doi:10.1038/nature04209

LETTERS

Towards a proteome-scale map of the human protein–protein interaction network

Jean-François Rual[1]*, Kavitha Venkatesan[1]*, Tong Hao[1], Tomoko Hirozane-Kishikawa[1], Amélie Dricot[1], Ning Li[1], Gabriel F. Berriz[2], Francis D. Gibbons[2], Matija Dreze[1,3], Nono Ayivi-Guedehoussou[1], Niels Klitgord[1], Christophe Simon[1], Mike Boxem[1], Stuart Milstein[1], Jennifer Rosenberg[1], Debra S. Goldberg[2], Lan V. Zhang[2], Sharyl L. Wong[2], Giovanni Franklin[2], Siming Li[1]†, Joanna S. Albala[1]†, Janghoo Lim[4], Carlene Fraughton[1], Estelle Llamosas[1], Sebiha Cevik[1], Camille Bex[1], Philippe Lamesch[1,3], Robert S. Sikorski[5], Jean Vandenhaute[3], Huda Y. Zoghbi[4], Alex Smolyar[1], Stephanie Bosak[6], Reynaldo Sequerra[6], Lynn Doucette-Stamm[6], Michael E. Cusick[1], David E. Hill[1], Frederick P. Roth[2] & Marc Vidal[1]

Systematic mapping of protein–protein interactions, or 'interactome' mapping, was initiated in model organisms, starting with defined biological processes[1,2] and then expanding to the scale of the proteome[3–7]. Although far from complete, such maps have revealed global topological and dynamic features of interactome networks that relate to known biological properties[8,9], suggesting that a human interactome map will provide insight into development and disease mechanisms at a systems level. Here we describe an initial version of a proteome-scale map of human binary protein–protein interactions. Using a stringent, high-throughput yeast two-hybrid system, we tested pairwise interactions among the products of ~8,100 currently available Gateway-cloned open reading frames and detected ~2,800 interactions. This data set, called CCSB-HI1, has a verification rate of ~78% as revealed by an independent co-affinity purification assay, and correlates significantly with other biological attributes. The CCSB-HI1 data set increases by ~70% the set of available binary interactions within the tested space and reveals more than 300 new connections to over 100 disease-associated proteins. This work represents an important step towards a systematic and comprehensive human interactome project.

Our working definition of a human interactome map is the complete collection of binary protein–protein interactions detectable in one or more exogenous assay. This definition excludes dynamic and functional properties of these interactions (Supplementary Data I). Thus, we treat interactome maps as 'scaffold' information, from which increasingly detailed and reliable biological models can be generated by integrating other functional genomic and proteomic data sets[10] (Supplementary Data II).

The currently available information on the human interactome network originates from either literature-curated (LC) interactions[11–15], or from 'interologs' (that is, potential interactions predicted from interactome data available for model organisms given evolutionary conservation of two known partners)[2,16]. This information needs to be complemented by systematic experimental mapping approaches that are: (1) not biased towards any particular biological interest (that is, without 'inspection bias'), as is the case for LC data sets; (2) more complete; and (3) supported by experiments rather than predictions. We are mapping the human interactome network systematically in successive versions, with each version defined by the availability of recombinationally cloned open reading frames (ORFs) in the human 'ORFeome'[17].

In this initial version, we use human ORFeome v1.1 (ref. 17), a resource containing ~8,100 Gateway-cloned ORFs (generated using the Mammalian Gene Collection, as previously described[17]) that correspond to ~7,200 distinct protein-coding genes (Supplementary Table S1). Thus, our initial 'search space' (Space-I) encompasses protein pairs encoded by a 7,200 × 7,200 matrix of genes. Future interactome versions can be generated by successively increasing the search space as additional versions of the human ORFeome become available (Supplementary Data III). Accepting a total of ~22,000 protein-coding genes in the human genome[18] and excluding polymorphic and splice variants, Space-I corresponds to ~10% of the total search space for a comprehensive human interactome map (Fig. 1a). Currently, 4,067 binary LC interactions are available in Space-I (LCI interactions; Supplementary Table S2).

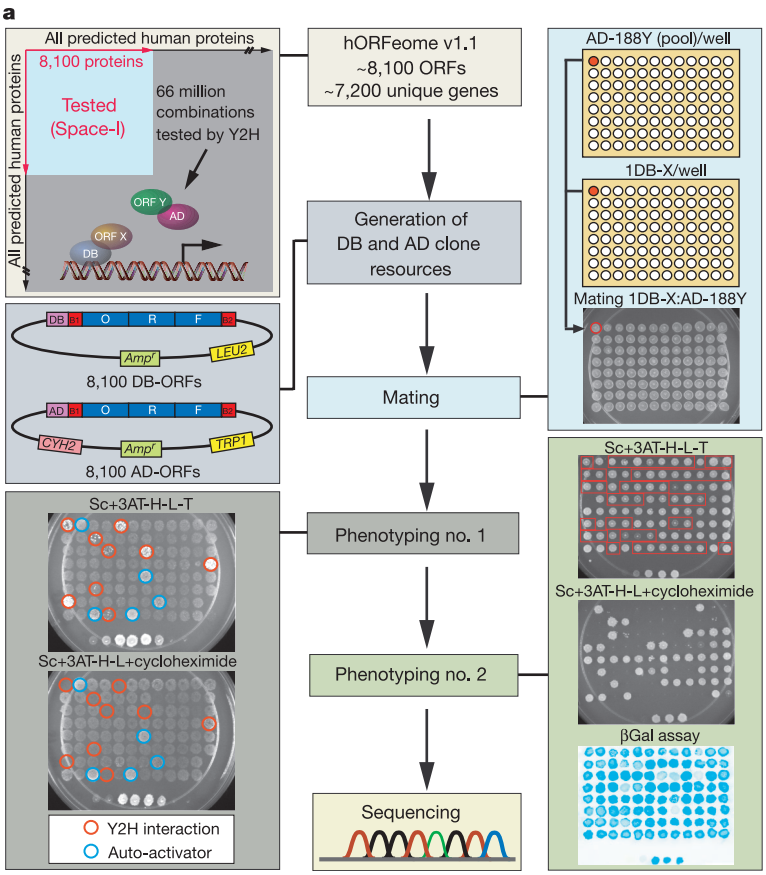
Our high-throughput yeast two-hybrid system is highly specific, benefiting from the following features, which were not uniformly present in earlier large-scale studies: relatively low levels of expression of both Gal4 DNA-binding domain (DB) and Gal4 activation domain (AD) hybrid proteins (DB-X and AD-Y, or DB-ORF and AD-ORF); three different yeast two-hybrid inducible reporter genes; and a plasmid-shuffling counter selection to eliminate systematically *de novo* auto-activators[19] (Supplementary Data IV). We tested each of the ~8,100 individual DB-X proteins against 45 mini-libraries, each containing a pool of 188 AD-Y fusion proteins (AD-188Ys), by yeast-mating in a 96-well format (Fig. 1a). Such small pools offer high sensitivity, because positive clones are less likely to be masked by other AD-Y clones within the same pool. Indeed, our overall reproducibility rate was ~55%, close to that observed in proteome-scale affinity purification followed by mass spectrometry experiments[20] (Supplementary Data V).

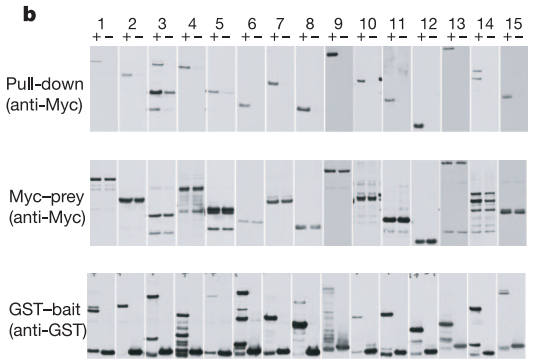
In our Space-I yeast two-hybrid matrix, we identified ~65,000 primary positive colonies, of which 12,251 scored positive after

[1]Center for Cancer Systems Biology and Department of Cancer Biology, Dana-Farber Cancer Institute and Department of Genetics, Harvard Medical School, 44 Binney Street, Boston, Massachusetts 02115, USA. [2]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 250 Longwood Ave, Boston, Massachusetts 02115, USA. [3]Unité de Recherche en Biologie Moléculaire, Facultés Notre-Dame de la Paix, 61 Rue de Bruxelles, 5000 Namur, Belgium. [4]Howard Hughes Medical Institute, and Departments of Pediatrics, Neurology, Neuroscience, and Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. [5]Arcbay, Inc., 6 Whittier Place, Suite 7J, Boston, Massachusetts 01915, USA. [6]Agencourt Bioscience Corporation, 500 Cummings Center, Suite 2450, Beverly, Massachusetts 01915, USA.
†Present addresses: ArQule, Inc., 19 Presidential Way, Woburn, Massachusetts 01081, USA (S.L.); Departments of Cancer Biology, and Otolaryngology, Head and Neck Surgery, University of California Davis, 2521 Stockton Blvd, Suite 7200, Sacramento, California 95817, USA (J.S.A.).
*These authors contributed equally to this work.

stringent phenotype testing; that is, we only retained clones that were positive for at least two yeast two-hybrid reporter assays and we controlled for auto-activation (Fig. 1a). Both DB-X and AD-Y fragments from these two-reporter-positive colonies were amplified by polymerase chain reaction (PCR) and sequenced to generate 10,597 pairs of interaction sequence tags (ISTs) (Supplementary Fig. S1a). We then collapsed yeast two-hybrid ISTs corresponding to the same pair of genes and removed lower confidence interactions (Supplementary Data VI). This resulted in a data set containing 2,754 yeast two-hybrid interactions—the Center for Cancer Systems Biology Human Interactome version 1 (CCSB-HI1) (Supplementary Fig. S1a and Supplementary Table S2. All data is publicly available (Supplementary Data VII). In all, CCSB-HI1 provides interaction information for 1,549 proteins, ~21% of the proteins tested in Space-I.

To measure the specificity of CCSB-HI1, we considered technical and biological false positives separately (Supplementary Data VIII). Technical false positives arise from experimental errors that can and should be avoided. To estimate our technical false positive rate, representative samples of interactions were verified by *in vivo* co-affinity purification glutathione *S*-transferase (GST) pull-down assay in human 293T cells (Supplementary Data IX). The verification rates for co-affinity purification were ~78% for yeast two-hybrid-only interactions, ~62% for LCI-only interactions and ~81% for interactions present in both LCI and yeast two-hybrid data sets (Fig. 1b; see also Supplementary Fig. S1b, c and Supplementary Tables S2 and S3). Given these results and the fact that co-affinity purification GST pull-down assays are not perfectly sensitive, we argue that CCSB-HI1 is largely free of technical false positives, and comparable in reliability to LCI interactions, supporting the validity of our improvements of





	Y2H-only	LCI-only	Y2H and LCI	Y2H/LCI-negative
Total	117	8	8	9
co-AP$^+$	100	6	7	3
co-AP$^-$	17	2	1	6
Success rate	85.5%	75%	87.5%	33.3%
Adjusted success rate	78.2%	62.5%	81.3%	NA

Figure 1 | Towards the generation of a proteome-scale human yeast two-hybrid map. a, Schema of the high-throughput yeast two-hybrid pipeline. Individual steps (middle column) and representative examples (flanking left and right columns) are indicated. The top panel of the left column represents the matrix of all protein pairs. All available ORFs from human ORFeome v1.1 were transferred into both DB and AD vectors by recombinational cloning (middle panel of left column). The top panel of the right column shows the mating process, with each bait mated to individual pools of 188 AD-ORFs. Initial phenotypic testing evaluated growth of diploid cells on selective medium in response to enhanced levels of the *GAL1::HIS3* selective marker (bottom panel of left column). All positive diploids from phenotyping no. 1 (red circles) were subsequently tested for activation of both *GAL1::HIS3* and *GAL1::lacZ* reporter genes. Auto-activators were identified by growth on medium containing cycloheximide (bottom panels of left and right columns). Positive colonies from phenotyping no. 2 (outlined in red) were isolated and used to PCR-amplify both DB-ORF and AD-ORF fragments for sequencing. **b**, Verification of yeast two-hybrid interactions by co-affinity purification assays. Fifteen representative examples of co-affinity purification-positive assays are shown. The middle and bottom panels show expression controls of Myc–prey and GST–bait fusion proteins, respectively. Each lane pair in the top panels shows presence or absence of Myc–prey fusions after affinity purification, demonstrating binding to GST–bait fusion proteins (+) or to GST alone (−). The Table summarizes the data obtained for four different classes of protein pairs. 'Y2H and LCI' describes interactions reported in both the yeast two-hybrid and LCI data sets. 'Y2H/LCI-negative' describes pairs of proteins that were not reported to interact either in the yeast two-hybrid or in the LCI data sets. Rows indicate the total number of interactions tested and considered for scoring (Total), the number of interactions not verified by co-affinity purification (co-AP$^-$), the number of interactions verified by co-affinity purification (co-AP$^+$), the proportion of co-affinity purification-positive interactions (success rate), and the adjusted success rate (which accounts for the observation that one-third of all co-affinity purification experiments yield an apparently positive result without regard to whether or not the protein pair truly interacts; see Supplementary Data IX). Identities, lane positions and scoring of all protein pairs tested by co-affinity purification are provided in Supplementary Tables S2 and S3.

the yeast two-hybrid methodology. Estimating biological false positive interactions, which are genuinely observed in one or more assay but do not occur *in vivo*, is more difficult. We partially addressed this by examining the correlation of CCSB-HI1 data with other biological information (see below).

To measure the sensitivity of CCSB-HI1, we selected two high-confidence subsets from among all 4,067 LCI direct binary interactions. LCI-core contains 624 interactions supported by at least two PubMed entries. LCI-hypercore contains 275 interactions supported by at least two PubMed entries and present in at least two curated databases (Supplementary Table S2). Overall, the fractions of LCI, LCI-core and LCI-hypercore interactions found in CCSB-HI1 are 2.3%, 4.6% and 8.4%, respectively (Fig. 2a). These overlaps are larger than expected by chance ($P < 6 \times 10^{-56}$) and are similar to those found for interactome maps in *Caenorhabditis elegans* and *Drosophila melanogaster*[7,21]. That the fraction of CCSB-HI1 interactions increases markedly with increasingly confident subsets of LCI suggests that literature-derived interactions are variable in quality and should not necessarily be interpreted as a 'gold standard'. Because Space-I represents ~10% of the human network (without accounting for alternative splice variants), and because we detected ~10% of LCI-hypercore interactions, we conclude that the CCSB-HI1 data set contains ~1% of the human interactome (Supplementary Data X).

We represented the union of all CCSB-HI1 and LCI interactions in a network graph in which nodes are proteins and edges are interactions. The main component of this network contains 2,784 nodes and 6,438 edges (Fig. 2b), and shows interactions largely segregated into two neighbourhoods: one enriched for CCSB-HI1 interactions (red edges) and the other enriched for LCI interactions (blue edges). To explore this hypothesis, we calculated, for each node, the fraction of yeast two-hybrid edges within paths of length 1, 2 and 3 (that is, within '1-hop', '2-hop' and '3-hop' neighbourhoods). The distribution of this fraction (Fig. 2c; see also Supplementary Fig. S2) confirms the evidence-type segregation apparent in Fig. 2b. One explanation for this phenomenon is that different biases exist in the CCSB-HI1 and LCI data sets. For example, certain protein classes (such as those involved in cancer) are studied more extensively than others, resulting in an inherent inspection bias in LC data (Supplementary Table S4). Furthermore, the methodologies used to detect interactions (including yeast two-hybrid) each have different biases for example, under-representation of membrane proteins (Supplementary Data X).

The novelty of CCSB-HI1 interactions was evaluated by systematically searching the PubMed and Google Scholar literature databases for co-occurrence of the corresponding gene symbols. More than 85% of the CCSB-HI1 pairs (as compared with only 25% of

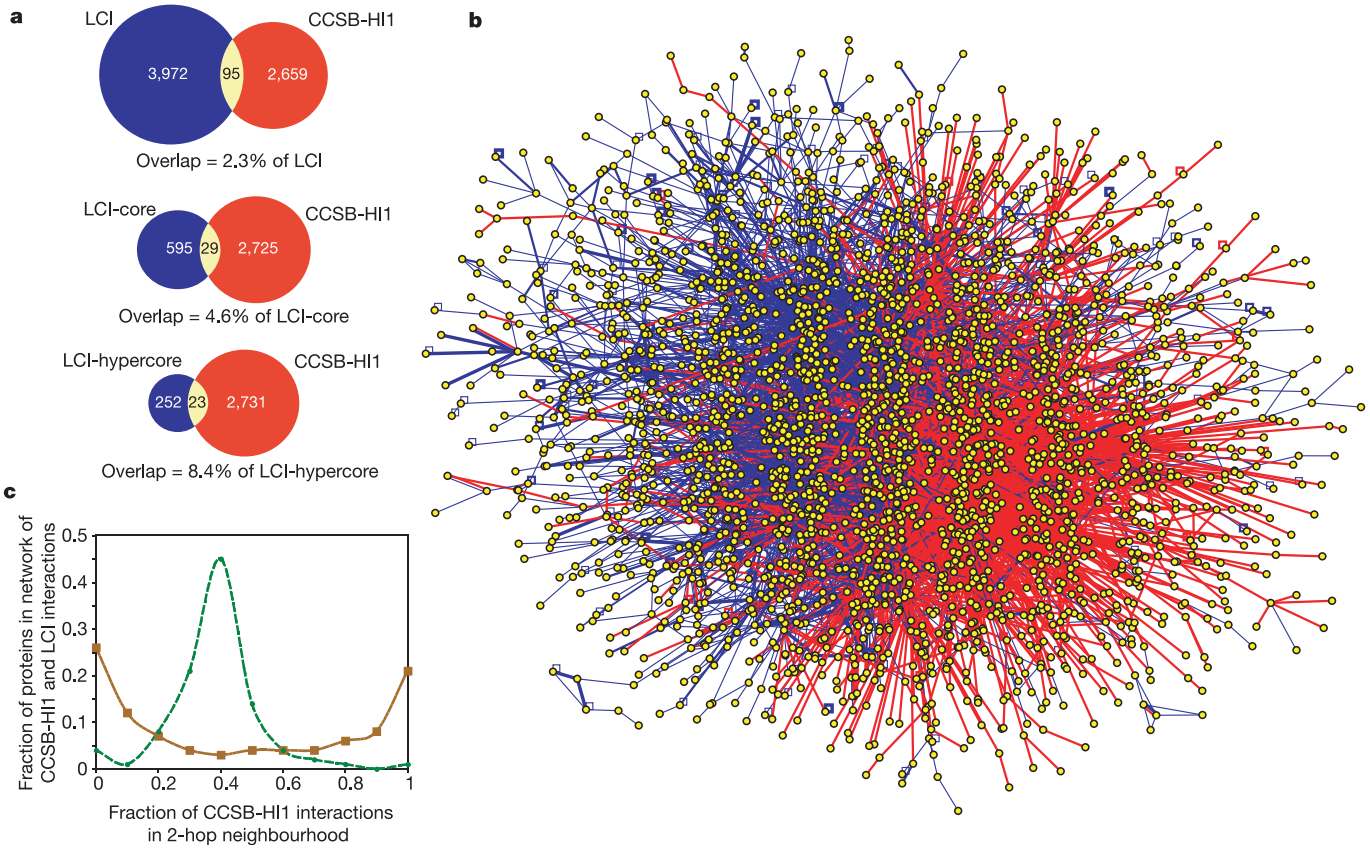


Figure 2 | Overlap of CCSB-HI1 with existing literature-curated (LC) data. **a**, Overlap between CCSB-HI1 and LC interactions in Space-I (LCI). The top, middle and bottom panels represent the overlap between CCSB-HI1 and LCI, LCI-core and LCI-hypercore, respectively. **b**, Network graph of the union of all CCSB-HI1 and LCI interactions. Proteins are shown as yellow nodes and CCSB-HI1 and LCI interactions are shown as red and blue edges, respectively. Blue edges with increasing thickness indicate LCI-non-core, LCI-core and LCI-hypercore, respectively. The apparent banding pattern of the yellow nodes is an artefact of the graph layout algorithm (Supplementary Data). Importantly, the layout algorithm was not informed by type of supporting evidence and therefore does not explain the evident separation of blue and red edges. **c**, Bias in 2-hop network neighbourhood for either CCSB-HI1 or LCI interactions. The frequency of nodes with a given proportion of CCSB-HI1 interactions in their 2-hop neighbourhood is depicted for the interactome network graph in **b** (solid curve) and for a network in which the types of supporting evidence (CCSB-HI1 or LCI) are randomly permuted among edges (dashed curve). The solid curve indicates that most of the proteins in the network of **b** have either only CCSB-HI1 or only LCI interactions in their 2-hop neighbourhood. In contrast, neighbourhoods are well mixed when evidence labels are randomly permuted among edges.

Table 1 | Overlap of protein interactions with other gene- or protein-pair characteristics

Protein pairs	Share mouse phenotype		Share upstream motif		Have correlated expression	
	F(*C*)	*P*-value	*F*(*C*)	*P*-value	*F*(*C*)	*P*-value
All possible within Space-I	0.128	NA	0.086	NA	0.063	NA
CCSB-HI1	0.257	2.53×10^{-3}	0.115	1.14×10^{-4}	0.130	2.14×10^{-7}
LCI	0.336	4.91×10^{-43}	0.146	9.05×10^{-20}	0.204	5.45×10^{-56}
LCI-core	0.471	7.53×10^{-20}	0.137	3.77×10^{-3}	0.243	3.57×10^{-12}
LCI-non-core	0.306	2.54×10^{-27}	0.147	3.65×10^{-18}	0.198	7.78×10^{-46}

Protein pairs	Share GO component		Share GO function		Share GO process	
	F(*C*)	*P*-value	*F*(*C*)	*P*-value	*F*(*C*)	*P*-value
All possible within Space-I	0.059	NA	0.021	NA	0.036	NA
CCSB-HI1	0.488	1.49×10^{-28}	0.250	5.49×10^{-20}	0.233	2.68×10^{-28}
LCI	0.656	6.47×10^{-139}	0.228	5.72×10^{-120}	0.410	1.74×10^{-405}
LCI-core	0.870	5.90×10^{-43}	0.270	1.70×10^{-32}	0.616	3.40×10^{-137}
LCI-non-core	0.610	3.84×10^{-100}	0.218	1.33×10^{-89}	0.368	4.03×10^{-280}

F(*C*) represents the fraction of gene- or protein-pairs (defined for each row) the given characteristic *C*. Assessed characteristics include shared mouse phenotype, shared upstream motif, correlated expression[30] and shared Gene Ontology annotation. 'All possible within Space-I' represents all possible gene- or protein-pairs in Space-I for which information regarding *C* is available. For each analysis of a shared characteristic, only gene- or protein-pairs for which both members had some annotation for that characteristic were considered. For analysis of correlated expression, only gene pairs with expression measurements for both genes were considered.

Figure 3 | Interaction network of disease-associated CCSB-HI1 proteins. The network has 121 OMIM disease-associated proteins (green nodes) and 424 CCSB-HI1 interactions involving them (red edges), along with known LC interactions (solid blue edges represent binary LCI interactions and dashed blue edges represent non-binary interactions). Proteins without an OMIM disease association are depicted as yellow nodes, and blue edges with increasing thickness indicate LCI-non-core, LCI-core and LCI-hypercore interactions, respectively. We note that 94 out of the 424 CCSB-HI1 interactions involve the Ewing sarcoma related protein (EWSR1; also known as EWS).

LCI pairs) showed no linkage of the corresponding gene symbols (Supplementary Fig. S3). These results indicate that most of our yeast two-hybrid interactions are novel.

To determine whether messenger RNAs corresponding to interacting protein pairs are likely to be co-expressed, we used Pearson correlation coefficients of the corresponding gene pairs in the CCSB-HI1 and LCI data sets from four expression studies in diverse human and mouse tissues (Supplementary Data XI). LCI pairs were enriched for correlated expression in all four cases ($P < 3 \times 10^{-17}$) and CCSB-HI1 pairs were enriched in three of the four cases ($P < 3 \times 10^{-5}$) (Table 1; see also Supplementary Fig. S4a and Supplementary Tables S2 and S5). In addition, CCSB-HI1 interactions are more enriched than would be expected by chance for (1) presence of a common upstream DNA sequence that is conserved across human, mouse, rat and dog genomes[22] ($P = 1 \times 10^{-4}$), (2) orthologous genes in mouse having a specific phenotype in common[23] ($P = 3 \times 10^{-3}$), and (3) annotation with the same Gene Ontology (GO) terms[24] ($P < 6 \times 10^{-20}$ for all three GO branches) (Table 1; see also Supplementary Tables S2 and S5 and Supplementary Data XI).

The higher likelihood of LCI interactions to share other biological attributes is not surprising given inspection bias and potential circularity where functional annotation has been derived from an LCI interaction. That the CCSB-HI1 interaction pairs (which do not have such biases) yield statistically significant correlation supports their biological relevance. In all, 357 CCSB-HI1 interactions are supported by at least one additional characteristic and thus represent particularly appealing hypotheses of functional relatedness (Supplementary Fig. S4b). Lack of additional biological evidence is not an argument against any interaction (Supplementary Data XII). Importantly, complementary information from the interactome and other functional genomic data can be integrated to formulate biological models[25].

The CCSB-HI1 network has topological properties that are similar to other sampled interactome networks, such as an approximately power-law degree distribution, hierarchical organization and a tendency for highly connected (hub) proteins to interact with less highly connected proteins (Supplementary Data XIII and Supplementary Fig. S5a–d). Surprisingly, although the CCSB-HI1 network has a small characteristic path length, it does not exhibit high clustering, seemingly contradicting findings from the sparsely sampled networks of other organisms that protein interaction networks are 'small world'[6,7,26] (that is, have a short characteristic path length and a high clustering coefficient). Possible explanations for this apparent discrepancy are discussed in Supplementary Data XIII.

To gain an insight into the evolution of the interactome, we classified proteins in the CCSB-HI1 network as 'eukaryotic', 'metazoan', 'mammalian' or 'human', and asked whether proteins specific to different evolutionary classes tend to interact with one another. The CCSB-HI1 network appears to be enriched for interactions between proteins of the same evolutionary class but not for interactions between proteins from two different evolutionary classes (Supplementary Table S6). This suggests that the human interactome has evolved through the preferential addition of interactions between lineage-specific proteins. Further investigation may provide hypotheses for mechanisms underlying interactome evolution.

To detect densely connected subgraphs potentially representing biological modules, we applied the MCODE graph clustering algorithm[27] to the CCSB-HI1 and to the combined CCSB-HI1–LCI and CCSB-HI1–LC networks (Supplementary Fig. S6, Supplementary Table S7 and Supplementary Data XIV). We identified functionally enriched MCODE complexes using FuncAssociate[28]. Out of 172 complexes (Supplementary Table S7) containing at least one CCSB-HI1 interaction, we identified 102 in which at least one GO term was significantly over-represented ($P \leq 0.05$), ten times the number expected by chance alone. The enriched functional terms we identified may also apply to unannotated proteins present in the complex ('guilt by association' predictions).

CCSB-HI1 represents a repository of novel biological hypotheses for genes implicated in human diseases. We compared all CCSB-HI1 proteins to the list of genes associated with human diseases in the Online Mendelian Inheritance in Man (OMIM) database and identified 424 interacting pairs for which at least one partner had been previously associated with a human disease (Supplementary Table S8). In a query of PubMed and Google Scholar, searching for gene symbols, 352 of the 424 interaction pairs appeared to be new based on the absence of any hit in either database. Along with 79 LC interactions (including LCI and non-binary LC interactions) among proteins in this space, the resulting network contains 484 interactions among 417 proteins (Fig. 3).

In one example, we found an interaction between RTN4, a neurite outgrowth inhibitor, and SPG21, the spastic paraplegia 21 protein. Mutations in SPG21 cause an autosomal recessive motor disorder called Mast syndrome. SPG21 protein localizes to intracellular endosomal/trans-Golgi transportation vesicles and is thought to function in protein transport and sorting. Although the function of its interacting partner, RNT4, remains elusive, RNT4 belongs to a family of proteins that localize to the endoplasmic reticulum and are markers of neuroendocrine differentiation[29]. In addition, RNT4 shares two regulatory motifs with SPG21 that are conserved across mammalian genomes[22] and may have a role in Mast syndrome. This and other examples (Supplementary Data XV) suggest that CCSB-HI1 can be used to connect biological processes in order to understand further network and disease relationships.

Although the CCSB-HI1 data set is far from comprehensive, and incomplete sampling limits conclusions regarding some network properties, our data provide useful hypotheses and can guide further studies of the expanded network. Currently, CCSB-HI1 is a static graph. Eventually the dynamics of the human interactome network will need to be considered to address where and when interactions take place and how they are regulated. The functional consequences of these physical interactions will also have to be studied to understand the logic of complex biological networks. Just as the first drafts of the human genome changed strategies for disease gene identification, the emerging human interactome will greatly further the understanding of human health and disease.

METHODS

The CCSB-HI1 data set was generated using a high-throughput version of the yeast two-hybrid system. First, all 8,100 cloned ORFs of the human ORFeome v1.1 were transferred from Entry clones to both AD and DB vectors by Gateway recombinational cloning. Resulting constructs were transformed in haploid yeast cells. After mating, diploids were tested on selective media for their ability to grow in an AD-Y-dependent manner. The identity of the interactors was determined after PCR amplification and sequencing of the AD and DB inserts from positive colonies. Resulting interactions were re-tested by the yeast two-hybrid system, individually, assessed for quality by co-affinity purification assays and analysed for correlation with other biological information. For a detailed description of the various methods, see Supplementary Data.

Received 21 July; accepted 8 September 2005.
Published online 28 September 2005.

1. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16,** 277–282 (1997).
2. Walhout, A. J. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122 (2000).
3. Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
4. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98,** 4569–4574 (2001).
5. Reboul, J. *et al. C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nature Genet.* **34,** 35–41 (2003).
6. Giot, L. *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* **302,** 1727–1736 (2003).
7. Li, S. *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* **303,** 540–543 (2004).

8. Jeong, H., Mason, S. P., Barabasi, A. L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411,** 41–42 (2001).
9. Han, J. D. *et al.* Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430,** 88–93 (2004).
10. Vidal, M. A biological atlas of functional maps. *Cell* **104,** 333–339 (2001).
11. Xenarios, I. *et al.* DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30,** 303–305 (2002).
12. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Lett.* **513,** 135–140 (2002).
13. Bader, G. D., Betel, D. & Hogue, C. W. BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* **31,** 248–250 (2003).
14. Peri, S. *et al.* Development of human protein reference database as an initial platform for approaching systems biology in humans. *Genome Res.* **13,** 2363–2371 (2003).
15. Pagel, P. *et al.* The MIPS mammalian protein-protein interaction database. *Bioinformatics* **21,** 832–834 (2005).
16. Lehner, B. & Fraser, A. A first-draft human protein-interaction map. *Genome Biol.* **5,** R63 (2004).
17. Rual, J. F. *et al.* Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res.* **14,** 2128–2135 (2004).
18. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* **431,** 931–945 (2004).
19. Walhout, A. J. & Vidal, M. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* **9,** 1128–1134 (1999).
20. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415,** 141–147 (2002).
21. Formstecher, E. *et al.* Protein interaction mapping: *A Drosophila* case study. *Genome Res.* **15,** 376–384 (2005).
22. Xie, X. *et al.* Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345 (2005).
23. Eppig, J. T. *et al.* The Mouse Genome Database (MGD): from genes to mice—a community resource for mouse biology. *Nucleic Acids Res.* **33,** D471–D475 (2005).
24. Camon, E. *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* **32,** D262–D266 (2004).
25. Gunsalus, K. C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436,** 861–865 (2005).
26. Wagner, A. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Mol. Biol. Evol.* **18,** 1283–1292 (2001).
27. Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4,** 2 (2003).
28. Berriz, G. F., King, O. D., Bryant, B., Sander, C. & Roth, F. P. Characterizing gene sets with FuncAssociate. *Bioinformatics* **19,** 2502–2504 (2003).
29. Huang, X. *et al.* Overexpression of human reticulon 3 (hRTN3) in astrocytoma. *Clin. Neuropathol.* **23,** 1–7 (2004).
30. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302,** 2141–2144 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This paper is dedicated to the memory of Stan Korsmeyer. We thank members of the Vidal laboratory and the participants of the ORFeome Meeting for discussions; the sequencing staff at Agencourt Biosciences for technical assistance; E. Smith for his help with the figures; C. McCowan, A. Bird, T. Clingingsmith and C. You for administrative assistance; and E. Benz, S. Korsmeyer, D. Livingston, P. McCue, J. Song, B. Rollins and the DFCI Strategic Planning Initiative for support. Our human interactome project is supported by the DFCI High-Tech Fund (S. Korsmeyer), an Ellison Foundation grant awarded to M.V., an NIH/NCI grant awarded to S. Korsmeyer, S. Orkin, G. Gilliland and M.V., an 'interactome mapping' grant from NIH/NHGRI and NIH/NIGMS awarded to F.P.R. and M.V., and a W.M. Keck Foundation grant awarded to E. Benz, J. Marto, F.P.R. and M.V. Other support includes Taplin Funds for Discovery (F.P.R., F.D.G. and G.F.B), a 2003 NSF Fellowship (D.S.G) and funding from the Fonds National de la Recherche Scientifique, Belgium (M.D.).

Author Contributions Experiments and data analyses were coordinated by J.F.R., T.H. and K.V. High-throughput ORF cloning and yeast two-hybrid screens were performed by J.F.R., T.H.K., A.D., N.L., N.A.G., J.R. and J.L. J.F.R developed the high-throughput yeast two-hybrid strategy. Computational analyses were performed by T.H., K.V., G.F.B., F.D.G., N.K., P.L., D.S.G., L.V.Z., S.L.W. and G.F. Co-affinity purification experiments were performed by M.D., C.S., J.F.R., S.M., M.B., S.L. and J.S.A. C.F., E.L., S.C. and C.B. provided laboratory support. R.S.S., J.V., H.Y.Z., A.S. and M.E.C. helped with the overall interpretation of the data. DNA sequencing was performed by S.B., R.S. and L.D.S. The manuscript was written by J.F.R., K.V., M.E.C., D.E.H., F.P.R. and M.V. The project was conceived by M.V. and co-directed by D.E.H., F.P.R. and M.V.

Author Information Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.V. (marc_vidal@dfci.harvard.edu), F.P.R. (fritz_roth@hms.harvard.edu) or D.E.H. (david_hill@dfci.harvard.edu).