AI in Drug Discovery 2022 – Aspects of Validation, Data, and Where We are on the Hype Cycle
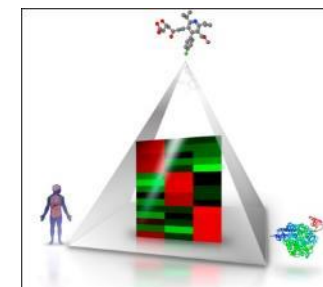
Andreas Bender, PhD
Natural Philosopher for Molecular Informatics
Department of Chemistry, University of Cambridge

Chief Scientific Officer, Terra Lumina, Berlin

Co-Founder of Healx, Ltd. and PharmEnable, Ltd.

UNIVERSITY OF CAMBRIDGE

Any statements made during this talk are
in my capacity as an academic

Further reading: Artificial Intelligence in Drug Discovery – What is Realistic, What are Illusions? (Parts 1 and 2)

Andreas Bender and Isidro Cortes-Ciriano

Drug Discovery Today 2021

Contents

1. Current state of AI in drug discovery

2. How do we know that a method *works*? What is 'validation'?

3. The Achilles heel of AI in drug discovery: data & proxy assays

4. Psychology, the hype cycle & the translational gap of methods

5. OK… and now?

1. Current state: The 3rd wave of computers in drug discovery (80s, 2000, today) – time for realistic assessment has come

Fortune cover 1981

Recent headlines (2018-2020)

Funding going into AI in drug discovery 2021: ~$4bn VC funding, $16bn total
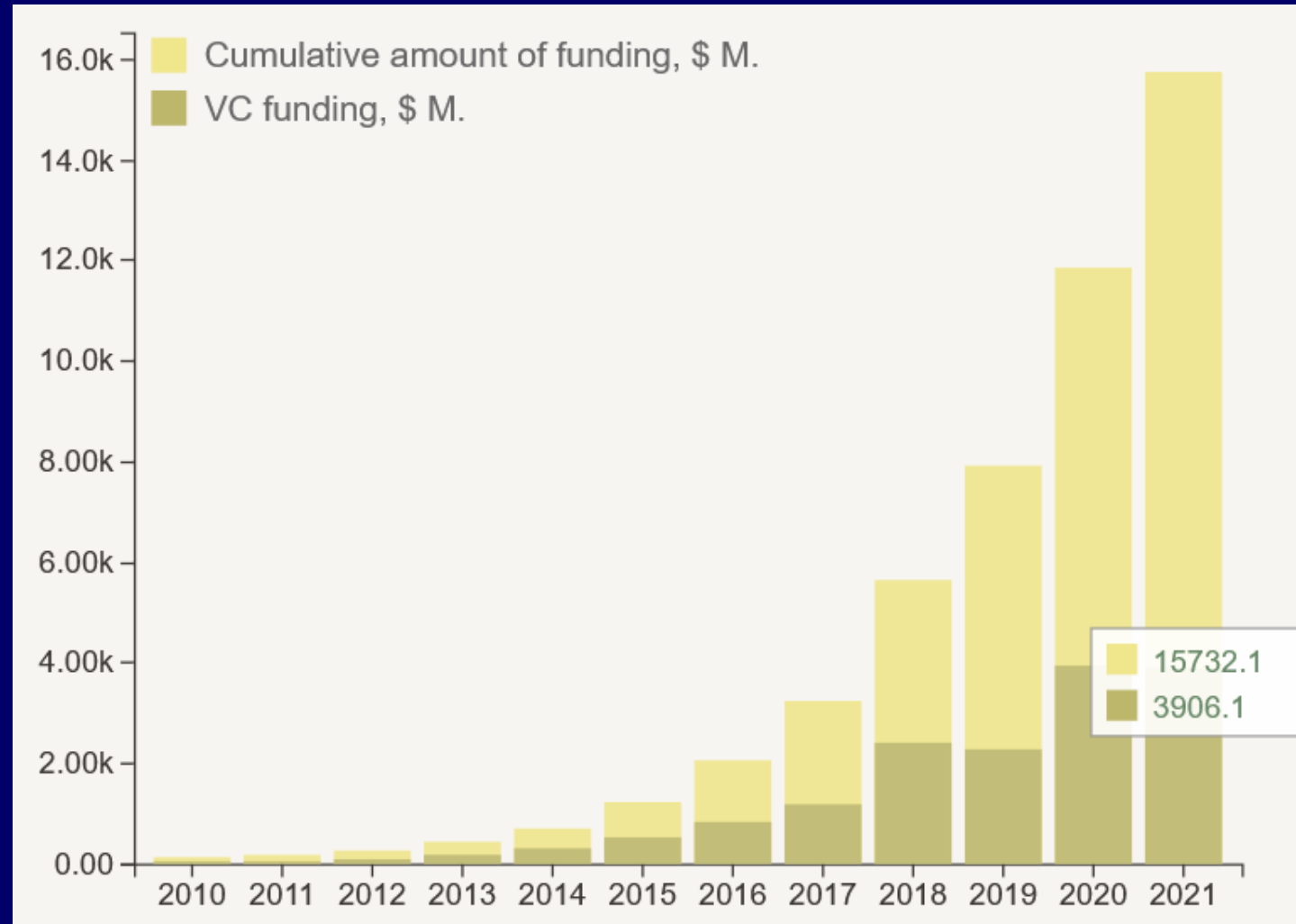
Current discovery pipeline: AI-based start-ups vs big pharma

'AI-native companies'   Top 20 pharma
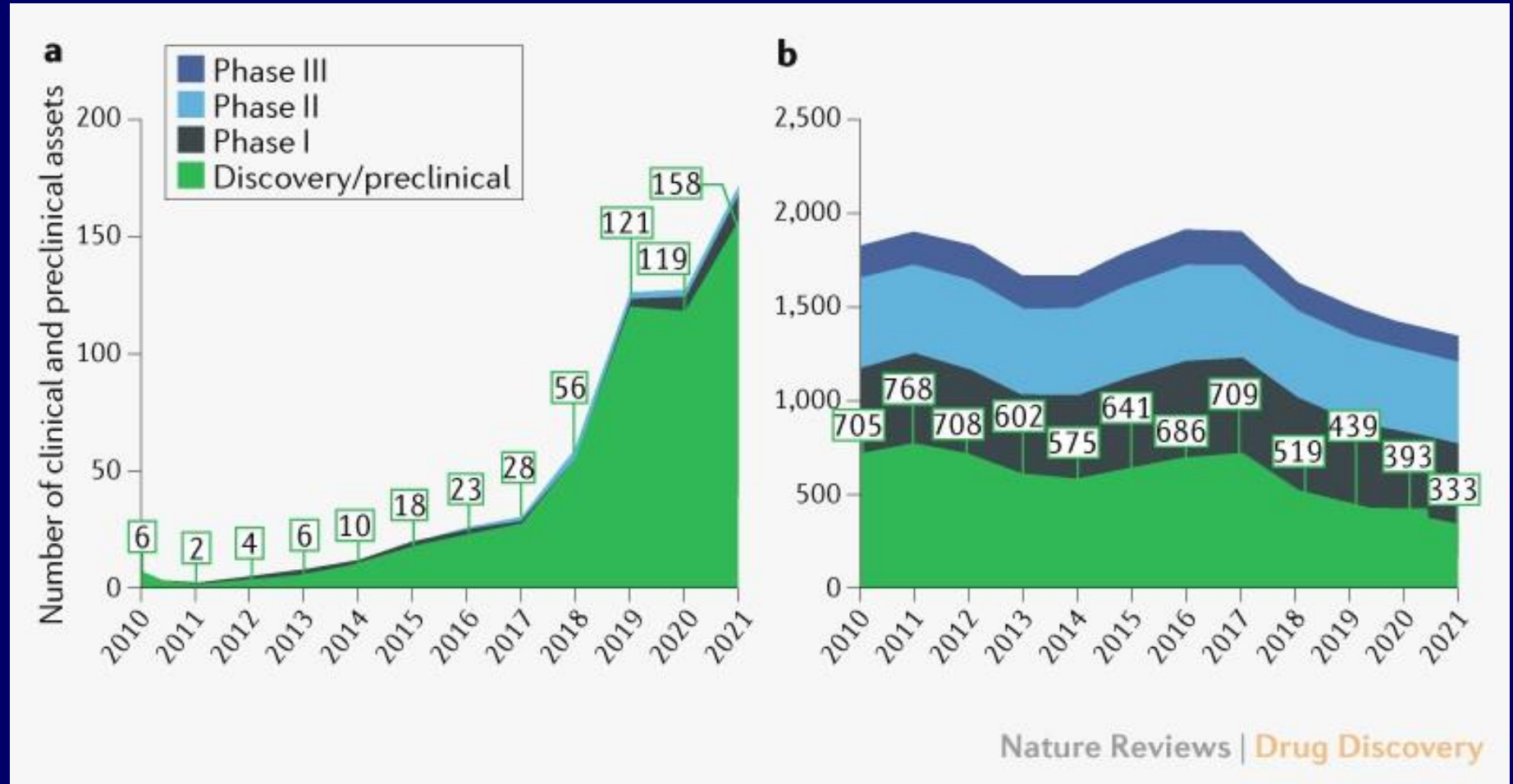
Significant *number* of *discovery/ preclinical* programs of AI companies (~160 vs ~330)

Very little Phase 1, less Phase 2, no Phase 3

Nature Reviews | Drug Discovery

-> Little *in vivo* safety (Phase 1) data *yet*; virtually no *in vivo* efficacy (Phase 2/3) data *yet*

Distribution of target profile similar, but focus on areas of more data, less complex target pharmacology

More kinases and enzymes in AI-driven companies:
(a) Quite *data-rich*
(b) Less complex pharmacology than other target classes

+ Transcription factors
- No ion channels, NHRs and transporters
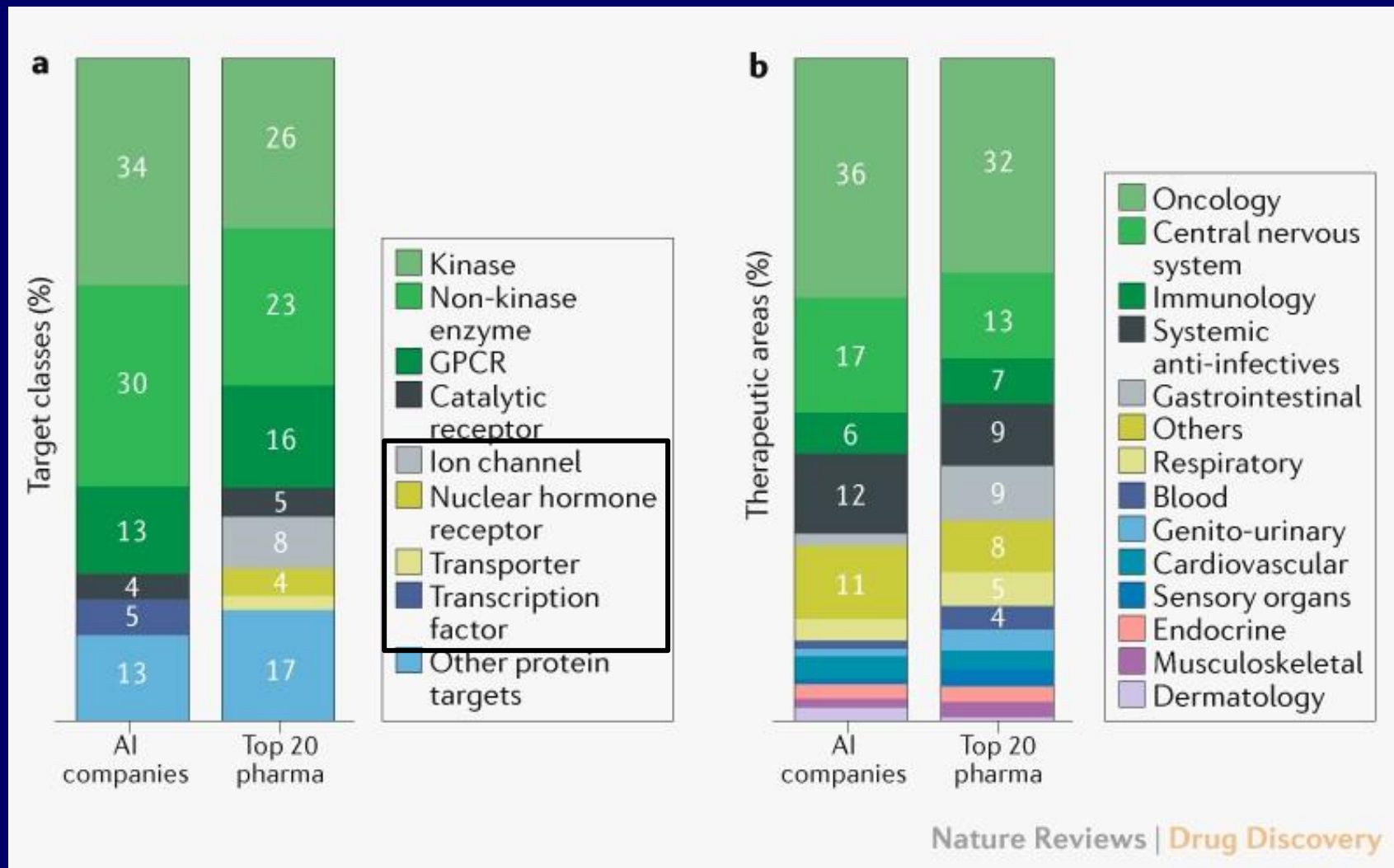
Little (but useful?) experimentation on chemistry level
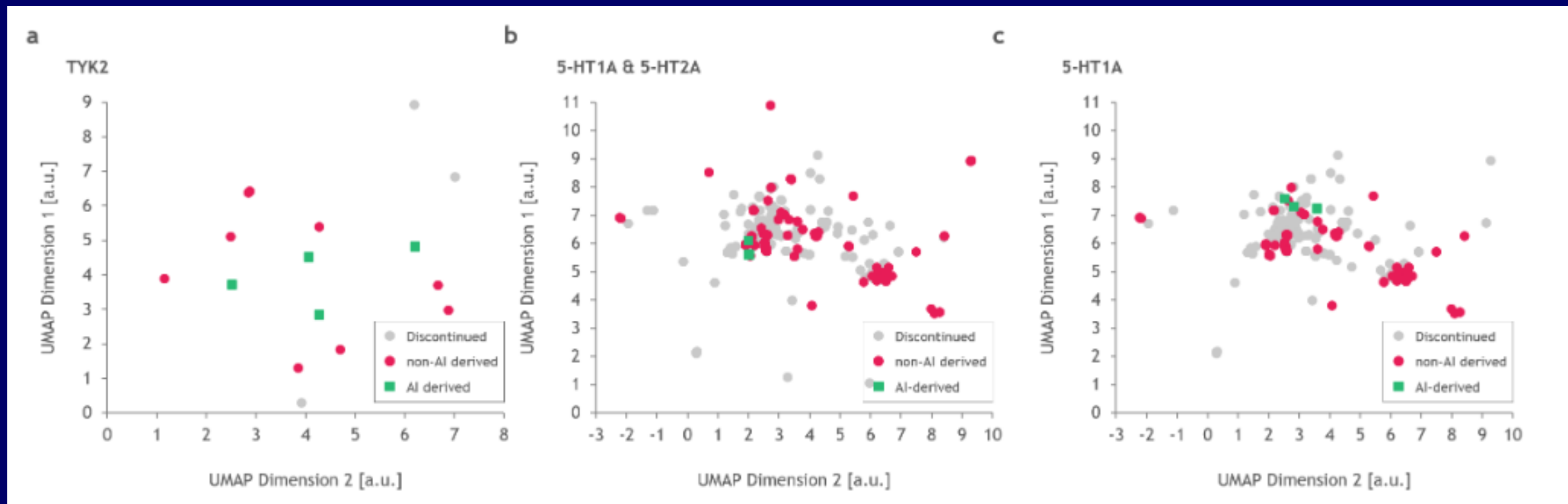
- Red: Non-AI derived; green AI-derived; grey: discontinued
- Relatively little chemical novelty; *but* sometimes superior selectivity
- *Be careful what you interpret into UMAP plots*, chemical space is high-dimensional; but when looking at structures you will come to similar conclusions

Conclusion about the world as it is

- Lots of activity in early stage pipeline of AI-first companies, but often already explored targets, close analogues
- Appropriate question to ask: Where is the novelty?
- Data is often limiting factor – in both chemical and target space (leads to work on well-explored targets, with more data, less complex pharmacology)

- Is *input* (e.g. funding) success, or *output*?
- The first 'AI-designed drug' will be celebrated by the media, but…

… tens of billions went into funding AI in drug discovery, so even the null model would lead to an expected tens of approved drugs
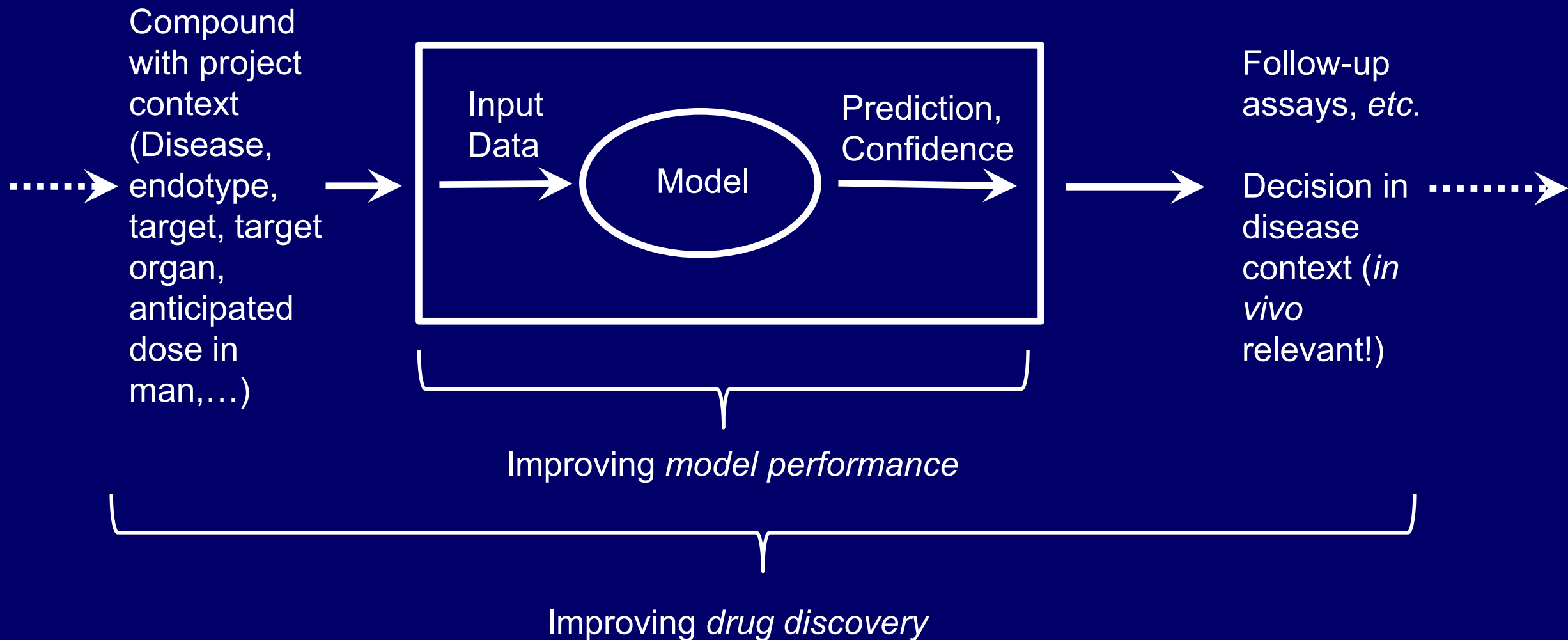
2. How do we know that something *works*? What is 'validation'?

- Core question in science, core question for start-ups

- In *theory* we establish a method, use a benchmark, and know how well the method works

- In *practice* this doesn't really work in *drug discovery* –

  - Labels are either mostly only *in vitro*-relevant, or conditional ('depend' on dose, *etc*)

  - Validation is costly (phase II studies for efficacy; *plus controls*), so *little prospective data*

  - Difficult to sample distribution in chemistry/'project' space well (diversity, number), so performance *depends heavily on test set*

- Retrospective validation is equally futile (no prospective discovery, predictivity for future projects unknown, all behave differently)

- *Core reasons for problem: In chemical space proper sampling impossible, underlying distribution unknown; conditionality of in vivo data*

What to watch out for in validation – and why the model, *embedded into the process* matters

- 'Proof by example' abounds, without baseline

- Irrelevant endpoints abound (numerical improvements, endpoints that don't directly translate into *in vivo*-relevant decision making)

- *Validation that matters* includes the *process and not only the model* in the validation (!)

- Further discussion of model validation in my blog:

    http://www.DrugDiscovery.NET/HowToLie

- Nature Reviews Chemistry article on 'validation' appearing shortly

Model validation vs process validation (e.g. ligand structure-based property predictions)

Compound with project context (Disease, endotype, target, target organ, anticipated dose in man,…)

Input Data

Model

Prediction, Confidence

Follow-up assays, *etc.*

Decision in disease context (*in vivo* relevant!)

Improving *model performance*

Improving *drug discovery*

Conclusion: So *did* AI contribute something to drug discovery?

- Probably in some areas yes (e.g. target prediction, digital pathology, …), but very difficult to quantify related to *process*
- After ~$20bn VC funding into AI in drug discovery and ~$50bn total funding we better see some successes!
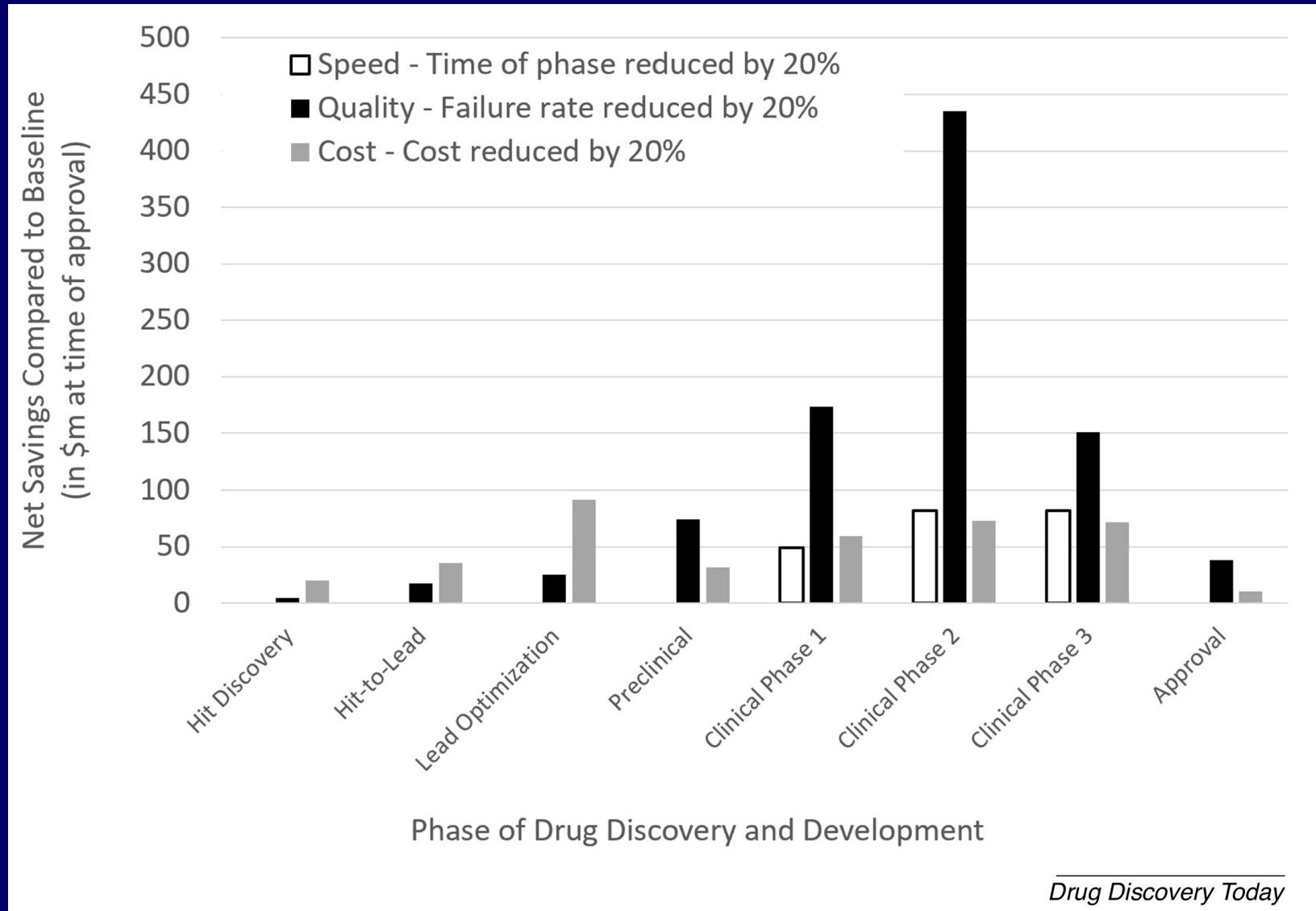

… and to round off this section on validation:

Q: "It works in practice, but does it work in theory?"

On the other hand of course: "The difference between theory and practice is bigger in theory, than in practice"

3. The Achilles heel of AI in DD: Data and proxy assays

"…it's the data, stupid!"

The *quality* of *in vivo-relevant* decisions matters more than *speed* and *cost*!

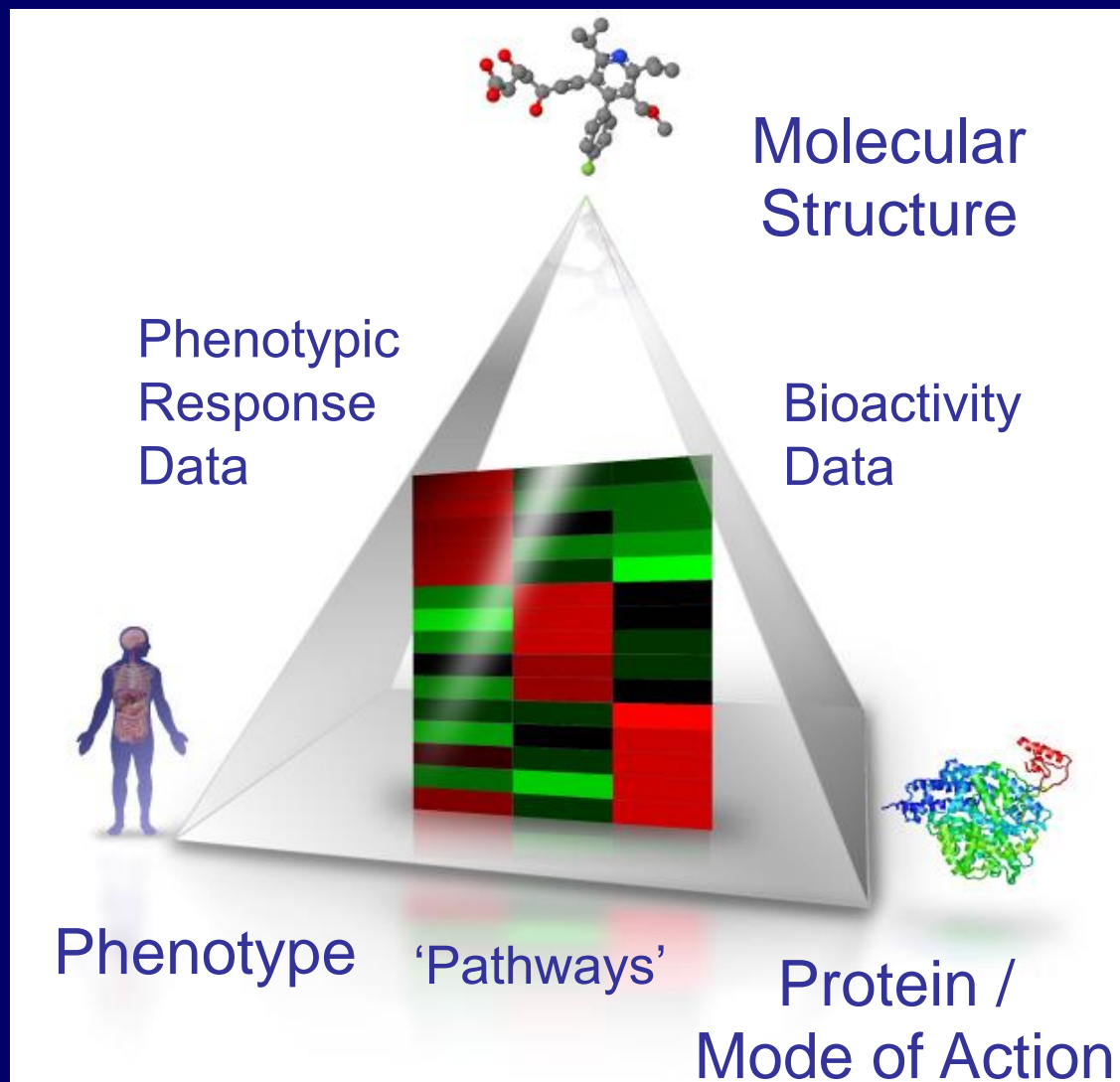Bender and Cortes, Drug Discovery Today 2021

In vivo-relevant decisions matter most!
But… is this where our *data* for models is?

- Chemical and biological data: The flat-earth (~'in vitro') view
  - And where a flat earth is great!
- Chemical and biological data: The round-earth (~'in vivo') view
  - Drug discovery data and its complexity (... the elephant in the room…)

- Why algorithms from image and speech recognition don't really translate to *drug* discovery

A simple view on the world: Linking Chemistry, Phenotype, Targets / Mode of Action (myself, until *ca.* 2010)

Molecular Structure

Phenotypic Response Data

Bioactivity Data

Phenotype   'Pathways'

Protein / Mode of Action

a.k.a. "The world is flat"

= "We believe our labels"

"Compound A is toxic", "Compound B binds target X", "Compound C treats disease Y", …

Works in cases where data is large-scale, and homogenous, and we have meaningful labels

Does not consider data conditionality, e.g. dose, PK, translatability from model system to *in vivo* setup, endotype, genotype, *etc. etc.*

BUT…The world is not flat. What now?

- Links between drugs/targets/diseases are quantitative, incompletely characterized

- Subtle differences in eg compound effects (partial vs full agonists, off-targets, residence times, biased signalling, etc.)

- 'Pathways' from very heterogenous underlying information; dynamic elements not captured etc.

- Effects are state-dependent (variation between individuals, age, sex, co-medication…) – PK is often rather neglected in AI approaches

- Phenotyping is sparse, subjective (deep phenotyping?)

- We don't understand biology ('the system'), we don't know what we *should* label, and measure, hence …

- We label what we *can* measure: 'T*echnology push*' vs '*science pull*' (!)

- **Are our labels – 'drug treats disease X', 'ligand is active against target Y', … - meaningful?**

- **Conditionality: Causality, confidence, quantification, ….?**

- **Computer science is tremendously powerful… but is our data?**
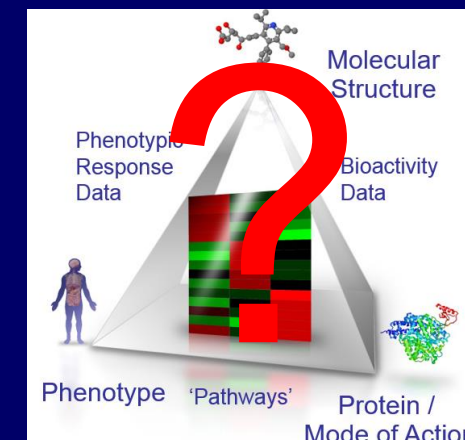
Are our understanding and data good enough? The many facets of ketamine
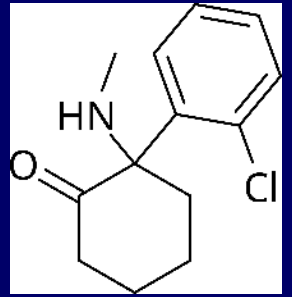
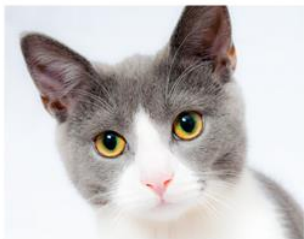- Ketamine both used as (rather safe) anaesthetic (iv 2mg/kg), approved since 1970, as well as a street drug

- In 2000 effect as antidepressant, when dosed significantly lower, also bronchodilator (acute asthma); iv 0.5mg/kg

- Ketamine long been thought to act via blocking the NMDA receptor - *but* other NMDA blockers such as memantine and lanicemine have not been successful in clinical trials (as antidepressants)

- Also the opioid system implicated in action of ketamine (naltrexone/opioid antagonist influences its effects)

- Furthermore, a metabolite of ketamine has recently been found to be active in animal models of depression

- … *etc. etc.* (disease endotype, co-medication, accumulation, …)

- **If it's not in the data (or hidden by conditionality!), it won't be in the model!**

Das, J. Repurposing of Drugs–The Ketamine Story. *J. Med. Chem.* 2020 (ASAP Article)

Example of conditional labels: adverse reactions

- *"Does drug Y cause adverse reaction Z? Yes, or no?"*
- Pharmacovigilance Department: Yes, *if* we have…
  - A patient with this *genotype* (which is generally unknown)
  - Who has this *disease endotype* (which is often insufficiently defined)
  - Who takes *dose X* of *drug Y* (but sometimes also forgets to take it)
  - With known targets 1...n, but also unknown targets (n+1…z)
  - Then we see *adverse reaction (effect) Z* …
  - But only in x*% of all cases* and
  - With *different severity* and
  - *Mostly if co-administered with a drug from class C*, and then
  - More frequently in *males* and
  - Only *long-term*
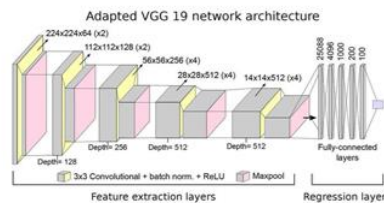 - (Etc.)
- **So – does drug Y cause adverse event Z?**

Object

Representation

Model

Object Label

Image Domain

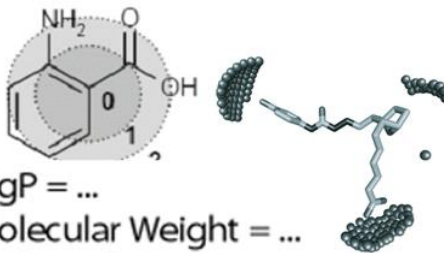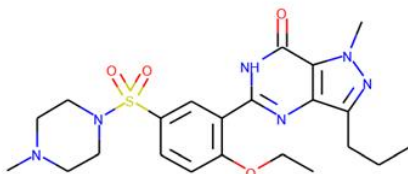Adapted VGG 19 network architecture

ResNet?
AlexNet?
CapsuleNet?

Cat

Largely *Unconditional* labels

Representation and model are *intrinsically linked* (ie, model uses native object representation by pixels)

Drug Discovery: Chemical Domain

?

logP = ...
Molecular Weight = ...
Molar Refractivity =

?

Artificial Neural Network/DNN? Support Vector Machine? Random Forest? Bayesian
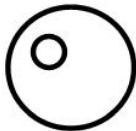
Property A

Conditional labels (eg dependent on assay system, genotype, dose, endotype, sex, age, comedications, lifestyle, ...)

?

1. Molecules *are* no graphs! You *can* use the connectivity table to derive *a representation* of it though, which in some cases can be suitable

2. Learned representations good *for large-scale, homogenous* data; but still suffer from *conceptual problem* of data conditionality in drug discovery, and lack of *in vivo-relevant* data
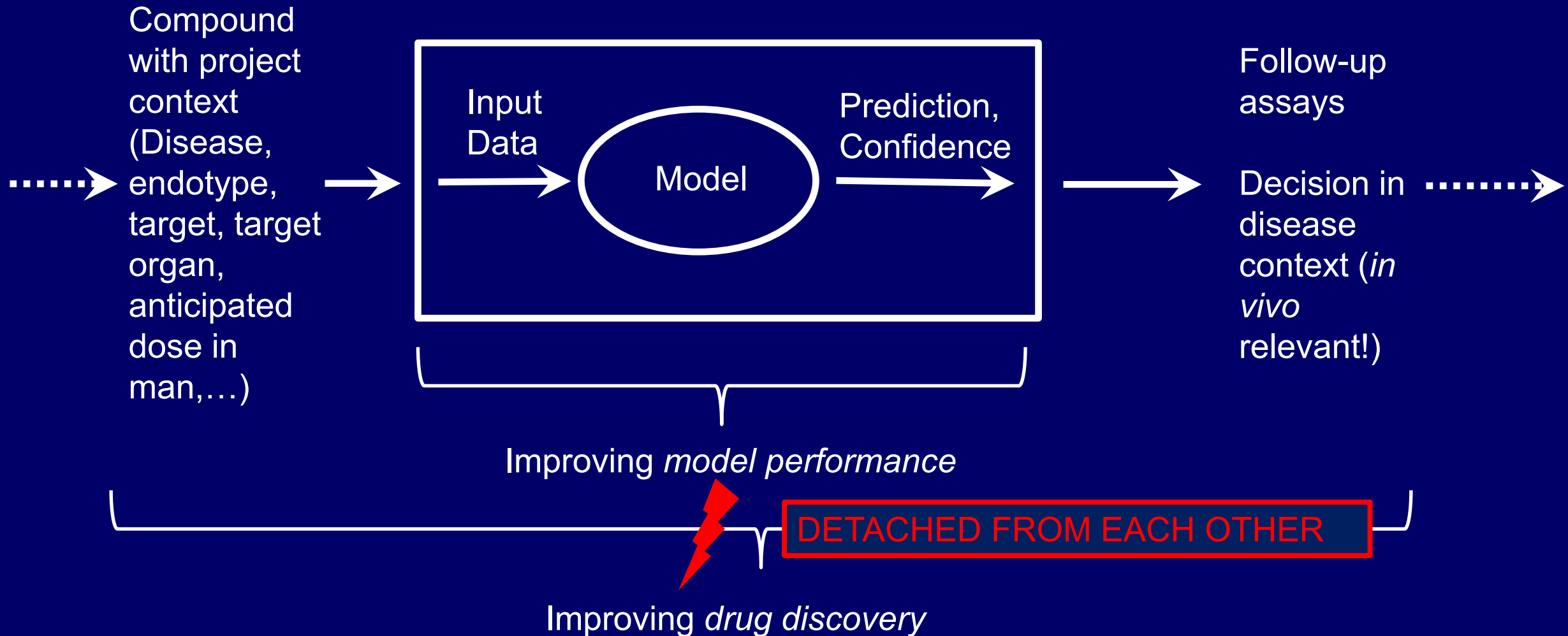
State/Effect B

Heavily conditional labels (eg dependent on genotype, dose, endotype, sex, age, comedications, lifestyle, ...)

?

Both representation and modelling approach are *largely trial and error* (in particular the information content of biological readouts has only been established for particular cases)

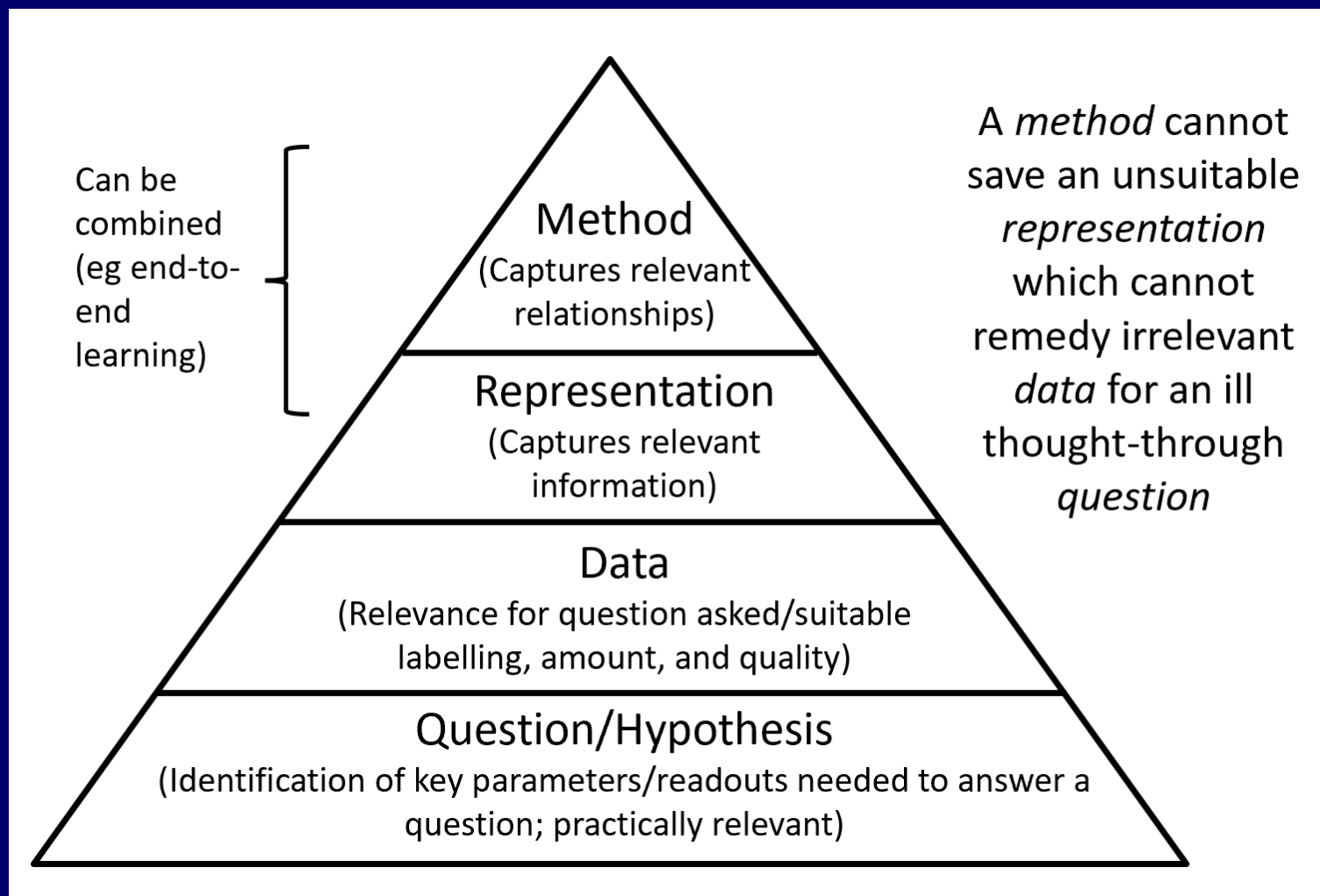Bender & Cortes
Drug Discovery Today 2021

Much of the data we have has been generated with proxy assays. Why is this a problem for AI in drug discovery?

- There is *what we are really interested in -* say, mitochondrial safety, Drug-Induced Liver Injury (DILI), …

- And there is what we *measure as an assay endpoint* – say, cytotoxicity in a Glu/Gal (differential cytotoxicity) assay to *approximate* mitochondrial safety; Bile Salt Export Pump (BSPE) inhibition to *approximate* DILI, …

- Take-away: 'Proxy' assays measure only part of reality, in a particular assay, with particular conditions

- Not to be confused with property itself!!!

- Problem: Proxy endpoint (a) taken as 'ground truth' in AI in drug discovery, (b) embedding into project context neglected
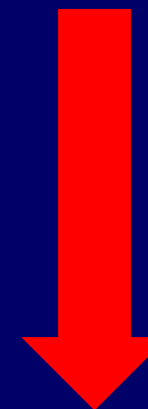
Why meeting the proxy endpoint (and any derived models) is neither sufficient (nor necessary!) for success in a drug discovery project

The *question* needs to come first… and then the data, then the representation, and then the method
http://www.DrugDiscovery.NET/HowToLie

Can be combined (eg end-to-end learning)

Method (Captures relevant relationships)

Representation (Captures relevant information)

Data (Relevance for question asked/suitable labelling, amount, and quality)

Question/Hypothesis (Identification of key parameters/readouts needed to answer a question; practically relevant)

A *method* cannot save an unsuitable *representation* which cannot remedy irrelevant *data* for an ill thought-through *question*

Lots of attention currently here…

But we need to care more about this

Bender and Cortes, Drug Discovery Today 2021

4. Psychology, the hype cycle and a methods translational gap

The bigger picture: 'AI' is where it is due in no small part due to human psychology

- Hype brings you money and fame – realism is boring
- FOMO ('the others also do it!') and 'beliefs' often drive decisions ('maybe they *really* have the secret sauce?')
- 'Everyone needs a winner' ('*after investing X million we need to show success to the CEO/VP/our investors/…*')
- Selective reporting of successes leads to everyone declaring victory (but in reality no one knows what's actually going on)
- Difficult to really 'advance a field' with little real comparison of methods

- NB: Multiple levels, individual psychology, as well as organizational psychology matter

AI on the Hype cycle (Gartner, 2021)
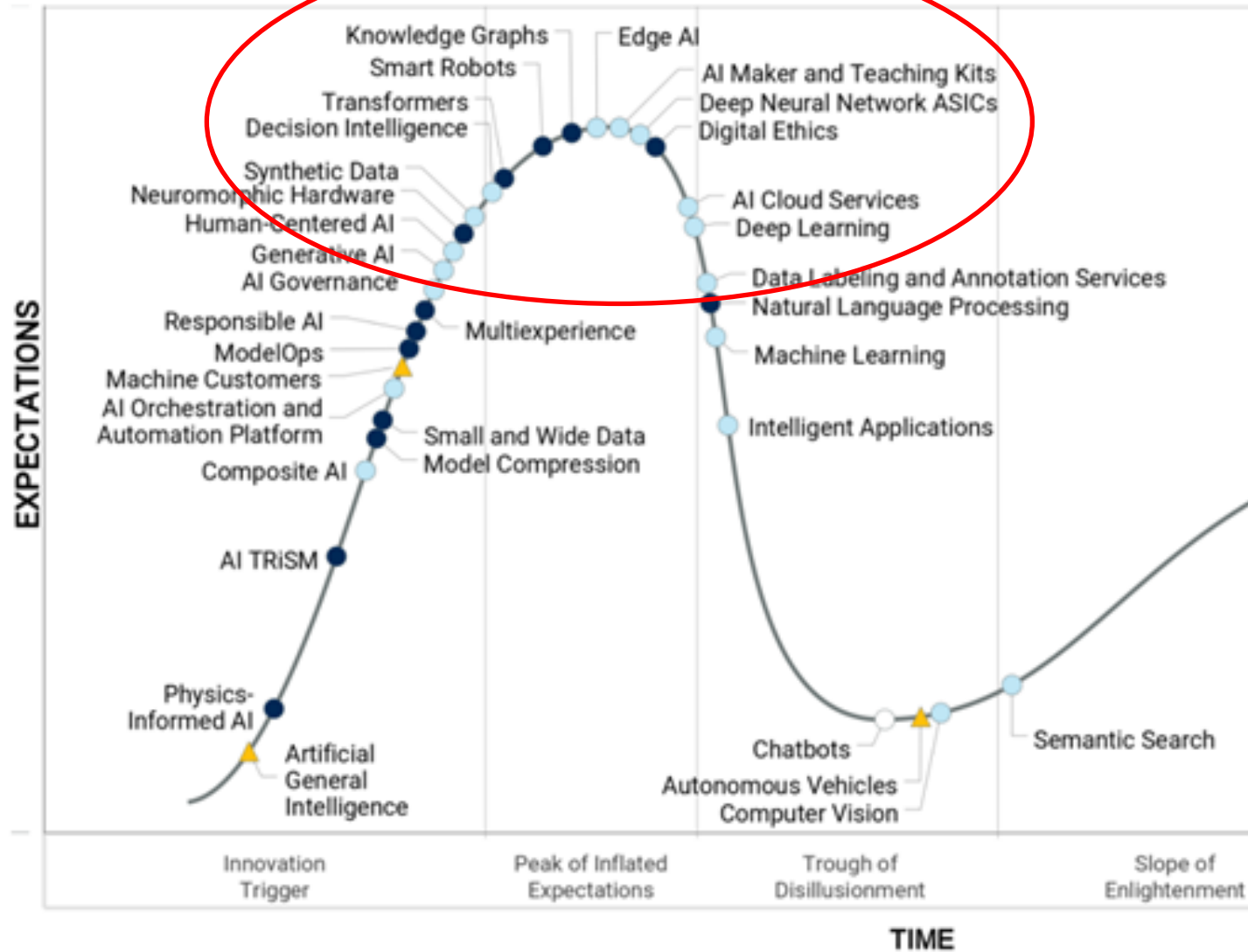
Notes:
- Y axis are expectations, not 'results'
- *Does not exist in this form*, only in *perception, with huge spread in the details*
- Agree with general place; *but aspects* clearly working (DL for images, ML for target prediction, cloud services useful in practice, etc etc.)
- Near future will further explore applicability of given method in a given context

On the disconnect between research/academia and drug discovery

- Current translation (and translatability) of research into drug discovery is relatively meagre

- Different objectives of 'publishing a paper' (methods-related), and 'practical impact' (process-related)

- Directly linked to disconnect between *model validation* and *process validation*

- *In silico* models need to be embedded into decision making pipeline; not 'AUC of 1' but success in real-world projects needs to be the goal

- Consortia (including experimental design) needed for *relevant benchmarking*

- Some coming (e.g. CACHE, for ligand identification), but needs more activity to really advance the field
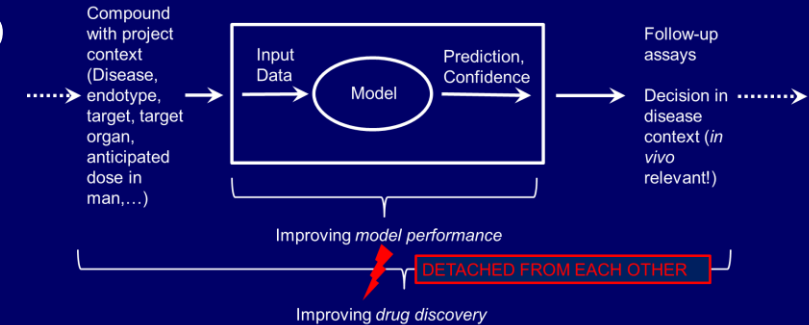
My look into the crystal ball (a few days ago…)

- Q1/2022: Inflation increasing (e.g. UK in 2021 5.5%)
- Central banks increase interest rates (money gets more expensive); pressure on asset prices; Ukraine war; …
- Return of the safe haven (gold etc.) within 1-3 (?) years

-> Less VC money available in the system

- *AI in drug discovery needs to deliver soon (in the next ~2-3 years?)*
- If you are a start-up, get funding into place soon

5. Ok… and now?

- We need *relevant* data (predictive for the *in vivo* situation), which is *possible to generate large-scale*
  - 'omics data: *Yes, but* experimental conditions (e.g. cell line)/dose/time point often don't extrapolate to relevant situations
  - Cellular morphology data: *Yes, but* we need to understand better what the applicability domain is/which interventions are visible in the readout
  - Organ-on-a-chip: *Yes (!), but* still under heavy development, details to be seen

- Probably industry-wide precompetitive consortia *involving experimental design and data generation* needed to establish best-in-class approaches across endpoints
- Required due to (a) large size of chemical/mode of action space, (b) high number and dimensionality of readouts that can be generated, and (c) large number of *in vivo* endpoints we are interested in

… there are always multiple ways to claim a 'win':

- Scientifically (broad, *meaningful* benchmarking);
- Using individual success cases; or
- Economically ('I found someone who bought my stuff'… which is mostly psychology)
- …

Summary

- We need to analyse our data (as we did for many years before), absolutely!

- 'AI' *is a valuable tool* in the toolbox

- The *real* game changer for translation to patients will come only once we understand biology/biological data better (and generate it, and encode it, and analyse it)

- From the data side, consortia on even larger scale are needed (for targeted data *generation*, not just sharing what is there already)

- Methods need to *translate into reality*, we need to go *from model validation to process validation*

Thank you for listening!
Any questions?

Contact: ab454@cam.ac.uk
Personal email: mail@andreasbender.de
Web: http://www.DrugDiscovery.NET
Twitter: @AndreasBenderUK