

Interactome-regulome integrative approach for genome-wide screening data analysis in a breast cancer stem cells study

C Rioualen^{1,2,3,4}, Q Da Costa^{1,2,3,4}, G Pinna⁵, A Harel-Bellan⁵, E Charafe-Jauffret^{1,2,3,4}, C Ginestier^{1,2,3,4} and G Bidaut^{1,2,3,4}
contact: claire.rioualen@inserm.fr

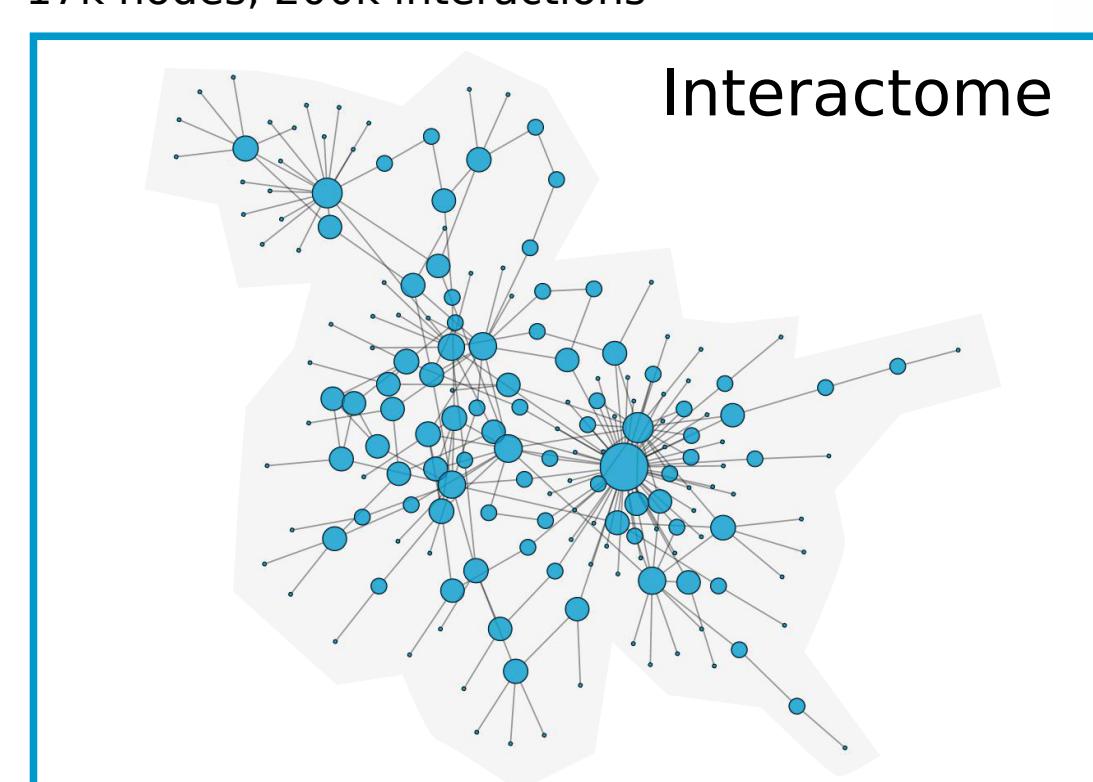
1. Centre de Recherche en Cancérologie de Marseille, Inserm U1068, Marseille, 13009, France.
2. CNRS UMR7258, Marseille, 13009, France.
3. Aix-Marseille Université, Marseille, 13009, France.
4. Institut Paoli-Calmettes, Marseille, 13009, France.
5. Plateforme ARN interference ParL, CEA Saclay, Gif-sur-Yvette, France.

Breast cancer is the deadliest cancer in women worldwide¹. It is believed that **cancer stem cells** (CSC) could explain its recurrence, due to their **resistance to conventional treatments** and their ability to generate metastases².

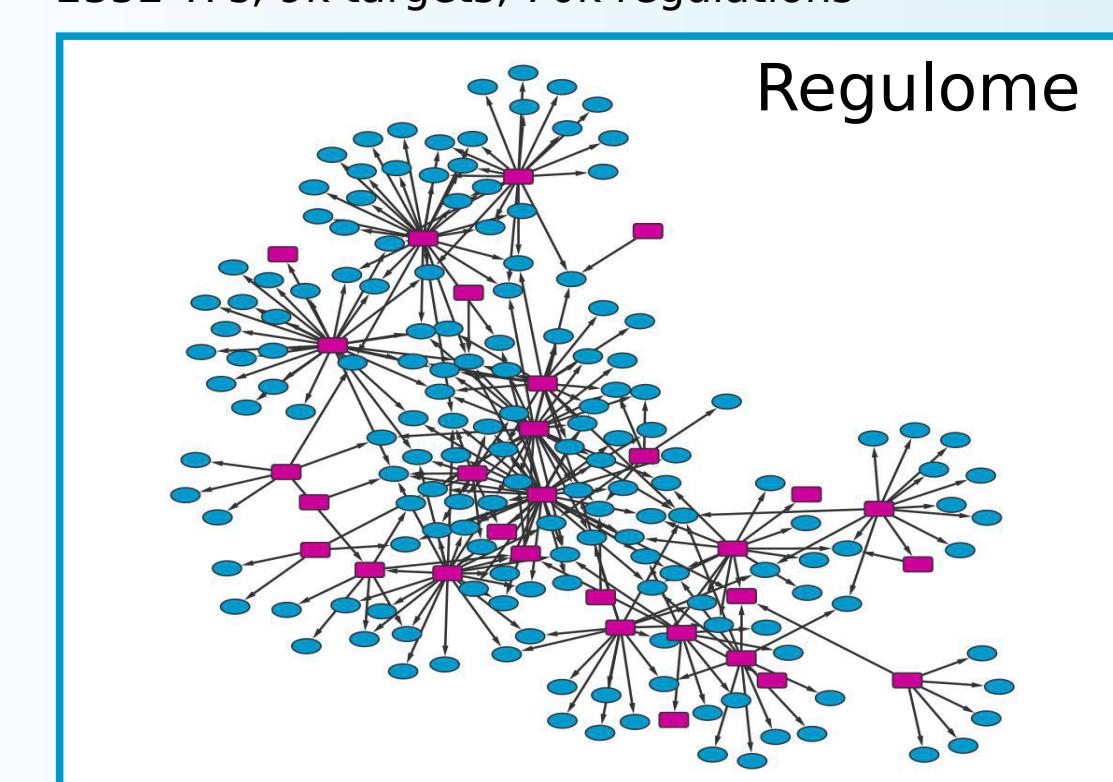
Input data

Networks building

HPRD, MINT, INTACT, DIP, Proteinpedia, I2D
17k nodes, 200k interactions



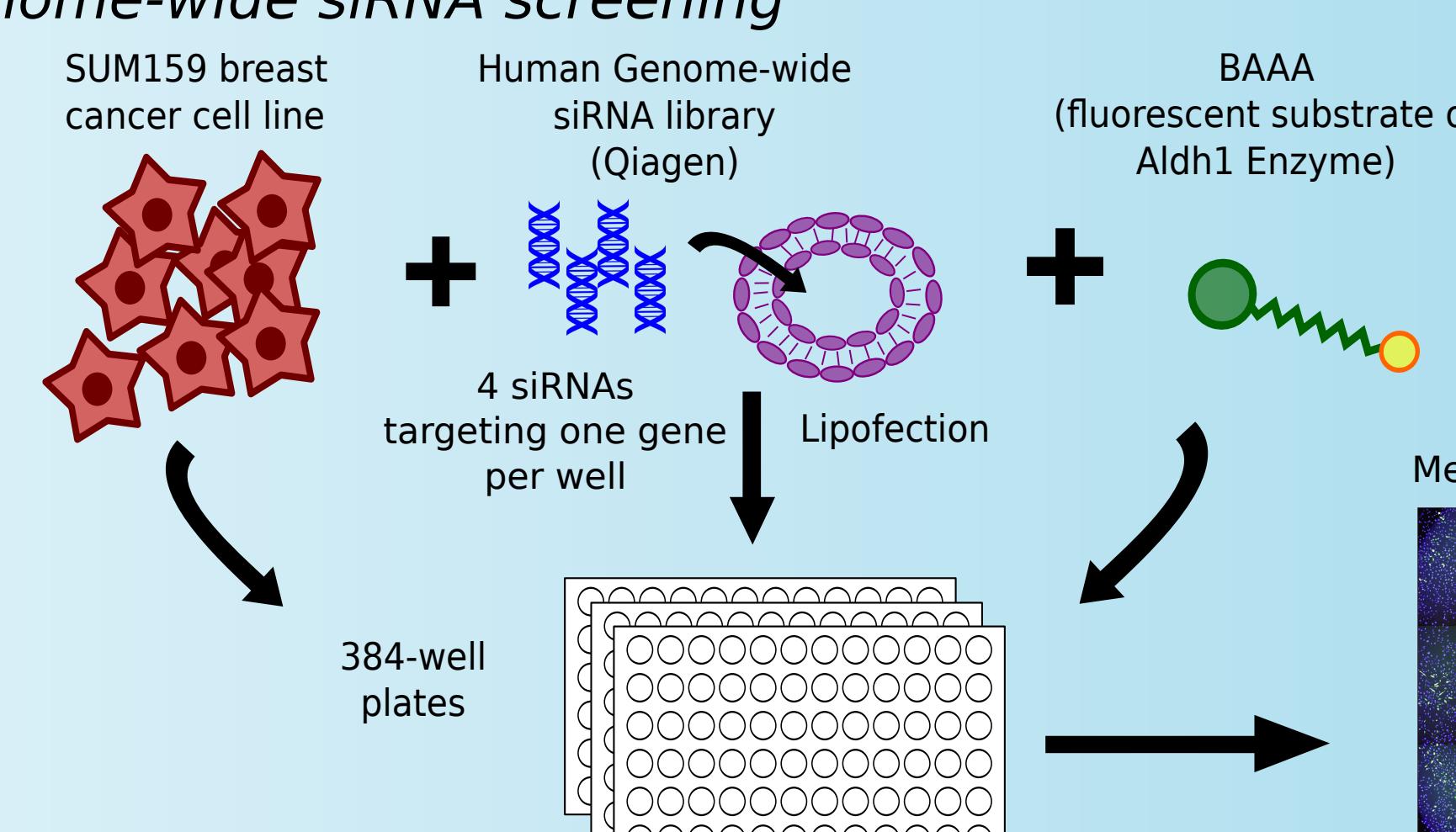
TRANSFAC, TRED, ITFP, PAZAR, ORegAnno
2352 TFs, 9k targets, 70k regulations



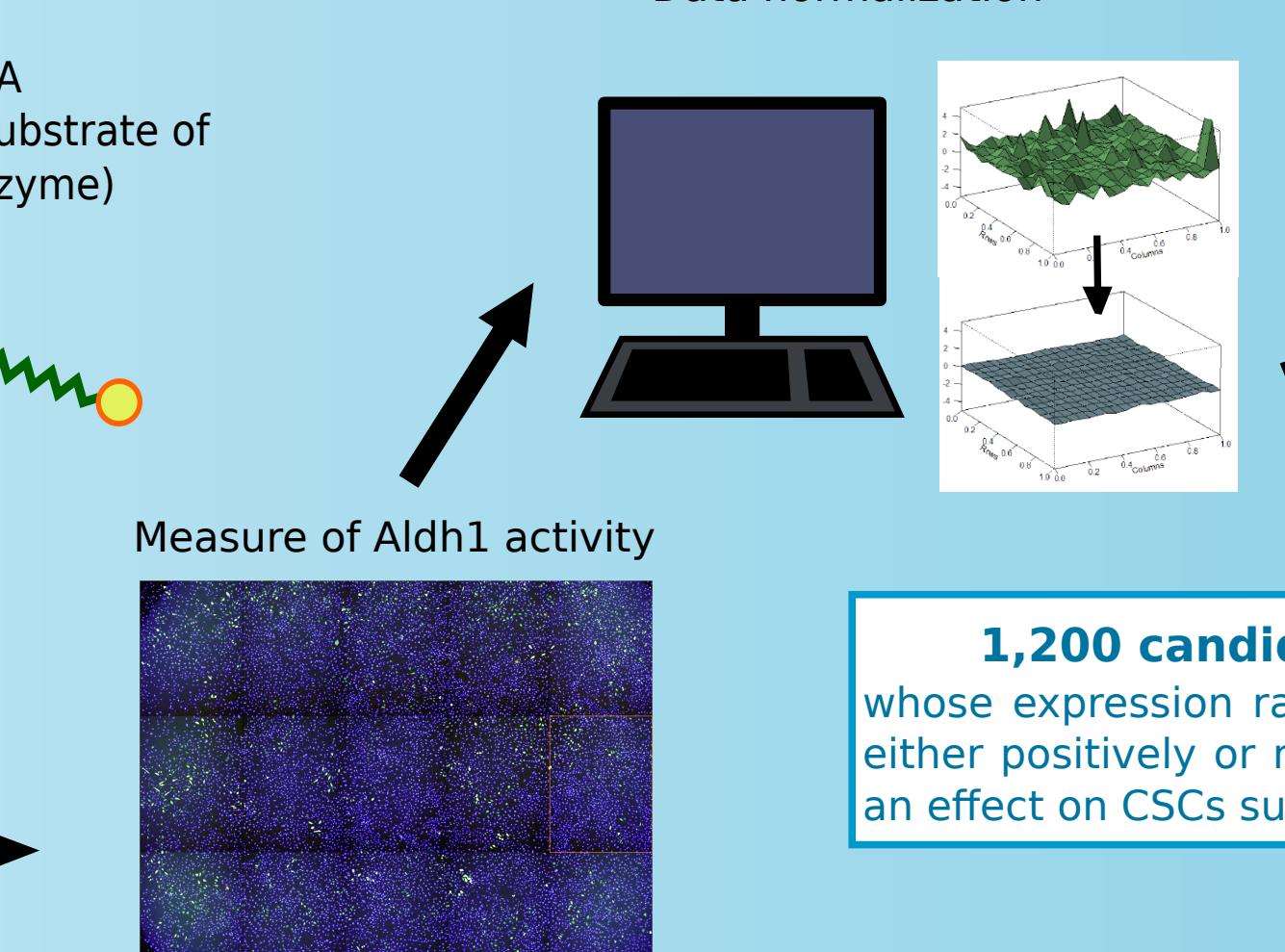
Public databases were parsed, and two big networks were built: the interactome, made of protein-protein interactions, and the regulome, made of TF-target relationships. Above are subsets of these networks.

In order to unravel their mechanisms, a genome-wide siRNA screening was performed. We decided to incorporate a **network-based approach** in order to validate *in vitro* experiments, identify key signaling pathways, and better exploit the secondary screening for future drug-testing.

Genome-wide siRNA screening



Data normalization



1,200 candidate genes
whose expression rates were affected, either positively or negatively, causing an effect on CSCs survival or death

CSCs are associated with a high level of Aldh1 enzymatic activity³. The increase or decrease of this population was measured by lipofecting 4-siRNA pools targeted at one gene per well, then adding a substrate associated with a fluorescent dye. 18,500 genes were scored according to the ratio Aldh⁺/Aldh⁻ per well, showing the effect on CSCs survival or death.

Data integration & Networks analysis

Subnetworks detection algorithm

Subnetworks are recursively aggregated from a seed, by using the ITI algorithm^{4,5} on the interactome and the regulome separately.

1. Subnetwork Seed Detection

Node * is the seed. If this gene showed a good score in the primary screening, its neighbours are investigated too.

3. Subnetwork Completion

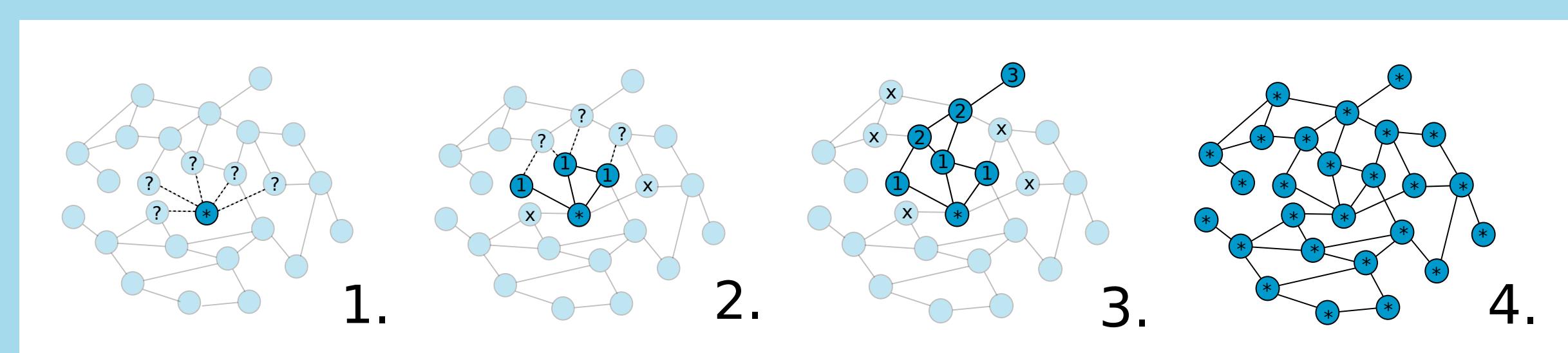
A node that is dismissed is never considered again. Once the score cannot be improved, the subnetwork is complete.

4. Subnetwork Statistical Validation

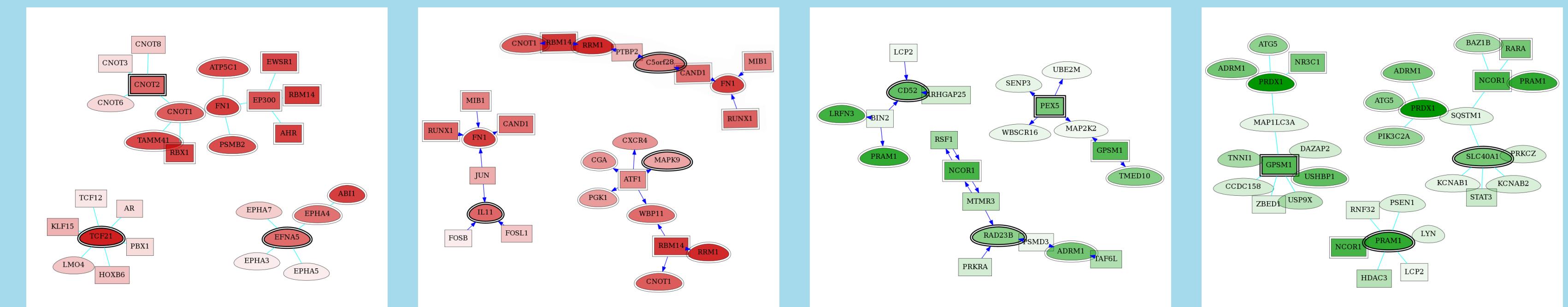
Every node is investigated as a seed. All kept subnetworks are statistically validated by drawing 3 null distributions generated by shuffling interaction data, shuffling expression data and generating random subnetworks.

2. Neighbors Exploration

Nodes that increase the average score of the subnetwork are kept. Their neighbours are investigated recursively too.



Subnetworks involved in CSC population growth or death



62 subnetworks were found in total. The 1,200 genes kept after the primary screening are highlighted by double circles in the figures. Green nodes are proteins coded by genes associated with a growth of the CSC population, while red ones are associated with CSC death. Squares are transcription factors.

The first step of the integration is consistent with the primary screening, most genes being common between the two methods.

Genes that are not impacted much by the knock-down could be involved with subsets of the regulome & the interactome that are more deregulated.

Subnetworks integration

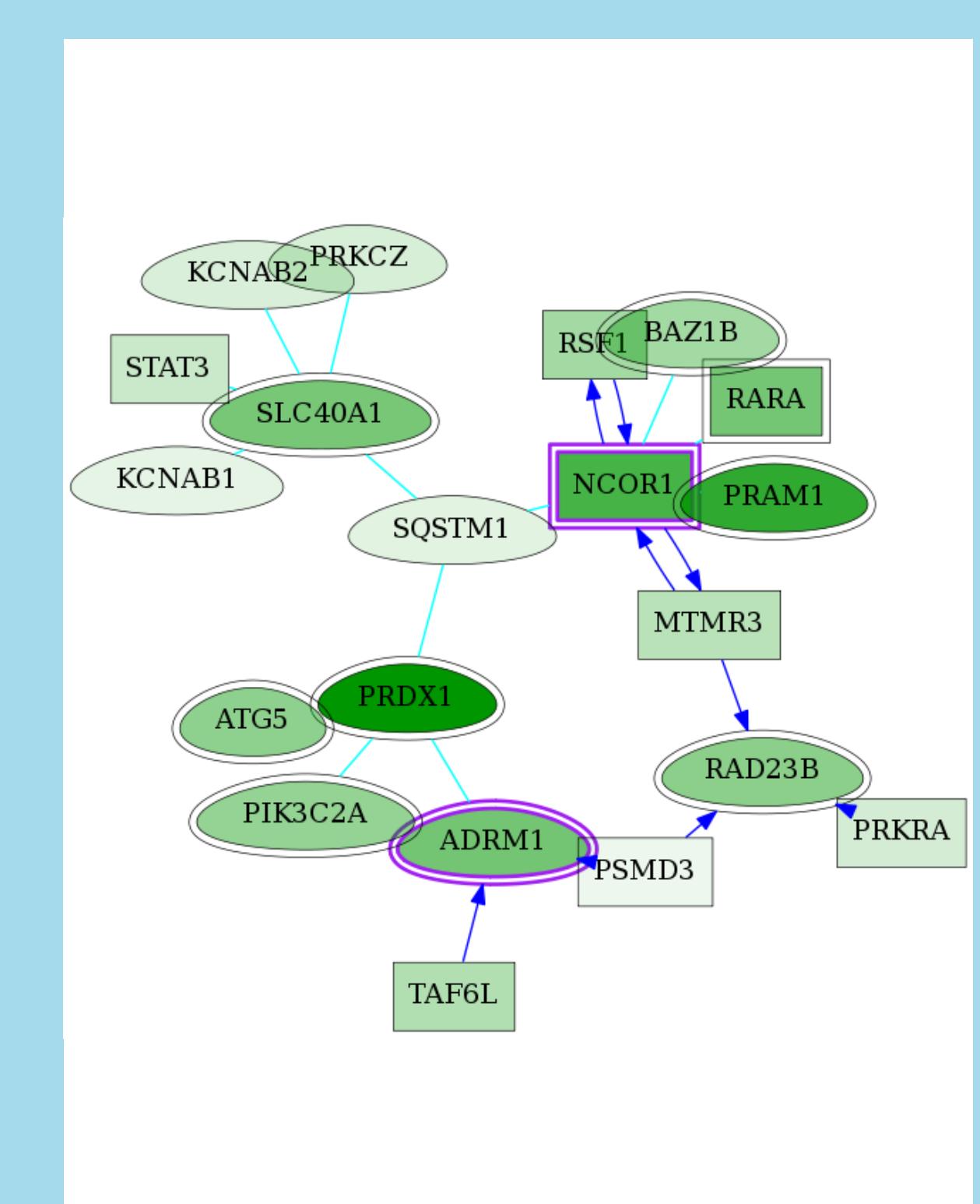
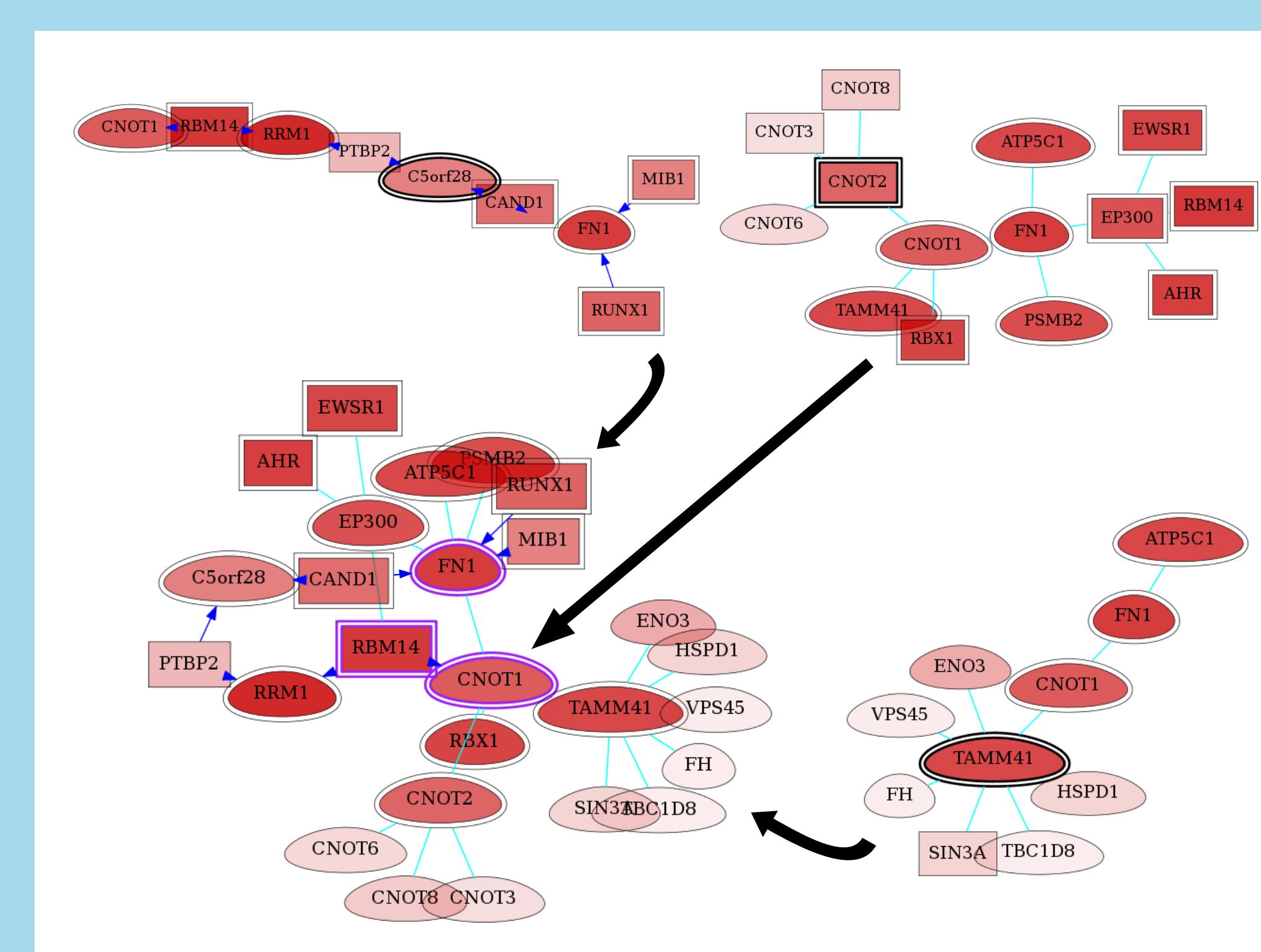
In order to visualize pathways of regulation that involve several levels of regulation, we integrated the interactome subnetworks (groups of interacting proteins) with regulome subnetworks (transcription factors and targets).

RUNX1, which was already described as a master regulator of leukemic stem cells, seems to be a regulator of breast CSCs. The integration confirms the fact that this transcription works with **EP300**. Our analysis identifies new potential breast CSCs regulators, such as **TAMM41**, which might be interesting to further investigate.

NCOR1/RARA were used as a positive control since they are known for being Aldh1 transcription factors.

Secondary screening

A secondary screening was performed on the 1,200 previously selected genes, and 300 genes were isolated. About 20% of them were already found by the *in silico* approach. The other genes detected by integration constitute interesting targets for the upcoming *in vitro* experiments.



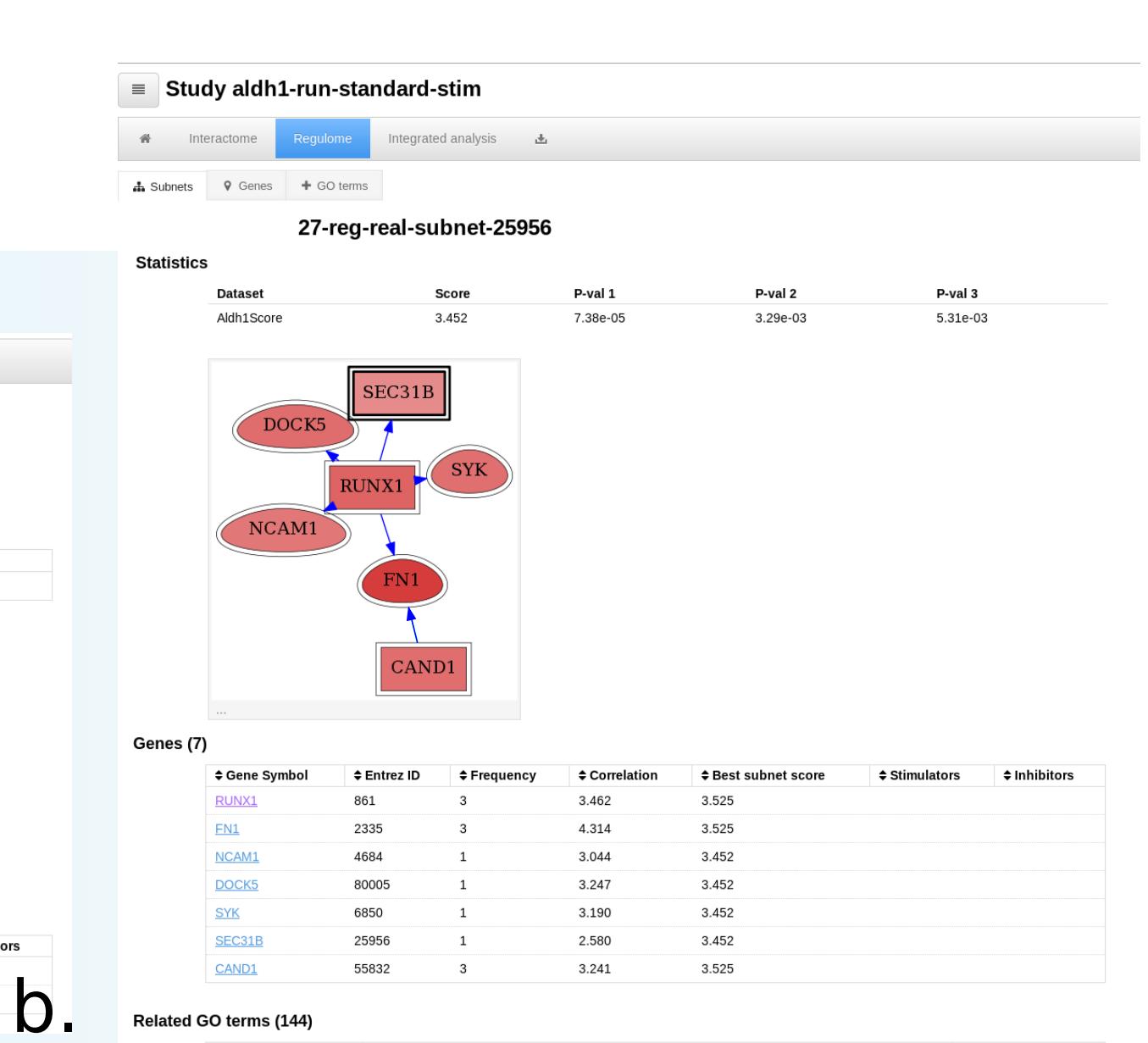
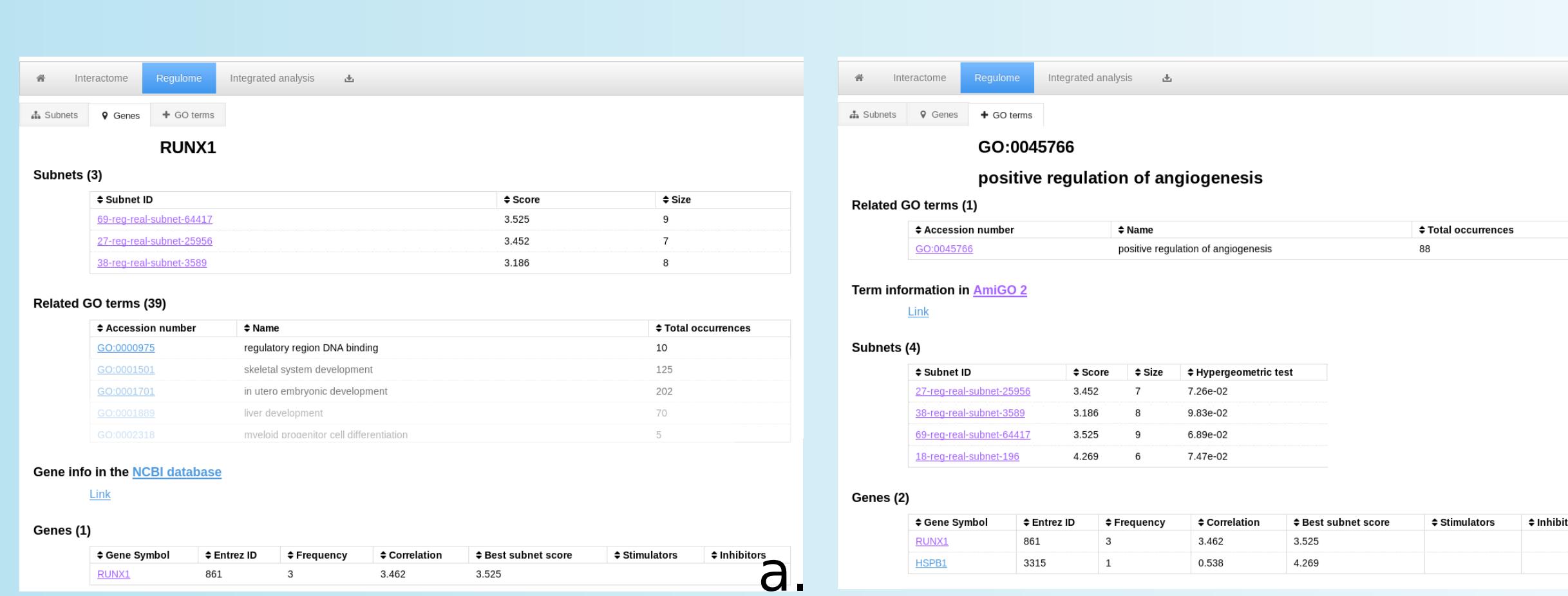
Online database of subnetworks

Our pipeline includes the possibility of generating full **HTML reports**.

Detected subnetworks are stored and can be investigated gene by gene (a). This visualization also includes all the information on Gene Ontology terms enriched in the subnetworks (b). Each subnetwork has its own page, compiling all the related information (c).

All the data can be downloaded from the site in flat format, and subnetwork files are formatted in *.nnf format, thus enabling their analysis in Cytoscape.

This new integrative approach yielded very promising results by both **confirming *in vitro* experiments, and unravelling new genes of interest**. As opposed to a single-gene approach allowed by the screening, the integrated method constitutes robust regulation chains that can be further investigated *in vitro*. It could ultimately allow us to identify **new targets in cancer treatment**. Elaborating a new drug targeting CSCs could greatly improve breast cancer clinical outcome.



- World Cancer Report. International Agency for Research on Cancer. 2008.
- Targeting breast cancer stem cells. Liu S & Wicha MS. J Clin Oncol. 2010 Sep 1;28(25):4006-12. doi: 10.1200/JCO.2009.27.5388.
- ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Ginestier C et al., Cell Stem Cell. 2007 Nov;1(5):555-67. doi: 10.1016/j.stem.2007.08.014.
- Interactome-transcriptome integration for predicting distant metastasis in breast cancer. Garcia et al., Bioinformatics. 2012 Mar 1;28(5):672-8.
- Detection of driver protein complexes in breast cancer metastasis by large-scale transcriptome-interactome integration. Garcia et al., Methods Mol Biol. 2014;1101:67-85.
- TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Matys et al., Nucleic Acids Res. 2006 Jan 1;34(Database issue):D108-10.

Acknowledgments

Support to this project is coming from the Institut National du Cancer grant Number INCA_5911 to CG. Additional support for QDC obtained from Plan Cancer 2013-2016 Action 3.6 to GB.