D8.2

Network models for molecular target identification

Project number	826121
Project acronym	iPC
Project title	individualizedPaediatricCure: Cloud-based virtual-patient models for precision paediatric oncology
Start date of the project	1st January, 2019
Duration	53 months
Programme	H2020-SC1-DTH-2018-1

Deliverable type	Report
Deliverable reference number	SC1-DTH-07-826121 / D8.2/ 1.0
Work package contributing to the deliverable	WP8
Due date	31st May, 2022
Actual submission date	30th May, 2022

Responsible organisation	UKL-HD
Editor	Attila Gabor, Julio Saez-Rodriguez
Dissemination level	PU
Revision	V1.0

Abstract	We focused on the development of patient specific signalling networks using prior knowledge about the molecular events and CRISPR perturbation datasets and associated the activity of the nodes of signaling network with drug response data to find molecular targets.
Keywords	Signaling network; Mechanistic modelling; CRISPR; drug response

The project iPC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826121.

Editor

Attila Gabor (UKL-HD)

Julio Saez-Rodriguez (UKLHD)

Contributors (ordered according to beneficiary numbers)

Rosa Hernansaiz Ballesteros (UKL-HD)

Disclaimer

Executive Summary

D8.2 provides a detailed overview on how causal inference can be used to find potential drug targets in cancer cell lines and in pediatric cancer patients samples.

D8.2 discusses how to derive patient specific signaling network from gene expression and genomics data

D8.2 provides guidelines how to associate the signaling network activity with drug sensitivity and CRISPR-based gene essentiality

D8.2 provides a list of biomarkers (active signaling proteins) that are found to positively respond to treatments with specific drugs

D8.2 showcase a web based tool (FUNKI), which provides an easy to use, graphical interface to footprint-based omics analysis tools.

Table of Content

Chapter 1 Introduction 1

Chapter 2 Association of personalised signalling networks with drug response and CRISPR essentiality scores 3

2.1 Methods 3

2.1.1 Processing of CRISPR data 3

2.1.2 Generation of CARNIVAL input data 3

2.1.3 Processing of CARNIVAL networks 4

2.2 Results 5

2.2.1 Replacement of PROGENy-based weights with CRISPR-based ones 5

2.2.2 Association of node activities with drug sensitivity data 5

2.2.3 Association of node activities with gene essentiality data 7

2.3 Discussion 8

Chapter 3 Association of contextualised networks with drug sensitivity data in patients' samples 10

3.1 Introduction 10

3.2 Results 10

Chapter 4 FUNKI: Interactive functional footprint-based analysis of omics data 14

4.1 Introduction 14

4.2 Features 14

4.3 Conclusion 15

Chapter 5 Summary and Conclusion 16

Chapter 6 Bibliography 17

List of Figures

Figure 1: Inference of cell line specific signalling networks and association to drug sensitivity data 1
Figure 2: The association of supplied weights and resulting node activities 5
Figure 3. Associations of nodes and drug response. 6
Figure 4: The association of EPAS1 and SCH772984 6
Figure 5: Associations of node activity and knockouts 7
Figure 6: Association of ETV4 node activity and ARPC4 essentiality 8
Figure 7: A) Progeny pathway signature – drug response correlation. Pearson correlation coefficient between the drug score and the pathway activity. B) Correlation matrix between transcription factors and sDSSasym score 11
Figure 8: Overview of the patient specific CARNIVAL analysis workflow. 12
Figure 9: CARNIVAL topology network top associations with drug response data 13
Figure 10: Graphical overview of analysis and visualization features provided by FUNKI. FUNKI provides a user interface to upload omics data, and then run DoRothEA, PROGENy, KinAct, CARNIVAL and COSMOS to estimate the activity of pathways, TFs and kinases. This figure was published in Hernansaiz-Ballesteros et al. 2022 15

List of Abbreviations

Abbreviation	Translation
CARNIVAL	CAusal Reasoning for Network identification using Integer VALue programming
DoRothEA	Discriminant Regulon Expression Analysis
FUNKI	FUNctional analysis toolKIt
INDEL	Insertion and Deletions
MANOVA	multivariate analysis of variance
PKN	Prior knowledge network
PROGENy	Pathway RespOnsive GENes
SNV	Single Nucleotide Variants

Chapter 1 Introduction

The aim of this deliverable is to report on the results of finding molecular targets in individuals using network-based methods.

Cancer is often studied in biological model systems such as immortalised cell lines. These in vitro cell line models capture the characteristics of different tumour types, resample tumour heterogeneity and have provided a fertile ground for cancer research. Nowadays, many cancer cell lines are well characterised by their gene expression, metabolite abundance and genomic variations (van der Meer et al. 2018). Further, the cell lines' response to drug perturbations (Yang et al. 2012) and CRISPR-technology based dropout (Pacini et al. 2021) is also available.

Further, we have curated multi-omics and drug response profiling data from the INFORM DSP (INdividualized Therapy FOr Relapsed Malignancies in Childhood, functional Drug Response Profiling) dataset with a freezing date (15.01.2022). This dataset included 74 patient samples that passed the screening quality control criteria, where the drug response was quantified using the selective drug sensitivity score (ElHarouni et al 2022).

Traditional statistical associations between the cancer sample features (gene expression, genomic variations etc.) and drug sensitivity or gene knockout (Behan et al. (2019)) can identify potential targets on the cancer cells, however, these pure statistical associations are prone to find non-causal interactions. Using prior knowledge, the same data can be used to derive mechanistic patterns which provide better insights into the state of said biological systems. These patterns require sophisticated methods and a framework of gene sets, pathways or even complete signaling networks.

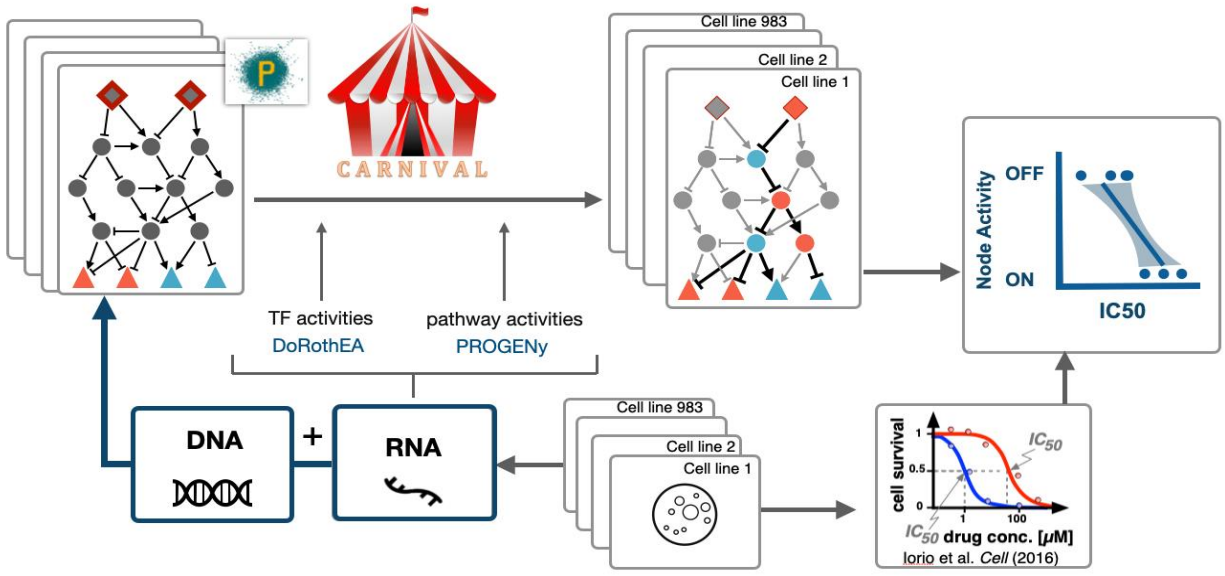

Figure 1: Inference of cell line specific signalling networks and association to drug sensitivity data

To this end, a plethora of tools was developed in recent years: For instance, over-representation analysis (ORA) accumulates the information about multiple genes in an unordered gene set to make statements about general biological processes (Boyle et al. 2004). However, these approaches share the limitation that they use gene expression as a proxy for protein activity. Thereby, they disregard post-translational modifications, especially phosphorylations that can dramatically change the activity of a protein (Schubert et al. 2018).

Signature-based methods that use network knowledge to predict transcription factors (TF: DoRothEA) or pathway activity (PROGENy) that cause the observed gene expression data (Dugourd and Saez-Rodriguez 2019) can overcome the above-mentioned limitations. These methods describe more accurately the pathways and expression programs that drive the state of a biological system, in particular the progression of a tumour or susceptibility to a treatment (Figure 1). In order to further elucidate the topology of signalling processes of an expression profile, the causal inference tools CARNIVAL (Liu et al. 2019) have been developed. Using the inferred TF activities, a defined perturbation and a signed and directed protein interactions network (or simply referenced as prior knowledge network PKN), CARNIVAL generates a set of contextualised networks that explain the transmission of a perturbation signal to downstream TFs. Finally, the cell line specific, contextualised networks are associated with drug perturbations, which can reveal potential drug targets.

Chapter 2 Association of personalised signalling networks with drug response and CRISPR essentiality scores

We used CARNIVAL (Liu et al 2019) to contextualise the transcriptomic data available from Cancer Dependency Map from Sanger and Broad Institutes to create signalling networks. In particular, we compare the influence of two different PKNs, a general one that contains all protein-protein interactions (PKN) reported in Omnipath database (Turei et al 2016), and a personalized PKN specific to the cell, by removing the non-expressed genes and accounting for mutations. Furthermore, we supply different types of weights to the network optimization to increase the expressiveness of the resulting signalling networks. Finally, said signalling networks are correlated to drug response and CRISPR knockout data in order to gain insights in how the inferred activities of a given gene is connected to the sensitivity to a drug or knockout. Ultimately, we aim to identify a set of genes whose activity determines the response of cancer cells to certain drugs or knockouts. Hence, genes that could potentially act as predictive biomarkers for therapeutic interventions. When experimentally validated, these insights can help advance precision cancer therapy and treat more individual patients based on their individual tumour.

2.1 Methods

We used publicly available data, namely gene expression, mutation, CNA and metadata from Cell Model Passports. In addition, integrated CRISPR essentiality scores were obtained from Pacini et al. (2021) and drug sensitivity data was downloaded from the Genomics of Drug Sensitivity in Cancer Website.

2.1.1 Processing of CRISPR data

CRISPR essentiality scores are obtained from Pacini et al. 2021 and cell lines with missing data were removed. CRISPR essentiality scores represent fold changes of proliferation between cells without and with a particular gene present. In total, CRISPR scores are available for 17487 genes and 897 cell lines.

We utilise the CRISPR essentiality scores as alternative weights to guide the generation of contextualized networks. Due to the nature of the objective function of CARNIVAL (Liu et al. 2019), only weights within the interval of $[-1; 1]$ should be used. PROGENy scores fulfil this condition by design. To comply with this constraints, the CRISPR scores are scaled to this interval by scaling with the most extreme value.

2.1.2 Generation of CARNIVAL input data

There are certain inputs that CARNIVAL requires to run:

1. TF activity measurements: From the publicly available gene expression data, z-scores are calculated for each gene to represent a DEG between a particular sample and the overall mean of all samples. These z-scores are supplied to DoRothEA to infer TF activities for each sample in relation to the population mean.

2. Prior knowledge network (PKN): The underlying protein interaction network is downloaded from OmniPath, a meta-database that stores prior knowledge from a variety of external sources (Türei, Korcsmáros, and Saez-Rodriguez 2016). Interactions in the prior knowledge network were filtered based on gene expression and mutation.
3. Perturbations/input nodes: These are inferred from the PKN itself. As the network is directed, we identified all nodes with an outgoing and no incoming edge and used this subset as perturbations.
4. Weights (optional): The pathway-based weights were generated using PROGENy and the same gene expression z-scores as DoRothEA. The CRISPR-based weights were generated as described above.

CARNIVAL uses the proprietary CPLEX solver for the optimization problem. For better reproducibility, CPLEX runs in deterministic mode and the sequence of inputs, PKN edges and TF activities are randomized using ten different, reproducible seeds and ten CARNIVAL networks are generated based on that.

CARNIVAL uses a weighted network in the optimization for which we examined the following sets of configurations:

1. no weights: The control includes only the mandatory inputs of CARNIVAL and no weights at all.
2. pscores: The first examined configuration includes the 33 progeny-derived weights (progeny scores) as in the original CARNIVAL workflow.
3. CRISPR-progeny: The second examined configuration replaces progeny scores directly by CRISPR-derived weights, thus 33 weights for the same nodes are given.
4. CRISPR-CS-top33: The third examined configuration selects 33 weights with extreme values (17 positive and 16 negative) for context-specific (CS) essential genes.

The last configuration requires additional subsetting of the 17487 available CRISPR scores for which the context-specific essential genes defined in Behan et al. (2019) was chosen.

Apart from the different weights that are supplied in these configurations, also different PKNs are evaluated: The general PKN contains 50725 known signed and directed interactions between human proteins. Additionally, there are specific PKNs for every cell line which contain only genes that are expressed and not mutated in the specific model. Depending on the cell line, the specific PKNs have a reduced size between 35177 and 45087 interactions.

2.1.3 Processing of CARNIVAL networks

CARNIVAL infers up to 100 solution networks that connect the measured TF activities with the given perturbations. From them, it creates a weighted summary network which essentially unites the nodes and edges from all solutions. The “weighted activities” are then based on the number of fraction of solutions that contained a certain node or edge.

In order to summarize the ten randomized versions of each configuration, the edges and node tables are merged and the means of node- and edge-activities calculated. The result is used to correlate node activity with drug sensitivity and gene knockouts.

Analogous to the approach of the GDSC project, which correlated genetic and genomic traits to drug sensitivity, we perform ANOVA between two groups of samples to compare their respective mean drug IC50 values. The two groups each contain samples where a specific node is either up- or downregulated.

2.2 Results

2.2.1 Replacement of PROGENy-based weights with CRISPR-based ones

The first objective was to implement and assess the usage of CRISPR-based gene essentiality scores as alternative to PROGENy-based weights to guide the generation of CARNIVAL signalling networks.

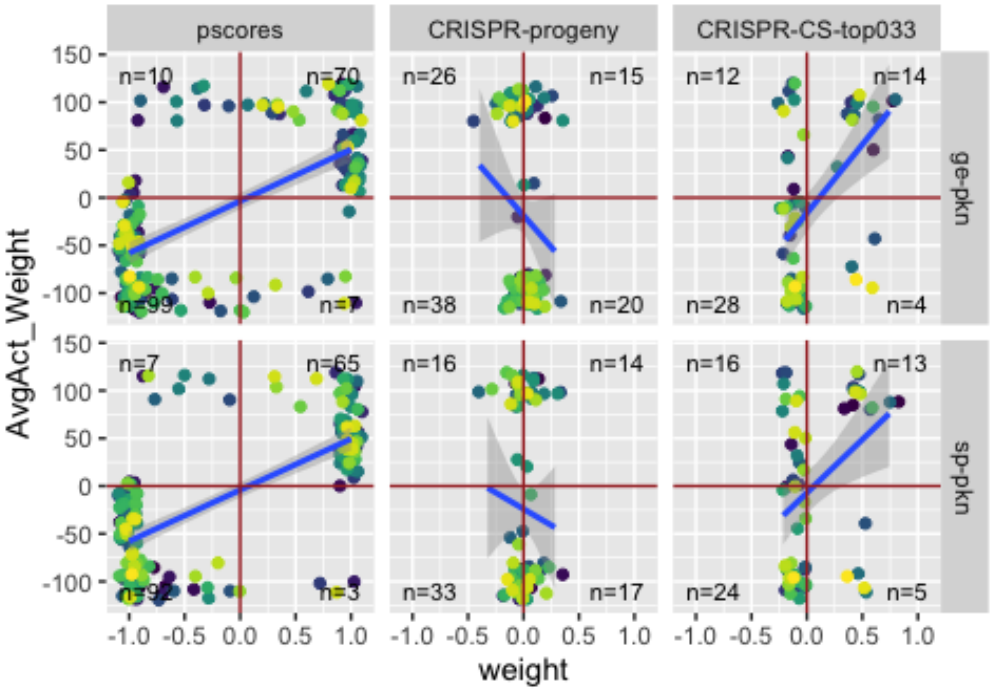

Figure 2: The association of supplied weights and resulting node activities

Across the investigated configurations, the results diverge widely. For the PROGENy-based pathway scores, the correlation coefficients of almost 0.6 and the favoured fraction of more than 90% indicate that the supplied weights have an impact on the weighted nodes and accordingly on the generation of the signalling networks. However, for the CRISPR-based weights, correlation coefficients are closer to zero and favoured fractions closer to 0.5 (relates to equal fractions).

In summary, the PROGENy-based pathway scores have clearly the most impact on the network generation and a bare replacement of them with CRISPR-derived weights performs poorly. However, an equally sized individual set of extreme CRISPR essentiality scores for every model can have some impact on the optimization. This highlights that the optional weights should ideally use a larger range within the given interval of $[-1; 1]$ and that the used transformation, while preserving the differences between genes and models, might not be optimal to maximize the weight importance.

2.2.2 Association of node activities with drug sensitivity data

In order to gain more insights in how cell-line-specific signalling networks and activity of certain nodes relate to drug sensitivity, the CARNIVAL results are associated with the CancerRxgene dataset (Yang et al. 2012). This dataset contains IC50 values for hundreds of drugs administered to hundreds of cell lines.

We associated the average node activities to drug sensitivity. In particular, the node activity and drug sensitivity data is joined by the cell line identifiers and the node activities discretized to groups of up-regulated (average node activity $a_{avg} > 10$) and down-regulated ($a_{avg} < -10$). Comparisons with few samples ($N < 3$ per group) are considered unreliable and thus excluded.

The results of the computed ANOVAs can be visualized for the different configurations as volcano plots (Figure 3). Therein, significant and thus interesting combinations of drugs and nodes appear close to the top corners.

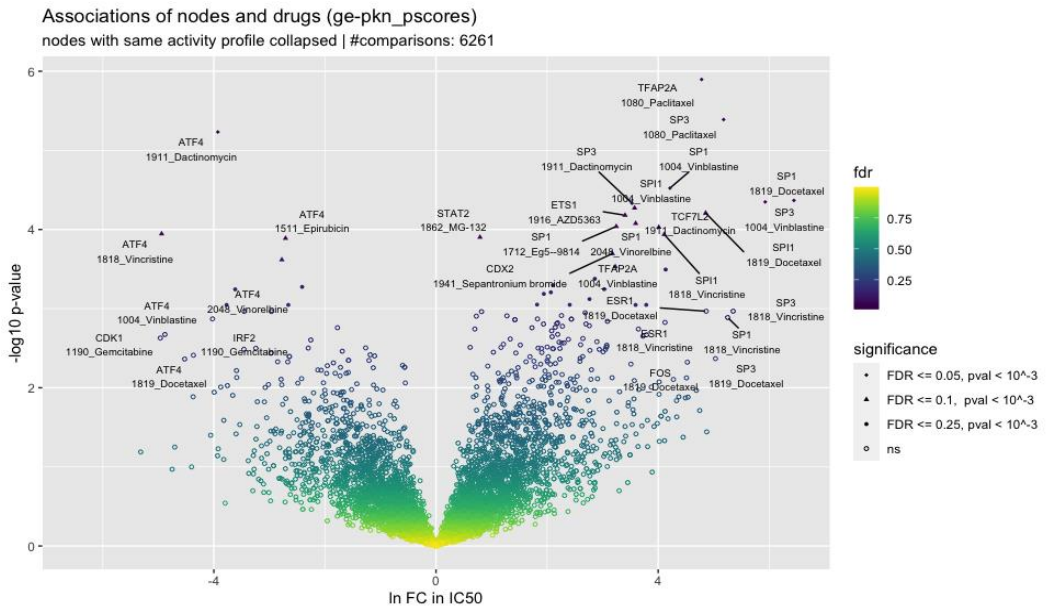

Figure 3. Associations of nodes and drug response.

To further assess the validity of the reported associations, the groups of cell lines that are characterised by a specific drug-node-combination are further analyzed. In total, 84 associations are significant in at least one configuration (type of PKN as explained above), but only 32 associations are significant in at least two configurations.

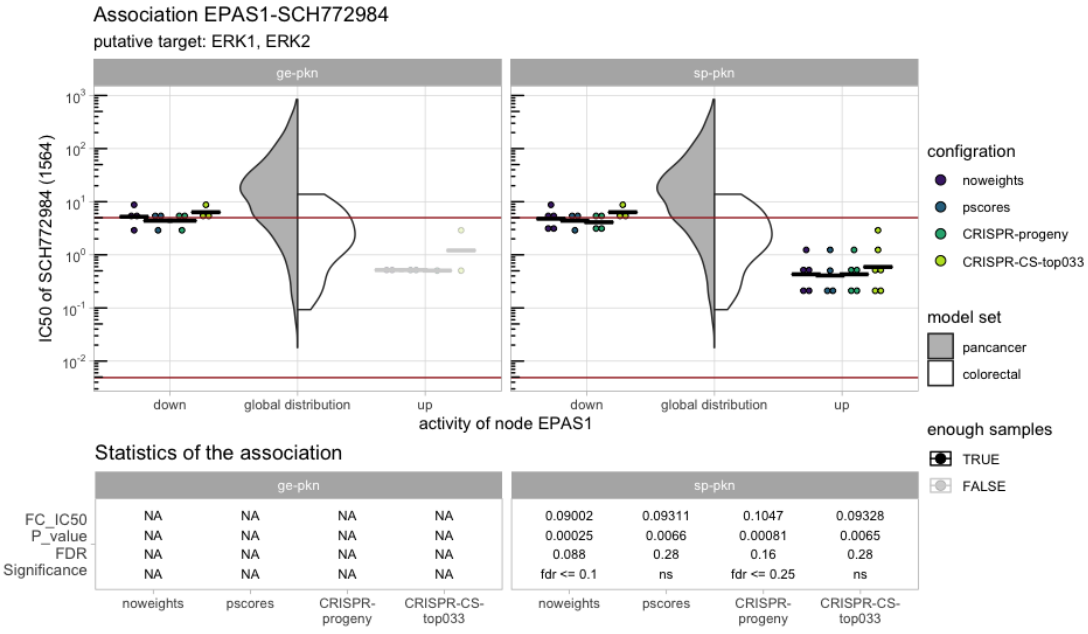

Figure 4: The association of EPAS1 and SCH772984

For example, we observed positive relationship between EPAS1 activity and sensitivity to SCH772984 (Figure 4). Since the cells with upregulated EPAS1 are more sensitive to SCH772984 treatment than other pancancer cell lines, the gene could also qualify as a predictive marker for sensitivity to the treatment. SCH772984 is a specific inhibitor of ERK1/2, thus it blocks MAPK signalling and proliferation in cancers resistant to inhibitors of upstream proteins BRAF and MEK

(Morris et al. 2013). EPAS1 is a part of the transcription factor HIF-2 and responsible for cellular response to low-oxygen environments. This qualifies the protein as a potential therapeutic target in cancer (Patel and Simon 2008). Interestingly, EPAS1 is phosphorylated and thus activated by ERK1/2 in hypoxic conditions (Gkotinakou et al. 2019). When it acts as a driver of angiogenesis and cancer progression, treatment with SCH772984 could have a strong negative effect on cancer cells.

2.2.3 Association of node activities with gene essentiality data

Similarly to the drug sensitivity data, to investigate how the activity of certain nodes relates to changes in gene dependencies, the node activities are associated with genome-wide CRISPR knockout screening data (Pacini 2021). Since the dataset contains essentiality scores of more than 17.000 genes, the gene dependency data is subset to include only the known context-specific genes in large intestine tumours (Behan et al. 2019), that were already used to select the CRISPR-based weights. The results are again visualized for the separate configurations as volcano plots.

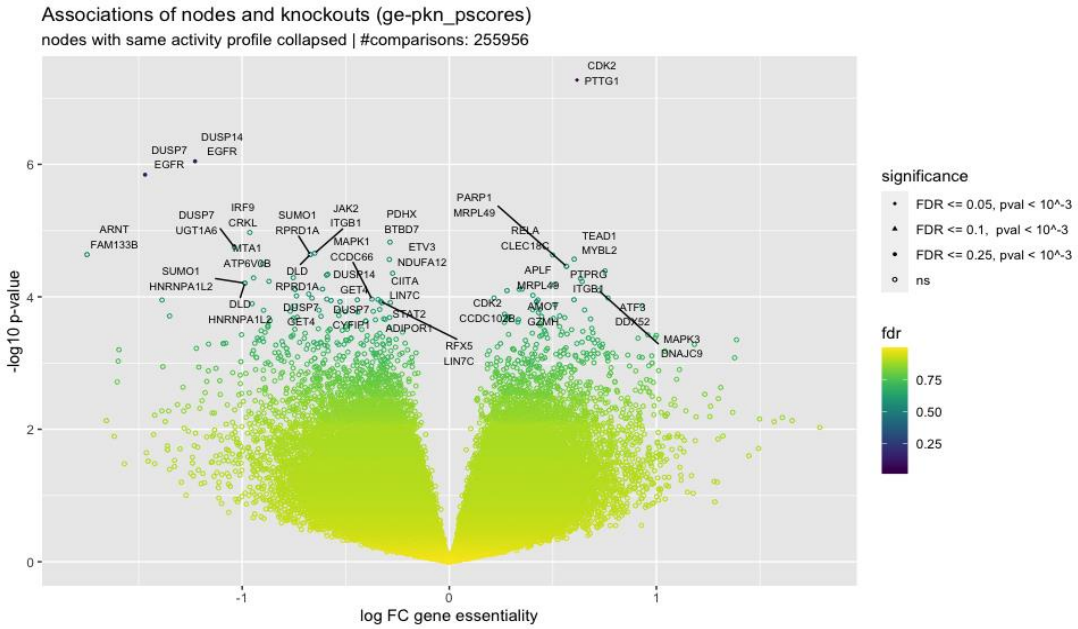

Figure 5: Associations of node activity and knockouts

Although the highest ranking significant associations have similar P-values as the node-drug associations, the much larger number of comparisons results in stricter multiple-testing correction and less significant associations between nodes activities and gene essentiality scores. In total, 15 associations are reported as significant in at least one configuration, seven of which are consistent across the specific PKN-based configurations.

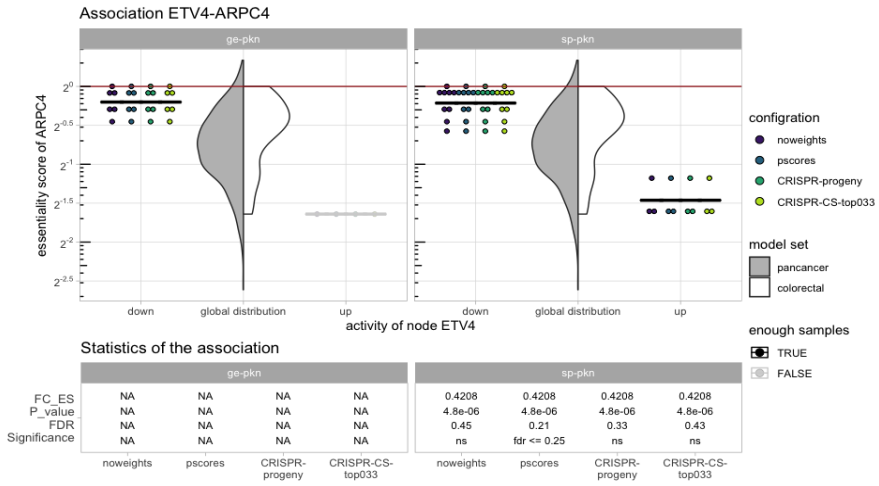

Figure 6: Association of ETV4 node activity and ARPC4 essentiality

Essentiality of ARPC4 is positively associated with upregulation of ETV4 (Figure 6). The former protein is a member of the Arp2/3 actin polymerization complex that is important for cellular motility (Welch et al. 1997). In bladder cancer, ARPC4 expression was reported to promote tumour invasion and migration (Xu et al. 2019). The transcription factor ETV4 promotes synthesis of matrix metalloproteases, secreted enzymes that degrade extracellular matrix (ECM) proteins. Additionally, there are reports that ETV4 expression promotes tumour cell invasion and metastasis, but the exact mechanisms are still unclear Cai et al. (2020). Both genes are tightly involved in cellular motility, and act synergistically in the invasive tumour stage. Our results indicate that the upregulation of ETV4 corresponds to increased essentiality of ARPC4. Assuming that the activity of ETV4 is part of an invasive transcriptional program, logically, the cells need ARPC4 and the actin polymerization complex to drive motility and invasion in this stage.

In total, four of the seven valid gene-knockout-associations can be confirmed with existing literature on different levels. For the remaining associations, a direct interaction between node activity and essentiality of another gene are less obvious or cannot be shown with the current state of research.

2.3 Discussion

It was demonstrated before that the genetic configuration of a tumour determines its sensitivity to an anticancer treatment (Yang et al. 2012; Iorio et al. 2016; Tang and Gottlieb 2021). The established methods, including the GDSC, statistically associate drug sensitivity screening data with transcriptional and mutational datasets to find significantly diverging reactions to a drug between groups that differ in a specific molecular feature. Many statistically relevant associations have been reported to date and are followed up on with experimental validation efforts. However, the behaviour of a cell is not only determined by its genetic profile, but also by other factors like its current environment, cellular and molecular interactions. As such factors are hard to measure, we infer the signalling behaviour of the investigated cells by applying the logic modeling tool CARNIVAL. That way, we elucidate some of the associations between the state of cellular signalling and drug sensitivity or gene essentiality in different colorectal cancer cell lines.

The study is centered around CARNIVAL, a computational tool that uses causal reasoning to predict signalling networks which explain signal transduction between specific stimuli and the gene expression profile of a biological system under the given stimulus. The generated signalling networks depict signalling paths and nodes that differ from canonical wiring in the form of up- and downregulation, thus characterising the behaviour of the specific system.

In order to gain insights from the generated CARNIVAL networks, their node activities are statistically analysed and associated with drug sensitivity and gene essentiality data. Importantly, we limit our analysis to the average activities of nodes and disregard their locations and wiring. For the statistical

associations, different approaches were compared and ANOVA was selected, similarly to the GDSC database which associates cell line drug sensitivities with genetic variants like coding mutations and CNAs.

Based on the aforementioned selection criteria, we report two sets of highly significant drug-node and gene-node associations. Although over 40 times more combinations of nodes and genes were analysed than node-drug pairs, the former yielded only 7 valid associations while the latter produced 32. This coincides with the respective p-value distributions, which had much more prominent peaks for the node-drug pairs and is likely rooted in the stricter multiple-testing correction for knockout-related associations. Also, the drug screenings were performed individually for each compound while knockout screenings covered all coding proteins in one experiment. Together with the larger range of values for drug sensitivity, one could expect clearer and more concise results for the first approach.

Almost all valid associations can be mapped to important cancerous processes. Drug-based associations prominently affect the cell cycle and DNA integrity while few targeted drugs are directed against specific signalling pathways that are related to proliferation and survival. The reported and confirmed associations represent a useful starting point for further experimental research on the involved nodes. The strict selection criteria with respect to statistical significance and consistency enable a clear prioritization of a defined set of interactions.

Chapter 3 Association of contextualised networks with drug sensitivity data in patients' samples

3.1 Introduction

Within the INFORM (INdividualized Therapy FOr Relapsed Malignancies in Childhood) registry study over 1200 cases were molecularly profiled using next generation sequencing (WGS, WES, RNAseq) and microarray-based technologies (methylome, transcriptome) (Worst et al 2016). The molecular diagnostics platform aimed to identify therapeutic targets. Nevertheless, high evidence targets (such as BRAFV600E mutations or NTRK fusions) are only detected in 5% of cases and further only 50% of the patients were identified with druggable pathways. The remaining cases lacked druggable alterations which mainly included brain tumors, sarcomas and neuroblastomas (van Tilburg et al 2021). Thus, an *ex-vivo* functional drug response profiling (DSP) platform for paediatric solid tumours has been established by the Translational Drug Screening Unit, KiTZ, DKFZ to complement the molecular profiles of the INFORM study with functional drug sensitivity data for each patient. The INFORM drug profiling platform aims to identify biomarkers and molecular mechanisms associated with drug response profiles.

3.2 Results

We have curated 74 samples from the multi-omics and drug response profiling data from the INFORM DSP dataset (ElHarouni et al 2022). The processed features within the omics profiling include: Functional somatic Single Nucleotide Variants (SNV)/ Insertion and Deletions (INDEL) calling, gene fusions, and quantified gene expression.

We have used CARNIVAL (CAusal Reasoning for Network identification using Integer VALue programming) (Liu et al 2019) to identify associations between functional drug response profiles and causal pathways from gene fusion and expression footprints. PROGENy (Schubert et al 2018). response signatures for 14 pathways were calculated from gene expression (TPM) values. Known signaling pathway correlation could be identified (eg: MEK inhibitors with MAPK signaling pathway), however the correlation coefficient for all drug-pathway analysis was not of high strength, where the maximum Pearson correlation coefficient found was 0.66 (Figure 7A). Transcription factors activity scores derived from gene expression data was estimated using DoRothEA (Garcia-Alonso et al 2019). The activity scores were correlated to the $sDSS_{asym}$ quantitative measurement. Different clusters of various drug mechanistic actions showed a strong correlation to the functional drug response profiles of the samples (Figure 7B).

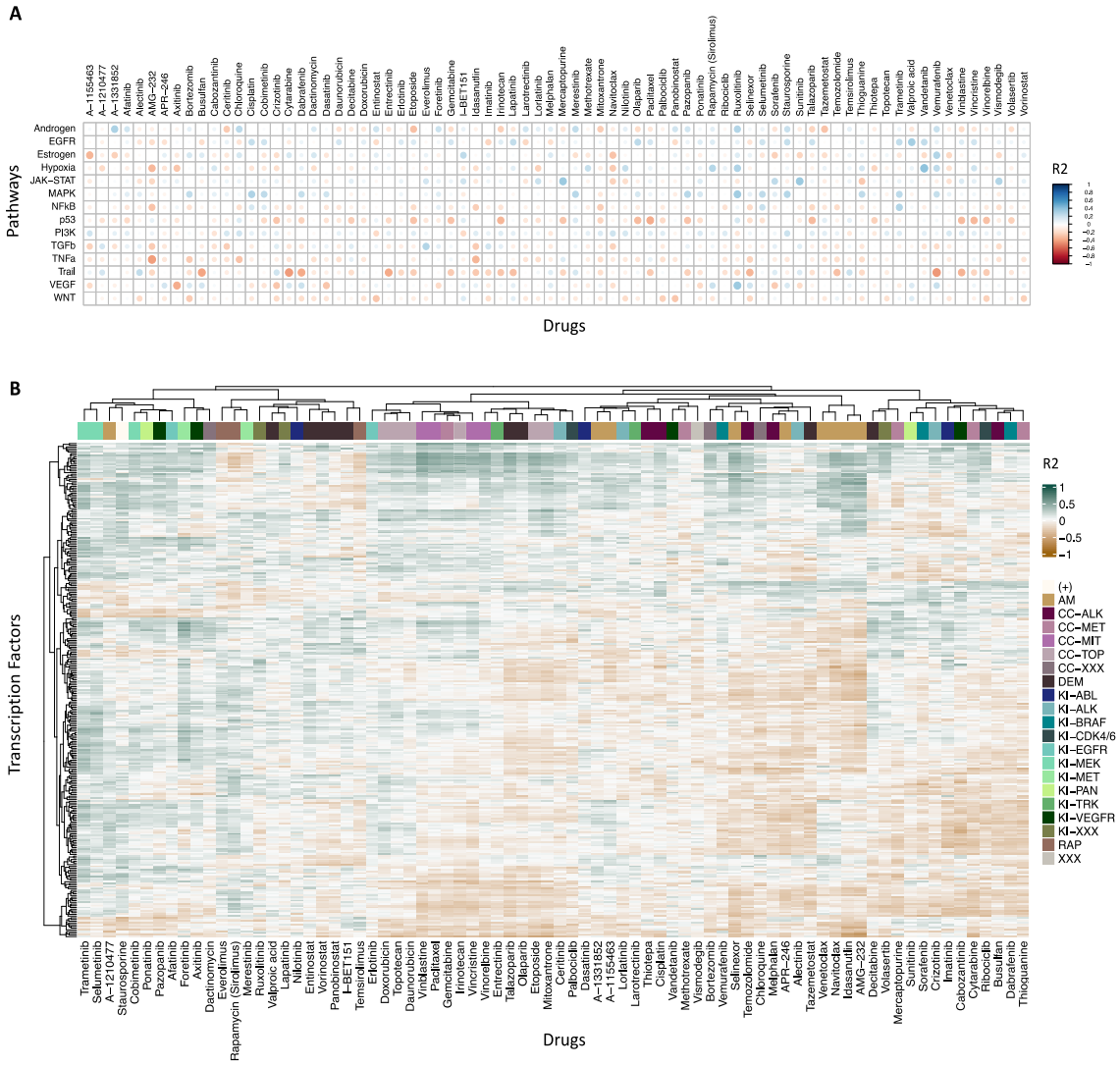

Figure 7: A) Progeny pathway signature – drug response correlation. Pearson correlation coefficient between the drug score and the pathway activity. B) Correlation matrix between transcription factors and sDSSasym score

Both pathway and transcription factor signatures can be used to identify a subset of interactions from a prior knowledge network that represent potential regulated pathways linking known or potential targets of perturbations towards active transcription factors derived from gene expression data using the CARNIVAL pipeline. Hence, a sample specific prior knowledge network was constructed using a constraint setting based on the patient's genomic profile (functional SNV and INDEL mutations). Since paediatric cancers have a low mutational burden, recurrent fusion genes and their regulatory signalling pathway can be explored as actionable biomarkers. Therefore, we have integrated fusion genes to the prior knowledge network to be able to explore the associations between active nodes of the derived network and drug response data. CARNIVAL inputs for the mentioned settings were prepared and used to identify the sub-network topology. CARNIVAL network interactions were

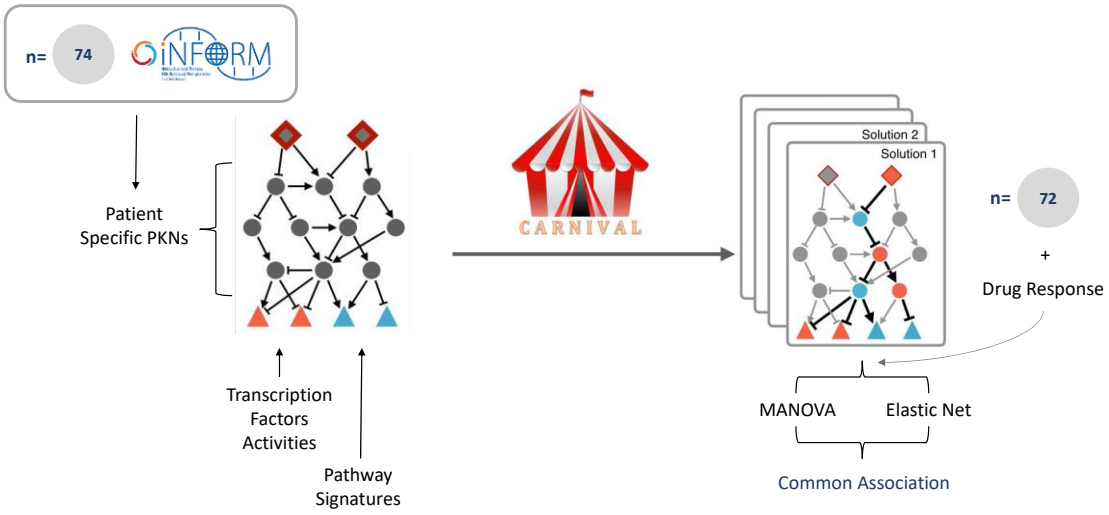

Figure 8: Overview of the patient specific CARNIVAL analysis workflow.

The CARNIVAl pipeline requires an input of a prior knowledge network, transcription factor activity scores, and progeny pathway signature scores. Inputs of 74 INFORM samples are used, analyzed and optimized using integer linear programing with the cplex solver. 72 samples had predicted signaling network topology solution with node activities. The Interactions are furtherly statistically tested against drug response data using MANOVA and ElasticNet

statistically tested for association with drug response outputs using both the multivariate analysis of variance (MANOVA) and ElasticNet. A summary of the analysis workflow using the CARNIVAL pipeline is shown in Figure 8**Fehler! Verweisquelle konnte nicht gefunden werden.**.

Figure 9 reveals the common associations detected by both statistical methods (MANOVA and ElasticNet), where the top weighted association captured is between the BCL2 inhibitor (Venetoclax) and the interaction between CTNNB1 and MMP9. The interaction and downstream TF signaling is shown in Figure 9B, where further mapping with OmniPath interactions showed a direct link between BCL2 and CTNNB1 confirming the finding.

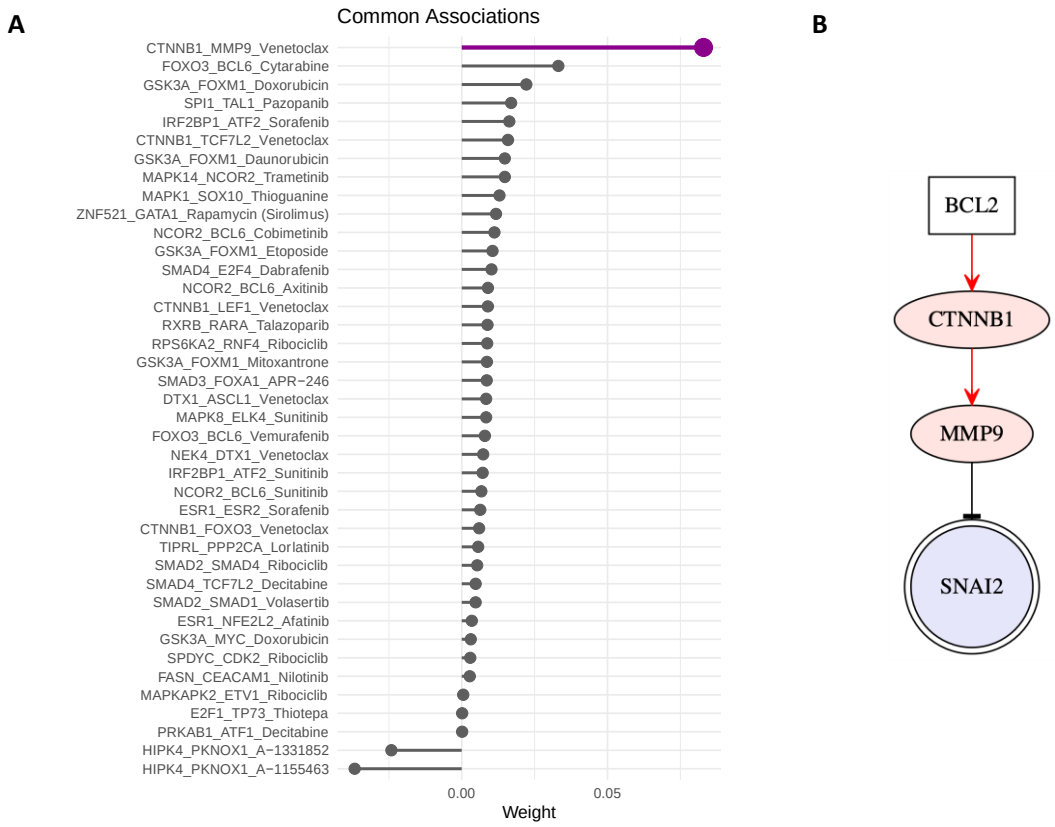

Figure 9: CARNIVAL topology network top associations with drug response data

A) Common association results between MANOVA and ElsaticNet ranked by the weight of the association. The top association reported is the CTNNB1_MMP9 interaction with the selective DSS (sDSSasym) of Venetoclax. B) The topology network of the top associated interaction affecting the SANI2 transcription factor. A direct link between the drug target associated (BCL2) and the interaction of CTNNB1_MMP is captured by the OmniPath interaction mapping.

Chapter 4 FUNKI: Interactive functional footprint-based analysis of omics data

4.1 Introduction

Multiple methods are conceived to infer the activities of specific processes or molecules using the abundance of known targets from omic data. We call them footprint-based methods (Dugourd and Saez-Rodriguez, 2019), and we have developed such tools for transcription factor (TF) from transcripts of target genes (Garcia-Alonso et al., 2019), kinases from phosphorylated sites (Wirbel et al., 2018) and pathways from transcripts of downstream responsive genes (Schubert et al., 2018). These activities can then be used to contextualize large signalling networks by identifying paths that can explain the changes in activities via reverse causal reasoning (Dugourd et al., 2021; Liu et al., 2019), and be further linked to changes observed at the level of metabolite abundances (Dugourd et al., 2021; Liu et al., 2019).

FUNKI (FUNctional analysis toolKIt) is an user-friendly interface developed in R language, and designed using Shiny (Chang et al., 2020), to analyze omics data using footprint methods. This application provides an interface for the R implementations (Bioconductor packages) for the aforementioned tools. All methods run on bulk data, and we have shown that they can also be applied to single-cell transcriptomics (Holland et al., 2020a), not only for humans but also for mouse samples (Holland et al., 2020b).

4.2 Features

The footprint methods implemented in FUNKI allow users to recover functional insight from several omics data without notions of programming. This application also enhances the analysis with an extended graphic visualization of the results. Thus, the typical FUNKI pipeline comprises three steps: (i) import the user's omic data, (ii) select the analysis tool according to the data and question and (iii) visualize the results in tables and graphical representations (Figure 10). Currently, the following tools are implemented:

DoRothEA (Discriminant Regulon Expression Analysis) is a resource that links TFs with their downstream targets (Garcia-Alonso et al., 2019). TF activities are computed from gene expression where the regulons (the collection of transcriptional targets for each TF) are the underlying gene sets.

PROGENy (Pathway RespOnsive GENes) is a footprint method developed to infer pathway activity from gene expression data (Schubert *et al.*, 2018). The scores are calculated using a linear model with weights based on consensus gene signatures obtained from publicly available perturbation experiments.

KinAct estimates kinase activities based on abundance changes measures in target phosphorylation sites (Wirbel *et al.*, 2018) using the same algorithm as DoRothEA. Instead of TF-target interactions, KinAct uses collections of kinase–substrate interactions via OmniPath (Türei *et al.*, 2016) and phosphoproteomic data instead of transcriptomic data.

CARNIVAL (CAusal Reasoning for Network identification using Integer VALue programming) reconstructs signalling networks from downstream TF activities by finding the upstream regulators (Dugourd *et al.*, 2021; Liu *et al.*, 2019). **COSMOS** is an extension of CARNIVAL that provides a multiomic network to connect different types of omic data together, including transcriptomics, metabolomics and phosphoproteomics. Both methods identify coherent mechanistic hypotheses (subnetworks) that explain how the measured deregulation may be reached.

FUNKI is a Shiny application developed using R programming language. It is directly accessible in the cloud through https://saezlab.shinyapps.io/funki/. The source code is freely available at https://github.com/saezlab/ShinyFUNKI, and it can be run locally in any platform (Windows, macOS and Linux) either downloading the repository or running it directly from GitHub (see https://saezlab.github.io/ShinyFUNKI/ for details).

4.3 Conclusion

FUNKI provides an intuitive user-friendly interface to run footprint methods from different omics. Together with the analysis implementation, FUNKI also incorporates several graphical representations to explore the results from different perspectives. Users with programming skills can take advantage of an extended script-based version of FUNKI for transcriptomic data (https://github.com/saezlab/transcriptutorial).

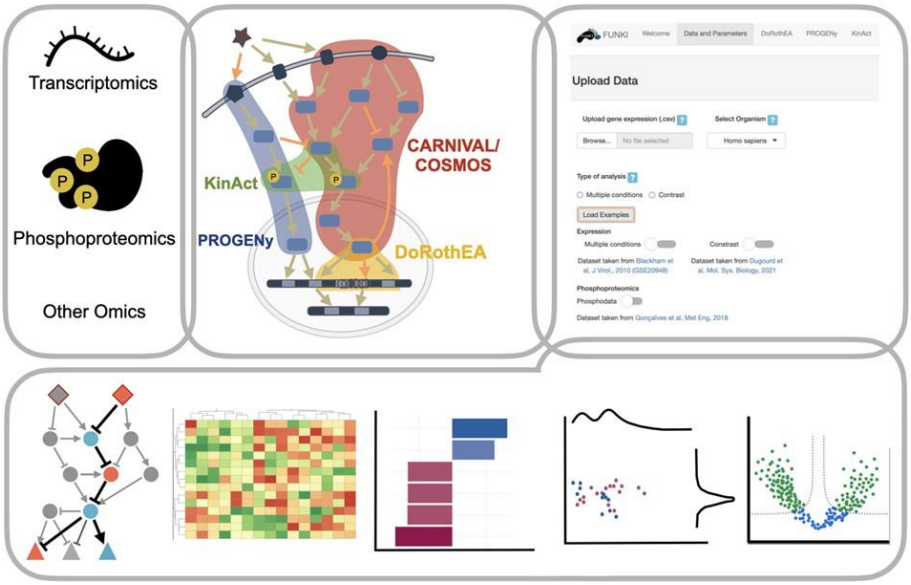

Figure 10: Graphical overview of analysis and visualization features provided by FUNKI. FUNKI provides a user interface to upload omics data, and then run DoRothEA, PROGENy, KinAct, CARNIVAL and COSMOS to estimate the activity of pathways, TFs and kinases. This figure was published in Hernansaiz-Ballesteros et al. 2022

Chapter 5 Summary and Conclusion

We fully accomplished the objective of D8.2 to develop computational pipelines that predicts potential drug targets for cell lines and paediatric cancer patients. First (see Chapter 2), we obtained a general prior knowledge network (PKN) of protein-protein interactions from Omnipath. Then, we used genomics data (copy number alteration, deletion and insertion) to tailor the general PKN to individual cell lines (see Chapter 2) and cancer patient samples (see Chapter 3). We utilized CARNIVAL, a logic based causal inference tool, to infer the activity state of the signaling network of the cell line and tumour samples using the prior knowledge network and gene expression data. Finally, we associated the estimated activity of the signaling nodes with CRISPR-based gene essentiality scores and drug sensitivity data.

The pipeline resulted in potential molecular targets of therapies. In the cell line-based study (Chapter 2) we found 7 significant gene-essentiality - signaling node activity and 32 drug – signaling node activity associations out of which 4 and 7 were found in the literature, respectively. Further, we also found significant associations between the drug response of patients' data (Chapter 3) and their signaling nodes' activity.

To make the above computational pipeline broadly available, we also developed an interactive tool which facilitates footprint-based omics analysis (Chapter 4). It can be used for the estimation of transcription factor and pathway activity from expression data via DoRothEA and PROGENy, estimation of kinase activity from phosphoproteomics data using KinAct and for the inference of signaling and multi-omics network using CARNIVAL and COSMOS. The tool can be used from a web-browser and requires little to no computational skills from the users

Chapter 6 Bibliography

Behan, Fiona M., Francesco Iorio, Gabriele Picco, Emanuel Gonçalves, Charlotte M. Beaver, Giorgia Migliardi, Rita Santos, et al. 2019. “prioritisation of Cancer Therapeutic Targets Using CRISPRCas9 Screens.” *Nature* 568 (7753): 511–16. https://doi.org/10.1038/s41586-019-1103-9.

Boyle, E. I., S. Weng, J. Gollub, H. Jin, D. Botstein, J. M. Cherry, and G. Sherlock. 2004. “GO::TermFinder–Open Source Software for Accessing Gene Ontology Information and Finding Significantly Enriched Gene Ontology Terms Associated with a List of Genes.” *Bioinformatics* 20 (18): 3710–15. https://doi.org/10.1093/bioinformatics/bth456.

Cai, L.-S., Q.-X. Chen, S.-Y. Fang, M.-Q. Lian, M.-J. Lian, and M.-Z. Cai. 2020. “ETV4 promotes the progression of gastric cancer through regulating KDM5D.” *European Review for Medical and Pharmacological Sciences* 24 (5): 2442–51. https://doi.org/10.26355/eurrev_202003_20511.

Dugourd, Aurelien, and Julio Saez-Rodriguez. 2019. “Footprint-Based Functional Analysis of Multiomic Data.” *Current Opinion in Systems Biology* 15 (June): 82–90. https://doi.org/10.1016/j.coisb.2019.04.002.

Dugourd, Aurelien et al. 2021. “Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses”. *Mol. Syst. Biol.*, 17, e9730. https://dx.doi.org/10.15252/msb.20209730.

ElHarouni, D. et al. 2022, “iTReX: Interactive exploration of mono- and combination therapy dose response profiling data,” *Pharmacological Research*, vol. 175, p. 105996, https://doi.org/10.1016/j.phrs.2021.105996.

Garcia-Alonso, L, Holland, C. H., Ibrahim, M. M., Turei, D. and Saez-Rodriguez, J. 2019, “Benchmark and integration of resources for the estimation of human transcription factor activities,” *Genome Research*, vol. 29, no. 8, pp. 1363–1375, https://doi.org/10.1101/gr.240663.118.

Gkotinakou, Ioanna-Maria, Christina Befani, George Simos, and Panagiotis Liakos. 2019. “Erk1/2 Phosphorylates HIF-2alpha and Regulates Its Activity by Controlling Its Crm1-Dependent Nuclear Shuttling.” *Journal of Cell Science* 132 (7). https://doi.org/10.1242/jcs.225698.

Holland, C. H., Tanevski, J., Perales-Patón, J., Gleixner, J., Kumar, M. P., Mereu, E., ... & Saez-Rodriguez, J. 2020. “Robustness and applicability of transcription factor and pathway analysis tools on single-cell RNA-seq data”. *Genome biology*, 21(1), 1-19. http://dx.doi.org/10.1186/s13059-020-1949-z.

Holland, Christian H., Bence Szalai, and Julio Saez-Rodriguez. 2020. "Transfer of regulatory knowledge from human to mouse for functional genomics analysis." *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* 1863, no. 6: 194431. http://dx.doi.org/10.1016/j.bbagrm.2019.194431

Iorio, Francesco, Theo A. Knijnenburg, Daniel J. Vis, Graham R. Bignell, Michael P. Menden, Michael Schubert, Nanne Aben, et al. 2016. “A Landscape of Pharmacogenomic Interactions in Cancer.” *Cell* 166 (3): 740–54. https://doi.org/10.1016/j.cell.2016.06.017.

“Lifetime Risk of Developing or Dying from Cancer.” 2020. *American Cancer Society*. https://www.cancer.org/cancer/cancer-basics/lifetime-probability-of-developing-or-dying-from-cancer.html.

Liu, Anika, Panuwat Trairatphisan, Enio Gjerga, Athanasios Didangelos, Jonathan Barratt, and Julio Saez-Rodriguez. 2019. “From Expression Footprints to Causal Pathways: Contextualizing Large signalling Networks with CARNIVAL.” *Npj Systems Biology and Applications* 5 (1). https://doi.org/10.1038/s41540-019-0118-z.

Morris, Erick J., Sharda Jha, Clifford R. Restaino, Priya Dayananth, Hugh Zhu, Alan Cooper, Donna Carr, et al. 2013. “Discovery of a Novel ERK Inhibitor with Activity in Models of Acquired Resistance

to BRAF and MEK Inhibitors." *Cancer Discovery* 3 (7): 742–50. https://doi.org/10.1158/2159-8290.cd-13-0070.

Pacini, Clare. 2021. "Integrated CRISPR." https://doi.org/10.6084/M9.FIGSHARE.C.5289226.V1.

Pacini, Clare, Joshua M. Dempster, Isabella Boyle, Emanuel Gonçalves, Hanna Najgebauer, Emre Karakoc, Dieudonne van der Meer, et al. 2021. "Integrated Cross-Study Datasets of Genetic Dependencies in Cancer." *Nature Communications* 12 (1). https://doi.org/10.1038/s41467-021-21898-7.

Patel, S A, and M C Simon. 2008. "Biology of Hypoxia-Inducible Factor-2alpha in Development and Disease." *Cell Death & Differentiation* 15 (4): 628–34. https://doi.org/10.1038/cdd.2008.17.

Schubert, Michael, Bertram Klinger, Martina Klünemann, Anja Sieber, Florian Uhlitz, Sascha Sauer, Mathew J. Garnett, Nils Blüthgen, and Julio Saez-Rodriguez. 2018. "Perturbation-Response Genes Reveal signalling Footprints in Cancer Gene Expression." *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-017-02391-6.

Tang, Yi-Ching, and Assaf Gottlieb. 2021. "Explainable Drug Sensitivity Prediction Through Cancer Pathway Enrichment." *Scientific Reports* 11 (1). https://doi.org/10.1038/s41598-021-82612-7.

Türei, Dénes, Tamás Korcsmáros, and Julio Saez-Rodriguez. 2016. "OmniPath: Guidelines and Gateway for Literature-Curated signalling Pathway Resources." *Nature Methods* 13 (12): 966–67. https://doi.org/10.1038/nmeth.4077.

van der Meer, Dieudonne, Syd Barthorpe, Wanjuan Yang, Howard Lightfoot, Caitlin Hall, James Gilbert, Hayley E Francies, and Mathew J Garnett. 2018. "Cell Model Passportsa Hub for Clinical, Genetic and Functional Datasets of Preclinical Cancer Models." *Nucleic Acids Research* 47 (D1): D923–29. https://doi.org/10.1093/nar/gky872.

van Tilburg, C. M. et al., 2021. "The Pediatric Precision Oncology INFORM Registry: Clinical Outcome and Benefit for Patients with Very High-Evidence Targets," *Cancer Discovery*, vol. 11, no. 11, pp. 2764–2779. https://doi.org/10.1158/2159-8290.CD-21-0094.

Welch, Matthew D., Angela H. DePace, Suzie Verma, Akihiro Iwamatsu, and Timothy J. Mitchison. 1997. "The Human Arp2/3 Complex Is Composed of Evolutionarily Conserved Subunits and Is Localized to Cellular Regions of Dynamic Actin Filament Assembly." *Journal of Cell Biology* 138 (2): 375–84. https://doi.org/10.1083/jcb.138.2.375.

Wirbel, Jakob, Pedro Cutillas, and Julio Saez-Rodriguez. 2018. "Phosphoproteomics-based profiling of kinase activities in cancer cells." *Cancer Systems Biology*, pp. 103-132. Humana Press, New York, NY, 2018. http://dx.doi.org/10.1007/978-1-4939-7493-1

Worst, B. C. et al. 2016, "Next-generation personalised medicine for high-risk paediatric cancer patients – The INFORM pilot study," *European Journal of Cancer*, 65, pp. 91–101, https://doi.org/10.1016/j.ejca.2016.06.009.

Xu, Ning, Gen-Yi Qu, Yu-Peng Wu, Yun-Zhi Lin, Dong-Ning Chen, Xiao-Dong Li, Shao-Hao Chen, et al. 2019. "Arpc4 Promotes Bladder Cancer Cell Invasion and Is Associated with Lymph Node Metastasis." *Journal of Cellular Biochemistry* 121 (1): 231–43. https://doi.org/10.1002/jcb.29136.

Yang, Wanjuan, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, et al. 2012. "Genomics of Drug Sensitivity in Cancer (GDSC): A Resource for Therapeutic Biomarker Discovery in Cancer Cells." *Nucleic Acids Research* 41 (D1): D955–61. https://doi.org/10.1093/nar/gks1111.

Yau, Tung On. 2019. "Precision Treatment in Colorectal Cancer: Now and the Future." *JGH Open* 3 (5): 361–69. https://doi.org/10.1002/jgh3.12153.