# Reconstruction of human protein-coding gene functional association network based on machine learning Supplementary Figures and Tables
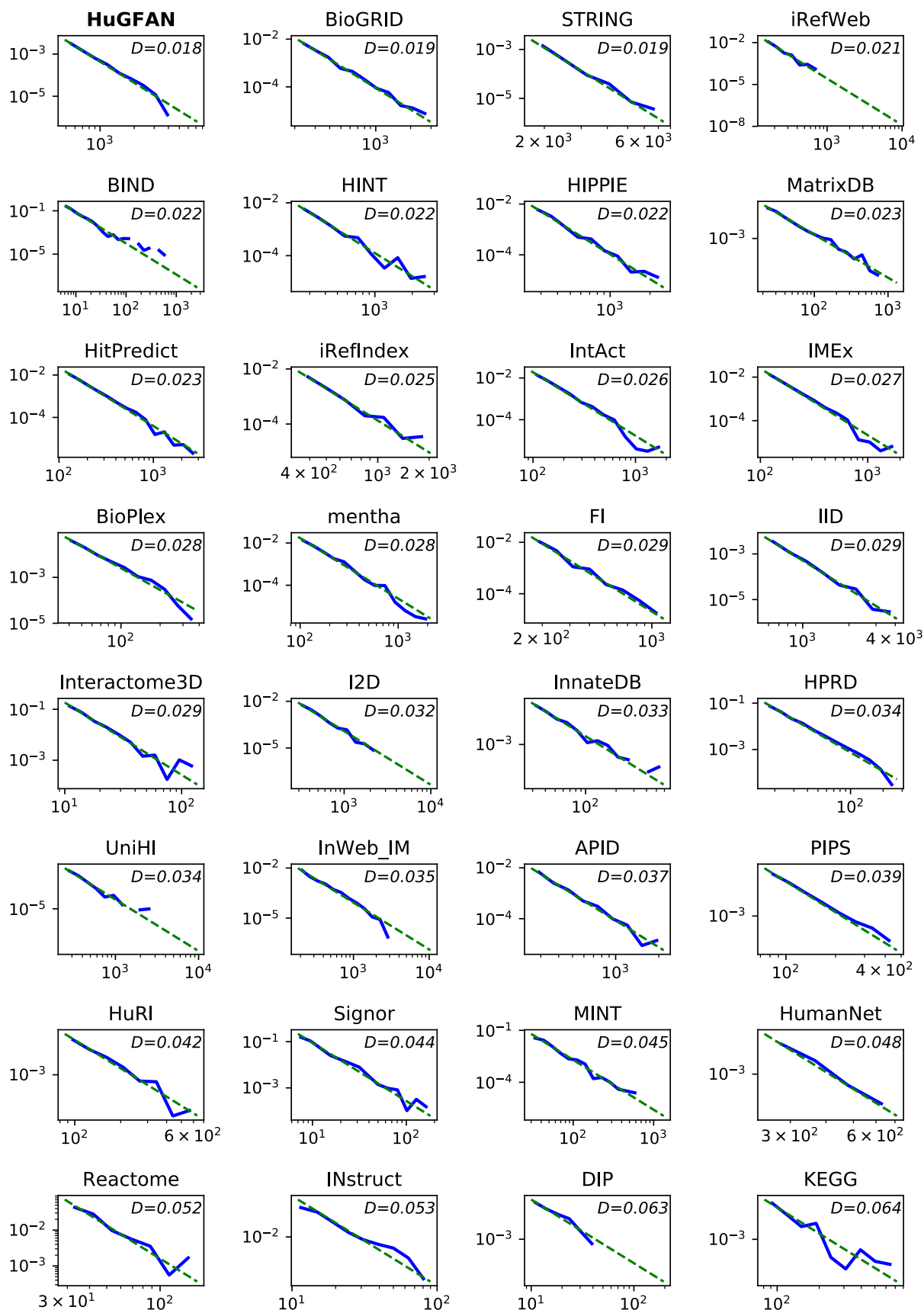
HUANG Xiaotai*, Songwei Jia and Lin Gao*

Figure S1: The power law fitting of HuGFAN and other networks. Ranked by Kolmogorov-Smirnov distance (D).
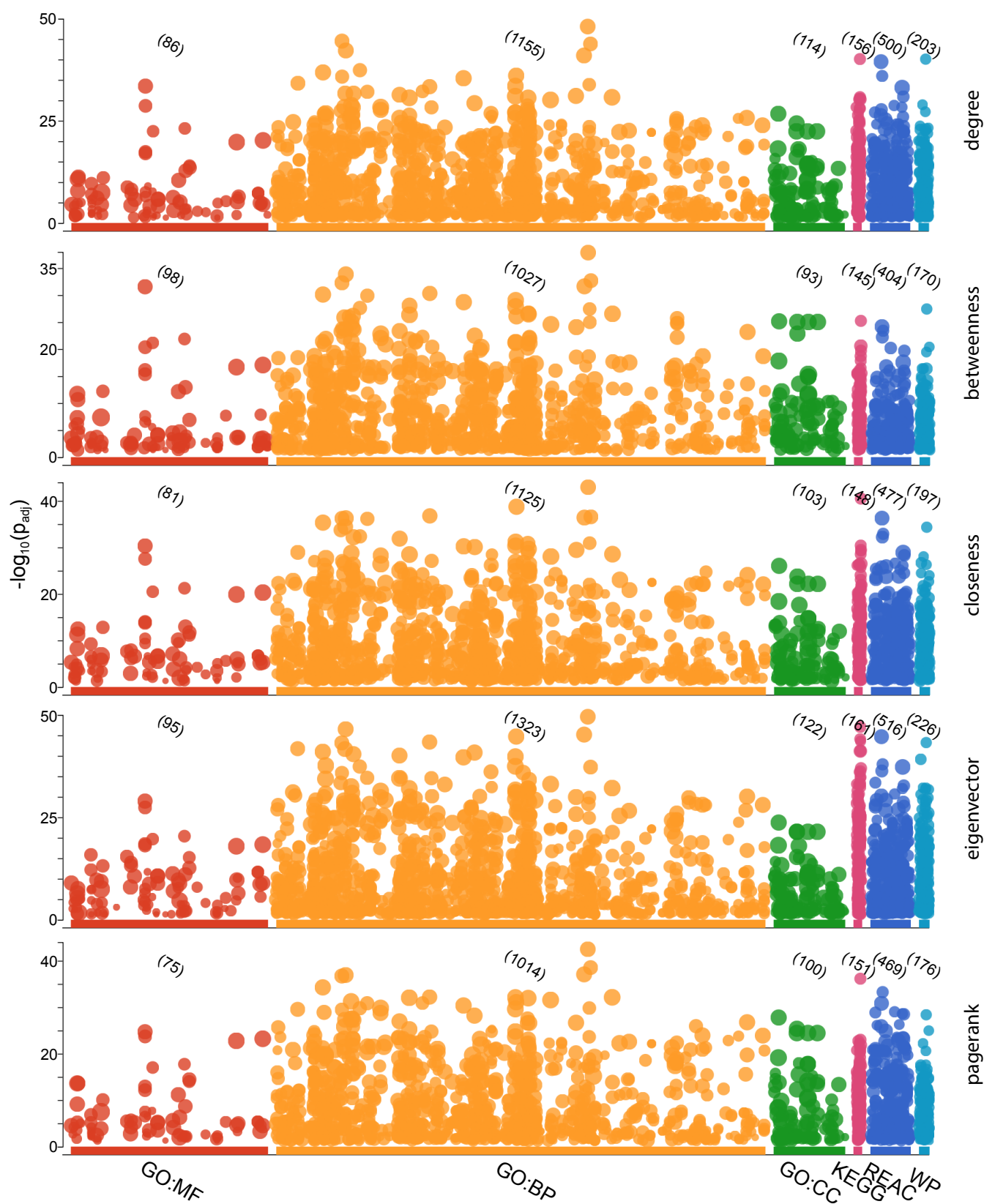
Figure S2: The enrichment results of GO terms and pathways for HuGFAN top 100 genes in rankings of the five centralities.

Table S1: Network topology metrics comparison between HuGFAN and other networks.

| Networks | Avg. CC | Avg. Degree | Avg. Betweenness | Avg. Closeness | Avg. Eigenvector | Avg. Pagerank |
|---|---|---|---|---|---|---|
| STRING | 0.203 | 594.173 | 9669.360 | 0.491 | 0.088 | 5.39E-05 |
| **HuGFAN** | 0.296 | 116.315 | 16968.672 | 0.380 | 0.031 | 4.91E-05 |
| IID | 0.243 | 106.690 | 15784.024 | 0.372 | 0.039 | 5.46E-05 |
| InWeb_IM | 0.317 | 74.086 | 14235.404 | 0.389 | 0.025 | 5.68E-05 |
| HumanNet | 0.287 | 58.590 | 18702.214 | 0.328 | 0.020 | 5.58E-05 |
| HINT | 0.150 | 51.588 | 16413.065 | 0.352 | 0.032 | 5.71E-05 |
| HitPredict | 0.118 | 47.435 | 16494.430 | 0.347 | 0.025 | 5.81E-05 |
| BioGRID | 0.129 | 45.907 | 17893.537 | 0.347 | 0.024 | 5.35E-05 |
| iRefIndex | 0.185 | 45.366 | 17677.007 | 0.334 | 0.020 | 5.74E-05 |
| HIPPIE | 0.133 | 42.254 | 17553.609 | 0.343 | 0.030 | 5.57E-05 |
| UniHI | 0.328 | 41.192 | 15226.191 | 0.375 | 0.014 | 5.65E-05 |
| mentha | 0.109 | 40.305 | 17454.641 | 0.334 | 0.026 | 5.80E-05 |
| APID | 0.125 | 39.441 | 17310.106 | 0.332 | 0.027 | 5.93E-05 |
| FI | 0.422 | 38.021 | 15772.794 | 0.309 | 0.023 | 7.39E-05 |
| I2D | 0.254 | 35.390 | 13429.346 | 0.386 | 0.014 | 6.11E-05 |
| IntAct | 0.098 | 30.455 | 18397.342 | 0.319 | 0.020 | 5.91E-05 |
| IMEx | 0.093 | 30.176 | 18276.813 | 0.320 | 0.020 | 5.91E-05 |
| iRefWeb | 0.254 | 20.678 | 13159.471 | 0.370 | 0.010 | 6.74E-05 |
| PIPS | 0.341 | 19.807 | 8636.076 | 0.285 | 0.029 | 1.55E-04 |
| BioPlex | 0.112 | 16.914 | 19757.916 | 0.265 | 0.012 | 7.17E-05 |
| KEGG | 0.271 | 16.909 | 9535.088 | 0.249 | 0.023 | 1.72E-04 |
| HuRI | 0.082 | 13.946 | 11867.599 | 0.280 | 0.024 | 1.12E-04 |
| MatrixDB | 0.137 | 13.647 | 14383.743 | 0.290 | 0.017 | 8.76E-05 |
| Reactome | 0.600 | 9.408 | 5334.450 | 0.350 | 0.019 | 2.19E-04 |
| HPRD | 0.136 | 7.761 | 14483.688 | 0.258 | 0.022 | 1.06E-04 |
| MINT | 0.154 | 6.853 | 11402.708 | 0.264 | 0.009 | 1.28E-04 |
| Signor | 0.105 | 5.345 | 6416.482 | 0.272 | 0.025 | 2.30E-04 |
| InnateDB | 0.185 | 4.759 | 8088.615 | 0.263 | 0.021 | 1.81E-04 |
| Interactome3D | 0.270 | 4.672 | 10387.688 | 0.296 | 0.014 | 1.70E-04 |
| BIND | 0.288 | 4.571 | 7772.695 | 0.324 | 0.013 | 1.41E-04 |
| DIP | 0.283 | 3.269 | 4946.810 | 0.290 | 0.008 | 3.29E-04 |
| INstruct | 0.197 | 3.260 | 2876.147 | 0.389 | 0.014 | 3.82E-04 |

*Note:* Abbreviations used in the table are: **CC**: Clustering Coefficient, **Avg.**: average.

Table S2: Prediction performance comparison between HuGFAN and other networks with DriverNet.

| Networks | GBM | | | MT | | | OV | | | TN | | | Avg. |
| | p-value | ACC | Rank | p-value | ACC | Rank | p-value | ACC | Rank | p-value | ACC | Rank | Rank |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **HuGFAN** | 0.001252 | 0.788 | **1** | 0.049148 | 0.779 | **1** | 0.000844 | 0.808 | **1** | 0.000859 | 0.911 | 2 | **1.25** |
| IID | 0.003487 | 0.688 | 2 | 0.123308 | 0.759 | 2 | 0.127765 | 0.749 | 8 | 0.024179 | 0.898 | 4 | 4 |
| STRING | 0.028464 | 0.679 | 3 | 0.305066 | 0.358 | 14 | 0.005747 | 0.781 | 2 | 0.025183 | 0.903 | 3 | 5.5 |
| InWeb_IM | 0.010117 | 0.654 | 5 | 0.571521 | 0.719 | 3 | 0.040270 | 0.764 | 3 | 0.364030 | 0.893 | 12 | 5.75 |
| HINT | 0.107288 | 0.627 | 9 | 0.519966 | 0.700 | 4 | 0.128572 | 0.702 | 14 | 0.057298 | 0.908 | 5 | 8 |
| FI | 0.013890 | 0.661 | 4 | 0.830629 | 0.640 | 9 | 0.012889 | 0.744 | 4 | 0.432827 | 0.863 | 17 | 8.5 |
| BioGRID | 0.027654 | 0.617 | 6 | 0.845064 | 0.509 | 10 | 0.357777 | 0.694 | 15 | 0.075735 | 0.897 | 8 | 9.75 |
| HitPredict | 0.095833 | 0.609 | 10 | 0.741535 | 0.672 | 5 | 0.376187 | 0.677 | 17 | 0.097226 | 0.890 | 13 | 11.25 |
| HumanNet | 0.179893 | 0.510 | 16 | 1 | 0.094 | 23 | 0.003997 | 0.723 | 5 | 0.000357 | 0.912 | **1** | 11.25 |
| APID | 0.088798 | 0.511 | 15 | 0.935337 | 0.643 | 8 | 0.597470 | 0.673 | 18 | 0.196577 | 0.904 | 6 | 11.75 |
| UniHI | 0.142188 | 0.593 | 12 | 0.611380 | 0.177 | 17 | 0.484708 | 0.729 | 11 | 0.302609 | 0.896 | 9 | 12.25 |
| I2D | 0.186039 | 0.523 | 14 | 0.644045 | 0.200 | 16 | 0.485799 | 0.738 | 9 | 0.068290 | 0.882 | 15 | 13.5 |
| iRefIndex | 0.185131 | 0.600 | 11 | 0.728953 | 0.148 | 18 | 0.054469 | 0.736 | 10 | 0.199221 | 0.874 | 16 | 13.75 |
| HIPPIE | 0.507726 | 0.469 | 19 | 0.800213 | 0.670 | 6 | 0.384832 | 0.657 | 21 | 0.094043 | 0.895 | 10 | 14 |
| INstruct | 1 | 0.667 | 8 | 1 | 0.250 | 15 | 0.500000 | 0.750 | 7 | 1 | 0.500 | 26 | 14 |
| BIND | 1 | 0.500 | 17 | 1 | 0.667 | 7 | 1 | 0.692 | 16 | 0.333333 | 0.833 | 19 | 14.75 |
| mentha | 0.153352 | 0.488 | 18 | 0.621680 | 0.429 | 12 | 0.226005 | 0.662 | 19 | 0.079393 | 0.894 | 11 | 15 |
| DIP | 1 | 0.667 | 8 | 1 | 0.500 | 11 | 0.700000 | 0.600 | 22 | 1 | 0.750 | 23 | 16 |
| iRefWeb | 0.211888 | 0.500 | 17 | 1 | 0.097 | 22 | 0.339559 | 0.724 | 12 | 0.212721 | 0.861 | 18 | 17.25 |
| IntAct | 0.456140 | 0.421 | 20 | 1 | 0.139 | 20 | 0.408214 | 0.571 | 23 | 0.114678 | 0.903 | 7 | 17.5 |
| KEGG | 1 | 0.250 | 25 | 0.752632 | 0.381 | 13 | 0.093255 | 0.707 | 13 | 0.075475 | 0.789 | 21 | 18 |
| PIPS | 1 | 0.364 | 22 | 1 | 0.125 | 21 | 0.033214 | 0.717 | 6 | 0.302150 | 0.692 | 25 | 18.5 |
| IMEx | 1 | 0.350 | 23 | 1 | 0.147 | 19 | 0.867323 | 0.478 | 25 | 0.165607 | 0.888 | 14 | 20.25 |
| Interactome3D | 1 | 0.714 | 7 | 1 | 0 | 25 | 1 | 0.286 | 29 | 0.109091 | 0.818 | 20 | 20.25 |
| MatrixDB | 1 | 0.400 | 21 | 1 | 0.091 | 24 | 0.577421 | 0.659 | 20 | 0.437835 | 0.833 | 19 | 21 |
| HuRI | 1 | 0.571 | 13 | 1 | 0 | 25 | 0.821429 | 0.5 | 24 | 0.450000 | 0.750 | 23 | 21.25 |
| HPRD | 1 | 0.500 | 17 | 1 | 0 | 25 | 0.761206 | 0.435 | 27 | 1 | 0.773 | 22 | 22.75 |
| MINT | 1 | 0.500 | 17 | 1 | 0 | 25 | 1 | 0.231 | 30 | 1 | 0.700 | 24 | 24 |
| BioPlex | 1 | 0.143 | 26 | 1 | 0.200 | 16 | 0.828431 | 0.333 | 28 | 0.961539 | 0.231 | 30 | 25 |
| Signor | 1 | 0.400 | 21 | 1 | 0 | 25 | 0.803030 | 0.455 | 26 | 1 | 0.286 | 29 | 25.25 |
| Reactome | 1 | 0.333 | 24 | 1 | 0 | 25 | 1 | 0.111 | 32 | 0.892857 | 0.375 | 28 | 27.25 |
| InnateDB | 1 | 0 | 27 | 1 | 0 | 25 | 1 | 0.2 | 31 | 1 | 0.400 | 27 | 27.5 |

*Note:* Abbreviations used in the table are: **GBM**: glioblastoma, **TN**: triple negative breast cancers, **MT**: breast tumors, **OV**: serous ovarian cancers.

Table S3: Prediction performance comparison between HuGFAN and other networks with HotNet2.

| Networks | # Subnetworks | Max_sub | # Genes | # CGC genes | Pan-cancer enrich. | | KEGG cancer enrich. | | Avg. Rank |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | # Pan-cancer_sub | Rank | # Cancer_path | Rank | |
| **HuGFAN** | 29 | 246 | 389 | 94 | 13 | **1** | 17 | **1** | **1** |
| KEGG | 36 | 111 | 264 | 86 | 6 | 6 | 12 | 3 | 4.5 |
| HPRD | 33 | 19 | 140 | 48 | 10 | 2 | 7 | 7 | 4.5 |
| DIP | 17 | 25 | 103 | 52 | 5 | 7 | 12 | 3 | 5 |
| BioGRID | 29 | 95 | 200 | 58 | 8 | 4 | 8 | 6 | 5 |
| BioPlex | 118 | 27 | 541 | 96 | 4 | 8 | 12 | 3 | 5.5 |
| Interactome3D | 53 | 73 | 333 | 98 | 5 | 7 | 10 | 4 | 5.5 |
| HitPredict | 31 | 76 | 191 | 51 | 8 | 4 | 7 | 7 | 5.5 |
| InWeb_IM | 12 | 126 | 170 | 52 | 8 | 4 | 7 | 7 | 5.5 |
| INstruct | 18 | 60 | 154 | 53 | 2 | 10 | 13 | 2 | 6 |
| Reactome | 23 | 45 | 139 | 54 | 5 | 7 | 9 | 5 | 6 |
| HINT | 27 | 76 | 178 | 47 | 8 | 4 | 6 | 8 | 6 |
| UniHI | 20 | 80 | 153 | 47 | 9 | 3 | 5 | 9 | 6 |
| STRING | 136 | 284 | 848 | 93 | 7 | 5 | 6 | 8 | 6.5 |
| MatrixDB | 28 | 32 | 158 | 48 | 4 | 8 | 8 | 6 | 7 |
| mentha | 29 | 82 | 185 | 49 | 8 | 4 | 4 | 10 | 7 |
| PIPS | 67 | 32 | 325 | 57 | 0 | 12 | 12 | 3 | 7.5 |
| IntAct | 20 | 77 | 144 | 47 | 6 | 6 | 5 | 9 | 7.5 |
| HIPPIE | 25 | 96 | 192 | 53 | 7 | 5 | 4 | 10 | 7.5 |
| FI | 88 | 120 | 488 | 76 | 8 | 4 | 3 | 11 | 7.5 |
| HuRI | 134 | 46 | 758 | 91 | 3 | 9 | 7 | 7 | 8 |
| APID | 31 | 81 | 196 | 48 | 7 | 5 | 3 | 11 | 8 |
| iRefWeb | 22 | 43 | 129 | 44 | 7 | 5 | 3 | 11 | 8 |
| I2D | 30 | 68 | 189 | 47 | 7 | 5 | 2 | 12 | 8.5 |
| IMEx | 7 | 24 | 51 | 37 | 7 | 5 | 2 | 12 | 8.5 |
| HumanNet | 21 | 66 | 152 | 42 | 5 | 7 | 3 | 11 | 9 |
| IID | 39 | 111 | 295 | 49 | 6 | 6 | 2 | 12 | 9 |
| iRefIndex | 28 | 96 | 211 | 55 | 6 | 6 | 2 | 12 | 9 |
| MINT | 15 | 42 | 89 | 41 | 7 | 5 | 1 | 13 | 9 |
| BIND | 18 | 27 | 114 | 52 | 2 | 10 | 5 | 9 | 9.5 |
| Signor | 16 | 65 | 126 | 57 | 1 | 11 | 4 | 10 | 10.5 |
| InnateDB | 16 | 34 | 124 | 49 | 1 | 11 | 1 | 13 | 12 |

*Note:* Abbreviations used in the table are: **# (number sign)**: number of, **Max_sub**: maximum size of subnetworks, **enrich.**: enrichment, **Pan-cancer_sub**: the identified subnetworks in [1], **Cancer_path**: KEGG pathways in cancer.

# References

[1] Mark DM Leiserson, Fabio Vandin, Hsin-Ta Wu, Jason R Dobson, Jonathan V Eldridge, Jacob L Thomas, Alexandra Papoutsaki, Younhun Kim, Beifang Niu, Michael McLellan, et al. Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature genetics*, 47(2):106–114, 2015.