

Phenologs

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

BCH339N Systems Biology / Bioinformatics
Edward Marcotte, Univ of Texas at Austin

There was an interesting spat 2 years over data sharing:



The NEW ENGLAND
JOURNAL of MEDICINE

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

EDITORIAL

Data Sharing

Den L. Longo, M.D., and Jeffrey M. Drazen, M.D.
N Engl J Med 2016; 374:276-277 | January 21, 2016 | DOI: 10.1056/NEJMe1518564

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>

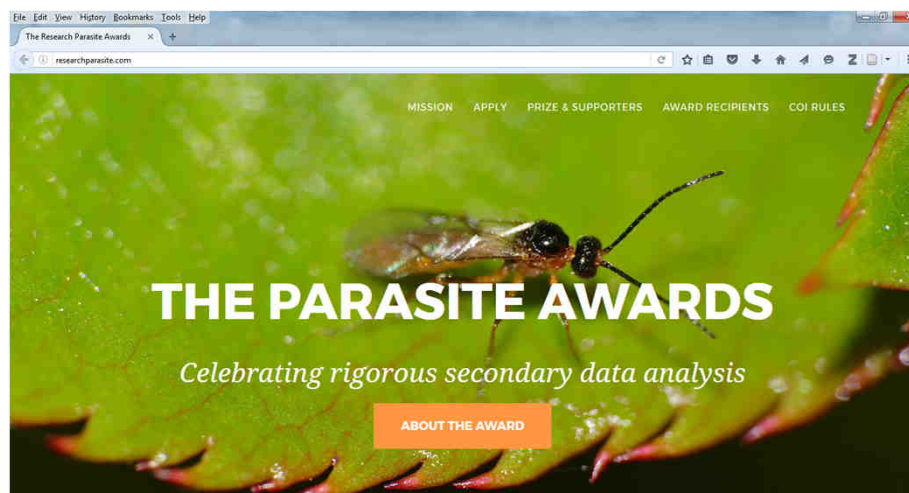
My opinion, FWIW, is that “research parasites” are

- 1. Independent and often highly rigorous scientists**
- 2. Essential to the scientific process, especially when they**
- 3. Independently test the original authors’ analyses. Often,**
- 4. They approach analyses with different starting biases, so**
- 5. Can contribute entirely new interpretations of the original studies, and**
- 6. Find entirely unanticipated uses for published data**

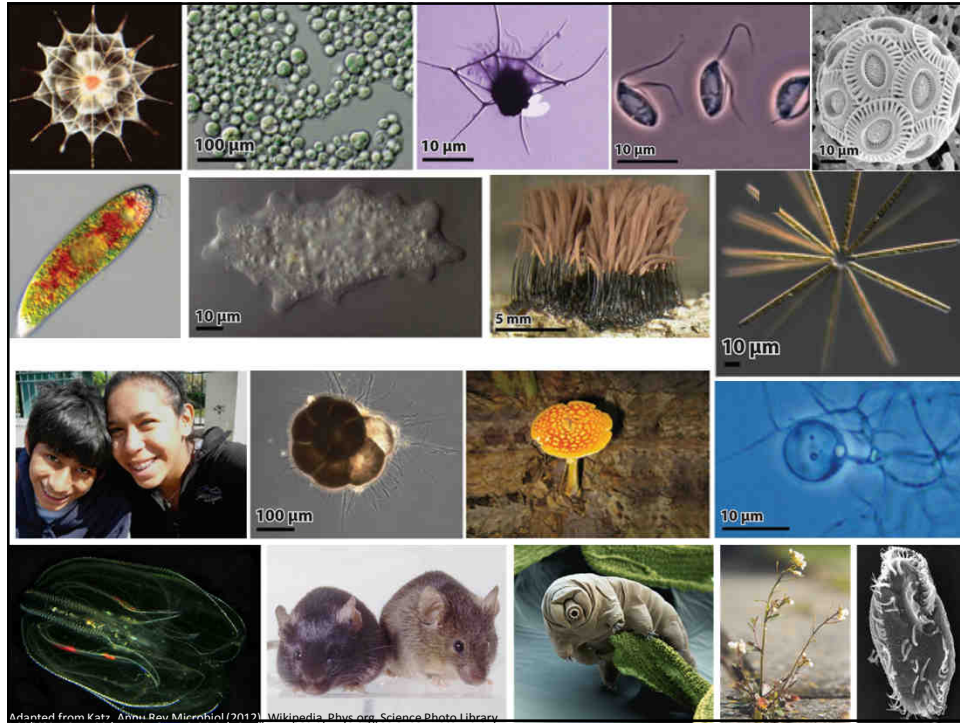
IMO, the act of publishing data in a peer-reviewed journal commits you to release that data for public inspection, reproducibility studies, re-analysis, and many unanticipated new uses.

Science is improved when this happens.

Plus, you could win big!



<http://researchparasite.com/>



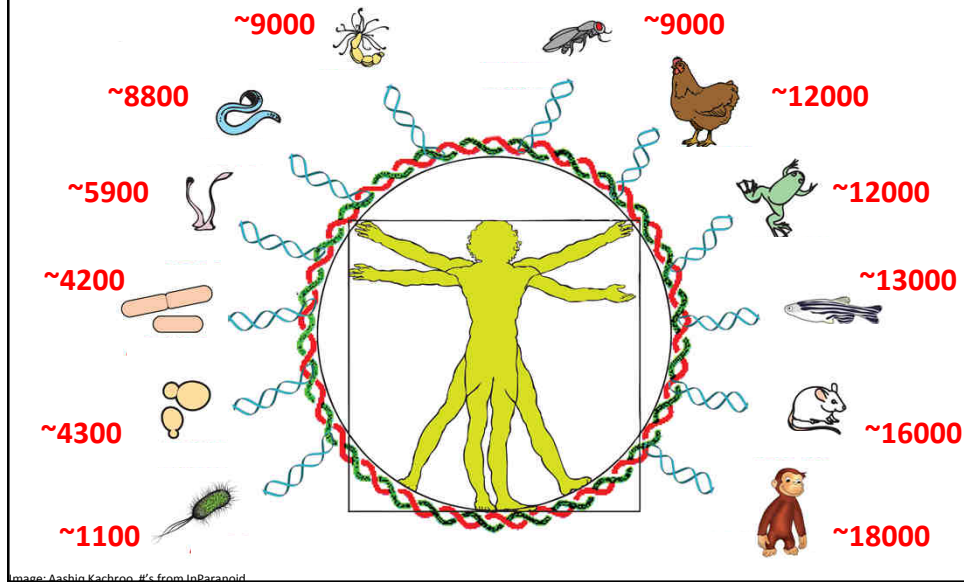
Adapted from Katz, *Annu Rev Microbiol* (2012), Wikipedia, Phys.org, Science Photo Library



Extraordinary Relatedness

Adapted from Katz, *Annu Rev Microbiol* (2012), Wikipedia, Phys.org, Science Photo Library

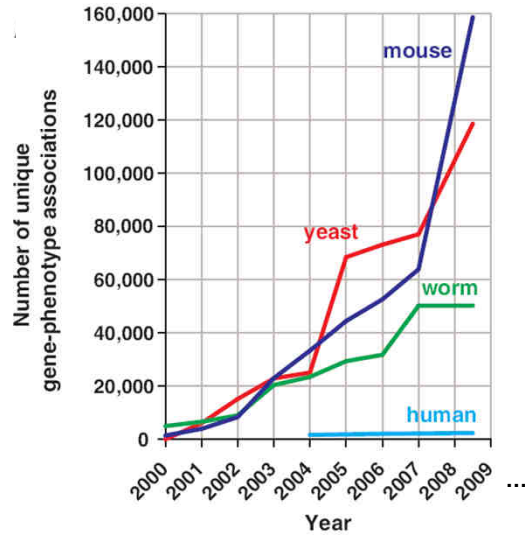
We share genes with almost every known organism



All genetic traits and diseases affect molecular structures that are evolutionarily conserved.

Illustration by Kathryn Weir

We know far more about genes & traits in lower organisms than in us. Can we transfer that knowledge to us?



Can these tell us about us?

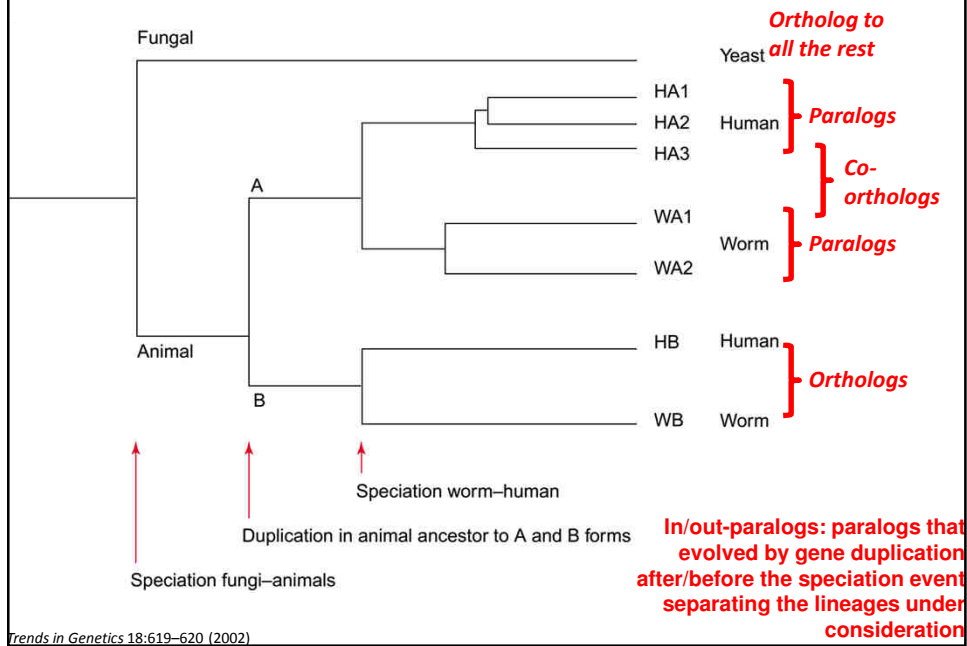
McGary, Park et al. *PNAS* 107:6544-9 (2010)

Comparative evolution studies rely on finding orthologs

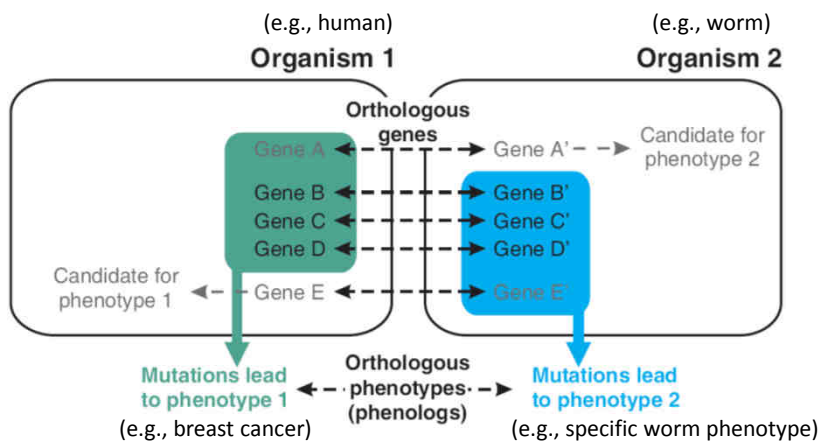
Orthologs = genes from different species that derive from a single gene in the last common ancestor of the species

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

Comparative evolution studies rely on finding orthologs

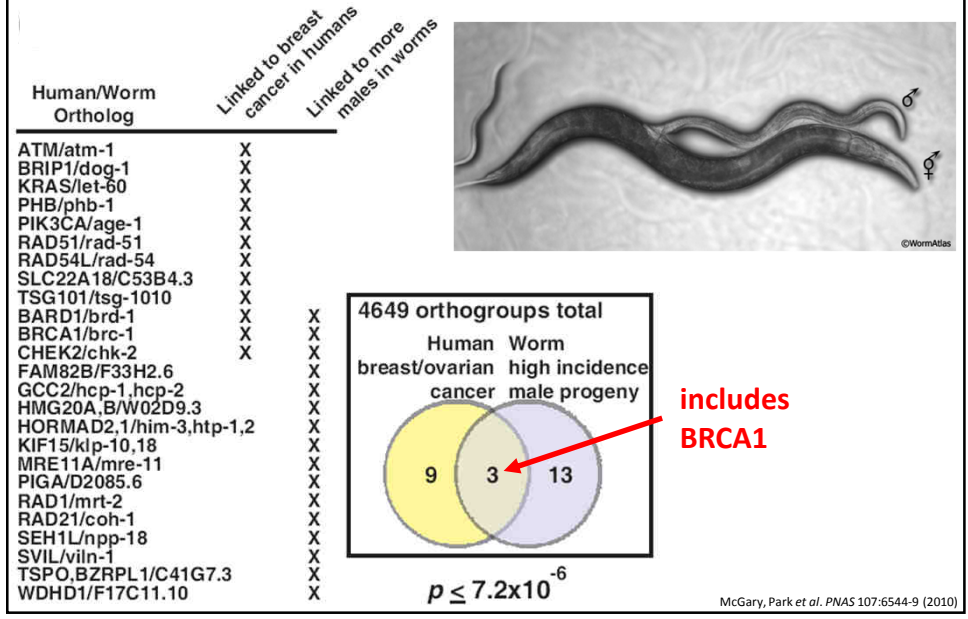


Phenologs = significantly overlapping sets of orthologous genes, such that each gene in a given set gives rise to the same phenotype in that organism



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

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



Building & searching a collection of phenotypes

Mining available databases + manual collection from the primary literature



gene-phenotype associations

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

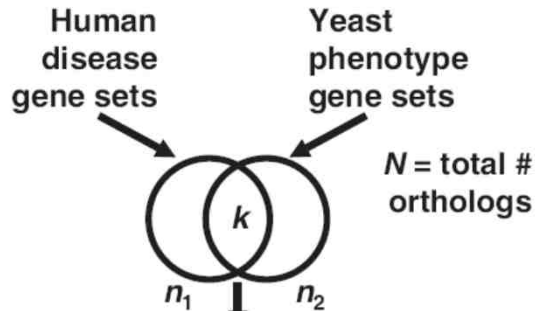
Spanning ~300 human diseases, >7,000 model organism mutational phenotypes



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)

Discovering phenologs



Measure $p(\text{overlap} \geq k \mid n_1, n_2, N)$ for each disease-phenotype pair, considering only human-yeast orthologs

Identify all significant phenologs by permutations or reciprocal best hits

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

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)



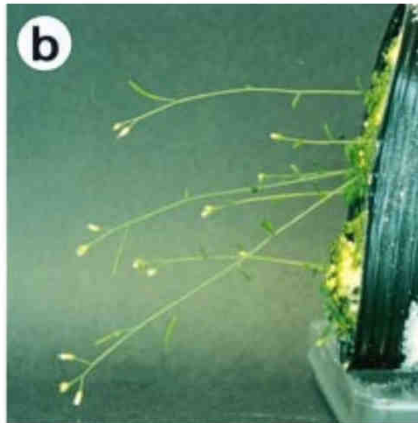
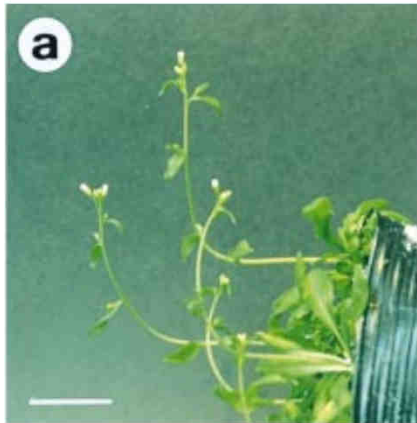
Michael Murphy, M.D.

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



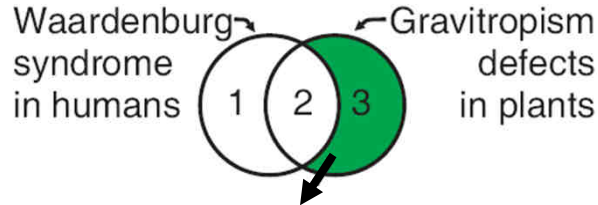
Associated websites: incl. www.uspaxwell.com/waardenburg-syndrome-10498992, <http://stephanmarrasandlaga.blogspot.com/>

Plants sense and respond to gravity → gravitropism

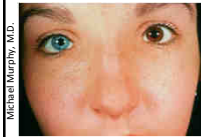


Fukaki et al., *The Plant Journal* 14, 425-430 (1998)

Plant gravitropism predicts genes for Waardenburg syndrome, a human congenital deafness syndrome

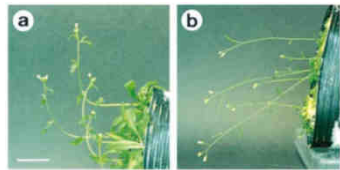


The human versions of these plant genes are candidate Waardenburg genes



Waardenburg syndrome

≈



Gravitropism defects

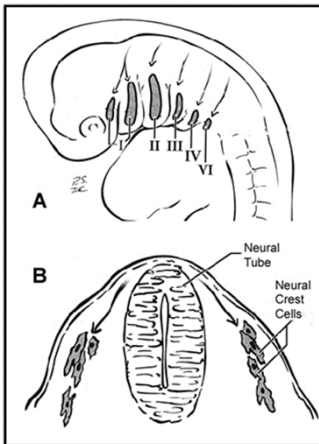
Michael Murphy, M.D.

Assorted websites

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

Waardenburg syndrome is a defect of neural crest cells

Neural crest cells migrate during embryonic development



Heike & Hing, Gene Reviews (2009)

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



www.petplanet.co.uk

Variations in the Blenheim spot
Cavalier King Charles Spaniels

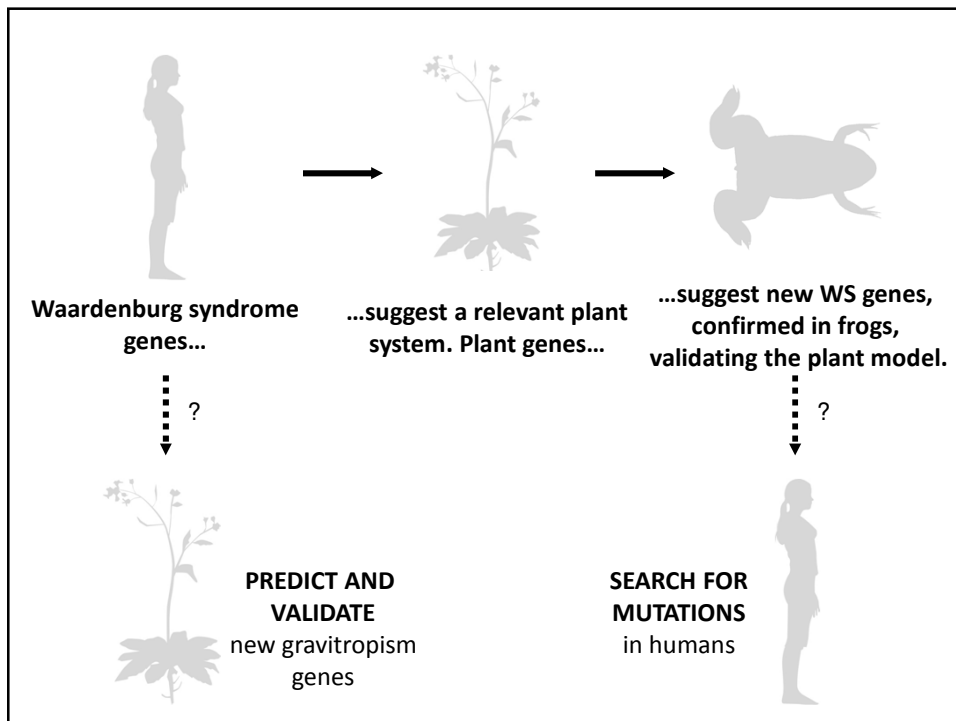
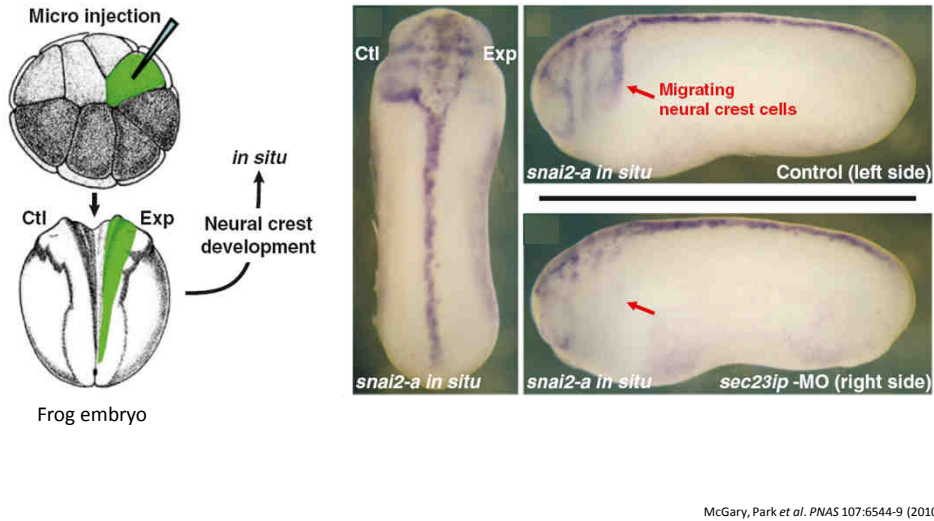


www.silvercea.co.uk

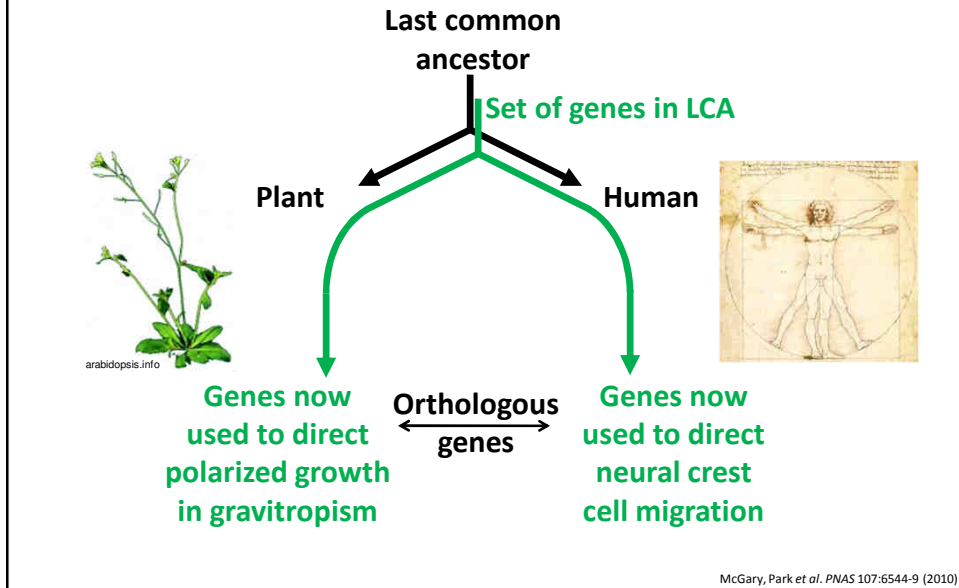
Association between white blue-eyed cats and deafness (noted by Darwin in 1859)

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

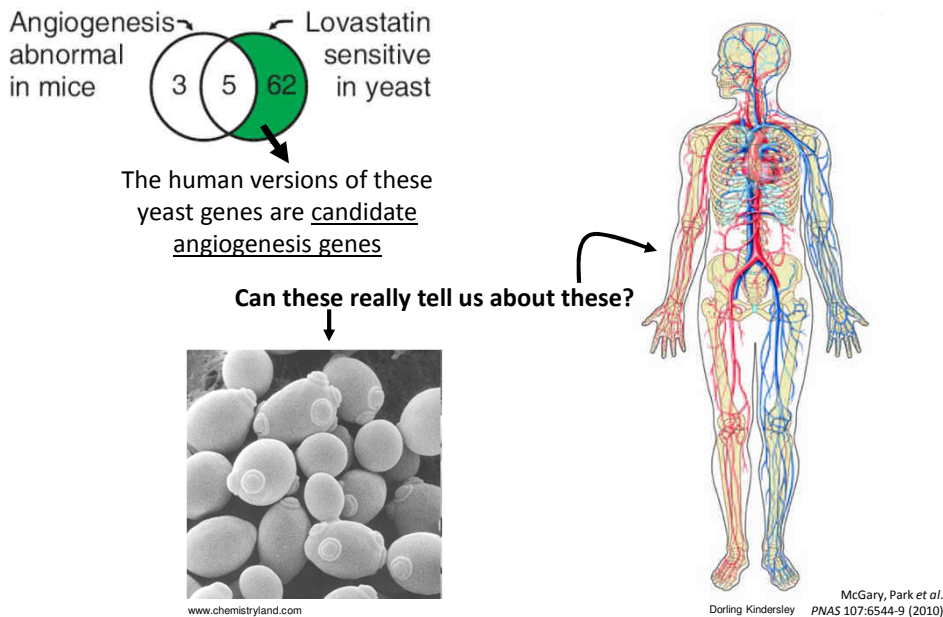
**Sure enough, inactivating one of the genes—
predicted from plants—in a tadpole disrupts neural
crest cells, consistent with Waardenburg syndrome**



Phenologs identify evolutionarily conserved systems of proteins relevant to particular traits/diseases.



Example #3: Yeast genes linked to statin sensitivity predict blood vessel defects

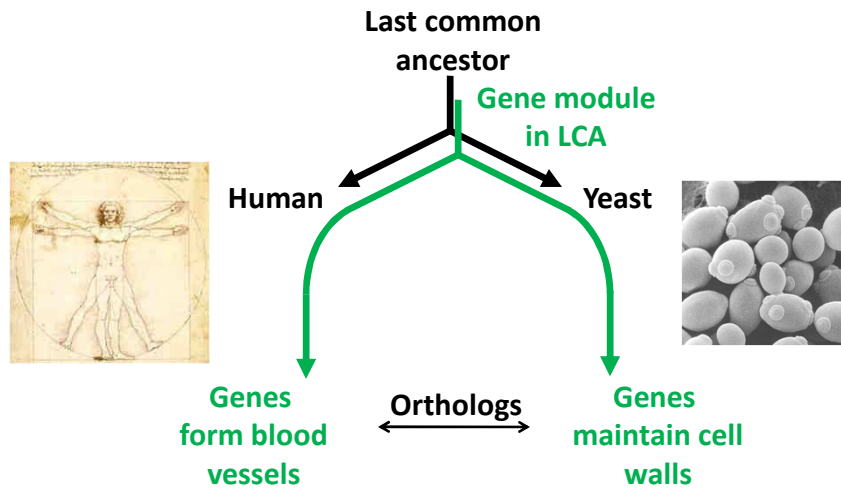


Disrupting the SOX13 gene causes strong blood vessel defects



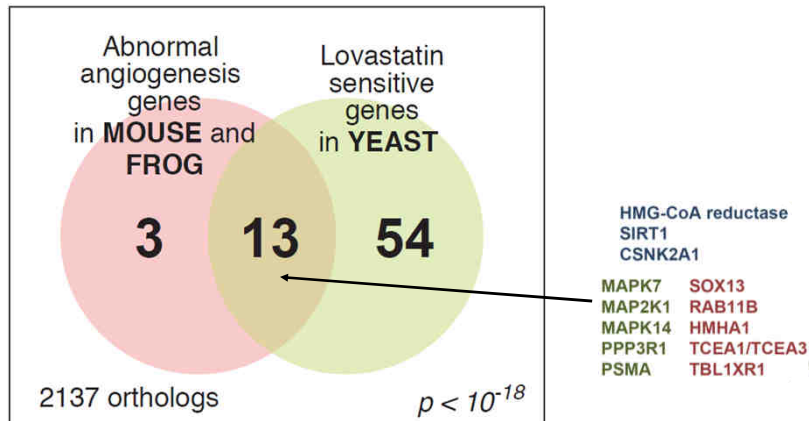
McGary, Park et al.
PNAS 107:6544-9 (2010)

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

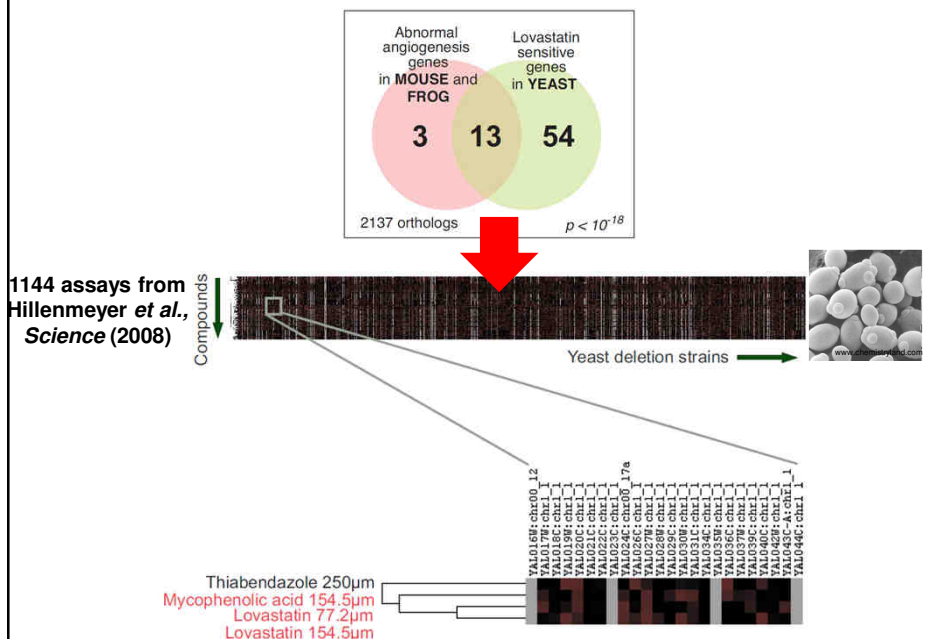


McGary, Park et al. *PNAS* (2010)

The yeast/angiogenesis gene module



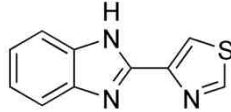
Chemicals that interact genetically with this module are candidate angiogenesis inhibitors



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

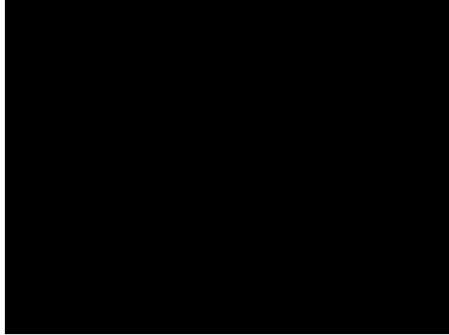
- Fungicide and parasiticide
- Not mutagenic or carcinogenic; 2 year dog safety trials
- Off-patent, marketed as a generic

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

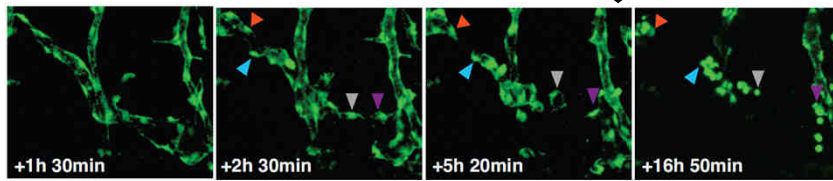
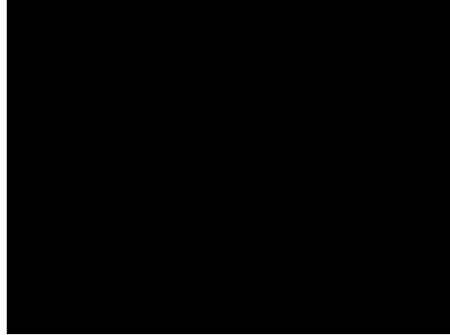


TBZ disrupts vascular integrity, making vascular endothelial cells retract & round up

Control (DMSO carrier)

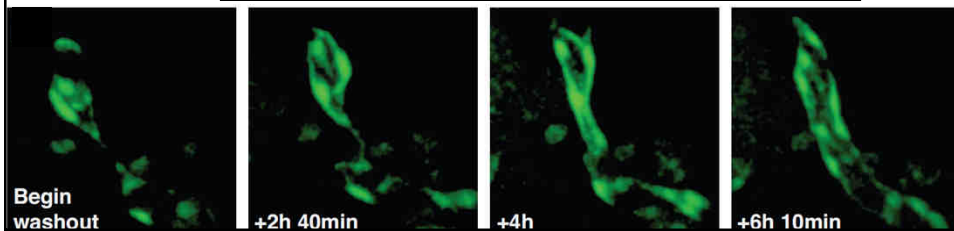
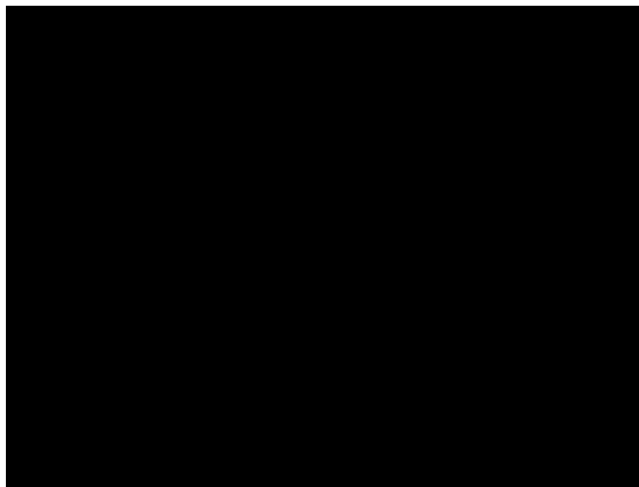


+ TBZ

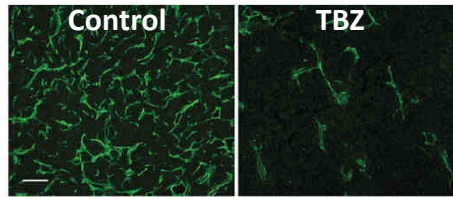
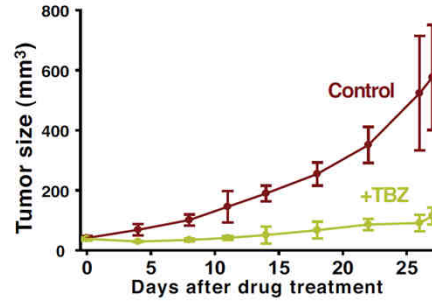
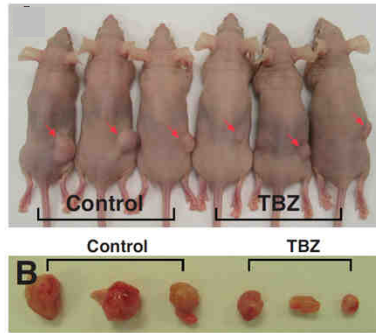


Cha et al., *PLoS Biology* (2012)

reversibly...



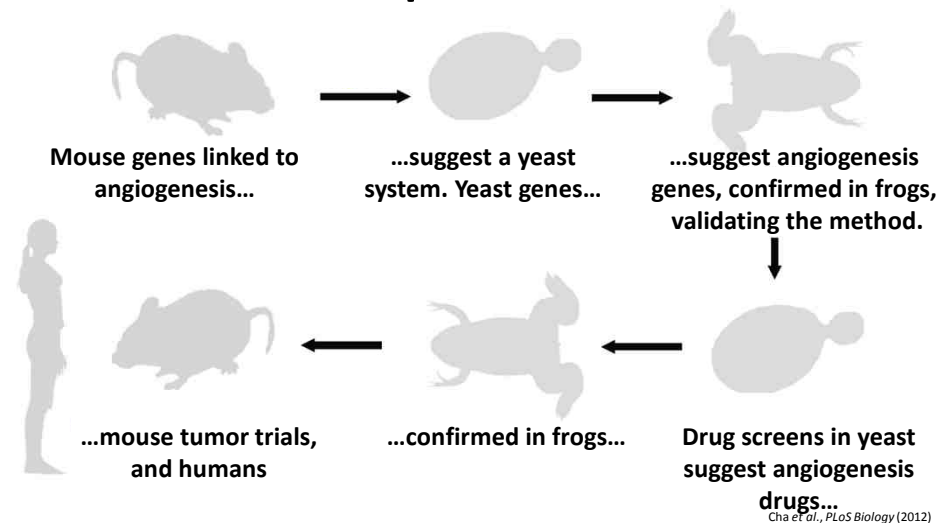
TBZ slows human fibrosarcoma tumors transplanted into immune-compromised mice



Vasculature in tumor sections

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

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



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

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