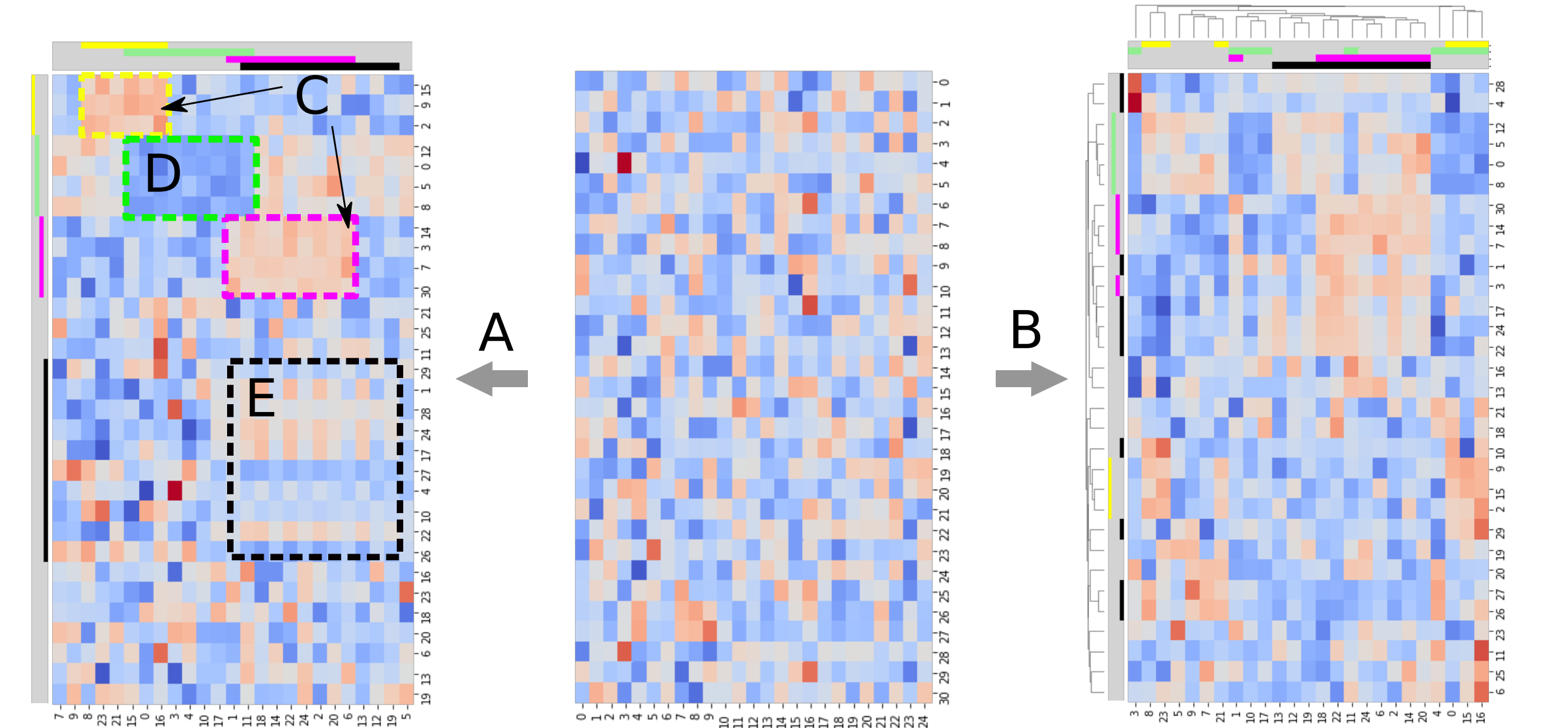


DESMOND 2.0: Identification of differentially expressed biclusters and investigation of their network properties

O. Zolotareva<sup>1,2</sup>, O. I. Isaeva<sup>3</sup>, Z. Chervontseva<sup>4</sup>, A. Savchik<sup>4</sup>, E. Zotova<sup>5</sup>, O. Lazareva<sup>1</sup>, D. B. Blumenthal<sup>1,6</sup>, N. K. Wenke<sup>2</sup>, M. Ester<sup>7,8</sup>, J. Baumbach<sup>2,9</sup>

1 - Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany; 2 - Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany; 3 - Division of Molecular Oncology & Immunology and Division of Tumor Biology & Immunology, The Netherlands Cancer Institute, Amsterdam, the Netherlands; 4 - A.A.Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences (RAS), Moscow, Russia; 5 - Independent scholar, Moscow, Russia; 6 - Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander University Erlangen-Nürnberg, Erlangen, Germany; 7 - School of Computing Science, Simon Fraser University, Canada; 8 - Vancouver Prostate Centre, Canada; 9 - Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark.

BICLUSTERING VS CLUSTERING



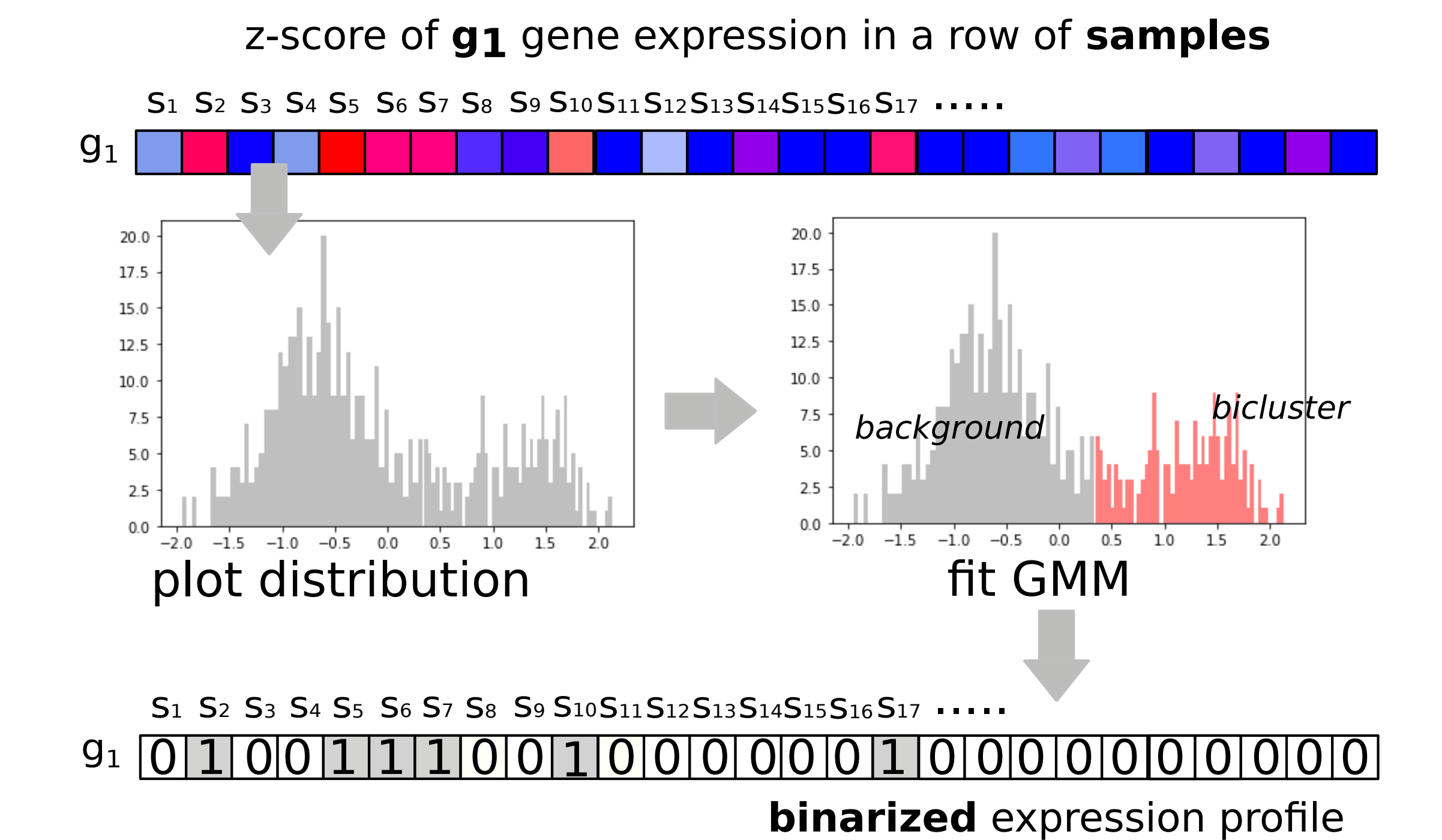
Biclustering (A) and conventional clustering (B). Differentially expressed up-regulated (C) and down-regulated (D) biclusters. Differentially co-expressed bicluster (E).

METHOD

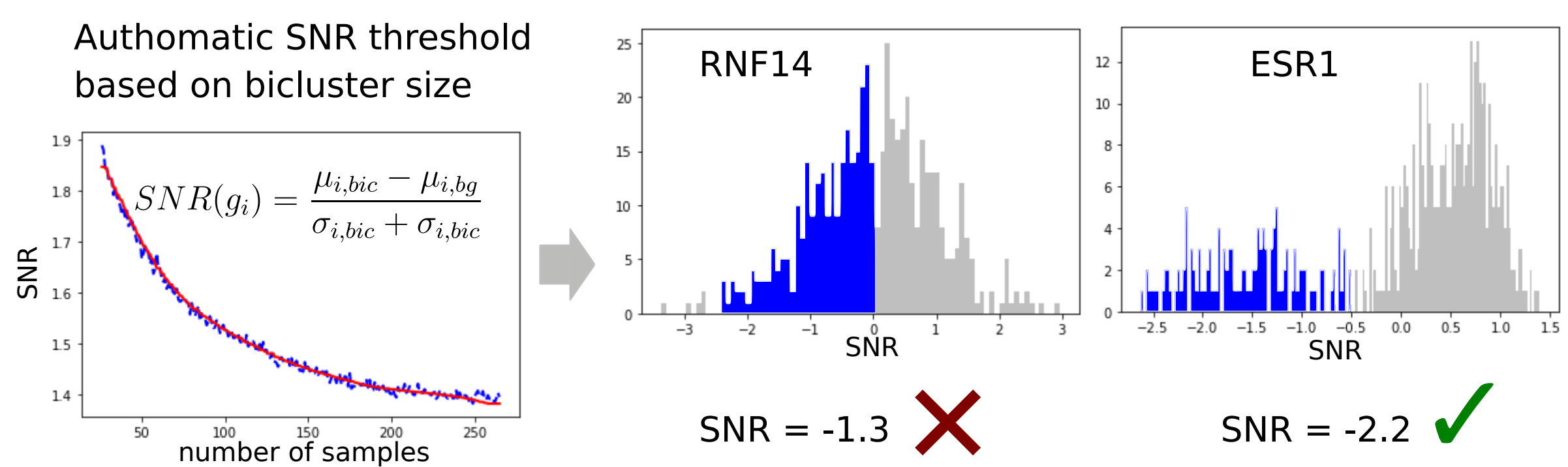
• **DESMOND** — a new method for identification of differentially expressed network-constrained biclusters (Zolotareva et al., 2020)

• **DESMOND2** is an **unconstrained** version

① Binarization of gene expressions



Only genes that split patients in two groups are kept. Genes are filtered by signal-to-noise ratio:



● **DESMOND2** identifies differentially expressed biclusters, biologically significant and reproducible in independent datasets

● Only small fractions of bicluster genes directly interact in homogeneous networks

● **Composite** networks may be a **better constraint** for biclustering and active sub-network detection methods than traditionally used **PPI networks**



ozolotareva/DESMOND2



zolotareva@wzw.tum.de

② Probabilistic gene clustering

**I. Gibbs sampling**

**- Initialization:** each gene is assigned to an individual bicluster

**- Iterations:** each gene joins a bicluster with a certain probability

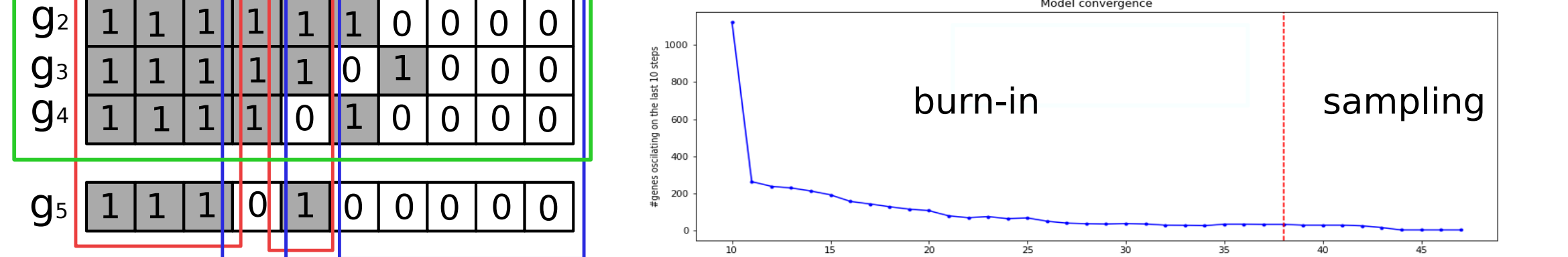
$$P(s_j = k | X, s_{-j}, \alpha, \beta) \propto \prod_{i: x_{ij}=1} \left[ \frac{\alpha/2 + \sum_{l: s_l=k, l \neq j} x_{li}}{\alpha + |\{l : s_l = k, l \neq j\}|} \right] \times \prod_{i: x_{ij}=0} \left[ \frac{\alpha/2 + \sum_{l: s_l \neq k, l \neq j} (1 - x_{li})}{\alpha + |\{l : s_l \neq k, l \neq j\}|} \right] \times \frac{|\{l : s_l = k, l \neq j\}| + \beta/K}{m - 1 + \beta}$$

1-match term

0-match term

module size term

**- Convergence:** the number of oscillating genes reaches a plateau



**II. Gene clustering:** each gene is assigned to all biclusters where it spent at least **f** time during the sampling phase

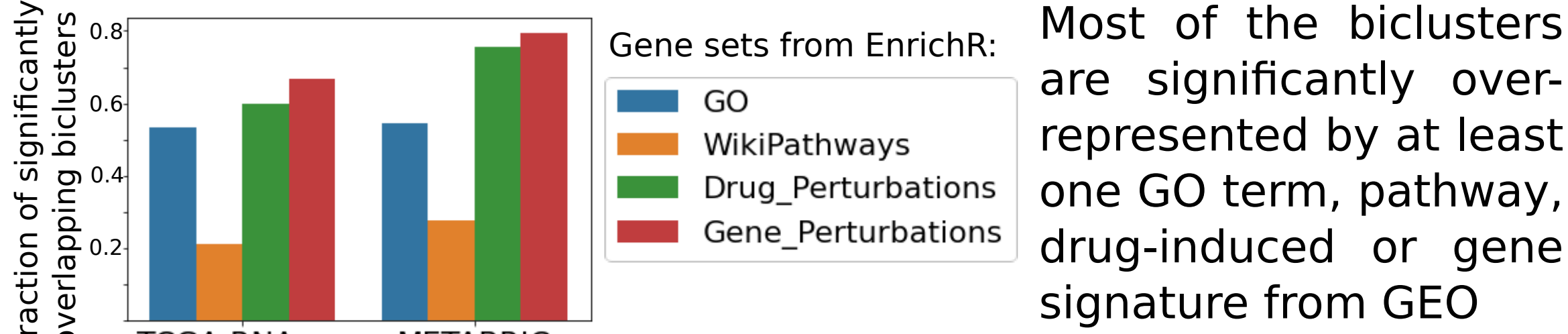
③ Postprocessing

**I.** Keeping **non-redundant** biclusters (Jaccard <0.5)

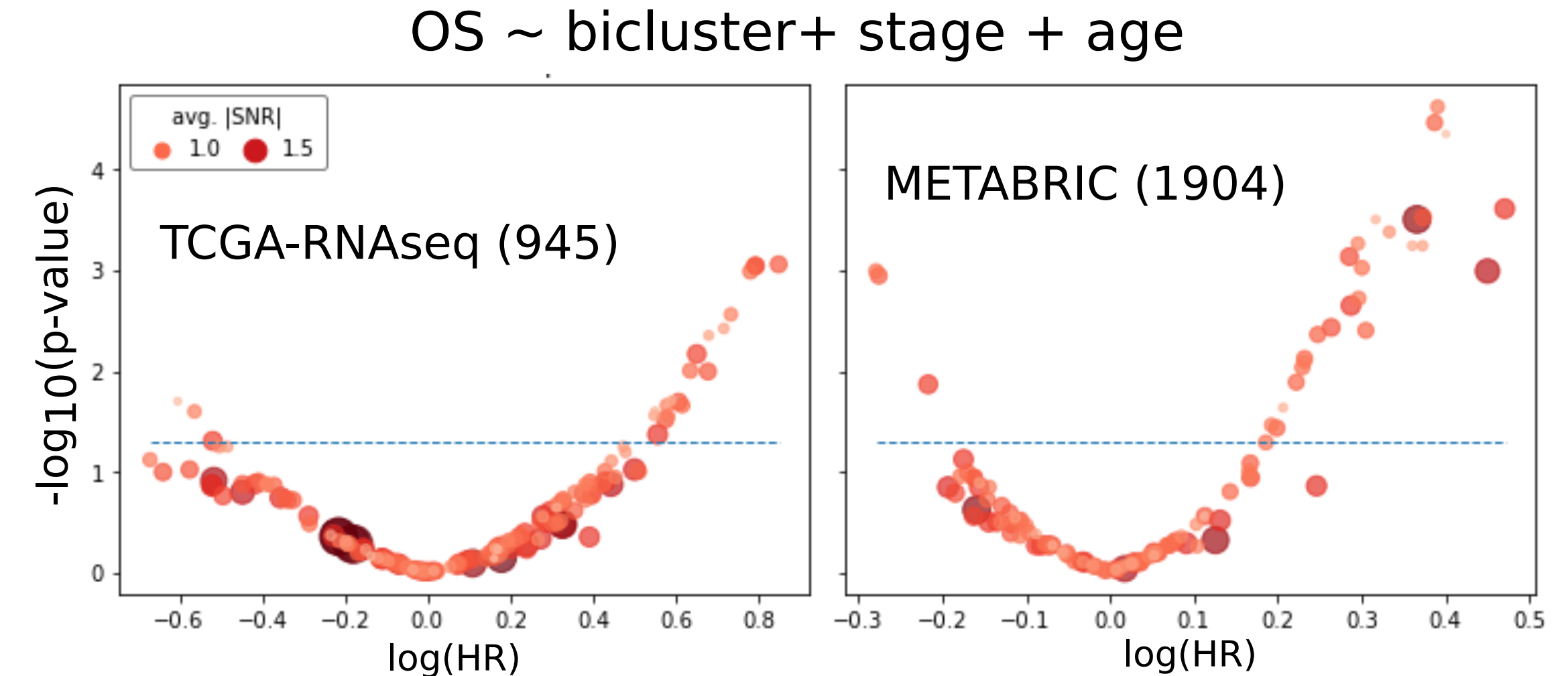
**II. Permutations:** filtering out biclusters with avg.SNR lower than random

EVALUATION

Gene set overrepresentation

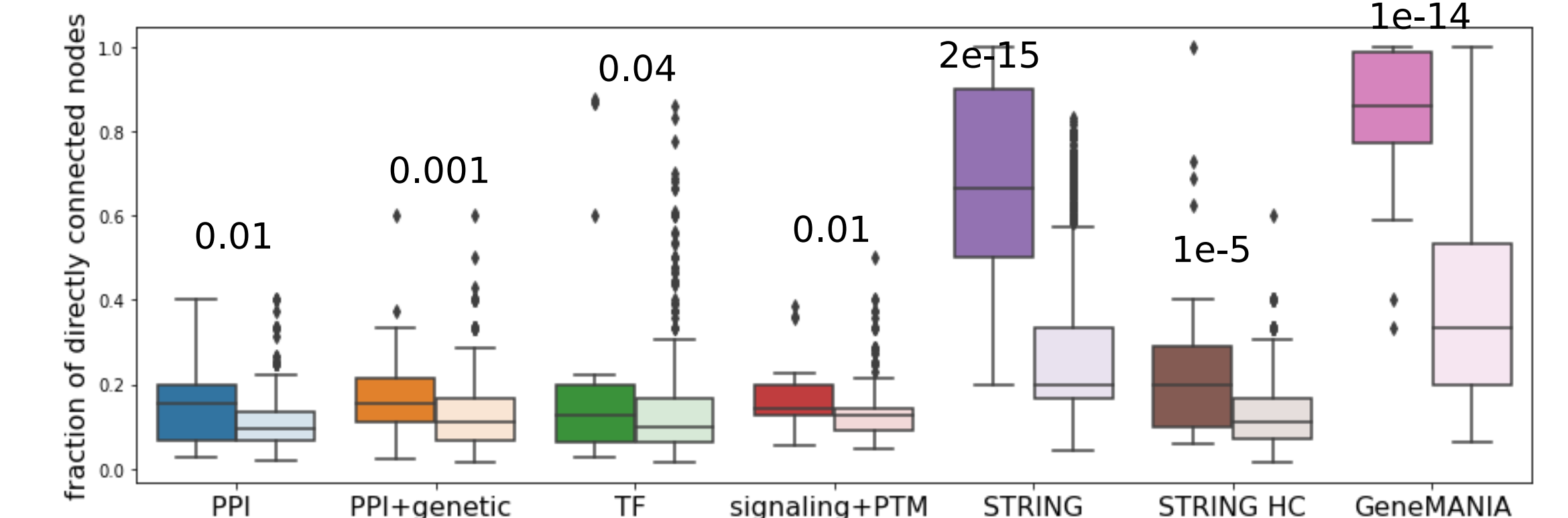


Survival analysis



Many biclusters identified by DESMOND2 are significantly associated with overall survival

CONNECTIVITY OF BICLUSTER GENES



**Fractions of directly connected bicluster genes.** Bright boxplots show the distributions for biclusters **matching between TCGA and METABRIC** (27-31 gene sets of 5-65 genes), pale boxplots - the distributions for 100 sets of size-matching random gene sets. Mann-Whitney test p-values are shown above each pair of boxplots.