## Building of the human protein interactome

To build a comprehensive human protein-protein interactome we combined total 15 databases with various kinds of experimental evidences as currently available. Specifically, we focus one the high-quality PPIs with four types of data.

(1) binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems: we combined binary PPIs tested from two public available high-quality Y2H datasets [[1](#_ENREF_1),[2](#_ENREF_2)], and one unpublished datasets but available from our website(<http://ccsb.dfci.harvard.edu/web/www/ccsb/Research/interactomegroup.html>). (2) kinase-substrate interactions by literature-derived low-throughput and high-throughput experiments from KinomeNetworkX [[3](#_ENREF_3)], Human Protein Resource Database (HPRD) [[4](#_ENREF_4)], PhosphoNetworks [[5](#_ENREF_5),[6](#_ENREF_6)], PhosphositePlus [[7](#_ENREF_7)], DbPTM 3.0 [[8](#_ENREF_8)], and Phospho. ELM [[9](#_ENREF_9)].

(3) carefully literature-curated PPIs identified by affinity purification followed by mass spectrometry (AP-MS), Y2H and by literature-derived low-throughput experiments, and protein three-dimensional structures from BioGRID [[10](#_ENREF_10)], PINA [[11](#_ENREF_11)], Instruct [[12](#_ENREF_12)], HPRD [[4](#_ENREF_4)], MINT [[13](#_ENREF_13)], IntAct [[14](#_ENREF_14)], and InnateDB [[15](#_ENREF_15)].

(4) signaling network by literature-derived low-throughput experiments as annotated in SignaLink2.0 [[16](#_ENREF_16)].

All data were downloaded in December, 2015. The genes were mapped to their Entrez ID based on the NCBI database [[17](#_ENREF_17)] as well as their official gene symbols based on GeneCards (http://www.genecards.org/). Duplicated pairs were removed. Hence, the data from inferred data, such as evolutionary analysis, gene expression data, and metabolic associations were excluded. The current updated human interactome includes 243,603 protein-protein interactions (PPIs) connecting 16,677 unique proteins, with over 40% increasing size in comparison with our previously used human interactome [[18](#_ENREF_18)].

1. Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, et al. (2014) A proteome-scale map of the human interactome network. Cell 159: 1212-1226.

2. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Towards a proteome-scale map of the human protein-protein interaction network. Nature 437: 1173-1178.

3. Cheng F, Jia P, Wang Q, Zhao Z (2014) Quantitative network mapping of the human kinome interactome reveals new clues for rational kinase inhibitor discovery and individualized cancer therapy. Oncotarget 5: 3697-3710.

4. Peri S, Navarro JD, Kristiansen TZ, Amanchy R, Surendranath V, et al. (2004) Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res 32: D497-501.

5. Newman RH, Hu J, Rho HS, Xie Z, Woodard C, et al. (2013) Construction of human activity-based phosphorylation networks. Mol Syst Biol 9: 655.

6. Hu J, Rho HS, Newman RH, Zhang J, Zhu H, et al. (2014) PhosphoNetworks: a database for human phosphorylation networks. Bioinformatics 30: 141-142.

7. Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, et al. (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res 43: D512-520.

8. Lu CT, Huang KY, Su MG, Lee TY, Bretana NA, et al. (2013) DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res 41: D295-305.

9. Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, et al. (2011) Phospho.ELM: a database of phosphorylation sites--update 2011. Nucleic Acids Res 39: D261-267.

10. Chatr-Aryamontri A, Breitkreutz BJ, Oughtred R, Boucher L, Heinicke S, et al. (2015) The BioGRID interaction database: 2015 update. Nucleic Acids Res 43: D470-478.

11. Cowley MJ, Pinese M, Kassahn KS, Waddell N, Pearson JV, et al. (2012) PINA v2.0: mining interactome modules. Nucleic Acids Res 40: D862-865.

12. Meyer MJ, Das J, Wang X, Yu H (2013) INstruct: a database of high-quality 3D structurally resolved protein interactome networks. Bioinformatics 29: 1577-1579.

13. Licata L, Briganti L, Peluso D, Perfetto L, Iannuccelli M, et al. (2012) MINT, the molecular interaction database: 2012 update. Nucleic Acids Res 40: D857-861.

14. Orchard S, Ammari M, Aranda B, Breuza L, Briganti L, et al. (2014) The MIntAct project--IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42: D358-363.

15. Breuer K, Foroushani AK, Laird MR, Chen C, Sribnaia A, et al. (2013) InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. Nucleic Acids Res 41: D1228-1233.

16. Fazekas D, Koltai M, Turei D, Modos D, Palfy M, et al. (2013) SignaLink 2 - a signaling pathway resource with multi-layered regulatory networks. BMC Syst Biol 7: 7.

17. Coordinators NR (2016) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 44: D7-19.

18. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, et al. (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. Science 347: 1257601.