

Improved gene co-expression network quality through expression dataset down-sampling and network aggregation

Supplementary Information

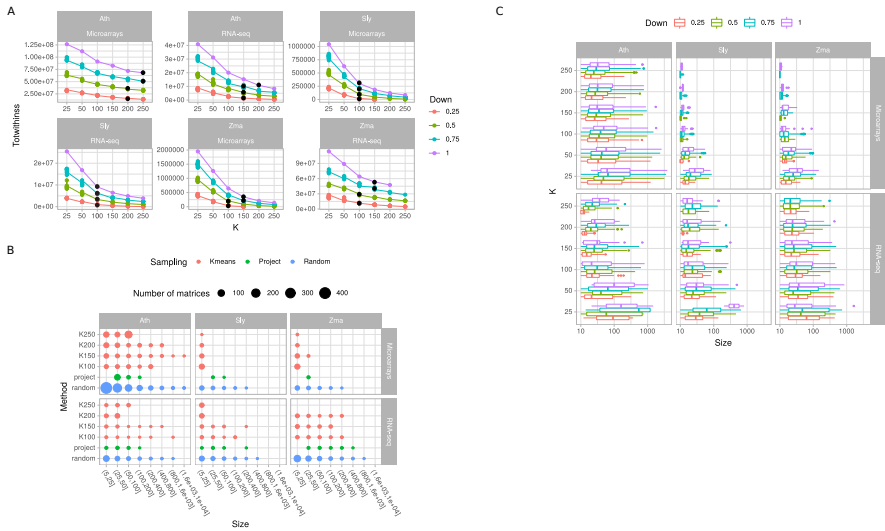
Franziska Liesecke^{1,*}, Johan-Owen de Craene¹, Sébastien Besseau¹, Vincent Courdavault¹, Marc Clastre¹, Valentin Vergès¹, Nicolas Papon², Nathalie Giglioli-Guivarc'h¹, Gaëlle Glévarec¹, Olivier Pichon¹, and Thomas Dugé de Bernonville^{1,*}

¹ *Université de Tours, EA2106 Biomolécules et Biotechnologies végétales, Tours, 37200, France*

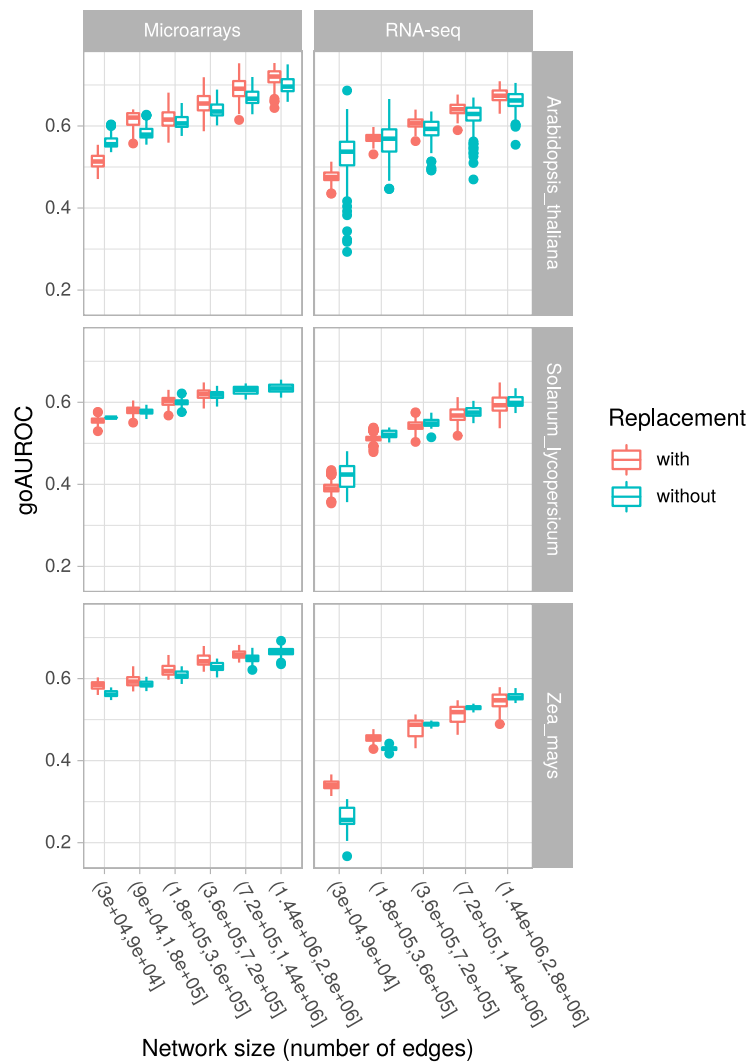
² *EA3142 GEIHP, Université d'Angers, Université Bretagne-Loire, Angers, 49100, France*

^{*} *Co-corresponding authors: franzi.liesecke@gmail.com, thomas.duge@univ-tours.fr*

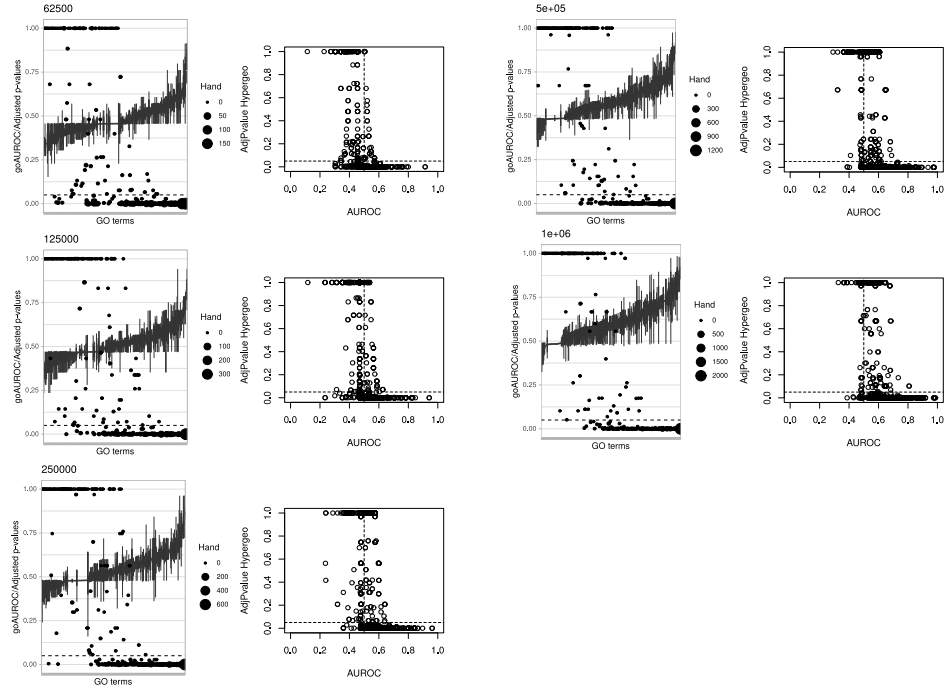
Supplemental Fig1: *k*-means down-sampling. Full or down-sampled (at 25%, 50% or 75%, multiple random sample combinations were tested, see Suppl Table 1) expression matrices were subjected to *k*-means partitioning for different *k* values. A, the total within cluster sum of squares is plotted against *k* values. Each dot represent one expression matrix (a total of 1,462 matrices). Black dots indicate *k* values that have been retained for further network construction and characterization. B, number and size of subsets in the full datasets partitioned with the *k*-means, project or random approaches. C, size of subsets obtained after applying *k*-means to randomly down-sampled full datasets.



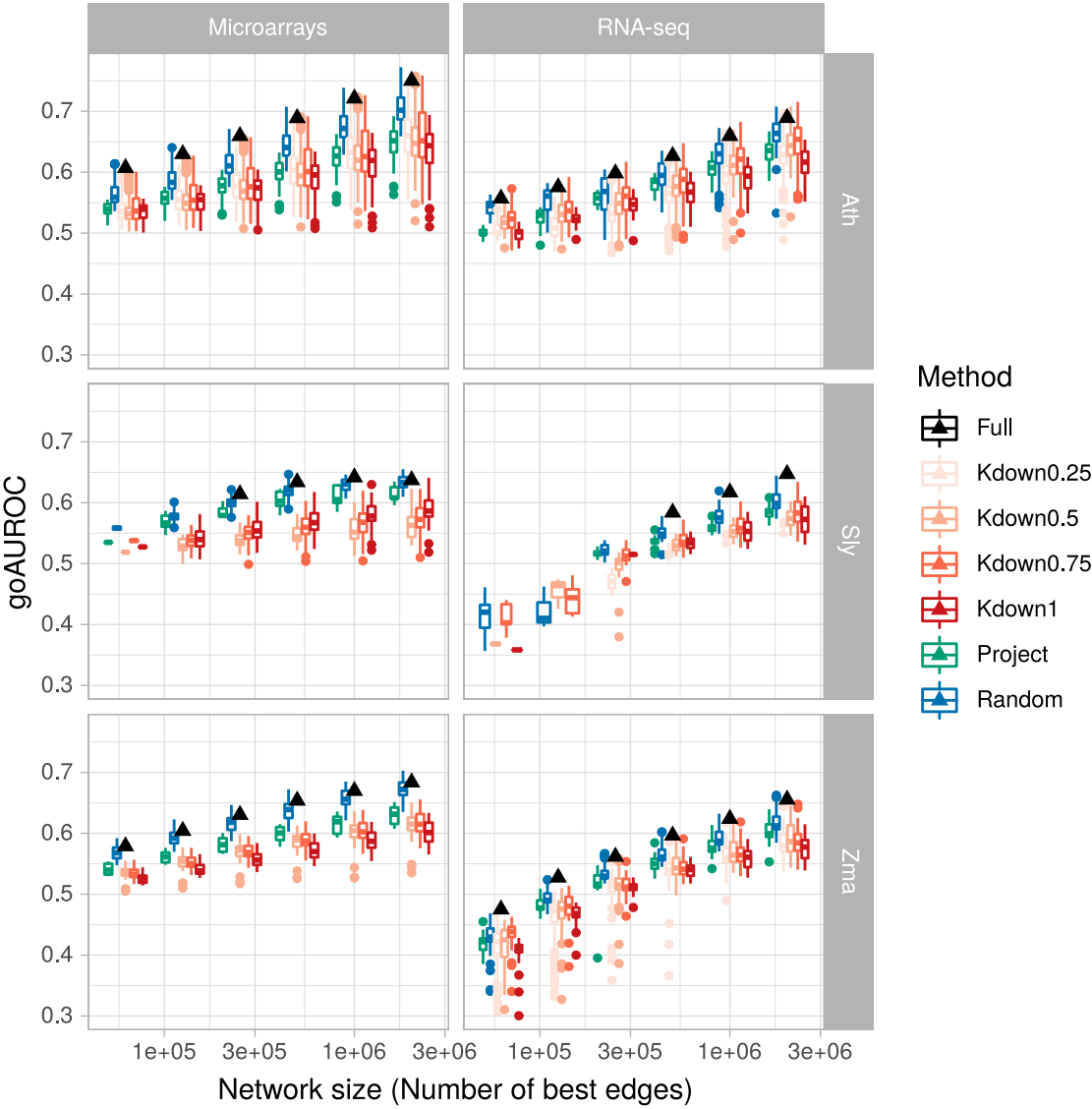
Supplemental Fig2: Effect of sample replacement on the performance of networks constructed from randomly sampled datasets.



Supplemental Fig3: GO AUROC values in cross-validation. Data presented here came from the *Solanum lycopersicum* RNA-seq expression matrix. Different network sizes were tested ranging from 62,500 to 1 million edges. Left part: for each GO term found in the network, the 3 AUROC calculated within the cross-validation step are represented as boxplots. Points correspond to the adjusted p-value calculated from an hypergeometric test and their sizes are scaled to the gene number in each GO term in the network. Right part, comparison between GO AUROC and p-value of the hypergeometric tests. Each point corresponds to one GO term.

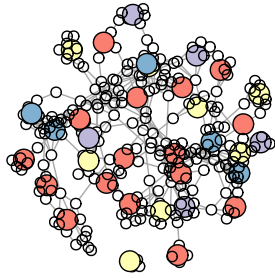


Supplemental Fig4: Effect of replacement in random sampling. Network performance is compared from subsets prepared by randomly selecting samples in the initial dataset, with or without replacement. Replacement implies than one sample may be used in different subsets.

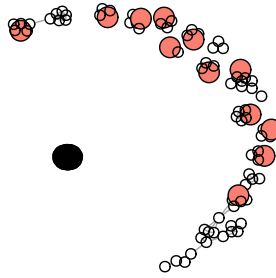


Supplemental Fig5: Estimation of PLC quality. The examples are drawn from different subsets of the Ath RNA-seq dataset. PLC quality may be evaluated by its transitivity (left), its ability to correctly associate query genes in communities (measured by a normalized Chi-squared test, center) or by its performance in capturing GO terms (right). Examples are given for low value (top) and high value (bottom) metrics.

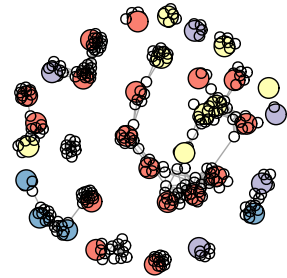
Tr=0.31;Chistat=0.10;Auroc=0.53



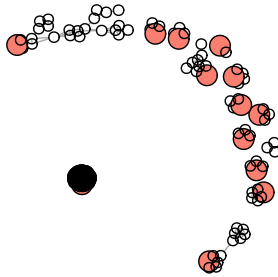
Tr=0.98;Chistat=0.058;Auroc=0.55



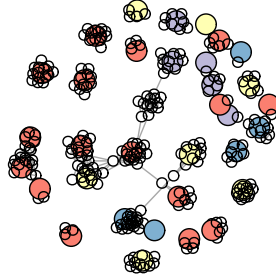
Tr=0.62;Chistat=0.20;Auroc=0.52



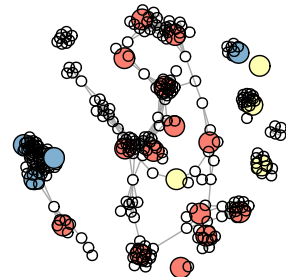
Tr=0.98;Chistat=0.05;Auroc=0.55



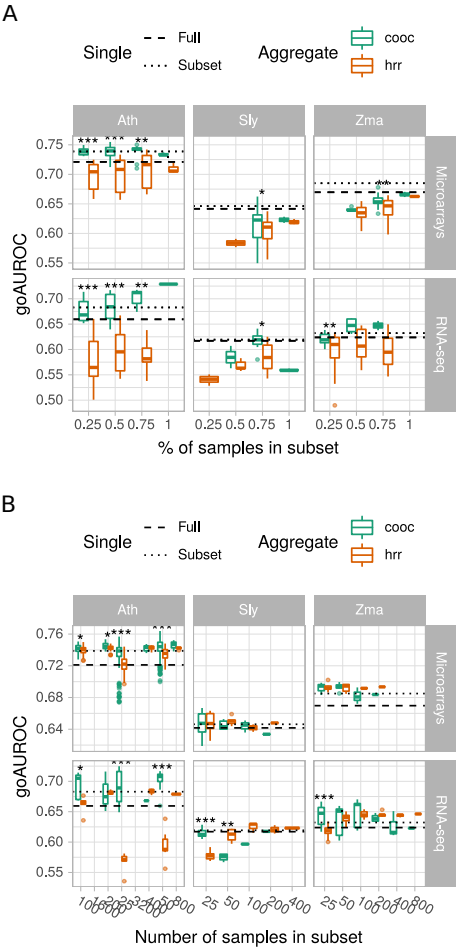
Tr=0.59;Chistat=0.49;Auroc=0.56



Tr=0.70;Chistat=0.20;Auroc=0.70



Supplemental Fig6: Effect of down-sampling methods on aggregate performance. Asteriks denote a statistical difference between co-occurrence and hrr aggregation methods (Wilcoxon rank sum test; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$). A, aggregates of networks obtained after applying k-means clustering full datasets or subsets containing 75, 50 or 25% of the initial samples. One aggregate contains all networks from one partition. B, aggregates of networks obtained after randomly selecting set of samples.



Supplemental Fig7: PLCs from aggregates. Performance of PLC was measured with normalized Chi-squared (A) and GO AUROC (B). Asteriks denote a statistical difference between co-occurrence and hrr aggregation methods (Wilcoxon rank sum test; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$).

