BioPlex protein-protein interaction networks

Ludwig Geistlinger

Robert Gentleman

Center for Computational Biomedicine

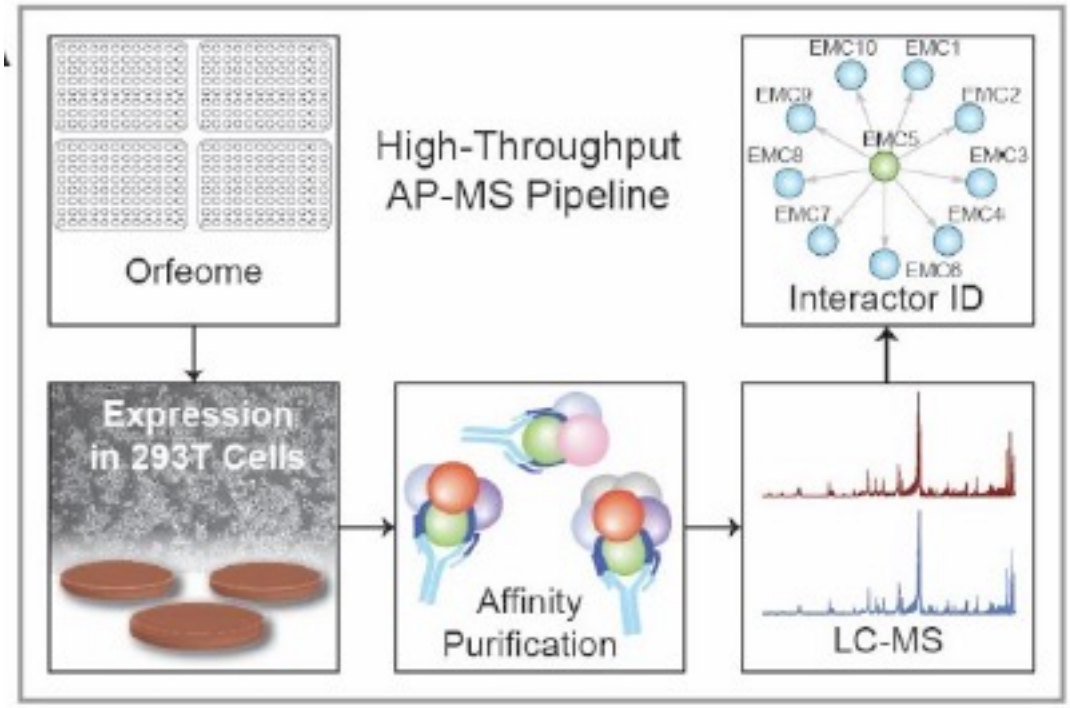
Harvard Medical School

Goal

- to find groups of interested participants to work on a coordinated analysis of this data in one area
- Vince, Laurant, Robert happy to help organize and explain the data and how to access it
- Several vignettes produced by Ludwig Geistlinger and Roger Vargas that can be jumping off points
- Spend 4+ hours investigating the data and making some progress – ideally contributing to the vignettes so others can build off of them
- What you get:
  - chance to work across modalities
  - learn to work in a group
  - learn some interesting tools

Experimental procedures

- uses affinity purification–mass spectrometry methodology
- protein interaction networks of more than 70% of protein-coding genes from the human genome

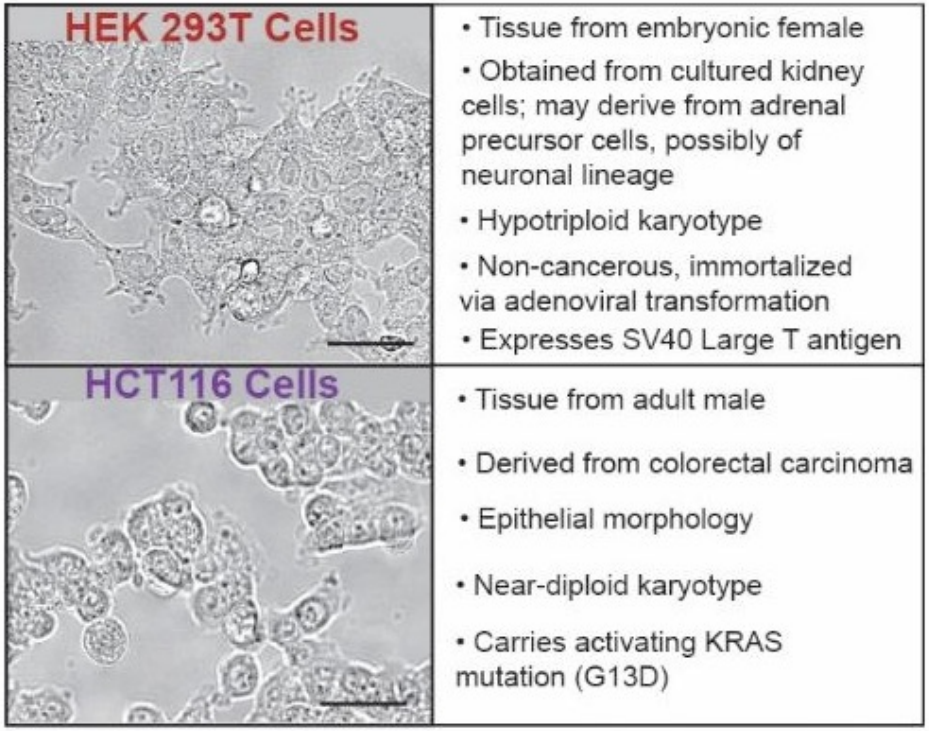


Relevant literature

- Huttlin et al, *Cell*, 2015: Bioplex 1.0 (293T cells)
- Huttlin et al, *Nature*, 2017: Bioplex 2.0 (293T cells)
- Huttlin et al, *Cell*, 2021: Bioplex 3.0 (293T cells & HCT116 cells)

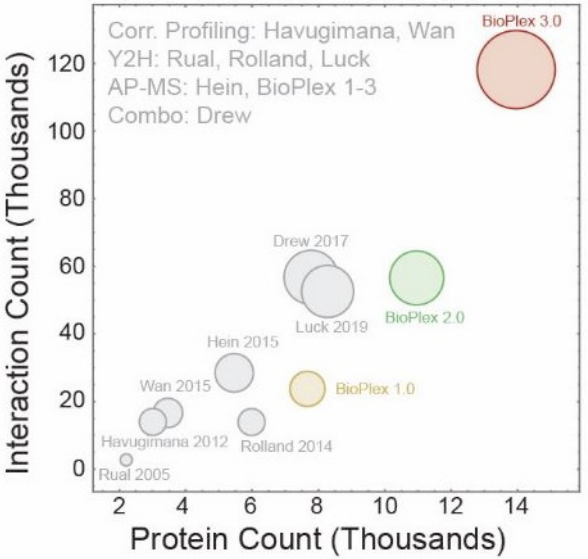
Cell lines

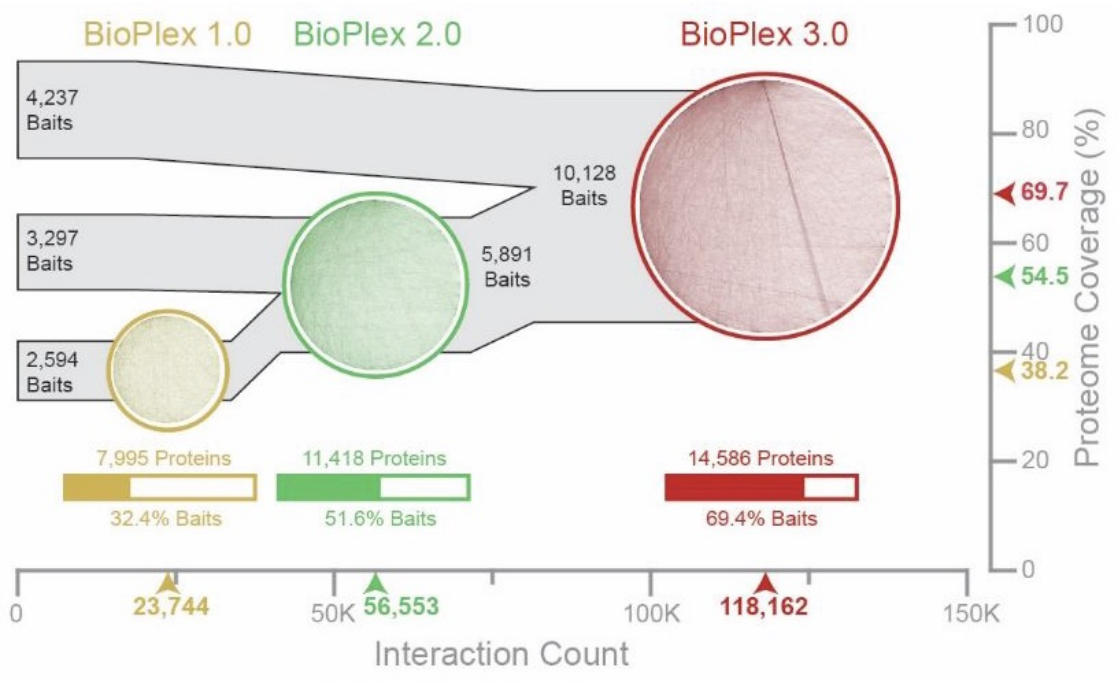
- **293T**: highly transfectable derivative of Human Embryonic Kidney 293 (HEK293) cell line
- **HCT116**: Human Colon Tumor-derived cell line initiated from an adult male



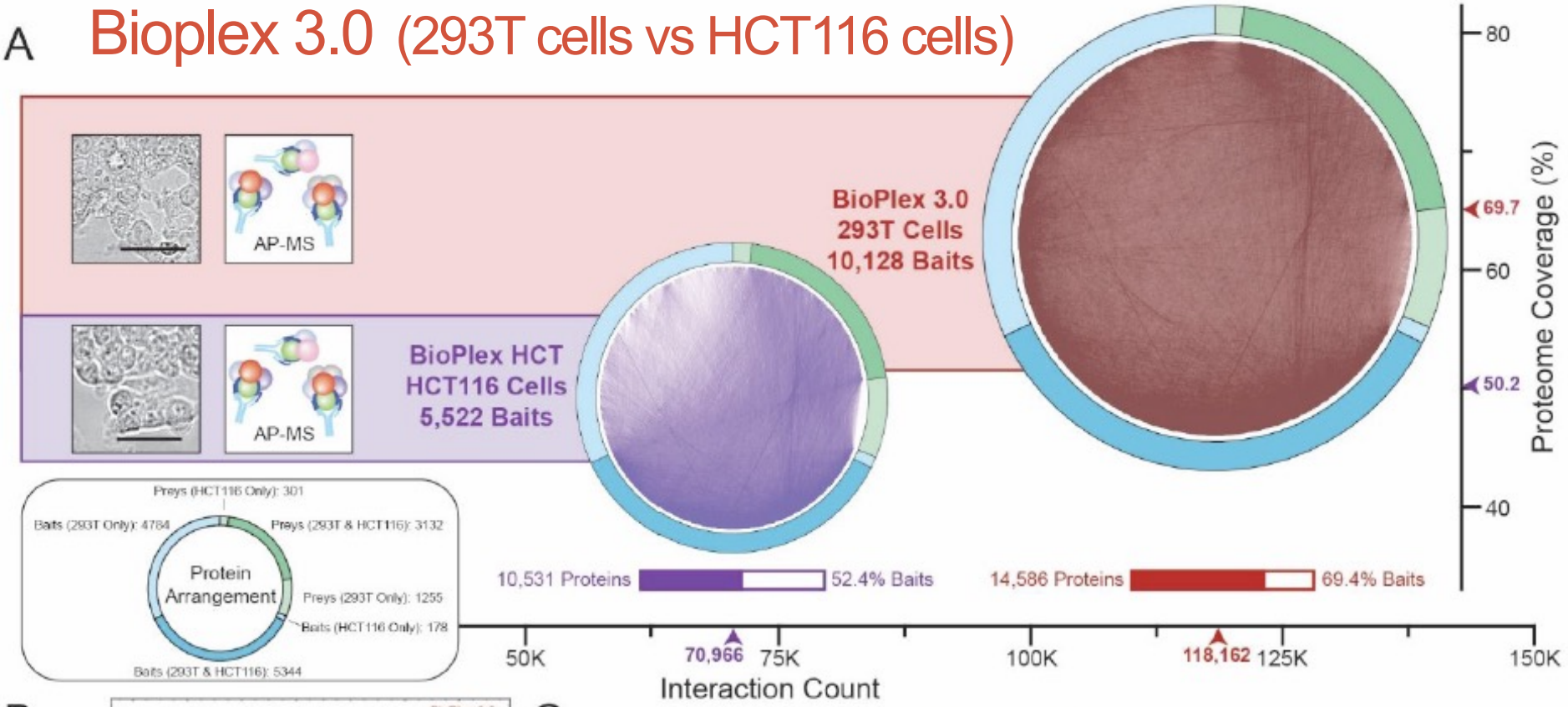
Bioplex PPIs (293T cells)

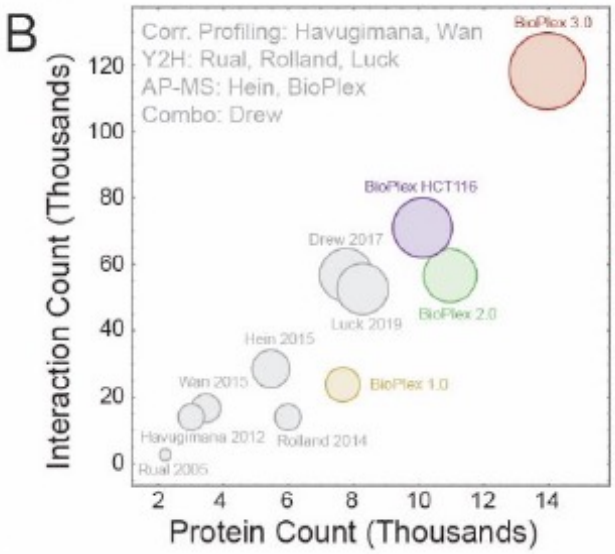
Version	Year	#Proteins	#Interactions
1.0	2015	8k	24k
2.0	2017	11k	57k
3.0	2021	15k	120k

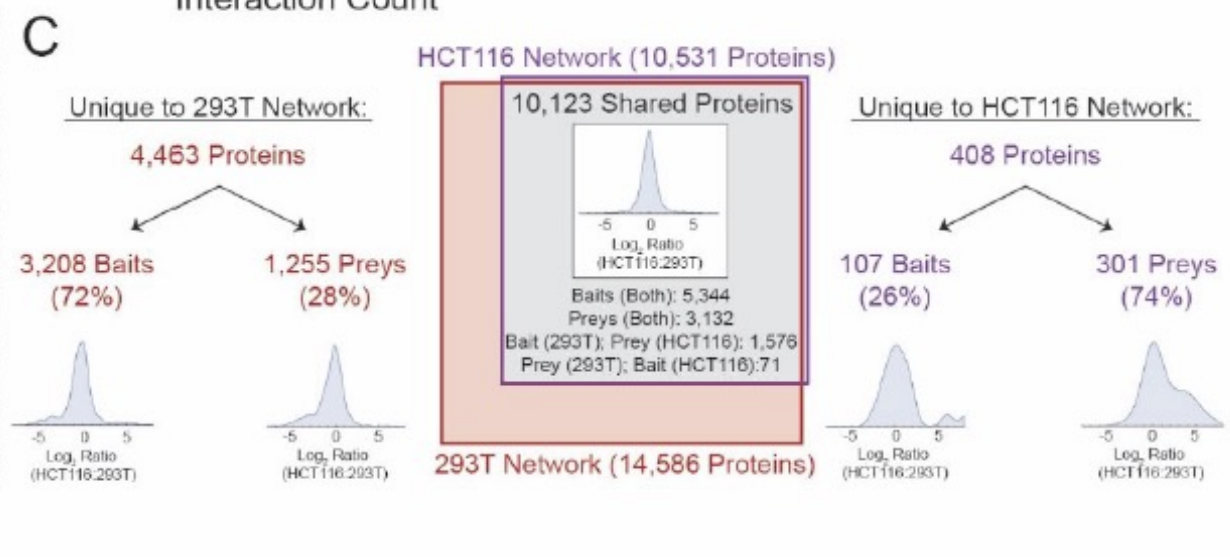




A Bioplex 3.0 (293T cells vs HCT116 cells)







Agreement with CORUM protein complexes

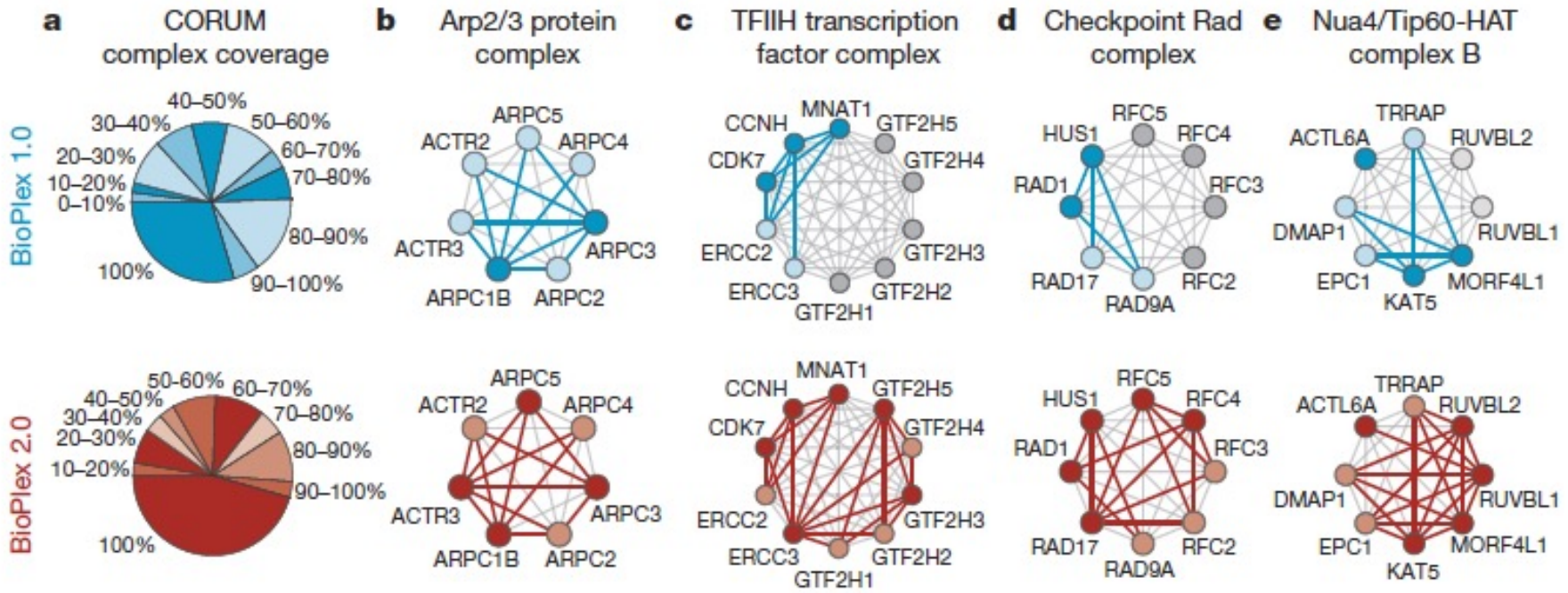


Figure 2 | BioPlex 2.0 maps protein complexes with increased resolution. **a**, Agreement among BioPlex networks and CORUM complexes. Pie charts indicate the fraction of CORUM complexes that attained the indicated protein coverage. Compared with BioPlex 1.0 (blue), BioPlex 2.0 (red) provides substantially improved coverage. **b–e**, Network coverage achieved by BioPlex 1.0 (blue) and BioPlex 2.0 (red) for selected CORUM complexes. Dark and light shades depict bait and prey proteins, respectively, while grey proteins were not observed in the network. Red and blue edges represent detected protein interactions.

Data resources

1. BioPlex PPI networks (293T & HCT116)
2. CORUM protein complexes
3. PFAM protein domains
4. Transcriptome data (293T & HCT116)
5. Proteome data (293T & HCT11

BioPlex – Bioconductor Package
(https://bioconductor.org/packages/release/bioc/html/BioPlex.html)
https://ccb-hms.github.io/BioPlexAnalysis/articles/BioNet.html

https://github.com/ccb-hms/BioPlex

https://ccb-hms.github.io/BioPlex/articles/BioPlex.html

BioPlex PPI networks (Huttlin et al., *Cell*, 2021)

- 293T cells:
  - Versions: 1.0, 2.0, 3.0
  - 120k interactions, 15k proteins
- HCT116 cells:
  - Versions: 1.0
  - 70k interactions, 10k proteins
- data.frame, graphNEL
- Challenge: GraphFrames backend

Transcriptome data

- 293T cells:
  - GSE122425 (Sun et al., *Epigenomics*, 2019)
    - 3 WT samples & 3 *NSUN2*-KO samples
    - Raw RNA-seq read counts and RPKMs
- HCT116 cells:
  - Genentech (Klijn et al., *Nat Biotechnol*, 2015)
    - 675 cancer cell lines (incl. HCT116); raw read counts
  - Cancer Cell Line Encyclopedia (CCLE, Ghandi et al. *Nature*, 2019)
    - 934 cancer cell lines (incl. HCT116); raw read counts
- Alignment / splicing → challenge / project

Proteome data

- HCT116 cells:
  - Cancer Cell Line Encyclopedia (CCLE, Nusinov et al., *Cell*, 2020)
  - 375 cancer cell lines (incl HCT116)
  - Normalized log2 expression levels of 12,755 proteins (MS)

- 293T cells & HCT116 cells:
  - Part of BioPlex 3.0 (Huttlin et al., *Cell*, 2021)
  - 5 samples each for both cell lines
  - Relative abundance of 9,604 proteins (sums to 1 across samples)

- Challenge: transcriptome/proteome data integration

Data checks

1. Extract BioPlex PPIs for a CORUM complex
2. Identify interacting PFAM domains
3. Expressed genes showing up as prey
4. Relationship: prey frequency vs. prey expression

…

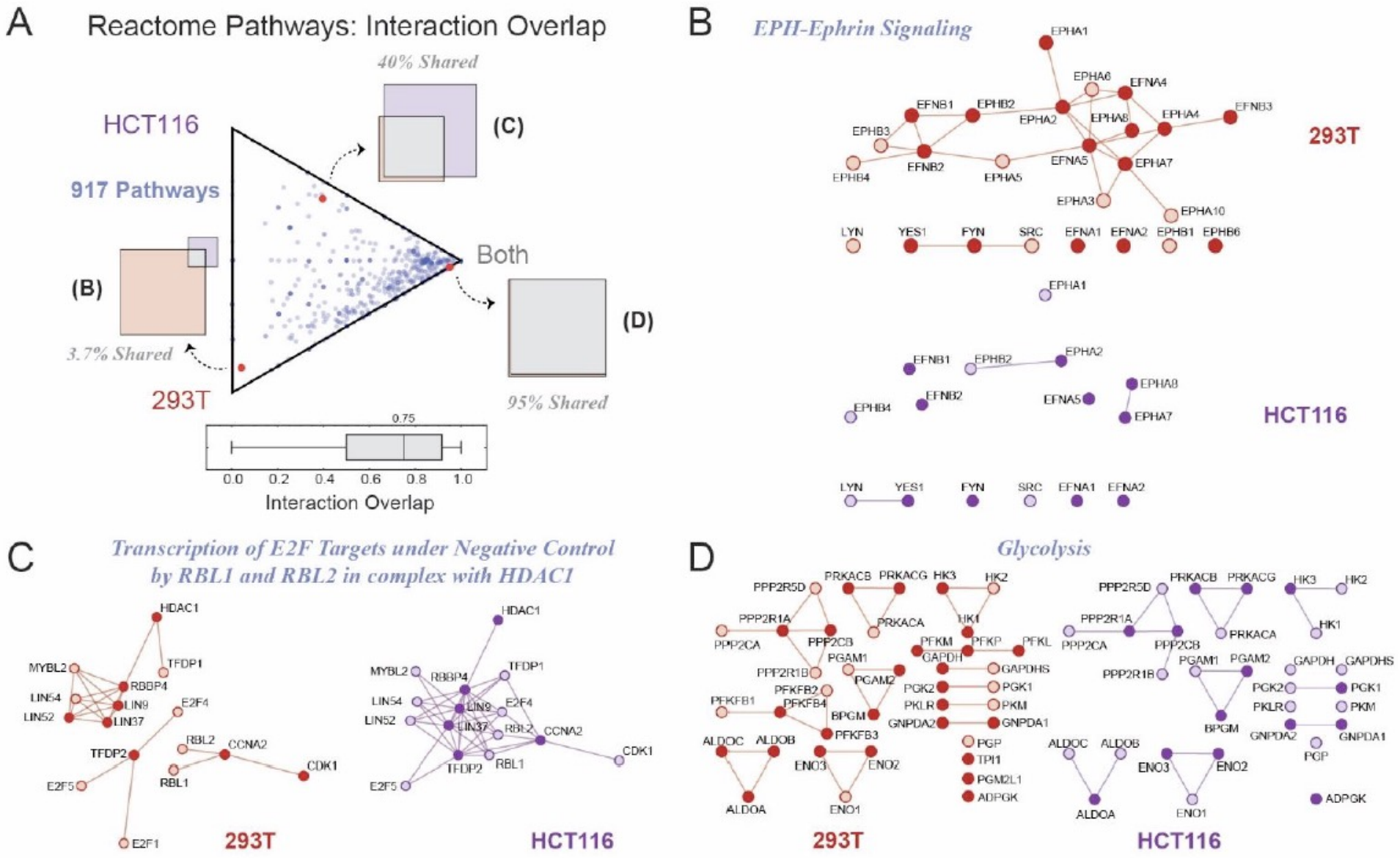
https://ccb-hms.github.io/BioPlex/articles/BasicChecks.html

Applications

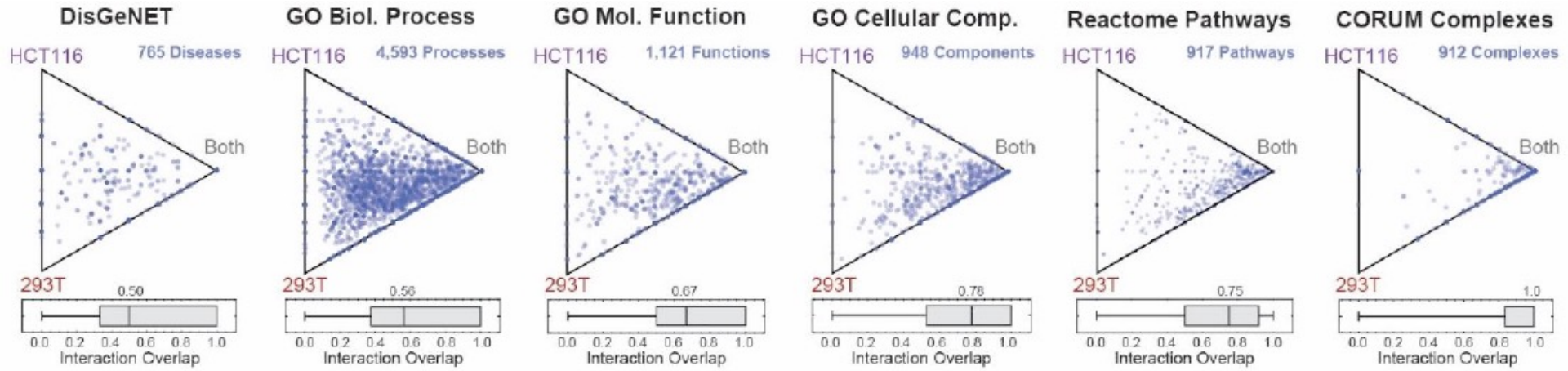
1. Maximum scoring subnetwork analysis
2. PFAM domain-domain association analysis
3. Transcriptome / proteome data integration (Challenge)

https://github.com/ccb-hms/BioPlexAnalysis

Overlap with gene set DBs (GO, Reactome, DisGeNET, …)



Overlap with gene set DBs (GO, Reactome, DisGeNET, …)



Less preserved → More preserved