DysRegNet: Patient-specific and confounder-aware dysregulated network inference

Olga Lazareva[1,5,6,7*], Zakaria Louadi[1,2*], Johannes Kersting[1], Jan Baumbach[2,3], David B. Blumenthal[4], and Markus List[1,✉]

[1]Chair of Experimental Bioinformatics, TUM School of Life Sciences, Technical University of Munich, Freising, Germany
[2]Institute for Computational Systems Biology, University of Hamburg, Hamburg, Germany
[3]Department of Mathematics and Computer Science, University of Southern Denmark, Odense, Denmark
[4]Department Artificial Intelligence in Biomedical Engineering (AIBE), Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen, Germany
[5]Division of Computational Genomics and Systems Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany
[5]Junior Clinical Cooperation Unit Multiparametric methods for early detection of prostate cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany
[6]European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany
[*]These authors contributed equally

Gene regulation is frequently altered in diseases in unique and often patient-specific ways. Hence, personalized strategies have been proposed to infer patient-specific gene-regulatory networks. However, existing methods do not focus on disease-specific dysregulation or lack assessments of statistical significance. Moreover, they do not account for clinically important confounders such as age, sex or treatment history.

To overcome these shortcomings, we present DysRegNet, a novel method for inferring patient-specific regulatory alterations (dysregulations) from gene expression profiles. We compared DysRegNet to state-of-the-art methods and demonstrated that DysRegNet produces more interpretable and biologically meaningful networks. Independent information on promoter methylation and single nucleotide variants further corroborate our results. We apply DysRegNet to eleven TCGA cancer types and illustrate how the inferred networks can be used for downstream analysis. We show that unique as well as cancer-type-specific dysregulation patterns exist and highlight immune-related mechanisms that are not obvious when focusing on individual genes rather than their interactions.

DysRegNet is available as a Python package (https://github.com/biomedbigdata/DysRegNet_package) and analysis results for eleven TCGA cancer types are further available through an interactive web interface (https://exbio.wzw.tum.de/dysregnet).

Correspondence: *markus.list@tum.de*

Introduction

Gene regulatory network (GRN) inference methods model the relation between different genes based on their co-expression, relying on measures such as (conditional) mutual information or (partial) correlation to infer edges [1]. Since this does not allow for distinguishing between direct and indirect effects, a common strategy is to consider the interaction of transcription factors (TFs) with putative target genes to create a directed GRN. While methods such as GENIE3 [2] or ARACNE [3] identify static GRNs from gene-expression data, dynamic methods have also been developed to compare the co-expression in different conditions. Such methods were successful in identifying differential gene regulation programs [4], disease modules and perturbation in regu-

lation. For example, apoptosis was shown to be activated or repressed in cancer, depending on the tumor state and environment [5].

Methods for differential expression and co-expression analysis are designed to compare between two groups or more (e.g. disease and control or treated and untreated patients). However, because of the heterogeneous nature of complex diseases such as cancer, these approaches are limited in their ability to identify disease subgroups or to describe patient-specific dysregulatory patterns. As a result, we often lack specific biological or therapeutic insights for individual patients. Few recent studies have shown that individual differences in gene expression can lead to new insights that cannot be captured by the general groups comparison [6, 7]. These methods identify patient-specific aberrations of gene expression in a one-against-all comparison, where sample-specific outlier genes are reported.

While capturing aberrantly expressed genes at the patient level is useful, such approaches cannot pinpoint the source of the dysregulation since they do not account for indirect effects and effects that affect only one of the interaction partners. For instance, if a mutated TF retains a normal expression pattern but leads to a change of the expression of its target gene, such approaches would not capture this. These observations show the importance of considering co-expression at the single patient level. Only few approaches have been proposed to this end which are mainly based on gene-gene correlation [8–11]. Specifically, these methods calculate the Pearson correlation between two genes before and after adding/removing one sample. Some approaches such as SSN [11] evaluate the significance of this difference using transformations to z-scores or p-values. Other frameworks like LIONESS [10] do not offer any significance evaluation. Nakazawa et al. [12] define an edge contribution value to extract subnetworks from Bayesian networks inferred from all samples. Following this approach, they successfully identified novel and known cancer subtypes. A limitation of above mentioned approaches is that they do not correct for confounders such as sex, age, and origin of the sample which can greatly impact the analysis at a single sample level.
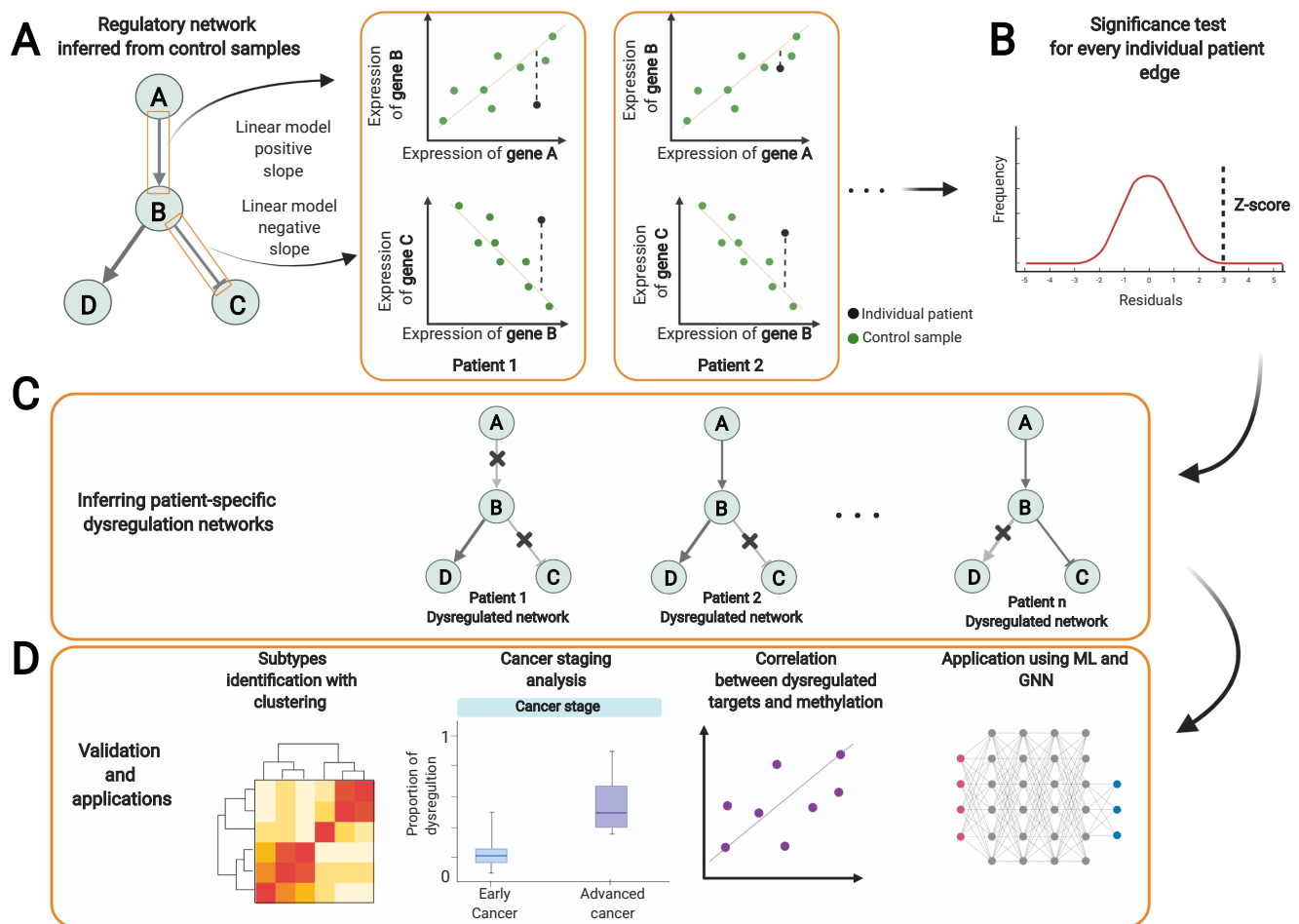
Fig. 1. Overview of the method. (A) We first infer a baseline gene regulatory network using only control samples. For illustration, we assume our network has only 3 edges with two activating and one repressing interaction. For all edges, we fit a linear model using all control samples and then test each individual patient sample, one at a time. (B) By comparing the observed value with the expected value from the linear model, we calculate patient-specific residual for every interaction; the latter is then transformed to a z-score and a p-value assuming the residual normality. (C) In the final step, the corrected p-values are used to infer patient-specific dysregulation. For instance, if gene A is an activator for gene B, we notice that for patient 1, the observed expression of gene B is significantly lower than the expected value from the positive slope of the linear model. This can be a sign of dysregulation of the activation. However, for patient 2, we notice that the activation is normal. After testing for every edge in the initial network, the patient-specific dysregulation networks are inferred. (D) We validate the generated networks using known cancer subtypes and other omics data. We also show how they can be used as input for machine learning algorithms such as support vector machines or graph neural networks. Created with BioRender.com

To mitigate this, we propose a novel method called DysReg-Net. DysRegNet employs linear models, using the TF expression as an explanatory variable and its target as a response variable. These linear models are inferred from all available control samples. Subsequently, we test for each patient sample if the co-expression pattern deviates from the expected value obtained from the linear model by considering its residual. Using a linear model allows us to correct for known covariates and allows fast computation of model residuals to identify dysregulated TF-gene interactions. This is an advantage compared to methods such as LIONESS or SSN which need to compute TF-gene correlations with and without each of sample of interest and cannot account for covariates. Note that our model can in principle be used for one-against-all comparisons where control samples for a baseline model are missing, where we would still retain the advantage of accounting for covariates.

To validate our approach, we perform an extensive pan-cancer analysis and showed that DysRegNet outperforms correlation-based methods in terms of biological relevance

and speed. Next, we investigate to what extent our approach can be informative at the patient level using validation from other omics data such as mutation and promoter methylation. We show that dysregulated edges are associated with mutated TF complexes and methylated targets. We further provide multiple use cases demonstrating the value of patient-specific networks, ranging from identifying known subtypes of cancer and driver genes to improve cancer stage prediction and providing new insights into cancer progression. Finally, we show how the topological features of the patient-specific dysregulated networks can be used to train a graph neural network to identify active modules related to cancer progression.

Results

A. Overview of the method. DysRegNet requires an initial GRN, which can be inferred from the control samples using methods such as GENIE3 [2] or ARACNE [3] or include experimentally verified interactions such as in HTRIdb
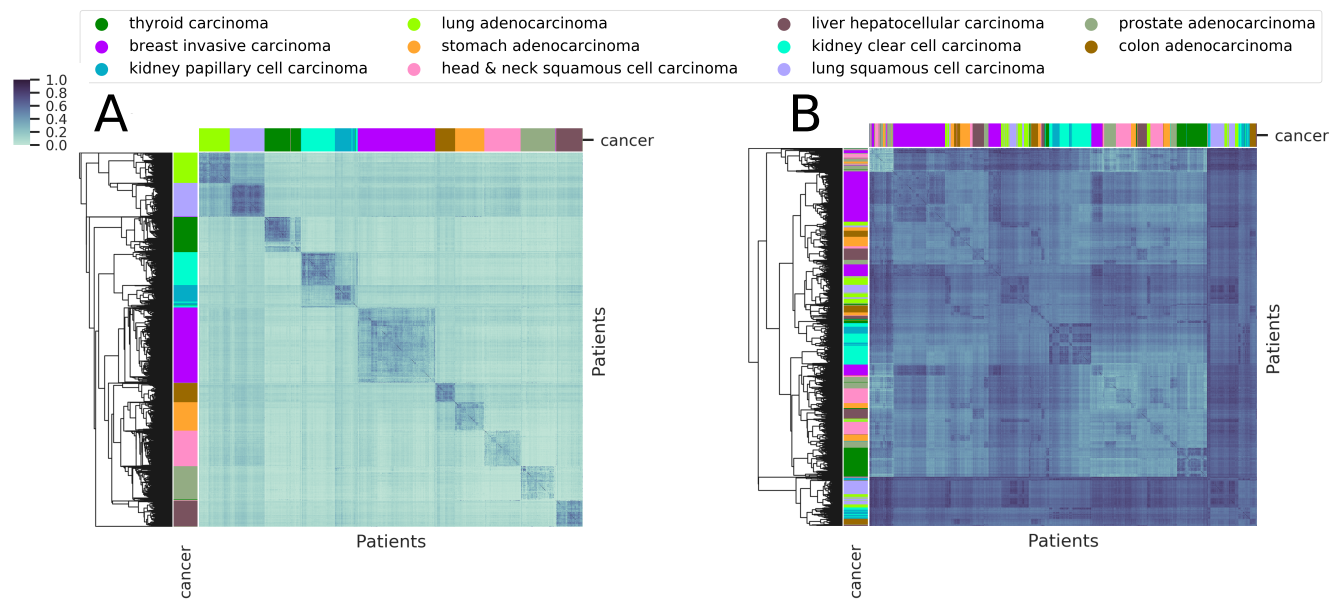
Fig. 2. Heatmaps and hierarchical clusterings based on pairwise patient similarities. A: Pairwise patient similarities are computed as overlap coefficients of edges contained in dysregulatory networks. B: Pairwise patient similarities are computed as overlap coefficients of nodes contained in dysregulatory networks. Here we observe that patients stratification based on dysregulated edges is superior to stratification based only on nodes.

[13]. The motivation for using a template network is to define feasible interactions and to reduce the number of false positives by relying on experimental evidence. The template GRN contains interactions of TFs with their target genes that could either represent an activation (positive correlation) or repression (negative correlation). When describing our own method we use networks inferred with the GENIE3 approach. Depending on the context, we use the GENIE3 individual network, which is derived from control samples for each cancer type individually, or the GENIE3 shared network, which comprises the intersection of edges across GENIE3 cancer networks. When we compare DysRegNet performance to other methods, we also use experimentally validated TF-target regulations from HTRIdb [13] and the STRING network [14], the latter of which is not limited to gene-regulatory interactions but considers protein-protein interactions. We included a PPI network to the benchmark analysis for a fair comparison with the SSN method that was evaluated using the STRING network. We fit a linear model for every edge in the network using all control samples. For illustration purposes, we consider gene A as an explanatory variable together with known covariates such as age, sex and ethnicity Figure 1A. After estimating the model parameters using Ordinary Least Squares (see Methods), we test every patient sample individually by comparing the expected value of gene B with the observed value specific to the patient sample. Since we verified normality of the error terms (residuals), we can calculate a z-score specific to the test sample using a standardized residual. This technique is comparable with an outlier detection task in regression analysis. After evaluating all patients, the z-scores are then transformed to p-values and corrected for multiple testing (see Methods). After modeling every edge in the initial GRN, the output of our method is a list of predicted dysregulated edges for every patient which can be integrated into a network with one or several connected components. It is important to note that previous studies used the term "patient- (or sample-) specific regulatory network". We prefer to call it a patient-specific dysregulated network since, using current approaches, we can only identify outliers w.r.t. the original GRN but not learn new edges or gain of function specific to one sample.

B. Pan-cancer analysis for assessment of biological relevance. To assess the biological relevance of the patient-specific dysregulatory networks computed by DysRegNet, we analysed the networks obtained for eleven cancer types available in TCGA. Three types of analysis were carried out, namely, patient clustering based on the computed networks, promoter methylation analysis for dysregulated targets, and comparison of dysregulation scores for the targets of mutated TFs. As detailed below, all three analyses generated evidence indicating that the networks computed by DysRegNet indeed capture biologically relevant signal.

B.1. Validation via clustering of dysregulated patient networks. Pan-cancer analysis was performed to identify similarities among different types of cancers regarding patient-specific dysregulations. Our primary assumption is that patients with the same type of cancer have similar dysregulated edges, while patients with different cancers do not have large edge overlaps. Some cancer types in our analysis originate from the same organ: lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), kidney renal clear cell carcinoma (KIRC), and kidney renal papillary cell carcinoma (KIRP) and, therefore, we expect them to be more similar to one another compared to cancer types of different origin. Detailed comparison of methods for KIRC, KIRP, LUAD and LUSC is provided in the Supplementary Material. To make sure that we focus on similarities between patients
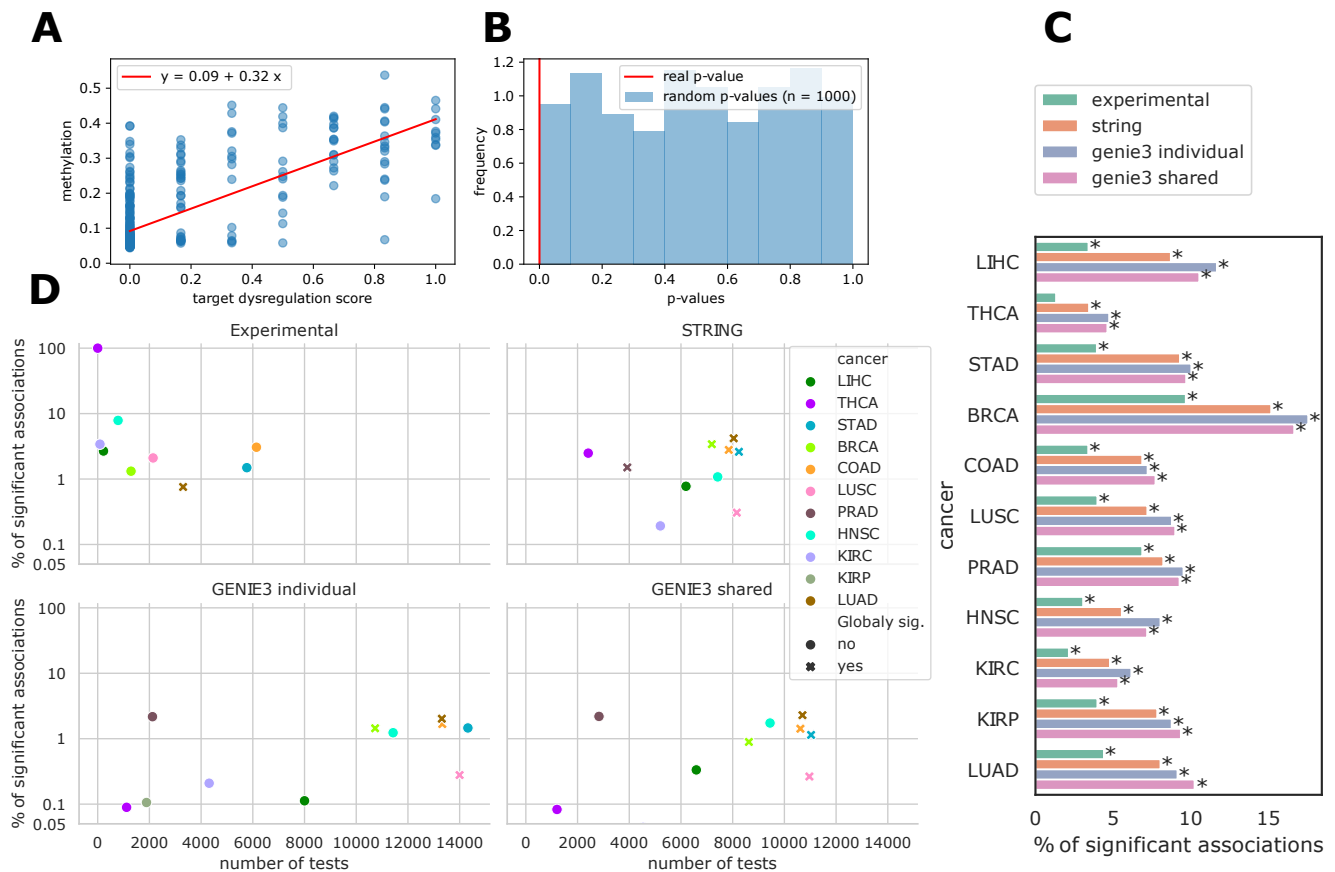
Fig. 3. Evaluation of DysRegNet with additional methylation and mutation data. A: Relationship between normalized degree of ZNF549 and its promoter methylation. B: p-value for the slope coefficient in the global methylation model with all targets vs. slope coefficients p-values for models where the degrees of the target genes have been shuffled. C: Significant associations between promoter methylation and target gene dysregulation. Bars with a star indicate that a global model (with all promoters) showed significant association between promoter methylation and target dysregulation. D: Associations between mutations in TFs and target genes dysregulation. TFs-target complexes were tested only when at least seven patients have a mutation in a TF. The x-axis shows how many TFs-target complexes have been tested. The y-axis shows the percentage of significant association out of all tested complexes.

based on cancer-type-specific dysregulation and not based on tissue-specific expression, we used the shared GENIE3 network that excludes tissue-specific edges and only comprises edges that are common in all cancer types. We evaluate the similarity between individual patient networks by computing a pairwise overlap coefficient for the set of edges (see Methods). To emphasize that our analysis focuses on differences in edges (dysregulations) and not abnormal single gene expression, we also computed the overlap between node sets. Figure 2 illustrates that node clustering shows far worse stratification of patients in different cancer types. Additionally, edge clustering demonstrates that LUAD and LUSC as well as KIRC and KIRP are their respective closest neigbours, as expected. In comparison, the other methods performed poorly and failed to identify similarities between patients of the same subtype (Supplementary Figure 1). Note that in order to reduce the effect of tissue-specificity, this comparison was performed on subtypes from the same (or similar) tissue. A more detailed quantitative comparison with these approaches is presented in subsection C.

B.2. Genes with methylated promoters are more likely to be dysregulated . DNA methylation plays a crucial role in controlling gene expression. For instance, methylation of cy-

tosines (in a CpG) context in the promoter region is associated with gene repression as TFs cannot bind anymore [15]. Gene promoter methylation is thus also used as a diagnostic and prognostic cancer biomarker [16].

Even though changes in promoter methylation represent only one possible cause of dysregulation, we hypothesize that promoter methylation across a large number of samples should be correlated with dysregulation of a target gene. To test this hypothesis, we evaluated all targets individually (local model) and all targets together (global model). The local model allowed us to evaluate how many targets exhibit a correlation between the node degree and promoter methylation (an example of such a target is shown in Figure 3 A).

We reason that a high node in-degree of a target gene (i.e. many dysregulated edges) indicates that TFs generally lost the ability to regulate the target as we would expect if they can no longer bind to the promoter. Note that while this would not affect TFs binding to enhancer regions, it suffices to show that dysregulation is related to changes in DNA methylation. We subsequently build a global model that allowed us to evaluate if this pattern holds in general in a given condition (for details see Methods sections O.2, O.3). An example of an empirical p-value distribution obtained from fitting the global model after repeatedly shuffling the target

Lazareva *et al.* | DysRegNet

gene's node degrees is shown in Figure 3 B. An illustration of differences between the global and the local model can be found in the Supplementary Material (Figure 3).

The local model showed that for all 11 cancer types tested, a significant association of node degree and promoter methylation was observed for at least 5% of all targets (for THCA) and at most 17% of all targets (for BRCA). The global model demonstrated that the assumption holds for all 11 considered cancer types (Figure 3 C).

B.3. Dysregulated targets are associated with mutated TFs.

Mutations in TFs are associated with many types of cancer: lung cancer [17], prostate cancer [18], breast cancer [19], and many others [20]. We hypothesize that TF mutations lead to a higher target gene dysregulation. We performed analysis on a local level for individual targets and on a global level for all target dysregulations in every cancer type to verify this hypothesis.

In the local model, we compared dysregulation scores (i.e. the percentage of dysregulated edges per node - see Methods) for targets where at least one of the regulating TFs was mutated with dysregulation scores of targets without mutated TFs using the Mann-Whitney U test and report p-values after correction for multiple hypothesis testing (see Methods for details). The global model was built as a regression where mutations in a set of co-regulating TFs were predicted based on a dysregulation score of a target gene. We then tested if the dysregulation score was significantly associated with TFs mutations for every cancer type individually.

Figure 3 D shows results of the local- and global-scale analyses for all types of cancers and networks. Based on the GE-NIE3 individual network, the most prominent connection (p-value 1.2^{-11}) was between RPL23 and TP53 in breast cancer. RPL23 was widely studied as it links oncogenic RAS signaling to p53-mediated tumor suppression [21]. Ribosomal proteins bind to and inhibit MDM2, a potentially oncogenic E3 ubiquitin ligase that interacts with and promotes the degradation of the TP53 tumor suppressor [22]. On the protein level, physical interaction between TP53 and RPL23 is confirmed by affinity chromatography [23]. As the analysis was performed on the GENIE3 inferred network, overlap with known and experimentally validated interactions and regulatory mechanisms creates additional confidence in our results.

C. Comparison against state-of-the-art methods.

Patient-specific network inference has the potential to unravel disease mechanisms for each patient individually following the precision medicine paradigm. Several published methods attempted to provide this information considering how the correlations between genes is affected by adding or removing individual patient samples.

We compared the performance of DysRegNet to two state-of-the-art methods: LIONESS and SSN. Both of these methods construct patient-specific networks by considering how the correlation between genes is affected by adding or removing individual patient samples. For all methods, we used all four template networks: GENIE3 individual and shared,
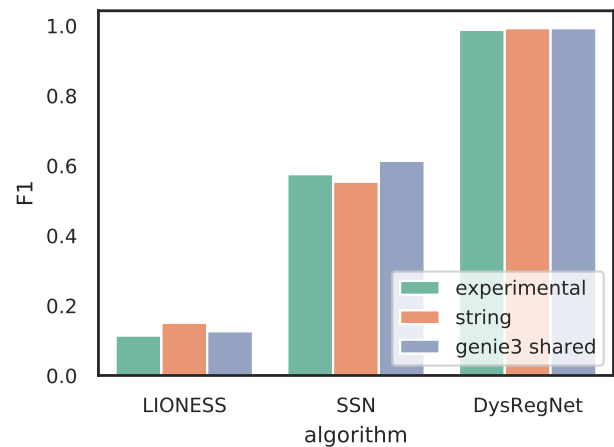

Fig. 4. F1 score for clustering performed on pair-wise similarities among patient-networks.

HTRIdb and STRING. We did not use an approach proposed by Nakazawa et al. [12] due to lack of publicly available code.

C.1. Ability of methods to distinguish known cancer types.

We assessed the ability of all tested methods to derive similar networks for patients with the same cancer type. For each pair of patients, we computed an overlap coefficient between their corresponding edge sets. Every network-algorithm pair was represented as a symmetric patient-by-patient similarity matrix that was clustered using spectral clustering. The obtained clusters were then mapped to the known classes using the Hungarian algorithm (i.e., maximizing F1 score for each class-cluster mapping) [24]. The final F1 score for every method-network pair is shown on Figure 4. The evaluation was made for GENIE3 shared network, HTRdb and STRING network. Individual (cancer-specific) GENIE3 network were excluded from the assessment due to high tissue specificity that makes it impossible to conclude if the observed dysregulation is specific for a cancer type. For the other three networks, DysRegNet consistently outperformed SSN which in turn outperformed LIONESS.

C.2. Cancer-specific gene set enrichment.

We investigated if the studied methods are capable of extracting cancer-related genes for each cancer individually. For each of the eleven studied cancers, we considered the 100 most common edges for all patients and checked if they were enriched in DisGeNet gene sets. Most TCGA cancers can be directly mapped to DisGenNet gene sets, e.g., TCGA 'lung squamous cell carcinoma' can be mapped to DisGeNet 'Squamous cell carcinoma of lung' gene set. The full mapping of TCGA cancers and DisGeNet gene sets are available in the Supplementary Material.

Figure 5 demonstrates that the method performance is highly dependent on the selected template network and studied cancer type. Prior-knowledge based networks (experimental and STRING) perform significantly better as they can potentially limit the search space to well-studied genes. SSN and LIONESS perform best for breast cancer, possibly due to the

lack of influence of the sex covariate.

D. Case study 1: Identifying the regulatory sub-networks linked to cancer progress in thyroid cancer. Disruption of TFs is a hallmark of cancer. However, one would expect more disruption in advanced stages than in the earlier stages of cancer. To investigate this assumption and to characterize the link between dysregulation and tumor-specific progression, we analyzed thyroid cancer (THCA) using patient stages annotations. We have chosen THCA because of the availability of sufficiently many samples for all stages.

We first split 511 THCA patients into two groups: early-stage and advanced stages (see Methods). Then, for every TF, we tested if the percentage of dysregulated edges is greater in the advanced stages group than in the early stage group using a one-sided Mann–Whitney U test. Out of 1,341 TFs in the initial GRN of THCA, 495 were significantly more dysregulated after multiple testing correction (Benjamini/Hochberg adjusted p-value ≤ 0.05). In comparison, the same analysis performed on the graphs generated from LIONESS and SSN led to only one significant TF for each method. As expected, using the three methods, none of the edges were more significantly dysregulated in early cancer compared to late stages. An example from DysRegNet is shown on the Figure 6, the median dysregulation of Zinc finger protein 687 (ZNF687) is higher in advanced stages. Interestingly, ZNF687 was previously associated with promoting hepatocellular carcinoma recurrence [25], but it was not fully investigated for other types of cancer.

Since the dysfunction of most TFs is expected and common in all cancer types and stages, we next focused on TF-target edges instead of single TFs. We hypothesize that only some edges of these TFs are linked with severe THCA. To examine this hypothesis, we checked for over-representation of dysregulated edges between the two groups. We compared the number of times an edge is dysregulated in each group using a one-sided Fisher's exact test. Indeed, out of the 47,236 edges from the initial list of significantly dysregulated TFs, 646 were significantly over-represented in late stages after multiple testing correction (Benjamini/Hochberg adjusted p-value ≤ 0.05). Most of the edges were linked to the TF FOXP3, a crucial transcriptional regulator for the development and inhibitory function of regulatory T-cells. FOXP3 is also a well-known tumor suppressor and plays an important role in cancer development [26]. Interestingly, the most significant edge corresponds to FOXP3 activation of the PYHIN1 gene, The encoded protein belongs to HIN-200 family which is involved in cell differentiation, and apoptosis [27]. This protein is also known to act as tumor suppressor [28]. To assess if the observed changes in immune function were related to differences in the cell type composition of the tumor microenvironment, we checked for differences in cell type enrichment between these two groups using the method xCell [29]. The results did not show any significant difference between the two groups (see Methods), except for CD4+ naive T-cells and CD8+ T-cells where a few samples exhibit a slightly higher cell-type enrichment in early stages (Supplementary Figure 8).

The observed dysregulation of this activation in a set of patients could be a sign of a disrupted immune response consistent with the progressive cancer state for these patients. Overall the over-represented edges were enriched in the pathways: "T cell receptor signaling", "Inflammatory Response Pathway" and "Hematopoietic Stem Cell Differentiation" (Figure 6). This enrichment confirms that the disruption of expression of tumor suppression genes and alterations in the regulatory T cells are associated with advanced cancer.

To explore the possibility of using dysregulated edges for patient stage predictions and as potential biomarkers, we trained an SVM model with patient-specific edges as features. An edge had a value of 1 if it was dysregulated in a patient or 0 otherwise (see Methods). To establish a baseline, we also trained a model using gene expressions for all patients and another model using TFs dysregulation percentages as features. It is important to note that even if the information at the edge level and TF level is derived from gene expression, it still contains an additional layer from the conditional expression of a gene given the expression of another gene (co-expression). All hyperparameters were selected using grid-search in nested cross-validation (see Methods). The model based on gene expression yields a mean and a standard deviation of AUC of 0.74 ± 0.05. The TFs dysregulation model performed similarly with an AUC of 0.70 ± 0.09. In contrast, the model based on binary edge features yields a higher AUC of 0.78 ± 0.03. This shows that TF dysregulation is at least as informative as the expression for stage predictions and offers a more detailed interpretation.

These observations are consistent with our initial hypothesis and confirm the importance of co-expression analysis at a single patient level as an additional layer of information. The analysis also demonstrates an opportunity to extend our understanding of complex diseases such as cancer and find new potential therapeutic targets.

E. Case study 2: Interpretable Graph Neural Networks identifies patients specific cancer stage. In the previous sections, we demonstrated that patient-specific dysregulation networks can be efficiently used in the downstream analysis for supervised and unsupervised learning. However, up to this point, every edge was treated as a binary feature, and therefore we did not take full advantage of the network topology. While application, calibration, and interpretation of graph neural networks (GNNs) is not the focus of this study, the following use case exemplifies how patient-specific GNNs can be employed for classification in the downstream analysis.

We used dysregulated networks for all cancer patients with early- and late-stage cancers and trained GNN for stage classification. As input, we used the patient-specific dysregulatory networks with the direction of regulation (1 or -1 corresponding to activation or repression) as edge features. Additionally, we encoded gene sequences and used them as node features to indicate positions of the same genes across patient graphs (more details in the Supplementary material).

Supplementary Figure 5 shows the convergence of GNN and its performance on an unseen validation set. Supplementary
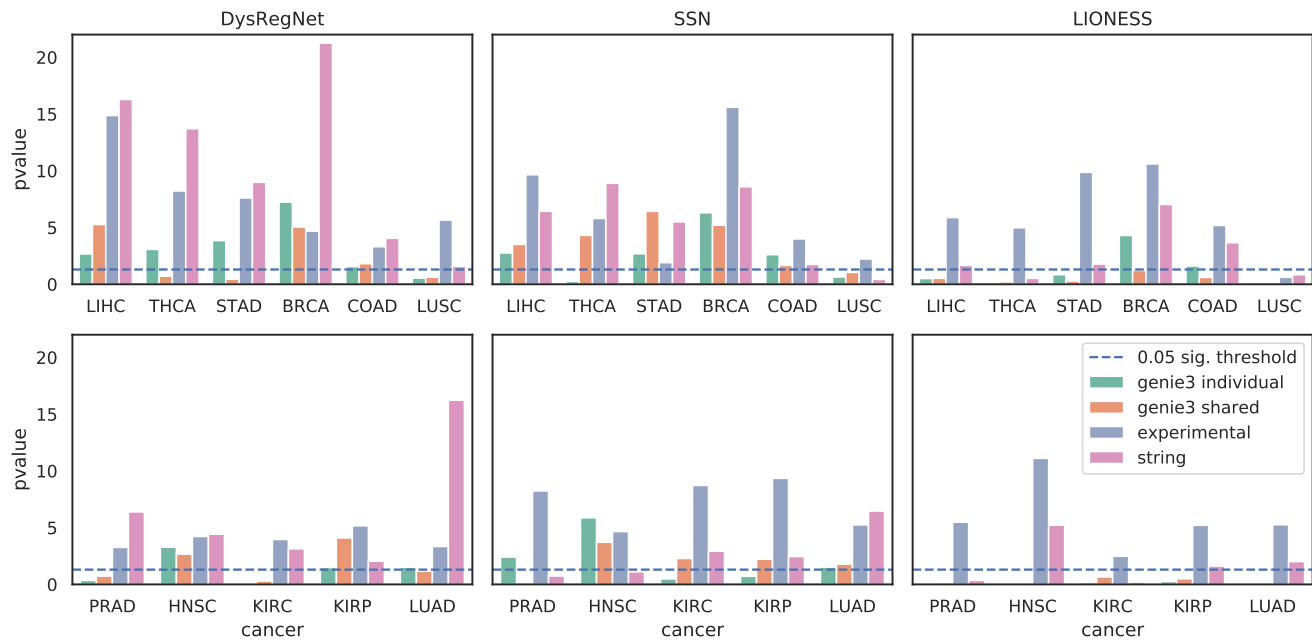
Lazareva *et al.* | DysRegNet

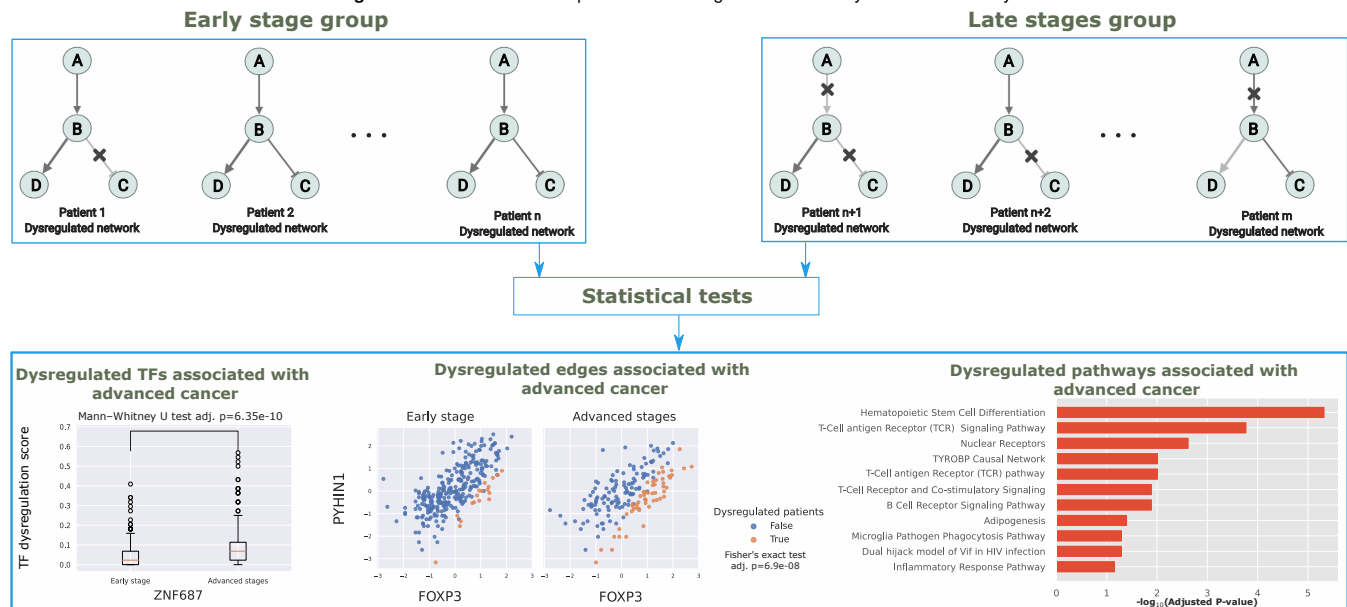**Fig. 5.** Enrichment of cancer-specific DisGeNet gene sets for every cancer individually.

Fig. 6. The regulatory sub-networks linked to cancer progress

Figure 6 shows the performance of GNN across cancer types. Mean validation accuracy is 0.68.

GNNs can be interpreted using various interpretability frameworks such as Captum [30]. We applied the integrated gradients approach [31] to obtain edge attribution, i.e., evaluate the predictive power of every edge in each patient-specific network. We extracted a cancer progression subnetwork (Supplementary Figure 7) using the top 10 the most predictive edges per cancer type.

The interpretation of the GNN results is consistent with the first case study, since FOXP3 again shows the most predictive edges. The second large component includes the E2F4 transcription factor and its targets. E2F4 is widely studied

in the context of cancer as a crucial regulator in liver cancer [32], breast cancer [33, 34], hepatocellular carcinoma [32], prostate cancer [35], skin cancer [36] and other [37]. Genes from the induced subnetwork showed significant enrichment for transcriptional dysregulation in the cancer KEGG pathway (adjusted p-value 0.000488).

This high-level analysis shows the potential of GNN application to patient-specific dysregulated networks. GNNs architectures are very flexible and can be enriched with other omics data as node features (gene expression, methylation, mutations, copy number variations, etc) and different edge types (PPIs, healthy regulations, etc). We provide implementation and code details in the Supplementary Materials.

Discussion

F. Influence of different template networks. The initial template network is necessary for DysRegNet and all other reviewed methods. In our analysis, we have considered three different template networks: computationally inferred regulatory network from GENIE3, experimentally validated network from HTRIdb, and a combined network from STRING. All networks were compared in 4 different contexts: networks clustering evaluation (i), edge enrichment evaluation (ii), the association between target dysregulation and promoter methylation (iii), the association between target dysregulation and TFs mutations (iv).

Given the different origins of the networks, we cannot conclude that one of them is universally better than others, and therefore the choice should depend on the downstream analysis. Some types of comparison may penalize false positive edges in networks (likely present in computationally inferred GENIE3 networks), while others penalize false negative edges (likely present in the experimental and STRING networks). We obtained encouraging results for the network clustering task, where DysRegNet performed equally good on all 4 networks. This implies that if a user wants to use patient networks for clustering or classification, most likely performance will not depend on the network selection. Although the interpretation of the results strongly depends on the template network choice.

Gene set enrichment worked best for the experimental and STRING network for all methods. This suggests that GENIE3 template networks introduce more false-positive edges but may be better suited for hypothesis generation. This is supported by the better performance of the GENIE3 inferred networks in the DNA methylation-dysregulation association analysis where we observed a stronger global trend. We hypothesize that the observed trend is due to understudied changes in DNA methylation that the GENIE3 inferred network allows us to test.

The mutation-dysregulation association analysis showed the best results for the STRING network. This is possibly due to the STRING network density as there are many connections between frequently mutated TFs and target genes. In general, STRING (as a PPI network) might not be the ideal source of a template network as it does not capture TF-target interactions. Interactions on a protein level can only portray indirect regulatory effects.

G. Dysregulation scenarios. The ability of a TF to predict expression of a target gene is often considered an indication of a regulatory link [2]. We consider a regulatory link dysregulated, if a model trained on control samples is not predictive of a target gene expression in a (disease) sample of interest. This implies four possible scenarios of dysregulation shown in Supplementary Figure 2: suppressed activation (1), amplified activation (2), amplified repression (3), suppressed repression (4).

We follow the assumption that changes in the expression of genes encoding TFs are followed by expression changes of the target genes they regulate [38]. Therefore, if regulation is disturbed, then changes of TFs expression do not lead to changes in target expression any longer. This pattern corresponds to scenarios 1 and 4. Scenarios 2 and 3 can be interpreted by a stronger response of the target gene, e.g. due to increased affinity of a TF to the target gene's regulatory motif. While we observed this type of correlation pattern in our data (although in far fewer cases than the other two scenarios), there is comparably little evidence in the literature to support these scenarios. In our analysis, we thus considered only scenarios 1 and 4 here and plan to follow up on the other two scenarios in the future. The DysRegNet python package allows to study all four scenarios.

H. Benchmark analysis. We benchmarked two popular methods for patient-specific network inference: LIONESS and SSN. While we only used two methods for comparison, other published methods use similar assumptions. Park et al. [8] use a similar approach to the SSN method, but instead of control data, they use gene expression data from healthy subjects in GTEx [39] . Lee et al. [9] use precisely the same approach as SSN for patient-specific network inference and further expand it to CNV, methylation, and miRNA. Due to the high similarity of methods and lack of public source code, we did not include those methods in the benchmark.

The use of control samples brings a clear advantage to a method (such as SSN or DysRegNet) as it allows to observe dysregulations common among patients. LIONESS is not using any control data, and while this has a clear advantage for experiment designs with no controls, it does not allow the observation of common dysregulation mechanisms among patients with similar phenotypes. SSN uses correlation-based approach similar to LIONESS, but the incorporation of control data produces significantly better clustering and enrichment results.

We hypothesize that better performance of DysRegNet in terms of clustering and gene set enrichment can be explained by using covariates in our statistical model. Very basic characteristics such as age [40], ethnicity [41] and sex [42] can be extremely strong confounders in certain types of cancer. We observed that SSN and LIONESS performed better for breast cancer (BRCA) analysis than other cancer types (Figure 5). A possible explanation is a lack of sex covariate influence.

I. Limitations and outlook. We evaluated DysRegNet in different scenarios to emphasize its contribution to the understanding of cancer dysregulation mechanisms at single-patient resolution. Cancer is known to be highly heterogeneous, and therefore, single-patient network extraction can be particularly advantageous for precision medicine. Additionally, TCGA data allowed us to verify our analysis on 11 different cancer types showing that all of them can be studied with DysRegNet.

Recent studies have shown that the vast majority of human genes have been studied in the context of cancer [43]. Cancer tends to cause severe perturbation to the regulatory mechanism, and therefore it is hard to disentangle the disease-driving mechanisms from their consequences. To gain a bet-

ter understanding of DysGeNet results, we will focus on applying DysRegNet to a wider range of complex diseases.

Additionally, users of the Python package should carefully consider data prepossessing strategies. We applied our method to well-studied, high-quality TCGA data that did not require extensive quality control. Users should keep in mind that the method is built based on the assumptions: (1) target gene expression can be predicted based on TF expression, (2) the residuals of the linear model are normally distributed. The first assumption typically does not hold for PPI networks. It also sometimes does not for regulatory networks, as not all TF-target pairs exhibit correlation. For those pairs, our method cannot be applied. A user needs to evaluate the goodness of fit before considering an edge as potentially dysregulated. Our python package allows users to apply an R2 filtering strategy to eliminate links where a TF does not have predictive power. The second assumption was verified for our analysis, but it is not guaranteed to hold for all gene expression data. We implement a test for normality [44, 45] of a residual distribution in our python package and encourage users to perform the tests for their data. If the normality of the residuals assumption is violated, a user might need to consider different pre-processing strategies or non-linear modeling techniques.

One promising direction for further research is the connection between dysregulations and mutations. We have investigated whether mutations in TFs co-regulating a gene can cause target gene dysregulation. Unlike the DNA methylation association analysis, the association between dysregulations and mutations was more complex to interpret. We found this effect in different cancer types, but the percentage of affected targets was not large (10 % at most). On the other hand, the strongest association was from the TP53 pathway in breast cancer, which suggests that a more focused analysis is needed to understand the mechanism at play. Our analysis did not consider that not all (somatic) mutations affect regulatory mechanisms. For example, a mutation could cause a gain of function instead of loss of interactions (new interactions for the TF). An example of a more specific analysis could focus on a known driver mutation within a DNA binding domain or, alternatively, on mutations of target gene TF binding motifs. Furthermore, somatic mutations in cancer are frequently affecting the splice-site and possibly cause isoform switches [46, 47]. Our analysis was performed at the gene level but a deeper analysis at the isoform or transcript level would help explaining a larger fraction of the identified dysregulated edges.

Another interesting application of the method is in studying rare and undiagnosed diseases, where the focus is often on the unique differences of a single sample. Up to date, the current rate of genetically diagnosed rare disorders is approximately 25 to 50% [6]. Thus, DysRegNet provides a novel opportunity to expand our knowledge on such disorders.

Additionally, we note that we cannot guarantee dysregulated edges to be cancer-related as they might also represent dysregulation that is already present in the healthy tissue of a patient. To discern such edges, it is necessary to profile both healthy and cancer samples which is only the case for a subset of samples in TCGA.

J. Conclusion. Aberrant TF regulation is an important mechanism in the development and progression of complex diseases such as cancer. Rather than focusing on the aberrant expression of TFs or their target genes, it is worthwhile to study which specific interactions of a TF are affected to gain a more detailed insight into the underlying pathomechanisms. A multitude of molecular changes can lead to the same outcome and hence, it is of importance to study dysregulation in a patient- or sample-specific manner. With DysRegNet, we present a novel approach that delineates such individual TF-regulatory changes in relation to a control cohort. In contrast to competing methods, DysRegNet uses linear models to account for confounders and residual-derived z-scores to assess significance. Due to the latter, DysRegNet scales efficiently to an arbitrary number of samples. We have shown that DysRegNet results are robust across template networks and produce meaningful insights into cancer biology, thus serving as an important systems medicine tool for data exploration and hypothesis generation in oncology and beyond.

Materials and Methods

K. Data preprocessing. All data from The Cancer Genome Atlas Program (TCGA) is acquired from the XENA browser (https://xena.ucsc.edu/) [48]. We included 11 cancers with at least 50 control samples (labeled as: solid tissue normals). As a gene expression dataset, we used TPM values for the PANCAN cohort. We retained only genes expressed in 80 % of the patients of the same cancer type, followed by z-scoring. Illumina 450k DNA methylation array data was filtered for CpGs associated with promoter regions (according to Illumina's annotation) which were aggregated using the mean. Somatic mutations were mapped to their transcripts/genes. Only missense mutation were included in our analysis since they are more likely to affect TF functions.

L. Statistical model behind DysRegNet. We define a template network $N = (G, T, E)$, where G is a set of genes, $T \subseteq G$ is a set of TFs, and $E \subseteq T \times G$ is a set of edges connecting TFs $t \in T$ to target genes $g \in G$. The role of the template network is to limit the search space for potentially dysregulated edges and to provide prior information about expected healthy regulations. We discuss possible choices for the template network in the subsection M.

For every pair of connected nodes (t_i, g_j), the relationship between expression profiles of a TF t_i and a target gene g_j can be modeled as:

$$\hat{E}^H(g_j) = \beta_0^H + \beta_1^H \cdot E^H(t_i) + \sum_{l=2}^{L} (\beta_l^H \cdot C_l^H) \quad \textbf{(1)}$$

where $E^H(t_i)$ is the expression of a TF t_i in a cohort of healthy controls, $\hat{E}^H(g_j)$ is the expected expression of a target gene g_j in a cohort of control samples, $\{C_2^H, .., C_L^H\}$ is

a set of available covariates such as age, ethnicity and sex, $\{\beta_0^H, .., \beta_L^H\}$ are coefficients estimated with an ordinary least squares model.

An edge $e^k = (t_i^k, g_j^k)$ is dysregulated for a patient k if the edge exists in the template network N and, for patient k, the expression of g_j^k cannot be reliably estimated using the model from Equation 1. Formally, this means that expected expression of $\hat{E}^k(g_j) = \beta_0^H + \beta_1^H \cdot E^k(t_i) + \sum_{l=2}^{L}(\beta_l^k \cdot C_l^k)$ is *significantly* different from the actual value of $E^k(g_j)$. This difference can be defined as a residual of the model, i.e. $r_k = \hat{E}^k(g_j) - E^k(g_j)$. We can convert r_k to a z-score using the following transformation:

$$z_k = \frac{r_k - \bar{r}_H}{\sigma(r_H)} \quad (2)$$

where $r_H = \hat{E}^H(g_j) - E^H(g_j)$ are residuals of Equation 1 model and $\bar{r}_H$ is the mean of r_H, $\sigma(r_H)$ is the standard deviation of r_H. According to our evaluations, residuals of the linear model are normally distributed, and therefore, z-scores can be further converted to probabilities $P(Z < z_k)$, which are then corrected for multiple hypothesis testing using Bonferroni correction at the 0.01 significance level.

An additional condition that we enforce for dysregulated edges is:

$$r_k \cdot \beta_1^H > 0 \quad (3)$$

This condition implies that two out of four possible scenarios of dysregulation are considered:

1. TF is an activator, but no activation of target is observed for a patient ($r_H > 0$ and $\beta_1^H > 0$).

2. TF is an activator and target is activated higher than expected ($r_H < 0$ and $\beta_1^H > 0$).

3. TF is a repressor, but target is repressed more than expected ($r_H > 0$ and $\beta_1^H < 0$).

4. TF is a repressor, but target is not repressed ($r_H < 0$ and $\beta_1^H < 0$).

Equation 6 implies that only scenario 1 and 4 are considered (all four scenarios are shown in Supplementary Material, Figure S1).

M. Template networks. The primary role of a template network is to provide prior information about possible regulations in an organism. The network can be derived from transcriptomics data using computational methods such as GENIE3 [2] or derived from ATAC-seq, ChIP-seq, and other appropriate technologies. In our analyses, we used three template networks: a in silico inferred network from GENIE3, an experimentally derived regulatory network from the Human Transcriptional Regulation Interactions database (HTRIdb) [13], and STRING [14].

GENIE3 GENIE3 uses an ensemble of trees to estimate the strength of the regulatory relationship between all possible TF-target pairs. The list of 1639 human TFs was used from Lambert et al. [49] (http://humantfs.ccbr.

utoronto.ca/). Then, the top 100 000 important regulations were used in the further analysis as a template network. In all PANCANCER analyses, we considered two possible applications of the template network: each cancer has its own template network, based on cancer-specific controls (i), a shared template network for all cancers (ii). Cancer-specific networks were computed by running GENIE3 individually on each set of controls. To obtain a shared template network, we summed up edge importance scores from every cancer-specific template network and retained the 100 000 most important edges in a shared template network.

HTRIdb The Transcriptional Regulation Interactions database (http://www.lbbc.ibb.unesp.br/htri) is an open-access database of experimentally validated TF-target gene interactions. The database provides information about regulation interactions among 284 TFs and 18302 TGs detected by 14 distinct techniques [13]. Namely, chromatin immunoprecipitation, concatenate chromatin immunoprecipitation, CpG chromatin immunoprecipitation, DNA affinity chromatography, DNA affinity precipitation assay, DNase I footprinting, electrophoretic mobility shift assay, southwestern blotting, streptavidin chromatin immunoprecipitation, surface plasmon resonance and yeast one-hybrid assay, chromatin immunoprecipitation coupled with microarray (ChIP-chip) or chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq).

STRING The STRING database (http://string-db.org/) is dedicated to protein-protein interactions. It was included in our assessment for a fair comparison between DysRegNet, and SSN [11] (see subsubsection P.1), which was originally evaluated using the STRING network. Following the described methodology, we also considered high-confidence interactions with a combined score larger than 0.9. The combined score is computed as an average of 7 channels (neighborhood, fusion, co-occurrence, co-expression, experimental, database, text mining).

N. Patient networks similarity. We evaluate the similarity between patient-specific networks by computing a pairwise overlap coefficient for the set of edges, i.e.:

$$o(p_i, p_j) = \left(\frac{|E_{p_i}^+ \bigcap E_{p_j}^+|}{min(|E_{p_i}^+|, |E_{p_j}^+|)} + \frac{|E_{p_i}^- \bigcap E_{p_j}^-|}{min(|E_{p_i}^-|, |E_{p_j}^-|)} \right) \cdot \frac{1}{2}, \quad (4)$$

where $E_{p_i}^+$ and E^-p_i are positive (dysregulated activation) and negative (dysregulated repression) edge sets for patient i, respectively.

O. Hypothesis testing for mutation and methylation analysis.

O.1. Dysregulation scores. We compute dysregulation scores to assess how much gene (TF or target) is affected by a condition in question. Dysregulation score represent the proportion of dysregulated patient-specific edges out of all edges towards target g (or from a TF t) in the template

network. Thus, for a target gene g, dysregulation score is defined as $\tilde{d}_s(g) = d_s^{in}(g)/d_T^{in}(g)$ where $d_s^{in}(g)$ is the in-degree of g in a dysregulated network of patient s and $d_T^{in}(g)$ is the in-degree of g in the template network. For a transcription factor t, dysregulation score is defined as $\tilde{d}_s(t) = d_s^{out}(t)/d_T^{out}(t)$ where $d_s^{out}(t)$ is the out-degree of t in a dysregulated network of patient s and $d_T^{out}(g)$ is the out-degree of t in the template network.

O.2. DNA Methylation local model. To model relationship between promoter DNA methylation and target dysregulation, the following model was used:

$$\widehat{me}(g) = \beta_0^{me} + \beta_1^{me} \cdot \tilde{d}(g) \tag{5}$$

Here, $me(g) = [me_1(g), \ldots, me_P(g)]$ is the vector of average (across CPGs) promoter DNA methylations of target g for all P patients and $\tilde{d}(g) = [\tilde{d}_1(g), \ldots, \tilde{d}_P(g)]$ is the vector of g's dysregulation scores.
Then the slope coefficient β_1^{me} was tested for significance of the association with null hypothesis $H_0 : \beta_1^{me} = 0$. P-values were then corrected using the Benjamini-Hochberg method.

O.3. DNA Methylation global model. While Equation 5 allows to test every target individually, we also applied a linear mixed effect model to all targets in the template network to test for the global effect. We chose the model with a random intercept coefficient for each target assuming different baseline methylation levels:

$$me(G) = \beta_0^{me*} + \beta_1^{me*} \cdot \tilde{D}_G + \gamma^{me}, \tag{6}$$

where $me(G)$ is methylation for any of $g \subseteq G$, $\tilde{D}_G = [\tilde{d}_1(g1), \ldots, \tilde{d}_P(g1), \tilde{d}_1(g2), \ldots, \tilde{d}_P(g2), \ldots]$ are target dysregulation scores across all targets and patients, and γ^{me} is a random intercept.

O.4. Mutation local model. We tested if mutations in any of TFs affect the dysregulation of a target gene. For that, we compared the dysregulation score of a target gene between patients without mutations in any TFs to and patients where at least one TF was mutated. We compared the two distributions using the Mann-Whitney U test if there were at least 7 patients with mutated TFs. P-values were then corrected using the Benjamini-Hochberg method.

O.5. Mutation global model. To evaluate the global relationship between TF mutations and targets dysregulation, we used a linear mixed effect model with a random intercept coefficient for each target assuming different baseline mutation load for each TFs-target complex:

$$mu(t_i) = \beta_0^{mu} + \beta_1^{mu} \cdot \tilde{D}_{gi} + \gamma_i^{mu} \tag{7}$$

where $mu(t_i)$ indicates if any of TFs that regulate g_i are mutated, $\tilde{D}_{gi}$ are target dysregulation scores, γ^{me} is a random intercept for every TFs-target complex. β_0^{mu} and β_1^{mu} were estimated using linear mixed effect model and then β_1^{mu} was tested for significance of association with a null hypothesis $H_0 : \beta_1^{mu} = 0$.

P. Benchmark analysis. To evaluate the DysRegNet method in comparison to other sample-specific methods, we performed a comparison with two methods: LIONESS (Linear Interpolation to Obtain Network Estimates for Single Samples) [10], and SSN (Single Sample Network) [11]. All tools for the benchmark were implemented in Python (code is available at https://github.com/biomedbigdata/DysRegNet). Both methods require a template network, i.e., a set of edges that will be tested for dysregulation. We used GENIE3 cancer-specific template network and GENIE3 shared template network (see subsection M) for both methods. Additionally, we employed STRING network with confidence scores higher than 0.9 as it was used by Liu et al. [11] for the SSN method.

P.1. SSN. SSN described by Liu et al. [11] is based on differences in correlations introduced by the addition of one case sample to a set of control samples. For each case sample and each pair of genes, the following score is computed:

$$Z = \frac{\Delta PCC_n}{(1 - PCC_n^2)/(n-1)} \tag{8}$$

, where $\Delta PCC_n = PCC_{n+1} - PCC_n$ is the difference in Pearson correlation coefficients between two genes in control samples (PCC_n) and control samples with one case sample(PCC_{n+1}), n is the number of control samples. According to Liu et al., the Z value can be further converted to a p-value based on normal distribution probability density function. Next, p-values were corrected for multiple hypothesis testing. We set a cut-off that defines a dysregulated edge at 0.005 significance level in order to have a comparable number of edges for all the methods.

P.2. LIONESS. LIONESS, just like SSN, is a correlation-based method, but instead of investigating the influence of the addition of a case sample to a set of controls, the LIONESS strategy is to investigate the withdrawal of one case sample from all other case samples. Formally, for each patient and each pair of genes, LIONESS performs the following scoring:

$$s = m(PCC^m - PCC^{m-1}) + PCC^{m-1} \tag{9}$$

, where PCC^m is the correlation between two genes based on m case samples and PCC^{m-1} is the correlation between two genes when one sample is withdrawn, m is the number of case samples.
LIONESS scores cannot be converted to p-values; therefore, we set a cut-off such that top 1% of the most highly dysregulated edges are preserved in each control sample.

Q. Hypothesis testing for cancer stage analysis. We separated the samples of THCA cancer into two groups: the early stage group (287 tumor samples annotated with stage I) and the advanced group (170 samples annotated with either stage III or stage IV). For a better separability, we excluded stage II samples from this analysis since they pose an intermediate setting that can resembles either stage I or III too closely. For every TF in the GENIE3 inferred network for THCA, we computed the dysregulated score (or

the normalized out-degree), which is the percentage of TF-specific edges that are dysregulated. A score of 1 means that all interactions of the TF are dysregulated in the sample, while 0 means that none of the interactions are affected. We then compare the dysregulated scores for the two patient groups with one one-sided Mann–Whitney U test and corrected for multiple testing using the Benjamini/Hochberg method. To test if some dysregulated edges are more likely in advanced stages, we performed a one-sided Fisher exact test, and further corrected for multiple testing in a similar fashion. Cell type enrichment analysis for TCGA data were pre-computed using xCell [29] and directly obtained from (https://xcell.ucsf.edu/). We tested for the difference between the two groups defined above using two-sided Mann–Whitney U test and corrected for multiple testing using the Benjamini/Hochberg method.

Q.1. Support vector machine classification. To train SVM classification models, we used a 5 fold nested cross-validation for hyper-parameter selection. All models were trained with the same hyper-parameter search space using the grid search implementation in *Scikit-learn* [50]. The search space includes the kernel type (either linear, polynomial, rbf or sigmoid), the regularization parameter L2 and the kernel coefficient. Since the number of edges considered in the GRN is very large compared to the number of samples available, and to reduce the high-dimensional feature space for the edge level model, we filtered out edges that are dysregulated in only 10 patients or less. The TF level model was trained using the dysregulated score, as described above. The gene expression model was trained using z-score normalized TPM values.

Q.2. Graph Neural network. The GNN architecture consisted of 2 graph convolutional layers, a mean pooling layer, and two fully connected layers. After each of the convolutional layers, we used batch normalization to improve stability. Each graph node (gene) had five features corresponding to gene sequence low-dimensional representation. The representation was obtained by making a k-mer (k = 4) frequency matrix over all gene sequences and then applying UMAP to get a 5-dimensional representation of each gene. The motivation behind this step was to indicate locations of identical genes across patient graphs. A Google collab notebook is available to reproduce our results using PyTorch and Captum (https://colab.research.google.com/drive/1La0NVGZjqIq_1T5EhulU9tHPNQ_o1ek0?usp=sharing).

R. Web interface. The results can be interactively explored using a web interface (https://exbio.wzw.tum.de/dysregnet), which is build with Plotly Dash v2.0.0 (https://plotly.com/dash/), the Cytoscape.js [51] wrapper Dash Cytoscape v0.2.0 (https://dash.plotly.com/cytoscape), Dash Bio v0.8.0 (https://dash.plotly.com/dash-bio) and a Neo4j v3.5.3 database (https://neo4j.com/). Since the underlying network is vast and highly connected,

the interface is centered around individually selected query genes. Only the regulatory connections between those genes and their targets or sources are displayed to keep the resulting network compact and tidy. Further query genes can be added to expand the graph in directions of interest.

We display the fraction of patients with a dysregulation for each regulatory connection, which is directly depicted by the corresponding edge in the graph network. This metric can also be compared visually between different cancer types. Furthermore, the web interface incorporates information about the gene mutation frequency and mean promoter DNA methylation. Heatmaps allow the investigation of the DNA methylation status and the significance of a dysregulation on the patient level.

To prevent the underlying graph structure from becoming too large, the maximum number of displayed regulations is capped, and the regulations can be filtered by their fraction of dysregulated patients and their type. In case a user is interested in the full, unfiltered graph, it can be downloaded as a CSV file.

Acknowledgements

Conflict of interest

No conflicts of interests are declared.

Data Availability

The datasets and computer code produced in this study are available in the following databases:

- RNA-Seq data: XENA browser ((https://xena.ucsc.edu/)
- PPI network: STRING database (http://string-db.org/)

- Experimentally validated regulatory network: Transcriptional Regulation Interactions database (http://www.lbbc.ibb.unesp.br/htri)

- Pyhton package: GitHub (https://github.com/biomedbigdata/DysRegNet_package)

- Modeling computer scripts: GitHub (https://github.com/biomedbigdata/DysRegNet).

References

1. Lian En Chai, Swee Kuan Loh, Swee Thing Low, Mohd Saberi Mohamad, Safaai Deris, and Zalmiyah Zakaria. A review on the computational approaches for gene regulatory network construction. *Computers in biology and medicine*, 48:55–65, 2014.

2. Vân Anh Huynh-Thu, Alexandre Irrthum, Louis Wehenkel, and Pierre Geurts. Inferring regulatory networks from expression data using tree-based methods. *PloS one*, 5(9):e12776, 2010.

3. Adam A Margolin, Ilya Nemenman, Katia Basso, Chris Wiggins, Gustavo Stolovitzky, Riccardo Dalla Favera, and Andrea Califano. Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. In *BMC bioinformatics*, volume 7, pages 1–15. Springer, 2006.

4. Dharmesh D Bhuva, Joseph Cursons, Gordon K Smyth, and Melissa J Davis. Differential co-expression-based detection of conditional relationships in transcriptional data: comparative analysis and application to breast cancer. *Genome biology*, 20(1):1–21, 2019.

5. Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.

6. Beryl B Cummings, Jamie L Marshall, Taru Tukiainen, Monkol Lek, Sandra Donkervoort, A Reghan Foley, Veronique Bolduc, Leigh B Waddell, Sarah A Sandaradura, Gina L O'Grady, et al. Improving genetic diagnosis in mendelian disease with transcriptome sequencing. *Science translational medicine*, 9(386):eaal5209, 2017.

7. Laura S Kremer, Daniel M Bader, Christian Mertes, Robert Kopajtich, Garwin Pichler, Arcangela Iuso, Tobias B Haack, Elisabeth Graf, Thomas Schwarzmayr, Caterina Terrile, et al. Genetic diagnosis of mendelian disorders via rna sequencing. *Nature communications*, 8(1):1–11, 2017.

8. Byungkyu Park, Wook Lee, Inhee Park, and Kyungsook Han. Finding prognostic gene pairs for cancer from patient-specific gene networks. *BMC medical genomics*, 12(8):1–14, 2019.

9. Wook Lee, De-Shuang Huang, and Kyungsook Han. Constructing cancer patient-specific and group-specific gene networks with multi-omics data. *BMC medical genomics*, 13(6):1–12, 2020.

10. Marieke Lydia Kuijjer, Matthew George Tung, GuoCheng Yuan, John Quackenbush, and Kimberly Glass. Estimating sample-specific regulatory networks. *Iscience*, 14:226–240, 2019.

11. Xiaoping Liu, Yuetong Wang, Hongbin Ji, Kazuyuki Aihara, and Luonan Chen. Personalized characterization of diseases using sample-specific networks. *Nucleic acids research*, 44(22):e164–e164, 2016.

12. Mai Adachi Nakazawa, Yoshinori Tamada, Yoshihisa Tanaka, Marie Ikeguchi, Kako Higashihara, and Yasushi Okuno. Novel cancer subtyping method based on patient-specific gene regulatory network. *Scientific Reports*, 11(1):1–11, 2021.

13. Luiz A Bovolenta, Marcio L Acencio, and Ney Lemke. Htridb: an open-access database for experimentally verified human transcriptional regulation interactions. *BMC genomics*, 13(1):1–10, 2012.

14. Lars J Jensen, Michael Kuhn, Manuel Stark, Samuel Chaffron, Chris Creevey, Jean Muller, Tobias Doerks, Philippe Julien, Alexander Roth, Milan Simonovic, et al. String 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic acids research*, 37(suppl_1):D412–D416, 2009.

15. Theresa Phillips et al. The role of methylation in gene expression. *Nature Education*, 1(1):116, 2008.

16. Emmanouil Bouras, Meropi Karakioulaki, Konstantinos I Bougioukas, Michalis Aivaliotis, Georgios Tzimagiorgis, and Michael Chourdakis. Gene promoter methylation and cancer: An umbrella review. *Gene*, 710:333–340, 2019.

17. Venkateshwar G Keshamouni. Excavation of fosl1 in the ruins of kras-driven lung cancer, 2018.

18. Xiaodong Sun, Henry F Frierson, Ceshi Chen, Changling Li, Qimei Ran, Kristen B Otto, Brandi M Cantarel, Robert L Vessella, Allen C Gao, John Petros, et al. Frequent somatic mutations of the transcription factor atbf1 in human prostate cancer. *Nature genetics*, 37(4):407–412, 2005.

19. Motoki Takaku, Sara A Grimm, Bony De Kumar, Brian D Bennett, and Paul A Wade. Cancer-specific mutation of gata3 disrupts the transcriptional regulatory network governed by estrogen receptor alpha, foxa1 and gata3. *Nucleic acids research*, 48(9):4756–4768, 2020.

20. John H Bushweller. Targeting transcription factors in cancer—from undruggable to reality. *Nature Reviews Cancer*, 19(11):611–624, 2019.

21. Xuan Meng, Nicole R Tackmann, Shijie Liu, Jing Yang, Jiahong Dong, Congying Wu, Adrienne D Cox, and Yanping Zhang. Rpl23 links oncogenic ras signaling to p53-mediated tumor suppression. *Cancer research*, 76(17):5030–5039, 2016.

22. James M Dolezal, Arie P Dash, and Edward V Prochownik. Diagnostic and prognostic implications of ribosomal protein transcript expression patterns in human cancers. *BMC cancer*, 18(1):1–14, 2018.

23. Mu-Shui Dai and Hua Lu. Inhibition of mdm2-mediated p53 ubiquitination and degradation by ribosomal protein l5. *Journal of Biological Chemistry*, 279(43):44475–44482, 2004.

24. Harold W Kuhn. The hungarian method for the assignment problem. *Naval research logistics quarterly*, 2(1-2):83–97, 1955.

25. T Zhang, Y Huang, W Liu, W Meng, H Zhao, Q Yang, SJ Gu, CC Xiao, CC Jia, B Zhang, et al. Overexpression of zinc finger protein 687 enhances tumorigenic capability and promotes recurrence of hepatocellular carcinoma. *Oncogenesis*, 6(7):e363–e363, 2017.

26. F Martin, S Ladoire, G Mignot, L Apetoh, and F Ghiringhelli. Human foxp3 and cancer. *Oncogene*, 29(29):4121–4129, 2010.

27. Michele Mondini, Silvia Costa, Simone Sponza, Francesca Gugliesi, Marisa Gariglio, and Santo Landolfo. The interferon-inducible hin-200 gene family in apoptosis and inflammation: implication for autoimmunity. *Autoimmunity*, 43(3):226–231, 2010.

28. Hirohito Yamaguchi, Yi Ding, Jin-Fong Lee, Ming Zhang, Ashutosh Pal, William Bornmann, Duen-Hwa Yan, and Mien-Chie Hung. Interferon-inducible protein ifixα inhibits cell invasion by upregulating the metastasis suppressor maspin. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, 47(10):739–743, 2008.

29. Dvir Aran, Zicheng Hu, and Atul J Butte. xcell: digitally portraying the tissue cellular heterogeneity landscape. *Genome biology*, 18(1):1–14, 2017.

30. Narine Kokhlikyan, Vivek Miglani, Miguel Martin, Edward Wang, Bilal Alsallakh, Jonathan Reynolds, Alexander Melnikov, Natalia Kliushkina, Carlos Araya, Siqi Yan, and Orion Reblitz-Richardson. Captum: A unified and generic model interpretability library for pytorch, 2020.

31. Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, pages 3319–3328. PMLR, 2017.

32. Qiuxian Zheng, Qiang Fu, Jia Xu, Xinyu Gu, Haibo Zhou, and Chen Zhi. Transcription factor e2f4 is an indicator of poor prognosis and is related to immune infiltration in hepatocellular carcinoma. *Journal of Cancer*, 12(6):1792, 2021.

33. Sari S Khaleel, Erik H Andrews, Matthew Ung, James DiRenzo, and Chao Cheng. E2f4 regulatory program predicts patient survival prognosis in breast cancer. *Breast Cancer Research*, 16(6):1–14, 2014.

34. Kenneth MK Mark, Frederick S Varn, Matthew H Ung, Feng Qian, and Chao Cheng. The e2f4 prognostic signature predicts pathological response to neoadjuvant chemotherapy in breast cancer patients. *BMC cancer*, 17(1):1–11, 2017.

35. A Waghray, M Schober, F Feroze, F Yao, J Virgin, and YQ Chen. Identification of differentially expressed genes by serial analysis of gene expression in human prostate cancer. *Cancer research*, 61(10):4283–4286, 2001.

36. Dawei Wang, Jamie Russell, Hui Xu, and David G Johnson. Deregulated expression of dp1 induces epidermal proliferation and enhances skin carcinogenesis. *Molecular Carcinogenesis: Published in cooperation with the University of Texas MD Anderson Cancer Center*, 31(2):90–100, 2001.

37. Jenny Hsu and Julien Sage. Novel functions for the transcription factor e2f4 in development and disease. *Cell Cycle*, 15(23):3183–3190, 2016.

38. Adam B Zaborowski and Dirk Walther. Determinants of correlated expression of transcription factors and their target genes. *Nucleic acids research*, 48(20):11347–11369, 2020.

39. John Lonsdale, Jeffrey Thomas, Mike Salvatore, Rebecca Phillips, Edmund Lo, Saboor Shad, Richard Hasz, Gary Walters, Fernando Garcia, Nancy Young, et al. The genotype-tissue expression (gtex) project. *Nature genetics*, 45(6):580–585, 2013.

40. Ezio Laconi, Fabio Marongiu, and James DeGregori. Cancer as a disease of old age: changing mutational and microenvironmental landscapes. *British journal of cancer*, 122(7):943–952, 2020.

41. Berna C Özdemir and Gian-Paolo Dotto. Racial differences in cancer susceptibility and survival: more than the color of the skin? *Trends in cancer*, 3(3):181–197, 2017.

42. Hae-In Kim, Hyesol Lim, and Aree Moon. Sex differences in cancer: epidemiology, genetics and therapy. *Biomolecules & therapeutics*, 26(4):335, 2018.

43. João Pedro de Magalhães. Every gene can (and possibly will) be associated with cancer. *Trends in Genetics*, 2021.

44. Ralph B d'Agostino. An omnibus test of normality for moderate and large size samples. *Biometrika*, 58(2):341–348, 1971.

45. RALPH D'AGOSTINO and Egon S Pearson. Tests for departure from normality. empirical results for the distributions of b 2 and b. *Biometrika*, 60(3):613–622, 1973.

46. Héctor Climente-González, Eduard Porta-Pardo, Adam Godzik, and Eduardo Eyras. The functional impact of alternative splicing in cancer. *Cell reports*, 20(9):2215–2226, 2017.

47. Zakaria Louadi, Kevin Yuan, Alexander Gress, Olga Tsoy, Olga V Kalinina, Jan Baumbach, Tim Kacprowski, and Markus List. Digger: exploring the functional role of alternative splicing in protein interactions. *Nucleic acids research*, 49(D1):D309–D318, 2021.

48. Mary J Goldman, Brian Craft, Mim Hastie, Kristupas Repečka, Fran McDade, Akhil Kamath, Ayan Banerjee, Yunhai Luo, Dave Rogers, Angela N Brooks, et al. Visualizing and interpreting cancer genomics data via the xena platform. *Nature biotechnology*, 38(6):675–678, 2020.

49. Samuel A Lambert, Arttu Jolma, Laura F Campitelli, Pratyush K Das, Yimeng Yin, Mihai Albu, Xiaoting Chen, Jussi Taipale, Timothy R Hughes, and Matthew T Weirauch. The human transcription factors. *Cell*, 172(4):650–665, 2018.

50. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

51. Max Franz, Christian T. Lopes, Gerardo Huck, Yue Dong, Onur Sumer, and Gary D. Bader. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics*, 32(2):309–311, 09 2015.

Supplementary Note 1: Performance comparison for same organ cancers

Stratification of different cancer types can be challenging to interpret due to vast differences between them. One might argue that performance of DisRegNet can be explained by tissue/organ specific differences while other methods are less biased. We want to demonstrate results of patients stratification in a homogeneous setting where potential tissue specific differences are minimized. The figure below (Figure 1) demonstrates superior performance of DysRegNet in cancer subtypes stratification.
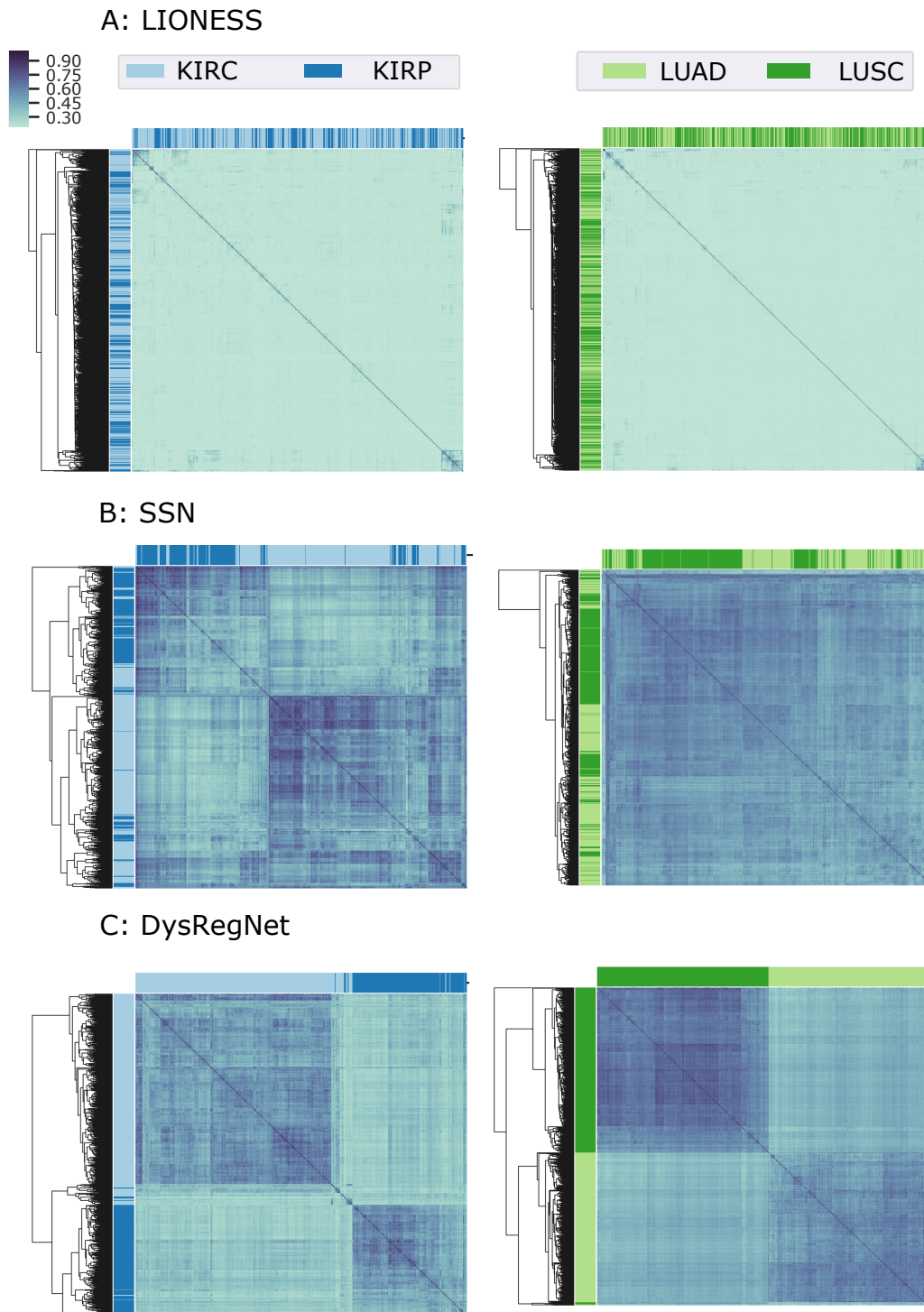
Supplementary Figure 1. Performance comparison for same-organ cancers

Supplementary Note 2: Dysregulation scenarios
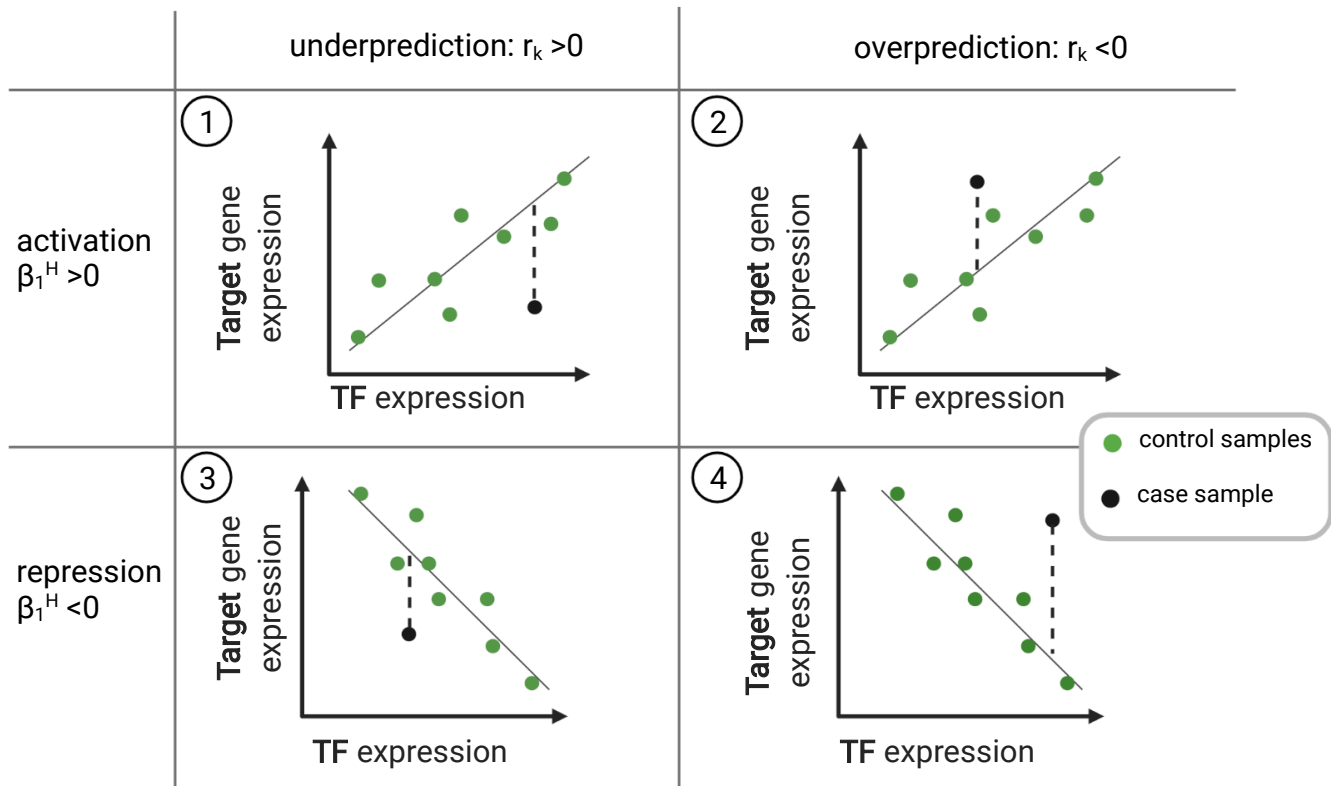
Supplementary Figure 2. Different scenarios of dysregulation

Supplementary Note 3: Global versus local model explanation
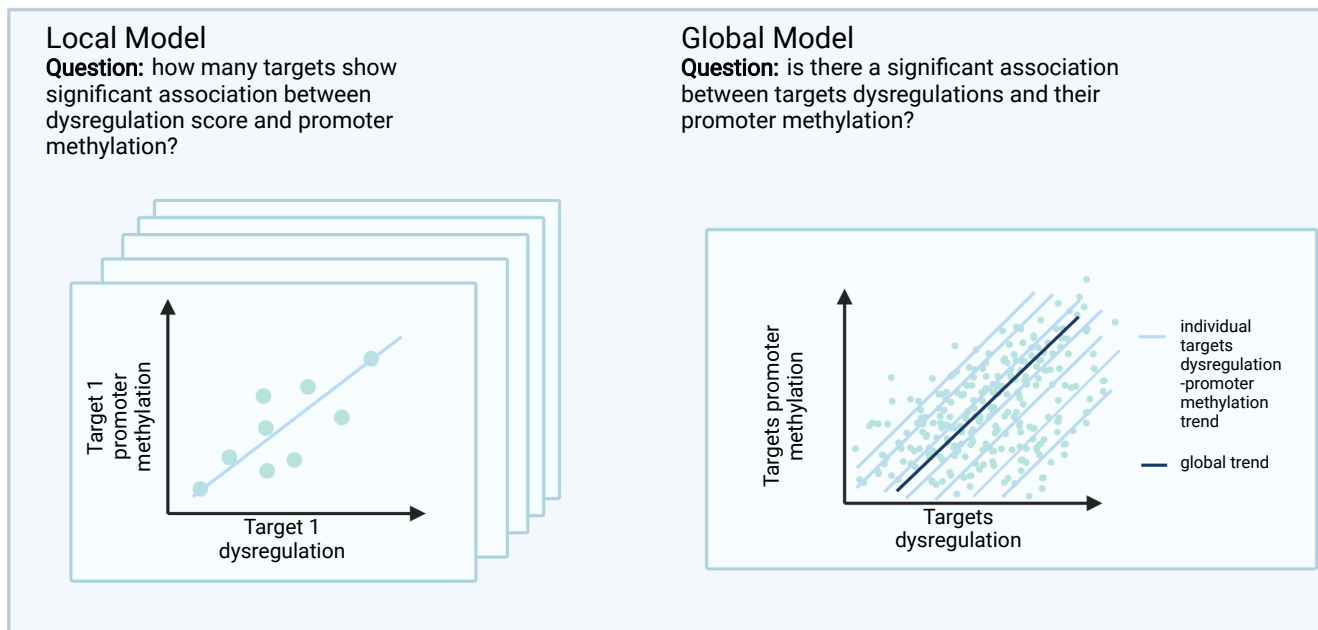
Supplementary Figure 3. Local and global models for methylation-dysregulation association studies.
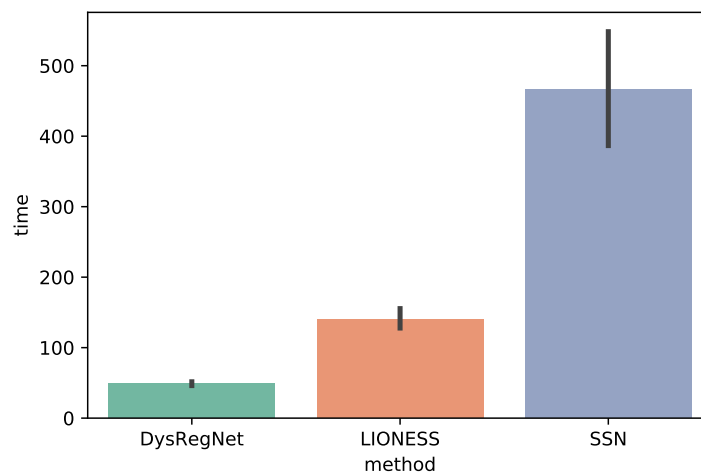
Supplementary Note 4: Run-time comparison

Run-time analysis was performed on Linux machine with 160 cores. All methods were implemented in Python 3.9. While LIONESS has an R implementation, we did not use it maintain our pipeline completely in python. Thus, our run-time estimation might not be completely accurate with respect to LIONESS.

If we consider an input gene expression matrix $X \in \mathbb{R}^{n \times m}$, where n is a number of genes and m is a number of patients then complexity of LIONESS is $\mathcal{O}(nm^3)$. This is a rough estimate assuming that correlation matrix can be obtained with complexity $\mathcal{O}(nm^2)$ and this procedure should be repeated for every patient.

SSN also computes correlation matrix for every patient, but the matrix itself is computed based on controls samples (m_* control samples). Additionally, for every patient, SSN requires to raise the matrix to a power of 2. Thus, SSN complexity can be estimated as $\mathcal{O}(nm_*^2)$ to obtain the correlation matrix and then $\mathcal{O}(n^3)$ for powering it. Repeating this procedure for every patient the final complexity of SSN is $\mathcal{O}(m(nm_*^2 + n^3) \sim \mathcal{O}(mn^3)$

DysRegNet relies on ordinary least squares model, where the number of features is equal to 2 (genes) + several covariates (c). Thus, the complexity is $\mathcal{O}(c^2 m)$. Given that the number of patients dominates the number of covariates (only 5 in our case), the expression can be simplified to $\mathcal{O}(m)$. Since this procedure needs to be repeated for every edge set (at most 100 000 in our case), the final asymptotic complexity is $\mathcal{O}(100000 \times m)$.

In the conducted experiment we used data with 50 case samples, 50 control samples, 10 000 potential regulatory edges and 6689 genes. Average run-time of DysRegNet was 49 seconds, LIONESS performed in 140 seconds and SSN in 466 seconds (Figure 4). Given the estimated algorithm complexity, larger number of patients will lead to even bigger difference in run-time between DysGeNet and other methods.

**Supplementary Figure 4.** Run-time comparison (in seconds) for 10 runs of every method

## Supplementary Note 5: Mapping of TCGA cancer abbreviations to DisGeNet terms
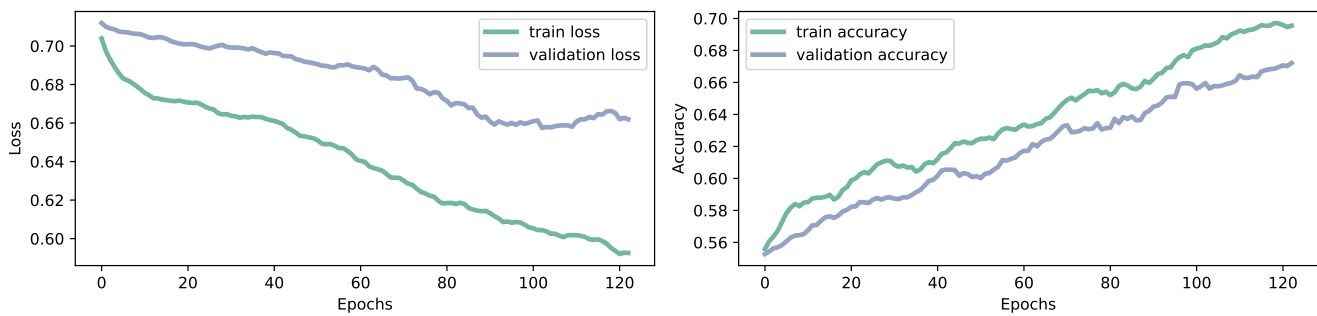
Supplementary Figure 5. Performance of GNN during training.

**Supplementary Table 1.** Mapping of TCGA cancers to closest DisGeNet terms

Abbreviation	Full name	DisGeNet term
LIHC	Liver hepatocellular carcinoma	Liver carcinoma (C2239176)
THCA	Thyroid carcinoma	Thyroid carcinoma (C0549473)
STAD	Stomach adenocarcinoma	Stomach Carcinoma (C0699791)
BRCA	Breast invasive carcinoma	Breast Carcinoma (C0678222)
COAD	Colon adenocarcinoma	Adenocarcinoma of colon (C0338106)
LUSC	Lung squamous cell carcinoma	Squamous cell carcinoma of lung (C0149782)
PRAD	Prostate adenocarcinoma	Adenocarcinoma of prostate (C0007112)
HNSC	Head & neck squamous cell carcinoma	Squamous cell carcinoma of the head and neck (C0007112)
KIRC	Kidney clear cell carcinoma	Malignant neoplasm of kidney (C0740457)
KIPR	Kidney papillary cell carcinoma	Malignant neoplasm of kidney (C0740457)
LUAD	Lung adenocarcinoma	Adenocarcinoma of lung (C0152013)
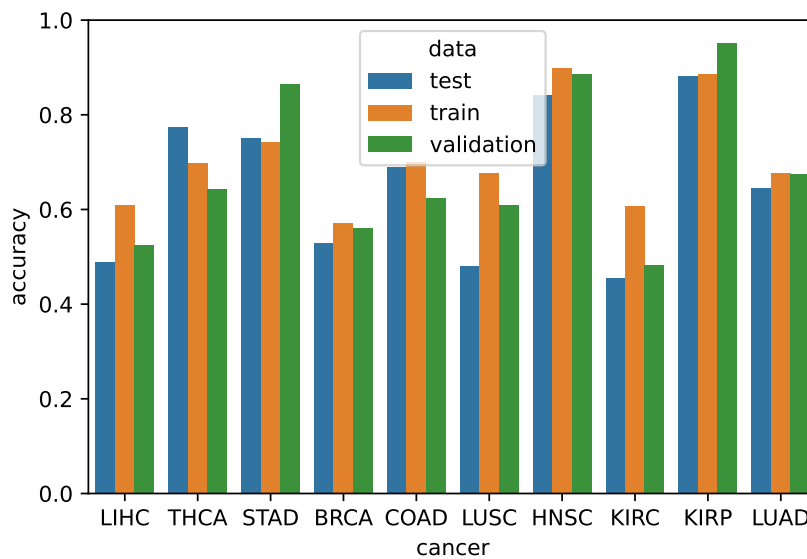
Supplementary Figure 6. Final GNN accuracy for different cancer types.
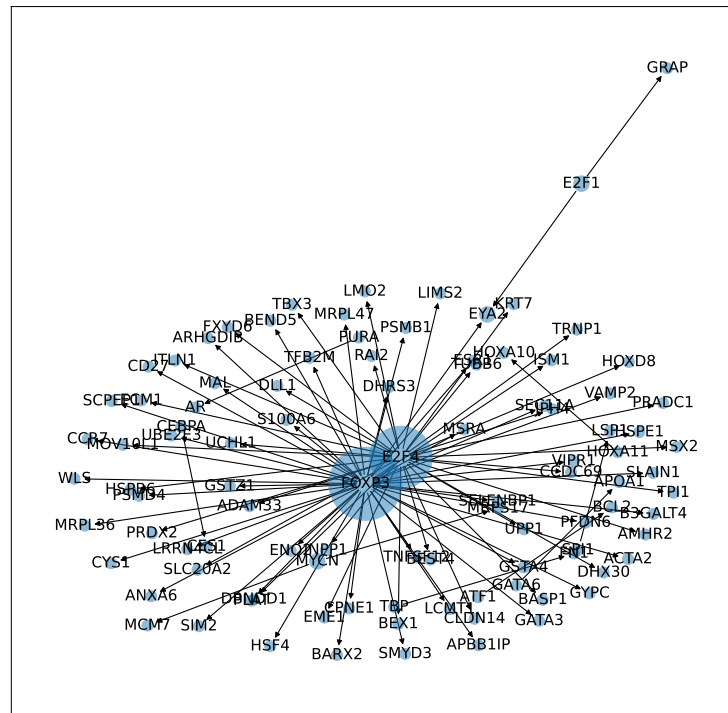
Supplementary Figure 7. Aggregated dysregulatory network that is the most predictive of advance stages across cancer types
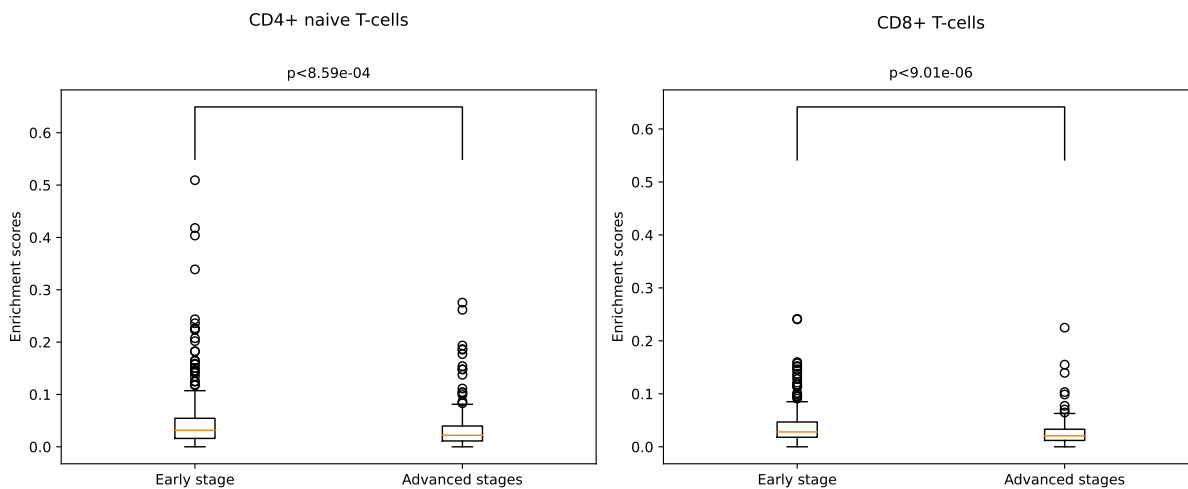
Supplementary Figure 8. CD4+ naive T-cells and CD8+ T-cells composition in early and advanced THCA cancer patients. No big effect, is observed except for few outliers.