A Scored Human Protein-Protein Interaction Network to Catalyze Genomic Interpretations

Taibo Li[1,4,6], Rasmus Wernersson[3,5], Rasmus Borup Hansen[3], Heiko Horn[2,4], Johnathan Mercer[2,4], Christopher Workman[5], Olga Rigina[5], Kristoffer Rapacki[5], Hans-Henrik Stærfeldt[5], Søren Brunak[5], Thomas Skøt[3], Kasper Lage[1,4]

[1]Stanley Center, [2]Broad Institute, [3]Intomics, [4]Department of Surgery, MGH, [5]Center for Biological Sequence Analysis, Technical University of Denmark, [6]Department of EECS, MIT

InWeb_InBioMap (InWeb_IM) Pipeline

InWeb_IM integrates interactions from 8 databases using stringent orthology transfer to map data from 34 organisms and provides global scores.

> Obtain data from **BioGRID, DIP, IntAct, Reactome, WikiPathways, BIND, MatrixDB**, and **NetPath**

> Stringent orthology transfer by majority voting using **eggNOGG, Ensembl, HomoloGene, InParanoid, Gene, OrthoDB, KEGG,** and **HOGENOM**

> Calculate confidence score using >68,000 publications and calibrate scores using high-confidence pathway interactions
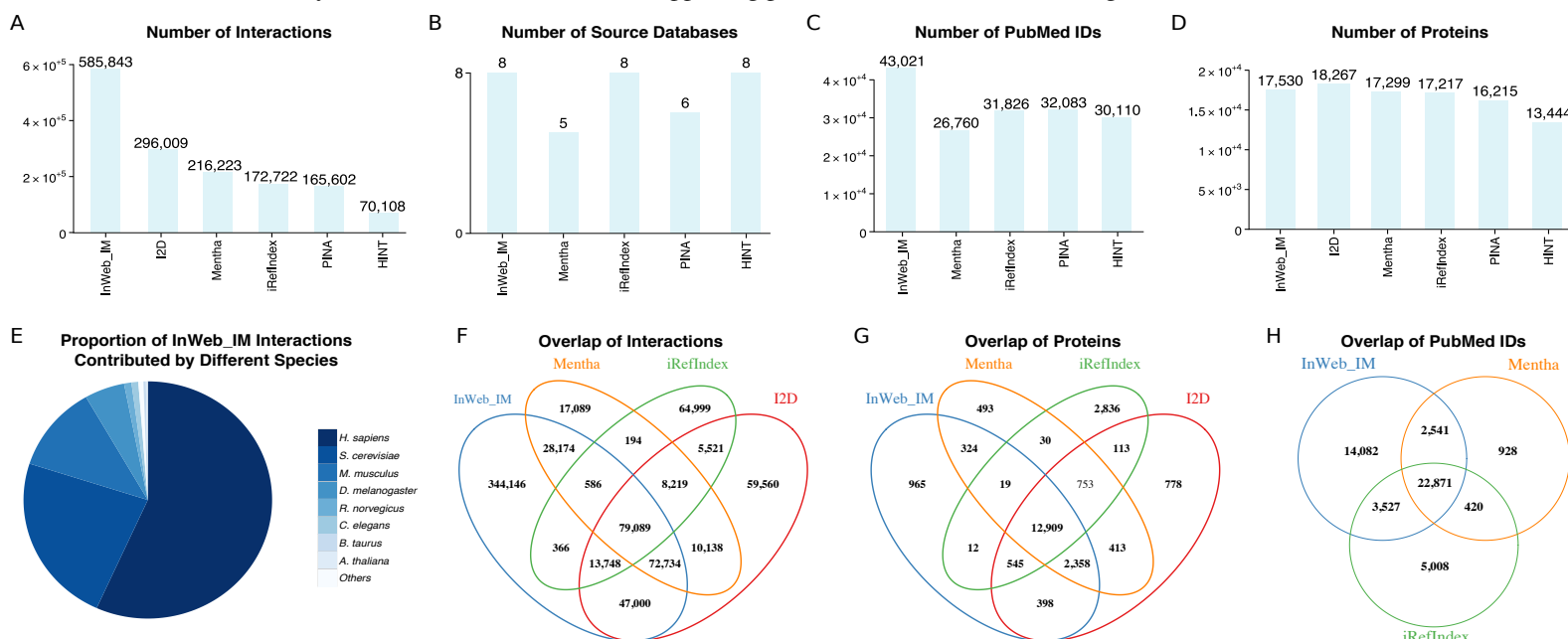
Qualitative Comparison

InWeb_IM provides the most transparent information compared to five other widely used resources.

|	Score	PMID	Source Database	Orthology Transfer
InWeb_IM	■	■	■	■
I2D			■	■
Mentha	■	■		
iRefIndex		■	■	
PINA		■	■	
HINT		■		

Quantitative Comparison

InWeb_IM has twice as many interactions and 34.1% more supporting publications than the second-largest resources, I2D and PINA.

A. Number of Interactions — InWeb_IM 585,843; I2D 296,009; Mentha 216,223; iRefIndex 172,722; PINA 165,602; HINT 70,108

B. Number of Source Databases — InWeb_IM 8; Mentha 5; iRefIndex 8; PINA 6; HINT 8

C. Number of PubMed IDs — InWeb_IM 43,021; Mentha 26,760; iRefIndex 31,826; PINA 32,083; HINT 30,110

D. Number of Proteins — InWeb_IM 17,530; I2D 18,267; Mentha 17,299; iRefIndex 17,217; PINA 16,215; HINT 13,444

E. Proportion of InWeb_IM Interactions Contributed by Different Species — H. sapiens, S. cerevisiae, M. musculus, D. melanogaster, R. norvegicus, C. elegans, B. taurus, A. thaliana, Others

F. Overlap of Interactions (InWeb_IM, Mentha, iRefIndex, I2D)

G. Overlap of Proteins (InWeb_IM, Mentha, iRefIndex, I2D)

H. Overlap of PubMed IDs (InWeb_IM, Mentha, iRefIndex)

Biological Signal of InWeb_IM

InWeb_IM is consistently better at learning pathway structures, annotating cancer driver genes, and corroborating quantitative proteomic experimental results.

MSigDB Canonical Pathways; Cancer5000 Stringent Set; Quantitative Proteomic Experiment

Networks: InWeb_IM, I2D, Mentha, iRefIndex, PINA, HINT
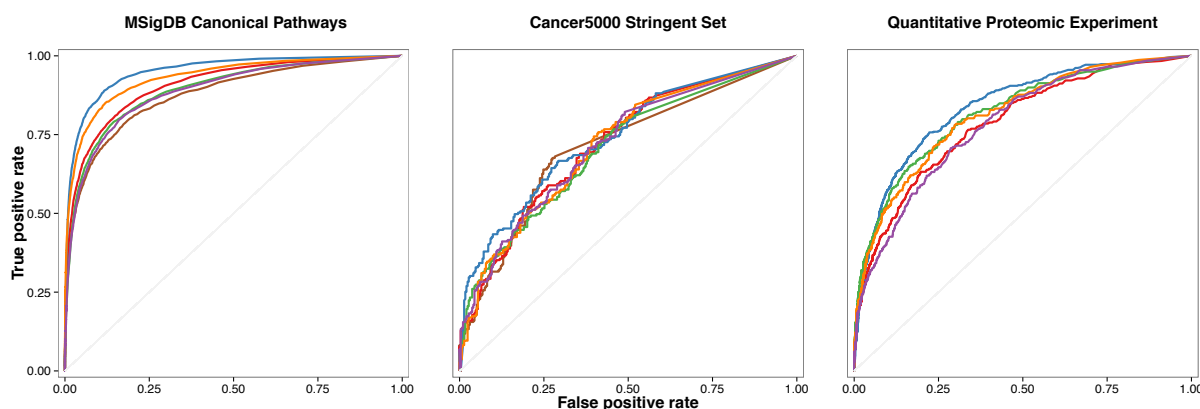
STANLEY CENTER FOR PSYCHIATRIC RESEARCH

BROAD INSTITUTE

MGH 1811 · MIT

CENTER FOR BIOLOGICAL SEQUENCE ANALYSIS CBS

intomics — from data to biology