Mathematical and Computational Biology In Drug Discovery (2022)

Dr. Jitao David Zhang

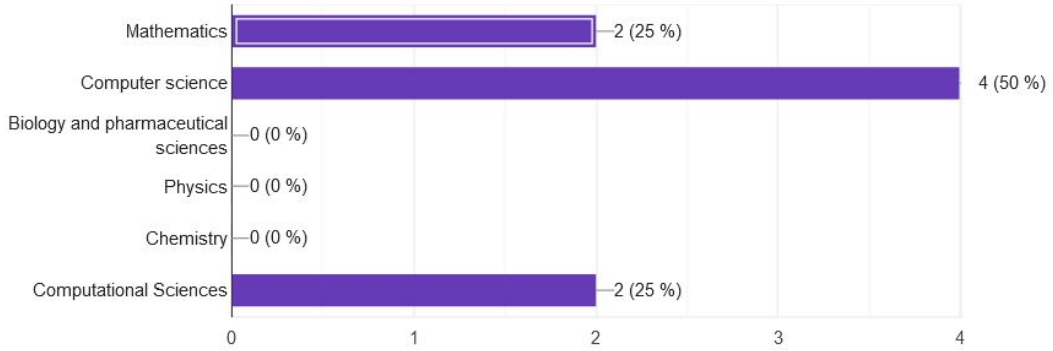
[1] Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

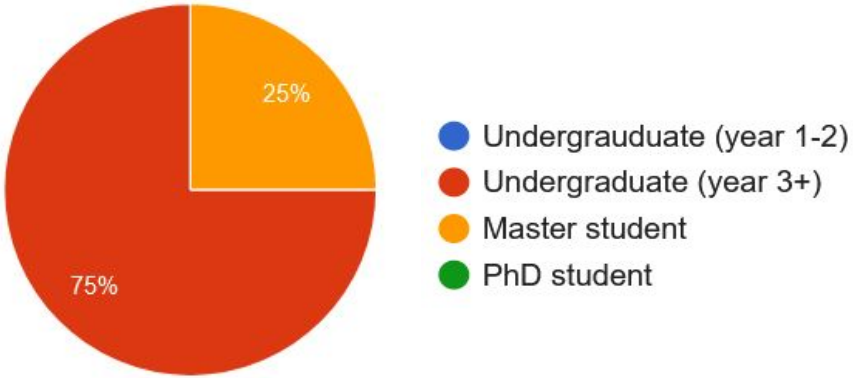
[2] Department of Mathematics and Computer Science, University of Basel

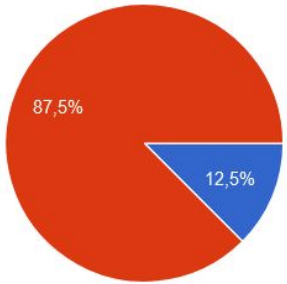
Administrivia

- **Last chance to fill the pre-course survey!**
- **Grades are given by offline activities (50%) and projects in teams of two (50%)**:
  - **Option 1**: Write a target (or screening) proposal for a disease of your choice, using publicly available data and tools/algorithms that we learn about to support your arguments.
  - **Option 2**: Write a report analysing data from the *Drug Central* database, raising your own scientific questions about drug-target associations and answering them with analysis.
  - Grade is given both by peer review and by the lecturer.
- I hope the course is a seminar more than a lecture: share your question and let's discuss!
- Attendance is preferred, however asynchronous learning is also encouraged if necessary.
- For catching-up students: let me know if you need access codes for AMIDD videos.
- **Any more questions?**

We have a great mix to learn from each other







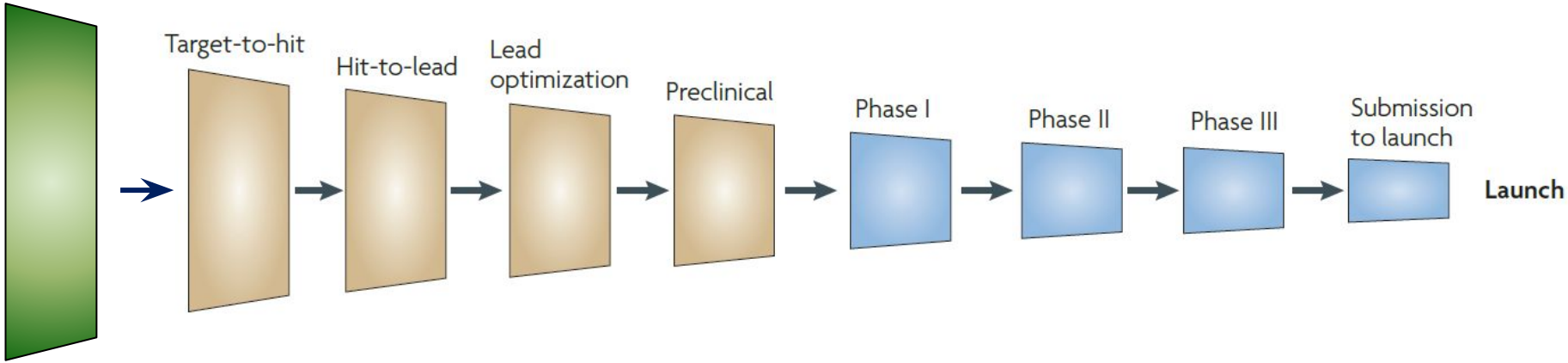
Working experience with mathematical and computational biology

Yes
No

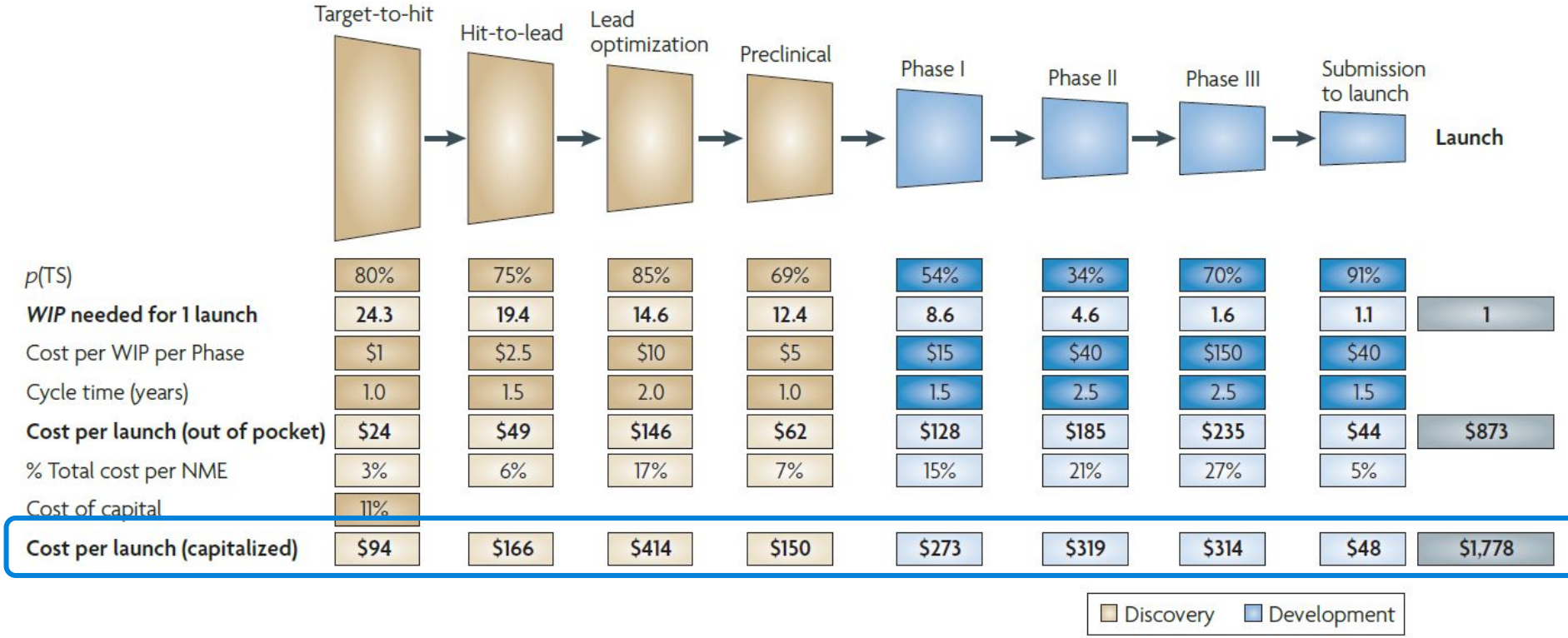
Your interest and motivation *motivate* me

A linear view of drug discovery

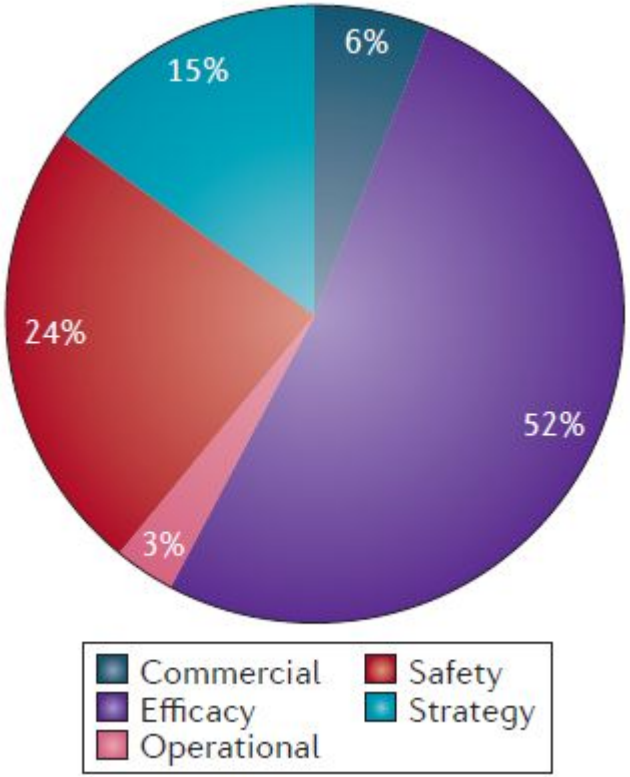
Target identification & assessment



Discussion: conclusions from the figures?



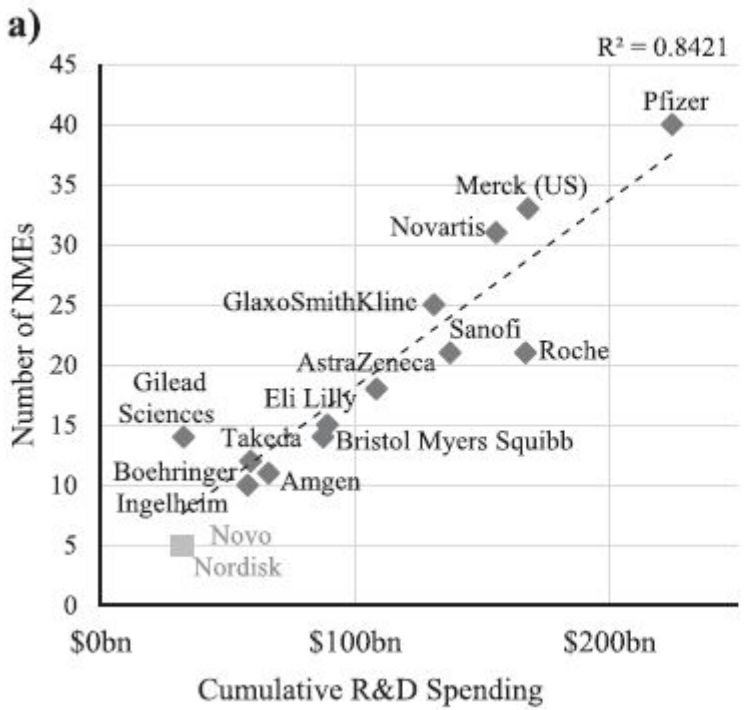
Failure analysis: 2013-2015

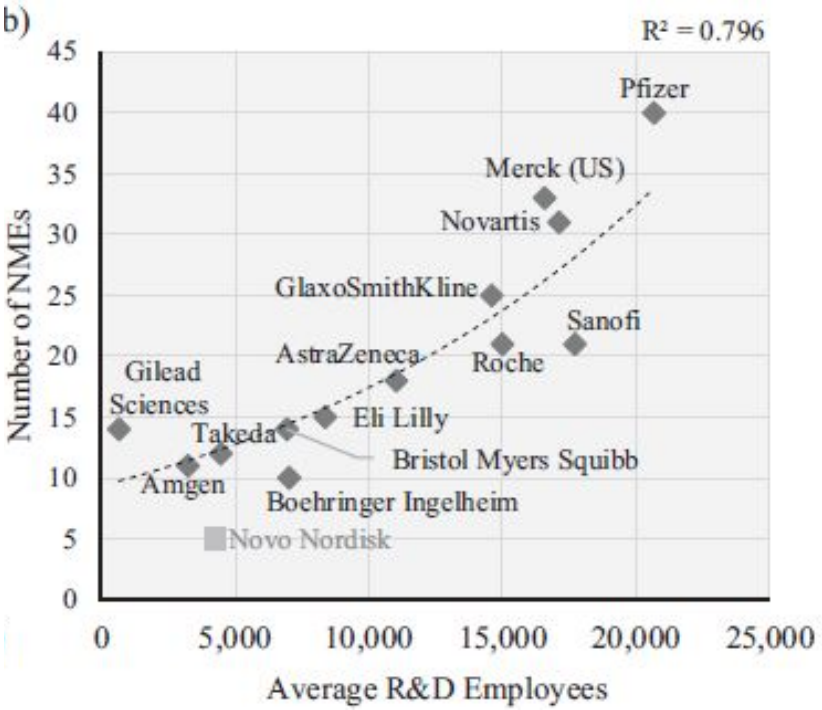


Learnings from numbers

1. Cost of target assessment and identification is not explicit.
2. Clinical studies are expensive, but picking a wrong target is twice as expensive.
3. It is probably wise to *infer* efficacy and safety profiles of drugs as accurately as possible.

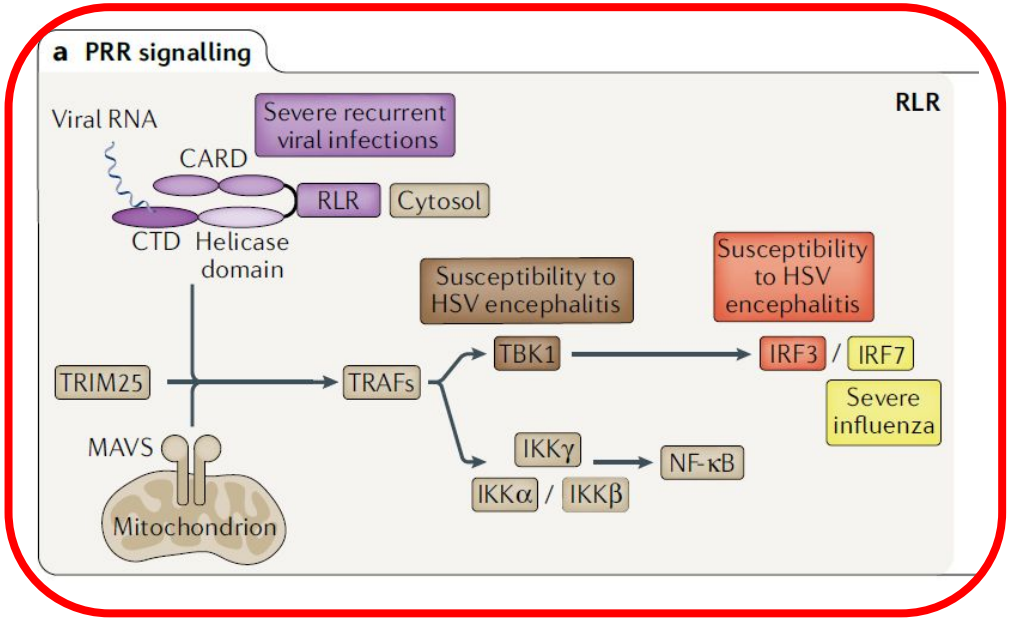
Investment and collaboration are necessary

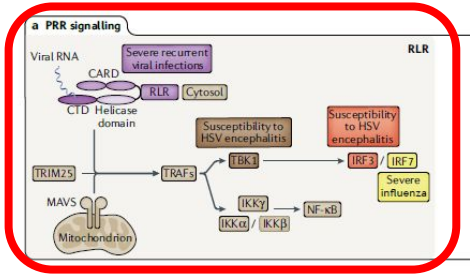


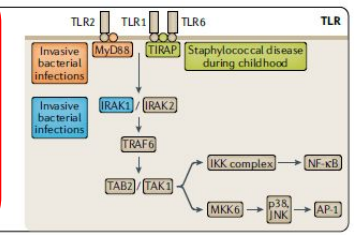


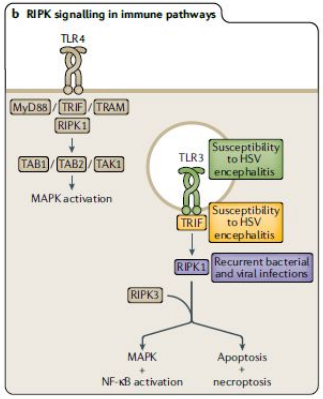
Complex Adaptive System

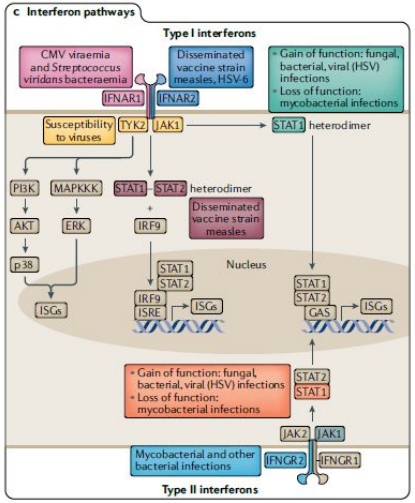
1. Parallel information channels
2. Conditional actions (if/then)
3. Modularity

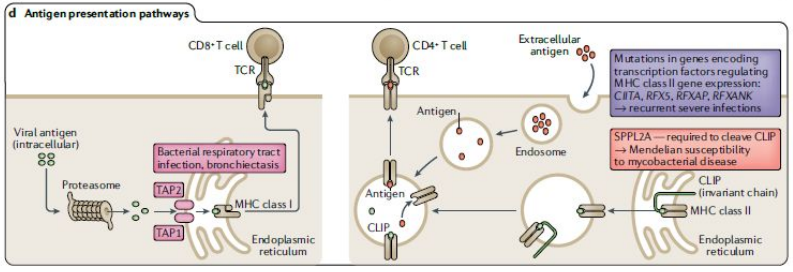












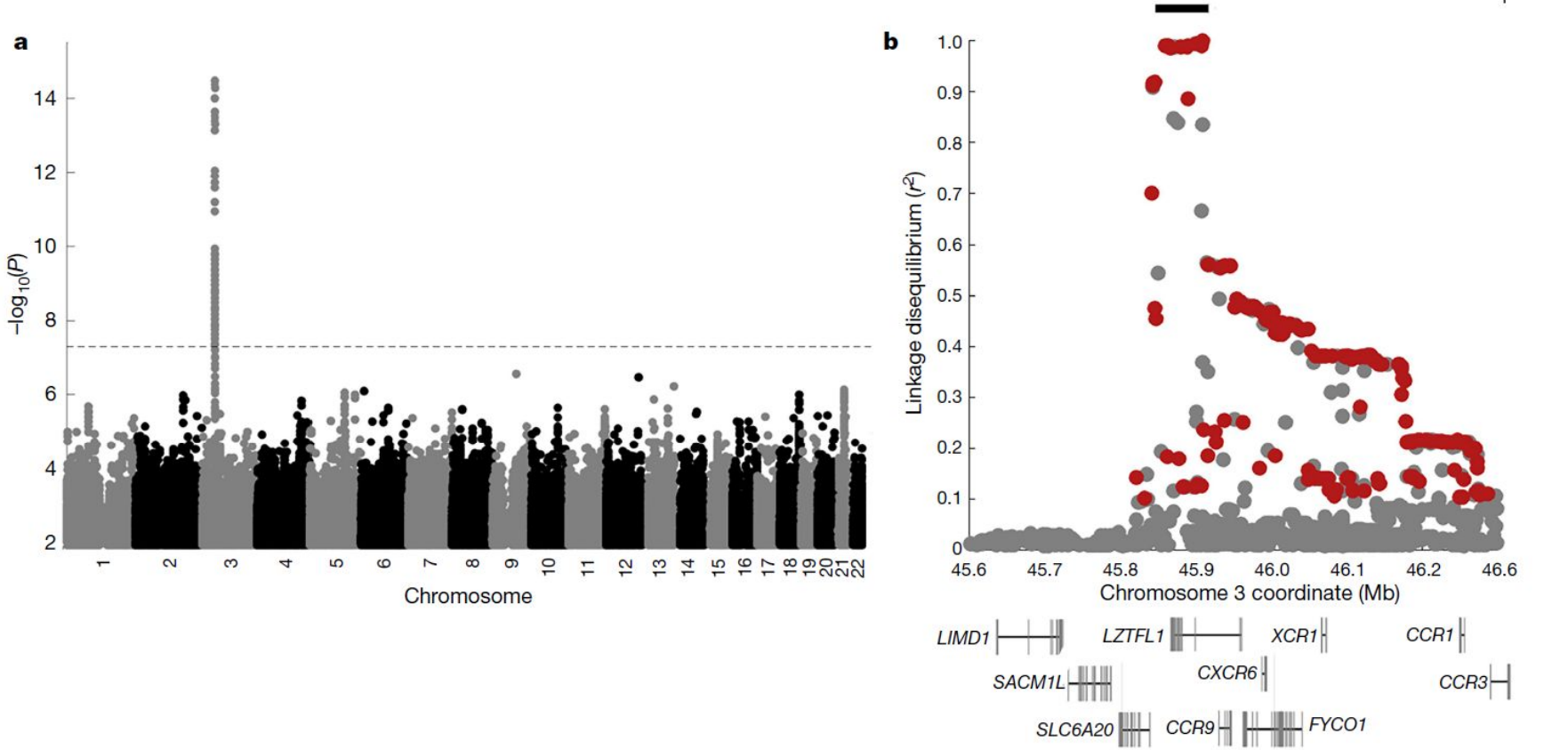
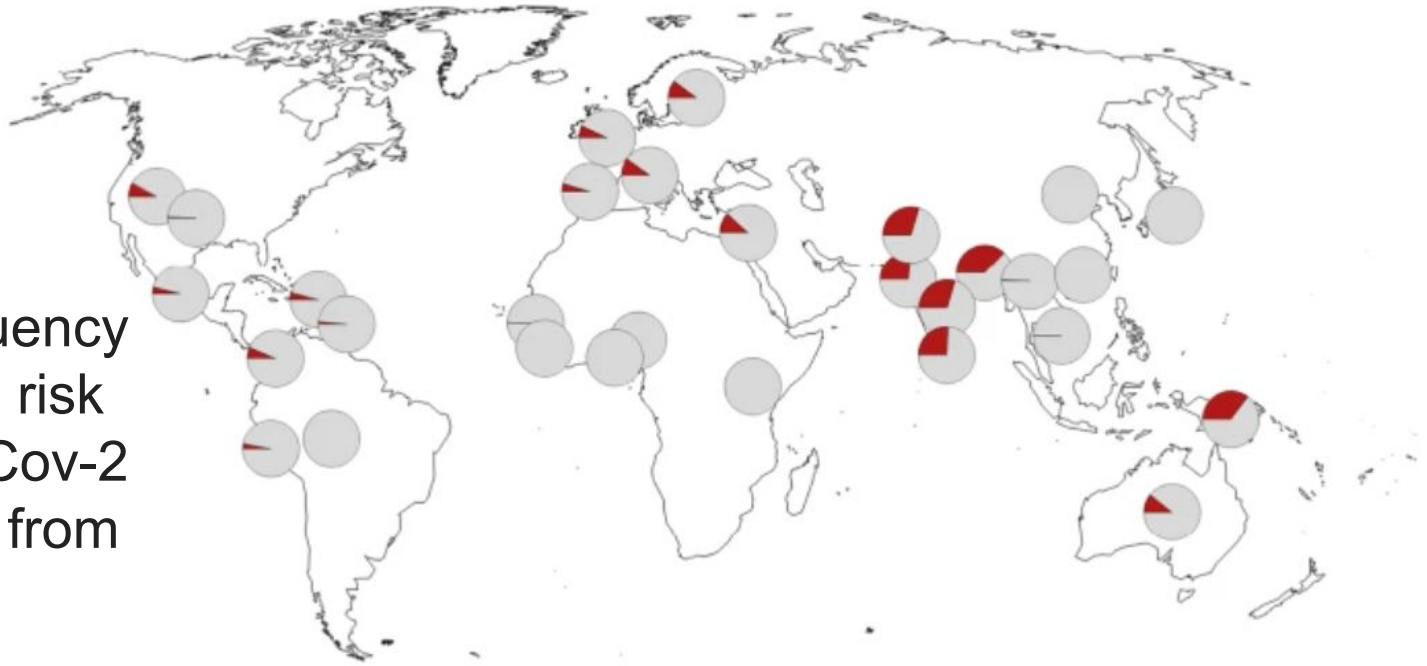


Fig. 1 | Genetic variants associated with severe COVID-19. a, Manhattan plot of a genome-wide association study of 3,199 hospitalized patients with COVID-19 and 897,488 population controls. The dashed line indicates genome-wide significance ($P = 5 \times 10^{-8}$). Data were modified from the COVID-19 Host Genetics Initiative[2] (https://www.covid19hg.org/). **b**, Linkage disequilibrium between the index risk variant (rs35044562) and genetic variants in the 1000 Genomes Project. Red circles indicate genetic variants for which the alleles are correlated to the risk variant ($r^2 > 0.1$) and the risk alleles match the Vindija 33.19 Neanderthal genome. The core Neanderthal haplotype ($r^2 > 0.98$) is indicated by a black bar. Some individuals carry longer Neanderthal-like haplotypes. The location of the genes in the region are indicated below using standard gene symbols. The *x* axis shows hg19 coordinates.

Complex Adaptive System

4. Adaptation and evolution



Minor allele frequency at rs35044562, a risk allele for SARS-Cov-2 that we inherited from Neanderthals.

Propositions about the course

1. Human (disease) biology is a hierarchical complex adaptive system.
2. Drug discovery aims at identifying *agents* that change the system's behaviour with acceptable benefit and risk profiles.
3. We use mathematical and computational biology to study the system in order to modulate it.

Complementary views of biological systems

- Metabolism
- Energy
- Information machine
- Evolution
- Computing machine
- Network
- ...

An example of complementary views

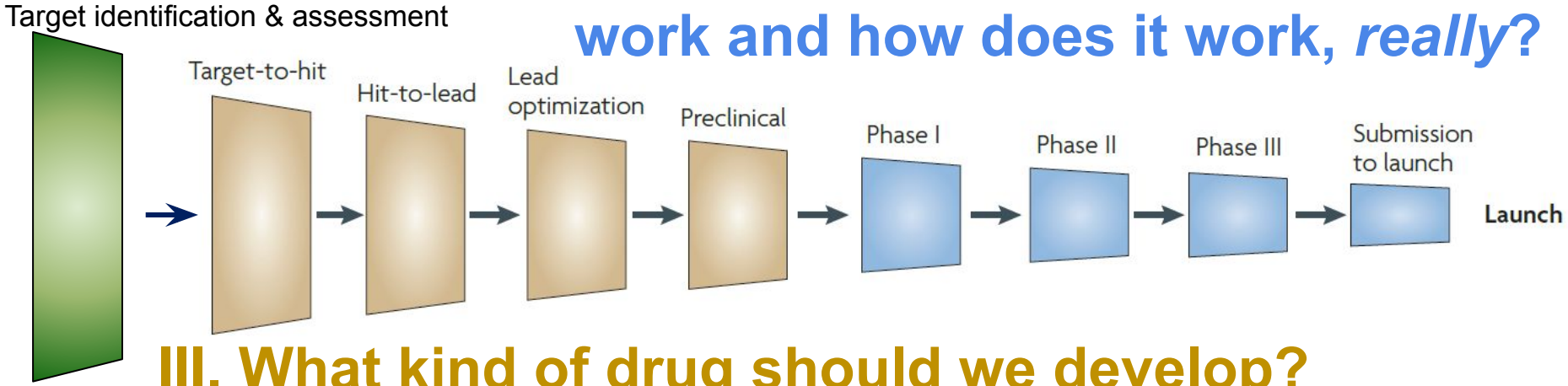
We want to work on hepatocarcinoma (liver cancer) and have the following information about a potential target X:

- X is a receptor expressing on the surface of most cell types;
- Upon binding ligands, X activates innate immune response;
- Gene sequence of X is conserved in primates but *not* in rodents;
- Protein X interacts with protein Y, which is essential, namely Y knockout causes lethal embryos;
- Asian population has a unique genetic variant in the non-coding region of *X*;

Discussion: what are the consequences of having these information?

Questions that we will address in this course

V: For which patients will the drug work and how does it work, *really*?



III. What kind of drug should we develop?

IV. What efficacy and safety profiles can we expect?

I. What makes a good drug target?

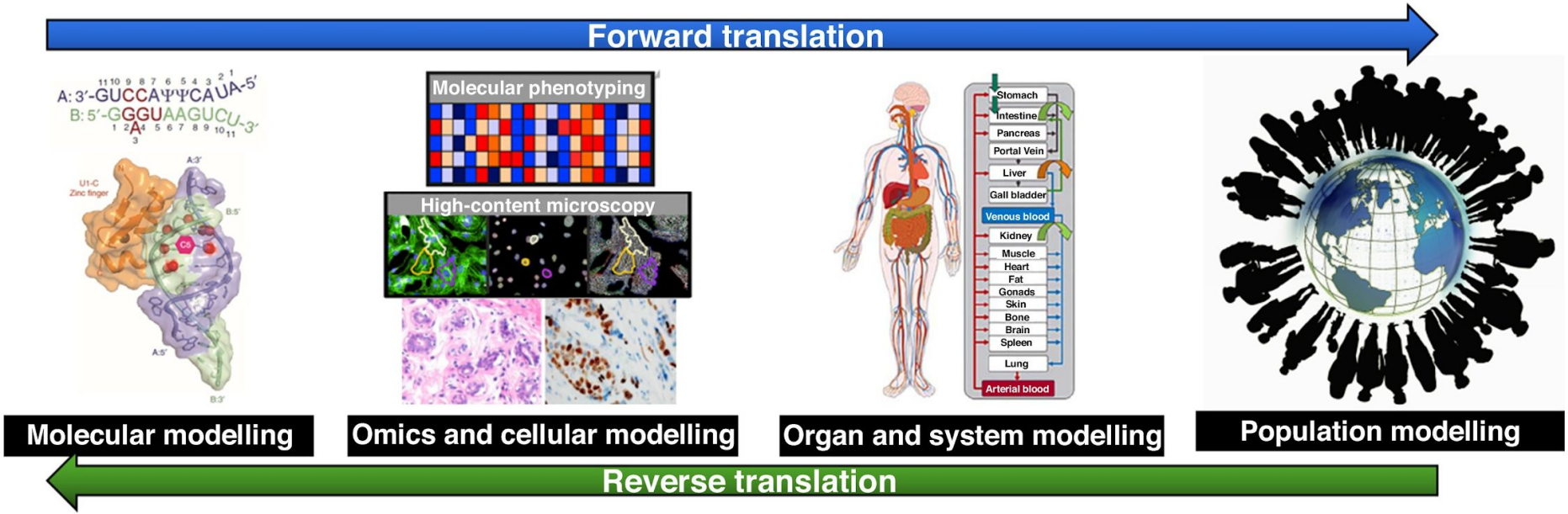
II. What can we do if there are no good targets?

Drug Discovery	Biology	Math./Comp.
Target identification, assessment, and phenotypic screening	• Genomics • Genetics • Gene expression • Chemical biology	• Statistical modelling • Machine learning • Mechanistic modelling
Drug modality and preclinical modelling	• RNA, antisense oligonucleotides, and antibodies • Gene expression • Network analysis	• Dynamic programming • Monte-Carlo methods • Clustering
Biomarker, clinical modelling and reverse translation	• Population genetics • Gene expression • Pharmacokinetics and pharmacodynamics	• Causal analysis • Learning algorithms • Agent-based modelling

Common modelling approaches

- **Statistical modelling**
- **ODE/PDEs (compartment & transport models)**
- **Agent-based models (particle models)**
- **Networks (graphical and boolean models)**
- *Evolutionary game theory*
- *Fractional calculus*

A multiscale-modelling view of drug discovery



Drug Discovery Today

Exercise

Right target

- Strong link between target and disease
- Differentiated efficacy
- Available and predictive biomarkers

Right tissue

- Adequate bioavailability and tissue exposure
- Definition of PD biomarkers
- Clear understanding of preclinical and clinical PK/PD
- Understanding of drug–drug interactions

Right safety

- Differentiated and clear safety margins
- Understanding of secondary pharmacology risk
- Understanding of reactive metabolites, genotoxicity and drug–drug interactions
- Understanding of target liability

Right patient

- Identification of the most responsive patient population
- Definition of risk–benefit for a given population

Right commercial potential

- Differentiated value proposition versus future standard of care
- Focus on market access, payer and provider
- Personalized health-care strategy, including diagnostics and biomarkers

Where do you think mathematical and computational biology will make a difference?

Nature Reviews | Drug Discovery

Take-home messages

- Drug discovery identifies agents modulating human disease biology as a hierarchical complex adaptive system.
- Mathematical and computational biology studies interactions within the system and help to build predictive models.
- Reproducible computational research help ourselves and others build a sustainable working environment.

Offline activities

1. **Last chance to fill the pre-course survey**, and (optionally) fill the *anonymous,* post-lecture survey.
2. Read 'Improving target assessment in biomedical research: the GOT-IT recommendations' by Emmerich *et al.* (NRDD, 2021). Think about the question that we will discuss next time: **what is a good drug target**?
3. (Optional) We celebrate 20 years of the human genome in 2021. Much promise was made and how much was fulfilled? Read the article '*Complicated legacies: the human genome at 20*' (Science, 2021).

References

1. Paul *et al.* “How to Improve R&D Productivity: The Pharmaceutical Industry’s Grand Challenge.” Nature Reviews Drug Discovery, 2010.
2. Ringel, Michael S., Jack W. Scannell, Mathias Baedeker, and Ulrik Schulze. 2020. “Breaking Eroom’s Law.” Nature Reviews Drug Discovery 19 (12): 833–34. https://doi.org/10.1038/d41573-020-00059-3.
3. Morgan, Paul, Dean G. Brown, Simon Lennard, Mark J. Anderton, J. Carl Barrett, Ulf Eriksson, Mark Fidock, et al. 2018. “Impact of a Five-Dimensional Framework on R&D Productivity at AstraZeneca.” Nature Reviews Drug Discovery 17 (3): 167–81. https://doi.org/10.1038/nrd.2017.244.
4. Harrison, Richard K. 2016. “Phase II and Phase III Failures: 2013–2015.” Nature Reviews Drug Discovery 15 (November): 817–18. https://doi.org/10.1038/nrd.2016.184.
5. Schuhmacher, Alexander, Lucas Wilisch, Michael Kuss, Andreas Kandelbauer, Markus Hinder, and Oliver Gassmann. “R&D Efficiency of Leading Pharmaceutical Companies – A 20-Year Analysis.” Drug Discovery Today 26, no. 8 (August 1, 2021): 1784–89. https://doi.org/10.1016/j.drudis.2021.05.005.
6. Zhang, Jitao David, Lisa Sach-Peltason, Christian Kramer, Ken Wang, and Martin Ebeling. 2020. “Multiscale Modelling of Drug Mechanism and Safety.” Drug Discovery Today 25 (3): 519–34. https://doi.org/10.1016/j.drudis.2019.12.009.
7. Holland, John H. 2006. “Studying Complex Adaptive Systems.” Journal of Systems Science and Complexity 19 (1): 1–8. https://doi.org/10.1007/s11424-006-0001-z.
8. Kwok, Andrew J., Alex Mentzer, and Julian C. Knight. 2021. “Host Genetics and Infectious Disease: New Tools, Insights and Translational Opportunities.” Nature Reviews Genetics 22 (3): 137–53. https://doi.org/10.1038/s41576-020-00297-6.
9. Zeberg, Hugo, and Svante Pääbo. 2020. “The Major Genetic Risk Factor for Severe COVID-19 Is Inherited from Neanderthals.” Nature 587 (7835): 610–12. https://doi.org/10.1038/s41586-020-2818-3.
10. Sturm, Gregor. 2020. “Hallmarks of Good Scientific Software”. https://grst.github.io/bioinformatics/2020/07/16/hallmarks-scientific-software.html

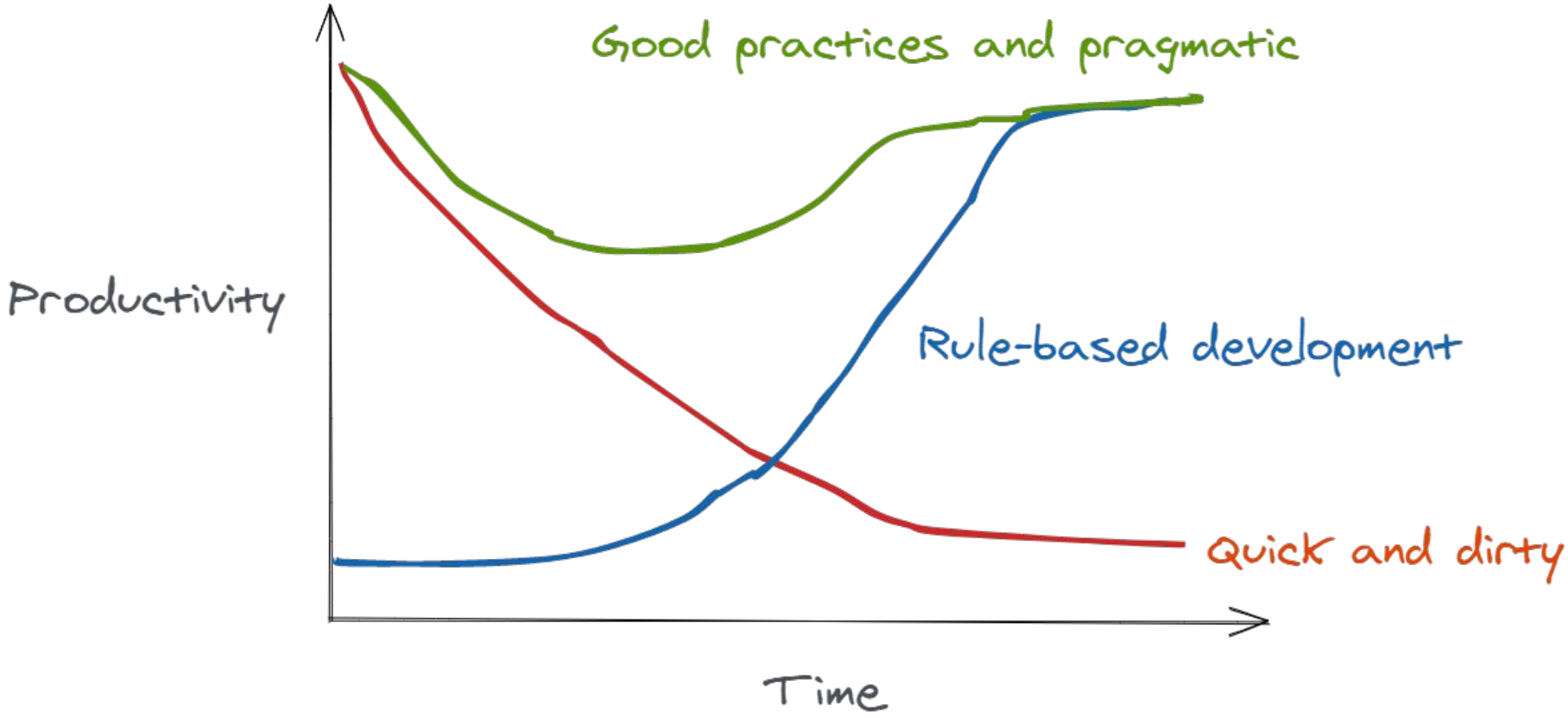
Backup and License

Nine steps toward reproducible research

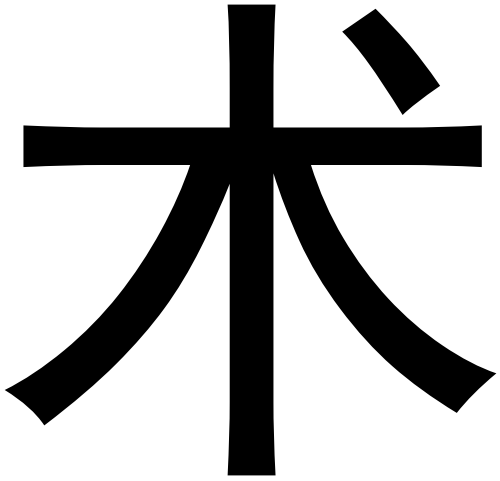
1. Version control (*git*)
2. Don’t Repeat Yourself (DRY)
3. Keep It Simple, Stuipid (KISS)
4. Automatic testing (*pytest/Hypothesis*, *testthat*, *GitHub Actions*)
5. Documentation (*sphinx, pkdown*)
6. Dependency Management (*conda*, *packrat*)
7. Containerization (*Docker/Singularity*, *Bioconda/conda-forge*)
8. Pipelining (*Snakemake*, *NextFlow*, *drake*)
9. Self-reporting analysis (*Jupyter Notebook*, *Rmarkdown*)

Arguments for reproducible research

- Egoism and altruism
- *You will have to do it again*
- Sustainable long-term work


Good practices and pragmatic
Productivity
Rule-based development
Quick and dirty
Time

Tao, Path, or Way

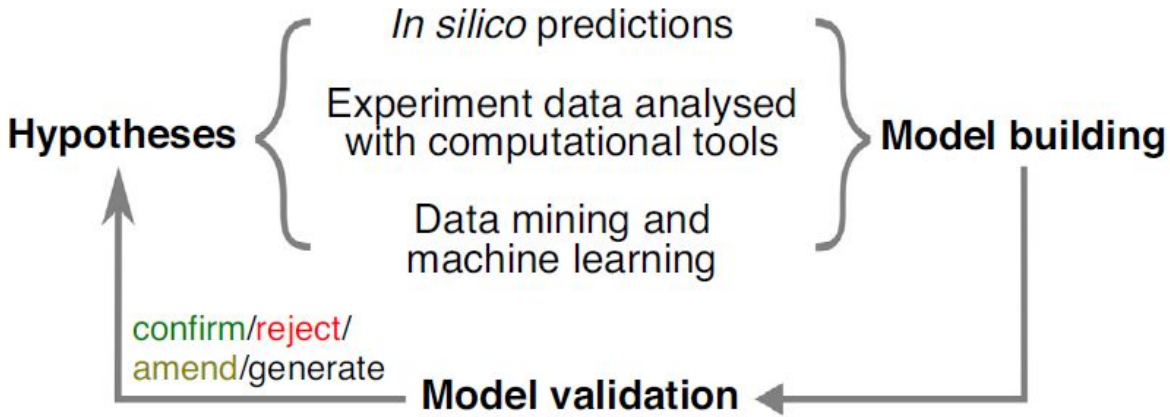


Shu, Technique, or Art

Learn more about reproducible research

- Software Carpentry (Unix Shell, Git, Python & R)
- Genomics Workshop of Data Carpentry
- *Clean Code* by Robert C. Martin
- Open-source tutorials of respective tools, such as sphinx, Snakemake, conda, or docker. Videos or podcasts work just as fine.

Mathematical and computational biologists are part of an interdisciplinary team



One-compartment model, oral dosing

For oral dosing, an extra gut compartment (right) is often sufficient to model the absorption phase

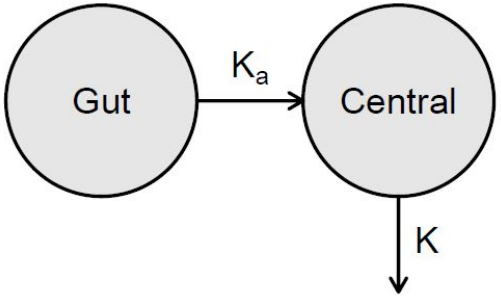
$$\frac{dA_{gut}}{dt} = -K_a \cdot A_{gut}$$

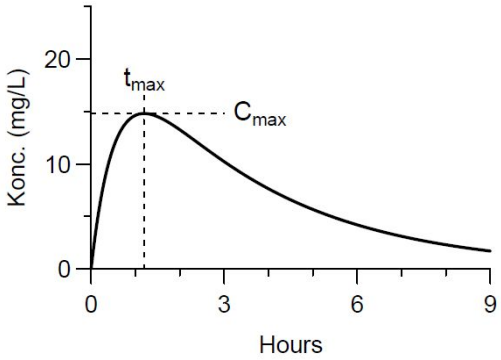
Suppose rate the absorption of the drug is faster than the elimination process (K*a*>*K*), we can model the concentration in the central compartment as

$$\frac{dA}{dt} = \overbrace{F \cdot K_a \cdot A_{gut}}^{\text{from gut}} - \overbrace{K \cdot A}^{\text{elimination}}$$

In reality, we cannot easily assess the concentration of drug in the gut. Is it possible to derive the relationship between central-compartment concentration *A* and time *t* given the initial condition?

Yes: we can find the expression of *A*(t) analytically in a closed form using *Laplace transform*, which translates a function of a continuous variable (*e.g.* time) to a function of a complex variable (frequency) (see backup).





Contact the author