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

There was an interesting spat in 2016 over data sharing:

"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

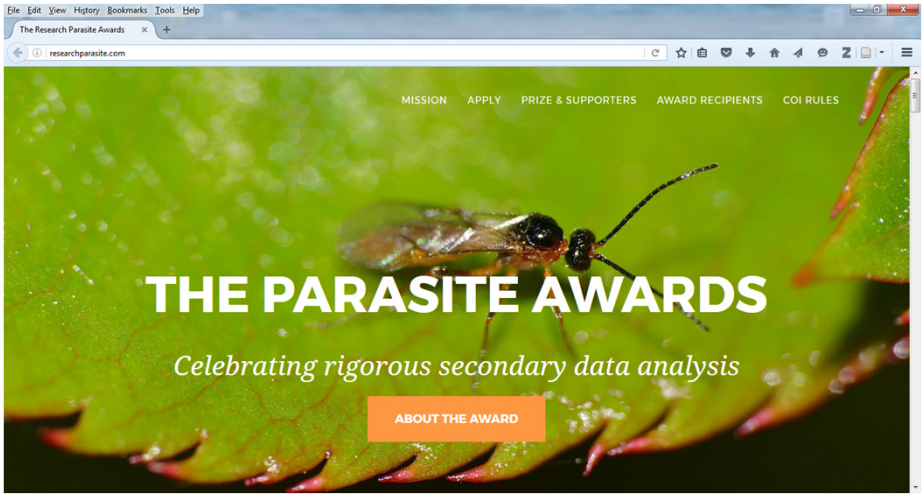
My opinion, FWIW, is that "research parasites" are

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.

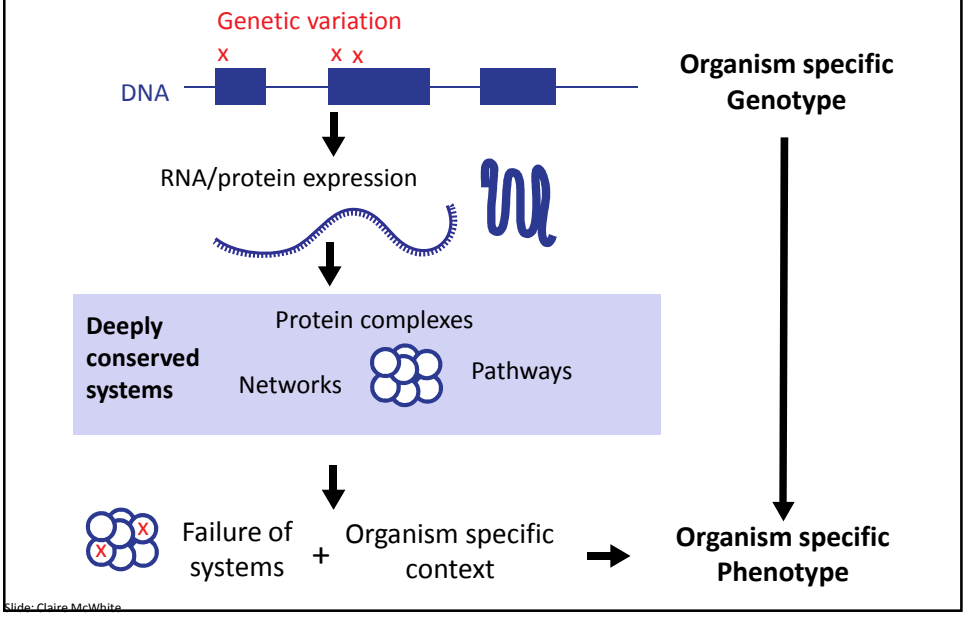
Science is improved when this happens.

Plus, you could win big!



http://researchparasite.com/

Conserved systems are key to connecting genotype to phenotype



Slide: Claire McWhite

We share genes with almost every known organism

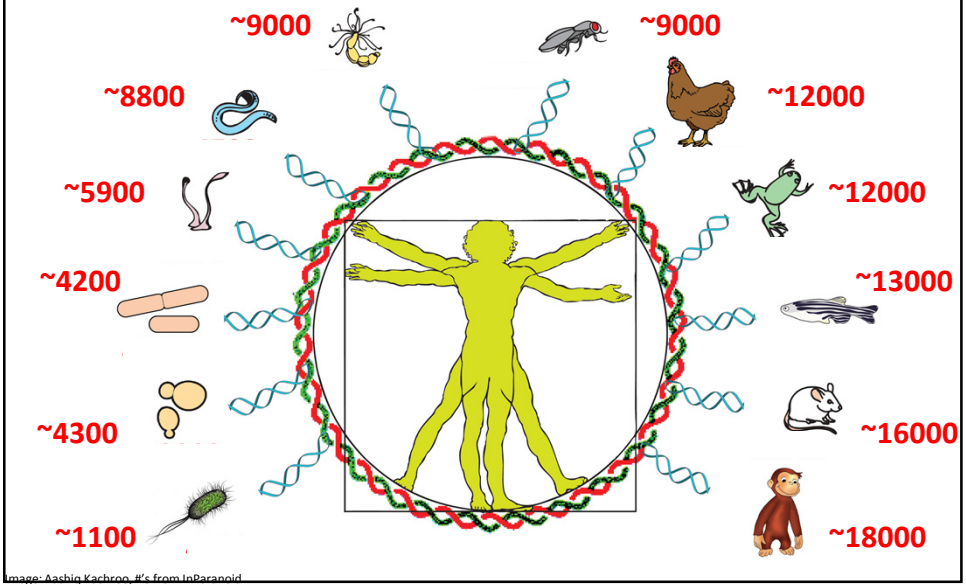


Image: Aashiq Kachroo, #'s from InParanoid

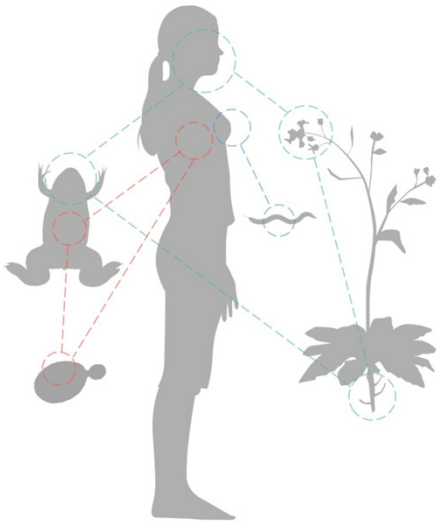


Illustration by Kathryn Weir

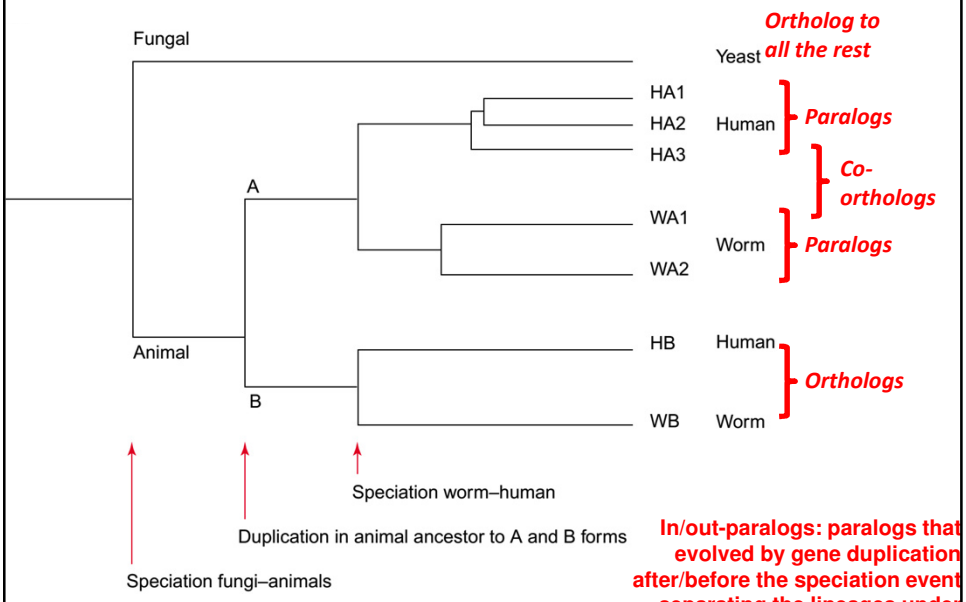
Corollary:
All genetic traits and diseases affect molecular structures that are evolutionarily conserved.

Comparative evolution studies rely on finding orthologs

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

<u>Paralogs</u> = genes that derive from a single gene that was duplicated within a genome

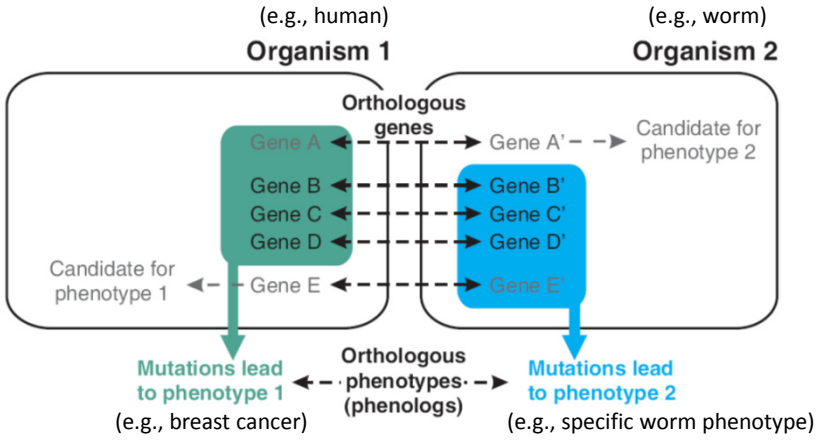
Comparative evolution studies rely on finding orthologs



In/out-paralogs: paralogs that evolved by gene duplication after/before the speciation event separating the lineages under consideration

Trends in Genetics 18:619–620 (2002)

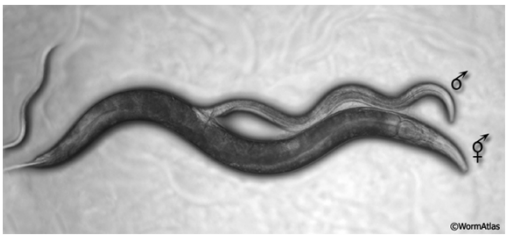
Phenologs = significantly overlapping sets of orthologous genes, such that each gene in a given set gives rise to the same phenotype in that organism

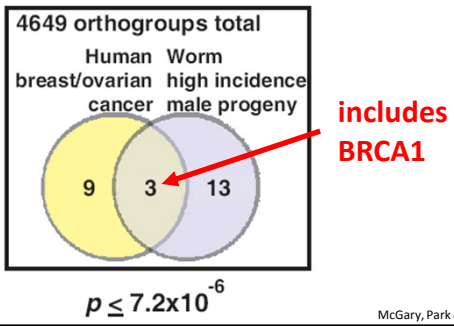


McGary, Park *et al*. *PNAS* 107:6544-9 (2010)

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

Human/Worm Ortholog	Linked to breast cancer in humans	Linked to more males in worms
ATM/atm-1	X	
BRIP1/dog-1	X	
KRAS/let-60	X	
PHB/phb-1	X	
PIK3CA/age-1	X	
RAD51/rad-51	X	
RAD54L/rad-54	X	
SLC22A18/C53B4.3	X	
TSG101/tsg-1010	X	
BARD1/brd-1	X	X
BRCA1/brc-1	X	X
CHEK2/chk-2	X	X
FAM82B/F33H2.6		X
GCC2/hcp-1,hcp-2		X
HMG20A,B/W02D9.3		X
HORMAD2,1/him-3,htp-1,2		X
KIF15/klp-10,18		X
MRE11A/mre-11		X
PIGA/D2085.6		X
RAD1/mrt-2		X
RAD21/coh-1		X
SEH1L/npp-18		X
SVIL/viln-1		X
TSPO,BZRPL1/C41G7.3		X
WDHD1/F17C11.10		X





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

Building & searching a collection of phenotypes

Mining available databases +
manual collection from the primary literature

↓

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

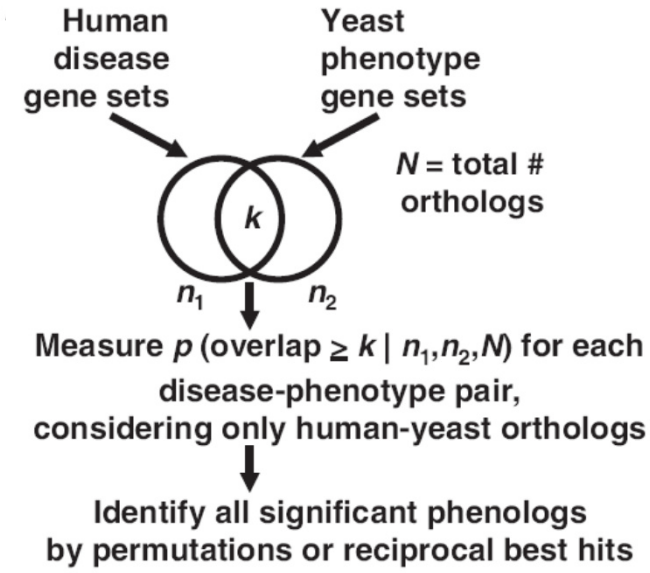
Spanning ~300 human diseases,
>7,000 model organism mutational phenotypes

↓

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)

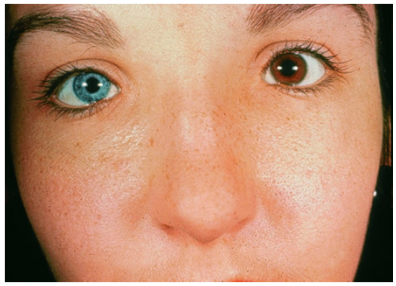
Discovering phenologs



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)

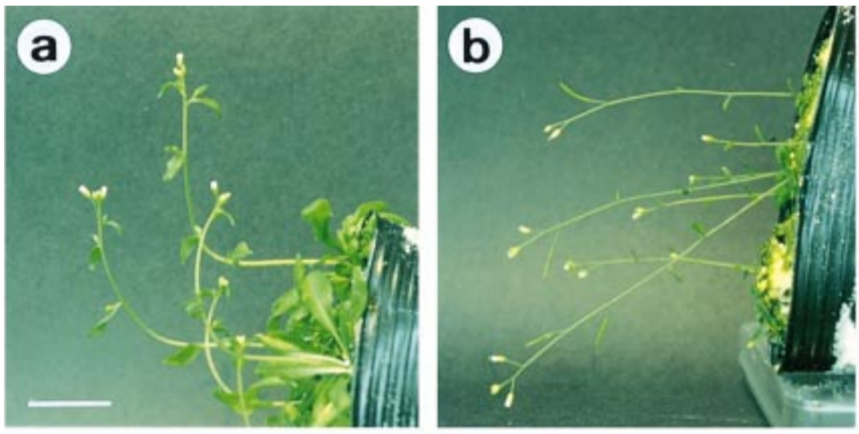


Michael Murphy, M.D.

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

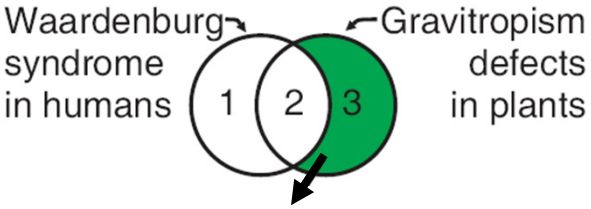
Assorted websites, incl. www.verywell.com/waardenburg-syndrome-1048952, http://stephaniemariesanchez.blogspot.com/

Plants sense and respond to gravity → gravitropism

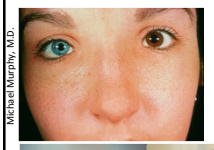


Fukaki *et al., The Plant Journal* 14, 425–430 (1998)

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



The human versions of these plant genes are <u>candidate Waardenburg genes</u>

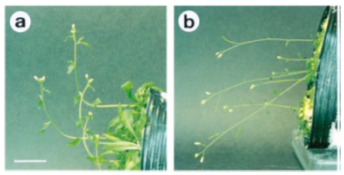


Michael Murphy, M.D.

Assorted websites

Waardenburg syndrome

≈

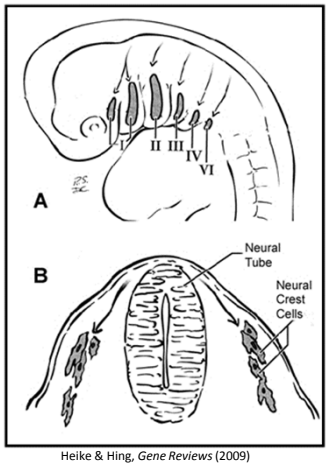


Fukaki *et al., The Plant Journal* 14, 425–430 (1998)

Gravitropism defects

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

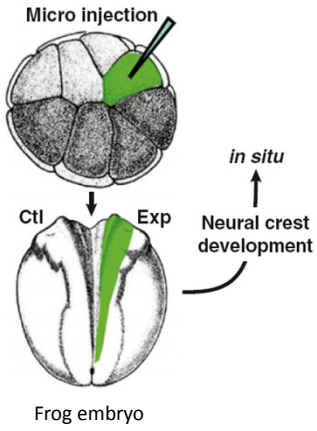
www.silvarcea.co.uk

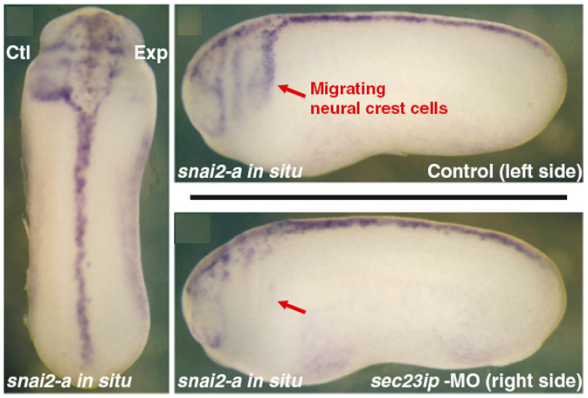
Association between white blue-eyed cats and deafness (noted by Darwin in 1859)

White forelock and deafness/bowel blockage in foals

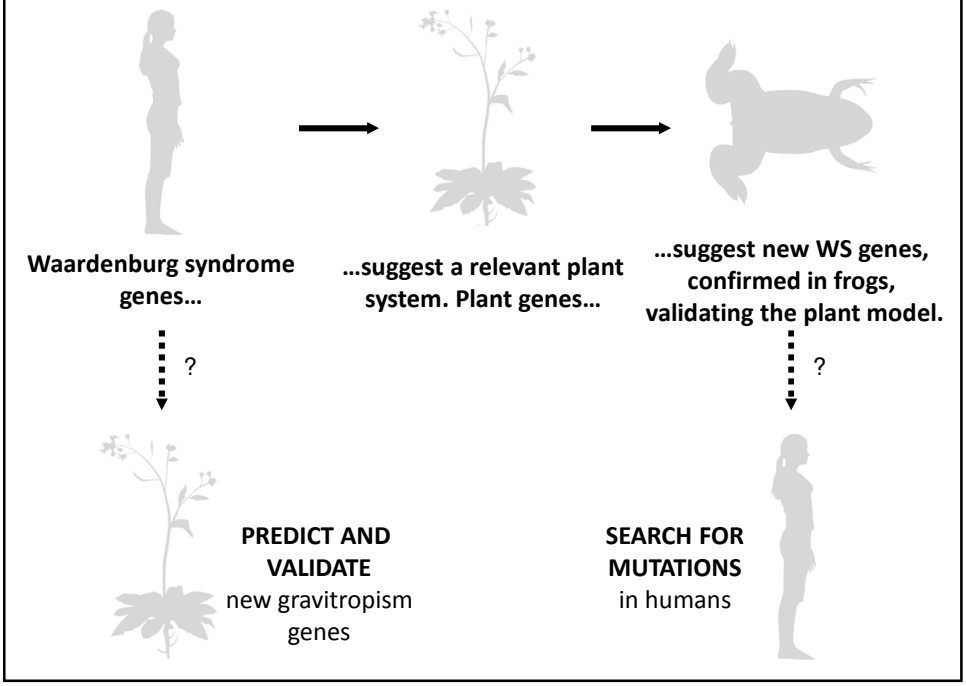
& many more...

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

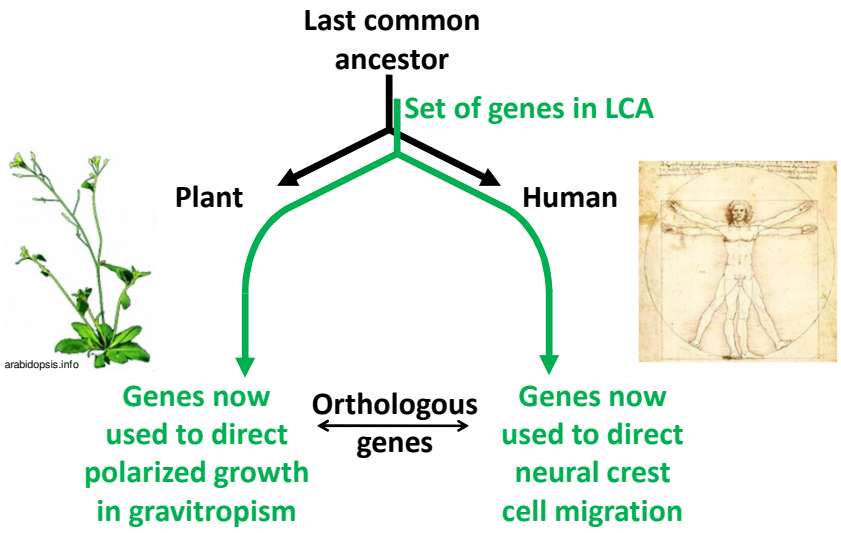




McGary, Park *et al*. *PNAS* 107:6544-9 (2010)



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



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

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

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

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?*

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 could 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?**

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 pain-stakingly curated relevant data, large-scale screens reporting easy-to-process data.**

Let's step through this particular discovery process:

5. **We started inventing/evaluating statistical models/algorithms, exploring the data and thinking about <u>how</u> to search for the relevant trends. We iterated steps 4/5 until we thought we understood the problem better.**

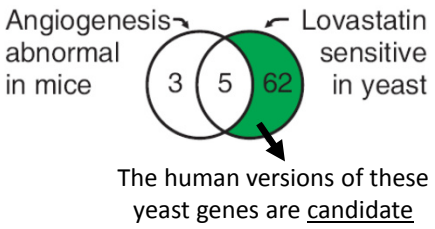
6. **At some point, the lab bet a 6 pack of beer on the outcome:**

Can we discover plant models of human disease?

Let's step through this particular discovery process:

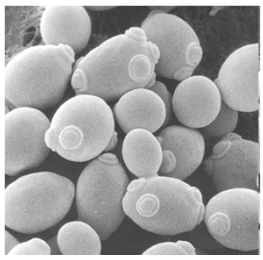
7. **The algorithms predicted some remarkable and crazy results. We had no option but to test or reject the new predictions, so began testing, thanks to collaborators in the Wallingford lab willing to sink a few weeks into high-risk experiments.**
8. **Some tests worked, some didn't. We went back & thought 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…**

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

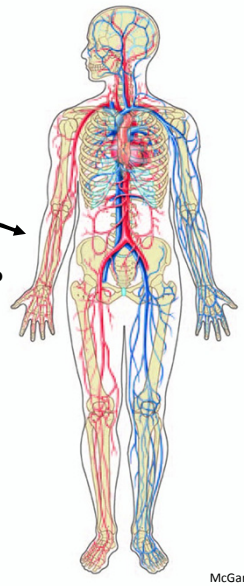


The human versions of these yeast genes are <u>candidate angiogenesis genes</u>

Can these really tell us about these?



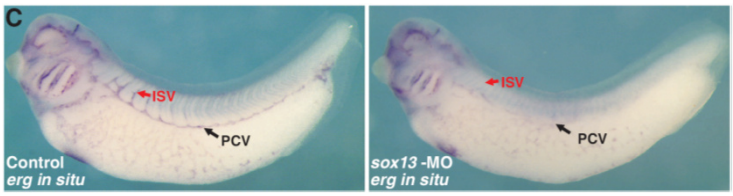
www.chemistryland.com



Dorling Kindersley

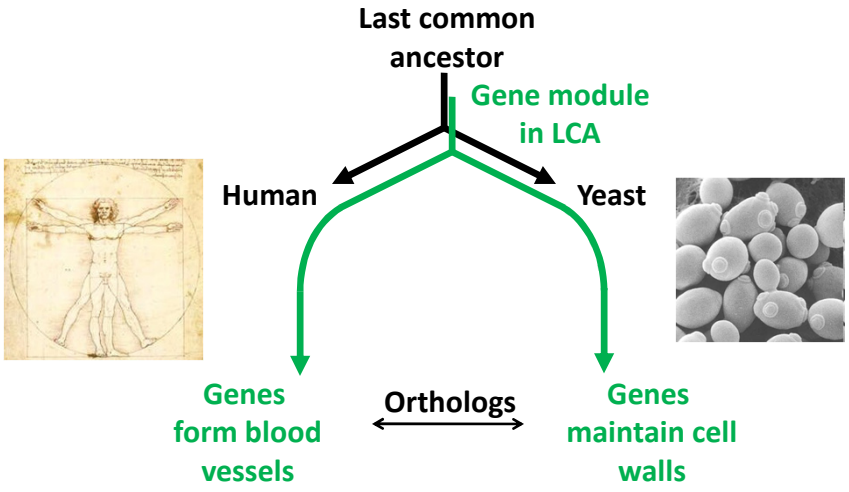
McGary, Park *et al.* *PNAS* 107:6544-9 (2010)

Disrupting the SOX13 gene causes strong blood vessel defects



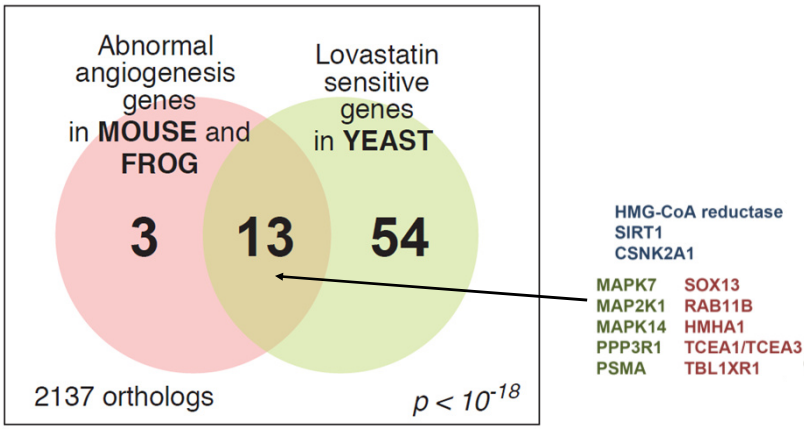
McGary, Park *et al.*
PNAS 107:6544-9 (2010)

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

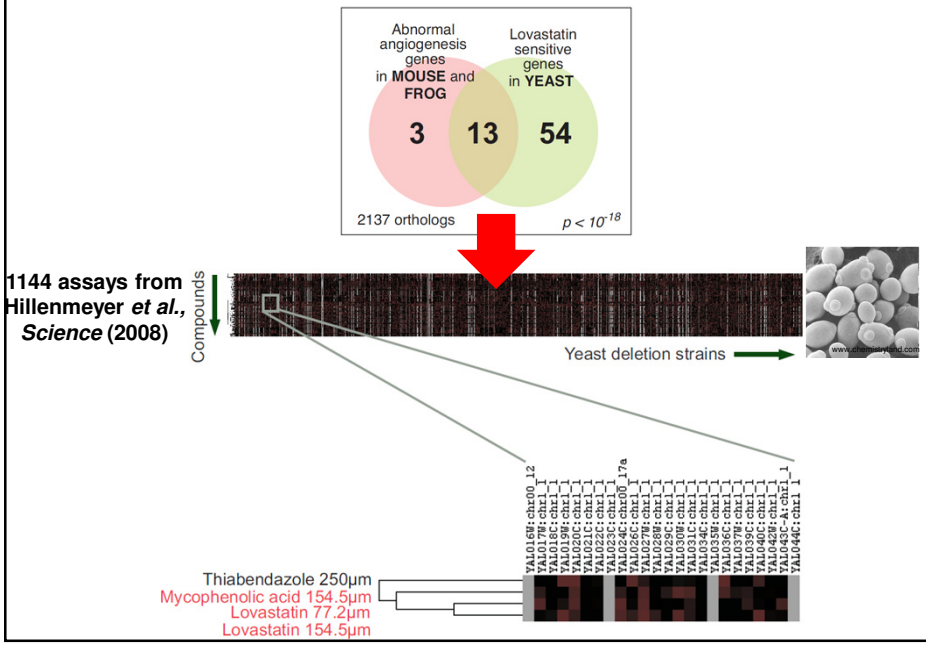


McGary, Park *et al. PNAS* (2010)

The yeast/angiogenesis gene module

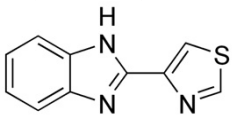


Chemicals that interact genetically with this module are candidate angiogenesis inhibitors



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

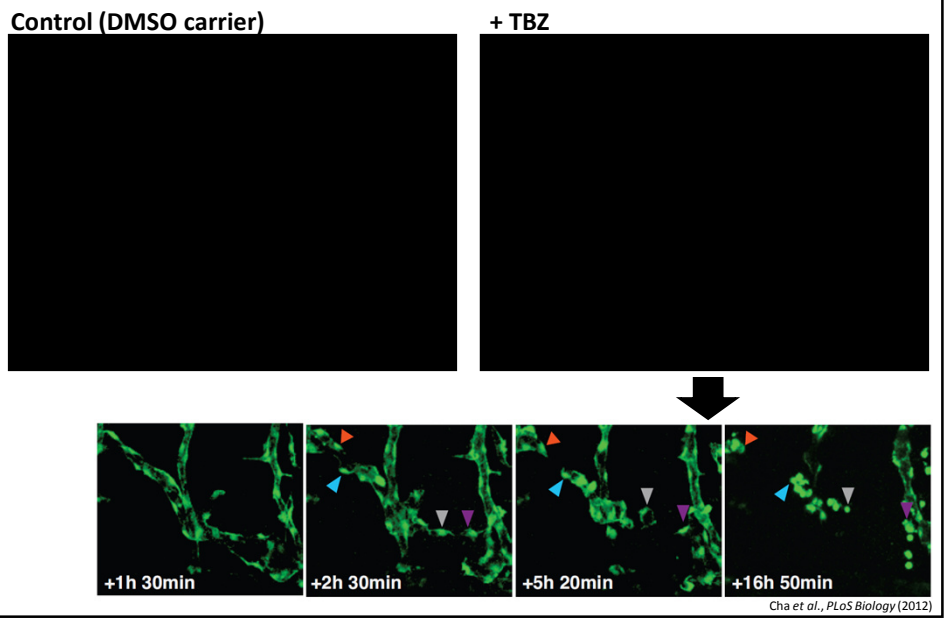


- **Approved by U.S. Food and Drug Administration in 1967**

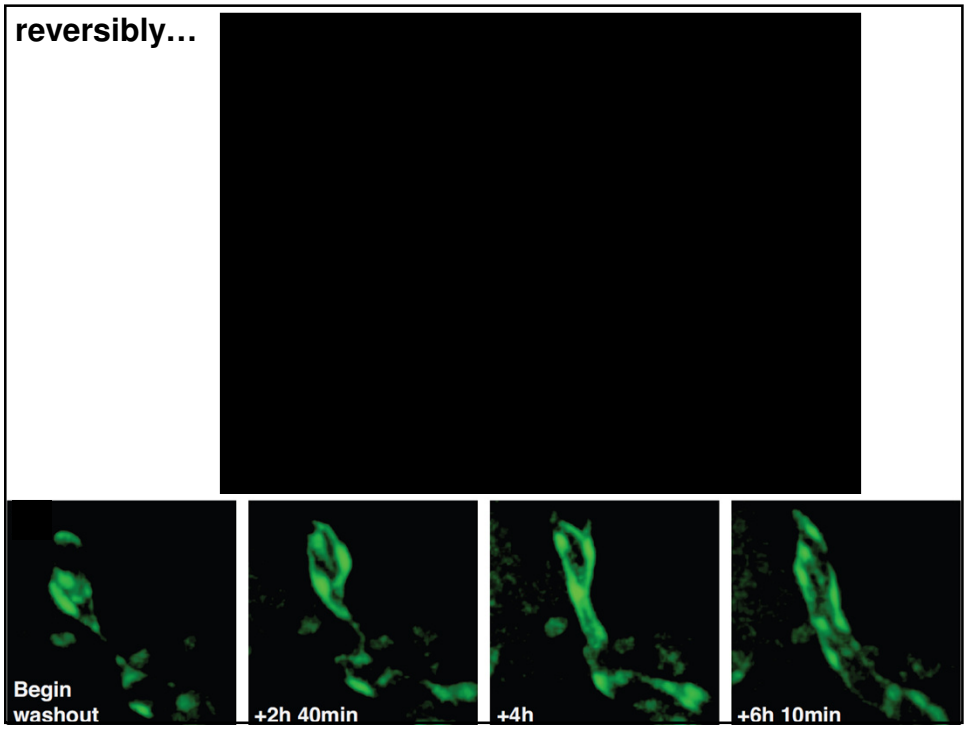
- Fungicide and parasiticide
- Not mutagenic or carcinogenic; 2 year dog safety trials
- Off-patent, marketed as a generic

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

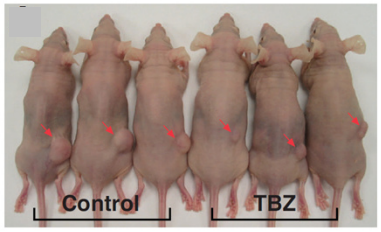
TBZ disrupts vascular integrity, making vascular endothelial cells retract & round up

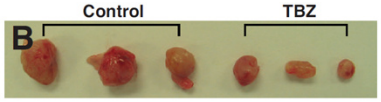


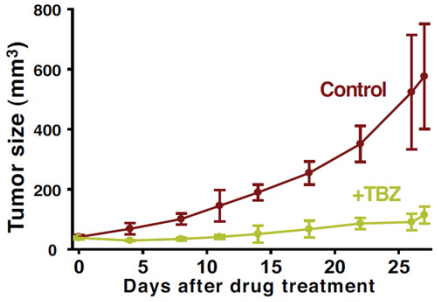
Cha *et al.*, *PLoS Biology* (2012)

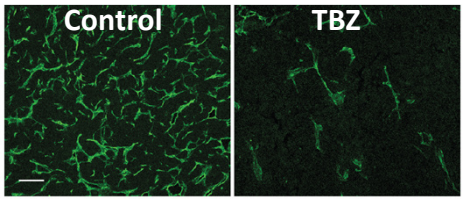


TBZ slows human fibrosarcoma tumors transplanted into immune-compromised mice





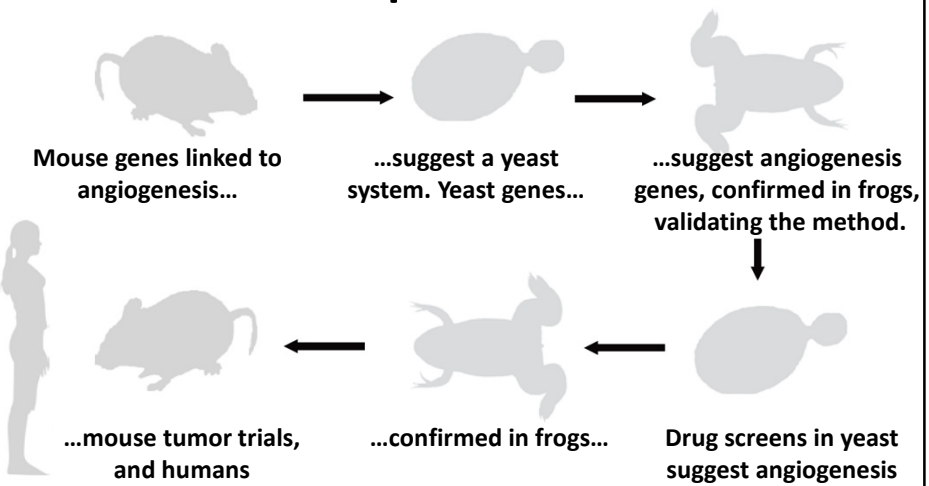




Vasculature in tumor sections

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