Intro to Drug Development and Appreciating Complexity

Rebecca Laposa

University of Toronto

Teaching team


Chris
CS instructor


Rebecca
PT instructor


Martin
PT instructor


Mica
TA


Ella
TA

Stef
TA

About Rebecca:

- Assistant professor, Pharmacology and Toxicology (Temerty Faculty of Medicine)
- Graduate training in Toxicology (drug safety), UCSF postdoc in cancer research
- Co-ordinator of graduate Drug Development course for PCL, industry-partnered
- Founder of spin-out venture in cancer drug discovery, national cross-functional team
- AI drug discovery partnerships since 2020
- I know very little about ML/AI and am hoping to learn more from you! (summer 2025 ML Bootcamp)

About this course

- Two sister course codes (CSC2541H / (PCL3107H, PCL3108H)), one course
- Reversed instruction – Martin and I am lecturing in CS and the CS instructor is lecturing in PCL
- **We're building the plane as it flies**

- First six weeks—lectures, notebooks, and project proposal (aka LOI)
- Second six weeks—co-working on a project
- Course website has all of the important information, including syllabus and schedule

Course Learning Objectives

- Our objective is to
  - prepare you for a future of working **collaboratively across disciplines** in both the biological sciences and AI
  - prepare you to brainstorming across disciplines while contributing your subject matter expertise
  - provide a **launchpad** to prepare you to turn inter-disciplinary ideas into falsifiable hypotheses and design an experimental approach

- We will not
  - teach you how to code
  - train you to be experts in AI or machine learning
  - train you in a certain wet lab technique
  - teach you in depth about how to develop or design drugs
  - teach you in depth about biology

Assessments

- **Code notebooks (collaboration highly encouraged).** Designed to be more approachable with a lower barrier to entry in terms of programming background. The biological context may be unfamiliar but we have great TAs that will be around to help and your peers in the PCL coursecode.

- **LOI (teams of 3~4 with 2-3 CS 1 PT).** A project proposal, which you can think of as a letter of intent for a grant application. Answers the question, what project are you hoping to work on?

- **Project report (same team as LOI).** A project report, which you can think of more as a grant proposal with some preliminary results. Think of this as a *launchpad* that you could in principle turn into an exciting collaborative project. You should have some initial results and a hypothesis, but it doesn't have to wrapped up with a bow.

- We'll have more details in the coming weeks.

Waitlist and class size

- There is a big waitlist on the CS side and the course is designed around collaborative research teams of 3-4 people.
- If for any reason
  - you are unable to attend this course in person or
  - think it's possible that you will drop it midterm or
  - are not interested in working towards a collaborative research project

- please drop the course very soon to allow students from the waitlist in.
- If you need to miss a few classes, please email the teaching team so we can discuss if we can accommodate.
- We look forward to an exciting term with you all!

Outline

1. Why Drug Development is Necessary
2. Why Drug Development Needs You
3. Scope of Drug Development “Pipeline”
4. Preclinical Drug Development
5. Clinical Drug Development
6. Case studies: Complexity and Opportunities
7. Business/Financial Aspects and Value Inflection Points

Why Drug Development is Necessary

Unmet medical needs

- First step in drug development – what do we even tackle? Narrow our focus
- Unmet medical needs are
  - Medical conditions that are not adequately treated by current therapies
- Current therapies may be:
  - Absent
  - Less than maximally effective
  - Undesirable side effects / toxicity
- Often the biggest obstacle in drug development is understanding the mechanisms of the disease in the first place (biomedical sciences)

- In your opinion, which diseases are currently unmet?

World Health Organization Disease Research and Development Priorities

World Health Organization | Health Topics | Countries | Newsroom | Emergencies | Data | About WHO

Priority diseases

Communicable diseases

Ebola virus disease

Communicable diseases

Lassa fever

Communicable diseases

Crimean-Congo haemorrhagic fever

Communicable diseases

Middle East respiratory syndrome coronavirus (MERS-CoV)

Communicable diseases

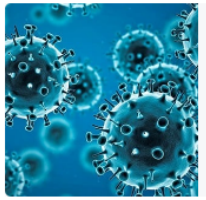
Nipah virus infection

Communicable diseases

Rift Valley fever

Diseases and conditions

Zika virus disease

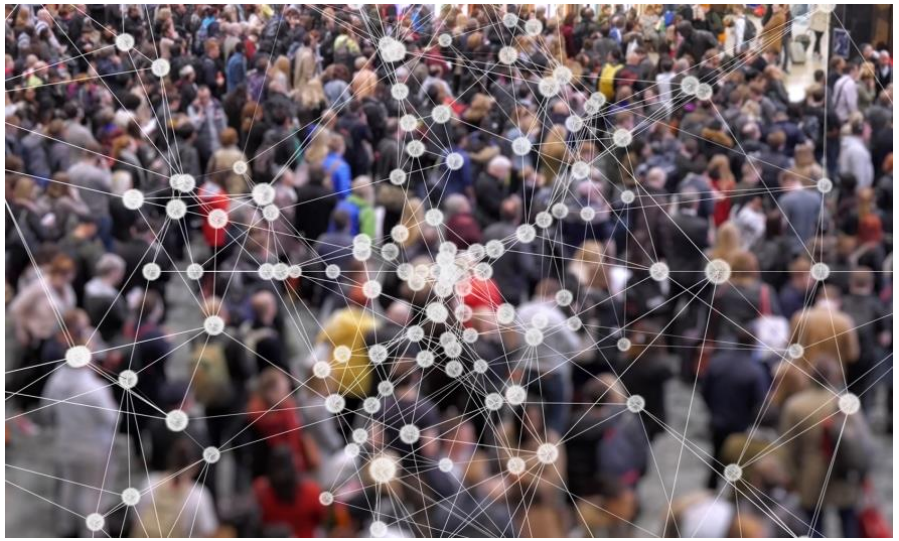


Communicable diseases

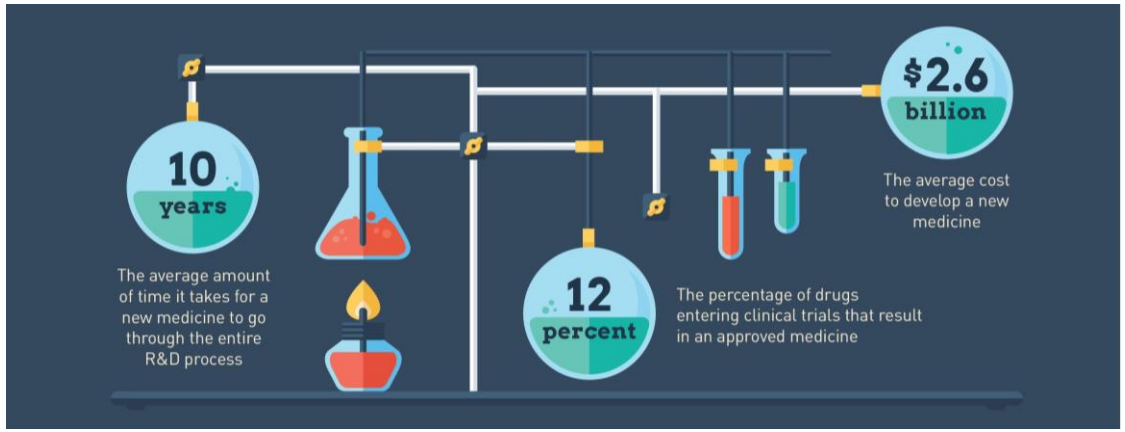
Coronavirus disease (COVID-19)

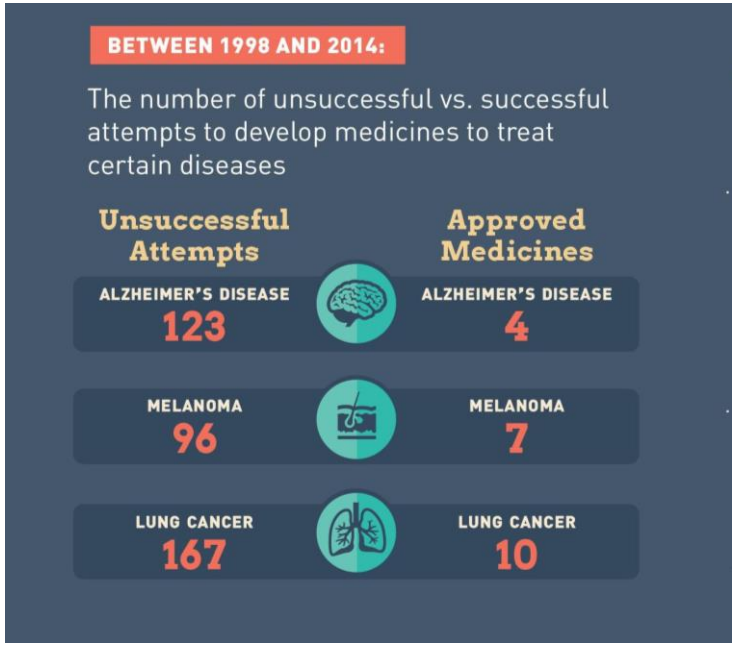
Disease X

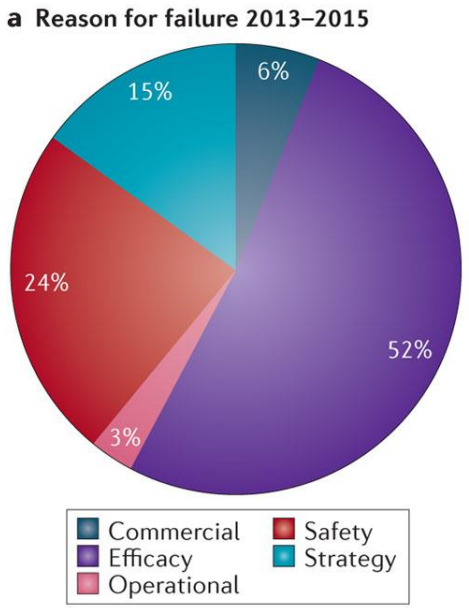
- *Disease X represents the knowledge that a serious international epidemic could be caused by a pathogen currently unknown to cause human disease.*
- *The WHO R&D Blueprint explicitly seeks to enable early cross-cutting R&D preparedness that is also relevant for an unknown "Disease X"*



Current Drug Development – Extensive Failure







Opportunity to improve the failure rate in DD

PhRMA, Harrison Nat Rev Drug Discovery

Why Drug Development Needs You

Efficiency and Effectiveness

- Efficiency
  - “doing things right”
  - Maximizing resources, minimizing waste
  - Cost, time
- Effectiveness
  - “doing the right thing”
  - In DD, actually tackling the essential mechanisms of disease
  - In DD, providing true benefit to patients

Drug Development opportunities in AI/ML

- Need and opportunity for dual skills in AI and life sciences

Dual-skilled professionals can:

- appreciate complexity of disease
- address meaningful drug targets
  - what we “aim” our drugs against
- Understand how data in drug development is processed, stored, what are the algorithms
- Create meaningful AI models based on an integrated understanding

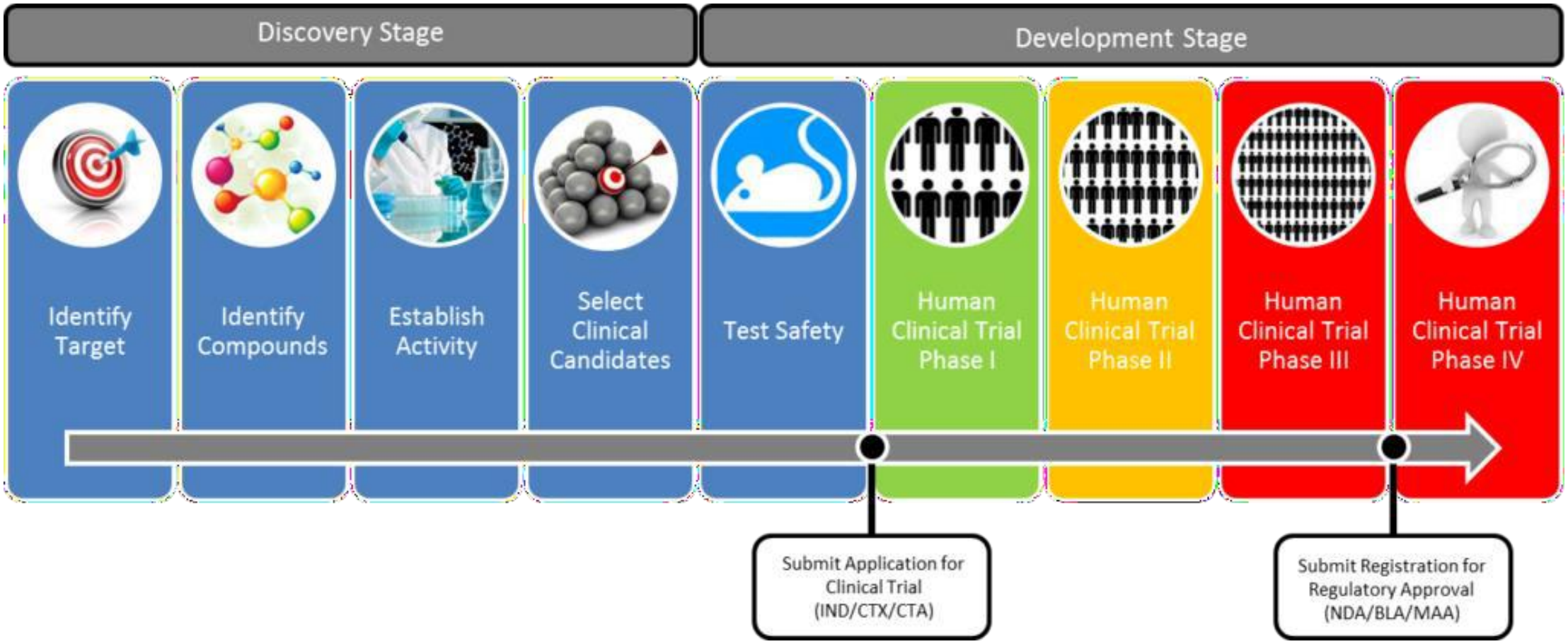
- Be both efficient AND EFFECTIVE

Drug Development Pipeline Overview

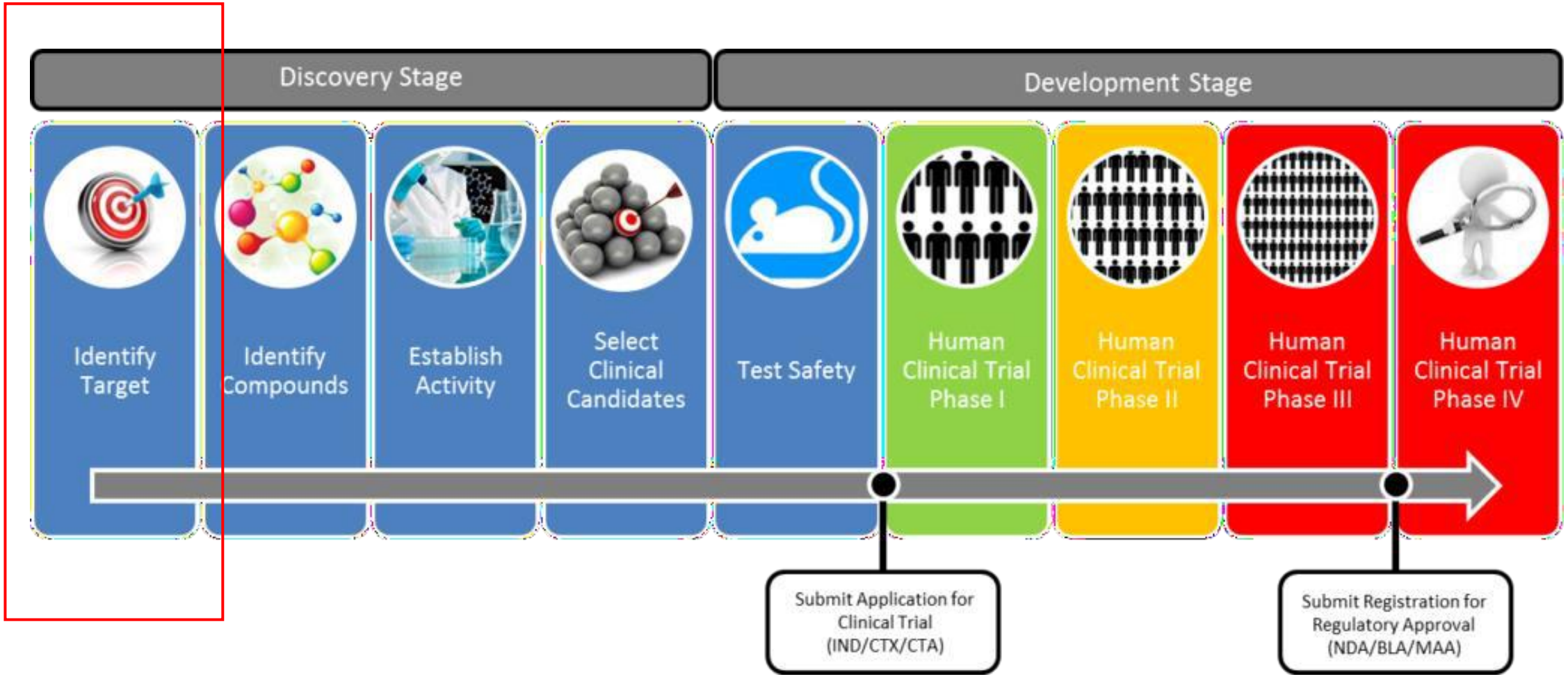
Drug development is a **sequential decision pipeline**

- Drug development: a series of decisions made under uncertainty
  - “Go – No Go decisions” to progress from one step to the next
- Each step compresses reality
  - complex diseases → biochemical pathway → single protein → single drug
- Pipeline is generally unidirectional
  - backtracking is possible (but inefficient)

Drug Development “Pipeline”

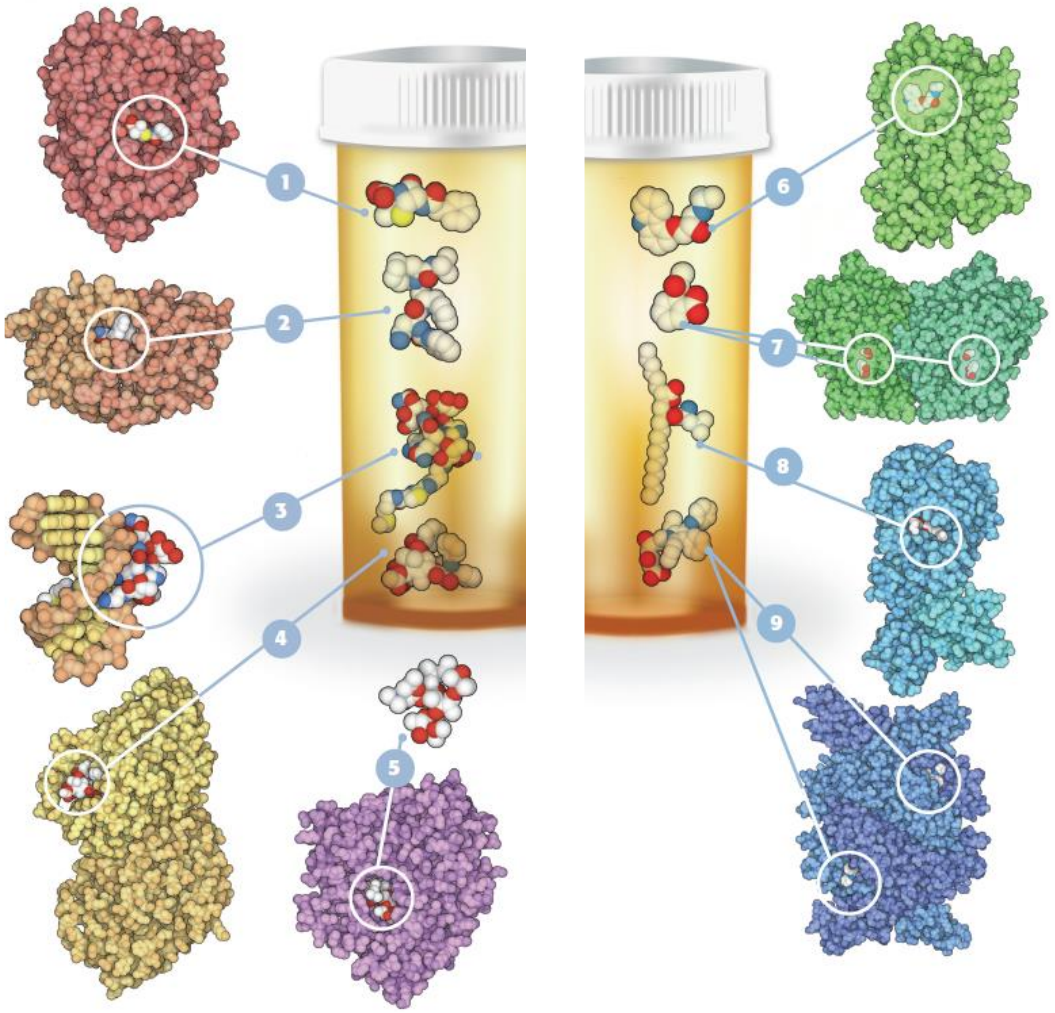


Drug Development “Pipeline”



Drug Target

- A specific protein that the drug is designed to modulate



Target Identification

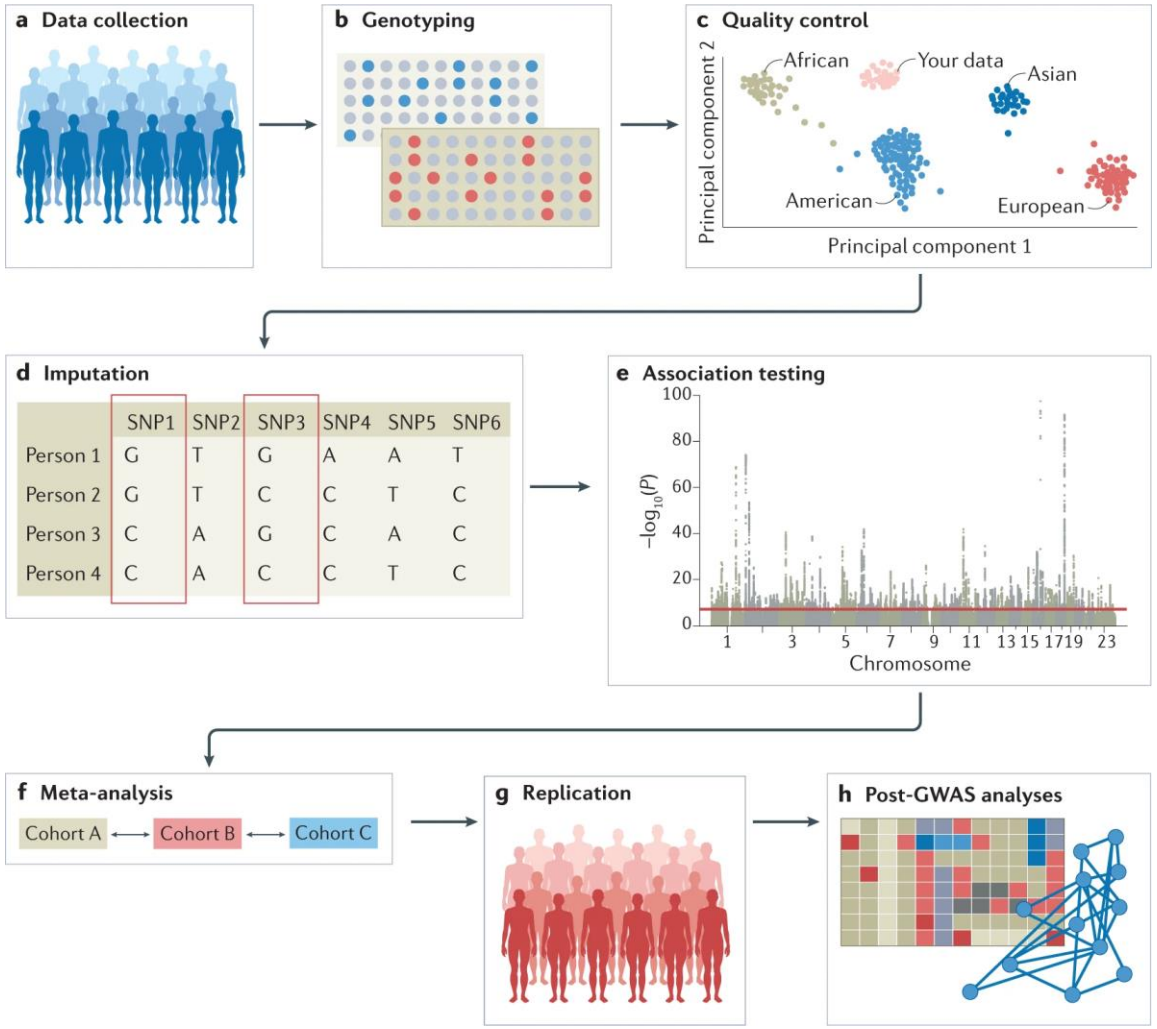
Choosing which protein to modulate

- Target identification is….
- Hypothesis:
  - Modulating a particular molecular entity (eg a specific protein) will change disease state
- Sources of evidence:
  - Genetics (GWAS, Mendelian disease)
  - Omics correlations
  - Animal models
  - Literature priors
- You're choosing a variable you believe is **causal**, not just predictive.

Key assumption

- Perturbing this variable is **sufficient** to control disease.
  - This is the strongest and riskiest assumption in the entire pipeline.

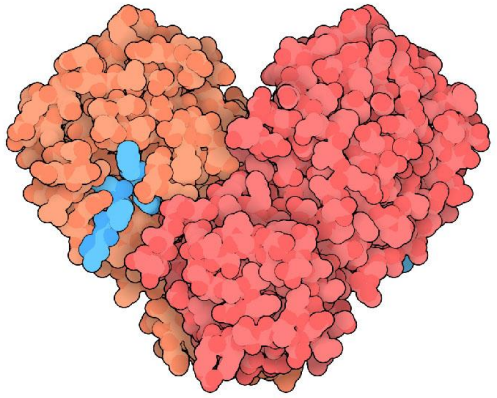
Genome-wise association studies (GWAS) connect disease status with DNA sequence variation



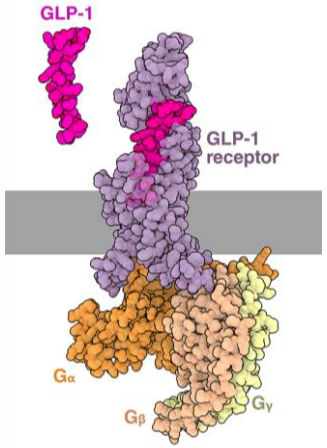
Uffelmann et al., Nat Rev Methods Primers 2021

Target Identification examples

Disease	Drug Target
SARS-CoV-2	Protease enzyme required by the virus for its life cycle
High blood pressure	Adrenaline receptor in the heart
Depression	Serotonin transporter
Diabetes	Glucacon-like-peptide 1 (GLP-1) receptor



SARS-CoV2 main protease



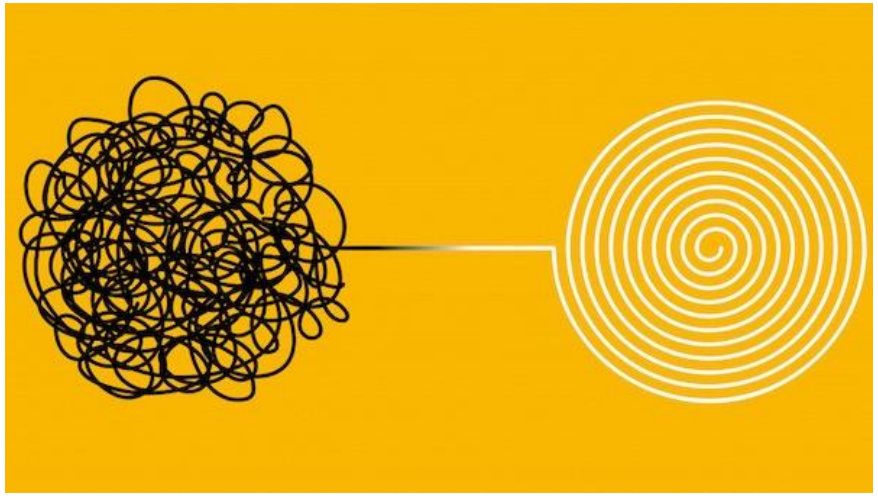
GLP-1 receptor

Why might this focus on single targets be problematic?

- Oversimplification of disease complexity → drugs that don’t work
- Examples later today of failed drugs

Complexity is not noise – we can contend with it

- Complex biological systems are not random.
- Complex biological systems are:
  - structured
  - constrained
  - learnable
- Ways to think about biological complexity:
  - biological networks → graphs
  - biochemical pathways → interacting modules
  - cell appearance and behaviour → system states
- The hypothesis class needs to be richer



- An opportunity for **MODERN AI models in DD** to effectively address biological complexity

Image: Boston College

*Why **complexity** matters for AI in drug development*

- AI will not rescue drug discovery if we collapse biology too much *before* the model ever sees it.
- Our algorithms can only learn from the assumptions in what we measure and evaluate (what we “bake in”)
- The future of drug development isn’t single targets versus complexity.
- It’s choosing the right level of abstraction and the appropriate balance between reductionism and accurately representing “ground truth”

Target Identification in this course

- Substantial focus on target identification in this course – due to the importance of this step in DD overall

- Drug target proteins and how to modulate them with small molecules: Module 1
- Identifying targets from cell structure and behaviour: Module 2
- Identifying targets from genome and RNA sequences in patients and animal models: Module 3

After Target Identification, need ***Target Validation*****:**
Testing whether the variable actually controls the system

- Experimental question:
  If I intervene here, does the system move in the direction I want?
- This is **causal inference**, not prediction.
- Experimental approaches:
  - Knockout / knockdown of gene
  - Overexpression of gene
  - Chemical probe to inhibit protein
  - Disease models (artificial/simple)

Target Validation is usually done in *simplified systems*

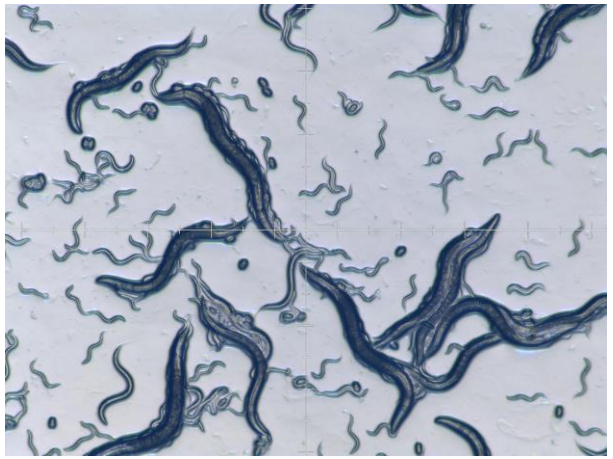
- Cell lines
- Transgenic mice

Researchers are validating causality in a **reduced** state space.

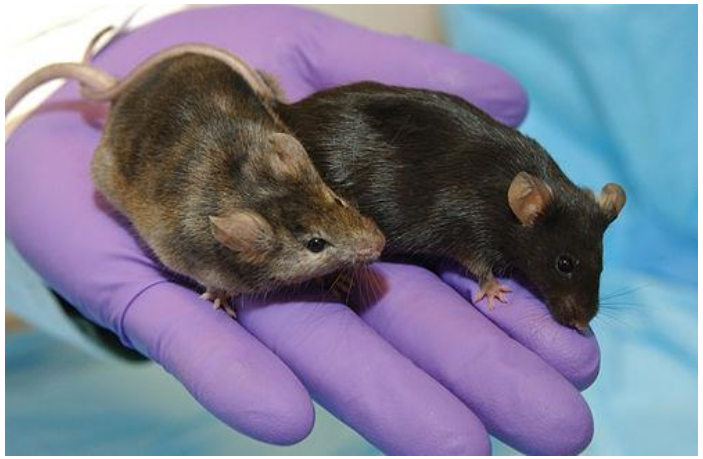
Example **model organisms** used in drug development

Experimental question in target validation:

- does changing the target protein (e.g. by genetic approaches) change the organism's health?



roundworms

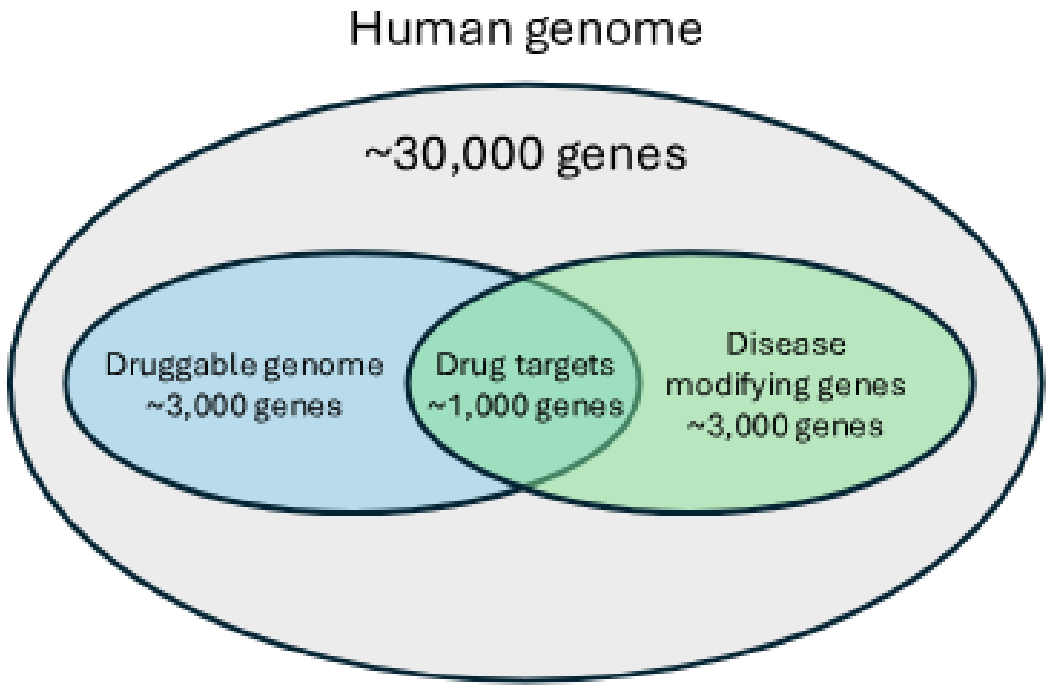


mice

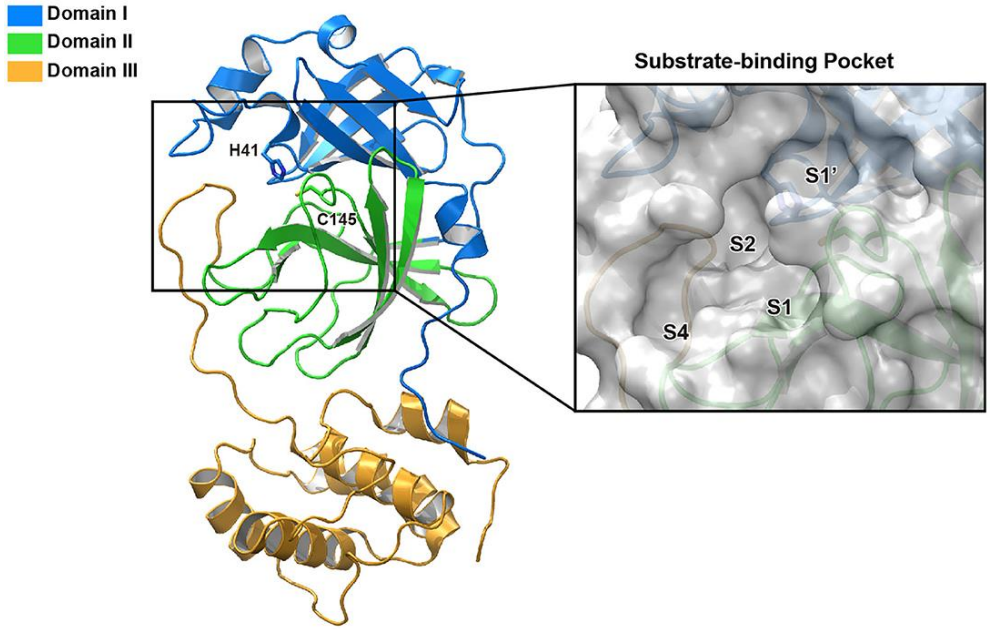
zebrafish

Druggability

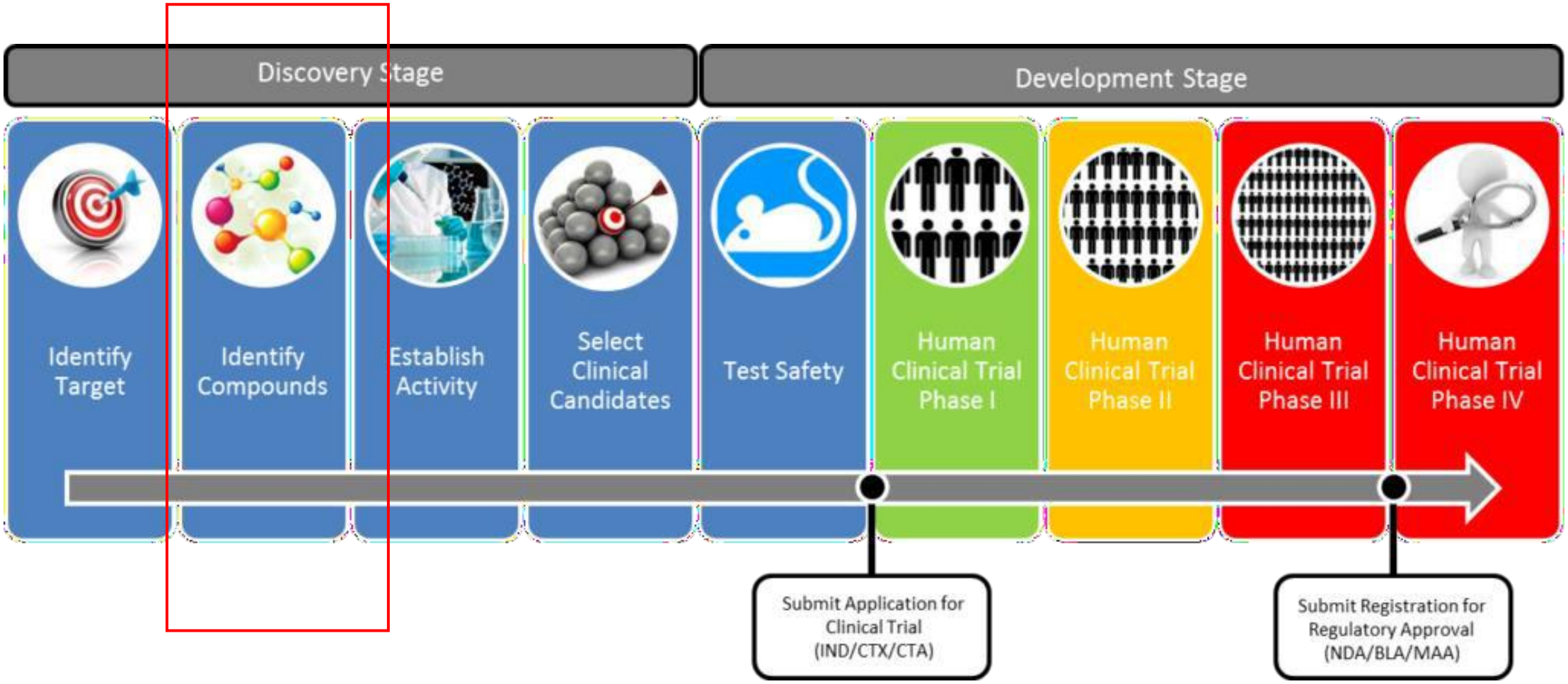
- A **druggable protein** can be functionally modified by a drug
- Has a 3D shape that is amenable to drug binding and action



SARS CoV2 protease binding pocket

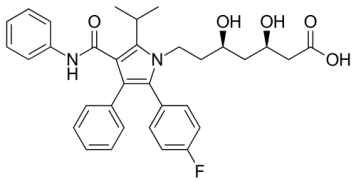


Drug Development “Pipeline”

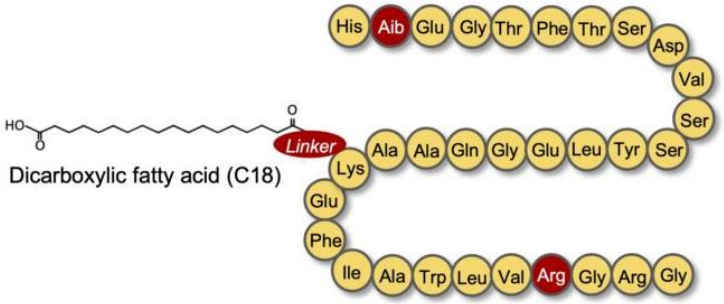


Kinds of drugs: small molecules and biologics

Small molecules (chemicals)

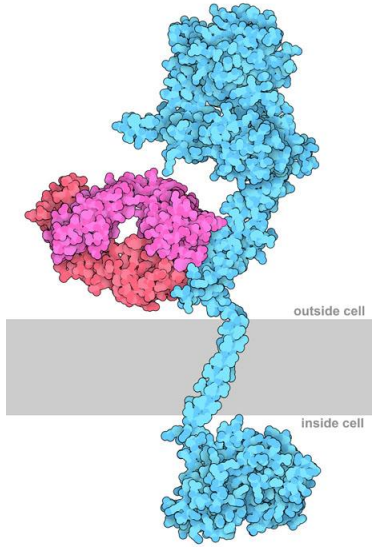


peptides



Semaglutide (165 h)

Biologics such as antibodies



Example:
statin drugs for blood pressure

Example:
Ozempic and other GLP-1 drugs

Example:
Herceptin and antibody drugs

Biologics – 30% of new drug approvals

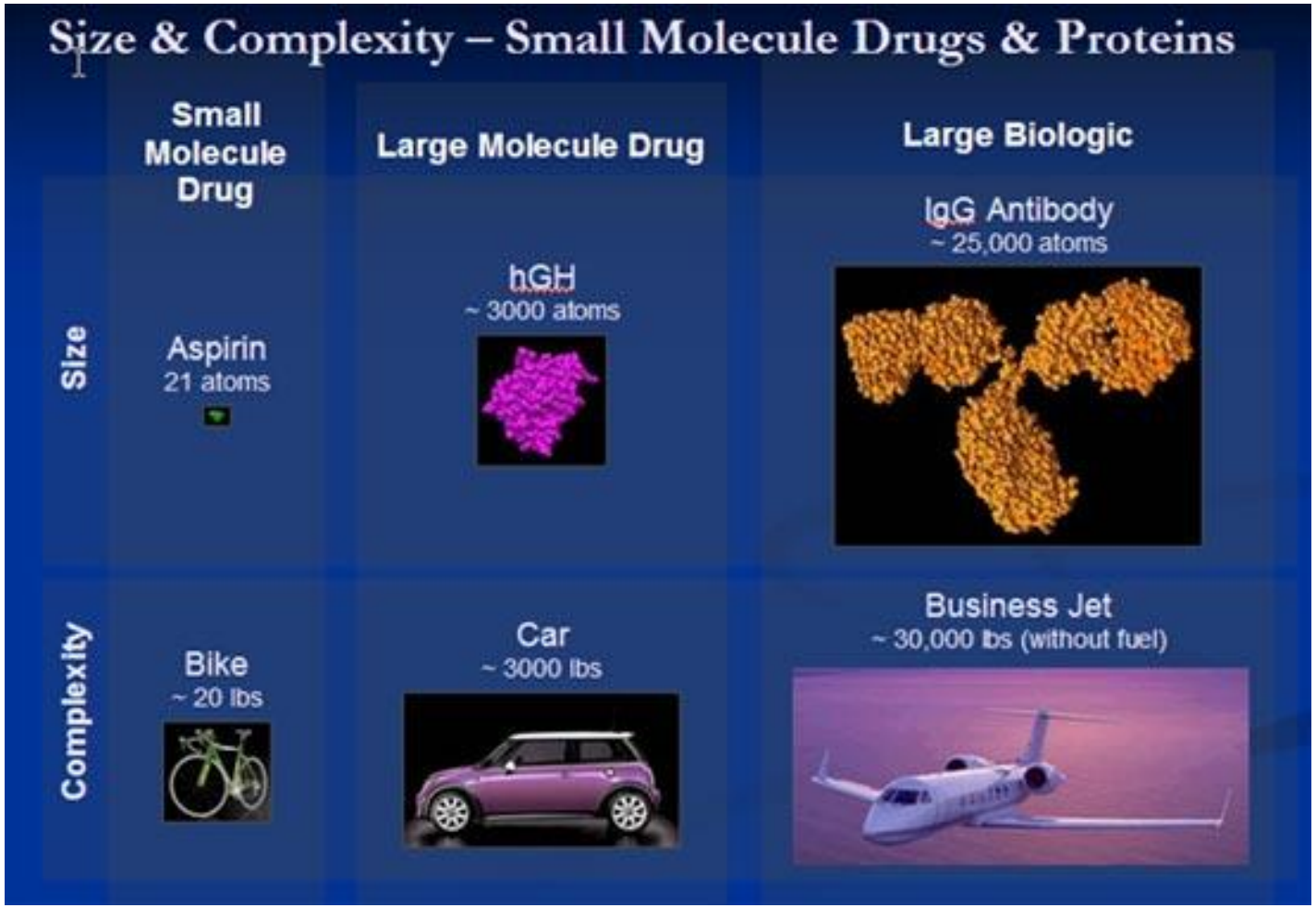
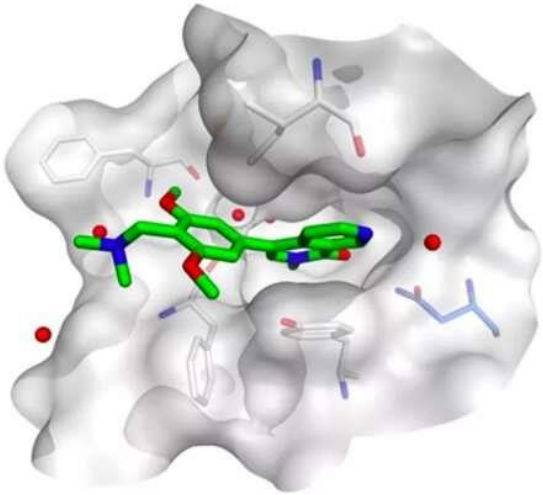
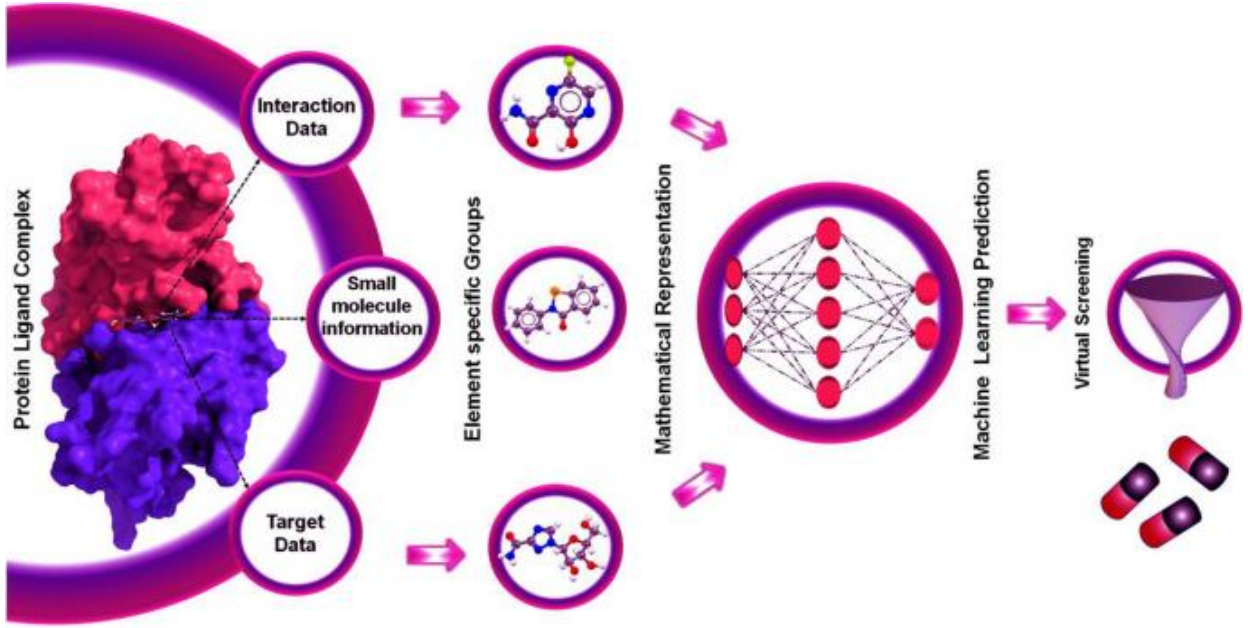


Image: AZBio

Predicting interactions between chemicals and proteins in silico

- Focus of Protein Lectures (next week) and Module 1 Google Collab

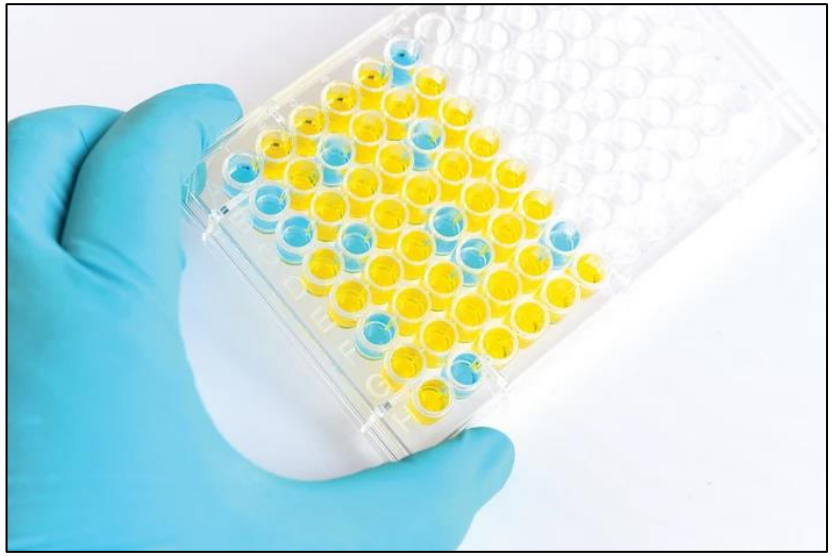




Selveraj et al., 2021

Experimental testing of drugs: Assay Development

- Chemicals predicted in silico often need to be followed up by experimental testing
- Average success rate of this: <<100%
- Before we test whether different drugs alter the target, we need an assay that provides a readout of “success”
- Usually in a simplified system



Chemical and Engineering News

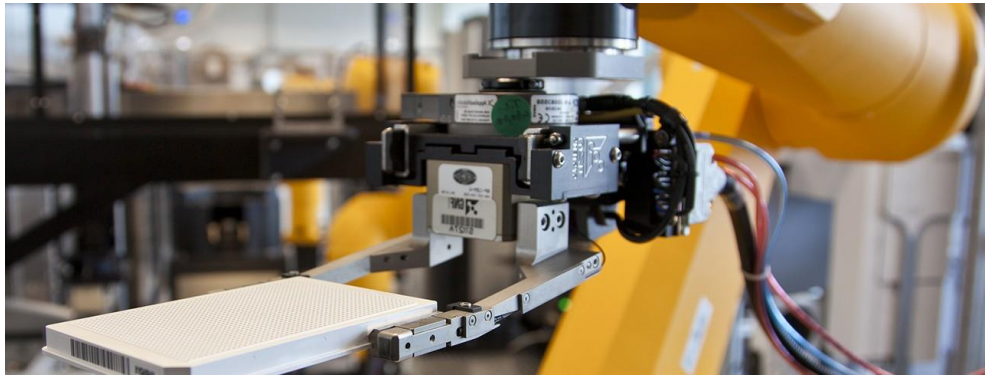
Assay Development:
Defining the objective function

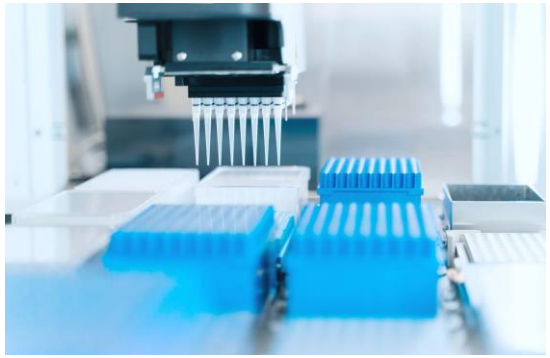
- An assay is how we decide whether a compound is ‘good.’
- Examples:
  - Enzyme inhibition
  - Reporter gene readout
  - Cell viability (Module 2)
  - Binding affinity (Module 1)

- An assay is a **surrogate evaluation**.
  - True objective: drug to improve patient outcome
  - Surrogate: drug to change assay readout
- Assays are chosen because they’re measurable, scalable, and reproducible, not because they perfectly represent disease.

High-throughput screening

- Robotic testing of hundreds of thousand of chemicals for their ability to modulate protein function
- Simple assays (see Assay Dev)
- “Hit rate” ~ 1%

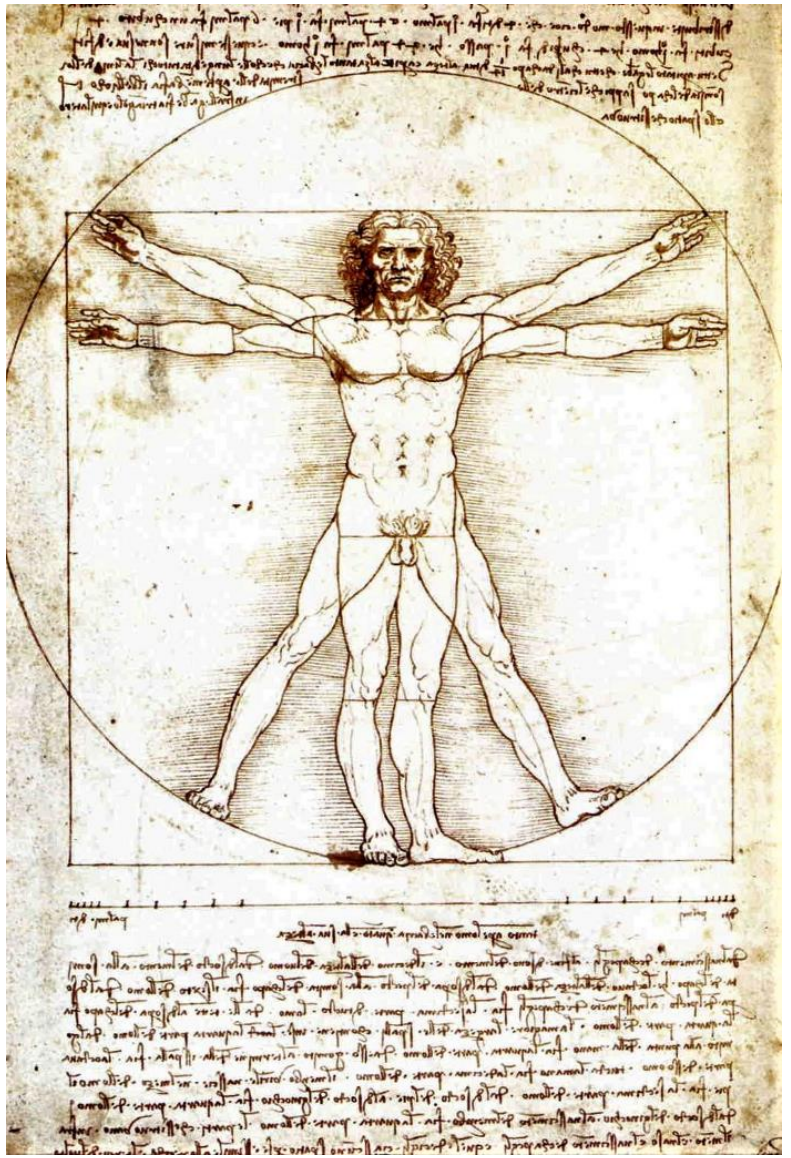




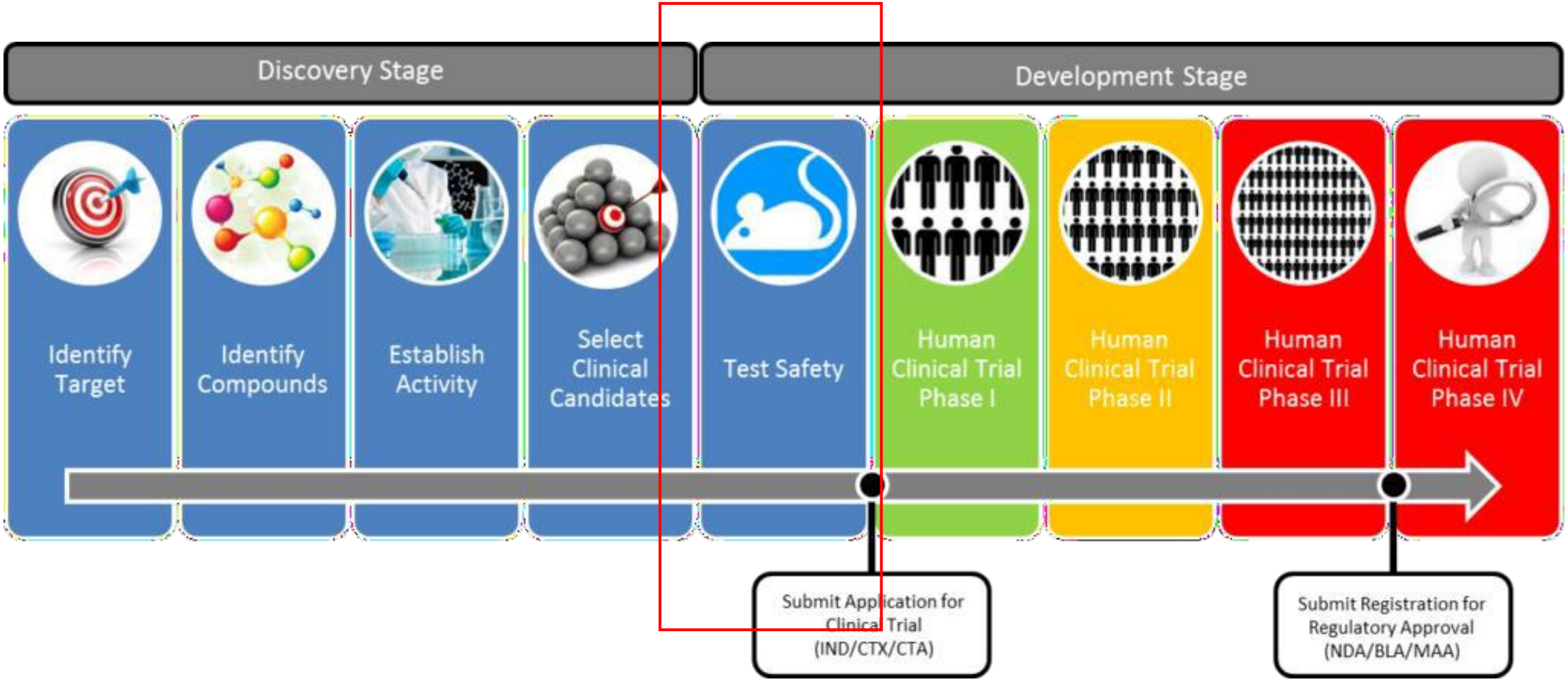
What does a drug need to do?

- After HTS, we have “hit compounds” (starting point)
- But a drug needs to achieve far more than just bind to and change a protein
- An orally-administered **drug** must:
  - Cross the gut and get into the bloodstream
  - Move to the target organ
  - Stay at the target protein for a reasonable length of time
  - Be safe
  - Be eliminated/removed from the body

- Chemists optimize the chemical structure of a hit (protein binder) to change these properties
- Chemical structure → biological effects



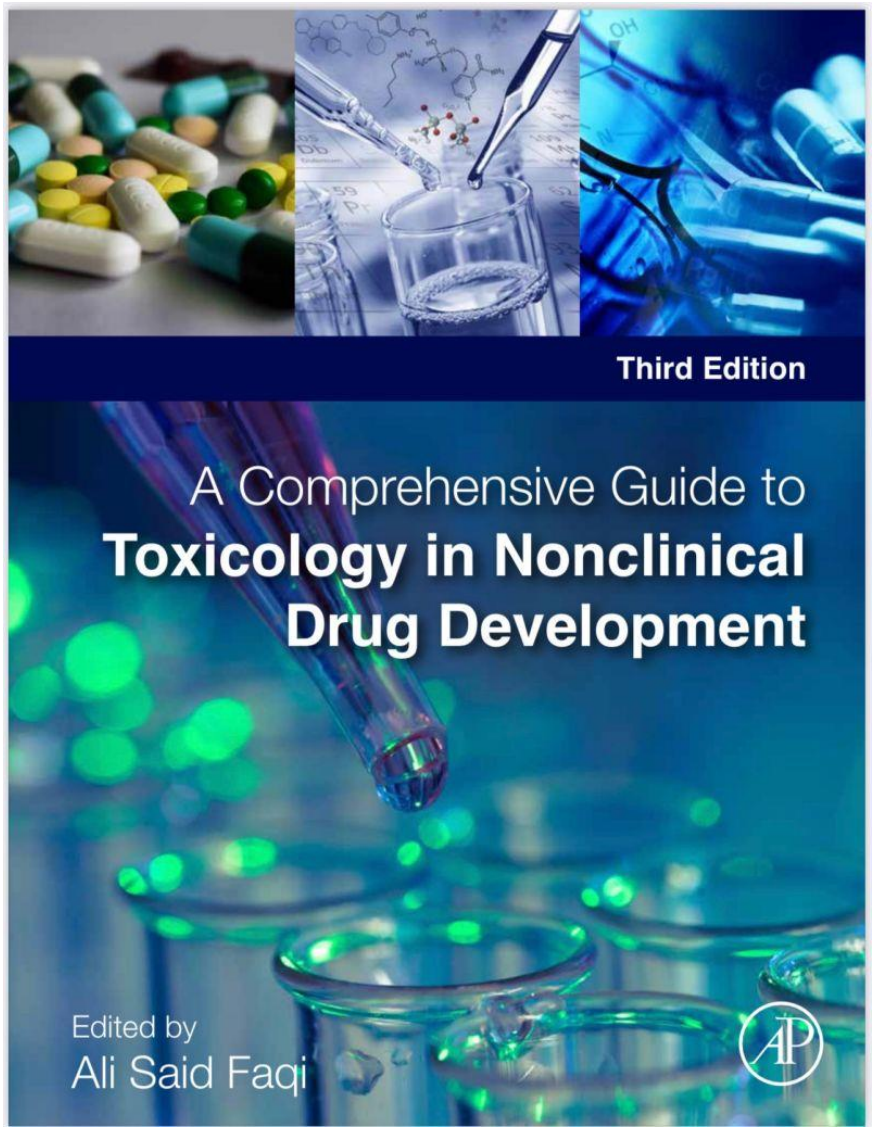
Drug Development “Pipeline”



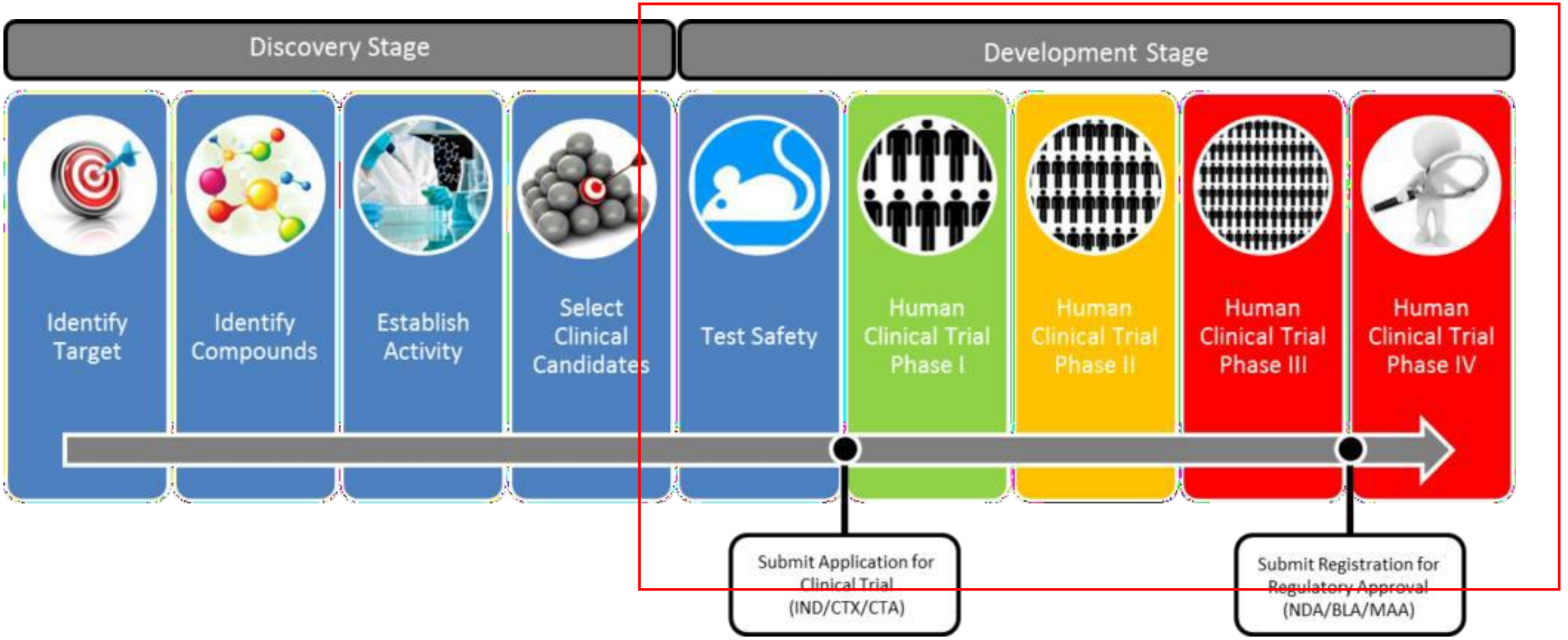
Before testing in humans

- Animal testing for efficacy
  - Does the drug “work”
- Animal testing for safety
  - Safety is very difficult to predict
  - Regulations require 2 or more species

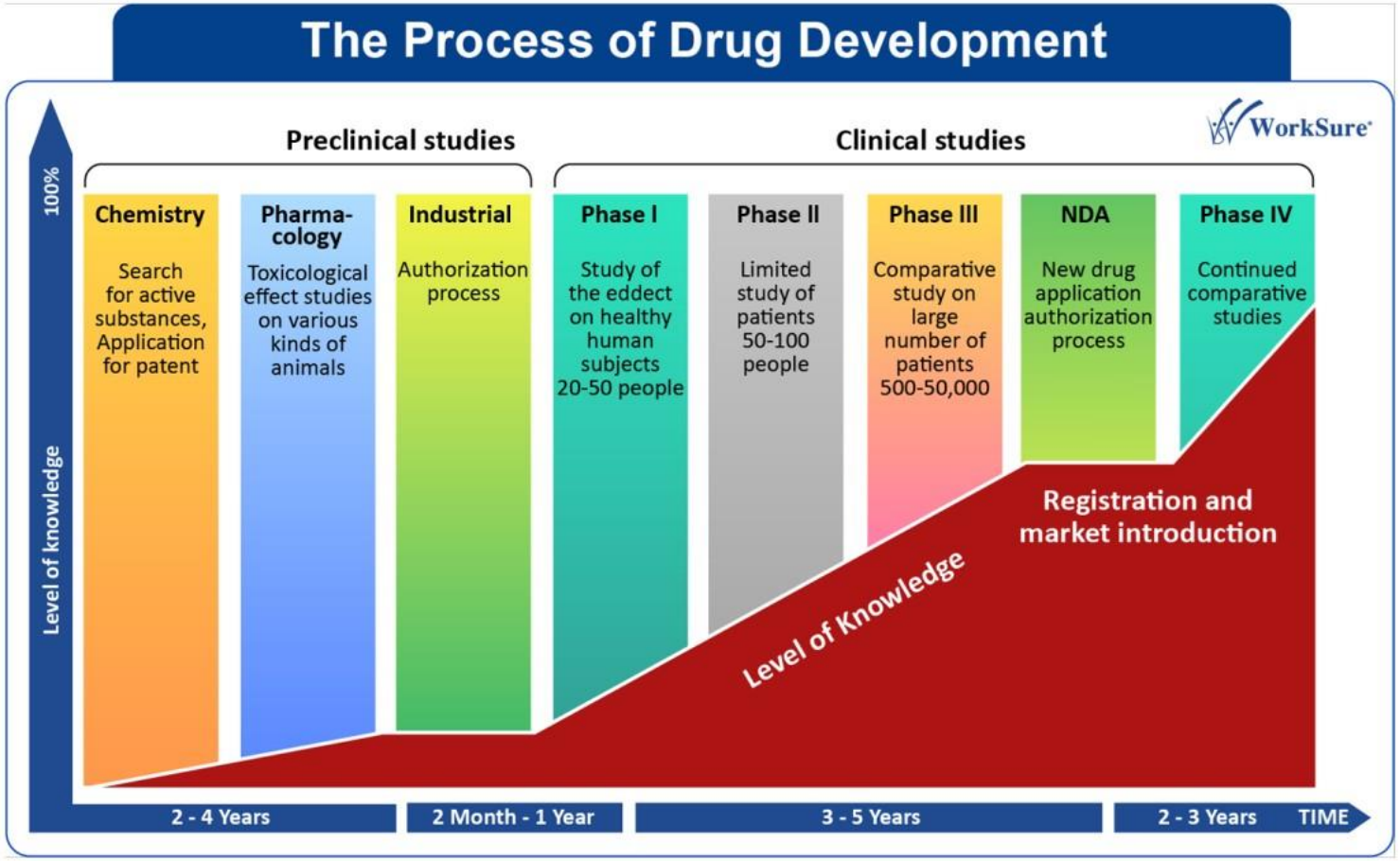
- Permission to test drug in humans (federal government)



Drug Development “Pipeline”

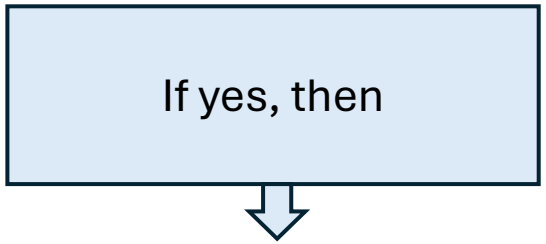


Clinical trials – stepwise de-risking of the drug

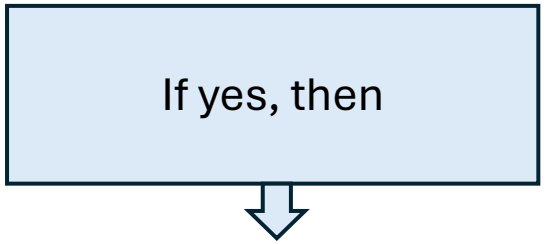


Phases of clinical trials: Stepwise de-risking

- Phase 1: Is the drug safe?
  - Healthy volunteers
  - < 50 people



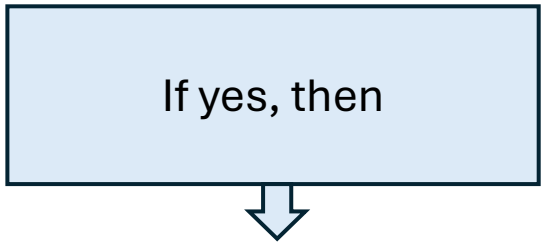
- Phase 2: Does the drug “work” to treat disease?
  - Patients (first time in patients)
  - Short-term efficacy (often biomarkers)
  - Several hundred people



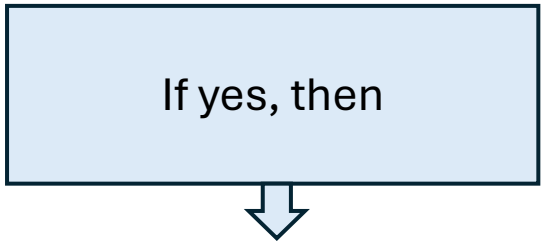
- Phase 3: Does the drug “work” with statistical confidence?
  - Patients
  - Much larger population (thousands, global sites)
  - Longer-term efficacy, still follow safety
  - Trial results used by governments to approve drug if data and risk-benefit are sound

Phases of clinical trials: Stepwise de-risking

- Phase 1: Is the drug safe?
  - Healthy volunteers
  - < 50 people



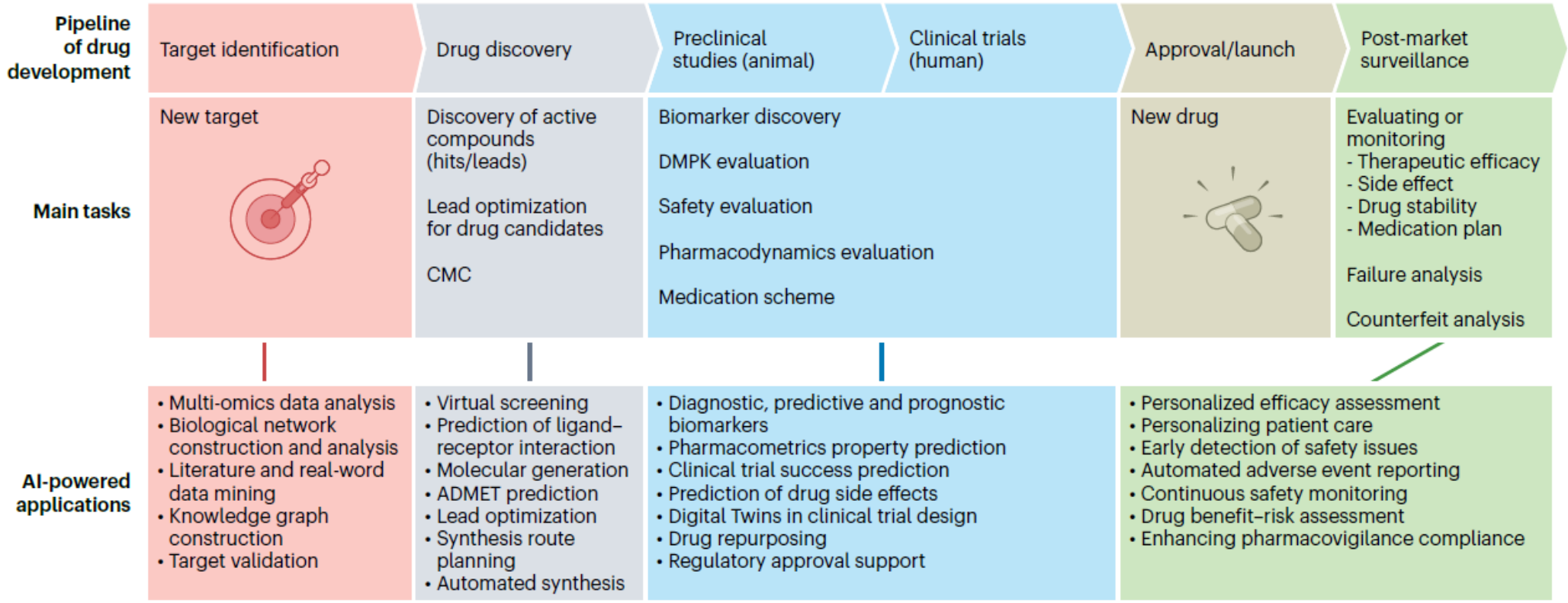
- Phase 2: Does the drug “work” to treat disease?
  - Patients (first time in patients)
  - Short-term efficacy (often biomarkers)
  - Several hundred people



- Phase 3: Does the drug “work” with statistical confidence?
  - Patients
  - Much larger population (thousands, global sites)
  - Longer-term efficacy, still follow safety
  - Trial results used by governments to approve drug if data and risk-benefit are sound

Federal Drug Approval:
drug becomes available in pharmacies

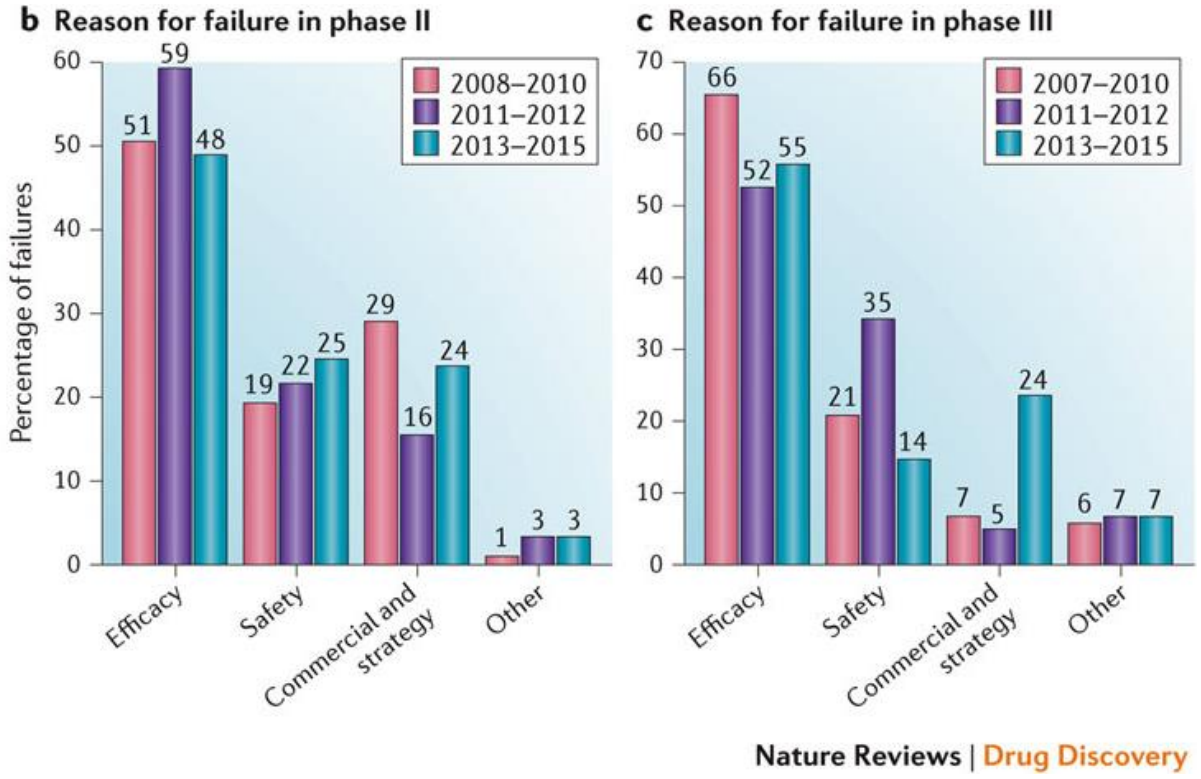
Some applications of AI in the Drug Dev Pipeline



Zhang et al., Nature Medicine 2025

Why do drugs fail in clinical trials?

- Major reason for failure is lack of EFFICACY. Drug doesn't "work"



AI for drug discovery attempts to improve this kind of failure, by acting much earlier in the DD pipeline

Harrison Nat Rev DD 2016

Ethical aspects of clinical trials.
Behind every approved drug are thousands of clinical research participants. Our “medical ancestors”

AI for DD: an opportunity to serve patients and clinical trial participants better

Case studies

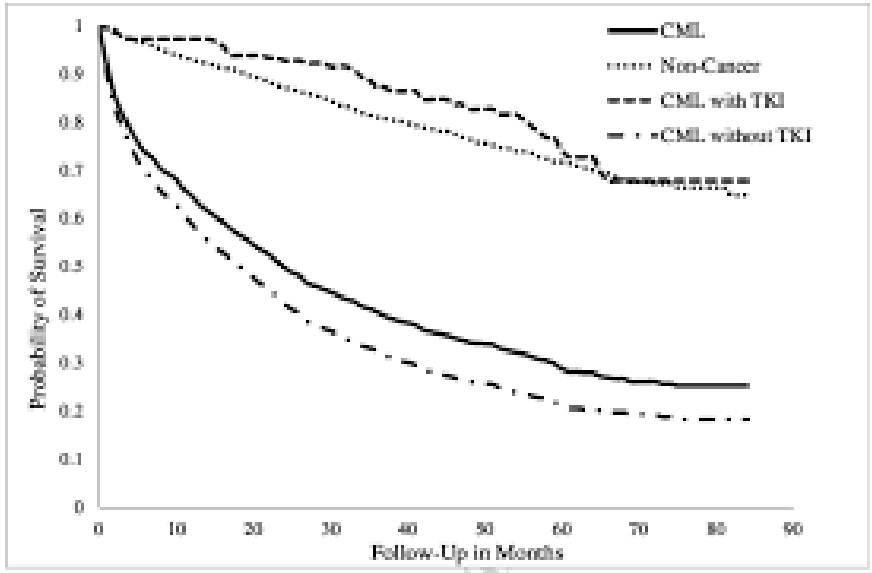
- Now that you understand how much simplification happens *before* a drug ever reaches a human,

- let’s look at what happens when the disease itself refuses to be simple.

Case Study 1: Gleevec (2001)

- Chronic myeloid leukemia (CML)

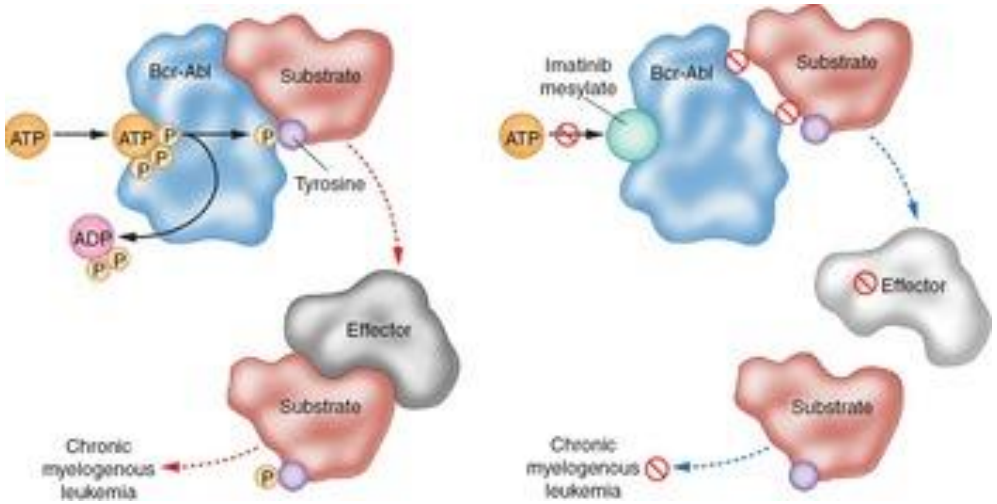
- One dominant genetic driver.
  - DNA rearrangement to create new gene
- One protein target.
  - Unique protein formed only in CML
- One drug - Gleevec
- Dramatic clinical effect.



Reductionism *as an implicit model*

- Reductionism: a modeling assumption.
- Reductionism assumes that disease causality flows through a **small number of dominant variables**.
- There is:
  - low-dimensional latent structure
  - strong signal, weak interactions

- When the modeling assumption of reductionism is correct, single-target drugs (like Gleevec) work beautifully.



Gleevec (imatinib) mechanism:
Inhibits a **single, disease-relevant enzyme**

Case Study 2: **Why “Perfect Targets” Fail**.

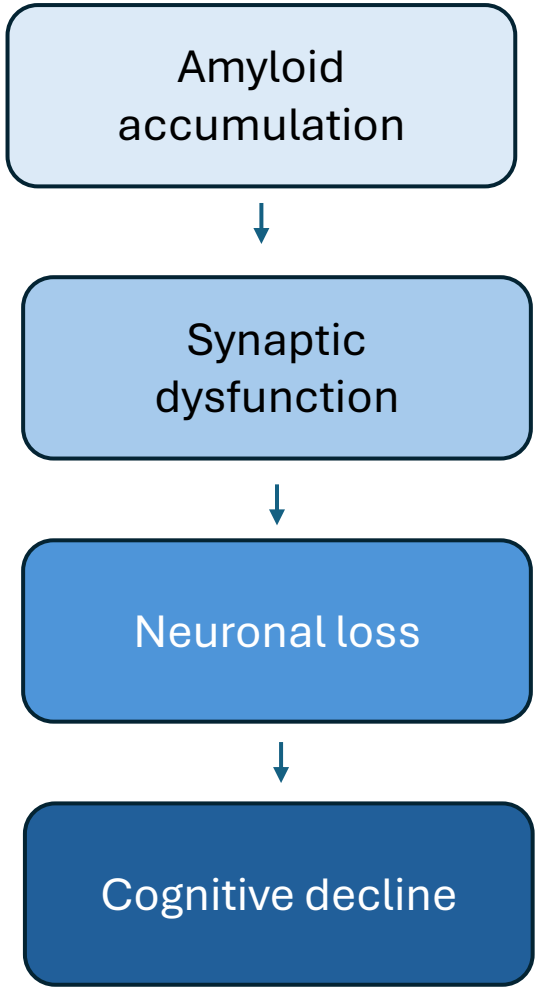
Alzheimer’s Disease and the Cost of Oversimplification

Alzheimer's Disease Model (~1995)

- Disease **model**

- It generated **correct predictions**
  - Genetics supported it
  - Biomarkers correlated
  - Pathology was reproducible

- From a modeling standpoint, this was a very reasonable hypothesis.
- But is was not a good causal model
  - predicts correlation not cause

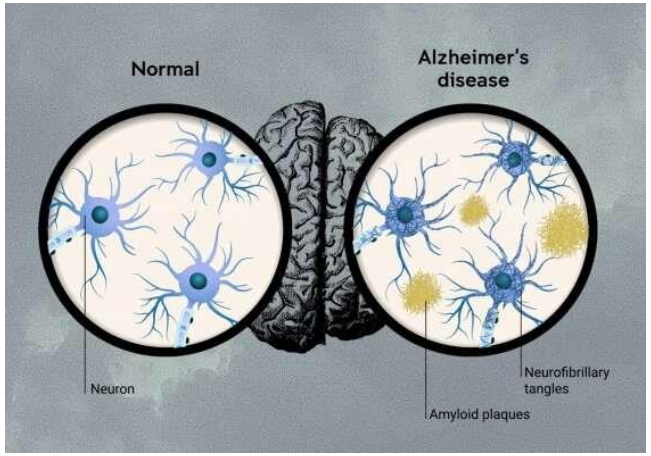


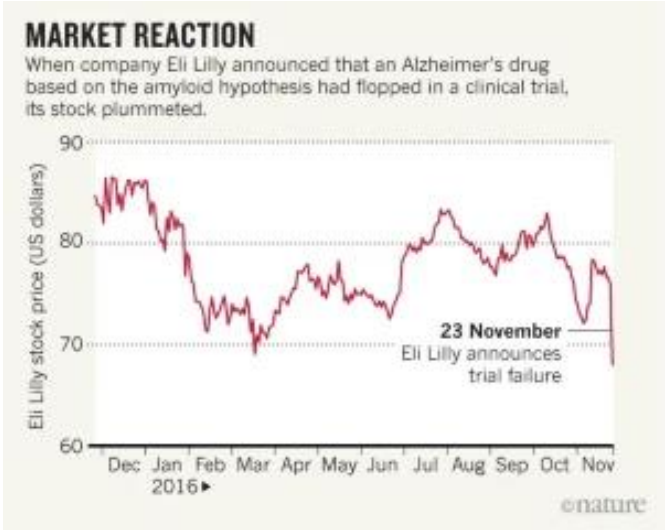
Amyloid Hypothesis: Intervention and Outcome

- **Intervention**
- Many companies developed drugs that removed amyloid
  - Variety of drugs with variety of mechanisms

- **Outcome**
  - Amyloid levels dropped
  - Cognition **did not** improve

- This is a **model failure**, not a drug failure

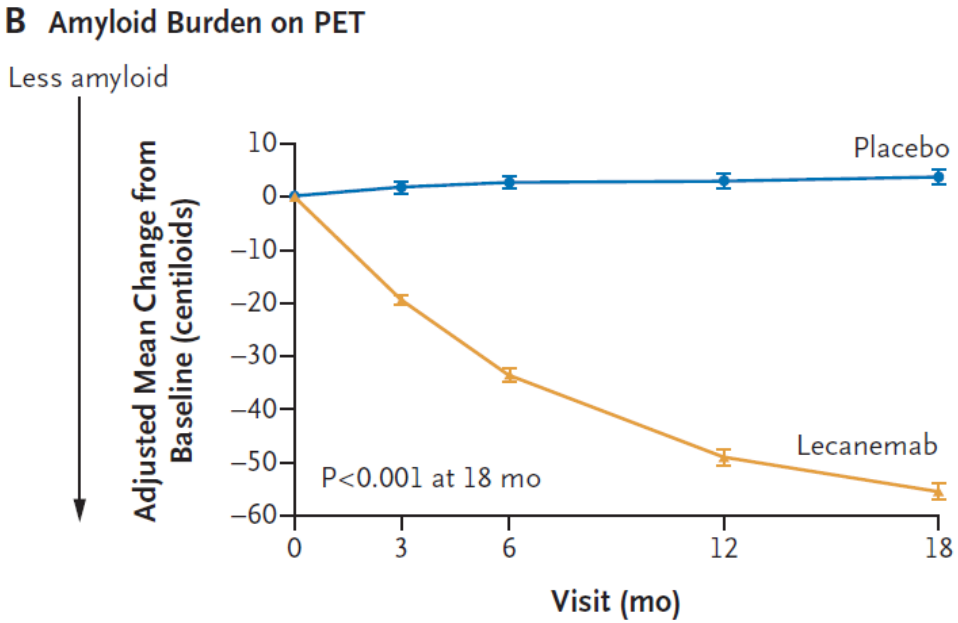




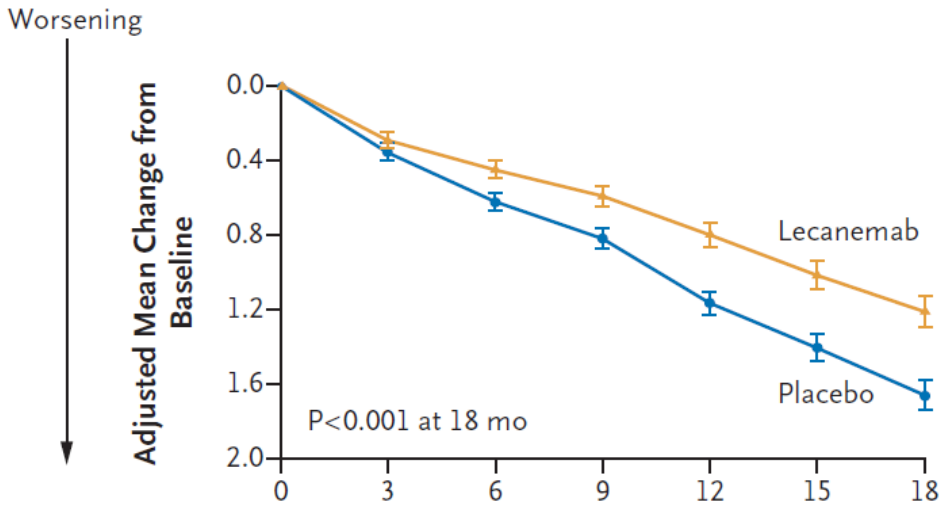
Images: Medical Xpress, Nature

Discordance between drug's effect on protein target and on Alzheimer's disease symptoms

Protein target

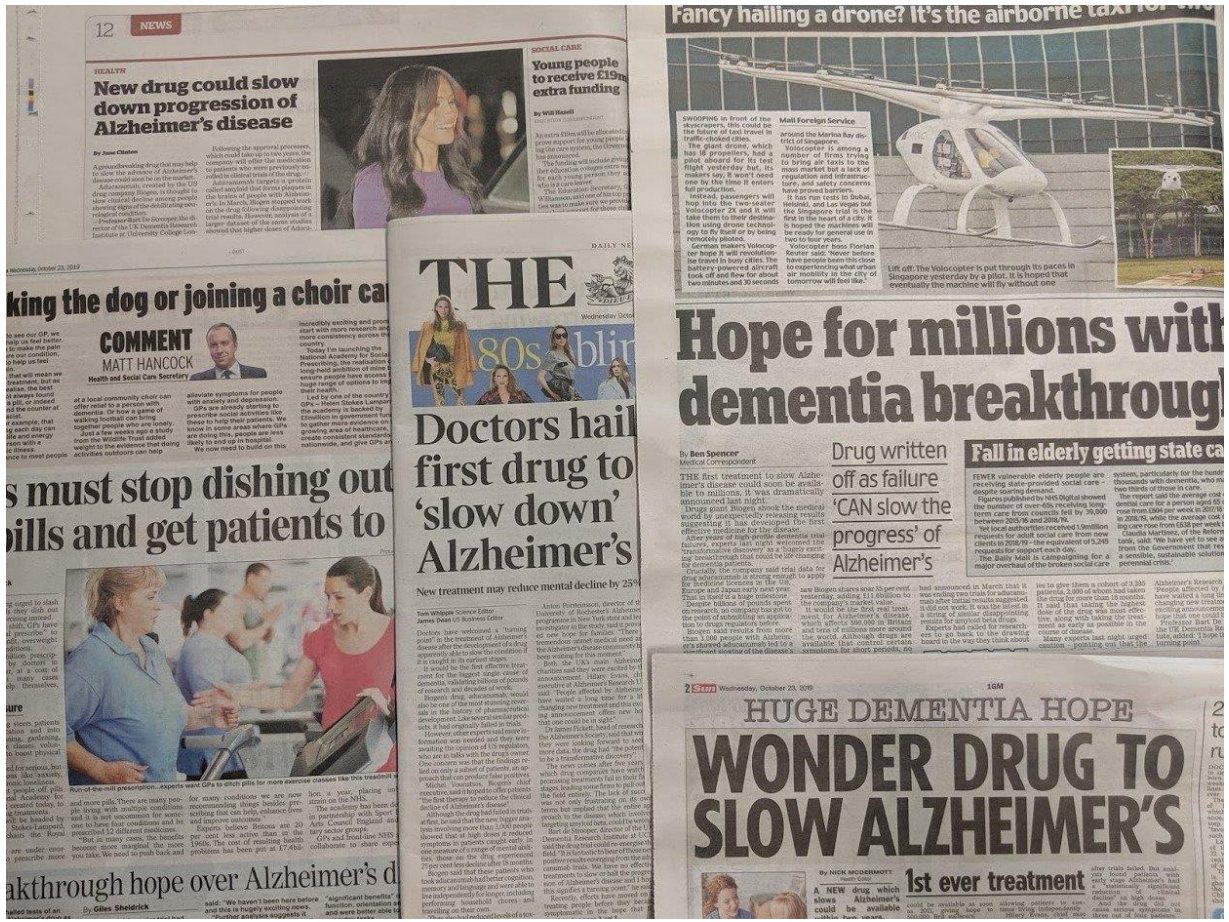


Clinical AD score (from 0 to 18)



New Engl J Med

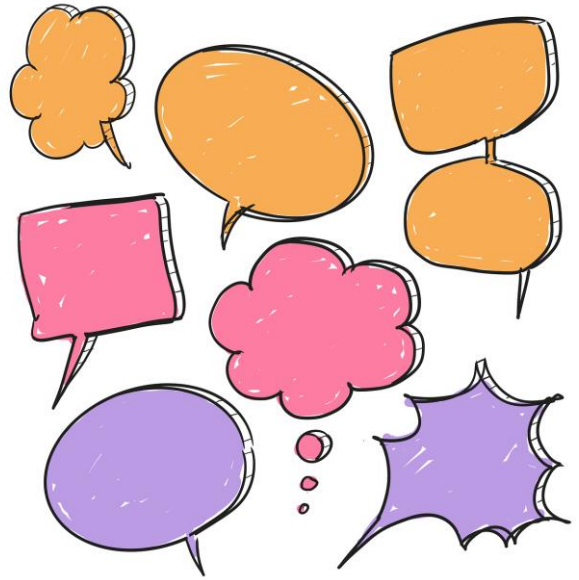
Watch out for the hype of “wonder drugs”



Researchers and physicians comment..

- “The amyloid hypothesis is dead…It’s a very simplistic hypothesis that was reasonable to propose 25 years ago. It is not a reasonable hypothesis any longer.”

- “We’re flogging a dead horse…There’s no sign of anybody getting better, even for a short period, and that suggests to me that you have the wrong mechanism.”



Nature, 2016

What went wrong?

1. Objective misalignment
2. Hidden variables and unmodeled state
3. Non-stationarity in individuals and populations
4. Feedback and adaptation

Objective misalignment

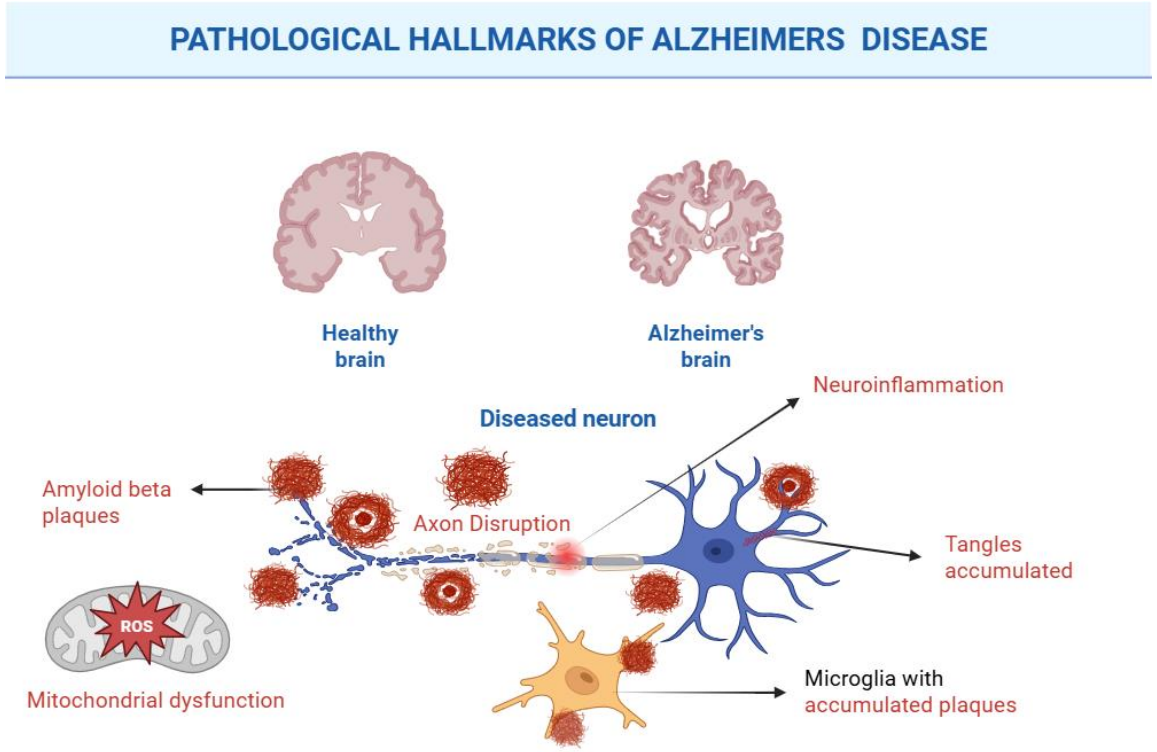
- Researchers optimized what was measurable, not what mattered.
- Amyloid was a convenient surrogate - not an actionable driver of disease

- ML analogy:
- This is optimizing a benchmark while deployment is different.

Hidden variables and unmodeled state

- The system had latent variables researchers collapsed or ignored.
- Examples of latent variables:
  - Tau pathology
  - Neuroinflammation
  - Vascular dysfunction
  - Aging dynamics

- ML analogy:
- Researchers trained a model with missing confounders and expected causal control.

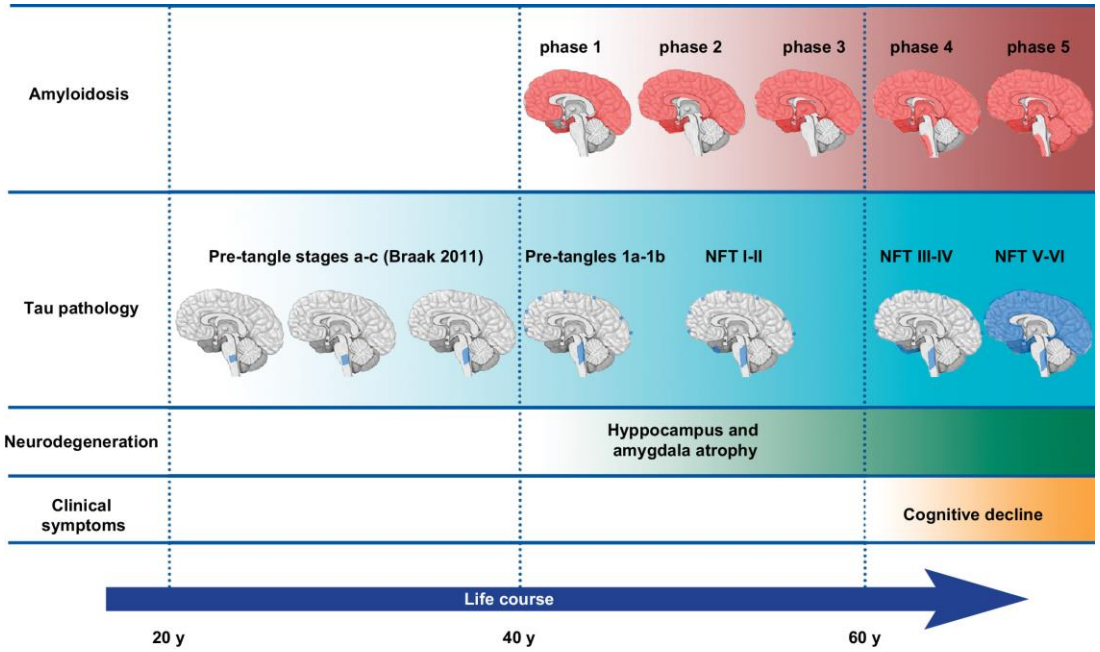


Images: BioRender

Non-stationarity

- Alzheimer's is not one disease but rather a trajectory
- Early disease and late disease are different systems

- ML analogy:
- Training and test distributions were not the same



Timecourse of Alzheimer's Disease

Feedback and adaptation

- Biological systems respond to intervention
- Removing amyloid changed immune signaling and neural networks
- System is not static

*The cost of **oversimplification***

- When we force a complex disease into a single-target frame, we lose:
- **Control** (correlation ≠ causation)
- **Robustness** (systems reroute)
- **Generalization** (patients are heterogeneous)
- **Explanatory power** (why the drug fails)

- None of these are exclusively pharmacology problems.
  They're also modeling problems.

Opportunities for integrateing complexity in modern AI for DD

Reductionism asks

- What happens if I inhibit this protein?

Complexity asks

- What system state am I moving in and where does it settle?

Three ways forward

- Better state inference (richer observables)
- Better interventions (multi-node control)
- Better objectives (state trajectories, not snapshots)

Financial aspects of Drug Development

- Drug Development: a sequential decision pipeline
- Financially, the drug development pipeline it has a modular structure to sequentially de-risk assets

MODULAR DRUG DISCOVERY AND DEVELOPMENT

Each stage can be funded, executed and monetized independently

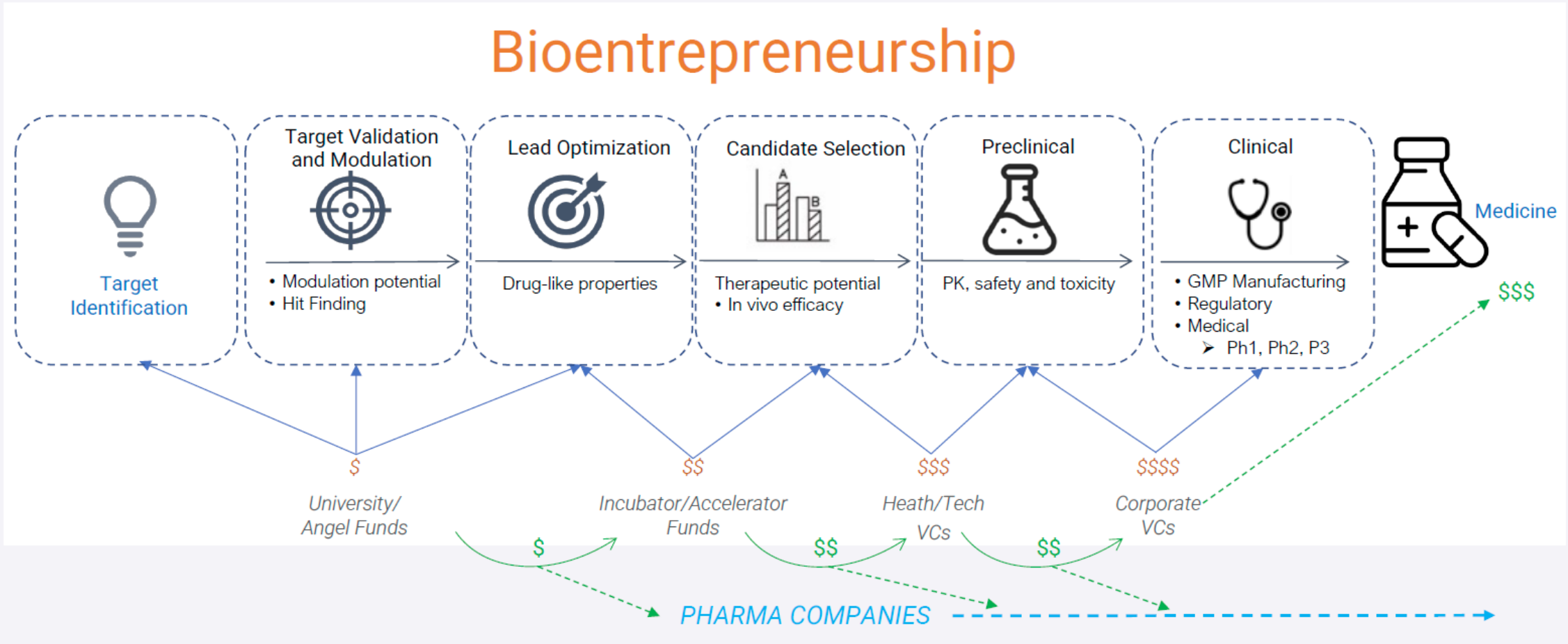


Image: A. Phinney

BIOENTREPRENEURS

Value of Early-stage Partnerships and Deals

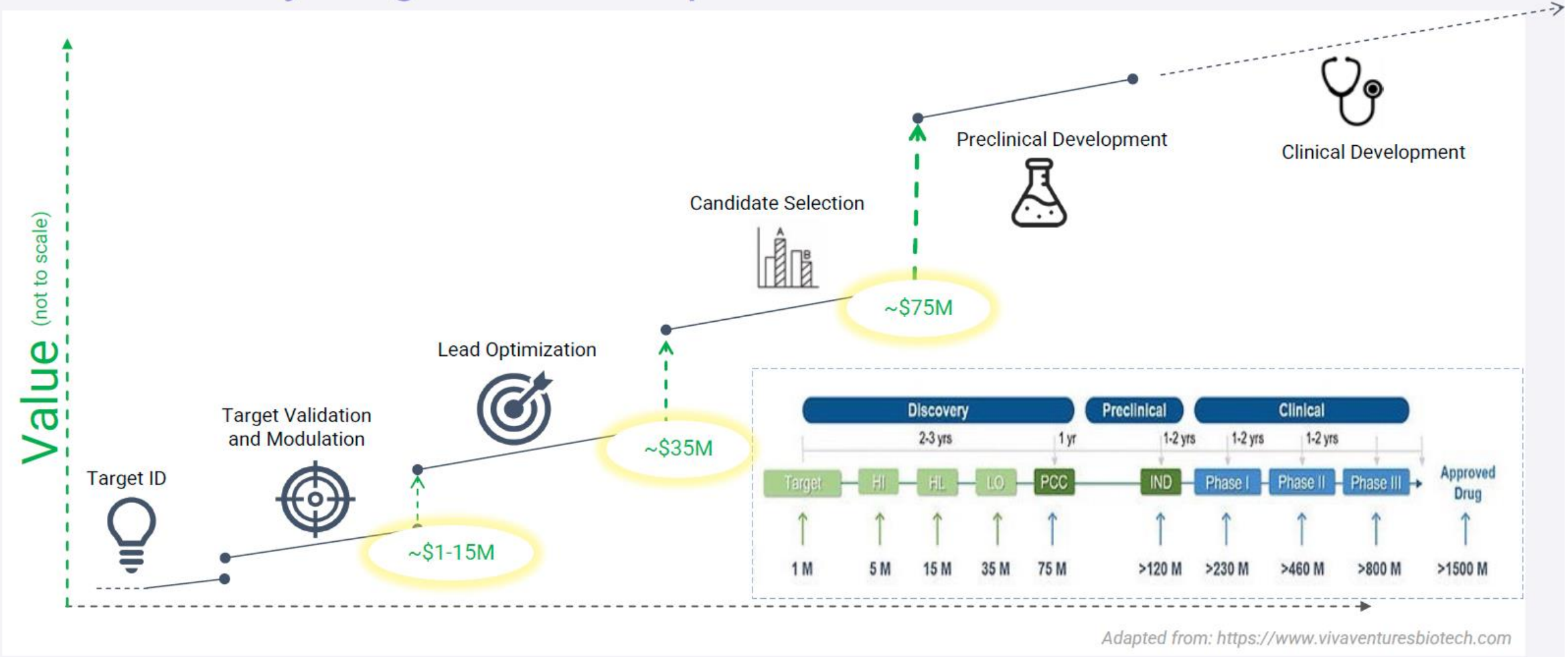


Image: A. Phinney

UNDERSTAND VALUE DRIVERS- THE SCIENCE

Align the stage of the science to the stage of the business

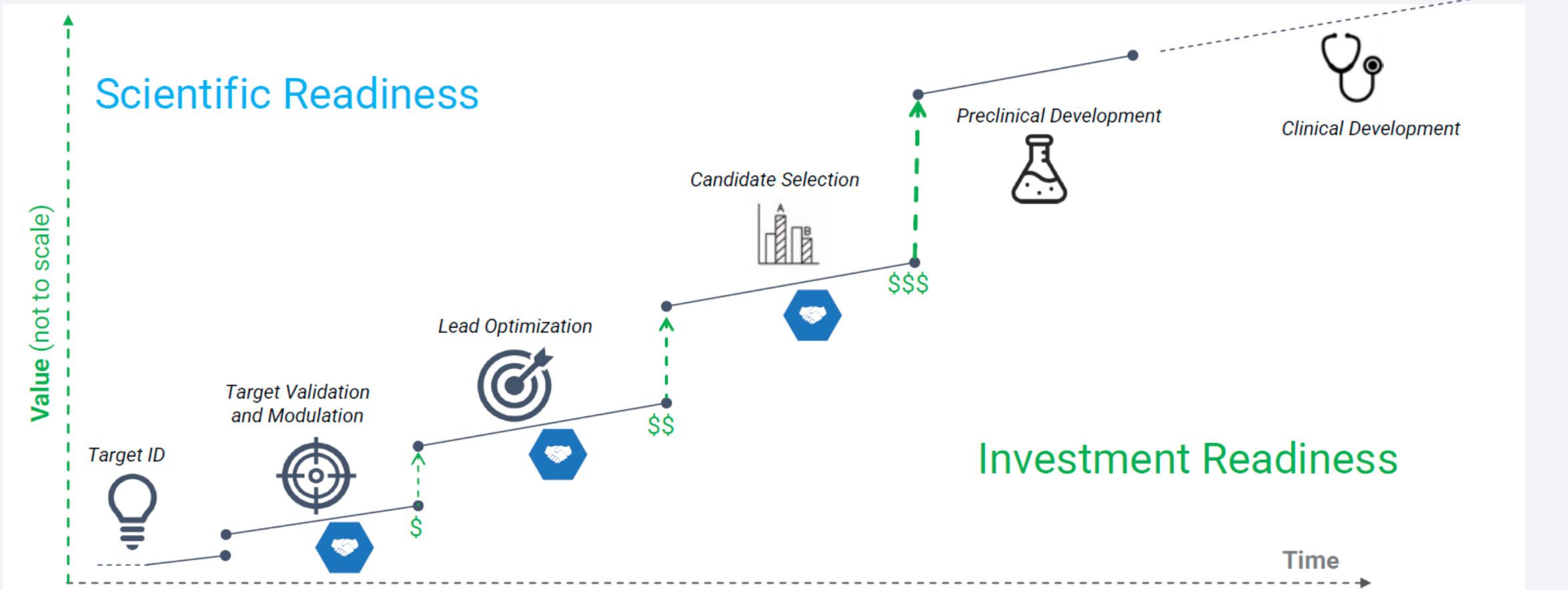


Image: A. Phinney

UNDERSTAND VALUE DRIVERS- THE SCIENCE

Align the stage of the science to the stage of the business

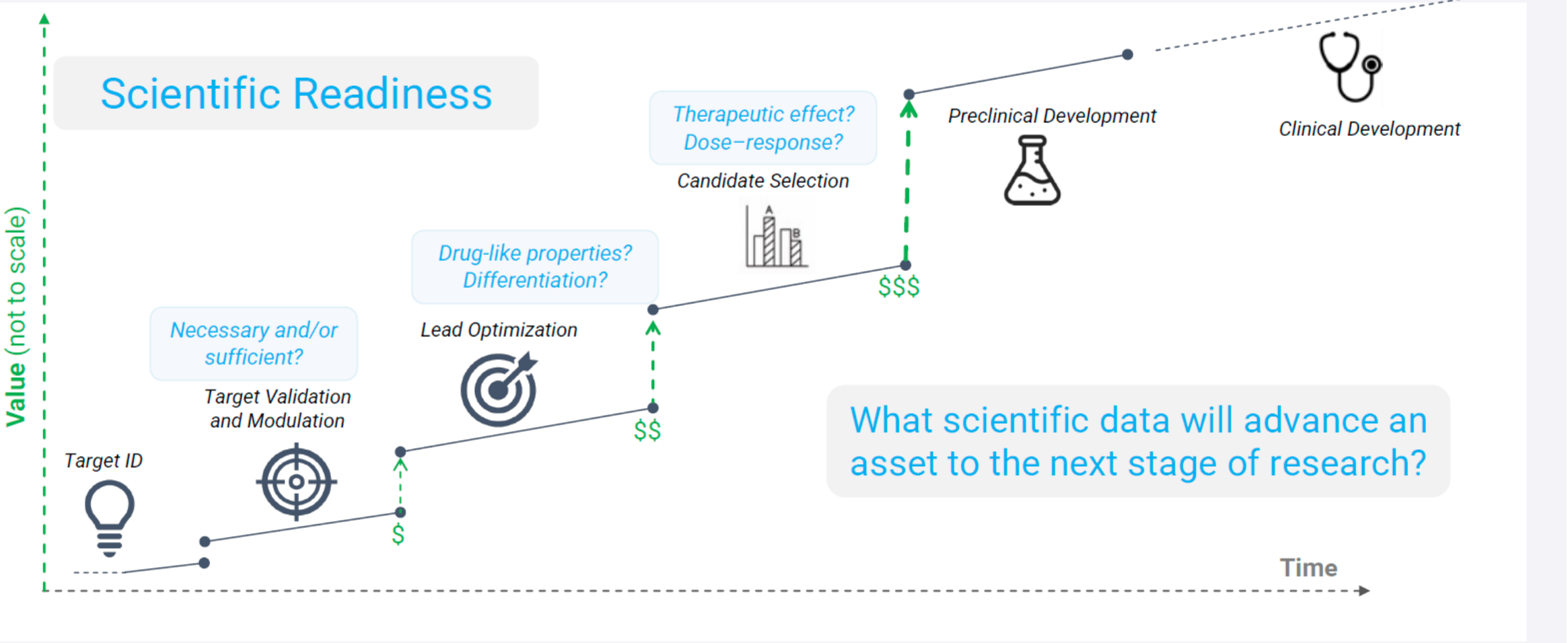


Image: A. Phinney

POLARIS
MARKET RESEARCH

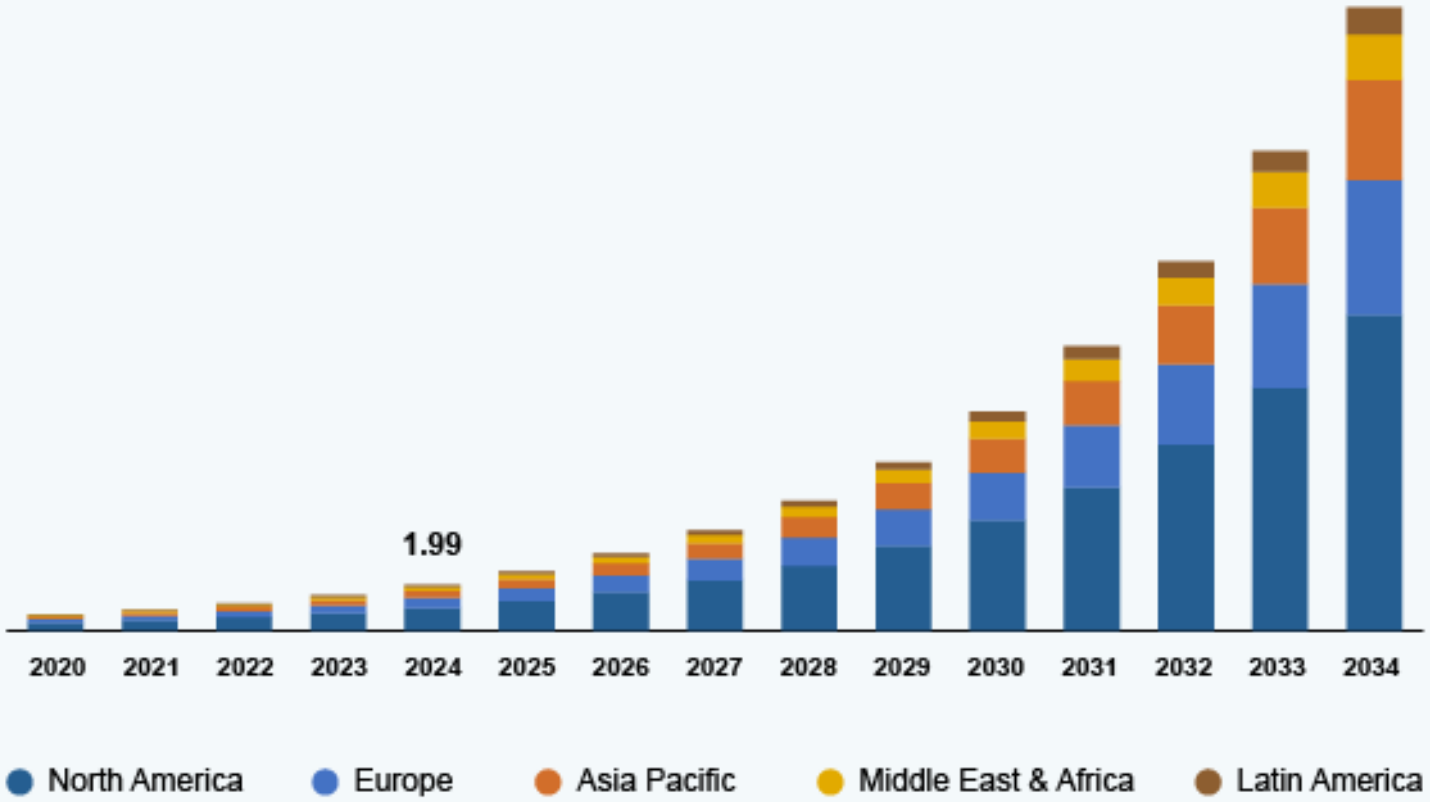
AI in Drug Discovery Market

Size, By Region, 2020 - 2034 (USD Billion)

The Global AI in drug discovery market is estimated to reach 35.42 billion by 2034

29.6%

Global Market CAGR
2025-2034



Source:www.polarismarketresearch.com

Note: The images shown are for illustration purposes only and may not be an exact representation of the data.

Why Drug Development Needs You