ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων για την αποκρυπτογράφηση μηχανισμών σε ανθρώπινες ασθένειες

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΔΗΜΗΤΡΙΟΥ ΜΕΣΣΗΝΗ

Διπλωματούχου Μηχανολόγου Μηχανικού Ε.Μ.Π.

ΕΠΙΒΛΕΠΩΝ:

Λ. ΑΛΕΞΟΠΟΥΛΟΣ

Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Απρίλιος 2018

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων για την αποκρυπτογράφηση μηχανισμών σε ανθρώπινες ασθένειες

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΔΗΜΗΤΡΙΟΥ ΜΕΣΣΗΝΗ

Διπλωματούχου Μηχανολόγου Μηχανικού Ε.Μ.Π.

ΤΡΙΜΕΛΗΣ ΣΥΜΒΟΥΛΕΥΤΙΚΗ ΕΠΙΤΡΟΠΗ:
Λ. ΑΛΕΞΟΠΟΥΛΟΣ, Επ. Καθ. Ε.Μ.Π.
Χ. ΠΡΟΒΑΤΙΔΗΣ, Καθ. Ε.Μ.Π.
Ν. ΧΡΟΝΗΣ, Αν. Καθ. U Michigan

ΕΠΤΑΜΕΛΗΣ ΕΞΕΤΑΣΤΙΚΗ ΕΠΙΤΡΟΠΗ:
Λ. ΑΛΕΞΟΠΟΥΛΟΣ, Επ. Καθ. Ε.Μ.Π.
Χ. ΠΡΟΒΑΤΙΔΗΣ, Καθ. Ε.Μ.Π.
Ν. ΧΡΟΝΗΣ, Αν. Καθ. U Michigan
Π. ΚΟΛΛΙΑ, Αν. Καθ. Ε.Κ.Π.Α.
Δ. ΧΑΤΖΗΑΒΡΑΜΙΔΗΣ, Καθ. Ε.Μ.Π.
Γ. ΜΑΤΣΟΠΟΥΛΟΣ, Αν. Καθ. Ε.Μ.Π.
Δ. ΚΕΚΟΣ, Καθ. Ε.Μ.Π.

Αθήνα, Απρίλιος 2018

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

Στη γιαγιά μου
Νίκη

Πρόλογος

Ο σκοπός της παρούσας διδακτορικής διατριβής είναι να αποκρυπτογραφήσει μηχανισμούς σε ανθρώπινες ασθένειες, μέσω της συστημικής μοντελοποίησης και της ανάλυσης μεγάλων δεδομένων. Η διδακτορική διατριβή διεξήχθη στο εργαστήριο του Επ. Καθηγητή ΕΜΠ. Λεωνίδα Αλεξόπουλου, του τομέα Μηχανολογικών Κατασκευών και Αυτομάτου Ελέγχου της σχολής Μηχανολόγων Μηχανικών του Εθνικού Μετσόβιου Πολυτεχνείου. Κατά τη διάρκεια της διατριβής του, ο συγγραφέας φιλοξενήθηκε κατά διαστήματα στο Metabolic Engineering and Systems Biology Lab (MESBL) του FORTH/ICE-HT στην Πάτρα, Ελλάδα και στο U.S. Food & Drug Administration (FDA) στην Washington DC, ΗΠΑ.

Ο συγγραφέας θέλει να ευχαριστήσει για την βοήθειά τους, τους ακόλουθους ανθρώπους. Τον Χριστόφορο Προβατίδη, Καθ. ΕΜΠ και τον Νικόλαο Χρόνη, Αν. Καθ. του Πανεπιστημίου του Michigan, για τις συμβουλές τους καθ' όλη την διάρκεια των διδακτορικών σπουδών του υποψηφίου. Την Μαρία Κλάπα, Principal Researcher του MESBL, για την φιλοξενία της στο εργαστήριό της και την καθοδήγησή της στις τεχνικές της Μεταβολομικής. Την Jane Bai, Regulatory Review Scientist - Systems Pharmacology του U.S. Food & Drug Administration, για την φιλοξενία της στο FDA και την καθοδήγησή της στην έρευνα της Καρδιομυοπάθειας που προκαλείται από φάρμακα. Την Παναγούλα Κόλλια, Αν. Καθ. ΕΚΠΑ, τον Δημήτριο Χατζηαβραμίδη, Καθ. ΕΜΠ, τον Γεώργιο Ματσόπουλο, Αν. Καθ. ΕΜΠ και τον Δημήτριο Κέκο, Καθ. ΕΜΠ, μέλη της επταμελούς επιτροπής του υποψηφίου για την εξέταση και εποικοδομητική κριτική της διατριβής του. Όλα τα μέλη του εργαστηρίου Εμβιομηχανικής και Συστημικής Βιολογίας του Δρ. Λεωνίδα Αλεξόπουλου για την πολύ εποικοδομητική συνεργασία μας και συνεχή υποστήριξη, με ειδική αναφορά στους: Δρ. Ιωάννη Μελά, Δανάη Κυρλή-Φλώρου, Δανάη-Στέλλα Ζαρείφη, Θεόδωρο Σακελλαρόπουλο, Νίκο Καβαλόπουλο, Γεώργιο Κανακάρη και Asier Antoranz για την στενή μας συνεργασία. Τέλος, ο συγγραφέας θέλει να ευχαριστήσει τον επιβλέποντα Δρ. Λεωνίδα Αλεξόπουλο για την καθοδήγηση, υποστήριξη και έμπνευση στον υποψήφιο διδάκτορα καθ' όλη τη διάρκεια των διδακτορικών του σπουδών.

Πίνακας Περιεχομένων

Περίληψη

Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων για την αποκρυπτογράφηση μηχανισμών σε ανθρώπινες ασθένειες

Ο σκοπός της παρούσας διδακτορικής διατριβής είναι να αποκρυπτογραφήσει μηχανισμούς σε ανθρώπινες ασθένειες, και συγκεκριμένα στην Οστεοαρθρίτιδα, στην μετάφραση βιολογικών δεδομένων από τον αρουραίο στον άνθρωπο, στην Σκλήρυνση κατά πλάκας, στον καρκίνο του ήπατος, στην καρδιομυοπάθεια που προκαλείται από φάρμακα και στην μη αλκοολική λιπώδη νόσο του ήπατος (ΜΑΛΝΗ). Στα παρακάτω κεφάλαια περιγράφονται τα ευρήματα σχετικά με την κάθε ανθρώπινη πάθηση και οι αντίστοιχες εργασίες.

Τα τελευταία 4 χρόνια, ο Δημήτρης Μεσσήνης έχει συγγράψει 6 δημοσιεύσεις σε επιστημονικά περιοδικά και 14 περιλήψεις - ανακοινώσεις σε διεθνή συνέδρια. Έχει μία δημοσίευση ως πρώτος συγγραφέας και έχει επιτύχει δείκτη h-index 6. Μέσω των προσπαθειών για την αποκρυπτογράφηση των μηχανισμών διαφορετικών ανθρώπινων ασθενειών, ο υποψήφιος διδάκτωρ έχει αποκτήσει ευρεία κατανόηση στη Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων, ανέπτυξε τις δικές του πειραματικές και υπολογιστικές μεθόδους και συνέβαλε σε αρκετά σημαντικά ευρήματα που περιγράφονται παραπάνω. Το έργο του στην Σκλήρυνση κατά πλάκας τον έκανε να επιλεχθεί μεταξύ 6 από 110 εργασιών για να δώσει μια σύντομη ομιλία στο International Conference on Systems Biology of Human Disease στο Harvard το 2014. Το 2016 συνεργάστηκε με το U.S. Food and Drug Administration για την ανάπτυξη μιας νέας υπολογιστικής μεθόδου για την πρόβλεψη της καρδιομυοπάθειας που προκαλείται από φάρμακα με ακρίβεια 88%, έρευνα που δημοσιεύθηκε στο CPT: Pharmacometrics & Systems Pharmacology. Η εργασία του πάνω στη ΜΑΛΝΗ αποκάλυψε έναν πολυπαραγοντικό μηχανισμό σηματοδότησης και κέρδισε το 1ο βραβείο από 56 εργασίες στο συνέδριο της Εταιρείας Ήπατος-Παγκρέατος-Χοληφόρων το 2017.

Εκτεταμένη Περίληψη

Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων για την αποκρυπτογράφηση μηχανισμών σε ανθρώπινες ασθένειες

Ο σκοπός της παρούσας διδακτορικής διατριβής είναι να αποκρυπτογραφήσει μηχανισμούς σε ανθρώπινες ασθένειες, και συγκεκριμένα στην Οστεοαρθρίτιδα, στην μετάφραση βιολογικών δεδομένων από τον αρουραίο στον άνθρωπο, στην Σκλήρυνση κατά πλάκας, στον καρκίνο του ήπατος, στην καρδιομυοπάθεια που προκαλείται από φάρμακα και στην μη αλκοολική λιπώδη νόσο του ήπατος (ΜΑΛΝΗ). Τα κύρια ευρήματα σχετικά με την κάθε ανθρώπινη πάθηση και εργασία περιγράφονται παρακάτω.

Μέσω της έρευνας στην οστεοαρθρίτιδα αποδείχθηκε ότι τα υγιή χονδροκύτταρα μπορούν να έχουν ισχυρή φλεγμονώδη - και όχι προστατευτική - ανταπόκριση σε διάφορα ερεθίσματα. Με αυτό τον τρόπο, η δημιουργία ενός φλεγμονώδους περιβάλλοντος στην άρθρωση διατηρείται, επειδή ο χόνδρος είναι ένας ιστός χωρίς αγγεία, ο οποίος στερείται σημαντικών αντιφλεγμονωδών συστατικών του περιφερικού αίματος και αυτό μπορεί τελικά να οδηγήσει στην αποικοδόμηση του ιστού.
Για πρώτη φορά εντοπίστηκαν τα συμβάντα σηματοδότησης που οδηγούν στην αύξηση των προ-φλεγμονωδών σημάτων κατά την διέγερση του TLR, αποκαλύπτοντας δύο σημαντικές φλεγμονώδεις οδούς: Το DEFB1 δίνει σήμα μέσω του υποδοχέα του προς RAC1, στα MAPKs και τελικά ενεργοποιεί το HSP27. Το Flagellin μεταδίδεται μέσω του TLR5 σε MYD88 και στη συνέχεια συγχωνεύεται με το μονοπάτι IL1 μέσω των IRAK, TIFA, TRAF6 και τελικά ενεργοποιεί τα σήματα IKB, MAPK14 και HSP27.
Η διέγερση των χονδροκυττάρων με τους φλεγμονώδους μεσολαβητές IL1B και Flagellin οδηγεί επίσης σε υπερ-ενεργοποίηση των σχετικών με την ανάπτυξη σημάτων CREB και MAP2K1 και στην απελευθέρωση προ-αναπτυξιακών κυτοκινών, όλες συνδεδεμένες με την υπερτροφία των χονδροκυττάρων και την οστεογέννεση.

Η ισχυρή ομοιότητα μεταξύ μηνίσκου και αρθρικού ιστού παρατηρήθηκε κατά την διέγερση αυτών των ιστών με διάφορα ερεθίσματα, υποστηρίζοντας την υπόθεση ότι υπάρχει σημαντική επικοινωνία μεταξύ αυτών των δύο διαμερισμάτων του γόνατος και οι αντιφλεγμονώδεις θεραπείες πρέπει να λαμβάνουν υπόψη και τους δύο ιστούς. Διεξήχθη μεταβολική ανάλυση ανθρώπινου οστεοαρθριτικού αρθρικού υγρού και εντοπίστηκαν μεταβολίτες που συνδέονται με οστεοαρθρίτιδα.

Για το σκοπό της έρευνας ομοιοτήτων ανθρώπου / αρουραίου, δημιουργήθηκε μια βάση δεδομένων πολλαπλών στρωμάτων που περιλάμβανε δεδομένα phosphoproteomics, transcriptomics και cytokine release, προερχόμενα από φυσιολογικά βρογχικά επιθηλιακά κύτταρα ανθρώπου και αρουραίου τα οποία εκτέθηκαν παράλληλα σε περισσότερα από 50 διαφορετικά ερεθίσματα υπό πανομοιότυπες συνθήκες.
Σημαντικές οδοί σηματοδότησης διατηρήθηκαν μεταξύ ειδών ανθρώπων και αρουραίων με αποκλίνοντα μόνο απομονωμένα συστατικά. Μια εξαίρεση ήταν οι στόχοι των μεταγραφικών παραγόντων, που φαινόταν πιο δύσκολο να προβλεφθούν. Ο παράγοντας μεταγραφής CREB1 έδειξε την πιο παρόμοια συμπεριφορά, αλλά η σύνδεση από το RSK1 ήταν παρούσα μόνο στο ανθρώπινο δίκτυο, γεγονός που μπορεί να εξηγηθεί από το γεγονός ότι οι ανθρώπινες ισομορφές του RSK1 δεν είναι λειτουργικές (RSK2, RSK3, RSK4). Αντίθετα, αυτό πιθανότατα δεν συμβαίνει στους αρουραίους. Ο Zeniou et al. αναφέρει ότι τα γονίδια RSK1 και RSK3 του αρουραίου μπορεί να μην είναι σε θέση να αντισταθμίσουν πλήρως την έλλειψη λειτουργίας του RSK2.

Δημιουργήθηκε ένα τεράστιο σύνολο δεδομένων για τη μελέτη της Σκλήρυνσης κατά πλάκας (MS). Απομώθηκαν PBMCs από 250 δότες (190 MS, 60 Υγιείς) και στη συνέχεια εκτέθηκαν σε 20 ερεθίσματα που συμπεριλάμβαναν 4 MS φάρμακα και μετρήθηκε η απόκριση 17 φωσφοπρωτεϊνών (5 'και 25') και 22 εκκρινόμενων κυτοκινών (24 ώρες). Αυτή η προσέγγιση επέτρεψε τον χαρακτηρισμό των δικτύων σηματοδότησης με τρόπο συγκεκριμένο για τον κάθε ασθενή και την πρόβλεψη νέων στόχων για συνδυαστική θεραπεία για MS.

Ο συνδυασμός fingolimod είτε με αναστολέα TAK1 είτε με EGCG επικυρώθηκε επίσης σε ένα ζωικό μοντέλο.

Ανάλυση με χρήση 3 κυτταροσειρών 3 ηπατικού καρκίνου (HCC) παρουσίασε νέες μηχανιστικές αντιλήψεις για τις στοχευμένες αντιφλεγμονώδεις δράσεις τριών υποσχόμενων nutraceuticals, epigallocatechin gallate (EGCG), fisetin (FIS) και eryodictyol (ERI). Το EGCG ήταν ο πιο αποτελεσματικός διαμορφωτής της έκκρισης φλεγμονώδων κυτοκινών (που ακολουθείται από FIS και ERI) και τα κύτταρα HEP3B ήταν τα καλύτερα ανταποκρινόμενα. Παρά την προηγούμενη εκτενή βιβλιογραφία, αυτή ήταν η πρώτη μελέτη που δείχνει την εξαιρετική ικανότητα αυτής της ένωσης να μειώνει ταυτόχρονα ένα ευρύ φάσμα κυτοκινών που εκκρίνουν οι HCC.

Με μια εφαρμογή στην καρδιομυοπάθεια που προκαλείται από φάρμακα, αποδείχθηκε ότι η κατασκευή συγκεκριμένων μονοπατιών σηματοδότησης μπορεί να συλλάβει υπολογιστικά τον τρόπο δράσης ενός φαρμάκου και να αυξήσει την ακρίβεια της πρόβλεψης της καρδιομυοπάθειας από 79% σε 88%, σε σύγκριση με τη χρήση μόνο των transcriptomic δεδομένων.
Χρησιμοποιώντας Elastic Net, εντοπίστηκαν 33 πρωτεΐνες / γονίδια, που προβλέπουν καλύτερα την ταξινόμηση της τοξικότητας στην καρδιά προκαλούμενης από φάρμακα.
Τα microRNAs που ρυθμίζουν την έκφραση των 6 κορυφαίων predictors είναι διαγνωστικής αξίας για φυσική καρδιακή ανεπάρκεια ή επαγόμενη από Doxorubicin καρδιομυοπάθεια. Μεταξύ αυτών, για τα miR193-3p και miR26b-5p που ρυθμίζουν 4 και 3 predictors αντίστοιχα, αξίζει να διεξαχθούν κλινικές μελέτες προκειμένου να καθοριστεί εάν αυτά τα microRNAs είναι χρήσιμα ως in vivo βιοδείκτες για καρδιομυοπάθεια που προκαλείται από φάρμακα.

Παρατηρήθηκε ότι η ΜΑΛΝΗ έχει πολυπαραγοντική φύση και δεν υπάρχει ενιαία θεραπεία για όλους τους υποτύπους της, υπογραμμίζοντας την ανάγκη για συστηματική προσέγγιση και εξατομικευμένες θεραπευτικές παρεμβάσεις για την καλύτερη κατανόηση και θεραπεία της. Αυτή ήταν η πρώτη φορά που μια μελέτη στοχεύει στην κατανόηση της πολυπαραγοντικής φύσης της ΜΑΛΝΗ στο επίπεδο

σηματοδότησης με τη μελέτη 5 μοντέλων επαγωγής ΜΑΛΝΗ σε πρωτογενή ανθρώπινα ηπατοκύτταρα.
Τα αποτελέσματα επιβεβαίωσαν ένα μεγάλο εύρος βιβλιογραφικών ευρημάτων για μηχανισμούς σηματοδότησης στην ΜΑΛΝΗ. Επιπλέον, οι CHK2 και EPOR έχουν αναδειχθεί ως πιθανά σημαντικά σήματα που μπορεί να είναι ενδιαφέρον να μελετηθούν περαιτέρω αφού είναι επίσης σημαντικοί παράγοντες στην αναγέννηση του ήπατος.

Τα τελευταία 4 χρόνια, ο Δημήτρης Μεσσήνης έχει συγγράψει 6 δημοσιεύσεις σε επιστημονικά περιοδικά και 14 περιλήψεις - ανακοινώσεις σε διεθνή συνέδρια. Έχει μία δημοσίευση ως πρώτος συγγραφέας και έχει επιτύχει δείκτη h-index 6. Μέσω των προσπαθειών για την αποκρυπτογράφηση των μηχανισμών διαφορετικών ανθρώπινων ασθενειών, ο υποψήφιος διδάκτωρ έχει αποκτήσει ευρεία κατανόηση στη Συστημική μοντελοποίηση και ανάλυση μεγάλων δεδομένων, ανέπτυξε τις δικές του πειραματικές και υπολογιστικές μεθόδους και συνέβαλε σε αρκετά σημαντικά ευρήματα που περιγράφονται παραπάνω. Το έργο του στην Σκλήρυνση κατά πλάκας τον έκανε να επιλεχθεί μεταξύ 6 από 110 εργασιών για να δώσει μια σύντομη ομιλία στο International Conference on Systems Biology of Human Disease στο Harvard το 2014. Το 2016 συνεργάστηκε με το U.S. Food and Drug Administration για την ανάπτυξη μιας νέας υπολογιστικής μεθόδου για την πρόβλεψη της καρδιομυοπάθειας που προκαλείται από φάρμακα με ακρίβεια 88%, έρευνα που δημοσιεύθηκε στο CPT: Pharmacometrics & Systems Pharmacology. Η εργασία του πάνω στη ΜΑΛΝΗ αποκάλυψε έναν πολυπαραγοντικό μηχανισμό σηματοδότησης και κέρδισε το 1ο βραβείο από 56 εργασίες στο συνέδριο της Εταιρείας Ήπατος-Παγκρέατος-Χοληφόρων το 2017.

Καρδιομυοπάθεια προκαλούμενη από φάρμακα

Η ασθένεια

Οι κλινικές εκδηλώσεις των καρδιακών παρενεργειών που σχετίζονται με τα αντικαρκινικά φάρμακα είναι ποικίλες και μπορεί να κυμαίνονται από οξείες επαγόμενες καρδιακές αρρυθμίες έως παράταση διαστήματος Q-T, μεταβολές στη στεφανιαία αγγειοπάθεια με διαδοχική ισχαιμία του μυοκαρδίου, μυοκαρδίτιδα, περικαρδίτιδα, σοβαρή συστολική δυσλειτουργία και δυνητικά θανατηφόρα συγκοπή.[1]

Πειραματικό Μοντέλο

Βασική μελέτη (Καρδιομυοκύτταρα)

- Καρδιομυοκύτταρα που προέρχονται από ανθρώπινα επαγόμενα πολυδύναμα βλαστοκύτταρα (Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes)
- Καρδιομυοκύτταρα που προέρχονται από ανθρώπινα εμβρυϊκά βλαστοκύτταρα (Human Embryonic Stem Cell-Derived Cardiomyocytes)
- Πρωτογενή ανθρώπινα καρδιομυοκύτταρα, απομονωμένα από τις κοιλίες της καρδιάς ενηλίκων (Primary Human Cardiomyocytes, isolated from the ventricles of the adult heart)

Μελέτη επαλήθευσης (καρκινικέ κυτταρικές σειρές από CMAP)

- MCF7
- HL60
- PC3

Πειραματικά εργαλεία

Πρόκειται για υπολογιστικό έργο όπου δεν χρησιμοποιήθηκαν πειραματικά εργαλεία.

Υπολογιστικά εργαλεία

- Integer Linear Programming formulation για τον συνδυασμό δεδομένων έκφρασης γονιδίων με προηγούμενη γνώση συνδεσιμότητας πρωτεϊνών για την κατασκευή μονοπατιών σηματοδότησης.
- Elastic Net Regularization και άλλες μέθοδοι μηχανικής μάθησης.

Σημαντικά Αποτελέσματα

Η κατασκευή μονοπατιών σηματοδότησης μπορεί να καταγράψει υπολογιστικά τον τρόπο δράσης ενός φαρμάκου και να αυξήσει την ακρίβεια της πρόβλεψης από 79% σε 88%, σε σύγκριση με τη χρήση μόνο των δεδομένων έκφρασης γονιδίων. Αυτό πιθανότατα οφείλεται στην προηγούμενη γνώση των αλληλεπιδράσεων των βιολογικών πρωτεϊνών και τους στόχους των φαρμάκων που λαμβάνονται επίσης υπόψη για την κατασκευή του ειδικού δικτύου.

Χρησιμοποιώντας Elastic Net regularization, καταφέραμε να εξαγάγουμε 33 predictors που προβλέπουν καλύτερα την ταξινόμηση της τοξικότητας φαρμάκων που προκαλούν καρδιοτοξικότητας (είτε τοξικό για> 0,1% κλινική επίπτωση είτε μη τοξικό για <0,1%).

Τα microRNAs που ρυθμίζουν την έκφραση των έξι κορυφαίων predictors μας έχουν διαγνωστική αξία για την φυσική καρδιακή ανεπάρκεια ή την καρδιομυοπάθεια που προκαλείται από Doxorubicin. Μεταξύ αυτών, το miR193-3p και το miR26b-5p ρυθμίζουν περισσότερους predictors από τα άλλα μικροRNAs (τέσσερις και τρεις από τους κορυφαίους predictors μας, αντίστοιχα). Μπορεί να αξίζει να διεξαχθούν κλινικές μελέτες για να προσδιοριστεί εάν το miR193-3p και το miR26b-5p είναι χρήσιμοι in vivo βιοδείκτες για την καρδιομυοπάθεια που προκαλείται από φάρμακα.

Πρόβλεψη Καρδιομυοπάθειας που προκαλείται από φάρμακα με μεθόδους Μεταφραστικής Συστημικής Φαρμακολογίας

Περίληψη

Η καρδιομυοπάθεια που προκαλείται από φάρμακα συμβάλλει στα υψηλά ποσοστά αποτυχίας κατά την ανάπτυξη φαρμάκων. Συγκρίναμε δύο μεθόδους πρόβλεψης με μοντελοποίηση: (1) εφαρμογή Elastic Net (EN) σε διαφορικά εκφρασμένα γονίδια (Differentially Expressed Genes - DEG) μετά από εφαρμογή φαρμάκων. (2) εφαρμογή ακέραιου γραμμικού προγραμματισμού (Integer Linear Programing - ILP) για την κατασκευή σηματοδοτικών μονοπατιών για κάθε φαρμάκο ξεκινώντας από τους στόχους του φαρμάκου προς τις πρωτεΐνες του μονοπατιού, στους παράγοντες μεταγραφής και στα DEG τους σε ανθρώπινα καρδιομυοκύτταρα και στη συνέχεια υποβάλλοντας τα γονίδια / πρωτεΐνες στα δίκτυα σηματοδότησης σε EN regression. Κατατάξαμε 31 φάρμακα με διαθεσιμότητα DEG σε 13 τοξικά και 18 μη τοξικά φάρμακα με βάση κλινική ένδειξη καρδιομυοπάθειας 0,1%. Η βελτιωμένη με ILP μοντελοποίηση αύξησε την ακρίβεια πρόβλεψης από 79% σε 88% (ευαισθησία: 88%, ειδικότητα: 89%) με Leave-one-out cross validation (LOOCV). Τα μονοπάτια σηματοδότησης των φαρμάκων που κατασκευάστηκαν ήταν καλύτεροι predictors από τα DEG. Σύμφωνα με την βιβλιογραφία, τα microRNAs που αναφέρεται ότι ρυθμίζουν την έκφραση των έξι κορυφαίων predictors μας έχουν διαγνωστική αξία για φυσική καρδιακή ανεπάρκεια ή καρδιομυοπάθεια που προκαλείται από το Doxorubicin. Αυτό το μεταφραστικό μοντέλο πρόβλεψης μπορεί να αποκαλύψει πιθανούς βιοδείκτες.

Εισαγωγή

Σοβαρές και απειλητικές για τη ζωή ανεπιθύμητες ενέργειες προκαλούν την αποτυχία σε διάφορα στάδια της ανάπτυξης φαρμάκου ή κατά την τροποποίηση θεραπευτικών αγωγών. Για παράδειγμα, οι ανθρακυκλίνες, αν και είναι αποτελεσματικές για τη θεραπεία καρκίνων, είναι γνωστό ότι προκαλούν μη αναστρέψιμη, εξαρτώμενη από τη

δόση καρδιοτοξικότητα (τοξικότητα σχετιζόμενη με το contractility).[2] Πρόσφατα, η στοχευμένη θεραπεία με αναστολείς κινάσης τυροσίνης (Tyrosine Kinase Inhibitors - TKIs) επίσης προκάλεσε τέτοια τοξικότητα.[2] Η πρόβλεψη της προκαλούμενης από φάρμακα καρδιοτοξικότητας μπορεί να μειώσει το ποσοστό αποτυχίας κατά την ανάπτυξη φαρμάκων και να προωθήσει την ιατρική ακριβείας.

Έχουν γίνει προσπάθειες πρόβλεψης με μοντελοποίηση για ανεπιθύμητες αντιδράσεις φαρμάκων μέσω της ενσωμάτωσης πληροφοριών από βάσεις δεδομένων και βάσεις γνώσης των βιολογικών δραστηριοτήτων, της χημείας και των ανεπιθύμητων ενεργειών φαρμάκων.[3–5] Ωστόσο, δεν έχουν κατασκευαστεί μοντέλα πρόβλεψης της καρδιομυοπάθειας που προκαλείται από φάρμακα, χρησιμοποιώντας πληροφορίες σηματοδοτικών μονοπατιών. Ο Harpaz et al.[5] υπογράμμισε τη σημασία της αξιοποίησης πολλαπλών πηγών γνώσης, βιολογικών πληροφοριών και βιοϊατρικής βιβλιογραφίας για την πρόβλεψη της τοξικότητας φαρμάκων. Σύμφωνα με αυτή την ιδέα, στο παρόν μοντέλο πρόβλεψης ενσωματώνουμε προηγούμενη γνώση, στόχους φαρμάκων και εμπειρικά δεδομένα, ώστε το μοντέλο μας να μπορεί να προσδιορίσει βασικούς predictors από τον μηχανισμό δράσης ενός φαρμάκου και να έχει τη δυνατότητα να βοηθήσει τον προσδιορισμό και την ανάπτυξη νέων φαρμάκων.

Για να συμπληρώσουμε το κενό, καταρτίσαμε έναν κατάλογο 31 τοξικών και μη τοξικών φαρμάκων τα οποία έχουν εξεταστεί πάνω σε ανθρώπινα καρδιομυοκύτταρα και έχουμε το μεταγραφικό τους προφίλ[6–8]. Επιθεωρήσαμε την λίστα και προσθέσαμε την κλινική εμφάνιση της καρδιομυοπάθειας που σχετίζεται με το κάθε φάρμακο. Έπειτα διεξήγαμε μοντελοποίηση για να προβλέψουμε την προκαλούμενη από τα φάρμακα αυτά καρδιομυοπάθεια. Δύο μοντέλα πρόβλεψης συγκρίθηκαν: (1) εφαρμογή Elastic Net (EN) σε διαφορικά εκφρασμένα γονίδια (Differentially Expressed Genes - DEG) μετά από εφαρμογή φαρμάκων. (2) εφαρμογή ακέραιου γραμμικού προγραμματισμού (Integer Linear Programing - ILP) για την κατασκευή σηματοδοτικών μονοπατιών για κάθε φαρμάκο ώστε να αντικατοπτρίζει τον μηχανισμό δράσης του,[9] και στη συνέχεια υποβολή των κόμβων των σε EN regression. Ο αλγόριθμος ILP formulation[9] περιηγείται σε ένα δίκτυο προηγούμενης γνώσης για την επικοινωνία πρωτεϊνών-πρωτεϊνών, πρωτεϊνών-μεταγραφικών παραγόντων (TF) και αλληλεπιδράσεων γονιδίων-TF και ταυτοποιεί τις οδούς που συνδέουν τους

στόχους του φαρμάκου με τα DEG. Το ILP όχι μόνο βελτιστοποιεί τη λύση εύρεσης διαδρομών σηματοδότησης ενός φαρμάκου αλλά επίσης ενισχύει την απόδοση του μοντέλου πρόβλεψης, επιτρέποντας την ταυτοποίηση του υποσυνόλου DEGs που είναι λειτουργικά σχετικά με τον τρόπο δράσης του φαρμάκου. Αναφέρουμε επίσης την βιβλιογραφία για τα microRNAs που είναι διαγνωστικά για την καρδιακή ανεπάρκεια και για την καρδιομυοπάθεια που προκαλείται από φάρμακα, καθώς επίσης και για τη ρύθμιση της έκφρασης των predictors μας με την ελπίδα ότι θα ρίξουμε φως στα πιθανά microRNAs ως in vivo βιοδείκτες καρδιομυοπάθειας που προκαλείται από φάρμακα.

Μέθοδοι

Σύνταξη καταλόγου φαρμάκων και κλινική εμφάνιση καρδιομυοπάθειας που προκαλείται από φάρμακα.

Προκειμένου να καταρτίσουμε τον κατάλογο των εγκεκριμένων φαρμάκων που προκαλούν καρδιομυοπάθεια που σχετίζεται με κάποια θεραπεία, αναφερόμαστε στα National Institutes of Health Common Terminology Criteria for Adverse Events (έκδοση 4.03)[10] και στο Medical Dictionary for Regulatory Activities[11] για όρους που σχετίζονται με την καρδιομυοπάθεια ώστε να κάνουμε text-mine εγκεκριμένες ετικέτες φαρμάκων. Οι χρησιμοποιούμενοι όροι περιλάμβαναν καρδιομυοπάθεια, καρδιακή ανεπάρκεια, συμφορητική καρδιακή ανεπάρκεια, , δυσλειτουργία της αριστερής κοιλίας, αποτυχία της αριστερής κοιλίας και μείωση του κλάσματος εξώθησης της αριστερής κοιλίας (cardiomyopathy, heart failure, congestive heart failure, cardiac failure, left ventricular dysfunction, left ventricular failure, and reduction in left ventricular ejection fraction). Τα τρέχοντα αρχεία PDF της ετικέτας του κάθε φαρμάκου (Drugs@FDA[2]) υποβλήθηκαν σε ανάλυση με μία μέθοδο εξόρυξης κειμένου (text mining), όπως δημοσιεύθηκε προηγουμένως.[12] Οι μεμονωμένες συχνότητες εμφάνισης για καρδιομυοπάθεια αναγνωρίστηκαν με χειροκίνητη απομόνωση ετικετών φαρμάκων, δημοσιευμένες αναθεωρήσεις νέων φαρμάκων (Drugs@FDA), καθώς και δημοσιευμένες κλινικές μελέτες.

Μοντέλο πρόβλεψης

Ροή εργασιών και κυριώτερα σημεία της EN και ILP.

Όπως φαίνεται στο **Figure P6-1**, συγκρίναμε δύο μεθόδους πρόβλεψης με μοντελοποίηση: (1) εφαρμογή Elastic Net (EN) σε διαφορικά εκφρασμένα γονίδια (Differentially Expressed Genes - DEG) μετά από εφαρμογή φαρμάκων. (2) εφαρμογή ακέραιου γραμμικού προγραμματισμού (Integer Linear Programing - ILP) για την κατασκευή σηματοδοτικών μονοπατιών για κάθε φαρμάκο και στη συνέχεια υποβάλλοντας τα γονίδια / πρωτεΐνες στα δίκτυα σηματοδότησης σε EN regression. Το EN είναι χρήσιμο στα μοντέλα πρόβλεψης όταν οι predictors ξεπερνούν κατά πολύ τις παρατηρήσεις ενώ παράλληλα είναι σε θέση να εντοπίσει στατιστικά σημαντικούς predictors [13] Το EN regularization είναι χρήσιμο για την γονιδιωματική ανάλυση ευαισθησίας φαρμάκου στον καρκίνο.[14]

Εφαρμόσαμε την ILP σε στόχους DEG και πρωτεϊνών ενός φαρμάκου για να μοντελοποιήσουμε τον τρόπο δράσης του. Αυτά τα δύο επίπεδα πληροφοριών συνδέονται μέσω μεταγωγής σήματος, όπου το σήμα προέρχεται από στόχους φαρμάκων, διαδίδεται ενδοκυτταρικά μέσω ενός σύνθετου δικτύου σηματοδοτικών σημάτων, διέρχεται από το στρώμα των TFs και τελικά φθάνει το μεταγραφικό επίπεδο DEG. Σχεδιάσαμε τις αλληλεπιδράσεις στο δίκτυο γνώσης χρησιμοποιώντας τον λογικό φορμαλισμό,[15] ο οποίος προσδιόρισε το ελάχιστο υποσύνολο του δικτύου για να επιτύχει την επιθυμητή συνδεσιμότητα. Κατασκευάσαμε το συγκεκριμένο δίκτυο σηματοδότησης για κάθε φάρμακο χρησιμοποιώντας ILP, όπως δημοσιεύθηκε προηγουμένως.[9]

Το ILP θα ενισχύσει την προβλεπτική απόδοση επειδή έχει την ικανότητα να συλλάβει τις κυτταρικές αποκρίσεις σε ένα φάρμακο, να αναγνωρίσει το υποσύνολο σημαντικών λειτουργικών DEG και να βοηθήσει στη διαφοροποίηση μεταξύ των φαρμάκων και να μεταφράσει όλα αυτά σε βελτιωμένες επιδόσεις πρόβλεψης.

Ομαλοποίηση ονόματος φαρμάκου.

Τα ονόματα φαρμάκων αρχικά εξομαλύνθηκαν και προσδιορίστηκαν με το αναγνωριστικό του PubChem για να διασφαλιστεί η συνοχή κατά τη λήψη δεδομένων από το Map Connectivity Map (CMap),[16] το Drug Toxicity Signature Generation

Center (DToxS),[6] το Search Tool for Interactions of Chemicals (STITCH),[17] και την βιβλιογραφία.

Συγκέντρωση στόχων φαρμάκων

Συγκεντρώσαμε τους στόχους των μεμονωμένων φαρμάκων από το STITCH,[17] και τη βάση δεδομένων "chemicalprotein links" και επιλέξαμε μόνο ανθρώπινες πρωτεΐνες. Οι πρωτεΐνες ταυτοποιήθηκαν από το αναγνωριστικό SwissProt / EnsEMBL και μεταφράστηκαν σε γονιδιακά σύμβολα της επιτροπής ονοματολογίας γονιδίων HUGO,[18] χρησιμοποιώντας το πακέτο R biomaRt, προκειμένου να ταιριάζουν με τους κόμβους του δικτύου προηγούμενης γνώσης [19] Χρησιμοποιήσαμε το αρχείο του STITCH “interaction types for links”, από όπου προσδιορίσαμε τα φάρμακα ως ότι ενεργοποιούν ή αναστέλουν τις μεμονωμένες πρωτεΐνες-στόχους. Χρησιμοποιήσαμε μόνο εκείνα που συνδέουν δεσμούς μεταξύ ζευγών πρωτεϊνών-φαρμάκων με βαθμολογία απόδοσης 0,7.

Πηγές δεδομένων γονιδιακής έκφρασης και χειρισμός τους.

Όπου υπήρχαν διαθέσιμα δεδομένα ως Affymetrix probe IDs, τα probe IDs (Affymetrix GeneChip Human Genome U133A Array) μεταφράστηκαν σε HUGO Gene Nomenclature Committee gene symbols χρησιμοποιώντας το biomaRt package[20] και το hgu133a2[21] package στην γλώσσα προγραμματισμού R. Σε όλες τις λίστες γονιδίων διατηρήσαμε μόνο εκείνα τα γονίδια με fold change > 2 και P value < 0.05 σύμφωνα με ένα two-tailed, two-sample, unequal variance Student’s t-test, προσαρμοσμένο χωριστά για τις λίστες γονιδίων υπερ- και υπο-έκφρασης με διόρθωση Bonferroni (P value προσαρμοσμένη για πολλαπλές συγκρίσεις).

Ένας κατάλογος 75 φαρμάκων με προκαλούμενα από φάρμακο DEGs που είναι διαθέσιμα από καρκινικά κύτταρα [16] στο CMap χρησιμοποιήθηκαν για διερευνητική μοντελοποίηση (βλέπε **Table P6-S1**). Για τη διεξαγωγή ισχυρής προγνωστικής μοντελοποίησης, εξαντλήσαμε τη βιβλιογραφία και τις βάσεις δεδομένων και βρήκαμε έναν κατάλογο 31 φαρμάκων από τα οποία ήταν διαθέσιμα DEGs επαγόμενα από φάρμακα σε ανθρώπινα καρδιομυοκύτταρα και καρδιομυοκύτταρα που προέρχονται από βλαστοκύτταρα Οι δύο πηγές δεδομένων για διαταραχή της γονιδιακής έκφρασης σε καρδιομυοκύτταρα που προκαλείται από φάρμακα ήταν: (1) 30

φάρμακα από DToxS, όπου χρησιμοποιήθηκαν πρωτογενή ανθρώπινα καρδιομυοκύτταρα ενηλίκων, και (2) στοιχεία από τη βιβλιογραφία του Doxorubicin μελετημένο σε καρδιομυοκύτταρα που προέρχονται από ανθρώπινα βλαστοκύτταρα. Το μέγεθος κάθε συνόλου δεδομένων περιορίστηκε κυρίως από τη διαθεσιμότητα δεδομένων DEG. Για τα δεδομένα του DToxS, κατεβάσαμε τα δεδομένα έκφρασης γονιδίου επιπέδου δύο, υπολογίσαμε τα fold changes, κρατήσαμε μόνο τα DEGs με fold change > 2 και τα συγχωνεύσαμε από διαφορετικούς δότες, υπολογίζοντας το μέσο όρο των fold changes και αποκλείοντας οποιαδήποτε DEG με αντίθετες κατευθύνσεις έκφρασης.

Το Doxorubicin μελετάται ευρέως για τη δοσοεξαρτώμενη καρδιακή της τοξικότητα και συνήθως χορηγείται σε δόση 40–60 mg/m. Μετά από ενδοφλέβια χορήγηση 60 mg/m^2, η μέγιστη συγκέντρωση στο πλάσμα (C_{max}) ήταν 630 ng/mL (1,159 nM).[22] Βλέπε **Table P6-S2** για μερικές μελέτες των μεταγραφικών προφίλ του doxorubicin. Για τη μοντελοποίησή μας, συμπεριλάβαμε τα δεδομένα από καρδιομυοκύτταρα που προέρχονται από ανθρώπινα επαγόμενα πολυδύναμα βλαστοκύτταρα από τους Chaudhari et al.[8] και Burridge et al.[7] Συμπεριλάβαμε τα δεδομένα γονιδιακής έκφρασης από Burridge et al. στα 100, 1,000, and 10,000 nM και του Chaudhari et al. στα 156 nM (βλέπε Supplementary Table S3 at DOI: 10.1002/psp4.12272 για το σκεπτικό).

<u>Προσδιορισμός του τρόπου δράσης ενός φαρμάκου με χρήση ILP.</u>

Κατασκευάσαμε αρχικά ένα δίκτυο προηγούμενης γνώσης ως ικρίωμα για την κατασκευή ενός σηματοδοτικού δικτύου φαρμάκου με τη λήψη από την Reactome της τελευταίας έκδοσης (έκδοση 2015) των "Λειτουργικών αλληλεπιδράσεων που προέρχονται από το Reactome". Όπως δημοσιεύσαμε προηγουμένως,[9] συγχωνεύσαμε αυτές τις αλληλεπιδράσεις με παράγοντες μεταγραφής και καταλήξαμε με ένα δίκτυο που καλύπτει τα επίπεδα των πρωτεϊνών, των παραγόντων μεταγραφής και των γονιδίων, το οποίο περιείχε 64.801 αντιδράσεις, 2.585 πρωτεΐνες σηματοδότησης και 12.376 γονίδια. Εφαρμόσαμε το ILP για τη βελτιστοποίηση του δικτύου σηματοδότησης ενός φαρμάκου παρέχοντας ως είσοδο το ικρίωμα που αναφέρθηκε παραπάνω και τους στόχους του φαρμάκου.

Η διαμόρφωση ILP επιλύθηκε χρησιμοποιώντας το στούντιο βελτιστοποίησης IBM

ILOG CPLEX με στόχο την βελτιστοποίηση του δικτύου για κάθε φάρμακο. Με βάση τους περιορισμούς που μιμούνται τη μεταγωγή σήματος[23] και προσαρμόζοντας στην ειδική περίπτωση πολύ μεγάλων (> 10.000 κόμβων) δικτύων, ο αλγόριθμος ελαχιστοποίησε την αναντιστοιχία μεταξύ των δεδομένων των μετρήσεων έκφρασης γονιδίων και της τοπολογίας της προηγούμενης γνώσης των μονοπατιών. Η έξοδος ήταν το βέλτιστο δίκτυο σηματοδότησης για κάθε φάρμακο, προσδιορίζοντας τις μοριακές αλληλεπιδράσεις που φαίνονταν να είναι λειτουργικές με βάση την είσοδο των DEG και των στόχων των φαρμάκων. Ήμασταν σε θέση να επιλέξουμε το ελάχιστο μέρος ενός δικτύου προηγούμενης γνώσης για κάθε φάρμακο που θα μπορούσε να εξηγήσει τα δεδομένα μας. Δείτε το Supplementary Document-ILP at DOI: 10.1002/psp4.12272 για την κατανόηση του παραδείγματος του δικτύου σηματοδότησης του φαρμάκου Methotrexate, όπως συλλαμβάνεται από το ILP (επίσης στο **Figure P6-S1**).

<u>Συγκρίνοντας τα μοντέλα πρόβλεψης: εφαρμόζοντας EN στα DEGs ενός φαρμάκου vs. στο σηματοδοτικό μονοπάτι του φαρμάκου όπως προκύπτει από το ILP.</u>

Για να κατασκευαστεί ένας πίνακας για το EN regression, κάθε φάρμακο σημειώθηκε με 0 εάν είναι μη τοξικό και με 1 εάν ταξινομείται ως τοξικό. Κατατάξαμε τα φάρμακα αναφερόμενοι στις εγκεκριμένες ετικέτες με τα κριτήρια: «Συχνές ανεπιθύμητες ενέργειες είναι εκείνες που συμβαίνουν σε μία ή περισσότερες περιπτώσεις σε τουλάχιστον 1/100 ασθενείς. Σπάνιες ανεπιθύμητες ενέργειες είναι εκείνες που εμφανίζονται σε 1/100 έως 1/1000 ασθενείς. Σπάνια γεγονότα είναι εκείνα που συμβαίνουν σε λιγότερους από 1/1000 ασθενείς."Αναφερόμενοι στον ορισμό των σπάνιων συμβάντων που χρησιμοποιούνται στην επισήμανση των φαρμάκων και λαμβάνοντας υπόψη την κατανομή των κλινικών περιπτώσεων, τον αριθμό των φαρμάκων με διαθέσιμα δεδομένα γονιδιακής έκφρασης και την ετερογένεια των κλινικών μελετών, ταξινομήσαμε τα φάρμακα σε δύο κατηγορίες, τοξικά για συχνότητα >0,1% και μη τοξικά για συχνότητα <0,1%.

Δημιουργήθηκε μια στήλη "καρδιοτοξικότητας" με κλίμακα κλινικής επίπτωσης: 1 για "τοξική" και 0 για "μη τοξική". Κάθε στήλη αντιστοιχεί σε ένα μόνο γονίδιο που εκφράζεται σε τουλάχιστον μία από τις λίστες DEG. Στα επιμέρους δείγματα ενός φαρμάκου δόθηκε μια τιμή 1, -1 ή 0 για να αντικατοπτρίζει την προς τα πάνω

ρυθμισμένη, προς τα κάτω ρυθμισμένη ή μη αναφερόμενη έκφραση, αντίστοιχα (μέθοδος 1). Οι ίδιες συνθήκες εφαρμόστηκαν στους κόμβους στο δίκτυο σηματοδότησης ILP κάθε φαρμάκου (μέθοδος 2).

Κατά τη μοντελοποίησή μας χρησιμοποιήσαμε EN regression και πιο συγκεκριμένα ένα μοντέλο γραμμικής παρεμβολής με μια ποινή EN που προσδιορίστηκε χρησιμοποιώντας το πακέτο R glmnet.[24] Η κανονικοποίηση της EN ορίζεται από δύο παραμέτρους, alpha και lambda. Το EN είναι ένα μίγμα LASSO και ridge regression και συνδυάζει τους δύο όρους ποινής τους για την παράμετρο alpha. Όταν το alpha είναι 0, το EN δρα ως ridge regression και όταν το alpha είναι 1, το EN δρα ως LASSO. Στο EN, η παράμετρος λάμδα αντικατοπτρίζει τη συρρίκνωση των συντελεστών του μοντέλου. Όταν το lambda ισούται με 0, δεν γίνεται συρρίκνωση των συντελεστών του μοντέλου, αλλά οι συντελεστές μειώνονται προς το 0 (αν και δεν είναι ακριβώς ίσο με 0) όσο αυξάνεται η τιμή του. Δοκιμάσαμε ένα εύρος τιμών για το alpha από 0 έως 1 με βήμα 0.01 και επιλέξαμε αυτό που ελαχιστοποίησε το μέσο τετραγωνικό σφάλμα. Για εκείνη την τιμή alpha, επιλέξαμε την τιμή του lambda που έδωσε το ελάχιστο μέσο σφάλμα cross validation.

Για να επικυρώσουμε κάθε μοντέλο, χρησιμοποιήσαμε leave-one-out cross validation (LOOCV) αφήνοντας κάθε φορά εκτός τα δεδομένα ενός φαρμάκου (είτε DEG είτε δίκτυο σηματοδότησης που κατασκευάστηκε από το ILP) και το κάναμε αυτό σε ολόκληρη τη λίστα φαρμάκων. Κάθε φορά υπολογίσαμε την ακρίβεια, την ευαισθησία και την ειδικότητα για ένα μοντέλο πρόβλεψης και επιλέξαμε και αναφέραμε το μοντέλο με την υψηλότερη ακρίβεια μαζί με την ευαισθησία και την ειδικότητά του. Από το επιλεγμένο μοντέλο πρόβλεψης, προέκυψαν οι predictors (γονίδια / πρωτεΐνες) που προέβλεπαν καλύτερα την προκαλούμενη από φάρμακα καρδιοτοξικότητα. Το receiver operating characteristic (ROC) και οι καμπύλες precision-recall δημιουργήθηκαν χρησιμοποιώντας το αντίστοιχο πακέτο της R.

Μέθοδος 1 – εφαρμόζοντας EN στα DEGs.

Τα αποτελέσματα των 75 φαρμάκων με DEGs από την CMap συνοψίζονται στο Supplementary Document-CMap at DOI: 10.1002/psp4.12272. Μεταξύ αυτών των 75 φαρμάκων με τα DEG τους από το CMap, 24 φάρμακα ήταν τοξικά και τα υπόλοιπα

51 φάρμακα ήταν μη τοξικά.

Ένα model matrix κατασκευάστηκε χρησιμοποιώντας δεδομένα από τα καρδιομυοκύτταρα, με τις 34 παρατηρήσεις (κατάταξη τοξικότητας) ως σειρές και τις 15,016 μεταβλητές (έκφραση γονιδίων) ως στήλες. Το γραμμικό μοντέλο πρόβλεψης κατασκευάστηκε έχοντας ως είσοδο όλες αυτές τις μεταβλητές για την κανονικοποίηση EN. Δοκιμάστηκαν όλα τα πιθανά διαφορετικά σενάρια cutoff (δείτε το φύλλο "summary" στο Supplementary Table S4 at DOI: 10.1002/psp4.12272 για τα αποτελέσματα των δοκιμών psccm_34_gen_heart και τα λεπτομερή αποτελέσματα των 18 μοντέλων με διαφορετικά cutoffs στο φύλλο "9"). Για παράδειγμα, cutoff 10 σήμαινε ότι τρέξαμε το μοντέλο χρησιμοποιώντας μόνο εκείνα τα γονίδια που εκφράστηκαν σε τουλάχιστον 10 από τις 34 λίστες, πράγμα που σημαίνει ότι η ανάλυση άρχισε με 3.508 γονίδια, ενώ ένα cutoff των 15 ξεκίνησε την ανάλυση με τα γονίδια που εμφανίστηκαν σε τουλάχιστον 15 από τις 34 λίστες, δηλαδή 464 γονίδια χρησιμοποιήθηκαν ως cutoff.

Μέθοδος 2 – εφαρμόζοντας EN στους κόμβους - γονίδια/πρωτεΐνες των σηματοδοτικών μονοπατιών που κατασκευάστηκαν με ILP.

Αρχικά διεξήγαμε εξερευνητική μοντελοποίηση χρησιμοποιώντας μια λίστα 75 φαρμάκων με δεδομένα γονιδιακής έκφρασης που είναι διαθέσιμα στο CMap και καταλήξαμε στο συμπέρασμα ότι τα δίκτυα σηματοδότησης των φαρμάκων που προέρχονται από ILP ξεπέρασαν τα DEG τους κατά την εφαρμογή της EN κανονικοποίησης (βλέπε **Figure P6-S2** για ROC και precision-recall καμπύλες).

Μπορέσαμε να βρούμε τις λύσεις του ILP για φάρμακα με δεδομένα γονιδιακής έκφρασης σε καρδιομυοκύτταρα (Supplementary Table S5 at DOI: 10.1002/psp4.12272) εκτός από cefuroxime, domperidone και olmesartan. Αυτά τα τρία φάρμακα απομακρύνθηκαν από αυτή την άσκηση μοντελοποίησης. Στο τέλος, είχαμε 31 σηματοδοτικά δίκτυα από 28 φάρμακα (15 μη τοξικά φάρμακα και 13 τοξικά φάρμακα). Βλέπε Supplementary Table S5 at DOI: 10.1002/psp4.12272 για τους κόμβους γονιδίων / πρωτεϊνών στο δίκτυο σηματοδότησης κάθε μεμονωμένου φαρμάκου. Κατασκευάσαμε ένα model matrix για τα 31 μονοπάτια / δίκτυα σηματοδότησης χρησιμοποιώντας προφίλ γονιδιακής έκφρασης από

καρδιομυοκύτταρα και θέτοντας 1 αν ένας κόμβος οδού ήταν ρυθμισμένος προς τα πάνω, -1 αν ήταν ρυθμισμένος προς τα κάτω και 0 εάν δεν υπήρχε στο βελτιστοποιημένο δίκτυο σηματοδότησης ενός φαρμάκου. Βλέπε το φύλλο "summary" του Supplementary Table S4 at DOI: 10.1002/psp4.12272 για τα αποτελέσματα των δοκιμών psccm_34_ILP_heart και τα λεπτομερή αποτελέσματα των 31 μοντέλων με διαφορετικά cut-offs στο φύλλο "10."

Βιολογική σημασία των predictors

Ερευνήσαμε τη βιβλιογραφία για microRNAs που έχουν αποδειχθεί ότι είναι διαγνωστικοί δείκτες καρδιακής ανεπάρκειας και επίσης εμπλέκονται στη ρύθμιση της γονιδιακής έκφρασης. Χρησιμιοποιήσαμε εξόρυξη κειμένου στη βιβλιογραφία και στην miRTarBase,[25] μια βάση δεδομένων με πειραματικά επικυρωμένες αλληλεπιδράσεις microRNA-στόχων, για μια λίστα με microRNAs, τα οποία έχουν αναφερθεί να ρυθμίζουν την έκφραση των κορυφαίων predictors γονιδίων / πρωτεϊνών μας και επίσης έχουν ανιχνευθεί στην κυκλοφορία ασθενών με καρδιακή ανεπάρκεια με διαφορετικό βαθμό σοβαρότητας[26,27] ή ασθενών με καρδιομυοπάθεια που προκαλείται από το Doxorubicin.[28]

Αποτελέσματα

Η λίστα φαρμάκων και το προφίλ τοξικότητας

Ο κατάλογος των 31 φαρμάκων με τα κλινικά τους προφίλ της καρδιομυοπάθειας που σχετίζεται με τη θεραπεία συνοψίζεται στο **Table P6-1**. Διεξήχθη επίσης έρευνα βιβλιογραφίας για να συμπληρωθεί η κλινική συχνότητα εμφάνισης καρδιομυοπάθειας, αν οι εγκεκριμένες ετικέτες φαρμάκων και οι δημοσιευμένες αναθεωρήσεις εφαρμογών [2] δεν είχαν τέτοιες πληροφορίες. Μεταξύ των 31 φαρμάκων, υπήρχαν 13 τοξικά φάρμακα (41,9%) και υπήρχαν 18 μη τοξικά φάρμακα (59,1%). Για τα φάρμακα για τα οποία δεν αναφερόταν η τοξικότητα που σχετίζεται με την καρδιομυοπάθεια στις ετικέτες τους σε κανένα από τα τμήματα κλινικών μελετών, τις εμπειρίες μετά την κυκλοφορία, τις προειδοποιήσεις και τις

προφυλάξεις, ερευνήσαμε επίσης βιβλιογραφικές και δημοσιευμένες αναθεωρήσεις[2] για να καταλήξουμε στο συμπέρασμα ότι είναι μη τοξικά φάρμακα.

Μοντέλο πρόβλεψης

Μέθοδος 1 – εφαρμόζοντας EN στα DEGs.

Χρησιμοποιώντας LOOCV σε ολόκληρη τη λίστα των 30 φαρμάκων και των προφίλ έκφρασης γονιδίων τους, επιτύχαμε ακρίβεια 79% και precision 75%, με 80% ευαισθησία και 79% ειδικότητα όταν χρησιμοποιήσαμε αυτά τα γονίδια που εκφράστηκαν σε τουλάχιστον 11 από τις 34 υπογραφές (cutoff 11 στο φύλλο "9" του Supplementary Table S4 at DOI: 10.1002/psp4.12272). Τα αποτελέσματα της κανονικοποίησης της EN φαίνονται στο **Figure P6-2**a, **2**c, και τα γονίδια / πρωτεΐνες με μη μηδενικούς συντελεστές είναι PHF19, HSPA8, RIF1, CD46, MXRA7, RAB27A, TOMM20, MYO6 και CCNA2. Οι καμπύλες ROC και οι καμπύλες precision-recall φαίνονται στο **Figure P6-3** και **Figure P6-S2,** αντίστοιχα.

Μέθοδος 2 – εφαρμόζοντας EN στους κόμβους - γονίδια/πρωτεΐνες των σηματοδοτικών μονοπατιών που κατασκευάστηκαν με ILP.

Με την εφαρμογή του EN και του LOOCV, μπορέσαμε να αυξήσουμε την ακρίβεια στο 88%, με ευαισθησία 88% και ειδικότητα 89%, σε σύγκριση με τα αποτελέσματα από την EN regression των DEG (Supplementary Table S4 at DOI: 10.1002/psp4.12272). Η κανονικοποίηση EN φαίνεται στο **Figure P6-2**b, **2**d. Τα αποτελέσματα για τις δοκιμές psccm_34_ILP_heart βρίσκονται στο φύλλο "summary" και τα λεπτομερή αποτελέσματα των 31 μοντέλων με cutoff που κυμαίνεται από 1 (5.012 γονίδια / πρωτεΐνες σε τουλάχιστον 1 φάρμακο) έως 31 (5 γονίδια / πρωτεΐνες σε τουλάχιστον 31 υπογραφές δικτύου) βρίσκονται στο υπολογιστικό φύλλο "10" του Supplementary Table S4 at DOI: 10.1002/psp4.12272. Η υψηλότερη ακρίβεια, ευαισθησία και εξειδίκευση επιτεύχθηκαν σε cutoff 10 με 189 γονίδια / πρωτεΐνες από τουλάχιστον 10 δίκτυα σηματοδότησης φαρμάκων. Οι καμπύλες ROC και precision-recall φαίνονται στο **Figure P6-3** και **Figure P6-S2**, αντίστοιχα.

Καταλήξαμε στο συμπέρασμα ότι το EN-ILP (μέθοδος 2) ξεπέρασε μόνο το EN (μέθοδος 1) όταν εφαρμόστηκε στην ίδια ομάδα DEG.

Η σημασία στην καρδιά των κορυφαίων predictors

Χρησιμοποιώντας EN κανονικοποίηση, μπορέσαμε να εξαγάγουμε τους predictors - πρωτεΐνες / γονίδια που προβλέπουν καλύτερα την ταξινόμηση τοξικότητας της καρδιοτοξικότητας που προκαλείται από φάρμακα (είτε τοξικό για> 0,1% κλινική συχνότητα είτε μη τοξικό για <0,1%). Οι 33 predictors μαζί με τους μεμονωμένους τους συντελεστές συνοψίζονται στο **Table P6-2**. Το δίκτυο των κορυφαίων 15 predictors που επιλέγονται από το μοντέλο παρουσιάζεται στο **Figure P6-4**. Η καρδιακή συνάφεια αυτών των predictors εξετάστηκε και συνοψίστηκε στο Supplementary Table S6 at DOI: 10.1002/psp4.12272. Οι predictors που προσδιορίστηκαν από το EN-ILP αντανακλούσαν τους βασικούς κυτταρικούς βιολογικούς παράγοντες για καρδιοτοξικότητα που προκαλείται από φάρμακα. Η κανονικοποίηση EN στο μοντέλο μας επέλεξε τους predictors που πρόβλεψαν καλύτερα την προκαλούμενη από φάρμακα καρδιοτοξικότητα.

Χρησιμοποιήσαμε εξόρυξη κειμένου σε μια βάση δεδομένων για την τεκμηρίωση των microRNAs[25] για εκείνα που αναφέρεται ότι ρυθμίζουν τους κορυφαίους predictors μας και επίσης αναζητήσαμε στη βιβλιογραφία για να περιορίσουμε τη λίστα σε αυτά που έχουν αναφερθεί ότι έχουν διαγνωστική αξία για καρδιακή ανεπάρκεια. Συνοψίζονται στο **Table P6-3** οι κορυφαίοι 10 predictors και τα μεμονωμένα ρυθμιστικά microRNAs που έχουν αναφερθεί ότι έχουν διαγνωστική αξία για φυσική καρδιακή ανεπάρκεια[26,27] ή για καρδιομυοπάθεια που προκαλείται από το Doxorubicin.[28]

Συζήτηση

Με τα κλινικά συμβάντα της προκαλούμενης από φάρμακο καρδιομυοπάθειας ως εξαρτημένης μεταβλητής, η προσομοίωση EN-ILP αύξησε την ακρίβεια της πρόβλεψης από 79% σε 88%, σε σύγκριση με τη μοντελοποίηση με EN και μόνο με DEG. Αυτή η βελτιωμένη πρόβλεψη έδειξε την ικανότητα του ILP να συλλάβει υπολογιστικά τον τρόπο δράσης ενός φαρμάκου μέσω της κατασκευής των σηματοδοτικών μονοπατιών του με σκοπό την μοντελοποίηση για προβλέψεις. Το ILP

προσφέρει το πλεονέκτημα της ενσωμάτωσης των προηγούμενων γνώσεών μας σχετικά με τις αλληλεπιδράσεις βιολογικών πρωτεϊνών και τους στόχους φαρμάκων (Reactome και STITCH), τους παράγοντες μεταγραφής και τα DEGs σε προγνωστικά μοντέλα. Το ILP βελτιστοποιεί επίσης το μέγεθος του προφίλ-δικτύου ενός φαρμάκου. Λαμβάνοντας το Lapatinib ως παράδειγμα, είχε 2.265 DEGs από τα καρδιομυοκύτταρα, ενώ από αυτό το σύνολο DEG, το ILP δίκτυο του περιελάμβανε 1.923 κόμβους, συμπεριλαμβανομένων των στόχων του, πρωτεϊνών που εμπλέκονταν στην μεταγωγή σηματοδότησης, παράγοντες μεταγραφής και λειτουργικά DEGs.

Οι 33 predictors μαζί με τους μεμονωμένους θετικούς ή αρνητικούς συντελεστές θα μπορούσαν να χρησιμοποιηθούν για την πρόβλεψη της «τοξικότητας» ή «μη τοξικότητας» για ένα φάρμακο χρησιμοποιώντας τα επίπεδα γονιδιακής έκφρασης από το δίκτυο σηματοδότησης που κατασκευάστηκε με το ILP. Η προβλεπτική ισχύς του μοντέλου φαρμακολογίας αυτού θα αυξηθεί με το ποσό των δεδομένων στο σύνολο εκπαίδευσης.

Μεταξύ των 31 φαρμάκων που χρησιμοποιήθηκαν για τη διεξαγωγή προβλέψεων, η κατανομή των τοξικών (n513) έναντι των μη τοξικών (n518) ήταν αποδεκτή, αν και όχι ιδανική. Μεταξύ αυτών, υπήρχαν 18 αναστολείς κινάσης (17 TKIs και 1 αναστολέας κινάσης σερίνης / θρεονίνης), κάτι που φαίνεται να είναι εκτός ισορροπίας από την άποψη της ποικιλομορφίας της κατηγορίας φαρμάκων. Το vemurafenib είναι ένας αναστολέας κινάσης σερίνης / θρεονίνης και δεν είναι τοξικός. Η κατανομή των τοξικών (n = 8) και μη τοξικών (n = 9) φαρμάκων μεταξύ των 17 TKIs ήταν αποδεκτή. Οι TKIs, γενικά, δεν έχουν ειδικό στόχο, έχουν πολλαπλούς στόχους και έχουν σχεδιαστεί για να διαταράξουν τις οδούς σηματοδότησης που είναι ζωτικής σημασίας για την επιβίωση των καρκινικών κυττάρων.[29] Δυστυχώς, αρκετές από αυτές τις οδούς σηματοδότησης παίζουν επίσης έναν κρίσιμο ρόλο στη βιολογία των καρδιομυοκυττάρων[30]; Κατά συνέπεια, αρκετά TKIs εξασθενίζουν την καρδιακή λειτουργία. Σε αυτό το πλαίσιο, το μοντέλο πρόβλεψης που δημιουργήσαμε θα μπορούσε να είναι χρήσιμο για την πρόβλεψη καρδιακής τοξικότητας για μελλοντικές νέες χημικές οντότητες.

Όλοι οι 15 κορυφαίοι predictors έχουν σχετικές καρδιακές λειτουργίες εκτός του ZNP 823 (Supplementary Table S5 at DOI: 10.1002/psp4.12272). Είναι ενδιαφέρον το γεγονός ότι το CYP3A4 ήταν ένας σημαντικός predictor Αν και το CYP3A4 δεν έχει

βιολογικές αλληλεπιδράσεις με άλλους προγνωστικούς παράγοντες, όπως φαίνεται στο **Figure P6-4**, είναι ένα σημαντικό ένζυμο που μεταβολίζει φάρμακα. Μεταξύ των 31 φαρμάκων, 10 από τα 13 (85%) τοξικά φάρμακα και 11 από τα 18 (61%) μη τοξικά φάρμακα μεταβολίστηκαν από το CYP3A4. Τα τοξικά φάρμακα που μεταβολίζονται κατά κύριο ή εκτεταμένο τρόπο από το CYP3A4 συμπεριέλαβαν τα: amiodarone, axitinib, cytrabine, dasatinib, doxorubicin, imatinib, ponatinib, sorafenib, sunitinib, και vandetanib.[2,31–33] Για τα μη τοξικά φάρμακα, η λίστα είναι: bosutinib, crizotinib, cyclosporine, domperidone, erlotinib, gefitinib, lapatinib, regorafenib, ruxolitinib, tofacitinib, και ursodeoxycholic acid.[2,34]

Ορισμένοι κορυφαίοι predictors σχετίζονται βιολογικά με την focal adhesion kinase (FAK), μια τυροσινική κινάση που δεν είναι υποδοχέας και η οποία εμπλέκεται στο δίκτυο των σηματοδοτικών πρωτεϊνών που συνδέονται με το κυτταρόπλασμα.[35] Focal adhesion συμπλέγματα διαδραματίζουν κρίσιμο ρόλο στον τρόπο ανταπόκρισης των καλλιεργημένων καρδιομυοκυττάρων σε μηχανικά και των νευροσωμικά ερεθίσματα και στην ανάπτυξη της καρδιακής ανεπάρκειας.[36] Η ενεργοποίηση του FAK διαδραματίζει ένα ρόλο στην προσαρμοστική απάντηση σε cardiac afterload και στην ανάπτυξη των μυοκυττάρων μέσω της κινάσης πρωτεΐνης Β / στόχου θηλαστικών της οδού ραπαμυκίνης.[37] Η διάσπαση του FAK προκαλείται από την οικογένεια CASP3 η απόπτωση των ανθρώπινων φυσιολογικών κυττάρων,[38] και συμβαίνει με την ενεργοποίηση της EPHA2 και της p38 mitogen-activated protein kinase κατά τη διάρκεια της επαγόμενης από δοξαζοσίνη απόπτωσης μιας καρδιακής κυτταρικής γραμμής.[39] Η FAK ενεργοποιεί την STAT1 κατά την προσκόλληση κυττάρων,[40] και διαδραματίζει σημαντικό ρόλο στη μετανάστευση κυττάρων με μία από τις ενέργειές της να συνδέεται με το σύμπλεγμα σηματοδότησης υποδοχέα αυξητικού παράγοντα από αιμοπετάλια (PDGFR).[41] Εν ολίγοις, οι κορυφαίοι predictors είναι σημαντικοί για να διατηρηθεί ή φυσιολογική καρδιακή λειτουργία.

Σύμφωνα με τη βιβλιογραφία, ορισμένα microRNAs που ρυθμίζουν την έκφραση των predictors μας έχουν επίσης αποδειχθεί ότι είναι διαγνωστικής αξίας για καρδιακή ανεπάρκεια με διαφορετικό βαθμό σοβαρότητας (**Table P6-3**).[26,27] Μεταξύ αυτών, τα miR193-3p και miR26b-5p ρυθμίζουν περισσότερους predictors από ότι άλλα μικροRNAs, τέσσερις και τρεις από τους κορυφαίους μας, αντίστοιχα. Μπορεί να αξίζει να διεξαχθούν περαιτέρω κλινικές μελέτες για να προσδιοριστεί εάν το miR193-

3p και το miR26b-5p είναι χρήσιμοι in vivo βιοδείκτες για καρδιομυοπάθεια που προκαλείται από φάρμακα. Η αναζήτηση στην βιβλιογραφία αποκάλυψε μια πρόσφατη μελέτη που διερεύνησε κυκλοφορούντα microRNAs σε παιδιά με επαγόμενη από ανθρακυκλίνες οξεία καρδιακή βλάβη.[28] Αυξημένα miR-29b και miR-499 στην κυκλοφορία φάνηκε να συσχετίζονται με την αύξηση της τροπονίνης σε αυτά τα παιδιά και ταυτοποιήθηκαν ως δυνητικοί βιοδείκτες καρδιομυοπάθειας.[28] Αυτή η παρατήρηση της αύξησης του miR-29b στην καρδιομυοπάθεια που προκαλείται από το Doxorubicin διέφερε από την παρατήρηση της μειωμένης έκφρασης του miR-29b-3p στο αίμα των στεφανιαίων κόλπων των ασθενών με καρδιακή ανεπάρκεια.[27] Το MiR-29b-3p ρυθμίζει την έκφραση ενός από τους κορυφαίους 10 predictors, του PDGFR-A. Απαιτούνται περαιτέρω μελέτες για τη διερεύνηση του ρόλου του miR-29b στην καρδιομυοπάθεια που προκαλείται από φάρμακα ή στη φυσική καρδιακή ανεπάρκεια. Παρόλο που το miR-27b αναφέρεται ότι ρυθμίζει την CYP3A4,[25,42] η αναζήτηση βιβλιογραφίας δεν αποκάλυψε οποιεσδήποτε αναφορές που πρότειναν ότι το miR-27b είναι διαγνωστικής αξίας για καρδιομυοπάθεια που προκαλείται από φάρμακα.

Η ενσωμάτωση των κλινικών συμβάντων με τους τρόπους δράσης ενός φαρμάκου, που απεικονίζεται ως το δίκτυο σηματοδότησής του για πρόβλεψη μοντέλων είναι ένα δυνατό σημείο της μελέτης μας. Υπάρχουν, ωστόσο, ορισμένοι περιορισμοί στην προσέγγιση μας: (1) Τα μη τοξικά ήταν ελαφρώς περισσότερα από τα τοξικά φάρμακα. (2) ο περιορισμός του ILP, δηλαδή το ότι δεν εξετάστηκαν οι βρόγχοι βιολογικής ανατροφοδότησης και ότι έγιναν παραδοχές που υιοθετήθηκαν στη διατύπωση του προβλήματος, (3) Τα DEGs του Doxorubicin στα καρδιομυοκύτταρα ήταν από διαφορετικές πηγές από τα υπόλοιπα 30 φάρμακα και (4) η περιορισμένη διαθεσιμότητα μεταγραφικών δεδομένων σε καρδιομυοκύτταρα. Επιπλέον, η μελέτη μας κληρονόμησε τις αδυναμίες που σχετίζονται με τις βάσεις δεδομένων και τις βάσεις γνώσεων που χρησιμοποιήθηκαν για τη μοντελοποίησή μας. Η επίδραση των ενδείξεων της νόσου επί των κλινικών συμβάντων και της σοβαρότητας της καρδιομυοπάθειας που σχετίζεται με τη θεραπεία δεν είναι καλά χαρακτηρισμένη.

Το μοντέλο πρόβλεψης που δημιουργήσαμε το οποίο ενσωματώνει την κλινική επίπτωση της καρδιομυοπάθειας που προκαλείται από φάρμακα με το δίκτυο σηματοδότησης τοξικών και μη τοξικών φαρμάκων είναι όχι μόνο χρήσιμη για την

περαιτέρω βελτίωση της προγνωστικής ισχύος, αλλά επίσης εντοπίζει σημαντικούς predictors - γονίδια/πρωτεΐνες που έχουν σχετικές καρδιακές βιολογικές λειτουργίες. Πάνω απ 'όλα, οι κορυφαίοι predictors γονιδίων/πρωτεϊνών ρυθμίζονται σύμφωνα με την βιβλιογραφία από ειδικά microRNAs που έχουν αποδειχθεί ότι έχουν διαγνωστική αξία για καρδιακή ανεπάρκεια ή καρδιομυοπάθεια που προκαλείται από φάρμακα. Αυτοί οι predictors θα μπορούσαν να είναι χρήσιμοι για να φέρουν στο φως πιθανά microRNAs ως in νίνο βιοδείκτες της προκαλούμενης από φάρμακα καρδιομυοπάθειας.

Summary

Systems Biology and Big Data Modeling to decipher mechanisms in Human Disease

The purpose of this PhD Thesis is to decipher mechanisms in human disease, more specifically Osteoarthritis, Human/Rat Translational Research, Multiple Sclerosis, Hepatocellular Carcinoma, Drug-Induced Cardiomyopathy and Non-Alcoholic Fatty Liver Disease (NAFLD). Main findings regarding each human condition and project are discussed below.

Through research in Osteoarthritis it was demonstrated that healthy chondrocytes can have a strong inflammatory -rather than protective- response to various stimuli. In this manner, the generation of an inflammatory environment in the joint is sustained, because cartilage is an avascular tissue, lacking important anti-inflammatory components of peripheral blood and this can eventually lead to the degradation of the tissue.

For the first time, the signaling events that lead to the up-regulation of pro-inflammatory signals upon stimulation of the TLR were identified, uncovering two major inflammatory pathways: DEFB1 signals via its receptor to RAC1, to the MAPKs and ultimately activates HSP27. Flagellin signals through TLR5 to MYD88 and then merges with the IL1 pathway signaling through IRAK, TIFA, TRAF6 and activates the IKB, MAPK14 and HSP27 signals.

Stimulation of chondrocytes with inflammatory mediators IL1B and Flagellin also leads to over-activation of growth-related signals CREB and MAP2K1 and the release of pro-growth cytokines, all connected to facilitating chondrocyte hypertrophy and bone ossification.

Strong similarity between meniscus and cartilage cytokine releases was observed upon stimulation of those tissues with various stimuli, supporting the hypothesis that significant crosstalk between these two knee compartments exists and anti-inflammatory therapies should take into consideration both tissues.

Metabolomic analysis of human osteoarthritic synovial fluid was performed and metabolites connected to Osteoarthritis were identified.

For the purpose of Human/Rat translational research, a multi-layer systems biology dataset was generated that was comprised of phosphoproteomics, transcriptomics and cytokine data derived from normal human and rat bronchial epithelial cells exposed in parallel to more than 50 different stimuli under identical conditions.
Major signaling pathways were conserved between human and rat species with only isolated components diverging. An exception was the targets of transcription factors, which seemed more difficult to predict.
Transcription factor CREB1, showed the best consensus for the edges upstream of it but the connection from RSK1 was present only in the human consensus network, which might be explained by the fact that human isoforms of RSK1 have functional redundancy (i.e. RSK2, RSK3, RSK4). In contrast, this is most likely not the case in rodents; Zeniou et al. reported that the mouse RSK1 and RSK3 genes may not be able to fully compensate for the lack of RSK2 function.

A huge dataset was created to study Multiple Sclerosis (MS). 250 donors were recruited (190 MS, 60 Healthy) and PBMCs were isolated from them, then treated with 20 stimuli including 4 MS drugs and the response of 17 phosphoproteins (5' and 25') and 22 secreted cytokines (24h) was measured. This approach allowed to characterize the signaling networks in a patient-specific manner and to predict new targets for combination therapy for MS.
The combination of fingolimod with either a TAK1 inhibitor or EGCG was also validated in an animal model.

Analysis using 3 Hepatocellular Carcinoma (HCC) cell lines presented new mechanistic insights into the targeted anti-inflammatory actions of three promising nutraceuticals, epigallocatechin gallate (EGCG), fisetin (FIS), and eriodictyol (ERI). EGCG was the most effective modulator of inflammatory cytokine secretion (followed by FIS and ERI) and HEP3B cells were the best responders. Despite previous

extensive literature, this was the first study showing the outstanding capability of this compound to concurrently reduce a wide range of HCC-secreted cytokines.

With an application on Drug-Induced Cardiomyopathy, it was shown that constructing specific signaling pathways can computationally capture a drug's mode of action and increase cardiomyopathy prediction accuracy from 79% to 88%, compared to just using the transcriptomic data at hand.
Using Elastic Net regularization, 33 protein/gene predictors were extracted, that best predict the toxicity classification of drug-induced cardiotoxicity.
The microRNAs that reportedly regulate expression of the 6 top predictors are of diagnostic value for natural heart failure or doxorubicin-induced cardiomyopathy. Among them, miR193-3p and miR26b-5p reportedly regulated 4 and 3 predictors, respectively, therefore it might be worthy of clinical studies to determine whether those micro RNAs are useful in vivo biomarkers for drug-induced cardiomyopathy.

It was observed that Non-Alcoholic Fatty Liver Disease (NAFLD) has a multifactorial nature and there is no single treatment for all subtypes of NAFLD, highlighting the need for a systemic approach and personalised therapeutic interventions to better understand and treat NAFLD. This was the first time that a study aims to understand the multifactorial nature of NAFLD at the signaling level by studying 5 NAFLD induction models in primary human hepatocytes.
The results confirmed a large body of literature findings for NAFLD signaling mechanisms. Furthermore, CHK2 and EPOR have emerged as potential NAFLD players that may be interesting to study further since they are important factors in liver regeneration.

Over the last 4 years, Dimitris Messinis has co-authored 6 peer-reviewed publications and 14 international conference abstracts. He has one first-author publication and has attained an h-index of 6. Through the efforts to decipher the mechanisms of different human conditions, the PhD candidate has gained a broad understanding in Systems Biology and Big Data Modeling, developed his own experimental and computational methods and contributed to several important findings described

above. His work in Multiple Sclerosis got him invited among 6 out of 110 abstracts to give a lightning talk at the International Conference on Systems Biology of Human Disease at Harvard in 2014. In 2016, he worked with the U.S. Food and Drug Administration developing a novel computational method to predict drug-induced Cardiomyopathy with 88% accuracy published in CPT: Pharmacometrics & Systems Pharmacology. His Cue-signal-response analysis in NAFLD revealed a multifactorial signaling mechanism and won the 1st prize among 56 submissions at the Panhellenic conference of Hepato-Pancreato-Biliary Association in 2017.

Osteoarthritis

The disease

Globally, it is estimated that 242 million people have osteoarthritis.
Osteoarthritis (OA) is the single most common cause of disability in older adults.

OA is a long-term chronic disease characterized by the deterioration of cartilage in joints which results in bones rubbing together and creating stiffness, pain, and impaired movement. The disease most commonly affects the joints in the knees, hands, feet, and spine and is relatively common in shoulder and hip joints. While OA is related to ageing, it is also associated with a variety of both modifiable and non-modifiable risk factors, including: obesity, lack of exercise, genetic predisposition, bone density, occupational injury, trauma, and gender.[43]

In normal joints, a firm, rubbery material called cartilage covers the end of each bone. Cartilage provides a smooth, gliding surface for joint motion and acts as a cushion between the bones. In OA, the cartilage breaks down, causing pain, swelling and problems moving the joint. As OA worsens over time, bones may break down and develop growths called spurs. Bits of bone or cartilage may chip off and float around in the joint. In the body, an inflammatory process occurs and cytokines and enzymes develop that further damage the cartilage. In the final stages of OA, the cartilage wears away and bone rubs against bone leading to joint damage and more pain.

Osteoarthritis and Metabolomics

The best OA biomarker candidates are generally molecules or molecular fragments present in cartilage, bone or synovium and may be specific to one type of joint tissue or common to them all.
Many currently investigated biomarkers are associated with collagen metabolism in cartilage or bone, or aggrecan metabolism in cartilage. Other biomarkers are related

to non-collagenous proteins, inflammation and/or fibrosis. Future avenues for research include exploration of underlying mechanisms of disease and development of new biomarkers; technological development; the 'omics' (genomics, metabolomics, proteomics and lipidomics); design of aggregate scores combining a panel of biomarkers and/or imaging markers into single diagnostic algorithms; and investigation into the relationship between biomarkers and prognosis.[44]

Complete metabolomic analysis could greatly benefit clinical Osteoarthritis (OA) by providing both diagnostic and prognostic biomarkers, as well as elucidate upon the underlying pathophysiological processes.[45]

Metabolomics has been employed to detect metabolic perturbations in the urine, blood, synovium, and synovial fluid (SF) of animal models and patients with OA. Many of the studies of venous plasma or urine identify metabolites that may be related to aging, altered muscle mass, and other factors that may confound the unique signature of a pathologic OA joint. For these reasons, SF may yield the most accurate, real-time, and joint-specific metabolic profile.[46]

OA Synovial Fluid Metabolic Profile

In 2014, a targeted metabolomics study[47] testing osteoarthritic Synovial Fluid (SF) using Biocrates AbsoluteIDQ p180 kit and UPLC MS FIA was the first to demonstrate that OA is an heterogeneous disease, consisting of at least three metabolically distinct subgroups, which are likely due to the differences in carnitine, lipid and collagen metabolism. More specifically, between the groups of patients there were significant differences in the ratio of acylcarnitines to free carnitine (C0), the concentrations of 24 glycerophospholipids and the concentrations of 9 sphingolipids (6 SM, 3 SM(OH)). The limitation was the absence of healthy SF samples.

SF metabolomics in different forms of arthritis were assessed by NMR in a study[48] analysing SF samples from patients with OA, gout, calcium pyrophosphate disease, spondylarthritis, septic arthritis and rheumatoid arthritis. The abstract reports that a

distinctive metabolism is observed in septic arthritis whereas metabolites in OA are similar to those in inflammatory arthritis.

A similar approach was performed in a study[49] using usual biochemical tests and proton magnetic resonance spectroscopy (MRS) analysis, where OA samples showed a characteristically increased level of N-glucosamine ($p < 0.04$) and creatine ($p < 0.04$) compared to other forms of arthritis (RA, gout, seronegative spondylarthropathies and septic arthritis).

In a most recent study[50], in 2015, testing both normal and OA SF samples, 11 metabolites were found to be significantly responsible for the separation between OA samples and the controls in the OPLS-DA model based on an NMR/GC-MS combined dataset. Fructose and citrate were increased in OA SF samples, while O-acetyl-carnitine, N-phenylacetylglycine, methionine, ethanol, creatine, malate, ethanolamine, 3-hydroxybutyrate and hexanoylcarnitine were decreased in OA compared to normal SF.

Examining the role of proteomics in OA pathogenesis[51], the proteins ALFA (Fructose-biphosphate aldolase A) and ENOA (Alpha enolase) are identified to play a role in cellular metabolism through the glycolysis pathway, specifically 'fructose biphosphate breakage' and 'pyruvate synthesis' respectively.

OA Synovial Tissue Metabolic Profile

End-stage OA versus early/no OA cultured synovial tissue samples were examined in a study[52] with the following results:
The metabolites (and their associated pathways) increased in end-stage OA were: Pro-hydroxyproline (Collagen degradation), Acetylcarnitine (Lipid, carnitine, TCA cycle), Myo-inositol (Lipid, inositol), N-acetylornithine (Urea cycle, AA), Succinate (TCA cycle), Glutamine (AA, TCA cycle), Urea (Urea cycle, AA), Beta-alanine (AA, pyrimidine degradation), Uracil (Pyrimidine), Arabitol (Carbohydrate), Catechol sulfate (Benzoate), P-cresol sulfate (AA).

The metabolites decreased in end-stage OA were: Gamma-glutamylleucine (Peptide, glutathione), 4-methyl-2-oxopentanoate (BCAA), 5-oxoproline (AA, glutathione), Phenylacetlyglycine (AA), 3-methyl-2-oxobutyrate (BCAA), Ornithine (Urea cycle, AA), Pyridoxate (Vitamin B6), 2-methylbutyroylcarnitine (BCAA).

OA Serum Metabolic Profile

Two sera BCAA metabolite ratios distinguish OA in the knee, in particular, the ratio of valine to histidine and xleucine to histidine differentiate OA from healthy controls with a p-value of 0.002.[53]

Healthy Synovial Fluid donors

Most studies do not use healthy controls when studying SF. (71% of the searched literature). The ones who do, collect healthy SF through 3 main ways: post mortem, non-OA controls with other knee disorders and healthy volunteers. Some examples are: 4h post mortem[50], 24h post mortem[54], patients undergoing ligament or meniscal repair with little or no evidence of OA[52], contralateral uninjured knees of patients undergoing unilateral ACL reconstruction, normal subjects undergoing arthroscopy, one healthy volunteer[55], asymptomatic individuals without radiographic OA paid to undergo arthrocentesis[56], non-OA controls with other knee disorders: meniscus injury or discoid meniscus without cartilage defect[57], donors with no prior history within the past 8 weeks of shoulder surgery, blood dyscrasias, cancer, chondrocalcinosis or corticosteroid injection[58], healthy volunteers[59]

Biomarkers in Arthroplasty

Increased / Decreased based on concentration compared to controls.

Osteolysis

URINE

Increased

Cross-linked N-telopeptides of type I collagen [60]

SERUM

Decreased

IL-11 [61]

Increased

C-terminal telopeptides of type I collagen [62]

Aseptic Prosthetic Loosening

URINE

Increased

Cross-linked N-telopeptides of type I collagen [63] [64]

Deoxypyridinoline [63] (but [64] found no significant difference)

Pyridinoline [63]

SERUM / PLASMA

Decreased

C-terminal propeptide of type I procollagen (PICP) [65]

Increased

Cross-linked N-terminal telopeptide (NTx) [65]

ICTP [66] [67] (but [64] found no significant difference)

IL6 [68] (but [61] found no significant difference)

TRAP5b [67]

TNFA [69] (but [70] found no significant difference) (plasma)

IL1B [69] (but [70] found no significant difference)

GMCSF [70]

Osteocalcin [63] (but [64] found no significant difference)

Osteoprotegerin (OPG) [71]

RANKL [71]

hyaluronic acid [72]

IL8 [69] (plasma)

SYNOVIAL FLUID

Increased

IL1B [73] [74] (both studies compared to OA)

TNFA [73]

YKL-40 [75] (compared to OA)

TRAP [74] (compared to OA)

IL8 [76] (compared to OA)

IL10 [76] (compared to OA)

Decreased

Osteocalcin [76] (compared to OA)

B-crosslaps [76] (compared to OA)

bone ALP [76] (compared to OA)

SNPs

Increased

GNAS1, TNF-238 A allele, TNF-α promoter (-308G→A) transition, IL6-174 G allele, interleukin (IL)-6 (-597) and (-572), MMP-1-promoting gene, C/C genotype for the MMP1, MT1-MMP, MMP-2, transforming growth factor-beta1 signal sequence (29T→C) transitions, A/A genotype for the OPG-163, and MBL [77]

Prosthetic Infections

C-reactive protein, erythrocyte sedimentation rate, and interleukin-6

SERUM

Increased

CRP [78] [79] [80]

Erythrocyte sedimentation rate (ESR) [78] [79] [80]

IL6 [81]

IL6 + CRP [82]

IL6 [78 79] (compared to aseptic prosthetic revisions)

TNFA [79] (compared to aseptic prosthetic revisions)

Procalcitonin [79] (compared to aseptic prosthetic revisions)

SYNOVIAL FLUID

Increased

IL6 [80 81]

TNFA [80]

IL1B [80]

Conclusions

Osteolysis markers: Increased Cross-linked N-telopeptides of type I collagen (**NTx**) in urine and C-terminal telopeptides of type I collagen (**CTX**) in serum as well as decreased IL-11 in serum

Loosening markers: Increased NTx and Pyridinoline in urine, increased NTx, PICP, TRAP5b, GMCSF, Osteoprotegerin, RANKL, Hyaluronic acid, IL8 in serum, increased IL1B, TNFA, YKL-40, TRAP, IL8, IL10 and decreased Osteocalcin, B-crosslaps, bone ALP in synovial fluid when compared to OA patients. Several cytokines have inconsistent associations

Prosthetic joint infection markers: **CRP** and **ESR** *[both of which are currently used clinically in the diagnosis and follow-up of patients with prosthetic joint infections*[83]*]* compared to aseptic joint revision controls

Experimental Models

- Cell culture of Primary Human Chondrocytes
- Human Synovial Fluid
- Human Plasma
- Human Meniscus Tissue
- Human Cartilage Tissue

Experimental Tools

- Cartilage Tissue, Synovial Fluid and Plasma collection from patients
- Chondrocyte isolation and cell culture
- Stimulation with 78 compounds, lysis and collection of cell supernatant
- Measurement with bead-based sandwich ELISA of 17 phosphoproteins and 55 cytokines
- Gas Chromatography - Mass Spectrometry (GC-MS) Metabolomics

Computational Tools

- Integer Linear Programming formulation to combine proteomic data at hand with prior knowledge of proteins' connectivity to construct specific signaling pathways
- Gas Chromatography - Mass Spectrometry analysis to identify metabolites

Main Findings

We demonstrated that healthy chondrocytes can have a strong inflammatory response to various stimuli (rather than protective). In this manner, the generation of an inflammatory environment in the joint is facilitated and given that cartilage is an avascular tissue, lacking important anti-inflammatory components of peripheral blood[84], this environment is sustained, eventually leading to the degradation of the tissue.

For the first time, we were able to identify the signaling pathways of underreported pro-inflammatory mediators such as DEFB1 and Flagellin (TLR ligands). Even though the role of TLR in cartilage inflammation has been studied before, this is the first attempt to identify the signaling events that lead to the up-regulation of NFKB, HSP27 and other pro-inflammatory signals upon stimulation of the TLR. Two major

inflammatory pathways were uncovered. We identified that DEFB1 signals via its receptor to RAC1, to the MAPKs and ultimately activates HSP27. Flagellin signals through TLR5 to MYD88 and then merges with the IL1 pathway signaling through IRAK, TIFA, TRAF6 and activates the IKB, MAPK14 and HSP27 signals.

Stimulation with IL1B and Flagellin (two major inflammatory mediators) apart from up-regulation of inflammatory signals, also leads to over-activation of CREB and MAP2K1, two growth related signals that are connected to hypertrophy[85,86]. IL1B and Flagellin apart from directly activating pro-growth signals, also lead to the release of pro-growth cytokines, facilitating chondrocyte hypertrophy and bone ossification indirectly.

Strong similarity (up to 73%) between meniscus and cartilage cytokine releases was observed upon stimulation of those tissues with various stimuli. Our results confirm that IL-1a, IL-1b and TNF-a, as proinflammatory cytokines, promote systemic inflammation and thus release of IL-6, IL-8 and GROa in both cartilage and meniscus tissue. The above findings suggest that meniscus is affected by its inflammatory environment and responds to it as actively as cartilage, supporting the hypothesis that significant crosstalk between these two knee compartments exists and anti-inflammatory therapies should take into consideration both tissues.

We performed metabolomic analysis of human osteoarthritic synovial fluid and identified several metabolites that are also mentioned in Osteoarthritis literature.

Publication Timeline

[P1] 10 February 2014, Osteoarthritis and Cartilage
Modeling of signaling pathways in chondrocytes based on phosphoproteomic and cytokine release data

[A1] 17-20 May 2014, Prague, Czech Republic
Meniscus – Cartilage paracrine crosstalk in osteoarthritis

[A4] 18-20 September 2014, Patras, Greece
An integrated proteomic and metabolomic approach to investigate cartilage degeneration

[A6] 30 April 3 May 2015, Seattle, WA, USA
The role of inflammatory mediators in meniscus and cartilage crosstalk in osteoarthritis

Co-authored Research: Modeling of signaling pathways in chondrocytes based on phosphoproteomic and cytokine release data

Please visit DOI of paper [P1] for more information on this chapter.

Abstract

Objective: Chondrocyte signaling is widely identified as a key component in cartilage homeostasis. Dysregulations of the signaling processes in chondrocytes often result in degenerative diseases of the tissue. Traditionally, the literature has focused on the study of major players in chondrocyte signaling, but without considering the cross-talks between them. In this paper, we systematically interrogate the signal transduction pathways in chondrocytes, on both the phosphoproteomic and cytokine release levels.
Methods: The signaling pathways downstream 78 receptors of interest are interrogated. On the phosphoproteomic level, 17 key phosphoproteins are measured upon stimulation with single treatments of 78 ligands. On the cytokine release level, 55 cytokines are measured in the supernatant upon stimulation with the same treatments. Using an Integer Linear Programming (ILP) formulation, the proteomic data is combined with a priori knowledge of proteins' connectivity to construct a mechanistic model, predictive of signal transduction in chondrocytes.

Results: We were able to validate previous findings regarding major players of cartilage homeostasis and inflammation (e.g., IL1B, TNF, EGF, TGFA, INS, IGF1 and IL6). Moreover, we studied pro-inflammatory mediators (IL1B and TNF) together with pro-growth signals for investigating their role in chondrocytes hypertrophy and highlighted the role of underreported players such as Inhibin beta A (INHBA), Defensin beta 1 (DEFB1), CXCL1 and Flagellin, and uncovered the way they cross-react in the phosphoproteomic level.
Conclusions: The analysis presented herein, leveraged high throughput proteomic data via an ILP formulation to gain new insight into chondrocytes signaling and the pathophysiology of degenerative diseases in articular cartilage.

Introduction

Articular cartilage is a connective tissue covering the ends of bones in a joint, responsible for bearing loads with minimum wear and friction. Cartilage homeostasis is orchestrated by a complex interplay of anabolic and catabolic processes that take place in chondrocytes. Chondrocytes are meant to maintain the structure of the tissue by synthesizing collagen (mostly of type II) and proteoglycans. However, in pathological situations, they release matrix metalloproteinases (MMPs) that degrade the collagen content of the tissue, leading to loss of its structural integrity. Protein signaling plays a central role to chondrocytes capability to either synthesize cartilage and maintain its homeostasis or degrade cartilage and promote the inflammatory response seen in pathologic situations. For example, up-regulation of the SOX9 transcription factor induced by TGFB or FGF stimulation leads to collagen synthesis. On the other hand, over-activation of NFKB induced by several pathways (e.g., Inflammation related pathways or bone development processes[87–93]) leads to the release of MMPs and collagen degradation. Thus, chondrocytes have the potential for both anabolic and catabolic roles in articular cartilage, and the role they eventually assume is the one dictated by their signaling mechanisms as a response to their biochemical microenvironment.
The importance of protein signaling in cartilage homeostasis has been interrogated before, with most of the studies focusing on the deconvolution of signaling processes

in chondrocytes and the identification of catabolic mediators. Traditionally, most of these approaches revolve around few major pathways (such as IL1A/B, TNFA, TGFA/B, etc.), without taking into account other less known players. Considering that chondrocyte function and response results as an aggregate of numerous processes and that even the slightest crosstalk can eventually affect cell behavior, the systematic study of chondrocytes signaling is of the utmost importance for uncovering the etiology of degenerative diseases and facilitate the development of novel therapies[94–97]. To this end, high throughput proteomic measurements combined with computational modeling offer a promising solution. Proteomics technologies allow the multiplexed quantification of proteins, while computational modeling post-processes the results in a way that interpretable conclusions can be extracted for the interrogated system.

Proteomics have been used in the study of degenerative diseases of cartilage in the past, addressing mostly the following: (1) direct analysis of cartilage protein content[98–101], (2) analysis of cartilage related biological fluids[102,103] (Synovial fluid) and (3) study of chondrocytes secretion upon treatment with catabolic mediators[104,105]. In more detail, proteomic analysis of cartilage explants and chondrocytes has led to the identification of hundreds of proteins in articular cartilage, as well as characterization of their expression patterns in normal (control) and disease patients. Findings of this analysis lead to better understanding of the etiology underlying degenerative diseases and potential drug targets. Study of chondrocyte secretory behavior may lead to deeper understanding of the cells' plasticity to mount and overcome an inflammatory response; while proteomic analysis of the synovial fluid and plasma has identified proteins differentially regulated in normal and dis- ease patients and aims mostly at biomarker discovery.

The applications of proteomics mentioned above, form the basis towards a systems-level understanding of the processes taking place in chondrocytes; however, they do not interrogate their signaling mechanisms. To study chondrocytes signaling, phosphoproteomic data must also be incorporated and by implementing a mathematical formalism to model how signal propagates from one protein to the next, construct predictive models of their function.

In this paper, we interrogate the signal transduction mechanisms of primary

chondrocytes on both the phosphoproteomic and the cytokine release levels upon stimulation with 78 ligands, including some major players of osteoarthritic pathophysiology, as well as ligands well characterized for promoting inflammation in other cells types. Our approach is two-fold. On the experimental front, we use the xMAP technology to measure the activation level of 17 key phosphoproteins and the release of 55 cytokines in the cells' supernatant, upon stimulation with single treatments of the 78 ligands. Even though the xMAP technology does not provide for signal multiplexability as high as other proteomic technologies, fast turnaround times using the Luminex equipment and low requirements in protein content allows the design of the experiment on 96-well plates, leading to high sample-throughput[106,107]. On the computational front, an Integer Linear Programming (ILP) formulation is used to fit a prior knowledge network (PKN) to the proteomic data, resulting in a mechanistic model, predictive of the function and response of human chondrocytes[108]. The proposed ILP approach is based on Boolean logic to model signal transduction in the network. Boolean logic assumes protein activation can take only binary values (ON/OFF) and then uses Boolean gates (AND/OR/NOT) to model protein connectivity in the signaling network. By adopting Boolean logic and using a PKN e obtained from literature citations of signaling reactions e as a scaffold, we construct an initial model of the signal transduction network. Subsequently, the ILP formulation is used to train this model to the proteomic data, by removing reactions that contradict the data at hand. In this manner, the optimized model best captures the signaling patterns of human chondrocytes.

Methods

Chondrocyte isolation, culture and stimulation

Cartilage tissue was obtained from the femoral heads of patients undergoing total hip arthroplasty because of subcapital femoral fractures. Tissue was isolated using standard methods. Chondrocytes were isolated by sequential enzymatic digestion with pronase for 15 min at 37C and 0.4% collagenase Type II for 4 h at 37C as previously reported[109]. Enzymes were diluted in serum-free Dulbecco's modified

Eagle's medium supplemented with 50 mg/ml gentamycin, 100 mg/ml Kanamycin and 1.25 mg/ml Fungizone Amphotericin B. Isolated chondrocytes were washed, strained with 70 mm nylon meshes, and finally seeded at high-density (50,000 cells/well) in 96-well plates resuspended in DMEM high glucose, supplemented with 10% FBS, 100 U/ml Penicillin Streptomycin, 10 mM HEPES and 1% MEM Non-Essential Amino Acids Solution. Cells were cultured for 24 h (60% confluency) and were then starved overnight at 37C, 5% CO2, in serum-free DMEM high glucose; stimulated with single treatments of 78 ligands (see Supplementary Material 1 at DOI: 10.1016/j.joca.2014.01.001) and were finally lysed at 5 and 25 min after stimulation. The lysates were pooled in a 1:1 ratio before executing the Luminex assay. The library of 78 ligands was compiled to include some (but not all) of the major players of cartilage biology together with many underreported ligands, in an attempt to discover new players of cartilage pathophysiology.

Samples from various donors were used for developing the experimental protocols and for preliminary studies. However, the computational analysis that follows requires minimum biological variation for maximum signal to noise ratio; thus, for the final experiment only samples from a single donor were considered. We expect donor to donor variability, but that should be in the same levels as with previous studies[107].

Luminex assay

Phosphoproteins

A Luminex 200 system was used, to measure the activation of 17 phosphoproteins (AKT, JUN, CREB, ERK, GSK3, HISTH3, HSP27, IKB, IRS1S, JNK, MAP2K1, MAPK14, TP53, RPS6KB1, RPS6KA1, STAT3, STAT6) 5 and 25 min after stimulation, as described in[107]. The 17 phosphoproteins were chosen based on assay availability and quality controls performed at early stages of the experimental setup.

Cytokine releases

A library of 55 cytokine releases including most major inflammatory mediators such as cytokines, chemokines, growth factors, as well as degenerative enzymes (MMPs) (CCL27, CCL11, FGF2, CSF3, CSF2, CXCL1, HGF, ICAM, IFNA2, IFNG, IL10, IL12,

IL12P40, IL13, IL15, IL16, IL17, IL18, IL1B, IL1RA, IL2, IL2R, IL3, IL4, IL5, IL6, IL7, IL8, IL9, CXCL10, LIF, CCL2, CCL7, CSF1, MIF, CXCL9, CCL4L1, CCL4L2, NGF, PDGFB, CCL5, KITLG, SCGFB, CXCL 12, TNF, LTA, TNFSF 10, VCAM1, VEGF121, MMP1, MMP10, MMP13, MMP3, MMP8, MMP9) was measured in the supernatant 24 h after stimulation.

Data normalization

Both phosphoprotein and cytokine release data are measured in fluorescent units and is dependent on the antibody pair used for detection. For instance, MAP2K1 ranged from 280 units (untreated condition) to 6500 units (under EGF), while GSK3 ranged from 500 units (untreated condition) to 1500 units. Variations such as these do not necessarily reflect that MAP2K1 is more activated than GSK3 but may be attributed to protein abundance or assay calibration issues. Consequently, two challenges emerge, firstly, identifying whether a signal is activated or not, and secondly, normalizing the raw data in a way that the optimization algorithm is not biased in favor of the highest values[110]. Herein we implemented the normalization procedure introduced in[111]: We fitted the measurements of each signal to a bimodal distribution and for each datapoint, we formed the ratio of the frequencies with respect to the two modes and passed it to a hill function to map it to [0, 1]. The normalized data were used by the ILP algorithm.

Construction and pre-processing of the PKN

A PKN was constructed downstream the 78 receptors of interest,based on literature citations of signaling reactions[111]. Several online databases were queried (Reactome[112], PathwayCommons[113], KEGG[114]), but most of the reactions were obtained from Ingenuity (http://www.ingenuity.com/). The PKN was constructed in such a manner that it includes all interrogated receptors and measured phosphoproteins.

Upon its construction, the PKN was pre-processed to remove non-observable and non-controllable parts of it according to[110]. Non-observable, are nodes in the pathway

whose activation state cannot be inferred based on the measured phosphoproteins (i.e., Nodes, downstream of which there are no measured signals). Non- controllable are nodes whose activation state cannot be controlled by the imposed perturbations (i.e., Nodes with no upstream stimuli). Removing non-observable, non-controllable parts of the pathway facilitates the optimization process by reducing the size of the pathway.

Optimization of signaling pathways to proteomic data via an ILP formulation

An ILP formulation was used to fit the PKN to proteomic data by removing reactions that contradict the data at hand. See also[23].

Assume a signaling network G defined as a set of reactions $i = 1,\ldots, n_r$ and species $j = 1,\ldots,n_s$. Each reaction i is described by three index sets; the set of reactants R_i, the set of inhibitors H_i and the set of products P_i; R_i, H_i, $P_i \subset \{1,\ldots,n_s\}$. We also define a set of experiments $k = 1,\ldots,n_e$. In each experiment k a set of species I_j^k are perturbed with $I_j^k \in \{0,1\}; \quad j = 1, \ldots, n_s; \quad k = 1, \ldots, n_e$. $I_j^k = 1$ if species j in experiment k is set to active (ON); $I_j^k = 0$ if j is not perturbed. Moreover, we define variables $x_j^k \in \{0,1\}$ to denote the predicted activation state of species j in experiment k. $x_j^k = 1$ if species j is active in experiment k; $x_j^k = 0$ otherwise. If species j is also measured and it is found to be active in experiment k, then $x_j^{k,m} = 1$, else if it is found to be inactive $x_j^{k,m} = 0$. We also introduce variables $z_i^k \in \{0,1\}; \quad i = 1, \ldots, n_r; \quad k = 1, \ldots, n_e$ to denote whether reaction i is active $(z_i^k = 1)$ or not $(z_i^k = 0)$ in experiment k. Finally, we define variables $y_i \in \{0,1\}; \quad i = 1, \ldots, n_r$ to denote whether reaction i is present in the network. $y_i = 1$ if reaction i is present in the network; $y_i = 0$ otherwise.

Starting from the perturbed nodes, the signal is propagated downstream following the rules of Boolean logic. Thus,

1. A reaction *i* will take place in experiment *k* ($z_i^k = 1$) if and only if it is present in the network ($y_i = 1$), all reactants are present (i.e., $x_j^k = 1; \quad \forall j \in R_i$) and no inhibitors are present (i.e., $x_j^k = 0; \quad \forall j \in H_i$).
2. If a reaction *i* takes place all downstream species will be activated (i.e., $x_j^k = 1; \quad \forall j \in P_i;\ k = 1, \ldots, n_e$)
3. A species *j* will be active (i.e., $x_j^k = 1$) if and only if a reaction *i* takes place (i.e., $z_i^k = 1$) where this species is a product (i.e., $j \in P_i$); otherwise *j* will be inactive (i.e., $x_j^k = 0$).

The rules above may be formulated as linear constraints in the following manner (math equation 1):

$$
\begin{aligned}
& x_j^k \geq I_j^k; \quad j = 1, \ldots, n_s;\ k = 1, \ldots, n_e \\
& z_i^k \leq x_j^k; \quad \forall j \in R_i;\ i = 1, \ldots, n_r;\ k = 1, \ldots, n_e \\
& z_i^k \leq 1 - x_j^k; \quad \forall j \in H_i;\ i = 1, \ldots, n_r;\ k = 1, \ldots, n_e \\
& z_i^k \geq y_i + \sum_{j \in R_i} \left(x_j^k - 1 \right) - \sum_{j \in H_i} \left(x_j^k \right); \quad i = 1, \ldots, n_r;\ k = 1, \ldots, n_e \\
& x_j^k \geq z_i^k; \quad \forall j \in P_i;\ i = 1, \ldots, n_r;\ k = 1, \ldots, n_e \\
& x_j^k \leq \sum_{i = 1, \ldots, n_r; j \in P_i} z_i^k + I_j^k; \quad j = 1, \ldots, n_s;\ k = 1, \ldots, n_e
\end{aligned}
$$

The aim of the formulation is to identify the optimal values of y_i variables to minimize the mismatch between experimental data and model predictions. Additionally, because a number of solutions may exist with the same optimal objective value, the size of the solution is also minimized to harvest the one incorporating the fewest reactions see also[108]. Thus, the following objective function is used (Math equation 2).

$$\sum_{j,k} a_j^k \left| x_j^k - x_j^{k,m} \right| + \sum_i \beta_i y_i$$

Equations ((1), (2)) define an ILP, where (1) are the constraints of the formulation and (2) is the objective function to be minimized. The ILP is solved using Gurobi29 under GAMS30 (General Algebraic Modeling System, http://www.gams.com/). Finally, variables x and z may be relaxated to [0, 1] to speed up the optimization procedure without affecting the optimal solution, see[108].

Results

Phosphoproteomic level

Phosphoproteomic measurements

Chondrocytes were stimulated with single treatments of 78 ligands while measuring the activation level of 17 key phosphoproteins via xMAP technology. The phosphoproteomic dataset is plotted in **Figure P1-1** via DataRail toolbox[115]. **Figure P1-1** is a collection of subplots representing the time course of the 17 signals from the unstimulated state to the average early response under each of the imposed ligands. The filling color in each subplot corresponds to the normalized value of the signal. Activated signals are plotted in red.

As shown in **Figure P1-1**, a large number of stimuli raised a significant response in chondrocytes, activating at least one phosphoprotein signal. As positive control observations, well known players such as IL1B, TNF, EGF, TGFA, INS, IGF1 and IL6 responded as expected from previous studies. The pro-inflammatory mediators IL1B and TNF activated IKB, HSP27, MAPK14 (p38) and JUN, already known to promote inflammation in cartilage, together with pro-growth signals such as CREB, ERK, GSK3, IRS1 and MAP2K1 (MEK12), validating their role in chondrocyte hypertrophy[116]. On the other hand, pro-growth stimuli such as EGF, TGFA and INS activated only anabolic pathways, leaving inflammation related signals unaffected. Apart from the major players, a number of underreported stimuli are found to affect chondrocytes signaling including INHBA (Inhibin beta A), ADIPOQ (Adiponectin), DEFB1 (Defensin beta 1), BTC (Betacellulin), CXCL1, HBEGF, IL19, CXCL10, ODN2006 (Toll Like Receptor (TLR)9 ligand), NOG (Noggin) and Flagellin. A detailed description of the screened stimuli and their role in cartilage physiology is given in [Supplementary Material 1 at DOI: 10.1016/j.joca.2014.01.001]. See also the Discussion section.

Pathway construction based on the ILP formulation

The ILP formulation is used to train the PKN to the phosphoproteomic data of **Figure**

P1-1, resulting in an integrative model, predictive of the signal transduction mechanisms of human chondrocytes. In more detail, the formulation removed all reactions from the PKN that contradicted the data at hand in an attempt to minimize the difference between model predictions and experimental data. The optimized network is shown in **Figure P1-2**. With thick black edges, we plotted the reactions that are validated by the data and with gray edges the reactions that contradicted the data and were removed by the ILP algorithm. An in-depth analysis of the ILP performance and experimental validation of model predictions in the neighborhood of MAPK14, HSP27 and IKB signals is shown in [Supplementary Material 2 at DOI: 10.1016/j.joca.2014.01.001].

Figure P1-2 validates the key findings of **Figure P1-1** and previous reports in the literature: Major inflammatory mediators such as IL1B and TNF, signal through their receptors to IKB, MAPK14, HSP27 and JUN. IL1B also activates CREB and MAP2K1 (growth related signals) via TRAF6, in good accordance to the phosphoproteomic data of **Figure P1-1** and the literature[117]. Pro-growth stimuli such as TGFA, BTC, EGF, IGF1, INS and FGF2 signal through GRB2 to SOS, RAS and from there either to MAP2K1 (MEK12) via RAF1, or signal through PI3K to AKT and to CREB. IL6 activates mostly STAT3 via JAK1. On the other hand, the signaling pathways of CXCL1, HBEGF, DEFB1, Flagellin and INHBA were uncovered for the first time in chondrocytes. CXCL1, a small cytokine of the CXC family, binds to CXCR2 and activates RPS6KA1. HBEGF, a ligand of the EGFR, signals via the same path- ways as EGF, BTC and TGFA. DEFB1, a TLR ligand, signals through TLR4 to RAC1 and from there to the MAPKs and finally activates HSP27 demonstrating pro-inflammatory action. Flagellin, also a TLR ligand, signals through TLR5 to MYD88 and then merges with the IL1 pathway activating major inflammatory signals, CREB and MAP2K1. INHBA, a ligand of the TGFBR, signals via the MAPKs to activate JNK and P53.

Experimental validation of the HSP27, MAPK14, and IKB connectivity in the solution.

To validate the results of the ILP formulation, we chose a neighborhood of the optimized network and performed follow up experiments; this is the neighborhood around the HSP27, MAPK14 and IKB signals. As shown in **Figure P1-2**, HSP27, MAPK14 and IKB are activated by two pathways that overlap at NIK, one originating

from IL1B and Flagellin that signals via IRAK, TIFA, TRAF6 and NIK, and a second one originating from TNF that signals via TNFRSF1B, TRAF2 and NIK. Then, from NIK a pathway activates IKB via IKBKB and another activates MAPK14 and HSP27 via MAP2K6 and MAPKAPK5 (node "OR" is an auxiliary node). Note that HSP27 is downstream of MAPK14, while it is also activated by an independent pathway via MAPK11, MAP2K4, MAP3K5 and RAC1, but this pathway is only functional under the DEFB1 stimuli. For the rest of the stimuli (IL1B, Flagellin, TNF), HSP27 is activated via MAPK14.

As a follow-up experiment we perturbed the cells with major activators of HSP27, MAPK14 and IKB signals, in combination with small molecular inhibitors to block key signaling proteins, while monitoring HSP27 and IKB activation in an attempt to validate the connectivity of these three nodes in the solution. As activators, we used IL1B and TNF. (1) IL1B to stimulate the first pathway to NIK and (2) TNF to stimulate the second pathway to NIK as described above. As inhibitors, we used a potent MAPK14 inhibitor (PHA- 818637 at 100 nM) and a PI3K inhibitor (PI-103 at 10 mM). The activation of HSP27 and IKB was measured at three time points (5, 15 and 25 min), using the Luminex xMAP technology as described in the Methods section. Results are plotted in **Figure P1-4**.

Figure P1-4 (A) shows how the inhibition of MAPK14 affects HSP27. In good accordance to model predictions, inhibition of MAPK14 caused a significant decrease of HSP27 activation, since the ILP placed the HSP27 signal directly downstream of MAPK14. Note that in the initial network (gray edges in **Figure P1-2**) a number of alternative pathways are included for HSP27 activation upon stimulation with IL1B and TNF that do not go through MAPK14 (e.g., IL1B / IL1R2 / IRAK / TIFA / TRAF6 / MAP3K5 / MAP2K4 / MAPK11 / MAPKAPK5 / HSP27, or TNF / TNFRSF1B / TRAF2 / NIK / MAP2K3 / MAPK11 / MAPKAPK5 / HSP27), however, these were removed by the ILP algorithm as non-functional in the interrogated cell type, and HSP27 was placed directly downstream of MAPK14. On the other hand, the inhibition of PI3K had no effects on HSP27 activation, also in good accordance to model predictions, where the MAPK14, HSP27 and IKB pathway is completely disconnected from PI3K. Finally, **Figure P1-4** (B) shows how IKB activity is affected by MAPK14 and PI3K inhibition. In good accordance to model predictions, IKB is not

affected by either MAPK14 or PI3K inhibition, since IKB is regulated by a different pathway than MAPK14 and is also disconnected from PI3K. Thus, the follow up experiments of **Figure P1-4** validate the connectivity of HSP27, MAPK14 and IKB in the optimized network.

Cytokine release level

Chondrocytes were stimulated with single treatments of 78 ligands while measuring the release of 55 cytokines. The cytokine release data is shown in **Figure P1-3**, which is a collection of subplots representing the time course of the 55 signals from the unstimulated state to 24 h, for each of the imposed ligands. The filling color in each subplot corresponds to the normalized value of the signal.

As shown in **Figure P1-3**, a large number of the imposed stimuli promoted significant cytokine release in chondrocytes. The strongest inducers were FSTL1 (a TLR 6/2 agonist), GDF5, HKSA (TLR2 ligand), IL1A, IL1B, IL6, LPS, PAM3CSK4 (TLR 1/2 agonist), POLYIC (TLR3 agonist), Flagellin (TLR5 agonist), TNF and LTA (a member of the TNF superfamily). Most of them have already been proven to exhibit pro-inflammatory action; however, there are a number of underreported players such as GDF5 and LTA. As positive controls, we observe that major inflammatory mediators such as IL1A and IL1B raised an extensive inflammatory response inducing the release of most of the measured cytokines. Moreover, IL1A, IL1B together with many of the inflammatory mediators mentioned above were found to induce the release of MMPs (mainly MMP13 and MMP1) known to degrade the collagen content of the tissue. Finally, we observe that pro-growth stimuli did not induce any significant cytokine releases.

Discussion

In this paper, we have presented a rigorous approach for the study of signal transduction in chondrocytes on a systems level. We interrogated their signaling mechanisms, downstream 78 receptors of interest, on both the phosphoproteomic and the cytokine release levels and also employed an ILP formulation to construct a

predictive model of their signaling processes. On the experimental front, we adopted the xMAP technology to measure 17 key phosphoproteins and the release of 55 cytokines in the supernatant, upon stimulation with single treatments of the 78 ligands. These ligands were selected to stimulate major osteoarthritic degradation pathways. On the computational front, we implemented an optimization formulation, based on the modeling of signal transduction via Boolean logic, to train a PKN to the proteomic data by removing reactions that appear not to be functional in primary chondrocytes.

The analysis presented herein, was able to validate previous findings regarding major players of cartilage homeostasis and inflammation (e.g., IL1B, TNF, EGF, TGFA, INS, IGF1 and IL6); highlight the role of underreported players such as INHBA, DEFB1, CXCL1 and Flagellin in chondrocyte signaling; identify their signaling pathways and uncover the way they cross-react in the phosphoproteomic level. Even though most of the effects of the interrogated stimuli on the measured phosphoproteins or cytokine releases have been reported before, this is the very first attempt to our knowledge to leverage high throughput proteomic data via a bioinformatics approach and identify which of the signaling pathways reported in literature are functional in primary human chondrocytes and orchestrate their response to external perturbations. On this front regarding the underreported players DEFB1 and Flagellin: We identified that DEFB1 signals via its receptor to RAC1, to the MAPKs and ultimately activates HSP27. Flagellin signals through TLR5 to MYD88 and then merges with the IL1 pathway signaling through IRAK, TIFA, TRAF6 and activates the IKB, MAPK14 and HSP27 signals. Moreover, we were able to validate part of the DEFB1 and Flagellin pathways in the neighborhood around the MAPK14, HSP27 and IKB signals via an independent follow up experiment (see Supplementary Material 2 at DOI: 10.1016/j.joca.2014.01.001), proving the predictive power of the proposed ILP algorithm and reliability of model predictions. On the other hand, the lack of data in the remaining part of these pathways implies the ambiguity of the solution from the TLRs to the neighborhood of MAPK14, HSP27 and IKB, requiring tedious follow up experiments to validate that exceeds the scope of this work. Below we discuss our findings in the context of degenerative diseases such as osteoarthritis (OA).

Inflammatory mechanisms

OA has a strong inflammatory component[90]. Our analysis of cytokine release data (see **Figure P1-3**) extensively validates the role of chondrocytes in inflammation: In **Figure P1-3**, chondrocytes are found to release cytokines and MMPs upon stimulation with major inflammatory mediators (IL1B and TNF), in good accordance to literature reports; but also, upon stimulation with underreported ligands. Moreover, taking into account that the tissues used for our analysis came from healthy donors, our findings suggest even healthy cells have the potential to demonstrate a strong inflammatory response (rather than protective). In this manner, the generation of an inflammatory environment in the joint is facilitated and given that cartilage is an avascular tissue, lacking important anti-inflammatory components of peripheral blood[84], this environment is sustained, eventually leading to the degradation of the tissue.

The inflammatory mechanisms of chondrocytes were also studied on the phosphoproteomic level. Using an ILP formulation, for the first time we were able to identify the signaling pathways of underreported pro-inflammatory mediators such as DEFB1 and Flagellin (TLR ligands). Even though the role of TLR in cartilage inflammation has been studied before, this is the first attempt to identify the signaling events that lead to the up-regulation of NFKB, HSP27 and other pro-inflammatory signals upon stimulation of the TLR. In more detail, two major inflammatory pathways were uncovered: the first activating IKB via NIK, and the second activating HSP27 via MAPK11 and MAPK14 dependent mechanisms. Moreover, cross-talks of the TLR with other major inflammatory and pro-growth pathways were uncovered.

Chondrocyte hypertrophy and proliferation

During OA, proliferation and hypertrophic differentiation of chondrocytes occur. These processes resemble to skeletal development by endochondral ossification mechanisms[118]. They refer to the gradual differentiation of chondrocytes and subsequent release of BMPs that leads to matrix remodeling. Our analysis of the phosphoproteomic data sheds light into these mechanisms. As seen in **Figure P1-1**,

stimulation with IL1B and Flagellin (two major inflammatory mediators) apart from up-regulation of inflammatory signals, also leads to over-activation of CREB and MAP2K1, two growth related signals that are connected to hypertrophy[85,86] (IL1B also activates ERK, GSK3 and IRS1 but this is not shown in the network).

The mechanism of CREB and MAP2K1 activation was uncovered by the ILP algorithm and is shown in the network of **Figure P1-2**. IL1B and Flagellin signal through overlapping pathways to IRAK, TIFA and then to TRAF6. From TRAF6 either activate pro-inflammatory signals or go through RAC1 to activate CREB and MAP2K1.

Moreover, inspection of the cytokine release data of **Figure P1-3** reveals another potential mechanism IL1B and Flagellin induce chondrocyte hypertrophy: Both IL1B and Flagellin induce the release of FGF2 known to play a role in endochondral ossification[118] and CXCL1, a chemokine that induces chondrocyte hypertrophy[119,120]. IL1B additionally induces the release of CSF3 (another pro-growth factor). Thus, IL1B and Flagellin apart from directly activating pro-growth signals via the pathways discussed above, also lead to the release of pro-growth cytokines, facilitating chondrocyte hypertrophy and bone ossification indirectly.

Innate immune response in OA TLR signaling

The role of TLR signaling in OA is becoming of increasing significance as the community attempts to identify the etiology underlying the pathogenesis of the disease. TLRs belong to the family of pattern-recognition receptors and play a pivotal role in the activation of the innate immune system in response to invading microbial components[121]. TLR stimulation elicits strong release of pro-inflammatory cytokines[122]; while only recently human articular chondrocytes were shown to express TLRs[123,124]. In this paper, we addressed TLR signaling extensively by screening 10 TLR ligands (DEFB1, HKSA, Imiquimod, Flagellin, LPS, ODN2006, PAM3CSK4, POLYIC, SSRNA40 and FSTL1). On the cytokine release level 6 of them (FSTL1, HKSA, LPS, PAM3CSK4, POLYIC and Flagellin) were found to raise a strong inflammatory response, validating the significance of TLR stimulation in joint inflammation. However, on the phosphoproteomic level only DEFB1 and Flagellin signaled through pathways monitored by the 17 phosphoprotein signals. DEFB1

signaled via TLR4 to RAC1 and from there to the MAPKs, eventually activating HSP27. Flagellin signaled via a pathway overlapping with IL1B and as discussed above, activated most inflammatory signals together with MAP2K1 and CREB.

Overall, the proposed approach successfully addressed the construction of a predictive model of the signaling mechanisms in human chondrocytes. We interrogated signal transduction on both the phosphoproteomic and cytokine release levels, upon stimulation with 78 factors of interest, including most major players of osteoarthritic pathophysiology, as well as ligands important for other cell types. Our analysis validated previous findings in chondrocytes signaling, elucidated the role of underreported players and identified cross-talks on the phosphoproteomic level, highlighting the pleiotropic role of major players in cartilage homeostasis and inflammation.

Meniscus – Cartilage paracrine crosstalk in osteoarthritis

Please visit abstract [A1] in section “Abstracts in International Conferences” for more information on this chapter.

Introduction

Meniscus plays an essential role in knee joint function providing stability and load transmission. In osteoarthritis (OA), a joint disease characterized by chronic synovitis and cartilage degeneration, pathological changes in the menisci are observed. However, whether menisci contribute to the progression of OA, the underlying mechanism for meniscus-cartilage communication is still unclear. In this study, we analyzed systematically the response of meniscus and cartilage explants to a number of inflammatory mediators, in order to reveal their response similarity and highlight potential crosstalks and interactions.

Methods

OA cartilage and the lateral meniscus were harvested from two patients undergoing total knee arthroplasty. Meniscus and cartilage disks (3 mm diameter) were stimulated with inflammatory mediators ((IL-1α, IL-1β, IL-12α, CSF2) (50 ng/ml), (TNF-α, IL-6, CXCL7, IL-8, CCL2, CXCL10, IFN-γ, IL-3, MIA2, IL-4) (100 ng/ml) and GROα (500 ng/ml)) for 24 h. For each condition the release of different proteins (IL-1α, IL-1β, TNF-α, IL-6, CXCL7, GROα, IL-8, CCL2, CXCL10, IFN-γ, IL-3, IL-12α, MIA-2, CSF2, IL-4) was measured in the supernatant using custom multiplexed assays on a Luminex FlexMap 3D instrument.

Results

In both tissues the major inflammatory players (IL-1α, IL-1β, TNFα) were the strongest stimuli as expected. Meniscus responses were the same up to 73 and 50% with the cartilage ones for the first and the second donors respectively. Interestingly, meniscus under certain stimuli (IL-1α, IFN-γ, CSF2, IL-8) responded differently than cartilage by releasing five different cytokines while cartilage did not.

Conclusions

Our results indicate that meniscus is affected by its inflammatory environment and responds to it as actively as cartilage. Moreover, the release of different cytokines from meniscus and cartilage suggests that meniscus can be an active player in the progression of OA. These data support the hypothesis that significant crosstalk between these two knee compartments exist and anti-inflammatory therapies should take into consideration both tissues.

An integrated proteomic and metabolomic approach to investigate cartilage degeneration

Please visit abstract [A4] in section “Abstracts in International Conferences” for more information on this chapter.

Cartilage degeneration as a feature of osteoarthritis (OA) is one of the most common causes of pain and disability in middle-aged and older people. The percentage of population above 45 years of age affected with OA is estimated to increase from 26.6% (2012) to 29.5% by 2032 due to the aging of the population and the obesity epidemic. Most research studies in OA focus on single therapeutic targets, disease processes or level of molecules, thus omics data have never been integrated in a systematic way.

Articular cartilage and synovium explants, whole cell extract of articular chondrocytes or mesenchymal stem cells, supernatant of articular cartilage explants, articular chondrocytes, osteoclasts or synovium-derived cells in culture, articular cartilage vesicles, synovial fluid (SF), plasma, serum and urine have all been studied on the proteomic front and a subset of those for their metabolomic signature. However, no more than three of those types of samples have been examined in a single study using the same assay type. By using samples from various tissues, we avoid focusing only on the biochemical changes that occur in the joints.

As a case study, we combined proteomic and metabolomic assays (multiplex bead-based sandwich ELISA xMAP technology and gas chromatography mass spectrometry), to measure the phosphoproteomic signature, cytokine release and metabolomic signature of untreated and treated with IL1B cultured chondrocytes, along with the proteomic and metabolomic signature of the same donor's synovial fluid and plasma.

Following this multi-omics approach, we had to include additional quality control check points to ensure the highest possible dataset quality and therefore the validity of results. In particular, we need to be able to address challenges such as acquiring multiple samples from the same donor at the same time, using the exact same samples for all experiments (cell line passage, timelines in cell culture), merging protocols of all omics assays while meeting the sample handling prerequisites for

every assay and having strictly fixed parameters across assays, such as time points and doses of any applied stimuli.

Besides ensuring robustness of the dataset, in the case of OA, we have to take into account the scarce availability of healthy human cartilage, synovium and SF samples and examine the clinical profile data of donors in order to expand the dataset while gaining knowledge about factors such as medication that can alter the omics data.
In conclusion, we believe that omics data acquired by multiple samples while avoiding the above-mentioned pitfalls, can be integrated in a multi-level dataset which may help us understand complex biological systems and multifactorial diseases such as OA.

Metabolomics analysis of Synovial Fluid samples

Synovial Fluid samples (The Donors used were donor #68 and donor #94 - please see all sample information in **Table A4-1**) were run in a Gas Chromatography Mass Spectrometry system and analyzed to specify which metabolites were present in the samples.

The Metabolomic analysis process involves creating a table matching your chromatograph with the lab's history of identified Metabolite Derivates. Marked as dm_X are unknown metabolites that have possibly not identified before from this lab. The full table is presented in **Table A4-2.**

We identified several metabolites also mentioned in the chapter of this thesis: "Osteoarthritis and Metabolomics" > "OA Synovial Tissue Metabolic Profile". All identified metabolites are shown on **Table A4-3**. With green and red color are shown similarities with literature, for upregulation and downregulation of those metabolites respectively.

Sample collection from Chondrocyte cell culture

We also isolated and prepared chondrocytes for metabolomics measurement, but we did not complete measurement and analysis. The procedure is described below.

The following protocol was based on the “Adherent Hela cells - Collection” protocol of Metabolic Engineering and Systems Biology Lab (MESBL), FORTH/ICE-HT.

1) Seed chondrocytes on a 10cm Petri dish (~78.5 cm2 cell growth area) and continue culture until reaching 80% confluency.
2) At the end of the stimulation period (if any), place the dish on ice and remove 1ml of the supernatant. Add this 1ml to 1ml ice-cold HPLC-grade MtOH. This is the cell culture supernatant sample.
3) Wash 3 times with 10ml ice-cold PBS and remove 1ml of the 3rd wash. Add this 1ml to 1ml ice-cold HPLC-grade MtOH. This is the 3rd PBS wash sample, acquired for quality control purposes.
4) Completely aspirate the PBS and add 2ml ice-cold HPLC-grade MtOH to the cells.
5) Move the liquid across the plate for ~1’ and then collect the 2ml.
6) Place all samples at -80C.

Notes:
1) The flasks’ surface should not be tissue-culture treated and ideally should not have any coating substances. For this preliminary experiment we used Corning #430641 flasks (75 cm2 cell growth area, tissue culture treated according to this process: http://www.corning.com/lifesciences/us_canada/en/technical_resources/surfaces/culture/stc_treated_polystyrene.aspx).
2) The chondrocytes should be always from the exact same passage for all Donors. For this preliminary experiment we used freshly thawed P3 cells (P4) from Donor #94.
3) The cells must be scrapped. For this preliminary experiment we didn’t scrape the cells.
4) Normally, for each experiment we culture one more Petri dish for RIPA - Bradford assay to evaluate protein content. For this preliminary experiment we did not do that.

For this preliminary experiment, we seeded freshly thawed P3 chondrocytes (P4) isolated from Donor #94 into 2 flasks (Corning #430641) with 20ml culturing medium (DMEM, 10% FBS, 1% PS) per flask, changing the medium every 2-3 days until reaching confluency 80%. Then, we changed the cell supernatant of one flask with medium containing 50ng/ml IL1B and the other flask with the normal culturing medium, incubated the cultures for 24h and then proceeded with the above protocol by placing the flasks on ice and following the rest of the steps.

Metabolite extraction for chondrocyte whole cell extracts

1) Add 0.1ug ribitol per mg of protein content in the sample.
2) Add 0.2ug glucose per mg of protein content in the sample.
3) Gently mix.
4) Place in waterbath at 70-75C for 20 minutes. Precipitation of big molecules is observed.
5) Add HPLC-grade H2O, same volume as the MtOH.
6) Gently mix.
7) Centrifuge at 10000g for 10 minutes at 4C.
8) Carefully collect the supernatant without taking any of the precipitated proteins.
9) Transfer the supernatant in a glass, pre-weighted tube.
10) Weigh the tube and calculate the mass of dry metabolites
11) Seal the tube with parafilm and keep at 4C until measurement

Derivatization of metabolites and transfer for measurement

There are two tips for each chemical in the hood: tips for transfer and tips for samples. The bold blue line of the tip goes to the syringe.
Keep samples with aluminum foil around them and keep them open as less time as possible.

Preparation of MeOx
MeOx 90’ 40C

Alfa Aesar A19188, CAS 593-56-6
weigh powder, dilute at 20ug/ml with Pyridine
easily evaporates

Preparation of MSTFA
MSTFA 9h+ 30C
Alfa Aesar A13141, CAS 24589-78-4
Transfer with tip for transfer from bottle into your vial.

Derivatization
1) Run sample in SpeedVac for 30 minutes.
2) Add MeOx using MeOx tip for samples and leave in incubator at 40C for 90 minutes.
3) Add MSTFA using MSTFA tip for samples and leave in incubator at 40C for 9 hours.
4) Gently mix tapping with finger.

Transfer
1) Wash syringe with pyridine
2) Transfer the sample without disturbing the precipitant into the dark vial
3) Cap the vial firmly using crimper

Vials (2ml)
Agilent Technologies
Part No.: 5181-3376 (vials)
Part No.: 5182-0871 (caps)
Part No.: 5183-2085 (inserts)
11mm Crimper
11mm Decapper

Samples transferred to MESBL

The samples transferred to MESBL for potential metabolomic analysis were:

1) supernatant of donor #94 chondrocytes treated with 50ng/ml IL1B for 24h [total vol. 2ml]
2) 3rd wash PBS of case 1 [total vol. 2ml]
3) MtOH lysate of case 1 [total vol. 2ml]
4) supernatant of donor #94 chondrocytes untreated [total vol. 2ml]
5) 3rd wash PBS of case 4 [total vol. 2ml]
6) MtOH lysate of case 4 [total vol. 2ml]
7) culturing medium (DMEM, 10% FBS, 1% PS) [total vol. 2ml]
8) synovial fluid from Donor #68 [total vol. 1.5ml]
9) synovial fluid from Donor #94 [total vol. 1.5ml]
10) blood plasma from Donor #68 [total vol. 1.5ml]
11) blood plasma from Donor #94 [total vol. 1.5ml]

- Samples 1-6 were prepared according to the protocol described above “Sample collection from Chondrocyte cell culture“.
- Sample 7 was prepared by adding 1ml of the culturing medium to 1ml ice-cold HPLC-grade MtOH.
- Samples 8-9 did not have MtOH added. Those samples were prepared according to the following steps:

Synovial fluid samples preparation

1) The samples are travelling from Larissa to Athens in RT conditions during winter or icepack during summer (which can also be considered RT). We get the samples in syringe or urobox, usually 24h after acquisition from the donor
2) Transfer 1.5ml in an eppendorf microcentrifuge tube
3) Spin down at 14000G, 15’, 4C
4) Observe a small pellet which is the synoviocytes. Transfer the supernatant into an eppendorf microcentrifuge tube
5) Store at -80C

- Samples 10-11 did not have MtOH added. Those samples were prepared according to the following steps:

Plasma samples preparation

1) The samples are travelling from Larissa to Athens in RT conditions during winter or icepack during summer (which can also be considered RT). We get the samples as 2ml blood in the pink cap vacutainer (EDTA), usually 24h after acquisition from the donor
2) Dilute the blood 1:1 with PBS (W/O Ca+) by adding 2ml PBS into a 15ml Falcon tube
3) Carefully add the PBS-blood mix on top of 3ml Ficoll present in another 15ml Falcon tube
4) Spin down at 400G, 15', 4C
5) Plasma is in the supernatant, which is the first of 3 phases. Transfer 1.5ml in an eppendorf microcentrifuge tube
6) Store at -80C

Equipment and reagents

VARIAN CP-3800 GAS CRHROMATOGRAPH
VARIAN Saturn 2200 GC / MS / MS
HPLC grade H2O
Water (UV-HPLC) PAI-ACS
Panreac 361074.1612
MtOH
Methanol (HPLC-gradient grade) PAI-ACS
Panreac 221091.1612
Pyridine
Vendor: SDS CAS: 110.86.1 Catalog: 0671016

The role of inflammatory mediators in meniscus and cartilage crosstalk in osteoarthritis

Please visit abstract [A6] in section “Abstracts in International Conferences” for more information on this chapter.

Purpose

Osteoarthritis (OA) is a common degenerative joint disorder causing cartilage degradation, pain and disability. Lately it is characterized as a whole joint disease as it affects the structure and functionality of all tissue components, such as the menisci. Meniscus plays an essential role in knee joint function providing stability and load transmission. During OA the balance between catabolic and anabolic processes in the cartilage tissue is disturbed favouring catabolism through the cytokines and Matrix Metalloproteinases (MMPs) that are present in the synovial fluid of patients with mild or severe OA. In this study, we examine how meniscus and cartilage explants react to a number of inflammatory mediators, in order to reveal their response similarity and highlight potential crosstalks and interactions.

Methods

OA cartilage and the menisci were harvested from patients undergoing total knee arthroplasty. Meniscus and cartilage disks (3 mm diameter) were stimulated with inflammatory mediators [IL-1a, IL-1b, IL12a, CSF2 (50 ng/ml), TNF-a, IL-6, CXCL7, IL-8, CCL2, CXCL10, IFN-g, IL-3, MIA2, IL-4 (100 ng/ml) and GROa (500 ng/ml)] for 24 h. For each condition the release of different proteins [IL-1a, IL-1b, TNF-a, IL-6, CXCL7, GROa, IL8, CCL2, CXCL10, IFN-g, IL-3, IL-12a, MIA-2, CSF2, IL-4] was measured in the supernatant after 24 hours using custom multiplexed assays on a Luminex FlexMap 3D instrument. Histological sections were prepared in order to assess the osteoarthritic grade of the tissues used. Datarail toolbox was used for data handling and visualization.

Results

In general, there was strong similarity between meniscus and cartilage releases; the major inflammatory players (IL-1a, IL-1b, TNFa) were the strongest stimuli as expected, in both tissues. Under these cytokines cartilage disks released IL-6, CXCL7, GROa, IL-8, and CCL2. Meniscus responses were the same up to 73% and 50% with the cartilage ones for the first and the second donors, respectively. Meniscus under certain stimuli (IL-1a, IFN-g, CSF2, IL-8) responded differently than cartilage by releasing five different cytokines that cartilage did not.

Conclusions

Until now, little is known about meniscus reaction to cartilage protein expression or to the high cytokines concentration of the synovial fluid after trauma or OA. Our results confirm that IL-1a, IL-1b and TNF-a, as proinflammatory cytokines, promote systemic inflammation and thus release of IL-6, IL-8 and GROa in both cartilage and meniscus tissue. The above findings suggest that meniscus is affected by its inflammatory environment and responds to it as actively as cartilage, supporting the hypothesis that significant crosstalk between these two knee compartments exists and anti-inflammatory therapies should take into consideration both tissues. Further investigations should be done to unveil the exact communication mechanisms between meniscus and the other joint tissues.

Human - Rat Translational Research

The disease

Animal models are important tools in drug discovery and for understanding human biology in general. However, many drugs that initially show promising results in rodents, fail in later stages of clinical trials. Understanding the commonalities and differences between human and rat cell signaling networks can lead to better experimental designs, improved allocation of resources and ultimately better drugs.

Experimental Model

- Culture of Normal Rat Bronchial Epithelial (NRBE) Cells
- Culture of Normal Human Bronchial Epithelial (NHBE) Cells

Experimental Tools

- Cell culture of NRBE and NHBE
- Stimulation with 52 compounds, lysis and collection of cell supernatant
- Total RNA was isolated from NHBE and NRBE cells using the QIAGEN RNeasy 96 Kit
- Measurement with bead-based sandwich ELISA of 17 phosphoproteins and 55 cytokines
- Measurement of mRNA samples with Affymetrix® HG-U133 Plus2 or Rat 230 2.0 GeneChips

Computational Tools

- We crowdsourced the problem to combine the data at hand with prior knowledge of proteins' connectivity to construct specific signaling pathways

for human and rat organisms.

Main Findings

A multi-layer systems biology dataset was generated that was comprised of phosphoproteomics, transcriptomics and cytokine data derived from normal human (NHBE) and rat (NRBE) bronchial epithelial cells exposed in parallel to more than 50 different stimuli under identical conditions. This unique multi-omics dataset is of great value for the computational community to develop new modelling capabilities to address the important topic of species translatability at different molecular levels of the human and rat bronchial epithelial cellular system. A better understanding of the range of applicability of the translation concept will impact the predictability of signaling responses, mode of action and efficacy of drugs in the field of systems pharmacology as well as increase the confidence in the estimation of human risk from rodent data in the context of toxicological risk assessment.

Animal models are important tools in drug discovery and for understanding human biology in general. However, many drugs that initially show promising results in rodents, fail in later stages of clinical trials. Understanding the commonalities and differences between human and rat cell signaling networks can lead to better experimental designs, improved allocation of resources and ultimately better drugs. Our analysis showed that major signaling pathways were conserved between the two species with only isolated components diverging, as in the case of RSK1. Overall, the consensus between inferred edges was relatively high with the exception of the downstream targets of transcription factors, which seemed more difficult to predict. Transcription factor *CREB1*, showed the best consensus for the edges upstream of it but with a couple of differences between human and rat: the connection from *RSK1* was present only in the human consensus network[125], whereas the connection from *PRKACA* was present only in the rat consensus network[126]. The prevalence of *RSK1* interactions in human might be explained by the fact that human isoforms of RSK1 have functional redundancy (i.e. *RSK2*, *RSK3*, *RSK4*). In contrast, this is most likely not the case in rodents; Zeniou et al.[127] reported that the mouse *RSK1* and *RSK3*

genes may not be able to fully compensate for the lack of *RSK2* function.

Publication Timeline

[P2] 10 June 2014, Nature Scientific Data
The species translation challenge—A systems biology perspective on human and rat bronchial epithelial cells

[P3] 7 October 2014, Bioinformatics
A crow-sourcing approach for the construction of species-specific cell signaling networks

Co-authored Research: The species translation challenge - A systems biology perspective on human and rat bronchial epithelial cells

Please visit DOI of paper [P2] for more information on this chapter.

Abstract

The biological responses to external cues such as drugs, chemicals, viruses and hormones, is an essential question in biomedicine and in the field of toxicology and cannot be easily studied in humans. Thus, biomedical research has continuously relied on animal models for studying the impact of these compounds and attempted to 'translate' the results to humans. In this context, the SBV IMPROVER (Systems Biology Verification for Industrial Methodology for PROcess VErification in Research) collaborative initiative, which uses crowd-sourcing techniques to address fundamental questions in systems biology, invited scientists to deploy their own computational methodologies to make predictions on species translatability. A multi-layer systems biology dataset was generated that was comprised of phosphoproteomics, transcriptomics and cytokine data derived from normal human

(NHBE) and rat (NRBE) bronchial epithelial cells exposed in parallel to more than 50 different stimuli under identical conditions. The present manuscript describes in detail the experimental settings, generation, processing and quality control analysis of the multi-layer omics dataset accessible in public repositories for further intra- and inter-species translation studies.

Background & Summary

Animal models have been used intensively to understand biological mechanisms associated with diseases and to unravel toxic effects of drugs or environmental agents. Biological processes in mice or rats have been generally assumed to reflect biological processes in humans under analogous conditions. A natural question in this context is the degree to which biological perturbations observed in rodents can be translated to humans. Such knowledge is important since it can reduce uncertainties in species extrapolations (**Figure P2-1**).

The Systems Biology Verification for Industrial Methodology for Process Verification in Research (SBV IMPROVER) initiative[128,129] (https://www.sbvimprover.com/) opened a challenge called Species Translation Challenge (STC) to the scientific community to identify compound-specific biological mechanisms of actions (MoA) that are common to different species, in this case, humans and rats. The challenge consisted of four sub-challenges whereby the interspecies pathway perturbation prediction challenge sought to explore whether responsive gene sets and related processes in humans can be inferred based upon the corresponding data in rats.
To address the question of species translatability at different molecular layers of the biological system in the context of STC, an experiment was designed to generate human and rat multi-layer datasets consisting of phosphoproteomics, transcriptomics and cytokine level measurements. To ensure that the generated datasets were comparable and that the proof of concept predictions across species was valid, experiments with well-controlled conditions were designed and conducted using an *in vitro* system. This chemical testing strategy is aligned with the effort to 'Replace, Reduce, Refine' animal experiments (the '3R' approach)

(http://ihcp.jrc.ec.europa.eu/our_activities/alt-animal-testing-safety-assessment-chemicals/alternative-testing-strategies-progress-report-2009.-replacing-reducing-and-refining-use-of-animals-in-research) and to use more appropriate cell-based assays that have the potential to provide more relevant data on the effects of short- and long-term exposure to toxicants. Primary normal human bronchial epithelial cells (NHBE) and primary normal rat bronchial epithelial cells (NRBE) were exposed in parallel to various types of stimuli, which were selected ensuring a broad perturbation spectrum of the cellular system, under identical experimental conditions (duration of exposure, concentration of stimuli and cell culture parameters).

The challenge aimed to investigate whether the phosphorylation signals could be inferred from gene expression data within species (reverse engineering) and the translatability of phosphorylation signals across species, and also to better understand the level at which translation across species is more robust (e.g., individual molecules, predefined gene sets representative of canonical pathways or higher-order processes). These questions have been articulated around four sub-challenges proposed to the scientific community (https://www.sbvimprover.com/challenge-2/challenge-2-challenge). The second SBV IMPROVER symposium was held in Greece at the end of October 2013 to announce the results of the Species Translation Challenge and to discuss the topic extensively with all participants (http://www.bio-itworld.com/2013/11/8/sometimes-you-can-trust-rat.html; http://www.genomeweb.com/informatics/improver-species-translation-challenge-results-released; http://www.americanlaboratory.com/913-Technical-Articles/149618-Results-are-in-for-the-Second-sbv-IMPROVER-Challenge-on-Species-Translation/).

The present manuscript describes the experimental design, optimization steps and data quality checks necessary to generate a multi-layer systems biology data compendium suitable for computational crowd-sourcing challenges such as the Species Translation Challenge. The experimental settings and protocols as well as the generation, processing and quality control analysis of the raw data are detailed. The raw data (168 and 164 CEL files for human and rat respectively) and processed data (e.g., normalized gene expression data) are freely available in public repositories such as ArrayExpress for transcriptomics data (Data Citation 1: ArrayExpress E-

MTAB-2091). Human and rat proteomics data are deposited in the figshare public repository (Supplementary Table 1 at DOI: 10.1038/sdata.2014.9) (Data Citation 2: Figshare http://dx.doi.org/10.6084/m9.figshare.960097).

The unique multi-omics dataset presented in this manuscript is of great value for the computational community to develop new modelling capabilities to address the important topic of species translatability at different molecular levels of the human and rat bronchial epithelial cellular system. A better understanding of the range of applicability of the translation concept will impact the predictability of signaling responses, mode of action and efficacy of drugs in the field of systems pharmacology as well as increase the confidence in the estimation of human risk from rodent data in the context of toxicological risk assessment. It provides a unique translational compendium with applicability in systems biology and toxicology, fully aligned with the Tox21 initiatives[130].

Methods

Cell culture

NHBE cells were purchased from Lonza (Catalog number CC-2540, Lonza Inc., Switzerland). These cells, obtained from different Caucasian, disease-free and non-smoker donors, were isolated from airway (tracheal/bronchial) epithelial tissue located above the bifurcation of the lungs. NRBE cells were purchased from CHI Scientific Inc. (Catalog number 4-61391, Maynard, Maryland, USA) and isolated from pooled tracheobronchial tissue of adult inbred AGA rats. The stocks of NHBE and NRBE cells were stored in liquid nitrogen with 10% (v/v) dimethylsulfoxyde (DMSO). Vials of stock cells were rapidly thawed and diluted in 20ml of bronchial epithelial cell growth medium with supplements (Lonza, BulletKit CC-3170). Both cell types were seeded in flasks (T75) coated with rat tail collagen type I from BD (catalog number: 354236) and grown in the same growth medium with supplements at 37.0±1°C in a humidified incubator with 5.0±0.5% CO2 in air. After 24h, the medium was changed and cells were regularly checked during proliferation using a microscope. Once reaching confluence, cells were split into subcultures. Briefly, cells were washed with HEPES Buffered Saline Solution, then trypsinized with Trypsin/EDTA that was

neutralized using a Trypsin Neutralizing Solution (TNS) (The 3 solutions are included in Clonetics™ ReagentPack™ from Lonza; catalog number CC-5034). Cells were expanded for 10 days (including 1 split) to reach the final number needed for screening or main experiments. Cells were seeded into pre-coated rat tail collagen type I 96-well plates (BD BioCoat™, catalog number: 356649) testing different cell densities ranging from 2,500 to 50,000 cells per well (in 100µl). The range of optimal seeding densities was determined by microscopic inspection to be 25,000–35,000 cells/well (80–90% confluence). However, optimal yield of RNA extraction used for transcriptomics analysis was obtained with 50,000 cells/well corresponding to 100% confluency. For the screening of the main experiment, cells were re-suspended in bronchial epithelial cell growth medium with supplements and seeded at a cell density of 50,000 cells/well in pre-coated rat tail collagen type I 96-well plates (BD BioCoat™, catalog number: 356649). After 24h, NHBE and NRBE were treated in parallel with selected stimuli or DME.

Systems biology data generation

Due to the high number of stimuli and experimental conditions described above, the main phase was conducted in two experiments (40 stimuli used for the experiment 1 and 12 stimuli for the experiment 2) to generate all samples required to produce the entire systems biology dataset.

Measurements of phosphoproteomics and cytokines using xMAP beads

For phosphoproteomics measurements, NRBE and NHBE cell cultures were removed from the incubator and placed on ice. The cells were washed with 100µl of ice-cold phosphate-buffered saline (PBS) and cells were lysed using 60µl of Tris-HCL supplemented with inhibitors of proteases and phosphatases in a 96-well plate on ice for 20min. The plates were incubated overnight at –20°C and then rapidly thawed in a 37°C water bath for 2min, followed by sonication. Cell debris was removed following a centrifugation at 2700×*g* for 20min at 4°C. For cytokines measurements, cell supernatants were collected 24h post-treatment. For the bead-based enzyme-linked immunosorbent assay (ELISA) procedure, 50µl of non-diluted cell lysates or supernatants were incubated with the xMAP beads (4000 beads/well for each protein)

for 1.5h to capture target proteins with specific antibodies coupled to the beads. The beads were washed twice with 100µl of PBS. The beads were then incubated for 1.5h with 20µl of detection antibodies targeting different epitopes than the bead-coupled capture antibodies (average concentration, 1µg/ml) followed by washing steps. Subsequently, 50µl of streptavidin-phycoerythrin (PE) (at a final concentration of 5µg/µl) were added and the mixture was incubated for 20min. The beads were then washed and re-suspended in 130µl of PBS-bovine serum albumin (BSA) assay buffer.

Transcriptomics

Total RNA was isolated from NHBE and NRBE cells using the QIAGEN RNeasy 96 Kit (Catalog number 74181). For each sample, isolated RNA was quantified using the Nanodrop 1000 Spectrophotometer (Thermo Scientific) and quality checked using the Agilent 2100 Bioanalyzer. Twenty nanograms of total RNA were reverse-transcribed into cDNA and amplified using the NuGEN™ Ovation™ RNA Amplification System V2 (Catalog number 3100-A01). The cDNA was then purified using magnetic beads (Agencourt RNAClean XP, Catalog number A63987 from Beckman Coulter GmbH, Krefeld, Germany) to remove unincorporated nucleotide triphosphates, salts, enzymes and inorganic phosphates. Purified cDNA was quantified, quality checked and fragmented (at least 3.75µg of cDNA is needed) with a combined chemical and enzymatic reaction, and finally labeled using enzymatic attachment of nucleotides coupled to biotin. Fifty microliters of fragmented and labeled cDNA were added to 170µl of Master Mix Hybridization Cocktail Assembly (Affymetrix GeneChip® Hybridization, Wash, and Stain Kit; Catalog number 900720). After denaturation reaction (2min at 99°C and 5min at 45°C) followed by centrifugation at Vmax for 1min, 200µl of the cDNA cocktail were hybridized on Affymetrix® HG-U133 Plus2 or Rat 230 2.0 GeneChips. The arrays were incubated in the GeneChip® Hybridization Oven 645 (Catalog number 00-0331) for 18±2h at 45°C with a rotation speed of 60rpm. After the hybridization step, the arrays were washed and stained on a Fluidics Station FS450 (Catalog number 00-0335) using Affymetrix® GeneChip Command Console™ Software (AGCC software version 3.2) with protocol FS450_0004. Finally, the arrays were scanned using the GeneChip® Scanner 3000 7G (Catalog number 00-0210).

Raw images from the scanner were saved as DAT files. Using the AGCC Viewer software application, each image was checked for artifacts, overall intensity distribution, checkerboards at the corners, a central cross to ensure adequate grid alignment and readability of the array name. The AGCC Viewer software automatically gridded the DAT file image and extracted probe cell intensities into a CEL file. The CEL files were further processed (MAS5.0) with Affymetrix® Expression Console™ software (version Build 1.3.1.187) for a first quality check of the data. Materials and reagent kits were purchased from Affymetrix, Inc. (Santa Clara, CA, USA), NuGen (San Carlos, CA, USA) and QIAGEN GmbH (Hilden, Germany).

Due to the high number of samples collected for experiment 1, it was not possible to process all samples at once. Therefore, mRNA samples were processed in three batches (samples were randomized within each batch). Each batch contained human and rat mRNAs (in triplicate) for a subset of randomly selected stimuli among those tested (**Figure P2-2**). The same DME control mRNA samples (four replicates) were re-hybridized for each batch. For experiment 2, all mRNA samples were processed together at a single point in time. This included the DME control mRNA samples (four replicates) obtained for this second experiment (**Figure P2-2**).

Data Records

Phosphoproteomics and cytokine data

Raw data processing, normalization and active signals analysis

Phosphoproteomics and cytokine release data were measured using xMAP technology on a Luminex FlexMAP3D® system and the software used was the Luminex xPONENT® for FLEXMAP3D®, Version 4.2. Custom software was developed to analyse the raw data following the standard LXB format data extraction (http://cran.r-project.org/web/packages/lxb/README.html). Following data acquisition, the raw measurements corresponding to the fluorescence intensity of each bead for each individual analyte (protein), were exported. At least 100 events (counts) were measured for each analyte. The median statistic (median fluorescence intensity, MFI) less sensitive to outliers was chosen to summarize data as a

representative value of the protein measurements upon Luminex recommendations. To remove the effects of non-specific binding of proteins to beads in lysates, negative control 'naked' beads (BSA-coated beads devoid of antibody that corresponds to Control B) were prepared using standard coupling procedures. Phycoerythrin-coated beads were also prepared and used as positive control (Control A). Both positive and negative control beads were mixed with the other beads in the multiplex assay. The signal intensities of the negative control beads were found to positively correlate with the signal intensities of the phosphoproteins, which were corrected using a robust linear regression on all replicates[131]. The dependent variable was the signal intensity of a phosphoprotein across stimuli and DME controls (including replicates), and the independent variable was the signal intensity of the 'naked' bead (robust Tukey biweight regressions were calculated with data from experiments 1 and 2, independently). The final normalized signal intensity values for phosphoproteins were taken as the ratio between the residuals and the Root Mean Square Error (RMSE) that resulted from the regression fit. The cytokine data corresponded to the median of the distribution of bead signal intensities measured for each protein in all supernatant samples. In the context of the supernatant, the chance of non-specific binding was reduced when compared to the cell lysate context. Therefore, it was not necessary, to use 'naked' beads (control B) to control for this effect. The median signal intensity values were normalized by calculating z-scores for each cytokine across all stimuli including DME controls. This score was independently calculated for experiment 1 and 2 by taking the ratio of difference between the signal and the mean as well as and the standard deviation calculated for one cytokine across stimuli. For phosphoproteins and cytokines, normalized values beyond 3 standard deviations were considered as active signals.

Quality control analysis

Both individual signals and clusters of signals were examined to achieve the highest quality of the datasets. For each analyte, the final reported signal value was the median of the distribution of individual bead counts because it is less sensitive to outliers and distribution skewness. The minimum number of beads that should be counted for each analyte was also an important parameter to ensure robustness and

reliability of the reported median. The effect of the minimum number of beads required to detect a robust signal was investigated by bootstrapping analysis. The analysis showed that the bead count could greatly affect the reported median, particularly for those with low protein concentrations. Therefore, to further increase robustness, the minimum bead count was increased from 25–50 beads to 100 beads. Furthermore, the distribution of raw (bead signal) measurements for each analyte was examined to evaluate skewness and bi-modality. If the distribution of a bead signal was significantly distorted, the analyte was excluded from the dataset. Finally, to evaluate the precision as well as the robustness of the dataset, each measurement was performed in triplicate while the measurement of the control state (basal level—no treatment) that is crucial in determining the fold increase of the signal from the basal level, was performed in six-plicate. The variability of the replicates of each signal (expressed as median coefficient of variation (CV) across all conditions) served as an estimate of the measurement precision (Supplementary Figure 1 at DOI: 10.1038/sdata.2014.9). RPS6 was excluded for further analysis due to high median CV (Supplementary Figure 1 at DOI: 10.1038/sdata.2014.9).

Data storage

The data are reported as the median of bead signal intensities for each phosphoprotein or cytokine of the panel that have been measured in each sample. For each stimulus, at least 3 sample replicates have been measured for the main experimental phase (Supplementary Table 1 at DOI: 10.1038/sdata.2014.9). The DME control included 5 to 6 sample replicates depending on the experiment. For the screening phase, a single sample was measured for each phosphoprotein and stimulus (Supplementary Table 2 at DOI: 10.1038/sdata.2014.9). Supplementary Table 1 at DOI: 10.1038/sdata.2014.9 (Results of proteomics data including phosphoproteomics and cytokine level measurements for NHBE and NRBE cells exposed to 52 stimuli) and 2 (Results of phosphoprotein measurements in NHBE and NRBE using antibody-bead based assays for the experimental screening of 270 stimuli) have been deposited in the Figshare public repository (Data Citation 2: Figshare http://dx.doi.org/10.6084/m9.figshare.960097).

Gene expression data

Raw data processing and normalization

For each species, all CEL files were processed and normalized together using GC robust multiarray averaging (GCRMA)[132,133]. The data were processed using the GCRMA R package (v2.32) from Bioconductor.

Quality control analysis and differential gene expression analysis

The quality of the chip was assessed at the probe- and probeset-levels by generating different diagnostic plots (chip images, probe-signal intensity distribution, pseudo-images, NUSE (Normalized Unscaled Standard Error) and RLE (Relative Log Expression) plots, correlation matrix) (Supplementary Figure 2 at DOI: 10.1038/sdata.2014.9). Chips exceeding a NUSE median value of 1.05 were considered to be outliers and excluded. Remaining CEL files were re-normalized together per species using GCRMA. The Principal Component Analysis (PCA) of normalized expression data revealed batch effects in both the human and rat gene expression datasets, which were expected, because the samples were processed as distinct batches (Supplementary Figure 2i and j at DOI: 10.1038/sdata.2014.9). No batch correction was done. Instead, batches were treated separately in all analyses. This was made possible by the presence of corresponding 'DME control' samples (at least four replicates) within each batch. Differentially expressed genes were identified by comparing normalized data from DME control with data from each stimulus using limma R-package from Bioconductor. Provided as Supplementary Figure 3 and 4 at DOI: 10.1038/sdata.2014.9, volcano plots indicate the magnitude and the confidence of gene expression regulation for each stimulus (relative to DME control) in human and rat cells, respectively.

Data storage

The raw (CEL. files) and processed (matrices of human and rat gene expression normalized separately; values correspond to log2 expression) gene expression data (Data Citation 1: ArrayExpress E-MTAB-2091) have been submitted to ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) and are available with the accession

number E-MTAB-2091. Metadata were stored in a MAGE-TAB file (SDRF and IDF tabs) supportive of MIAME format for microarray data[134].

Technical Validation

The large multi-omics dataset was generated from *in vitro* cultures to feasibly test a large number of stimuli that were needed to perturb various biological pathways under controlled conditions in both human and rat systems.

Immortalized cell lines have been used in the scientific community for decades in different cellular assays due to their commercial availability at very affordable prices and the ease to culture them. While immortalized cell lines often originate from primary cells/tissues, they have gone through significant mutations, leading to genotypic and phenotypic drifting and eventually to the loss of tissue specific function. In a systems biology perspective, genetic and phenotypic modifications of cell lines have an impact on genome-wide expression profiles and probably also on other large-scale omics approaches and thus could bias data aimed to understand how cellular responses may translate from one species to another[135–139]. For example, it has been shown that the expression profile of primary airway epithelial cells and immortalized cells were different since their expression profiles did not group together using an unsupervised hierarchical clustering approach[140]. Therefore, despite the wide use of immortalized cells, it was decided to work with primary cells that constitute more suitable *in vitro* models to mimic *in vivo* behaviour.

Bronchial epithelial cells were selected as the cell system used for our experiments. The choice for these primary cell types was driven by the fact that these primary cells are at the critical interface between the body and the external environment and were commercially available in both species.

The detailed experimental workflow is described in **Figure P2-3a** starting from the optimization phase to the execution of the main experiment that generated the final datasets for the challenge. The experimental workflow involves various steps, including (i) optimization of the cell culture and experimental conditions; (ii) validation of protein assays (**Figure P2-3b**); (iii) identification, screening and selection of stimuli; and (iv) generation, processing and quality control of omics data.

I-Optimization experiments

Optimization of the cell culture and experimental conditions

Adaptation and optimization of the cell culture conditions originally provided by the vendor for both NHBE and NRBE cells were conducted to avoid spurious differences not associated with the origin of the cells as described in detail in the 'Methods' section.

Bead assay optimization for phosphoprotein and cytokine measurements

Using Luminex's xMAP technology (Luminex Corp, Austin, TX, USA) and ProtATonce multiplex assay optimization (ProtATonce, Athens, Greece), sandwich antibody multiplex assays were employed for the acquisition of both phosphoproteomics and cytokine data. Distinct sets of colour-coded beads with a unique colour-ID formed the solid support for antibody coupling to enable the binding of specific sample proteins on the beads (Supplementary Table 3 at DOI: 10.1038/sdata.2014.9). A biotinylated detection antibody and a streptavidin-reporter dye (phycoerythrin) completed the sandwich enzyme-linked immunosorbent assay (ELISA). The Luminex analyser works as a fluorescence-activated-cell-sorting instrument that simultaneously measures the intensity of the reporter dye and identifies the colour-ID of the bead. The xMAP technology enables the measurement of up to 500 different analytes in a single sample but antibody cross-reactivity limits the simultaneous measurements of analytes to a few dozen. Because the quality of the data is dependent on the quality of antibodies with minimum cross-reactivity (high specificity and minimal background noise), a large number of antibodies from several vendors was purchased and first validated for the xMAP technology according to a six-step process: 1) antibodies were screened to identify optimal antibody pairs, 2) antibodies were captured on different colour-coded magnetic microspheres and the capturing efficiency was confirmed, 3) detection antibodies were biotinylated, and successful biotinylation was confirmed, 4) the concentrations of the capture and detection antibodies were optimized, 5) multiplexability issues caused by the cross-reactivity

of antibodies were identified, 6) the assays were optimized to specific sample requirements (**Figure P2-3b**).

The quality of the antibodies was validated by performing cross-reactivity (assess the specificity of an antibody) experiments in which single purified recombinant proteins were measured using the whole panel of beads (http://www.protatonce.com/#!assay-development/czoq). The antibody selection was based on an optimization algorithm that selected the maximum number of best-performing antibodies and retained the largest possible multiplexability without compromising the signal-to-noise ratio (SNR) (calculated as the ratio between the signal measured for a single purified recombinant protein and the average of signals measured in wells that do not contain the recombinant protein corresponding to background signal). Every antibody was tested against every possible antigen/antibody substrate to create a large matrix representing the specificity of each antibody to each substrate. An optimization problem was formulated to identify antibody pairs with the lowest possible off-target specificity. So if $x_i \in \{0, 1\}$ is the decision whether to include antibody i in the final assay and $C_{i,j}$ is the specificity of antibody i for substrate j, then the problem to solve is: $\min_x \sum_{i,j} x_i x_j C_{i,j}$. The problem was bound to yield a multiplex assay of size N ($\sum_i x_i = N$) and iteratively solve the problem for every N. Finally, the largest multiplex assay that yielded an acceptable background signaling level was selected. The antibodies selected were then tested for their sensitivity to their target protein and those that gave large signal to noise ratios were selected for the final experiments.

This procedure was only possible for phosphoproteomics experiments for which recombinant phosphorylated proteins were available. An alternative solution was to use cell lysates generated from cells exposed to prototypical stimuli known to modulate the phosphorylation of measured proteins. Signal-to-noise ratio (SNR) was used as an assay quality indicator. SNR values, which ranged from 5 to 1700 (no unit), strongly depend on the affinity and concentration of capture and detection antibodies. Multiplexed assays were optimized for the bronchial epithelial cell lysates and supernatants. When low signals were obtained, the concentration of detection antibody was increased to compensate for the low signals. In the absence of signal for all treatments and conditions, the assay was removed. In total, 41 different multiplexed assays for phosphoproteomics were evaluated and 19 met the criteria

described above and were further used in the main experiment (**Table P2-1**). Eighty cytokine assays were evaluated out of which 22 cytokine assays were selected for the main experiment (**Table P2-2**).

Phosphoproteomics and cytokine assay variability assessment for NHBE cells

Potential sources of variability when measuring both phosphoproteins and cytokines were investigated including the inter-donor variability (for human derived primary cells) as well as different factors contributing to the technical variability[141].

Inter-donor variability

Inter-donor variability was investigated in NHBE cells from four different donors under the same conditions. These cells were stimulated with human TNF-alpha (100ng/ml) for 20min to measure phosphoprotein HSP27 levels or PolyI:C (10µg/ml) and for 4h to measure the secretion of CXCL10 protein (n=8 wells of treated cells per donor and n=4 wells of untreated cells per donor). The experiment demonstrated that the donor-to-donor variability was low. Coefficients of variation equal to 11 and 24% for HSP27 and CXCL10, respectively, were calculated together with the mean and standard deviation of the signal values measured in lysates of NHBE cells from the four donors. The main source of variability originated from one particular donor with systematic lower signals for HSP27 and CXCL10. NHBE cells from two donors, which gave similar results for HSP27 and CXCL10, were pooled 1+1 to obtain sufficient cells for the main phase.

Technical variability

As variability may arise from pipetting errors during cell plating, cell lysis or various ELISA technique steps, for each donor, samples from eight different wells were processed identically in parallel. The measurement of the phosphorylated HSP27 and CXCL10 level in these samples were used to quantify well-to-well variability from the same donor. The coefficients of variation for HSP27 measurement after TNF-alpha treatment ranged from 6 to 33%. For CXCL10 measurement, the coefficients of

variation ranged from 7 to 14%.

To assess the variability of the ELISA technique, HSP27 was measured in 12 samples derived from a mixed pool of NHBE cell lysates that were prepared from cells treated with human TNF-alpha (100ng/ml) for 20min. A coefficient of variation equal to 7% was observed.

Reading the same well several times lowered the fluorescence intensity due to a photo-bleaching effect which prevented to determine the instrument variability.

Overall, the technical variability resulting from sample handling and the methods used to determine phosphoprotein and cytokine levels was lower than biological variability, ensuring robustness of data.

Optimization of the experimental design for the phosphoproteomics measurements

To capture the highest number of protein phosphorylation events linked to pathways perturbed upon stimulus exposure, it was essential to determine the optimal time points to harvest both cell types for the main experimental phase. To select these time points, NHBE and NRBE cells were cultured in parallel under the optimized conditions as previously determined followed by exposure to seven prototypical stimuli: TNF-alpha, TGF-alpha, insulin, IL-6, IL1-alpha, IL1-beta, IFN-gamma and a vehicle control (Dulbecco's Modified Eagle's Medium; DME). The concentration of each stimulus was chose based on literature review as described in Step 2 of the study workflow (**Figure P2-3a**). Five different time points were selected (0, 5, 15, 20 and 25min) to measure 41 different phosphoproteins that were plotted using a modified version of DataRail[115] (**Figure P2-4**). For each of the time point, the fold increase of each signal was calculated as compared to the basal level. The two time points (5 and 25min) with the maximum number of activated signals and the largest fold increase for both cell types were selected as the optimal time points for phosphoproteomics measurements in the main experiment. The selection of 5 and 25min was also consistent with earlier studies done with hepatocytes[142].

II-Stimulus selection process

The goal of the Species Translation challenge was to understand and provide insight on how far the translation concept can be applied between rodents and humans at various layers of biological molecules measured *in vitro*. This implied to activate or repress a large range of pathways/biological functions to ensure broad perturbation coverage of the biological system. Therefore, **Figure P2-4** illustrates our strategy in selecting various stimuli to perturb as many pathways as possible and how an initial experiment was performed to screen various phosphorylated proteins following stimuli exposure. This strategy provided a final selection of stimuli active in human, rat or both species to be used for the main experiment.

Stimulus selection by in silico analysis and literature review

The following criteria were considered for the initial selection of potential candidate stimuli, including 1) stimuli that modulate the activity of transcription factors/regulators; 2) classical stimuli known to target specific pathways; and 3) stimuli with heterogeneous downstream effects. Computational and manual curation approaches were undertaken to achieve an appropriate selection (**Figure P2-5**).

Stimuli that modulate the activity of transcription factors/regulators

Transcription factors/regulators directly regulate the transcription of target genes. Querying of databases containing biological knowledge, such as the Ingenuity database (Ingenuity® Systems, www.ingenuity.com), enabled the identification of compounds that could modulate the activity of transcription factors/regulators expressed in the tissues/cells (lung, lung cells, lung cell lines, lung tissue, small airway epithelial cells, airway epithelium and airway epithelial cells) and organisms of interest (human, mouse and rat). Overall, 710 compounds were identified that could modulate the activity of 182 transcription factors/regulators. Compounds that modulate the activity of many different transcription regulators (e.g., beta-estradiol) were prioritized and retained in the initial stimulus list.

Stimuli known to target specific pathways

Prototypical stimuli that have been extensively used as proxy tools to perturb

(activate or inhibit) specific pathways were identified from the literature and selected (e.g., rapamycin, an inhibitor of the mTor pathway; lipopolysaccharide, an activator of the NFκB signaling pathway; and tunicamycin, an inducer of the unfolded protein response). Some cytokines and growth factors were also selected because they target very specific pathways.

Stimuli with heterogeneous downstream effects

Analysis of the connectivity map (CMAP) large-scale expression compendium enabled the selection of compounds that induce heterogeneous downstream effects[143]. The CMAP dataset is a collection of genome-wide gene expression profiles that represent the transcriptional responses of five different human cell lines (HL60, MCF7, PC3, SKMEL5 and ssMCF7) to 1,309 different compounds (small active molecules) or control vehicles following 6h of exposure (12h for a specific subset of compounds)[143]. Interestingly, Lorio and colleagues have constructed a 'drug network' partitioned into communities using the CMAP dataset[144]. Drugs within a community were clustered together on the basis of similar regulation patterns of gene expression, suggesting an analogous mode of actions. This 'drug network' was leveraged to identify drugs/compounds with heterogeneous downstream effects by computing a between- *versus* within-community (B/W) average distance ratio. Compounds with the highest ratios were prioritized during the screening of the stimuli. Including the review of the scientific literature, a total set of 270 stimuli were selected for *in vitro* testing as described below.

In vitro stimulus screening

In vitro screening was performed to identify a subset of stimuli that could elicit responses in NHBE and NRBE cells, which then would be used for the main experiment (**Figure P2-5**). This *in vitro* screening was performed by measuring the phosphorylation levels of proteins from the lysates of NHBE and NRBE cells that were exposed to 270 compounds for 5min. For each of the 270 compounds, the concentration was manually curated from the literature or through a semi-automated literature-mining approach (Supplementary Table 2 at DOI: 10.1038/sdata.2014.9

deposited in the Figshare public repository). Briefly, the Google search engine was used to query the names and common aliases of the 270 compounds found in the HUGO database (for cytokines and growth factors) followed by the concentration units (i.e., 'mg/ml', 'ng/ml', 'mM', 'μM', 'nM'). The concentrations were automatically extracted from the top 100 results and plotted on a histogram. A rounded value 20% above the median of the histogram was chosen as the final concentration.

Final stimulus selection for the main experiment

Out of 270 compounds that were originally used in the screening of phosphoproteomics, the most potent compounds were selected for the main experiment. These compounds, including activators/inhibitors, were chosen based on (i) the number of phosphorylation signals that were affected, (ii) the strength of their responses (maximum fold increase of the activated signals), (iii) the diversity of the affected pathways, and (iv) and downstream gene expression changes (see paragraph 'Stimuli with heterogeneous downstream effect') (Supplementary Table 2 at DOI: 10.1038/sdata.2014.9 deposited in the Figshare public repository). With respect to these criteria for the final selection of stimuli, the following analysis was performed. The screening phase included one biological replicate for each stimulus/protein/species, therefore, a statistical analysis could not be performed and an alternative approach was followed. For each protein, a fold change of the signal was calculated comparing the phosphorylation signals measured for compound-treated cells and for unstimulated cells (control). Subsequently, a number of thresholds ranging from 1.25 to 2.0-fold (i.e. 1.25, 1.3, 1.35, 1.4, 1.45, 1.5, 1.55, 1.60, 1.65, 1.7, 1.75, 1.8, 1.85, 1.9, 1.95 and 2.0-fold) were used to binarize signals as active or non-active. For example, a threshold of 2.0 implies that an increase of 2-fold or higher was required for the signal to be considered activated. For each threshold, the number of signals considered as up- and down-regulated was calculated for each stimulus across all phosphoproteins. This number was used as a score to sort the stimuli from the most to the less potent. The most potent stimuli were prioritized also with respect to the other criteria (mentioned above) for the final selection of stimuli (Supplementary Table 4 at DOI: 10.1038/sdata.2014.9).

III-Main experiment

The cell culture and experimental conditions established for NHBE and NRBE cells during the optimization phase were applied in the main experiment. Both NHBE and NRBE cells were seeded in 96-well plates on day one and incubated overnight at 37°C in 5% CO2. The cells were starved in bronchial epithelial cell basal medium (Lonza) for 4h and then exposed, in parallel, to 52 different stimuli or to a control medium (DME), which is the standard culture medium for these cells. The cells were collected and lysed at different time points (5 and 25min for phosphoproteins measurement, 6h for gene expression measurement) and supernatants were collected at 24h for cytokine release measurement in the supernatants. The cells were exposed to each stimulus in triplicate, or in 5-plicates and 6-plicates for the DME controls. To avoid spatial confounding effects, the stimuli and DME controls were randomly distributed throughout the 96-well plate.
Datasets for the main experiment were generated from two sets of independent experiments: 75% of the stimuli (40 compounds) were tested first; whereas, the remaining 25% (12 compounds) were tested in a second experiment. For each experiment DME controls were included. The final STC compendium contains a collection of phosphoproteomics, transcriptomics, and cytokine data.

Co-authored Research: A crowd-sourcing approach for the construction of species-specific cell signaling networks

Please visit DOI of paper [P3] for more information on this chapter.

Abstract

Motivation: Animal models are important tools in drug discovery and for understanding human biology in general. However, many drugs that initially show promising results in rodents, fail in later stages of clinical trials. Understanding the

commonalities and differences between human and rat cell signaling networks can lead to better experimental designs, improved allocation of resources and ultimately better drugs.
Results: The SBV IMPROVER Species-Specific Network Inference challenge was designed to use the power of the crowds to build two species-specific cell signaling networks given phosphoproteomics, transcriptomics and cytokine data generated from NHBE and NRBE cells exposed to various stimuli. A common literature-inspired reference network with 220 nodes and 501 edges was also provided as prior knowledge from which challenge participants could add or remove edges but not nodes. Such a large network inference challenge not based on synthetic simulations but on real data presented unique difficulties in scoring and interpreting the results. Because any prior knowledge about the networks was already provided to the participants for reference, novel ways for scoring and aggregating the results were developed. Two human and rat consensus networks were obtained by combining all the inferred networks. Further analysis showed that major signaling pathways were conserved between the two species with only isolated components diverging, as in the case of ribosomal S6 kinase RPS6KA1. Overall, the consensus between inferred edges was relatively high with the exception of the downstream targets of transcription factors, which seemed more difficult to predict.

Introduction

Unveiling the inner workings of cell signaling networks is one of the long-standing challenges of systems biology. Small-scale versions of these networks have been built edge by edge using classic laboratory techniques such as immunoprecipitation, which has resulted in a large body of literature describing various gene and protein interactions. Although successful in their initial scope, these methods do not scale up to the genome level and are difficult to combine into a larger network, because of the different contexts in which they were originally reported. Organism, cell type, experiment timing and other conditions are crucial for determining whether an edge exists in a signaling network.
The advent of large-scale assays that can simultaneously measure the activity of

thousands of genes has circumvented these aforementioned issues by enabling purely data-driven methods to infer large-scale networks. Various algorithms have been developed, including models based on Bayesian networks[145], mutual information[146], regression[147], neural networks[148], Boolean networks[23] and differential equations[149]. Despite these advances, there is no clear best method. Each method has strengths and limitations influenced by how the methodology addresses the fact that network inference is inherently an underdetermined problem in the majority of cases[150,151]. However, it has been observed that the aggregation of different network inference methods generates high-quality robust results[150].

Efforts to catalog and compare network inference algorithms have occurred in the form of data prediction competitions such as the ones organized by the Dialogue for Reverse Engineering Assessments and Methods (DREAM) consortium[152]. DREAM challenges participants to reconstruct cell signaling networks from gene expression datasets. Predicted networks are then evaluated based on a subset of known interactions, or the complete network in cases where the corresponding gene expression data were generated *in silico* (i.e. simulated).

DREAM is part of a larger group of successful crowd-sourcing initiatives in systems biology alongside CASP [critical assessment of protein structure prediction[153]], CAFA [critical assessment of function annotation[154]], CAPRI [critical assessment of prediction of interactions[155]], FlowCAP [critical assessment of automated flow cytometry data analysis techniques[156]] and Foldit [predicting protein structure with a multiplayer online game[157]]. In the same spirit as these academic initiatives, sbv IMPROVER is a crowd-sourcing-based methodology for the verification of research in an industrial setting[158]. In its second instalment, it challenges the research community to solve four problems related to the translation of molecular biology findings between rat and human model systems[159].

Here we present the analysis of the Species-Specific Network Inference challenge, part of the sbv IMPROVER Species Translation set of challenges (https://www.sbvimprover.com). For this challenge, participants were asked to infer human- and rat-specific networks given phosphoprotein, gene expression and cytokine data (**Figure P3-1**). The organizers also provided a common reference network from which participants had to generate the two networks by adding and

removing edges. The purpose of this challenge was to augment and refine the reference map in a species-specific manner using data-driven approaches.

Methods

Evaluation of inferred networks

Most of the prior knowledge regarding the interactions between elements in the cell signaling network was already incorporated in the reference map provided to the challenge participants. Hence, this information could not have been used as a gold standard against which to evaluate inferred networks. To circumvent this issue, we proposed that the true ranking of the submissions be viewed as a prediction problem in itself by combining different scoring strategies. Rank-based aggregation of individual predictions has been shown to provide robust results on par with the best performing methods in other data prediction challenges[150,160,161]. Drawing from this result, the predicted networks were evaluated using softer methods that did not involve the use of a gold standard where the final ranks were derived by simply averaging the ranks obtained using the different scoring strategies.

The first scoring method involved the use of a published network inference algorithm[23] to generate a 'silver standard' network against which all submissions were evaluated. This is mainly a pruning algorithm, hence only the subnetworks that intersected the reference network were scored. The following metrics were considered for this purpose: the z-score of the Jaccard similarity (JS), Matthews correlation coefficient (MCC) and the difference between the true-positive rate and false-positive rate (TPR-FPR)[162]. In addition, two versions of the silver standard were generated: one that was trained on only the data available to the participants and one that also made use of part of the dataset that was kept hidden from participants and used as the gold standard in the other Species Translation challenges.

For another scoring method, the write-ups describing the methodology used for making the predictions were scored based on the following criteria: rigor, defined as the soundness of the proposed methodology based on valid statistics, arguments and premises without gaps in a logical, well-defined sequence of procedural steps;

originality, defined as novelty in concept when compared with existing methods and typical approaches in the field; and practical implementation, defined as the ability to instantiate the proposed methods with existing or clearly described novel algorithms and commonly used computer architectures, the use of data sources commonly available to the field and a reasonable execution time. Three independent evaluators blindly assigned scores ranging from 1 (very poor) to 5 (very good) for each criterion, and then the final score was obtained by adding these points and then averaging among the evaluators.

The reference network

The reference network represents an ensemble of canonical pathways and was built following a top-down multi-layer hierarchical architecture starting with the stimulus layer through multiple signaling cascades to the transcription factor (TF) and secreted cytokine layers (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). Only stimuli with known mode of action present in subset A (training dataset) were included in the reference network.

The signaling cascade layer connected stimuli to latent (i.e. not measured) and measured (phosphoproteins) nodes representative of a membrane-to-nucleus protein signaling cascade (i.e., from stimuli to TF via kinase proteins). The identification and prioritization of latent nodes and edges (connectivity between stimuli, latent nodes and measured nodes) were conducted using various biological pathway databases (e.g. KEGG, Biocarta) and the ensemble network published by[163], embodying the union of several online pathway databases. The network was traversed using a depth-first search algorithm, computing its transitive closure and identifying paths. Latent nodes that were not transitively connected to a stimulus or a measured node were removed. Additional latent nodes were identified based on topological features of the ensemble network. These highly connected nodes (counting the largest amount of incoming and outgoing connections) with a minimum overlap between them were identified using standard k-means clustering and integrated in the reference network. For the TF layer, TFs corresponding to a subset of measured nodes and selected latent nodes were connected to a subset of target genes. These genes were identified

using the Transcriptional Regulatory Element Database[164,165].

The cytokine layer was constructed by connecting target genes to corresponding measured cytokines. The final step included a manual verification and curation of the reference network to prune and refine it using literature reviews and various pathway databases (e.g. Biocarta, KEGG).

The silver standard network

The construction of the silver standard networks was based on a method developed by Mitsos *et al.* The outline of this approach is to use Boolean logic to model signal transduction and integer linear programming (ILP) to fit the model to the data. In particular, Boolean logic was used to represent signal transduction in a prior knowledge network (i.e. reference network) to create a model capable of predicting the state of a node in a given experiment. Because Boolean models are limited to qualitative predictions, discretization of the experimental data was necessary. The discretization of the datasets was done by use of double threshold functions. In particular, the thresholds were set at ±2-fold changes for the gene expression data, ±3 standard residuals for the phosphoproteomics data and ±2 standard residuals for the cytokine data. The initial choice of thresholds was done in accordance with the processing methods used for the different data types as described in[166]. In addition, a sensitivity analysis was performed to ensure that the final network would be robust on slight variations of the thresholds.

ILP was further used to combine the Boolean model with the experimental data by formulating an optimization problem that sought to minimize the mismatches between the predictions derived from the final network and the data at hand. The optimization procedure was performed by removing reactions from the reference network that were contradicted by the data and thus created a smaller data-specific network. More details about the silver standard model are available in the Supplementary Data at DOI: 10.1093/bioinformatics/btu659.

The consensus network

The predictions from each of the M challenge participants can be organized as a binary vector xj=(x1j,x2j,...,xNj) where j=1...M and N is the total number of possible edges, while the unknown gold standard network is represented as the vector y=(y1,y2,...,yN). Each element xij or yi can either be 1 (edge exists) or 0 (edge does not exist).

Let PT be the probability that a method predicts the existence of an edge given that the edge exists, and PF the probability of predicting the existence of an edge given that the edge does not exist. If Xi and Y are random variables with realizations xj and y, respectively, then Xi|Y=1∼Bernoulli(PT) and Xi|Y= 0 ∼Bernoulli (PF). Assuming that the predictions are independent given the true edge label, then the conditional distributions of the sum X=X1+X2+...+XM are modeled by the Binomial distributions:

$$\Pr(X=k|Y=1)=(Mk)PkT(1-PT)M-k \quad (1)$$

$$\Pr(X=k|Y=0)=(Mk)PkF(1-PF)M-k \quad (2)$$

where k is effectively the number of 'votes' received by an edge. Therefore, the probability density function that k teams picked the same edge is as follows:

$$\Pr(X=k)=ENPr(X=k|Y=1)+N-ENPr(X=k|Y=0) \quad (3)$$

where E is the number of true edges.

The Equations (1) and (2) assume the performance of predictions is constant, modeled by parameters PT and PF; however, this is not true in practice. The variation in prediction performance between different algorithms can be modeled by imposing PT and PF to follow Beta distributions, normally used to model random variables limited to intervals of finite length. Consequently, the conditional probability functions in Equations (1) and (2) become the Beta-Binomial compound distributions:

$$\Pr(X=k|Y=1) = (Mk)B(k+a1, M-k+b1)B(a1,b1) \quad (4)$$

$$\Pr(X=k|Y=0) = (Mk)B(k+a2, M-k+b2)B(a2,b2) \quad (5)$$

where PT follows the beta distribution B (a1, b1) with shape parameters a1 and b1, and PF follows the beta distribution B(a2,b2) with shape parameters a2 and b2.

The model described by the Equations (3), (4) and (5) can be fitted to the distribution of the data comprising the number of times each edge was present among the different proposed networks. An optimal consensus network can be reconstructed using all the predictions by finding the minimum number of votes per edge that

satisfies the condition Pr(Y = 1|X = k) > Pr(Y = 0|X = k). This threshold can be easily found by numerically solving the following equation:

r(Y=1|X=k)Pr(Y=0|X=k)=Pr(X=k|Y=1)Pr(Y=1)Pr(X=k|Y=0)Pr(Y=0)=1 (6)

where Pr(Y = 1) and Pr(Y = 0) are prior probabilities related to the true number of edges in the network:

Pr(Y=1) = EN (7)

Pr(Y=0) = 1−EN (8)

Results

The described methodology for building the reference network created a directed graph with 220 nodes and 501 edges organized into cascading layers where the edges are oriented from the top to the bottom layers. At the top is the stimulus layer that contains a subset of the compounds used to generate the training data, followed by receptor, adaptor, signaling, TF, target and cytokine layers (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). It is interesting to note that not all the TFs are reachable from all the stimuli nodes. The addition of a top stimuli layer to an otherwise generic network introduces the notion of context to pathways that are only active under certain conditions.

By mapping the nodes from the reference network to the genes from the canonical pathways listed in the Molecular Signature Database v3.1[167], we observe a diverse representation of cellular processes. Among the most common were cell growth and survival (*EGF*, *INSULIN*, *PDGF* and *RAS*), interleukin (*IL1R*, *IL3* and *IL4*), inflammatory response (*NFKB*) and cell signaling (*MAPK*) as shown in **Figure P3-2**.

Comparison of predicted networks

Challenge participants were allowed to add or remove edges from the reference network, although they were not allowed to add extra nodes. The purpose of this was to make submissions comparable and to put some boundaries that were relevant to the experiments performed. It is interesting to note that most proposed networks were built by removing edges from the reference network rather than adding

additional interactions, which led to a bigger consensus on the existence of edges that were already part of the reference network. (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). The median number of edges of the proposed networks was 406 for human and 429 for rat compared with 501 edges of the reference network.

In the case of the silver standard, two versions of the networks were considered: one that relied only on the training dataset and numbered 131 edges for human and 175 for rat, and one that used the full dataset (training and testing sets) and numbered 114 edges for human and 162 edges for rat. The JS between the two silver standards was 0.50 for human and 0.67 for rat. However, when using the two silver standard versions to evaluate the submissions, the scores obtained were very similar (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). This led to the decision to use only the first proposal, which used the same data as the challenge participants. The heatmaps in **Figure P3-3** show the similarity between predicted networks together with the silver standard using MCC in the space of the reference network edges. Both panels suggest an emerging pattern where a few of the networks are more similar to each other and to the silver standard. The same can be observed when looking at the number of edges that overlaps between the different networks (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). These are the networks that were ranked higher independent of the scoring metric used (i.e., JS, MCC or TPR-FPR) (Supplementary Data at DOI: 10.1093/bioinformatics/btu659).

The second method for evaluating submissions used the scores obtained by the accompanying write-ups describing the algorithms used to build the species-specific networks. The scores listed in Supplementary Data at DOI: 10.1093/bioinformatics/btu659 are separated by criterion (originality, rigor, practical implementation) and show remarkable consistency between reviewers. In the end, the final ranking was calculated by averaging the ranks obtained by each team for the two scoring methods and are listed in Supplementary Data at DOI: 10.1093/bioinformatics/btu659.

All the predicted networks can be used to construct a consensus network by keeping the edges chosen by at least a predetermined number of teams. Supplementary Data at DOI: 10.1093/bioinformatics/btu659 shows how all the participating teams would

have fared against a consensus network constructed using different thresholds from three to seven teams. Because two of the teams had similar entries (**Figure P3-3**), one of them was discarded (Team 93) to avoid bias when the consensus network was built. It is worth noting that the top performing teams determined by consensus scoring using large thresholds (Teams 116 and 55) were the same as the ones that were the challenge best performers according to Supplementary Data at DOI: 10.1093/bioinformatics/btu659. In contrast, the performance of lower ranked teams was less consistent between the different scoring strategies.

Optimal consensus network

The optimal threshold for building the consensus network was determined by fitting the model described in Section 2 (Equations 3, 4, and 5) followed by solving Equation 6. The data used for the fit were assembled by counting the number of 'votes' received by each edge in the reference network from the participating teams (excluding Team 93) and the silver standard network. This was performed separately for human and rat networks, and then the resulting datasets were mixed to improve the fit. Maximizing the log likelihood function of the mixture of two beta-binomial distributions (Equation 3) for different mixing constants led to $\Pr(Y = 1) = 0.16$, $\Pr(Y = 0) = 0.84$ (**Figure P3-4**A) and shape parameters a_1 = 8.77e+06, b_1 = 1.95e+06, a_2 = 3.46e+06 and b_2 = 1.57e+06. Using this result and after solving Equation 6, it was found that it takes approximately eight votes to verify the condition $\Pr(Y = 1|X = k) > \Pr(Y = 0|X = k)$. This result can also be visualized in **Figure P3-4**B by tracing the intersection of the two mixture components depicted in black.

The model was tested on two additional datasets and showed a good overall fit (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). In the first case, the edge counts shown in **Figure P3-4**B were extended to all possible edges, including the ones not present in the reference network. In the second case, a completely new set of network predictions was obtained from DREAM 3[168]. For this challenge, 27 participants had to predict *de novo* a synthetic network with 50 nodes and 82 edges from simulated gene expression data without knowing the identity of the nodes.

Conservation and divergence of human and rat cell signaling networks

Using the threshold determined in the previous section, two consensus networks were built for human and rat using the networks predicted by participants together with the silver standard. The individual edges that resulted are depicted in Supplementary Data at DOI: 10.1093/bioinformatics/btu659 and color-coded based on their presence in the human, rat or both consensus networks. The number and the size of the resulting connected components are listed in Supplementary Data at DOI: 10.1093/bioinformatics/btu659. Two of these subnetworks are shown in **Figure P3-5** panels A and B as examples of predicted differences between human and rat cell signaling networks. Although there were plenty of edges that were active only in human or rat, these differences were rather isolated. The differences between human and rat did not scale up to the level of pathways or other higher levels of organization, as will be reinforced in the following analysis.

For any group of edges, a consensus score can be calculated by averaging the individual scores associated with each edge, which is simply the percentage of times the edge was predicted to exist. Here we assume that consensus between participants regarding an edge is associated with higher probability that the edge is real. The panel C in **Figure P3-5** shows the average consensus scores of the edges between consecutive layers for human and rat together with the associated standard errors. Although there were no significant differences between human and rat, the overall consensus for the edges downstream of TFs seemed to be much lower than the rest, suggesting that these edges were more difficult to predict. The consensus scores of the edges in the canonical pathways listed in **Figure P3-2** also showed no significant differences between human and rat (Supplementary Data at DOI: 10.1093/bioinformatics/btu659).

The conservation of phosphoprotein activity was measured by calculating the average consensus score of all edges adjacent to a phosphoprotein node (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). From all the proteins measured, RPS6KA1 had a significantly higher consensus score in human (P-value = 0.0161) and WNK1 had a significantly higher consensus score in rat (P-value = 0.0498). Similarly, the conservation of TF activity was assessed by calculating the

consensus score of the edges upstream of a TF (Supplementary Data at DOI: 10.1093/bioinformatics/btu659) and then downstream of it (Supplementary Data at DOI: 10.1093/bioinformatics/btu659). Edges upstream of *STAT1* had a higher consensus score in human than in rat (*P*-value = 0.0004), whereas edges downstream of *MYC* also showed a higher consensus score in human (*P*-value = 0.0287). Significantly higher consensus scores in rat were found for edges downstream of *TCF3*, *GLI2* and *SMAD3* (*P*-values = 8.8-e06, 0.0287 and 0.0156).

Discussion

The scope of sbv IMPROVER Species Translation challenges was to assess the limits of using rat models to predict human biology in the specific context of bronchial epithelial cells exposed to various stimuli. Along these lines, the rationality behind the Network Inference challenge was to build two species-specific cell signaling networks starting from a generic literature-inspired network and using high-throughput proteomics and transcriptomics data to add or reject edges. This challenge differed from other challenges because it did not come with a gold standard (i.e. the true human and rat networks are unknown) and this posed difficulties in scoring and interpreting the results. The current work details how the aforementioned issues were addressed together with the lessons learned from organizing and curating such a challenge.

Despite the apparent top-down organization of the reference network, some feedback loops were present consistent with the structure of known pathways. However, the challenge experiments were designed to capture a broad area of the signaling network and not feedback mechanisms. The latter would have required a different experimental setup with more samples collected at later time points, as feedback loops tend to be more prominent at longer time scales.

Without a gold standard, individual scoring criteria can potentially be useful in separating poor performers from good performers but can also have flaws. The silver standard is biased by the choice of algorithm used to generate it, and the quality of the write-ups does not always predict the best performing algorithms. It is thus advisable to combine the rankings resulting from individual scoring methods to

reduce bias. The best performers obtained in this manner were the same as the ones obtained by comparing predictions with a consensus network built by aggregating the submissions from all participants. This result suggests that consensus scoring could be used as a legitimate scoring strategy for future challenges where a gold standard is absent.

The network aggregation procedure described in this article provides a statistically sound way of merging predicted networks or any other binary predictions given a sufficiently large sample space. This is especially useful when a clear way of assessing the best performing method is absent. However, even when one can accurately determine the best algorithm for performing a specific task, the result might be context dependent. It has been shown that disease classifiers vary greatly in performance when applied to different datasets[169]. Aggregating multiple predictions has been proven to generate a more robust outcome on par with the best performing methods[150,160].

The generation of a consensus prediction can potentially have benefits beyond that of robustness and performance, particularly in the absence of a gold standard. The data shown in Supplementary Data at DOI: 10.1093/bioinformatics/btu659 suggest that predictions can be scored against a consensus network instead of using a silver standard, with similar top rankings when an appropriate threshold is used. Consensus scoring can thus avoid any bias caused by the choice of algorithm for the silver standard; however, it could be sensitive to outliers (e.g. predictions that are much better than the rest), or multiple correlated predictions caused by collaborating teams or the use of similar methods.

The predicted networks were aggregated using a mixture of two beta-binomial distributions as shown in Section 2. To find the optimal threshold for determining the existence of an edge, a two-step process was used. First, the distribution in Equation 3 was fitted to the consensus data; then the minimum number of teams k was determined for which $\Pr(Y = 1|X = k) > \Pr(Y = 0|X = k)$. From the first step, the value of the mixture constant $\Pr(Y = 1)$ (Equation 7) can give an indication of the proportion of true edges in the reference network which in this case was 16%. Despite this, the solution to the second step resulted in consensus networks with 7.4% edges for human and 6.7% edges for rat out of all the reference network edges. This result

suggests that less than half of the potential regulatory connections were discovered and more challenge participants were needed to increase statistical power and reconcile the two estimates of the number of true edges.

Despite these limitations, the consensus network shown in Supplementary Data at DOI: 10.1093/bioinformatics/btu659 displays some interesting patterns, some of which are shown in **Figure P3-5**A and B. Overall, the cAMP-responsive element-binding protein 1, also known as *CREB1*, showed the best consensus for the edges upstream of it (Supplementary Data at DOI: 10.1093/bioinformatics/btu659) but with a couple of differences between human and rat: the connection from *RPS6KA1* was present only in the human consensus network[125], whereas the connection from *PRKACA* was present only in the rat consensus network[126]. The prevalence of *RPS6KA1* (a.k.a. *RSK1*) interactions in human (Supplementary Data at DOI: 10.1093/bioinformatics/btu659) might be explained by the fact that human isoforms of RSK1 have functional redundancy (i.e. *RPS6KA3* [*RSK2*]; *RPS6KA2* [*RSK3*]; and *RPS6KA6* [*RSK4*]). In contrast, this is most likely not the case in rodents; Zeniou et al.[127] reported that the mouse *RSK1* and *RSK3* genes may not be able to fully compensate for the lack of *RSK2* function.

The consensus results also suggest a preference for *JAK1* activation through *EGFR* for human and the *PDGFR* complex for rat. Direct interaction between *JAK1* and *IRS1* has been reported in cultured human peripheral blood T cells[170]. In rat, however, the interaction seems to be indirect through proteins *SOCS2*, *SOCS3* and *JAK2*[171,172]. Other conserved interactions include *IFNGR1* to *JAK2* and *JAK2* to *STAT5A*, which are parts of the interferon gamma pathway known to be conserved across vertebrate species[173].

Additional references are provided for the majority of edges from the consensus networks and are available as Supplementary Data at DOI: 10.1093/bioinformatics/btu659. These references are categorized by organism and tissue context as follows: airway cells, non-lung cells and lung cancer epithelial cells. Although numerous pathway databases are widely available, they are too generic and lack specific context when displaying an interaction. The purpose of this challenge was to fine tune one of these generic networks based on data collected from bronchial primary cells exposed to specific stimuli (compounds). When comparing

the resulting consensus network to networks obtained from the Ingenuity Pathway Analysis tool (IPA: www.ingenuity.com), we observe a steady increase in precision as the number of votes required for an edge increases (Supplementary Data at DOI: 10.1093/bioinformatics/btu659), culminating at eight votes as predicted by the model in Equation 6. The maximum precision obtained is 0.33 for the human network and 0.09 for the rat network. However, this could be explained by the relatively few edges identified in IPA for human (69) and especially rat (26) (the number of edges drastically decreased if a filter on cell/tissue type was applied), as well as the lack of proper context provided by tissue specificity and stimuli. The IPA networks as well as the consensus networks are available as Supplementary Data at DOI: 10.1093/bioinformatics/btu659.

Overall, the fact that fewer suitable edge additions existed in most inferred networks (Supplementary Data at DOI: 10.1093/bioinformatics/btu659) indicates that the reference network contains probably most of the true active pathways in both species. However, as observed by the large number of edge removals, it also contains many inactive pathways. In other words, the phosphoproteins represented by network nodes were less responsive to some stimuli than expected from the reference network. Furthermore, because most participants (and all top performers) used the reference network in their models it is likely that expert/prior knowledge was critical for optimal network construction.

The methods used by the participants to solve the challenge were varied and included Bayesian networks, Boolean networks, mutual information, lasso and elastic net, ANOVA and various heuristics (more details on the individual algorithms are available in the Supplementary Data at DOI: 10.1093/bioinformatics/btu659). It is interesting to note that different flavors of the same method, in this case Bayesian networks, do not perform similarly when applied to the same problem. When designing a prediction algorithm, a multitude of choices were made, ranging from various constants and priors to learning criteria and regularization options, which can lead to vastly different outcomes. This justifies efforts, such as the sbv IMPROVER challenges or any of the other crowd-sourcing initiatives such as DREAM or CASP, to try and establish best practices in the ever-changing field of computational biology.

Multiple Sclerosis

The disease

Multiple Sclerosis (MS) is an autoimmune disease that affects the brain and spinal cord. An estimated 2,500,000 around the world have MS and there is not yet a cure for the disease. Even though significant progress is currently being made in MS research, the pathogenesis of the disease has not been comprehensively understood. A great number of pathological mechanisms responsible for the disease have been described, involving hundreds of genes and proteins altering multiple processes and signaling pathways. By understanding how current MS therapies work in biological networks, more effective therapies can be designed.

Experimental Model

- Peripheral blood mononuclear cells (PBMCs)

Experimental Tools

- Blood from 255 donors was collected.
- PBMCs isolation and culture.
- Stimulation with 20 compounds, lysis and collection of cell supernatant.
- Measurement with bead-based sandwich ELISA of 17 phosphoproteins at 5 and 25 minutes and 22 secreted cytokines at 24 hours post stimulus.

Computational Tools

The PhD candidate was not involved in the computational analysis of this dataset.

Main Findings

We created a huge dataset for Multiple Sclerosis: 250 donors (190 MS, 60 Healthy), 20 stimuli (Pro-inflammatory or pro-oxidant stimuli, immunomodulatory stimuli, neuroprotectants or anti-oxidants, disease modifying drugs from MS: Dimethyl Fumarate, Fingolimod, Teriflunomide, IFNB1a/Rebif), 17 phosphoproteins (5' and 25'), 22 cytokines (24h). Our approach allows to characterize the signaling networks in a patient-specific manner and to predict new targets for combination therapy for MS.

The combination of fingolimod with either a TAK1 inhibitor or EGCG was validated in an animal mode.

Publication Timeline

[A2] 17-19 June 2014, Boston MA, USA
Construction of a drug-induced phosphoprotein/cytokine dataset in clinical samples for Multiple Sclerosis

[P4] 1 February 2015, Multiple Sclerosis Journal
Signaling networks in MS: A systems-based approach to developing new pharmacological therapies

[A7] 5-8 May 2015, Berlin, Germany
Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis

[A8] 8-9 June 2015, Berlin, Germany
Multiplex High Throughput screening for Drug Discovery: Application for mechanism-based biomarker discovery in Multiple Sclerosis

[A9] 1-3 July 2015, Avignon, France

Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis

[A10] 6-8 July 2015, Heidelberg, Germany
Patient-specific signalling pathway analysis for Multiple Sclerosis

Construction of a drug-induced phosphoprotein/cytokine dataset in clinical samples for Multiple Sclerosis

Please visit abstract [A2] in section “Abstracts in International Conferences” for more information on this chapter.

Multiple Sclerosis (MS) is an autoimmune disease that affects the brain and spinal cord. An estimated 2,500,000 around the world have MS and there is not yet a cure for the disease. Even though significant progress is currently being made in MS research, the pathogenesis of the disease has not been comprehensively understood. A great number of pathological mechanisms responsible for the disease have been described, involving hundreds of genes and proteins altering multiple processes and signaling pathways. By understanding how current MS therapies work in biological networks, more effective therapies can be designed. On this front, the CombiMS consortium (http://combims.eu/) is developing computational and experimental tools to improve the therapeutic options of MS in the future.
A milestone in this consortium is to evaluate how current and “promising” MS drugs work at the signaling level in different patient populations. Peripheral blood mononuclear cells (PBMCs) from approximately 255 donors were collected from several European medical centers. The PBMCs were plated in 96 well plates and 20 stimuli & drugs were applied. Using custom multiplex assays, 17 phosphoproteins plus 2 control beads were measured at 5 and 25 minutes and 22 cytokines plus two control beads at 24 hours post stimulus.
To ensure the highest possible data quality, a kit was developed to include all reagents needed in order to isolate PBMCs from one donor and then plate, stimulate and lyse them. Sample collection controls were applied in the stimuli set to evaluate

errors in sample processing whereas two custom xMAP bead sets were used to evaluate errors in the bead-based ELISA procedure and instrument measuring variability.

Phosphoproteomic and cytokine data will be combined with SNP data and clinical profiles (i.e. responders, non-responders, therapeutic intervention) in a computational framework which will help to understand MS more thoroughly and systematically. As a first step, the phosphoproteomic dataset will be used for the construction of a detailed map of the signaling pathway differences between MS and healthy donors, which can help generate a model of MS pathogenesis and improve our understanding of the disease.

Co-authored Research: Signaling networks in MS: A systems-based approach to developing new pharmacological therapies

Please visit DOI of paper [P4] for more information on this chapter.

Abstract

The pathogenesis of multiple sclerosis (MS) involves alterations to multiple pathways and processes, which represent a significant challenge for developing more effective therapies. Systems biology approaches that study pathway dysregulation should offer benefits by integrating molecular networks and dynamic models with current biological knowledge for understanding disease heterogeneity and response to therapy. In MS, abnormalities have been identified in several cytokine-signaling pathways, as well as those of other immune receptors. Among the downstream molecules implicated are Jak/Stat, NF-Kb, ERK1/3, p38 or Jun/Fos. Together, these data suggest that MS is likely to be associated with abnormalities in apoptosis/cell death, microglia activation, blood-brain barrier functioning, immune responses, cytokine production, and/or oxidative stress, although which pathways contribute to the cascade of damage and can be modulated remains an open question. While current MS drugs target some of these pathways, others remain untouched. Here,

we propose a pragmatic systems analysis approach that involves the large-scale extraction of processes and pathways relevant to MS. These data serve as a scaffold on which computational modeling can be performed to identify disease subgroups based on the contribution of different processes. Such an analysis, targeting these relevant MS-signaling pathways, offers the opportunity to accelerate the development of novel individual or combination therapies.

Introduction

Multiple sclerosis (MS) is a chronic disease that is known to involve both inflammatory and neurodegenerative responses. Despite the significant progress made in recent decades, we are still relatively far from achieving a comprehensive understanding of the pathogenesis of this disease. The revolution in molecular biology, immunology and genetics, along with the development of new high-throughput technologies, has driven the production of large amounts of data in recent years. However, while numerous genes and proteins have been associated with the disease, significant gaps remain in the quest to understand the pathological mechanisms responsible for MS. Although there is still hope that new studies will reveal specific genes, proteins or cells that will explain an important proportion of the causes of the disease, what is urgently needed is to integrate the available and any future data into a comprehensive dynamic picture of MS[174].

Unfortunately, knowing that a gene or cell type is associated with MS is far from providing an explanation about the disease. This is related to the fact that to date no mutated genes associated with MS have been found, although new sequencing studies are under way that may change this fact, and therefore such genes do play the physiological role expected of them, complicating the analysis[175]. Moreover, in a complex disease such as MS, genes, proteins and cells dynamically interact with each other in response to the stimuli and challenges the immune and nervous system face[176]. This quantitative and dynamic information is extremely difficult to capture from patients and even from animal models. Second, each individual harbors a different genetic background and also the development of the immune and nervous system is customized for their environment during development, being one of the

bases of disease heterogeneity. Therefore, without the integration of molecular information in pathways and considering molecular and cellular heterogeneity, it will be difficult to achieve a good understanding of the pathogenesis of MS. At the individual patient level, it will be critical to collect personalized data to customize the analysis to pave the way toward stratified medicine.

Systems biology approaches may offer important benefits integrating current biological knowledge with clinical information and data on therapeutic responses, thereby allowing models to be generated that might help explain the pathogenesis of the disease[176,177]. Therefore, in this review we will focus on how a systems biology approach applied to medicine (systems medicine) from the pathway perspective, incorporating molecular information about MS pathogenesis and drug targets, could improve our understanding of the disease and help in the development or identification of new improved therapies.

Pathways regulating MS pathogenesis: A puzzleof the immune system, the central nervous system (CNS) and missing pieces

Decades of cellular and molecular research in the field of MS have revealed many genes, proteins and cell subpopulations of the immune system associated with the disease, and such information has expanded massively with the new omics technologies. In order to identify the pathways involved in a given disease, abundant information is available in databases such as the Gene-Disease Association Database, the Protein Sequence Database, the Comparative Toxicogenomics Database, the Online Mendelian Inheritance in Man, the Genetic Association Database, or the Literature- derived Human Gene-Disease Network. Moreover, genetic susceptibility for MS has been revealed by genome-wide association studies (GWAS) and the ImmunoChip study in MS, which have identified more than 100 single-nucleotide polymorphisms (SNPs) associated with the disease[178,179], which have been implicated mainly in immune system pathways (leukocyte activation, apoptosis, and positive regulation of macro-molecule metabolism, Janus kinase/ signal transducers and activators of transcription (Jak/Stat) signaling pathway, acute myeloid leukemia, and T cell receptor signaling)[180]. Moreover, several data- bases

containing information about pathways are available, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, the PathwayCommons or the ConsensusPathDB. Finally, chemoinformatic resources such as DrugBank, ChEMBL and Drug Information Online contain information about drugs, including their targets within human pathways. By combining the information available in these databases, more than 40 pathways associated with MS can be found (**Figure P4-1**(a)). The overall picture obtained reveals the involvement of a wide range of cellular processes and pathways implicated: apoptosis/cell death, microglia activation, blood-brain barrier functioning, immune response, cytokine production, or oxidative stress. In addition, the search of these cellular processes can be combined with the targets of MS treatments, such as fingolimod, dimethyl-fumarate or interferon-beta (IFNβ). The targets of these therapies can be included by identifying the pathways that link them to the processes above, namely lipid-mediated signaling and its crosstalk with survival and nuclear factor (NF)-Kb pathways, anti- oxidant pathways and the Stat-mediated IFNβ response, respectively. Interestingly, pathway analysis revealed certain processes that are not yet targeted by current therapies, e.g. the Notch pathway[181–183]. Another pathway related to MS is vitamin D metabolism, and at present there are several trials testing the efficacy of vitamin D supplementation[184].

New roles in MS of components in known pathways

Once data are retrieved from databases, bioinformatic tools allow identification of interactions between genes, proteins and cells that can be used as hypotheses (**Figure P4-1**(b)). For instance, and of pivotal clinical importance, these tools help to study the involvement of CNS pathways in MS pathogenesis. MS is a condition associated with substantial neuronal and axonal damage, and this neurodegeneration probably drives long-term neurological disability[185]. Hence, pathways that are related to neuronal death and axonal damage may be of particular interest for the development of neuroprotective therapies, an approach pursued for decades without success to date, including apoptosis, oxidative stress, or microglia activation. Potential neuroprotective therapies under development are aimed to target such pathways such as green tea catechin epigallocatechin-gallate, trophic factors,

methylthioadenosine or drugs enhancing remyelination[186–189].

Patient-to-patient genetic variabilityhinders understanding of signaling pathways involved in MS

It is striking that the recent massive genetic studies (e.g. GWAS, ImmunoChip) explain so little of individual disease risk[178,179]. It is also surprising that no single MS therapeutic yet has had well-validated genetic stratification. We believe that this fact, common to many complex diseases, is at least in part due to the lack of a functional, network-based perspective of the pathogenesis of the disease. Signaling networks are very robust to variation in cells and their environment to enable cellular functioning. For instance, it has been shown that even clonal populations strongly vary in the concentration of the same protein[190]. The cellular function regulated by that protein needs to remain unaltered for healthy cellular behavior, therefore a number of network motifs grant robustness to signaling networks such as negative feedback loops[191]. Other variations to which signaling pathways can be robust are genetic polymorphisms. Therefore, including genotyping data when modeling signaling pathways of MS patients is key to understanding MS pathogenesis.

Furthermore, some pathways may not be etiologically relevant because they are associated due to co-segregation of alleles with diseases. One approach for integrating the role of genetic susceptibility in systems biology methods is by considering that risk alleles mildly modify the parameters governing the functioning of the pathways. Therefore, one single allele may not have a significant effect in a given pathway, but the collection of all the risk alleles in a given individual may influence the function of immune pathways to the level of producing autoimmune activation. These considerations may help to improve the prediction of autoimmune response at the individual level. The same reasoning may apply to the fact that no drug stratifies by any single risk allele, although this may change if new therapies target MS-associated genes and one of the alleles modulates the biological effects of the drug.

In summary, understanding how each individual's genetic polymorphisms lead to their specific signaling network activity would enable characterization of the different

MS phenotypes. This would, however, suggest a further question: How does patient-to-patient variability in signaling activity affect drug efficacy? To answer this question, signaling networks need to be elucidated not only depending on donor genetic variability, but also in a cell-specific manner, thereby determining how each pathway contributes to the cell phenotype (pro-inflammatory, degeneration, repair), and to identify the missing steps (molecules and interactions) that participate in such MS pathways. For example, interleukin (IL)-10 is one of the main immunosuppressor cytokines but clinical trials with IL-10 failed to show benefits in MS, probably because IL-10 receptor signaling is at least partially deregulated in immune cells[192], and this can influence the individual response. We envision that a fine mapping of specific pathways such as cytokines pathways in specific cell types (e.g. CD4, CD8, B cells) in parallel with large high-throughput studies will allow us to improve pathway annotation. Coupled computational modeling and experimental validation will enable characterization of signaling networks in a cell-type-, donor- and genetic variant-specific manner, as reviewed in detail below.

Other challenges in pathway analysis

The fact that database searches identify many path- ways associated with MS in immune cells might suggest that there is significant cross-talk between the major pathways within the same cell, with important proteins participating in several signaling cascades (Jak/Stat, NF-Kb, extracellular signal-regulated kinase (ERK)1/2, p38, Jun/Fos). Crosstalk within pathways in the same cell is complex and thus difficult to study based only on existing experiments. Second, there is a substantial gap in our understanding of how such crosstalk interactions are translated into a cell-type-specific response at the system level (e.g. IFNb produces different effects on macrophages and T cells, which are related to different clinical effects). Third, it is particularly difficult to make sense out of the existing MS data, since it is a disease that affects arguably the two most complex tissues/ organ systems in our body, i.e. the immune system and the CNS, as described above. Fourth, annotation of gene function is still incomplete, and the role of the same genes in the CNS is often even less well under- stood or as yet unknown. For example, tumor necrosis factor-alpha

(TNFα) may have detrimental effects on the immune system in MS but it might also be beneficial in the CNS during remyelination[193]. Indeed, TNFα promotes oligodendrocyte progenitor proliferation, as well as remyelination, which probably explains why the application of the TNF-antagonist Lenercept produced an unexpected deterioration of MS[194]. Finally, a principal limitation in pathway analysis using existing data for functional annotation is that these approaches do not provide a mechanistic model that can be simulated, hence hindering the prediction of novel signaling mechanisms. To solve the challenges described here, combined analysis of newly acquired experimental data and mathematical models can be used[195]. Next, we review in detail such predictive models.

Predictive and mechanistic models to understandMS pathogenesis and therapies

The past decade has seen an explosion in the information regarding the cellular networks that transmit and process signals from the cell's environment. To gain novel understanding of the basic mechanisms that the cell uses to integrate these signals, as well as how such mechanisms are impaired by disease, mechanistic—mathematical—models are a powerful tool[196]. The first step to mathematical modeling is a literature search to gather the current understanding regarding the molecular process of interest, in this case MS. To that end, we can query public resources, a process that yields the pathways known to be involved in MS, which in turn are combined to form a signaling network. Signaling networks can be used as initial scaffolds on which we can formulate mechanistic hypotheses and evaluate similarity with experimental data and disease-driven changes (**Figure P4-2**, upper row). Therefore, experimental data need to be acquired that measure as many readouts as possible relevant to the disease of study, i.e. present in the signaling network. To that end, phosphoproteomic measurements are key in the analysis of signaling pathways because measuring abundance of phosphorylated proteins closely indicates propagation of a signal through a pathway and can be used in functional models[197]. Previous work in the field has provided clear examples that phosphoproteomic analysis is able to provide accurate models of some pathways in cells such as

hepatocytes[198]. Bead-based enzyme-linked immunosorbent assays (ELISAs) of xMAP technology (Luminex, Austin, TX) are well suited for this task, enabling measurement of the abundance of a large number of phosphorylated proteins in the above-mentioned MS pathways in immune cells of individual patients of different cohorts. Combining phosphoproteomics with genotyping in mathematical models, both the activity of MS pathways and the genetic variability that may explain the patient- to-patient difference in terms of response to treatment can be studied (**Figure P4-2**, upper row). Once a signaling network has been assembled via literature search, and the data to compare it have to been measured, mathematical approaches enable formalization of the net- work as a mechanistic model. Intuitively, the formulation as a mathematical model of such a signaling network addresses two limitations: They are neither specific to individual patients (or even often to specific cell types), nor are they computable, i.e. can be used to predict the outcome of perturbations with drugs and ligands.

Several mathematical modeling approaches have become well established in the field of systems biology and can be applied to signaling pathways, ranging from logic to physicochemical models[199]. The lack of quantitative information for building the models can be bridged by using logic (Boolean) modeling, which includes only causal information and that, because of this simplicity, has many fewer parameters (quantitative properties) to evaluate. This advantage can be used to represent large signaling networks that can be generated with limited data[200]. To implement logic models, tools such as CellNOpt[110] enable formalization of the signaling network as a logic model and subsequent simulation. Next, these tools enable calibration of the model, which is performed by changing the network topology, i.e. the shape of the network in terms of the interactions between the present signaling intermediates. These changes consist of introducing or removing interactions, and systematic comparison of the simulations on different topologies against the experimental data predicts the topology that best fits the data. The simplicity in logic modeling that enables simulation of large networks at the same time hinders highly detailed modeling of small networks. In more detailed analyses, appropriate tools are physicochemical models that describe the underlying biochemical reactions explicitly[195,201,202]. Here the model parameters are quantitative characteristics such as

kinetic rates of the reactions that they represent, which are revealed by model calibration against the experimental data. In both modeling approaches, the main challenge lies in calibrating the model in order to make the model specific for MS, while at the same time determining the factors contributing to patient-to-patient variability. To address patient-to-patient variability, instead of starting from a single signaling network, one solution is to calibrate an ensemble of networks featuring a high number of different starting topologies in order to test many different hypotheses that are compared separately to the xMAP and genotype of individual patients[203], including in these signaling topologies the mechanisms that grant robustness to signaling pathways, such as negative feedback loops. Thereby, we could determine which of the ensemble of topologies best fits each individual patient (**Figure P4-2**, middle row). Overall, the modeling of signaling pathways in MS, using either logic networks or mathematical models, offers the opportunity to predict new signaling mechanisms that help us better understand disease pathogenesis. For example, a recent mathematical model of the type 1 IFN pathway revealed the translocation of Stat-1 to the nucleus as the most critical step in the signaling of IFNb, a finding that could not be predicted solely based on molecular analysis but required dynamic simulations[204].

Drug development and combination therapy in predictive models of MS

One obvious question is if recent technological developments have provided a large amount of data about MS, why is drug discovery still so complex and provides such limited results? Although the limitations of the drug-discovery process have been reviewed in detail[205], several specific issues regarding how biological information is translated into models of the disease and pathways are of importance. In the process of developing useful pathway models for drug development it is critical to take into consideration many aspects that at present are not well covered, such as (i) the availability of quantitative and kinetic data from human/patients, (ii) integration of individual heterogeneity and genetic background for defining the response to therapy, or (iii) the need to develop approaches for integrating and simulating complex networks of not just cells but also tissues. As described in **Figure P4-2**, here we

propose that coupling several omics and genotyping to mathematical models of signaling networks can address these issues. Further, existent drugs can be repurposed to target MS-related components by including their targets in a signaling network that can then be formalized as a mathematical model and simulated. This would enable the testing of millions of different options in terms of topology of signaling networks, therapeutic regimens, and drug/target combinations. This should allow the prediction of the signaling mechanisms by which these existing drugs could be repurposed to MS, discarding therapeutic approaches that may not work and pointing to the ones that deserve careful experimental and clinical testing.

Given the complexity and heterogeneity of MS, combination therapies that modulate various pathogenic pathways simultaneously are an attractive treatment strategy[206]. A synergistic effect of two drugs with different mechanisms of action may potentially improve efficacy, safety and tolerability. By contrast, defining the optimal combination of drugs requires a more comprehensive understanding of the networks of pathways in different cells initiating and driving the progression of MS, an effort that can be addressed using systems biology techniques[207]. The integration of clinical, biological and pharmaceutical data in computational models that reproduce the complexity of such diseases can be used to identify synergistic effects by evaluating the downstream effects of drugs[208].

Finally, another significant challenge in improving drug development is predicting side effects of therapies.

Predictive toxicity was something highly theoretical until recently, but in the last years new significant insights have been provided by developing new algorithms combining drug databases and safety databases. Prior knowledge extracted from such databases can be introduced in mathematical models (**Figure P4-2**) that are starting to provide useful predictions regarding potential side effects that can be tested in preclinical or early clinical phases of drug development[209,210]. However, this complex issue is still far from being solved.

Conclusions

The pathogenesis of MS is complex, involving hundreds of genes and proteins that

act in numerous path- ways and evolve along a time and disease progression, each of which can contribute to the phenotype. These genes and proteins may respond distinctly to different therapies, and even behave differently in different patients. In order to integrate current knowledge and generate a comprehensive model of MS pathogenesis, pathway analysis represents a promising strategy. Combining experimental and medical data with distinct systems biology approaches should provide new insights on disease pathogenesis, allowing us to screen in silico new drugs for repurposing, as well as testing combinations of drugs, before exposing patients to therapy.

Prediction of combination therapy based on logic modeling of the immune cells signaling network in Multiple Sclerosis

Signal transduction deregulation is a hallmark of many complex diseases, such as multiple sclerosis (MS), and subsequently a therapeutic target. Here, we performed ex vivo multiplexed phosphoproteomic assays in PBMCs from 190 MS patients either untreated or treated with fingolimod, natalizumab, interferon-beta, or glatiramer acetate or the experimental therapy epigallocatechin gallate (EGCG); and 60 matched healthy controls. Fitting a prior knowledge-based network of MS related pathways using logic modeling yielded a signaling network specific for the disease and for each drug. We then established a co-druggability score based in the pathway topology for each drug's signaling network and identified kinase interactions whose activity could be reverted to healthy-like status by combination therapy. Our method predicted several combinations with approved MS drugs, being the interaction of IKKB by TAK1 co-druggable with all approved drugs. The combination of fingolimod with either a TAK1 inhibitor or EGCG was validated in the animal model of MS. Our approach allows to characterize the signaling networks in a patient-specific manner and to predict new targets for combination therapy for MS and other complex diseases.

Subjects and clinical cohorts

We recruited 255 subjects including 190 patients with MS and 60 healthy controls, matched for age and sex with the RRMS group, in a multi-centric study in four MS centers (Hospital Clinic Barcelona – IDIBAPS (n=69), Karolinska Institute (n=64), University of Zurich (n=40) and Charite University (n=82)) (Table S6). Patients fulfilled McDonald 2005 criteria (37) and disease subtype was defined using Lublin criteria (38). Patients were allowed to receive any therapy and should be stable in such therapy for the previous 6 months. Patients were recruited by their neurologist after signing informed consent. The study was approved by the IRB of each clinical center.

Processing and measurement of samples

A unified standard operating procedure for PBMCs isolation, stimulation and lysis, as well as sample storing and shipping was developed along with a kit (plates) with reagents and buffers that were produced in a single facility (ProtAtOnce) and ship to all participating centers. The reagents were prepared from a single batch and plates were prepared from a single batch for each stimulus. Quality controls were carried out to ensure that the reagents remain stable for 3 months.

XMAP assays were developed by ProtAtOnce (Athens, Greece) and were standardized to minimize error. We optimized assays from a list of 70 candidates and obtained a final list of 17 phosphoproteins which display a good signal to noise ratio to be measured in the in vitro assays: AKT1, CREB1, FAK1, GSK3A, HSPB1, IKBA, JUN, MK03, MK12, MP2K1, PTN11, STAT1, STAT3, STAT5A, STAT6, TF65, WNK1. We used a set of 21 stimuli, which included pro-inflammatory or pro-oxidant stimuli (Anti-CD3, concanavaline A (conA), H2O2, IFNG, IL1A, IL6, LPS, NaCl, PolyIC, TNFA), immunomodulatory stimuli (IFNB1a, S1P, vitD3) neuroprotectants or anti-oxidants (BDNF, EGCG, INS), disease modifying drugs from MS (DMF, FTY, Teriflunomide, IFNB1a (Rebif), and a culture media as control. Samples were collected at baseline (time 0) and after 5 and 25 min.

Multiplex High Throughput screening for Drug Discovery: Application for mechanism-based biomarker discovery in Multiple Sclerosis

Please visit abstracts [A7-10] in section “Abstracts in International Conferences” for more information on this chapter.

High throughput screening (HTS) has become an invaluable tool in drug discovery. Current HTS platforms are based on the measurement of very few reporter assays and lose the mechanistic picture on how compounds work on cells. Multiplex HTS can measure a large number of intracellular and extracellular signals and thus provide mechanism-based insight in biomarker discovery.

Multiple Sclerosis (MS) is an autoimmune disease affecting the brain and spinal cord. There is not yet a cure for the disease while 2,5 million people around the world have MS. Several pathological mechanisms for MS have been described, involving alterations in multiple processes and signalling pathways.

Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective therapies can be designed that take into account the uniqueness of each patient’s response in treatment and biomarkers can be developed to stratify patients.

For the production of this dataset, we used High-Throughput Screening (HTS) with custom cell signalling assays (by ProtATonce), that allow screening of thousands of samples at the proteomic and phosphoproteomic level at a very small fraction of the cost compared to off-the-shelf reagents. Strict quality control check points were embedded to the collection-to-measurement pipeline in order to evaluate errors in sample collection, sample preparation, ELISA procedure and instrument variation.

Collection kits were prepared and shipped to 4 clinical centres across EU in order to collect peripheral blood mononuclear cells (PBMCs) from 255 donors. The cells were plated and stimulated at 3 time points with 20 compounds and drugs. We collected

cell lysates and cell supernatants to simultaneously quantify in the samples 17 phosphoproteins and 22 secreted proteins respectively.

This dataset is analysed with systems biology algorithms in order to construct patient-specific signalling pathways and develop biomarkers for MS. Machine learning and optimization algorithms were employed to quantify patient-based drug efficacy and predict responders from non-responders. In summary, multiplex high throughput screening and systems biology algorithms introduce a mechanistic insight in biomarker discovery and pave a new way for pharmaceutical research.

Hepatocellular Carcinoma

The disease

Hepatocellular carcinoma (HCC) is the fifth most common malignancy worldwide, with a high rate of metastasis. Despite major improvements in HCC management (resection, transplantation, radiofrequency ablation, chemoembolization, sorafenib therapy), long-term survival remains poor[211,212]. Therefore, there is an urgent need to identify novel HCC chemopreventive and/or therapeutic agents able to protect populations at high risk and/or improve prognosis of patients following curative treatment[213].

Experimental Model

- Human Hepatocellular Carcinoma cell lines HEP3B, HEPG2 and HUH7
- Human umbilical vein endothelial cells (HUVEC)

Experimental Tools

- Matrigel for migration and invasion assays
- MTT assay for viability assessment
- Cell culture of standard cell lines
- Stimulation, lysis and collection of cell supernatant
- Measurement with bead-based sandwich ELISA
-

Computational Tools

- Integer Linear Programming formulation to combine proteomic data at hand with prior knowledge of proteins' connectivity to construct specific signaling

pathways

Main Findings

Our integrated analysis presents new mechanistic insights into the targeted anti-inflammatory actions of three promising nutraceuticals, epigallocatechin gallate (EGCG), fisetin (FIS), and eriodictyol (ERI), thus setting the basis for innovative HCC chemopreventive and/or therapeutic interventions. Our results demonstrated EGCG as the most effective modulator of inflammatory cytokine secretion (followed by FIS and ERI) and HEP3B cells as the best responders. Despite previous extensive literature[214,215], this is the first study to our knowledge showing the outstanding capability of this compound to concurrently reduce a wide range of HCC-secreted cytokines.

Publication Timeline

[P5] 1 June 2015, CPT: Pharmacometrics & Systems Pharmacology
Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action

Co-authored Research: Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action

Please visit DOI of paper [P5] for more information on this chapter.

Abstract

Chronic inflammation is associated with the development of human hepatocellular carcinoma (HCC), an essentially incurable cancer. Anti-inflammatory nutraceuticals

have emerged as promising candidates against HCC, yet the mechanisms through which they influence the cell signaling machinery to impose phenotypic changes remain unresolved. Herein we implemented a systems biology approach in HCC cells, based on the integration of cytokine release and phosphoproteomic data from high-throughput xMAP Luminex assays to elucidate the action mode of prominent nutraceuticals in terms of topology alterations of HCC-specific signaling networks. An optimization algorithm based on SigNetTrainer, an Integer Linear Programming formulation, was applied to construct networks linking signal transduction to cytokine secretion by combining prior knowledge of protein connectivity with proteomic data. Our analysis identified the most probable target phosphoproteins of interrogated compounds and predicted translational control as a new mechanism underlying their anti-cytokine action. Induced alterations corroborated with inhibition of HCC-driven angiogenesis and metastasis.

Introduction

Hepatocellular carcinoma (HCC) is the fifth most common malignancy worldwide, with a high rate of metastasis. Despite major improvements in HCC management (resection, transplantation, radiofrequency ablation, chemoembolization, sorafenib therapy), long-term survival remains poor[211,212]. Therefore, there is an urgent need to identify novel HCC chemopreventive and/or therapeutic agents able to protect populations at high risk and/or improve prognosis of patients following curative treatment[213].

Compelling evidence has established that inflammation plays a critical role in tumor progression[216,217]. HCC, a typical example of inflammation-related cancer, slowly progresses on a background of chronic inflammation mainly triggered by exposure to infectious agents (hepatotropic viruses), toxic compounds (ethanol), or dietary carcinogens (aflatoxins, nitrosamines). Inflammatory cytokines have been shown to play a prominent role in mediating changes within the tumor or tumor microenvironment through abnormal regulation of signaling pathways that influence crucial cancer-promoting processes such as cell proliferation, survival, angiogenesis, and metastasis[218]. On the other hand, the neoplastic cell itself can advance the

development of local inflammation, as many of the most frequently activated oncogenes (*RAS*, *MYC*) could elicit a transcriptional program leading to the expression of various inflammatory mediators that lead to tumor immune escape and expansion[216].

At the genome level, HCC is characterized by several aberrations which underlie dysregulation of multiple steps in cell signaling pathways[219]. Therefore, modulation of a single gene product or signaling cascade is unlikely to mediate an efficient therapeutic outcome. Hence, current research efforts have been focused on developing multi-targeted therapies using novel high-throughput technologies[220]. Nutraceuticals (a term coined from "nutrition" and "pharmaceutical") i.e., compounds in dietary sources with disease chemopreventive/chemotherapeutic activities, have been proved to possess such multi-targeting properties and reduced side effects, thus providing a suitable alternative in achieving alleviation of various cancers[214,221]. Despite the emerging social and economic interest in nutraceuticals[220], there is not yet a global mechanistic understanding of their action mode to provide a scientifically solid support for their clinical use.

Identifying the detailed mode of action of active compounds is a major endeavor today, as is evident by the large number of publications on off-target effects of drugs that made it to the market even decades ago and whose off-target effects remained unknown[222,223]. Typically, the identification of a drug mode of action is carried out either via bioactivity assays, which screen the effect of the compound on key cellular processes (proliferation/viability, etc.) or via kinase assays, which screen the binding of the interrogated compound on a kinase panel[224–226]. However, even if the drug interactions are easy to obtain, the chain effects happening in the cells because of the inhibition of a certain kinase is not trivial to predict, as it is orchestrated to a great extent by the properties and robustness of the signaling mechanisms of the specific cell type. Herein, we developed a novel methodology that leverages multi-combinatorial proteomic data from multiplex antibody assays and screens the effects of selected nutraceuticals on three human HCC model cell lines both on the basal level of cytokines release and phosphoproteins activity, but also on perturbation with prototypical cell growth and inflammation signaling stimuli, in an attempt to capture alterations induced by each interrogated compound on key signaling pathways. To

this end, we employ an optimization algorithm based on the *SigNetTrainer* software that combines the proteomic data with prior knowledge of protein interactions (e.g., from online pathway databases), identifies the signaling reactions that are functional based on the data at hand, thus deconvoluting compound effects, and facilitates their mechanistic interpretation as topology alterations of cell-type-specific signaling pathways[15]. Furthermore, we use experimental models of cancer metastasis and angiogenesis to ensure the biological importance of compound-imposed alterations. Our integrated analysis presents new mechanistic insights into the targeted anti-inflammatory actions of three promising nutraceuticals, epigallocatechin gallate (EGCG), fisetin (FIS), and eriodictyol (ERI), thus setting the basis for innovative HCC chemopreventive and/or therapeutic interventions.

Methods

Plant-derived nutraceuticals

Nutraceuticals (Supplementary Table S1 at DOI: 10.1002/psp4.40) were purchased from Carl Roth (Karlsruhe, Germany), or Sigma (St. Louis, MO). Stocks were prepared in dimethyl sulfoxide (DMSO; Sigma) and stored at –20°C. Working dilutions contained up to 0.1% v/v DMSO.

Cell cultures

HEP3B, HEPG2 (ATCC, Manassas, VA), and HUH7 (provided by J. Wands, Brown University) cells were maintained as described before[227]. To obtain HCC conditioned medium (HCC-CM), confluent cultures were treated with test compound or DMSO for 24 hours, thoroughly washed twice with phosphate-buffered saline (PBS) and incubated for 24 hours in starvation medium, i.e., serum-free medium containing 0.25% bovine serum albumin (BSA). Supernatant was then collected and stored at −80 °C. Human umbilical vein endothelial cells (HUVEC) were isolated and maintained up to passage four as previously described[228].

To assess the potential cytotoxicity of test compounds, cells were treated as specified for each of the following biological assays and the number of viable cells was measured by the MTT method as described before[229].

xMAP assays

HCC cells, seeded the day before at 3 × 10^4 cells/well into a 96-well plate, were serum-starved for 4 hours and then treated with selected nutraceuticals or DMSO at the indicated concentration (**Table P5-1**) for 2 hours. Cells were then exposed to stimuli (PeproTech, Rocky Hill, NJ) at saturated levels[227]: interleukin (IL)-6 (0.1 μg/ml), tumor necrosis factor A (TNFA) (0.1 μg/ml), IL1A (0.02 μg/ml), IL1B (0.01 μg/ml), tumor growth factor A (TGFA) (0.2 μg/ml), or insulin (INS) (1.72 μg/ml), for either 22 hours (cytokine measurements) or 15 minutes (phosphoprotein measurements) as indicated by previous studies[227] and pilot experiments (Supplementary Figure S1 at DOI: 10.1002/psp4.40). Cell supernatants and lysates were then collected and stored at −80°C until use. Total protein concentrations were quantified using the BioRad Dc Protein Assay kit (Hercules, CA).

xMAP assays were performed on a Luminex-200 platform (Luminex, Austin, TX) using custom cytokine/ phosphoprotein antibody-coupled beads (ProtATonce, Athens, Greece). A custom 28-plex was used to detect the levels of selected cytokines in cell supernatant (see Supplementary Table S2 at DOI: 10.1002/psp4.40 for full name, classification, and biological role). A custom 15-plex was used to determine in cell lysates the levels of test phosphoproteins: ribosomal protein S6 kinase alpha-1 (RSK1), heat shock protein beta-1 (HSPB1, alternate name HSP27), cAMP-responsive element-binding protein-1 (CREB1), protein kinase B (AKT1), p38 mitogen-activated protein kinase (P38MAPK), mammalian target of rapamycin (TOR), glycogen synthase kinase-3 beta (GSK3B), dual specificity mitogen-activated protein kinase kinase-1 (MEK1), extracellular signal-regulated kinase-1 (ERK1), Src homology 2 domain-containing protein-tyrosine phosphatase-2 (SHP2), c-Jun N-terminal kinase-2 (JNK2), dual specificity mitogen-activated protein kinase kinase 6 (MP2K6), ribosomal protein S6 (RPS6), p70 ribosomal S6 kinase (P70S6K), and nuclear factor-kappa B (NFKB). Custom antibody-coupled beads were technically validated as

described before[166] (Supplementary Figures S2-S5, Supplementary Table S3 at DOI: 10.1002/psp4.40).

xMAP data processing

Signaling datasets were analyzed and plotted using Datarail, an opensource MATLAB (MathWorks, Natick, MA) toolbox[110]. For modeling, the fold change of the signals relatively to the unstimulated state, i.e., DMSO-MEM, was computed and then data were discretized to [0, −1,1], where 1 denotes signal increase ≥1.5 or 2-fold, −1 denotes signal decrease >1.5 or 2-fold, and 0 denotes any signal activation in between. The thresholds were selected based on the platforms sensitivity (see Supplementary Material at DOI: 10.1002/psp4.40) and previous studies[230,231]. Data discretization is vital before proceeding with the construction of compound-specific signaling networks, since the applied *SigNetTrainer* methodology implements a qualitative approach in the modeling of signal transduction and can only handle the above-mentioned values for protein activation [−1, 0, 1].

Construction of compound-specific signaling networks

A canonical pathway was constructed downstream of the five stimuli and in the neighborhood of 15 measured phosphoproteins. First, the stimuli receptors were identified, and then canonical pathways downstream of these receptors were extracted from several online databases including Pathway Commons, KEGG, and Ingenuity, with most of the interactions obtained from Ingenuity. Subsequently, the receptor-specific pathways were merged together into a signaling network, and the Floyd-Warshall algorithm was used to identify the observable-controllable part of this network. Observables are defined as all network nodes that are upstream of the measured phosphoproteins (thus, their activation value can be inferred based on the value of the signals downstream). Controllables are defined as all network nodes that are downstream of stimuli used (thus, their activation value can be controlled by the stimuli upstream). All nodes and corresponding reactions that are either

nonobservable or noncontrollable were removed from the network, because the Integer Linear Programming (ILP) algorithm cannot handle them properly.
Subsequently, the canonical pathway was augmented with compound→target interactions obtained from PubChem (Supplementary Table S5 at DOI: 10.1002/psp4.40) and with hypothetical phosphoprotein→cytokine release interactions to result in an integrated network that describes both levels (intracellular and extracellular) of signal transduction. Finally, the signaling network was trained to compound-specific data via an adapted variant of *SigNetTrainer,* an ILP formulation that detects and removes inconsistencies between network predictions and data at hand. Details of the formulation, model code, and datasets are provided as Supplementary Material at DOI: 10.1002/psp4.40.

Migration, invasion, and tube-like formation assays

Cell invasion and migration were evaluated as described before[229] using an 8 μm-pore size membrane BioCoat Matrigel Invasion (BD Biocoat, BD Biosciences, Franklin Lakes, NJ) and migration (ThinCerts, Greiner Bio-One International, Kremsmuenster, Austria) chambers, respectively. Serum-starved cells ($3–5 \times 10^4$cells/well) migrated towards HCC-CM or starvation medium for 24 hours (HCC) or 6 hours (HUVEC). Migrating/invading cells were photographed and manually counted. Formation of HUVEC tube-like structures in the presence of HCC-CM or starvation medium was assessed on growth factor-reduced Matrigel and quantified as previously[232].

Statistical analysis

Data were analyzed by Mann–Whitney test with significance levels of $P < 0.05$, using GraphPad Prism 5.00 for Windows (GraphPad Software, San Diego, CA). Data are presented as means±SEM of the indicated number of observations.

Results

Collection of released cytokine data

Since HCC cells have been shown to express a number of proinflammatory cytokines and growth factors, we examined whether nutraceuticals with known anti-inflammatory properties might be able to modulate their secretion pattern. HCC-effective compounds were chosen from a library of 23 nutraceuticals based initially on preliminary HCC cell proliferation/viability experiments determining the range of effective non-cytotoxic concentrations (Supplementary Table S1 at DOI: 10.1002/psp4.40) and next on pilot xMAP assays evaluating the effect of 14 chosen compounds (selected as potent HCC cell proliferation inhibitors that represented major structural classes of nutraceuticals), on the levels of 13 cytokines upon induction with a stimuli mixture (Supplementary Figure S6 at DOI: 10.1002/psp4.40). Finally, nutraceuticals found to be active in at least one cell line were included in a more extensive xMAP analysis at selected concentrations (**Table P5-1**), which although high in some cases (POH, NAR, EGCG) did not affect normal endothelial cell proliferation (unpublished observations). Multiplex cytokine assays concomitantly determined the release of 28 cytokines (Supplementary Table S2 at DOI: 10.1002/psp4.40), constitutively or upon stimulation with IL6, TNFA, IL1B, TGFA, or INS. As shown in **Figure P5-1**, some cytokines (CXCL16, IL8, GROA, CCL20, NGAL, HAVR1) were highly expressed by HCC cells, whereas administration of stimuli, especially IL1B and TNFA, further enhanced their basal levels or triggered the secretion of numerous other cytokines. EGCG treatment proved to be the most effective in reducing the expression (basal and/or induced) of 13 out of 28 test cytokines, without affecting others (e.g., CCL20, PLGF, HAVR1, MIA2, NGAL, PDGF), thus excluding nonspecific cytotoxicity. FIS, ERI, QUE, CUR, POH, and NAR displayed (in order of efficacy) a less broad inhibitory action. Overall, we noticed that HEP3B were the most sensitive cells to compound treatment (see Supplementary Table S4 at DOI: 10.1002/psp4.40 for a summary of data).

Collection of signaling phosphoprotein data

Next, we analyzed the impact of EGCG, FIS, and ERI (selected based on their cytokine-reducing activities; Supplementary Table S4 at DOI: 10.1002/psp4.40) on the levels of 15 phosphoproteins. As shown in **Figure P5-2**, (**a**) HEP3B, (**b**) HEPG2, and (**c**) HUH7 cells generally displayed a remarkable divergence regarding their response to both compounds and stimuli. In HEP3B cells, which were again found to be the most responsive, EGCG, FIS, and to a lesser extent ERI downregulated the basal levels of phosphorylated AKT1, TOR, GSK3B, P70S6K, RPS6, P38MAPK, JNK2, MP2K6, and NFKB. Furthermore, EGCG decreased both the constitutive and stimulated phosphorylation of SHP2, whereas it increased the basal and induced expression of phosphorylated CREB1 and HSPB1. In HEPG2, FIS was found to attenuate most of the phosphoproteins (SHP2, AKT1, GSK3B, TOR, P70S6K, RPS6, MP2K6, P38MAPK, JNK2, NFKB). Finally, in HUH7 only a weak inhibition in certain phosphoproteins was observed in response to FIS and ERI, whereas, unexpectedly, EGCG upregulated the levels of phospho-AKT1, GSK3B, RPS6, P38MAPK, and NFKB. Immunoblotting performed in cell lysates from EGCG-treated HCC cells for phospho-AKT1 confirmed these multiplex data (Supplementary Figure S7a at DOI: 10.1002/psp4.40). Furthermore, as the active non-phosphorylated GSK3B is a known negative regulator of beta-catenin stability[233], we predicted and experimentally confirmed that the levels of the stable active form of beta-catenin were altered by EGCG according to the phospho-AKT1/GSK3B modification pattern (Supplementary Figure S7b at DOI: 10.1002/psp4.40) in each HCC cell line, thus indicating a differential targeting of the WNT/beta-catenin signaling[234].

Integrated data analysis to decipher compound action mechanisms

Since HEP3B cells provided the broadest datasets due to their high sensitivity to imposed perturbations, we subsequently constructed compound-specific signaling networks that combined existing knowledge with HEP3B phosphoproteomic and cytokine release data employing an adapted version of our SigNetTrainer methodology and a sensitivity analysis for a range of thresholds (see Methods and Supplementary Material at DOI: 10.1002/psp4.40).

As a result, we visualized how EGCG (**Figure P5-3**), FIS (**Figure P5-4**), and ERI (**Figure P5-5**), modulated signal transduction and cytokine secretion in interrogated cells under basal conditions (**Figures P5-3a**, **4a**, and **5a**) or following subsequent exposure to stimuli (**Figures P5-3b**, **4b**, and **5b**). Our network-based analysis confirmed or rejected existing information (black and gray-colored edges, respectively) and furthermore enriched the canonical pathway with new interactions (blue-colored edges), thus predicting potential nutraceutical-targeted pathways. A major mechanism designated for EGCG and FIS is through inhibition of the PI3K/PIP3/PDK1/P70S6K/RPS6 pathway with potential involvement of AKT1/TOR signaling blockade (**Figures P5-3, 4**, solid and dashed black edges, respectively), in accordance with existing literature (Supplementary Table S5 at DOI: 10.1002/psp4.40) and current data showing attenuation of AKT1, TOR, P70S6K, and RPS6 activation by these compounds. Regarding ERI, the inhibition of the same pathway downstream of PDK1 (blue edge, **Figures P5-5a**) emerged for the first time in HEP3B cells. Most important, the fact that the blockade of this particular signaling cascade by all test compounds is well correlated with their inhibitory effects on the release of a number of cytokines (EGCG: CXCL7, CXCL16, CXCL10, IL17A, FIS: CXCL7, CXCL10, CXCL11, ERI: CXCL10) pointed to the prediction of a common cytokine downregulating mechanism through eventual inhibition of the translation factor RSP6 (blue edges, **Figures P5-3**, **4**, and **5**). This major network prediction was further validated by independent real-time quantitative reverse transcription-polymerase chain reaction (RT-PCR) and western blot analyses (Supplementary Figure S7c, d at DOI: 10.1002/psp4.40) focusing on CXCL10, as this cytokine was connected with RPS6 in the networks of all three compounds. We found that nutraceutical treatment: (a) did not reduce CXCL10 mRNAs levels, (b) did not further decrease compared to vehicle control the total CXCL10 protein levels under conditions of protein synthesis blockade, hence indicating no effect on protein degradation machinery, while (c) reduced the total CXCL10 levels under conditions of proteasome-mediated protein degradation blockade, thus indicating an effect on new protein synthesis. Overall, these results supported the conclusion that the observed reduction in CXCL10 protein levels is not attributed to either a decrease in gene transcription or protein degradation but is rather causally linked to a

translational downregulation of cytokine expression via RPS6 inhibition, as predicted by our network analysis.

On the other hand, only signaling alterations specific for the examined cancer cell type were maintained in the network. For instance, the PAK/RAF1/MEK1/ERK1 pathway was removed from the relevant network (**Figure P5-3b**) as not being involved in the action mode of EGCG in HEP3B cells because prior knowledge (Supplementary Table S5 at DOI: 10.1002/psp4.40) was not confirmed by our data for MEK1 and ERK1 in the case of EGCG (gray edges, **Figure P5-3a**).

Prevention of HCC-induced metastasis and angiogenesis

To assess the biological significance of changes imposed by EGCG, FIS, and ERI on HCC-secreted factors (including cytokines shown in **Figure P5-1**, Supplementary Table S4 at DOI: 10.1002/psp4.40), we next applied a chemotaxis model of metastasis, where naïve HCC cells were allowed to migrate and/or invade Matrigel-coated membranes towards HCC-CM collected from compound pre-treated cultures. As shown in **Figure P5-6 a**, **b**, HCC-CM from vehicle-treated cells (CTL) significantly induced cancer cell migration and Matrigel invasion, respectively, compared to basal starvation medium, thus confirming that factors released from HCC cells act as strong stimulators of the metastatic process. However, this effect was reversed in HCC-CM collected from cells pre-exposed to compounds. Likewise, we examined the responsiveness of HUVEC to HCC-secreted signaling mediators. We found that HCC-CM from vehicle-treated cells (CTL) strongly induced HUVEC motility (**Figure P5-6c**), invasiveness (**Figure P5-6d**) and differentiation into tube-like structures (**Figure P5-6e**), compared to basal starvation medium. However, these proangiogenic attributes were significantly abrogated when HUVEC were challenged with HCC-CM from compound pre-treated cultures. Overall, the cytokine-reducing activity of EGCG, FIS, and ERI (**Figure P5-1**, Supplementary Table S4 at DOI: 10.1002/psp4.40) correlated with their capability to inhibit tumor-driven prometastatic and proangiogenic phenotypes (**Figure P5-6**).

Discussion

Plant-derived compounds have always been an important source for the development of new drugs in pharmaceutical research[235]. However, their clinical use is limited by the fact that, in the majority, these compounds do not have a known set of target proteins but they act on many different levels, thus making the identification of their detailed action mechanism very difficult if at all possible. In this work by using the Luminex xMAP system as a drug-discovery platform, we screened several anti-inflammatory nutraceuticals for their ability to modulate the cytokine release and phosphoproteomic response of HCC cells, constitutively or upon perturbation with signaling stimulators that are actively involved in human hepatocarcinogenesis[236]. Then, by applying an integrated multi-combinatorial analysis of the high-throughput proteomic data we provided novel mechanistic insights into the mode of action of most promising compounds.

Like the majority of neoplastic cells, HCC secrete an array of inflammatory cytokines that exert their multiple, partially overlapping functions (mitogenic, motogenic, angiogenic) in a para-/autocrine manner via binding to their appropriate receptors[218]. Our results demonstrated EGCG as the most effective modulator of inflammatory cytokine secretion (followed by FIS and ERI) and HEP3B cells as the best responders. Despite previous extensive literature[214,215], this is the first study to our knowledge showing the outstanding capability of this compound to concurrently reduce a wide range of HCC-secreted cytokines, including a group of C-X-C motif ELR-positive (CXCL7, CXL16, GROA, IL8) as well as ELR-negative (CXL10, CXCL11) chemokines and interleukins (IL4 and IL17A), all known to be critical modulators of tumor microenvironment. In fact, CXCL7, GROA, and IL8, identified as CXCR2 ligands, have been shown to strongly promote tumor growth and angiogenesis[237], whereas soluble CXCL16 acting through the CXCL16/CXCR6 axis has been proved to induce cancer invasion and metastasis[238]. Overexpression of CXL10 by malignant cells has been reported to desensitize CXCR3 in lymphocytes from HCC patients resulting in tumor escaping from host defense mechanisms[239], while activation of the CXCL11/CXCR7 pathway has been related to HCC progression[240]. Concerning IL4 and IL17A, they

have been connected with suppression of cancer immunosurveillance and promotion of metastasis[241–243].

At the phosphoproteomic level, xMAP measurements in treated cells, besides confirming previous knowledge, thus supporting the validity of the applied method, uncovered some new interesting molecular targets such as SHP2, CREB, and HSPB1. Particularly, SHP2 tyrosine phosphatase has emerged as a key nodal point in cytokine and growth factor-induced signaling; therefore, its inhibition by EGCG and FIS in HEP3B and HEPG2 cells is expected to drastically influence the activation state of several downstream effectors, including JAK, STAT, and PI3K[244]. In consistency, EGCG and FIS were found capable of inhibiting the phosphorylation of important components of PI3K axis (AKT1, GSK3B, TOR, P70S6K, RPS6), which is known to be abnormally activated in various cancers, including HCC, promoting cell survival, invasion and angiogenesis[245]. Regarding phospho-CREB and HSPB1 upregulation, although data interpretation merits further experimental investigation, the available literature supports that phosphorylated CREB can competitively inhibit NFKB activation, thereby restricting proinflammatory responses[246], whereas phosphorylated HSPB1 can mediate growth suppression in human HCC[247,248].

At the computational front, we adapted our previously described *SigNetTrainer* formulation to construct compound-specific signaling networks in HEP3B cells that linked intracellular activity (phosphosignaling) to cellular function (cytokine release), thus revealing mechanisms of compound actions on the basis of topology alterations of key signaling pathways. In contrast to previous mechanistic studies using kinase assays, the proposed methodology is able to capture not only the compound effects on specific target kinases but also how the signaling machinery in its entirety is affected by this compound. A novel outcome of our computational analysis, which we further validated experimentally, is the prediction of a common negative control mechanism underlying the downregulation of several of the aforementioned cytokines, through ultimate inhibition of the translational factor RPS6 via PI3K/PIP3/PDK1/P70S6K and/or AKT1/TOR signaling obstruction. Since aberrant mRNA translation plays a pivotal role in cancer progression, modulating the activation status of essential components of the protein synthetic machinery, such as RPS6, is expected to have a more general impact on cancer homeostasis[249]. Because the

construction of networks was based on data restricted in the neighborhood of five receptors and 15 measured phosphoproteins, some potential nutraceutical targets with importance for cancer growth, such as, for example, the AMP-dependent protein kinase[250], were omitted. Despite this limitation, our proposed systems methodology is the first effort to obtain, on a systems scale, mechanistic cues into the mode of action of HCC inhibitory nutraceuticals.

At the phenotypic level, changes in HCC-secreted factors mediated by EGCG, FIS, and HES (including the above-mentioned reduction of crucial prometastatic and proangiogenic cytokines), were found eventually to be essential, as they restricted *in vitro* the capability of treated cancer cells to transmit prometastatic and proangiogenic signals to other cancer cells, as well as to their context tumor vasculature, respectively. These results further support the potential effectiveness of these compounds to prevent HCC expansion through restriction of tumor neovascularization and metastasis[251].

In conclusion, in this study by combining high-throughput protein profiling and network-based analysis of inflammatory and phosphoproteomic HCC responses, we were able to provide important preclinical evidence and molecular insight for use of most promising nutraceuticals in novel chemopreventive and/or therapeutic interventions for HCC.

Drug-Induced Cardiomyopathy

The disease

The clinical manifestations of anti-cancer drug associated cardiac side effects are diverse and can range from acutely induced cardiac arrhythmias to Q–T interval prolongation, changes in coronary vasomotion with consecutive myocardial ischemia, myocarditis, pericarditis, severe contractile dysfunction, and potentially fatal heart failure.[1]

Experimental Model

Primary study (Cardiomyocytes)

- Human Induced Pluripotent Stem Cell-Derived Cardiomyocytes
- Human Embryonic Stem Cell-Derived Cardiomyocytes
- Primary Human Cardiomyocytes, isolated from the ventricles of the adult heart

Verification study (Cancer cell lines from CMAP)

- MCF7
- HL60
- PC3

Experimental Tools

This is a computational project where no experimental tools were used.

Computational Tools

- Integer Linear Programming formulation to combine gene expression data at hand with prior knowledge of proteins' connectivity to construct specific

signaling pathways

- Elastic Net Regularization and other methods of predictive modelling and machine learning
- Text mining

Main Findings

Constructing specific signaling pathways can computationally capture a drug's mode of action and increase prediction accuracy from 79% to 88%, compared to just using the transcriptomic data at hand. This is probably because of prior knowledge of biological protein interactions and drug targets are taken also into account to construct the drug-specific network.

Using EN regularization, we were able to extract 33 protein/gene predictors that best predict the toxicity classification of drug-induced cardiotoxicity (either toxic for >0.1% clinical incidence or nontoxic for <0.1%).

The microRNAs that reportedly regulate expression of our six top predictors are of diagnostic value for natural heart failure or doxorubicin-induced cardiomyopathy. Among them, miR193-3p and miR26b-5p reportedly regulated more predictors than other microRNAs, and regulated four and three of our top predictors, respectively. It might be worthy of clinical studies to determine whether miR193-3p and miR26b-5p are useful in vivo biomarkers for drug-induced cardiomyopathy.

Publication Timeline

[P6] 17 January 2018, CPT: Pharmacometrics & Systems Pharmacology
Translational systems pharmacology-based predictive assessment of drug-induced cardiomyopathy

Co-authored Research: Translational Systems Pharmacology-Based Predictive Assessment of Drug-Induced Cardiomyopathy

Please visit DOI of paper [P6] for more information on this chapter.

Abstract

Drug-induced cardiomyopathy contributes to drug attrition. We compared two pipelines of predictive modeling: (1) applying elastic net (EN) to differentially expressed genes (DEGs) of drugs; (2) applying integer linear programming (ILP) to construct each drug's signaling pathway starting from its targets to downstream proteins, to transcription factors, and to its DEGs in human cardiomyocytes, and then subjecting the genes/proteins in the drugs' signaling networks to EN regression. We classified 31 drugs with availability of DEGs into 13 toxic and 18 nontoxic drugs based on a clinical cardiomyopathy incidence cutoff of 0.1%. The ILP-augmented modeling increased prediction accuracy from 79% to 88% (sensitivity: 88%; specificity: 89%) under leave-one-out cross validation. The ILP-constructed signaling networks of drugs were better predictors than DEGs. Per literature, the microRNAs that reportedly regulate expression of our six top predictors are of diagnostic value for natural heart failure or doxorubicin-induced cardiomyopathy. This translational predictive modeling might uncover potential biomarkers.

Introduction

Serious and life-threatening drug-induced adverse events cause drug attrition at various stages of drug development or modification of treatment regimens. For instance, anthracyclines, although effective to treat cancers, are known to cause irreversible, dose-dependent cardiotoxicity (contractility-related toxicity).[2] Most recently, targeted therapy with tyrosine kinase inhibitors (TKIs) also cause such toxicity.[2] The ability to predict drug-induced cardiotoxicity may reduce drug attrition

and advance precision medicine.

Predictive modeling of adverse drug reactions by integrating information across databases and knowledgebase of biological activities, chemistry, and adverse drug reactions has been undertaken.[3–5] However, no predictive models of drug-induced cardiomyopathy utilizing signaling network information have been constructed. Harpaz et al.[5] stressed the importance of harnessing multiple sources of knowledge, biological information, and biomedical literature for predicting drug toxicity. In line with this notion, we reported herein predictive modeling by integrating prior knowledge, drug targets, and empirical data in order to enable the model to identify key predictors from a drug's mode of action, and to have the potential to inform lead identification and development.

To fill in the gap, we compiled a list of 31 toxic and nontoxic drugs that were transcriptomically profiled in human cardiomyocytes[6–8]; manually curated and compiled their clinical incidence of treatment-related cardiomyopathy; and conducted predictive modeling of drug-induced cardiomyopathy. Two predictive models were compared: (1) applying elastic net (EN) to gene expression data; and (2) applying integer linear programming (ILP) to construct a drug's signaling network to reflect its mechanism of action,[9] and then subjecting the nodes in individual drugs' signaling networks to EN regression. The ILP formulation[9] navigates a prior knowledge network of protein-protein, protein-transcription factor (TF), and TF-gene interactions, and identifies the pathways that connect a drug's targets to its differentially expressed genes (DEGs). The ILP not only optimizes the solution of finding a drug's signaling pathways but also enhances performance of predictive modeling by enabling identification of the subset of DEGs that are functionally relevant to a drug's mode of action. We further referenced literature for the microRNAs, which are reportedly of diagnostic value for heart failure and for drug-induced cardiomyopathy, as well as also regulating the expression of our predictors in hopes of shedding light on potential microRNAs as in vivo drug-induced cardiomyopathy biomarkers.

Methods

Compilation of drugs and their clinical incidence of drug-induced cardiomyopathy.

To compile the list of approved drugs that cause treatment-related cardiomyopathy, we referenced the National Institutes of Health Common Terminology Criteria for Adverse Events (version 4.03)[10] and the Medical Dictionary for Regulatory Activities[11] for cardiomyopathy-related terms to text-mine approved drug labels. The terms used included cardiomyopathy, heart failure, congestive heart failure, cardiac failure, left ventricular dysfunction, left ventricular failure, and reduction in left ventricular ejection fraction. The current drug label PDF files (Drugs@FDA[2]) were processed using a text-mining analysis pipeline, as published previously.[12] Individual rates of occurrence for cardiomyopathy were extracted by manual curation of drug labels, published redacted new drug application reviews (Drugs@FDA), as well as published clinical studies.

Predictive modeling

Workflow and highlights of EN and ILP.

As shown in **Figure P6-1**, we compared two pipelines of predictive modeling. For pipeline 1, we applied EN to DEGs of a drug. For pipeline 2, we applied ILP to construct each drug's signaling pathway, and then subjected the genes/proteins in each drug's signaling network to EN regression.

The EN is useful for predictive modeling when predictors greatly outnumber observations while simultaneously being able to identify statistically significant predictors.[13] The EN regularization is useful for analyzing genomics of drug sensitivity in cancer.[14]

We applied ILP to a drug's DEGs and protein targets to model its mode of action. These two levels of information are connected via signal transduction where the signal originates at drug targets, propagates intracellularly via a complex network of signaling cascades, passes through the layer of TFs, and finally reaches the transcriptomic level of DEGs. We modelled the interactions in the knowledge network by using the logic formalism,[15] which identified the minimum subset of the network to

achieve the desired connectivity. We constructed the specific signaling network for each drug using an ILP formulation, as published previously.[9]

The ILP will enhance predictive performance because it has the ability to capture cellular responses to a drug, to identify the subset of important functional DEGs, and to help differentiate between compounds and translate into improved performance.

Drug name normalization.

Drug names were first normalized and identified by the PubChem compound identifier to ensure consistency when downloading data from Connectivity Map (CMap),[16] Drug Toxicity Signature Generation Center (DToxS),[6] Search Tool for Interactions of Chemicals (STITCH),[17] and literature.

Compilation of drug targets.

We compiled the targets of individual drugs from STITCH,[17] and the "chemicalprotein links" database and selected only human proteins. The proteins were identified by the SwissProt/EnsEMBLidentifier, and translated into HUGO Gene Nomenclature Committee gene symbols,[18] using the R biomaRt package, in order to match with the nodes in the prior-knowledge network.[19] We used STITCH's "interaction types for links" data file, from where we identified the drugs as activating or inhibiting individual target proteins. We used only those associating links between protein-drug pairs with an evidence score of 0.7.

Gene expression data sources and handling.

Wherever data were available in Affymetrix probe IDs, the probe IDs (Affymetrix GeneChip Human Genome U133A Array) were translated into HUGO Gene Nomenclature Committee gene symbols using the biomaRt package[20] and hgu133a2[21] packages in R, an open source statistical computing graphics systems. Across all the gene lists, we kept only those genes with fold change > 2 and P value < 0.05 by a two-tailed, two-sample, unequal variance Student's t-test, adjusted separately for the up and down gene lists with Bonferroni correction (P value adjusted for multiple comparisons).

A list of 75 drugs with drug-induced DEGs available from cancer cells[16] in CMap were

used for exploratory modeling (see **Table P6-S1**). To conduct robust predictive modeling, we exhausted literature and databases and found a list of 31 drugs of which drug-induced DEGs in human cardiomyocytes and stem cells-derived cardiomyocytes were available. The two data sources for drug-induced perturbation of gene expression in cardiomyocytes were: (1) 30 drugs from DToxS, where primary human adult cardiomyocytes were used; and (2) literature data of doxorubicin studied in human stem cell-derived cardiomyocytes. The size of each dataset was mainly constrained by the availability of DEGs data. For DToxS data, we downloaded the level two gene expression data, calculated the fold changes, kept only those DEGs with achange > 2 and merged them from different donors by averaging the fold changes while excluding any DEGs with opposite directions of fold change among donors.

Doxorubicin is widely studied for its dose-dependent cardiac toxicity, and is commonly dosed at 40–60 mg/m. Following intravenous 60 mg/m^2, its peak plasma concentration (C_{max}) was 630 ng/mL (1,159 nM).[22] See **Table P6-S2** for a few studies of transcriptomic profiles of doxorubicin. For our modeling, we included the data from human-induced pluripotent stem cells-derived cardiomyocytes by Chaudhari et al.[8] and Burridge et al.[7] We included the gene expression data by Burridge et al. were at 100, 1,000, and 10,000 nM and those by Chaudhari et al. were at 156 nM (see Supplementary Table S3 at DOI: 10.1002/psp4.12272 for the rationale).

Identifying a drug's mode of action using ILP.

We first built a prior-knowledge network as a scaffold for constructing a drug's signaling network by downloading from Reactome the latest version (version 2015) of the "Functional interactions derived from Reactome." As published previously,[9] we merged those interactions with transcription factors and obtained a network across the protein, transcription factor, and gene levels, which contained 64,801 reactions, 2,585 signaling proteins, and 12,376 genes. We applied ILP to optimize a drug's signaling network by providing as the input the scaffold mentioned above and its targets.

The ILP formulation was solved using IBM ILOG CPLEX optimization studio for the objective of optimizing a drug's network. Based on the constraints that mimic signal

transduction[23] and adjustment to the specific case of very large (>10,000 nodes) networks, the algorithm minimized the mismatch between the data of gene expression measurements and the prior knowledge pathway topology. The output was the optimal signaling network of a drug, identifying the molecular interactions that seemed to be functional based on the input of DEGs and drug targets. We were able to select the minimum part of a prior-knowledge network for each drug that could explain the data in hand. See Supplementary Document-ILP at DOI: 10.1002/psp4.12272 for understanding the example of methotrexate signaling network captured by ILP (also in **Figure P6-S1**) and how the proposed ILP formulation works.

Comparing predictive modeling by applying EN to a drug's DEGs vs. to its ILP signaling network.

To construct a matrix for EN regression, a drug was marked with 0 if classified as nontoxic and marked 1 if classified as toxic. We classified drugs by referencing approved labels for the criteria of "frequent adverse events being those occurring on one or more occasions in at least 1/100 patients; infrequent adverse events being those occurring in 1/100 to 1/1,000 patients; rare events being those occurring in less than 1/1,000 patients." Referencing the definition of rare events used in drug labelling and considering the distribution of clinical incidence, the number of drugs with gene expression data available, and the heterogeneity of clinical studies, we classified drugs into two classes, toxic for those with incidence >0.1% and nontoxic for those with incidence <0.1%.

A column of "cardiotoxicity" was created with the clinical incidence score: 1 for "toxic" and 0 for "nontoxic." Each column corresponded to a single gene expressed in at least one of the DEGs signatures. Individual DEGs of a drug were assigned a value of 1, -1, or 0 to reflect upregulated, downregulated, or not reported, respectively (pipeline 1). The same assignments were applied to the nodes in each drug's ILP signaling network (pipeline 2).

In our modeling, we used EN regression, and more specifically a linear regression model with an EN penalty determined using the R package glmnet.[24] The EN regularization is defined by two parameters, alpha and lambda. The EN regression is

a mixing of LASSO and ridge regression and combines their two penalty terms for the alpha parameter. When alpha equals 0, EN performs as ridge regression and when alpha equals 1, EN performs as LASSO. In EN, the lambda parameter reflects shrinkage of the model's coefficients. When lambda equals 0, no shrinkage of the model's coefficients is performed but the coefficients decrease toward 0 (although not exactly equal 0) as its value increases. We tried a range of values for alpha from 0 to 1 by a 0.01 step and selected the one that minimized the mean squared error. For that alpha value, we selected the value of lambda that gave the minimum mean cross-validated error.

To validate each model, we used leave-one-out cross validation (LOOCV) by leaving a drug's signature out one at a time (either DEGs or signaling network constructed from ILP) and did so across the whole list of drugs. Each time we calculated the accuracy, sensitivity, and specificity for a predictive model, and selected and reported the model with the highest accuracy along with its precision, sensitivity, and specificity. From the chosen predictive model, we extracted the predictors (genes/proteins) that best predicted drug-induced cardiotoxicity. The receiver operating characteristic (ROC) and precision-recall curves using the R package with the former plotted in smooth curve.

Pipeline 1 – applying EN to DEGs

The results of 75 drugs with DEGs from CMap are summarized in Supplementary Document-CMap at DOI: 10.1002/psp4.12272. Among these 75 drugs with their DEGs from CMap, 24 drugs were toxic and the remaining 51 drugs were nontoxic.

A model matrix was constructed using cardiomyocyte data, with the 34 observations (toxicity classification) as rows and 15,016 variables (gene expression) as columns. The predictive linear model was constructed by having as input all these variables for EN regularization. We tried all possible different cutoff scenarios (see the spreadsheet "summary" of Supplementary Table S4 at DOI: 10.1002/psp4.12272 for the results of the psccm_34_gen_heart trials and the detailed results of 18 models with different cutoffs in the spreadsheet "9"). For example, a cutoff of 10 meant that we ran the model by using only those genes that were expressed in at least 10 of the 34

signatures, meaning that the analysis started with 3,508 genes, whereas a cutoff of 15 started the analysis with the genes that appeared in at least 15 of the 34 signatures, meaning 464 genes were used as the cutoff.

Pipeline 2 – applying EN to gene/protein nodes in ILP-constructed signaling networks.

We first performed exploratory modeling using a list of 75 drugs with gene expression data available in CMap and concluded that signaling networks of drugs derived from ILP outperformed their DEGs when applying EN regularization (see **Figure P6-S2** for ROC and precision-recall curves).

We were able to find the ILP solutions for drugs with gene expression data in cardiomyocytes (Supplementary Table S5 at DOI: 10.1002/psp4.12272) except cefuroxime, domperidone, and olmesartan. These three drugs were removed from this modeling exercise. At the end, we had 31 signaling networks from 28 drugs (15 nontoxic drugs and 13 toxic drugs). See Supplementary Table S5 at DOI: 10.1002/psp4.12272 for the gene/protein nodes in the signaling network of each individual drugs. We built a model matrix for the 31 signaling pathways/networks by using gene expression profiles from cardiomyocytes and by assigning 1 if a pathway node was upregulated, -1 if it was downregulated, and 0 if it was not present in a drug's optimized signaling network. See the spreadsheet "summary" of Supplementary Table S4 at DOI: 10.1002/psp4.12272 for the results of the psccm_34_ILP_heart trial and the detailed results of 31 models with different cut-offs in the spreadsheet "10."

Biological context of predictors

To gain translational insight, we searched literature for microRNAs that have been shown to be diagnostic markers of heart failure and also involved in regulation of gene expression. We mined literature and miRTarBase,[25] a database of experimentally validated microRNA-target interactions, for a list of microRNAs, which have been individually reported to regulate expression of our top gene/protein predictors, and

also been reportedly detected in the circulation of patients with heart failure with a varying degree of severity[26,27] or of patients with doxorubicin-induced cardiomyopathy.[28]

Results

The list of drugs and toxicity profile.

The list of 31 drugs with their clinical profiles of treatment-related cardiomyopathy is summarized in **Table P6-1**. Literature search was also conducted to supplement clinical incidence of cardiomyopathy, if approved drug labels and published application reviews[2] did not have such information. Among the 31 drugs, there were 13 toxic drugs (41.9%) and there were 18 nontoxic drugs (59.1%). For those drugs without mention of cardiomyopathy-related toxicity described in their labels throughout the sections of clinical studies, post-marketing experiences, and warnings and precautions, we also searched literature and published reviews[2] to reach the conclusion that they are nontoxic drugs.

Predictive modeling

Applying EN to DEGs (pipeline 1).

Using LOOCV across the whole list of 30 drugs and their gene expression signatures, we achieved 79% accuracy and 75% precision, with 80% sensitivity and 79% specificity when using those genes that were expressed in at least 11 of the 34 signatures (a cutoff of 11 in spreadsheet "9" of Supplementary Table S4 at DOI: 10.1002/psp4.12272). The results of EN regularization are shown in **Figure P6-2**a, **2**c, and the genes/proteins with non-zero coefficients are PHF19, HSPA8, RIF1, CD46, MXRA7, RAB27A, TOMM20, MYO6, and CCNA2. The ROC curves and precision-recall curves are shown in **Figure P6-3** and **Figure P6-S2,** respectively.

Applying EN to the gene/protein nodes in ILP-constructed signaling networks (pipeline 2).

By applying EN regression and LOOCV, we were able to increase both prediction accuracy and precision to 88%, with 88% sensitivity, and 89% specificity, compared with the results from EN regression of DEGs (Supplementary Table S4 at DOI: 10.1002/psp4.12272). The EN regularization is shown in **Figure P6-2**b, **2**d. The result for the psccm_34_ILP_heart trial is in the spreadsheet "summary" and the detailed results of 31 models with a cutoff ranging from 1 (5,012 genes/proteins in at least 1 drug) to 31 (5 genes/proteins in at least 31 network signatures) are in spreadsheet "10" of Supplementary Table S4 at DOI: 10.1002/psp4.12272. The highest accuracy, sensitivity, and specificity were achieved at cutoff of 10 with 189 genes/proteins from at least 10 drugs' signaling networks. The ROC and precision-recall curves are shown in **Figure P6-3** and **Figure P6-S2**, respectively.

We concluded that EN-ILP (pipeline 2) outperformed EN alone (pipeline 1) when applied to the same set of DEGs.

Cardiac context of top predictors

Using EN regularization, we were able to extract the protein/gene predictors that best predict the toxicity classification of drug-induced cardiotoxicity (either toxic for >0.1% clinical incidence or nontoxic for <0.1%). The 33 protein/ gene predictors along with their individual coefficients are summarized in **Table P6-2**. The network of the top 15 genes/ proteins selected by the model is presented in **Figure P6-4**. Cardiac relevance of these predictors was reviewed and summarized in Supplementary Table S6 at DOI: 10.1002/psp4.12272. The protein and gene predictors identified by EN-ILP reflected the key cellular biological factors for drug-induced cardiotoxicity. The EN regularization in our predictive modeling selected the protein/gene predictors that best predicted drug-induced cardiotoxicity.

We mined an evidence-based database of microRNAs[25] for those that reportedly regulate our top predictors, and also referenced literature to narrow the list to those that are reportedly of diagnostic value for heart failure. Summarized in **Table P6-3** are our top 10 predictors and their individual regulating microRNAs that have reportedly been of diagnostic value for natural heart failure[26,27] or for doxorubicin-induced cardiomyopathy.[28]

Discussion

With the clinical incidence of drug-induced cardiomyopathy as a dependent variable, ILP-enhanced predictive modeling increased prediction accuracy from 79% to 88%, compared to modeling with EN and DEGs alone. This improved prediction signified the ability of ILP to computationally capture a drug's mode of action through constructing its signaling pathways for the purpose of predictive modeling. ILP offers the advantage of integrating our prior knowledge of biological protein interactions and drug targets (Reactome and STITCH), transcription factors, and DEGs into predictive modeling. ILP also optimizes the size of a drug's network signature in addition to capturing the signaling pathways of a drug. Take Lapatinib as an example, it had 2,265 DEGs from cardiomyocytes, whereas from this set of DEGs, its ILP network consisted of 1,923 nodes, including its targets, proteins involved in its signaling transduction, transcription factors, and functional DEGs.

The 33 gene/protein predictors along with their individual positive or negative coefficients could be used to predict "toxic" or "nontoxic" for a drug by linear summation using their individual levels of expression (either upregulation (1) or downregulation (-1)) from its ILP-constructed signaling network. The predictive power of this system's pharmacology predictive model will increase with the amount of data in the training set.

Among the 31 drugs used to conduct predictive modeling, the distribution of toxic (n513) vs. nontoxic (n518) classification was acceptable, although not ideal. Among them, there were 18 kinase inhibitors (17 TKIs and 1 serine/ threonine kinase inhibitor), which might seemingly be off-balance from the perspective of the diversity of drug class. Vemurafenib is a serine/threonine kinase inhibitor and not toxic. The distribution of toxic (n = 8) and nontoxic (n = 9) drugs among the 17 TKIs was acceptable. TKIs, in general, lack target specificity, have multiple targets, and were designed to disrupt the signaling pathways that are vital to cancer cell survival.[29] Unfortunately, several of these signaling pathways also play a critical role in cardiomyocyte biology[30]; consequently, several TKIs impair cardiac function. Within this context, our predictive modeling could be useful for predicting cardiac toxicity

for future new chemical entities.

All top 15 gene/protein predictors have relevant cardiac functions except ZNF 823 (Supplementary Table S5 at DOI: 10.1002/psp4.12272). Interestingly, CYP3A4 was an important predictor. Although CYP3A4 does not have biological interactions with other predictors, as shown in **Figure P6-4**, it is a major drug metabolizing enzyme. Among the 31 drugs, 10 of 13 (85%) toxic drugs and 11 of 18 (61%) nontoxic drugs were metabolized by CYP3A4. The toxic drugs that are primarily or extensively metabolized by CYP3A4, included amiodarone, axitinib, cytrabine, dasatinib, doxorubicin, imatinib, ponatinib, sorafenib, sunitinib, and vandetanib.[2,31–33] For nontoxic drugs, they are bosutinib, crizotinib, cyclosporine, domperidone, erlotinib, gefitinib, lapatinib, regorafenib, ruxolitinib, tofacitinib, and ursodeoxycholic acid.[2,34]

Some top predictors are biologically associated with focal adhesion kinase (FAK), a nonreceptor protein-tyrosine kinase, which is involved in the cytoskeleton-associated network of signaling proteins.[35] Focal adhesion complexes play a critical role in how cultured cardiomyocytes respond to mechanical and neurohormonal stimuli, and in the development of heart failure.[36] FAK activation plays a role in the adaptive response to cardiac afterload and in myocyte growth via the protein kinase B/mammalian target of rapamycin pathway.[37] The FAK cleavage is mediated by CASP3 family during apoptosis of human normal cells,[38] and occurs with activation of EPHA2 and p38 mitogen-activated protein kinase during doxazosin-induced apoptosis of a cardiac cell line.[39] FAK activates STAT1 during cell attachment,[40] and plays a role in cell migration with one of its actions being associated with platelet-derived growth factor receptor (PDGFR) signaling complex.[41] In short, the top predictors are important to maintain normal cardiac function.

Per literature, some microRNAs that reportedly regulated expression of our predictors have also been shown to be of diagnostic value for heart failure with a varying degree of severity (**Table P6-3**).[26,27] Among them, miR193-3p and miR26b-5p reportedly regulated more predictors than other microRNAs, and regulated four and three of our top predictors, respectively. It might be worthy of clinical studies to determine whether miR193-3p and miR26b-5p are useful in vivo biomarkers for drug-induced cardiomyopathy. Literature search uncovered a recent study that investigated circulating microRNAs in children with anthracycline-induced acute heart injury.[28]

Elevated miR-29b and miR-499 in the circulation seemed to correlate with troponin elevation in these children, and were identified as potential cardiomyopathy biomarkers.[28] This observation of miR-29b elevation in doxorubicin-induced cardiomyopathy differed from an observation of decreased expression of miR-29b-3p in the coronary sinus blood of patients with heart failure.[27] The MiR-29b-3p regulates expression of one of our top 10 predictors, PDGFR-A. Further studies are needed to investigate the role of miR-29b in drug-induced cardiomyopathy or in natural heart failure. Even though miR-27b reportedly regulated CYP3A4,[25,42] literature search did not uncover any reports that suggested miR-27b to be of diagnostic value for drug-induced cardiomyopathy.

Integrating clinical incidence with the modes of action of a drug, which is depicted as its signaling network, for predictive modeling is a strength of our study. There are, however, some limitations in our approach: (1) nontoxic slightly outnumbered toxic drugs; (2) limitation of ILP where no biological feedback controls were considered and assumptions adopted in ILP formulation; (3) DEGs of doxorubicin in cardiomyocytes were from different sources than the rest of 30 drugs; and (4) availability of transcriptomic profiling data in cardiomyocytes. Furthermore, our study inherited the shortcomings associated with the databases and knowledge base used for our modeling. The impact of disease indications on the incidence and severity of treatment-related cardiomyopathy is not well characterized.

Our predictive modeling of integrating clinical incidence of drug-induced cardiomyopathy with the signaling network of toxic and nontoxic drugs not only is useful for further improving its predictive power, but also identifies important gene/protein predictors that have relevant cardiac biological functions. Above all, the top genes/protein predictors are reportedly regulated by specific microRNAs that have been shown to be of diagnostic value for heart failure or drug-induced cardiomyopathy. These predictors might be useful for shedding light on potential microRNAs as in vivo biomarkers of drug-induced cardiomyopathy.

Non-Alcoholic Fatty Liver Disease

The disease

Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, estimated to be by 2030 the most frequent indication for liver transplantation. NAFLD is a multifactorial disease and its cause and progression mechanisms are still not completely understood.

Experimental Model

- Primary Human Hepatocytes

Experimental Tools

- Collection of liver from patients
- Isolation of Primary Human Hepatocytes and cell culture
- Stimulation, lysis and collection of lysates
- Measurement with bead-based sandwich ELISA for phosphoproteins
- High content screening: Nile Red stain with Hoechst 33342 as a counterstain

Computational Tools

- Integer Linear Programming formulation to combine proteomic data at hand with prior knowledge of proteins' connectivity to construct specific signaling pathways
- Enrichment analysis

Main Findings

Since 2014, researchers have suggested in-depth study of signal transductions in order to provide novel solutions in curing or preventing NAFLD.[252] Today, NAFLD has become a threat to public health mainly because of the obesity epidemic and it is well understood that it's a multifactorial disease. To our knowledge, this is the first study that aims to understand the multifactorial nature of NAFLD at the signaling level by studying 5 NAFLD induction models in primary human hepatocytes.

Our results confirm a large body of literature findings for NAFLD signaling mechanisms. Furthermore, CHK2 and EPOR have emerged as potential NAFLD players that may be interesting to study further since they are important factors in liver regeneration[253].

The observed multifactorial nature of the disease suggests that there is no single treatment for all subtypes of NAFLD, highlighting the need for a systemic approach and personalised therapeutic interventions to better understand and treat NAFLD.

Publication Timeline

[A11] 3-5 March 2017, Athens, Greece
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.

[A12] 7 April 2017, Athens, Greece
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.

[A13] 7-8 November 2017, Amsterdam, Netherlands
Development of a new drug repositioning platform for Non-Alcoholic Fatty Liver disease through network analysis.

[A14] 9-11 November 2017, Rome, Italy
Network based drug repositioning for Non-Alcoholic Fatty Liver Disease.

[Submitted Manuscript]

Cue-signal-response analysis reveals multifactorial signalling mechanism of NAFLD in vitro.

Co-authored Research: Cue-signal-response analysis reveals multifactorial signaling mechanism of NAFLD in vitro.

Regarding the study presented in this chapter, Informed consent to participate in the study was obtained from all participants. Ethics approval and approval of the study was obtained from Research Committee National & Kapodistrian University of Athens – Medical School – Aretaieion University Hospital. The authors declare no conflict of interest. Leonidas G Alexopoulos has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreements No 642295 and No 675585. Authors: Dimitris E Messinis, Danai-Stella Zareifi, Angeliki Minia, Vaia Pliaka, Eirini V Pantiora, Andreas A Polydorou, Antonios I Vezakis, Georgios P Fragulidis, Manoussos M Konstadoulakis, Konstantinos J Bramis, Efstathios A Antoniou, Leonidas G Alexopoulos. Authors' Contributions: Conceptualization, LGA; Methodology, DEM and DZ; Software, DEM; Formal Analysis, DEM; Investigation, DEM and DZ; Resources (tissue samples from patients and their preparation for cell culture), AM, VP, EVP, AAP, AIV, GPF, MMK, KJB, EAA; Writing - Original Draft, DEM and DZ; Visualization, DEM and DZ; Supervision, LGA; Funding Acquisition, LGA.

Abstract

Background: Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, estimated to be by 2030 the most frequent indication for liver transplantation. NAFLD is a multifactorial disease and its cause and progression mechanisms are still not completely understood. The main aim of this study is to unveil different potential signaling mechanisms of NAFLD.

Results: Amiodarone, Free Fatty Acids, Tamoxifen, Tetracycline and Valproic acid were employed as 5 NAFLD induction models in an *in vitro* platform to study the disease using primary human hepatocytes. NAFLD induction was monitored by Nile red staining. For each NAFLD-induction model, the signaling mechanism was interrogated by measuring 17 phosphorylated protein targets and pathway optimization algorithm was employed to construct signaling networks. Pathway findings were integrated with gene expression enrichment analysis to shed light into the mechanism of each *in vitro* approach to induce NAFLD. 5 NAFLD induction models were developed and showed strong lipid accumulation as evident by Nile red staining. Significant phosphoproteomic deregulations of CREB1, ERK1, MEK1, P53 and NFKB came in agreement with *in vivo* and *in vitro* literature findings whereas CHK2 and EPOR have arisen as important in signaling pathways of NAFLD and are related to hepatic regeneration.
Conclusions: In this study, we induced NAFLD on primary human hepatocytes with 5 models that correspond to clinical causes of the disease. All models revealed strong NAFLD phenotype that was originated from diverse signaling mechanisms. Our results suggest a multifactorial mechanism in NAFLD progression involving significant signaling pathway deregulations.

Introduction

Non-alcoholic fatty liver disease (NAFLD) has a prevalence of up to 30% in developed countries[254] and is considered as the hepatic manifestation of the metabolic syndrome. NAFLD is defined by the presence of hepatic steatosis in the absence of excess alcohol consumption (<21 units in men, <14 units in women, were 1 unit = 10ml) and represents a spectrum of clinical conditions, ranging from simple steatosis and Non-Alcoholic Steatohepatitis (NASH) to fibrosis, cirrhosis and hepatocellular carcinoma[255]. The cause and disease progression mechanisms of NAFLD are still not completely understood[256],[257]. NAFLD can be induced by lifestyles such as a high-fat diet but also by drugs and there is a great interest for reliable *in vitro* models to capture all clinical phenotypes of NAFLD.

Several *in vitro* models have been employed to study NAFLD and explore its steatotic mechanisms. Free Fatty Acids (FFA) has been used extensively. On this front, primary human hepatocytes are exposed to elevated concentrations of FFA that results in increased intracellular accumulation of lipid droplets similar to those observed in patients with NAFLD and NASH[258]. Induction of reversible intrahepatic lipid accumulation has been also connected with drugs from a range of therapeutic categories, including antiarrhythmics such as Amiodarone (AMI), antibiotics such as Tetracycline (TET), estrogen receptor modulators like Tamoxifen (TMX), and also antiepileptics such as Valproic Acid (VPA).[259] Regarding VPA, it is the most prescribed antiepileptic drug worldwide and it is the chemical analogue of valeric acid. VPA causes NAFLD through weight gain, hyperinsulinemia and insulin resistance.[260] It has been suggested that VPA causes microvesicular steatosis through VPA-CoA inhibition of hepatic carnitine palmitoyl-transferase I, an important enzyme in mitochondrial fatty acid beta-oxidation. Thus, this mechanism, supporting the evidence of weight gain frequently observed in patients under VPA therapy, has been described as potentially pivotal in drug-induced hepatotoxicity.[261] AMI is the most commonly prescribed antiarrhythmic drug.[262] Long-term Amiodarone administration is known to cause several hepatic side effects through an ambiguous mechanism, such as dose dependent effect on mitochondria.[263] It is known that mitochondrial β-oxidation and oxidative phosphorylation are inhibited by AMI which results in microvesicular steatosis, apoptosis and necrosis of the hepatocytes.[264] Interestingly, a similar nut less hepatotoxic drug, Dronedarone, has been proven to also inhibit the mitochondrial β-oxidation at the same extent,[265] suggesting that other mechanisms are involved. Another marker of AMI's steatotic effect is phospholipidosis, caused by increased triglyceride influx.[266] TET was among the first drugs found to cause dose-related microvesicular steatosis by increasing triglyceride biosynthesis and inhibiting mitochondrial fatty acid β-oxidation.[267] TMX is a selective estrogen receptor, considered as a gold standard for breast cancer therapy. TMX is known to cause macrovesicular steatosis and phospholipidosis by decreasing triglyceride secretion, impairing mitochondrial β-oxidation and progressively depleting hepatic mitochondrial DNA.[268]

A range of cellular systems have been employed to capture NAFLD clinical phenotypes *in vitro*. HepG2s have been used extensively in liver research and AMI, TET and VPA have been assessed as steatogenic compounds by flow cytometry[269]. In addition, primary rat hepatocytes and gene expression analysis[270] have been used to analyse the effect of TET, TMX and VPA. A canine primary hepatocyte culture model was also developed[271] to detect drug-induced steatosis from TET.

Signaling pathways and the phosphorylation of proteins during signal propagation are proximal to many disease-causing signaling mechanisms[197], thus phosphoproteomic measurements are becoming increasingly important in drug discovery. On this front, systems pharmacology approaches are becoming increasingly important to decipher NAFLD mechanisms through multi-omic data generation and analysis[272]. In order to construct signaling pathways from signaling data a prior knowledge network, experimental data and an optimization algorithm are needed[230]. Prior knowledge networks can be found in databases such as Reactome[19]. The ILP formulation[231] is an optimization algorithm that can analyze such a "prior knowledge" network of protein interactions along with phosphoproteomic experimental data, and identify the pathways that best describe the data at hand. This process has been shown to lead to a better understanding of several diseases' mechanisms[273] and provide useful insights relevant to a drug's mode of action.[274]

It has been suggested[252] that in-depth understanding of liver cells signaling in the NAFLD pathogenesis process will contribute to the prevention and treatment of NAFLD. In this study, we developed 5 NAFLD-induction mechanisms on primary human hepatocytes and we combined multi-omics data and computational tools to unveil different potential mechanisms of NAFLD. Our cue-signal-response analysis comprised of 5 cues for NAFLD induction, gene expression and phosphoproteomics data for signal analysis, and nile-red staining for monitoring hepatocyte steatotic response. Our platform sheds a light on the multifactorial nature of NAFLD mechanisms.

Methods

Workflow

Primary human hepatocytes were isolated from histologically normal liver fragments of adult donors undergoing liver resection with standard operating procedures and approved ethics protocols. A summary of the experimental and computational workflow is visualized on **Figure P7-1**. The isolated hepatocytes were cultured and NAFLD induction was performed via 24h treatment with 5 different steatogenic compounds: Amiodarone (AMI), Free Fatty Acids (FFA), Tetracycline (TET), Tamoxifen (TMX) or Valproic Acid (VPA). Following assessment of cell viability, verification of lipid droplet formation was assessed. To assess signaling pathway activity, NAFLD-induced cells were treated with a cocktail of activators of major signaling pathways (EGF, IL1A, IL6, LPS) and lysed for protein isolation. xMAP assay was performed to quantify phosphorylated proteins in each sample and the results were analyzed with a pathway optimization tool in order to construct signaling networks and delve into the mechanism of each in *vitro* NAFLD-induction model. This work was followed by enrichment of the pathways and comparison analysis with gene expression data from literature. The detailed steps of the process is listed below.

Primary Human Hepatocytes Isolation

Primary human hepatocytes were isolated by collagenase perfusion of histologically normal liver fragments using standard operating procedures[275] from adult donors undergoing resection for liver metastasis from colorectal cancer, and approved ethics protocols. Informed consents were obtained from all subjects (2nd Department of Surgery “Aretaieio” Hospital, 1st Department of Propaedeutic Surgery, Hippokration General Hospital and 2nd Department of Propaedeutic Surgery, Laiko Hospital, School of Medicine, National and Kapodistrian University of Athens, Greece).

Cell Culture and NAFLD induction

Primary human hepatocytes were thawed and seeded in 96-well plates (Corning® Costar®, 3599) coated with Rat Tail Collagen I (BD Biosciences) at a plating concentration of 50,000 cells per well in cell culture medium (Biopredic International).

All compounds were prepared and diluted according to the manufacturer's protocol. In detail, palmitic acid, oleic acid and sodium valproate were diluted in 100% ethanol and added to cell culture medium at 1% v/v. Tamoxifen citrate (Cayman Chemicals, 11629), Amiodarone hydrochloride (Cayman Chemicals, 15213) and Tetracycline hydrochloride (Cayman Chemicals, 14328) were diluted in DMSO and added to cell culture medium at 0,1% v/v. A mixture of exogenous FFA in molar ratio 1:2 Palmitic Acid: Oleic Acid (Cayman Chemicals, 90260, 10006627 respectively) and Valproic acid Sodium Salt (Cayman Chemicals 13033) were diluted in Ethanol. Primary Human Hepatocytes were exposed for 24h to 100 – 1000 μM FFA, 1 – 10 mM Valproic Acid, 2 – 10 μM Tamoxifen, 1 – 10 μM Amiodarone, 20 – 400 μM Tetracycline and their vehicles respectively.

High Content Screening

The formation of intracellular lipid droplets was verified using high content screening; lipid droplets were stained with 2μg/ml Nile Red fluorescent probe (Thermo Fisher Scientific, N1142). Hoechst 33342 (Thermo Fisher Scientific, H3570) was used for counterstaining the cell nucleus at 5μg/ml concentration. Images were acquired by JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek), using the channels DAPI (Excitation 390/40, Emission 452/45) and RFP (Excitation 525/50, Emission 580), under 20x optical magnification.

Resazurin Viability Assay - Viability dosage curves

Resazurin is a non-fluorescent compound and is reduced by the intracellular NADH of live cells to resorufin, which is highly fluorescent[276]. Resazurin was added to cell culture medium at 10 μg/mL and incubated for 2 hours in a 37C, 5% CO2, humidified incubator. Fluorescence was measured at wavelength Ex560nm/Em590nm using Varioskan™ LUX multimode microplate reader (Thermo Scientific™). Cell viability was expressed as a percentage of treated to untreated cells. Dose-viability curves and IC_{10} calculation was performed with a 4-parameter logistic model using the

GraphPad Prism 7.0 software. IC_{10} compound concentrations (cell viability greater than 90%) were selected for further experiments.

Bicinchonic Acid Assay to measure total protein content

Micro BCA™ Protein Assay Kit (Thermo Fisher Scientific, 23235) was used to measure the protein content of the lysates in order to adjust their concentrations to the same levels across samples before phosphoproteomic measurement. BSA (Bovine Serum Albumin) was used for standard curve construction. Absorbance was measured at 562nm using the Varioskan™ LUX multimode microplate reader (Thermo Scientific™).

Phosphoproteomic sample preparation and multiplex antibody-based ELISA

In order to prepare samples for a systematic study of the cell signaling networks, based on previous research[142], a mix of stimuli EGF, IL6, IL1A and LPS was used to activate different hepatocellular pathways simultaneously. 24 hours before this stimulation the cells were treated with the five NAFLD-inducing compounds and their vehicles, DMSO and Ethanol. 25 minutes after stimulation the cells were lysed. The selection of 25minutes was consistent with numerous earlier studies done, for example with primary hepatocytes[231] or other cell lines[277]. Lysis buffer optimized for phosphoproteomic measurements (ProtAtOnce Ltd) was used along with protease/phosphatase inhibitor mix (ProtAtOnce Ltd) and Phenylmethanesulfonyl fluoride (PMSF; SIGMA, P4626). xMAP assays were performed on a Luminex FlexMAP 3D platform (Luminex, Austin, TX). 17 phosphorylated protein targets were measured in this multiplex antibody-based ELISA. The custom phosphoprotein antibody-coupled beads were developed by ProtAtOnce Ltd. (Athens, Greece). The phosphorylated proteins' UNIPROT names and their phosphorylation residues targeted by our assays were: AKT1 (S473), CHK2 (T68), CREB1 (S133), EPOR (Y426), ESR1 (S118), HSPB1 (S78/S82), IKBA (S32/S36), JUN (S63), KS6B1 (T389), MK03

(T202/Y204), MP2K1 (S217/S221), P53 (S46), PRGR (S294), PTN11 (Y542), STAT3 (Y705), STAT6 (Y641), TF65 (S536).

Construction and analysis of signaling pathways

A prior-knowledge network including all signaling pathway information was downloaded from Reactome. The top 25 targets for Amiodarone, Free Fatty Acids (Palmitic Acid & Oleic Acid), Tamoxifen and Valproic Acid were acquired from STITCH[17], a platform pooling together all interactions between proteins and small molecules and rating them based on the available evidence. Only the targets relative to organism *Homo sapiens* were selected. Then, all targets, enzymes and transporters were acquired from DrugBank, following a search in the extended STITCH database for proteins that were not in the top 25 but were reported in DrugBank. Based on the descriptions “Activation”, “Inhibition”, “Binding” or “Catalysis” reported in STITCH, and the descriptions “Agonist”, “Antagonist”, “Inhibitor”, “Substrate” or “Inducer” reported in DrugBank, each target as was characterized as “activated” or “inhibited” from the specific compound. The targets with the higher statistical significance of the stimuli used were also acquired from STITCH. Once the target lists and prior-knowledge networks were compiled, results from experimental phosphoproteomics measurements for each compound were applied to an integer linear programming (ILP) formulation, to create compound-specific signaling networks. The ILP formulation was solved using IBM ILOG CPLEX optimization studio[231].

Enrichment analysis and signature comparison visualization

In order to visualize the high-dimensional data, multidimensional scaling was performed using cmdscale from R package ‘stats’[278] to construct a 2D plot of the compounds’ mechanisms for visualization purposes.

Expression data for NAFLD patients with Gene Expression Omnibus (GEO) accession number GSE49541 were used[279]. The data were created from liver biopsy tissue RNA analysis. 40 gene expression profiles with mild NAFLD (fibrosis stage 0-1) and 32 with

advanced NAFLD (fibrosis stage 3-4) were analyzed with GEO2R to find differentially expressed genes. Gene Ontology (GO) using the Term Enrichment Service, powered by PANTHER[280] provides 8 different annotation data sets, which were all used to collect annotation terms for our analysis: PANTHER Pathways, Protein Class, GO-Slim Molecular Function (MF), GO-Slim Biological Process (BP), GO-Slim Cellular Component (CC) and the GO complete ontologies MF, BP, CC.

Results

Dose response curves and calculation of IC10

Resazurin viability assay was used for the construction of dose-viability curves and tha calculation of IC_{10} for all five compounds used to induce NAFLD (**Figure P7-S1**): 90.56 ± 1.099 μM for Amiodarone, 591.07 ± 1.389 μM for Free Fatty Acids, 150.3 ± 1.483 μM for Tetracycline, 17.71 ± 1.488 μM for Tamoxifen and 6200.8 ± 2.109 μM for Valproic Acid.

Confirmation of NAFLD induction via High Content Screening

The formation of intracellular lipid droplet formation was verified high content screening, more specifically Nile Red stain with Hoechst 33342 as a counterstain (**Figure P7-2**).

xMAP phosphoproteomic assays results

Student's t-test was performed across the four donors' sample measurements, comparing the DMSO or Ethanol vehicles with the Amiodarone, Free Fatty Acids, Tetracycline, Tamoxifen and Valproic Acid treatments. Fold increase was calculated by dividing the average across the four donors treated cells by the respective average of control.

Phosphoproteomic results are summarized on **Figure P7-3**A. Amiodarone significantly increased (1.27-fold, p-value 0.029) phosphorylated Erythropoietin

receptor (EPOR). The Free Fatty Acids stimulation significantly increased (1.32-fold, p-value 0.012) the phosphorylated Transcription factor p65 (TF65), while Tamoxifen-stimulated cells were observed to have increased phosphorylated cellular tumor antigen p53 (P53) at 1.47-fold, p-value 0.007, Serine/threonine-protein kinase Chk2 (CHK2) at 1.27-fold, p-value 0.027 and cyclic AMP-responsive element-binding protein 1 (CREB1) at 1.37-fold, p-value 0.005. Finally, Valproic Acid treatment significantly increased phosphorylation of TF65 at 1.41-fold, p-value 0.016, Dual specificity mitogen-activated protein kinase kinase 1 (MP2K1) at 1.42-fold, p-value 0.015 and Mitogen-activated protein kinase 3 (MK03) at 1.64-fold, p-value 0.010.

Signaling Pathways Analysis

The NAFLD-inducing compounds' targets from STITCH and DrugBank, a prior-knowledge network from Reactome and our phosphoproteomic measurements formulated an optimization problem solved with an ILP formulation. The algorithm minimized the mismatch between our experimental data and the prior knowledge network. Each signaling pathway (**Figure P7-S2**) describe the compound's mechanism of action. The proteins involved in those pathways and their corresponding networks are combined and presented on **Figure P7-4**. Those networks visualize the identified molecular interactions that appear to be functional based on our experimental results.

Signature comparison and enrichment analysis

Using all experimental data available for each of the five compounds, a 2D plot was created using multidimensional scaling in order to visualize how closely related they are (**Figure P7-3**B).

Regarding the enrichment analysis, 72 gene expression profiles with mild or advanced NAFLD from liver biopsy tissue RNA dataset GSE49541[279] were compared with GEO2R to find differentially expressed genes important in the progression of NAFLD disease. Then, the list of top 250 genes of highest fold change were analyzed with Gene Ontology (GO) enrichment tool to find GO annotations. The gene lists

corresponding to the pathway nodes of AMI, FFA, TET, TMX, VPA were also analyzed for GO annotations. The GO annotations that are present in the analyzed dataset GSE49541 and at least one of the compounds' pathways are shown on **Table P7-S1** including their fold enrichment. GO:0030335 (positive regulation of cell migration), GO:2000147 (positive regulation of cell motility), GO:0051272 (positive regulation of cellular component movement) and GO:0040017 (positive regulation of locomotion) were significantly enriched only in the GSE49541, FFA and VPA lists. GO:0071229 (cellular response to acid chemical) and GO:0001101 (response to acid chemical) were significantly enriched in GSE49541 and TMX lists. GO:0042060 (wound healing) and GO:0009611 (response to wounding) were significantly enriched in GSE49541 and VPA lists.

Discussion

NAFLD has been described as an extremely complex disease, representing the convergence of several signaling networks and various phenotypes that will require study from many perspectives.[281] In this study, 5 clinical phenotypes of NAFLD were translated into 5 different in vitro models of primary human hepatocytes. Following a Systems Pharmacology experimental design, our experiments revealed irregular phosphorylation patterns caused by AMI, FFA, TMX and VPA stimulation of hepatocytes and significant (p value < 0.05) deregulations (>1.25-fold increase or decrease) were observed in the phosphorylation of intracellular proteins TF65, P53, CHK2, CREB1, EPOR, MP2K1 and MK03. The phosphorylation patterns were in good agreement with the literature: Phosphorylated TF65, also known as NFKB or RELA was significantly increased under FFA and VPA guided induction of NAFLD and came in agreement with activation patterns of TF65 that have been observed in obese patients[282] and rats with induced NAFLD[283]. VPA stimulation also significantly increased phosphorylation of MP2K1 and MK03, both very important components of the MAPK pathway. MP2K1, also known as MAP2K1 or MEK1 is an essential component of the MAPK/RAS-signaling module. Interestingly, MAP2K1 has been shown to export PPARG from the nucleus and PPARs are modulating hepatic triglyceride accumulation, a hallmark of NAFLD development.[284] Also, inhibition of

MP2K1 pathway in the liver has been shown to ameliorate Insulin resistance.[285] MK03, also known with its gene names MAPK3 or ERK1, has been shown to be significantly increased upon lipid accumulation induction.[286] Two further interesting phosphoproteomic deregulations were observed in the TMX stimulation group. Phosphorylated P53 and CREB1 were both significantly increased under TMX treatment. P53 is coded by the gene TP53, a tumor suppressor gene. It can induce growth arrest or apoptosis depending on the cell type and physiological circumstances. P53 becomes activated in response to DNA damage, oxidative stress and many other situations of stress. It is involved in cell cycle regulation, negatively regulating cell division by controlling a set of genes required for the process. It has been proved that the stage of liver steatosis increases the expression of pro-apoptotic proteins in hepatocytes, mainly P53.[287] CREB1 is a transcription factor involved in different cellular processes. Experiments have indicated a potential role for CREB1 antagonists as therapeutic agents in enhancing insulin sensitivity in the liver.[288] Other researchers have suggested that CREB1 is a potential therapeutic target for hepatic insulin resistance associated with NAFLD and they were also able to show that CREB knockdown in rodent models decreased hepatic triglyceride content.[289] On top of the phosphoproteomic results that were in good agreement of other researchers' work, two further irregular phosphorylation patterns were identified for CHK2 and EPOR which our findings suggest to be linked to NAFLD mechanisms. Phosphorylated CHK2 and EPOR were found significantly increased under TMX and AMI induction of NAFLD respectively. CHK2 is coded by the gene CHEK2 and is required for checkpoint-mediated cell cycle arrest, apoptosis in response to the presence of DNA double-strand breaks and activation of DNA repair. The gene CHEK2 gene can be seen on a pathway figure depicting the protein-protein interaction network for the differentially expressed genes between healthy and obese samples and normal controls in a publication about NAFLD but is not further explored by the authors.[290] EPOR mainly mediates erythropoietin-induced erythroblast proliferation and differentiation. EPOR was one of four hepcidin regulators whose expression was found altered in mice fed with a high-fat diet, but not further studied in the publication.[291] Further investigation is needed to uncover how those phosphorylated proteins may play a role in NAFLD progression.

As expected due to the multifactorial nature of NAFLD, significant differences were observed among the compounds used to induce NAFLD. On **Figure P7-3**B, an attempt in visualization of those differences using all collected phosphoproteomic data and multi-dimensional scaling to plot them in 2 dimensions. VPA appears to be closer to FFA than to the other compounds. VPA is an analog of the natural fatty acid valeric acid, therefore it is expected to cluster together with FFA in this graph. The similarities observed on AMI and TET can be attributed to the fact that both these compounds disrupt mitochondrial functions that lead to blockage of triglyceride secretion. On the contrary, TMX's steatogenic effects arise from increased fatty acid biosynthesis and subsequently increased triglyceride formation. Nevertheless, further investigations are needed to describe the observed similarities on NAFLD induction of these compounds.

The enrichment of GSE49541 data[279] (differentially expressed genes important to advancing NAFLD progression) was also compared with the enrichment of AMI, FFA, TET, TMX and VPA pathways. Four GO terms from three different GO biological process categories (cellular process, localization and locomotion) were enriched only in the GSE49541, FFA and VPA lists, all pointing to the same process of cell movement, while migration is known to be enhanced by FFA treatment.[292] Two GO terms about response to acid chemical and three GO terms regarding development of the circulatory system were enriched only in the GSE49541 and TMX lists. TMX has been linked with circulatory system regulation and is known to have cardiovascular effects.[293] Finally, there was a response to wound enrichment of two GO terms, appearing in the GSE49541 and VPA lists. Interestingly, VPA has been proven to induce cutaneous wound healing in vivo.[294] It should also be noted that EPOR plays an important role during wound healing[295] while CHK2 expression correlates with the peak cellular activity during liver regeneration.[253]

Conclusions

Since 2014, researchers have suggested in-depth study of signal transductions in order to provide novel solutions in curing or preventing NAFLD.[252] Today, NAFLD has become a threat to public health mainly because of the obesity epidemic and it is well

understood that it's a multifactorial disease. To our knowledge, this is the first study that aims to understand the multifactorial nature of NAFLD at the signaling level by studying 5 NAFLD induction models in primary human hepatocytes. Our results confirm a large body of literature findings for NAFLD signaling mechanisms. Furthermore, CHK2 and EPOR have emerged as potential NAFLD players that may be interesting to study further since they are important factors in liver regeneration[253]. The observed multifactorial nature of the disease suggests that there is no single treatment for all subtypes of NAFLD, highlighting the need for a systemic approach and personalised therapeutic interventions to better understand and treat NAFLD.

List of Abbreviations

NAFLD: Non-alcoholic Fatty Liver Disease
NASH: Non-Alcoholic Steatohepatitis
FFA: Free Fatty Acids
AMI: Amiodarone
TET: Tetracycline
TMX: Tamoxifen
VPA: Valproic Acid
EGF: Epidermal growth factor
IL1A: Interleukin 1A
IL6: Interleukin 6
LPS: Lipopolysaccharide
DMSO: Dimethyl sulfoxide
GEO: Gene Expression Omnibus
GO: Gene Ontology
MF: Molecular Function
BP: Biological Process
CC: Cellular Component
EPOR: Erythropoietin receptor
TF65: Transcription factor p65
P53: Cellular tumor antigen p53

CHK2: Serine/threonine-protein kinase Chk2
CREB1: Cyclic AMP-responsive element-binding protein 1
MP2K1: Dual specificity mitogen-activated protein kinase kinase 1
MK03: Mitogen-activated protein kinase 3

Development of a new drug repositioning platform for Non-Alcoholic Fatty Liver disease through network analysis.

Please visit abstract [A13] in section “Abstracts in International Conferences” for more information on this chapter.

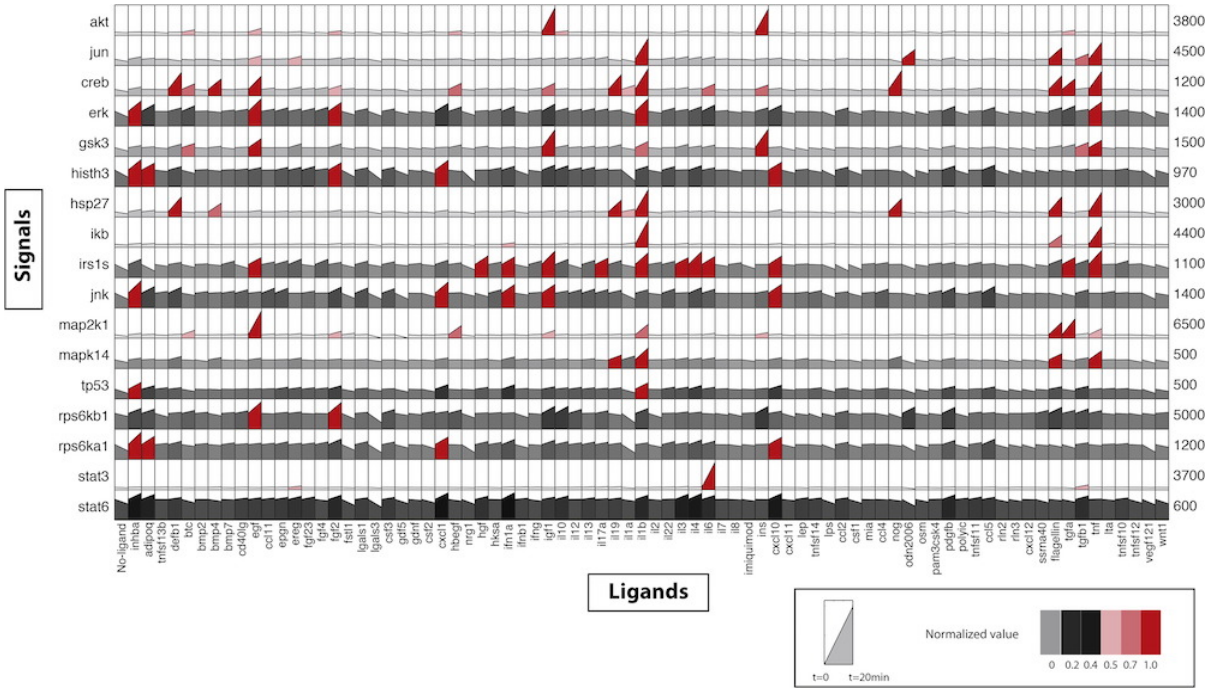
Non-alcoholic fatty liver disease (NAFLD) is defined as the presence of hepatic steatosis in the absence of excess alcohol consumption and is considered the hepatic manifestation of the metabolic syndrome. It is considered to be the most common pathological condition of the liver, but its cause and progression mechanisms are still not completely understood, partially because of the lack of sufficient in vitro models. There is no therapy approved specifically for NAFLD by FDA). The aim of this project is to develop a platform for drug repositioning in NAFLD/NASH by combining novel in vitro models of primary human hepatocytes with network-based analysis of gene expression and xMAP data.
For the in vitro induction of NAFLD, primary human hepatocytes (pHH) were exposed to free fatty acids (palmitic acid, oleic acid) and to the steatogenic compounds amiodarone, tamoxifen, tetracycline and valproic acid. The presence of steatosis was assessed using high content screening; lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. In addition, 17 phosphorylated protein targets were measured and signaling networks were constructed, in an attempt to shed a light into the mechanism of each in vitro approach to induce NAFLD.
A network-based computational approach was employed to suggest compounds for NAFLD. NAFLD-related networks were identified through (i) gene set analysis (GSA) and (ii) multiplex phosphoproteomic data. The common pathways with the aforementioned steatogenic compounds, used to induce NAFLD in vitro, were found

through DrugBank and MSig database. To suggest compounds that reverse the disease mechanism, the steatogenic compounds were used with the Connectivity Map database. The most promising compounds for drug repositioning were identified in the intersection between patient-derived and drug-derived networks.

In this work, we have successfully developed NAFLD-induced in vitro models driven by steatogenic compounds. xMAP phosphoproteomic data and pathway analysis identified deregulations of CREB1, ERK1, MEK1, P53 and NFKB that came in agreement with literature whereas the involvement of CHK2 and EPOR pathways seems to be unknown mechanisms for steatosis. Our drug reposition platform suggested 10 potential candidates that were tested in our in-vitro models. Interestingly, 5 of them have shown that can reverse in-vitro NAFLD. 3 of them are already in clinical trials whereas 2 more show very promising therapeutic potential.

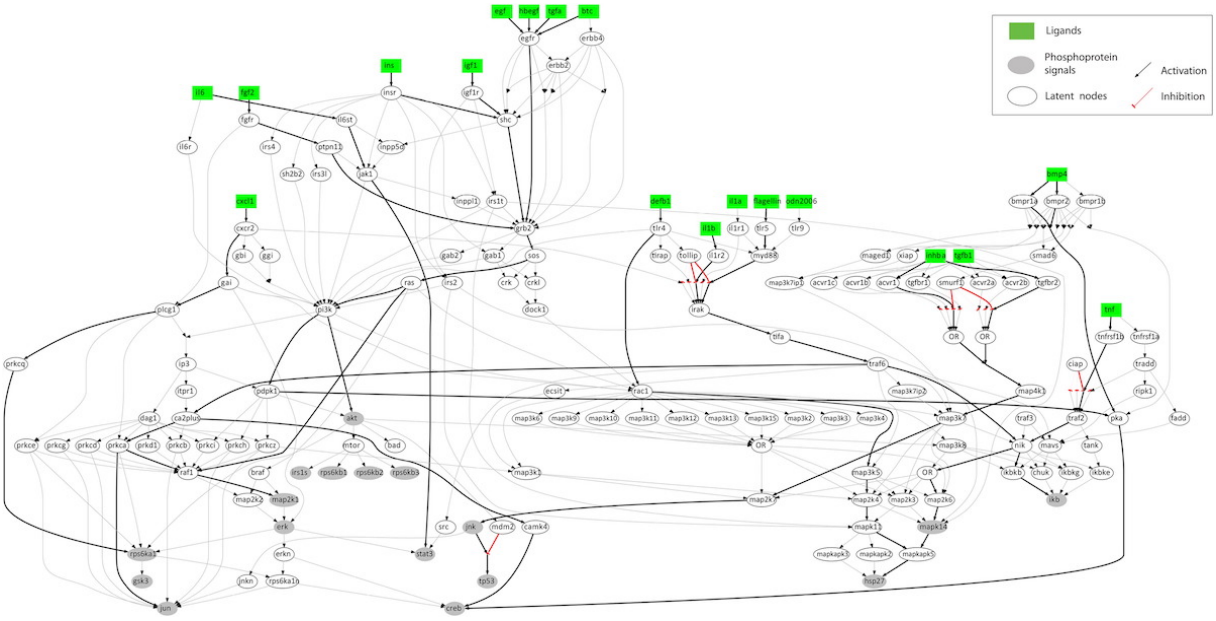
Εικόνες | Figures

Figure P1-1



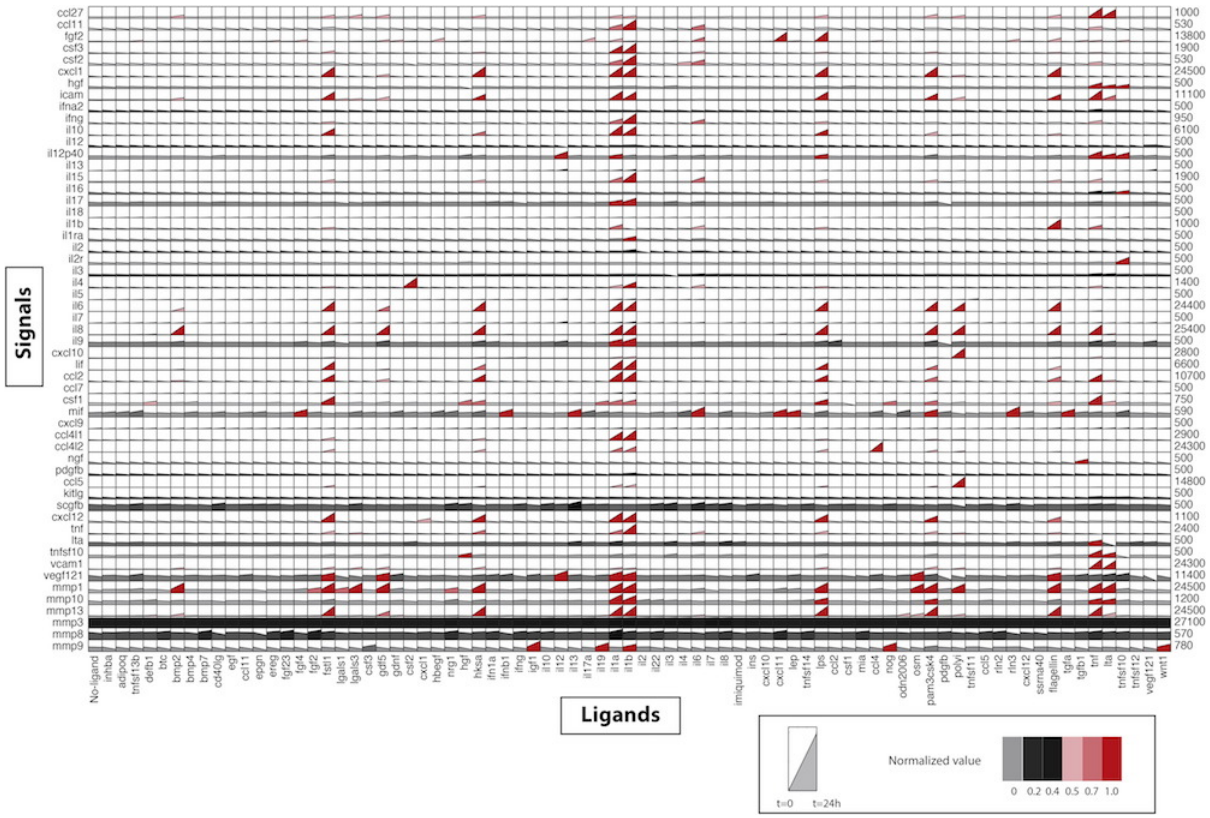
Phosphoproteomic data: The time course of the phosphoprotein signals from the unstimulated state to the average early response is illustrated. The rows correspond to the 17 phosphoproteins measured and the columns to the 79 ligand treatments (including the No-ligand treatment). In each subplot, the first point shows the unstimulated activity of the respective signal (zero time point); the second point shows the raw measurement of the signal (in fluorescent units) 5 þ 25 min after stimulation; while the color code corresponds to the normalized value (between 0 and 1) of the signal. The numbers on the right-hand side of the figure show the maximum phosphorylation value of each signal in fluorescent units.

Figure P1-2



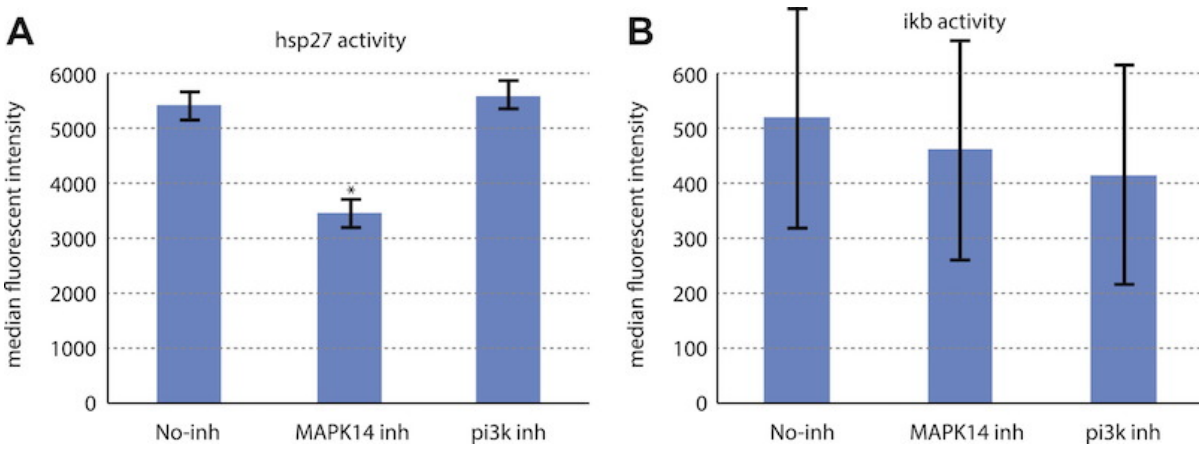
Optimized pathway: The signal transduction network upon optimization via the ILP formulation. Green rectangular nodes correspond to the imposed ligands, gray elliptic nodes to the measured phosphoproteins, gray rectangle nodes to the measured cytokine releases and clear (white) elliptic nodes to latent signaling proteins in the network. The optimization procedure trains the PKN by removing reactions that contradict the proteomic data at hand. Thick black edges denote the reactions validated by the data; gray edges denote the reactions that contradicted the data and were removed by the ILP algorithm.

Figure P1-3



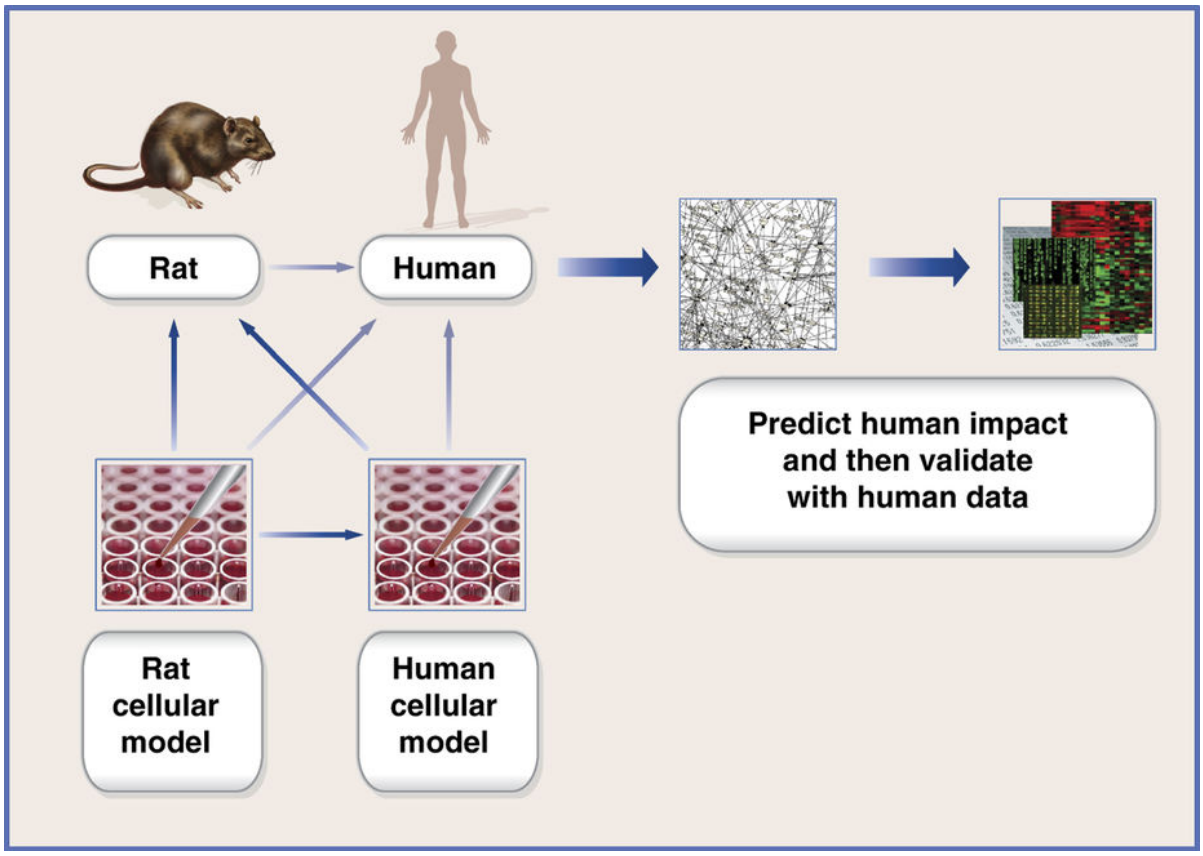
Cytokine release data: The time course of the cytokine releases from the unstimulated state to 24 h is illustrated. The rows correspond to the 55 cytokine releases measured in the supernatant and the columns to the 79 ligand treatments (including the No-ligand treatment). In each subplot, the first point shows the unstimulated concentration of the respective cytokine in the supernatant (zero time point); the second point shows the raw measurement of the signal (in fluorescent units) 24 h after stimulation; while the color code corresponds to the normalized value (between 0 and 1) of the signal. The numbers on the right-hand side of the figure show the maximum phosphorylation value of each signal in fluorescent units.

Figure P1-4



Effects of the MAPK14 and PI3K inhibition on the activity of IKB and HSP27 upon stimulation with IL1B and TNF: Inhibition of MAPK14 clearly decreases HSP27 activity upon IL1B and TNF stimulation, validating that HSP27 is activated in a MAPK14 dependent manner, in good accordance to model predictions. IKB activity is not blocked neither by MAPK14 nor PI3K inhibition validating that IKB is activated by an independent pathway. The numbers on the left correspond to the raw phosphorylation value of the measured signals in fluorescent units.

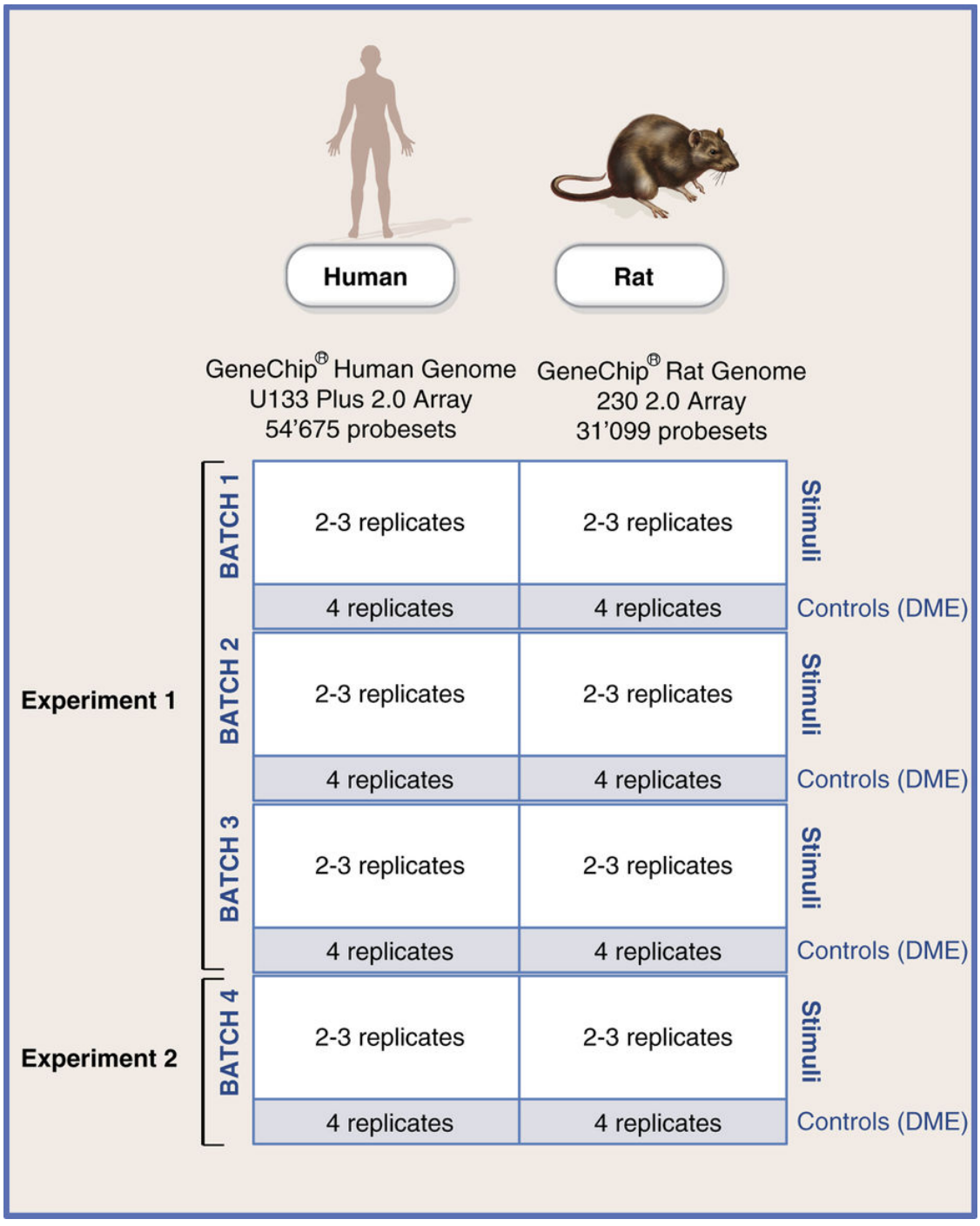
Figure P2-1



Concept of translatability.

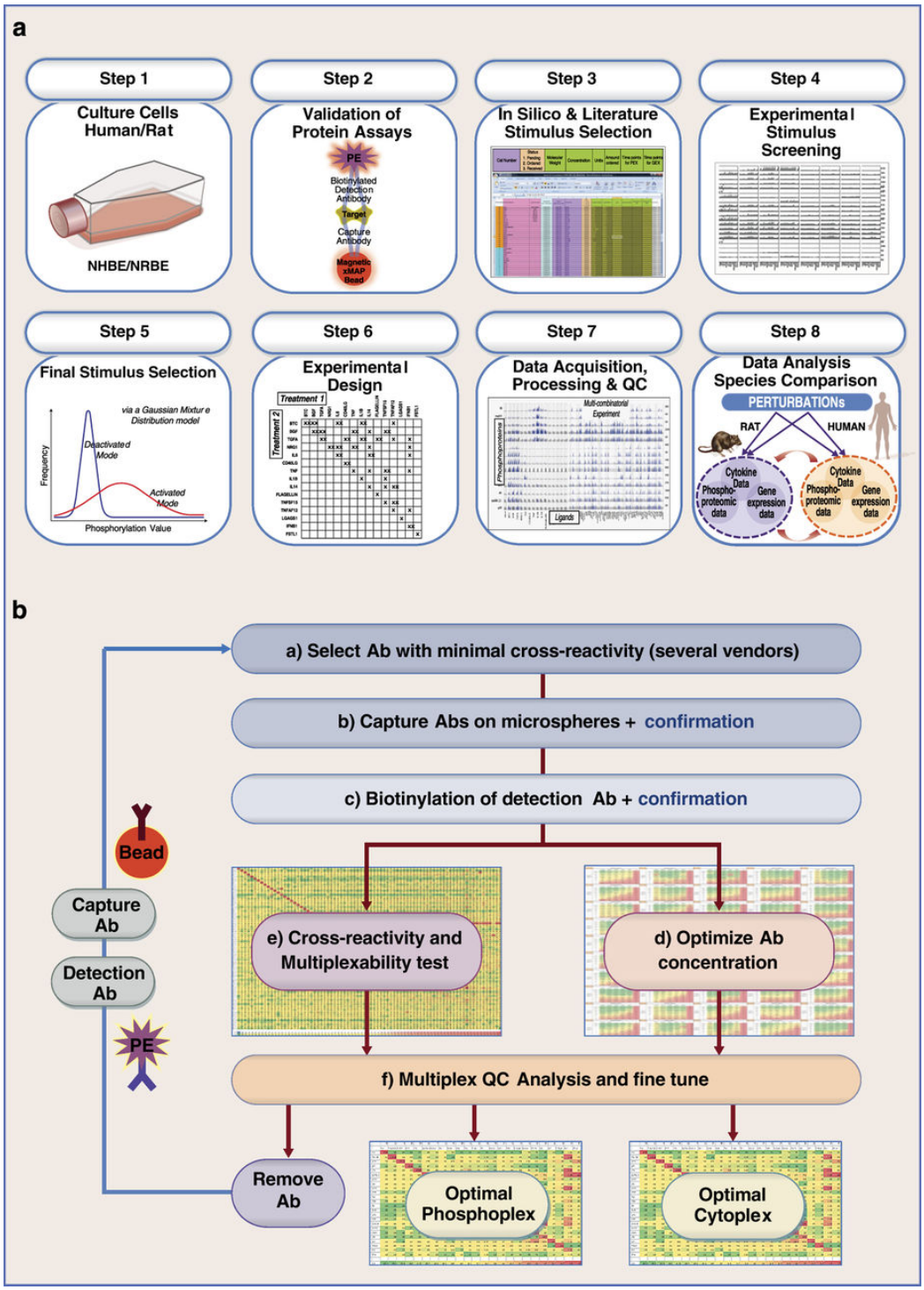
The arrows indicate potential routes of translation between in vitro and in vivo systems and/or across species.

Figure P2-2



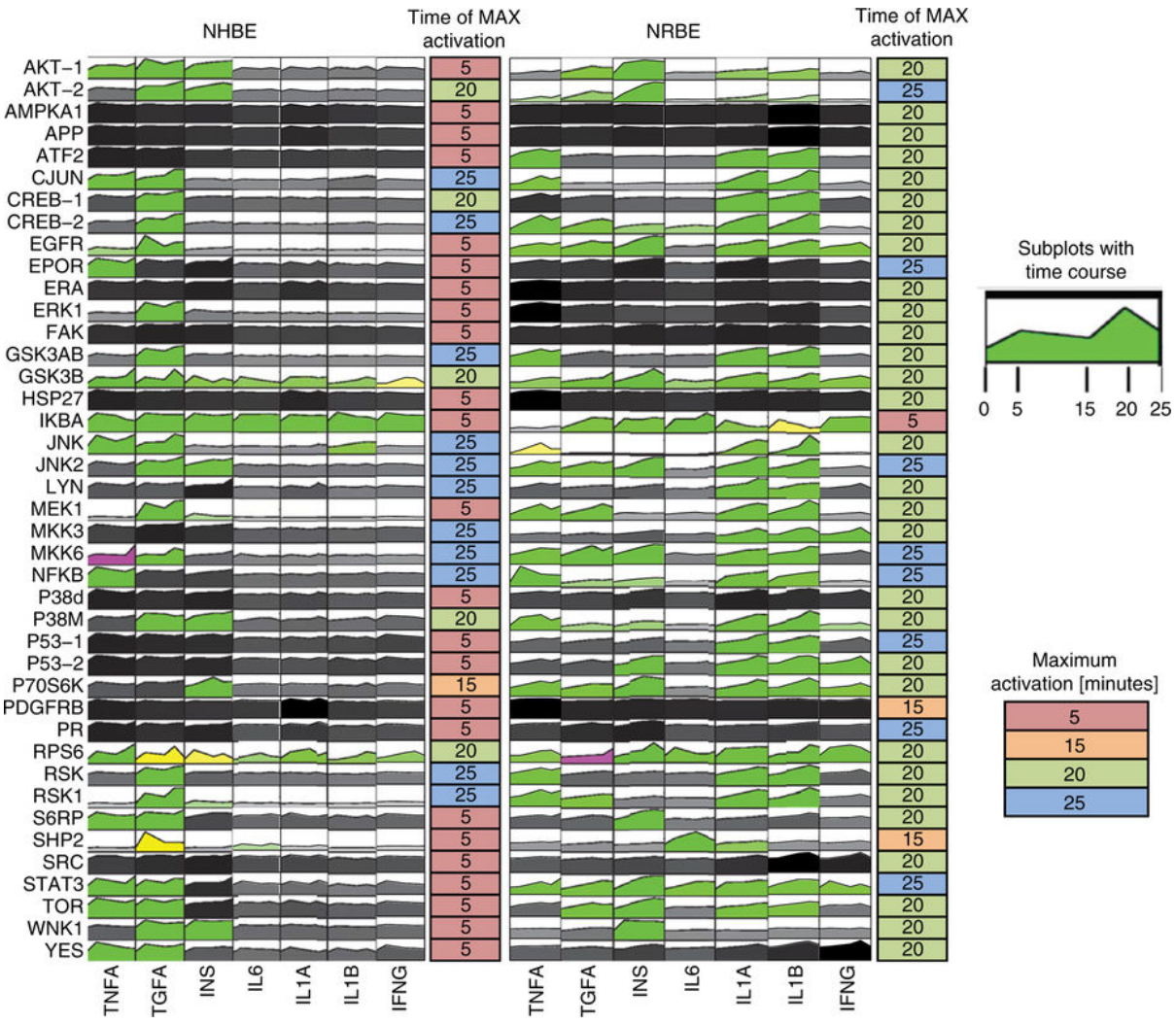
Schema of the mRNA processing to generate the gene expression dataset avoiding confounding effects between species and between DME controls and treatment conditions.

Figure P2-3



Overall experimental workflow. (a)- Experimental steps followed to generate the STC multi-layer omics dataset compendium for translational systems biology. (b)- Pipeline for the development and optimization of antibody-based multiplexed assays (detailed description of step 2 'Validation of protein assays').

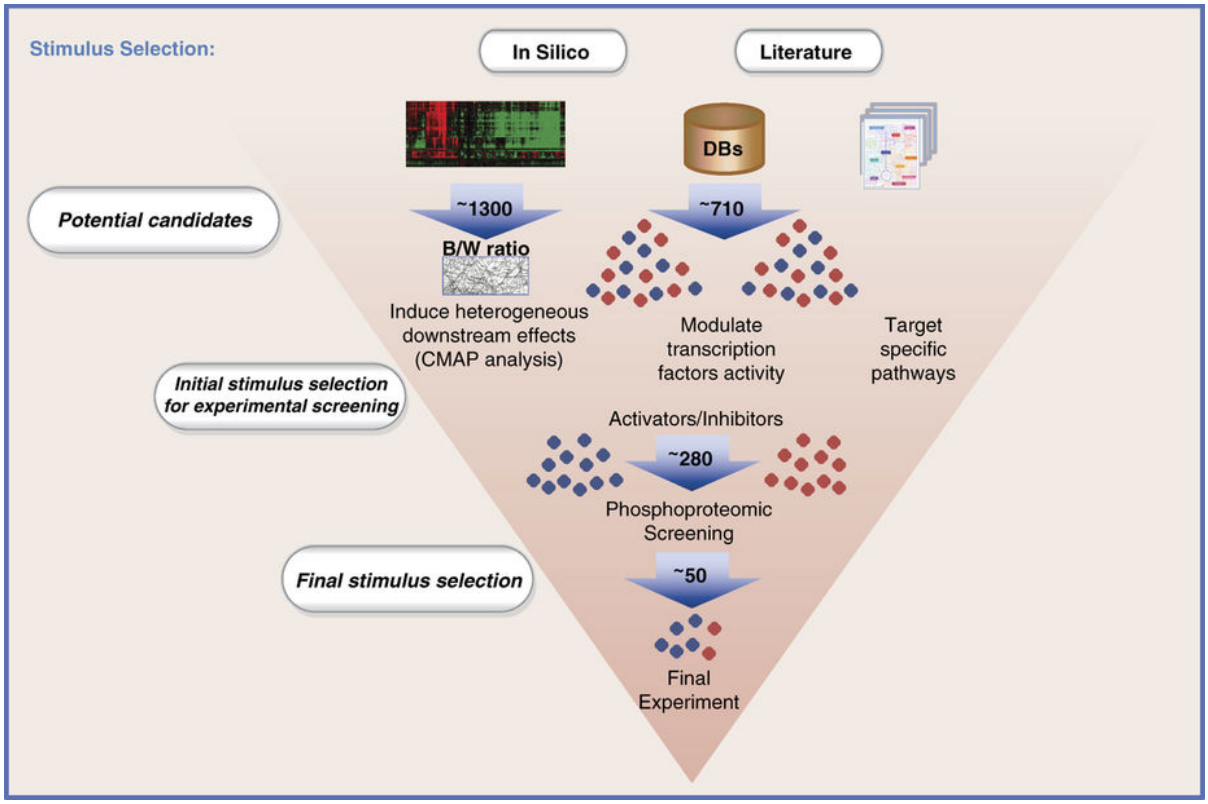
Figure P2-4



The determination of optimal time points for phosphoproteomics measurements in NHBE and NRBE: Human and rat bronchial epithelial cells were treated with seven stimuli at five different time points (0, 5, 15, 20 and 25min). The time course of the raw data (fluorescent intensity: FI) for each phosphoprotein was plotted in subplots using a modified version of DataRail. The solid fill colours (yellow, green, purple, grey/black) of the time course correspond to different signal behaviour over time according to the DataRail colouring scheme. Yellow colour corresponds to transient activity (FI increases and then decreases), green colour corresponds to sustained activity (FI increases and remains active), purple colour corresponds to late activity (FI starts stable and then increases) and grey/black to no change (FI increase/decrease compare to basal level at 0 time point less than 50% across all

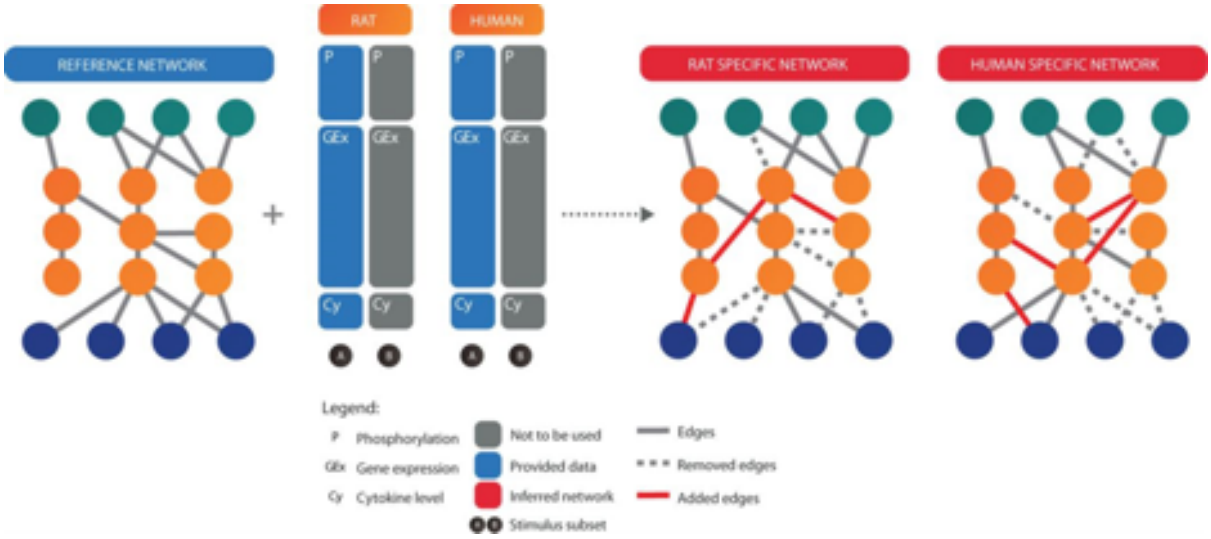
time points - the darker the grey colour the bigger the average FI). In the majority of experiments, maximum phosphoprotein activation in NHBE cells was found at 5 (red) and 25 (blue) minutes, whereas NRBE cells were maximally activated at 20 (green) and 25 (blue) minutes. Thus, 5 and 25min were selected as the optimal time points for both cell types.

Figure P2-5



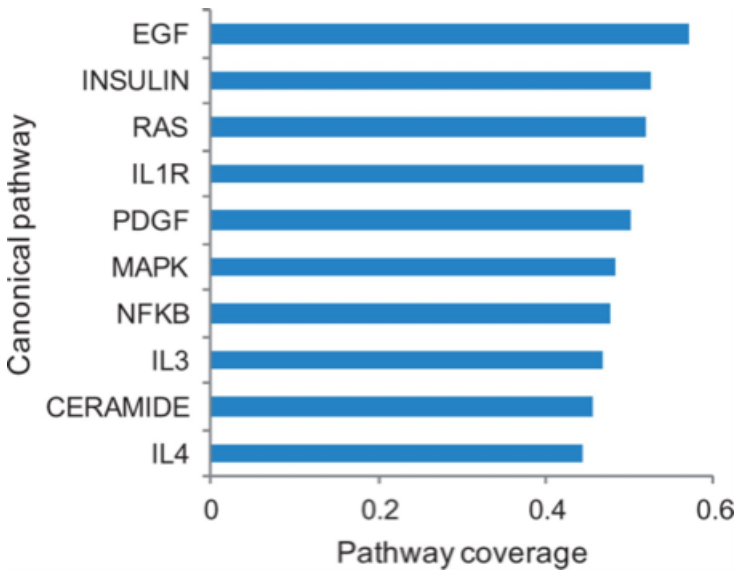
The process of selection of the stimuli used to generate the dataset for the Species Translation Challenge: The selection processes involve various steps, including in silico analysis, literature review, and phosphoproteomics screening.

Figure P3-1



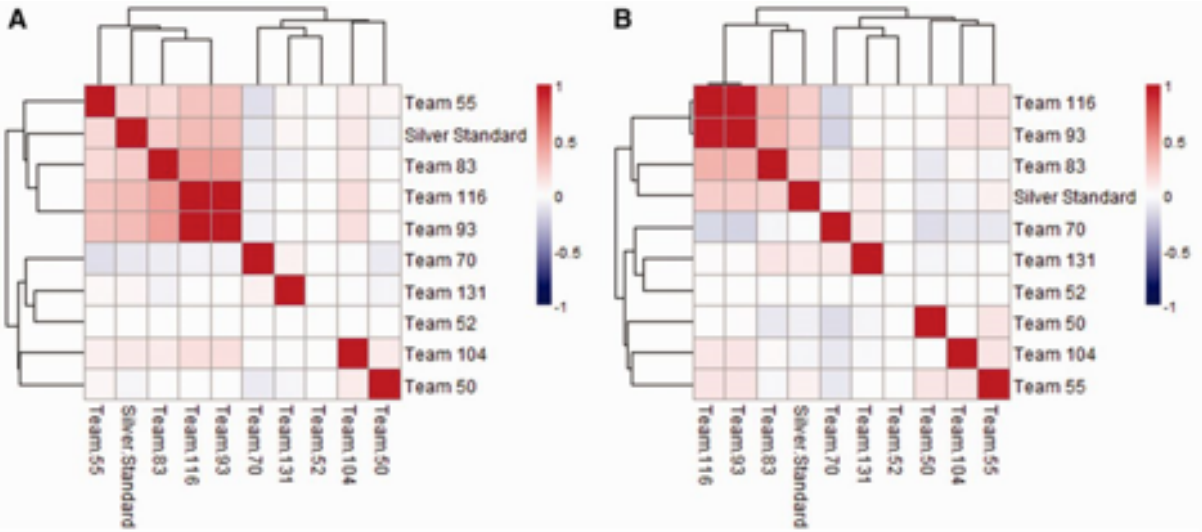
Overview of the Network Inference Challenge. Participants are provided with a reference network together with Affymetrix gene expression and Luminex phosphoproteomics and cytokine data derived from human and rat bronchial epithelial cells. The goal is to generate two separate networks for human and rat by adding and removing edges from the reference network using the data provided

Figure P3-2



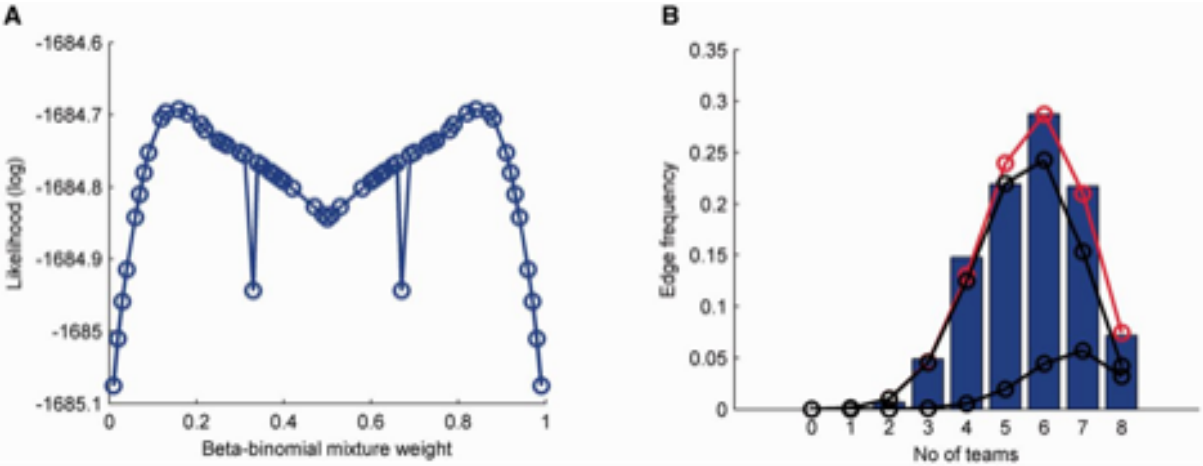
The top 10 canonical pathways represented in the reference network. The pathways are ordered by the proportion of genes present in the reference network

Figure P3-3



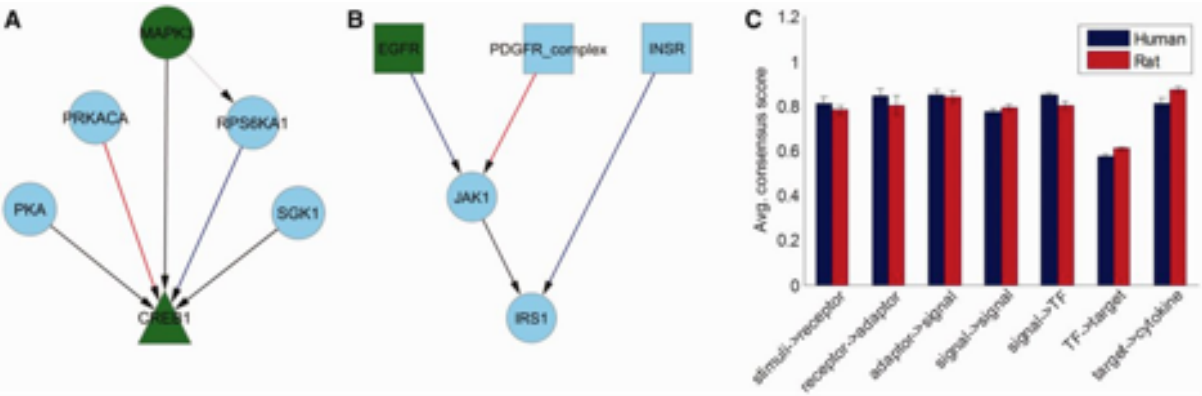
The predicted networks for human (A) and rat (B) were compared with the silver standard and against each other using MCC. Only edges present in the reference network were considered

Figure P3-4



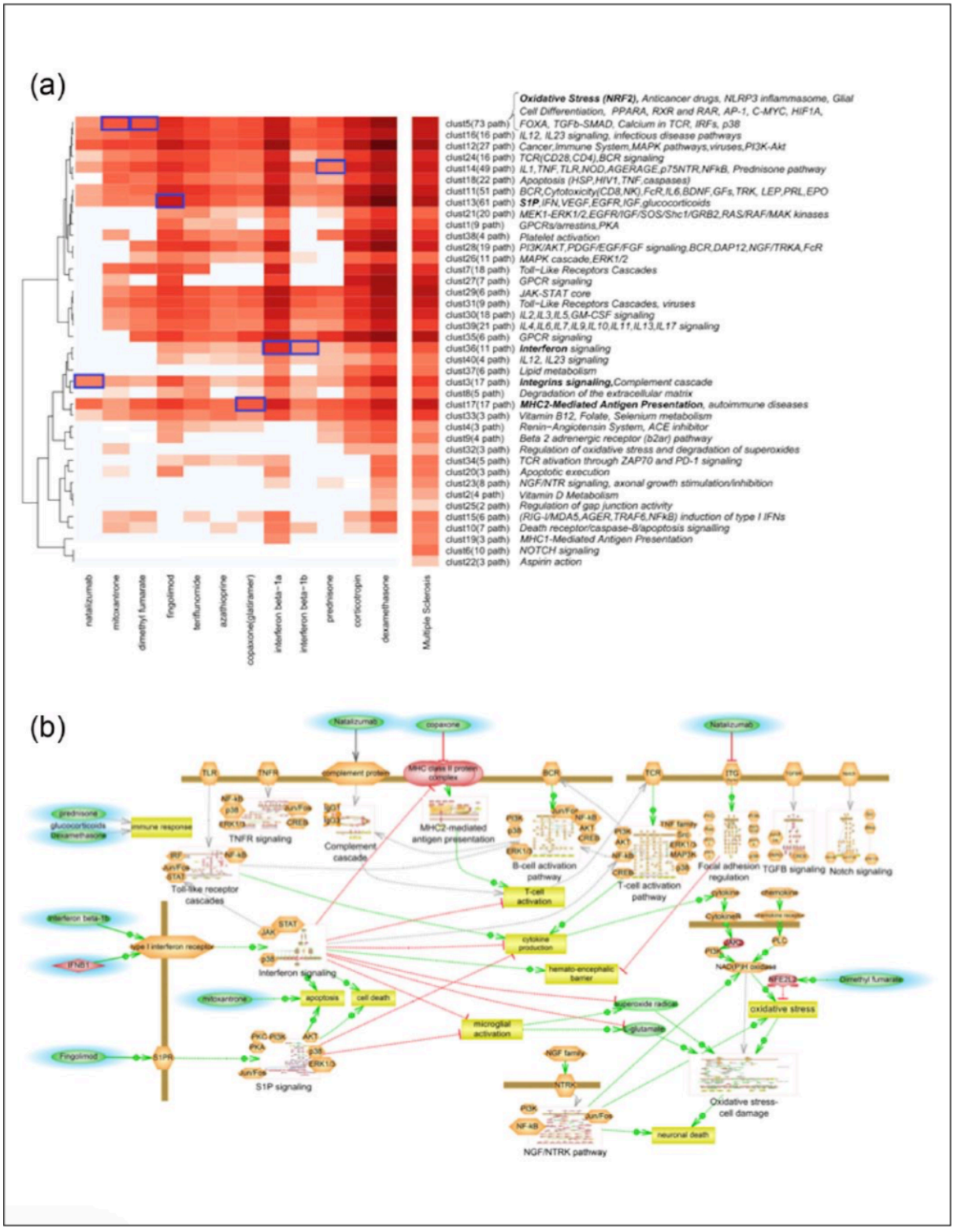
(A) The beta-binomial mixture weight can be calculated by maximizing the log-likelihood function. (B) Using this value, the fitted mixture is shown in red together with the individual-weighted components in black. Only edges present in the reference network were used in this case

Figure P3-5



Panels A and B show two example subnetworks of the consensus network where in blue are human-specific edges, in red rat-specific edges and in black edges common to both species. Depicted in gray are edges from the original reference network that did not gather sufficient consensus between participants. Panel C shows the average consensus score of the edges between a layer and the next one downstream from it for human and rat networks

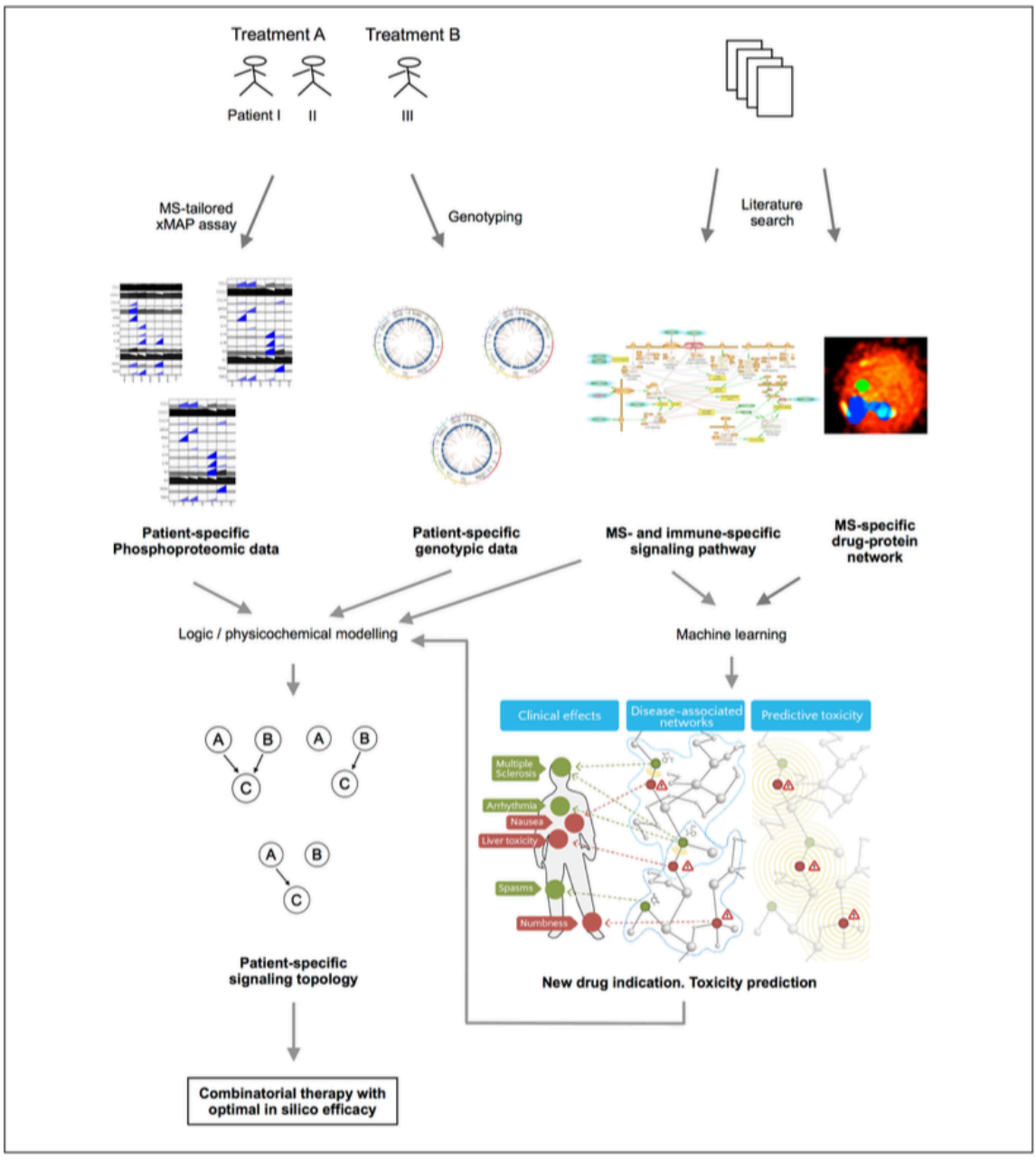
Figure P4-1



Signaling pathways associated with multiple sclerosis (MS). (a) An example of the pathways implicated in MS that can be identified from databases. The right column

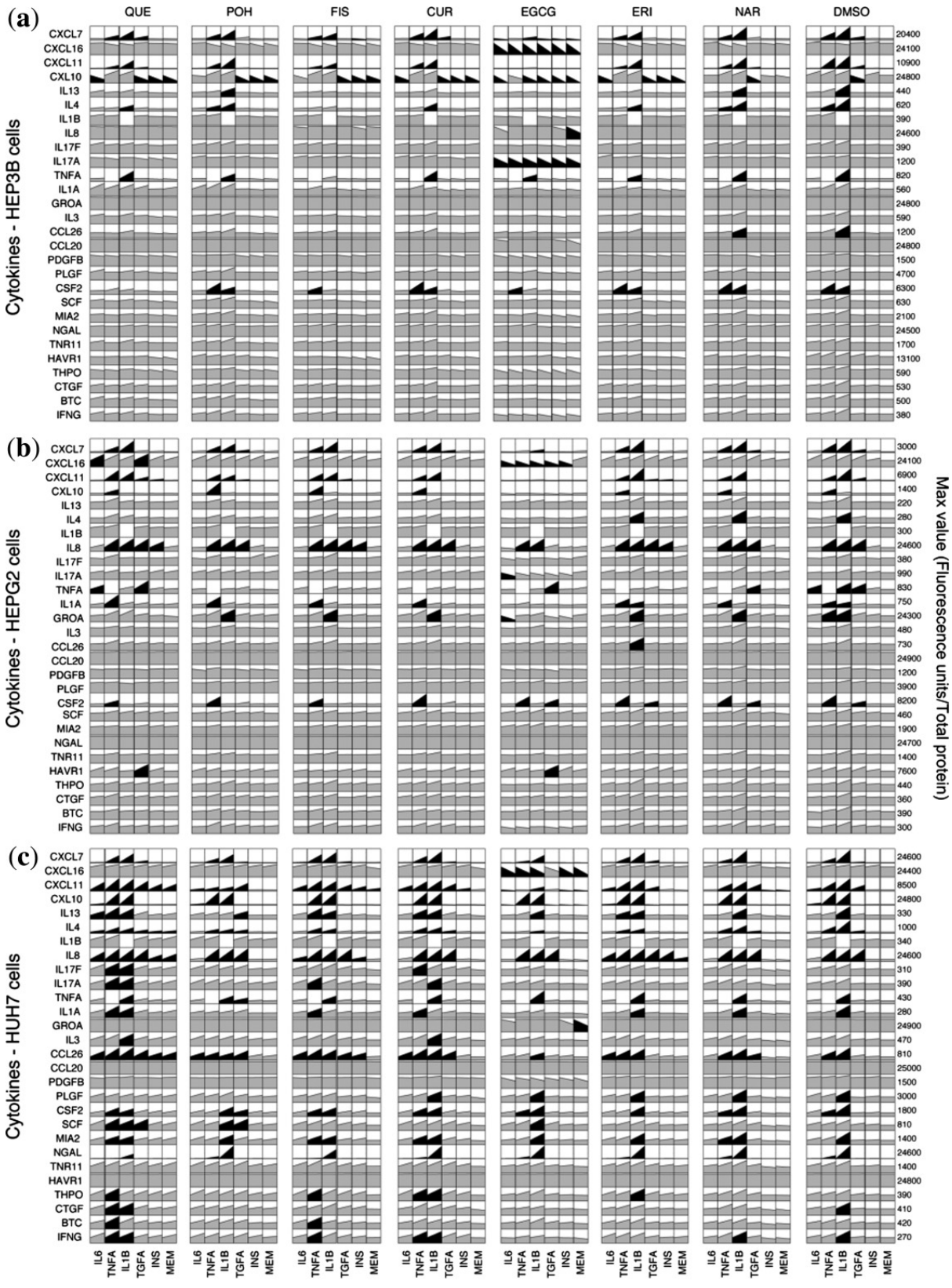
shows the heat map of the pathways associated with MS and the left columns the heat map of the pathways associated with each of the drugs. On the right, each cluster is followed by the list of the pathways it includes (in bold the pathways specifically target by a drug). Blue squares: known mechanisms of action for a given drug. (b) Integration of signaling pathways implicated in MS in network models: The genes/proteins associated with MS are displayed in orange, drugs are in green, and the main MS pathways targeted by therapies are in yellow.

Figure P4-2



Pipeline for the identification of new therapies based on the modeling of signaling pathways associated with multiple sclerosis (MS) and MS drugs. Flow from first to second row panels: Experimental set-ups, such as proteomics and genotyping, can be tailored to interrogate MS-specific signatures in terms of phosphoproteomics (rows, phosphorylation profile of specific proteins e.g. xMAP assays; Columns, MS-related treatments) and the risk variants. A literature search enables MS- and immune-specific pathways to be compiled and drug-protein networks can be assembled (the hot scale shows the density of proteins in the signaling pathways, and the upper layer shows green and blue clusters of proteins targeted by MS-related drugs). Flow from second to third row panels: Logic and dynamic models can be constructed based on MS- and immune-specific signaling pathways. In order to study how signaling is deregulated in MS, one model can be calibrated against a patient-specific dataset, thereby yielding an ensemble of patient-specific models that enables common signaling mechanisms and those that explain patient-to-patient variability to be discriminated. In parallel, the signaling pathways and drug-protein networks can be used as an input for machine-learning approaches in order to reposition existing drugs that can be used to infer new drug indications or to predict toxicity. From new drug indication to model: Literature search of existing drugs and their targets, combined with search of those targets in MS-specific networks yields MS-specific drug-protein networks, which suggests new drug indications by identifying the interactions from drug targets to MS networks. These newly indicated drugs can then be introduced in the predictive models to understand their mechanisms of action in order to select those drugs with the best potential efficacy.

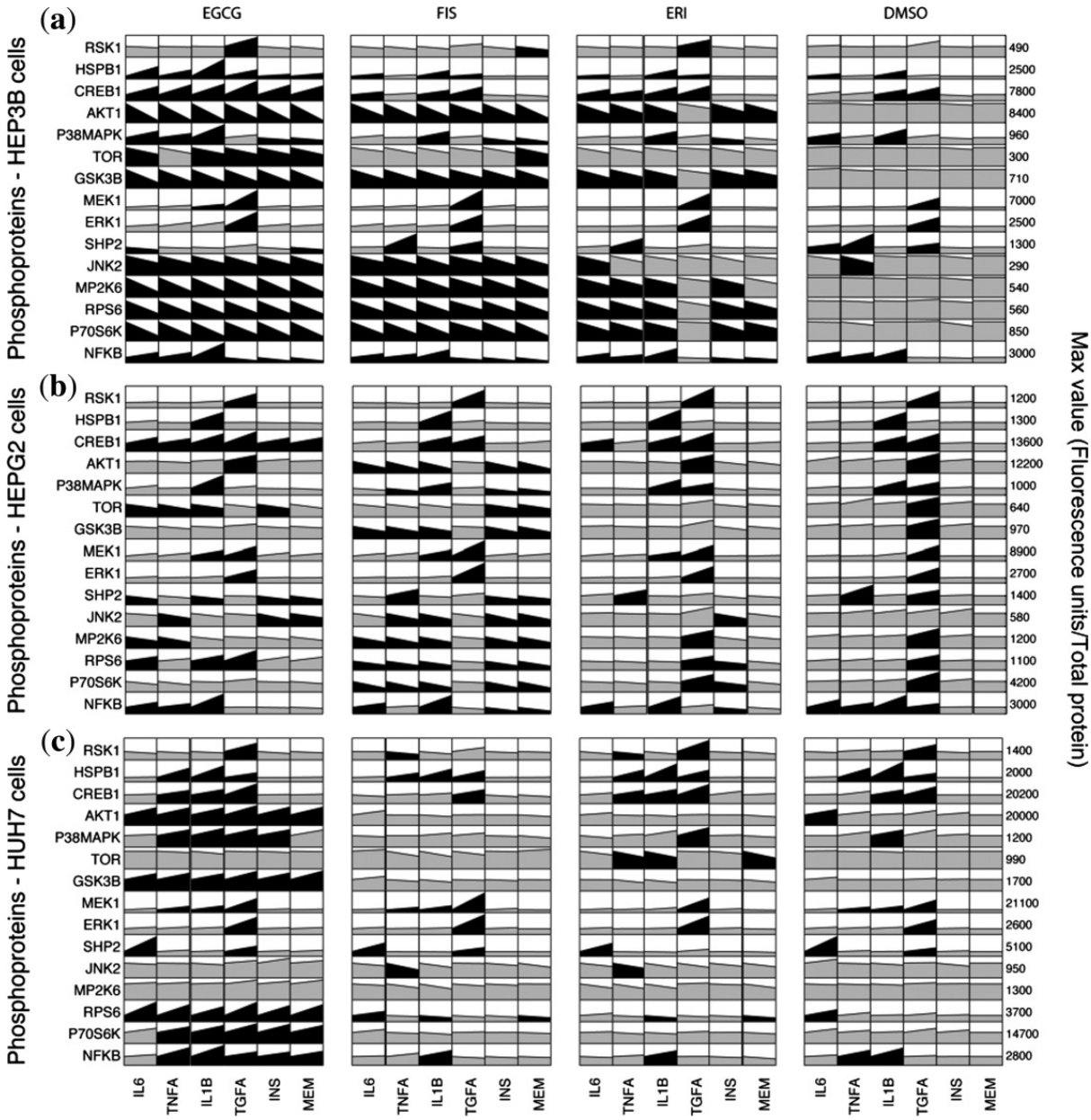
Figure P5-1



Cytokine xMAP datasets from (a) HEP3B, (b) HEPG2, and (c) HUH7 cells. Rows

represent measurements (mean values) of secreted cytokines 24 hours posttreatment. Columns correspond to treatment with indicated stimulus or minimum essential medium (MEM). Column blocks designate treatment with indicated compound or DMSO vehicle. Black color highlights signals that exceeded a 2-fold threshold alteration compared to DMSO-MEM. Mean values±SEM (2–3 experiments) and data transformation of fluorescence units to pg/ml are provided in Appendices I and II of the Supplementary Material, respectively.

Figure P5-2



Phosphosignaling xMAP datasets from (a) HEP3B, (b) HEPG2, and (c) HUH7 cells. Rows represent measurements (mean values) of phosphoproteins at 15 minutes following stimulation. Columns correspond to treatment with indicated stimulus or minimum essential medium (MEM). Column blocks designate treatment with indicated compound or DMSO vehicle. Black color highlights signals that exceeded a 1.5-fold threshold alteration compared to DMSO-MEM. Mean values±SEM (2–3 experiments) are provided in Appendix III of Supplementary Material.

Figure P5-3

Compound-specific signaling networks in HEP3B cells treated with EGCG under (a) basal and (b) stimulated conditions. Compound-phosphoprotein expression-cytokine release pathways were constructed from xMAP data and a reference network with canonical pathways using an adapted SigNetTrainer method. Black, opaque edges correspond to interactions that were found to be functional based on the data at hand and were conserved in the solution. Gray edges correspond to interactions that were found to contradict with the data and were thus removed from the solution. Black dashed edges correspond to interactions that may be functional or not, depending on the data discretization threshold used. The thickness of the dashed lines corresponds to the number of solutions (for different thresholds) that support the respective interaction. Blue edges correspond to interactions that were added by the algorithm for explaining patterns in the data that could not be fitted with solely removing interactions from the canonical pathway. Abbreviations are explained in Appendix IV of the Supplementary Material.

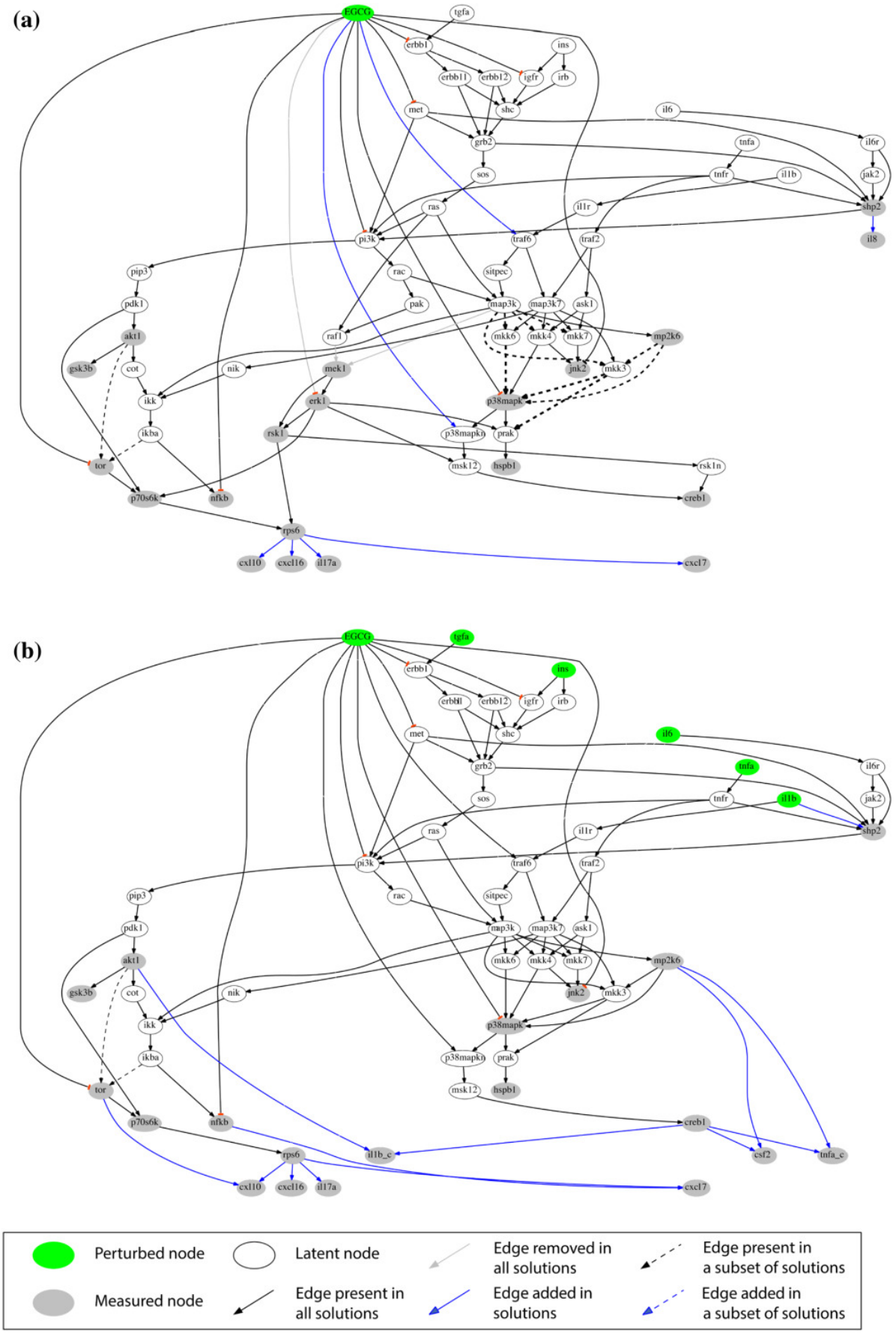
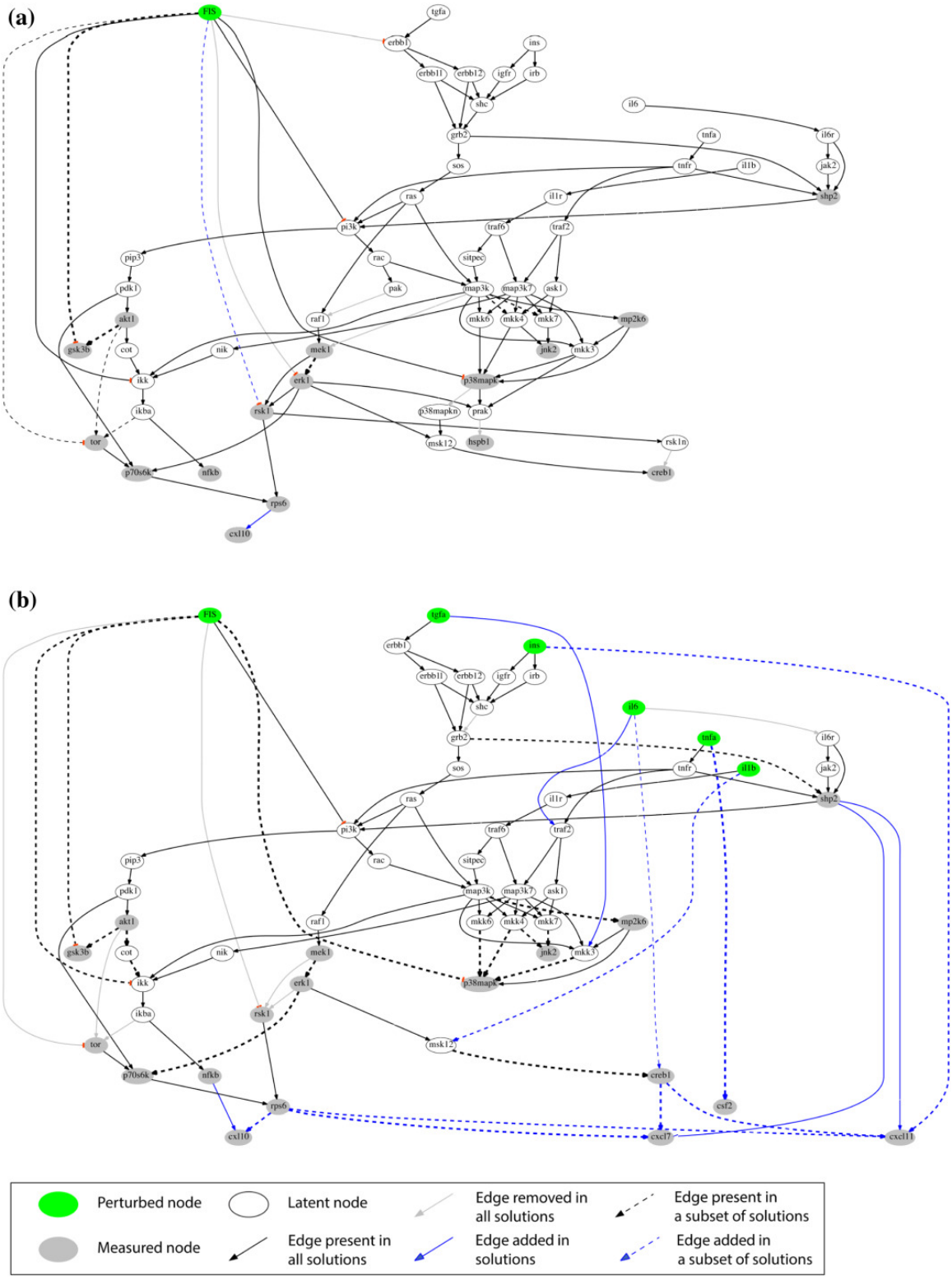
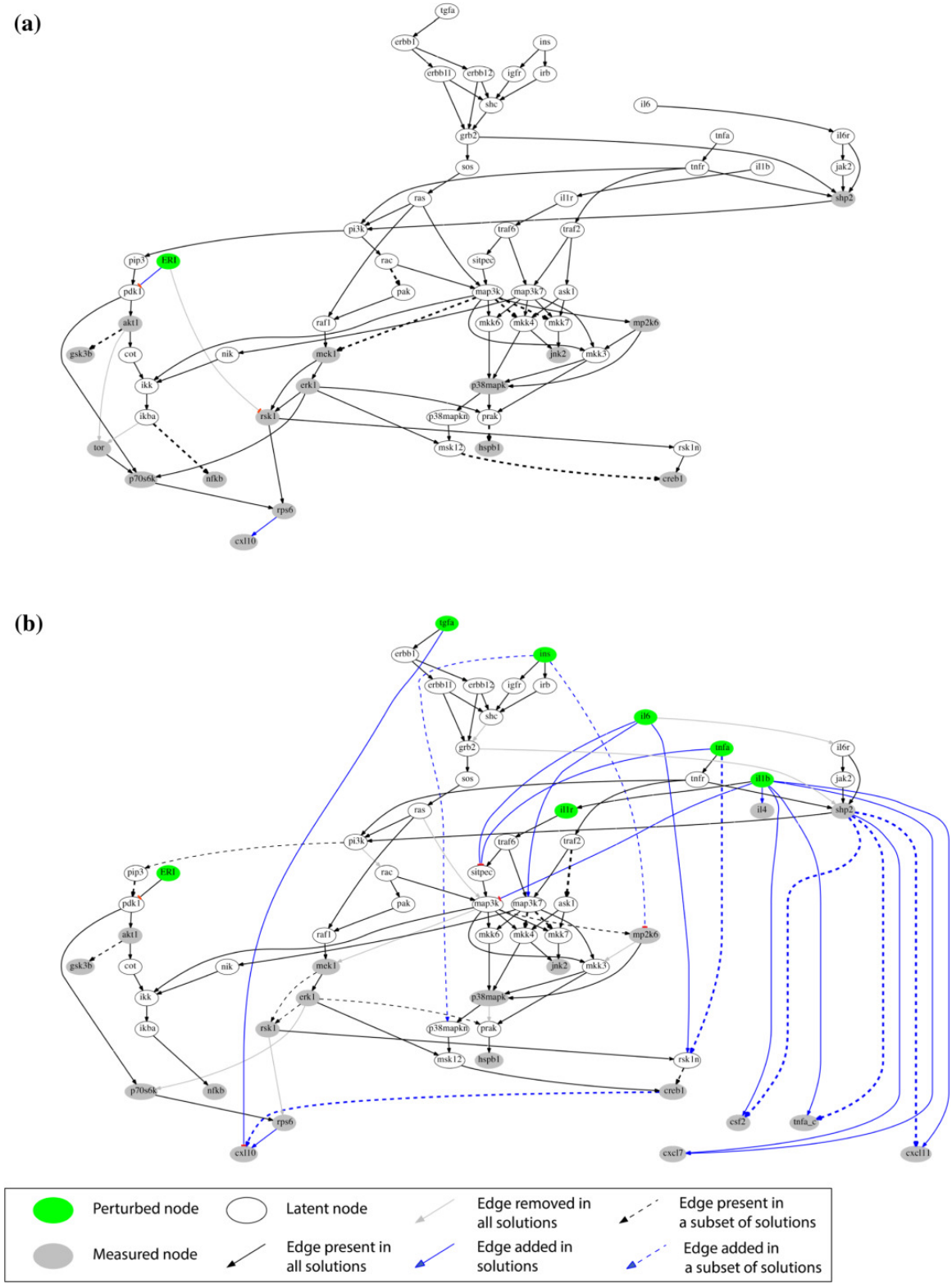

(a)
EGCG
tgfa
erbb1
ins
erbb11
erbb12
igfr
irb
shc
met
il6
grb2
tnfa
il6r
sos
tnfr
il1b
jak2
ras
il1r
shp2
pi3k
traf6
traf2
il8
pip3
rac
sitpec
pdk1
pak
map3k
map3k7
ask1
akt1
raf1
mkk6
mkk4
mkk7
mp2k6
gsk3b
cot
nik
mek1
jnk2
mkk3
ikk
erk1
p38mapk
ikba
rsk1
p38mapkb
prak
tor
msk12
hspb1
rsk1n
p70s6k
nfkb
creb1
rps6
cxl10
cxcl16
il17a
cxcl7
(b)
EGCG
tgfa
ins
erbb1
erbb11
erbb12
igfr
irb
shc
met
il6
grb2
tnfa
il6r
sos
tnfr
il1b
jak2
ras
il1r
shp2
pi3k
traf6
traf2
pip3
rac
sitpec
pdk1
map3k
map3k7
ask1
akt1
mkk6
mkk4
mkk7
mp2k6
gsk3b
cot
nik
jnk2
mkk3
ikk
p38mapk
ikba
p38mapkb
prak
tor
msk12
hspb1
p70s6k
nfkb
creb1
il1b_c
csf2
tnfa_c
rps6
cxl10
cxcl16
il17a
cxcl7
Perturbed node
Latent node
Edge removed in all solutions
Edge present in a subset of solutions
Measured node
Edge present in all solutions
Edge added in solutions
Edge added in a subset of solutions

Figure P5-4



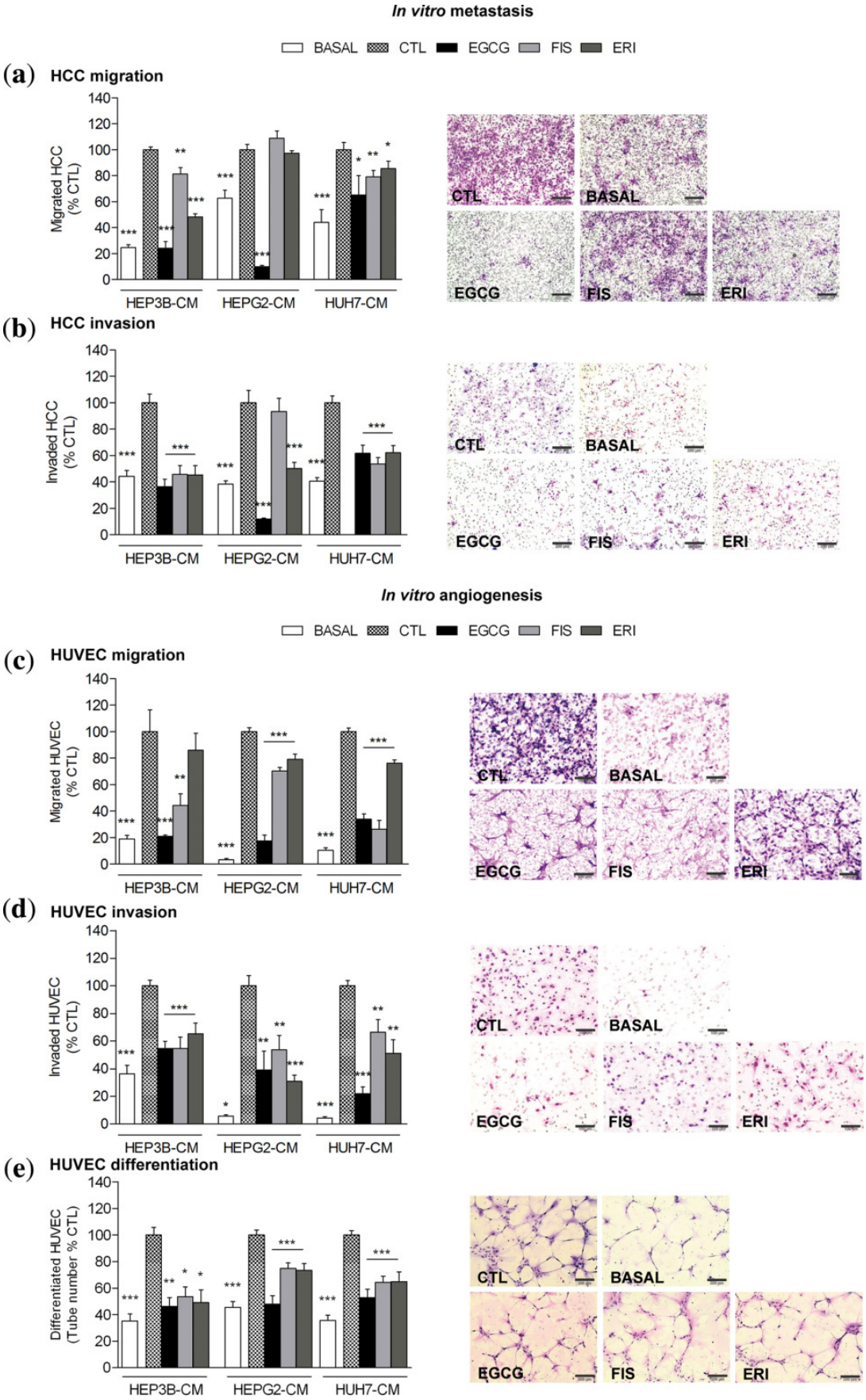
Compound-specific signaling networks in HEP3B treated with FIS under (a) basal and (b) stimulated conditions (see legend to **Figure P5-3** for details).

Figure P5-5



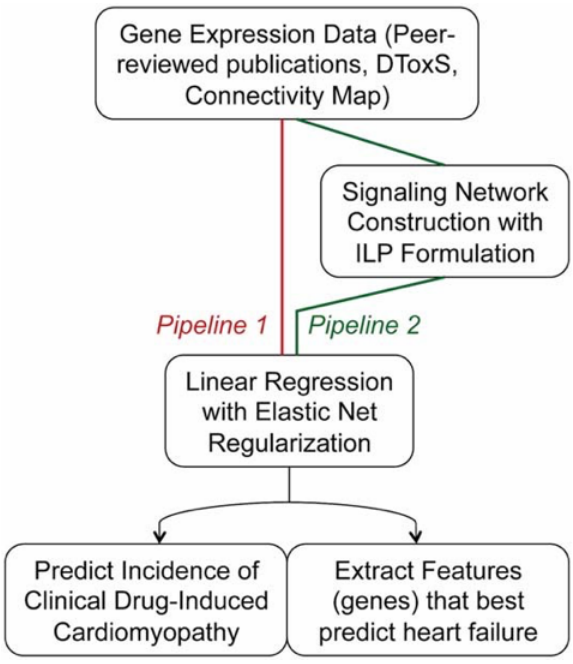
Compound-specific signaling networks in HEP3B treated with ERI under (a) basal and (b) stimulated conditions (see legend to **Figure P5-3** for details).

Figure P5-6


In vitro metastasis
BASAL CTL EGCG FIS ERI
(a) HCC migration
Migrated HCC (% CTL)
HEP3B-CM
HEPG2-CM
HUH7-CM
CTL
BASAL
EGCG
FIS
ERI
(b) HCC invasion
Invaded HCC (% CTL)
In vitro angiogenesis
(c) HUVEC migration
Migrated HUVEC (% CTL)
(d) HUVEC invasion
Invaded HUVEC (% CTL)
(e) HUVEC differentiation
Differentiated HUVEC (Tube number % CTL)

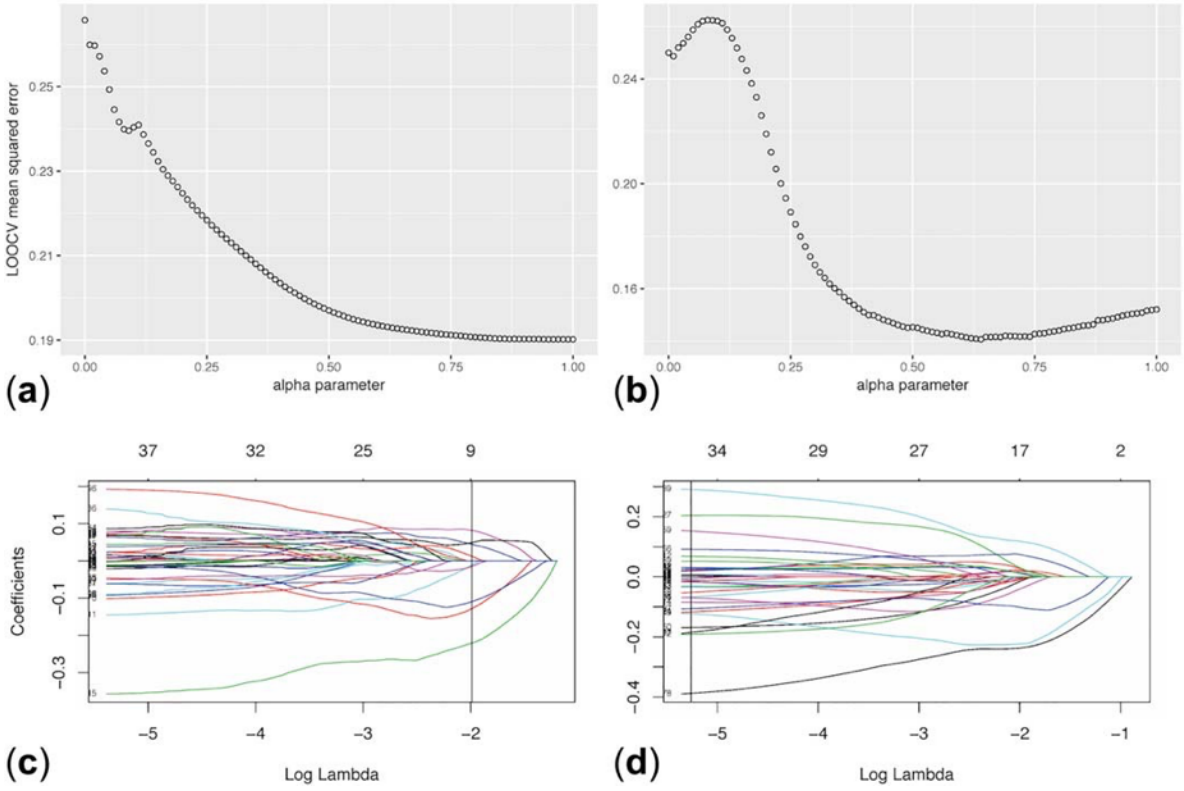
Preventive potential of test compounds on HCC-induced metastasis and angiogenesis. HCC cells (a,b) or HUVEC (c,d) loaded onto migration or Matrigel invasion chambers chemotactically moved towards starvation medium (BASAL) or HCC-CM from DMSO-vehicle (CTL), EGCG, FIS, or ERI pretreated cancer cells. (e) HUVEC differentiation on Matrigel-coated plates in presence of starvation medium (BASAL) or HCC-CM from DMSO-vehicle (CTL), EGCG, FIS, or ERI pretreated cells. Results show mean percentage of CTL±SEM of migrating/invading cells (n=10) or tube-like structures (n=15; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). Representative microphotographs from HEP3B metastasis (a,b) and HEP3B-induced angiogenesis (c–e) assays are shown on the right panels; magnification ×100 (a,b,e) or ×200 (c,d); scale bars=200 μm (a,b,e) or 100 μm (c,d).

Figure P6-1



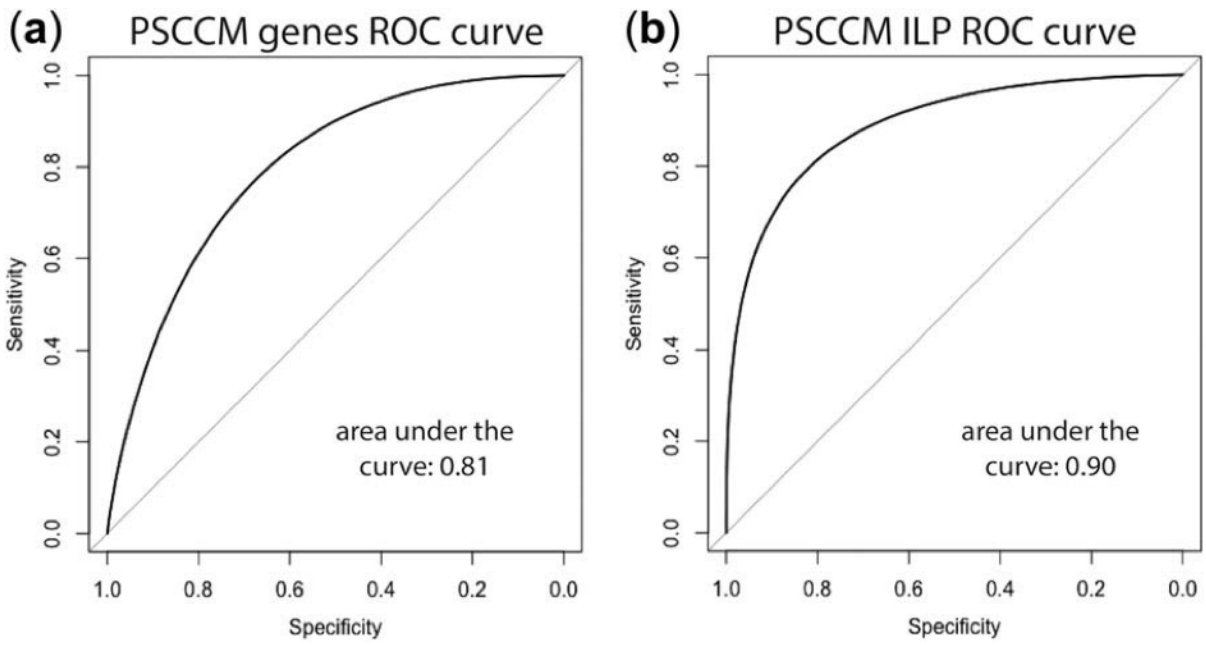
Workflow of predictive modeling. We built datasets using gene expression data and we compared two pipelines to predict clinical drug-induced cardiomyopathy and extract features that best predict such toxicity. Running the Gene Expression Data at hand through a linear regression model with elastic net regularization or constructing signaling networks from the data before modeling using an integer linear programming (ILP) formulation. DToxS, Drug Toxicity Signature Generation Center.

Figure P6-2



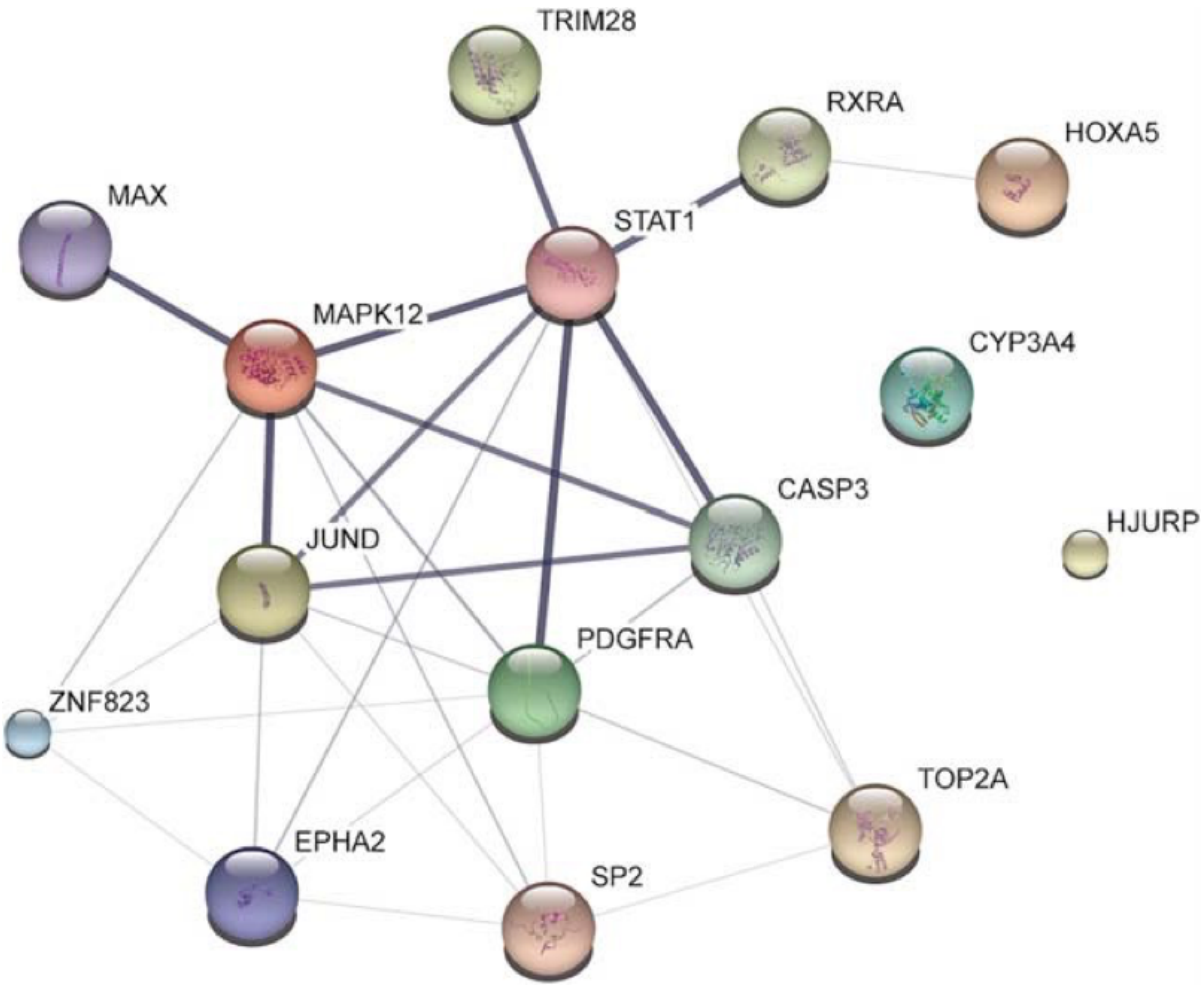
Plots of elastic net regularization results. (a and b) Show selection of the alpha parameter in the elastic net regularization by minimizing the leave-one-out cross validation (LOOCV) mean squared error to extract the features (genes) that best predict clinical incidence of cardiomyopathy. (c and d) Show the number of variables kept in the model, with a vertical line showing the optimal number for maximization of accuracy. a and c refer to the results of analyzing gene expression data only, whereas b and d correspond to the results of analyzing drugs' signaling networks obtained from integer linear programming formulation analysis. Each of the plotted lines in c and d corresponds to a variable (for example, a specific gene's expression) and shows how its coefficient changes with the log lambda parameter of elastic net. The vertical line shows the optimal number of parameters kept and their coefficients for maximization of accuracy.

Figure P6-3



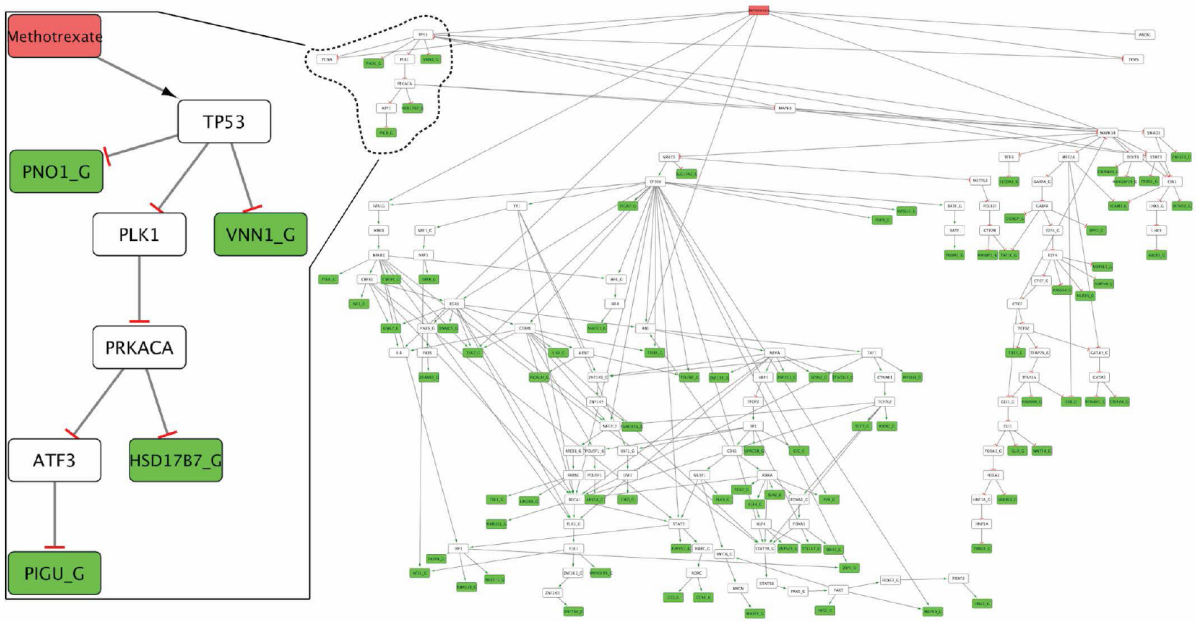
Receiver operating characteristic (ROC) curves. (a) ROC curve from modeling differentially expressed genes (DEGs) using elastic net (EN) and (b) ROC curve from modeling by subjecting these DEGs to integer linear programming (ILP) to construct their individual drugs' signaling networks and then subject these networks to EN. PSCCM, human cardiomyocytes.

Figure P6-4



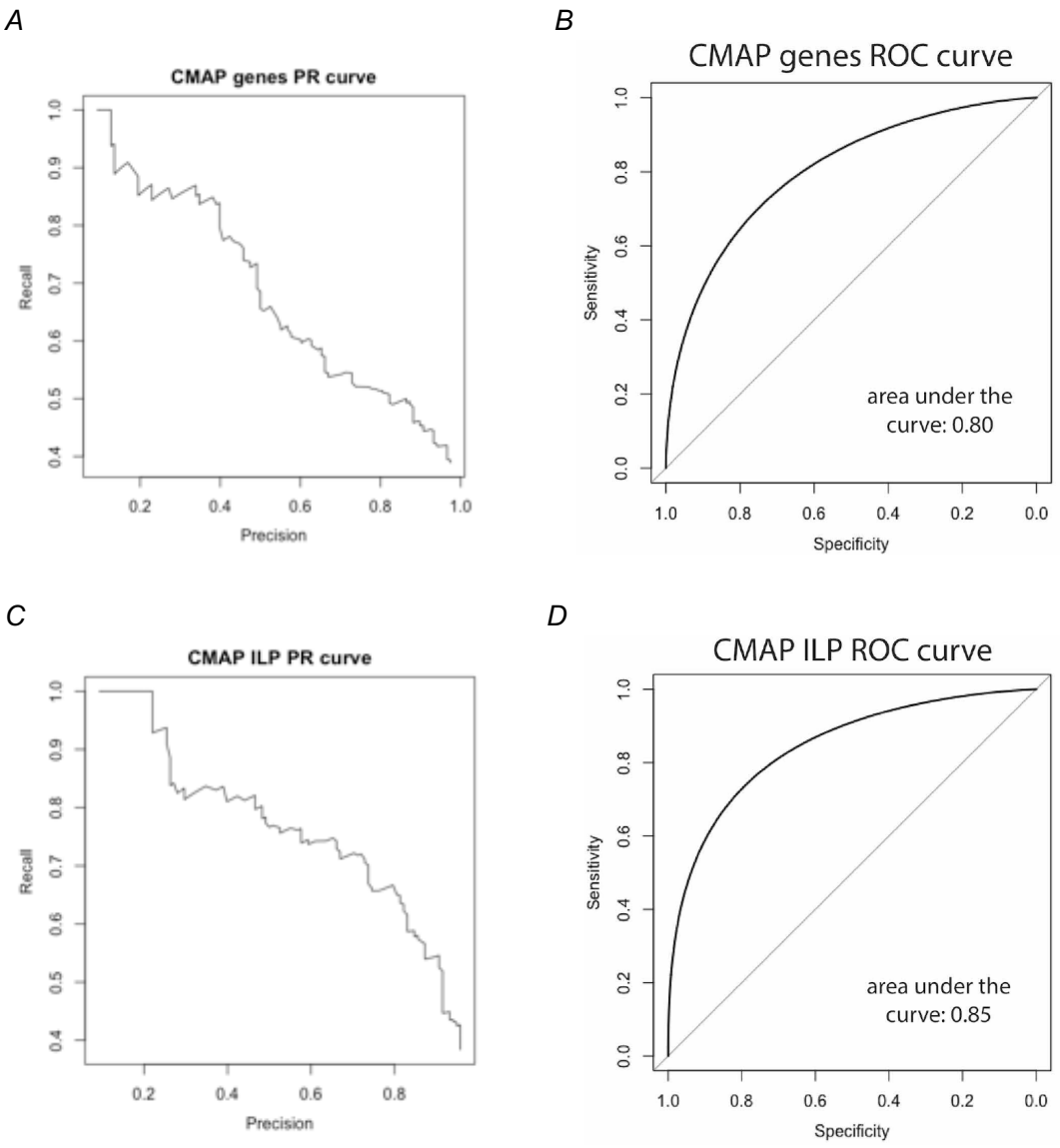
Interactions among the top 15 gene/protein predictors. Interactions among the top 15 genes/proteins selected by our model to best predict cardiomyopathy using cardiomyocytes data are depicted as a network using the STITCH website for visualization. Small nodes correspond to protein of unknown 3D structure and large nodes to known or predicted. Edges represent protein-protein associations and the intensity of the line is proportional to the confidence score of each association. The confidence score is calculated by combining the probabilities from all evidence channels and is corrected for random observation probability.

Figure P6-S1



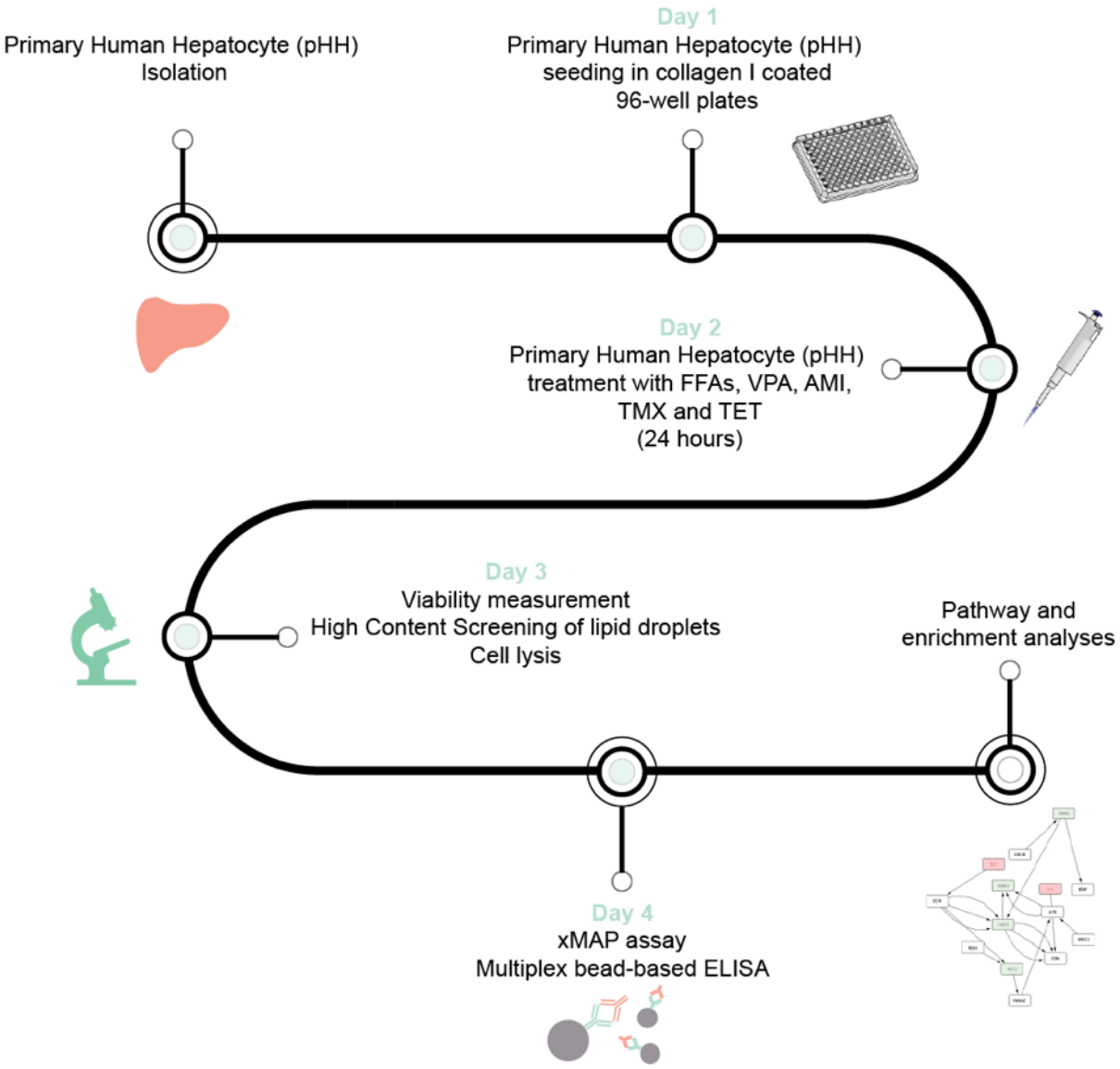
To better illustrate how the proposed ILP formulation works, we examined the mode of action of methotrexate (MTX). Let's focus on a small part of it involving TP53, PRKACA and ATF3 transcription factors (a similar rationale applies to the rest of its network). MTX activated TP53, and TP53 in turn blocked PNO1 and VNN1 genes via direct interactions, and blocked PIGU and HSD17B7 genes via activating PLK1, PRKACA and ATF3. The above is driven by MTX-induced under-expression of PNO1, VNN1, HSD17B7 and PIGU genes. ILP tried to explain these under-expressions, navigate the prior knowledge network of protein interaction, and identify a pathway connecting MTX (via one of its targets) with these 4 genes.

Figure P6-S2



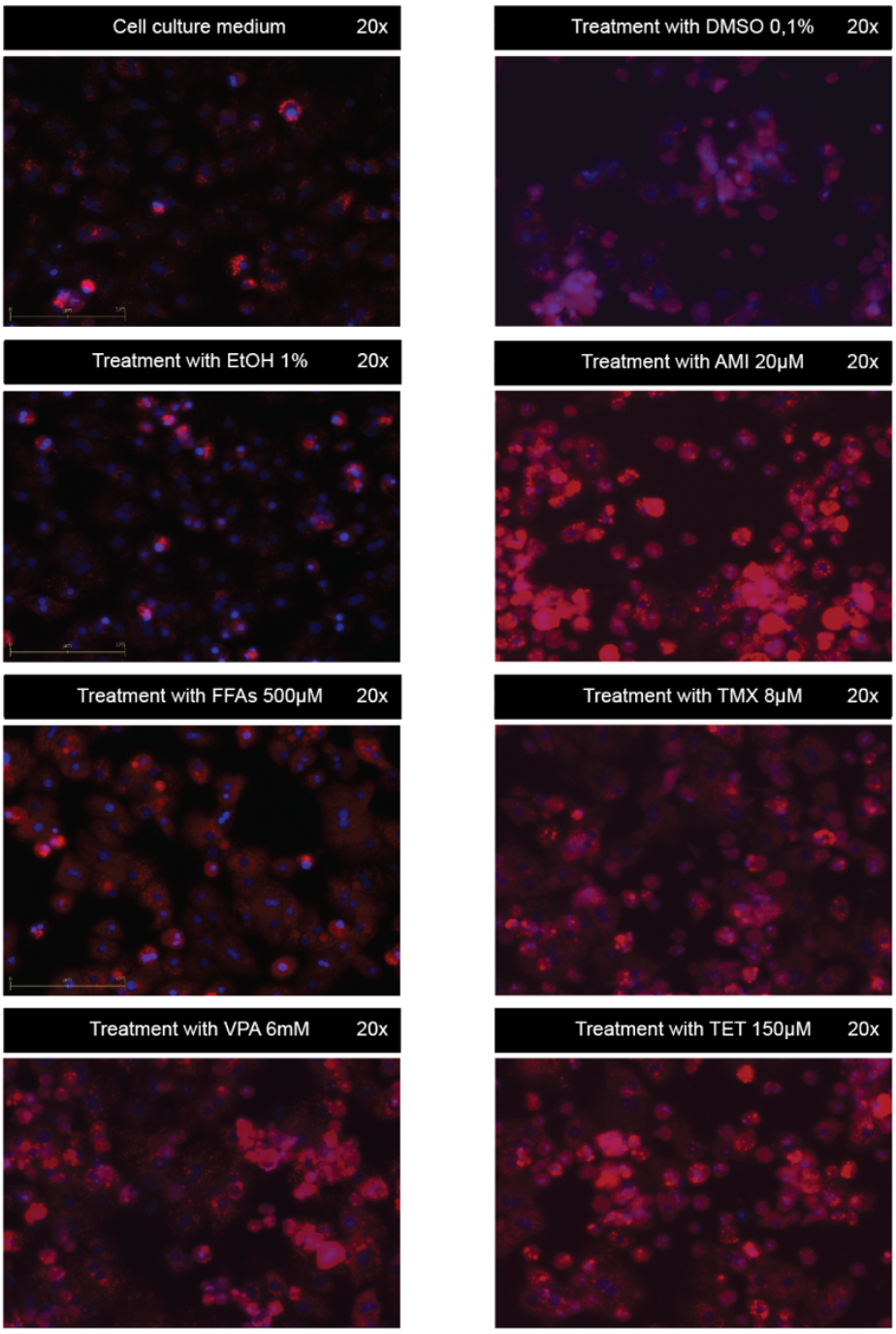
Panels A and C are the precision-recall curves for the same set of CMap differentially expressed genes that were modelled using Elastic Net (EN) alone and an integer linear programming(ILP)-EN combination), respectively. Panels B and D are the ROC curves for the same set of CMap genes that were modelled using EN alone and ILP-EN, respectively.

Figure P7-1



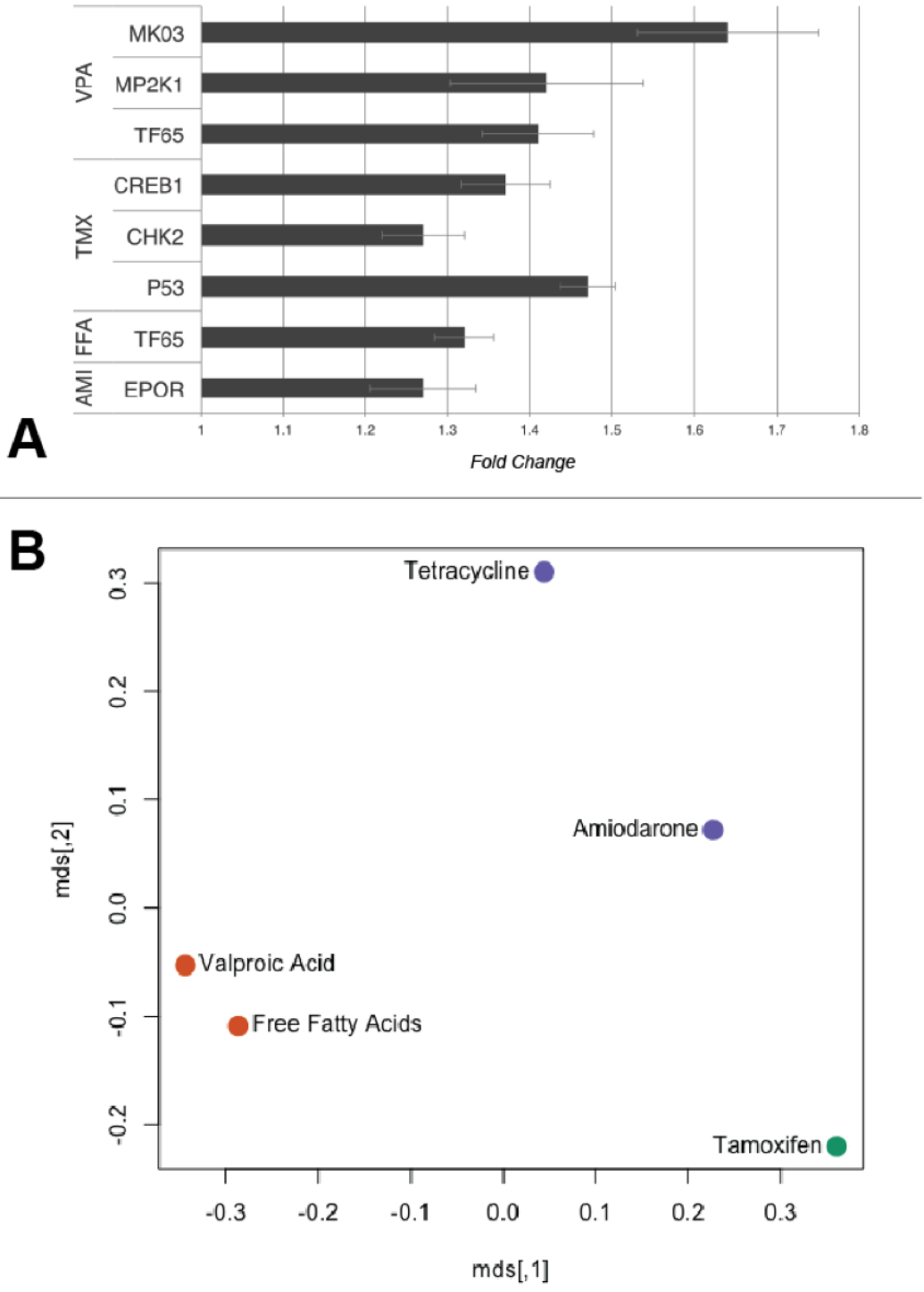
Workflow of experiments and computational analysis. Primary human hepatocytes were and cultured. NAFLD induction was performed using Amiodarone, Free Fatty Acids, Tetracycline, Tamoxifen or Valproic Acid. The cultures were assessed for viability and lipid droplet formation. Cells from all were treated with a combination of stimuli and lysed. xMAP assay was performed to quantify phosphorylated proteins and the results were analyzed to construct signaling pathways.

Figure P7-2



High content screening under 20x optical magnification: Intracellular lipid droplets stained with Nile Red with Hoechst 33342 counterstaining of the cell nuclei. The 8 images correspond to primary human hepatocyte culture treated with AMI (Amiodarone), FFA (Free Fatty Acids), TMX (Tamoxifen), VPA (Valproic Acid) and TET (Tetracycline) along with 3 diluent controls: Cell culture medium, DMSO and Ethanol.

Figure P7-3

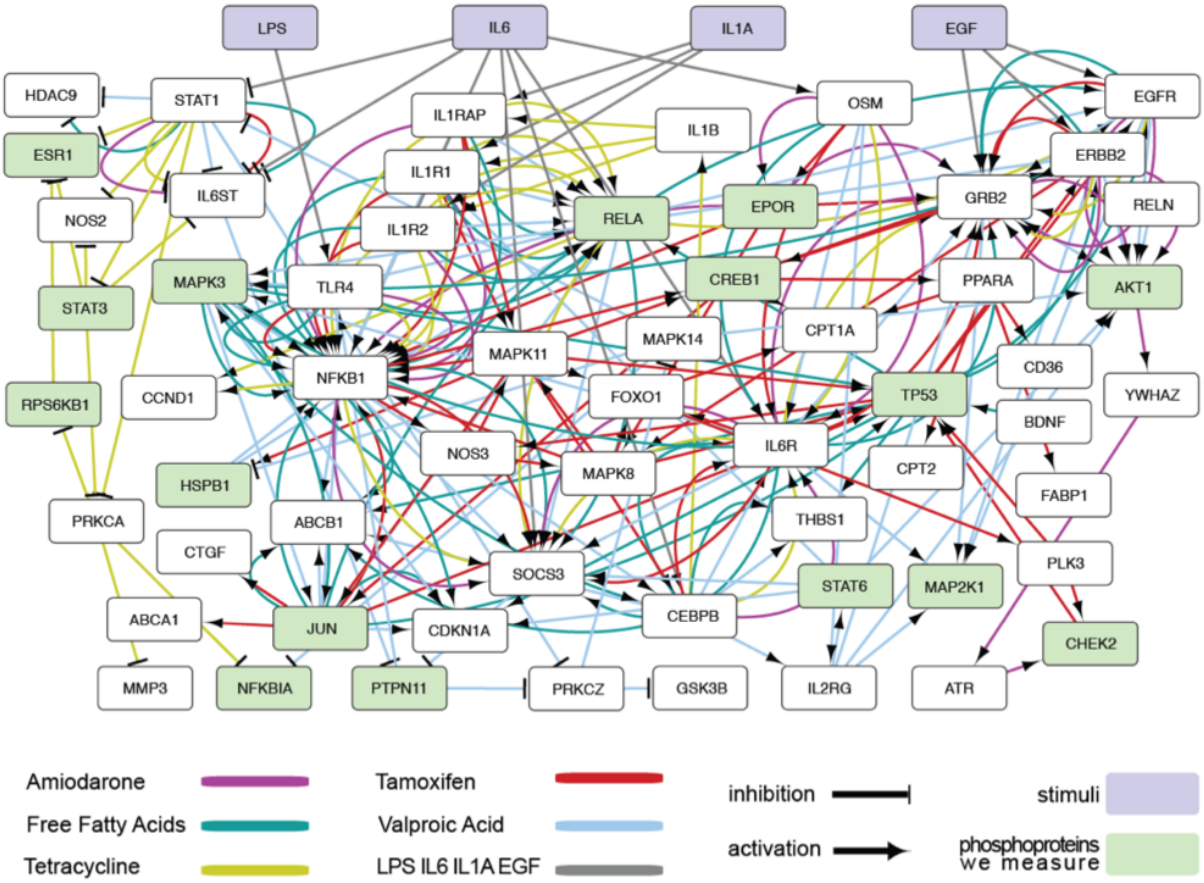


A: Phosphoproteomic results. Fold increase was calculated by dividing the average across the four donors treated cells by the respective average of control. Cells were treated with the NAFLD-inducing compounds AMI (Amiodarone), FFA (Free Fatty

Acids), TMX (Tamoxifen), VPA (Valproic Acid). The Uniprot names of phosphoproteins measured are shown on the table.

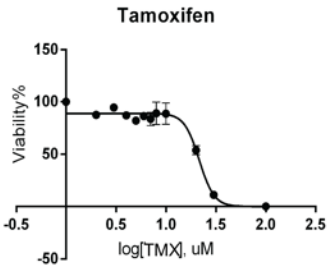
B: Comparison of NAFLD in vitro models. All experimental data available for each of the five compounds were used as a vector and multidimensional scaling (mds) was applied to create a 2D plot for visualization purposes.

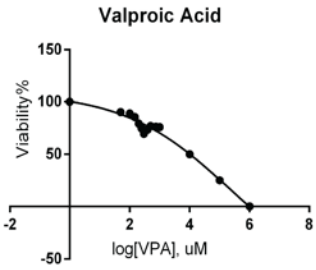
Figure P7-4

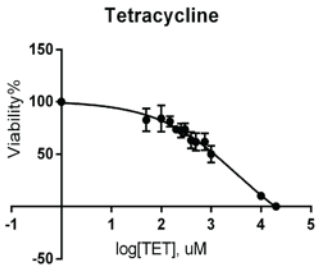


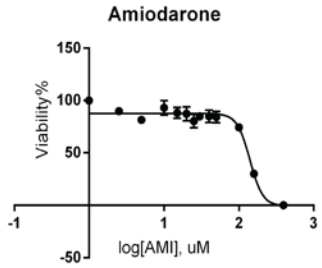
Signaling networks of Amiodarone, Free Fatty Acids, Tetracycline, Tamoxifen and Valproic Acid, presented on the same graph with a different edge color for each compound. Light purple is used for stimuli nodes and light green is used for phosphoprotein nodes we measured.

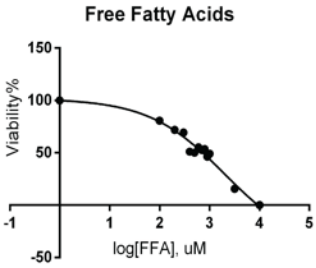
Figure P7-S1











Compound	IC_{50} (µM)	IC_{10} (µM)
Amiodarone	140 ± 1.0471	90.56 ± 1.099
Tamoxifen	21.53 ± 1.041	17.71 ± 1.488
Tetracycline	348.2 ± 2.628	150.3 ± 1.483
Valproic Acid	24498.2 ± 8.388	6200.8 ± 2.109
Free Fatty Acids	2027 ± 1.906	591.07 ± 1,389

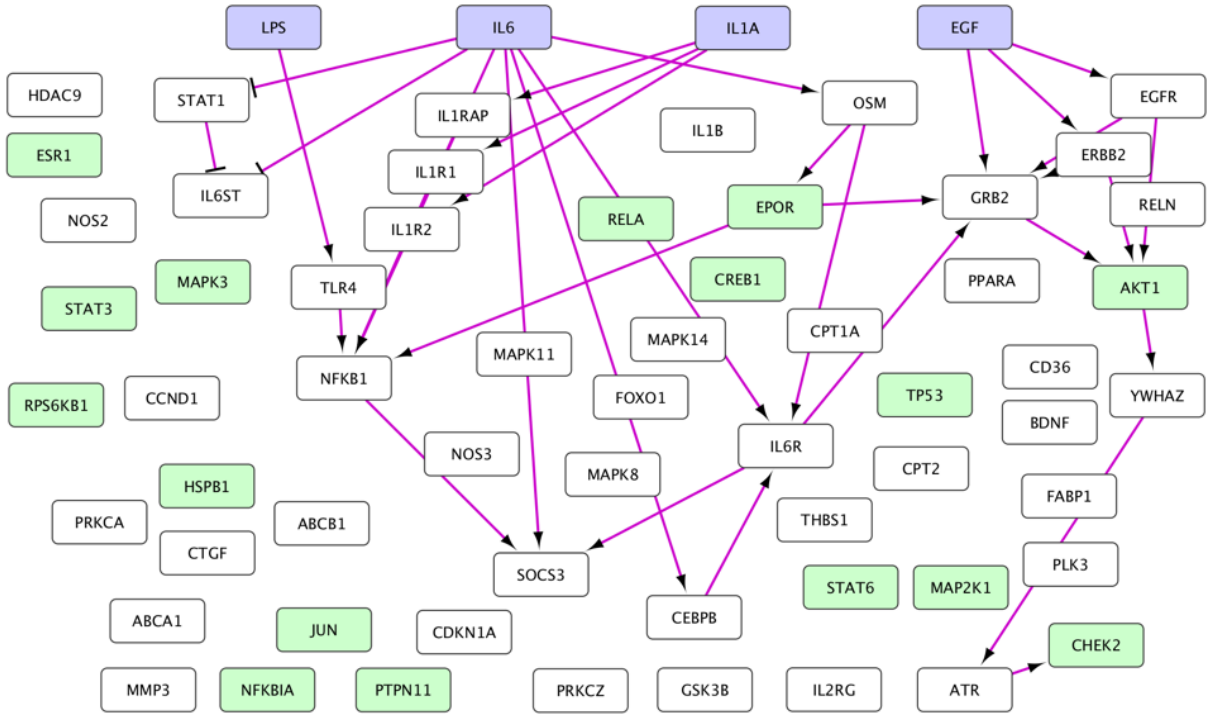
Values IC_{50} (±SEM), IC_{10} (±SEM)

Dose response curves and table summarizing the IC10 concentrations for the compounds Amiodarone (AMI), Tamoxifen (TMX), Tetracycline (TET), Valproic Acid (VPA) and Free Fatty Acids (FFA). Primary Human Hepatocytes were used in cell culture and Resazurin assay was used to assess viability, which was expressed as a percentage of treated to untreated cells.

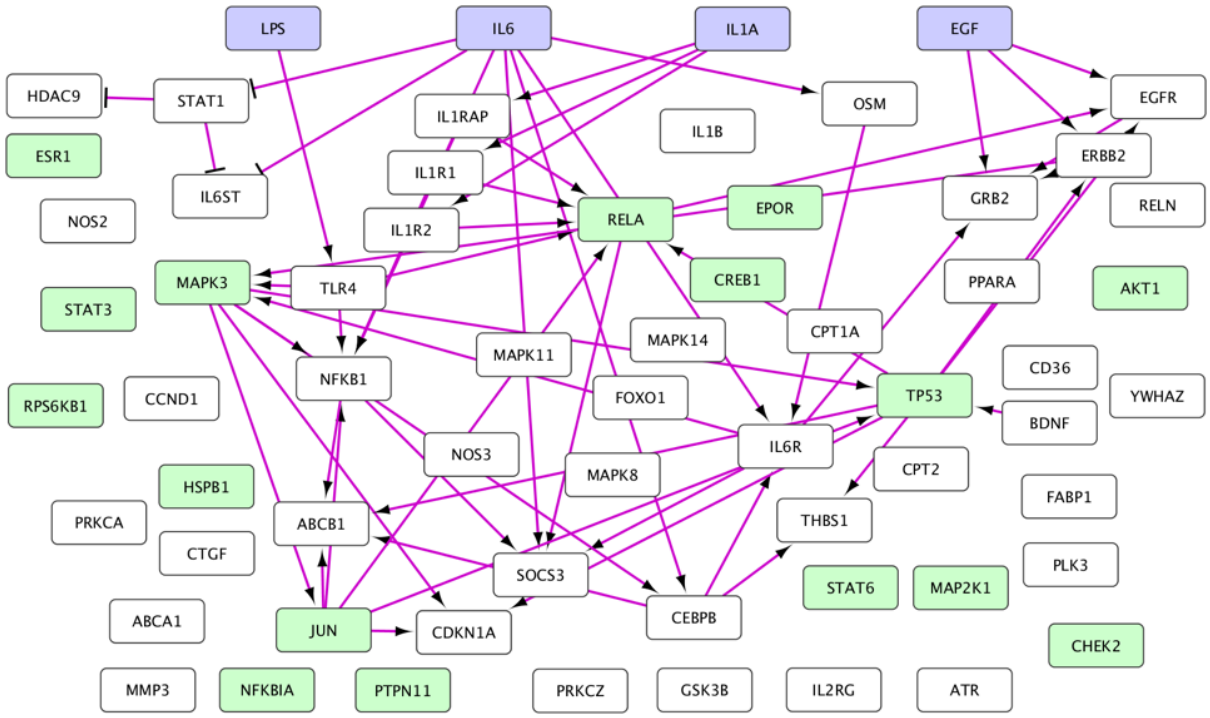
Figure P7-S2

Signaling networks visualizing the phosphoproteomic mechanism of each NAFLD-induction in vitro model. Networks A-E correspond to Amiodarone, Free Fatty Acids, Tetracycline, Tamoxifen and Valproic Acid. Light green nodes are the proteins we measure for phosphorylation activity.

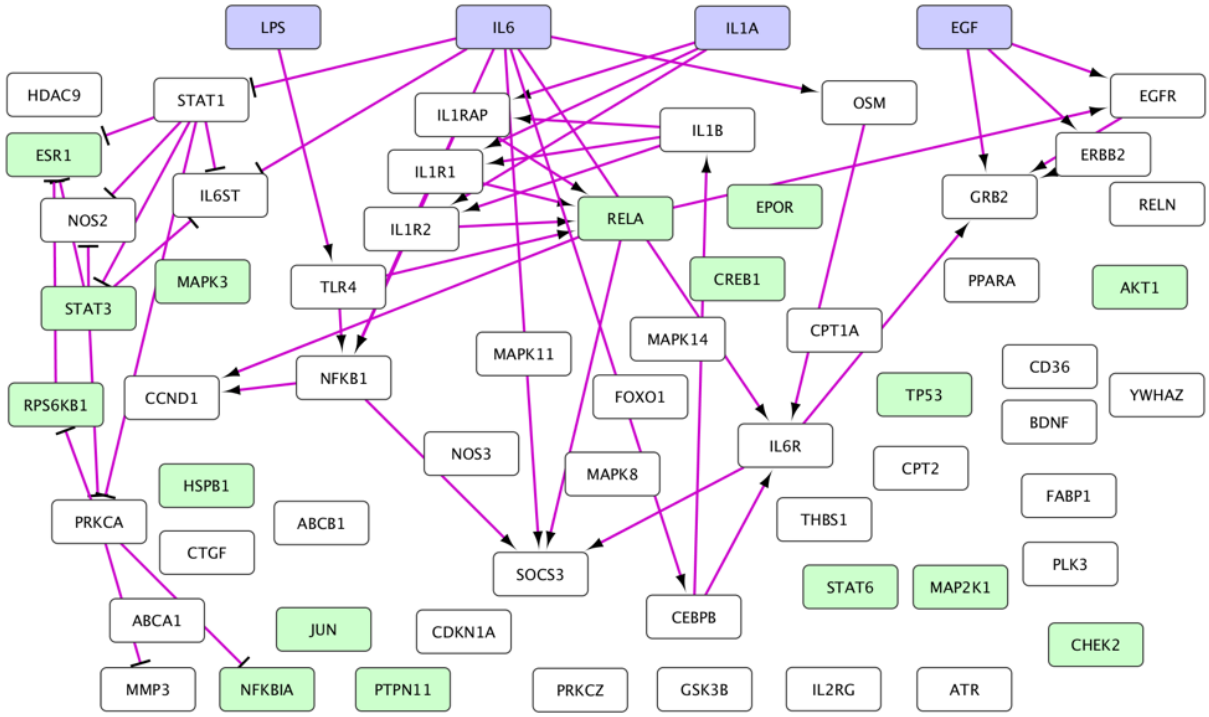
A: Amiodarone



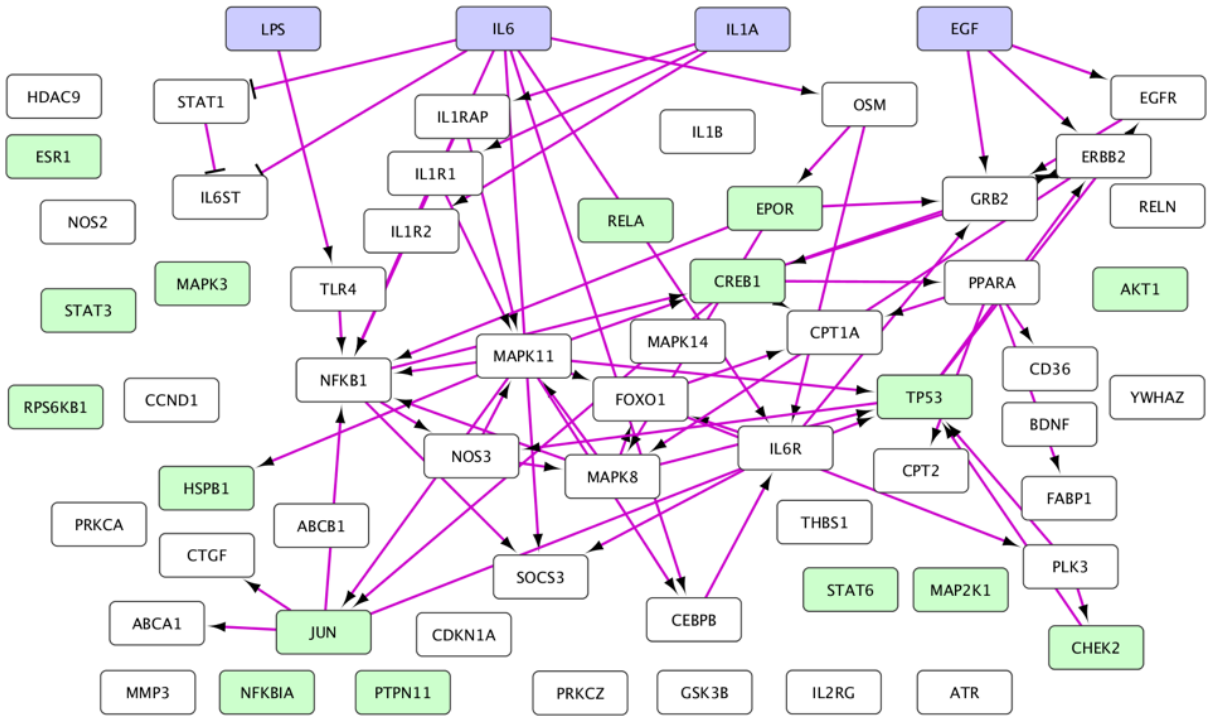
B: Free Fatty Acids



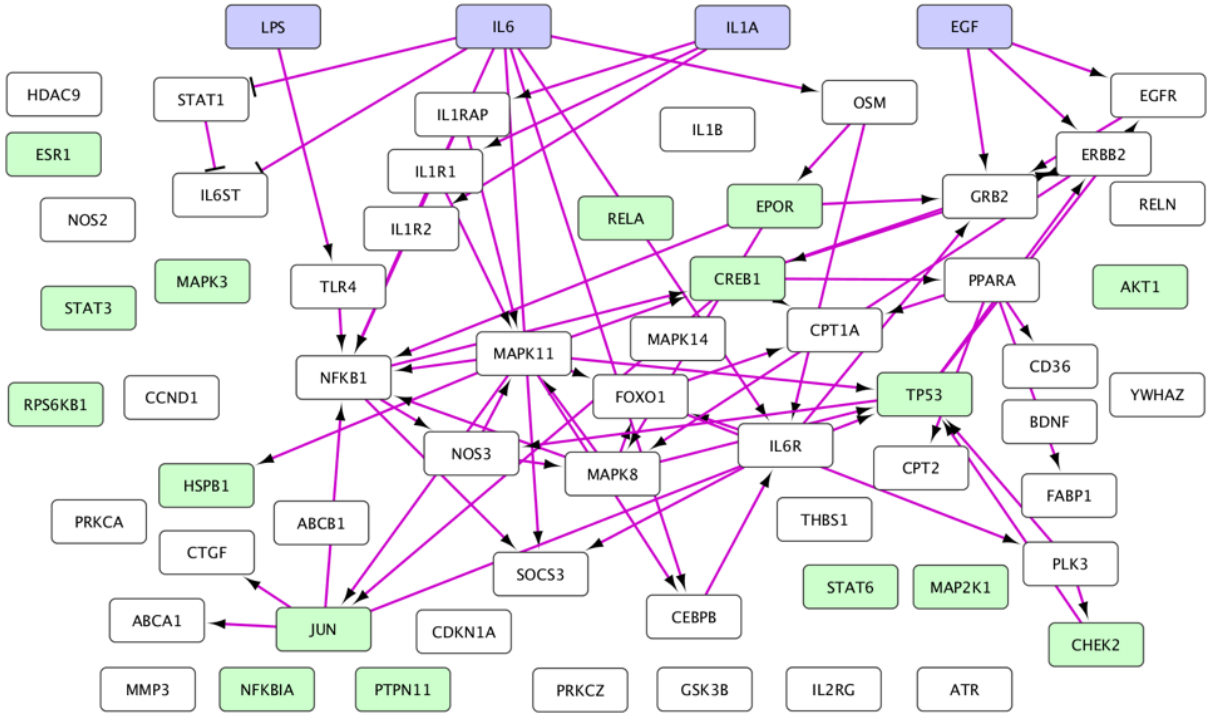
C: Tetracycline



D: Tamoxifen



E: Valproic Acid



Πίνακες | Tables

Table A4-1: The samples transferred to MESBL for potential metabolomic analysis

SMS or VIAL TAG	donor	cells/tissue	sample type	sample	MtOH	ribitol	glucose	H2O	total	used	empty tube	tube after	metabolites	MeOx	MSTFA	split ratio	deriv	217	319				notes
				ul	ul	ug	ug	ul	ul	ul	mg	mg	mg	ul	ul		hrs	ribitol	ribitol	Glu1	Glu2	K	
IL1B_CELLS_17h	94	chondrocytes 24h 50ng/ml IL1B	whole cell extract	0	2000	0.5	1	2000	4000	~3500	13007.7	13012.2	4.5	150	300	1:40	17	2869	935	3437	632	5.44	
IL1B_CELLS_18h																	18	3096	933	3219	627	5.13	
IL1B_CELLS_19h																	19	3189	923	3067	684	4.48	
CTRL_CELLS_12h	94	chondrocytes untreated	whole cell extract	0	2000	0.5	1	2000	4000	~3500	13094.0	13097.8	3.8	150	300	1:40	12	2866	933	3210	577	5.56	
CTRL_CELLS_13h																	13	2991	890	3295	582	5.66	
CTRL_CELLS_14h																	14	3167	1046	3631	596	6.09	
IL1B sup	94	chondrocytes 24h 50ng/ml IL1B	supernatant	1000	3000	10	20	3000	7000	1500	13034.9	13039.0	4.1										
IL1B sup B										1500	13030.7	13034.1	3.4										
CTRL sup	94	chondrocytes untreated	supernatant	1000	3000	10	20	3000	7000	1500	13078.1	13082.5	4.4										
CTRL sup B										1500	12927.9	12931.1	3.2										
IL1B wash	94	chondrocytes 24h 50ng/ml IL1B	PBS wash	1000	3000	10	20	3000	7000	1500	13069.2	13070.9	1.7										
IL1B wash B										1500	13068.3	13070.4	2.1										
CTRL wash	94	chondrocytes untreated	PBS wash	1000	3000	10	20	3000	7000	1500	13156.5	13157.4	0.9										
CTRL wash B										1500	13052.6	13056.2	3.6										
P94_1	94	blood	plasma	100	300	0.1	0.2	300	700	~500	12955.5	12957.1	1.6										
P94_10h	94	blood	plasma	100	300	0.1	0.2	300	700	~500	13013.9	13015.2	1.3	150	300	1:40	10						
P94_11h																	11						
P94_12h																	12						
P68_16h	68	blood	plasma	100	300	0.1	0.2	300	700	~500	13081.5	13085.6	4.1	150	300	1:40	16						hemolysis
P68_17h																	17						
P68_18h																	18						
P68_2	68	blood	plasma	100	300	0.1	0.2	300	700	~500	13021.8	13025.2	3.4										hemolysis
SF94_1	94	OA synovial fluid	synovial fluid	100	300	0.1	0.2	300	700	~300	13025.2	13025.1	-0.1										
SF94_100_9h	94	OA synovial fluid	synovial fluid	100	300	0.1	0.2	300	700	~300	13174.2	13174.2	0.0	150	300	1:40	9	425	115	1084	257	4.22	
SF94_100_10h																	10	441	119	1226	205	5.98	
SF94_200_13h	94	OA synovial fluid	synovial fluid	200	300	0.2	0.4	300	800	~300	12924.2	12925.4	1.2	150	300	1:40	13	812	222	2051	465	4.41	
SF94_200_14h																	14	750	181	2106	383	5.50	
X	68	OA synovial fluid	synovial fluid	100	300	0.1	0.2	300	700	X	12960.5	X	X	X	X								not enough
SF68_2	68	OA synovial fluid	synovial fluid	100	300	0.1	0.2	300	700	~200	12776.3	12775.8	-0.5										SF68_1+2
SF68_200_17h	68	OA synovial fluid	synovial fluid	200	300	0.2	0.4	300	800	~300	13125.7	13125.2	-0.5	150	300	1:40	17	415	117	1379	248	5.56	
SF68_200_18h																	18	356	117	1518	258	5.88	
SF_68_100.200_8h	68	OA synovial fluid	synovial fluid	100	500	10	20	500	1100					100	200	1:40	8	3420	945	5197	798	6.51	run by Kate
SF_68_100.200_9h																	9	3581	996	4182	828	5.05	
SF_68_100.200_10h																	10	3597	974	4940	755	6.54	
SF_94_1h	94	OA synovial fluid	synovial fluid	100	500	10	20	500	1100					50	100	1:40	1	9491	2844	11354	2201	5.16	run by Kate
SF_94_2h																	2	11090	3022	13325	2501	5.33	
SF_94_6h																	6	10191	2999	11757	2488	4.73	
SF_94_7h																	7	11036	3077	12740	2830	4.50	
SF_94_9h																	9	5122	1481	9747	1863	5.23	
SF_94_10h																	10	316	113	3955	819	4.83	

Table A4-2

LEGEND													
Unknown peak with matching ions		Ions filled with color when they are the same with either Christoniki's or Harin's lists. QuanIon filled with color when matching my Ion1.											yellow: used that as QuanIon in Method
Quality Control (trailing tail, derivatization agents, IS)													analyzed: sf_94_7h
Known Metabolite Derivative													

Name / Metabolite Derivative	RT	QuanIon	Ion 1	Ion 2	Ion 3	Ion 4	Ion 5	Ion 6	Ion 7	Ion 8	Cat	Type	Comments
dm_1	4.085	144.8	144.8	114.9								unknown	
dm_2	4.541	68.8	68.8	82.9								unknown	
dm_3	4.876	54.9	54.9	96.9								unknown	
m_3 / C_008	4.992	183.9	183.9	133.8								unknown	same ions as t=5.558 (MSTFA)
dm_4	5.082	155.0	155.0									unknown	
trailing tale	5.314	128.0	74.9	128.0	102.9	99.9	145.9	83.8			1	trailing	
MSTFA	5.558	183.9	183.9	77.0	133.8	213.9					1	derivatization	
kwn_100	6.010	114.9	114.9	149.0	129.9							unknown	same ions as t=7.198 (Lactate), matches kwn (115, 130), 46.8% NIST:50401
dm_5	6.513	355.0	355.0	267.0	356.1							unknown	
K_6	6.745	145.9	145.9	133.0	249.0	118.9						unknown	133 ion same as t=7.198 (Lactate)
Lactate 2 TMS	7.198	117.0	117.0	190.9	218.8	59.0	133.0	131.0	175.0	381.0	1	known	
P0743 / Unknown 97	7.470	122.8	122.8	92.9	124.8	94.9					10	unknown	
Alanine 2 TMS	8.037	116.0	116.0	190.0							3	known	
P0901 / Unknown 125	9.068	203.9	203.9	149.0							11	unknown	double peak
dm_6	9.097	131.0	131.0									unknown	double peak
Valine 1 TMS	9.377	71.9	71.9	54.9	155.9						3	known	
Pyruvate 2 TMS	9.605	173.9	173.9	89.0	59.0	98.9	158.0	114.9			1	known	
2-hydroxybutanoic Acid	10.182	190.9 + 117.0	190.9	117.0	232.8						1	known	
P1091 / Unknown 99	10.945	220.0	133.0	220.0	234.9	204.9	190.0	160.0	59.0	89.0	1	unknown	
P1099 / f_12	11.022	174.0	174.0								10	unknown	very close to previous peak
Valine 2 TMS	11.163	218.0	144.0	218.0							3	known	
Ethanolamine 3 TMS	11.357	174.0	85.9	74.9	170.0	174.0	341.0	429.0			3	known	341 is from an artifact peak coming next in Harin's list
dm_7	11.949	212.0	212.0	521.8	122.0							unknown	
Glycerol 3 TMS	12.194	204.8	117.0	204.8	133.0	103.0	175.0	217.9	59.0	191.0	1	known	
P1229 / Unknown 126	12.326	228.0	228.0	110.0	183.9	77.0	133.8	285.0			10	unknown	
Leucine 2 TMS	12.815	158.0	158.0	102.0	232.1	218.0					3	known	
P1295	13.138	70.0	70.0	102.9	171.9	144.0					3	unknown	
Isoleucine 2 TMS	13.460	158.0	158.0	258.0	218.0						3	known	
Glycine 3 TMS	13.614	174.1	174.1	248.1	58.9	85.9	99.9	133.0	275.9		3	known	
dm_8	14.132	216.0	200.0	216.0	89.0	188.8						unknown	Harin's P1391 has the ions 89 + 189
Serine 2 TMS	14.287	116.0	116.0	131.9	143.9						3	known	
Proline 1 TMS	14.479	142.0 + 216.0	142.0								1	known	216 is sharp with count 19
Phosphoric Acid 4 TMS	14.700	387.0	299.3	314.0	283.4	387.0	211.2	225.3			1	known	
Threonine 2 TMS	14.881	130.0 + 219.0	130.0	117.0	219.0						3	known	
Urea	15.103	189.1	189.1	173.2	205.0	351.2	131.2	99.2	59.0		1	known	
dm_9	15.466	292.1	292.1	202.9	220.0							unknown	
P1552 / f_105	15.543	174.1	74.0	117.0	133.8	183.9	284.9	174.1			3	unknown	
dm_10	15.698	281.0	281.0	415.0	327.0							unknown	
dm_11	15.801	199.0	199.0									unknown	carry over 74 and 117
Threonine 3 TMS	15.904	217.9	217.9	291.2							3	known	Harin quanion: 291.2 + 217.9
Fumarate 2 TMS	16.059	245.0	245.0								1	known	
Succinate 2 TMS	16.240	246.8	215.0	246.8	128.9						1	known	
dm_12	16.511	117.0	117.0	89.0	59.0	100.8						known artifact	known artifact (kwn_31)
B-alanine 3 TMS	16.925	248.0	248.0	174.0							3	known	
f_75	17.234	228.0	228.0	217.0	110.0	133.8	146.9	142.9	183.9			unknown	
3,4-Dihydroxybutyrate 3 TMS	17.480	188.8	188.8	232.9	202.8	321.0					1	known	
P1753	17.570	174.0 + 248.1	174.0	248.1	373.1						3	unknown	
Erythritol 4 TMS	17.815	217.0	217.0	228.0	117.0	292.9	320.1	190.9	188.8	102.8	1	known	
dm_13	17.957	146.0	146.0	89.9	130.0							unknown	
P1863? / Unknown 108	18.667	116.8 + 229.1	229.1	116.8							3	unknown	is it Harin's P1863?? / first of sth like a double peak
K_136	18.719	116.0	116.0	190.0	159.9						3	unknown	
Isoleucine 3 TMS	19.378	127.0	231.9	70.0	127.0	99.9	162.9	334.0	262.0		3	known	
dm_14	19.533	355.0	355.0	401.0	327.0	385.1						unknown	
dm_15	19.674	292.0	292.0	217.0	245.0	265.0						unknown	
a_34 / f_79	19.765	217.9	292.0	320.0	217.9	248.1	174.0					unknown	

Hydroxyproline 2 TMS	20.229	259.9	67.9	158.0							3	known	does not open with supposed Quanlon 260 , but matches all ions and time.
Threonic Acid 4 TMS	20.449	319.0	292.0	217.0	219.9	245.0	204.8	117.0	319.0	408.8	1	known	supposed Quanlon is not good
dm_16	20.552	56.0	56.0	103.9	60.9	141.0						unknown	
a_39 / x_2 / U_032	21.068	263.2	263.2									unknown	
dm_17	21.107	221.0	221.0									unknown	
Methionine 2 TMS	21.172	176.0	176.0	128.0							3	known	
x_3 (carry over)	21.197	71.9	71.9	210.9								unknown	
Cysteine NNO	21.327	220.0	220.0								3	known	
dm_18	21.417	177.0	177.0	207.0								unknown	
x_4	21.753	114.9	114.9	329.1	314.1							unknown	(creatinine 3 TMS)
Ribitol (217)	22.139	217.0	217.0	319.0	117.1	129.1	332.0	205.0	243.0	189.1	1	IS	
Ribitol (319)	22.139	319.0	217.0	319.0	117.1	129.1	332.0	205.0	243.0	189.1	1	IS	
Xylitol (NQ)	22.243	217.0	217.0	319.0	117.1	129.1	332.0	205.0			1	unknown	same ions as Ribitol. Christoniki: "Xylitol (NQ) dn iparxei"
N-acetoglutamic acid 2 TMS	22.346	174.1	174.1	83.9	229.2	580.9					1	known	Christoniki's name: N-acetoglutamate
K_47	22.436	188.0	188.0	217.0	99.8							unknown	third of three peaks after ribitol, having 217 in their ions
Glutamate 3 TMS	22.694	246.1	246.1	128.0	84.0	348.1	363.0	56.0	156.0	230.0	3	known	Ions 156, 230 higher at t = 23.069 (Pyroglutamate 2 TMS)
P2284	22.862	173.0	173.0	355.1	428.9	262.8	341.0	147.0				unknown	
Pyroglutamate 2 TMS	23.069	155.9	155.9	230.0	258.0	346.1	214.1	147.0			3	known	Harin quanlon: 230.0 + 155.9 / Ion 147 higher at t = 22.139 (Ribitol)
Rhamnose MeOX2 4 TMS	23.237	277.0	277.0	116.8	159.9						2	known	not good peaks
K_52	23.689	316.1	316.1	171.0	417.9	290.0						unknown	
Phenylalanine 2 TMS	23.793	217.9	217.9	192.1	99.9	91.0	266.1				3	known	
P2427 / Unknown 120	24.050	292.0	292.0	120.0	217.0	333.0	197.9				1	unknown	similar ions to t = 24.295 / found at expected Harin time +200. Christoniki ok
dm_19	24.217	221.0	221.0	368.8								unknown	
P2445 / f_21	24.295	292.0	292.0	217.0	333.0								similar ions to t = 24.050 / found at expected Harin time +200. Christoniki ok
P2433	24.385	142.0	142.0	185.8	216.0	288.0						unknown	
Asparagine 3 TMS	24.824	231.0 +258.0	115.9	188.0	131.9	231.0						known	258 is double peak and Quanlon is almost double peak
dm_20	24.953	132.8	132.8	102.8	217.0							unknown	
Ornithine 4 TMS	25.147	142.0	142.0	174.0	420.2	200.0	216.0	258.1	214.0		3	known	
P25541	25.675	210.9	210.9	304.0								unknown	304 has a smaller but much better peak
Fructose MeOX 1	25.997	217.0	217.0	103.0	277.0	306.8	133.0				2	known	Harin's Quanlon: 307 (also good)
Sorbitol	26.063	319.0	319.0	129.1	205.0	157.0	159.9				1	known	also has 217.0 from previous peak and there is a peak following with the same ions
Fructose MeOX 2	26.331	306.8	217.0	103.0	277.0	306.8	133.0				2	known	also opens with 217 as Fructose 1
Glucose MeOX 1	26.526	319.2	319.2								2	known	
[U-13C] Glucose MeOX 1	26.526	323.0	323.0								2	IS	
Glucopyranose MeOX 1	26.653	204.0	204.0	191.1							2	known	
Citrate	26.783	273.1	273.1	347.0	375.0						1	known	this peak is hidden in the following peak
Glucose MeOX 2	26.796	319.2	319.2								2	known	
[U-13C] Glucose MeOX 2	26.796	323.0	323.0								2	IS	
Glutamine 3 TMS	27.031	156.0	156.0	245.0	229.1	347.0	203.0					known	
Lysine 4 TMS	27.159	174.1	156.1	174.1	230.0	128.2	329.2	434.1			3	known	
unknown K_62 / f_94	27.315	174.1	174.1	200.0								unknown	
Galacturonic acid MeOX 1	27.881	333.1	333.1	305.0	159.9							known	double peak
dm_21	27.945	492.2	189.0	157.0	200.9	273.1	492.2					unknown	opens nicely with 492.2
Gluconate 6 TMS	28.048	333.1	292.1	333.1	305.0								double peak
chiro- or scyllo- inositol	28.280	318.1	318.1	305.1							1	unknown	small peak on the hill of the next big peak
Glucopyranose MeOX 2	28.345	204.0	204.0	191.1	217.1						2	known	
dm_22	28.537	205.0	205.0	189.1	272.9							unknown	a big peak, opens with 205,6,7,8 but it's a bit far to be the [U-13C] glucopyranose
P2893	28.939	393.1	393.1	332.9							1	unknown	a bit rough / Christoniki has different ions but the same name (P2893)
myo-inositol 6 TMS	29.093	305.2	305.2	318.2	217.2	191.1	432.1	507.3	265.2	291.2	1	known	
P2922	29.223	361.0	217.0	361.0	157.0	129.0	244.0	332.1	149.1	133.0		unknown	only in Harin's list / 361 opens better indeed
f_28	29.428	284.0	284.0	299.2								unknown	
P2957	29.571	204.1	204.1	319.1	217.0	157.0	220.0	149.2	129.1			unknown	
Tyrosine 3 TMS	29.637	217.9	217.9	179.0	280.1	354.2	382.2	99.9				known	flat peak
dm_23	29.765	369.0	369.0	295.0	221.0	355.0						unknown	221 and 355 from previous peak
P2997	29.957	361.0	157.0	217.0	189.0	129.0	102.9	319.0	361.0	133.0		unknown	
dm_24	30.063	186.9	186.9	508.2	262.0							unknown	a bit noisy
dm_25	30.243	293.1	293.1	237.0	422.6							unknown	
Tyrosine 2 TMS	30.411	179.0	179.0	208.1	310.1						3	known	
P3046	30.450	203.9	203.9	319.0	217.0							unknown	
dm_26	30.798	221.0	221.0									unknown	221 is the only ion that separates it from the next peak. There are also 281, 355, 429

dm_27	30.824	428.8	428.8	355.0								unknown	
31248_Hexadecanoic acid_1 TMS	30.889	117.0	117.0	313.2	74.9	129.0	185.0	55.0	57.0		11	known	same ions at the next peak (30.979) which is almost double with this
dm_28	31.443	266.1	266.1	237.0	327.0							unknown	179 and 369.1 peaking a bit behind
P3188	31.870	327.0	327.0	74.9	92.9	96.9	82.9	110.9				unknown	
dm_29	31.973	149.0	149.0									unknown	
kwn_109	32.244	221.0	221.0	368.9	295.0	207.0						unknown	281 falls a bit behind
P3263 / Unknown 115	32.618	117.0	128.9	117.0	327.2						1	unknown	Christoniki ion: 128.2
dm_30	32.851	443.1	443.1									unknown	
dm_31	32.915	217.1	217.1	189.0								unknown	205.2 falls a bit behind
P3307	33.083	429.0	429.0	503.0	221.0	281.0	299.2					unknown	P3307 has quanion 418, which is completely absent but timing is perfect
dm_32	33.315	202.0	202.0	547.1								unknown	319.1 a bit behind, 217 not so good
dm_33	33.651	547.4	547.4	551.4	83.8	217.1						unknown	
dm_34	34.064	149.0	149.0									unknown	
P3423	34.218	441.2	441.2	456.2	382.2	367.2						unknown	P3423 has quanion 361.1, which is completely absent but timing is perfect
dm_35	34.284	117.0	117.0	129.0	341.2	74.9	185.0	355.9	55.0	297.1			a smaller peak is following (34.401) with the same ions
Linoleic acid	34.580	337.2	74.9	337.2	67.0	81.0	95.0	107.0	262.2	245.2		known	
dm_36	35.216	355.0	355.0										bad peak
dm_37	35.616	202.0	202.0										
dm_38	36.158	202.0	202.0	369.2									
dm_39	37.217	165.0	165.0										
dm_40	37.281	429.0	429.0										very bad peak
Eicosanoic acid 1 TMS	37.436	117.0	369.2	117.0	129.0								117 is very weak for Quanlon
K_105 (Arachidonic acid)	37.720	117.0	78.9	54.8	67.0	90.9	117.0	175.0	120.0	77.0		unknown	
dm_41	37.771	202.0	202.0										
Sucrose	38.403	437.1	361.0	217.0	437.1							known	217 not so good, 437.1 is really bad to be Quanlon
dm_42	38.920	259.1	259.1	204.0	217.0	361.0							
a_78	39.423	285.2	285.2									unknown	not sure
P3960	39.578	371.2	371.2	203.0								unknown	
P3967	39.694	480.4	361.0	480.4								unknown	480.4 is a bit weak for Quanlon
P3991	39.888	239.1	239.1									unknown	
P4232	41.991	203.9	203.9	361.0	217.0							unknown	
b_35 / f_50 / d_68 / A_227 (u_086)	42.314	399.2	399.2	203.0								unknown	
dm_43	47.876	179.0	179.0	105.0	91.0	103.0	77.0	217.1	171.0	157.0			
Tocopherol 1 TMS	48.712	502.6	502.6	237.0							1	known	
Cholesterol	48.829	368.3	368.3	129.1	353.3	458.2	329.4	255.3			1	known	

			chromatograph	SF_94_3h	SF_94_4h		SF94_100_9h	SF94_100_10h		SF94_100_13h	SF94_100_14h						SF68_200_17h	SF68_200_18h	
			Donor ->	94	94		94	94		94	94		68	68	68		68	68	
			sample (ul) ->	100	100		100	100		200	200		100	100	100		200	200	
			MtOH (ul) ->	500	500		300	300		300	300		500	500	500		300	300	
			ribitol (ug) ->	10	10		0.1	0.1		0.2	0.2		10	10	10		0.2	0.2	
			glucose (ug) ->	20	20		0.2	0.2		0.4	0.4		20	20	20		0.4	0.4	
			MeOX (ul) ->	50	50		150	150		150	150		100	100	100		150	150	
			MSTFA (ul) ->	100	100		300	300		300	300		200	200	200		300	300	
			deriv (hrs) ->	6	7		9	10		13	14		8	9	10		17	18	
Peak #	RT	Peak Name	Quan Ions		REF	AVG			AVG			AVG				AVG			AVG	SUM AVG
88	26.520	Glucose MeOX 1	319.2	107.27	107.74	107.51	1306.52	1110.15	1208.33	1071.79	1410.16	1240.97	131.85	121.48	131.82	128.38	1288.47	1265.97	1277.22	3962.42
35	15.027	Urea	189.1	79.67	80.03	79.85	841.74	729.17	785.46	666.62	871.64	769.13	61.37	58.74	63.28	61.13	556.78	538.04	547.41	2242.97
92	26.797	Glucose MeOX 2	319.2	20.53	20.43	20.48	247.45	206.73	227.09	192.92	262.98	227.95	25.18	23.99	26.83	25.33	237.55	234.61	236.08	736.93
11	7.166	Lactate 2 TMS	117	28.40	28.40	28.40	253.27	222.42	237.84	186.21	251.41	218.81	15.56	14.44	15.43	15.14	128.91	117.90	123.40	623.60
33	14.701	Phosphoric Acid 4 TMS	299.3	25.87	26.17	26.02	254.38	199.35	226.86	97.91	141.10	119.50	29.45	26.36	28.93	28.24	122.04	128.32	125.18	525.81
25	12.315	P1229 / Unknown 126	228	0.54	0.54	0.54	145.09	118.98	132.03	64.52	85.49	75.00	1.38	1.37	1.50	1.42	113.51	117.12	115.31	324.31
91	26.773	Citrate	273.1	7.87	7.98	7.93	48.23	43.56	45.90	40.94	56.25	48.59	7.96	7.03	8.00	7.66	45.17	41.30	43.24	153.32
18	10.168	2-hydroxybutanoic Acid	190.9+117.0	6.92	6.98	6.95	66.81	55.89	61.35	51.18	68.13	59.66	1.89	1.78	2.06	1.91	16.41	15.98	16.19	146.06
29	13.612	Glycine 3 TMS	174.1	4.58	4.28	4.43	37.69	32.52	35.11	28.99	38.45	33.72	5.51	5.15	5.58	5.41	43.90	41.23	42.57	121.23
73	23.064	Pyroglutamate 2 TMS	155.9	4.14	4.17	4.15	28.54	26.57	27.56	24.12	32.69	28.41	4.78	4.45	4.94	4.73	39.28	42.79	41.04	105.88
107	29.569	P2957	204.1	2.09	2.35	2.22	26.34	23.53	24.93	28.56	39.18	33.87	3.96	3.86	4.16	3.99	38.48	34.83	36.65	101.67
20	11.019	P1099 / f_12	174	0.86	0.89	0.87	43.74	38.73	41.24	20.38	24.59	22.49	2.42	2.27	2.56	2.42	35.56	25.22	30.39	97.40
19	10.924	P1091 / Unknown 99	220	1.31	1.51	1.41	24.48	18.89	21.69	30.95	39.83	35.39	9.13	8.49	9.94	9.19	24.96	23.99	24.47	92.15
12	7.470	P0743 / Unknown 97	122.8	0.08	0.08	0.08	31.67	27.67	29.67	14.38	19.15	16.77	0.26	0.25	0.24	0.25	24.22	25.01	24.62	71.38
8	6.010	kwn_100	114.9	0.16	0.10	0.13	30.32	24.68	27.50	12.99	19.95	16.47	0.44	0.40	0.43	0.42	23.87	24.01	23.94	68.47
101	28.345	Glucopyranose MeOX 2	204	2.05	2.16	2.10	40.29	34.61	37.45	15.48	20.27	17.87	2.28	2.07	2.26	2.20	0.97	0.82	0.90	60.53
86	26.052	Sorbitol	319	2.05	2.11	2.08	19.88	17.38	18.63	16.76	22.13	19.45	1.95	1.70	2.00	1.88	17.07	15.84	16.45	58.49
104	29.100	myo-inositol 6 TMS	305.2	2.39	2.37	2.38	21.09	17.30	19.20	16.35	22.54	19.44	1.74	1.74	1.82	1.77	14.25	13.77	14.01	56.79
87	26.514	Fructose MeOX 2	306.8	0.08	0.09	0.08	33.75	0.45	17.10	0.64	23.87	12.26	2.87	2.08	2.46	2.47	0.30	35.11	17.70	49.61
16	9.387	Valine 1 TMS	71.9	1.72	1.86	1.79	12.05	11.83	11.94	11.85	16.85	14.35	1.08	1.18	0.84	1.03	9.65	13.72	11.69	40.80
90	26.647	Glucopyranose MeOX 1	204	1.55	1.57	1.56	25.70	21.82	23.76	9.38	14.22	11.80	1.52	1.65	1.83	1.67	1.65	2.15	1.90	40.68
89	26.516	[U-13C] Glucose MeOX 1	323	3.74	4.01	3.88	9.35	9.29	9.32	8.80	12.76	10.78	4.86	3.64	4.59	4.36	10.44	11.33	10.88	39.22
95	27.161	Lysine 4 TMS	174.1	2.99	2.65	2.82	10.23	10.07	10.15	12.02	15.70	13.86	3.36	3.45	3.90	3.57	14.87	2.16	8.52	38.92
1	4.070	dm_1	144.8	0.24	0.19	0.21	15.61	12.75	14.18	7.00	10.18	8.59	0.61	0.54	0.52	0.56	11.63	11.66	11.65	35.19
15	9.082	dm_6	131	1.43	1.61	1.52	13.79	11.55	12.67	10.74	14.80	12.77	0.96	1.00	0.98	0.98	5.89	7.76	6.83	34.76
76	23.788	Phenylalanine 2 TMS	217.9	1.93	1.81	1.87	9.47	8.46	8.97	8.90	12.52	10.71	1.38	1.37	1.38	1.37	8.80	8.87	8.83	31.76
117	30.885	31248_Hexadecanoic acid_1	117	2.94	3.50	3.22	10.68	10.05	10.36	5.46	4.85	5.16	6.72	5.77	6.28	6.25	4.45	7.90	6.17	31.17
24	12.179	Glycerol 3 TMS	204.8	1.01	1.05	1.03	10.41	8.37	9.39	7.64	11.28	9.46	0.88	0.88	0.93	0.90	8.28	8.75	8.52	29.30
85	25.991	Fructose MeOX 1	217	1.60	1.61	1.60	6.91	5.89	6.40	12.45	15.90	14.17	0.52	0.49	1.47	0.83	5.13	5.84	5.48	28.49
17	9.587	Pyruvate 2 TMS	173.9	1.30	1.40	1.35	8.02	6.98	7.50	9.06	12.48	10.77	0.70	0.66	0.62	0.66	5.17	4.35	4.76	25.04
21	11.149	Valine 2 TMS	144	1.38	0.79	1.09	8.92	7.47	8.20	6.80	7.77	7.29	1.87	1.73	1.82	1.81	7.68	4.48	6.08	24.46
13	8.004	Alanine 2 TMS	116	1.10	0.53	0.81	9.10	7.57	8.34	7.46	8.80	8.13	1.48	1.47	1.59	1.51	6.27	3.58	4.92	23.72
130	34.287	dm_35	117	3.04	3.14	3.09	4.53	5.34	4.93	4.75	9.02	6.89	3.70	3.42	3.67	3.59	3.91	6.00	4.95	23.46
22	11.358	Ethanolamine 3 TMS	85.9	0.87	0.71	0.79	5.53	4.64	5.09	6.25	8.28	7.27	0.75	0.72	0.59	0.69	5.55	7.03	6.29	20.13
34	14.889	Threonine 2 TMS	130.0+219.0	0.68	0.77	0.72	5.95	4.72	5.33	5.38	6.76	6.07	0.83	0.86	0.89	0.86	6.38	6.40	6.39	19.37
66	22.135	Ribitol (217)	217	3.62	3.77	3.70	3.91	3.56	3.73	3.78	5.07	4.42	3.55	3.48	3.60	3.54	3.66	3.49	3.58	18.97
31	14.287	Serine 2 TMS	116	0.71	0.68	0.69	5.89	5.11	5.50	4.88	6.92	5.90	0.31	0.63	0.78	0.57	5.75	6.15	5.95	18.61
7	5.506	MSTFA	183.9	0.27	0.25	0.26	7.86	6.36	7.11	2.91	3.80	3.35	0.67	0.65	0.68	0.67	6.43	6.44	6.44	17.83
108	29.638	Tyrosine 3 TMS	217.9	1.67	1.61	1.64	5.35	5.45	5.40	6.05	8.01	7.03	0.76	0.57	0.94	0.76	5.61	0.00	2.81	17.64
94	27.029	Glutamine 3 TMS	156	1.77	1.63	1.70	2.83	2.77	2.80	4.06	6.02	5.04	1.31	1.39	1.60	1.44	4.69	5.16	4.93	15.90
96	27.317	unknown K_62 / f_94	174.1	0.60	0.87	0.73	4.98	2.02	3.50	2.98	0.00	1.49	0.54	0.87	0.54	0.65	5.23	10.28	7.76	14.13
5	5.068	dm_4	155	0.29	0.27	0.28	4.45	4.05	4.25	2.36	3.03	2.70	0.84	0.83	0.92	0.86	4.08	4.08	4.08	12.17
71	22.688	Glutamate 3 TMS	246.1	0.74	0.74	0.74	3.98	3.09	3.54	3.18	4.22	3.70	0.61	0.60	0.67	0.63	4.02	2.92	3.47	12.07
10	6.745	K_6	145.9	0.09	0.00	0.05	7.59	0.00	3.80	2.89	3.40	3.14	0.00	0.00	0.27	0.09	5.24	4.19	4.72	11.79
68	22.135	Xylitol (NQ)	217	0.19	0.18	0.18	3.91	3.58	3.74	2.30	1.96	2.13	3.55	0.17	0.39	1.37	3.11	3.49	3.30	10.73
14	9.036	P0901 / Unknown 125	203.9	0.83	1.05	0.94	3.63	3.29	3.46	1.91	2.72	2.31	1.20	1.25	1.58	1.35	3.95	1.28	2.62	10.67

Table A4-3: Metabolomics analysis of Synovial Fluid samples - Identified Metabolites

26	12.804	Leucine 2 TMS	158	0.70	0.38	0.54	3.26	2.80	3.03	2.88	3.28	3.08	0.89	0.85	1.03	0.92	3.48	2.46	2.97	10.54
114	30.450	P3046	203.9	0.08	0.09	0.08	1.77	1.45	1.61	2.36	3.76	3.06	0.30	0.30	0.38	0.33	4.37	4.28	4.32	9.40
150	48.855	Cholesterol	368.3	4.58	4.43	4.50	0.00	0.00	0.00	0.00	0.00	0.00	4.89	4.09	5.07	4.68	0.00	0.00	0.00	9.19
83	25.143	Ornithine 4 TMS	142	0.52	0.49	0.50	1.96	1.77	1.86	1.89	2.66	2.28	0.84	0.90	0.99	0.91	3.73	2.54	3.13	8.68
45	17.231	f_75	228	0.24	0.25	0.24	2.81	2.64	2.73	1.29	1.76	1.53	0.65	0.61	0.66	0.64	2.20	2.69	2.45	7.58
52	19.378	Isoleucine 3 TMS	231.9	0.14	0.23	0.18	3.63	3.02	3.33	1.74	2.33	2.04	0.07	0.21	0.17	0.15	1.80	1.74	1.77	7.46
129	34.218	P3423	441.2	0.90	0.62	0.76	1.09	1.41	1.25	3.77	5.18	4.47	0.00	0.04	0.08	0.04	1.60	0.00	0.80	7.33
93	26.796	[U-13C] Glucose MeOX 2	323	0.80	0.89	0.85	2.20	1.55	1.87	2.00	2.32	2.16	0.78	0.72	0.70	0.74	1.85	0.88	1.36	6.98
148	47.908	dm_43	179	3.90	3.89	3.90	0.00	0.00	0.00	0.00	0.00	0.00	2.48	1.17	2.38	2.01	0.00	0.00	0.00	5.91
80	24.385	P2433	142	0.30	0.58	0.44	2.66	1.95	2.31	1.22	1.67	1.45	0.22	0.21	0.33	0.25	1.50	1.32	1.41	5.86
48	17.815	Erythritol 4 TMS	217	0.20	0.20	0.20	1.82	1.68	1.75	1.56	2.13	1.84	0.06	0.14	0.17	0.12	1.33	1.56	1.44	5.36
28	13.460	Isoleucine 2 TMS	158	0.32	0.16	0.24	1.59	1.55	1.57	1.49	1.65	1.57	0.37	0.39	0.46	0.40	1.64	1.04	1.34	5.12
67	22.136	Ribitol (319)	319	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	5.00
105	29.223	P2922	361	0.37	0.38	0.38	1.64	1.45	1.55	1.25	1.68	1.47	0.28	0.32	0.35	0.32	1.39	1.00	1.19	4.90
57	20.449	Threonic Acid 4 TMS	292	0.22	0.22	0.22	2.07	1.55	1.81	1.60	2.07	1.83	0.00	0.05	0.05	0.03	0.66	0.64	0.65	4.55
61	21.172	Methionine 2 TMS	176	0.10	0.04	0.07	1.35	1.20	1.28	1.15	1.84	1.50	0.24	0.22	0.24	0.23	1.16	1.46	1.31	4.39
118	31.443	dm_28	266.1	0.26	0.29	0.27	1.55	1.10	1.33	1.24	1.81	1.52	0.15	0.17	0.16	0.16	0.70	0.96	0.83	4.11
40	15.904	Threonine 3 TMS	217.9	0.06	0.05	0.06	1.29	1.14	1.22	0.63	1.16	0.90	0.17	0.15	0.22	0.18	1.83	1.48	1.65	4.00
54	19.674	dm_15	292	0.14	0.14	0.14	1.33	0.99	1.16	1.01	1.40	1.20	0.11	0.09	0.12	0.11	1.05	0.89	0.97	3.59
102	28.537	dm_22	205	3.20	3.31	3.26	0.00	0.00	0.00	0.00	0.00	0.00	0.19	0.18	0.25	0.21	0.00	0.00	0.00	3.46
97	27.881	Galacturonic acid MeOX 1	333.1	0.18	0.17	0.18	0.00	1.30	0.65	1.26	1.47	1.36	0.14	0.15	0.05	0.11	1.17	1.14	1.15	3.46
4	4.992	m_3/C_008	183.9	0.04	0.10	0.07	1.03	1.26	1.14	0.74	1.00	0.87	0.21	0.22	0.23	0.22	0.68	1.28	0.98	3.29
59	21.068	a_39/x_2/U_032	263.2	0.00	0.11	0.06	2.14	0.00	1.07	1.26	0.00	0.63	0.00	0.18	0.16	0.11	0.00	1.97	0.99	2.85
55	19.765	a_34/f_79	217.9	0.04	0.03	0.04	1.53	1.56	1.54	0.46	0.70	0.58	0.11	0.05	0.12	0.09	0.13	0.00	0.06	2.32
37	15.543	P1552/f_105	174.1	0.02	0.03	0.02	0.89	0.71	0.80	0.49	0.69	0.59	0.00	0.00	0.00	0.00	0.76	0.78	0.77	2.18
47	15.570	P1753	174.0+248.1	0.02	0.03	0.02	0.88	0.72	0.80	0.00	0.70	0.35	0.00	0.00	0.00	0.00	0.77	0.80	0.78	1.95
113	30.411	Tyrosine 2 TMS	179	0.38	0.37	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.86	1.43	1.80
72	22.862	P2284	173	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.17	0.21	0.11	0.11	0.15	1.32	1.41	1.36	1.68
131	34.580	Linoleic acid	337.2	0.22	0.22	0.22	0.57	0.39	0.48	0.26	0.31	0.28	0.13	0.15	0.09	0.12	0.00	0.45	0.22	1.32
121	32.244	kwn_109	221	0.14	0.15	0.14	0.34	0.39	0.37	0.22	0.15	0.18	0.33	0.33	0.34	0.33	0.00	0.27	0.13	1.16
99	28.048	Gluconate 6 TMS	292.1	0.04	0.04	0.04	0.17	0.30	0.23	0.00	0.58	0.29	0.04	0.02	0.09	0.05	0.60	0.44	0.52	1.13
79	24.295	P2445/f_21	292	0.08	0.10	0.09	0.22	0.32	0.27	0.42	0.63	0.53	0.00	0.00	0.02	0.01	0.00	0.15	0.07	0.97
77	24.050	P2427/Unknown 120	292	0.09	0.11	0.10	0.00	0.36	0.18	0.39	0.64	0.51	0.00	0.00	0.00	0.00	0.08	0.22	0.15	0.94
30	14.132	dm_8	216	0.05	0.05	0.05	0.00	0.36	0.18	0.28	0.49	0.39	0.00	0.03	0.05	0.03	0.29	0.25	0.27	0.91
36	15.466	dm_9	292.1	0.07	0.05	0.06	0.13	0.00	0.06	0.37	0.50	0.44	0.00	0.00	0.00	0.00	0.28	0.30	0.29	0.85
60	21.107	dm_17	221	0.07	0.07	0.07	0.00	0.39	0.20	0.00	0.29	0.15	0.19	0.20	0.12	0.17	0.35	0.14	0.25	0.83
139	37.771	dm_41	202	0.42	0.30	0.36	0.00	0.60	0.30	0.00	0.00	0.00	0.12	0.18	0.16	0.15	0.00	0.00	0.00	0.82
100	28.280	chiro- or scyllo- inosito	318.1	0.11	0.10	0.10	0.00	0.00	0.00	1.09	0.00	0.54	0.00	0.00	0.09	0.03	0.00	0.00	0.00	0.68
103	28.939	P2893	393.1	0.00	0.08	0.04	0.00	0.00	0.00	1.15	0.00	0.58	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.61
32	14.479	Proline 1 TMS	142.0+216.0	0.24	0.12	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.30	0.37	0.47	0.38	0.00	0.00	0.00	0.56
140	38.403	Sucrose	361	0.12	0.05	0.08	0.00	0.54	0.27	0.00	0.13	0.07	0.01	0.03	0.09	0.04	0.00	0.00	0.00	0.46
27	13.138	P1295	70	0.48	0.44	0.46	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.46
78	24.217	dm_19	221	0.11	0.09	0.10	0.00	0.00	0.00	0.37	0.00	0.19	0.22	0.16	0.11	0.16	0.00	0.00	0.00	0.45
106	29.428	f_28	284	0.06	0.08	0.07	0.44	0.00	0.22	0.00	0.26	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.42
43	16.511	dm_12	117	0.32	0.45	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38
53	19.533	dm_14	355	0.14	0.11	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.31	0.31	0.14	0.25	0.00	0.00	0.00	0.38
42	16.240	Succinate 2 TMS	246.8	0.02	0.02	0.02	0.15	0.26	0.20	0.19	0.00	0.09	0.03	0.02	0.00	0.02	0.00	0.00	0.00	0.34
143	39.578	P3960	371.2	0.19	0.28	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.03	0.06	0.07	0.00	0.00	0.00	0.30
112	30.243	dm_25	293.1	0.02	0.04	0.03	0.54	0.00	0.27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.30
9	6.513	dm_5	355	0.08	0.06	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.22	0.21	0.10	0.18	0.00	0.00	0.00	0.25
110	29.957	P2997	361	0.17	0.18	0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.04	0.03	0.04	0.00	0.00	0.00	0.22
38	15.698	dm_10	281	0.10	0.13	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.10	0.00	0.00	0.00	0.21
119	31.870	P3188	327	0.18	0.17	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.05	0.03	0.00	0.00	0.00	0.20
149	48.712	Tocopherol 1 TMS	502.6	0.00	0.28	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.02	0.00	0.00	0.00	0.16
145	39.888	P3991	239.1	0.13	0.16	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14
50	18.667	P1863/Unknown 108	116.8+229.1	0.00	0.14	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.21	0.07	0.00	0.00	0.00	0.14

109	29.765	dm_23	369	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.06	0.08	0.08	0.00	0.00	0.00	0.13
56	20.229	Hydroxyproline 2 TMS	67.9	0.12	0.12	0.12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12
116	30.824	dm_27	428.8	0.05	0.06	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.06	0.12	0.07	0.00	0.00	0.00	0.12
125	33.083	P3307	429	0.04	0.03	0.03	0.00	0.00	0.00	0.05	0.00	0.03	0.07	0.06	0.04	0.06	0.00	0.00	0.00	0.12
58	20.552	dm_16	56	0.10	0.09	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
81	24.824	Asparagine 3 TMS	231.0+258.0	0.07	0.06	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.08
123	32.851	dm_30	443.1	0.01	0.03	0.02	0.00	0.00	0.00	0.08	0.00	0.04	0.00	0.00	0.04	0.01	0.00	0.00	0.00	0.08
23	11.949	dm_7	212	0.03	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.04	0.05	0.00	0.00	0.00	0.08
70	22.436	K_47	188	0.03	0.03	0.03	0.00	0.00	0.00	0.07	0.00	0.03	0.03	0.00	0.00	0.01	0.00	0.00	0.00	0.07
39	15.801	dm_11	199	0.01	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.08	0.11	0.06	0.00	0.00	0.00	0.07
126	33.315	dm_32	202	0.06	0.07	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07
44	16.925	B-alanine 3 TMS	248	0.07	0.06	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
62	21.197	x_3 (carry over)	71.9	0.08	0.03	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.06
147	42.314	b_35 / f_50 / d_68 / A_22	399.2	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.05
134	36.158	dm_38	202	0.04	0.06	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05
69	22.346	N-acetoglutamic acid 2 TM	174.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.04	0.02	0.00	0.00	0.01	0.00	0.00	0.00	0.04
51	18.719	K_136	116	0.00	0.08	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
64	21.417	dm_18	177	0.04	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
41	16.059	Fumarate 2 TMS	245	0.00	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.01	0.00	0.00	0.00	0.04
49	17.957	dm_13	146	0.03	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
63	21.327	Cysteine NNO	220	0.01	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.03
127	33.651	dm_33	547.4	0.02	0.02	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
74	23.237	Rhamnose MeOX2 4 TMS	277	0.00	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
133	35.616	dm_37	202	0.00	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02
46	17.480	3,4-Dihydroxybutyrate 3 T	188.8	0.03	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
98	27.945	dm_21	492.2	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
115	30.798	dm_26	221	0.00	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
132	35.216	dm_36	355	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	0.00	0.01	0.00	0.00	0.00	0.01
2	4.541	dm_2	68.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	4.876	dm_3	54.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	5.314	trailing tale	74.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
65	21.753	x_4	114.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
75	23.689	K_52	316.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
82	24.953	dm_20	132.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
84	25.675	P25541	210.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
111	30.063	dm_24	186.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
120	31.973	dm_29	149	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
122	32.618	P3263 / Unknown 115	117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
124	32.915	dm_31	217.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
128	34.064	dm_34	149	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
135	37.217	dm_39	165	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
136	37.281	dm_40	429	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
137	37.436	Eicosanoic acid 1 TMS	117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
138	37.720	K_105 (Arachidonic acid)	117	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
141	38.920	dm_42	259.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
142	39.423	a_78	285.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
144	39.694	P3967	480.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
146	41.991	P4232	203.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table P2-1

Common name (Target)	Residue	Uniprot ID Human	Uniprot ID Rat	Protein name
AKT1	S473	P31749	P47196	RAC-alpha serine/threonine-protein kinase
CREB1	S133	P16220	P15337	Cyclic AMP-responsive element-binding protein 1
EGFR	Y1068	P00533	Q9WTS1	Epidermal growth factor receptor
ERK1 (MAPK3)	T202/Y204	P27361	P21708	Mitogen-activated protein kinase 3
FAK1	Y397	Q05397	O35346 (FADK 1)	Focal adhesion kinase 1
GSK3B	S21/S9	P49841	P18266 (GSK-3 beta)	Glycogen synthase kinase-3 beta
HSP27 (HspB1)	S78	P04792	P42930 (HspB1)	Heat shock protein beta-1
IKBA	S32/S36	P25963	Q63746 (IkB-alpha)	NF-kappa-B inhibitor alpha
JNK2 (MAPK9)	T183/Y185	P45984	P49186 (MAPK 9)	Mitogen-activated protein kinase 9
MEK1 (MAPKK1)	S217/S221	Q02750	Q01986 (MAPKK 1)	Dual specificity mitogen-activated protein kinase kinase 1
MKK6 (MAPKK6)	S207/T211	P52564	Q925D6 (-)	Dual specificity mitogen-activated protein kinase kinase 6
NFKB	S536	Q04206	O88619 (-)	Transcription factor p65
p38MAPK	T180/Y182	Q16539 (MAPK 14)/Q15759 (MAPK 11)	P70618 (MAPK 14)/(MAPK11)	Mitogen-activated protein kinase 14/11
P53	S46	P04637	P10361	Cellular tumor antigen p53
RPS6KB1 (P70S6K, S6K1)	T421/S424	P23443	P67999	Ribosomal protein S6 kinase beta-1
RPS6	S235/S236	P62753	P62755	40S ribosomal protein S6
RPS6KA1 (RSK1)	S380	Q15418	Q63531	Ribosomal protein S6 kinase alpha-1
SHP2	Y542	Q06124	P41499	Tyrosine-protein phosphatase non-receptor type 11
WNK1	T60	Q9H4A3	Q9JIH7	Serine/threonine-protein kinase WNK1

Final phosphoproteomics assay panel.

Table P2-2

Target	Uniprot ID Human	Uniprot ID Rat	Name
CCL2 (MCP-1)	P13500	P14844	C-C motif chemokine 2
CCL20 (MIP3-alpha)	P78556	P97884	C-C motif chemokine 20
CCL3 (MIP1-alpha)	P10147	P50229	C-C motif chemokine 3
CCL5	P13501	P50231	C-C motif chemokine 5
CNTF	P26441	P20294	Ciliary neurotrophic factor
CRP	P02741	P48199	C-reactive protein
CXL10 (IP10)	P02778	P48973	C-X-C motif chemokine 10
EGF	P01133	P07522	Pro-epidermal growth factor
GROA (CXCL1)	P09341	P14095	Growth-regulated alpha protein
HAVR1	Q96D42	O54947	Hepatitis A virus cellular receptor 1 (Human) Hepatitis A virus cellular receptor 1 homolog (Rat)
ICAM1	P05362	Q00238	Intercellular adhesion molecule 1
IFNG	P01579	P01581	Interferon gamma
IL10	P22301	P29456	Interleukin-10
IL1A	P01583	P16598	Interleukin-1 alpha
IL1B	P01584	Q63264	Interleukin-1 beta
IL6	P05231	P20607	Interleukin-6
LYAM1	P14151	P30836	L-selectin
NGF	P01138	P25427	Beta-nerve growth facto
AGER (RAGE)	Q15109	Q63495	Advanced glycosylation end product-specific receptor
TNFA	P01375	P16599	Tumor necrosis factor
VEGFB	P49765	O35485	Vascular endothelial growth factor B
X3CL1	P78423	O55145	Fractalkine

Final cytokine assay panel.

Table P5-1

Nutraceutical	Major functions	Reference	HCC-effective concentrations	xMAP assay selected concentrations
QUE (*flavonol*)	• antioxidant • antiproliferative • anti-inflammatory • antimetastatic	10, 21	10–50 μM	25 μM
POH (*monoterpene*)	• antiproliferative • anti-inflammatory • proapoptotic • antiangiogenic	22	0.25–1 mM	500 μM
FIS (*flavonol*)	• antiproliferative • anti-inflammatory • antioxidant • antiangiogenic	10	10–50 μM	10 μM
CUR (*polyphenol*)	• anti-inflammatory • antioxidant	10	10–25 μM	15 μM
EGCG (*polyphenol*)	• anti-inflammatory • antioxidant • antiproliferative • antiangiogenic	23, 24	50–200 μM	50–200 μM
ERI (*flavonone*)	• anti-inflammatory • antioxidant • antiproliferative	25	10–50 μM	25 μM
NAR (*flavonone*)	• antiproliferative • anti-inflammatory • proapoptotic	21	10–200 μM	100 μM

Selected nutraceuticals tested against HCC cells. A complete list of tested nutraceuticals is shown in Supplementary Table S1, at DOI: 10.1002/psp4.40.

Table P6-1

Drug name	Classification	Reference[a]
Afatinib	0	Drugs@FDA and literature search
Alendronate	0	Drugs@FDA and literature search
Amiodarone	1	Drugs@FDA
Axitinib	1	Drugs@FDA
Bosutinib	0	Drugs@FDA and literature search
Cefuroxime	0	Drugs@FDA and literature search
Crizotinib	0	Drugs@FDA and literature search
Cyclosporine	0	Drugs@FDA and literature search
Cytarabine	1	NIH DailyMed
Dasatinib	1	Drugs@FDA
Diclofenac	1	Drugs@FDA
Domperidone	0[a]	Not approved by FDA
Doxorubicin	1	Drugs@FDA
Diethylpropion	0	Drugs@FDA and literature search
Erlotinib	0	Drugs@FDA and literature search
Gefitinib	0	Drugs@FDA and literature search
Imatinib	1	Drugs@FDA
Lapatinib	0	Drugs@FDA
Methotrexate	0	Drugs@FDA and literature search
Olmesartan	0	Drugs@FDA and literature search
Paroxetine	1	Drugs@FDA
Ponatinib	1	Drugs@FDA
Regorafenib	0	Drugs@FDA and literature search
Ruxolitinib	0	Drugs@FDA and literature search
Sorafenib	1	Drugs@FDA
Sunitinib	1	Drugs@FDA
Tofacitinib	0	Drugs@FDA and literature search
Trametinib	1	Drugs@FDA and literature search
Ursodeoxycholic acid	0	Drugs@FDA and literature search
Vandetanib	1	Drugs@FDA
Vemurafenib	0	Drugs@FDA and literature search

FDA, US Food and Drug Administration; NIH, National Institutes of Health.
Note: Toxic: 1 (clinical incidence $\geq$ 0.1%), and nontoxic: 0 (clinical incidence $<$0.1%). https://dailymed.nlm.nih.gov/dailymed/

[a]Domperidone was profiled by Drug Toxicity Signature Generation Center (DtoxS) and toxicity information was from http://www.hc-sc.gc.ca/dhp-mps/medeff/reviews-examens/domperidone-eng.php.

The list of drugs with gene expression in cardiomyocytes and their cardiotoxicity classification

Table P6-2

Gene/protein	Coefficient	Gene/protein	Coefficient	Gene/protein	Coefficient	Gene/protein	Coefficient
CYP3A4	−0.39	TOP2A	−0.11	FLI1	−0.03	H2AFX	−0.01
ZNF823	0.29	MAX	0.09	TCF12	−0.03	IRF1	−0.011
CASP3	0.20	JUND	−0.08	AHR	0.03	MAP3K5	0.01
HJURP	−0.19	MAPK12	−0.07	BCR	0.03	E2F1	0.01
EPHA2	−0.19	RXRA	0.07	GATA3	0.03	SMOC2	0.01
STAT1	−0.17	HOXA5	−0.07	SMC3	0.02	CYP2D6	−0.01
SP2	0.15	STAT5A	−0.05	EDN1	0.02		
PDGFR-A	−0.12	TCF7L2	0.05	FOXF2	−0.02		
TRIM28	−0.12	NR4A2	−0.03	CTCFL	−0.02		

Nodes from drugs' signaling networks constructed using integer linear programming (ILP) included proteins (targets and protein-protein interactions) and genes (differentially expressed). The gene/protein nodes from ILP were then subjected to elastic net regularization.

Predictors with non-zero coefficients from modeling/analysis of cardiomyocyte data

Table P6-3

Predictors	Regulating microRNAs[a] that are of diagnostic value
CYP3A4	No information
ZNF823	miR193-3p (↓)
CASP3	miR-375[b], miR-26b-5p (↓); miR-30e-5p (↓), let-7a-5p (↑)
HJURP	miR-671-5p (↑)
EPHA2	miR-26b-5p (↓), miR-193b-3p (↓); miR-16-5p (↓)
STAT1	miR 145-5p (↓)
SP2	miR-29a-3p (↓), miR-638[b]
PDGFR-A	miR-140-5p (↓); miR-26b-5p (↓); miR-29b-3p (↓); 181a-5p (↑); miR-1233 (↑)
TRIM28	miR-423-5p (inconsistent reports), miR-193b-3p (↓), miR-183-3p (↓), miR-92a-3p (↓)
TOP2A	miR-193b-3p (↓), miR-21-5p (↑)

Top 10 predictors and their corresponding regulating microRNAs that are reportedly of diagnostic value for heart failure

(a) Regulating microRNAs are from Chou et al.[235] (http://mirtarbase.mbc.nctu.edu.tw).

(b) Differentiating heart failure with reduced ejection fraction from heart failure with preserved ejection fraction. ↑ and ↓ represent elevation and decrease, respectively, compared to healthy controls.

Table P6-S1

Alprostadil	Fluoxetine	Propofol
Amiodarone	Flurbiprofen	Propranolol
Amitriptyline	Gefitinib	Raloxifene
Azacitidine	Hydralazine	Reserpine
Budesonide	Iloprost	Riluzole
Bupropion	Imatinib	Risperidone
Capsaicin	Imipramine	Rosiglitazone
Carbamazepine	Indometacin	Selegiline
Carmustine	Irinotecan	Simvastatin
Chloroquine	Ketoconazole	Sulfasalazine
Chlorpromazine	Levodopa	Sulindac
Cimetidine	Menadione	Tamoxifen
Citalopram	Mesalazine	Testosterone
Clomipramine	Metformin	Tetracycline
Clotrimazole	Methotrexate	Thalidomide
Clozapine	Miconazole	Theophylline
Colchicine	Minocycline	Thioridazine
Daunorubicin	Mitoxantrone	Tretinoin
Decitabine	Naloxone	Trifluoperazine
Desipramine	Nifedipine	Troglitazone
Diclofenac	Omeprazole	Valproic Acid
Dinoprostone	Orlistat	Verapamil
Doxorubicin	Pioglitazone	Vinblastine
Doxycycline	Prednisolone	Vorinostat
Etoposide	Probenecid	Zidovudine

The list of 75 drugs used for exploratory modeling.

Table P6-S2

Year	First Author	CMs	Doxorubicin nM	DOX h	wash d	expression	data level availability	Original Name
2016	Maillet	hESC	1000, 2500	16	0	HTS (Illumina HiSeq 2500)	not available	Modeling Doxorubicin-induced Cardiotoxicity in Human Pluripotent Stem Cell Derived-Cardiomyocytes
2016	Burridge	hESC	100, 1000, 10000	24	0	array (HG-U133_Plus_2)	up/down reg, FC, p-value	Human induced pluripotent stem cell-derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. https://www.ebi.ac.uk/arrayexpress/experiments/E-GEOD-79413/
		hiPSC	1000			HTS (Illumina HiSeq 2000)	gene expression	
2016	Holmgren	hESC	50, 150, 450			microRNA		MicroRNAs as potential biomarkers for doxorubicin-induced cardiotoxicity
2016	Chaudhari	hiPSC	156			microRNA		MicroRNAs as early toxicity signatures of doxorubicin in human-induced pluripotent stem cell-derived cardiomyocytes
2015	Chaudhari	hiPSC	156	48, 144	0, 8, 12	array (HG-U133_Plus_2)	up/down regulation	Identification of genomic biomarkers for anthracycline-induced cardiotoxicity in human iPSC-derived cardiomyocytes: an in vitro repeated exposure toxicity approach for safety assessment
				48	0	RT-PCR (84 target genes)	up/down reg, FC, p-value	
2015	Holmgren	hESC	50, 150, 450	24, 48	12	array (HuGene ST 2.0)	deregulation	Identification of novel biomarkers for doxorubicin-induced toxicity in human cardiomyocytes derived from pluripotent stem cells.

Studies of transcriptomic profiles of doxorubicin.

Table P7-S1

	TOP250	AMI	FFA	TET	TMX	VPA	
GO:0019838	7.7	28.08	30.94	26.52	31.05	28.93	growth factor binding
P00005	4.38	0	20.12	13.41	17.66	32.92	Angiogenesis
GO:0042127	2.5	5.43	6.22	8.42	6.11	6.59	regulation of cell proliferation
GO:0051241	2.49	5.87	6.65	7.2	5.84	6.35	negative regulation of multicellular organismal process
GO:0051240	2.19	5.75	6.59	6.2	6.13	7.29	positive regulation of multicellular organismal process
GO:0051094	2.34	0	6.63	6.16	7.49	7.75	positive regulation of developmental process
GO:0030335	4.38	0	10.98	0	0	12.35	positive regulation of cell migration
GO:2000147	4.59	0	10.54	0	0	11.85	positive regulation of cell motility
GO:0051272	4.47	0	10.26	0	0	11.54	positive regulation of cellular component movement
GO:0032879	2.11	3.56	4.93	4.26	5.51	5.13	regulation of localization
GO:0040017	4.46	0	9.83	0	0	11.05	positive regulation of locomotion
GO:2000145	4.14	0	6.75	6.75	0	7.36	regulation of cell motility
GO:0051239	1.91	3.76	4.6	5.02	4.22	5.13	regulation of multicellular organismal process
GO:2000026	2.33	0	5.45	5.45	4.22	6.29	regulation of multicellular organismal development
GO:0050793	2.04	0	4.92	5.16	5.4	5.63	regulation of developmental process
GO:0071229	6.02	0	0	0	15.6	0	cellular response to acid chemical
GO:0001101	4.88	0	0	0	15.52	0	response to acid chemical
GO:0022008	2.19	0	6.56	4.63	0	6.63	neurogenesis
GO:0065008	1.74	3.48	3.61	3.28	3.75	3.63	regulation of biological quality
GO:0048522	1.58	3.1	3.28	3.75	3.81	3.36	positive regulation of cellular process
GO:0032502	1.98	2.69	3.48	2.75	4.17	2.77	developmental process
GO:0040012	3.79	0	0	6.88	0	6.75	regulation of locomotion
GO:0048518	1.53	2.98	3.02	3.44	3.38	3.07	positive regulation of biological process
GO:0072359	3.78	0	0	0	6.98	5.91	circulatory system development
GO:0007275	1.88	2.71	3.17	2.93	2.57	3.19	multicellular organism development
GO:0048856	1.84	2.64	2.95	2.84	2.49	2.97	anatomical structure development
GO:0002376	2.91	0	3.89	4.11	0	4.49	immune system process
GO:0048869	1.81	0	3.57	3.57	2.85	3.58	cellular developmental process
GO:0048731	2	0	3.47	3.06	2.68	3.64	system development
GO:0051128	2.1	0	4.25	0	3.51	4.29	regulation of cellular component organization
GO:0042060	3.81	0	0	0	0	10.22	wound healing
GO:0032501	1.56	2.21	2.53	2.61	2.29	2.57	multicellular organismal process
GO:0001568	4.22	0	0	0	8.87	0	blood vessel development
GO:0009611	3.34	0	0	0	0	9.41	response to wounding
GO:0030154	1.83	0	3.64	3.48	0	3.66	cell differentiation
GO:0001944	4.18	0	0	0	8.43	0	vasculature development
GO:0007399	2.64	0	4.98	0	0	4.93	nervous system development
GO:0072358	4.1	0	0	0	8.26	0	cardiovascular system development
GO:0048513	1.87	0	3.32	3.32	0	3.67	animal organ development
GO:0048468	2.5	0	4.69	0	0	4.8	cell development
GO:0040011	6.18	0	0	0	0	5.5	locomotion
GO:0030334	4.21	0	0	0	0	7.25	regulation of cell migration
R-HSA-42247	3.58	0	0	0	0	7.02	Axon guidance
GO:0051270	3.88	0	0	0	0	6.72	regulation of cellular component movement
GO:0000902	3.35	0	0	0	0	6.8	cell morphogenesis
GO:0005515	1.27	1.72	1.78	1.83	1.7	1.79	protein binding
GO:0065007	1.26	1.7	1.65	1.75	1.71	1.67	biological regulation
GO:0044087	2.88	0	0	0	6.01	0	regulation of cellular component biogenesis
GO:0044459	1.75	3.87	0	0	3.21	0	plasma membrane part
GO:0030030	2.54	0	0	0	0	5.37	cell projection organization
GO:0006928	2.25	0	0	0	0	5.2	movement of cell or subcellular component
GO:0009888	2.45	0	0	0	0	4.61	tissue development
GO:0009653	2.69	0	0	0	0	4.27	anatomical structure morphogenesis
GO:0009987	1.18	0	1.4	1.4	1.4	1.4	cellular process
GO:0005488	1.17	1.42	0	1.42	0	1.38	binding
GO:0005737	1.34	0	0	0	0	1.56	cytoplasm

List of GO annotation terms (1st column) and their description (last column). The other columns contain the fold enrichment of the specific annotation term for the GSE49541 dataset analyzed with GEO2R (please see Results for further details) labelled “TOP250” and the gene lists corresponding to the pathway nodes of AMI, FFA, TET, TMX and VPA.

Βιβλιογραφικές Αναφορές | References

1. Zuppinger, C., Timolati, F. & Suter, T. M. Pathophysiology and diagnosis of cancer drug induced cardiomyopathy. *Cardiovasc. Toxicol.* **7,** 61–66 (2007).
2. FDA Drugs@FDA: FDA Approved drug products. <http://www.https://www.accessdata.fda. gov/scripts/cder/daf/>. Accessed July 2016.
3. Pouliot, Y., Chiang, A. P. & Butte, A. J. Predicting Adverse Drug Reactions Using Publicly Available PubChem BioAssay Data. *Clin. Pharmacol. Ther.* **90,** 90–99 (2011).
4. Liu, M. *et al.* Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J. Am. Med. Inform. Assoc.* **19,** e28-35 (2012).
5. Harpaz, R. *et al.* Novel Data-Mining Methodologies for Adverse Drug Event Discovery and Analysis. *Clin. Pharmacol. Ther.* **91,** 1010–1021 (2012).
6. DtoxS - Drug Toxicity Signature Generation Center - Data & Resources. at <https://martip03.u.hpc.mssm.edu/data.php>
7. Burridge, P. W. *et al.* Human induced pluripotent stem cell–derived cardiomyocytes recapitulate the predilection of breast cancer patients to doxorubicin-induced cardiotoxicity. *Nat. Med.* **22,** 547–556 (2016).
8. Chaudhari, U. *et al.* Identification of genomic biomarkers for anthracycline-induced cardiotoxicity in human iPSC-derived cardiomyocytes: an in vitro repeated exposure toxicity approach for safety assessment. *Arch. Toxicol.* 1–15 (2015). doi:10.1007/s00204-015-1623-5
9. Melas, I. N. *et al.* Identification of drug-specific pathways based on gene expression data: application to drug induced lung injury. *Integr. Biol. (Camb).* **7,** 904–20 (2015).
10. National Cancer Institute. Common Terminology Criteria for Adverse Events (CTCAE) version 4.03. <https://ctep.cancer.gov/protocolDevelopment/electronic_applications/ctchtm>. Accessed July 2016.

11. International Council for Harmonisation of Technical Requirements for Pharmaceuticals (ICH). Medical Dictionary for Regulatory Activities (MeDRA). <http://www.meddra. com/>. Accessed July 2016.
12. Hur, J., Guo, A. Y., Loh, W. Y., Feldman, E. L. & Bai, J. P. F. Integrated systems pharmacology analysis of clinical drug-induced peripheral neuropathy. *CPT pharmacometrics Syst. Pharmacol.* **3,** e114 (2014).
13. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. B* **67,** 301–320 (2005).
14. Institute WTS. Genomics of Drug Sensitivity in Cancer. <http://www.cancer.rxgene.org/ news>. Accessed August 2017.
15. Melas, I. N., Samaga, R., Alexopoulos, L. G. & Klamt, S. Detecting and removing inconsistencies between experimental data and signaling network topologies using integer linear programming on interaction graphs. *PLoS Comput. Biol.* **9,** e1003204 (2013).
16. Connectivity Map (CMAP). <https://www.broadinstitute.org/connectivity-map-cmap>. Accessed August 2016.
17. Szklarczyk, D. *et al.* STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44,** D380–D384 (2016).
18. (EMBL-EBI) EBI. HUGO Gene Nomenclature Committee. <http://www.genenames. org/>. Accessed August 2016.
19. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37,** D619-22 (2009).
20. Gentleman, R. annotate: Annotation for microarrays. R package version 1.50.0.
21. Carlson, M. hgu133a2.db: Affymetrix Human Genome U133A 2.0 Array annotation data (chip hgu133a2). R package version 3.2.3. (2016).
22. Barpe, D. R., Rosa, D. D. & Froehlich, P. E. Pharmacokinetic evaluation of doxorubicin plasma levels in normal and overweight patients with breast cancer and simulation of dose adjustment by different indexes of body mass. *Eur. J. Pharm. Sci.* **41,** 458–63 (2010).
23. Mitsos, A. *et al.* Identifying Drug Effects via Pathway Alterations using an

Integer Linear Programming Optimization Formulation on Phosphoproteomic Data. *PLoS Comput. Biol.* **5,** e1000591 (2009).

24. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33,** 1–22 (2010).
25. Chou, C.-H. *et al.* miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res.* **44,** D239–D247 (2016).
26. Schulte, C., Westermann, D., Blankenberg, S. & Zeller, T. Diagnostic and prognostic value of circulating microRNAs in heart failure with preserved and reduced ejection fraction. *World J. Cardiol.* **7,** 843–60 (2015).
27. Marques, F. Z., Vizi, D., Khammy, O., Mariani, J. A. & Kaye, D. M. The transcardiac gradient of cardio-microRNAs in the failing heart. *Eur. J. Heart Fail.* **18,** 1000–8 (2016).
28. Leger, K. J. *et al.* Circulating microRNAs: Potential Markers of Cardiotoxicity in Children and Young Adults Treated With Anthracycline Chemotherapy. *J. Am. Heart Assoc.* **6,** e004653 (2017).
29. Karaman, M. W. *et al.* A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* **26,** 127–132 (2008).
30. Force, T., Krause, D. S. & Van Etten, R. A. Molecular mechanisms of cardiotoxicity of tyrosine kinase inhibition. *Nat. Rev. Cancer* **7,** 332–344 (2007).
31. Fabre, G., Julian, B., Saint-Aubert, B., Joyeux, H. & Berger, Y. Evidence for CYP3A-mediated N-deethylation of amiodarone in human liver microsomal fractions. *Drug Metab. Dispos.* **21,** 978–85
32. Zientek, M. A. *et al.* In Vitro Kinetic Characterization of Axitinib Metabolism. *Drug Metab. Dispos.* **44,** 102–114 (2015).
33. Scripture, C. D. & Figg, W. D. Drug interactions in cancer therapy. *Nat. Rev. Cancer* **6,** 546–558 (2006).
34. Bodin, K., Lindbom, U. & Diczfalusy, U. Novel pathways of bile acid metabolism involving CYP3A4. *Biochim. Biophys. Acta* **1687,** 84–93 (2005).
35. Schlaepfer, D. D., Hauck, C. R. & Sieg, D. J. Signaling through focal adhesion kinase. *Prog. Biophys. Mol. Biol.* **71,** 435–78 (1999).

36. Samarel, A. M. Focal adhesion signaling in heart failure. *Pflügers Arch. - Eur. J. Physiol.* **466,** 1101–1111 (2014).
37. Clemente, C. F. M. Z. *et al.* Focal adhesion kinase governs cardiac concentric hypertrophic growth by activating the AKT and mTOR pathways. *J. Mol. Cell. Cardiol.* **52,** 493–501 (2012).
38. Grossmann, J. *et al.* Hierarchical cleavage of focal adhesion kinase by caspases alters signal transduction during apoptosis of intestinal epithelial cells. *Gastroenterology* **120,** 79–88 (2001).
39. Jehle, J. *et al.* Regulation of apoptosis in HL-1 cardiomyocytes by phosphorylation of the receptor tyrosine kinase EphA2 and protection by lithocholic acid. *Br. J. Pharmacol.* **167,** 1563–72 (2012).
40. Xie, B. *et al.* Focal adhesion kinase activates Stat1 in integrin-mediated cell migration and adhesion. *J. Biol. Chem.* **276,** 19512–23 (2001).
41. Sieg, D. J. *et al.* FAK integrates growth-factor and integrin signals to promote cell migration. *Nat. Cell Biol.* **2,** 249–56 (2000).
42. Pan, Y.-Z., Gao, W. & Yu, A.-M. MicroRNAs regulate CYP3A4 expression via direct and indirect targeting. *Drug Metab. Dispos.* **37,** 2112–7 (2009).
43. Haq, I., Murphy, E. & Dacre, J. Osteoarthritis. *Postgr. Med J* **79,** 377–383 (2003).
44. Lotz, M. *et al.* Value of biomarkers in osteoarthritis: current status and perspectives. *Ann. Rheum. Dis.* **72,** 1756–63 (2013).
45. Priori, R. *et al.* Metabolomics in rheumatic diseases: The potential of an emerging methodology for improved patient diagnosis, prognosis, and treatment efficacy. *Autoimmun. Rev.* **12,** 1022–1030 (2013).
46. Adams, S. B., Setton, L. a & Nettles, D. L. On the Horizon From the ORS The Role of Metabolomics in Osteoarthritis Research. **21,** 63–64 (2013).
47. Zhang, W. *et al.* Classification of osteoarthritis phenotypes by metabolomics analysis. *BMJ Open* **4,** e006286 (2014).
48. Hügle, T. *et al.* Synovial fluid metabolomics in different forms of arthritis assessed by nuclear magnetic resonance spectroscopy. *Clin. Exp. Rheumatol.* **30,** 240–5
49. Ciurtin, C. *et al.* Correlation between different components of synovial fluid

and pathogenesis of rheumatic diseases. *Rom. J. Intern. Med.* **44,** 171–81 (2006).

50. Mickiewicz, B. *et al.* Metabolic analysis of knee synovial fluid as a potential diagnostic approach for osteoarthritis. *J. Orthop. Res.* (2015). doi:10.1002/jor.22949
51. Ruiz-Romero, C. & Blanco, F. J. The role of proteomics in osteoarthritis pathogenesis research. *Curr. Drug Targets* **10,** 543–556 (2009).
52. Adams, S. B. *et al.* Global metabolic profiling of human osteoarthritic synovium. *Osteoarthritis Cartilage* **20,** 64–7 (2012).
53. Zhai, G. *et al.* Serum branched-chain amino acid to histidine ratio: a novel metabolomic biomarker of knee osteoarthritis. *Ann. Rheum. Dis.* **69,** 1227–31 (2010).
54. Tsuchida, A. I. *et al.* Cytokine profiles in the joint depend on pathology, but are different between synovial fluid, cartilage tissue and cultured chondrocytes. *Arthritis Res. Ther.* **16,** 441 (2014).
55. Wang, S. *et al.* Identification of α2-macroglobulin as a master inhibitor of cartilage-degrading factors that attenuates the progression of posttraumatic osteoarthritis. *Arthritis Rheumatol. (Hoboken, N.J.)* **66,** 1843–53 (2014).
56. Ritter, S. Y. *et al.* Proteomic analysis of synovial fluid from the osteoarthritic knee: comparison with transcriptome analyses of joint tissues. *Arthritis Rheum.* **65,** 981–92 (2013).
57. Liao, W. *et al.* Proteomic analysis of synovial fluid: insight into the pathogenesis of knee osteoarthritis. *Int. Orthop.* **37,** 1045–53 (2013).
58. Wanner, J. *et al.* Proteomic profiling and functional characterization of early and late shoulder osteoarthritis. *Arthritis Res. Ther.* **15,** R180 (2013).
59. Pan, X., Huang, L., Chen, J., Dai, Y. & Chen, X. Analysis of synovial fluid in knee joint of osteoarthritis:5 proteome patterns of joint inflammation based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. *Int. Orthop.* **36,** 57–64 (2012).
60. Antoniou, J., Huk, O., Zukor, D., Eyre, D. & Alini, M. Collagen crosslinked N-telopeptides as markers for evaluating particulate osteolysis: a preliminary study. *J. Orthop. Res.* **18,** 64–7 (2000).

61. Fiorito, S., Magrini, L. & Goalard, C. Pro-inflammatory and anti-inflammatory circulating cytokines and periprosthetic osteolysis. *J. Bone Joint Surg. Br.* **85,** 1202–6 (2003).
62. Arabmotlagh, M., Sabljic, R. & Rittmeister, M. Changes of the biochemical markers of bone turnover and periprosthetic bone remodeling after cemented hip arthroplasty. *J. Arthroplasty* **21,** 129–34 (2006).
63. Schneider, U. *et al.* Increased urinary crosslink levels in aseptic loosening of total hip arthroplasty. *J. Arthroplasty* **13,** 687–92 (1998).
64. Wilkinson, J. M., Hamer, A. J., Rogers, A., Stockley, I. & Eastell, R. Bone mineral density and biochemical markers of bone turnover in aseptic loosening after total hip arthroplasty. *J. Orthop. Res.* **21,** 691–6 (2003).
65. Savarino, L. *et al.* Systemic cross-linked N-terminal telopeptide and procollagen I C-terminal extension peptide as markers of bone turnover after total hip arthroplasty. *J. Bone Joint Surg. Br.* **87,** 571–6 (2005).
66. Li, M. G., Thorsen, K. & Nilsson, K. G. Increased bone turnover as reflected by biochemical markers in patients with potentially unstable fixation of the tibial component. *Arch. Orthop. Trauma Surg.* **124,** 404–9 (2004).
67. Lawrence, N. R., Jayasuriya, R. L., Gossiel, F. & Wilkinson, J. M. Diagnostic accuracy of bone turnover markers as a screening tool for aseptic loosening after total hip arthroplasty. *Hip Int.* (2015). doi:10.5301/hipint.5000253
68. Hernigou, P., Intrator, L., Bahrami, T., Bensussan, A. & Farcet, J. P. Interleukin-6 in the blood of patients with total hip arthroplasty without loosening. *Clin. Orthop. Relat. Res.* 147–54 (1999). at <http://www.ncbi.nlm.nih.gov/pubmed/10627728>
69. Hundrić-Haspl, Z., Pecina, M., Haspl, M., Tomicic, M. & Jukic, I. Plasma cytokines as markers of aseptic prosthesis loosening. *Clin. Orthop. Relat. Res.* **453,** 299–304 (2006).
70. Granchi, D. *et al.* Bone-resorbing cytokines in serum of patients with aseptic loosening of hip prostheses. *J. Bone Joint Surg. Br.* **80,** 912–7 (1998).
71. Granchi, D. *et al.* Serum levels of osteoprotegerin and receptor activator of nuclear factor-kappaB ligand as markers of periprosthetic osteolysis. *J. Bone Joint Surg. Am.* **88,** 1501–9 (2006).

72. Moreschini, O., Fiorito, S., Magrini, L., Margheritini, F. & Romanini, L. Markers of connective tissue activation in aseptic hip prosthetic loosening. *J. Arthroplasty* **12,** 695–703 (1997).
73. Nivbrant, B., Karlsson, K. & Kärrholm, J. Cytokine levels in synovial fluid from hips with well-functioning or loose prostheses. *J. Bone Joint Surg. Br.* **81,** 163–6 (1999).
74. Kovacik, M. W., Gradisar, I. A., Haprian, J. J. & Alexander, T. S. Osteolytic indicators found in total knee arthroplasty synovial fluid aspirates. *Clin. Orthop. Relat. Res.* 186–94 (2000). at <http://www.ncbi.nlm.nih.gov/pubmed/11039806>
75. Kawasaki, M., Hasegawa, Y., Kondo, S. & Iwata, H. Concentration and localization of YKL-40 in hip joint diseases. *J. Rheumatol.* **28,** 341–5 (2001).
76. Sypniewska, G., Lis, K. & Bilinski, P. J. Bone turnover markers and cytokines in joint fluid: analyses in 10 patients with loose hip prosthesis and 39 with coxarthrosis. *Acta Orthop. Scand.* **73,** 518–22 (2002).
77. Del Buono, A., Denaro, V. & Maffulli, N. Genetic susceptibility to aseptic loosening following total hip arthroplasty: a systematic review. *Br. Med. Bull.* **101,** 39–55 (2012).
78. Di Cesare, P. E., Chang, E., Preston, C. F. & Liu, C. Serum interleukin-6 as a marker of periprosthetic infection following total hip and knee arthroplasty. *J. Bone Joint Surg. Am.* **87,** 1921–7 (2005).
79. Bottner, F. *et al.* Interleukin-6, procalcitonin and TNF-alpha: markers of peri-prosthetic infection following total joint replacement. *J. Bone Joint Surg. Br.* **89,** 94–9 (2007).
80. Nilsdotter-Augustinsson, A. *et al.* Inflammatory response in 85 patients with loosened hip prostheses: a prospective study comparing inflammatory markers in patients with aseptic and septic prosthetic loosening. *Acta Orthop.* **78,** 629–39 (2007).
81. Randau, T. M. *et al.* Interleukin-6 in serum and in synovial fluid enhances the differentiation between periprosthetic joint infection and aseptic loosening. *PLoS One* **9,** e89045 (2014).
82. Ettinger, M. *et al.* Circulating Biomarkers for Discrimination Between Aseptic

Joint Failure, Low-Grade Infection, and High-Grade Septic Failure. *Clin. Infect. Dis.* **61,** 332–41 (2015).

83. Mertens, M. T. & Singh, J. A. Biomarkers in arthroplasty: a systematic review. *Open Orthop. J.* **5,** 92–105 (2011).
84. Serhan, C. N. *et al.* Resolution of inflammation: state of the art, definitions and terms. *FASEB J.* **21,** 325–332 (2007).
85. Long, F., Schipani, E., Asahara, H., Kronenberg, H. & Montminy, M. The CREB family of activators is required for endochondral bone development. *Development* **128,** 541–50 (2001).
86. Murakami, S. *et al.* Constitutive activation of MEK1 in chondrocytes causes Stat1-independent achondroplasia-like dwarfism and rescues the Fgfr3-deficient mouse phenotype. *Genes Dev.* **18,** 290–305 (2004).
87. Berenbaum, F. Signaling transduction: target in osteoarthritis. *Curr. Opin. Rheumatol.* **16,** 616–22 (2004).
88. Chakraborti, S., Mandal, M., Das, S., Mandal, A. & Chakraborti, T. Regulation of matrix metalloproteinases: an overview. *Mol. Cell. Biochem.* **253,** 269–85 (2003).
89. Baldwin, A. S. THE NF-κB AND IκB PROTEINS: New Discoveries and Insights. *Annu. Rev. Immunol.* **14,** 649–681 (1996).
90. Marcu, K. B., Otero, M., Olivotto, E., Borzi, R. M. & Goldring, M. B. NF-kappaB signaling: multiple angles to target OA. *Curr. Drug Targets* **11,** 599–613 (2010).
91. Guilak, F. *et al.* The role of biomechanics and inflammation in cartilage injury and repair. *Clin. Orthop. Relat. Res.* 17–26 (2004). at <http://www.ncbi.nlm.nih.gov/pubmed/15232421>
92. Luyten, F. P., Tylzanowski, P. & Lories, R. J. Wnt signaling and osteoarthritis. *Bone* **44,** 522–527 (2009).
93. Lories, R. J. & Luyten, F. P. Bone morphogenetic proteins in destructive and remodeling arthritis. *Arthritis Res. Ther.* **9,** 207 (2007).
94. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nat. Chem. Biol.* **4,** 682–690 (2008).
95. Mobasheri, A. Applications of Proteomics to Osteoarthritis, a Musculoskeletal

Disease Characterized by Aging. *Front. Physiol.* **2,** 108 (2011).

96. Wilson, R., Whitelock, J. M. & Bateman, J. F. Proteomics makes progress in cartilage and arthritis research. *Matrix Biol.* **28,** 121–128 (2009).
97. Shigemizu, D. *et al.* Using Functional Signatures to Identify Repositioned Drugs for Breast, Myelogenous Leukemia and Prostate Cancer. *PLoS Comput. Biol.* **8,** e1002347 (2012).
98. Garcia, B. A. *et al.* Protein profile of osteoarthritic human articular cartilage using tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **20,** 2999–3006 (2006).
99. Wu, J. *et al.* Comparative proteomic characterization of articular cartilage tissue from normal donors and patients with osteoarthritis. *Arthritis Rheum.* **56,** 3675–3684 (2007).
100. Vincourt, J.-B. *et al.* Establishment of a Reliable Method for Direct Proteome Characterization of Human Articular Cartilage. *Mol. Cell. Proteomics* **5,** 1984–1995 (2006).
101. Lambrecht, S., Verbruggen, G., Verdonk, P. C. M., Elewaut, D. & Deforce, D. Differential proteome analysis of normal and osteoarthritic chondrocytes reveals distortion of vimentin network in osteoarthritis. *Osteoarthr. Cartil.* **16,** 163–173 (2008).
102. Sinz, A. *et al.* Mass spectrometric proteome analyses of synovial fluids and plasmas from patients suffering from rheumatoid arthritis and comparison to reactive arthritis or osteoarthritis. *Electrophoresis* **23,** 3445–3456 (2002).
103. Liao, H. *et al.* Use of mass spectrometry to identify protein biomarkers of disease severity in the synovial fluid and serum of patients with rheumatoid arthritis. *Arthritis Rheum.* **50,** 3792–3803 (2004).
104. Catterall, J. B. *et al.* Development of a novel 2D proteomics approach for the identification of proteins secreted by primary chondrocytes after stimulation by IL-1 and oncostatin M. *Rheumatology* **45,** 1101–1109 (2006).
105. Haglund, L., Bernier, S. M., Önnerfjord, P. & Recklies, A. D. Proteomic analysis of the LPS-induced stress response in rat chondrocytes reveals induction of innate immune response components in articular cartilage. *Matrix Biol.* **27,** 107–118 (2008).

106. Saez-Rodriguez, J., Alexopoulos, L. G. & Stolovitzky, G. Setting the Standards for Signal Transduction Research. *Sci. Signal.* **4,** pe10-pe10 (2011).
107. Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A. & Sorger, P. K. Networks Inferred from Biochemical Data Reveal Profound Differences in Toll-like Receptor and Inflammatory Signaling between Normal and Transformed Hepatocytes. *Mol. Cell. Proteomics* **9,** 1849–1865 (2010).
108. Mitsos, A. *et al.* Identifying Drug Effects via Pathway Alterations using an Integer Linear Programming Optimization Formulation on Phosphoproteomic Data. *PLoS Comput. Biol.* **5,** e1000591 (2009).
109. Quinn, T. M., Schmid, P., Hunziker, E. B., Grodzinsky, A. J. & Mueller, M. E. Proteoglycan deposition around chondrocytes in agarose culture: Construction of a physical and biological interface for mechanotransduction in cartilage. *Biorheology* **39,** 27–37 (2002).
110. Saez-Rodriguez, J. *et al.* Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.* **5,** 331 (2009).
111. Melas, I. N. *et al.* Construction of large signaling pathways using an adaptive perturbation approach with phosphoproteomic data. *Molecular BioSystems* **8,** 1571 (2012).
112. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33,** D428–D432 (2004).
113. Cerami, E. G. *et al.* Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **39,** D685–D690 (2011).
114. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
115. Saez-Rodriguez, J. *et al.* Flexible informatics for linking experimental data to mathematical models via DataRail. *Bioinformatics* **24,** 840–847 (2008).
116. Goldring, M. B. Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Ther. Adv. Musculoskelet. Dis.* **4,** 269–285 (2012).
117. Scanzello, C. R., Plaas, A. & Crow, M. K. Innate immune system activation in osteoarthritis: is osteoarthritis a chronic wound? *Curr. Opin. Rheumatol.* **20,**

565–572 (2008).

118. Dreier, R. Hypertrophic differentiation of chondrocytes in osteoarthritis: the developmental aspect of degenerative joint disorders. *Arthritis Res. Ther.* **12,** 216 (2010).
119. Merz, D., Liu, R., Johnson, K. & Terkeltaub, R. IL-8/CXCL8 and growth-related oncogene alpha/CXCL1 induce chondrocyte hypertrophic differentiation. *J. Immunol.* **171,** 4406–15 (2003).
120. Olivotto, E. *et al.* Chondrocyte hypertrophy and apoptosis induced by GROα require three-dimensional interaction with the extracellular matrix and a co-receptor role of chondroitin sulfate and are associated with the mitochondrial splicing variant of cathepsin B. *J. Cell. Physiol.* **210,** 417–427 (2007).
121. Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* **388,** 394–397 (1997).
122. Kim, H. A. *et al.* The catabolic pathway mediated by Toll-like receptors in human osteoarthritic chondrocytes. *Arthritis Rheum.* **54,** 2152–2163 (2006).
123. Su, S.-L., Tsai, C.-D., Lee, C.-H., Salter, D. M. & Lee, H.-S. Expression and regulation of Toll-like receptor 2 by IL-1β and fibronectin fragments in human articular chondrocytes. *Osteoarthr. Cartil.* **13,** 879–886 (2005).
124. Liu-Bryan, R., Pritzker, K., Firestein, G. S. & Terkeltaub, R. TLR2 signaling in chondrocytes drives calcium pyrophosphate dihydrate and monosodium urate crystal-induced nitric oxide generation. *J. Immunol.* **174,** 5016–23 (2005).
125. Fang, X., Stachowiak, E. K., Dunham-Ems, S. M., Klejbor, I. & Stachowiak, M. K. Control of CREB-binding Protein Signaling by Nuclear Fibroblast Growth Factor Receptor-1. *J. Biol. Chem.* **280,** 28451–28462 (2005).
126. Wang, Y. *et al.* Strain-induced fetal type II epithelial cell differentiation is mediated via cAMP-PKA-dependent signaling pathway. *Am. J. Physiol. Cell. Mol. Physiol.* **291,** L820–L827 (2006).
127. Zeniou, M., Ding, T., Trivier, E. & Hanauer, A. Expression analysis of RSK gene family members: the RSK2 gene, mutated in Coffin-Lowry syndrome, is prominently expressed in brain structures essential for cognitive function and

learning. *Hum. Mol. Genet.* **11,** 2929–40 (2002).

128. Meyer, P. Verification of systems biology research in the age of collaborative competition. *Nat. Biotechnol.* **29,** 811–815 (2011).
129. Meyer, P. Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. *Bioinformatics* **28,** 1193–1201 (2012).
130. Bhattacharya, S., Zhang, Q., Carmichael, P. L., Boekelheide, K. & Andersen, M. E. Toxicity testing in the 21 century: defining new risk assessment approaches based on perturbation of intracellular toxicity pathways. *Plo S One* **6,** e20887 (2011).
131. Huber, P. J. *No Title*. *Robust Statistics* (, 1981).
132. Wu, Z., Irizarry, R. A., Gentleman, R., Murillo, F. M. & Spencer, F. A model based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.* **99,** 909–917 (2004).
133. Irizarry, R. A. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
134. Rayner, T. F. A simple spreadsheet-based, MIAME-supportive format for microarray data: MAGE-TAB. *BMC Bioinformatics* **7,** 489 (2006).
135. Bartelt, R. R., Cruz-Orcutt, N., Collins, M. & Houtman, J. C. Comparison of T cell receptor-induced proximal signaling and downstream functions in immortalized and primary T cells. *Plo S One* **4,** e5430 (2009).
136. Boerma, M. Comparative expression profiling in primary and immortalized endothelial cells: changes in gene expression in response to hydroxy methylglutaryl-coenzyme A reductase inhibition. *Blood Coagul. Fibrin.* **17,** 173–180 (2006).
137. Czekanska, E. M., Stoddart, M. J., Ralphs, J. R., Richards, R. G. & Hayes, J. S. A phenotypic comparison of osteoblast cell lines versus human primary osteoblasts for biomaterials testing. *J. Biomed. Mater. Res. A* (2013).
138. Hou, A., Voorhoeve, P. M., Lan, W., Tin, M. & Tong, L. Comparison of gene expression profiles in primary and immortalized human pterygium fibroblast cells. *Exp. Cell Res.* **319,** 2781–2789 (2013).
139. Pan, C., Kumar, C., Bohl, S., Klingmueller, U. & Mann, M. Comparative

proteomic phenotyping of cell lines and primary cells to assess preservation of cell type-specific functions. *Mol. Cell. Proteomics* **8,** 443–450 (2009).

140. Pezzulo, A. A. The air-liquid interface and use of primary cell cultures are important to recapitulate the transcriptional profile of in vivo airway epithelia. *Am. J. Physiol. Lung C* **300,** L25–L31 (2011).
141. Clarke, D. C., Morris, M. K. & Lauffenburger, D. A. Normalization and statistical analysis of multiplexed bead-based immunoassay data using mixed-effects modeling. *Mol. Cell. Proteomics* **12,** 245–262 (2013).
142. Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A. & Sorger, P. K. Networks Inferred from Biochemical Data Reveal Profound Differences in Toll-like Receptor and Inflammatory Signaling between Normal and Transformed Hepatocytes. *Mol. Cell. Proteomics* **9,** 1849–1865 (2010).
143. Lamb, J. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science (80-.).* **313,** 1929–1935 (2006).
144. Iorio, F. Discovery of drug mode of action and drug repositioning from transcriptional responses. *Proc. Natl Acad. Sci. Usa* **107,** 14621–14626 (2010).
145. Perrin, B.-E. *et al.* Gene networks inference using dynamic Bayesian networks. *Bioinformatics* **19 Suppl 2,** ii138-48 (2003).
146. Margolin, A. A. *et al.* ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **7,** S7 (2006).
147. Bonneau, R. *et al.* The Inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biol.* **7,** R36 (2006).
148. Rui Xu, R., Wunsch, D. C. & Frank, R. L. Inference of Genetic Regulatory Networks with Recurrent Neural Network Models Using Particle Swarm Optimization. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* **4,** 681–692 (2007).
149. Chen, T., He, H. L. & Church, G. M. Modeling gene expression with differential equations. *Pac. Symp. Biocomput.* 29–40 (1999). at <http://www.ncbi.nlm.nih.gov/pubmed/10380183>
150. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat.*

Methods **9,** 796–804 (2012).

151. De Smet, R. & Marchal, K. Advantages and limitations of current network inference methods. *Nat. Rev. Microbiol.* **8,** 717–729 (2010).
152. STOLOVITZKY, G., MONROE, D. & CALIFANO, A. Dialogue on Reverse-Engineering Assessment and Methods: The DREAM of High-Throughput Pathway Inference. *Ann. N. Y. Acad. Sci.* **1115,** 1–22 (2007).
153. Moult, J., Pedersen, J. T., Judson, R. & Fidelis, K. A large-scale experiment to assess protein structure prediction methods. *Proteins Struct. Funct. Genet.* **23,** ii–iv (1995).
154. Radivojac, P. *et al.* A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10,** 221–227 (2013).
155. Janin, J. *et al.* CAPRI: A Critical Assessment of PRedicted Interactions. *Proteins Struct. Funct. Genet.* **52,** 2–9 (2003).
156. Aghaeepour, N. *et al.* Critical assessment of automated flow cytometry data analysis techniques. *Nat. Methods* **10,** 228–238 (2013).
157. Cooper, S. *et al.* Predicting protein structures with a multiplayer online game. *Nature* **466,** 756–760 (2010).
158. Meyer, P. *et al.* Industrial methodology for process verification in research (IMPROVER): toward systems biology verification. *Bioinformatics* **28,** 1193–1201 (2012).
159. Rhrissorrakrai, K. *et al.* Understanding the limits of animal models as predictors of human biology: lessons learned from the sbv IMPROVER Species Translation Challenge. *Bioinformatics* **31,** 471–483 (2015).
160. Bilal, E. *et al.* Improving Breast Cancer Survival Analysis through Competition-Based Multidimensional Modeling. *PLoS Comput. Biol.* **9,** e1003047 (2013).
161. Margolin, A. A. *et al.* Systematic Analysis of Challenge-Driven Improvements in Molecular Prognostic Models for Breast Cancer. *Sci. Transl. Med.* **5,** 181re1-181re1 (2013).
162. Jaccard, P. THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1. *New Phytol.* **11,** 37–50 (1912).
163. Kirouac, D. C. *et al.* Creating and analyzing pathway and protein interaction compendia for modelling signal transduction networks. *BMC Syst. Biol.* **6,** 29

(2012).

164. Jiang, C., Xuan, Z., Zhao, F. & Zhang, M. Q. TRED: a transcriptional regulatory element database, new entries and other development. *Nucleic Acids Res.* **35,** D137–D140 (2007).
165. Zhao, F., Xuan, Z., Liu, L. & Zhang, M. Q. TRED: a Transcriptional Regulatory Element Database and a platform for in silico gene regulation studies. *Nucleic Acids Res.* **33,** D103–D107 (2004).
166. Poussin, C. *et al.* The species translation challenge - A systems biology perspective on human and rat bronchial epithelial cells. *Sci. Data* **1,** (2014).
167. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* **102,** 15545–15550 (2005).
168. Prill, R. J. *et al.* Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges. *PLoS One* **5,** e9202 (2010).
169. Tarca, A. L. *et al.* Strengths and limitations of microarray-based phenotype prediction: lessons learned from the IMPROVER Diagnostic Signature Challenge. *Bioinformatics* **29,** 2892–2899 (2013).
170. Johnston, J. A. *et al.* Interleukins 2, 4, 7, and 15 stimulate tyrosine phosphorylation of insulin receptor substrates 1 and 2 in T cells. Potential role of JAK kinases. *J. Biol. Chem.* **270,** 28527–30 (1995).
171. Calegari, V. C. *et al.* Suppressor of Cytokine Signaling 3 Is Induced by Angiotensin II in Heart and Isolated Cardiomyocytes, and Participates in Desensitization. *Endocrinology* **144,** 4586–4596 (2003).
172. Park, E. S. *et al.* Thyrotropin Induces SOCS-1 (Suppressor of Cytokine Signaling-1) and SOCS-3 in FRTL-5 Thyroid Cells. *Mol. Endocrinol.* **14,** 440–448 (2000).
173. Pestka, S., Krause, C. D. & Walter, M. R. Interferons, interferon-like cytokines, and their receptors. *Immunol. Rev.* **202,** 8–32 (2004).
174. Villoslada, P. & Baranzini, S. Data integration and systems biology approaches for biomarker discovery: Challenges and opportunities for multiple sclerosis. *J. Neuroimmunol.* **248,** 58–65 (2012).
175. Hauser, S. L., Chan, J. R. & Oksenberg, J. R. Multiple sclerosis: Prospects

and promise. *Ann. Neurol.* **74,** 317–327 (2013).

176. Villoslada, P., Steinman, L. & Baranzini, S. E. Systems biology and its application to the understanding of neurological diseases. *Ann. Neurol.* **65,** 124–39 (2009).
177. Clermont, G. *et al.* Bridging the gap between systems biology and medicine. *Genome Med.* **1,** 88 (2009).
178. International Multiple Sclerosis Genetics Consortium, S. *et al.* Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. *Nature* **476,** 214–9 (2011).
179. International Multiple Sclerosis Genetics Consortium (IMSGC), A. H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nat. Genet.* **45,** 1353–60 (2013).
180. Baranzini, S. E. *et al.* Network-Based Multiple Sclerosis Pathway Analysis with GWAS Data from 15,000 Cases and 30,000 Controls. *Am. J. Hum. Genet.* **92,** 854–865 (2013).
181. Palacios, R. *et al.* A network analysis of the human T-cell activation gene network identifies JAGGED1 as a therapeutic target for autoimmune diseases. *PLoS One* **2,** e1222 (2007).
182. Zhang, Y. *et al.* Notch1 signaling plays a role in regulating precursor differentiation during CNS remyelination. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 19162–7 (2009).
183. Bassil, R. *et al.* Notch ligand delta-like 4 blockade alleviates experimental autoimmune encephalomyelitis by promoting regulatory T cell development. *J. Immunol.* **187,** 2322–8 (2011).
184. Ascherio, A. & Marrie, R. A. Vitamin D in MS: A vitamin for 4 seasons. *Neurology* **79,** 208–210 (2012).
185. Lassmann, H. Axonal and neuronal pathology in multiple sclerosis: what have we learnt from animal models. *Exp. Neurol.* **225,** 2–8 (2010).
186. Mähler, A. *et al.* Epigallocatechin-3-gallate: a useful, effective and safe clinical approach for targeted prevention and individualised treatment of neurological diseases? *EPMA J.* **4,** 5 (2013).
187. Sühs, K.-W. *et al.* A randomized, double-blind, phase 2 study of erythropoietin

in optic neuritis. *Ann. Neurol.* **72,** 199–210 (2012).

188. Moreno, B. *et al.* Differential neuroprotective effects of 5'-deoxy-5'-methylthioadenosine. *PLoS One* **9,** e90671 (2014).
189. Rudick, R. A., Mi, S. & Sandrock, A. W. LINGO-1 antagonists as therapy for multiple sclerosis: in vitro and in vivo evidence. *Expert Opin. Biol. Ther.* **8,** 1561–70 (2008).
190. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444,** 643–646 (2006).
191. Blüthgen, N. & Legewie, S. Robustness of signal transduction pathways. *Cell. Mol. Life Sci.* **70,** 2259–2269 (2013).
192. Martinez-Forero, I. *et al.* IL-10 suppressor activity and ex vivo Tr1 cell function are impaired in multiple sclerosis. *Eur. J. Immunol.* **38,** 576–86 (2008).
193. Arnett, H. A. *et al.* TNFα promotes proliferation of oligodendrocyte progenitors and remyelination. *Nat. Neurosci.* **4,** 1116–1122 (2001).
194. TNF neutralization in MS: results of a randomized, placebo-controlled multicenter study. The Lenercept Multiple Sclerosis Study Group and The University of British Columbia MS/MRI Analysis Group. *Neurology* **53,** 457–65 (1999).
195. Aldridge, B. B., Saez-Rodriguez, J., Muhlich, J. L., Sorger, P. K. & Lauffenburger, D. A. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput. Biol.* **5,** e1000340 (2009).
196. Kholodenko, B. N., Hancock, J. F. & Kolch, W. Signalling ballet in space and time. *Nat. Rev. Mol. Cell Biol.* **11,** 414–26 (2010).
197. Morris, M. K., Chi, A., Melas, I. N. & Alexopoulos, L. G. Phosphoproteomics in drug discovery. *Drug Discov. Today* **19,** 425–432 (2014).
198. Saez-Rodriguez, J. *et al.* Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. *Cancer Res.* **71,** 5400–11 (2011).
199. Kholodenko, B., Yaffe, M. B. & Kolch, W. Computational Approaches for Analyzing Information Flow in Biological Networks. *Sci. Signal.* **5,** re1-re1 (2012).
200. MacNamara, A., Terfve, C., Henriques, D., Bernabé, B. P. & Saez-Rodriguez,

J. State–time spectrum of signal transduction logic models. *Phys. Biol.* **9,** 45003 (2012).

201. Tegnér, J. N. *et al.* Computational disease modeling – fact or fiction? *BMC Syst. Biol.* **3,** 56 (2009).
202. Spencer, S. L. & Sorger, P. K. Measuring and Modeling Apoptosis in Single Cells. *Cell* **144,** 926–939 (2011).
203. Gomez-Cabrero, D., Compte, A. & Tegner, J. Workflow for generating competing hypothesis from models with parameter uncertainty. *Interface Focus* **1,** 438–49 (2011).
204. Pertsovskaya, I., Abad, E., Domedel-Puig, N., Garcia-Ojalvo, J. & Villoslada, P. Transient oscillatory dynamics of interferon beta signaling in macrophages. *BMC Syst. Biol.* **7,** 59 (2013).
205. Scannell, J. W., Blanckley, A., Boldon, H. & Warrington, B. Diagnosing the decline in pharmaceutical R&D efficiency. *Nat. Rev. Drug Discov.* **11,** 191–200 (2012).
206. Conway, D. & Cohen, J. A. Combination therapy in multiple sclerosis. *Lancet. Neurol.* **9,** 299–308 (2010).
207. Fitzgerald, J. B., Schoeberl, B., Nielsen, U. B. & Sorger, P. K. Systems biology and combination therapy in the quest for clinical efficacy. *Nat. Chem. Biol.* **2,** 458–466 (2006).
208. Lehár, J. *et al.* Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.* **3,** 80 (2007).
209. Zhao, S. *et al.* Systems Pharmacology of Adverse Event Mitigation by Drug Combinations. *Sci. Transl. Med.* **5,** 206ra140-206ra140 (2013).
210. Klipp, E., Wade, R. C. & Kummer, U. Biochemical network-based drug-target prediction. *Curr. Opin. Biotechnol.* **21,** 511–516 (2010).
211. El-Serag, H. B. Hepatocellular Carcinoma. *N. Engl. J. Med.* **365,** 1118–1127 (2011).
212. Villanueva, A., Hernandez-Gea, V. & Llovet, J. M. Medical therapies for hepatocellular carcinoma: a critical view of the evidence. *Nat. Rev. Gastroenterol. Hepatol.* **10,** 34–42 (2013).
213. Singh, S., Singh, P. P., Roberts, L. R. & Sanchez, W. Chemopreventive

strategies in hepatocellular carcinoma. *Nat. Rev. Gastroenterol. Hepatol.* **11,** 45–54 (2014).

214. Gupta, S. K. *et al.* Estimation of intraocular pressure in rabbits using Non Contact Tonometer: A comparative evaluation with Schiotz Tonometer. *Methods Find. Exp. Clin. Pharmacol.* **29,** 405 (2007).
215. Darvesh, A. S. & Bishayee, A. Chemopreventive and Therapeutic Potential of Tea Polyphenols in Hepatocellular Cancer. *Nutr. Cancer* **65,** 329–344 (2013).
216. Candido, J. & Hagemann, T. Cancer-Related Inflammation. *J. Clin. Immunol.* **33,** 79–84 (2013).
217. Hanahan, D. & Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **144,** 646–674 (2011).
218. Berasain, C. *et al.* Inflammation and Liver Cancer. *Ann. N. Y. Acad. Sci.* **1155,** 206–221 (2009).
219. Aravalli, R. N., Steer, C. J. & Cressman, E. N. K. Molecular mechanisms of hepatocellular carcinoma. *Hepatology* **48,** 2047–2063 (2008).
220. Kalra, E. K. Nutraceutical--definition and introduction. *AAPS PharmSci* **5,** E25 (2003).
221. Lee, M. S. & Jang, H.-S. Two case reports of the acute effects of Qi therapy (external Qigong) on symptoms of cancer: short report. *Complement. Ther. Clin. Pract.* **11,** 211–213 (2005).
222. Wang, S., Wilkes, M. C., Leof, E. B. & Hirschberg, R. Imatinib mesylate blocks a non-Smad TGF-β pathway and reduces renal fibrogenesis in vivo. *FASEB J.* **19,** 1–11 (2005).
223. Denard, B., Lee, C. & Ye, J. Doxorubicin blocks proliferation of cancer cells through proteolytic activation of CREB3L1. *Elife* **1,** e00090 (2012).
224. Ma, H., Deacon, S. & Horiuchi, K. The challenge of selecting protein kinase assays for lead discovery optimization. *Expert Opin. Drug Discov.* **3,** 607–621 (2008).
225. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40,** D1100–D1107 (2012).
226. Wang, Y. *et al.* PubChem BioAssay: 2014 update. *Nucleic Acids Res.* **42,** D1075–D1082 (2014).

227. Alexopoulos, L. G., Saez-Rodriguez, J., Cosgrove, B. D., Lauffenburger, D. A. & Sorger, P. K. Networks Inferred from Biochemical Data Reveal Profound Differences in Toll-like Receptor and Inflammatory Signaling between Normal and Transformed Hepatocytes. *Mol. Cell. Proteomics* **9,** 1849–1865 (2010).
228. Hatziapostolou, M., Katsoris, P. & Papadimitriou, E. Different inhibitors of plasmin differentially affect angiostatin production and angiogenesis. *Eur. J. Pharmacol.* **460,** 1–8 (2003).
229. Loutrari, H., Magkouta, S., Papapetropoulos, A. & Roussos, C. Mastic Oil Inhibits the Metastatic Phenotype of Mouse Lung Adenocarcinoma Cells. *Cancers (Basel).* **3,** 789–801 (2011).
230. Melas, I. N., Mitsos, A., Messinis, D. E., Weiss, T. S. & Alexopoulos, L. G. Combined logical and data-driven models for linking signalling pathways to cellular response. *BMC Syst. Biol.* **5,** 107 (2011).
231. Melas, I. N. *et al.* Construction of large signaling pathways using an adaptive perturbation approach with phosphoproteomic data. *Mol. Biosyst.* **8,** 1571 (2012).
232. Michailidou, M. *et al.* Microvascular Endothelial Cell Responses in vitro and in vivo: Modulation by Zoledronic Acid and Paclitaxel? *J. Vasc. Res.* **47,** 481–493 (2010).
233. Ban, K. C., Singh, H., Krishnan, R. & Seow, H. F. GSK-3beta phosphorylation and alteration of beta-catenin in hepatocellular carcinoma. *Cancer Lett.* **199,** 201–8 (2003).
234. Polakis, P. Drugging Wnt signalling in cancer. *EMBO J.* **31,** 2737–2746 (2012).
235. Newman, D. J. & Cragg, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **75,** 311–335 (2012).
236. Capece, D. *et al.* The Inflammatory Microenvironment in Hepatocellular Carcinoma: A Pivotal Role for Tumor-Associated Macrophages. *Biomed Res. Int.* **2013,** 1–15 (2013).
237. Keeley, E. C., Mehrad, B. & Strieter, R. M. Chemokines as mediators of tumor angiogenesis and neovascularization. *Exp. Cell Res.* **317,** 685–690 (2011).
238. O'Hayre, M., Salanga, C. L., Handel, T. M. & Allen, S. J. Chemokines and cancer: migration, intracellular signalling and intercellular communication in

the microenvironment. *Biochem. J.* **409,** 635–649 (2008).

239. Liu, Y.-Q. *et al.* Desensitization of T lymphocyte function by CXCR3 ligands in human hepatocellular carcinoma. *World J. Gastroenterol.* **11,** 164–70 (2005).
240. Monnier, J. *et al.* CXCR7 is up-regulated in human and murine hepatocellular carcinoma and is specifically expressed by endothelial cells. *Eur. J. Cancer* **48,** 138–148 (2012).
241. Hallett, M. A., Venmar, K. T. & Fingleton, B. Cytokine Stimulation of Epithelial Cancer Cells: The Similar and Divergent Functions of IL-4 and IL-13. *Cancer Res.* **72,** 6338–6343 (2012).
242. Li, J. *et al.* Interleukin 17A Promotes Hepatocellular Carcinoma Metastasis via NF-kB Induced Matrix Metalloproteinases 2 and 9 Expression. *PLoS One* **6,** e21816 (2011).
243. Ma, S. *et al.* IL-17A Produced by T Cells Promotes Tumor Growth in Hepatocellular Carcinoma. *Cancer Res.* **74,** 1969–1982 (2014).
244. Qu, C.-K. Role of the SHP-2 tyrosine phosphatase in cytokine-induced signaling and cellular response. *Biochim. Biophys. Acta* **1592,** 297–301 (2002).
245. Chen, J. *et al.* Involvement of PI3K/PTEN/AKT/mTOR pathway in invasion and metastasis in hepatocellular carcinoma: Association with MMP-9. *Hepatol. Res.* **39,** 177–186 (2009).
246. Wen, A. Y., Sakamoto, K. M. & Miller, L. S. The Role of the Transcription Factor CREB in Immune Function. *J. Immunol.* **185,** 6413–6419 (2010).
247. Matsushima-Nishiwaki, R. *et al.* Phosphorylated Heat Shock Protein 27 Represses Growth of Hepatocellular Carcinoma via Inhibition of Extracellular Signal-regulated Kinase. *J. Biol. Chem.* **283,** 18852–18860 (2008).
248. Yasuda, E. *et al.* Attenuated phosphorylation of heat shock protein 27 correlates with tumor progression in patients with hepatocellular carcinoma. *Biochem. Biophys. Res. Commun.* **337,** 337–342 (2005).
249. Topisirovic, I. & Sonenberg, N. mRNA Translation and Energy Metabolism in Cancer: The Role of the MAPK and mTORC1 Pathways. *Cold Spring Harb. Symp. Quant. Biol.* **76,** 355–367 (2011).
250. Lea, C. H. Chemical and nutritional aspects of oxidised and heated fats. *Chem. Ind.* **6,** 244–8 (1965).

251. Hernandez–Gea, V., Toffanin, S., Friedman, S. L. & Llovet, J. M. Role of the Microenvironment in the Pathogenesis and Treatment of Hepatocellular Carcinoma. *Gastroenterology* **144,** 512–527 (2013).
252. Zeng, L., Tang, W. J., Yin, J. J. & Zhou, B. J. Signal transductions and nonalcoholic fatty liver: a mini-review. *Int. J. Clin. Exp. Med.* **7,** 1624–31 (2014).
253. Lu, S., Shen, K. C., Wang, Y., Brooks, S. C. & Wang, Y. A. Impaired hepatocyte survival and liver regeneration in Atm-deficient mice. *Hum. Mol. Genet.* **14,** 3019–3025 (2005).
254. Smith, B. W. & Adams, L. A. Non-alcoholic fatty liver disease. *Crit. Rev. Clin. Lab. Sci.* **48,** 97–113 (2011).
255. Willebrords, J. *et al.* Strategies, models and biomarkers in experimental non-alcoholic fatty liver disease research. *Prog. Lipid Res.* **59,** 106–125 (2015).
256. Kumashiro, N. *et al.* Cellular mechanism of insulin resistance in nonalcoholic fatty liver disease. *Proc. Natl. Acad. Sci.* **108,** 16381–16385 (2011).
257. Pettinelli, P., Obregón, A. M. & Videla, L. A. Molecular mechanisms of steatosis in nonalcoholic fatty liver disease. *Nutr. Hosp.* **26,** 441–50
258. Than, N. N. & Newsome, P. N. A concise review of non-alcoholic fatty liver disease. *Atherosclerosis* **239,** 192–202 (2015).
259. Amacher, D. & Chalasani, N. Drug-Induced Hepatic Steatosis. *Semin. Liver Dis.* **34,** 205–214 (2014).
260. Verrotti, A., D'Egidio, C., Mohn, A., Coppola, G. & Chiarelli, F. Weight gain following treatment with valproic acid: pathogenetic mechanisms and clinical implications. *Obes. Rev.* **12,** e32–e43 (2011).
261. Aires, C. C. P. *et al.* Inhibition of hepatic carnitine palmitoyl-transferase I (CPT IA) by valproyl-CoA as a possible mechanism of valproate-induced steatosis. *Biochem. Pharmacol.* **79,** 792–799 (2010).
262. Dusman, R. E. *et al.* Clinical features of amiodarone-induced pulmonary toxicity. *Circulation* **82,** 51–9 (1990).
263. Varbiro, G. *et al.* Concentration dependent mitochondrial effect of amiodarone. *Biochem. Pharmacol.* **65,** 1115–28 (2003).
264. Fromenty, B. *et al.* Amiodarone inhibits the mitochondrial beta-oxidation of

fatty acids and produces microvesicular steatosis of the liver in mice. *J. Pharmacol. Exp. Ther.* **255,** 1371–6 (1990).

265. Felser, A., Blum, K., Lindinger, P. W., Bouitbir, J. & Krähenbühl, S. Mechanisms of Hepatocellular Toxicity Associated with Dronedarone—A Comparison to Amiodarone. *Toxicol. Sci.* **131,** 480–490 (2013).
266. Guigui, B. *et al.* Amiodarone-induced hepatic phospholipidosis: a morphological alteration independent of pseudoalcoholic liver disease. *Hepatology* **8,** 1063–8
267. Yin, H.-Q. *et al.* Hepatic Gene Expression Profiling and Lipid Homeostasis in Mice Exposed to Steatogenic Drug, Tetracycline. *Toxicol. Sci.* **94,** 206–216 (2006).
268. Larosche, I. *et al.* Tamoxifen Inhibits Topoisomerases, Depletes Mitochondrial DNA, and Triggers Steatosis in Mouse Liver. *J. Pharmacol. Exp. Ther.* **321,** 526–535 (2007).
269. Donato, M. T. *et al.* Cytometric analysis for drug-induced steatosis in HepG2 cells. *Chem. Biol. Interact.* **181,** 417–423 (2009).
270. Fujimura, H., Murakami, N., Kurabe, M. & Toriumi, W. *In vitro* assay for drug-induced hepatosteatosis using rat primary hepatocytes, a fluorescent lipid analog and gene expression analysis. *J. Appl. Toxicol.* **29,** 356–363 (2009).
271. Amacher, D. E. & Martin, B. A. Tetracycline-induced steatosis in primary canine hepatocyte cultures. *Fundam. Appl. Toxicol.* **40,** 256–63 (1997).
272. Wruck, W. *et al.* Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. *Sci. Data* **2,** 150068 (2015).
273. Kotelnikova, E. *et al.* Signaling networks in MS: A systems-based approach to developing new pharmacological therapies. *Mult. Scler. J.* 1352458514543339- (2014). at <http://msj.sagepub.com/content/early/2014/08/04/1352458514543339.abstract>
274. Michailidou, M. *et al.* Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action. *CPT Pharmacometrics Syst. Pharmacol.* **4,** 350–361 (2015).
275. Godoy, P. *et al.* Gene networks and transcription factor motifs defining the

differentiation of stem cells into hepatocyte-like cells. *J. Hepatol.* **63,** 934–942 (2015).

276. O'Brien, J., Wilson, I., Orton, T. & Pognan, F. Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity. *Eur. J. Biochem.* **267,** 5421–5426 (2000).
277. Poussin, C. *et al.* The species translation challenge-a systems biology perspective on human and rat bronchial epithelial cells. *Sci. data* **1,** 140009 (2014).
278. R Foundation for Statistical Computing. R: A language and environment for statistical computing. (2016). at <https://www.r-project.org/>
279. Moylan, C. A. *et al.* Hepatic gene expression profiles differentiate presymptomatic patients with mild versus severe nonalcoholic fatty liver disease. *Hepatology* **59,** 471–82 (2014).
280. Mi, H., Poudel, S., Muruganujan, A., Casagrande, J. T. & Thomas, P. D. PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res.* **44,** D336-42 (2016).
281. Brunt, E. M. *et al.* Nonalcoholic fatty liver disease. *Nat. Rev. Dis. Prim.* **1,** 15080 (2015).
282. Videla, L. A. *et al.* Liver NF-κB and AP-1 DNA Binding in Obese Patients. *Obesity* **17,** 973–979 (2009).
283. dela Peña, A. *et al.* NF-κB Activation, Rather Than TNF, Mediates Hepatic Inflammation in a Murine Dietary Model of Steatohepatitis. *Gastroenterology* **129,** 1663–1674 (2005).
284. Tailleux, A., Wouters, K. & Staels, B. Roles of PPARs in NAFLD: Potential therapeutic targets. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* **1821,** 809–818 (2012).
285. Ueyama, A. *et al.* Inhibition of MEK1 Signaling Pathway in the Liver Ameliorates Insulin Resistance. *J. Diabetes Res.* **2016,** 8264830 (2016).
286. Wu, H.-T. *et al.* The role of Hepassocin in the development of non-alcoholic fatty liver disease. *J. Hepatol.* **59,** 1065–1072 (2013).
287. Panasiuk, A., Dzieciol, J., Panasiuk, B. & Prokopowicz, D. Expression of p53, Bax and Bcl-2 proteins in hepatocytes in non-alcoholic fatty liver disease.

World J. Gastroenterol. **12,** 6198–202 (2006).

288. Herzig, S. *et al.* CREB controls hepatic lipid metabolism through nuclear hormone receptor PPAR-γ. *Nature* **426,** 190–193 (2003).
289. Erion, D. M. *et al.* Prevention of hepatic steatosis and hepatic insulin resistance by knockdown of cAMP response element-binding protein. *Cell Metab.* **10,** 499–506 (2009).
290. Wang, R., Wang, X. & Zhuang, L. Gene expression profiling reveals key genes and pathways related to the development of non-alcoholic fatty liver disease. **15,** 190–199 (2016).
291. Hasebe, T. *et al.* Bone morphogenetic protein-binding endothelial regulator of liver sinusoidal endothelial cells induces iron overload in a fatty liver mouse model. *J. Gastroenterol.* **52,** 341–351 (2017).
292. Byon, C. H. *et al.* Free fatty acids enhance breast cancer cell migration through plasminogen activator inhibitor-1 and SMAD4. *Lab. Invest.* **89,** 1221–8 (2009).
293. Reis, S. E. *et al.* Cardiovascular effects of tamoxifen in women with and without heart disease: breast cancer prevention trial. National Surgical Adjuvant Breast and Bowel Project Breast Cancer Prevention Trial Investigators. *J. Natl. Cancer Inst.* **93,** 16–21 (2001).
294. Lee, S.-H., Zahoor, M., Hwang, J.-K., Min, D. S. & Choi, K.-Y. Valproic acid induces cutaneous wound healing in vivo and enhances keratinocyte motility. *PLoS One* **7,** e48791 (2012).
295. Haroon, Z. A., Amin, K., Jiang, X. & Arcasoy, M. O. A Novel Role for Erythropoietin During Fibrin-Induced Wound-Healing Response. *Am. J. Pathol.* **163,** 993–1000 (2003).
296. Sookoian, S. *et al.* Heat Shock Protein 27 is down-regulated in Ballooned Hepatocytes of Patients with Nonalcoholic Steatohepatitis (NASH). *Sci. Rep.* **6,** 22528 (2016).

Παράρτημα Ι: Χρονοδιάγραμμα Δημοσιεύσεων | Publication Timeline

Green: Peer-reviewed publications

Blue: Abstracts in international conferences

[P1] 10 February 2014, Osteoarthritis and Cartilage

Modeling of signaling pathways in chondrocytes based on phosphoproteomic and cytokine release data

[A1] 17-20 May 2014, Prague, Czech Republic

Meniscus – Cartilage paracrine crosstalk in osteoarthritis

[P2] 10 June 2014, Nature Scientific Data

The species translation challenge—A systems biology perspective on human and rat bronchial epithelial cells

[A2] 17-19 June 2014, Boston MA, USA

Construction of a drug-induced phosphoprotein/cytokine dataset in clinical samples for Multiple Sclerosis

[A4] 18-20 September 2014, Patras, Greece

An integrated proteomic and metabolomic approach to investigate cartilage degeneration

[P3] 7 October 2014, Bioinformatics

A crow-sourcing approach for the construction of species-specific cell signaling networks

[P4] 1 February 2015, Multiple Sclerosis Journal

Signaling networks in MS: A systems-based approach to developing new pharmacological therapies

[A6] 30 April 3 May 2015, Seattle, WA, USA
The role of inflammatory mediators in meniscus and cartilage crosstalk in osteoarthritis

[A7] 5-8 May 2015, Berlin, Germany
Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis

[P5] 1 June 2015, CPT: Pharmacometrics & Systems Pharmacology
Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action

[A8] 8-9 June 2015, Berlin, Germany
Multiplex High Throughput screening for Drug Discovery: Application for mechanism-based biomarker discovery in Multiple Sclerosis

[A9] 1-3 July 2015, Avignon, France
Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis

[A10] 6-8 July 2015, Heidelberg, Germany
Patient-specific signalling pathway analysis for Multiple Sclerosis

[A11] 3-5 March 2017, Athens, Greece
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.

[A12] 7 April 2017, Athens, Greece
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.

[A13] 7-8 November 2017, Amsterdam, Netherlands

Development of a new drug repositioning platform for Non-Alcoholic Fatty Liver disease through network analysis.

[A14] 9-11 November 2017, Rome, Italy

Network based drug repositioning for Non-Alcoholic Fatty Liver Disease

[P6] 17 January 2018, CPT: Pharmacometrics & Systems Pharmacology

Translational systems pharmacology-based predictive assessment of drug-induced cardiomyopathy

Παράρτημα ΙΙ: Περιλήψεις σε Διεθνή Συνέδρια | Abstracts in International Conferences

[A1] 17-20 May 2014, Prague, Czech Republic
European Calcified Tissue Society Conference
Meniscus – Cartilage paracrine crosstalk in osteoarthritis
Published in "Bone Abstracts" DOI:10.1530/boneabs.3.PP183
Authors: Stavroula Samara, Elisavet Chatzopoulou, Ioannis Melas, Dimitris Messinis, Zoe Dailiana, Panagoula Kollia, Leonidas Alexopoulos

Introduction: Meniscus plays an essential role in knee joint function providing stability and load transmission. In osteoarthritis (OA), a joint disease characterized by chronic synovitis and cartilage degeneration, pathological changes in the menisci are observed. However, whether menisci contribute to the progression of OA, the underlying mechanism for meniscus-cartilage communication is still unclear. In this study we analyzed systematically the response of meniscus and cartilage explants to a number of inflammatory mediators, in order to reveal their response similarity and highlight potential crosstalks and interactions.
Methods: OA cartilage and the lateral meniscus were harvested from two patients undergoing total knee arthroplasty. Meniscus and cartilage disks (3 mm diameter) were stimulated with inflammatory mediators ((IL-1α, IL-1β, IL-12α, CSF2) (50 ng/ml), (TNF-α, IL-6, CXCL7, IL-8, CCL2, CXCL10, IFN-γ, IL-3, MIA2, IL-4) (100 ng/ml) and GROα (500 ng/ml)) for 24 h. For each condition the release of different proteins (IL-1α, IL-1β, TNF-α, IL-6, CXCL7, GROα, IL-8, CCL2, CXCL10, IFN-γ, IL-3, IL-12α, MIA-2, CSF2, IL-4) was measured in the supernatant using custom multiplexed assays on a Luminex FlexMap 3D instrument.
Results: In both tissues the major inflammatory players (IL-1α, IL-1β, TNFα) were the strongest stimuli as expected. Meniscus responses were the same up to 73 and 50% with the cartilage ones for the first and the second donors respectively. Interestingly, meniscus under certain stimuli (IL-1α, IFN-γ, CSF2, IL-8) responded differently than cartilage by releasing five different cytokines while cartilage did not.

Conclusions: Our results indicate that meniscus is affected by its inflammatory environment and responds to it as actively as cartilage. Moreover, the release of different cytokines from meniscus and cartilage suggests that meniscus can be an active player in the progression of OA. These data support the hypothesis that significant crosstalk between these two knee compartments exist and anti-inflammatory therapies should take into consideration both tissues.

[A2] 17-19 June 2014, Boston MA, USA
International Conference on Systems Biology of Human Disease
Construction of a drug-induced phosphoprotein/cytokine dataset in clinical samples for Multiple Sclerosis
Authors: Vicky Pliaka, Dimitris E Messinis, Theodore Sakellaropoulos, Ekaterina Kotelnikova, Tomas Olsson, Jesper Tegner, Roland Martin, Dimitris Tzeranis, Friedemann Paul, Julio Saez-Rodiguez, Marti Bernardo-Faura, Ioannis N Melas, Jose Manuel Mas, Laura Artigas, Elena Schwartz, Ilya Mazo, Sophia Stamatatou, Mar Masso, Albert Zamora, Pablo Villoslada, Leonidas G Alexopoulos

Multiple Sclerosis (MS) is an autoimmune disease that affects the brain and spinal cord. An estimated 2,500,000 around the world have MS and there is not yet a cure for the disease. Even though significant progress is currently being made in MS research, the pathogenesis of the disease has not been comprehensively understood. A great number of pathological mechanisms responsible for the disease have been described, involving hundreds of genes and proteins altering multiple processes and signaling pathways. By understanding how current MS therapies work in biological networks, more effective therapies can be designed. On this front, the CombiMS consortium (http://combims.eu/) is developing computational and experimental tools to improve the therapeutic options of MS in the future.
A milestone in this consortium is to evaluate how current and “promising” MS drugs work at the signaling level in different patient populations. Peripheral blood mononuclear cells (PBMCs) from approximately 255 donors were collected from several European medical centers. The PBMCs were plated in 96 well plates and 20 stimuli & drugs were applied. Using custom multiplex assays, 17 phosphoproteins

plus 2 control beads were measured at 5 and 25 minutes and 22 cytokines plus two control beads at 24 hours post stimulus.

To ensure the highest possible data quality, a kit was developed to include all reagents needed in order to isolate PBMCs from one donor and then plate, stimulate and lyse them. Sample collection controls were applied in the stimuli set to evaluate errors in sample processing whereas two custom xMAP bead sets were used to evaluate errors in the bead-based ELISA procedure and instrument measuring variability.

Phosphoproteomic and cytokine data will be combined with SNP data and clinical profiles (i.e. responders, non-responders, therapeutic intervention) in a computational framework which will help to understand MS more thoroughly and systematically. As a first step, the phosphoproteomic dataset will be used for the construction of a detailed map of the signaling pathway differences between MS and healthy donors, which can help generate a model of MS pathogenesis and improve our understanding of the disease.

[A4] 18-20 September 2014, Patras, Greece

MET-GR III workshop: Metabolic and Protein Network analysis in Systems Biology

An integrated proteomic and metabolomic approach to investigate cartilage degeneration

Authors: Dimitris E Messinis, Vaia Pliaka, Stavroula Samara, Zoe Dailiana, Panagoula Kollia, Maria I Klapa, Leonidas G Alexopoulos

Cartilage degeneration as a feature of osteoarthritis (OA) is one of the most common causes of pain and disability in middle-aged and older people. The percentage of population above 45 years of age affected with OA is estimated to increase from 26.6% (2012) to 29.5% by 2032 due to the aging of the population and the obesity epidemic. Most research studies in OA focus on single therapeutic targets, disease processes or level of molecules, thus omics data have never been integrated in a systematic way.

Articular cartilage and synovium explants, whole cell extract of articular chondrocytes or mesenchymal stem cells, supernatant of articular cartilage explants, articular

chondrocytes, osteoclasts or synovium-derived cells in culture, articular cartilage vesicles, synovial fluid (SF), plasma, serum and urine have all been studied on the proteomic front and a subset of those for their metabolomic signature. However, no more than three of those types of samples have been examined in a single study using the same assay type. By using samples from various tissues, we avoid focusing only on the biochemical changes that occur in the joints.

As a case study, we combined proteomic and metabolomic assays (multiplex bead-based sandwich ELISA xMAP technology and gas chromatography mass spectrometry), to measure the phosphoproteomic signature, cytokine release and metabolomic signature of untreated and treated with IL1B cultured chondrocytes, along with the proteomic and metabolomic signature of the same donor's synovial fluid and plasma.

Following this multi-omics approach, we had to include additional quality control check points to ensure the highest possible dataset quality and therefore the validity of results. In particular, we need to be able to address challenges such as acquiring multiple samples from the same donor at the same time, using the exact same samples for all experiments (cell line passage, timelines in cell culture), merging protocols of all omics assays while meeting the sample handling prerequisites for every assay and having strictly fixed parameters across assays, such as time points and doses of any applied stimuli.

Besides ensuring robustness of the dataset, in the case of OA, we have to take into account the scarce availability of healthy human cartilage, synovium and SF samples and examine the clinical profile data of donors in order to expand the dataset while gaining knowledge about factors such as medication that can alter the omics data.

In conclusion, we believe that omics data acquired by multiple samples while avoiding the above mentioned pitfalls, can be integrated in a multi-level dataset which may help us understand complex biological systems and multifactorial diseases such as OA.

[A6] 30 April - 3 May 2015, Seattle, WA, USA

OARSI World Congress on Osteoarthritis 2015

The role of inflammatory mediators in meniscus and cartilage crosstalk in osteoarthritis

Published in Osteoarthritis and Cartilage DOI: 10.1016/j.joca.2015.02.488

Authors: Z.H. Dailiana, E.I. Chatzopoulou, S. Samara, I.M. Melas, M. Hantes, S. Varitimidis, D. Messinis, P. Kollia, L.G. Alexopoulos.

Purpose: Osteoarthritis (OA) is a common degenerative joint disorder causing cartilage degradation, pain and disability. Lately it is characterized as a whole joint disease as it affects the structure and functionality of all tissue components, such as the menisci. Meniscus plays an essential role in knee joint function providing stability and load transmission. During OA the balance between catabolic and anabolic processes in the cartilage tissue is disturbed favouring catabolism through the cytokines and Matrix Metalloproteinases (MMPs) that are present in the synovial fluid of patients with mild or severe OA. In this study we examine how meniscus and cartilage explants react to a number of inflammatory mediators, in order to reveal their response similarity and highlight potential crosstalks and interactions. Methods: OA cartilage and the menisci were harvested from patients undergoing total knee arthroplasty. Meniscus and cartilage disks (3 mm diameter) were stimulated with inflammatory mediators [IL-1a, IL-1b, IL12a, CSF2 (50 ng/ml), TNF-a, IL-6, CXCL7, IL-8, CCL2, CXCL10, IFN-g, IL-3, MIA2, IL-4 (100 ng/ml) and GROa (500 ng/ml)] for 24 h. For each condition the release of different proteins [IL-1a, IL-1b, TNF-a, IL-6, CXCL7, GROa, IL8, CCL2, CXCL10, IFN-g, IL-3, IL-12a, MIA-2, CSF2, IL-4] was measured in the supernatant after 24 hours using custom multiplexed assays on a Luminex FlexMap 3D instrument. Histological sections were prepared in order to assess the osteoarthritic grade of the tissues used. Datarail toolbox was used for data handling and visualization.

Results: In general, there was strong similarity between meniscus and cartilage releases; the major inflammatory players (IL-1a, IL-1b, TNFa) were the strongest stimuli as expected, in both tissues. Under these cytokines cartilage disks released IL-6, CXCL7, GROa, IL-8, and CCL2. Meniscus responses were the same up to 73% and 50% with the cartilage ones for the first and the second donors, respectively. Meniscus under certain stimuli (IL-1a, IFN-g, CSF2, IL-8) responded differently than

cartilage by releasing five different cytokines that cartilage did not. Conclusions: Until now, little is known about meniscus reaction to cartilage protein expression or to the high cytokines concentration of the synovial fluid after trauma or OA. Our results confirm that IL-1a, IL-1b and TNF-a, as proinflammatory cytokines, promote systemic inflammation and thus release of IL-6, IL-8 and GROa in both cartilage and meniscus tissue. The above findings suggest that meniscus is affected by its inflammatory environment and responds to it as actively as cartilage, supporting the hypothesis that significant crosstalk between these two knee compartments exists and anti-inflammatory therapies should take into consideration both tissues. Further investigations should be done to unveil the exact communication mechanisms between meniscus and the other joint tissues. Acknowledgments: Funded by European Union (European Social Fund e ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) e Research Funding Program: Thalis. Investing in knowledge society through the European Social Fund.

[A7] 5-8 May 2015, Berlin, Germany
European Pharma Summit
Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis
Authors: Dimitris E Messinis, Vicky Pliaka, Sophia Stamatatou, Theodore Sakellaropoulos, combiMS consortium, Leonidas G Alexopoulos

Multiple Sclerosis (MS) is an autoimmune disease affecting the brain and spinal cord. There is not yet a cure for the disease while 2,5 million people around the world have MS. Several pathological mechanisms for MS have been described, involving alterations in multiple processes and signalling pathways.
Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective therapies can be designed that take into account the uniqueness of each patient’s response in treatment and biomarkers can be developed to stratify patients.

For the production of this dataset, we used High-Throughput Screening (HTS) with custom cell signalling assays (by ProtATonce), that allow screening of thousands of samples at the proteomic and phosphoproteomic level at a very small fraction of the cost compared to off-the-shelf reagents. Strict quality control check points were embedded to the collection-to-measurement pipeline in order to evaluate errors in sample collection, sample preparation, ELISA procedure and instrument variation.
Collection kits were prepared and shipped to 4 clinical centres across EU in order to collect peripheral blood mononuclear cells (PBMCs) from 255 donors. The cells were plated and stimulated at 3 time points with 20 compounds and drugs. We collected cell lysates and cell supernatants to simultaneously quantify in the samples 17 phosphoproteins and 22 secreted proteins respectively.
This dataset was analysed with pathway optimization tools for the construction of detailed signalling pathway maps for each donor which can help reveal the drugs' mode of action and efficacy in correlation with differences in the patient population. The therapeutic potential of compounds that their phosphoproteomic signature matches that of existing drugs will be evaluated and the data will be combined with SNP data and clinical profiles in an attempt to develop biomarkers to distinguish between responder and non-responder to treatment patients.

[A8] 8-9 June 2015, Berlin, Germany
16th Annual Drug Discovery Leaders Summit and 3rd Annual Discovery Chemistry & Drug Design Congress
Multiplex High Throughput screening for Drug Discovery: Application for mechanism-based biomarker discovery in Multiple Sclerosis
Authors: Dimitris E Messinis, Vicky Pliaka, Sophia Stamatatou, Theodore Sakellaropoulos, combiMS consortium, Leonidas G Alexopoulos

High throughput screening (HTS) has become an invaluable tool in drug discovery. Current HTS platforms are based on the measurement of very few reporter assays and lose the mechanistic picture on how compounds work on cells. Multiplex HTS can measure a large number of intracellular and extracellular signals and thus provide mechanism-based insight in biomarker discovery.

In this study, we employed a dual phosphoprotein and cytokine multiplex high throughput screening using a Luminex FlexMap 3D system with custom assays by ProtATonce. Seeking biomarkers in Multiple Sclerosis (MS), Luminex assays were developed to monitor the mode of action of 20 drugs in 255 donors at the signalling (17plex) and cytokine (22plex) level. Collection kits were prepared and shipped to clinical centres across EU in order to collect peripheral blood mononuclear cells (PBMCs) from donors. This dataset is analysed with systems biology algorithms in order to construct patient-specific signalling pathways and develop biomarkers for MS. Machine learning and optimization algorithms were employed to quantify patient-based drug efficacy and predict responders from non-responders. In summary, multiplex high throughput screening and systems biology algorithms introduce a mechanistic insight in biomarker discovery and pave a new way for pharmaceutical research.

[A9] 1-3 July 2015, Avignon, France

51st International Conference on Medicinal Chemistry RICT Drug Discovery and Selection Understanding Targets and Mechanisms

Signalling pathway-based screening for drug discovery: an application in Multiple Sclerosis

Authors: Dimitris E Messinis, Vicky Pliaka, Sophia Stamatatou, Theodore Sakellaropoulos, combiMS consortium, Leonidas G Alexopoulos

Multiple Sclerosis (MS) is an autoimmune disease affecting the brain and spinal cord. There is not yet a cure for the disease while 2,5 million people around the world have MS. Several pathological mechanisms for MS have been described, involving alterations in multiple processes and signalling pathways.

Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective therapies can be designed that take into account the uniqueness of each patient's response in treatment and biomarkers can be developed to stratify patients.

For the production of this dataset, we used High-Throughput Screening (HTS) with custom cell signalling assays (by ProtATonce), that allow screening of thousands of samples at the proteomic and phosphoproteomic level at a very small fraction of the cost compared to off-the-shelf reagents. Strict quality control check points were embedded to the collection-to-measurement pipeline in order to evaluate errors in sample collection, sample preparation, ELISA procedure and instrument variation.
Collection kits were prepared and shipped to 4 clinical centres across EU in order to collect peripheral blood mononuclear cells (PBMCs) from 255 donors. The cells were plated and stimulated at 3 time points with 20 compounds and drugs. We collected cell lysates and cell supernatants to simultaneously quantify in the samples 17 phosphoproteins and 22 secreted proteins respectively.
This dataset was analysed with pathway optimization tools for the construction of detailed signalling pathway maps for each donor which can help reveal the drugs' mode of action and efficacy in correlation with differences in the patient population. The therapeutic potential of compounds that their phosphoproteomic signature matches that of existing drugs will be evaluated and the data will be combined with SNP data and clinical profiles in an attempt to develop biomarkers to distinguish between responder and non-responder to treatment patients.

[A10] 6-8 July 2015, Heidelberg, Germany
International Conference on System Biology of Human Disease
Patient-specific signalling pathway analysis for Multiple Sclerosis
Authors: Dimitris E Messinis, Vicky Pliaka, Sophia Stamatatou, Theodore Sakellaropoulos, combiMS consortium, Leonidas G Alexopoulos

Multiple Sclerosis (MS) is an autoimmune disease affecting the brain and spinal cord. There is not yet a cure for the disease while 2,5 million people around the world have MS. Several pathological mechanisms for MS have been described, involving alterations in multiple processes and signalling pathways.
Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective

therapies can be designed that take into account the uniqueness of each patient's response in treatment and biomarkers can be developed to stratify patients.
Collection kits were prepared and shipped to 4 clinical centres across EU in order to collect peripheral blood mononuclear cells (PBMCs) from 255 donors. The cells were plated and stimulated at 3 time points with 20 compounds and drugs. We collected cell lysates and cell supernatants to simultaneously quantify in the samples 17 phosphoproteins and 22 secreted proteins respectively. Strict quality control check points were embedded to the collection-to-measurement pipeline in order to evaluate errors in sample collection, sample preparation, ELISA procedure and instrument variation.
The resulting dataset was analysed with pathway optimization tools for the construction of detailed signalling pathway maps for each donor which can help reveal the drugs' mode of action and efficacy in correlation with differences in the patient population. The therapeutic potential of compounds that their phosphoproteomic signature matches that of existing drugs will be evaluated and the data will be combined with SNP data and clinical profiles in an attempt to develop biomarkers to distinguish between responder and non-responder to treatment patients.

[A11] 3-5 March 2017, Athens, Greece
Hepato-Pancreato-Biliary Greek Association Conference
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.
Authors: Danai Stella Zareifi, Dimitris E Messinis, Angeliki Minia, Vaia Pliaka, Jan Rožanc, Manoussos M Konstadoulakis, Konstantinos J Bramis, Efstathios A Antoniou, Leonidas G Alexopoulos

Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, with a prevalence of up to 30% in developed countries. NAFLD is defined by the presence of hepatic steatosis in the absence of excess alcohol consumption and represents a spectrum of disease, from simple steatosis and Non-Alcoholic Steatohepatitis (NASH) to fibrosis, cirrhosis and hepatocellular carcinoma. The cause and disease progression mechanisms of NAFLD are still not completely understood.

Researchers have tried to understand the disease through multi-omic approaches but there is no study -to our knowledgeexploring the signal transduction level, while phosphoproteomic measurements are becoming very important in drug discovery. In vitro exposure of primary human hepatocytes to increasing concentrations of free fatty acids (FFA) results in increased intracellular accumulation of lipid droplets similar to those observed in patients with NAFLD and NASH. We prepared a mixture of exogenous FFA in ethanol in molar ratio 1:2 Palmitic Acid: Oleic Acid respectively, as it has been done before in a study using Huh7 cells. However, instead of cancer cells, we cultured primary human hepatocyte cells, isolated from healthy resected human liver tissue. The cells were exposed to increasing concentrations of the FFA mixture (100-1000uM). The formation of intracellular lipid droplets was verified using high content screening; lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. After a 24h treatment, we lysed the cells, measured the protein content of the lysates and adjusted the samples to the same concentration. After pooling 5 replicates of each FFA concentration, we measured 19 phosphorylated protein targets using Luminex technology (multiplex antibody-based ELISA). We found that the phosphorylation level of heat shock protein beta-1 (HSPB1) was three-fold less than the control. Other scientists have recently shown that HSPB1 is down-regulated in NASH. Moreover, we show irregular phosphorylation patterns in IKBA, AKT1, WNK1, FAK1 and STAT6, most of which play a role in NAFLD/NASH mechanism. The present study pinpoints the signaling alterations of FFA-induced NAFLD.

[A12] 7 April 2017, Athens, Greece
Panhellenic Congress of Biomedical Technology
Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics.
Authors: Danai Stella Zareifi, Dimitris E Messinis, Angeliki Minia, Vaia Pliaka, Jan Rožanc, Manoussos M Konstadoulakis, Konstantinos J Bramis, Efstathios A Antoniou, Leonidas G Alexopoulos

Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, with a prevalence of up to 30% in developed countries. NAFLD is defined by

the presence of hepatic steatosis in the absence of excess alcohol consumption and represents a spectrum of disease, from simple steatosis and Non-Alcoholic Steatohepatitis (NASH) to fibrosis, cirrhosis and hepatocellular carcinoma. The cause and disease progression mechanisms of NAFLD are still not completely understood. Researchers have tried to understand the disease through multi-omic approaches but there is no study -to our knowledgeexploring the signal transduction level, while phosphoproteomic measurements are becoming very important in drug discovery. In vitro exposure of primary human hepatocytes to increasing concentrations of free fatty acids (FFA) results in increased intracellular accumulation of lipid droplets similar to those observed in patients with NAFLD and NASH. We prepared a mixture of exogenous FFA in ethanol in molar ratio 1:2 Palmitic Acid: Oleic Acid respectively, as it has been done before in a study using Huh7 cells. However, instead of cancer cells, we cultured primary human hepatocyte cells, isolated from healthy resected human liver tissue. The cells were exposed to increasing concentrations of the FFA mixture (100-1000uM). The formation of intracellular lipid droplets was verified using high content screening; lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. After a 24h treatment, we lysed the cells, measured the protein content of the lysates and adjusted the samples to the same concentration. After pooling 5 replicates of each FFA concentration, we measured 19 phosphorylated protein targets using Luminex technology (multiplex antibody-based ELISA). We found that the phosphorylation level of heat shock protein beta-1 (HSPB1) was three-fold less than the control. Other scientists have recently shown that HSPB1 is down-regulated in NASH[296]. Moreover, we show irregular phosphorylation patterns in IKBA, AKT1, WNK1, FAK1 and STAT6, most of which play a role in NAFLD/NASH mechanism. The present study pinpoints the signaling alterations of FFA-induced NAFLD.

[A13] 7-8 November 2017, Amsterdam, Netherlands
Conference of Luminex, 3rd xMAP Connect
Development of a new drug repositioning platform for Non-Alcoholic Fatty Liver disease through network analysis.

Authors: Danae S. Zareifi, Nafsika Chala, Dimitris E. Messinis, Asier Antonaz, Angeliki Minia, Vaia Pliaka, Jan Rozanc, Manoussos M. Konstadoulakis, Konstantinos J. Bramis, Efstathios A. Antoniou, Eirini Pantiora, Andreas Polydorou, Antonios Vezakis, Georgios Fragulidis, Leonidas G. Alexopoulos

Non-alcoholic fatty liver disease (NAFLD) is defined as the presence of hepatic steatosis in the absence of excess alcohol consumption and is considered the hepatic manifestation of the metabolic syndrome. It is considered to be the most common pathological condition of the liver, but its cause and progression mechanisms are still not completely understood, partially because of the lack of sufficient in vitro models. There is no therapy approved specifically for NAFLD by FDA). The aim of this project is to develop a platform for drug repositioning in NAFLD/NASH by combining novel in vitro models of primary human hepatocytes with networkbased analysis of gene expression and xMAP data.

For the in vitro induction of NAFLD, primary human hepatocytes (pHH) were exposed to free fatty acids (palmitic acid, oleic acid) and to the steatogenic compounds amiodarone, tamoxifen, tetracycline and valproic acid. The presence of steatosis was assesed using high content screening; lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. In addition, 17 phosphorylated protein targets were measured and signaling networks were constructed, in an attempt to shed a light into the mechanism of each in vitro approach to induce NAFLD.

A network-based computational approach was employed to suggest compounds for NAFLD. NAFLD-related networks were identified through (i) gene set analysis (GSA) and (ii) multiplex phosphoproteomic data. The common pathways with the aforementioned steatogenic compounds, used to induce NAFLD in vitro, were found through Drugbank and MSig database. To suggest compounds that reverse the disease mechanism, the steatogenic compounds were used with the Connectivity Map database. The most promising compounds for drug repositioning were identified in the intersection between patient-derived and drug-derived networks.

In this work, we have successfully developed NAFLD-induced in vitro models driven by steatogenic compounds. xMAP phosphoproteomic data and pathway analysis

identified deregulations of CREB1, ERK1, MEK1, P53 and NFKB that came in agreement with literature whereas the involvement of CHK2 and EPOR pathways seems to be unknown mechanisms for steatosis. Our drug reposition platform suggested 10 potential candidates that were tested in our in-vitro models. Interestingly, 5 of them have shown that can reverse in-vitro NAFLD. 3 of them are already in clinical trials whereas 2 more show very promising therapeutic potential.

[A14] 9-11 November 2017, Rome, Italy
First NAFLD Summit, European Association of Liver Disease
Network based drug repositioning for Non-Alcoholic Fatty Liver Disease
Authors: Danae S. Zareifi, Nafsika Chala, Dimitris E. Messinis, Asier Antonaz, Angeliki Minia, Vaia Pliaka, Jan Rozanc, Manoussos M. Konstadoulakis, Konstantinos J. Bramis, Efstathios A. Antoniou, Eirini Pantiora, Andreas Polydorou, Antonios Vezakis, Georgios Fragulidis, Leonidas G. Alexopoulos

Introduction: Non-alcoholic fatty liver disease (NAFLD) is the most common pathological condition of the liver. Many pharmacological agents have been tested for the management of the disease, but there is no therapy approved specifically for NAFLD by the US Food and Drug Administration. On this front, drug repositioning (DR) offers an accelerated route for drug discovery. Aims: The aim of this project is to suggest a platform for drug repositioning in NAFLD by combining novel in vitro models of primary human hepatocytes with network-based analysis of gene expression data from NAFLD patients.
Material and Methods: To induce NAFLD in vitro, primary human hepatocytes were exposed to free fatty acids (FFAs, palmitic and oleic acid) and to the steatogenic compounds amiodarone (AMI), tamoxifen (TMX), tetracycline (TET) and valproic acid (VPA). The formation of intracellular lipid droplets was verified using high content screening; lipid droplets were stained with Nile Red fluorescent probe, Hoechst 33342 was used for staining cell nucleus. The intracellular ROS production was measured using the fluorescent substrate CM-H2DCFDA. A network-based computational approach was employed to suggest compounds for NAFLD.

Briefly, gene expression networks derived from NAFLD patients were matched with drug-induced networks in an effort to identify drugs that affect the NAFLD-mechanisms. NAFLD-related networks were identified through gene set analysis (GSA) of two microarray datasets from GEO-NCBI. Common pathways with the steatogenic compounds used to induce NAFLD in vitro were found through Drugbank and MSig databases. To suggest compounds that reverse the disease mechanism, the steatogenic compounds were used with the Connectivity Map database. The promising compounds for DR are considered to belong in the intersection of GSA-derived and drug-derived networks.

Results: Lipid droplet accumulation and ROS production was present in all in vitro models, as it is observed in patients with NAFLD. The in silico analysis for DR identified Naftifine, Pralidoxime, Fusidic acid, Raloxifene, Oxprenolol, Dipivrefin, Metoprolol, Estrone sulfate, Physostigmine, Cefmetazole and Diflorasone as promising compounds for NAFLD treatment.

Conclusions: We have successfully developed NAFLD-induced in vitro models using the steatogenic compounds FFAs, VPA, TMX, TET and AMI. We have identified mechanisms of NAFLD pathogenesis through gene set analysis. The efficacy of the resulting compounds from the DR platform is being tested on the NAFLD in vitro models.

Abstract A2

Construction of a drug-induced phosphoprotein/cytokine dataset in clinical samples for Multiple Sclerosis

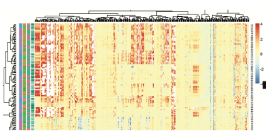
Vicky Pliaka[8*], Dimitris E Messinis[7,8*], Theodore Sakellaropoulos[7], Ekaterina Kotelnikova[1], Tomas Olsson[2], Jesper Tegner[3], Roland Martin[4], Dimitris Tzeranis[7,8], Friedemann Paul[5], Julio Saez-Rodiguez[6], Marti Bernardo-Faura[6], Ioannis N Melas[6,7,8], Jose Manuel Mas[9], Laura Artigas[9], Elena Schwartz[10], Ilya Mazo[10], Sophia Stamatatou[8], Mar Masso[11], Albert Zamora[11], Pablo Villoslada[1], Leonidas G Alexopoulos[7,8]

*Equal contributors / [1]IDIBAPS – Hospital Clinic of Barcelona, Spain / [2]Department of Neurology and [3]Unit of Computational Medicine, Department of Medicine, Center for Molecular Medicine, Karolinska Institutet, Karolinska University Hospital, Sweden / [4]University of Zurich, Switzerland / [5]NeuroCure Clinical Research Center, Charité University Medicine Berlin, Berlin, Germany / [6]European Molecular Biology Laboratory, European Bioinformatics Institute, UK / [7]National Technical University of Athens, Greece / [8]ProtATonce Ltd, Athens, Greece / [9]Anaxomics Biotech, Barcelona, Spain / [10]Ariadne Diagnostic, Rockville, MD, US / [11]Bionure Farma SL, Barcelona, Spain.

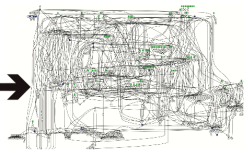
Multiple sclerosis is an autoimmune disease that affects the brain and spinal cord. About two and a half million people around the world have MS and there is not yet a cure for the disease. Current therapies for MS are not effective and they target only part of the immune response.

Scope of the project

In order to be able to develop combination therapies for MS, we first need to understand how current MS therapies work in biological networks.
For this purpose, by adopting a systems approach, we tried to build an as extensive and multi-level dataset as possible.



Dataset development



understanding of MS Biological Networks

Combination Therapies for Multiple Sclerosis

Pipeline for Dataset Development

255 Donors x **20 compounds** x **22 cytokines @24h 17 phosphoproteins @ 5′ & 25′** = **300000 data points**

Extraction of peripheral blood mononuclear cells (PBMCs) from **255** MS and healthy blood donors → Stimulation with **20** currently in-use and potential drugs → Collection of supernatants at 24h post stimulus and lysates at 5′ & 25′ post stimulus → Measurement of all samples for **22** cytokines and **17** phosphoproteins

Those three steps were all completed using the kit developed

Kit Development

In order to provide every medical center with reagents of the exact same batch, we created kits for PBMC isolation, plating, stimulation and lysis.

We got foam sheets and cut them in a way to fit all the tubes and vials needed for 1 donor's PBMCs isolation, gluing them together to create foam cases for safe shipment and we also designed boxes for -80 degrees, 4 degrees and room temperature shipments.

All reagents for isolation and drugs for stimulation were prepared in a single batch and were split into 300 kits.

Finally, the kits were assembled and sent to our medical partners while we also organized a training workshop where we followed the agreed protocol to make sure that all experimentalists are on the same page and are familiar with the kits.

Quality Control Check Points

During this project we had to have a great variety of quality control check points in order to ensure the highest possible data quality of the dataset

- Preparation of all reagents and stimuli in a single batch and subsequent split into 300 kits delivered to 4 countries
- Training workshop to ensure that all experimentalists follow the exact same protocol during isolation, stimulation and lysis
- Randomization of all samples before measurement
- Various hidden positive control stimuli among the drugs to detect potential experimentalists errors during stimulation
- Control signals during measurements that can detect biological noise or instrument error
- Exclusion of any donors that did not provide the required number of PBMCs during plating of cells
- Exclusion of any measurements that during the data acquisition process were considered by our algorithms to not be robust

Bulk Assay Development & Measurements

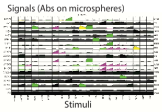
We used Luminex xMAP assays, developed by ProtATonce for phosphoproteomic and cytokine signals specific for Multiple Sclerosis. Those signals were optimized for the measurement of PMBC samples. By performing bulk assay development for 100s of plates we were able to produce this dataset at a fraction of the cost of commercial reagents.

CCL20	IFNG	IL8
CCL3	IL12A	PROK1
CCL5	IL1A	TNF10
CXL10	IL1B	TNF12
CXL11	IL20	TNFA
CXL16	IL3	TNR9
GROA	IL4	
ICAM1	IL6	

Custom Cytokine Panel

AKT1 (S473)	JUN (S63)	STAT3 (Y705)
CREB1 (S133)	MK12 (T183/Y185)	STAT5 (Y694)
FAK1 (Y397)	MK03 (T202/Y204)	STAT6 (Y641)
GSK3A (S21)	MP2K1 (S217/S221)	TF65 (S536)
HSPB1 (S78/S82)	PTN11 (Y542)	WNK1 (T60)
IKBA (S32/S36)	STAT1 (Y701)	

Custom Phosphoproteomic Panel



combiMS measurements

Final Dataset & Analysis

The resulting 200 Donors that passed every quality control test and were measured both for cytokines and phosphoproteins, resulted in a dataset of more than **300000 data points**. The assays for such a dataset could cost more than a million euros in terms of commercial reagents cost.

Those phosphoproteomic and cytokine data will be combined with SNP data and clinical profiles in a computational framework which will help to understand MS more thoroughly and systematically.

As a first step, the phosphoproteomic dataset will be used for the construction of a detailed map of the signaling pathway differences between MS and healthy donors, which can help generate a model of MS pathogenesis and improve our understanding of the disease.

Acknowledgements

Funded by European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding
Program: ERC. Investing in knowledge society through the European Social Fund.

European Union
European Social Fund

OPERATIONAL PROGRAMME EDUCATION AND LIFELONG LEARNING
investing in knowledge society
MINISTRY OF EDUCATION & RELIGIOUS AFFAIRS
MANAGING AUTHORITY

NSRF 2007-2013
EUROPEAN SOCIAL FUND

Co-financed by Greece and the European Union

Abstract A4

An integrated proteomic and metabolomic approach to investigate cartilage degeneration

Dimitris E Messinis[1,2,3], Vicky Pliaka[2], Stavroula Samara[4], Zoe Dailiana[5], Panagoula Kollia[4], Maria I Klapa[3], Leonidas G Alexopoulos[1,2]

[1]Mechanical Engineering Department, National Technical University of Athens, Athens, Greece
[2]ProtATonce Ltd, Athens, Greece
[3]Metabolic Engineering and Systems Biology Laboratory, Institute of Chemical Engineering Sciences, Foundation for Research and Technology-Hellas (FORTH/ICE-HT), Patras, Greece
[4]Department of Genetics & Biotechnology, Faculty of Biology, National and Kapodistrian University of Athens, Athens, Greece
[5]Department of Orthopedic Surgery, School of Medicine, University of Thessaly, Larissa, Greece

Introduction

Cartilage degeneration as a feature of osteoarthritis (OA) is one of the most common causes of pain and disability in middle-aged and older people. [1]
Most research studies in OA focus on single therapeutic targets, disease processes or level of molecules, thus omics data have never been integrated in a systematic way.
Many different types of samples have been studied in osteoarthritis research (articular cartilage and synovium explants, whole cell extract of articular chondrocytes or mesenchymal stem cells, supernatant of articular cartilage explants, articular chondrocytes, osteoclasts or synovium-derived cells in culture, articular cartilage vesicles, synovial fluid, plasma, serum and urine), but no study has integrated data from more than three samples examined with the same assay, generating the need for a more systematic, multi-omics, multi-sample investigation of osteoarthritis.

Sources of data

Towards an integration of several omics data shown below, in this case study we measured samples using proteomic (phosphoproteomic and cytokine release) and metabolomic approaches.

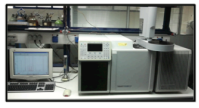
A. Genomics

B. Transcriptomics

C. Proteomics

Using ProtATonce's multiplex custom assays based on xMAP technology, we were able to measure 17 phosphoproteins in cell lysates and 36 secreted proteins in culture supernatants using only 50ul of sample.

D. Metabolomics



Using MESBL's Saturn 2200 GC-ion trap MS (Varian Inc, currently Bruker/Agilent), we were able to measure the metabolomic signatures of chondrocyte whole cell extracts, culture supernatants, plasma and synovial fluid samples of osteoarthritic patients.

E. Clinical Profiles

By examining the clinical profile data of donors we can further expand the dataset while gaining knowledge about factors such as medication that can alter the omics data.

Quality Control

When following a systematic approach, it is very important to include sufficient quality control check points to ensure the highest possible dataset quality and therefore the validity of results. In particular, we need to be able to address challenges such as:

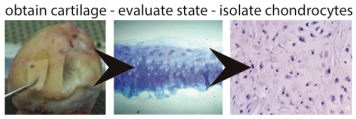
- acquiring multiple samples from the same donor at the same time
- using the exact same samples for all experiments (cell line passage, timelines in cell culture)
- merging protocols of all omics assays while meeting the sample handling prerequisites for every assay
- having strictly fixed parameters across assays, such as time points and doses of any applied stimuli

Sources of samples

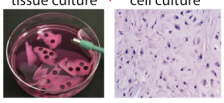
By using samples from various tissues, we avoid focusing only on the biochemical changes that occur in the joints. Some of the samples relative to osteoarthritis research are presented below:

A. Cartilage

obtain cartilage - evaluate state - isolate chondrocytes

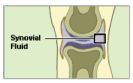


tissue culture → cell culture



Having cartilage explants in tissue culture or chondrocytes in cell culture, we can collect the culture's supernatant or lyse the cells to collect lysate.

B. Synovial Fluid

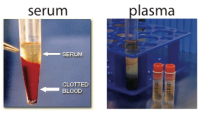


Synovial fluid can be found in the cavities of synovial joints and can be collected during surgery or by arthrocentesis.

C. Blood

Both blood serum and blood plasma can be used as samples. In this case study we examined blood plasma from osteoarthritic donors.

serum plasma

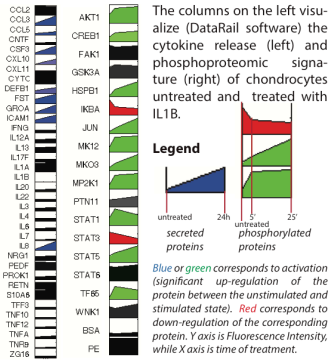


D. Urine

Metabolomic analysis of urine reveals differences between osteoarthritic and healthy donors. [2]

Our case study

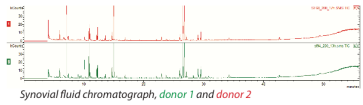
As a case study, we combined proteomic and metabolomic assays (multiplex bead based sandwich ELISA - xMAP technology and gas chromatography - mass spectrometry), to measure the phosphoproteomic signature, cytokine release and metabolomic signature of untreated and treated with IL1B cultured chondrocytes, along with the proteomic and metabolomic signature of the same donor's synovial fluid and plasma.



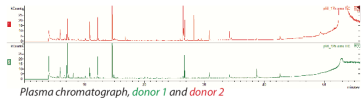
The columns on the left visualize (DataRail software) the cytokine release (left) and phosphoproteomic signature (right) of chondrocytes untreated and treated with IL1B.

Blue or green corresponds to activation (significant up-regulation of the protein between the unstimulated and stimulated state). Red corresponds to down-regulation of the corresponding protein. Y axis is Fluorescence Intensity, while X axis is time of treatment.

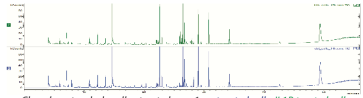
The OA donor used in proteomic experiments is donor 1. We also analyzed another OA donor 's synovial fluid and plasma, which is donor 2.



Synovial fluid chromatograph, donor 1 and donor 2



Plasma chromatograph, donor 1 and donor 2



Chondrocyte whole cell extract, donor 1, untreated cells, IL1B treated cells

Conclusions

We believe that omics data acquired by multiple samples while ensuring that all quality control check points have been met, can be integrated in a multi-level dataset that may help us understand complex biological systems and multifactorial diseases such as OA.
For this purpose, we intend to focus on gathering proteomic, metabolomic and other omics data examining in parallel as many samples of tissues relative to OA as possible and then integrating those data in a computational framework which will help to understand OA more thoroughly and systematically.

Acknowledgements

Co-financed by Greece and the European Union

Funded by European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding. Program: ERC. Investing in knowledge society through the European Social Fund.

We would also like to kindly thank Caterina Vasilopoulou, Eleni Kafkia, Ilona Binenbaum, Maria-Konstantina Ioannidi and the MESBL team for their valuable help in the progress of this study.

References

[1] *Buckwalter, J.A. and H.J. Mankin, Instr Course Lect, 1998. 47: p. 487-504.*
[2] *Lamers, R.J., et al., Osteoarthritis Cartilage, 2005. 13(9): p. 762-8.*

Signalling pathway-based screening for Drug Discovery: an application in Multiple Sclerosis

Dimitris E Messinis[1,2], Vicky Pliaka[2], Sophia Stamatatou[2] ,Theodore Sakellaropoulos[1], CombiMS consortium[3], Leonidas G Alexopoulos[1,2]

[1] National Technical University of Athens - [2] ProtATonce Ltd. - [3] CombiMS consortium (http://www.combims.eu)

contact: di.messinis@gmail.com
+306936767643

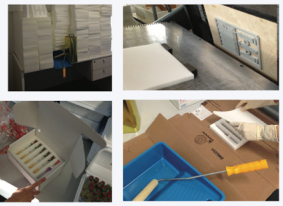
Multiple Sclerosis (MS) is an autoimmune disease affecting the brain and spinal cord. There is not yet a cure for the disease while 2,5 million people around the world have MS. Several pathological mechanisms for MS have been described, involving alterations in multiple processes and signalling pathways.

goal Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective therapies can be designed that take into account the uniqueness of each patient's response in treatment and biomarkers can be developed to stratify patients.

sample collection

kit development

In order to provide every medical center with reagents of the exact same batch



technicians workshop

We organized a training workshop where we followed the agreed protocol to make sure that all experimentalists are on the same page and are familiar with the kits.

blood sample acquisition

plating of PBMCs

compound stimulation

cell supernatant collection

lysate collection

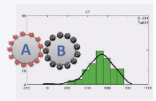
shipping of samples

sample measurement

255 donors of PBMCs
20 compounds for PBMC stimulation
22 secreted proteins measured @ 5' 25'
17 phosphoproteins measured @ 24h

CCL20	IFNG	IL8
CCL3	IL12A	PROK1
CCL5	IL1A	TNF10
CXL10	IL1B	TNF12
CXL11	IL20	TNFA
CXL16	IL3	TNR9
GROA	IL4	
ICAM1	IL6	

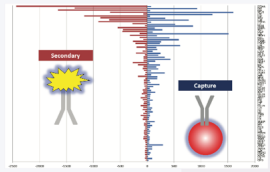
AKT1 (S473)	JUN (S63)	STAT3 (Y705)
CREB1 (S133)	MK12 (T183/Y185)	STAT5 (Y694)
FAK1 (Y397)	MK03 (T202/Y204)	STAT6 (Y641)
GSK3A (S21)	MP2K1 (S217/S221)	TF65 (S536)
HSPB1 (S78/S82)	PTN11 (Y542)	WNK1 (T60)
IKBA (S32/S36)	STAT1 (Y701)	



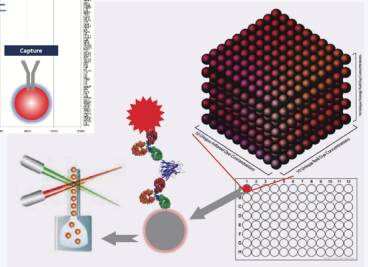
ProtATonce platform

custom Luminex assays for selected targets

multiplex high-throughput screening



matrix optimization

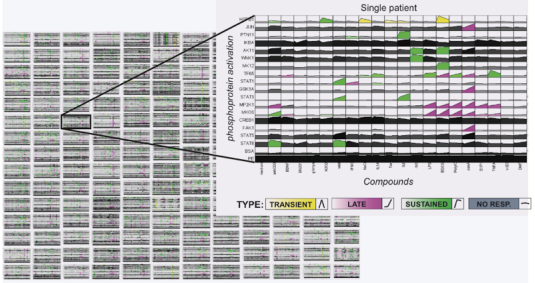


control beads

fraction of the cost of commercial reagents

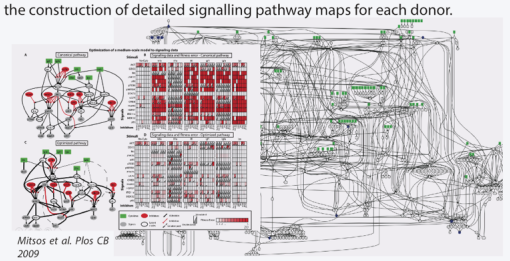
data analysis

patient-specific signalling response



pathway analysis

This dataset is being analyzed with pathway optimization tools for the construction of detailed signalling pathway maps for each donor.



Mitsos et al. Plos CB 2009

Funded by European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding
Program: ERC. Investing in knowledge society through the European Social Fund.

Abstract A7

Abstract A10 (similar with Abstract A8, A9)

Patient-specific signalling pathway analysis for Multiple Sclerosis

Dimitris E Messinis[1,2], Vicky Pliaka[2], Sophia Stamatatou[2] ,Theodore Sakellaropoulos[1], CombiMS consortium[3], Leonidas G Alexopoulos[1,2]

[1] National Technical University of Athens - [2] ProtATonce Ltd. - [3] CombiMS consortium (http://www.combims.eu)

introduction

Multiple sclerosis is an autoimmune disease that affects the brain and spinal cord. About two and a half million people around the world have MS and there is not yet a cure for the disease. Current therapies for MS are not effective and they target only part of the immune response. Our goal is to evaluate how current MS drugs and compounds with a therapeutic potential work at the signalling level in different patient populations. By understanding how current MS drugs work on patient-specific biological networks, more effective therapies can be designed that take into account the uniqueness of each patient's response in treatment and biomarkers can be developed to stratify patients.

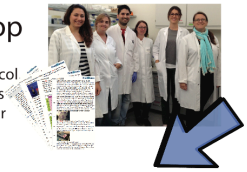
sample collection and processing

kit development

In order to provide every medical center with reagents of the exact same batch

technicians workshop

We organized a training workshop where we followed the agreed protocol to make sure that all experimentalists are on the same page and are familiar with the kits.



blood sample acquisition

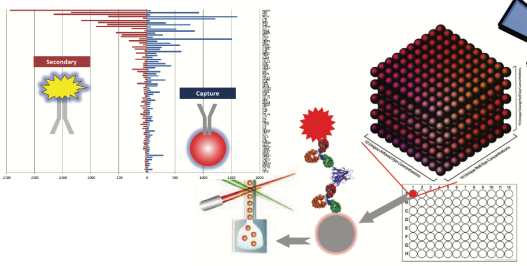
plating of PBMCs

compound stimulation

assay development

custom Luminex assays for selected targets

We used Luminex xMAP assays, developed by ProtATonce for phosphoproteomic and cytokine signals specific for MS. Those signals were optimized for the measurement of PMBC samples.



cell supernatant collection

lysate collection

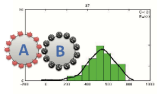
shipping of samples

sample measurement

255 donors of PBMCs
20 compounds for PBMC stimulation
22 secreted proteins measured @ 5' 25'
17 phosphoproteins measured @ 24h

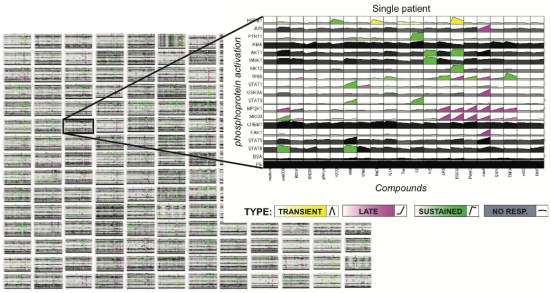
CCL20	IFNG	IL8
CCL3	IL12A	PROK1
CCL5	IL1A	TNF10
CXL10	IL1B	TNF12
CXL11	IL20	TNFA
CXL16	IL3	TNR9
GROA	IL4	
ICAM1	IL6	

AKT1 (S473)	JUN (S63)	STAT3 (Y705)
CREB1 (S133)	MK12 (T183/Y185)	STAT5 (Y694)
FAK1 (Y397)	MK03 (T202/Y204)	STAT6 (Y641)
GSK3A (S21)	MP2K1 (S217/S221)	TF65 (S536)
HSPB1 (S78/S82)	PTN11 (Y542)	WNK1 (T60)
IKBA (S32/S36)	STAT1 (Y701)	



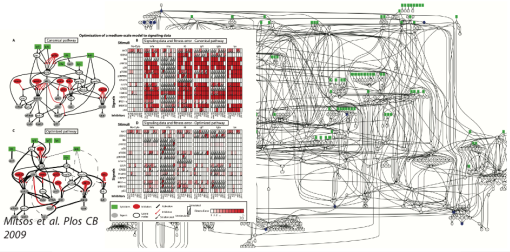
data analysis

patient-specific signalling response



pathway analysis

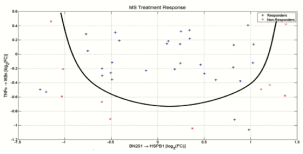
This dataset is being analyzed with pathway optimization tools for the construction of detailed signalling pathway maps for each donor.



Mitsos et al. Plos CB 2009

SVM models

We developed SVM models to predict a patient's response to treatment with Multiple Sclerosis drugs.



Funded by European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding
Program: ERC. Investing in knowledge society through the European Social Fund.

Abstract A12

Primary human hepatocyte models for NAFLD/NASH based on phosphoproteomics

Danai Stella Zareifi*[a], Dimitris E Messinis*[ab], Angeliki Minia[b], Vaia Pliaka[b], Jan Rozanc[b], Manoussos M Konstadoulakis[c], Konstantinos J Bramis[c], Efstathios A Antoniou[d], Leonidas G Alexopoulos[ab]

* equal contributors, [a] School of Mechanical Engineering, National Technical University of Athens, [b] ProtATonce Ltd, [c] 1st Department of Propaedeutic Surgery, Hippokration General Hospital, Medical School, University of Athens, [d] 2nd Department of Propaedeutic Surgery, Laiko Hospital, University of Athens,School of Medcine, Athens, Greece

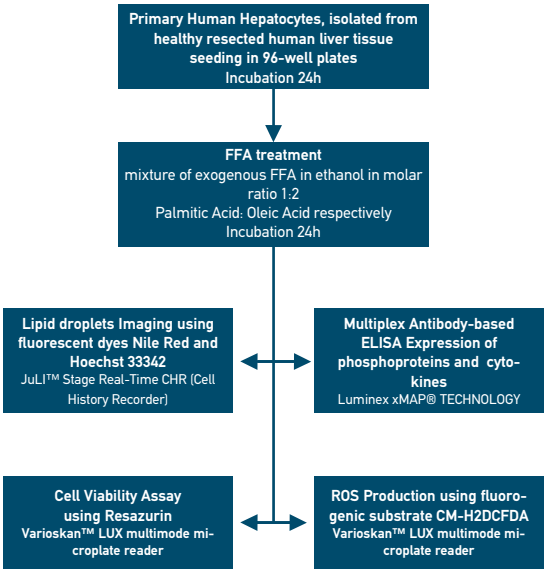
Introduction

Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, with a prevalence of up to 30% in developed countries[1]. NAFLD is defined by the presence of hepatic steatosis in the absence of excess alcohol consumption and represents a spectrum of disease, from simple steatosis and Non-Alcoholic Steatohepatitis (NASH) to fibrosis, cirrhosis and hepatocellular carcinoma[2]. The cause and disease progression mechanisms of NAFLD are still not completely understood[3,4]. Researchers have tried to understand the disease through multi-omic approaches[5] but there is no study -to our knowledge- exploring the signal transduction level, while phosphoproteomic measurements are becoming very important in drug discovery[6]. In vitro exposure of primary human hepatocytes to increasing concentrations of free fatty acids (FFA) results in increased intracellular accumulation of lipid droplets similar to those observed in patients with NAFLD and NASH[7].

Aim

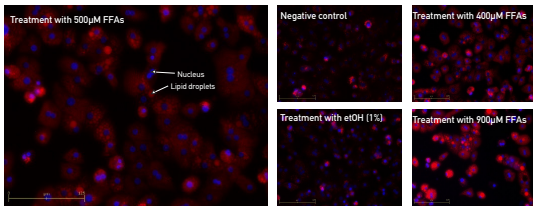
In vitro induction of NAFLD/NASH in a new in vitro model of primary human hepatocytes, by exposing to increasing concentrations of free fatty acids (FFA), in order to conduct a proteomic profiling of the disease's underlying mechanism.

Methods



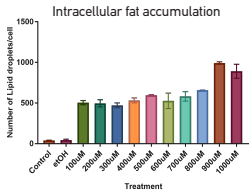
Results

1. High Content Screening

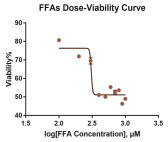


Verification of formation of intracellular lipid droplets; lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus.

2. Image Analysis



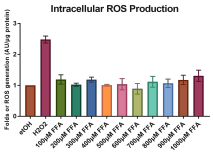
3. Assessment of viability



2. Dose dependent intracellular fat accumulation was quantified from the obtained images using CellProfiler open source software. Data are expressed as mean±SEM of three independent experiments.

3. Dose-Viability curves constracted using 4 parameter logistic regression (EC50=307±1.377). Data are expressed as mean±SEM of three independent experiments.

4. Assessment of ROS production



4.Fluorescence was normalized by μg of protein. Treatment with H2O2 was used as positive control. Data are expressed as mean±SEM of three independent experiments.

5. Proteomics

Secreted Proteins

FFA Treatment (μM)	CCL5	CXL10	IFNG	IL1A	IL20	IL6	IL8	TNF10	TNFA
100	0.87	0.94	0.70	0.93	0.91	0.71	1.09	0.71	0.88
200	0.97	1.01	0.67	0.96	0.98	1.11	1.16	0.75	0.83
300	0.89	0.76	0.75	1.00	1.00	0.99	1.19	0.78	0.83
400	0.90	0.66	0.67	0.77	0.76	0.48	0.95	0.65	0.96
500	0.89	0.76	0.72	0.95	0.98	0.73	1.36	0.78	0.87
600	0.91	0.80	0.69	0.96	0.96	0.70	1.34	0.68	0.82
700	0.85	0.78	0.71	0.94	0.92	0.79	1.43	0.71	0.77
800	0.85	0.71	0.61	0.91	0.93	0.72	1.44	0.65	0.77
900	0.84	0.66	0.65	0.97	0.90	0.68	1.41	0.66	0.71
1000	1.03	0.77	0.73	0.99	0.99	0.78	1.47	0.75	0.77

LEGEND	
value	CV < 20%
value	< 0.75

Phosphoproteins

FFA Treatment (μM)	AKT1	FAK1	HSBP1	IKBA	JUN	STAT6	WNK1
100	0.67	0.88	0.66	0.60	0.73	0.75	0.66
200	0.65	0.83	0.85	0.69	0.76	0.75	0.64
300	0.72	0.86	0.71	0.80	0.80	0.81	0.66
400	0.64	0.81	0.67	0.67	0.73	0.67	0.54
500	0.66	0.81	0.64	0.66	0.73	0.68	0.57
600	0.67	0.86	0.67	0.62	0.75	0.66	0.59
700	0.69	0.85	0.70	0.68	0.78	0.73	0.66
800	0.67	0.85	0.76	0.73	0.85	0.71	0.61
900	0.67	0.87	0.72	0.70	0.79	0.73	0.60
1000	0.62	0.91	0.75	0.70	0.74	0.74	0.60

Fold change of the proteomic measurements of three biological replicates at increasing concentrations of FFAs.

Differences between groups were compared by using Students t test.

Discussion

In the present work we describe an in vitro model able to mimic most of the processes observed in vivo in NAFLD/NASH, as shown by the increased intracellular fat accumulation, and intracellular ROS production that reflects defected antioxidant defenses.

We observed significant reduction of the phosphorylation of AKT1, that has been previously reported to downregulate in liver steatosis.[8] Moreover, we show irregular phosphorylation patterns in IKBA, JUN, STAT6 and WNK1 as well as in the secretion of TNFA, TNF10 ad TNFG.

References

1.Smith, B. W. & Adams, L. A. Non-alcoholic fatty liver disease. Crit. Rev. Clin. Lab. Sci. 48, 97–113 (2011).
2.Willebrords, J. et al. Strategies, models and biomarkers in experimental non-alcoholic fatty liver disease research. Prog. Lipid Res. 59, 106–125 (2015).
3.Kumashiro, N. et al. Cellular mechanism of insulin resistance in nonalcoholic fatty liver disease. Proc. Natl. Acad. Sci. 108, 16381–16385 (2011).
4.Pettinelli, P., Obreg n, A. M. & Videla, L. A. Molecular mechanisms of steatosis in nonalcoholic fatty liver disease. Nutr. Hosp. 26, 441–50
5.Wruck, W. et al. Multi-omic profiles of human non-alcoholic fatty liver disease tissue highlight heterogenic phenotypes. Sci. Data 2, 150068 (2015).
6.Morris, M. K., Chi, A., Melas, I. N. & Alexopoulos, L. G. Phosphoproteomics in drug discovery. Drug Discov. Today 19, 425–432 (2014).
7. Than, N. N. & Newsome, P. N. A concise review of non-alcoholic fatty liver disease. Atherosclerosis 239, 192–202 (2015).
8. Hur, W. et al. Downregulation of microRNA-451 in non-alcoholic steatohepatitis inhibits fatty acid-induced proinflammatory cytokine production through the AMPK/AKT pathway. Int. J. Biochem. Cell Biol. 64, 265–276 (2015).

Abstract A13

Construction of a new drug repositioning platform for Non-Alcoholic Fatty Liver disease through network analysis

D. ZAREIFI[1], N. CHALA[1], D. E. MESSINIS[1] , A. ANTORANZ[1], A. MINIA[2] b, V. PLIAKA[2], J. ROŽANC[2], M. M. KONSTADOULAKIS[3], K. J. BRAMIS[3], E. A. ANTONIOU[4], E. PANTIORA[5], A. POLYDOROU[5], A. VEZAKIS[5], G. FRAGULIDIS[5], L. G. ALEXOPOULOS[1,2]

[1]School of Mechanical Engineering, National Technical University of Athens, Athens, Greece, [2] ProtATonce Ltd, Athens, Greece, [3] 1st Department of Propaedeutic Surgery, Hippokration General Hospital, Medical School, University of Athens, Athens, Greece, [4] 2nd Department of Propaedeutic Surgery, Laiko Hospital, University of Athens, School of Medicine, Athens, Greece, [5] 2nd Department of Propaedeutic Surgery, Aretaieio Hospital, University of Athens, School of Medicine, Athens, Greece

INTRODUCTION

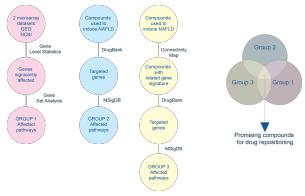
Non-alcoholic fatty liver disease (NAFLD) is the most common pathological condition of the liver, with a prevalence of up to 30% in developed countries[1,2]. Many pharmacological agents have been tested for the management of the disease, but there is no therapy approved specifically for NAFLD/NASH by the US Food and Drug Administration (FDA)[2,3]. On this front, drug repositioning (DR) offers an accelerated route for drug discovery.

AIM

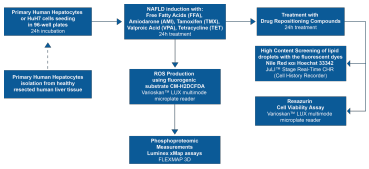
The aim of this project is to suggest a platform for drug repositioning in NAFLD/NASH by combining novel *in vitro* models of primary human hepatocytes with network-based analysis of gene expression data from NAFLD patients.

METHOD

A)Network-based drug repositioning workflow

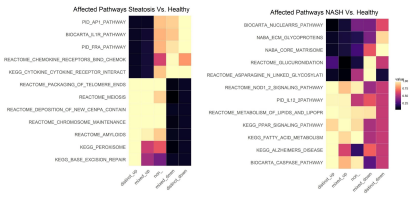


B) NAFLD *in vitro* models – Experimental Design



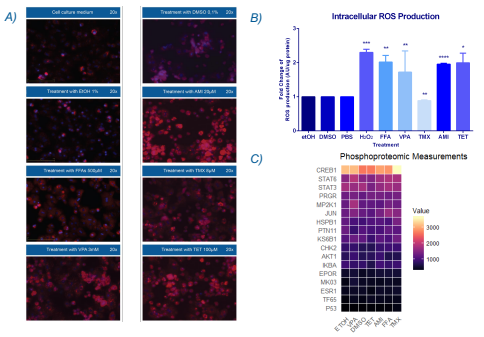
RESULTS

Pathway Analysis from microarray data sets



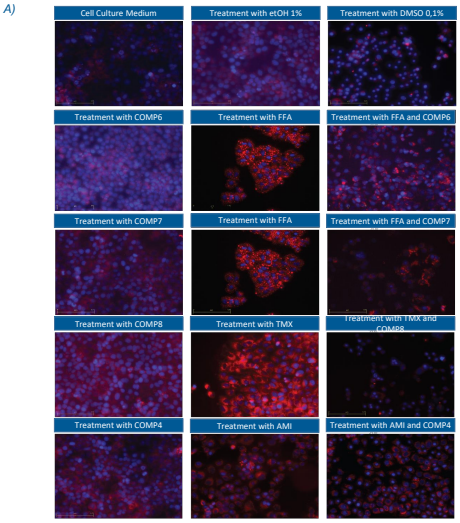
Example of heat map of affected pathways derived from Functional Class Scoring of 2 microarray datasets[4,5] (GEO-NCBI). Rows represent known pathways acquired from the annotated databases. For each Gene Set Analysis and pathways 5 p-value classes are obtained: 1) "distinct_up" , of pathways with differentially up-regulated genes, 2) "mixed_up", of pathways with less differentially up-regulated genes, 3) "non_directional", of pathways differentially regulated, 4) "mixed_down", of pathways with less differentially down-regulated genes, and 5) "distinct_down", of pathways with differentially down-regulated genes. Data are presented as St(-log$_{10}$(Pvalue) of a pathway for each p-value class.

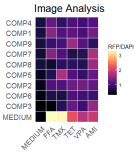
NAFLD induction on Primary Human Hepatocytes



***A)** Intracellular lipid droplet accumulation after 24h exposure of Primary Human Hepatocytes to Free Fatty Acids (FFAs; oleic and plamitic acid), Amiodarone (AMI), Tamoxifen (TMX), Tetracycline (TET) and Valproic Acid (VPA). Lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. Images were acquired by JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek) under 20x optical magnification.*
***B)** Fold change of intracellular ROS production after 24h exposure of primary human hepatocytes to FFAs, VPA, TMX, AMI and TET. The intracellular ROS production was measured using the fluorescent substrate CM-H_2DCFDA. Fluorescence was measured using Varioskan™ LUX multimode microplate reader (Thermo Scientific™) and normalized by µg of protein. Treatment with H_2O_2 was used as positive control. Data are expressed as mean ±SEM of three independent experiments.*
***C)** Heat map of fold change of 12 phosphoproteins compared to the controls. Data are presented as fold change of Mean Fluorescence intensity (MFI) compared to the respective controls. Protein expression was quantified using Luminex xMap assays in FLEXMAP 3D (Luminex).*

Screening of Compounds from Drug Repositioning Platform





C)

NAFLD model	Compounds from Drug repositioning								
	COMP 1	COMP 2	COMP 3	COMP 4	COMP 5	COMP 6	COMP 7	COMP 8	COMP 9
FFA	X	X	√	X	√	√	√	√	√
AMI	X	?	X	?	?	√	X	X	?
TMX	X	X	X	√	√	?	√	√	?
TET	X	X	X	X	?	X	√	√	?
VPA	?	X	?	√	?	?	√	?	?

√ → no steatosis present
X → no effect
? → under further investigation

***A)** Intracellular lipid droplet presence after 24h exposure of HuH7 to FFAs, AMI, TMX, TET and VPA followed by 24h treatment with the drug repositioning compounds (coded COMPx) from LOPAC^{1280} library (Sigma-Aldrich). Lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. Images were acquired by JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek) under 20x optical magnification.* ***B)** Heatmap of image analysis results (MATLAB). Data are presented as RFP intensity/DAPI intensity of lipid droplets and cell nuclei respectively.* ***C)** Summary of the preliminary screening of drug repositioning compounds.*

CONCLUSIONS

- Development of NAFLD-induced *in vitro* models of Primary Human Hepatocytes and hepatocellular cell lines driven by Free Fatty Acids (FFAs; palmitic and oleic acid) and the steatogenic compounds Valproic acid (VPA), Tamoxifen (TMX), Tetracycline (TET) and Amiodarone (AMI).
- Identification of NAFLD pathogenesis mechanisms through gene set analysis of microarray datasets.
- Development of a network-based drug repositioning platform for NAFLD, that could be expanded for other diseases.
- Identification of 56 compounds for drug repositioning:
  1) 33 of these compounds are known hepatotoxic compounds
  2) 7 of these compounds are already on clinical trials or have been used *in vivo* on mouse models for the treatment of NAFLD (Fusidic acid, Quinacrine, Resveratrol, Sirolimus, Estrone sulfate, Raloxifene, Ivermectin)
  3) 16 of these compounds are novel findings
- The efficacy of the 23 compounds is being tested on the NAFLD-induced *in vitro* models.

ACKNOWLEDGEMENTS

Leonidas G Alexopoulos has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreements No 642295 and No 675585. We would like to thank Jan Rožanc for his help in liver sample acquisition.

REFERENCES

1. **Smith, B. W. & Adams, L. A.** Non-alcoholic fatty liver disease. *Crit. Rev. Clin. Lab. Sci. 48, 97–113 (2011)*
2. **Willebrords, J. et al.** Strategies, models and biomarkers in experimental non-alcoholic fatty liver disease research. *Prog. Lipid Res. 59, 106–125 (2015)*
3. **Than, N. N. & Newsome, P. N.** A concise review of non-alcoholic fatty liver disease. *Atherosclerosis 239, 192–202 (2015)*
4. The data discussed in this poster have been deposited in NCBI's Gene Expression Omnibus (**Frades I. et al., 2014**) and are accessible through GEO Series accession number **GSE63067** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63067)
5. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (**Allard JP et al., 2016**) and are accessible through GEO Series accession number **GSE89632** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89632)

EASL | FIRST NAFLD SUMMIT
The Home of Hepatology

P02-06YI

Network based drug repositioning for Non-Alcoholic Fatty Liver Disease

D. ZAREIFI[1], N. CHALA[1], D. E. MESSINIS[1], A. ANTORANZ[1], A. MINIA[2], V, PLIAKA[2], J. ROŽANC[2], M. M. KONSTADOULAKIS[3], K. J. BRAMIS[3], E. A. ANTONIOU[4], E. PANTIORA[5], A. POLYDOROU[5], A. VEZAKIS[5], G. FRAGULIDIS[5], L. G. ALEXOPOULOS[1,2]

[1]School of Mechanical Engineering, National Technical University of Athens, Athens, Greece, [2] ProtATonce Ltd, Athens, Greece, [3] 1st Department of Propaedeutic Surgery, Hippokration General Hospital, Medical School, University of Athens, Athens, Greece, [4] 2nd Department of Propaedeutic Surgery, Laiko Hospital, University of Athens, School of Medicine, Athens, Greece, [5] 2nd Department of Propaedeutic Surgery, Aretaieio Hospital, University of Athens, School of Medicine, Athens, Greece

PROT AT ONCE
THE ART TO MULTIPLEXED ASSAYS

HELLENIC REPUBLIC
National and Kapodistrian University of Athens

INTRODUCTION

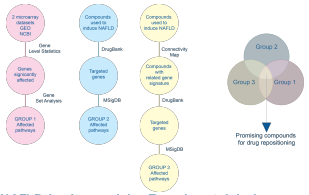
Non-alcoholic fatty liver disease (NAFLD) is the most common pathological condition of the liver, with a prevalence of up to 30% in developed countries[1,2]. Many pharmacological agents have been tested for the management of the disease, but there is no therapy approved specifically for NAFLD/NASH by the US Food and Drug Administration (FDA)[2,3]. On this front, drug repositioning (DR) offers an accelerated route for drug discovery.

AIM

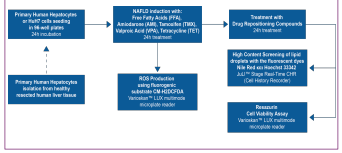
The aim of this project is to suggest a platform for drug repositioning in NAFLD/NASH by combining novel *in vitro* models of primary human hepatocytes with network-based analysis of gene expression data from NAFLD patients.

METHOD

Network-based drug repositioning workflow

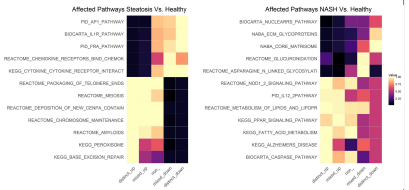


NAFLD *in vitro* models - Experimental design



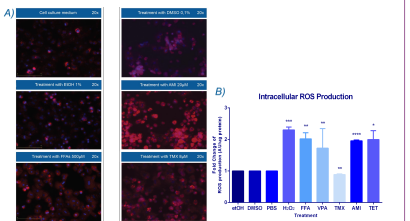
RESULTS

Pathway Analysis from microarray data sets



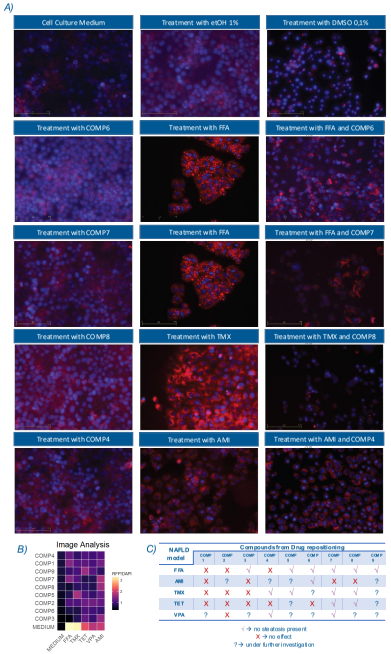
Example of heat map of affected pathways derived from Functional Class Scoring of 2 microarray datasets[4,5] (GEO-NCBI). Rows represent known pathways acquired from the annotated databases. For each Gene Set Analysis and pathways 5 p-value classes are obtained: 1) "distinct_up", of pathways with differentially up-regulated genes, 2) "mixed_up", of pathways with less differentially up-regulated genes, 3) "non_directional", of pathways differentially regulated, 4) "mixed_down", of pathways with less differentially down-regulated genes, and 5) "distinct_down", of pathways with differentially down-regulated genes. Data are presented as SI(-log$_{10}$(Pvalue) of a pathway for each p-value class.

NAFLD induction on Primary Human Hepatocytes



***A)** Intracellular lipid droplet accumulation after 24h exposure of Primary Human Hepatocytes to Free Fatty Acids (FFAs; oleic and plamitic acid), Amiodarone (AMI), Tamoxifen (TMX), Tetracycline (TET) and Valproic Acid (VPA). Lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. Images were acquired by JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek) under 20x optical magnification.*
***B)** Fold change of intracellular ROS production after 24h exposure of primary human hepatocytes to FFAs, VPA, TMX, AMI and TET. The intracellular ROS production was measured using the fluorescent substrate CM-H$_2$DCFDA. Fluorescence was measured using Varioskan™ LUX multimode microplate reader (Thermo Scientific™) and normalized by μg of protein. Treatment with H$_2$O$_2$ was used as positive control. Data are expressed as mean ±SEM of three independent experiments.*

Screening of Compounds from Drug Repositioning Platform



***A)** Intracellular lipid droplet presence after 24h exposure of HuH7 to FFAs, AMI, TMX, TET and VPA followed by 24h treatment with the drug repositioning compounds (coded COMPx) from LOPAC^{1280} library (Sigma-Aldrich). Lipid droplets were stained with Nile Red fluorescent probe and Hoechst 33342 was used for counterstaining cell nucleus. Images were acquired by JuLI™ Stage Real-Time CHR (Cell History Recorder) (NanoEnTek) under 20x optical magnification.* ***B)** Heatmap of image analysis results (MATLAB). Data are presented as RFP intensity/DAPI intensity of lipid droplets and cell nuclei respectively.* ***C)** Summary of the preliminary screening of drug repositioning compounds.*

CONCLUSIONS

- Development of NAFLD-induced *in vitro* models of Primary Human Hepatocytes and hepatocellular cell lines (HuH7, Hep3B, HepG2, FOCUS) driven by Free Fatty Acids (FFAs; palmitic and oleic acid) and the steatogenic compounds Valproic acid (VPA), Tamoxifen (TMX), Tetracycline (TET) and Amiodarone (AMI).
- Identification of NAFLD pathogenesis mechanisms through gene set analysis of microarray datasets.
- Development of a network-based drug repositioning platform for NAFLD, that could be expanded for other diseases.
- Identification of 56 compounds for drug repositioning:
  1) 33 of these compounds are known hepatotoxic compounds
  2) 7 of these compounds are already on clinical trials or have been used *in vivo* on mouse models for the treatment of NAFLD (Fusidic acid, Quinacrine, Resveratrol, Sirolimus, Estrone sulfate, Raloxifene, Ivermectin)
  3) 16 of these compounds are novel findings
- The efficacy of the 23 compounds is being tested on the NAFLD-induced *in vitro* models.
- More microarray data sets are under analysis.

ACKNOWLEDGEMENTS

Leonidas G Alexopoulos has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreements No 642295 and No 675585. We would like to thank Jan Rožanc for his help in liver sample acquisition.

REFERENCES

1. **Smith, B. W. & Adams, L. A.** Non-alcoholic fatty liver disease. *Crit. Rev. Clin. Lab. Sci. 48, 97–113 (2011)*
2. **Willebrords, J. et al.** Strategies, models and biomarkers in experimental non-alcoholic fatty liver disease research. *Prog. Lipid Res. 59, 106–125 (2015)*
3. **Than, N. N. & Newsome, P. N.** A concise review of non-alcoholic fatty liver disease. *Atherosclerosis 239, 192–202 (2015)*
4. The data discussed in this poster have been deposited in NCBI's Gene Expression Omnibus (**Frades I. et al., 2014**) and are accessible through GEO Series accession number **GSE63067** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63067)
5. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (**Allard JP et al., 2016**) and are accessible through GEO Series accession number **GSE89632** (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE89632)

CONTACT INFORMATION

Leonidas G. Alexopoulos: leo@ntua.gr
Danae Zarifi: danae.zar@gmail.com

Abstract A14

Παράρτημα IV: Δημοσιεύσεις | Peer-reviewed Publications

Dimitris E Messinis (DEM), as a PhD candidate, contributed as an author in the following publications, numbered P1 to P6. For P1, DEM constructed the prior knowledge network. For P2, DEM wrote the manuscript, performed the experiment and generated the data. For P3, DEM performed the experiment and generated the data. For P4, DEM designed the research, performed the experiment and generated the data. For P5, DEM performed the research and analysed the data. For P6, DEM wrote the manuscript, designed the research and performed the research. P7 manuscript has been submitted and DEM wrote the manuscript, designed the research and performed the research.
For source of funding, conflict of interest declarations, other author contributions and disclaimers please visit the publications using the DOI provided.

[P1] IN Melas*, AD Chairakaki*, EI Chatzopoulou*, DE Messinis, T Katopodi, V Pliaka, S Samara, A Mitsos, Z Dailiana, P Kollia, LG Alexopoulos “Modeling of signaling pathways in chondrocytes based on phosphoproteomic and cytokine release data”, * equal contributors
Osteoarthritis and Cartilage 2014 DOI: 10.1016/j.joca.2014.01.001

Objective: Chondrocyte signaling is widely identified as a key component in cartilage homeostasis. Dysregulations of the signaling processes in chondrocytes often result in degenerative diseases of the tissue. Traditionally, the literature has focused on the study of major players in chondrocyte signaling, but without considering the cross-talks between them. In this paper, we systematically interrogate the signal transduction pathways in chondrocytes, on both the phosphoproteomic and cytokine release levels.
Methods: The signaling pathways downstream 78 receptors of interest are interrogated. On the phosphoproteomic level, 17 key phosphoproteins are measured upon stimulation with single treatments of 78 ligands. On the cytokine release level,

55 cytokines are measured in the supernatant upon stimulation with the same treatments. Using an Integer Linear Programming (ILP) formulation, the proteomic data is combined with a priori knowledge of proteins' connectivity to construct a mechanistic model, predictive of signal transduction in chondrocytes.
Results: We were able to validate previous findings regarding major players of cartilage homeostasis and inflammation (e.g., IL1B, TNF, EGF, TGFA, INS, IGF1 and IL6). Moreover, we studied pro-inflammatory mediators (IL1B and TNF) together with pro-growth signals for investigating their role in chondrocytes hypertrophy and highlighted the role of underreported players such as Inhibin beta A (INHBA), Defensin beta 1 (DEFB1), CXCL1 and Flagellin, and uncovered the way they cross-react in the phosphoproteomic level.
Conclusions: The analysis presented herein, leveraged high throughput proteomic data via an ILP formulation to gain new insight into chondrocytes signaling and the pathophysiology of degenerative diseases in articular cartilage.

[P2] C Poussin, C Mathis, LG Alexopoulos, DE Messinis, RHJ Dulize, V Belcastro, IN Melas, T Sakellaropoulos, K Rhrissorrakrai, E Bilal, P Meyer, M Talikka, S Boué, R Norel, JJ Rice, G Stolovitzky, NV Ivanov, MC Peitsch & J Hoeng "The species translation challenge—A systems biology perspective on human and rat bronchial epithelial cells"
Nature Scientific Data 2014 DOI: 10.1038/sdata.2014.9

The biological responses to external cues such as drugs, chemicals, viruses and hormones, is an essential question in biomedicine and in the field of toxicology and cannot be easily studied in humans. Thus, biomedical research has continuously relied on animal models for studying the impact of these compounds and attempted to 'translate' the results to humans. In this context, the SBV IMPROVER (Systems Biology Verification for Industrial Methodology for PROcess VErification in Research) collaborative initiative, which uses crowd-sourcing techniques to address fundamental questions in systems biology, invited scientists to deploy their own computational methodologies to make predictions on species translatability. A multi-layer systems biology dataset was generated that was comprised of

phosphoproteomics, transcriptomics and cytokine data derived from normal human (NHBE) and rat (NRBE) bronchial epithelial cells exposed in parallel to more than 50 different stimuli under identical conditions. The present manuscript describes in detail the experimental settings, generation, processing and quality control analysis of the multi-layer omics dataset accessible in public repositories for further intraand inter-species translation studies.

[P3] E Bilal*, T Sakellaropoulos*, Challenge Participants, IN Melas, DE Messinis, V Belcastro, K Rhrissorrakrai, P Meyer, R Norel, A Iskandar, E Blaese, JJ Rice, MC Peitsch, J Hoeng, G Stolovitzky, LG Alexopoulos & C Poussin “A crow-sourcing approach for the construction of species-specific cell signaling networks”, *equal contributors
Bioinformatics 2014 DOI: 10.1093/bioinformatics/btu659

Motivation: Animal models are important tools in drug discovery and for understanding human biology in general. However, many drugs that initially show promising results in rodents fail in later stages of clinical trials. Understanding the commonalities and differences between human and rat cell signaling networks can lead to better experimental designs, improved allocation of resources and ultimately better drugs.
Results: The sbv IMPROVER Species-Specific Network Inference challenge was designed to use the power of the crowds to build two species-specific cell signaling networks given phosphoproteomics, transcriptomics and cytokine data generated from NHBE and NRBE cells exposed to various stimuli. A common literature-inspired reference network with 220 nodes and 501 edges was also provided as prior knowledge from which challenge participants could add or remove edges but not nodes. Such a large network inference challenge not based on synthetic simulations but on real data presented unique difficulties in scoring and interpreting the results. Because any prior knowledge about the networks was already provided to the participants for reference, novel ways for scoring and aggregating the results were developed. Two human and rat consensus networks were obtained by combining all the inferred networks. Further analysis showed that major signaling pathways were

conserved between the two species with only isolated components diverging, as in the case of ribosomal S6 kinase RPS6KA1. Overall, the consensus between inferred edges was relatively high with the exception of the downstream targets of transcription factors, which seemed more difficult to predict.

[P4] E Kotelnikova*, M Bernardo-Faura*, G Silberberg, NA Kiani, DE Messinis, IN Melas, L Artigas, E Schwartz, I Mazo, M Masso, LG Alexopoulos, JM Mas, T Olsson, J Tegner, R Martin, A Zamora, F Paul, J Saez-Rodriguez & P Villoslada “Signaling networks in MS: A systems-based approach to developing new pharmacological therapies”, * equal contributors
Multiple Sclerosis Journal 2015 DOI: 10.1177/1352458514543339

The pathogenesis of multiple sclerosis (MS) involves alterations to multiple pathways and processes, which represent a significant challenge for developing more-effective therapies. Systems biology approaches that study pathway dysregulation should offer benefits by integrating molecular networks and dynamic models with current biological knowledge for understanding disease heterogeneity and response to therapy. In MS, abnormalities have been identified in several cytokine-signaling pathways, as well as those of other immune receptors. Among the downstream molecules implicated are Jak/Stat, NF-Kb, ERK1/3, p38 or Jun/Fos. Together, these data suggest that MS is likely to be associated with abnormalities in apoptosis/cell death, microglia activation, blood-brain barrier functioning, immune responses, cytokine production, and/or oxidative stress, although which pathways contribute to the cascade of damage and can be modulated remains an open question. While current MS drugs target some of these pathways, others remain untouched. Here, we propose a pragmatic systems analysis approach that involves the large-scale extraction of processes and pathways relevant to MS. These data serve as a scaffold on which computational modeling can be performed to identify disease subgroups based on the contribution of different processes. Such an analysis, targeting these relevant MS-signaling pathways, offers the opportunity to accelerate the development of novel individual or combination therapies.

[P5] M Michailidou, IN Melas, DE Messinis, S Klamt, LG Alexopoulos, FN Kolisis and H Loutrari “Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action”
CPT: Pharmacometrics & Systems Pharmacology 2015 DOI: 10.1002/psp4.40

Chronic inflammation is associated with the development of human hepatocellular carcinoma (HCC), an essentially incurable cancer. Anti-inflammatory nutraceuticals have emerged as promising candidates against HCC, yet the mechanisms through which they influence the cell signaling machinery to impose phenotypic changes remain unresolved. Herein we implemented a systems biology approach in HCC cells, based on the integration of cytokine release and phospoproteomic data from highthroughput xMAP Luminex assays to elucidate the action mode of prominent nutraceuticals in terms of topology alterations of HCC-specific signaling networks. An optimization algorithm based on SigNetTrainer, an Integer Linear Programming formulation, was applied to construct networks linking signal transduction to cytokine secretion by combining prior knowledge of protein connectivity with proteomic data. Our analysis identified the most probable target phosphoproteins of interrogated compounds and predicted translational control as a new mechanism underlying their anticytokine action. Induced alterations corroborated with inhibition of HCC-driven angiogenesis and metastasis.

[P6] DE Messinis, IN Melas, J Hur, N Varshney, LG Alexopoulos and JPF Bai “Translational systems pharmacology-based predictive assessment of drug-induced cardiomyopathy”
CPT: Pharmacometrics & Systems Pharmacology 2018 DOI: 10.1002/psp4.12272

Drug-induced cardiomyopathy contributes to drug attrition. We compared two pipelines of predictive modeling: (1) applying elastic net (EN) to differentially expressed genes (DEGs) of drugs; (2) applying integer linear programming (ILP) to construct each drug’s signaling pathway starting from its targets to downstream proteins, to transcription factors, and to its DEGs in human cardiomyocytes, and then subjecting the genes/proteins in the drugs’ signaling networks to EN regression. We

classified 31 drugs with availability of DEGs into 13 toxic and 18 nontoxic drugs based on a clinical cardiomyopathy incidence cutoff of 0.1%. The ILP-augmented modeling increased prediction accuracy from 79% to 88% (sensitivity: 88%; specificity: 89%) under leave-one-out cross validation. The ILP-constructed signaling networks of drugs were better predictors than DEGs. Per literature, the microRNAs that reportedly regulate expression of our six top predictors are of diagnostic value for natural heart failure or doxorubicin-induced cardiomyopathy. This translational predictive modeling might uncover potential biomarkers.

[P7 - Submitted] Cue-signal-response analysis reveals multifactorial signalling mechanism of NAFLD in vitro.

Authors: Dimitris E Messinis, Danai-Stella Zareifi, Angeliki Minia, Vaia Pliaka, Eirini V Pantiora, Andreas A Polydorou, Antonios I Vezakis, Georgios P Fragulidis, Manoussos M Konstadoulakis, Konstantinos J Bramis, Efstathios A Antoniou, Leonidas G Alexopoulos

Non-alcoholic fatty liver disease (NAFLD) is the most common liver condition in the world, estimated to be by 2030 the most frequent indication for liver transplantation. NAFLD is a multifactorial disease and its cause and progression mechanisms are still not completely understood. The main aim of this study is to unveil different potential signaling mechanisms of NAFLD.

Amiodarone, Free Fatty Acids, Tamoxifen, Tetracycline and Valproic acid were employed as 5 NAFLD induction models in an in vitro platform to study the disease using primary human hepatocytes. NAFLD induction was monitored by Nile red staining. For each NAFLD-induction model, the signaling mechanism was interrogated by measuring 17 phosphorylated protein targets and pathway optimization algorithm was employed to construct signaling networks. Pathway findings were integrated with gene expression enrichment analysis to shed light into the mechanism of each in vitro approach to induce NAFLD.

5 NAFLD induction models were developed and showed strong lipid accumulation as evident by Nile red staining. Significant phosphoproteomic deregulations of CREB1, ERK1, MEK1, P53 and NFKB came in agreement with in vivo and in vitro literature

findings whereas CHK2 and EPOR have arisen as important in signaling pathways of NAFLD and are related to hepatic regeneration.

In this study, we induced NAFLD on primary human hepatocytes with 5 models that correspond to clinical causes of the disease. All models revealed strong NAFLD phenotype that was originated from diverse signaling mechanisms. Our results suggest a multifactorial mechanism in NAFLD progression involving significant signaling pathway deregulations.

Συνοπτικό Βιογραφικό Σημείωμα

Dimitris E. Messinis

di.messinis@gmail.com | +30 6936767643

I am a biotech scientist, social activist and technology enthusiast. Over the past 9 years, I have been highly involved in various types of experiences, including research, industry, nonprofit, and leadership opportunities. I am very persistent and curious, deeply enjoying learning new skills. I also enjoy traveling, cooking, playing tennis, and playing music (bass guitar, drums, and piano).

Career Highlights

- **Founder** at **Mindspace**, nonprofit aiming to educate students in technology entrepreneurship
- Data Scientist at the U.S. Food & Drug Administration (**FDA**) in Washington DC
- Systems Pharmacology Postdoctoral Fellow at **Philip Morris** R&D in Neuchâtel, Switzerland
- International Visitor Leadership Program (**IVLP**) alumnus, U.S. Department of State
- **PhD** in **Systems Bioengineering**, NTUA (h-index 6)
- Led two national startup competitions (operations in 6 cities) and designed the educational program and business meetings in Boston and New York for the winners
- Selected by the European Institute of Innovation & Technology to a group of 24 innovators to solve healthcare challenges posed by General Electric Healthcare

Professional Experience

- May 2018 - now: Postdoctoral Fellow, **Philip Morris International, Switzerland**
- Jul 2016 - Nov 2016: Data Scientist, **U.S. Food & Drug Administration (FDA), Washington DC** Completed a computational project to predict drug-induced cardiotoxicity
- Nov 2011 - Jan 2014: Founding Member & Research Scientist, **Protavio Ltd** Joined as the company's first employee and worked there for 2 years, establishing the company as an advanced Systems Pharmacology solutions provider.
- Aug 2009 - Oct 2011: Research Assistant, **Systems Bioengineering Group NTUA**

Social Activism

- July 2017 - Aug 2017: International Visitor Leadership Program (**IVLP**) "Social and Economic Entrepreneurship for Young Leaders", by the **U.S. Department of State**
- Aug 2017 - present: Shaper, **Global Shapers** - a global network of young people driving dialogue, action and change, initiated by the **World Economic Forum**
- Feb 2015 - present: **Founder**, and Chairman (Feb 2015 – Jul 2016), **Mindspace** - a nonprofit aiming to inspire and educate students to start a business

Education

- PhD, Mechanical Engineering NTUA, Jan 2014 – Apr 2018

  Advisor: Dr. Leonidas Alexopoulos, "Systems Biology and Big Data Modeling to decipher mechanisms in Human Disease"
- Diploma, Mechanical Engineering, NTUA, 2011

Honors & Awards

- U.S. Exchange Alumnus: Invited to IVLP by the U.S. Department of State, 2017
- Selected to join European Institute of Innovation & Technology (EIT) Health StarShip, 2017
- 1st prize among 56 submissions, Hepato-Pancreato-Biliary Association Conference, 2017
- Invited among 6 (out of 110) abstracts to give a lightning talk at SBHD, Harvard, 2014
- Best Poster First Prize award at Planet xMAP Multiplexing Congress in Vienna, 2011
- Second Prize award at Pan-University Innovative Design Competition, 2011
- Thomaidio award for scientific publication as undergraduate student, 2011

Peer-Reviewed Publications

[1] DE Messinis, IN Melas, J Hur, N Varshney, LG Alexopoulos and JPF Bai "Translational systems pharmacology-based predictive assessment of drug-induced cardiomyopathy"
CPT: Pharmacometrics & Systems Pharmacology 2018 DOI: 10.1002/psp4.12272
[2] M Michailidou, IN Melas, DE Messinis, S Klamt, LG Alexopoulos, FN Kolisis and H Loutrari "Network-Based Analysis of Nutraceuticals in Human Hepatocellular Carcinomas Reveals Mechanisms of Chemopreventive Action"
CPT: Pharmacometrics & Systems Pharmacology 2015 DOI: 10.1002/psp4.40
[3] E Kotelnikova*, M Bernardo-Faura*, G Silberberg, NA Kiani, DE Messinis, IN Melas, L Artigas, E Schwartz, I Mazo, M Masso, LG Alexopoulos, JM Mas, T Olsson, J Tegner, R Martin, A Zamora, F Paul, J Saez-Rodriguez & P Villoslada "Signaling networks in MS: A systems-based approach to developing new pharmacological therapies", * equal contributors
Multiple Sclerosis Journal 2015 DOI: 10.1177/1352458514543339
[4] E Bilal*, T Sakellaropoulos*, Challenge Participants, IN Melas, DE Messinis, V Belcastro, K Rhrissorrakrai, P Meyer, R Norel, A Iskandar, E Blaese, JJ Rice, MC Peitsch,

J Hoeng, G Stolovitzky, LG Alexopoulos & C Poussin “A crow-sourcing approach for the construction of species-specific cell signaling networks”, *equal contributors
Bioinformatics 2014 DOI: 10.1093/bioinformatics/btu659
[5] C Poussin, C Mathis, LG Alexopoulos, DE Messinis, RHJ Dulize, V Belcastro, IN Melas, T Sakellaropoulos, K Rhrissorrakrai, E Bilal, P Meyer, M Talikka, S Boué, R Norel, JJ Rice, G Stolovitzky, NV Ivanov, MC Peitsch & J Hoeng “The species translation challenge—A systems biology perspective on human and rat bronchial epithelial cells”
Nature Scientific Data 2014 DOI: 10.1038/sdata.2014.9
[6] IN Melas*, AD Chairakaki*, EI Chatzopoulou*, DE Messinis, T Katopodi, V Pliaka, S Samara, A Mitsos, Z Dailiana, P Kollia, LG Alexopoulos “Modeling of signaling pathways in chondrocytes based on phosphoproteomic and cytokine release data”, * equal contributors
Osteoarthritis and Cartilage 2014 DOI: 10.1016/j.joca.2014.01.001
[7] IN Melas, A Mitsos, DE Messinis, TS Weiss, J Saez-Rodriguez, LG Alexopoulos “Construction of large signaling pathways using an adaptive perturbation approach with phosphoproteomic data”
Molecular BioSystems 2012 DOI: 10.1039/c2mb05482e
[8] IN Melas*, A Mitsos*, DE Messinis, TS Weiss, LG Alexopoulos “Combined logical and data-driven models for linking signalling pathways to cellular response”, * equal contributors
BMC Systems Biology 2011 DOI: 10.1186/1752-0509-5-107