

NETWORK-GUIDED INTEGRATION OF MULTI-OMICS DATA

Towards a comprehensive
view of cancer

Laura Cantini

Computational Systems Biology Team

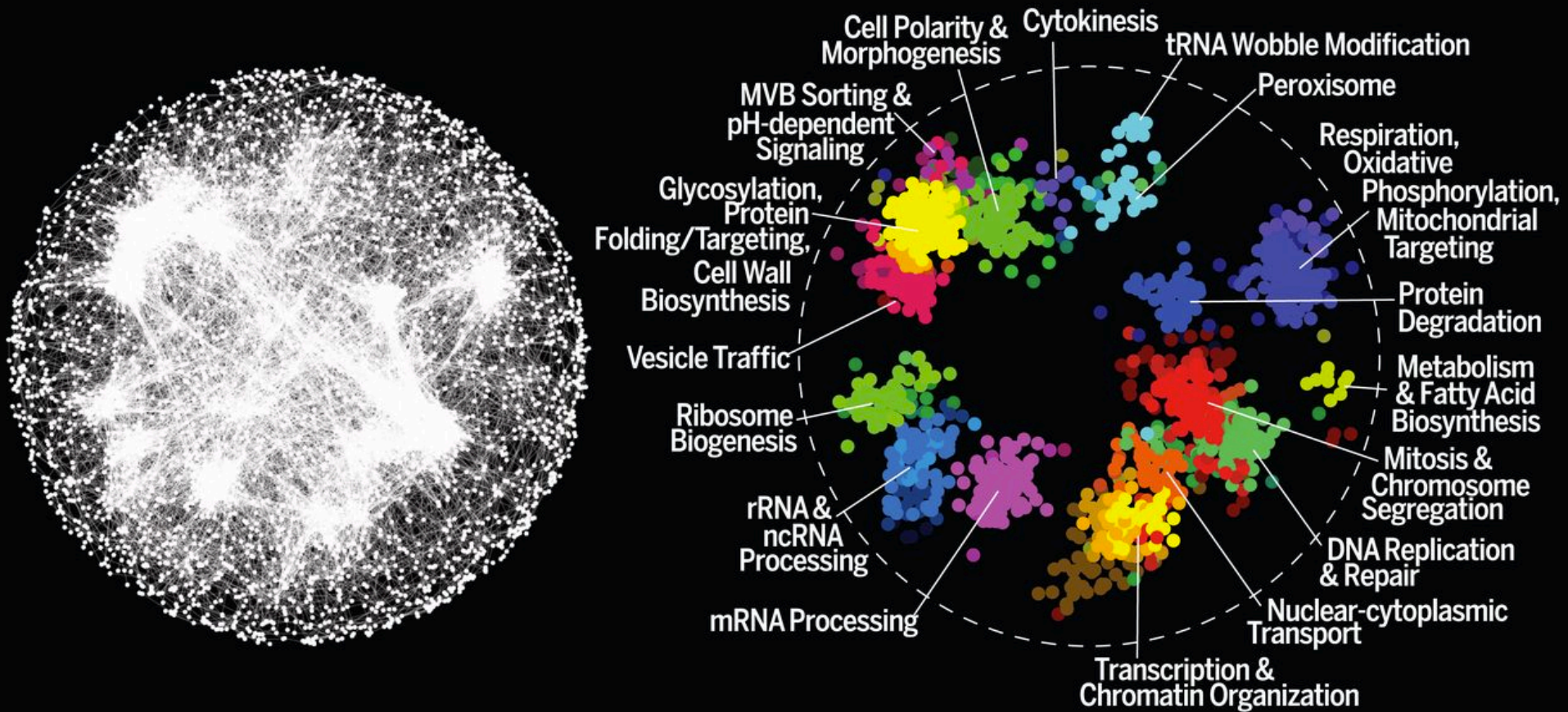
IBENS, Paris

09-04-2019

Bioinfo4Women, BSC



The cell is a complex system

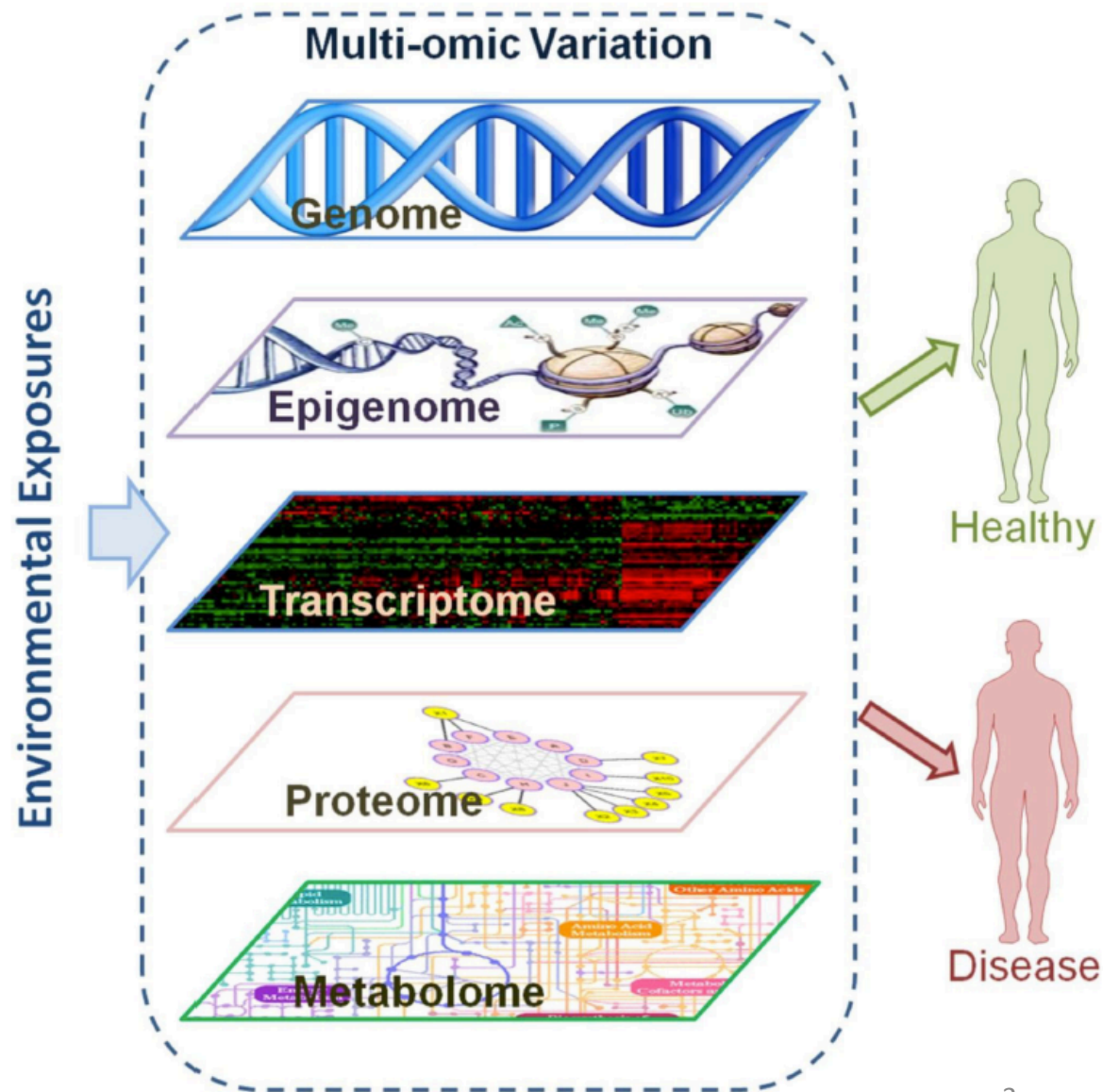


We cannot explain the genotype-phenotype connection or diseases (e.g. cancer) with a single molecular layer

Multi-omics integration in cancer

Lot of omics data available (more in the future)

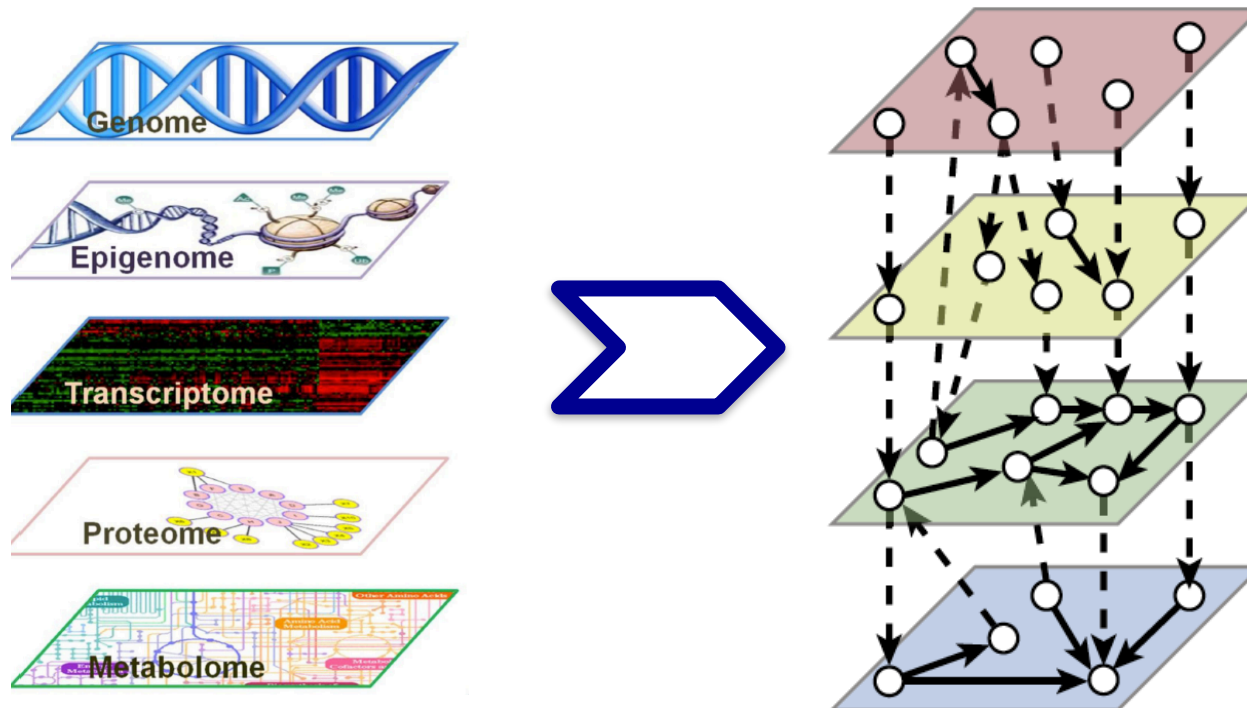
We miss integrative methodologies



Multi-omics integration in cancer

Challenges: large number and heterogeneity of variables, noise of data

Methodological aim: to combine dimensionality reduction and enhance Network-based approaches



Multi-omics integration in cancer

1. MicroRNA-mRNA interactions in colorectal cancer subtypes
2. Cancer data integration with multiplexes
3. Dimensionality reduction to reconstruct cancer processes

MicroRNA-mRNA interactions in colorectal cancer subtypes:

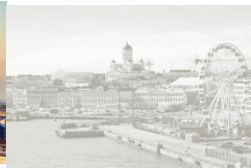
FLORENCE



TURIN



HELSINKI



PARIS



Math
background

PhD
Complexity in
life sciences

Visiting PhD
Network
science

Postdoc
Institut Curie

CNRS
Permanent
Researcher

Enzo Medico's lab
Candiolo Cancer Institute

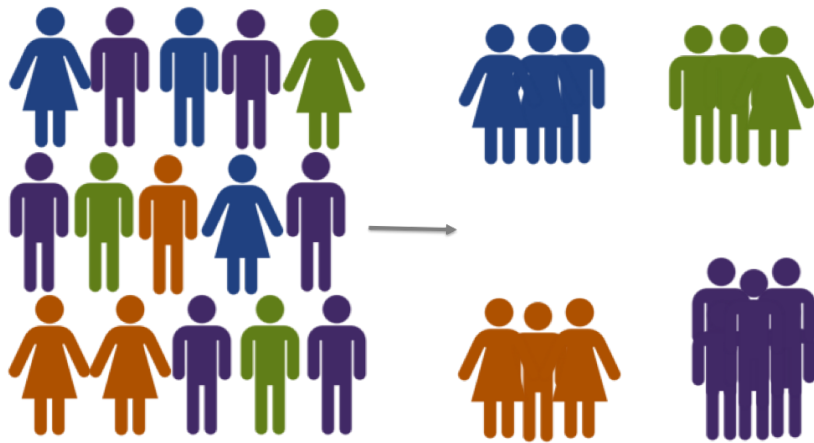


Michele Caselle
Physics
Department



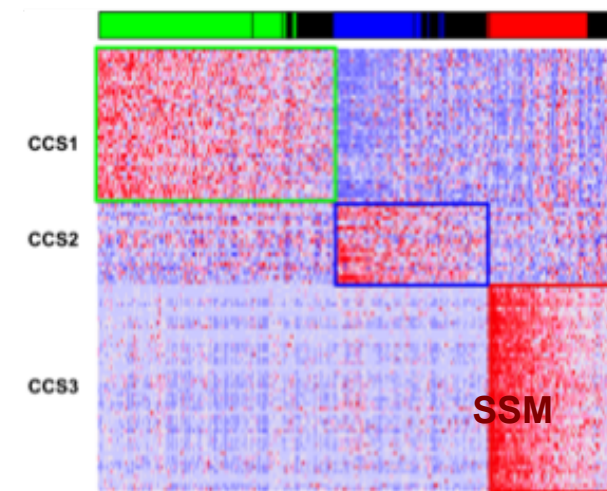
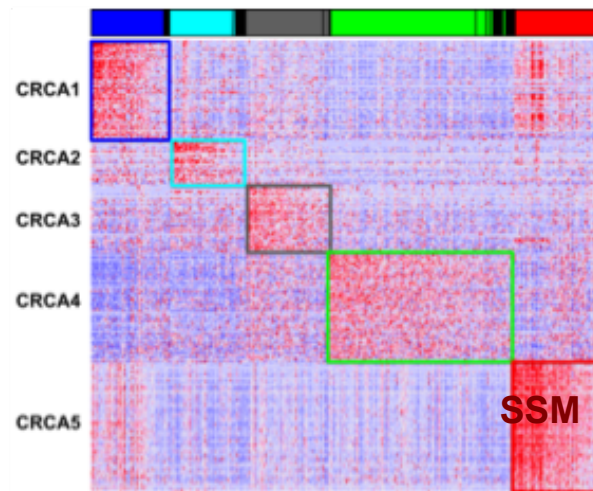
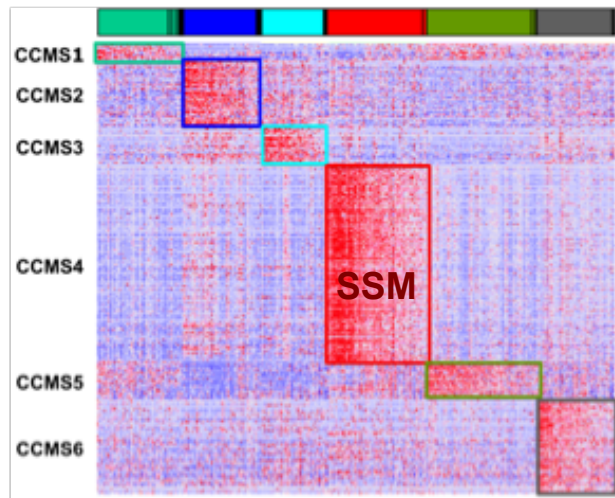
Elisa Ficarra
Computer
Science
Department

Colorectal Cancer (CRC) subtypes



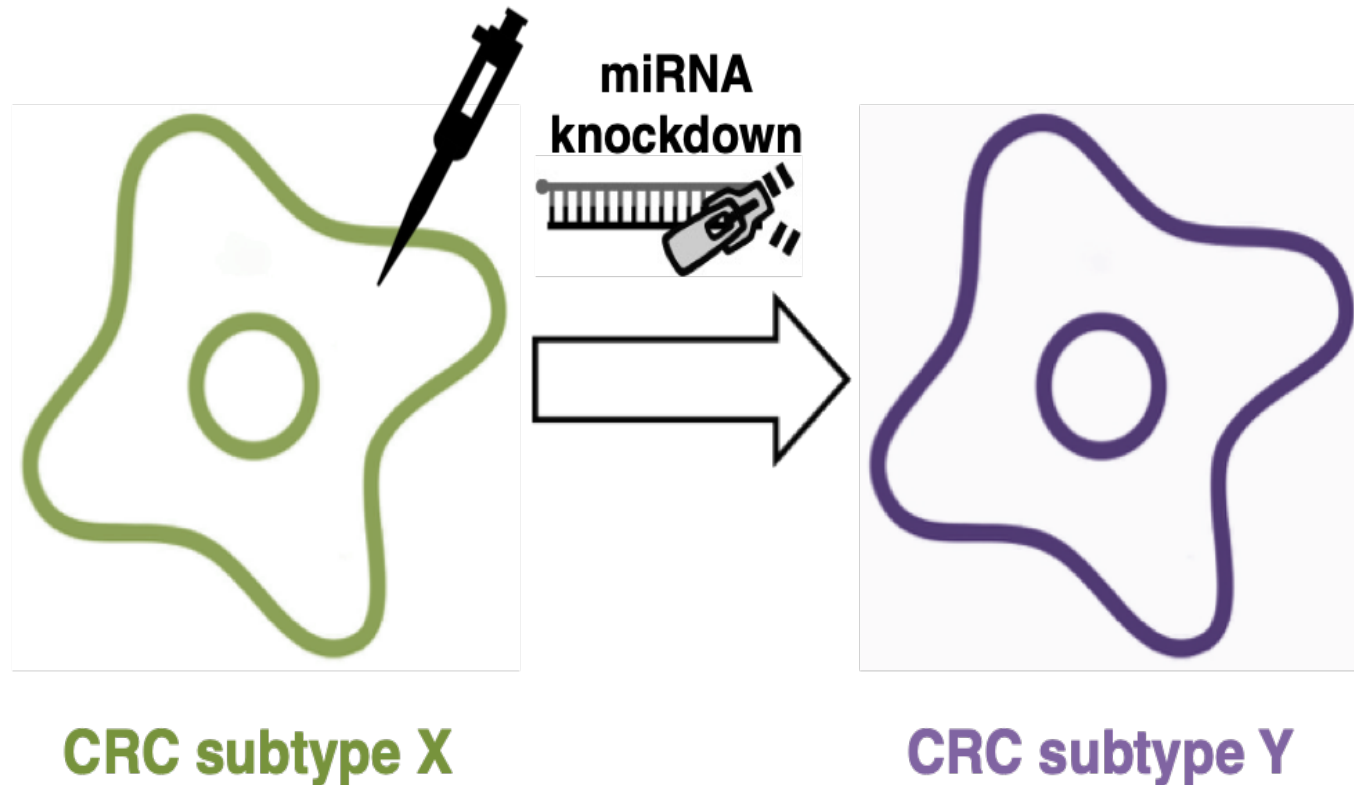
A cancer subtype is a set of homogeneous patients.

Cancer subtyping is important for clinical management.



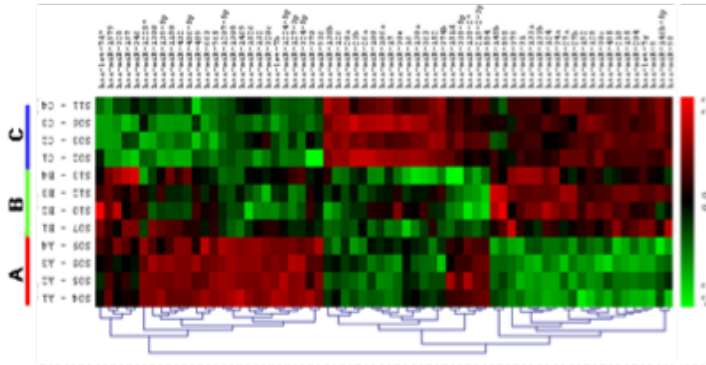
SSM Stem Serrated mesenchymal CRC subtype

Identify microRNAs driving the CRC subtypes

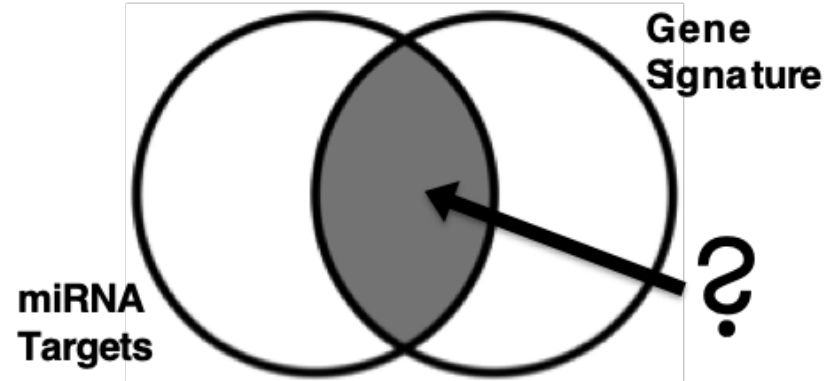


MicroRNA Master Regulator Analysis (MMRA)

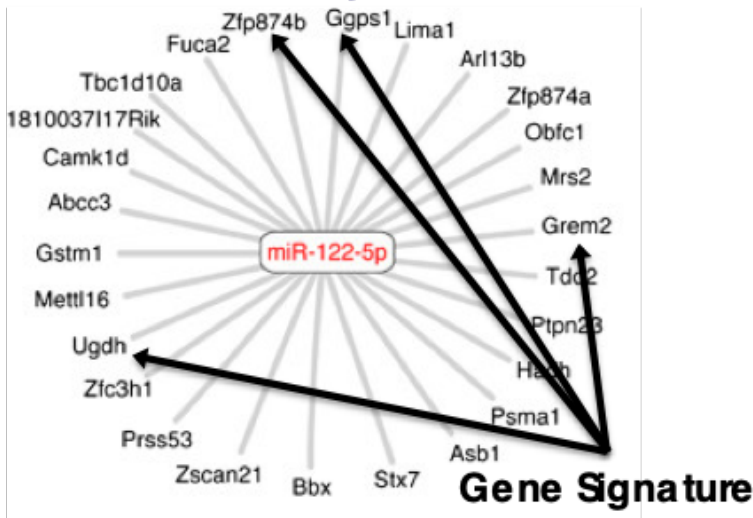
1. MicroRNA differential expression analysis



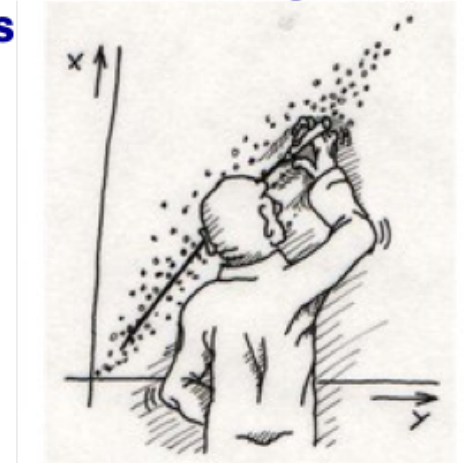
2. MicroRNA target enrichment analysis



3. Network analysis

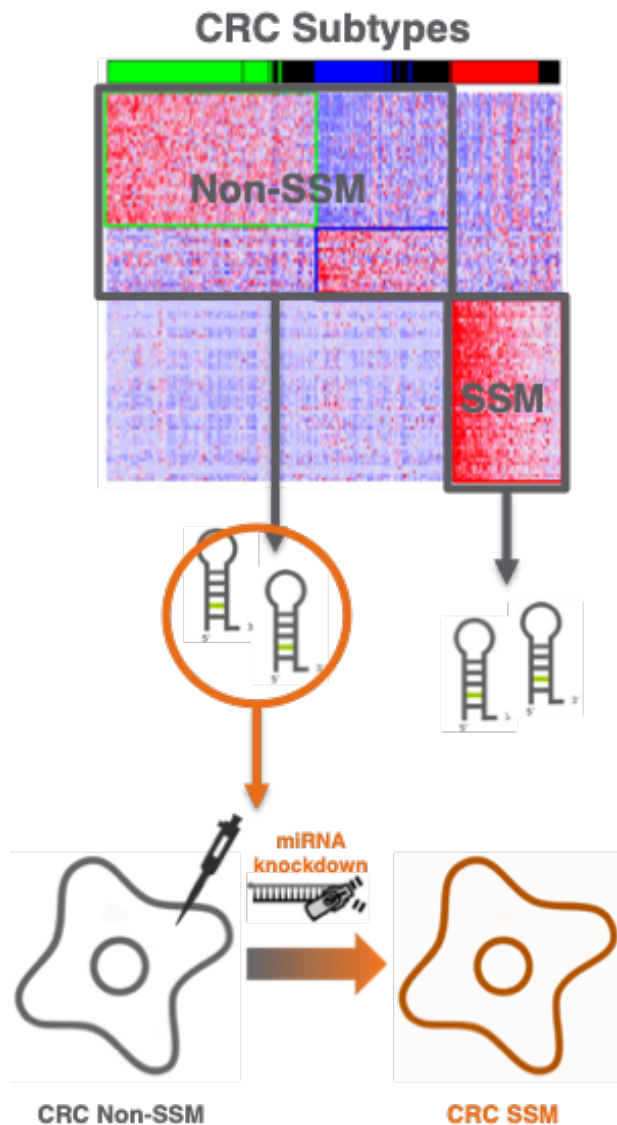


4. Stepwise Linear Regression (SLR) analysis



$$\text{Gene}_x = a_1 \text{mir}_y + a_2 \text{mir}_z + \dots$$

Many microRNAs regulate SSM



24 microRNAs underlying the CRC subtypes identified from TCGA data

Results confirmed on a panel of ~100 cell lines

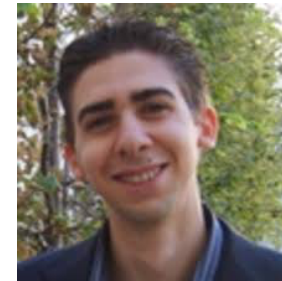
Experimental validations then performed to test *in silico* predictions

Validations in cell lines

Cell line	HT29			NCIH508				SW403
Targeted microRNA	mir-194	mir-200b	mir-429	mir-194	mir-200b	mir-429	mir-203	mir-429
Upregulated genes (total)	252	567	163	20	6	20	32	104
Fold enrichment in SSM genes	5.23	3.01	2.85	5.9	19.7	59.2	33.3	7.55
Enrichment <i>P</i> value	2.4E-08	2.9E-06	1.4E-02	0.14	0.049	2.2E-16	3.8E-12	0.0001
Downregulated genes (total)	244	411	115	6	1	2	11	83
Fold enrichment in non-SSM genes	6.65	1.67	0.54	8.8	0	5.9	1.8	4.4
Enrichment <i>P</i> value	1.6E-14	0.034	0.29	9.3E-05	0.983	2.9E-04	0.158	0.002



Gabriele Picco



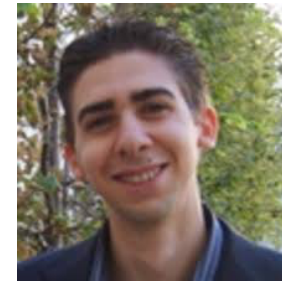
Consalvo Petti

Validations in cell lines

Cell line	HT29			NCIH508				SW403
Targeted microRNA	mir-194	mir-200b	mir-429	mir-194	mir-200b	mir-429	mir-203	mir-429
Upregulated genes (total)	252	567	163	20	6	20	32	104
Fold enrichment in SSM genes	5.23	3.01	2.85	5.9	19.7	59.2	33.3	7.55
Enrichment <i>P</i> value	2.4E-08	2.9E-06	1.4E-02	0.14	0.049	2.2E-16	3.8E-12	0.0001
Downregulated genes (total)	244	411	115	6	1	2	11	83
Fold enrichment in non-SSM genes	6.65	1.67	0.54	8.8	0	5.9	1.8	4.4
Enrichment <i>P</i> value	1.6E-14	0.034	0.29	9.3E-05	0.983	2.9E-04	0.158	0.002
Original subtype	SSM			Non-SSM				Non-SSM
Original FDR	0.88			0.28				0.002
New subtype	SSM	SSM	SSM	SSM	SSM	SSM	SSM	Non-SSM
New FDR	0.200	0.001	0.005	0.004	0.005	0.004	0.002	0.002



Gabriele Picco



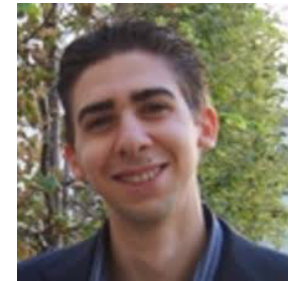
Consalvo Petti

Validations in cell lines

Cell line	HT29			NCIH508				SW403
Targeted microRNA	mir-194	mir-200b	mir-429	mir-194	mir-200b	mir-429	mir-203	mir-429
Upregulated genes (total)	252	567	163	20	6	20	32	104
Fold enrichment in SSM genes	5.23	3.01	2.85	5.9	19.7	59.2	33.3	7.55
Enrichment <i>P</i> value	2.4E-08	2.9E-06	1.4E-02	0.14	0.049	2.2E-16	3.8E-12	0.0001
Downregulated genes (total)	244	411	115	6	1	2	11	83
Fold enrichment in non-SSM genes	6.65	1.67	0.54	8.8	0	5.9	1.8	4.4
Enrichment <i>P</i> value	1.6E-14	0.034	0.29	9.3E-05	0.983	2.9E-04	0.158	0.002
Original subtype	SSM			Non-SSM				Non-SSM
Original FDR	0.88			0.28				0.002
New subtype	SSM	SSM	SSM	SSM	SSM	SSM	SSM	Non-SSM
New FDR	0.200	0.001	0.005	0.004	0.005	0.004	0.002	0.002



Gabriele Picco



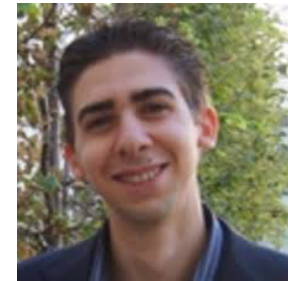
Consalvo Petti

Validations in cell lines

Cell line	HT29			NCIH508				SW403
Targeted microRNA	mir-194	mir-200b	mir-429	mir-194	mir-200b	mir-429	mir-203	mir-429
Upregulated genes (total)	252	567	163	20	6	20	32	104
Fold enrichment in SSM genes	5.23	3.01	2.85	5.9	19.7	59.2	33.3	7.55
Enrichment <i>P</i> value	2.4E-08	2.9E-06	1.4E-02	0.14	0.049	2.2E-16	3.8E-12	0.0001
Downregulated genes (total)	244	411	115	6	1	2	11	83
Fold enrichment in non-SSM genes	6.65	1.67	0.54	8.8	0	5.9	1.8	4.4
Enrichment <i>P</i> value	1.6E-14	0.034	0.29	9.3E-05	0.983	2.9E-04	0.158	0.002
Original subtype	SSM			Non-SSM				Non-SSM
Original FDR	0.88			0.28				0.002
New subtype	SSM	SSM	SSM	SSM	SSM	SSM	SSM	Non-SSM
New FDR	0.200	0.001	0.005	0.004	0.005	0.004	0.002	0.002



Gabriele Picco



Consalvo Petti

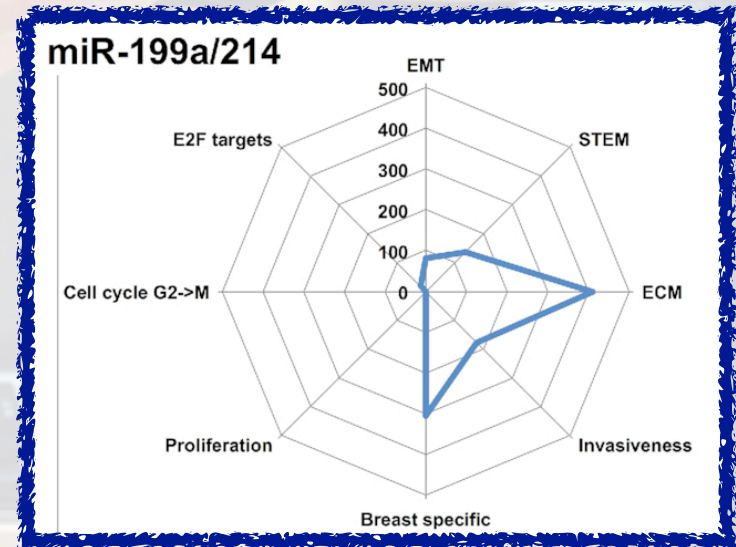
From MMRA to clustMMRA

MMRA has been extended to clustMMRA (<https://github.com/lcan88/clustMMRA>)

clustMMRA has been applied to two independent BRCA datasets

11 genomic clusters of microRNAs have been identified as putative driver of BRCA subtypes

miR-199a/214 validated in cell lines

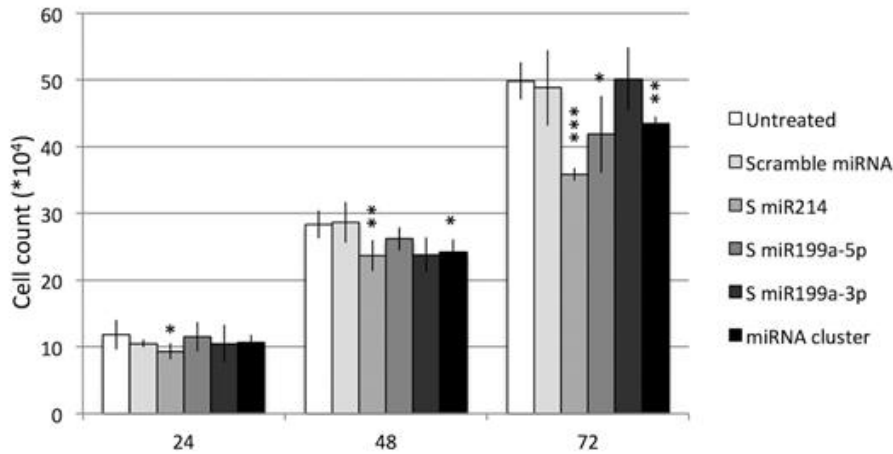


miR-199a/214 validated in MDA-MB-231

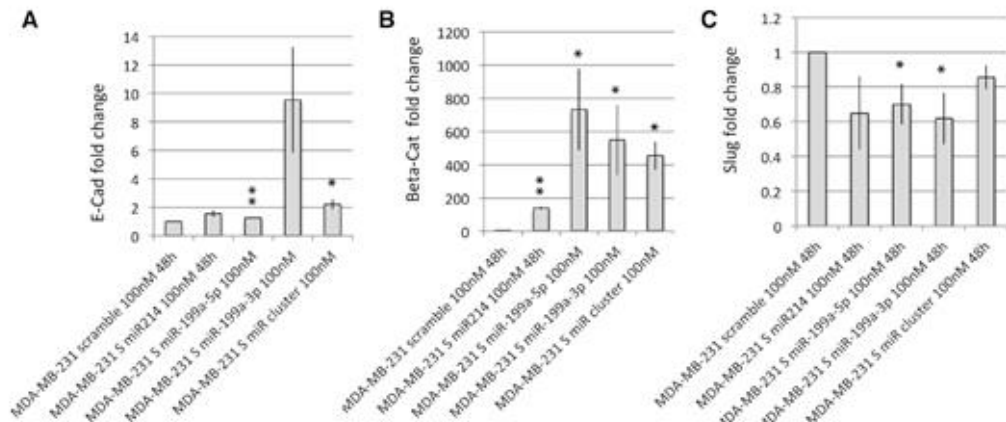


Gloria Bertoli
Isabella Castiglioni
Claudia Cava
CNR, Milan.

Proliferation



EMT



Cancer data integration with multiplexes:

FLORENCE



TURIN



HELSINKI



PARIS



Math
background

PhD
Complexity in
life sciences

**Visiting PhD
Network
science**

Postdoc
Institut Curie

CNRS
Permanent
Researcher

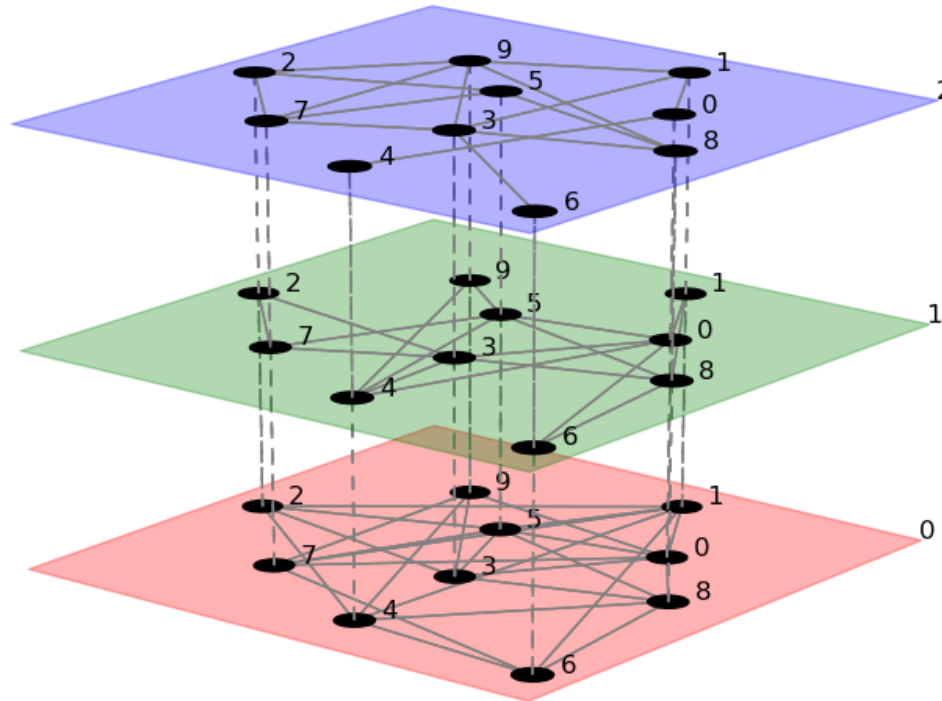


Santo Fortunato
Indiana
University



Michele Caselle
Physics
Department

The multiplex framework



Layers of multiplex:

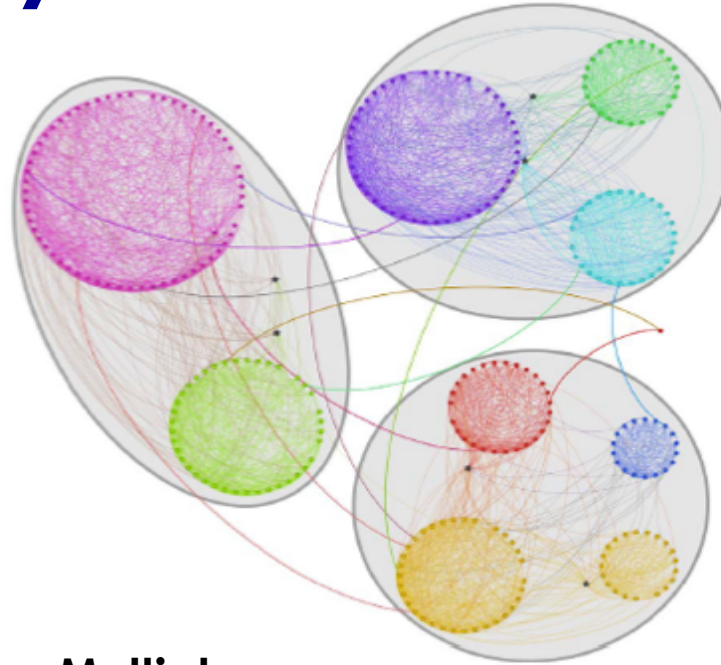
1: Gene co-expression network

2: Regulatory network (Transcription Factors)

3: Regulatory network (microRNAs)

4: Protein-protein interactions (PPIs)

Community detection on multiplex



Community detection on Multiplex:

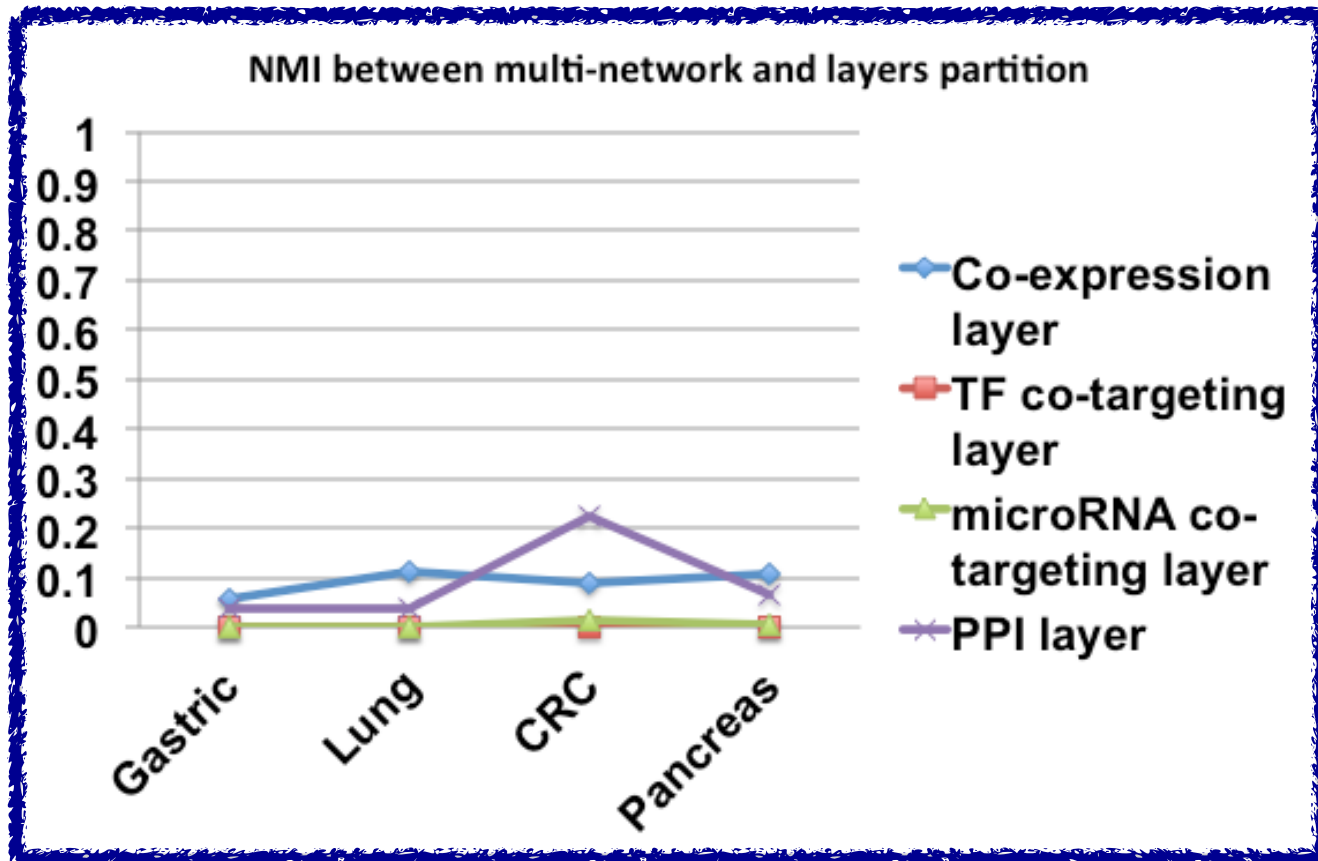
1. Community detection within each layer (Infomap, OSLOM , Label propagation, Louvain, Modularity optimization via simulated annealing)
2. Consensus clustering (Lancichinetti et al., Scientific reports, 2012).

Summarized into Gene4X (<https://github.com/lcan88/Gene4x>)

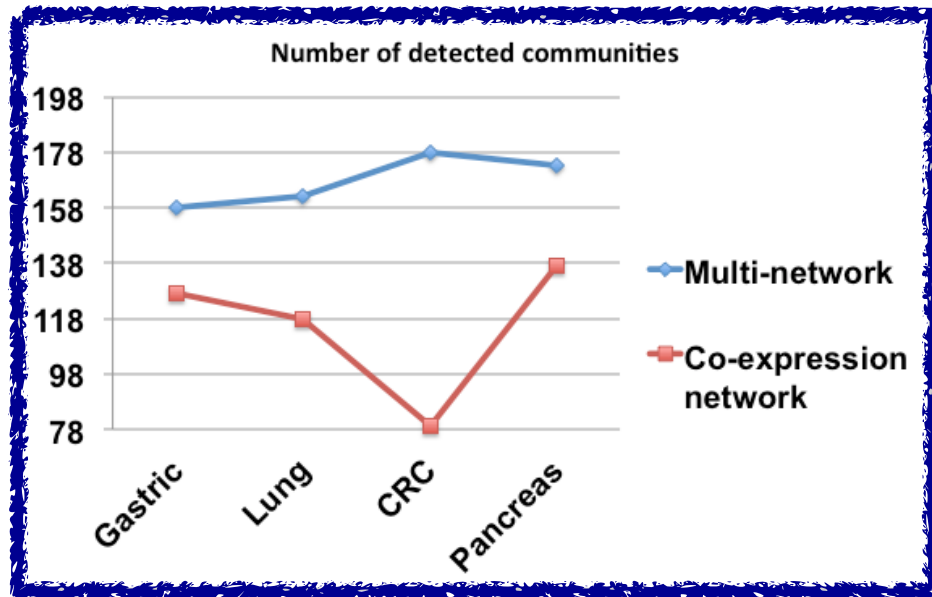
Gene4x results in cancer

The approach has been applied to four cancers (gastric, lung, arc and pancreas)

Multiplex results are not driven by a specific layer:

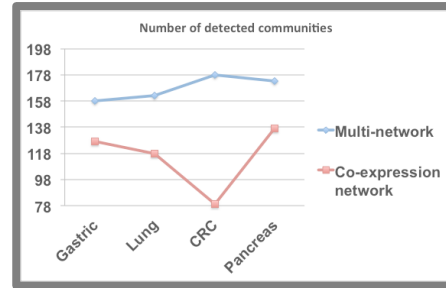


Multiplex vs. single network

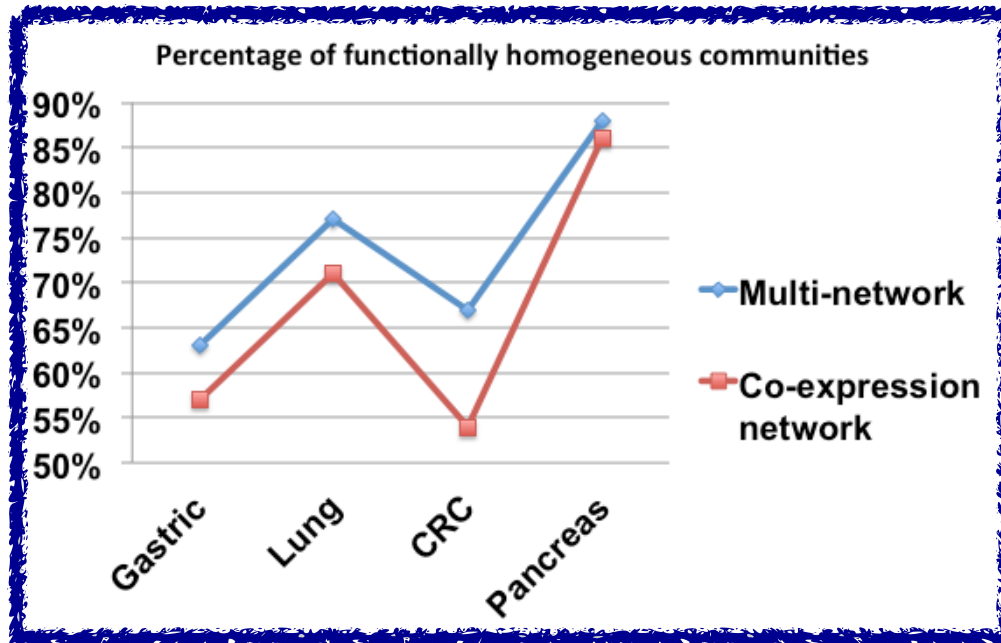


1. The multiplex detects more communities than the co-expression network.

Multiplex vs. single network

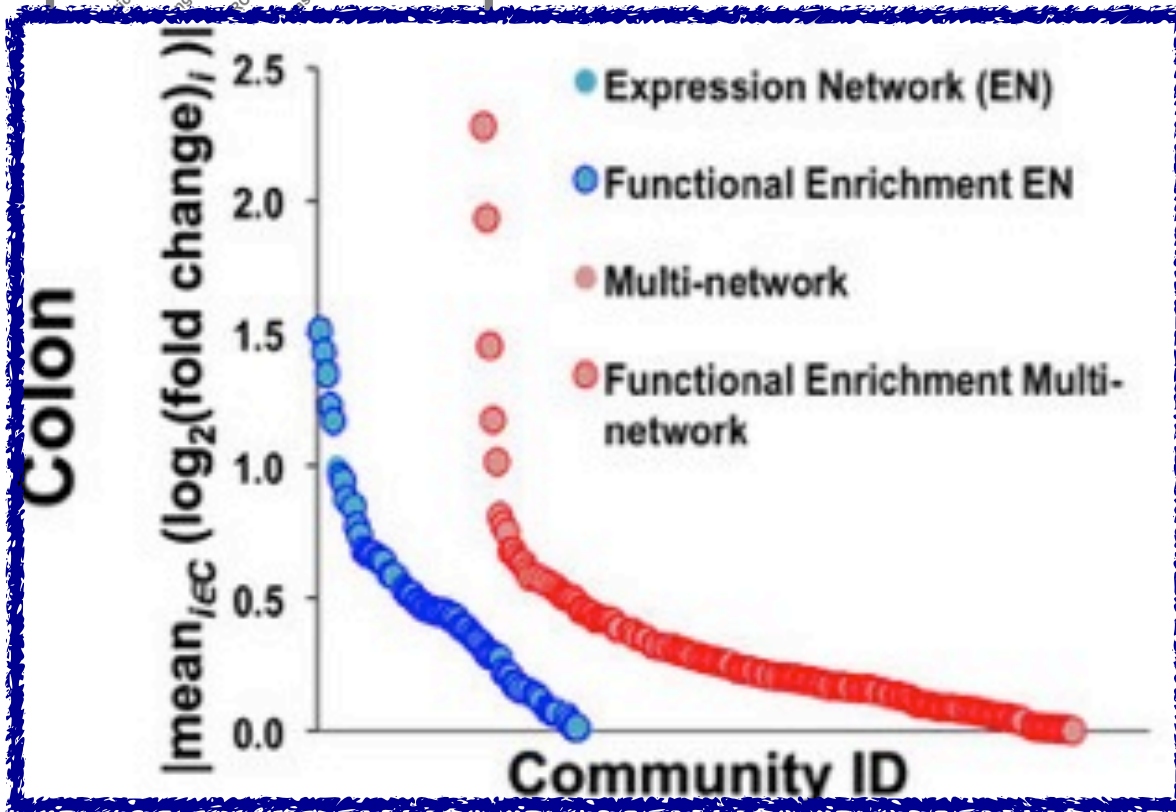
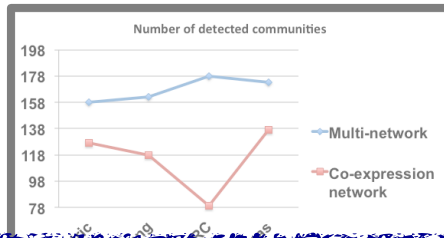


1. The multiplex detects more communities than the co-expression network.



2. The communities of the multiplex more frequently reflect functional modules.

Multiplex vs. single network



1. The multiplex detects more communities than the co-expression network.

2. The communities of the multiplex more frequently reflect functional modules.

3. The communities of the multiplex more frequently reflect tumor vs. normal differential modules.

Dimensionality reduction to reconstruct cancer processes:

FLORENCE



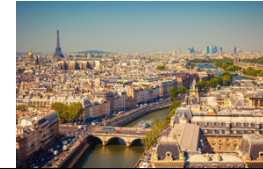
TURIN



HELSINKI



PARIS



Math
background

PhD
Complexity in
life sciences

Visiting PhD
Network
science

Postdoc
Institut Curie

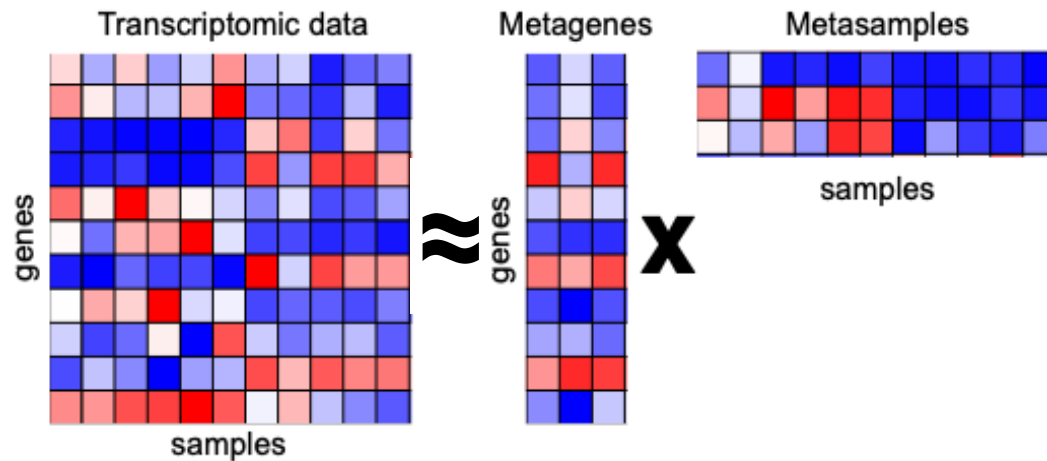
CNRS
Permanent
Researcher

Sysbio Institut Curie



Matrix Factorization (MF)

Current omic datasets are huge, with MF we reduce their dimensionality without losing biological information



$$X \approx AS$$

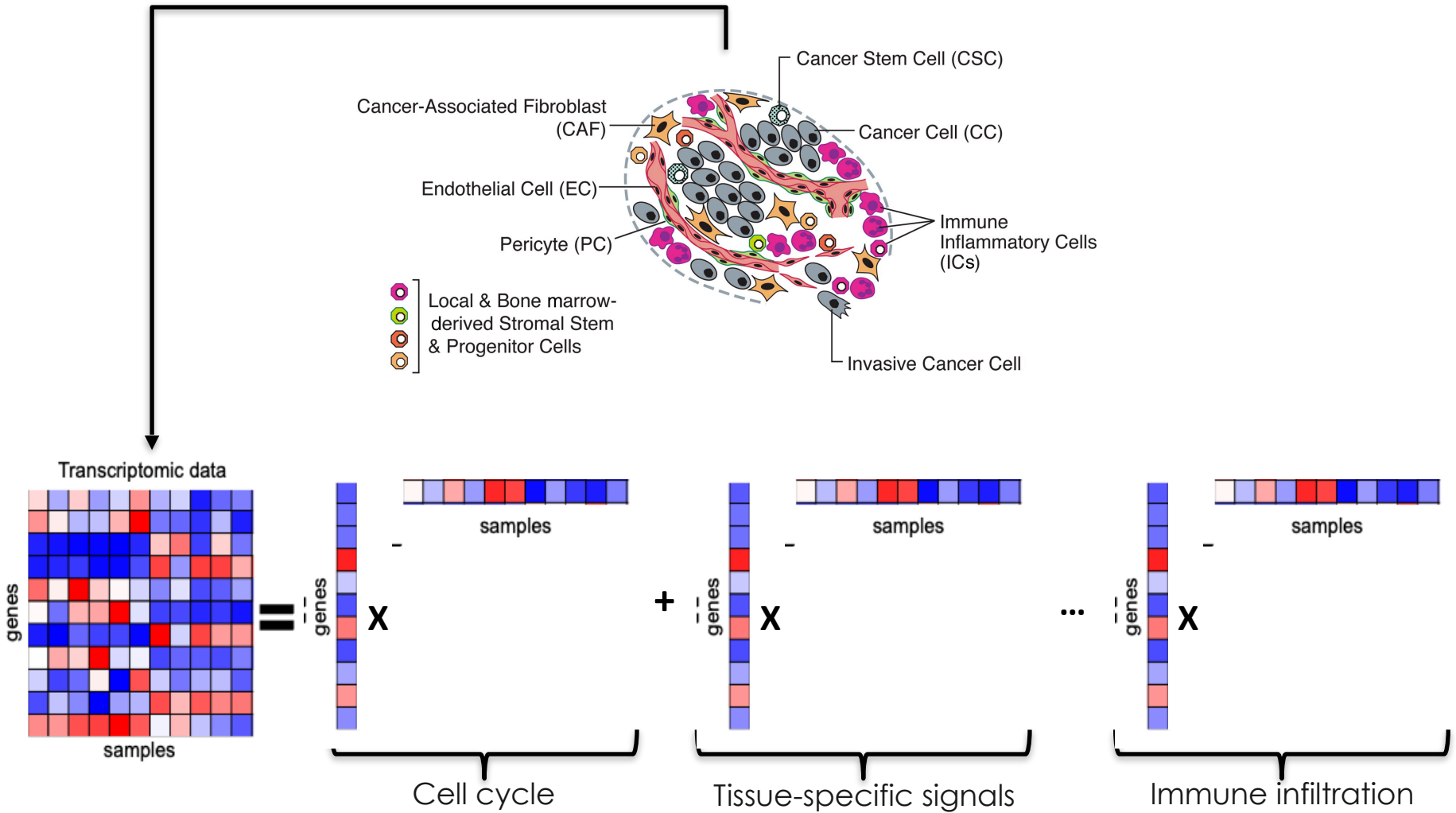
with

X transcriptomic dataset (genes x samples)

A matrix of metagenes (genes x #components)

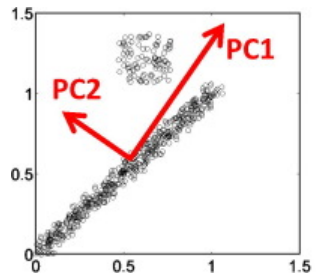
S matrix of meta samples (samples x #components)

Matrix Factorization (MF)

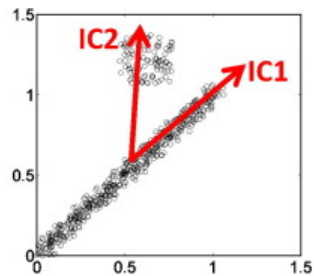


Matrix Factorization (MF)

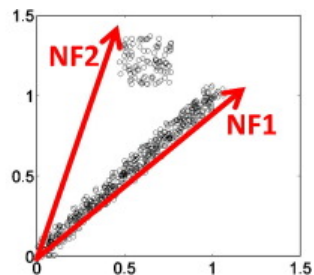
State-of-art MF approaches:



**Principal Component
Analysis (PCA)**



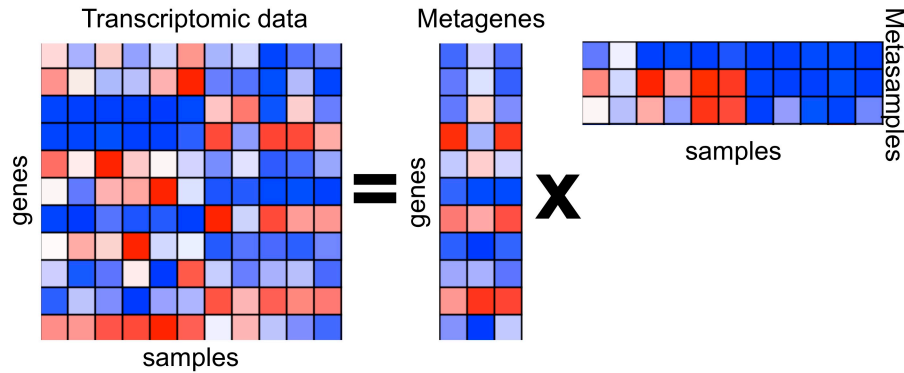
**Independent Component
Analysis (ICA)**



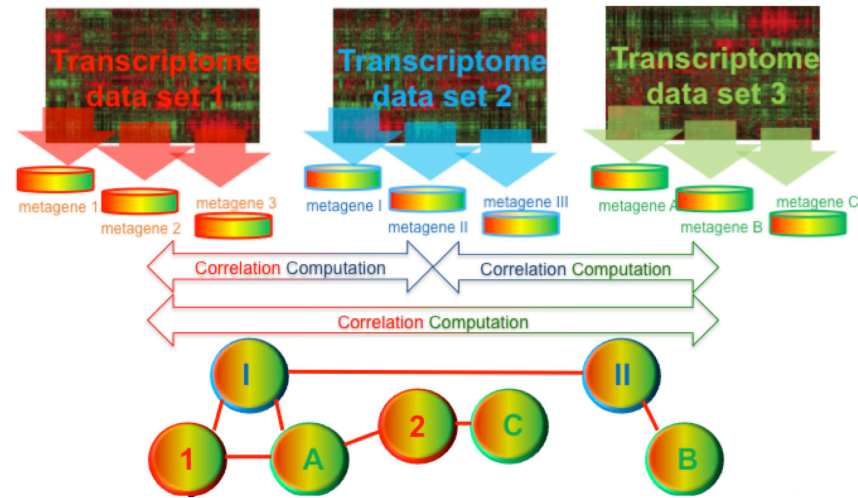
**Non-negative Matrix
Factorization (NMF)**

Assessing reproducibility of MF

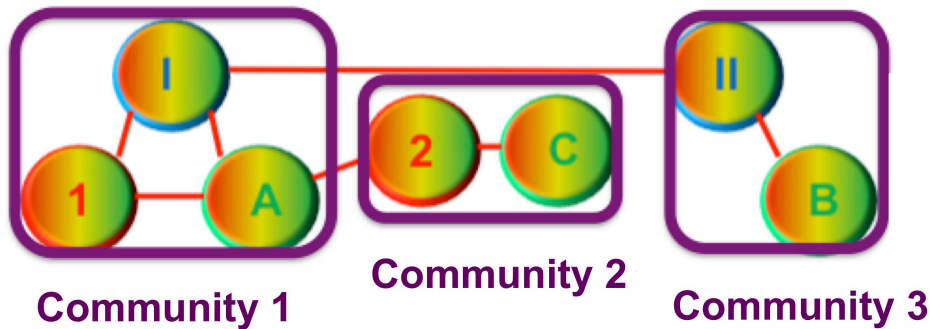
Step1. Decomposition of each dataset



Step2. Construction of the RBH network



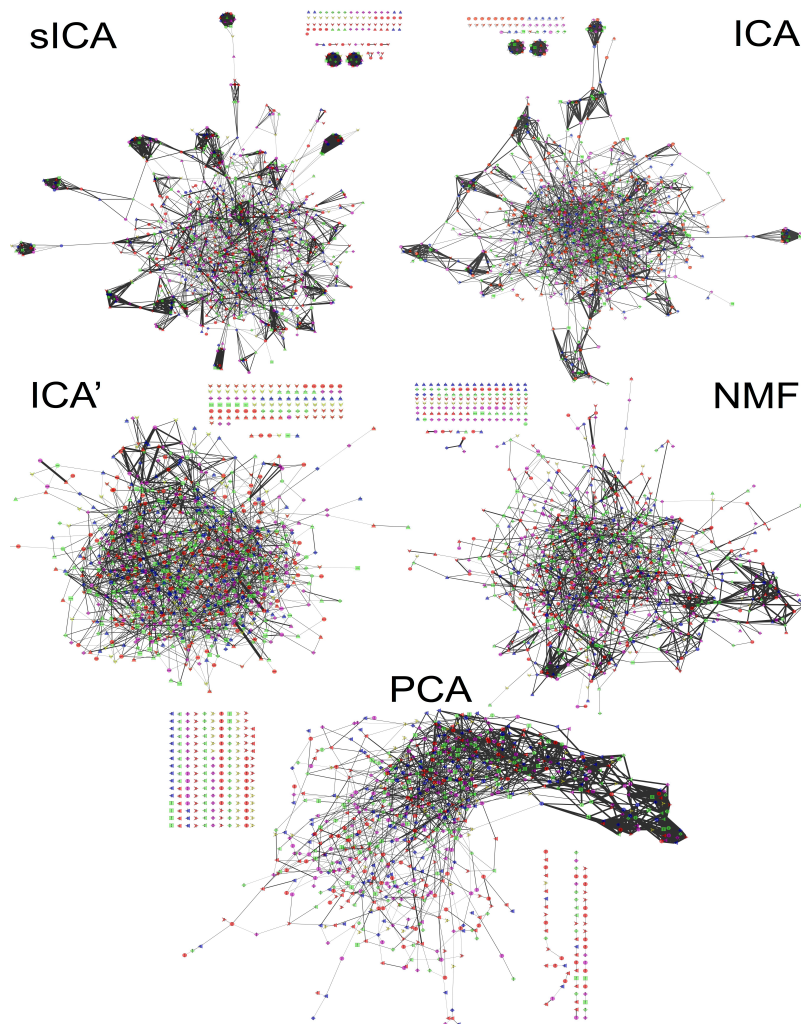
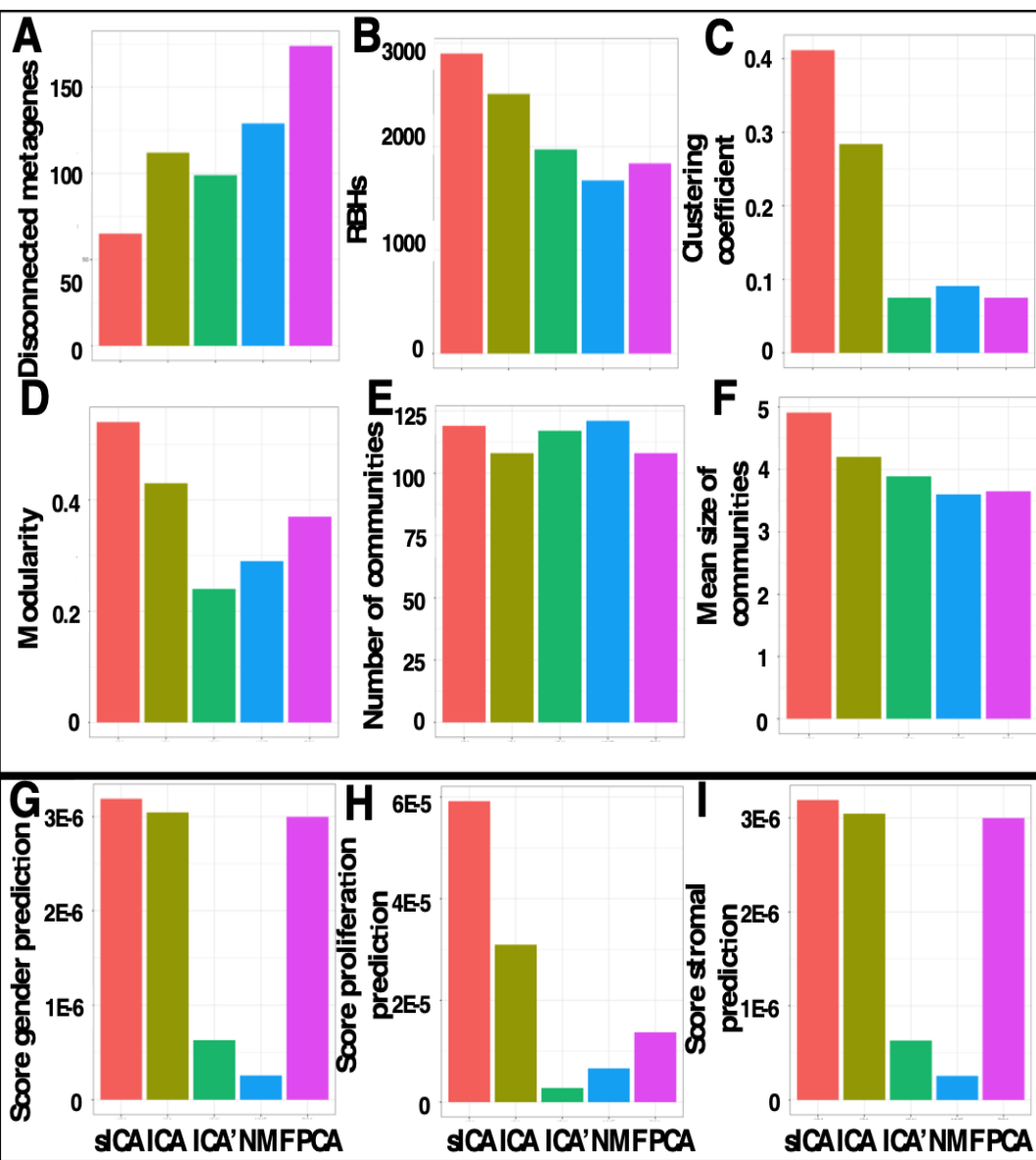
Step3. Community detection



Step4. Comparison of MF methods

- Reproducibility in at least another dataset
- Wide across-datasets reproducibility
- Tightness of the community structure in the RBH graph
- Biological content and specificity of the components.

Assessing reproducibility of MF

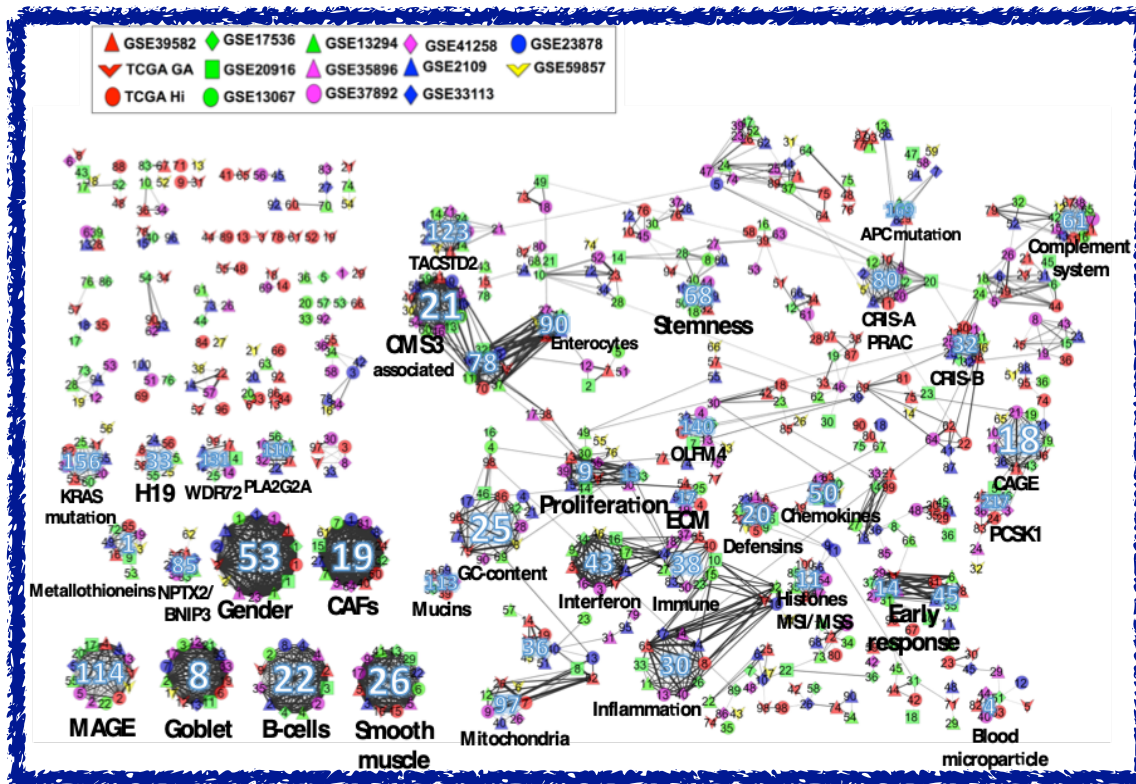


New direction: CRC Consortium

Consensus Molecular Subtypes of CRC (CMS) have been proposed

Few is known about their molecular pathways

Use consensus of different methods to characterize CMS subtypes



Sage Bionetworks

Justin Guinney

Michael Mason

Brian White

Universitair Ziekenhuis Leuven

Sabine Tejpar

Emre Etlioglu

VHIO

Rodrigo Dienstmann

Swiss Institute of Bioinformatics

Mauro Delorenzi

Asa Wirapati

Ligue Nationale Contre le Cancer

Aurélien de Reyniès

Institut Curie

Andrei Zinovyev

Institut de Biologie de l'École Normale

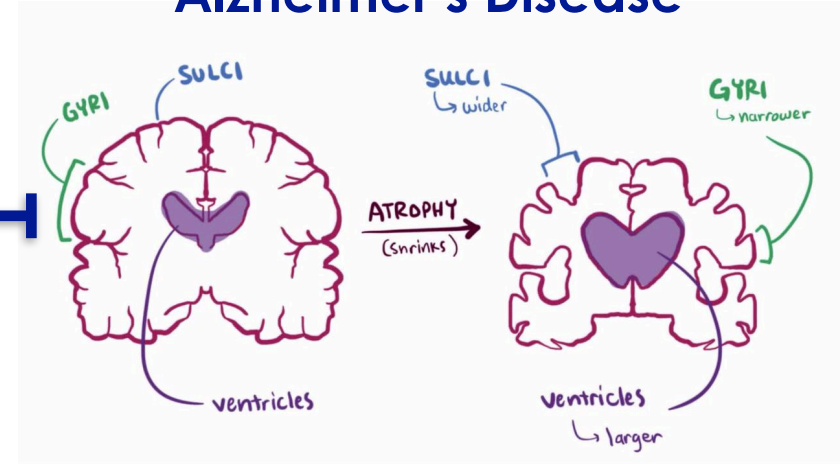
Laura Cantini

New direction: Inverse comorbidity

Lung Cancer



Alzheimer's Disease

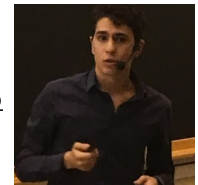


Possible causes



Institut de Biologie de l'Ecole Normale
Laura Cantini

Univeristy of Turin
Michele Caselle
Alessandro Greco



Institut Curie
Andrei Zinovyev
Emmanuel Barillot

Barcelona Supercomputing Center
Jon Sanchez Valle
Alfonso Valencia

CRCT Toulouse
Vera Pancaldi

I2M Marseille
Anaïs Baudot

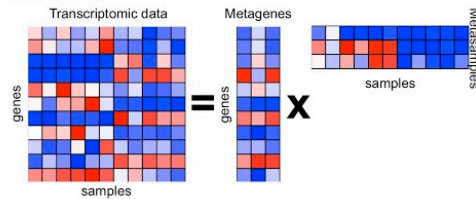
New direction: Inverse comorbidity

Lung C

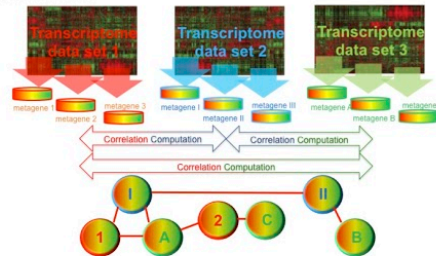


Original Framework

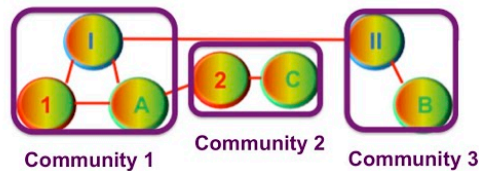
Step1. Decomposition of each dataset



Step2. Construction of the RBH network



Step3. Community detection

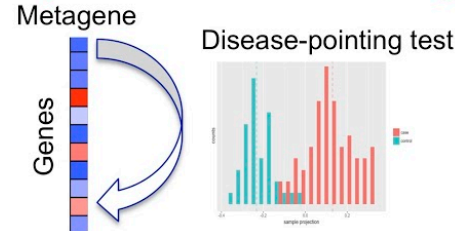


Step4. Comparison of MF methods

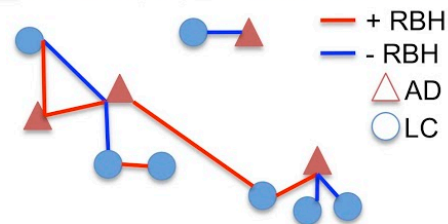
- Reproducibility in at least another dataset
- Wide across-datasets reproducibility
- Tightness of the community structure in the RBH graph
- Biological content and specificity of the components.

Novelties

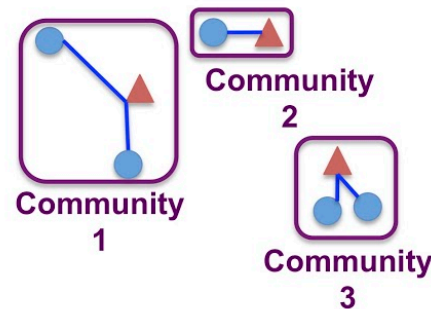
Step1. Orientation of the metagenes



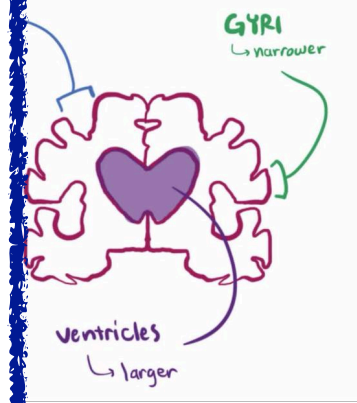
Step2. (+/-)RBH construction



Step3. Extraction of -RBH:AD-LC sub-network and community detection



Disease

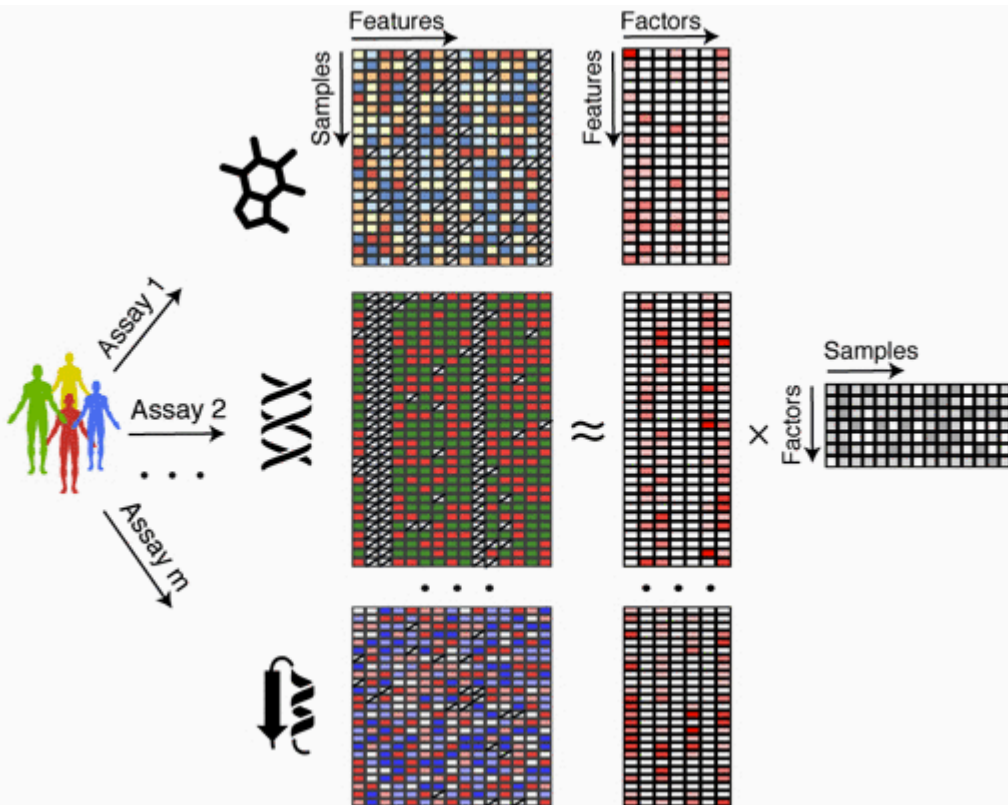


gie de l'Ecole Normale



in
e
CO
v
lot
ercomputing Center
alle
ia

New direction: Multi-omics dimensionality reduction



Few multi-omics dimensionality reduction approaches currently exist.

Our aim:

1. To assess the performances of existing methods (bulk and single-cell)
2. To explore new solutions based on combination of network-theory and matrix factorization

MIT

Manolis Kellis
Alvin Shi
Genevieve Boland

IBENS

Laura Cantini
Denis Thieffry

I2M Marseille

Anaïs Baudot
Elisabeth Remy
Pooya Zakeri

Take home message:

Complex systems (e.g. cancer) cannot be explained by looking at its parts in isolation.

With **omics data** we have good the inputs to achieve this aim.

Methodologies able to insightfully combine the omics layers are now fundamental.

Combining complementary approaches, such as **networks and dimensionality reduction**, can be the key to achieve this aim.

ACKNOWLEDGEMENTS

Part I

Enzo Medico's Lab Candiolo
cancer Institute



Michele Caselle



Elisa Ficarra



Loredana Martignetti
Gloria Bertoli
Isabella Castiglioni
Claudia Cava
CNR, Milan.



Part II

Santo Fortunato



Michele Caselle

Part III

Sysbio,
Institut Curie



Current Team

CSB Team IBENS



Denis THIEFFRY
Group leader,
Prof. ENS



Morgane THOMAS-CHOLLIER
Assoc. Prof. ENS,
IUF Junior
Member



Pierre VINCENS
Assoc. Prof. ENS
and head of the
IBENS computing
platform



Laura CANTINI
Research Scientist
CNRS (CR)



Marika Kapsimali
Research Scientist
INSERM (CR)



Aurélien NALDI
Post doc



Céline HERNANDEZ
Bioinformatics
Engineer / PhD
student



Swann Floc'hlay
PhD student



Nathalie LEHMANN
Engineer



Olivier COLLIN
Assoc. Prof. ENS



Hatim El Jazouli
Master2 student
/Apprentice



Walter Santana
Master2 student



Anais Baudot



Elisabeth Remy





Thanks for the attention

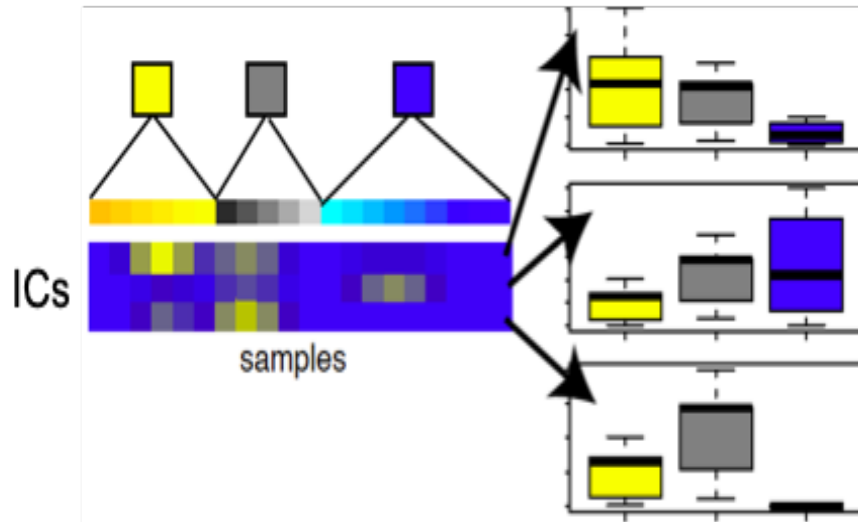
“Behind complexity, there is always simplicity to be revealed.

Inside simplicity, there is always complexity to be discovered.”

Gang Yu

Biological interpretation of sICs

1. Association of the ICs to clinical annotations and cancer subtypes



2. Association of the ICs to pathway collections

