



## Celebrating Similarities—Embracing Differences

Laurie Goodman

*Genome Res.* 1999 9: 797-800

Access the most recent version at doi:[10.1101/gr.9.9.797](https://doi.org/10.1101/gr.9.9.797)

---

### References

This article cites 20 articles, 12 of which can be accessed free at:  
<http://genome.cshlp.org/content/9/9/797.full.html#ref-list-1>

### Creative Commons License

This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genome.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported License), as described at <http://creativecommons.org/licenses/by-nc/3.0/>.

### Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

---

To subscribe to *Genome Research* go to:  
<http://genome.cshlp.org/subscriptions>

---

# Celebrating Similarities—Embracing Differences

Laurie Goodman<sup>1</sup>

Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724 USA

Using model systems to gain a better understanding of human disease and the underlying biology is standard in biological research. The power of model systems, however, has never been so great as in recent years with the growing availability of a wealth of genomic data from a variety of animals and plants and the ever more-advanced tools for probing, classifying, and characterizing these resources. There are few better examples illustrating the usefulness of model systems as the material presented at the “Yeast Genetics and Human Disease II” meeting, which took place 3 months ago in Vancouver, Canada (June 24–27, 1999). The meeting, which included talks that covered far more than just yeast as a model system, was undeniably successful. Meeting attendees’ primary complaint, in fact, was the press for time; poster sessions started at 7:45 a.m. and talks ended at 10 p.m. This is a relative standard day for most meetings, except for the very early rise. The complaint, however, was due to the fact that most attendees sat through every single talk, rather than taking breaks here and there. A quick dash out into the hallway for a bathroom break or a coffee refill showed an empty corridor, rather than what is usually seen at meetings—a hallway containing several clustered groups talking about previous presentations, waiting for talks that will come up later in the day, or discussing current or potential collaborations with colleagues.

The information presented at the meeting was compelling, though not necessarily for its newness but, rather, for the context of the presentations. The composition of talks and attendees varied widely in terms of system and subtopic studied, but the theme in each ses-

sion formed a cohesiveness that enabled listeners to appreciate the broader perspective and to see clearly the power gained in utilizing information from a broader spectrum of organisms and tools. Like the first meeting held 2 years ago in Baltimore, this conference held forth the strength of yeast (for the most part, especially in this report, *Saccharomyces cerevisiae*) as a system to use to understand human disease; but with 2 years time to look at the first completely sequenced eukaryotic genome, there was also an increasing awareness of the differences.

Sequence and, potentially, functional homology of genes and proteins are the primary reasons why model systems are at all useful. That genes, proteins, and biological systems share a level of similarity in all organisms can often be taken for granted but just how much similarity different organisms share can sometimes be surprising given their evolutionary distances. Even given this surprising similarity, differences do exist, and as such, extrapolation from one system to another requires caution. The power in using model systems comes from being able to define both the similarities and the differences, and to use these in an intelligent way to best direct the research.

Functional similarity has numerous applications, but the relevance of that similarity from one organism to another may be far from complete. Worse, as was pointed out by Geoff Duyk, when considering a particular model organism’s overall relevance to humans, the ease in which that organism can be worked with in the laboratory is generally inversely proportionate to its relevance. But this is *overall* relevance. Growing understanding of functional similarity and/or homology and the use of this as a tool to direct research in a pinpointed

fashion can help one pick and choose when to utilize a system that allows ease at getting at specific aspects of a biological pathway and when a whole organism approach is required to better understand the implications of a genetic change on the organism as a whole, and how that might relate to humans.

As previously indicated, the presentations at the Vancouver meeting were classified by their area in biology, for example, cell cycle, RNA metabolism and translation, chromosome and genome stability, and so on. But there were other ways to classify these talks, ways that transcended their biological classifications—making each talk useful from a conceptual standpoint to anyone, no matter their “biological sect.” Experiments from any biological discipline could really fall into three categories: use of a model system to define a disease gene (or unknown gene) function; use of a model system to dissect cellular mechanisms; and use of large data sets to classify and characterize novel genes in known biological systems, pathways, or components.

Given yeast, with its complete genome, in conjunction with the plethora of other model organisms, some also with complete genomes (*Caenorhabditis elegans* and numerous bacteria and archaea), as well as growing sequence data for mouse, human, *Drosophila*, *Arabidopsis*, and so on, current advances are and should continue to cross species and disciplinary boundaries, as this can only increase our understanding of general biological systems and our own place within these.

## Gene and Protein Homology

Gavin Sherlock (Stanford University) and Eugene Koonin (National Institutes of Health) provided meeting attendees

<sup>1</sup>E-MAIL [goodman@cshl.org](mailto:goodman@cshl.org); FAX (516) 367-8334.

with the greatest overview for what is found when identifying sequence homology when comparing genomes. Sherlock focused on an analysis of the complete yeast genome compared with last year's completed *C. elegans* genome sequence (Chervitz et al. 1998), whereas Koonin presented a series of comparative studies from a variety of organisms (many still with incomplete genomes), providing examples of sequence homology from particular gene families and protein domains studies, such as nucleotidyltransferases (Aravind and Koonin 1999) or BRCT domains (identified originally in *BRCA*) (Bork et al. 1997) present in bacteria, yeast, *C. elegans*, and humans. The main message from both talks was that while complete proteins themselves are not necessarily conserved (and in most cases, unlikely to be conserved), functional domains persist over evolution. A domain, as defined by Koonin, must fulfill three criteria: (1) It should be statistically significant, that is there is a low expectation that the domain would occur by chance alone; (2) it should have a level of individuality; a presence in different contexts in different proteins; and (3) it should have an evolutionary resilience of the conserved features and a correlation of these features with structural and functional cognates. Both Koonin and Sherlock showed that while functional domains are conserved, the number and overall architecture of proteins from one species to another is not conserved. Generally, as the complexity of an organism increases, the complexity (in terms of number and position of functionally conserved domains) and the overall number of increasingly complex proteins increase. Work now proceeds to define as many functional domains as possible to generate a complete list of all these domains in every organism. Clearly, such a catalog would be invaluable; a point made infinitely apparent in other talks at the meeting, where studies relied heavily on both the use and understanding of gene homologues of different organisms in order to answer biological questions.

### Defining Disease Gene Function

The identification of human disease

genes continues to become more straightforward as more sequence data, higher quality maps, and better resources and tools become available. But the identification of the gene is really only the starting point for understanding the cause of a disease. Obviously studying blindness or brain disease in organisms such as yeast or *C. elegans* may seem to lack the main components for getting at such diseases—the presence of eyes or brains. Comparative studies, however, provide one of the greatest tools for getting at the mechanisms underlying disease at the cellular level. Although such studies are unlikely to give the whole story, they can provide vital clues for where to search in understanding the role of the human homolog.

Several nice examples of these types of studies were presented at the Vancouver meeting. One presentation was the analysis of the yeast Batten disease gene homolog. Batten disease is a recessive, progressive neurodegenerative disorder, commonly associated with blindness, loss of mental abilities, motor skills, increasing number and severity of seizures, and, finally, premature death. One of the cellular features of the disease is an accumulation of material in the lysosomes. The gene underlying this disorder, *CLN3*, was identified in 1995 by The International Batten Disease Consortium. Analysis of the gene indicated that it was completely novel and not much more beyond that. Moving to yeast, David Pearce and colleagues (University of Rochester) identified a *CLN3* yeast homolog, *BTN1*, and determined that its gene product is located in the yeast vacuole. Then, using microarray analysis to compare the expression of all yeast genes in *BTN1* and *btn1* yeast cells, the researchers found a change in expression of two genes: *HSP30* expression decreased and *BIN2* expression increased in *bin1*. Of interest, *HSP30* expression was known to increase in response to increased plasma membrane ATPase activity when vacuolar membrane pH regulation is altered. Based on these findings, Pearce and colleagues propose that Batten disease may be caused by a defect in vacuolar (or in hu-

mans, lysosomal) pH control. (Pearce et al. 1999; Pearce and Sherman 1999). Focus should move to investigate the potential role pH may play in Batten disease cells.

In a similar type of study, Carla Koehler (University of Basel) analyzed yeast homologs to gain a better understanding of the cellular mechanisms underlying Mohr-Tranebjaerg syndrome, which is a human deafness dystonia. The disease is caused by a defect in another gene of unknown function, *DDP1* (Tranebjaerg et al 1995; Jin et al. 1996). *DDP1* is on the X chromosome. The disease, however, does have some similarities to mitochondrial disorders, as it especially affects nerve and muscle tissues. Koehler and colleagues, investigating protein import into mitochondria, identified several small proteins (Tim8, Tim9, Tim10, Tim12, and Tim13) that are similar to *DDP1* (it had greatest similarity to Tim8). They also found that *DDP1* is a mitochondrial protein. Tim8 appears to mediate the import of a particular subset of proteins into the mitochondria (Koehler et al. 1999). Further studies on the activity of Tim8 and on *DDP1* should aid in elucidating the exact biochemical role the protein plays in the cell.

Both of these examples, along with many more at the meeting, provide a clear indication of the way in which work with yeast homologs can provide useful clues for determining the function of a disease gene. Obvious care must always be taken as to how far the similarity in function might extend. Recall that while studies do indicate a conservation of functional domains, with domain swapping and the creation of new and more complex proteins, purported homologs may share some of the basic biochemical activities; but whether this similarity extends to any biological relevance remains suspect. It is vital to consider additional points, such as the extent of the similarity between protein domain architecture, parallels in cellular location, and even protein-protein interactions of your protein of interest and those of its homolog. Homology studies can give important clues, but these must be used in a biological and evolutionary context.

## Dissecting Cellular Mechanisms

Given a high similarity of function between two genes or proteins of different organisms, the organism that is simpler to work with is the system of choice for studying details of cellular mechanisms. Several talks demonstrated the strength of the yeast system, both from a genetic and a biochemical standpoint for looking at the details of a variety of cellular mechanisms. Both Reed Wickner (The National Institutes of Health) and Sue Lindquist (University of Chicago) (Lindquist et al. 1998; Zhou et al. 1999) presented data on their studies of prion-like proteins in yeast. While it is (hopefully) unlikely that newspapers will suddenly begin crying “Mad Yeast Disease,” producing a rabid public concern about the dangers in eating bread, there are several yeast genes that produce protein products with prion-like behavior, making this a potentially terrific system to study prions.

There are a number of yeast genes that produce a normal form of their protein, but also an altered form that aggregates, and can aggregate their normal counterparts. Wickner presented data on [URE3], the prion form of the *URA2* gene (Edskes et al. 1999, Maddelein and Wickner 1999, Taylor et al. 1999), and Lindquist presented data on [PSI], the prion form of the *SUP35* gene. Both Wickner and Lindquist presented data from systems they had developed to express particular domains of the proteins, which allowed them to define which regions of the protein were required or sufficient for prion activity. Given these data, Wickner pointed out that the yeast system could be used as a sort of prion “Ames test,” allowing researchers to look for prion-inducing and curing agents. Yeast could also be used for molecular screening for protein domains that are able to produce infectious proteins.

What might be considered standard cellular activity can also be easily studied in yeast and applied to other organisms from a mechanistic standpoint. For example, Diane Cox (University of Alberta) and Jonathan Gitlin presented details of the use of yeast for studying cop-

per transport and homeostasis in yeast. As a model, the data provided information on how copper transport affected its utilization in humans and the potential roles in disease by human homologs (Forbes et al. 1999, Suzuki and Gitlin 1999). Other cellular activities, such as replication and the mechanisms involved have a long history of being well-served by initial analyses in yeast. Doug Koshland (Carnegie Institute of Washington) presented data on sister chromatid cohesion that continue to highlight the strength of the yeast system. In the work by Koshland and colleagues, the yeast two-hybrid system helped determine some of the proteins that bind to centromeres and cohesion sites. They were also able to map the region within the yeast centromere that these proteins bind and to investigate the sequences required for such binding. The data indicated that the centromeric region is required to load these proteins and that they spread from there. This “loading region” is also required to keep these proteins on the sister chromatids during cell division (Megee and Koshland 1999).

## Large-Scale Identification of Novel Genes

While yeast has been and clearly continues to be useful for helping to decipher unknown gene function and to more easily study cellular mechanisms, the now complete sequence of the yeast genome allows researchers to apply new tools to better understand the function of genes and gene families in a large-scale fashion. The primary focus at the meeting for classifying gene function in large-scale studies included, for example, studies based on gene expression, in which the complete array of yeast genes were analyzed using DNA array technology, and the expression patterns were examined and characterized. Bruce Futcher (Cold Spring Harbor Laboratory) presented data from which he and his colleagues looked for genes whose expression varied with cell cycle, providing a set of ~700 genes that are likely to be involved in the cell cycle (Spellman et al. 1998). In addition to

gene expression analysis, Eric Phizicky (University of Rochester) presented a method to identify genes based on the activity of their proteins. Phizicky and colleagues expressed all the yeast genes as tagged proteins and then assayed these tagged proteins for particular biochemical activities. The genes for these proteins, once linked with some known biochemical function, can be identified and analyzed further, and their genes classified based on these biochemical functions. In a similar fashion, proteins (and the genes encoding them) can be identified based on broader biological bases. James Broach (University of Princeton), for example, presented a method whereby yeast could be used as a system to pull out unknown human G-protein-coupled receptors (Klein et al. 1998). In this last example, once again due to similarity of cellular systems from one organism to the next, yeast can be used as a tool in which to analyze proteins from other organisms.

The use of expression data for classifying and characterizing large numbers of genes was a main theme at the meeting. The usefulness of such analyses for tackling the large data sets that come from completely sequenced genomes is undeniable, as these classifications provide the first clues as to what all of these genes are doing and how they work together to maintain all cellular functions in all environments. Equally clear is that a public database that allows researchers to dip into such expression (or biochemical activity) databases would be an incredibly useful tool for the community. It would allow new data by anyone to be compared and contrasted with data already accumulated and stored. As an example, Douglas Bassett and Matthew Marton (both at Rosetta Inpharmatics) presented information on Rosetta's private expression database (Marton et al. 1998). Their database stores and allows the comparison and analysis of thousands of yeast gene-expression profiles from DNA microarray experiments. The data appeared to provide quite powerful and useful comparisons of changes in gene expression from different genetic perturbations, drug treatments, responses to different growth conditions,

and so on, allowing researchers using the database to make connections between, for example, genetic alterations and environmental changes, or as an aid to drug target selection or even drug selection for different genotypes. Seeing these studies, and recognizing the power in many of the other expression studies presented at the meeting, the development of a public expression database will certainly be as useful as current gene and protein sequence databases are for classifying genetic unknowns based on sequence similarity.

### Conclusions

All in all, with all the advances being made, there is still a long way to go. More genomes await their completion—although there is growing excitement about the *Drosophila* genome sequence to be released this month from Celera in collaboration with researchers at The University of California, Berkeley; the completion of the *Arabidopsis* genome next year; and increasing prospects from the expeditious work on the human and mouse genomes. These soon-to-come completely sequenced genomes certainly raise our anticipation about what we will be able to discover about ourselves and the world around us, but before we can begin with our future, there remains a great deal more to do with sequence in hand.

“In yeast, we have had the whole genome complete for 2 years now, and we don’t know where transcripts start, or even the definition of a gene. As a community, we really need to focus on this,” Ira Herskowitz pointed out at the

meeting, in consideration of Munira Basrai’s presentation on the difficult matter of identifying the often ignored small ORFs (<100 codons) that are present in the yeast genome—ORFs too small to identify as potential genes when using current gene-finding computer programs, a problem since some of these small ORFs clearly have biological relevance (such as the 36-amino acid **a** factor precursor protein).

That said, there remains an equal amount to celebrate. The collaborative spirit at the Vancouver meeting, the high interest the attendees had for seemingly disparate studies, and the ease with which different disciplines shared data and ideology indicate an exciting road ahead with abundant data, new tools, more ideas, and increasing collaborations spanning organisms and disciplines.

The “Yeast and Human Disease III” meeting, 2 years hence should, if anything, only be more interesting.

### REFERENCES

- Aravind, L. and E.V. Koonin. 1999. *Nucleic Acids Res.* **27**: 1609–1618.
- Bork, P., K. Hofmann, P. Bucher, A.F. Neuwald, S.F. Altschul, and E.V. Koonin. 1997. *FASEB J.* **11**: 68–76.
- Chervitz, S.A., L. Aravind, G. Sherlock, C.A. Ball, E.V. Koonin, S.S. Dwight, M.A. Harris, K. Dolinski, S. Mohr, T. Smith, S. Weng, J.M. Cherry, and D. Botstein. 1998. *Science* **282**: 2022–2028.
- Edskes, H.K., V.T. Gray, and R.B. Wickner. 1999. *Proc. Natl. Acad. Sci.* **96**: 1498–1503.
- Forbes, J.R., G. Hsi, and D.W. Cox. 1999. *J. Biol. Chem.* **274**: 12408–12413.
- The International Batten Disease Consortium. 1995. *Cell* **82**: 949–957.
- Jin, H., M. May, L. Tranebjaerg, E. Kendall, G. Fontan, J. Jackson, S.H. Subramony, F. Arena, H. Lubs, S. Smith et al. 1996. *Nat. Genet.* **14**: 177–180.
- Klein, C., J.I. Paul, K. Sauve, M.M. Schmidt, L. Arcangeli, J. Ransom, J. Trueheart, J.P. Manfredi, J.R. Broach, and A.J. Murphy. 1998. *Nat. Biotechnol.* **16**: 1334–1337.
- Koehler, C.M., D. Leuenberger, S. Merchant, A. Renold, T. Junne, and G. Schatz. 1999. *Proc. Natl. Acad. Sci.* **96**: 2141–2146.
- Lindquist, S., S.K. DebBurman, J.R. Glover, A.S. Kowal, J.J. Liu, E.C. Schirmer, and T.R. Serio. 1998. *Biochem. Soc. Trans.* **26**: 486–490.
- Maddelein, M.L. and R.B. Wickner. 1999. *Mol. Cell Biol.* **19**: 4516–4524.
- Marton M.J., J.L. DeRisi, H.A. Bennett, V.R. Iyer, M.R. Meyer, C.J. Roberts, R. Stoughton, J. Burchard, D. Slade et al. 1998. *Nat. Med.* **4**: 1293–1301.
- Megee, P.C. and D. Koshland. 1999. *Science* **285**: 254–257.
- Pearce, D.A., T. Ferea, S.A. Nosel, B. Das, and F. Sherman. 1999. *Nat. Genet.* **22**: 55–58.
- Pearce, D.A. and F. Sherman. 1999. *Mol. Genet. Metab.* **66**: 314–319.
- Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. 1998. *Mol. Biol. Cell* **9**: 3273–3297.
- Suzuki, M. and J.D. Gitlin. 1999. *Pediatr. Int.* **41**: 436–442.
- Taylor, K.L., N. Cheng, R.W. Williams, A.C. Steven, and R.B. Wickner. 1999. *Science* **283**: 1339–1343.
- Tranebjaerg, L., C. Schwartz, H. Eriksen, S. Andreasson, V. Ponjavic, A. Dahl, R.E. Stevenson, M. May, F. Arena, D. Barker et al. 1995. *J. Med. Genet.* **32**: 257–263.
- Zhou, P., I.L. Derkatch, S.M. Uptain, M.M. Patino, S. Lindquist, and S.W. Liebman. 1999. *EMBO J.* **18**: 1182–1191.