Genome **Biology**

RESEARCH HIGHLIGHT

Phenotypic connections in surprising places

Bolan Linghu[1] and Charles DeLisi[2]*

Abstract

Surprising correlations between human disease phenotypes are emerging. Recent work now reveals startling phenotype connections between species, which could provide new disease models.

The human disease landscape

The past few years have witnessed a growing number of well documented connections between and among human disease phenotypes, whose relationship would not have been obvious within the current disease classi-fication framework. The evidence stems from a variety of sources, spanning clinical epidemiology, computational genomics and various model systems [1-5]. The implica-tions are potentially so fundamental to disease etiology, drug development and the general diagnostic paradigm that there have been calls for an NIH Roadmap (large trans-institute transformational grants) focused on delineating the human disease landscape - a quantitative bipartite correlation map relating disease phenotypes and their genetic structures. A new study of phenotype correlations [6] now takes this idea further.

One of the earliest comprehensive studies of human phenotype correlations was by Rzhetsky *et al.* [5], who established a correlation network between 161 diseases and disorders using evidence of comorbidities obtained from some 1.5 million patient records. Among other results, they found suggestive evidence for genetic rela-tions between autism, which manifests in childhood, and several late onset diseases, including bipolar disorder and schizophrenia.

A different approach was taken by Goh *et al.* [1], who constructed a human disease network by linking genetic disorders that are known to share causative genes. The network was based on an analysis of the Online Mendelian Inheritance in Man (OMIM) database, the

most comprehensive compendium of well established associations between human disorders and their asso-ciated genes [7]. Among their findings was a large subnet formed by 516 of the 1,284 disorders studied, clearly showing many-to-many relationships between pheno-types and genes. For example, *KRAS* (encoding a small GTPase), *BRCA1* and *BRCA2* (both encoding tumor suppressors) are all involved in breast cancer, but *KRAS* is also implicated in pancreatic cancer, whereas *BRCA1* and *BRCA2* are associated with papillary carcinoma and prostate cancer, respectively [7]. The results of these and other studies support the concept that genes underlying a disorder tend to be functionally related; for example, they could be part of a particular protein complex, a particular pathway or process, or a particular set of coexpressed genes. A disorder can then be viewed as a phenotype that emerges from dysfunction of one or more components of a functionally coherent gene module. An important aspect of modularization is that any gene in the module that was not previously identified with the phenotype is a candidate for association with it [8]. Moreover, a particular module can, to varying extents, underlie related pheno-types, opening up the possibility of linking phenotypes not only on the basis of shared genes, but also on the basis of functionally related genes that are not shared.

Inferring disease connectivity through functional linkage starts with a network of genes whose functions are correlated; that is, each pair of nodes (genes or proteins) is connected by one or more sources of evidence supporting its functional coherence, such as physical interaction, correlated expression, adjacency in the same metabolic pathway or genetic interaction [2-4]. Connections can then be inferred between the diseases whose associated genes are linked in the gene network. For instance, we and our colleagues [3] constructed a network for the human genome by integration of diverse types of evidence using a Bayesian model, and annotated it with all OMIM disease genes for diseases known to be associated with five or more genes. Genes that are most tightly linked to those known to be associated with a given disease are immediate candidates for association with that disease. Furthermore, connections between disease pairs can be quantitatively identified on the basis of the magnitude of the functional linkage between their disease-associated genes. Thus, two diseases can be

*Correspondence: delisi@bu.edu
[2]Departments of Biomedical Engineering and Physics, Program in Bioinformatics and Systems Biology, Boston University, Boston, MA 02215, USA
Full list of author information is available at the end of the article

linked even if they do not share disease-associated genes. The associations found ranged from phenotypically disparate disease pairs, such as multiple sclerosis with malaria, to phenotypically similar pairs, such as muscular dystrophy with myopathy. Such results suggest that the current disease classification may be much less informative than is commonly believed.

Recognition of molecular connections between disease phenotypes provides immediate insight into the molecular mechanisms underlying different diseases and can therefore generate novel hypotheses for therapeutic strategies. This is especially valuable if one disease is well studied and the other is not; it might also be valuable if viable drug targets have been found for one of the diseases but not the other. This prospect of drug repositioning could accelerate the introduction of therapies by years [9].

Human disease models and phenotypic connections in surprising places

Recent work from Edward Marcotte's laboratory by McGary *et al.* [6] takes these concepts, and indeed the entire field of phenomics, to an entirely new level by developing, establishing and exploring a quantitative method to find non-obvious phenotypic connections, not just within a species, but across species. Such cross-species phenotypic connections stem from the evolutionary conservation of the underlying associated gene modules [8,10]. The importance of the work reported by McGary *et al.* [6] can be glimpsed by recalling that whereas OMIM, which is the most extensive database of well established human gene-phenotype associations, contains approximately 5,400 unique associations [7], the numbers for some model organisms are 1 to 2 orders of magnitude higher [6]. Some of these relations - such as between obesity and its implicated genes in mice - have obvious equivalences in humans, and when they do, the laboratory model can serve as a useful surrogate to study the disease or disorder. Other phenotypic pairs are entirely non-obvious, such as the fact that retinal cancer in humans and ectopic vulvae in the nematode can both be caused by disruption of the human retinoblastoma 1 (*RB1*) gene or its nematode ortholog [6]. Because such non-obvious relations can occur with surprising frequency, a method that can rapidly and reliably link human genes to non-obviously related phenotypes in model systems offers the prospect of radically accelerating the rate at which we can explore disease landscapes in both humans and model organisms.

McGary *et al.* [6] provide more than a glimpse at the possibilities. They begin by defining 'phenologs' as cross-species mutant phenotypes that share a significant number of orthologous genes. The statistical significance of phenologs, whose emergence results from disruption of orthologous genes, can be estimated as the probability that the observed number of orthologues common to the two phenotypes would be found by chance and correcting for multiple hypotheses. In this way, using approximately 300 human diseases and over 6,000 phenotypes in model organisms, including mouse, worm, yeast and *Arabidopsis*, they identify 4,390 significant phenologs. As one of the positive controls, the authors note that the 3,755 mouse-human phenologs identified contain many of the known disease models - including cataracts, deafness and retinal disease - all at *P*-values well below 10-8. Given that phenologs map gene-phenotype associations across the phyla, an association known in one species can be used to find non-established relations in another. Cross-validations presented by McGary *et al.* [6] show that phenologs can predict genes associated with about a third to a half of tested human diseases.

The work [6] offers tantalizing evidence for several counterintuitive mammalian disease models, including reduced growth rate of yeast deletion strains in medium enriched with the cholesterol-lowering drug lovastatin as a model for abnormal angiogenesis in mice, and negative gravitropism defects in *Arabidopsis* as a model for human Waardenburg syndrome (which causes deafness with defects in neural-crest-derived tissues). Furthermore, using the yeast model, they demonstrate that *SOX13* (encoding a transcription factor related to the sex-determining gene *SRY*) is a new gene that regulates angiogenesis, and using the *Arabidopsis* model they show that *SEC23IP* (encoding a protein that interacts with SEC23, a component of the COPII complex that controls endoplasmic reticulum-to-Golgi trafficking) is probably a new Waardenburg syndrome gene. Notwithstanding the fact that many functionally coherent gene modules are conserved across different species [10], such demonstrations - especially the identification of phenologs that predate kingdom divergence - seem to mark one of those uncommon occasions in science in which intuition built on years of experience fails completely.

Additional connections

As staggering as these results [6] are, it seems possible that many phenologs have been missed, because the method is confined to connections based only on phenotype pairs sharing known orthologous genes. Gene-phenotype association data may, however, be far from complete; if so, many orthologous phenotypes will be missed. One possible way to increase the discovery rate would be to consider the functional relatedness for the genes associated with each phenotype. As described earlier, gene-gene functional relatedness has been successfully used to identify phenotypic connections within the same species [2-4], and it is possible that the same principle can be applied to identify phenotype connections between different species.

A deeper and far more important connection relates to the impact of McGary *et al.* [6] on efforts to develop a more complete picture of the topology of the human phenome. It is evident that the way phenomics looks today, the amount of information and the incredible interconnectivities, could not have been imagined even 3 years ago. It seems likely that the methods developed and demonstrated by McGary *et al.* [6], and their inevitable extensions, will add unprecedented knowledge and quantitative detail to the interrelated landscapes of mutant phenotypes for humans and other species, and this will offer many more surprising correlations between and among human diseases. As our picture of the human disease landscape continues to take shape, perhaps the one thing that we should not be surprised about will be the need to fundamentally rethink the current disease classification and its associated diagnostic paradigm.

Author details
[1]Translational Sciences Department, Novartis Institutes for BioMedical Research, Cambridge, MA 02139, USA. [2]Departments of Biomedical Engineering and Physics, Program in Bioinformatics and Systems Biology, Boston University, Boston, MA 02215, USA.

References
1. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network.** *Proc Natl Acad Sci USA* 2007, **104:**8685-8690.
2. Lee DS, Park J, Kay KA, Christakis NA, Oltvai ZN, Barabasi AL: **The implications of human metabolic network topology for disease comorbidity.** *Proc Natl Acad Sci USA* 2008, **105:**9880-9885.
3. Linghu B, Snitkin ES, Hu Z, Xia Y, Delisi C: **Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network.** *Genome Biol* 2009, **10:**R91.
4. Park J, Lee DS, Christakis NA, Barabasi AL: **The impact of cellular networks on disease comorbidity.** *Mol Syst Biol* 2009, **5:**262.
5. Rzhetsky A, Wajngurt D, Park N, Zheng T: **Probing genetic overlap among complex human phenotypes.** *Proc Natl Acad Sci USA* 2007, **104:**11694-11699.
6. McGary KL, Park TJ, Woods JO, Cha HJ, Wallingford JB, Marcotte EM: **Systematic discovery of nonobvious human disease models through orthologous phenotypes.** *Proc Natl Acad Sci USA* 2010, **107:**6544-6549.
7. Amberger J, Bocchini CA, Scott AF, Hamosh A: **McKusick's Online Mendelian Inheritance in Man (OMIM).** *Nucleic Acids Res* 2009, **37:**D793-D796.
8. Oti M, Brunner HG: **The modular nature of genetic diseases.** *Clin Genet* 2007, **71:**1-11.
9. Campas C: **Drug repositioning summit: finding new routes to success.** *Drug News Perspect* 2009, **22:**126-128.
10. Kelley BP, Sharan R, Karp RM, Sittler T, Root DE, Stockwell BR, Ideker T: **Conserved pathways within bacteria and yeast as revealed by global protein network alignment.** *Proc Natl Acad Sci USA* 2003, **100:**11394-11399.