spongEffects: ceRNA modules offer patient-specific insights into the miRNA regulatory landscape

Fabio Boniolo[1,2,*], Markus Hoffmann[1,3,*], Norman Roggendorf[1], Bahar Tercan[4], Jan Baumbach[5,6], Mauro Castro[7], A. Gordon Robertson[8], Dieter Saur[2], and Markus List[1,**]

[1]Big Data in BioMedicine Group, Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Munich, Germany
[2]Chair of Translational Cancer Research and Institute of Experimental Cancer Therapy, Klinikum Rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany
[3]Institute for Advanced Study (Lichtenbergstrasse 2 a, D-85748 Garching, Germany), Technical University of Munich, Germany
[4] Institute for Systems Biology, Seattle, WA, USA
[5]Chair of Computational Systems Biology, University of Hamburg, Hamburg, Germany
[6]Computational BioMedicine Lab, University of Southern Denmark, Odense, Denmark
[7]Bioinformatics and Systems Biology Laboratory, Universidade Federal do Paraná, Curitiba, Brazil
[8]BC Cancer Genome Sciences Centre, Vancouver, Canada; Dxige Research Inc., Courtenay BC Canada

**The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.*
*** corresponding author: markus.list@tum.de*

ABSTRACT

Cancer is one of the leading causes of death worldwide. Despite significant improvements in prevention and treatment, mortality remains high for many cancer types. Hence, innovative methods that use molecular data to stratify patients and identify biomarkers are needed. Promising biomarkers can also be inferred from competing endogenous RNA (ceRNA) networks that capture the gene-miRNA-gene regulatory landscape. Thus far, the role of these biomarkers could only be studied globally but not in a sample-specific manner. To mitigate this, we introduce spongEffects, a novel method that infers subnetworks (or modules) from ceRNA networks and calculates patient- or sample-specific scores related to their regulatory activity. Notably, these module scores can be inferred from gene expression data alone and can thus be applied to cohorts where miRNA expression information is lacking. We show how spongEffects can be used for downstream machine learning tasks such as tumor classification and for identifying subtype-specific regulatory interactions. In a concrete example of breast cancer subtype classification, we prioritize modules impacting the biology of the different subtypes. In summary, spongEffects can perform classification tasks and prioritize ceRNA modules as biomarkers, hence offering insights into the miRNA regulatory landscape from gene expression data alone.

INTRODUCTION

Cancer is one of the leading causes of death worldwide, with an estimated 19.3 million new cases and almost 10 million deaths in 2020 [1]. Despite recent advances in screening, diagnosis, and prognosis, and an increased understanding of the mechanisms driving tumorigenesis, progression, and maintenance, cancer deaths are estimated to be ~400,000/year by 2040 in the US alone [2], highlighting the need for innovative approaches for the diagnosis, disease monitoring, and treatment.

The paradigm of precision medicine offers an alternative to the traditional "one-fits-all" approach by exploiting the wealth of molecular and clinical data to stratify patients into clinically relevant subgroups with shared molecular disease mechanisms and, hence, offering the potential for mechanism-based treatment strategies [3,4].

In addition to the traditional study of genome variation and gene expression, microRNAs (miRNAs) have been assessed as potential biomarkers, given their important roles in regulating gene expression both in physiological conditions and in cancer [5,6] and their stability in biological fluids [7]. Mature miRNAs are short non-coding RNAs with lengths of ~20-23 nucleotides that play an important role in regulating gene product abundance [8]. miRNAs are involved in regulating at least half the genes in the human genome [9] and are dysregulated in many diseases [10]. They function via the targeting of RNAs that will be degraded or whose translation will be impeded [11]. miRNA-target recognition is facilitated via matching the seed region at the 5' end of the miRNA with the 3' binding site of the target (canonical targeting) or via the action of additional regions outside of the seed that can contribute to the recognition of the target (non-canonical targeting) [12,13]. Salmena et al. proposed the competitive endogenous RNA (ceRNA) hypothesis [14], which suggests that RNAs with miRNA binding sites compete for a limited pool of miRNAs, giving rise to a complex gene regulatory network [15]. Past studies have identified genes that act as ceRNAs [14,16,17] and have associated them with different RNA classes, like messenger RNAs (mRNAs), circular RNAs, pseudogenes, transcripts of 3' untranslated regions (UTRs), and long non-coding RNAs (lncRNAs) [18]. In coding genes, effective miRNA targeting (i.e., reducing transcript levels) largely occurs in 3'UTRs [19]. lncRNAs have recently received particular attention in the framework of ceRNAs, suggesting the key role these molecules may play in binding miRNAs and indirectly regulating the expression of protein-coding genes sharing the same miRNA binding site region [20]. It is important to note that the key actors in these modules are not the miRNAs but rather the ceRNAs that exert influence over the expression of other ceRNAs using miRNAs as a limited resource.

While laboratory experiments designed to identify condition-specific miRNA-target interactions are usually labor- and cost-intensive, a range of computational methods has been developed to identify potential miRNA-target interactions based on different inputs, such as gene expression data, protein-protein interaction (PPI) networks, and sequence information [21]. Such methods often result in a large number of inferred interactions that, in return, suggest complex ceRNA networks that are typically difficult to explore. A key problem is to dissect these networks into functional units, or modules, that could play a role in specific tissues or diseases [22]. The identification of ceRNA modules could highlight the biological relevance of miRNA regulation in specific processes, while potentially identifying new prognostic biomarkers and therapeutic targets for clinical applications.

Multiple tools able to infer ceRNA modules have been designed [23]. While the ability to identify functional modules with biological relevance is challenging, several drawbacks hinder experimental validation and use in clinical settings. For example, module identification algorithms often result in a small number of modules that contain up to hundreds of genes [24] that are related to very general cancer-related hallmarks, e.g., EMT transition or cancer metabolism. Importantly, none of the ceRNA module identification methods proposed to date is able to summarize the information content of the modules in a patient- or sample-specific way. A sample-specific measure of ceRNA regulatory activity would be useful for the interpretation of its biological function and make it accessible for downstream analysis tasks such as clustering and classification. Similarly, previous work

focused on the design of methods able to infer personalized ceRNA networks [20,25]. Specifically, the statistical association of known ceRNA pairs was evaluated where deviations from the overall pattern in a specific sample are scored. Thus, a sample-specific ceRNA network retains those edges that deviate from the norm. Instead of focusing on these individual aberrant edges, we are more interested in quantifying the overall activity of ceRNAs in a sample. To this end, we seek to divide the ceRNA network into modules that reflect the activity of important ceRNAs driving tumor biology.

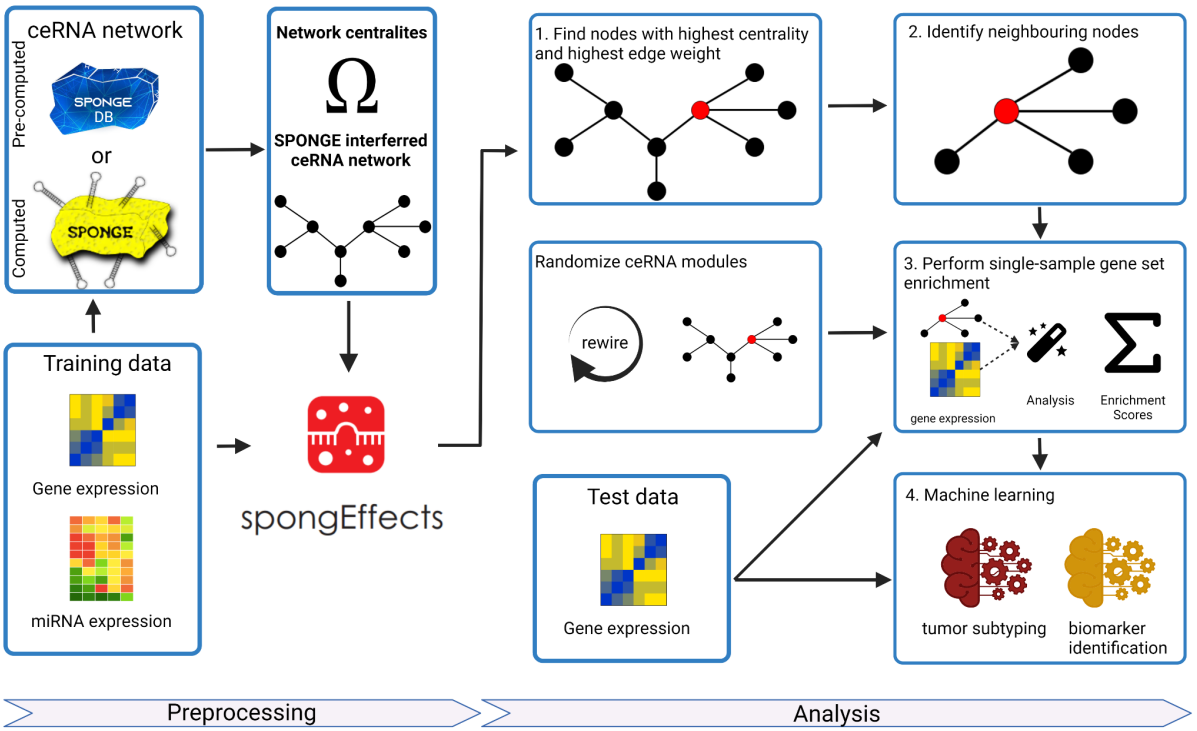

Figure 1: The workflow of spongEffects. SpongEffects accepts a gene expression matrix and a ceRNA network as input. Subsequently, it 1) filters the network and calculates weighted centrality scores to identify important nodes, 2) identifies first neighbors, 3) runs single-sample gene set enrichment, and 4) prepares the output for further downstream tasks (e.g., machine learning-based classification and extraction of mechanistic biomarkers).

Here, we introduce spongEffects, a new method designed to extract ceRNA modules from previously inferred ceRNA networks. A particular strength of spongEffects is that it can assess the regulatory activity of a module using enrichment scores as surrogates for module activity. We refer to these as spongEffects scores. We calculate such scores at the single patient- or sample level, thus allowing the study of ceRNA effects across groups of samples or patients or even in personalized analysis. The spongEffects method is based on our previous work, Sparse Partial correlation ON Gene Expression (SPONGE), a computational method that systematically infers ceRNA interactions from gene and miRNA expression data [26]. With SPONGE, we constructed 23 ceRNA networks (22 cancer types and a pan-cancer network), which we subsequently made publicly available as part of SPONGEdb, a pan-cancer resource for competing endogenous RNA interactions [27].

The general workflow of spongEffects is shown in Figure 1. Given gene expression data and a pre-computed ceRNA network (e.g., as generated by SPONGE [26]), spongEffects performs several steps: (1) it filters for significant interactions with meaningful effect size

and identifies the most important ceRNAs via network centrality analysis; (2) for the subset of nodes with the highest centrality values, it constructs ceRNA modules by incorporating first-degree ceRNA network neighbors; (3) it performs single-sample gene set enrichment using the identified nodes, thus obtaining spongEffect scores; and (4) it uses the spongEffects scores to perform downstream machine learning tasks for classification and biomarker identification.

We show that taking pre-computed ceRNA networks from SPONGEdb and breast cancer gene expression data from TCGA as inputs, spongEffects scores can be used to classify cancer subtypes with high accuracy. Using randomized input networks, we show that random modules do not have a comparable predictive value. Focusing on modules with high predictive power further allowed us to identify ceRNAs and miRNAs that may play important roles in cancer biology. In contrast to the ceRNA networks that are typically inferred from matched gene and miRNA expression data, spongEffects scores are inferred exclusively from gene expression data, making spongEffects applicable to data sets for which miRNA expression profiles are missing. As we show using independent data from the METABRIC consortium [28], the enrichment methods underlying spongEffects are robust to missing values (i.e. genes not measured or annotated) and across different technologies (e.g., microarray). This is a particular advantage for studying long non-coding RNAs which are often not included in gene expression data due to technical limitations.

MATERIALS AND METHODS

Data processing

We used the pre-computed inferred ceRNA networks and the network centralities data from SPONGEdb [27]. ceRNA networks were previously inferred for TCGA cancer types with > 100 samples, resulting in 22 cancer type-specific networks and one pan-cancer ceRNA network that we used here as input data for spongEffects. We downloaded log2(TPM + 0.001) transformed TCGA expression data (RNAseq) from the XENA Browser [29]. In addition, we downloaded log intensity levels (Illumina Human v3 microarray) gene expression data from the 1st and 2nd METABRIC (Molecular Taxonomy of Breast Cancer International Consortium) cohorts [28] from cBioPortal [30] and used them as external datasets to validate our findings.

Implementation

spongEffects was implemented in R (version 3.6.2) and is provided to the community as a new function in the SPONGE package in Bioconductor (see Availability Section).

Defining ceRNA modules

Centrality measures provide essential information about the organization of complex systems in network analysis [31,32]. Classical centrality measures such as those first formalized by Freeman et al. [33], i.e., degree, closeness, and betweenness, can extract important information in biological networks by identifying hub and bottleneck nodes [34–36]. The definitions of such centrality measures have been generalized for applications in weighted networks [37–39] to acknowledge that not all interactions are equally important. In particular, the definition of degree centrality has been shifted from the sum of edges of a node to the sum of the edge weights (node strength). These can be formalized as

$$\text{Equation 1: } Degree_i = \sum_{j}^{N} x_{ij}$$

$$\text{Equation 2: } Strength_i = \sum_{j}^{N} w_{ij}$$

$$\text{Equation 3: } Centrality_i^{\alpha} = Degree_i \, x \left(\frac{Strength_i}{Degree_i}\right)^{\alpha} = (Degree_i)^{1-\alpha} x \, Strength_i^{\alpha}$$

Equation 1 describes the definition of the degree of centrality of node *i* in an undirected network of size N. x_{ij} is the binary adjacency matrix describing the connection of nodes *i* and *j* (i.e., x_{ij}= 1 if *i* and *j* are connected in the graph and 0 otherwise).

Equation 2 formalizes the weighted degree centrality measure, or node strength of node *i* in a network of size N. In particular, w_{ij}>0 if nodes *i* and *j* are connected and it corresponds to the weight of the edge between *i* and *j*. Centrality is defined in Equation 3. Opsahl et al., describes α as "a positive tuning parameter that can be set according to the research setting and data. If this parameter is between 0 and 1, then having a high degree is taken as favorable, whereas if it is set above 1, a low degree is favorable" [40].

In this work, we exploit the definition of weighted degree centrality offered in [40] as implemented in the R package (version 3.0.16) *tnet* (Equation 3) [41] with the alpha parameter set to 1. In the context of a ceRNA network, we thus do not consider the number of interactions of the nodes (ceRNAs) but consider nodes as important that have a large sum of weights (using multiple sensitivity correlation (mscor) as effect size).

We define ceRNA modules as subnetworks consisting of ceRNA genes with highest weighted centrality scores (central node) and their first-degree neighbors, i.e. nodes directly connected to the central (ceRNA) node.

spongEffects scores

We implemented three different gene set enrichment approaches to aggregate the information associated with genes belonging to sponge modules into unique scores, which we call spongEffects scores. These methods belong to the class of unsupervised and single-sample assessment tools [42], in that they do not require initial knowledge of existing phenotype groups, which is often unavailable, and they yield patient-specific scores for each module, which is here interpreted as a gene set. They include single sample Gene Set Enrichment Analysis (ssGSEA) and Gene Set Variation analysis (GSVA) algorithms as implemented in the *GSVA* package (version 1.34.0) [42], and Overall Expression (OE) [43].

Importantly, these approaches allow the calculation of the spongEffects even when some of the genes in the modules are not present in the input expression matrix, and result in module-by-sample spongEffects matrices that can be used for downstream analyses. When comparing results between GSVA, ssGSEA, and OE (See Supplementary Figure 1) we did not observe any major differences in the results. Nevertheless, the original GSVA publication highlights that the choice of the optimal single-sample enrichment tool is highly dependent on the specific task and dataset and cannot be defined a priori.

Random Forest for subtype classification

Random forests are ensemble tree-based models that classify samples via majority voting [44]. In this work, we used Random Forest as implemented in the *caret* R package (version 6.0.90) [45] to classify tumor samples in the respective subtypes, using spongEffects scores as input. The optimal number of features (i.e., spongEffects modules) was randomly sampled as each split was defined via grid search in a repeated (3x) 5-fold cross-validation procedure. We calculated the Gini index to identify the modules that drove the predictions as implemented in the *randomForest* package (version 4.6.14) [46].

Randomization of the ceRNA networks

To validate our findings and evaluate whether the previously defined modules are robust, we sampled random modules of equal size (i.e. equal number of genes) using randomly chosen nodes (ceRNAs) from the ceRNA network. For each original module, we sample a random module of the same size such that the size distribution of the random modules matches the real modules exactly. Subsequently, we trained a classifier as previously described.

RESULTS AND DISCUSSION

SPONGE modules predict breast cancer subtypes

Breast cancer is the most commonly diagnosed cancer type worldwide (11.7% of newly diagnosed cases) and the leading cause of cancer-related deaths in women [1]. Its etiology is complex, involving the dysregulation and disruption of multiple processes at the genetic and epigenetic levels [47]. Such complexity results in a heterogeneous disease characterized by subgroups with different genetic prognostic profiles. The previous gene expression-based subtyping efforts identified five breast cancer subtypes (Luminal A, Luminal B, Basal, HER2-positive, and Normal) [48] that have been associated with distinct genetic drivers and patterns of variation at multiple levels, and, most importantly, present significant differences in aggressiveness and survival [49] that influence treatment choice [7].

Mechanisms of miRNA dysregulation are known to play a role in breast cancer [50], thus offering the chance to validate miRNA mature strands as potential biomarkers [51]. Here, we exploit our newly designed method, spongEffects, to study differences in miRNA regulation at the patient-specific level by exploiting publicly available breast cancer datasets from The Cancer Genome Consortium [49] (TCGA-BRCA, number of samples (n) = 1,063) and the Molecular Taxonomy of Breast Cancer International Consortium [28] (METABRIC cohorts 1 and 2, n = 1,905). In both cohorts, we selected all patients with tumor stages I, II, III, and IV, and removed all samples for which the subtype was missing or not identified as one of the five classes of interest (LumA, LumB, Her2, Basal, Normal-like). This step resulted in two final datasets: (1) TCGA-BRCA (n = 944), which we used for model training, and (2) METABRIC (n = 1,699), which we used as an external validation set.

SPONGE modules were calculated based on the TCGA-BRCA ceRNA network publicly available in SPONGEdb (http://sponge.biomedical-big-data.de/) [27]. The original network, containing ~$3x10^7$ edges, was filtered (mscor > 0.1, adjusted p-value < 0.01) down to 702,026 edges to preserve significant associations with non-negligible effect sizes (for further information see List et al. [26]). Weighted centrality measures were calculated for each node

in the network (See Material and Methods). For the purpose of this analysis, we chose to focus on lncRNAs as central nodes in the sponge modules, motivated by the importance these molecules have in ceRNA networks, their potential as biomarkers [20], and their relevance in breast cancer subtypes [52]. The 750 lncRNAs with the highest weighted centrality measures were selected and used as central nodes to define the sponge modules.

Once the modules were defined, we applied a filtering step to preserve all modules containing 10 and 200 genes after removing genes missing in the expression matrices, as previously suggested [42].

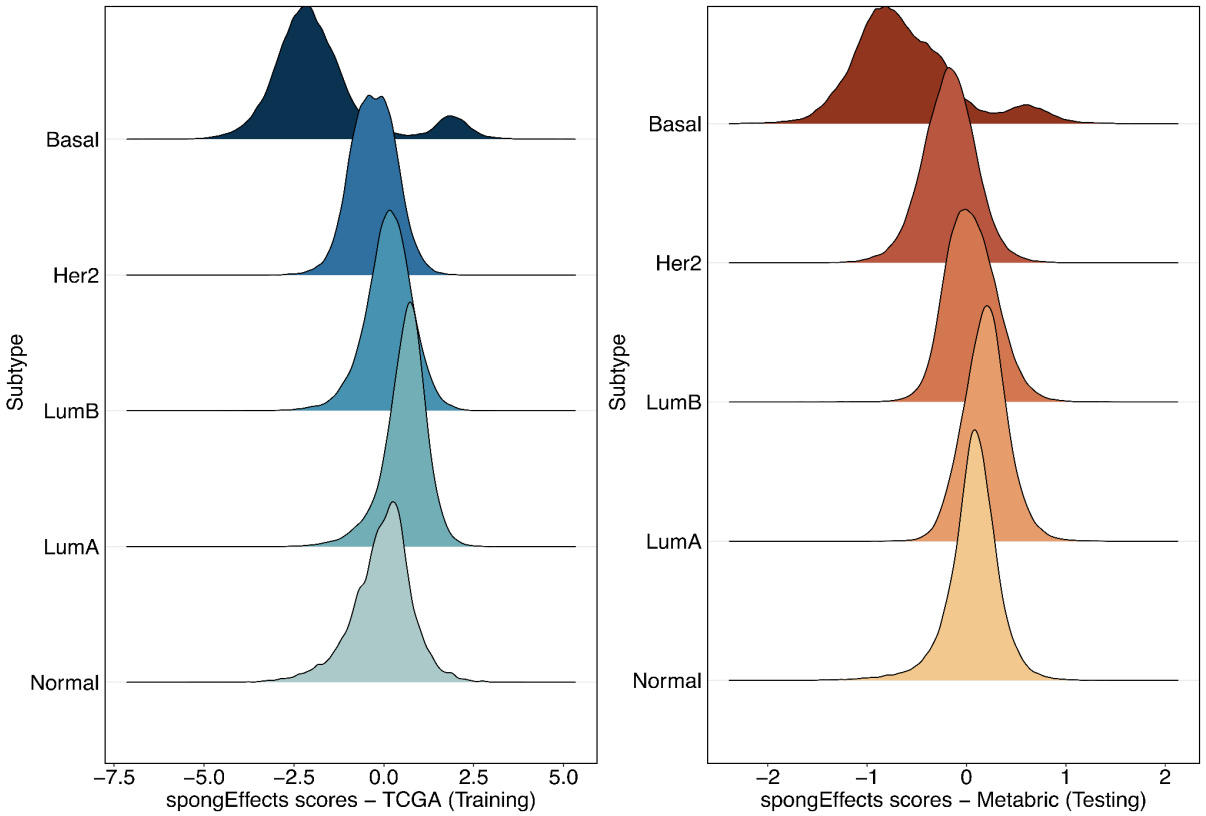

Figure 2: Distribution of the spongEffects scores in the TCGA-BRCA (training) and METABRIC (testing) dataset. With the exception of Basal breast cancer, the scores for the various breast cancer subtypes follow a normal distribution. For the Basal subtype, the bimodality of the distribution hints at the existence of a subgroup of patients with characteristics different from the rest of the class (see Supplementary Figure 2).

Enrichment scores of the ceRNA modules, referred to as spongEffects scores, were calculated for the two datasets separately (see Materials and Methods for details). To define the best single-sample enrichment tool in this setting, we compared the performances in training and testing on models calibrated on spongEffects scores calculated via OE, GSVA, and ssGSEA. The three different approaches resulted in models with similar subtype classification performances (Supplementary Figure 1), showing the robustness of the ceRNA modules. Here, we show the results for spongEffects scores calculated via OE, given its previous application in similar bulk transcriptional profiles [43].

As described in the original publication [43], the divergence from a normal-like distribution of enrichment scores calculated with this approach might hint at different subgroups of patients within the same class that can be modeled as a mixture of Gaussians. spongEffects is able to capture such heterogeneity, as shown by the bimodality of the Basal samples, where a subgroup of patients, identified via the R package *mixtools* (version 1.2.0) *[53]*, is over-enriched in stroma-high tumors (Figure 2 and Supplementary Figure 2.a), as calculated via ESTIMATE [54], and enriched in extracellular matrix (ECM)-related genes, which have been associated with Basal invasion programs under the regulation of KRT14 related genes [55] (Supplementary Figure 2.b). Analogous studies have identified similar Basal subgroups from different data types and cohorts [56], laying the ground for the identification of clinically relevant targets beyond the typical subtypes. This highlights the ability of spongEffects to identify a subset of samples with potential prognostic value.

We trained a random forest classifier on the TCGA-BRCA dataset and implemented a repeated 5-fold cross-validation to tune the optimal number of variables randomly sampled at each split (*mtry*). We then validated the resulting model on the METABRIC dataset.

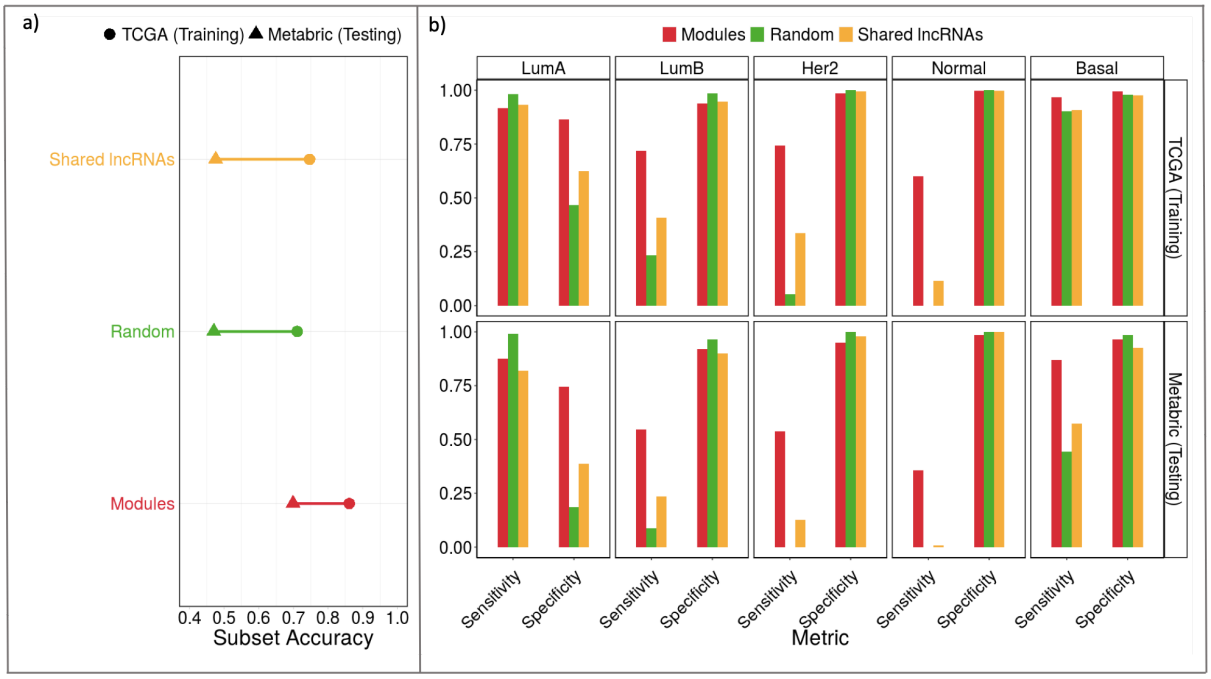

Figure 3: Overview of performance metrics for the random forest models calibrated on SPONGE modules (red), randomly defined modules (green), and lncRNAs common to both the TCGA (training) and METABRIC (testing) datasets (yellow). a) Subset accuracy (the proportion of samples that have all their labels classified correctly) values in training and testing for the three models. b) Sensitivity and specificity for the three models across breast cancer subtypes.

We repeated this analysis using different numbers of lncRNAs (200, 250, 500, 750, 1000, 1500, and 2000) to study the robustness of our approach and to highlight the relatively small

influence of the number of initial lncRNAs (and thus modules) considered on the final results (Supplementary Figure 3).

We compared our results to a control model trained on randomly defined modules (see Methods for details) and one baseline model trained on the expression values (not on spongEffects scores) of the 25 lncRNAs common to both datasets to verify the meaningfulness of the modules and the model itself. The subset accuracy metric (also known as exact match ratio, typically used in multiclass classification tasks [57]) was used to evaluate the performance of the models, together with standard single-class evaluation metrics such as sensitivity and specificity (confusion matrices in Supplementary Figure 4). The spongEffects-based classification model outperformed the random module-based and the lncRNA-based models both in training and testing (Figure 3.a) and preserved good specificity and sensitivity across all breast cancer subtypes (Figure 3.b).

Interpretation of spongEffects scores

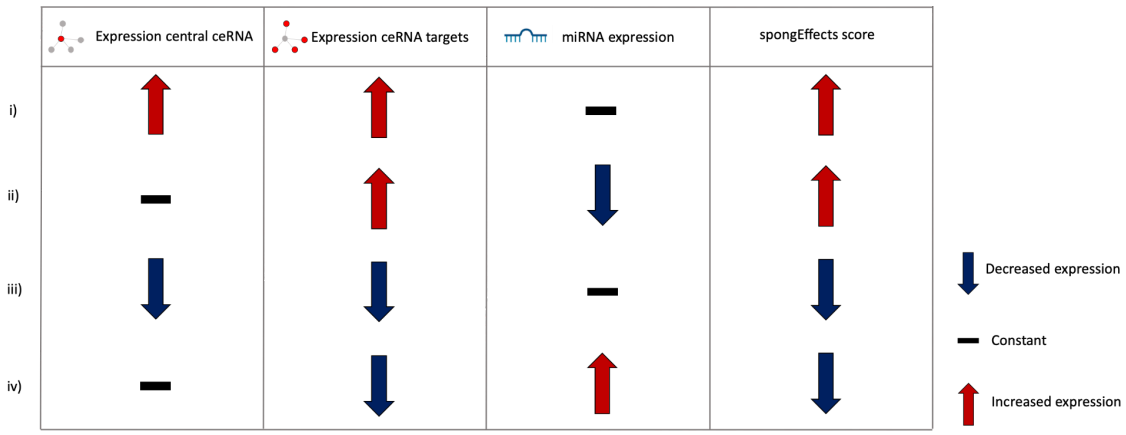

Figure 4: Understanding changes in spongEffects scores. Higher spongEffects scores can result from: i) an increase in the expression of a module's central ceRNA while miRNA expression levels remain constant, leading the to upregulation of the module's target ceRNAs, and/or ii) a decrease in the expression of a module's miRNAs that results in higher expression of the the module ceRNAs. Lower spongEffects scores can result from: iii) a decrease of expression in the module's central ceRNAwhile while miRNA expression levels remain constant, leading to the downregulation of the module's target ceRNAs, and/or iv) an increase of expression in a module's miRNAs that results in lower expression of the the module's target ceRNAs.

We expect spongEffects scores to reflect two overlapping regulatory mechanisms. The first is miRNA-mediated ceRNA-target regulation, where an increase in the expression of a ceRNA leads to miRNA depletion, relieving other target genes from miRNA repression and, conversely, a decrease in ceRNA expression leaves more miRNA copies that can act as repressors of their target genes (Figure 4 top). The second mechanism is the classical miRNA

regulation, where an increase in the expression of a miRNA leads to the repression of its target genes and *vice versa* (Figure 4 bottom). While it is possible to verify the former in specific settings, e.g., in the presence of miRNA expression data (as shown in two examples below), it is generally challenging to disentangle the two effects computationally. Single-sample enrichment tools estimate the enrichment or depletion of the ceRNAs part of the sponge modules with respect to the expression of all the other genes in the expression matrix. We view our method as a way to explain the calculated enrichment scores via the contributions of the regulatory mechanisms explained above.

ceRNA modules identify key biological mechanisms

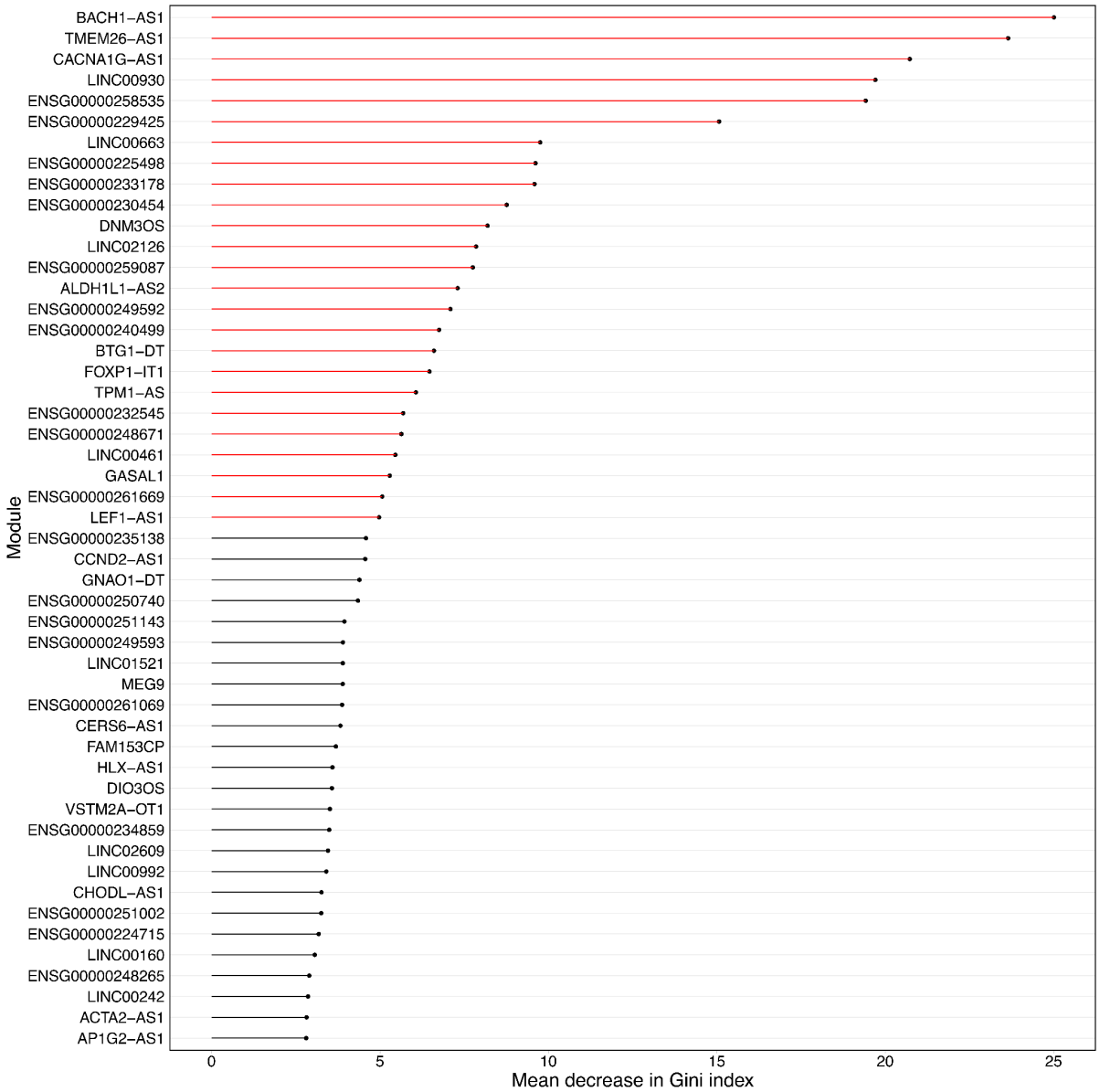

Figure 5: Most-predictive modules ranked by the Gini index. We selected the top 25 most predictive modules (red) for downstream analysis.

We identified the 25 most important ceRNA modules driving subtype classification by ordering the modules based on the mean decrease in the Gini index, a commonly used measure of feature importance in a random forest (Figure 5). Modules are designated by the name of the lncRNA central to the module.

Figure 6 shows differences in spongEffects scores for the top 25 ceRNA modules across breast cancer subtypes in both datasets. Interestingly, Basal samples showed low scores of low scores for modules centered around lncRNAs that were previously suggested to act as miRNA sponges, highlighting their potential as biomarkers in various cancer subtypes. For example, CACNA1G-AS1 was shown to decrease proliferation, EMT transition, and migration in hepatocellular carcinoma [58] and non-small cell lung cancer [59], and to downregulate p53 levels in colorectal cancer [60]. On the contrary, DNM3OS has been shown to promote tumor progression and SNAIL-mediated EMT in gastric cancers [61] and to confer resistance to radiotherapy by mediating DNA damage response mechanisms in esophageal squamous cell carcinoma. TPM1-AS has been shown to regulate alternative splicing of TPM1 and regulate TPM1-mediated migration of cancer cells in esophageal cancer [62]. Modules such as LINC00461 show strong enrichment in Basal samples. LINC00461 has been shown to promote migration and invasion in breast cancer [63,64] and to play a role in progression in other cancer types such as hepatocellular carcinoma [65] and renal cell carcinoma [66], or in mediating radiosensitivity in lung adenocarcinoma [67].

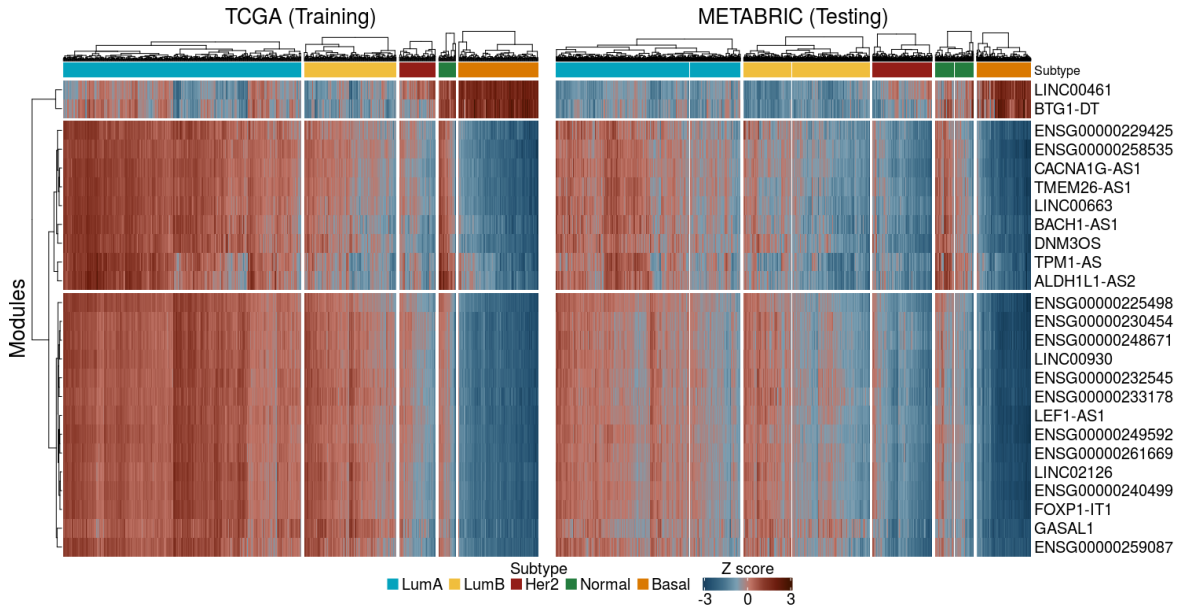

Figure 6: Heatmaps showing the standardized spongEffects scores for the top 25 most predictive modules calculated on the TCGA and METABRIC datasets. The heatmaps show a clear separation of the Basal subtype from the remaining ones, potentially pointing to the role of miRNA-regulated modules in the aggressive and proliferative phenotype observed in this subtype.

Predictive SPONGE modules are regulated by important miRNA families

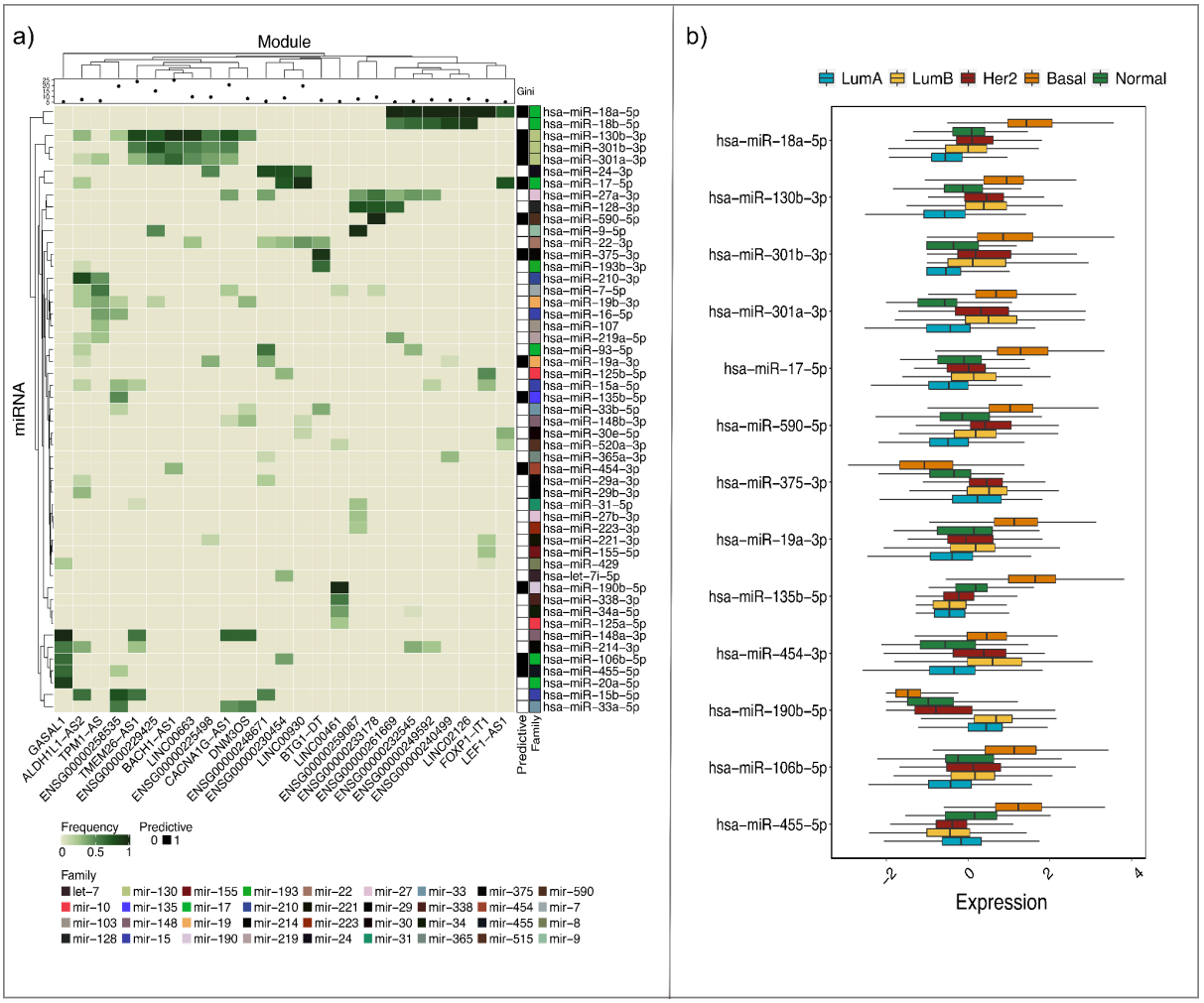

Figure 7: spongEffects scores can be partially explained by miRNA regulation. a) Fraction of a module's genes that are targeted by the miRNA in the top 25 ceRNA modules. Additionally, miRNA families are shown to indicate miRNAs with a shared seed sequence. We further indicate which of the miRNAs are predictive of breast cancer subtypes. b) Expression levels of the miRNA mature strands driving classification of breast cancer subtypes.

We were further interested in investigating the effect of miRNA regulation on the spongEffects scores of the 25 most-predictive modules. This was possible only for the TCGA-BRCA dataset, given the presence of patient-matched miRNA and gene expression data. We ranked the miRNAs based on how many times they were predicted by SPONGE to target the genes in the modules. The results, depicted in Figure 7.a as frequencies (the number of genes predicted to be targeted by a miRNA in a module divided by the size of the module) and further described below, highlight several miRNA families (characterized by shared seed sequences) that are common between modules. Importantly, 13 of the 51 miRNAs that were frequently associated with modules were also considered highly predictive in a subtype classification model based on miRNA expression data from the TCGA-BRCA dataset (Supplementary Figure 5) and showed clear differences in expression between subtypes (Figure 7.b). This highlights how our approach offers insights into miRNA and ceRNA

regulation even when miRNA expression data are unavailable. Focusing on the miRNA expression differences between subtypes offers a potential explanation for the spongEffects scores we have observed here.

a) miR-17~92-1 cluster, mir-106a~363 cluster

Gregorova et al. proposed that the miR-17~92-1 and mir-106a~363 clusters should be considered together since they are frequently upregulated in solid tumors and hematologic malignancies [68,69]. We can see a significant visual collection around hsa-miR-18a-5p. hsa-miR-18a-5p in the miR-17~92a-1 cluster that consists of hsa-miR-17, hsa-miR-18a-5p, among other miRNAs [68,69]. hsa-miR-18b-5p is part of the mir-106a~363 Cluster. hsa-miR-17 is involved in the ceRNA interactions in three of the 25 modules (ALDH1L1-AS2, ENSG00000230454, LINC00930). hsa-miR-18a-5p and hsa-miR-18b-5p are involved in an significant part of the ceRNA interactions in respectively seven and five out of the 25 modules (ENSG00000261669, ENSG00000232545, ENSG00000249592, ENSG00000240499, LINC02126, (FOXP1-IT1, LEF1-AS1)). hsa-miR-18a-5p and hsa-miR-18b-5p, among other miRNAs of the mir-17~92-1 and mir-106a~363 cluster, were identified as important tumor suppressors in pancreatic ductal adenocarcinoma [70,71]. Additionally, decreased levels of hsa-miR-18b-5p were detected in melanoma, inferring the p53 pathway [72] that is crucial in cancer biology [73].

b) miR-130b-3p/301-3p/454-3p family

One of the most significant visual collections in Figure 7 is around hsa-miR-130b-3p and hsa-miR-301b-3p. These miRNAs are as a whole or as a part involved in a large part of the ceRNA interactions in 10 of the 25 modules: ALDH1L1-AS2, TPM1-AS, ENSG00000258535, TMEM26-AS1, ENSG00000229425, BACH1-AS1, LINC00663, ENSG00000225498, CACNA1G-AS1, and DNM3OS. Gregorova et al. described the miR-130-3p/301-3p/454-3p family as significant in cancer biology [69]. hsa-miR-130b-3p was confirmed to directly target colony-stimulating factor 1 (CSF-1), downregulate CSF-1 expression, and result in decreased sensitivity to anticancer drugs [74]. hsa-miR-130b-3p belongs to the same 22q11.21 miRNA cluster as hsa-miR-301b-3p, which have the same seed sequence and thus share similar target genes [75]. Downregulation of this cluster was shown to be inversely correlated with cell proliferation [76], a cancer hallmark [77,78]. Furthermore, the cluster was described to play an oncogenic role by Fort et al. [79]. hsa-miR-454-3p, the last of this miRBase miRNA family, is not part of the significant collection in Figure 7. hsa-miR-454-3p is described as significant only in the BACH-AS1 module. Decreased levels of hsa-miR-454-3p are associated with less apoptotic cell death and increased cell proliferation [80,81].

c) miR-27a-3p

hsa-miR-27a-3p (19p13.12) is involved in ceRNA interactions in seven out of the 25 modules (ENSG00000259087, CACNA1G-AS1, ENSG00000248671, ENSG00000233178, ENSG00000261669, ENSG00000232545, ENSG00000249592). hsa-miR-27a-3p can function as both an oncogene or tumor-suppressor, depending on the cellular context [82].

d) miR-34a-5p

hsa-miR-34a-5p (1p36.22) is active in the LINC00461 module and strongly impacts the p53 pathway. hsa-miR-34a-5p inactivation may substitute for loss of p53 function in cancer biology [83].

e) miR-190b-5p

hsa-miR-190b-5p (1q21.3) is strongly active in the LINC00461 module. Xie et al. showed in a Kaplan-Meier survival analysis that hsa-miR-190b-5p is significantly correlated with survival in Hepatocellular Carcinoma [84]. Inline, Dai et al. suggested that hsa-miR-190b-5p is a potential biomarker for breast cancer overall survival [85]. Li et al. showed that high expression of hsa-miR-190b-5p is linked to a good prognosis, while low expression has been shown to drive the malignant progression of pancreatic cancer [86].

Predictive SPONGE modules can identify potential biomarkers in breast cancer

A deeper analysis of the ceRNA modules identified here could suggest important layers of regulation in breast cancer, the roles of ceRNAs and miRNAs in the disease, and could identify lncRNAs as potential biomarkers or therapeutic targets. As described previously, we identified two possible factors contributing (independently or jointly) to the spongEffects scores: ceRNA regulation via sponging and miRNA regulation. Due to CACNA1G-AS1's known roles in different cancer types (see above), we describe the CACNA1G-AS1 module and miRNAs that have been experimentally validated in more detail. Specifically, we investigated its role in the Basal subtype, one of the most aggressive breast cancer subtypes for which effective targeted therapies are still missing [87,88].

Genes in module CACNA1G-AS1 (Supplementary Table 1) have lower expression levels in the Basal subtype when compared to the other subtypes in TCGA-BRCA and METABRIC (Figures 7.a and 7.b) and contribute to the biology of this subtype. For example, expression of TBC1D9 has been shown to be inversely correlated to proliferation and grading in Basal breast cancer [89], while genes such as MYB and ZBTB16 are tumor suppressors whose promoter regions are often hypermethylated in Basal breast cancers [90,91].

In order to check whether we could exploit our modules to explain such changes, we further investigated the most frequent miRNAs in the CACNA1G-AS1 module (Figure 7) and checked their expression in the TCGA-BRCA dataset. Interestingly, miR-301b-3p and miR-130b-3p have higher expression in the Basal subtypes, giving a potential explanation for the spongEffects score of the genes they target (Figure 8. c).

Figure 8: Expression of genes in the CACNA1G-AS1 and LINC00461 modules and miRNAs predicted to target them. These genes have been experimentally validated to play a role in Basal breast cancers. a) Three genes of interest in the CACNA1G-AS1 module and their shared miRNAs. b) Log2-transformed and normalized expression values of the three genes of interest from the CACNA1G-AS1 module, TBC1D9, ZBTB16, and MYB, stratified by subtype. c) Log2-transformed and normalized expression values of the three miRNAs targeting the genes in panel b, miR-301a-3p, miR-130b-3p, and miR-301b-3p, stratified by subtype. d) Three genes of interest in the LINC00461 module and of the shared miRNAs. e) Log2-transformed and normalized expression values of the four genes of interest from the LINC00461 module, IGF2BP2, CDK6, RARRES1, and BCL11A, stratified by subtype. f) Log2-transformed and normalized expression values of the miR-190b-5p, which targets the genes in panel b, stratified by subtype.

The same approach can be applied to module LINC00461 (see Supplementary Table 1), where genes belonging to these modules are known to be upregulated in Basal cancers and contribute to its highly proliferative and aggressive phenotype (Figures 8.d and 8.e). For example, CRYAB has been associated with poor survival of Basal cancers [92,93] and is

hypothesized to be a potential therapeutic target [94]. Other genes in the modules, e.g., RARRES1 [95], BCL11A [96], IGF2BP2 [97], and CDK6 [98], have been shown to contribute to the overall shorter survival in Basal breast cancer. Genes in this module are regulated by miRNA miR-190b-5p (1q21.3), which shows lower expression in Basal samples when compared to the remaining subtypes (Figure 8.f).

CONCLUSION AND OUTLOOK

We present spongEffects, a new method for extracting ceRNA modules from pre-computed ceRNA networks. To the best of our knowledge, spongEffects is the first method that offers a sample- or patient-specific score that can quantify the regulatory activity of a ceRNA and its associated miRNAs. Using SPONGE-inferred ceRNA networks and RNA-seq data for TCGA breast cancer as inputs, we demonstrate how spongEffects scores can capture important aspects of cancer biology and be used in downstream machine learning tasks such as subtype prediction and biomarker identification. Using independent microarray-based gene expression data from the METABRIC consortium, we show that the inferred modules generalize to different cohorts. A particular strength of our approach is that spongEffects modules, once inferred from a ceRNA network, can also be computed for datasets or cohorts that lack miRNA data. For the modules that were predictive in subtype classification, spongEffects scores capture contributions of two different regulatory mechanisms: ceRNA regulation and miRNA regulation (see Interpretation of spongEffects scores). While it is challenging to separate the two, we observe that, in practice, our method returns a robust representation of microRNA-mediated regulatory effects.

Many gene regulatory mechanisms of lncRNAs are poorly understood and need to be experimentally discovered [99]. The spongEffects modules could help prioritize lncRNAs that have an expected impact on cancer biology and miRNA regulation, informing experimental validation and unraveling currently uncharacterized gene regulatory mechanisms of lncRNAs in combination with miRNAs. Of particular interest here is to address the question of lncRNAs' modes of action. For instance, it is currently debated if lncRNAs are exported outside of the nucleus, which would be a requirement for the established Argonaute-dependent mechanism of miRNA regulation [13]. lncRNAs with an experimentally verified role as ceRNAs could then serve as important biomarkers and potential therapeutic drug targets. We note that while we have applied spongEffects here to ceRNA networks inferred with SPONGE, our method is generally applicable to arbitrary ceRNA networks and potentially also to gene-regulatory networks, where similar concepts have already been explored [100,101]. We foresee two interesting directions for further research. First, recent studies on the inference of regulons [100,102] offer the chance to integrate methods for the combined analysis of transcription factors and miRNA regulation. Second, the surge of single-cell sequencing technologies allows the study of regulatory mechanisms at a higher resolution than what was possible before [103–105], paving the way to the design of methods with the potential of uncovering new insights into the complexity of tumor biology.

We have calculated spongEffects scores for samples from 22 different cancer types from TCGA and made the scores publicly available so that the community can use them to infer ceRNA and miRNA regulatory activities and gain insights into the gene regulatory landscape where the cohort size is too small for robustly inferring ceRNA networks or where supporting miRNA expression data is missing.

DATA AVAILABILITY

spongEffects source code is available at
https://github.com/biomedbigdata/SPONGE/tree/spongEffects

SPONGEdb is available at
http://sponge.biomedical-big-data.de/

We made spongEffects scores calculated for TCGA datasets available to the community at: https://doi.org/10.6084/m9.figshare.19328885.v1

SUPPLEMENTARY DATA

Supplementary Data is available at biorxiv online.

FUNDING

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 777111. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains. This work was further supported by the German Federal Ministry of Education and Research (BMBF) within the framework of the e:Med research and funding concept (grant 01ZX1908A).

CONFLICT OF INTEREST DISCLOSURE

The authors have no conflicts of interest to declare.

ACKNOWLEDGEMENTS

We want to thank Christina Trummer for designing the spongEffects logo.
The results published and shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga
Images were created with https://www.biorender.com. Parts of images were designed using resources from Flaticon.com.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: A Cancer Journal for Clinicians. 2021. pp. 209–249. doi:10.3322/caac.21660

2. Rahib L, Wehner MR, Matrisian LM, Nead KT. Estimated Projection of US Cancer Incidence and Death to 2040. JAMA Netw Open. 2021;4: e214708.

3. Reardon B, Moore ND, Moore NS, Kofman E, AlDubayan SH, Cheung ATM, et al. Integrating molecular profiles into clinical frameworks through the Molecular Oncology Almanac to prospectively guide precision oncology. Nature Cancer. 2021;2: 1102–1112.

4. Nogales C, Mamdouh ZM, List M, Kiel C, Casas AI, Schmidt HHH. Network pharmacology: curing causal mechanisms instead of treating symptoms. Trends Pharmacol Sci. 2021;0. doi:10.1016/j.tips.2021.11.004

5. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016;1: 15004.

6. Liu H, Lei C, He Q, Pan Z, Xiao D, Tao Y. Nuclear functions of mammalian MicroRNAs in gene regulation, immunity and cancer. Mol Cancer. 2018;17: 64.

7. Fernandez-Rozadilla C, Simões AR, Lleonart ME, Carnero A, Carracedo Á. Tumor Profiling at the Service of Cancer Therapy. Front Oncol. 2020;10: 595613.

8. Kartha RV, Subramanian S. Competing endogenous RNAs (ceRNAs): new entrants to the intricacies of gene regulation. Front Genet. 2014;5: 8.

9. Friedman RC, Farh KK-H, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009;19: 92–105.

10. Jiang Q, Wang Y, Hao Y, Juan L, Teng M, Zhang X, et al. miR2Disease: a manually curated database for microRNA deregulation in human disease. Nucleic Acids Res. 2009;37: D98–104.

11. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009. Available: https://www.sciencedirect.com/science/article/pii/S0092867409000087

12. Chipman LB, Pasquinelli AE. miRNA Targeting: Growing beyond the Seed. Trends Genet. 2019;35: 215–222.

13. McGeary SE, Lin KS, Shi CY, Pham TM, Bisaria N, Kelley GM, et al. The biochemical basis of microRNA targeting efficacy. Science. 2019;366. doi:10.1126/science.aav1741

14. Salmena L, Poliseno L, Tay Y, Kats L, Pandolfi PP. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? Cell. 2011;146: 353–358.

15. Tsang JS, Ebert MS, van Oudenaarden A. Genome-wide dissection of microRNA functions and cotargeting networks using gene set signatures. Mol Cell. 2010;38: 140–153.

16. Arvey A, Larsson E, Sander C, Leslie CS, Marks DS. Target mRNA abundance dilutes microRNA and siRNA activity. Molecular Systems Biology. 2010. p. 363. doi:10.1038/msb.2010.24

17. Tay Y, Rinn J, Pandolfi PP. The multilayered complexity of ceRNA crosstalk and competition. Nature. 2014;505: 344–352.

18. Dykes IM, Emanueli C. Transcriptional and Post-transcriptional Gene Regulation by Long Non-coding RNA. Genomics Proteomics Bioinformatics. 2017;15: 177–186.

19. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015;4. doi:10.7554/eLife.05005

20. Wang P, Guo Q, Qi Y, Hao Y, Gao Y, Zhi H, et al. LncACTdb 3.0: an updated database of experimentally supported ceRNA interactions and personalized networks contributing to precision medicine. Nucleic Acids Res. 2022;50: D183–D189.

21. Muniategui A, Pey J, Planes FJ, Rubio A. Joint analysis of miRNA and mRNA expression data. Brief Bioinform. 2013;14: 263–278.

22. Choobdar S, Ahsen ME, Crawford J, Tomasoni M, Fang T, Lamparter D, et al. Assessment of network module identification across complex diseases. Nat Methods. 2019;16: 843–852.

23. Zhang J, Liu L, Xu T, Zhang W, Li J, Rao N, et al. Time to infer miRNA sponge modules. Wiley Interdiscip Rev RNA. 2021; e1686.

24. Lazareva O, Baumbach J, List M, Blumenthal DB. On the limits of active module identification. Brief Bioinform. 2021;22. doi:10.1093/bib/bbab066

25. Wang P, Guo Q, Hao Y, Liu Q, Gao Y, Zhi H, et al. LnCeCell: a comprehensive database of predicted lncRNA-associated ceRNA networks at single-cell resolution. Nucleic Acids Res. 2021;49: D125–D133.

26. List M, Dehghani Amirabad A, Kostka D, Schulz MH. Large-scale inference of competing endogenous RNA networks with sparse partial correlation. Bioinformatics. 2019;35: i596–i604.

27. Hoffmann M, Pachl E, Hartung M, Stiegler V. SPONGEdb: a pan-cancer resource for competing endogenous RNA interactions. Narodonaselenie. 2021. Available: https://academic.oup.com/narcancer/article-abstract/3/1/zcaa042/6066568

28. Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature. 2012;486: 346–352.

29. Goldman MJ, Craft B, Hastie M, Repečka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38: 675–678.

30. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6: l1.

31. Rodrigues FA. Network Centrality: An Introduction. A Mathematical Modeling Approach from Nonlinear Dynamics to Complex Systems. 2019. pp. 177–196. doi:10.1007/978-3-319-78512-7_10

32. Zheng F, Wei L, Zhao L, Ni F. Pathway Network Analysis of Complex Diseases Based on Multiple Biological Networks. Biomed Res Int. 2018;2018: 5670210.

33. Freeman LC. Centrality in social networks conceptual clarification. Soc Networks. 1978;1: 215–239.

34. He X, Zhang J. Why do hubs tend to be essential in protein networks? PLoS Genet. 2006;2: e88.

35. Zhu X, Gerstein M, Snyder M. Getting connected: analysis and principles of biological networks. Genes Dev. 2007;21: 1010–1024.

36. Lu X, Jain VV, Finn PW, Perkins DL. Hubs in biological interaction networks exhibit low changes in expression in experimental asthma. Mol Syst Biol. 2007;3: 98.

37. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. Proc Natl Acad Sci U S A. 2004;101: 3747–3752.

38. Newman ME. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. Phys Rev E Stat Nonlin Soft Matter Phys. 2001;64: 016132.

39. Brandes U. A faster algorithm for betweenness centrality. J Math Sociol. 2001;25: 163–177.

40. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. Soc Networks. 2010;32: 245–251.

41. Opsahl T. Structure and evolution of weighted networks. Queen Mary, University of London. 2009. Available: https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.507253.

42. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013;14: 7.

43. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su M-J, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell. 2018;175: 984–997.e24.

44. Breiman L. Random Forests. Mach Learn. 2001;45: 5–32.

45. Kuhn M. Building Predictive Models inRUsing thecaretPackage. J Stat Softw. 2008;28. doi:10.18637/jss.v028.i05

46. Liaw A, Wiener M, Others. Classification and regression by randomForest. R news. 2002;2: 18–22.

47. Dumitrescu RG. Interplay Between Genetic and Epigenetic Changes in Breast Cancer Subtypes. In: Dumitrescu RG, Verma M, editors. Cancer Epigenetics for Precision Medicine : Methods and Protocols. New York, NY: Springer New York; 2018. pp. 19–34.

48. Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406: 747–752.

49. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490: 61–70.

50. Mulrane L, McGee SF, Gallagher WM, O'Connor DP. miRNA dysregulation in breast cancer. Cancer Res. 2013;73: 6554–6562.

51. Hamam R, Hamam D, Alsaleh KA, Kassem M, Zaher W, Alfayez M, et al. Circulating microRNAs in breast cancer: novel diagnostic and prognostic biomarkers. Cell Death Dis. 2017;8: e3045.

52. Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget. 2014;5: 9864–9876.

53. Benaglia T, Chauveau D, Hunter DR, Young DS. mixtools: An R Package for Analyzing Mixture Models. J Stat Softw. 2010;32: 1–29.

54. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 2013;4: 2612.

55. Hanley CJ, Henriet E, Sirka OK, Thomas GJ, Ewald AJ. Tumor-Resident Stromal Cells Promote Breast Cancer Invasion through Regulation of the Basal Phenotype. Mol Cancer Res. 2020;18: 1615–1622.

56. Asleh K, Negri GL, Spencer Miko SE, Colborne S, Hughes CS, Wang XQ, et al. Proteomic analysis of archival breast cancer clinical specimens identifies biological subtypes with distinct survival outcomes. Nat Commun. 2022;13: 896.

57. Ghamrawi N, McCallum A. Collective multi-label classification. Proceedings of the 14th ACM international conference on Information and knowledge management. New York, NY, USA: Association for Computing Machinery; 2005. pp. 195–200.

58. Yang J, Li C, Li H, E C. LncRNA CACNA1G-AS1 facilitates hepatocellular carcinoma progression through the miR-2392/C1orf61 pathway. J Cell Physiol. 2019;234: 18415–18422.

59. Yu P-F, Kang A-R, Jing L-J, Wang Y-M. Long non-coding RNA CACNA1G-AS1 promotes cell

migration, invasion and epithelial-mesenchymal transition by HNRNPA2B1 in non-small cell lung cancer. Eur Rev Med Pharmacol Sci. 2018;22: 993–1002.

60. Wei L-J, Bai D-M, Wang Z-Y, Liu B-C. Upregulated lncRNA CACNA1G-AS1 aggravates the progression of colorectal cancer by downregulating p53. Eur Rev Med Pharmacol Sci. 2020;24: 130–136.

61. Wang S, Ni B, Zhang Z, Wang C, Wo L, Zhou C, et al. Long non-coding RNA DNM3OS promotes tumor progression and EMT in gastric cancer by associating with Snail. Biochem Biophys Res Commun. 2019;511: 57–62.

62. Huang G-W, Zhang Y-L, Liao L-D, Li E-M, Xu L-Y. Natural antisense transcript TPM1-AS regulates the alternative splicing of tropomyosin I through an interaction with RNA-binding motif protein 4. Int J Biochem Cell Biol. 2017;90: 59–67.

63. Zhang Q, Jin X, Shi W, Chen X, Pang W, Yu X, et al. A long non-coding RNA LINC00461-dependent mechanism underlying breast cancer invasion and migration via the miR-144-3p/KPNA2 axis. Cancer Cell Int. 2020;20: 137.

64. Dong L, Qian J, Chen F, Fan Y, Long J. LINC00461 promotes cell migration and invasion in breast cancer through miR-30a-5p/integrin β3 axis. J Cell Biochem. 2019;120: 4851–4862.

65. Ji D, Wang Y, Li H, Sun B, Luo X. Long non-coding RNA LINC00461/miR-149-5p/LRIG2 axis regulates hepatocellular carcinoma progression. Biochem Biophys Res Commun. 2019;512: 176–181.

66. Chen Y, He J, Su C, Wang H, Chen Y, Guo W, et al. LINC00461 affects the survival of patients with renal cell carcinoma by acting as a competing endogenous RNA for microRNA-942. Oncol Rep. 2019;42: 1924–1934.

67. Hou J, Wang Y, Zhang H, Hu Y, Xin X, Li X. Silencing of LINC00461 enhances radiosensitivity of lung adenocarcinoma cells by down-regulating HOXA10 via microRNA-195. J Cell Mol Med. 2020;24: 2879–2890.

68. Khuu C, Utheim TP, Sehic A. The Three Paralogous MicroRNA Clusters in Development and Disease, miR-17-92, miR-106a-363, and miR-106b-25. Scientifica . 2016;2016: 1379643.

69. Gregorova J, Vychytilova-Faltejskova P, Sevcikova S. Epigenetic Regulation of MicroRNA Clusters and Families during Tumor Development. Cancers . 2021;13. doi:10.3390/cancers13061333

70. Wong KK. DNMT1 as a therapeutic target in pancreatic cancer: mechanisms and clinical implications. Cell Oncol . 2020;43: 779–792.

71. Zagorac S, Alcala S, Bayon GF, Kheir TB, Schoenhals M, González-Neira A, et al. DNMT1 Inhibition Reprograms Pancreatic Cancer Stem Cells via Upregulation of the miR-17-92 Cluster. Cancer Research. 2016. pp. 4546–4558. doi:10.1158/0008-5472.can-15-3268

72. Dar AA, Majid S, Rittsteuer C, de Semir D, Bezrookove V, Tong S, et al. The role of miR-18b in MDM2-p53 pathway signaling and melanoma progression. J Natl Cancer Inst. 2013;105: 433–442.

73. Harris SL, Levine AJ. The p53 pathway: positive and negative feedback loops. Oncogene. 2005;24: 2899–2908.

74. Yang C, Cai J, Wang Q, Tang H, Cao J, Wu L, et al. Epigenetic silencing of miR-130b in ovarian cancer promotes the development of multidrug resistance by targeting colony-stimulating factor 1. Gynecol Oncol. 2012;124: 325–334.

75. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47: D155–D162.

76. Ramalho-Carvalho J, Graça I, Gomez A, Oliveira J, Henrique R, Esteller M, et al. Downregulation of miR-130b~301b cluster is mediated by aberrant promoter methylation and impairs cellular senescence in prostate cancer. J Hematol Oncol. 2017;10: 43.

77. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100: 57–70.

78. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144: 646–674.

79. Fort RS, Mathó C, Oliveira-Rizzo C, Garat B, Sotelo-Silveira JR, Duhagon MA. An integrated view of the role of miR-130b/301b miRNA cluster in prostate cancer. Exp Hematol Oncol. 2018;7: 10.

80. Bao X, Ren T, Huang Y, Sun K, Wang S, Liu K, et al. Knockdown of long non-coding RNA HOTAIR increases miR-454-3p by targeting Stat3 and Atg12 to inhibit chondrosarcoma growth. Cell Death Dis. 2017;8: e2605.

81. Li B-L, Lu W, Lu C, Qu J-J, Yang T-T, Yan Q, et al. CpG island hypermethylation-associated silencing of microRNAs promotes human endometrial cancer. Cancer Cell Int. 2013;13: 44.

82. Barros-Silva D, Costa-Pinheiro P, Duarte H, Sousa EJ, Evangelista AF, Graça I, et al. MicroRNA-27a-5p regulation by promoter methylation and MYC signaling in prostate carcinogenesis. Cell Death Dis. 2018;9: 167.

83. Vogt M, Munding J, Grüner M, Liffers S-T, Verdoodt B, Hauk J, et al. Frequent concomitant inactivation of miR-34a and miR-34b/c by CpG methylation in colorectal, pancreatic, mammary, ovarian, urothelial, and renal cell carcinomas and soft tissue sarcomas. Virchows Archiv. 2011. pp. 313–322. doi:10.1007/s00428-010-1030-5

84. Xie Y, Wang Y, Xue W, Zou H, Li K, Liu K, et al. Profiling and integrated analysis of differentially expressed microRNAs as novel biomarkers of hepatocellular carcinoma. Front Oncol. : 5917.

85. Dai W, He J, Zheng L, Bi M, Hu F, Chen M, et al. miR-148b-3p, miR-190b, and miR-429 Regulate Cell Progression and Act as Potential Biomarkers for Breast Cancer. J Breast Cancer. 2019;22: 219–236.

86. Li Y, Wang Z, Zhao F, Zeng J, Yang X. MicroRNA-190b expression predicts a good prognosis and attenuates the malignant progression of pancreatic cancer by targeting MEF2C and TCF4. Oncol Rep. 2022;47. doi:10.3892/or.2021.8223

87. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010;363: 1938–1948.

88. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J Clin Invest. 2011;121: 2750–2767.

89. Kothari C, Clemenceau A, Ouellette G, Ennour-Idrissi K, Michaud A, C-Gaudreault R, et al. TBC1D9: An Important Modulator of Tumorigenesis in Breast Cancer. Cancers . 2021;13. doi:10.3390/cancers13143557

90. He J, Wu M, Xiong L, Gong Y, Yu R, Peng W, et al. BTB/POZ zinc finger protein ZBTB16 inhibits breast cancer proliferation and metastasis through upregulating ZBTB28 and antagonizing BCL6/ZBTB27. Clin Epigenetics. 2020;12: 82.

91. Roll JD, Rivenbark AG, Sandhu R, Parker JS, Jones WD, Carey LA, et al. Dysregulation of the epigenome in triple-negative breast cancers: basal-like and claudin-low breast cancers express aberrant DNA hypermethylation. Exp Mol Pathol. 2013;95: 276–287.

92. Moyano JV, Evans JR, Chen F, Lu M, Werner ME, Yehiely F, et al. AlphaB-crystallin is a novel oncoprotein that predicts poor clinical outcome in breast cancer. J Clin Invest. 2006;116: 261–270.

93. Sitterding SM, Wiseman WR, Schiller CL, Luan C, Chen F, Moyano JV, et al. AlphaB-crystallin: a novel marker of invasive basal-like and metaplastic breast carcinomas. Ann Diagn Pathol. 2008;12: 33–40.

94. Chen Z, Ruan Q, Han S, Xi L, Jiang W, Jiang H, et al. Discovery of structure-based small molecular inhibitor of αB-crystallin against basal-like/triple-negative breast cancer development in vitro and in vivo. Breast Cancer Res Treat. 2014;145: 45–59.

95. Coyle KM, Murphy JP, Vidovic D, Vaghar-Kashani A, Dean CA, Sultan M, et al. Breast cancer subtype dictates DNA methylation and ALDH1A3-mediated expression of tumor suppressor RARRES1. Oncotarget. 2016;7: 44096–44112.

96. Khaled WT, Choon Lee S, Stingl J, Chen X, Raza Ali H, Rueda OM, et al. BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nat Commun. 2015;6: 5987.

97. Barghash A, Helms V, Kessler SM. Overexpression of IGF2 mRNA-Binding Protein 2 (IMP2/p62) as a Feature of Basal-like Breast Cancer Correlates with Short Survival. Scandinavian journal of immunology. 2015. pp. 142–143.

98. Hsu Y-H, Yao J, Chan L-C, Wu T-J, Hsu JL, Fang Y-F, et al. Definition of PKC-α, CDK6, and MET as therapeutic targets in triple-negative breast cancer. Cancer Res. 2014;74: 4822–4835.

99. Statello L, Guo C-J, Chen L-L, Huarte M. Gene regulation by long non-coding RNAs and its biological functions. Nat Rev Mol Cell Biol. 2021;22: 96–118.

100. Castro MAA, de Santiago I, Campbell TM, Vaughn C, Hickey TE, Ross E, et al. Regulators of genetic risk of breast cancer identified by integrative network analysis. Nat Genet. 2016;48: 12–21.

101. Chagas VS, Groeneveld CS, Oliveira KG, Trefflich S, de Almeida RC, Ponder BAJ, et al. RTNduals: an R/Bioconductor package for analysis of co-regulation and inference of dual regulons. Bioinformatics. 2019;35: 5357–5358.

102. Fletcher MNC, Castro MAA, Wang X, de Santiago I, O'Reilly M, Chin S-F, et al. Master regulators of FGFR2 signalling and breast cancer risk. Nat Commun. 2013;4: 2464.

103. Van de Sande B, Flerin C, Davie K, De Waegeneer M, Hulselmans G, Aibar S, et al. A scalable SCENIC workflow for single-cell gene regulatory network analysis. Nat Protoc. 2020;15: 2247–2276.

104. Ma A, Wang C, Chang Y, Brennan FH, McDermaid A, Liu B, et al. IRIS3: integrated cell-type-specific regulon inference server from single-cell RNA-Seq. Nucleic Acids Res. 2020;48: W275–W286.

105. Gibbs CS, Jackson CA, Saldi G-A, Tjärnberg A, Shah A, Watters A, et al. High performance single-cell gene regulatory network inference at scale: The Inferelator 3.0. bioRxiv. 2022. p. 2021.05.03.442499. doi:10.1101/2021.05.03.442499