



Implicación de los miRNAs en la enfermedad de Parkinson desde RNA-seq en sangre periférica: Enfoque basado en Machine learning

José Javier Herranz Bañón

Tutores:

Juan Antonio Botía Blaya (Universidad de Murcia)

Laura Ibáñez Lladó (Washington Saint Louis University)

Universidad de Murcia, Facultad de Biología
Máster en Bioinformática

2024-07-19

Contenido

| | |
|---|----------|
| Resumen: | 1 |
| Introducción: | 1 |
| Parkinson: | 1 |
| Epidemiología: | 1 |
| Etiología: | 1 |
| Sintomatología: | 2 |
| Funcionamiento: | 2 |
| Mecanismos moleculares: | 2 |
| Base genética: | 4 |
| Micro ARNs: | 4 |
| Biogénesis: | 5 |
| Regulación de la producción de ARN mensajero: | 5 |
| Micro ARNs en sangre o Micro ARNs circulantes: | 6 |
| Materiales: | 6 |
| Software y paquetes: | 6 |
| Datasets: | 7 |
| Métodos: | 7 |
| Selección de micro ARNs mediante modelos predictivos: | 7 |
| Preprocesado: | 7 |
| Criterios de evaluación de modelos: | 7 |
| Precisión balanceada: | 7 |
| Prueba de Mc Nemar: | 8 |
| Tasa de no información: | 8 |
| P-valor: | 8 |
| LASSO: | 8 |
| Parámetros de LASSO: | 8 |
| Algoritmo ranger(bosques aleatorios): | 9 |
| Hiperparámetros del bosque aleatorio: | 9 |
| Análisis de enriquecimiento funcional: | 9 |
| GSEA: | 10 |

| | |
|---|-----------|
| Resultados: | 12 |
| Rendimiento de los modelos: | 12 |
| Resultados del análisis GSEA: | 12 |
| Solapamiento de micro ARNs entre modelos y Prueba exacta de Fisher: | 14 |
| Discusión: | 14 |
| Los modelos: | 14 |
| Micro ARNs usados por el modelo de LASSO con lambda óptima: | 14 |
| hsa-miR-6864-5p: | 15 |
| hsa-miR-1229-5p: | 15 |
| hsa-miR-20b-3p: | 15 |
| hsa-miR-3074-3p: | 15 |
| hsa-miR-3182: | 15 |
| hsa-miR-6726-5p: | 16 |
| hsa-miR-6787-3p: | 16 |
| hsa-miR-4733-5p: | 16 |
| hsa-miR-93-3p: | 16 |
| Análisis GSEA: | 16 |
| Micro ARNs de GSEA más relevantes: | 17 |
| Conclusión: | 17 |
| Bibliografía: | 18 |

Lista de tablas

| | | |
|---|---|----|
| 1 | Valores de las matrices de confusión. | 12 |
| 2 | micro ARNs usados por el modelo de LASSO con lambda menor | 13 |
| 3 | Resultados de GSEA para el modelo de LASSO balanceado con más genes | 13 |

Lista de figuras

| | | |
|---|--|----|
| 1 | Diagrama de flujo de trabajo | 11 |
|---|--|----|

Repositorio de archivos: <https://github.com/jjherranz/Implicaci-n-de-los-miRNAs-en-la-enfermedad-de-Parkinson-desde-RNA-seq-en-sangre-perif-rica.git>

Resumen:

La enfermedad de Parkinson es la segunda enfermedad neurodegenerativa más común, por detrás del Alzheimer. Es una afección cerebral que causa trastornos motrices como la bradicinesia y la rigidez muscular, así como no motrices, como depresión y trastornos del sueño. Su diagnóstico temprano y tratamiento efectivo todavía son un reto a día de hoy. En este trabajo se pretende utilizar el método de machine learning LASSO junto con la expresión de los micro ARNs en plasma de sangre periférica, para encontrar grupos de micro ARNs cuyas funciones e implicaciones en el Parkinson permitan obtener información útil. Para ello, se entrenaron varios modelos con LASSO, intentando mejorar la capacidad discriminatoria en cada intento hasta obtener un conjunto de micro ARNs que diese lugar a una clasificación precisa. Después se intentó obtener información sobre las posibles vías metabólicas en las que participan estos micro ARNs mediante un análisis GSEA. Tras obtener los resultados del modelo con menor cantidad de micro ARNs y analizar los resultados de GSEA, se investigó en la bibliografía disponible los posibles papeles o relaciones de los micro ARNs en el Parkinson.

Introducción:

Parkinson:

La enfermedad del Parkinson consiste en un trastorno neurodegenerativo descrito inicialmente por el doctor James Parkinson en 1817 (Parkinson 2002).

Epidemiología:

El Parkinson es el desorden motor más común y la segunda enfermedad degenerativa del sistema nervioso central más común. El desarrollo de la enfermedad se da usualmente a una edad de entre 65 y 70 años, se considera que la enfermedad afecta al 1% de la población por encima de los 60 años. El desarrollo de la enfermedad anterior a la edad de 40 años es inferior al 5% de los casos de las cohortes basadas en población. Sin embargo existe el desarrollo temprano de la enfermedad por variantes genéticas. Se piensa que los factores genéticos de Parkinson están involucrados en entre el 5-10% de los casos. La enfermedad es ligeramente más frecuente en hombres. La prevalencia está situada en un rango entre 100 y 200 cada 100000 personas y la incidencia alrededor de 15 cada 100000 (Meara and Hobson 2018). La mortalidad no incrementa en pacientes de Parkinson en relación a los individuos no afectados para la primera década desde el diagnóstico de la enfermedad, sin embargo incrementa después (Pinter et al. 2015). Se espera que con el envejecimiento de la población, la prevalencia de Parkinson se incremente de manera importante (Dorsey et al. 2018).

Etiología:

La enfermedad del Parkinson es una enfermedad idiopática para la que, sin embargo, se han identificado factores de riesgo genéticos, teniendo los familiares en primer grado de los pacientes 2 o 3 veces más riesgo de desarrollar la enfermedad comparado con la población general o los controles (Savica, Cannon-Albright, and Pulst 2016).

Se piensa que la mayoría de los casos probablemente tienen una etiología multifactorial, que combina factores genéticos y ambientales. Los factores ambientales conocidos son la exposición a sustancias químicas tóxicas, las heridas de cabeza y factores de estilo de vida. Genéticamente, del 5% a el 10% de los casos son causados por mutaciones identificables en ciertos genes, pero en la mayoría de casos estas están ausentes. Además, las mutaciones más comúnmente asociadas a Parkinson tienen penetrancia incompleta, por lo que el desarrollo de la enfermedad dependerá de otros factores también, ya sean genéticos o ambientales (Simon, Tanner, and Brundin 2020).

Un estudio que compara las tasas de concordancia en gemelos monocigóticos y dicigóticos estima que la heredabilidad de Parkinson es solo el 30%, lo que indica que la mayoría del riesgo está relacionado con otros factores (Goldman et al. 2019).

Sintomatología:

La sintomatología motriz incluye temblores asimétricos progresivos en reposo, bradicinesia, rigidez muscular, deterioro de la postura y el equilibrio, pérdida de movimientos automáticos, cambios en el habla y en la escritura. Sin embargo, se pueden desarrollar otros síntomas no motrices años antes del desarrollo de la enfermedad, como depresión, desorden del sueño REM, estreñimiento y pérdida de olfato (Sung and Nicholas 2013).

Funcionamiento:

Parkinson se define por la pérdida de neuronas dopaminérgicas en la parte compacta de la sustancia negra (sustantia nigra) del mesencéfalo (SNm) con despigmentación, y se asocia a la presencia de los cuerpos de Lewy (CL), que consisten en inclusiones citoplasmáticas intraneuronales, eosinofílicas. También se caracteriza por una patología extendida a otras regiones cerebrales e involucra neuronas no dopaminérgicas (Simon, Tanner, and Brundin 2020).

Los CL están constituidos por un núcleo de hialina y un halo periférico pálido compuesto de más de 90 proteínas. Sus principales componentes son α -sinucleína y ubiquitina. Esta última tiene propensión a plegarse mal, se vuelve insoluble y forma agregados amiloides ricos en láminas beta. Los compuestos intermedios de este proceso de agregación son las formas tóxicas oligoméricas y protofibrilares que interfieren con las funciones mitocondriales, lisosomales y del proteasoma, dañan las membranas biológicas y el citoesqueleto, alteran las funciones sinápticas y causan degeneración neuronal (Balestrino and Schapira 2020).

Mecanismos moleculares:

Agregación de α -sinucleína: La agregación de α -sinucleína es uno de los más importantes y característicos. Esta agregación es una de las hipótesis más importantes que explican la muerte de neuronas nigroestriadas en Parkinson (Maries et al. 2003). La α -sinucleína es una proteína neuronal abundante en terminales presinápticas, bajo condiciones fisiológicas se une a las membranas de las vesículas sinápticas y funciona en la liberación de neurotransmisores (M. Sharma and Burré 2023). Esta puede volverse neurotóxica cuando su monómero forma oligómeros, que se combinan para, en última instancia, formar fibrilas grandes e insolubles (C. Kim and Lee 2008). Se piensa que el declive de las defensas proteolíticas en el cerebro por la edad contribuye a la acumulación de esta proteína (Kaushik and Cuervo 2015). Además la degradación de los sistemas de ubiquitina-proteasoma, autofagia lisosomal y las proteasas extracelulares pueden contribuir también a la acumulación (Dong-Chen et al. 2023).

Estrés oxidativo: El estrés oxidativo también es un proceso importante en Parkinson, ya que con el envejecimiento daña al sistema nervioso central (Hajam et al. 2022). Este proceso ocurre cuando los radicales libres sobrepasan la capacidad de la actividad antioxidante de la célula. Esto genera compuestos citotóxicos que al acumularse provocan el colapso de proteínas, fallo enzimático, rotura de los lípidos y muerte celular en neuronas (Dong-Chen et al. 2023). Esto contribuye a la patogénesis de Parkinson (Dorszewska et al. 2021). La oxidasa de NADPH (NOX) se considera el generador de radicales libres más importante y tiene un papel en el desencadenamiento de desordenes neurodegenerativos (Zilberter, Tabuena, and Zilberter 2023). Por otro lado las mitocondrias también son grandes productoras de radicales libres. Los complejos I y III de la cadena de electrones son donde más radicales libres se producen, siendo los radicales superóxidos los principales (Dong-Chen et al. 2023). Un estrés oxidativo incrementado puede bajar los niveles de lisosomas y dañar el sistema lisosomal de autofagia, provocando la acumulación de α -sinucleína (Navarro-Romero, Montpeyó, and Martínez-Vicente 2020). Otra hipótesis defiende que la dopamina extra citosólica se puede

oxidar para dar lugar a quinonas dopaminérgicas, que modifican la α -sinucleína que exhibe parcialmente autofagia mediada por chaperonas, lo que causa el autoensamblaje de la α -sinucleína (Belarbi et al. 2017). Además los agregados intracelulares de α -sinucleína incrementan el estrés oxidativo mitocondrial (Dong-Chen et al. 2023).

Ferroptosis: Otro mecanismo molecular es la ferroptosis, una muerte celular controlada dependiente del hierro. Es causado por un metabolismo de hierro aberrante y peroxidación lipídica severa. Este mecanismo ha sido relacionado con la muerte de neuronas dopaminérgicas en Parkinson (Do Van et al. 2016). Se ha encontrado que moléculas de hierro coexisten con la α -sinucleína en los cuerpos de Lewy en el cerebro de pacientes de Parkinson (Castellani et al. 2000). Esta última cambia de conformación cuando se une a el hierro, dando lugar a la agregación de la misma (Dong-Chen et al. 2023). Una posible causa de ferroptosis es NOX, que se ha observado que produce estrés oxidativo en microglia activa que causando ferroptosis en neuronas dopaminérgicas (N. Sharma and Nehru 2016). También se encontró que la ceruloplasmina (que tiene actividad ferroxidasa y permite la salida del hierro de las células) tenía una expresión reducida en pacientes de Parkinson, lo que sugiere que la acumulación de hierro tenga un papel en la muerte de neuronas en Parkinson (Dong-Chen et al. 2023).

Disfunción mitocondrial: Se ha observado que la disfunción mitocondrial induce neurodegeneración dopaminérgica y produce radicales libres del oxígeno (N. Bhattacharjee and Borah 2016), y que tiene un papel significativo en Parkinson (Bose and Beal 2016). Hay varios factores que pueden influir en la disfunción mitocondrial. Por ejemplo, los ratones que sobreexpresan α -sinucleína son más susceptibles a las toxinas mitocondriales, lo que podría indicar que una acumulación de esta proteína en la mitocondrial provoca degeneración de neuronas dopaminérgicas (Klivenyi et al. 2006). Se cree que la desregulación de los factores de transcripción y los consecuentes cambios en la biogénesis de la mitocondrial son una de las principales causas de disfunción mitocondrial (Chaturvedi and Beal 2008). Específicamente, el PGC-1 α (transcription factor coactivator peroxisome proliferator activated receptor gamma coactivator-1 α) parece ser un regulador clave de la biogénesis mitocondrial. En ratones knockout de este factor las células dopaminérgicas son más sensibles a MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine), una toxina que provoca un síndrome idéntico a Parkinson. Mientras que la sobreexpresión de este factor de transcripción da lugar a protección contra neurotoxicidad (Dong-Chen et al. 2023). Además existen genes cuyas mutaciones patológicas provocan disfunción mitocondrial, como Parkin, LRRK2, PINK1 y DJ-1, cuyo rol en el Parkinson son explicados en la sección de bases genéticas de más adelante.

Neuroinflamación: La neuroinflamación es una respuesta inmune a elementos patógenos o daños en el tejido, que tiene la función de proteger la parenquima del sistema nervioso central y promover la reparación de tejidos (Köhlker-Frers et al. 2021). Sin embargo, cuando la lesión es severa y las reacciones inmunes son mantenidas prolongadamente, pueden provocar la descompensación del sistema nervioso central, dando lugar a la acentuación de la muerte neuronal, lo que provoca un déficit en la capacidad de mantenimiento celular, la disrupción de la barrera entre la sangre y el cerebro, y alcanzando una respuesta inflamatoria crónica y excesiva (Araújo et al. 2022). La neuroinflamación ha sido reconocida como un factor clave en Parkinson (Grottemeyer et al. 2022), tanto la relación con la respuesta inmune innata como la adaptativa. Sin embargo no está muy claro como ocurre la neuroinflamación crónica en Parkinson (Dong-Chen et al. 2023). En cuanto a la activación de las microglías, existen varios posibles mecanismos por los que se dé lugar a la neuroinflamación en Parkinson. La α -sinucleína es un patrón molecular asociado a daños (DAMPs, danger associated molecular pattern), por lo que cuando entra en las células por a través del receptor similar a toll 2, puede provocar cambios proinflamatorios (Ferreira and Romero-Ramos 2018). Por otro lado, las células moribundas o dañadas pueden liberar DAMPs, IL-1 α , o radicales libres que activen la respuesta inmune al interactuar con receptores de reconocimiento de patrones. La consecuente activación de NLRP3 eleva la síntesis de β , lo que inicia más respuestas inmunes innatas (Dinarello 2018). En modelos animales con neurodegeneración inducida, las microglías eran repolarizadas al fenotipo proinflamatorio M1. Esto provocaba la transcripción de procaspasa-1 e interleukina. Estos procesos forman el inflamasoma NLRP3 que, junto

con la caspasa-1, activan IL-1 β , una proteína proinflamatoria. Además hay otras proteínas proinflamatorias que son liberadas de las células M1 (como iNOS y TNF) que participan en la neurodegeneración en Parkinson (Dong-Chen et al. 2023). TLR-4 también contribuye a la neuroinflamación, demostrado por un estudio knockout que daba lugar a la bajada de los niveles de microglías activadas y la protección contra neurodegeneración de células dopaminérgicas del sistema nervioso central (Noelker et al. 2013). En cuanto a la respuesta inmune adaptativa, varios estudios indican que las subpoblaciones de células T contribuyen a la patofisiología del Parkinson (Sommer et al. 2018). Siendo importantes las células T CD4 y CD8, que se infiltran significativamente en pacientes de Parkinson (Dong-Chen et al. 2023; Galiano-Landeira et al. 2020).

Base genética:

Un gen con importancia patogénica en Parkinson es SNCA, el gen que codifica la alfa-sinucleína. Hay descubiertas mutaciones patogénicas relacionadas con Parkinson para este gen. Estas confieren a la alfa-sinucleína una tendencia mayor a plegarse mal y agregarse. Otras mutaciones afectan a la cantidad de alfa-sinucleína (como duplicaciones, triplicaciones, alteraciones de su expresión o su eliminación), y alteran sus modificaciones post-transcripcionales y/o su interacción con otros orgánulos y sistemas de transporte. La sobre expresión de este gen se ha observado en pacientes con Parkinson idiopático (Balestrino and Schapira 2020).

Otros genes implicados en el Parkinson familiar regulan las funciones mitocondriales. Como PINK1 y Parkin, que tienen roles en las vías de control de calidad mitocondrial. PINK1 es una quinasa de serina/treonina que marca mitocondrias dañadas y activa la vía de la mitofagia mediante el reclutamiento de Parkin. DJ-1 tiene un rol en la regulación del flujo de calcio en la mitocondria, protegiendo a la célula del estrés oxidativo de la actividad de las neuronas dopaminérgicas y de la toxicidad por dopamina (Balestrino and Schapira 2020).

Otro elemento clave en el Parkinson es la disfunción en las vías de eliminación celulares, dado que se han asociado genes relacionados con la autofagia con Parkinson. Mutaciones en LRRK2 interfieren con la autofagia, reduciendo la degradación de alfa-sinucleína, lo que contribuye a su acumulación. En ATP13A2 mutante se provoca disfunción lisosomal y causa el síndrome de Kufor Rakeb, parkinsonismo juvenil. ATP13A2 se encuentra sobre expresado en neuronas dopaminérgicas supervivientes en parkinson idiopático, lo que sugiere un rol neuroprotector (Balestrino and Schapira 2020).

GBA1 codifica para la glucocerebrosidasa, una enzima lisosomal que metaboliza la glucosilceramida, y sus mutaciones constituyen el factor genérico de riesgo más importante para Parkinson que se conoce. Estas mutaciones son altamente prevalentes en pacientes de Parkinson (entre un 5% y un 25% de ellos porta alguna) y tiene una razón de probabilidad del 5.43. Este gen interactúa con diferentes vías relacionadas con la patogénesis de Parkinson. Tiene una relación recíproca con la acumulación de la alfa-sinucleína, el estrés del retículo endoplasmático y la disfunción mitocondrial. El Parkinson relacionado con GBA es indistinguible del Parkinson esporádico, con diferencias en que tienen un inicio más temprano, un deterioro más rápido y un riesgo más alto de desarrollar impedimentos cognitivos (Balestrino and Schapira 2020; Balestrino and Schapira 2018).

Finalmente, otras proteínas relacionadas con la red trans de Golgi han sido asociadas a Parkinson, como nueve variantes raras de LRP10, VPS35 y GGA1. Pero sus roles en la enfermedad aún no se conocen. (Balestrino and Schapira 2020; Williams, Chen, and Moore 2017; Quadri et al. 2018; Einem et al. 2017).

Micro ARNs:

Los miARNs (micro ácidos ribonucleicos) consisten en pequeños grupos no codificantes de 19 a 24 nucleótidos. Tienen funciones regulatorias de la expresión de genes diana mediante las rutas de ribointerferencia. El perfilado transcriptómico ha permitido el descubrimiento de micro ARNs que se expresan diferencialmente en tejidos afectados por enfermedades. Las firmas de micro ARNs permiten detectar y clasificar distintas enfermedades, así como predecir la severidad de la enfermedad (Ross, Carlson, and Brock 2007; Pillai 2005).

El genoma humano codifica cerca de 2600 micro ARNs según la versión de miRBase 22, y de acuerdo a datos de la versión 29 de GENCODE mas de 200000 transcritos(Plotnikova, Baranova, and Skoblov 2019).

Biogénesis:

Los micro ARNs son transcritos por la ARN polimerasa II (Pol II), y ocasionalmente por la ARN polimerasa III. Su transcripción produce los transcritos primarios de micro ARN(Cullen 2004). Estos son largos y pueden contener modificaciones 5' y 3' idénticas a las presentes de ARN mensajero. Muchos micro ARNs tienen sus secuencias codificantes localizadas dentro de genes anotados a ARN mensajero o otros ARNs, entonces estos genes son llamados genes hospedadores. Cuando la secuencia codificante del micro ARN se encuentra situada cerca de un gen, se asume que la transcripción del mismo produce el transcrito del micro ARN, sin embargo es posible que este tenga su propio promotor. También hay veces que los micro ARNs tienen su propia unidad transcripcional, aunque es posible que compartan transcritos con otros genes también(Graves and Zeng 2012).

Los micro ARNs se dividen en dos clases, canónica y no canónica. Estas se basan en como se procesa el micro ARN primario. En la vía canónica, el transcrito primario es cortado por Drosha (una proteína de la familia RNasa III), mediante la unión a su subunidad DGCR8, lo que libera en el núcleo el precursor (pre-miRNA) que tiene la estructura de horquilla debido a poseer secuencias complementarias en un mismo transcrito. Después es transportado al citoplasma a través de las Exportina 5. En su nueva localización es cortado por Dicer (otra RNasa de la misma familia que Drosha) para producir un duplex intermediario de micro ARN de alrededor de 22 pares de bases. Entonces la hebra funcional se incorpora a el complejo miRISC (miRNA-induced silencing complex), ensamblado por Dicer1, TRBP (trans-activation-responsive RNA-binding protein) y las proteínas argonautas (AGO1, AGO2, AGO3, AGO4), siendo entonces funcional para ejercer su regulación postraducciona. El complejo incluye. Cual hebra se descarta depende de la estabilidad termodinámica(Chong et al. 2010; Graves and Zeng 2012; Pu et al. 2019). Después el complejo Ago:ARN La vía no canónica no requiere las proteínas anteriores. En algunos casos, como los genes mirtron, los micro ARNs son producidos por splicing. En otros casos algunas de las proteínas no son usadas como con los siARNs (Ácidos ribonucleicos pequeños endógenos de interferencia), que son procesados por Dicer en su fase de micro ARN primario(Graves and Zeng 2012).

Regulación de la producción de ARN mensajero:

Los Micro ARNs tienen un papel de reguladores postranscripcionales en el citoplasma, localizándose en los cuerpos P o cuerpos de procesamiento, un conjunto de enzimas involucradas en el recambio del ARN mensajero. Además se han observado en orgánulos citoplásmicos como la mitocondria y tiene unas serie de papeles la regulación de la función celular. Por otro lado, también se ha detectado micro ARN en el núcleo celular, teniendo este también funciones de silenciamiento y o activación(Pu et al. 2019). Existen varios mecanismos de regulación de la expresión génica por micro ARNs, aquí se comentan aquellos que ocurren en animales.

Desadenilación, descaperuzar y decaimiento: Este proceso empieza con el recortado preliminar de la cola poli-A del ADN mensajero objetivo por el complejo PAN2-PAN3 (subunidad PAN2/3 de ribonucleasa específica de la cola poli-A), seguido por la subsecuente desadenilación por parte del complejo CCR4-NOT (carbon catabolite repression-negative on TATA-less). El proceso continúa con el descaperuzamiento del extremo 5' por diferentes proteínas como son DCP2, DCP1, PATL1, DDX6 (DEAD box helicase 6), LSM1-7 y LSM14. Esto expone el ARNm a la degradación por exoribonucleasa de XRN-1(Frédéric and Simard 2022).

Inhibición de la traducción: miRISC puede inhibir la traducción previamente a los eventos de desadenilación, pudiendo ser dependiente de este proceso o no. miRISC realiza esta inhibición en el paso de iniciación de la traducción con ayuda del complejo GW182 y CCR4-NOT. Además la traducción puede ser inhibida

con independencia de GW182 con la ayuda de otros elementos. Finalmente, también se puede inhibir la traducción mediante la unión a la región codificante (CDS) del ARNm, lo bloquea al ribosoma (Frédéric and Simard 2022).

Degradación de miARN dirigida por ARN: Este proceso ocurre cuando se produce una unión extensa del ARN del complejo miRISC con la región no traducida 3' (3' UTR). Esto provocará el reclutamiento del complejo ZSWIM8 (complejo ligador de ubiquitinas), lo que a su vez permitirá la poliubiquitinación de AGO. Esto la dirigirá a su degradación proteasomal y expondrá el miARN a RNAsas, degradando también el ARN diana (Hiers et al. 2024).

Micro ARNs en sangre o Micro ARNs circulantes:

Se descubrió que los miARNs son secretados por diferentes tipos celulares bajo condiciones normales y patológicas. Esto son altamente estables y se pueden entregar a las células de destino de manera funcional (Ortiz-Quintero 2016). Estas moléculas fueron aisladas en 2008 del serum y plasma de individuos sanos, y se demostró que eran estables y que estaban protegidos de las RNAsas (Mitchell et al. 2008), lo que los hace viables como biomarcadores potenciales. Además en este estudio estableció que se podía distinguir entre pacientes sanos y pacientes con cáncer de próstata a través de miR-141.

Micro ARNs circulantes en Parkinson: La desregulación de miARNs está implicada en diferentes enfermedades neurodegenerativas, como el Alzheimer (Van den Hove et al. 2014), esclerosis múltiple (Junker et al. 2009) y Parkinson (Patil et al. 2019; Martins et al. 2011), contribuyendo a la neurodegeneración y la progresión de la enfermedad (Oliveira et al. 2020).

Existen diferentes micro ARNs que han sido identificados como expresados diferencialmente en pacientes de Parkinson y controles sanos. miR-34a fue detectado diferencialmente en plasma y se tiene un rol clave en la inflamación en enfermedades neurodegenerativas, en cuanto a Parkinson a través de la regulación de la expresión de TREM2 (Nuzziello and Liguori 2019; Jay, Saucken, and Landreth 2017; S. Bhattacharjee, Zhao, and Lukiw 2014). También fueron detectados con niveles diferenciales en PBMCs (célula mononuclear de sangre periférica) miR-146a y miR-155, que son micro ARNs inflamatorios que han sido asociados a Parkinson (Caggiu et al. 2018; Chen et al. 2008), y en suero y PBMC miR-335 (Patil et al. 2019; Martins et al. 2011), cuyas versiones 3p y 5p además otro estudio encontró regulados negativamente (Oliveira et al. 2020). Otros micro ARNs que han sido asociados a Parkinson pero que no se han descubierto diferencialmente expresados en sangre en pacientes de Parkinson son: miR-21, miR-34b, miR-34c (Oliveira et al. 2020).

Materiales:

Software y paquetes:

Para programar los scripts con los que preprocesar, entrenar los modelos y analizarlos se usó el lenguaje de programación R (R Core Team 2023) y el programa RStudio (RStudio Team 2020) con los paquetes caret (Kuhn and Max 2008), glmnet (Friedman, Tibshirani, and Hastie 2010; Tay, Narasimhan, and Hastie 2023) para realizar LASSO, tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016) y el paquete ranger (Wright and Ziegler 2017) para bosques aleatorios.

Para hacer el enriquecimiento funcional se usó la herramienta web miEAA (Aparicio-Puerta et al. 2023) y su opción para GSEA. Dado que esta realiza estos análisis para micro ARNs.

Para realizar la prueba exacta de Fisher se utilizó la función específica para esta prueba del paquete "stats", versión 4.3.3. (R Core Team 2023)

Datasets:

Se seleccionaron las muestras contenidas en el fichero “muestrasCA_CO_LAST.csv” del data set sobre transcriptómica de la cohorte Parkinson’s Progression Markers Initiative (PPMI)(Marek et al. 2011) “std_quantification_rpmmm_norm_mirna.final_ids.csv”, proveniente del archivo actualizado de todos los fenotipos con prodromales de Julio de 2022. Este archivo contiene las lecturas de 2656 micro ARNs en 4443 pacientes de la cohorte PPMI.

Métodos:

Selección de micro ARNs mediante modelos predictivos:

Las técnicas de Machine Learning permiten abordar la identificación de micro ARNs como un problema selección de características. Por ello este estudio está orientado a la selección de un subconjunto de predictores que, en un modelo, resuelva satisfactoriamente la clasificación de pacientes entre sanos y control. Este proceso hará uso de modelos basados en LASSO (Least Absolute Shrinkage and Selection Operator)(Ranstam and Cook 2018) y en segundo lugar (con intención de establecer un contraste) en Bosques Aleatorios(Rigatti 2017). Tras obtener los micro ARNs que son usados por los modelos, estos pueden ser examinados para la búsqueda de posibles funciones o relaciones con el Parkinson.

Preprocesado:

Los datos ya habían sido normalizados por los creadores del dataset, así como limpiados de valores ausentes. Se usaron varias iteraciones del preprocesamiento para distintos modelos. Para los modelos basados en el uso del algoritmo LASSO de glmnet, primero se eliminaron las variables con varianza cercana a cero (frecuencia del valor mas común respecto al segundo 95/5 y un 10% de porcentaje de corte de valores únicos) y las altamente correlacionadas (punto de corte 0.9), lo que las redujo a 591 micro ARNs, y se escogieron de aquellos pacientes coincidentes en el archivo muestras CA_CO_LAST.csv que fuesen enfermos de Parkinson o controles sanos (Anotados como PD y HC respectivamente) y que cuyas entradas fuese de la primera visita (anotadas como BL), lo que dio lugar a una matriz con 659 muestras. Para el segundo modelo se realizó adicionalmente al paso anterior un submuestreo para balancear el data set, que tenía originalmente 515 enfermos y 144 controles sanos, y quedó en 144 enfermos. El tercer modelo se realizó con la matriz completa balanceada (sin reducir variables). Esta última matriz fue usada también para entrenar un modelo basado en un algoritmo de bosques aleatorios de glmnet.

Criterios de evaluación de modelos:

Las principales estadísticas valoradas en los modelos para evaluar su capacidad fueron la precisión balanceada, el p-valor del test de McNemar y el p-valor sobre la hipótesis alternativa de que la precisión es significativamente mayor que la tasa de predicción sin información, así como esta última tasa en si. Para la ejecución de los scripts de los modelos se usó la semilla de aleatoriedad 124.

Precisión balanceada:

Consiste en la media aritmética de la sensibilidad y la especificidad, típicamente usada cuando se trata con datos desbalanceados.

$$PrecisinBalanceada = (Sensibilidad + Especificidad)/2$$

Prueba de Mc Nemar:

En este contexto esta prueba comprueba si hay diferencia en que clase de errores comete el error, si tiende a clasificar verdaderos negativos como positivos o verdaderos positivos como negativos. La hipótesis nula es que no hay tendencia en el tipo de error que comete.

Tasa de no información:

Consiste en la proporción de la clase más observada e indicaría la precisión si el modelo siempre clasifica como esa clase.

P-valor:

Este p-valor es referido al resultado de un test binomial cuya hipótesis nula es que la precisión y la tasa de no información son las mismas, y la alternativa de que la precisión es mayor. Este test evalúa si la precisión del modelo es significativa respecto a un modelo que solo clasifica como la clase más numerosa.

LASSO:

LASSO(Least Absolute Shrinkage and Selection Operator) es un un método de análisis de regresión que pretende identificar las variables y los correspondientes coeficientes de regresión que dan lugar a un modelo que minimiza el error de la predicción. Esto se logra imponiendo una penalización en las variables predictoras del modelo, que encoge los coeficientes hacia cero. Esto se consigue forzando que la suma del valor absoluto de los coeficientes de regresión sea menos que una valor fijado (λ). Las variables cuyo coeficientes sea 0 quedan eliminadas. Esta eliminación reduce el riesgo de sobreajuste y simplifica los modelos (Ranstam and Cook 2018).

Se entrenaron 3 modelos con el algoritmo de glmnet(Friedman, Tibshirani, and Hastie 2010; Tay, Narasimhan, and Hastie 2023) para red elástica usando solamente LASSO (es decir con α igual a 1 en el algoritmo de glmnet) y usando validación cruzada de 10 pliegues para seleccionar la λ optima. Se usó el error cuadrático medio como medida para la selección de λ .

Se comprobó la capacidad predictiva de los modelos finales mediante matrices de confusión realizadas con predicciones del modelo y valores originales de la variable respuesta (de un conjunto de test que se dejó fuera del entrenamiento).

Inicialmente se usó la λ de error mínimo para construir los tres primeros modelos de LASSO, pero para poder obtener un mayor número de micro ARNs como input para las herramientas de enriquecimiento funcional, se eligió otra λ para el cuarto modelo de LASSO, lo que permitió obtener hasta 33 micro ARNs con el valor de λ de 0.0654724.

Parámetros de LASSO:

alpha(α): Este parámetro controla el equilibrio en el uso de las regularizaciones L1 y L2, es decir entre LASSO y regresión de Ridge. Un valor de 1 provoca que solo se aplique LASSO y un valor de 0 que solo se aplique Ridge.

lambda(λ): Consiste en el término de penalización que se aplica a los coeficientes, en el caso de LASSO es una regularización L1. Cuanto más grande sea λ mayor será la penalización que puede llegar a eliminar variables al reducir los coeficientes de estas a 0.

nfolds: Este parámetro indica cuantos pliegues tiene la validación cruzada usada en el entrenamiento del modelo.

type: Cuando se realiza una predicción con la función “predict”, este parámetro especifica el formato de los resultados de la predicción.

Algoritmo ranger(bosques aleatorios):

Los bosques aleatorios usan aleatorización para crear una gran cantidad de árboles de decisión. El resultado de estos árboles es agregado usando un mecanismo de votación para problemas de clasificación (o haciendo media en los problemas de regresión). La aleatorización se implementa de dos maneras. Primero mediante muestreo con reemplazamiento (bootstrap sampling). En segundo lugar la aleatorización ocurre en los nodos de decisión. En cada nodo se eligen cierta cantidad de predictores, usualmente se usa una cantidad de predictores equivalente al redondeo de la raíz cuadrada del número total de predictores. Una vez elegidos, se test todos los posibles umbrales para las variables seleccionadas y entonces selecciona la combinación de variables y umbral que genera la mejor separación de las muestras (en nuestro caso aquel que separa controles y enfermos de Parkinson mejor). Este procedimiento continúa hasta alcanzar nodos puros (aquellos contienen solo uno de los dos tipos de respuesta solo) o cuando llega a otro punto terminal predefinido. El proceso se repite desde el principio hasta alcanzar el número de árboles deseado, dando lugar a un bosque (Rigatti 2017).

Para realizar modelos de bosques aleatorios se usó el algoritmo de caret “ranger”(Wright and Ziegler 2017), usando una validación cruzada de 5 pliegues (por tiempo de computación). Se usó una red de búsqueda, que es una parrilla de hiperparámetros que el entrenamiento debe usar generando diferentes combinaciones para generar diferentes modelos, entre los cuales elegir el que mejor prediga, usando la precisión.

Hiperparámetros del bosque aleatorio:

splitrule: Esta es la métrica que el algoritmo usará para generar los distintos árboles. Se usó la impureza de Gini.

mtry: El número de predictores que son usados en cada nodo. Se usaron valores que son fracciones de la raíz cuadrada redondeada del número de variables(50). Los valores son 5, 25, 50 y 75, 100.

Tamaño mínimo de nodo: El número de predictores a partir del cual se para el crecimiento del árbol. Se usaron los valores 5, 15, 30, 60 y 90.

metric: La métrica usada para valorar los modelos. Fue usada la precisión (Número de verdaderos positivos dividido entre la suma de los positivos verdaderos y falsos), “Accuracy”.

num.trees: La cantidad de árboles que va a tener el bosque. Se fijó en 500 árboles.

Análisis de enriquecimiento funcional:

Las herramientas de enriquecimiento funcional facilitan el análisis de una tanda de genes para identificar los procesos biológicos en los que están implicados en el organismo.

Con el fin de identificar información acerca de los posibles caminos metabólicos y grupos de genes implicados en enfermedades que estuviesen representados por los micro ARNs de los modelos, se usaron las herramientas de enriquecimiento funcional de la página miEAA(Aparicio-Puerta et al. 2023), que funcionan específicamente con micro ARNs y sus precursores. Los análisis de enriquecimiento de esta web están basados en el programa GeneTrail3(Stöckel et al. 2016).

Tanto para los micro ARNs seleccionados por LASSO(cuarto modelo, sin reducir matriz, con submuestreo y λ reducida) como para aquellos seleccionados mediante bosques aleatorios, se realizó un GSEA, ordenados mediante el coeficiente en valor absoluto en el caso de LASSO y los 100 micro ARNs con mayor puntuación de importancia, y un análisis ORA. Sin embargo para cada modelo solo una de las técnicas dio resultados significativos. ORA para bosques aleatorios y GSEA para LASSO.

GSEA:

GSEA (Gene set enrichment analysis) surge como una técnica de análisis de enriquecimiento funcional similar a ORA (Overrepresentation analysis) pero que tiene en cuenta el hecho de que hay genes con mayor o menor cantidad de expresión. Consiste en una técnica que tiene como input una lista de los genes ordenados por sus valores de interés, típicamente niveles de expresión diferencial (aunque existen otros tipos de información válidos para ordenar). Esto se realiza en tres pasos. Primero se calcula la puntuación de enriquecimiento, para ello se realiza un test similar a la prueba de Kolmogorov-Smirnov para comprobar que los miembros del set de genes a priori están distribuidos aleatoriamente en los genes de la lista o si se encuentran en parte inferior o superior (los sets relacionados con la lista no tendrán una distribución aleatoria). Esto se calcula para cada gen de arriba de la lista, generando una suma continuada, la magnitud del incremento que aporta un gen depende de su correlación con el gene set. La puntuación final de enriquecimiento es la desviación máxima de cero en el gráfico de la suma. El segundo paso consiste en estimar el nivel de significancia de la puntuación de enriquecimiento. Para ello se recalcula la puntuación de enriquecimiento permutando las etiquetas de fenotipo, lo que genera una distribución nula para la puntuación de enriquecimiento. Entonces el p-valor de la puntuación es calculado en función de esta distribución nula. El tercer paso es el ajuste por pruebas múltiples. Primero se normaliza la puntuación de enriquecimiento y luego se realiza el ajuste (En el caso de este trabajo FDR) (Subramanian et al. 2005).

Parámetros de la herramienta de miEAA:

Las categorías usadas fueron: Pathways (KEGG), Diseases (miRWalk), Gene Ontology (miRWalk), Pathways (miRWalk), Diseases (MNR) y Published Diseases. El método de ajuste del p valor por pruebas múltiples fue FDR (Benjamini-Hochberg), el rango de ajuste del p valor fue a cada categoría independientemente. El nivel de significancia de corte usado fue 0.05 y el número de coincidencias mínimas por categoría 2.

En primer lugar se realizó con el modelo de LASSO sin matriz reducida y con balanceado de clases, de λ con menor error cuadrático. Este modelo no dio lugar a resultados significativos en el análisis. Por eso se realizó en segundo lugar otro modelo con una λ (0.0654724) que permitiera el uso de una mayor cantidad de micro ARNs, que si dio lugar a los resultados que se exponen en la sección de resultados.

GSEA también fué usado con los micro ARNs usados por el modelo de bosque aleatorio (aquellos micro ARNs que tienen una puntuación de importancia superior a 0).

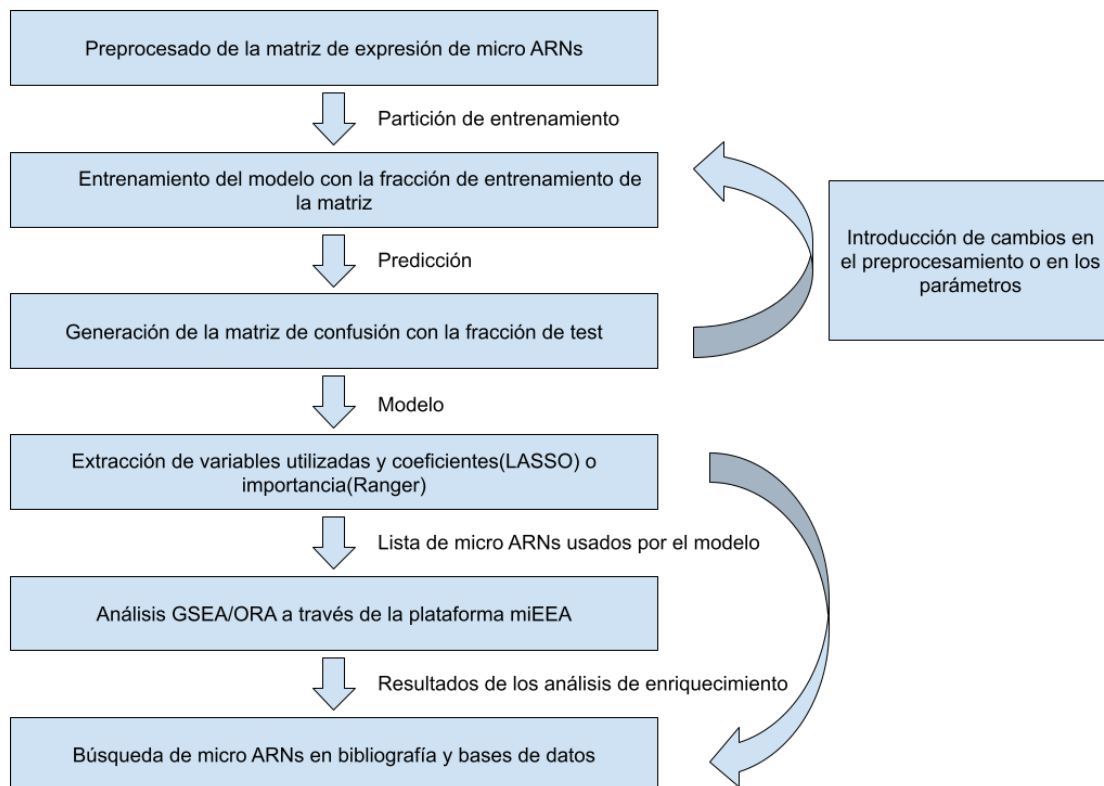


Figura 1: Diagrama de flujo de trabajo

Resultados:

Rendimiento de los modelos:

Como se puede ver en la tabla 1, los dos primeros modelos hechos con LASSO el no permitían descartar que la tasa de no información no fuese significativamente distinta a la precisión (P.valor [Acc > NIR]), siendo NIR (No information rate) muy alta en el primer modelo. Esto es debido a que la clase de enfermos de parkinson es mucho mayor que la de controles sanos. Se observa que el submuestreo del data set reduce el P.valor, aunque sin usar la matriz completa sigue siendo demasiado alto. Con el uso del balanceo y sin reducir el número de variables se obtienen mejores resultados, teniendo una precisión balanceada mayor (0.6537) y un p-valor inferior al 0.05. El bosque aleatorio, por otro lado, muestra los mejores valores de todos los modelos, teniendo la mayor precisión balanceada y el menor p-valor. Además la utilidad de los dos primeros modelos no es significativa debido a que usan solo los interceptos. En cuanto al p.valor de la prueba de McNemar, esta indica si las frecuencias marginales son iguales para muestras pareadas (McNemar 1947). En este caso indica si el modelo tiene una tendencia a cometer un tipo de error u otro, o si por el contrario no son significativamente diferentes. En los dos primeros modelos que hay una diferencia entre como clasifica cada clase, teniendo estos una tendencia a clasificar sanos como enfermos del parkinson. En el caso de los modelos con matriz completa y sobremuestreo, así como con el modelo de bosque aleatorio, no hay una diferencia significativa en los errores al clasificar una clase u otra.

Resultados del análisis GSEA:

El análisis de GSEA con el primer modelo LASSO balanceado no se obtuvieron resultados significativos para los parámetros usados (descritos en métodos), pero tras elegir una λ que diese lugar a un modelo con 34 micro ARNs (el segundo modelo LASSO balanceado) se obtuvieron resultados significativos con el análisis GSEA. Los resultados se muestran en la tabla 3. Como se puede observar hay dos micro ARNs en común en todas las categorías, hsa-miR-93-3p y hsa-miR-34a-5p. Además estos dos micro ARNs estaban presentes en 991 categorías de 1003, siendo también los únicos micro ARNs asociados a 971 categorías, estando presentes en aquellas con el p valor ajustado más bajo, estando más enriquecidos. Esto indica su importancia en el análisis. Si embargo, todas estas categorías utilizan micro ARNs con coeficientes bajos, con la única excepción de la categoría “Cáncer del tracto biliar”, la cual usa micro ARNs con coeficiente alto quedando estos por encima de la mitad de la lista ordenada de mayor coeficiente a menor. También cabe mencionar que 21 categorías eran de algún tipo de cáncer, pero con micro ARNs de coeficientes bajos.

Tabla 1: Valores de las matrices de confusión. Modelo: indica de que modelo son estos resultados; Precisión balanceada: Consiste en la media de la sensibilidad y la especificidad, usada para datos desbalanceados; P.valor: Este p.valor es relativo a la hipótesis nula de que la precisión no sea significativamente distinta de la tasa de no información; NIR: No information ratio, tasa de no información, indica que precisión se obtendría en caso de elegir clasificar como la clase más numerosa; N° de variables: La cantidad de variables que usa el modelo para predecir; *: El modelo de random forest usó 1742 variables pero solo 26 sobrepasan la puntuación de importancia de 50.

| Modelo | Precisión balanceada | P-valor | NIR | N° de variables | P-valor de McNemar |
|---|----------------------|----------|--------|-----------------|--------------------|
| LASSO sin submuestreo y con matriz reducida | 0.5000 | 0.5427 | 0.8030 | Intercepto | 1.166e-09 |
| LASSO con submuestreo y matriz reducida | 0.5287 | 0.3746 | 0.5402 | Intercepto | 8.636e-08 |
| LASSO con matriz completa y submuestreo | 0.6537 | 0.006244 | 0.5116 | 10 | 0.100348 |
| LASSO con matriz completa, submuestreo y lambda menor | 0.6770 | 0.003197 | 0.5233 | 34 | 0.344704 |
| Bosque Aleatorio | 0.8785 | <2e-16 | 0.5000 | 1742* | 0.1763 |

Tabla 2: micro ARNs usados por el modelo de LASSO con lambda menor

| miRNA | Coeficientes | miRNA | Coeficientes |
|-------------------|--------------|---------------------|--------------|
| ‘hsa-miR-6864-5p’ | 9.3730699 | ‘hsa-miR-3922-5p’ | 0.4086642 |
| ‘hsa-miR-1229-5p’ | 2.6344518 | ‘hsa-miR-3679-3p’ | 0.3942538 |
| ‘hsa-miR-4733-5p’ | 2.5255517 | ‘hsa-miR-1825’ | 0.3828701 |
| ‘hsa-miR-6726-5p’ | 2.2339524 | ‘hsa-miR-6829-5p’ | 0.3447844 |
| ‘hsa-miR-6857-5p’ | 2.0348835 | ‘hsa-miR-3175’ | 0.3307051 |
| ‘hsa-miR-3182’ | 1.5240102 | ‘hsa-miR-508-5p’ | 0.2661245 |
| ‘hsa-miR-6888-3p’ | 1.3879994 | ‘hsa-miR-6769b-3p’ | 0.2517903 |
| ‘hsa-miR-6799-5p’ | 0.9292071 | ‘hsa-miR-135a-2-3p’ | 0.2328511 |
| ‘hsa-miR-190a-3p’ | 0.9195580 | ‘hsa-miR-3074-3p’ | 0.2277414 |
| ‘hsa-miR-5586-3p’ | 0.8758889 | ‘hsa-miR-3684’ | 0.2193223 |
| ‘hsa-miR-4638-5p’ | 0.8388170 | ‘hsa-miR-365b-5p’ | 0.1944159 |
| ‘hsa-miR-4688’ | 0.8230123 | ‘hsa-miR-6763-5p’ | 0.1289912 |
| ‘hsa-miR-6787-3p’ | 0.7530091 | ‘hsa-miR-664b-5p’ | 0.0946445 |
| ‘hsa-miR-381-5p’ | 0.7416059 | ‘hsa-miR-6509-3p’ | 0.0175910 |
| ‘hsa-miR-20b-3p’ | 0.7111125 | ‘hsa-let-7i-3p’ | 0.0161223 |
| ‘hsa-miR-1321’ | 0.5284157 | ‘hsa-miR-93-3p’ | 0.0014805 |
| | | ‘hsa-miR-34a-5p’ | 0.0010678 |

Tabla 3: Resultados de GSEA para el modelo de LASSO balanceado con más genes. Categoría: El término enriquecido en el análisis GSEA; FDR: p-valor ajustado por el método Benjamini-Hochberg; ARNs: Los micro ARNs que pertenecen a esa categoría; Relación: Aquello por lo que está relacionado con el Parkinson, generalmente un tipo de patología, proceso o organulo.

| Categoría | FDR | ARNs | Relación |
|---|-----------|--|-------------------------|
| Peroxisome | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Ferroptosis |
| peroxisomal membrane | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Ferroptosis |
| iron ion binding | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Ferroptosis |
| mitochondrial outer membrane translocase complex | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| mitochondrial inner membrane | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| mitochondrial respiratory chain | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| mitochondrial intermembrane space | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| protein targeting to mitochondrion | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| mitochondrial membrane | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| mitochondrial nucleoid | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| Mitochondrial LC Fatty Acid Beta Oxidation | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Disfunción mitocondrial |
| Inflammation mediated by chemokine and cytokine signaling pathway | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Neuroinflamación |
| positive regulation of inflammatory response | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Neuroinflamación |
| Apoptosis Modulation by HSP70 | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Apoptosis |
| Apoptosis | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Apoptosis |
| positive regulation of neuron apoptotic process | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Apoptosis |
| negative regulation of neuron apoptotic process | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Apoptosis |
| Parkinson disease | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | |
| golgi membrane | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Aparato de Golgi |
| golgi apparatus | 0.0061202 | hsa-miR-93-3p; hsa-miR-34a-5p | Aparato de Golgi |
| Neurodegenerative diseases | 0.0369428 | hsa-miR-20b-3p; hsa-miR-508-5p; hsa-let-7i-3p; hsa-miR-93-3p; hsa-miR-34a-5p | |
| Cadherin signaling pathway | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Neuroinflamación |
| Axon guidance by netrin | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Apoptosis |
| p38 MAPK signaling pathway | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Autofagia |
| IL 2 signaling pathway | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Neuroinflamación |
| TCR signaling Pathway | 0.0034479 | hsa-miR-93-3p; hsa-miR-34a-5p | Neuroinflamación |

Solapamiento de micro ARNs entre modelos y Prueba exacta de Fisher:

Trás comparar los micro ARNs usados por los modelos de bosque aleatorio y LASSO con menor λ , hay 32 micro ARNs comunes a ambos modelos (de los de LASSO queda fuera del solapamiento solo hsa-miR-1825).

Para comprobar que el solapamiento de micro ARNs es significativo estadísticamente se usó la prueba exacta de Fisher (Fisher 1922). Esta consiste en una prueba estadística que se usa para determinar la significación de la asociación entre dos tipos de clasificación. Es usada en el caso de solapamiento de genes para cuantificarlo entre dos listas de genes. Para ello se usa una tabla de contingencia con los siguientes términos: Genes que se encuentran en ambas lista, los que se encuentran en la primera pero no en la segunda, los que se encuentran en la segunda pero no en la primera y los que no se encuentran en ninguna lista. Tras esto se usa la prueba exacta de Fisher para calcular un p-valor (Fury et al. 2006).

Tras realizar la prueba exacta de Fisher para el solapamiento de micro ARNs entre el último modelo de LASSO con más genes y bosque aleatorio (con la hipótesis alternativa de que la razón de momios sea mayor que 1), se obtuvo un p-valor de $1.495e-5$, por lo que, para un nivel de significación del 0.05, habría evidencia estadística de que los grupos de micro ARNs de LASSO y bosques aleatorios guarden relación y el solapamiento es significativo. Es decir, que ambos modelos hayan escogido estos micro ARNs no es debido al azar. La misma operación realizada entre bosque aleatorio y el modelo con λ más óptima (que tenía 9 micro ARNs en común, es decir todos) resultó en un p-valor de 0.02222, también por debajo de un nivel de significancia del 0.05.

Discusión:

Los modelos:

Se puede observar que el uso de una mayor cantidad de micro ARNs para el entrenamiento de modelos con LASSO permite una mejora en la capacidad de clasificación del modelo. Esto puede ser debido a que las 1019 variables altamente correlacionadas (con un punto de corte de 0.9) y 1158 variables con varianza cercana a 0 contengan información útil.

Por otro lado, el balanceo de clases (en este caso submuestreo) ha resultado esencial para obtener un modelo útil (que use más variables que el intercepto).

En comparación con el uso de bosques aleatorios, el resultado es más interpretable y permite la selección de variables de una forma más intensa (es decir que se descartan más micro ARNs utilizando solo los más útiles para el modelo), pero como se ha mencionado anteriormente es posible la pérdida de información útil. Si bien la precisión balanceada era inferior con los modelos realizados mediante LASSO, el modelo que usa bosques aleatorios realiza una selección de variables que deja un gran número de micro ARNs.

Micro ARNs usados por el modelo de LASSO con λ óptima:

En esta sección comentaremos algunos de los diferentes hallazgos sobre los micro ARNs que participan en el modelo LASSO con λ óptima y también en el que tiene más micro ARNs. En general se ha buscado los artículos que puedan relacionarlos con el Parkinson directamente, y luego aquellos artículos que hablen de su funcionalidad o su posible papel en procesos que son importantes en Parkinson, con la intención de identificar aquellos micro ARNs relacionados con la enfermedad. Las búsquedas se realizaron a través de los buscadores del NCBI (National Center for Biotechnology Information) (“National Center for Biotechnology Information (NCBI) Search Engine,” n.d.), Google Académico, GeneCaRNA (Barshir et al. 2021), RNAcentral (“RNAcentral 2021: Secondary Structure Integration, Improved Sequence Search and New Member Databases” 2021) y miRBase (Kozomara, Birgaoanu, and Griffiths-Jones 2019). Ni hsa-miR-4733-5p, ni has-miR-6787-3p

hsa-miR-6864-5p:

La sobreexpresión de hsa-miR-6864-5p reduce los niveles de la proteína TRIM13 y su ARN mensajero. Esta proteína está involucrada en la retrotranslocación y reemplazo de proteínas secretoras y de membrana en el retículo endoplasmático, regulando la autofagia inducida por estrés (Lai et al. 2023). Este proceso actúa sobre proteínas mal plegadas así como en la degradación regulada de aquellas que están bien plegadas. Un defecto en este micro ARN o unos niveles descompensados podrían dar lugar a una reducción de la autofagia de proteínas en el retículo endoplasmático, lo que a su vez podría estar relacionado con la acumulación de α -sinucleína que no es eliminada a través de este mecanismo. TRIM13 también participa en la respuesta inmune estimulando la actividad de NF-kappa-B en el camino de señalización TLR2. NF-kappa-B (Huang and Baek 2017), que a su vez está relacionada con la apoptosis (Sheikh and Huang 2003). Por tanto una mala regulación inicial del micro ARN puede llevar a el desequilibrio en la actividad de NF-kappa-B, promoviendo la apoptosis, posiblemente de las células dopaminérgicas.

hsa-miR-1229-5p:

La versión 3p de hsa-miR-1229-5p regula la expresión de SORL1, que a su vez se asocia con la disfunción lisosomal y la agregación de proteínas mal plegadas en desordenes neurodegenerativos. Un estudio en una cohorte del norte de china concluyó que dos polimorfismos de este gen (rs1010159 y rs2298813) podrían estar implicados en el Parkinson, por lo menos en la población de estudio (Y. Wang et al. 2022). Además, otros polimorfismos de este gen se han correlacionado con el susceptibilidad al Alzheimer, otra enfermedad neurológica amiloidea (Liu et al. 2017). Si bien la versión 5p puede no tener función y ser degradada, o tener una función distinta (o distintos genes diana).

hsa-miR-20b-3p:

En cuanto a hsa-miR-20b-3p, miR-20b tiene un papel en la respuesta inflamatoria durante inducida por isquemia cerebral teniendo como diana la proteína NLRP3. Su regulación negativa inhibe el camino de señalización de NLRP3 y los niveles downstream de IL-1 β y IL-18, también reduciendo los niveles de ATP y de radicales libres, mitigando el daño inflamatorio tras las isquemia en células cerebrales (Zhao et al. 2019). Por tanto se encuentra relacionado con la neuroinflamación, pudiendo tener relación con la patología del Parkinson.

hsa-miR-3074-3p:

Un estudio utilizó diferentes métodos bioinformáticos, como el enriquecimiento con Gene Ontology y el mapeado funcional de los genes en la enciclopedia de Kyoto de genes y genomas, para averiguar que procesos están enriquecidos en hsa-miR-3074-3p, encontrando que podría actuar en la inhibición de la proliferación y diferenciación neuronal, en un contexto de alteraciones de los micro ARNs tras el tratamiento con células precursoras de oligodendrocitos a ratas con daños en la médula espinal (J. Yang et al. 2018). Su versión 5p promueve la apoptosis en osteoblastos (células MC3T3) bajo sobrecarga de hierro (Feng et al. 2021). También promueve la apoptosis en trofoblastos extravelosos (células HTR8/SVneo) cuando es sobreexpresado (Gu et al. 2018). Si bien no hay evidencia directa o indirecta de su implicación en Parkinson, su capacidad de inhibición en neuronas y la relación de su versión 5p con la apoptosis en diferentes células, podría tener un papel no descubierto en Parkinson.

hsa-miR-3182:

Se descubrió que cuando este micro ARN es sobreexpresado, E2F7 (E2F transcription factor 7) es regulado negativamente. Este factor de transcripción tiene un rol esencial en el control del ciclo celular y en la inhibición de la apoptosis (Khazaei-Poul et al. 2023). Por tanto es posible que una alteración en los niveles

de miR-3182 dé lugar a la apoptosis. Lo que lo pondría como candidato temprano para buscar si tiene un rol en la patología del Parkinson.

hsa-miR-6726-5p:

Un estudio reveló que la versión 3p de este micro ARN empeoraba la viabilidad celular ante la exposición de MPP+ (methyl-4-phenylpyridinium) en células SHSY5Y(Pallarès-Albanell et al. 2019). Esta neurotoxina provoca parkinsonismo matando neuronas productoras de dopamina en la sustancia nigra en primates(Burns et al. 1983) y es usada en modelos del Parkinson. *In vivo* se produce por una doble oxidación en del precursor MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) por la monamina oxiase B, y ocurre en principalmente en los astrocitos del cerebro(Kopin 1987). Posiblemente la desregulación de las cantidades de este micro ARN junto con la presencia de cierta cantidad de esta neurotoxina dé lugar a una mayor tasa de muerte celular.

hsa-miR-6787-3p:

Un estudio encontró que se encuentra regulado positivamente de manera diferencial en Alzheimer(Lu et al. 2021). Si bien esto no lo relaciona con Parkinson, si lo hace con una enfermedad amiloide, y es posible que forme parte de un mecanismo común en ambas enfermedades. También es incluido por otro estudio en un grupo de microARNs sobreexpresados en enfermedades neurológicas(Lui et al. 2024).

hsa-miR-4733-5p:

El gen diana de este micro ARN es KLF7 (kruppel like factor 7) que tiene un papel importante en la morfogénesis neuronal en ciertas regiones del sistema nervioso, optimizando las moléculas que promueven el crecimiento del axón(Laub et al. 2005).

hsa-miR-93-3p:

Se ha demostrado que miR-93 reduce los daños neuronales y suprime el proceso de inflamación mediante la regulación de la expresión de STAT3 (signal transductor and transcription activator 3) en ratones con Parkinson inducido por MPTP(X. Wang, Liu, and Wang 2021). STAT3 cuando es expresado constitutivamente tiene efectos anti-apoptóticos. También puede ser un participante clave en la patofisiología de la Esclerosis Múltiple(Q. Yang, Pan, and Qian 2017).

Análisis GSEA:

Se pueden ver enriquecidas distintas categorías que se relacionan con procesos y sistemas de importancia en el Parkinson. La Ferroptosis se ve representada a través de la unión de iones de hierro y los peroxisomas, que son clave en la patología. Específicamente, la unión de iones de hierro a la α -sinucleína y la peroxidación lipídica severa, que provocan la ferroptosis. También se ve enriquecidos los elementos mitocondriales en este análisis, que puede estar relacionado con el Parkinson en el contexto de la disfunción mitocondrial. Adicionalmente, los defectos en la oxidación β mitocondrial de los ácidos grasos son partícipes de enfermedades neurodegenerativas(Kastaniotis, Autio, and R. Nair 2021). Más específicamente, se ha demostrado que hay un incremento en aminoácidos de cadena larga en pacientes de Parkinson, sugiriendo que hay insuficiencia en la oxidación β en el Parkinson temprano(Saiki et al. 2017). La inflamación también está enriquecida en estos micro ARNs por la regulación positiva de la misma y por la inflamación mediada por la vía de quimiocinas y citoquinas, que están aumentadas en las microglías de muestras de cerebro póstumas de pacientes de Parkinson(Sawada, Imamura, and Nagatsu 2006). La apoptosis también está enriquecida (en sí misma y como proceso modulado por la proteína de golpe térmico 70, así como su regulación positiva y negativa). Este es un proceso esencial en la patología del Parkinson, por la apoptosis de las células dopaminérgicas. La

proteína HSP70 (heat shock protein 70) tiene un papel neuroprotector cuando se exponen células modelo de Parkinson (patología inducida por rotenona) a un incremento en esta proteína (Li et al. 2019). Su papel en la modulación de la apoptosis es la de inhibirla y promover la supervivencia (Lanneau et al. 2008). Otros elementos que aparecen aparecen enriquecidos son el aparato de Golgi y su membrana. Se sabe que de este hay variantes de sus proteínas LRP10, VPS35 y GGA1 que están asociadas a Parkinson (Balestrino and Schapira 2020). Una vía potencialmente relevante en el Parkinson es la de señalización de la e-cadherina, se ha demostrado que su sobreexpresión puede ser un factor limitante del daño neuroinflamatorio (Cataldi et al. 2016). Otra vía que ha resultado estar implicada en el Parkinson es la del guiado del axon por Netrina-1. La deficiencia de esta netrina promueve la apoptosis dopaminérgica en Parkinson (Ahn et al. 2020). La vía de señalización de p38 MAPK (mitogen-activated protein kinase) puede ser clave en la patología del Parkinson, se ha propuesto que pueda regular la autofagia (Obergasteiger et al. 2018) y tiene un papel en la enfermedades como el Alzheimer, la esclerosis amiotrófica lateral y el Parkinson (E. K. Kim and Choi 2010; Jha et al. 2015). La vía de señalización de la interleukina 2, cuando esta es inyectada en ratones modelo de Parkinson, provoca la expansión de las células T reguladoras neuroprotectoras, lo que aumenta la neuroprotección de las neuronas dopaminérgicas nigroestriales (Markovic et al. 2022). Finalmente, también se ve enriquecida la vía de la señalización por TCR (receptor de antígeno de célula T). Esta vía se ve implicada en Parkinson en modelos de ratón porque, en el caso del TCR β , su deficiencia atenúa la neuroinflamación y la neurodegeneración (Pacheco 2021).

En cuanto al término cáncer de tracto biliar, no está claro si tiene puntos en común con el Parkinson, aparte de tener un ciclo celular aberrante. Valorar esto sale merece un estudio más a fondo que escapa a los objetivos de este TFM.

Cabe señalar que el hecho de que las categorías usen micro ARNs con coeficientes pequeños puede generar un problema por la tendencia de LASSO a incluir variables de coeficientes pequeños sin que necesariamente posean una relevancia significativa.

Micro ARNs de GSEA más relevantes:

Además también se ha buscado información acerca de los dos micro ARNs más usados por las categorías con gran diferencia que son hsa-miR-34a y hsa-miR-93-3p y también están presentes en todas las categorías relevantes para el Parkinson, aunque estos dos micro ARNs no tienen un coeficiente alto dentro del modelo LASSO. Como del segundo ya se ha hablado, ahora solo nos centraremos en el primero.

hsa-miR-34a: Aparte de lo comentado en la introducción de este trabajo, tiene una función proapoptótica. El gen que lo codifica es MIR34A. Su sobreexpresión provoca un incremento en la apoptosis y su inactivación atenúa la apoptosis mediada por p53 (Raver-Shapira et al. 2007). También tiene una expresión incrementada en vesículas extracelulares pequeñas en Parkinson temprano, y ha sido usado como variable discriminadora para diferenciar pacientes de Parkinson de pacientes sanos con buena precisión (Grossi et al. 2021).

Conclusión:

El uso de LASSO como método para la selección de variables en el contexto de la selección de micro ARNs con un rol en el Parkinson es viable, habiéndose descubierto 7 candidatos en mayor o menor grado por el modelo de lambda optima, que es el que más variables ha descartado. Sin embargo, como elimina variables debido a la multicolinealidad, es posible que se hayan perdido otros candidatos, dado que de los micro ARNs previamente asociados a Parkinson mediante su expresión diferencial ninguno ha sido usado en el modelo de lambda óptima. En cuanto al modelo de lambda reducida solo incluyó miR-34a como micro ARN conocido en su asociación con Parkinson y expresado diferencialmente en sangre. En resumen, este trabajo ha podido obtener una lista de micro ARNs que permite la discriminación de pacientes de Parkinson respecto de los controles sanos (si bien la precisión obtenida no permite su uso diagnóstico), y varios de estos micro ARNs tienen relación con el Parkinson o con vías y procesos relacionados con su patología. Por tanto, es posible

usar LASSO como técnica de selección de micro ARNs con la intención de identificar posibles candidatos a examinación posterior en relación al Parkinson, como forma de cribado.

Bibliografía:

- Ahn, Eun Hee, Seong Su Kang, Qi Qi, Xia Liu, and Keqiang Ye. 2020. “Netrin1 Deficiency Activates MST1 via UNC5B Receptor, Promoting Dopaminergic Apoptosis in Parkinson’s Disease.” *Proceedings of the National Academy of Sciences* 117 (39): 24503–13.
- Aparicio-Puerta, Ernesto, Pascal Hirsch, Georges P Schmartz, Fabian Kern, Tobias Fehlmann, and Andreas Keller. 2023. “miEAA 2023: Updates, New Functional microRNA Sets and Improved Enrichment Visualizations.” *Nucleic Acids Research*, gkad392.
- Araújo, Bruna, Rita Caridade-Silva, Carla Soares-Guedes, Joana Martins-Macedo, Eduardo D Gomes, Susana Monteiro, and Fábio G Teixeira. 2022. “Neuroinflammation and Parkinson’s Disease—from Neurodegeneration to Therapeutic Opportunities.” *Cells* 11 (18): 2908.
- Balestrino, Roberta, and AHV Schapira. 2020. “Parkinson Disease.” *European Journal of Neurology* 27 (1): 27–42.
- Balestrino, Roberta, and Anthony HV Schapira. 2018. “Glucocerebrosidase and Parkinson Disease: Molecular, Clinical, and Therapeutic Implications.” *The Neuroscientist* 24 (5): 540–59.
- Barshir, Ruth, Simon Fishilevich, Tsippi Iny-Stein, Ofer Zelig, Yaron Mazor, Yaron Guan-Golan, Marilyn Safran, and Doron Lancet. 2021. “GeneCaRNA: A Comprehensive Gene-Centric Database of Human Non-Coding RNAs in the GeneCards Suite.” *Journal of Molecular Biology* 433 (11): 166913.
- Belarbi, Karim, Elodie Cuvelier, Alain Destée, Bernard Gressier, and Marie-Christine Chartier-Harlin. 2017. “NADPH Oxidases in Parkinson’s Disease: A Systematic Review.” *Molecular Neurodegeneration* 12: 1–18.
- Bhattacharjee, Nivedita, and Anupom Borah. 2016. “Oxidative Stress and Mitochondrial Dysfunction Are the Underlying Events of Dopaminergic Neurodegeneration in Homocysteine Rat Model of Parkinson’s Disease.” *Neurochemistry International* 101: 48–55.
- Bhattacharjee, Surjyadipta, Yuhai Zhao, and Walter J Lukiw. 2014. “Deficits in the miRNA-34a-Regulated Endogenous TREM2 Phagocytosis Sensor-Receptor in Alzheimer’s Disease (AD); an Update.” *Frontiers in Aging Neuroscience*. Frontiers Media SA.
- Bose, Anindita, and M Flint Beal. 2016. “Mitochondrial Dysfunction in Parkinson’s Disease.” *Journal of Neurochemistry* 139: 216–31.
- Burns, R Stanley, Chuang C Chiueh, Sanford P Markey, Michael H Ebert, David M Jacobowitz, and Irwin J Kopin. 1983. “A Primate Model of Parkinsonism: Selective Destruction of Dopaminergic Neurons in the Pars Compacta of the Substantia Nigra by n-Methyl-4-Phenyl-1, 2, 3, 6-Tetrahydropyridine.” *Proceedings of the National Academy of Sciences* 80 (14): 4546–50.
- Caggiu, Elisa, Kai Paulus, Giuseppe Mameli, Giannina Arru, Gian Pietro Sechi, and Leonardo A Sechi. 2018. “Differential Expression of miRNA 155 and miRNA 146a in Parkinson’s Disease Patients.” *Eneurologicalsci* 13: 1–4.
- Castellani, Rudy J, Sandra L Siedlak, George Perry, and Mark A Smith. 2000. “Sequestration of Iron by Lewy Bodies in Parkinson’s Disease.” *Acta Neuropathologica* 100: 111–14.
- Cataldi, Samuela, Michela Codini, Stéphane Hunot, François-Pierre Légeron, Ivana Ferri, Paola Siccu, Angelo Sidoni, et al. 2016. “E-Cadherin in 1-Methyl-4-Phenyl-1, 2, 3, 6-Tetrahydropyridine-Induced Parkinson Disease.” *Mediators of Inflammation* 2016 (1): 3937057.
- Chaturvedi, Rajnish K, and M Flint Beal. 2008. “Mitochondrial Approaches for Neuroprotection.” *Annals of the New York Academy of Sciences* 1147 (1): 395–412.
- Chen, Xi, Yi Ba, Lijia Ma, Xing Cai, Yuan Yin, Kehui Wang, Jigang Guo, et al. 2008. “Characterization of microRNAs in Serum: A Novel Class of Biomarkers for Diagnosis of Cancer and Other Diseases.” *Cell Research* 18 (10): 997–1006.
- Chong, Mark MW, Guoan Zhang, Sihem Cheloufi, Thomas A Neubert, Gregory J Hannon, and Dan R Littman. 2010. “Canonical and Alternate Functions of the microRNA Biogenesis Machinery.” *Genes & Development* 24 (17): 1951–60.

- Cullen, Bryan R. 2004. "Transcription and Processing of Human microRNA Precursors." *Molecular Cell* 16 (6): 861–65.
- Dinareello, Charles A. 2018. "Overview of the IL-1 Family in Innate Inflammation and Acquired Immunity." *Immunological Reviews* 281 (1): 8–27.
- Do Van, Bruce, Flore Gouel, Aurélie Jonneaux, Kelly Timmerman, Patrick Gelé, Maud Pétrault, Michèle Bastide, et al. 2016. "Ferroptosis, a Newly Characterized Form of Cell Death in Parkinson's Disease That Is Regulated by PKC." *Neurobiology of Disease* 94: 169–78.
- Dong-Chen, Xu, Chen Yong, Xu Yang, ShenTu Chen-Yu, and Peng Li-Hua. 2023. "Signaling Pathways in Parkinson's Disease: Molecular Mechanisms and Therapeutic Interventions." *Signal Transduction and Targeted Therapy* 8 (1): 73.
- Dorsey, E, Todd Sherer, Michael S Okun, and Bastiaan R Bloem. 2018. "The Emerging Evidence of the Parkinson Pandemic." *Journal of Parkinson's Disease* 8 (s1): S3–8.
- Dorszewska, Jolanta, Marta Kowalska, Michał Prendecki, Thomas Piekut, Joanna Kozłowska, and Wojciech Kozubski. 2021. "Oxidative Stress Factors in Parkinson's Disease." *Neural Regeneration Research* 16 (7): 1383–91.
- Einem, Bjoern von, Judith Eschbach, Martin Kiechle, Anke Wahler, Dietmar R Thal, Pamela J McLean, Jochen H Weishaupt, Albert C Ludolph, Christine AF von Arnim, and Karin M Danzer. 2017. "The Golgi-Localized, Gamma Ear-Containing, ARF-Binding (GGA) Protein Family Alters Alpha Synuclein (α -Syn) Oligomerization and Secretion." *Aging (Albany NY)* 9 (7): 1677.
- Feng, Yi, Pei-Yan He, Wei-Dong Kong, Wan-Jing Cen, Peng-Lin Wang, Chang Liu, Wu Zhang, Shu-Shu Li, and Jian-Wei Jiang. 2021. "Apoptosis-Promoting Properties of miR-3074-5p in MC3T3-E1 Cells Under Iron Overload Conditions." *Cellular & Molecular Biology Letters* 26: 1–15.
- Ferreira, Sara A, and Marina Romero-Ramos. 2018. "Microglia Response During Parkinson's Disease: Alpha-Synuclein Intervention." *Frontiers in Cellular Neuroscience* 12: 247.
- Fisher, Ronald A. 1922. "On the Interpretation of χ^2 from Contingency Tables, and the Calculation of p." *Journal of the Royal Statistical Society* 85 (1): 87–94.
- Frédéric, Pierre-Marc, and Martin J Simard. 2022. "Regulation and Different Functions of the Animal microRNA-Induced Silencing Complex." *Wiley Interdisciplinary Reviews: RNA* 13 (4): e1701.
- Friedman, Jerome, Robert Tibshirani, and Trevor Hastie. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1–22. <https://doi.org/10.18637/jss.v033.i01>.
- Fury, Wen, Franak Batliwalla, Peter K Gregersen, and Wentian Li. 2006. "Overlapping Probabilities of Top Ranking Gene Lists, Hypergeometric Distribution, and Stringency of Gene Selection Criterion." In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 5531–34. IEEE.
- Galiano-Landeira, Jordi, Albert Torra, Miquel Vila, and Jordi Bove. 2020. "CD8 t Cell Nigral Infiltration Precedes Synucleinopathy in Early Stages of Parkinson's Disease." *Brain* 143 (12): 3717–33.
- Goldman, Samuel M, Kenneth Marek, Ruth Ottman, Cheryl Meng, Kathleen Comyns, Piu Chan, Jinghong Ma, et al. 2019. "Concordance for Parkinson's Disease in Twins: A 20-Year Update." *Annals of Neurology* 85 (4): 600–605.
- Graves, Paul, and Yan Zeng. 2012. "Biogenesis of Mammalian microRNAs: A Global View." *Genomics, Proteomics and Bioinformatics* 10 (5): 239–45.
- Grossi, Ilaria, Annalisa Radeghieri, Lucia Paolini, Vanessa Porrini, Andrea Pilotto, Alessandro Padovani, Alessandra Marengoni, et al. 2021. "MicroRNA-34a-5p Expression in the Plasma and in Its Extracellular Vesicle Fractions in Subjects with Parkinson's Disease: An Exploratory Study." *International Journal of Molecular Medicine* 47 (2): 533–46.
- Grottemeyer, Alexander, Rhonda Leah McFleder, Jingjing Wu, Jörg Wischhusen, and Chi Wang Ip. 2022. "Neuroinflammation in Parkinson's Disease—Putative Pathomechanisms and Targets for Disease-Modification." *Frontiers in Immunology* 13: 878771.
- Gu, Yan, Yan Shi, Qian Yang, Wen-Wen Gu, Ya-Ping He, Hua-Jun Zheng, Xuan Zhang, Jian-Mei Wang, and Jian Wang. 2018. "miR-3074-5p Promotes the Apoptosis but Inhibits the Invasiveness of Human Extravillous Trophoblast-Derived HTR8/SVneo Cells in Vitro." *Reproductive Sciences* 25 (5): 690–99.
- Hajam, Younis Ahmad, Raksha Rani, Shahid Yousuf Ganie, Tariq Ahmad Sheikh, Darakhshan Javaid, Syed Sanobar Qadri, Sreepoorna Pramodh, et al. 2022. "Oxidative Stress in Human Pathology and Aging: Molecular Mechanisms and Perspectives." *Cells* 11 (3): 552.

- Hiers, Nicholas M, Tianqi Li, Conner M Traugot, and Mingyi Xie. 2024. "Target-Directed microRNA Degradation: Mechanisms, Significance, and Functional Implications." *Wiley Interdisciplinary Reviews: RNA* 15 (2): e1832.
- Huang, Bin, and Suk-Hwan Baek. 2017. "Trim13 Potentiates Toll-Like Receptor 2-Mediated Nuclear Factor κ b Activation via K29-Linked Polyubiquitination of Tumor Necrosis Factor Receptor-Associated Factor 6." *Molecular Pharmacology* 91 (4): 307–16.
- Jay, Taylor R, Victoria E von Saucken, and Gary E Landreth. 2017. "TREM2 in Neurodegenerative Diseases." *Molecular Neurodegeneration* 12: 1–33.
- Jha, Saurabh Kumar, Niraj Kumar Jha, Rohan Kar, Rashmi K Ambasta, and Pravir Kumar. 2015. "P38 MAPK and PI3K/AKT Signalling Cascades in Parkinson's Disease." *International Journal of Molecular and Cellular Medicine* 4 (2): 67.
- Junker, Andreas, Markus Krumbholz, Sylvia Eisele, Hema Mohan, Florian Augstein, Robert Bittner, Hans Lassmann, Hartmut Wekerle, Reinhard Hohlfeld, and Edgar Meinl. 2009. "MicroRNA Profiling of Multiple Sclerosis Lesions Identifies Modulators of the Regulatory Protein CD47." *Brain* 132 (12): 3342–52.
- Kastaniotis, Alexander J, Kaija J Autio, and Remya R. Nair. 2021. "Mitochondrial Fatty Acids and Neurodegenerative Disorders." *The Neuroscientist* 27 (2): 143–58.
- Kaushik, Susmita, and Ana Maria Cuervo. 2015. "Proteostasis and Aging." *Nature Medicine* 21 (12): 1406–15.
- Khazaei-Poul, Yalda, Seyed Amir Mirmotalebisohi, Hakimeh Zali, Zahra Molavi, and Samira Mohammadi-Yeganeh. 2023. "Identification of miR-3182 and miR-3143 Target Genes Involved in the Cell Cycle as a Novel Approach in TNBC Treatment: A Systems Biology Approach." *Chemical Biology & Drug Design* 101 (3): 662–77.
- Kim, Changyoun, and Seung-Jae Lee. 2008. "Controlling the Mass Action of α -Synuclein in Parkinson's Disease." *Journal of Neurochemistry* 107 (2): 303–16.
- Kim, Eun Kyung, and Eui-Ju Choi. 2010. "Pathological Roles of MAPK Signaling Pathways in Human Diseases." *Biochimica Et Biophysica Acta (BBA)-Molecular Basis of Disease* 1802 (4): 396–405.
- Klivenyi, Peter, Donald Siwek, Gabrielle Gardian, Lichuan Yang, Anatoly Starkov, Carine Cleren, Robert J Ferrante, Neil W Kowall, Asa Abeliovich, and M Flint Beal. 2006. "Mice Lacking Alpha-Synuclein Are Resistant to Mitochondrial Toxins." *Neurobiology of Disease* 21 (3): 541–48.
- Köllicker-Frers, Rodolfo, Lucas Udovin, Matilde Otero-Losada, Tamara Kobiec, María Inés Herrera, Jorge Palacios, Gabriela Razzitte, and Francisco Capani. 2021. "Neuroinflammation: An Integrating Overview of Reactive-Neuroimmune Cell Interactions in Health and Disease." *Mediators of Inflammation* 2021 (1): 9999146.
- Kopin, IJ. 1987. "MPTP: An Industrial Chemical and Contaminant of Illicit Narcotics Stimulates a New Era in Research on Parkinson's Disease." *Environmental Health Perspectives* 75: 45–51.
- Kozomara, Ana, Maria Birgaoanu, and Sam Griffiths-Jones. 2019. "miRBase: From microRNA Sequences to Function." *Nucleic Acids Research* 47 (D1): D155–62.
- Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. <https://doi.org/10.18637/jss.v028.i05>.
- Lai, Mingshuang, Meiliang Liu, Deyuan Li, Ruirui Zhang, Lijun Wang, Xiaofei Li, Sixian Chen, et al. 2023. "circELMOD3 Increases and Stabilizes TRIM13 by Sponging miR-6864-5p and Direct Binding to Inhibit HCC Progression." *Iscience* 26 (10).
- Lanneau, D, M Brunet, E Frisan, Eric Solary, M Fontenay, and Carmen Garrido. 2008. "Heat Shock Proteins: Essential Proteins for Apoptosis Regulation." *Journal of Cellular and Molecular Medicine* 12 (3): 743–61.
- Laub, Friedrich, Lei Lei, Hideaki Sumiyoshi, Daisuke Kajimura, Cecilia Dragomir, Silvia Smaldone, Adam C Puche, et al. 2005. "Transcription Factor KLF7 Is Important for Neuronal Morphogenesis in Selected Regions of the Nervous System." *Molecular and Cellular Biology* 25 (13): 5699–5711.
- Li, Haining, Juan Yang, Yanbai Wang, Qiang Liu, Jiang Cheng, and Feng Wang. 2019. "Neuroprotective Effects of Increasing Levels of HSP70 Against Neuroinflammation in Parkinson's Disease Model by Inhibition of NF- κ b and STAT3." *Life Sciences* 234: 116747.
- Liu, Guiyou, Jing-Yi Sun, Meiling Xu, Xiao-Yi Yang, and Bao-Liang Sun. 2017. "SORL1 Variants Show Different Association with Early-Onset and Late-Onset Alzheimer's Disease Risk." *Journal of Alzheimer's*

- Disease* 58 (4): 1121–28.
- Lu, Liu, Wen-Zhuo Dai, Xi-Chen Zhu, and Tao Ma. 2021. “Analysis of Serum miRNAs in Alzheimer’s Disease.” *American Journal of Alzheimer’s Disease & Other Dementias* 36: 15333175211021712.
- Lui, Austin, Timothy Do, Omar Alzayat, Nina Yu, Su Phyu, Hillary Joy Santuya, Benjamin Liang, et al. 2024. “Tumor Suppressor MicroRNAs in Clinical and Preclinical Trials for Neurological Disorders.” *Pharmaceuticals* 17 (4): 426.
- Marek, Kenneth, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, et al. 2011. “The Parkinson Progression Marker Initiative (PPMI).” *Progress in Neurobiology* 95 (4): 629–35.
- Maries, Eleonora, Biplob Dass, Timothy J Collier, Jeffrey H Kordower, and Kathy Steece-Collier. 2003. “The Role of α -Synuclein in Parkinson’s Disease: Insights from Animal Models.” *Nature Reviews Neuroscience* 4 (9): 727–38.
- Markovic, Milica, Pravin Yeapuri, Krista L Namminga, Yaman Lu, Maamoon Saleh, Katherine E Olson, Howard E Gendelman, and R Lee Mosley. 2022. “Interleukin-2 Expands Neuroprotective Regulatory t Cells in Parkinson’s Disease.” *Neuroimmune Pharmacology and Therapeutics* 1 (1): 43–50.
- Martins, Madalena, Alexandra Rosa, Leonor C Guedes, Benedita V Fonseca, Kristina Gotovac, Sara Violante, Tiago Mestre, et al. 2011. “Convergence of miRNA Expression Profiling, α -Synuclein Interacton and GWAS in Parkinson’s Disease.” *PloS One* 6 (10): e25443.
- McNemar, Quinn. 1947. “Note on the Sampling Error of the Difference Between Correlated Proportions or Percentages.” *Psychometrika* 12 (2): 153–57.
- Meara, Jolyon, and Peter Hobson. 2018. “Epidemiology of Parkinson’s Disease.” *Parkinson’s Disease in the Older Patient*, 30–38.
- Mitchell, Patrick S, Rachael K Parkin, Evan M Kroh, Brian R Fritz, Stacia K Wyman, Era L Pogossova-Agadjanian, Amelia Peterson, et al. 2008. “Circulating microRNAs as Stable Blood-Based Markers for Cancer Detection.” *Proceedings of the National Academy of Sciences* 105 (30): 10513–18.
- “National Center for Biotechnology Information(NCBI) Search Engine.” n.d. <https://www.ncbi.nlm.nih.gov/search/>.
- Navarro-Romero, Alba, Marta Montpeyó, and Marta Martinez-Vicente. 2020. “The Emerging Role of the Lysosome in Parkinson’s Disease.” *Cells* 9 (11): 2399.
- Noelker, Carmen, Lydie Morel, Thomas Lescot, Anke Osterloh, Daniel Alvarez-Fischer, Minka Breloer, Carmen Henze, et al. 2013. “Toll Like Receptor 4 Mediates Cell Death in a Mouse MPTP Model of Parkinson Disease.” *Scientific Reports* 3 (1): 1393.
- Nuzziello, Nicoletta, and Maria Liguori. 2019. “The microRNA Centrism in the Orchestration of Neuroinflammation in Neurodegenerative Diseases.” *Cells* 8 (10): 1193.
- Obergasteiger, Julia, Giulia Frapporti, Peter P Pramstaller, Andrew A Hicks, and Mattia Volta. 2018. “A New Hypothesis for Parkinson’s Disease Pathogenesis: GTPase-P38 MAPK Signaling and Autophagy as Convergence Points of Etiology and Genomics.” *Molecular Neurodegeneration* 13 (1): 40.
- Oliveira, Sara R, Pedro A Dionísio, Leonor Correia Guedes, Nilza Gonçalves, Miguel Coelho, Mário M Rosa, Joana D Amaral, Joaquim J Ferreira, and Cecília MP Rodrigues. 2020. “Circulating Inflammatory miRNAs Associated with Parkinson’s Disease Pathophysiology.” *Biomolecules* 10 (6): 945.
- Ortiz-Quintero, Blanca. 2016. “Cell-Free microRNAs in Blood and Other Body Fluids, as Cancer Biomarkers.” *Cell Proliferation* 49 (3): 281–303.
- Pacheco, Rodrigo. 2021. “T-Cell Based Immunotherapies for Parkinsons Disease.” *Exploration of Neuroprotective Therapy* 1 (2): 72–85.
- Pallarès-Albanell, Joan, M Teresa Zomeño-Abellán, Georgia Escaramís, Lorena Pantano, Aroa Soriano, Miguel F Segura, and Eulàlia Martí. 2019. “A High-Throughput Screening Identifies MicroRNA Inhibitors That Influence Neuronal Maintenance and/or Response to Oxidative Stress.” *Molecular Therapy-Nucleic Acids* 17: 374–87.
- Parkinson, James. 2002. “An Essay on the Shaking Palsy.” *The Journal of Neuropsychiatry and Clinical Neurosciences* 14 (2): 223–36.
- Patil, Ketan S, Indranil Basak, Ingvald Dalen, Esthelle Hoedt, Johannes Lange, Kristin A Lunde, Ying Liu, et al. 2019. “Combinatory microRNA Serum Signatures as Classifiers of Parkinson’s Disease.” *Parkinsonism & Related Disorders* 64: 202–10.
- Pillai, Ramesh S. 2005. “MicroRNA Function: Multiple Mechanisms for a Tiny RNA?” *Rna* 11 (12): 1753–

- Pinter, Bernadette, Anja Diem-Zangerl, Gregor Karl Wenning, Christoph Scherfler, Willi Oberaigner, Klaus Seppi, and Werner Poewe. 2015. "Mortality in Parkinson's Disease: A 38-Year Follow-up Study." *Movement Disorders* 30 (2): 266–69.
- Plotnikova, Olga, Ancha Baranova, and Mikhail Skoblov. 2019. "Comprehensive Analysis of Human microRNA–mRNA Interactome." *Frontiers in Genetics* 10: 933.
- Pu, Mengfan, Jing Chen, Zhouteng Tao, Lingling Miao, Xinming Qi, Yizheng Wang, and Jin Ren. 2019. "Regulatory Network of miRNA on Its Target: Coordination Between Transcriptional and Post-Transcriptional Regulation of Gene Expression." *Cellular and Molecular Life Sciences* 76: 441–51.
- Quadri, Marialuisa, Wim Mandemakers, Martyna M Grochowska, Roy Masius, Hanneke Geut, Editio Fabrizio, Guido J Breedveld, et al. 2018. "LRP10 Genetic Variants in Familial Parkinson's Disease and Dementia with Lewy Bodies: A Genome-Wide Linkage and Sequencing Study." *The Lancet Neurology* 17 (7): 597–608.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ranstam, Jonas, and Jonathan A Cook. 2018. "LASSO Regression." *Journal of British Surgery* 105 (10): 1348–48.
- Raver-Shapira, Nina, Efi Marciano, Eti Meiri, Yael Spector, Nitzan Rosenfeld, Neta Moskovits, Zvi Bentwich, and Moshe Oren. 2007. "Transcriptional Activation of miR-34a Contributes to P53-Mediated Apoptosis." *Molecular Cell* 26 (5): 731–43.
- Rigatti, Steven J. 2017. "Random Forest." *Journal of Insurance Medicine* 47 (1): 31–39.
- "RNAcentral 2021: Secondary Structure Integration, Improved Sequence Search and New Member Databases." 2021. *Nucleic Acids Research* 49 (D1): D212–20.
- Ross, Jeffrey S, J Andrew Carlson, and Graham Brock. 2007. "miRNA: The New Gene Silencer." *American Journal of Clinical Pathology* 128 (5): 830–36.
- RStudio Team. 2020. *RStudio: Integrated Development Environment for r*. Boston, MA: RStudio, PBC. <http://www.rstudio.com/>.
- Saiki, Shinji, Taku Hatano, Motoki Fujimaki, Kei-Ichi Ishikawa, Akio Mori, Yutaka Oji, Ayami Okuzumi, et al. 2017. "Decreased Long-Chain Acylcarnitines from Insufficient β -Oxidation as Potential Early Diagnostic Markers for Parkinson's Disease." *Scientific Reports* 7 (1): 7328.
- Savica, Rodolfo, Lisa A Cannon-Albright, and Stefan Pulst. 2016. "Familial Aggregation of Parkinson Disease in Utah: A Population-Based Analysis Using Death Certificates." *Neurology Genetics* 2 (2).
- Sawada, M, K Imamura, and T Nagatsu. 2006. "Role of Cytokines in Inflammatory Process in Parkinson's Disease." *Parkinson's Disease and Related Disorders*, 373–81.
- Sharma, Manu, and Jacqueline Burré. 2023. " α -Synuclein in Synaptic Function and Dysfunction." *Trends in Neurosciences* 46 (2): 153–66.
- Sharma, Neha, and Bimla Nehru. 2016. "Apocyanin, a Microglial NADPH Oxidase Inhibitor Prevents Dopaminergic Neuronal Degeneration in Lipopolysaccharide-Induced Parkinson's Disease Model." *Molecular Neurobiology* 53: 3326–37.
- Sheikh, M Saeed, and Ying Huang. 2003. "Death Receptor Activation Complexes: It Takes Two to Activate TNF Receptor 1." *Cell Cycle* 2 (6): 549–51.
- Simon, David K, Caroline M Tanner, and Patrik Brundin. 2020. "Parkinson Disease Epidemiology, Pathology, Genetics, and Pathophysiology." *Clinics in Geriatric Medicine* 36 (1): 1–12.
- Sommer, Annika, Franz Marxreiter, Florian Krach, Tanja Fadler, Janina Grosch, Michele Maroni, Daniela Graef, et al. 2018. "Th17 Lymphocytes Induce Neuronal Cell Death in a Human iPSC-Based Model of Parkinson's Disease." *Cell Stem Cell* 23 (1): 123–31.
- Stöckel, Daniel, Tim Kehl, Patrick Trampert, Lara Schneider, Christina Backes, Nicole Ludwig, Andreas Gerasch, et al. 2016. "Multi-Omics Enrichment Analysis Using the GeneTrail2 Web Service." *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btv770>.
- Subramanian, Aravind, Pablo Tamayo, Vamsi K Mootha, Sayan Mukherjee, Benjamin L Ebert, Michael A Gillette, Amanda Paulovich, et al. 2005. "Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles." *Proceedings of the National Academy of Sciences* 102 (43): 15545–50.
- Sung, Victor W, and Anthony P Nicholas. 2013. "Nonmotor Symptoms in Parkinson's Disease: Expanding

- the View of Parkinson’s Disease Beyond a Pure Motor, Pure Dopaminergic Problem.” *Neurologic Clinics* 31 (3): S1–16.
- Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie. 2023. “Elastic Net Regularization Paths for All Generalized Linear Models.” *Journal of Statistical Software* 106 (1): 1–31. <https://doi.org/10.18637/jss.v106.i01>.
- Van den Hove, Daniel L, Konstantinos Kompotis, Roy Lardenoije, Gunter Kenis, Jonathan Mill, Harry W Steinbusch, Klaus-Peter Lesch, Carlos P Fitzsimons, Bart De Strooper, and Bart PF Rutten. 2014. “Epigenetically Regulated microRNAs in Alzheimer’s Disease.” *Neurobiology of Aging* 35 (4): 731–45.
- Wang, Xiufeng, Zhijun Liu, and Fang Wang. 2021. “MicroRNA-93 Blocks Signal Transducers and Activator of Transcription 3 to Reduce Neuronal Damage in Parkinson’s Disease.” *Neurochemical Research* 46 (7): 1859–68.
- Wang, Ying, Mengting Luan, Li Xue, Jianing Jin, and Anmu Xie. 2022. “Evaluation of the Relationship Between SORL1 Gene Polymorphism and Parkinson’s Disease in the Chinese Population.” *Neuroscience Letters* 778: 136602.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Williams, Erin T, Xi Chen, and Darren J Moore. 2017. “VPS35, the Retromer Complex and Parkinson’s Disease.” *Journal of Parkinson’s Disease* 7 (2): 219–33.
- Wright, Marvin N., and Andreas Ziegler. 2017. “ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R.” *Journal of Statistical Software* 77 (1): 1–17. <https://doi.org/10.18637/jss.v077.i01>.
- Yang, Jin, Liu-Lin Xiong, You-Cui Wang, Xiang He, Ling Jiang, Song-Jun Fu, Xue-Fei Han, Jia Liu, and Ting-Hua Wang. 2018. “Oligodendrocyte Precursor Cell Transplantation Promotes Functional Recovery Following Contusive Spinal Cord Injury in Rats and Is Associated with Altered microRNA Expression.” *Molecular Medicine Reports* 17 (1): 771–82.
- Yang, Qinghe, Wei Pan, and Liwei Qian. 2017. “Identification of the miRNA–mRNA Regulatory Network in Multiple Sclerosis.” *Neurological Research* 39 (2): 142–51.
- Zhao, Jingru, Hebo Wang, Lipeng Dong, Sujuan Sun, and Litao Li. 2019. “miRNA-20b Inhibits Cerebral Ischemia-Induced Inflammation Through Targeting NLRP3.” *International Journal of Molecular Medicine* 43 (3): 1167–78.
- Zilberter, Yuri, Dennis R Tabuena, and Misha Zilberter. 2023. “NOX-Induced Oxidative Stress Is a Primary Trigger of Major Neurodegenerative Disorders.” *Progress in Neurobiology*, 102539.