

Supplementary Data:

Case studies:

Case 1, Retinitis pigmentosa 28 (RP28), linkage interval 2p15-p11. Retinitis pigmentosa is the common name of a clinically and genetically heterogeneous group of disorders. Common traits are night blindness, constricted visual field, and dystrophy of the retina. Linkage to 2p15-p11 has been shown in an Indian family suffering from autosomal recessive retinitis pigmentosa ¹. This genomic area has 250 candidate genes in which the predictor has identified a candidate in this area with a score of 0.5232, approximately one out of three candidates with this score is correct in the benchmarking cases. The candidate gene (LOC130951) is an uncharacterized and highly conserved gene, that on a protein level interacts with a single protein coded by the CRX gene. This is a homeobox transcription factor, known to be involved in Leber congenital amaurosis, rod cone dystrophy and retinitis pigmentosa (<http://www.genecards.org/cgi-bin/carddisp.pl?gene=CRX&search=CRX>). The interaction between these proteins is reproducible in several organisms and has been shown both in yeast ² and a recent large scale human interaction experiment ³. If it can be verified that this candidate is the underlying cause of retinitis pigmentosum in the Indian family, this is a good example of the potential for unexpected and novel discoveries that our pipeline can lead to. This candidate is by no means obvious, but turns up as a likely candidate based on the automated integration of large scale experiments from multiple organisms, and the first clue towards the function of the gene could be provided by our method.

Case 2, Epithelial ovarian cancer susceptibility locus, 3p25-p22. Epithelial ovarian cancer (EOC) arises as a result of genetic alterations in the ovarian surface epithelium. Clinically it is characterized by loco-regional dissemination in the peritoneal cavity and the rare incidence of visceral metastases. This disease is the leading cause of death from gynecologic malignancy⁴. Amongst 200 candidates, the Bayesian predictor points out fanconi anemia group D2 protein (FANCD2) with a score of 0.9981. This protein is placed in a complex with breast cancer type 2 susceptibility protein (BRCA2), breast cancer type 1 susceptibility protein (BRCA1) and nibrin isoform 1 (NBN), all known to be involved in ovarian cancer, breast cancer, and/or chromosomal instability disorders⁵⁻⁸. Furthermore other proteins involved in cancer and or DNA repair mechanisms can be identified in the complex (i.e. Fanconi anemia group E protein (FANCE)⁹, and p53 binding protein (TP53BP1)¹⁰) (**Supplementary Fig. 4** online). Literature studies show that FANCD2 is involved in the mechanisms of the BRCA pathway in cisplatin sensitive cells¹¹. To our knowledge a mutation in this gene has never been shown to be the cause of the disease in these families, perhaps because this gene was first characterized at the time of publication of the linkage analysis¹², and the gene was first associated discovered to be involved in the same pathway as BRCA1 and BRCA2 two years after the publication of the linkage analysis¹¹. We consider it to be a very likely candidate in EOC, in the families with linkage to 3p22-p25.

Case 3, Inflammatory bowel disease (IBD), linkage interval 6p. Inflammatory bowel disease (IBD) is characterized by a chronic relapsing intestinal inflammation, and is subdivided into Crohn disease and ulcerative colitis. A susceptibility locus has been

identified on 6p by linkage analysis^{13,14}. This genomic area has 500 candidate genes. Amongst these genes the Bayesian Predictor has assigned a score of 0.9984 to one of the candidates the RIPK1 protein. Approximately one out of five candidates in the benchmarking set with this score is correct. The candidate protein is a Receptor-interacting serine/threonine protein kinase, and can be placed in a module with Tumor necrosis factor receptor 2 (TNFRSF1B), Tumor necrosis factor precursor (TNF) and Tumor necrosis factor receptor precursor (TNFRSF1A), all of which are known to be involved in inflammation and/or inflammatory bowel disease (**Supplementary Fig. 5** online). Additionally this candidate is interesting as it is known to be involved in the induction of TNF receptor I-mediated apoptosis by taking part in a two step signaling cascade together with TNFRSF1B, TNFRSF1A and TNF¹⁵, rendering it possible that defects in the candidate leads to defective signaling through this mechanism and ultimately inflammatory bowel disease. Furthermore other proteins involved in inflammatory and immune responses can be observed in the module. FADD is known to be involved in an innate immune mechanism in mammalian cells,¹⁶ and NEMO is a protein that has been associated with inflammatory bowel disease potentially by being a hub central for the regulation of the inflammatory response¹⁷. All these facts place the RIPKI protein central in a network of proteins involved in inflammatory responses and inflammatory bowel disease, and we consider it to be a very interesting candidate in the disorder. We note that RIPK1 lies 20,6 Mb from the closest high resolution marker published. However, due to the fact that all of 6q was screened for candidates, and that several genes lying far from the published markers are most likely true predictions in **Table 1**, we still find it likely that RIPK1 is involved in IBD.

Case 4, Amyotrophic lateral sclerosis (ALS) with frontotemporal dementia, linkage interval 9q21-q22. Amyotrophic Lateral Sclerosis (ALS) is a motor neurodegenerative disorder characterized by muscular atrophy, progressive motor neuron function loss and bulbar paralysis. This disorder has been observed in a large number of families combined with frontotemporal dementia and linkage has been shown to a area on 9q21-q22¹⁸ containing 167 candidate genes. The predictor has identified a candidate in this area with a score of 0.4351, approximately one out of three candidates with this score is correct in the benchmarking cases.. The candidate is cytoplasmic Isoleucyl-tRNA synthetase (IARS) (**Supplementary Fig. 6** online). This candidate is associated with the disorder by being in a module with Superoxide dismutase (SOD1), a protein in which mutations are known to be involved in amyotrophic lateral sclerosis¹⁹). Other interesting proteins such as an ubiquitin activating enzyme, and two molecular chaperones HSC70 (part of the Hsp70 gene family²⁰) and HSP 90-beta, are also present in the module with IARS and SOD1. The interaction data underlying this module comes from 9 different studies in different organisms (including human, plant and yeast) and out of the 92 interactions leading to the module 76 have been shown in more than one (and up to 8) different publications and/or species²¹⁻²⁹. Thus, the interaction data underlying the module is highly conserved, reproducible and well substantiated. It is known that mutated SOD1 inhibits chaperone activity³⁰, and a recent publication shows that SOD1 aggregations, exceeding the degenerative capacity of the proteasome and molecular Hsp70 chaperones, have recently been shown to play a role in the molecular aetiology of ALS¹⁹. Since IARS

can in this module be associated to the same proteasomal and molecular chaperone components as SOD1 (from the constitutively expressed subgroup of the Hsp70 chaperone family), it seems plausible that IARS could be involved in the disease, in a manner similar to that of SOD1, in the families studied by Hosler et al. This mechanism could also lead to neurodegeneration in the frontotemporal cortex leading to the dementia observed in the families of the linkage studies by Hosler et al. Additionally, an interesting observation is that Lewy bodies (aggregates observed in Parkinson disease) are known to sequester ubiquitin-activating enzyme (E1)³¹. This protein is also found in the module with SOD1 and IARS, further adding to the speculations of IARS' role in ALS combined with dementia.

1. Gu, S., Kumaramanickavel, G., Srikumari, C.R., Denton, M.J. & Gal, A. Autosomal recessive retinitis pigmentosa locus RP28 maps between D2S1337 and D2S286 on chromosome 2p11-p15 in an Indian family. *J Med Genet* **36**, 705-7 (1999).
2. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* **98**, 4569-74 (2001).
3. Rual, J.F. et al. Towards a proteome-scale map of the human protein-protein interaction network. *Nature* (2005).
4. Sellar, G.C. et al. OPCML at 11q25 is epigenetically inactivated and has tumor-suppressor function in epithelial ovarian cancer. *Nat Genet* **34**, 337-43 (2003).
5. Demuth, I. et al. An inducible null mutant murine model of Nijmegen breakage syndrome proves the essential function of NBS1 in chromosomal stability and cell viability. *Hum Mol Genet* **13**, 2385-97 (2004).
6. Matsuura, S. et al. Positional cloning of the gene for Nijmegen breakage syndrome. *Nat Genet* **19**, 179-81 (1998).
7. Castilla, L.H. et al. Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer. *Nat Genet* **8**, 387-91 (1994).
8. Lancaster, J.M. et al. BRCA2 mutations in primary breast and ovarian cancers. *Nat Genet* **13**, 238-40 (1996).
9. Meetei, A.R. et al. A human ortholog of archaeal DNA repair protein Hef is defective in Fanconi anemia complementation group M. *Nat Genet* **37**, 958-63 (2005).
10. Yoo, E. et al. 53BP1 is associated with replication protein A and is required for RPA2 hyperphosphorylation following DNA damage. *Oncogene* **24**, 5423-30 (2005).
11. Taniguchi, T. et al. Disruption of the Fanconi anemia-BRCA pathway in cisplatin-sensitive ovarian tumors. *Nat Med* **9**, 568-74 (2003).
12. Timmers, C. et al. Positional cloning of a novel Fanconi anemia gene, FANCD2. *Mol Cell* **7**, 241-8 (2001).
13. Dechairo, B. et al. Replication and extension studies of inflammatory bowel disease susceptibility regions confirm linkage to chromosome 6p (IBD3). *Eur J Hum Genet* **9**, 627-33 (2001).
14. Hampe, J. et al. Linkage of inflammatory bowel disease to human chromosome 6p. *Am J Hum Genet* **65**, 1647-55 (1999).
15. Micheau, O. & Tschopp, J. Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* **114**, 181-90 (2003).
16. Balachandran, S., Thomas, E. & Barber, G.N. A FADD-dependent innate immune mechanism in mammalian cells. *Nature* **432**, 401-5 (2004).
17. Burns, K.A. & Martinon, F. Inflammatory diseases: is ubiquitinated NEMO at the hub? *Curr Biol* **14**, R1040-2 (2004).
18. Hosler, B.A. et al. Linkage of familial amyotrophic lateral sclerosis with frontotemporal dementia to chromosome 9q21-q22. *Jama* **284**, 1664-9 (2000).

19. Koyama, S. et al. Alteration of familial ALS-linked mutant SOD1 solubility with disease progression: Its modulation by the proteasome and Hsp70. *Biochem Biophys Res Commun* **343**, 719-30 (2006).
20. Tavaría, M. et al. Localization of the gene encoding the human heat shock cognate protein, HSP73, to chromosome 11. *Genomics* **29**, 266-8 (1995).
21. Ho, Y. et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180-3 (2002).
22. Blagoev, B. et al. A proteomics strategy to elucidate functional protein-protein interactions applied to EGF signaling. *Nat Biotechnol* **21**, 315-8 (2003).
23. Bouwmeester, T. et al. A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat Cell Biol* **6**, 97-105 (2004).
24. Nakayama, M., Kikuno, R. & Ohara, O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. *Genome Res* **12**, 1773-84 (2002).
25. Gavin, A.C. et al. Proteome survey reveals modularity of the yeast cell machinery. *Nature* (2006).
26. Fu, H., Reis, N., Lee, Y., Glickman, M.H. & Vierstra, R.D. Subunit interaction maps for the regulatory particle of the 26S proteasome and the COP9 signalosome. *Embo J* **20**, 7096-107 (2001).
27. Zhao, R. et al. Navigating the chaperone network: an integrative map of physical and genetic interactions mediated by the hsp90 chaperone. *Cell* **120**, 715-27 (2005).
28. Conibear, E. & Stevens, T.H. Vps52p, Vps53p, and Vps54p form a novel multisubunit complex required for protein sorting at the yeast late Golgi. *Mol Biol Cell* **11**, 305-23 (2000).
29. Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-7 (2002).
30. Tummala, H. et al. Inhibition of chaperone activity is a shared property of several Cu,Zn-superoxide dismutase mutants that cause amyotrophic lateral sclerosis. *J Biol Chem* **280**, 17725-31 (2005).
31. Olanow, C.W., Perl, D.P., DeMartino, G.N. & McNaught, K.S. Lewy-body formation is an aggresome-related process: a hypothesis. *Lancet Neurol* **3**, 496-503 (2004).