



NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

**SCHOOL OF SCIENCE
DEPARTMENT OF INFORMATICS AND TELECOMMUNICATIONS**

**POSTGRADUATE PROGRAM
"INFORMATION TECHNOLOGIES IN MEDICINE AND BIOLOGY"**

MASTER THESIS

**A method for parsing clinical outcomes and combining them
with phosphoproteomic and genomic data for predicting
drug efficacy: Application in hepatocellular carcinoma**

Orfeas S. Aidonopoulos

Supervisor: Stavros Perantonis, PhD, Scientific Director, N.C.S.R. Demokritos

**ATHENS
NOVEMBER 2014**



ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

**ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ**

**ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
"ΤΕΧΝΟΛΟΓΙΕΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΣΤΗΝ ΙΑΤΡΙΚΗ ΚΑΙ ΤΗ ΒΙΟΛΟΓΙΑ"**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Ανάπτυξη μεθόδου για πρόβλεψη κλινικής αποδοτικότητας
φαρμάκων μέσω ενσωμάτωσης γενομικών, κλινικών και
φωσφοπρωτεομικών δεδομένων: Εφαρμογή στον ηπατικό
καρκίνο**

Ορφέας Σ. Αηδονόπουλος

Επιβλέπων: Σταύρος Περαντώνης, Διευθυντής Ερευνών – ΕΚΕΦΕ Δημόκριτος

ΑΘΗΝΑ

ΝΟΕΜΒΡΙΟΣ 2014

MASTER THESIS

A method for parsing clinical outcomes and combining them with phosphoproteomic and genomic data for predicting drug efficacy: application in hepatocellular carcinoma

Orfeas S. Aidonopoulos

A.M.: PIV099

SUPERVISOR: Stavros Perantonis, PhD, Scientific Director, N.C.S.R. Demokritos

**EXAMINATION
COMMITTEE:**

Stavros Perantonis, PhD, Scientific Director,
N.C.S.R. Demokritos

Leonidas Alexopoulos, Assistant Professor
NTUA

George Vernikos, PhD - Investigator

November 2014

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Ανάπτυξη μεθόδου για πρόβλεψη κλινικής αποδοτικότητας φαρμάκων μέσω ενσωμάτωσης γενομικών, κλινικών και φωσφοπρωτεομικών δεδομένων: Εφαρμογή στον ηπατικό καρκίνο

Ορφέας Σ. Αηδονόπουλος

A.M.: ΠΙΒ099

ΕΠΙΒΛΕΠΩΝ: **Σταύρος Περαντώνης**, Διευθυντής Ερευνών – ΕΚΕΦΕ
Δημόκριτος

ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ: **Σταύρος Περαντώνης**, Διευθυντής Ερευνών
ΕΚΕΦΕ Δημόκριτος
Λεωνίδας Αλεξόπουλος, Επίκουρος Καθηγητής
ΕΜΠ
Γιώργος Βερνίκος, Διδάκτωρ - Ερευνητής

Νοέμβριος 2014

ABSTRACT

Systems biology has become an essential component of drug discovery, attempting to combine experimental data with computational modeling to capture different levels of cellular function (such as signaling, transcription, and regulation) and integrate them in predictive models. These models are then used to best understand the drug mode of action (MOA), identify new targets and predict clinical drug efficacy and toxicity. In this study, we tried to identify signaling pathways related to drug efficacy in one of the most lethal malignancies worldwide, hepatocellular carcinoma (HCC). Particularly, gene expression data were collected for various HCC cell lines treated with anticancer compounds of known clinical efficacy. Each compound had been categorized with a 'pass' or 'fail' label according to their success or failure in human clinical trials. For labeling each drug we constructed a graphical user interface that parses clinical trials databases for clinical outcomes containing the respective compound. Thus, having available a dataset consisting of labeled drugs as observations and genes as features a supervised learning method was applied (feature selection) to identify genes predictive of the drugs' clinical efficacy. Finally, using the extracted data as an input to a pathway construction algorithm, we were able to infer signaling networks on the proteomic level that best fit the measured gene expression signatures. We identified reactions and pathways playing an important role as accurate predictors for the efficacy of nine drugs in HCC.

SUBJECT AREA: Systems Biology, Bioinformatics

KEYWORDS: Machine learning, signaling pathways, pathway construction, gene expression, data analysis

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια ο τομέας της συστημικής βιολογίας αποτελεί ένα από τα κυριότερα πεδία έρευνας για ανακάλυψη νέων φαρμακευτικών ουσιών. Η ανάλυση των πειραματικών δεδομένων με υπολογιστικές μεθόδους μπορεί να χαρτογραφήσει διαφορετικές κυτταρικές λειτουργίες (όπως ενδοκυτταρική σηματοδότηση, μεταγραφή, ρυθμιστικοί παράγοντες) και να τις ενσωματώσει σε προγνωστικά μοντέλα (predictive models). Στη συνέχεια, τα μοντέλα αυτά χρησιμοποιούνται προκειμένου να ερμηνευθεί όσο καλύτερα και πιο συγκεκριμένα γίνεται ο μηχανισμός δράσης των φαρμάκων (mode of action – MOA) με τελικό στόχο την πρόβλεψη της αποδοτικότητας και τοξικότητας της δράσης τους.

Αντικείμενο της παρούσας διπλωματικής εργασίας ήταν ο εντοπισμός σηματοδοτικών μονοπατιών που σχετίζονται με την κλινική αποδοτικότητα φαρμάκων σε μία από τις πιο συνηθισμένες παγκοσμίως ασθένειες, τον ηπατικό καρκίνο (hepatocellular carcinoma – HCC). Αρχικά, συλλέχθηκαν δεδομένα γονιδιακής έκφρασης από διάφορες κυτταρικές σειρές υπό την παρουσία 9 αντικαρκινικών ουσιών και εφαρμόστηκε μία μέθοδος επιλογής χαρακτηριστικών (feature selection) έτσι ώστε να επιλεγθεί ένα υποσύνολο γονιδίων που παίζουν καθοριστικό ρόλο στο MOA των ουσιών αυτών. Για την κατηγοριοποίηση (labeling) των δειγμάτων κατασκευάστηκε ένα εργαλείο-ταξινομητής με το οποίο κάποιος χρήστης επιλέγει ένα οποιοδήποτε φάρμακο ή ουσία και το ταξινομεί με ένα label PASS/FAIL με βάση τις κλινικές δοκιμές που έχουν γίνει για το επιλεγμένο φάρμακο. Το εργαλείο περιέχει, επίσης, ενσωματωμένο έναν web browser όπου φορτώνονται online οι κλινικές δοκιμές που έχουν πραγματοποιηθεί αλλά και έγγραφα από το 'orange book' του οργανισμού FDA (Food and Drug Administration) το οποίο περιέχει πληροφορίες για τα εγκεκριμένα στην αγορά φάρμακα (labels and approval letters). Στη συνέχεια, δίνοντας τα επιλεγμένα γονίδια (gene signature) ως είσοδο σε έναν αλγόριθμο κατασκευής μονοπατιών (pathway construction) μας δίνεται η δυνατότητα να δημιουργήσουμε σηματοδοτικά δίκτυα που περιγράφουν τη μετάδοση κυτταρικού σήματος (από το γενομικό μέχρι το φωσφοπρωτεομικό επίπεδο). Έτσι, με την πλειοψηφία των σύγχρονων φαρμάκων να δρουν στο φωσφοπρωτεομικό επίπεδο, μπορούμε να εντοπίσουμε εκείνες τις φωσφοπρωτεΐνες που παίζουν κύριο ρόλο στην κλινική αποδοτικότητα των υποκείμενων φαρμάκων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Συστημική Βιολογία, Βιοπληροφορική

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Μηχανική μάθηση, σηματοδοτικά μονοπάτια, βιολογικά δίκτυα, γονιδιακή έκφραση, ανάλυση δεδομένων