

Informing cell models with genomics analysis (and vice versa)

James Eddy – Sage Bionetworks
2018 mini-DREAM :: Module 7

July 31st 2018



What is a (cell) model?



“A model is a simplification or approximation of reality...”
~ Burnham & Anderson

Why build models?

“A model is a simplification or approximation of reality and hence will not reflect all of reality”

~ Burnham & Anderson

“Essentially, all models are wrong, but some are useful”

~ George Box

Models are not a novel concept in systems and computational biology! Biologists (of all types) have been using them for centuries!

What are some examples of models in science?

Word cloud!

(this is more confusing than I intended...)

A word cloud composed of various scientific terms and names, including components, bertalanffy, natural, bohr, random, persistent, walks, growth, selection, ribosomal, and model.

Scientific theories, cellular components, specific physical models, modeling classes – all great answers

Also worth noting: animal (e.g., mouse) models, cell lines, epidemiological models, anatomy diagrams

What are some things/processes we can model in biology?

Some examples

“Intracellular signaling”

“protein structure”

“biological pathways”

“growth curves”

“DNA”

“Cancer cell migration”

“Cell migration”

“blood flow”

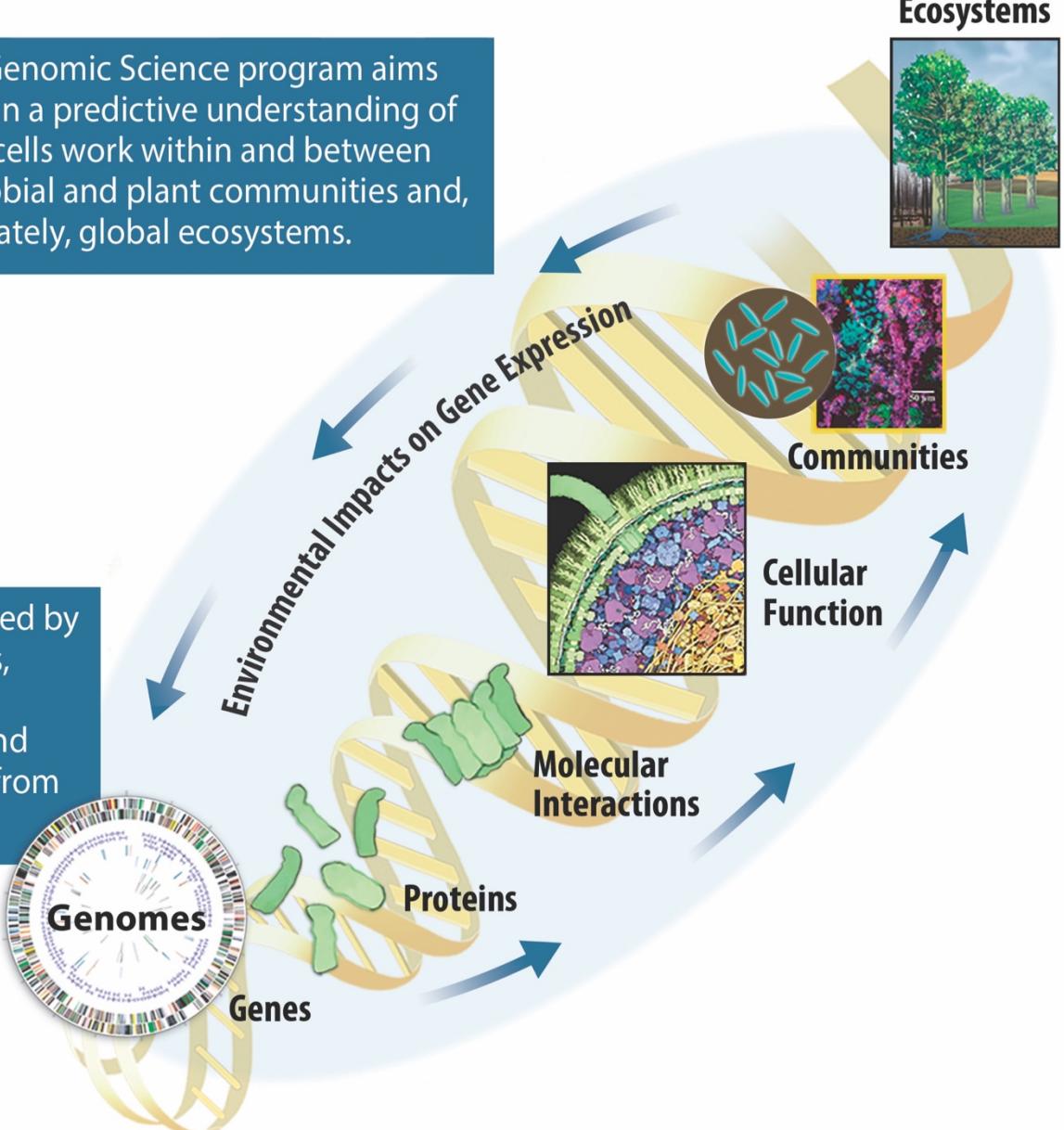
Many, many more!

Modeling different biological scales

Defining the “system”

The Genomic Science program aims to gain a predictive understanding of how cells work within and between microbial and plant communities and, ultimately, global ecosystems.

The genome, influenced by environmental factors, determines dynamic biological structure and function at all scales, from genes to ecosystems.



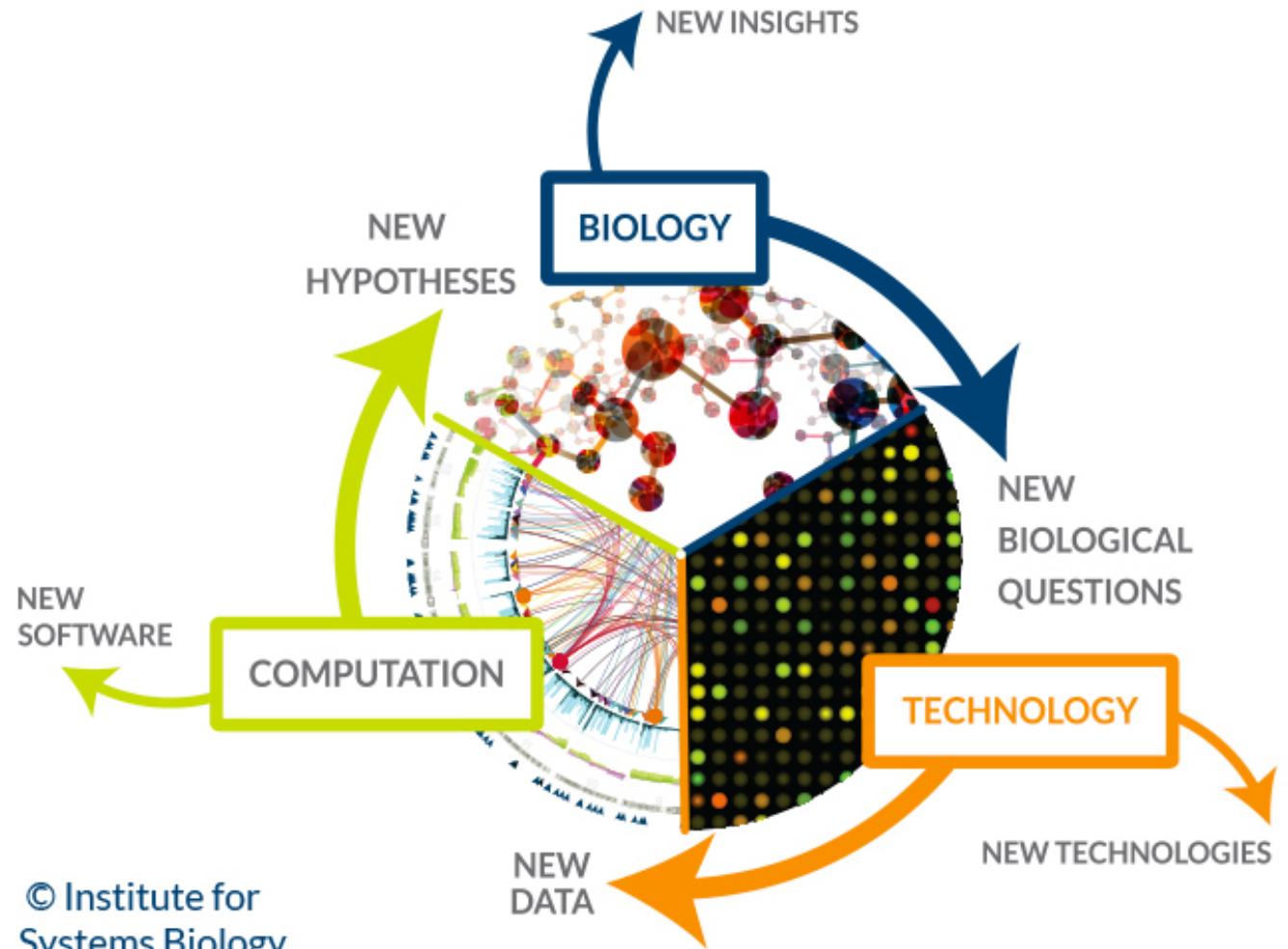
What is systems biology?

Toolbox for models of non-trivial scale

Beyond a few components, we lack the ability to effectively build or interpret models by hand/intuition

Mathematics and computation required, systems biology embodies this regime (large scales, complex)

Still very much reliant on good experimental data (and can also drive hypothesis generation)



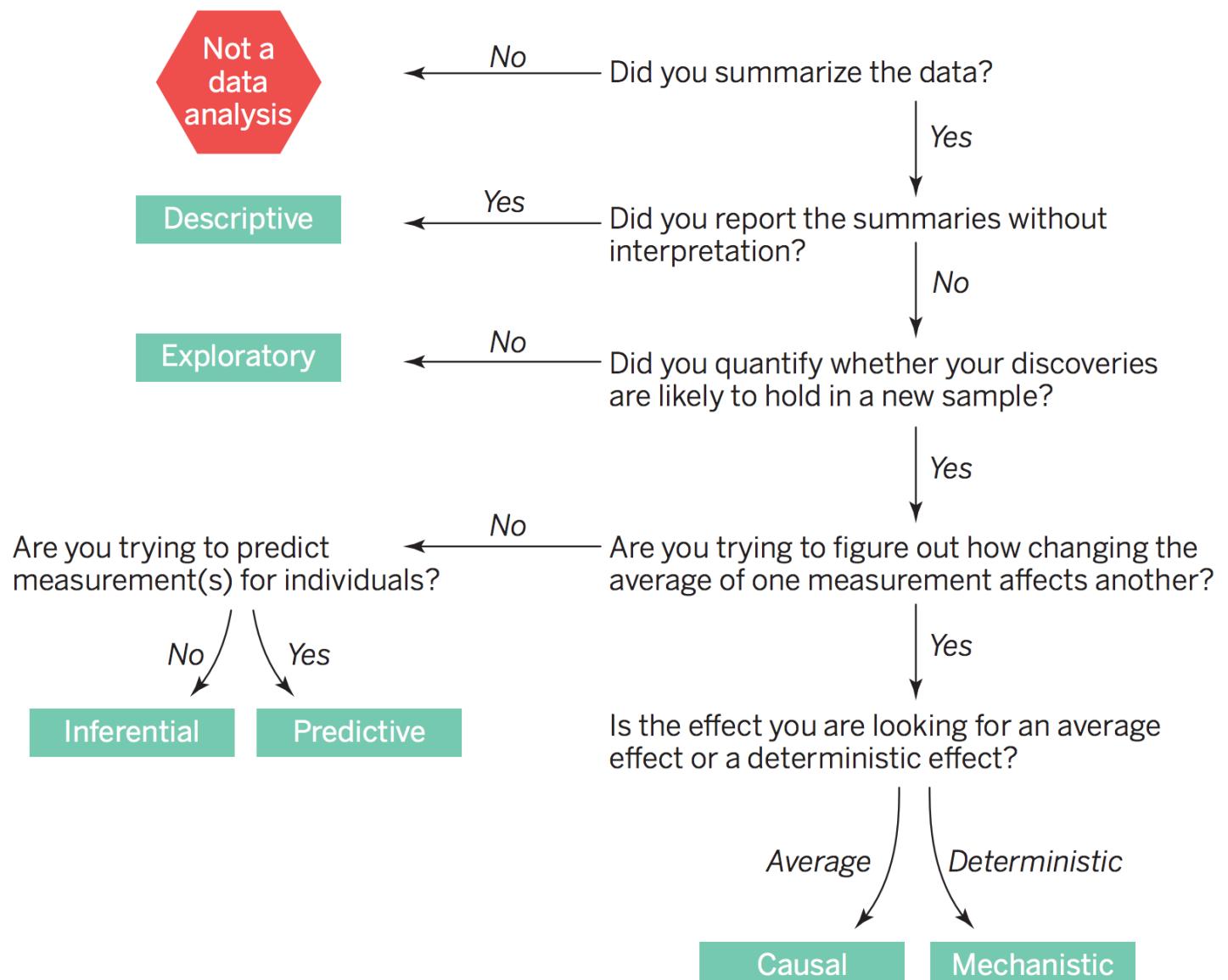
© Institute for Systems Biology

Stages of understanding in modeling

What is the question?

Answers you can hope to get from a model depend on the data used to build it (and also how it was designed)

Generalizable understanding of mechanism is a great goal, but not often feasible – especially in biology



Bottom-up modeling is hard

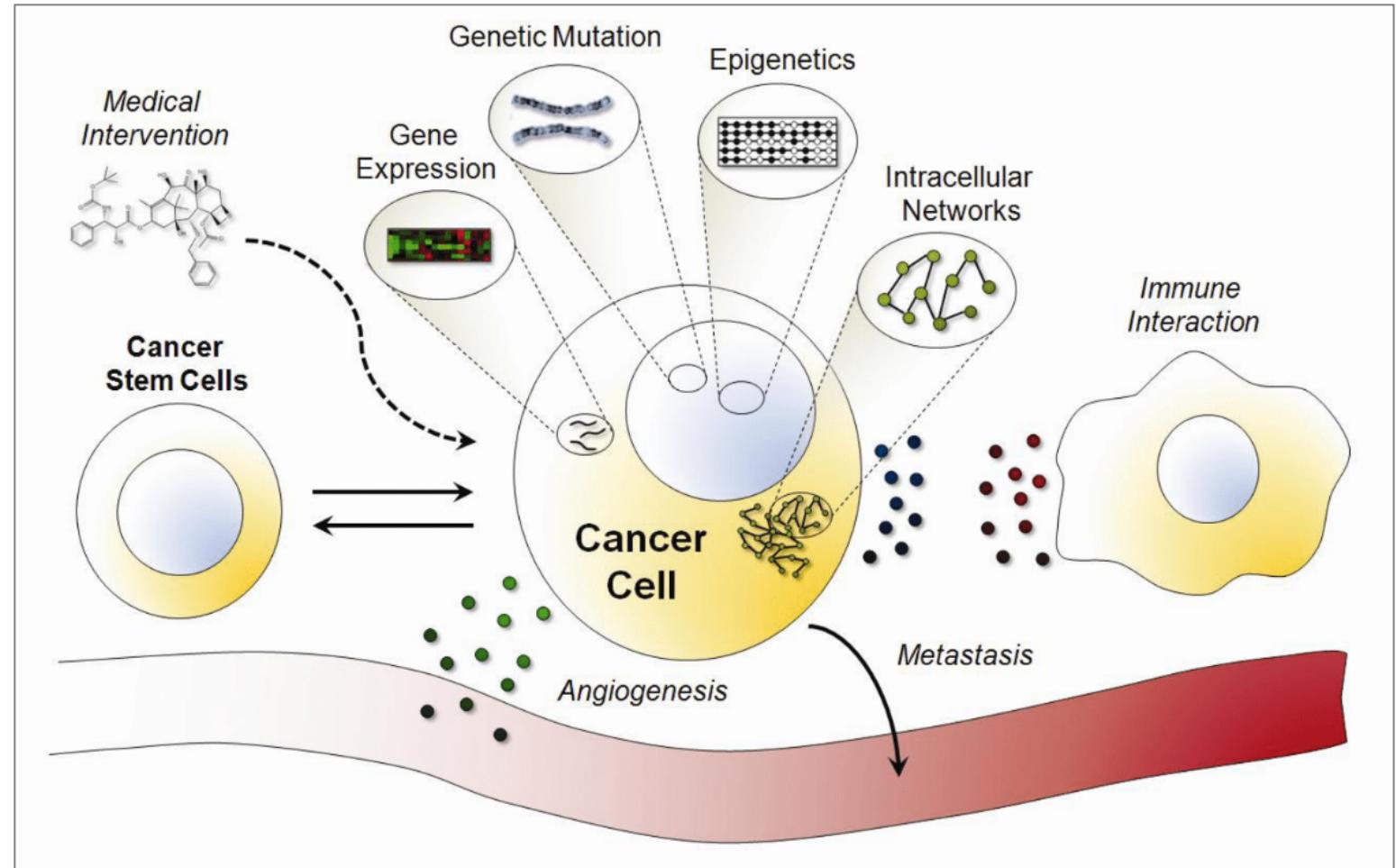
Why not build detailed mechanistic models of everything?

Lack of data

Impossible to simultaneously measure everything

Context-specific behavior

Computational complexity



Top-down modeling is limited

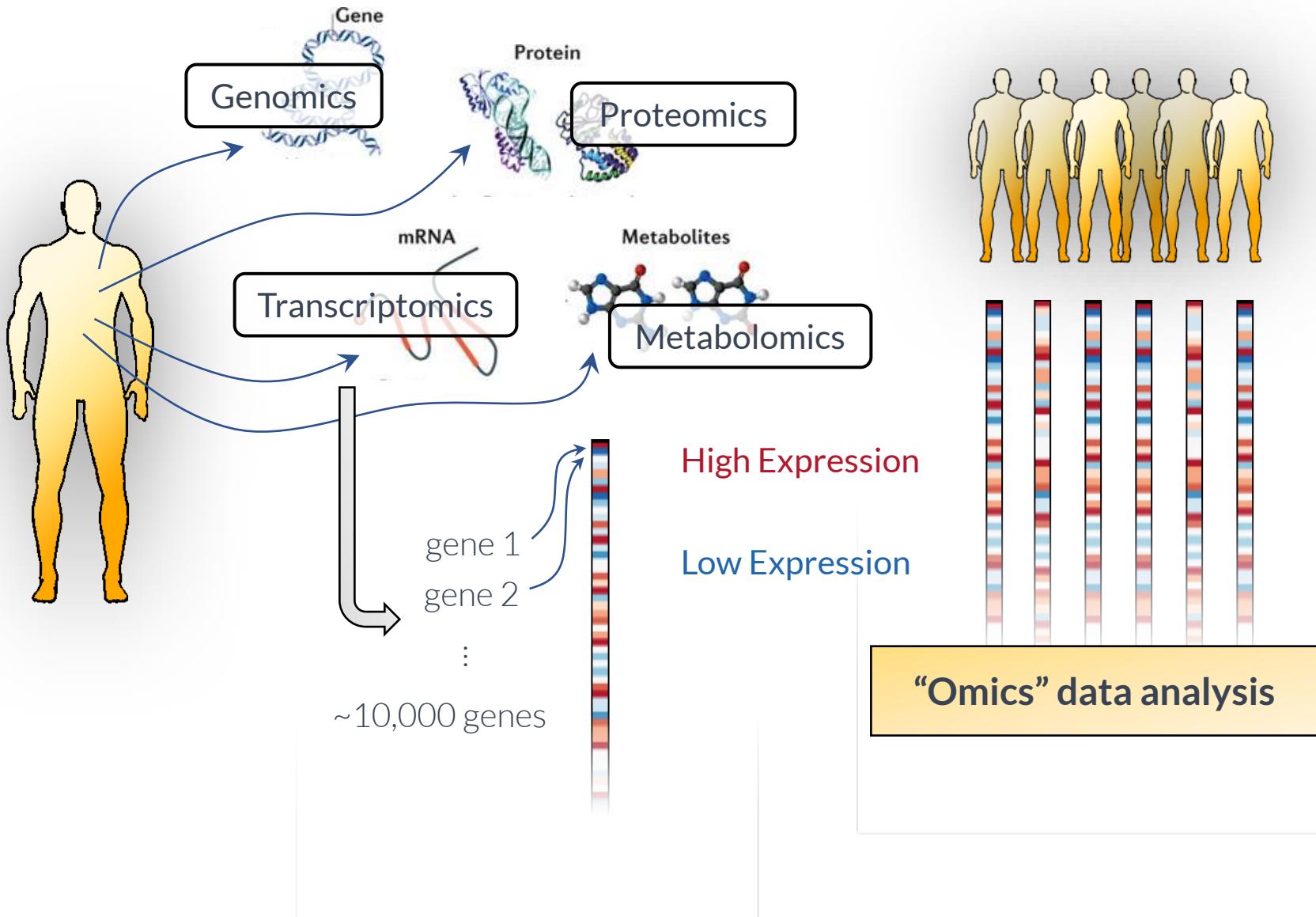
Why not model everything from data?

Lack of data (still)

Measurements often indirect, noisy

Still just a snapshot

Need carefully designed experiments (and analyses) to determine causality

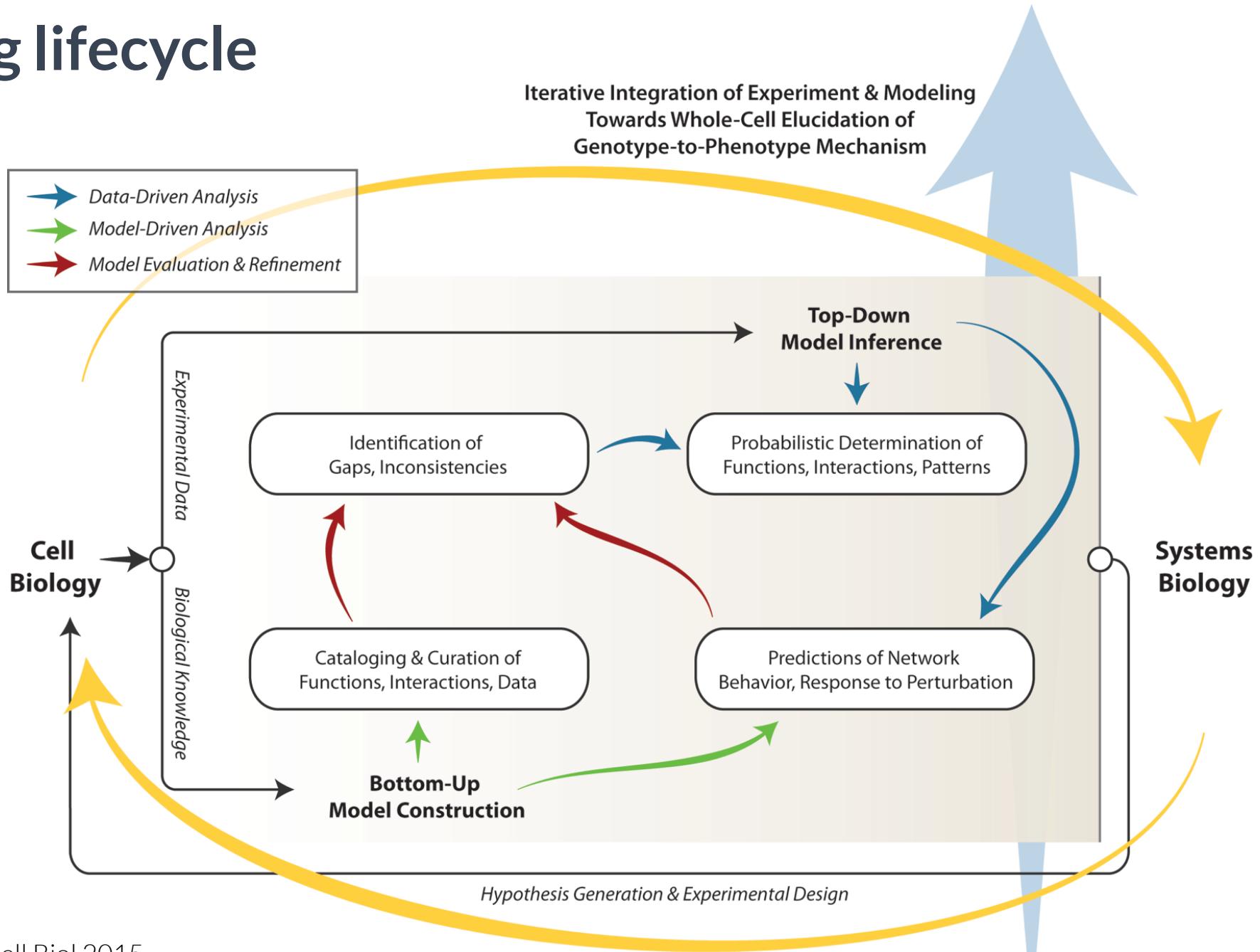


Middle-out modeling?

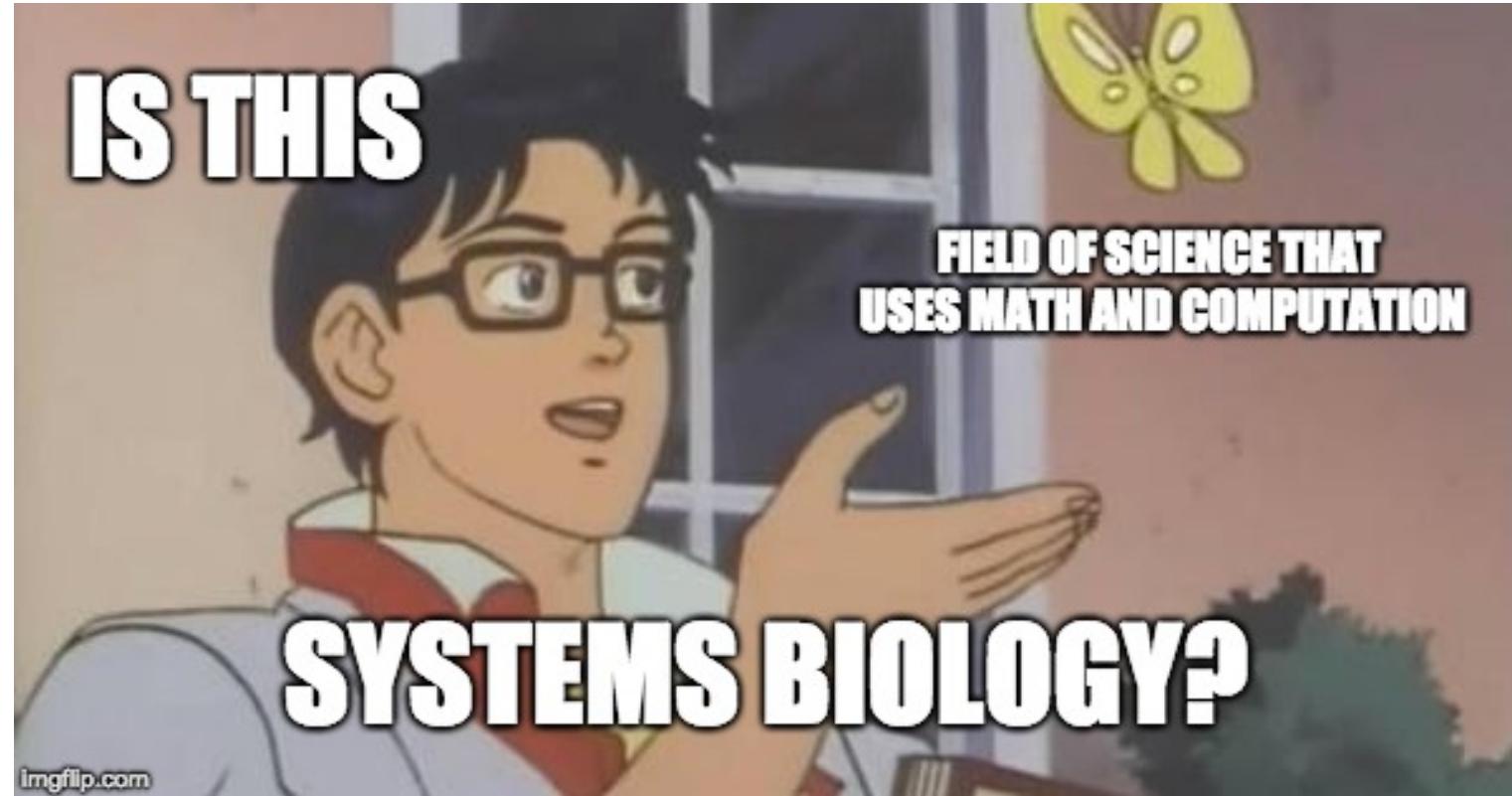


No need to choose bottom-up or top-down, can leverage the strengths of both in iterative process (aka. “science”)

Modeling lifecycle



What's in a name?



Which of these areas do you think qualifies as "systems biology"?

Bioinformatics **A**

Computational biology **B**

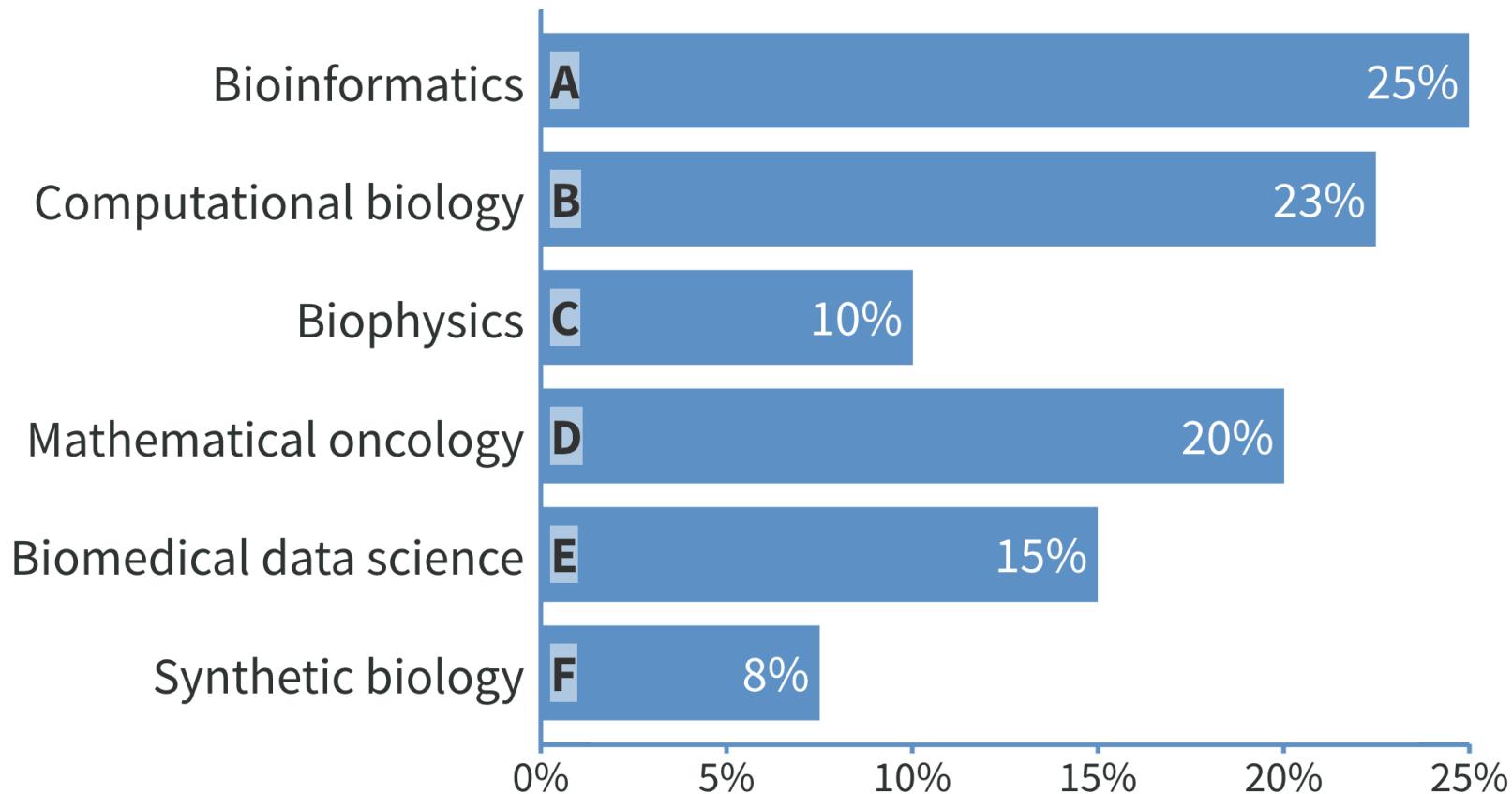
Biophysics **C**

Mathematical oncology **D**

Biomedical data
science **E**

Synthetic biology **F**

Answer: all of the above!

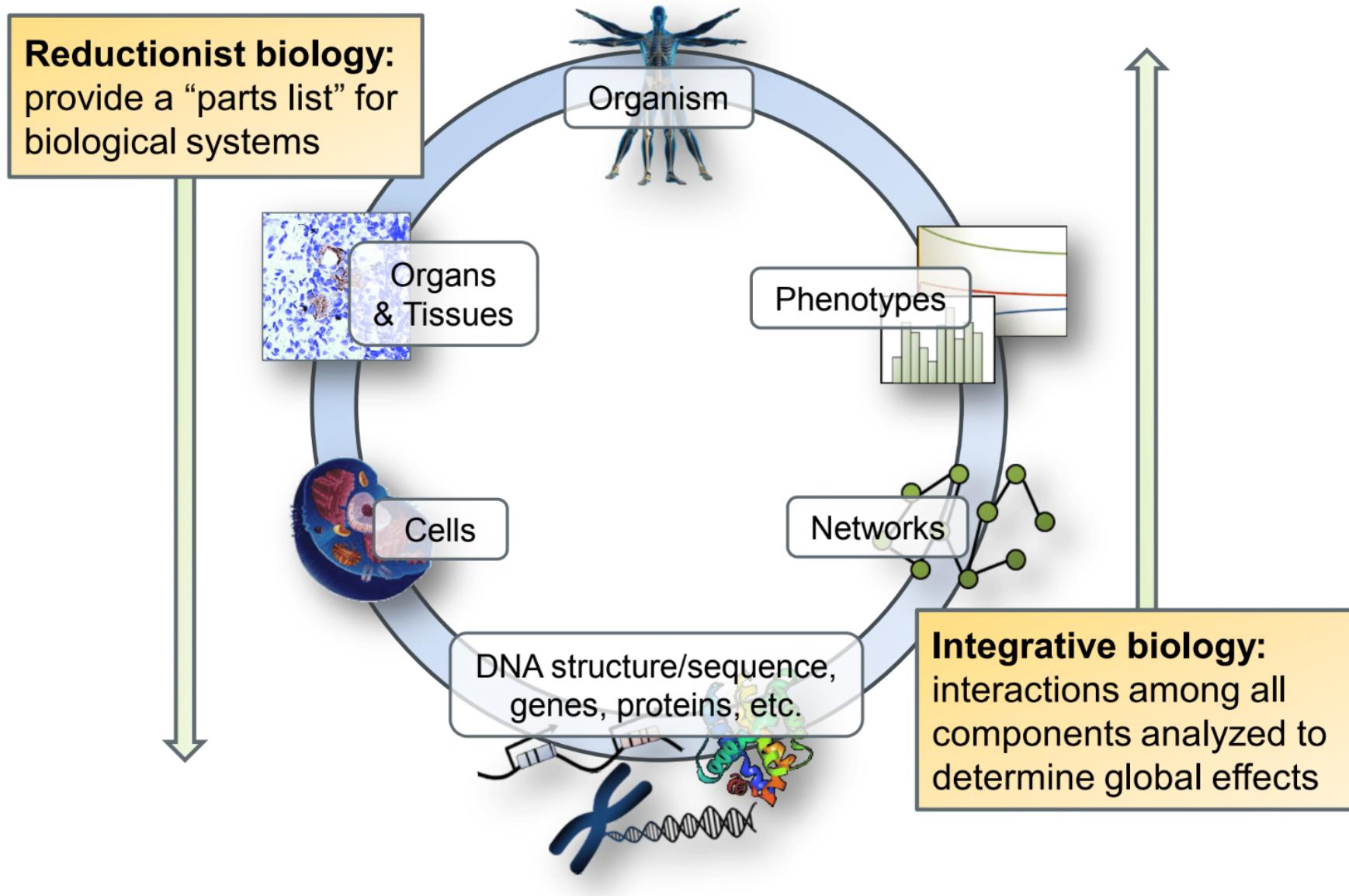


Emphasis on “toolbox”

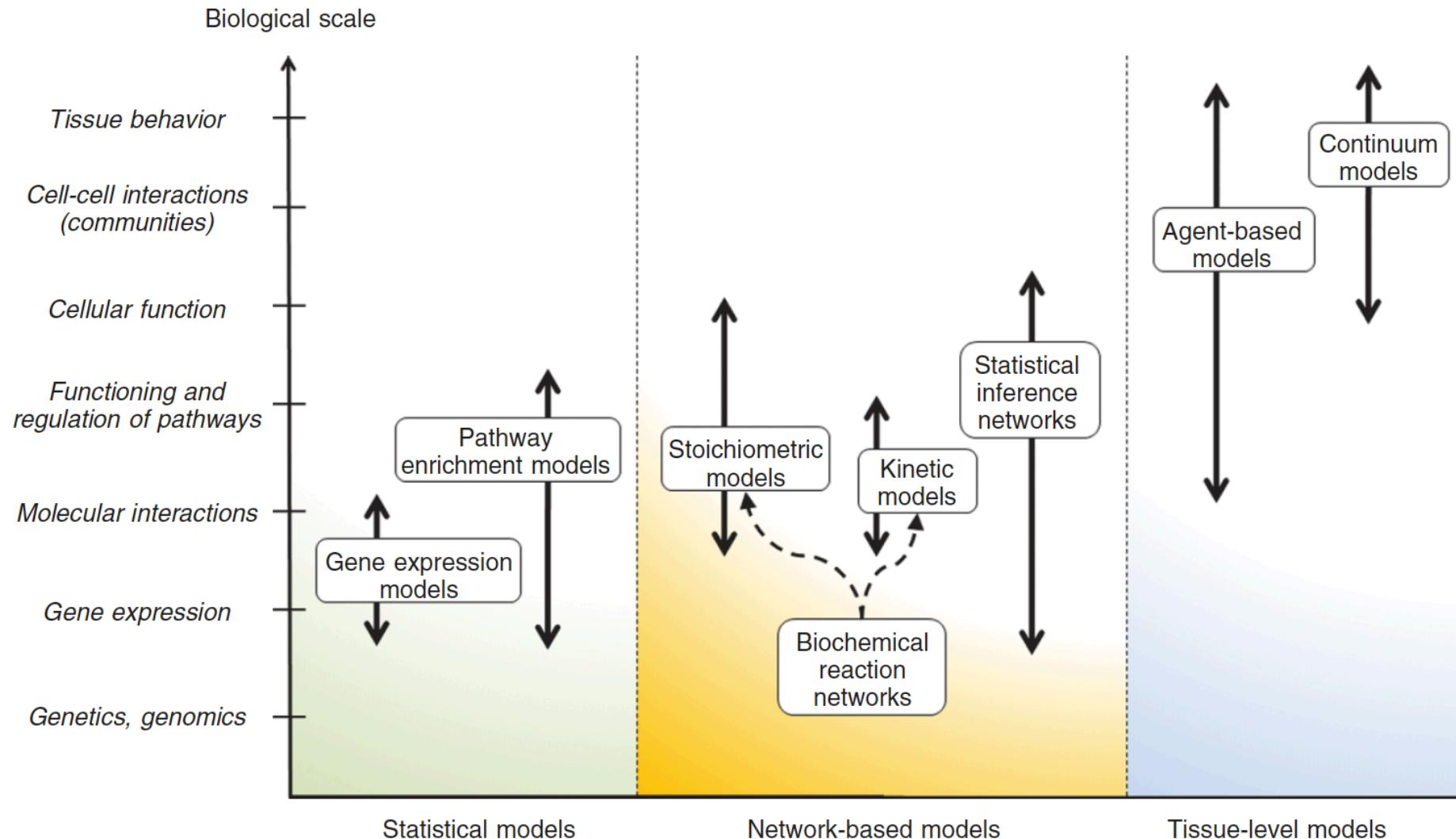


Lots of complementary tools and approaches...
don't get hung up on semantics!

Integrative vs. reductionist biology



Flavors of (computational) systems biology models

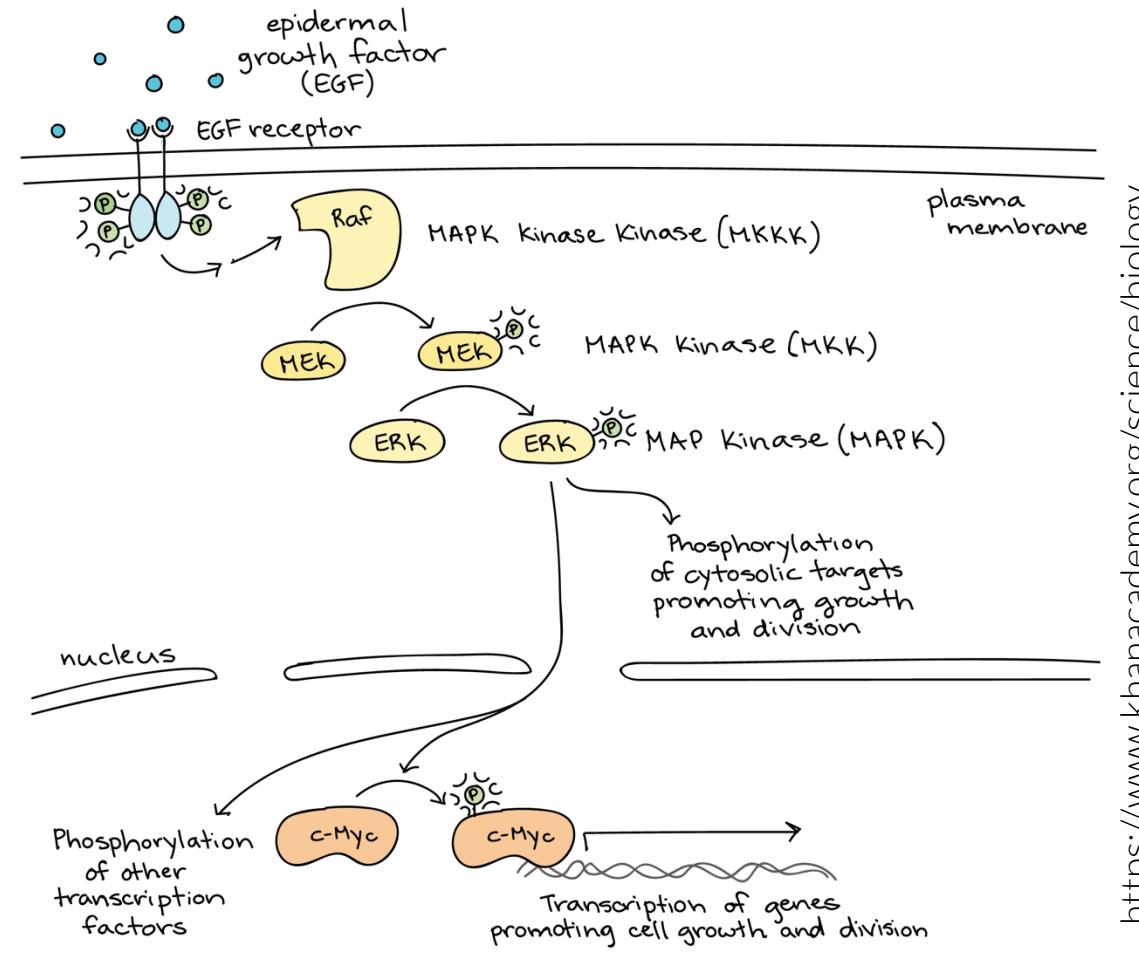


Kinetic/dynamic modeling

Mathematical modeling of physical & chemical properties

- Systems of equations (e.g., ODE, PDE) describing reaction rates based on **conservation laws, regulation patterns**, and (sometimes) probability
- Used to simulate **continuous** changes in time and space
- More intuitive (direct) representation of biomolecular systems – difficult to build, get prohibitive to compute

$$\frac{dRaf^*}{dt} = V_{raf} \cdot RPKC \cdot Raf - K_{dc} \cdot Raf^* \cdot RMAPK$$
$$\frac{dRaf}{dt} = -\frac{dRaf^*}{dt}$$
$$\frac{dMAPKK^*}{dt} = V_{mapkk} \cdot MAPKK \cdot RRaf - V_{dmapkk} \cdot MAPKK^*$$
$$\frac{dMAPKK}{dt} = -\frac{dMAPKK^*}{dt}$$
$$\frac{dMAPK^*}{dt} = V_{mapk} \cdot MAPK \cdot RMAPKK - V_{dmapk} \cdot MAPK^*$$
$$\frac{dMAPK}{dt} = -\frac{dMAPK^*}{dt}$$
$$RRaf = \frac{Raf^*}{\alpha_1 \cdot Raf_i}$$
$$RMAPKK = \frac{MAPKK^*}{\alpha_2 \cdot MAPKK_i}$$
$$RMAPK = \frac{MAPK^*}{\alpha_3 \cdot MAPK_i}$$



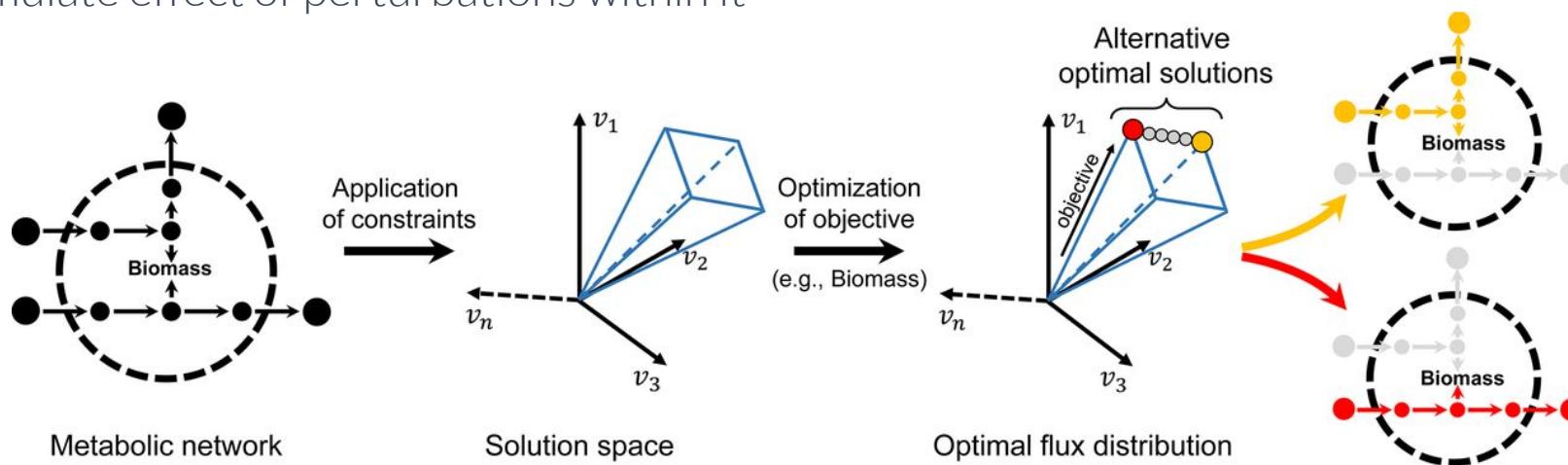
Constraint-based modeling

“Once you eliminate the impossible, whatever remains, no matter how improbable, must be the truth”

~ Sherlock Holmes (Arthur Conan Doyle)



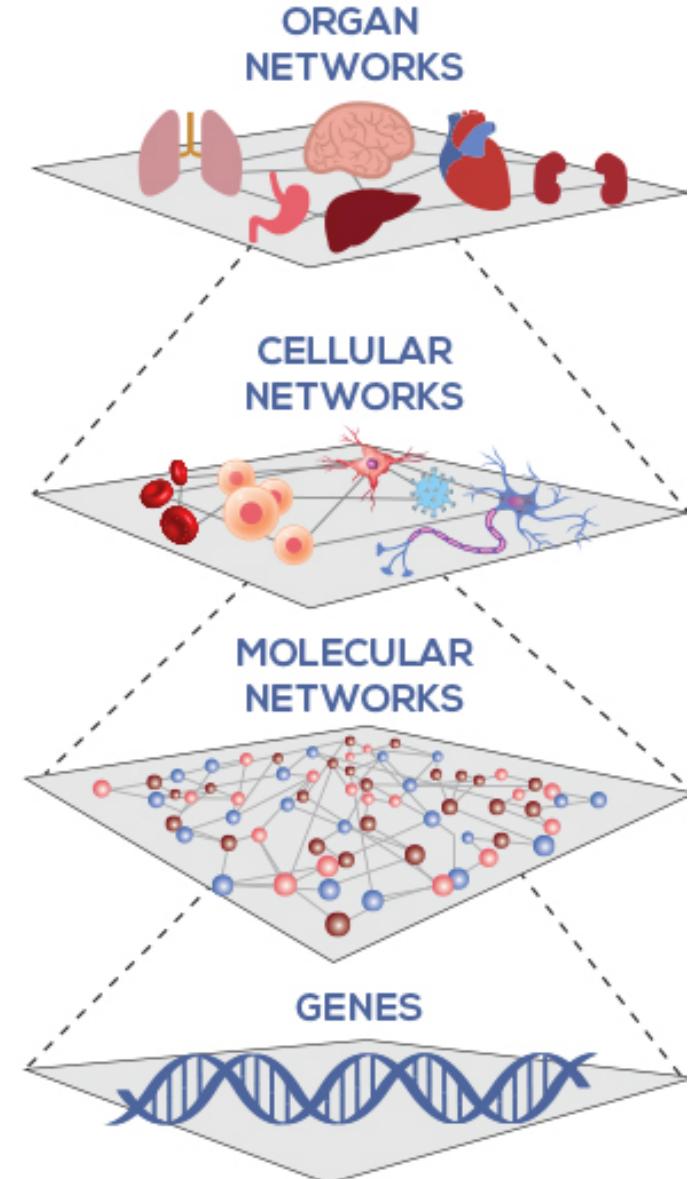
- Kinetic rules abstracted to **stoichiometric** reactions (number of molecules in and out); conservation of mass used to solve for reaction **fluxes** through system
- Emphasis on what's not feasible (invalid reactions, unreachable flux ranges); explore **bounded solution space** and simulate effect of perturbations within it



Network inference

Learning networks & network activity from data

- Relationships **inferred** from data or depicted in simplified (not strictly stoichiometric) ways
- Relationships between molecules, cells, etc. defined as **binary events** (binding), **association** (correlation, mutual information), **influence** (activation/inhibition) – or other forms, depending on the question
- Often at least some biological knowledge (e.g., TF binding motifs) used to **constrain** statistical inference

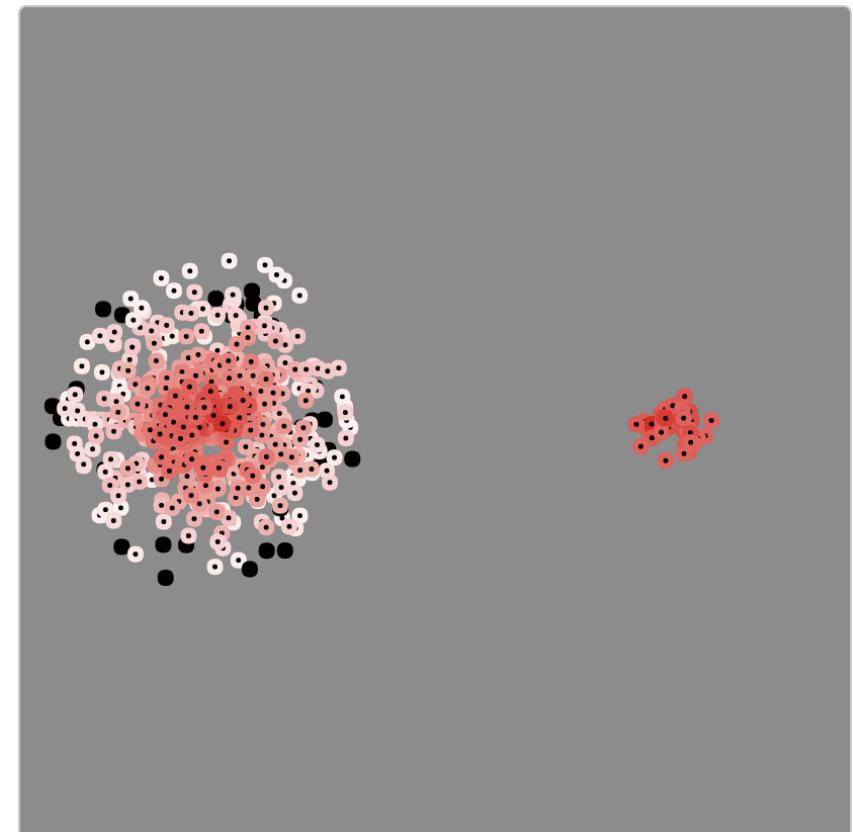


Agent-based modeling

Emergent phenomena from interacting agents

- **Agents:** units with defined rules governing their behavior or response – can interact with other agents or environment
- Systems simulated in **discrete** time and space
- Global behaviors or patterns arise through “self-organization” – often could not have been characterized based on knowledge of individual agents
- Powerful tool for modeling **communities** of molecules or cells

NetLogo Tumor model

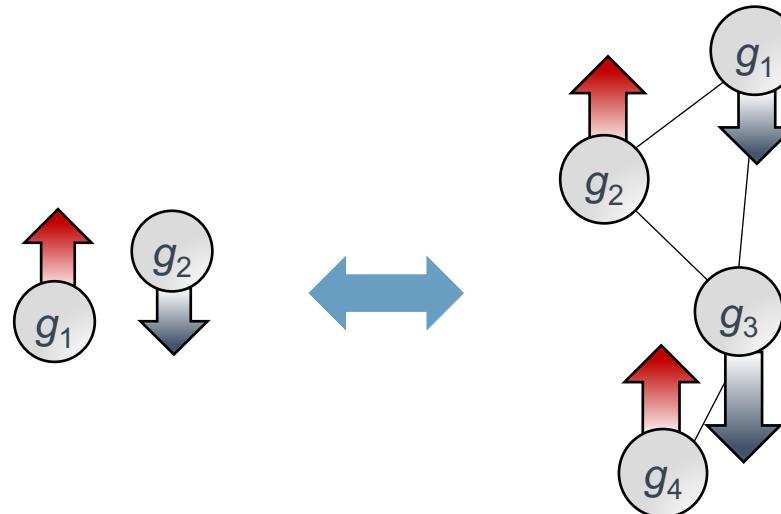


<http://ccl.northwestern.edu/netlogo/models/Tumor>

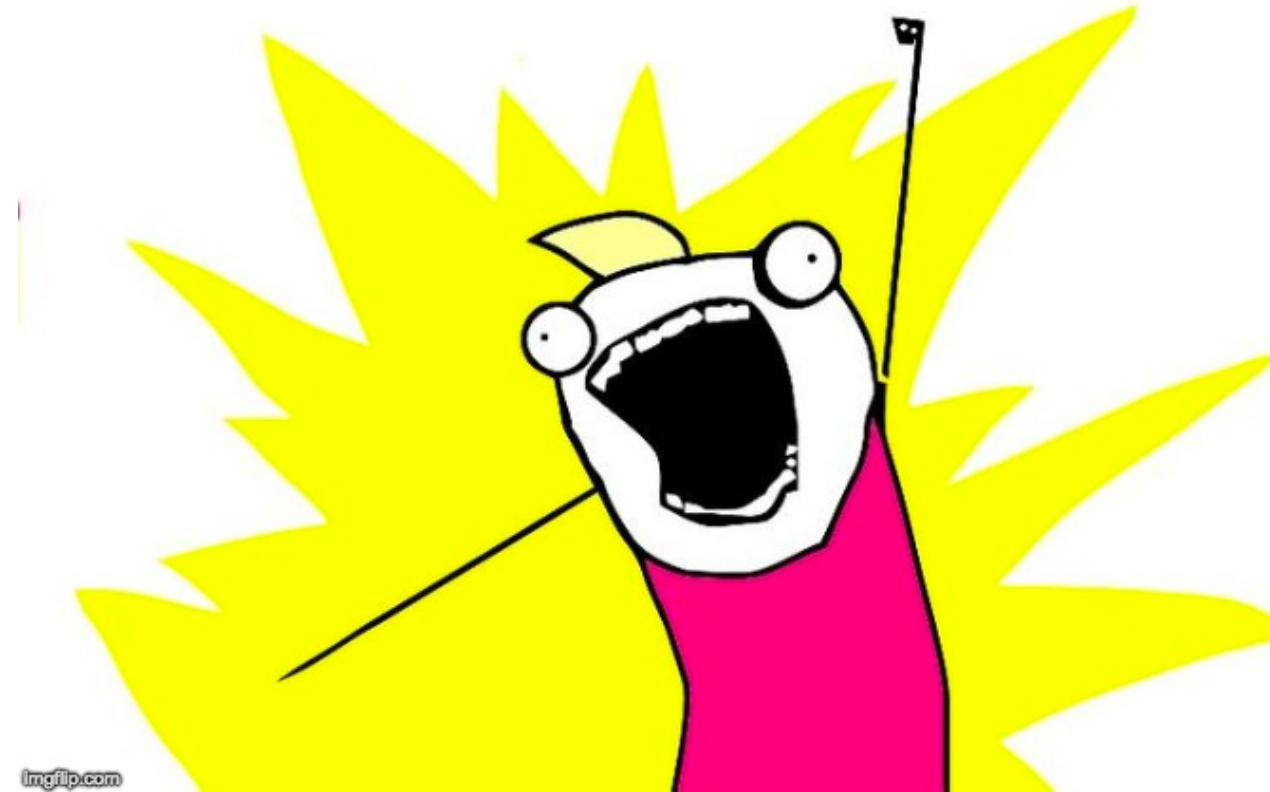
Multivariate statistical & machine learning

Modeling combinatorial patterns and interactions

- Rather than model the association between phenotypes and single variables (genes), use **weighted combination of molecular features**
- Statistical models tend to emphasize interpretation, while machine learning models emphasize predictive performance
- Both can be informed by known biology, or examined **in functional contexts** (e.g., pathways) after the fact



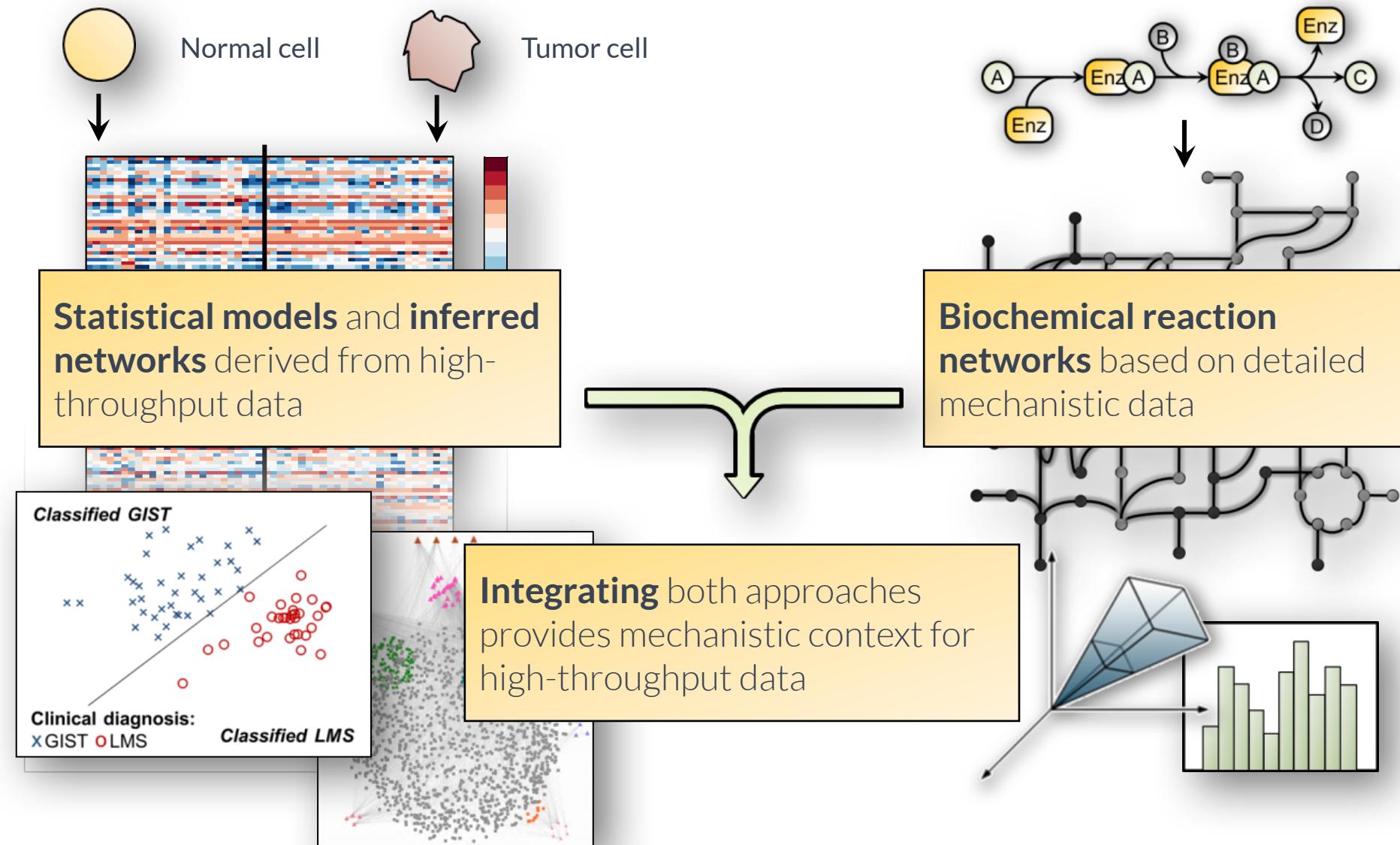
INTEGRATE ALL THE MODELS



imgflip.com

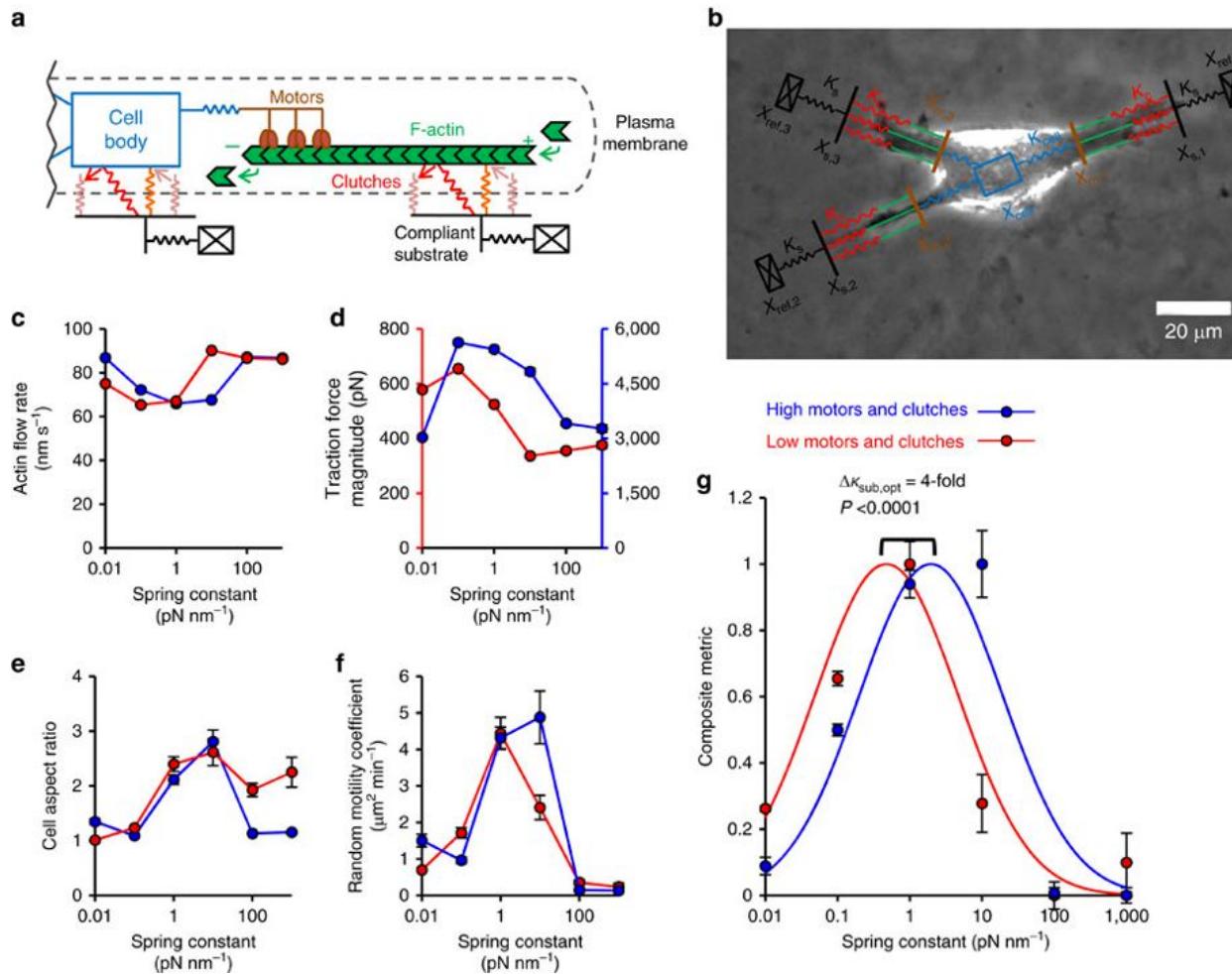
From integrative models to model integration

Statistical and mechanistic models



Statistical and mechanistic models

Cell Migration Simulator (CMS): mathematical model of cell adhesion dynamics



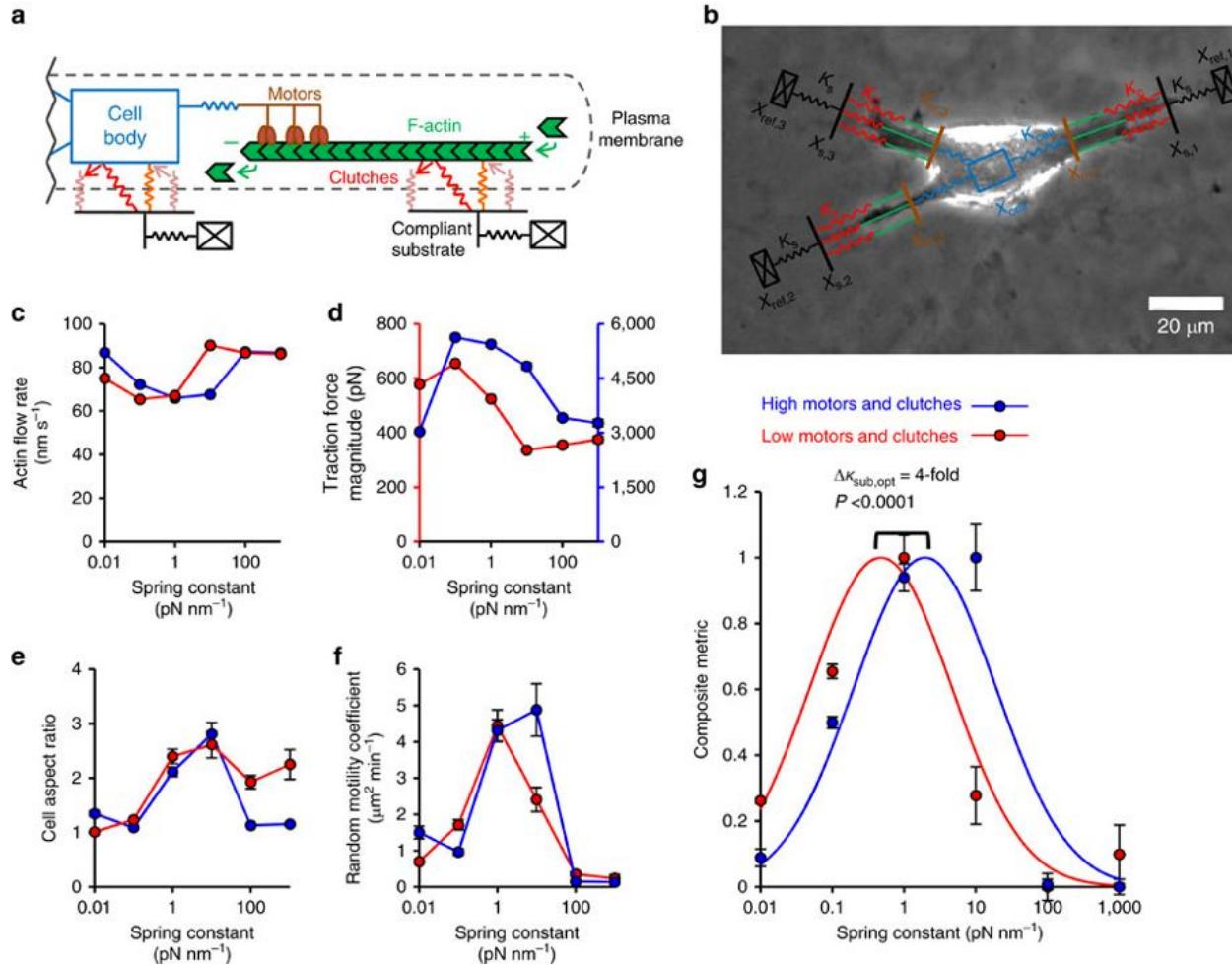
What are the steps that one might take to test model predictions experimentally, using the CMS predictions and the PS-ON data?

Q1. Using the CMS results in **Figure 6f** (regarding random motility coefficient, i.e. migration speed), identify a condition or set of conditions from the PSON motility data set that may be used to test the model prediction.

Q2. Make a hypothesis for a trend you would expect to see in your experimental condition(s) based on the CMS motility results (**Figure 6f**).

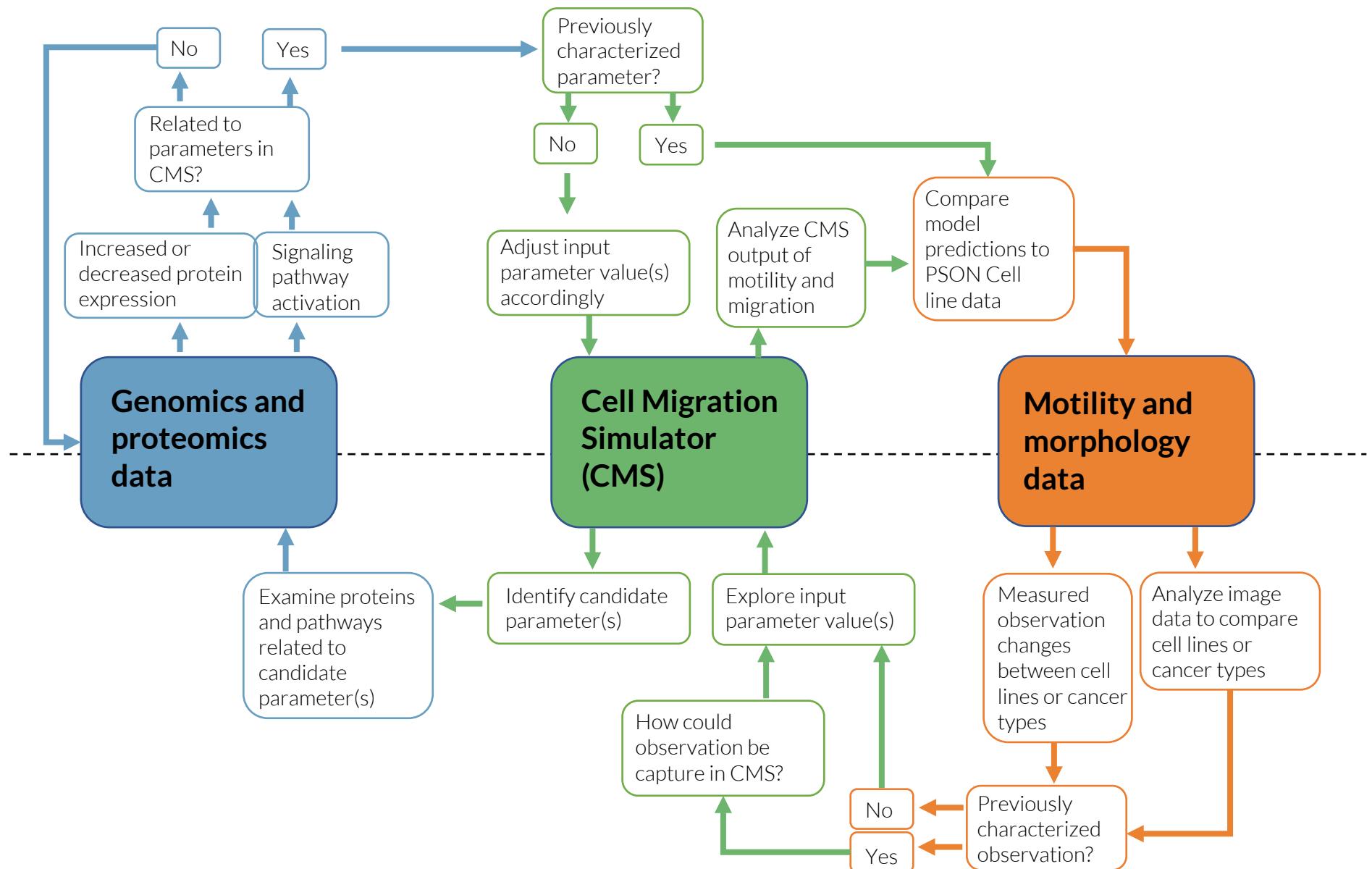
Statistical and mechanistic models

Cell Migration Simulator (CMS): mathematical model of cell adhesion dynamics



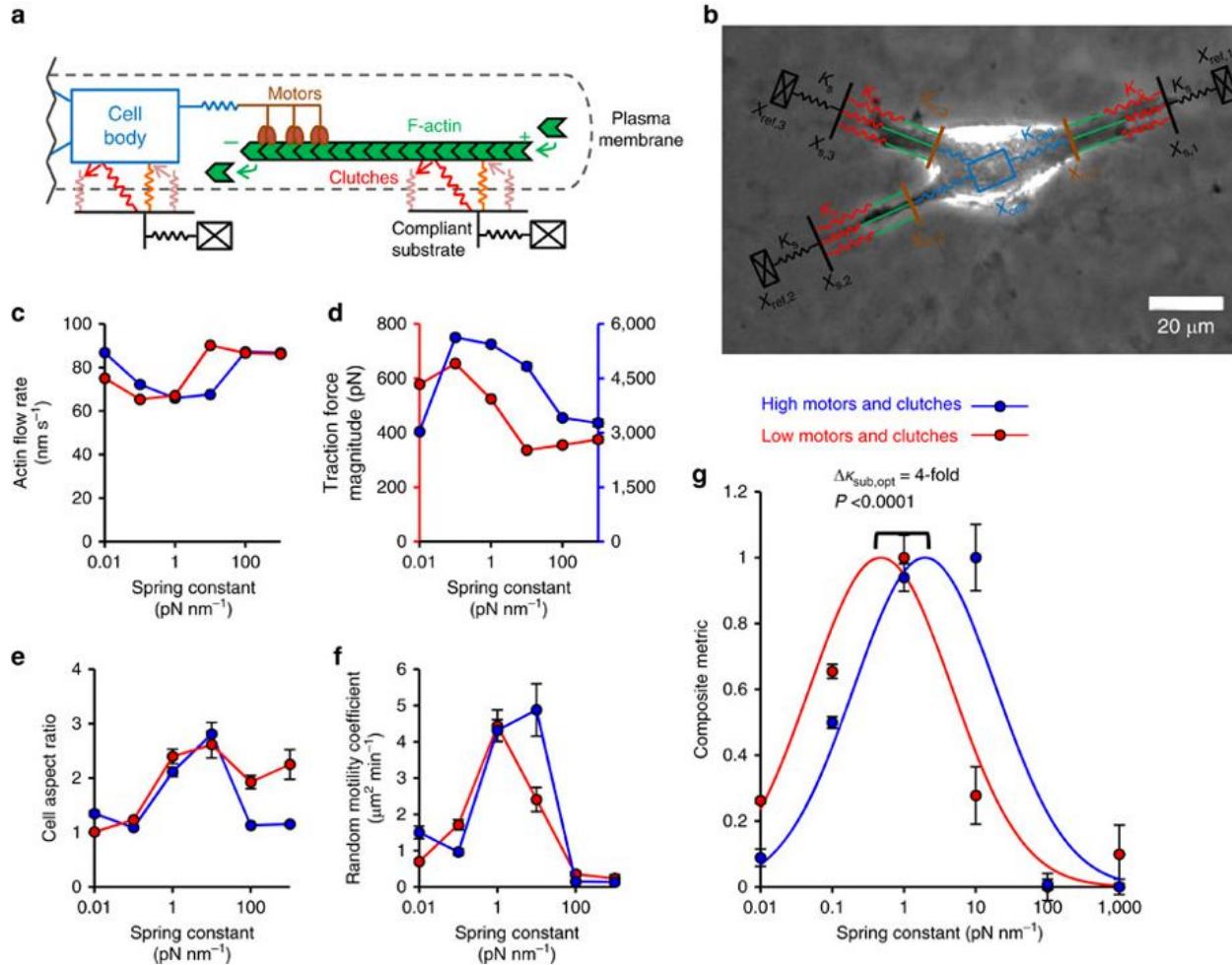
- The number of motors/clutches (high vs. low) – as represented by blue and red lines in the CMS plots – can be estimated by the expression of associated genes or pathways, in particular, those related to myosin and cell adhesion proteins.
- Spring force is representative of the stiffness of the experimental growth substrate/surface. For example some of the PS-ON data is on 500 Pa or 30 kPa, providing two different stiffness conditions. Glass is stiffer than both hyaluronic acid and polyacrylamide. (Also worth noting: the published CMS results were for "polyacrylamide hydrogels," which is one of the conditions included in the PS-ON data – so an attractive choice!).
- The motility coefficient is a summary metric than can be considered analogous to speed/distance from the PS-ON motility imaging data.

Starting from omics data



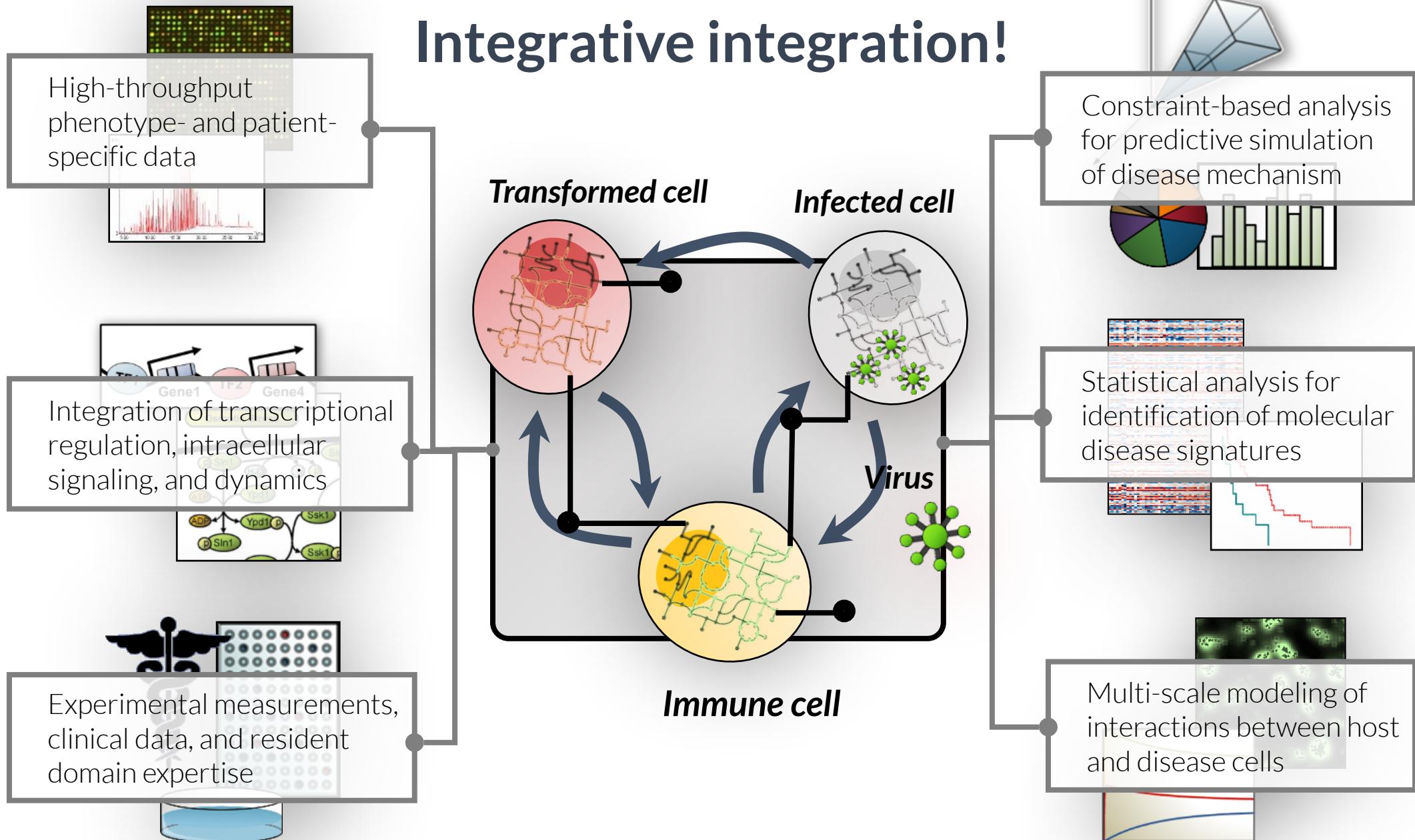
Next time!

Cell Migration Simulator & PS-ON modeling activity



Exploring our hypotheses
using PS-ON data and
statistical analysis in R!

Integrative integration!



Takeaways

- Systems biology complements cell biology – it doesn't replace it
- Top-down and bottom-up modeling approaches each have advantages, depending on context and data available
- Idealized cycle includes experimental investigation, data generation, systems-level modeling, and mechanistic characterization (rinse and repeat as needed)
- Don't get tripped up or discouraged by semantics and naysayers
- Data* — and the question you're trying to ask with it — are always the most important things!

* *open science, data sharing means better models for everyone!*



Thanks!

Questions?

Email: james.eddy@sagebionetworks.org

Twitter: @jamesaeddy

Want to learn more?

Resources on the mini-DREAM Synapse page

(see *Additional Resources*)

[Systems Biology & Bioinformatics](#)

Courses, tutorials, and other resources to learn more about the use of mathematics, statistics, and programming for biomedical research applications.