SEMMELWEIS EGYETEM

DOKTORI ISKOLA

Ph.D. értekezések

3480.

SCHULC KLÁRA

Patobiokémia

című program

Programvezető: Dr. Csala Miklós, egyetemi tanár

Témavezető: Dr. Csermely Péter, egyetemi tanár

Dr. Veres Dániel, PhD

NETWORK BIOLOGY APPROACHES TOWARDS THE IDENTIFICATION OF PREDICTIVE CANCER BIOMARKERS

PhD thesis

Klára Schulc, M.D.

Molecular Medicine Doctoral School

Semmelweis University

Supervisors:	Péter Csermely, Ph.D., D.Sc. Dániel Veres, M.D., Ph.D.
Official reviewers:	Bence Czakó, Ph.D. János Tibor Fekete, Ph.D.
Head of the Complex Examination Committee:	Csaba Bödör, Ph.D., Habil.
Members of the Complex Examination Committee:	Gábor Turu, Ph.D. Máté Szalay-Bekő, Ph.D.

Budapest
2025

Table of Contents

List of Abbreviations

AUC: Area Under Curve

BPS: Biomarker Probability Score

CSN: Human Cancer Signalling Network

DMEM: Dulbecco's Modified Eagle's Medium

DMSO: dimethyl sulfoxide

F1-score: Harmonic mean of the equally weighted precision and recall

FBS: Foetal Bovine Serum

GO: Gene Ontology Consortium

HCC: Hepatocellular Carcinoma

IDP: intrinsically disordered protein

lncRNA: long non-coding ribonucleic acid

LOOCV: Leave-One-Out-Cross-Validation

MAPK: mitogen-activated protein kinase

ML: machine learning

MPAS: MAPK Pathway Activity Score

NSCLC: Non-Small Cell Lung Cancer

RF: Random Forest algorithm

RNA: ribonucleic acid

RNAseq: RNA sequencing

RT-qPCR: Real-Time Quantitative Polymerase Chain Reaction

s*iRNA:* small interfering RNA

TCGA: The Cancer Genome Atlas

XGB: XGBoost algorithm

1. Introduction

"Down to their innate molecular core, cancer cells are hyperactive, survival-endowed, scrappy, fecund, inventive copies of ourselves."

— Siddhartha Mukherjee, The Emperor of All Maladies: A Biography of Cancer

1.1. The development of cancer

Cancer is a disease emerging from cumulated genetic changes leading to sustained proliferation. Besides proliferation, other common properties of cancer cells and the tumour microenvironment have been observed, categorised as the hallmarks of cancer [1]. These include angiogenesis, escaping the immune system, evading apoptosis, invasion and metastasis, all contributing to sustain the growth of cancers. These phenotypic changes through dysregulated signalling pathways effected by genetic events and environmental factors, may lead to somatic evolution and cell state changes [2]. The dysregulation can be at different regulatory levels, such as genetic changes, epigenetic silencing or enhancements, overexpression, and posttranslational regulations effecting protein translocation or degradation [3]. As we have more information about these complex changes, understanding may require system-level tools to understand the changes in key signalling pathways of cancer cells [4].

One core example is the MAPK (mitogen activated protein kinase) pathway, a conserved signalling cascade where extracellular mitogens activate RAS, leading to subsequent phosphorylation of RAF, MEK, and ERK proteins to regulate cell proliferation, survival, and differentiation through downstream effectors [5]. Dysregulation of this pathway often sets the stage for malignant transition, causing excessive proliferation. Consequently, there are approved drugs targeting this pathway, such as the inhibitors of MEK and KRAS, most frequently given in melanoma [6]. Even though MAPK mutations can foreshadow the intrinsic effectivity of these treatments, there are no biomarkers available for longitudinal clinical decision support. Further, expression-level biomarkers would be necessary to monitor the efficacy of these treatments and quickly inform about the drug's emerging infectivity before clinical complications occur.

Even though MAPK pathway inhibitors are not routinely applied in colon cancer, the pathway is still essential in colon carcinogenesis [7]. This path of healthy cells to acquire

the cancerous phenotype is thoroughly researched in colon cancer, including detailed genetic events from healthy colon through colon adenoma to colon adenocarcinoma through years or even decades. The classic normal–adenoma–carcinoma sequence is governed by stepwise accumulation of driver mutations, beginning with the loss of APC or other Wnt‑pathway disruption in normal colon cells [8]. Changes in cancer cell signalling come alongside with changes in the tumour microenvironment and cell-cell interactions [9]. This process initiates small adenomas, while further activating KRAS mutations in the adenoma cells promote growth into larger, dysplastic adenomas, as constitutively active KRAS induces the MAPK pathway. Finally, alterations such as TP53 and SMAD4 inactivation and chromosomal instability enable malignant transformation into invasive carcinoma, requiring not only surgical but oncological intervention [10].

Even though the mutation sequence seems like a clear-cut process, these single mutations might lead to global network signalling changes underscoring the system-level concept of carcinogenesis. Previously, system biology approaches were made to understand differences between normal colon and colon carcinoma [11], and even from adenoma to carcinoma [12]. However, comprehensive network modelling of colon carcinogenesis involving colon adenoma as an important tumorous but not malignant stage of cancer development was not yet made. Adding colon adenoma to a 3-step model of colon carcinogenesis can potentially highlight what global network changes are truly happening between the normal-adenoma and the adenoma-carcinoma development [13].

1.2. Therapy resistance

Therapy resistance is one of the key challenges in medical oncology today, causing ineffective treatments and unnecessary toxicity, while conferring a financial burden on the medical system [14]. This resistance can be the internal attribute of the given patient or cancer type, i.e., intrinsic resistance, or acquired after starting the treatment, i.e., acquired resistance. Intrinsic resistance may be caused by pre-existing mutations or signalling patterns, or the pharmacokinetics of the drug, such as fast metabolism or inefficient transfer of the drug [15]. Various underlying mechanisms leading to acquired drug resistance were discovered, from newly acquired mutations to drug efflux pumps and microenvironmental changes to activated cross-talks in signalling pathways redirecting information flow [16]. Regardless of the details of the resistance mechanism,

the consequent changes affect the whole cellular signalling machinery [17]. Detecting these global signalling changes or their proxies, i.e., key resistance-indicating proteins in time can lead to early detection of resistance, sparing the patient from clinically significant progression and unnecessary toxicities.

"Forgetting", i.e., rearrangement of cellular networks is a framework able to describe therapy resistance. As it was recently proposed, the cell 'forgets' drug-related pathways during drug adaptation and re-learns new ways of signalling [18]. Intrinsically disordered proteins (IDPs) participate in the treatment adaptation process with their transitional conformational changes. However, as discussed above, adaptation to a drug treatment can take many forms, and it is not always trivial which pathways and cross-talks the cell 'learns'. Thus, it is important to identify key markers of different adaptation styles, as they can help to predict treatment responses. One famous example is breast cancer, where it is a frequent treatment adaptation strategy to shift between subtypes and change the expression of hormone receptors and HER2 [19]. This change comes with clinical consequences, as the tumour usually loses sensitivity to the previous treatment and may require another one. Motivated by this well-known example, defining and describing further adaptation states with preferably sensitive and specific markers is a key goal in current cancer research. However, the identification of these markers is not always straightforward, setting the stage to system-level analyses. The question is how we can detect the cellular forgetting process and what therapy changes we can implement to tackle the adaptation by rather causing the "forgetting", i.e., circumventing the unwanted signalling processes [20].

1.3. Biomarkers in cancer therapy

One of the possible ways to tackle therapy resistance is identifying predictive biomarkers and signatures indicating intrinsic or acquired therapy resistance in time, to avoid or quit ineffective treatments as soon as possible [21]. As these decisions deeply affect patient survival, the biomarker must be trusted by the patient and the medical personnel for clinical decision-making. Thus, a good predictive biomarker should be sensitive, specific, and preferably easily and cheaply detectable. Biomarkers can be selected from all kinds of biological entities, from DNA mutations through RNA and protein expression to metabolic changes. Protein immunohistochemistry and gene mutation analysis are the

most often performed [6, 19] and are among the cheapest biomarker diagnostic methods. Other biomarker types such as RNA expression, are less frequently used in clinical settings, however, they are still important for exploratory analyses [22]. Traditional biomarkers can be detected from cancer tissue acquired with invasive biopsies. Less invasive methods such as liquid biopsy allow more frequent testing but often lack sensitivity and cost-effectivity [23]. The selection method of the biomarkers can extend from high-throughput sequencing and bioinformatic screening to smaller laboratory experiments and their tactical, consequent combination to validate the findings [24]. In the end, clinical validation showing the selectiveness, sensitivity and the added value of the biomarker is always necessary before medical use. However, validating biomarkers on patients is costly and comes with some risk. Thus, effective selection and pre-testing of biomarkers is needed with cost-effective tools such as screening with bioinformatic tools or cell lines [25].

1.4. Intrinsically disordered proteins and long non-coding RNAs in cancer

Intrinsically disordered proteins highly contribute to cancer signalling and therapy adaptation. They are essential in cancer signalling due to their flexible structure [26]. Their relaxed conformation introduces conformational noise into cellular networks and promotes cells to switch phenotypes due to genetic and non-genetic effects. This phenotype switching can even be reversible through similar mechanisms, changing between a drug sensitive and resistant state [26]. Their adaptive binding pattern allows them to engage in various interactions, leading to pathological processes. Perturbations, such as therapeutic effects or outside stressors, often result in an increased expression of IDPs. This brings flexibility in the signalling network, allowing it to discover the network search space to find the optimal phenotype. Due to this adaptive potential, they often have central signalling roles in disease networks, including cancer.

From a signalling standpoint, IDPs would make good drug targets due to their regulatory role in networks and importance in cancer. However, due to their changing structure, they have been considered undruggable and targeting them still poses significant challenges today [27]. A few approaches exist to target this, such as targeting the non-disordered regions of IDPs, interactions with other proteins or DNA, or the most hydrophobic parts

of the protein. This can be combined with the screening of large molecular libraries to find effective binding partners or designing partners for their binding surfaces. One example is c-Myc, a central hub in the signalling of aggressive cancers, which has been historically difficult to target [27]. Currently, a few potent c-Myc inhibitors exist, but no drug has been approved yet for clinical use.

Drug targeting is not the only way to exploit the importance of IDPs in cancer signalling. Disordered proteins can generally fit the criteria of a cancer biomarker due to their flexible structure allowing phenotype switching during adaptation. Consequently, they should be screened as biomarkers, especially for treatment response prediction. Cancer testis antigens, often disordered themselves, were already mentioned in the literature as good biomarker candidates, supporting these claims [28]. Testing this hypothesis with a high-throughput screening of intrinsically disordered proteins may help to discover effective biomarkers, while learning more about the role of protein structure in signalling networks.

Long non-coding RNAs (lncRNAs), defined as non-coding RNAs longer than 200 nucleotides, were previously considered 'junk' RNAs, as their function was not identified at the time [29]. These transcripts might have been the subjects of weaker selective pressure, thus providing an opportunity for innovative regulatory functions to emerge. This phenomenon is called constructive neutral evolution, which similarly to other somatic evolution mechanisms, might be present in cancer environments in a smaller scale. Aberrant lncRNA expression has been previously observed in cancer patients, supporting that they are affected by the signalling dysregulation of cancer cells [30]. Evidence also shows that they might be able to regulate oncogene expression by their transcriptional suppression function, as anti-sense transcripts often interfere with their sense pairs, and sense transcripts might be related to chromatin remodelling [29].

LncRNAs are attractive cancer biomarkers because they are often secreted by tumour cells or released by necrosis into bodily fluids. They often sponge protein and miRNA partners to avoid degradation. Many of the secreted lncRNAs are stably present and quantifiable in easily accessible body fluids like peripheral blood with RT-qPCR, enabling minimally invasive and sensitive detection of tumour-associated expression changes. Frequent liquid biopsies for lncRNA biomarker measurement can help early

detection of therapy resistance, and continuous decision support during a treatment course [30]. Even though the role and function of lncRNAs in signalling is not always clear, we hypothesise that with thorough testing, it is possible to identify pan-cancer pathway-specific lncRNAs dysregulated in cancer without knowing their detailed role in regulation. These lncRNAs or lncRNA panels, if detectable in tumour or peripheral blood samples, may be suitable for longitudinal clinical decision support during pathway-targeting cancer treatment. To establish this, the pan-cancer pathway-specificity of lncRNAs must be established in simple models such as cell lines, to later continue further testing in animal experiments and clinical settings with the most successful candidates.

1.5. *In silico* methods to understand cancer signalling

Network topology, i.e., descriptive network science can provide insights into therapy resistance and aid the discovery of network-defined biomarkers *via* analysing static local and global network attributes [17]. Among these, network motifs, i.e., small subnetworks with higher abundance in natural networks compared to random networks, serve important roles in perturbation propagation [31]. Recently, it was shown that co-regulation in network motifs is stronger than a simple connection [32], which may help to more precisely detect the signalling changes with the blockade of the target protein, attributed to sensitive and specific biomarkers. Network centrality values such as betweenness centrality can show the most influential nodes in the network. Bridgeness centrality values show the key signal transductor proteins between modules i.e. functional units of the network [33]. Network topological studies also have shown that IDPs are major contributors to information flow within signalling networks and tend to interact with the most central nodes, creating an interface between these network hubs and the rest of the network [34]. However, more detailed network topological studies about IDPs are not yet conducted. All this calls for the screening of network motifs to test if they indeed contain targets and biomarkers. Other annotations, such as centrality values or protein disorder content might help to narrow down the search field and increase efficiency. This also raises the question: which features and annotations, including motif analysis, centrality values, disorder content values or other biological or network topological annotations lead to the most effective biomarker prediction, i.e., determine

what a good biomarker is. Identifying these key features can possibly help to understand biomarkers from a network scientific perspective.

Machine learning methods can effectively contribute to high-throughput biomarker detection methods by ranking findings and predicting new biomarkers based on previous knowledge, as shown before [35-37]. However, machine learning and other artificial intelligence methods are often viewed as 'black box' methods, where the reasoning behind the decisions is not always clear. Tree-based methods like XGBoost [38] and Random Forest [39] combine effectivity with interpretability by providing the learnt decision tree, which is highly valuable when planning for clinical approval. Feature importance analysis methods like the SHAP package [40] using the Shapley values of game theory to rank the most important features in decision making further helps to see the rationale behind the predictions. Still, it is important to thoroughly select the data used to train the machine learning models to avoid overfitting and misleading predictions [41]. Also, applying machine learning methods such as cross-validation and feature selection to test if there is any overfitting is necessary [42]. Harmonising the results of multiple models may also help to create more robust predictions [43]. It is also important to check that the predictions are logical and medically relevant, to ensure the validity of new predictions before further, expensive experimental and clinical testing. One way is to look up if any of the predicted biomarkers from the neighbour-target pairs have been described as biomarkers to their target pairs before but did not end up in the training set. Confirming that the model predicts actual biomarkers validates the method.

The question, whether applying the above principles to high-throughput predictive biomarker prediction results in an effective, high-throughput tool and a narrowed list of potential biomarkers, still must be validated. It is also a question whether using no cancer-specific experimental or expression data, just the derived network structure and annotations, is enough for prediction and can result in an effective pan-cancer model. We hypothesised and proved that the curated combination of rich network topology data and biological attributes can lead to effective and medically relevant biomarker predictions [44]. This may provide a selection of potentially predictive biomarker proteins ready for further testing or bring us closer to understand the behaviour of predictive cancer biomarkers at a network science level.

2. Objectives

2.1. Overall objective

The main objective of the study was to understand signalling changes in cancer better with the tools of network science, machine learning and experimental biology, while identifying predictive biomarkers for cancer targeted therapeutics.

2.2. Specific objectives

As this PhD research included collaborative work and an international research scholarship, multiple different projects tackling cancer signalling and biomarker detection were conducted with different methodologies. We highlight the value of these methodologies by using examples from the MAPK pathway throughout this thesis, which is a core signalling pathway governing proliferation in cancer cells [45].

2.2.1. Network topology changes in cancer development

For the colon carcinogenesis modelling project, the goal was to assess system-level changes during transition from healthy colon through colon adenoma to colon carcinoma from a network topology standpoint.

2.2.2. Network-based tool to predict cancer biomarkers

The goal of the MarkerPredict project was to create a network-based biomarker prediction tool using network topology data and biological attributes, including IDP annotations for targeted cancer therapies. The goal was also to assess the following hypotheses: a) network motif members are more likely to be biomarkers of their neighbours b) IDP proteins are more likely to be biomarkers.

2.2.3. LncRNA biomarker signature for MAPK inhibitor cancer therapies

The goal of the liquid biopsy lncRNA project was to find a pan-cancer lncRNA biomarker signature for MAPK pathway activity to detect MAPK inhibitor therapy effectiveness.

3. Methods

In this section, we describe the various commonly used and self-developed methods used during the various projects mentioned in this thesis. Many of these methods were published or presented by the candidate and her collaborators [13, 44, 46].

3.1. Network construction and topology analysis

3.1.1. Network construction

A network's definition relies on its nodes and edges [47]. In case of cell biology networks, they are often proteins and their connections, established based on previous experiments. For protein-protein interaction networks, the interaction of two proteins is enough, while for signalling networks, only regulatory connections are included [48]. Since there are many curated open access network databases published, this research mainly customises these for downstream analysis instead of *de novo* network construction. The Human Cancer Signalling Network (CSN) only contains genes frequently mutated in cancer [49]. It contains both signalling regulations and protein interactions, so the latter were not included for signalling motif analysis and biomarker identification. Likewise, the signed subnetworks (i.e., positive and negative edges) of the larger SIGNOR [50] and ReactomeFI [51] were selected as well to enhance the pool of possible new biomarkers. The global topology parameters of these networks differed, and this research aimed to exploit this diversity to enable robust biomarker prediction that do not depend on the specific network structure (see Table 1). Network density was calculated with the following equation:

(1)

$$\eta = \frac{2|E|}{|V|(|V|-1)}$$

where η is network density, $|V|$ is the number of vertices i.e. nodes, and $|E|$ is the number of edges in the graph. Triangle motifs were identified with the FANMOD [52] program (see *Section 3.1.4*). Node identification was unified between the networks using UniProt IDs from the UniProt [53] database.

Table 1. The global topological attributes of the three signed subnetworks used for motif-based biomarker prediction. *These three models have differing number of nodes and edges, density (highlighted as percentage values) and triangles, providing opportunity for network-independent biomarker prediction. This table is in the Supplementary Material in the candidate's open access publication and included here with the permission of the co-authors [44].*

Network	CSN	SIGNOR	ReactomeFI
No. of nodes	*1 229*	*4 486*	*9300*
No. of edges	*3 144*	*10 935*	*128 658*
Network density	*0.417%*	*0.109%*	*0.298%*
No. of triangles	*976*	*2318*	*840 505*

3.1.2. Edge weight calculations

To analyse the network topological changes of colon cancer development, the CSN network was weighted using healthy colon, colon adenoma and colon adenocarcinoma datasets (n=128, 131 and 178, respectively) from the Gene Expression Omnibus database [22, 54-57]. The data contained mainly left-sided colon cancer to avoid heterogeneity; however, all clinical stages were included to show general colon cancer characteristics. These datasets were collected with Borbála Kovács. Then, RNA normalization and median aggregation were conducted on this dataset, conducted by our collaborators Sebestyén Kamp and János Molnár. Edge weights were considered as the logarithmic of the product of RNA abundance values (developed with Borbála Kovács):

(2)

$$Link\ weight_{AB} = \log_{10}\ (Abundance_A x\ Abundance_B)$$

This formula considers the higher chance of protein interaction with higher abundance – as mRNA expression was used as proxy for proteomics in this study. The logarithmic transformation can prevent outlying data points dominating the dataset. Following these steps resulted in a network model for normal colon, colon adenoma and colon adenocarcinoma.

3.1.3. Topology parameters

Multiple topology parameters were used to characterise global and local properties of the network models. These parameters were calculated with the *NetworkX [58]* network analysis Python package and the *ModuLand [33] Cytoscape* [59] plugin suitable for module/community identification. Beside these, the direction and sign of edges between the drug target and its neighbours in the motifs were also included in the machine learning training dataset.

3.1.3.1. Network diameter

The network diameter is defined as the longest shortest path between any two nodes in the network. Average shortest path length was also computed. The network diameter was calculated with *a)* considering the network directed *b)* undirected *c)* mixture of the two, keeping the original directions for positive and negative edges and replacing the undirected neutral edges with bidirectional edges. Isolated network components were removed due to the limitations of diameter calculations. Since *NetworkX* considers edge weights to be path lengths, negative logarithmic transformation, reciprocal transformation or inversion (maximum edge weight – actual edge weight) was performed. Shortest paths were computed *via* Dijkstra's algorithm [60]. This script was firstly developed by Zsolt Nagy, and the candidate adapted the script.

3.1.3.2. Centrality values

Centrality values are helpful to assess importance of certain nodes and edges in the networks, providing an important input to the machine learning model. Bridgeness and betweenness values were calculated for the three signed subnetworks. The **bridgeness centrality** value was defined by *ModuLand* [33]: for a given node, it is the sum of the smaller of the two modular assignments for every module pairs. This value is high if the node equally belongs to two modules, so the node can behave as a “bridge” between them. For **betweenness centrality**, the traditional 1977 definition was used, calculating the fraction of shortest paths passing through the given node [61].

3.1.4. Triangles

Three-nodal motifs were identified with the FANMOD program [52] with the discrimination of positive and negative edges. Fully connected motifs (**triangles**) were

further analysed to find rare regulatory motifs: **unbalanced triangles** (containing odd number of negative edges) and **cycles** (allowing a walk around the motif from at least one direction) as defined in social sciences [62].

3.1.5. Modules

Modules were identified with the *ModuLand* plugin (for the colon carcinogenesis project, run by Borbála Kovács) [33], which has the advantage of focusing on overlapping modules and hierarchical layers. It calculates the assignment of the nodes to the discovered modules based on influence functions and community centrality, allowing module overlaps. Nodes having the largest community centrality and local influence organise and integrate the module, allowing us to estimate the biological function of the given module.

3.2. Protein annotation in triangles

In the biomarker prediction project, proteins in the identified triangles were annotated to select protein pairs for further analysis and to add IDP features as inputs of the machine learning model. Annotations were performed through UniProt ID matching.

3.2.1. IDP annotations

Disordered regions and their ratio in proteins were determined by three different IDP annotation methods. *DisProt* [63] is considered as a gold-standard validated database of disordered regions. The *AlphaFold* [64] database mainly contains structural data, but disorder prediction was possible using pLLDT (B-factor) values from downloadable *.cif* files. Average pLLDT and disorder content (the ratio of pLLDT < 50) was calculated. With the *IUPred* [65] Python package, disordered regions were predicted on FASTA files from *UniProt* [53]. Short, long, and globular scores were calculated, but globular scores were like long scores, so they were later excluded. Average IUPred scores and disorder content values (the ratio of IUPred score > 50) were considered.

3.2.2. Onco-therapeutic target definition

Onco-therapeutic targets were established with the *My Cancer Genome* [66] database. We considered target proteins of approved targeted therapies in oncology and haematology and immunological drugs in anticancer clinical trials. Each target was individually

analysed for multi-kinase inhibitors. The Target Central Resource Database (TCRD/PHAROS) [67] was used to identify preclinical targets, filtering to exclude drugs in clinical trials based on TCRD and ChEMBL [68].

3.2.3. Neighbour-target pairs

Neighbour-target triangles were defined as triangles that included at least one separate approved target member. All the 4550 neighbour-target pair i.e., drug targets paired with every possible member from their motifs was identified in the triangles and were used to construct the training dataset for machine learning along with their annotation and network topological values.

The *CIViCmine* [69] text-mining database was used to find biomarkers, separately annotating prognostic, predisposing, diagnostic, and predictive biomarkers. *CIViCmine* defined predictive biomarker as a protein helping to predict sensitivity or response to a certain drug by its expression level or mutational status. When the neighbour of a target served as a predictive biomarker for the drug inhibiting the target, we manually reviewed the text-mined articles. These protein pairs were treated as positive controls in the final input dataset.

We defined a negative control set from neighbours that were not present in *CIViCmine* as any type of biomarker. This was supplemented with randomly selected pairs for the smaller DisProt-only dataset. We tried a completely randomised approach but resulted in worse metrics potentially due to the involvement of strong biomarker candidates in the negative set. A cancer hallmark-based approach was also previously applied and showed good performance, however, following reviewer feedback regarding the biological assumptions underlying this strategy, we replaced it with the current negative set definition. The list of positive and negative control was the training dataset, while the remaining pairs were classified to predict new potential predictive biomarkers.

3.3. Machine learning

Three different binary classification algorithms were tested to predict new potential predictive biomarkers; Support Vector Machine and Random Forest (RF) from the *Scikit-learn* package and XGBoost (XGB) [38, 39]. Optimal hyperparameters for Random

Forest and XGBoost were acquired through competitive halving random search. The training dataset consisted of the positive and negative controls (see *Section 3.2.3*).

The training dataset includes:

- Topological parameters
    - Networks containing the pair and their triangles
    - Motif and rare motif (unbalanced triangles, cycles) counts
    - Edge types between the neighbour and targets
    - Centrality values (bridgeness, betweenness)
- Biological annotations
    - IDP definitions and the ratio of disordered regions
    - Preclinical and approved drug targets
    - Drug types (antibody or small molecule)

Performance was evaluated with multiple data split techniques, such as the 70:30 data split 5-fold cross-validation and leave-one-out-cross-validation (LOOCV) [70]. Multiple metrics such as receiver operation characteristic area under curve (ROC AUC), precision, recall, their harmonic mean (F1-score) and accuracy were calculated [71]. We excluded the SVM model from further analysis due to underperformance. Leave-one-out-cross-validation models were trained on separate feature groups, such as centrality values or biological features to show their predicting value. In a separate experiment, we tried the systematically cross-classification of networks by training on one network and testing on another one to see the pan-network applicability. Final predictions were made on unclassified neighbour-target pairs. Combined classification was done for all networks and network-specific model. In the same way, combined and IDP annotation-specific predictions were also conducted. The predicted class and the probability of each class were exported. Feature importances were identified with the game-theory based *SHAP* [40] Python package with a *gbtree* booster. The **MarkerPredict** GitHub package contains all the codes and input data (https://github.com/klari98/MarkerPredict). This package is applicable to other networks and biological problems with the appropriate annotation. Gábor Kovács reviewed the codes and helped to solve package dependencies.

3.4. Biomarker Probability Score (BPS) calculation

We defined Biomarker Probability Score (BPS) as a ranking-based probability score estimating the neighbour's biomarker potential for the target's inhibitor in neighbour-target pairs. It was calculated from the biomarker probabilities of the final machine learning prediction from 32 separate models: using XGB and RF on combined and network-specific data. The Biomarker Probability Score (BPS) of each neighbour-target pair was defined as the normalised average of the rank of the predictions (descending order):

(3)

$$BPS(p) = 1 - \frac{\left(\frac{\sum_{i=1}^{n} R_{p,i}}{n}\right)}{m-1}$$

where *R* is the rank of the probability of class 1 for the neighbour-target pair in the given prediction *i*, *n* is the total number of predictions and *m* is the number of the neighbour-target pairs. BPS of 1 means the highest biomarker probability and BPS of 0 means lowest biomarker probability. 4 BPS scores were calculated for the combined and each IDP-annotation-specific prediction. BPS correlates well with the original training label (see Figure 1).

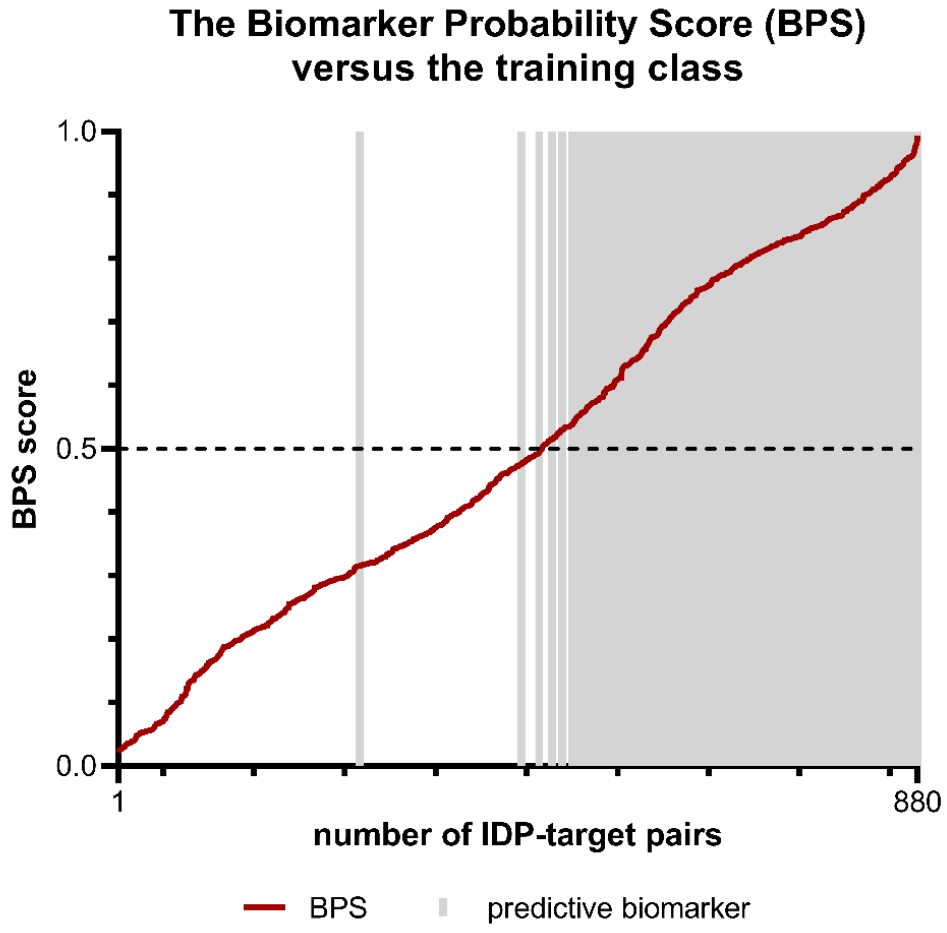

Figure 1. The BPS score versus the original training label of the training dataset. *The BPS score correlates well with the original labels especially at top and bottom values, showing that the score system is suitable to use on the new predictions. This figure is in the candidate's open access publication as Figure 1E and included here with the permission of the co-authors [44].*

3.5. *In vitro* and *in silico* methods used in the pathway specific lncRNA identification

3.5.1. Cell culture and viability

Lung (A549 [72], HT1299 [73]) and liver (HuH7 [74], Hep3B [75]) maintained by Patrizia Birner) cancer cell lines were provided by the laboratory of Prof. Kai Breuhahn. Cell lines were cultured in DMEM medium supplemented with 10% foetal bovine serum (FCS) and 1% penicillin/streptomycin. Cells were kept at 37 °C with 5% CO2 in a humidified atmosphere. They were routinely tested for mycoplasma contamination. Viability detection was performed after 1h incubation with 10% dilution of rezonium.

3.5.2. Drug treatments, siRNA transfection

For treatments and transfection, an equal number of cells were seeded after manual cell counting. Drug treatments were performed with the MAPK pathway inhibitors trametinib and binimetinib (MEK inhibitors; [76], [77]) and ulixertinib and MK-8353 (ERK inhibitors; (ERK inhibitors; [78], [79]) alongside with the AKT inhibitor AKTVIII, one drug per treatment. A matching DMSO-treated control was used. Cells were treated with gene-specific siRNAs as follows. First, cells were washed with Opti-MEM and incubated at room temperature for 8 minutes. The siRNA–transfection reagent complex was prepared and allowed to mix for 20 minutes at room temperature. Following mixing, 800 µl Opti-MEM were added to the cells, and 200 µl of the siRNA–reagent solution (solution A/B) were then added to each well. Cells were incubated for 4 hours, after which 1 ml of complete growth medium was added. The medium was replaced again after 24 hours to ensure optimal conditions for gene silencing. As control, non-silencing siRNA was used.

3.5.3. Western blotting, RNA isolation and real-time quantitative PCR (RT-qPCR)

Protein isolation, Western blotting, RNA extraction and RT-qPCR were conducted as recently described by Prof. Kai Breuhahn [80]. Multiple newly designed lncRNA primers were tested for each target, designed preferably on exon-exon junctions. Melting curves were used for appropriate primer selection. LncRNA candidates with working primers were used for further analysis.

3.5.4. LncRNAseq and bioinformatic analyses

RNA sequencing including lncRNAs was performed on one NSCLC (A549) and one HCC (Hep3B) cell line after 3 hours of trametinib or MK-8353 treatment or DMSO control by the BGI company (Hong Kong). RNAseq preprocessing was conducted by Carsten Sticht from the Medical Research Centre of Mannheim, Germany. On the pre-processed RNAseq data, lncRNA candidates were identified among the overlap of downregulated lncRNAs in both cell lines and treatments. The survival prediction potential of the lncRNA candidates was evaluated on TCGA HCC and NSCLC data.

4. Results

4.1. Key network topology changes in the development of colon adenocarcinoma

In this project, network science tools were used to model the development of colon cancer using the cancer-specific *Human Cancer Signalling Network* (CSN) [49] and open-access RNA microarray data [22, 54-57] of healthy colon, colon adenoma and left-sided colon adenocarcinoma, aggregating >100 samples per each. Changes in global and local network attributes highlighted important changes in the carcinogenesis process. This project was conducted in collaboration with Borbála Kovács, the diameter calculation method was first developed by Zsolt Nagy, and the data normalization was conducted by Sebestyén Kamp and János Molnár. This project was previously published in an open-access article alongside with other results [13].

4.1.1. Network diameter

Network diameter is a global attribute of signalling networks, defined as the sum of the edge weights within the longest shortest path between two nodes [81], describing signal propagation in networks. Average shortest path length was also calculated for further characterization.

Using the normal, adenoma and carcinoma network models, multiple calculations were conducted (see Table 2), handling the networks as a) undirected, b) directed and c) the mixture of the two. As the calculated edge strengths represent connection strengths and the used Dijkstra algorithm [60] handles them as distance, multiple approaches were used to reverse the values, such as a) negative logarithmic transformation, b) reciprocal transformation, c) inversion (for further details, see Methods). The original script was written by Zsolt Nagy, while adapting the code and the evaluation of results was performed by the candidate.

Table 2. Various network diameter calculations in the healthy colon, colon adenoma and colon adenocarcinoma network models. *Network diameter and average shortest path lengths were calculated with the Dijkstra algorithm and NetworkX [58, 60]. Calculations were conducted considering the networks as a) undirected, b) directed, c) as a mixed graph (see Methods). As network diameters represent distance in a network and the edge calculation represent connection strength, various methods were used for reversing edge weights, such as negative logarithmic transformation, reciprocal transformation and inversion (extracting the edge weight from the maximum edge weight). Negative logarithmic network diameter was also calculated with adding 5% noise to the edge weights to assess the robustness of the results. The base negative logarithmic values were evaluated with Friedman ANOVA, resulting in a p=0.0278. These results were shown in the candidate's article as Table 1 and S2, shown with the co-authors' approval [13].*

Network diameters with negative logarithmic mapping:

	Undirected		**Directed**		**Mixed graph**	
	Network diameter	**Average path length**	**Network diameter**	**Average path length**	**Network diameter**	**Average path length**
Normal	34.361	*9.984*	36.988	*12.940*	36.435	*11.397*
Adenoma	36.365	*10.667*	40.051	*13.790*	39.152	*12.198*
Carcinoma	29.753	*8.224*	32.499	*10.798*	30.769	*9.423*

Network diameters with negative logarithmic mapping and 5% added noise:

	Undirected		**Directed**		**Mixed graph**	
	Network diameter	**Average path length**	**Network diameter**	**Average path length**	**Network diameter**	**Average path length**
Normal	34.295	*9.885*	37.346	*12.822*	36.435	*11.399*
Adenoma	36.682	*10.573*	40.407	*13.680*	38.876	*12.092*
Carcinoma	29.362	*8.122*	32.779	*10.676*	30.552	*9.314*

Network diameters with reciprocal and inverted data transformation:

	Undirected		**Directed**		**Mixed graph**	
	Reciprocal	**Inverted**	**Reciprocal**	**Inverted**	**Reciprocal**	**Inverted**
Normal	44 090	*29 260*	23 739	*21 987*	44 447	*29 697*
Adenoma	55 649	*38 724*	32 983	*29 065*	57 834	*38 724*
Carcinoma	24 523	*20 035*	21 235	*21 235*	31 757	*21 783*

Across all three networks, the adenoma network consistently showed the largest, and the carcinoma network the smallest network diameter, regardless of the calculation method. The average shortest path lengths changed in parallel. These findings were reproduced with two independent approaches and remained robust under added noise. The increased diameter in the adenoma network suggests a less compact network architecture, while the

reduced diameter and shorter paths in the carcinoma network may reflect an adaptive strategy to enhance information flow—an effect also reported in other topological studies. This pattern may represent a cellular stress response in the cancer environment, offering additional perspectives beyond standard carcinogenesis-related changes.

4.1.2. Therapy-related pathways

For metastatic colon cancer, targeted therapies such as EGFR-, VEGFR-inhibitors and immune checkpoint inhibitors are available, the latter is based on mismatch repair status [82]. These pathways were analysed in the normal colon, adenoma and carcinoma network models, defined as the proteins belonging to the respective Gene Ontology terms: *'epidermal growth factor receptor signalling pathway'*, *'vascular endothelial growth factor receptor signalling pathway'* and *'mismatch repair'* [83].

Ultimately, approximately half of the proteins in the EGFR and VEGFR pathways, but only about one quarter of those in the mismatch repair pathway were represented in the network (see Table 3). This discrepancy likely reflects the composition of the Human Cancer Signalling Network [49], which seems to be enriched for signalling-related proteins and therefore may include fewer components of the mismatch repair pathway. Regardless, as a next step, network topology analysis was conducted on these pathways' nodes.

Table 3. Number of identified proteins and their median abundance. *Pathway nodes were identified based on Gene Ontology [83] terms. As the data did not pass the Shapiro-Wilk normality testing, Wilcoxon-test was used on the abundance values of the pathway proteins to assess significance. Significant values are highlighted. Published as Table 2 with the candidate's other work, shown with the co-authors' approval [13].*

	No. of proteins in GO	**No. of proteins in the network**	**Median normal abundance**	**Median adenoma abundance**	**Median carcinoma abundance**	***p-value (normal - adenoma)***	***p-value (adenoma - carcinoma)***	***p-value (normal - carcinoma)***
EGFR signalling	121	58	7.061	6.665	7.058	*0.018*	*0.018*	*0.670*
VEGFR signalling	93	49	6.662	7.383	7.608	*0.842*	*<0.001*	*0.018*
mismatch repair	38	8	8.533	9.931	8.987	*0.080*	*1.000*	*0.234*

To assess the importance of these pathways in normal, adenoma and carcinoma networks, both median abundance (see Table 3) and weighted degree (see Figure 2) were calculated. The median abundance of the VEGFR pathway-associated proteins were the highest in the carcinoma dataset (significant for VEGFR), while the median weighted degree was the highest in the normal network for EGFR and VEGFR. This may be due to the higher-expressed neighbours of the pathway proteins, but it is also known that EGFR is important in the healthy alimentary tract too [84]. The adenoma network has low values for both (significantly for abundances), in the EGFR pathway, showing that this pathway does not drive the early stage of carcinogenesis [85]. However, the mismatch repair pathway has the highest abundance and weighted degree values for adenoma and relatively low for carcinoma, which may be related to the selection criteria in this study: mainly left-sided colon cancer samples were included where mismatch-repair deficiency is less common, and only 8 nodes was found in the network (see Methods) [86].

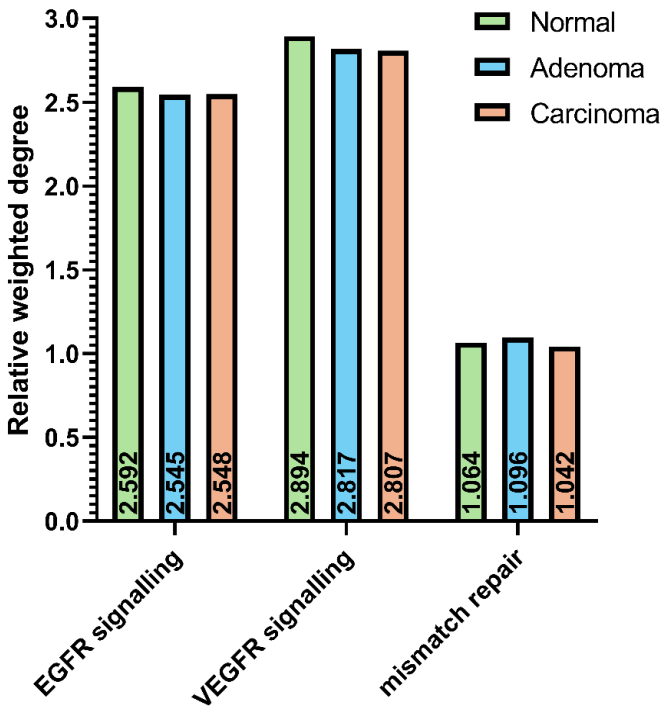

Figure 2. Median relative weighted degrees of targeted and immunotherapy pathway related nodes in the normal, adenoma and carcinoma network. *Weighted degrees were calculated according to the traditional formula, such as the sum of the edge weights belonging to each node. Relative weighted degrees were considered as the quotients of the weighted degrees of each of the signalling pathways and the whole network. Pairwise comparisons significant with Wilcoxon tests*

were EGFR normal-adenoma and VEGFR adenoma-carcinoma. The data was published as Table S10 in the candidate's publication, shown with the co-authors' approval [13].

The modular affiliation of each node was studied in all three networks (with ModuLand, run by Borbála Kovács). Most nodes clustered into the RAC1 module, the largest module in the network, whose biological functions cannot be clearly defined. This suggests that these pathways do not form distinct modules because they are highly intertwined with the rest of the graph, preventing their separation by the modularization algorithm (see Figure 3). This might be because the EGFR and VEGFR pathways contain essential proliferation-associated proteins that can be activated by other upstream processes and cross-talks, thus forming the backbone of the network, while the mismatch repair pathway contains only 8 proteins, too few to form a module. Edges associated with targeted pathways were found among both the strongest and weakest interactions in the network.

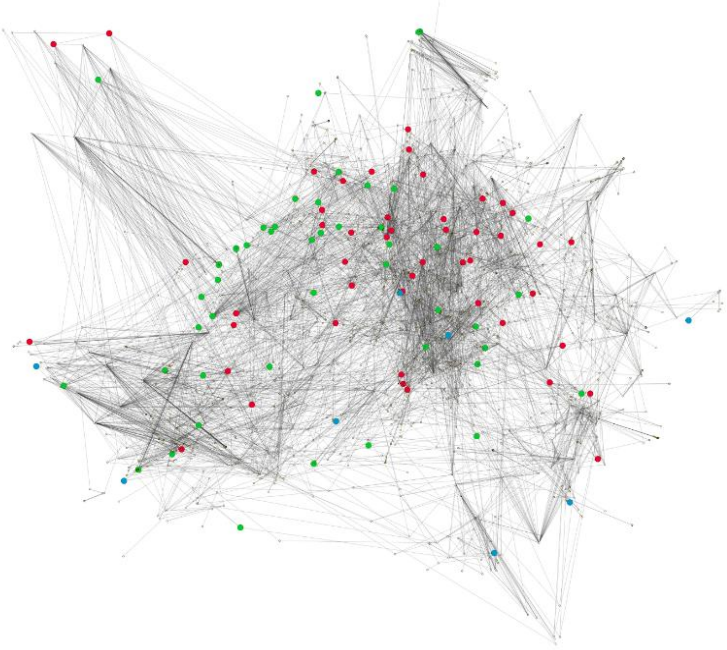

Figure 3. The EGFR-, VEGFR- and mismatch repair pathway proteins in the carcinoma network model. *The image was generated using the EntOpt entropy optimalization-based visualisation tool [87]. The EGFR pathway nodes are green, the VEGFR nodes are red, and the mismatch repair pathway nodes are blue. These pathways do not form different modules, and they are distributed in the network. This visualisation was published as Figure S22 in the candidate's work, shown with the co-authors' approval [13].*

4.1.3. The MAPK pathway during colon carcinogenesis

As about half of colon cancer patients have one or more mutations in the MAPK pathway [7], it was analysed in the normal colon, colon adenoma and carcinoma network models. However, we note that the effect of MAPK pathway changes like the frequent KRAS mutation [8] might affect posttranslational and downstream expression levels, which is difficult to assess here due to the limitation of this model. These results were not part of the original publication.

The proteins belonging to the *'MAPK cascade'* Gene Ontology [83] term were analysed. Out of the 189 proteins, 93 were found in the CSN network, about 80% of them belonging to the largest RAC1 module in the networks. Interestingly, the relative median weighted degree (the quotient of the median weighted degree of the MAPK pathway and the whole network) of these proteins is slightly decreasing throughout the modelled carcinogenesis (normal: 1.80, adenoma: 1.76, carcinoma: 1.74). A narrower GO term *'MAP kinase activity'* was also analysed, with 13 strictly MAPK pathway members in the CSN network out of the 16 – all belonging to the RAC1 module. Their median relative weighted degree also slightly decreased (normal: 3.81, adenoma: 3.75, carcinoma: 3.66). Notably, several key MAPK components' (p38, MEK1, JNK, ERK2, c-Myc, c-JUN, BRAF) weighted degree steadily increased from normal to adenoma to carcinoma.

The higher importance of the pathway in the normal colon model is likely because MAPK signalling is essential in normal, proliferating colon cells as well [88]. The mutational activation of KRAS and the whole pathway might also lead to negative feedback on the expression level [89]. However, we concluded that the overall high relative weighted degree of this pathway shows that it is generally important in colon mucosa-derived signalling networks regardless of tumorigenesis stage.

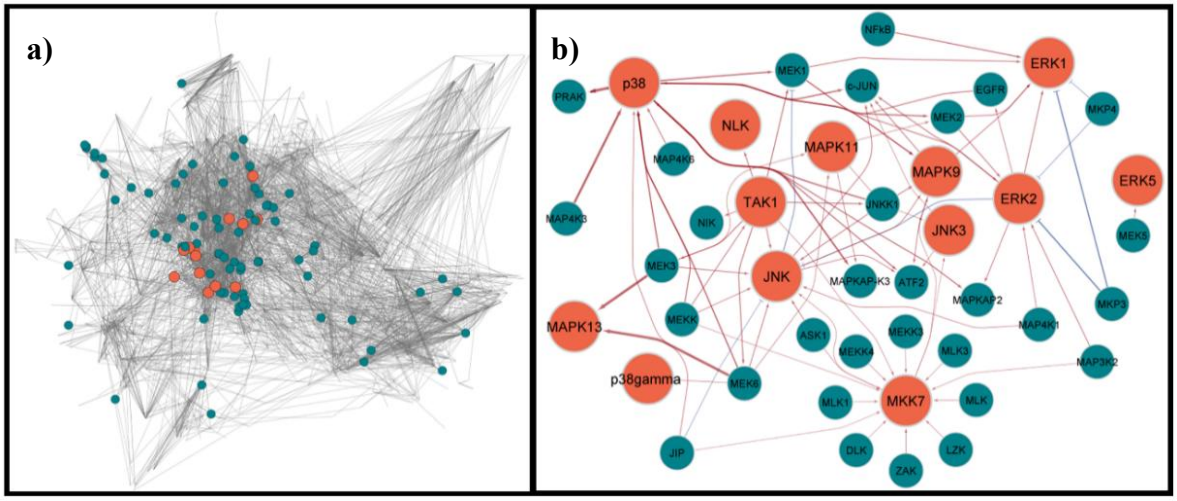

Figure 4. The MAPK pathway in the colon carcinoma network. *Blue nodes belong to the 'MAPK cascade' and orange nodes to the narrower 'MAP kinase activity' GO terms.* ***a)*** *EntOpt visualisation of the network with the highlighted pathway.* ***b)*** *The MAPK cascade subnetwork.*

4.2. MarkerPredict: A machine learning model predicting new biomarkers to targeted cancer drugs

To see how network science and machine learning can be combined to identify key determinants of therapy response, graph-based machine learning models were run after analysing network motifs containing cancer therapy targets (see Figure 5). The resulting **Biomarker Probability Score (BPS)** is a powerful tool in finding predictive biomarkers for targeted cancer therapeutics. These ideas were co-developed with Dániel Veres and Péter Csermely, and the final codes were reviewed by Gábor Kovács, who also helped in the development of the Python package, which can be found at https://github.com/klari98/MarkerPredict. The results are published in an open-access article [44].

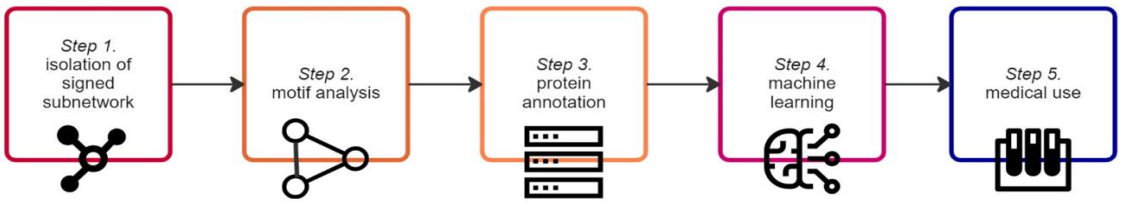

Figure 5. A flowchart of the MarkerPredict process. *After network construction and motif-analysis, annotation was conducted to build the input dataset for the machine learning model. The final predictions were annotated with rationale and medical relevance. This figure was published as Figure 1A in the candidate's article, shown with the co-authors' approval [44].*

4.2.1. The enrichment of intrinsically disordered proteins in triangles of therapeutic targets

Three signed subnetworks — derived from the Human Cancer Signalling Network (CSN), SIGNOR, and ReactomeFI [49-51] — exhibiting markedly different topological features (see Table 1) were used for motif identification. Fully connected three-node motifs, i.e. triangles were identified with the FANMOD program (see Methods). Rare regulatory structures such as unbalanced triangles and cycles were also analysed due to their specific roles in signalling networks.

After obtaining the lists of IDPs (as defined in DisProt [63]) and onco-therapeutic targets (see Methods), their occurrences within triangles were determined. Triangles simultaneously containing DisProt-defined IDPs and drug targets were analysed as potential signalling "hot spots." Unbalanced triangles were significantly overrepresented among them, whereas cycles were only overrepresented in the CSN and SIGNOR networks. Importantly, IDP–target triangles appeared far more (11.91x in ReactomeFI) frequently than expected by random chance. This enrichment held true not only for DisProt-defined IDPs but also for AlphaFold-based (pLLDT < 50 [64]) and IUPred-predicted (score > 0.5 [65]) IDPs (a range of 3-11.91x enrichment in total).

Sharing a common motif typically indicates a close regulatory relationship between two nodes, and motif structures can even be used to predict new drug targets [32]. Because an ideal predictive biomarker should reflect conditions that influence the efficacy of a drug, a strong regulatory connection between an IDP and a drug target—such as co-occurrence in the same triangular motif—may indicate predictive biomarker potential. In our framework, signalling relationships are represented through simplified three-node. Since such motifs usually correspond to a local fragment of a signalling pathway, we hypothesised that IDPs found in the same triangle as a drug target (**"IDP–target pairs"**) may serve as predictive biomarkers for drugs aimed at that target. To expand the analysis, we also included all neighbours of drug targets (**"neighbour–target pairs"**).

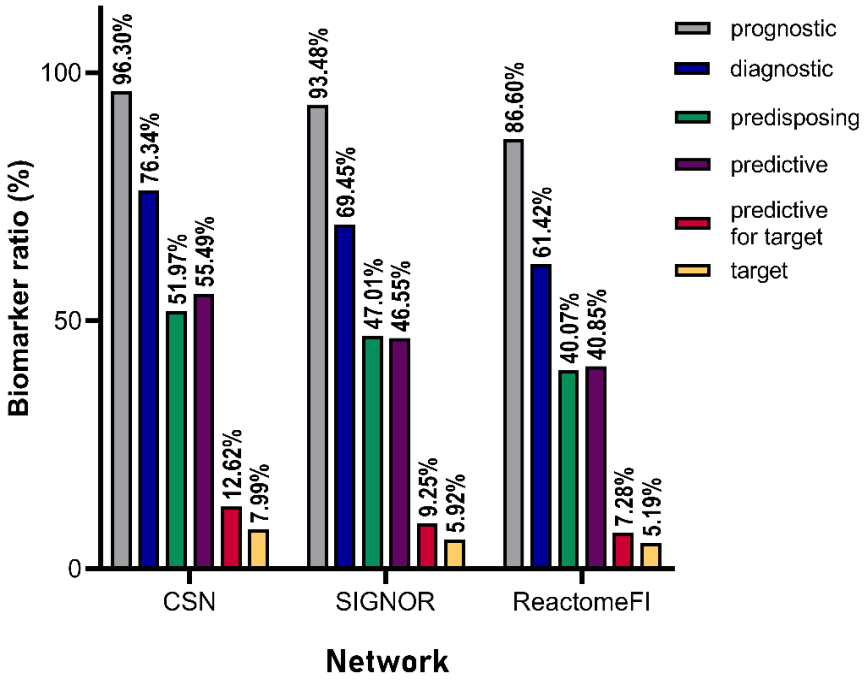

Figure 6. Biomarker and drug target properties of neighbours in target-neighbour triangles. *Most of the target neighbours [86.6% - 96.3%) fit the given biomarker definition of the CIViCmine database. A considerable ratio is a predictive biomarker of a drug specifically targeting their neighbour. This figure was published as Figure 1C in the candidate's article, shown with the co-authors' approval [44].*

To assess whether IDPs show biomarker properties, we annotated them using the CIViCmine database (see Figure 6 [69]). Across all three networks, more than 86% of IDPs were labelled as prognostic biomarkers, and substantial proportions were predisposing, diagnostic, or predictive biomarkers. Cases in which a protein served as a predictive biomarker for a drug targeting its triangle partner were manually reviewed and designated as positive controls for machine-learning training. In total, 332 of the 4,550 neighbour–target pairs represented confirmed predictive biomarker relationships—an impressive number given the specificity of this condition. For the remaining 4,223 neighbour–target pairs, a negative control dataset was constructed from neighbours absent from CIViCmine and additional random pairs. This data was used to train binary classification models.

4.2.2. MarkerPredict is a powerful tool for biomarker prediction

Random Forest and XGBoost binary classifiers [38, 39] were trained on the network-specific datasets as well as on the combined data from the three (CSN, SIGNOR,

ReactomeFI [49-51] signalling networks for the 880 annotated neighbour-target pairs. Final biomarker predictions were conducted on the remaining unclassifiable 3670 pairs. Separate models were also trained on each of the three IDP database and prediction method (DisProt, AlphaFold, IUPred [63-65], as well as on their combination, resulting in thirty-two high-performing models in total. Optimal hyperparameters were selected using competitive random halving.

4.2.2.1. Validation methods show high performance

Model performance was evaluated using leave-one-out cross-validation (LOOCV), 5-fold cross-validation (see Figure 7), and a 70:30 train–test split. All models achieved strong metrics with the optimised hyperparameters; even those trained on the smaller DisProt dataset performed acceptably. Random Forest models performed slightly worse than XGBoost, and models trained on the CSN network [49] showed lower performance overall, likely due to its smaller size. The LOOCV accuracy values were between 0.7-0.96. All 32 models were included for the final prediction. To harmonise prediction probabilities, we defined the normalised average of ranked probabilities as the Biomarker Probability Score (BPS), calculated four times for every IDP database and their combination, every score including 8 models. With validation, it shows good correlation with the original training label on the training data (see Methods).

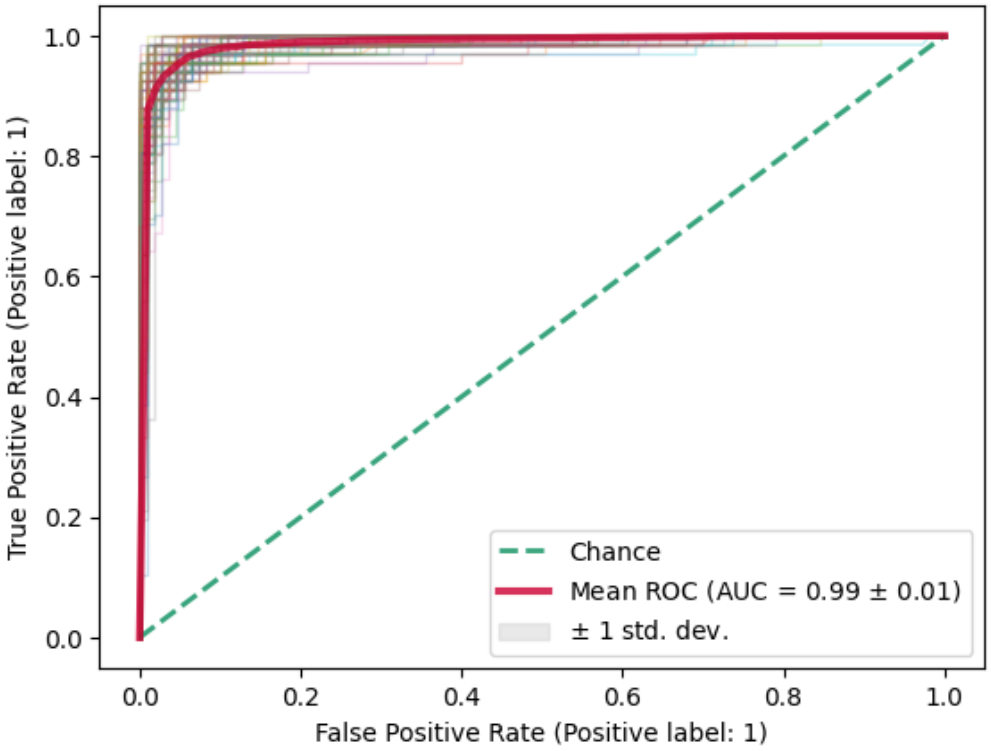

Figure 7. The ROC curve of 100 5-fold cross-validation of the XGBoost model trained on the combined data of all networks and databases. *The model reached excellent performance, with an AUC of 0.99±0.01 (shown with red). Other models also showed high performance. This was published as Figure 1D in the candidate's article, shown with the co-authors' approval [44].*

4.2.2.2. Protein disorder is important in biomarker prediction

Feature importance was evaluated with the SHAP package [40], which computes game-theory-based Shapley value estimations to quantify each feature's contribution. The SHAP values indicate that features derived from the IDP databases and prediction methods are among the most influential inputs. Centrality measures and motif analysis (i.e., number of triangles with the given neighbour-target pair) also showed strong contributions, with higher feature values corresponding to higher SHAP values. Interestingly, some neighbour–target pairs with low or moderate disorder content were more likely to be classified as predictive biomarkers, whereas for the neighbour's IUPred short predictions, a score system developed for short, disordered regions, higher disorder levels were more likely. In summary, intrinsically disordered proteins may be more likely to be biomarkers, and disorder-related features provide important predictive value. However, the exact translation of this may varies on the disorder prediction tool in question.

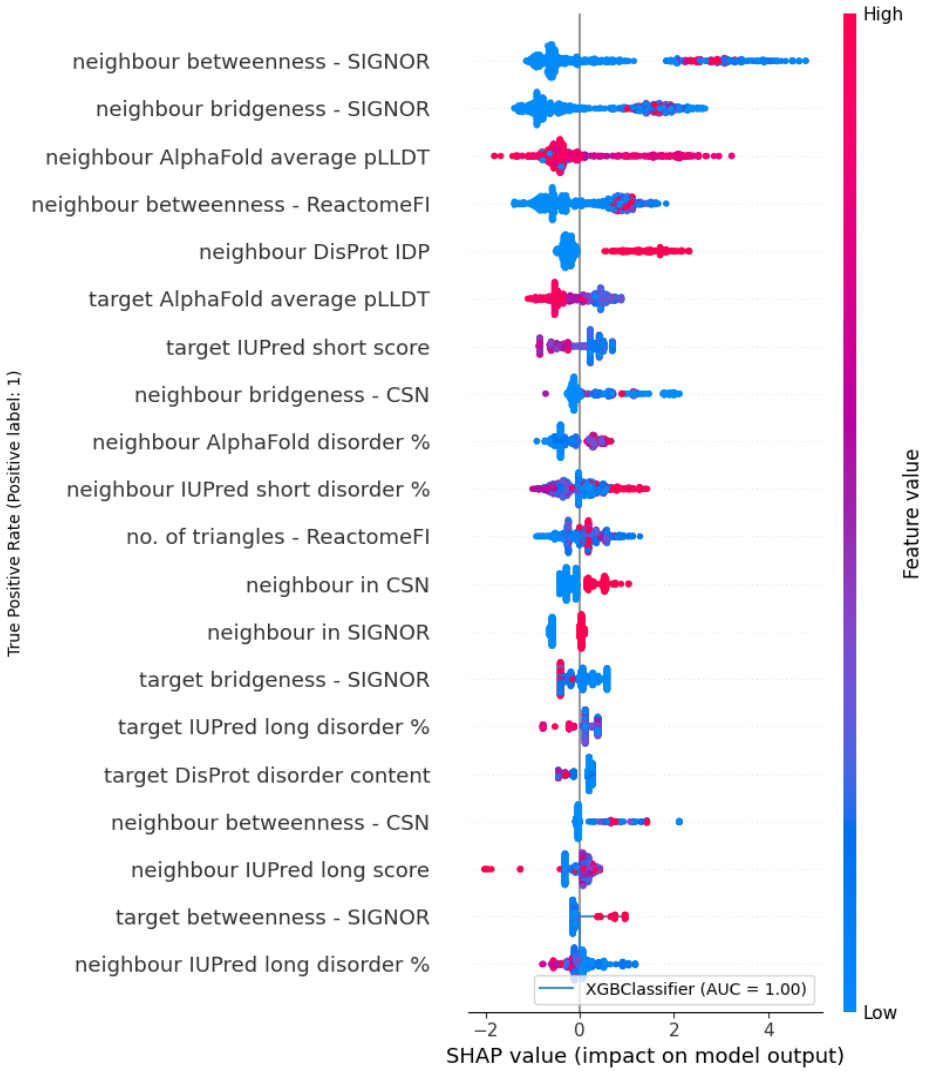

Figure 8. The SHAP beeswarm plot of the XGBoost model trained on the combined data of all networks and databases. *The plot shows the impact of the given feature's value on the model's output, listed in order of significance. Centrality values, disorder annotations and motif analysis results are among the most important features. This figure was published as Figure S6A in the candidate's article, shown with the co-authors' approval [44].*

4.2.3. Predicted biomarkers with biomedical relevance

After generating predictions with all eight models, we calculated the Biomarker Predictability Score (BPS) for each neighbour–target pair across all IDP annotations (see Sections 3.3–4). The BPS effectively unified probability estimates across the XGBoost and Random Forest models and across networks. Using a threshold of BPS > 0.5 in at least one of the four calculations, 2084 neighbour–target pairs were classified as potential predictive biomarkers, and 426 pairs exceeded this threshold in all four IDP-prediction method-specific and combined analyses.

4.2.3.1. The strongest biomarker predictions

For many high-BPS pairs, preclinical or clinical evidence supported the predictions (see Table 4). The top predictions mainly involved targets of small-molecule receptor inhibitors or HER2 inhibitors like trastuzumab [90], typically with neighbours exhibiting low to moderate – but not zero disorder content. For many predictions, clinical or preclinical evidence can be found, however, there are some new strong predictions such as LCK-PDGFRB, SRC-RAF1, ERK1-BCL-2, TRAF6-HER2 and TRAF6-EGFR.

Table 4. Top 5 predictive biomarker predictions per IDP annotation method and for all annotations. *BPS was calculated from the 8 models (two machine learning algorithms on three signalling networks and their total, see Methods). Disorder content was added from the DisProt [63] database and the AlphaFold [64] and IUPred [65] prediction methods, while targeted therapies were listed from the My Cancer Genome [66] database. Supporting literature is from PubMed with the keywords of the drug and the potential biomarker neighbour names. Articles where the response to the drug could be forecasted knowing the mutation status or expression level of the biomarker. Preclinical and clinical articles were handled separately. This table was published as Table 1 in the candidate's article, shown with the co-authors' approval [44].*

No	Bio-marker Proba-bility Score (BPS)	Neigh-bour name	Annotation disorder content (%)				Target name	Inhibitor(s) of target	Supporting literature
			DisProt	*AlphaFold*	*IUPred (long)*	*IUPred (short)*			
ALL ANNOTATIONS									
1.	*0.998*	**LCK**	*7.1%*	*11%*	*3.1%*	*3.9%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	*Clinical:* Pizzamiglio et al., 2021 [91]
2.	*0.991*	**LCK**	*7.1%*	*11%*	*3.1%*	*3.9%*	**PDGFRB**	regorafenib, axitinib	
3.	*0.990*	**ITGB1**	*5.9%*	*3.8%*	*5%*	*6.6%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	*Preclinical:* Gu et al., 2025 [92]
4.	*0.990*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**FGFR3**	ponatinib	*Preclinical:* Matsuda et al., 2016 [93]
5.	*0.987*	**STAT1**	*5.5%*	*6.1%*	*6.8%*	*6.9%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	*Preclinical:* Yang et al., 2025 [94] *Clinical:* Srivastava et al., 2016 [95]

Table 4. is continued on the next page.

Table 4. continued from the previous page.

DISPROT									
1.	*0.991*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**BRAF**	dabrafenib, vemurafenib	*Preclinical:* *King et al., 2013 [96]*
2.	*0.988*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**RAF1**	regorafenib, sorafenib	*Preclinical:* Cao et al., 2006 [97]
3.	*0.987*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**HER2**	trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib, neratinib, afatinib	*Preclinical:* Li et al., 2018 [90]
4.	*0.984*	**SRC**	*16%*	*14%*	*16%*	*14%*	**RAF1**	regorafenib, sorafenib	
5.	*0.980*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**BCL-2**	venetoclax	
ALPHAFOLD									
1.	*0.979*	**TRAF6**	*N/A*	*12%*	*4.4%*	*4.2%*	**HER2**	trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib, neratinib, afatinib	
2.	*0.976*	**PRKCA**	*N/A*	*8.6%*	*9.1%*	*11%*	**HER2**	trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib, neratinib, afatinib	*Preclinical:* Pandya et al., 2016 [98]
3.	*0.976*	**JNK**	*N/A*	*14%*	*17%*	*19%*	**HER2**	trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib, neratinib, afatinib	*Preclinical:* Itah et al., 2023 [99]
4.	*0.974*	**NFkB**	*N/A*	*26%*	*20%*	*15%*	**MTOR**	everolimus, temsirolimus	*Preclinical:* Xie et al., 2007 [100]
5.	*0.973*	**LCK**	*7.1%*	*11%*	*3.1%*	*3.9%*	**PDGFRB**	regorafenib, axitinib	
IUPRED									
1.	*0.993*	**FYN**	*N/A*	*15%*	*7.3%*	*8.2%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	*Preclinical* Kim et al., 2025 [101]
2.	*0.990*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**HER2**	trastuzumab, pertuzumab, ado-trastuzumab emtansine, lapatinib, neratinib, afatinib	*Preclinical:* Li et al., 2018 [90]
3.	*0.987*	**TRAF6**	*N/A*	*12%*	*4.4%*	*4.2%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	
4.	*0.986*	**ERK1**	*6.1%*	*6.5%*	*6.6%*	*12%*	**JAK2**	ruxolitinib	*Preclinical:* Stivala et al., 2019 [102]
5.	*0.983*	**STAT1**	*5.5%*	*6.1%*	*6.8%*	*6.9%*	**EGFR**	afatinib, cetuximab, erlotinib, gefitinib, lapatinib, necitumumab, osimertinib, panitumumab, vandetanib	*Preclinical:* Yang et al., 2025 [94] *Clinical:* Srivastava et al., 2016 [95]

4.2.3.2. Potential predictive biomarkers of the MAPK pathway with MarkerPredict

To show how the MAPK pathway can be analysed with the MarkerPredict approach, the newly predicted MAPK pathway biomarkers were reviewed. As there are already approved cancer drugs for BRAF and MEK1/2, there are clinically established predictive biomarkers too, mainly canonical MAPK pathway activators. In particular, BRAF *V600* mutations are the primary indication for BRAF inhibitors [6], and NRAS/KRAS mutations strongly predict sensitivity to MEK inhibitors like binimetinib alongside with KRAS inhibitors [103].

To find other potential MAPK pathway biomarkers, the network neighbours of these proteins were considered as potential predictive biomarkers of the MAPK pathway. As MEK1 and MEK2 were filtered out from the previously published analysis due to the limitations of the used drug target list, the 139 neighbour-MEK1/2 pairs and their annotations were added to the input tables. The MarkerPredict script were re-ran and the results were evaluated, with recalculating the BPS scores. For clarity, these new BPS scores are only used in this subchapter.

After re-running the MarkerPredict package with the added MEK1/2 pairs, several potential predictive biomarkers were identified for both BRAF and MEK inhibitors. ERK1 had a high BPS score for both BRAF, MEK1 and MEK2, corresponding with its role as a frequent MAPK pathway activity readout in experiments [45]. Several other MAPK pathway members and associated proteins were also predicted, such as RAF1, ERK2, SHC1, SOS1, SRC and RAC1. This may not be surprising, as MAPK pathway is a strictly regulated, evolutionary conserved core pathway, with many parallelisms and co-regulations [5].

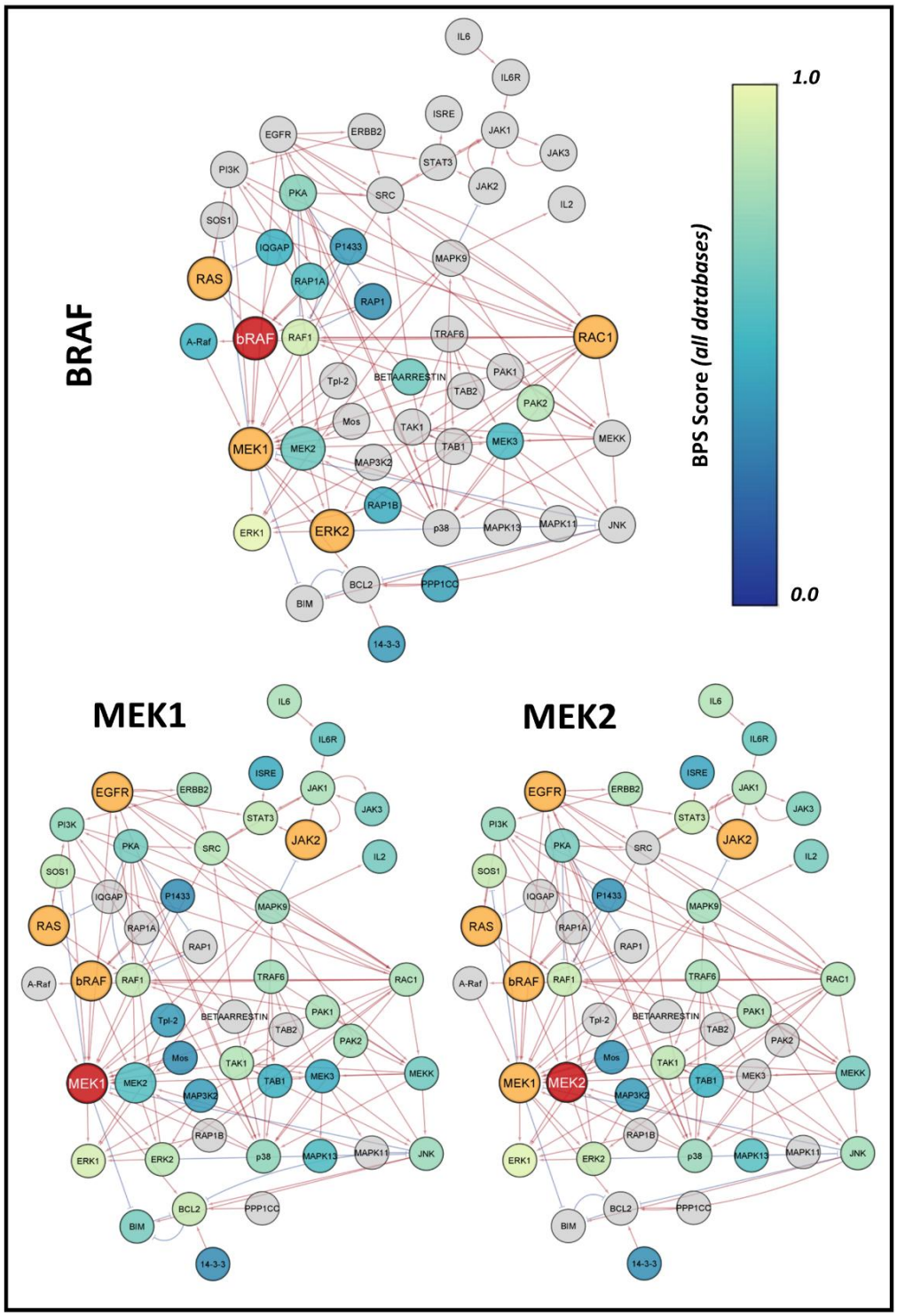

Figure 9. The BPS scores of potential predictive biomarkers for MAPK pathway inhibitors. *The CSN network was selected for visualisation, the subnetwork of MEK1/2 and BRAF motifs (union of all 3 networks) were created. Positive edges are shown with red; negatives are with blue. Targets are shown with red, their defined predictive biomarkers with orange (from the CIViCmine [69] database). 3 separate visualisations were created to show the BPS scores (calculated from predictions with all IDP annotations) of the pairs of BRAF, MEK1 and MEK2. Nodes with high BPS scores shown with yellow, low with blue (see colour scale), non-pairs with grey. The edges with YWHAZ were not present in the CSN network, thus not visualised here.*

Even though the predicted biomarkers were not present in the CIViCmine [69] database, several of the predicted biomarkers have some kind of evidence supporting that they modulate MAPK inhibitor sensitivity. ERBB2 has been shown to drive resistance to BRAF inhibitors in BRAF mutant tumours, where ERBB2 activation restores downstream signalling despite BRAF inhibition [104]. GSK3B is another hit with experimental support: high GSK3B activity has been associated with reduced vemurafenib response in melanoma models [105]. Protein kinase C family members, such as PRKCA, B, D and E were also in the predictions and evidence shows that PRKCA expression might be a good marker of MEK inhibitor sensitivity in low-grade serous ovarian cancer models [106]. PIK3R1, the PI3K p85α regulatory subunit, has also supporting evidence, as its loss enhances MEK dependency and increases MEK-inhibitor sensitivity in breast cell lines [107]. These literature findings support the validity of the predictions made by the MarkerPredict method.

Among the novel predictions, YWHAZ (14-3-3ζ) as a member of the 14-3-3 adaptor family, stands out as it is mechanistically relevant, but not a typical co-regulated MAPK pathway kinase [108]. It is the only 14-3-3 protein predicted as biomarker for both BRAF and MEK1/2 inhibitors, as BPS_{alldb} and $BPS_{DisProt}$ scores are in the range of 0.53-0.66. The 14-3-3 proteins are adaptor proteins that bind BRAF/MEK complexes and regulate MAPK signalling [109]. Studies show that the binding happens in both the autoinhibited and the active status, using different conformations to regulate BRAF activation [110]. Thus, differing protein levels affected by mutation or expression changes could affect the pathway's activity. Among 14-3-3 proteins, YWHAZ is especially relevant in cancers [111]. As a prognostic biomarker, its overexpression associates with worse survival in various cancers like lung, breast, head and neck tumours and glioblastoma. It promotes the signalling of MAPK-associated upstream proteins, shown by the genetic blockage of YWHAZ, which disrupts ERBB2 and PI3K signalling [112].

There are efforts to design inhibitors for YWHAZ, however, there are no effective drugs in clinical phase yet. This could be connected to the fact that it is an established IDP according to the DisProt [63] IDP database. Even though it has small, 4.9% disorder content, it is historically difficult to design drugs for disordered proteins and regions [113]. Despite these challenges, the role of YWHAZ in cancers and MAPK pathway

activation might be important as a good predictive biomarker candidate for BRAF and MEK inhibition, which needs validation with clinical and preclinical studies.

4.3. A pan-cancer lncRNA signature specific to MAPK pathway activation

In this project, biomarker lncRNAs of the MAPK pathway were identified using experimental techniques and RNA sequencing. First, we searched the optimal cell lines, drugs and timepoints of optimal MAPK pathway downregulation. As a next step, lncRNAseq was performed in an external laboratory, followed by the identification of lncRNA candidates. Confirmation experiments supported the MAPK pathway specificity of this panel. This project was supervised by Prof. Kai Breuhahn, and the experiments were split with our lab technician Patrizia Birner. The general pipeline of lncRNA identification was initially developed by Fabian Rose, a PhD student of the Breuhahn lab on another pathway, yet to be published. These results are also unpublished at the moment of writing this thesis; however, a preliminary report was already presented at a conference [46].

4.3.1. Optimal cell lines, treatments, doses and incubation time for effective MAPK inhibition

As the goal was to identify a pan-cancer, or at least cancer-type-independent lncRNA signature, cell lines of two different cancer types (lung (A549 [72], HT1299 [73]), liver (HuH7 [74], Hep3B [75]) were used to find the conditions of the most effective MAPK pathway inhibition. To ensure that the results are specific to the whole MAPK pathway, drugs on two separate points of interference (trametinib and binimetinib (MEK inhibitors; [76], [77]), ulixertinib and MK-8353 (ERK inhibitors; [78], [79]) were tested. Viability testing was used to find doses of effective but not lethal inhibition. Western blotting was performed to show the effectivity and time-dependence of the pathway inhibition (see Figure 10).

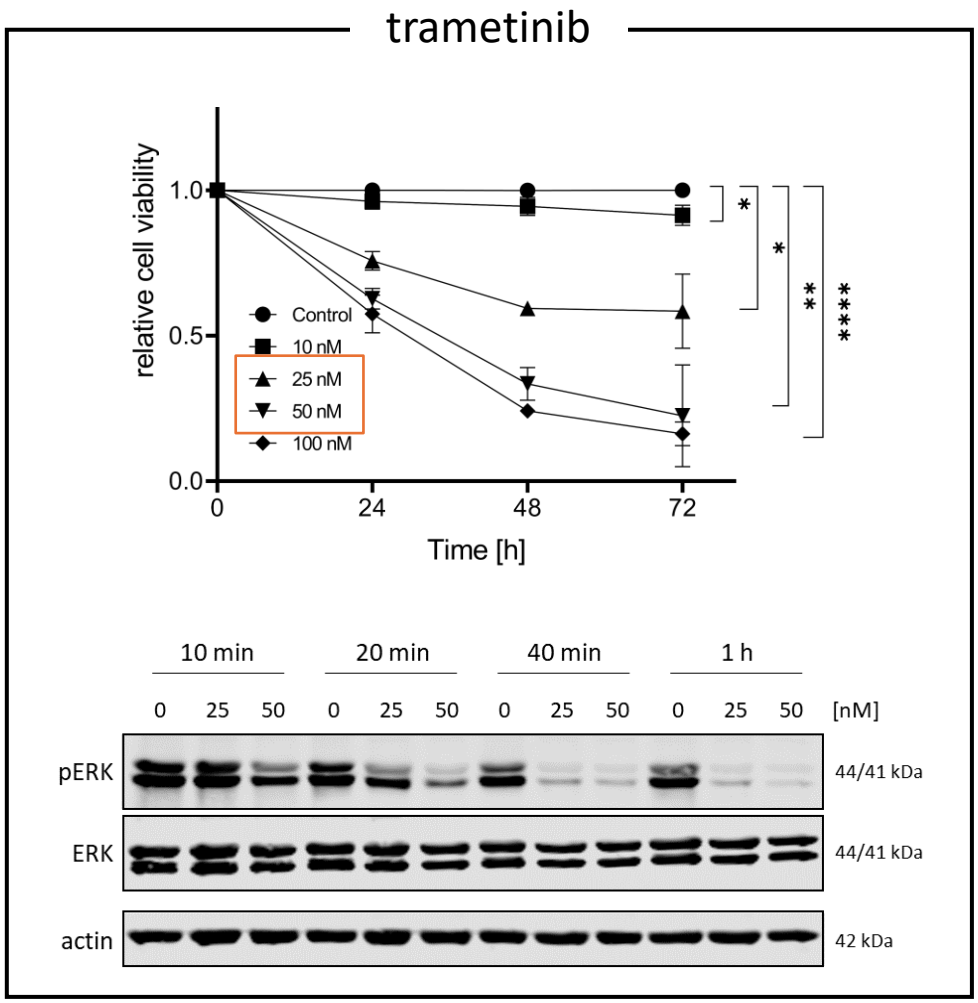

Figure 10. The effect of trametinib inhibition on Hep3B HCC cell line. *Four different doses were used for viability testing with a DMSO control. The doses with medium viability decrease were selected. The effective time-dependent ERK and phospho-ERK inhibition was confirmed with Western blotting. Actin was used as internal loading control. The same experiments were performed with MK-8353, binimetinib and ulixertinib and with the NSCLC cell line A549. Brown-Forsythe ANOVA was performed on viability results to assess significance. These experiments were conducted to determine the conditions achieving the strongest MAPK pathway inhibition to later use these conditions to create MAPK pathway inhibited samples for lncRNA sequencing. Brown-Forsythe ANOVA was used to test the significance of viability testing.*

To assess the transcriptional effect of the MAPK pathway inhibitions on the cells, RT-qPCR was performed on A549 and Hep3B cells using all the four pre-selected drugs to detect the MAPK target mRNAs DUSP6 and SPRY2 [114], showing effective downregulation starting from 3 hours of treatment. Thus, the conditions of the most effective MAPK pathway inhibition were the following: *a)* cell lines A549 and Hep3B, *b)* trametinib 25nM, MK-8353 300 nM, *c)* treatment time of 3 hours. Samples were prepared with these treatment conditions and sent to lncRNA sequencing.

4.3.2. LncRNAseq data analysis

On the pre-processed RNAseq data, classical data analysis and lncRNA candidate selection was performed. There was an effective downregulation of MAPK pathway-associated genes and lower MAPK Pathway Activity Score (MPAS) [115] supporting that the MAPK inhibition worked. Thus, it is rightful to assume that the downregulated lncRNAs may also be regulated by the MAPK pathway. 35 commonly downregulated lncRNAs were found among HCC and NSCLC cell lines and ERK and MEK inhibition (see Figure 11), possibly acting as biomarkers of MAPK pathway activity.

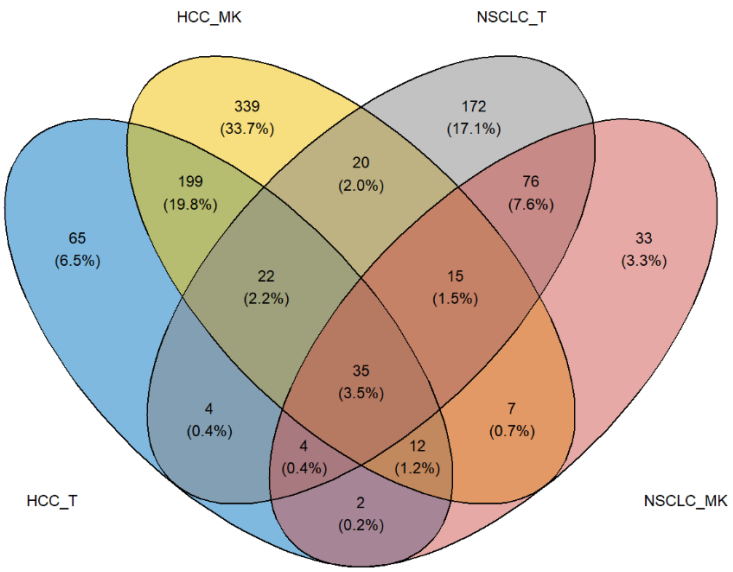

Figure 11. Overlapping downregulated lncRNAs among cancer types and treatments. *RNA sequencing was performed using the A549 NSCLC and the Hep3B HCC cell lines and two MAPK pathway inhibitors. After pre-processing and confirming MAPK pathway inhibition, lncRNAs downregulated in all samples were identified and defined as MAPK pathway activity lncRNA biomarkers.*

4.3.3. Experimental confirmation of lncRNA candidates

To confirm these findings, primers were designed for the lncRNA candidates, and for those where working primers could be designed, confirmatory RT-qPCRs were performed to detect their downregulation (see Figure 12, names hidden on purpose due to the possibility of a patent). These results were confirmed on the previously used four different cell lines and treatments. The effect of ERK inhibition on the lncRNA panel was also confirmed with siRNA inhibition of ERK1 and ERK2. Starvation experiments also

confirmed these findings - MAPK pathway inhibition stopped the growth factor caused increase of the lncRNA panel expression.

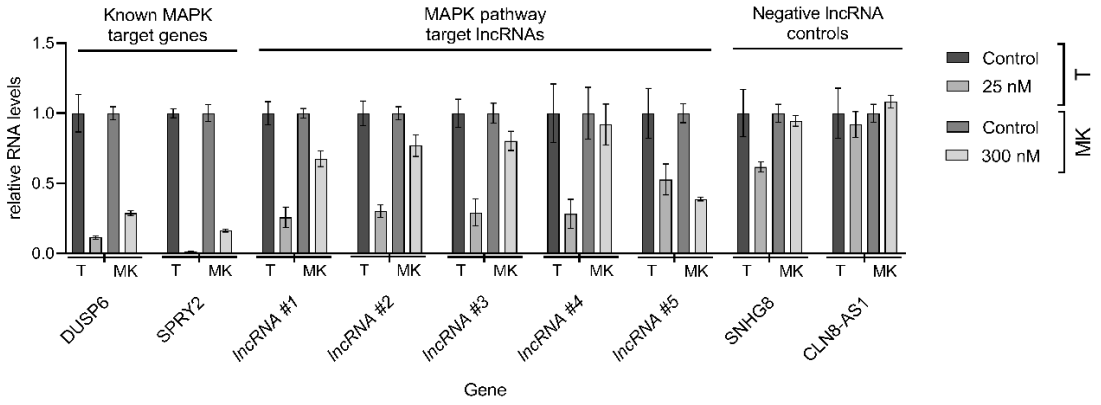

Figure 12. Confirmation RT-qPCR of the lncRNA biomarker panel. *Names are hidden due to the possibility of a patent. This experiment confirms the downregulation of the lncRNA candidates due to MAPK pathway inhibition compared to negative controls. This experiment was repeated for the four cell lines and four treatments, sixteen replicates altogether.*

The pathway specificity of the lncRNA panel was confirmed using the AKTVIII Akt-inhibitor and YAP siRNA inhibition. These results showed no downregulation, supporting the MAPK pathway specificity of these lncRNAs. 16 known transcription factors regulated by the MAPK pathway [116] were tested to find which one is responsible for the downregulation of our lncRNAs. No transcription factors could be identified for some lncRNAs, but for example, NFAT4 [117] and STAT3 [118] were shown to regulate *lncRNA #1*. Altogether, a MAPK-pathway specific lncRNA-panel was identified, with potential further use to evaluate clinical effectivity of anti-MAPK treatment.

4.3.4. Survival analysis on TCGA data

The prognostic value of our lncRNA biomarker panel was evaluated on publicly available HCC and NSCLC data. The results shown here are partially based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga [119]. The results showed that tumours have significantly higher expression than healthy tissues, however, the survival difference of the high and low expressing patients were not significant (see Figure 13).

However, there is no information about the treatments of these patients, so we cannot evaluate the effectivity of this panel as a predictive biomarker for MAPK inhibition

treatment based on this data – and our goal was to establish a predictive lncRNA panel, not necessary a prognostic one. Besides, higher expression in tumours compared to normal tissues may help detectability with liquid biopsy.

This panel needs further validation on patient samples and animal experiments to establish the specificity, sensitivity and overall detectability of these biomarkers, and to assess if these lncRNAs could be detected in peripheral blood samples too, which would make them feasible for long-term therapeutic decision making.

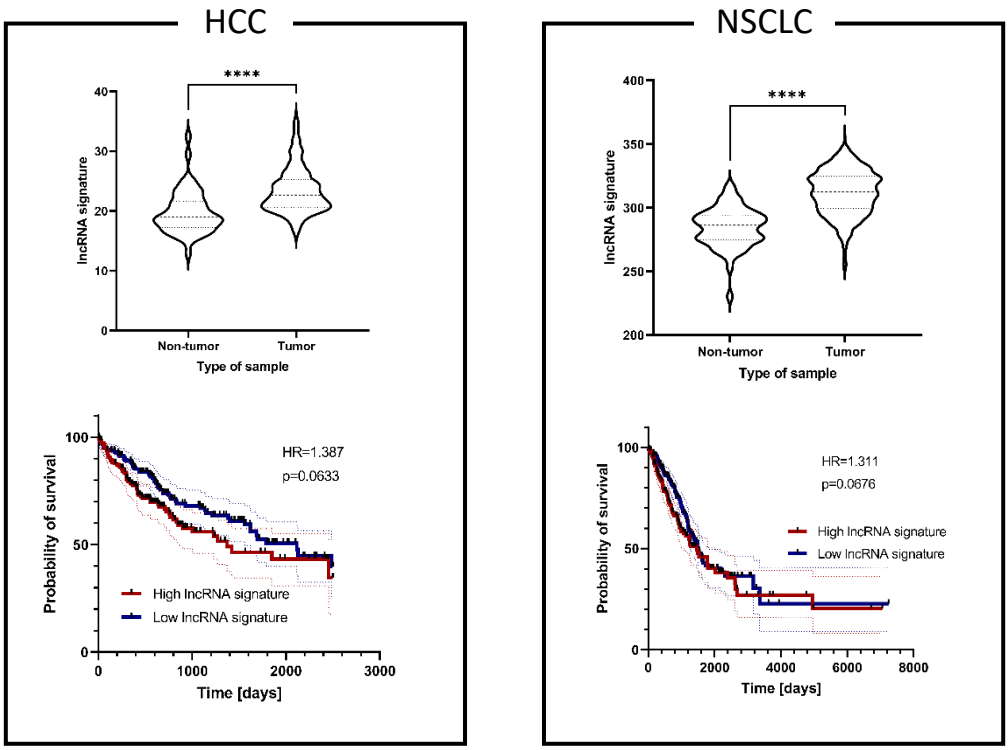

Figure 13. The expression and prognostic value of the MAPK lncRNA signature in HCC and NSCLC TCGA data. *For the lncRNAs found in the TCGA HCC and NSCLC datasets, violin plots are showing their expression in non-tumour versus tumour samples. The survival of the highly and lowly expressing patients is visualised on Kaplan-Meier curves. Mann-Whitney tests were used to assess significance.*

5. Discussion

Even though the era of polyhistors is long gone, science still has a huge need for effective information exchange between the experts of different methodologies, such as experimental and computational scientists and clinical researchers. However, one often solely specialises in one or the other disciplines, leading to separate conversations within these fields. As computational researchers use the data produced by experimental teams, and clinicians rely on all these fields for effective drug development, we think that it is important to develop the general understanding to be able to ask questions when faced with the research of others. However, this is generally missing between many disciplines, thus limiting the cross-applicability of findings. In this work, different approaches were used to work on problems related to signalling changes in cancer, with their own strengths and limitations, combined into a more complex understanding of altered signalling in malignancies.

Looking back at the results of the colon cancer development modelling study, these results show interesting and previously never captured global network changes between normal colon, colon adenoma and colon carcinoma. Even though previous network studies analysed diameter changes between normal and cancer networks [11], they did not analyse precancerous states alongside with normal and cancer samples. The reduction in network diameter from normal to carcinoma has been repeatedly reported across multiple network topological modelling approaches both before and after publishing this study. Studies based on differential co-expression [120, 121], association-type [122], interaction networks [123-125], and ceRNA networks [126] all describe a shift toward more compact, shorter-path networks in carcinoma. Notably, Wen et al. [123] reported a similar decrease in diameter using miRNA networks built by overlaying expression data onto a literature-derived network skeleton, a strategy conceptually aligned with our approach. In addition, a comparative multi-cancer analysis identified colon cancer as one of the tumour types in which the transition from normal to carcinoma is associated with a shrinking diameter [11].

Studies published after our study also describe definite network topology changes during carcinogenesis with numerous modelling techniques with various methodologies. A single-cell transcriptomic study showed a denser T-cell co-expression network in colon

cancer versus normal tissues [127]. This study also showed an increased number of cell-cell interactions in colon adenoma and carcinoma epithelial cells. However, no network diameter was calculated from these. In contrast, a study published in 2025 calculated network diameter on multiple cancer-specific protein-protein interaction networks, showing that they have larger diameters than random power-law graphs with similar node and edge counts [128]. A lung cancer RNAseq-based immune correlation network topology analysis study also showed larger network diameter in normal compared with cancer networks on one sample set, while an opposite pattern was observed on the samples in another dataset [129]. Lung adenocarcinoma topology values, including network diameter, were closer to normal tissue than lung squamous cell carcinoma, which showed smaller network diameter on bulk data. This study also contained network analysis of a flow cytometry cell sorting dataset, confirming that CD45+ immune cells, CD10+ fibroblasts, and even CD31+ endothelial cells have larger network diameter than cancer cells [129]. The exact diameter values were 11-12 for immune cells, 6-8 for fibroblasts, 6 for endothelial cells and 5 for lung cancer, showing that immune cells have much less compact signalling networks [129]. This raises interest in comprehensive analysis of network diameter differences between cell types and cancer histology subtypes. It might be worthwhile to further assess the normal tissue with higher network diameter, if they may contain inflammation or precancerous tissue. In the absence of precancerous networks to compare in these studies, our study still stays unique with the conclusion of de-compacting then re-compacting of signalling networks during carcinogenesis.

The newly described higher network diameter of the colon adenoma model was never described before in literature. It is the sign of a less regulated, less compact transitional signalling network, which lost the old structure, but it is not under the harsh environmental pressure cancer cells face. Similarly, the decrease in the EGFR pathway significance during adenoma might be a sign of EGFR-independent proliferation, while losing the role of EGFR signalling in healthy cells [84], similarly to the MAPK pathway in the highly proliferating colon endothelium [88]. The gradual increase of the VEGFR pathway's expression might be attributed to the hypoxic conditions in colon adenoma and carcinoma [130]. However, these pathways would need further, more detailed analysis to subdivide the exact cross-talks and processes taking place. Additionally, another network

study confirmed the importance of other pathways such as Wnt signalling and matrix metalloproteases, well-known in colon carcinogenesis, while constructing networks from differently expressed genes of colon adenoma and carcinoma [12]. As a feedback of the methodology, these classic pathways should be also analysed with our model.

One of the limitations of this study is the lack of patient-level resolution or the comparison of patients with different mutational profile or outcome. As mainly left-sided colon cancer samples were included, the attributes of right-sided, immune-governed colon cancer is also missing from this analysis, which may have different results especially in the mismatch repair pathway analysis due to its frequent mismatch repair deficiency [86]. As quality patient- and even cell-level expression data is more accessible today, these network topology analyses could be repeated with patient- or even cell-level resolution, contributing to the recent high resolution network topology studies [127, 129], but concentrating more on the discovered network diameter phenomenon. A systematic high-resolution network topology study could be one of the future works stemming from these results, stratifying patients based on global network attributes, and finding association with cancer subtype or outcome. Analysing other cancer types with well-known precancerous states would also contribute to this study, and it would allow the assessment of whether the adenoma-related changes are a colon-cancer-specific, or a general phenomenon. Another future scope might be to analyse these topological changes during different treatments, such as EGFR- or VEGFR-inhibitors, to assess therapy-dependent global network changes. Using RNAseq data instead of microarray in the future would also be beneficial due to the increased accuracy of the method [131].

To move the focus of the network analysis approach from a global to a more local perspective, three-nodal motif members were analysed as potential target-biomarker pairs in the MarkerPredict project [44]. As to our current knowledge, MarkerPredict is the first network-topology-based predictive biomarker prediction tool in the literature. However, recently published methods exist, which combine network topology with machine learning for diagnostic and prognostic biomarker prediction [132-134]. Our results, like the cited articles, showed that the number of triangles was among the most important input features in machine learning prediction, together with network centrality values. In both cases, large numbers representing higher connectivity pushed the model towards predicting the given protein as predictive biomarker, confirming the hypothesised central

role of biomarkers, similar to the also central IDPs [34]. This was not as simple for the other hypothesis, as even though IDPs were enriched in the motifs of therapeutic targets, higher IDP scores and disorder content percent did not mean more probable biomarker prediction in most of the score system, however, for the IUPred short score, which was developed to detect short disorder regions, higher score meant more likely to be predicted as a predictive biomarker [65]. Supporting these claims, even the top predictions by the model have low to moderate disorder content. We hypothesise that this might be because the presence of the non-disordered regions allows the protein to participate in key signalling processes, while the presence of disordered regions allows conformational changes in adaptation processes, like in a cancer cell under therapeutic pressure. When confirmed by other methods, this can lead to a new understanding of the role of disorder in cell signalling: to the hypothesis that partially disordered proteins are the key regulators of adaptation processes. Due to the limitations of our solely bioinformatic study, this would require further experimental confirmation, for example comparing the effects of more- and less-disordered variants of the same proteins.

Overall, multiple particularly effective binary classification machine learning models were developed, which passed validation steps with high accuracy, and learns on meaningful features. We theorize that the interconnectivity of the networks may facilitate model training. The strong AUC values may also partly reflect the use of protein pairs rather than individual proteins as predictors, increasing the dimensionality and diversity of the feature space and potentially enabling decision tree–based models to identify informative feature combinations more effectively. Similar pairwise approaches have been applied in previous studies with acceptable but not outstanding metrics; usually not limiting the search on network neighbours [135, 136]. Finally, 2084 neighbour pairs of therapeutic targets, such as YWHAZ to BRAF and MEK1/2 were predicted to be good predictive biomarker candidates of the respective drugs. One of the limitations of this study is that strongly regulated, but more downstream pathway effectors like DUSP6 and SPRY2 in the MAPK pathway cannot be predicted as this study is limited to three-nodal motifs. Thus, extending the study to larger motif structures is an option. However, strongly co-regulated neighbouring proteins would be more useful to assess intrinsic therapy resistance. Nonetheless, the promising biomarker predictions require further systematic literature review and validation, as many of them are already established or

show promise to be a good predictive biomarker for cancer therapeutics. For these promising candidates, the next step would be conducting experimental confirmation followed by clinical studies, to overcome the limitations of bioinformatics and result in clinically useful longitudinal decision support systems.

To conduct detailed signalling network analysis, an established network structure is necessary, derived from plenty of experimental data confirming each interaction [50]. For example, during the lncRNA MAPK signature project, it was not possible to perform network modelling on the lncRNA candidates, as there was not enough information on their function and interaction, even though preliminary lncRNA networks exist [137]. With open access data, the prognostic value and tumour specificity of these candidates could be analysed (see the TCGA data, https://www.cancer.gov/tcga; [119]). To validate the predictive value of this lncRNA panel, it is also necessary to detect these candidates in animal and patient samples, showing that they are indeed as sensitive, specific and detectable biomarkers in real-life conditions as on cell lines. This work was started with preliminary *in situ* hybridization experiments, however, it is yet to be finished as the funding has expired. Another potential limitation of this study is that there were no studies done on the function of these lncRNAs, as it is not required for their potential use in therapeutic decision making. Instead, focusing on extending this straightforward lncRNA signature identification pipeline, it is possible to define selective biomarker panels for other key signalling pathways targeted in oncology. This work was started by the candidate for the PI3K pathway, however, for technical reasons, good downstream readouts of the pathway activity could not be identified. However, as a preceding study, a yet unpublished YAP-pathway activity dependent lncRNA signature was already established by Fabian Rose, a PhD student of the Breuhahn laboratory. He also developed many techniques used in this study [138]. Overall, an unprecedented list of pan-cancer MAPK-pathway-specific lncRNAs were identified, waiting for further experimental confirmation and peer review. Here we conclude that despite their limitations, different bioinformatic and experimental methods all led to effective identification of key cancer signalling changes and biomarkers.

6. Conclusions

To summarise, here we showed that experimental and *in silico* methods are both helpful to identify key signalling changes and cancer biomarkers. However, while experimental research typically requires a previously well-established question, the high-throughput characteristic of network science and machine learning allowed us to perform mass identification of promising predictive cancer biomarker candidates.

1.) Key signalling changes such as the less-organised then re-organised information flow were identified with the modelling of colon cancer development, represented by growing then decreasing network diameter and average shortest path length in the normal-adenoma-carcinoma sequence.

2.) MarkerPredict, a pan-cancer machine learning biomarker prediction tool was created using network topology data and protein annotation, and tested with various cross-validation methods, showing high performance. A Biomarker Probability Score ranking predictions of 32 different models identified 2084 potential cancer biomarkers.

3.) MarkerPredict showed that more co-regulatory motifs a neighbour-target pair participates in, the more likely the neighbour can be a predictive biomarker for the target's inhibitor. High centrality values are also attributes of cancer biomarkers. IDPs with low to moderate disorder content can be effective biomarkers.

4.) The identified 2084 predictive biomarker candidates show promise in clinical use, thus require further research and clinical validation to finally help therapeutic decision-making and personalised oncology. Their potential was shown through the example of the 14-3-3ζ protein, a potential biomarker of MAPK pathway inhibitors.

7. Summary

Cancer is a disease where malignantly transformed cells use the available resources for self-replication instead of their original functions, evading immune response and therapeutic efforts. Therapy resistance is one of the largest challenges in oncology today, leading to unnecessary treatments and larger mortality. Signalling changes marking therapy resistance in cancer can be studied in multiple ways. In this work, we used *in silico* approaches to model cancer signalling in carcinogenesis and therapy resistance, and both *in vitro* and *in silico* methods to identify predictive biomarkers.

Network topology analysis was used to model colon cancer development, using open-source RNA microarray data to create normal, adenoma and carcinoma network models from a pre-published network skeleton. The results show a larger network diameter indicating loosened signalling structure in colon adenoma, and a smaller network diameter pointing to strictly organised signalling in colon carcinoma. While the VEGFR pathway becomes more expressed with carcinogenesis, the EGFR and MAPK pathway models have higher weighted degree and median abundance in normal network models than in the adenoma network models.

Network motifs are small, frequently present network structures indicating strong co-regulation of the participating nodes. Here, we showed that intrinsically disordered proteins, difficult to target, but central players in signalling networks, are enriched in the triangle motifs containing onco-therapeutic targets. Using network topology features such as motif information and protein annotations including IDP databases, a powerful biomarker prediction tool called MarkerPredict was established on target-neighbour pairs. Validation showed high performance, and 2084 potential biomarkers were predicted using the established Biomarker Probability Score. These predictions are shown to be highly relevant, waiting for further validation before clinical use.

LncRNAs are promising biomarker candidates, as they can not only be detected in tumours, but in bodily fluids like peripheral blood, promising minimal-invasive therapeutic decision-support. A pan-cancer MAPK-pathway-specific lncRNA signature was established with RNA sequencing on lung and liver cancer cell lines with effective but not lethal MAPK pathway inhibition, ready for validation on patient samples.

8. References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100(1):57-70.
2. Gerlinger M, McGranahan N, Dewhurst SM, Burrell RA, Tomlinson I, Swanton C. Cancer: evolution within a lifetime. Annu Rev Genet. 2014;48:215-36.
3. Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. 2022;12(1):31-46.
4. Janes KA, Lazzara MJ. Systems Biology of the Cancer Cell. Annu Rev Biomed Eng. 2025;27(1):1-28.
5. Plotnikov A, Zehorai E, Procaccia S, Seger R. The MAPK cascades: signaling components, nuclear roles and mechanisms of nuclear translocation. Biochim Biophys Acta. 2011;1813(9):1619-33.
6. Ravnan MC, Matalka MS. Vemurafenib in patients with BRAF V600E mutation-positive advanced melanoma. Clin Ther. 2012;34(7):1474-86.
7. Costigan DC, Dong F. The extended spectrum of RAS-MAPK pathway mutations in colorectal cancer. Genes Chromosomes Cancer. 2020;59(3):152-9.
8. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. Cell. 1990;61(5):759-67.
9. Peddareddigari VG, Wang D, Dubois RN. The tumor microenvironment in colorectal carcinogenesis. Cancer Microenviron. 2010;3(1):149-66.
10. Smit WL, Spaan CN, Johannes de Boer R, Ramesh P, Martins Garcia T, Meijer BJ, et al. Driver mutations of the adenoma-carcinoma sequence govern the intestinal epithelial global translational capacity. Proc Natl Acad Sci U S A. 2020;117(41):25560-70.
11. Rahman KMT, Islam MF, Banik RS, Honi U, Diba FS, Sumi SS, Kabir S, Akhter MD. Changes in protein interaction networks between normal and cancer conditions: Total chaos or ordered disorder? Network Biology. 2013;3(1):15.

12. Jiang Y, Song F, Hu X, Guo D, Liu Y, Wang J, Jiang, L., Huang, P., Zhang, Y. Analysis of dynamic molecular networks: the progression from colorectal adenoma to cancer. J Gastrointest Oncol. 2021;12(6):2823-37.

13. Schulc K, Nagy ZT, Kamp S, Molnar J, Veres DV, Csermely P, Kovács BM. Modular Reorganization of Signaling Networks during the Development of Colon Adenoma and Carcinoma. J Phys Chem B. 2021;125(7):1716-26.

14. Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;13(10):714-26.

15. Asic K. Dominant mechanisms of primary resistance differ from dominant mechanisms of secondary resistance to targeted therapies. Crit Rev Oncol Hematol. 2016;97:178-96.

16. Lei ZN, Tian Q, Teng QX, Wurpel JND, Zeng L, Pan Y, Chen ZS. Understanding and targeting resistance mechanisms in cancer. MedComm (2020). 2023;4(3):e265.

17. Kerestely M, Keresztes D, Szarka L, Kovacs BM, Schulc K, Veres DV, Csermely P. System level network data and models attack cancer drug resistance. Br J Pharmacol. 2025:1–13.

18. Keresztes D, Kerestely M, Szarka L, Kovacs BM, Schulc K, Veres DV, Csermely P. Cancer drug resistance as learning of signaling networks. Biomed Pharmacother. 2025;183:117880.

19. Lim SK, Lee MH, Park IH, You JY, Nam BH, Kim BN, Ro J, Lee KS, Jung SY, Kwon YM, Lee ES. Impact of Molecular Subtype Conversion of Breast Cancers after Neoadjuvant Chemotherapy on Clinical Outcome. Cancer Res Treat. 2016;48(1):133-41.

20. Veres T, Kerestely M, Kovacs BM, Keresztes D, Schulc K, Seitz E, Seitz E, Vassy Z, Veres DV, Csermely P. Cellular forgetting, desensitisation, stress and ageing in signalling networks. When do cells refuse to learn more? Cell Mol Life Sci. 2024;81(1):97.

21. Zhou Y, Tao L, Qiu J, Xu J, Yang X, Zhang Y, Tian X, Guan X, Cen X, Zhao Y. Tumor biomarkers for diagnosis, prognosis and targeted therapy. Signal Transduct Target Ther. 2024;9(1):132.

22. de Sousa EMF, Colak S, Buikhuisen J, Koster J, Cameron K, de Jong JH, et al. Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. Cell Stem Cell. 2011;9(5):476-85.

23. Senna KM, Zimmermann IR, Costa M, Tura BR, Evangelista J, Santos M. Cost-effectiveness assessment of liquid biopsy for early detection of lung cancer in Brazil. PLoS One. 2025;20(7):e0328631.

24. Ou FS, Michiels S, Shyr Y, Adjei AA, Oberg AL. Biomarker Discovery and Validation: Statistical Considerations. J Thorac Oncol. 2021;16(4):537-45.

25. Kramer A, van Schaik LF, van den Broek D, Meijer GA, Gutierrez Ibarluzea I, Galnares Cordero L, et al. Towards Recommendations for Cost-Effectiveness Analysis of Predictive, Prognostic, and Serial Biomarker Tests in Oncology. Pharmacoeconomics. 2025;43(5):483-97.

26. Kulkarni V, Kulkarni P. Intrinsically disordered proteins and phenotypic switching: Implications in cancer. Prog Mol Biol Transl Sci. 2019;166:63-84.

27. Santofimia-Castano P, Rizzuti B, Xia Y, Abian O, Peng L, Velazquez-Campoy A, Neira JL, Iovanna J. Targeting intrinsically disordered proteins involved in cancer. Cell Mol Life Sci. 2020;77(9):1695-707.

28. Kulkarni P, Uversky VN. Cancer/Testis Antigens: "Smart" biomarkers for diagnosis and prognosis of prostate and other cancers. Int J Mol Sci. 2017;18(4):740.

29. Palazzo AF, Koonin EV. Functional Long Non-coding RNAs Evolve from Junk Transcripts. Cell. 2020;183(5):1151-61.

30. Beylerli O, Gareev I, Sufianov A, Ilyasova T, Guang Y. Long noncoding RNAs as promising biomarkers in cancer. Noncoding RNA Res. 2022;7(2):66-70.

31. Bao X, Hu Q, Ji P, Lin W, Kurths J, Nagler J. Impact of basic network motifs on the collective response to perturbations. Nat Commun. 2022;13(1):5301.

32. Zhang XD, Song J, Bork P, Zhao XM. The exploration of network motifs as potential drug targets from post-translational regulatory networks. Sci Rep. 2016;6:20558.

33. Szalay-Beko M, Palotai R, Szappanos B, Kovacs IA, Papp B, Csermely P. ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality. Bioinformatics. 2012;28(16):2202-4.

34. Rangarajan N, Kulkarni P, Hannenhalli S. Evolutionarily conserved network properties of intrinsically disordered proteins. PLoS One. 2015;10(5):e0126729.

35. C L, S P, Kashyap AH, Rahaman A, Niranjan S, Niranjan V. Novel biomarker prediction for lung cancer using random forest classifiers. Cancer Inform. 2023;22:11769351231167992.

36. Acharjee A, Larkman J, Xu Y, Cardoso VR, Gkoutos GV. A random forest based biomarker discovery and power analysis framework for diagnostics research. BMC Med Genomics. 2020;13(1):178.

37. Toth R, Schiffmann H, Hube-Magg C, Buscheck F, Hoflmayer D, Weidemann S, et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. Clin Epigenetics. 2019;11(1):148.

38. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA. New York, NY, USA: Association for Computing Machinery; 2016. p. 785-94.

39. Pedregosa F, V. G, al. Scikit-learn: machine learning in Python. Journal of Machine Learning Research. 2011;12:2825-30.

40. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017:pp. 4768 - 77.

41. Charilaou P, Battat R. Machine learning models and over-fitting considerations. World J Gastroenterol. 2022;28(5):605-7.

42. Iwagami M, Matsui H. Introduction to Clinical Prediction Models. Ann Clin Epidemiol. 2022;4(3):72-80.

43. Dietterich TG, editor Ensemble Methods in Machine Learning. Multiple Classifier Systems; 2000 2000//; Berlin, Heidelberg: Springer Berlin Heidelberg.

44. Veres DV, Csermely P, Schulc K. MarkerPredict: predicting clinically relevant predictive biomarkers with machine learning. NPJ Syst Biol Appl. 2025;11(1):132.

45. Guo YJ, Pan WW, Liu SB, Shen ZF, Xu Y, Hu LL. ERK/MAPK signalling pathway and tumorigenesis. Exp Ther Med. 2020;19(3):1997-2007.

46. Schulc K. lncRNA signatures as biomarkers for the activity of oncogenic signaling pathways. Trilaterales Symposium; 09.24.; Heidelberg 2024.

47. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5(2):101-13.

48. Signorini LF, Kupiec M, Sharan R. Signing protein-protein interaction networks. Bioinformatics. 2026;42(1).

49. Cui Q, Ma Y, Jaramillo M, Bari H, Awan A, Yang S, et al. A map of human cancer signaling. Mol Syst Biol. 2007;3:152.

50. Lo Surdo P, Iannuccelli M, Contino S, Castagnoli L, Licata L, Cesareni G, Perfetto L. SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. Nucleic Acids Res. 2023;51(D1):D631-D7.

51. Wu G, Dawson E, Duong A, Haw R, Stein L. ReactomeFIViz: a Cytoscape app for pathway and network-based data analysis. F1000Res. 2014;3:146.

52. Wernicke S, Rasche F. FANMOD: a tool for fast network motif detection. Bioinformatics. 2006;22(9):1152-3.

53. UniProt C. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D31.

54. Skrzypczak M, Goryca K, Rubel T, Paziewska A, Mikula M, Jarosz D, Pachlewski J, Oledzki J, Ostrowski J. Modeling oncogenic signaling in colon tumors by multidirectional analyses of microarray data directed for maximization of analytical reliability. PLoS One. 2010;5(10):e13091.

55. Galamb O, Gyorffy B, Sipos F, Spisak S, Nemeth AM, Miheller P, Tulassay Z, Dinya E, Molnár B.. Inflammation, adenoma and cancer: objective classification of

colon biopsy specimens with gene expression signature. Dis Markers. 2008;25(1):1-16.

56. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. Mol Cancer Res. 2007;5(12):1263-75.

57. Valcz G, Patai AV, Kalmar A, Peterfia B, Furi I, Wichmann B, Műzes G, Sipos F, Krenács T, Mihály E, Spisák S, Molnár B, Tulassay Z. Myofibroblast-derived SFRP1 as potential inhibitor of colorectal carcinoma field effect. PLoS One. 2014;9(11):e106143.

58. Aric A. Hagberg DASaPJS. Exploring network structure, dynamics, and function using NetworkX. In: Gäel Varoquaux TV, and Jarrod Millman, editor. Proceedings of the 7th Python in Science Conference; Pasadena, CA USAAug 2008. p. 11-5.

59. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498-504.

60. Dijkstra E. A note on two problems in connexion with graphs, Numerisce Mathematik 1. 1 Numerishe Mathematik. 1959;1:269–71.

61. Freeman LC. A Set of Measures of Centrality Based on Betweenness. Sociometry. 1977;40(1):35-41.

62. Giscard P-L, R. P, Wilson RC. Evaluating balance on social networks from their simple cycles. Journal of Complex Networks. 2017;5:750-75.

63. Quaglia F, Meszaros B, Salladini E, Hatos A, Pancsa R, Chemes LB, et al. DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation. Nucleic Acids Res. 2022;50(D1):D480-D7.

64. Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 2022;50(D1):D439-D44.

65. Erdos G, Pajkos M, Dosztanyi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. Nucleic Acids Res. 2021;49(W1):W297-W303.

66. Holt ME, Mittendorf KF, LeNoue-Newton M, Jain NM, Anderson I, Lovly CM, Osterman T, Micheel C, Levy M. My Cancer Genome: Coevolution of precision oncology and a molecular oncology knowledgebase. JCO Clin Cancer Inform. 2021;5:995-1004.

67. Sheils TK, Mathias SL, Kelleher KJ, Siramshetty VB, Nguyen DT, Bologa CG, et al. TCRD and Pharos 2021: mining the human proteome for disease biology. Nucleic Acids Res. 2021;49(D1):D1334-D46.

68. Mendez D, Gaulton A, Bento AP, Chambers J, De Veij M, Felix E, et al. ChEMBL: towards direct deposition of bioassay data. Nucleic Acids Res. 2019;47(D1):D930-D40.

69. Lever J, Jones MR, Danos AM, Krysiak K, Bonakdar M, Grewal JK, Culibrk L, Griffith OL, Griffith M, Jones SJM. Text-mining clinically relevant cancer biomarkers for curation into the CIViC database. Genome Med. 2019;11(1):78.

70. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365.

71. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One. 2015;10(3):e0118432.

72. Giard DJ, Aaronson SA, Todaro GJ, Arnstein P, Kersey JH, Dosik H, Parks WP. In vitro cultivation of human tumors: establishment of cell lines derived from a series of solid tumors. J Natl Cancer Inst. 1973;51(5):1417-23.

73. Mitsudomi T, Steinberg SM, Nau MM, Carbone D, D'Amico D, Bodner S, et al. p53 gene mutations in non-small-cell lung cancer cell lines and their correlation with the presence of ras mutations and clinical features. Oncogene. 1992;7(1):171-80.

74. Nakabayashi H, Taketa K, Miyano K, Yamane T, Sato J. Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. Cancer Res. 1982;42(9):3858-63.

75. Knowles BB, Howe CC, Aden DP. Human hepatocellular carcinoma cell lines secrete the major plasma proteins and hepatitis B surface antigen. Science. 1980;209(4455):497-9.

76. Gilmartin AG, Bleam MR, Groy A, Moss KG, Minthorn EA, Kulkarni SG, et al. GSK1120212 (JTP-74057) is an inhibitor of MEK activity and activation with favorable pharmacokinetic properties for sustained in vivo pathway inhibition. Clin Cancer Res. 2011;17(5):989-1000.

77. Bendell JC, Javle M, Bekaii-Saab TS, Finn RS, Wainberg ZA, Laheru DA, et al. A phase 1 dose-escalation and expansion study of binimetinib (MEK162), a potent and selective oral MEK1/2 inhibitor. Br J Cancer. 2017;116(5):575-83.

78. Germann UA, Furey BF, Markland W, Hoover RR, Aronov AM, Roix JJ, et al. Targeting the MAPK Signaling Pathway in Cancer: Promising Preclinical Activity with the Novel Selective ERK1/2 Inhibitor BVD-523 (Ulixertinib). Mol Cancer Ther. 2017;16(11):2351-63.

79. Moschos SJ, Sullivan RJ, Hwu WJ, Ramanathan RK, Adjei AA, Fong PC, et al. Development of MK-8353, an orally administered ERK1/2 inhibitor, in patients with advanced solid tumors. JCI Insight. 2018;3(4):e92352.

80. Weiler SME, Bissinger M, Rose F, von Bubnoff F, Lutz T, Ori A, Schirmacher P, Breuhahn K. SEPTIN10-mediated crosstalk between cytoskeletal networks controls mechanotransduction and oncogenic YAP/TAZ signaling. Cancer Letters. 2024;584:216637.

81. Amini H, Lelarge M. The diameter of weighted random graphs. Ann Appl Probab. 2015;25(3):1686-727.

82. Cervantes A, Adam R, Rosello S, Arnold D, Normanno N, Taieb J, Yoshino T, Wyrwicz L, Ducreux M; ESMO Guidelines Committee. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34(1):10-32.

83. Gene Ontology C. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015;43(Database issue):D1049-56.

84. Tang X, Liu H, Yang S, Li Z, Zhong J, Fang R. Epidermal Growth Factor and Intestinal Barrier Function. Mediators Inflamm. 2016;2016:1927348.

85. Flora M, Piana S, Bassano C, Bisagni A, De Marco L, Ciarrocchi A, et al. Epidermal growth factor receptor (EGFR) gene copy number in colorectal adenoma-carcinoma progression. Cancer Genet. 2012;205(12):630-5.

86. Keshinro A, Ganesh K, Vanderbilt C, Firat C, Kim JK, Chen CT, et al. Characteristics of Mismatch Repair-Deficient Colon Cancer in Relation to Mismatch Repair Protein Loss, Hypermethylation Silencing, and Constitutional and Biallelic Somatic Mismatch Repair Gene Pathogenic Variants. Dis Colon Rectum. 2023;66(4):549-58.

87. Ágg B, Császár A, Szalay-Bekő M, Veres DV, Mizsei R, Ferdinandy P, Csermely P, Kovács IA. The EntOptLayout Cytoscape plug-in for the efficient visualization of major protein complexes in protein-protein interaction and signalling networks. Bioinformatics (Oxford, England). 2019 Nov;35(21):4490-4492.

88. Zhang Z, Zhang Q, Li F, Xin Y, Duan Z. Contributions of HO-1-Dependent MAPK to Regulating Intestinal Barrier Disruption. Biomol Ther (Seoul). 2021;29(2):175-83.

89. Lam KK, Tang CL, Tan E, Wong SH, Cheah PY. KRAS mutation-independent downregulation of MAPK/PI3K signaling in colorectal cancer. Mol Oncol. 2022;16(5):1171-83.

90. Li F, Zhang L, Feng F, Zheng K, Li Y, Wang T, Ren G. Livin participates in resistance to trastuzumab therapy for breast cancer through ERK1/2 and AKT pathways and promotes EMT-like phenotype. RSC Adv. 2018;8(50):28588-601.

91. Pizzamiglio S, Ciniselli CM, Triulzi T, Gargiuli C, De Cecco L, de Azambuja E, et al. Integrated Molecular and Immune Phenotype of HER2-Positive Breast Cancer and Response to Neoadjuvant Therapy: A NeoALTTO Exploratory Analysis. Clin Cancer Res. 2021;27(23):6307-13.

92. Gu X, Huang Q, Huang Y, Ding D, Zhong J, Lin S, et al. Identification of dynamic network biomarker ITGB1 for erlotinib pre-resistance using single-cell differential covariance entropy. Mol Ther Oncol. 2025;33(2):200993.

93. Matsuda Y, Yamauchi T, Hosono N, Uzui K, Negoro E, Morinaga K, Nishi R, Yoshida A, Kimura S, Maekawa T, Ueda T. Combination of panobinostat with ponatinib synergistically overcomes imatinib-resistant CML cells. Cancer Sci. 2016;107(7):1029-38.

94. Yang G, Ding C, Yang X, Jiang J, He S, Shao Y, et al. NDRG1 enhances the sensitivity to Cetuximab by promoting Stat1 ubiquitylation in colorectal cancer. J Adv Res. 2025;72:555-69.

95. Srivastava RM, Trivedi S, Concha-Benavente F, Hyun-Bae J, Wang L, Seethala RR, Branstetter BF, Ferrone S, Ferris RL. STAT1-Induced HLA Class I Upregulation Enhances Immunogenicity and Clinical Response to Anti-EGFR mAb Cetuximab Therapy in HNC Patients. Cancer Immunol Res. 2015;3(8):936-45.

96. King AJ, Arnone MR, Bleam MR, Moss KG, Yang J, Fedorowicz KE, et al. Dabrafenib; preclinical characterization, increased efficacy when combined with trametinib, while BRAF/MEK tool combination reduced skin lesions. PLoS One. 2013;8(7):e67583.

97. Liu L, Cao Y, Chen C, Zhang X, McNabola A, Wilkie D, Wilhelm S, Lynch M, Carter C. Sorafenib blocks the RAF/MEK/ERK pathway, inhibits tumor angiogenesis, and induces tumor cell apoptosis in hepatocellular carcinoma model PLC/PRF/5. Cancer Res. 2006;66(24):11851-8.

98. Pandya K, Wyatt D, Gallagher B, Shah D, Baker A, Bloodworth J, et al. PKCalpha Attenuates Jagged-1-Mediated Notch Signaling in ErbB-2-Positive Breast Cancer to Reverse Trastuzumab Resistance. Clin Cancer Res. 2016;22(1):175-86.

99. Itah Z, Chaudhry S, Raju Ponny S, Aydemir O, Lee A, Cavanagh-Kyros J, Tournier C, Muller WJ, Davis RJ. HER2-driven breast cancer suppression by the JNK signaling pathway. Proc Natl Acad Sci U S A. 2023;120(4):e2218373120.

100. Xie C, Patel R, Wu T, Zhu J, Henry T, Bhaskarabhatla M, et al. PI3K/AKT/mTOR hypersignaling in autoimmune lymphoproliferative disease engendered by the epistatic interplay of Sle1b and FASlpr. Int Immunol. 2007;19(4):509-22.

101. Kim T, Park BS, Heo S, Jeon H, Kim J, Kim D, Kook Lee S, Jung SY, Kong SY, Lu T. Combinatorial CRISPR screen reveals FYN and KDM4 as targets for synergistic drug combination for treating triple negative breast cancer. Elife. 2025;13:RP93921.

102. Stivala S, Codilupi T, Brkic S, Baerenwaldt A, Ghosh N, Hao-Shen H, et al. Targeting compensatory MEK/ERK activation increases JAK inhibitor efficacy in myeloproliferative neoplasms. J Clin Invest. 2019;129(4):1596-611.

103. Dummer R, Ramelyte E, Schindler S, Thurigen O, Levesque MP, Koelblinger P. MEK inhibition and immune responses in advanced melanoma. Oncoimmunology. 2017;6(8):e1335843.

104. Kwong LN, Boland GM, Frederick DT, Helms TL, Akid AT, Miller JP, et al. Co-clinical assessment identifies patterns of BRAF inhibitor resistance in melanoma. J Clin Invest. 2015;125(4):1459-70.

105. Crisan D, Schatz S, Hainzl H, Zhao F, Paschen A, Crisan M, Baican A, Scharffetter-Kochanek K, Basu A. GSK3beta activation is a key driver of resistance to Raf inhibition in BRAF mutant melanoma cells. Oncotarget. 2025;16:257-9.

106. Fernandez ML, Dawson A, Hoenisch J, Kim H, Bamford S, Salamanca C, et al. Markers of MEK inhibitor resistance in low-grade serous ovarian cancer: EGFR is a potential therapeutic target. Cancer Cell Int. 2019;19:10.

107. Turturro SB, Najor MS, Yung T, Portt L, Malarkey CS, Abukhdeir AM, Cobleigh MA. Somatic loss of PIK3R1 may sensitize breast cancer to inhibitors of the MAPK pathway. Breast Cancer Res Treat. 2019;177(2):325-33.

108. Morrison DK. The 14-3-3 proteins: integrators of diverse signaling cues that impact cell fate and cancer development. Trends Cell Biol. 2009;19(1):16-23.

109. Tzivion G, Avruch J. 14-3-3 proteins: active cofactors in cellular regulation by serine/threonine phosphorylation. J Biol Chem. 2002;277(5):3061-4.

110. Sun Q, Wang W. Structures of BRAF-MEK1-14-3-3 sheds light on drug discovery. Signal Transduct Target Ther. 2019;4:59.

111. Gan Y, Ye F, He XX. The role of YWHAZ in cancer: A maze of opportunities and challenges. J Cancer. 2020;11(8):2252-64.

112. Pennington KL, Chan TY, Torres MP, Andersen JL. The dynamic and stress-adaptive signaling hub of 14-3-3: emerging mechanisms of regulation and context-dependent protein-protein interactions. Oncogene. 2018;37(42):5587-604.

113. Qin C, Wang YL, Zheng J, Wan XB, Fan XJ. Current perspectives in drug targeting intrinsically disordered proteins and biomolecular condensates. BMC Biol. 2025;23(1):118.

114. Joseph EW, Pratilas CA, Poulikakos PI, Tadi M, Wang W, Taylor BS, et al. The RAF inhibitor PLX4032 inhibits ERK signaling and tumor cell proliferation in a V600E BRAF-selective manner. Proc Natl Acad Sci U S A. 2010;107(33):14903-8.

115. Wagle MC, Kirouac D, Klijn C, Liu B, Mahajan S, Junttila M, et al. A transcriptional MAPK Pathway Activity Score (MPAS) is a clinically relevant biomarker in multiple cancer types. NPJ Precis Oncol. 2018;2(1):7.

116. Wagner EF, Nebreda AR. Signal integration by JNK and p38 MAPK pathways in cancer development. Nat Rev Cancer. 2009;9(8):537-49.

117. Macian F. NFAT proteins: key regulators of T-cell development and function. Nat Rev Immunol. 2005;5(6):472-84.

118. Tkach M, Rosemblit C, Rivas MA, Proietti CJ, Diaz Flaque MC, Mercogliano MF, et al. p42/p44 MAPK-mediated Stat3Ser727 phosphorylation is required for progestin-induced full activation of Stat3 and breast cancer growth. Endocr Relat Cancer. 2013;20(2):197-212.

119. Tomczak K, Czerwinska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol (Pozn). 2015;19(1A):A68-77.

120. Deng SP, Zhu L, Huang DS. Predicting Hub Genes Associated with Cervical Cancer through Gene Co-Expression Networks. IEEE/ACM Trans Comput Biol Bioinform. 2016;13(1):27-35.

121. Kugler KG, Mueller LA, Graber A, Dehmer M. Integrative network biology: graph prototyping for co-expression cancer networks. PLoS One. 2011;6(7):e22843.

122. Lee YS, Hwang SG, Kim JK, Park TH, Kim YR, Myeong HS, Kwon K, Jang CS, Noh YH, Kim SY. Topological network analysis of differentially expressed genes in cancer cells with acquired gefitinib resistance. Cancer Genomics Proteomics. 2015;12(3):153-66.

123. Wen J, Hall B, Shi X. A network view of microRNA and gene interactions in different pathological stages of colon cancer. BMC Med Genomics. 2019;12(Suppl 7):158.

124. Verma Y, Yadav A, Katara P. Mining of cancer core-genes and their protein interactome using expression profiling based PPI network approach. Gene Reports. 2020;18:100583.

125. Schramm G, Kannabiran N, Konig R. Regulation patterns in signaling networks of cancer. BMC Syst Biol. 2010;4:162.

126. Zhou M, Wang X, Shi H, Cheng L, Wang Z, Zhao H, Yang L, Sun J. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer. Oncotarget. 2016;7(11):12598-611.

127. Song C. Single-cell transcriptomic reveals network topology changes of cancer at the individual level. Comput Biol Chem. 2025;117:108401.

128. Fadhal E. Network Topology Similarities Across Cancer Types: Identifying Central Protein Hubs for Drug Discovery. OBM Genetics. 2025;9:1-34.

129. Kim KS, Kang T, Jekarl DW. Immune Gene Networks from Lung Cancer Patients Treated with Immune Checkpoint Inhibitors. Biomedicines. 2024;12(3):628.

130. Greijer AE, Delis-van Diemen PM, Fijneman RJ, Giles RH, Voest EE, van Hinsbergh VW, Meijer GA. Presence of HIF-1 and related genes in normal mucosa, adenomas and carcinomas of the colorectum. Virchows Arch. 2008;452(5):535-44.

131. Rai MF, Tycksen ED, Sandell LJ, Brophy RH. Advantages of RNA-seq compared to RNA microarrays for transcriptome profiling of anterior cruciate ligament tears. J Orthop Res. 2018;36(1):484-97.

132. Keikha F, Wang C, Yang Z, Liu ZP. TransMarker: Unveiling dynamic network biomarkers in cancer progression through cross-state graph alignment and optimal transport. PLoS Comput Biol. 2025;21(11):e1013743.

133. Liu Y, Huse J, Kannan K. Expression graph network framework for biomarker discovery. Brief Bioinform. 2025;26(5):bbaf559.

134. Dhillon A, Singh A, Bhalla V. Biomarker identification and cancer survival prediction using random spatial local best cat swarm and Bayesian optimized DNN. Applied Soft Computing. 2023;146:110649.

135. Kim S, Lin CW, Tseng GC. MetaKTSP: a meta-analytic top scoring pair method for robust cross-study validation of omics prediction analysis. Bioinformatics. 2016;32(13):1966-73.

136. Kaur P, Schlatzer D, Cooke K, Chance MR. Pairwise protein expression classifier for candidate biomarker discovery for early detection of human disease prognosis. BMC Bioinformatics. 2012;13:191.

137. Guo Q, Liu Q, He D, Xin M, Dai Y, Sun R, et al. LnCeCell 2.0: an updated resource for lncRNA-associated ceRNA networks and web tools based on single-cell and spatial transcriptomics sequencing data. Nucleic Acids Res. 2025;53(D1):D107-D15.

138. Pellegrino R, Castoldi M, Ticconi F, Skawran B, Budczies J, Rose F, et al. LINC00152 Drives a Competing Endogenous RNA Network in Human Hepatocellular Carcinoma. Cells. 2022;11(9).

9. Bibliography of the candidate's publications

1. **Schulc K**, Nagy ZT, Kamp S, Molnar J, Veres DV, Csermely P, Kovács BM. Modular Reorganization of Signaling Networks during the Development of Colon Adenoma and Carcinoma. J Phys Chem B. 2021;125(7):1716-26. , IF= 2.9, Q1
2. Veres DV, Csermely P, **Schulc K***. MarkerPredict: predicting clinically relevant predictive biomarkers with machine learning. NPJ Systems Biology Applications 2025;11(1):132. , IF= 3.5, D1

   ** Corresponding author*

3. Kerestely M, Keresztes D, Szarka L, Kovacs BM, **Schulc K**, Veres DV, Csermely P. System level network data and models attack cancer drug resistance. Br J Pharmacol. 2025:1–13., IF= 7.7, D1
4. Keresztes D, Kerestely M, Szarka L, Kovacs BM, **Schulc K**, Veres DV, Csermely P. Cancer drug resistance as learning of signaling networks. Biomedical Pharmacotherapy. 2025; 183:117880. , IF= 7.5, D1
5. Veres T, Kerestely M, Kovacs BM, Keresztes D, **Schulc K**, Seitz E, Vassy Z, Veres DV, Csermely P. Cellular forgetting, desensitisation, stress and ageing in signalling networks. When do cells refuse to learn more? Cell Mol Life Sci. 2024;81(1):97. , IF= 6.2, D1
6. **Schulc K**. lncRNA signatures as biomarkers for the activity of oncogenic signaling pathways. Trilaterales Symposium; 09.24.; Heidelberg 2024.

10. Acknowledgements

I thank my wonderful supervisors Prof. Péter Csermely and Dániel Veres for introducing me to science and being my partners in generating innovative ideas. I am grateful for the LINK Group community for providing a platform of professional development. I am thankful for Borbála Kovács for the collaborative work on the colon carcinogenesis project, especially her work of highlighting the modular changes during cancer development. Within that project, I thank Sebestyén Kamp and János Molnár (Turbine Ltd.) for preparing the microarray dataset, and for Zsolt Nagy for developing the diameter calculation script. I thank Gábor Kovács (Turbine Ltd.) for reviewing my codes for the MarkerPredict package. I am grateful for the Kerpel-Fronius Ödön Talent Support Program and the Jellinek Harry fellowship for allowing me to study for 10 months at the University of Heidelberg and learn many things about experimental biology. I thank Prof. Kai Breuhahn and the Breuhahn lab for guiding me in the wet lab experiments. The experiments in the lncRNA project were partially performed by Patrizia Birner, our excellent lab technician. Finally, I thank my wonderful family and friends for supporting me along the way.