pharma Research

## Biology is not too complex to be modelled

| HPC Simulations | AI/ML | HPC Simulation + AI/ML |
| --- | --- | --- |
| Predict Biological Activity<br><br>Analyzed large set of genomic data<br><br>Population wide Healthcare & wellness data<br><br>Virtual Twins | **Extract knowledge from**<br><br>Patient's Clinical Records<br>Literatures<br>Large scale experiments<br>Clinical trials<br>Wearable Devices<br>… | Use synthetic data to train AI to enhance the quality of:<br>ML models<br>Computer Simulations<br><br>Generate completely novel hypothesis |

Public

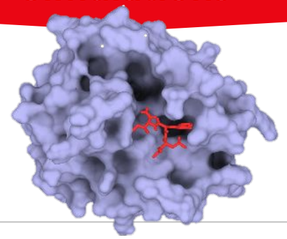
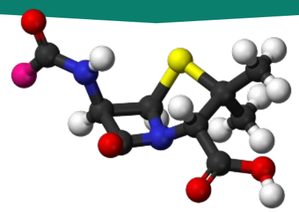


# … Starting form How We Design Novel Molecules

# High Performance Computing in Pharma Research

## Data Explosion in Drug Discovery

**In the last 5 years we faced an expansion of orders of magnitude in datasets,**
**Is required a redesign of the architectures for storage and data analytics**

| Biology | Simulations | Chemistry | Drug Profiling | Scientific Knowledge |
|---|---|---|---|---|
| |  |  | | |
| RCSB PDB PROTEIN DATA BANK | Schrödinger | PubChem | Corporate Data | PubMed |
| 0.2 Million | 10 Million | 0.1 Billion | ~1K | 30 Million |
| Google DeepMind | | Enamine | GOSTAR Global Online Structure Activity Relationship Database | Health Care Data |
| 200 Million | >1 Trillion | 6 Billion | 28 Million | Billions |

# High Performance Computing in Pharma Research

## Use virtual chemical space for real world applications

**Tangible Chemical Space is the most challenging in terms of data flow and manipulation. To search trillion of data in seconds we decide to build up a custom appliance for exploration of extremely large chemical spaces.**

| CTTD | LiGen | TCS | ProfhEX | KEX |
|---|---|---|---|---|
| | | Enamine<br>6 Billion<br>E4 COMPUTER ENGINEERING, MariaDB<br>>10 Trillion | | KnowledgEX<br>(under development) |

Public

# Tangible Chemical Space Appliance

## Key Features

Hardware

Software

**STORAGE CLUSTER**
**MINIO**: High Performance Object Storage [140TB NVMe]
GPUs Chemical Cartridge

**DATABASE QUERY CLUSTER**
**NVMesh**: disk aggregation [128TB NVMe]
1 query parallel on 216 cores

**HIGH SPEED CONNECTION**
2 InfiniBand Mellanox 200Gb/s & RDMA

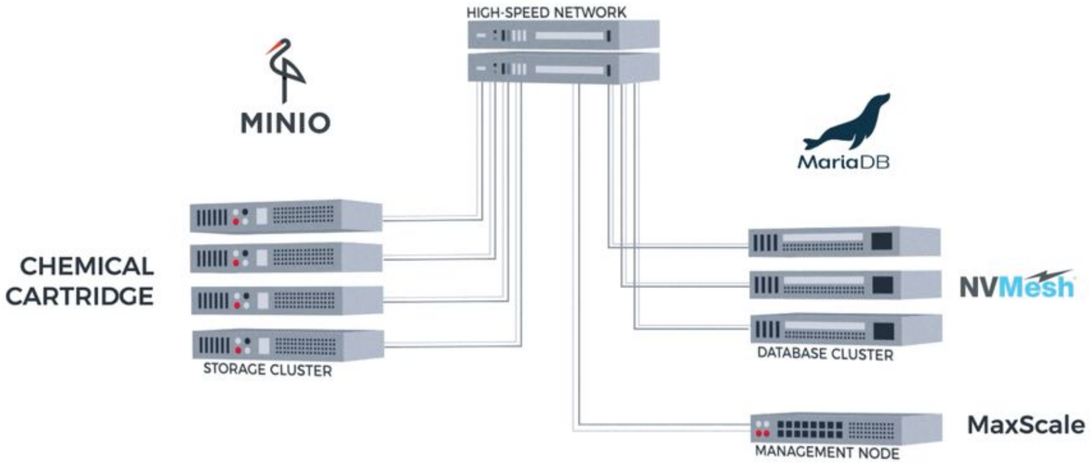
Multiple Storage Engines

Thread pooling

SQL compatibility

Virtual columns

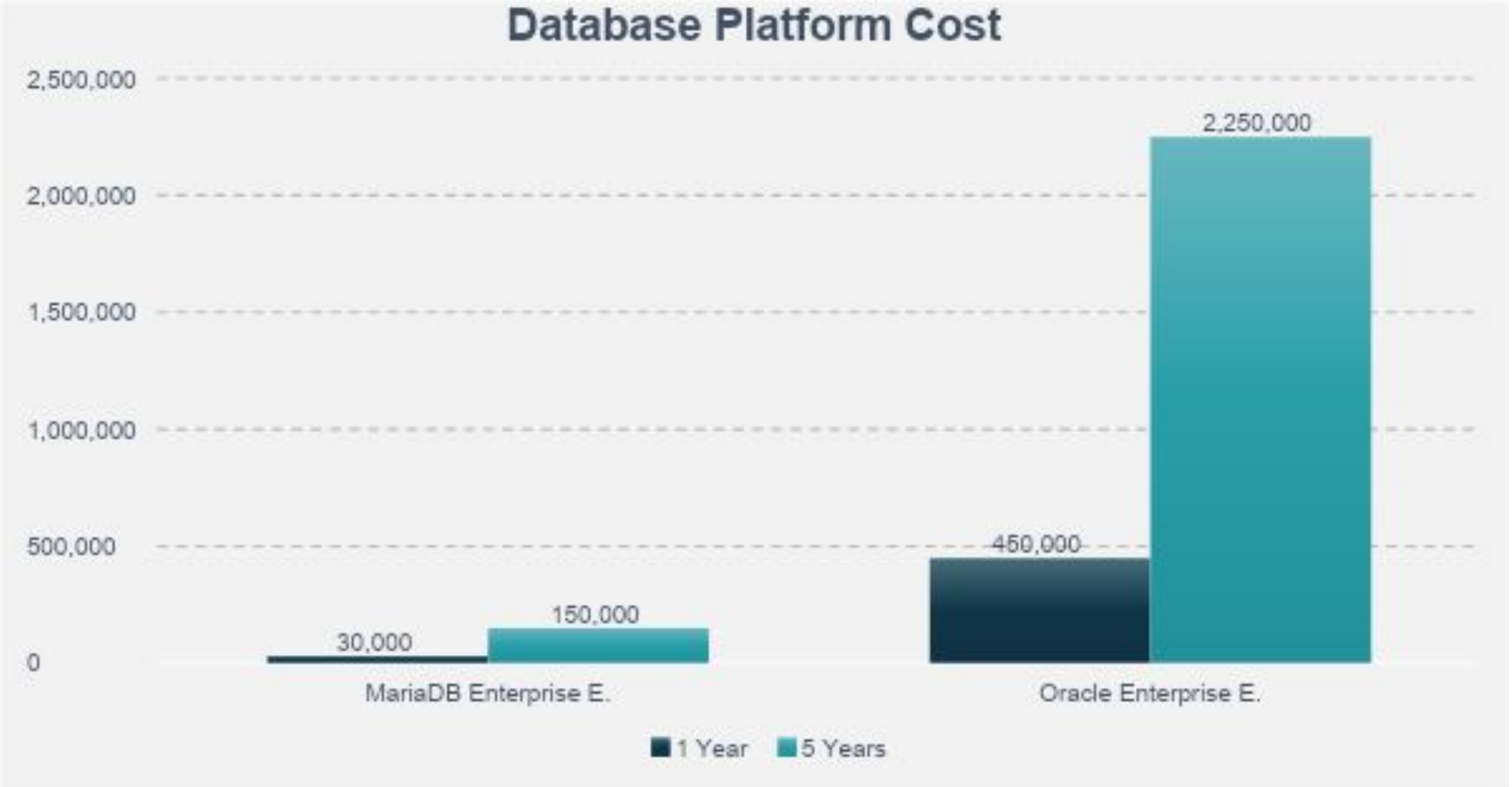
Parallel query execution




Public

# Tangible Chemical Space Appliance

## Total Cost of Ownership




Public

# Tangible Chemical Space Appliance

## Future Development

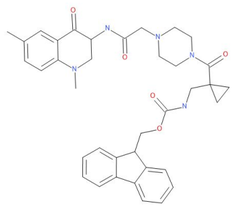
Customization

Ad Hoc intrinsic compression of Text Molecular Representation

Graph isomorphism search directly from DB

Fingerprint similarity Search & Clustering directly from DB

**The further expansion of the TCS leverage on the competence of the MariaDB development team. MariaDB supported us in the current implementation and will guide us in the deep integration of our algorithms in the DB architecture to achieve the highest performance**



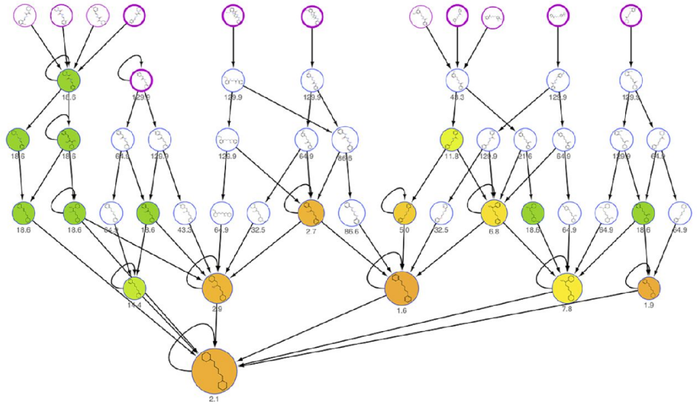
CN1CC(NC(=O)CN2CCN(CC2)C(=O)C3(CNC(=O)OCC4c5ccccc5c6ccccc46)CC3)C(=O)c7cc(C)ccc17

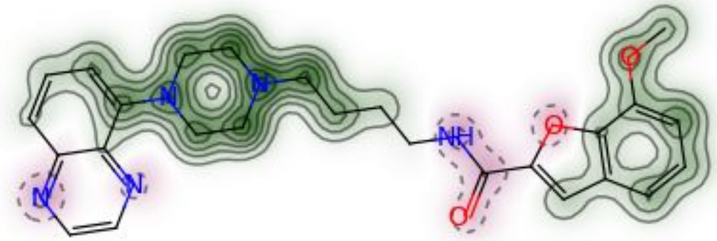
C□b?!?&by3L?$p!?.'?

>70% Single Record Intrinsic Lossless TXT compression

C□b?!?&by3L?$p!?.'?

Still compressible by MariaDB





Public

Thank You for Kind Attention !
Any Questions ?
exscalate.com

MariaDB

OPENWORKS

# THANK YOU

PLEASE DON’T FORGET TO GIVE US YOUR FEEDBACK ON THIS SESSION.
YOU CAN ACCESS THE SURVEY IN THE ENGAGEMENT PANEL
ON THE RIGHT SIDE OF YOUR SCREEN.