ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Εύρεση της σχέσης μετάλλαξης::miRNA::mRNA με νευρολογικές και ψυχιατρικές παθήσεις μέσω ανάλυσης άρθρων και χρήσης γνωστών εργαλείων πρόβλεψης στόχων

Ληφούση Άννα

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ
Υπεύθυνος
Χατζηγεωργίου Άρτεμις-Γεωργία
Καθηγήτρια

Λαμία, 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΜΕ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΒΙΟΙΑΤΡΙΚΗ

Εύρεση της σχέσης μετάλλαξης::miRNA::mRNA με νευρολογικές και ψυχιατρικές παθήσεις μέσω ανάλυσης άρθρων και χρήσης γνωστών εργαλείων πρόβλεψης στόχων

Ληφούση Άννα

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Επιβλέπουσα
Χατζηγεωργίου Άρτεμις-Γεωργία
Καθηγήτρια

Λαμία, 2022

Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις [(1)], που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. *Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί* ***χωρίς να τα περικλείω σε εισαγωγικά*** *και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάσθηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.*

2. *Δέχομαι ότι η αυτολεξεί* ***παράθεση χωρίς εισαγωγικά****, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.*

3. *Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια*

4. *Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.*

Ημερομηνία: ……/…../20……

Ο – Η Δηλ.

(Υπογραφή)

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.

Εύρεση της σχέσης μετάλλαξης::miRNA::mRNA με νευρολογικές και ψυχιατρικές παθήσεις μέσω ανάλυσης άρθρων και χρήσης γνωστών εργαλείων πρόβλεψης στόχων

Ληφούση Άννα

Τριμελής Επιτροπή:

Χατζηγεωργίου Άρτεμις-Γεωργία, Καθηγήτρια (επιβλέπουσα)

Παντελεήμων Μπάγκος, Καθηγητής

Μπράλιου Γεωργία, Επίκουρος Καθηγήτρια

Ευχαριστίες

Με την παρούσα πτυχιακή εργασία ολοκληρώθηκε ο κύκλος των προπτυχιακών σπουδών μου στο τμήμα Πληροφορικής με εφαρμογές στην Βιοϊατρική της σχολής Θετικών Επιστημών του Πανεπιστημίου Θεσσαλίας, ένας όμορφος αλλά όχι εύκολος κύκλος μέσα από τον οποίο έμαθα πάντα να αναζητώ και να ερευνώ το επιστημονικά ορθό και εξελήχθηκα τόσο προσωπικά όσο κι ακαδημαϊκά. Θα ήθελα λοιπόν να ευχαριστήσω την επιβλέπουσα καθηγήτρια κυρία Χατζηγεωργίου Γ. Άρτεμις που μου έδωσε την ευκαιρία να ασχοληθώ με έναν τόσο ενδιαφέρον τομέα που παντα επιθυμούσα, να μάθω νέα πράγματα και να ανακαλύψω τα όρια μου. Επιπλέον, θα ήθελα να ευχαριστήσω θερμά την υποψήφια διδάκτορα κύρια Ζαχαροπούλου Ελίζα για την πολύτιμη καθοδήγηση της κατά την διάρκεια της εκπόνησης της εργασίας. Στην συνέχεια, θα ήθελα να ευχαριστήσω από καρδιάς την οικογένεια μου και τους αγαπημένους μου ανθρώπους για την στήριξη τους καθ'όλη την διάρκεια αυτού του ταξιδιού. Είμαι ευγνώμων που τους έχω δίπλα μου να μου υπενθιμίζουν πόσο σημαντικό είναι να κάνω αυτό αγαπάω και να κυνηγάω τα όνειρα μου ακόμα και αν αυτό προυποθέτει να ακολουθήσω τον δύσκολο δρόμο. Μου δίνει αμέτρητο κουράγιο και δύναμη να έχω αυτούς τους ανθρώπους δίπλα μου που χαίρονται με κάθε μου επιτυχία.

Στην συνέχεια, θα ήθελα να ευχαριστήσω από καρδιάς την οικογένεια μου και τους αγαπημένους μου ανθρώπους για την στήριξη τους καθ'όλη την διάρκεια αυτού του ταξιδιού. Είμαι ευγνώμων που τους έχω δίπλα μου να μου υπενθιμίζουν πόσο σημαντικό είναι να κάνω αυτό αγαπάω και να κυνηγάω τα όνειρα μου ακόμα και αν αυτό προυποθέτει να ακολουθήσω τον δύσκολο δρόμο. Μου δίνει αμέτρητο κουράγιο και δύναμη να έχω αυτούς τους ανθρώπους δίπλα μου που χαίρονται με κάθε μου επιτυχία.

Στην γιαγιά μου…

Περίληψη

Τα microRNAs (miRNAs) αποτελούν μικρά μόρια RNA, μήκους περίπου 22 βάσεων τα οποία παίζουν σημαντικό ρόλο στην ρύθμιση της γονιδιακής έκφρασης. Πιο συγκεκριμένα, αυτά στοχεύουν τα mRNAs, με αποτέλεσμα να ρυθμίζουν την έκφραση των γονιδίων. Τυχόν πολυμορφισμοί που επηρεάζουν την αλληλεπίδραση miRNA::mRNA μπορούν να συμβάλλουν στην δημιουργία παθολογιών που ευθύνονται για την ανάπτυξη καρκίνων ή άλλων νοσημάτων, μεταξύ άλλων και νευρολογικές και ψυχιατρικές ασθένειες. Η νόσος Αλτσχάιμερ, η νόσος Πάρκισον και η Σχιζοφρένια αποτελούν πολυπαραγοντικές παθήσεις που η ανάπτυξη τους φαίνεται να επηρεάζεται από τέτοιους πολυμορφισμούς που επεμβαίνουν στην αλληλεπίδραση των miRNA με τα mRNA.

Στην παρούσα εργασία αναπτύχθηκε ένα σύστημα αυτόματης αναζήτησης στο Entrez για ένα δεδομένο ερώτημα (query) και στην συνέχεια μέσω επεξεργασίας κείμενου με την βοήθεια μοντέλων ανάλυσης ονομάτων-οντοτήτων του scispaCy, τα άρθρα που επιστράφηκαν φιλτραρίστηκαν με στόχο την απομόνωση των άρθρων που περιείχαν πληροφορία σχετική με την σχέση μετάλλαξης::miRNA::mRNA με τις νόσους Alzheimer, Parkinson και την Σχιζοφρένεια. Έπειτα, τα άρθρα που πέρασαν τα 2 φίλτρα που αναπτύχθηκαν, σχολιάστηκαν χειροκίνητα και συγκεντρώθηκαν μεταδεδομένα όπως το όνομα του συγγραφέα και το ID του άρθρου και οι πληροφορίες της συσχέτισης όπως το miRNA, η μετάλλαξη και η ασθένεια που αφορά.Τα γονίδια που είχαν μονονουκλεοτιδικούς πολυμορφισμούς στην 3’UTR (3’ Untranslated Region) ή την CDS (Coding Sequence) χρησιμοποιήθηκαν ως είσοδος (με και χωρίς τους πολυμορφισμούς) στον αλγόριθμο πρόβλεψης στόχων microT-CNN προς επιβεβαίωση της βιβλιογραφίας.

Με την αναζήτηση στο Entrez λήφθηκαν 179 άρθρα, εκ των οποίων τα 125 ήταν διαθέσιμα στην Pubmed Central και τα 81 από αυτά είχαν διαθέσιμο κείμενο, που ήταν απαραίτητο για την επεξεργασία ονομάτων οντοτήτων. Τα 53 από αυτά φιλτραρίστηκαν, ενώ μόνο τα 27 κατέληξαν στον τελικό κατάλογο του χειροκίνητου σχολιασμού. Στο γονιδίωμα εισόδου για τον αλγόριθμο microT-CNN κατέληξαν 10 γονίδια για τα οποία εξετάστηκε η σχέση τους με 10 miRNAs. Εν κατακλείδι, οι τελικές εγγραφές της βάσης ανέρχονται στις 5 και αποτελούν τον συνδυασμό της ερευνητικής βιβλιογραφίας με τα αποτελέσματα του in silico αλγορίθμου.

Στόχος της παρούσας εργασίας είναι η αυτοματοποιημένη και αποδοτική εξαγωγή πολύτιμης πληροφορίας από τον πλέον τεράστιο όγκο της βιβλιογραφίας. Τα αποτελέσματα της μεθοδολογίας που ακολουθήθηκε μπορούν να αποτελέσουν ένα σύνολο βάσης που θα αφορά την επιστημονική κοινότητα και θα προσφέρει δεδομένα είτε από την αρθογραφία είτε και από in silico αλγορίθμους πρόβλεψης στόχων των miRNAs.

Λέξεις κλειδιά: miRNA, mRNA, μονονουκλεοτιδικοί πολυμορφισμοί, μεταλλάξεις, Νόσος Alzheimer, νόσος του Parkinson, Σχιζοφρένεια

Abstract

MicroRNAs are small non coding RNAs, almost 22 bases long, which play an important role in the regulation of gene expression. More specifically, microRNAs bind at binding sites on mRNAs and as a result they regulate gene expression. Genetic polymorphisms which interfere with miRNA::mRNA have the ability to contribute at the development of certain types of cancer and diseases, such as neurological and psychiatric diseases. Alzheimer's disease, Parkinson's disease and Schizophrenia are some examples of diseases whose development is connected to such kind of polymorphisms.

In this study, we developed an automated search system in Python programming language. In this system, for a specific query, a search in PubMed was created so as to retrieve publications based at this query. Afterwards, the main text of this publication was analyzed by name-entity recognition analysis with scispaCy library in Python. Next, the main text of each publication was filtered so as to contain the 3 main terms (variant, miRNA, mRNA, disease). The filtered publications were manually curated and the following information were collected: PubMed id, article's title, journal's title, author, year, abstract, gene, gene biotype, miRNA, variant id, variant region, association, risk, disease, population, study, cell line, comments, sentences. For the execution of the MicroT-CNN algorithm, only the genes which contained SNPs on 3'UTR and CDS were selected. Finally, we executed the algorithm two times, one for the reference genome and the other for the genome with SNPs intergraded.

The search on PubMed returned 179 publications and the 125 of them were available at PubMed Central. 53 publications passed the filters and only 27 of them were selected by the manual curation. We examined the interaction of 10 genes with 10 miRNA with MicroT-CNN algorithm. In conclusion, the final catalogue contained 5 records which are combination of the research literature with the results of the in silico algorithm.

The aim of this work is the automated and efficient extraction of valuable information from the vast volume of literature. The results of the methodology followed can form a base set that will concern the scientific community and will provide data either from the literature or from in silico miRNA target prediction algorithms.

Key words: miRNA, mRNA, SNPs, Alzheimer's disease, Parkinson's disease , Schizophrenia

Περιεχόμενα

ΚΕΦΆΛΑΙΟ 1: ΕΙΣΑΓΩΓΉ **14**

1.1.Η ΝΌΣΟΣ ALZHEIMER 14
1.1.1. Τα συμπτώματα της νόσου *14*
1.1.2. Τα αίτια της νόσου *14*
1.1.3. Τα στάδια της νόσου *16*
1.5. Επιδημιολογία *18*
1.2.Η ΝΌΣΟΣ ΤΟΥ PARKINSON 18
1.2.1. Τα συμπτώματα της νόσου *18*
1.2.3. Τα στάδια της νόσου *20*
1.2.4. Παράγοντες κινδύνου *20*
1.2.5. Επιδημιολογία *21*
1.3. ΣΧΙΖΟΦΡΈΝΕΙΑ 21
1.3.1. Τα συμπτώματα της νόσου *21*
1.3.2.Τα στάδια της νόσου *22*
1.3.3.Παράγοντες κινδύνου *22*
1.3.4. Τα αίτια της νόσου *23*
1.3.5. Επιδημιολογία *24*
1.4. MICRORNA 25
1.4.1. Τι είναι τα microRNAs; *25*
1.4.2. Βιογένεση *25*
1.5. MICROT-CNN 28
1.6. ΣΤΌΧΟΣ ΕΡΓΑΣΊΑΣ 29

ΚΕΦΆΛΑΙΟ 2: ΜΈΘΟΔΟΙ **30**

2.1. ΕΡΏΤΗΜΑ (QUERY) ΤΗΣ ΑΝΑΖΉΤΗΣΗΣ 30
2.2. ΑΥΤΌΜΑΤΗ ΑΝΑΖΉΤΗΣΗ ΣΤΟ ENTREZ ΜΈΣΩ ΤΗΣ PYTHON 31
2.3. ΣΥΛΛΟΓΉ ΒΑΣΙΚΏΝ ΠΛΗΡΟΦΟΡΙΏΝ ΤΩΝ ΔΗΜΟΣΙΕΎΣΕΩΝ 32
2.4. ΑΝΆΛΥΣΗ ΑΝΑΓΝΏΡΙΣΗΣ ΟΝΟΜΆΤΩΝ-ΟΝΤΟΤΉΤΩΝ (NAMED-ENTITY RECOGNITION) 33
2.4.1. Τι είναι η Ανάλυση Αναγνώρισης ονομάτων-οντοτήτων *33*
2.4.2. Αναζήτηση στην PMC *33*
2.4.3. Ανάλυση Αναγνώρισης ονομάτων-οντοτήτων με την χρήση του πακέτου scispaCytης Python *34*
2.5. ΦΙΛΤΡΆΡΙΣΜΑ ΑΠΟΤΕΛΕΣΜΆΤΩΝ 36
2.6.ΧΕΙΡΟΚΊΝΗΤΟΣ ΣΧΟΛΙΑΣΜΌΣ ΤΩΝ ΑΠΟΤΕΛΕΣΜΆΤΩΝ (MANUAL CURATION) 37
2.7. ΥΠΟΛΟΓΙΣΤΙΚΉ ΈΥΡΕΣΗ ΣΤΌΧΩΝ ΜΕ ΤΗΝ ΧΡΉΣΗ ΤΟΥ ΑΛΓΟΡΊΘΜΟΥ MICROT-CNN 40
2.7.1. Χρήση των πληροφορίων του χειροκίνητου σχολιασμού ως είσοδο του αλγορίθμου *40*
2.7.2. Εκτέλεση του αλγόριθμου για το γονιδίωμα με ενσωματωμένους πολυμορφισμούς *40*

ΚΕΦΆΛΑΙΟ 3: ΑΠΟΤΕΛΈΣΜΑΤΑ **42**

3.1.ΑΠΟΤΕΛΈΣΜΑΤΑ ΑΝΑΖΉΤΗΣΗΣ ΣΤΗΝ ENTREZ ΚΑΙ ΦΙΛΤΡΑΡΊΣΜΑΤΟΣ 42
3.2. ΑΠΟΤΕΛΈΣΜΑΤΑ ΧΕΙΡΟΚΊΝΗΤΟΥ ΣΧΟΛΙΑΣΜΟΎ 45
3.3. ΑΠΟΤΕΛΈΣΜΑΤΑ ΤΟΥ ΑΛΓΌΡΙΘΜΟΥ MICROT-CNN 50

ΚΕΦΆΛΑΙΟ 4: ΣΥΖΉΤΗΣΗ **52**

BIBLIOGRAPHY **54**

Εικόνες

ΕΙΚΌΝΑ 1: ΤΟ ΓΟΝΊΔΙΟ BACE1 ΚΩΔΙΚΟΠΟΙΕΊ ΤΗΝ Β-ΣΕΚΡΕΤΆΣΗ, Ο ΟΠΟΊΑ ΔΙΑΣΠΆ ΤΗΝ APP ΣΕ ΈΝΑ ΘΡΆΥΣΜΑ ΤΩΝ 99 ΑΜΙΝΟΞΈΩΝ, ΤΟ C99, ΚΑΙ ΈΝΑ ΤΩΝ 83, ΤΟ C83. Η Γ-ΣΕΚΡΕΤΆΣΗ ΜΕ ΤΗΝ ΣΕΙΡΆ ΤΗΣ ΕΠΕΞΕΡΓΆΖΕΤΑΙ ΤΟ ΠΡΩΤΟ ΚΑΙ ΈΧΟΥΜΕ ΩΣ ΑΠΟΤΕΛΕΣΜΑ ΑΥΤΗΣ ΤΗΣ ΔΙΑΔΙΚΑΣΙΑΣ ΤΗΝ ΕΞΩΚΥΤΤΑΡΙΚΉ ΈΚΚΡΙΣΗ Β-ΑΜΥΛΟΕΙΔΌΥΣ. 15
ΕΙΚΌΝΑ 2: Η ΔΡΆΣΗ ΤΩΝ ΝΕΥΡΟΪΝΙΔΙΑΚΏΝ ΣΩΡΏΝ ΚΑΙ ΤΩΝ ΑΜΥΛΟΕΙΔΏΝ ΠΛΑΚΏΝ [2] 15
ΕΙΚΌΝΑ 3: *Η ΣΤΑΔΙΑΚΉ ΕΞΆΠΛΩΣΗ ΤΗΣ ΝΌΣΟΥ ΤΟΥ ALZHEIMER* [3] 16
ΕΙΚΌΝΑ 4: *ΣΩΜΆΤΙΑ LEWY ΣΤΗΝ ΜΈΛΑΙΝΑ ΟΥΣΊΑ ΑΣΘΕΝΉ ΜΕ ΤΗΝ ΝΌΣΟ ΤΟΥ PARKINSON* [10]..... 19
ΕΙΚΌΝΑ 5: *ΟΙ ΑΥΞΗΜΈΝΕΣ ΣΕ ΌΓΚΟ ΠΛΕΥΡΙΚΈΣ ΚΟΙΛΊΕΣ ΤΟΥ ΕΓΚΕΦΆΛΟΥ* 24
ΕΙΚΌΝΑ 6: *Η ΒΙΟΓΈΝΕΣΗ ΤΩΝ MICRORNAS* [29] 25
ΕΙΚΌΝΑ 7: *ΤΟ ΣΧΉΜΑ ΦΟΥΡΚΈΤΑΣ ΤΟΥ ΠΡΩΤΑΡΧΙΚΟΎ ΜΟΡΊΟΥ MIR-1 ΣΤΟΝ ΆΝΘΡΩΠΟ* 26
ΕΙΚΌΝΑ 8: *ΠΑΡΆΔΕΙΓΜΑ ΕΓΓΡΑΦΉΣ ΣΤΟ ΑΡΧΕΊΟ ENTINTY_BI.CSV*..... 35
ΕΙΚΌΝΑ 9: ΠΑΡΆΔΕΙΓΜΑ ΕΓΓΡΑΦΉΣ ΣΤΟ ΑΡΧΕΊΟ META_INFO.CSV..... 35
ΕΙΚΌΝΑ 10: *ΠΑΡΆΔΕΙΓΜΑ ΕΓΓΡΑΦΉΣ ΣΤΟ ΑΡΧΕΊΟ FINALRESULT.TXT*..... 37
ΕΙΚΌΝΑ 11: *ΤΟ ΠΟΣΟΣΤΌ ΤΩΝ ΆΡΘΡΩΝ ΠΟΥ ΉΤΑΝ ΔΙΑΘΈΣΙΜΑ ΣΤΗΝ PMC* 42
ΕΙΚΌΝΑ 12: *ΤΟ ΠΟΣΟΣΤΌ ΤΩΝ ΆΡΘΡΩΝ ΜΕ PMCID ΠΟΥ ΕΊΧΑΝ ΔΙΑΘΈΣΙΜΟ ΚΕΊΜΕΝΟ ΕΝΤΌΣ ΤΟΥ XML* 43
ΕΙΚΌΝΑ 13: *ΤΟ ΠΟΣΟΣΤΌ ΤΩΝ ΆΡΘΡΩΝ ΠΟΥ ΠΈΡΑΣΑΝ ΚΑΙ ΤΑ ΔΎΟ ΦΊΛΤΡΑ*..... 43
ΕΙΚΌΝΑ 14: *ΔΙΆΓΡΑΜΜΑ ΡΟΉΣ ΤΟΥ ΑΡΙΘΜΟΥ ΤΩΝ ΆΡΘΡΩΝ ΠΟΥ ΕΠΙΛΕΧΘΗΚΑΝ ΣΕ ΚΆΘΕ ΒΉΜΑ* 44

Πίνακες

ΠΊΝΑΚΑΣ 1: *ΤΑ ΜΟΝΤΈΛΑ ΤΟΥ SCISCPACY, ΤΑ ΣΎΝΟΛΑ ΣΤΑ ΟΠΟΊΑ ΕΚΠΑΙΔΕΎΤΗΚΑΝ ΚΑΙ ΟΙ ΚΛΆΣΕΙΣ ΠΟΥ ΚΑΤΗΓΟΡΙΟΠΟΙΟΎΝ ΤΙΣ ΟΝΤΌΤΗΤΕΣ.* 34
ΠΊΝΑΚΑΣ 2: *ΟΙ ΚΑΝΟΝΙΚΈΣ ΕΚΦΡΆΣΕΙΣ ΠΟΥ ΧΡΗΣΙΜΟΠΟΙΉΘΗΚΑΝ ΓΙΑ ΤΗΝ ΈΥΡΕΣΗ ΤΩΝ* 36
ΠΊΝΑΚΑΣ 3: *Ο ΠΊΝΑΚΑΣ ΠΕΡΙΛΑΜΒΆΝΕΙ ΤΑ ΓΟΝΊΔΙΑ ΚΑΙ ΤΑ MICRORNAS ΠΟΥ ΑΠΟΤΈΛΕΣΑΝ ΕΊΣΟΔΟ ΣΤΟΝ ΑΛΓΌΡΙΘΜΟ, ΚΑΘΏΣ ΕΠΊΣΗΣ ΚΑΙ ΤΟΥΣ ΜΟΝΟΝΟΥΚΛΕΟΤΙΔΙΚΟΎΣΠΟΛΥΜΟΡΦΙΣΜΟΎΣ ΠΟΥ ΕΝΣΩΜΑΤΏΘΗΚΑΝ ΣΤΙΣ ΑΛΛΗΛΟΥΧΊΕΣ ΚΑΤΆ ΤΗΝ ΔΕΎΤΕΡΗ ΕΚΤΈΛΕΣΗ. ΝΑ ΣΗΜΕΙΏΘΕΊ ΟΤΙ ΟΙ ΑΛΛΑΓΈΣ ΤΩΝ ΑΛΛΗΛΌΜΟΡΦΩΝ (ALLELES) ΚΑΘΏΣ ΚΑΙ ΟΙ ΘΈΣΕΙΣ ΤΩΝ ΑΛΛΑΓΏΝ ΑΝΑΓΡΆΦΟΝΤΑΙ ΌΠΩΣ ΣΤΗΝ DBSNP.* 41
ΠΊΝΑΚΑΣ 4: *ΤΑ ΜΕΤΑΔΕΔΟΜΈΝΑ ΠΟΥ ΣΥΛΛΈΧΘΗΚΑΝ ΑΠΌ ΤΑ ΆΡΘΡΑ (ΜΕΡΟΣ Α)*..... 46
ΠΊΝΑΚΑΣ 5: ΤΑ ΜΕΤΑΔΕΔΟΜΈΝΑ ΠΟΥ ΣΥΛΛΈΧΘΗΚΑΝ ΑΠΌ ΤΑ ΆΡΘΡΑ (ΜΈΡΟΣ Β)..... 47
ΠΊΝΑΚΑΣ 6: *ΤΑ ΔΕΔΟΜΈΝΑ ΠΟΥ ΣΥΓΚΕΝΤΡΏΘΗΚΑΝ ΚΑΤΆ ΤΟΝ ΧΕΙΡΟΚΊΝΗΤΟ ΣΧΟΛΙΑΣΜΌ (ΜΕΡΟΣ Α)* 48
ΠΊΝΑΚΑΣ 7: *ΤΑ ΔΕΔΟΜΈΝΑ ΠΟΥ ΣΥΓΚΕΝΤΡΏΘΗΚΑΝ ΚΑΤΆ ΤΟΝ ΧΕΙΡΟΚΊΝΗΤΟ ΣΧΟΛΙΑΣΜΌ (ΜΕΡΟΣ Β)*..... 49
ΠΊΝΑΚΑΣ 8: *ΤΑ ΤΕΛΙΚΆ ΑΠΟΤΕΛΈΣΜΑΤΑ* 51

Κεφάλαιο 1: Εισαγωγή

1.1.Η Νόσος Alzheimer

1.1.1. Τα συμπτώματα της νόσου

Η νόσος του Alzheimer είναι μία νευροεκφυλιστική νόσος, δηλαδή μία νόσος που προσβάλει τους νευρώνες του εγκεφάλου, ενώ αποτελεί την πρώτη νευροεκφυλιστική νόσο στην πρόκληση άνοιας. Ο όρος «άνοια» αναφέρεται σε εξασθένηση των γνωστικών λειτουργιών του εγκεφάλου, όπως η διαταραχή λόγου, η διαταραχή μνήμης, η διαταραχή στην αντίληψη του χώρου, καθώς επίσης και η απραξία, σε σημείο που παρεμποδίζεται η ανεξάρτητη διαβίωση του ατόμου. Στα συμπτώματα της άνοιας συμπεριλαμβάνονται, εκτός των άλλων, συμπεριφορικές και ψυχικές διαταραχές, όπως κατάθλιψη, άγχος, ανάρμοστη συμπεριφορά ή επιθετικότητα, οπτικές ψευδαισθήσεις, διαταραχές ύπνου κ.α. Η νόσος πρωτοπεριγράφηκε από τον Γερμανό ψυχίατρο Alois Alzheimer το 1906. [1]

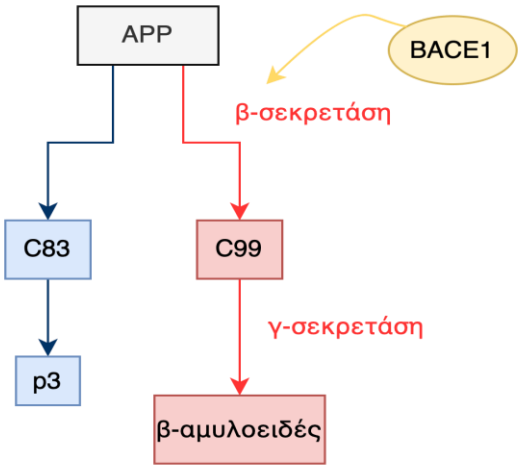
1.1.2. Τα αίτια της νόσου

Η νόσος του Alzheimer χαρακτηρίζεται από την βαθμιαία ανάπτυξη παθολογικών καταστάσεων στο κεντρικό νευρικό σύστημα (ΚΝΣ), που έχουν ως συνέπεια την καταστροφή των νευρικών κυττάρων και την ατροφία του εγκεφάλου. Παρά το γεγονός ότι ο ακριβής μηχανισμός πρόκλησης της ασθένειας δεν είναι γνωστός, η επικρατέστερη υπόθεση θέλει την νόσο να οφείλεται σε συσσώρευση πρωτεϊνών, σχηματίζοντας έτσι δύο βασικές δομές:

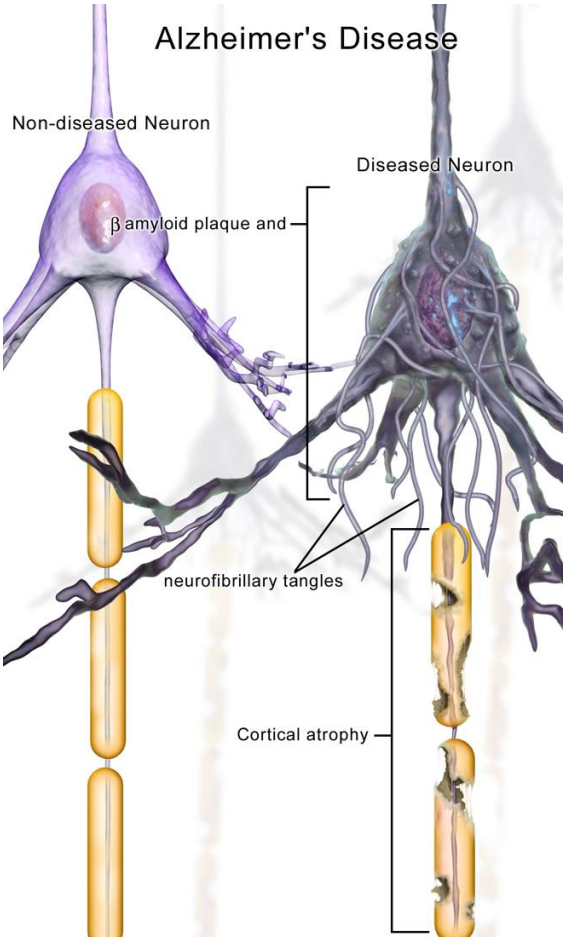
- Πλάκες αμυλοειδούς (amyloid plaques): ο σχηματισμός τους οφείλεται σε υπερπαραγωγή του β-αμυλοειδούς (Αβ), που έχει ως αποτέλεσμα την δημιουργία ολιγομερών, η συσσώρευση των οποίων σχηματίζει τις πλάκες αμυλοειδούς. Το β-αμυλοειδές συμμετέχει στο μονοπάτι επεξεργασίας της πρόδρομης πρωτεΐνης αμυλοειδούς (amyloid precursor protein-APP). Στην εικόνα περιγράφεται συνοπτικά το μονοπάτι επεξεργασίας της APP.

- Νευροϊνιδιακοί σωροί (neurofibrillary tangles): σχηματίζονται από την πρωτεΐνη tau η οποία βρίσκεται σε αφθονία εντός του ΚΝΣ και παίζει σημαντικό ρόλο στην λειτουργία των νευρώνων.[20]

Τόσο οι πλάκες αμυλοειδούς, όσο και η πρωτεΐνη tau μέσω των νευροϊνιδιακών σωρών, μπορούν να συμπεριφερθούν σαν πρίονες (prion-like proteins), δηλαδή μπορούν να μεταδώσουν την μη φυσιολογική αναδίπλωση τους σε άλλες φυσιολογικές πρωτεΐνες ίδιου τύπου. Με αυτόν τον τρόπο προκαλείται έλλειψη της πλαστικότητας των νευρικών συνάψεων, αλλά και καταστροφή των τελευταίων, με αποτέλεσμα την έκπτωση των γνωστικών λειτουργιών αποτέλεσμα την έκπτωση των γνωστικών λειτουργιών. [1]



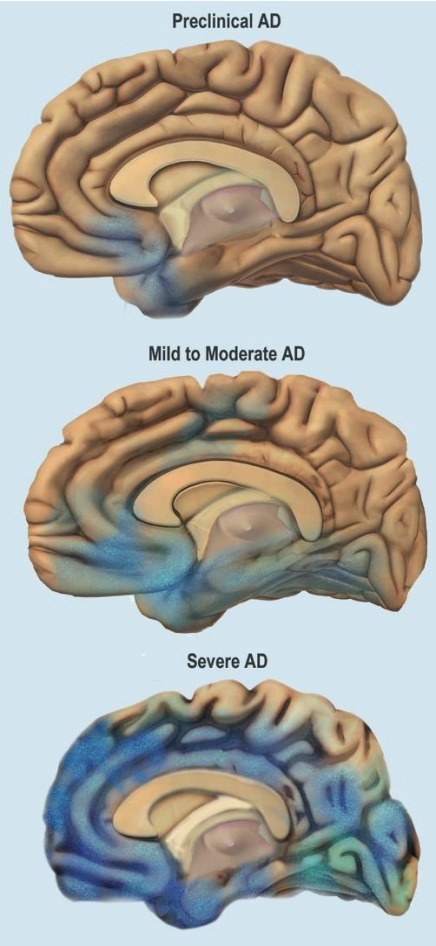
Εικόνα 1: το γονίδιο BACE1 κωδικοποιεί την β-σεκρετάση, ο οποία διασπά την APP σε ένα θραύσμα των 99 αμινοξέων, το C99, και ένα των 83, το C83. Η γ-σεκρετάση με την σειρά της επεξεργάζεται το πρώτο και έχουμε ως αποτέλεσμα αυτής της διαδικασίας την εξωκυτταρική έκκριση β-αμυλοειδούς.



Εικόνα 2: *Η δράση των νευροϊνιδιακών σωρών και των αμυλοειδών πλακών [2]*

1.1.3. Τα στάδια της νόσου

Η νόσος του Alzheimer εμφανίζεται σταδιακά και πιο συγκεκριμένα σε τρία στάδια. Κατά το προκλινικό στάδιο, παρά το γεγονός ότι έχει ξεκινήσει η συσσώρευση των πλακών αμυλοειδούς και των νευροϊνιδικών σωρών, καθώς επίσης και η πρόκληση μερικών συναπτικών δυσλειτουργιών, αυτά δεν είναι σοβαρής έκτασης ώστε ο ασθενής να εμφανίσει συμπτώματα. Το προκλινικόστάδιο είναι το μεγαλύτερο χρονολογικά, καθώς μπορεί να έχει διάρκεια πάνω από 10 έτη. Έπειτα, κατά το δεύτερο στάδιο, έχουμε την εμφάνιση των πρώτων συμπτωμάτων. Συγκεκριμένα, αυτό το στάδιο χαρακτηρίζεται από την ήπια γνωστική διαταραχή (Mild Cognitive Impairment-MCI), χωρίς αυτή όμως να εμποδίζει την καθημερινότητα του ατόμου. Η διάρκεια του δεύτερου σταδίου της νόσου ποικίλει σε κάθε άτομο. Στο τελικό στάδιο της νόσου έχουμε την κλινική εκδήλωση της νόσου, με τα γνωστικά συμπτώματα να είναι πλέον σοβαρά και να συνοδεύονται από συμπεριφορικάκαι ψυχικά συμπτώματα. Σε αυτό το στάδιο ο πάσχων αδυνατεί να ολοκληρώσει ακόμα και τις βασικές δραστηριότητες της καθημερινής διαβίωσης.



Εικόνα 3: *η σταδιακή εξάπλωση της νόσου του Alzheimer* [3]

1.1.4. Παράγοντες κινδύνου

1.1.4.1. Γενετικοί παράγοντες

Η εμφάνιση της νόσου πριν τα 60-65 έτη, που καλείται ως νόσος Alzheimer πρώιμης έναρξης (early onset Alzheimerdisease), θεωρείται πολυπαραγοντική, δηλαδή προκαλείται από έναν συνδυασμό γενετικών και περιβαλλοντικών παραγόντων. Έχουν βρεθεί μεταλλάξεις σε τρία γονίδια που προκαλούν την αυτοσωμική επικρατή μορφή της νόσου του Alzheimer. Τα γονίδια αυτά συμβάλλουν στην παραγωγή και την επεξεργασία του β-αμυλοειδούς και είναι η πρόδρομη πρωτεΐνη αμυλοειδούς (APP) στο χρωμόσωμα 21, η πρεσενιλίνι 1 (presenilin 1, PSEN1) στο χρωμόσωμα 14 και η πρεσενιλίνι 2 (presenilin 2, PSEN2) στο χρωμόσωμα 1. Συνδυασμός των τριών μεταλλάξεων φαίνεται να προκαλεί το 11% των περιπτώσεων πρώιμης έναρξης της νόσου και το 0.6% όλων των περιπτώσεων συνολικά. [4]

Η όψιμη εκδήλωση της νόσου υπολογίζεται να οφείλεται κατά 60% με 80% σε κληρονομικούς παράγοντες, με το υπόλοιπο ποσοστό να αποδίδεται στους περιβαλλοντικούς παράγοντες. Το 27% των κληρονομικών γενετικών παραγόντων αφορούν την απολιποπρωτεΐνη Ε (ApoE). Η παραπάνω πρωτεΐνη συμμετέχει στην μεταφορά λιπιδίων αλλά και στον μεταβολισμό, ενώ στην νόσο του Alzheimer συνεντοπίζεται με τα ολιγομερή Αβ, ενισχύοντας την καταστροφή των συνάψεων. [4]

1.1.4.2. Περιβαλλοντικοί παράγοντες και παθήσεις

Οι περιβαλλοντικοί παράγοντες είναι ένας σημαντικός παράγοντας κινδύνου για την εμφάνιση της νόσου. Αρχικά, οι κρανιοεγκεφαλικές κακώσεις, ιδιαιτέρως αν αυτές είναι πολλαπλές, αυξάνουν την πιθανότητα εμφάνισης της νόσου. Επιπλέον, παθήσεις όπως η στεφανιαία νόσος, ο ζακχαρώδηςδιαβήτης και η αρτηριακή υπέρταση, έχουν συνδεθεί με την εμφάνιση της νόσου.

1.1.4.3. Ο ρόλος των microRNAs

Έρευνες έχουν δείξει πως τα microRNAs (miRNAs)μπορούν να παίξουν σημαντικό ρόλο στην εμφάνιση της νόσου Alzheimer. Πιο συγκεκριμένα έχουν βρεθεί microRNAs που ρυθμίζουν την έκφραση των APP και BACE1, με αποτέλεσμα να συμμετέχουν και στα παθολογικά μονοπάτια εμφάνισης της νόσου. Παραδείγματος χάριν, για τα miR-106a και miR-29a έχουν βρεθεί μέρη πρόσδεσης (binding sites) στα γονίδια APP και BACE1 αντίστοιχα, με αποτέλεσμα τα πρώτα να ρυθμίζουν την έκφραση αυτών των γονιδίων. [5] Είναι πασιφανές λοιπόν ότι οποιαδήποτε μεταλλαγή ή μετάλλαξη στα microRNAs ή στα μέρη πρόσδεσης αυτών στα γονίδια μπορεί να οδηγήσει σε νευροεκφυλιστικές διαδικασίες.
Συνεπώς η ύπαρξη κάποιων miRNA σε ιστούς του ΚΝΣ, το εγκεφαλονωτιαίο υγρό και τον εγκέφαλο μπορεί να χρησιμοποιηθεί ως βιοδείκτης της νόσου.

1.5. Επιδημιολογία

Η εμφάνιση της νόσου είναι ανεξάρτητη φυλετικής καταγωγής και σχετίζεται κυρίως με την ηλικία και το φύλο. Η συχνότητα εμφάνισης της νόσου στην Ευρώπη είναι 11.08 ανά 1000 ανθρωποέτη. Με βάση το φύλο, η συχνότητα εμφάνισης της νόσου στις γυναίκες είναι 13.25 ανά 1000 ανθρωποέτη και στους άντρες 7.02 ανά 1000 ανθρωποέτη. Ο επιπολασμός της νόσου, δηλαδή το ποσοστό του πληθυσμού που νοσεί σε μία συγκεκριμένη χρονική στιγμή, είναι στο 0.97% για άτομα ηλικίας 65–74, στο 7.66% για άτομα ηλικίας75–84, και στο 22.53% για άτομα που ξεπερνάνε την ηλικία των 85, σύμφωνα με δεδομένα μίας μετα-ανάλυσης του επιπολασμού στην Ευρώπη. [6]

1.2.Η νόσος του Parkinson

1.2.1. Τα συμπτώματα της νόσου

Η νόσος του Parkinson είναι μία χρόνια προοδευτική ασθένεια του νευρικού συστήματος που θεωρείται πως είναι η δεύτερη πιο συχνή νευροεκφυλιστική νόσος μετά την νόσοτου Alzheimer. Η εκδήλωση της ασθένειας περιλαμβάνει τόσο κινητικά όσο και μη κινητικά συμπτώματα. Πιο συγκεκριμένα στα κινητικά συμπτώματα συμπεριλαμβάνονται η μυϊκή ακαμψία, η βραδυκινησία,άκαμπτα άκρα, σύρσιμο των ποδιών, απώλεια της ισορροπίαςκαι ο τρόμος σε κατάσταση ηρεμίας, ενώ τα μη κινητικά συμπτώματα αφορούν συμπεριφορικές και ψυχικές διαταραχές, όπως άνοια, κατάθλιψη, άγχος και διαταραχές ύπνου. Η περιγραφή της νόσου πρωτοδημοσιεύτηκε από τον Άγγλο γιατρό James Parkinson το 1817. [7] [8]

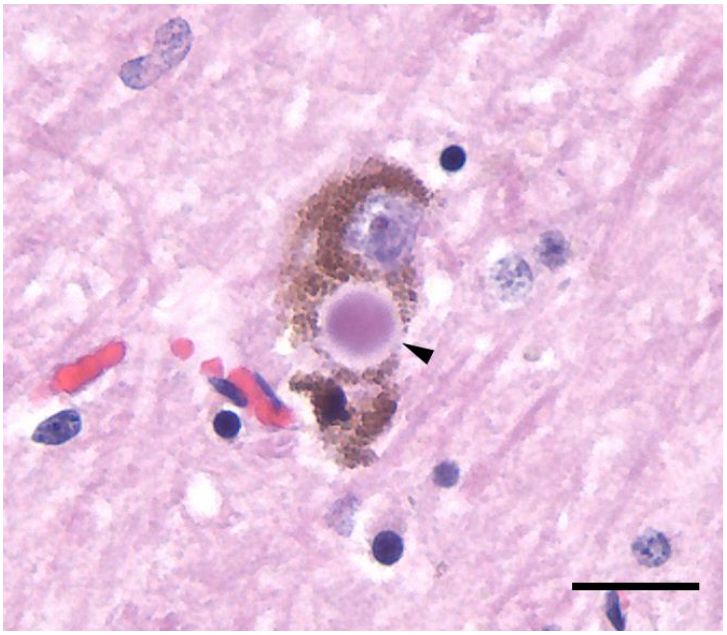
1.2.2. Τα αίτια της νόσου

1.2.2.1. Νευροχημικά αίτια

Όπως κάθε νευροεκφυλιστική νόσος, η νόσος του Parkinson σχετίζεται με σταδιακή ανάπτυξη μοριακών και κυτταρικών αλλοιώσεων, που οδηγούν ύστερα σε καταστροφή των νευρικών κυττάρων και κατά συνέπεια των νευρώνων. Η νευροχημική παθολογία της νόσου του Parkinson χαρακτηρίζεται από μη φυσιολογική παραγωγή και λειτουργία της ντοπαμίνης, έναν νευροδιαβιβαστή του εγκεφάλου που παράγεται από τους κεντρικούς νευρώνες του μεσεγκεφάλου. Οι παραπάνω ανωμαλίες που αφορούν την ντοπαμίνη προκαλούνται από καταστροφή ή εκφυλισμό τον ντοπαμινενεργικών νευρώνων στη μέλαινα ουσία του μεσεγκέφαλου. Πιο συγκεκριμένα, Διαταραχές στους ντοπαμινενεργικούς νευρώνες μπορούν να προκαλέσουν δυσλειτουργία στα βασικά γάγγλια του εγκεφάλου που είναι υπεύθυνα για τις ακούσιες και εκούσιες κινήσεις. [9]

1.2.2.2. Γενετικά αίτια και σωμάτια Lewy

Όσον αφορά τα γενετικά αίτια της νόσου, γενετικές μεταλλάξεις σε βασικές πρωτείνες του ΚΝΣ, μπορούν να οδηγήσουν σε καταστροφή των νευρώνων. Μεταλλάξεις στο γονίδιο της α-συνουκλεΐνης, μίας νευρωνικής πρωτεΐνης, έχουν συνδεθεί με την ανάπτυξη της νόσου. Πιο συγκεκριμένα, η μη φυσιολογική ρύθμιση της α-συνουκλεΐνης μπορεί να οδηγήσει σε συσσώρευση αυτής σε σωμάτια Lewy. Τα σωμάτια αυτά σχηματίζονται στους νευρώνες και οδηγούν στην μη φυσιολογική λειτουργία τους. Τα σωμάτια Lewy αποτελούν χαρακτηριστικό γνώρισμα της νόσου του Parkinson. Αυτές η παθολογίες πιθανολογείται ότι μεταδίδονται σε υγιή κύτταρα μέσω ενός μηχανισμού που δρα σαν πρίονες. [7]



Εικόνα 4: *Σωμάτια Lewy στην μέλαινα ουσία ασθενή με την νόσο του Parkinson* [10]

1.2.2.3. Οξειδωτικό στρες και μιτοχόνδρια

Το οξειδωτικό στρες αλλά και η μιτοχονδριάκηδυσλειτουργία συμβάλλουν στην περαιτέρω απόπτωση των νευρώνων και έχουν συνδεθεί με την ανάπτυξη της νόσου. Το οξειδωτικό στρες μπορεί να ενισχυθεί με πολλούς τρόπους. Ο μεταβολισμός της ντοπαμίνης στην μέλαινα ουσία του εγκεφάλου παράγει τοξικές ελεύθερες ρίζες, με αποτέλεσμα οποιαδήποτε ανωμαλία στην ποσότητα ντοπαμίνης να ενισχύει το οξειδωτικό στρες. Επιπλέον, οι διεργασίες που ενεργοποιούνται κατά την φλεγμονή σε περιοχές του εγκεφάλου εμπλέκονται στην παθολογία της νόσου του Parkinson μέσω της ενίσχυσης του οξειδωτικού στρες. Πιο συγκεκριμένα, κατά την φλεγμονή έχουμε αύξηση των διαμεσολαβητών ιντερλευκίνης και TNF-α, οι οποίοι ενεργοποιούν την λειτουργία των μικρογαγλιακών κυττάρων και εκκρίνουν νιτρικό οξύ (NO). Έτσι, έχουμε ως

αποτέλεσμα την επιδείνωση του οξειδωτικού στρες και την καταστροφή των κυττάρων.

Τα μιτοχόνδρια είναι ζωτικής σημασίας οργανίδια που συμβάλλουν στην παραγωγή ενέργειας στο κύτταρο. Έρευνες έχουν συνδέσει την έλλειψη του συμπλόκου 1 της βασικής αναπνευστικής αλυσίδας στα μιτοχόνδρια με την νόσο του Parkinson, ενώ το γεγονός φέρεται να είναι μοναδικό χαρακτηριστικό της νόσου και δεν συνδέεται με άλλη νευροεκφυλιστική νόσο. Αυτή η έλλειψη οδηγεί σε κυτταρικό θάνατο μέσω έκκριση ειδικών αποπτικών παραγόντων και επιπλέον ενίσχυσης του οξειδωτικού στρες. [11]

1.2.3. Τα στάδια της νόσου

Η εμφάνιση της νόσου του Parkinson θεωρείται προοδευτική και μπορεί να διαχωριστεί σε 3 φάσεις. Κατά την φάση 1, την προκλινική φάση της νόσου του Parkinson, το άτομο δεν εμφανίζει κανένα σύμπτωμα και η ύπαρξη της νόσου μπορεί να υποστηριχθεί μόνο από τον εντοπισμό μόνο κάποιον βιοδεικτών. Κατά την φάση 2, έχουμε την εμφάνιση των πρώτων μη-κινητικών συμπτωμάτων, που οφείλονται σε μικρές παθολογίες στην μέλαινα ουσία του εγκεφάλου. Η 3^{η}και τελευταία φάση χαρακτηρίζεται από την έναρξη των κινητικών συμπτωμάτων και την επιδείνωση των μη κινητικών, καθώς έχουμε πολύ μεγαλύτερες αλλοιώσεις στην μέλαινα ουσία. [12]

1.2.4. Παράγοντες κινδύνου

1.2.4.1. Γενετικοί παράγοντες

Η νόσος του Parkinson είναι μία πολυπαραγοντική νόσος και η εμφάνιση της μπορεί να ενισχυθεί τόσο από γενετικούς όσο και από περιβαντολογικούς παράγοντες. Όσον αφορά τους γενετικούς παράγοντες, μεταλλάξεις ή μεταλλαγές σε σημαντικά γονίδια του ΚΝΣ μπορεί να συμβάλλουν στην εμφάνιση της νόσου. Μεταλλάξεις στο γονίδιο της παρκινίνης PRKN και στο PARK1 έχουν συνδεθεί με την μιτοχονδριακή δυσλειτουργία και την πρόωρη εμφάνιση της ασθένειας. Παρά το γεγονός ότι η νόσος του Parkinson δεν θεωρείται κληρονομική, τέτοιου είδους μεταλλάξεις έχουν ως αποτέλεσμα μία γενετική προδιάθεση του ατόμου, εάν αυτό έχει συγγενείς πρώτου βαθμού με την νόσο και ιδιαίτερα μεταξύ των αδερφών. Πιο συγκεκριμένα, μεταλλάξεις στο γονίδιο της α-συνουκλεΐνη στο χρωμόσωμα 4 συνδέονται με κληρονομική προδιάθεση αυτής. [9] [13]

1.2.4.2. Περιβαλλοντικοί παράγοντες

Οι περιβαλλοντικοί παράγοντες είναι πολύ σημαντικοί παράγοντες κινδύνου για την εμφάνιση της νόσου. Συγκεκριμενά η έκθεση σε τοξίνες φαίνεται να ευνοεί τον σχηματισμό των παθολογιών της νόσου. Τα φυτοφάρμακα, τα παρασιτοκτόνα, τοξίνες μέσα σε τροφές ή υφάσματα μπορούν να προκαλέσουν εγκεφαλική φλεγμονή, η οποία με την σειρά την να αυξήσει την πιθανότητα εμφάνισης της νόσου. [14]

1.2.4.3. Φύλο και ηλικία

Η ηλικία είναι ο πιο ισχυρός παράγοντας κινδύνου, με μέση ηλικία εμφάνισης της νόσου να είναι τα 50 με 60 έτη. Το φύλο του ατόμου σε συνδιασμό με γενετικούς και περιβαλλοντικούς παράγοντες αποτελεί επίσης έναν παράγοντα που φαίνεται να ευνοεί την εμφάνιση της νόσου, καθώς οι άντρες έχουν αυξημένες πιθανότητες νόσησης σε σχέση με τις γυναίκες. [8]

1.2.4.4. Ο ρόλος των microRNAs

Τα microRNAs (miRNAs) συμμετέχουν σε μονοπάτια πρόκλησης του οξειδωτικού στρες με αποτέλεσμα να παίζουν ρόλο στην ανάπτυξη της νόσου Πάρκινσον, ενώ μπορούν να χρησιμοποιηθούν και ως βιοδείκτες αυτής. Μεταξύ των πολλών microRNAs που έχουν συνδεθεί με την νόσο, τα miR-34b και miR-34c εντοπίζονται σε αυξημένα επίπεδα σε ασθενείς που εμφάνισαν πρόωρα ή μη την νόσο, γεγονός που δικαιολογείται καθώς τα miR-34b και miR-34c στοχεύουν το γονίδιο PRKN, το οποίο όπως έχει ήδη αναφερθεί έχει συνδεθεί με την ασθένεια.

1.2.5. Επιδημιολογία

Η νόσος του Parkinson υπολογίζεται ότι προσβάλει 1 στα 2 άτομα για έναν τύχαιο πληθυσμό 1000 ατόμων. Η εμφάνισης της νόσου πριν την ηλικία των 50 είναι σπάνια, καθώς τα πρώτα συμπτώματα κατά μέσο όρο εμφανίζονται μετά την ηλικία των 60. Ο επιπολασμός μετά τα 60 έτη υπολογίζεται στο 1% και ανέρχεται στο 4% για μεγαλύτερες ηλικίες. Όσον αφορά το φύλο, τα άτομα αρσενικού φύλουέχουν κατά 1.5 φορά υψηλότερες πιθανότητες να εμφανίσουν την νόσο σε σχέση με τα άτομα θηλυκού φύλου. [8]

1.3. Σχιζοφρένεια

1.3.1. Τα συμπτώματα της νόσου

Η Σχιζοφρένεια είναι μία χρόνια ψυχική νόσος που η έναρξη των πρώτων συμπτωμάτων ξεκινά κατά μέσο όρο στα πρώτα χρόνια της ενήλικης ζωής. Τα συμπτώματα της ασθένειας είναι διαφορετικής έντασης και φύσης στο κάθε άτομο, αλλά κυρίως αφορούν συμπεριφορικές ανωμαλίες, γνωστικές δυσλειτουργίες και συναισθηματικές παρεκκλίσεις. Πιο ειδικά, τα συμπτώματα της νόσου διαχωρίζονται σε θετικά και αρνητικά ή υπολειμματικά. Στα θετικά συμπτώματα περιλαμβάνονται οι ψευδαισθήσεις, οι παραληρηματικές ιδέες και οι ψυχοκινητικές αναταραχές και αντιμετωπίζονται επιτυχώς από τις φαρμακευτικές αγωγές. [15] Τα αρνητικά συμπτώματα αφορούν την κοινωνική απόσυρση, την απάθεια και την έλλειψη κίνητρου. Τέλος, τα άτομα με την νόσο αντιμετωπίζουν και γνωστικά συμπτώματα, όπως η δυσλειτουργία της λειτουργικής μνήμης και η μειωμένη προσοχή. [16] Η νόσος περιγράφηκε αρχικά από τον Ελβετό ψυχίατρο Έουγκεν Μπλόιλερ.

1.3.2.Τα στάδια της νόσου

Η νόσος δεν εμφανίζεται σε ξεκάθαρα στάδια. Παρόλ΄αυτά αν θα έπρεπε να διαχωρίσουμε την νόσο σε στάδια, αυτά θα ήταν το πρόδρομο και το ενεργές στάδιο. Κατά την πρόδρομη φάση, η νόσος δεν έχει εκδηλωθεί πλήρως αλλά το άτομο αντιμετωπίζει τα πρώτα συμπτώματα, όπως άγχος, συμπεριφορικές διαταραχές και κάποιες ελαφριάς μορφής παραισθήσεις. Κατά το ενεργές στάδιο η νόσος έχει κάνει την πλήρη εμφάνιση της και το άτομο αντιμετωπίζει μία ποικιλία συμπτωμάτων, όπως αυτά που αναφέρθηκαν στην ενότητα 1.3.1. [17]

1.3.3.Παράγοντες κινδύνου

1.3.3.1. Κληρονομικότητα

Η σχιζοφρένεια είναι μία πολυπαραγοντική νόσος η οποία φαίνεται να οφείλεται στον συνδυασμό γενετικών και περιβαλλοντικών παραγόντων. Το θετικό στην νόσο οικογενειακό ιστορικό είναι ένα παράγοντας κινδύνου εμφάνισης της ασθένειας, γεγονός το οποίο οφείλεται σε μεταλλαγές ή μεταλλάξεις που κληρονομούνται. Συγκεκριμένα, ενώ το ρίσκο εμφάνισης της νόσου είναι 1% για τον γενικό πληθυσμό, αυτό ανέρχεται στο 6.5% αν το άτομο έχει συγγενείς 1^{ου} βαθμού με την νόσο. [16] Σημειώνεται ότι το ρίσκο αυτό φτάνει στο 40% με 60% αν ο συγγενής αυτός είναι μονοζυγωτικό δίδυμο του ατόμου και στο 0% με 28% αν είναι διζυγωτικό. [17] Να σημειωθεί ότι τα μονοζυγωτικά δίδυμα μοιράζονται ταυτόσημο γονότυπο, ενώ να διζυγωτικά όχι.

1.3.3.2. Επιπλοκές κατά την προγεννητική περίοδο

Ένας εξίσου σημαντικός παράγοντας είναι οι επιπλοκές κατά την προγεννητική περίοδο. Πιο συγκεκριμένα, η πρόωρη γέννα, η περιγεννητική ασφυξία κατά την γέννα και ο εμβρυακός υποσιτισμός μπορούν να οδηγήσουν σε ψυχωτικές διαταραχές. [18]Επιπλέον, η έκθεση της μητέρας σε ιούς έχει συνδεθεί με την μετέπειτα εμφάνιση της ασθένειας στο παιδί. Σε αυτούς τους ιούς ανήκουν ο ιός της γρίπης, το πρωτόζωο τοξόπλασμα και ο HSV-2 (Herpes simplex virus 2). Γενικότερα, αυξημένοι δείκτες φλεγμονής κατά την κύηση, όπως η ιντερλευκίνη-8 και η C-αντιδρώσα πρωτεΐνησυνδέονται με την μετέπειτα εμφάνιση της νόσου. [19]

1.3.3.3. Κατάχρηση ουσιών

Η κατάχρηση ουσιών φαίνεται πως μπορεί μεταγενέστερα να οδηγήσει σε ψυχωτικές διαταραχές. Ο λόγος γίνεται κυρίως για ψυχοδιεγερτικά ναρκωτικά, δηλαδή ψυχοτρόπα φάρμακα που έχουν την δυνατότητα να διεγείρουν το νευρικό σύστημα. Τέτοια ναρκωτικά είναι η κοκαΐνη και μεθαμφεταμίνη. Επιπροσθέτως, η κατάχρηση του αλκοόλ και της κάνναβις φαίνεται να παίζουν ρόλο στην ανάπτυξη ψύχωσεων. Ωστώσο, το ποσοστό των ατόμων που εμφανίζουν σχιζοφρένεια έπειτα από κατάχρηση κάνναβις ή αλκοολ είναι χαμηλό. [20]

1.3.3.4. Γονιδιακοί παράγοντες και microRNAs

Πολλά γονίδια έχουν θεωρηθεί υποψήφια για την παθογένεση της νόσου. Ένα από αυτά είναι το MIR137HG το οποίο κωδικοποιεί το miR-137 και πολυμορφισμοί σε αυτό το γονίδιο οδηγούν σε παραγωγή μη φυσιολογικής ποσότητας του microRNA, γεγονός το οποίο έχει συνδεθεί με την νόσο. Τα ποσοστά του microRNA, όπως το miR-137, μπορούν να χρησιμοποιηθούν και ως βιοδείκτες. Επιπλέον, μεταλλαγές στο γονίδιο μεταβολισμού της ντοπαμίνης κατεχολ-Ο-μεθυλοτρανσφεράσης (COMT) σε συνδυασμό με την χρήση κάνναβις, είναι μία αλληλεπίδραση γονιδίου-περιβάλοντοςπου φαίνεται να ευνοεί την εμφάνιση σχιζοφρένειας. [21]

1.3.3.5. Κοινωνικοί παράγοντες

Πολλοί παράγοντες που αφορούν την κοινωνική ζωή του ατόμου έχουν συνδεθεί με αύξηση του ποσοστού του ρίσκουεμφάνισης. Τα ψυχικά τραύματα και οι αντιξοότητες, όπως κακοποίηση οποιασδήποτε μορφής, ο χαμός ενός αγαπημένου προσώπου, αλλά και ο σχολικός εκφοβισμός κατά την παιδική αλλά και ενήλικη ζωή μπορούν να θεωρηθούν παράγοντες κινδύνου. Τέλος, πολλές έρευνες υποστηρίζουν ότι η κοινωνική απομόνωση, η κατώτερη κοινωνική τάξη, η κατοίκηση σε μεγάλες πόλεις, αλλά και η μετανάστευση συνδέονται με περιπτώσεις εμφάνισης σχιζοφρένειας. [20] [22]

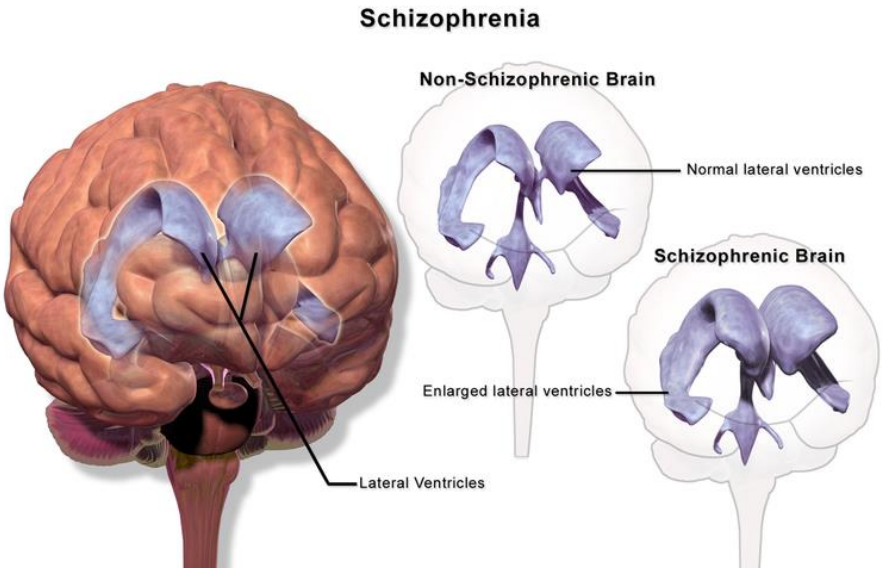
1.3.4. Τα αίτια της νόσου

1.3.4.1. Βιοχημικά αίτια

Υπάρχουν πάνω από μία υποθέσεις που αφορούν την παθολογία της σχιζοφρένειας. Η επικρατέστερη αφορά την περίσσεια ντοπαμίνης στο ραβδωτό σώμα του εγκεφάλου και συγκεκριμένα συνδέεται με τα θετικά συμπτώματα της νόσου. Μία άλλη θεωρία βασισμένη σε αυτήν της περίσσειας ντοπαμίνης, αναφέρεται στο ότι ο αξόνας υποθαλάμου-υπόφυσης-επινεφριδίων (hypothalamus–pituitary–adrenal /HPA) μπορεί να προκαλέσει πληθώρα συμβάντων που επηρεάζουν τις νευρικές λειτουργίες, όπως η δραστηριότητα των υποδοχέων της ντοπαμίνης. Αυτό το νευροενδροκρινικόσύστημα παίζει σημαντικό ρόλο στην αντίδραση του οργανισμού στο άγχος και συμμετέχει στο μονοπάτι παραγωγής της κορτιζόλης. Τα υψηλά επίπεδα κορτιζόληςευνοούν την εκδήλωση ψυχώσεων, ιδιαίτερα σε ομάδες με ήδη αυξημένο κίνδυνο. Μία ακόμα θεωρία εμπλέκει τους υποδοχείς NMDA γλουταμινικού, καθώς βρέθηκε πως φάρμακα που καταστέλλουν την δράση του γλουταμινικού, όπως η κεταμίνη και η φαινκυκλιδύνη (PCP) έχουν επιδράσεις παρόμοιες με αυτή της νόσου. [23]

1.3.4.2. Ανατομικά αίτια

Τεχνικές απεικόνισης του εγκεφάλου, όπως μαγνητική τομογραφία (MRI), σε ασθενείς με σχιζοφρένεια έχουν υποδείξει ανατομικές ανωμαλίες σε μέρη του εγκεφάλου, όπως ο μειωμένος όγκος της αμυγδαλής και του ιππόκαμπου, αυξημένος όγκος των πλευρικών κοιλιών του εγκεφάλου, καθώς επίσης και δομικές ανωμαλίες στα βασικά γάγγλια, στο προμετωπιαίο φλοιό, στον εμπρόσθιο φλοιό και στον ενδοκρινικό φλοιό. [24] [25]



Εικόνα 5: *οι αυξημένες σε όγκο πλευρικές κοιλίες του εγκεφάλου ατόμου με σχιζοφρένεια σε σύγκριση με αυτές ενός φυσιολογικού εγκεφάλου* [26]

1.3.4.5. Επιγενετική

Η έκθεση σε περιβαντολογικούς παράγοντες κινδύνου όπως η χρήση ουσίων, το στρες αλλά και η αντίστοιχοι προγεννητικοί παράγοντες που αναφέρθηκαν στην ενότητα3.Γ.ΙΙΙ, φαίνεται να προκαλούν επιγενετικές τροποποιήσεις,δηλαδή τροποποιήσεις στο τελικό προϊόν του γονιδίου χωρίς αλλαγή στην αλληλουχία του DNA, σε γονίδια που ρυθμίζουν την λειτουργία του γ-αμινοβουτυρικού οξέος (GABA), και κυρίως το GAD1. Το γ-αμινοβουτυρικό οξύ αποτελεί έναν νευροδιαβιβαστή που παίζει σημαντικό ρόλο στην ανάπτυξη του εγκεφάλου αλλά και την σωστή λειτουργία του. [21]

1.3.5. Επιδημιολογία

Η σχιζοφρένεια αφορά το 1% του γενικού πληθυσμού, με την επίπτωση να είναι ίση με 15,2 στα 100.000 άτομα. Η διάγνωση της νόσου σε άτομα με αρσενικό φύλο είναι συχνότερη από αυτήν στα άτομα με θηλυκό φύλο, με την αναλογία να υπολογίζεται στο 1,4:1 αντίστοιχα. [16] [27]Η μέση ηλικία εμφάνισης της νόσου είναι κατά τα πρώτα έτη της ενήλικης ζωής στα αρσενικά άτομα, ενώ τα θηλυκά άτομα εμφανίζουν μεταγενέστερα την νόσο, κατά μέσο όρο στα 25 με 35 έτη. Τα άτομα με σχιζοφρένεια έχουν κατά μέσο όρο μικρότερο προσδόκιμο ζωής. Η υψηλή

θνησιμότητα της νόσου έχει ως επακόλουθο την μείωση του προσδόκιμου ζωής των ατόμων με την νόσο, έως και 20 έτη. [28]

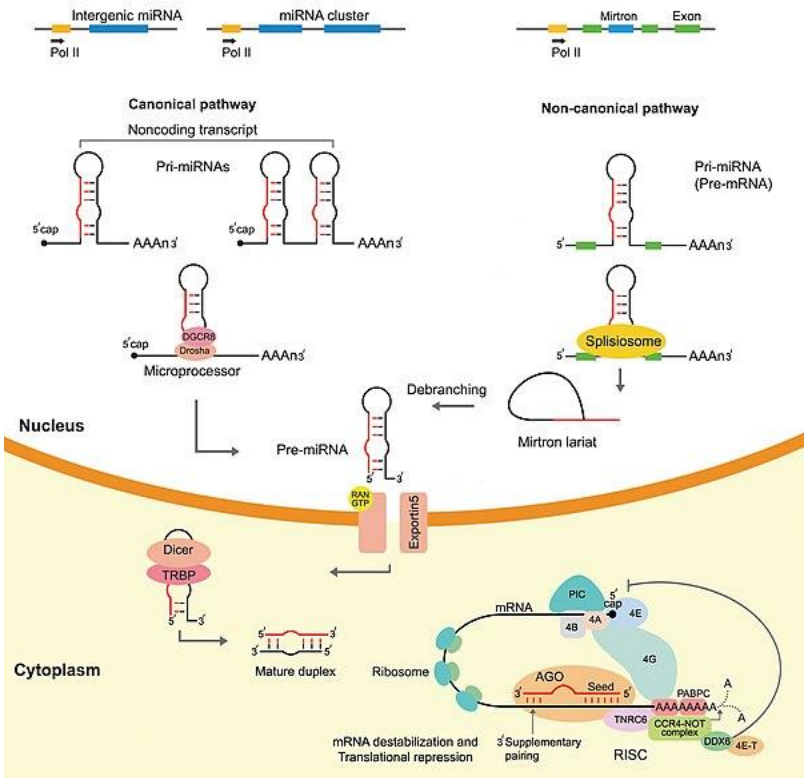
1.4. MicroRNA

1.4.1. Τι είναι τα microRNAs;

Τα microRNAs (miRNAs) αποτελούν μικρά μη κωδικά RNA, μήκους περίπου 22 βάσεων. Πολλές μελέτες έχουν δείξει ότι τα microRNAs έχουν πολλαπλές λειτουργίες σε ένα ευρύ φάσμα βιολογικών διεργασιών, όπως ο πολλαπλασιασμός, η απόπτωση, η διακοπή κυτταρικού κύκλου, η μετανάστευση κυττάρων και η εισβολή. Τα miRNAs μπορούν να προκαλέσουν αποικοδόμηση και/ή μεταφραστική καταστολή του mRNA με δέσμευση στην 3 'αμετάφραστη περιοχή (3' UTR), την κωδικοποιητική περιοχή (CDS) και σε μερικές περιπτώσεις την 5' αμετάφραστη περιοχή (5' UTR) του mRNA στόχου. Συνεπώς, τα miRNAs έχουν την δυνατότητα ρύθμισης της γονιδιακής έκφρασης μέσω της αλληλεπίδρασης τους με τα mRNAs.

1.4.2. Βιογένεση

Η βιογένεση των microRNAs (miRNAs) ολοκληρώνεται σε τρία στάδια: την μεταγραφή των γονιδίων σε πρωταρχικά μόρια miRNA, την επεξεργασία των πρώιμων μορίων miRNA(precursor miRNA/pre-miRNA) στον πυρήνα και την δημιουργία ώριμων μορίων miRNA στο κυτταρόπλασμα.



Εικόνα 6: *η βιογένεση των microRNAs* [29]

1.4.2.1. Στάδιο μεταγραφής

Η πλειοψηφία των microRNAs γονιδίων μεταγράφονται από την RNA πολυμεράση II (RNA polymerase ii/Pol II) σχηματίζοντας μεγάλα μετάγραφα , μερικών χιλιάδων νουκλεοτιδίων, σε σχήμα φουρκέτας, τα πρωταρχικά miRNAs(primary miRNAs/pri-miRNAs). Η πλειοψηφία των γονιδίων που κωδικοποιούν miRNAs εντοπίζονται σε διαγονιδιακέςπεριοχές, σχηματίζοντας ανεξάρτητες μεταγραφικές μονάδες. Μπορούν επίσης να εντοπιστούν σε ιντρονικές περιοχές άλλων γονιδίων με αποτέλεσμα να μεταγράφονται ως μέρος του γονιδίου στου οποίου την ιντρονική περιοχή βρίσκονται. [30]

Εικόνα 7: *το σχήμα φουρκέτας του πρωταρχικού μορίου miR-1 στον άνθρωπο*

1.4.2.2. Δεύτερο στάδιο βιογένεσης στον πυρήνα

Κατά το δεύτερο στάδιο της βιογένεσης των microRNAs, έχουμε την διάσπαση των πρωταρχικών μορίων miRNA σε μόρια πρώιμου miRNA (pre-miRNA) μήκους 60 με 100 νουκλεοτιδίων. Η διάσπαση αυτή γίνεται από ένα μικροεπεξεργαστικό σύμπλοκο που αποτελείται από ένα ένζυμο κλάσης 2 ριβονουκλεάσης III, το DROSHA και δύο υπομονάδες της πρωτεΐνης δέσμευσης δίκλωνου RNA DGCR8. Αρχικά, πρέπει να γίνει αναγνώριση του μορίου pri-miRNA καθώς υπάρχει πιθανότητα ύπαρξης και άλλωνμορίων σε σχήμα φουρκέτας. Υπάρχουν δύο βασικά αναγνωριστικά πάνω στο μόριο pri-miRNA που επιτρέπουν στο σύμπλοκο να το αναγνωρίσει. Το πρώτο αναγνωριστικό αποτελείται από ένα αταίριαστο μοτίβο GHG ακολουθούμενο από τρεις κύριες ακολουθίες, ένα UG μοτίβο στην βάση της φουρκέτας και ένα UGUG μοτίβο στην κορυφή της. Το δεύτερο αναγνωριστικό είναι η παρουσία μίας N6 μεθυλ-αδενοσίνης κοντά στο μόριο pri-miRNA, η οποία αναγνωρίζεται από ένα σύμπλοκο RNA και πρωτεΐνης (ετερογενής πυρινικό ριβονουκλεοπρωτεϊνικό μόριο/Heterogeneous ribonucleoprotein particle/ hnRNPs), το οποίο με την σειρά του αντιδρά με το διμερές DGCR8, διεγείροντας έτσι την επεξεργασία του miRNA. [31] [32]

Εφόσον γίνει η αναγνώριση του μορίου, το διμερές DGCR8 αλληλεπιδρά με τα αναγνωριστικά στοιχεία στο στέλεχος και στην κορυφή του μορίου μέσω των περιοχών δέσμευσης δίκλωνου RNA, οδηγώντας σε ακριβή επεξεργασία. Από την άλλη πλευρά, το DROSHA κόβει το μόριο σε μία απόσταση 11 ζεύγη βάσεων από την βάση της φουρκέτας (δηλαδή εκεί που έχουμε ένωση του μονόκλωνου και δίκλωνου τμήματος της φουρκέτας), απομακρύνοντας με αυτόν τον τρόπο τις μονόκλωνες ουρές. Είναι πολύ σημαντικό να σημειωθεί ότι ο προσανατολισμός του μορίου διατηρείται, καθώς το DROSHA και το διμερές DGCR8 αναγνωρίζουν

τα UG και UGUG μοτίβα, τα οποία βρίσκονται στην πλευρά του 5΄ άκρου. Έπειτα από αυτή την διαδικασία έχουμε την δημιουργία του πλέον πρώιμου μορίου miRNA (pre-miRNA). [31]

1.4.2.3. Εξαγωγή από τον πυρήνα

Το πρώιμο μόριο miRNA θα πρέπει να μεταφερθεί στο κυτταρόπλασμα, όπου και θα ολοκληρωθεί η ωρίμανση του. Η μεταφορά του μορίου γίνεται με την βοήθεια μίας πρωτεΐνης, της εξπορτίνης 5 (exportin 5/XPO5). Η πρωτεΐνη αυτή σχηματίζει ένα μόριο μεταφοράς μαζί με την πυρηνική πρωτεΐνη δέσμευσης GTP RAN•GTP, το οποίο έχει σχήμα ημισωλήνα. Εντός αυτού του μορίου προσδένεται και μεταφέρεται με ασφάλεια το pre-miRNA. Μόλις το σύμπλοκοβρεθεί στο κυτταρόπλασμα, το GTP υδρολύεται και το σύμπλοκο αποδιατάσσεται, με αποτέλεσμα την απελευθέρωση του πρώιμου μορίου miRNA. [32]

1.4.2.4. Τρίτο στάδιο επεξεργασίας στο κυτταρόπλασμα

Στο κυτταρόπλασμα το πρώιμο μόριο miRNA υφίσταται την επόμενη επεξεργασία από το ένζυμο Dicer, το οποίο ανήκει στην οικογένεια των ριβονουκλεασών III (RNase III) και στους ανθρώπους κωδικοποιείται από το γονίδιο DICER1. Το Dicer προσδένεται με το pre-miRNA με μία προτίμηση στην προεξοχή μήκους 2 νουκλεοτιδίων στο 3΄ άκρο που κατασκευάστηκε από το Drosha. Τα σημεία τομής στα θηλαστικά επιλέγονται με ένα μηχανισμό κατά τον οποίο το Dicer συνδέεται στο 5΄φωσφορυλιωμένο άκρο του πρώιμου miRNA και το κόβει σε μία απόσταση 22 νουκλεοτιδίων από το 5΄άκρο του. Ως αποτέλεσμα έχουμε την απομάκρυνση του τερματικού βρόγχου της φουρκέτας και την δημιουργία ενός δίκλωνο μόριο RNA μήκους 22 νουκλετιδίων, δηλαδή του ώριμου miRNA (mature miRNA). [32]

1.4.2.5. Δημιουργία του RISC

Το ώριμο δίκλωνο μόριο miRNA που κατασκευάστηκε από το Dicer φορτώνεται σε μία πρωτεΐνη αργοναύτη (Argonaute/AGO) έτσι ώστε να γίνει ο σχηματισμός τουΕπαγομένου από RNA Συμπλόκου Αποσιώπησης (RNA Induced Silencing Complex/RISC). Η φόρτωση του miRNAστην πρωτεΐνη αποτελεί το πρώτο βήμα για τον σχηματισμό του συμπλόκου RISC και το μόριο το οποίο σχηματίζεται σε αυτό το στάδιο καλείται ως πρώιμο RISC (pre-RISC). Κατά το δεύτερο και τελευταίο στάδιο, έχουμε την απομάκρυνση της μίας αλυσίδας του δίκλωνου μορίου. Το πλέον μονόκλωνο miRNA του ώριμου μορίου RISC αποτελεί οδηγό για την πρόσδεση του RISC στους στόχους των mRNAs (messengerRNAs). [33] [34]

1.4.2.6. Ρύθμιση της έκφρασης και μεταλλάξεις

Στην στη 3΄ αμετάφραστη περιοχή (3΄UTR) των mRNA-στόχων περιλαμβάνονται μέρη πρόσδεσης των miRNA(miRNA binding sites). Τα miRNA αναγνωρίζουν τα μέρη πρόσδεσης στα mRNA μέσω της περιοχής seed (seed region), μίας αλληλουχίας μήκους 2-8 νουκλεοτιδίων ξεκινώντας από το 5΄άκρο, η οποία είναι συμπληρωματική βάσει της

συμπληρωματικότητας Watson-Crick με τα μέρη αυτά(κανονικός τύπος πρόσδεσης/canonical binding type).Επιπλέον, υπάρχει και η περίπτωση πρόσδεσης χωρίς να υπάρχει συμπληρωματικότητα των βάσεων (μη κανονικός τύπος πρόσδεσης/ non canonical binding type). Η πρόσδεση ευνοείται και από άλλα στοιχεία αναγνώρισης, όπως μία μη ζευγαρωμένη αδενοσίνη στην αλληλουχία στόχο του mRNA. Με αυτόν τον τρόπο τα miRNAs μέσω του RISC ρυθμίζουν την έκφραση των γονιδίων. [35]

Τα miRNA αποτελούν βασικούς ρυθμιστές της γονιδιακής έκφρασης καθοδηγώντας την πληροφορία ενός γονιδίου, το οποίο με την σειρά του θα γίνει μία λειτουργική πρωτεΐνη. Είναι λοιπόν προφανές οτι τα miRNAs αποτελούν αναπόσπαστο κομμάτι της έκφρασης των γονιδίων παίζοντας με αυτόν τον τρόπο έναν ρόλο κλειδί στις λειτουργίες του οργανισμού.

Αλλαγές, ή μεταβολές τόσο στην έκφραση των miRNAs, όσο και στις σημαντικές αλληλουχίες πρόσδεσης με tamRNAs, δηλαδή
στα binding sites των mRNAs και στα seedregions των miRNAs, έχουν συνδεθεί με πολλές ασθένειες του ανθρώπου, όπως ο καρκίνος, οι μεταβολικές παθήσεις αλλά και παθήσεις του εγκεφάλου. Διάφοροι τύποι μεταλλάξεων, όπως η εισαγωγή ή η διαγραφή βάσεων (insertion or deletion/ indels), η επανάληψη βάσεων (copynumber variation/CNV) σε τέτοιες σημαντικές για την πρόσδεση περιοχές μπορούν να επηρεάσουν την αλληλεπίδραση miRNA::mRNA. Ένας από αυτούς τους τύπους μεταλλάξεων είναι οι μονονουκλεοτιδικοίπολυμορφισμοί (single-nucleotide polymorphism/SNP), οι οποίοι αποτελούν μικρές αλλαγές στην αλληλουχία του DNAόπου η μία αντικαθίσταται με μία άλλη σε συγκεκριμένες θέσεις του γονιδιώματος. Τα SNPs δεν επηρεάζουν άμεσα την υγεία του ατόμου, αλλά την επηρεάζουν έμμεσα καθώς οι αλλαγές αυτές μπορεί να οδηγήσουν στην προδιάθεση πολυπαραγοντικών ασθενειών όπως είναι οι νευρολογικές ασθένειες. [36]

1.5. MicroT-CNN

Ο MicroT-CNN αποτελεί έναν αλγόριθμο, και πιο συγκεκριμένα
ένα νευρωνικό δίκτυο, πρόβλεψης στόχων miRNA. Σύμφωνα με τα ευρήματα πειραμάτων υψηλής απόδοσης, όπως HITS-CLIP και PAR-CLIP, υποστηρίζεται η ύπαρξη κανονικών και μη κανονικών θέσεων πρόσδεσης miRNA τόσο
στην 3΄αμετάφραστη περιοχή (3΄ untranslatedregion/3΄UTR) όσο και
στις κωδικές αλληλουχίες (codingsequences/CDS). Για αυτόν τον λόγο, ο MicroT-CNN έχει εκπαιδευτεί ξεχωριστά για την 3΄UTR και την CDS, χρησιμοποιώντας πληροφορίες σχετικά με την αλληλουχία, την αναδίπλωση του RNA, την δημιουργία υβριδίου και την συντήρηση. Η σύγκριση του με άλλους γνωστούς αλγόριθμους πρόβλεψης στόχων, όπως microT-CDS, Targetscan v7, MBStar, και miRAW, έδειξε μεγαλύτερα ποσοστά ευαισθησίας για τον ίδιο αριθμό προβλέψεων.

1.6. Στόχος εργασίας

Την δεδομένη χρονική στιγμή η Pubmed διαθέτει 34 εκατομμύρια εγγραφές, πράγμα το οποίο καθιστά δύσκολη και χρονοβόρα την ανάκτηση πληροφοριών από αυτή χειροκίνητα. Έτσι λοιπόν, η αυτοματοποιημένη εξαγωγή πολύτιμης πληροφορίας είναι η γρηγορότερη και πιο αποδοτική λύση στον όλο και αυξανόμενο όγκο της βιβλιογραφίας.

Για να μπορέσει, λοιπόν, να βρεθεί η σχέση μεταλλάξεων που αφορούν microRNA και γονίδια με νευρολογικές και ψυχιατρικές ασθένειες, όπως η νόσος Αλτσχάιμερ, η νόσος του Πάρκινσον και η Σχιζοφρένια μέσω της χρήσης εργαλείων πρόβλεψης στόχων, ήταν απαραίτητη η ανάκτηση σχετικής πληροφορίας από την διαθέσιμη αρθρογραφία. Για αυτόν τον λόγο αναπτύχθηκε ένα σύστημα αυτόματης αναζήτησης στο Entrez και φιλτραρίσματος των άρθρων που ανακτήθηκαν, έτσι ώστε να μελετηθεί η προς επιβεβαίωση από εργαλεία πρόβλεψης στόχων πληροφορία.

Κεφάλαιο 2: Μέθοδοι

2.1. Ερώτημα (query) της αναζήτησης

Η αναζήτηση στο Entrez υποστηρίζει σύνθετες αναζητήσεις οι οποίες αποτελούνται από τους όρους της αναζήτησης, τις ετικέτες και τους λογικούς τελεστές Η΄(OR), ΚΑΙ (AND) και ΟΧΙ (NOT). Η αναζήτηση των επιστημονικών άρθρων έγινε σύμφωνα με τρία χαρακτηριστικά τα οποία είναι απαραίτητα να αναφέρονται μέσα στο άρθρο και αποτελούνται από μία εκ των τριών νευρολογικών και ψυχιατρικών ασθενειών, μία μετάλλαξη και ένα microRNA.
Η αναζήτηση στο Entrez έγινε μέσω του παρακάτω ερωτήματος (query) στο σύστημα:

```
'("alzheimer"[Title/Abstract] OR "alzheimer s"[Title/Abstract] OR
"parkinson"[Title/Abstract] OR "schizophrenia"[Title/Abstract])
AND ("variant"[Title/Abstract] OR "mutation"[Title/Abstract] OR
"polymorphism"[Title/Abstract]  OR "snp"[Title/Abstract] OR
"variation"[Title/Abstract] OR "mutant"[Title/Abstract]) AND
("microrna"[Title/Abstract] OR "mirna"[Title/Abstract]) NOT
Review[ptyp]'
```

Για κάθε χαρακτηριστικό χρησιμοποιήθηκαν εναλλακτικά ονόματα που μπορούν να βρεθούν στην βιβλιογραφία και τα οποία εισήχθησαν μεταξύ του λογικού τελεστή OR, καθώς είναι απαραίτητο να εμφανίζεται στο άρθρο τουλάχιστον μία φορά αυτός ο όρος. Για την εύρεση των νευρολογικών και ψυχιατρικών ασθενειών χρησιμοποιήθηκαν οι όροι ***alzheimer***, ***alzheimer s***, ***parkinson***και ***schizophrenia***. Για την εύρεση της μετάλλαξης χρησιμοποιήθηκαν οι εναλλακτικοί όροι ***variant***, ***mutation***, ***polymorphism***, ***snp***, ***variation***, ***mutant***, ενώ για την εύρεση των microRNAs οι ***microrna*** και ***mirna***. Οι εναλλακτικοί όροι ομαδοποιήθηκαν βάσει του εκάστοτε χαρακτηριστικού (νευρολογική ή ψυχιατρική ασθένεια, μετάλλαξη, microRNA) και οι ομάδες μεταξύ τους χωρίστηκαν με τον λογικό τελεστή AND, διότι απαιτείται να βρεθούν και τα τρία χαρακτηριστικά. Για να γίνει περισσότερο ακριβής η αναζήτηση κάθε όρος αναζητήθηκε για το αν υπάρχει στον τίτλο ή το abstract του άρθρου. Αυτό επιτεύχθηκε μέσω της ετικέτας ***[Title/Abstract]***, που κατά κανόνα τοποθετείται μετά τον όρο. Επιπροσθέτως, τα τελικά δεδομένα που θα συλλεχθούν από τα άρθρα θέλουμε να έχουν βρεθεί μέσω κάποιας μελέτης, π.χ. μελέτη περιπτώσεων (case-controlstudy). Γι΄αυτόν τον λόγο στο query προστέθηκε ο όρος ***NOTReview*** με ετικέτα ***[ptyp]*** , η οποία αναφέρεται στον τύπο της δημοσίευσης (publication type) και τον οποίο περιορίζουμε στο να μην είναι δημοσίευση ανασκόπησης (review), και συνεπώς να μας επιστραφούν οι δημοσιεύσεις κατά τις οποίες υλοποιήθηκε μία μελέτη.

2.2. Αυτόματη αναζήτηση στο Entrez μέσω της Python

Για την ανάκτηση των δεδομένων από το Entrezχρησιμοποιήθηκε το πακέτο **Bio.Entrez** της Python, το οποίο δίνει πρόσβαση στο NCBI μέσω του WWW (World WideWeb). Τα βήματα για την ανάκτηση των δεδομένων ξεκινάνε με την εισαγωγή του πακέτου στο πρόγραμμα με την εντολή **from *Bio* import *Entrez*** και έπειτα τον ορισμό της μεταβλητής **email** του πακέτου, η οποία είναι απαραίτητη για την πρόσβαση στο Entrez, καθώς είναι αναγνωριστικό του χρήστη που έχει πρόσβαση μέσω του κώδικα και δεν έχει κάποια προκαθορισμένη τιμή. Έπειτα καλούμε την συνάρτηση **esearch()** με ορίσματα την βάση δεδομένων που θέλουμε να αναζητήσουμε στο Entrez η οποία ορίστηκε να είναι η PubMed, τον μέγιστο αριθμό αποτελεσμάτων που θέλουμε να επιστραφούν, ο οποίος επιλέχθηκε να είναι ίσος με το 100.000 έτσι ώστε να επιστραφούν όσο το δυνατόν περισσότερα αποτελέσματα και τέλος το όρισμα *term* μέσα στο οποίο καταχωρήθηκε το query. Η **esearch()** αναζητά το *term* που της έχει δοθεί στην βάση δεδομένων που της έχει δοθεί και κάνει ανάκτηση των βασικών κλειδιών τους (Primary IDs). Τα αποτελέσματα αυτής της συνάρτησης επιστρέφονται πάντα σε μορφή XML (Extensible Markup Language) και εισάγονται ως είσοδο στην συνάρτηση **read()**η οποία κάνει την ανάλυση των δεδομένων (parse). Ησυνάρτηση **read()** με την σειρά της επιστρέφει μία πολυεπίπεδη δομή δεδομένων αποτελούμενη από λίστες της Python της μορφής:

{Count: ΄΄, RetMax: ΄΄ , RetStart: ΄΄, IdList:[], TranslationSet: [], TranslationStack: [], QueryTranslation: ΄΄}

Στην συνέχεια καλείται η συνάρτηση **efetch()** η οποία ανακτά τα δεδομένα που της ζητήθηκαν βάσει ορισμάτων. Πιο συγκεκριμένα, ως πρώτο όρισμα έχουμε το ***db*** το οποίο αφορά την βάση δεδομένων από την οποία θα ανακτηθούν τα δεδομένα και ορίστηκε να είναι η PubMed. Το δεύτερο όρισμα είναι το ***resetmode*** το οποίο ορίστηκε ως *xml* και αφορά την μορφή με την οποία θα επιστραφούν τα δεδομένα, ενώ το επόμενο όρισμα αφορά το id που θα αναζητηθεί στην βάση δεδομένων που ορίστηκε και το οποίο είναι ίσο με *record['IdList']*, έτσι ώστε η συνάρτηση **efetch()** να αναζητήσει τα βασικά κλειδιά που επιστράφηκαν από την **esearch()** μέσω της **read()**. Τέλος το όρισμα ***rettype*** ορίστηκε ως *full* καθώς θέλουμε να μας επιστραφεί η αναφορά (citation) και η σύνοψη (abstract) εξίσου. Τα τελικά αποτελέσματα αποθηκεύτηκαν στο αρχείο *articles.xml*.

Παρακάτω παρατίθεται ο κώδικας που περιγράφηκε παραπάνω:

```
from Bio import Entrez

#indentifing the connected user
Entrez.email = 'janedoe@mail.com'

Entrez.tool = 'Demoscript'
#searching for the query in Entrez
query= '("alzheimer"[Title/Abstract] OR "alzheimer
s"[Title/Abstract] OR "parkinson"[Title/Abstract] OR
"schizophrenia"[Title/Abstract]) AND
("variant"[Title/Abstract] OR "mutation"[Title/Abstract]
OR "polymorphism"[Title/Abstract]  OR
"snp"[Title/Abstract] OR "variation"[Title/Abstract] OR
"mutant"[Title/Abstract]) AND ("microrna"[Title/Abstract]
OR "mirna"[Title/Abstract]) NOT Review[ptyp]'
info =
Entrez.esearch(db="pubmed",retmax=100000, term=query)
#parsing the XML data
record = Entrez.read(info)

#retriving records in XML format
fetch = Entrez.efetch(db='pubmed',
                    resetmode='xml',
                    id= record['IdList'],
                    rettype= 'full')

#writing records in XML file
with open('articles.xml', 'wb') as f:
   f.write(fetch.read())
```

2.3. Συλλογή βασικών πληροφοριών των δημοσιεύσεων

Όπως αναφέρθηκε παραπάνω, έγινε αναζήτηση του query που κατασκευάστηκε στο Entrez και τα δεδομένα που επιστράφηκαν σε μορφή XML βρίσκονται εντός του αρχείου *articles.xml*. Για δική μας διευκόλυνση δημιουργήθηκε ένα αρχείο τύπου *.xlsx* το οποίο περιέχει τα εξής μεταδεδομένα: Pubmed ID, τίτλος δημοσίευσης, τίτλος περιοδικού που δημοσιεύτηκε, συγγραφείς, έτος δημοσίευσης, σύνοψη δημοσίευσης. Για να συλλεχθούν αυτά τα μεταδεδομένα από το XML αρχείο χρησιμοποιήθηκε το πακέτο bs4.Beautifulsoup το οποίο εξάγει δεδομένα από ένα XMLαρχείο βάσει του αναλυτή (parser) που έχει επιλεχθεί και ο οποίος παρέχει ιδιωματικούς τρόπους αναζήτησης. Στην δεδομένη περίπτωση επιλέχθηκε ο lxml parser. Με την χρήση την συνάρτησης **findAll()** η οποία αναζητά στο αρχείο το όρισμα ‘pubmedarticle’ επιστρέφονται όλες οι δημοσιεύσεις μία προς μία, καθώς κάθε άρθρο περικλείεται από τις ετικέτες (tags) <*pubmedarticle*> και <*/pubmedarticle*>. Για την εξαγωγή των δεδομένων χρησιμοποιήθηκε ο τελεστής ‘.’ της Python ο οποίος μας επιτρέπει την πρόσβαση σε ένα χαρακτηριστικό (attribute) ενός αντικειμένου (object).

Παρακάτω παρουσιάζεται ενδεικτικά ένα κομμάτι της ανάλυσης
που περιγράφηκε παραπάνω:

```
#opening and parsing the xml file
with open('articles.xml', "r") as xml_file:
   soup = BeautifulSoup(xml_file,  'lxml')

#collecting all the articles
all = soup.findAll('pubmedarticle')

#finding the information for every article
for article in all:

   #searcing for pubmed id
   pmid = article.pmid.text

   # searcing for article's title
   title = article.articletitle.text
```

Για την εγγραφή των δεδομένων σε αρχείο τύπου .xlsx έγινε χρήση της βιβλιοθήκης **xlsxwriter** της Python.

2.4. Ανάλυση Αναγνώρισης ονομάτων-οντοτήτων (Named-entity recognition)

2.4.1. Τι είναι η Ανάλυση Αναγνώρισης ονομάτων-οντοτήτων

Παρά το γεγονός ότι το ερώτημα που έγινε στο Entrezήταν κατασκευασμένο να βρίσκει και τα τρία ζητούμενα χαρακτηριστικά, αυτό δεν εξασφαλίζει οτι όλες οι δημοσιεύσεις που επιστράφηκαν περιέχουν αυτή την πληροφορία. Η εξασφάλιση του οτι κάθε δημοσίευση περιλαμβάνει τους όρους που μας ενδιαφέρουν μπορεί να επιτευχθεί μέσω της διαδικασίας Αναγνώρισης ονομάτων-οντοτήτων (Named-entity recognition). Η Αναγνώριση ονομάτων-οντοτήτων είναι μία διαδικασία που έχει ως στόχο τον εντοπισμό και την ταξινόμηση επώνυμων οντοτήτων σε προκαθορισμένες ομάδες που αναφέρονται μέσα σε ένα κείμενο. Για να γίνει αυτή η ανάλυση θα πρέπει τα δεδομένα να είναι σε μορφή απλού κειμένου.

2.4.2. Αναζήτηση στην PMC

Σε αυτό το σημείο είναι απαραίτητο να θυμηθούμε το είδος των δεδομένων που υπάρχουν διαθέσιμα στην PubMed. Η PubMed αποτελεί μία βιβλιογραφική βάση δεδομένων η οποία περιλαμβάνει μόνο τις αναφορές και τις συνόψεις κάθε καταχώρησης. Είναι προφανές οτι στα δεδομένα που έχουν συλλεχθεί κατά την ανάλυση που περιγράφηκε στην ενότητα 2.2 δεν μπορεί να εφαρμοστεί η διαδικασία Αναγνώρισης ονομάτων-οντοτήτων, καθώς θα ήταν πιο εύλογο αυτή να εφαρμοστεί σε ολόκληρο το κείμενο μίας δημοσίευσης. Από την άλλη πλευρά
η PubMed Central (PMC) είναι ένα ψηφιακό αρχείο της βιβλιογραφίας που αφορά τις βιοϊατρικέςεπιστήμες και επιστήμες της ζωής. Στην PMC βρίσκεται διαθέσιμο το

σύνολο του κειμένου της εκάστοτε δημοσίευσης. Έτσι, έγινε αναζήτηση στο αρχείο *articles.xml*των άρθρων που έχουν διαθέσιμα PMCids. Η εύρεση των άρθρων αυτών έγινε με διαδικασία παρόμοια αυτής που ακολουθήθηκε στην ενότητα 2.3 και τα δεδομένα καταγράφηκαν στο αρχείο pmc_articles.xml. Τα PMCidsαναζητήθηκαν με τον εξής τρόπο:

```
pmcidlist =
article.find_all("articleid", attrs={"idtype": "pmc"})
```

Τέλος, στο αρχείο pmc_articles.xml υπολογίστηκαν ο αριθμός των άρθρων με διαθέσιμο PMCid, αναζητώντας την ετικέτα <*body*>.

2.4.3. Ανάλυση Αναγνώρισης ονομάτων-οντοτήτων με την χρήση του πακέτου scispaCyτης Python

Για την αναγνώριση ονομάτων-οντοτήτων χρησιμοποιήθηκε το πακέτο scispaCy της Python, το οποίο περιλαμβάνει μοντέλα της spaCy , μιας βιβλιοθήκης για επεξεργασία φυσικής γλώσσας, με στόχο την επεξεργασία βιοϊατρικών κειμένων. Πιο συγκεκριμένα χρησιμοποιήθηκαν δύο μοντέλα αναγνώρισης ονομάτων-οντοτήτων τα οποία παρουσιάζονται στον παρακάτω πίνακα:

Όνομα	Σύνολο εκπαίδευσης	Τύπος οντοτήτων
en_ner_bionlp13cg_md	BIONLP13CG	AMINO_ACID, ANATOMICAL_SYSTEM, CANCER, CELL, CELLULAR_COMPONENT, DEVELOPING_ANATOMICAL_STRUCTURE, GENE_OR_GENE_PRODUCT, IMMATERIAL_ANATOMICAL_ENTITY, MULTI-TISSUE_STRUCTURE, ORGAN, ORGANISM, ORGANISM_SUBDIVISION, ORGANISM_SUBSTANCE, PATHOLOGICAL_FORMATION, SIMPLE_CHEMICAL, TISSUE
en_ner_bc5cdr_md	BC5CDR	DISEASE, CHEMICAL

Πίνακας 1: *Τα μοντέλα του sciscpaCy, τα σύνολα στα οποία εκπαιδεύτηκαν και οι κλάσεις που κατηγοριοποιούν τις οντότητες*.

Το μοντέλο ***en_ner_bionlp13cg_md*** χρησιμοποιήθηκε για την εύρεση γονιδίων και γονιδιακών προϊόντων, καθώς επίσης και πολυμορφισμών με την μορφή rsid. Οποιαδήποτε από αυτές τις οντότητες αναμένεται να αναγράφουν στην κλάση GENE_OR_GENE_PRODUCT. Από την άλλη πλευρά, το μοντέλο ***en_ner_bc5cdr_md*** χρησιμοποιήθηκε για την αναγνώριση των ασθενειών, οι οποίες αναμένεται να κατηγοριοποιηθούν στην κλάση DISEASE.

Ως είσοδος στα μοντέλα επιλέχθηκε το κείμενο από τα άρθρα της PMC από το αρχείο pmc_articles.xml. Πιο συγκεκριμένα, έγινε αναζήτηση στο xml αρχείο των ετικετών *<p>*, έτσι ώστε να απομονωθεί το κείμενο. Έπειτα, αυτό το κείμενο χρησιμοποιήθηκε ως είσοδος στην ανάλυση από τα μοντέλα του πακέτου scispaCy.

Τα αποτελέσματα της παραπάνω ανάλυσης αποθηκεύτηκαν στο αρχείο Entity_bi το οποίο είναι αποτελούμενο από τις στήλες *ID,* στην οποία εγγράφεται το PMC id του άρθρου, την *Entity,* στην οποία αναγράφεται η λέξη που έχει κατηγοριοποιηθεί και την *Class,* στην οποία αναφέρεται η ομάδα που κατηγοριοποιήθηκε η λέξη.

ID,Entity,Class
8869484,SYT11,GENE_OR_GENE_PRODUCT
8866491,Alzheimer,DISEASE
8866491,miR-298,GENE_OR_GENE_PRODUCT
8866491,cDNA,CELLULAR_COMPONENT

Εικόνα 8: *παράδειγμα εγγραφής στο αρχείο Entinty_bi.csv*

Τέλος, επιλέχθηκε μία πλήρη σωλήνωση (pipeline) της βιβλιοθήκης spaCy για επεξεργασία βιοϊατρικών δεδομένωνμε στόχο την δημιουργία ενός αρχείου με όλες τις προτάσεις των δημοσιεύσεων καθώς επίσης και των οντοτήτων που βρέθηκαν εντός των προτάσεων, με απώτερο σκοπό το φιλτράρισμα και την εύρεση προτάσεων που περιέχουν τουλάχιστον δύο από την οντότητες που μας ενδιαφέρουν. Η σωλήνωση (pipeline) παίρνει ως είσοδο μία πρόταση και έπειτα από μία επεξεργασία φυσικής γλώσσας επιστρέφει μία λίστα με οντότητες βιοϊατρικού ενδιαφέροντος που βρέθηκαν εντός των προτάσεων. Για την τμηματοποίηση του κειμένουπου επιλέχθηκε από το κάθε άρθρο (μέσω της απομόνωνσηςτου κειμένου εντός των ετικετών *<p>)* σε προτάσεις χρησιμοποιήθηκε η βιβλιοθήκη nltk (Natural LanguageToolkit) της python. Τα αποτελέσματα αποθηκεύτηκαν στο αρχείο meta_info.csv το οποίο αποτελείται από τις στήλες *PMCID,* στην οποία εγγράφεται το PMC id του άρθρου, την *TITLE* για τον τίτλο του άρθρου, την *SETENCE,* στην οποία αποθηκεύεται η εκάστοτε πρόταση και την *DOCENTS,* στην οποία περιλαμβάνεται η λίστα οντοτήτων που επιστράφηκε από την ανάλυση.

PMCID,TITLE,SENTENCE,DOCENTS
8866491, Effects of microRNA-298 on APP and BACE1 translation differ according to cell type and 3´-UTR variation, "In this study, we demonstrated that miR-298 significantly reduced APP and BACE1 levels in human astrocytes at both protein and mRNA levels.", "(study, miR-298, reduced, APP, BACE1, levels, human, astrocytes, protein, mRNA levels)"

Εικόνα 9: παράδειγμα εγγραφής στο αρχείο meta_info.csv

2.5. Φιλτράρισμα αποτελεσμάτων

Για την τελική εξαγωγή των άρθρων με την απαιτούμενη πληροφορία κατασκευάστηκε ένα φίλτρο το οποίο θα ελέγχει αν μία δημοσίευση περιέχει όλες τις απαραίτητες οντότητες, δηλαδή αν περιέχει κάποιο rsid (αποτελεί μία χαρακτηριστική αναφορά σε μία συστάδα μονονουκλεοτιδικώνπολυμορφισμών), ένα γονίδιο (gene), ένα microRNA καθώς επίσης και αν αναφέρεται σε μία εκ των τριών νευρολογικών και ψυχιατρικών ασθενειών. Έπειτα θα πρέπει να ελεγχθεί αν υπάρχει τουλάχιστον μία πρόταση εντός αυτής της δημοσίευσης που να αναφέρει τουλάχιστον δύο από τις παραπάνω οντότητες.
Το φίλτρο λαμβάνει ως είσοδο τα αρχεία *Entity_bi.csv*και *meta_info.csv* τα οποία δημιουργήθηκαν κατά την ανάλυση της υποενότητας 2.4.3. Αρχικά το φίλτρο ελέγχει από το αρχείο *Entity_bi.csv* κάθε id με την σειρά για το αν πληροί τις προϋποθέσεις. Έτσι, κατασκευάστηκαν τρεις συνθήκες ελέγχου, μία για κάθε όρο που αφορά το γονίδιο, το microRNA ή κάποια από τις ασθένειες, εντός των οποίων αν η συνθήκη είναι αληθής τότε η αντίστοιχη σημαία (flag) γίνεται αληθής. Έπειτα υπάρχει άλλη μία συνθήκη ελέγχου κατά την οποία ελέγχεται αν τα τρία παραπάνω flags είναι αληθή και αν ναι, τότε το id της δημοσίευσης που ελέγχεται καταχωρείται εντός της λίστας *id_list[]*. Η εύρεση του microRNA και της ασθένειας έγιναν με την βοήθεια κανονικών εκφράσεων (regular expression), τα οποία αναζητήθηκαν σε κάθε γραμμή της στήλης Entity του αρχείου *Entity_bi.csv*. Παρακάτω αναφέρονται οι κανονικές εκφράσεις που χρησιμοποιήθηκαν για την εύρεση κάθε όρου:

Όρος	Κανονική έκφραση
microRNA	**mir|let-**
ασθένεια	**schizophre|alzheim|dementia|senile|presenile|parkin|paralysisagitans |familial|onset|/^scz$/\|/^sd$/|/^ad$/\|/^atd$/|/^fad$/**

Πίνακας 2: *Οι κανονικές εκφράσεις που χρησιμοποιήθηκαν για την εύρεση των miRNA και των ασθενειών εντός των αρχείων*

Ελέγχθηκε επιπλέον αν έχει βρεθεί κάποιο γονίδιο εντός των όρων που βρέθηκαν για κάθε δημοσίευση. Η ταυτοποίηση των γονιδίων έγινε με την βοήθεια ενός αρχείου που περιλαμβάνει τα ονόματα και τους συμβολισμούς για όλα τα γονίδια τα οποία έχουν βρεθεί στον άνθρωπο και το οποίο ανακτήθηκε από την βάση δεδομένων HGNC (HUGO Gene Nomenclature Committee) που έχει ως στόχο την ανάθεση ενός μοναδικού ονόματος και συμβολισμού για κάθε ανθρώπινο γονίδιο. Κάθε γραμμή της στήλης Entity του csvαρχείου συγκρίθηκε με την σειρά της με όλους τους συμβολισμούς στο αρχείο της HGNC έτσι ώστε να βρεθούν τα γονίδια.

Στην συνέχεια γίνεται ένας δεύτερος έλεγχος στο αρχείο meta_info.csv και μόνο για τις δημοσιεύσεις που πέρασαν τον πρώτο έλεγχο, δηλαδή μόνο για τις δημοσιεύσεις των οποίων τα id βρίσκονται εντός της λίστας id_list[]. Σε αυτόν τον έλεγχο ελέγχεται αν εντός της κάθε λίστας που είναι αποθηκευμένη στην στήλη DOCENTS του αρχείου αναφέρεται κάποιο rsid, microRNA ή ασθένεια. Για να είναι ητελική συνθήκη ελέγχου αληθής είναι απαραίτητο στην λίστα να υπάρχει τουλάχιστον ένα rsid και έπειτα ένας τουλάχιστον όρος που να αναφέρεται σε κάποιο microRNA ή και κάποια ασθένεια. Η εύρεση το rsid έγινε με την βοήθεια της κανονικής έκφρασης *rs[0-9]+*. Αν η τελική συνθήκη είναι αληθής τότε επιστρέφεται η πρόταση η οποία ακολουθεί τους περιορισμούς. Ως τελικό αποτέλεσμα λαμβάνουμε ένα αρχείο τύπου txt, το finalresults.txt στο οποίο αναγράφονται μόνο οι δημοσιεύσεις που πέρασαν με επιτυχία το φίλτρο και για τις οποίες αναφέρονται το PMC id, ο τίτλος καθώς επίσης και μία περίληψη η οποία αποτελείται από τις προτάσεις που επιστράφηκαν από το δεύτερο βήμα του φίλτρου.

ID:8024493
Title: In silico Analysis of Polymorphisms in microRNAs Deregulated in Alzheimer Disease
A variant in miR-101-2 (rs138231885) has the most negative ΔG (−3.1) with a high expression rate of mature miRNA, while another SNPs (rs188892061) in miR-328 has the most ΔG (5.8) with a low expression rate of mature miRNA.
has declared that the level of miRNA-1229-3p which has been confirmed to regulate post-transcriptionally SORL1, is increased in the rs2291418 pre-miRNA-1229 variant. According to the evidence, COX2, an inductive enzyme which catalyzes the conversion of arachidonic acid to prostanoids, plays a vital role in the plasticity of neurons and memory acquisition It seems that variant rs138231885, which is predicted to increase the expression of the mature form of miR-101-2 (performing biological function), is likely to be associated to disease risk.(2014) has found that miR-125 and its SNPs (rs12976445) have a negative relationship with Graves' disease (GD) and Hashimoto's disease (HD); moreover, not only the expression of miRNA-125 but also its efficacy has been reduced. They also found rs2291418 in the miR-1229 precursor to being significantly associated with Alzheimer's disease, consistent with our data.

Εικόνα 10: *παράδειγμα εγγραφής στο αρχείο finalresult.txt*

2.6.Χειροκίνητος σχολιασμός των αποτελεσμάτων (Manual Curation)

Σε αυτό το σημείο, τα άρθρα του τελικού αρχείου της παραπάνω ανάλυσης μελετήθηκαν κάθε ένα ξεχωριστά με στόχο την δημιουργία ενός καταλόγου με τις βασικές πληροφορίες που συγκεντρώθηκαν από την επιλεγμένη βιβλιογραφία. Αυτή η διαδικασία διεξάχθηκε παρά το γεγονός ότι τα άρθρα φιλτραρίστηκαν υπολογιστικά έτσι ώστε να είναι περισσότερο στοχευμένες οι τελικές πληροφορίες. Οι δημοσιεύσεις που κατέληξαν στον τελικό κατάλογο επιλέχθηκαν με βάση το αν περιέχουν πειραματικά αποδεδειγμένη σχέση ανάμεσα σε γονίδιο-miRNA-πολυμορφισμό-ασθένεια. Θα πρέπει να σημειωθεί οτι από το τελικό φίλτρο πέρασαν και άρθρα τα οποία ανέφεραν ένα miRNA ή έναν πολυμορφισμό ή μία από

τις τρεις ασθένειες ενδιαφέροντος στο κεφάλαιο της Εισαγωγής (Introduction) ή της Συζήτησης (Discussion), χωρίς να συμπεριλαμβάνονται στο πείραμα που διεξάχθηκε, καθώς αναγνωρίστηκαν από το μοντέλο ως λέξεις κλειδιά. Τα συγκεκριμένα άρθρα παρά το γεγονός οτι έγιναν αποδεκτά από το τελικό φίλτρο δεν συμπεριλήφθηκαν στον τελικό κατάλογο καθώς η σχέση των -απλά- αναφερόμενων όρων δεν μελετήθηκε πειραματικά. Ωστόσο, επιλέχθηκε να συμπεριληφθούν τα άρθρα που δεν συμπεριλαμβάνουν στο πείραμα κάποιο γονίδιο και ελέγχουν την σχέση μόνο μεταξύ των miRNA-πολυμορφισμός-ασθένεια.

Ο κατάλογος αποτελείται από στήλες στις οποίες αναγράφονται οι απαραίτητες πληροφορίες έτσι ώστε να διεξαχθούν τα τελικά συμπεράσματα. Πιο συγκεκριμένα, δομείται από τις στήλες:

• **Pmid**: το αναγνωριστικό του άρθρου στην Pubmed

• **Article's title**: ο τίτλος του άρθρου

• **Journal's title**: ο τίτλος του περιοδικού δημοσίευσης

• **Author's name**: το όνομα του πρώτου συγγραφέα

• **Year**: το έτος δημοσίευσης του άρθρου

• **Abstract**: η περίληψη του άρθρου

• **miRNA**: το όνομα του miRNA σε μορφή συμβατή με την miRbase

• **Gene**: το σύμβολο του γονιδίου όπως αυτό αναφέρεται στην HGNC. Αν στο άρθρο δεν αναφέρεται κάποιο γονίδιο τότε στο κελί αναγράφεται μία παύλα.

• **Association**: παίρνει τις τιμές TRUE/FALSE αντίστοιχα για το αν αποδεικνύεται η σχέση ή μη του πολυμορφισμού με την έκφραση των miRNA ή των γονιδίων στην ανάπτυξη της ασθένειας.

• **Risk**: παίρνει τις τιμές INCREASED/DECREASED/NO ASSOCIATION αντίστοιχα για το αν αυξάνεται ή μειώνεται το ρίσκο εμφάνισης της ασθένειας. Εφόσον το παραπάνω πεδίο έχει την τιμή FALSE τότε η τιμή της στήλης Risk θα είναι no association.

• **Gene Biotype**: εδώ σημειώνεται η ομάδα του γονιδίου, δηλαδή αν κωδικοποιεί πρωτεΐνη (proteincoding),
είναι ψευδογονίδιο (pseudogene), αν κωδικοποιεί lncRNA ή ncRNA κ.α. Αν στο άρθρο δεν αναφέρεται κάποιο γονίδιο τότε στο κελί αναγράφεται μία παύλα.

• **Varian**t: το id του πολυμορφισμού όπως αναγράφεται στην dbSNP.

• **Variant region**: η περιοχή που εντοπίζεται ο πολυμορφισμός στο miRNA ή το γονίδιο. Μαζί με την περιοχή αναγράφεται και το αν βρίσκεται στο miRNA ή το γονίδιο, π.χ. μία πιθανή τιμή μπορεί να είναι Gene-3΄UTR.

• **Disease**: η ασθένεια για την οποία διεξάχθηκε το πείραμα

• **Population**: ο πληθυσμός για τον οποίο διεξάχθηκε το πείραμα

• **Study**: ο τύπος της έρευνας ή του πειράματος που διεξάχθηκε

• **Cell line**: η κυτταρική σειρά στην οποία εκφράζεται το γονίδιο ή το miRNA

• **Sentences**: οι προτάσεις που ανακτήθηκαν στην ανάλυση που περιγράφηκε στην ενότητα 2.5.

• **Comments**: επιπλέον σχόλια

Να σημειωθεί ότι τα δεδομένα που συμπληρώθηκαν στις στήλες Article’s title, Journal’s title, Author’s name, Year και abstract είναι αυτά που ανακτήθηκαν από την ανάλυση της ενότητας 2.3.

2.7. Υπολογιστική εύρεση στόχων με την χρήση του αλγορίθμου MicroT-CNN

2.7.1. Χρήση των πληροφορίων του χειροκίνητου σχολιασμού ως είσοδο του αλγορίθμου

Κατά την διάρκεια του χειροκίνητου σχολιασμού των άρθρων ανακτήθηκαν πληροφορίες σχετικά με την αλληλεπίδραση ενός το γονιδίου και ενός miRNA, και πιο συγκεκριμένα την αλληλεπίδραση mRNA::miRNA όταν υπάρχει κάποια μετάλλαξη στο miRNA ή το γονίδιο, και κατά πόσο η γονιδιακή ρύθμιση με την ύπαρξη της μετάλλαξης επηρεάζει την εμφάνιση της ασθένειας ενδιαφέροντος. Αυτή η πληροφορία συλλέχθηκε με στόχο την είσοδο της στον αλγόριθμο MicroT-CNN, έτσι ώστε να επιβεβαιωθούν υπολογιστικά οι αλληλεπιδράσεις με την ύπαρξη ή μη της μετάλλαξης.

2.7.2. Εκτέλεση του αλγόριθμου για το γονιδίωμα με ενσωματωμένους πολυμορφισμούς

Στο δεύτερο βήμα δημιουργήθηκαν εκ νέου τα 3 παραπάνω αρχεία αλλά αυτή την φορά προσαρμόστηκαν οι μεταλλάξεις. Σε αυτό το σημείο να θυμηθούμε ότι ο microT-CNN βρίσκει αλληλεπιδράσεις μόνο για τις 3΄αμετάφραστη περιοχή (3΄ Untranslated region/ 3’ UTR) και την κωδικήαλληλουχία (Coding sequence/CDS), συνεπώς από το τελικό αρχείο των δεδομένων του σχολιασμού των άρθρων επιλέχθηκαν μόνο μεταλλάξεις που αφορούν γονίδια και βρίσκονται εντός της 3’UTR ή της CDS. Έπειτα από αυτές τις μεταλλάξεις επιλέχθηκαν μόνο αυτές που αφορούσαν μονοκλουτεοτιδικούς πολυμορφισμούς (Single-nucleotidepolymorphisms). Για να γίνει η αντικατάσταση των βάσεων που αφορούν τον μονονουκλεοτιδικό πολυμορφισμό θα πρέπει να βρεθεί η θέση του μέσα στο εξόνιο. Έτσι λοιπόν, επιλέχθηκαν οι αλληλουχίες των εξονίων μαζί με την πληροφορία για την αρχή και το τέλος τους. Επίσης, από την dbSNP ανακτήθηκαν οι θέσεις των SNPs καθώς επίσης και σηματοδοτικές αλληλουχίες (flanks) που βρίσκονται πριν από την βάση που αλλάζει στον πολυμορφισμό. Θα πρέπει να σημειωθεί οτι η αλληλουχία των εξονίων που βρίσκονται στην αρνητική αλυσίδα είναι ανεστραμένη και συμπληρωματική σε σχέση με την αλληλουχία των σηματοδοτικών αλληλουχιών που δίνονται στην dbSNP βάση του Genome Reference Consortium Human Build 38(GRCh38). Συνεπώς, τόσο η σηματοδοτική αλληλουχία πριν την βάση που αλλάζει όσο και η ίδια η βάση στις αλληλουχίες των εξονίων στην αρνητική αλυσίδα θα είναι συμπληρωματικές βάσει της συμπληρωματικότητας των βάσεων με τα δεδομένα που αναφέρονται στην dbSNP. Άρα, αν στην dbSNP αναφέρεται η αλλαγή της βάσης T με την βάση G, τότε στο αρχείο με τις αλληλουχίες των γονιδίων θα γίνει η αλλαγή από A σε C. Επιπλέον, η θέση της βάσης της μετάλλαξης υπολογίστηκε ως *end position* –

*variant position*καθώς η αλληλουχία εκτός από συμπληρωματική είναι και ανεστραμμένη. Από την άλλη πλευρά, για τα εξόνια στην θετική αλυσίδα η θέση του πολυμορφισμού υπολογίστηκε ως *variant position – start position*, καθώς η αλυσίδα είχε την ίδια «φορά» με αυτή της αλυσίδας αναφοράς. Παραδείγματος χάριν, αν ο πολυμορφισμός αναφέρεται στην dbSNP ότι βρίσκεται στην θέση 50.000.030 και η αρχή του γονιδίου είναι στο 50.000.000 ενώ το τέλος του στο 50.000.100, τότε αν το εξόνιο αφορά της θετική αλυσίδα η αλλαγή της βάσης θα γίνει στην θέση 50.000.030 – 50.000.000 = 30 της αλληλουχίας του εξονίου , ενώ αν αφορά την αρνητική αλυσίδα η αλλαγή θα γίνει στην θέση 50.000.100 – 50.000.030 = 70 της αλληλουχίας του εξονίου. Τέλος, ακολούθησε εκτέλεση του αλγόριθμου για το γονιδίωμα που ενσωματώθηκαν οι πολυμορφισμοί.

Index	Variant	Gene	miRNA	Variant type	Alleles	Chromosome	Position	Disease
1	rs7143400	FERMT2	hsa-miR-4504	SNV	A>C	14	52858173	AD
2	rs9909	NUP160	hsa-miR-1185-1-3p	SNV	C>G	11	47778223	AD
3	rs12720208	FGF20	hsa-miR-433	SNV	G>A	8	16992890	PD
4	rs550067317	EFNB2	hsa-miR-137	SNV	T>G	13	106491980	SCZ
5	rs1060120	H3-3B	hsa-miR-616	SNV	C>T	17	75776919	SCZ
6	rs9722	S100B	hsa-miR-6827-3p	SNV	G>A	21	46599326	AD
7	rs896	VIPR1	hsa-miR-525-5p	SNV	T>C	3	42536843	AD
8	rs113810300	NCSTN	hsa-miR-186	SNV	T>G	1	160358894	AD
9	rs1130354	DRD2	hsa-miR-326	SNV	G>C	11	113410198	SCZ
10	rs1045881	NRXN1	1274a / hsa-miR-339-5p	SNV	C>T	2	49921834	SCZ

Πίνακας 3: *Ο πίνακας περιλαμβάνει τα γονίδια και τα microRNAs που αποτέλεσαν είσοδο στον αλγόριθμο, καθώς επίσης και τους μονονουκλεοτιδικούςπολυμορφισμούς που ενσωματώθηκαν στις αλληλουχίες κατά την δεύτερη εκτέλεση. Να σημειωθεί οτι οι αλλαγές των αλληλόμορφων (alleles) καθώς και οι θέσεις των αλλαγών αναγράφονται όπως στην dbSNP.*

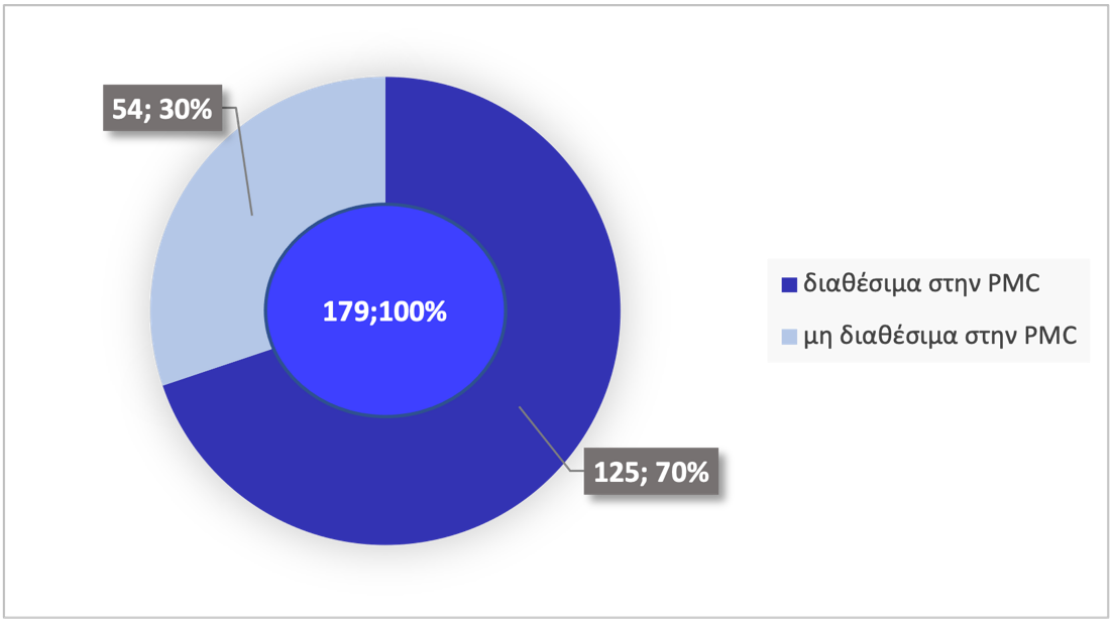
Κεφάλαιο 3: Αποτελέσματα

3.1.Αποτελέσματα αναζήτησης στην Entrez και φιλτραρίσματος

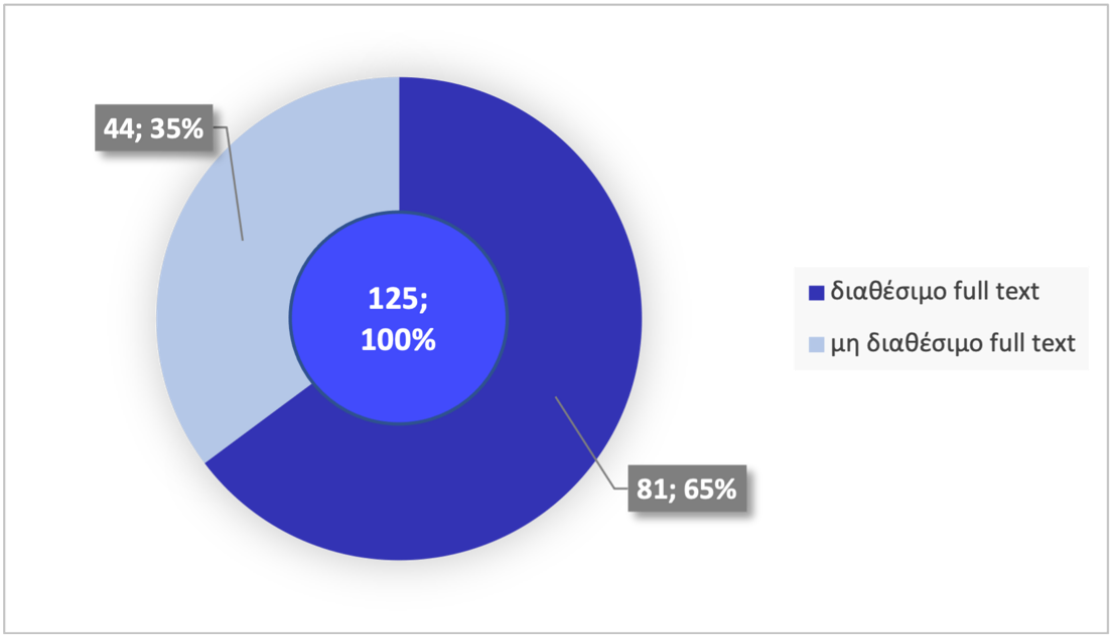
Με την αυτόματη αναζήτηση στο entrez με το query που δόθηκε επιστράφηκαν 179 δημοσιεύσεις. Στην συνέχεια επιλέχθηκαν τα άρθρα που ήταν διαθέσιμα στην PMC και ο αριθμός αυτών ανέρχεται στα 125. Έπειτα, ελέγχθηκε πως τα 81 από τα 125 άρθρα είχαν διαθέσιμο πλήρες κείμενο που ήταν απαραίτητο για την ανάλυση ονομάτων-οντοτήτων.

Η ανάλυση ονομάτων-οντοτήτων έγινε με την χρήση 2 μοντέλων και ενός pipeline του πακέτου scispaCy της Python. Τα αποτελέσματα της ανάλυσης με τα δύο μοντέλα εγγράφτηκαν στο αρχείο τύπου CSV, *το entity_bi.csv*, και συνολικά για τις 81 δημοσιεύσεις τα μοντέλα αναγνώρισαν 30.497 οντότητες. Παρόμοια, τα αποτελέσματα της ανάλυσης με το pipeline εγγράφτηκαν στο αρχείο *meta_info.csv*, το οποίο αποτελείται από 22.077 εγγραφές, οι οποίες αντιπροσωπεύουν τις προτάσεις που βρέθηκαν στις 81 δημοσιεύσεις.

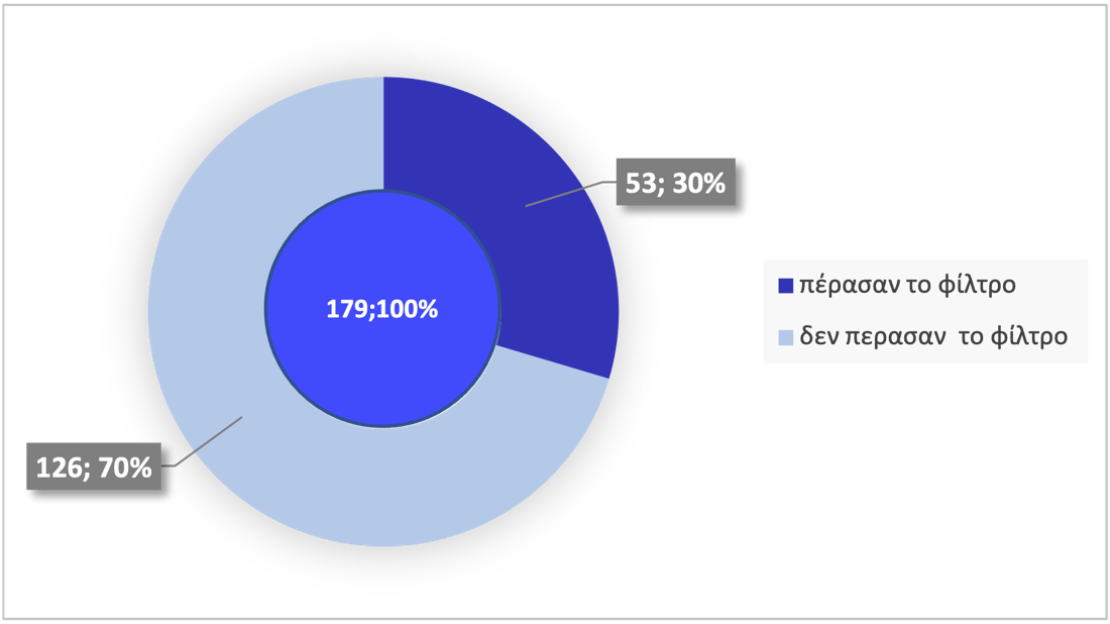
Το τελικό φιλτράρισμα έγινε πάνω στα δύο παραπάνω αρχεία με τα κριτήρια: 1. το κείμενο της δημοσίευσης να περιλαμβάνει και του 4 όρους ενδιαφέροντος, miRNA, γονίδιο, ασθένεια, πολυμορφισμός, και 2. να περιέχουν έστω και μία πρόταση με έστω δύο εκ των παραπάνω όρων. Το τελικό αρχείο τύπου .txt, το *finalresults.txt* ήταν αποτελούμενο από 53 δημοσιεύσεις.



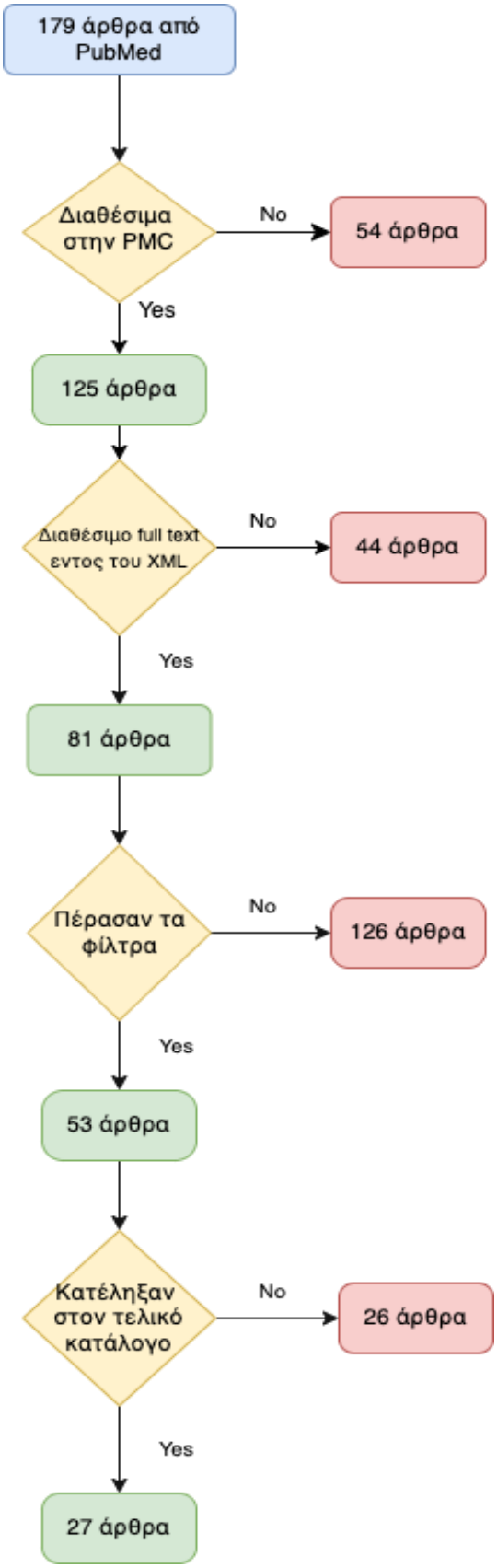
Εικόνα 11: *Το ποσοστό των άρθρων που ήταν διαθέσιμα στην PMC*



Εικόνα 12: *Το ποσοστό των άρθρων με PMCid που είχαν διαθέσιμο κείμενο εντός του XML*



Εικόνα 13: *Το ποσοστό των άρθρων που πέρασαν και τα δύο φίλτρα*



Εικόνα 14: *Διάγραμμα ροής του αριθμού των άρθρων που επιλέχθηκαν σε κάθε βήμα*

3.2. Αποτελέσματα χειροκίνητου σχολιασμού

Κατά τον χειροκίνητο σχολιασμό, 27 από τις δημοσιεύσεις είχαν την απαραίτητη πληροφορία και κατέληξαν στον τελικό κατάλογο. Παρακάτω παρατίθονταιπλήρως τα αποτελέσματα που συγκεντρώθηκαν από τον χειροκίνητο σχολιασμό. Να σημειωθεί ότι για την καλύτερη παρουσίαση των πινάκων παραλήφθηκε η στήλη που περιελάμβανε τις προτάσεις που συγκεντρώθηκαν από το δεύτερο φίλτρο του συστήματος που αναπτύχθηκε στην python, όπως αναφέρθηκε στην ενότητα 2.5.

index	ID	Article Title	Journal Titl	Author	Year	Abstract
1	34607594	Association study b	BMC psychiatr	Li Dai Liu Lin	2021	Schizophrenia is a polygenic disease;
2	34607594	Association study b	BMC psychiatr	Li Dai Liu Lin	2021	Schizophrenia is a polygenic disease;
3	34607594	Association study b	BMC psychiatr	Li Dai Liu Lin	2021	Schizophrenia is a polygenic disease;
4	27215977	miRNA-dependent	Alzheimer's re	Delay Grenier	2016	A growing body of evidence suggests that
5	27215977	miRNA-dependent	Alzheimer's re	Delay Grenier	2016	A growing body of evidence suggests that
6	26856603	Primate-specific mi	Aging	Zhang Lu Liu	2016	Alzheimer's disease (AD) is a serious
7	26856603	Primate-specific mi	Aging	Zhang Lu Liu	2016	Alzheimer's disease (AD) is a serious
8	26856603	Primate-specific mi	Aging	Zhang Lu Liu	2016	Alzheimer's disease (AD) is a serious
9	18252210	Variation in the miR	American jour	Wang van der	2008	Parkinson disease (PD) is a common
10	33708240	Sex-Specific Associ	Frontiers in ge	Yin Luo Peng	2021	Objective: To investigate the effects of microRNA-
11	33708240	Sex-Specific Associ	Frontiers in ge	Yin Luo Peng	2021	Objective: To investigate the effects of microRNA-
12	26095531	A single nucleotide	Molecular mec	Zhang Wang	2015	Alzheimer's disease (AD) is a progressive
13	27650867	MicroRNA-137 Inhi	EBioMedicine	Wu Zhang Nie	2016	MicroRNAs (miRNAs) are a class of
14	26609515	Association betwee	FEBS open bio	Sun Yu Zhu S	2015	Schizophrenia is one of the most common mental
15	26609515	Association betwee	FEBS open bio	Sun Yu Zhu S	2015	Schizophrenia is one of the most common mental
16	31991575	Characterization of	International jo	Imperatore Th	2020	Alzheimer's disease (AD), the most common
17	34938351	Polymorphisms in M	Computationa	Zhang Li Ding	2021	Schizophrenia (SCZ) is a common and complex
18	34938351	Polymorphisms in M	Computational	Zhang Li Ding	2021	Schizophrenia (SCZ) is a common and complex
19	34938351	Polymorphisms in M	Computationa	Zhang Li Ding	2021	Schizophrenia (SCZ) is a common and complex
20	25250332	Association of a miR	BioMed resear	Ma Yin Fu Luo	2014	Both genome wide association study
21	31198450	The Relationship be	Open access M	Cao Yang Kar	2019	To investigate the relationship between the
22	31198450	The Relationship be	Open access M	Cao Yang Kar	2019	To investigate the relationship between the

Πίνακας 4: *Τα μεταδεδομένα που συλλέχθηκαν από τα άρθρα (μέρος α)*

index	ID	Article Title	Journal Titl	Author	Year	Abstract
23	31198450	The Relationship be	Open access M	Cao Yang Kar	2019	To investigate the relationship between the
24	24586483	A functional polymo	PloS one	Cui Li Ma War	2014	miR146a is well known for its regulatory role in
25	24586483	A functional polymo	PloS one	Cui Li Ma War	2014	miR146a is well known for its regulatory role in
26	29529098	Validation of a micro	PloS one	Manley Morea	2018	Despite much progress, few genetic findings for
27	33982673	S100B gene polym	Aging	Wang Zhou Y	2021	To examine the role of S100B in genetic
28	27328823	Genome-wide ident	Scientific repor	Ghanbari Ikrar	2016	MicroRNAs (miRNAs) serve as key post-
29	26429811	A GWAS SNP for Sc	Schizophrenia	Warburton Bre	2016	Single nucleotide polymorphisms (SNPs)
30	25390694	Single nucleotide p	PloS one	Paladini Porci	2014	Complex immune and neurodegenerative
31	25100943	MicroRNAs targetin	Frontiers in mo	Delay Dorval	2014	Despite the growing number of genome-wide
32	25100943	MicroRNAs targetin	Frontiers in mo	Delay Dorval	2014	Despite the growing number of genome-wide
33	29142105	The NDE1 genomic	Open biology	Bradshaw Ukl	2017	Genetic studies of familial schizophrenia in
34	29142105	The NDE1 genomic	Open biology	Bradshaw Ukl	2017	Genetic studies of familial schizophrenia in
35	26836412	Polymorphisms in M	Translational p	Wright Gupta	2016	Evidence suggests that microRNA-137 (miR-137)
36	24675081	MicroRNA-9 and mi	The Journal of	Shi Leites He	2014	The human dopamine receptor D2 (DRD2) has
37	26257337	Polymorphisms in M	Scientific repo	Zhang Fan Wa	2015	Disturbances in glutamate signaling
38	22850735	Impact of a microRN	Neuropsychop	Heather C Wha	2012	A recent 'mega-analysis' combining genome-wide
39	21687627	Neurexin-1 and fror	PLoS One	Aristotle N Voi	2011	Background: Structural variation in the neurexin-
40	21687627	Neurexin-1 and fror	PLoS One	Aristotle N Voi	2011	Background: Structural variation in the neurexin-
41	30026708	Association of MIR1	Frontiers in psy	Kandratsenka	2018	MicroRNA-137 (miRNA-137; miR-137) is one of
42	30914946	Serum miRNAs Exp	Frontiers in ag	Agostini Manc	2019	MicroRNAs (miRNAs) are small non-coding
43	30914946	Serum miRNAs Exp	Frontiers in ag	Agostini Manc	2019	MicroRNAs (miRNAs) are small non-coding
44	30914946	Serum miRNAs Exp	Frontiers in ag	Agostini Manc	2019	MicroRNAs (miRNAs) are small non-coding

Πίνακας 5: Τα μεταδεδομένα που συλλέχθηκαν από τα άρθρα (μέρος β)

index	miRNA	Gene	Associatio	Risk	Gene Biotype	Variant	Variant region	Disease	Population	Study	Cell line	comments
1	hsa-miR-219a-1	HLA-DPB2	TRUE	increased	pseudogene	rs383711	miRNA-2kb upstream	SCZ	Han Chinese	case-control study of 455 cases, 445 controls	neurons	association with rs2021722 located within the HLA locus
2	hsa-miR-219a-1	HLA-DPB2	TRUE	increased	pseudogene	rs439205	miRNA-2kb upstream	SCZ	Han Chinese	case-control study of 455 cases, 445 controls	neurons	association with rs2021722 located within the HLA locus
3	hsa-miR-219a-1	HLA-DPB2	TRUE	increased	pseudogene	rs421446	miRNA-2kb upstream	SCZ	Han Chinese	case-control study of 455 cases, 445 controls	neurons	association with rs2021722 located within the HLA locus
4	hsa-miR-4504	FERMT2	TRUE	decreased	protein coding	rs7143400	Gene-3' UTR	AD	-	in silico analysis 220 genes of hg19in 29 AD-associated	-	-
5	hsa-miR-1185-1-3p	NUP160	TRUE	decreased	protein coding	rs9909	Gene-3' UTR	AD	-	in silico analysis 220 genes of hg19 in 29 AD-associated	-	-
6	hsa-miR-603	E2F1	TRUE	decreased	protein coding	rs11014002	miRNA-precursor	AD	North American	case-control study of 365 NC, 772 MCI,301 AD	-	-
7	hsa-miR-603	LRP1	TRUE	decreased	protein coding	rs11014002	miRNA-precursor	AD	North American	case-control study of 365 NC, 772 MCI,301 AD	-	-
8	hsa-miR-603	LRPAP1	TRUE	decreased	protein coding	rs11014002	miRNA-precursor	AD	North American	case-control study of 365 NC, 772 MCI,301 AD	-	mild cognitive impairment subjects (MCI) and
9	hsa-miR-433	FGF20	TRUE	increased	protein coding	rs12720206	Gene-3' UTR	PD	-	case-control study of 1089 cases, 1165 controls	brain-spinal cord	-
10	hsa-miR-137	-	TRUE	increased	-	rs1198588	miRNA-41.1 kb upstream	SCZ	Han Chinese	case-control study of 1116 cases, 1039 healthy controls	-	female-specific association of MIR137
11	hsa-miR-137	-	TRUE	increased	-	rs2660304	miRNA-promoter	SCZ	Han Chinese	case-control study of 1116 cases, 1039 healthy controls	-	female-specific association of MIR137
12	hsa-miR-146a	TLR2	TRUE	increased	protein coding	rs2910164	miRNA-precursor	AD	Han Chinese	case-control study of 103 cases, 206 controls	-	-
13	hsa-miR-137	EFNB2	TRUE	increased	protein coding	rs5500673	Gene-3' UTR	SCZ	Chinese	case-control study of 44 cases, 44 controls	brain	-
14	hsa-mir-219-1	-	TRUE	increased	-	rs107822	miRNA-2kb upstream	SCZ	Chinese	case-control study of 589 patients, 622 controls	brain tissue	-
15	hsa-miR-137	-	FALSE	o associatio	-	rs1625579	miRNA-host gene	SCZ	Chinese	case-control study of 589 patients, 622 controls	brain tissue	-
16	hsa-miR-1229	SORL1	TRUE	increased	protein coding	rs2291418	miRNA-precursor	SCZ	-	biophysical characterization experiments with chemically produced	neurons	AD-associated SNP rs2291418 pre-miR-1229
17	hsa-miR-107	-	TRUE	increased	-	rs2296616	miRNA-2kb upstream	SCZ	Chinese	case-control study of 150 cases,150 controls	brain tissue	MECT lowers the mirna levels
18	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-host gene	SCZ	Chinese	case-control study of 150 cases,150 controls	-	-
19	hsa-miR-34a	-	TRUE	increased	-	rs6577555	miRNA-2kb upstream	SCZ	Chinese	case-control study of 150 cases,150 controls	-	-
20	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-primary transcript	SCZ	Han Chinese	case-control study of 611 cases, 628 controls	-	-
21	hsa-miR-605	FGA	FALSE	o associatio	protein coding	rs2043556	miRNA-precursor	SCZ	Han Chinese	case-control study of 513 cases, 513 controls	-	-
22	hsa-miR-499a	FGA	FALSE	o associatio	protein coding	rs4909237	miRNA-precursor	SCZ	Han Chinese	case-control study of 513 cases, 513 controls	-	-

Πίνακας 6: *Τα δεδομένα που συγκεντρώθηκαν κατά τον χειροκίνητο σχολιασμό (μέρος α)*

index	miRNA	Gene	Associatio	Risk	Gene Biotype	Variant	Variant region	Disease	Population	Study	Cell line	comments
23	hsa-miR-499b	FGA	FALSE	o associatio	protein coding	rs4909237	miRNA-precursor	SCZ	Han Chinese	case-control study of 513 cases, 513 controls	-	-
24	hsa-miR146a	IL-1b	TRUE	increased	protein coding	rs5709532	miRNA-stem region	AD	Chinese	case-control study of 292 cases, 300 controls	-	-
25	hsa-miR146a	IL-6	TRUE	increased	protein coding	rs5709532	miRNA-stem region	AD	Chinese	case-control study of 292 cases, 300 controls	-	-
26	hsa-miR-616	H3F3B	TRUE	increased	protein coding	rs1060120	Gene-3' UTR	SCZ	Celtic,German	case-control study of 85 cases, 40 controls	brain	-
27	hsa-miR-6827-3p	S100B	TRUE	increased	protein coding	rs9722	Gene-3' UTR	AD	Han Chinese	case-control study of 280 cases, 400 controls	neurons	-
28	hsa-miR-1229-3p	SORL1	TRUE	increased	protein coding	rs2291418	miRNA-precursor	AD	-	in-silico analysis, in-vitro miRNA express	human brain	-
29	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-promoter	SCZ	-	haplotype tagging-SNP analysis, reporter gene analysis	brain	-
30	hsa-miR-525-5p	VIPR1	TRUE	increased	protein coding	rs896	Gene-3' UTR	AD	-	differential expression analysis	-	-
31	hsa-miR-186	NCSTN	TRUE	increased	protein coding	rs1418494	Gene-3' UTR	AD	Canadian	case-control study of 511 cases, 631 co	-	-
32	hsa-miR-186	NCSTN	TRUE	increased	protein coding	rs1138103	Gene-3' UTR	AD	Canadian	case-control study of 511 cases, 631 co	brain	-
33	hsa-miR-484	NDE1	TRUE	increased	protein coding	rs1050162	Gene- intron	SCZ	Finnish	cohort study of 458 families	-	-
34	hsa-miR-484	NDE1	TRUE	increased	protein coding	rs2242549	Gene-intron	SCZ	Finnish	cohort study of 458 families	-	-
35	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-host gene	SCZ	Caucasian	case-control study of 89 cases, 132 controls	-	from Mind Clinical Imaging Consortium (MCIC) shared
36	hsa-miR-326	DRD2	TRUE	increased	protein coding	rs1130354	Gene-3' UTR	SCZ	European	1870 cases and 2002 controls	-	-
37	hsa-miR-219	-	TRUE	increased		rs107822	miRNA-2kb upstream	SCZ	Han Chinese	case-control study of 1041 cases, 953 c	-	-
38	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-host gene	SCZ	Scotish	case-control study of 44 SCZ high-risk subjects ,81 controls	-	-
39	hsa-miR-1274a	NRXN1	TRUE	increased	protein coding	rs1045881	Gene-3' UTR	SCZ	-	genotyping 53 healthy subjects in 11 SNPs of the gene, in silico analysis	brain-frontal cortex	-
40	hsa-miR-339-5p	NRXN1	TRUE	increased	protein coding	rs1045881	Gene-3' UTR	SCZ	-	genotyping 53 healthy subjects in 11 SNPs of the gene, in silico analysis	brain-frontal cortex	-
41	hsa-miR-137	-	TRUE	increased	-	rs1625579	miRNA-host gene	SCZ	Belarusian	case-control study of 150 SCZ patients,	adult brain	-
42	miR181a-5p	SNAP25	TRUE	increased	protein coding	rs363050	Gene- intron	AD	Italian	case-control study 2 cohorts. Study coho	neurons	mild cognitive impairment (MCI)
43	miR-23a-3p	SNAP25	TRUE	increased	protein coding	rs363050	Gene- intron	AD	Italian	case-control study 2 cohorts. Study coho	neurons	mild cognitive impairment (MCI)
44	miR-27b-3p	SNAP25	TRUE	increased	protein coding	rs363050	Gene- intron	AD	Italian	case-control study 2 cohorts. Study coho	neurons	mild cognitive impairment (MCI)

Πίνακας 7: *Τα δεδομένα που συγκεντρώθηκαν κατά τον χειροκίνητο σχολιασμό (μέρος β)*

3.3. Αποτελέσματα του αλγόριθμου MicroT-CNN

Από τα δεδομένα που συγκεντρώθηκαν κατά τον χειροκίνητο σχολιασμό, οι 10 πολυμορφισμοί αφορούσαν την 3΄UTR περιοχή του γονιδίου και καμία την CDS, ενώ όλοι ήταν μονονουκλεοτιδικοί. Η σχέση των 10 γονιδίων που περιείχαν αυτούς τους πολυμορφισμούς ελέγχθηκε για 10 microRNAs. Να σημειωθεί ότι ο πολυμορφισμός με idrs1045881αφορούσε τα microRNAs hsa-miR-1274a και hsa-miR-339-5p, αλλά το hsa-miR-1274a δεν συμπεριλήφθηκε στο γονιδίωμα αναφοράς καθώς δεν υπήρχε για αυτό διαθέσιμη εγγραφή στην BioMart. Συνεπώς, για αυτόν τον λόγο ο αριθμός των microRNAs για τα οποία ελέγχθηκε αλληλεπίδραση είναι 10 και όχι 11.

Τα αποτελέσματα του αλγόριθμου συγκρίθηκαν πριν και μετά την ενσωμάτωση του πολυμορφισμού. Πιο συγκεκριμένα, ελέγχθηκε το mre score του αλγόριθμου με κατώφλι το 0.5, δηλαδή θεωρούμε οτι αν το σκορ του αλγόριθμου είναι μεγαλύτερο του 0.5 τότε ο αλγόριθμος βρίσκει αλληλεπίδραση για το miRNA και μετάγραφο που του δόθηκαν. Όσο μεγαλύτερο είναι το mre σκορ με τόση μεγαλύτερη εμπιστοσύνη μπορούμε να πούμε ότι το miRNAκαι το μετάγραφο του γονιδίου αλληλέπιδρουν. Παραδείγματος χάριν, η βιβλιογραφία για το γονίδιο FERMT2 αναφέρει ότι παρουσία του μονονουκλεοτιδικούπολυμορφισμού το γονίδιο χάνει την σύνδεση του με το hsa-miR-4504 αυξάνοντας έτσι την έκφραση του γονιδίου. Αυτό επιβεβαιώνεται από τα αποτελέσματα του αλγόριθμου, καθώς χωρίς τον πολυμορφισμό το mre score είναι ίσο με 0.868 (> 0.5) και με τον πολυμορφισμό είναι ίσο με 0.3474 (<0.5). Ένα ακόμα παράδειγμα είναι αυτό του γονιδίου DRD2, για το οποίο η βιβλιογραφία αναφέρει οτι χάνεται η αλληλεπίδραση με το hsa-miR-326 όταν το μετάγραφο του γονιδίου έχει τον πολυμορφισμό, γεγονός που αποδεικνύεται και από τα αποτελέσματα του αλγόριθμου (mre_score_without_snp = 0.3119 <0.5; mre_score_with_snp = 0.6375 > 0.5). Για τις εγγραφές που το mre score τους ήταν πάνω από 0.5 με ή χωρίς τον πολυμορφισμό, ακόμα και αν υπήρχε διαφορά στην τιμή του, δεν θεωρήθηκε ότι υπάρχει ή όχι αλληλεπίδραση. Για παράδειγμα, η εγγραφή για το γονίδιο NRXN1 σημειώθηκε πως δεν επιβεβαιώνει την υπόθεση της βιβλιογραφίας οτι ο πολυμορφισμός καταργεί την σύνδεση με το hsa-miR-339-5p, καθώς και οι δύο τιμές είναι μεγαλύτερες από το κατώφλι ανεξάρτητα από το αν εντοπίζεται αύξηση (mre_score_without_snp = 0.7824 <0.5; mre_score_with_snp= 0.9712 > 0.5). Επιπλέον, θα πρέπει να σημειωθεί ότι δόθηκε περισσότερη βαρύτητα στο mre score έναντι του score, διότι το mre score είναι πιο ευαίσθητο σε μικρότερες αλλαγές καθώς αφορά συγκεκριμένα την περιοχή της σύνδεσης. Να σημειωθεί σε αυτό το σημείο οτι ελέγχθηκαν τα αποτελέσματα του μεγαλύτερου σε μήκος μετάγραφου, καθώς η βιβλιογραφία δεν αναφερόταν σε κάποιο συγκεκριμένο ενώ οι πολυμορφισμοί αφορούσαν πολλαπλά μετάγραφα. Παρακάτω παρατίθενται τα αναλυτικά αποτελέσματα του αλγόριθμου.

index	transcript id	mre score with snp	score with snp	binding type with snp	mirna	mre score without snp	score without snp	binfing type without snp	Confirmed
1	ENSG00000073712@FERMT2 @ENST00000395631	0.3474	0.719384	non-canonical	hsa-miR-4504	0.868	0.719384	canonical	TRUE
2	ENSG00000030066@NUP160 @ENST00000694866	0.749	0.687564	non-canonical	hsa-miR-1185-1-3p	0.0	0.0	-	TRUE
3	ENSG00000078579@FGF20 @ENST00000180166	0.137	0.266380	non-canonical	hsa-miR-433	0.1368	0.266380	canoninal	FALSE
4	ENSG00000125266@EFNB2 @ENST00000646441	0.0	0.0	-	hsa-miR-137	0.886	0.787916	non-canonical	TRUE
5	ENSG00000132475@H3-3B@ENST00000254810	0.0777	0.494449	non-canonical	hsa-miR-616	0.0656	0.494449	non-canonical	FALSE
6	ENSG00000160307@S100B@ ENST00000367071	0.8766	0.728801	non-canonical	hsa-miR-6827-3p	0.9238	0.735159	canoninal	FALSE
7	ENSG00000114812@VIPR1@ ENST00000543411	0.1498	0.543078	non-canonical	hsa-miR-525-5p	0.1489	0.543078	non-canonical	FALSE
8	ENSG00000162736@NCSTN @ENST00000699528	0.9643	0.784498	canonical	hsa-miR-186	0.0	0.0	-	TRUE
9	ENSG00000179915@NRXN1 @ENST00000401669	0.7824	0.847052	non-canonical	hsa-miR-339-5p	0.9712	0.853880	canonical	FALSE
10	ENSG00000149295@DRD2@ ENST00000362072	0.3119	0.577126	non-canonical	hsa-miR-326	0.577126	0.6375	canonical	TRUE

Πίνακας 8: *Τα τελικά αποτελέσματα των δύο εκτελέσεων του MicroT-CNN*

Κεφάλαιο 4: Συζήτηση

Τα microRNAs λαμβάνουν σημαντικό ρόλο στην γονιδιακή ρύθμιση με αποτέλεσμα να ερευνώνται ως πιθανοί θεραπευτικοί στόχοι και βιοδείκτες για μορφές καρκίνου αλλά και πολλές ασθένειες, συμπεριλαμβανομένου και νευρολογικές ασθένειες. Όλο και περισσότερες έρευνες δείχνουν πως μεταλλάξεις του αφορούν την αλληλεπίδραση miRNA::mRNA παίζουν σημαντικό ρόλο στην ανάπτυξη πολλών νευρολογικών και ψυχιατρικών παθήσεων. Οι επιδράσεις των μεταλλάξεων αυτών πάνω στην δράση των microRNA με τα γονίδια είναι απαραίτητο να κατανοηθούν σε βάθος με στόχο την διάγνωση και θεραπεία τέτοιων ασθενειών. Για αυτόν τον λόγο, οποιαδήποτε μελέτη πάνω σε αυτά τα ζητήματα είναι εξίσου σημαντική στο να κατανοήσουμε την ρύθμιση των microRNAs και τα μονοπάτια συμμετοχής τους που οδηγούν στην εμφάνιση των ασθενειών.

Όσον αφορά την διαχείριση κειμένου που διεξάχθηκε στην παρούσα εργασία, θα μπορούσαν να γίνουν περαιτέρω βελτιώσεις έτσι ώστε να συλλεχθεί περισσότερη και πιο στοχευμένη πληροφορία. Πιο συγκεκριμένα, για να απομονωθούν οι δημοσιεύσεις που περιέχουν με μεγαλύτερη εμπιστοσύνη την απαραίτητη πληροφορία, με στόχο το μεγαλύτερο ποσοστό αυτών να καταλήξουν στον τελικό κατάλογο των δεδομένων, θα μπορούσε να προστεθεί ένα επιπλέον κριτήριο που θα αφορούσε τον εντοπισμό των όρων απαραίτητα μέσα στις ενότητες των Μεθόδων, Αποτελεσμάτων ή Περίληψης μίας δημοσίευσης. Αυτό θα μπορούσε να επιτευχθεί με τον διαχωρισμό του κύριου κειμένου σε ενότητες μέσω των αντίστοιχων ετικετών στο xml αρχείο, με αποτέλεσμα το κείμενο που θα εισαχθεί στα μοντέλα της ανάλυσης ονομάτων-οντοτήτων να αφορούν συγκεκριμένα κεφάλαια της δημοσίευσης με στόχο να εξεταστεί το επιπλέον κριτήριο. Επιπλέον, θα μπορούσαν να χρησιμοποιηθούν εναλλακτικοί τρόποι αναγνώρισης και κατηγοριοποίησης των οντοτήτων, έναντι της ανάλυσης ονομάτων-οντοτήτων. Μία από αυτές, είναι η εξαγωγή σχέσεων μεταξύ οντοτήτων (Relation Extraction), μία διαδικασία πρόβλεψης των ιδιοτήτων και των σχέσεων μεταξύ των οντοτήτων μίας πρότασης.

Στην παρούσα εργασία, τα δεδομένα που συλλέχθηκαν από την διαχείριση των δημοσιεύσεων επιβεβαιώθηκαν υπολογιστικά μέσω του αλγόριθμου πρόβλεψης στόχωνMicroT-CNN. Ο συγκεκριμένος αλγόριθμος αποτελεί έναν πρωτοποριακό αλγόριθμο, ο οποίος προβλέπει αλληλεπιδράσεις που αφορούν όχι μόνο την 3’UTR αλλά και την CDS περιοχή του γονιδίου. Ωστόσο, για την διεξαγωγή πιο έγκυρων αποτελεσμάτων θα μπορούσαν να χρησιμοποιηθούν και άλλοι αλγόριθμοι πρόβλεψης στόχων, όπως ο TargetScan ή ο MBSTAR, έτσι ώστε να συγκριθούν τα αποτελέσματα του εκάστοτε αλγόριθμου και να διεξαχθεί ένα ενιαίο συμπέρασμα.

Στην παρούσα εργασία, ακολουθήθηκε μία ολοκληρωμένη διαδικασία συγκέντρωσης και ανάλυσης βιολογικών δεδομένων που αφορούσαν την σχέσημετάλλαξης::miRNA::mRNA με νευρολογικές και ψυχιατρικές παθήσεις. Τα στάδια αφορούσαν πρώτα την ανάπτυξη ενός αυτοματοποιημένου συστήματος αναζήτησης σε μεγάλες σχολιαστικές βάσεις δεδομένων και φιλτραρίσματος των άρθρων που επιστράφηκαν, έπειτα τον χειροκίνητο σχολιασμό των άρθρων με στόχο

την συγκέντρωση της πληροφορίας ενδιαφέροντος και τέλος, την υπολογιστική επιβεβαίωση αυτών με την εκτέλεση του αλγόριθμου MicroT-CNN. Οι βελτιώσεις που αναφέρθηκαν παραπάνω θα βοηθήσουν στην συγκέντρωση πιο στοχευμένηςπληροφορίας και συνεπώς την διεξαγωγή καλύτερων αποτελεσμάτων, με στόχο την συμβολή στη γενικότερη μελέτη των microRNAs και την σχέση τους με νευρολογικές και ψυχιατρικές ασθένειες. Στο μέλλον, θα μπορούσε το παραπάνω σύστημα να χρησιμοποιηθεί εναλλακτικά και για άλλες ασθένειες που θεωρείται πως οφείλονται σε μεταλλάξεις που αφορούν microRNAs, και ακολουθώντας την ίδια διαδικασία για διαφορετικά δεδομένα να βρεθούν νέα ευρήματα.

Βιβλιογραφία

[1] J. Soria Lopez, H. Gonzalez and G. Leger, "Alzheimer's Disease," *Handb Clin Neurol,* 2019.

[2] BruceBlaus, "CC BY 3.0," [Online]. Available: https://creativecommons.org/licenses/by/3.0.

[3] N. I. o. Aging, "National Institutes of Health," 2016. [Online]. Available: https://www.flickr.com/photos/nihgov/24524716351.

[4] A. Armstrong, "Risk factors for Alzheimer's disease," *Folia Neuropatho,* 2019.

[5] M. Wang, L. Qin and B. Tang, "MicroRNAs in Alzheimer's Disease," *Front Genet,* 2019.

[6] Alzheimer's Association, "2019 Alzheimer's disease facts and figures," *Alzheimer's and Dementia,* vol. 15, no. 3, pp. 321-387, 2019.

[7] B. JM, "Parkinson's disease: a review," *Front Biosci (Schol Ed),* 2014.

[8] C. S, L. Mus and F. Blandini, "Parkinson's Disease in Women and Men: What's the Difference?," *J Parkinsons Dis,* 2019.

[9] M. G and D. Stewart, "Parkinson's disease – pathology, aetiology and diagnosis," *Reviews in Clinical Gerontology,* 2012.

[10] Tulemo, "CC BY-SA 4.0," [Online]. Available: https://creativecommons.org/licenses/by-sa/4.0.

[11] S. Subramaniam and M. Chesselet, "Mitochondrial dysfunction and oxidative stress in Parkinson's disease," *Prog Neurobiol,* 2013.

[12] L. Kalia and A. Lang, "Parkinson's disease," *Lancet,* 2015.

[13] A. Pickrell and R. Youle, "The roles of PINK1, parkin, and mitochondrial fidelity in Parkinson's disease," *Neuron,* 2015.

[14] A. Ascherio and M. Schwarzschild, "The epidemiology of Parkinson's disease: risk factors and prevention," *Lancet Neurol,* 2016.

[15] A. P. Association, "Task Force on DSM-IV," in *Diagnostic and statistical manual of mental disorders,* American Psychiatric Pub, 2000, p. 299.

[16] M. Picchioni and R. Murray, "Schizophrenia," *BMJ,* 2007.

[17] A. Priol, L. Denis, G. Boulanger, M. Thépaut, M. Geoffray and S. Tordjman, "Detection of Morphological Abnormalities in Schizophrenia: An Important Step to Identify Associated Genetic Disorders or Etiologic Subtypes," *Int J Mol Sci,* 2021.

[18] S. Stilo and R. Murray, "Non-Genetic Factors in Schizophrenia," *Curr Psychiatry Rep,* 2019.

[19] A. Brown, "The environment and susceptibility to schizophrenia," *Prog Neurobiol,* 2011.

[20] J. Vilain, A. Galliot, J. Durand-Roger, M. Leboyer, P. Llorca, F. Schürhoff and A. Szöke, "Les facteurs de risque environnementaux de la schizophrénie [Environmental risk factors for schizophrenia: a review]," *Encephale,* 2013.

[21] J. Richetto and U. Meyer, "Epigenetic Modifications in Schizophrenia and Related Disorders: Molecular Scars of Environmental Exposures and Source of Phenotypic Variability," *Biol Psychiatry,* 2021.

[22] K. Mueser and . S. McGurk, "Schizophrenia," *Lancet,* 2014.

[23] T. Schwartz, S. Sachdeva and S. Stahl, "Glutamate neurocircuitry: theoretical underpinnings in schizophrenia," *Front Pharmacol,* 2012.

[24] W. Carpenter and R. Buchanan, "Schizophrenia," *N Engl J Med,* 1994.

[25] C. Zhuo, G. Li, X. Lin, D. Jiang, Y. Xu, H. Tian, W. Wang and X. Song, "Strategies to solve the reverse inference fallacy in future MRI studies of schizophrenia: a review," *Brain Imaging Behav,* 2021.

[26] BruceBlaus, " CC BY-SA 4.0," [Online]. Available: https://creativecommons.org/licenses/by-sa/4.0.

[27] J. McGrath, S. Saha, D. Chant and J. Welham, "Schizophrenia: a concise overview of incidence, prevalence, and mortality," *Epidemiol Rev,* 2008.

[28] T. Laursen, M. Nordentoft and P. Mortensen, "Excess early mortality in schizophrenia," *Annu Rev Clin Psychol,* 2014.

[29] Kioomars743, "CC BY-SA 4.0," [Online]. Available: https://creativecommons.org/licenses/by-sa/4.0.

[30] Y. Lee, M. Kim, J. Han, K. Yeom, S. Lee, S. Baek and V. Kim , " MicroRNA genes are transcribed by RNA polymerase II," *EMBO J,* 2004.

[31] G. Michlewski and J. Cáceres, "Post-transcriptional control of miRNA biogenesis," *RNA,* 2019.

[32] M. Ha and Kim VN, "Regulation of microRNA biogenesis," *Nat Rev Mol Cell Biol,* 2014.

[33] N. S, "The AGO proteins: an overview," *Biol Chem,* 2018.

[34] H. Kobayashi and Y. Tomari, "RISC assembly: Coordination between small RNAs and Argonaute proteins," *Biochim Biophys Acta,* 2016.

[35] K. Saliminejad, H. Khorram Khorshid , S. Soleymani Fard and S. Ghaffari, "An overview of microRNAs: Biology, functions, therapeutics, and analysis methods," *J Cell Physiol,* 2019.

[36] H. SM, "An overview of microRNAs," *Adv Drug Deliv Rev,* 2015.