Supplementary text

Permutation tests

Since multiple testing takes place while constructing the biclusters (we look through multiple node pairs to decide which of them are to be added to the bicluster), there could be some false positives. To address this issue and demonstrate that biclusters produced by DESMOND outperform null models, we calculated permutation-based empirical p-values to test two null hypotheses.

- To test the null hypothesis that DESMOND reports completely random biclusters, we shuffled TCGA-micro expression data 200 times in two modes and ran DESMOND on each of the shuffled datasets. The shuffling modes were the following: shuffling of the expression values in each sample across genes or shuffling the expression values for each gene across samples. The p-value for each particular bicluster from the original dataset was calculated as a rate of biclusters with the same or higher value among all the biclusters in the shuffled data. For each bicluster of a particular size, we considered only those biclusters in randomized data which were not smaller in terms of genes and samples than the original one, because smaller biclusters have more chances to demonstrate higher average SNR.

  No biclusters comprising four or more genes were found by DESMOND in shuffled data. However, when expressions were shuffled across samples within each gene, DESMOND detected a certain amount of 3-gene biclusters. Of 226 three-gene biclusters found in TCGA-micro dataset, 62 had empirical p-values of 0.05 and above.

- We demonstrated that DESMOND is able to identify biclusters with more pronounced differential expression than biclusters built on randomly selected subnetworks. For that, for every bicluster of n genes size, 1000 randomly chosen connected subnetworks of genes were considered. Then, for each random subnetwork, two groups of samples were identified by 2-means in a subspace of its genes. The $avg.|SNR|$ of the original bicluster was compared to the resulting empirical distribution of $avg.|SNR|$ computed for biclusters induced by random subnetworks. Same as above, we considered only non-smaller biclusters when computed empirical p-values for $avg.|SNR|$. All biclusters found in TCGA-micro, TCGA-RNAseq, and METABRIC had permutation p-values less than 0.05.

Evaluation on yeast datasets

To show that DESMOND is applicable to datasets other than human cancer, we run it and the baselines on two yeast datasets: DREAM5 and on the dataset from Hughes et al. DREAM5 datasets included 5667 genes profiled in 536 samples by Affymetrix Yeast Genome S98 Array. Preprocessed expression values were downloaded from supplementary

files of Marbach et al., 2012. Hughes et al. dataset comprised 300 samples subjected to different mutations and chemical perturbations and 5346 genes measured by a two-color microarray. This dataset was used for evaluation of DeBi method and therefore we took gene expressions preprocessed by Serin et al. deposited at https://github.com/esebesty/debi/blob/master/samples/yeast/hugues.txt.
All gene expressions were normalized and standardized prior to biclustering. Yeast PPI network obtained from BioGRID and included 4177 nodes and 74022 edges representing physical interactions validated in multiple experiments. Gene names were mapped using YeastMine (https://yeastmine.yeastgenome.org/) (Balakrishnan et al., 2012).

Again, all methods found biclusters of different shapes and SNR (Supplementary Figure S11). Same as on breast cancer data, QUBIC and DESMOND found biclusters with stronger differential expression and DESMOND produced more connected gene sets compared to other methods.

The functional coherence of the identified biclusters was evaluated in the same way as on breast cancer data (Supplementary Figure 12, left panels). Gene sets for GO and KEGG overrepresentation tests were obtained from YeastEnrichr (https://amp.pharm.mssm.edu/YeastEnrichr/).

Although 90.4% and 96.9% of DESMOND biclusters found in DREAM5 and Hughes datasets were enriched by at least one GO term, fractions of GO-enriched biclusters in outputs of ISA (on Hughes dataset) and QUBIC (on DREAM5 dataset) were even higher. DESMOND produced the highest percentage of biclusters overrepresented by KEGG pathways on Hughes et al. data, but lost to QUBIC on DREAM5.

In addition to GO and KEGG enrichment, we tested all found biclusters for co-regulation by transcription factors (TFs). Information on gene-TF relationships was obtained from *Saccharomyces* Genome Database ((Cherry et al., 2012)) via YeastMine (Balakrishnan et al., 2012). We used two types of TF-specific gene sets:

- genes regulated by individual TFs (264 gene sets)
- genes co-regulated by couples of TFs and genes regulated by individual TFs (6674 gene sets)

Only about 60% of DESMOND biclusters overlapped with known sets of co-regulated genes, which is inferior to QUBIC, ISA on both datasets, and COALESCE and DeBi on Hughes data (Supplementary Figure 12, right panels). This limited ability to identify putative regulatory modules is explained by the fact DESMOND is aimed at differentially expressed biclusters, rather than differentially co-expressed.

Investigation of OS-associated biclusters

Of all methods, only DESMOND, DeBi and QUBIC identified OS-associated biclusters in both TCGA-RNAseq and METABRIC. All OS-associated biclusters are listed in Supplementary Table 5.

QUBIC

QUBIC applied with default parameters found only two isolated OS-associated biclusters, 27x53 in TCGA-RNAseq (HR 3.07) and 18x101 in METABRIC (HR 1.84). Both biclusters had high avg.|SNR| of 0.9 and 1.26, and shared 8 genes (Jaccard similarity 0.22): *ESR1*, *CA12, AGR3, FOXA1, XBP1, GATA3, C1orf64, MLPH*.

Estrogen receptor alpha (ESR1) expression is a primordial feature in breast cancer, which strongly influences treatment selection (Reinert et al., 2019). CA12 and AGR3 expression was shown to be correlated with ESR1 expression (Barnett et al., 2008; Garczyk et al., 2015). CA12 contributes to the acidification of the tumor microenvironment (Barnett et al., 2008), and AGR3 promotes breast cancer cell migration (Obacz et al., 2019).

FOXA1 expression correlates with luminal subtype and represses basal subtype-associated genes, suppressing breast cancer cells migration (Bernardo et al., 2013). GATA3 is also associated with luminal subtype (Shaoxian et al. 2017) and is necessary for ESR1 transcription ("The Significance of GATA3 Expression in Breast Cancer: A 10-Year Follow-up Study," 2009). XBP1 promotes triple-negative breast cancer and is associated with unfavourable outcome in breast cancer (X. Chen et al., 2014). C1orf64 tends to be coexpressed with androgen receptor in breast cancer and is involved in the interaction of estrogen and androgen receptors (Naderi, 2017). MLPH is a Rab effector protein included in the ER-positive breast cancer prognostic signature (Thakkar et al., 2010). Overall, these genes are upregulated or downregulated in ER-positive breast tumors and have prognostic significance in breast cancer.

With optimized parameters, QUBIC found 3 OS-associated biclusters with similar HR of 1.26-1.3 and overlapping in samples and genes (Jaccard similarities 0.74-0.91 and 0.44-0.79 for gene and samples overlaps respectively). In TCGA-RNAseq only one OS-associated bicluster of 501 genes and 219 samples was found. It was similar to three biclusters detected in METABRIC, shared with them 143-264 genes (Jaccard similarities 0.25-0.36), and had HR of 1.66.

All OS-associated biclusters found by QUBIC were significantly enriched by samples with Basal subtype. Interestingly, Basal subtype itself demonstrates weaker associations with OS in these two cohorts than biclusters found by QUBIC (HR 1.18 and 1.26 in TCGA-RNAseq and METABRIC, p-values less than 0.50 and 0.06 respectively).

DeBi

With default parameters, DeBi found 37 biclusters in TCGA-RNAseq and 11 in METABRIC. Biclusters found in METABRIC were very small in samples (10-16) and had low avg. |SNR| 0.52-0.63, compared to 20-106 samples and avg.|SNR| of 0.53-0.94 in TCGA-RNAseq. Only

2 of 11 biclusters OS-associated biclusters showed significant associations with molecular subtypes: one with Basal and the other with LumA. Most TCGA-RNAseq biclusters demonstrated significant overlap with Basal (10) and LumB (7) subtypes, or under-representation of LumA (13).

The biggest overlap between OS-associated biclusters found in TCGA and METABRIC was observed between biclusters of sizes 445x62 and 306x13 and included 72 genes (J=0.11, 11 shared genes are expected on average). These genes included cell division cycle-associated proteins, kinesins, DNA replication licensing factors, centromere proteins, cell cycle regulators, etc. They are involved in mitosis and cell cycle, which is associated with proliferation, a known prognostic factor in breast cancer (van Diest et al., 2004). Both biclusters demonstrated relatively low avg.|SNR| of 0.59. The bicluster of 442 genes and 62 samples found in TCGA-RNAseq included 34 samples with Basal subtype, when its best match from METABRIC demonstrated no significant overlap with any subtype.

With optimized parameters DeBi found 23 biclusters in TCGA-RNAseq and 5 in METABRIC. These biclusters had higher avg.|SNR| than biclusters found with default parameters (0.73-1.23 and 0.63-1.1 in TCGA-RNAseq and METABRIC). The biggest overlap was between biclusters of sizes 15x15 (TCGA-RNAseq, HR=5.92; p-value<2.40e-02) and 112x68 (METABRIC, HR=2.24;p-value<1.33e-03) which shared 7 genes (0.14 expected by chance): *FHOD1, GGT1, KYNU, LBP, PNLIPRP3, POR, TRPV6.* Both of these overlapping biclusters were enriched by Her2 subtype. Hazard ratios for patient groups defined by these biclusters were much higher than for the whole group of patients with Her2-positive disease (1.94 in TCGA-RNAseq and 1.54 in METABRIC).

FHOD1 is associated with EMT and is overexpressed in basal-like triple-negative breast cancer (Heuser et al., 2018). GGT1 serum level may be used as a breast cancer biomarker, however, not much is known about GGT1 expression in the tumor (Shackshaft et al., 2017). KYNU is involved in tryptophan metabolism and is associated with basal breast cancer subtype (Heng et al., 2016). PNLIPRP3 is a part of CBFB and RUNX1-influenced prognostic signature for breast cancer (Malik et al., 2019). TRPV6 is a calcium channel overexpressed in ER-negative breast cancer and associated with decreased survival (pubmeddev & Peters AA, n.d.). Thus, the overlapping genes are important for breast cancer biology but do not seem to be functionally connected or specifically associated with Her2 subtype.

DESMOND

The majority of OS-associated biclusters found by DESMOND were associated with LumA subtype. Absolute coefficients in Cox proportional hazards models fitted for patient groups defined by these biclusters were nearly the same as for LumA subtype: about 0.52 in TCGA-RNAseq and 0.28 in METABRIC. When tested OS of LumA subtype versus all other patients, coefficients had negative signs and HR values were below 1. This means that patients with LumA subtype of breast cancer demonstrate improved OS, compared to other patients. Interestingly, many of OS-associated biclusters found by DESMOND, significantly overlapped with the complement of LumA sample set, had positive coefficients and HR more than 1.

For example, the largest overlap was observed between two LumA-associated biclusters: of 132 genes and 482 samples in TCGA-RNA (SNR=0.86) and 66 genes and 840 samples (SNR=0.84) in METABRIC shared 47 genes (J=0.31): *ANLN, ATAD2, AURKA, AURKB, BIRC5, BUB1, CASC5, CBX2, CCNA2, CCNB1, CCNB2, CDC20, CDC25A, CDC25C, CDCA2, CDCA5, CDCA8, CENPA, CENPE, CENPO, CKS2, FOXM1, HIST1H2BG, INCENP, KIF18A, KIF20A, KIF23, KIF2C, KIF4A, MAD2L1, MCM10, NDC80, NEK2, NUF2, NUSAP1, PLK1, PLK4, PRC1, RACGAP1, SPC25, TOP2A, TPX2, TRIP13, TROAP, TTK, UBE2C, ZWINT.* The first bicluster found in TCGA included 376 samples annotated with LumA (adjusted hypergeometric p-value < 1.37e-100). Its samples belonged to patients with better OS (HR=0.59, adjusted p-value < 4.72e-02). Its best match by genes in METABRIC, on the opposite, included samples from patients with lower OS (HR=1.32, adjusted p-value < 4.72e-02). Samples annotated with LumA were depleted in this bicluster.

The overlapping genes are associated with cell cycle and mitosis. Anilin is involved in cytokinesis. ATAD2 is involved in reading epigenetic marks during DNA replication (Koo et al., 2016). AURKA and AURKB are involved in chromosome segregation and are shown to be overexpressed in breast cancer (Tang et al., 2017). Survivin (BIRC5) is also involved in cell division, as well as in apoptosis inhibition (Jha et al., 2012). CDCA2, CDCA5, and CDCA8 are cell division cycle-associated proteins, and their higher expression was shown to be associated with reduced survival in breast cancer (Phan et al., 2018). CENPA, CENPE and CENPO are centromere proteins involved in cell division by participating in the centromere formation. CENPA was shown to be associated with negative patient prognosis in breast cancer (Sun et al., 2016).

CKS2 is a subunit of cyclin-dependent kinases and is associated with poor prognosis in breast cancer (Huang et al., 2019). FOXM1 is also involved in cell division and is shown to be overexpressed in primary breast cancer (Wonsey & Follettie, 2005). INCENP is a centromere protein interacting with AURKB (Xu et al., 2010). KIF18A, KIF20A, KIF23, KIF2C and KIF4A are kinesin superfamily members involved in transport of chromosomes during mitosis. Some of them are associated with poor prognosis in breast cancer (Lucanus & Yip, 2018). MAD2L1 is a component of the mitotic spindle assembly checkpoint. MCM10 is a DNA replication licensing factor overexpressed in breast cancer (Mahadevappa et al., 2018). NDC80 is a kinetochore protein associated with negative prognosis in breast cancer (Meng et al., 2015). NUF2 and SPC25 are also components of the NDC80 complex.

NEK2 is a cell cycle-associated protein kinase important for breast cancer proliferation (Cappello et al., 2014). NUSAP1 is a nucleolar spindle-associated protein and has a negative prognostic significance in breast cancer (L. Chen et al., 2015). PLK1 and PLK4 are known cell cycle regulators (Lee et al., 2014). PRC1 is involved in cytokinesis (Jiang et al., 1998). RACGAP1 is also essential for cytokinesis and was shown to have prognostic significance in luminal breast cancer (Milde-Langosch et al., 2013).

TOP2A encodes for DNA topoisomerase II alpha and was shown to be associated with proliferation in breast cancer (An et al., 2018). TPX2 is a microtubule-associated protein that is involved in mitotic spindle organization (Wittmann et al., 2000). TRIP13 is involved in spindle assembly checkpoint and DNA repair and is overexpressed in breast cancer (Lu et al., 2019). TROAP is required for spindle assembly and is involved in breast cancer

proliferation (Li et al., 2019). TTK is a component of the spindle assembly complex (Kaistha et al., 2014; Yang et al., 2018). UBE2C is involved in protein degradation and has a negative prognostic significance in breast cancer (Mo et al., 2017). ZWINT is involved in the kinetochore formation (Yang et al., 2018).

One of the primordial features of the luminal A subtype tumors is low proliferation marker expression (Hashmi et al., 2018). Thus, the overlap of the luminal A-correlated gene sets found by DESMOND containing exclusively mitosis- and proliferation-associated genes may serve as a biological confirmation of the method validity and of the benefit of the network-constrained approach.

References

An, X., Xu, F., Luo, R., Zheng, Q., Lu, J., Yang, Y., Qin, T., Yuan, Z., Shi, Y., Jiang, W., & Wang, S. (2018). The prognostic significance of topoisomerase II alpha protein in early stage luminal breast cancer. *BMC Cancer*, *18*(1), 331.

Balakrishnan, R., Park, J., Karra, K., Hitz, B. C., Binkley, G., Hong, E. L., Sullivan, J., Micklem, G., & Michael Cherry, J. (2012). YeastMine—an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. In *Database* (Vol. 2012). https://doi.org/10.1093/database/bar062

Barnett, D. H., Sheng, S., Charn, T. H., Waheed, A., Sly, W. S., Lin, C.-Y., Liu, E. T., & Katzenellenbogen, B. S. (2008). Estrogen receptor regulation of carbonic anhydrase XII through a distal enhancer in breast cancer. *Cancer Research*, *68*(9), 3505–3515.

Bernardo, G. M., Bebek, G., Ginther, C. L., Sizemore, S. T., Lozada, K. L., Miedler, J. D., Anderson, L. A., Godwin, A. K., Abdul-Karim, F. W., Slamon, D. J., & Keri, R. A. (2013). FOXA1 represses the molecular phenotype of basal breast cancer cells. *Oncogene*, *32*(5), 554–563.

Boppana, R., & Halldórsson, M. M. (1992). Approximating maximum independent sets by excluding subgraphs. In *BIT* (Vol. 32, Issue 2, pp. 180–196). https://doi.org/10.1007/bf01994876

Cappello, P., Blaser, H., Gorrini, C., Lin, D. C. C., Elia, A. J., Wakeham, A., Haider, S.,

Boutros, P. C., Mason, J. M., Miller, N. A., Youngson, B., Done, S. J., & Mak, T. W. (2014). Role of Nek2 on centrosome duplication and aneuploidy in breast cancer cells. *Oncogene*, *33*(18), 2375–2384.

Chen, L., Yang, L., Qiao, F., Hu, X., Li, S., Yao, L., Yang, X.-L., & Shao, Z.-M. (2015). High Levels of Nucleolar Spindle-Associated Protein and Reduced Levels of BRCA1 Expression Predict Poor Prognosis in Triple-Negative Breast Cancer. *PloS One*, *10*(10), e0140572.

Chen, X., Iliopoulos, D., Zhang, Q., Tang, Q., Greenblatt, M. B., Hatziapostolou, M., Lim, E., Tam, W. L., Ni, M., Chen, Y., Mai, J., Shen, H., Hu, D. Z., Adoro, S., Hu, B., Song, M., Tan, C., Landis, M. D., Ferrari, M., … Glimcher, L. H. (2014). XBP1 promotes triple-negative breast cancer by controlling the HIF1α pathway. *Nature*, *508*(7494), 103–107.

Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., … Wong, E. D. (2012). Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research*, *40*(Database issue), D700–D705.

Garczyk, S., von Stillfried, S., Antonopoulos, W., Hartmann, A., Schrauder, M. G., Fasching, P. A., Anzeneder, T., Tannapfel, A., Ergönenc, Y., Knüchel, R., Rose, M., & Dahl, E. (2015). AGR3 in breast cancer: prognostic impact and suitable serum-based biomarker for early cancer detection. *PloS One*, *10*(4), e0122106.

Hashmi, A. A., Aijaz, S., Khan, S. M., Mahboob, R., Irfan, M., Zafar, N. I., Nisar, M., Siddiqui, M., Edhi, M. M., Faridi, N., & Khan, A. (2018). Prognostic parameters of luminal A and luminal B intrinsic breast cancer subtypes of Pakistani patients. *World Journal of Surgical Oncology*, *16*(1), 1.

Heng, B., Lim, C. K., Lovejoy, D. B., Bessede, A., Gluch, L., & Guillemin, G. J. (2016).

Understanding the role of the kynurenine pathway in human breast cancer immunobiology. *Oncotarget*, *7*(6), 6506–6520.

Heuser, V. D., Mansuri, N., Mogg, J., Kurki, S., Repo, H., Kronqvist, P., Carpén, O., & Gardberg, M. (2018). Formin Proteins FHOD1 and INF2 in Triple-Negative Breast Cancer: Association With Basal Markers and Functional Activities. *Breast Cancer: Basic and Clinical Research*, *12*, 1178223418792247.

Huang, N., Wu, Z., Hong, H., Wang, X., Yang, F., & Li, H. (2019). Overexpression of CKS2 is associated with a poor prognosis and promotes cell proliferation and invasion in breast cancer. *Molecular Medicine Reports*, *19*(6), 4761–4769.

Jha, K., Shukla, M., & Pandey, M. (2012). Survivin expression and targeting in breast cancer. In *Surgical Oncology* (Vol. 21, Issue 2, pp. 125–131). https://doi.org/10.1016/j.suronc.2011.01.001

Jiang, W., Jimenez, G., Wells, N. J., Hope, T. J., Wahl, G. M., Hunter, T., & Fukunaga, R. (1998). PRC1: a human mitotic spindle-associated CDK substrate protein required for cytokinesis. *Molecular Cell*, *2*(6), 877–885.

Kaistha, B. P., Honstein, T., Müller, V., Bielak, S., Sauer, M., Kreider, R., Fassan, M., Scarpa, A., Schmees, C., Volkmer, H., Gress, T. M., & Buchholz, M. (2014). Key role of dual specificity kinase TTK in proliferation and survival of pancreatic cancer cells. *British Journal of Cancer*, *111*(9), 1780–1787.

Koo, S. J., Fernández-Montalván, A. E., Badock, V., Ott, C. J., Holton, S. J., von Ahsen, O., Toedling, J., Vittori, S., Bradner, J. E., & Gorjánácz, M. (2016). ATAD2 is an epigenetic reader of newly synthesized histone marks during DNA replication. *Oncotarget*, *7*(43), 70323–70335.

Lee, S.-Y., Jang, C., & Lee, K.-A. (2014). Polo-like kinases (plks), a key regulator of cell cycle and new potential target for cancer therapy. *Development & Reproduction*, *18*(1), 65–71.

Li, K., Zhang, R., Wei, M., Zhao, L., Wang, Y., Feng, X., Yang, Y., Yang, S., & Zhang, L. (2019). TROAP Promotes Breast Cancer Proliferation and Metastasis. *BioMed Research International*, *2019*, 6140951.

Lucanus, A. J., & Yip, G. W. (2018). Kinesin superfamily: roles in breast cancer, patient prognosis and therapeutics. *Oncogene*, *37*(7), 833–838.

Lu, S., Qian, J., Guo, M., Gu, C., & Yang, Y. (2019). Insights into a Crucial Role of TRIP13 in Human Cancer. *Computational and Structural Biotechnology Journal*, *17*, 854–861.

Mahadevappa, R., Neves, H., Yuen, S. M., Jameel, M., Bai, Y., Yuen, H.-F., Zhang, S.-D., Zhu, Y., Lin, Y., & Kwok, H. F. (2018). DNA Replication Licensing Protein MCM10 Promotes Tumor Progression and Is a Novel Prognostic Biomarker and Potential Therapeutic Target in Breast Cancer. *Cancers*, *10*(9). https://doi.org/10.3390/cancers10090282

Malik, N., Yan, H., Moshkovich, N., Palangat, M., Yang, H., Sanchez, V., Cai, Z., Peat, T. J., Jiang, S., Liu, C., Lee, M., Mock, B. A., Yuspa, S. H., Larson, D., Wakefield, L. M., & Huang, J. (2019). The transcription factor CBFB suppresses breast cancer through orchestrating translation and transcription. *Nature Communications*, *10*(1), 2071.

Meng, Q.-C., Wang, H.-C., Song, Z.-L., Shan, Z.-Z., Yuan, Z., Zheng, Q., & Huang, X.-Y. (2015). Overexpression of NDC80 is correlated with prognosis of pancreatic cancer and regulates cell proliferation. *American Journal of Cancer Research*, *5*(5), 1730–1740.

Milde-Langosch, K., Karn, T., Müller, V., Witzel, I., Rody, A., Schmidt, M., & Wirtz, R. M. (2013). Validity of the proliferation markers Ki67, TOP2A, and RacGAP1 in molecular subgroups of breast cancer. *Breast Cancer Research and Treatment*, *137*(1), 57–67.

Mo, C.-H., Gao, L., Zhu, X.-F., Wei, K.-L., Zeng, J.-J., Chen, G., & Feng, Z.-B. (2017). The clinicopathological significance of UBE2C in breast cancer: a study based on immunohistochemistry, microarray and RNA-sequencing data. *Cancer Cell International*, *17*, 83.

Naderi, A. (2017). C1orf64 is a novel androgen receptor target gene and coregulator that interacts with 14-3-3 protein in breast cancer. *Oncotarget*, *8*(34), 57907–57933.

Obacz, J., Sommerova, L., Sicari, D., Durech, M., Avril, T., Iuliano, F., Pastorekova, S., Hrstka, R., Chevet, E., Delom, F., & Fessart, D. (2019). Extracellular AGR3 regulates breast cancer cells migration via Src signaling. *Oncology Letters*, *18*(5), 4449–4456.

Phan, N. N., Wang, C.-Y., Li, K.-L., Chen, C.-F., Chiao, C.-C., Yu, H.-G., Huang, P.-L., & Lin, Y.-C. (2018). Distinct expression of CDCA3, CDCA5, and CDCA8 leads to shorter relapse free survival in breast cancer patient. *Oncotarget*, *9*(6), 6977–6992.

pubmeddev, & Peters AA, E. al. (n.d.). *Calcium channel TRPV6 as a potential therapeutic target in estrogen receptor-negative breast cancer. - PubMed - NCBI*. Retrieved December 2, 2019, from https://www.ncbi.nlm.nih.gov/pubmed/22807578

Reinert, T., Coelho, G. P., Mandelli, J., Zimermann, E., Zaffaroni, F., Bines, J., Barrios, C. H., & Graudenz, M. S. (2019). Association of ESR1 Mutations and Visceral Metastasis in Patients with Estrogen Receptor-Positive Advanced Breast Cancer from Brazil. In *Journal of Oncology* (Vol. 2019, pp. 1–5). https://doi.org/10.1155/2019/1947215

Shackshaft, L., Van Hemelrijck, M., Garmo, H., Malmström, H., Lambe, M., Hammar, N., Walldius, G., Jungner, I., & Wulaningsih, W. (2017). Circulating gamma-glutamyl transferase and development of specific breast cancer subtypes: findings from the Apolipoprotein Mortality Risk (AMORIS) cohort. *Breast Cancer Research: BCR*, *19*(1), 22.

Sun, X., Clermont, P.-L., Jiao, W., Helgason, C. D., Gout, P. W., Wang, Y., & Qu, S. (2016). Elevated expression of the centromere protein-A(CENP-A)-encoding gene as a prognostic and predictive biomarker in human cancers. *International Journal of Cancer. Journal International Du Cancer*, *139*(4), 899–907.

Tang, A., Gao, K., Chu, L., Zhang, R., Yang, J., & Zheng, J. (2017). Aurora kinases: novel therapy targets in cancers. *Oncotarget*, *8*(14), 23937–23954.

Thakkar, A. D., Raj, H., Chakrabarti, D., Ravishankar, Saravanan, N., Muthuvelan, B., Balakrishnan, A., & Padigaru, M. (2010). Identification of Gene Expression Signature in Estrogen Receptor Positive Breast Carcinoma. In *Biomarkers in Cancer* (Vol. 2, p. BIC.S3793). https://doi.org/10.4137/bic.s3793

The significance of GATA3 expression in breast cancer: a 10-year follow-up study. (2009). *Human Pathology*, *40*(4), 489–495.

van Diest, P. J., van der Wall, E., & Baak, J. P. A. (2004). Prognostic value of proliferation in invasive breast cancer: a review. *Journal of Clinical Pathology*, *57*(7), 675–681.

Wittmann, T., Wilm, M., Karsenti, E., & Vernos, I. (2000). Tpx2, a Novel Xenopus Map Involved in Spindle Pole Organization. *The Journal of Cell Biology*, *149*(7), 1405.

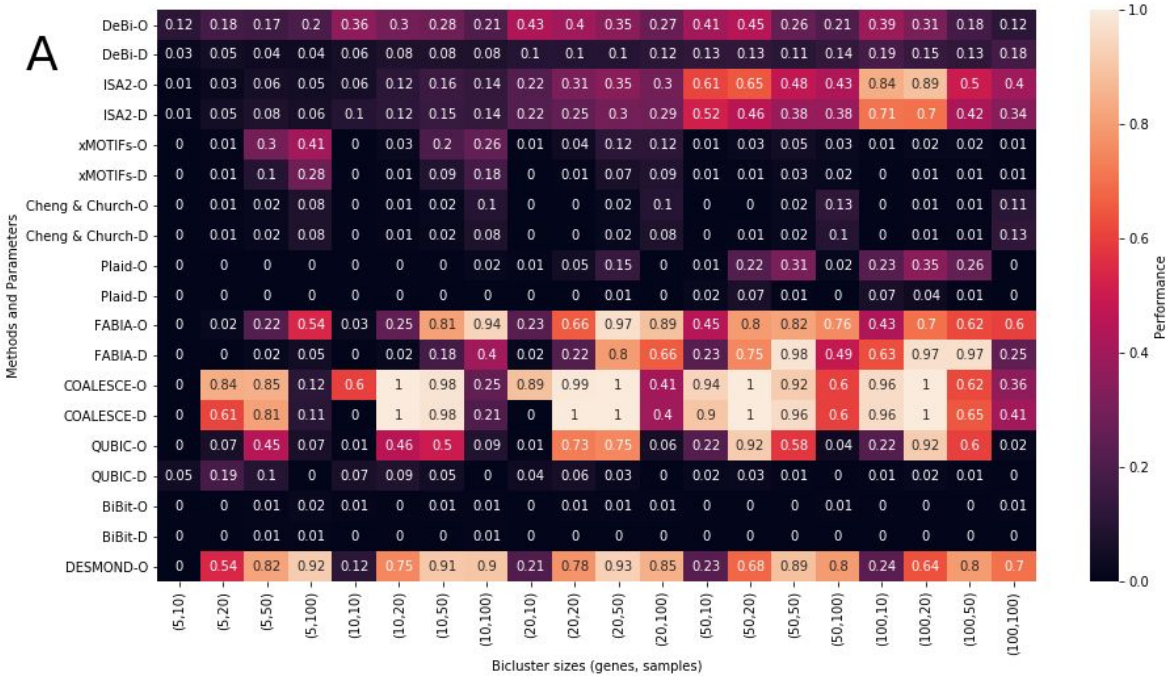
Wonsey, D. R., & Follettie, M. T. (2005). Loss of the forkhead transcription factor FoxM1 causes centrosome amplification and mitotic catastrophe. *Cancer Research*, *65*(12), 5181–5189.

Xu, Z., Ogawa, H., Vagnarelli, P., Bergmann, J. H., Hudson, D. F., Ruchaud, S., Fukagawa, T., Earnshaw, W. C., & Samejima, K. (2010). INCENP–aurora B interactions modulate kinase activity and chromosome passenger complex localization. In *The Journal of Cell Biology* (Vol. 188, Issue 4, pp. 611–611). https://doi.org/10.1083/jcb.20090605320100202c

Yang, X.-Y., Wu, B., Ma, S.-L., Yin, L., Wu, M.-C., & Li, A.-J. (2018). Decreased Expression of ZWINT is Associated With Poor Prognosis in Patients With HCC After Surgery. *Technology in Cancer Research & Treatment*, *17*, 1533033818794190.

Shaoxian, T., Baohua, Y., Xiaoli, X., Yufan, C., Xiaoyu, T., Hongfen, L., ... Wentao, Y. (2017). Characterisation of GATA3 expression in invasive breast cancer: Differences in histological subtypes and immunohistochemically defined molecular subtypes. *Journal of Clinical Pathology*, *70*(11), 926-934.

Supplementary Figures



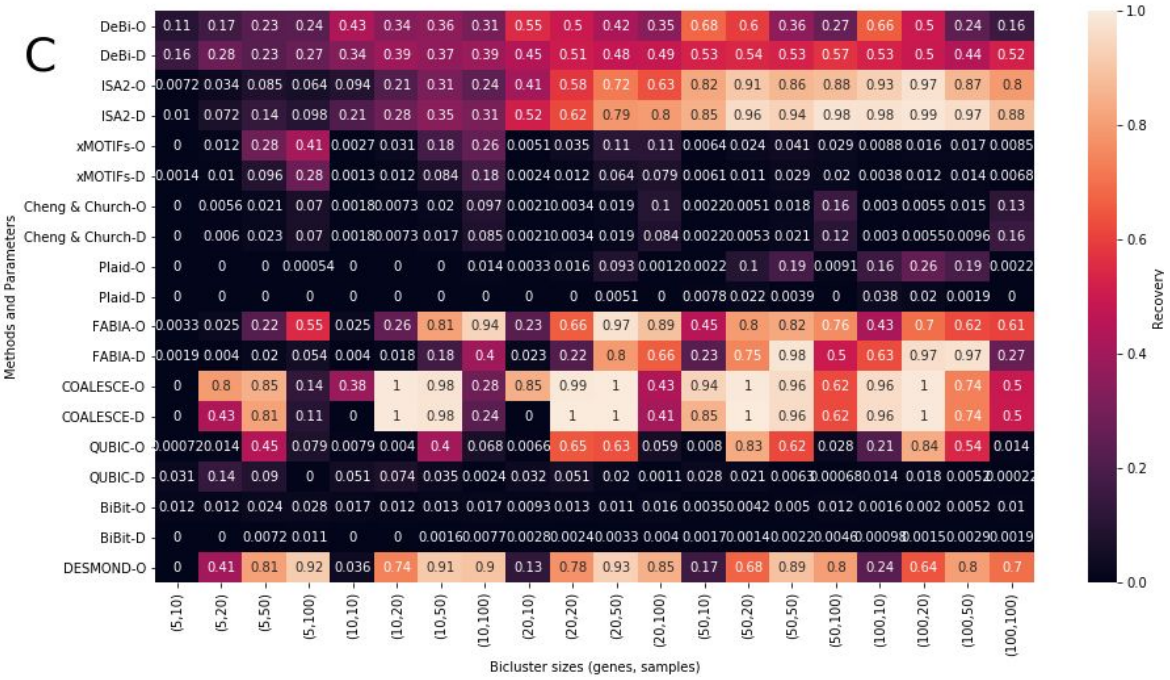


Figure S1. *Performance (A), Relevance (B), and Recovery (C) scores demonstrated by DESMOND and baseline methods on 20 synthetic datasets containing biclusters of different shapes. For non-deterministic methods, average performance in 10 runs is reported. For each of baselines, performance scores for default (D) and optimized (O) parameters are reported.*

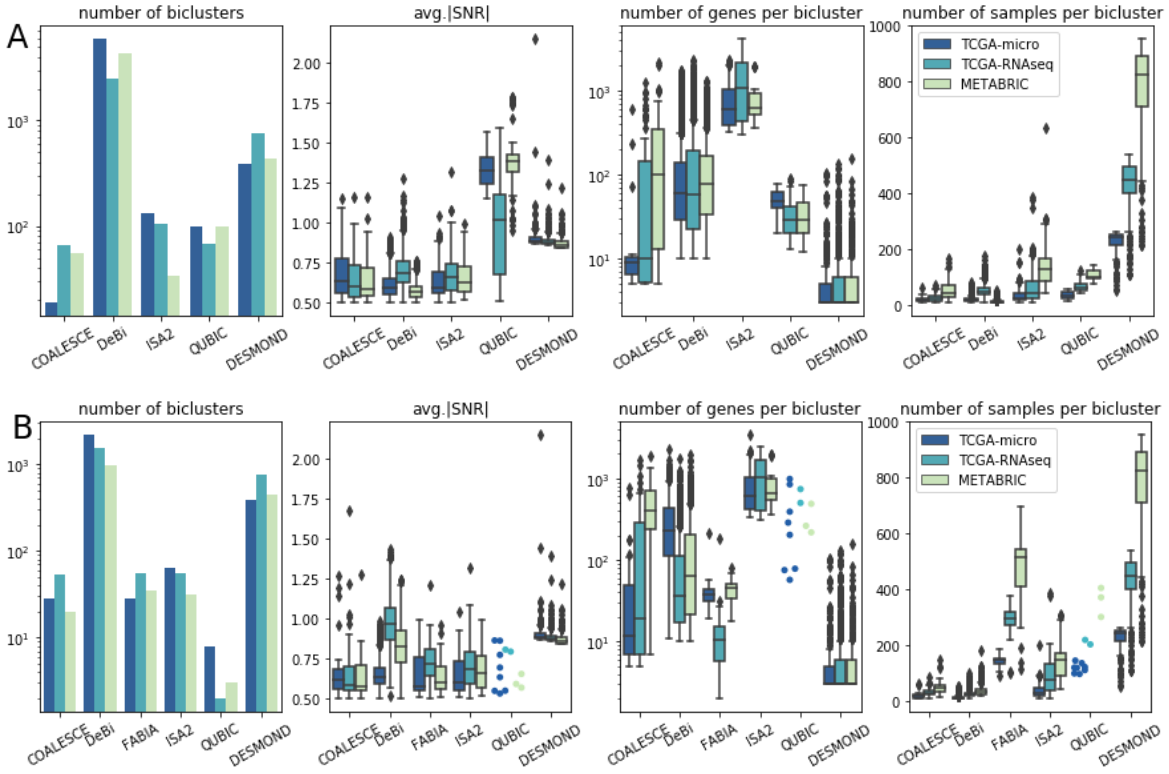


Figure S2. *Characteristics of differentially expressed biclusters produced by DESMOND and baseline methods in TCGA and METABRIC data with default (A) and optimized (B) parameters. Since QUBIC produced less than 10 biclusters on all real datasets with optimized parameters, its results are represented by dots instead of boxplots.*

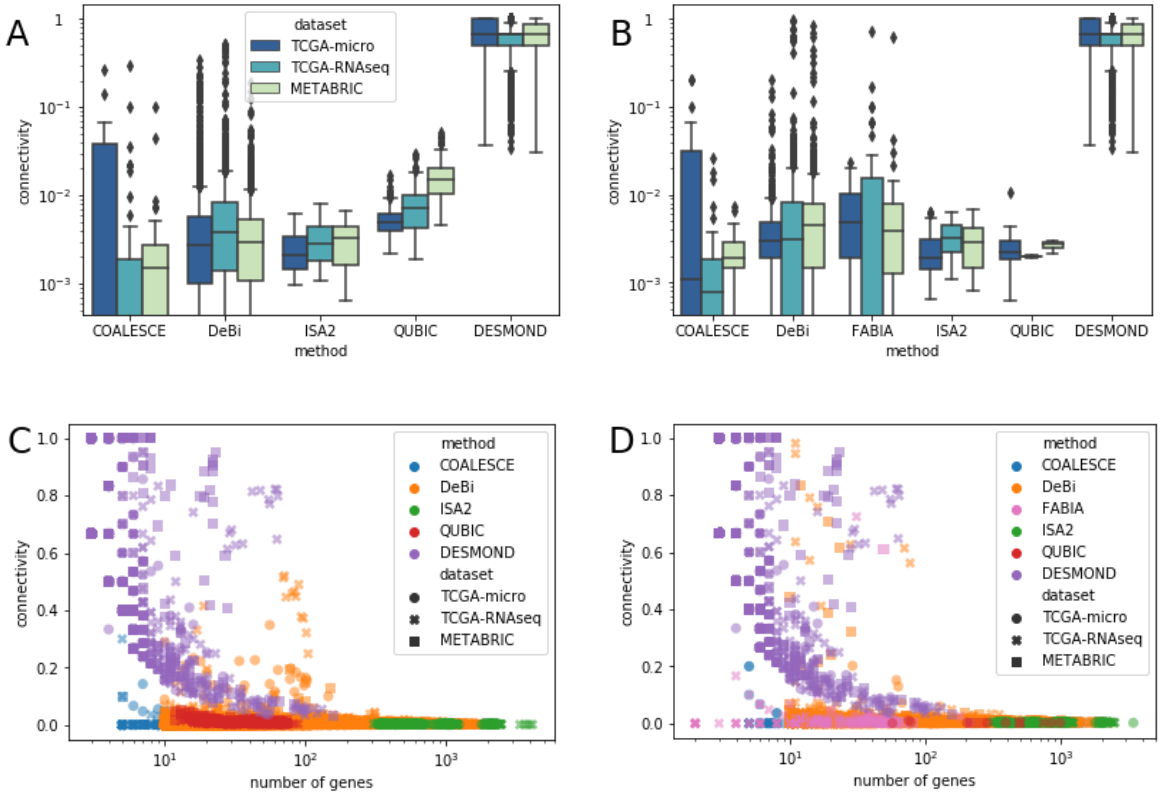


Figure S3. *Connectivity of biclusters found with default (A, C) and optimized (B, D) parameters. Subnetwork connectivity was measured as a ratio of the actual edge number to the maximal possible number of edges in a fully connected subnetwork with the same number of nodes. Connectivity varies between 0 when all nodes are disconnected and 1 when the network is fully connected. As expected, biclusters produced by DESMOND demonstrated higher connectivity on average than biclusters found by all other methods. Panels C and D demonstrate the dependency between connectivity and size in terms of genes for biclusters found by all compared methods.*

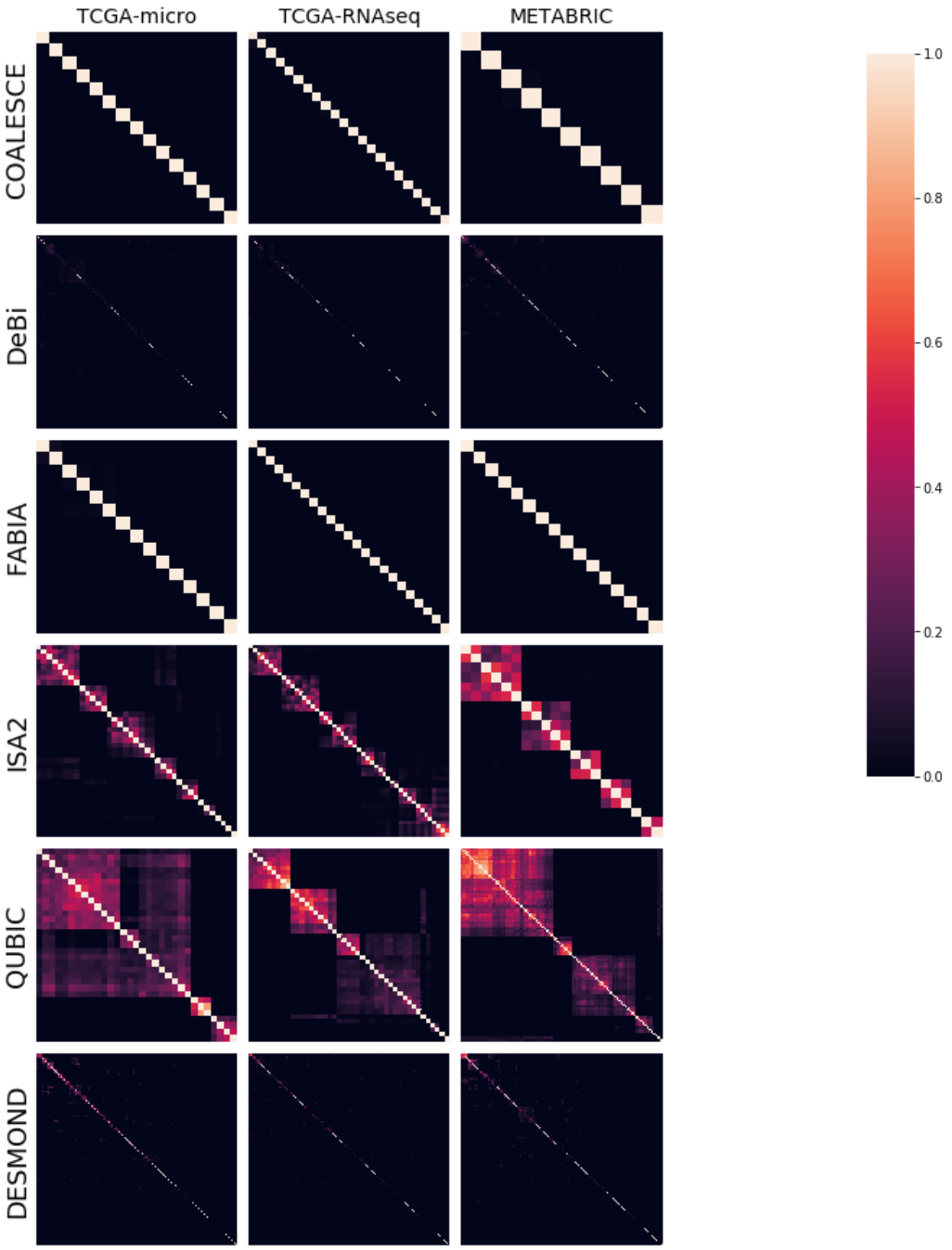


Figure S4. *Redundancy of biclusters discovered by each method in TCGA-micro, TCGA-RNAseq and METABRIC. Each heatmap shows Jaccard similarities computed for each pair of biclusters, considering genes and samples:* $J = \frac{|G_1 \cap G_2||S_1 \cap S_2|}{|G_1 \backslash G2||S_1| + |G_1 \cap G_2||S_1 \cup S_2| + |G_2 \backslash G_1||S_2|}$*, where* G_1, G_2 *are gene sets, and* S_1*,* S_2 *are sample sets from biclusters 1 and 2. QUBIC and ISA2 identify biclusters strongly overlapping in genes and samples. Biclusters found by other methods are non-redundant or almost non-redundant.*

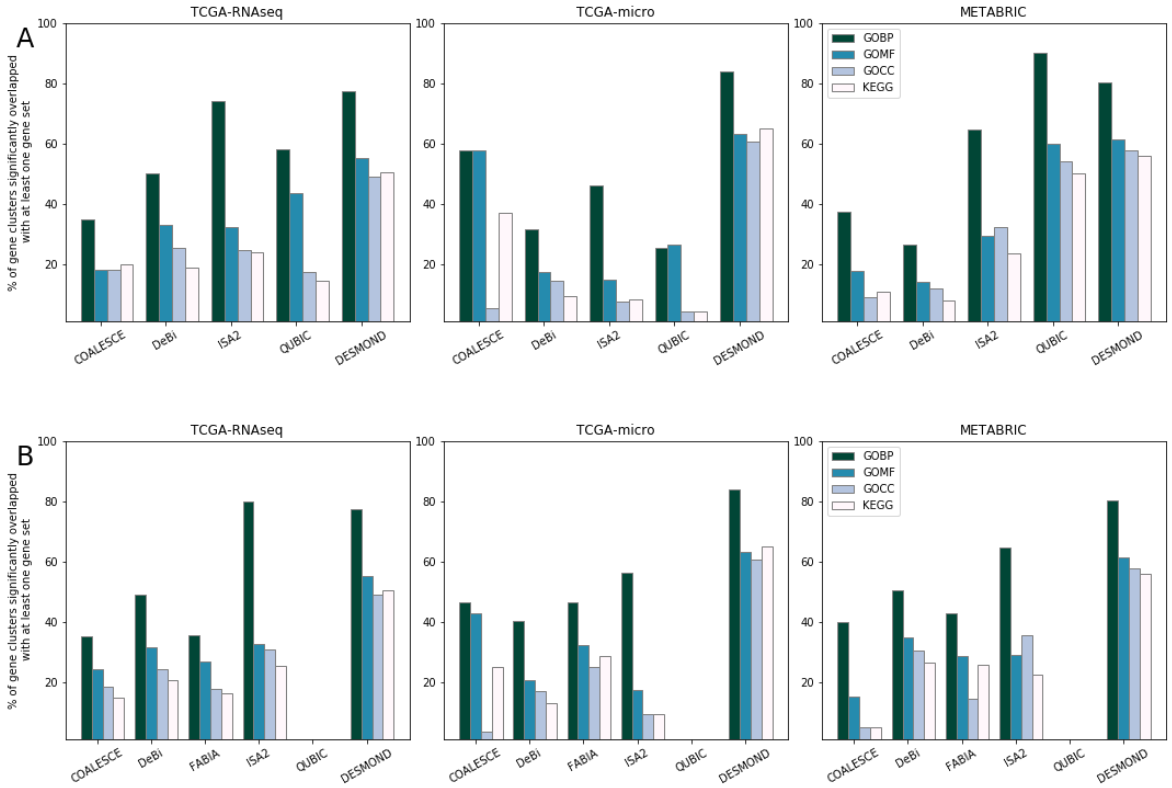
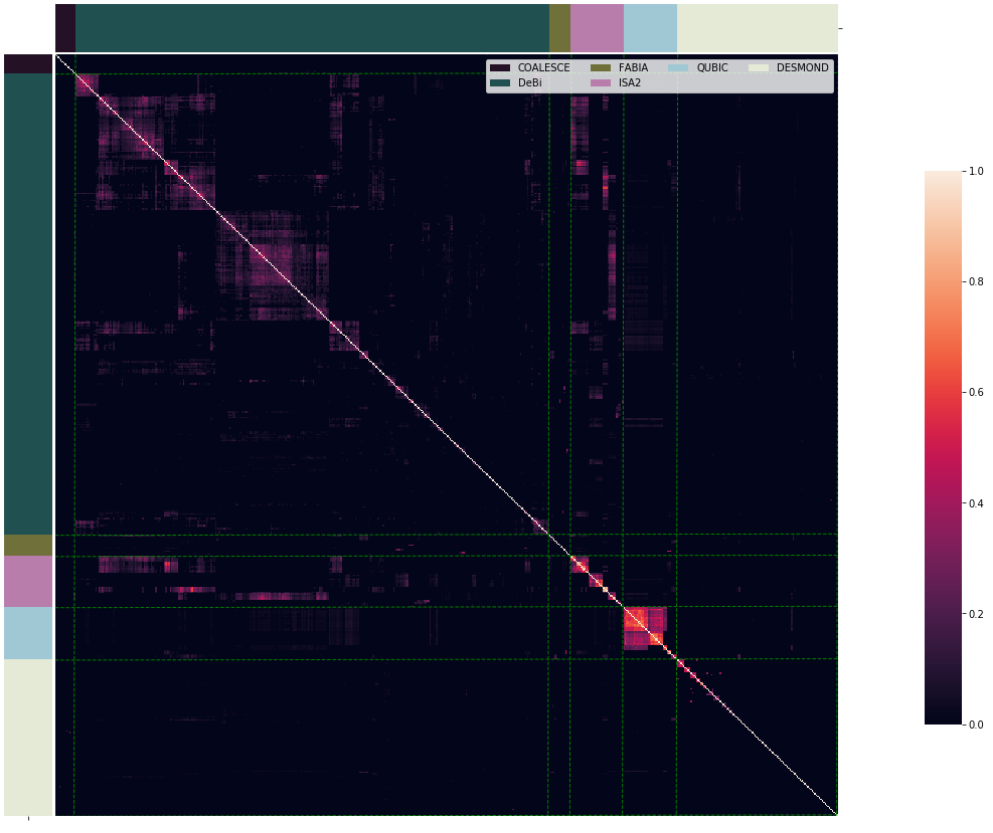
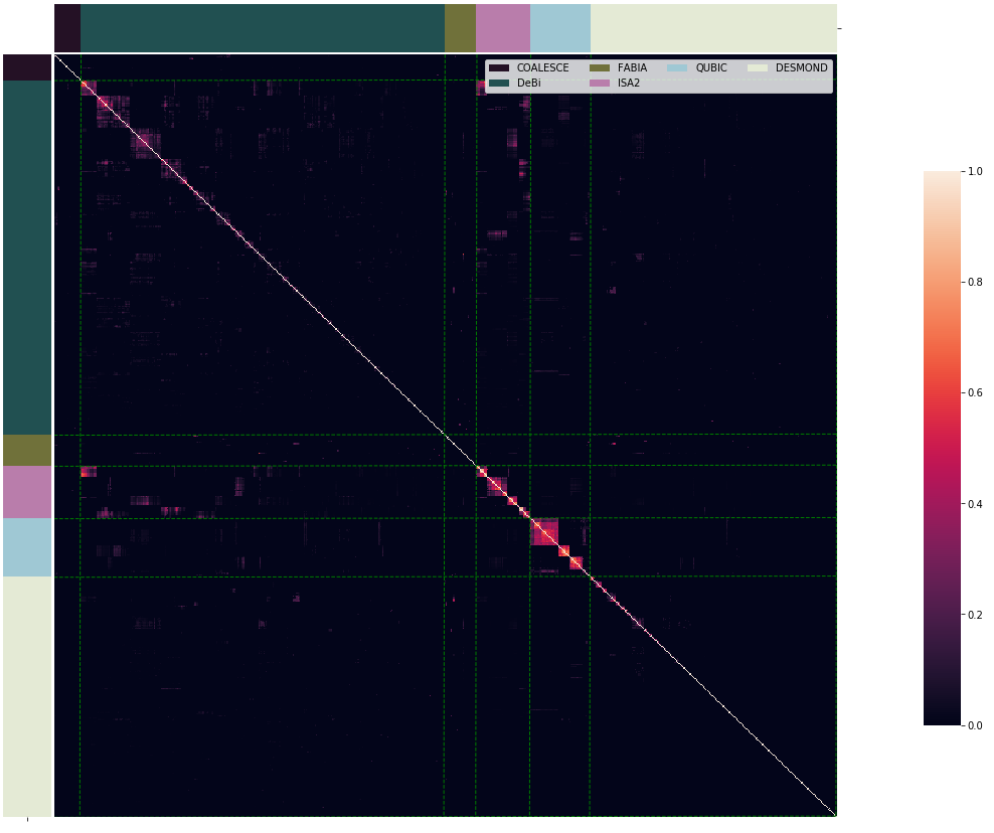


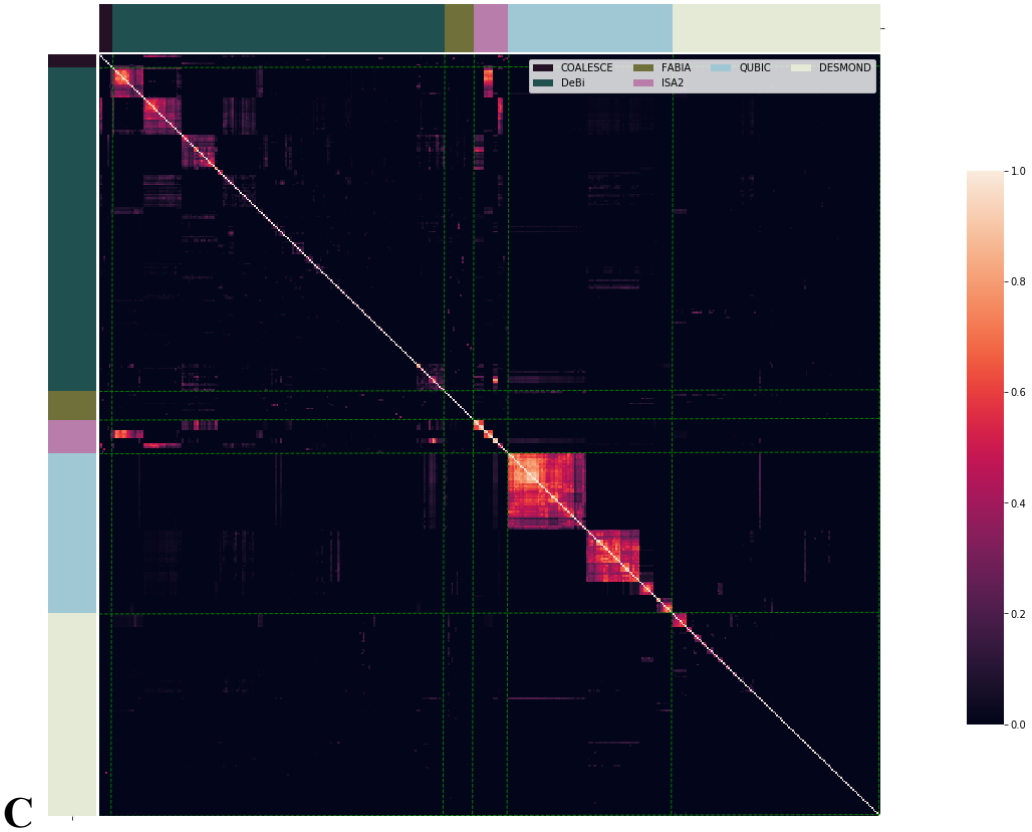
Figure S5. *Percentage of gene clusters significantly (BH-adjusted p-value<0.05) overlapping with at least one functionally related gene set from GO Biological Process (GOBP), GO Molecular Function (GOMF) and GO Cellular Component (GOCC). Results obtained with default (**A**) and tuned (**B**) parameters.*


COALESCE
DeBi
FABIA
ISA2
QUBIC
DESMOND
1.0
0.8
0.6
0.4
0.2
0.0

A


COALESCE
DeBi
FABIA
ISA2
QUBIC
DESMOND
1.0
0.8
0.6
0.4
0.2
0.0

B



C

Figure S6. *Similarities of functionally coherent gene sets determined by biclusters found by different methods in TCGA-micro (A), TCGA-RNAseq (B), METABRIC (C) datasets. For two gene sets similarity equals Jaccard index computed for these two sets if their overlap is significant (BH-adjusted p-value in hypergeometric test < 0.05), or 0 otherwise. For the convenience of visualisation, only randomly chosen one third of all biclusters is shown for DeBi and DESMOND. ISA2 and DeBi output many biclusters strongly overlapping in genes. The examples of alike gene sets found by other pairs of methods are rather isolated. Of all functionally coherent biclusters output by DESMOND, 18%, 21%, and 14% have zero similarity in genes with all biclusters found by other methods in TCGA-micro, TCGA-RNAseq, and METABRIC respectively.*

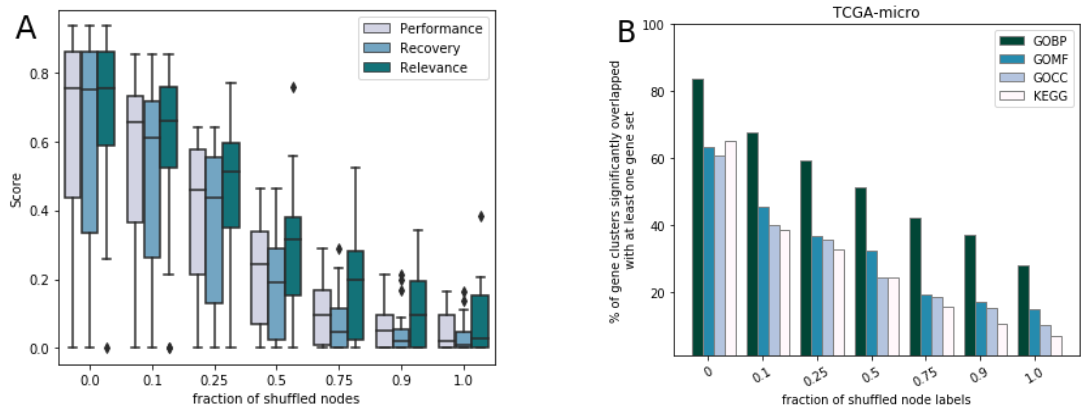


Figure S7. *The effect of the network randomization on the results. A. Performance, Recovery, and Relevance scores computed on 20 synthetic datasets, averaged in 10 runs. B. The percentage of biclusters overrepresented by GO terms and KEGG pathways on TCGA-micro.*

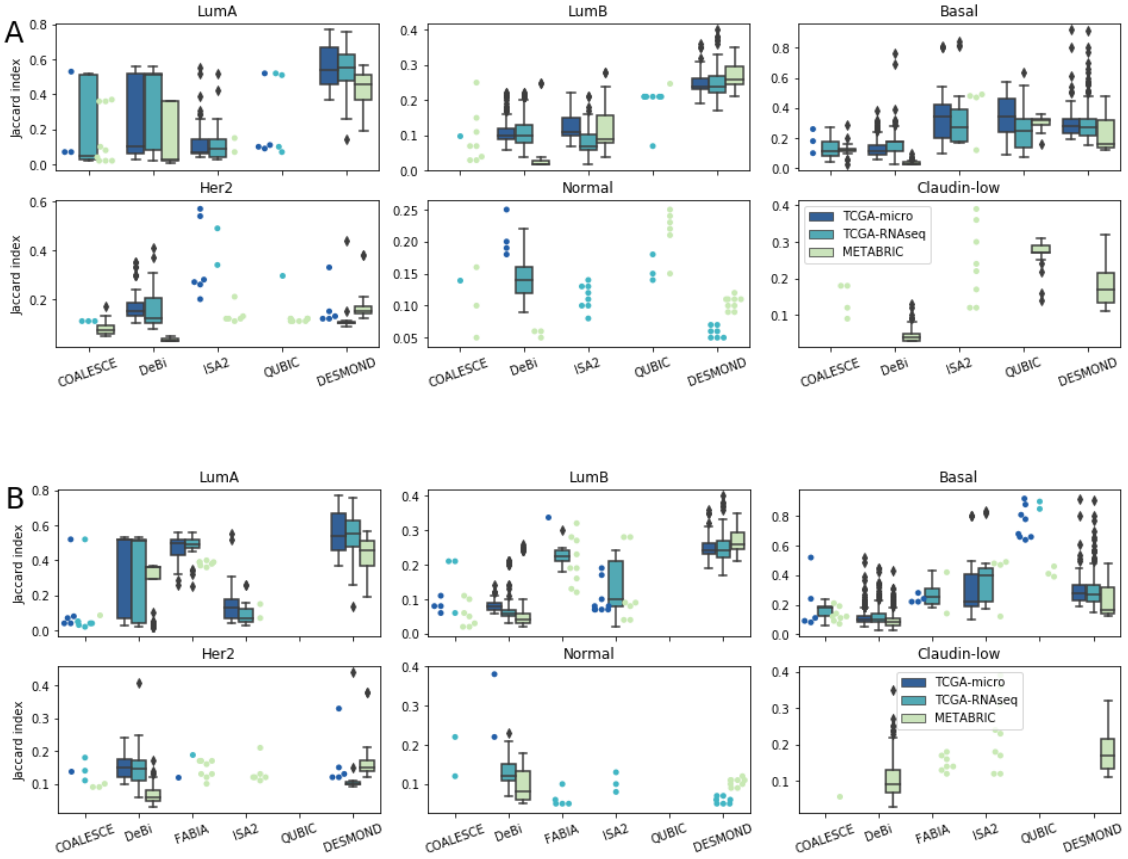


Figure S8. *Distributions of Jaccard similarities of known breast cancer subtypes and sample sets defined by biclusters produced by each method. For each bicluster, over- and under-representation of each subtype was evaluated using the hypergeometric test. Each bicluster was annotated with the subtype based on a minimal adjusted p-value passing threshold of 0.05. The results obtained with default parameters and with parameters optimized on synthetic data are shown in figures A and B respectively. When the group contains less than 10 biclusters, the results are shown as dots instead of a boxplot. Claudin-low subtype was annotated only in METABRIC dataset and therefore biclusters found in TCGA data sets were not tested for overlap with this subtype. All methods except QUBIC with optimized parameters found biclusters overlapping with Jaccard similarities about 0.5 and higher with Luminal A (LumA) subtype in TCGA. ISA2 and DESMOND found biclusters strongly overlapping with Basal subtype in TCGA datasets (Jaccard similarity above 0.8). ISA2 applied with default but not with optimized parameters identified biclusters strongly (Jaccard about 0.5) overlapping with Her-2 subtype in TCGA. For all other subtypes in TCGA and all subtypes in METABRIC, overlaps with the most significantly enriched biclusters were even weaker. DESMOND managed to find biclusters with stronger overlaps with LumA, LumB and Basal subtypes in TCGA than its competitors.*

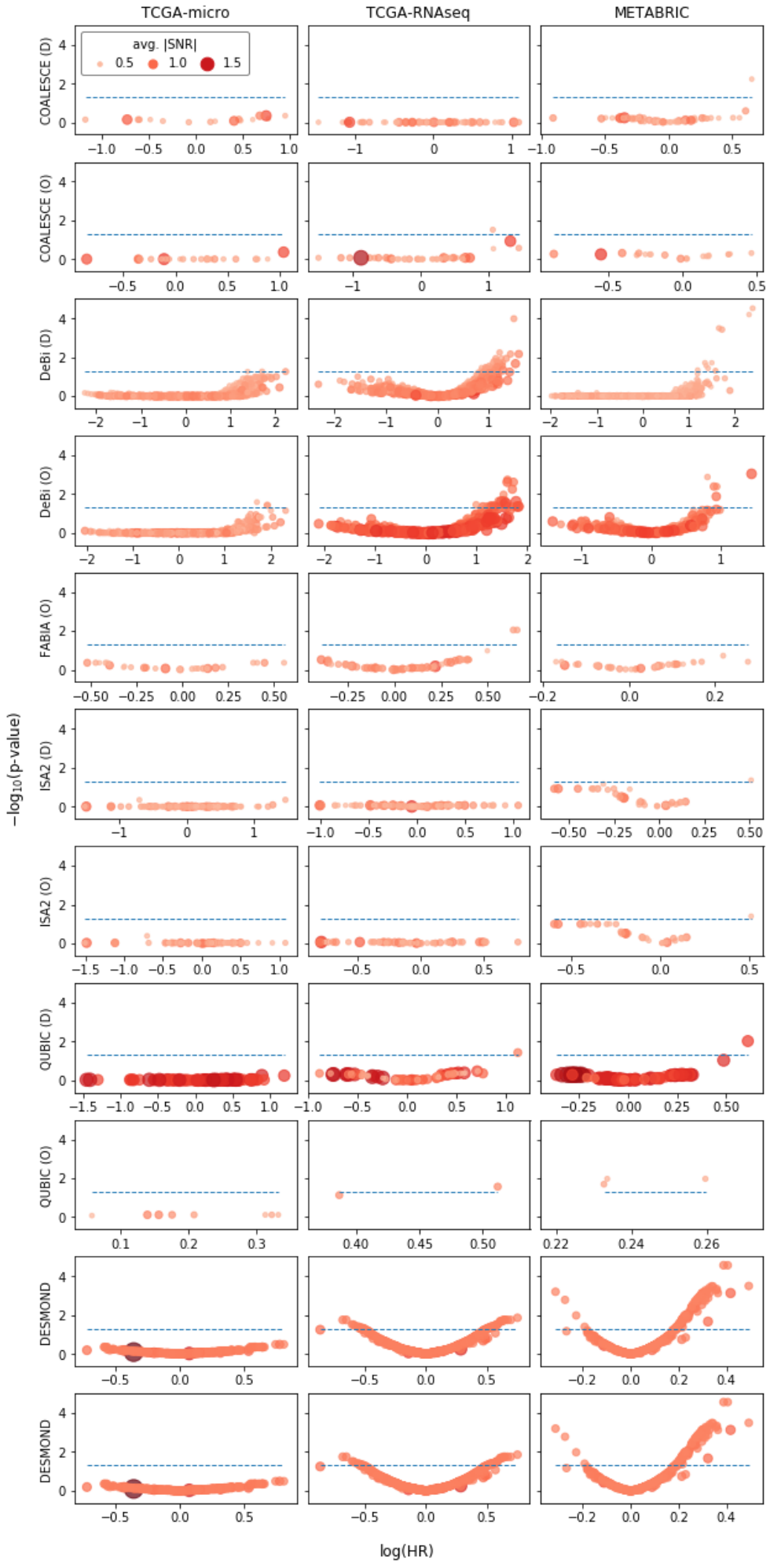

TCGA-micro
TCGA-RNAseq
METABRIC
avg. |SNR|
0.5
1.0
1.5
COALESCE (D)
COALESCE (O)
DeBi (D)
DeBi (O)
FABIA (O)
ISA2 (D)
ISA2 (O)
QUBIC (D)
QUBIC (O)
DESMOND
DESMOND
−log10(p-value)
log(HR)

Figure S9. *Association of biclusters found by DeBi and DESMOND with overall survival. Every circle represents a bicluster, with size and color intensity proportional to avg.|SNR|. The X and Y axes show a negative logarithm of adjusted p-values and coefficients (logarithm of Hazard Ratio) of Cox regression models fitted for patient sets defined by biclusters. The best biomarkers have higher average SNR and larger positive or negative regression coefficients.*

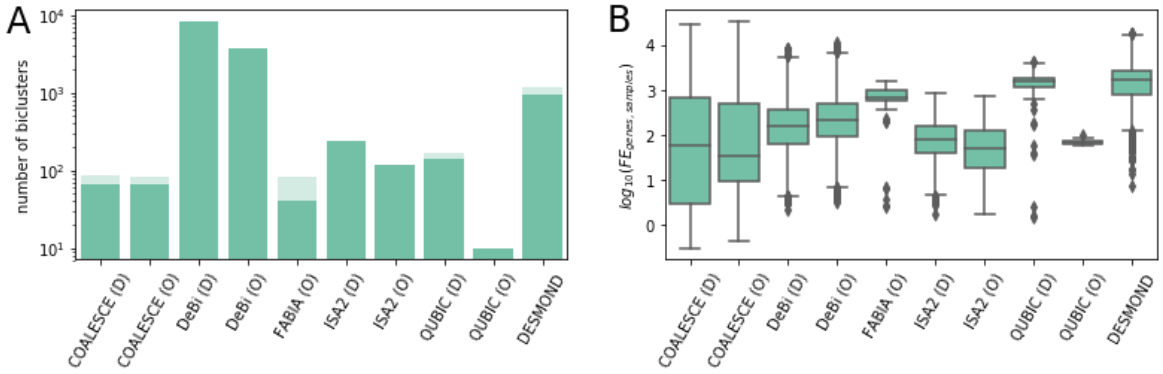


Figure S10. Similarities of biclusters found in TCGA-BRCA datasets profiled by RNA-seq and microarrays computed considering genes and shared samples. (A) Total number of biclusters tested. Transparent parts of the bars show a total number of biclusters including those which had no best match. (B) Distributions of log-transformed fold-enrichments of Jaccard similarities computed for best matches.

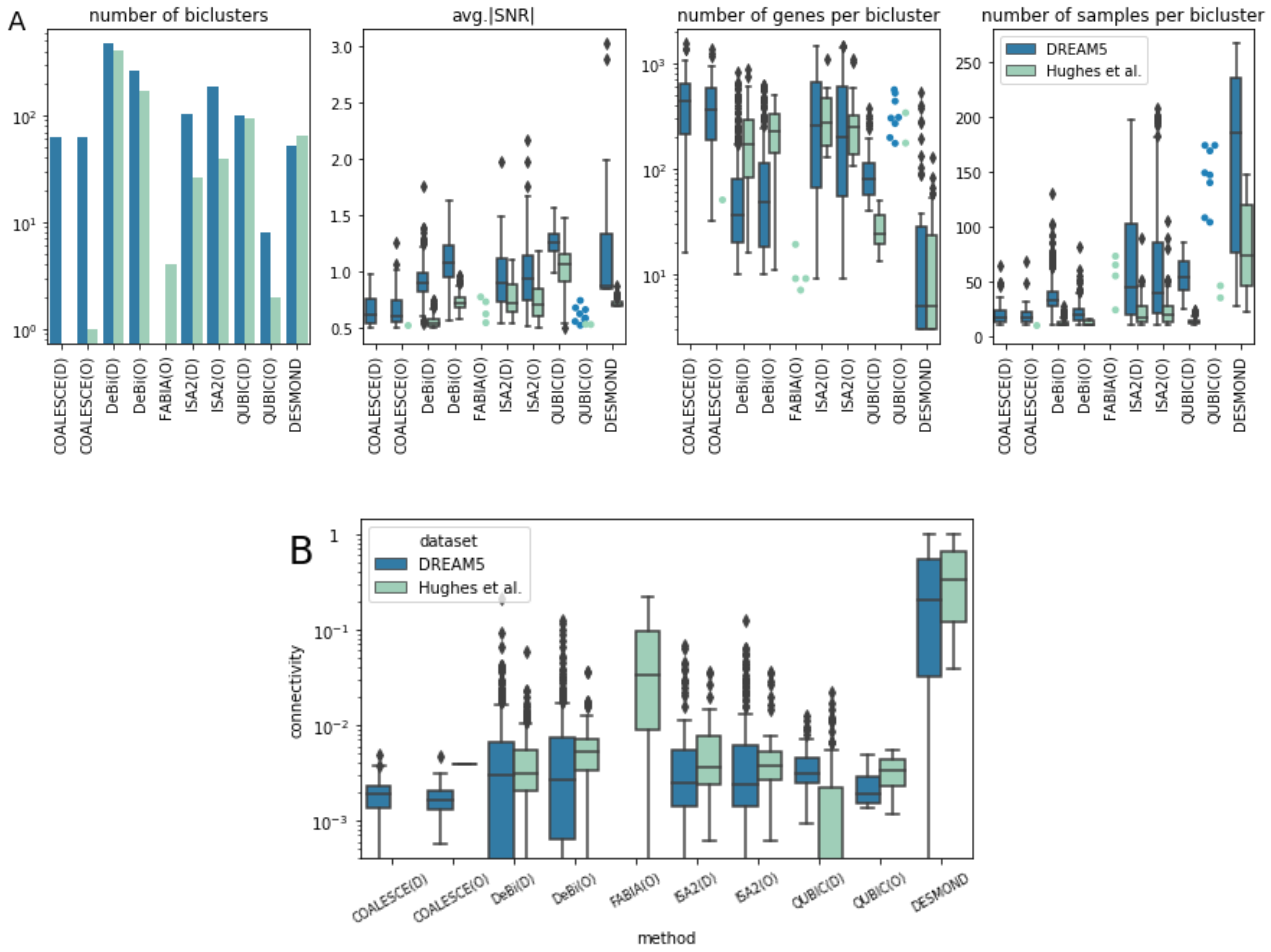


Figure S11. *Characteristics of biclusters found by DESMOND and baseline methods in two yeast datasets with default (D) and optimized (O) parameters.* ***A.*** *Total number, average absolute SNR, and sizes in genes and samples.* ***B.*** *Connectivity of found biclusters. FABIA(D) outputs no biclusters on both datasets and therefore is not shown.*

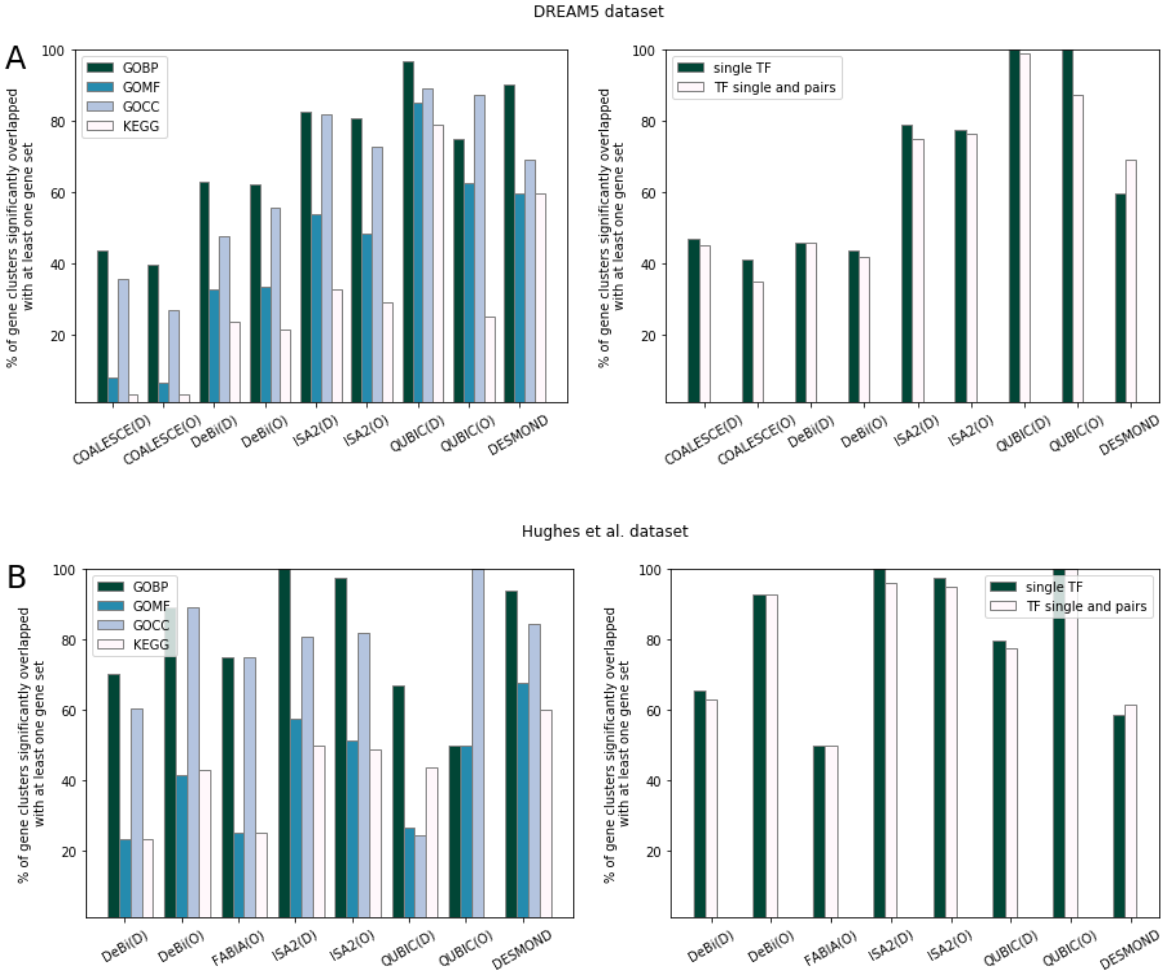


Figure S12. *The percentage of biologically significant biclusters found on DREAM5 (A) and Hughes et al. datasets (B). D and O indicate whether the method was applied with default or optimized parameters. Only methods that produced at least one enriched bicluster are shown. Left panels show the percentages of functionally coherent biclusters produced by each method. right panels demonstrate the percentages of biclusters significantly overlapping by sets of genes regulated by the same single TF or by the same pair of TFs.*