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

BCH364C/394P Systems Biology / Bioinformatics
Edward Marcotte, Univ of Texas at Austin

There was an interesting spat last year 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."

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/
Extraordinary Relatedness
We share genes with almost every known organism.

All genetic traits and diseases affect molecular structures that are evolutionarily conserved.
We know far more about genes & traits in lower organisms than in us. Can we transfer that knowledge to us?

![Graph showing number of unique gene-phenotype associations across different species over years.]

Can these tell us about us?


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.

Orthologs are genes from different species that evolved from a single ancestral gene before the divergence of those species.

Paralogs are genes that have evolved from a single ancestral gene within a single species.

Co-orthologs are orthologs that have evolved from the same ancestral gene in two different species.

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

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

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

<table>
<thead>
<tr>
<th>Human/Worm Ortholog</th>
<th>Linked to breast cancer in humans</th>
<th>Linked to norm male in worms</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATM/atm-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BRIP1/dog-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>KRAS/let-50</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PHB/phb-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PIK3CA/age-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RAD51/rad-51</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RAD54L/red-54</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SLC22A18/C53B4.3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TSG101/tsg-101</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BARD1/brd-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>BRCA1/brc-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>CHEK2/chk-2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>FAIM5B/F33H2.6</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>GCC2/hcp-1,hcp-2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>HRD20A,BV02D9.3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>KIF15/hlp-10,18</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>MRE11A/mre-11</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PIGA/D2085.6</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RAD1/mrt-2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>RAD21/coh-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SEH1L/np-18</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SSVL/visn-1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>TSPO.BZNPL1/C41G7.3</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>WDH1/F17C11.10</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

4649 orthogroups total

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

Building & searching a collection of phenotypes

Mining available databases + manual collection from the primary literature

Organism associations

<table>
<thead>
<tr>
<th>Organism</th>
<th>associations</th>
</tr>
</thead>
<tbody>
<tr>
<td>human</td>
<td>1,923</td>
</tr>
<tr>
<td>mouse</td>
<td>74,250</td>
</tr>
<tr>
<td>worm</td>
<td>27,065</td>
</tr>
<tr>
<td>yeast</td>
<td>86,383</td>
</tr>
<tr>
<td>Arabidopsis</td>
<td>22,921</td>
</tr>
</tbody>
</table>

Includes BRCA1

$p \leq 7.2 \times 10^{-6}$

Discovering phenologs

Human disease gene sets  Yeast phenotype gene sets

$N = \text{total} \# \text{ orthologs}$

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

Computationally, we find many genes shared between human diseases and mouse, yeast, worm, and even plant traits
Waardenburg syndrome accounts for ~2-5% of cases of deafness

Plants sense and respond to gravity → gravitropism

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

Waardenburg syndrome in humans

Gravitropism defects in plants

The human versions of these plant genes are candidate Waardenburg genes

Waardenburg syndrome

Gravitropism defects

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...

Heike & Hing, Gene Reviews (2009)
Sure enough, inactivating one of the genes—predicted from plants—in a tadpole disrupts neural crest cells, consistent with Waardenburg syndrome.

Waardenburg syndrome genes... → ...suggest a relevant plant system. Plant genes... → ...suggest new WS genes, confirmed in frogs, validating the plant model.

PREDICT AND VALIDATE new gravitropism genes

SEARCH FOR MUTATIONS in humans
Phenologs identify evolutionarily conserved systems of proteins relevant to particular traits/diseases.

Last common ancestor

Set of genes in LCA

Plant

Genes now used to direct polarized growth in gravitropism

Orthologous genes

Genes now used to direct neural crest cell migration

Human


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

Angiogenesis abnormal in mice

Lovastatin sensitive in yeast

The human versions of these yeast genes are candidate angiogenesis genes

Can these really tell us about these?


www.chemistryland.com

Dorling Kindersley
Disrupting the SOX13 gene causes strong blood vessel defects


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

*kdr:GFP transgenic* *Xenopus laevis*
TBZ disrupts vascular integrity, making vascular endothelial cells retract & round up

Control (DMSO carrier) + TBZ

Cha et al., PLoS Biology (2012)

reversibly...

Begin washout +2h 40min +4h +6h 10min
TBZ slows human fibrosarcoma tumors transplanted into immune-compromised mice

![Graph showing tumor size over days after drug treatment with TBZ compared to control.]

Vasculature in tumor sections

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

Mouse genes linked to angiogenesis... → ...suggest a yeast system. Yeast genes... → ...suggest angiogenesis genes, confirmed in frogs, validating the method.

...mouse tumor trials, and humans → ...confirmed in frogs... → Drug screens in yeast suggest angiogenesis drugs...

[Images and text from 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