

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ**

**ΤΜΗΜΑ ΕΠΙΣΤΗΜΗΣ ΥΠΟΛΟΓΙΣΤΩΝ**

**ΠΑΡΟΥΣΙΑΣΗ / ΕΞΕΤΑΣΗ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ**

**Κερκεντζές Κωνσταντίνος**

**Μεταπτυχιακός Φοιτητής**

**Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

Επόπτης Μεταπτ. Εργασίας: Αναπλ. Καθηγητής, Ιωάννης Τσαμαρδίνος

**Τρίτη, 24 Μαρτίου 2015, 12:00**

**Αίθουσα K206, Τμήμα Επιστήμης Υπολογιστών, Πανεπιστήμιο Κρήτης**

**" Ανάλυση «παλαιών» δεδομένων μικροσυστοιχιών καρκίνου του πνεύμονα  
αποκαλύπτει κρυμμένα κομμάτια βιολογικής γνώσης "**

#### **ΠΕΡΙΛΗΨΗ**

Ο καρκίνος του πνεύμονα ιστολογικά και κλινικά είναι μια ετερογενή ομάδα κακοήθων όγκων με αδιευκρίνιστη μοριακή βάση. Επί του παρόντος, χιλιάδες σύνολα μοριακών δεδομένων από καρκίνο του ανθρώπου είναι διαθέσιμα δημοσίως, τα οποία συνήθως αναλύονται μόνο κατά την πρωταρχική τους δημοσίευση. Νέες στατιστικές μέθοδοι και οι όλο και πιο ακριβείς γονιδιακοί χαρακτηρισμοί μπορούν να μετατρέψουν αυτά τα «παλαιά» βιολογικά δεδομένα σε ανανεωμένες πηγές γνώσης με πιθανή κλινική σημασία. Στην εργασία αυτή, παρέχουμε μια *in silico* απόδειξη της ιδέας μας με την εξαγωγή νέων πληροφοριών από ένα υψηλής ποιότητας σύνολο δεδομένων έκφρασης

mRNA, που αρχικά δημοσιεύθηκε το 2001, μέσω της χρήσης σύγχρονων βιοπληροφορικών προσεγγίσεων.

Τα δεδομένα αποτελούνται από ιστολογικά καθορισμένες περιπτώσεις αδενοκαρκινώματος του πνεύμονα, καρκινώματος των πλακώδων κυττάρων του πνεύμονα, μικροκυτταρικού καρκίνου του πνεύμονα, καρκινικών όγκων, μετάστασεων (αδενοκαρκινώματος του μαστού και του παχέους εντέρου) και δείγματα φυσιολογικού πνεύμονα (203 δείγματα συνολικά). Μια ευρεία γκάμα στατιστικών ελέγχων πραγματοποιήθηκε για τον προσδιορισμό γονιδίων με διαφορεική έκφραση, διαγνωστικών και προγνωστικών γονιδίων, εμπλουτισμένων γονιδιακών οντολογιών και βιοχημικών μονοπατιών. Επιπλέον, επικύρωσαμε την προσέγγιση της ανάλυσής μας εφαρμόζοντάς την σε άλλα δύο σύνολα δεδομένων και ελέγχοντας την αντιστοιχία των αποτελεσμάτων μεταξύ και των τριών.

Τα αποτελέσματα της εκ νέου ανάλυσης αυτού του δημόσια διαθέσιμου συνόλου δεδομένων παρουσίασαν γνωστά βιολογικά χαρακτηριστικά των υποκατηγοριών καρκίνου του πνεύμονα. Επιπλέον, αποκαλύφθηκαν νέα μονοπάτια με δυνητικά κλινική σημασία. Τέλος, τα ευρήματα υποστηρίζουν επίσης την υπόθεσή μας ότι ακόμη και τα παλιά «-ωματικά» δεδομένα υψηλής ποιότητας μπορούν να αποτελέσουν μια πηγή σημαντικών βιολογικών πληροφοριών όταν χρησιμοποιούνται κατάλληλες σύγχρονες μέθοδοι της βιοπληροφορικής.

**“Analysis of "old" microarray data of lung cancer reveals hidden pieces of new biological knowledge”**

**ABSTRACT**

Lung cancer is histologically and clinically an heterogeneous group of malignant tumours with obscure molecular basis. Currently, thousands of molecular datasets on human cancer are publicly available, which usually are only analysed when they are first published. Novel statistical methods and increasingly more accurate gene annotations can transform "old" biological data into a renewed source of knowledge with potential clinical relevance. In this work, we provide an in-silico proof-of-concept by extracting

novel information from a high quality mRNA expression dataset, originally published in 2001, using state-of-the-art bioinformatics approaches.

The dataset consists of histologically defined cases of lung adenocarcinoma, squamous cell carcinoma, small-cell lung cancer, carcinoid, metastasis (breast and colon adenocarcinoma) and normal lung specimens (203 samples in total). A battery of statistical tests was used for the identification of differential gene expressions, diagnostic and prognostic genes, enriched gene ontologies and signaling pathways. Furthermore, a validation of our procedure was performed by analysing two more datasets and testing the concordance among the results of all three of them.

The results of the reanalysis of this public dataset displayed the known biological features of lung cancer subtypes. Moreover, novel pathways of potentially clinical importance were revealed. Finally, the findings also support our hypothesis that even old omics data of high quality can be a source of significant biological information when appropriate bioinformatics methods are used.