Centre de Recherche en Cancérologie de Marseille
Unité Mixte de Recherche UMR 891

Biomarker Discovery in Breast Cancer by Interactome-Transcriptome Integration

Maxime Garcia (PhD Student)
Integrative Bioinformatics
Marseille Cancer Research Center Inserm U891
Supervisor : Ghislain Bidaut, PhD

Instituts thématiques Inserm Institut national de la santé et de la recherche médicale
INSTITUT PAOLI-CALMETTES Centre régional de lutte contre le cancer Provence-Alpes-Côte d'Azur
UNIVERSITÉ DE LA MÉDITERRANÉE AIX-MARSEILLE II

Why search for a signature predicting metastatic relapse in breast cancer (BC) ?

Breast cancer is the most common and deadly cancer type in women,

- Incidence rate (2005) : 101.5 per 100,000 women
- All cancer types (2005) : 251.9 per 100,000 women
- Mortality rate (2005) : 17.7 per 100,000 women
- All cancer types (2005) : 80 per 100,000 women

Most patients diagnosed with node-negative early breast cancer undergo adjuvant chemotherapy
which could be avoided in 70-80% of cases (Bertucci & Birnbaum, 2008)

Biomarkers found using genomic approaches may refine therapeutic decision

Talk outline

Goal : signature predicting metastatic BC relapse at five years

Integration of interaction and gene expression data (ITI algorithm)

Breast Cancer Compendium assembly

Integration of interaction datasets

Algorithm to discover discriminative subnetworks

Generation of Subnetwork database

SVM-based classification : separation in good and bad prognosis groups

Conclusion and perspectives

State of the art

(2002) Van't veer *et al*. : 117 patients → **70 gene** signature
Currently undergoing clinical trial (MINDact)

(2005) Wang *et al*. : 286 patients → **76 gene** signature
only 3 genes in common

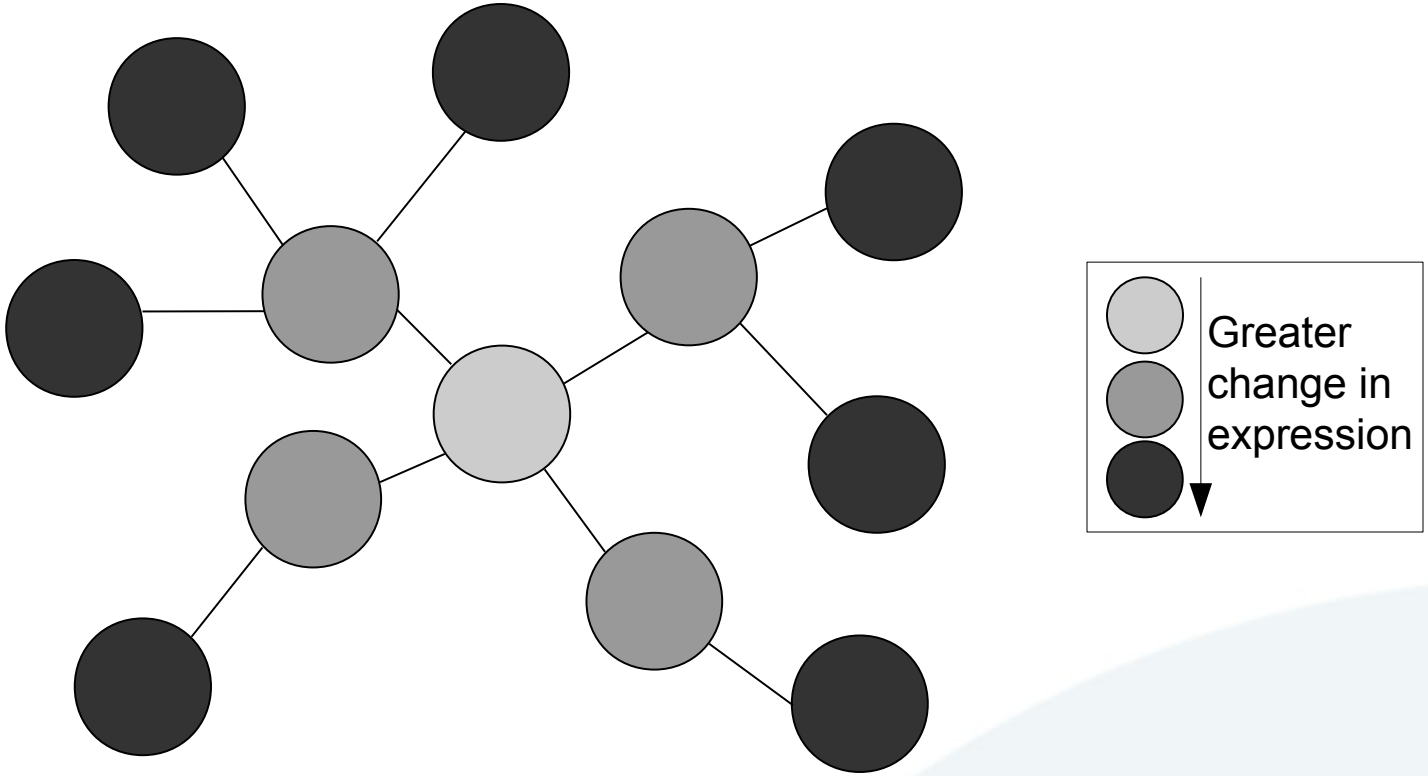
(2005) Michiels *et al*. : Signatures were unstable and not generalizable, however, cross-validation wasn't used (Fan *et al*., 2010)

(2006) Ein-Dor *et al*. : **more than one** 70 genes signature exist

(2007) Ein-Dor *et al*. : Theoretically **thousands of samples** are needed to generate a robust list.

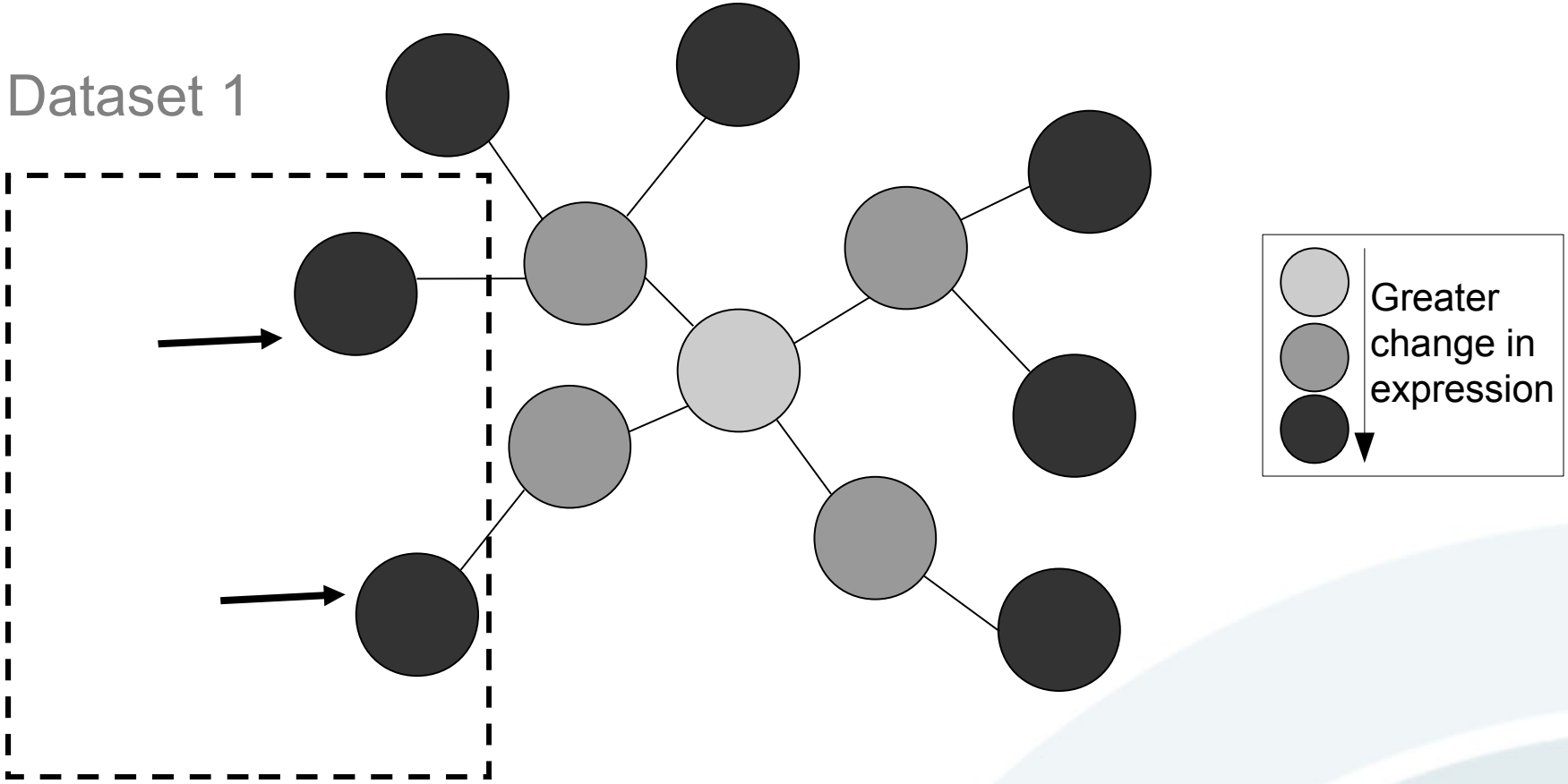
(2008) Chuang *et al*. : Integration of interactome-transcriptome results in more robust biomarkers (but on only one dataset)

Integration of interaction data and gene expression data



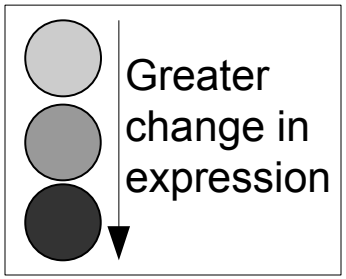
Example of an interaction network

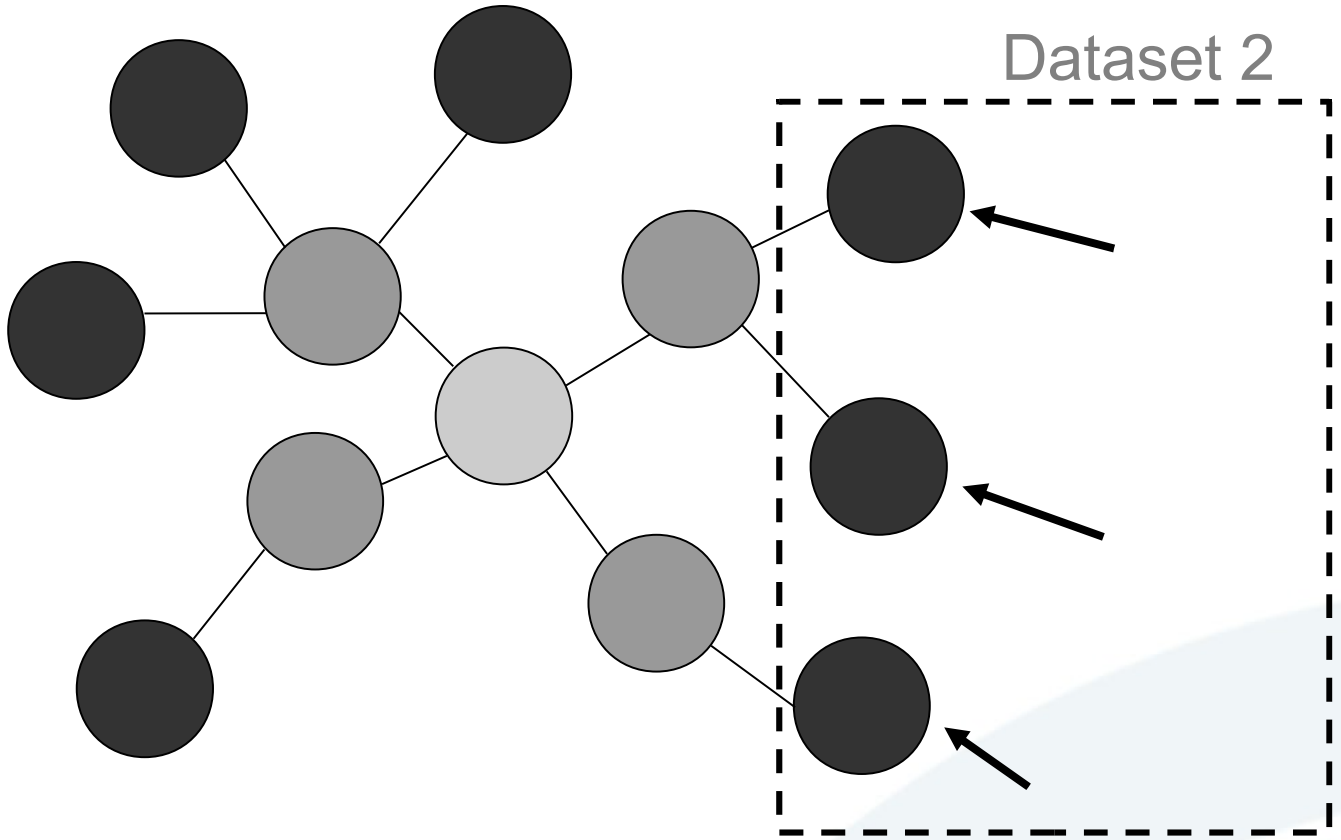
Integration of interaction data and gene expression data



One study may find these effectors with the greatest changes in expression

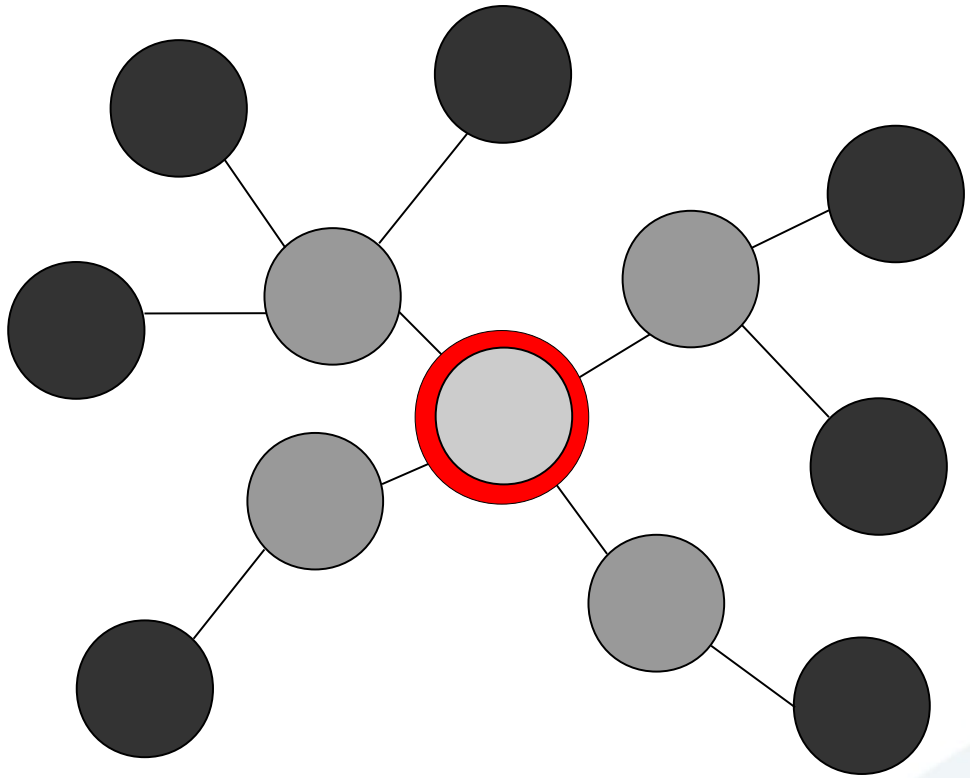
Integration of interaction data and gene expression data





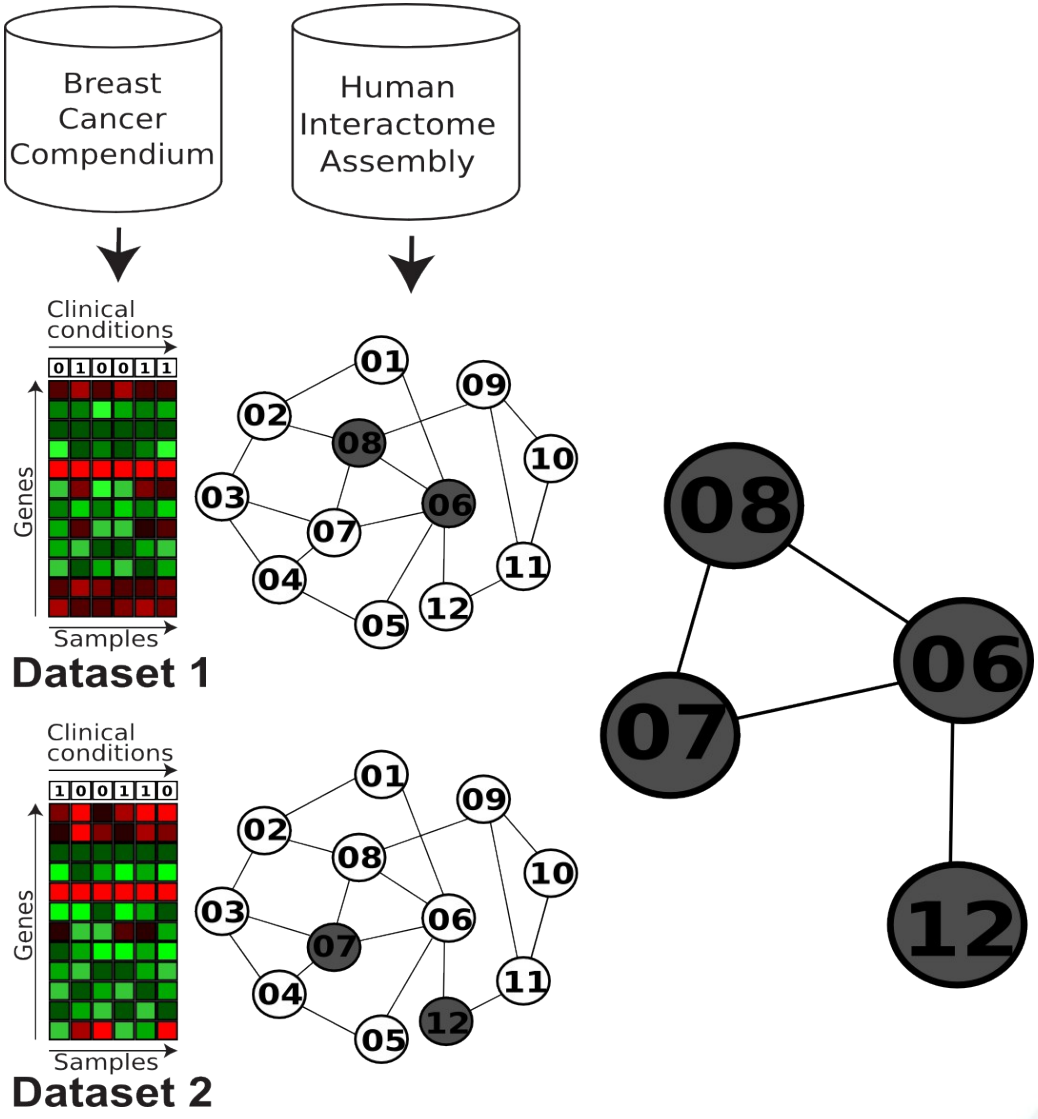
Another study may reveal other effectors

Integration of interaction data and gene expression data



The upstream gene at the origin of metastasis has only subtle expression changes

Interactome-Transcriptome Integration (ITI)



2 types of data

+ clinical annotation

(DMFS status)

Measure of correlation

of GEP*

with patient DMFS status

Detection of

discriminating

subnetworks

* Gene expression profile

Breast Cancer compendium assembly

Dataset	Samples	Samples Kept
Desmedt	198	190
IPC	266	31
Ivshina	249	
Jezequel	252	
Kreike	59	
Loi	414	101
Miller	251	
Parker	225	
Pawitan	159	
Perou	84	
Schmidt	200	182
Sorlie	85	
Sotiriou	189	
van de Vijver	295	
Wang	286	276
Zhang	136	
Zhou	54	

Datasets Kept

Datasets Removed

- 5 datasets selected on status availability
- 780 samples selected on 5 years following time

Human Protein-Protein Interaction (PPI) data assembly

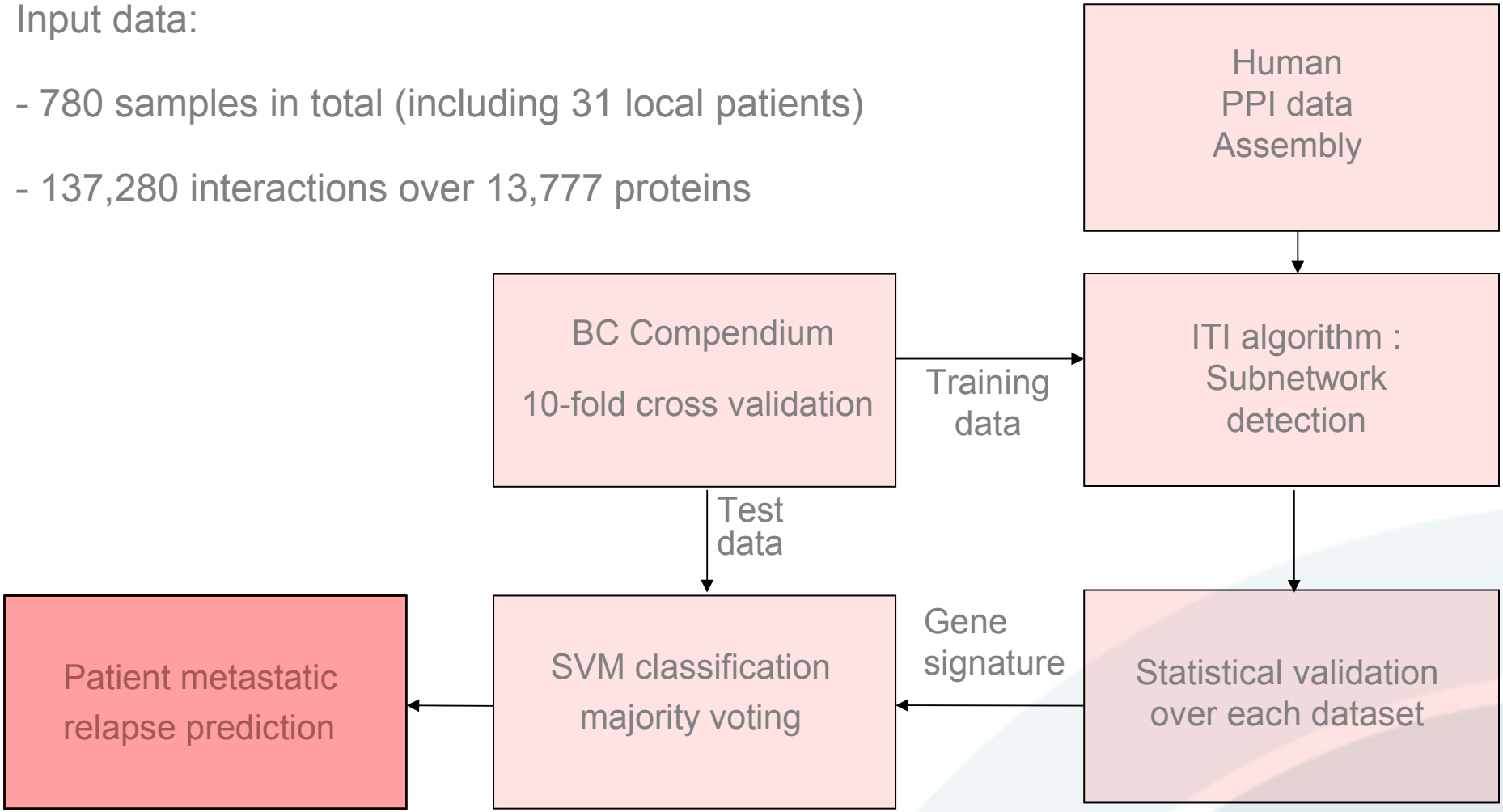
	Proteins	Interactions
Reactome	3.531	69.800
DIP	918	810
MINT	5.559	12.143
IntAct	7.471	25.616
Marcotte	7.568	31.353
HPRD	9.386	36.577
Assembly	13.777	137.280

Data workflow

Input data:

- 780 samples in total (including 31 local patients)

- 137,280 interactions over 13,777 proteins



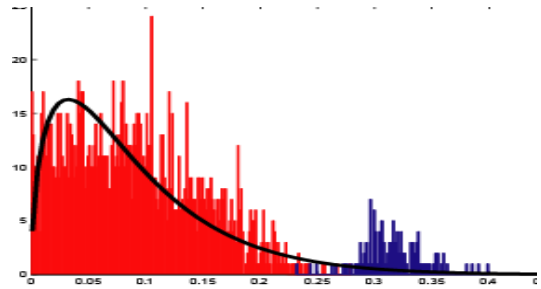
Statistical validation yields 107 subnetworks

2,331 subnetworks found with ITI algorithm

- A random subnetworks 527 subnetworks (1.10^{-2} on 1 dataset)
- B shuffled expression data 122 subnetworks (1.10^{-3} on 1 dataset)
- C random interactome 230 subnetworks (1.10^{-1} on 1 dataset)

Intersection => 122 subnetworks left

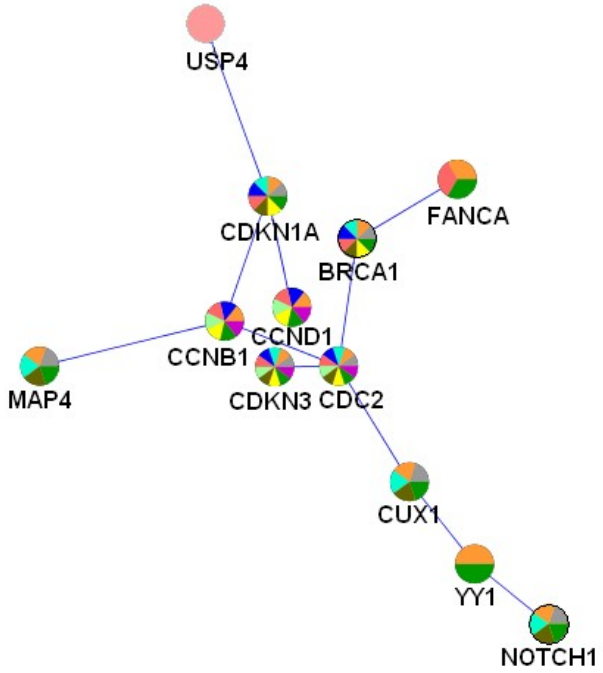
Score distribution on Loi et al



80% overlap filtering

107 subnetworks kept

Subnetworks database generation : ITI bioinformatics ressource



GOID: 51726		Regulation of cell cycle
GOID: 7049		Cell cycle
GOID: 51325		Interphase
GOID: 74		Regulation of progression through cell cycle
GOID: 51329		Interphase of mitotic cell cycle
GOID: 43118		Negative regulation of physiological process
GOID: 50791		Regulation of physiological process
GOID: 48523		Negative regulation of cellular process
GOID: 51243		Negative regulation of cellular physiological process
GOID: 51244		Regulation of cellular physiological process

Made with Cytoscape and GOlorize

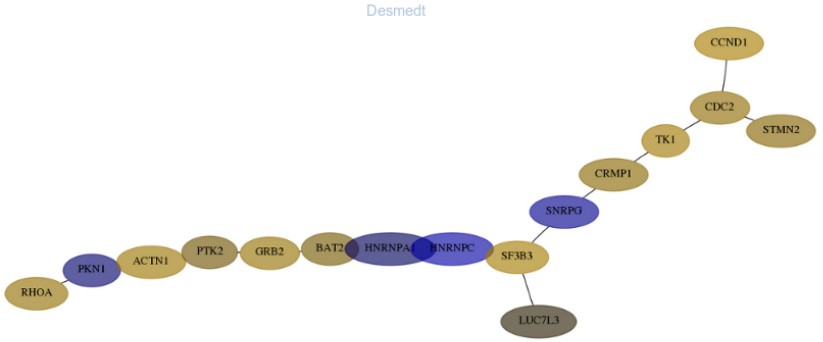
Result: 107 subnetworks containing 419 genes

Bioinformatics resource linking interactome and phenotype

(http://bioinformatique.marseille.inserm.fr/iti)

Subnetwork structure for each dataset

- Desmedt IPC-NIBC-129 Ivshina_GPL96-GPL97 Loi_GPL570



Score for each gene in subnetwork 387-4-real in each dataset

Gene Symbol	Links	Frequency	Frequency Rank	Global rank	Desmedt	IPC-NIBC-129	Ivshina_GPL96-GPL97	Loi_GPL570
crmp1		2	104	18	0.065	0.155	-0.056	0.060
stmn2		29	8	3	0.057	0.106	0.005	0.052
rhoa		1	164	31	0.080	0.054	0.083	-0.107
bat2		1	164	31	0.039	0.182	-0.001	-0.013
luc7l3		7	38	5	0.006	0.111	-0.046	-0.016
ccnd1		102	2	2	0.127	-0.183	0.062	0.190
pkn1		1	164	31	-0.031	0.056	0.096	0.119
hnrnpa1		4	60	9	-0.021	0.209	0.088	-0.020
sf3b3		7	38	5	0.119	0.241	0.143	-0.019
tk1		6	43	7	0.109	0.149	0.190	-0.017
cdc2		103	1	1	0.073	0.247	0.206	0.172
snrpg		1	164	31	-0.068	0.217	0.120	0.108
ptk2		1	164	31	0.035	0.139	-0.059	0.033
actn1		4	60	9	0.100	0.223	0.031	-0.052
hnrnpc		1	164	31	-0.110	0.101	0.116	0.114
grb2		9	27	4	0.084	-0.013	0.086	-0.031

Discovery of a set of subnetworks biologically relevant to metastatic BC

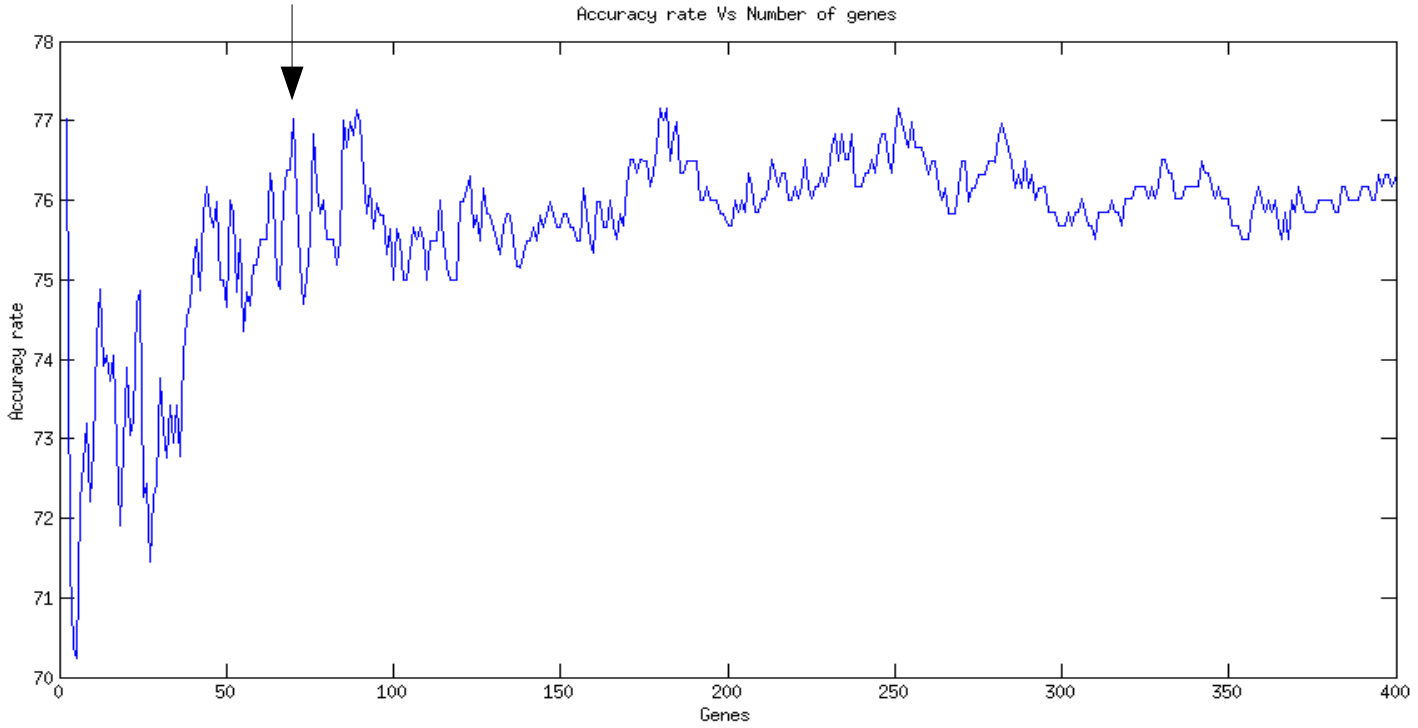
Apoptosis regulation	[2 Subnetworks]
Cellular adhesion	[1 Subnetwork]
Cell cycle regulation	[4 Subnetworks]
Immune response	[3 Subnetworks]
Development, polarity	[7 Subnetworks]
Metabolism	[7 Subnetworks]

Genes

CDC2
CCND1
STMN2
GRB2
LUC7L3
SF3B3
TK1
TSC1
HNRNPA1
ACTN1
HSPB1
MAPKAPK2
AGTPBP1
CYCS
BAX
PPFIA1
SFN
CRMP1
PRKCI
YWHAZ

Successful classification based on Support Vector Machine

77% accuracy



Majority vote is weighted by the number of samples
10 fold cross-validation

Conclusion and perspectives

Conclusion

- Construction of Breast Cancer Compendium and Human PPI assembly
- Subnetwork discovery using the ITI algorithm
- Validation of Subnetworks using 3 statistical methods
- Linking ITI interactome results and phenotype – web database available
- SVM-classification with 77% successful majority vote

Perspectives

- Currently : Specific biology of ER+ and ER- patients
- Currently : Application to inflammatory breast cancer
- Short term : Application to other diseases
- Long term : Add other data types (CGH-array...)

Publication

•Garcia *et al.* (*in press*) *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications*

Acknowledgments

ITI project collaborators

G. Bidaut,PhD (CRCM)
D. Birnbaum, MD, PhD (IPC, CRCM)
F. Bertucci, MD, PhD (IPC)
S. Carpentier, MSc (Ipsogen)
M. Chaffanet, PhD (IPC)
P. Finetti, MSc (IPC)

Integrative Bioinformatics team members

G. Bidaut, PhD (Team leader)
O. Stahl, MSc (Bioinformatics engineer)
P. Rouillier, PhD (Post-doctoral researcher)
A. Guille (graduate student)
F. Blondin (graduate student)

http://bioinformatique.marseille.inserm.fr/