Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Current Opinion in Genetics & Development

Edgotype: a fundamental link between genotype and phenotype

Nidhi Sahni[1,2], Song Yi[1,2], Quan Zhong[1,2], Noor Jailkhani[1,2], Benoit Charloteaux[1,2], Michael E Cusick[1,2] and Marc Vidal[1,2]

Classical 'one-gene/one-disease' models cannot fully reconcile with the increasingly appreciated prevalence of complicated genotype-to-phenotype associations in human disease. Genes and gene products function not in isolation but as components of intricate networks of macromolecules (DNA, RNA, or proteins) and metabolites linked through biochemical or physical interactions, represented in 'interactome' network models as 'nodes' and 'edges', respectively. Accordingly, mechanistic understanding of human disease will require understanding of how disease-causing mutations affect systems or interactome properties. The study of 'edgetics' uncovers specific loss or gain of interactions (edges) to interpret genotype-to-phenotype relationships. We review how distinct genetic variants, the genotype, lead to distinct phenotypic outcomes, the phenotype, through edgetic perturbations in interactome networks altogether representing the 'edgotype'.

Addresses
[1] Center for Cancer Systems Biology (CCSB) and Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2] Department of Genetics, Harvard Medical School, Boston, MA 02115, USA

Corresponding author: Vidal, Marc (marc_vidal@dfci.harvard.edu)

Current Opinion in Genetics & Development 2013, **23**:649–657

This review comes from a themed issue on **Genetics of system biology**

Edited by **Shamil Sunyaev** and **Fritz Roth**

For a complete overview see the Issue and the Editorial

Available online 26th November 2013

0959-437X/$ – see front matter, © 2013 Elsevier Ltd. All rights reserved.

http://dx.doi.org/10.1016/j.gde.2013.11.002

Introduction

Genotype-to-phenotype relationships are far more complicated than the 'one-gene/one-enzyme/one-function' paradigm introduced decades ago [1]. Different variants of the same gene may cause different functional defects of the corresponding gene product and as a consequence different diseases (allelic heterogeneity), while the same disease can be caused by mutations in different genes (genetic heterogeneity) [2,3]. The confounding phenomena of incomplete penetrance and variable expressivity are encountered far more often than expected [4]. With increasing number of genomic variants potentially associated with disease being identified by genome-wide association studies [5] and next-generation sequencing [6], it is more imperative than ever to work out underlying principles of genotype-to-phenotype relationships [7].

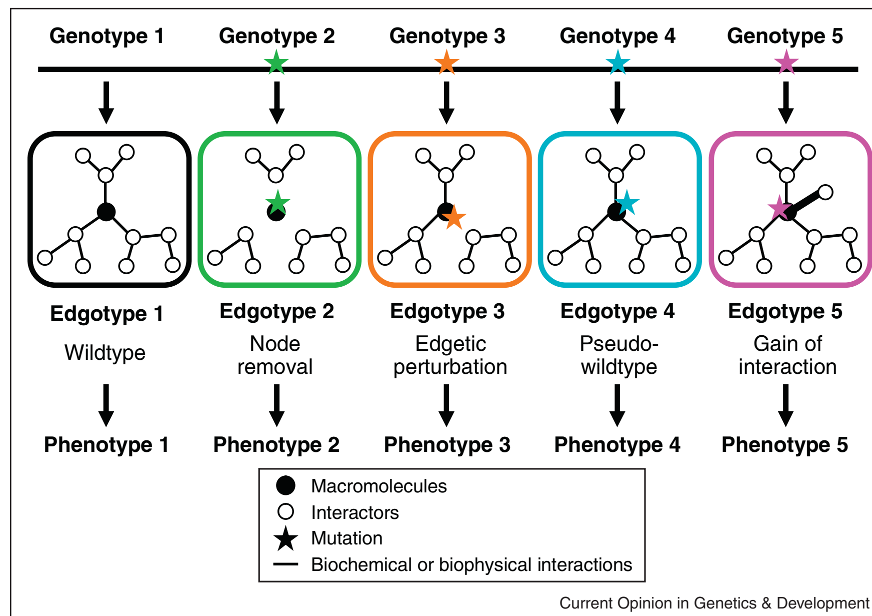
Genes and gene products do not act in isolation but rather interact with each other within intricate and dynamic 'interactome' networks, depicted as nodes and edges representing individual molecules and their mutual interactions, respectively [7,8]. Interactome networks provide an informative platform to investigate functional properties of cellular systems [9]. Comprehensive mapping of protein–protein interaction (PPI) networks has been informative for several human diseases, ataxia [10–12], autism [13], Huntington disease [14,15] and breast cancer [16–18] included.

In network representations, a genotypic variation can be modeled either as knockout or knockdown of gene function, leading to removal of a node and all of its edges, or alternatively, as interaction-specific 'edgetic' perturbation, leading to the removal or addition of specific interactions while other edges remain unperturbed [19••] (Figure 1). These specific perturbations of interactome networks arising from genetic variants can give rise to distinct phenotypic outcomes [20•]. Edgetic network perturbation models, which emphasize the disruption of specific edges, complement classic gene-centric paradigms [21], which assess the effects of deleting or overexpressing genes, but with few exceptions [22] neglect the influence of genetic variation [23••]. Edgetics can help make sense of confounding genetic heterogeneity, and puts forth direct mechanistic connections from genotype to phenotype [19••,24••,25]. Edgetic modeling is not limited to protein-protein interactions but can be applied to any type of biomolecular interaction. Edgetic perturbation models are emerging as a powerful strategy for interpretation of genotype-to-phenotype relationships.

High-quality interactome networks

Before knowing which interactions are perturbed by particular mutations in a particular gene it is necessary to know the interactions of the wild-type non-mutated protein. Therefore, building comprehensive 'reference' interactome networks is clearly the first step for studying edgetic perturbations. High-throughput experimental approaches generate systematic and well-controlled data. They either test all binary combinations of possible protein pairs to determine which ones interact directly [26], or identify protein membership of protein complexes isolated

Figure 1



Genetic variant-induced perturbations in network properties give rise to altered phenotypes, such as disease. Distinct genetic variants of the same gene can exhibit different interaction profiles: loss of all interactions (node removal), loss of some interactions (edgetic), no loss of interactions (pseudo-wildtype), or gain-of-interaction. Nodes represent macromolecules, and edges represent biochemical or biophysical interactions between them. The stars denote a disease-associated variant or mutation. The profile of edgetic perturbations defines the edgotype, providing the explanatory connections between genotype and phenotype.

from cells, that is, co-complex associations [27]. Mapping of the binary interactome is carried out primarily by enhanced variants of yeast two-hybrid methodologies followed by orthogonal assays for validation [28,29]. Mapping of the co-complex interactome is carried out primarily by affinity purification followed by mass spectrometry [30].

Previous high-quality binary protein–protein interaction mapping efforts have identified an appreciable fraction of interactome networks for human [31,32] as well as for model organisms [33–37]. A new generation of binary interactome mapping is underway with enhanced network completeness and resolution. That enhancement comes partly from forceful implementation of an empirical framework that quantitatively assesses the quality, coverage and size of interactome maps [38–40]. Empirical determination of quality has most readily been applied to binary interactome mapping methods [36,41] but adaptations to measure co-complex interactome quality have been pursued [42].

Network perturbations underlie genotype-to-phenotype relationships

Mutations can alter interactome networks due to either node removal or edgetic perturbations [19••,43•]. Truncating mutations, including out-of-frame indels and nonsense mutations, are most likely node removal perturbations, although small in-frame indels could potentially be edgetic [19••]. In-frame missense mutations could be either edgetic, disrupting a protein interaction, or they could destabilize the protein, in effect being node removal (Figure 1). The proportion in each category is unclear as yet. Some reports argue that nearly all missense mutations would be destabilizing [21,44,45]. Other reports estimate that a considerable portion of known Mendelian missense mutations could be edgetic, especially single amino acid changes in protein-binding interfaces [19••,20•,24••,46]. An analysis of protein–protein interaction defects of 29 disease-causing missense alleles identified both node removal and edgetic perturbations in roughly equal proportion [19••].

In accounting how distinct types of network perturbation, node removal *vs.* edgetic, can result in distinct phenotypic outcomes, edgetic models shed new light on perplexing genotype–phenotype correlations in human disease. For example, von Hippel-Lindau (VHL) syndrome is an autosomal dominant disorder that manifests as diverse tumors. Mutations associated with Type 1 VHL syndrome (hemangioblastomas and renal cell carcinoma) are typically truncating mutations. In contrast, mutations associated with Type 2 VHL syndrome (pheochromocytoma) are typically missense mutations. Missense mutations on the surface of the VHL protein present a

much higher risk for pheochromocytoma than those in the core of the protein [47], in agreement with a model that these surface mutations are disrupting specific edges (PPIs) which are then responsible for the development of pheochromocytoma [48].

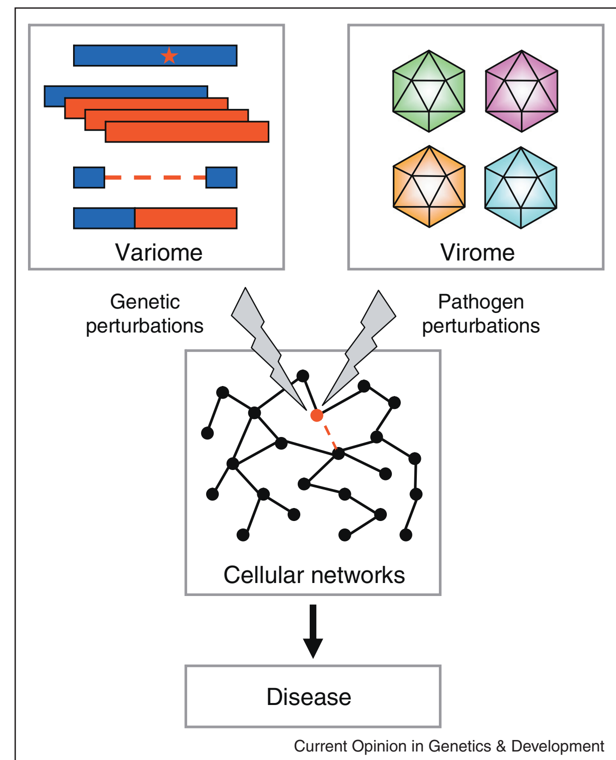
Edgetic perturbations verified so far in human disease represent loss-of-interaction, but theoretically gain-of-interaction alleles (Figure 1), akin to gain-of-function alleles in classical genetics [49], are also possible. Consistent with an edgetic gain-of-interaction is the interaction of nardilysin (NRD1) with the p53 tumor suppressor. Only the R273H mutant of p53 associates with NRD1, not wild-type p53 or the R280K or R273C mutants of p53 [50]. The p53 R273H-NRD1 interaction is critical for promotion of cellular invasion. Systematic searches for edgetic gain-of-interactions, although technically more challenging, are feasible in the near future.

Virus–host protein interactions as a surrogate for edgetic perturbations

Similar to human genetic variation, pathogens induce disease states by perturbing host cellular networks [51,52]. This common observation has inspired examinations of the global landscape of host perturbations by proteins from diverse viruses [42,53••,54]. These efforts have uncovered novel disease gene associations. Systematic interaction mapping of human papillomavirus (HPV) proteins to host proteins provided evidence that supported genome-sequencing observations implicating mutations in *MAML1*, a Notch pathway component, in tumorigenesis [53••,55]. *MAML1* has since been confirmed as a gene mutated in cancer [56,57].

Systematic viral-host perturbation mapping functionally characterized FAM111A protein as a previously unidentified host range restriction factor specifically targeted by Simian Virus 40 (SV40) large T antigen [58]. In a validation of the hypothesis that viral-protein interactions can act as surrogates to identify proteins mutated in human disease (Figure 2) [53••,59], a set of *de novo* arising heterozygous point mutations in FAM111A lie behind two phenotypically related developmental disorders of previously unknown molecular etiology, Kenny-Caffey syndrome (KCS) and osteocraniostenosis (OCS) [60•]. These mutations all map to the same minimal region of FAM111A that is required for large T antigen binding [58], predicting that pathogenesis is due to these mutations disrupting the binding of FAM111A to as yet undiscovered cellular proteins. Experiments to identify these interactors, and then to uncover how the newly identified FAM111A mutations perturb these interactions, are readily envisioned.

Figure 2



Human genetic variations (the variome) and pathogenic viral proteins (the virome) similarly influence local and global properties of networks to induce disease states.

Edgetic approaches to study genetic variants

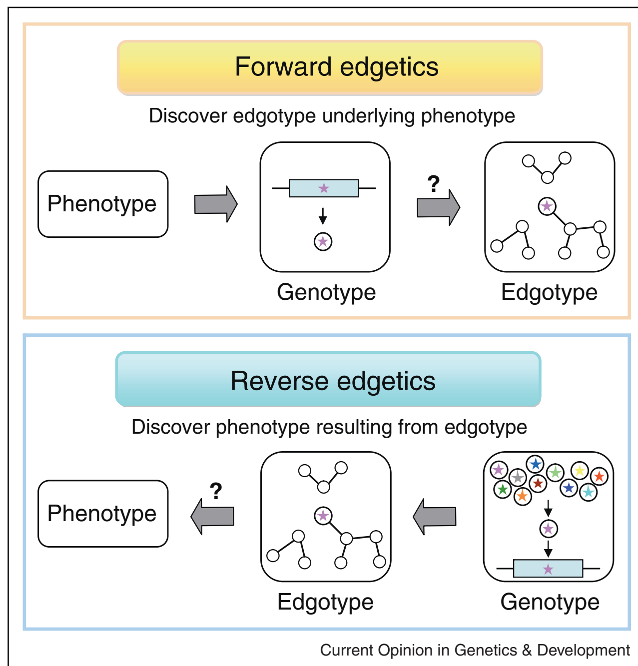
Two complementary strategies can identify edgetic perturbations resulting from particular genetic variants: forward edgetics and reverse edgetics (Figure 3) [43•], analogous to the longstanding distinction between classical forward genetics (generate a phenotype then identify the responsible genetic variants) and modern reverse genetics (generate specific mutations in proteins then identify the resulting phenotype) [61].

Forward edgetics

Starting from known mutations associated with particular phenotypes, primarily human disease phenotypes, forward edgetics investigates interaction profile changes in proteins constructed to contain these known mutations (Figure 3). Such mutations can be readily introduced by high-throughput technologies of recombinational DNA cloning [62] and site-directed mutagenesis [63,64].

Edgetic analysis may seem to be in its infancy, the term reaching acceptance only recently [60•,65•,66], but the principle of interaction profiling for phenotypic resolution has actually been around for considerable time [67•,68,69,70,71]. These early studies introduced a series of point mutations into a single protein then tested by yeast two-hybrid (Y2H) for loss or maintenance of binary interactions against known interactors. An appealing

Figure 3



Forward and reverse edgetics to functionally characterize genomic variants. Forward edgetics studies the underlying edgotype for a given phenotype (disease), introducing known disease-causing mutations to study mutation-mediated loss or maintenance of known protein interactions and to relate the corresponding edgotype to a disease phenotype. Reverse edgetics introduces novel mutations into proteins of interest, finding those mutations that cause loss or maintenance of interactions against known interactors. The obtained mutations can then be introduced *in vivo* to characterize the resulting phenotype.

example is the detailed mapping of the interaction between the mammalian retromer complex components Vps26 and Vps35. Based on structural modeling 38 distinct mutations in Vps26 were tested for functional interaction with Vps35 by yeast two-hybrid [72], the results meticulously delineating the binding site between the two proteins.

The earliest study that applied edgetic profiling systematically, that is, testing more than one interactor at-a-time, characterized 35 actin mutants constructed by charged-to-alanine scanning mutagenesis for binary protein interaction defects with five known actin interactors [67•]. Three classes of perturbation were identified: lethal—removed binding of all tested interactors (in current network terms 'node removal'); differential—removed binding of some but not all interactors (by current terms edgetic); and 'unchanged'—did not alter binding of any interactors (by current terms pseudo-wildtype) (Figure 1).

Forward edgetics has helped elucidate disease mechanisms for cancer and other diseases [73,74•]. The *ATXN1* gene is frequently mutated in ataxia. *ATXN1* mutants containing an increased polyglutamine repeat produce a protein that preferentially interacts with RBM17, which promotes neurotoxicity, whereas there is reduced interaction of mutant ATXN1 with the protein Capicua and reduced neuroprotection [11]. Another telling example of how edgetic modeling can clarify disease etiology is the interaction between mutant Mdm2 and ribosomal proteins (RPs) L5 and L11 in cancer. In response to ribosomal stress, these RPs interact with and inhibit Mdm2, resulting in the stabilization and activation of p53, and inhibition of c-Myc induced lymphomagenesis. The cancer associated C305F missense mutation in the acidic zinc finger domain of Mdm2 results in the loss of Mdm2 binding to L5 and L11. This edgetic perturbation causes loss of the ribosomal stress response and an increase in c-Myc induced tumorigenesis [75]. A last example concerns the Fbxw7 protein, a component of the SCF ubiquitin ligase complex that as part of the SCF complex binds to and degrades the c-Myc transcription factor. Mutations in *FBXW7* have been found in multiple cancers including T-cell acute lymphoblastic leukemia, with the R465C missense mutation being the most common. Fbxw7-R465C protein is unable to bind and ubiquitylate c-Myc, resulting in increased c-Myc protein stability, acting to augment the Leukemia Initiating Cell (LIC) population [76]. The R465C mutation specifically affects cancer cells and not normal hematopoietic stem cells, which rarely express c-Myc [76].

Putative edgetic alleles of the same gene, but associated with different diseases, tend to be located in distinct interaction domains and thus likely perturb distinct interactions [24••]. A compelling example of distinct domain to distinct disease relationships comes from TP63 [77] in which two clinically distinct developmental disorders, ectrodactyly ectodermal dysplasia (EEC) and ankyloblepharon ectodermal dysplasia (AEC), are caused by mutations in two separate domains, one predicted to bind DNA and the other to mediate protein–protein interactions [19••].

Technology has now advanced to the point where thousands of alleles and hundreds of interactors can readily be tested for allelic perturbations [19••]. Quality of the resulting edgetic interactomic data would be assessed by implementation of the empirical framework already applied to reference interactomic data. Large-scale experimental edgetic profiling, assessing variant-associated interaction changes at large-scale with respect to the corresponding wild-type counterpart, is needed to thoroughly test edgetic models. For such efforts wild-type protein interaction partners can be retrieved from available and ongoing proteome-wide interactome mapping projects [26].

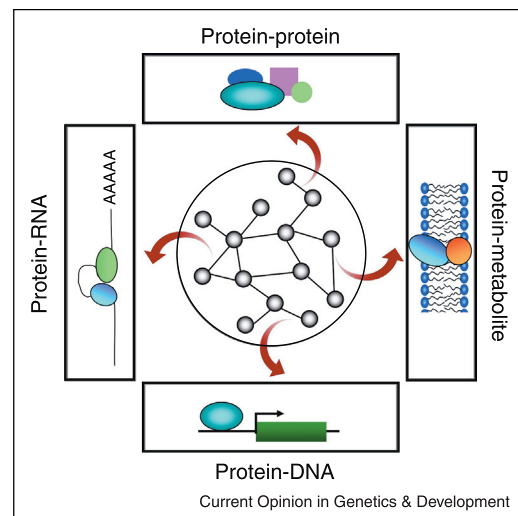
Reverse edgetics

The reverse edgetics approach systematically searches for specific alleles encoding mutant proteins with desired

interaction defects (Figure 3). A library of alleles is first generated by random mutagenesis, and interaction-defective alleles are then isolated from the library using reverse yeast two-hybrid selections [78,79]. Interaction profiling of the mutant proteins then identifies edgetic alleles, which are further studied functionally to investigate the phenotypic consequences [43•,80••]. The earliest application of reverse edgetics profiling targeted the E2F–DP1 interaction, a key interaction in regulation of cell proliferation, and obtained single amino acid mutations identifying a putative helix in a region conserved among E2F family members as the critical determinant for the interaction [79].

A seminal investigation of the *C. elegans* anti-apoptotic protein CED-9 found that systematic isolation of edgetic alleles is both practical and a fruitful strategy for investigation of gene function [80••]. CED-9 interactors were mapped by Y2H and co-affinity purification, and interaction-defective alleles were selected based on their edgetic perturbation profiles. Structural analyses confirmed that mutations in edgetic alleles specifically affect protein binding sites. *In vivo* characterization of CED-9 edgetic alleles demonstrated that distinct alleles were associated with distinct phenotypes.

Figure 4



Edgetics can apply to all types of biomolecular interaction networks, including but not limited to protein–protein, protein–DNA, protein–RNA, and protein–metabolite interactions. Perturbations of these distinct types of interactions have been shown to be a factor in human disease.

Mapping domain–domain interactions

Analysis of resolved structures of many protein complexes shows that physical interactions can be grouped into two broad classes: domain–domain interactions and domain–motif interactions [81]. Most proteins are composed of multiple domains that bind to specific partners [82,83]. The structures of many interaction domains, including SH2, SH3, PTB, and PDZ domains, have been solved [83,84].

Fragmenting an interacting protein into domain-sized pieces and then determining which region of the protein is responsible for the interaction has been practiced almost since yeast two-hybrid technologies for binary interaction mapping were first implemented [85–87]. Fragmentation for determination of the 'minimal region of interaction' is commonplace because of the ease and convenience of the DNA-based technology behind Y2H. Such fragmentation can now be done systematically at large scale for hundreds if not thousands of proteins.

A first systematic experimental mapping of protein interaction domains used a high-throughput, fragment-based Y2H approach to identify interaction domains for ~200 *C. elegans* early embryogenesis proteins [88••]. This approach greatly increased the sensitivity of interaction detection while maintaining high specificity, leading to a fuller interactome network for embryogenesis..

Edgetics beyond protein–protein interactions

In cellular network models the nodes represent not only proteins, as discussed so far, but also DNA or RNA entities or metabolites, while the edges are physical or biochemical interactions between them (Figure 4). The edgetics strategy is accordingly not limited to protein–protein interactions. Edgetic perturbations of any type of biomolecular interaction can have relevance for improved understanding of disease etiologies.

Protein–DNA interactions

For an example of edgetic perturbation of protein–DNA interactions (PDI) possibly causative of human disease, take the S128F mutation in the T-box transcription factor Tbx19, which leads to adrenocorticotropin hormone deficiency due to virtually complete loss of DNA binding affinity [89]. Another example is the I156V mutation of the transcription factor H-Twist, leading to Baller-Gerold syndrome. This mutation resides in the highly conserved Helix II domain of this gene, and disrupts the interaction interface with DNA targets [90]. Missense mutation H275R in *KLF3*, encoding a Krüppel family zinc finger transcription factor, is associated with various cardiovascular defects. This point mutation, by abolishing a highly conserved His residue, impairs Klf3 binding to its canonical DNA binding sequence [91].

Protein–RNA interactions

Allelic perturbations of protein–RNA interactions may also play a role in disease etiology. The *PRKRA* gene, which encodes the protein PACT/RAX, is an important regulator for ear and craniofacial development. PACT mediates the assembly of the RNA Induced Silencing Complex (RISC), and is required for the biogenesis of

miRNAs which are involved in RNA silencing. A mutant protein carrying the missense mutation S130P is unable to bind dsRNA, although otherwise seems normal in tested biochemical functions. This mutant shows defective ear and craniofacial development [92,93].The expression of a unique type of selenoproteins requires the translational recoding of the UGA stop codon to selenocysteine. This process is regulated by a selenocysteine insertion sequence (SECIS) and SECIS-binding protein 2 (SBP2). An R540Q amino acid change in SBP2 alters its RNA binding activity, resulting in abnormal hormone signaling [94].

Protein–metabolite interactions

Defects in protein–lipid interactions might also be edgetic. The E17K missense mutation in the gene *AKT1* is associated with various cancers (breast, colorectal and ovarian). This mutation is localized in the lipid-binding pocket of AKT1, and therefore alters the electrostatic interactions and forms new hydrogen bonds with a phosphoinositide ligand. As a consequence, the mutant protein disrupts lipid binding, stimulating downstream signaling to induce cancer transformation [95]. Another interesting example is the association of nephrin with signaling microdomains, also known as lipid rafts. Lipid rafts spatially organize glomerular structures under physiological conditions [96]. The two mutations C265R and V822M in nephrin lead to a dysfunctional complex due to defective cell surface targeting and ineffective association with lipid rafts, predisposing to a relapsing phenotype [97].

Conclusions

While the usefulness of interactome maps to identify new candidate disease genes and modifier genes has become evident [12,16,31,32], the task of investigating the impact of diverse genomic variants on interactome networks is just underway. Edgetic profiling has already proven insightful in deciphering molecular etiologies of several Mendelian diseases. Deeper edgetic studies on disease mechanisms can now be done at large-scale and at high-throughput [43•,98].

While applied so far mostly to Mendelian disorders, as illustrated by the examples provided, there is no reason why the edgetics strategy could not be extended to complex trait disorders [23••]. Still, it has been frequently noted that even 'simple' Mendelian traits are highly 'complex' [99]. The line between Mendelian disorders and complex disorders is a gray one indeed. For either, the edgetics paradigm should guide prioritizations of disease-causing variants emerging from GWAS and next-generation sequencing projects, and provide explanations of mutation-specific disease outcome. Functional characterization of the effects of genetic variants on edgotypes will undoubtedly facilitate understanding of how network perturbations relate genotypes to phenotypes.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

• of special interest

•• of outstanding interest

1. Beadle GW, Tatum EL: **Genetic control of biochemical reactions in Neurospora**. *Proc Natl Acad Sci USA* 1941, **27**:499-506.

2. Fearon ER: **Human cancer syndromes: clues to the origin and nature of cancer**. *Science* 1997, **278**:1043-1050.

3. Weatherall DJ: **Phenotype–genotype relationships in monogenic disease: lessons from the thalassaemias**. *Nat Rev Genet* 2001, **2**:245-255.

4. Perlis RH, Smoller JW, Mysore J, Sun M, Gillis T, Purcell S, Rietschel M, Nothen MM, Witt S, Maier W *et al.*: **Prevalence of incompletely penetrant Huntington's disease alleles among individuals with major depressive disorder**. *Am J Psychiatry* 2010, **167**:574-579.

5. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery**. *Am J Hum Genet* 2012, **90**:7-24.

6. Majewski J, Rosenblatt DS: **Exome and whole-genome sequencing for gene discovery: the future is now!** *Hum Mutat* 2012, **33**:591-592.

7. Vidal M, Cusick ME, Barabási AL: **Interactome networks and human disease**. *Cell* 2011, **144**:986-998.

8. Barabási AL, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nat Rev Genet* 2004, **5**:101-113.

9. Ideker T, Sharan R: **Protein networks in disease**. *Genome Res* 2008, **18**:644-652.

10. Kahle JJ, Gulbahce N, Shaw CA, Lim J, Hill DE, Barabási AL, Zoghbi HY: **Comparison of an expanded ataxia interactome with patient medical records reveals a relationship between macular degeneration and ataxia**. *Hum Mol Genet* 2011, **20**:510-527.

11. Lim J, Crespo-Barreto J, Jafar-Nejad P, Bowman AB, Richman R, Hill DE, Orr HT, Zoghbi HY: **Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1**. *Nature* 2008, **452**:713-718.

12. Lim J, Hao T, Shaw C, Patel AJ, Szabo G, Rual JF, Fisk CJ, Li N, Smolyar A, Hill DE *et al.*: **A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration**. *Cell* 2006, **125**:801-814.

13. Sakai Y, Shaw CA, Dawson BC, Dugas DV, Al-Mohtaseb Z, Hill DE, Zoghbi HY: **Protein interactome reveals converging molecular pathways among autism disorders**. *Sci Transl Med* 2011, **3**:ra49.

14. Kaltenbach LS, Romero E, Becklin RR, Chettier R, Bell R, Phansalkar A, Strand A, Torcassi C, Savage J, Hurlburt A *et al.*: **Huntingtin interacting proteins are genetic modifiers of neurodegeneration**. *PLoS Genet* 2007, **3**:e82.

15. Goehler H, Lalowski M, Stelzl U, Waelter S, Stroedicke M, Worm U, Droege A, Lindenberg KS, Knoblich M, Haenig C *et al.*: **A protein interaction network links GIT1, an enhancer of Huntingtin aggregation, to Huntington's disease**. *Mol Cell* 2004, **15**:853-865.

16. Pujana MA, Han JD, Starita LM, Stevens KN, Tewari M, Ahn JS, Rennert G, Moreno V, Kirchhoff T, Gold B *et al.*: **Network modeling links breast cancer susceptibility and centrosome dysfunction**. *Nat Genet* 2007, **39**:1338-1349.

17. Maxwell CA, Benitez J, Gómez-Baldó L, Osorio A, Bonifaci N, Fernández-Ramires R, Costes SV, Guinó E, Chen H, Evans GJ

et al.: **Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer**. *PLoS Biol* 2011, **9**:e1001199.

18. Taylor IW, Linding R, Warde-Farley D, Liu Y, Pesquita C, Faria D, Bull S, Pawson T, Morris Q, Wrana JL: **Dynamic modularity in protein interaction networks predicts breast cancer outcome**. *Nat Biotechnol* 2009, **27**:199-204.

19. •• Zhong Q, Simonis N, Li QR, Charloteaux B, Heuze F, Klitgord N, Tam S, Yu H, Venkatesan K, Mou D *et al.*: **Edgetic perturbation models of human inherited disorders**. *Mol Syst Biol* 2009, **5**:321.
This paper explicitly defines the edgetic concept for delineation of genotype-to-phenotype relationships in inherited human disease.

20. • Schuster-Bockler B, Bateman A: **Protein interactions in human genetic diseases**. *Genome Biol* 2008, **9**:R9.
This paper computationally predicts human diseases that arise from mutational perturbations of protein–protein interactions.

21. Botstein D, Risch N: **Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease**. *Nat Genet* 2003, **33**:228-237.

22. Raj A, Rifkin SA, Andersen E, van Oudenaarden: **Variability in gene expression underlies incomplete penetrance**. *Nature* 2010, **463**:913-918.

23. •• Chakravarti A, Clark AG, Mootha VK: **Distilling pathophysiology from complex disease genetics**. *Cell* 2013, **155**:21-26.
This perspective argues forcefully that increased insight into the pathophysiology of complex traits requires a turn towards working out functional impacts of genetic variation in candidate disease genes.

24. •• Wang X, Wei X, Thijssen B, Das J, Lipkin SM, Yu H: **Three-dimensional reconstruction of protein networks provides insight into human genetic disease**. *Nat Biotechnol* 2012, **30**:159-164.
This paper uses protein structural information to evaluate what proportion of human diseases may potentially arise from mutations that disrupt protein–protein interactions.

25. Csermely P, Korcsmaros T, Kiss HJ, London G, Nussinov R: **Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review**. *Pharmacol Ther* 2013, **138**:333-408.

26. Braun P: **Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays**. *Proteomics* 2012, **12**:1499-1518.

27. Gingras AC, Raught B: **Beyond hairballs: the use of quantitative mass spectrometry data to understand protein–protein interactions**. *FEBS Lett* 2012, **586**:2723-2731.

28. Chien CT, Bartel PL, Sternglanz R, Fields S: **The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest**. *Proc Natl Acad Sci USA* 1991, **88**:9578-9582.

29. Dreze M, Monachello D, Lurin C, Cusick ME, Hill DE, Vidal M, Braun P: **High-quality binary interactome mapping**. *Methods Enzymol* 2010, **470**:281-315.

30. Walzthoeni T, Leitner A, Stengel F, Aebersold R: **Mass spectrometry supported determination of protein complex structure**. *Curr Opin Struct Biol* 2013, **23**:252-260.

31. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, Berriz GF, Gibbons FD, Dreze M, Ayivi-Guedehoussou N *et al.*: **Towards a proteome-scale map of the human protein–protein interaction network**. *Nature* 2005, **437**:1173-1178.

32. Stelzl U, Worm U, Lalowski M, Haenig C, Brembeck FH, Goehler H, Stroedicke M, Zenkner M, Schoenherr A, Koeppen S *et al.*: **A human protein-protein interaction network: a resource for annotating the proteome**. *Cell* 2005, **122**:957-968.

33. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, Li Y, Hao YL, Ooi CE, Godwin B, Vitols E *et al.*: **A protein interaction map of *Drosophila melanogaster***. *Science* 2003, **302**:1727-1736.

34. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, Boxem M, Vidalain PO, Han JD, Chesneau A, Hao T *et al.*: **A map of the interactome network of the metazoan *C. elegans***. *Science* 2004, **303**:540-543.

35. Yu H, Braun P, Yildirim MA, Lemmons I, Venkatesan K, Sahalie J, Hirozane-Kishikawa T, Gebreab F, Li N, Simonis N *et al.*: **High quality binary protein interaction map of the yeast interactome network**. *Science* 2008, **322**:104-110.

36. *Arabidopsis* Interactome Mapping Consortium: **Evidence for network evolution in an *Arabidopsis* interactome map**. *Science* 2011, **333**:601-607.

37. Das J, Vo TV, Wei X, Mellor JC, Tong V, Degatano AG, Wang X, Wang L, Cordero NA, Kruer-Zerhusen N *et al.*: **Cross-species protein interactome mapping reveals species-specific wiring of stress response pathways**. *Sci Signal* 2013, **6**:ra38.

38. Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI *et al.*: **An empirical framework for binary interactome mapping**. *Nat Methods* 2009, **6**:83-90.

39. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, De Smet A-S *et al.*: **An experimentally derived confidence score for binary protein–protein interactions**. *Nat Methods* 2009, **6**:91-97.

40. Cusick ME, Yu H, Smolyar A, Venkatesan K, Carvunis AR, Simonis N, Rual JF, Borick H, Braun P, Dreze M *et al.*: **Literature-curated protein interaction datasets**. *Nat Methods* 2009, **6**:39-46.

41. Simonis N, Rual JF, Carvunis AR, Tasan M, Lemmons I, Hirozane-Kishikawa T, Hao T, Sahalie JM, Venkatesan K, Gebreab F *et al.*: **Empirically-controlled mapping of the *Caenorhabditis elegans* protein–protein interactome network**. *Nat Methods* 2009, **6**:47-54.

42. Varjosalo M, Sacco R, Stukalov A, van Drogen A, Planyavsky M, Hauri S, Aebersold R, Bennett KL, Colinge J, Gstaiger M *et al.*: **Interlaboratory reproducibility of large-scale human protein-complex analysis by standardized AP-MS**. *Nat Methods* 2013, **10**:307-314.

43. • Charloteaux B, Zhong Q, Dreze M, Cusick ME, Hill DE, Vidal M: **Protein–protein interactions and networks: forward and reverse edgetics**. *Methods Mol Biol* 2011, **759**:197-213.
This contribution was the first to forthrightly distinguish forward edgetics from reverse edgetics.

44. Ferrer-Costa C, Orozco M, de la Cruz X: **Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties**. *J Mol Biol* 2002, **315**:771-786.

45. Yue P, Li Z, Moult J: **Loss of protein structure stability as a major causative factor in monogenic disease**. *J Mol Biol* 2005, **353**:459-473.

46. Guo Y, Wei X, Das J, Grimson A, Lipkin SM, Clark AG, Yu H: **Dissecting disease inheritance modes in a three-dimensional protein network challenges the "guilt-by-association" principle**. *Am J Hum Genet* 2013.

47. Ong KR, Woodward ER, Killick P, Lim C, Macdonald F, Maher ER: **Genotype–phenotype correlations in von Hippel-Lindau disease**. *Hum Mutat* 2007, **28**:143-149.

48. Nordstrom-O'Brien M, van der Luijt RB, van Rooijen E, van den Ouweland AM, Majoor-Krakauer DF, Lolkema MP, van Brussel A, Voest EE, Giles RH: **Genetic analysis of von Hippel-Lindau disease**. *Hum Mutat* 2010, **31**:521-537.

49. Strachan TR, Reid A: *Human Molecular Genetics*. edn 2. New York: John Wiley & Sons Inc; 1999.

50. Coffill CR, Muller PA, Oh HK, Neo SP, Hogue KA, Cheok CF, Vousden KH, Lane DP, Blackstock WP, Gunaratne J: **Mutant p53 interactome identifies nardilysin as a p53R273H-specific binding partner that promotes invasion**. *EMBO Rep* 2012, **13**:638-644.

51. DeCaprio JA: **How the Rb tumor suppressor structure and function was revealed by the study of Adenovirus and SV40**. *Virology* 2009, **384**:274-284.

52. White EA, Howley PM: **Proteomic approaches to the study of papillomavirus–host interactions**. *Virology* 2013, **435**:57-69.

53. •• Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M,

Tavares M *et al.*: **Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins**. *Nature* 2012, **487**:491-495.

This paper presents and interprets a large-scale systematic mapping of viral protein perturbations of host interactome networks, and relates those perturbations to human disease.

54. Jäger S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K *et al.*: **Global landscape of HIV–human protein complexes**. *Nature* 2011, **481**:365-370.

55. White EA, Kramer RE, Tan MJ, Hayes SD, Harper JW, Howley PM: **Comprehensive analysis of host cellular interactions with human papillomavirus E6 proteins identifies new E6 binding partners and reflects viral diversity**. *J Virol* 2012, **86**:13174-13180.

56. Tan MJ, White EA, Sowa ME, Harper JW, Aster JC, Howley PM: **Cutaneous β-human papillomavirus E6 proteins bind Mastermind-like coactivators and repress Notch signaling**. *Proc Natl Acad Sci USA* 2012.

57. Brimer N, Lyons C, Wallberg AE, Vande Pol SB: **Cutaneous papillomavirus E6 oncoproteins associate with MAML1 to repress transactivation and NOTCH signalling**. *Oncogene* 2012, **31**:4639-4640.

58. Fine DA, Rozenblatt-Rosen O, Padi M, Korkhin A, James RL, Adelmant G, Yoon R, Guo L, Berrios C, Zhang Y *et al.*: **Identification of FAM111A as an SV40 host range restriction and adenovirus helper factor**. *PLoS Pathog* 2012, **8**:e9491002.

59. Gulbahce N, Yan H, Dricot A, Padi M, Byrdsong D, Franchi R, Lee D-S, Rozenblatt-Rosen O, Mar JC, Calderwood MA *et al.*: **Viral perturbations of host networks reflect disease etiology**. *PLoS Comput Biol* 2012, **8**:e531 1002.

60. • Unger S, Gorna MW, Le Bechec A, Do Vale-Pereira S, Bedeschi MF, Geiberger S, Grigelioniene G, Horemuzova E, Lalatta F, Lausch E *et al.*: **FAM111A mutations result in hypoparathyroidism and impaired skeletal development**. *Am J Hum Genet* 2013.

In this paper newly discovered mutations in FAM111A causative of two developmental disorders are predicted to be edgetic in their action, that is, disrupt critical interactions of FAM111A with cellular proteins.

61. Lehner B: **Genotype to phenotype: lessons from model organisms for human genetics**. *Nat Rev Genet* 2013, **14**:168-178.

62. Festa F, Steel J, Bian X, Labaer J: **High-throughput cloning and expression library creation for functional proteomics**. *Proteomics* 2013, **13**:1381-1390.

63. Suzuki Y, Kagawa N, Fujino T, Sumiya T, Andoh T, Ishikawa K, Kimura R, Kemmochi K, Ohta T, Tanaka S: **A novel high-throughput (HTP) cloning strategy for site-directed designed chimeragenesis and mutation using the Gateway cloning system**. *Nucleic Acids Res* 2005, **33**:e109.

64. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R, Ishioka C: **Understanding the function–structure and function–mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis**. *Proc Natl Acad Sci USA* 2003, **100**:8424-8430.

65. • Furlong LI: **Human diseases through the lens of network biology**. *Trends Genet* 2012, **29**:150-159.

This contribution comprehensively surveys the impact of network biology for interpreting complex genotype to-phenotype relationships in disease.

66. Pavlicev M, Wagner GP: **A model of developmental evolution: selection, pleiotropy and compensation**. *Trends Ecol Evol* 2012, **27**:316-322.

67. • Amberg DC, Basart E, Botstein D: **Defining protein interactions with yeast actin in vivo**. *Nat Struct Biol* 1995, **2**:28-35.

This paper represents the first systematic edgetic study, methodically investigating the effect of single amino acid changes on protein interactions of actin.

68. Schwartz H, Alvares CP, White MB, Fields S: **Mutation detection by a two-hybrid assay**. *Hum Mol Genet* 1998, **7**:1029-1030.

69. Vojtek AB, Hollenberg SM, Cooper JA: **Mammalian Ras interacts directly with the serine/threonine kinase Raf**. *Cell* 1993, **74**:205-214.

70. Young K, Lin S, Sun L, Lee E, Modi M, Hellings S, Husbands M, Ozenberger B, Franco R: **Identification of a calcium channel modulator using a high throughput yeast two-hybrid screen**. *Nat Biotechnol* 1998, **16**:946-950.

71. Ishioka C, Ballester R, Engelstein M, Vidal M, Kassel J, The I, Bernards A, Gusella JF, Friend SH: **A functional assay for heterozygous mutations in the GTPase activating protein related domain of the neurofibromatosis type 1 gene**. *Oncogene* 1995, **10**:841-847.

72. Shi H, Rojas R, Bonifacino JS, Hurley JH: **The retromer subunit Vps26 has an arrestin fold and binds Vps35 through its C-terminal domain**. *Nat Struct Mol Biol* 2006, **13**:540-548.

73. De Nicolo A, Parisini E, Zhong Q, Dalla Palma M, Stoeckert KA, Domchek SM, Nathanson KL, Caligo MA, Vidal M, Cusick ME *et al.*: **Multimodal assessment of protein functional deficiency supports pathogenicity of BRCA1 p.V1688del**. *Cancer Res* 2009, **69**:7030-7037.

74. • Wang W, Zhong Q, Teng L, Bhatnagar N, Sharma B, Zhang X, Luther W 2nd, Haynes LP, Burgoyne RD, Vidal M *et al.*: **Mutations that disrupt PHOXB interaction with the neuronal calcium sensor HPCAL1 impede cellular differentiation in neuroblastoma**. *Oncogene* 2013 http://dx.doi.org/10.1038/onc.2013.290.

In this paper PHOX2B variants associated with neuroblastoma pathogenesis are shown to lose the ability to bind to the key interacting protein HPCAL1, representing a forthright demonstration of the path from genotype through edgotype to phenotype.

75. Macias E, Jin A, Deisenroth C, Bhat K, Mao H, Lindstrom MS, Zhang Y: **An ARF-independent c-MYC-activated tumor suppression pathway mediated by ribosomal protein–Mdm2 Interaction**. *Cancer Cell* 2010, **18**:231-243.

76. King B, Trimarchi T, Reavie L, Xu L, Mullenders J, Ntziachristos P, Aranda-Orgilles B, Perez-Garcia A, Shi J, Vakoc C *et al.*: **The ubiquitin ligase FBXW7 modulates leukemia-Initiating cell activity by regulating MYC stability**. *Cell* 2013, **153**:1552-1560.

77. van Bokhoven H, Brunner HG: **Splitting p63**. *Am J Hum Genet* 2002, **71**:1-13.

78. Vidal M, Brachmann RK, Fattaey A, Harlow E, Boeke JD: **Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions**. *Proc Natl Acad Sci USA* 1996, **93**:10315-10320.

79. Vidal M, Braun P, Chen E, Boeke JD, Harlow E: **Genetic characterization of a mammalian protein-protein interaction domain by using a yeast reverse two-hybrid system**. *Proc Natl Acad Sci USA* 1996, **93**:10321-10330.

80. •• Dreze M, Charloteaux B, Milstein S, Vidalain PO, Yildirim MA, Zhong Q, Svrzikapa N, Romero V, Laloux G, Brasseur R *et al.*: **'Edgetic' perturbation of a *C. elegans* BCL-2 ortholog**. *Nat Methods* 2009, **6**:843-849.

This paper presents a proof-of-principle demonstration of the reverse edgetics strategy for introduction of specific edgetic alleles that disrupt critical functional interactions of a key apoptosis protein.

81. Liddington RC: **Structural basis of protein–protein interactions**. *Methods Mol Biol* 2004, **261**:3-14.

82. Han JH, Batey S, Nickson AA, Teichmann SA, Clarke J: **The folding and evolution of multidomain proteins**. *Nat Rev Mol Cell Biol* 2007, **8**:319-330.

83. Pawson T, Nash P: **Assembly of cell regulatory systems through protein interaction domains**. *Science* 2003, **300**:445-452.

84. Bhattacharyya RP, Remenyi A, Yeh BJ, Lim WA: **Domains, motifs, and scaffolds: the role of modular interactions in the evolution and wiring of cell signaling circuits**. *Annu Rev Biochem* 2006, **75**:655-680.

85. Staub O, Dho S, Henry P, Correa J, Ishikawa T, McGlade J, Rotin D: **WW domains of Nedd4 bind to the proline-rich PY**

motifs in the epithelial Na+ channel deleted in Liddle's syndrome. *EMBO J* 1996, **15**:2371-2380.

86. Guglielmi B, van Berkum NL, Klapholz B, Bijma T, Boube M, Boschiero C, Bourbon HM, Holstege FC, Werner M: **A high resolution protein interaction map of the yeast Mediator complex**. *Nucleic Acids Res* 2004, **32**:5379-5380.

87. Albers M, Kranz H, Kober I, Kaiser C, Klink M, Suckow J, Kern R, Koegl M: **Automated yeast two-hybrid screening for nuclear receptor-interacting proteins**. *Mol Cell Proteomics* 2005, **4**:205-213.

88. •• Boxem M, Maliga Z, Klitgord N, Li N, Lemmens I, Mana M, de Lichtervelde L, Mul JD, van de Peut D, Devos M *et al.*: **A protein domain-based interactome network for *C. elegans* early embryogenesis**. *Cell* 2008, **134**:534-545.

This paper introduces large-scale systematic fragmentation of proteins for delineation of minimal region of interaction for hundreds of proteins, providing critical information on their functional domain organization and domain–domain interactions.

89. Pulichino AM, Vallette-Kasic S, Couture C, Gauthier Y, Brue T, David M, Malpuech G, Deal C, Van Vliet G, De Vroede M *et al.*: **Human and mouse TPIT gene mutations cause early onset pituitary ACTH deficiency**. *Genes Dev* 2003, **17**:711-716.

90. Seto ML, Lee SJ, Sze RW, Cunningham ML: **Another TWIST on Baller-Gerold syndrome**. *Am J Med Genet* 2001, **104**:323-330.

91. Kelsey L, Flenniken AM, Qu D, Funnell AP, Pearson R, Zhou YQ, Voronina I, Berberovic Z, Wood G, Newbigging S *et al.*: **ENU-induced mutation in the DNA-binding domain of KLF3 reveals important roles for KLF3 in cardiovascular development and function in mice**. *PLoS Genet* 2013, **9**:e6121003.

92. Dickerman BK, White CL, Chevalier C, Nalesso V, Charles C, Fouchecourt S, Guillou F, Viriot L, Sen GC, Herault Y: **Missense mutation in the second RNA binding domain reveals a role for Prkra (PACT/RAX) during skull development**. *PLoS ONE* 2011, **6**:e7 2853.

93. Lee Y, Hur I, Park SY, Kim YK, Suh MR, Kim VN: **The role of PACT in the RNA silencing pathway**. *EMBO J* 2006, **25**:522-532.

94. Bubenik JL, Driscoll DM: **Altered RNA binding activity underlies abnormal thyroid hormone metabolism linked to a mutation in selenocysteine insertion sequence-binding protein 2**. *J Biol Chem* 2007, **282**:34653-34660.

95. Carpten JD, Faber AL, Horn C, Donoho GP, Briggs SL, Robbins CM, Hostetter G, Boguslawski S, Moses TY, Savage S *et al.*: **A transforming mutation in the pleckstrin homology domain of AKT1 in cancer**. *Nature* 2007, **448**:439-444.

96. Simons M, Schwarz K, Kriz W, Miettinen A, Reiser J, Mundel P, Holthofer H: **Involvement of lipid rafts in nephrin phosphorylation and organization of the glomerular slit diaphragm**. *Am J Pathol* 2001, **159**:1069-1070.

97. Shono A, Tsukaguchi H, Kitamura A, Hiramoto R, Qin XS, Doi T, Iijima K: **Predisposition to relapsing nephrotic syndrome by a nephrin mutation that interferes with assembly of functioning microdomains**. *Hum Mol Genet* 2009, **18**:2943-2950.

98. Pineda-Alvarez DE, Roessler E, Hu P, Srivastava K, Solomon BD, Siple CE, Fan CM, Muenke M: **Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its ligand, SHH**. *Hum Genet* 2012, **131**:301-310.

99. Dipple KM, McCabe ER: **Phenotypes of patients with ''simple'' Mendelian disorders are complex traits: thresholds, modifiers, and systems dynamics**. *Am J Hum Genet* 2000, **66**:1729-1730.