

## Open Schools Journal for Open Science

Vol 5, No 1 (2022)

Open Schools Journal for Open Science - Special Issue - Πρότυπο ΓΕΛ Ευαγγελικής Σχολής Σμύρνης



### Γενετική Επιδημιολογία

Αθανασιος Ζευαρίδης

doi: [10.12681/osj.30336](https://doi.org/10.12681/osj.30336)

Copyright © 2022, Αθανασιος Ζευαρίδης



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/).

### To cite this article:

Ζευαρίδης Α. (2022). Γενετική Επιδημιολογία: Μελέτη πληθυσμών και αίτια εκδήλωσης ασθενειών. *Open Schools Journal for Open Science*, 5(1). <https://doi.org/10.12681/osj.30336>

# ΓΕΝΕΤΙΚΗ ΕΠΙΔΗΜΙΟΛΟΓΙΑ

*Α. Ζευγαρίδης, Ομιλος Σύνδεση Εκπαίδευσης και Εργασίας, Πρότυπο Γενικό Λύκειο Ευαγγελικής Σχολής Σμύρνης*

*Π. Δαλέζιος, Ομιλος Σύνδεση Εκπαίδευσης και Εργασίας, Πρότυπο Γενικό Λύκειο Ευαγγελικής Σχολής Σμύρνης*

## Περίληψη

Η παρούσα εργασία υλοποιήθηκε στα πλαίσια του ομίλου «Σύνδεση Εκπαίδευσης και Εργασίας» του Πρότυπου Γενικού Λυκείου Ευαγγελικής Σχολής Σμύρνης. Το κίνητρο για την μελέτη της γενικής επιδημιολογίας είναι το ενδιαφέρον μου να ασχοληθώ με την Βιολογία στην μελλοντική επαγγελματική μου σταδιοδρομία. Μέσω μιας σειράς σύντομων παραδειγμάτων και αναλύσεων, θα επιχειρήσουμε, λοιπόν, να καταστήσουμε απτές τις θεωρητικές βάσεις του αντικειμένου της Γενετικής Επιδημιολογίας και να εξοικειώσουμε τα άτομα της ηλικίας μας με αυτήν την ενδιαφέρουσα επιστήμη. Όλα αυτά με τη χρήση θεμελιωδών μαθηματικών γνώσεων περί στατιστικής, αλλά και ορισμένων στοιχείων σχετικών με την γενετική βιολογία και τις ιδιότητες του ανθρώπινου γενετικού υλικού.

## Εισαγωγή

Η μελέτη του ανθρώπινου γονιδιώματος, προκειμένου να γίνει αντιληπτή η σύνδεση διαφόρων ασθενειών με τον γονότυπο και ακολούθως να βρεθούν μέσα ανίχνευσης και δυναμικής θεραπείας τέτοιων νοσημάτων είναι, αδιαμφισβήτητα, μια από τις σημαντικότερες εφαρμογές της βιολογίας στις ιατρικές επιστήμες και πρόκειται για έναν πολλά υποσχόμενο κλάδο. Η κατανόηση, εντούτοις, αυτού του θέματος, έστω και σε ουσιώδη μόνο βαθμό, προαπαιτεί την επαφή με ορισμένα στοιχεία και βάσεις του αντικειμένου ώστε το άτομο να κατανοήσει επακριβώς τους στόχους και κυρίως τη μεθοδολογία αυτής της πρακτικής, η οποία έτσι κι αλλιώς χαρακτηρίζεται από μια πολυπλοκότητα αλλά και από την πάγια ανάγκη για μεγάλη προσοχή όσον αφορά φαινομενικές “λεπτομέρειες”, οι οποίες είναι παρόλα αυτά κρίσιμες για την εξαγωγή ορθών συμπερασμάτων.

**Λέξεις-Κλειδιά:** *Γονότυπος, Πιθανότητα, Πληθυσμός*

## Δραστηριότητα 1.1

Συγκρίνοντας τους πίνακες μπορούμε να καταλάβουμε ότι όσοι έχουν συγκεκριμένο γονότυπο (για παράδειγμα CC) φαίνεται να είναι πιο επιρρεπείς να έχουν τη νόσο αντί για άτομα με γονότυπο GG, τα οποία φαίνεται να είναι ασφαλή. Στην πραγματικότητα, κρίνοντας αποκλειστικά από τα παραδείγματα που δίνονται, μπορούμε να συμπεράνουμε ότι τα άτομα με γονότυπο CC είναι εγγυημένα ότι θα έχουν τη νόσο, ενώ τα άτομα με GG δεν είναι πιθανό να την αναπτύξουν. Αυτό σημαίνει ότι υπάρχει απόλυτη σχέση μεταξύ των συγκεκριμένων γονότυπων και της ανάπτυξης ή μη της νόσου.

Σε αυτό το παράδειγμα η κατάσταση φαίνεται να είναι πιο περίπλοκη. Αυτή τη φορά, τα άτομα με CC είναι πιο επιρρεπή να αναπτύξουν τη νόσο, ενώ τα άτομα με AA φαίνεται να έχουν μεγαλύτερες πιθανότητες να μην την αναπτύξουν. Ωστόσο, η συσχέτιση δεν είναι απόλυτη. Τα άτομα με CC μπορεί ακόμα να μην πάσχουν από τη νόσο, αφού τα 3/10 των ατόμων που έχουν CC δεν την έχουν. Ομοίως, τα άτομα με AA μπορούν επίσης να αναπτύξουν τη νόσο, αλλά έχουν λιγότερες πιθανότητες να το κάνουν καθώς τα 2/20 με AA έχουν τη νόσο.

## Δραστηριότητα 1.2

Από τη σύγκριση των πινάκων καταλαβαίνουμε ότι η εκδήλωση της ασθένειας δεν οφείλεται μονάχα στον γονότυπο των ατόμων. Παρόλο που οι πιθανότητες για εμφάνιση της ασθένειας διαφέρουν και εδώ ανάλογα με τον γονότυπο ( οι περισσότεροι καπνιστές με CC έχουν την ασθένεια, κανείς καπνιστής με AA δεν την έχει), ο εξωτερικός παράγοντας του καπνίσματος φαίνεται να επηρεάζει τις πιθανότητες που έχει κάποιος να αναπτύξει την ασθένεια και ειδικότερα κανείς μη καπνιστής δεν εμφάνισε την ασθένεια, επομένως το κάπνισμα παρουσιάζεται ως το αποκλειστικό αίτιο της ασθένειας.

Πράγματι πιστεύω ότι τόσο ο περιβαλλοντικός παράγοντας του καπνίσματος όσο και ο γενετικός παίζουν ρόλο στην εμφάνιση της συγκεκριμένης ασθένειας. Το κάπνισμα είναι ο λόγος που εμφανίζεται η ασθένεια, ενώ ο γονότυπος καθορίζει τις πιθανότητες ασθένειας στους καπνιστές, αφού κάποιος με CC είναι πιθανό να ασθενήσει εφόσον καπνίζει, ενώ κάποιος με AA (κρίνοντας μόνο από τα δείγματα που εξετάσαμε) δεν έχει πιθανότητες να εμφανίσει ασθένεια.

## Δραστηριότητα SNPs (SNP: Single-Nucleotide Polymorphism)

Η γνωστή ασθένεια Alzheimer οφείλεται, τόσο όταν εμφανίζεται σε νέα αλλά και σε γηραιότερα άτομα (>65) επηρεάζεται από περισσότερα από 150 εκατομμύρια γονίδια. Παράδειγμα γονιδίου

με μεγάλη σημασία για την εμφάνιση της ασθένειας σε γηραιότερα άτομα είναι το PLD3 rs7412 που διπλασιάζει τις πιθανότητες εμφάνισης της ασθένειας.

Διάφορα αντικαταθλιπτικά φάρμακα όπως το duloxetine ή το desipramine επιδρούν διαφορετικά σε άτομα με διαφορές στον γονότυπό τους. Παράδειγμα είναι η διαφορά στον γονότυπο ABCB1 και συγκεκριμένα στο rs2032583. [1]

### Δραστηριότητα Promethease Report

Περιγράφηκε ένα είδος τεχνολογίας που αντιμετωπίζει κάθε άνθρωπο ξεχωριστά, μελετώντας τον γονότυπο του και με βάση τα δεδομένα του καθενός και τις πληροφορίες για την πιθανότητα εμφάνισης ασθενειών ή ανταπόκρισης σε διάφορους τύπους θεραπείας, αποφασίζει για τη μέθοδο που θα ακολουθηθεί ώστε να θεραπευθεί-προληφθεί κάθε ασθένεια στα άτομα. Θεωρώ ότι αυτή η τεχνολογία είναι πολλά υποσχόμενη, αφού εγγυάται την εξατομικευμένη και καλύτερη δυνατή μεταχείριση κάθε ασθενούς και τη μεγιστοποίηση των οφελών μιας θεραπείας που σε άλλη περίπτωση ενδεχομένως να οδηγούσε σε σπατάλη χρόνου, χρημάτων και προσπάθειας.

Η παραπάνω τεχνολογία προϋποθέτει τη μελέτη του DNA των ατόμων και συνεπώς την εξαγωγή δεδομένων για το πόσο ευάλωτος είναι κάποιος σε ορισμένες ασθένειες αλλά και την ανταπόκριση του σε κάποια φάρμακα- μεθόδους θεραπείας. Αυτά τα δεδομένα, εφόσον διατεθούν σε άλλους ανθρώπους, μπορεί να αποδειχθούν επικίνδυνα. Αρχικά, μπορεί να βληθεί το προσωπικό απόρρητο του κάθε ανθρώπου μέσω της κοινοποίησης της κατάστασης του οργανισμού του σε ορισμένα πράγματα. Αυτά τα δεδομένα θα μπορούσαν, σε άλλη περίπτωση, να χρησιμοποιηθούν από εταιρείες ή οργανισμούς με αρνητικές συνέπειες για το άτομο. Επίσης το άτομο μπορεί να αγχωθεί αφού μάθει πράγματα και κινδύνους για τον εαυτό του.

Προσωπικά θα έκανα την αναφορά που περιγράφεται, αφού θα με βοηθούσε να πληροφορηθώ για τη πιθανότητα να εμφανίσω ασθένειες και συνεπώς να προφυλαχθώ καλύτερα από αυτές, με όποιο μέσο μπορώ, αλλά και να μάθω για τον καλύτερο τρόπο θεραπείας σε περίπτωση που εν τέλει αποκτήσω κάποια ασθένεια. Άλλοι άνθρωποι ενδεχομένως να μην επιθυμούσαν τέτοια προσέγγιση αφού θα έβρισκαν δυσάρεστη την γνωστοποίηση των ασθενειών που πιθανώς πρόκειται να τους απειλήσουν αλλά και τη γνωστοποίηση των δεδομένων που παρουσιάστηκαν σε άλλα άτομα που θα έχουν τη δυνατότητα να τα εκμεταλλευτούν για προσωπικό όφελος.[2]

## Δραστηριότητα 2.1

Από τους παραπάνω πίνακες, μπορούμε να συμπεράνουμε την σχέση μεταξύ των γονοτύπων ενός δείγματος σε έναν συγκεκριμένο πληθυσμό με την εμφάνιση μίας ασθένειας, αλλά και την πιθανή σχέση της καταγωγής ενός ατόμου με την ασθένεια αυτή καθ' αυτή. Τα αποτελέσματα μάς υποδηλώνουν την ύπαρξη αντιστοιχίας μεταξύ του γονιδιώματος, και συγκεκριμένα του αλληλόμορφου που σχετίζεται με την ασθένεια, με την πιθανότητα εμφάνισης της ασθένειας, αλλά και την ύπαρξη σχέσης μεταξύ της εθνικότητας και της γενικότερης καταγωγής του ατόμου με την γενετική προδιάθεση εμφάνισης της ασθένειας. Βλέπουμε πως σχηματίζεται μια σχέση αιτίας-αιτιατού με την καταγωγή και την ασθένεια, κάτι που οφείλεται όμως στο συγκεκριμένο δείγμα μας και δεν πρέπει να θεωρείται ασφαλές και αξιόπιστο συμπέρασμα. Στην πραγματικότητα, η εμφάνιση μιας ασθένειας μπορεί να οφείλεται ταυτόχρονα στον γονότυπο αλλά και την καταγωγή του ατόμου ή και να οφείλεται σε κάτι τελείως διαφορετικό όπως οι περιβαλλοντικοί παράγοντες κ.α.

## Δραστηριότητα 2.2

Αν υποθέταμε ότι ο συγκεκριμένος γονότυπος που ερευνήσαμε (TT ή GG) έχει επίδραση στην εμφάνιση της ασθένειας στα άτομα, η γνώση της αποκλειστικής εξάρτησής τους από την εθνικότητα των ατόμων θα άλλαζε τα συμπεράσματά μας για την κατάσταση. Θα καταλαβαίναμε, αρχικά, ότι το παραπάνω αλληλόμορφο δεν οδηγεί στην εκδήλωση της ασθένειας, συνεπώς δεν πρέπει να ανησυχεί τα άτομα, ούτε έχει κάποια σημασία κατά τη διερεύνηση του γονότυπου του ατόμου με στόχο τον προσδιορισμό αιτιών της ασθένειας. Το παραπάνω παράδειγμα μας διδάσκει πως η διερεύνηση του γονότυπου κάποιου είναι δύσκολη και απαιτητική, αφού πρέπει να εξακριβώνουμε ποια στοιχεία αφορούν την ασθένεια που εξετάζουμε κάθε φορά και ποια δεν έχουν κάποια σχέση με αυτήν. Στην περίπτωση μας, η λήψη δειγμάτων από διαφορετικές πληθυσμιακές ομάδες ήταν αυτή που μας γνωστοποίησε τον ρόλο του SNP στην υπόθεση.

## Εξάσκηση 1

Υπάρχουν διάφορα γνωρίσματα βάσει των οποίων θα μπορούσαμε να χωρίσουμε την τάξη στην ανάλυσή μας, εκτός από την εθνικότητά τους. Μερικά από αυτά είναι το φύλο, η σωματική διάπλαση (ύψος, βάρος, ποσοστό λίπους), συνήθειες όπως το πρόγραμμα ύπνου, η διατροφή και η χρήση της τεχνολογίας.

## Εξάσκηση 2

Τα αποτελέσματα αυτής της κατηγοριοποίησης, βεβαίως, δεν πρέπει να ληφθούν ως απόλυτα, αφού ως προς ορισμένες παραμέτρους δεν υπάρχει τρόπος να εξακριβώσουμε την ορθότητα της κατηγοριοποίησης μας. Αυτό ισχύει περισσότερο σε στοιχεία που δεν μπορούν να επαληθευθούν από εμάς, κυρίως λόγω της προσωπικής τους φύσης. Οι απαντήσεις που μας έδωσαν οι μαθητές για το πρόγραμμα ύπνου τους ενδέχεται να μην ανταποκρίνεται στην πραγματικότητα και να μας οδηγήσει σε λανθασμένα δεδομένα, τόσο εξαιτίας της αβεβαιότητας των μαθητών για αυτά τα στοιχεία όσο και εξαιτίας της επιθυμίας τους να μην κοινοποιήσουν προσωπικά στοιχεία. Ακόμη και αν εξαλείψουμε τους παραπάνω παράγοντες σφάλματος, βέβαια, είναι ακόμη δύσκολο να κατηγοριοποιήσουμε αποτελεσματικά την ομάδα που μελετάμε. Είναι σημαντικό να χωρίσουμε σωστά τις κατηγορίες ανάλογα με κριτήρια που θα οδηγήσουν σε εύστοχα συμπεράσματα (αν χωρίσουμε το δείγμα ανάλογα με το ποσοστό λίπους και ορίσουμε λίγες κατηγορίες, μπορεί να μην φτάσουμε σε σωστά συμπεράσματα επειδή δεν καταφέραμε να απομονώσουμε σωστά τις περιοχές όπου φαίνεται η επίδραση κάποιου SNP).

## Εξάσκηση 3

Μια κατηγορία στην οποία θα μπορούσαμε να χωρίσουμε τους μαθητές με βεβαιότητα είναι το φύλο τους, αφού δεν υπάρχει περιθώριο λήψης λανθασμένων στοιχείων. Μια κατηγορία που δε θα οδηγούσε σε απόλυτα δεδομένα, από την άλλη, θα ήταν η ικανότητα των μαθητών σε κάποια φυσική δραστηριότητα, όπως ο αριθμός συνεχόμενων έλξεων, εφόσον οι απαντήσεις του δείγματος μπορεί να είναι ανακριβείς λόγω των παραγόντων που αναφέρθηκαν στην προηγούμενη απάντηση.

## Εξάσκηση 4

Εάν λαμβάναμε τα δεδομένα της ερώτησης, θα υποθέταμε ότι τα SNP's και το χρώμα μαλλιού στους μαθητές δεν έχουν σχέση εξάρτησης μεταξύ τους, αφού το CC και το AA εντοπίζεται και στις δύο κατηγορίες και ο αριθμός τους είναι ανάλογος με το μέρος του πληθυσμού κάθε φορά.

## Εξάσκηση 5

Στην περίπτωση που η σχέση των αλληλομόρφων και του χρώματος μαλλιών εξεταζόταν σύμφωνα με την κατηγοριοποίηση που δίνεται στο ερώτημα, θα εξαγάγαμε διαφορετικά αποτελέσματα. Φαίνεται, τώρα, ότι ανάλογα με τον πληθυσμό το αλληλόμορφο επηρεάζει το

χρώμα των μαλλιών και μάλιστα διαφορετικά στις δύο ομάδες, αφού στην πρώτη το CC φαίνεται να οδηγεί σε μεγάλη πιθανότητα για κόκκινα μαλλιά και το AA σε μικρή τέτοια πιθανότητα, ενώ στην δεύτερη ισχύει το αντίθετο. Καταλαβαίνουμε, λοιπόν, ότι είναι σημαντική η κατηγοριοποίηση του δείγματος σχετικά με πολλές παραμέτρους ώστε να εντοπίσουμε σωστά τη σχέση που αναζητούμε.

### Ενότητα 3

Αν τα δεδομένα μιας έρευνάς μας, περιέχουν ατελείς δείκτες περιεκτικότητας αλληλόμορφων στον πληθυσμό που μελετάμε, μπορούμε να χρησιμοποιήσουμε αυτούς τους δείκτες για να κατατάξουμε τον πληθυσμό με βάση το κριτήριο της καταγωγής. Αυτό συμβαίνει κατά τον ίδιο τρόπο όπως αν είχαμε τέλειους δείκτες. Έτσι χαρακτηρίζουμε την παρουσία ενός αλληλόμορφου, ως χαρακτηριστικό μιας εθνικής ομάδας μέσα στο συνολικό εν μελέτη πληθυσμό. Όμως σε αυτή την περίπτωση πρέπει να κατατάξουμε, βάσει των πιθανοτήτων που μας δίνονται, τον πληθυσμό σε εθνικές ομάδες, θέτοντας ένα όριο πιθανότητας παρουσίας ενός αλληλόμορφου (π.χ. 50%) που θα χαρακτηρίζει μία εθνική ομάδα.

Το θέμα που προκύπτει σχετίζεται με την πιθανότητα να κατατάξουμε λανθασμένα ένα άτομο σε μία εθνική ομάδα επειδή παρουσιάζει τα στοιχεία αυτής της εθνικής ομάδας, ενώ έχει διαφορετική εθνικότητα. Για παράδειγμα, αν κατατάξουμε έναν άνθρωπο ως Ευρωπαίο επειδή περιέχει το T αλληλόμορφο, ενδέχεται να ανήκει σε διαφορετική εθνικότητα (π.χ. Αφρικανός) και να έχει κατ' εξαίρεση το αλληλόμορφο T λόγω μετάλλαξης σε αντίθεση με την πλειονότητα της εθνικής του ομάδας.

Στην περίπτωση που μπορούσαμε να χρησιμοποιήσουμε μόνο ένα πίνακα για να καθορίσουμε την καταγωγή ενός ατόμου, θα ήταν επιστημονικά ορθό να χρησιμοποιήσουμε τον δεύτερο πίνακα. Αυτό οφείλεται στο γεγονός ότι μειώνονται οι πιθανότητες να ανήκει κάποιος σε μία εθνική ομάδα και να παρουσιάζει στοιχεία διαφορετικής εθνικότητας, ελαχιστοποιώντας στον μεγαλύτερο δυνατό βαθμό τις πιθανότητες λάθους.

#### Δραστηριότητα 3.1

Σε περίπτωση που κάποιος άνθρωπος δεν αντιστοιχίζεται σε καμία πληθυσμιακή ομάδα, αυτό θα οφειλόταν σε δύο πιθανούς παράγοντες. Ο πρώτος είναι αυτό το άτομο να ανήκει σε πληθυσμιακή ομάδα για την οποία δεν έχουμε δεδομένα και συνεπώς δεν μπορούμε να κατατάξουμε άτομα σε αυτήν. Ο δεύτερος πιθανός παράγοντας θα ήταν τα λανθασμένα δεδομένα που θα εξήγαμε για το άτομο. Η κατηγοριοποίησή του με βάση την παρουσία ή όχι

ενός αλληλόμορφου δε θα αντιστοιχούσε σε κάποια ομάδα, δηλαδή, διότι σε κάποιο από αυτά ή και σε περισσότερα δε θα άνηκε στην πλειοψηφία μιας ομάδας.

Η παραπάνω περίπτωση καταδεικνύει την ανασφάλεια που προκύπτει από την παραπάνω μέθοδο κατηγοριοποίησης ατόμων. Καταλαβαίνουμε ότι είναι πολύ εύκολο να συμβεί κάποιο λάθος σε μεμονωμένες περιπτώσεις ατόμων. Ωστόσο, αν εξετάζουμε μεγάλα μέρη του πληθυσμού και δεν ενδιαφερόμαστε για μεγάλη ακρίβεια στα αποτελέσματά μας, η μέθοδος φαίνεται να είναι κατάλληλη.

Στην δεύτερη περίπτωση, κατανοούμε ότι ορισμένοι δείκτες είναι περισσότερο χρήσιμοι από άλλους στην εξαγωγή συμπερασμάτων για την κατηγοριοποίηση κάποιου ατόμου. Συγκεκριμένα, όσοι δείκτες παρουσιάζουν ακραία (σχετικά με τις άλλες περιπτώσεις) ποσοστά για την παρουσία αλληλόμορφου στον πληθυσμό είναι περισσότερο πιθανό να οδηγήσουν σε σωστές κατηγοριοποιήσεις ατόμων σε ομάδες από ότι άλλοι δείκτες των οποίων οι τιμές κυμαίνονται πιο κεντρικά. Πέρα από αυτό, βέβαια, ισχύει επίσης ότι για συγκεκριμένα άτομα ορισμένοι δείκτες βοηθούν σημαντικά περισσότερο από άλλους στο να εξάγουμε ασφαλή αποτελέσματα. Αυτό συμβαίνει για περιπτώσεις όπου κάποιος δείκτης περιέχει την τιμή του ατόμου σε μια μονάχα περίπτωση, συνεπώς δε μας δημιουργούν αμφιβολίες για την κατάταξή του σε σχέση με ένα συγκεκριμένο αλληλόμορφο. Το αποτέλεσμα μας είναι ακόμα πιο ασφαλές, βεβαίως, στην περίπτωση που οι τιμές είναι ακραίες.

Είναι δυνατόν, πράγματι, δύο άτομα με διαφορετικό γονότυπο να εμφανίζουν ίδια αλληλουχία 0 και 1, δηλαδή ίδιες τιμές στους δείκτες που αναλύουμε. Αυτό θα συμβεί εάν έχουν αμφότεροι τον ίδιο δείκτη, αλλά ο ένας από τους δύο έχει στον γονότυπό του δύο διαφορετικές αζωτούχες βάσεις. Πχ εάν αναλύουμε άτομα σε σχέση με την παρουσία ή όχι A σε μια συγκεκριμένη θέση, δύο άτομα με AA και AT αντίστοιχα θα εμφανίσουν ίδια τιμή στον δείκτη A για αυτό το αλληλόμορφο παρόλο που έχουν διαφορετικό γονότυπο.

Είναι λογικό πως όσες περισσότερες φορές ένα άτομο αντιστοιχίζεται σε έναν πληθυσμό με βάση τους δείκτες που αναλύουμε, οι πιθανότητες να ανήκει πράγματι σε αυτόν αυξάνονται. Αυτό συμβαίνει επειδή, όσες περισσότερες φορές γίνεται η αντιστοίχιση σε μια ομάδα, μειώνεται η πιθανότητα λάθους λόγω απόκλισης του γονότυπου του ατόμου από την πλειοψηφία. Εξυπακούεται, βεβαίως, ότι οι δείκτες με ακραίες τιμές έχουν μεγαλύτερη "αξία" ως προς την βεβαιότητα που μας προσδίδουν σχετικά με το ότι αντιστοιχίσαμε σωστά ένα άτομο.



Ως προς τους δείκτες, είναι προφανές ότι όσο περισσότερους εξετάζουμε τόσο πιο βέβαιοι θα είμαστε για τη σωστή κατηγοριοποίηση ενός ατόμου, ενώ το ίδιο θα ισχύει και ενώ αυξάνουμε την ποσότητα και την οξύτητα δεικτών με ακραίες τιμές.

Ως προς τη μετάφραση των δεδομένων σε 1 και 0, είναι αναμενόμενο ότι αν διατηρήσουμε τις αληθείς τιμές του κάθε δείκτη, αυτό θα μας βοηθήσει στην εκτίμηση της βεβαιότητας που έχουμε για κάποια αντιστοιχία, ενώ επίσης θα μας οδηγήσει στην επιλογή της πιο ασφαλούς επιλογής αν σε μια περίπτωση διχαστούμε σχετικά με την σωστή ομάδα για ένα άτομο.[3]

### Δραστηριότητα 5.1

Από την ανάλυση του πίνακα καταλαβαίνουμε ότι, στην περίπτωση που έχουμε τρεις πληθυσμιακές ομάδες, οι ελάχιστοι δείκτες αζωτούχων βάσεων που πρέπει να έχουμε ώστε να τις ξεχωρίσουμε είναι 2, αφού αυτή είναι η περίπτωση όπου θα υπάρχουν τρεις ξεχωριστές τιμές για κάθε πληθυσμό, πχ 11, 00, 01. Στην προκειμένη περίπτωση, μπορούμε να αφαιρέσουμε πχ. την μέση τιμή του T, αφού και έτσι κάθε πληθυσμιακή ομάδα θα έχει διαφορετική “ταυτότητα” από τις άλλες δύο (πληθυσμός 1: 00, πληθυσμός 2:10 , πληθυσμός 3:01)

Γενικά, στο παρόν παράδειγμα οι επιπρόσθετοι δείκτες για τις βάσεις δεν μας βοήθησαν εφόσον είχαμε ήδη τοποθετήσει τα άτομα σε ομάδες με βάση τα αρχικά δεδομένα, παρά στην περίπτωση του ατόμου του οποίου τα δεδομένα αρχικά δεν οδήγησαν στην τοποθέτησή του σε κάποια βάση. Επιπρόσθετα, οι επιπλέον δείκτες αυξάνουν τη σιγουριά που έχουμε για την σωστή τοποθέτηση κάποιου σε μια ομάδα, εφόσον όσο περισσότερο αντιστοιχίζεται σε κάποια με βάση πλήθος δεικτών, τόσο αποκτούμε ασφάλεια ότι δεν πρόκειται για σφάλμα. Κατανοούμε ξανά ότι αυτή η μέθοδος προσδιορισμού της εθνικότητας κάποιου δεν στηρίζεται στην ακρίβεια της αντιστοιχίας στα υπάρχοντα δεδομένα, ιδιαίτερα στις περιπτώσεις που δεν έχουμε μεγάλο πλήθος δεικτών που μας προσφέρουν μεγάλη βεβαιότητα. Αντιθέτως, πρόκειται για μια μέθοδο που στηρίζεται στη “συνολική εικόνα” των δεδομένων του κάθε ατόμου, αφού αποκλίσεις από τον “πρότυπο γενότυπο” πάντα υπάρχουν και το νόημα έγκειται στο να παίρνουμε αποφάσεις με τη βοήθεια επαρκούς αριθμού βάσεων.

### Δραστηριότητα Hamming Distance

Εάν η “απόσταση” δύο αλυσίδων DNA είναι 0, αυτό σημαίνει ότι δεν διαφέρουν πουθενά στο περιεχόμενο των αζωτούχων βάσεων. Για δύο αλυσίδες με σημαντικό αριθμό βάσεων, δηλαδή μεγάλο μήκος, η χρήση υπολογιστή θα διευκόλυne την διαπίστωση της απόστασης Hamming μεταξύ τους. Ειδικά, ενδεχομένως να λαμβάναμε υπόψιν μόνο ένα μέρος από τις δύο

αλυσίδες και να γενικεύαμε τα συμπεράσματά μας και για το υπόλοιπο, μέθοδος που ενδεχομένως επιφυλάσσει σοβαρές ανακρίβειες.

## Δραστηριότητα 5.2

Από τα δεδομένα του πίνακα ενδεχομένως θα συμπεραίναμε ότι η ύπαρξη της A βάσης στο σημείο του γονότυπου που εξετάζουμε αυξάνει σημαντικά τις πιθανότητες για εμφάνιση ελκώδους κολίτιδας, αφού 20 από τους 70 με CC έχουν την ασθένεια ενώ 80 από τους 130 με AC ή AA έχουν την ασθένεια.

POP 1	Positive	Negative
CC	15	3
AA or AC	65	17

POP 2	Positive	Negative
CC	5	42
AA or AC	15	33

Τελικά συμπεραίνουμε ότι η πληθυσμιακή ομάδα είναι αυτή που σαφώς καθορίζει σε σημαντικότερο βαθμό την ύπαρξη ή όχι της ασθένειας, αφού στην πρώτη ομάδα η μεγάλη πλειοψηφία νοσεί ενώ στη δεύτερη μόνο μια μειοψηφία νοσεί. Βεβαίως, η ύπαρξη ή όχι A στο εξεταζόμενο σημείο συνεχίζει να επηρεάζει τις πιθανότητες να αναπτύξει κάποιος την ασθένεια, αλλά σε πολύ λιγότερο αποφασιστικό βαθμό από ότι η εθνικότητα, αφού απλά μεταβάλλει για λίγο τη πιθανότητα εμφάνισης της ασθένειας. Βεβαίως, αυτή η μικρή ανισορροπία μεταξύ A και μη A μπορεί και να οφείλεται αποκλειστικά ή και σε απλώς μεγάλο βαθμό στο δείγμα μας, που δεν χαρακτηρίζεται από αριθμητική ισχύ που να καθιστά τα συμπεράσματά μας αδιαμφισβήτητα. Καταλαβαίνουμε συνεπώς ότι από κάθε άποψη το προηγούμενο συμπέρασμά μας στην 1) ήταν εσφαλμένο, γεγονός που αποδεικνύει τη βαρύτητα της εξέτασης της εθνικότητας των ατόμων στη διάρκεια μιας έρευνας.

Είδαμε ότι η 6η ομάδα δεν αντιστοιχίζεται σε κάποια από τις δύο ομάδες για τις οποίες έχουμε τον "πρότυπο γονότυπο", αφού η απόσταση Hamming μεταξύ αυτής της ομάδας και των

προτύπων είναι μεγαλύτερη από 2 και στις δύο περιπτώσεις ( πρέπει αυτή η απόσταση να είναι το πολύ 1 ώστε να θεωρούμε ότι κάποια ομάδα ανήκει στον πληθυσμό 1 ή 2). Αυτή η αναντιστοιχία μπορεί να οφείλεται είτε στο γεγονός ότι η 6η ομάδα απλώς έτυχε να έχει μεγαλύτερη απόσταση Hamming από την “πρότυπη” αλυσίδα του πληθυσμού που ανήκει και λανθασμένα θεωρήσαμε ότι δεν ανήκει στην 1η ή στην 2η πληθυσμιακή ομάδα. Η άλλη περίπτωση είναι να ανήκει σε κάποια εντελώς διαφορετική πληθυσμιακή ομάδα που δε λάβαμε υπόψη στην έρευνα. Και στις δύο περιπτώσεις, συνειδητοποιούμε τη σημασία της σωστής προετοιμασίας των δεδομένων μας πριν προχωρήσουμε στην ανάλυση κάποιου δείγματος.

## Εξάσκηση 2

Στην ανάπτυξη προγράμματος για την διαπίστωση της απόστασης Hamming μεταξύ δύο αλυσίδων θα μπορούσαμε ενδεχομένως να ζητάμε από τον χρήστη να καταχωρήσει τους δείκτες κάθε αλυσίδας και στη συνέχεια να αφαιρεί τις αντίστοιχες τιμές, δηλαδή την  $\alpha_1$  με την  $\alpha_2$ , την  $\alpha_2$  και τη  $\beta_2$  κτλ.

## Δραστηριότητα 6

Παρατηρούμε ότι και στους τρεις δείκτες, ισχύει ότι τα άτομα στα οποία η εθνικότητα δεν είναι αμιγώς ευρωπαϊκή ή αφρικανική εμφανίζουν τιμές που βρίσκονται πιο κοντά στις πρότυπες τιμές στην “κύρια” εθνικότητά τους, π.χ. οι Αφροαμερικανοί του Detroit που είναι κυρίως Αφρικανοί στον τρίτο δείκτη 0.133, τιμή που πλησιάζει την αφρικανική αντίστοιχη τιμή του τρίτου δείκτη που είναι 0 και λιγότερο την ευρωπαϊκή που είναι 1 (παρόλα αυτά φαίνεται στον γονότυπό τους η αναμειγμένη εθνικότητά τους).

Παρατηρώντας τα δεδομένα των δύο εικόνων υποθέτουμε ότι πρόκειται για την περιγραφή του γονότυπού τους με βάση τον βαθμό στον οποίο τα SNP's, σχετικά με την εθνικότητα, τους ταυτίζονται με αυτά του “πρότυπου” γονότυπου των ευρωπαίων ή των Αφρικανών ή των αυτοχθόνων Αμερικανών.

## Τελικό συμπέρασμα

Ύστερα από την μελέτη και επεξεργασία του φύλλου εργασίας που δόθηκε, καταλήξαμε στα εξής συμπεράσματα. Πρωτίστως, καθίσταται σαφής η άμεση σχέση του ανθρώπινου γονιδιώματος με την εμφάνιση διαφόρων ασθενειών (π.χ. Alzheimer), το οποίο και παρατηρείται σε ποικίλα παραδείγματα μελέτης πληθυσμού, τα οποία αναγράφονταν στο φύλλο εργασίας. Παράλληλα, καθοριστικό παράγοντα για την συχνότητα εμφάνισης μιας

ασθένειας στον υπό μελέτη πληθυσμό, πρέπει να αποτελούν οι περιβαλλοντικές και εν γένει εξωτερικές συνθήκες διαβίωσης του κάθε ατόμου. Οι περιβαλλοντικοί παράγοντες στο σύνολό τους διαδραματίζουν σημαντικό ρόλο και συσχετίζονται με την εμφάνιση ασθενειών και διαφόρων παθολογιών, με χαρακτηριστικότερο παράδειγμα το κάπνισμα που παρατηρήσαμε και σε μελέτη του φύλλου εργασίας. Συνεπώς, καταλήξαμε, με τη βοήθεια του δοσμένου φύλλου εργασίας, στο συμπέρασμα πως η έρευνα για την αναζήτηση των αιτιών που προκαλούν την εμφάνιση διαφόρων ασθενειών, οφείλει να λαμβάνει υπόψιν της τόσο γενετικούς και κληρονομικούς όσο και περιβαλλοντολογικούς παράγοντες, αυξάνοντας έτσι την πολυπλοκότητα, αλλά και τη γοητεία της επιστήμης της γενετικής επιδημιολογίας. Στόχος της παραπάνω εργασίας συνιστούσε, στα πλαίσια του ομίλου “Σύνδεσης Εκπαίδευσης και Εργασίας”, να εργαστούμε και να ερευνήσουμε τις πτυχές του κλάδου της γενετικής επιδημιολογίας, σε μια προσπάθεια να διευρύνουμε τις γνώσεις μας στο αντικείμενο, που θα μας παρέχουν πολύτιμες πληροφορίες για την επιλογή του επαγγέλματος που θα ακολουθήσουμε στο μέλλον, ένα θέμα το οποίο ταλανίζει τους νέους διαχρονικά.

#### Πηγές αναφοράς:

- ✚ [1] United States Department of Energy Office of Biological and Environmental Research. (2001, July). Human Genome Working Guides: First Edition Travel Guides. Human Genome News, 11(3-4), 3-4.
- ✚ [2] SNPedia. Promethease Page. Wiki site
- ✚ [3] Parra, E.J., Marcini, A., Akey, J., et. al. 1998. Estimating African American Admixture Proportions by Use of PopulationSpecific Alleles. American Journal of Human Genetics, 63(6), 1839-1851, 1998.
- ✚ [4] Malkevitch, J., Froelich, G., and Froelich, D. Codes Galore. (HistoMAP Unit #18). Bedford MA: COMAP.

