



Α.Π. : 61977
Αθήνα, 8/1/23

ΚΟΣΜΗΤΟΡΑΣ

Προς
τα Μέλη ΔΕΠ της Σχολής Μηχανολόγων Μηχανικών ΕΜΠ

Π Ρ Ο Σ Κ Λ Η Σ Η

Παρακαλείστε να παρευρεθείτε στην παρουσίαση της Διδακτορικής Διατριβής που εκπόνησε στον Τομέα Μηχανολογικών Κατασκευών & Αυτομάτου Ελέγχου του ΕΜΠ, ο Υ.Δ. κ. **ΑΣΙΜΟΜΗΤΗΣ Γεώργιος του Ανδρέα**, Διπλωματούχου Ηλεκτρολόγου Μηχανικού & Μηχανικών Υπολογιστών ΕΜΠ (Integrated Master) και κατόχου Μεταπτυχιακού Διπλώματος Υπολογιστικής Βιολογίας και Βιοπληροφορικής από το Ομοσπονδιακό Ινστιτούτο Τεχνολογίας της Ζυρίχης (ETH Zurich).

Η παρουσίαση θα πραγματοποιηθεί **τη Δευτέρα 22 Ιανουαρίου 2024, ώρα 17:00**, διαδικτυακά στο κάτωθι link:

<https://meetmsk.zoom.us/j/8424053211>

Ο ελληνικός τίτλος της Διδακτορικής Διατριβής είναι ο εξής:

«Ανάπτυξη αναλυτικών υπολογιστικών πλαισίων για τη σύνδεση βιομοριακών δεδομένων με στόχο την κατανόηση της βιολογίας των μυελογενών κακοηθειών»

Και ο αγγλικός τίτλος έχει ως εξής:

«Development of analytical frameworks for the integration of molecular data views to understand the biology of myeloid malignancies»

Ο Κοσμήτορας της Σχολής

Ι. Αντωνιάδης
Καθηγητής Ε.Μ.Π



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Μηχανολόγων Μηχανικών
Τομέας Μηχανολογικών Κατασκευών & Αυτομάτου Ελέγχου

***Development of analytical frameworks
for the integration of molecular data
views to understand the biology of
myeloid malignancies***

Διδακτορική Διατριβή

Γεώργιος Ασιμομήτης

ΔΙΠΛΩΜΑΤΟΥΧΟΥ ΗΛΕΚΤΡΟΛΟΓΟΥ ΜΗΧΑΝΙΚΟΥ
ΕΜΠ

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Λ. Αλεξόπουλος, Καθ. ΕΜΠ (Επιβλέπων)

Ε. Παπαεμμανουήλ, Αν. Καθ. MSKCC

Ι. Κοτσιανίδης, Καθ. ΔΠΘ

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:

Λ. Αλεξόπουλος, Καθ. ΕΜΠ (Επιβλέπων)

Ε. Παπαεμμανουήλ, Αν. Καθ. MSKCC

Ι. Κοτσιανίδης, Καθ. ΔΠΘ

Χ. Μανόπουλος, Επ. Καθ. ΕΜΠ

Ε. Παπαπέτρου, Καθ. Mt Sinai

Α. Σπυριδωνίδης, Καθ. Παν. Πατρών

Α. Τσιρίγος, Καθ. NYU

Athens, January 2024

Thesis Abstract

Myeloid malignancies consist of a heterogeneous spectrum of clonal stem cell disorders driven by genetic and epigenetic alterations resulting in dysregulated hematopoiesis. The investigation of the mechanisms underpinning myeloid neoplasia relies on experimental models of disease biology and the phenotyping of primary patient samples using emerging genomic technologies. In recognition of the increasing complexity, scale and dimensionality of the datasets generated by these approaches, this thesis develops analytical frameworks within and across different omics modalities (genomics, transcriptomics, epigenomics) and sequencing techniques (bulk and single cell), aiming to enhance the understanding of the underlying biology of myeloid neoplasms. Specifically, this work deploys principles from multi-view data fusion and interconnection to analyze signals in an integrative manner, setting out to elucidate molecular landscapes and assist the study of phenotypes at a genetic level.

Particularly, chapter 2 investigates the transcriptional repertoire and chromatin profile of SF3B1 mutated Myelodysplastic Syndromes (MDS), leveraging bulk RNA and ATAC sequencing data from patient-derived genetically matched normal and SF3B1 mutated induced pluripotent stem cell (iPSC) lines. We introduce a multi-stage fusion framework that merges signals from diverse data layers derived from transcriptome sequencing (splicing, transcript usage, gene expression). The analytical framework developed as part of this work leads to the derivation of a splicing signature linked to 34 genes, which associates with the SF3B1 mutational status of primary MDS patient cells. Additional unimodal chromatin accessibility analysis showed increased priming of SF3B1 hematopoietic progenitors toward the megakaryocyte - erythroid lineage, as well as the enrichment of motifs from the TEA (TEAD) domain in accessible regions linked to genes with upregulated expression. Overall, chapter 2, applies a multi-stage fusion approach on transcriptomic data views to prioritize mis-spliced gene targets, and concurrently provides a formal overview of the SF3B1 mutated chromatin landscape and nominates transcriptional programs with putative roles in MDS disease biology.

Chapter 3 examines if single cell gene expression signals together with the computational capacity of neural networks are able to predict a cell's malignant status and subsequently its genotype for specific abnormalities in IDH1/2 mutated Acute Myeloid Leukemia (AML). To this end, using single cell RNA sequencing data from 6 AML patients and 4 healthy individuals, a feedforward neural network was trained to predict the cell's malignant or wild-type (WT) status in a binary fashion, achieving an accuracy of 98% on the holdout test-set. Furthermore, in a multi-label setting, this work deploys a similar architecture to predict the mutational status of specific genomic abnormalities at the single cell level, showing a macro-average AUC ROC=0.84 and NRAS mutational status prediction AUC ROC=0.83 on the holdout test-set. Altogether, chapter 3 applies deep learning in a supervised context to explore the connection between single cell gene expression profiles and genotypes in IDH1/2 AML and shows the potential of such modeling approaches in capturing meaningful genotypic-phenotypic relationships.

Περίληψη

Οι μυελογενείς κακοήθειες αποτελούν ετερογενείς διαταραχές κλωνικών βλαστοκυττάρων, που οφείλονται σε γενετικές και επιγενετικές αλλοιώσεις και οδηγούν σε ελαττωματική αιμοποίηση. Η έρευνα των μηχανισμών των μυελογενών νεοπλασμάτων βασίζεται σε πειραματικά μοντέλα βιολογίας και στο φαινοτυπικό προσδιορισμό πρωτογενών δειγμάτων ασθενών μέσω ανερχόμενων γονιδιωματικών τεχνολογιών. Στο περιθώριο της αυξανόμενης πολυπλοκότητας, του όγκου και της διαστασιμότητας των δεδομένων που δημιουργούνται από αυτές τις πρακτικές, η εν λόγω διατριβή αναπτύσσει υπολογιστικά πλαίσια χρησιμοποιώντας διαφορετικά ομικά προφίλ (γονιδιωματικά, μεταγραφωματικά, επιγονιδιωματικά) και είδη τεχνικών αλληλουχίας (μαζικής και μεμονωμένων κυττάρων), με στόχο να ενισχύσει την κατανόηση της υποκείμενης βιολογίας των μυελογενών νεοπλασμάτων. Συγκεκριμένα, η εργασία αυτή στηρίζεται στη σύμπραξη και τη διασύνδεση πολλαπλών όψεων δεδομένων για την ολιστική ανάλυση σημάτων, αποσκοπώντας να αποσαφηνίσει μοριακά τοπία και να επικουρήσει τη μελέτη φαινότυπων σε γενετικό επίπεδο.

Το Κεφάλαιο 2 ερευνά τα τοπία μεταγραφώματος και χρωματίνης των SF3B1 μεταλλαγμένων Μυελοδυσπλαστικών Συνδρόμων (MDS), αξιοποιώντας δεδομένα μαζικής αλληλουχίας RNA και ATAC από ισογονικές υγιείς και SF3B1 μεταλλαγμένες σειρές επαγόμενων πολυδύναμων βλαστικών κυττάρων. Ειδικά, υλοποιούμε ένα αναλυτικό πλαίσιο συγχώνευσης πληροφοριών από διαφορετικά επίπεδα δεδομένων αλληλουχίας RNA (μάτισμα, χρήση μεταγραφημάτων, έκφραση γονιδίου). Το πλαίσιο αυτό ξεχωρίζει ένα σύνολο συμβάντων ματίσματος από 34 γονίδια, το οποίο σχετίζεται με την κατάσταση της μετάλλαξης SF3B1 σε πρωτογενή δείγματα ασθενών με MDS. Παράλληλα, η ανάλυση της προσβασιμότητας της χρωματίνης δείχνει αυξημένη παρουσία μοτίβων TEAD σε ένα σύνολο ανοιχτών περιοχών της καθώς και αυξημένη κλίση των αιμοποιητικών προγονικών κυττάρων SF3B1 προς την κατεύθυνση των μεγακαρυοκυττάρων - ερυθροειδών. Συνολικά, το κεφάλαιο αυτό, συνδυάζει προβολές από δεδομένα RNA για να κατηγοριοποιήσει γονιδιακούς στόχους με ελαττωματικό μάτισμα, ενώ επίσης παρέχει μια επισκόπηση του τοπίου της χρωματίνης SF3B1 MDS και προτείνει προγράμματα μεταγραφής με πιθανούς ρόλους στη βιολογία της νόσου MDS.

Το Κεφάλαιο 3 εξετάζει εάν η έκφραση γονιδίων μεμονωμένων κυττάρων μαζί με την υπολογιστική ικανότητα των νευρωνικών δικτύων μπορούν να προβλέψουν χαρακτηριστικά του κυτταρικού γονότυπου στην IDH1/2 μεταλλαγμένη Οξεία Μυελογενή Λευχαιμία (AML). Συγκεκριμένα, χρησιμοποιώντας δεδομένα αλληλουχίας RNA μεμονωμένων κυττάρων από 6 ασθενείς με AML και 4 υγιή άτομα, εκπαιδεύτηκε ένα νευρωνικό δίκτυο που προβλέπει την κακοήθη ή υγιή κατάσταση του κυττάρου με ακρίβεια 98%. Στη συνέχεια, δοκιμάστηκε μια παρόμοια αρχιτεκτονική για την ταυτόχρονη πρόβλεψη συγκεκριμένων γονιδιωματικών ανωμαλιών σε μεμονωμένα κύτταρα, η οποία παρουσίασε macro-average AUC ROC=0.84 και AUC ROC=0.83 για την πρόβλεψη της μετάλλαξης NRAS. Εν κατακλείδι, το κεφάλαιο 3, μέσω επιβλεπόμενου deep learning, συνδέει τα προφίλ γονιδιακής έκφρασης και γονότυπου μεμονωμένων κυττάρων στην IDH1/2 μεταλλαγμένη AML και δείχνει την προοπτική τέτοιων προσεγγίσεων μοντελοποίησης στην αποτύπωση σχέσεων γονότυπου-φαινότυπου.