NNNPrECISE

D5.4

Lists of possible drug targets (intervention points), and of individual and combination of drugs

Project number:	668858
Project acronym:	PrECISE
Project title:	PrECISE: Personalized Engine for Cancer Integrative Study and Evaluation
Start date of the project:	1st January, 2016
Duration:	36 months
Programme:	H2020-PHC-02-2015

Deliverable type:	Report
Deliverable reference number:	PHC-668858 / D5.4/ V1.0
Work package contributing to the deliverable:	WP 5
Due date:	December 2018– M36
Actual submission date:	21st December 2018

Responsible organisation:	UKAACHEN
Editor:	Luis Tobalina
Dissemination level:	PU
Revision:	V1.0

Abstract:	This deliverable provides a list of potential therapeutic targets based on the analysis of personalized logical models and a description of how these were obtained.
Keywords:	Boolean models, target prediction, cell-line specific models, patient specific models

Editor

Luis Tobalina (UKAACHEN)

Contributors (ordered according to beneficiary numbers)

Montagud, Arnau (CI)

Szalai, Benze (UKAACHEN)

Čuklina, Jelena (ETH)

Disclaimer

The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The users thereof use the information at their sole risk and liability.

Executive Summary

One of the objectives of WP5 is to provide a list of candidate drugs and combinations of drugs for cell lines and for specific patients. To get to this point, we leverage the work described in deliverables D5.1, D5.2 and D5.3 and run a series of computational analysis that aim to provide a list of targets of interest in prostate cancer. This deliverable describes how the prostate cancer Boolean model is mixed with prostate cancer related experimental data to derive cell line and patient specific target candidates. As an alternative to Boolean modeling and for comparison purposes, a data based approach to finding cell line specific targets is also used.

The final list of proposed targets constitutes a starting point for the design of validation experiments in D6.5.

Contents

List of Figures

List of Tables

Chapter 1   Introduction

The current array of options for prostate cancer treatment is limited. When careful observation, surgical intervention and chemical castration fail, there are only a handful of treatment options left. Currently approved drugs by the FDA for prostate cancer can be organized in four types of inhibitors: androgen receptor (AR), testosterone production, tubulin and gonadotrophin-releasing hormone (GnRH) inhibitors (https://www.cancer.gov/about-cancer/treatment/drugs/prostate). One of these approved drugs is Enzalutamide, which is given to those patients that have become castration resistant (i.e. they do not respond to treatments designed to lower testosterone levels anymore). However, the drugs used at later stages usually end up failing when the cancer develops a resistance mechanism against them. For this reason, finding new targets and drugs to act upon them constitutes an urgent need.

Given the heterogeneity of the disease observed in each patient, it is likely that personalized treatments will be needed to effectively treat patients. However, as trying different treatments on a single patient is not a feasible nor an ethical option, an in-silico approach needs to be devised to meet this goal. To evaluate the efficacy of different drugs, we need to start by working with cell lines instead of patients. Using cell lines that, for example, represent an Enzalutamide sensitive and a resistant stage of prostate cancer, we can learn important things about its biology and obtain clues to better treatments. The use of cell lines allows the design of experiments in which time-course data can be obtained under perturbations that may include combination of treatments. This type of setting provides input data for dynamic models and the exploration of steady state responses.

This deliverable is focused on the use of Boolean model techniques mixed with prostate cancer related experimental data to get closer to meeting these goals. Using these models, we can propose cell line and patient specific target candidates, which can later enter a stage of experimental validation provided they meet additional criteria established during drug development processes. In this regard, the list of proposed targets in this deliverable constitutes a starting point for the design of validation experiments in D6.5.

Chapter 2   Cell-line-specific drug targets

Cell lines are particularly important for the modelling of drug response in patients, as the data characterizing their molecular profiles in basal state or upon perturbation with drugs and their combination is easier to acquire than analogous data for patients. Also, multiple drugs in different doses can be tested on the cell lines, while such an option is not conceivable with patient research. *In silico* methods also allow to investigate the mechanisms of drug sensitivity/resistance, improving the quality of the predicted drug.

Here, the goal was to identify drug targets specific for each cell-line of interest enabled by the developments attained in tailoring our model to a given set of data. Thus, the present deliverable builds on top of D5.3 to be able to provide candidate drugs and combinations of drugs for cell lines (Task 5.4).

Each cell line is conceptually identical to a single patient. Thus, predicting cell line response from molecular data, mimics personalization of patient drug treatment, informed by molecular profiling of patient tissues.

There are multiple approaches to predict drugs that could benefit prostate cancer patients. One is to identify the key protein that can be targeted, and these can be identified with logical models; another is to build the regression model, predicting the sensitivity to the drug from gene expression data. In the following section, we will describe these approaches in detail.

2.1   Use of cell-line-specific Boolean models

2.1.1   *Generic logical model of prostate cancer*

The generic logical model of prostate cancer delivered in D5.1 was integrated with data to have tailored models, delivered in D5.3. The results from this set of models capture patient- and cell-line-specific behaviours.

2.1.2   *Public transcriptomic cell line data used to tailor the model*

We used data from Genomics of Drug Sensitivity in Cancer (GDSC) database (Yang et al. 2013) that is the largest public resource for information on drug sensitivity in cancer cells and molecular markers of drug response. GDSC currently contains drug sensitivity data for almost 75 000 experiments, describing response to 138 anticancer drugs across almost 700 cancer cell lines.

We obtained the data for all the 1001 GDSC cell lines and selected the 4 prostate-specific cell lines for which we had copy number alterations (CNA), mutations and RNA data: BPH-1, DU-145, LNCaP-Clone-FGC and PC-3. All of them are derived from carcinomas, except for BPH-1 that has its origin in a benign prostate hyperplasia.

2.1.3   *Obtaining cell-line-specific Boolean models*

Following the methodology presented in D5.3, the recipe (the methodology used to combine data in the model) selected for the prostate cell lines was using transcriptomic data as initial conditions and transition rates. Using mutation or CNA data was not deemed useful in this case as their recipes did not further discriminate among cell lines' behaviours.

The personalization of the prostate model into cell-line-specific models generates different phenotype probabilities among these and captures some of their differences. For instance,

PC-3 has a high glycolysis rate, BPH-1 has a very low Invasion and a high Proliferation and DU-145 has a higher Proliferation than the rest of carcinomas (Figure 1).
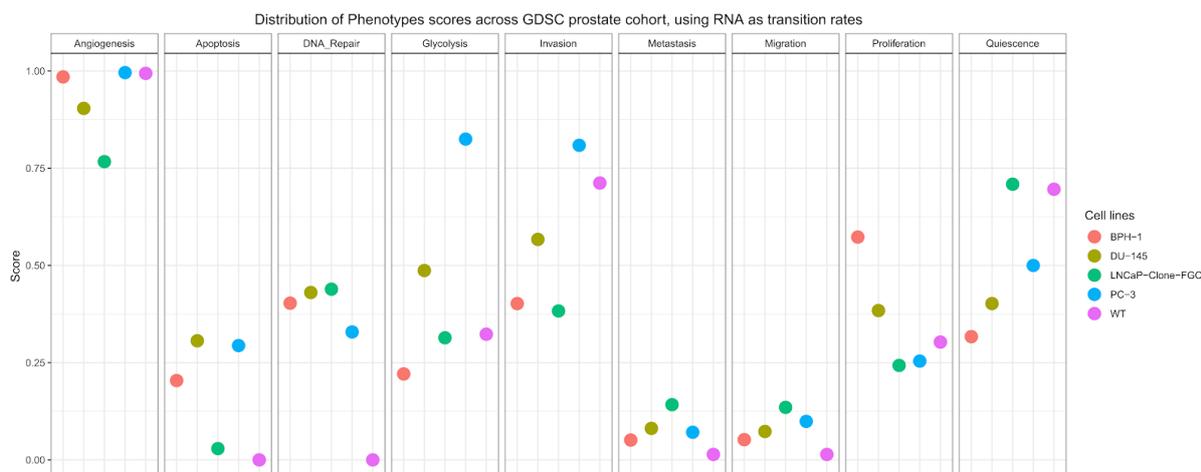

Figure 1: Phenotype simulation results across GDSC prostate-cell-line-specific Boolean models' simulation with random initial conditions. WT stands for wild type model, the original prostate model with no personalization.

As deliverable D5.2 provided additional data for the LNCaP-clone-FGC cell line, we chose this cell line model to study its genetic interactions and its uses for drug discovery. This LNCaP-specific model has 24 stable states that can be grouped in 6 main phenotypes: *Quiescence, Proliferation, Apoptosis, Invasion, Glycolysis* and *Hypermethylation*. Thanks to MaBoSS software (Stoll et al. 2017) we were able to assign probabilities to each one of these phenotypes.

2.2 Model-based identification of drug targets specific to cell-lines

Using MaBoSS, the LNCaP model, which corresponds to LNCaP-clone-FGC cell line data, was simulated with increasing node inhibition values to mimic the effect of drugs on the model genes. Six simulations were done per each inhibited node, with 100% of node activity (no inhibition), 80%, 60%, 40%, 20% and 0% (proper knock-out), under four different initial conditions, a nutrient-rich media with androgen, with EGF, with both and with none. For instance, *beta_catenin* node was sequentially inhibited under these four condition and their *Invasion* output scores were compared to the WT scores (Figure 2). Inhibition of *beta_catenin* node is dependent on *Androgen* presence in the growth media and the inhibition is much more efficient when *Androgen* is not present (*00* and *EGF* conditions, Figure 3).

Likewise, this analysis was extended to knock-out perturbations and to the 10 outputs of the model and this resulted in a list of 26 genes that hampered *Proliferation* and/or promoted *Apoptosis* under one or several of the aforementioned growth conditions:

ANAPC1, BAX, BRCA1, CASP9, **CDH1**, ETS1, FOXO1, **HSP90AA1**, MAPK1, MAPK3, **MAX**, NCOR1, NFKB1, NFKB2, **NOX1**, **NOX2**, NOX3, **NOX4**, REL, RELA, RELB, RUNX2, **SPOP**, **TERT**, TNFRSF1A, **TP53**

These 26 genes corresponded to 18 nodes in the model (*BAX, BRCA1, Caspase9, Cdh1, ERK, ETS1, E_cadherin, FOXO, HSPs, MAX, NCOR1, NF_kB, ROS, SPOP, TERT, TNFalpha, beta_catenin, p53*).

This list of targets was compared to analyses performed using other datasets from the present project, such as PPP1, PC39 and the panel of genes used for targeted sequencing in WP1. Depicted in **bold** are genes found in most of these studies. Most of the genes of this gene list were identified as being significantly characteristic of PPP1 patients (21 out of 26:

ANAPC1, BAX, BRCA1, CASP9, CDH1, ETS1, FOXO1, **HSP90AA1**, MAPK1, MAPK3, **MAX**, NFKB1, **NOX1, NOX3, NOX4**, RELA, RUNX2, **SPOP**, **TERT**, TNFRSF1A, **TP53**). Also, 2 genes of this gene list were identified as being significantly characteristic of PC39 patients (**CDH1**, **HSP90AA1**). Lastly, 3 genes of this gene list were also found in the panel of genes used for targeted sequencing in WP1 (**CDH1**, **SPOP**, **TP53**).

Additionally, these analyses can be combined to study the interaction of inhibiting two nodes simultaneously. For instance, *beta_catenin* and *AKT* nodes combined inhibition can be studied in Figure 4 and their *Invasion* score is mainly affected by *beta_catenin* inhibition under non-androgenic conditions (*00* or *EGF*) and responds to both *beta_catenin* and *AKT* nodes inhibition under androgenic conditions (*AR* or *AR_EGF*). Lastly, drug synergies can be studied using the single and combined simulations (Figure 5). In present work, Bliss Independence synergies were calculated for the combined inhibition of *beta_catenin* and *AKT* nodes (Figure 6). This score compares the combined effect of two drugs with the effect of each one of them, with a synergy when the value of this score is lower than 1. As it can be seen in Figure 6, there are synergies among these drugs in phenotypes such as *Invasion* (for *beta_catenin*) and *Glycolysis* and *Proliferation* for *AKT*. Again, these synergies depend on the presence of androgen in the growth media.
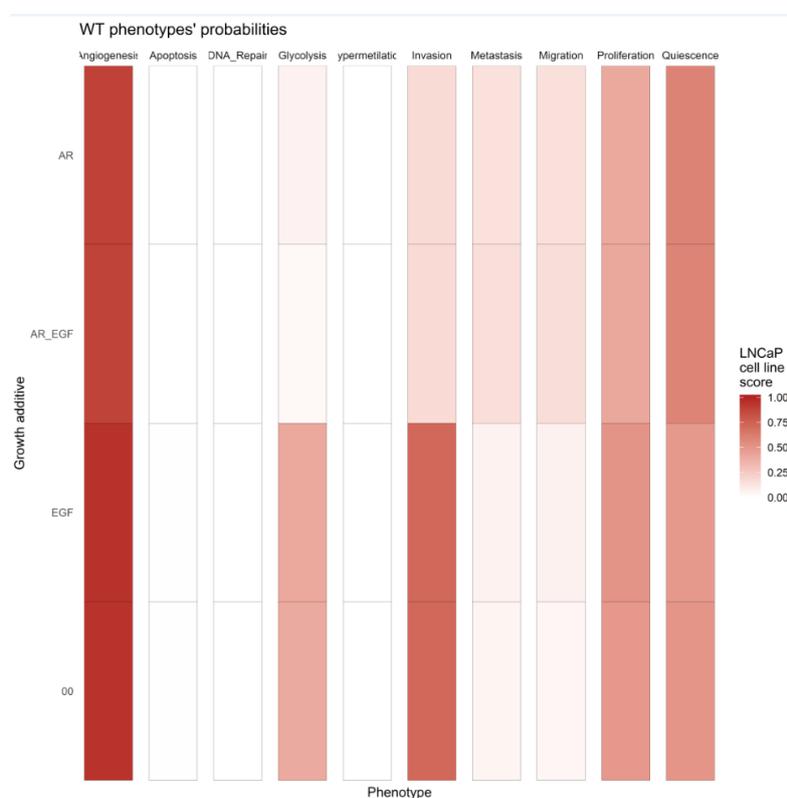
Figure 2: LNCaP-specific model phenotype probability variations under four different growth conditions. AR stands for androgen presence, EGF for EGF presence.
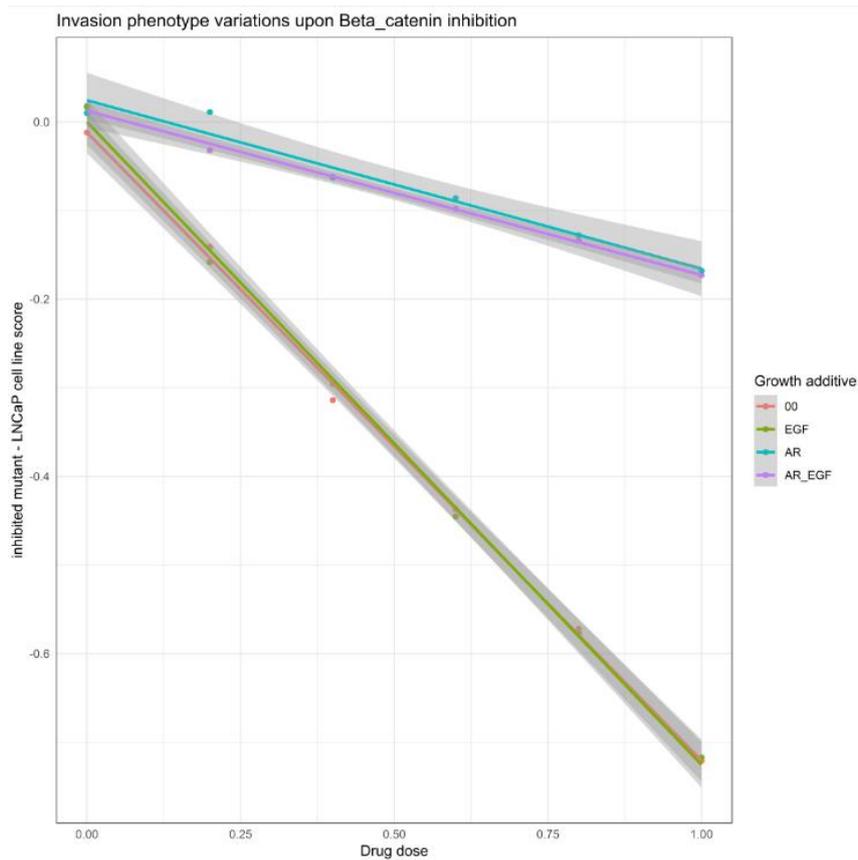
Figure 3: Invasion phenotype probability variations upon *beta_catenin* node inhibition under four different growth conditions.
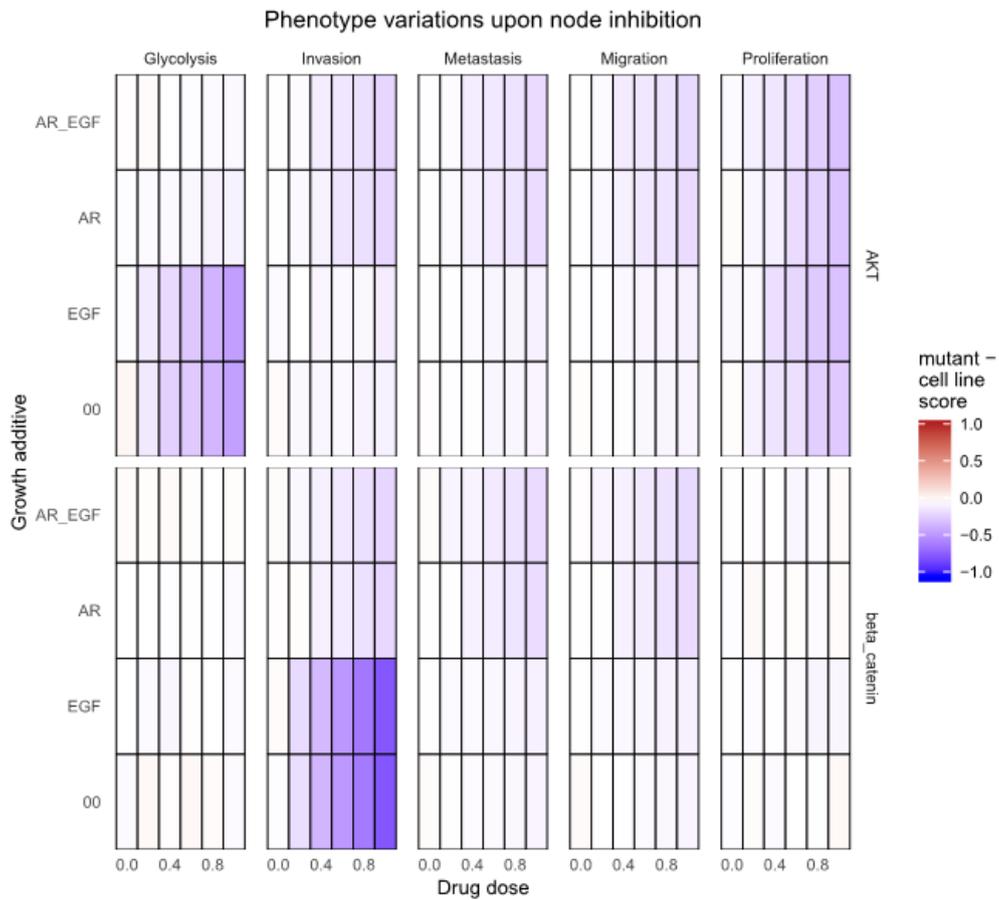
Figure 4: Phenotype score variations upon nodes inhibition under four different growth conditions for *beta_catenin* and *AKT* inhibitions.
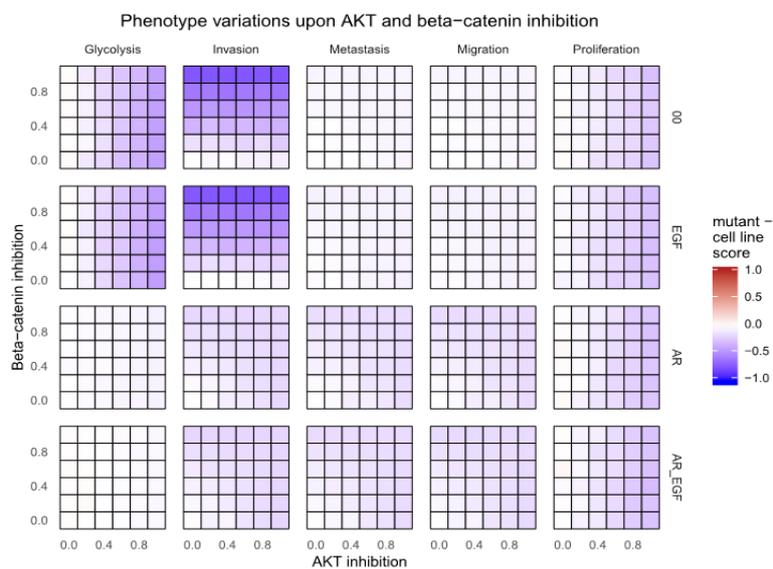
Figure 5: Phenotype score variations upon combined nodes inhibition under four different growth conditions.
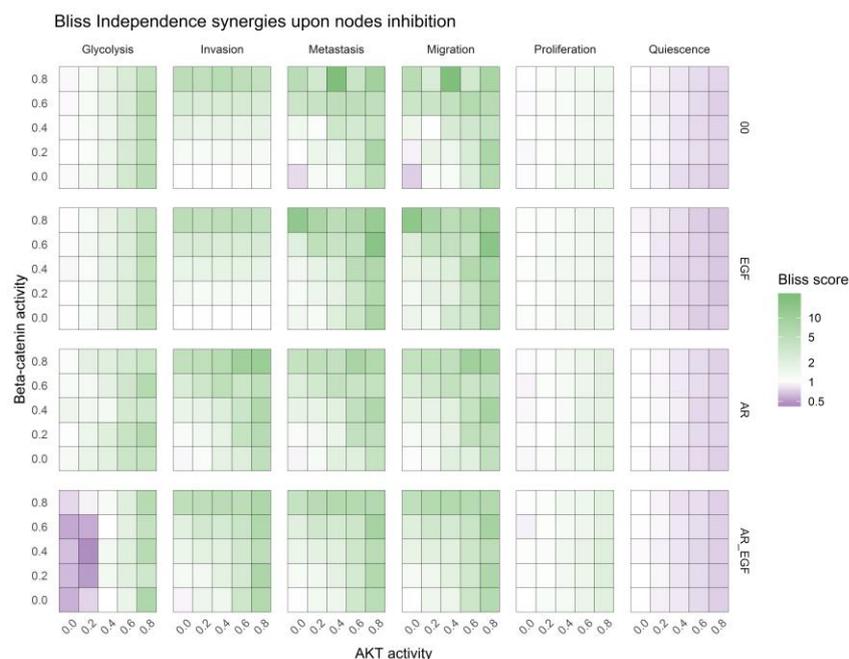
Figure 6: Bliss Independence synergies scores variations upon combined nodes inhibition under four different growth conditions. Combination Index < 1 means synergy.

2.2.1 List of targets proposed

As a result of these analyses, we selected the 5 mutated nodes that hampered *Proliferation* and/or promoted *Apoptosis* the most (*HSPs* as **HSP90AA1** gene, ***SPOP, MAX, TERT*** and *ROS* as **NOX1**, NOX2, **NOX3** and **NOX4** genes) and we completed it with several genes considered relevant in cancer progression, such as *AKT, AR, EGFR, PI3K* and *MEK*. Inhibition of different nodes affected differently the output probabilities and some of them were androgen dependent, such as *beta_catenin* (Figure 4, where values of scores are depicted with a colour gradient). Thus, the final gene list proposed for validation experiments given by this part of the analysis for the validation experiments of T6.3 and to constitute the starting point of D6.5 was:

HSP90AA1, SPOP, MAX, TERT, NOX1, NOX2, NOX3, NOX4, AKT, AR, EGFR, PI3K, MEK

2.3 Phosphoproteomic profile prostate cancer cell lines

With the aim of better characterizing the signalling response of prostate cancer cells, manifested mostly as the change of phosphorylation state of signalling proteins, changing their post-translational modification (PTM) status, in deliverable D5.2 we described phospho-proteomic SWATH-MS measurements on two prostate cancer cell lines, LNCaP and LNCaP-abl upon perturbation with different ligands and inhibitors (Figure 7). The cell lines chosen are closely related, but differ in a major component – enzalutamide sensitivity, major component of prostate cancer aggressiveness. This setup allows to address specifically the changes that are related to the enzalutamide resistant growth of tumors and is associated with the latest and most aggressive state of the disease.

After removing outliers, the PCA plot after batch correction (using a linear model that accounts for the effect of covariates) showed a clear difference between the two cell lines used, but the difference between perturbations was not immediately obvious (Figure 8).

The main goal of this data was to use it in combination with the prostate cancer Boolean model delivered in D5.1 and CellNOpt software (Terfve et al. 2012) in order to improve the Boolean model. We started by using CellNOpt and then moved on to using PHONEMeS (Terfve et al. 2015), which also uses Boolean modelling but it is better prepared to handle high coverage data such as the one provided by SWATH-MS technology.

PHONEMeS finds a network that connects the perturbations (inhibitors or ligands) to the phosphosites that the data shows as changing under the performed experiments. The link between them is explained by an underlying Boolean model. We used a background network of post-translational modifications (PTMs) retrieved from Omnipath (introduced in D3.1) (Türei et al. 2016) in combination with the experimental data (D5.2). The network obtained for the LNCaP cell line data (Figure 9) is smaller than the one recovered for the LNCaP-abl cell line data (Figure 10). It can be interpreted that the LNCaP-abl cell line shows a greater activity in response to the applied perturbations. Differences between both networks are highlighted in Figure 11.
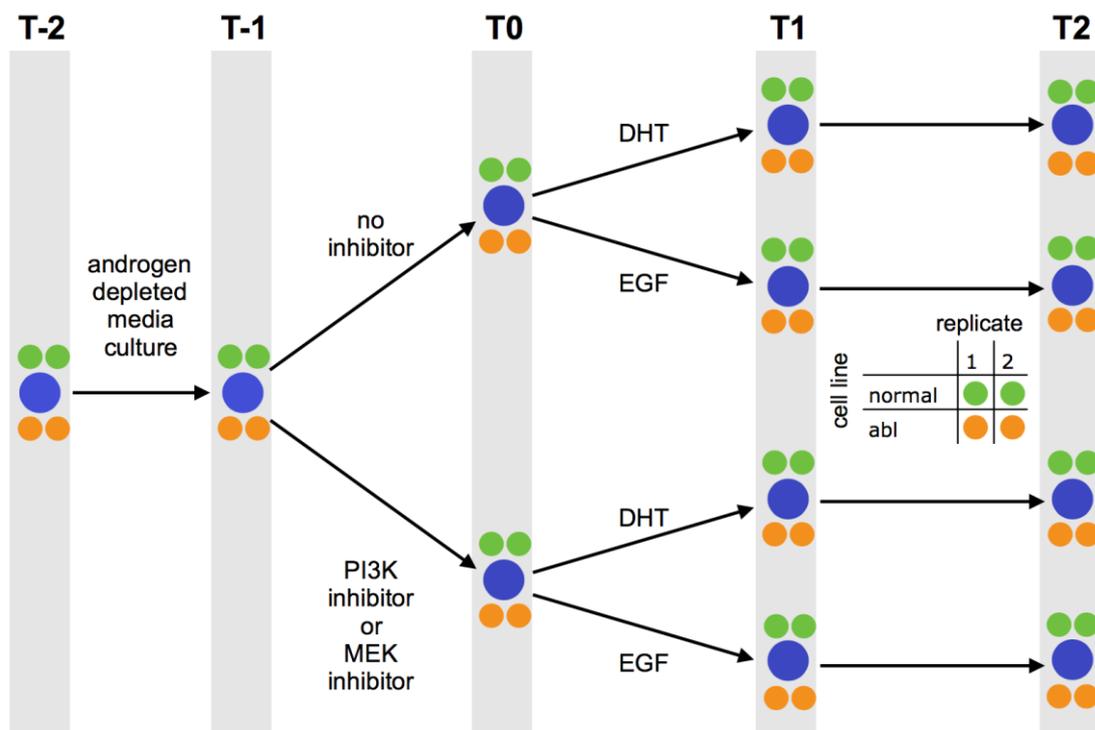
Figure 7: Schematic representation of collected phospho-proteomic samples. The cell lines used are LNCaP and LNCaP-abl, exposed to different combinations of ligands (EGF or DHT) and inhibitors (no inhibitor, PI3K inhibitor or MEK inhibitor) and measured at different time points.
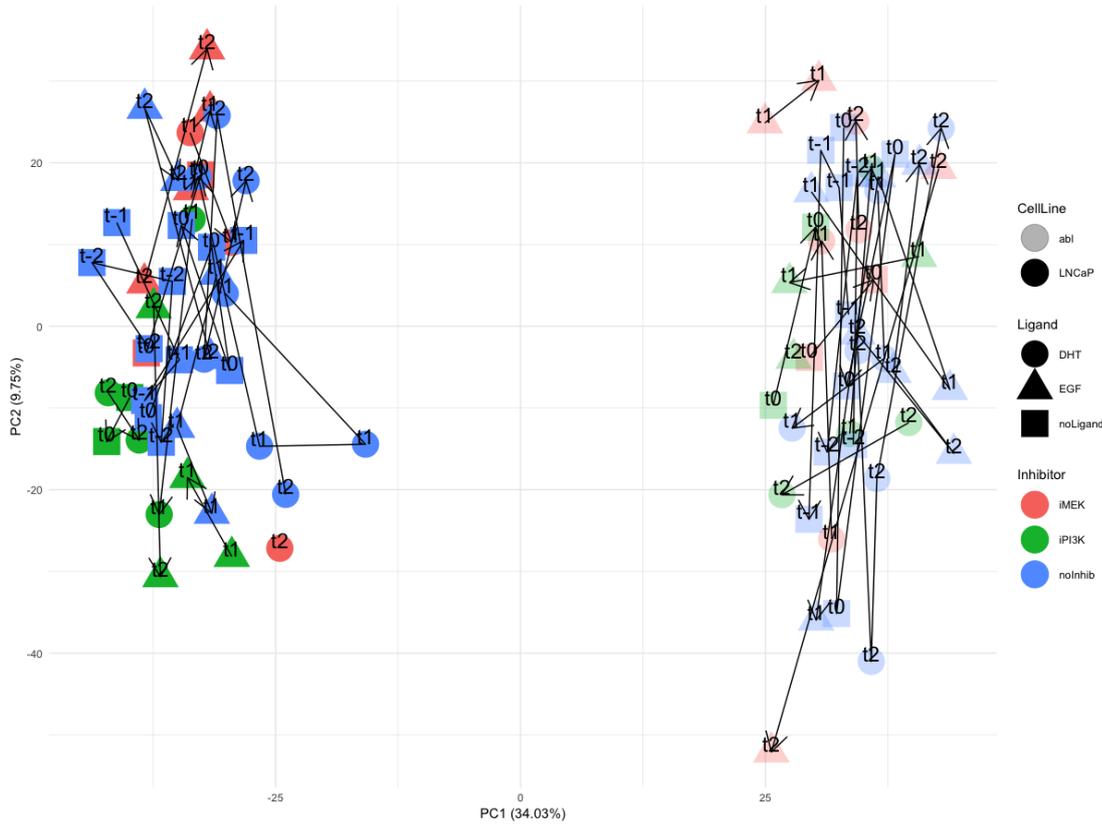
Figure 8: PCA plot of the corrected phospho-proteomic data. Clear differences between cell lines can be seen, but the effect of the different perturbations is not immediately evident. The cell lines are distinguished using different shading. Ligands are distinguished using different shapes. Inhibitors are distinguished using different colors. Lines connect replicate samples, with the arrow pointing to the last replicate.
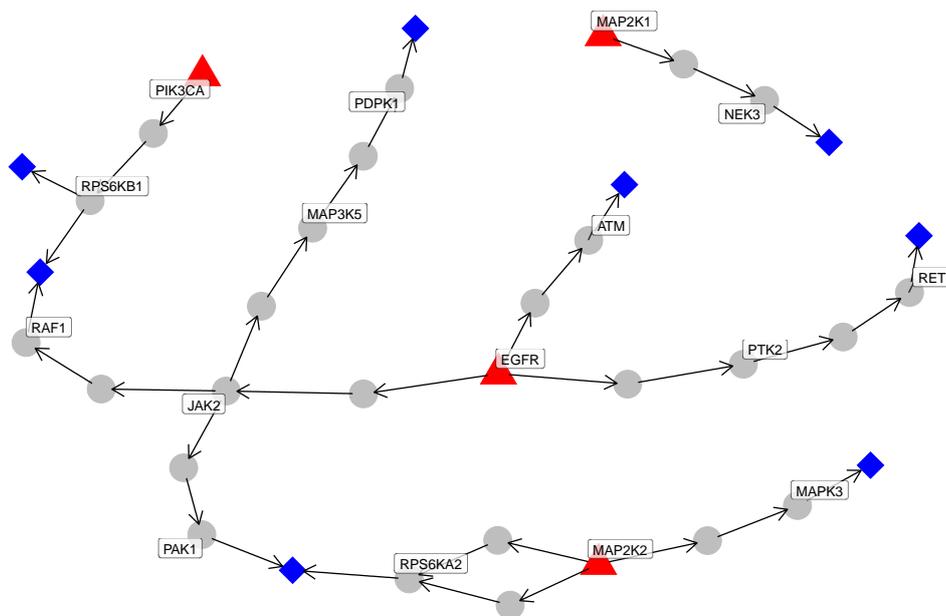
Figure 9: PHONEMeS network for the LNCaP cell line.
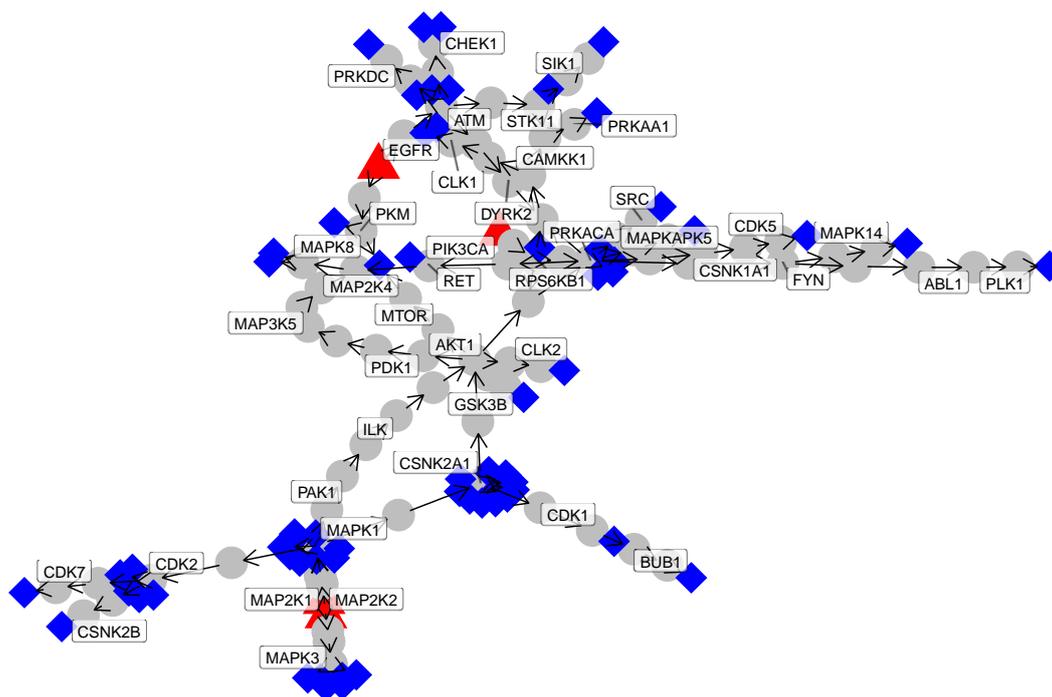
Figure 10: PHONEMeS network for the LNCaP-abl cell line.

Figure 11: All the gene symbols represented in nodes found in the LNCaP network recovered by
PHONEMeS are included in this figure. The LNCaP-abl network recovered by PHONEMeS covers a
total of 97 gene symbols, while the LNCaP network covers only 24 gene symbols.

2.3.1 List of targets proposed

Given the amount of activity difference shown by the two network around MAPK proteins, (the ones related to MEK signalling), it would be of special interest to test that inhibition of some MAPK proteins produces a different effect in LNCaP and LNCaP-abl cell lines. CDK genes could also be targets of interest given their abundance in the LNCaP-abl network in comparison with the LNCaP network.

2.4 Targets based on statistical model of perturbation data

The use of Boolean models is not the only way one could get a list of therapeutic drug target candidates. In fact, there might be interesting targets outside the list of genes covered by the logic model. We considered important to briefly explore the use of a data analysis approach to target discovery in order to provide a comparison between both.

For this part, we made use of the LINCS-L1000 dataset (Subramanian et al. 2017). In this dataset cell lines (~10 different) are perturbed with different compounds (>20,000) with different concentrations and gene expression changes are measured after different elapsed times (e.g. after 6, 24 or 96 hours). This results in a very large (>1,000,000) number of perturbation signatures collected in the form of transcriptomic data. We found that from these perturbation signatures, cell death / toxicity can be effectively predicted. In essence, we matched the LINCS-L1000 signatures with data from cell viability screens (e.g. The Cancer Therapeutics Response Portal (CTRP) or project Achilles) and trained a linear model to predict cell viability from perturbation signature. The models show very nice (Pearson correlation 0.5-0.6) cross-validation performance. The reason behind this is that cell death leads to a unique signature (changes of apoptosis, cell cycle etc. related genes).

These models are trained based on a small subset (~20,000 signatures, ~300 compounds) of LINCS-L1000 data, so it is possible to predict cell viability for the rest of the data, actually using LINCS-L1000 as a "cell viability dataset". The reason behind this is that in the LINCS-L1000 data a large number of non-anticancer drugs are used, so there is a chance that we can find some interesting hits (i.e. drugs that are not used in cancer but in other conditions, and are having cell specific toxicity). We made the cell viability predictions and a computational validation by comparing the results with the NCI60 drug screen. We used NCI60 screen because it does not only include traditional cancer drugs, and because it also involves a large number of compounds (>20,000). The models had a good performance in this validation (ROC AUC>0.7 for detecting toxic compounds), comparable with the performance of experimentally measured values used from different drug screening database ("ground truth").

The LINCS-L1000 dataset contains two prostate cancer cell lines (VCAP and PC3), so we were able to predict toxicity for the drugs used on these cell lines in the LINCS data. Figure 12 shows the prediction results. Each point is a compound, X axis is the predicted toxicity for VCAP and Y axis is the predicted toxicity for PC3 (toxicity values are provided in arbitrary units, with lower values meaning higher cell death). Color code is the highest predicted toxicity for the other (non prostate cancer) cell lines in the LINCS dataset. The most interesting drugs are the ones that demonstrate selective toxicity for prostate cancer cell lines, meaning low score (high toxicity) for either PC3 or VCAP, and high score (low toxicity, blue color) for other cell lines. Marked in red are the most interesting candidates for experimental validation, given their predicted selective toxicity between these two prostate cancer cell lines. These candidate targets proposed for experimental validation are also reported in Table 1.
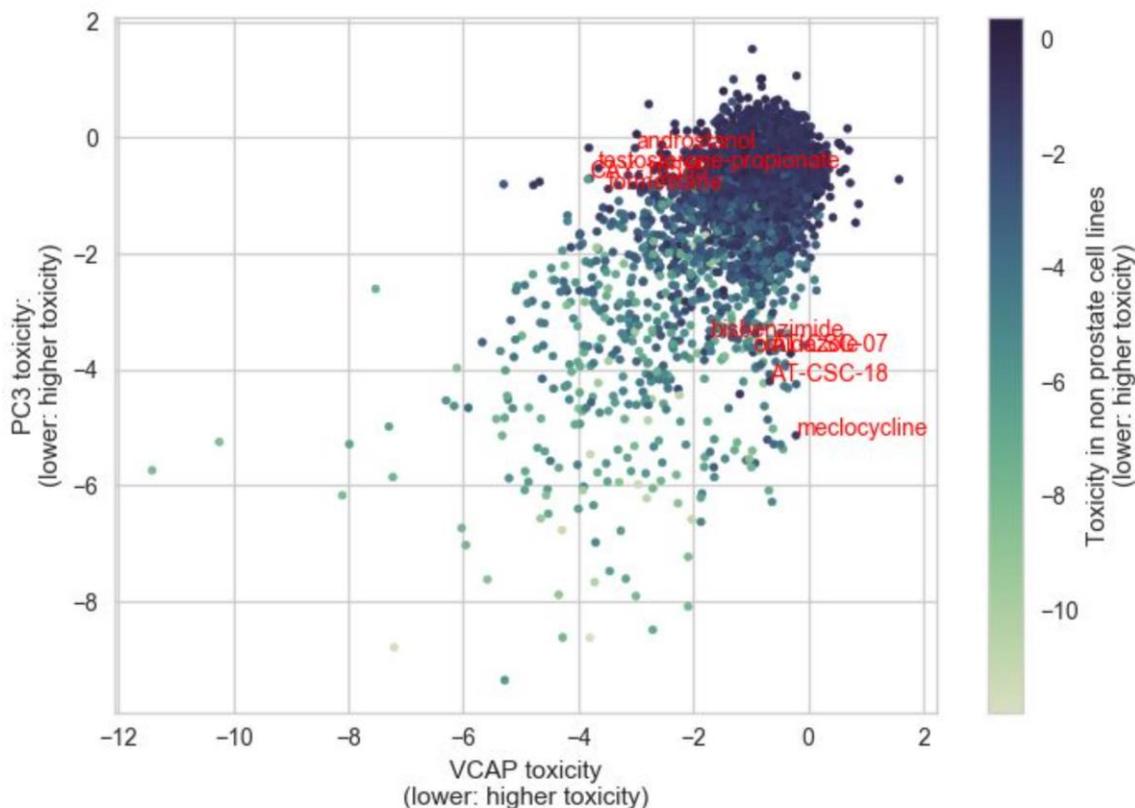
Figure 12: Prediction results of the statistical model of perturbation data. Each point is a compound, X axis is the predicted toxicity for VCAP and Y axis is the predicted toxicity for PC3. Toxicity values are provided in arbitrary units, with lower values meaning higher cell death. Color code is the highest predicted toxicity for the other (non prostate cancer) cell lines in the LINCS dataset. The most interesting drugs are the ones that demonstrate selective toxicity for prostate cancer cell lines, i.e. low score (high toxicity) for either PC3 or VCAP, and high score (low toxicity, blue color) for other cell lines. Marked in red are the most interesting candidates for experimental validation, given their predicted selective toxicity between these two prostate cancer cell lines.

Table 1: List of drug targets and cell lines predicted by the linear statistical model from matching
transcriptomic cell line perturbation data from LINCS-L1000 dataset.

Drug	Selective toxicity in cell	Note
CAY-10585	VCAP	HIF1A inhibitor, some literature about HIF1Ai in prostate cancer
androstanol	VCAP	AR agonist, interesting that it kills prostate cancer cells, some literature about this
testosterone-propionate	VCAP	Same as above
formestane	VCAP	Aromatase inhibitor, also related to AR/steroids
ornidazole	PC3	Anti-protozoa antibiotic
meclocycline	PC3	Tetracycline antibiotic

Chapter 3   Patient-specific drug targets

Using the same methodology as in previous Chapter 2, sections 2.1 and 2.2, we simulated models tailored to patients to be able to identify candidate drugs and combinations of drugs for these specific patients (Task 5.5).

3.1   Use of patient-specific Boolean models

3.1.1   Generic logical model of prostate cancer

As before, the generic logical model of prostate cancer delivered in D5.1 was integrated with data to have tailored models, delivered in D5.3. The results from this set of models capture patient- and cell-line-specific behaviours.

3.1.2   Data from patients used to tailor the model

We used data from The Cancer Genome Atlas (TCGA) database (Cancer Genome Atlas Research Network 2015), which is the largest public resource for information on cancer patients' data. We obtained the data for all the 487 TCGA prostate-cancer patients for which we had CNA, mutations and RNA data available.

In addition to that, we used proteomic profiles of prostate cancer patient tissue punches. The proteome constitutes the majority of cellular machinery that is closet level to the phenotype, such as disease or normal. Proteomic data of sufficient depth are still rare. In this project, PC39 and PPP1 datasets, provided by UZH and ETH, have been extensively used in computational modelling. Both PC39 and PPP1 have profile the proteome of fresh frozen prostate tissue, both normal and tumor, upon radical prostatectomy. PC39 dataset is a multi-level profile of 39 patients from ProCOC cohort, for which exome, transcriptome and proteome data are available. PPP1 data profiles 248 patients from the same cohort, resulting in 1566 samples available for analysis. PPP1 data, described in D6.1 -D6.3 has been acquired 3 years after PC39. Thus, PPP1 dataset serves as validation cohort for results, obtained for PC39 data.

3.1.3   Obtaining patient-specific Boolean models

Following the methodology presented in D5.3, the recipe selected for the prostate TCGA patients was using mutations and CNA as node variants and RNA data as initial conditions and transition rates. In this case, we used mutation or CNA data as their recipes allowed to further discriminate among patients' behaviours.

To validate the method, we matched the patients to available clinical data, e.g. Gleason's score. The results (Figure 13) show good correlations between the Gleason's scores and two of the model phenotypes: Apoptosis and Proliferation. We chose to study those for their importance in cancer progression and because they are thoroughly described in the model.
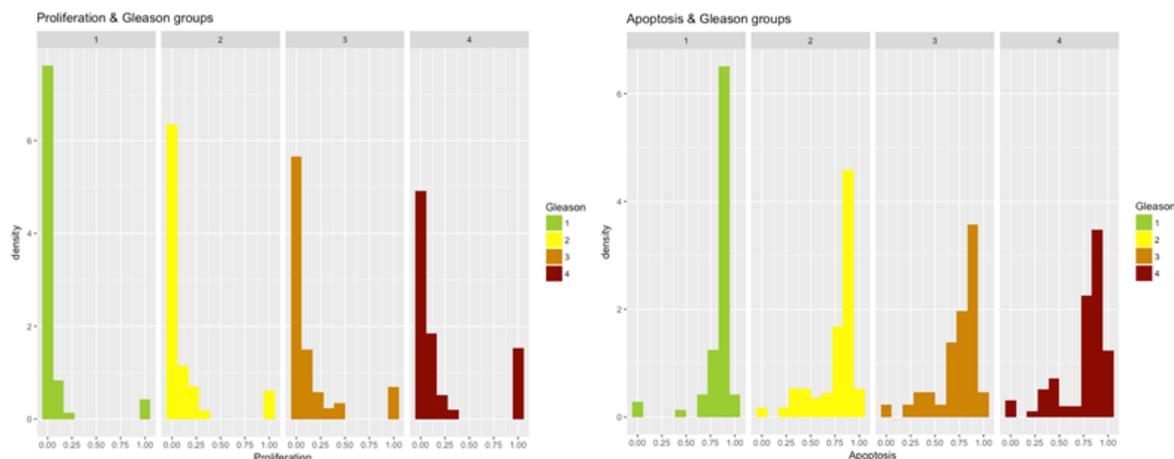
Figure 13: Associations between simulations and Gleason groups (GG). Distribution histograms of Proliferation (a) and Apoptosis (b) scores according to GG; each vignette corresponds to a specific sub-cohort with a fixed GG (1, 2, 3 and 4 from left to right).

3.2 Model-based identification of drug targets specific to patients

As most of the drugs inactivate genes, we looked in each patient for genes that, when inhibited, would hamper *Proliferation* or that promote *Apoptosis* in the model. Interestingly, we found several genes that were found as suitable points of intervention in most of the 487 patients (HSP90 and SHH were identified in more than 80% of the cases) (Figure 14 and Figure 15), but others were specific to only some of the patients (FADD and p21 were identified in only 10% of them).

This list of 322 targets was compared to analyses performed using other datasets from the present project, such as genes identified as targets in cell lines (from Section 2), PPP1, PC39 and the panel of genes used for targeted sequencing in WP1. PPP1 and PC39 are datasets used in present project. All of the 26 genes identified as targets in cell lines are recovered in present analysis. Then, most of the genes of this gene list were identified as being significantly characteristic of PPP1 patients (211 out of 487). Also, 8 genes of this gene list were identified as being significantly characteristic of PC39 patients (AR, CYCS, CDH1, HSP90AA1, HSP90AB1, HSP90B1, HSPB1, LDHA). Lastly, 18 genes of this gene list were also found in the panel of genes used for targeted sequencing in WP1 (AKT1, AR, ERG, ATM, ATR, CHEK2, CDH1, EP300, EZH2, FOXA1, MED12, MYC, CDKN1B, TP53, PIK3CA, PTEN, RB1, SPOP).
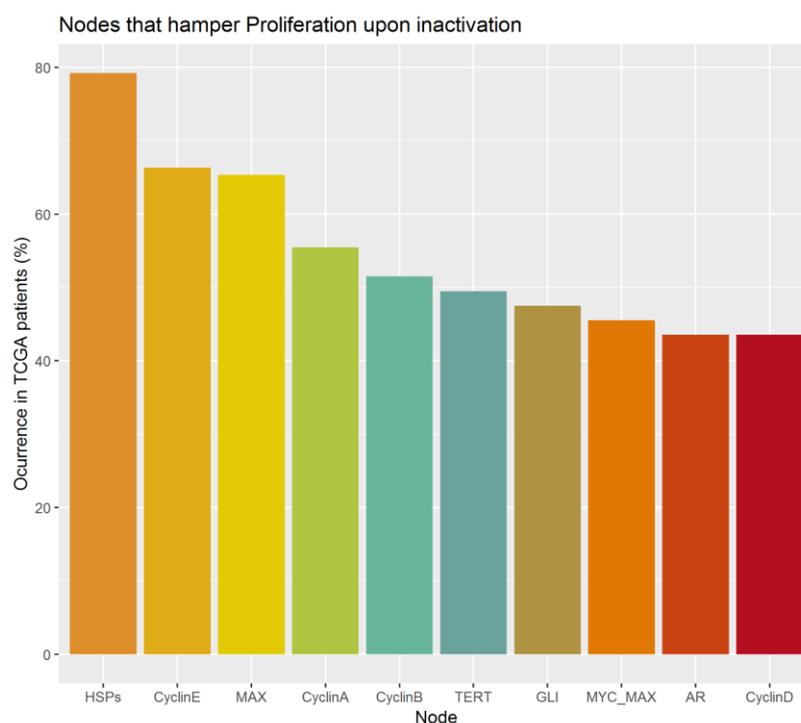
Figure 14: Nodes in the Boolean model that hamper *Proliferation* upon inactivation
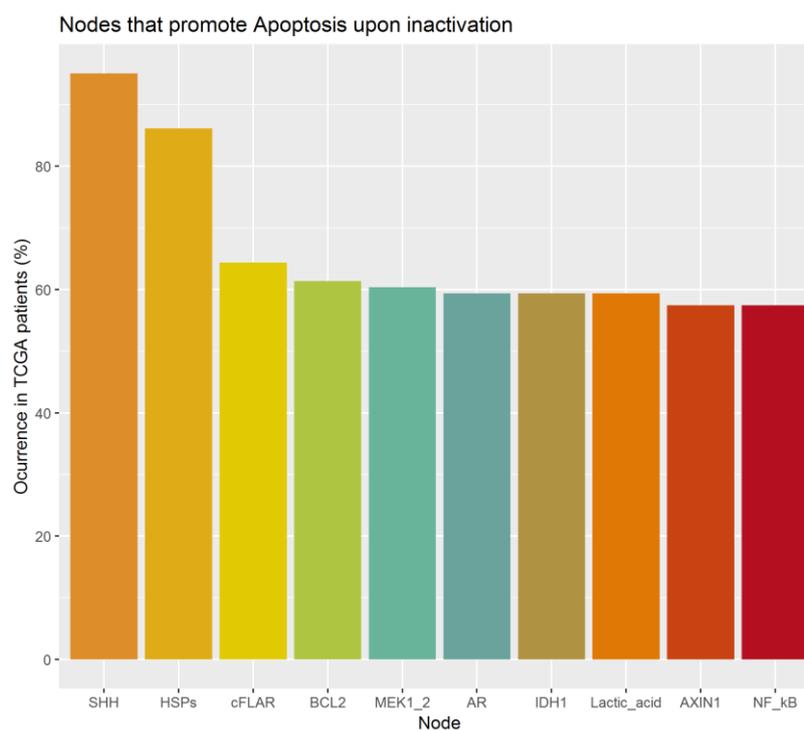
Figure 15: Nodes in the Boolean model that promote *Apoptosis* upon inactivation

Chapter 4   Summary and Conclusion

There are different approaches to drug and target prediction. In this proposal, we have leveraged the use of Boolean modelling approaches (e.g. MaBoSS, CellNOpt and PHONEMeS) in combination with prostate cancer related network information and public and newly generated experimental data. Some Boolean modelling techniques (i.e. CellNOpt and PHONEMeS) require of a specific type of data, phosphoproteomic measurements after perturbation, which is not widely available at the moment. Some of this data has been generated during PrECISE and serves as a proof of concept for this approaches (on top of the value they contain by themselves). Future increased availability of this type of data promises to lead to better models and insights. Notwithstanding, as shown in this deliverable, there are other approaches (i.e. MaBoSS) that provide valuable outputs using other types of data already available. In addition, we have considered a statistical modelling approach for comparison purposes and increasing the coverage of the predictions outside the domain covered by the network models.

The final list of target genes and drugs to be validated in T6.3 is given in Table 2. As can be observed, some of the targets suggested by one model can also be found in another model or when analysing independent datasets. This strengthens the interest in testing these predictions. Furthermore, the models provide the possibility of inspecting them to better understand the predictions and suggest additional experiments if necessary.

The models and strategies used in this deliverable have been focused in prostate cancer. They constitute a pilot case study, but their principles are general and can be applied to any tumour type, opening up the possibility to improved therapies in other fields of oncology.

Table 2: List of proposed drug and targets to be tested in validation experiments.

Target	Type	Notes
HSP90AA1	Gene	Highlighted by the Boolean model and found in other datasets used within PrECISE.
SPOP	Gene	Highlighted by the Boolean model and found in other datasets used within PrECISE.
MAX	Gene	Highlighted by the Boolean model and found in other datasets used within PrECISE.
TERT	Gene	Highlighted by the Boolean model and found in other datasets used within PrECISE.
NOX1, NOX2, NOX3, NOX4	Gene	Highlighted by the Boolean model and found in other datasets used within PrECISE.
AKT	Gene	Relevant in cancer progression.
AR	Gene	Relevant in cancer progression and a target of some of the drugs proposed by the statistical model of perturbation data.
EGFR	Gene	Relevant in cancer progression.

Target	Type	Notes
PI3K	Gene	Relevant in cancer progression.
MEK	Gene	Highlighted by the use of PHONEMeS in the phospho-proteomic data because of the MAPK proteins, some of which also highlighted by the Boolean model. Also relevant in cancer progression.
CDK1, CDK2, CDK5, CDK7	Gene	Highlighted by the use of PHONEMeS in the phospho-proteomic data.
CAY-10585	Drug	HIF-1 inhibitor.
androstanol	Drug	Agonist of the androgen receptor (AR).
testosterone-propionate	Drug	Agonist of the androgen receptor (AR).
formestane	Drug	Aromatase inhibitor, also related to AR/steroids.
ornidazole	Drug	Anti-protozoa antibiotic.
meclocycline	Drug	Tetracycline antibiotic.

Chapter 5   List of Abbreviations

CNA	Copy number alterations
AR	androgen receptor
GnRH	gonadotrophin-releasing hormone
WT	Wild type
PCA	Principal Component Analysis
EGF	Epidermal Growth Factor
ROC	Receiver-Operator Curve
AUC	Area Under the Curve

List of cell lines used in this deliverable

Cell line	Molecular profile source
LNCaP (LNCaP-Clone-FGC)	GDSC, Phosphoproteomic profile by ETH and UZH
LNCaP-abl	Phosphoproteomic profile by ETH and UZH
DU-145	GDSC
PC3	GDSC, LINCS-L1000
VCAP	LINCS-L1000
BPH-1	GDSC

Chapter 6   Bibliography

[1] Cancer Genome Atlas Research Network. 2015. "The Molecular Taxonomy of Primary Prostate Cancer." *Cell* 163 (4): 1011–25.

[2] Stoll, Gautier, Barthélémy Caron, Eric Viara, Aurélien Dugourd, Andrei Zinovyev, Aurélien Naldi, Guido Kroemer, Emmanuel Barillot, and Laurence Calzone. 2017. "MaBoSS 2.0: An Environment for Stochastic Boolean Modeling." *Bioinformatics* 33 (14): 2226–28.

[3] Subramanian, Aravind, Rajiv Narayan, Steven M. Corsello, David D. Peck, Ted E. Natoli, Xiaodong Lu, Joshua Gould, et al. 2017. "A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles." *Cell* 171 (6): 1437–52.e17.

[4] Terfve, Camille, Thomas Cokelaer, David Henriques, Aidan MacNamara, Emanuel Goncalves, Melody K. Morris, Martijn van Iersel, Douglas A. Lauffenburger, and Julio Saez-Rodriguez. 2012. "CellNOptR: A Flexible Toolkit to Train Protein Signaling Networks to Data Using Multiple Logic Formalisms." *BMC Systems Biology* 6 (October): 133.

[5] Terfve, Camille D. A., Edmund H. Wilkes, Pedro Casado, Pedro R. Cutillas, and Julio Saez-Rodriguez. 2015. "Large-Scale Models of Signal Propagation in Human Cells Derived from Discovery Phosphoproteomic Data." *Nature Communications* 6 (September): 8033.

[6] Türei, Dénes, Tamás Korcsmáros, and Julio Saez-Rodriguez. 2016. "OmniPath: Guidelines and Gateway for Literature-Curated Signaling Pathway Resources." *Nature Methods* 13 (12): 966–67.

[7] Yang, Wanjuan, Jorge Soares, Patricia Greninger, Elena J. Edelman, Howard Lightfoot, Simon Forbes, Nidhi Bindal, et al. 2013. "Genomics of Drug Sensitivity in Cancer (GDSC): A Resource for Therapeutic Biomarker Discovery in Cancer Cells." *Nucleic Acids Research* 41 (Database issue): D955–61.