ProLific

ODYSSEA
ensemble contre le cancer du sein

Inserm

Amad DIOUF, Gustave Roussy Institute - CBIO

Project :

ATIP3 protein-related biomarker discovery in breast cancer.

Gene lists using SIGs on 7 Feature selections of LASSO regression

GUSTAVE ROUSSY
CANCER CAMPUS
GRAND PARIS

MINES ParisTech

CBIO

Equipe Dr Clara NAHMIAS
Directrice de recherche CNRS

Dr Chloé-Agathe Azencott
Chargée de recherche CBIO

Feature selection analysis

New definition of a SIG :

- **(high_stringency condition) :** score s (ratio #_majority_sign / # all_signs) = 1 on each dataset where the gene has been selected
- **(100% conservation of sign condition) :** score s_all (score s over all the coefficients produced for the gene) = 1

We lack at this point a way to further restrict the gene lists…

Solution : Divide the SIGs found in 4 Class (from Class 4 to Class 1, a gene is present in more feature selection ie has been "tested more times"

- **Class 1 :** the gene is always selected in all the FSs, in a stable cond., and even selected in the MT
- **Class 2 :** the gene is always selected in all the FSs, in stable cond., but just not selected in the MT
- **Class 3 :** As in Class 2, the gene is not selected in the MT, but for the rest of the datasets, if we allow 1 absence from a selection, the rest is selected in stable cond.
- **Class 4 :** a gene stable in every feature selection where it is selected, but it is absence of at least 2 feature selections

PrOlific

ODYSSEA
ensemble contre le cancer du sein

Inserm

Thank you

GUSTAVE
ROUSSY
CANCER CAMPUS
GRAND PARIS

MINES
ParisTech

CBIO

Feature selection analysis: Part 1 - Restriction of Feature Selections