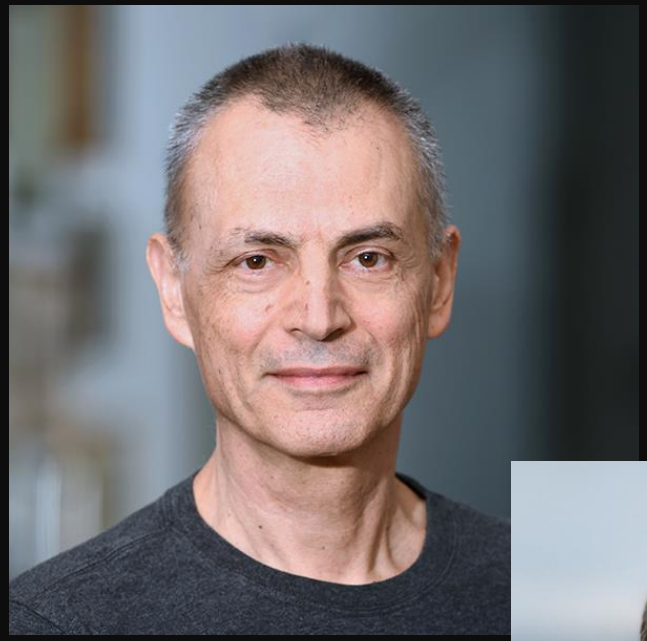
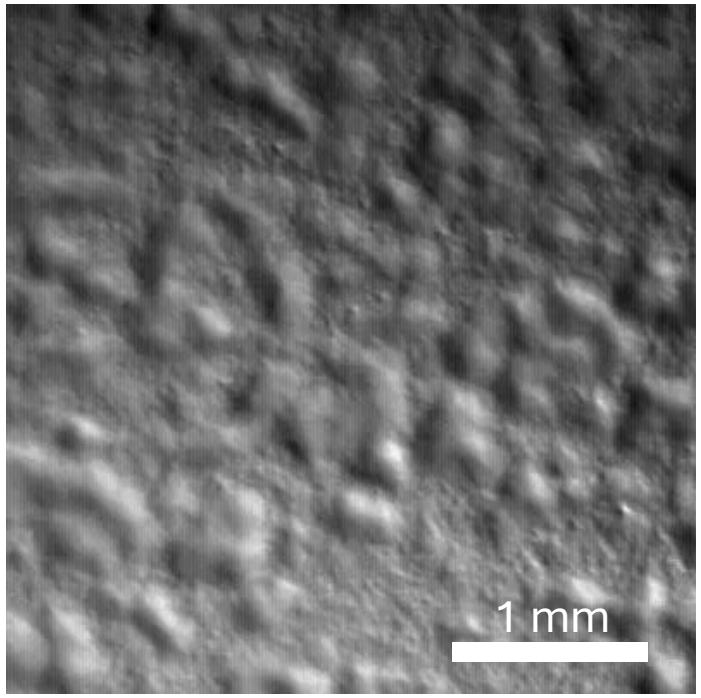
Deciphering *Dictyostelium* Biology: From Genetic Pathways to AI Systems for Knowledge and Discovery

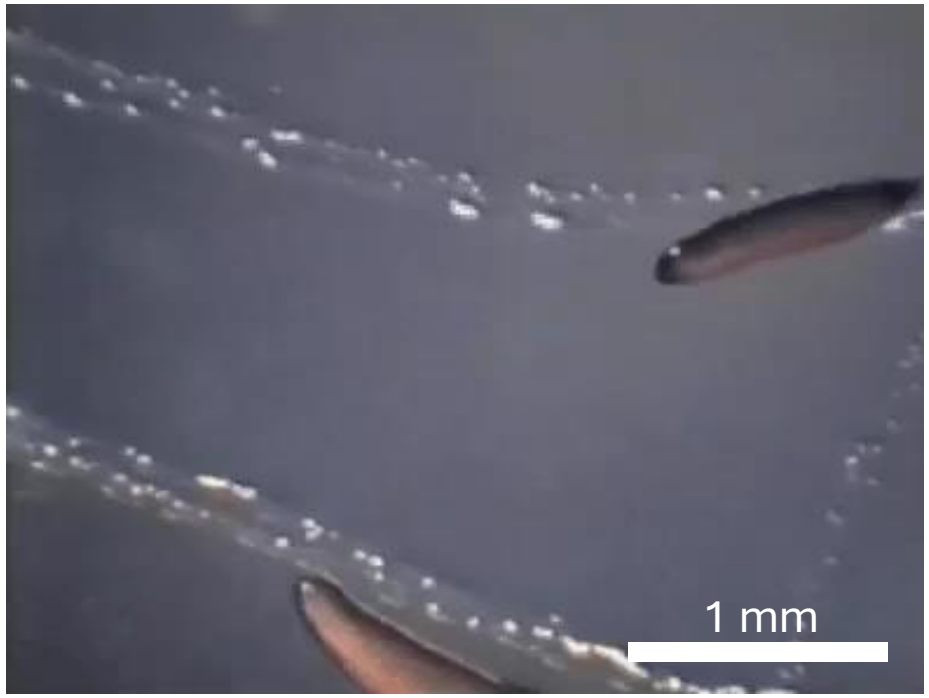
A collaborative journey of a biologist and a data scientist





Dictyostelium amoebae aggregate into mounds of thousands of cells





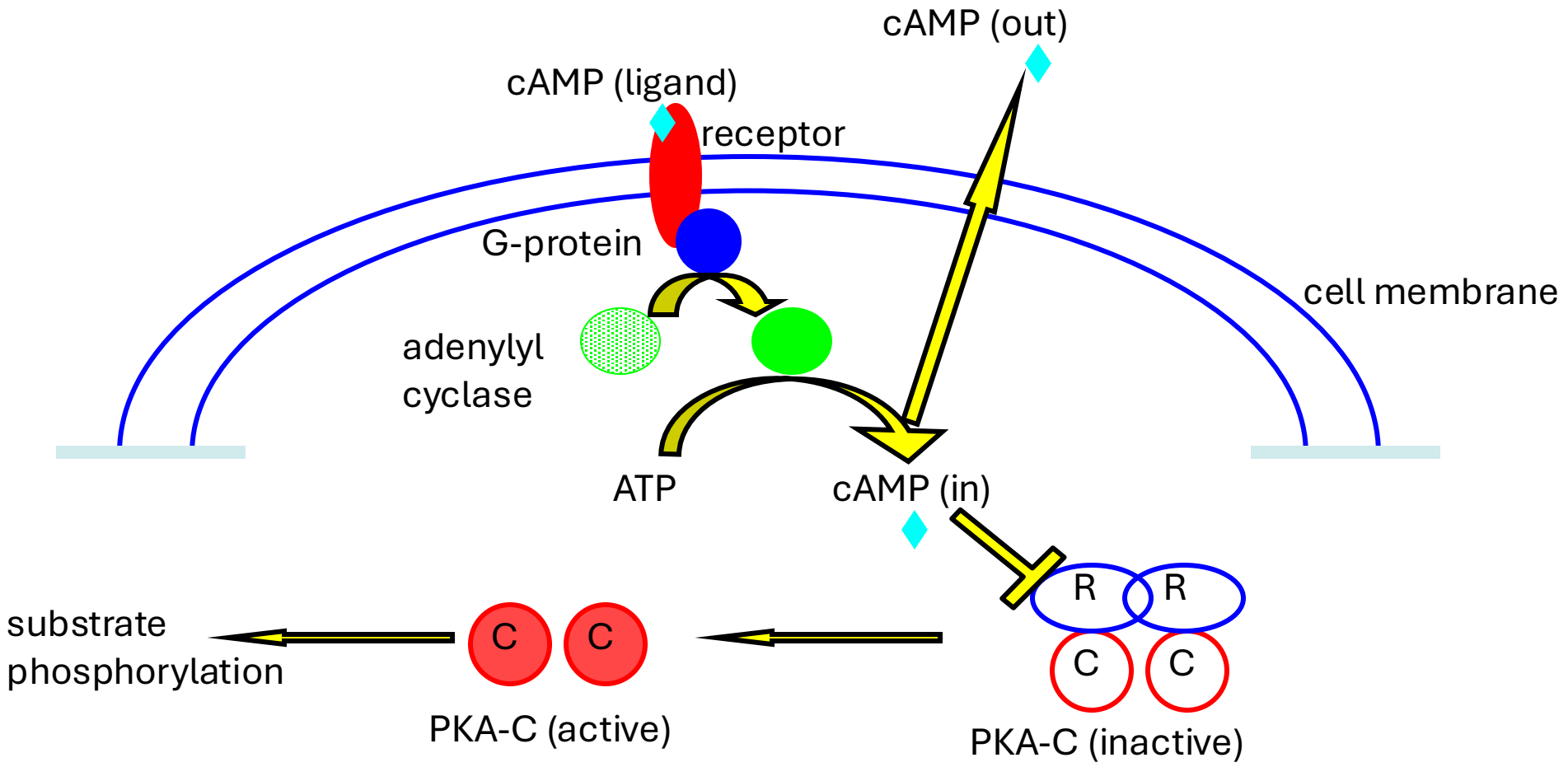
cAMP chemotaxis



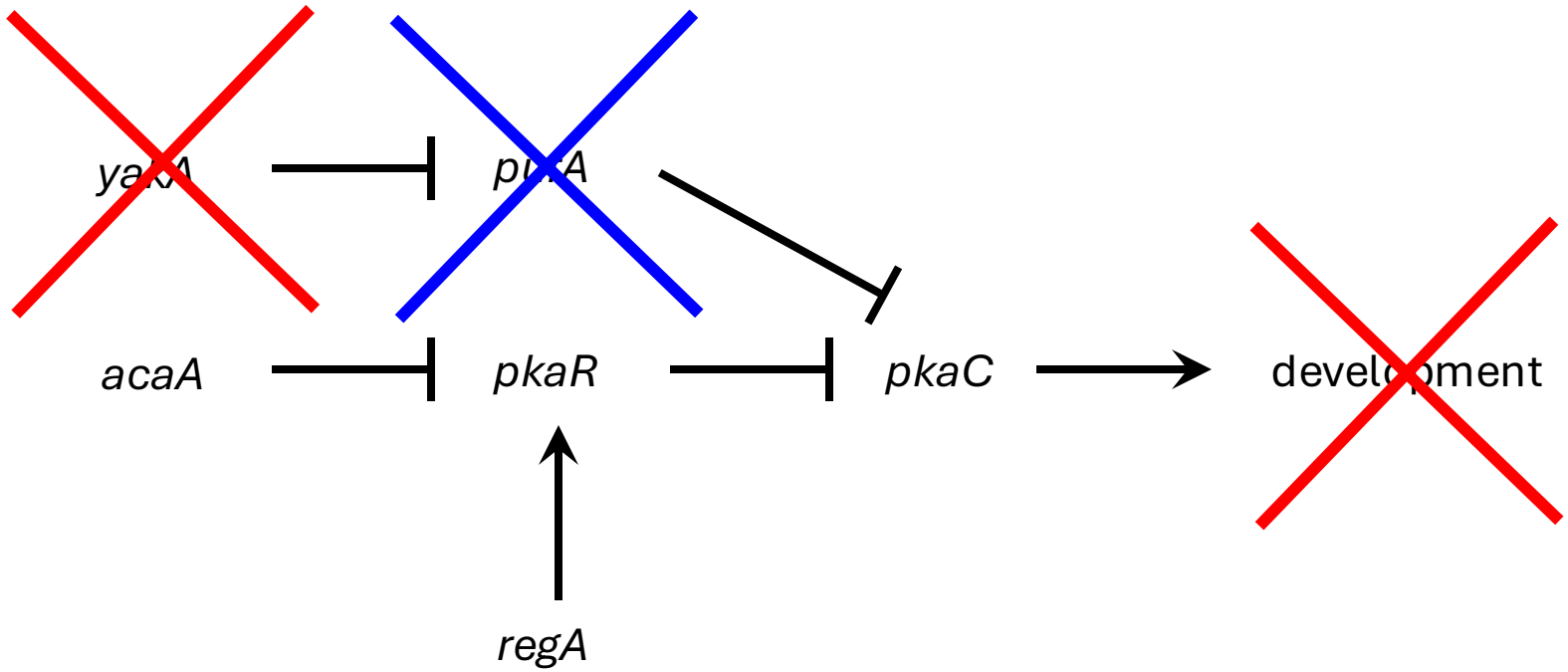


Movies by S. Lee and R. Firtel, UCSD

cAMP signaling



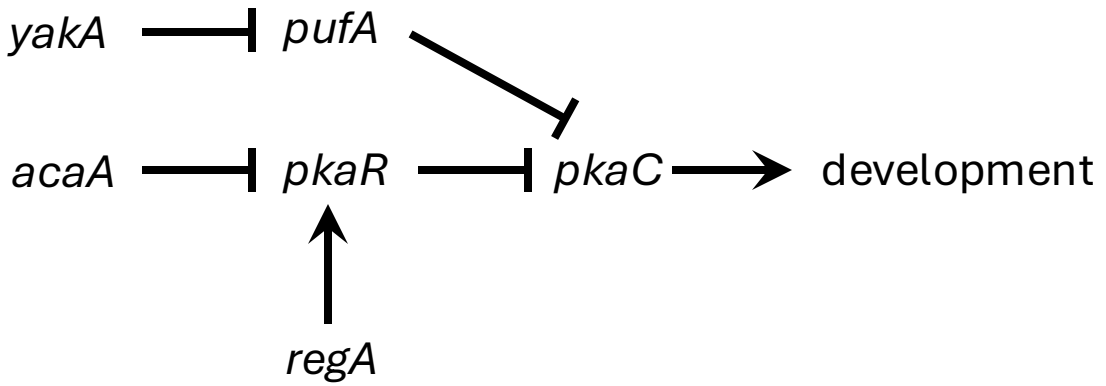
Epistasis in the PKA pathway



Souza, G. M., Lu, S. & Kuspa, A. *Development* **125**, 2291-2302 (1998).
Souza, G. M., da Silva, A. M. & Kuspa, A. *Development* **126**, 3263-3274 (1999).

Epistasis rules (simplified)

Mut A - P_A *yakA*$^-$ - no development
Mut B - P_B *pufA*$^-$ - yes development
Mut AB - P_{AB} *yakA*$^-$ *pufA*$^-$ - yes development



In a linear signaling pathway:

If $P_{AB} = P_B$ then B is epistatic ($A \longrightarrow B$)

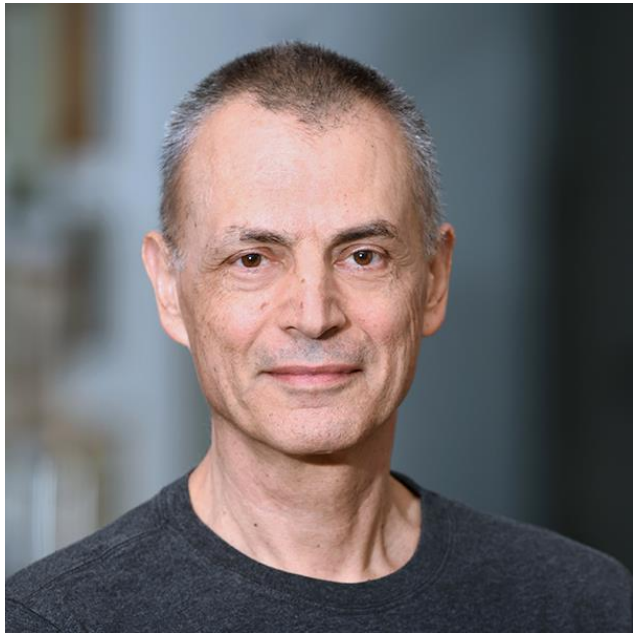
If ($P_{AB} \neq P_A$) and ($P_{AB} \neq P_B$) then A||B (parallel paths)

Can a computer do it?



Can I try?
What is epistasis? What is it good for?
What is a gene?
What is a protein?
What is a mutation?
What is a phenotype? Is it always true?
How many possible pathways are there?
Are genes always in pathways?
How to you measure a phenotype?

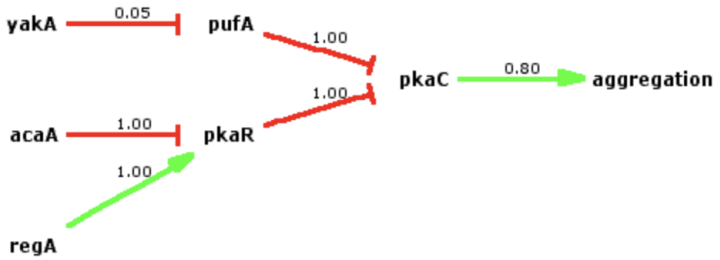
I don’t think anybody tried



OK

Yes, a computer can do epistasis analysis

New/Open | Save/Close | Options | Description | Genes | Prior Knowledge | Experimental Data | Expression Data
Relations | Plan Experiments | Genetic Network | What If Analysis | Notebook
Dictyostelium Aggregation | Dictyostelium Aggregation



Experimental data:

ID	Gene 1	Gene 2	aggregation	Confidence	Comments	Ignore	Edit
E1			+	1.00		I	E
E2	yakA-		-	0.50		I	E
E3	pufA-		++	0.50		I	E
E4	pkaR-		++	0.50		I	E
E5	pkaC-		-	0.50		I	E
E6	acaA-		-	0.50		I	E
E7	regA-		++	0.50		I	E
E8	acaA+		++	0.50		I	E
E9	pkaC+		++	0.50		I	E
E10	pkaC-	regA-	-	0.20		I	E
E11	yakA-	pufA-	++	0.20		I	E
E12	yakA-	pkaR-	±	0.20		I	E
E13	yakA-	pkaC-	-	0.20		I	E
E14	pkaC-	yakA+	-	0.20		I	E
E15	yakA-	pkaC+	++	0.20		I	E

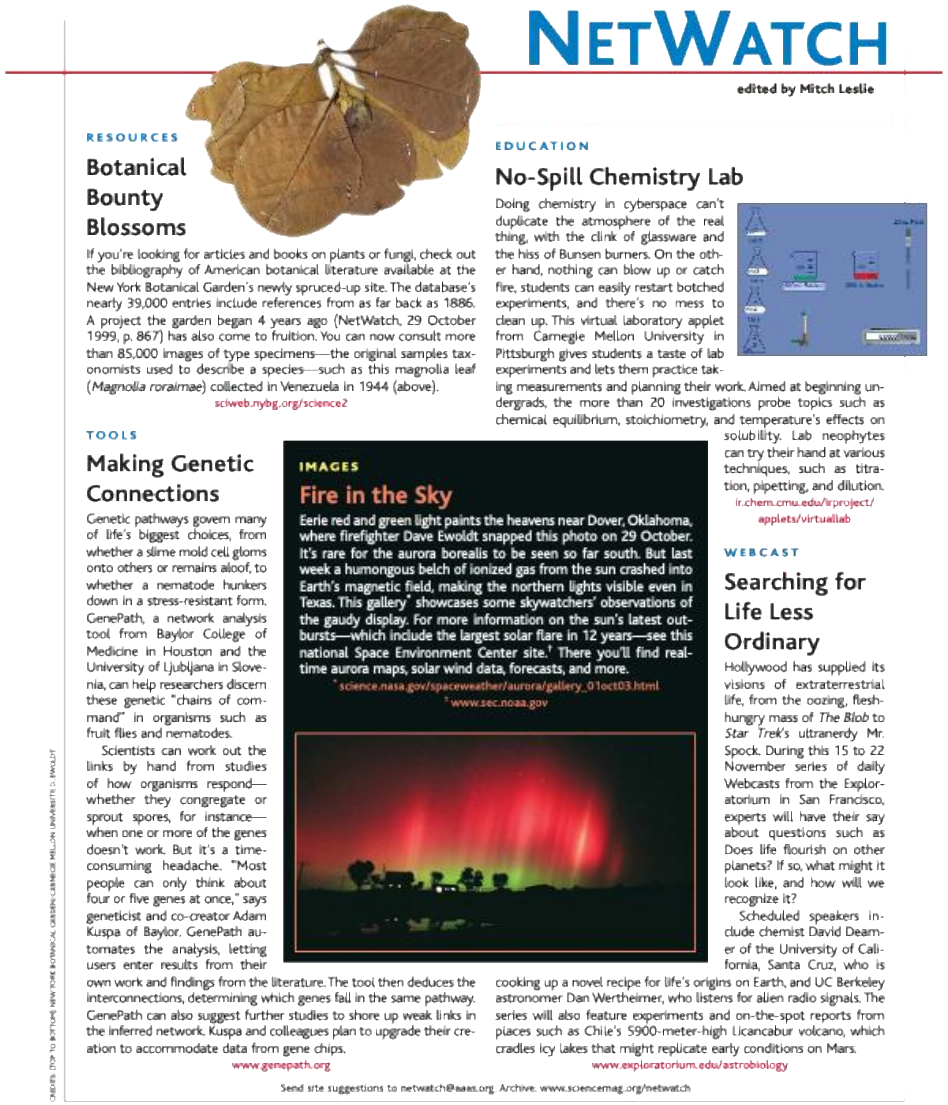
Prior knowledge:

ID	Gene	Influence	Bio. Entity	Confidence	Comments	Ignore	Edit
P1	pkaR	-|	pkaC	1.00		I	E
P2	acaA	-|	pkaR	1.00		I	E
P3	regA	->	pkaR	1.00		I	E
P4	pufA	-|	pkaC	1.00		I	E

www.genepath.org

Zupan B., et al., *Bioinformatics* 19(3), pp.383-389, 2003.
Zupan B., et al., *Artificial Intelligence in Medicine* 29(1-2), pp.107-130, 2003.

Epistasis is commonly used in biomedical sciences



NETWATCH

edited by Mitch Leslie

RESOURCES

Botanical Bounty Blossoms

If you're looking for articles and books on plants or fungi, check out the bibliography of American botanical literature available at the New York Botanical Garden's newly spruced-up site. The database's nearly 39,000 entries include references from as far back as 1886. A project the garden began 4 years ago (NetWatch, 29 October 1999, p. 867) has also come to fruition. You can now consult more than 85,000 images of type specimens—the original samples taxonomists used to describe a species—such as this magnolia leaf (*Magnolia roraimae*) collected in Venezuela in 1944 (above).

sciweb.nybg.org/science2

EDUCATION

No-Spill Chemistry Lab

Doing chemistry in cyberspace can't duplicate the atmosphere of the real thing, with the clink of glassware and the hiss of Bunsen burners. On the other hand, nothing can blow up or catch fire, students can easily restart botched experiments, and there's no mess to clean up. This virtual laboratory applet from Carnegie Mellon University in Pittsburgh gives students a taste of lab experiments and lets them practice taking measurements and planning their work. Aimed at beginning undergrads, the more than 20 investigations probe topics such as chemical equilibrium, stoichiometry, and temperature's effects on solubility. Lab neophytes can try their hand at various techniques, such as titration, pipetting, and dilution.

ir.chem.cmu.edu/irproject/applets/virtuallab

TOOLS

Making Genetic Connections

Genetic pathways govern many of life's biggest choices, from whether a slime mold cell gloms onto others or remains aloof, to whether a nematode hunkers down in a stress-resistant form. GenePath, a network analysis tool from Baylor College of Medicine in Houston and the University of Ljubljana in Slovenia, can help researchers discern these genetic "chains of command" in organisms such as fruit flies and nematodes.

Scientists can work out the links by hand from studies of how organisms respond—whether they congregate or sprout spores, for instance—when one or more of the genes doesn't work. But it's a time-consuming headache. "Most people can only think about four or five genes at once," says geneticist and co-creator Adam Kuspa of Baylor. GenePath automates the analysis, letting users enter results from their own work and findings from the literature. The tool then deduces the interconnections, determining which genes fall in the same pathway. GenePath can also suggest further studies to shore up weak links in the inferred network. Kuspa and colleagues plan to upgrade their creation to accommodate data from gene chips.

www.genepath.org

IMAGES

Fire in the Sky

Eerie red and green light paints the heavens near Dover, Oklahoma, where firefighter Dave Ewoldt snapped this photo on 29 October. It's rare for the aurora borealis to be seen so far south. But last week a humongous belch of ionized gas from the sun crashed into Earth's magnetic field, making the northern lights visible even in Texas. This gallery* showcases some skywatchers' observations of the gaudy display. For more information on the sun's latest outbursts—which include the largest solar flare in 12 years—see this national Space Environment Center site.† There you'll find real-time aurora maps, solar wind data, forecasts, and more.

* science.nasa.gov/spaceweather/aurora/gallery_01oct03.html
† www.sec.noaa.gov

WEBCAST

Searching for Life Less Ordinary

Hollywood has supplied its visions of extraterrestrial life, from the oozing, flesh-hungry mass of *The Blob* to *Star Trek*'s ultranerdy Mr. Spock. During this 15 to 22 November series of daily Webcasts from the Exploratorium in San Francisco, experts will have their say about questions such as Does life flourish on other planets? If so, what might it look like, and how will we recognize it?

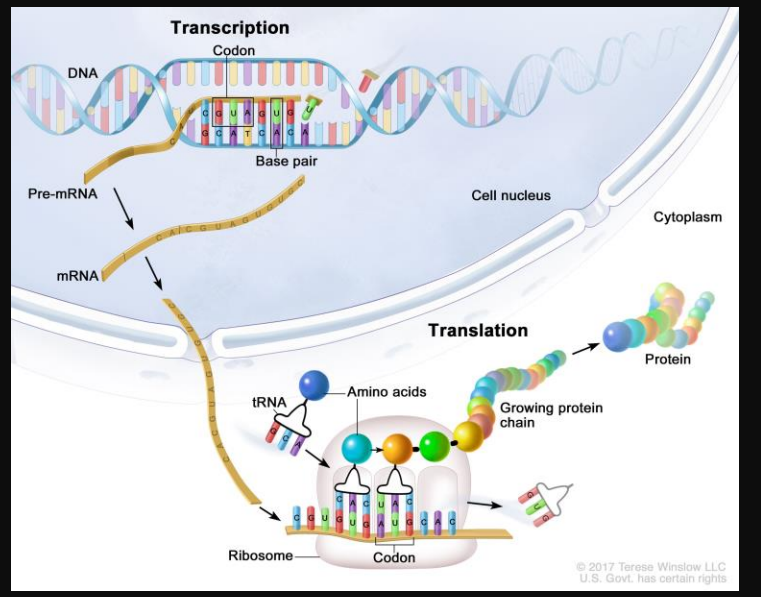
Scheduled speakers include chemist David Deamer of the University of California, Santa Cruz, who is cooking up a novel recipe for life's origins on Earth, and UC Berkeley astronomer Dan Wertheimer, who listens for alien radio signals. The series will also feature experiments and on-the-spot reports from places such as Chile's 5900-meter-high Licancabur volcano, which cradles icy lakes that might replicate early conditions on Mars.

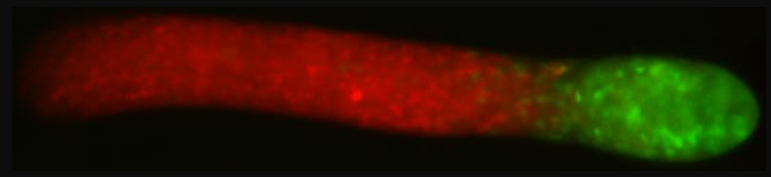
www.exploratorium.edu/astrobiology

Send site suggestions to netwatch@aaas.org. Archive: www.sciencemag.org/netwatch

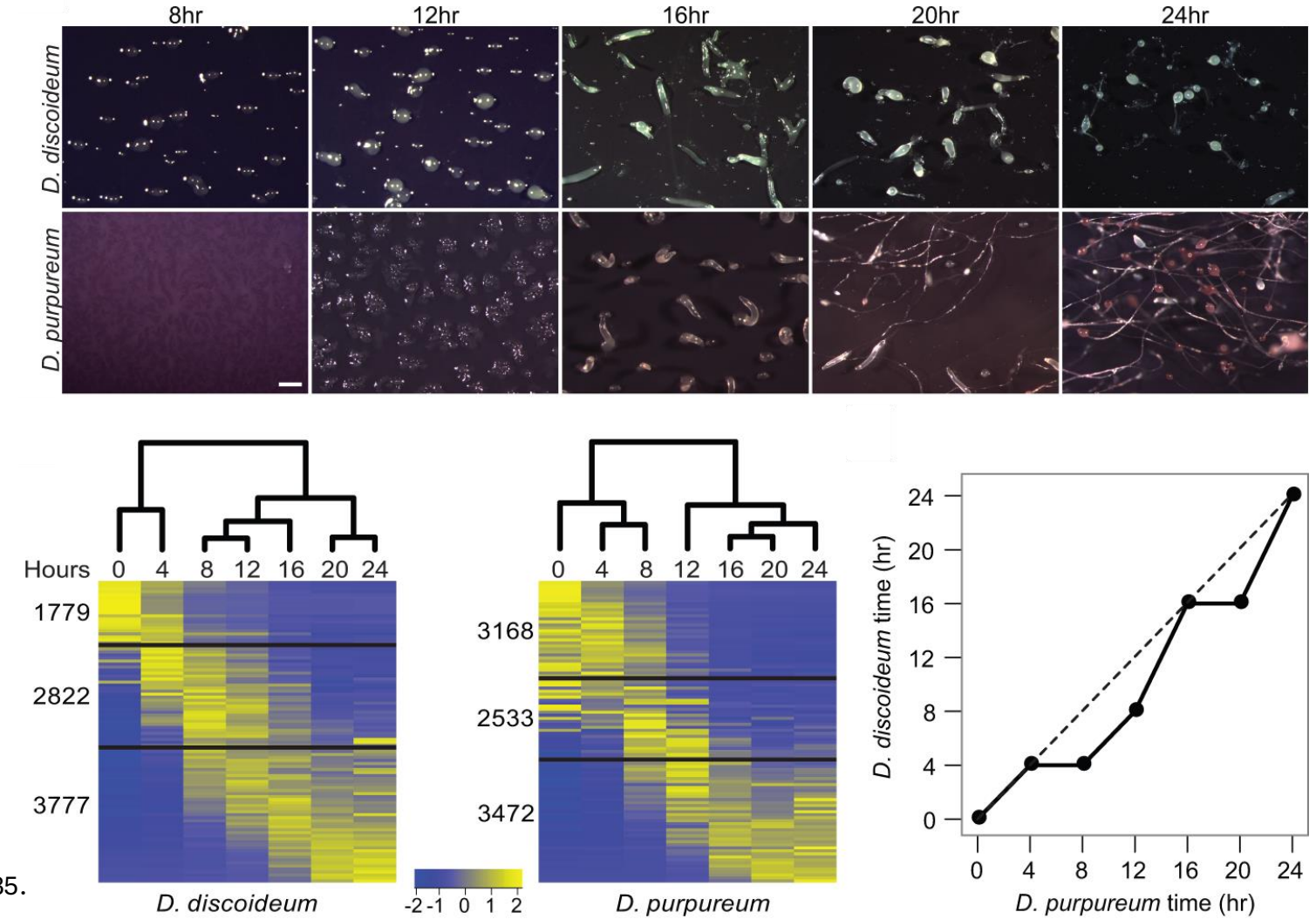
Transcriptome analysis

- *Dictyostelium* has 13,892 genes
- DNA – RNA – Protein
- Some mRNAs are always present
- Some are present under certain conditions



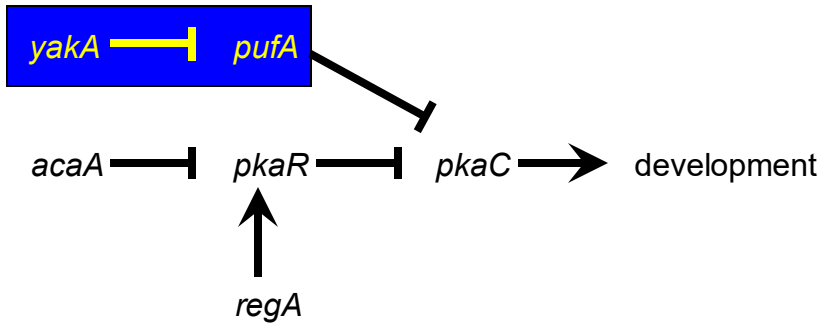


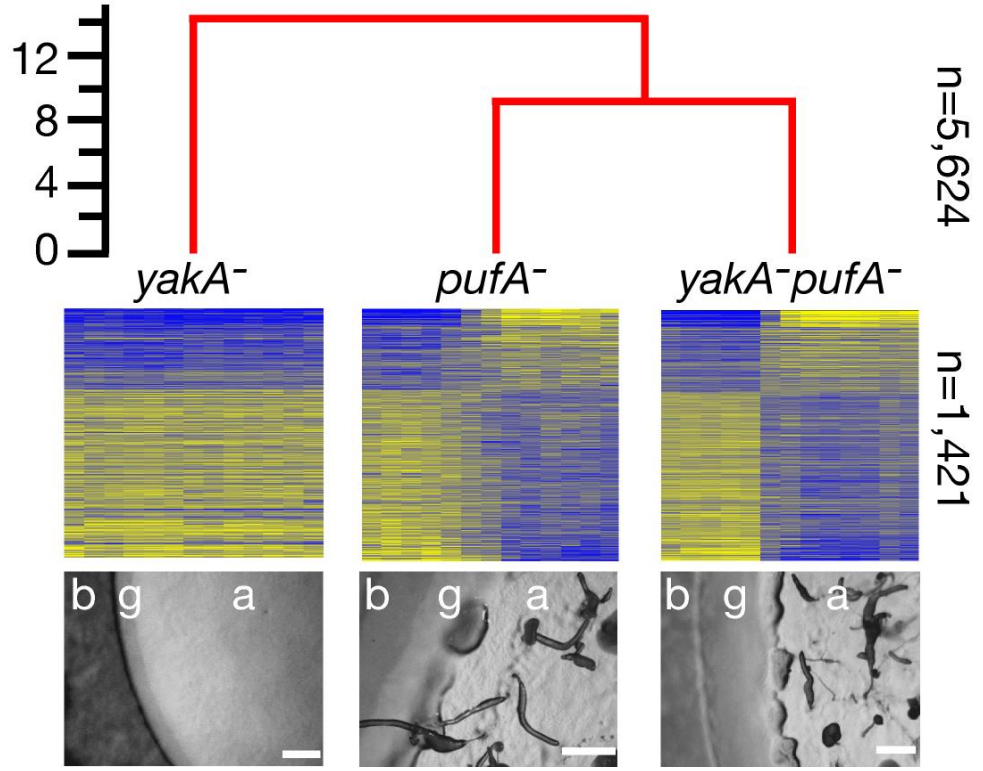
We can measure the expression of all the genes at once



Parikh et al. (2010). Genome Biol *11*, R35. 10.1186/gb-2010-11-3-r35.

Transcriptome phenotyping in epistasis analysis

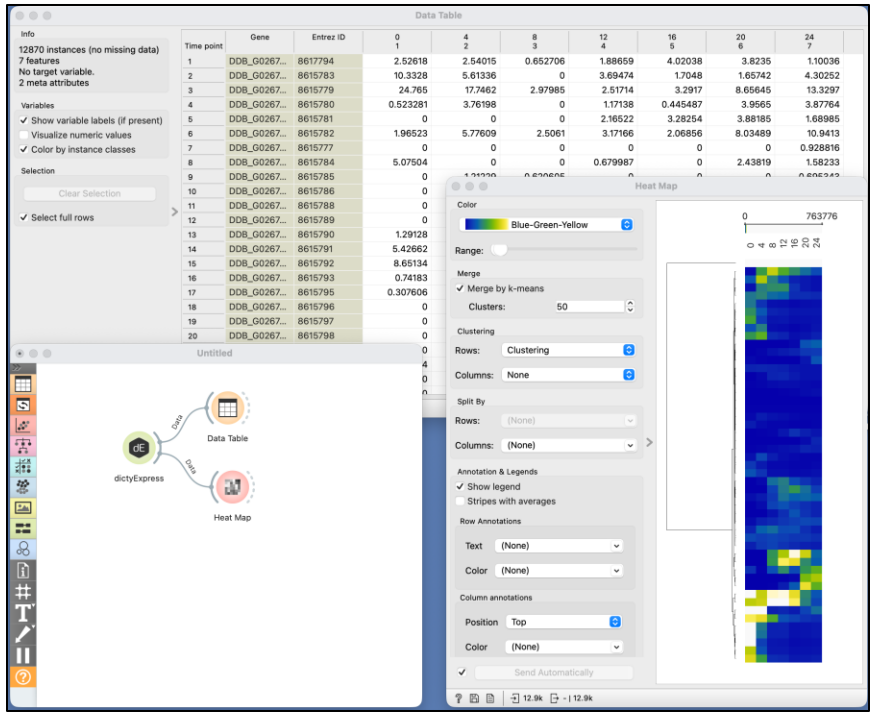




Van Driesshe et al., (2005). *Nature Genetics* 37, 471-477.

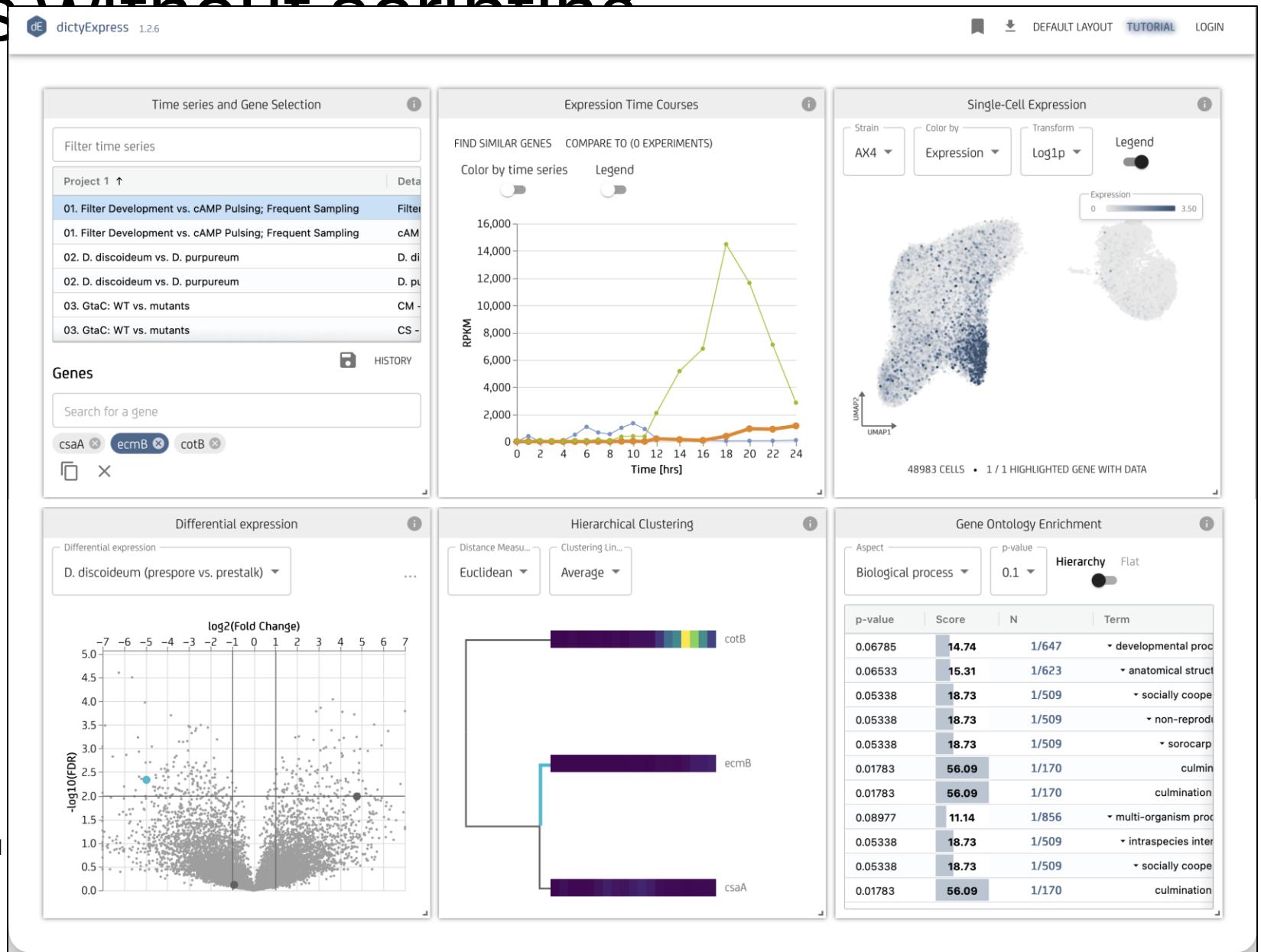
Visual analytics without scripting

Orange



Curk et al., (2005). Bioinformatics *21*, 396-398.

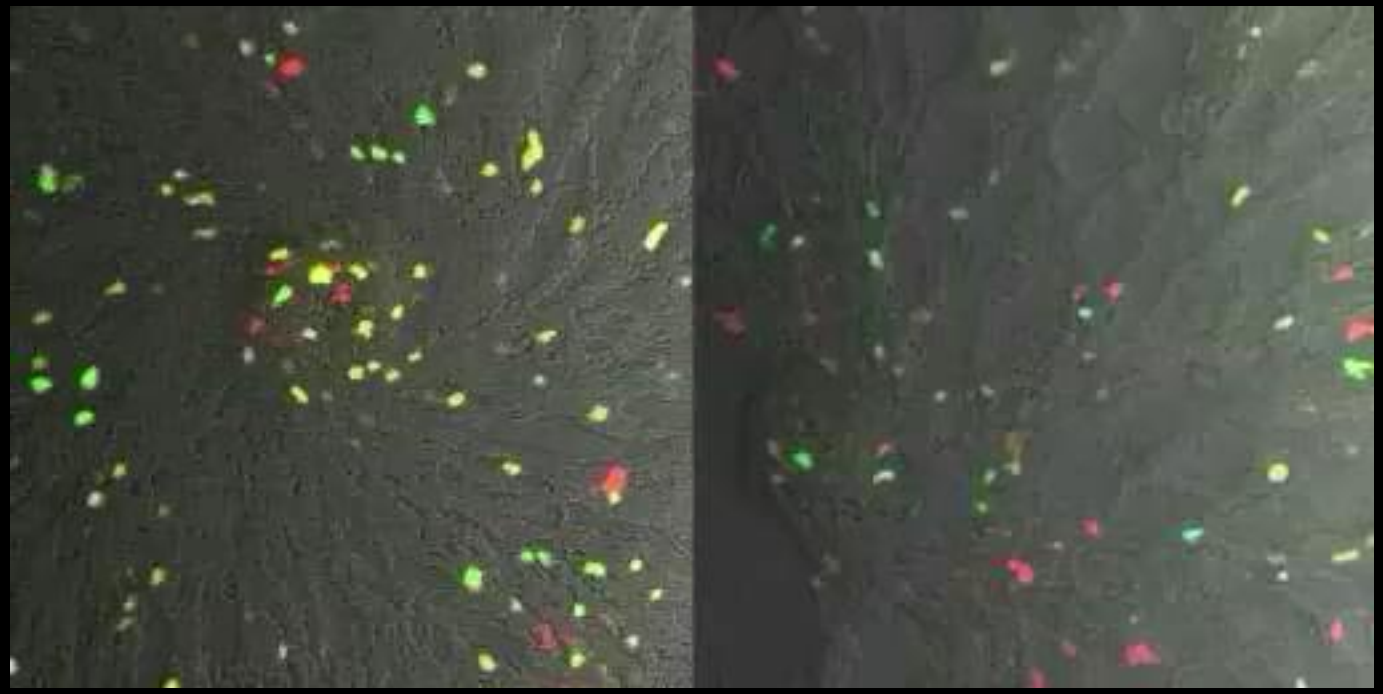
Rot et al., (2009) BMC Bioinformatics *10*, 265
Stajdohar et al., (2017). BMC Bioinformatics *18*, 291
Trnovec et al., submitted



Allorecognition

Compatible

Incompatible



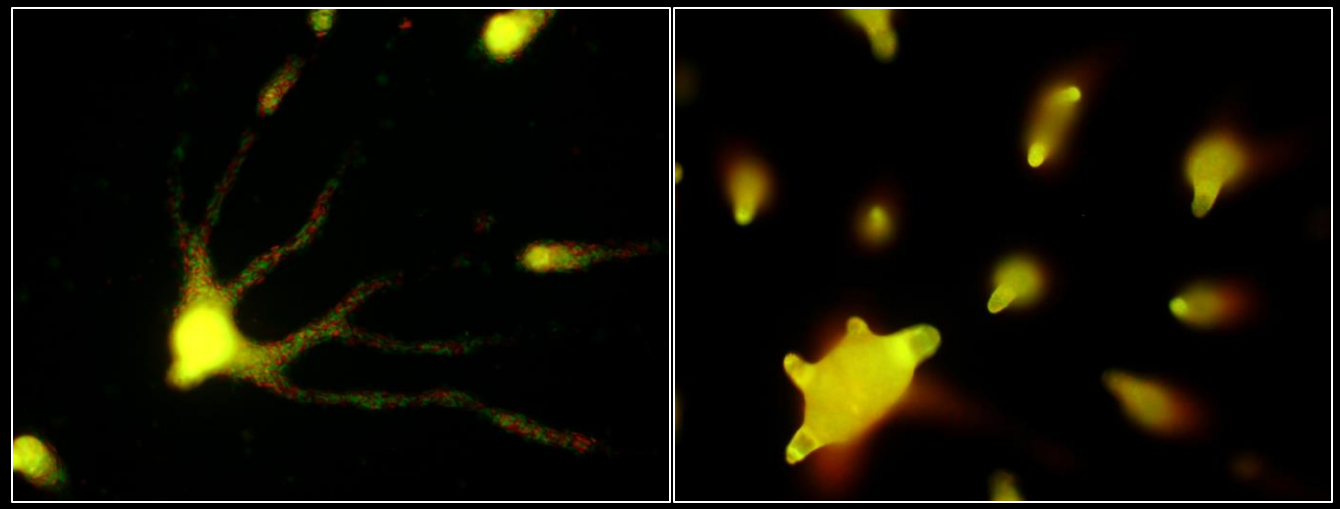
tgrB1^{AX4} *tgrC1*^{AX4}
+ AX4

tgrB1^{QS38} *tgrC1*^{QS38}
+ AX4

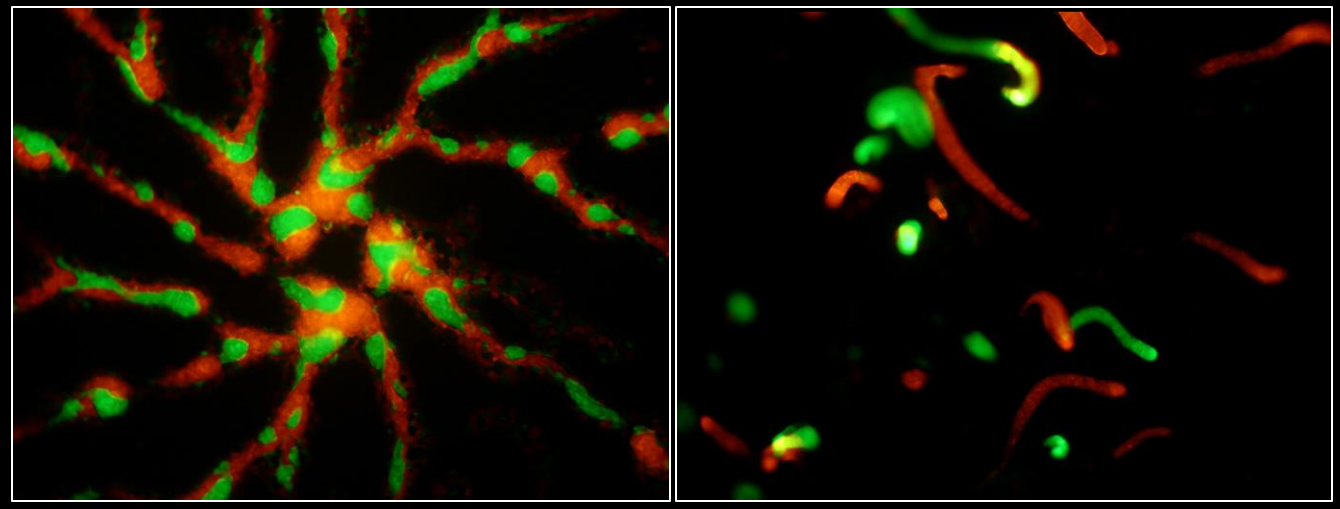
Hirose *et al.*, Science 2011

Allorecognition

Same allotype

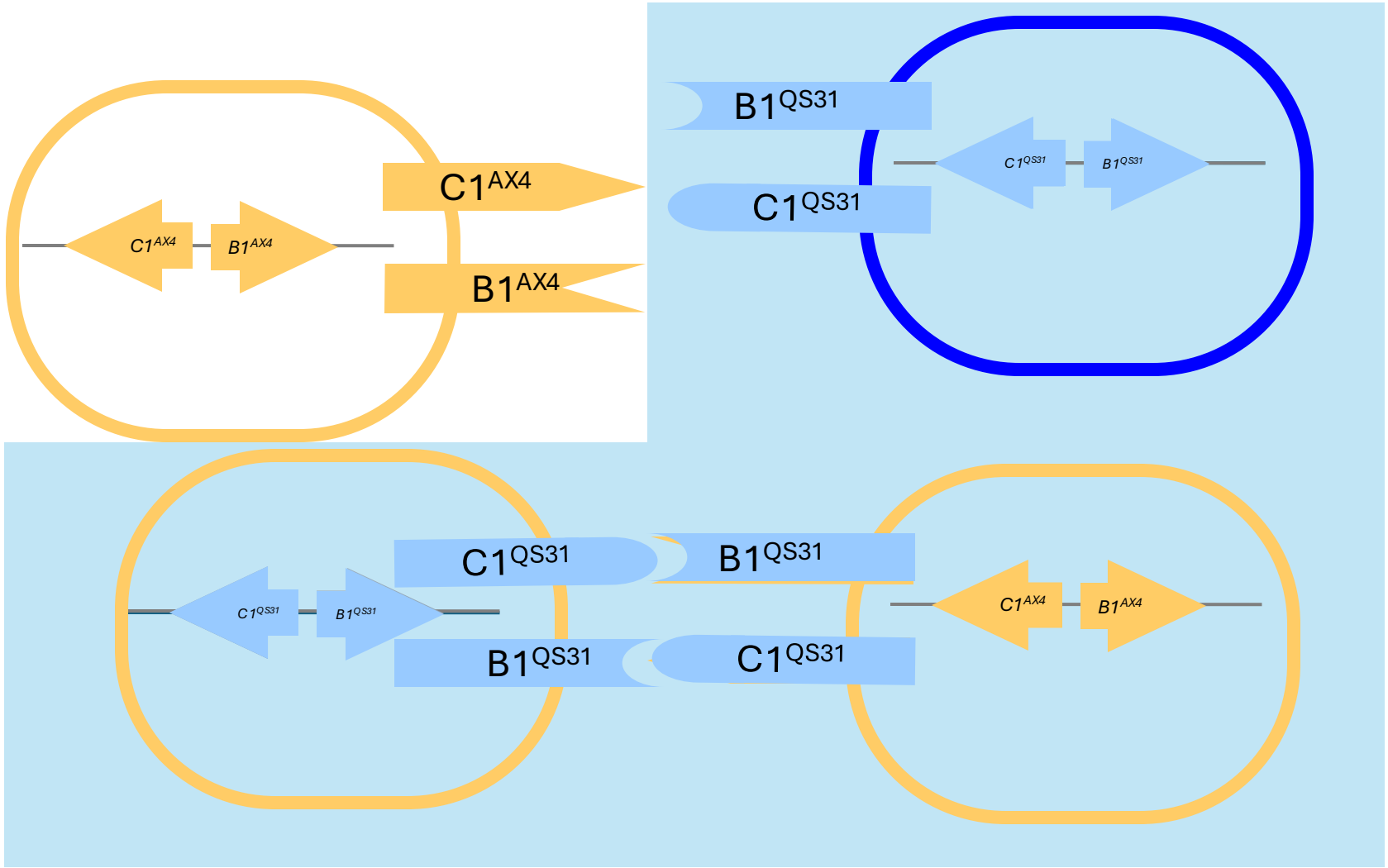


Different allotypes



Hirose *et al.*, Science 2011

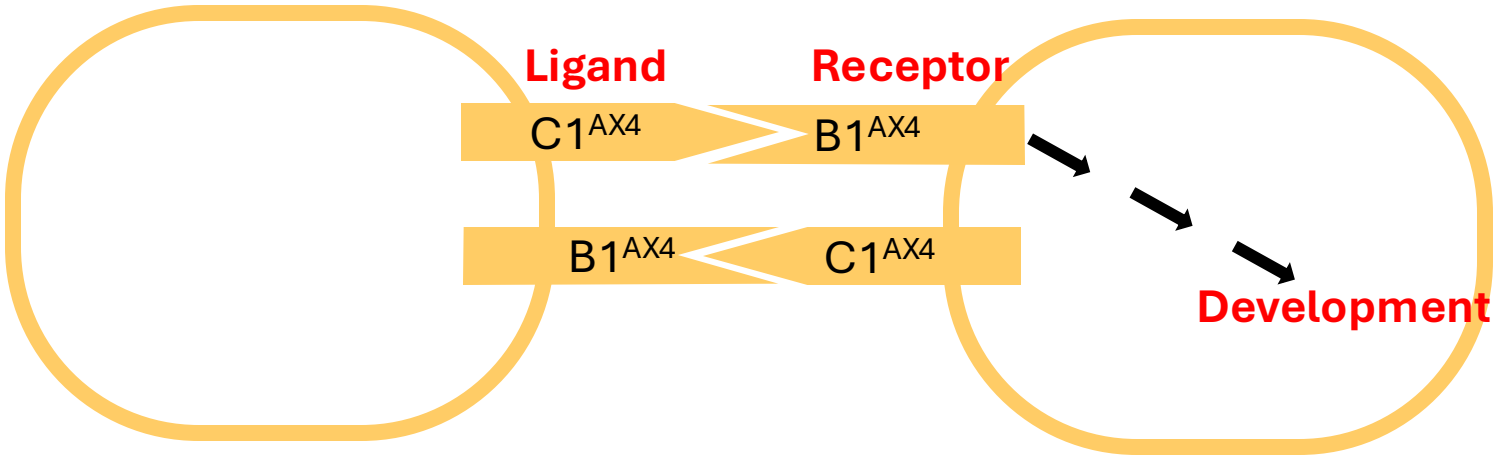
TgrB1 and TgrC1 mediate allorecognition



Hirose *et al.*, Science 2011

TgrB1-C1 signal transduction

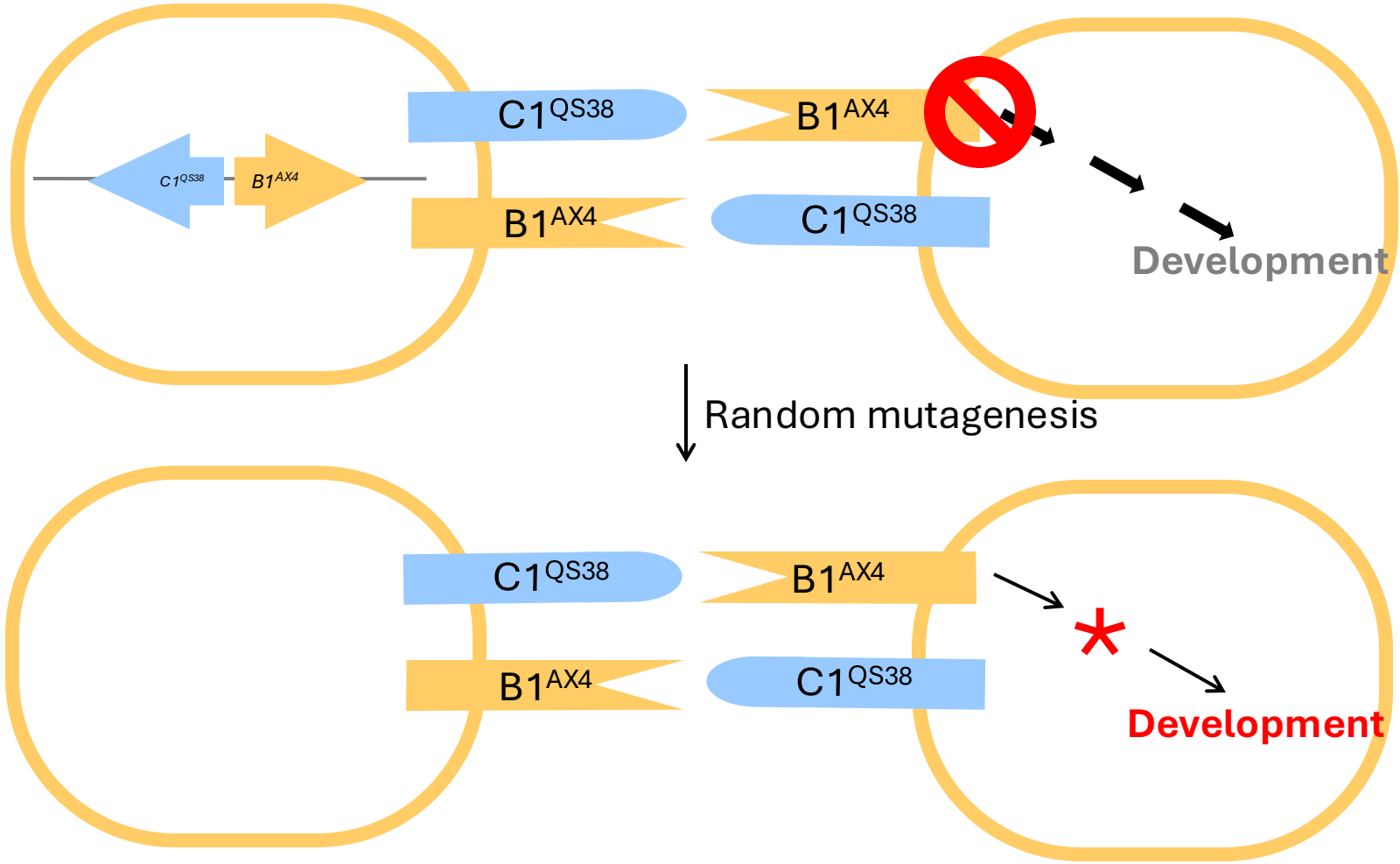
Ligand-Receptor interaction



Hirose *et al.*, J. Cell Sci. 2017

TgrB1-C1 signal transduction

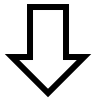
genetic screen for components



Li *et al.*, Genome Res. 2016

Suppressor screen

63 mutant strains



Whole-genome shotgun sequencing
~20X coverage



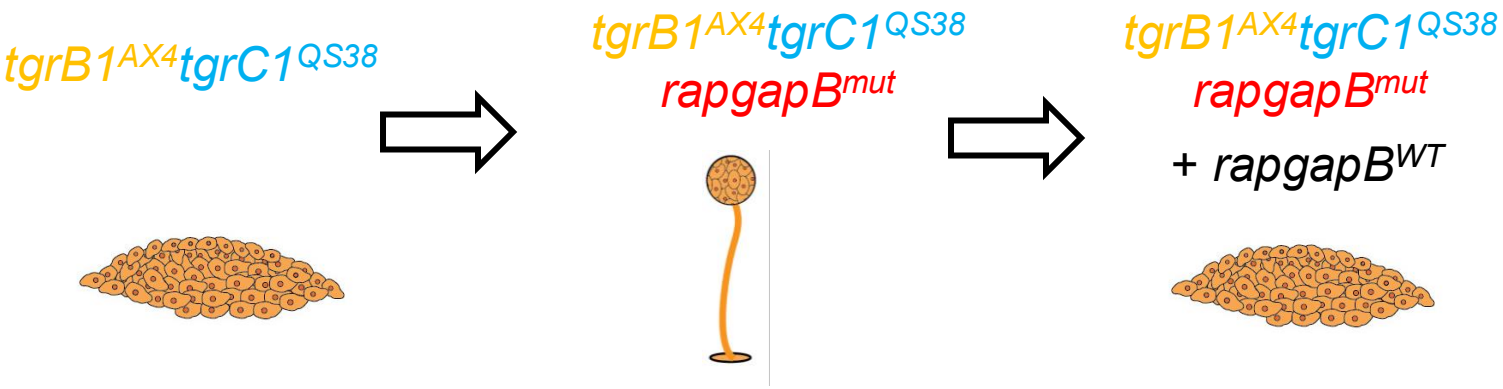
20 mutations per genome

rapgapB	19 strains,	17 alleles
tgrB1	9 strains,	6 alleles
others	>30 strains,	>2 alleles ea.

Li *et al.*, Genome Res. 2016

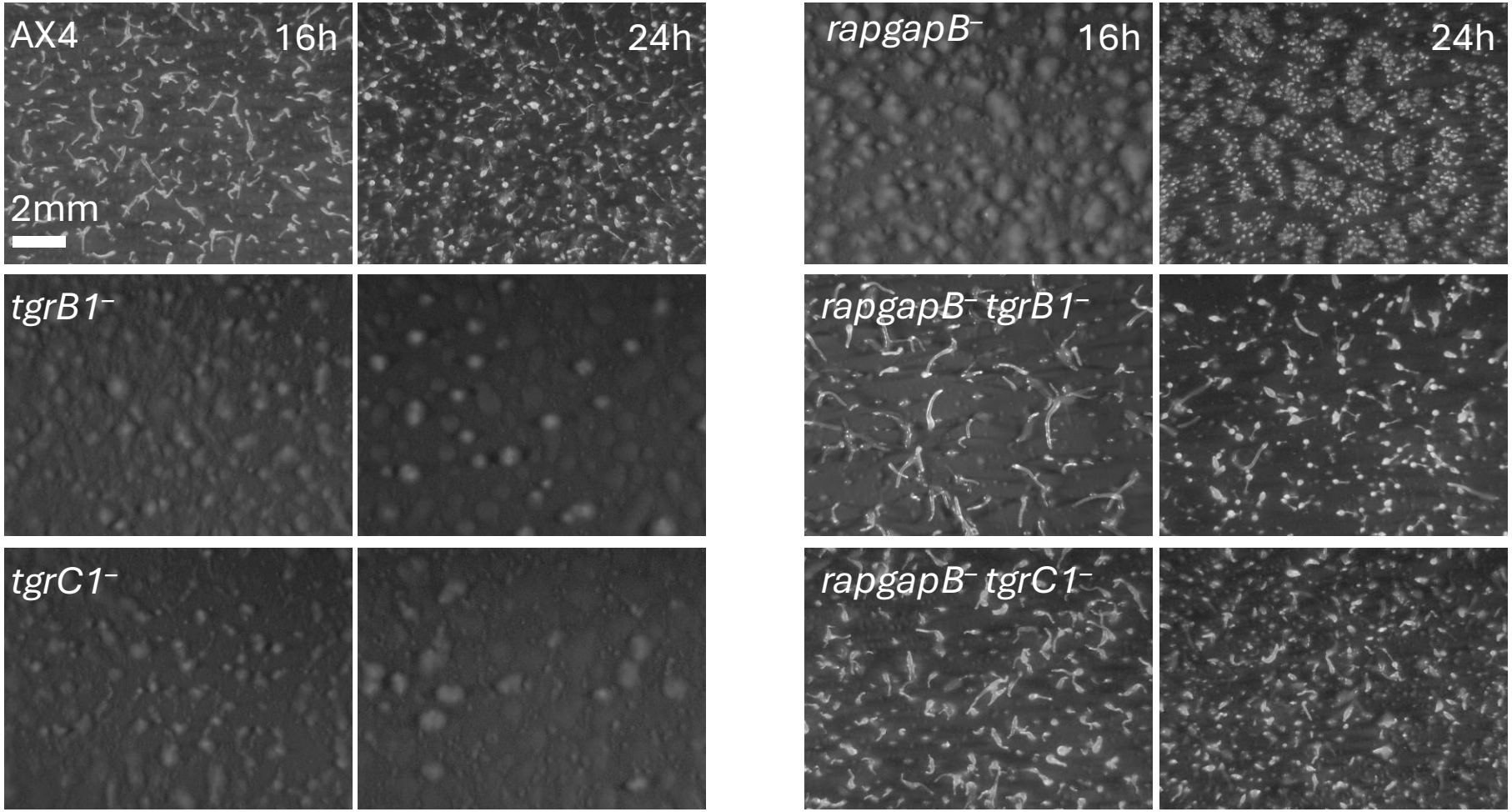
Causative mutations in *rapgapB*





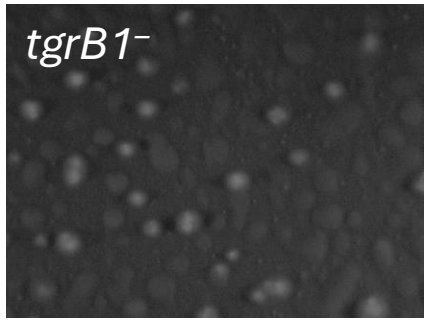
Li *et al.*, Genome Res. 2016

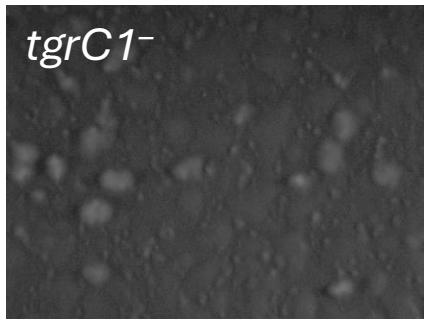
Mutual suppression between *tgrB1/C1* and *rapgapB*

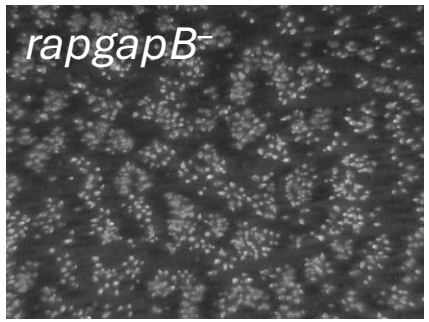


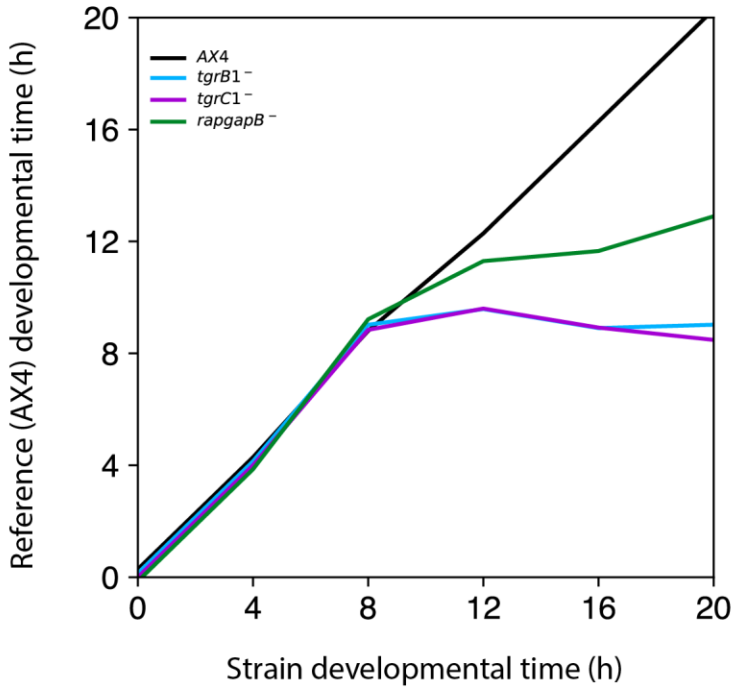
Lehmann *et al*. (2024). iScience, 27, 111125.

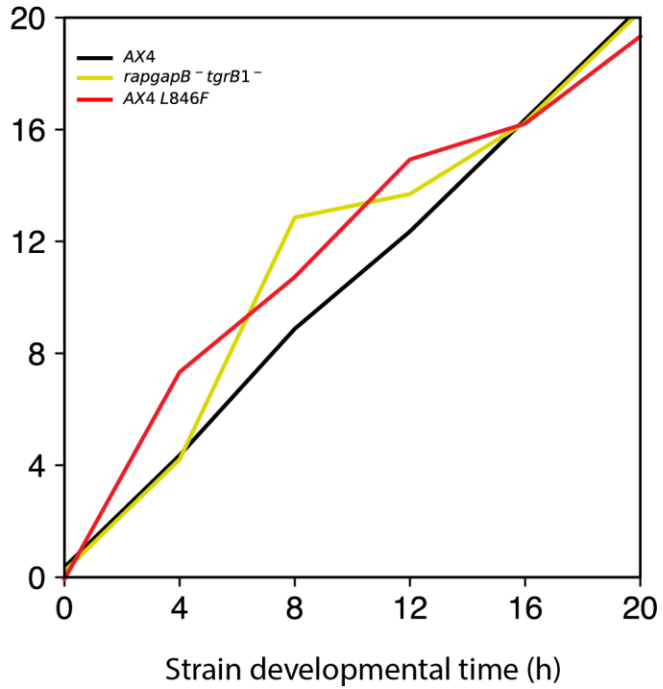
Mutual suppression restores WT development

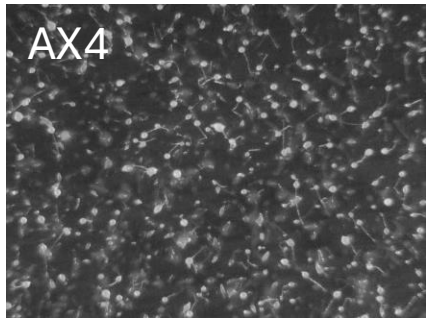


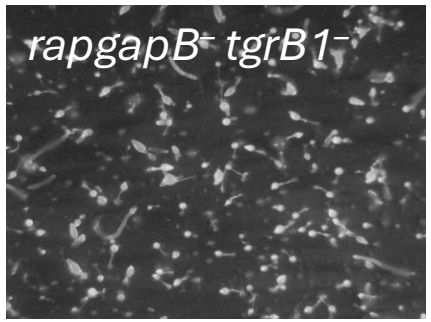


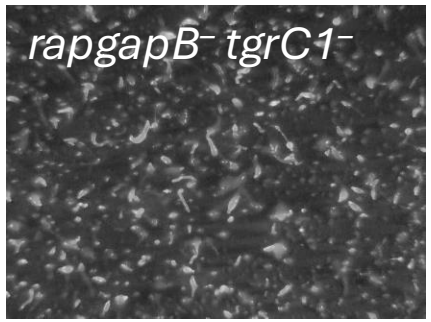






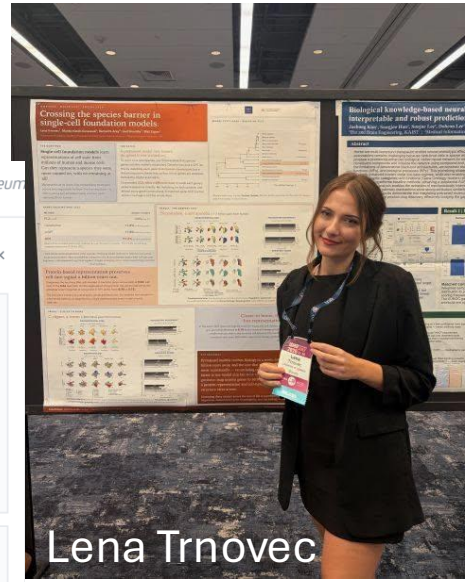






Katoh-Kurasawa et al., (2025) *BMC Genomics* 26:563.

Predicting TF regulation with AI



dictyExpress TF

Dictyostelium discoideum

Candidate ranking

How to read this

STRONGEST CANDIDATES FOR A VISIBLE PHENOTYPE top 10

PRIORITY / TF	AVERAGE RANK	PHENOTYPE ENRICHMENT	PHENOTYPE GROUPS
#1 gtaH Target genes	#2.25 Range #1–3	2.45× 54 hits / ~22 expected	▶ 6 categories
#2 bzpE Target genes	#3.00 Range #1–5	2.36× 52 hits / ~22 expected	▶ 5 categories
#3 jcdJ Target genes	#4.25 Range #1–10	2.50× 55 hits / ~22 expected	▶ 7 categories

Top 5 phenotype-relevant targets 52 anchor hits

DG1040 12–24 h	fruiting body	RESPONSE top 0.10%	MOTIF top 7% no ATAC peak	Whole population →
dimA 12–24 h	fruiting body	RESPONSE top 0.2%	MOTIF top 19% ATAC in 2 reps	Prestalk branch →
yakA 12–24 h	fruiting body	RESPONSE top 0.2%	MOTIF top 22% no ATAC peak	Whole population →
amiB 0–8 h	aggregation acaA rescue · supported	RESPONSE top 0.4%	MOTIF top 5% no ATAC peak	Prespore branch →
atg1 0–8 h	aggregation	RESPONSE top 0.4%	MOTIF top 6% no ATAC peak	Prespore branch →

Ask about a TF, target gene, phenotype, or regulatory relation...

DNA MOTIF COMPATIBILITY

bzpE → dimA

TF MOTIF

TARGET PROMOTER

TACAACATTCA **CACATACACAC** ATACACACCCA

Expected C A A A T A C G T A T

Target C A C A T A C A C A C

Motif **top 19%** Position **874 bp upstream** Strand **-** Open chromatin **2 ATAC reps** Occupancy **not tested**

Function **not tested**

▼ Show genomic context

−2,000 bp *dimA* start

Motif · 874 bp upstream Open chromatin · 2 replicates

SINGLE-CELL MAPS

bzpE → dimA

bzpE TF *dimA* target

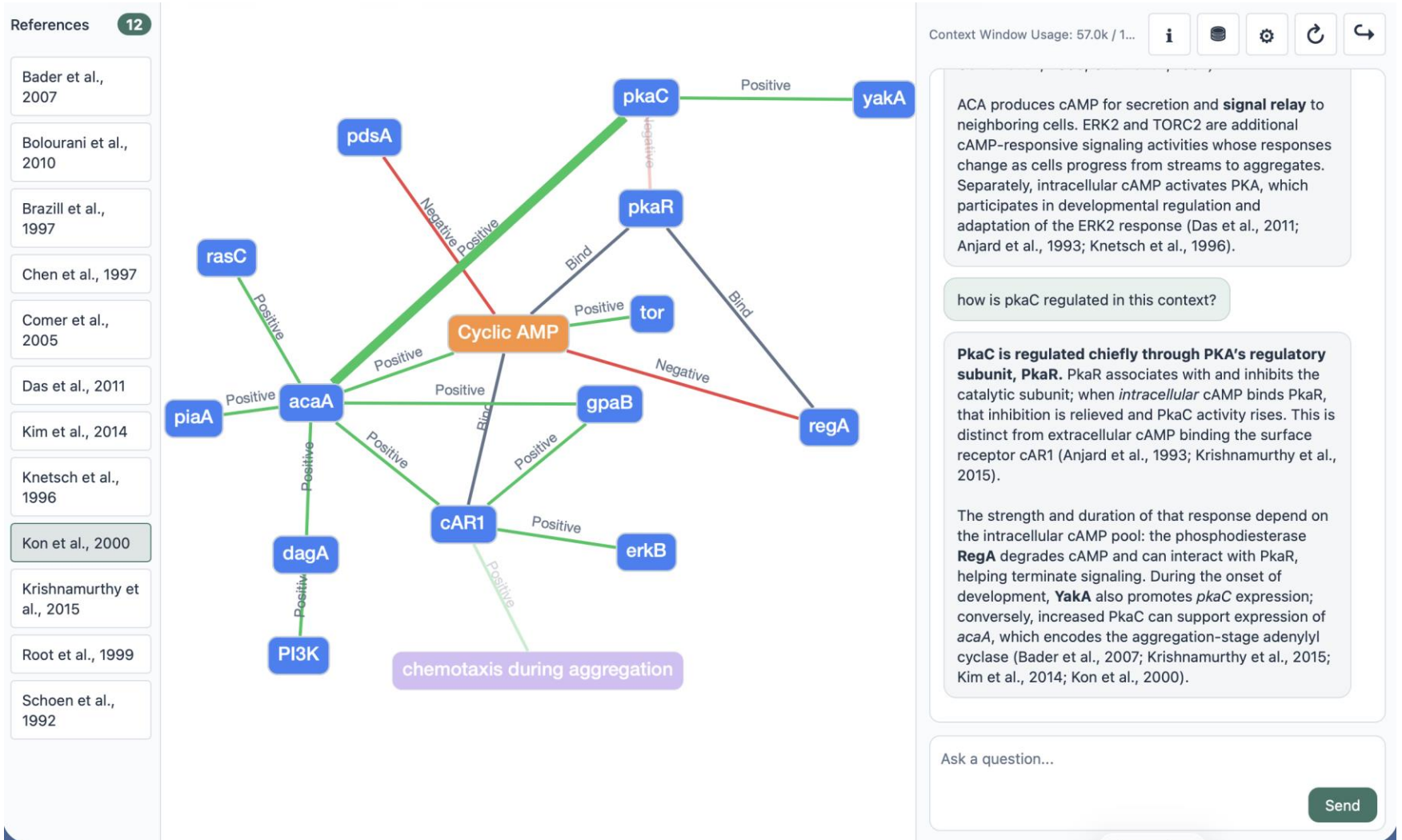
prestalk branch

prestalk branch

undifferentiated

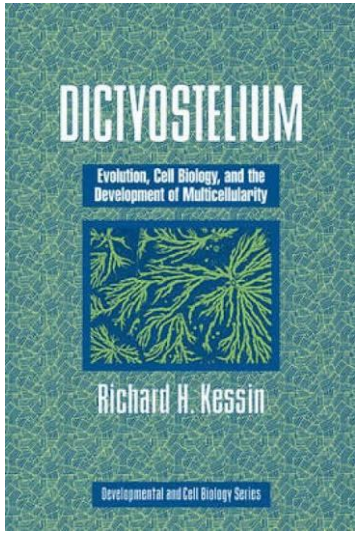
prestalk branch

dictyNet - An interactive tool for biological knowledge exploration through natural language

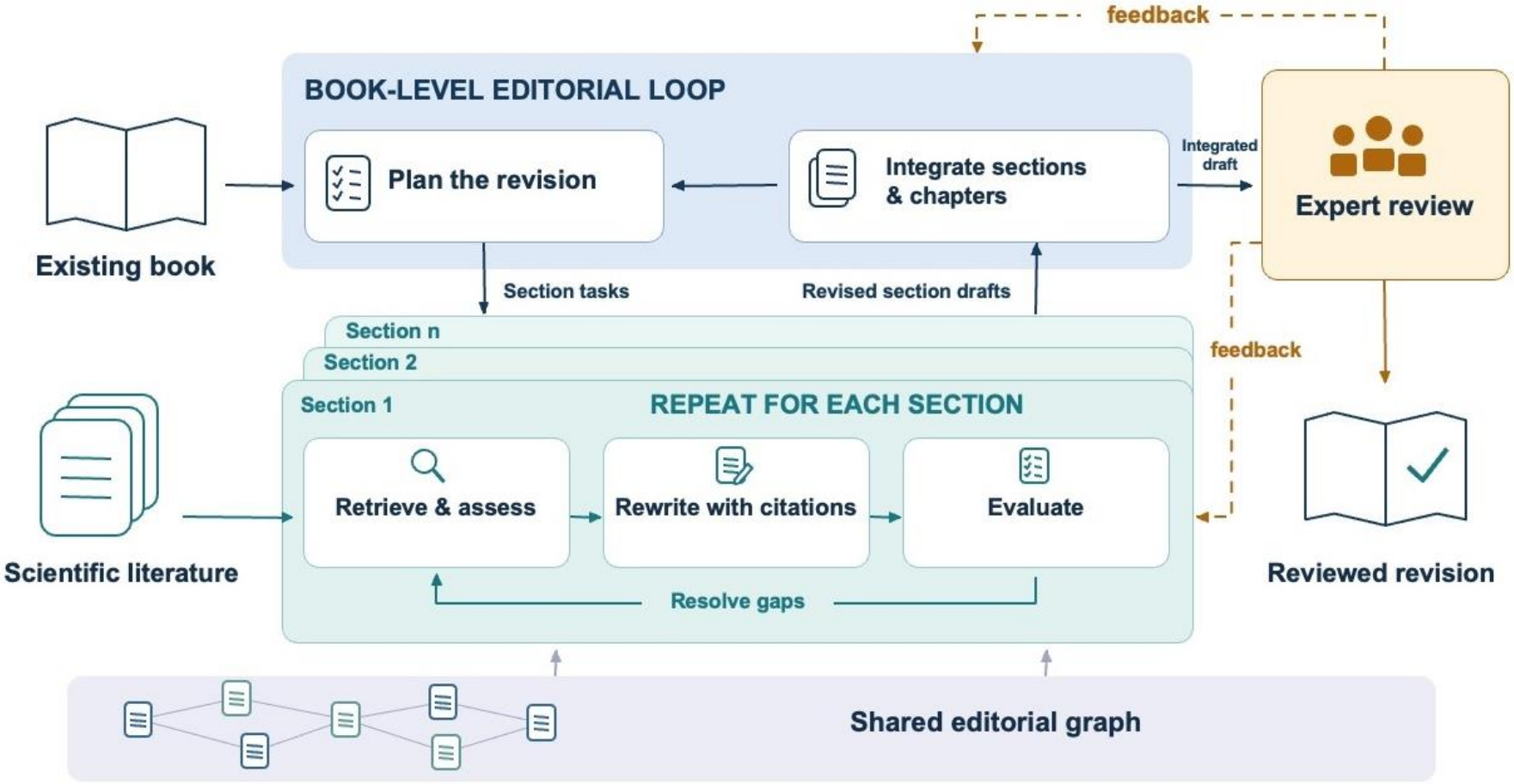




AI-assisted framework for updating scientific books with evidence from the literature



Kessin, 2001





Collaboration across disciplines

Mutual respect
Communication
Open minds
Patience

Democratization