

Phenologs

A case study of using bioinformatics to find new genes for genetic traits

BCH394P/364C Systems Biology / Bioinformatics

Edward Marcotte, Univ of Texas at Austin

1

Are *you* a research parasite?



The NEW ENGLAND
JOURNAL of MEDICINE

HOME ARTICLES & MULTIMEDIA ISSUES SPECIALTIES & TOPICS FOR AUTHORS CME >

EDITORIAL

Data Sharing

Dan L. Longo, M.D., and Jeffrey M. Drazen, M.D.
N Engl J Med 2016; 374:276-277 | January 21, 2016 | DOI: 10.1056/NEJMe1516564

Share    

Article References Citing Articles (52) Letters Metrics

The aerial view of the concept of data sharing is beautiful. What could be better than having high-quality information carefully reexamined for the possibility that new nuggets of useful data are lying there, previously unseen? The potential for leveraging existing results for even more benefit pays appropriate increased tribute to the patients who put themselves at risk to generate the data. The moral imperative to honor their collective sacrifice is the trump card that takes this trick.

However, many of us who have actually conducted clinical research, managed clinical studies and data collection and analysis, and curated data sets have concerns about the details. The first concern is that someone not involved in the generation and collection of the data may not understand the choices made in defining the parameters. Special problems arise if data are to be combined from independent studies and considered comparable. How heterogeneous were the study populations? Were the eligibility criteria the same? Can it be assumed that the differences in study populations, data collection and analysis, and treatments, both protocol-specified and unspecified, can be ignored?

“The aerial view of the concept of data sharing is beautiful.”

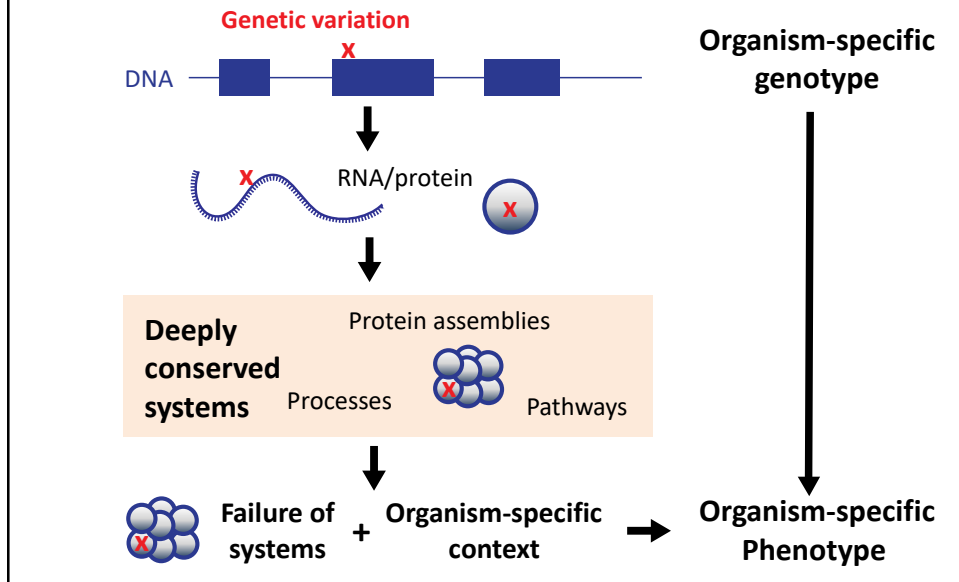
[but!]

A ... concern ... is that a new class of research person will emerge...the system will be taken over by ...
“research parasites.”

<http://www.nejm.org/doi/full/10.1056/NEJMe1516564>

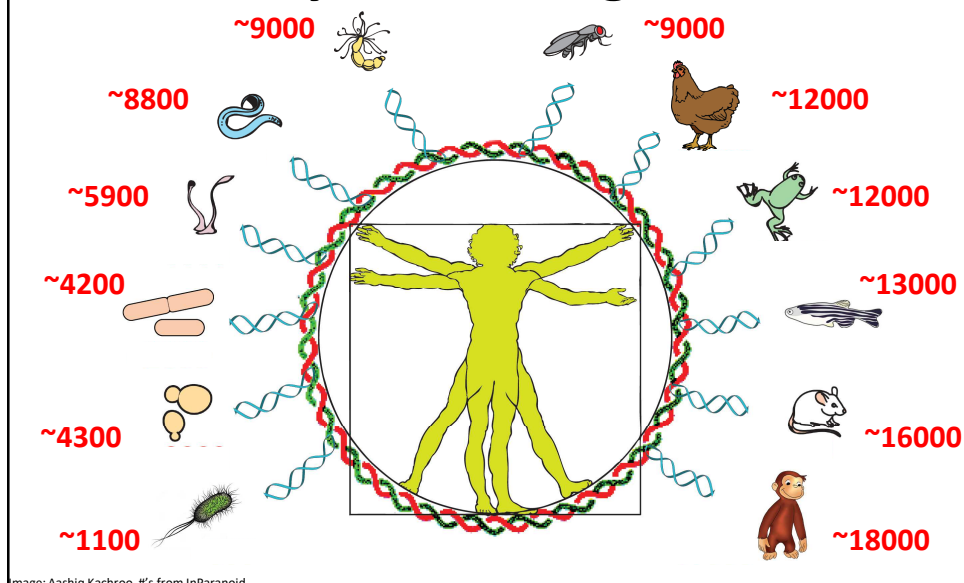
2

Conserved systems intermediate between organism-specific genotype and phenotype

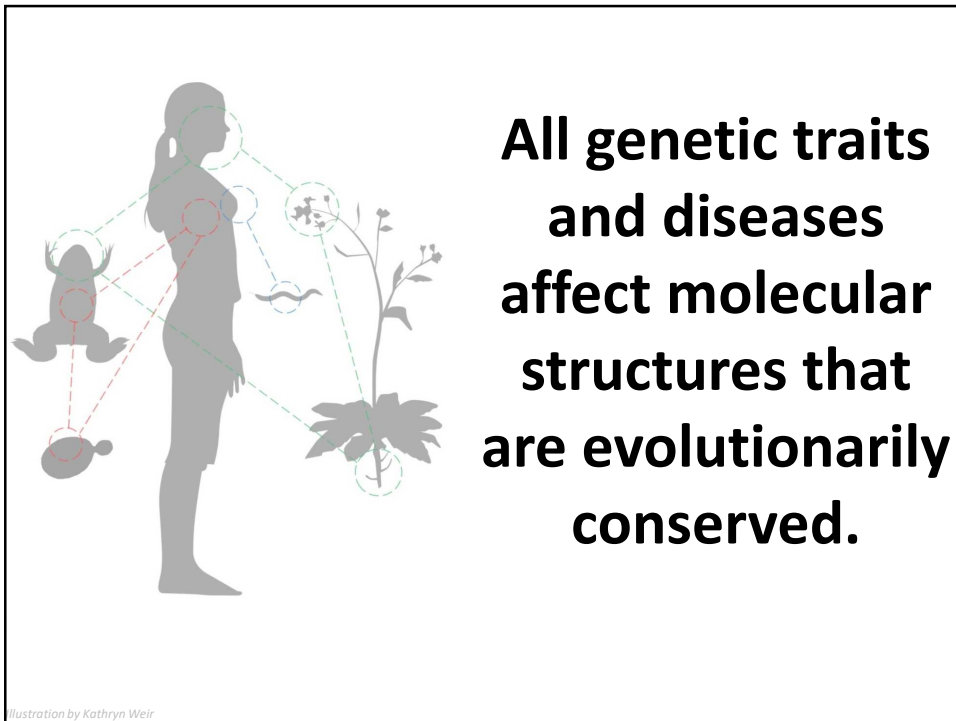


3

We share genes with almost every known organism



4



5

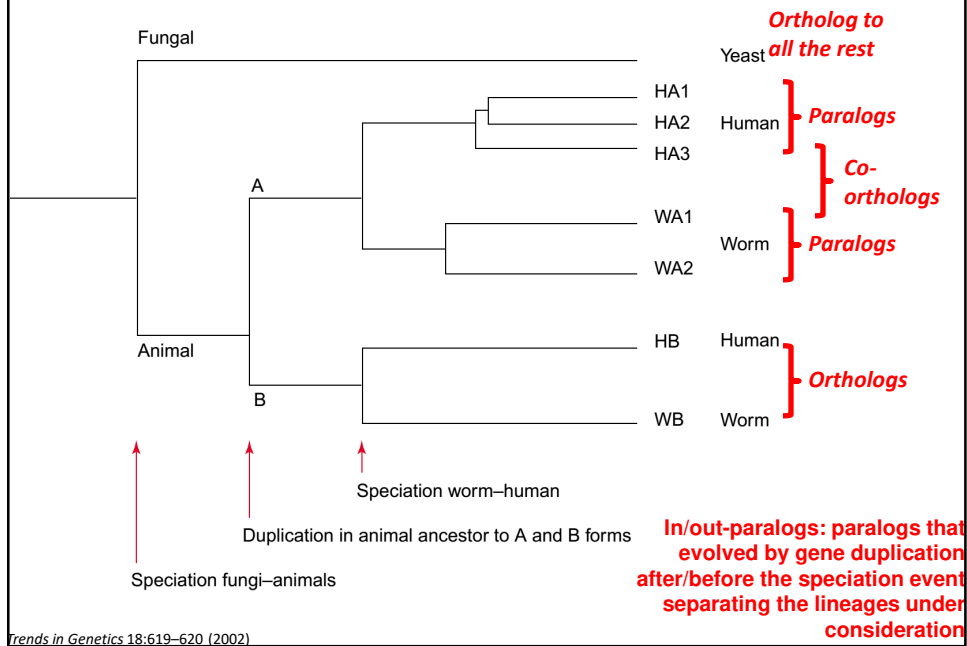
Comparative evolution studies rely on finding orthologs

Orthologs = genes from different species that derive from a single gene in the last common ancestor of the species

Paralogs = genes that derive from a single gene that was duplicated within a genome

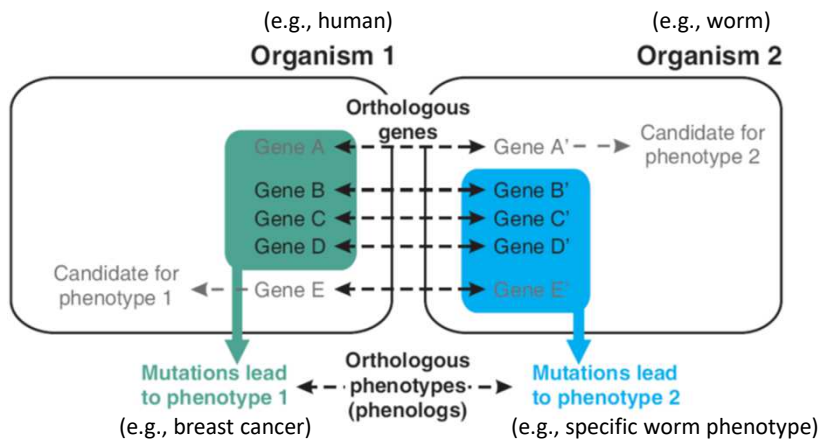
6

Comparative evolution studies rely on finding orthologs



7

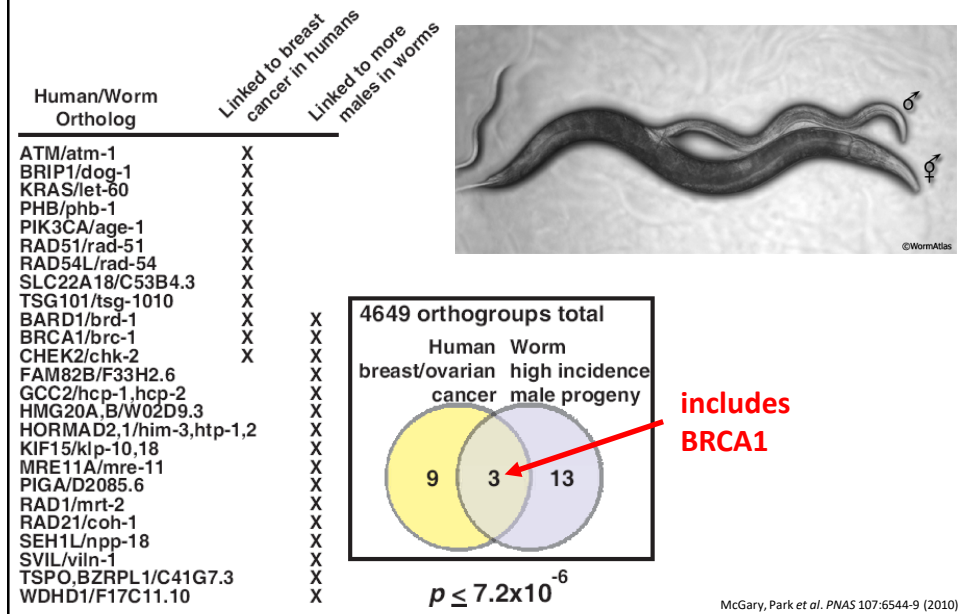
Phenologs = significantly overlapping sets of orthologous genes, such that each gene in a given set gives rise to the same phenotype in that organism



McGary, Park et al. *PNAS* 107:6544-9 (2010)

8

E.g., 'high incidence of male' *C. elegans* genes predict human breast/ovarian cancer genes



9

Building & searching a collection of phenotypes

Mining available databases +
manual collection from the primary literature



gene-phenotype
associations

<u>Organism</u>	<u># gene-phenotype associations</u>
human	1,923
mouse	74,250
worm	27,065
yeast	86,383
<i>Arabidopsis</i>	22,921

Spanning ~300 human diseases,
>7,000 model organism mutational phenotypes

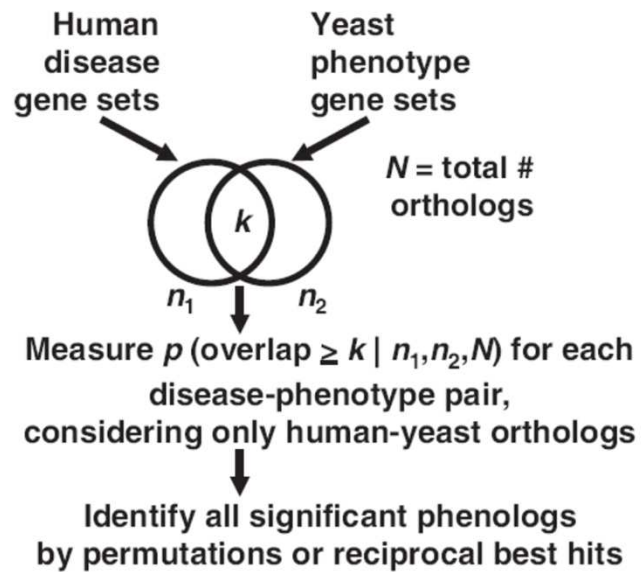


Computational scan phenotypes for novel models of a disease of interest,
identify significant phenologs using permutation tests

McGary, Park et al. PNAS 107:6544-9 (2010)

10

Discovering phenologs



McGary, Park et al. *PNAS* 107:6544-9 (2010)

11

Computationally, we find many genes shared between human diseases and mouse, yeast, worm, and even plant traits

McGary, Park et al. *PNAS* 107:6544-9 (2010)
Woods, Blom et al. *BMC Bioinformatics*, 14:203 (2013)

12



**Waardenburg syndrome
accounts for ~2-5% of
cases of deafness**

Michael Murphy, M.D.

Associated websites: <http://www.werewolf.com/Waardenburg-syndrome-1048980>, <http://stephanieanisarchez.blogspot.com/>



13

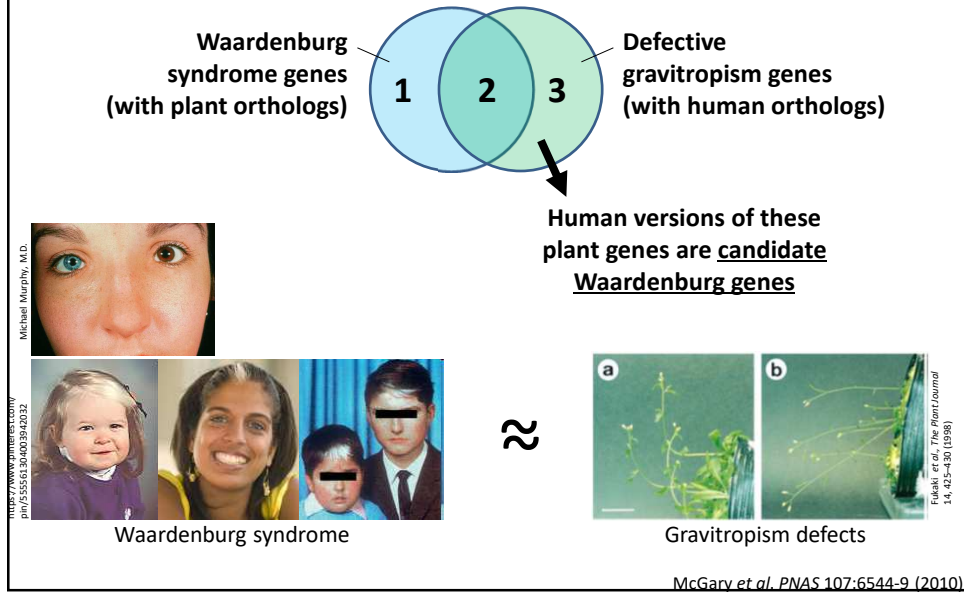
Plants sense and respond to gravity → gravitropism



Fukaki et al., *The Plant Journal*
14, 425-430 (1998)

14

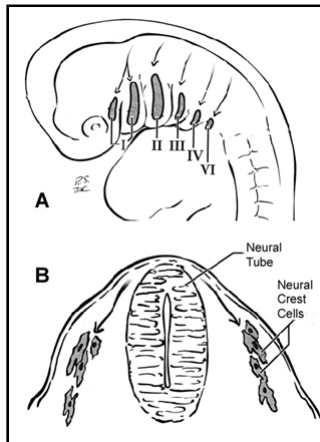
Plant gravitropism genes predict Waardenburg syndrome, a human congenital deafness syndrome



15

Waardenburg syndrome is a defect of neural crest cells

Neural crest cells migrate during embryonic development



Some WS correlates in other animals:
Deafness in Dalmatian dogs (22% unilaterally deaf)



Variations in the Blenheim spot
Cavalier King Charles Spaniels

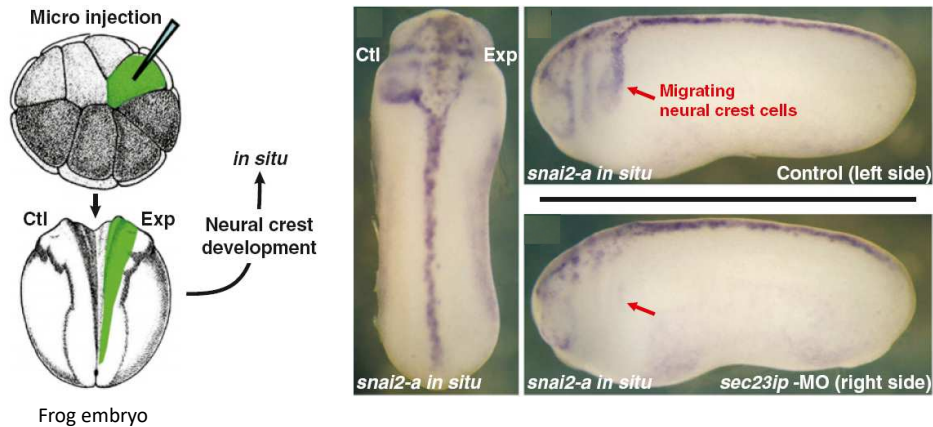


Association between white blue-eyed cats and deafness (noted by Darwin in 1859)

White forelock and deafness/bowel blockage in foals
& many more...

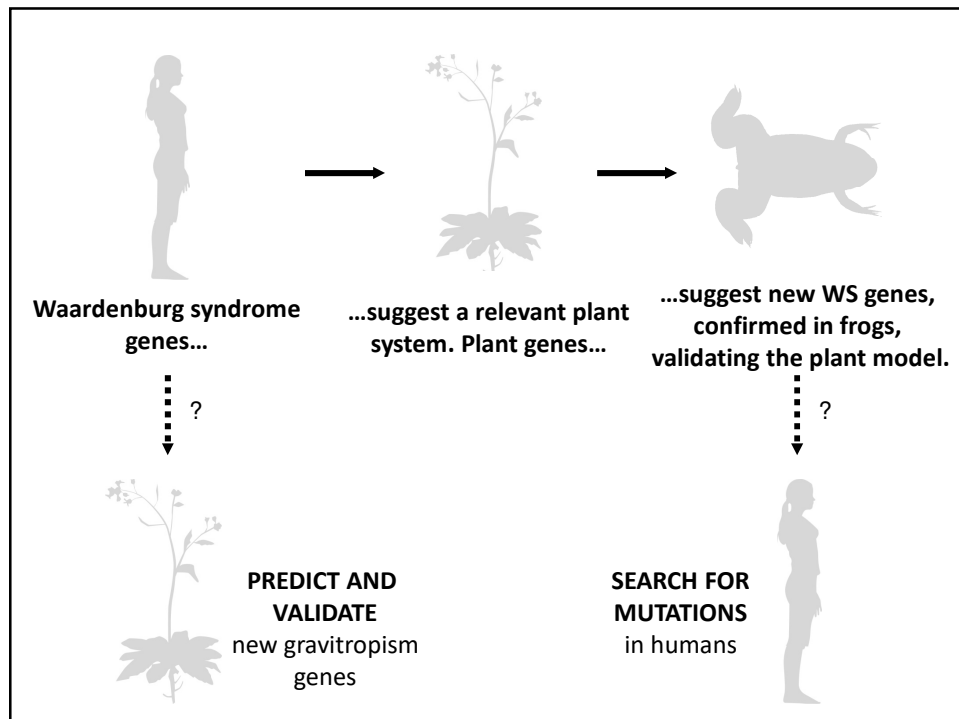
16

**Sure enough, inactivating one of the genes—
predicted from plants—in a tadpole disrupts neural
crest cells, consistent with Waardenburg syndrome**



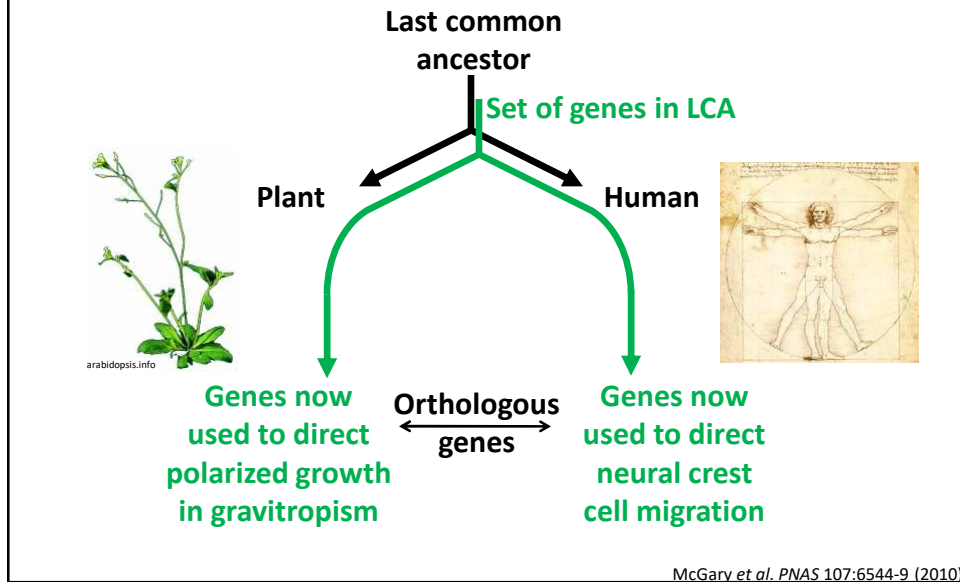
McGary, Park et al. PNAS 107:6544-9 (2010)

17



18

Conservation and “repurposing” of core cell machinery across deep evolutionary time



19

Let's talk about how such projects play out in practice.

**How are discoveries made?
How do you computationally explore ideas?**

20

Let's step through this particular discovery process:

- 1. We had an idea, based on a puzzling observation:**

Why do mutations in worm retinoblastoma genes induce ectopic vulva while a mutation in the human ortholog causes eye cancer?

We weren't interested in specific mechanism here, but rather the impact of organismal context on conserved systems. In particular, *how do ever-more distant evolutionary models inform us about human disease?*

21

Let's step through this particular discovery process:

- 2. We thought about how this might be part of a large trend—does it illustrate a general principle? Could we look for new cases systematically?**
- 3. We thought about other examples, mentally assembling what could serve as positive and negative control cases. i.e. how to we decide if a systematic approach is working?**

22

Let's step through this particular discovery process:

- 4. A grad student (Kris McGary) started assembling relevant datasets. We took heavy advantage of existing resources: model organism databases that had already painstakingly curated relevant data, large-scale screens reporting easy-to-process data.**

23

Let's step through this particular discovery process:

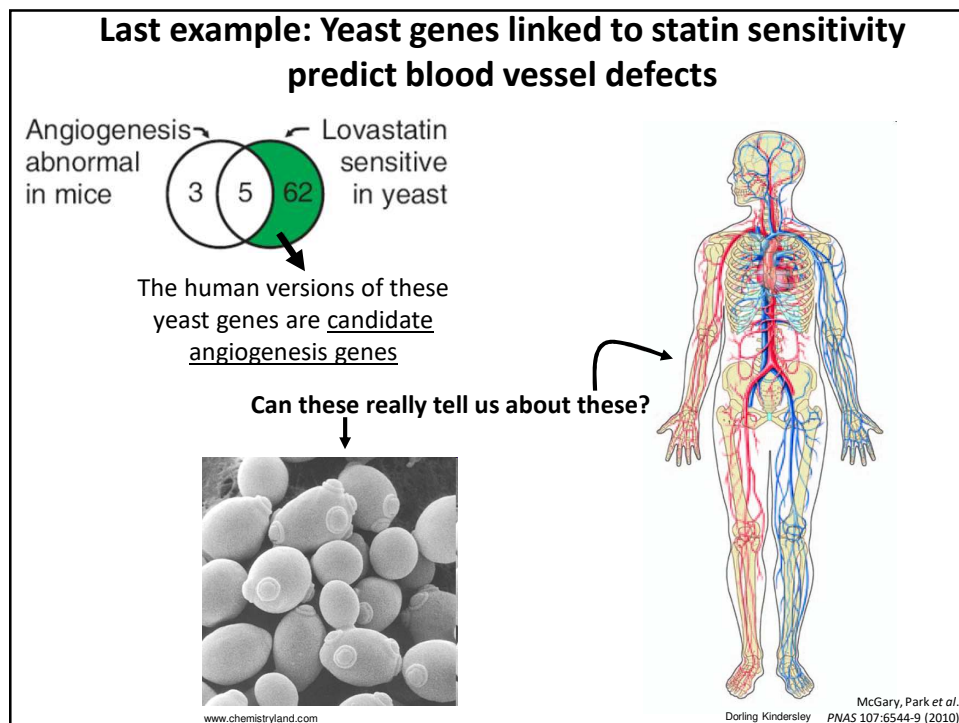
- 5. We started inventing/evaluating statistical models/algorithms, exploring the data and thinking about how to search for the relevant trends. We iterated these steps until we thought we understood the problem better.**
- 6. At some point, the lab bet a 6 pack of beer on the outcome:
Can we discover plant models of human disease?**

24

Let's step through this particular discovery process:

- 7. The algorithms predicted some remarkable and crazy results. We had no option but to test or reject the new predictions, so began testing, thanks to collaborators in the Wallingford lab willing to sink a few weeks into high-risk experiments.**
- 8. Some tests worked, some didn't. We went back & thought about the ones that didn't and refined how we prioritized the results.**
- 9. Iterate, iterate. Jackpot! A plant model of deafness! Shouting in the halls...**

25



26

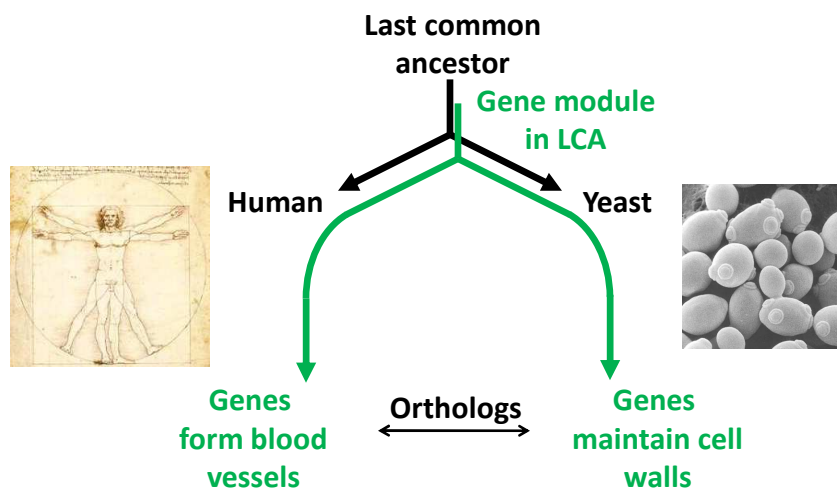
Disrupting the SOX13 gene causes strong blood vessel defects



McGary, Park *et al.*
PNAS 107:6544-9 (2010)

27

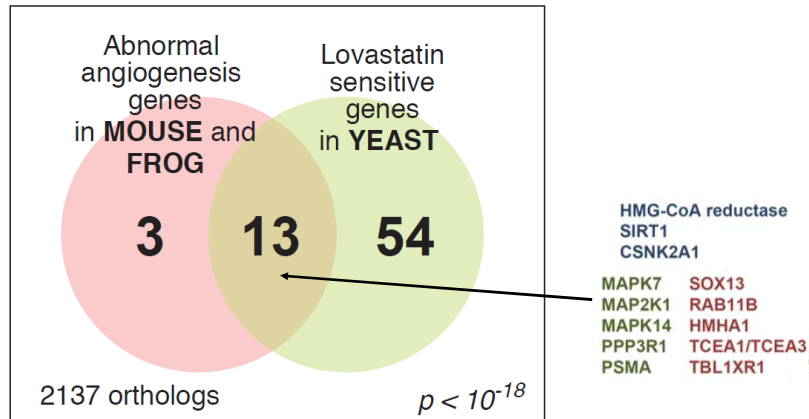
A yeast model of angiogenesis = example of a deeply conserved, but “repurposed” gene module



McGary, Park *et al.* *PNAS* (2010)

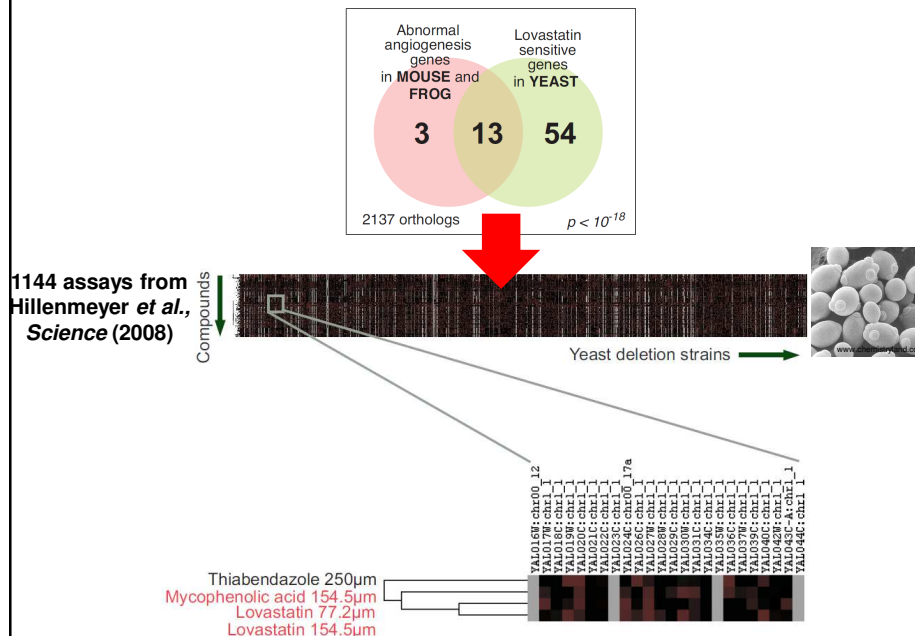
28

The yeast/angiogenesis gene module



29

Chemicals that interact genetically with this module are candidate angiogenesis inhibitors

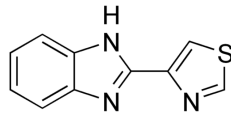


30

Screening for drugs that interact genetically with this yeast module led us to identify a new angiogenesis inhibitor

TBZ = thiabendazole

FDA-approved antifungal drug with 40 years of safety data

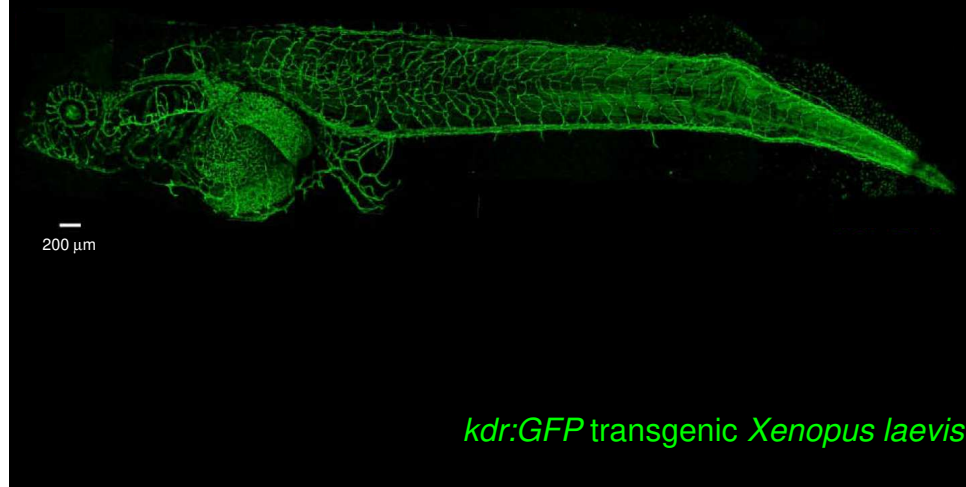


- Approved by U.S. Food and Drug Administration in 1967

- Fungicide and parasiticide
- Not mutagenic or carcinogenic; 2 year dog safety trials
- Off-patent, marketed as a generic

31

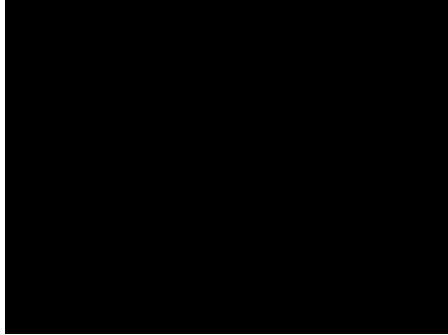
Imaging the blood vessels of a living, transgenic tadpole in a dish of water



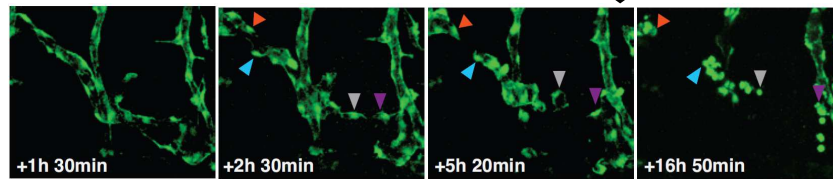
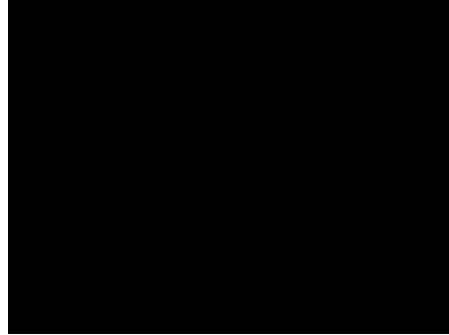
32

TBZ disrupts vascular integrity, making vascular endothelial cells retract & round up

Control (DMSO carrier)



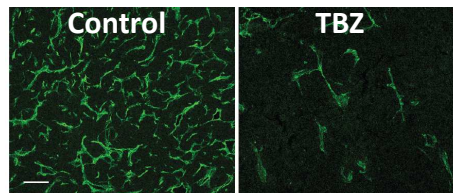
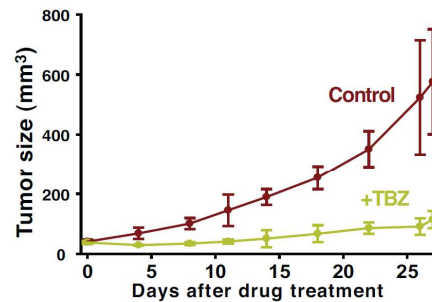
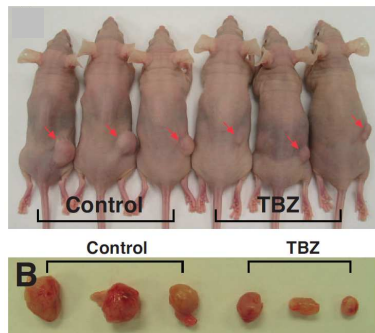
+ TBZ



Cha et al., *PLoS Biology* (2012)

33

TBZ slows human fibrosarcoma tumors transplanted into immune-compromised mice

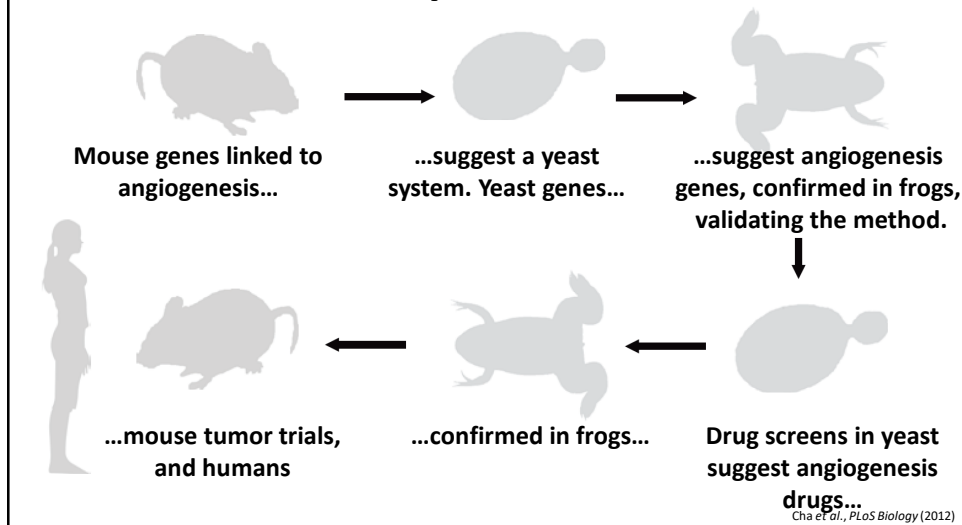


Vasculature in tumor sections

Cha et al., *PLoS Biology* (2012)

34

“Road map” to a new vascular disrupting agent, by mapping phenotypes across species



35

Try it out yourself!
<http://www.phenologs.org>

You can start by rediscovering the plant model of
Waardenburg syndrome:

Search known diseases for "Waardenburg", or enter the
human genes linked to Waardenburg (Entrez gene IDs 4286,
5077, 6591, 7299) to start.

Tools for finding orthologs are linked on the class website

36