

Multi-omics data integration and enrichment

Cenna Doornbos

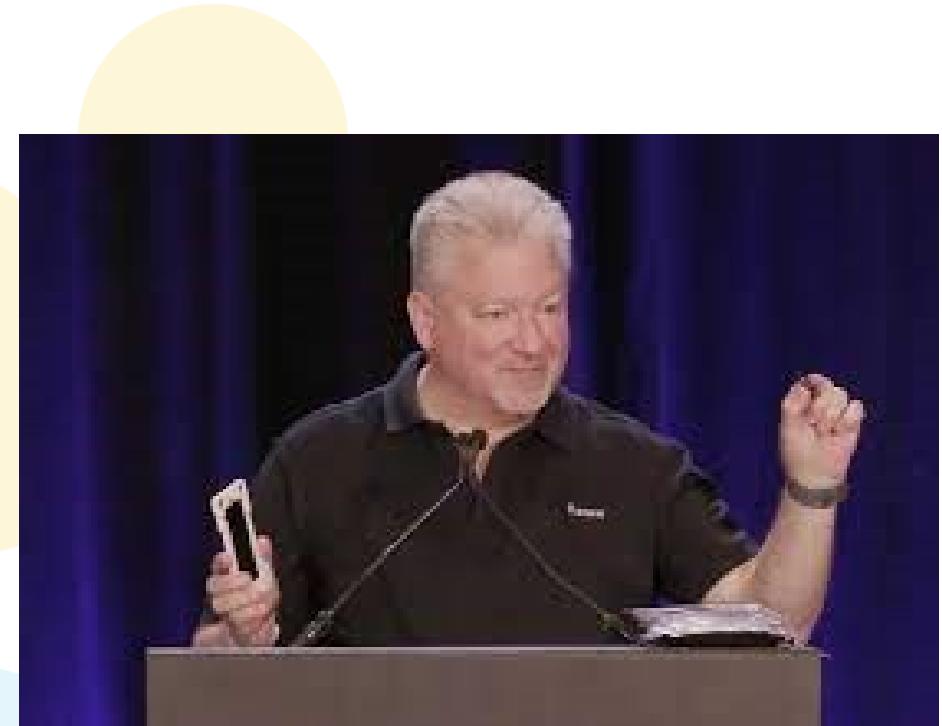
✓ Call Variants

✓ Annotate and filter Variants

✓ Polygenic Risk Scores

✓ WikiPathways

Multi-omics?



Gary Schroth



“When we talk about translational and clinical applications, multiomics is showing us how complicated biology really is, and that is a little bit daunting.”



TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Medical BioSciences
Radboudumc



Genomics

Epigenomics

Proteomics

Transcriptomics

Metabolomics

FISH

Blood
values

HPO

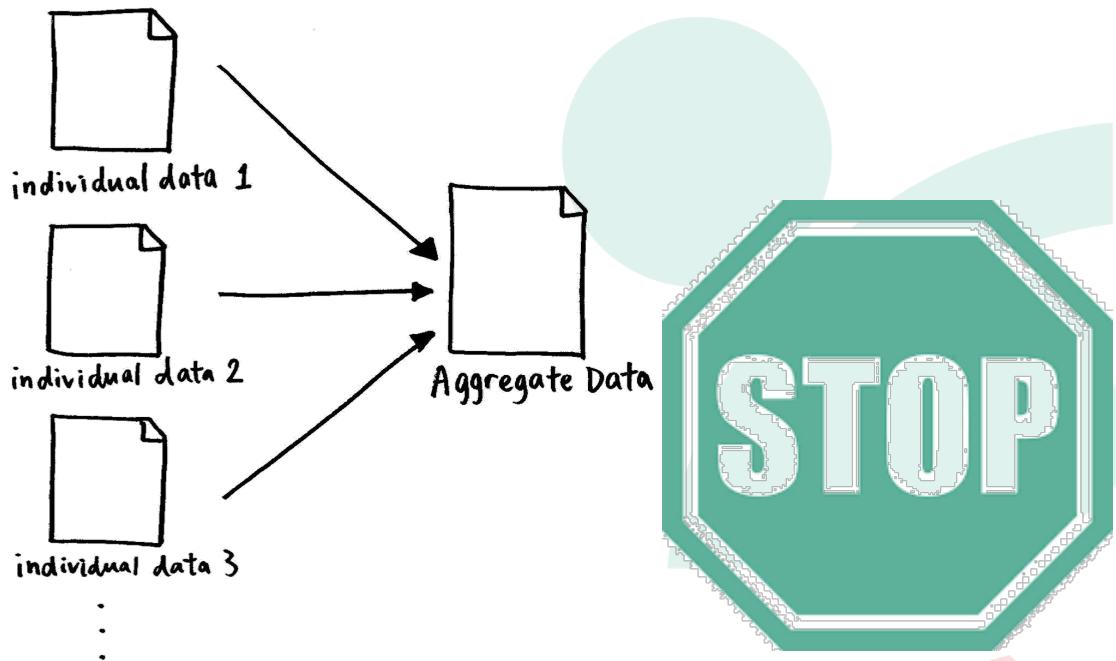
MRI

CT

ECG

Pedigree





A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS
1	POS	Name	Age	T	EXP	OVR	POT	S	M	C	PS	PM	PC	FB	CH	CV	SI	SL	Sp	Ct	Fo	Cc	Sc	Kc	Kn	V	ST	GR	FC	%D	W	L	SV	IP	ERA	DEM	Bang\$Buck	#P	LEAD	WORK	INT	INJ	Ethics	
2	SP	T. Thompson	34	R	14	74	74	5	7	6	5	7	6	-	5	-	6	6	-	-	-	-	-	-	96	6	71%	4	43.27	43.27	100%	5	8	1	4	4	Fragile	9						
3	MR	T. Poole	32	R	15	73	73	7	5	6	5	7	6	-	5	-	-	-	-	-	-	-	-	-	95	5	56%	7	42.80	42.80	100%	11	9	1	1	33%	4.78	\$7,000,000						
4	MR	D. Perez	32	R	10	74	74	6	6	5	6	5	6	-	-	-	-	-	-	-	-	-	-	-	96	3	71%	3	42.17	42.17	100%	5	5	2	3	3	Normal	11						
5	SP	E. Harrow	28	R	10	74	74	6	5	7	6	5	7	-	7	6	-	-	-	-	-	-	-	-	97	5	55%	4	41.98	41.98	100%	15	11	2	228%	3.39	\$19,000,000							
6	MR	Z. Phillips	32	R	11	70	70	5	7	6	5	7	6	-	-	-	-	-	-	-	-	-	-	-	92	3	68%	3	41.92	41.92	100%	6	7	1	1	35%	4.98	\$7,000,000						
7	SP	B. Lawson	36	R	10	74	74	5	6	7	5	6	7	-	6	-	-	-	-	-	-	-	-	-	88	5	65%	4	41.88	41.88	100%	9	1	1	112%	2.32	\$20,000,000							
8	SP	B. Koplow	35	R	14	72	72	5	6	6	5	6	6	5	-	4	6	-	-	-	-	-	-	-	96	4	68%	4	41.87	41.87	100%	8	13	0	174%	4.39	\$8,000,000							
9	SP	Z. Fricano	36	R	16	73	73	5	6	6	5	6	5	5	3	-	-	-	-	-	-	-	-	-	93	5	74%	2	41.87	41.83	100%	9	9	0	185%	2.41	\$13,000,000							
10	MR	C. Behan	28	L	11	61	63	7	6	4	7	6	5	6	-	7	-	-	-	-	-	-	-	-	94	3	56%	8	40.14	41.34	97%	1	9	0	66%	4.75	\$3,000,000							
11	SP	D. Hogan	37	R	17	68	68	5	6	5	5	6	5	-	6	-	-	-	-	-	-	-	-	-	96	6	66%	8	41.07	41.07	100%	9	11	0	155%	1.50	\$10,500,000							
12	SP	B. Langston	34	R	17	68	68	7	4	6	6	6	6	-	7	-	-	-	-	-	-	-	-	-	95	5	54%	6	41.01	41.00	100%	18	7	0	192%	4.41	\$10,000,000							
13	MR	P. Thibault	37	L	20	51	51	4	6	6	5	6	7	-	4	-	5	5	-	-	-	-	-	-	86	6	71%	7	38.60	40.60	95%	7	1	1	44%	2.45	\$3,200,000							
14	MR	R. Vara	32	R	11	63	63	4	7	6	4	7	6	-	2	-	5	5	-	-	-	-	-	-	89	7	70%	5	40.49	40.49	100%	3	0	1	58%	0.34	\$2,000,000							
15	SP	J. Kerstetter	29	R	12	60	61	5	7	4	5	7	4	-	5	6	-	-	-	-	-	-	-	-	94	4	66%	3	39.84	39.94	100%	7	0	0	166%	0.66	\$9,500,000							
16	SP	J. Kessler	28	R	11	59	60	6	6	6	5	6	7	-	4	-	6	-	-	-	-	-	-	-	97	5	57%	4	39.58	39.68	100%	10	9	0	163%	0.48	\$10,500,000							
17	SP	G. Catron	30	R	13	53	53	6	6	5	6	5	6	-	5	6	-	-	-	-	-	-	-	-	98	6	54%	8	39.59	39.59	100%	12	5	1	170%	1.38	\$6,500,000							
18	SP	J. Baird	31	R	11	57	57	5	6	5	6	5	6	-	5	6	-	-	-	-	-	-	-	-	94	5	54%	6	39.24	39.24	100%	7	11	0	156%	0.19	\$17,000,000							
19	SP	P. Milesell	32	L	14	58	58	6	5	5	6	5	7	3	6	-	-	5	-	-	-	-	-	-	98	4	53%	8	39.20	39.20	100%	8	3	0	95%	2.32	\$11,500,000							
20	SP	J. Sessions	34	R	17	59	59	5	6	5	5	6	5	-	6	-	-	-	-	-	-	-	-	-	95	6	61%	5	38.80	38.96	100%	10	10	0	185%	0.48	\$8,000,000							
21	SP	D. Hubert	33	R	16	56	56	5	6	6	5	6	6	-	3	5	-	-	-	-	-	-	-	-	95	6	57%	3	38.80	38.86	100%	9	12	0	191%	0.43	\$2,000,000							
22	SP	J. Chernoff	29	R	12	56	57	5	6	5	5	6	5	-	5	-	-	-	-	-	-	-	-	-	96	8	54%	7	38.57	38.67	100%	10	17	0	237%	0.35	\$1,500,000							
23	SP	N. Bouchet	36	R	16	50	50	5	6	5	6	5	6	-	3	5	-	-	-	-	-	-	-	-	92	4	53%	9	37.62	38.62	97%	5	13	0	156%	2.414	\$4,800,000							
24	MR	D. Bush	38	R	16	59	59	5	6	5	5	6	5	-	4	-	5	-	-	-	-	-	-	-	98	5	65%	5	38.58	38.58	100%	13	6	1	116%	2.354	\$7,000,000							
25	SP	R. Fialkowski	32	R	12	47	47	5	6	5	5	6	5	-	4	5	-	5	-	-	-	-	-	-	96	3	65%	6	38.47	38.47	100%	4	12	0	128%	0.591	\$1,100,000							
26	MR	P. Pekarsky	36	L	14	55	55	7	5	5	5	6	7	-	6	-	-	-	-	-	-	-	-	-	94	6	48%	6	39.14	38.14	100%	9	11	0	182%	0.361	\$6,500,000							
27	MR	T. Matthews	33	L	13	54	54	5	5	6	5	6	5	-	4	5	-	-	-	-	-	-	-	-	90	6	55%	8	38.00	38.00	100%	3	5	3	88%	0.348	\$7,000,000							
28	SP	R. Garnica	35	L	17	50	50	4	6	5	4	5	5	-	3	5	-	-	-	-	-	-	-	-	93	6	67%	6	37.63	37.63	100%	10	4	0	117%	2.367	\$5,000,000							
29	SP	T. Simms	31	R	14	47	47	6	6	5	6	6	6	-	5	-	-	-	-	-	-	-	-	-	97	6	49%	4	37.58	37.58	100%	11	10	1	185%	1.461	\$4,800,000							
30	SP	D. Richards	29	R	9	37	40	5	6	4	5	6	5	-	5	5	-	-	-	-	-	-	-	-	93	7	66%	6	36.23	37.53	97%	13	14	0	220%	0.380	\$9,500,000							
31	MR	P. Chandler	36	R	19	47	47	6	5	6	5	6	5	-	5	-	-	-	-	-	-	-	-	-	98	5	48%	8	37.49	37.49	100%	1	3	0	25.2	2.45	\$4,200,000							
32	MR	J. Adams	31	R	13	47	47	6	5	6	5	5	5	-	5	6	-	-	-	-	-	-	-	-	96	4	54%	6	37.47	37.47	100%	4	2	0	58.2	3.99	\$7,000,000							
33	SP	J. Brower	35	R	15	46	46	6	4	7	5	6	6	-	6	5	-	-	-	-	-	-	-	-	94	6	44%	5	37.04	37.04	100%	8	17	0	212.1	4.871	\$1,000,000							



Incorrect Duplicates
Missing data
Incorrect IDs
Low quality



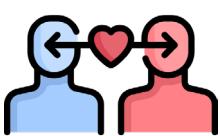
F

Findable



A

Accessible



i

Interoperable



R

Reusable



GO FAIR

go-fair.org/fair-principles

T R I V I A

APC?

- A) Adenomatous Polyposis Coli protein?
- B) Anaphase-Promoting Complex
- C) Another Possibly Complex name

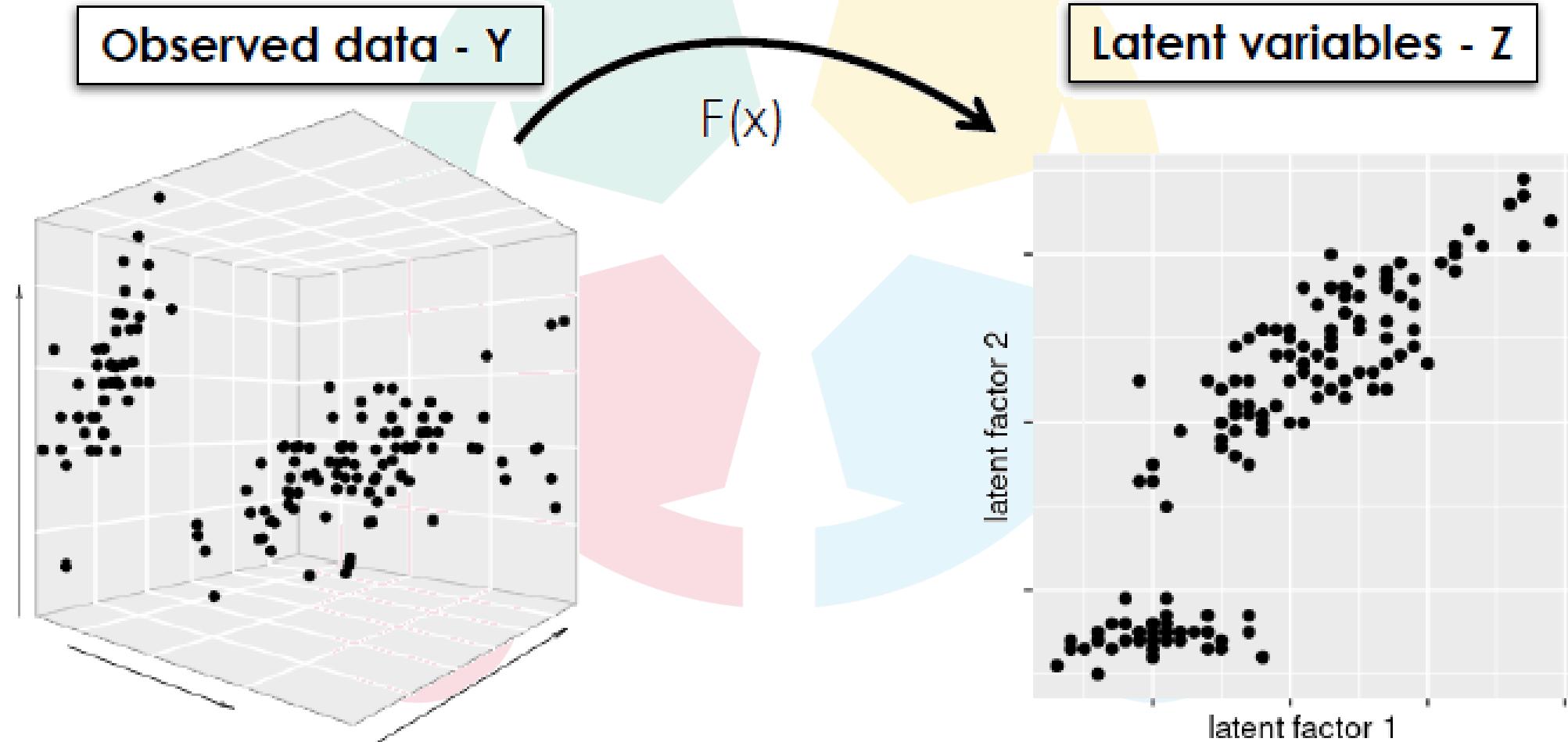
Medical BioSciences
Radboudumc



Funded by the
European Union
GA n°825575

High dimensionality → Low dimensionality

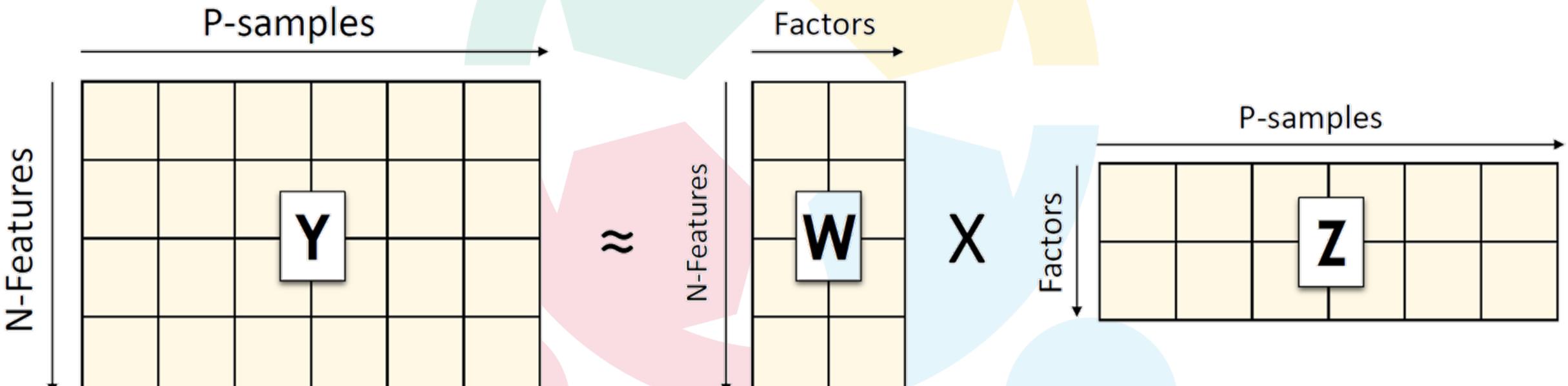
PCA



High dimensionality → Low dimensionality

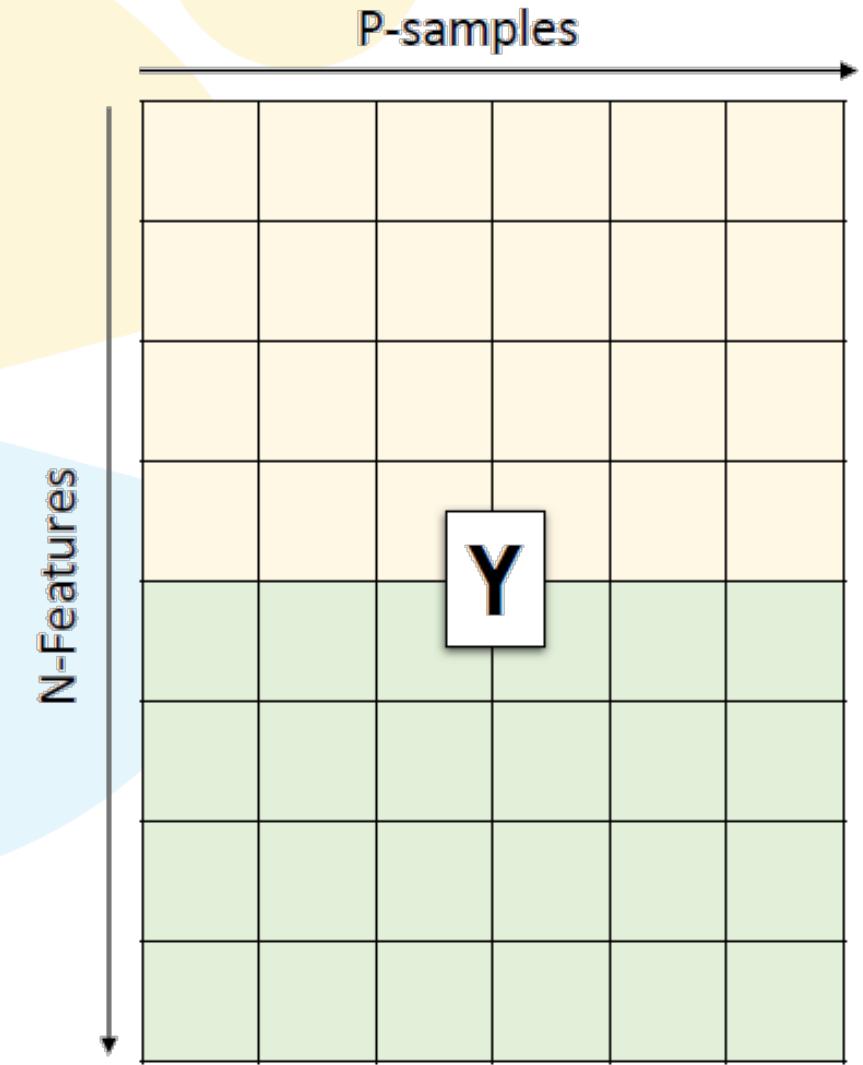
PCA

Matrix factorization



Challenges in multiomics data integration

- Source variance
- Data types and distributions
- Missing data
- Heterogeneity

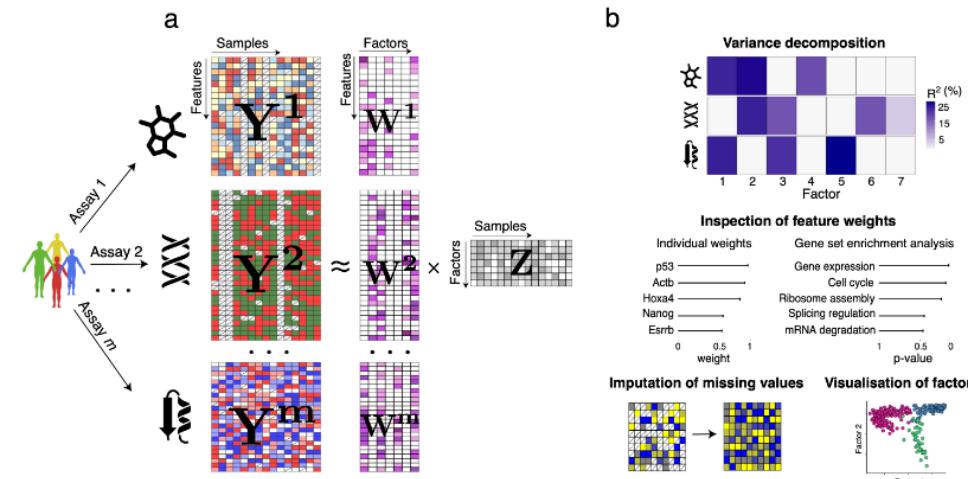


Biofam.github.io/MOFA2

Unsupervised
Bayesian model

Overview

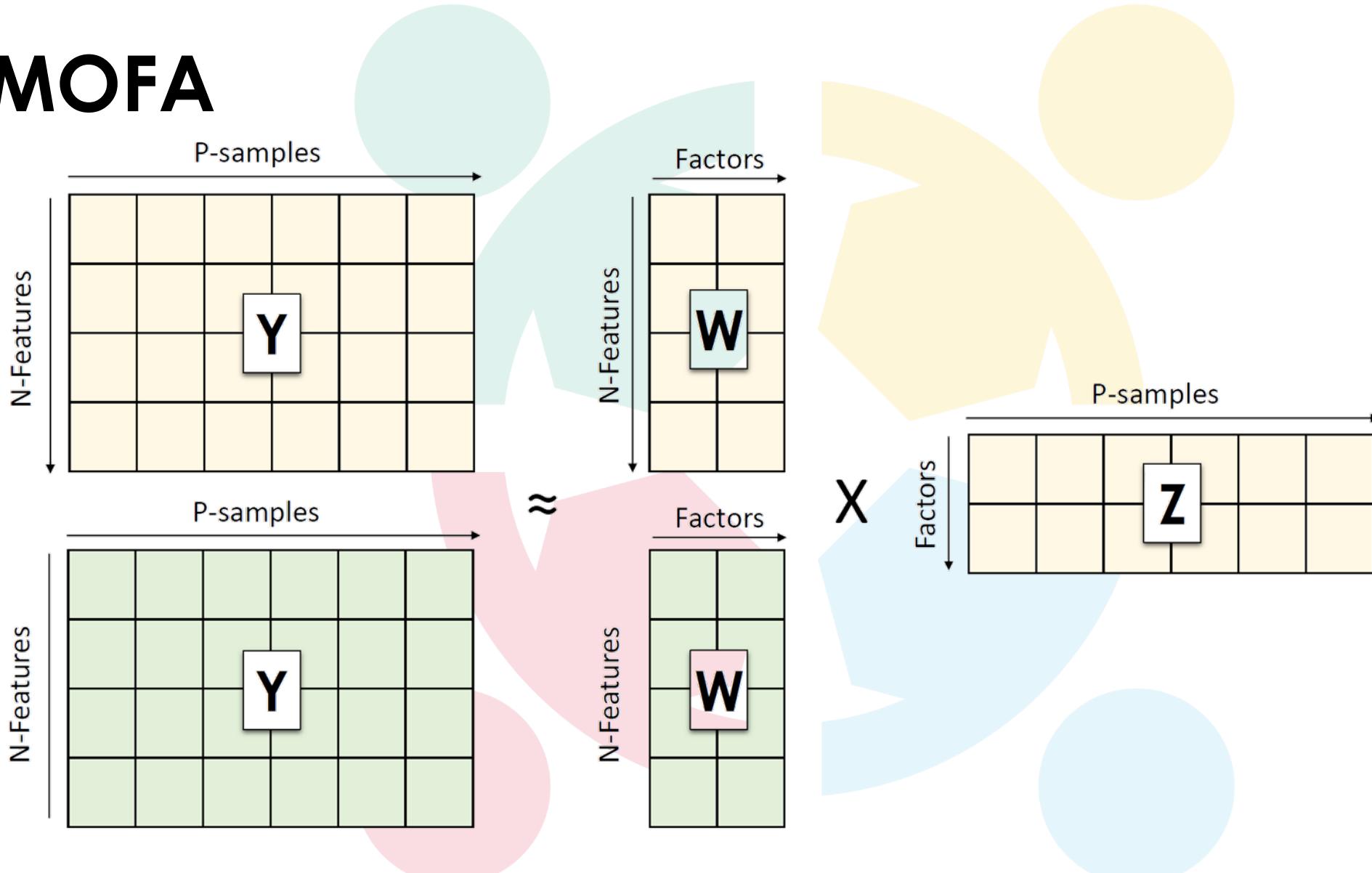
MOFA is a factor analysis model that provides a **general framework for the integration of multi-omic data sets** in an unsupervised fashion. Intuitively, MOFA can be viewed as a versatile and statistically rigorous generalization of principal component analysis to multi-omics data. Given several data matrices with measurements of multiple -omics data types on the same or on overlapping sets of samples, MOFA infers an **interpretable low-dimensional representation in terms of a few latent factors**. These learnt factors represent the driving sources of variation across data modalities, thus facilitating the identification of cellular states or disease subgroups.



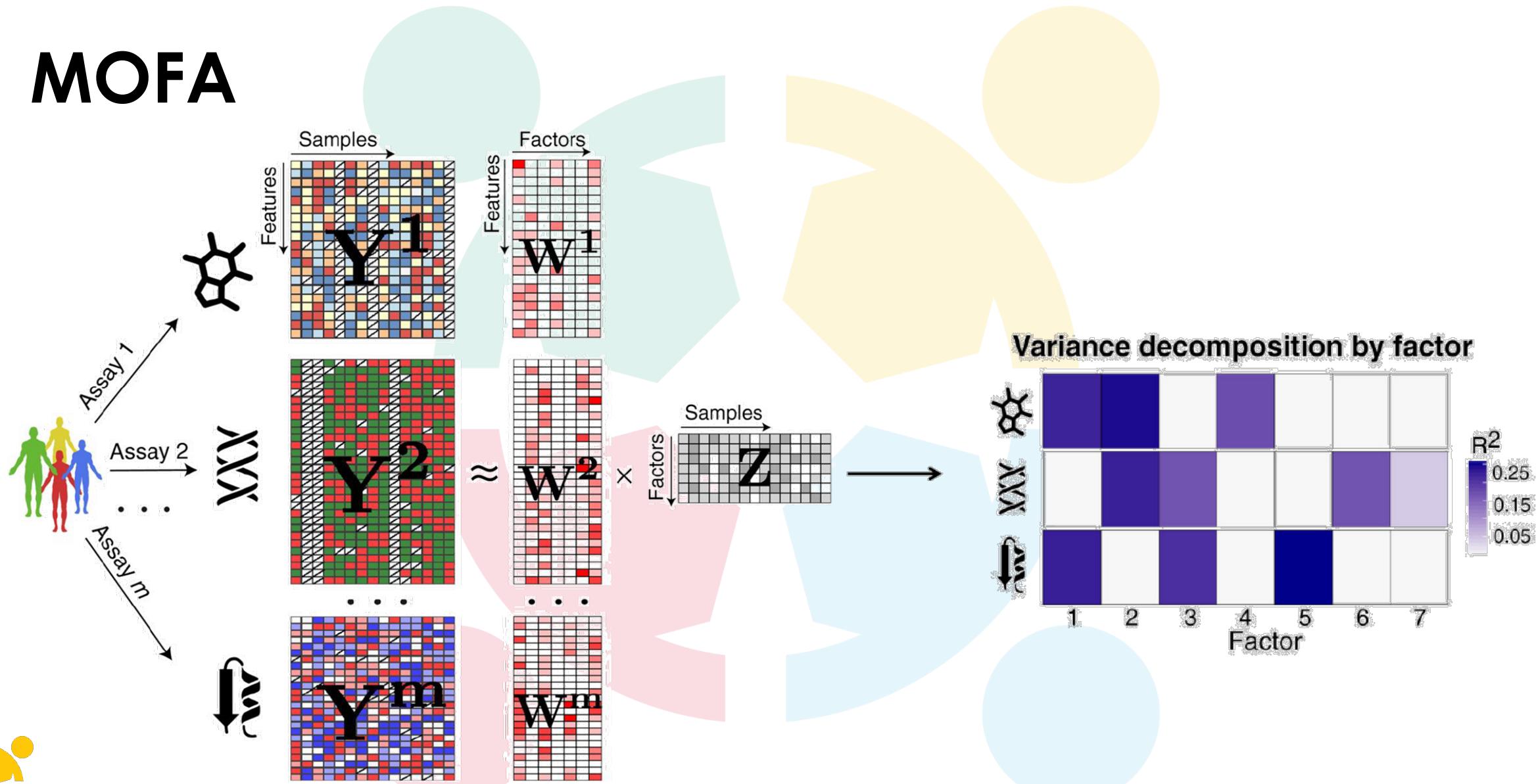
For more details you can read our papers:

- general framework: published in *Molecular Systems Biology*
- multi-group framework and single cell applications: **MOFA+**, published in *Genome Biology*
- temporal or spatial data: **MEFISTO**, published in *Nature Methods*

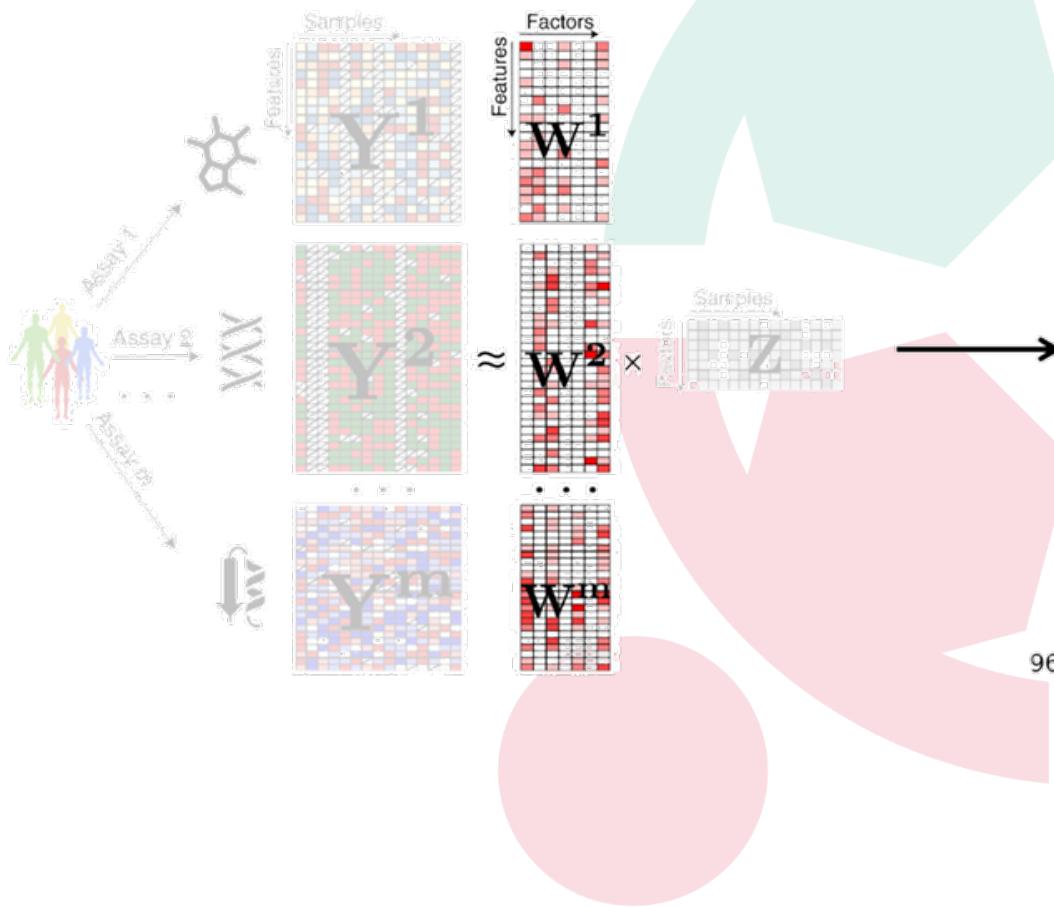
MOFA



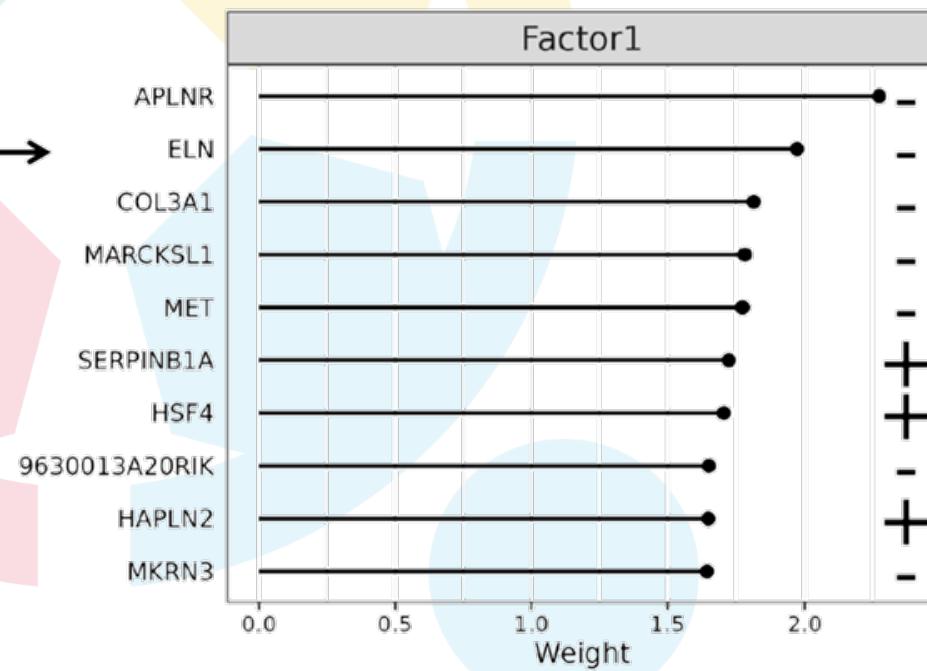
MOFA



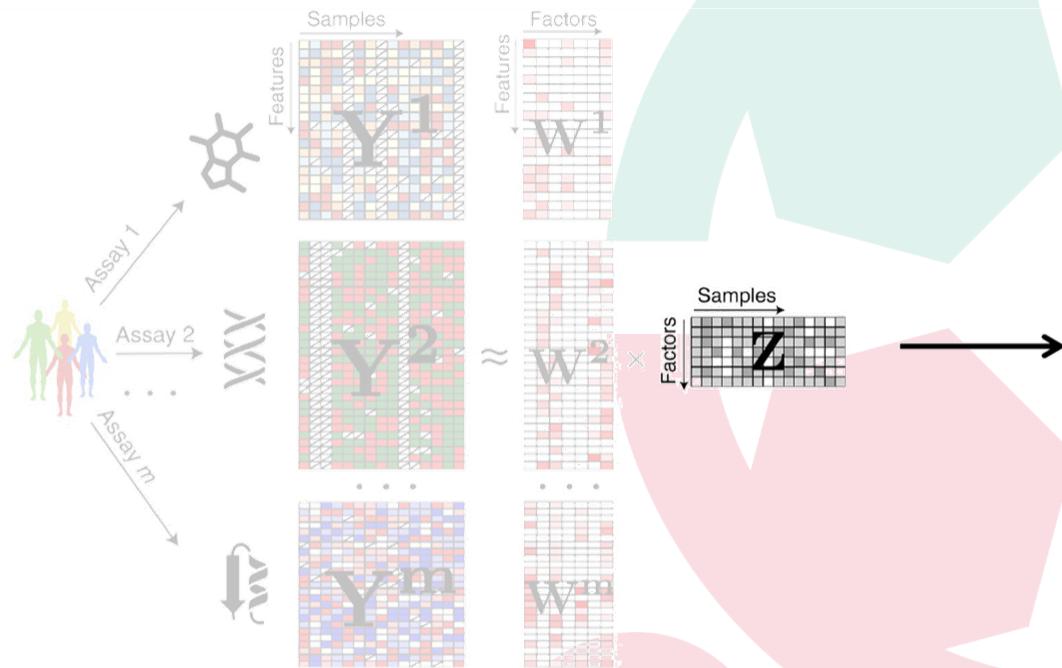
MOFA



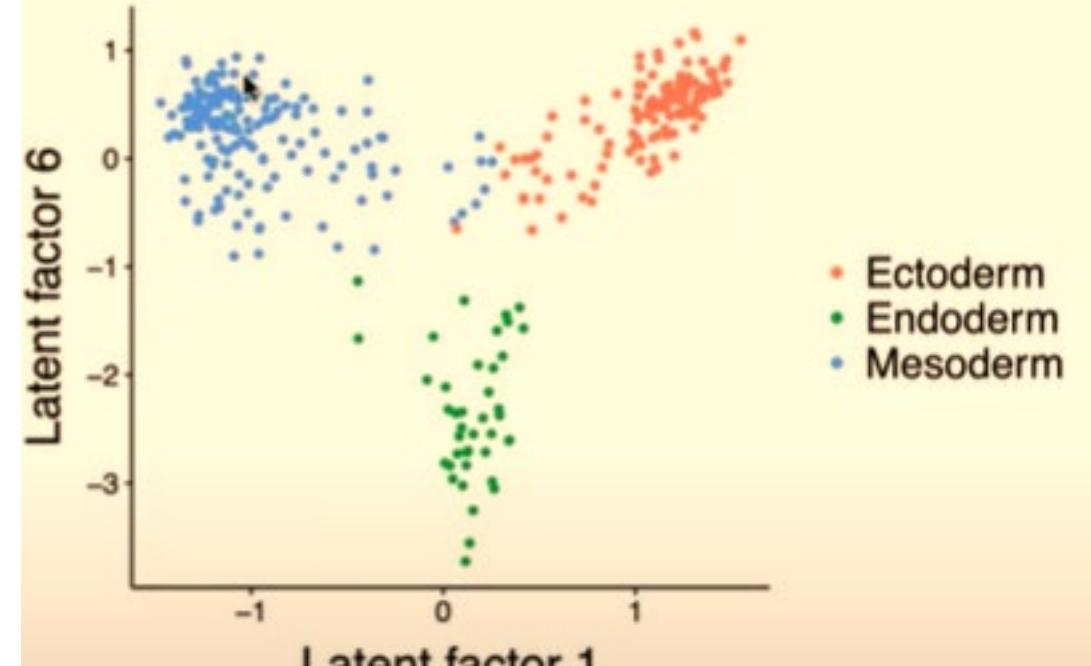
Feature weights exploration
- Top features



MOFA



Visualisation of samples in the latent space

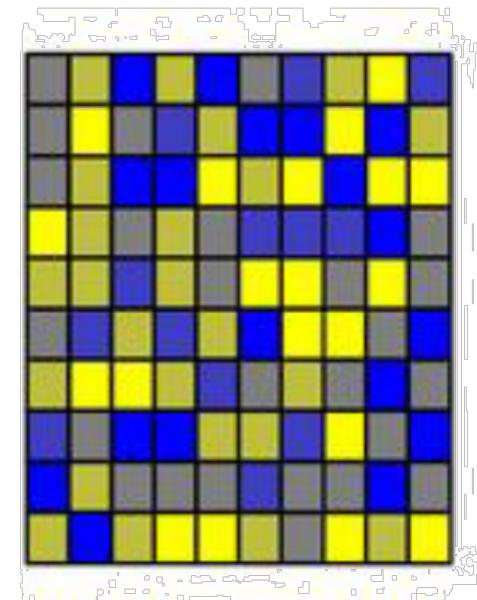
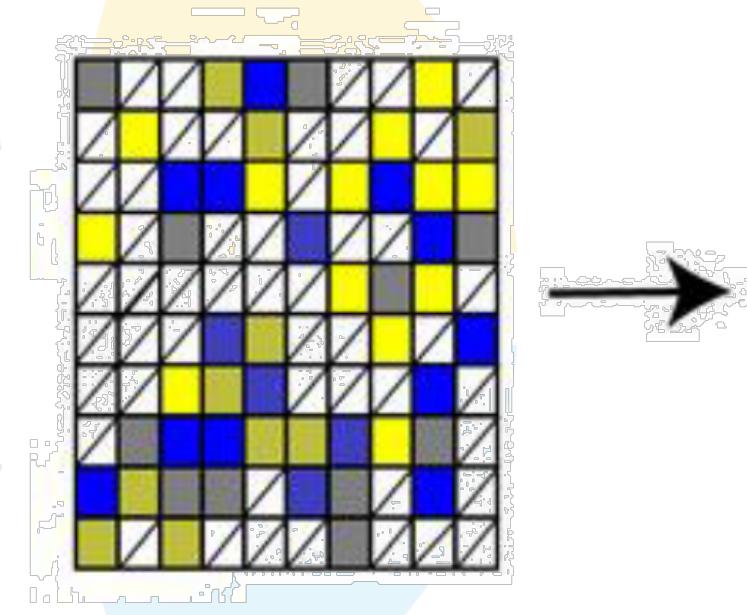
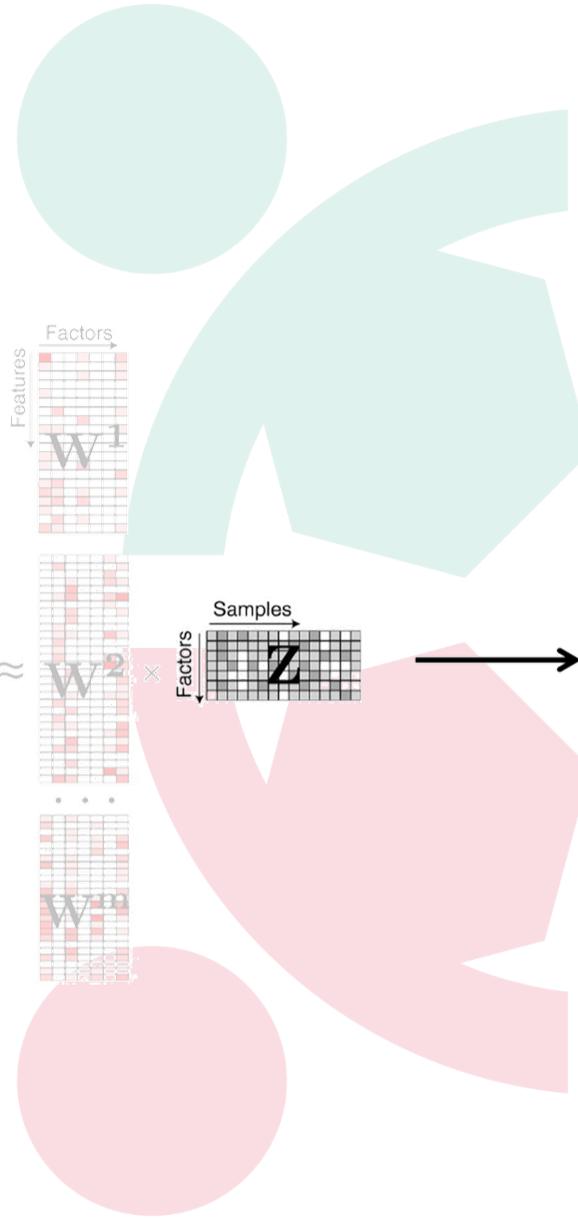
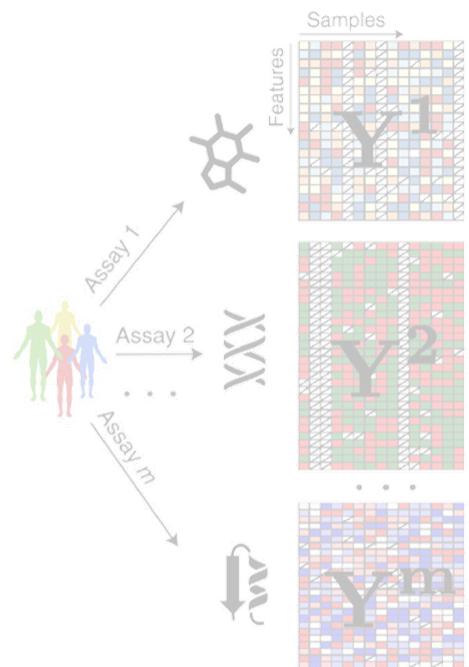


MOFA

Data should be:

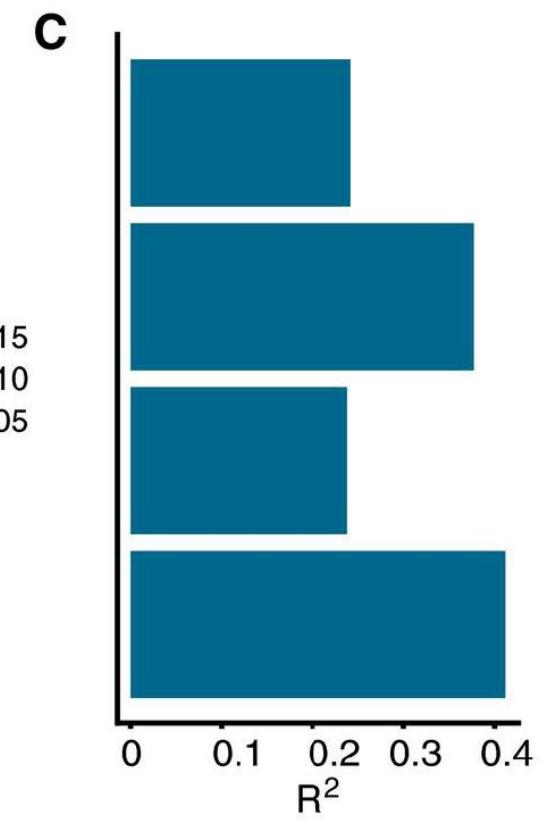
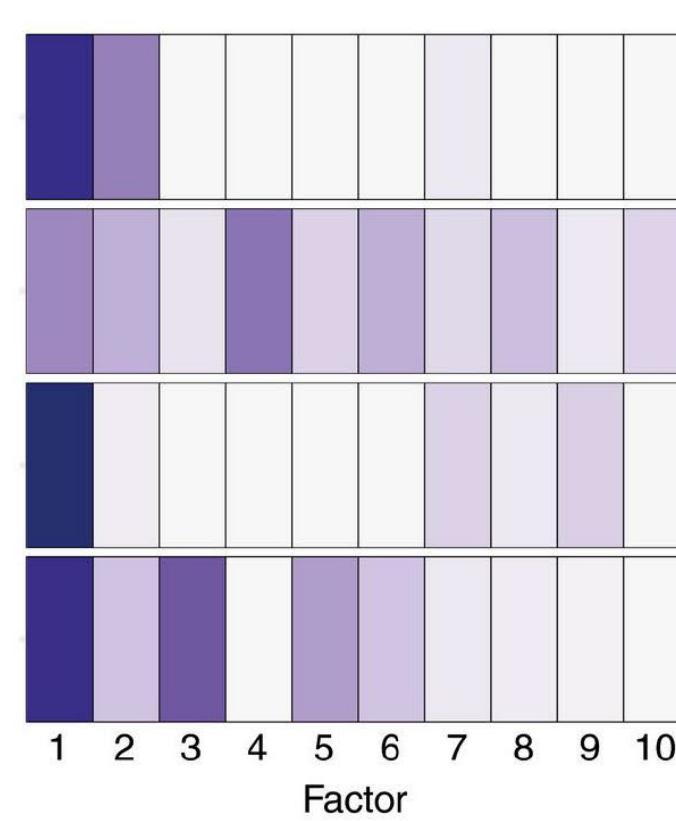
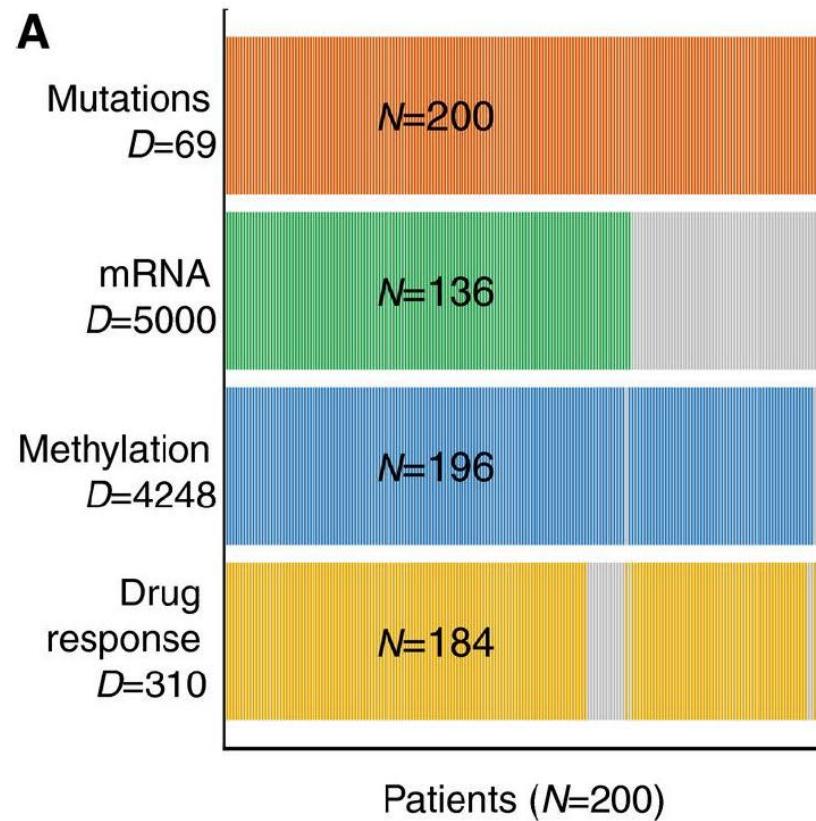
- Filtered
- Normalized/ Scaled
- Paired

MOFA



$$Y_m = ZW^{mT}$$

MOFA



MOFA

Paper:

<https://www.embopress.org/doi/full/10.15252/msb.20178124>

Website: <https://biofam.github.io/MOFA2/>

DOMINO

Welcome to **DOMINO** Web Executor



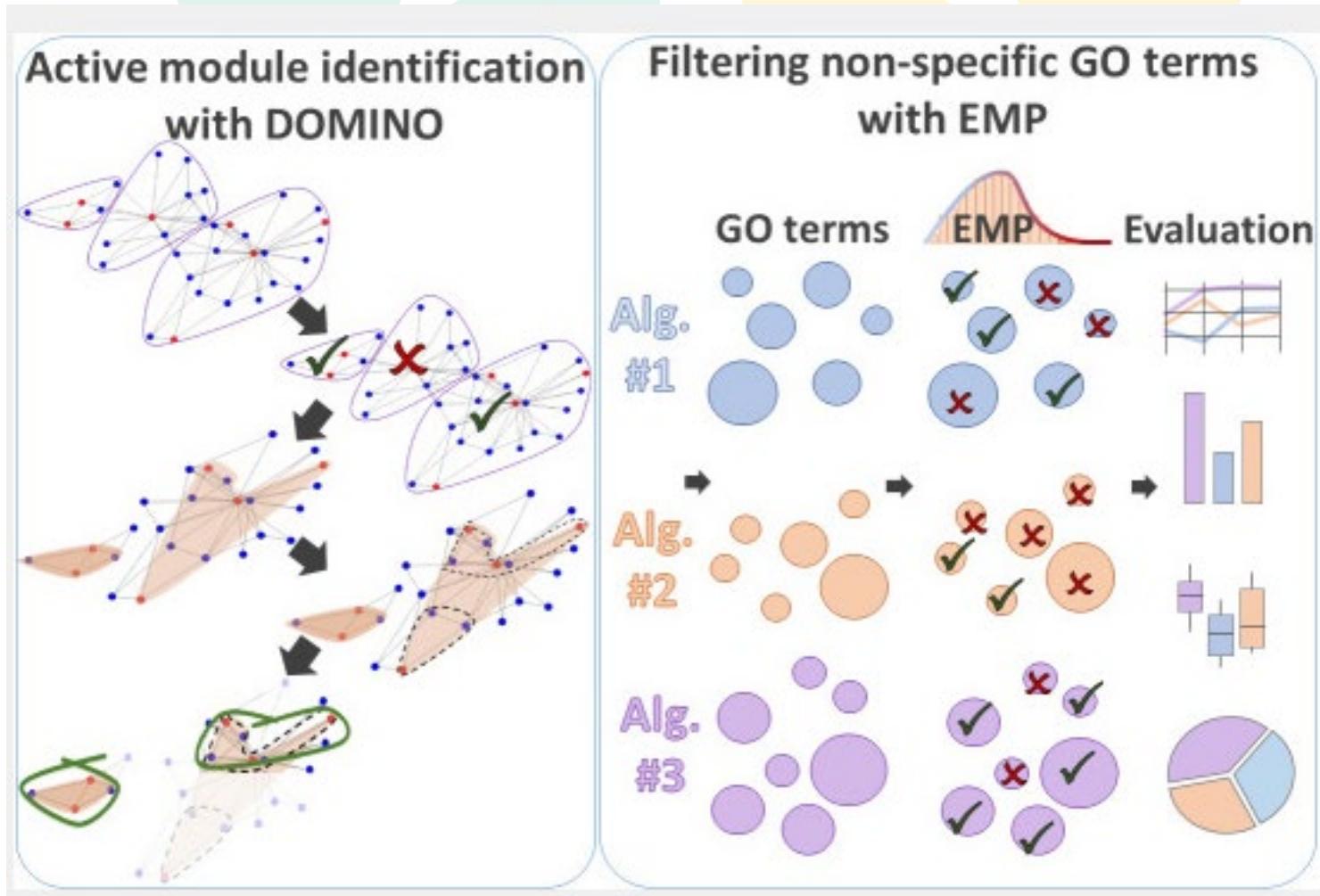
DOMINO receives sets of active genes and reports Active Modules, i.e., connected subnetworks that are enriched for the active genes.

Run DOMINO

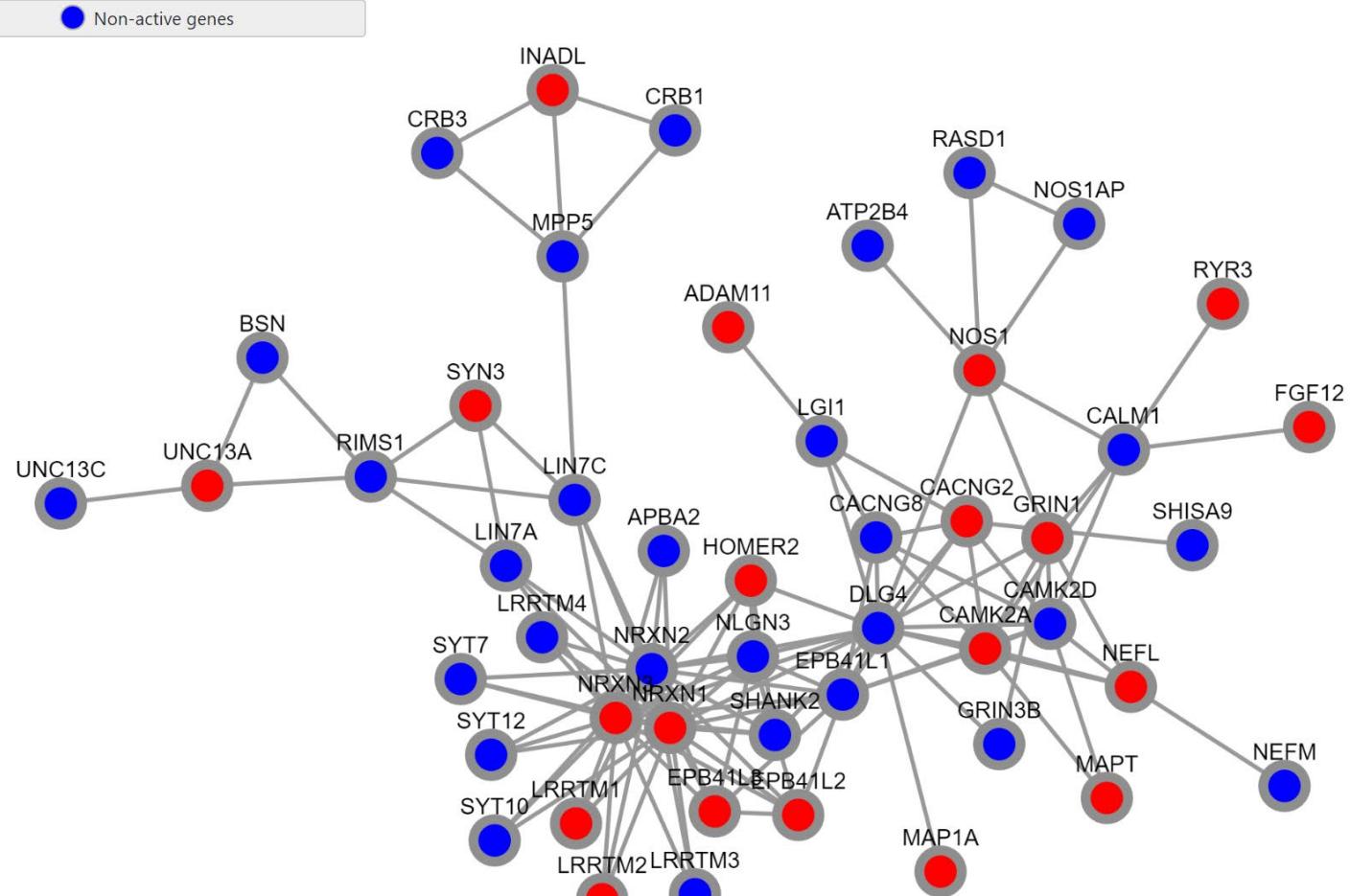
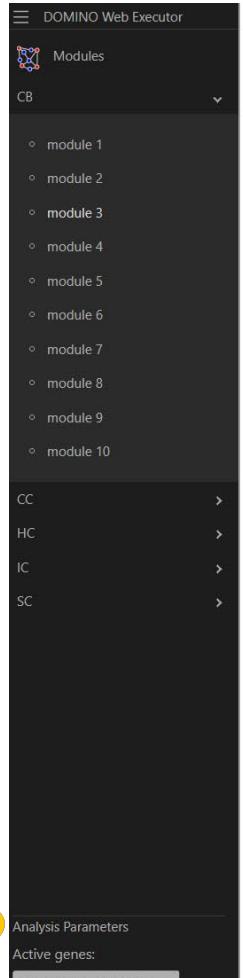


Funded by the
European Union
GA n°825575

DOMINO



DOMINO



DOMINO

[show details for all genes/nodes](#)

Total number of genes: 49

Show 10 ↓ entries

Search:

Gene name	ENSG id
CACNG2	ENSG00000166862
EPB41L1	ENSG0000088367
NOS1AP	ENSG00000198929
NOS1	ENSG00000089250
BSN	ENSG00000164061
UNC13A	ENSG00000130477
EPB41L3	ENSG0000082397
NLGN3	ENSG00000196338
LRRTM2	ENSG00000146006

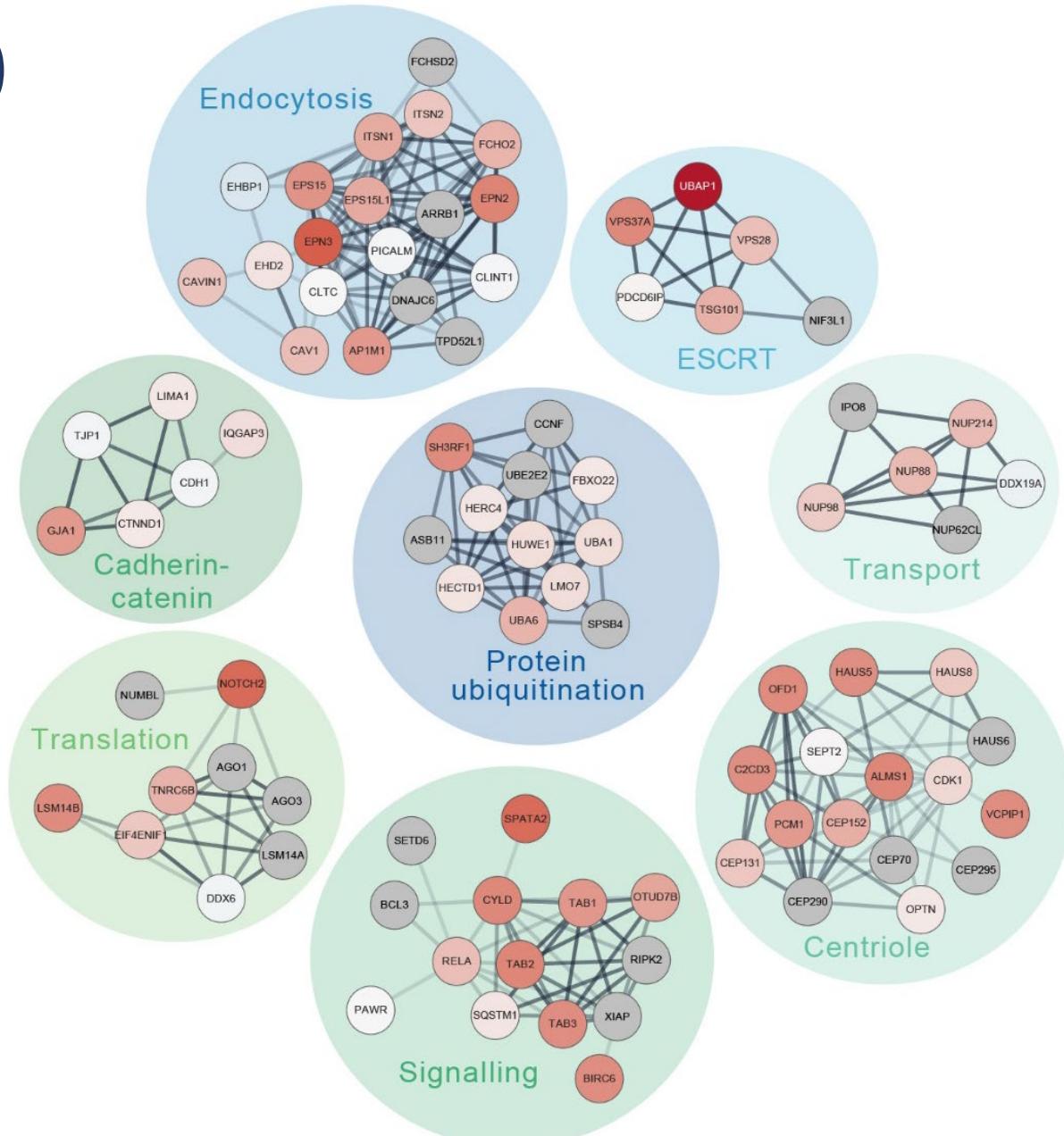
Total number of enriched groups (qval<0.05): 293

Show 10 ↓ entries

Search:

index	GO id	GO term	Ontology	pval	qval
1	GO:0045202	synapse	CC	5.217e-20	9.855e-17
2	GO:0030054	cell junction	CC	1.921e-19	1.814e-16
3	GO:0050804	modulation of chemical synaptic transmission	BP	1.296e-18	1.014e-14
4	GO:0099177	regulation of trans-synaptic signaling	BP	1.373e-18	1.014e-14
5	GO:0099643	signal release from synapse	BP	3.778e-15	1.861e-11
6	GO:0050808	synapse organization	BP	3.203e-14	1.183e-10

DOMINO



DOMINO

Paper:

<https://www.embopress.org/doi/full/10.15252/msb.20209593>

Website: <http://domino.cs.tau.ac.il/>

THANK YOU

www.ejprarediseases.org

coordination@ejprarediseases.org

helpdesk@ejprarediseases.org

Follow us on social media



@EJPRarediseases



The EJP RD initiative has received funding from the European Union's
Horizon 2020 research and innovation programme under grant
agreement N°825575