PACZKOWSKA, M., BARENBOIM, J., SINTUPISUT, N. ET AL.

INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA

“Integrative pathway enrichment analysis helps distill thousands of high-throughput measurements to a smaller number of pathways and biological themes that are most characteristic of the experimental data at hand, ideally leading to mechanistic insights and candidate genes for follow-up studies. In particular, a joint analysis of complementary datasets often leads to insights that are unavailable in any particular dataset.“

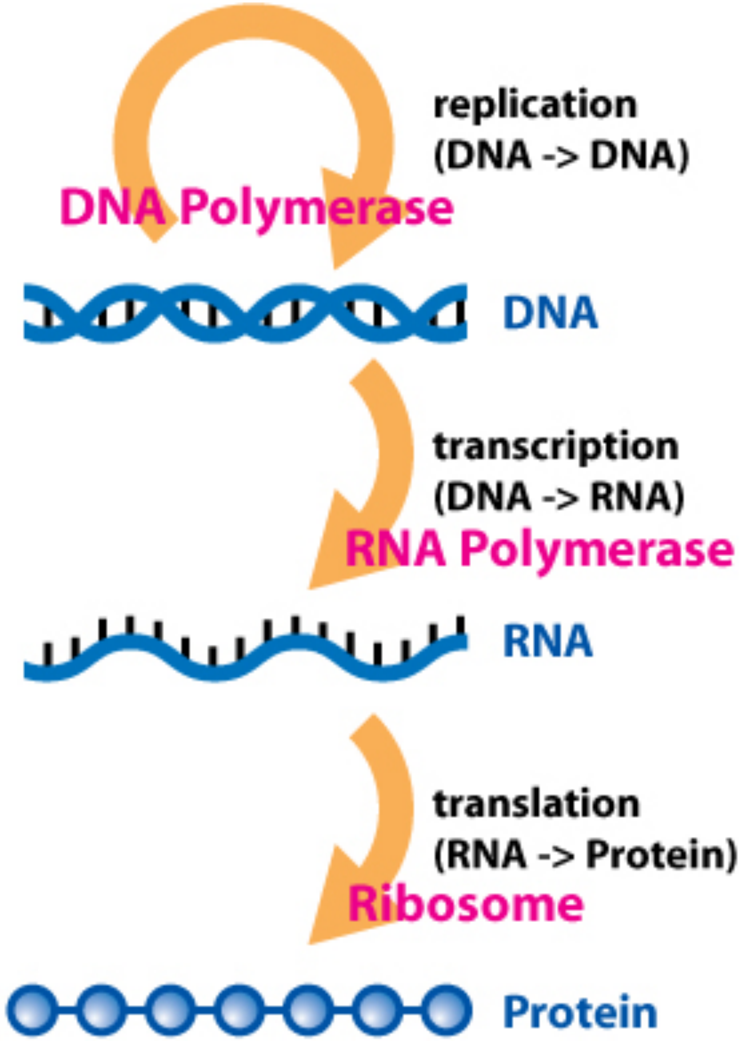
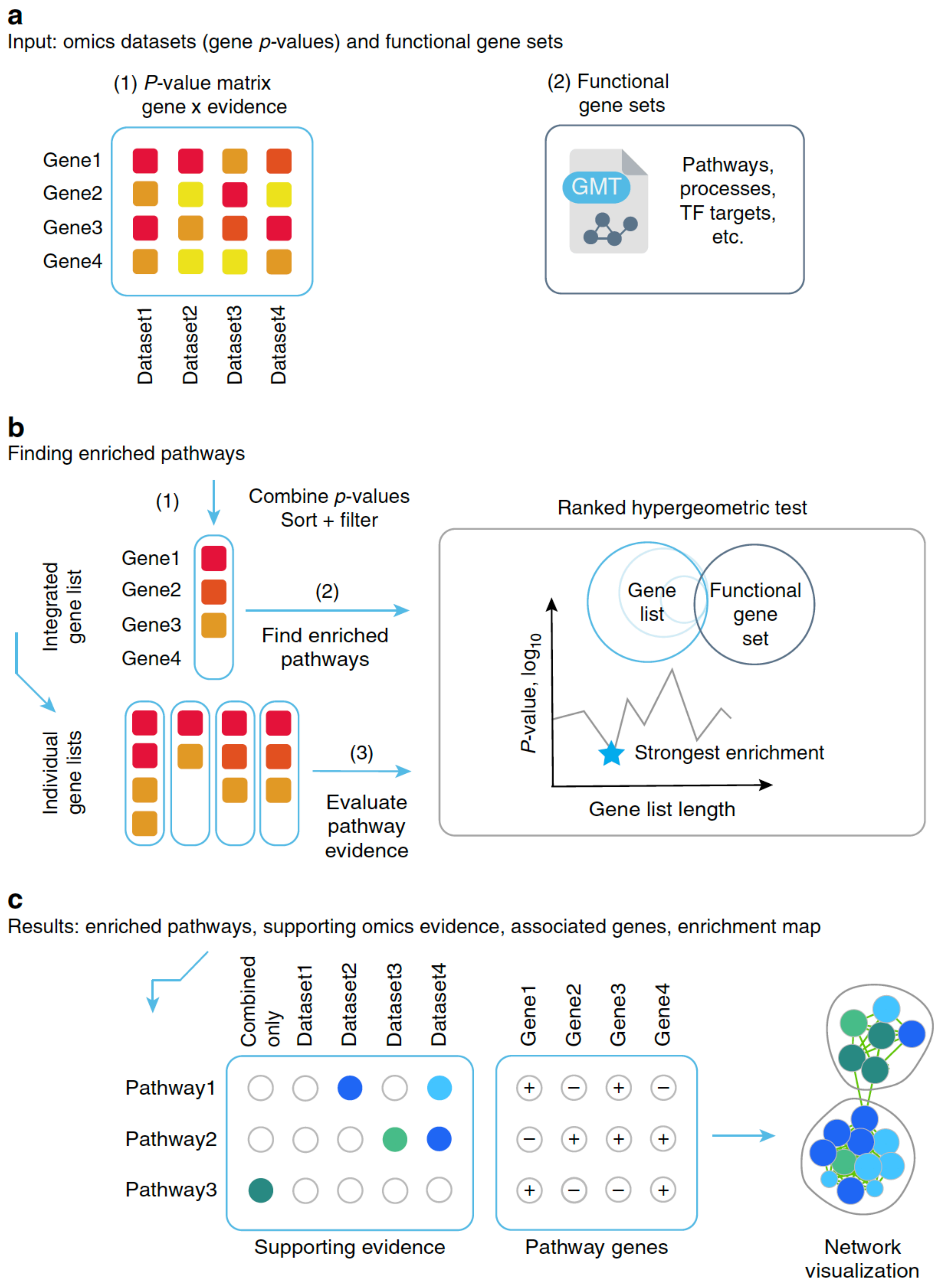
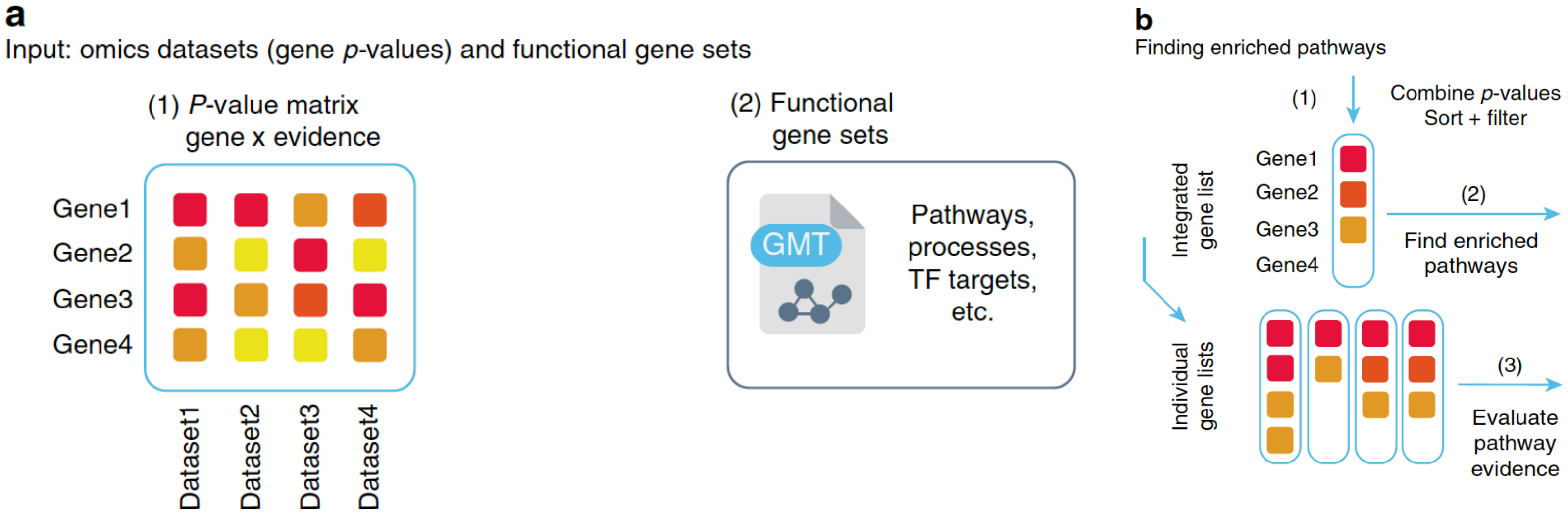


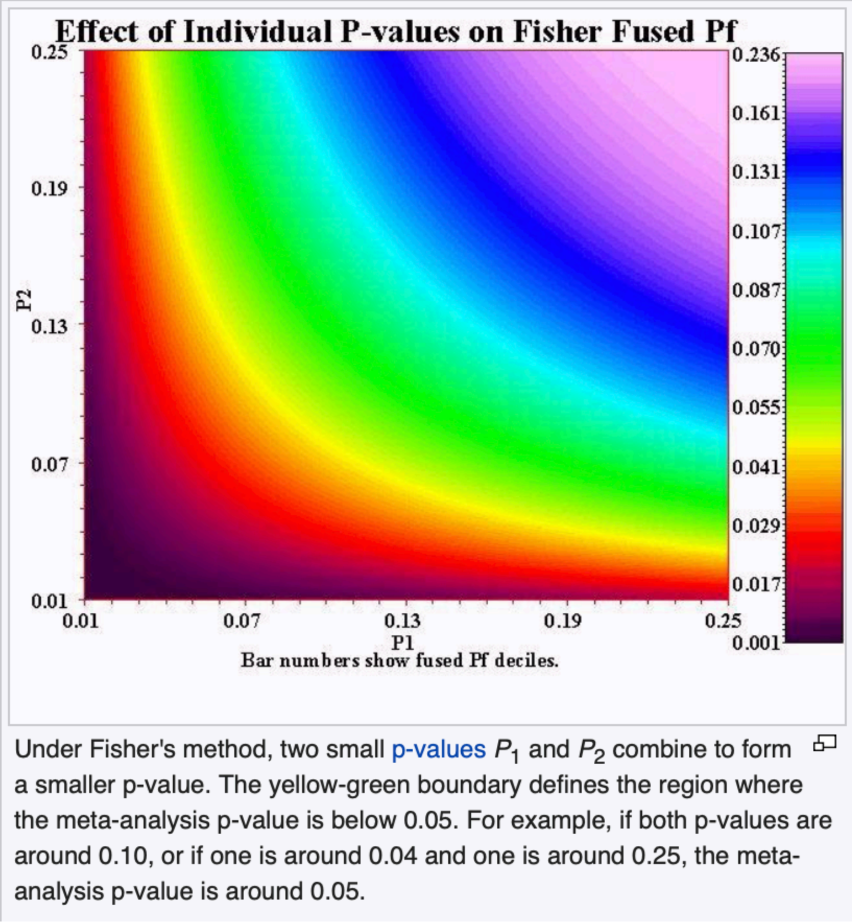
Image courtesy: https://commons.wikimedia.org/wiki/File:Central_Dogma_of_Molecular_Biochemistry_with_E

METHOD OVERVIEW



INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA

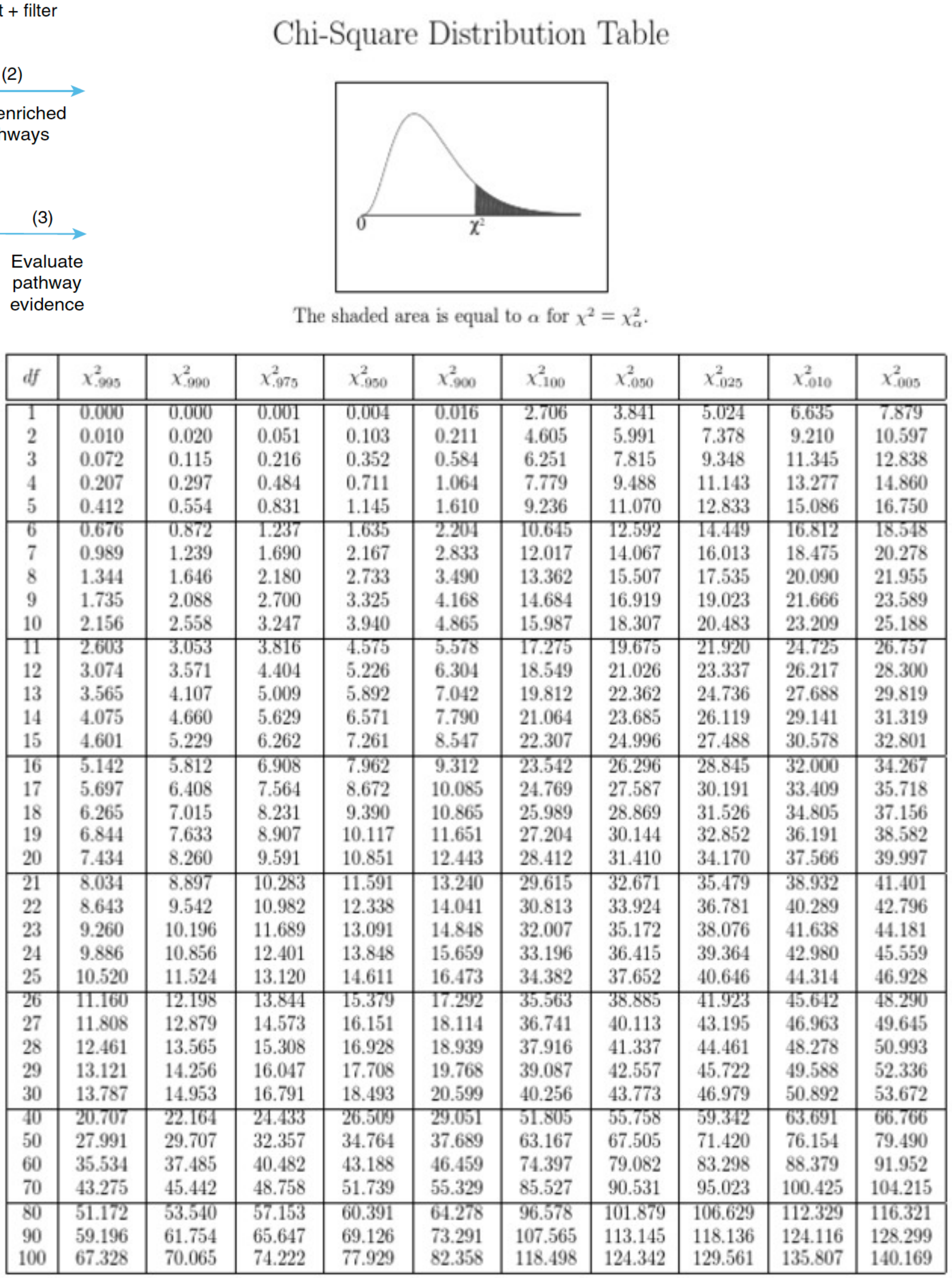




Under Fisher's method, two small p-values P_1 and P_2 combine to form a smaller p-value. The yellow-green boundary defines the region where the meta-analysis p-value is below 0.05. For example, if both p-values are around 0.10, or if one is around 0.04 and one is around 0.25, the meta-analysis p-value is around 0.05.

$$X^2_{2k} \sim -2\sum_{i=1}^{k} \ln(p_i),$$

Fisher fused Pf: https://en.wikipedia.org/wiki/Fisher%27s_method



Chi-Square Distribution Table

The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.169

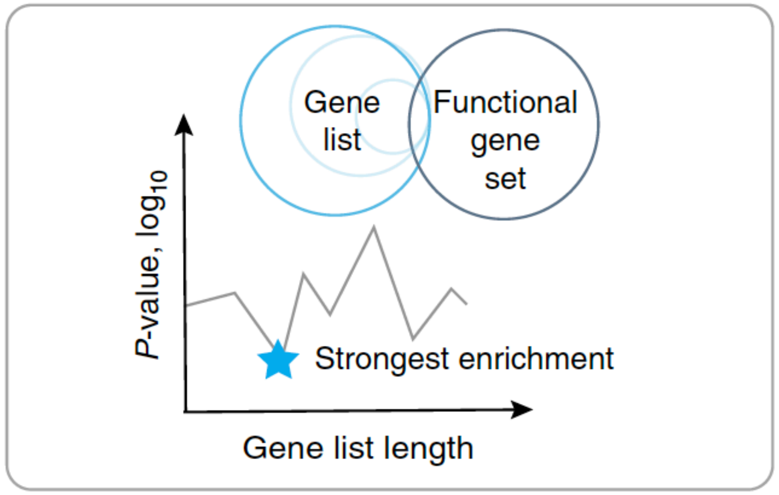
https://www.academia.edu/10107363/Chi-square-table

INPUTS TO ACTIVE PATHWAYS

Integrated Gene list	P-value
Int Gene 1	
Int Gene2	
Int Gene3	
Int Gene4	

Pathway Gene Set
PGS Gene1
PGS Gene2
PGS Gene3
PGS Gene4

Ranked hypergeometric test



ITERATION 1

Integrated Gene	P-value
Int Gene 1	

Pathway Gene Set
PGS Gene1
PGS Gene2
PGS Gene3
PGS Gene4

G=1, K=4

ITERATION 2

Integrated Gene list	P-value
Int Gene 1	
Int Gene 2	

Pathway Gene Set
PGS Gene1
PGS Gene2
PGS Gene3
PGS Gene4

G=2, K=4

$$\left(P_{\text{pathway}}, G\right) = \left\{\min, \operatorname{argmin}_n\right\} \sum_{x=k}^{\min(n,K)} \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}},$$

Ppathway stands for the hypergeometric P-value of the pathway enrichment at the optimal sub-list of the significance-ranked candidate genes

G represents the length of the optimal sub-list, i.e., the number of top genes from the input gene list,

N is the number of protein-coding genes with annotations in the pathway database, i.e., in Gene Ontology and Reactome,

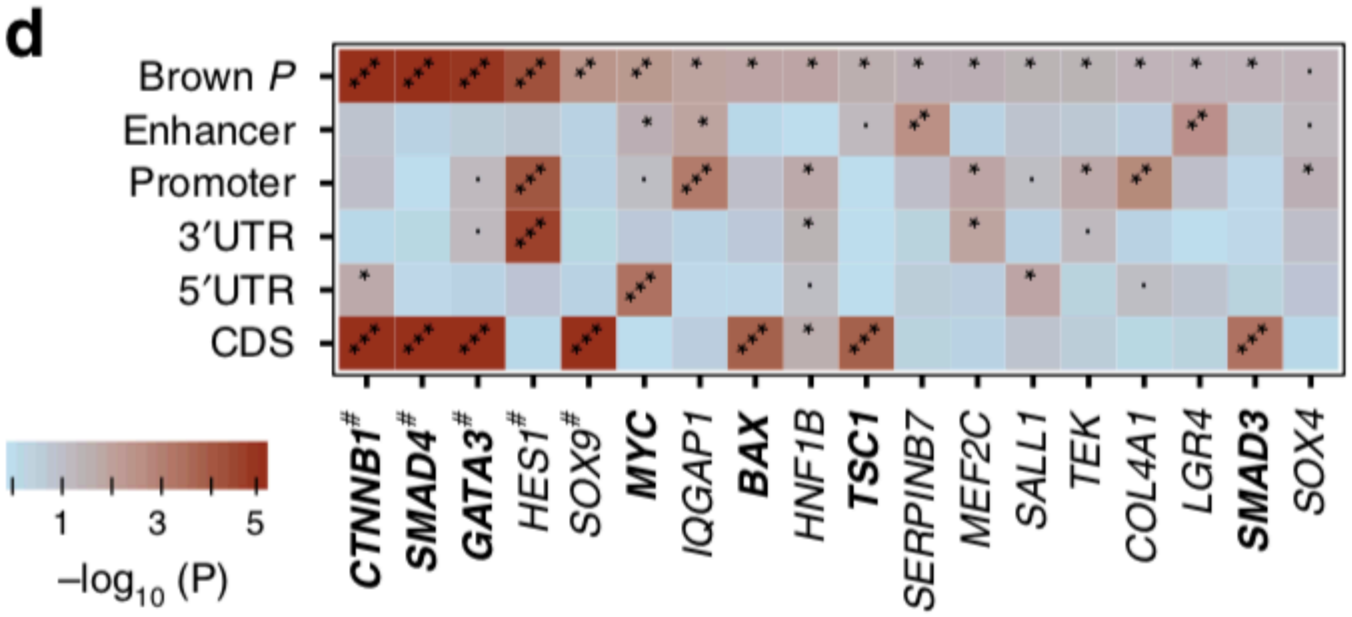
K is the total number of genes in a given pathway

n is the number of genes in a given gene sub-list considered

k is the number of pathway genes in the considered sub-list.

To obtain candidate genes involved in the pathway of interest, we intersect pathway genes with the optimal sub-list of candidate genes.

INPUT 1



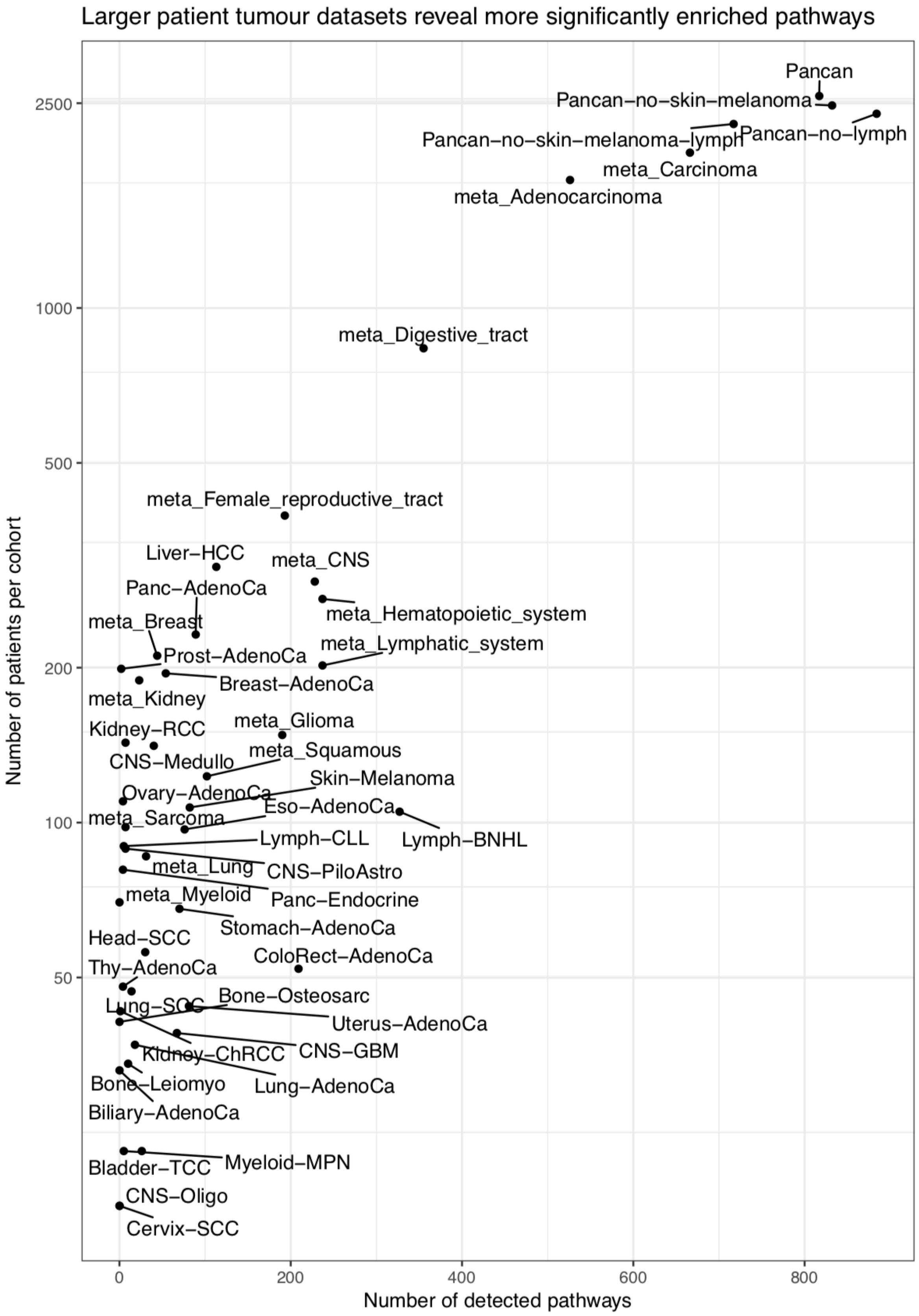
INPUT 2

PathwayX Gene Set
PXGS Gene1
PXGS Gene2
PXGS Gene3
PXGS Gene4

PathwayY Gene Set
PYGS Gene1
PYGS Gene2
PYGS Gene3
PYGS Gene4

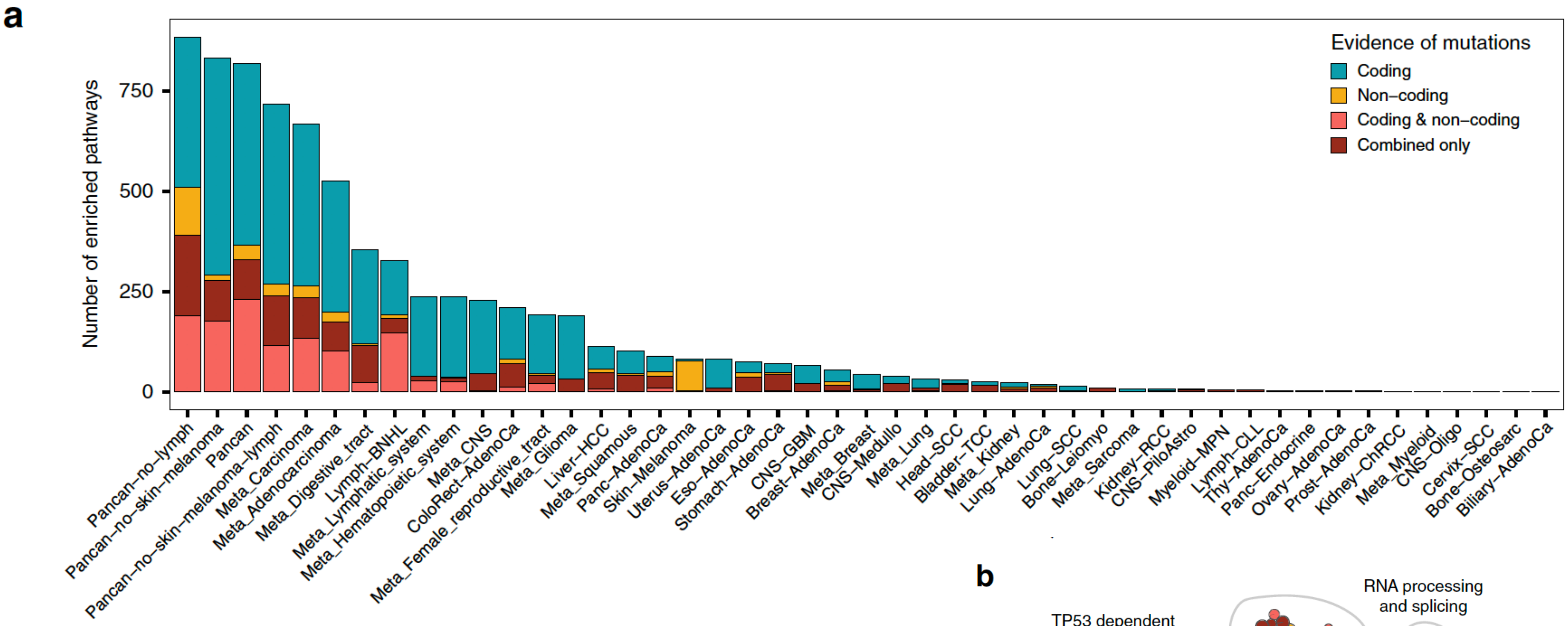
ANALYSIS OF CANCER DRIVER GENES

Data from 47 cohorts



Larger patient tumour datasets reveal more significantly enriched pathways

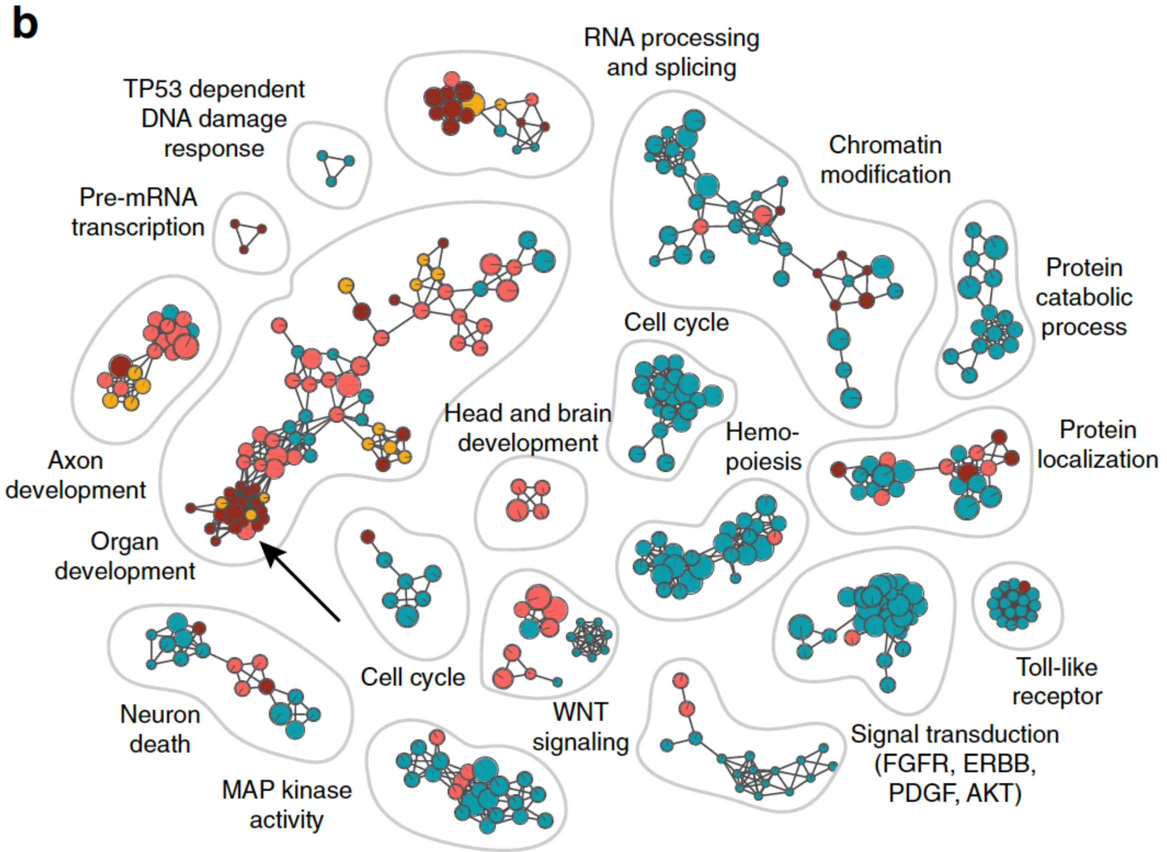
INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA

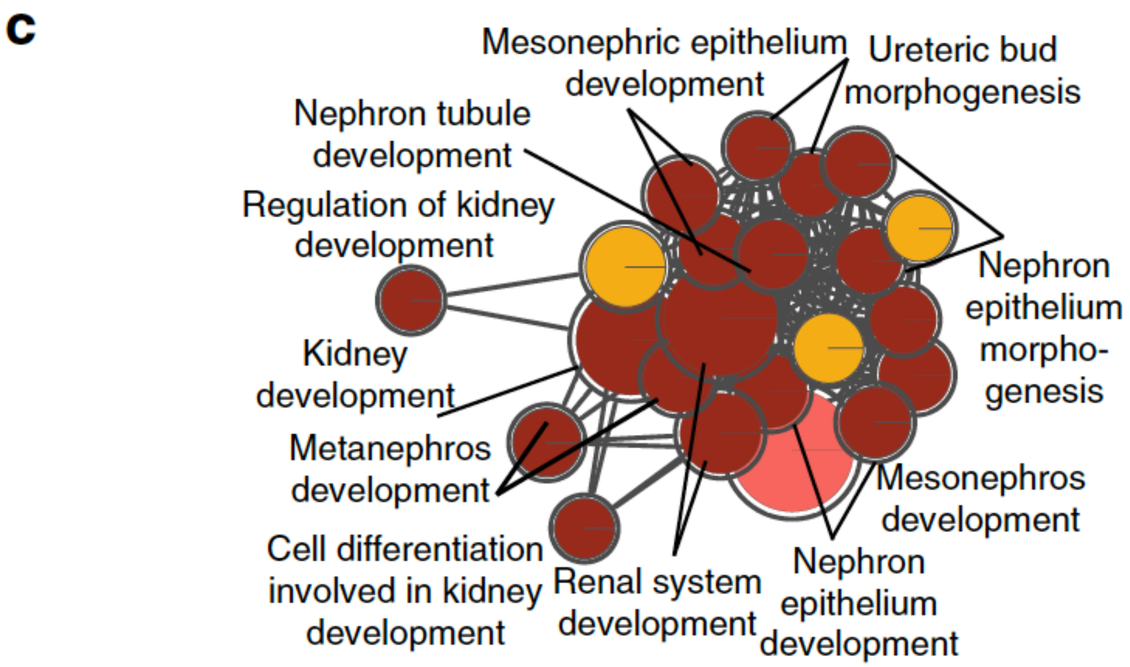


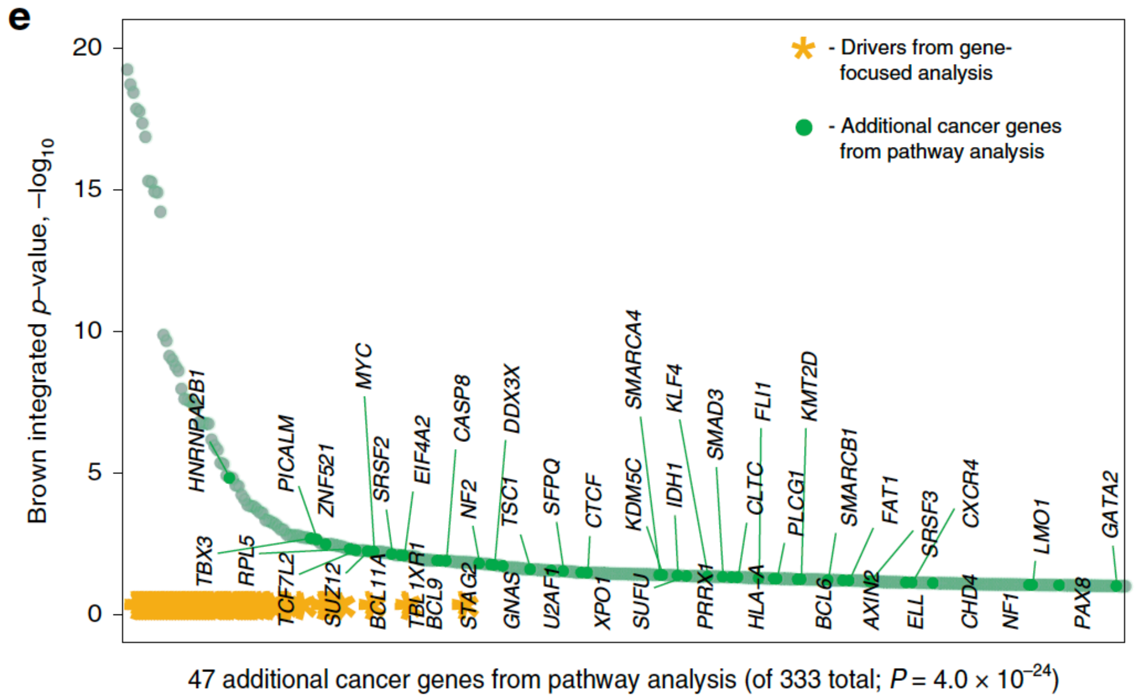
One significantly enriched process or pathway in 89% of 47 cohorts

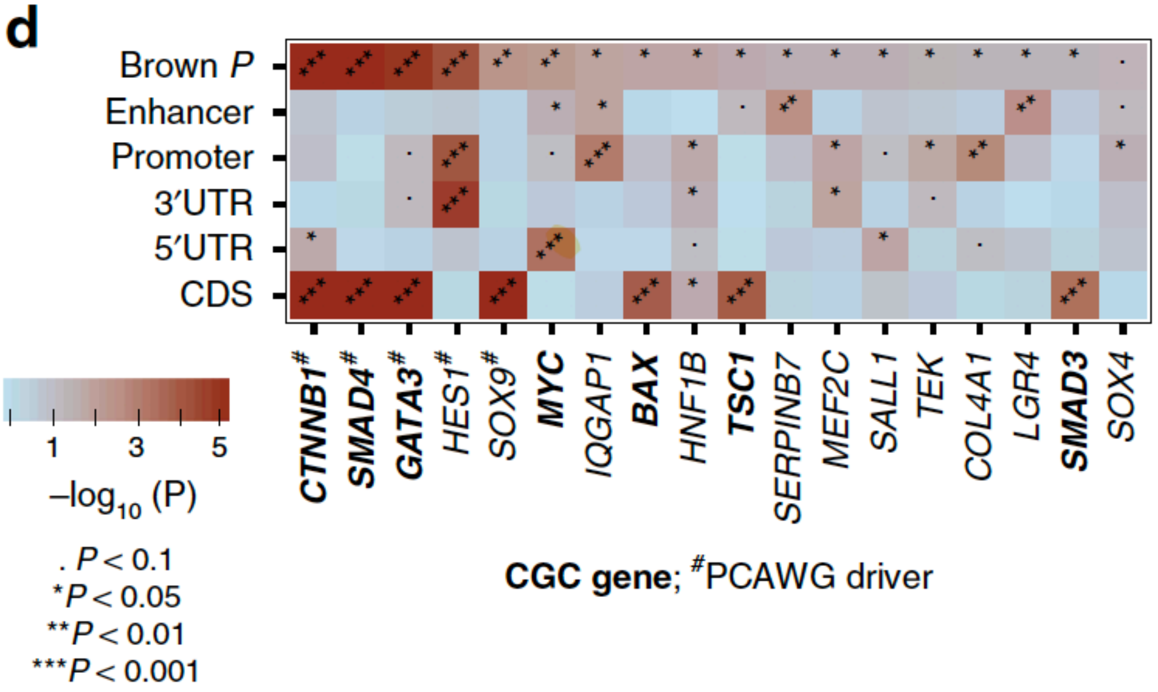
37/47 or 79% of cohorts showed enrichments in pathways supported by protein- coding mutations in genes

“24/47 cohorts (51%) showed significantly enriched pathways that were apparent when only analyzing non-coding driver scores corresponding to UTRs, promoters or enhancers.”




c
Mesonephric epithelium development
Ureteric bud morphogenesis
Nephron tubule development
Regulation of kidney development
Nephron epithelium morpho-genesis
Kidney development
Metanephros development
Mesonephros development
Cell differentiation involved in kidney development
Renal system development
Nephron epithelium development





18 genes involved in kidney development processes discovered, only five of them were predicted as driver genes in the consensus driver analysis of the PCAWG project

“Certain genes were upgraded through the data fusion procedure as a single stronger P-value per gene was derived by combining multiple weaker P-values corresponding to the coding regions, promoters, UTRs, enhancers of the gene. This affected 17/333 pathway-associated genes including six known cancer genes (HNRNPA2B1, STAG2, TCF7L2, SUZ12, CLTC, and ZNF521)”

INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA

INPUT 1

Gene Set	Tumor Cell mRNA	Tumor adjacent cell mRNA	Gene copy number alteration data
Gene1			
Gene2			
Gene3			
Gene4			

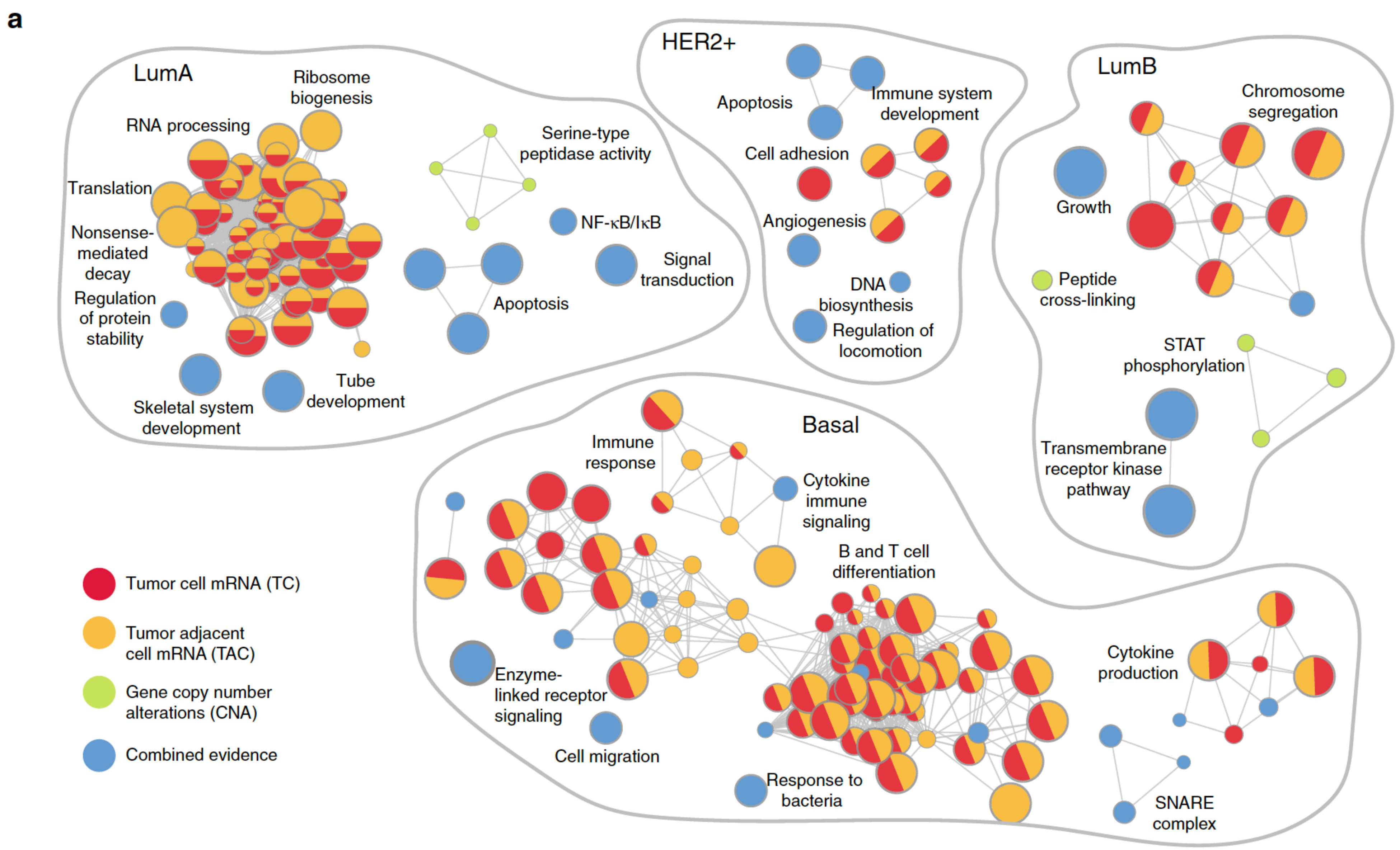
INPUT 2

PathwayX Gene Set
PXGS Gene1
PXGS Gene2
PXGS Gene3
PXGS Gene4

PathwayY Gene Set
PYGS Gene1
PYGS Gene2
PYGS Gene3
PYGS Gene4

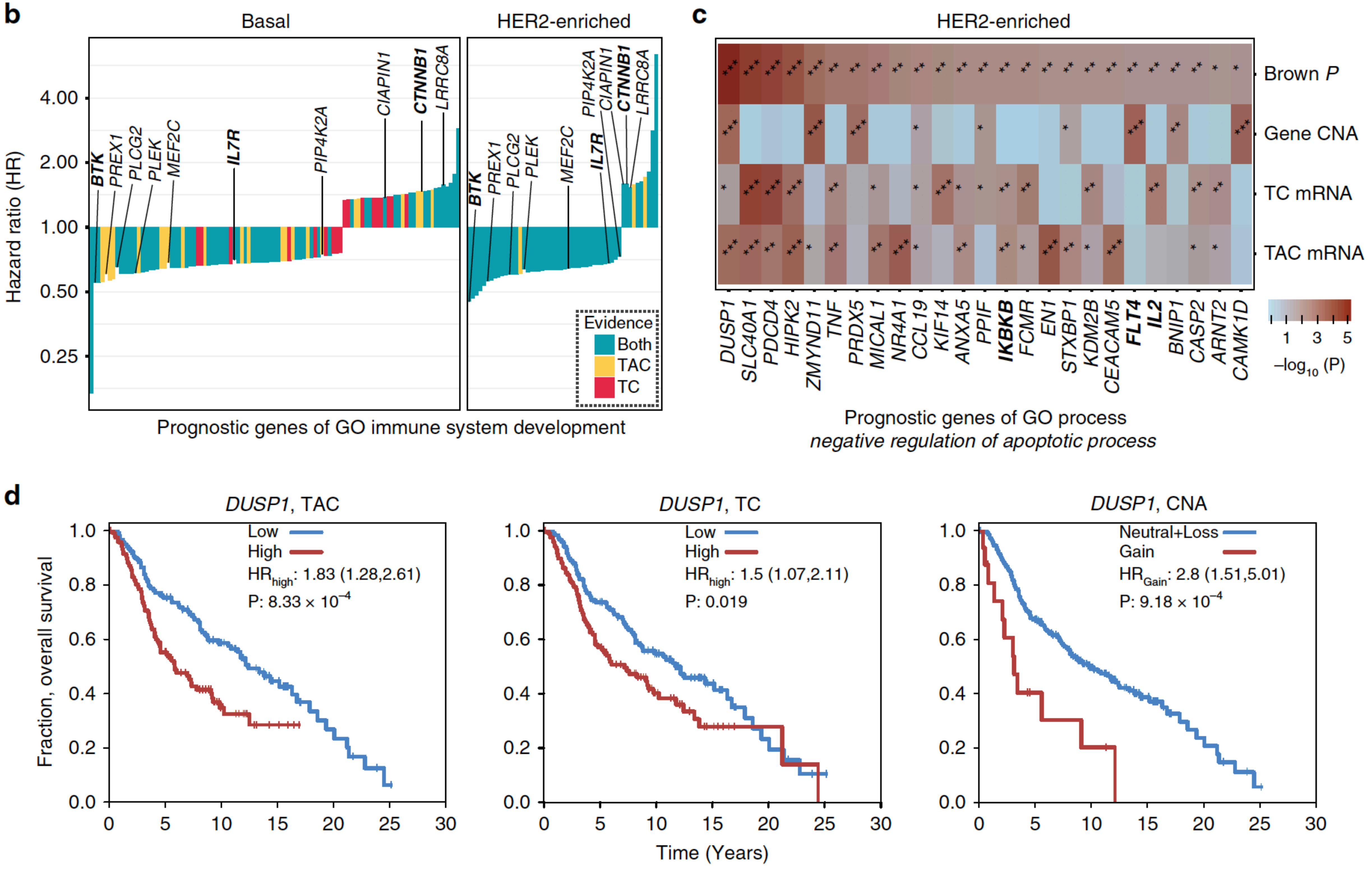
ANALYSIS OF BREAST CANCER DATA

330 basal-like breast cancers, 238 HER2-enriched breast cancers, 721 luminal-A breast cancers, 491 luminal-B breast cancers



192 significantly enriched pathways across the four subtypes of breast cancer
“Nine pathways were enriched in multiple cancer subtypes and 33 pathways were only apparent through the integrative pathway analysis but not in any of the CNA or mRNA datasets alone.”
“The major findings enriched in prognostic signatures in breast cancer subtypes involved the processes and pathways of immune response, apoptosis, ribosome biogenesis and chromosome segregation”

INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA



"DUSP1 encodes a phosphatase signaling protein of the MAPK pathway that is over-expressed in malignant breast cancer cells and inhibits apoptotic signaling"

INPUT 1

Gene Set	YAP transcriptional target	TAZ transcriptional target	YAP ChIP data
Gene1			
Gene2			
Gene3			
Gene4			

INPUT 2

PathwayX Gene Set
PXGS Gene1
PXGS Gene2
PXGS Gene3
PXGS Gene4

PathwayY Gene Set
PYGS Gene1
PYGS Gene2
PYGS Gene3
PYGS Gene4

HIPPO SIGNALING PATHWAY

INPUT 1

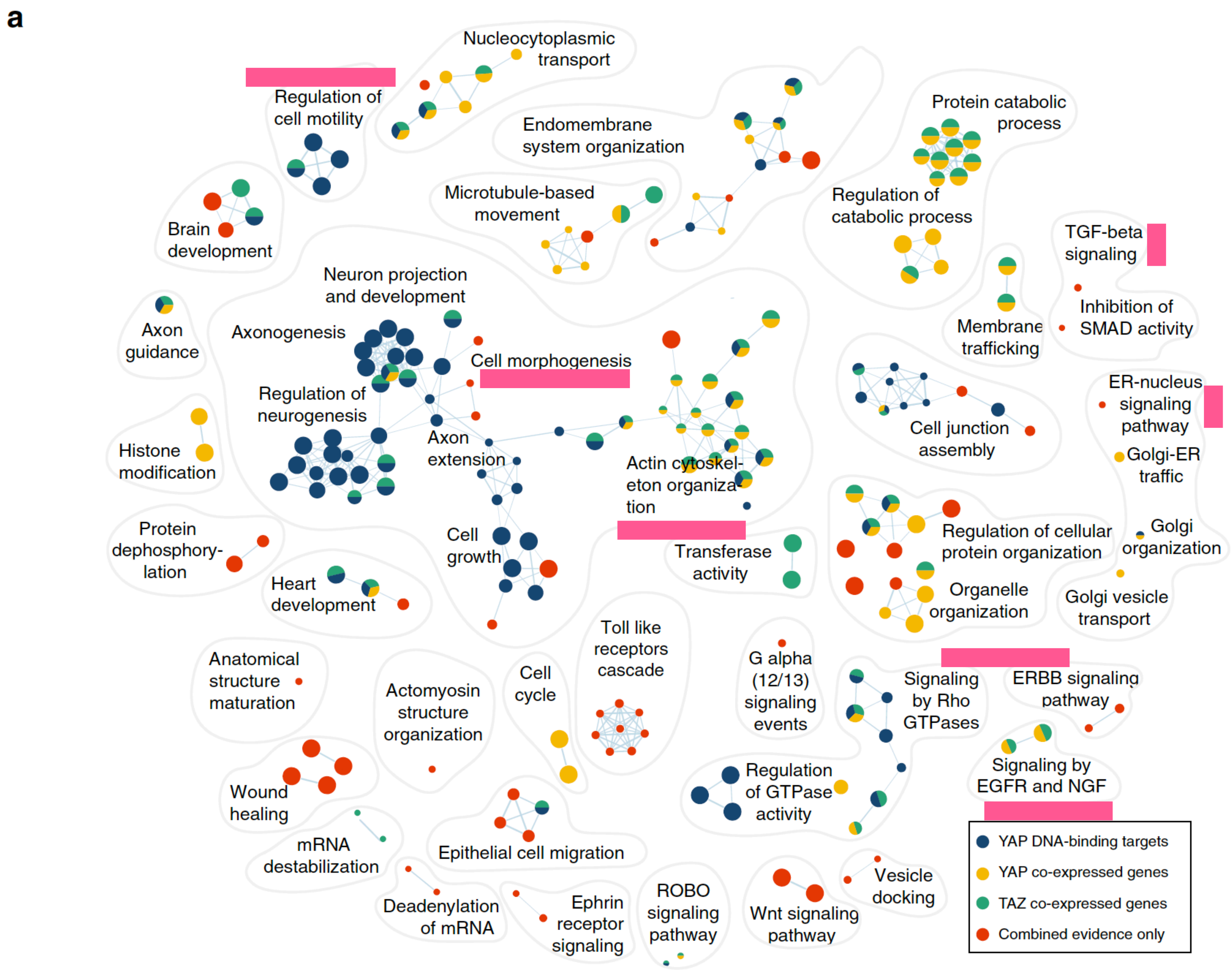
Gene Set	YAP transcriptional target	TAZ transcriptional target	YAP ChIP data
Gene1			
Gene2			
Gene3			
Gene4			

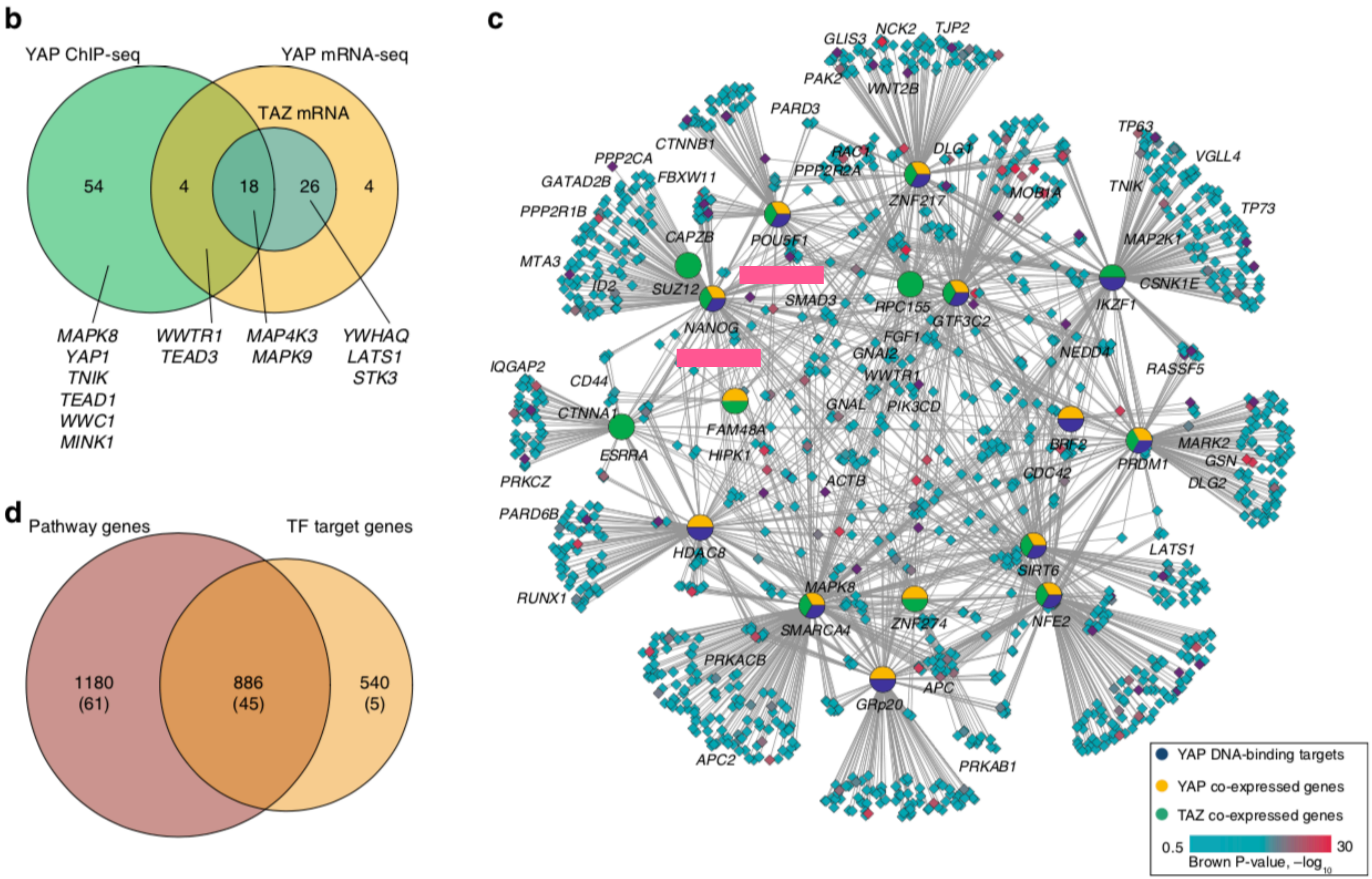
INPUT 2

TFX Target Gene Set
TFXGS Gene1
TFXGS Gene2
TFXGS Gene3
TFXGS Gene4

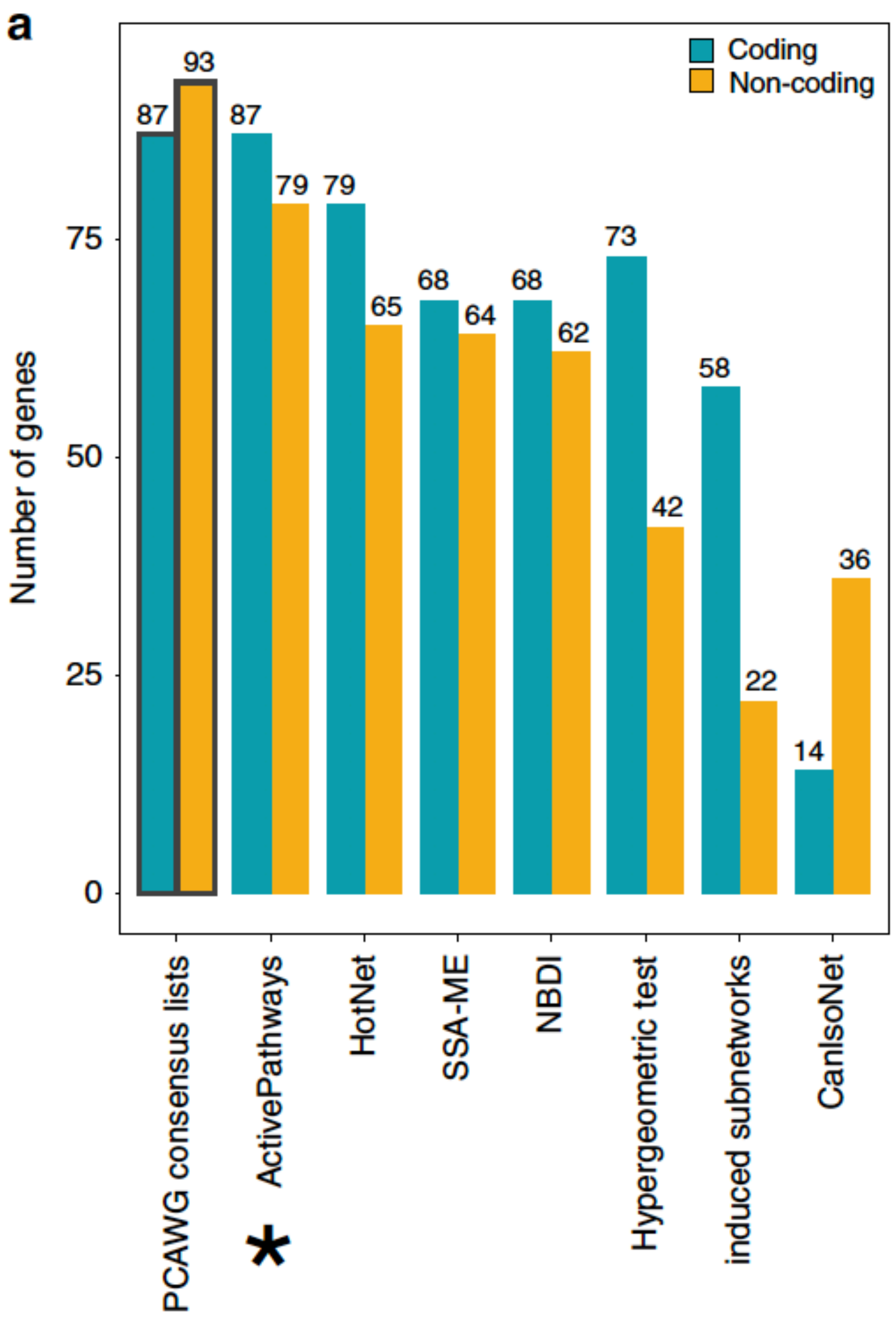
TFY Target Gene Set
TFYGS Gene1
TFYGS Gene2
TFYGS Gene3
TFYGS Gene4

INTEGRATIVE PATHWAY ENRICHMENT ANALYSIS OF MULTIVARIATE OMICS DATA


a
Nucleocytoplasmic transport
Regulation of cell motility
Endomembrane system organization
Protein catabolic process
Brain development
Microtubule-based movement
Regulation of catabolic process
TGF-beta signaling
Neuron projection and development
Axon guidance
Axonogenesis
Cell morphogenesis
Membrane trafficking
Inhibition of SMAD activity
Regulation of neurogenesis
Histone modification
Axon extension
Actin cytoskeleton organization
Cell junction assembly
ER-nucleus signaling pathway
Golgi-ER traffic
Protein dephosphorylation
Cell growth
Regulation of cellular protein organization
Golgi organization
Heart development
Transferase activity
Organelle organization
Golgi vesicle transport
Toll like receptors cascade
Anatomical structure maturation
Actomyosin structure organization
Cell cycle
G alpha (12/13) signaling events
Signaling by Rho GTPases
ERBB signaling pathway
Wound healing
Regulation of GTPase activity
Signaling by EGFR and NGF
mRNA destabilization
Epithelial cell migration
Deadenylation of mRNA
Ephrin receptor signaling
ROBO signaling pathway
Wnt signaling pathway
Vesicle docking
● YAP DNA-binding targets
● YAP co-expressed genes
● TAZ co-expressed genes
● Combined evidence only



Hippo signaling pathway is involved in organ size control, tissue homeostasis and cancer.


a
Coding
Non-coding
Number of genes
0
25
50
75
87
93
87
79
79
65
68
64
68
62
73
42
58
22
14
36
PCAWG consensus lists
ActivePathways
*
HotNet
SSA-ME
NBDI
Hypergeometric test
induced subnetworks
CanIsoNet

DISCUSSION

- Widely applicable to different datasets
- Databases have variable coverage, rely on frequent data updatesand may miss sparsely annotated candidate genes
- “Pathway information is highly redundant and analyses of rich molecular datasets often result in many significant results reflecting the same underlying pathway”
- No gene-gene interactions