

# ELIXIR-GR: ΥΠΟΔΟΜΕΣ ΕΡΓΑΛΕΙΩΝ ΓΟΝΙΔΙΩΜΑΤΙΚΗΣ ΝΕΑΣ ΓΕΝΙΑΣ

Τσέτσος Φώτιος<sup>1</sup>, Τσιφιντάρης Μαργαρίτης<sup>1</sup>, Δημόπουλος Αλέξανδρος<sup>2</sup>, Λιάμπα Ειρήνη<sup>3</sup>, Μανώλης Απόστολος<sup>4</sup>, Δε Λαστίκ Έκτορ Χαβιέ<sup>3</sup>, Χατζηγιάννου Αριστοτέλης<sup>3</sup>, Γαλάνης Αλέξιος<sup>1</sup>, Τσαμαρδίνος Ιωάννης<sup>5</sup>, Χατζηγεωργίου Άρτεμις<sup>6,7</sup>, Παπαγεωργίου Αριστοτέλης<sup>4</sup>, Πάσχου Περιστέρα<sup>1</sup>

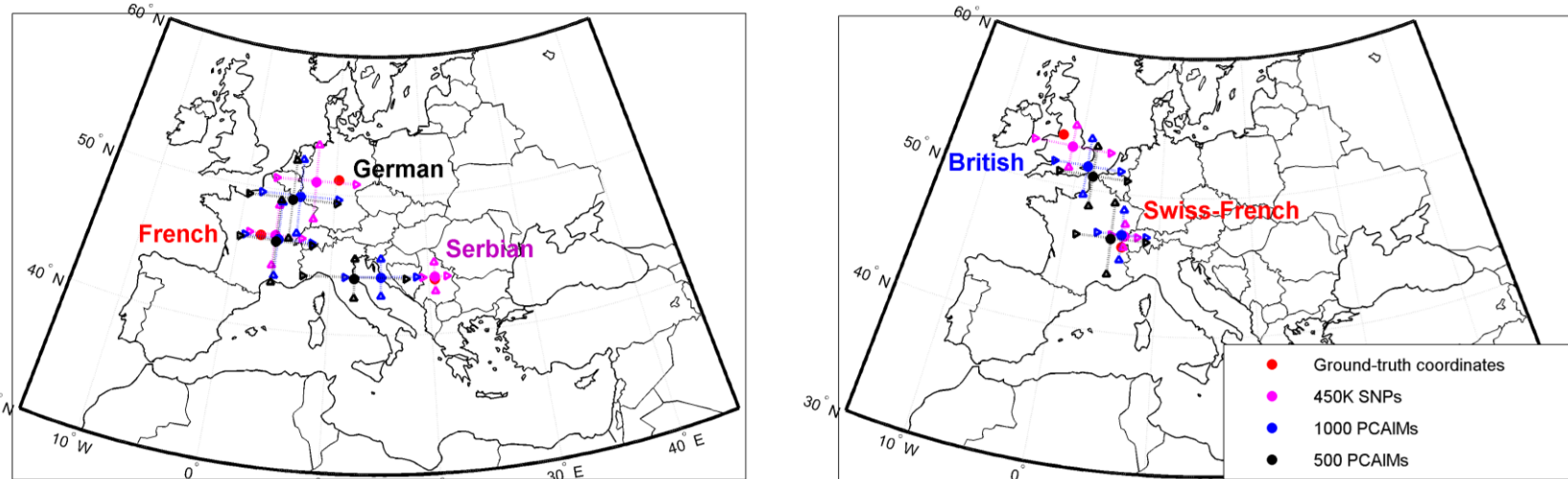
<sup>1</sup>Τμήμα Μοριακής Βιολογίας και Γενετικής, Δημοκρίτειο Πανεπιστήμιο Θράκης <sup>2</sup>Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών “Αλέξανδρος Φλέμινγκ” <sup>3</sup>Ινστιτούτο Βιολογίας, Φαρμακευτικής Χημείας και Βιοτεχνολογίας, Εθνικό Ίδρυμα Ερευνών <sup>4</sup>Τμήμα Δασολογίας και Διαχείρισης Περιβάλλοντος και Φυσικών Πόρων, Δημοκρίτειο Πανεπιστήμιο Θράκης <sup>5</sup>Τμήμα Πληροφορικής, Πανεπιστήμιο Κρήτης <sup>6</sup>Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας <sup>7</sup>Ελληνικό Ινστιτούτο Παστέρ

## Εισαγωγή

Η μεγάλη ποσότητα των δεδομένων αλληλούχησης έχει δημιουργήσει το πεδίο της γονιδιωματικής, έρευνα που χρησιμοποιεί υπολογιστικά εργαλεία για να αναζητήσει και να αναλύσει τα πρότυπα στα πλήρη γονιδιώματα των οργανισμών. Η γονιδιωματική μπορεί επίσης να θεωρηθεί ως υποπεδίο της βιοπληροφορικής, η οποία χρησιμοποιεί υπολογιστικές προσεγγίσεις για την ανάλυση μεγάλων συνόλων βιολογικών δεδομένων. Στην παρούσα εργασία παρουσιάζουμε δυο διαδικτυακές πλατφόρμες (Variant Ranker, Network Analyzer)<sup>[1]</sup> και μία μέθοδο (PCASNP)<sup>[2]</sup> για την ανάλυση γενετικών δεδομένων υψηλής απόδοσης, μαζί με ένα πλαίσιο προγραμματιστικών διοχετεύσεων.

## Υλικά και Μέθοδοι

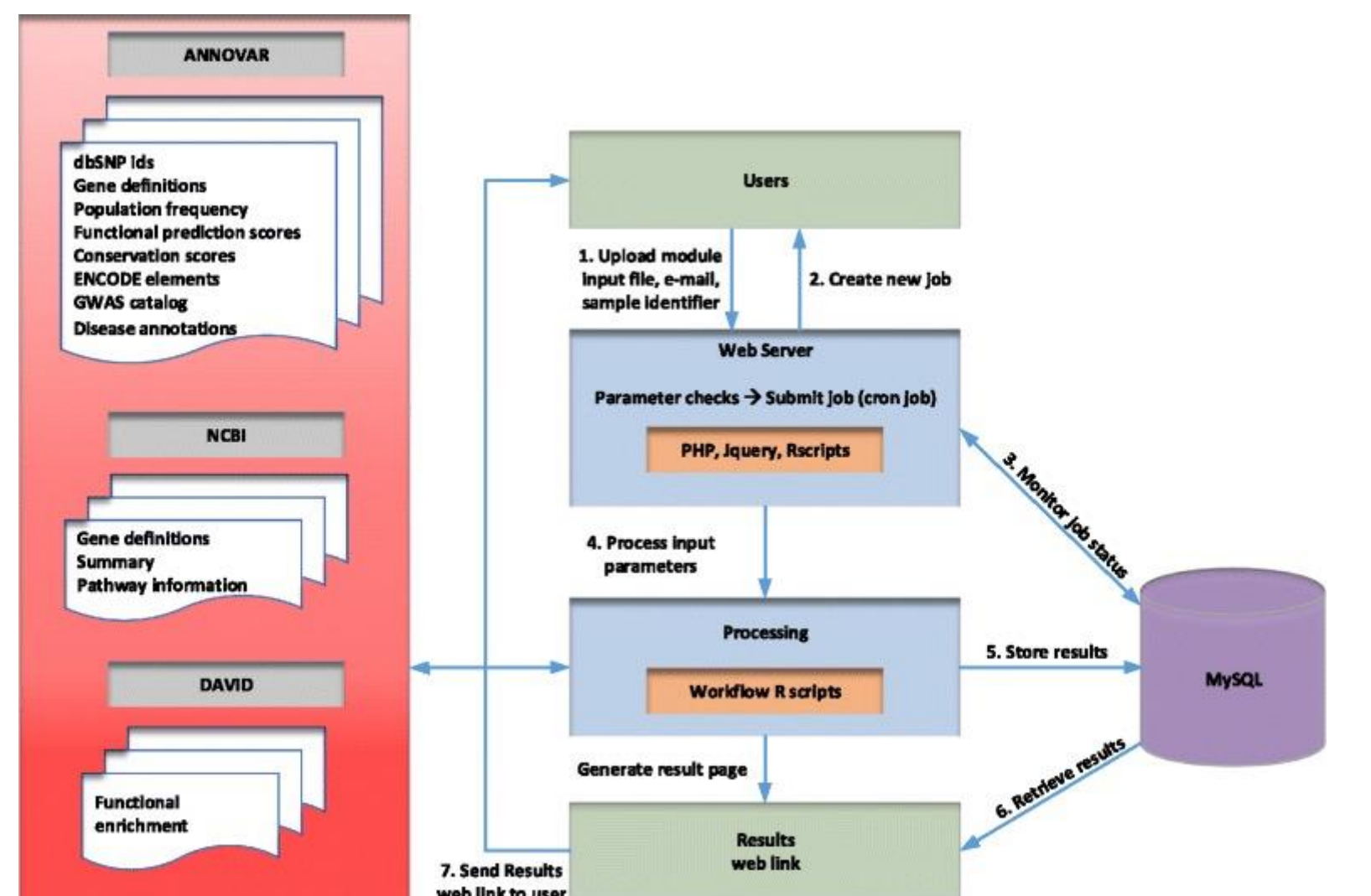
Για την προετοιμασία των δεδομένων αλλά και για τη δημιουργία των γραφημάτων χρησιμοποιήθηκαν script σε γλώσσες προγραμματισμού όπως η Python, η R και η Bash. Στην ανάλυση συμπεριλαμβάνονται μόνο οι πολυμορφισμοί ενός νουκλεοτιδίου (SNPs) και πιο συγκεκριμένα, όσοι έχουν 2 αλληλόμορφα. Για την μέθοδο PCASNP έχει δημιουργηθεί ένας αλγόριθμος ο οποίος υπολογίζει τον ελάχιστο αριθμό των SNPs που απαιτούνται για τον προσδιορισμό της καταγωγής σε μία ανάλυση<sup>[2]</sup>.



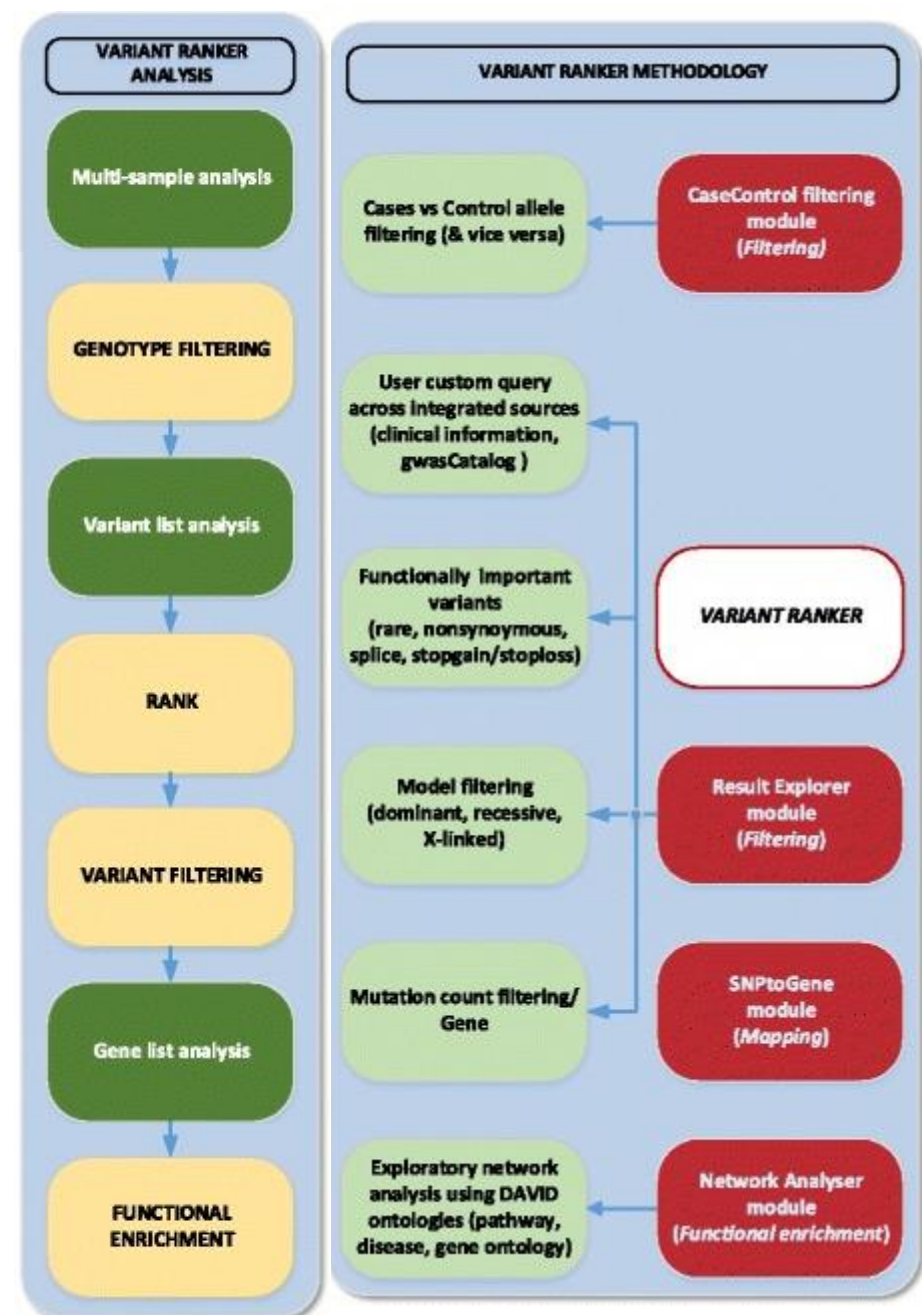
**Εικόνα 1.** Ανάλυση Κυρίων Συνιστωσών (PCA) σε συγκεκριμένους πλυθησμούς και προσδιορισμός της καταγωγής τους χρησιμοποιώντας διαφορετικά μεγέθη δεικτών<sup>[3]</sup>.

Τα εργαλεία VariantRanker και NetworkAnalyzer είναι διαθέσιμα μέσω μιας διαδικτυακής πλατφόρμας ακολουθώντας τον παρακάτω σύνδεσμο.

<http://paschou-lab.mbg.duth.gr/index/login/homeNR.php>



**Εικόνα 2.** Αρχιτεκτονική της πλατφόρμας Variant Ranker και διάγραμμα ροής.



**Εικόνα 3.**

Λειτουργικότητα της πλατφόρμας διαμέσου διαφόρων ενοτήτων (modules) για ανάλυση πολυμορφισμών και γονιδίων.

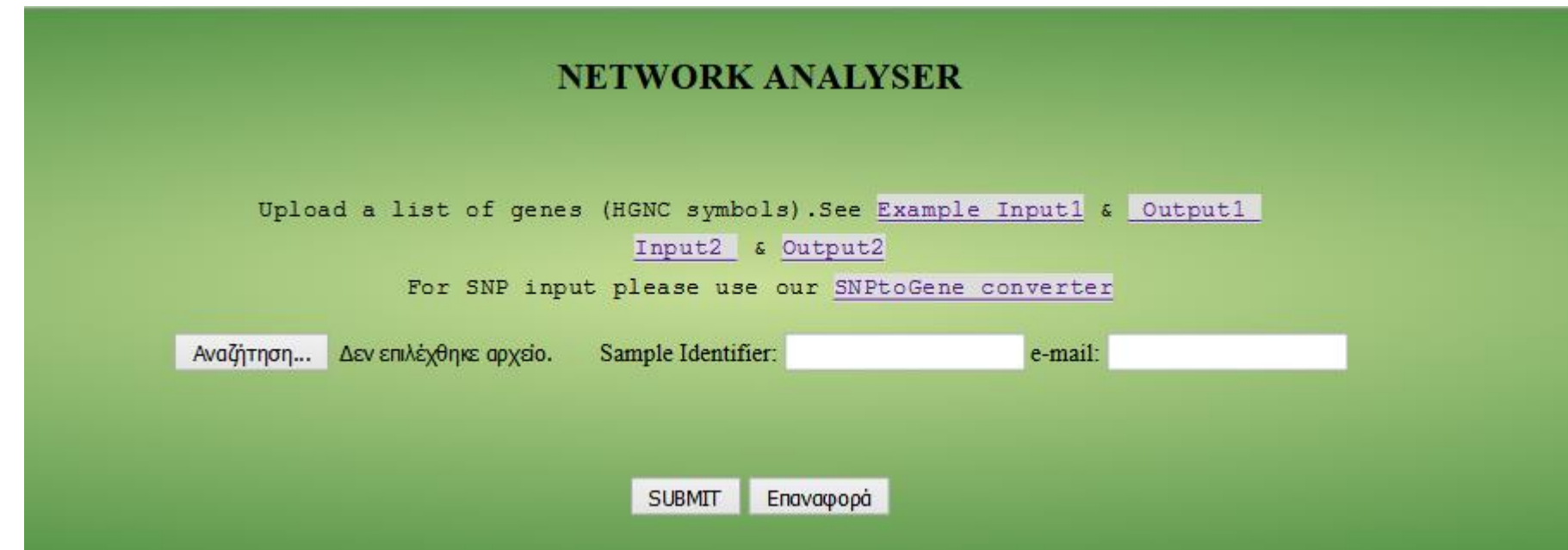
## Αποτελέσματα

Το **VariantRanker** είναι ένα εργαλείο για την κατάταξη και την επισημείωση κωδικεύουσων και μη κωδικεύουσων μεταλλαγών και διευκολύνει την αναγνώριση των αιτιουσών μεταλλαγών, βασιζόμενο σε καινοτόμες πληροφορίες μοριακών επιδράσεων και επισημειώσεις. Οι χρήστες μπορούν να αναζητήσουν και να φιλτράρουν μεγάλα σύνολα δεδομένων υψηλής απόδοσης και να διερευνήσουν διάφορα μοντέλα κληρονομότητας.



**Εικόνα 4.** (Α) Προκαθορισμένες τιμές για κάθε παράμετρο του VariantRanker. Οι τιμές αυτές μπορούν να τροποποιηθούν από τον χρήστη. (Β) Τα 20 κορυφαία υποψήφια γονίδια από την ανάλυση εξώνων ενός ατόμου με ιδιοπαθή αιμολυτική αναιμία (IHA) για την οποία το PKLR αναγνωρίστηκε ως το πιο πιθανό παθογόνο γονίδιο.

Το **NetworkAnalyzer** δημιουργεί λειτουργικά δίκτυα από λίστες ταυτοποιημένων γονιδίων. Οι χρήστες μπορούν να υποβάλουν τα κορυφαία υποψήφια γονίδια και να εντοπίσουν αλληλεπικαλυπτόμενα σύνολα από διαφορετικές λειτουργικά εμπλουτισμένες κατηγορίες όπως μοριακά μονοπάτια - οντολογίες - ασθένειες. Η εργαλειοθήκη συνοδεύεται επίσης από ένα εξαιρετικά εξελιγμένο πλαίσιο γονιδιωματικών διοχετεύσεων σχεδιασμένο για την απλή, γρήγορη και ακριβή ανάλυση των πειραμάτων γονιδιωματικής υψηλής απόδοσης, εστιάζοντας κυρίως στα DNA-seq, RNA-seq, ChIP-seq, CLIP-seq και DNAase-seq.



**Εικόνα 5.** Η διαδικτυακή πλατφόρμα Network Analyzer

Τα **PCASNP** είναι μια ισχυρή και ευρέως δημοφιλής μέθοδος για την ταυτοποίηση δεικτών ενδεικτικών για την γενετική καταγωγή για διάφορους πληθυσμούς. Μπορεί να χρησιμοποιηθεί για την ταξινόμηση δειγμάτων σύμφωνα με την γεωγραφική περιοχή προέλευσής τους με εφαρμογές στην ιατροδικαστική αλλά και για να διορθώσει τη γενετική διαστρωμάτωση σε γενετικές μελέτες συσχετισμού και να διερευνήσει τις ιστορικές σχέσεις των ανθρώπινων πληθυσμών.

## Συμπεράσματα

Το VariantRanker, όπως και το NetworkAnalyzer είναι διαθέσιμα από τον ίδιο διαδικτυακό τόπο. Είναι γενικά εύχρηστα εργαλεία, διαθέτουν γραφικό περιβάλλον και μπορούν να χρησιμοποιηθούν έχοντας πρόσβαση από οποιοδήποτε υπολογιστή με διαδίκτυο ανεξαρτήτων επιδόσεων καθώς οι αναλύσεις γίνονται στον δικό τους server. Ακολουθώντας την μέθοδο PCASNP μπορεί να μειωθεί σημαντικά ο χρόνος για μια ανάλυση. Ωστόσο, δεν διατίθεται ακόμα ως μια διαδικτυακή πλατφόρμα όπως τα δυο προηγούμενα εργαλεία. Μελλοντικός στόχος μας είναι η ανάπτυξη νέων εργαλείων στο χώρο της γονιδιωματικής που θα διευκολύνουν τον χρήστη αλλά και παράλληλα θα του δίνουν δυνατότητες που δεν είναι διαθέσιμες στο παρόν.

## Βιβλιογραφικές Αναφορές

- Alexander J, Mantzaris D, Georgitsi M, Drineas P, Paschou P. Variant Ranker: a web-tool to rank genomic data according to functional significance. BMC Bioinformatics. 2017;18(1):341. doi:10.1186/s12859-017-1752-3
- Paschou P, Drineas P, Lewis J, et al. Tracing sub-structure in the European American population with PCA-informative markers. Pritchard JK, ed. *PLoS Genet.* 2008;4(7):e1000114. doi:10.1371/journal.pgen.1000114
- Drineas P, Lewis J, Paschou P. Inferring Geographic Coordinates of Origin for Europeans Using Small Panels of Ancestry Informative Markers. Relethford J, ed. *PLoS One.* 2010;5(8):e11892. doi:10.1371/journal.pone.0011892