



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

# ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΥΤΤΑΡΙΚΗΣ ΑΠΟΚΡΙΣΗΣ ΜΕ ΜΕΘΟΔΟΥΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ  
**Θεόδωρος Σακελλαρόπουλος**  
ΔΙΠΛΩΜΑΤΟΥΧΟΥ ΜΗΧΑΝΟΛΟΓΟΥ ΜΗΧΑΝΙΚΟΥ ΕΜΠ

**ΕΠΙΒΛΕΠΩΝ:**  
Λ.Γ. Αλεξόπουλος  
Αν. Καθηγητής, ΕΜΠ

Αθήνα, Ιούλιος 2018





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

# ΜΟΝΤΕΛΟΠΟΙΗΣΗ ΚΥΤΤΑΡΙΚΗΣ ΑΠΟΚΡΙΣΗΣ ΜΕ ΜΕΘΟΔΟΥΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

**Θεόδωρος Σακελλαρόπουλος**

ΔΙΠΛΩΜΑΤΟΥΧΟΥ ΜΗΧΑΝΟΛΟΓΟΥ ΜΗΧΑΝΙΚΟΥ ΕΜΠ

**ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ  
ΕΠΙΤΡΟΠΗ:**

Λ ΑΛΕΞΟΠΟΥΛΟΣ, Αν. Καθ. ΕΜΠ  
(Επιβλέπων)

Χ. ΠΡΟΒΑΤΙΔΗΣ, Καθ. ΕΜΠ

Σ. ΤΣΑΓΓΑΡΗΣ, Ομ. Καθ. ΕΜΠ

**ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ  
ΕΠΙΤΡΟΠΗ:**

Λ ΑΛΕΞΟΠΟΥΛΟΣ, Αν. Καθ. ΕΜΠ  
(Επιβλέπων)

Χ. ΠΡΟΒΑΤΙΔΗΣ, Καθ. ΕΜΠ

Ζ. ΚΟΥΡΝΙΑ, Ερευν. Γ. ΠΒΕΑΑ

Ν. ΧΟΝΔΡΟΓΙΑΝΝΗ, Ερευν. Β. ΕΙΕ

Γ. ΜΑΤΣΟΠΟΥΛΟΣ, Αν. Καθ. ΕΜΠ

Α. ΤΣΙΡΙΓΟΣ, Αν. Καθ. NYU

Α. ΧΑΤΖΗΩΑΝΝΟΥ, Ερευν. Β. ΕΙΕ

**Αθήνα, Ιούλιος 2018**



Η παρούσα εργασία χορηγείται με άδεια Creative Commons Attribution - Share Alike 4.0 International. Αντίγραφο της άδειας βρίσκεται στην ιστοσελίδα: <http://creativecommons.org/licenses/by-sa/4.0/>

---

Η έγκριση της διδακτορικής διατριβής από την Ανώτατη Σχολή Μηχανολόγων Μηχανικών του Ε.Μ. Πολυτεχνείου δεν υποδηλώνει αποδοχή των γνώμων του συγγραφέα (Ν. 5343/1932, Άρθρο 202)



## Πρόλογος

Η έρευνα για την εν λόγω διατριβή πραγματοποιήθηκε στο εργαστήριο του Αναπληρωτή Καθηγητή της σχολής Μηχανολόγων Μηχανικών του Ε.Μ.Π Λεωνίδα Γ. Αλεξόπουλου στον τομέα κατασκευών και αυτομάτου ελέγχου. Κατά την διάρκεια του διδακτορικού μου επισκέφθηκα και συνεργάστηκα με τα εξής εργαστήρια/οργανισμούς: Ευρωπαϊκό κέντρο βιοπληροφορικής (EBI) στο Κέμπριτζ του Ηνωμένου Βασιλείου, ομοσπονδιακός οργανισμός τροφίμων και φαρμάκων των ΗΠΑ (FDA), Illumina Inc., και νοσοκομείο Langone του πανεπιστημίου της Νέας Υόρκης (NYU Langone Health). Η χρηματοδότηση της έρευνα έγινε κυρίως μέσω του προγράμματος ERC “Investing in knowledge society through the European Social Fund” (Grant no. ERC-11/MIS:374071) καθώς και με πόρους του εκάστοτε οικοδεσπότη οργανισμού.

Θα ήθελα να ευχαριστήσω πολύ τους ακόλουθους ανθρώπους για την βοήθεια τους. Αρχικά την διεθνή επιστημονική κοινότητα και την κοινότητα ανοιχτού λογισμικού για την ανιδιοτελή προσφορά τους σε εργαλεία και δεδομένα. Όλα τα μέλη του εργαστηρίου Εμβιομηχανικής της σχολής Μηχανολόγων Μηχανικών του ΕΜΠ μεταξύ 2012 και 2018 και ιδιαίτερα τους Δημήτρη Μεσσίνη, Δανάη Κιρλή, Γιώργο Κανακάρη, Δημήτρη Τζεράνη, Σταυρούλα Σαμαρά, Ilona Binenbaum, Κατερίνα Σκορδά, Αγγελική Μήνια, Ορφέα Αηδονόπουλο, Jan Rozanc και Asier Antoranz Martinez για την συνεργασία ή/και τις εποικοδομητικές συζητήσεις μας καθώς και τις Σοφία Σταματάτου και Γεωργία Μοσχοπούλου για την υποστήριξη. Από το Εθνικό Ίδρυμα Ερευνών (ΕΙΕ) τους ερευνητές Αριστοτέλη Χατζηνιάννου και Νίκη Χονδρογιάννη για τις συμβουλές τους και τη συνεργασία μας. Από την IBM τον Erhan Bilal και από την PMI την Carine Poussin για τη συνεργασία μας στον διαγωνισμό SBV Improver. Όλα τα μέλη του Saez-Rodriguez group στο EBI και ιδιαίτερα τον Καθηγητή Julio Saez-Rodriguez για την φιλοξενία και την καθοδήγηση του καθώς και τον μεταδιδακτορικό φοιτητή Marti Bernado Faura για τη συνεργασία μας. Από τον FDA την Jane Bai για τη φιλοξενία της και τον Timothy Herod για τη συνεργασία μας. Όλα τα μέλη του τμήματος βιοπληροφορικής της Illumina Inc. και ιδιαίτερα τους ερευνητές Μίαιο He, Jennifer Becq, και Επαμεινώνδα Φριτζίλα για την καθοδήγηση τους. Όλα τα μέλη των Aifantis και Tsirigos Lab στο NYU Langone Health και ιδιαίτερα τους Ιannis Aifantis για την φιλοξενία, Αριστοτέλη Τσιρίγο για την καθοδήγηση καθώς και τους διδακτορικούς και μεταδιδακτορικούς φοιτητές Igor Dolgalev, Nicolas Coudray, Χαρίλαο Λαζάρη, Yixiao Gong, Matthew Witkowski, Anastasia Tikhonova για τη συνεργασία μας. Τους καθηγητές της σχολής Εφαρμοσμένων Μαθηματικών και Φυσικών Επιστη-

μών του Ε.Μ.Π, Δημήτρη Φουσκάκη και Μιχαήλ Λουλάκη για τη συνεργασία μας και την κατανόηση που μου έδειξαν. Τα μέλη της τριμελούς συμβουλευτικής επιτροπής Χριστόφορο Προβατίδη και Σωκράτη Τσαγγάρη για την υποστήριξη τους και τα μέλη της επταμελούς εξεταστικής επιτροπής για την κριτική τους. Τα μέλη της γραμματείας της σχολής Μηχανολόγων Μηχανικών και ειδικά τις Δήμητρα Πουλά και Δήμητρα Δαρδαμάνη για την υποστήριξη τους όλα αυτά τα χρόνια.

Τέλος, θα ήθελα να ευχαριστήσω ιδιαίτερα τους ακόλουθους: Τον επιβλέποντα καθηγητή μου Λεωνίδα Αλεξόπουλο, για την καθοδήγηση και την πνευματική ελευθερία που μου παρείχε. Τον Ιωάννη Μελά, με τον οποίο συνεργάστηκα στο ΕΜΠ, ΕΒΙ και FDA, για τις συζητήσεις μας, που διάνθισαν όλες τις ιδέες της παρούσας εργασίας. Την οικογένεια μου και τη Μαρίνα που με στήριξαν υλικά και ψυχικά καθ' όλη την διάρκεια των σπουδών μου. Όλοι οι προαναφερθέντες της παραγράφου έχουν προσφέρει, εν γνώσει τους, πολύ περισσότερα από όσα θα μπορούσα να τους ανταποδώσω και για αυτό τους είμαι ευγνώμων. Η παρούσα εργασία δεν θα ήταν δυνατή χωρίς αυτούς.

**Θεόδωρος Σακελλαρόπουλος**  
Αθήνα, Μάιος 2018



## Περίληψη

Η εν λόγω διδακτορική διατριβή πραγματεύεται την μοντελοποίηση κυτταρικών σηματοδοτικών μονοπατιών με χρήση μεθόδων βελτιστοποίησης, με σκοπό την κατανόηση της διαφοροποίηση των κυττάρων μεταξύ φυσιολογικών και παθολογικών καταστάσεων και εφαρμογές στη συστημική φαρμακολογία. Ο υποψήφιος διδάκτωρ ανέπτυξε μεθόδους Ακέραιου Γραμμικού Προγραμματισμού (ΑΓΠ/ILP) για να εξερευνήσει συστηματικά την συνδεσμολογία των κόμβων που απαρτίζουν το δίκτυο μονοπατιών και να προτείνει μοντέλα που περιγράφουν συσχετίσεις που έχουν παρατηρηθεί πειραματικά. Οι αλγόριθμοι που αναπτύχθηκαν παρέκκλιαν μερικούς σημαντικούς περιορισμούς που είχαν παραδοσιακά οι σχετικοί αλγόριθμοι που έκαναν χρήση ILP τεχνικών όπως η αδυναμία μοντελοποίησης πειραμάτων με άγνωστη αρχική κατάσταση, η αδυναμία να εξερευνήσουν κυκλικά δίκτυα καθώς και η έμμεση μοντελοποίηση της παρουσίας ή μη σχετικών πρωτεϊνών. Κατά τη διάρκεια της διατριβής του ο υποψήφιος διδάκτωρ χρησιμοποίησε τους αλγόριθμους αυτούς για να μελετήσει ποικίλα βιολογικά φαινόμενα όπως η μεταφρασσιμότητα των σηματοδοτικών δικτύων μεταξύ ποντικών και ανθρώπων, οι μηχανισμοί με τους οποίους διάφορα φάρμακα προκαλούν βλάβη στον πνεύμονα, κοινά μοτίβα τοξικότητας φαρμάκων σε διαφορετικά όργανα καθώς και η κυτταρική γήρανση. Αποτελέσματα της έρευνας αυτής δημοσιεύτηκαν σε έγκριτα επιστημονικά περιοδικά και διεθνή συνέδρια.



# Εκτεταμένη Περίληψη

## Εισαγωγή

*Συστημική Βιολογία* είναι ένα μοντέλο βιολογικής έρευνας που προάγει την ολιστική περιγραφή των υπό μελέτη μοντέλων εν αντιθέσει με την “παραδοσιακή” Βιολογία που εστιάζει την περιγραφή των επιμέρους στοιχείων. Η φιλοδοξία για μία ολιστική κατανόηση των βιολογικών συστημάτων δεν είναι καινούργια, αλλά μόνο τις τελευταίες δύο δεκαετίες με την ραγδαία ανάπτυξη των τεχνολογιών μοριακής μέτρησης κατέστη δυνατή η συλλογή του απαραίτητου όγκου πληροφορίας που απαιτεί ένα τέτοιο εγχείρημα. Ο Δρ. Κιτάνο, ένας εκ των “πατέρων” της σύγχρονης αναγέννησης του τομέα, υποδεικνύει 4 κατευθύνσεις αυτής της έρευνας:

- αναγνώριση της δομής του υπό μελέτη συστήματος
- προσδιορισμός των δυναμικών του ιδιοτήτων
- ανάπτυξη τεχνικών για τον έλεγχο του συστήματος
- ανάπτυξη μεθόδων για τον σχεδιασμό νέων συστημάτων

Η παρούσα διατριβή εστιάζει στην μελέτη της κυτταρικής συμπεριφοράς μέσω της αναγνώρισης της δομής των σηματοδοτικών δικτύων του κυττάρου.

Το κύτταρο μοντελοποιείται ως ένα σύστημα πρωτεϊνών και γονιδίων. Η κατάσταση του συστήματος περιγράφεται από τη συγκέντρωση και τη μορφή των πρωτεϊνών καθώς και την έκφραση των γονιδίων που το αποτελούν. Οι πρωτεΐνες είναι μακρομόρια, η μορφή και η συγκέντρωση των οποίων ρυθμίζει τις μακροσκοπικές ιδιότητες του κυττάρου, ενώ τα γονίδια είναι κομμάτια του γενετικού υλικού του κυττάρου, η έκφραση των οποίων ρυθμίζει την παραγωγή πρωτεϊνών. Το σύστημα αυτό βρίσκεται μέσα σε ένα περιβάλλον από το οποίο δέχεται “σήματα” τα οποία επεξεργάζεται και ανταποκρίνεται αλλάζοντας την κατάσταση του.

Τα *σηματοδοτικά δίκτυα* είναι ο μηχανισμός μέσω του οποίου το κύτταρο επεξεργάζεται και ανταποκρίνεται σε πληροφορίες από το περιβάλλον του. Συγκεκριμένα, το κύτταρο επικοινωνεί με το περιβάλλον του κυρίως μέσω ειδικών πρωτεϊνών που καλούνται *υποδοχείς*. Οι πρωτεΐνες αυτές, κατά πλειονότητα, προεξέχουν εκατέρωθεν της κυτταρικής μεμβράνης και όταν έρθουν σε επαφή με κάποιο περιβαλλοντικό ερέθισμα αλλάζουν μορφή. Η αλλαγή αυτή αναγνωρίζεται με τη σειρά της από τις γύρω πρωτεΐνες και διαδίδεται μέσω μίας αλληλουχίας

βιοχημικών αντιδράσεων μέχρι τον πυρήνα του κυττάρου όπου επηρεάζει την έκφραση των γονιδίων και εν τέλει αλλάζει τη συνολική κατάσταση του κυττάρου.

Μία από τις σημαντικότερες σηματοδοτικές αντιδράσεις είναι αυτή της (απο)-φωσφορυλίωσης, κατά την οποία μία πρωτεΐνη αποκτά (χάνει) μία φωσφορική ομάδα με αποτέλεσμα να αλλάξει η μορφή της. Η αλλαγή αυτή, “ενεργοποιεί” (σηματοδοτικά) την πρωτεΐνη. Οι αντιδράσεις αυτές καταλύονται από ειδικά ένζυμα: τις κινάσες για την φωσφορυλίωση και τις φωσφατάσες για την αποφωσφορυλίωση. Για την παρούσα διατριβή, η μέτρηση των φωσφορυλιωμένων πρωτεϊνών έγινε με τη τεχνική bead-based sandwich ELISA. Η τεχνική αυτή επιτρέπει την μέτρηση ενός προκαθορισμένου πλήθους πρωτεϊνών σε πολλά δείγματα.

Ο όρος *γονιδιακή έκφραση* αναφέρεται στη διαδικασία κατά την οποία μέρος του γενετικού κώδικα (γονίδιο) μεταγράφεται σε RNA και βγαίνει από τον πυρήνα με σκοπό να μεταφραστεί σε πρωτεΐνη. Ως *ένταση* της έκφρασης ορίζεται το πλήθος των αντιγράφων RNA το οποίο είναι ανάλογο της συγκέντρωσης της παραγόμενης πρωτεΐνης. Για τις ανάγκες αυτής της διατριβής χρησιμοποιήθηκαν μετρήσεις τύπου μικροσυστοιχίας DNA από δημόσιες βάσεις δεδομένων. Οι τεχνικές αυτές έχουν τη δυνατότητα να μετρούν την έκφραση ενός προκαθορισμένου αριθμού γονιδίων σε πολλά δείγματα.

Η αναγνώριση της δομής των σηματοδοτικών δικτύων συνίσταται στην περιγραφή των σηματοδοτικών αντιδράσεων μεταξύ πρωτεϊνών και γονιδίων. Προς αυτή την κατεύθυνση, έχουν αναπτυχθεί πολλές τεχνικές. Οι σημαντικότερες για την αναγνώριση πρωτεϊνικών αντιδράσεων είναι η yeast-two-hybrid (Y2H) και οι AP-MS, CoFrac-MS που βασίζονται στην τεχνολογία της φασματομετρίας μάζας (Mass Spectrometry). Για τον προσδιορισμό των αντιδράσεων μεταξύ πρωτεϊνών (μεταγραφικοί παράγοντες) και γονιδίων επίσης έχουν αναπτυχθεί πολλές τεχνολογίες οι σημαντικότερες των οποίων βασίζονται στην τεχνική της ανοσοκατακρήμνισης χρωματίνης (chromatin immunoprecipitation (ChIP-X)).

Οι τεχνικές αυτές έχουν δεχθεί κριτική για την ακρίβεια και την ευαισθησία τους, αλλά ο πιο σημαντικός περιορισμός τους δεν είναι τεχνικός αλλά θεμελιώδης. Τα σηματοδοτικά δίκτυα δεν είναι στατικές οντότητες οι οποίες μπορούν να μετρηθούν αν χρησιμοποιηθούν αρκετοί πόροι, αλλά μεταβάλλονται και προσαρμόζονται ανά κύτταρο και στις συνθήκες του περιβάλλοντος.

## Μέθοδοι

Έχουν αναπτυχθεί πολλές μέθοδοι που χρησιμοποιούν τον κατάλογο των γνωστών αντιδράσεων σαν ένα δίκτυο πρότερης γνώσης (Prior Knowledge Network – PKN) και το προσαρμόζουν σε πειραματικά δεδομένα που περιγράφουν την απόκριση, ως αλλαγή στην κατάσταση, των υπό μελέτη κυττάρων σε εξωτερικά ερεθίσματα. Με αυτό το τρόπο επιτυγχάνουν όχι μόνο την αναγνώριση των ειδικών σηματοδοτικών δικτύων, αλλά και περιορίζουν το θόρυβο των μετρήσεων μέσω της σύμπλεξης τους.

Για την παρούσα διατριβή, αναπτύχθηκαν δύο τέτοιες μέθοδοι. Η μέθοδος Συνέπειας Επιρροών (Sign Consistency Method – SCM) και η μέθοδος Αντίστροφου Αιτιακού Λογισμού (Reverse Causal Reasoning – RCR). Οι μέθοδοι αυτές μοντελοποιούν τη διαδικασία σηματοδότησης σαν ένα σύστημα λογικών εξισώσεων. Όπως όλα τα μοντέλα που βασίζονται στη λογική, στοχεύουν σε μία ποιοτική περιγραφή του συστήματος και αντισταθμίζουν τη “θυσία” στη διακριτική τους ικανότητα με το γεγονός ότι δε χρειάζονται πολύ μεγάλο όγκο δεδομένων συγκριτικά με άλλες πιο ρεαλιστικές μεθόδους.

Οι μέθοδοι που αναπτύχθηκαν γενικεύουν και εξελίσσουν μία σειρά μεθόδων στις οποίες έχει συμβάλει το εργαστήριο Συστημικής Βιοϊατρικής της σχολής Μηχ. Μηχ. του ΕΜΠ. Συγκεκριμένα, και οι δύο μέθοδοι χρησιμοποιούν Γράφους Επιρροής (interaction or influence graph (IG)) για να μοντελοποιήσουν το δίκτυο και το μοντέλο της συνέπειας επιρροών για τον μηχανισμό διάδοσης σήματος. Τέλος, όπως και στις προηγούμενες δημοσιεύσεις, χρησιμοποιείται η τεχνική του ακέρατου γραμμικού προγραμματισμού για την επίλυση των προβλημάτων βελτιστοποίησης που σχεδιάζονται.

Οι Γράφοι Επιρροής είναι προσημασμένοι κατευθυνόμενοι γράφοι  $G = (N, A, \sigma)$  όπου:  $N$  το σύνολο των κόμβων,  $A$  το σύνολο των ακμών, και  $\sigma : A \rightarrow \{+1, -1\}$ . Οι κόμβοι του γράφου αντιστοιχούν σε πρωτεΐνες ή γονίδια ενώ οι ακμές σε αντιδράσεις μεταξύ του. Το πρόσημο της ακμής υποδεικνύει την επιρροή που ασκεί ο μητρικός κόμβος στον θυγατρικό. Συγκεκριμένα, οι κόμβοι του δικτύου μπορούν να βρεθούν σε μία εκ των τριών πιθανών καταστάσεων: υπερδιέγερσης (+1) που σημαίνει αύξηση στην συγκέντρωση των φωσφοπρωτεϊνών ή στην έκφραση των γονιδίων, υποδιέγερσης (-1) που είναι η αντίστροφη κατάσταση, ή στην βασική κατάσταση (0) που υποδηλώνει καμία αλλαγή στην κατάσταση του κόμβου ως αποτέλεσμα της σηματοδότησης. Η επιρροή που ασκεί ένας μητρικός κόμβος στους θυγατρικούς του εξαρτάται από το πρόσημο της συνδέουσας ακμής  $\sigma(a)$ .

Για να είναι *συνεπείς* δύο κόμβοι που συνδέονται με ακμή πρέπει να βρίσκονται στην ίδια κατάσταση αν η ακμή είναι θετικά προσημασμένη και αντίθετα αν είναι αρνητικά. Για να είναι ένας γράφος *ασθενώς συνεπής*, τότε πρέπει κάθε κόμβος να είναι συνεπής με *τουλάχιστον έναν* από τους μητρικούς του κόμβους. Ένας συνεπής γράφος ερμηνεύει ποιοτικά τη διαδικασία διάδοσης σήματος καθώς για κάθε κόμβο υπάρχει ένα μονοπάτι αιτιακών σχέσεων που εξηγεί την κατάσταση του με βάση κάποιο αρχικό ερέθισμα. Επομένως, στόχος και των 2 προτεινόμενων μεθόδων είναι να βρουν το συνεπή γράφο ο οποίος ερμηνεύει καλύτερα τα πειραματικά αποτελέσματα.

Η μέθοδος SCM διατυπώνει το πρόβλημα της ανεύρεσης του βέλτιστου συνεπή γράφου ως πρόβλημα ακέρατου γραμμικού προγραμματισμού. Συγκεκριμένα, ψάχνει όλα τα υποσύνολα του PKN για να βρει αυτό το οποίο μπορεί να προσομοιώσει βέλτιστα τις καταστάσεις των διαφόρων κόμβων όπως αυτές μετρήθηκαν πειραματικά. Οι περιορισμοί του προβλήματος κωδικοποιούν το μοντέλο συνεπών επιρροών και εξασφαλίζουν ότι ο γράφος μπορεί να σημανθεί με τρόπο συνεπή ενώ η αντικειμενική συνάρτηση κωδικοποιεί την απόσταση μεταξύ της

προσομοιωμένης και της πειραματικά προσδιορισμένης κατάστασης των κόμβων που μετρήθηκαν.

Σε προηγούμενες διατυπώσεις του προβλήματος, η αναζήτηση για το βέλτιστο υπο-δίκτυο γινόταν μόνο επί του συνόλου των ακμών. Στη διατύπωση που αναπτύχθηκε για αυτή τη διατριβή, η αναζήτηση γίνεται και επί του συνόλου των κόμβων. Η αλλαγή αυτή δεν επηρεάζει το σύνολο των λειτουργικά ισοδύναμων λύσεων. Τα βασικά οφέλη που αποφέρει είναι ότι πρώτον, επιτρέπει στον ερευνητή μεγαλύτερη ελευθερία να ενσωματώσει την πρότερη γνώση του στο σύστημα κατευθύνοντας τη λύση, και δεύτερον, καθιστά τη διαδικασία της βελτιστοποίησης πιο ρεαλιστική καθώς υπάρχουν φυσικά γεγονότα που μπορούν να οδηγήσουν στην πλήρη αδρανοποίηση μίας πρωτεΐνης ή γονιδίου.

Ένας άλλος σημαντικός περιορισμός αυτής της μεθόδου ήταν η αδυναμία της να διερευνήσει δίκτυα που είχαν βρόχους ανατροφοδότησης (feedback loops), διότι βρόχοι μπορούν να μεταβούν σε μία μη βασική κατάσταση χωρίς να λάβουν σήμα από κάποιο ερέθισμα από το υπόλοιπο δίκτυο. Στη διατύπωση που αναπτύχθηκε για αυτή τη διατριβή, επιπλέον περιορισμοί έχουν εισαχθεί οι οποίοι αποτρέπουν αυτή την παρενέργεια.

Η δεύτερη μέθοδος που αναπτύχθηκε RCR λύνει το πρόβλημα της ανεύρεσης του βέλτιστου συνεπή γράφου για την περίπτωση που το αρχικό ερέθισμα είναι άγνωστο. Για την μέθοδο SCM, απαιτούνται πειράματα κατά τα οποία τα κύτταρα διεγείρονται τεχνητά, συνήθως με κυτοκίνες, και εν συνεχεία μετράται η απόκρισή τους, ως αλλαγή στην συγκέντρωση κάποιων φωσφοπρωτεϊνών ή την έκφραση κάποιων γονιδίων. Παρ'όλα αυτά, πολλές φορές η γενεσιουργός αιτία της κυτταρικής απόκρισης δεν είναι γνωστή αλλά η "απόκριση" μπορεί να μετρηθεί σαν διαφορά μεταξύ ενός πληθυσμού ελέγχου και ενός "αποκρίνοντος", για παράδειγμα μεταξύ υγιών και καρκινικών κυττάρων. Η μέθοδος RCR γενικεύει περαιτέρω την μέθοδο SCM ώστε να μπορεί να μοντελοποιήσει και αυτά τα πειράματα. Συγκεκριμένα, η αρχική διέγερση, η οποία θεωρείτο γνωστή προηγουμένως, μετατρέπεται σε μεταβλητή η οποία συμμετέχει στην διαδικασία εύρεσης του βέλτιστου γράφου. Επειδή αυτή η αλλαγή αυξάνει πολύ τους βαθμούς ελευθερίας του προβλήματος, απαιτείται η χρήση επιπλέον περιορισμών για να διασφαλιστεί ότι οι λύσεις δεν θα είναι τετριμμένες. Οι περιορισμοί αυτοί έχουν στόχο να περιορίσουν τον αριθμό των πιθανών "αιτιών" και το μέγεθος του δικτύου και μπορούν να επιβληθούν ως αυστηρά όρια ή/και ως επιπλέον κόστη στη αντικειμενική συνάρτηση.

## Εφαρμογές

Στη συνέχεια περιγράφονται 3 εφαρμογές στις οποίες έγινε χρήση των μεθόδων που περιγράφηκαν για την ανίχνευση σηματοδοτικών δικτύων. Τα δίκτυα αυτά χρησιμοποιήθηκαν στη συνέχεια για να διερευνηθούν οι σημαντικότεροι συντελεστές της κυτταρικής απόκρισης ή για να προταθούν πιθανές φαρμακευτικές αγωγές που θα επαναφέρουν τα κύτταρα σε υγιή κατάσταση.

Η πρώτη εφαρμογή που περιγράφεται αφορά τη διαδικασία της κυτταρικής γήρανσης. Για αυτή την εφαρμογή, ως βιολογικό μοντέλο χρησιμοποιήθηκαν ανθρώπινοι πρωτογενείς ινοβλάστες. Μετρήθηκε η συγκέντρωση 18 φωσφοπρωτεϊνών σε νεαρά και γηρασμένα κύτταρα κατόπιν διέγερσης των κυττάρων με 6 διαφορετικές κυτοκίνες. Τα νεαρά κύτταρα διπλασιάζονταν κάθε μέρα ενώ τα γηρασμένα δεν είχαν διπλασιαστεί για ένα μήνα. Για τις φωσφοπρωτεΐνες αυτές, προηγούμενες έρευνες είχαν δείξει ότι συμμετέχουν στη γήρανση και στον πολλαπλασιασμό των κυττάρων. Η κατάσταση κάθε πρωτεΐνης προσδιορίστηκε κατόπιν σύγκρισης της συγκέντρωσής της με την αντίστοιχη κυττάρων που δεν είχαν διεγερθεί με κάποιο ερέθισμα. Σε συνδυασμό με το PKN από τη βάση αντιδράσεων MetaCore χρησιμοποιήθηκε μια παραλλαγή της μεθόδου SCM για να ανιχνευτεί το σηματοδοτικό δίκτυο των 2 κυτταρικών πληθυσμών. Η παραλλαγή συνίστατο στην σύζευξη των αντιδράσεων των 2 δικτύων. Συγκεκριμένα, θεωρήθηκε ότι οι διαφοροποιήσεις κατά τη γήρανση οφείλονται σε μείωση της έκφρασης των γονιδίων οπότε το σύνολο των αντιδράσεων θα έπρεπε να είναι κοινό και τα δίκτυα να διαφέρουν μόνο ως προς το σύνολο των κόμβων. Οι διαφορές των δύο δικτύων επαλήθευσαν προηγούμενες μελέτες που κατέδειξαν υποτονική απόκριση στην παρουσία ινσουλίνης και ισχυρή ενεργοποίηση του NFκB και των συναφών λειτουργιών κατά την διέγερση με IL1A και TNF. Επιπλέον, η δομή των δικτύων είχε διαφορές που δεν είχαν αναφερθεί στη σχετική βιβλιογραφία όπως ο κεντρικός ρόλος που παίζει η SRC, η οποία φαίνεται να είναι ανενεργή στα γηρασμένα κύτταρα, στη ρύθμιση πολλών πρωτεϊνών που σχετίζονται με τον πολλαπλασιασμό των κυττάρων.

Για τη δεύτερη εφαρμογή που περιγράφεται χρησιμοποιήθηκε η μέθοδος RCR για να αναλυθεί ο “μηχανισμός λειτουργίας” της φαρμακολογικής τοξικότητας στον πνεύμονα και να προταθούν πιθανές θεραπείες. Συγκεκριμένα, επιλέχθηκαν 200 φάρμακα από τη βάση δεδομένων Pneumotox τα οποία έχουν τη δυνατότητα να προκαλέσουν ασθένεια του πνεύμονα. Στη συνέχεια οι πιθανοί στόχοι αυτών των φαρμάκων ανακτήθηκαν από την βάση δεδομένων STICH. Τέλος, χρησιμοποιήθηκε η βάση CMap, η οποία περιγράφει διαφορές στην γονιδιακή έκφραση 1000 γονιδίων σε 5 κυτταροσειρές πριν και μετά την επώασή τους με αυτά τα φάρμακα. Με βάση αυτές τις πληροφορίες και σε συνδυασμό με ένα PKN από τη βάση Reactome, δημιουργήθηκε ένα σηματοδοτικό δίκτυο για κάθε φάρμακο το οποίο ερμηνεύει την παρατηρούμενη αλλαγή στην γονιδιακή έκφραση. Επειδή και τα 200 φάρμακα μπορούν να προκαλέσουν ασθένεια του πνεύμονα, αναλύθηκε η τομή των 200 δικτύων και ανιχνεύτηκαν 2 υπο-δίκτυα τα οποία ήταν συστηματικά υπέρ ή υπό διεγερμένα. Για την επαλήθευση των αποτελεσμάτων, χρησιμοποιήθηκε ο αλγόριθμος GUIDE για να κατασκευάσει έναν ανεξάρτητο δίκτυο με βάση τα ίδια δεδομένα και παρατηρήθηκε ότι τα 2 αποτελέσματα είχαν στατιστικά σημαντική αλληλοκάλυψη. Στη συνέχεια, με παρόμοιο τρόπο κατασκευάστηκαν σηματοδοτικά δίκτυα και για τα υπόλοιπα, μη-τοξικά, φάρμακα που ήταν κοινά στις STICH και CMap έτσι ώστε να βρεθούν φάρμακα που να επάγουν την αντίστροφη κατάσταση στα υπο-δίκτυα που αναφέρθηκαν. Τα κα-

λύτερα 40 προτεινόμενα φάρμακα, ελέγχθηκαν από φαρμακολόγο η οποία παρατήρησε ότι τα περισσότερα έχουν τη δυνατότητα να βελτιώσουν την κατάσταση κυρίως λόγω της αντιφλεγμονώδους δράσης τους.

Για την τρίτη εφαρμογή, ακολουθήθηκε η ίδια διαδικασία όπως και στη δεύτερη αλλά αυτή τη φορά για να προταθούν φάρμακα για την αντιμετώπιση της μόλυνσης από το βακτήριο του άνθρακα. Χρησιμοποιήθηκαν τα ίδια μη-τοξικά φάρμακα όπως και στη δεύτερη εφαρμογή. Για την μελέτη των συνεπειών της μόλυνσης από άνθρακα στην γονιδιακή έκφραση χρησιμοποιήθηκαν δεδομένα από 4 έρευνες από την βάση GEO. Αφού κατασκευάστηκε το σηματοδοτικό δίκτυο που μοντελοποιεί την κυτταρική απόκριση στις τοξίνες που εκκρίνει το βακτήριο, με χρήση της μεθόδου RCR, αναζητήθηκαν φάρμακα που να αντιστρέφουν την απόκριση αυτή. Η αναζήτηση έγινε συγκρίνοντας τα δίκτυα των φαρμάκων και του άνθρακα, και τα φάρμακα ταξινομήθηκαν από το καλύτερο στο χειρότερο ανάλογα με την απόστασή τους από την πλήρη αντιστροφή του δικτύου του άνθρακα. Για επαλήθευση έγινε και πάλι βιβλιογραφική έρευνα για τα φάρμακα που ήταν στη κορυφή και στον πάτο της κατάταξης καταλληλότητας. Και για τις δύο κατηγορίες τα αποτελέσματα κρίθηκαν λογικά, καθώς πολλά από τα φάρμακα που ήταν ψηλά στην λίστα είχαν ήδη προταθεί και από άλλες μελέτες, ενώ αντίθετα φάρμακα που ήταν χαμηλά στη λίστα είχαν παρόμοιους στόχους με αυτούς του βακτηρίου οπότε αναμένεται να είχαν παρόμοιες συνέπειες στα κύτταρα.



# Contents

Πρόλογος . . . . .	v
Περίληψη . . . . .	vii
Εκτεταμένη Περίληψη . . . . .	ix
<b>1 Biological Background</b>	<b>1</b>
1.1 Systems Biology . . . . .	1
1.2 Signaling Networks . . . . .	2
1.2.1 Protein Phosphorylation . . . . .	3
1.2.2 Gene Expression . . . . .	5
1.2.3 Protein-Protein Interactions . . . . .	5
1.2.4 Protein-Gene Interactions . . . . .	6
1.3 Data Bases . . . . .	7
<b>2 Mathematical Modeling</b>	<b>9</b>
2.1 Literature Review . . . . .	9
2.1.1 Logic Models . . . . .	11
2.2 Contributions . . . . .	13
2.2.1 Sign Consistency Method . . . . .	13
2.2.2 Reverse Causal Reasoning Method . . . . .	18
2.3 Expansion and Future Work . . . . .	20
2.3.1 Probabilistic Networks . . . . .	20
2.3.2 Hybrid & Dynamic Networks . . . . .	21
<b>3 Replicative Senescence</b>	<b>23</b>
3.1 Introduction . . . . .	23
3.2 Materials and methods . . . . .	24
3.2.1 Experiments . . . . .	24
3.2.2 Network Construction . . . . .	25
3.3 Results . . . . .	25
3.3.1 Sensitivity Analysis . . . . .	28
3.4 Discussion . . . . .	29
<b>4 Drug Induced Lung Disease</b>	<b>33</b>
4.1 Introduction . . . . .	33
4.2 Analysis . . . . .	35

4.2.1	Workflow . . . . .	35
4.2.2	Lung-toxic Drugs . . . . .	37
4.2.3	Identification of drug's MoA . . . . .	39
4.2.4	Construction of the DILD network . . . . .	41
4.2.5	Drug Candidates for DILD . . . . .	44
4.2.6	Validations . . . . .	46
4.3	Conclusions . . . . .	46
4.4	Methods . . . . .	48
4.4.1	ILP formulation . . . . .	48
4.4.2	Construction of Prior Knowledge Network . . . . .	48
<b>5</b>	<b>Anthrax Infection</b>	<b>51</b>
5.1	Introduction . . . . .	51
5.2	Analysis . . . . .	54
5.2.1	Anthrax Functional Network . . . . .	54
5.2.2	Repurposing Candidates . . . . .	55
5.3	Validation . . . . .	55
5.4	Discussion . . . . .	57
<b>A</b>	<b>Replicative Senescence</b>	<b>63</b>
A.1	Data Description . . . . .	63
A.2	Quality Control . . . . .	64
A.3	Data Normalization . . . . .	67
<b>B</b>	<b>Drug Induced Lung Disease</b>	<b>71</b>
B.1	Top Drug Candidates . . . . .	71
B.2	Enrichment Tables . . . . .	74

# List of Figures

1.1	Schematic representation of signaling via tyrosine kinase receptor as it was known in 1990 and 2010. . . . .	3
1.2	Protein Measuring Technologies . . . . .	4
1.3	Schematic Representation of Y2H technique . . . . .	6
2.1	Example of combinatorial perturbation experiment . . . . .	10
2.2	Arc Decomposition . . . . .	16
2.3	Illustrative example of the Reverse Causal Reasoning Method . . . .	19
3.1	Cell Response to different stimuli as they age . . . . .	26
3.2	Signaling network of young and senescent cells . . . . .	27
3.3	Predictions of cell response . . . . .	28
3.4	Sensitivity analysis of aging networks . . . . .	29
4.1	Drug induced lung disease workflow . . . . .	36
4.2	Identification of Drug's Mode of Action . . . . .	37
4.3	Imatinib mode of action . . . . .	40
4.4	Drug Induced Lung Disease network & analytics . . . . .	42
4.5	Consistently up-regulated module of the DILD network . . . . .	43
4.6	Consistently down-regulated module of the DILD network . . . . .	49
4.7	Ciclosporin mode of action . . . . .	50
5.1	Workflow: Drug reposition for Anthrax infection . . . . .	53
5.2	Anthrax functional network . . . . .	54
A.1	Aging raw data overview . . . . .	65
A.2	Aging QC: Bead count per analyte . . . . .	66
A.3	Aging QC: Positive Control . . . . .	66
A.4	Aging QC: Negative Control . . . . .	67
A.5	Aging: Summary of fold change per analyte . . . . .	68
A.6	Aging: Fold change per analyte and stimulus . . . . .	69
A.7	Aging: Overview of the activity of every analyte . . . . .	70



# List of Tables

5.1	Top 10 drug candidates for Anthrax infection . . . . .	60
5.2	Bottom 10 drug candidates for Anthrax infection . . . . .	61
A.1	Aging data sample . . . . .	63
B.1	GO terms enriched in up-regulated DILD nodes . . . . .	74
B.2	GO terms enriched in down-regulated DILD nodes . . . . .	75
B.3	Pharmacological effects of Pneumotox Drugs . . . . .	76



# Chapter 1

## Biological Background

This chapter is a non-technical summary of the biological concepts and techniques that are discussed in the following chapters. Its main goal is to present the ideas of Systems Biology and how they are implemented in practice. It also offers a quick overview of the experimental data and techniques utilised for this thesis.

This chapter is not a full and detailed description of the subjects it discusses. Tomes of Biology, which the author does not claim to have read or understand fully, are summarized in a few paragraphs! Biology researchers may find this chapter useful in order to understand the abstractions that the author is assuming. Readers without a biology training but interested in learning are referred to the cited literature or the molecular biologist of their area.

### 1.1 Systems Biology

The term “Systems Biology” was initially coined by Hiroaki Kitano [1], [2] to describe a holistic model for biological research. Contrary to “traditional” molecular biology, this new model does not focus on characterizing individual molecular entities, like genes and proteins, but instead on how these entities interact to form a biological system<sup>1</sup>, whose functional characteristics cannot be attributed to any particular entity but are emergent. Systems Biology is not a new idea [3], but before the advent of modern molecular technologies it was not possible to gather the required mass and granularity of data required for this type of research. The major challenge in the road for a systems level understanding of biology is that it requires insights from a diverse range of disciplines like biology [4], mathematical modeling [5], control theory [6] as well as specialized knowledge for the application at hand, like pharmacology [7], [8].

Kitano, inspired by control theory, indicated 4 requirements and/or research directions towards a system level understanding of biology:

---

<sup>1</sup>The definition of “system” depends on the scale of the phenomena studied. For example, an organism is a system of tissues and each tissue is a system of cells etc.

1. identification of system structure
2. determination of system dynamic properties
3. develop methods to control them
4. develop techniques to design or alter systems

System structure comprises of interaction networks between the molecular entities involved as well as a high-level description of how these networks control the system's properties and behavior.

This thesis focuses on the cell as the system under study and develops methods for the identification of its structure. The cell is modelled as a complex system inside an unknown biochemical environment. It receives signals from the environment which it processes and responds by changing its state<sup>2</sup>. As a first approximation, the cell is composed of proteins and genes. Proteins are macromolecules whose shape and concentration controls the macroscopic properties of the cell, while genes are pieces of genetic material that determine the shape of the proteins. Consequently, the state of the cell is described by the concentration and shape of the proteins and genes that comprise it.

## 1.2 Signaling Networks

The processing of environmental information is carried out by the signaling network of the cell, i.e. a series of biochemical interactions between the proteins inside it. In more detail, the cell communicates with its environment either via small molecules that can permeate its membrane or through specialized proteins, called receptors, that protrude from the membrane in both directions. In both cases, the information is recorded as a change in the shape of the proteins that came in touch with the environment, which constitutes a partial differentiation of the cell's state. This change affects other proteins in proximity and thus propagates inside the cell affecting proteins and genes until the system reaches a new steady state. This process constitutes the cell's response to the environmental change<sup>3</sup>.

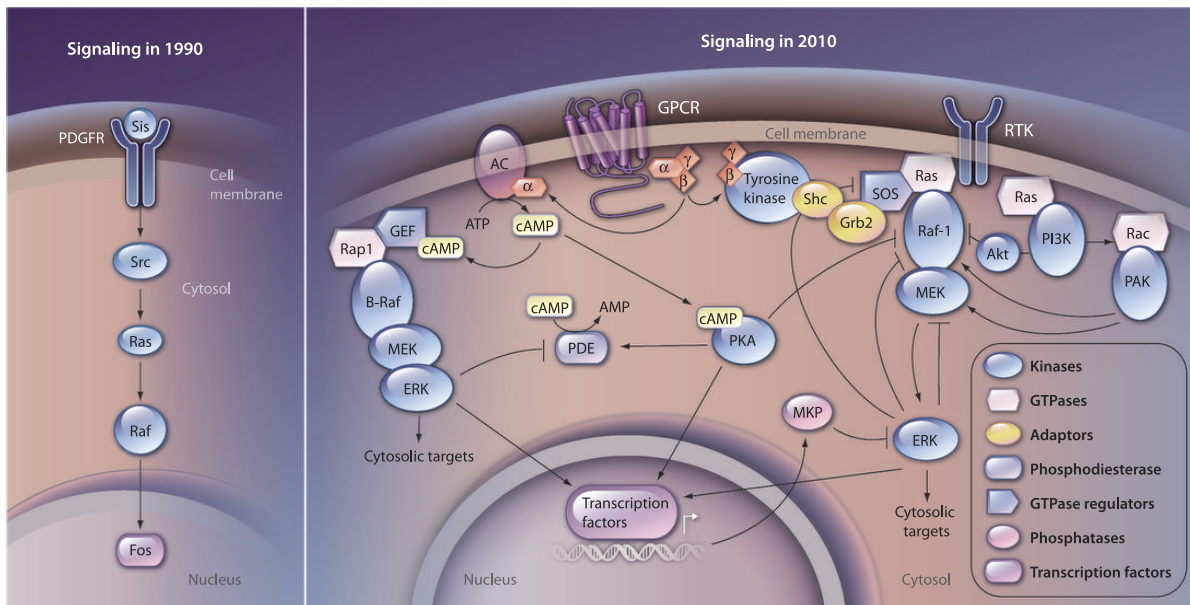
A reductionist approach to cell response focuses on identifying signaling pathways controlling an elementary function of the cell. The term "signaling pathway" describes a roughly linear cascade of interactions between a small number of proteins. A Systems Biology approach, focusing on the *networks* that control more complex functions, became attainable only during the last decade (Fig. 1.1) through the emergence of modern high-throughput technologies that can measure multiple cellular components (multi-omics) in an unbiased way. The following sections describe the technologies used for the applications of this thesis.

---

<sup>2</sup>This model is evidently naive since it overlooks the fact that cells typically live in colonies thus their state and the state of their environment is coupled to a degree and feedback interactions occur.

<sup>3</sup>It is interesting to note that information is "processed" and "stored" in the same medium, namely the network of proteins inside the cell. For a more in depth treatment of this idea [see 9]





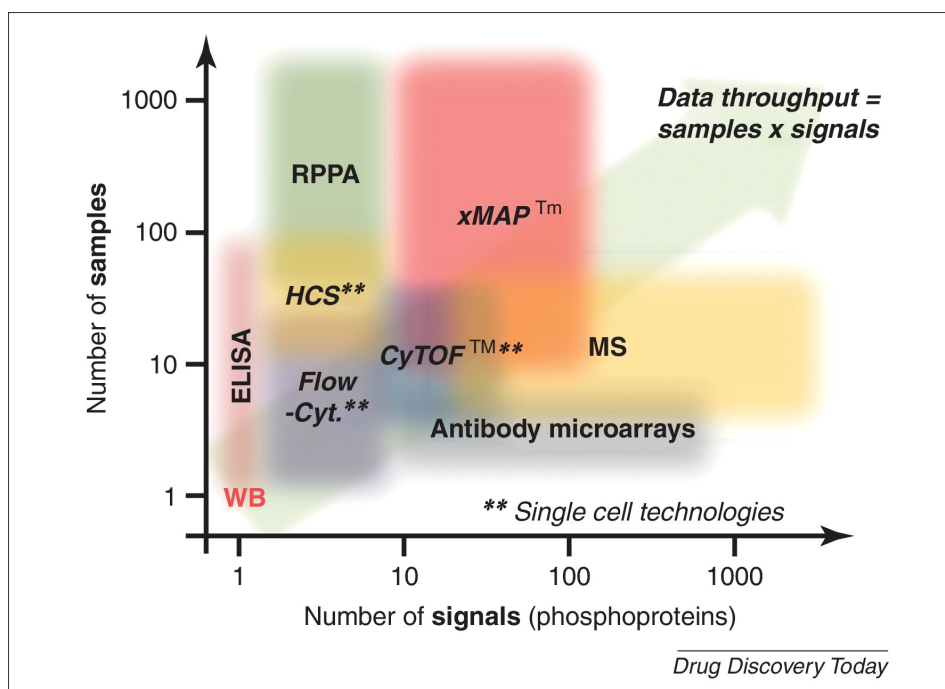
**Figure 1.1: Schematic representation of signaling via tyrosine kinase receptor as it was known in 1990 and 2010.** Reprinted from [10]

### 1.2.1 Protein Phosphorylation

The term “phosphorylation” refers to the process of appending one or more phosphate groups ( $\text{PO}_4^{3-}$ ) to a protein’s structure. Phosphorylation is catalyzed by enzymes called kinases. The reverse process is called dephosphorylation and is catalyzed by enzymes called phosphatases. Because of the strong negative charge of a phosphate group, (de)phosphorylation has significant consequences on a protein’s shape and consequently its function and ability to interact with other proteins [11, Chapter 3]. It’s estimated that between 1/3 and 2/3 of all proteins can be phosphorylated. The scale and intensity of its effects makes phosphorylation one of the most important modifications that a protein can undergo during the signaling process, especially for eukaryotic cells. Moreover, disruptions in the phosphorylation process have been linked with several pathologies [12] and many drugs intervening in the processes by targeting kinases or phosphatases have been approved, especially for targeted cancer therapies (e.g. Sorafenib, Crizotinib, Erlotinib)

Technologies to measure phosphorylated proteins can be grouped into two main categories: MS-based and tag-based. The two categories span the trade-off between high-throughput and multiplexability 1.2. MS-based technologies can measure every protein in a sample but their ability to measure multiple samples is hindered by the complexity of the protocol and post-processing analysis. On the other hand, tag-based technologies use antibodies or aptamers<sup>4</sup> to measure a pre-specified set of proteins, but they can measure them across multiple samples in parallel. For more details about protein measuring technologies [see 13].

<sup>4</sup>Antibodies are proteins produced by the immune system of an organism in order to attach



**Figure 1.2: Protein Measuring Technologies** Reprinted from [13]. Schematic representation of the operating range of the major protein measuring technologies. Abbreviations not discussed in the text: Western Blot (WB), Enzyme-Linked Immunosorbent Assay (ELISA), Reverse Phase Protein Array (RPPA), High-Content Screening (HCS), fluorescence-based Flow Cytometry (Flow-Cyt.), Mass Spectrometry (MS), Mass Cytometry (CyTOF)

For the applications presented in this thesis the xMAP™ technology was utilized. xMAP is tag-based technology based on antibodies. In particular, it uses two sets of antibodies per protein. One set is coupled with a fluorescent dye (phycoerythrin) to render the proteins detectable while the other set is immobilized on color-coded, magnetic micro-beads which are used to discriminate the proteins and pull them down from the sample. At the moment of writing, the theoretical upper limit to the number of proteins that can be measured in parallel is 500, as many as the different bead colors, but in practice this number is limited by the quality of available antibodies. Antibodies that bind to multiple proteins (low specificity) can affect their measurement by other antibodies and increase the noise. Finally, xMAP can distinguish between phosphorylated and unphosphorylated proteins by using antibodies specific to either form of the same protein.

---

and identify to pathogenic molecules (antigens). Aptamers are small molecular chains design to recognize other molecules with a complementary structure.

## 1.2.2 Gene Expression

The genetic material (DNA) of a cell resides in its nucleus and it contains the information necessary for the creation of the proteins that make up cell behavior. This information is stored in segments of the DNA called genes<sup>5</sup>. The translation of genes to proteins is an elaborate process that involves transcription of the information of the gene onto RNA, a part of the genetic material that can wander inside the cell. A gene is “expressed” when it is transcribed to RNA in order to produce the encoded protein. The intensity of the gene expression is defined to be the number of RNA copies produced which is proportional to the production of the protein [14].

Gene expression measuring technologies present the same trade-offs between high-throughput and multiplexability as their protein counterparts. Here, sequencing technologies allow the unbiased measurement of every RNA fragment in the sample, at the cost of fewer samples, while DNA microarrays can quantify a pre-specified set of genes in more samples. Measuring protein concentration is also an indirect way of measuring gene expression, but direct measurement technologies are more mature and genes are more stable entities and easier to amplify than proteins and thus easier to quantify.

For the application presented in this thesis, publicly available data from microarray experiments were utilized. The working principle of this technology is the same with that of tag-based technologies for protein quantification. Instead of antibodies or aptamers, oligonucleotides, small sequences of DNA code, are used to immobilize their complementary sequence on a substrate and then a fluorescent dye is used to quantify the captured fragments. In fact, the xMAP technology can be used to measure gene expression, although in this case the limit of 500 bead colors is more important.

## 1.2.3 Protein-Protein Interactions

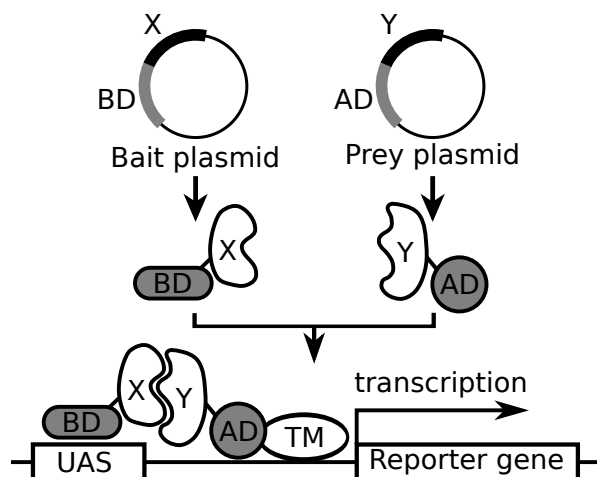
The three most important technologies for the identification of protein-protein interactions (PPIs) in large scale are: yeast-two-hybrid (Y2H) [15] and the two mass spectrometry (MS) based techniques affinity purification (AP-MS) [16] and co-fractionation (CoFrac-MS) [17].

The Y2H technology is used to screen for binary protein-protein interactions. Its working principle (see Figure 1.3) is to link the presence or absence of an interaction with the expression of a reporter gene. In particular, the two interrogated proteins are slightly modified (hybrids) by being linked to the two domains of a transcription factor. One of them (Bait) is linked with the DNA binding domain and the other (Prey) with the activation domain. If the two proteins interact then the transcription factor is assembled and can activate the gene expression process.

---

<sup>5</sup>All the proteins are encoded by genes, but not all genes encode proteins. Moreover, there are post-translational modifications that further affect the final form of the protein

**Figure 1.3: Schematic Representation of Y2H technique.** Adapted from [18, fig. 3.21]. X and Y are the interrogated proteins. Plasmids are used to modify the proteins and append the binding domain (BD) and the activation domain (AD). Abbreviations: UAS: Upstream Activating Sequence, TM: Transcription Machinery.



Traditionally, the expression of the reported gene was linked to the survival of a yeast colony, but nowadays other technologies can be used. An important limitation of this method is that the interactions have to take place in yeast and not the host organism of interest. This is because yeast survival is used as a marker for the interaction.

On the other hand, MS-based technologies can detect interaction inside any cell without the need to modify them. For AP-MS proteins are purified<sup>6</sup> from the cell and tagged. Then they are mixed with other purified proteins to form protein complexes which are then passed through an MS to be identified. For CoFrac-MS, protein complexes are extracted directly from the cells through extensive fractionation with different biochemical techniques and then passed through an MS to identify their components. The main limitation of these methods is that not all pairs of proteins in a complex interact and there is no way of discriminating between protein interactions and protein associations.

Over the last decade, these techniques have lead to a dramatic increase of the number of known or suspected interactions. From 5000 in 2005 [19] to 14000 in 2014 [20] using Y2H and more than 20000 in 2015 using AP-MS [21], [22].

This increase in size has raised concerns because the resulting datasets are characterized by low repeatability that some researchers attribute to high rates of false positive calls [23]–[25], while other argue that is due to low sensitivity and complementarity of the methods [26]. Common countermeasures for this problem include running validation experiments, usually with low-throughput technologies of higher accuracy, and/or manual curation of the results by experts.

## 1.2.4 Protein-Gene Interactions

As explained in Section 1.2.2, genes are “expressed” by being transcribed to RNA. This process is carried out by a specialized protein complex called *RNA polymerase* which controls transcription. However, RNA polymerase by itself cannot bind to the DNA molecule to initiate the process. Other proteins, collectively known as

*transcription factors*, help initiate the process by binding to the DNA and providing a platform for RNA polymerase. The term transcription factor is even broader than that and includes every protein that can regulate gene expression either by promoting or suppressing it. Transcription factors are part of the signaling process and the most common form of gene control.

Many technologies have been developed to identify transcription factors and their regulating genes. The most prominent one is ChIP-seq [27] and its variants like ChIP-PET [28], ChIP-chip [29], and DamID [14]. The operating principle of all these technologies is to “freeze” the DNA at a specific time point, then fragment it and then use antibodies to pull down the proteins attached to it along with the DNA fragments they were attached to. Proteins are identified by the antibodies used while the DNA fragments are sequenced (for ChIP-seq) and aligned.

## 1.3 Data Bases

The following databases were utilised for the work presented in this thesis:

- Protein interactions
  - **Metacore** private database. Entries are manually selected and annotated by experts.
  - **Reactome** [30] public database compiled from open access sources. Entries are checked by experts.
- Transcription factors
  - **ChEA** [31] open access list of protein genes interactions compiled from literature.
  - **Transfac** [32] private database. Partial access is granted free of charge for academic users
  - **Jaspar** [33] public database of protein gene interactions across multiple organisms.
- Drugs
  - **Pneumotox** [34] list of 892 drugs and chemical compounds associated with drug-induced lung disease
- Gene expression
  - **Connectivity Map** [35] pilot study measuring gene expression for 5 cell-lines upon perturbation with drugs and other chemical compounds.
  - Gene Expression Omnibus (**GEO**) [36] public database where everyone can upload primary and processed data from his/her study.



# Chapter 2

## Mathematical Modeling

In Chapter 1, the main components of the cell's signaling system, as well as the technologies shedding light upon them, were described. This chapter focuses on combining the wealth of generated data into models that can "explain" and predict cell behavior. The presentation begins with a short review of the relevant literature and culminates with the description of the two models formulated by the author. The goal of the review is not to provide an unbiased and complete description of the literature but to illustrate the trajectory of the ideas that morphed into the models presented in this thesis. In the following chapters, 3 real world applications of these models are presented.

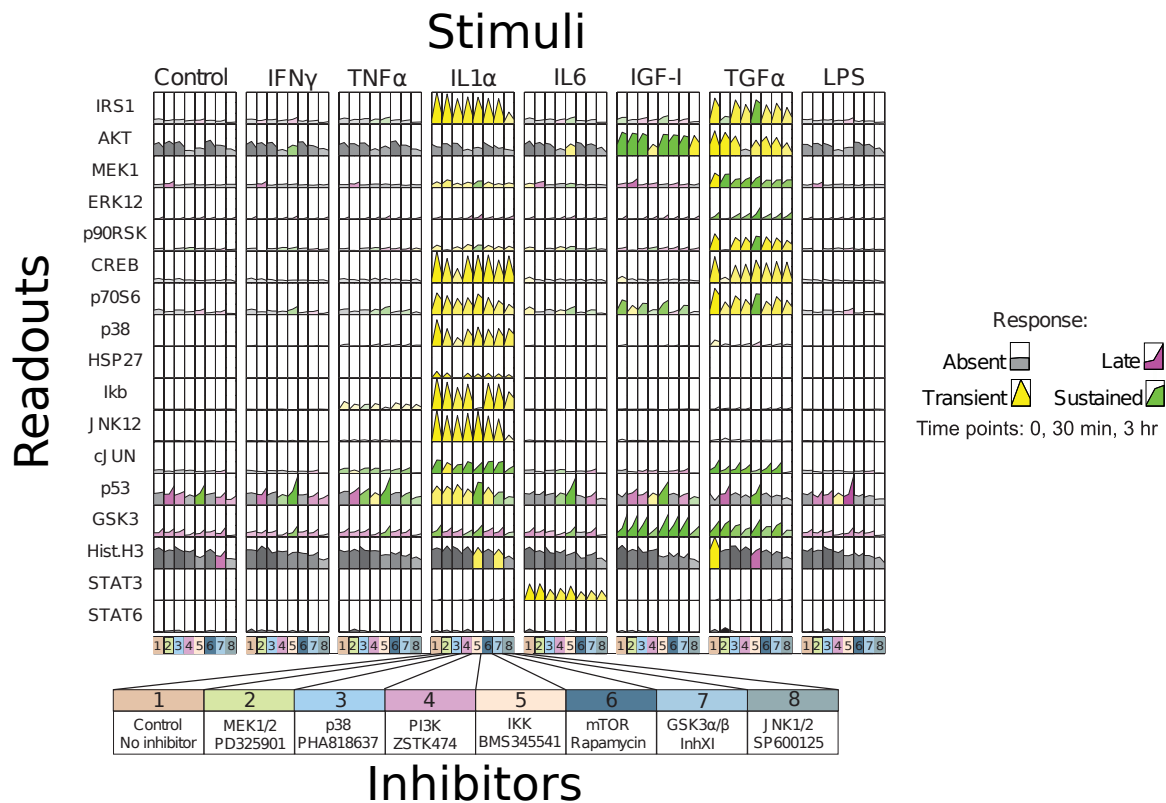
### 2.1 Literature Review

Mathematical models serve a dual role with respect to signaling networks. On the one hand, they explain cell behavior in terms of how the different components interact to give rise to it. In this regard, they can be used to make prediction about future responses to changes in the environment or suggest possible interventions when the signaling process is deregulated. On the other hand, they act as a noise filter. As explained in Chapter 1, the technologies used to probe the cell can be prone to high rates of false positive and/or negative errors. Fitting a model, that is conceptually consistent, to data, that may not be, can act as a noise filter, provided that Nature is consistent with the model's expectations of her.

Models are of particular importance in deciphering the network structure of the signaling process. As described in Section 1.2.3, all the major technologies for identifying interactions have accuracy issues. But, even if the accuracy was increased, signaling networks would still be context-dependent and no static/generic list of interactions would suffice to describe them.

As a result, many algorithms have been proposed for deciphering the network structure by adapting the set of known interactions (interactome), to data generated by perturbation experiments in order to produce a network specific to the cells under investigation [10]. Perturbation experiments are designed to highlight the

dependencies between the nodes of a network. Their design consists of perturbing the cells, with cytokines<sup>1</sup> and/or other biochemical stimuli, and measuring the response of some key nodes of the signaling network (example Figure 2.1). The algorithms then proceed to reconstruct the connectivity of the signaling network to satisfy the observed dependencies using the interactome to guide them through and narrow down the search space.



**Figure 2.1: Example of combinatorial perturbation experiment** Adapted from [37, Fig. 2]. Cells were perturbed with a combination of cytokines and kinase inhibitors, and the cells' response in terms of 17 proteins was measured. Rows represent measurements of 17 intracellular proteins (readouts) assayed at  $t = 0, 30\text{min},$  and  $3\text{h}$  (relative to stimulus addition), and columns represent different stimuli. For each combination of stimulus and readout, one of seven inhibitors was applied as indicated in the schematic below the data. Data are coded to highlight no induction (relative to basal activity; gray), transient induction (peaking at 30min; yellow), sustained induction (equal at 30min and 3h; green), or late induction (peaking at 3h; purple).

The first step towards developing a model for the signaling process is selecting the appropriate formalism. A major factor contributing to this decision is the granularity of the available data and of the desired mechanistic description.

<sup>1</sup>Cytokines are small proteins secreted by cells in order to communicate with their environment. They bind to cell receptors and initiate the signaling process.



For this work, two logic-based models were developed together with appropriate optimization methods for fitting them (Section 2.2). Logic models occupy a middle ground with the respect to the granularity of mechanistic description, between the naturalistic ODE models and pure correlation based models [10, Fig. 2]. Although logic models can describe some aspects of the dynamic behavior of the network, the models presented in this thesis aim for a qualitative description of the static structure of the network. To do so, they use phosphoproteomic or gene expression data to refine the topology of a prior knowledge network assembled from generic interactions available in databases (see Section 1.3). Both models are based on the same signal transduction mechanism and, when fitted, yield a network of protein-protein and protein-gene interactions that best describes the observed cell response. In the applications presented in the following chapter, these networks are then analysed and compared in order to identify key components characterizing cell behavior.

### 2.1.1 Logic Models

Logic networks have been one of the first formalisms used to model cell systems even before the rise of modern Systems Biology [38]. As an abstraction they have been appealing because they are easy to compose and analyse compared to the more realistic system of differential equations. They provide a qualitative description of the system and as a result have less free parameters to determine than their quantitative counter-parts and thus they are robust to noise and do not require large volumes of data [39]. However, reconstructing Boolean networks from experiments on a proteome scale is a challenging discrete optimization problem and has only been pursued over the last decade.

One of the first studies on Boolean networks was by Saez-Rodriguez et al. [40]. There the authors started with a directed graph, which they used to model the state of every node as an unknown Boolean function of its predecessors. In order to identify the functions, they used perturbation experiments and a genetic algorithm (GA) to search the space of all Boolean functions using AND and OR gates. NOT gates, necessary for modeling inhibitory effects, have to be pre-specified but the user. Upon convergence, the connectivity of the network was computed by connecting nodes that were linked by a Boolean function. In the same year, Mitsos et al. [41] proposed an Integer Linear Programming (ILP) formulation for the same problem. The main difference between the two formulations, is that the ILP formulation does not search over all possible Boolean function but instead selects from a list of pre-specified candidates. However, it is significantly faster than the GA approach, due to its non-random search strategy, and can provide optimality guarantees. Both approaches also included a form of (soft) sparsity constraints to penalize large networks. Later, Sharan and Karp [42] suggested an ILP formulation that can learn any Boolean function between a node and its predecessors. Other approaches used to tackle the same problem include Constraint Based Modeling

(CBM) [43] and Answer Set Programming (ASP) [44], both of which can learn any Boolean function.

All the models described so far use *controlled* perturbation experiments in which the stimuli that initiate the signal transduction process are known *a priori*. Fraenkel, Huang, and their collaborators [45]–[47] have developed a family of models using the theory of prize-collecting Steiner tree that reconstructs a Boolean network as well as the most probable source of the signals based only on the change of state of some nodes. The prize-collecting Steiner tree problem begins with a weighted tree and a set of terminal nodes and the goal is to identify that tree that optimizes a cost function involving the weights of the arcs and the “prizes” of the nodes. The original formulation of the problem considers only undirected graphs but the authors have extended to the directed case as well [47], however at the cost of increasing the complexity and thus compromising the optimality guarantees of their approach. Another limitation of this approach is that it cannot learn arbitrary Boolean functions, each arc is treated as an independent IF-THEN clause (the most common case). Instead, its main advantage is that it can be used to infer the cause of change between two set of cells, usually normal and diseased, by using their differences as the result of a signaling cascade.

Apart from Boolean networks, another logic-based tool for modeling signaling networks is *interaction graphs* (IG) [48]. Interaction graphs are *signed directed graphs* where each arc is labeled with a sign to indicate the relationship between the joining nodes. Negative arcs model events like dephosphorylation or protein cleavage. This abstraction, uses a more expressive 3-value logic to model the state of each node: up-, down-, and un-regulated (+1, -1, and 0 respectively). Melas et al. [49] used IGs together with an ILP formulation to learn a signaling network by refining a prior knowledge IG to simulate a set of perturbation experiments. Instead of Boolean functions, they used *sign consistency* to determine the state of a node from its potential predecessors. Sign consistency is a set of rules describing the valid configuration of node values for an IG. The general framework is discussed in [50] but the authors in [49] used a simpler version (described in Section 2.2.1). A major difference between Boolean functions and the sign consistency rules is that the latter is not deterministic but allows for some ambiguity in the results. Finally, the authors also proposed an ILP formulation to identify the *Minimal Correction Set*, i.e. a set of arcs that when included in the recovered network can reproduce the experimental outcome exactly.

Most of the method described in this section, except those by Fraenkel and Huang, require as input a directed acyclic graph (DAG). This is a fundamental limitation for this type of problem. A cycle in the network can self-activate, due to circular logic, regardless of the event that initiated the signaling process which renders the signaling network unidentifiable because there is no way to decouple the two events. In most cases, the authors manually curated the networks to remove cycles before passing them to the algorithm. In [43], the authors developed a heuristic approach to remove them during the preprocessing. Removing cycles

during the preprocessing, either manually or through heuristics, is not a principled approach because it can block the access to some section of the search space and thus lead to biased results. A more principled approach is to include the constraint of “no-cycles” into the problem’s constraints and let the optimization process handle it. In [44] the authors do follow this approach. In all cases, including those of Fraenkel and Huang, the resulting networks are acyclic.

## 2.2 Contributions

The main contribution of this thesis is the development of two new methods for reconstructing logic network using perturbation experiments and prior knowledge about the connectivity of the involved nodes. Both models build upon and expand the work of Melas et al. [49]. In particular, they use IGs and the rules of sign consistency to model signal transduction and an ILP formulation to learn the connectivity from the data. Despite the fact the ILP problems are known to be NP-hard, they have been proven efficient for this type of problems. The names of the two methods are:

- Sign Consistency Method (SCM)
- Reverse Causal Reasoning Method (RCRM)

### 2.2.1 Sign Consistency Method

This method, which was published in [51], builds and expands upon the work of [49]. In the original formulation, the algorithm searches over the power set<sup>2</sup> of all arcs to find the set that defines the sign consistent network that best describes the experimental data. This formulation expands the search space to include the power set of nodes. Although this change appears to increase the space a lot, the new solutions are “functionally equivalent” to the previous ones, in the sense that an “outside observer” that only sees the network’s response to external stimuli cannot tell whether the algorithm removed a node or all the edges incident to it.

From the optimizers point of view, this change allows it to make bigger leaps inside the search space and thus converge faster to an area of higher payoffs. On the other hand, once it gets to this area, the convergence may slow down because of the multiplicity of equally good solutions. However, in practice, many times the optimization process is terminated before the optimizer converges close to the global minimum. Moreover, for interactome scale networks, the user may take advantage of this feature in combination with the ambiguity allowed by the sign consistency formulation and optimize only over the nodes, of which there are typically much fewer than arcs, and achieve a good approximation of the optimal network that can be then optimized separately.

---

<sup>2</sup>The set of all subsets of a *set* including the empty set and the *set* itself

From a researcher’s point of view, there are two main benefits from introducing the nodes into the optimization process. First, nodes can be assigned weights to bias the solution towards our prior knowledge about their presence in the network or not. Second, the framework is more flexible and can include more complicated scenarios like knocking-down nodes or coupling the reactions/nodes of one cell type to another and optimize only over the nodes/reactions (see application in Chapter 3). Because the optimizer has two ways to search for a solution, imposing hard constraints on one of them is less likely to render the problem unsolvable.

### Preliminaries

This method assumes that the prior knowledge network (PKN) is given in the form of a sign directed graph  $G = (N, A, \sigma)$ . Sign directed graphs, also known as interaction graphs, extend the notion of directed graphs by assigning a sign to every arc  $\sigma : A \rightarrow \{-1, +1\}$ . The nodes ( $N$ ) of an IG can assume one of three possible states: up-regulated (+1), down-regulated (-1) and unperturbed/basal state (0). The sign of an arc represents the effect or influence the parent-node exerts on the child-node (see 2.2.1). To allow for more complicated scenarios to be modeled, each node and arc is associated with a possibly empty set of inhibitors  $I_n$  and  $I_a$  respectively. These may correspond to actual inhibitors, like drugs and small molecules which can inhibit particular reactions or other types of inhibition such as shRNA or CRISPR which can knock-down nodes completely. Inhibitors function as switches that force their associated elements to their basal state when activated. To keep the problem description concise the inhibitor sets are defined as  $I_n, I_a \subseteq N \times \{-1, 1\}$ , ie that they can be described by the state of some nodes of the graph.

Apart from prior knowledge about the connectivity, the method also requires a set of experimentally observed dependencies in order to refine the PKN so as to simulate them as closely as possible. Dependencies can be observed via perturbation experiments where some nodes of the network are perturbed (set in a non-basal state) and the state of some other nodes is recorded. In the context of signaling networks, these experiments correspond to measuring the phosphorylation status of intra-cellular proteins upon perturbing the cells. It’s assumed that these experiments capture a *steady-state shift*.

Experiments are modeled as a set  $E$  of tuples. Each tuple  $e \in E$  consists of two functions; a perturbation function  $p_e : N \rightarrow \{-1, 0, 1\}$  indicating which nodes are initially perturbed and  $m_e : M_e \rightarrow [-1, 1]$  indicating the state of the measured nodes in experiment  $e$   $M_e \subseteq V$ . A perturbation of zero indicates an unperturbed node, to force a node to assume the state zero an inhibitor for the corresponding node must be added. On the other hand, a measurement of zero is not equivalent to no measurement, that is why  $m_e$  are not defined over all nodes  $N$ . Another important distinction between  $p_e$  and  $m_e$  is that the first maps to discrete values indicating that perturbations are known with certainty while measurements are not. The image of  $m_e$  is a “fuzzy” relaxation of the three distinct states where

the distance from each discrete value  $\{-1, 0, 1\}$  is inversely proportional to the modeler's belief that the node measured was in the corresponding state.

### Sign Consistency

Signal transduction is simulated via the rules of sign consistent labeling of nodes. A sign consistent labeling  $l : N \rightarrow \{-1, 0, 1\}$  abides by a set of rules that restrict the possible state-combinations of adjacent nodes. First, in a perturbation experiment, every perturbed node must be labeled with the intended sign.

$$\forall n \in N : p_e(n) \neq 0 \Rightarrow l(n) = p_e(n) \quad (2.1)$$

Second, the state of every non-perturbed node must be equal to the state of **at least one** of its predecessors multiplied by the sign of the connecting arc.

$$\forall v \in N : p_e(v) = 0 \Rightarrow \exists a = (u, v) \in A : l(v) = \sigma(a)l(u) \quad (2.2)$$

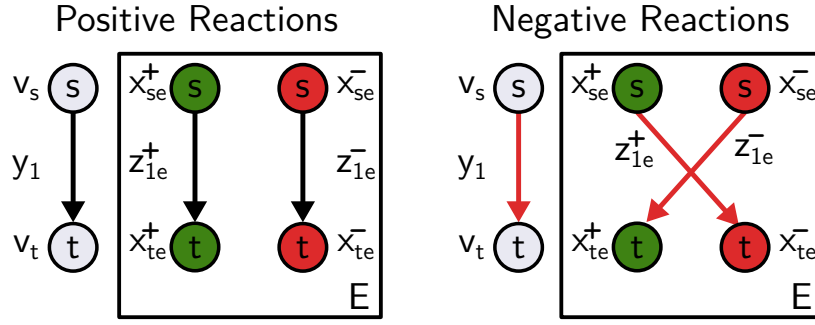
These rules produces a “weakly consistent” network.

### ILP Formulation

This section describes a mixed integer program to prune the original PKN in order to identify a subnetwork that can be labeled sign consistently in a way that minimizes the divergence between labels and measured states. The main idea of the formulation is to use the constraints in order to force a sign consistent labeling and then use decision (binary) variables to select nodes and arcs in order to minimize the objective function

Variables are indexed by  $n \in N$  to indicate the node,  $a \in A$  to indicate the arc and  $e \in E$  to indicate the experiment. For each node  $n$ , a binary variable  $v_n$  is introduced to model whether it is part of the optimal sub-network. To model its state in experiment  $e$ , two binary variables are introduced  $x_{n,e}^+$  and  $x_{n,e}^-$  modeling whether the node is up or down regulated respectively. For each arc  $a$ , a binary decision variable  $y_a$  is introduced to model whether it belongs to the optimal sub-network and two continuous variables  $z_{a,e}^+$  and  $z_{a,e}^-$  to model whether, in experiment  $e$ , the signal originated from an up or a down regulated node. The asymmetry in the  $x$  and  $z$  variables, the former being a decision variable while the latter not, is meant to accommodate the degrees of freedom that sign consistency allows. In particular, the fact that each node does not have to be consistent with all of its predecessors allows the optimizer to select the label that would result in a better overall fit. The relationship between the  $x$  and  $z$  variables is summarized visually in Figure 2.2.

The constraints modeling sign consistency can be grouped into two main groups, those regulating the state of the nodes and those regulating the state of the arcs. For notational convenience, the symbols  $s_a$  and  $t_a$  are used to indicate the source



**Figure 2.2: Reaction Modeling** Positive and negative effects are modeled by different variables and the dependencies between them characterize the reaction.

and target node of an arc  $a$  and  $\pm$  to indicate “double” constraints both for the positive and negative variables involved.

Arc constraints:

$$z_{a,e}^+ + z_{a,e}^- \leq y_a \quad (2.3a)$$

$$z_{a,e}^\pm \leq x_{s_a,e}^\pm \quad (2.3b)$$

$$z_{a,e}^\pm \leq 1 - x_{i,e}^s, \quad \forall i, s \in I_a \quad (2.3c)$$

$$z_{a,e}^\pm \geq y_a + x_{s_a,e}^\pm - 1 - \sum_{i,s \in I_a} x_{i,e}^s \quad (2.3d)$$

$$0 \leq z_{a,e}^\pm \leq 1 \quad (2.3e)$$

These must hold for every arc  $a \in A$  in every experiment  $e \in E$ . The first four constraints implement an AND gate of necessary and sufficient conditions for an arc  $a$  to transduce a signal in experiment  $e$ . The three conditions are; first, the arc must be included in the final sub-network, second, the source node must be in the appropriate state and third no inhibitors should be active. If all of these conditions are satisfied then the fourth, constraint ensures that the arc will become active. The first and the last constraints also ensure that  $z^+$  and  $z^-$  behave as mutually exclusive binary variables. The reason  $z$ s do not have to be declared as binary variables is that their value is uniquely determined for every combination of the actual decision variables of the problem.

For every node  $n \in N$  and every experiment  $e \in E$  the following constraints must hold:

Node constraints:

$$x_{n,e}^+ + x_{n,e}^- \leq v_n \quad (2.4a)$$

$$x_{n,e}^\pm \leq \sum_{a \in A: t_a = n} z_{a,e}^{\pm\sigma(a)} + \mathbb{1}(p_e(n) = \pm 1) \quad (2.4b)$$

$$x_{n,e}^\pm \leq 1 - x_{i,e}^s, \quad \forall i, s \in I_v \quad (2.4c)$$

$$x_{n,e}^+ + x_{n,e}^- \geq v_n + z_{a,e}^{\pm\sigma(a)} - 1 - \sum_{i,s \in I_v} x_{i,e}^s, \quad \forall a \in A : t_a = n \quad (2.4d)$$

$$x_{n,e}^\pm \geq \mathbb{1}(p_e(n) = \pm 1) \quad (2.4e)$$

where  $\mathbb{1}$  is the identity function (equals 1 when the condition is met).

Again, the first four constraints implement an AND gate of necessary and sufficient conditions for a node to be up or down regulated. These conditions are; first, the node must be included in the final sub-network, second, unless it's perturbed, at least one reaction must activate it and third, no inhibitor should be active. If all three requirements are met, then the fourth constraint will force the node to assume a non-basal state. The first constraint also enforces mutual exclusivity between the two active states of a node, as it cannot be both up and down regulated. The fifth constraint guarantees that the perturbed nodes have the intended state and initiates the signaling process.

The sum at the left-hand side of the fourth constraint allows the optimizer an extra degree of freedom in labeling a node. In particular, unless forced by another constraint, a node can become either up or down regulated regardless of the sign of  $z$ , and the optimizer has to decide the best option. If there are no conflicting influences then there is no decision to be made since the second constraint would have already forced one of the two  $x$  variables to zero. Because of this constraint,  $x$ s have to be binary variables, unlike the  $z$ s.

Finally, to produce the sub-graph that best simulates the data an appropriate objective function must be defined. A common choice is that of least-squared difference between the observed  $m_e(n)$  and the predicted value  $x_{ne} = x_{n,e}^+ - x_{n,e}^-$  of node  $n$  in experiment  $k$ . The objective is then to minimize the function  $(x_{n,e} - m_e(n))^2$  across all experiments, which is equivalent to the following linear function:

$$\begin{aligned} \sum_{e \in E} \sum_{n \in M_e} (x_{n,e} - m_e(n))^2 = \\ \sum_{e \in E} \sum_{n \in M_e} (1 - 2m_e(n)) x_{n,e}^+ + (1 + 2m_e(n)) x_{n,e}^- \end{aligned} \quad (2.5)$$

Where the square terms are dropped because the  $x$ s only assumes the values 0 and 1. Extra terms can be added to incentivize sparsity on the solution vectors or prioritize the inclusion of some arcs and/or nodes.

## Cycle Removal

The final subnetwork will always be an acyclic graph rooted at the perturbed nodes and branching towards the measured nodes. Nonetheless, the prior knowledge network may contain loops. Loops present a challenge for logic-based methods since they can be self-activated without external perturbation. Because the up and down regulated states are mutually exclusive, negative feedback loops are automatically handled by the constraints presented thus far. However, for the case of positive feedback loops an extra set of constraints and variables have to be introduced. In particular, for every node  $n \in N$  a variable  $d_n$  and the following constraints are introduced:

Cycle constraints:

$$d_{t_a} \geq d_{s_a} + 1 + (y_a - 1)M \quad \forall a \in A \quad (2.6a)$$

$$0 \leq d_n \leq M \quad \forall n \in N \quad (2.6b)$$

In the presence of a cycle, these constraints will push the  $d$  variables involved to grow without bound. Thus, by bounding them up by  $M$ , the optimizer is forced to break the cycle by removing some arcs.  $M$  represents a big constant and can be thought of as the longest distance between any pair of nodes and thus the best “uninformative” value is  $M = |N| - 1$ , corresponding to a path spanning all the nodes. It’s also worth mentioning that every node will have a larger  $d$  variable than all its predecessors and thus the order of the  $d$  variables corresponds to a topological ordering of the nodes which is only defined for acyclic graphs.

## 2.2.2 Reverse Causal Reasoning Method

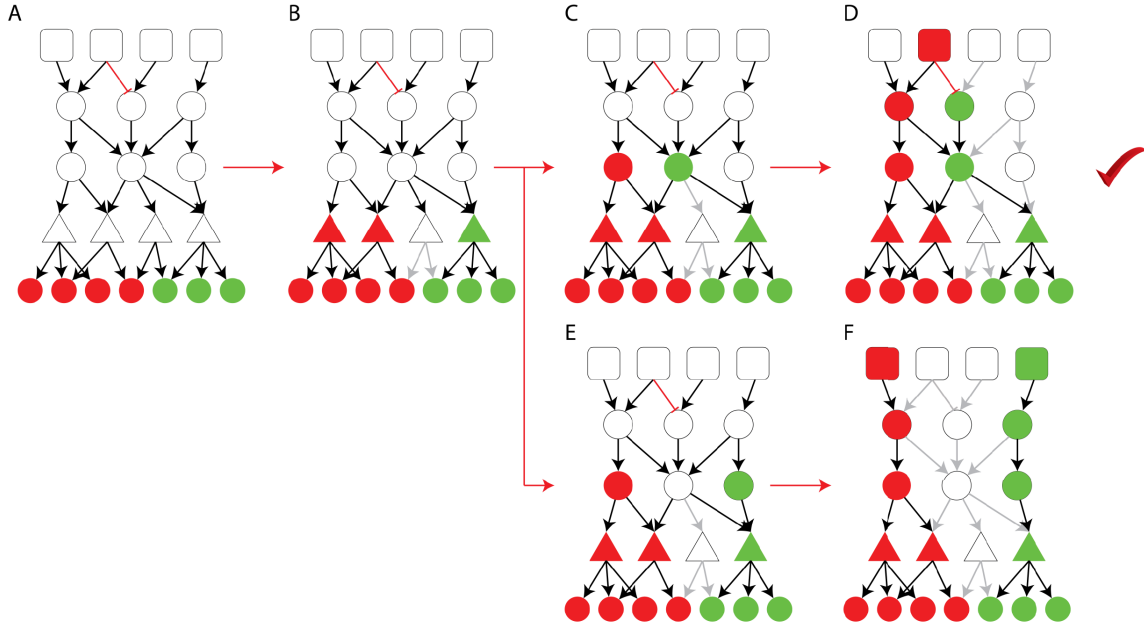
This method was first described in [52], here it is slightly adjusted to its more general form and also some variables are renamed for notational consistency. It is based on the same ILP formulation as the Sign Consistency Method (Section 2.2.1) with slight adjustments in order to address the scenario where the source of perturbation is unknown. The main adjustment consists of introducing another set of variables to indicate whether a specific node is an initiator of a signaling cascade. Figure 2.3 depicts a toy example that illustrates the main ideas of the method.

### ILP formulation

All the parameters, variables, and constraints from Section 2.2.1 are still in place. This Section describes only the adjustments made to the previous formulation.

In the Sign Consistency Model, the initial perturbation in experiment  $e$  was known *a priori* as the function  $p_e(n)$ . For this method, the function  $p_e(n)$  is replaced with two sets of binary variables  $p_{n,e}^+$  and  $p_{n,e}^-$  indicating whether node  $n$  initiated a signaling cascade in experiment  $e$  by being up- or down-regulated respectively.





**Figure 2.3: Illustrative example of the Reverse Causal Reasoning Method** (A) A PKN and a set of measured nodes is provided. Red arcs represent negative interactions while node shape indicates different levels of biological hierarchy. Squares represent source candidates. (B) The algorithm leverages searches for a sign consistent labeling to simulate the observed states. (C-F) The algorithm continues its search over all the possible paths and selects the one that makes the fewest assumptions (minimum number of nodes and sources)

These variables replace the perturbation function in constraints 2.4b and 2.4e as follows:

$$x_{n,e}^{\pm} \leq \sum_{a \in A: t_a = n} z_{a,e}^{\pm \sigma(a)} + p_{n,e}^{\pm}$$

$$x_{n,e}^{\pm} \geq p_{n,e}^{\pm}$$

Another constraint, that is implicitly imposed, is that of mutual exclusivity. This is taken care through the constraints already present for the  $x$  variables. In particular, the scenario where both  $p_{n,e}^+$  and  $p_{n,e}^-$  are set to 1 would result in forcing both  $x_{n,e}^+$  and  $x_{n,e}^-$  to 1 which has already been ruled out as a possibility by constraint 2.4a.

Sparsity constraints are important for this formulation. If the number of active or perturbed nodes is not restricted somehow, the optimizer would opt for the trivial solution of simply perturbing all the observed nodes to achieve a perfect score. These constraints can be imposed on either the  $x$  or the  $p$  variables in a soft manner as weights in the objective function, as is the case with the Sign

Consistency formulation. So the objective function becomes:

$$\sum_{e \in E} \sum_{n \in M_e} (x_{n,e} - m_e(n))^2 + \lambda_x |x_{n,e}| + \lambda_p |p_{n,e}| \quad (2.7)$$

where  $|x_{n,e}| = x_{n,e}^+ + x_{n,e}^-$  and  $|p_{n,e}| = p_{n,e}^+ + p_{n,e}^-$

$\lambda_x$  and  $\lambda_p$  define the trade-off between accuracy and sparsity. Typically,  $\lambda_p$  is set higher than  $\lambda_x$  to allow the optimizer some leeway to form the network but not include many “source” nodes. It is also typical to have some idea of which -few- nodes are the possible stimuli, like known drug targets for example, in which case all the other  $p$  variables can be fixed at 0.

## 2.3 Expansion and Future Work

This section describes some practical issues and limitations encountered when implementing the 2 models and how they can be addressed.

### 2.3.1 Probabilistic Networks

For both models, a solution consists of an interaction graph together with the appropriate sign-consistent labeling. When sparsity incentives are included, this solution is usually the unique optimal solution. However, focusing on a single solution can lead to overconfident conclusions. Thus many times some *solution averaging* scheme is required.

The simplest such scheme simply averages the value of each decision variable ( $v$ ,  $y$ ,  $x$ ,  $p$ ) over a set of available solutions weighted by their objective values. Obviously, the objective values have to undergo some transformation to become proper weights since smaller ones must contribute more and they may have different signs. The simplest of these transformation is a linear map to  $[0, 1]$ , where the smallest values are mapped to 1 and the largest ones to 0 or some small constant to avoid ignoring them completely. Moreover, care must be taken in order to make sure that the networks are not “functionally equivalent”. For example, in absence of sparsity incentives, any node or arc unconnected to an active node can be added to the solution and yield an equally good solution, since the objective function only depends on the  $x$  values of the measured nodes. For this case, extra constraints can be added of the form:

$$y_a \leq \sum z_a^+ + z_a^-, \quad \forall a \in A$$

$$v_n \leq \sum x_n^+ + x_n^-, \quad \forall n \in N$$

to ensure that arcs and nodes are included only when required.

There are 2 ways to generate a set of solutions. First, the solutions generated during the optimization process can be stored<sup>3</sup>. Second, once the optimal solution has been reached, a new ILP problem can be formulated where the original objective function is constrained to be less than the optimal objective value, scaled by some relaxation factor. The new problem needs no objective function since the goal is just to find valid networks that are close to the optimal. In practice, most modern solvers have built-in routines that allow them to *populate the solution pool* according to multiple criteria.

Although this averaging scheme makes intuitive sense, users must be aware that this is not a proper probabilistic but a rather *ad hoc* approach. In particular, the sub-optimal solutions are generated via a deterministic process and the only guarantee of convergence is that the space of possible solutions is finite.

Thus, further study is required to understand the properties of these probabilistic networks. Some interesting vectors of research might include:

1. determining a rough estimate of how many solutions are required per variable for its probability to converge
2. studying how robust are the probabilities to changes in the normalization process that determines the objective function
3. taking into account the correlation between the variables and how they converge.

### 2.3.2 Hybrid & Dynamic Networks

The models described in this dissertation are based on two major simplifications: the response is discrete and the resulting network is acyclic. Generalizing these simplifications pushes the two models and their underlying framework to their limit. This sub-section briefly discusses alternatives to both of these.

As mentioned above, the acyclic constraints are necessary for the formulation to behave properly since cycles can be auto-activated. However, many real networks do have cycles and use them to regulate cell behavior [53]. In some applications, omitting the cycles makes sense because the cells are not allowed enough time to respond, for example, if the cells start in a hyper-inactive state, e.g. after they have been starved, and the response is measured only minutes after the original stimulation. In this case, it is reasonable to assume that the signal did not have time to traverse the whole cycle once.

An “easy” workaround, to include cycles in the final network, is to expand the state-space and include *simple* cycles as nodes. These nodes must have an arc to/from every other node down/up stream of the cycle. If the number of cycles is

---

<sup>3</sup>When the Simplex algorithm is used to solve the problem, the solution is recovered as the final term in a sequence of solutions of reducing cost.

small this approach would work but the number of cycles can be exponential in the number of nodes and thus prohibitive.

The most important challenge that cycles pose is not the technical problem of self-activation but the fundamental problem of “time”. An implicit assumption for both SCM and RCRM is that the cells transition between two steady-states in a single time-step, before and after the signal has been “processed”. Cycles break this assumption by encoding extra time steps, namely the time-points when the signal reaches one of their nodes for a second time. Thus it is the author’s belief that a proper solution to this problem cannot be achieved within the given framework, unless time is modeled explicitly. No such line of research is known to the author.

On the other hand, the discretization assumption is more nuanced. It can be thought of as an approximation of the sigmoid behavior of many biological processes which tend to saturate at very low or high stimulation. Consequently, it is only limiting in the gray “middle area”. A possible workaround could be to replace the threshold-activation-function with sigmoid-functions similar to the work of *Mitsos et al.* [54] for Boolean networks.

For the case of RCRM, an interesting approach is to allow the state of the measured nodes to be continuous and reformulate the problem as a network flow problem [55, Chapter 7], where the “demand” of the measured nodes has to be covered by some unknown perturbations. The challenge of this approach is that the states can assume both positive and negative values and thus the relevant theory and algorithms cannot be applied directly. In particular, network flow requires that the “flow of signal” coming in and going out of a node must be equal. So in the case where “negative signal” is allowed, a node can generate signal without being a designated stimulus ( $p_{ne} = 0$ ) by simply balancing the positive and negative signals.

The author does not know of any published method based on either of these ideas.

# Chapter 3

## Replicative Senescence

This chapter presents the work published in [51]. In this work, the authors utilize a special form of the Sign Consistency Model (SCM), presented in Section 2.2.1, in order to elucidate changes in the network of human primary fibroblast cells (HFL-1) as they age. The chapter follows the original work closely with only slight modifications to match the nomenclature and previous content of this thesis.

### 3.1 Introduction

As discussed in Chapters 1 and 2, models describing the cell's signaling process is a core aspect of Systems Biology research. A common approach is to combine prior knowledge about the (static) connectivity of the signaling proteins with high-throughput (dynamic) data in order to come up with a network specific to the activity under study. However, to the best of the author's knowledge, the methods proposed thus far in order to reconstruct signaling networks focus on selecting a subset of the known reactions that can optimally simulate the observed behavior. This approach is justified for two main reasons. First, as discussed Section 1.2.3, reactions are context dependent and some argue that currently available databases include a lot of false positive ones. Second, cells can be probed in the presence of inhibitors, like drugs or small molecules, that inhibit specific reactions and thus the effect of these interactions can be evaluated directly.

However, signaling networks can also be altered in cells by the presence or absence of the involved proteins, either because the cell under-expresses them or researchers have knocked them down, using CRISPR or shRNA for example. In this chapter, the Sign Consistency Model (SCM) described in Section 2.2.1 is used to reconstruct the signaling networks of human primary fetal lung fibroblast cells (HFL-1) as they undergo replicative senescence. In the case of aging, the reactions involved are expected to remain unaffected by the process and only the expression of the involved entities should be affected.

## 3.2 Materials and methods

### 3.2.1 Experiments

We maintained human primary fetal lung fibroblasts (HFL-1) in culture from a young stage until they reached a terminally senescent (old) stage after 50 replication cycles. In the young stage the cells were replicating every day while in the old stage they had not replicated for a month. Cells were seeded in 96-well plates at 20000 cells/well 24 hours prior to stimulation. After 24 hours cells were stimulated by adding the stimulants to the cell medium at a concentration calculated to reach the target concentration in the cell supernatant. Cell lysates were collected at 5 and 25 minutes following the cytokine stimulation. The 5 and 25 minute time points were identified in a preliminary experiment as the optimal time points to capture early phosphorylation activities. A panel of 18 phosphoproteins upon stimulation with 6 cytokines was used to interrogate the cells. The panel was selected to represent known proliferation and senescence regulating pathways. Stimulant concentrations were selected after pre-screening experiments. All measurements were done in triplicates. The cytokines used to stimulate the cells and their target concentrations were: EGF (100ng/ml), IL1A (50ng/ml), TGFA (200ng/ml), TNF (100ng/ml), IGF1 (200ng/ml) and INS (1000ng/ml). The interrogated proteins were: AKT1, CREB1, PTK2, GSK3A, HSPB1, NFKBIA, JUN, MAPK12, MAPK3, MAPK9, MAP2K1, TP53, PTPN11, STAT1, STAT3, STAT5, STAT6 and RELA. Basal protein phosphorylation was measured in absence of any stimulant in DME cell medium.

HFL-1 human primary embryonic lung fibroblasts were purchased from ECACC (ECACC 89071902). Cells were cultured in Dulbecco's modified Eagle's (DME) medium (Invitrogen, Carlsbad, CA, USA) supplemented with 10% fetal bovine serum (Invitrogen), 100 units/mL penicillin, 100 µg/mL streptomycin, and 2 mM glutamine (complete medium) and maintained in a Binder Incubator at 37°C, 95% air, 5% CO<sub>2</sub>. HFL-1 fibroblasts were seeded at a density of  $2 \times 10^5$  cells per 75 cm<sup>2</sup> flask, were subcultured at a split ratio of 1:2 when cells reached confluence, until they entered senescence at about 50 CPD, as described before [56]. In all experimental procedures described below early passage (young; CPD < 25) and late passage (senescent CPD > 50) HFL-1 cultures were used. Cells were fed approximately 16 hours prior to the assay and cell number were determined in duplicates using Coulter Z2 counter (Beckman Coulter, Nyon Switzerland). The cytokines used for stimulation were purchased from Peprotech (PeproTech, NJ, USA), IGF1 (100-11), EGF (AF-100-15), IL1A (200-01A) TGFA (100-16A), TNF (300-01A) and insulin (INS I9278) was obtained from Sigma-Aldrich (Sigma-Aldrich, Shanghai, P.R. China). Measurements were carried out in 96-well plates using Luminex's xMAP technology with custom assays by ProtATonce Ltd.

### 3.2.2 Network Construction

The median fluorescent intensity was used to summarize the results of the Luminex measurements. To map values to the  $[-1, 1]$  interval required by the algorithm, fold changes were computed between the cytokine-stimulated state and the basal state (DME) and passed through the Gaussian error function. The two time points (5 and 25 minutes) were aggregated into a single “early” time point by taking the maximum absolute value which is analogous to an OR gate. See A for more details about data pre-processing.

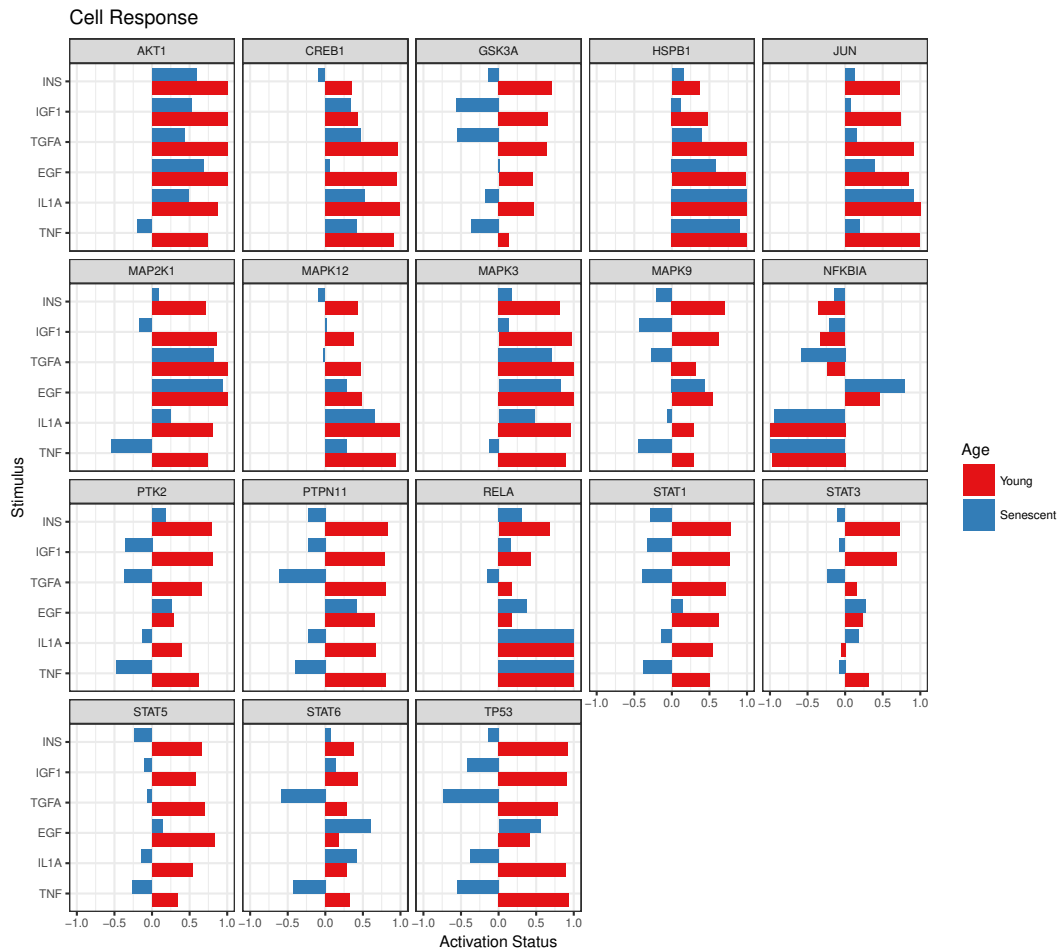
The prior knowledge network (PKN) was assembled by combining pathways from the MetaCore database. In particular, pathways that contain at least one of the stimuli used or readout measured were included. Then, nodes that are not controllable, meaning that there was neither path from a stimuli nor to a signal, were filtered out. These steps do not affect the final outcome since the solver would have eliminated them during the pre-processing because, from its perspective, their state is fixed at zero. The network was further curated by removing reactions with unknown effects or few evidence. These interventions do bias the solution but they were justified in order to avoid overfitting.

Finally, the SCM was used to reconstruct the signaling networks for young and senescent cells. The two cells were not fitted independently though. As explained in Section 3.1, the reactions involved are expected to remain unaffected by the process and only the expression of the involved entities should be affected. In order to achieve this effect, two instances of the SCM were formulated but only a single instance of  $y$  variables, representing the reactions included in the final networks, was used for both networks while for all other variables two instances, one for young and one for senescent cells, were defined. The ILP formulation is otherwise identical to the one presented in Section 2.2.1. The two networks were fitted together.

## 3.3 Results

Figure 3.1 summarizes the measured states  $m_e$  of every phosphoprotein measured upon stimulation with every cytokine both for young and senescent cells. As expected, young cells are much more responsive across the board. We observed that the signaling proteins that are involved in proliferation including AKT Serine/Threonine Kinase 1 (AKT1), cAMP Responsive Element Binding Protein 1 (CREB1), Mitogen-Activated Protein Kinase Kinase 1 (MAP2K1), Mitogen-Activated Protein Kinases 3 and 9 (MAPK3, MAPK9), Protein Tyrosine Kinase 2 (PTK2), Protein Tyrosine Phosphatase, Non-Receptor Type 11 (PTPN11) and Signal Transducer and Activator of Transcription (STAT) family members 1, 3 and 5 (STAT1, STAT3, STAT5) are substantially more activated in young than senescent cells. Senescent cells show a higher activation of STAT6 when stimulated by EGF or IL1A. Stimulation with EGF also provides a higher response in senescent than

young cells in the NFKB Inhibitor Alpha (NFKBIA), RELA, STAT3, STAT6 and Tumor Protein P53 (TP53) and overall it seems that EGF is the growth factor that senescent cells are most responsive to. Notably, Glycogen Synthase Kinase 3 Alpha (GSK3A), TP53, STAT1, STAT5 and MAPK9 respond in opposite ways in young and senescent cells across multiple stimulants. Heat Shock Protein Family B Member 1 (HSPB1) and Jun Proto-Oncogene (JUN) are more responsive in young cells as well.

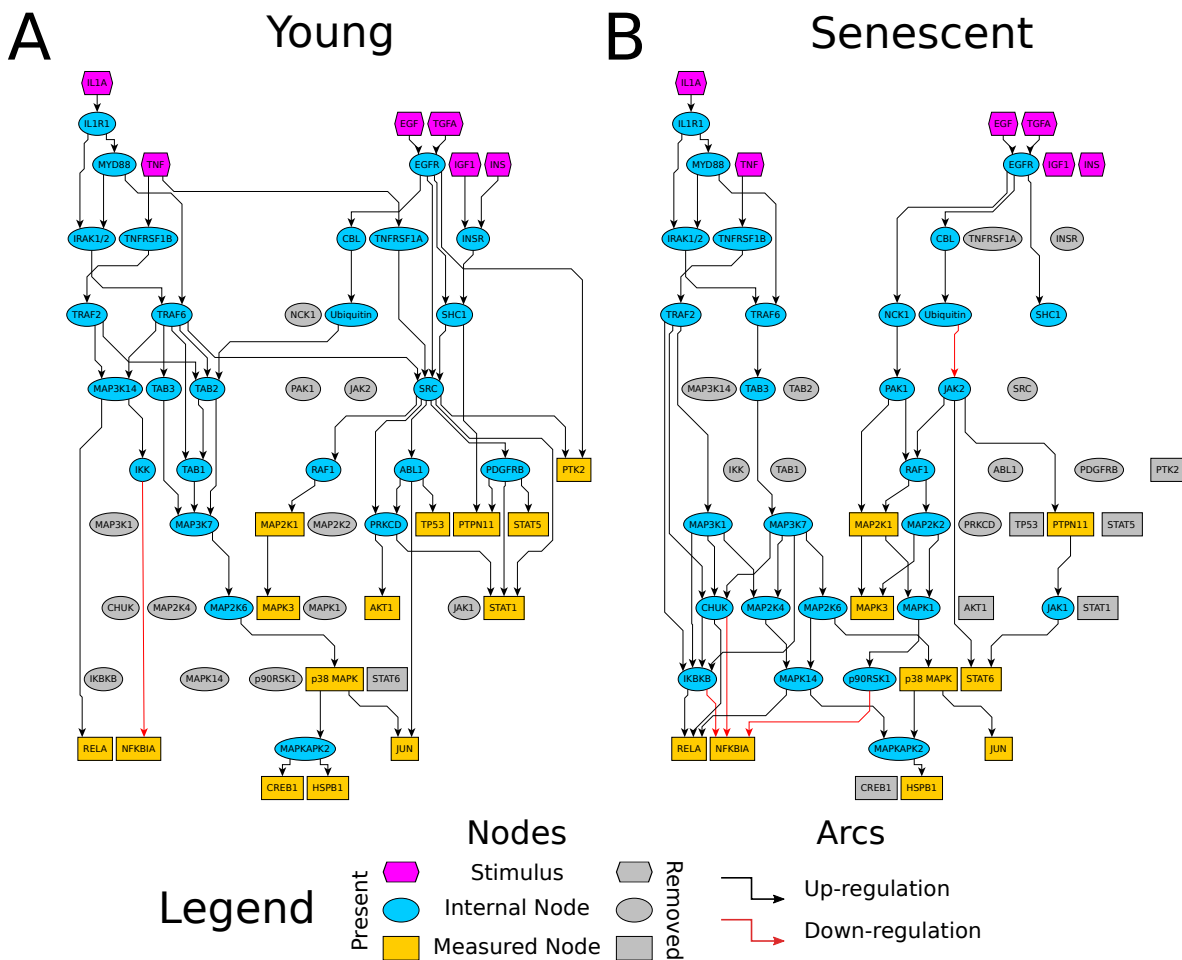


**Figure 3.1: Cell Response to different stimuli** Panels correspond to measure phosphoproteins. On the x-axis is their measured status (-1: down-regulated, 1: up-regulated) and on the y-axis the stimuli used in every experiment.

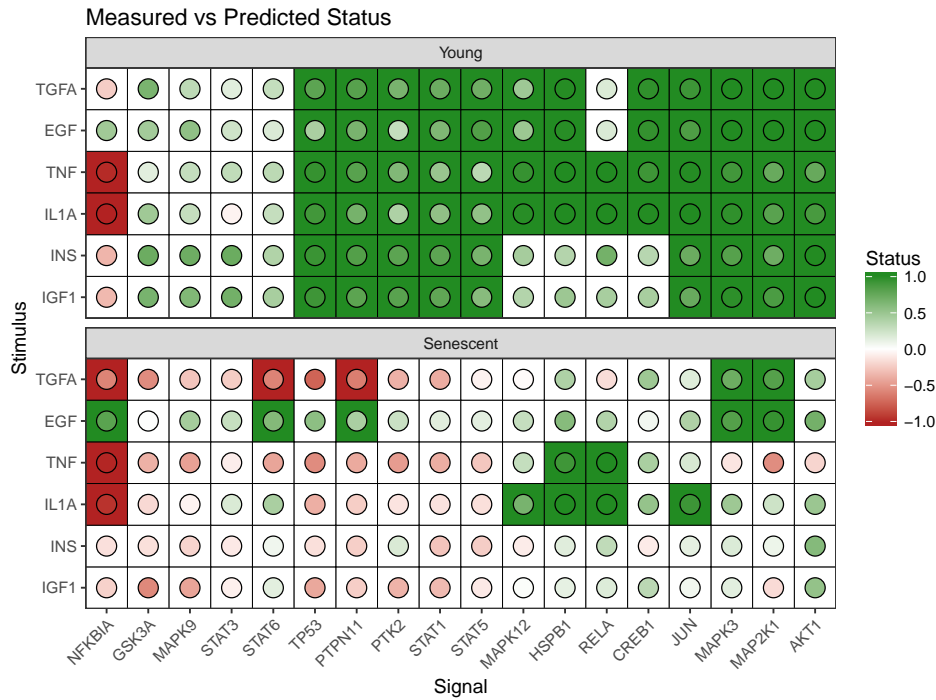
To gain a more mechanistic insight into what drives this change in response, we run our algorithm for young cells, replicating every day, and old cells that have entered the senescence state. The resulting networks are shown in Figure 3.2, while the predicted status of the signals is shown in Figure 3.3. As expected senescent cells appear unresponsive to insulin (INS) and insulin-like growth factor 1 (IGF1) because their insulin receptor (INSR) appears to be unresponsive (absent). Moreover, signaling through the TNF receptor 1 (TNFRSF1A) is also absent in



senescent cells. As a result, the SRC proliferative pathway that is active in young cells, leading to activation of proteins involved in cell growth and proliferation such as AKT1, STAT1, STAT5, and PTK2, is completely deactivated in senescent cells. On the other hand, NFKBIA is highly down-regulated by upstream inhibitors, Conserved Helix-Loop-Helix Ubiquitous Kinase (CHUK) and Inhibitor Of Nuclear Factor Kappa B Kinase Subunit Beta (IKBKB), and RelA/p65 is highly up-regulated in senescent cells which leads to activation of NFKB signaling that establishes and maintains cellular senescence. Both Janus Kinase 1 (JAK1) and Janus Kinase 2 (JAK2) appear to be inactive in young cells and active in senescent cells leading to inhibition of NFKBIA and activation of STAT6. Several MAP kinases are only active in senescent cells including MAP3K1, MAP2K2, MAP2K4, MAPK1 and MAPK14. Another important transcription factor that remains inactive in senescent cells is CREB1, known to play a significant role in cellular aging [57].



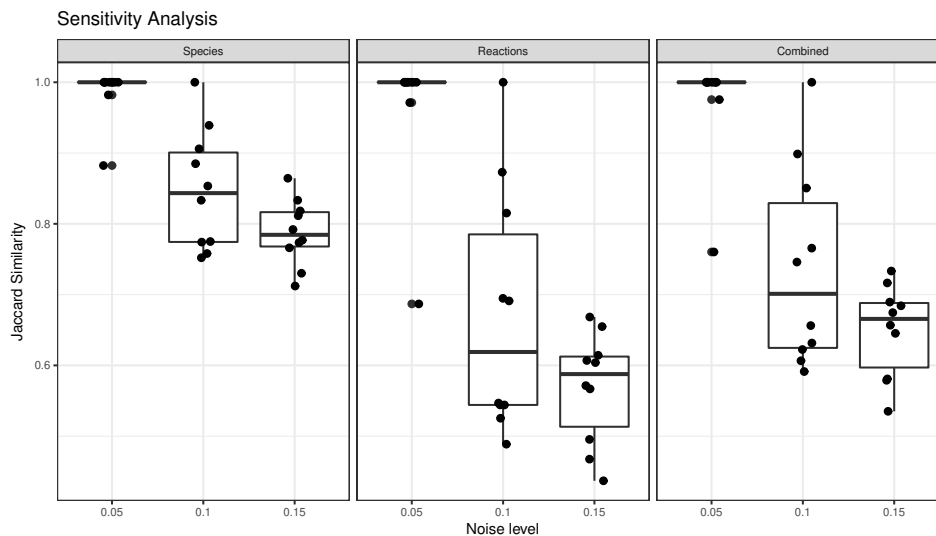
**Figure 3.2: Network comparison** Grayed out nodes are not present in the network but included for comparison. A: Signaling network for young cells. B: Signaling network for senescent cells.



**Figure 3.3: Network Predictions** Every tile corresponds to a stimulus-signal combination. The tile color corresponds to the predicted status of the signal while the color of the enclosed disk corresponds to the measured status. Predicted status is discrete while measured status is continuous.

### 3.3.1 Sensitivity Analysis

Technical variability for bead-based phosphorylation assays has been studied before in [58], where they estimated the coefficient of variance due to the experimental protocol and pipetting errors to average in the range of 5–15%. To assess the impact of such variability in our results, we generated artificial datasets by adding noise to our MFI readouts and rerun the analysis. We repeated this processes 10 times for three different levels of noise 5, 10 and 15%, and computed the Jaccard similarity index of the resulting network with the original network. We observed that a CV of 5% has virtually no effect on the results while for larger values the networks start to diverge but not significantly. The results are summarized in Figure 3.4.



**Figure 3.4: Sensitivity Analysis** 10 randomized datasets were generated for each noise level and the optimal network were reconstructed and compared with the original dataset. The comparison was based on the presence of reactions ( $y$ ), species ( $v$ ) or both.

## 3.4 Discussion

In this chapter, an adaptation of the SCM, described in Section 2.2.1, was presented. The changes made allowed the method to model an evolving system like cellular aging. For the case study presented, the reaction set was constrained to be the same across both young and senescent cells while the node set was allowed to vary. The opposite configuration could have been implemented just as easily. The configuration used allowed networks to differ only due to changes in the expression levels of the proteins while also allowing more data points to inform which reactions should be included. More generally, within the presented framework, one can impose soft or hard constraints on nodes and/or reactions without necessarily compromising the accuracy of the model, since the two mechanisms allow the optimizer to work around them if necessary.

The sensitivity analysis performed illustrated that the set of reactions included in the final network is more sensitive to noise than the set of nodes, despite the fact that reactions were constrained to be the same for the 2 populations. This is probably due to their moderate effect on the overall connectivity compared to the effect of removing a node from the network. Thus extending the scope of optimization to nodes as well as reactions renders the model less sensitive to overfitting.

As with the original SCM framework, the proposed formulation can handle both positive and negative interactions as well as cycles in the prior knowledge network transparently. However, the resulting networks are acyclic so it cannot capture dynamic behavior of the network. Also, as is the case with all logic-based approaches, it lacks the ability to describe the kinetic aspects of the network [39]. Because of the MIP implementation, the networks constructed are optimal with respect to the proposed framework or at least within a pre-specified distance from optimal [55]. Moreover, modern MIP libraries like CPLEX and Gurobi, allow the enumeration of multiple Pareto-optimal or near optimal, solutions. When tested with synthetic data that respected sign consistency, it was able to reconstruct the generated signatures in every occasion (data not shown).

The number of variables/constraints required is  $O(2k(N + A))$  where  $N, A$  is the number of nodes and interactions in the PKN and  $k$  is the number of experiments simulated. Although up and down regulation are not independent, they require their own variables and constraints to be implemented properly so that is why 2 appears in the model. Depending on the infrastructure and the number of experiments the proposed method can handle networks for thousands of species. Practical experience indicates the most important parameter, in terms of running time, is the number of experiments  $k$ . Lastly from a computational point of view, the number of proteins is typically only a fraction of the number of reactions resulting in a possibly significant shrinkage of the search space, if the optimization is carried over the proteins only.

To demonstrate this framework, the signaling networks of cells aging *in vitro* was reconstructed. In particular, 18 phosphoproteins were measured and their fold-change, in concentration upon stimulation with 6 cytokines, was computed for young and senescent cells. The MetaCore database was used to assemble the pool of possible interactions.

The two networks were compared and we observed classical changes associated with senescence such as insulin resistance and strong activation of the NF $\kappa$ B pathway in response to IL1A and TNF stimulation, which has been associated with the senescence-associated secretory phenotype (SASP) [59]. The activity of JAK1 and JAK2 in senescent cells is one potential driver of the SASP through inhibition of NF $\kappa$ BIA and activation of STAT6. STAT6 has been found to be involved in the induction of cellular senescence by [60]. Various studies have proposed that the SASP can be suppressed through inhibition of the JAK pathway in senescent cells [61]. Stress activated MAP kinases cascades are also known to play an important role in cellular senescence [62]. The insulin and IGF1 signaling (IIS) pathway is a highly studied pathway in aging. The attenuation of IIS has been shown to promote longevity and extend the lifespan of various organisms [63]. However, IIS is also known to decline during normal and accelerated aging [64]. The recovered networks suggest that the attenuation of the IIS is mediated through the inactivation of the INSR. A central driver of cell proliferation in young cells, SRC is completely inactive in senescent cells, leading to inactivation of various proteins

involved in cell proliferation. It is interesting to mention that the SRC pathway has been associated specifically with geriatric cancers, with a higher probability of being deregulated in elderly patients [65], [66]. Analysis of microarray data from the same experiment (paper in preparation, data not shown) confirms the perturbation of EGFR and INSR signaling, as well as perturbation of downstream signaling by SRC in senescent cells, while the cellular response to IL1A is increased. In young cells the transcriptomic data demonstrate JAK1 and JAK2 inhibition.



# Chapter 4

## Drug Induced Lung Disease

This chapter presents the work published in [52]. In this work, the authors utilize a special form of the Reverse Causal Reasoning framework (RCR), presented in Section 2.2.2, in order to elucidate the modes of action (MoA) of drug-induced lung disease. The chapter follows the original work closely with only slight modifications to match the nomenclature and previous content of this thesis.

### 4.1 Introduction

The identification and understanding of a drug's mode of action (MoA) is at the core of pharmacology. For the many drugs that target signal transduction processes, this requires an understanding of the MoA at the signaling level and in the specific tissue where the drug is to be used, along with other tissues that may be subject to off-target effects. Understanding this could have an enormous impact in many aspects of drug development and public health [67].

Ideally, one would have dedicated phosphoproteomic or chemoproteomic experiments [68], where thousands of proteins and their post-translational modifications, like phosphorylation, are measured upon perturbation with the interrogated drug and its MoA would be identified in terms of altered proteins. However, phospho- and chemoproteomic data are still scarce in the literature, and few datasets exist that screen drugs on different panels of cell lines in a consistent experimental setup, including culturing conditions, concentrations, time of stimulation etc.

In contrast, such data upon perturbation exist abundantly at the gene expression level [35], and they are an invaluable resource for comparative studies of drugs and cell lines, enabling the use of computational modeling for predicting drug efficacy or identifying potential drugs for repositioning [69]. Thus, the development of novel approaches that leverage gene expression datasets to identify drug modes of action, is an important question. Most computational methodologies for identifying drug modes of action based on gene expression data use one of the following two workflows.

The first workflow starts with differentially expressed genes identified upon

perturbation with the interrogated drug, and subsequently, enrichment analysis is performed to identify biological processes, signaling pathways, or other gene sets that are highly enriched in the differentially expressed genes and thus, are likely to be deregulated by the interrogated drug. The gene sets could be either GO terms or genes that are deregulated upon perturbation with known stimulants [70]. Because enrichment based strategies ignore the complex gene interactions that may drive cellular response, hybrid methods have also been proposed that take into account information from pathway maps to improve their prediction [71].

Other approaches are primarily based on the incorporation of prior knowledge of signaling networks, either from protein-protein interaction (PPI) data or causal links, or transcription regulation in addition to the gene expression data [45], [72], [73]. For example, in the work by *Zarrinhalam et al.* [73], the Selventa Knowledge-Base was used that includes causal, condition specific relationships between signaling proteins and gene expressions, and a Bayesian inference approach was used to identify subsets of this knowledge base that are most probably active in the specific biological context. They were able to identify the key regulators that govern gene expression, but they could only capture limited mechanistic aspects of the intermediates in signal transduction, i.e. how signal propagates from one protein to the next before translating into the gene expression level. In another work [74], a PPI network was used to represent protein connectivity, and an enrichment analysis method was implemented to infer the activity of transcription factors and signaling proteins based on the observed gene expression signatures. Similarly, *Huang et al.* [45] used a PPI network to represent protein connectivity, and implemented a Prize Collecting Steiner Tree (PCST) algorithm to identify minimum sub-trees of the PPI network that connect differentially expressed genes or proteins, discovering the backbone networks that are most probably functional in the specific biological context. In more detail, the PCST algorithm addresses the problem of connecting into a Steiner arborescence tree as much of the differentially expressed genes (or proteins) as possible, while minimizing the number of edges in the tree. The PCST does not impose the requirement that all differentially expressed genes/proteins are included in the solution, but the algorithm identifies only a subset of those, whose connectivity is also strongly supported by the network, thus offering robust predictions even when noisy data are used. Also, the PCST can be formulated as an integer linear programming (ILP) problem, which can be solved efficiently allowing the interrogation of genome-wide networks.

The use of PPI networks offers clear advantages over strictly data driven methods. Firstly, these methods combine gene expression data-sets with the wealth of published high throughput interaction data, making model predictions more biologically relevant. Secondly, the identification of network topologies implicated in drug response is easier to be interpreted, as it offers mechanistic insight into the drug's MoA. Finally, the use of networks allows the generation of predictions for signaling molecules that are not directly measured, for example the nearest neighbors of the measured genes/proteins. Nevertheless, the use of PPI networks has



its own shortcomings. PPIs are undirected; thus, the direction of signal flow from one protein to the next is not easily identifiable. While the original formulation of PCST considered undirected networks, extensions have been proposed [47] to include directionality in the networks and to generalize from a single tree that connects together all differentially expressed genes (or proteins), to forests, that permit different, unconnected neighborhoods of the PPI network to be functional at the same time. As more complexity is incorporated in the formulation, a global solution becomes intractable, forcing the use of heuristic methods in the optimization, risking a suboptimal solution. Moreover, even these PCST extensions cannot incorporate signed data and interactions (positive vs. negative effects), while these effects are key in defining the mechanisms underlying signal transduction.

In this chapter, a methodology is introduced for the identification of drug mode of action, based on gene expression data and prior knowledge of protein connectivity in the form of a genome wide, directed signaling network. In the heart of the methodology lies the reverse causal reasoning framework (RCR), described in Section 2.2.2, modified at key points to address the complexity of large-scale signaling networks. The methodology combines gene expression data with a Prior Knowledge Network (PKN) based on signed and directed causal interactions as those that can be curated from the literature, and it identifies sub-graphs of the PKN that appear to be functional based on the data at hand.

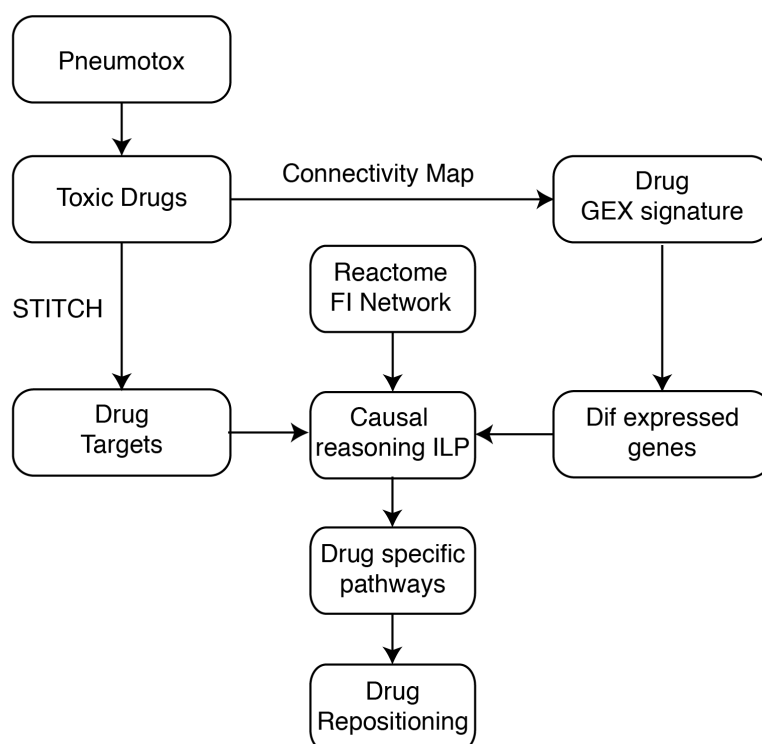
To illustrate the value of this approach, the identification of the modes of action of drugs that are known to induce lung injury is addressed. Drug induced lung injury is a major safety concern and more than 800 drugs are listed as potential inducers [34] of lung injury including asthma, fibrosis, and interstitial pneumonia. Thus, understanding the molecular mechanisms underlying Drug Induced Lung Disease (DILD) may have an impact in drug development and in public health. In this work, the MoA of 200 drugs that are known to induce respiratory problems is identified, in terms of signaling networks that start at the drug targets, go through the signaling level, and terminate at the genomic level with the regulation of genes that are differentially expressed upon perturbation with the toxic drugs. Subsequently, the drug specific pathways are merged together into a signaling network (i.e. DILD network) that captures the signaling mechanisms underlying DILD. Moreover, to demonstrate the predictive power of model predictions, its findings are used to suggest promising drugs for repositioning.

## 4.2 Analysis

### 4.2.1 Workflow

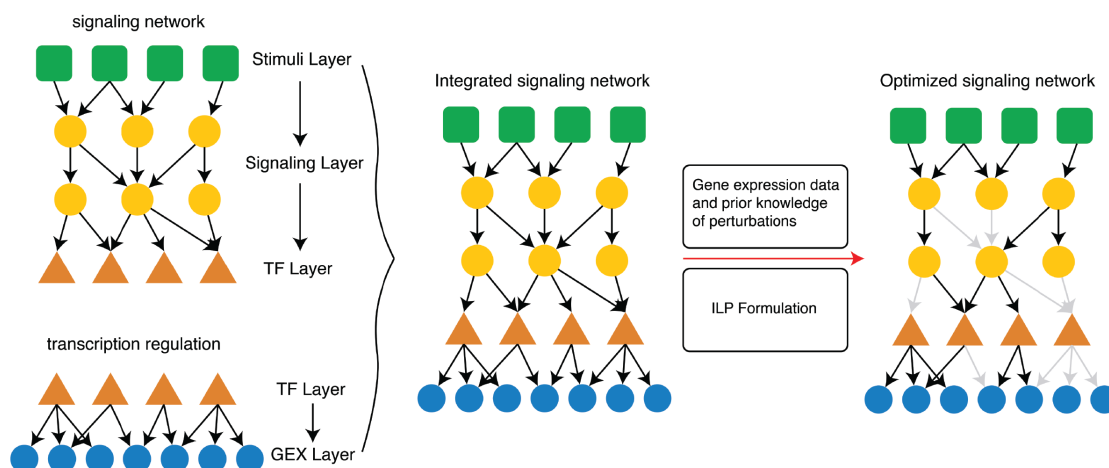
The proposed method attempts to identify the drug's MoA based on gene expression data and prior knowledge of drug targets, protein connectivity and transcription regulation. The workflow of the proposed method is shown in Figure 4.1. First, pharmacological targets are identified from the STITCH database [75], and

differentially expressed genes upon perturbation with the interrogated drugs are identified from the Connectivity Map [35]. Subsequently, an algorithm based on the RCR framework is used to identify functional interactions that model signal transduction from the drug targets to the differentially expressed genes. The identified pathways are functional sub-graphs of a genome wide signaling network, and originate at the drug targets, span across the signaling level, go through the affected transcription factors and terminate at the genomic level with the regulation of the differentially expressed genes (see also Figure 4.2).



**Figure 4.1: DILD Workflow.** Drugs that induce respiratory problems are extracted from Pneumotox. Pharmacological targets are identified from STITCH and their gene expression profiles from the Connectivity Map. Over- and under-expressed genes are identified using the rank matrix of the Connectivity Map. Then, the proposed ILP formulation is applied to identify signaling pathways connecting drug targets and over- and under-expressed genes. The drug specific signaling networks are merged into a DILD network that is subsequently used for suggesting potential drugs for repositioning to treat DILD.

In more detail, in the context of DILD, The Pneumotox database [34] is used to extract the drugs that cause lung injury. Pharmacological targets are extracted from STITCH and their gene expression profiles from the Connectivity Map, resulting in a list of 200 lung-toxic drugs with known drug-target interactions and gene



**Figure 4.2: Identification of drug’s MoA.** in terms of drug induced signaling network alterations via the proposed ILP algorithm. First, A Prior Knowledge Network (PKN) is constructed by merging prior knowledge of protein connectivity and transcription regulation. Then, the proposed ILP algorithm is implemented to identify subsets of the PKN that appear to be functional based on the data at hand. The resulting pathways start at the drug targets, span across the signaling level, go through the layer of transcription factors and terminate at the genomic level with the regulation of the differentially expressed genes.

expression profiles. Then, the Reactome Functional Interaction network [76] is used to connect drug targets, transcription factors and gene expressions as illustrated in Figure 4.1. Subsequently, the proposed ILP formulation identifies a functional sub-graph of the Reactome network, connecting drug targets and genes that are differentially expressed upon perturbation with the lung-toxic compounds.

In particular, the ILP constructs a signaling network per toxic compound. At the end the drug-specific pathways are pooled together into a signal transduction network that captures the molecular mechanisms underlying DILD (DILD network). Finally, to demonstrate the biological relevance of the DILD network, it is leveraged to identify potential drugs for repositioning that will reverse the disease phenotype. To this end, all remaining drugs in cMAP that are not lung-toxic are considered, and their targets are extracted from STITCH. If drug targets overlap with the DILD network, then the drug specific pathways of these drugs are computed using the ILP algorithm, and the drugs are ranked based on how much their pathways disrupt the DILD network. Drugs that significantly disrupt the DILD network are considered candidates for repositioning.

### 4.2.2 Lung-toxic Drugs

Lung toxic drugs were obtained from the Pneumotox database [34]. Pneumotox includes 892 chemicals known to induce lung injury. To obtain a better perspective

on the kind of drugs included in this list, chEMBL was used to extract their nominal pharmacological effects. Table B.3 includes the most frequent nominal pharmacological effects.

As a positive control observation, DNA inhibitors are at the top of the table. This is expected since DNA inhibitors are often used as anti-neoplastic agents and are inherently toxic. COX inhibitors are also at the top of the table. This is in good accordance with the literature where it has been reported that a range of COX inhibitors and other NSAIDs, frequently used as analgesics, may cause respiratory problems [77]. Beta-1 adrenergic receptor antagonists are also suspected of inducing respiratory distress, since Beta adrenergic receptors are found to be desensitized in lung injury [78]. Finally, Tubulin inhibitors may contribute to lung injury via inducing oxidative stress [79]. For the rest of the pharmacological effects there is no clear mechanism that elucidates the etiology underlying lung disease, and the adverse events are sporadic.

Next, the known targets of the toxic drugs were extracted from the STITCH database. STITCH includes all known targets of drugs, both the nominal pharmacological targets and other molecules they may interact with, based on direct experimental data, the available literature, or computational predictions. The identified drug targets will be used to model the interaction of the interrogated drugs with the cell's signaling machinery, and they are the potential starting points of the drug's MoA. Of the 892 toxic chemicals, only the ones that have known targets in STITCH and also known gene expression signatures in the Connectivity Map will be processed further, thus resulting in a total of 200 compounds/ drugs. The list is included in Supplementary Information of the original paper.

Finally, the gene expression signature of each drug was computed using the rank matrix of Connectivity Map (cMAP) dataset. For each toxic drug in Pneumotox, the top and bottom 1% of the genes were extracted from the rank matrix. All genes were pooled together and the frequency with which they are over- and under- expressed across all drugs was calculated. The 5% most frequently over- and under-expressed genes were extracted as the most significantly over- and under-expressed genes upon perturbation with the toxic compounds. The differentially expressed genes were used here as a readout of the cellular response upon perturbation with the interrogated drugs, and they are going to serve as the endpoints of the identified drug modes of action.

The lists of over-expressed and under-expressed genes are included in the Supplementary Information of the original paper. Subsequently, gene ontology (GO) enrichment analysis was performed to identify biological processes enriched in differentially expressed genes. Top results are included in the Appendix B.2 as well. Over-expressed genes are enriched in pro-apoptotic processes. This is expected, as a big part of the lung toxic compounds are chemotherapeutics or generally anti-neoplastic agents and are expected to be toxic. The list also included terms related to blood vessel development which is in good accordance with the literature, where vascular development has been reported to take place in Acute Lung Injury

(ALI) [80]. The VEGF gene in particular is over-expressed in 12% of the lung toxic drugs. On the other hand, the GO terms corresponding to the under-expressed genes are related to cell cycle, nuclear division, mitosis etc. These processes are expected to be disrupted by toxic compounds, resulting in lung injury.

### 4.2.3 Identification of drug's MoA

Here, the method described in Section 4.4.1 was employed to identify the mode of action of the 200 lung-toxic drugs in the Pneumotox list, based on their gene expression signatures from the Connectivity Map and prior knowledge of protein connectivity, drug targets and transcription regulation. At the end, all drug specific signaling networks were pooled together to obtain a signaling network that best captures the molecular mechanisms underlying drug induced lung injury i.e. DILD network.

Every drug's MoA was identified in terms of a signaling network starting at the drug targets, extracted from STITCH, spanning across the signaling level, going through the layer of transcription factors and terminating at the gene expression level with the regulation of the differentially expressed genes. The ILP algorithm identified the minimum subset of the PKN, that achieves the desired targets to gene expression connectivity. In this context, the drug targets correspond to the interface of the drugs with the cell's signaling machinery, and the differentially expressed genes represent the cellular response upon perturbation with the interrogated drugs. Thus, the identified signaling networks constitute Cue-Signal-Response models [81], capturing cells response to the toxic drugs.

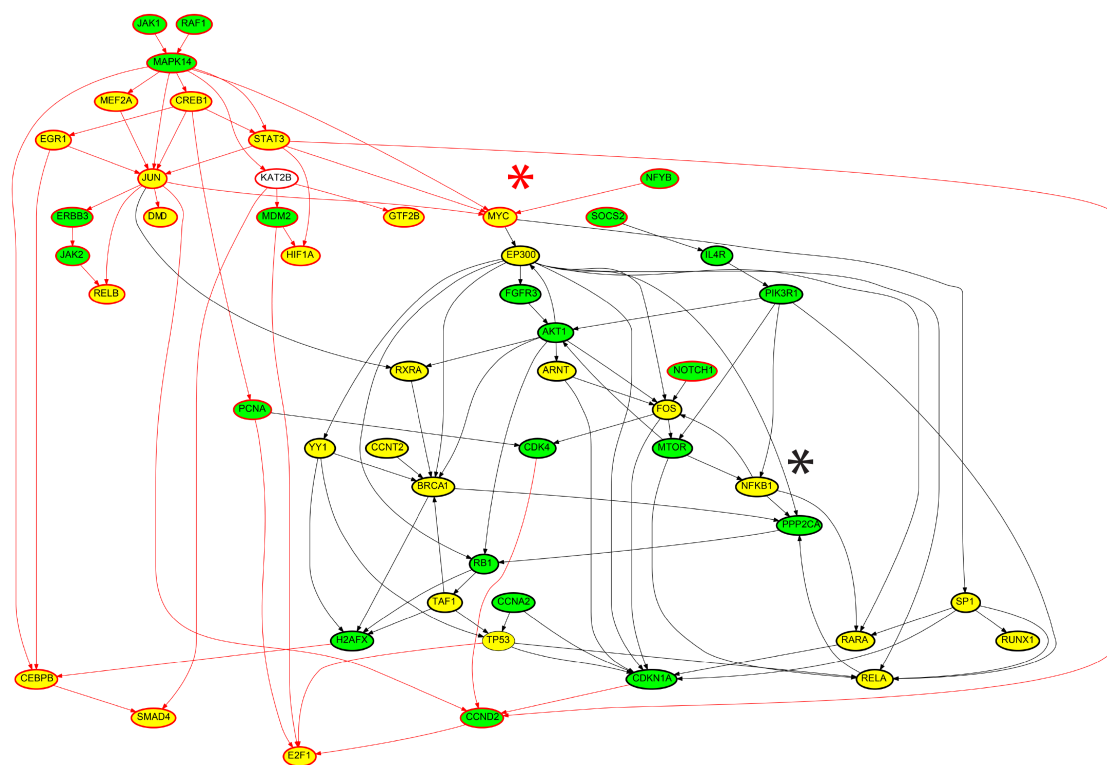
#### Case Study: Imatinib

A simple case study for Imatinib is presented, to best illustrate how the proposed method identifies the drug MoA, in terms of a signaling network that originates at the drug targets, spans across the proteomic level and terminates at the gene expression level.

Imatinib is a tyrosine kinase inhibitor used in the treatment of multiple cancers, and is also known to induce acute lung injury as one of its adverse effects. Its nominal pharmacological targets are BCR-ABL, PDGFR and cKIT. Imatinib has also been shown to interact with 296 proteins according to STITCH. Moreover, its gene expression signature is included in the Connectivity Map.

The resulting network is shown in Figure 4.3. Only 22 out of the 296 known targets were conserved in the solution. The method in an attempt to minimize the size of the network, conserved only the nodes that are required to propagate the signal from the drug targets to the differentially expressed genes. In this specific case, the observed gene expressions can be explained using only the 22 targets, thus the remaining drug targets were removed.

The transcription factors conserved in the Imatinib specific network were chosen



**Figure 4.3: Imatinib's MoA.** Nodes in green correspond to drug targets, nodes in yellow correspond to transcription factors. The black rings around the nodes indicate that the corresponding proteins are up-regulated and the red rings indicate that the corresponding proteins are down-regulated. The differentially expressed genes upon perturbation with Imatinib are not shown in the network for the sake of clarity, however, there are differentially expressed genes downstream of all the transcription factors in the network and their differential expression has the same sign as the regulation of the respective transcription factor.

in a similar way. The method conserved only the transcription factors that were required to propagate the signal to the differentially expressed genes. For example, NFKB1 was conserved with a positive sign, represented with a black asterisk in Figure 4.3, because downstream of NFKB1 there are 2 genes that are over-expressed upon perturbation with Imatinib (CFLAR and PIK3CD). Since NFKB1 is not one of the targets of Imatinib, the algorithm also conserved PIK3R1, which is known to interact with Imatinib, to activate NFKB1 via the  $PIK3R1 \rightarrow NFKB1$  interaction. Moreover, the  $NFKB1 \rightarrow FOS$  interaction was conserved to activate FOS inducing the expression of ETV5 and TNRC6B genes. NFKB1 also activates RARA. RARA serves to express the NCOA2 gene that according to the connectivity Map is over-expressed upon perturbation with Imatinib. FOS also interacts with MTOR and from there activates RELA facilitating the expression of PPARA and MMP14 genes.

Even though the reaction  $FOS \rightarrow MTOR$  is not necessary to activate MTOR, since MTOR is one of Imatinib targets, this reaction was present in the PKN and its presence cannot be disproved based on the experimental data. Moreover, the ILP minimizes the number of included nodes, not the number of reactions, thus the reaction  $FOS \rightarrow MTOR$  was conserved in the solution.

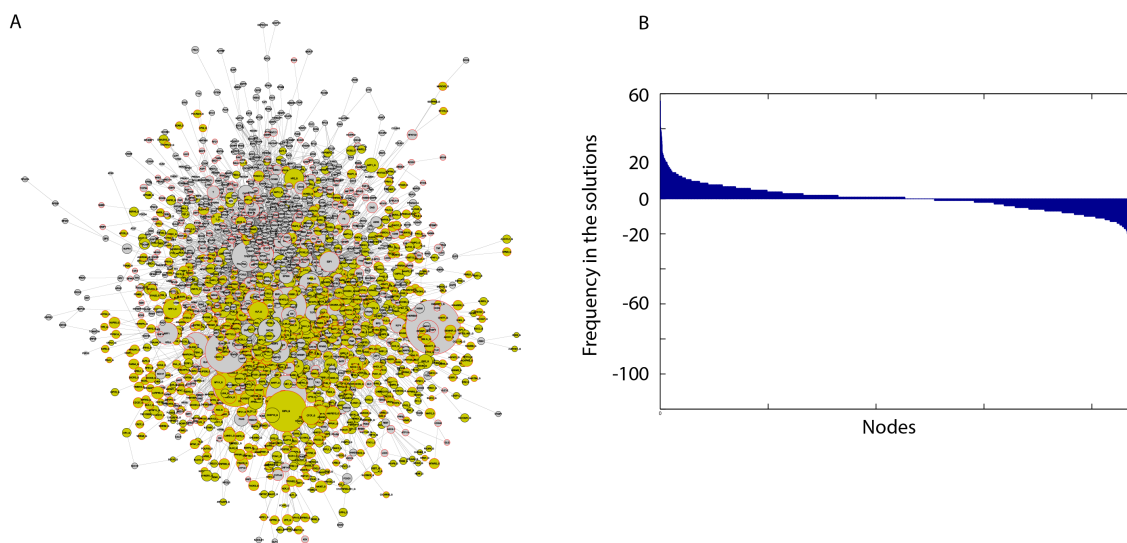
Similar logic applies to the down-regulated nodes in the Imatinib specific network. For example the transcription factor MYC was conserved with a negative sign, shown with red asterisk in Figure 4.3, because 12 genes downstream of MYC are under-expressed upon perturbation with Imatinib. MYC is not one of Imatinib targets, thus signal has to originate from another target upstream of MYC, such as MAPK14 (P38 protein). Down-regulation of MAPK14 also leads to the down-regulation of STAT3, CREB1, MEF2A and JUN, all of which are transcription factors and have downstream genes that are down-regulated upon perturbation with Imatinib. There are some interactions that appear to be redundant, for example the downregulation of MAPK14 from JAK1 and RAF1, however, because both proteins are down-regulated directly from Imatinib and there is an interaction between them in the PKN, the ILP cannot disprove the presence of that reaction, it was thus conserved in the solution. The rest of the nodes and interactions are justified in a similar way.

#### 4.2.4 Construction of the DILD network

All the drug specific signaling networks were pooled together to obtain a signaling network that best captures the molecular mechanisms underlying drug induced lung injury i.e. DILD network. The network is shown in Figure 4.4. It includes a total of 2197 nodes and 6480 reactions. In the same Figure, an analytic showing the significance of the included nodes is plotted. The nodes of the network correspond to different coordinates on the x-axis and the y-axis corresponds to the number of drug specific pathways where each node is either up- or down-regulated. Consistently up-regulated nodes at the proteomic level are placed on the left of the figure, while down-regulated nodes are placed on the right.

Even though there is significant drug to drug variability in the signaling pathways, there are nodes that are consistently up- or down-regulated, and the signaling processes related to these nodes may play a key role in drug induced lung injury. The network modules consisting of the strongly up- and down-regulated nodes, defined as nodes present in 5 or more of the drug specific pathways, were extracted from the DILD network and plotted separately in Figures 4.5 and 4.6 respectively.

As shown in the consistently up-regulated network module of Figure 4.5, a number of proteins related to DNA damage, apoptotic signaling, stress response and inflammation are present. For example TP53, CASP3, BCL2, BAX, CASP6, BCL2L1, CASP8, CASP9, BID, PARP1, CFLAR, GADD45A, FASLG, DDIT3, NFKB1, ATF2, ATF4, TNFRSF10A, TNFRSF10B, TNFAIP3, RIPK2, HSPD1, HSP90AA1,

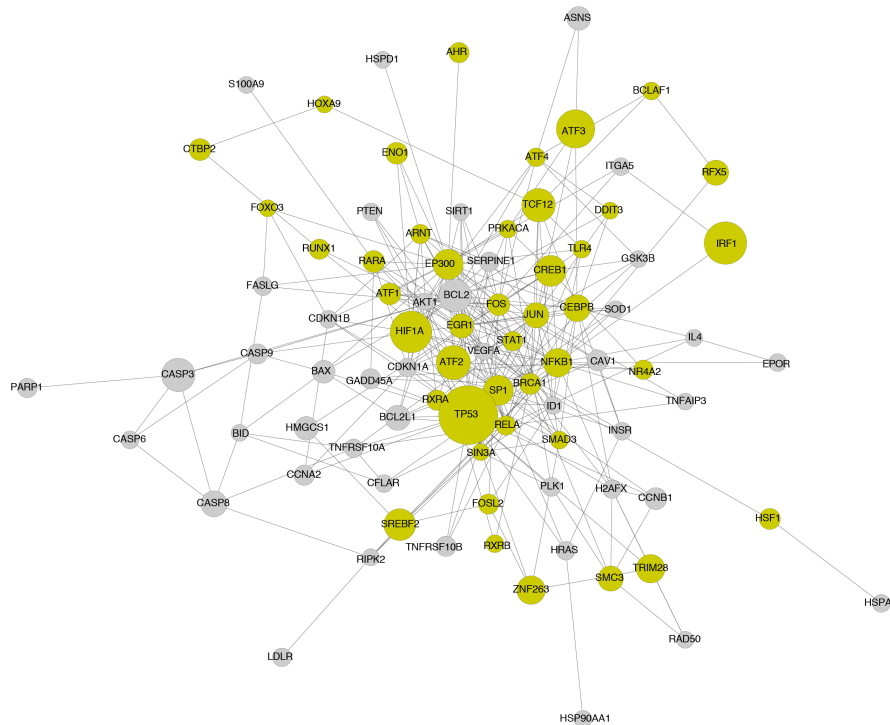


**Figure 4.4: DILD network and analytics.** (A). DILD network. It includes a total of 2,197 nodes and 6,480 reactions. Yellow nodes represent differentially expressed genes and grey nodes represent signaling proteins including drug targets and transcription factors. The size of the nodes corresponds to the number of solutions where this node is active. Thus, most significant nodes are plotted bigger. (B). An analytic showing the significance of the included nodes. The nodes of the network correspond to different coordinates of the x-axis and the y-axis corresponds to the number of drug specific pathways where each node is either up- or down-regulated. Consistently up-regulated nodes are placed on the left of the figure, while down-regulated nodes are placed on the right. There are nodes that are consistently up- or down-regulated, and the signaling processes related to these nodes may play a key role in drug induced lung injury.

HSF1, HSPA6, IFNG, HIF1A and PTEN. Moreover, proteins with a broad role in signaling including JUN, CREB and FOS are present. The above findings are expected and are in good accordance with the Gene Ontology enrichment analysis applied on the differentially expressed genes, as discussed in Section 4.2.2, where the list of over-expressed genes was highly enriched in biological processes related to cell death and apoptosis (see Table B.1).

The proposed algorithm leveraged the differential gene expressions and prior knowledge of protein connectivity and transcription regulation and identified the signaling networks underlying the observed gene expression signatures. Since the gene expression data revealed a strong correlation with biological processes related to cell death and apoptosis, the signaling networks that yield this response are DNA damage and apoptotic signaling networks as shown in Figure 4.5. Pro-apoptotic and response to DNA damage pathways are also known to be implicated in various forms of lung injury [82]. The agreement of the signaling networks of Figure 4.5 with the biological processes related to the differentially expressed genes (Table B.1), also validates that the breaching of the signaling and gene expression





**Figure 4.5: Consistently up-regulated module of the DILD network.** Including only the nodes that are up-regulated in five or more of the drug specific signaling networks. Transcription factors are plotted in yellow. Differentially expressed genes have been omitted from the figures for the sake of clarity. The size of the nodes corresponds to the number of drug-specific pathways where the respective node is up-regulated.

levels via the layer of transcription factors is accurate. Apart from the DNA damage and apoptosis pathways, proteins related to calcium signaling are present, such as FASLG, FOS, IL4 and JUN, which is in agreement with the literature where hypercalcemic activity has been observed in lung injury [83].

Since the ILP algorithm is agnostic to the biological function of the included proteins, and only uses the experimental data to identify subsets of the PKN that are functional in the specific biological context, it is expected that under-reported proteins appear. These may constitute novel findings, or they may be an artifact of the PKN structure, or the prior knowledge of transcription regulation. For example, EP300 is a transcription factor and it facilitates the expression of 108 differentially expressed genes in the DILD network. Moreover, it takes part in signaling and interacts with 170 proteins in the DILD network, thus, it appears to have a central role. The 108 differentially expressed genes connected to EP300 may also be expressed by another TF that is not included in the PKN, thus the ILP is forced to use the EP300 protein to fit these gene expressions, even though this is not the true mechanism.

These advantages and pitfalls exist in all unbiased approaches. In this context,

EP300 is known to play a role in the WNT/ $\beta$ -catenin pathway which is found to induce IL1B expression and be implicated in Interstitial Pulmonary Fibrosis, one of the lung injury phenotypes [84]. Moreover, a GO enrichment analysis on the target genes of EP300 found over-representation of programmed cell death and other apoptotic processes. In particular 42 genes related to apoptosis are expressed by EP300, which implies its potential role in pro-apoptotic response and consequently drug induced lung injury.

In Figure 4.6, the network module of the consistently down-regulated proteins is shown. A number of proteins related to pro-growth and pro-survival pathways are present, such as MYC, E2F1, E2F6, CDK1, RAF1, SRF, RPS6A3, and MAPK7. This is in good accordance with the Gene Ontology enrichment analysis performed on the differentially expressed genes, discussed above, where the list of under-expressed genes was highly enriched in biosynthetic and metabolic processes, and also processes affecting the cell cycle. Here, the signaling networks underlying these biological processes are shown, as these were computed by the ILP algorithm based on the gene expression data. The under-expression of pro-growth, pro-survival and cell cycle pathways in lung injury has also been reported in literature [85].

Apart from the major pro-growth pathways, TOP2A (DNA topoisomerase 2A) is also consistently down-regulated. This is expected as a large number of the lung toxic drugs are DNA inhibitors and target TOP2A. Finally, a number of proteins related to female hormone signaling are present in the network (ESR1, ESR2). This is in good accordance with the literature where estradiol and other estrogen receptor agonists are found to ameliorate the symptoms and protect against lung injury [86].

A number of unrelated proteins are also present in Figure 4.6, such as MEF2A and GABPA. MEF2A mediates cellular functions mostly in the skeletal and cardiac muscle development. However, it is also found to play a diverse role in controlling cell growth survival and apoptosis via the MAPK14 (P38) signaling network. In good accordance with the literature, in Figure 4.6, MEF2A is activated by MAPK14. Moreover, MEF2A facilitates the expression of 28 genes, and also participates in signaling by interacting with 19 other proteins. GABPA is also found to play a significant role in the DILD network, expressing 40 genes and interacting with 8 proteins.

### 4.2.5 Drug Candidates for DILD

This section demonstrates the predictive power of the proposed method and the biological relevance of its predictions, by leveraging the DILD network to identify potential drugs for repositioning to treat DILD, using the 1109 non-toxic drug of the Connectivity Map.

First, the targets and gene expression profiles of the non-toxic drugs were extracted from STICH and cMAP respectively. Then, drug specific signaling networks

were computed for all drugs whose targets overlap with the DILD network. Finally, the drugs were ranked according to how much their pathways disrupt the DILD network. A drug was considered to disrupt the DILD network if its signaling network up-regulates proteins that are down-regulated in the network or down-regulates proteins that are up-regulated in the network. The top 40, most highly ranked drugs are listed in the Appendix B.1 together with their indication and relevant information supporting their usability for treating DILD. The drugs at the top of the list are in good accordance with our previous reports and have also been shown to ameliorate the symptoms of lung injury.

For example, ciclosporin is an immunosuppressant drug widely used in organ transplantation. It reduces the activity of the immune system by interfering with the activity of T cells. It is also effective in rheumatoid arthritis and severe psoriasis, both of them are auto-immune disorders with a strong inflammatory component. Moreover, it has been shown to be an effective treatment for interstitial lung disease of unknown etiology [87]. Its signaling pathway is shown in Figure 4.7. There were proteins strongly up-regulated in the DILD network, and implicated in apoptotic and inflammatory processes, that are down-regulated by ciclosporin and vice versa. For example, the proteins TP53, TRIM28, RELA, HIF1A, FOS and JUN that were consistently up-regulated upon perturbation with the lung toxic drugs, they are down-regulated upon perturbation with ciclosporin. Moreover, the proteins RPS6KA3 and SRF, related to cell cycle and consistently down-regulated upon perturbation with the toxic compounds, are up-regulated upon perturbation with ciclosporin. In total, ciclosporin up-regulated 13 proteins that were down-regulated in the DILD network: CCNB2, ESR2, NFE2L2, NFYA, RAD21, RB1, REL, RPS6KA3, SRF, TAF1, TFDPI, YBX1, YY1, and down-regulated 20 proteins that were up-regulated in the DILD network: AURKA, BHLHE40, BUB1B, CCNA2, CCNG2, CTBP2, FOS, HBP1, HIF1A, JUN, POLR2E, RAD50, RELA, RUNX1, SMC3, STAT1, TCF12, TP53, TP53BP1, and TRIM28, indicating a potential disease modifying action.

Apart from ciclosporin, the flavonols quercetin and genistein, ranked third and sixth in the list, have strong anti-inflammatory action and have been shown to be beneficial in pulmonary disease. The protective effect of flavonoids in lung injury has been reported previously [88]. Resveratrol (ranked 4<sup>th</sup>) is another plant extract that has been shown to alleviate COPD injury in rats [89]. Tretinoin, ranked second in the list, is another immunosuppressant, and has been shown to ameliorate the symptoms of oxygen induced lung injury in the newborn rat [90]. However, it has also been reported in FactMed to have caused traumatic lung injury in at least one patient out of 957 reports of any other side effects of tretinoin. Of the 933 physicians that expressed their opinion on the report, 295 were highly suspicious of tretinoin as the cause of the incident.

Apart from the immunosuppressants and other anti-inflammatory drugs in the list, the estrogen diethylstilbestrol is also present (ranked 7<sup>th</sup>). Even though diethylstilbestrol has not been shown to treat DILD, it is expected to upregulate ESR1

and ESR2, that according to our predictions are consistently down-regulated in DILD (see Figure 4.6). Moreover, estradiol and other estrogen receptor agonists are found to ameliorate the symptoms and protect against lung injury [91], implying diethylstilbestrol could be a novel finding of this analysis [92]. In similar fashion, dinoprostone (prostaglandin E2), ranked 16<sup>th</sup> in the list, has been found to significantly disrupt the DILD network, which is in good accordance with our previous analysis where we identified that Prostaglandin-endoperoxide synthase 2 (also known as COX-2) is consistently down-regulated in lung injury. In addition, dinoprostone has been found to protect against lung fibrosis [93], [94]. Similar observations can be made for other drugs in the list.

### 4.2.6 Validations

For this section, an independent statistical method is applied to validate the DILD network. To this end, the GUIDE algorithm is used [95], [96]. GUIDE is an algorithm that builds a classification and regression tree model to predict the values of one or more response variables ( $Y_1, Y_2, \dots$ ) from the values of the predictor variables ( $X_1, X_2, \dots$ ). It can also produce an importance score for each  $X_i$ . Classification and regression trees were also shown to predict oral absorption in humans based on predictors of chemical substructures [97].

The drug targets from STITCH were used as predictor variables ( $X_i$ ) and the differential gene expressions for each drug were used as response variables ( $Y_i$ ). GUIDE was used to construct regression trees, modeling how drug targets correlate statistically with the differential gene expressions. Since GUIDE is agnostic to the protein connectivity in the PKN, it cannot construct functional mechanistic pathways, but can produce scores for the drug targets that represent their importance in predicting the observed gene expression signatures. Drug targets with importance score greater than 1 are considered significant and their overlap with the nodes in the DILD network is computed using the hypergeometric CDF.

Of the 78 drug targets identified by the GUIDE algorithm to be important for predicting differential gene expression, that are also present in the PKN, 71 of them have been conserved by the ILP algorithm in the DILD network. The DILD network includes 1150 signaling nodes, out of 2585 signaling nodes originally included in the PKN. Thus, the p-value of the enrichment of the DILD network in the drug targets identified by GUIDE, as calculated using the hypergeometric CDF to be  $1.2466 \times 10^{-19}$ . The significant overlap between the proposed method and GUIDE predictions, further establishes the statistical significance of the method results.

## 4.3 Conclusions

In this work, an approach is presented based on an ILP formulation that combines available gene expression data with prior knowledge of protein connectivity and

transcription regulation, in the form of a prior knowledge network, to identify subsets of the network that appear to be functional based on the data.

As a case study the identification of the MoA of drugs that are known to induce lung injury is addressed, and a signaling network is constructed spanning across both the signaling and gene expression levels, through a layer of transcription factors, that elucidates the signaling mechanisms underlying drug induced lung injury (DILD). Manual inspection of the DILD network revealed that pathways related to DNA damage, inflammation and apoptosis are consistently up-regulated, while pathways related to cell cycle are down-regulated. This is in good accordance with the GO enrichment analysis performed on the differentially expressed genes upon perturbation with the toxic compounds, that uncovered biological processes related to cell death highly enriched in the over-expressed genes, and processes related to cell cycle highly enriched in the under-expressed genes. This indicates the successful breach of the signaling and genomic levels through the layer of transcription factors. Pathways related to DNA damage, inflammation and apoptosis have also been reported in the literature to be implicated in lung injury [82], while pathways related to cell cycle have been reported to be under-expressed [85].

Moreover, an independent classification and regression trees (CART) analysis was performed using the GUIDE algorithm, that identified the most important drug targets for predicting the observed gene expressions. The CART predictions significantly overlapped with the ILP predictions, thus supporting the relevance of the drug specific signaling networks.

Finally, to demonstrate its usability, the DILD network produced was leveraged to identify suitable drugs for repositioning to treat lung injury. To this end, drugs whose targets overlap with the DILD network were considered, and their signaling networks were constructed via the same method. The drugs were ranked according to how much their pathways disrupt the DILD network, indicating a potential disease modifying action. Most drugs at the top of the list are good candidates for treating DILD. They have strong anti-inflammatory action and many of them have also been shown to ameliorate the symptoms and/or protect against lung injury. Nevertheless, before claiming that the candidate drugs treat DILD, follow up experiments should be performed to better characterize their action.

A key feature of our proposed method for reconstructing signaling networks based on gene expression data, and fundamental for its predictive power, is working with directed, signed signaling reactions, rather than undirected, unsigned PPIs and the ability of our ILP algorithm to efficiently handle this information. When working with PPI networks, the lack of directionality and sign of the interactions makes it difficult to interpret the results. In most cases connectivity metrics are employed such as node centrality, betweenness, communicability, etc. to evaluate the significance of every node in the network. However, these metrics fail to capture the mechanistic component of the signal flow.

## 4.4 Methods

### 4.4.1 ILP formulation

The ILP formulation used for this study is similar in nature to the reverse causal reasoning formulation described in Section 2.2.2. Slight adjustments were made in order to facilitate the fitting process due to the size of the PKN and the number of networks fitted. In particular, all the  $y$  variables, modeling the inclusion of arcs in the final network, were set to 1 and the optimization was carried over only over the power-set of the nodes.

Hard sparsity constraints were imposed on the  $p$  variables, modeling the candidates for roots of the drug networks, to include only known or suspected drug targets from STITCH. Moreover, soft sparsity constraints were imposed on all nodes as weights. The trade-off between accuracy and sparsity was set to 5:1, i.e. removing 5 nodes was valued equally to allowing 1 mismatch.

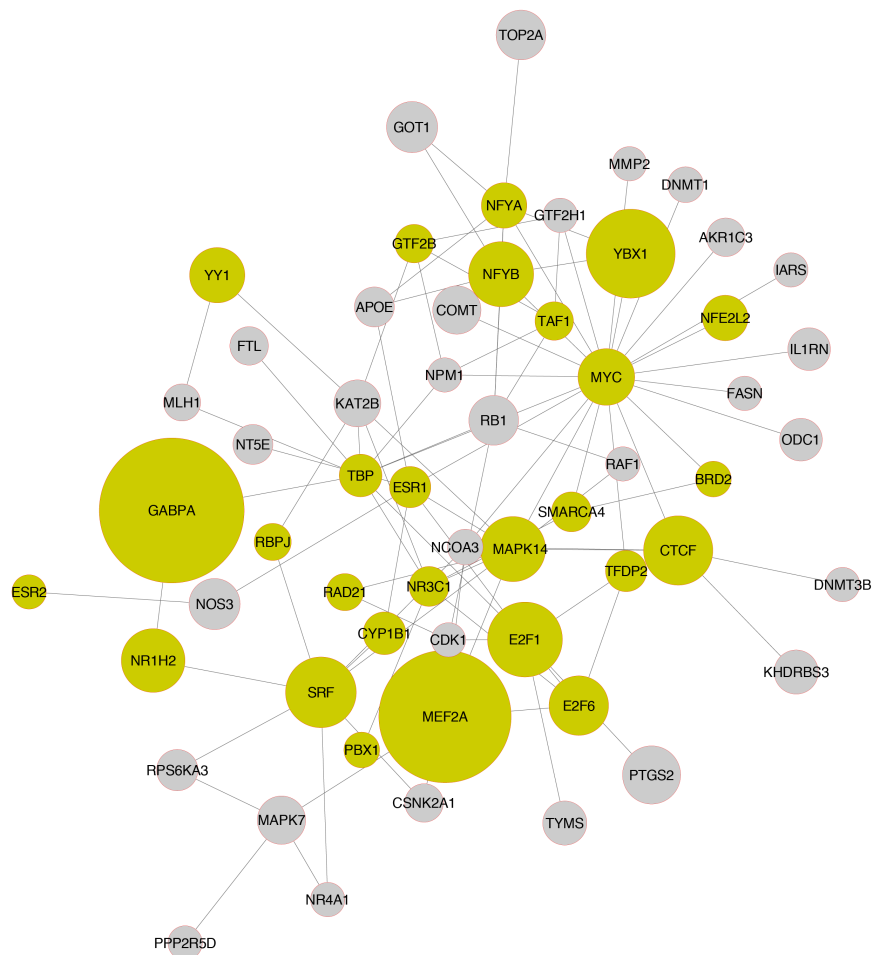
Finally, because there were not multiple experiments per network (one-shot learning) the distinction between activation ( $x$ ) and inclusion ( $v$ ) in the network did not make sense so the 2 variables were merged.

### 4.4.2 Construction of Prior Knowledge Network

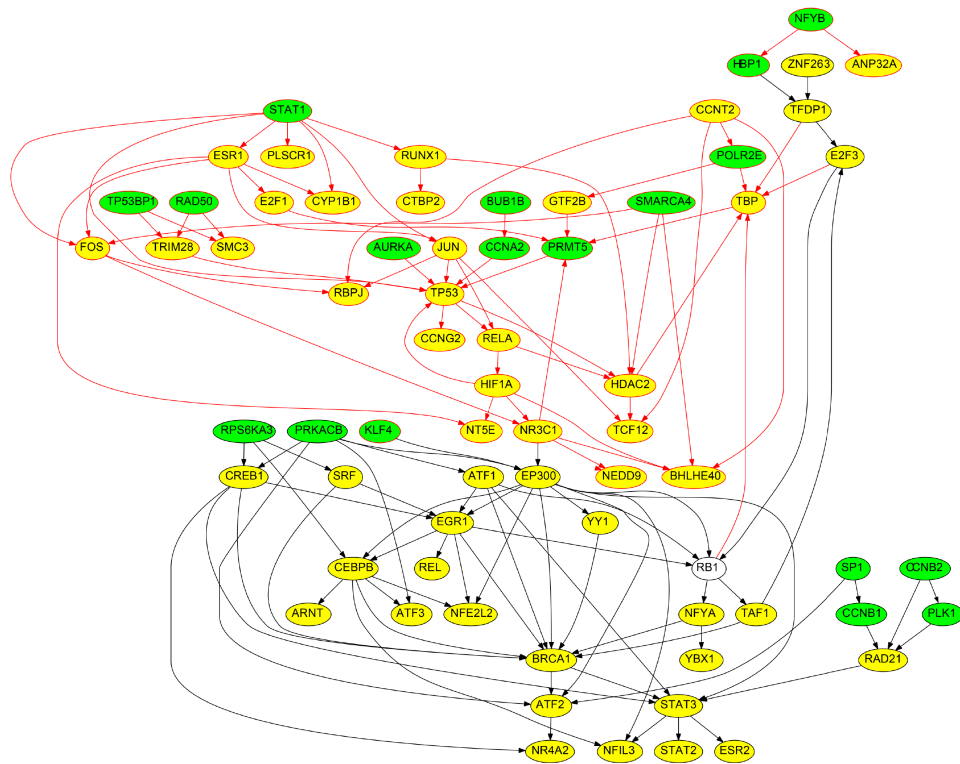
The Prior Knowledge Network (PKN) is used to represent prior knowledge of protein connectivity and transcription regulation and serves as a scaffold for the ILP algorithm presented above. It was constructed by merging the Functional Interaction Network (FIN) by Reactome [76] and information on transcription regulation in the form of set of transcription factor regulons, i.e. sets of targeted genes, assembled from public available resources, such as ChEA [31], Transfac [32], and Jaspar [33].

Before using the FIN other networks were considered including the Human Signaling Network [98], Signalink [99], and the network by [25]. The FIN was chosen because it offered the best coverage of the transcription factors for which there is an available regulon, while being the sparsest of all, facilitating the performance of the ILP algorithm. The FIN consists of 209988 functional interactions between 10956 proteins. The regulons implement 16043 interactions between 153 transcription factors and 12376 target genes.

Before merging, undirected and unsigned interactions were removed, as well as interactions that were predicted computationally without experimental validation. Interactions between proteins that appear not to be expressed, or take part in the signaling processes, in the lung tissue were also removed. The final PKN included only nodes that appear in the lung-specific network published by Guan et al. [100] The resulting PKN spans across the proteomic and genomic levels going through a layer of transcription factors and includes a total of 64801 reactions between 2585 signaling proteins and 12376 genes.



**Figure 4.6: Consistently down-regulated module of DILD network.** Module of the DILD network including only the nodes that are down-regulated in five or more of the drug specific signaling networks. Transcription factors are plotted in yellow. Differentially expressed genes have been omitted from the figures for the sake of clarity. The size of the nodes corresponds to the number of drug-specific pathways where the respective node is down-regulated.



**Figure 4.7: Ciclosporin MoA.** Nodes in green correspond to drug targets, nodes in yellow correspond to transcription factors. The black rings around the nodes indicate that the corresponding proteins are up-regulated and the red rings indicate that the corresponding proteins are down-regulated. The differentially expressed genes upon perturbation with ciclosporin are not shown in the network for the sake of clarity, however, there are differentially expressed genes downstream of all the transcription factors in the network and their differential expression has the same sign as the regulation of the respective transcription factor.



# Chapter 5

## Anthrax Infection

This chapter presents the work published in [101]. The chapter follows the original work closely, with slight changes to ensure a uniform style for this thesis.

This was a follow up study of the work presented in Chapter 4. The same method was used to identify the mode of action of pulmonary lung infection, and non lung-toxic drugs were screened and prioritized as candidates for repositioning, with an approach similar to the one described in Section 4.2.5. The results were validated with a literature review. The author was supported by an ORISE fellowship for this work.

### 5.1 Introduction

The potential use of *B. anthracis* (Gram positive) as a weapon of bioterrorism, combined with recent outbreaks and isolated cases of anthrax infection in the US [102], [103] and Europe [104], has focused the developed world's attention on this lethal bacterium. Notably, the mortality rate during invasive anthrax infection and the development of shock has been exceptionally high when compared to more commonly encountered bacteria [105], [106].

Production of lethal toxin (LT) and edema toxin (ET) is closely associated with the pathogenesis of *B. anthracis* infection [107] and the development of shock [108]. LT and ET are both binary type toxins consisting of protective antigen (the component necessary for host cell uptake of each toxin's toxic moiety through a membrane anthrax toxin receptor identified in human cells [109]) and lethal factor (LF) for LT and edema factor (EF) for ET [108]. Lethal factor is a metalloproteinase [110] which cleaves and inactivates mitogen-activated protein kinase kinases (MAPKK, including MAPKK's 1, 2, 3, 4, 6 and 7) [111], [112] and also activates host cell inflammasome formation [112], [113]. Edema factor has calmodulin dependent adenyl-cyclase activity and increases intracellular cAMP concentrations to high levels [114].

Increasing evidence suggests that both LT and ET target and disrupt the function both of mononuclear and macrophage (mononuclear) cells participating in the

host innate and adaptive immune responses and of endothelial cells maintaining vascular integrity and function. While disruption of macrophages by LT and ET is believed to play an important role in propagating early *B. anthracis* infection, disruption of endothelial cell function likely contributes to the highly resistant shock which can develop in some patients with anthrax.

In particular, for LT evidence suggest that:

- impairs adaptive immunity in dendritic cells [115]
- impairs innate and adaptive immunity, as well as vascular barrier integrity by inhibiting the MAPK pathway [116]
- induces a concentration and time dependent increase in vascular permeability [117]
- activates the host's innate immune responses in alveolar macrophages [118]
- damages endothelial barrier and vascular integrity by decreasing the activity of p38 and MK2 and reduced phosphorylation of HSP27 [119]
- causes apoptosis and reduces production of inflammatory cytokines in PBMCs [120]

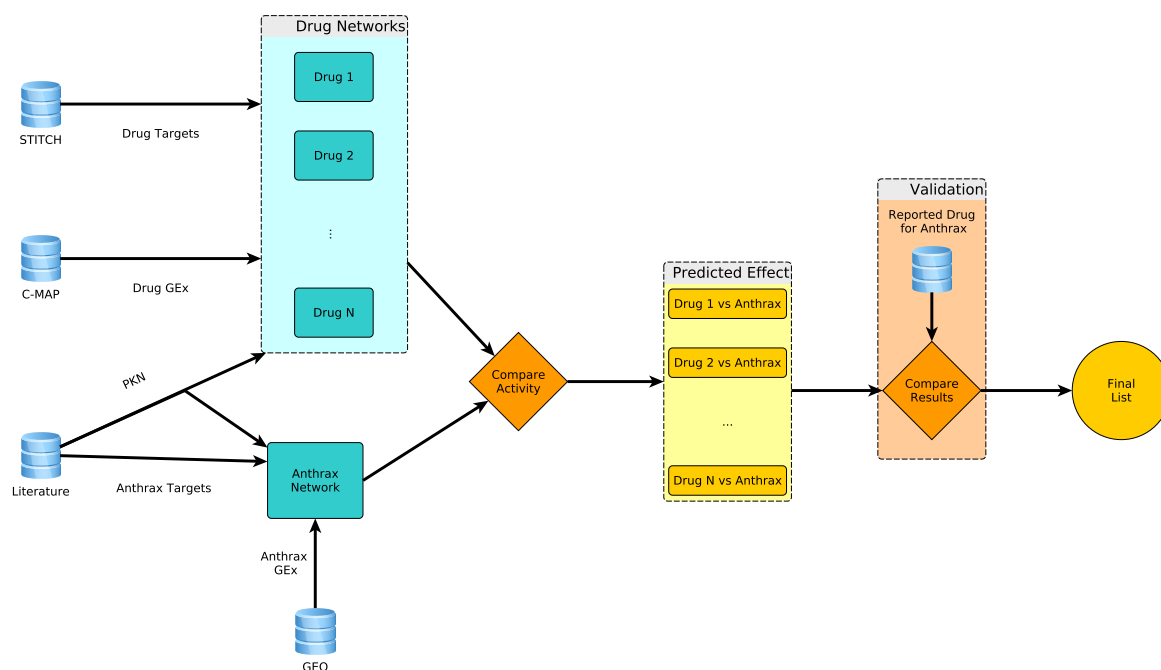
and for ET:

- induces cAMP accumulation and damages antimicrobial activity in human monocytes [121]
- impairs the immune response in cooperation with LT [122]
- inhibits T-cell activation in mice [123]
- inhibits endothelial cell chemotaxis via Epac [124]
- suppresses human macrophage phagocytosis by deregulating cAMP-dependent kinase pathways [125]
- induces trans-endothelial cell macro-aperture (TEM) tunnels by affecting cAMP signaling [126]

Identifying effective agents that are capable of blocking the pathogenic effects of LT and ET on mononuclear and endothelial cell function and could be applied clinically, would improve the management of this highly lethal infection. To achieve this goal, the present study was designed to utilize reported data regarding the known effects of LT and ET on intracellular molecular targets in either mononuclear or endothelial cells, as well as the mechanisms of action of drugs to identify the already approved drugs that might serve as therapies for anthrax.

The approach of Chapter 4 was employed to computationally construct the mode of action (signaling network) for each anthrax toxin and for each drug by

utilizing its molecular targets and differentially expressed genes, as well as the prior knowledge of protein interactions. We mined published studies and GEO [36] for the gene expression data generated from human macrophages or peripheral blood mononuclear cells (PBMCs) following exposure to *B. anthracis* spores or to lethal toxin or edema toxin. For drugs, we used the differentially expressed genes from Connectivity Map [35]. We then computationally identified and scored the drugs for their ability to reverse the actions of anthrax lethal and edema toxins. To validate our results, we searched the literature for evidences to determine whether the top 10 and bottom 10 drugs among the 474 drugs/compounds computationally ranked were supported by literature reports, see Figure 5.1 for workflow.

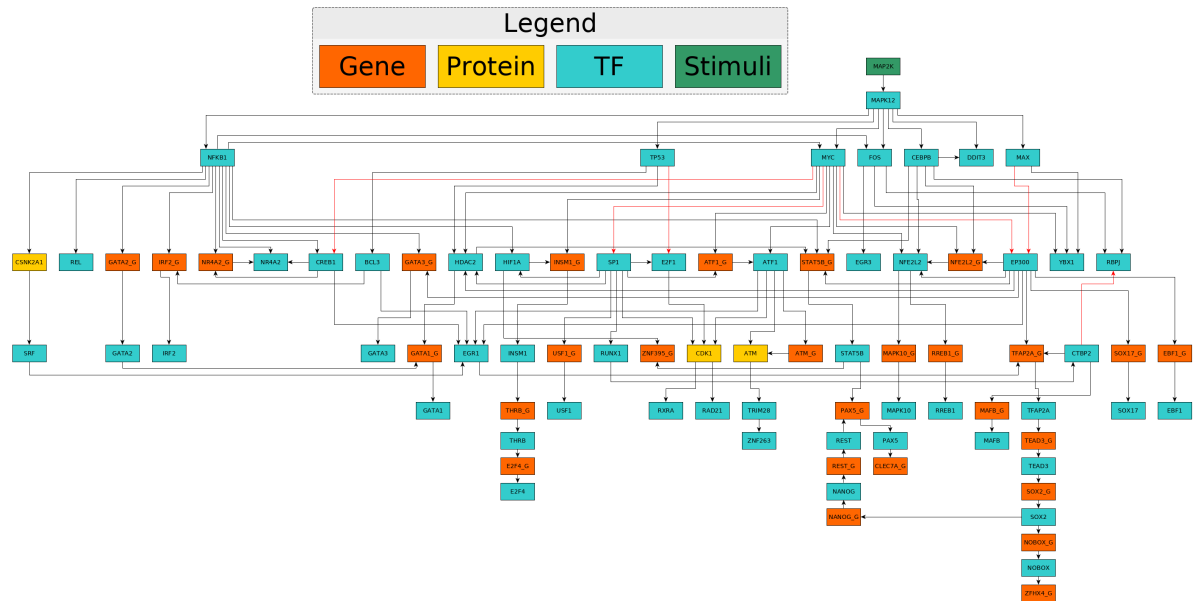


**Figure 5.1: Drug repositioning for Anthrax infection workflow.** First, construction of anthrax networks and individual drug networks. Second, computationally scoring individual drugs by computing the distance between anthrax networks and individual drug networks. Third validation of computed scores and rankings by referencing the literature. Gene expression data from GEO were used, prior knowledge network of protein interaction and molecular targets of anthrax toxins in host cells for anthrax networks. Individual drugs' networks were constructed with gene expression data from Connectivity Map, prior knowledge network of proteins and their respective targets from STITCH

## 5.2 Analysis

### 5.2.1 Anthrax Functional Network

To construct the signaling network describing anthrax toxicity, the method described in Section 4.4.1 was employed, where genes that were differentially expressed upon stimulation with anthrax toxins were combined with the PKN from Chapter 4 and molecular targets of anthrax toxins. The “core” of the resulting network is shown in Figure 5.2.



**Figure 5.2:** The anthrax network of lethal toxin and edema toxin constructed to reflect the modes of action of anthrax in human macrophages upon exposure to anthrax spores. For spatial convenience, the network doesn't include the genes that were used to fit it.

To identify differentially expressed genes, 6 GEO studies were collected, in which human cells were treated with anthrax of any form. These were:

1. GSE14390: Alveolar macrophages treated with anthrax spores
2. GSE34407: Peripheral monocytes treated with lethal toxin
3. GSE12131: Umbilical vein endothelial cells treated with lethal toxin
4. GSE17777: Microvascular endothelial cells treated with edema toxin
5. GSE4478: Peripheral monocytes treated with lethal toxin
6. GSE12533: Peripheral monocytes treated with protective antigen

The datasets used were by Dozmorov et al. [127], where HAM cell were treated with *B. anthracis* spore, and by Chauncey et al. [128] where PBMCs were treated directly with lethal toxin. The genes reported as differentially expressed by the authors were used directly (no meta-analysis) and all other included in the studies but not reported were considered to be at the basal state.

The HAM dataset consisted of 280 significantly perturbed genes in total, 205 up and 75 down regulated, while for the PBMC dataset consisted of 407, 309 up and 98 down regulated. The data are shown in the Supplementary Material of the published version. Genes that were not present in the PKN or were not reachable by any known target of anthrax toxins were eventually filtered out. The final lists for the two datasets consisted of 132 genes for the HAM dataset, 101 up and 31 down, and 320 genes for PBMC, 45 up and 275 down.

Anthrax targets were selected based on the studies cited in Section 5.1. For the HAM cells treated with anthrax spores, the targets of both LT and ET were included while for the PBMCs only those of ET. As mentioned above, LT is shown to cleave MAP2Ks, but it also targets NLRP1. NOD-like receptor (NLRP1) was reportedly targeted and cleaved by LT, consequently activating inflammasome in rodents [129]. Though the role of NLRP1 in activating the inflammasome in human cells is not clearly defined [130], NLRP1 was included because NLRP1 inflammasome activity is involved in human diseases [131]. ET's "nominal" target is cMAP which is not included in the PKN. However, cMAP regulates protein kinase A (PKA) [108] and Epac (a Rap1 guanine-nucleotide-exchange factor [132]), so these were used instead as targets.

### 5.2.2 Repurposing Candidates

This study focuses on the human alveolar macrophage (HAM) dataset because *B. anthracis* disables the host's immune defense system and is most deadly if inhaled. For the reference and comparison, the combined Peripheral Blood Monocytes (PBMC) datasets were analysed as well.

To identify drug candidates for repurposing, the drug networks developed for DILD in Chapter 4 were used. The development process is described in Section 4.2.2. Out of the 652 drug networks only 474 had some overlap with the anthrax network. Each drug's network activity, i.e. the state of its nodes, was compared to that of the anthrax activity. Drugs were ranked according to their similarity with the "reverse" anthrax activity. Reverse activity denotes the activity vector multiplied by  $-1$ . The squared Euclidean distance was used as a comparison metric, in order to penalize synergies, i.e. nodes with the same sign in both anthrax and drug networks. The top and bottom 10 drugs are shown in Tables 5.1 and 5.2 respectively.

## 5.3 Validation

The anthrax network of the HAM cells was compared with the DILD network up and down regulated modules (Figures 4.5 and 4.6) for overlap. Both activities take place in the lung and also cells of the immune system are engaged in activities across all organs. The overlap of the 2 was highly significant, the p-value of the

hypergeometric test was  $1.51 \times 10^{-11}$ . This piece of evidence, further supports the premise that drugs suitable to treat DILD can be used for anthrax infection as well.

The ranking of the drugs was validated with a literature review. In particular, literature reports were mined and referenced for the drugs/compounds that have been shown to have in-vitro effectiveness against anthrax or for relevant biological activities to evaluate our computational rankings of the top 10 and bottom 10 drugs. The main resource used is a report by Sanchez et al [140] in which a large number of known drugs were shown to render anthrax antitoxic protection of cells.

A list of drugs/compounds was compiled that has been reported to possess in-vitro potency against LT or ET-induced toxicity or increased animal survival. The drugs included

- Bepridil, a calcium channel blocker, and isotretinoin, sb-203580, propafenone, h-89, and sb-202190 [140]
- Chloroquine [155]
- Niclosamide [156]
- other calcium channel blockers like Verapamil and Nitrendipine [144]
- Simvastatin and Fluvastatin [133]
- MG-132 [157]
- antibiotics [158]–[160] like
  - Doxycycline
  - Ciprofloxacin
  - Chloramphenicol
  - Ampicillin
  - Penicillin
  - Clindamycin
  - Imipenem
  - Vancomycin
  - Clarithromycin
  - Rifampicin
  - Neomycin

Several other calcium channel blockers that were reported to protect cells from lethal toxin [140] were not on the list of 474 drugs/compounds that were ranked owing to lack of gene expression data.

Sanchez et al[140] showed a potency rank order of

1. Bepridil
2. Nicardipine
3. SB-203580, SB-202190
4. Isotretinoin, Propafenone, Retinoic acid, Tretinoin, H-89

Bepridil, SB-203580, SB-202190, Isotretinoin, Propafenone, and H-89 were among the 474 drugs/compounds considered, with all of them ranked top 100 except h-89

and sb-202190 that were ranked outside the top 100 drugs. In the proposed ranking their order is:

1. Bepridil
2. Isotretinoin
3. sb-203580
4. Propafenone
5. h-89
6. sb-202190.

Most interestingly, Bepridil is ranked 8<sup>th</sup> and simvastatin 4<sup>th</sup> out of the 474 drugs/compounds, their top rankings were in agreement with the reports that both drugs were potent in increasing cell survival following exposure to LT [133], [140].

In Table 5.1, next to the top 10 candidate drugs a possible mode of action for alleviating the effects of anthrax infection is listed as mined by the literature.

On the other hand, the modes of action of the bottom 10 drugs listed in Table 5.2 include activities similar to those of anthrax toxins, including reduction of MAPK activities or increase of intracellular level of cAMP. Notable exceptions are Monastrol, Ebselen and Genistei. Monastrol arrested cells in mitosis [147] which is unlikely to provide protection for cells under attack by anthrax toxins. It is not clear why Ebselen and Genistein were placed in the bottom 10. It could be due to an error of the approach or another unknown mechanism.

## 5.4 Discussion

Among the top 10 drugs (Table 5.1, Bepridil and Simvastatin have been reported to significantly increase in-vitro cell viability; the other 8 drugs had relevant biological evidence suggestive of their protective actions against anthrax toxin toxicities. On the other hand, 7 out of the bottom 10 drugs do not have the ability to increase cell survival upon exposure to anthrax toxins, for details see Table 5.2.

Both LT and ET damage endothelium integrity and cardiovascular function leading to fatal shock and tissue injury [161], [162]. Fenofibrate, the top drug candidate, is a cholesterol-modifying drug and activates PPAR $\alpha$  receptors, and could presumably protect the vascular system from anthrax LT and ET toxicities since PPAR $\alpha$  receptors play a key role in endothelial function [163], [164]. Most significantly, PPAR $\alpha$  and  $\gamma$  ligands activate MAPKs [165]. Fenofibrate's ranking indicates that the computational approach revealed its direct effect at counteracting the toxic actions of LT and ET at their molecular targets.

Simvastatin is reportedly a bactericidal, though at a concentration much higher than its therapeutic range observed in humans when used for lowering cholesterol [166]. Several statins increased macrophage viability upon exposure to LT through inhibiting Rho GTPase and activating MAPK signaling pathways [133]. Although simvastatin was ranked in the top 10, rankings of other statins beyond 300 by our approach require further investigations.

$\text{Ca}^{2+}$  is required for the toxicity of lethal toxin [144], and calcium channel blockers including Verapamil and Nitrendipine at 100  $\mu\text{M}$  and Bepridil at a concentration between 0.125 and 12.5  $\mu\text{M}$  increased cell survival when exposed to LT. Bepridil was reported to be more potent than Verapamil and Nitrendipine which is in agreement with the suggested ranking. Mephenytoin, an active analogue of Phenytoin that inhibited  $\text{Ca}^{2+}$  transport into the cells was also ranked high (6<sup>th</sup>). Finally, Ifosfamide, an antineoplastic agent, caused increased renal excretion of  $\text{Ca}^{2+}$  [146], which could lead to depletion of  $\text{Ca}^{2+}$ .

Cotinine was ranked among the top 10; such ranking seemed to be consistent with the reports that nicotine and its active metabolite, cotinine, activated the mitogen-activated protein kinase pathway [137], [138], an action opposite to that of LT [111], [167]. Cotinine reportedly increased intracellular  $\text{Ca}^{2+}$  concentration [168], which would contradict what has been observed for calcium blockers mentioned above, and could not have benefited its anti-anthrax toxin activities. However, its activation of MAPK pathway is right at the molecular targets of LT. Ranking Cotinine in the top 10 implied that immediate activation of MAPKs was important, which is consistent with the basis of our computational scoring of a drug/compound's anti-anthrax toxin ability.

Sotalol decreased cAMP, which is contrary to the toxic action of ET. Beta blockers including Betaxolol, Bisoprolol, Carvedilol, Metoprolol, Nadolol, Propranolol, Sotalol, and Timolol decreased cAMP accumulation [145]. Among these drugs, Sotalol, Metoprolol and Propranolol were among the 474 drugs/compounds ranked in this study. Sotalol was reported as a more potent reducer of the rate of cAMP accumulation than Metoprolol and Propranolol which were equal [145]. Since ET increased cAMP, the effects of these beta blockers on reducing cAMP accumulation could be beneficial to cell survival. Interestingly, in the proposed ranking Sotalol is 9<sup>th</sup>, Metoprolol 13<sup>th</sup> and Propranolol 109<sup>th</sup>. Further studies would be needed to understand whether they have other mechanisms that can be beneficial.

Dihydroergotamine is a 5-HT 1B/1D agonist [135]; stimulation of 5-HT 1B/1D receptors activated MAPK and reduced cAMP level [136], shedding light on a possible mechanism of dihydroergotamine in antagonizing anthrax toxins. Our ranking of Meloquine in the top 10 could be linked to the fact that it is an analogue of chloroquine and chloroquine reportedly protected cells from LT toxicity [155]. Amantadine reportedly cancelled activation of p38/MAPK [139]. Since p38/MAP kinase inhibitors, like SB-203580 and SB-202190, protected cells from LT-mediated cytotoxicity [140]; such action of Amantadine can render protective effects like p38/MAP kinase inhibitors.

On the opposite side, out of the 474 drugs/compounds, the bottom 10 compounds included Monastrol, Colforsin, Berberine, Withaferin-a, Arecoline, Ebselen, and Genisten.

As highlighted in Table 5.2, Berberine and Apigenin inhibited MAPK, thereby enhancing LT toxicity. Withaferin-a, Arecoline, and Beclomethasone also have the capability. Withaferin-a, Arecoline, and Beclomethasone reportedly activated



p38 MAPK, an action that is opposite to that of p38 MAPK inhibitors shown to increase cell survival upon exposure to LT [140].

Colforsin reportedly increased intracellular cAMP [148], an action similar to that of ET. Monastrol arrested mitosis [147], likely causing harm to the cells rather than increasing their survival. As for Enilconazole, it is not clear how our computation ranked it in the bottom 10.

Ebselen activated MAPK p44/42 [152] and so did Genisten [153]. Since MAPK p44/42 are downstream of the MAPK pathway; such a downstream activation might not offer any protection resulting in these two compounds being ranked in the bottom 10.

Ciprofloxacin is an antibiotic that has been approved for treating Anthrax disease. The antibiotics in the top 100 drugs included Amoxicillin, Spectinomycin, Ciprofloxacin, Alveparmycin, Amphotericin-b, Doxycycline, Troleandomycin, Benzylpenicillin, and Roxithromycin. Since the presented approach is designed to utilize the modes of action of drugs/compounds at the molecular level for their anthrax antitoxicity activities, the rankings of these antibiotics are indicative of molecular-level interactions between these antibiotics with anthrax activity networks in addition to their bactericidal effects on *B. anthracis*.

Considering the number of drugs reported by Sanchez et al [140] in the top 100 ranked drugs, the technical merit of our computational approach is worthy of further research. That been said, the computation also took into consideration the effect of ET while Sanchez only focused on LT, which could be the reason for their differences. Regardless of design differences, the extent of agreement of the computed results and the literature is encouraging. Most importantly, the top 10 drugs are worthy of study for their clinical utility in treating anthrax disease, and could be used with antibiotics for the best clinical treatment outcomes if proven therapeutically beneficial.

**Table 5.1: Top 10 drug candidates** sorted by their squared Euclidean distance to the reverse anthrax activity

Drug	Biological Evidences
Fenofibrate	Cross talks between mevalonate pathway and PPAR $\alpha$ ; inhibition of LT cytotoxicity by statins mediated via inhibiting Rho GTPase and activating PPAR $\alpha$ [133], [134]
Dihydroergotamine	5-HT 1B/1D agonist Stimulation of 5-HT 1B/1D receptors activated MAPK and reduced cAMP level [135], [136]
Cotinine	Activated mitogen-activated protein kinases [137], [138]
Simvastatin	Statins inhibited LT cytotoxicity by inactivating Rho GTPase [133]
Amantadine	Cancelled activation of p38MAPK. p38MAPK kinase inhibitors protected cells from LT-mediated cytotoxicity [139], [140]
Mephenytoin	a derivative of phenytoin which inhibits active transport of Ca <sup>2+</sup> via enterocytes, and Ca <sup>2+</sup> channel in the brain [141], [142]
Mefloquine	Mefloquine is an analogue of chloroquine that had in-vitro activity protecting cells from LT toxicity [143]
Bepidil	Calcium channel blocker; Ca <sup>2+</sup> is required for LT toxicity [140], [144]
Sotalol	Decreased intracellular accumulation of cAMP, an action that is opposite to that of ET [145]
Ifosfamide	Increased renal recreation of Ca <sup>2+</sup> that can lead to disturbance of Ca <sup>2+</sup> homeostasis and depletion of Ca <sup>2+</sup> [146]

**Table 5.2: Bottom 10 drug candidates** sorted by their squared Euclidean distance to the reverse anthrax activity

Drug	Biological Evidences
Monastrol	Arresting cells in mitosis [147]
Colforsin	An agonist of adenylyl cyclase that converts ATP to cAMP. Such action would increase intracellular cAMP and synergistically increase ET toxicity [148]
Berberine	Reduces activation of MAPK signaling by chikungunya virus [149]
Withaferin-a	Activates p38 MAPK [150]
Arecoline	Antagonist P38 MAPK inhibitors [151]
Ebselen	Inhibits ASK1-p38 MAPK-p35 and JNK signaling and activated MPAK p44/42 [152]
Genistein	Activates MAPK p44/42 [153]
Apigenin	Inhibits MAPK. An action similar to LT [119]
Beclometasone	Activates p38 MAPK [140], [154]
Enilconazole	Antifungal drug for animals [Merk veterinary manual]



# Appendix A

## Replicative Senescence

### A.1 Data Description

As described in the main manuscripts, data were collected using the Luminex Xmap technology. The cells response to 6 different stimuli (wells) was measured at 2 timepoints (5 and 25 minutes) with respect to 19 phosphoproteins. The experimental results are summarized in `luminex_measurements.csv`, available in the Supplementary Information of [51]. A sample of the table is shown in Table A.1.

**Table A.1:** Aging data sample

Age	Time	Stimulus	Bead	MFI	Count
p27r1	25	IGF1	PE	16424	512
p27r1	25	EGF	STAT3	1798	230
r30	25	IL1A	STAT1	834	170
p27r1	5	EGF	AKT1	9448	162
p27r1	5	IL1A	STAT5	1668	100
p27r1	5	TGFA	STAT5	2277	107
r30	25	IL1A	HSPB1	25515	229
p27r1	5	EGF	STAT1	3890	104
r30	25	EGF	STAT1	909	173
r30	5	IGF1	NFKBIA	5588	277

The columns of the table are:

- Age describing the “age” of the cells:
  - p27r1 means that cells have undergone 27 duplications and are replicating every day
  - r30 means that the cells haven’t replicated for 30 days. This happened after 51 duplications.

- Time is the time point of the measurement in minutes
- Stimulus the stimulus used to perturb the cells. DME corresponds to no stimulus.
- Bead the measured bead. Apart from the 19 phosphoproteins, there are 2 control beads PE and BSA (see later discussion)
- MFI the median fluorescent intensity of the measurement
- Count the number of beads comprising the measurement.

The readouts and stimuli used were:

- Stimuli: EGF, IL1A, TGFA, TNF, IGF1, INS, DME (control)
- Readouts: AKT1, CREB1, PTK2, GSK3A, HSPB1, NFKBIA, JUN, MAPK12, MAPK3, MAPK9, MAP2K1, TP53, PTPN11, PRKAR2A, STAT1, STAT3, STAT5, STAT6, RELA

An overview of the raw data is shown in Figure A.1:

## A.2 Quality Control

The quality of the data was evaluated based on three check-points:

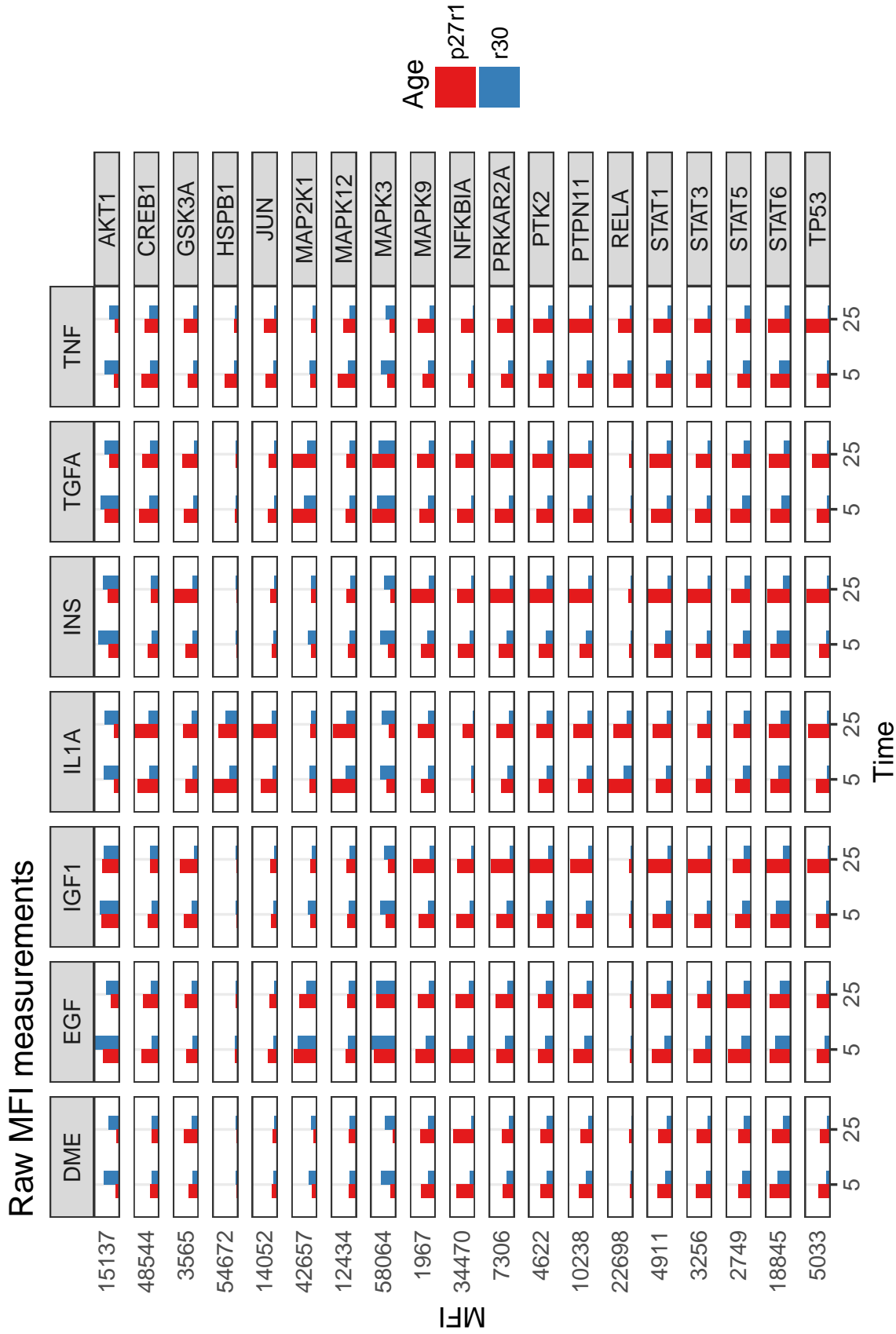
1. The sample size
2. A positive control bead (PE)
3. A negative control bead (BSA)

The bead count of a measurement represents its sample size. For this study, the experimental protocol specified the minimum bead count to 100 beads. This threshold was not met all the time and measurements with bead count less than 35 were removed from subsequent analysis.

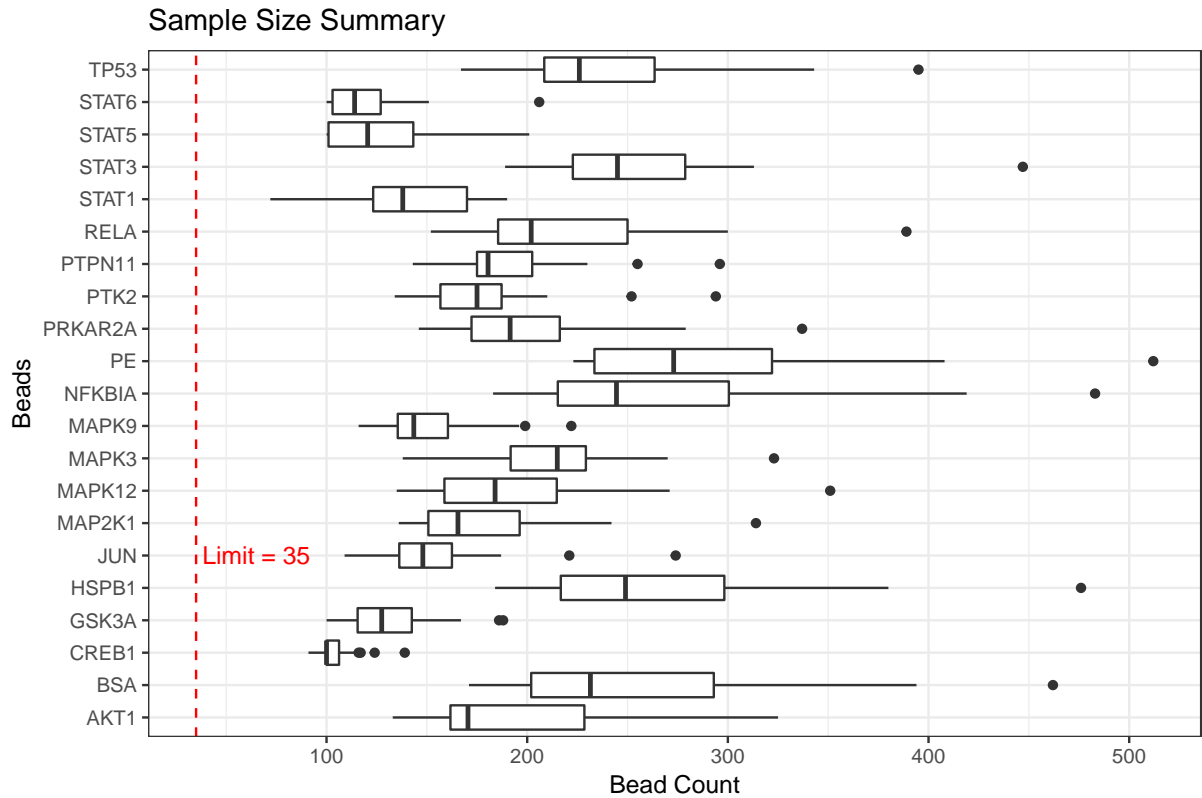
The distribution of every the sample sizes for every protein is presented as box-plots in Figure A.2. No measurements were below the required limit.

To evaluate possible fluctuations in the intensity of the detection-laser, a positive control bead was utilized. This bead was covered with *Phycoerythrin* (PE), the fluorescent proteins used to tag the captured proteins. The MFI distribution of the PE bead was examined for outliers and its overall stochasticity. The distribution of PE across wells is shown in Figure A.3. The Shapiro-Wilk test failed to reject ( $p\text{-val} = 0.81$ ) the normality hypothesis.

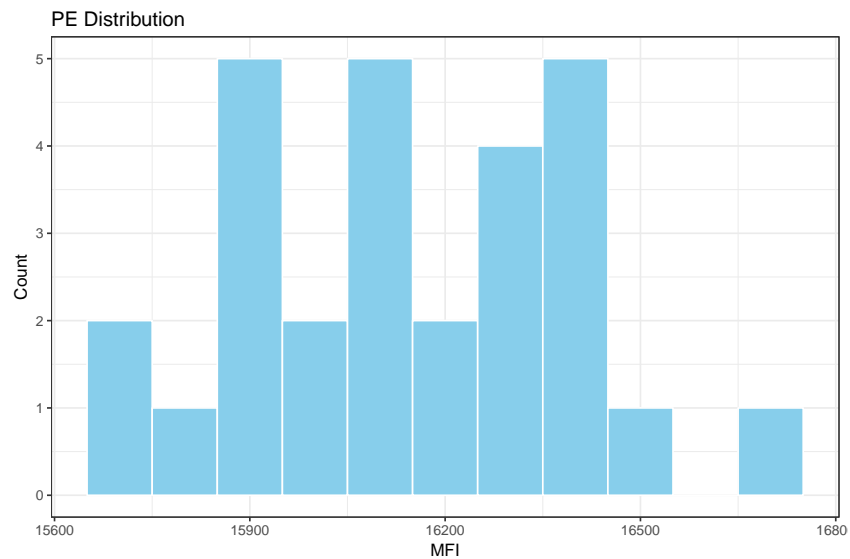
To evaluate the fluctuations of the non-specific binding, a negative control bead was utilized. This bead was covered with *Bovine Serum Albumin* (BSA), the protein used to coat the surface of the beads not covered by antibodies. The MFI



**Figure A.1: Aging raw data overview** Each panel/facet corresponds to a protein (row) stimulus (column) combination. Bars are plotted for the 2 time points (5 and 25 min) and their colors correspond to cell's age. Finally, the values on the right represent the maximum observed value in fluorescent units.



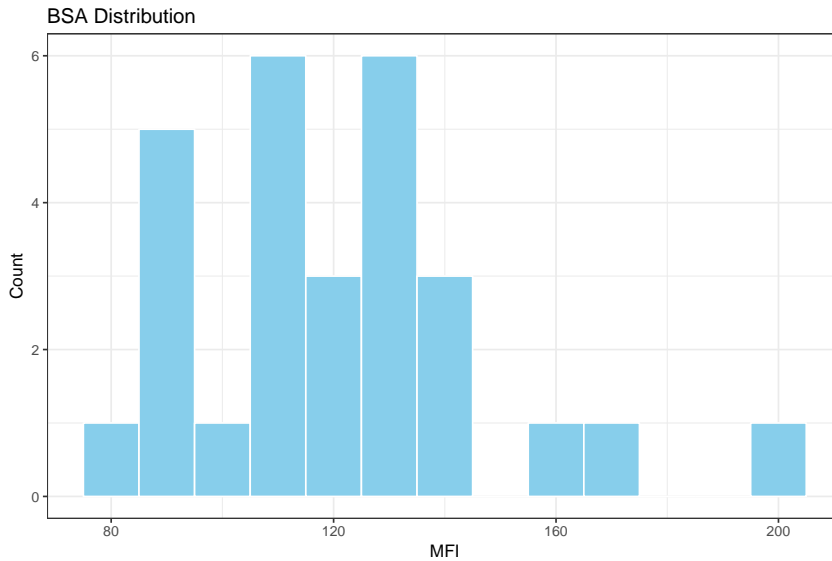
**Figure A.2:** Bead count per analyte the protocol specified 100 beads but sometimes they couldn't be retrieved.



**Figure A.3:** Distribution of positive control per well

distribution of BSA across wells is shown in Figure A.4. Again, the Shapiro-Wilk test fails to reject the normality hypothesis (pval = 0.08).





**Figure A.4: Distribution of negative control per well**

So overall, no wells were removed from the data-set and no corrections were made in this stage.

### A.3 Data Normalization

To run out algorithm, data need to be mapped into the  $[-1, 1]$  interval, where  $-1$  indicated down-regulation and  $1$  up-regulation.

To do this, we compared the effect stimuli had on cells against cells that were left untreated. For every condition, the fold change with respect to its control (*DME*) was computed. The two time-points were merged by discarding the smallest absolute fold change. The new single time-point was representative of the cells' "early response". An overview of the aggregated fold-changes is presented Figure A.5 and in more detail in Figure A.6.

Because, the input for the main algorithm must lie in the  $[-1, 1]$  interval, we normalize the fold changes using the *Gaussian Error function*. There is a variety of sigmoid function that can achieve the desired outcome but *erf* was selected because:

1. It's symmetric around 0
2. It's smooth
3. It respects some useful heuristic used in previous studies [40], [41], i.e. it's practically  $\pm 1$  for log-fold-changes close to  $\pm 1$  and close to  $\pm 0.5$  for log-fold-change close to  $\pm 0.5$ .

An overview of the final data-set is shown as a heatmap in Figure A.7.

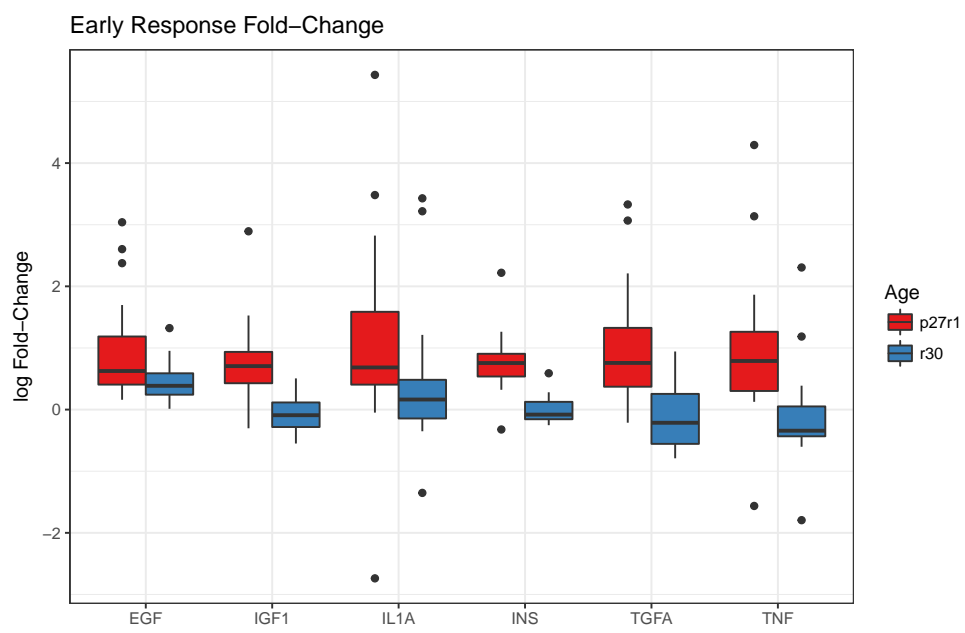


Figure A.5: Summary of fold change per analyte

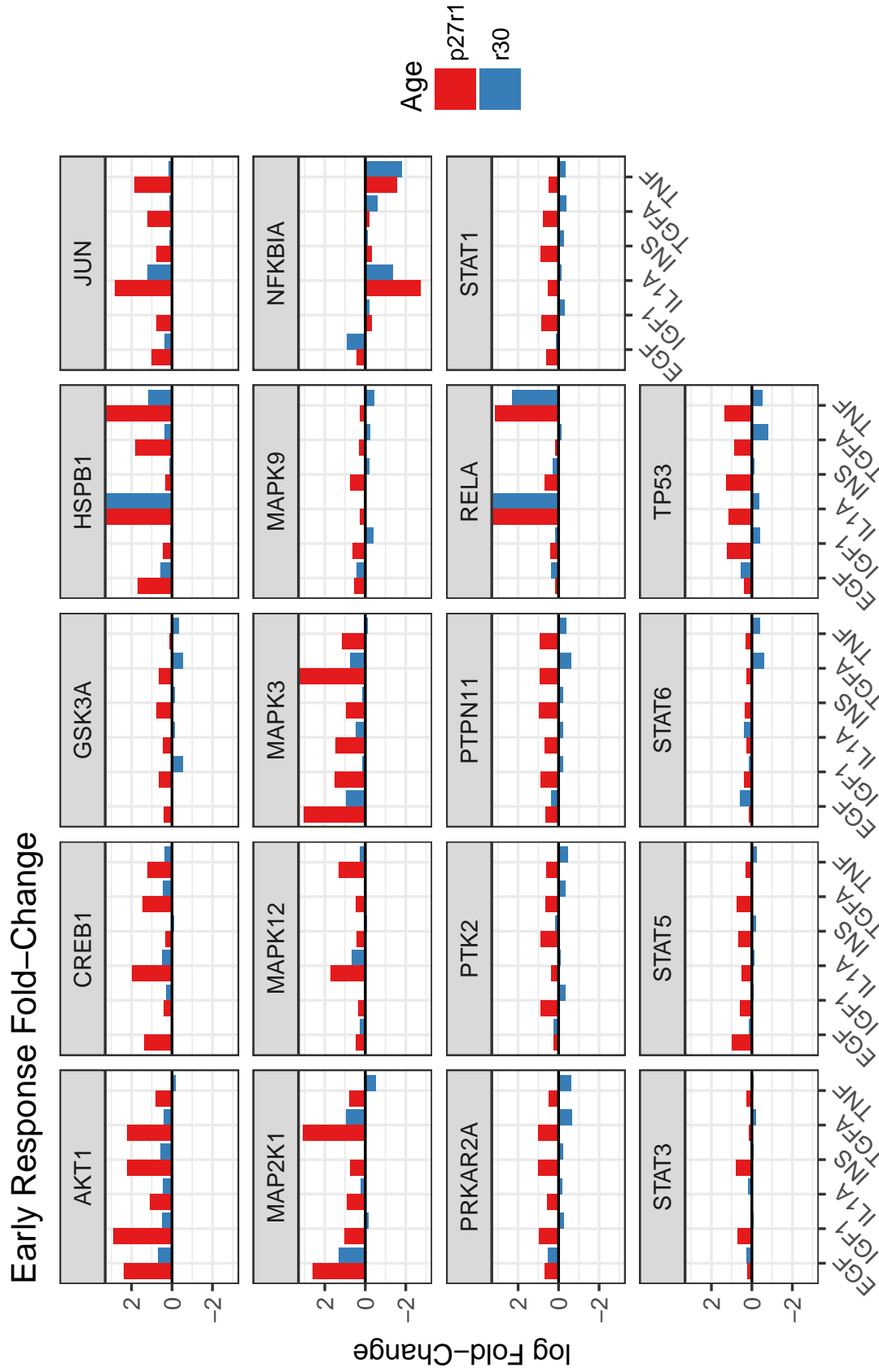
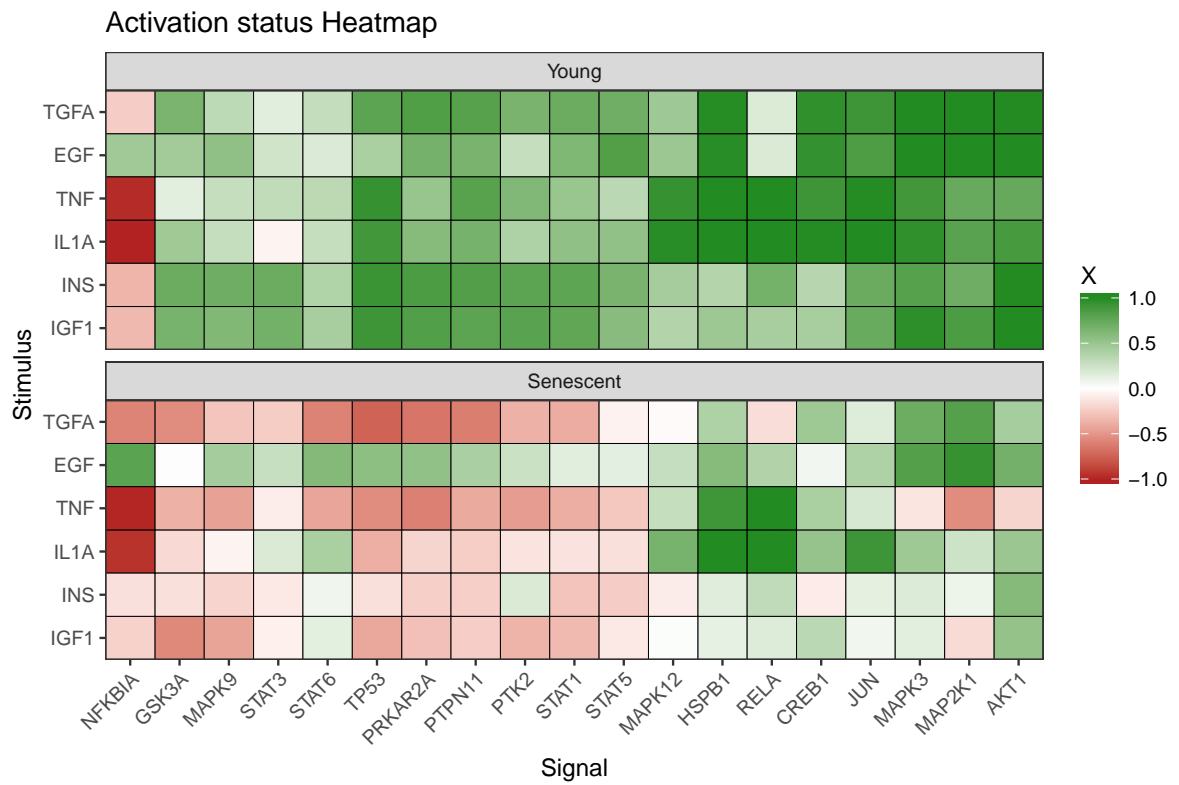


Figure A.6: Fold change per analyte and stimulus. y-axis is trimmed at  $\pm 3$ .



**Figure A.7: Overview of the activity of every analyte.** A heatmap representation of the  $m_e(n)$  function

# Appendix B

## Drug Induced Lung Disease

### B.1 Top Drug Candidates

1. **Ciclosporin** (119): Anti-inflammatory. For treatment of transplant (kidney, liver, and heart) rejection, rheumatoid arthritis, severe psoriasis. Shown to be effective treatment for interstitial lung disease of unknown etiology [87].
2. **Tretinoin** (115): Immunosuppressor. For the induction of remission in patients with acute promyelocytic leukemia (APL), French-American-British (FAB) classification M3 (including the M3 variant); For the topical treatment of acne vulgaris, flat warts and other skin conditions (psoriasis, ichthyosis congenita, ichthyosis vulgaris, lamellar ichthyosis, keratosis palmaris et plantaris, epidermolytic hyperkeratosis, senile comedones, senile keratosis, keratosis follicularis (Darier's disease), and basal cell carcinomas.); For palliative therapy to improve fine wrinkling, mottled hyperpigmentation, roughness associated with photodamage.
3. **Quercetin** (108): Flavonol. Has anti-inflammatory properties. Used to prevent the progression of obstructive pulmonary diseases [169].
4. **Resveratrol** (92): Experimental, being investigated for the treatment of Herpes labialis infections (cold sores). Has anti-inflammatory and antioxidant effects. Has been shown to alleviate COPD lung injury in rats [89].
5. **Paracetamol** (91): For temporary relief of fever, minor aches, and pains. Demonstrates weak anti-inflammatory action. Has been shown to be potentially induce asthma in long term use [170].
6. **Genistein** (88): Flavonoid. Has anti-inflammatory action. Currently being studied in clinical trials as a treatment for prostate cancer. Reverses Severe Pulmonary Hypertension and Prevents Right Heart Failure in Rats [171].
7. **Diethylstilbestrol** (78): Estrogen. For the treatment of hypertension, angina, and cluster headache prophylaxis.
8. **Copper sulfate** (78)

9. **Fulvestrant** (76): For the treatment of hormone receptor positive metastatic breast cancer in postmenopausal women with disease progression following anti-estrogen therapy.
10. **Wortmannin** (73): Used in research. Has been shown to reduce immediate-type allergic response and late phase pulmonary inflammation induced by allergen challenge in the ovalbumin-sensitized Brown Norway rat [172].
11. **ly-294002** (70): Potent inhibitor of phosphoinositide 3-kinases (PI3Ks). Has been shown to reduce allergic airway inflammation in rats. [173].
12. **Melatonin** (69): Used orally for jet lag, insomnia, shift-work disorder, circadian rhythm disorders in the blind, and benzodiazepine and nicotine withdrawal. Evidence indicates that melatonin is likely effective for treating circadian rhythm sleep disorders in blind children and adults. May be effective for treating sleep-wake cycle disturbances in children and adolescents with mental retardation, autism, and other central nervous system disorders. It may also improve secondary insomnia associated with various sleep-wake cycle disturbances. Demonstrates anti-inflammatory activity in the CNS. Reduces lung oxidative stress in patients with chronic obstructive pulmonary disease [174].
13. **Celastrol** (68): Plant extract. Potent antioxidant and anti-inflammatory drug.
14. **Cyclic Adenosine monophosphate** (65): Targets potassium/sodium HNC channel 2, PRKAR1A, PRKAR2B, Adenylate cyclase, Cyclic nucleotide-gated potassium channel mll3241, cAMP-activated global transcriptional regulator CRP. Decreases pulmonary edema in experimental acid-induced lung injury [175].
15. **sb-202190** (63): Targets P38MAPK
16. **Dopamine** (63): For the correction of hemodynamic imbalances present in the shock syndrome due to myocardial infarction, trauma, endotoxic septicemia, open-heart surgery, renal failure, and chronic cardiac decompensation as in congestive failure. Has immunomodulatory action. Has been shown inhibit pulmonary edema through the VEGF-VEGFR2 axis in a murine model of acute lung injury [176].
17. **Dinoprostone** (62): Prostaglandin E2. Up-regulation of PGE2 expression protects against the development of fibrosis after lung injury. [94].
18. **Acetylsalicylic acid** (62): Aspirin. For use in the temporary relief of various forms of pain, inflammation associated with various conditions (including rheumatoid arthritis, juvenile rheumatoid arthritis, systemic lupus erythematosus, osteoarthritis, and ankylosing spondylitis), and is also used to reduce the risk of death and/or nonfatal myocardial infarction in patients with a previous infarction or unstable angina pectoris. Was found to improve outcome in animal models of acute lung injury [177].
19. **sb-203580** (61): Targets P38MAPK

20. **Rottlerin** (61): Experimental, conductance potassium channel (BKCa++) opener. May cause pulmonary edema in vivo [178].
21. **Sulfinpyrazone** (60): For the treatment of gout and gouty arthritis
22. **Staurosporine** (60): LCK, PIM1, ITK/TSK, SYK, MAPK2, GSK3B, CSK, CDK2, PIK3CG, PDPK1, PRKCQ.
23. **Nocodazole** (59): Targets Hematopoietic prostaglandin D synthase
24. **Chrysin** (59): Plant extract. Suppresses inflammation. Attenuates allergic airway inflammation in mice. [179].
25. **Pirinixic acid** (58): Experimental, under investigation for prevention of severe cardiac dysfunction, cardiomyopathy and heart failure as a result of lipid accumulation within cardiac myocytes.
26. **Lidocaine** (58): A local anesthetic and cardiac depressant used as an antiarrhythmia agent. Demonstrates anti-inflammatory action. Attenuates acute lung injury induced by a combination of phospholipase A2 and trypsin [180].
27. **Ketoconazole** (58): For the treatment of the following systemic fungal infections: candidiasis, chronic mucocutaneous candidiasis, oral thrush, candiduria, blastomycosis, coccidioidomycosis, histoplasmosis, chromomycosis, and paracoccidioidomycosis. Has been tested for early treatment of acute lung injury and acute respiratory distress syndrome in a randomized controlled trial, but was ineffective. [181].
28. **Kanamycin** (58): For treatment of infections where one or more of the following are the known or suspected pathogens: *E. coli*, *Proteus* species (both indole-positive and indole-negative), *E. aerogenes*, *K. pneumoniae*, *S. marcescens*, and *Acinetobacter* species.
29. **Arachidonic acid** (58): Targets fatty acid-binding protein, Prostaglandin G/H synthase 1.
30. **Thioridazine** (57): For the treatment of schizophrenia and generalized anxiety disorder.
31. **Nortriptyline** (57): For the treatment of depression, chronic pain, irritable bowel syndrome, sleep disorders, diabetic neuropathy, agitation and insomnia, and migraine prophylaxis.
32. **Cycloheximide** (57): Experimental, inhibitor of protein biosynthesis in eukaryotic organisms.
33. **Furosemide** (56): For the treatment of edema associated with congestive heart failure, cirrhosis of the liver, and renal disease, including the nephrotic syndrome. Also for the treatment of hypertension alone or in combination with other antihypertensive agents.

34. **Clioquinol** (56): Withdrawn. Used as a topical antifungal treatment.
35. **Zaprinast** (55): unsuccessful clinical drug candidate that was a precursor to the chemically related PDE5 inhibitors, such as sildenafil, which successfully reached the market.
36. **Thiamazole** (55): For the treatment of hyperthyroidism, goiter, Graves disease and psoriasis. Has anti-inflammatory action.
37. **Ouabain** (55): For the treatment of atrial fibrillation and flutter and heart failure.
38. **Indometacin** (55): For moderate to severe rheumatoid arthritis including acute flares of chronic disease, ankylosing spondylitis, osteoarthritis, acute painful shoulder (bursitis and/or tendinitis) and acute gouty arthritis. Has been shown to attenuate lung injury in surfactant-deficient rabbits [182].
39. **Chenodeoxycholic acid** (55): For patients with radiolucent stones in well-opacifying gallbladders, in whom selective surgery would be undertaken except for the presence of increased surgical risk due to systemic disease or age. Chenodiol will not dissolve calcified (radiopaque) or radiolucent bile pigment stones.
40. **Kaempferol** (54): Targets UDP-glucuronosyltransferase 3A1. Has anti-inflammatory action. Has preventive and curative effects in TH2-driven allergic airway disease [183].

## B.2 Enrichment Tables

**Table B.1: GO terms enriched in genes from up-regulated DILD module.** Threshold for inclusion is p value less than 0.001.

Up-regulated module
Regulation of programmed cell death
Regulation of cell death
Regulation of apoptosis
Response to organic substance
Positive regulation of programmed cell death
Positive regulation of cell death
Regulation of transcription from RNA polymerase II promoter
Positive regulation of apoptosis
Response to inorganic substance
Negative regulation of apoptosis



**Table B.2: GO terms enriched in genes from down-regulated DILD module.**  
Threshold for inclusion is p value less than 0.001.

---

Down-regulated module
Cell cycle phase
Cell cycle process
Mitotic cell cycle
Cell cycle
m phase
Intracellular transport
Organelle fission
Interphase of mitotic cell cycle
Positive regulation of programmed cell death
Positive regulation of cell death
Interphase
Regulation of cell cycle
Positive regulation of apoptosis
M phase of mitotic cell cycle
Cell division
Nuclear division
Mitosis
Cellular response to stress

---

**Table B.3:** Most frequent nominal pharmacological effects of Pneumotox drugs from chEMBL. The frequency of the corresponding modes of action across all the drugs in cMAP is also shown.

Effect	Number of Drugs	
	Pneumotox	cMAP
DNA inhibitor	27	62
Cyclooxygenase 1,2 inhibitor	18	31
Sodium channel alpha subunit blocker	15	47
Serotonin 2a (5-HT2a) receptor antagonist	11	24
GABA-A receptor modulator	11	15
Norepinephrine transporter inhibitor	10	19
Serotonin transporter inhibitor	9	38
Glucocorticoid receptor agonist	9	50
Beta-1 adrenergic receptor antagonist	9	15
Mu opioid receptor agonist	8	8
Bacterial penicillin-binding protein inhibitor	8	36
Angiotensin-converting enzyme inhibitor	8	8
Bacterial 70S ribosome inhibitor	7	33
Tubulin inhibitor	7	12
PPAR agonist	7	9
D2-like dopamine receptor antagonist	7	17
Beta-2 adrenergic receptor antagonist	7	26
Progesterone receptor agonist	5	12
Voltage-gated L-type calcium channel blocker	5	15
Type-1 angiotensin II receptor antagonist	5	8
RNA inhibitor	5	6
Arachidonate 5-lipoxygenase inhibitor	4	4
Thymidylate synthase inhibitor	4	4
Serotonin 2c (5-HT2c) receptor antagonist	4	12
Serotonin 1d (5-HT1d) receptor agonist	4	4
Norepinephrine transporter releasing agent	4	18
HMG-CoA reductase inhibitor	4	10
FK506-binding protein 1A inhibitor	4	4
Dopamine transporter inhibitor	4	17
Dihydrofolate reductase inhibitor	4	7
Androgen Receptor agonist	3	8
Adrenergic receptor alpha-2 agonist	3	7
Vitamin k epoxide reductase inhibitor	3	4
Ferriprotoporphyrin IX inhibitor	3	6
Cytochrome P450 51 inhibitor	3	9
Bacterial dihydropteroate synthase inhibitor	3	16
Androgen Receptor antagonist	3	8

# Bibliography

- [1] H. Kitano, "Looking beyond that details: A rise in system-oriented approaches in genetics and molecular biology," *Current Genetics*, vol. 41, no. 1, pp. 1–10, 2002. doi: 10.1007/s00294-002-0285-z.
- [2] —, "Systems Biology: A Brief Overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, 2002. doi: 10.1126/science.1069492.
- [3] N. Wiener, *Cybernetics: Control and communication in the animal and the machine*. Wiley New York, 1948.
- [4] P. K. Sorger, "A reductionist's systems biology: Opinion," *Current Opinion in Cell Biology*, vol. 17, no. 1, pp. 9–11, Feb. 2005. doi: 10.1016/J.CEB.2004.12.012.
- [5] K. A. Janes and D. A. Lauffenburger, "A biological approach to computational models of proteomic networks.," *Current opinion in chemical biology*, vol. 10, no. 1, pp. 73–80, Feb. 2006. doi: 10.1016/j.cbpa.2005.12.016.
- [6] E. D. Sontag, "Some new directions in control theory inspired by systems biology.," *Systems biology*, vol. 1, no. 1, pp. 9–18, Jun. 2004.
- [7] E. C. Butcher, E. L. Berg, and E. J. Kunkel, "Systems biology in drug discovery," *Nature Biotechnology*, vol. 22, no. 10, pp. 1253–1259, Oct. 2004. doi: 10.1038/nbt1017.
- [8] E. C. Butcher, "Can cell systems biology rescue drug discovery?" *Nature Reviews Drug Discovery*, vol. 4, no. 6, pp. 461–467, May 2005. doi: 10.1038/nrd1754.
- [9] S. I. Walker and P. C. W. Davies, "The algorithmic origins of life," *Journal of The Royal Society Interface*, vol. 10, no. 79, pp. 20120869–20120869, Dec. 2012. doi: 10.1098/rsif.2012.0869.
- [10] B. Kholodenko, M. B. Yaffe, and W. Kolch, "Computational Approaches for Analyzing Information Flow in Biological Networks," *Science Signaling*, vol. 5, no. 220, re1–re1, Apr. 2012. doi: 10.1126/scisignal.2002961.
- [11] B. B. Alberts, A. D. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*. W. W. Norton & Company, 2014, ISBN: 815344325. [Online]. Available: <http://books.wwnorton.com/books/webad.aspx?id=4294997124>.
- [12] P. Cohen, "The role of protein phosphorylation in human health and disease.," *European Journal of Biochemistry*, vol. 268, no. 19, pp. 5001–5010, 2001. doi: 10.1046/j.0014-2956.2001.02473.x.

- [13] M. K. Morris, A. Chi, I. N. Melas, and L. G. Alexopoulos, "Phosphoproteomics in drug discovery," *Drug Discovery Today*, vol. 19, no. 4, pp. 425–432, 2014. doi: 10.1016/j.drudis.2013.10.010.
- [14] M. J. Vogel, D. Peric-Hupkes, and B. Van Steensel, "Detection of in vivo protein–dna interactions using damid in mammalian cells," *Nature protocols*, vol. 2, no. 6, p. 1467, 2007.
- [15] M. Dreze, D. Monachello, C. Lurin, M. E. Cusick, D. E. Hill, M. Vidal, and P. Braun, "High-Quality Binary Interactome Mapping," *Methods in Enzymology*, vol. 470, pp. 281–315, Jan. 2010. doi: 10.1016/S0076-6879(10)70012-4.
- [16] A. Bauer and B. Kuster, "Affinity purification-mass spectrometry," *European Journal of Biochemistry*, vol. 270, no. 4, pp. 570–578, Jan. 2003. doi: 10.1046/j.1432-1033.2003.03428.x.
- [17] P. C. Havugimana, G. T. Hart, T. Nepusz, H. Yang, A. L. Turinsky, Z. Li, P. I. Wang, D. R. Boutz, V. Fong, S. Phanse, M. Babu, S. A. Craig, P. Hu, C. Wan, J. Vlasblom, V.-N. Dar, A. Bezginov, G. W. Clark, G. C. Wu, S. J. Wodak, E. R. M. Tillier, A. Paccanaro, E. M. Marcotte, and A. Emili, "A census of human soluble protein complexes," *Cell*, vol. 150, no. 5, pp. 1068–81, Aug. 2012. doi: 10.1016/j.cell.2012.08.011.
- [18] E. Barillot, L. Calzone, P. Hupe, J.-P. Vert, and A. Zinovyev, *Computational Systems Biology of Cancer*. CRC Press, 2012.
- [19] U. Stelzl, U. Worm, M. Lalowski, C. Haenig, F. H. Brembeck, H. Goehler, M. Stroedicke, M. Zenkner, A. Schoenherr, S. Koeppen, J. Timm, S. Mintzlaff, C. Abraham, N. Bock, S. Kietzmann, A. Goedde, E. Toksöz, A. Droege, S. Krobitsch, B. Korn, W. Birchmeier, H. Lehrach, and E. E. Wanker, "A Human Protein-Protein Interaction Network: A Resource for Annotating the Proteome," *Cell*, vol. 122, no. 6, pp. 957–968, Sep. 2005. doi: 10.1016/j.cell.2005.08.029.
- [20] T. Rolland, M. Taşan, B. Charlotiaux, S. J. J. Pevzner, Q. Zhong, N. Sahni, S. Yi, I. Lemmens, C. Fontanillo, R. Mosca, A. Kamburov, S. D. D. Ghiassian, X. Yang, L. Ghamsari, D. Balcha, B. E. E. Begg, P. Braun, M. Brehme, M. P. P. Broly, A.-R. Carvunis, D. Convery-Zupan, R. Corominas, J. Coulombe-Huntington, E. Dann, M. Dreze, A. Dricot, C. Fan, E. Franzosa, F. Gebreab, B. J. J. Gutierrez, M. F. F. Hardy, M. Jin, S. Kang, R. Kiros, G. N. N. Lin, K. Luck, A. MacWilliams, J. Menche, R. R. R. Murray, A. Palagi, M. M. M. Poulin, X. Rambout, J. Rasla, P. Reichert, V. Romero, E. Ruysinck, J. M. M. Sahalie, A. Scholz, A. A. A. Shah, A. Sharma, Y. Shen, K. Spirohn, S. Tam, A. O. O. Tejada, S. A. A. Trigg, J.-C. Twizere, K. Vega, J. Walsh, M. E. E. Cusick, Y. Xia, A.-L. Barabási, L. M. M. Iakoucheva, P. Aloy, J. De Las Rivas, J. Tavernier, M. A. A. Calderwood, D. E. E. Hill, T. Hao, F. P. P. Roth, M. Vidal, J. De Las Rivas, J. Tavernier, M. A. A. Calderwood, D. E. E. Hill, T. Hao, F. P. P. Roth, and M. Vidal, "A Proteome-Scale Map of the Human Interactome Network," *Cell*, vol. 159, no. 5, pp. 1212–1226, Nov. 2014. doi: 10.1016/j.cell.2014.10.050.

- [21] M. Y. Hein, N. C. Hubner, I. Poser, J. Cox, N. Nagaraj, Y. Toyoda, I. A. Gak, I. Weisswange, J. Mansfeld, F. Buchholz, A. A. Hyman, and M. Mann, "A human interactome in three quantitative dimensions organized by stoichiometries and abundances," *Cell*, vol. 163, no. 3, pp. 712–23, Oct. 2015. doi: 10.1016/j.cell.2015.09.053.
- [22] E. L. Huttlin, L. Ting, R. J. Bruckner, F. Gebreab, M. P. Gygi, J. Szpyt, S. Tam, G. Zarraga, G. Colby, K. Baltier, R. Dong, V. Guarani, L. P. Vaites, A. Ordureau, R. Rad, B. K. Erickson, M. Wühr, J. Chick, B. Zhai, D. Kolippakkam, J. Mintseris, R. A. Obar, T. Harris, S. Artavanis-Tsakonas, M. E. Sowa, P. De Camilli, J. A. Paulo, J. W. Harper, and S. P. Gygi, "The BioPlex Network: A Systematic Exploration of the Human Interactome," *Cell*, vol. 162, no. 2, pp. 425–440, Jul. 2015. doi: 10.1016/j.cell.2015.06.043.
- [23] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork, "Comparative assessment of large-scale data sets of protein–protein interactions," *Nature*, vol. 417, no. 6887, pp. 399–403, May 2002. doi: 10.1038/nature750.
- [24] F. Ramirez, A. Schlicker, Y. Assenov, T. Lengauer, and M. Albrecht, "Computational analysis of human protein interaction networks," *Proteomics*, vol. 7, no. 15, pp. 2541–2552, 2007.
- [25] D. C. Kirouac, J. Saez-Rodriguez, J. Swantek, J. M. Burke, D. A. Lauffenburger, and P. K. Sorger, "Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks," *BMC Systems Biology*, vol. 6, no. 1, p. 29, 2012. doi: 10.1186/1752-0509-6-29.
- [26] K. Luck, G. M. Sheynkman, I. Zhang, and M. Vidal, "Proteome-Scale Human Interactomics," *Trends in Biochemical Sciences*, vol. 42, no. 5, pp. 342–354, May 2017. doi: 10.1016/J.TIBS.2017.02.006.
- [27] D. S. Johnson, A. Mortazavi, R. M. Myers, and B. Wold, "Genome-wide mapping of in vivo protein-dna interactions," *Science*, vol. 316, no. 5830, pp. 1497–1502, 2007. doi: 10.1126/science.1141319.
- [28] C.-L. Wei, Q. Wu, V. B. Vega, K. P. Chiu, P. Ng, T. Zhang, A. Shahab, H. C. Yong, Y. Fu, Z. Weng, J. Liu, X. D. Zhao, J.-L. Chew, Y. L. Lee, V. A. Kuznetsov, W.-K. Sung, L. D. Miller, B. Lim, E. T. Liu, Q. Yu, H.-H. Ng, and Y. Ruan, "A global map of p53 transcription-factor binding sites in the human genome," *Cell*, vol. 124, no. 1, pp. 207–219, 2006. doi: 10.1016/j.cell.2005.10.043.
- [29] V. R. Iyer, C. E. Horak, C. S. Scafe, D. Botstein, M. Snyder, and P. O. Brown, "Genomic binding sites of the yeast cell-cycle transcription factors sbf and mbf," *Nature*, vol. 409, no. 6819, p. 533, 2001.
- [30] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Hermjakob, L. Stein, and P. D'Eustachio, "The Reactome pathway knowledgebase," *Nucleic Acids Research*, vol. 42, no. D1, pp. D472–D477, Jan. 2014. doi: 10.1093/nar/gkt1102.

- [31] A. Lachmann, H. Xu, J. Krishnan, S. I. Berger, A. R. Mazloom, and A. Ma'ayan, "Chea: Transcription factor regulation inferred from integrating genome-wide chip-x experiments," *Bioinformatics*, vol. 26, no. 19, pp. 2438–2444, 2010. doi: 10.1093/bioinformatics/btq466.
- [32] V. Matys, E. Fricke, R. Geffers, E. Gößling, M. Haubrock, R. Hehl, K. Hornischer, D. Karas, A. E. Kel, O. V. Kel-Margoulis, *et al.*, "Transfac®: Transcriptional regulation, from patterns to profiles," *Nucleic Acids Research*, vol. 31, no. 1, pp. 374–378, 2003.
- [33] A. Mathelier, X. Zhao, A. W. Zhang, F. Parcy, R. Worsley-Hunt, D. J. Arenillas, S. Buchman, C.-y. Chen, A. Chou, H. Ienasescu, J. Lim, C. Shyr, G. Tan, M. Zhou, B. Lenhard, A. Sandelin, and W. W. Wasserman, "Jaspar 2014: An extensively expanded and updated open-access database of transcription factor binding profiles," *Nucleic Acids Research*, vol. 42, no. D1, pp. D142–D147, 2014. doi: 10.1093/nar/gkt997.
- [34] P. Camus, A. Fanton, P. Bonniaud, C. Camus, and P. Foucher, "Interstitial lung disease induced by drugs and radiation," *Respiration*, vol. 71, no. 4, pp. 301–326, 2004. doi: 10.1159/000079633.
- [35] J. Lamb, E. D. Crawford, D. Peck, J. W. Modell, I. C. Blat, M. J. Wrobel, J. Lerner, J.-P. Brunet, A. Subramanian, K. N. Ross, M. Reich, H. Hieronymus, G. Wei, S. A. Armstrong, S. J. Haggarty, P. A. Clemons, R. Wei, S. A. Carr, E. S. Lander, and T. R. Golub, "The connectivity map: Using gene-expression signatures to connect small molecules, genes, and disease," *Science*, vol. 313, no. 5795, pp. 1929–1935, 2006. doi: 10.1126/science.1132939.
- [36] T. Barrett, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, M. Holko, A. Yefanov, H. Lee, N. Zhang, C. L. Robertson, N. Serova, S. Davis, and A. Soboleva, "NCBI GEO: archive for functional genomics data sets—update," *Nucleic Acids Research*, vol. 41, no. D1, pp. D991–D995, 2013. doi: 10.1093/nar/gks1193.
- [37] L. G. Alexopoulos, J. Saez-Rodriguez, B. D. Cosgrove, D. A. Lauffenburger, and P. K. Sorger, "Networks inferred from biochemical data reveal profound differences in toll-like receptor and inflammatory signaling between normal and transformed hepatocytes," *Molecular & cellular proteomics : MCP*, vol. 9, no. 9, pp. 1849–65, Sep. 2010. doi: 10.1074/mcp.M110.000406.
- [38] S. a. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets.," *Journal of theoretical biology*, vol. 22, no. 3, pp. 437–467, 1969. doi: 10.1016/0022-5193(69)90015-0.
- [39] N. Le Novère, "Quantitative and logic modelling of molecular and gene networks.," *Nature reviews. Genetics*, vol. 16, no. 3, pp. 146–58, Mar. 2015. doi: 10.1038/nrg3885.
- [40] J. Saez-Rodriguez, L. G. Alexopoulos, J. Epperlein, R. Samaga, D. A. Lauffenburger, S. Klamt, and P. K. Sorger, "Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction.," *Molecular systems biology*, vol. 5, no. 331, p. 331, Dec. 2009. doi: 10.1038/msb.2009.87.

- [41] A. Mitsos, I. N. Melas, P. Siminelakis, A. D. Chairakaki, J. Saez-Rodriguez, and L. G. Alexopoulos, "Identifying Drug Effects via Pathway Alterations using an Integer Linear Programming Optimization Formulation on Phosphoproteomic Data," *PLoS Computational Biology*, vol. 5, no. 12, P. E. Bourne, Ed., e1000591, Dec. 2009. doi: 10.1371/journal.pcbi.1000591.
- [42] R. Sharan and R. M. Karp, "Reconstructing Boolean Models of Signaling," *Journal of Computational Biology*, vol. 20, no. 3, pp. 249–257, Mar. 2013. doi: 10.1089/cmb.2012.0241.
- [43] L. Vardi, E. Ruppín, and R. Sharan, "A Linearized Constraint-Based Approach for Modeling Signaling Networks," *Journal of Computational Biology*, vol. 19, no. 2, pp. 232–240, 2012. doi: 10.1089/cmb.2011.0277.
- [44] S. Videla, C. Guziolowski, F. Eduati, S. Thiele, M. Gebser, J. Nicolas, J. Saez-Rodriguez, T. Schaub, and A. Siegel, "Learning Boolean logic models of signaling networks with ASP," *Theoretical Computer Science*, vol. 599, pp. 79–101, Sep. 2015. doi: 10.1016/j.tcs.2014.06.022.
- [45] S.-S. C. Huang and E. Fraenkel, "Integrating proteomic, transcriptional, and interactome data reveals hidden components of signaling and regulatory networks," *Science Signaling*, vol. 2, no. 81, ra40, 2009. doi: 10.1126/scisignal.2000350.
- [46] N. Tuncbag, S. McCallum, S.-S. C. Huang, and E. Fraenkel, "SteinerNet: A web server for integrating 'omic' data to discover hidden components of response pathways," *Nucleic Acids Research*, vol. 40, no. W1, pp. 1–5, 2012. doi: 10.1093/nar/gks445.
- [47] N. Tuncbag, A. Braunstein, A. Pagnani, S.-S. C. Huang, J. Chayes, C. Borgs, R. Zecchina, and E. Fraenkel, "Simultaneous Reconstruction of Multiple Signaling Pathways via the Prize-Collecting Steiner Forest Problem," *Journal of Computational Biology*, vol. 20, no. 2, pp. 124–136, Feb. 2013. doi: 10.1089/cmb.2012.0092.
- [48] S. Klamt and A. von Kamp, "Computing paths and cycles in biological interaction graphs," *BMC Bioinformatics*, vol. 10, p. 181, 2009. doi: 10.1186/1471-2105-10-181.
- [49] I. N. Melas, R. Samaga, L. G. Alexopoulos, and S. Klamt, "Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs," *PLoS Computational Biology*, vol. 9, no. 9, e1003204, Sep. 2013. doi: 10.1371/journal.pcbi.1003204.
- [50] S. Thiele, L. Cerone, J. Saez-Rodriguez, A. Siegel, C. Guziolowski, and S. Klamt, "Extended notions of sign consistency to relate experimental data to signaling and regulatory network topologies," *BMC Bioinformatics*, vol. 16, no. 1, p. 345, Dec. 2015. doi: 10.1186/s12859-015-0733-7.
- [51] T. Sakellaropoulos, I. Binenbaum, M. Lefaki, A. Chatziioannou, N. Chondrogianni, and L. G. Alexopoulos, "Qualitative modeling of signaling networks in replicative senescence by selecting optimal node and arcs sets," Under Review.

- [52] I. N. Melas, T. Sakellaropoulos, F. Iorio, L. G. Alexopoulos, W.-Y. Loh, D. A. Lauffenburger, J. Saez-Rodriguez, and J. P. F. Bai, "Identification of drug-specific pathways based on gene expression data: application to drug induced lung injury.," *Integrative biology : quantitative biosciences from nano to macro*, vol. 7, no. 8, pp. 904–20, Aug. 2015. doi: 10.1039/c4ib00294f.
- [53] U. Alon, "Network motifs: theory and experimental approaches.," *Nat Rev Genet*, vol. 8, no. 6, pp. 450–61, 2007. doi: 10.1038/nrg2102.
- [54] A. Mitsos, I. N. Melas, M. K. Morris, J. Saez-Rodriguez, D. A. Lauffenburger, and L. G. Alexopoulos, "Non Linear Programming (NLP) Formulation for Quantitative Modeling of Protein Signal Transduction Pathways," *PLoS ONE*, vol. 7, no. 11, 2012. doi: 10.1371/journal.pone.0050085.
- [55] D. Bertsimas and J. N. Tsitsiklis, *Introduction to Linear Optimization*. Athena Scientific Belmont, MA, 1997, vol. 6.
- [56] N. Chondrogianni, F. L. L. Stratford, I. P. Trougakos, B. Friguet, A. J. Rivett, and E. S. Gonos, "Central Role of the Proteasome in Senescence and Survival of Human Fibroblasts," *Journal of Biological Chemistry*, vol. 278, no. 30, pp. 28 026–28 037, Jul. 2003, issn: 0021-9258. doi: 10.1074/jbc.M301048200.
- [57] P. Rolewska, A. Simm, R.-E. Silber, and B. Bartling, "Reduced Expression Level of the cAMP Response Element-Binding Protein Contributes to Lung Aging," *American Journal of Respiratory Cell and Molecular Biology*, vol. 50, no. 1, p. 130 830 092 407 005, Aug. 2013, issn: 1044-1549. doi: 10.1165/rcmb.2013-00570C.
- [58] C. Poussin, C. Mathis, L. G. Alexopoulos, D. E. Messinis, R. H. J. Dulize, V. Belcastro, I. N. Melas, T. Sakellaropoulos, K. Rhrissorakrai, E. Bilal, P. Meyer, M. Talikka, S. Boué, R. Norel, J. J. Rice, G. Stolovitzky, N. V. Ivanov, M. C. Peitsch, and J. Hoeng, "The species translation challenge-a systems biology perspective on human and rat bronchial epithelial cells.," en, *Scientific Data*, vol. 1, p. 140 009, Jan. 2014, issn: 2052-4463. doi: 10.1038/sdata.2014.9.
- [59] T. Tchkonina, Y. Zhu, J. van Deursen, J. Campisi, and J. L. Kirkland, "Cellular senescence and the senescent secretory phenotype: therapeutic opportunities.," *The Journal of clinical investigation*, vol. 123, no. 3, pp. 966–72, Mar. 2013, issn: 1558-8238. doi: 10.1172/JCI64098.
- [60] H. D. Kim, S.-J. Yu, H. S. Kim, Y.-J. Kim, J. M. Choe, Y. G. Park, J. Kim, and J. Sohn, "Interleukin-4 induces senescence in human renal carcinoma cell lines through stat6 and p38 mapk," *Journal of Biological Chemistry*, vol. 288, no. 40, pp. 28 743–28 754, 2013.
- [61] M. Xu, T. Tchkonina, and J. L. Kirkland, "Perspective: Targeting the JAK/STAT pathway to fight age-related dysfunction," *Pharmacological Research*, vol. 111, pp. 152–154, Sep. 2016, issn: 10436618. doi: 10.1016/j.phrs.2016.05.015.
- [62] J. Maruyama, I. Naguro, K. Takeda, and H. Ichijo, "Stress-activated MAP kinase cascades in cellular senescence," *Current Medicinal Chemistry*, vol. 16, no. 10, pp. 1229–1235, Apr. 2009. doi: 10.2174/092986709787846613.



- [63] H. Pan and T. Finkel, "Key proteins and pathways that regulate lifespan," *Journal of Biological Chemistry*, vol. 292, no. 16, pp. 6452–6460, Apr. 2017, issn: 0021-9258. doi: 10.1074/jbc.R116.771915.
- [64] B. Schumacher, I. van der Pluijm, M. J. Moorhouse, T. Kosteus, A. R. Robinson, Y. Suh, T. M. Breit, H. van Steeg, L. J. Niedernhofer, W. van IJcken, A. Bartke, S. R. Spindler, J. H. J. Hoeijmakers, G. T. J. van der Horst, and G. A. Garinis, "Delayed and Accelerated Aging Share Common Longevity Assurance Mechanisms," *PLoS Genetics*, vol. 4, no. 8, S. K. Kim, Ed., e1000161, Aug. 2008, issn: 1553-7404. doi: 10.1371/journal.pgen.1000161.
- [65] A. V. Rao, P. J. Valk, K. H. Metzeler, C. R. Acharya, S. A. Tuchman, M. M. Stevenson, D. A. Rizzieri, R. Delwel, C. Buske, S. K. Bohlander, A. Potti, and B. Löwenberg, "Age-Specific Differences in Oncogenic Pathway Dysregulation in Patients With Acute Myeloid Leukemia," *Journal of Clinical Oncology*, vol. 27, no. 33, pp. 5580–5586, Nov. 2009, issn: 0732-183X. doi: 10.1200/JCO.2009.22.2547.
- [66] C. K. Anders, C. R. Acharya, D. S. Hsu, G. Broadwater, K. Garman, J. A. Foekens, Y. Zhang, Y. Wang, K. Marcom, J. R. Marks, S. Mukherjee, J. R. Nevins, K. L. Blackwell, and A. Potti, "Age-Specific Differences in Oncogenic Pathway Deregulation Seen in Human Breast Tumors," *PLoS ONE*, vol. 3, no. 1, X. Wu, Ed., e1373, Jan. 2008, issn: 1932-6203. doi: 10.1371/journal.pone.0001373.
- [67] S. Zhao and R. Iyengar, "Systems pharmacology: Network analysis to identify multi-scale mechanisms of drug action," *Annual Review of Pharmacology and Toxicology*, vol. 52, no. 1, pp. 505–521, 2012. doi: 10.1146/annurev-pharmtox-010611-134520.
- [68] J. Li, U. Rix, B. Fang, Y. Bai, A. Edwards, J. Colinge, K. L. Bennett, J. Gao, L. Song, S. Eschrich, *et al.*, "A chemical and phosphoproteomic characterization of dasatinib action in lung cancer," *Nature chemical biology*, vol. 6, no. 4, p. 291, 2010. doi: 10.1038/nchembio.332.
- [69] F. Iorio, T. Rittman, H. Ge, M. Menden, and J. Saez-Rodriguez, "Transcriptional data: A new gateway to drug repositioning?" *Drug Discovery Today*, vol. 18, no. 7, pp. 350–357, 2013. doi: 10.1016/j.drudis.2012.07.014.
- [70] J. R. Parikh, B. Klinger, Y. Xia, J. A. Marto, and N. Blüthgen, "Discovering causal signaling pathways through gene-expression patterns," *Nucleic Acids Research*, vol. 38, no. suppl\_2, W109–W117, Jul. 2010. doi: 10.1093/nar/gkq424.
- [71] A. L. Tarca, S. Draghici, P. Khatri, S. S. Hassan, P. Mittal, J.-s. Kim, C. J. Kim, J. P. Kusanovic, and R. Romero, "A novel signaling pathway impact analysis," *Bioinformatics*, vol. 25, no. 1, pp. 75–82, 2009. doi: 10.1093/bioinformatics/btn577.
- [72] S. Jaeger, J. Min, F. Nigsch, M. Camargo, J. Hutz, A. Cornett, S. Cleaver, A. Buckler, and J. L. Jenkins, "Causal network models for predicting compound targets and driving pathways in cancer," *Journal of Biomolecular Screening*, vol. 19, no. 5, pp. 791–802, 2014. doi: 10.1177/1087057114522690.

- [73] K. Zarrinhalam, A. Enayetallah, A. Gutteridge, B. Sidders, and D. Ziemek, "Molecular causes of transcriptional response: A bayesian prior knowledge approach," *Bioinformatics*, vol. 29, no. 24, pp. 3167–3173, 2013. doi: 10.1093/bioinformatics/btt557. eprint: /oup/backfile/content\_public/journal/bioinformatics/29/24/10.1093\_bioinformatics\_btt557/3/btt557.pdf.
- [74] E. Y. Chen, H. Xu, S. Gordonov, M. P. Lim, M. H. Perkins, and A. Ma'ayan, "Expression2kinases: Mrna profiling linked to multiple upstream regulatory layers," *Bioinformatics*, vol. 28, no. 1, pp. 105–111, 2012. doi: 10.1093/bioinformatics/btr625.
- [75] M. Kuhn, D. Szklarczyk, S. Pletscher-Frankild, T. H. Blicher, C. von Mering, L. J. Jensen, and P. Bork, "Stitch 4: Integration of protein–chemical interactions with user data," *Nucleic Acids Research*, vol. 42, no. D1, pp. D401–D407, 2014. doi: 10.1093/nar/gkt1207.
- [76] G. Wu, X. Feng, and L. Stein, "A human functional protein interaction network and its application to cancer data analysis," *Genome Biology*, vol. 11, no. 5, R53, May 2010. doi: 10.1186/gb-2010-11-5-r53.
- [77] M. Schwaiblmair, W. Behr, T. Haeckel, B. Märkl, W. Foerg, and T. Berghaus, "Drug induced interstitial lung disease," *The Open Respiratory Medicine Journal*, vol. 6, p. 63, 2012. doi: 10.2174/1874306401206010063.
- [78] S. M. Kabir, S. Mukherjee, V. Rajaratnam, M. G. Smith, and S. K. Das, "Desensitization of  $\beta$ -adrenergic receptors in lung injury induced by 2-chloroethyl ethyl sulfide, a mustard analog," *Journal of Biochemical and Molecular Toxicology*, vol. 23, no. 1, pp. 59–70, 2009. doi: 10.1002/jbt.20265.
- [79] E. Kratzer, Y. Tian, N. Sarich, T. Wu, A. Meliton, A. Leff, and A. A. Birukova, "Oxidative Stress Contributes to Lung Injury and Barrier Dysfunction via Microtubule Destabilization," *American Journal of Respiratory Cell and Molecular Biology*, vol. 47, no. 5, pp. 688–697, 2012. doi: 10.1165/rcmb.2012-01610C.
- [80] A. R. L. Medford and A. B. Millar, "Vascular endothelial growth factor (VEGF) in acute lung injury (ALI) and acute respiratory distress syndrome (ARDS): paradox or paradigm?" *Thorax*, vol. 61, no. 7, pp. 621–626, 2006. doi: 10.1136/thx.2005.040204.
- [81] K. A. Janes, J. R. Kelly, S. Gaudet, J. G. Albeck, P. K. Sorger, and D. A. Lauffenburger, "Cue-signal-response analysis of TNF-induced apoptosis by partial least squares regression of dynamic multivariate data," *Journal of Computational Biology*, vol. 11, no. 4, pp. 544–561, Aug. 2004. doi: 10.1089/cmb.2004.11.544.
- [82] T. R. Martin, M. Nakamura, and G. Matute-Bello, "The role of apoptosis in acute lung injury," *Critical Care Medicine*, vol. 31, no. Supplement, S184–S188, Apr. 2003. doi: 10.1097/01.ccm.0000057841.33876.b1.
- [83] E. J. Seeley, P. Rosenberg, and M. A. Matthay, "Calcium flux and endothelial dysfunction during acute lung injury: a STIMulating target for therapy," *Journal of Clinical Investigation*, vol. 123, no. 3, pp. 1015–1018, Feb. 2013. doi: 10.1172/jci68093.

- [84] M. Kolb, P. J. Margetts, D. C. Anthony, F. Pitossi, and J. Gauldie, "Transient expression of il-1 $\beta$  induces acute lung injury and chronic repair leading to pulmonary fibrosis," *The Journal of Clinical Investigation*, vol. 107, no. 12, pp. 1529–1536, Jun. 2001. doi: 10.1172/JCI12568.
- [85] M. A. O'Reilly, "DNA damage and cell cycle checkpoints in hyperoxic lung injury: braking to facilitate repair," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 281, no. 2, pp. L291–L305, 2001. doi: 10.1152/ajplung.2001.281.2.L291.
- [86] S. P. McGee, H. Zhang, W. Karmaus, and T. Sabo-Attwood, "Influence of sex and disease severity on gene expression profiles in individuals with idiopathic pulmonary fibrosis," *International journal of molecular epidemiology and genetics*, vol. 5, no. 2, p. 71, 2014.
- [87] J. Moolman, P. Bardin, D. Rossouw, and J. Joubert, "Cyclosporin as a treatment for interstitial lung disease of unknown aetiology.," *Thorax*, vol. 46, no. 8, pp. 592–595, 1991.
- [88] J. Wang, H.-W. Xu, B.-S. Li, J. Zhang, and J. Cheng, "Preliminary study of protective effects of flavonoids against radiation-induced lung injury in mice," *Asian Pacific Journal of Cancer Prevention*, vol. 13, no. 12, pp. 6441–6446, 2012.
- [89] M. Zhou, J. He, S. Yu, R. Zhu, J. Lu, F. Ding, and G. Xu, "Effect of resveratrol on chronic obstructive pulmonary disease in rats and its mechanism," *Yao xue xue bao= Acta pharmaceutica Sinica*, vol. 43, no. 2, pp. 128–132, 2008.
- [90] E. A. Ozer, A. Kumral, E. Ozer, N. Duman, O. Yilmaz, S. Ozkal, and H. Ozkan, "Effect of retinoic acid on oxygen-induced lung injury in the newborn rat," *Pediatric Pulmonology*, vol. 39, no. 1, pp. 35–40, 2004. doi: 10.1002/ppul.20131.
- [91] M. K. Glassberg, R. Choi, V. Manzoli, S. Shahzeidi, P. Rauschkolb, R. Voswinckel, M. Aliniazei, X. Xia, and S. J. Elliot, "17 $\beta$ -estradiol replacement reverses age-related lung disease in estrogen-deficient c57bl/6j mice," *Endocrinology*, vol. 155, no. 2, pp. 441–448, Feb. 2014. doi: 10.1210/en.2013-1345.
- [92] H.-P. Yu, Y.-C. Hsieh, T. Suzuki, T. Shimizu, M. A. Choudhry, M. G. Schwacha, and I. H. Chaudry, "Salutary effects of estrogen receptor-beta agonist on lung injury after trauma-hemorrhage," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 290, no. 5, pp. L1004–L1009, May 2006. doi: 10.1152/ajplung.00504.2005.
- [93] V. Ivanova, O. B. Garbuzenko, K. R. Reuhl, D. C. Reimer, V. P. Pozharov, and T. Minko, "Inhalation treatment of pulmonary fibrosis by liposomal prostaglandin e2," *European Journal of Pharmaceutics and Biopharmaceutics*, vol. 84, no. 2, pp. 335–344, Jun. 2013. doi: 10.1016/j.ejpb.2012.11.023.
- [94] R. J. Hodges, R. G. Jenkins, C. P. D. Wheeler-Jones, D. M. Copeman, S. E. Bottoms, G. J. Bellingan, C. B. Nanthakumar, G. J. Laurent, S. L. Hart, M. L. Foster, and R. J. McAnulty, "Severity of lung injury in cyclooxygenase-2-deficient mice is dependent on reduced prostaglandin e2 production," *The American Journal of Pathology*, vol. 165, no. 5, pp. 1663–1676, Nov. 2004. doi: 10.1016/s0002-9440(10)63423-2.

- [95] W.-Y. Loh, "Regression trees with unbiased variable selection and interaction detection," *Statistica Sinica*, vol. 12, no. 2, pp. 361–386, 2002.
- [96] J. Hur, A. Y. Guo, W. Y. Loh, E. L. Feldman, and J. P. F. Bai, "Integrated systems pharmacology analysis of clinical drug-induced peripheral neuropathy," *CPT Pharmacometrics Syst. Pharmacol.*, vol. 3, no. 5, e114, May 2014. doi: 10.1038/psp.2014.11.
- [97] J. P. F. Bai, A. Utis, G. Crippen, H.-D. He, V. Fischer, R. Tullman, H.-Q. Yin, C.-P. Hsu, L. Jiang, and K.-K. Hwang, "Use of classification regression tree in predicting oral absorption in humans," *Journal of Chemical Information and Computer Sciences*, vol. 44, no. 6, pp. 2061–2069, Nov. 2004. doi: 10.1021/ci040023n.
- [98] N. Zaman, L. Li, M. L. Jaramillo, Z. Sun, C. Tibiche, M. Banville, C. Collins, M. Trifiro, M. Paliouras, A. Nantel, M. O'Connor-McCourt, and E. Wang, "Signaling network assessment of mutations and copy number variations predict breast cancer subtype-specific drug targets," *Cell Reports*, vol. 5, no. 1, pp. 216–223, Oct. 2013. doi: 10.1016/j.celrep.2013.08.028.
- [99] D. Fazekas, M. Koltai, D. Türei, D. Módos, M. Pálffy, Z. Dúl, L. Zsákai, M. Szalay-Bekó, K. Lenti, I. J. Farkas, T. Vellai, P. Csermely, and T. Korcsmáros, "Signalink 2 – a signaling pathway resource with multi-layered regulatory networks," *BMC Systems Biology*, vol. 7, no. 1, p. 7, 2013. doi: 10.1186/1752-0509-7-7.
- [100] Y. Guan, D. Gorenshteyn, M. Burmeister, A. K. Wong, J. C. Schimenti, M. A. Handel, C. J. Bult, M. A. Hibbs, and O. G. Troyanskaya, "Tissue-specific functional networks for prioritizing phenotype and disease genes," *PLoS Computational Biology*, vol. 8, no. 9, e1002694, Sep. 2012. doi: 10.1371/journal.pcbi.1002694.
- [101] J. Bai, T. Sakellaropoulos, and L. Alexopoulos, "A Biologically-Based Computational Approach to Drug Repurposing for Anthrax Infection," *Toxins*, vol. 9, no. 3, p. 99, Mar. 2017. doi: 10.3390/toxins9030099.
- [102] A. Navdarashvili, T. J. Doker, M. Geleishvili, D. L. Haberling, G. A. Kharod, T. H. Rush, E. Maes, K. Zakhshvili, P. Imnadze, W. A. Bower, H. T. Walke, S. V. Shadomy, and Anthrax Investigation Team, "Human anthrax outbreak associated with livestock exposure: Georgia, 2012.," *Epidemiology and Infection*, vol. 144, no. 1, pp. 76–87, Jan. 2016, issn: 1469-4409. doi: 10.1017/S0950268815001442. [Online]. Available: [http://www.journals.cambridge.org/abstract%7B%5C\\_%7DS0950268815001442%20http://www.ncbi.nlm.nih.gov/pubmed/26088361](http://www.journals.cambridge.org/abstract%7B%5C_%7DS0950268815001442%20http://www.ncbi.nlm.nih.gov/pubmed/26088361).
- [103] Committee on Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Bacillus Anthracis Mailings, *Review of the Scientific Approaches Used During the FBI's Investigation of the 2001 Anthrax Letters*, eng. Washington, D.C.: National Academies Press, Jun. 2011, ISBN: 9780309187190. doi: 10.17226/13098. [Online]. Available: <http://www.nap.edu/catalog/13098>.
- [104] M. Meselson, J. Guillemin, M. Hugh-Jones, A. Langmuir, I. Popova, A. Shelokov, and O. Yampolskaya, "The Sverdlovsk anthrax outbreak of 1979.," eng, *Science (New York, N.Y.)*, vol. 266, no. 5188, pp. 1202–1208, Nov. 1994, issn: 0036-8075 (Print).

- [105] F. A. Abramova, L. M. Grinberg, O. V. Yampolskaya, and D. H. Walker, "Pathology of inhalational anthrax in 42 cases from the sverdlovsk outbreak of 1979.," eng, *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 6, pp. 2291–2294, Mar. 1993, issn: 0027-8424 (Print).
- [106] L. M. Grinberg, F. A. Abramova, O. V. Yampolskaya, D. H. Walker, and J. H. Smith, "Quantitative pathology of inhalational anthrax I: quantitative microscopic findings.," eng, *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc*, vol. 14, no. 5, pp. 482–495, May 2001, issn: 0893-3952 (Print). doi: 10.1038/modpathol.3880337.
- [107] S. Liu, M. Moayeri, and S. H. Leppla, "Anthrax lethal and edema toxins in anthrax pathogenesis.," eng, *Trends in microbiology*, vol. 22, no. 6, pp. 317–325, Jun. 2014, issn: 1878-4380 (Electronic). doi: 10.1016/j.tim.2014.02.012.
- [108] C. W. Hicks, X. Cui, D. A. Sweeney, Y. Li, A. Barochia, and P. Q. Eichacker, "The potential contributions of lethal and edema toxins to the pathogenesis of anthrax associated shock.," eng, *Toxins*, vol. 3, no. 9, pp. 1185–1202, Sep. 2011, issn: 2072-6651 (Electronic). doi: 10.3390/toxins3091185.
- [109] K. A. Bradley, J. Mogridge, M. Mourez, R. J. Collier, and J. A. T. Young, "Identification of the cellular receptor for anthrax toxin," *Nature*, vol. 414, no. 6860, p. 225, 2001.
- [110] K. R. Klimpel, N. Arora, and S. H. Leppla, "Anthrax toxin lethal factor contains a zinc metalloprotease consensus sequence which is required for lethal toxin activity," *Molecular microbiology*, vol. 13, no. 6, pp. 1093–1100, 1994.
- [111] N. S. Duesbery and G. F. Woude, "Anthrax lethal factor causes proteolytic inactivation of mitogen-activated protein kinase kinase," *Journal of applied microbiology*, vol. 87, no. 2, pp. 289–293, 1999.
- [112] T. C. Levin, K. E. Wickliffe, S. H. Leppla, and M. Moayeri, "Heat shock inhibits caspase-1 activity while also preventing its inflammasome-mediated activation by anthrax lethal toxin," *Cellular microbiology*, vol. 10, no. 12, pp. 2434–2446, 2008.
- [113] K. E. Wickliffe, S. H. Leppla, and M. Moayeri, "Anthrax lethal toxin-induced inflammasome formation and caspase-1 activation are late events dependent on ion fluxes and the proteasome," *Cellular microbiology*, vol. 10, no. 2, pp. 332–343, 2008.
- [114] S. H. Leppla, "Anthrax toxin edema factor: a bacterial adenylate cyclase that increases cyclic AMP concentrations of eukaryotic cells," *Proceedings of the National Academy of Sciences*, vol. 79, no. 10, pp. 3162–3166, 1982.
- [115] A. Agrawal, J. Lingappa, S. H. Leppla, S. Agrawal, A. Jabbar, C. Quinn, and B. Pulendran, "Impairment of dendritic cells and adaptive immunity by anthrax lethal toxin," *Nature*, vol. 424, no. 6946, p. 329, 2003.
- [116] A. Agrawal and B. Pulendran, "Anthrax lethal toxin: A weapon of multisystem destruction," *Cellular and Molecular Life Sciences CMLS*, vol. 61, no. 22, pp. 2859–2865, 2004.
- [117] J. M. Warfel, A. D. Steele, and F. D'Agnillo, "Anthrax lethal toxin induces endothelial barrier dysfunction," *The American journal of pathology*, vol. 166, no. 6, pp. 1871–1881, 2005.

- [118] W. J. Ribot, R. G. Panchal, K. C. Brittingham, G. Ruthel, T. A. Kenny, D. Lane, B. Curry, T. A. Hoover, A. M. Friedlander, and S. Bavari, "Anthrax lethal toxin impairs innate immune functions of alveolar macrophages and facilitates *Bacillus anthracis* survival," *Infection and immunity*, vol. 74, no. 9, pp. 5029–5034, 2006.
- [119] T. Liu, E. Milia, R. R. Warburton, N. S. Hill, M. Gaestel, and U. S. Kayyali, "Anthrax lethal toxin disrupts the endothelial permeability barrier through blocking p38 signaling," *Journal of cellular physiology*, vol. 227, no. 4, pp. 1438–1445, 2012.
- [120] T. G. Popova, V. Espina, W. Zhou, C. Mueller, L. Liotta, and S. G. Popov, "Whole proteome analysis of mouse lymph nodes in cutaneous anthrax," *PLoS One*, vol. 9, no. 10, e110873, 2014.
- [121] D. L. Hoover, A. M. Friedlander, L. C. Rogers, I. K. Yoon, R. L. Warren, and A. S. Cross, "Anthrax edema toxin differentially regulates lipopolysaccharide-induced monocyte production of tumor necrosis factor alpha and interleukin-6 by increasing intracellular cyclic AMP.," *Infection and immunity*, vol. 62, no. 10, pp. 4432–4439, 1994.
- [122] J.-N. Tournier, A. Quesnel-Hellmann, J. Mathieu, C. Montecucco, W.-J. Tang, M. Mock, D. R. Vidal, and P. L. Goossens, "Anthrax edema toxin cooperates with lethal toxin to impair cytokine secretion during infection of dendritic cells," *The Journal of Immunology*, vol. 174, no. 8, pp. 4934–4941, 2005.
- [123] J. E. Comer, A. K. Chopra, J. W. Peterson, and R. König, "Direct inhibition of T-lymphocyte activation by anthrax toxins in vivo," *Infection and immunity*, vol. 73, no. 12, pp. 8275–8281, 2005.
- [124] J. Hong, R. C. Doebele, M. W. Lingen, L. A. Quilliam, W.-J. Tang, and M. R. Rosner, "Anthrax edema toxin inhibits endothelial cell chemotaxis via Epac and Rap1," *Journal of Biological Chemistry*, vol. 282, no. 27, pp. 19781–19787, 2007.
- [125] L. A. Yeager, A. K. Chopra, and J. W. Peterson, "Bacillus anthracis edema toxin suppresses human macrophage phagocytosis and cytoskeletal remodeling via the protein kinase A and exchange protein activated by cyclic AMP pathways," *Infection and immunity*, vol. 77, no. 6, pp. 2530–2543, 2009.
- [126] M. P. Maddugoda, C. Stefani, D. Gonzalez-Rodriguez, J. Saarikangas, S. Torrino, S. Janel, P. Munro, A. Doye, F. Prodon, M. Aurrand-Lions, *et al.*, "cAMP signaling by anthrax edema toxin induces transendothelial cell tunnels, which are resealed by MIM via Arp2/3-driven actin polymerization," *Cell host & microbe*, vol. 10, no. 5, pp. 464–474, 2011.
- [127] M. Dozmorov, W. Wu, K. Chakrabarty, J. L. Booth, R. E. Hurst, K. M. Coggeshall, and J. P. Metcalf, "Gene expression profiling of human alveolar macrophages infected by *B. anthracis* spores demonstrates TNF- $\alpha$  and NF- $\kappa$ b are key components of the innate immune response to the pathogen," *BMC infectious diseases*, vol. 9, no. 1, p. 152, 2009.
- [128] K. M. Chauncey, M. C. Lopez, G. Sidhu, S. E. Szarowicz, H. V. Baker, C. Quinn, and F. S. Southwick, "Bacillus anthracis' lethal toxin induces broad transcriptional responses in human peripheral monocytes," *BMC immunology*, vol. 13, no. 1, p. 33, 2012.

- [129] J. L. Levinsohn, Z. L. Newman, K. A. Hellmich, R. Fattah, M. A. Getz, S. Liu, I. Sastalla, S. H. Leppla, and M. Moayeri, "Anthrax lethal factor cleavage of Nlrp1 is required for activation of the inflammasome," *PLoS pathogens*, vol. 8, no. 3, e1002638, 2012.
- [130] M. Moayeri, I. Sastalla, and S. H. Leppla, "Anthrax and the inflammasome.," *Microbes and infection*, vol. 14, no. 5, pp. 392–400, May 2012, ISSN: 1769-714X. doi: 10.1016/j.micinf.2011.12.005. [Online]. Available: <http://www.ncbi.nlm.nih.gov/pubmed/22207185><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3322314>.
- [131] J. N. Finger, J. D. Lich, L. C. Dare, M. N. Cook, K. K. Brown, C. Duraiswami, J. J. Bertin, and P. J. Gough, "Autolytic proteolysis within the function to find domain (FIIND) is required for NLRP1 inflammasome activity," *Journal of Biological Chemistry*, vol. 287, no. 30, pp. 25 030–25 037, 2012.
- [132] J. De Rooij, F. J. T. Zwartkruis, M. H. G. Verheijen, R. H. Cool, S. M. B. Nijman, A. Wittinghofer, and J. L. Bos, "Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP," *Nature*, vol. 396, no. 6710, p. 474, 1998.
- [133] A. M. DeCathelineau and G. M. Bokoch, *Inactivation of Rho GTPases by Statins Attenuates Anthrax Lethal Toxin Activity*, eng, Jan. 2009. doi: 10.1128/IAI.01005-08.
- [134] G. Martin, H. Duez, C. Blanquart, V. Berezowski, P. Poulain, J.-C. Fruchart, J. Najib-Fruchart, C. Glineur, and B. Staels, "Statin-induced inhibition of the Rho-signaling pathway activates PPAR $\alpha$  and induces HDL apoA-I," *The Journal of clinical investigation*, vol. 107, no. 11, pp. 1423–1432, 2001.
- [135] V. Kayser, B. Aubel, M. Hamon, and S. Bourgoin, "The antimigraine 5-HT<sub>1B/1D</sub> receptor agonists, sumatriptan, zolmitriptan and dihydroergotamine, attenuate pain-related behaviour in a rat model of trigeminal neuropathic pain," *British journal of pharmacology*, vol. 137, no. 8, pp. 1287–1297, 2002.
- [136] J. M. Hinton, P. B. Hill, J. Y. Jeremy, and C. J. Garland, "Signalling pathways activated by 5-HT<sub>1B/5-HT1D</sub> receptors in native smooth muscle and primary cultures of rabbit renal artery smooth muscle cells," *Journal of vascular research*, vol. 37, no. 6, pp. 457–468, 2000.
- [137] G. W. Warren and A. K. Singh, "Nicotine and lung cancer," *Journal of carcinogenesis*, vol. 12, 2013.
- [138] J.-R. Tsai, I.-W. Chong, C.-C. Chen, S.-R. Lin, C.-C. Sheu, and J.-J. Hwang, "Mitogen-activated protein kinase pathway was significantly activated in human bronchial epithelial cells by nicotine," *DNA and cell biology*, vol. 25, no. 5, pp. 312–322, 2006.
- [139] P. C. Eckels, A. Banerjee, E. E. Moore, N. J. D. McLaughlin, L. M. Gries, M. R. Kelher, K. M. England, F. Gamboni-Robertson, S. Y. Khan, and C. C. Silliman, "Amantadine inhibits platelet-activating factor induced clathrin-mediated endocytosis in human neutrophils," *American Journal of Physiology-Cell Physiology*, vol. 297, no. 4, pp. C886–C897, 2009.

- [140] A. M. Sanchez, D. Thomas, E. J. Gillespie, R. Damoiseaux, J. Rogers, J. P. Saxe, J. Huang, M. Manchester, and K. A. Bradley, "Amiodarone and bepridil inhibit anthrax toxin entry into host cells," *Antimicrobial agents and chemotherapy*, vol. 51, no. 7, pp. 2403–2411, 2007.
- [141] M. von Borstel Smith, K. Crofoot, R. Rodriguez-Proteau, and T. M. Filtz, "Effects of phenytoin and carbamazepine on calcium transport in Caco-2 cells," *Toxicology in vitro*, vol. 21, no. 5, pp. 855–862, 2007.
- [142] M. Sitges, L. M. Chiu, and R. C. Reed, "Effects of levetiracetam, carbamazepine, phenytoin, valproate, lamotrigine, oxcarbazepine, topiramate, vinpocetine and sertraline on presynaptic hippocampal Na<sup>+</sup> and Ca<sup>2+</sup> channels permeability," *Neurochemical research*, vol. 41, no. 4, pp. 758–769, 2016.
- [143] A. J. Thompson, M. Lochner, and S. C. R. Lummis, "The antimalarial drugs quinine, chloroquine and mefloquine are antagonists at 5-HT<sub>3</sub> receptors," *British journal of pharmacology*, vol. 151, no. 5, pp. 666–677, 2007.
- [144] R. Bhatnagar, Y. Singh, S. H. Leppla, and A. M. Friedlander, "Calcium is required for the expression of anthrax lethal toxin activity in the macrophagelike cell line J774A. 1.," *Infection and immunity*, vol. 57, no. 7, pp. 2107–2114, 1989.
- [145] J. W. Wisler, S. M. DeWire, E. J. Whalen, J. D. Violin, M. T. Drake, S. Ahn, S. K. Shenoy, and R. J. Lefkowitz, "A unique mechanism of  $\beta$ -blocker action: carvedilol stimulates  $\beta$ -arrestin signaling," *Proceedings of the National Academy of Sciences*, vol. 104, no. 42, pp. 16 657–16 662, 2007.
- [146] P. T. C. Ho, K. Zimmerman, L. H. Wexler, S. Blaney, P. Jarosinski, L. Weaver-McClure, S. Izraeli, and F. M. Balis, "A prospective evaluation of ifosfamide-related nephrotoxicity in children and young adults," *Cancer*, vol. 76, no. 12, pp. 2557–2564, 1995.
- [147] J. C. Cochran, J. E. Gatial, T. M. Kapoor, and S. P. Gilbert, "Monastrol inhibition of the mitotic kinesin Eg5," *Journal of Biological Chemistry*, vol. 280, no. 13, pp. 12 658–12 667, 2005.
- [148] C. M. Johannessen, L. A. Johnson, F. Piccioni, A. Townes, D. T. Frederick, M. K. Donahue, R. Narayan, K. T. Flaherty, J. A. Wargo, D. E. Root, *et al.*, "A melanocyte lineage program confers resistance to MAP kinase pathway inhibition," *Nature*, vol. 504, no. 7478, p. 138, 2013.
- [149] F. S. Varghese, B. Thaa, S. N. Amrun, D. Simarmata, K. Rausalu, T. A. Nyman, A. Merits, G. M. McInerney, L. F. P. Ng, and T. Ahola, "The antiviral alkaloid berberine reduces chikungunya virus-induced mitogen-activated protein kinase signaling," *Journal of virology*, vol. 90, no. 21, pp. 9743–9757, 2016.
- [150] L.-H. Shi, X.-J. Wu, J.-S. Liu, and Y.-B. Gao, "Withaferin a activates stress signalling proteins in high risk acute lymphoblastic leukemia," *International journal of clinical and experimental pathology*, vol. 8, no. 12, p. 15 652, 2015.
- [151] Y.-T. Deng, H.-M. Chen, S.-J. Cheng, C.-P. Chiang, and M. Y.-P. Kuo, "Arecoline-stimulated connective tissue growth factor production in human buccal mucosal fibroblasts: Modulation by curcumin," *Oral oncology*, vol. 45, no. 9, e99–e105, 2009.



- [152] K. P. Sarker, K. K. Biswas, J. L. Rosales, K. Yamaji, T. Hashiguchi, K.-Y. Lee, and I. Maruyama, "Ebselen inhibits NO-induced apoptosis of differentiated PC12 cells via inhibition of ASK1-p38 MAPK-p53 and JNK signaling and activation of p44/42 MAPK and Bcl-2," *Journal of neurochemistry*, vol. 87, no. 6, pp. 1345–1353, 2003.
- [153] L. Yu, K. Ham, X. Gao, L. Castro, Y. Yan, G. E. Kissling, C. J. Tucker, N. Flagler, R. Dong, T. K. Archer, *et al.*, "Epigenetic regulation of transcription factor promoter regions by low-dose genistein through mitogen-activated protein kinase and mitogen-and-stress activated kinase 1 nongenomic signaling," *Cell Communication and Signaling*, vol. 14, no. 1, p. 18, 2016.
- [154] S. Boncompagni, L. Arthurton, E. Akujuru, T. Pearson, D. Steverding, F. Protasi, and G. Mutungi, "Membrane glucocorticoid receptors are localised in the extracellular matrix and signal through the MAPK pathway in mammalian skeletal muscle fibres," *The Journal of physiology*, vol. 593, no. 12, pp. 2679–2692, 2015.
- [155] A. W. Artenstein, S. M. Opal, P. Cristofaro, J. E. Palardy, N. A. Parejo, M. D. Green, and J. W. Jhung, "Chloroquine enhances survival in *Bacillus anthracis* intoxication," *Journal of Infectious Diseases*, vol. 190, no. 9, pp. 1655–1660, 2004.
- [156] P. J. Zhu, J. P. Hobson, N. Southall, C. Qiu, C. J. Thomas, J. Lu, J. Inglese, W. Zheng, S. H. Leppla, T. H. Bugge, *et al.*, "Quantitative high-throughput screening identifies inhibitors of anthrax-induced cell death," *Bioorganic & medicinal chemistry*, vol. 17, no. 14, pp. 5139–5145, 2009.
- [157] R. C. Squires, S. M. Muehlbauer, and J. Brojatsch, "Proteasomes control caspase-1 activation in anthrax lethal toxin-mediated cell killing," *Journal of Biological Chemistry*, vol. 282, no. 47, pp. 34 260–34 267, 2007.
- [158] I. Brook, "The prophylaxis and treatment of anthrax," *International journal of antimicrobial agents*, vol. 20, no. 5, pp. 320–325, 2002.
- [159] L. V. Lee, K. E. Bower, F.-S. Liang, J. Shi, D. Wu, S. J. Sucheck, P. K. Vogt, and C.-H. Wong, "Inhibition of the proteolytic activity of anthrax lethal factor by aminoglycosides," *Journal of the American Chemical Society*, vol. 126, no. 15, pp. 4774–4775, 2004.
- [160] J. Ågren, M. Finn, B. Bengtsson, and B. Segerman, "Microevolution during an anthrax outbreak leading to clonal heterogeneity and penicillin resistance," *PLoS One*, vol. 9, no. 2, e89112, 2014.
- [161] K. E. Remy, P. Qiu, Y. Li, X. Cui, and P. Q. Eichacker, "B. anthracis associated cardiovascular dysfunction and shock: the potential contribution of both non-toxin and toxin components," *BMC medicine*, vol. 11, no. 1, p. 217, 2013.
- [162] J. Brojatsch, A. Casadevall, and D. L. Goldman, "Molecular determinants for a cardiovascular collapse in anthrax," *Frontiers in bioscience (Elite edition)*, vol. 6, p. 139, 2014.
- [163] N. Silswal, N. Parelkar, J. Andresen, and M. J. Wacker, "Restoration of endothelial function in Ppar $\alpha$ -/- mice by tempol," *PPAR research*, vol. 2015, 2015.

- [164] A. Tabernero, K. Schoonjans, L. Jesel, I. Carpusca, J. Auwerx, and R. Andriantsitohaina, "Activation of the peroxisome proliferator-activated receptor  $\alpha$  protects against myocardial ischaemic injury and improves endothelial vasodilatation," *BMC pharmacology*, vol. 2, no. 1, p. 10, 2002.
- [165] O. S. Gardner, B. J. Dewar, H. S. Earp, J. M. Samet, and L. M. Graves, "Dependence of peroxisome proliferator-activated receptor ligand-induced mitogen-activated protein kinase signaling on epidermal growth factor receptor transactivation," *Journal of Biological Chemistry*, vol. 278, no. 47, pp. 46 261–46 269, 2003.
- [166] P. Bergman, C. Linde, K. Pütsep, A. Pohanka, S. Normark, B. Henriques-Normark, J. Andersson, and L. Björkhem-Bergman, "Studies on the antibacterial effects of statins-in vitro and in vivo," *PloS one*, vol. 6, no. 8, e24394, 2011.
- [167] G. Vitale, L. Bernardi, G. Napolitani, M. Mock, and C. Montecucco, "Susceptibility of mitogen-activated protein kinase family members to proteolysis by anthrax lethal factor.," *Biochemical Journal*, vol. 352, no. Pt 3, p. 739, 2000.
- [168] P. J. Vainio, K. Törnquist, and R. K. Tuominen, "Cotinine and nicotine inhibit each other's calcium responses in bovine chromaffin cells," *Toxicology and applied pharmacology*, vol. 163, no. 2, pp. 183–187, 2000.
- [169] S. Ganesan, A. Faris, A. Comstock, S. Chatteraj, A. Chatteraj, J. Burgess, J. Curtis, F. Martinez, S. Zick, M. Hershenson, and U. Sajjan, "Quercetin prevents progression of disease in elastase/LPS-exposed mice by negatively regulating MMP expression," *Respiratory Research*, vol. 11, no. 1, p. 131, 2010. doi: 10.1186/1465-9921-11-131.
- [170] S. O. Shaheen, J. A. C. Sterne, C. E. Songhurst, and P. G. J. Burney, "Frequent paracetamol use and asthma in adults," *Thorax*, vol. 55, no. 4, pp. 266–270, 2000. doi: 10.1136/thorax.55.4.266.
- [171] H. Matori, S. Umar, R. D. Nadadur, S. Sharma, R. Partow-Navid, M. Afkhami, M. Amjedi, and M. Eghbali, "Genistein, a soy phytoestrogen, reverses severe pulmonary hypertension and prevents right heart failure in rats," *Hypertension*, vol. 60, no. 2, pp. 425–430, 2012. doi: 10.1161/HYPERTENSIONAHA.112.191445.
- [172] B. Tigani, J. P. Hannon, L. Mazzoni, and J. R. Fozard, "Effects of wortmannin on airways inflammation induced by allergen in actively sensitised brown norway rats," *European Journal of Pharmacology*, vol. 433, no. 2, pp. 217–223, 2001. doi: [https://doi.org/10.1016/S0014-2999\(01\)01515-1](https://doi.org/10.1016/S0014-2999(01)01515-1).
- [173] W. Duan, A. M. A. Datiles, B. P. Leung, C. J. Vlahos, and W. F. Wong, "An anti-inflammatory role for a phosphoinositide 3-kinase inhibitor ly294002 in a mouse asthma model," *International Immunopharmacology*, vol. 5, no. 3, pp. 495–502, 2005. doi: <https://doi.org/10.1016/j.intimp.2004.10.015>.
- [174] de Matos Cavalcante Antonio George, de Bruin Pedro Felipe Carvalhedo, de Bruin Veralice Meireles Sales, N. D. Muniz, P. E. D. Barros, C. M. Medeiros, and A. G. Matos, "Melatonin reduces lung oxidative stress in patients with chronic obstructive pulmonary disease: A randomized, double-blind, placebo-controlled study," *Journal of Pineal Research*, vol. 53, no. 3, pp. 238–244, 2012. doi: 10.1111/j.1600-079X.2012.00992.x.

- [175] D. F. McAuley, J. A. Frank, X. Fang, and M. A. Matthay, "Clinically relevant concentrations of beta2-adrenergic agonists stimulate maximal cyclic adenosine monophosphate-dependent airspace fluid clearance and decrease pulmonary edema in experimental acid-induced lung injury," *Crit. Care Med.*, vol. 32, no. 7, pp. 1470–1476, Jul. 2004.
- [176] P. K. Vohra, L. H. Hoepfner, G. Sagar, S. K. Dutta, S. Misra, R. D. Hubmayr, and D. Mukhopadhyay, "Dopamine inhibits pulmonary edema through the vegf-vegfr2 axis in a murine model of acute lung injury," *American Journal of Physiology-Lung Cellular and Molecular Physiology*, vol. 302, no. 2, pp. L185–L192, 2012. doi: 10.1152/ajplung.00274.2010.
- [177] P. R. Tuinman, M. C. Müller, G. Jongsma, M. A. Hegeman, and N. P. Juffermans, "High-dose acetylsalicylic acid is superior to low-dose as well as to clopidogrel in preventing lipopolysaccharide-induced lung injury in mice," *Shock*, vol. 40, no. 4, pp. 334–338, 2013.
- [178] J. R. Klinger, J. D. Murray, B. Casserly, D. F. Alvarez, J. A. King, S. S. An, G. Choudhary, A. N. Owusu-Sarfo, R. Warburton, and E. O. Harrington, "Rottlerin causes pulmonary edema in vivo: A possible role for  $pkc\delta$ ," *Journal of Applied Physiology*, vol. 103, no. 6, pp. 2084–2094, 2007. doi: 10.1152/jappphysiol.00695.2007.
- [179] M. Su, "Chrysin attenuates allergic airway inflammation by modulating the transcription factors T-bet and GATA-3 in mice," *Molecular Medicine Reports*, Apr. 2012. doi: 10.3892/mmr.2012.893. [Online]. Available: <https://doi.org/10.3892/mmr.2012.893>.
- [180] Y. Kiyonari, K. Nishina, K. Mikawa, N. Maekawa, and H. Obara, "Lidocaine attenuates acute lung injury induced by a combination of phospholipase a2 and trypsin," *Critical care medicine*, vol. 28, no. 2, pp. 484–489, 2000.
- [181] T. A. N. A. for the ARDS Network, "Ketoconazole for early treatment of acute lung injury and acute respiratory distress syndrome: A randomized controlled trial," *JAMA*, vol. 283, no. 15, pp. 1995–2002, 2000. doi: 10.1001/jama.283.15.1995.
- [182] R. Burger, D. Fung, and A. C. Bryan, "Lung injury in a surfactant-deficient lung is modified by indomethacin," *Journal of Applied Physiology*, vol. 69, no. 6, pp. 2067–2071, Dec. 1990. doi: 10.1152/jappl.1990.69.6.2067.
- [183] K. Medeiros, L. Faustino, E. Borduchi, R. Nascimento, T. Silva, E. Gomes, M. Pi-vezam, and M. Russo, "Preventive and curative glycoside kaempferol treatments attenuate the TH2-driven allergic airway disease," *International Immunopharmacology*, vol. 9, no. 13-14, pp. 1540–1548, Dec. 2009. doi: 10.1016/j.intimp.2009.09.005.



# Theodoros SAKELLAROPOULOS

## PERSONAL DATA

---

PLACE AND DATE OF BIRTH: Athens, Greece | 30 July 1988  
ADDRESS: Margari 17, 11525, Athens, Greece  
PHONE: +30 210 7721516, +30 6980194834  
EMAIL: [teo.sakel@gmail.com](mailto:teo.sakel@gmail.com)  
LINKEDIN: [teo-sakel](#)

## WORK EXPERIENCE

---

JUL 2018 | PhD Student at National Technical University of Athens (NTUA)  
FEB 2014 | [Systems Bioengineering Group](#)  
Mathematical modeling of cell signaling networks

NOV 2017 | [Aifantis & Tsirigos Lab](#) – NYU Langone Health  
MAY 2017 | *Pre-doc Fellow of Pathology*

SEP 2016 | [Illumina, Inc.](#)  
MAY 2016 | *Bioinformatics Intern*  
Development of the *Encore* package. Cancer cohort analysis and reports

MAY 2017 | [Protavio Ltd](#)  
AUG 2013 | *Computational Scientist*  
Proteomic assay analysis and characterization.

AUG 2015 | [FDA Center for Drug Evaluation and Research](#)  
MAR 2015 | *ORISE Fellow*  
Computational analysis of adverse drug reactions. Drug repositioning to treat Anthrax infection.

FEB 2014 | Research Scientist at NTUA  
SEP 2012 | [Systems Bioengineering Group](#)  
Development, statistical analysis and optimization of proteomic assays

## EDUCATION

---

SEPTEMBER 2012 | Diploma in MECHANICAL ENGINEERING,  
**National Technical University of Athens**, Athens, Greece  
Thesis: “Noise Modeling and Optimization in multiplex ELISA experiments”,  
Advisor: Prof. Leonidas Alexopoulos | GPA: 7.53/10

APRIL-JULY 2014 | Visiting Student at Saez-Rodriguez Group  
European Bioinformatics Institute, Hinxton, Cambridge UK

SINCE 2012 | Massive Online Open Courses (Coursera, EdX)  
JULY 2006 | 2<sup>nd</sup> High School of Neo Psychiko  
Final Grade: 18.6/20, 1<sup>st</sup> overall.

## PEER REVIEWED PUBLICATIONS

---

“Qualitative modeling of signaling networks in replicative senescence by selecting optimal node and arc sets” **Sakellaropoulos T**, Binenbaum I, Lefaki M, Chatziioannou A, Chondrogianni N, Alexopoulos LG PLoS Computational Biology (Under Review)

“Classification and Mutation Prediction from Non-Small Cell Lung Cancer Histopathology Images using Deep Learning” Coudray N, Moreira AL, **Sakellaropoulos T**, Fenyo D, Razavian N, Tsirigos A bioRxiv

“Stratification of TAD boundaries reveals preferential insulation of super-enhancers by strong boundaries” Gong Y, Lazaris C, **Sakellaropoulos T**, Lozano A, Kambadur P, Ntziachristos P, Aifantis I, Tsirigos A Nature Communications 2018

“Mechanism-based biomarker discovery” Antoranz A, **Sakellaropoulos T**, Saez-Rodriguez J, Alexopoulos LG Drug Discovery Today 2017

“A biologically-based computational approach to drug repurposing for anthrax infection” Bai J, **Sakellaropoulos T**, Alexopoulos LG Toxins 2017

“Computed biological relations among 5 select treatment-related organ/tissue toxicities” **Sakellaropoulos T**, Herod T, Alexopoulos LG, Bai J. Chemical Research in Toxicology 2016 (Cover Page)

“Computational Approaches to Accelerating Novel Medicine and Better Patient Care from Bedside to Benchtop” **Sakellaropoulos T**, Hur J, Melas IN, Guo EY, Alexopoulos LG, Bohlooly M, Bai J. Advances in Protein Chemistry and Structural Biology 2016

“Identification of drug-specific pathways based on gene expression data: application to drug induced lung injury” Melas IN, **Sakellaropoulos T**, Iorio F, Alexopoulos LG, Loh WY, Lauffenburger DA, Saez-Rodriguez J, Bai J. Integrative Biology 2015

“Network Reconstruction Based on Proteomic Data and Prior Knowledge of Protein Connectivity Using Graph Theory”. Stavrakas V, Melas IN\*, **Sakellaropoulos T\***, Alexopoulos LG PLoS ONE. 2015 (\* Equal Contributors).

“A crowd-sourcing approach for the construction of species-specific cell signaling networks”. Erhan B\*, **Sakellaropoulos T\***, Challenge Participants, Melas IN, Messinis D, Belcastro V, Rhrissorakrai K, Meyer P, Norel R, Iskandar A, Blaese E, Rice JJ, Peitsch MC, Hoeng J, Stolovitzky G, Alexopoulos LG and Poussin C Bioinformatics. 2014 (\* Equal Contributors).

“The species translation challenge - A systems biology perspective on human and rat bronchial epithelial cells”. Poussin C, Mathis C, Alexopoulos LG, Messinis DE, Dulize R, Belcastro V, Melas IN, **Sakellaropoulos T**, Rhrissorakrai K, Bilal E, Meyer P, Talikka M, Boué S, Norel R, Rice JJ, Stolovitzky G, Ivanov NV, Peitsch MC, Hoeng J. Scientific Data. 2014.

“Modeling chondrocyte response using systems biology and biomechanics approach”. Alexopoulos LG, Chatzopoulou E, **Sakellaropoulos T**, Samara S, Katopodi T, Melas IN, Tzeranis T, Dailiana Z, Kolia P. Journal of Tissue Engineering and Regenerative Medicine. 2014