

ONLINE METHODS

Constructing a mammalian protein-protein interaction network from available resources.

The protein-protein and signaling networks we chose were all literature-based “legacy” direct biochemical mammalian interactions from low-throughput functional experiments extracted manually by expert biologists (literature-curated). We did not include interactions from high-throughput methods, orthologous interactions from lower organisms, or interactions predicted using *in silico* methods. We only considered direct biophysical binding or enzymatic interactions, and excluded interactions based on functional association. The following available protein-protein interaction datasets were used: DIP¹⁹ (updated to May 30, 2006); IntAct²⁰ (updated to June 12, 2006); MINT²¹ (updated to May 21, 2006); Ma’ayan *et al.*²² (updated to May 21, 2006); BIND²³ (updated to January 24, 2006); PDZBase²⁴ (updated to September 25, 2006). We choose these datasets because components in those networks were annotated with accession codes that permit data consolidation and those datasets were provided freely for analysis and reuse. All interactions from these databases claimed to be direct biochemical interactions determined experimentally, and include the PubMed reference of the research article that described the experiments used to identify the interactions. Consolidating interactions from the different network databases was accomplished by combining human/mouse/rat gene symbols using the xml version of Swiss Prot (June 21, 2006).

Algorithm used to generate a list of novel NS candidate genes.

Problem: Given a graph G in which a small subset of vertices S , $S \subset G$, are identified as seed nodes in this case known disease genes that cause NS, find a close to minimum connected

subgraph G' that includes the seed nodes in S while pruning out intermediate nodes and links that are not statistically significant for interacting with the seed list.

Algorithm:

1. Combine available mammalian protein-protein interaction networks using as described¹⁰.
2. Filter the merged network to prune out interactions from publications reporting high-throughput interaction data as described¹⁰.
3. Find all shortest paths²⁵ of length k_l between all pairs of vertices in the merged seed list $S' \cup S''$ of all known NS disease genes.
4. Find all edges between intermediate vertices identified in 3. Intermediate vertices, I , are vertices that fall on shortest paths between pairs between all pairs in $S' \cup S''$ such that $I \subset G$ and $I \notin \{S' \cup S''\}$.
5. Combine all nodes and links found in 3 and 4 to create the subnetwork G' .
6. Rank intermediates base on their links in background network vs. links in subnetwork using a Binomial proportions test as described¹⁰.

Subjects and mutation analysis. Genomic DNAs from a cohort of 96 subjects with NS or a phenotype suggestive of this disorder without mutation in previously identified disease genes (*PTPN11*, *SOS1*, *KRAS*, *HRAS*, *RAF1*, *BRAF*, *MEK1* and *MEK2*) were screened for the entire *SHOC2* coding region using high-throughput resequencing as previously described²⁶. All sequence variants identified were verified by manual inspection of the chromatograms and putative causative mutations were verified using another independent sequencing reaction. *SHOC2* was then analyzed in a panel of 410 mutation-negative individuals with NS or a clinically related phenotype with denaturing high-performance liquid chromatography and direct sequencing²⁶. In this cohort, clinical features for the majority of subjects satisfied standardized diagnostic criteria²⁷⁻³¹, but a few individuals who lacked sufficient features for a definitive diagnosis were also included. DNA from skin fibroblasts, hair bulbs and/or epithelial cells from

the oral mucosa was extracted using standard protocols. Samples were collected under Institutional Review Board-approved protocols, with informed consent. Permission was obtained to publish the photographs of subjects shown in Figure 1. When available, parental DNAs were sequenced to establish whether identified changes were *de novo*. Paternity was confirmed using the AmpF/STR Identifier PCR Amplification Kit (Applied Biosystems).

Functional analyses. *In silico* analysis of protein *N*-myristoylation was performed using the web available Myristoylator, TerminiNator, and NMT softwares. The nucleotide substitutions of interest were introduced in V5- and Myc-tagged (*C*-terminus) human *SHOC2* cDNA expression constructs by site-directed mutagenesis (QuikChange Site-Directed Mutagenesis Kit, Stratagene). Cos-1, 293-T and Neuro2A cells were maintained in DMEM (GIBCO) supplemented with 10% heat-inactivated FBS (Euroclone) and antibiotics, and transfected at 60-70% confluency, using Fugene6 (Roche) or Lipofectamine 2000 (Invitrogen). *N*-myristoylation was evaluated by [³H]myristic acid (30μCi/ml) incorporation as described elsewhere³². Proteins immunoprecipitated with an anti-V5 antibody from cell lysates were separated by SDS-PAGE. Gels were fixed, soaked in Amplify™ (Perkin Elmer) for 30 min, dried under a GelAir drying frame (BioRad), and exposed to x-ray film (Kodak) for two months. Cellular fractionation and ERK phosphorylation assays were performed on Cos-1 cells transiently expressing the V5-tagged *SHOC2*^{wt} or *SHOC2*^{S2G} using standard protocols^{26,33}. Cells were serum starved (16 h) and then stimulated with EGF (30 to 100 ng/ml) for the indicated intervals. In all experiments, a human *NMT1* cDNA expression construct (Origene) was co-transfected to ensure that the amount of endogenous NMT would not be limiting.

Confocal laser scanning microscopy. 3×10^3 cells were seeded on glass coverslips, transiently transfected, serum starved (16 h) and stimulated with EGF (30 ng/ml, 15 min). Cells were fixed with 3% paraformaldehyde (30 min, 4°C), permeabilized with 0.5% Triton X-100 (10 min, room temperature), and stained as described in the figure legends. Imaging was performed on a Leica TCS SP2 AOBS apparatus, utilizing excitation spectral laser lines at 405, 488 and 594 nm, tuned with an acousto-optical tunable filter. Image acquisition and processing were conducted by using the Leica Confocal Software (Leica Lasertechnik GmbH). Signals from different fluorescent probes were taken in sequential scanning mode.

Generation of *C. elegans* strains and phenotypic analysis. Culture, maintenance and genetic crosses for nematodes were as described³⁴. Nematode strains were provided by the *Caenorhabditis* Genetics Center (University of Minnesota, Minneapolis, MN). The following mutant alleles were used: *sur-8^{tf}*: *sur-8(ku167) IV*; *let-60^{gof}*: *let-60(n1046) IV*; *let-23^{tf}*: *let-23(sy1) II*. V5-tagged *SHOC2^{wt}* and *SHOC2^{S2G}* cDNA were subcloned into the heat shock inducible pPD49.83 vector (a gift of Andrew Fire, Stanford University School of Medicine, Stanford, CA). A chimeric SHOC2 protein, *myr::SHOC2^{wt}*, in which the first seven amino acid residues were substituted by the *N*-terminal myristoylation signal (MGSCIGK) of *src-2* was obtained via PCR amplification and cloned into the pPD49.83 vector. Germline transformation was performed as described³⁵. *elt-2::GFP* (pJM67, a gift from James D. McGhee, University of Calgary, Calgary, Canada), which drives GFP expression in intestinal cells, was used as co-injection marker. At least three independent lines for each construct were tested for the Pvl phenotype after heat shock. All the lines expressing *SHOC2^{S2G}* or *myr::SHOC2^{wt}* upon heat shock exhibited a Pvl phenotype. Only the lines carrying the following transgenes were scored quantitatively at the compound microscope and used for further analyses and crosses: *gbEx240[hsp16.2::SHOC2^{WT} ::V5*; *pelt-*

2::GFP], *gbEx208a*[*hsp16.2::SHOC2^{S2G}::V5*; *pelt-2::GFP*] and *gbEx209*[*hsp16.2::myr::SHOC2^{wt}::V5*; *pelt-2::GFP*]. Genetic crosses were performed according to standard methods. The presence of *sur-8(ku167)*, *let-60(n1046)* and *let-23(sy1)* alleles was confirmed by sequencing the appropriate region of genomic DNA from each transgenic strain. After each cross, isogenic worms that had lost the transgene were cloned separately and used as controls. Animals were scored blindly at the dissecting microscope to count the number of eggs *in utero* after cutting the mother (Egl), animals that had become bags of worms (Bag) and to check for the presence of multiple ectopic pseudovulvae (Muv). A subset of worms was also scored blindly at the compound microscope for vulva morphology and VPC induction phenotypes.

C. *elegans* heat shock experiments, microscopy and immunocytochemistry. At different developmental stages, worms carrying the transgenes were subjected to heat shock at 33 °C for 30 min and then kept at 30 °C for 1 h. Synchronized embryos were heat shocked to study the effects of transgene expression on embryonic and early larval development, while synchronized L1/L2 larvae were heat shocked to study the effects on later larval development, movement and fertility. To study VPC induction and vulva morphogenesis, hermaphrodites were heat shocked at early L3 stages and animals were scored for vulval induction at the L4 stage and for Pvl phenotype at the adult stage. Microscopy observations were performed with a Zeiss Axioskop equipped with epifluorescence and Differential Interference Contrast on live animals anesthetized and mounted on 2% agarose pads containing 10 mM Na-Azide. Images were collected with an Axiocam digital camera. Confocal analyses were performed using a Leica TCS SP2 microscope. For immunocytochemistry analyses, transgenic worms were heat shocked, and after 2 h were fixed with 2% PFA (R.T. 5 min, 1 h on ice). They were processed as reported³⁶, and then incubated overnight in a dilution of anti-V5 monoclonal antibody (1:200). After repeated washing (24 h), animals were incubated overnight with Texas-Red conjugated anti-mouse secondary antibody (1:100) (Invitrogen), washed and mounted for observation on microscope slides.

URLs

DIP, <http://dip.doe-mbi.ucla.edu/>; IntAct, <ftp://ftp.ebi.ac.uk/pub/databases/intact/current>; MINT, <http://mint.bio.uniroma2.it/mint-old/release/main.php>; Iyengar web resources, <http://www.mssm.edu/labs/iyengar/resources>; BIND, <http://www.bind.ca/>; PDZBase, <http://icb.med.cornell.edu/services/pdz/start>; Swiss Prot, <http://www.pir.uniprot.org/database/>; GenBank, <http://www.ncbi.nlm.nih.gov/nucleotide/>; Ensembl, <http://www.ensembl.org/index.html>; OMIM, <http://www.ncbi.nlm.nih.gov/omim>; Prosite, <http://www.expasy.ch/tools/scanprosite/>; Pfam database, <http://pfam.janelia.org/>; Myristoylator, <http://www.expasy.org/tools/myristoylator/>; TerminiNator, <http://www.isv.cnrs-gif.fr/terminator3/index.html>; NMT, <http://mendel.imp.ac.at/sat/myristate/index.html>.