

Γενετική, βιοπληροφορική και οντολογία: Από τον γονότυπο στον φαινότυπο

Γεώργιος Γκούτος

*Τμήμα Γενετικής,
Πανεπιστήμιο του Cambridge*

ΠΡΟΛΟΓΟΣ

Ο σημαντικότερος ίσως στόχος κάθε βιοϊατρικής έρευνας είναι η κατανόηση του μοριακού υπόβαθρου της ανθρώπινης ασθένειας. Κάτι τέτοιο σε ιδεατές συνθήκες θα εξισωνόταν με τη συσχέτιση μιας ομάδας συμπτωμάτων με πλήρως περιγεγραμμένες βιολογικές διαδικασίες, όπου κάθε μοριακή οντότητα καθώς και οι σχέσεις τους είναι γνωστές. Μια τέτοια γνώση όχι μόνο θα οδηγούσε σε νέες θεραπείες, αλλά και θα επέτρεπε στις ήδη υπάρχουσες να χρησιμοποιηθούν αποτελεσματικότερα. Μια τέτοια αντίληψη της υποκείμενης βιολογίας θα οδηγούσε στη βελτίωση της υγειονομικής περίθαλψης.

Με την καλύτερη κατανόηση της αίτιας, της επίδρασης, άλλα και του αποτελέσματος, η υγειονομική περίθαλψη μπορεί να εστιάσει ακριβώς στην καταλληλότερη θεραπεία για κάθε μεμονωμένο ασθενή με δύο πολύ συναφείς συνέπειες. Κατά πρώτο λόγο, εστιάζει στην καλύτερη κατανόηση των θεμελιω-

δών βιολογικών διαδικασιών που προκαλούν την ασθένεια, ενώ, μέσω της μελέτης των περιστάσεων των μεμονωμένων ασθενών, οδηγεί στην καλύτερη αντίληψη του βιολογικού πλαισίου μέσα στο οποίο εμφανίζεται η ασθένεια.

Είναι προφανές ότι στην ιατρική δεν υπάρχουν ολικές απαντήσεις. Πολλές θεραπείες είναι αποτελεσματικές σε ένα μονό μέρος του πληθυσμού και σε διαφορετική κλίμακα. Ένας από τους βασικότερους λογούς αυτών των διαφορών έγκειται στο γενετικό υλικό του κάθε ατόμου – στο σύνολο δηλαδή των γενετικών πληροφοριών ή αλλιώς στον γονότυπο του. Θα είναι σύντομα τεχνητά εφικτή η γρήγορη και με χαμηλό κόστος χαρτογράφηση του πλήρους γονότυπου οποιουδήποτε μεμονωμένου ατόμου – αξιολογώντας όχι μόνο ένα μεμονωμένο γονίδιο, άλλα χιλιάδες γονίδια ταυτόχρονα. Ωστόσο, αυτή η εξατομικευμένη γενετική γνώση θα αποβεί χρήσιμη μόνο αν συνοδευτεί από ανάλογη υποδομή, η οποία να μπορεί να συσχετίσει αυτή τη γνώση με την ελλοχεύουσα βιολογία της ασθένειας. Ο πιο αποτελεσματικός τρόπος για να επιτευχθεί ένας τέτοιος σκοπός είναι μέσω των δεδομένων που συλλέγονται από πειράματα που πραγματοποιούνται χρησιμοποιώντας οργανισμούς, η μελέτη των οποίων μπορεί να αποτελέσει μοντέλο της ανθρώπινης ασθένειας.

ΕΙΣΑΓΩΓΗ

Η σημαντικότερη απασχόληση των γενετιστών είναι τι ακριβώς κάνουν τα γονίδια, δηλαδή ποια είναι η λειτουργία κάθε μεμονωμένου γονιδίου μέσα στο πλαίσιο ενός ολόκληρου οργανισμού. Υπάρχουν πολλοί τρόποι με τους οποίους ένας γενετιστής μπορεί να ανακαλύψει τη «λειτουργία» ενός γονιδίου. Αυτές οι τεχνικές, είτε είναι σχετικές με αναλύσεις διαγραφής γονιδίων, πειράματα RNAi, τεχνικές υπερέκφρασης, ή άλλες τεχνικές, ακολουθούν μια στρατηγική που είναι ουσιαστικά ίδια: μετάλλαξη ενός γονιδίου και παρατήρηση του φαινοτυπικού αποτελέσματος της γενετικής αλλαγής. Με άλλα λόγια, συσχέτιση του γονότυπου ενός οργανισμού με τον φαινότυπό του – δηλαδή τα ορατά φυσικά ή βιοχημικά χαρακτηριστικά του, όπως αυτά καθορίζονται από τον συνδυασμό του γενετικού υλικού και των περιβαλλοντολογικών επιρροών του.

Πολλές μεταλλάξεις έχουν σαφείς φαινοτυπικές συνέπειες, πιο συγκεκριμένα επιφέρουν επιπτώσεις, όπως αλλαγές στην εμφάνιση ενός οργανισμού. Παραδείγματος χάριν, μια μεταλλαγή σε ένα ανθρώπινο γονίδιο που κωδικοποιεί το ένζυμο τυροσινάση προκαλεί τον φαινότυπο του αλμπινισμού, με άλλα λόγια, την απουσία χρωστικής ουσίας από το δέρμα, του τριχωτού του σώματος και τα μάτια του ασθενούς. Φυσικά, οι γενετιστές δεν μπορούν πειραματικά να προκαλέσουν τέτοιες μεταλλάξεις στους ανθρώπους, μπορούν όμως να τις επιφέρουν σε οργανισμούς, όπως οι μύγες και τα ποντίκια, η μελέτη των οποίων είναι δυνατόν να χρησιμοποιηθεί ως μοντέλο για τη λειτουργία γονιδίων στον άνθρωπο. Επί του παρόντος, διάφορα πειράματα πολύ μεγάλης κλίμακας λαμβάνουν χώρα σε αυτούς τους οργανισμούς και στοχεύουν στην συστηματική μετάλλαξη όλων των γονιδίων τους.

Δεν είναι όμως μόνο αρκετή η φαινοτυπική περιγραφή αυτών των μεταλλαγών, είναι απολύτως αναγκαίο αυτές οι περιγραφές να αντιπροσωπευθούν με συνέπεια σε υπολογιστικά συστήματα διαχείρισης, και να συσχετιστούν με άλλες ιδιότητες των γονιδίων (π.χ. τη φύση των πρωτεϊνών τους, τον χώρο και τον χρόνο που εκφράστηκαν κτλ.) που απαιτεί μια συστηματική μέθοδο περιγραφής των μεταλλαγών, συστηματική υπό την έννοια ότι η γλώσσα που χρησιμοποιείται για την περιγραφή είναι ελεγχόμενη και ότι οι περιγραφές αντιπροσωπεύονται με υπολογιστικά κατανοητό τρόπο.

Η ΠΑΡΟΥΣΑ ΚΑΤΑΣΤΑΣΗ

Η πρόοδος στη γενετική την τελευταία δεκαετία είναι εκπληκτική. Η σχεδόν ολοκληρωμένη χαρτογράφηση του ανθρώπινου γονιδιώματος και των άλλων μοντέλων, μας αποκαλύπτει ότι το «βιβλίο της ανθρωπότητας» είναι ακόμα πιο θαυμάσιο και μυστήριο απ' ό,τι προηγουμένως θεωρούσαμε. Τα πρώτα πειράματα μας αποκαλύπτουν/ δείχνουν ότι το ανθρώπινο γονιδίωμα κατέχει ένα εξαιρετικό καταστατικό των απαιτούμενων πληροφοριών για την ανθρωπινή ανάπτυξη, φυσιολογία και εξέλιξη. Γνωρίζουμε πια ότι υπάρχουν περίπου 30.000 με 40.000 γονίδια, ένας αριθμός πολύ χαμηλότερος από τον

αναμενόμενο και μόλις δυο φορές μεγαλύτερος από τον αριθμό γονιδίων πολύ απλούστερων οργανισμών, όπως τα σκουλήκια και οι μύκητες. Αυτό που σταδιακά αποκαλύπτεται είναι το γεγονός ότι πολλά γονίδια στον ανθρώπινο οργανισμό φαίνεται να παίρνουν μέρος στην οργάνωση της λειτουργίας άλλων γονιδίων δημιουργώντας, έτσι, περίπλοκα δίκτυα γονιδίων, η μελέτη των οποίων αποτελεί βασική ανάγκη.

Ως αποτέλεσμα, υπάρχουν εργαστήρια και προγράμματα μεγάλης κλίμακας σε παγκόσμιο επίπεδο [1, 2, 3, 4, 5, 6] που αναλαμβάνουν εξονυχιστικές έρευνες για να επιτύχουν την πλέον απαραίτητη εις βάθος κατανόηση των σχέσεων και διαδικασιών που συμμετέχει κάθε γονίδιο. Πολλά από αυτά τα προγράμματα έχουν σκοπό τη συστηματική μετάλλαξη όλων των γονιδίων των μοντέλων οργανισμών και την καταγραφή των φαινοτυπικών αποτελεσμάτων της.

Δεν είναι όμως αρκετή μόνο η φαινοτυπική περιγραφή αυτών των μεταλλαγών, αφού είναι απολύτως αναγκαίο αυτές οι περιγραφές να συσχετιστούν με άλλες ιδιότητες των γονιδίων (π.χ. τη φύση των πρωτεϊνών τους, τον χώρο και τον χρόνο που εκφράστηκαν κτλ.).

Η ΕΠΟΜΕΝΗ ΜΕΓΑΛΗ ΠΡΟΚΛΗΣΗ ΣΤΗ ΓΕΝΕΤΙΚΗ

Αυτές οι προσπάθειες οδήγησαν στη δημιουργία ενός τεράστιου αριθμού βιολογικών δεδομένων που αυξάνεται εκθετικά. Ο αριθμός αυτών των δεδομένων είναι μάλιστα τόσο μεγάλος που μας έκανε να αρχίζουμε να συνειδητοποιούμε ότι η πληροφορία είναι, κατά πρώτο λόγο, τόσο αξιόπιστη όσο και η πηγή απ' όπου προέρχεται και, κατά δεύτερο λόγο, ότι από μονή της είναι σχεδόν μη εκμεταλλεύσιμη. Συνεπώς, είναι αναγκαίες οι υπολογιστικές υποδομές που θα μας επιτρέπουν την αναζήτηση, την ανάλυση, άλλα και τη σύγκριση αυτών των πληροφοριών.

Τα τελευταία έξι έτη έχει σημειωθεί ιδιαίτερη πρόοδος στη χρήση των δομημένων ελεγχόμενων λεξιλογίων (γνωστά ως οντολογίες) από τις βάσεις δεδομένων μοντέλων οργανισμών για την αντιπροσώπευση και αποθήκευση

βιολογικών πληροφοριών με αποτέλεσμα οι σημαντικότερες από αυτές, αλλά και πολλά από τα μεγάλης κλίμακας προγράμματα γενετικών μεταλλάξεων να χρησιμοποιούν στην οντολογία γονιδίων, γνωστή ως Gene Ontology (GO) για την αντιπροσώπευση λειτουργικών γονιδιακών «ιδιοτήτων» με δύο σημαντικές συνέπειες. Η πρώτη είναι ότι τα δεδομένα είναι πια προσπα σε υπολογιστικές μεθόδους, όπως βεβαιώνεται από τη μεγάλη χρήση των GO σχολιασμών για την ανάλυση, παραδείγματος χάρη, microarray δεδομένων, και η δεύτερη ότι υπάρχει, τουλάχιστον σε αυτόν τον περιορισμένο τομέα, μια *de facto* ενσωμάτωση αυτών των δεδομένων από έναν μεγάλο αριθμό διάσπαρτων βάσεων δεδομένων. Ο συνδυασμός αυτός προκάλεσε ιδιαίτερο ενδιαφέρον εντός της κοινότητας για την επέκταση αυτού του μοντέλου και σε άλλες γνωστικές περιοχές, όπως στην οντολογία ακολουθίας (Sequence Ontology) που χρησιμοποιείται από διάφορες βάσεις δεδομένων για την αντιπροσώπευση ακολουθιών, αλλά και την οντολογία ανατομίας που είναι ανεξάρτητη συγκεκριμένου οργανισμού με συνεπεία την ενσωμάτωση δεδομένων από διάφορους πόρους. Στις αρχές του 2003, ο Ashburner και ο Γκούτος πρότειναν την ανάπτυξη ενός εννοιολογικού πλαισίου για την αυστηρή αντιπροσώπευση των φαινοτυπικών δεδομένων οδηγώντας στη δημιουργία ενός εννοιολογικού μοντέλου που αποκάλεσαν οντολογία φαινότυπου και γνωρίσματος (Phenotype & Trait Ontology - PATO).

ΟΝΤΟΛΟΓΙΑ ΚΑΙ ΦΙΛΟΣΟΦΙΑ

Η οντολογία ως κλάδος της φιλοσοφίας είναι η επιστήμη που εξετάζει την υπόσταση των διαφόρων ειδών και δομών των αντικειμένων, των ιδιοτήτων τους καθώς και των σχέσεών τους σε κάθε τομέα της πραγματικότητας. Ως όρος, προερχόμενος από την ελληνική γλώσσα, δημιουργήθηκε το 1613 ανεξάρτητα από δύο φιλοσόφους, τους Rudolf Gockel (*Lexicon philosophicum*) και Jacob Lorhard (*Theatrum philosophicum*), ενώ στη φιλοσοφία συνήθως αποτελεί τη μελέτη της υπόστασης των διάφορων ειδών και δομών των αντικειμένων, των ιδιοτήτων τους καθώς και των σχέσεών τους σε κάθε τομέα της πραγματικότητας. Η έννοια της οντολογίας χρησιμοποιείται συχνά από τους

φιλοσόφους ως συνώνυμο των μεταφυσικών, ενός όρου που χρησιμοποιήθηκε από τους πρώτους μαθητές του Αριστοτέλη για να αναφερθούν σε αυτό που ο ίδιος ο Αριστοτέλης αποκαλούσε Πρώτη Φιλοσοφία. Η οντολογία ασχολείται με το τι υπάρχει ή καλύτερα την έννοια της ύπαρξης, ενώ ορισμένες φορές κατά μια γενικότερη έννοια πραγματεύεται τη μελέτη του τι μπορεί να υπάρχει. Στη φιλοσοφία έχει τουλάχιστον τέσσερις καθιερωμένες έννοιες.

Την οντολογία υπό την καθαρή φιλοσοφική έννοια (που ασχολείται με το γιατί υπάρχει κάτι από το να μην υπάρχει τίποτα), την οντολογία ως τομέας γνώσεως που ασχολείται με τα φιλοσοφικά προβλήματα των οντοτήτων, την οντολογία ως κλάδο που πραγματεύεται τη μελέτη του αποτελέσματος της οντολογίας ως τομέα γνώσεως και, τέλος, την εφαρμοσμένη επιστημονική οντολογία που επικεντρώνεται στη μελέτη ενός συγκεκριμένου επιστημονικού τομέα γνώσεως που εξετάζει τι υπάρχει στην πραγματικότητα. Ο τελευταίος τομέας χωρίζεται στην οντολογία ως *θεωρητικό* κλάδο που παρέχει μια περιγραφή των πραγμάτων που υποτίθεται ότι υφίστανται σύμφωνα με μια ιδιαίτερη θεωρία, άλλα που δεν είναι αναγκαίο να είναι αληθινά και στην οντολογία, ως *υφιστάμενο* κλάδο που ασχολείται με τον πραγματικό κόσμο όλων των υπάρχουσών οντοτήτων που προσδιορίζονται από μια τελείως πλήρη εφαρμοσμένη οντολογική θεωρία.

ΟΝΤΟΛΟΓΙΑ ΚΑΙ ΠΛΗΡΟΦΟΡΙΚΗ

Αντιθέτως, η οντολογία στην πληροφορική έχει μια διαφορετική και πάρα πολύ πρακτική σημασία. Το λεξικό του Webster ορίζει την οντολογία στην πληροφορική ως:

«μια κατηγοριοποίηση όλων των σημαντικών κατηγοριών αντικείμενων ή εννοιών που υπάρχουν σε κάποιο τομέα γνώσης συμπεριλαμβάνοντας και τις αναμεταξύ τους σχέσεις. Όταν είναι πλήρης η οντολογία αποτελεί την κατηγοριοποίηση όλων των εννοιών σε κάποιο τομέα γνώσης, συμπεριλαμβανομένων των αντικείμενων και των ιδιοτήτων τους, καθώς και των λειτουργιών που απαιτούνται για να τα καθορίσουν πλήρως. Μια απλου-

στευμένη οντολογία μπορεί να περιέχει μόνο μια ιεραρχική ταξινόμηση που παρουσιάζει τις σχέσεις μεταξύ των εννοιών που περιέχει. Μια οντολογία μπορεί να απεικονιστεί ως μια θεωρητική γραφική παράσταση με τους κόμβους και τα τόξα που αντιπροσωπεύουν τα αντικείμενα και τις σχέσεις».

Σύμφωνα με τον ορισμό του Webster η οντολογία είναι μια συστηματική κατανομή, κατηγοριοποίηση των εννοιών, ιεραρχική ταξινόμηση κτλ., ενώ σύμφωνα με τον πλέον διαδεδομένο ορισμό του Gruber η οντολογία είναι ένας ρητός και τυπικός ορισμός μιας εννοιολογικής σύλληψης. Έτσι λοιπόν, στον τομέα της Τεχνητής Νοημοσύνης αλλά και της (βιο)πληροφορικής, η οντολογία παίζει έναν πολύ σημαντικό ρόλο εξυπηρετώντας τις τεχνολογικές ανάγκες εκμετάλλευσης της πληροφορίας. Συνεπώς, συχνά αναφερόμαστε σε μια ορολογία δανεισμένη από τη φιλοσοφία που δεν αποτελεί μια θεωρία όπως στη φιλοσοφία, άλλα έναν ορισμό μιας εννοιολογικής σύλληψης που αποσκοπεί στην αντιπροσώπευση της πληροφορίας. Αυτό όμως που γίνεται συνεχώς όχι μονό επιθυμητό, άλλα και απολύτως αναγκαίο κατά τον σχεδιασμό της οντολογίας, είναι η ανάγκη εκμετάλλευσης βασικών αρχών προσαρμοσμένων από την εφαρμοσμένη επιστημονική οντολογία. Αυτές οι φιλοσοφικές αρχές θα μας βοηθήσουν στην οργάνωση και εκμετάλλευση της γνώσης.

ΟΝΤΟΛΟΓΙΑ ΚΑΙ ΓΕΝΕΤΙΚΗ

Λαμβάνοντας υπόψη την ιστορία της έρευνας στον τομέα της γενετικής των τελευταίων 50 ετών, η συμβολή της φαινοτυπικής ανάλυσης των γενετικών μεταλλάξεων έπαιξε ρόλο ζωτικής σημασίας στην κατανόηση της *ομαλής* λειτουργίας των γονιδίων καθώς οι φαινοτυπικές αναλύσεις μιας μετάλλαξης, αλλά και των συνδυασμών τους, προσδιόρισαν όχι μόνο τον βιολογικό σκοπό των μεμονωμένων γονιδίων αλλά και τη λειτουργία σύνθετων βιολογικών διαδικασιών. Εντούτοις, έχουν υπάρξει, με ελάχιστες εξαιρέσεις, πολύ λίγες οργανωμένες προσπάθειες συστηματοποίησης της περιγραφής των φαινοτυπικών πληροφοριών, ακόμη και σε πολύ καλά μελετημένους οργανισμούς, όπως π.χ. τη *δροσόφylla*.

Η περιγραφή και απεικόνιση των φαινοτυπικών πληροφοριών παρουσιάζει ένα τεράστιο εννοιολογικό και πρακτικό πρόβλημα για πολλούς και διάφορους λόγους ξεκινώντας από το γεγονός ότι οι φαινότυποι μπορούν να περιγραφούν σε πολλά διαφορετικά επίπεδα, από το κυτταρικό ως την εξέταση ενός πληθυσμού. Το ότι παρατηρούνται και καταγράφονται με διάφορες μεθόδους καθώς επίσης το ότι καθορίζονται (τουλάχιστον στην κλασική γενετική) με μεθοδολογίες χαμηλής απόδοσης και περιγράφονται σε μη δομημένη ελεύθερη μορφή. Συνήθως, η οντολογική περιγραφή ενός φαινότυπου συνδέεται με μεμονωμένες βάσεις δεδομένων και έχει εξελιχθεί με εξειδικευμένες μεθοδολογίες που καλύπτουν διαφορετικές ανάγκες. Εννοιολογικά, η περιγραφή των φαινότυπων απαιτεί τους συνδυασμούς ορθογώνιων οντολογιών με τη δυνατότητα συσχέτισης παραγόντων ανάλογα με τις πειραματικές τιμές. Σε πρακτικό επίπεδο, εάν τα βιολογικά δεδομένα πρόκειται να αναλυθούν υπολογιστικά με αποτελεσματικό τρόπο, υπάρχει ανάγκη για συνέπεια μεταξύ των εκφράσεων που χρησιμοποιούνται από διαφορετικές φαινοτυπικές περιοχές καθώς επίσης και από διαφορετικούς οργανισμούς.

Η αντιμετώπιση αυτού του προβλήματος είναι, εντούτοις, αναγκαία και ενώ, ακόμη και πριν από μόλις μία δεκαετία, ένας ερευνητής που εργαζόταν στη γενετική του ανθρώπου ή του ποντικού μπορούσε εύλογα να αγνοήσει τις φαινοτυπικές πληροφορίες σχετικές με, παραδείγματος χάριν, τη *δροσόφυλλα*, αυτό δεν ισχύει πλέον. Αν μας δίδαξε κάτι η εξέλιξη της γενετικής στην ανάλυση των διαφόρων γονιδιωμάτων και μεμονωμένων γονιδίων είναι ότι πάρα πολλές σύνθετες βιολογικές διαδικασίες είναι πολύ παρόμοιες και χρησιμοποιούν τις ίδιες πρωτεΐνες, σε πολύ διαφορετικούς οργανισμούς. Λίγοι, ενδεχομένως, θα είχαν προβλέψει δέκα χρόνια πριν ότι τα γονίδια που, για παράδειγμα, περιλαμβάνονται στην ανάπτυξη των οφθαλμών ή της καρδιάς, στη *δροσόφυλλα*, θα περιλαμβάνονταν επίσης στις ίδιες διαδικασίες και στα θηλασικά ή ότι η ανοσοολογική απάντηση των εντόμων στα παθογόνα και η οξεία ανοσοολογική απάντηση των θηλασικών είναι ομόλογες διαδικασίες.

Η συνέπεια του προαναφερθέντος είναι ότι όλοι οι βιοϊατρικοί ερευνητές χρειάζονται εύκολη και συστηματοποιημένη πρόσβαση σε φαινοτυπικές πληροφορίες όλων των μοντέλων οργανισμών (και εδώ περιλαμβάνουμε και τον

άνθρωπο) με τη δυνατότητα διατύπωσης ερωτημάτων, όπως για παράδειγμα: λαμβάνοντας υπόψη ποιοι είναι οι φαινότυποι των μεταλλάξεων αλληλόμορφων γονιδίων ενός γονιδίου που εκφράζεται στον οργανισμό Α καθώς και έναν φαινότυπο, π.χ. υπερτροφία της καρδιάς, και ποιες μεταλλάξεις αλληλόμορφων γονιδίων σε οποιοδήποτε σχετικό οργανισμό έχουν τον ίδιο φαινότυπο. Είναι πια σαφές ότι ο μόνος τρόπος με τον οποίο αυτή η απαίτηση μπορεί να ικανοποιηθεί είναι η αυστηρή αντιπροσώπευση των φαινοτυπικών πληροφοριών όσον αφορά τη σημασιολογία αυτής της απεικόνισης. Κατά συνέπεια, η πρόκληση είναι καταρχήν στη δημιουργία ενός δομημένου ελεγχόμενου λεξιλογίου για την αντιπροσώπευση των φαινοτυπικών δεδομένων μέσα στο πλαίσιο των βάσεων δεδομένων μοντέλων οργανισμών, καθώς και στη καθιέρωση της χρήσης αυτού του λεξιλογίου για την επισημείωση πραγματικών δεδομένων.

Οι οντολογίες αποτελούν σημαντικό μέσο για την απεικόνιση βιολογικών πληροφοριών από τον καιρό του Linnaeus και με τη σχετικά πρόσφατη εμφάνιση της οντολογίας των γονιδίων (Gene Ontology GO) το 2000. Αυτές οι τεχνικές αυστηρής διευκρίνισης των σημασιολογικών σχέσεων μεταξύ ορισμών έχουν γίνει πρότυπα που χρησιμοποιούνται για τη υποστήριξη της αντιπροσώπευσης της γνώσης στον τομέα της γενετικής. Οι ιεραρχικές οντολογίες φυλάσσουν τις πληροφορίες για τη δομή μιας ιδιαίτερης γνωστικής περιοχής με ποικίλους βαθμούς λεπτομέρειας, επιτρέποντας κατά συνέπεια, την ενσωμάτωση των εννοιών και των περιγραφών σε διαφορετικά επίπεδα διάκρισης. Αυτή η προσέγγιση αποτελεί τη βάση νέων μεθόδων υπολογιστικής ανάλυσης βιολογικών πληροφοριών. Μια δικαιολογημένη ερώτηση είναι γιατί να χρησιμοποιήσουμε μια οντολογία και όχι μια σειρά από αποσυνδεδεμένους, τυποποιημένους ορισμούς, όπως αυτοί παρέχονται από ένα ελεγχόμενο λεξιλόγιο; Τα πλεονεκτήματα της χρήσης οντολογιών έχουν υποστηριχτεί εκτενώς, αλλά ο κύριος λόγος είναι ότι οι οντολογίες προσπαθούν να συλλάβουν την ακριβή έννοια των όρων, με τρόπο που μπορεί να είναι υπολογιστικά εκμεταλλεύσιμος και συμπεριλαμβάνει την αυτοματοποιημένη λογική συναγωγή συμπερασμάτων, την τυπική συλλογιστική, τον έλεγχο συνέπειας κτλ. Από τη

σκοπιά της φαινοτυπικής γνώσης, ο σημαντικότερος παράγοντας είναι η ανάγκη συνδυασμού πληροφοριών από διαφορετικές φαινοτυπικές επισημειώσεις ή από διαφορετικά πρωτόκολλα (βιολογικές δοκιμασίες), καθώς επίσης και τις σχέσεις μεταξύ τους.

Η περιεκτικότερη προσπάθεια απεικόνισης φαινοτυπικών πληροφοριών μέχρι σήμερα, είναι η οντολογία φαινότυπου των θηλαστικών υπό ανάπτυξη από το Jackson Laboratory που αποκαλείται Mammalian Phenotype (MP), με δομή OBO κατασκευασμένη στο OBO-edit, το κοινό εργαλείο της κοινότητας, επιτρέποντας μια ιεραρχική κατάταξη των όρων και των ορισμών τους. Αυτοί οι όροι περιλαμβάνουν έναν συνδυασμό οντοτήτων και τιμών, π.χ. ο κωδικός MP:0001509, που αντιστοιχεί στην *ανώμαλη θέση σώματος*, και προσφέρει ικανοποιητική, χαμηλής ευκρίνειας περιγραφή της φαινοτυπικής πληροφορίας. Αυτή η προσέγγιση επιτρέπει μια αντίστοιχης ευκρίνειας πρόσβαση στη γνώση που περιέχεται στην οντολογία, αλλά έχει ορισμένα σημαντικά μειονεκτήματα παρόμοια με αυτά που απαντώνται στην οντολογία γονιδίων (GO), με κυριότερο την ανεξέλεγκτη δημιουργία νέων ορισμών και την παραγκώνιση των σχέσεων προέλευσης τους για την κάλυψη ιδιομορφιών. Για παράδειγμα, δημιουργώντας μόνο τους αναγκαίους νέους όρους, συνοδευόμενους από την πρόσθεση των ποσοτικών ιδιοτήτων «αυξανόμενο» και «μειωμένο», το μέγεθος της οντολογίας αυξάνεται σε τεράστιο βαθμό, ενώ για να επιτραπεί μια συστηματική προσέγγιση του μοντέλου, είναι αναγκαία η δημιουργία συνδυασμών οντοτήτων που πιθανότητα δεν θα χρησιμοποιηθούν ποτέ. Για πρακτικούς λόγους, επίσης, θα πρέπει να τεθούν περιορισμοί στη δημιουργία νέων ορισμών, με αποτέλεσμα, αναπόφευκτα, να ληφθούν πιθανόν εσφαλμένες ή κακώς ενημερωμένες αποφάσεις ως προς το ποιοι συνδυασμοί οντοτήτων περιγράφουν καλύτερα τις διάφορες φαινοτυπικές παρατηρήσεις. Σημειώνουμε εδώ ότι η ανάπτυξη του MP έγινε κατ' αυτό τον τρόπο, λόγω πρακτικών αναγκών, παρόμοιων με αυτών του GO και οι υπεύθυνοι για τη δημιουργία του είναι μέρος του Phenotype Consortium, αναπτύσσοντας σκόπιμα την οντολογία τους κατά τέτοιο τρόπο ώστε να μπορεί να επεκταθεί εύκολα σύμφωνα με το προτεινόμενο μοντέλο του PATO.

Άλλη μια προσπάθεια είναι η οντολογία συμπεριφοράς των ποντικών που αναπτύχθηκε από τον Γκούτο μέσα στο πλαίσιο του μεγάλης κλίμακας ευρωπαϊκού προγράμματος EUMORPHIA. Η οντολογία αυτή βασίζεται στο προτεινόμενο μοντέλο και αποτελεί το πρώτο παράδειγμα του συνδυασμού μιας ανεξάρτητης οντολογίας με το ΡΑΤΟ (συνεπώς την πρώτη οντολογία φαινότυπων - ΡΟ) και είναι υπεύθυνη για την επισημείωση 600.000 φαινοτυπικών παρατηρήσεων.

ΡΑΤΟ - ΟΝΤΟΛΟΓΙΑ ΦΑΙΝΟΤΥΠΟΥ ΚΑΙ ΓΝΩΡΙΣΜΑΤΟΣ

Η οντολογία φαινότυπου και γνώρισματος αποσκοπεί στην παροχή ενός εννοιολογικού πλαισίου που θα διευκολύνει την αμοιβαία κατανόηση και διαλειτουργικότητα φαινοτυπικών πληροφοριών διάφορων ειδών και περιοχών γνώσεως μεταξύ ανθρώπων και υπολογιστών. Η κεντρική ιδέα του ΡΑΤΟ βασίζεται στο γεγονός ότι οι φαινότυποι μπορούν να περιγραφούν σε πολλές διαφορετικές διαστάσεις, π.χ., σε βιοχημικό επίπεδο («αυξημένη παρουσία αλκοολικής αφυδρογονάσης»), σε κυτταρικό επίπεδο («διακοπή της κυτταροδιαίρεσης στη μετάφραση»), σε ανατομικό επίπεδο («απουσία οφθαλμού»), ή σε επίπεδο συμπεριφοράς («αυξημένη επιθετικότητα»). Εντούτοις, σε όλα τα επίπεδα υπάρχουν κοινά χαρακτηριστικά που επιτρέπουν την αποσύνθεση της πλειονότητας των φαινοτυπικών περιγραφών σε τρία μέρη:

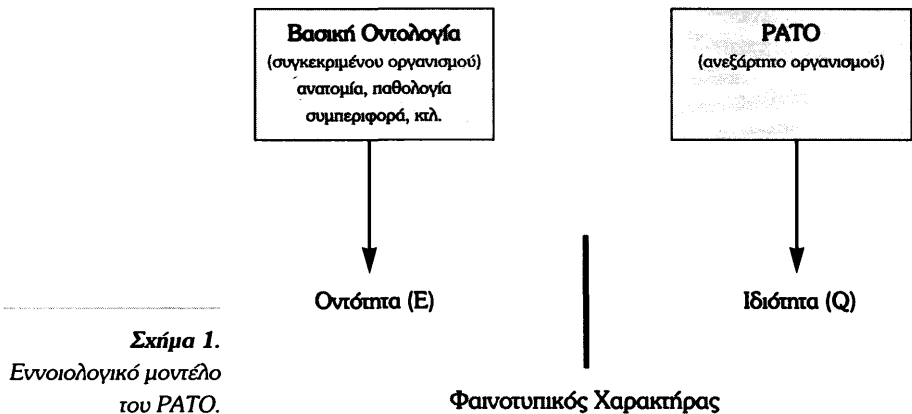
- Στην επηρεασμένη οντότητα (entity E) που μπορεί να είναι ένα ένζυμο, μια ανατομική δομή ή μια σύνθετη βιολογική διαδικασία.
- Στις ιδιότητες (quality Q) αυτής της οντότητας.
- Στην τιμή αυτών των ιδιοτήτων που μπορεί να είναι ποσοτική ή ποιοτική.

Οντότητες: Μέσα στο πλαίσιο της οντολογίας φαινότυπου (Phenotype Ontology - ΡΟ) ο ορισμός της «οντότητας» προέρχεται από μια ανεξάρτητη οντολογία ή μια βάση δεδομένων. Έτσι, η οντότητα αυτή μπορεί να προέρχεται από μια οντολογία βιολογικών διαδικασιών (π.χ. η οντολογία του GO biological process), μια βάση πρωτεϊνικών δεδομένων (π.χ. από UniProt), μια ανατομική οντολογία (όπως υπάρχει τώρα για διάφορα μοντέλα οργανισμούς) κτλ.

Ιδιότητες: Αυτές είναι οι κατηγορίες ιδιοτήτων της οντότητας που διαφέρουν ανάλογα με την οντότητα. Κατά συνέπεια, μια ανατομική οντότητα μπορεί να έχει τις ιδιότητες της «μορφής», του «μεγέθους», σε μερικές περιπτώσεις του «χρώματος», μια πρωτεϊνική οντότητα μπορεί να έχει την ιδιότητα της «σταθερότητας», ενώ μία βιολογικής διαδικασίας οντότητα μπορεί να πάρει την ιδιότητα του «ρυθμού».

Τιμές: Για οποιοσδήποτε δεδομένες ιδιότητες θα υπάρχει μια σειρά επιτρεπόμενων τιμών που μπορούν να είναι απόλυτες ή σχετικές και μπορούν να εκφραστούν με διάφορες δεδομένες επιτρεπόμενες μονάδες. Η ιδιότητα του χρώματος θα μπορούσε να πάρει τιμές, όπως «άσπρο | μαύρο | γκρι | κόκκινο |...», μια ανατομική δομή να έχει την ιδιότητα του «μεγέθους» με τιμές, όπως «μεγάλο | μικρό» ή απόλυτες τιμές «2 κιλ.» κτλ.

Η οντολογία ιδιοτήτων και τιμών (ως μέρος του PATO), η οποία βρίσκεται στην ιστοσελίδα του Open Biological Ontologies, περιλαμβάνει 1.562 όρους, και ο συνδυασμός του PATO με τις ανεξάρτητες οντολογίες (π.χ. την ανατομία) και τις βιολογικές δοκιμασίες αποκαλείται οντολογία φαινότυπου (Phenotype Ontology - PO). Το μοντέλο αυτό έχει εφαρμοστεί με μεγάλη επιτυχία σε πειραματικό επίπεδο στο πλαίσιο της FlyBase (P. Drysdale, *pers. comm.*), του FIN τη βάση δεδομένων του μοντέλου zebra fish (E. Segerdell, *pers. comm.*, http://www.cs.uoregon.edu/~eriks/Documents/ZFIN_PATO_trial_data.htm) και του MRC (Γκούτος *pers. comm.*).



Το εννοιολογικό μοντέλο του ΡΑΤΟ συνδυάζει τις βασικές οντολογίες, συγκεκριμένες κάποιου οργανισμού π.χ. ανατομία της Δροσόφylla, ανατομία του ποντικού κτλ. και που καθορίζουν τις σχετικές οντότητες, με την οντολογία του ΡΑΤΟ, που συλλέγει όλες τις ιδιότητες που παίρνουν μέρος σε μια φαινοτυπική έκφραση και το οποίο είναι ανεξάρτητο οποιαδήποτε οργανισμού. Αποκαλούμε τον συνδυασμό αυτών των δύο «φαινοτυπικό χαρακτήρα», που όχι μόνο είναι ικανός να δημιουργεί φαινοτυπικές οντολογίες συγκεκριμένου οργανισμού, αλλά ίσως κυριότερα επιτρέπει μια κοινή φαινοτυπική πλατφόρμα μεταξύ διάφορων οργανισμών.

Το μοντέλο του ΡΑΤΟ επιτρέπει την απεικόνιση του τρόπου που αντιλαμβανόμαστε τον βιολογικό κόσμο, καταγράφοντας παρατηρήσεις μιας βιολογικής πραγματικότητας σε μια συγκεκριμένη χρονική στιγμή, όπως, για παράδειγμα, ότι τα ποντίκια έχουν κάποιο βάρος ή ότι το χρώμα των ματιών της δροσόφylla είναι κόκκινο ή ότι τα άτομα που πάσχουν από διαβήτη έχουν ανεβασμένα επίπεδα σακχάρου κτλ. Από αυτές τις παρατηρήσεις εξάγουμε κάποιες σταθερές και αποικοδομούμε διάφορους τομείς γνώσεως καθώς και σχέσεις αναμεταξύ τους. Για να περιγράψουμε τους διάφορους φαινοτυπικούς χαρακτήρες συνδυάζουμε οντότητες προερχόμενες από τις διάφορες βασικές οντολογίες, όπως την ανατομία του ποντικού, της δροσόφylla κτλ. με ιδιότητες προερχόμενες από το ΡΑΤΟ, όπως το βάρος, το χρώμα, τη συγκέντρωση (Πίνακας 1).

Πίνακας 1. Παραδείγματα δημιουργίας φαινοτυπικών χαρακτήρων βάση του ΡΑΤΟ.

Φαινοτυπικός Χαρακτήρας	Οντότητα	Ιδιότητα
βάρος σώματος	ανατομία ποντικού: σώμα	ΡΑΤΟ: βάρος
χρώμα ματιών	Ανατομία Δροσόφylla: οφθαλμός	ΡΑΤΟ: χρώμα
συγκέντρωση γλυκόζης	ChEBI: γλυκόζη	ΡΑΤΟ: συγκέντρωση

ΠΡΟΣ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΕΝΟΣ ΦΑΙΝΟΤΥΠΙΚΟΥ ΣΗΜΑΣΙΟΛΟΓΙΚΟΥ ΙΣΤΟΥ

Υπάρχουν δεκάδες χιλιάδες γονίδια και ο αριθμός των πιθανών αλληλεξαρτήσεων μεταξύ τους είναι τεράστιος. Η ρύθμιση της μεταγραφής και μετάφρασης, τα *positive* και *negative controls*, τα *upstream* και *downstream effects*, και τα *redundant systems* αποτελούν τρόπους με τους οποίους τα γονίδια μπορούν να αλληλεπιδράσουν και ένα μεγάλο κομμάτι της έρευνας στη γενετική αποτελείται από προσπάθειες συγκέντρωσης τέτοιου είδους πληροφοριών. Αυτά τα σύνολα δεδομένων είναι μεγάλα και σύνθετα και, προφανώς, απαιτούν καλή διαχείριση καθώς και καλά ανεπτυγμένους και στοιχειοθετημένους τρόπους καταγραφής, ξεκινώντας από το εργαστηριακό επίπεδο. Υπάρχει όμως ένας πιο αναγκαίος λόγος για μια εξελιγμένη διαχείριση πληροφοριών, πέρα των αναγκών οποιασδήποτε ανεξάρτητης ερευνητικής ομάδας, και αυτός είναι η ανάπτυξη της βιολογικής έρευνας σε συλλογικό επίπεδο. Οι προσπάθειες μας, μέσω του OBO [22] και του CBIO [23], αποσκοπούν στη δημιουργία μιας σηματολογικής υποδομής βασισμένης σε μια οντολογική αντιπροσωπευση της γνώσης μας, που θα συνοδεύεται από συνεπή περιγραφή των βιοϊατρικών αποτελεσμάτων και πειραμάτων, σηματολογική διαλειτουργικότητα των resources, Grid υποδομές για τη συλλογή, σχολιασμό, αποθήκευση, ανάληψη, πρόσβαση και σύγκριση βιολογικών δεδομένων που επιτυγχάνει την πρώτη πραγματοποίηση της κρισιμότερης πρόκλησης στη βιοπληροφορική σήμερα - αυτή της ερμηνείας της έννοιας της βιοϊατρικής πληροφορίας. Στόχος μας είναι η ενσωμάτωση πληροφοριών, βάση οντολογικών αρχών, επιτρέποντας την ερμηνεία των δεδομένων και τη δημιουργία συμπερασμάτων που μπορούν να προκύψουν βασισμένα στην υπάρχουσα γνώση, ενώ θα χρησιμοποιήσει τις βιοϊατρικές πληροφορίες με τρόπο που να επιτρέπει στους επιστήμονες και στους υπολογιστικούς πόρους τη διαμόρφωση υποθέσεων, τη θέση προτεραιοτήτων, αλλά και τη συμμετοχή στη διαδικασία λήψης αποφάσεων, η οποία είναι έμφυτη σε οποιοδήποτε πείραμα, καθιστώντας, λοιπόν, εφικτό τον αυτόματο υπολογισμό που θα βοηθήσει στην αποφυγή πειραματικών αδιεξόδων αυξάνοντας τις πιθανότητες ανακάλυψης λανθασμένων δεδομένων.

Σύμφωνα με την ανάλυση της έρευνας στο γενικό τομέα της γενετικής, όπως αυτός υποστηρίζεται από το FP6 (έκθεση EYP-208111 2004 της Ευρωπαϊκής Επιτροπής), η ανταλλαγή βιοϊατρικών πληροφοριών, βασισμένη στα υποστηριγμένα από το FP5 έργα, τονίζεται ως ευρωπαϊκή ανάγκη αλλά η βιοπληροφορική κρίνεται ανεπαρκής και επισημαίνεται ως προτεραιότητα για το FP7. Ο όγκος των δεδομένων φαινοτυπικών μεταλλάξεων που προκύπτουν από ποικίλα μεγάλα ευρωπαϊκά προγράμματα (όπως το EUMORPHIA, EUCOMM, Europhenome, PRIME) και διάφορων άλλων προγραμμάτων χρηματοδοτούμενων από το FP6, καθώς επίσης και τα εξίσου μεγάλα προγράμματα υπό εξέλιξη στις ΗΠΑ (π.χ. The Mouse Phenome Project), τον Καναδά, την Αυστραλία και την Ιαπωνία, καθιστά απολύτως αναγκαία τη δημιουργία κατάλληλων υποδομών βιοπληροφορικής. Το εννοιολογικό πλαίσιο του PATO ενσωματώνει μια εφικτή, διεθνώς αποδεκτή και επικυρωμένη, προσέγγιση για μια τέτοια υποδομή. Ο αντίκτυπος της σημασιολογικής αντιπροσώπευσης των φαινοτυπικών δεδομένων (PATO), ο τρόπος που παράγονται και συλλέγονται (δομημένο λεξιλόγιο βιολογικών δοκιμασιών), η πραγματοποίηση μιας σημασιολογικής διαλειτουργικότητας των πόρων (η κατασκευή του σημασιολογικού πλέγματος για τους φαινότυπους), η υποδομή (βοηθούμενη από τα εργαλεία και παραδείγματα) για τη φαινοτυπική συλλογή δεδομένων, ο σχολιασμός, η αποθήκευση, η ανάλυση, η πρόσβαση και η σύγκριση βιολογικών δεδομένων, όχι μόνο θα βελτιώσει την ανταγωνιστικότητα της ευρωπαϊκής έρευνας σε παγκόσμιο επίπεδο, αλλά εξετάζει μερικά από τα τρέχοντα θεμελιώδη και τεχνολογικά προβλήματα. Σήμερα, η ευρωπαϊκή ακαδημαϊκή και βιομηχανική έρευνα στη γενετική, ειδικά η σχετική με το ποντίκι ως μοντέλο, υπερέχει σε σχέση με τον υπόλοιπο κόσμο. Η επίτευξη του προτεινόμενου μοντέλου θα καλύψει μια κρίσιμη ανάγκη (χαρακτηριζόμενη ως η επόμενη πρόκληση σε αυτό τον τομέα) για μια συνεπή υπολογιστική στρατηγική για τη δημιουργία σημασιολογικής υποδομής για τη διαχείριση δεδομένων. Η διαλειτουργικότητα των πόρων παραδίδει ενδεχομένως την ποθητή σύμπραξη της ολοκλήρωσης και ανάλυσης ενός εκτεταμένου φάσματος δεδομένων με τη συνακόλουθη προστιθέμενη αξία για τα ερευνητικά προγράμματα. Επιπλέον, οι θεωρητικές

και τεχνολογικές προκλήσεις που απαντώνται, μπορούν να εφαρμοστούν για την επίλυση πολλών βιοϊατρικών ζητημάτων και να αποτελέσουν παραδείγματα για την επίλυση αναγκών εκτός του βιολογικού χώρου.

Κατά συνέπεια, η συλλογική προσπάθεια του ΟΒΟ αναμένεται να ασκήσει πραγματική και σημαντική επίδραση στον τομέα της υγείας, διευκολύνοντας τη δημιουργία ηλεκτρονικών πόρων γνώσης που μπορούν να ενισχύσουν τον σχεδιασμό και την εκτέλεση πειραμάτων, την ανάλυση δεδομένων, τη σύνθεση πληροφοριών, τη δημιουργία ερευνητικών υποθέσεων και την αποκάλυψη νέων δεδομένων για τη μοριακή βάση της ανθρώπινης ασθένειας.

ΕΥΧΑΡΙΣΤΙΕΣ

Ο συγγραφέας θα ήθελε να ευχαριστήσει τη Δρ Λουκία Τσαπρούνη από το Imperial College του Λονδίνου, την Έλενα Γκούτου και τον Βαγγέλη Μιλιαράκη για τη βοήθεια τους στη μετάφραση και την απόδοση του κειμένου στην ελληνική γλώσσα.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. EUMORPHIA, see <http://www.eumorphia.org>.
2. EUREGENE - The European Renal Genome Project., see <http://www.euregene.org/>
3. PRIME - Priorities for Mouse Functional Genomic Research in Europe. Co-ordinator Steve Brown
4. The Mouse Phenome Project, see <http://aretha.jax.org/pub/cgi/phenome/mpdcgi?rtn=docs/home>
5. EUCOMM - European Conditional Mouse Mutagenesis Program. Co-ordinator Wolfgang Wurst
6. The European Mutant Mouse Archive EMMA (Co-ordinator Glauco Tocchini-Valentini), see <http://www.emma.rm.cnr.it/>
7. Gene Ontology www.geneontology.org; - The phenotype consortium

8. Eilbeck, K., Lewis, S.E., Mungall, C.J., Yandell, M., Stein, L., Durbin, R., Ashburner, M., The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol.*, 2005; 6(5):R44. PMID: 15892872.
9. XSPAN - A cross-species anatomy project, see <http://www.xspan.org>.
10. B. Smith, *Ontology, Blackwell Guide to the Philosophy of Computing and Information*, Oxford: Blackwell, 2003, 155–166.
11. D. Jacquette, *Ontology*, Acumen Books / Central Problems of Philosophy Series, 2002.
12. Webster Online Dictionary, [http:// www.websters-online-dictionary.org/](http://www.websters-online-dictionary.org/)
13. T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220, 1993.
14. T. R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Presented at the Padua workshop on Formal Ontology, March 1993.
15. Braver, N. B. The mutants of *Drosophilamelanogaster* classified according to body parts affected. Carnegie Inst. Wash. 1956; Pub. 552.
16. GO Consortium: The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res*, 2004, 32: 258-261.
17. Camon E., Magrane M., Barrell D., Lee V., Dimmer E., Maslen J., Binns D., Harte N., Lopez R., Apweiler R.: The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res*, 2004; 32, 262-266.
18. Mammalian Phenotype Ontology, Mouse Genome Informatics Web Site, The Jackson, Laboratory, Bar Harbor, ME. see http://www.informatics.jax.org/searches/MP_form.shtml
19. Richter J., Lewis S.: DAG-Edit see <http://www.geneontology.org/doc/GO.tools.html#dagedit>
20. UniProt, see <http://www.ebi.uniprot.org/index.shtml/>
21. Gkoutos, G.V. & Ashburner, M., *unpublished see*, <http://obo.sourceforge.net/cgi-bin/detail.cgi?poav>); Gkoutos G.V., *et.al*. Using Ontologies to describe mouse *phenoty*, *pes*, *Genome Biolog*, 6(1):R8, 2005, 1465-6914
22. Open Biological Ontologies, see <http://obo.sourceforge.net>
23. The National Center for Biomedical Ontology see <http://www.bioontology.org/>

24. European Conditional Mouse Mutagenesis – EUCOMM, see <http://www.eucomm.org/>

25. Europhenome see <http://www.eumorphia.org/europhenome/browser/servlet/europhenome.Query>

26. The Mouse Phenome Project, see <http://aretha.jax.org/pub-cgi/phenome/mpdcgi?rtn=docs/home>