# Systematic Discovery of Human Membrane Receptor Interactions 

# Presenter: Yanjun Qi <br> (Current: NEC Labs America) 

$$
12
$$

Qi Y, Dhiman HK, Bhola N, Budyak I, Kar S, Man D, Dutta A, Tirupula K, Carr BI, Grandis J, Bar-Joseph Z, Klein-Seetharaman J.(2009) Systematic prediction of human membrane receptor interactions. Proteomics. 9(23):5243-55.

## Protein interaction networks

- Critical for the comprehensive understanding of the cell
- Goal:
- To decipher the edges according to "real" interactions, i.e. direct physical contacts


Figure 3-83 Molecular Biology of the Cell 5/e (0 Garland Science 2008)

## Focus on Human Membrane Receptors



Signal Transduction Cascades

## From Membrane Receptors... <br> ....to Cellular Communication Mechanisms



## PPI Predictions for Human Membrane

 Receptors- A combined approach
- Binary classification
step 2:
predictions for all receptors
step 1:
feature extraction
step 3:
- Global analysiseceptor interactome
- Biological feedback \& validation identification
step 4:
interaction experiments



## Step 1: Feature Evidences

- High-throughput direct data
- yeast-2-hybrid, mass spectrometry
- Indirect data
- gene expression, protein-DNA binding, ...
- functional annotation data: Gene ontology annotation, ...
- sequence based data sources: Domain information, homology based protein protein interactions, ...



## Use implicit and direct data as evidence not as proof for an interaction

## Step 2: Supervised Classification

1. Describe the human protein pair with numeric features
2. Learn a function that maps a feature vector into one of the two classes
$(),)=\left[\right.$ feat $_{1}$, feat $_{2}, \ldots$. feat $\left._{m}\right]$
feat ${ }_{1}=$ Location similarity feat $_{2}=$ Functional similarity
...


## Step 2: Supervised Binary Classification

- To classify each protein pair:

Treat as a binary classification task
Target function: interacts or not?

- Feature Set
- Feature are heterogeneous
- Most features are noisy
- Most features have missing values
- Reference Set:
, Use the small positive set as positive training
- No negative (not interacting) set available
- Highly skewed class distribution
- Much more non-interacting pairs than interacting pairs


## Step 2: Random Forest Classifier

- A collection of independent decision trees (ensemble classifier) learning from the features
- Each tree is grown on a bootstrap sample of the training set
- Advantages of the Random Forest:
» - can handle heterogeneous features
-     - Is not as much affected by noisy features

म - Is not as much affected by correlated features
> - can estimate features with missing values

## Step 2: Random Forest Classifier



## Step 2: Classification Comparison



- Compare Classifiers
( 27 features extracted from 8 different data sources, modified with biological feedbacks)

- Receptor (subfamily) PPI task to general Human PPI prediction task


## Step 3: Global Analysis



- Degree distribution / Hub analysis
- Graph module (from bi-clustering study)
- Family based graph patterns (receptors / receptors subclasses / ligands / etc )


## Step 3: Receptor Hubs



## Step 3: Receptor-Receptor Interactions



Differences in signaling crosstalk mechanism

## Step 3: Receptor-Ligand Interactions


vs. HPRD


ISMB10-Highlight

Step 3: Overlap between our predictions (HMRI) and other methods

## PPI Data Set

Predicted HMRI 9144
HPRD
RhodeBioTech05[O1] 257
ScottBMCPPI07[O2] 505 STRING08[O3] 220
TAP-MSB07[O5] 3
EGFR-nature06 [O6] 50
(Four ERBB)
Lumier05 [O7]
2

## Step 4: Experimental Validation

- Three of our predictions (of EGFR) were chosen for experimentally tests
- EGFR with HCK (pull-down assay)
- EGFR with Dynamin-2 (pull-down assay + functional assay)
- EGFR with TGF-beta1(pull-down assay)
- Experiments @ U.Pitt School of Medicine


## Step 4: Epidermal Growth Factor Receptor Predictions



# EGFR has a total of 91 validated partners in HPRD (2007v); Choose among our top 200 predictions for EGFR 

## Step 4: EGFR Validation: Dynamin2

## Dnm2 Domain Structure Pull-Down Experiment



## Function:

## Receptor internalization



## Step 4: EGFR Validation: Hck

## Hck Domain Structure



Hck Functions:

- Binds and regulates Nef during HIV infection
- Function in signal transduction, but not well defined


## Pull-Down Experiment



Anti-His-blot


## Step 4: Functional Experiments



## Step 4: Summary of Validation Steps


F. Functional assay


## Web Server: HMRI

Address eethttp://flan.blm.cs.cmu.edui/MMRI/index.jsp
$\square \square$

Four Ways to Query our Human Receptor Interactome Database

Services:

## Check query protein (ClickToUse)

1
This senvice would perform a check on the input protein to see if the query is a receptor protein or not. If yes, we further check if the input belongs to the GPCR family or not. Please input the query as either NCBI Entrez gene ID or gene Name.
Query Database

Getting Help

## Search interaction partners for the Query protein (ClickToUse)

2 This senvice would search on the whole predicted receptor interactome to find interaction partners for the query. Another constraint is that the returned pairs C interaction scores are higher than an input score threshold. Please input the query protein with either NCBI Entrez gene ID or gene Name.

## Search interaction pairs among the set of input proteins (ClickToUse)

3
This senvice would search the potential pairs among a group of input proteins. The returned pairs a interaction scores are higher than an input score threshold. C The input proteins list use either NCBI Entrez gene ID or gene Name.

Check predicted scores for a pair of input proteins (ClickToUse)
4
 check if this pair existed in $\operatorname{HPRD}$ (2006 version) or not. I In addition, we also provide the features of this requested pair.

## Predictions/Code Downloads

www.cs.cmu.edu/~qyj/HMRI

- Download predictions for each receptor (each having its own predicted interaction score file)
- Includes HPRD label, related features, gene description, genetic disorder information
- Download family related subgraphs
- Download multiple kinds of hub protein lists
- Software Download (both source-code and runnable versions provided)


## Acknowledgements

## NSF \& NIH Grants

## CMU Computer Science

- Ziv Bar-Joseph
- Christos Faloutsos
- Jaime Carbonell
- System Biology Teammates
- BLM Teammates
- All the other collaborators

UPitt School of Medicine

- Judith Klein-Seetharaman
- Harpreet Dhiman
- Ivan Budyak
- Eric E. Gardner
- Oznur Tastan
- Neil Bhola
- All the other collaborators



## Step 4: Experimental Validation





D. EGFR/TGF- $\beta 1$

II. $\quad$ C225 ( $6 \mu \mathrm{~g} / \mathrm{ml}$ )

NoTx EGF TGFB1 NoTx EGF TGFB1
III.

ALK4 $(5 \mu \mathrm{M})$
NoTx EGF TGFs1 NoTx EGF TGFß1
Phospho-
MAPK
MAPK


## References

- O1. Rhodes DR, Tomlins SA, Varambally S, Mahavisno V, Barrette T, KalyanaSundaram S, Ghosh D, Pandey A, Chinnaiyan AM: Probabilistic model of the human protein-protein interaction network. Nat Biotechnol 2005, 8:951-959.
- O2. Scott MS, Barton GJ: Probabilistic prediction and ranking of human protein-protein interactions. BMC Bioinformatics 2007, 8:239.
- O3. von Mering C, Jensen LJ, Kuhn M, Chaffron S, Doerks T, Kruger B, Snel B, Bork P: STRING 7--recent developments in the integration and prediction of protein interactions. Nucleic Acids Res 2007, 35(Database issue):D358-362.
- O4. Ramani AK, Li Z, Hart GT, Carlson MW, Boutz DR, Marcotte EM: A map of human protein interactions derived from co-expression of human mRNAs and their orthologs. Mol Syst Biol 2008, 4:180.
- O5. Ewing RM, Chu P, Elisma F, Li H, Taylor P, Climie S, McBroom-Cerajewski L, Robinson MD, O'Connor L, Li M et al: Large-scale mapping of human proteinprotein interactions by mass spectrometry. Mol Syst Biol 2007, 3:89.
- O6. Jones RB, Gordus A, Krall JA, MacBeath G: A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 2006, 439(7073):168-174.
- O7. Barrios-Rodiles M, Brown KR, Ozdamar B, Bose R, Liu Z, Donovan RS, Shinjo F, Liu Y, Dembowy J, Taylor IW et al: High-throughput mapping of a dynamic signaling network in mammalian cells. Science 2005, 307(5715):1621-1625.
- O8. Miller JP, Lo RS, Ben-Hur A, Desmarais C, Stagljar I, Noble WS, Fields S: Large-scale identification of yeast integral membrane protein interactions. Proc Natl Acad Sci U S A 2005, 102(34):12123-12128.

