



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIERÍA
INDUSTRIAL VALENCIA

Curso Académico:

AGRADECIMIENTOS

En primer lugar, me gustaría agradecer a mis tutores los doctores Óscar Pastor López y Ana León Palacio, y al centro de investigación PROS, por el apoyo y consejos para llevar a cabo este TFG que tanto he disfrutado realizando. Gracias a vosotros he podido descubrir el apasionante mundo de la investigación relacionada con el análisis de datos genómicos, a la que espero poder dedicarme en un futuro.

A mis compañeras y amigas por hacer que estos últimos años hayan sido mucho más llevaderos, fáciles y sobre todo divertidos.

Agradecer también a mi familia, pues sin ellos no sería la persona que soy ahora.

Finalmente agradecer a mi pareja, Álvaro, el gran apoyo que ha sido para mí durante este año de convivencia. Gracias por todo lo que haces por mí cada día, estoy segura de que este trabajo no habría sido lo mismo sin tu apoyo.

RESUMEN

El Alzheimer es la forma de demencia más común en personas mayores, sin embargo, también existe una forma de Alzheimer que aparece en personas más jóvenes (*Early Onset Alzheimer Disease*, EOAD) y con una mayor dificultad de diagnóstico debido a su temprana manifestación. Debido a la gran influencia que tiene la componente genética en el desarrollo de esta enfermedad, identificar variaciones genéticas relevantes es crucial para poder facilitar su diagnóstico.

Estas variaciones, junto con la información relacionada, se almacenan en repositorios y bases de datos. Gracias a las técnicas de secuenciación de nueva generación, la información genética disponible ha crecido de forma masiva. Este crecimiento masivo, junto con la complejidad de la información y los problemas de calidad existentes (ausencia de interpretación de las variaciones, falta de evidencias sobre el criterio utilizado para interpretarlas, conflictos entre expertos, etc) hace necesario la utilización de métodos que faciliten la selección de la información relevante para el diagnóstico.

Con el objetivo de identificar variaciones genéticas relevantes para una enfermedad, en el centro de I+D PROS se ha desarrollado el método SILE. Este método busca sistematizar el proceso de identificación, analizando la información almacenada en las bases de datos y repositorios con el fin de seleccionar únicamente aquella que tenga la calidad suficiente para ser utilizada en un ámbito clínico. Este método fue aplicado en un trabajo previo sobre el Alzheimer y no se encontró ninguna variación genética relevante para el caso particular del Alzheimer temprano.

Debido a esto, como primer objetivo de este trabajo, se ha buscado realizar un análisis exhaustivo de la etapa de identificación de este método, tras lo que se encontraron algunos problemas para tratar casos específicos como el que aquí se aborda, caracterizados por su baja incidencia y marcado componente familiar. Derivado de este análisis se encuentra el segundo objetivo de este trabajo, donde se ha buscado mejorar la etapa de identificación de variaciones genéticas de SILE, adaptándolo por tanto a casos específicos como el Alzheimer temprano.

Finalmente, en un tercer objetivo se ha buscado evaluar el workflow diseñado mediante su aplicación al caso del Alzheimer temprano, obteniendo como resultado un total de 29 variaciones genéticas identificadas como relevantes (25 con una evidencia limitada y 4 con una evidencia moderada).

Palabras Clave: Método SILE, Alzheimer temprano, diagnóstico genético

RESUM

L'Alzheimer és la forma de demència més comuna en persones majors, no obstant això, també hi ha una forma d'Alzheimer que apareix en persones més joves (*Early Onset Alzheimer Disease*, EOAD) i amb una major dificultat de diagnòstic degut a la seua temprana manifestació. A causa de la gran influència que té la component genètica en el desenvolupament d'esta malaltia, identificar variacions genètiques rellevants és crucial per a poder facilitar el diagnòstic d'esta manera de la malaltia.

Estes variacions, junt amb la informació relacionada, s'emmagatzemen en repositories i bases de dades. Gràcies a les tècniques de seqüenciació de nova generació, la informació genètica disponible ha crescut de forma massiva. Este creixement massiu, junt amb la complexitat de la informació i els problemes de qualitat existents (absència d'interpretació de les variacions, falta d'evidències sobre el criteri utilitzat per a interpretar-les, conflictes entre experts, etc) fa necessari la utilització de mètodes que faciliten la selecció de la informació rellevant per al diagnòstic.

Amb l'objectiu d'identificar variacions genètiques rellevants per a una malaltia, en el centre d'I+D PROS s'ha creat el mètode SILE. Este mètode busca sistematitzar el procés d'identificació, analitzant la informació emmagatzemada en les bases de dades i reposadors a fi de seleccionar únicament aquella que tinga la qualitat suficient per a ser utilitzada en un àmbit clínic. Este mètode va ser aplicat en un treball previ sobre l'Alzheimer i no es va trobar cap variació genètica rellevant per al cas particular de l'Alzheimer d'aparició prematura.

A causa d'açò, com a primer objectiu d'este treball, s'ha buscat realitzar una anàlisi exhaustiva de l'etapa d'identificació d'este mètode, després del que es van trobar alguns problemes per a tractar casos específics com el que açí es tracta, caracteritzat per la seua baixa incidència i un marcat component familiar. Derivat d'esta anàlisi es troba el segon objectiu d'este treball, on s'ha buscat millorar l'etapa d'identificació de variacions genètiques de SILE adaptant-lo per tant a casos específics com l'Alzheimer d'aparició prematura.

Finalment, en un tercer objectiu s'ha buscat avaluar el workflow dissenyat amb la seua aplicació al cas de l'Alzheimer d'aparició prematura, obtenint com resultat un total de 29 variacions genètiques identificades com rellevants (25 amb una evidència limitada i 4 amb una evidència moderada).

Paraules Clau: Mètode SILE, Alzheimer d'aparició prematura, diagnòstic genètic

ABSTRACT

The Alzheimer disease (AD) is the most common type of dementia in the elderly, however, there is also an Early Onset form which is more difficult to diagnose precisely because it is early development. As the genetic component has a major influence in the development on the disease, determining relevant genetic variants is crucial to facilitate its diagnosis.

These variants together with their information, are stored in repositories and databases. Thanks to the new generation sequencing techniques, the available genetic information has grown massively. This massive growth, together with the information complexity and the existing quality problems (lack of interpretation of the variants, lack of evidence about the criteria used to interpret them, conflicts between experts, etc) makes it necessary to use methods that facilitate the selection of the relevant information for the diagnosis.

In order to identify relevant genetic variants for a disease, the SILE method has been developed at the PROS R&D centre. This method seeks to systematize the process of identification through the analysis of the information stored in databases and repositories, with the objective of selecting only the information that is of sufficient quality to be used in a clinical setting. This method was applied in a previous work on Alzheimer's and no relevant genetic variation was found for the particular case of Early Onset Alzheimer Disease (EOAD).

Due to this, as the first objective of this work, we have sought to carry out an exhaustive analysis of the identification stage of this method. Thanks to this analysis, some problems were found to treat specific cases such as EOAD, which is characterized by its low incidence and relevant family component. Derived from this analysis is the second objective of this work, which has sought to improve the identification stage of the SILE method for adapting it to specific cases as EOAD.

Finally, a third objective is to evaluate the workflow designed by its application to EOAD. As a result of this application, a set of 29 variants has been identified (25 variants accepted with a limited evidence a 4 accepted with moderate evidence).

Keywords: SILE Method, Early Onset Alzheimer Disease, Genetic Diagnosis.

ÍNDICE

Documentos Contenidos en el TFG

- Memoria
- Presupuesto
- Anexos

Índice de la Memoria

1. INTRODUCCIÓN.....	1
1.1. Marco de Aplicación: Medicina de Precisión.	1
1.2. Big Data vs. Smart Data.	2
1.3. Introducción al Método SILE.....	4
1.4. Trabajos Relacionados	5
1.5. Estructura del Trabajo	6
2. OBJETIVOS Y STAKEHOLDERS	7
2.1. Objetivos	7
2.2. Stakeholders	8
3. MÉTODO SILE	9
3.1. Búsqueda.....	9
3.2. Identificación.....	9
3.3. Carga	12
3.4. Explotación.....	13
4. FENOTIPO ESTUDIADO: ALZHEIMER TEMPRANO.....	14
4.1. Características del Alzheimer Temprano	15
4.2. Aplicación del Método SILE al EOAD.....	19
5. FUENTES DE INFORMACIÓN	20
5.1. Adecuación de ClinVar a los Criterios Establecidos en la Etapa de Búsqueda del Método SILE	22
5.2. Información Proporcionada por ClinVar	23
5.2.1. Tipos de Submitters	23
5.2.2. Significado Clínico (Clinical Significance)	24
5.2.3. Estado de Revisión (Review Status)	24
5.2.4. Tipo de Método Utilizado	25
6. ANÁLISIS Y PROPUESTA DE MEJORA DE LA ETAPA DE IDENTIFICACIÓN DEL MÉTODO SILE	26

6.1. Análisis de la Etapa de Identificación del Método SILE.....	26
6.2. Solución al problema 2. Definición de los Estudios Familiares.....	30
6.3. Nuevo Workflow Propuesto para la Etapa de Identificación del método SILE.....	33
6.3.1..Solución al problema 3. Tratamiento de las Variaciones con Significado Clínico Conflicto de Interpretación (pasos F10 a F12).....	36
6.3.2.Solución al problema 4.1. Selección de las Variaciones con los Submitters más Importantes (paso F3 y F7)	37
6.3.3.Solución al problema 4.2. Evaluación del Tipo de Método Utilizado para Obtener la Información de la Variación (paso F5)	38
6.3.4.Solución al problema 5. Comprobación de si el Gen donde se Encuentra Localizada la Variación Genética Está Relacionado con la Enfermedad (paso F4).....	39
6.3.5.Solución al problema 1 y 4.3. Verificar el Assertion Criteria Utilizado para la Interpretación de la Variación (pasos F6 y F8).....	40
6.3.6.Solución al problema 6. Evaluación de la Última Fecha en la que Ha Sido Revisada la Variación (F9).	43
6.3.7.Inclusión de Clasificación Seguimiento en el Nuevo Workflow	44
6.3.8.Evaluación de las Variaciones que No Dispongan de un Significado Clínico (paso F4, F13).....	45
7. APLICACIÓN DE LA MEJORA PROPUESTA PARA LA ETAPA DE IDENTIFICACIÓN: RESULTADOS Y DISCUSIÓN.....	47
7.1. Resultados de la Búsqueda de Variaciones Genéticas Relacionadas con el EOAD en ClinVar.....	47
7.2. Resultados de la Aplicación del Workflow Propuesto al Caso del EOAD	53
7.3. Discusión.....	57
8. CONCLUSIONES.....	60
9. TRABAJOS FUTUROS.....	62
10. BIBLIOGRAFÍA.....	64

Índice del Presupuesto

1. NECESIDAD DEL PRESUPUESTO.....	1
2. PRESUPUESTO DESGLOSADO.....	1
2.1. Coste de Personal	1
2.2. Costes de Software	2
2.3. Costes de Hardware.....	3
3. PRESUPUESTO TOTAL	3

Índice de los Anexos

Anexo 1. Variaciones Clasificadas Como Aceptadas Tras La Aplicación Del Workflow Propuesto..... 1

Índice de Figuras

Figura 1. Variabilidad genética	1
Figura 2. Número de variaciones en bases de datos conocidas	3
Figura 3. Evolución del Big Data al Smart data	3
Figura 4. Diagrama de flujo para la clasificación de artículos	10
Figura 5. Workflow para la identificación de variaciones genéticas relevantes en el método SILE ...	11
Figura 6. Número de fallecidos por Alzheimer al año	14
Figura 7. Fallecimientos por Alzheimer en 2018 en España distribuido por edades	15
Figura 8. Mecanismos de herencia autosómica dominante	16
Figura 9. Distribución de los diferentes tipos de Alzheimer	17
Figura 10. Divisiones dentro de la enfermedad de Alzheimer	17
Figura 11. Medio celular en el cerebro sin y con presencia de agregados proteicos.....	18
Figura 12. Menú principal de ClinVar	20
Figura 13. Buscador avanzado de ClinVar.....	21
Figura 15. Interpretación dada a la variación rs33949390 en ClinVar.....	27
Figura 16. Fenotipos (Conditions) asociados con la variación rs33949390.....	28
Figura 14. Modificación del diagrama de flujo para la clasificación de artículos.....	32
Figura 17. Workflow diseñado para la etapa de identificación	34
Figura 18. Detalle del workflow con los pasos tratados (F10-F12) resaltados	36
Figura 19. Detalle del workflow con los pasos tratados (F3 y F7) resaltados.....	38
Figura 21. Detalle del workflow con el paso tratado (F5) resaltado	39
Figura 20. Detalle del workflow con el paso tratado (F4) resaltado	40
Figura 22. Detalle del workflow con los pasos tratados (F6 y F8) resaltados.....	41
Figura 23. Combinación de criterios de las guías para la clasificación de las secuencias patogénicas	42
Figura 24. Detalle del workflow con el paso tratado (F9) resaltado	44
Figura 25. Detalle del workflow con los pasos tratados (F4 y F13) resaltados.....	45
Figura 26. Resultados de la búsqueda en ClinVar a fecha 18/05/2020	47

Figura 27. Número de variaciones para cada significado clínico.....	48
Figura 28. Número de variaciones para cada consecuencia molecular	49
Figura 29. Número de variaciones asociadas a cada tipo de método	51
Figura 30. Distribución del número de variaciones en función del gen en el que se localizan.....	51
Figura 31. Resultados de la aplicación del workflow propuesto al EOAD.....	53
Figura 32. Número de variaciones antes y después del paso F5	55
Figura 33. Clasificación final de las variaciones obtenidas en ClinVar.....	56

Índice de Tablas

Documento I: Memoria

Tabla 1. Estado de revisión de las variaciones en ClinVar.....	25
Tabla 2. Palabras clave en los Abstract de Estudios Familiares.....	31
Tabla 3. Número de variaciones obtenidas para cada clasificación posible para el EOAD	56

Documento II: Presupuesto

Tabla 4. Costes de personal	2
Tabla 5. Costes de software	2
Tabla 6. Costes de hardware.....	3
Tabla 7. Cálculo del presupuesto total.....	4

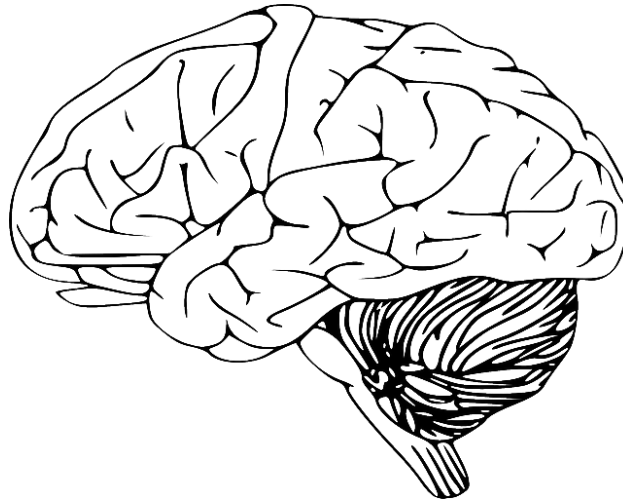
Documento III: Anexos

Tabla 8. Variaciones clasificadas como aceptadas tras la aplicación del workflow propuesto.....	3
--	---

MEMORIA

Diseño y aplicación de un método basado en SILE para la identificación de variaciones genéticas relevantes asociadas a la enfermedad de Alzheimer temprano

Documento I



Mireia Costa Sánchez

Grado en Ingeniería Biomédica

Curso académico 2019/2020

1. INTRODUCCIÓN

Dentro del amplio abanico de ramas de estudio de la ingeniería biomédica, uno de los ámbitos que más relevancia está adquiriendo es el campo de la bioinformática, concretamente, el análisis de datos genómicos.

En una sociedad cada vez más orientada hacia una medicina de precisión, el análisis del genoma, que es lo que nos diferencia como individuos, se ha convertido en un paso fundamental para alcanzar una medicina cada vez más adaptada a las características y necesidades de cada paciente.

En este contexto, este trabajo se ha desarrollado dentro del ámbito de análisis de datos genómicos, buscando determinar aquellas variaciones potencialmente causativas para un fenotipo concreto: el Alzheimer temprano.

A continuación, en este capítulo, se va a entrar en mayor detalle en el marco contextual del trabajo, para poder comprender mejor las dimensiones del problema que se está tratando.

1.1. Marco de Aplicación: Medicina de Precisión.

La medicina de precisión se puede definir como “Una nueva perspectiva de tratamiento y prevención, que tiene en cuenta la variabilidad individual en los genes, el ambiente, y el estilo de vida de cada individuo” [1].

La propia definición ya plantea la necesidad de evaluar cómo afecta la variabilidad genética interindividual en el desarrollo de una enfermedad. En este contexto, se entiende como variación genética una alteración en la secuencia más común del ADN (secuencia de referencia) [2]. Estas variaciones no son siempre patogénicas, pero en algunos casos pueden suponer un aumento en el riesgo de padecimiento de alguna enfermedad genética [3].

De los diferentes tipos de variaciones, los más comunes son los polimorfismos de nucleótido único (*Single Nucleotide Polymorphisms*, SNP), que consisten en un cambio de un único nucleótido en la secuencia de ADN. En la **Figura 1**, se muestra gráficamente el concepto de SNP.

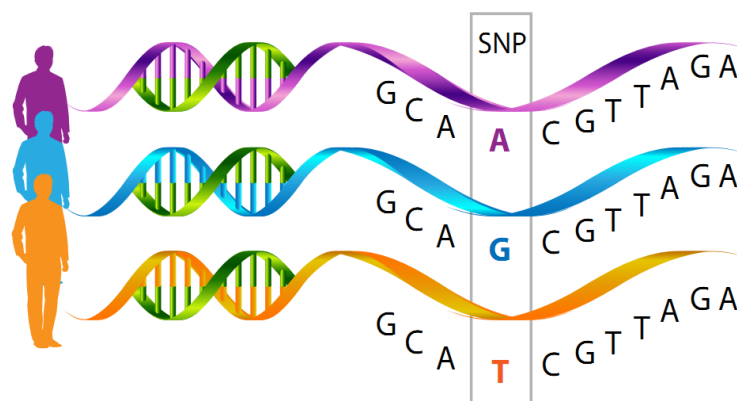


Figura 1. Variabilidad genética

Con esto, la evaluación de la presencia de variaciones genéticas (sean o no SNP), no solo tiene como fin la prevención o el tratamiento, sino que también va a tener una importante componente diagnóstica, que es la que se pretende explotar en este trabajo [4].

Desde que en 2003 se consiguiera secuenciar un genoma completo [5], y gracias a las técnicas de secuenciación de nueva generación que permiten hacerlo en un solo día de forma económica, la cantidad de información genética disponible ha crecido de forma masiva. Así, existen catálogos como ExAC en los que hay disponibles hasta 7.4 millones de variaciones genéticas junto con su frecuencia de aparición en determinadas poblaciones [6].

El problema, por tanto, no reside en la detección de las variaciones, sino en ir más allá e identificar cuál es el efecto que tienen estas variaciones sobre los individuos [7]. Existen variaciones genéticas que son comunes en la población, no estando relacionadas con el desarrollo de ningún fenotipo concreto, pero otras se pueden relacionar con el desarrollo de enfermedades, siendo estas la de interés a nivel clínico. Sin embargo, debido a la gran cantidad de información disponible y a la complejidad de esta, seleccionar aquellas variaciones genéticas relevantes a nivel clínico no resulta trivial. A ello, se añaden problemas en la calidad de la información almacenada en las bases de datos, lo que dificulta su uso en un ámbito clínico. Algunos ejemplos de estos problemas son la ausencia de interpretación de las variaciones, falta de evidencias sobre cuál ha sido el criterio utilizado cuando han sido interpretadas, conflictos entre expertos, etc.

Los problemas identificados en la información disponible limitan su uso a nivel clínico, ya que comprometen su calidad y la fiabilidad. Si se busca aplicar el concepto de medicina de precisión, este no puede sustentarse en información de poca calidad, pues esto podría tener un impacto real en la salud de los pacientes, pudiendo llevar a diagnósticos o tratamientos erróneos.

Por estos problemas nace la necesidad de trabajos como este, desarrollados en el centro de investigación de I+D PROS UPV, que buscan determinar estrategias para discriminar las variaciones que sean relevantes para el diagnóstico de una enfermedad, de las que no lo son.

Debido a las características de la información relacionada con las variaciones genéticas, siendo estas su gran volumen, variedad en la representación, así como su rápido crecimiento, se hace fundamental presentar el concepto de Big Data y Smart Data.

1.2. Big Data vs. Smart Data.

El concepto de Big Data hace referencia al uso del análisis de datos para extraer información que permita tomar mejores decisiones informadas [8]. Los datos que se analizan mediante Big Data se caracterizan por crecer a una gran velocidad y por poseer una gran variedad de información, conceptos que se conocen como las “tres V” (Velocidad, Volumen y Variedad).

Haciendo referencia al número de variaciones almacenadas en algunos de los repositorios o bases de datos de variaciones más conocidos como ClinVar [9], Alzforum [10], GWAS [11] y SNPedia [12] en la **Figura 2**, se puede entender que en el ámbito de desarrollo de este trabajo se va a tratar con un gran volumen de información que crece rápidamente y que presenta una gran variedad: se trata de un problema de Big Data.

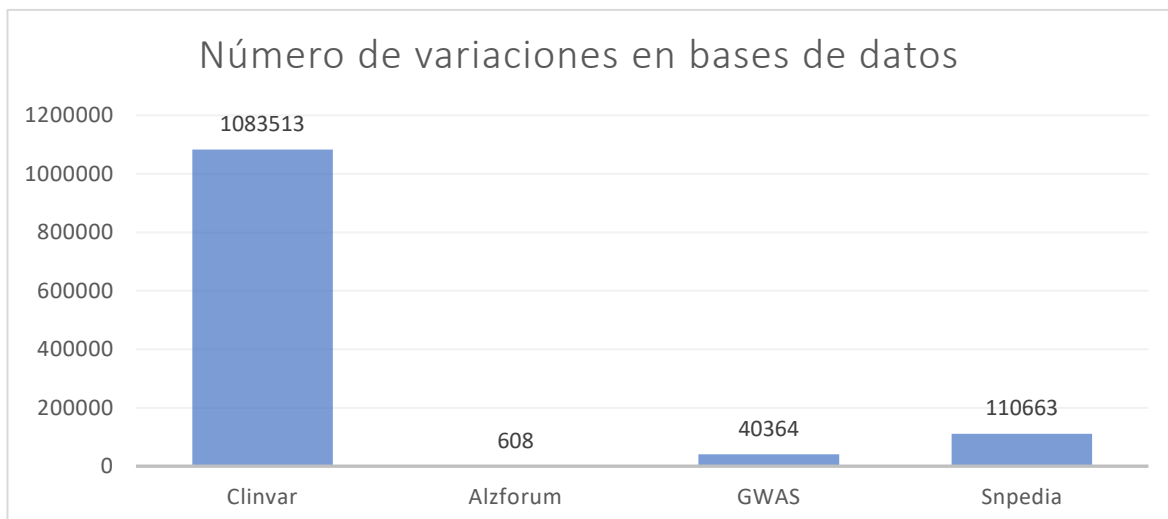


Figura 2. Número de variaciones en bases de datos conocidas

El problema radica, como se ha mencionado, en que no solo se busca procesar esta gran cantidad de datos, sino que también se busca seleccionar de toda esta información cuál es aquella que puede resultar relevante para su aplicación en clínica.

Con la visión de no solo poder procesar grandes cantidades de datos, sino de hacerlo de una forma inteligente para seleccionar solo la información relevante, nació el concepto de Smart Data [13]. Este concepto añade dos nuevas dimensiones al concepto de Big Data: el valor y la veracidad de la información, que contemplan la necesidad de distinguir de toda la información aquella que es verídica y cuyo uso aporta valor [14].

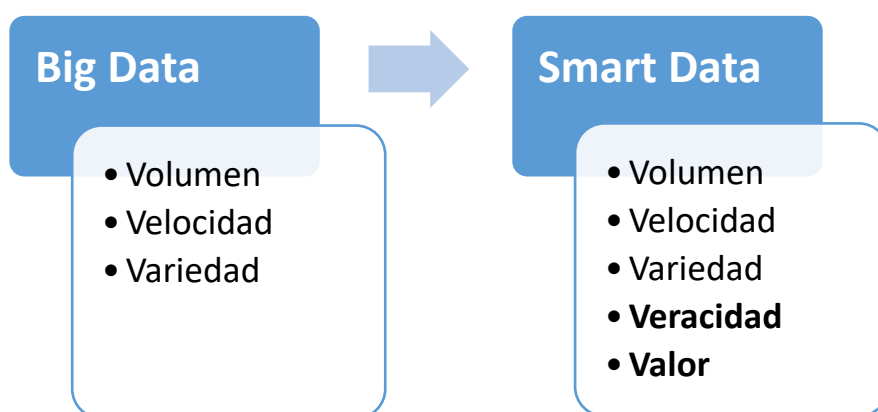


Figura 3. Evolución del Big Data al Smart data

En términos del trabajo desarrollado se va a aplicar el concepto de Smart Data para, de entre toda la información disponible en las bases de datos genómicas, seleccionar aquellas variaciones que estén relacionadas con el fenotipo de estudio y que sean relevantes para su uso en el diagnóstico genético, sustentando estos resultados en criterios de calidad, veracidad y valor de la información. En concreto, esto se llevará a cabo siguiendo el método SILE.

1.3. Introducción al Método SILE

Con la perspectiva de pasar de la aplicación del concepto de Big Data al de Smart Data en el dominio de los datos genómicos, en el centro de I+D PROS UPV se ha desarrollado el método SILE. El objetivo de este método es definir una estrategia para sistematizar el proceso de identificación de variaciones genéticas relevantes para una enfermedad, con la visión de obtener unos resultados lo suficientemente consistentes como para ser aplicados en el ámbito clínico.

El nombre del método hace referencia a las 4 etapas necesarias para alcanzar el objetivo propuesto (*Search - Identification - Load - Exploitation*). Debido a que el trabajo se ha desarrollado en el marco de este método, en la sección MÉTODO SILE de este documento se profundizará en sus fundamentos y características. Se define, brevemente, en qué consisten las diferentes etapas:

- **Search (Búsqueda).** Selección de los repositorios o bases de datos de los que se va a extraer la información.
- **Identification (Identificación).** Determinación de las variaciones genéticas que tengan una relación con la enfermedad suficientemente contrastada.
- **Load (Carga).** Carga de las variaciones en una base de datos.
- **Explotación (Explotación).** Uso de la información extraída con una finalidad clínica.

Este método fue aplicado en un trabajo previo sobre el Alzheimer, incluyendo también al Alzheimer temprano, pero no se obtuvo ninguna variación relevante [15]. El Alzheimer temprano, tal y como se tratará en mayor profundidad en la sección FENOTIPO ESTUDIADO: ALZHEIMER TEMPRANO del documento, presenta una serie de características, principalmente su baja incidencia y su importante componente familiar, que dificultan el análisis de las variaciones genéticas. Los resultados obtenidos en la aplicación del método SILE a esta enfermedad han planteado la necesidad de realizar un análisis exhaustivo de la etapa de identificación del método, para determinar cómo se podrían tratar adecuadamente casos específicos como el Alzheimer temprano.

De esta forma, el primer objetivo de este trabajo va a ser realizar una evaluación de la etapa de identificación del método SILE, buscando determinar los problemas existentes a la hora de realizar el análisis de las variaciones en casos particulares como el del fenotipo estudiado. De este proceso de análisis deriva el segundo objetivo del trabajo, que consiste en realizar una propuesta de mejora para la etapa de identificación del método SILE, con el objetivo de hacerla también utilizable en casos específicos como el Alzheimer temprano.

En la siguiente sección, se van a introducir los trabajos encontrados que presentan relación con el aquí desarrollado.

1.4.Trabajos Relacionados

Dentro del centro de investigación de I+D PROS UPV, se han desarrollado diferentes trabajos de Fin de Grado y Fin de Máster, que por el método seguido guardan relación con este Trabajo Fin de Grado.

- Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca: Aplicaciones a la medicina de precisión. Trabajo Fin de Máster en Ingeniería Biomédica. Nelvia del Cisne (2019) [16].
- Análisis de datos genómicos para la identificación de variantes genómicas relevantes en la enfermedad de Alzheimer. Trabajo Fin de Máster en Ingeniería Biomédica. Ignacio Pascual Fernández (2019) [15].
- Ciencia de datos para la medicina de precisión: Aplicaciones al diagnóstico genómico del Neuroblastoma. Trabajo Fin de Máster en Ingeniería y Tecnología de Sistemas Software. Paloma Cuenca Gil (2019) [17].
- Biotecnología para la medicina de precisión: Identificación de variaciones genómicas para el diagnóstico de cáncer de mama. Trabajo Fin de Grado en Ingeniería Biomédica. Alba Escalera Balsera (2018) [18].
- Un proceso para la identificación sistemática de variaciones genómicas: Aplicaciones a la medicina de precisión. Trabajo Fin de Máster en Ingeniería y Tecnología de Sistemas Software. Simranpreet Kahur (2018) [19].
- Exploración de bases de datos genómicas dirigida por modelos conceptuales. Trabajo Fin de Máster en Ingeniería y Tecnología de Sistemas Software. Vanessa Solís Cabrera (2018) [20].
- Diseño de un sistema de información genómica para el diagnóstico del neuroblastoma. Trabajo Fin de Grado en Ingeniería Biomédica. Clara Soler Pellicer (2016) [21].

El objetivo principal de estos trabajos consistía en aplicar el método SILE para la identificación de variaciones genéticas relevantes en diferentes enfermedades de estudio. En este trabajo, en cambio, no se busca realizar una aplicación directa del método, sino realizar un análisis crítico de la etapa de identificación para determinar si se ajusta a enfermedades con particularidades como la de estudio.

Este análisis permitirá plantear una propuesta de mejora para la etapa de identificación del método, buscando que el método sí pueda ajustarse a las enfermedades con características particulares. Una vez propuesta la mejora, se realizará la aplicación del método a la enfermedad de estudio, buscando realizar una evaluación de la solución propuesta.

1.5. Estructura del Trabajo

Se presenta a continuación la estructura que se va a seguir en el trabajo:

1. **Introducción.** Explicación de contexto en el que se ha desarrollado el presente trabajo.
2. **Objetivos y Stakeholders.** Se presentarán los objetivos concretos que se pretenden alcanzar, así como las diferentes preguntas de investigación derivadas de cada uno de estos objetivos.
3. **Método SILE.** Se explicará en que consiste el método SILE, desarrollado por el centro de investigación de I+D PROS UPV, y dentro del cual se encuentra contenido este trabajo.
4. **Fenotipo Estudiado: Alzheimer Temprano.** En este apartado se analizarán las características principales del fenotipo, así como las particularidades de este que puedan dificultar la determinación de las variables relevantes.
5. **Fuentes de Información.** En este apartado se realizará un análisis del repositorio que va a utilizarse en este trabajo, ClinVar, y se entrará en mayor detalle en sus características para facilitar la comprensión de las mejoras propuestas en la siguiente sección.
6. **Análisis y Propuesta de mejora de la Etapa de Identificación del Método SILE.** Tras justificar la necesidad de mejora de la etapa de identificación del método SILE, se realizará un análisis exhaustivo de esta etapa y se propondrá una mejora de esta
7. **Aplicación de la Mejora Propuesta para la Etapa de Identificación: Resultados y Discusión.** Se tratará como se ha realizado la aplicación del método SILE, con la etapa de identificación mejorada, y los resultados obtenidos. Se analizarán esos resultados y las limitaciones que se han tenido en su obtención.
8. **Conclusiones.** Conclusiones del trabajo.
9. **Trabajos Futuros.** Se detallarán todos los trabajos futuros que puedan derivar del presente trabajo.
10. **Bibliografía.** Fuentes de información consultadas durante el trabajo.

2. OBJETIVOS Y STAKEHOLDERS

2.1. Objetivos

Según lo expuesto, en este trabajo existen un total de 3 objetivos principales. El primero de ellos reside en hacer un análisis de la etapa de identificación del método SILE. Del resultado de este análisis deriva el segundo de los objetivos, que pretende proponer una mejora para esta etapa que permita tratar mejor el análisis de las variaciones en casos específicos como los del Alzheimer temprano. Finalmente, como tercer objetivo se va a buscar realizar una evaluación preliminar de la mejora propuesta mediante su aplicación al caso del Alzheimer temprano. Para la consecución de estos objetivos, se pueden establecer una serie de preguntas de investigación (*Research questions*, RQ) que marcarán las pautas para alcanzar cada uno de estos objetivos.

- **Objetivo 1.** Realizar un análisis de la etapa de identificación del método SILE para determinar si es apta para la evaluación de casos específicos como el Alzheimer temprano.
RQ1. ¿Qué criterios existen actualmente para la identificación de variaciones?
RQ2. ¿Los criterios que existen actualmente son aplicables al Alzheimer temprano?
RQ3. ¿Qué problemas se encuentran a la hora de aplicar los criterios existentes?
- **Objetivo 2.** Planteamiento de una mejora para la etapa de identificación del método SILE teniendo en cuenta los problemas detectados.
RQ4. ¿Qué mejoras se pueden plantear para resolver los problemas encontrados?
- **Objetivo 3.** Evaluación de la propuesta de mejora para la etapa de identificación mediante la aplicación del método SILE al caso del Alzheimer temprano.
RQ5. ¿Se consigue mejorar todos los problemas que se detecten?
RQ6. ¿La mejora realmente permite tratar las variaciones genéticas en casos particulares cómo la enfermedad de estudio?
RQ7. ¿Los resultados obtenidos son consistentes con los conocimientos disponibles de la enfermedad?
RQ8. ¿Qué garantías hay de que los resultados obtenidos con la aplicación de la mejora sean de calidad?
RQ9. ¿Se podrían emplear los resultados obtenidos en la práctica clínica?

2.2. Stakeholders

Los *stakeholders*, término utilizado típicamente en el ámbito empresarial, son aquellas personas o entidades que puedan verse afectadas por el tratamiento de un problema [22]. En este caso concreto, son aquellos individuos que puedan estar interesados en los resultados obtenidos por este Trabajo Fin de Grado. Concretamente, se considerarán stakeholders:

- Investigadores relacionados con el Alzheimer temprano.
- Personas que estén esperando un diagnóstico para el Alzheimer temprano.
- Médicos y clínicas que realicen test genéticos de diagnóstico.

A continuación, en la siguiente sección, se va a explicar en mayor profundidad el método SILE, introducido previamente en el apartado Introducción al Método SILE.

3. MÉTODO SILE

El método SILE, tal y cómo ha sido tratado en el apartado de Introducción al Método SILE, busca sistematizar el proceso de identificación de variaciones genéticas para una enfermedad, y adicionalmente, obtener unos resultados que tengan la consistencia suficiente como para ser aplicados en el ámbito clínico.

Debido a que el trabajo se ha desarrollado en el marco de este método, se presenta en más detalle en qué consisten cada una de sus 4 etapas.

3.1. Búsqueda

En esta etapa se busca determinar de qué fuentes se va a extraer la información de las variaciones genéticas relacionadas con la enfermedad de estudio. Para la selección de las fuentes adecuadas se plantean una serie de métricas que deben ser evaluadas en cada base de datos previo uso:

- **Credibilidad.** Se evalúan aspectos como si la base de datos ha sido revisada por expertos, o si hay controles sobre la calidad de los datos almacenados.
- **Relevancia.** Se determina si la base de datos contiene toda la información necesaria para el análisis de las variaciones.
- **Reputación.** Se evalúa si la base de datos es mantenida por instituciones, asociaciones o centros de investigación que sean conocidos y respetados.
- **Actualidad.** La base de datos debe ser actualizada con frecuencia, aportando siempre información sobre en qué versión se encuentra y sobre su última fecha de actualización.
- **Accesibilidad.** La información de la base de datos debe ser pública y descargable para poder ser utilizada.

A la hora de seleccionar las bases de datos, se buscará que se cumplan la mayoría de las métricas para asegurar que la información que se vaya a utilizar para analizar la relevancia de las variaciones sea fiable y de calidad [23].

3.2. Identificación

En la etapa de identificación, de todas las variaciones que se encuentren en las bases de datos seleccionadas en la etapa de búsqueda, se identificarán aquellas cuya relevancia para la enfermedad esté contrastada. Es decir, de todas las variaciones genéticas que aparezcan relacionadas con el fenotipo de estudio, solo se seleccionarán aquellas que tengan una evidencia lo suficientemente relevante como para que puedan ser consideradas útiles para su aplicación en el dominio clínico.

Previo al planteamiento de los diferentes criterios utilizados para determinar la relevancia de las variaciones, debido a que parte de estos se basan en el análisis de los estudios asociados, es necesario conocer cuáles son los estudios más comunes en relación con las variaciones genéticas:

- **Estudios de ligamiento.** Estos estudios permiten determinar el grado de influencia que tiene la genética en el desarrollo de la enfermedad. Se identifican con los términos MeSH¹ *linkage Disequilibrium* o *Genetic Linkage*.
- **Estudios funcionales.** Se centran en comprender los mecanismos moleculares que relacionan el genotipo con el fenotipo.
- **Estudios de asociación.** Estos estudios pretenden determinar que variaciones son de interés para una determinada enfermedad, en una determinada población. Estos estudios se subdividen en dos grupos: Estudios GWAS (término MeSH Genome-Wide Association Study) que buscan determinar las variaciones de relevancia para una población mediante la comparación del genoma completo de diferentes individuos de esta, y los estudios de cohortes (término MeSH Cohort Studies) que determinan relaciones entre la genética y el ambiente que llevan al desarrollo de una enfermedad.

Dentro del método SILE se ha desarrollado un diagrama de flujo, mostrado en la **Figura 4**, que permite en base a los términos MeSH indicados y a palabras clave identificadas, la clasificación de los artículos disponibles para cada variación. Distinguir el tipo de artículo es importante, pues como se ha mencionado, de esto dependen parte de los criterios definidos en SILE para determinar la relevancia de las variaciones genéticas.

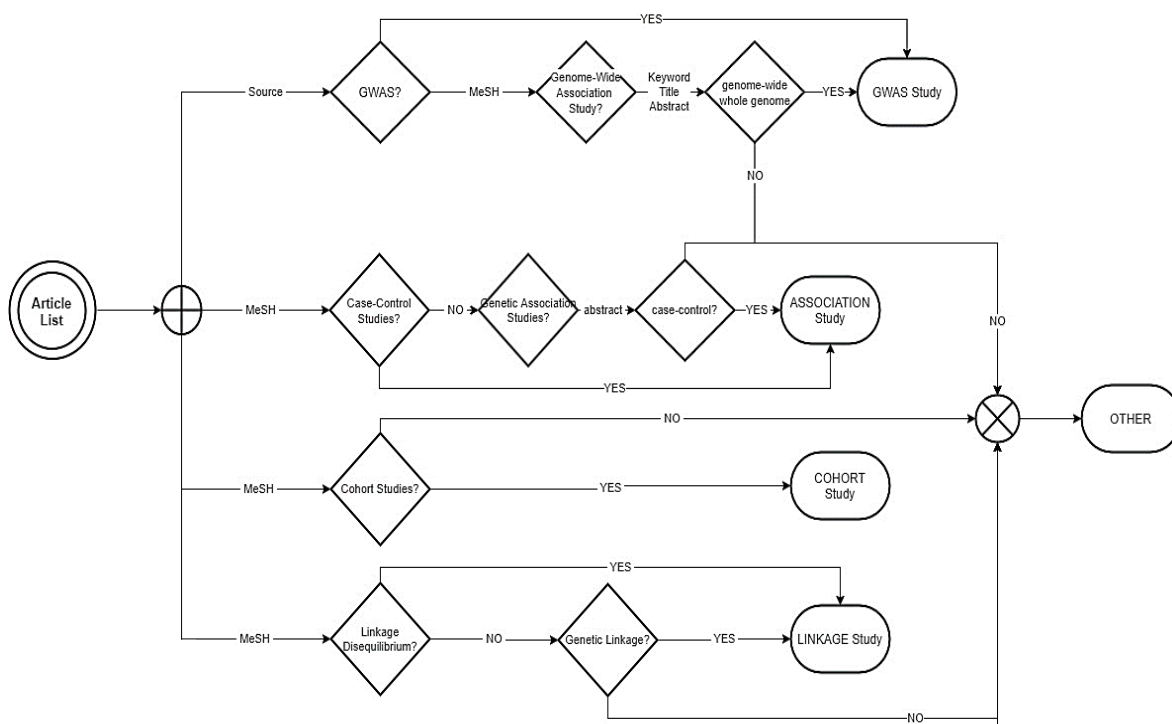


Figura 4. Diagrama de flujo para la clasificación de artículos

¹ Los términos MeSH (Medical Subject Headings) son los incluidos en el diccionario de sinónimos controlado por el NLM (National Library of Medicine), que son utilizado para indexar artículos para Pubmed [87], que contiene más de 30 millones de citaciones de literatura biomédica [88].

Este diagrama de flujo contiene 5 posibles clasificaciones para los artículos: *GWAS Study*, *ASSOCIATION Study*, *COHORT Study*, *LINKAGE Study*, y *OTHER*. Como se puede ver en la **Figura 4**, en primer lugar, se comprueba si los términos MeSH concuerdan con alguno de los tipos de artículos definidos, en caso contrario, se buscan palabras clave coincidentes con alguno de los tipos. Finalmente, si no se encuentra ningún término MeSH ni palabra clave que permita clasificar el artículo en alguno de los 4 tipos identificados, el artículo será clasificado como *OTHER*.

Una vez definidos los diferentes tipos de estudios considerados en el método, así como las características en base a las cuales pueden ser identificados, se pueden definir los diferentes criterios utilizados para determinar qué variaciones genéticas tienen la evidencia suficiente como para poder ser aplicadas en el ámbito clínico. Estos criterios, se han planteado en forma de flujo de trabajo (*workflow*), como se puede ver en la **Figura 5**.

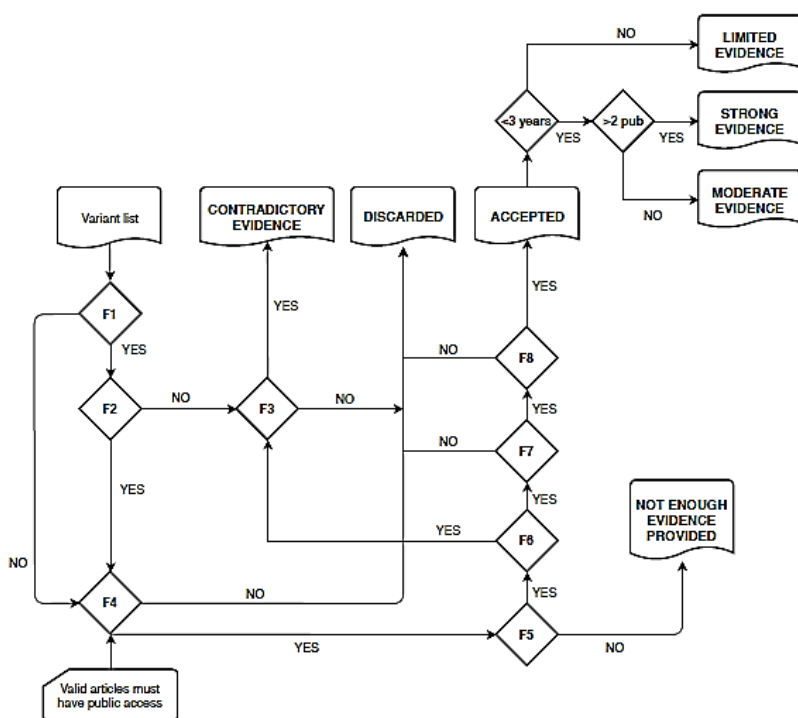


Figura 5. Workflow para la identificación de variaciones genéticas relevantes en el método SILE

Los diferentes pasos que constituyen este workflow son los siguientes:

- Evaluación de la relevancia clínica
 - Determinar si la variación ha sido interpretada (F1).
 - Analizar la utilidad de la interpretación para el diagnóstico clínico (F2): se consideran útiles aquellas variaciones que han sido interpretadas como patogénicas, probablemente patogénicas y factor de riesgo.
- Evaluación de la presencia de conflictos en la interpretación (F3).
- Determinación de si hay estudios publicados que apoyen la asociación de la variación con la enfermedad (F4).

- Evaluación de la relevancia de los estudios en términos de significancia estadística (F5): se consideran estudios relevantes aquellos realizados en humanos y que cumplan los siguientes criterios:
 - En los estudios de GWAS se requieren más de 700 casos y controles. En el caso de las variaciones patogénicas, los requisitos estadísticos son un *odd-ratio* mayor que uno, un intervalo de confianza también mayor que uno y un valor p menor que 5×10^{-8} .
 - En estudios de otro tipo, se requieren 500 participantes. Los requisitos estadísticos son un valor p inferior a 0,0001 y un intervalo de confianza que no contenga 1.
 - Es recomendable que los estudios sean replicables.
- Evaluación de la presencia de conflictos en la literatura asociada (F6).
- Evaluación del MAF (*Minor Allele Frequency*)²: El MAF de la variación debe ser menor o igual a la frecuencia del fenotipo en la población (F7).
- Evaluación de la consistencia de la variación con la enfermedad (F8): El patrón de herencia, la penetrancia y el mecanismo de la variación deben ser consistentes con la enfermedad.

Así, gracias a este workflow, se puede obtener una clasificación para cada una de las variaciones extraídas de los diferentes repositorios seleccionados en el apartado de búsqueda, que vendrá determinada por el nivel de evidencia del que disponga cada variación [23]. Las posibles clasificaciones para estas variaciones son las que siguen:

- **Contradictory evidence.** Existen evidencias contradictorias acerca del significado clínico de la variación.
- **Discarded.** Variaciones que, aunque si tienen artículos relevantes asociados, no cumplen con los requerimientos para el MAF o no se han encontrado evidencias que respalden la consistencia de la variación con la enfermedad. También serán clasificadas de este modo las variaciones que no sean útiles a nivel clínico.
- **Not enough evidence provided.** No se dispone de artículos relacionados con la variación que proporcionen la evidencia estadística requerida.
- **Accepted.** Variaciones que cumplen con todos los requerimientos del workflow. Existen tres niveles de evidencia, en función de cuánto hace que se revisó la variación y el número de artículos que tiene asociados (*limited, strong* y *moderate*).

3.3. Carga

Una vez se han determinado las variaciones relevantes gracias a la etapa de identificación, el próximo paso consiste en cargar las variaciones seleccionadas en la *Human Genome Database* (HGDB). Esta base de datos forma parte de un sistema de información genómica (*Genomic Information System, GeIS*), basado en el modelo conceptual del genoma humano (CSHG), también desarrollado en el centro PROS [24].

² El MAF se refiere a la frecuencia de aparición del segundo alelo más común en una determinada población, y es ampliamente utilizado debido a que permite diferenciar variaciones comunes de las de rara aparición.

Como paso previo a la carga de la información en la HGDB se debe asegurar que todas las inconsistencias entre los datos hayan sido resueltas. Algunas de las inconsistencias típicamente encontradas en los datos son: entradas duplicadas entre las diferentes bases de datos, claves inválidas, y cualquier otra restricción derivada de la tecnología utilizada para desarrollar la base de datos.

3.4. Explotación

En la última etapa del método, se pretende extraer conocimiento del sistema de información, buscando utilizar la información almacenada en la etapa anterior con fines diagnósticos. Tal y como se ha mencionado en la INTRODUCCIÓN, la determinación de la influencia de la variabilidad genética en el desarrollo de la enfermedad puede suponer un antes y un después en el diagnóstico y tratamiento gracias a la medicina de precisión.

Adicionalmente a esto, debido a que el dominio de datos genéticos es muy complejo, dentro del método SILE se ha desarrollado una herramienta, VarSearch [25], que permite comparar de forma intuitiva e interactiva las variaciones genéticas encontradas en el análisis genético de un paciente frente a las almacenadas en el HGDB, generando así un *report* que indica el riesgo de sufrir una determinada enfermedad.

Este trabajo se centra en la segunda etapa del método (identificación), pues es la que será objeto de análisis y mejora. En la siguiente sección, se van a analizar las diferentes características del Alzheimer temprano, así como los inconvenientes encontrados a la hora de aplicar la etapa de identificación del método SILE.

4. FENOTIPO ESTUDIADO: ALZHEIMER TEMPRANO

El Alzheimer es la forma de demencia más común en personas mayores [26], entendiendo la demencia como una pérdida grave de las funciones mentales, con un importante impacto en la vida diaria del paciente [27]. Algunos de los efectos más destacados del Alzheimer son: pérdida de memoria, dificultades en la concentración, disminución en la capacidad para tomar decisiones y cambios en la conducta [28].

La demencia está teniendo un mayor impacto en nuestra sociedad con el paso de los años, así mientras que en 2018 se estimó que habían 50 millones de afectados por la demencia en todo el mundo, en 2030 se espera que el número de afectados sea 82 millones, y en 2050 de 152 millones [29]. Debido a que el Alzheimer es la forma más común de demencia, se entiende la necesidad de poner medidas para facilitar los diagnósticos y tratamientos para la enfermedad.

En el caso concreto de España, también se ha observado una tendencia creciente de muertes debido a la enfermedad de Alzheimer. De acuerdo con las estadísticas realizadas por el INE (Instituto Nacional de Estadística), en el año 2000 se produjeron 5382 muertes debido a esta enfermedad, mientras que en el 2018 fueron 14929, lo que confirma la tendencia al alza que se está dando a nivel mundial [30].

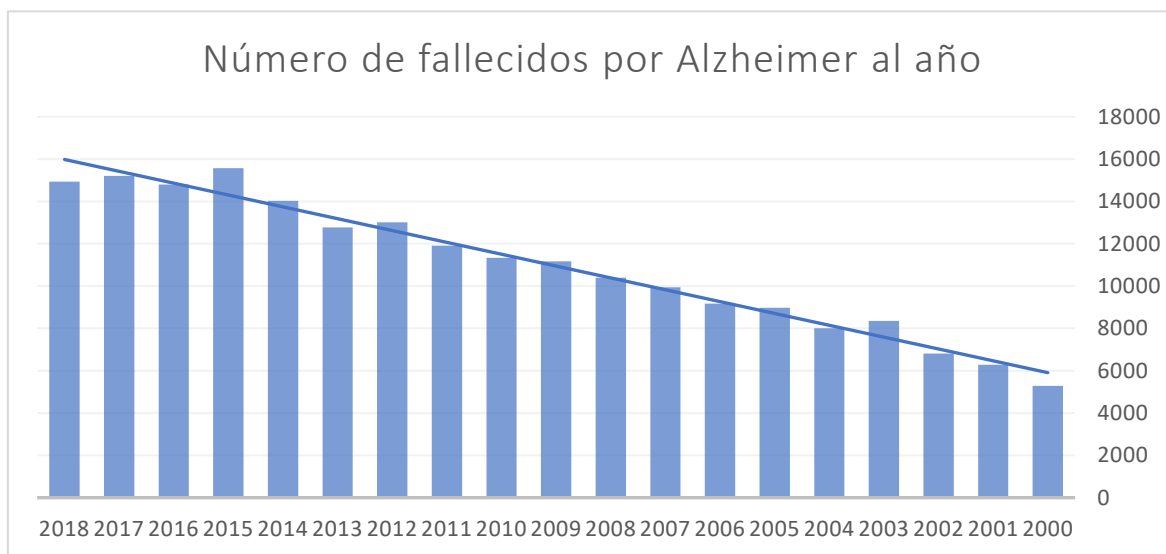


Figura 6. Número de fallecidos por Alzheimer al año

Analizando el número de muertes en 2018 por Alzheimer, pero en este caso viendo la distribución por edades (**Figura 7**), aparecen casos de pacientes que han fallecido por esta enfermedad entre los 45-65 años, unas edades inferiores a lo típicamente esperado para esta enfermedad.

Esto se debe a que existen dos tipos de Alzheimer, el Alzheimer tardío (*Late Onset Alzheimer Disease, LOAD*) que aparece en personas mayores de 65 años y el Alzheimer temprano (*Early Onset Alzheimer Disease, EOAD*) que aparece en menores de 65 años [31].

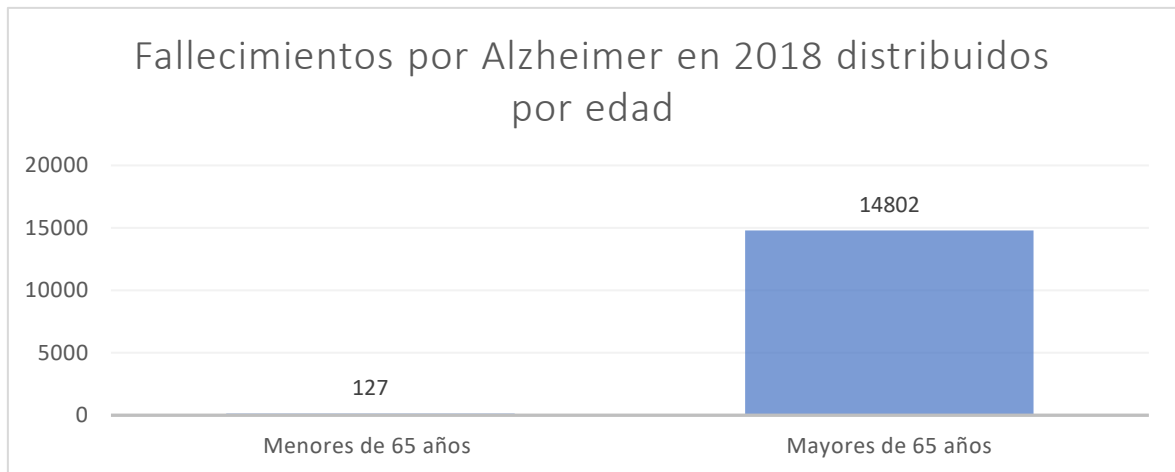


Figura 7. Fallecimientos por Alzheimer en 2018 en España distribuido por edades

En este trabajo van a buscarse variaciones genéticas que estén relacionadas con la forma más temprana de Alzheimer ya que, debido a que no se suele esperar encontrar Alzheimer en pacientes tan jóvenes, el diagnóstico puede ser largo y dificultoso.

En los próximos apartados, se van a analizar las características más relevantes del EOAD, así como las dificultades que van a encontrarse en el proceso de identificación de las variaciones relevantes.

4.1. Características del Alzheimer Temprano

Como se puede extraer de las estadísticas presentadas en la **Figura 7**, el porcentaje de incidencia del EOAD es mucho inferior al LOAD, siendo el total de casos de EOAD tan solo el 2-10% de todos los casos de Alzheimer [32]. Este porcentaje oscila en función del estudio considerado, de forma que a veces se sitúa entre el 0-1% [33], o entre el 4-5% [34], pero siempre supone menos del 10% de los casos totales de Alzheimer. Aunque no se pueda establecer un porcentaje fijo de la incidencia de la enfermedad, de todos los estudios se concluye que es una enfermedad con una baja incidencia.

Respecto a la influencia de la genética en el desarrollo de la enfermedad, se ha determinado que el EOAD en su mayoría está determinado por la componente genética, ya que presenta una heredabilidad entre el 92-100% [35]. Esto contrasta con el LOAD, que se caracteriza por ser una enfermedad de carácter poligénico, donde numerosos genes de riesgo y biomarcadores han sido identificados, aunque ningún gen se ha podido establecer como causativo [30]-[33].

El EOAD presenta dos subgrupos de la enfermedad, el EOAD esporádico (sEOAD) y el EOAD familiar (EOFAD). Se considera EOAD familiar cuando el paciente diagnosticado tiene al menos un familiar de primer grado afectado por Alzheimer (ya sea EOAD o LOAD), suponiendo esta forma de EOAD entre el 35-60% de los casos [35]. En contraposición, se considerará EOAD esporádico cuando el paciente no tenga una historia familiar positiva.

En el caso del EOFAD, se han determinado 3 genes cuyas variaciones presentan una herencia autosómica dominante y una penetrancia³ superior al 85%: PSEN1, PSEN2, APP. Debido a estas características, estos tres genes se consideran como biomarcadores de diagnóstico para esta enfermedad [31]. Esto indica que muchas de las variaciones que se encuentren en las bases de datos, relacionadas con el EOAD, estarán presentes en alguno de estos tres genes.

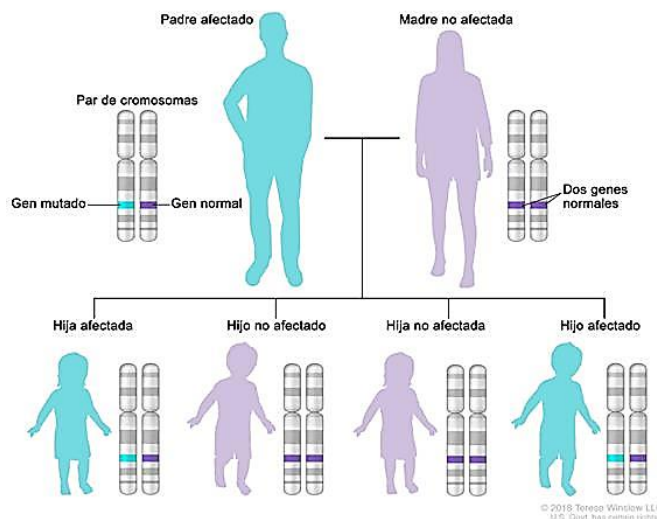


Figura 8. Mecanismos de herencia autosómica dominante

Aunque en un principio se determinó que el EOAD de carácter familiar solo respondía a una herencia autosómica dominante, se ha demostrado que tan solo entre el 10-15% del EOAD familiar presenta este tipo de herencia [35]. Sin embargo, sí se ha determinado que las variaciones en los genes APP, PSEN1 y PSEN2 explican la mayoría de los casos de EOAD con herencia autosómica dominante (explican entre el 23-88.2% de los casos de EOAD familiar autosómico dominante [35], según estudio).

Con esto, a pesar de que las variaciones genéticas presentes en estos tres genes son las más típicamente asociadas con esta enfermedad, con ellas solo se consigue explicar entre el 5-10% de los casos de EOAD [31][36]. Por lo que quedan entre un 90-95% de casos de EOAD que no se encuentran asociados a estos genes, y cuya causa genética todavía está por determinar.

Para clarificar la distribución de todos los tipos de Alzheimer mencionados a lo largo del apartado, en la **Figura 9** se muestran los diferentes tipos, así como su porcentaje de aparición sobre el total de enfermos de Alzheimer.

³ Proporción de individuos que, presentando una determinada variación genética, presentan síntomas de la enfermedad con la que se asocia la variación [89].

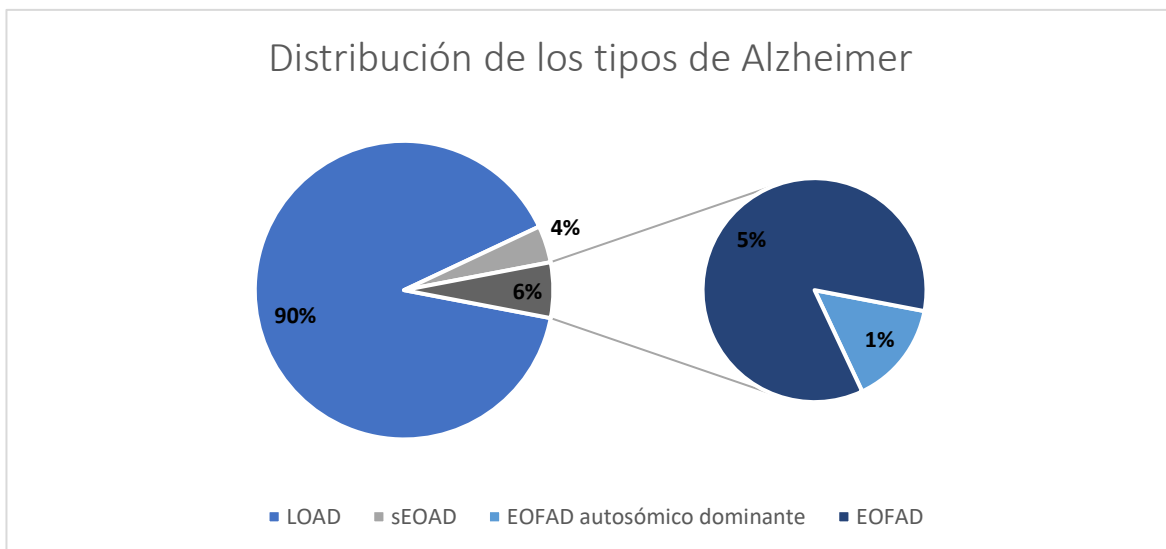


Figura 9. Distribución de los diferentes tipos de Alzheimer

Cuando se habla de EOAD familiar, en el aspecto autosómico dominante, se distinguen tres tipos de Alzheimer en función del gen en el que se encuentre la variación. Se contemplan para esta clasificación los genes más claramente asociados a la enfermedad que son el PSEN1, PSEN2 y el APP:

- **Alzheimer tipo 1.** Asociado al gen APP. También se incluyen los genes HFE, NOS3, PLAU, A2M, MPO [37].
- **Alzheimer tipo 3.** Asociado al gen PSEN1 y APOE [38].
- **Alzheimer tipo 4.** Asociado al gen PSEN2, que presenta una muy baja frecuencia de aparición de mutaciones [31][39].

La siguiente figura clarifica los diferentes subtipos que podemos encontrar de la enfermedad.

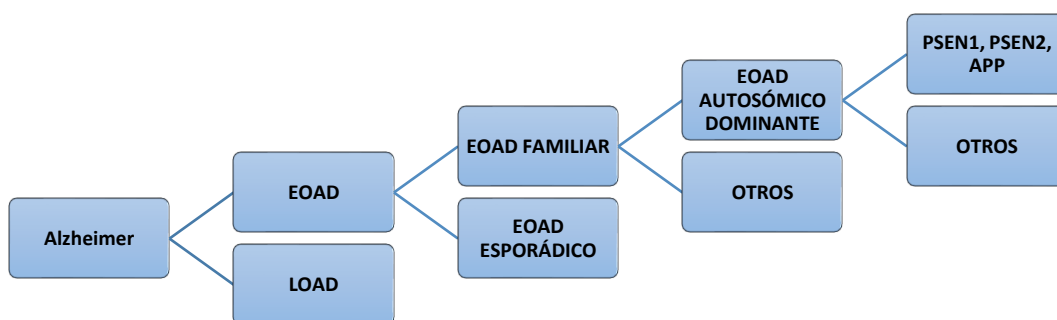


Figura 10. Divisiones dentro de la enfermedad de Alzheimer

Por el momento, se han tratado los diferentes tipos de EOAD, así como información relevante acerca de los mismos. Sin embargo, no se ha tratado el impacto de esta enfermedad a nivel molecular, algo que será importante para una mejor comprensión del apartado de Discusión de los resultados obtenidos.

Aunque existen diferentes consecuencias neuropatológicas de la enfermedad, dos de las más destacadas son la aparición de ovillos neurofibrilares en el interior de la neurona y las placas seniles en el exterior de esta.

Los ovillos neurofibrilares son acúmulos localizados en el interior neuronal que están constituidos por filamentos pareados helicoidales (PHFs), cuyo componente principal es una forma alterada de la proteína tau, que es clave para el mantenimiento interno de la arquitectura neuronal. Al perder la proteína tau esta capacidad para mantener la arquitectura, se producen graves daños en las neuronas que afectan a su capacidad de transmisión de información y que pueden causar la muerte neuronal [40]-[41].

Las placas seniles son agregados insolubles que aparecen en el medio extracelular y que están constituidos por el péptido $A\beta$, que deriva del péptido betaamiloide (APP, codificado por el gen del mismo nombre). Al igual que los ovillos neurofibrilares, estas placas también pueden ser responsables de la muerte neuronal [42].

La longitud del péptido betaamiloide presenta una longitud variable (40 o 42 aminoácidos), siendo la forma de 40 aminoácidos la más común. Se ha demostrado que la forma del péptido de 42 aminoácidos presenta una mayor facilidad de agregación, lo que justifica que las placas seniles estén formadas fundamentalmente por el péptido en la forma $A\beta_{42}$ [43].

Debido a la relación de las placas seniles con el desarrollo de la enfermedad, se ha demostrado que en pacientes con Alzheimer existe un incremento del ratio $A\beta_{42}/A\beta_{40}$, siendo este un indicativo de la presencia de la enfermedad [40]-[41].

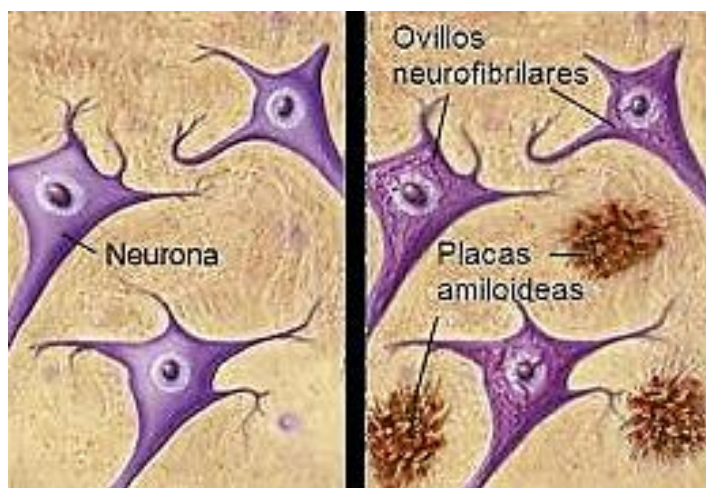


Figura 11. Medio celular en el cerebro sin y con presencia de agregados proteicos

Adicionalmente al impacto de la presencia de los ovillos neurofibrilares y las placas amiloides, existen teorías que relacionan los efectos que tiene el Alzheimer en el cerebro del paciente enfermo con un exceso de radicales libres, y por tanto con un proceso oxidativo. Este exceso se asocia con el envejecimiento y con la muerte neuronal, debido al daño oxidativo generado, siendo por tanto otro factor de interés para estudiar el impacto de la enfermedad [40].

De lo aquí expuesto, se traduce que en el cerebro de pacientes con Alzheimer se pueden detectar ovillos neurofibrilares, muerte/daño celular, placas seniles y un incremento del ratio $A\beta_{42}/A\beta_{40}$.

Una vez se han definido las características del EOAD en todas sus dimensiones, en la siguiente sección se va a hacer un análisis de los resultados que se obtuvieron en la aplicación del método SILE al caso del EOAD.

4.2. Aplicación del Método SILE al EOAD

En el apartado 3, donde se ha detallado en que consiste el método SILE, se han presentado diferentes criterios utilizados para determinar la relevancia de las variaciones, basados parte de ellos en la relevancia estadística proporcionada por los estudios relacionados.

Tal y como se ha mencionado en el apartado 4.1, una de las características principales del Alzheimer temprano es su baja incidencia. En una enfermedad con tan pocos pacientes afectados, difícilmente se encontrarán estudios que alcancen una evidencia estadística significativa tal como se especifica en la etapa de identificación de SILE, pues estos se llevarán a cabo con un número limitado de pacientes.

Si se observa el workflow presentado en la **Figura 5**, todas las variaciones que no dispongan de artículos con el número de pacientes requeridos, o con la relevancia estadística suficiente serán automáticamente clasificadas como *Not Enough Evidence Provided*. Esta hipótesis se confirma con la revisión de un trabajo llevado a cabo en el centro PROS UPV, en el que se buscaron variaciones responsables tanto del Alzheimer temprano como del de origen tardío, y no se encontró ninguna variación relevante para el EOAD con la aplicación de estos criterios [15].

Estos resultados implican que con este workflow no se contemplan casos específicos, como el EOAD, que presenten una baja incidencia y una marcada componente familiar, ya que ninguna de las variaciones asociadas podría llegar a ser clasificada como aceptada.

Por tanto, según lo expuesto, se plantea la necesidad de realizar un análisis exhaustivo de la etapa de identificación del método SILE, así como de una propuesta de mejora que sí permita tratar con casos específicos. Sin embargo, previa presentación de esta propuesta es necesario tener una visión general del repositorio del que va a extraerse la información. En el siguiente apartado, se realizará un análisis exhaustivo acerca del repositorio que se va a utilizar en este trabajo: ClinVar.

5. FUENTES DE INFORMACIÓN

Cómo se ha mencionado previamente en la sección 1.2 las diferentes variaciones genéticas, así como la información relacionada con las mismas, se almacena en diferentes bases de datos y repositorios cada uno con sus propia estructura y características.

Antes de realizar un análisis y propuesta de mejora de la etapa de identificación del método SILE es necesario entender la información con la que se va a trabajar, ya que esto facilitará la comprensión de las mejoras propuestas. Aunque existen múltiples bases de datos y repositorios disponibles, en este caso únicamente va a analizarse ClinVar pues debido a las características que presenta, que se van a ser tratadas a continuación, es el repositorio más adecuado para realizar este trabajo.

ClinVar es un repositorio de datos abiertos, que contiene la relación entre diversas variaciones genéticas y los fenotipos a los que se encuentran asociadas. La información es subida a la base de datos por diferentes usuarios, denominados *submitters*, que generan entradas o *submissions* en las variaciones para aportar un nuevo conocimiento sobre las mismas. Para cada una de estas submissions, el repositorio da la opción introducir el criterio que se ha utilizado para realizar la interpretación de la variación. Este concepto es representado en ClinVar como *assertion criteria* y hace referencia a cualquier publicación o documentación utilizada por el submitter para asignar a una variación un significado clínico u otro [44].

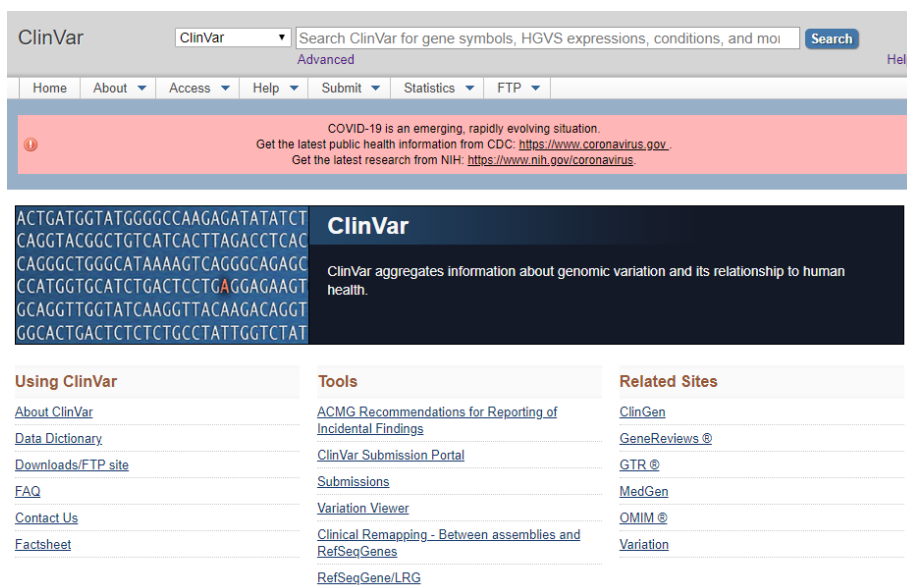


Figura 12. Menú principal de ClinVar

En ClinVar se realiza una validación de las diferentes submissions previa a la carga de los datos en el repositorio. Se resumen algunos de los aspectos más importantes que se tienen en cuenta para realizar esta validación:

- **Definición de la variación.** La variación se define como un cambio en la secuencia respecto a una referencia, y se verifica que el alelo indicado se produce en la posición indicada. Si esta información no se completa correctamente, la variación aparecerá como *non-validated*.
- **Normativa HGVS.** Se evalúa el formato HGVS (*Human Genome Variation Society*) asociado a la variación. La nomenclatura HGVS tiene como fin principal proporcionar un estándar para la descripción de las variaciones facilitando así el intercambio de información [45].
- **Evaluación de la consistencia de la información.** Se verifica que el identificador de la variación proporcionado (si es que se proporciona uno), que contiene información sobre la ubicación genómica, es consistente con la definición de la variación.
- **Análisis de los identificadores asociados al fenotipo.** En caso de que a la hora de subir a la base de datos una variación esta se relacione con algún fenotipo, ClinVar valida que los identificadores proporcionados sean válidos. Si además del identificador, se proporciona un nombre para el fenotipo, se comprueba que este nombre coincida con el concepto definido por el identificador.
- **Relación entre la variación y el gen en el que se localiza.** Todavía está en progreso, busca que cuando se indique el gen afectado, se pueda comprobar que la localización de la variación se encuentra dentro de este gen.
- **Evidencias proporcionadas.** Se validan identificadores de PubMed y otros identificadores de citas.
- **Interpretación.** Clinvar no interviene en la interpretación de las variaciones. Esto implica que, no valida la interpretación, no determina cual es la correcta cuando hay discrepancias en la interpretación entre submitters, y no se revisa el criterio que utilizan los submitters para validar la información.

Este proceso de validación previo de la información sirve como una garantía de que la información que se extraiga de la base de datos va a tener un mínimo de calidad. Debido a estas garantías, junto con que es un repositorio muy conocido, utilizado y que se encuentra en continua actualización (una vez al mes), esta será la fuente de información consultada para la extracción de variaciones genéticas. Además, tal y cómo se va a tratar en mayor profundidad en la sección 5.1 cumple con todos los requerimientos establecidos en la sección 3.1 para bases de datos (Credibilidad, Relevancia, Reputación, Actualidad y Accesibilidad).

Como ventaja adicional, ClinVar permite realizar búsquedas avanzadas en la base de datos que permiten acotar los resultados obtenidos para ajustarlos lo máximo posible a la información de interés. Se puede filtrar por aspectos como: tipo de submitter, significado clínico, estado de revisión, etc.

Builder

Disease/Phenotype [Show index list](#)

AND Gene Name [Show index list](#)

or [Add to history](#)

Figura 13. Buscador avanzado de ClinVar

Sin embargo, como se ha mencionado anteriormente, ClinVar no realiza ningún tipo de validación al respecto de la interpretación [46]. Esto justifica la necesidad de métodos como SILE, que buscan analizar si la información proporcionada como soporte para la interpretación de una variación es de la calidad suficiente para poder ser utilizada en el ámbito clínico.

En los siguientes apartados, primeramente, se va a justificar que ClinVar cumple con los requerimientos establecidos en la etapa de búsqueda del método SILE para ser considerada una base de datos adecuada para la extracción de información. Tras esto, se realizará un análisis de algunas características acerca de las variaciones sobre las que ClinVar proporciona información, que permitirán la mejor comprensión de la propuesta de mejora realizada en la sección 6.

5.1. Adecuación de ClinVar a los Criterios Establecidos en la Etapa de Búsqueda del Método SILE

Según lo expuesto en el apartado de Búsqueda del método SILE las bases de datos o repositorios que se utilicen como fuente de información deben cumplir con las siguientes características: Credibilidad, relevancia, reputación, actualidad y accesibilidad. Lo que se pretende en este apartado es, por tanto, verificar que ClinVar cumple con todas estas características.

Para la evaluación de la credibilidad debe tenerse en cuenta por una parte que la base de datos sea revisada por expertos, y por otra parte la existencia de controles de calidad que aseguren que la información proporcionada por la base de datos es correcta [23]. Según lo mencionado, ClinVar no realiza una validación de la interpretación de las variaciones, pero sí que presenta un proceso de curado por expertos. Adicionalmente, se ha determinado que se siguen numerosos controles de calidad a través de las características contempladas en la validación de las submissions.

En el caso de la relevancia, lo que se busca es determinar si la base de datos contempla toda la información necesaria para realizar la carga de las variaciones en la HGDB de acuerdo con el modelo conceptual (CSHG) [23]. Aunque no se han detallado los diferentes aspectos de los que dispone el CSHG, algunos a considerar respecto a las variaciones es el tipo de variación (inserción, delección, etc), el fenotipo al que se asocia, el identificador SNP, etc. En el caso de Clinvar, toda la información requerida está proporcionada y es accesible.

Respecto a la medida de la reputación, se debe evaluar si la base de datos está mantenida por una institución, asociación o centro de investigación que sea conocido y respetado [23]. En el caso de ClinVar, esta está mantenida por el NCBI (*National Center for Biotechnology Information*) que forma parte de la Biblioteca Nacional de Medicina de Estados Unidos, quedando por tanto garantizada la reputación de la base de datos.

Otro aspecto que considerar es la actualidad del repositorio. ClinVar actualiza su información de forma mensual, siendo uno de los repositorios con una mayor actualización de la información, y cumpliendo por tanto con este criterio.

Finalmente, el último aspecto para evaluar es la accesibilidad de la información. El repositorio ClinVar es de acceso público y también permite la descarga de toda la información relacionada con el resultado de la búsqueda de variaciones en un archivo en formato texto. Por tanto, se proporciona un fácil acceso a toda la información disponible.

Una vez se ha garantizado que ClinVar cumple con todos los requisitos establecidos en el método SILE, en el siguiente apartado se van a tratar diferentes aspectos que ClinVar proporciona acerca de las variaciones, y que son fundamentales para la comprensión del workflow que va a diseñarse para la identificación de variaciones genéticas relevantes.

5.2. Información Proporcionada por ClinVar

Existen diferentes aspectos que es importante conocer en detalle para mejorar la comprensión de los cambios introducidos en la etapa de identificación del método SILE, así como de los resultados obtenidos en su aplicación al EOAD. Para el correcto análisis de estos aspectos, se detalla la información en diferentes subapartados, que reflejan como se encuentran representados estos aspectos en ClinVar.

5.2.1. Tipos de Submitters

Tal y como se ha mencionado, toda la información contenida en ClinVar se obtiene a partir de submissions realizadas por submitters. Existen diferentes tipos de submitters, siendo algunos de los más comunes:

- Guías de práctica clínica.
- Paneles de expertos.
- Proyectos comunitarios.
- Laboratorios de test genéticos.
- Bases de datos Locus-específicas.
- Estudios de investigación.
- Flujos de datos semiautomáticos, como los proporcionados por OMIN y GeneReviews.
- Estudios de investigación.
- Otras bases de datos de públicas.

De entre todos los tipos de submitter, son las guías de práctica clínica y los paneles de expertos, los que dan una mayor garantía de la calidad y veracidad de la información [47]. Por este motivo, sus interpretaciones tienen prioridad sobre todas las realizadas por el resto de submitters.

5.2.2. Significado Clínico (*Clinical Significance*)

El significado clínico de una variación busca reflejar el impacto que tiene la variación sobre el desarrollo de una o varias enfermedades. En ClinVar se contemplan una serie de significados clínicos ya predefinidos, estando algunos de esos definidos de acuerdo a las recomendaciones de la ACMG/AMP (*American College of Medical Genomics/ Association for Molecular Pathology*) [48], mientras que otros han sido creados para cubrir las necesidades de información de los usuarios. Los significados clínicos relevantes para este trabajo son:

- **Benign (benigna).** La variación no es causativa de la enfermedad.
- **Likely Benign (probablemente benigna).** Se da este significado clínico a aquellas variaciones que con un 90% de certeza no son causativas de la enfermedad.
- **Uncertain Significance (significado incierto).** Cuando no se cumplen ni los criterios para variaciones benignas ni para patogénicas, o bien cuando criterios de ambas categorías dan lugar a contradicción.
- **Likely Pathogenic (probablemente patogénica).** Se da este significado clínico a aquellas variaciones que con un 90% de certeza son causativas de la enfermedad.
- **Pathogenic (patogénica).** La variación es causativa de la enfermedad.
- **Risk Factor (factor de riesgo).** Las variaciones con esa clasificación no causan la enfermedad, pero si aumentan el riesgo de padecerla.
- **Conflicting Interpretations (conflicto de interpretación).** Clinvar únicamente considera que se produce un conflicto de interpretación cuando para una misma variación existen con dos o más significados clínicos diferentes. Concretamente, significado incierto – benigna/probablemente benigna, significado incierto – patogénica/probablemente patogénica, benigna/probablemente benigna - patogénica/probablemente patogénica [49].

5.2.3. Estado de Revisión (*Review Status*)

ClinVar tiene internamente un sistema de clasificación basado en estrellas (0 a 4) que determina el nivel de revisión asociado a cada una de las variaciones. Para asociar la clasificación a cada variación, se tienen en cuenta aspectos como el tipo de submitter que ha realizado la submission, el número de submitters, si se ha proporcionado o no un assertion criteria, etc. En la **Tabla 1** se resumen los principales aspectos que se tienen en cuenta para la asignación del nivel de evidencia:

Número de estrellas	Estado de revisión
4 estrellas	El submitter es una guía de práctica clínica.
3 estrellas	El submitter es un panel de expertos.
2 estrellas	Hay más de un submitter y no existen conflictos entre la información proporcionada. Ninguno de estos submitters es una guía de práctica clínica o panel de expertos.
1 estrella	Dos posibles casos. Existe más de un submitter, pero aparecen conflictos en la interpretación, o bien solo aparece un único submitter. Ninguno de estos submitters es una guía de práctica clínica o panel de expertos.
0 estrellas	Diversos casos: el alelo únicamente aparece formando parte de un haplotipo o genotipo, no se proporciona un <i>assertion criteria</i> , o no se da ninguna información respecto a la variación.

Tabla 1. Estado de revisión de las variaciones en ClinVar

5.2.4. Tipo de Método Utilizado

El tipo de método (*method type*) indica la fuente de la información proporcionada acerca de la variación, es decir, indica de donde proviene la información en la que se ha basado el submitter para determinar la importancia clínica de la variación. Existen tres tipos de métodos: *Clinical Testing* (ensayo clínico), *Literature Only* (literatura), *Research* (investigación).

Cuando el método utilizado es el ensayo clínico, la información se obtiene en base a estudios clínicos desarrollados en pacientes. En el caso de las basadas en literatura, el submitter ha consultado diferente información disponible en la literatura científica, y en base a la información recopilada hace una interpretación de la variación. Finalmente, en el método basado en investigación toda la información proporcionada es el resultado de una investigación científica.

Una vez se han analizado las diferentes características de ClinVar, el siguiente paso es realizar el análisis de la etapa de identificación del método SILE y proponer mejoras a los problemas encontrados.

6. ANÁLISIS Y PROPUESTA DE MEJORA DE LA ETAPA DE IDENTIFICACIÓN DEL MÉTODO SILE

En el apartado de Aplicación del Método SILE al EOAD se determinó que el workflow actual del método SILE para la identificación de variaciones genéticas relevantes no permite tratar casos específicos como el EOAD, ya que en fenotipos con características similares al EOAD las variaciones serían clasificadas como *Not Enough Evidence Provided*.

Se plantea por tanto la necesidad de realizar un análisis de la etapa de identificación del método SILE, buscando concretar los aspectos susceptibles de mejora y las problemáticas existentes a la hora de evaluar las variaciones genéticas encontradas en las bases de datos o repositorios. En base a este análisis exhaustivo se diseñará un nuevo workflow, basado en el utilizado en el método SILE, como una propuesta de mejora para la identificación de las variaciones genéticas que contemple enfermedades con características particulares como el EOAD.

6.1. Análisis de la Etapa de Identificación del Método SILE

En la sección 4.2 se planteó el motivo principal por el que era necesario realizar una revisión de esta etapa del método SILE. El EOAD es una enfermedad con baja incidencia, lo que causa que los artículos relacionados con la enfermedad no se desarrollen en un gran número de pacientes y, por tanto, no alcancen una relevancia estadística significativa. Esto causaba que las variaciones relacionadas con el EOAD, o con otras enfermedades de características similares, fueran clasificadas según el workflow original del método SILE (**Figura 5**) como *Not Enough Evidence Provided*. Esta afirmación se confirma con los resultados obtenidos en una aplicación previa de este workflow al caso del Alzheimer, donde no se obtuvo ninguna variación relevante para el EOAD. Por tanto, son estos resultados los que han llevado al análisis de la etapa de identificación del método SILE que va a llevarse a cabo en este apartado.

El criterio de la relevancia estadística de los estudios es un criterio de uso extendido y aceptado, de forma que no se puede simplemente dejar de tener en consideración. Con esto, es necesario buscar criterios alternativos para la evaluación de la relevancia de las variaciones, que contemplen la relevancia estadística de los estudios asociados, pero sin descartar las variaciones únicamente porque no se cumpla este criterio.

Otro problema detectado para la aplicación de la etapa de identificación se relaciona con el tipo de estudios predominantes en esta enfermedad. Tal y como se ha indicado en el apartado 3, en el método SILE se contemplan principalmente 3 tipos de estudios: Estudios funcionales, estudios de asociación y estudios de ligamiento. En el apartado 4.1 se determinó que esta enfermedad presenta una componente familiar muy importante, de forma que los estudios predominantes son aquellos realizados en diferentes generaciones de una o varias familias. Este tipo de estudios, que serán definidos como estudios familiares, tienen el objetivo de determinar el impacto que tiene la herencia

de una variación en el desarrollo de una enfermedad, estudiando la segregación en una o varias familias.

Estos tipos de estudios no están siendo contemplados explícitamente hasta ahora en el método SILE ya que se clasificarían como *Other*, por lo que para el correcto estudio de la bibliografía asociada a las variaciones es necesario contemplarlos y definirlos para poder identificarlos correctamente. Con esto, se plantea la necesidad de realizar un análisis de los estudios familiares, y de incluirlos en el diagrama de flujo de identificación de artículos presentado en la **Figura 4**.

Realizando el análisis del workflow se han detectado otra serie de problemáticas, adicionales a la evaluación de la relevancia estadística de los estudios y de la no consideración de los estudios familiares, que son las que derivan más directamente de la aplicación del workflow del método SILE al caso del EOAD. A continuación, se van a tratar una a una estas problemáticas para después realizar una propuesta de solución en el siguiente apartado.

En primer lugar, un problema detectado se relaciona con las variaciones con interpretaciones conflictivas. En el workflow inicial (**Figura 5**), se puede ver como todas las variaciones que presenten algún conflicto en su interpretación serán clasificadas como *Contradictory Evidence*. El problema es que debido a cómo ClinVar gestiona las interpretaciones asociadas a una variación, específicamente cuando las variaciones han sido interpretadas para más de un fenotipo, algunas variaciones son incorrectamente clasificadas como un conflicto de interpretación. Se muestra un ejemplo de esta situación con la variación de ID rs33949390, que ha sido incorrectamente clasificada como un conflicto de interpretación.

Interpretation:	Conflicting interpretations of pathogenicity Benign(1);Likely benign(1);Uncertain significance(1)
Review status:	★☆☆☆ criteria provided, conflicting interpretations
Submissions:	8 (Most recent: Oct 22, 2019)
Last evaluated:	May 28, 2019
Accession:	VCV000039198.5
Variation ID:	39198
Description:	single nucleotide variant

Figura 14. Interpretación dada a la variación rs33949390 en ClinVar

En la **Figura 14** se puede ver como la variación ha sido clasificada como conflicto de interpretación debido a que existen asociados a esta variación los siguientes significados clínicos: Benigna, probablemente benigna y significado clínico incierto. Sin embargo, si se revisa en mayor detalle la información disponible acerca de esta variación, se observa que el conflicto de interpretación se produce para la enfermedad de Parkinson, pero no para el EOAD.

Interpreted condition	Interpretation	Number of submissions	Review status	Last evaluated	Variation/condition record
not provided	Benign	1	criteria provided, single submitter	Jan 22, 2019	RCV000873020.1
Parkinson disease 8, autosomal dominant	Conflicting interpretations of pathogenicity	4	criteria provided, conflicting interpretations	May 28, 2019	RCV000032472.4
Early onset Alzheimer disease with behavioral disturbance	Likely pathogenic	1	no assertion criteria provided	Feb 28, 2019	RCV000984891.1
Spinocerebellar atrophy	Pathogenic	1	no assertion criteria provided	Feb 28, 2019	RCV000984892.1
Klippel-Feil syndrome 1, autosomal dominant	Pathogenic	1	no assertion criteria provided	Feb 28, 2019	RCV000984893.1

Figura 15. Fenotipos (Conditions) asociados con la variación rs33949390

Por tanto, según lo mostrado en la **Figura 15** la variación rs33949390 debería haber sido clasificada como probablemente patogénica para el caso del EOAD, no como un conflicto de interpretación. Esto significa, por tanto, que al no considerar el workflow original del método SILE esta problemática, puede haber variaciones que resulten relevantes para el fenotipo de estudio y que sean incorrectamente clasificadas como *Contradictory Evidence*. Esta situación requiere el realizar una modificación del workflow inicial para contemplar correctamente estos casos.

El siguiente problema detectado es que en el workflow se revisan los artículos y la relevancia estadística de los mismos, pero en ningún momento se evalúa quien ha introducido la información en la base de datos (submitters), el tipo de método que se ha utilizado para extraer la información de la variación (method type), o que criterio se ha utilizado para asignarle el significado clínico a la variación (assertion criteria). En el caso del tipo de submitter, las submissions incluidas en las guías de práctica clínica o realizadas por paneles de expertos tienen prioridad sobre cualquier otra submission, debido a que estos tipos submitter son los más relevantes y los que proporcionan una mayor garantía de calidad de la información proporcionada. En lo referente al tipo de método, la información más fiable proviene de los ensayos clínicos, pues los tipos basados en literatura o investigación están claramente más sujetos a la interpretación del usuario que sube la información. El assertion criteria sirve para determinar en base a qué criterios se ha interpretado la variación, de forma que si este no ha sido proporcionado por el usuario no se tiene garantía de que la interpretación se ha hecho en base a criterios válidos o coherentes. Debido a que se busca determinar que la información relativa a la variación sea de calidad, se necesita conocer en base a qué criterios se ha realizado su interpretación.

Se ha justificado la importancia que tienen los aspectos mencionados, pero sin embargo ninguno de ellos se encuentra incluido dentro del workflow de la **Figura 5**. Por tanto, será necesario incluirlos en el workflow ya que son parámetros indicativos de la calidad de las submissions realizadas para cada variación.

Otro problema detectado es que en el workflow de la **Figura 5** no se hace ningún tipo de comprobación previa acerca de si los resultados que se han obtenido en la búsqueda en la base de datos son coherentes con el conocimiento que se dispone de la enfermedad. En las bases de datos se almacena una gran cantidad de información, y el usuario que vaya a trabajar con ellas debe ser capaz de determinar si la información proporcionada por la base de datos es coherente con sus conocimientos de la enfermedad. Es decir, se debe evaluar si las variaciones obtenidas en la base de datos tienen sentido dentro de los conocimientos de los que se dispone acerca de la enfermedad. Para evitar posibles errores, es recomendable incluir algún paso previo que sirva para realizar una verificación de los resultados obtenidos, como podría ser el comprobar si el gen en el que se localiza una determinada variación se encuentra relacionado con la enfermedad.

Finalmente, el último problema detectado consiste en cómo se ha introducido el criterio temporal en el workflow. Como se ve en la **Figura 5**, se introduce un criterio temporal (evidencias de hace menos de 3 años) para establecer el nivel de evidencia disponible para las variaciones aceptadas. El problema es que, con este criterio, variaciones con unos sólidos criterios de interpretación y una información relacionada de gran calidad podría ser clasificada como *Limited Evidence* solo por haber sido revisadas hace más de tres años. El criterio temporal es importante tenerlo en cuenta, pues la información relacionada con las variaciones genéticas está en continuo cambio y crecimiento, tal y como se vio en la sección 1.2, por lo que se deberá realizar un cambio en cómo se está evaluando este criterio para evitar que el nivel de evidencia alcanzado dependa principalmente del criterio temporal.

Teniendo en cuenta todo lo expuesto, se presentan un total de 6 problemas a resolver:

- **Problema 1.** Plantear una evaluación alternativa de la relevancia estadística de los estudios relacionados.
- **Problema 2.** Falta de consideración de los estudios familiares como un estudio independiente en el diagrama de clasificación de artículos de la **Figura 4**.
- **Problema 3.** Ausencia de un tratamiento adecuado de las variaciones incorrectamente clasificadas como un conflicto de interpretación.
- **Problema 4.** No consideración del tipo de método, el assertion criteria y el tipo de submitter como aspecto relevante para evaluar la calidad de la información. Con esto, tenemos por tanto tres subproblemas:
 - 4.1. No consideración del tipo de submitter.
 - 4.2. No consideración del tipo de método.
 - 4.3. No consideración del assertion criteria.
- **Problema 5.** No validación previa de que los resultados obtenidos son coherentes con los conocimientos disponibles de la enfermedad.
- **Problema 6.** Criterio temporal como criterio principal para determinar el nivel de evidencia.

Una vez se ha realizado el análisis del workflow y determinado las principales problemáticas a resolver, en las siguientes secciones se realizará una propuesta de mejora que permitirá solucionar todos los problemas identificados y además permitirá tratar casos específicos como los de EOAD, buscando con esto alcanzar el segundo de los objetivos definidos en la sección de Objetivos de este trabajo.

Esta propuesta se divide en dos apartados, en el primero de ellos se resolverá el segundo de los problemas, buscando definir correctamente los estudios familiares e incluirlos en diagrama de artículos de la **Figura 4**. En el siguiente apartado se planteará el diseño propuesto para un nuevo workflow para la identificación de variaciones genéticas relevantes, basado en el existente en SILE, donde se resolverán el resto de las problemáticas encontradas.

6.2. Solución al problema 2. Definición de los Estudios Familiares

Para comenzar a considerar los estudios familiares en el análisis de las variaciones genéticas relevantes, y resolver por tanto el segundo de los problemas identificados en la etapa de identificación del método SILE, es necesario realizar una correcta definición de estos para poder así distinguir inequívocamente cuando se está tratando con este tipo de estudios.

Se denominan estudios familiares a aquellos que buscan determinar la relevancia de una variación, centrándose en el impacto que tiene la herencia de esta en el desarrollo de la enfermedad. Se distinguen dos clases de estudios familiares:

- **Los que se realizan en una única familia.** Con estos estudios lo que se evalúa es el impacto de una variación genética en diferentes miembros de una misma familia.
- **Los que se realizan en múltiples familias.** En estos se estudian se dispone de pacientes pertenecientes a varias familias, teniendo siempre más de un paciente por cada familia de estudio. Así, se tendrán evidencias del efecto de la variación genética en diferentes familias, lo que puede dar más información acerca del impacto de la variación, y de si esta afecta igual en diferentes condiciones y ambientes.

Cabe destacar que existe un subtipo de estudio familiar que es aquel realizado en hermanos gemelos. Este subtipo de estudios familiares tiene por objetivo evaluar tanto la importancia de la genética, como la influencia que pueden tener diferentes ambientes en el desarrollo de la enfermedad

En el análisis de las evidencias proporcionadas por este tipo de estudios pueden aparecer diversos problemas. Primeramente, no suele disponerse de muchos miembros de la misma familia para la realización del estudio, algo causado por la temprana aparición de la enfermedad, pues puede darse la situación de que en el momento de realización del estudio no queden familiares del paciente vivos que padezcan la enfermedad. Por otra parte, debido al número limitado de pacientes disponibles para estos estudios, analizarlos en términos de relevancia estadística no suele dar buenos resultados. Sin embargo, estos estudios pueden proporcionar información muy valiosa acerca de si la variación es de novo (primera vez que se detecta en un individuo), de si existe cosegregación de la enfermedad en múltiples miembros de una misma familia en un gen conocido por causar la enfermedad, y determinación de si la variación es consistente con la enfermedad (paso F8 del workflow presentado en la **Figura 5**), entre otros.

Según el diagrama de flujo definido en la **Figura 4**, los estudios familiares serían clasificados como *Other*. Debido a la relevancia que tienen estos estudios para el estudio de las variaciones genéticas, tanto en esta como en otras enfermedades de características similares, se hace necesario su inclusión en el diagrama de flujo para la clasificación de artículos como un tipo de estudio concreto.

Una vez se ha definido en qué consisten este tipo de estudios, y sus diferentes subtipos, se deben concretar estrategias para poder identificarlos. En este caso, se ha optado por buscar diferentes palabras clave que aparezcan comúnmente en el abstract de estos estudios, y que sirvan para su identificación. Para la determinación de estas palabras, se ha buscado en diferentes tipos de estudios familiares asociados con el EOAD. Los estudios consultados se reflejan en la **Tabla 2**.

PMID y referencia del artículo	Palabras relevantes
PMID: 28350801, [50]	Families, Relatives, Parental DNA, Familial, Familial cases
PMID: 25948718, [51]	Family, Familial, Generations
PMID: 11710891, [52]	Familial, Family-based case, Families
PMID: 11796781, [53]	Family
PMID: 12433263, [54]	Families, Family history, Familial
PMID: 16033913, [55]	Families
PMID: 10441572, [56]	Generations, Families
PMID: 9521423, [57]	Familial, Kindreds, Families
PMID: 11030797, [58]	Family
PMID: 29404783, [59]	Familial, Family members
PMID: 28532646, [60]	Kindred, Generations

Tabla 2. Palabras clave en los Abstract de Estudios Familiares

De las palabras recogidas en la **Tabla2**, se considerarán como palabras clave las más comunes de las encontradas. Así, se clasificarán como estudios familiares aquellos que contengan algunos de los siguientes términos en el abstract o en el título: *Kindred, Familial, Family/Families, Relatives*.

Buscando complementar las palabras clave, se han determinado términos MeSH que sirvan para identificar los estudios familiares. Tras la búsqueda, se ha encontrado el término *Pedigree* [61] haciendo referencia a los estudios familiares en general, y el término *Twin Study* [62] haciendo referencia a los estudios realizados en hermanos gemelos.

Considerando estos términos, se ha modificado el diagrama de la clasificación de artículos presentado en la **Figura 4** para la inclusión de los estudios familiares, tal y como se muestra en la **Figura 16**.

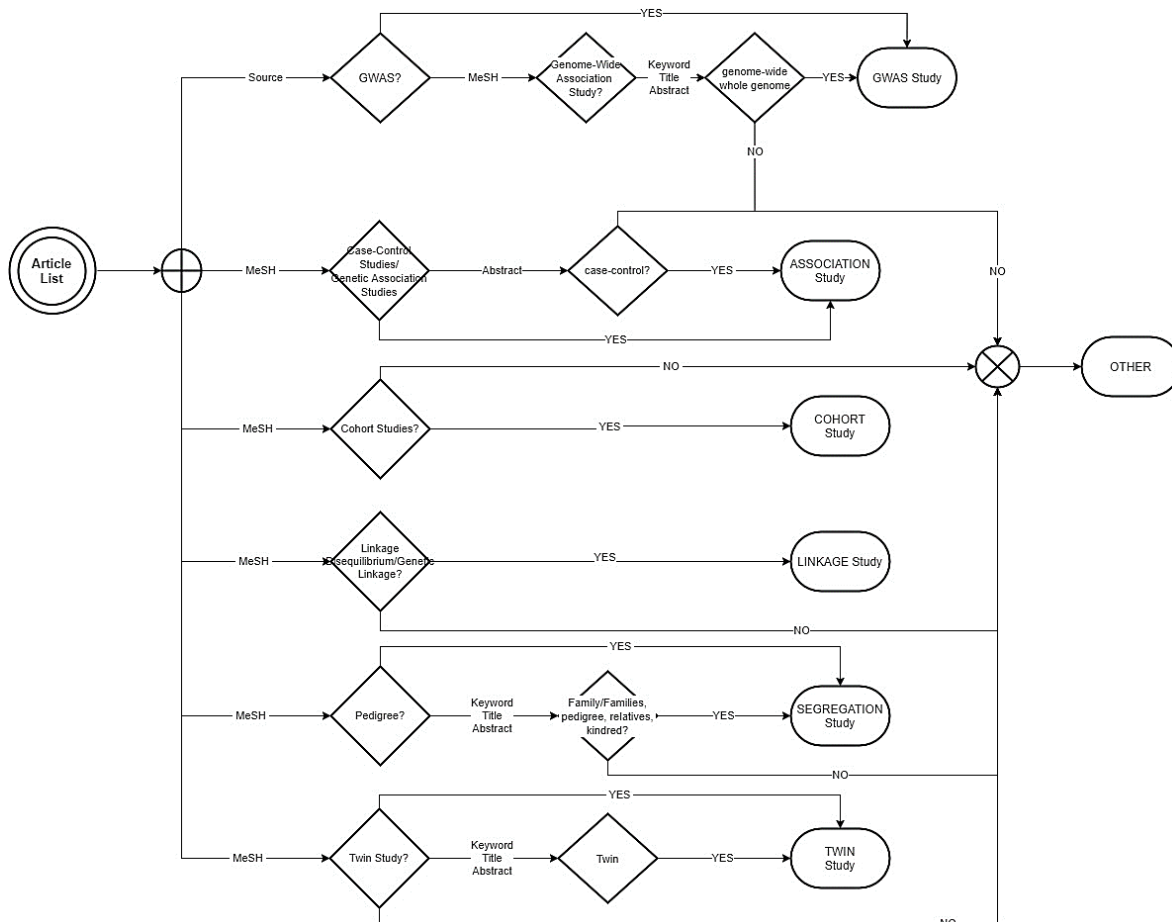


Figura 16. Modificación del diagrama de flujo para la clasificación de artículos

Del análisis derivado de los estudios familiares se han añadido dos tipos de estudio al diagrama de clasificación de artículos:

- **Segregation Study.** Aquellos que estudian la enfermedad en función de su segregación en una o varias familias. Estos constituyen el caso genérico de los estudios familiares.
- **Twin Study.** Caso particular de los estudios en familias en los que se estudia el desarrollo de la enfermedad en hermanos gemelos.

Según lo mostrado en el diagrama, para determinar si un estudio es del tipo familiar, se revisarán primeramente los términos MeSH. En caso de no encontrar ninguno de los términos asociados a este tipo de estudios, se revisará si alguna de las palabras clave determinadas en la **Tabla 2** aparecen en el título o en el abstract. Si no se encontrara ninguna de las palabras clave, el estudio asociado será clasificado como *Other*, o bien como alguno de los otros tipos de estudios considerados.

Una vez se ha realizado el reconocimiento de los estudios familiares como un tipo de estudio significativo, deberían plantearse criterios para poder determinar cuando estos estudios disponen de la relevancia suficiente para ser considerados en el workflow, pues los criterios definidos en SILE para la relevancia de los estudios están enfocados a estudios GWAS y otros estudios de asociación. Por tanto, aunque este tipo de estudios haya sido definido y evaluado, todavía es necesario plantear criterios de evaluación propios para estos. Esto requeriría un profundo análisis de este tipo de estudios para determinar cómo podrían evaluarse y la información a considerar, algo que queda fuera de los objetivos de este trabajo, pero será incluido en la sección de TRABAJOS FUTUROS.

A continuación, se procede a presentar el nuevo workflow propuesto para la identificación de variaciones genéticas relevantes.

6.3. Nuevo Workflow Propuesto para la Etapa de Identificación del método SILE

En el anterior apartado se ha realizado la inclusión de los estudios familiares como un estudio independiente para considerar en el diagrama de clasificación de artículos. Para resolver el resto de las problemáticas identificadas, y poder alcanzar el segundo de los objetivos planteados en este trabajo, en esta sección se va a realizar el diseño de un nuevo workflow, basado en el planteado del método SILE, que busca modificar el workflow actual para tratar casos específicos como el EOAD.

En primer lugar, se presenta el workflow que se ha diseñado teniendo en cuenta los problemas presentados en la sección 6.1:

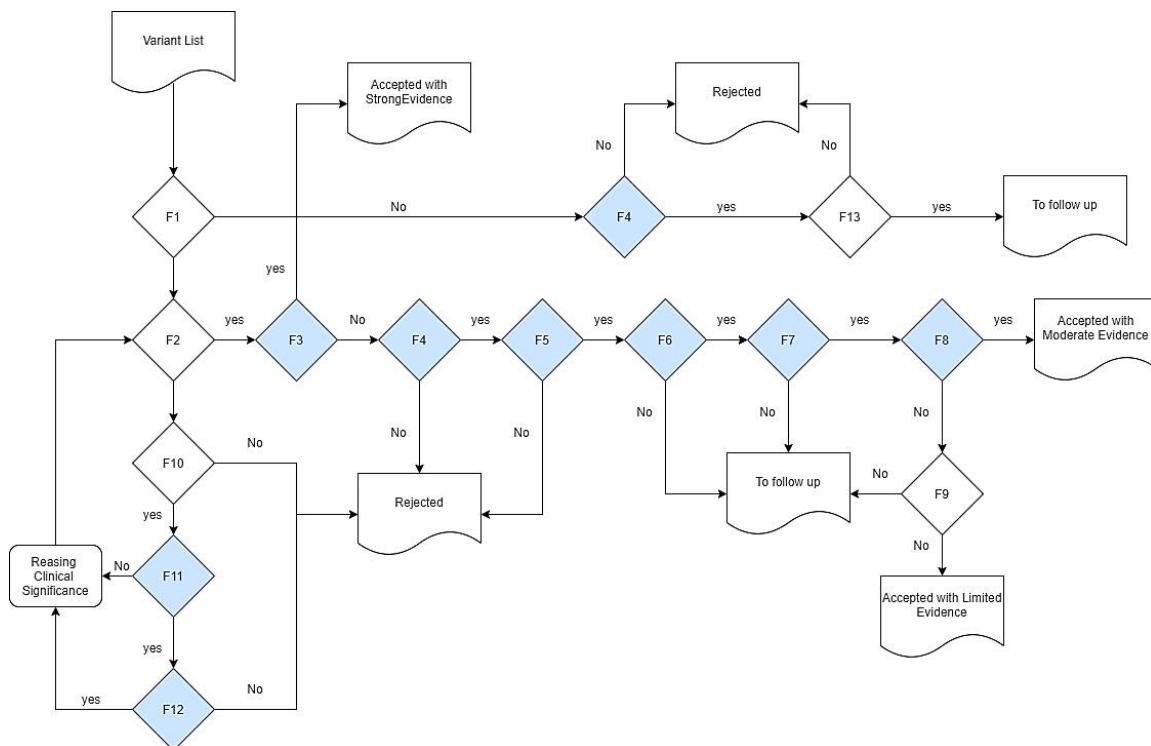


Figura 17. Workflow diseñado para la etapa de identificación

Con el fondo azul se han marcado los pasos que son nuevos respecto al workflow presentado en la Figura 5, mientras que los pasos que se mantienen con el fondo blanco son aquellos que se conservan del workflow inicial, o bien que derivan de otros presentes, pero modificando la forma de implementación.

A continuación, se detalla en que consiste cada una de las etapas del workflow diseñado:

- Evaluación de la relevancia clínica
 - Determinar si la variación ha sido interpretada (F1).
 - Analizar la validez clínica de la interpretación (F2): patogénica, probablemente patogénica y factor de riesgo.
- Selección de las variaciones de acuerdo con la relevancia de los submitters:
 - Guía de práctica clínica o panel de expertos (F3).
 - Submitters con al menos 200 submissions (F7).
- Comprobación de si el gen donde se encuentra localizada la variación genética está relacionado con la enfermedad (F4).
- Evaluación de los métodos y criterios utilizados para la interpretación:
 - Evaluación de si el tipo de método utilizado para obtener la información de la variación es ensayo clínico (F5).
 - Comprobar si el submitter ha proporcionado el assertion criteria utilizado para la interpretación de la variación (F6).
 - Verificar que el assertion criteria utilizado para la interpretación de la variación son las guías ACMG/AMP (F8).

- Evaluación de la última fecha en la que ha sido revisada la variación (F9).
 - La variación debe haber sido revisada hace menos de tres años.
- Tratamiento de las variaciones con significado clínico conflicto de interpretación:
 - Evaluación de la presencia de conflictos en la interpretación (F10).
 - Verificación de que el conflicto aparece en el fenotipo de interés (F11).
 - Evaluación de si al menos el 75% de los remitentes proporciona la misma importancia clínica (F12).
- Tratamiento de las variaciones que no han sido interpretadas:
 - Comprobación de si el gen donde se encuentra localizada la variación genética está relacionado con la enfermedad (F4).
 - Determinación de si hay estudios publicados que apoyen la asociación de la variación con la enfermedad (F13).

En el workflow presentado en la **Figura 17** las variaciones pueden tener ahora 5 clasificaciones:

- **Rechazada.** La variación carece de la evidencia suficiente para ser clasificada como aceptada.
- **Seguimiento.** La variación no tiene en la actualidad la evidencia suficiente para ser clasificada como aceptada, pero puede resultar de interés en un futuro cercano. Esta clasificación no estaba contemplada en el workflow inicial, de forma que será tratada en mayor detalle en la sección 6.3.7.
- **Aceptada con evidencia fuerte.** Solo serán clasificadas de esta forma las variaciones que tengan como submitter una guía práctica clínica o un panel de expertos.
- **Aceptada con evidencia moderada.** El submitter no es una guía de práctica clínica o un panel de expertos, pero la variación se encuentra en un gen conocido por su relación en la enfermedad, el tipo de método es un ensayo clínico, el submitter ha demostrado ser relevante y además como assertion criteria se ha utilizado las guías de la ACMP/AMG.
- **Aceptada con evidencia limitada.** Se cumplen todas las condiciones anteriores, pero en este caso no se han utilizado las guías de la ACMP/AMG, aunque sí que se ha proporcionado un assertion criteria. Además, se ha comprobado que la evidencia disponible es reciente.

Una vez se ha presentado el nuevo workflow propuesto, se puede ver que existen múltiples diferencias respecto al workflow original (**Figura 5**), que están orientadas a resolver todos los problemas identificados en la sección 6.1. En los siguientes apartados, se va a detallar la relación entre todos los problemas identificados y las diferencias presentes en el nuevo workflow respecto al workflow original. La solución a los problemas va a ser planteada en el orden en el que han sido resueltos en el workflow diseñado, buscando con esto además justificar la lógica utilizada para el diseño.

6.3.1. Solución al problema 3. Tratamiento de las Variaciones con Significado Clínico Conflictivo de Interpretación (pasos F10-F11-F12)

En tercero de los problemas identificados en el apartado 6.1 hacía referencia a las variaciones genéticas clasificadas como conflicto de interpretación. El problema detectado es que, en ocasiones, a las variaciones que habían sido interpretadas para más de un fenotipo, en ClinVar se les asignaba de forma incorrecta la clasificación de conflicto de interpretación, lo que ocasiona que variaciones que puedan ser relevantes para la enfermedad sean incorrectamente clasificadas por el workflow inicial de SILE (**Figura 5**).

La solución planteada para este problema se recoge en los pasos F10 a F12 del workflow presentado en la **Figura 17**. En el primer paso (F10) se seleccionarán solo las variaciones que tengan un significado clínico conflictivo. Debido a que en el paso F2 se seleccionan las variaciones con interpretaciones relevantes a nivel clínico (patogénicas, probablemente patogénicas y factor de riesgo) en el paso F10 se clasificarán como *Rechazada* aquellas variaciones de significado clínico benigno, probablemente benigno, y significado incierto.

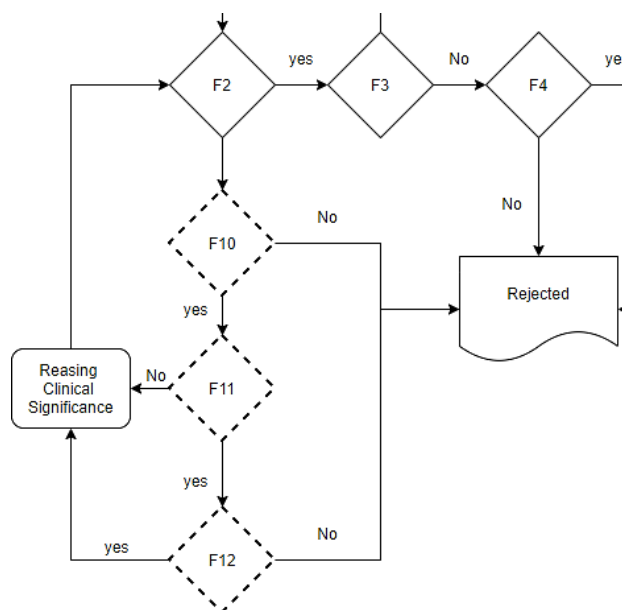


Figura 18. Detalle del workflow con los pasos tratados (F10-F12) resaltados

En el siguiente paso (F11), se evaluará si el conflicto de interpretación ocurre realmente en el fenotipo que es objeto de estudio. En el caso el conflicto no ocurra en el fenotipo estudiado, se realizará la reasignación de la interpretación al significado clínico correspondiente. Si el conflicto sí se da en el fenotipo de interés, se pasa al paso F12.

En el paso F12 la cuestión que se plantea es si realmente la variación debería ser considerada como un conflicto de interpretación, si el conflicto se produce en el fenotipo de interés, pero existe un acuerdo mayoritario en las interpretaciones que han hecho los submitters de la variación (sin ser

ninguno de los submitter una guía de práctica clínica o un panel de expertos). En estos casos, cuando el conflicto de interpretación sea en el fenotipo de interés, se comprobará si existe un acuerdo en la interpretación de la variación en al menos el 75% de los submitters, y en el caso de que así sea, se producirá la reasignación del significado clínico al mayormente aceptado. En caso contrario, se considerará que realmente existe un conflicto de interpretación y la variación será rechazada.

Una vez se ha reasignado el significado en las variaciones que corresponda, se deben seleccionar solo aquellas variaciones que sean relevantes a nivel clínico (paso F2). Tras esto, se deberá evaluar la calidad de la información disponible para cada una de estas variaciones, siendo el primer paso el análisis de la relevancia de los submitters (pasos F3 y F7). Con esto, se dará una solución a parte de la problemática resumida en el problema 4.

6.3.2. Solución al problema 4.1. Selección de las Variaciones con los Submitters más Importantes (paso F3-F7)

Otra de las problemáticas identificadas en el problema 4 se relacionaba con la ausencia de consideración en el workflow original (**Figura 5**) de la relevancia de los submitters como un criterio de calidad de la información relacionada con la variación. Otros de los aspectos identificados bajo este problema, falta de consideración del tipo de método y del assertion criteria, serán tratados en otros apartados.

Como se expone en la sección 5.2.1 las submissions realizadas por guías de práctica clínica o por paneles de expertos tienen prioridad sobre el resto de las interpretaciones realizadas por otros submitters. Esto significa que, si todos los submitters clasifican la variación como benigna, pero una guía de práctica clínica o panel de expertos la considera como patogénica, la variación será clasificada como patogénica.

Debido al gran impacto que tienen estos submitters sobre la veracidad y calidad de la interpretación de las variaciones, se ha considerado como fundamental incluir un paso que compruebe si el submitter es uno de los dos mencionados (F3). De esta forma, solo se clasifican como aceptadas con evidencia fuerte aquellas variaciones que han sido interpretadas por este tipo de submitters o el review status equivalente (3 ó 4 estrellas).

Existen submitters que son fiables y que proporcionan información de calidad, sin ser guías de práctica clínica o paneles de expertos, como por ejemplo grandes laboratorios especializados en análisis genético, que recopilan una gran cantidad de información acerca de variaciones genéticas relacionadas con diferentes tipos de enfermedades. Teniendo esto en consideración, se considerará también un criterio de calidad, pero con un menor nivel de evidencia, el hecho de que el submitter sea otro tipo de fuente de gran reputación (F7). Debido a que estos tipos de laboratorios tienen un gran número de submissions en las bases de datos, la evaluación de la relevancia del submitter en estos casos se hará a través del número de submissions, concretamente, serán necesarias más de 200 submissions en la base de datos.

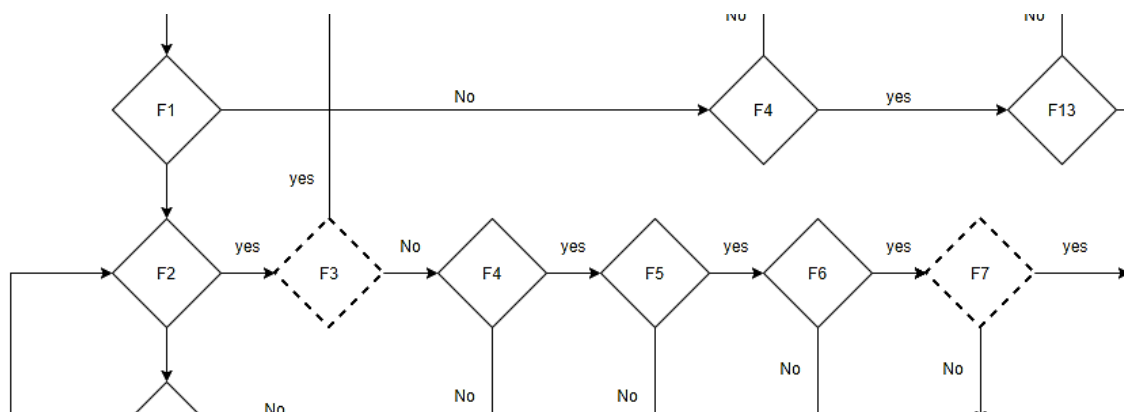


Figura 19. Detalle del workflow con los pasos tratados (F3 y F7) resaltados

Con el paso F3 se ha introducido el criterio a evaluar para que una variación sea clasificada como *Aceptada con evidencia fuerte*. En el siguiente apartado, se va a completar la solución al problema 4 incluyendo la consideración del tipo de método en el workflow mediante el paso F5. Este paso, junto con los pasos F4, F6, F8 y F9 forma parte de la evaluación de un nivel menor de evidencia para las variaciones que han sido clasificadas como aceptadas.

En el siguiente paso se buscará determinar cómo ha sido obtenida la información relacionada con la variación. Entender el origen de la información, tal y como se explica en el siguiente apartado, será útil como un primer paso para determinar su calidad.

6.3.3. Solución al problema 4.2. Evaluación del Tipo de Método Utilizado para Obtener la Información de la Variación (paso F5)

Otro de los aspectos que se contemplaban dentro del problema 4 era la falta de consideración del tipo de método utilizado para la identificación de variaciones genéticas relevantes. Tal y como se ha explicado en la sección 5.2.4 existen tres tipos de métodos que pueden ser utilizados por los submitters para obtener la información relativa a las variaciones genéticas: ensayo clínico, literatura e investigación.

El objetivo es seleccionar aquellas variaciones que estén más probablemente relacionadas con la enfermedad, para lo que se necesita garantizar que la información en la que se ha basado la interpretación de la variación es fiable y de calidad. El problema detectado con los tipos de método basados en literatura y los basados en investigación es que las interpretaciones que se hagan de las variaciones están más sujetas a cómo interprete el usuario una determinada información disponible, algo que no ocurre con el método de ensayo clínico, donde las interpretaciones se fundamentan en los resultados obtenidos en el ensayo clínico.

Debido a que se busca garantizar la fiabilidad y calidad de la información que esté relacionada con las variaciones, se ha considerado importante incluir este paso donde se seleccionarán únicamente las variaciones que tengan submissions con método ensayo clínico (F5).

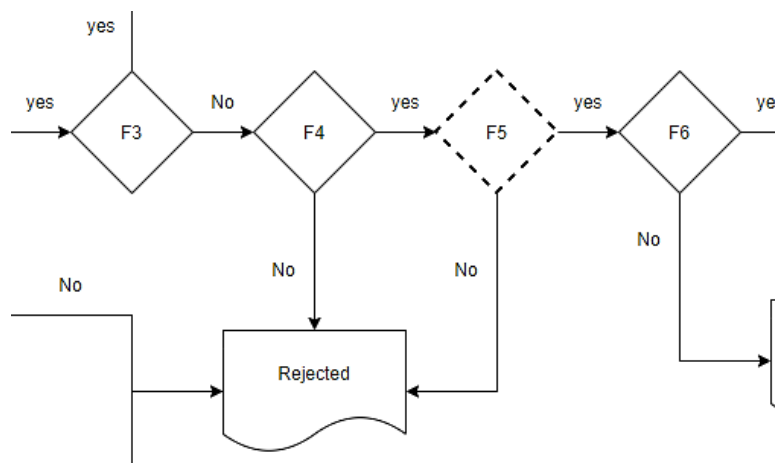


Figura 20. Detalle del workflow con el paso tratado (F5) resaltado

En caso de que la variación no disponga de ninguna submission con el método deseado, la variación será clasificada como *Seguimiento*.

En el siguiente apartado se va a dar solución al problema 5, donde se ha identificado como problema el hecho de no realizar una validación previa de los resultados obtenidos tras la búsqueda de variaciones genéticas en la base de datos.

6.3.4. Solución al problema 5. Comprobación de si el Gen donde se Encuentra Localizada la Variación Genética Está Relacionado con la Enfermedad (paso F4)

Buscando dar solución al problema 5, identificado en la sección 6.1, se va a buscar realizar una comprobación previa de la consistencia de los resultados de la búsqueda en la base de datos, mediante la comprobación del gen en el que se localiza la variación.

Como se mencionó en este apartado, existe una gran cantidad de información en las bases de datos que hace que, si no se busca la información de forma correcta, puedan encontrarse variaciones que no estén realmente relacionadas con el fenotipo de interés. De ahí nace la necesidad de incluir este paso (F4) para garantizar la consistencia de los resultados obtenidos con el fenotipo de estudio, en este caso mediante el análisis del gen en el que se localiza la variación.

Para determinar que genes están relacionados con la enfermedad de la que se quieren determinar las variaciones genéticas relevantes, es necesario realizar una investigación previa acerca de la enfermedad de estudio, tal y cómo se ha realizado en la sección 4 de este documento.

Como se muestra en el workflow diseñado en la **Figura 17**, en el caso de que las variaciones no se encuentren en un gen de interés para la enfermedad serán rechazadas.

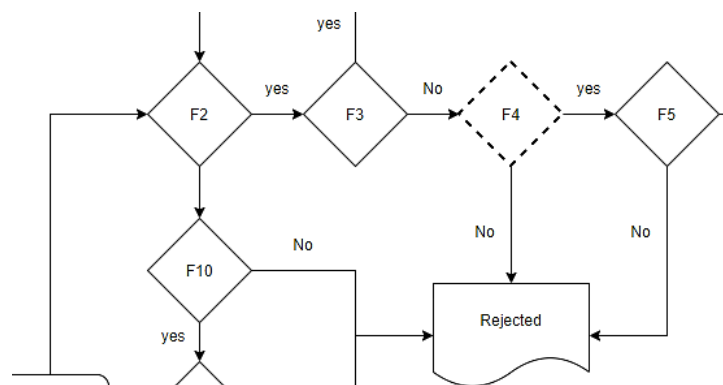


Figura 21. Detalle del workflow con el paso tratado (F4) resaltado

Con el paso F4, se ha verificado la relación del gen en el que se encuentra localizada la variación con el fenotipo de interés, mientras que con los pasos F5 y F7 se ha evaluado la calidad de la información utilizada para realizar la interpretación de la variación. Una vez se ha asegurado la calidad, es necesario evaluar en base a qué criterios se ha realizado la interpretación de la variación, buscando determinar si estos son o no fiables (pasos F6 y F8). A través del análisis de los criterios se va a dar solución a parte del problema 4, donde se destacaba la importancia de los criterios utilizados como forma de validación de la interpretación, y también al primero de los problemas pues la inclusión de estos pasos permitirá una evaluación diferente de la relevancia estadística de los estudios relacionados.

6.3.5. Solución al problema 1 y 4.3. Verificar el Assertion Criteria Utilizado para la Interpretación de la Variación (pasos F6-F8)

En la sección 6.1 se planteó como una parte problema 4 la ausencia de consideración del assertion criteria utilizado para la interpretación. En este apartado se va a exponer cómo se ha realizado la inclusión de la evaluación del assertion criteria en el nuevo workflow propuesto. Esto permitirá, concretamente con el paso F8, solucionar el primer problema identificado en la sección 6.1, donde se expuso la dificultad de la evaluación de la relevancia estadística de los estudios relacionados en casos particulares como el EOAD.

Para el análisis del criterio utilizado en la interpretación, en primer lugar, se determinará mediante el paso F6 si se ha proporcionado un criterio para la interpretación de la variación. En caso de que no se haya proporcionado un criterio, al no se dispusere de medios para determinar si la interpretación de la variación es o no fiable, la variación será clasificada como *Seguimiento*.

En segundo lugar, con el paso F8, se evaluará si para realizar la interpretación de la variación se han utilizado como referencia los criterios definidos en las guías ACMG/AMP. La ACMG/AMP

(American College of Medical Genomics/ Association for Molecular Pathology) publicó en el año 2015 *Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology* [63], unas guías que pretenden establecer una serie de criterios para la determinación de la interpretación más adecuada (patogénica, probablemente patogénica, benigna, probablemente benigna y significado incierto) para cada variación genética en enfermedades de herencia mendeliana.

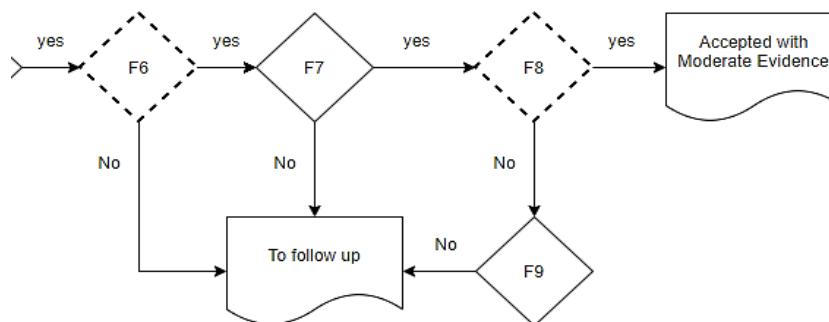


Figura 22. Detalle del workflow con los pasos tratados (F6 y F8) resaltados

Las guías permiten asignar, en base a unos criterios que evalúan la relación entre la variación y el fenotipo, la interpretación más adecuada a cada variación. Estos criterios se agrupan en 4 niveles de evidencia (*very strong*, *strong*, *moderate* y *supporting*), en función del grado de evidencia que proporciona cada uno de los criterios de cada nivel acerca de la patogenicidad de una variación.

Los criterios definidos en estas guías están aceptados internacionalmente de forma que, si son el criterio utilizado para realizar la interpretación de la variación, es una garantía de la fiabilidad del significado clínico de la variación.

La introducción de las guías como criterio para evaluar la relevancia de las variaciones aporta varias ventajas. Primeramente, algunos de los criterios incluidos en las guías se corresponden con los definidos en el workflow inicial (F5-F7-F8), de forma que con un único paso se pueden evaluar los siguientes pasos del workflow original del método SILE presentado en la **Figura 5**:

1. Evaluación de la relevancia de los estudios en términos de significancia estadística (F5). Este criterio se corresponde con el cuarto criterio de evidencia fuerte para la patogenicidad de la variación en las guías ACMG/AMP: *La prevalencia de la variación en los individuos afectados está aumentada significativamente respecto a los controles. Esto se determinará evaluando el odd-ratio >5 y el intervalo de confianza que no debe contener el 1.*
2. Evaluación del MAF (*Minor Allele Frequency*). En este caso, en las guías se evalúa este criterio para determinar la benignidad de la variación, buscando determinar si el MAF es mayor que el esperado para la enfermedad, siendo un criterio de evidencia fuerte: *La frecuencia del alelo es mayor que el esperado para la enfermedad.*

3. Evaluación de la consistencia de la variación con la enfermedad (F8). Debido a que la valoración de este criterio es más amplia, se puede establecer una relación aproximadamente análoga con varios de los criterios de las guías ACMG/AMP como puede ser: *Estudios de tipo funcional, ya sea in-vitro o in-vivo, que den apoyo al efecto dañino de la variación en el gen o en su producto* como criterio de evidencia fuerte de patogenicidad, *Variación nula (null variant) que causa la pérdida de función en un gen, cuya pérdida de función está típicamente relacionada con la enfermedad* como criterio de evidencia muy fuerte, etc.

En la sección 6.1 se identificó como uno de los problemas (problema 1) la forma en la que era evaluada la relevancia estadística. Cómo se puede ver en el primero de los puntos tratados, en las guías se mantiene la evaluación de los criterios estadísticos, ya que tal y cómo se ha mencionado es un criterio ampliamente aceptado y utilizado. La ventaja que aporta la evaluación a través de las guías y no de forma directa es que, aunque este criterio es de los que se consideran como relevantes en las guías, no es necesario que de forma irrevocable se cumpla para poder clasificar una variación como patogénica.

Explicando esto en mayor detalle, en la **Figura 23** se puede ver que para clasificar la variación como patogénica se evalúa el cumplimiento de una combinación de criterios de diferentes niveles de evidencia. Por tanto, aunque el criterio de la relevancia estadística se tiene en cuenta, no necesariamente por el no cumplimiento de este criterio será clasificada una variación como *Not Enough Evidence Provided* como ocurría en el workflow original.

Pathogenic	(i) 1 Very strong (PVS1) AND (a) ≥ 1 Strong (PS1-PS4) OR (b) ≥ 2 Moderate (PM1-PM6) OR (c) 1 Moderate (PM1-PM6) and 1 supporting (PP1-PP5) OR (d) ≥ 2 Supporting (PP1-PP5) (ii) ≥ 2 Strong (PS1-PS4) OR (iii) 1 Strong (PS1-PS4) AND (a) ≥ 3 Moderate (PM1-PM6) OR (b) 2 Moderate (PM1-PM6) AND ≥ 2 Supporting (PP1-PP5) OR (c) 1 Moderate (PM1-PM6) AND ≥ 4 supporting (PP1-PP5)
Likely pathogenic	(i) 1 Very strong (PVS1) AND 1 moderate (PM1-PM6) OR (ii) 1 Strong (PS1-PS4) AND 1-2 moderate (PM1-PM6) OR (iii) 1 Strong (PS1-PS4) AND ≥ 2 supporting (PP1-PP5) OR (iv) ≥ 3 Moderate (PM1-PM6) OR (v) 2 Moderate (PM1-PM6) AND ≥ 2 supporting (PP1-PP5) OR (vi) 1 Moderate (PM1-PM6) AND ≥ 4 supporting (PP1-PP5)

Figura 23. Combinación de criterios de las guías para la clasificación de las secuencias patogénicas [63]

Con lo expuesto, gracias al uso de las guías ACMG/AMP se ha solucionado uno de los principales problemas identificados en el workflow inicial, donde la forma de evaluar la relevancia estadística de los artículos no permitía la correcta identificación de variaciones relevantes en casos específicos como el EOAD.

En el caso de que no se hayan aplicado las guías de la ACMG/AMP, pero si se haya proporcionado otro criterio para la interpretación, se considerará la fecha en la que se realizó la interpretación para asegurar que se haya realizado recientemente, tal y como se explica en el siguiente apartado. Con la introducción de la fecha en la que se realizó la interpretación en este punto del workflow y no antes, se dará solución al problema número 6.

6.3.6. Solución al problema 6. Evaluación de la Última Fecha en la que Ha Sido Revisada la Variación (F9).

Otra modificación realizada es referente a la forma en la que se evalúa el criterio temporal establecido en el workflow original de la **Figura 5**, buscando con esto dar solución al problema 6 de los identificados. Según lo expuesto en la sección 6.1, el problema identificado se debe a que variaciones con unos criterios muy sólidos acerca de su patogenicidad podían ser clasificadas como *Limited Evidence* debido únicamente al criterio temporal.

Este criterio no puede simplemente descartarse, pues tal y cómo se ha mencionado la cantidad de información genética ha crecido de forma masiva en los últimos años. Este crecimiento de la información ha venido acompañado de un crecimiento también en las técnicas y criterios utilizados para determinar la relevancia de las variaciones genéticas. Esto hace que, si una variación ha sido recientemente revisada, sea más probable que los criterios utilizados para su interpretación estén basados en el conocimiento actual de la rama de conocimiento, y por tanto es más probable que la interpretación sea más fiable.

Adicionalmente a lo expuesto, el hecho de que los criterios utilizados para la interpretación no sean los determinados por las guías ACMG/AMP no necesariamente implica que los criterios utilizados no sean válidos. Con esto, se ha determinado el aceptar con una evidencia limitada aquellas variaciones revisadas hace menos de tres años, y cuyo assertion criteria no sean las guías ACMG/AMP (pero sí se haya proporcionado uno).

Así, aunque en la mejora del workflow planteada en la **Figura 17** también se evalúa lo reciente que ha sido realizada la interpretación, se da prioridad a criterios que evalúan la relevancia y significancia de evidencia disponible acerca de la variación a la hora de determinar el grado de aceptación de esta.

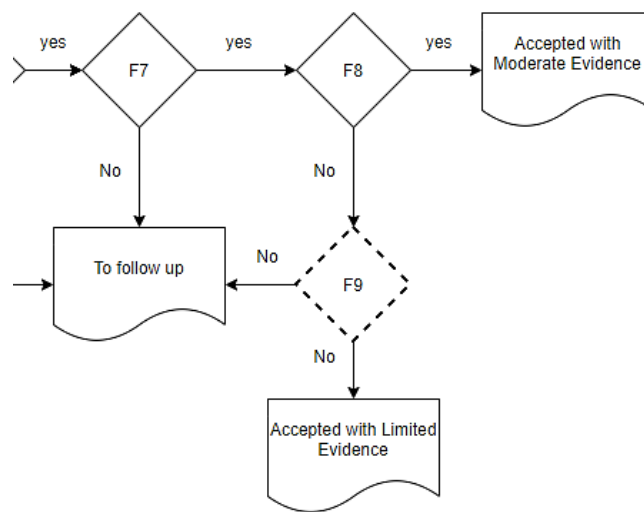


Figura 24. Detalle del workflow con el paso tratado (F9) resaltado

Con todo lo expuesto en los apartados 6.2, 6.3.1, 6.3.3, 6.3.4, 6.3.5 y 6.3.6 se ha conseguido dar solución a todos los problemas identificados acerca del workflow original definido en el método SILE. Sin embargo, hay dos diferencias respecto al workflow original, inclusión de la clasificación *Seguimiento* y el tratamiento de las variaciones que no han sido interpretadas que, aunque no responden a problemas concretos identificados, deben ser explicadas en mayor detalle para justificar su inclusión en el nuevo workflow.

6.3.7. Inclusión de Clasificación Seguimiento en el Nuevo Workflow

En la sección 1.2 se introdujo como las nuevas técnicas de secuenciación masiva han permitido un gran crecimiento de la información genética disponible. Debido a este crecimiento, la información relativa a las variaciones genéticas crece y cambia con el paso del tiempo. Esto significa que la clasificación de una variación puede cambiar con el tiempo, a medida que vayan apareciendo nuevas evidencias acerca de esta. Por tanto, según lo mencionado, la clasificación que se hace de las variaciones con el workflow presentado va a ser siempre dinámica, y puede cambiar con el paso del tiempo.

Teniendo esto en cuenta, habrá variaciones que no dispongan de la evidencia suficiente para ser consideradas como aceptadas en la actualidad, pero que potencialmente puedan disponer de la misma en un futuro cercano. Esto hace que haya variaciones que sea necesario ir revisando cada cierto tiempo para comprobar si se han añadido nuevas evidencias que puedan modificar la clasificación de la variación para que pase a ser aceptada.

Son este tipo de variaciones las que serán clasificadas en el workflow propuesto en la **Figura 17** como *Seguimiento*, pues son aquellas variaciones que pueden ser potencialmente relevantes en un futuro cercano, y que es necesario ir revisando con regularidad para determinar si ha habido algún cambio.

Finalmente, en la siguiente sección se va a exponer como se va a realizar el tratamiento de aquellas variaciones cuyo significado clínico no haya sido proporcionado, ya que en ocasiones pueden disponer de información asociada que las haga de interés para el fenotipo estudiado.

6.3.8. Evaluación de las Variaciones que No Dispongan de un Significado Clínico (pasos F1-F4-F13)

Algunas variaciones que no han sido interpretadas, es decir, en las que no se dispone de información acerca de si la variación es patogénica o benigna, pueden disponer de estudios estadísticamente relevantes que sirvan de apoyo a la relación entre la variación y la enfermedad.

Para determinar si existen evidencias para estas variaciones, en primer lugar, con el paso F1 se seleccionarán aquellas variaciones que no dispongan de una interpretación, tal y cómo se hacía en el workflow original del método SILE (**Figura 5**). Tras esto, al igual que con el resto de las variaciones en el workflow propuesto, se realizará una verificación de que la variación se encuentra a priori relacionada con el fenotipo de estudio mediante la comprobación de que el gen en el que se localiza la variación está relacionado con la enfermedad (paso F4). Este paso en este tipo de variaciones es especialmente importante, pues al disponer de información limitada acerca de la interpretación de la variación, debe comprobarse cualquier evidencia que sirva para establecer la relación con la enfermedad.

En el siguiente paso, F13, se evaluará si existen estudios que sean estadísticamente relevantes que indique la relación variación-enfermedad. Para considerar como relevante un estudio, se seguirán los mismos criterios que los utilizados en el workflow original del método SILE. El motivo de retomar la evaluación directa de la relevancia estadística es que, al disponer de información limitada acerca de estas variaciones, los pasos F5 a F8 del nuevo workflow propuesto no pueden ser evaluados, de forma que la relevancia estadística de los estudios cobra más protagonismo en este caso.

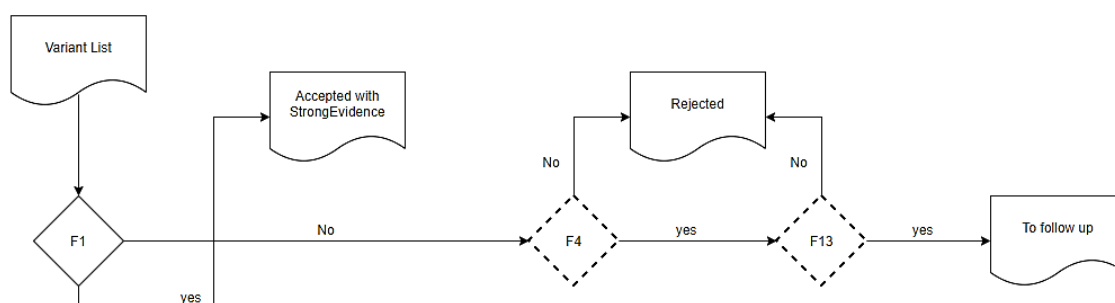


Figura 25. Detalle del workflow con los pasos tratados (F4 y F13) resaltados

Si se encuentran estudios relacionados con la variación, se tendrá un indicador de que la variación está posiblemente relacionada con la enfermedad, pero no se considera que se disponga de la evidencia necesaria para clasificar como aceptadas ninguna de estas variaciones. Debido a esto, estas variaciones serán clasificadas como *Seguimiento*.

Una vez se presentado el workflow diseñado para la identificación de variaciones relevantes, es necesario aplicarlo a la enfermedad de estudio, que en este caso es el EOAD. El objetivo de esa aplicación es realizar una evaluación del nuevo workflow propuesto, para determinar los resultados que da su aplicación a casos específicos como el EOAD.

7. APLICACIÓN DE LA MEJORA PROPUESTA PARA LA ETAPA DE IDENTIFICACIÓN: RESULTADOS Y DISCUSIÓN

Una vez ha sido propuesta la mejora para la etapa de identificación del método SILE, se va a realizar una evaluación de los resultados que proporciona, en este caso, su aplicación al EOAD. En primer lugar, se va a realizar un análisis general de los resultados que nos proporciona ClinVar en la búsqueda de variaciones genéticas relacionadas con el Alzheimer temprano. Tras esto, en base a los resultados obtenidos, se realizará la aplicación del nuevo workflow propuesto (**Figura 17**) para la identificación de variaciones genéticas, obteniendo así un listado reducido de variaciones relevantes para la enfermedad. Finalmente, se discutirán los resultados obtenidos buscando analizar en mayor profundidad las variaciones que han resultado aceptadas.

7.1. Resultados de la Búsqueda de Variaciones Genéticas Relacionadas con el EOAD en ClinVar

En primer lugar, previo a realizar la aplicación del workflow diseñado en el apartado 6.3, va a realizarse un análisis a nivel global de los resultados obtenidos para la búsqueda de variaciones genéticas relacionadas con el EOAD en ClinVar.

Tal y como se puede ver en la **Figura 26**, un total de 276 variaciones son obtenidas como resultado de la búsqueda en ClinVar a fecha 18/05/2020. La base de datos se actualiza de forma mensual, de forma que es importante considerar que el número de variaciones disponible, y por tanto los resultados de la aplicación del workflow obtenidos en este trabajo, pueden variar en función de la fecha de aplicación.

The screenshot shows the ClinVar search interface. The search query is: `((alzheimer[Disease/Phenotype] AND *early onset*[Disease/Phenotype])) OR ((alzheimer[Disease/Phenotype] AND (*type 1*[Disease/Phenotype] O))`. The results are displayed in a table with the following columns: Variation Location, Gene(s), Protein change, Condition(s), Clinical significance (Last reviewed), and a status column. The first two items are:

Variation Location	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Status
1. NM_019112.3(ABCA7)c.2128_2132del(p.Glu709fs) GRCa37: Chr19:1047508-1047514 GRCa38: Chr19:1047509-1047515	ABCA7	E709fs	not provided, Alzheimer disease, type 9, Early-onset Alzheimer's disease	Conflicting interpretations of pathogenicity, risk factor (Dec 31, 2019)	crit inte
2. NM_000484.4(APP)c.1090C>T(p.Leu364Phe)	APP	L364F, L308F	Alzheimer disease, Early-onset Alzheimer's disease	Uncertain significance (Oct 28, 2018)	crit sut

Figura 26. Resultados de la búsqueda en ClinVar a fecha 18/05/2020

En la parte izquierda de la **Figura 26** pueden apreciarse diferentes características a nivel global de las variaciones obtenidas, como es el número de variaciones para cada significado clínico, su consecuencia molecular, el tipo de método utilizado, el gen en el que se localizan, etc. Algunas de estas características van a permitir realizar un análisis general de los resultados proporcionados por la base de datos, y adicionalmente mejorar la comprensión de los resultados de la aplicación del workflow sobre el listado de variaciones, que se va a realizar en el siguiente apartado. Por esto, a continuación, van a analizarse las características que se consideran más relevantes.

Primeramente, haciendo referencia al significado clínico de las variaciones obtenidas, en la **Figura 27** se muestran el número de variaciones para cada uno de los posibles significados clínicos disponibles en la base de datos.

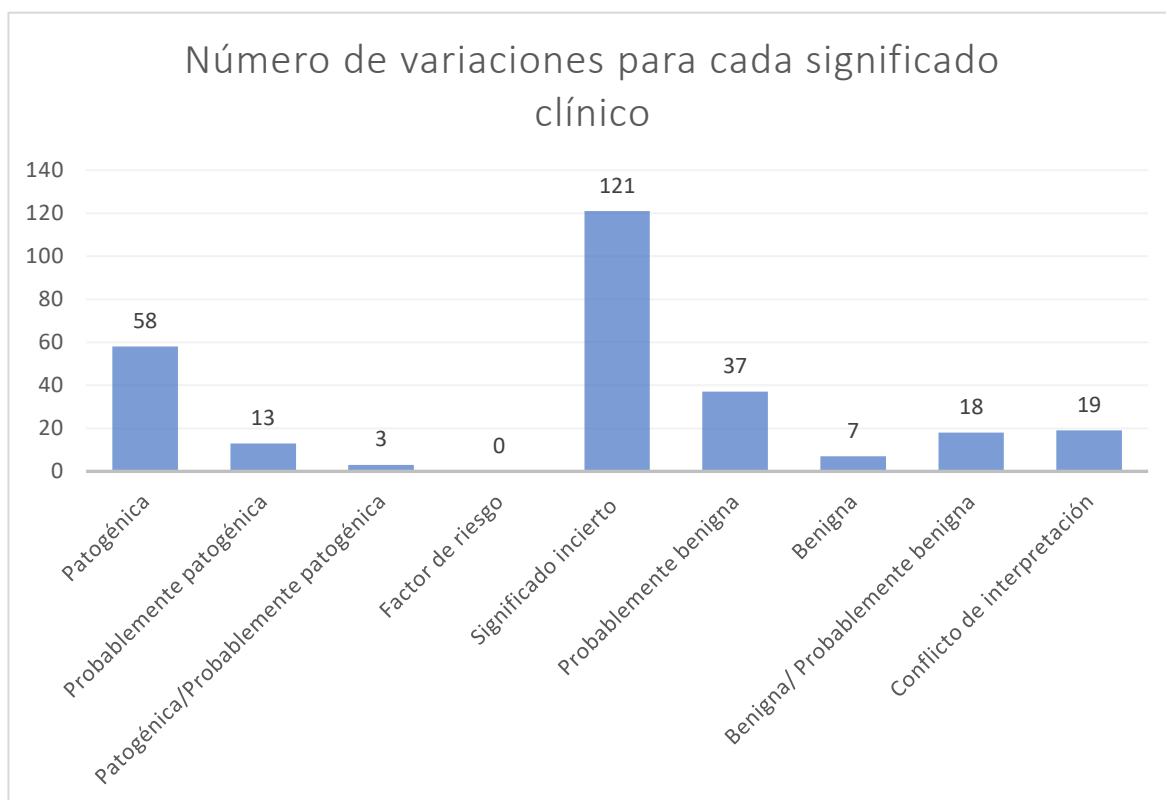


Figura 27. Número de variaciones para cada significado clínico

En la **Figura 27** se puede ver como aparecen dos significados clínicos que no habían sido mencionados previamente: benigna/probablemente benigna y patogénica/probablemente patogénica. Debido a que la combinación de significados clínicos benigno y probablemente benigno, y patogénico y probablemente patogénico no generan un conflicto en la interpretación, según se determinó en el apartado 5.2.2, habrá variaciones que puedan ser clasificadas como benigna-probablemente benigna y patogénica-probablemente patogénica por diferentes submitters, teniendo por tanto dos significados clínicos. Debido a que el significado clínico es uno de los factores considerados en el nuevo workflow propuesto, se analizará en mayor detalle en el siguiente apartado.

Otro aspecto que llama la atención de la **Figura 27** es que aproximadamente la mitad de las variaciones presenta un significado clínico incierto. Si analizamos por separado estas variaciones, destaca que la mayoría tienen un único submitter siendo este en la mayoría de las variaciones los laboratorios *Illumina Clinical Services Laboratory*. Estos laboratorios ofrecen servicios de diagnóstico del genoma completo, habiendo secuenciado el genoma de hasta 2000 adultos asintomáticos. Con esto, resulta probable que se hayan encontrado esas variaciones en pacientes con sospecha de padecer la enfermedad, pero en las que o bien no se ha podido confirmar la presencia de la enfermedad, o bien no se ha podido sustentar una relación variación-enfermedad y de ahí su significado incierto [64].

Otra característica para analizar en las variaciones es el tipo de consecuencia molecular, que se relaciona con el impacto que puede tener la variación sobre el gen en el que está localiza. Las variaciones encontradas en relación con el EOAD se distribuyen en 5 tipos de consecuencias molecular:

- **Missense.** Cambio puntual en la secuencia de ADN que se traduce en un cambio en un aminoácido de la proteína para la que codifica ese gen [65].
- **UTR.** Variaciones que se localizan en la sección no traducida del gen (sus extremos).
- **Splice site.** Se localizan en la zona de unión entre un intrón y un exón [66].
- **Frameshift.** Se produce la inserción o deleción de pares de bases en un número no múltiplo de tres, alterando el patrón de lectura del gen en su traducción a proteína [65].
- **ncRNA.** Son variaciones producidas en regiones que, aunque si son transcritas, no tiene capacidad para codificar en proteínas [67].

En la **Figura 28** se puede observar cómo se distribuye el número de variaciones para cada una de estas consecuencias moleculares:

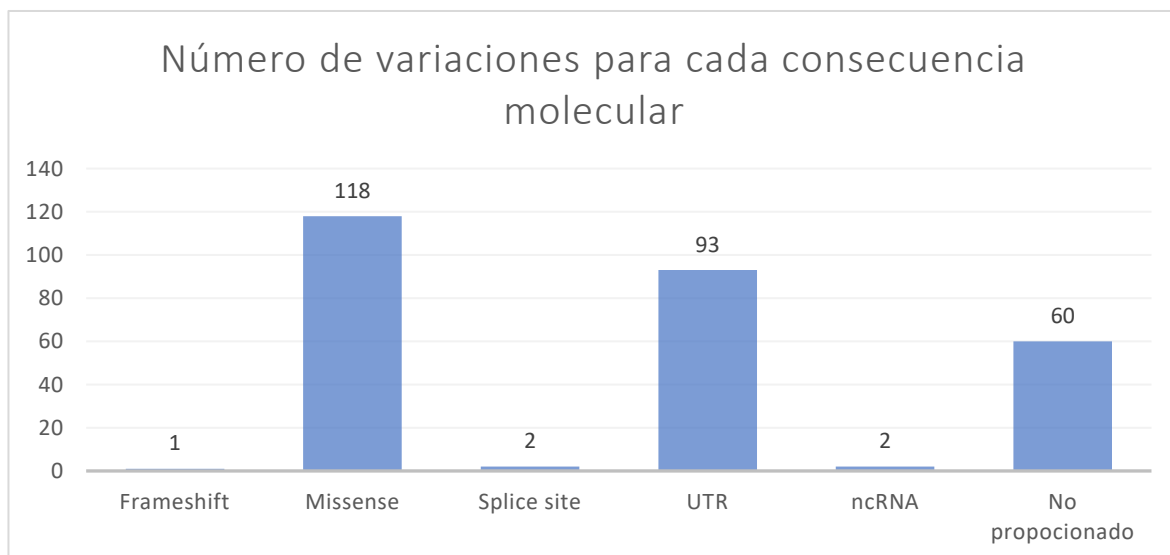


Figura 28. Número de variaciones para cada consecuencia molecular

Según los resultados reflejados en la **Figura 28**, la mayoría de las variaciones producen un cambio de aminoácido, pues 118 de las 276 variaciones son de tipo *missense*.

El siguiente tipo de variación más común según su consecuencia molecular es el tipo UTR, cuyo efecto difiere en función de si se localizan en el extremo 3' o en el extremo 5' del gen. En el extremo 5', las variaciones tienden a producir fallos en las síntesis de proteínas mientras que el caso del extremo 3', al tener el fragmento 3' UTR un papel fundamental en la traducción del mRNA, las variaciones en esta región pueden llevar a una desregulación de la traducción, aunque todavía se requiere una mayor investigación de las enfermedades relacionadas [68]. En el caso concreto de las localizadas en el fragmento 3'UTR, se han encontrado estudios que relacionan la presencia de variaciones en este extremo con el riesgo de padecer Alzheimer a través de cambios en la modulación de la expresión del gen APP, cuya relación con el EOAD ha sido determinada en la sección 4.1, modificando los niveles de proteína para la cual codifica este gen. En concreto, en el estudio publicado por Delay, C., Calon, F., Mathews, P. et al.[69], se hace referencia a 3 variaciones en esta región que pueden estar relacionadas con un aumento en el riesgo de padecer Alzheimer: A454G (reduce la expresión del APP), T171C (aumenta la expresión de APP) y A833C de la que no se proporciona información, pero ninguna de las variaciones a las que hace referencia el artículo han sido encontradas en ClinVar.

Los otros tipos de consecuencia molecular encontradas son menos comunes. En el caso de las variaciones del tipo *splice site*, al localizarse en el límite entre un exón y un intrón, estas pueden ocasionar alteraciones de la secuencia proteica debido a la introducción parte del intrón en la zona a traducir, o a la pérdida de parte del exón [66]. En el caso de las variaciones del tipo *frameshift*, la traducción de un gen con este tipo de variaciones suele dar lugar a proteínas que han perdido completamente su funcionalidad [70]. Finalmente, en el caso de las variaciones *ncRNA*, es complejo establecer un impacto a nivel global, ya que los diversos subtipos existentes en este tipo de variaciones difieren en su impacto [67].

Según lo expuesto, la consecuencia molecular refleja el impacto que tiene la presencia de una variación genética en cómo un gen determinado es traducido a proteínas. Debido a las diferencias en los efectos de cada una de las posibles consecuencias moleculares presentadas, a la hora de analizar las variaciones que resulten tener una calidad suficiente tras la aplicación del workflow en la siguiente sección, será un factor de interés que puede ayudar en la comprensión de cuál puede ser el impacto de la variación en el desarrollo de la enfermedad.

Otro de los aspectos importantes a considerar para las variaciones, y que además está incluido en un paso del nuevo workflow propuesto (**Figura 17**), es el tipo de método utilizado para obtener la variación. En la **Figura 29** se muestra cuántas variaciones se encuentran asociadas a cada tipo de método:

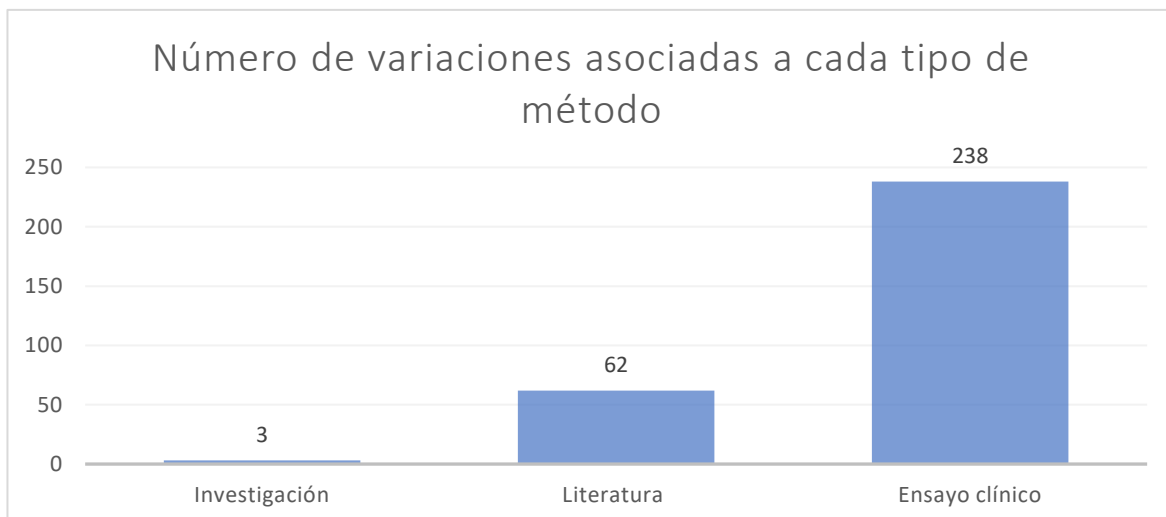


Figura 29. Número de variaciones asociadas a cada tipo de método

Sumando el número de variaciones de cada tipo se observa que suman un total de 303 variaciones, que es superior a las 276 que componen el resultado de la búsqueda en la base de datos. Esto se debe a que una variación puede haber sido interpretada utilizando diferentes métodos, de forma que en la **Figura 29** hay variaciones que aparecerán duplicadas.

Adicionalmente, otro aspecto que se tiene en cuenta en el nuevo workflow propuesto (**Figura 17**) y que por tanto es importante para entender los resultados de su aplicación al caso del EOAD, es el gen en el que se encuentran localizadas las variaciones obtenidas en la base de datos. En la **Figura 30** se muestra cómo se distribuyen las 276 variaciones obtenidas en función del gen en el que se localizan.

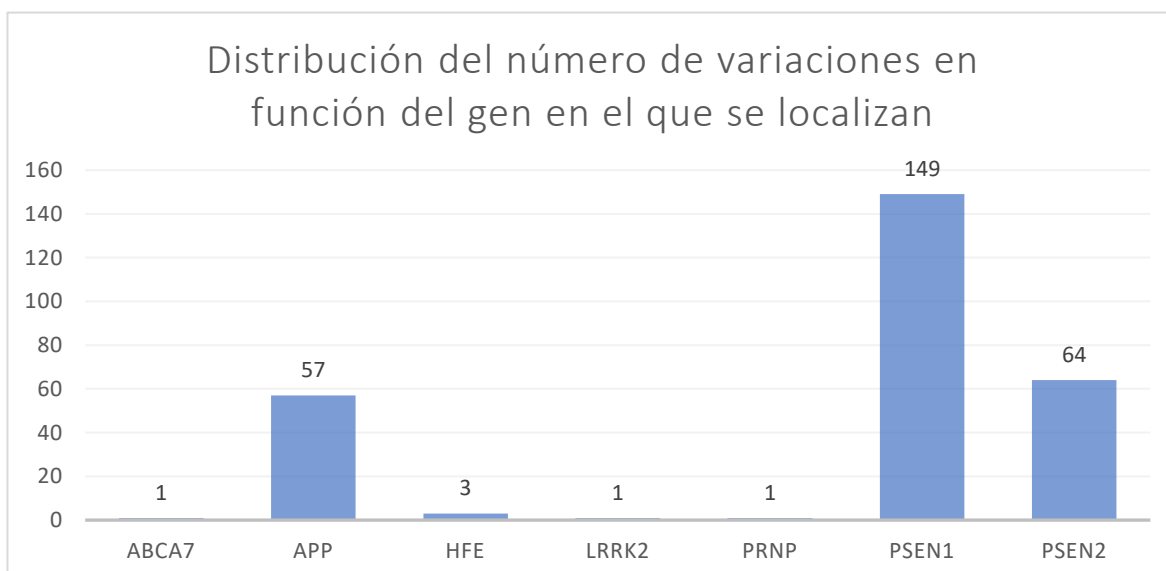


Figura 30. Distribución del número de variaciones en función del gen en el que se localizan

En la **Figura 30** se observa como la mayoría de las variaciones se concentran en tres genes: PSEN1, PSEN2, APP, algo con sentido pues tal y cómo se ha mencionado en la sección 4.1, son los genes que más típicamente se relacionan con el Alzheimer temprano. El gen HFE también se encuentra relacionado con el EOAD, concretamente con el mismo subtipo de Alzheimer que el gen APP.

Sin embargo, los genes ABCA7, LRRK2 y PRNP no se encuentran típicamente relacionados con el fenotipo de estudio. Para la búsqueda de esta relación, si es que existe, se ha determinado el realizar una revisión de la literatura disponible.

En el caso del gen PRNP, este codifica para la proteína PrP (*prion protein*) cuya función es desconocida [71]. A pesar de esto, existen diversos estudios que relacionan la patología de la enfermedad del Alzheimer con esta proteína [72]-[73].

Respecto al gen LRRK2, se han encontrado referencias a la variación encontrada en ClinVar (p.Arg1628Pro) en diferentes artículos, donde se la relaciona un aumento en el riesgo de desarrollo de Alzheimer [74]. Sin embargo, las variaciones en este gen no se encuentran comúnmente en pacientes con Alzheimer, y las evidencias indican que su posición (locus) ha sido relacionado previamente con el Alzheimer familiar, pero de desarrollo tipo tardío [75].

Finalmente, también se han encontrado relaciones entre el gen ABCA7 y el Alzheimer, pero la relación también ha sido encontrada para el Alzheimer de tipo tardío. La proteína codificada por este gen se ha relacionado mediante una correlación positiva con la presencia de placas amiloides en el cerebro, algo típicamente encontrado en cerebros de pacientes con Alzheimer [76]-[77]. En el caso de que la variación situada en este gen y la que está situada el gen LRRK2 pasaran a ser aceptadas por el nuevo workflow propuesto (**Figura 17**), debido a que en la literatura se asocia estos genes con el Alzheimer tardío, se requeriría una mayor revisión.

Finalmente, otro aspecto a destacar de los resultados obtenidos tras la búsqueda en la base de datos es los fenotipos que se suelen encontrar relacionados con las variaciones genéticas obtenidas. Como se ha comentado en varias ocasiones, las variaciones genéticas pueden aparecer asociadas a múltiples fenotipos. Los fenotipos más frecuentes encontrados son: *Early-onset familial Alzheimer disease*, *Alzheimer disease type 4*, *Alzheimer disease type 3*, *Alzheimer disease type 1*, *Dilated cardiomyopathy*, *Acne inversa familial*, *Pick's disease* y *Frontotemporal dementia*. Es de destacar que, aunque los 4 primeros tipos de fenotipos encontrados si se encuentran relacionados con la enfermedad según lo visto en la sección 4.1, los 4 últimos no guardan ningún tipo de relación con el Alzheimer.

Analizando los fenotipos que difieran con el Alzheimer en mayor profundidad, se observa que el fenotipo *Dilated cardiomyopathy* (DCM) aparece en todas las variaciones localizadas en los genes PSEN1 y PSEN2, pero no en localizadas en el gen APP. Esto se debe a que los genes PSEN1 y PSEN2 también se expresan en el corazón, siendo vitales para el desarrollo cardíaco, por lo que se han encontrado variaciones en estos genes que pueden ser causativas de la DCM [78].

También son comunes los fenotipos *Frontotemporal dementia* y *Pick's disease* pero en este caso asociados únicamente al gen PSEN1. La enfermedad de Pick es un subtipo de *Frontotemporal dementia* que es similar al Alzheimer, pero que es muchísimo menos común y sin una pérdida tan grande de memoria asociada [79].

Finalmente, también se ha encontrado en variaciones localizadas en el gen PSEN1 referencias al fenotipo *Acne inversa, familiar*. Se han encontrado estudios que relacionan esta enfermedad con los genes PSEN1 y APP debido al efecto que estos tienen sobre la gamma-secretasa [79], una proteína integral de membrana que también se relaciona con el Alzheimer a través de su papel en la formación de las placas seniles [80].

En este apartado se ha realizado un análisis general de los resultados obtenidos para la búsqueda de variaciones genéticas relacionadas con el EOAD en ClinVar. Sin embargo, no todas las 276 variaciones encontradas dispondrán de la calidad y fiabilidad suficiente para ser aplicadas en el ámbito clínico. Con el objetivo evaluar los resultados obtenidos, según el tercer objetivo planteado en este trabajo, en el siguiente apartado se va a realizar la aplicación del nuevo workflow propuesto (Figura 17) a la lista de 276 variaciones obtenidas de ClinVar.

7.2. Resultados de la Aplicación del Workflow Propuesto al Caso del EOAD

Una vez se han tratado los resultados de la búsqueda en la base de datos, es necesario aplicar el workflow diseñado (Figura 17) al listado de las 276 variaciones. En la Figura 31 se muestra una visión global de los resultados obtenidos con la aplicación, indicando el número de variaciones que pasan cada una de las etapas del workflow. Partiendo de esta perspectiva general de los resultados, se entrará en detalle en cada una de las etapas.

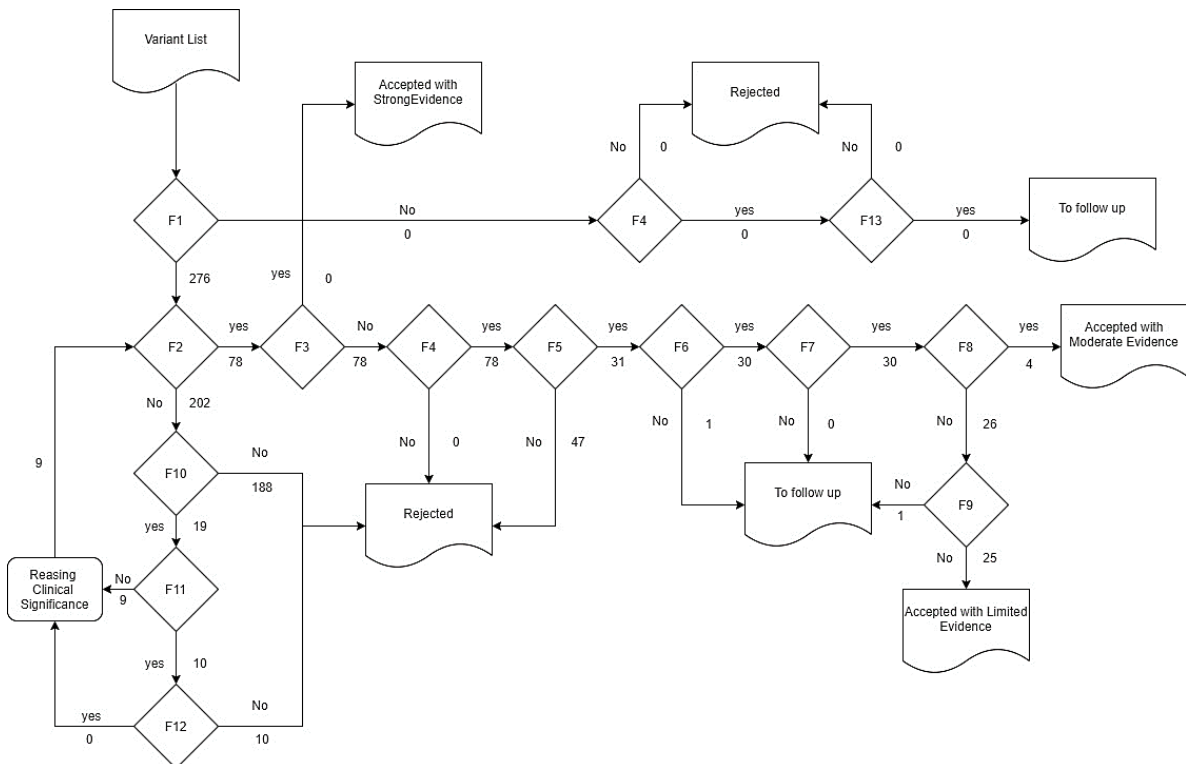


Figura 31. Resultados de la aplicación del workflow propuesto al EOAD

En primer lugar, el objetivo de los pasos F10 a F12 era asegurar que ninguna de las variaciones había sido incorrectamente clasificada por la base de datos como un conflicto de interpretación. En la base de datos un total de 19 variaciones habían sido clasificadas como un conflicto de interpretación, pero sin embargo con la aplicación del workflow se ha determinado que 9 de ellas habían sido incorrectamente clasificadas de esta forma. De las 9 variaciones mal clasificadas, 2 deberían haber sido clasificadas como un factor de riesgo, 4 como un significado incierto, 1 patogénica, 1 probablemente patogénica y 1 probablemente benigna. Las 2 variaciones clasificadas como factor de riesgo, la patogénica y la probablemente patogénica presentan un significado clínico que es de utilidad para el diagnóstico clínico, de forma que deberán ser consideradas en el paso F2 del workflow. El hecho de que 9 de las 19 variaciones estuvieran incorrectamente clasificadas como un conflicto de interpretación pone en valor la necesidad de incluir los pasos F10 a F12 del workflow.

Tras esto, el siguiente paso a aplicar sería la selección de las variaciones que tengan un significado clínico de utilidad (F2). Como se ha indicado previamente, estas variaciones son las que tienen como significado clínico patogénica, probablemente patogénica o factor riesgo. Teniendo en cuenta los resultados de la búsqueda en lo referente al significado clínico (**Figura 27**), junto con las 4 variaciones cuyo significado clínico ha sido reasignado a uno de los de interés, un total de 78 variaciones pasarían este filtro.

En el siguiente de los pasos (F3), se clasifican como aceptadas con evidencia fuerte aquellas variaciones que tienen como submitter guías de práctica clínica o paneles de expertos. Como se evidencia en los resultados de aplicación del workflow (**Figura 31**), ninguna de las variaciones alcanza la evidencia suficiente para ser clasificada como *Aceptada con evidencia fuerte*.

En la evaluación de la existencia de evidencias que relacionen el gen en el que se localiza la variación con la enfermedad (F4), se ha encontrado que todas las variaciones se localizan en genes cuya relación con la enfermedad ha sido establecida, como se ha explicado en la sección anterior

El siguiente paso es la selección de aquellas variaciones cuya información esté basada en un ensayo clínico. En la **Figura 32** se muestra el gran impacto que tiene la aplicación de este criterio sobre las 78 variaciones que han pasado el filtro anterior:

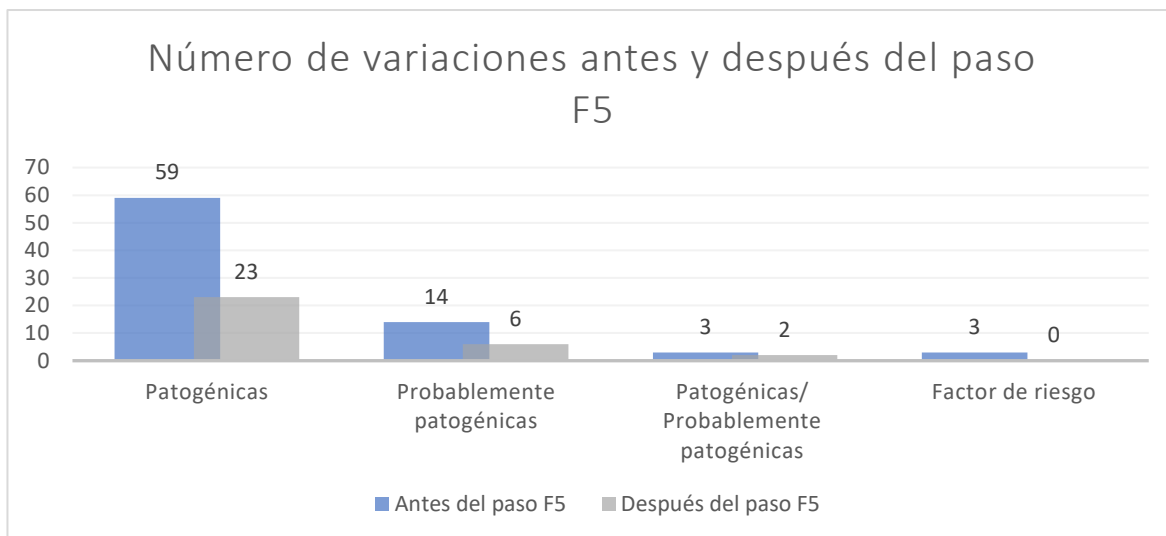


Figura 32. Número de variaciones antes y después del paso F5

Con lo reflejado en la **Figura 32**, tan solo 31 de las 78 variaciones que se tenían antes del paso F5 han conseguido pasar el filtro.

En el siguiente paso (F6) se buscaba comprobar si las variaciones seleccionadas en el apartado anterior tenían un *assertion criteria* proporcionado. Solo una variación con significado clínico probablemente patogénica no ha conseguido pasar este filtro, teniendo tras esto 23 variaciones patogénicas, 5 probablemente patogénicas y 2 patogénicas/ probablemente patogénicas. Respecto al paso F7, donde se evalúa la relevancia de los *submitters*, todas las variaciones del paso anterior consiguen pasar este filtro.

A continuación, con el paso F8 se va a evaluar si alguna de las variaciones cumple con los requerimientos necesarios para ser clasificada como *Aceptada con evidencia moderada*, mediante la determinación de si se han aplicado o no las guías ACMG/AMP como criterio de interpretación. Únicamente 4 de las 30 variaciones que entraban en este paso ha conseguido ser clasificada como *Aceptada con evidencia moderada*. Concretamente, se tratan de tres variaciones probablemente patogénicas y una patogénica.

En el caso de las 26 variaciones en las que el criterio utilizado no ha sido la aplicación de las guías ACMG/AMP, se evaluará si la última revisión ha sido realizada hace menos de tres años, buscando garantizar que la variación ha sido al menos revisada recientemente. Tan solo 1 de las 26 variaciones no cumple con este criterio, de forma que 25 variaciones serán clasificadas como *Aceptada con evidencia limitada*.

Finalmente, tras haber realizado la aplicación de la totalidad de los criterios en la **Tabla 3** se muestra un resumen de los resultados.

	N.º inicial de variaciones	Aceptadas con evidencia fuerte	Aceptadas con evidencia moderada	Aceptadas con evidencia limitada	Seguimiento	Rechazadas
Patogénicas	59	0	1	20	1	35
Probablemente patogénicas	14	0	3	3	1	7
Factor de riesgo	2	0	0	0	0	2
Patogénica/ Probablemente patogénica	3	0	0	2	0	1

Tabla 3. Número de variaciones obtenidas para cada clasificación posible para el EOAD

Adicionalmente, en la **Figura 33** se pueden ver como se distribuyen de forma global las variaciones en cada una de las posibles clasificaciones del workflow.

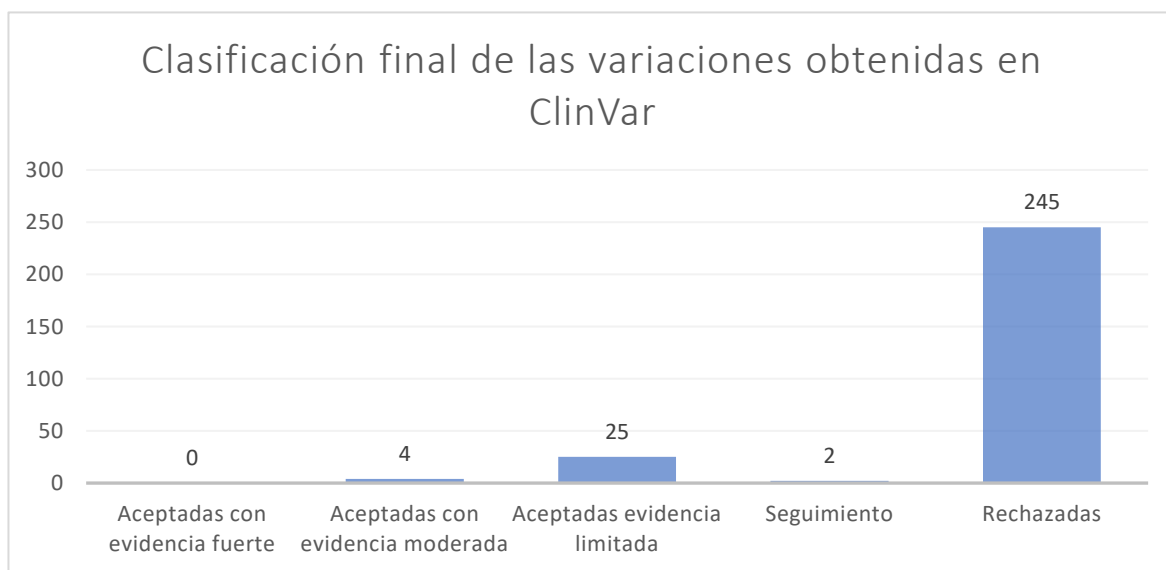


Figura 33. Clasificación final de las variaciones obtenidas en ClinVar

Buscando proporcionar más información acerca de las variaciones aceptadas, en el Anexo 1 se ha creado una tabla donde se encuentran todas las variaciones aceptadas, junto con información relevante acerca de las mismas.

Una vez han sido obtenidos los resultados tras la aplicación del workflow, en la siguiente sección se va a realizar el análisis y discusión de estos. Adicionalmente, se va a realizar un análisis más exhaustivo acerca de las 4 variaciones clasificadas como *Aceptada con evidencia moderada*, buscando comprobar que las variaciones de mayor evidencia realmente tienen un impacto en el desarrollo de la enfermedad, y por tanto buscando validar de forma preliminar los resultados del workflow.

7.3. Discusión

En la sección anterior se han expuesto los resultados de aplicación del workflow propuesto obteniendo un total de 29 variaciones que han sido clasificadas como aceptadas, algo que contrasta con los resultados obtenidos con la aplicación del workflow original, donde ninguna variación fue clasificada como aceptada. Gracias a la mejora planteada del workflow, se ha conseguido obtener variaciones para la enfermedad de estudio, por lo que se puede concluir que este nuevo workflow sí que permite tratar casos específicos como el EOAD.

Con todos los pasos incluidos en el workflow se ha buscado que todas las variaciones que se clasifiquen como aceptadas dispongan de una gran calidad, ya que se ha analizado tanto la información disponible considerada como relevante acerca de la variación (gen-tipo de método-interpretación-relevancia del submitter) como que la interpretación ha sido hecha en base a criterios de calidad. El exhaustivo análisis realizado por el workflow a cada una de las variaciones para poder ser clasificada como aceptada, sirve de garantía de que las variaciones obtenidas con el workflow diseñado siguen unos criterios de calidad y pueden ser suficientemente fiables para poder quizás utilizarlas en un ámbito clínico.

Analizando en mayor profundidad los resultados de las variaciones clasificadas como aceptadas, 26 de las 29 variaciones se localizan en el gen PSEN1, algo que concuerda con los conocimientos disponibles acerca del EOAD, donde se ha determinado que el gen PSEN1 es el mayor responsable del EOAD de carácter familiar [31]. Haciendo referencia a los fenotipos a los que se encuentran referidas estas 29 variaciones, el más común ha sido el *Alzheimer disease, type 3* algo que concuerda con los resultados obtenidos para los genes, pues este es el tipo de Alzheimer que se relaciona con el gen PSEN1. Además, también en 26 de las 29 variaciones se ha determinado que la consecuencia molecular es de tipo *missense*, constituyendo por tanto la mayoría de las variaciones aceptadas un cambio puntual en la secuencia de ADN que codifica para un aminoácido diferente. También han sido aceptadas las dos variaciones con la consecuencia molecular *splice site*, y una sin la consecuencia molecular proporcionada.

También resulta destacable que el *assertion criteria* más común en las variaciones aceptadas es *Invitae Variant Classification Sherlock (09022015)*, por lo que *Invitae*, un respetado laboratorio dedicado a ensayos clínicos es el submitter más común de las variaciones aceptadas.

Otro aspecto para destacar es que el hecho de que la mayoría de las variaciones estudiadas hayan sido revisadas recientemente (solo 1 variación es clasificada como *Seguimiento* por este motivo). En los casos de variaciones con más de una *submission*, se ha detectado que las *submissions* más antiguas estaban principalmente basadas en literatura, mientras que las *submissions* más nuevas

estaban principalmente basadas en ensayos clínicos. Esto puede significar que gracias a las nuevas tecnologías de NGS, que permiten secuenciar el genoma de una forma mucho más rápida y económica, los ensayos clínicos son menos difíciles de llevar a cabo que hace años, y por tanto se ha visto aumentado el número de submissions basadas en este tipo de método.

Una vez se han discutido las variaciones aceptadas en su generalidad, demostrando la coherencia de los resultados obtenidos, se van a analizar en mayor detalle las 4 variaciones que han sido aceptadas con evidencia moderada, buscando en la literatura referencias que respalden el efecto de estas variaciones para el desarrollo de la enfermedad. Las 4 variaciones presentan los siguientes identificadores: rs63749836, rs63751223, rs1799945, NM_000021.4(PSEN1): c.257T>G (p.Phe86Cys).

En el caso de la variación rs63749836 cuyo significado clínico es probablemente patogénica, se encuentra localizada en el gen PSEN1, y su consecuencia molecular es *missense*. La variación se encuentra asociada a varios fenotipos destacando principalmente el *Alzheimer disease type 3*, *Dilated cardiomyopathy* y *Acne inversa, familial*. Respecto a las evidencias disponibles en su relación con el EOAD, se han encontrado evidencias que relacionan la presencia de esta variación y un aumento de la muerte celular en el cerebro [81]. Adicionalmente, se han encontrado referencias a un incremento en el ratio A β 42/A β 40 [56], [82], lo que indica que la presencia de esta variación favorece la formación de las placas amiloides, estrechamente relacionadas con la enfermedad de Alzheimer como se ha visto en la sección 4.1.

Respecto a la variación rs63751223, presenta un significado clínico probablemente patogénico, se localiza en el gen PSEN1, y presenta una consecuencia molecular de tipo *missense*. En el estudio de la información disponible, se ha encontrado un estudio de imagen donde se realiza un análisis de retención PIB mediante PET. La retención positiva en PIB-PET es indicativa de formación de placas amiloides, pero cabe destacar que esta prueba, aunque presenta una gran sensibilidad, es muy poco específica [83]. Además, también se han encontrado referencias a un incremento en la concentración de la proteína dendrítica neurogranina⁴ en el líquido cefalorraquídeo debido a la presencia de la variación [84], estando este incremento asociado con la presencia de Alzheimer [85].

La variación rs1799945 presenta un significado clínico patogénico, y una consecuencia molecular tipo *missense*, además de encontrarse asociada con el fenotipo *Alzheimer disease type 1*. Se localiza en el gen HFE cuyas variaciones (incluido la aquí referida) se relacionan con un incremento de hierro en múltiples órganos, incluyendo el cerebro. Se ha demostrado que esta acumulación del hierro junto con el estrés oxidativo están relacionados con enfermedades neurodegenerativas como el Alzheimer, como se ha tratado en la sección 4.1, lo que confirma la relación variación-enfermedad [86].

Finalmente, en el caso de la variación NM_000021.4(PSEN1): c.257T>G (p.Phe86Cys), no se ha encontrado información adicional en la literatura de forma que solo se dispone de la información subida por el submitter a la base de datos. Se encuentra localizada en el gen PSEN1, presenta una consecuencia molecular de tipo *missense*, y aparece asociada al fenotipo *Alzheimer disease, type 3*. Respecto a su significado clínico, este es probablemente patogénica.

⁴ La neurogranina es una proteína abundante en el cerebro que interviene en procesos relacionados con la plasticidad neuronal, y que se relaciona con mecanismos de la memoria y el aprendizaje [90].

Con todo esto, excepto en una variación, se han podido encontrar referencias que respaldan el efecto patogénico/probablemente patogénico que tienen las variaciones aceptadas con evidencia moderada sobre el desarrollo de la enfermedad. Teniendo en cuenta las evidencias encontradas, se podría valorar el uso de estos resultados en la práctica clínica, especialmente de las variaciones que se clasificaran como aceptadas con evidencia fuerte o con evidencia moderada. Para ello, los resultados serán presentados a un grupo de expertos en diagnóstico genético y tratamiento del Alzheimer para su validación formal, siendo este uno de los puntos del trabajo futuro.

A continuación, en el siguiente apartado, van a presentarse las conclusiones.

8. CONCLUSIONES

Este trabajo se ha llevado a cabo siguiendo tres objetivos, según lo mostrado en la sección de OBJETIVOS Y STAKEHOLDERS, para cada uno de los cuales se habían establecido unas preguntas de investigación que han guiado la consecución de estos objetivos.

El primero de los objetivos consistía en realizar un análisis de la etapa de identificación del método SILE para determinar si permitía tratar casos específicos como el EOAD. En la sección 3.2, se analizaron los criterios existentes en el método SILE para la identificación de variaciones relevantes, respondiendo así a la primera pregunta de investigación propuesta. Una vez determinados los criterios existentes, en la sección de Aplicación del Método SILE al EOAD y en la de Análisis de la Etapa de Identificación del Método SILE se ha determinado que no todos los criterios eran evaluables en casos como el EOAD, específicamente el criterio de evaluación de la relevancia estadística de los artículos relacionados, entre otros, y además se han concretado los problemas existentes en su aplicación identificando problemáticas adicionales. Con esto, se ha dado respuesta a la segunda y la tercera de las preguntas de investigación planteadas.

Derivado del primero de los objetivos se encontraba el segundo, donde se busca plantear una mejora de la etapa de identificación del método SILE que tenga en cuenta todos los problemas identificados. Buscando responder a la cuarta pregunta de investigación, en la sección de Nuevo Workflow Propuesto para la Etapa de Identificación del método SILE se ha propuesto un nuevo workflow que contempla todas las mejoras necesarias para resolver los problemas detectados en la etapa de identificación.

Finalmente, en el tercero de los objetivos se ha buscado realizar una evaluación del nuevo workflow propuesto para la etapa de identificación mediante su aplicación al caso del EOAD. Las mejoras propuestas en la sección Nuevo Workflow Propuesto para la Etapa de Identificación del método SILE consiguen resolver todos los problemas detectados, según lo detallado en el apartado, consiguiendo resolver con esto a la quinta pregunta de investigación. Con la aplicación del nuevo workflow en la sección de Resultados de la Aplicación del Workflow Propuesto al Caso del EOAD se han encontrado variaciones genéticas relevantes, concluyendo con esto que con este nuevo workflow si se pueden tratar las variaciones genéticas en casos particulares como la enfermedad de estudio, respondiendo con esto a la sexta pregunta de investigación planteada. Adicionalmente, en la sección de Discusión se ha evaluado la relación entre las variaciones clasificadas como aceptadas con evidencia moderada y el EOAD demostrando que los resultados obtenidos son consistentes con los conocimientos disponibles de la enfermedad, obteniendo por tanto respuesta a la séptima pregunta de investigación.

Adicionalmente, con todos los pasos implementados en el workflow y que han sido justificados en la sección Nuevo Workflow Propuesto para la Etapa de Identificación del método SILE, se ha buscado tanto que la información sea de calidad, como que la interpretación de la variación sea confiable, y por tanto se puede determinar que los resultados disponen de calidad y fiabilidad, lo que responde a la octava pregunta de investigación.

Con todo esto, la última pregunta de investigación que se plantea es si estos resultados podrían ser o no aplicados a la práctica clínica. Aunque los resultados estén basados en la fiabilidad y calidad de la información disponible, deben ser todavía evaluados por un grupo de expertos para determinar su validez.

Finalmente, se puede concluir por tanto que todos los objetivos propuestos al principio del trabajo han sido alcanzados, y que se ha conseguido determinar variaciones genéticas relevantes para el caso de la enfermedad de Alzheimer temprano.

Este trabajo es solo una muestra de la necesidad de plantear métodos que permitan realizar una exploración inteligente de los datos disponibles en bases de datos y repositorios genéticos, ya que puede suponer un antes y un después en el tratamiento y diagnóstico de algunas enfermedades. Cómo se va a mostrar en la siguiente sección de TRABAJOS FUTUROS el campo en el que se ha desarrollado este trabajo todavía requiere de una gran investigación y desarrollo, ya que resulta fundamental para poder alcanzar un correcto desempeño del concepto de medicina de precisión.

9. TRABAJOS FUTUROS

El último punto para tratar en este trabajo consiste en determinar, una vez diseñado el nuevo workflow y tras haber hecho su aplicación al EOAD, cuáles son las posibles líneas de trabajo que pueden tratarse en un futuro. Por tanto, a lo largo de esta sección, se plantearán las diferentes líneas futuras que derivan del trabajo aquí presentado.

En primer lugar, aunque se ha realizado un análisis preliminar de los resultados obtenidos, es necesario que los resultados proporcionados por el workflow sean validados de manera formal para poder realmente afirmar que el nuevo workflow propuesto permite obtener variaciones genéticas relevantes. Como se ha mencionado en la sección anterior, para esta validación los resultados deben ser presentados a un grupo de expertos en diagnóstico genético y tratamiento del Alzheimer para su validación formal.

Dentro de los posibles trabajos, el que tiene una mayor continuidad con el llevado a cabo es la implementación de las etapas de carga y explotación del método SILE, ya que no han sido tratadas en este trabajo. El llevar a cabo estas etapas completaría todo el proceso, permitiendo finalmente el utilizar la información disponible acerca de las variaciones clasificadas como aceptadas en el ámbito clínico.

Por otra parte, tal y como se ha mostrado en las últimas secciones, con el workflow propuesto se han conseguido solucionar los problemas a la hora de analizar variaciones en casos específicos como el EOAD. En este trabajo se ha utilizado ClinVar, de forma que el workflow presentado permite tratar los casos específicos en variaciones genéticas que hayan sido extraídas de esta base de datos. Debido a la existencia de múltiples repositorios y bases de datos de las que extraer información, también es necesario realizar el análisis de estos en busca de variaciones genéticas relevantes. Esto plantea dos posibilidades, o bien se busca desarrollar un workflow genérico que tenga en cuenta todas las posibles bases de datos, o bien se desarrollan workflow(s) específicos para cada base de datos. La primera alternativa tiene como ventaja el uso de un único workflow, lo que facilitaría el proceso de aplicación de este. Sin embargo, al ser más genérico, podría ser más difícil de automatizar para cada una de las bases de datos. En cambio, en la segunda alternativa al poder ajustar los pasos a la información que proporciona cada base de datos, es más susceptible de ser automatizado el proceso. Como inconveniente, el proceso de diseño se dificultaría e incrementaría en el tiempo al haber una gran multitud de bases de datos disponibles.

Derivado de esto, otra posible línea de trabajo sería la automatización del workflow diseñado. Como se ha mencionado en la sección 5 ClinVar se actualiza de forma mensual, lo que significa que para tener unos resultados actualizados acerca de la enfermedad sería necesario aplicar el workflow a la lista de variaciones que nos devuelva la base de datos mensualmente. Debido al considerable número de variaciones, que es mucho mayor en el caso de otras enfermedades, aplicar el workflow mensualmente supone un gran esfuerzo que justifica la necesidad de automatizar el proceso. Esta automatización permitiría acelerar la identificación de variaciones relevantes y ajustarse a la evolución temporal de la información disponible.

En relación con esta automatización, otro posible trabajo futuro sería desarrollar una herramienta que permitiera realizar un análisis a posteriori de las variaciones aceptadas. En los casos con muchas

variaciones, puede ser interesante el poder agrupar los resultados por gen, por tipo de método, y poder hacer estadísticas acerca de los resultados obtenidos a nivel global. Esto puede ayudar a avanzar en el conocimiento de la enfermedad, ya que se podrían detectar patrones en la localización de variaciones, en la consecuencia molecular, detectar nuevas relaciones entre fenotipos, etc.

Otro aspecto a considerar es el uso de las guías ACMG/AMP que son aplicables a enfermedades con patrones de herencia mendeliana, por lo que no son evaluables en enfermedades de carácter complejo. En este tipo de enfermedades, donde para el desarrollo de la enfermedad influyen otros factores a parte de la genética, se deberán plantear en futuros trabajos alternativas que permitan validar la interpretación realizada de la variación.

Respecto al workflow diseñado, también existen factores a considerar en trabajos futuros. Por una parte, en la sección 6.2 se definieron estudios familiares, pero no criterios para poder utilizarlos en la evaluación de la relevancia de los estudios disponibles. Por tanto, para completar la inclusión de estos estudios dentro del método SILE y del workflow propuesto, deberían analizarse estos estudios y plantear criterios para los mismos. Además de esto, en la sección de Resultados de la Búsqueda de Variaciones Genéticas Relacionadas con el EOAD en ClinVar se ha expuesto la importancia que tiene el análisis de la consecuencia molecular, ya que determina el impacto que tendrá la variación en la pérdida total o parcial de la función de un gen. Debido a su importancia, para versiones posteriores del workflow diseñado se debe contemplar su inclusión como un nuevo criterio.

Finalmente, otra posibilidad de trabajo futuro sería el aplicar el workflow en otras enfermedades para analizar si los resultados obtenidos son también consistentes con lo que se conozca acerca de la enfermedad. Esto permitiría además validar el hecho de que el workflow debe permitir analizar las variaciones genéticas relevantes en cualquier tipo de enfermedad. Debido a que el workflow original ha sido aplicado a otras variaciones a parte del EOAD, también podría resultar de interés comparar los resultados obtenidos por ambos workflow, con el objetivo de determinar debido a que se producen las posibles diferencias. En suma, estos trabajos futuros estarían orientados a testear el workflow con el objetivo de validarlo y de determinar factores y aspectos susceptibles de mejora.

10. BIBLIOGRAFÍA

- [1] “What is precision medicine? - Genetics Home Reference - NIH.” [Online]. Available: <https://ghr.nlm.nih.gov/primer/precisionmedicine/definition>. [Accessed: 30-Mar-2020].
- [2] “Definition of variant - NCI Dictionary of Genetics Terms - National Cancer Institute.” [Online]. Available: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/genetic-variant>. [Accessed: 30-Mar-2020].
- [3] “Variabilidad genética | NHGRI.” [Online]. Available: <https://www.genome.gov/es/genetics-glossary/Variabilidad-genetica>. [Accessed: 31-Mar-2020].
- [4] “What are some potential benefits of precision medicine and the Precision Medicine Initiative? - Genetics Home Reference - NIH.” [Online]. Available: <https://ghr.nlm.nih.gov/primer/precisionmedicine/potentialbenefits>. [Accessed: 30-Mar-2020].
- [5] “Breve Historia Del Proyecto Del Genoma Humano | NHGRI.” [Online]. Available: <https://www.genome.gov/breve-historia-del-proyecto-del-genoma-humano>. [Accessed: 30-Mar-2020].
- [6] “ExAC: el mayor catálogo de variación genética humana aplicado a la investigación clínica - Genotipia.” [Online]. Available: https://genotipia.com/genetica_medica_news/exac-catalogo-variacion-genetica/. [Accessed: 30-Mar-2020].
- [7] S. J. Aronson and H. L. Rehm, “Building the foundation for genomics in precision medicine,” *Nature*, vol. 526, no. 7573. Nature Publishing Group, pp. 336–342, 14-Oct-2015, doi: 10.1038/nature15816.
- [8] W. Raghupathi and V. Raghupathi, “Big data analytics in healthcare: promise and potential,” *Heal. Inf. Sci. Syst.*, vol. 2, no. 1, Dec. 2014, doi: 10.1186/2047-2501-2-3.
- [9] “ClinVar.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/clinvar/>. [Accessed: 30-Apr-2020].
- [10] “Mutations | ALZFORUM.” [Online]. Available: <https://www.alzforum.org/mutations>. [Accessed: 09-May-2020].
- [11] “GWAS Catalog.” [Online]. Available: <https://www.ebi.ac.uk/gwas/diagram#>. [Accessed: 09-May-2020].
- [12] “SNPedia.” [Online]. Available: <https://snpedia.com/>. [Accessed: 30-Apr-2020].

- [13] “Big Data o Smart Data: El poder de la información - Cegos.” [Online]. Available: <https://www.cegosonlineuniversity.com/big-data-o-smart-data-el-poder-de-la-informacion/>. [Accessed: 06-May-2020].
- [14] “¿Qué es el Big Data? Las 5 V’s del Big Data - LEONUP.” [Online]. Available: <http://www.leonup.com/es/2018/05/04/big-data-las-5-vs-del-big-data/>. [Accessed: 31-Mar-2020].
- [15] I. J. Pascual Fernández, “Análisis de Datos Genómicos para la identificación de variantes genómicas relevantes en la enfermedad de Alzheimer,” Oct. 2019.
- [16] N. del C. Gonza Ajila, “Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca: aplicaciones a la medicina de precisión,” Dec. 2019.
- [17] P. Cuenca Gil, “Ciencia de datos para la medicina de precisión: Aplicaciones al diagnóstico genómico del Neuroblastoma,” Oct. 2019.
- [18] U. Politècnica De València, “Trabajo Fin de Grado BIOTECNOLOGÍA PARA LA MEDICINA DE PRECISIÓN: IDENTIFICACIÓN DE VARIACIONES GENÓMICAS PARA EL DIAGNÓSTICO DE CÁNCER DE MAMA Presentado por: Alba Escalera Balsera,” 2018.
- [19] Ó. Pastor López, A. León, and S. Kaur, “Trabajo Fin de Máster Un proceso para la identificación sistemática de variaciones genómicas: Aplicaciones a la medicina de precisión Septiembre 2018.”
- [20] V. Alexandra and S. Cabrera, “Exploración de Bases de Datos Genómicas Dirigida por Modelos Conceptuales Septiembre 2018.”
- [21] CLARA SOLER PELLICER, “DISEÑO DE UN SISTEMA DE INFORMACIÓN GENÓMICA PARA EL DIAGNÓSTICO DEL NEUROBLASTOMA,” 2016.
- [22] “Stakeholders: ejemplos para entender el concepto.” [Online]. Available: <https://obsbusiness.school/es/blog-investigacion/marketing-y-comunicacion/stakeholders-ejemplos-para-entender-el-concepto>. [Accessed: 02-Apr-2020].
- [23] A. León Palacio and Ó. Pastor López, “Smart Data for Genomic Information Systems: the SILE Method,” *Complex Syst. Informatics Model. Q.*, no. 17, pp. 1–23, 2018, doi: 10.7250/csimq.2018-17.01.
- [24] J. F. Reyes Román, Ó. Pastor, J. C. Casamayor, and F. Valverde, “Applying conceptual modeling to better understand the human genome,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in*

- Bioinformatics*), 2016, vol. 9974 LNCS, pp. 404–412, doi: 10.1007/978-3-319-46397-1_31.
- [25] J. F. Reyes Román, A. García, U. Rueda, and Ó. Pastor, “GenesLove.Me 2.0: Improving the Prioritization of Genetic Variations,” in *Communications in Computer and Information Science*, 2019, vol. 1023, pp. 314–333, doi: 10.1007/978-3-030-22559-9_14.
- [26] “Enfermedad de Alzheimer: MedlinePlus en español.” [Online]. Available: <https://medlineplus.gov/spanish/alzheimersdisease.html>. [Accessed: 04-Apr-2020].
- [27] “Demencia: MedlinePlus en español.” [Online]. Available: <https://medlineplus.gov/spanish/dementia.html>. [Accessed: 04-Apr-2020].
- [28] “Enfermedad de Alzheimer: síntomas y causas.” [Online]. Available: <https://www.mayoclinic.org/es-es/diseases-conditions/alzheimers-disease/symptoms-causes/syc-20350447>. [Accessed: 16-May-2020].
- [29] Alzheimer’s Disease International (ADI), “Informe mundial sobre el Alzheimer 2018,” 2018.
- [30] “Defunciones por causas (lista reducida) por sexo y grupos de edad(7947).” [Online]. Available: <https://www.ine.es/jaxiT3/Datos.htm?t=7947#!tabs-tabla>. [Accessed: 04-Apr-2020].
- [31] M. H. Dai, H. Zheng, L. D. Zeng, and Y. Zhang, “The genes associated with early-onset Alzheimer’s disease,” *Oncotarget*, vol. 9, no. 19, pp. 15132–15143, 2018, doi: 10.18632/oncotarget.23738.
- [32] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers, “The genetic landscape of Alzheimer disease: Clinical implications and perspectives,” *Genet. Med.*, vol. 18, no. 5, pp. 421–430, 2016, doi: 10.1038/gim.2015.117.
- [33] D. J. Selkoe, “Alzheimer ’ s Disease Genetics of Alzheimer ’ s Disease *,” vol. 2013, pp. 18295–18298, 1996, doi: 10.1074/jbc.271.31.18295.
- [34] A. Joshi, J. M. Ringman, A. S. Lee, K. O. Juarez, and M. F. Mendez, “Comparison of clinical characteristics between familial and non-familial early onset Alzheimer’s disease,” *Journal of Neurology*, vol. 259, no. 10, pp. 2182–2188, Oct-2012, doi: 10.1007/s00415-012-6481-y.
- [35] R. Cacace, K. Sleegers, and C. Van Broeckhoven, “Molecular genetics of early-onset Alzheimer’s disease revisited,” *Alzheimer’s Dement.*, vol. 12, no. 6, pp. 733–748, 2016, doi: 10.1016/j.jalz.2016.01.012.
- [36] C. Van Cauwenberghe, C. Van Broeckhoven, and K. Sleegers, “The genetic landscape of Alzheimer disease: Clinical implications and perspectives,” *Genetics in Medicine*, vol. 18,

- no. 5. Nature Publishing Group, pp. 421–430, 01-May-2016, doi: 10.1038/gim.2015.117.
- [37] “Alzheimer disease, type 1 (Concept Id: C2931257) - MedGen - NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/medgen/444014>. [Accessed: 16-May-2020].
- [38] “ALZHEIMER DISEASE 3; AD.” [Online]. Available: [https://omim.org/entry/607822?search=ALZHEIMER DISEASE 1&highlight=1 alzheimer disease](https://omim.org/entry/607822?search=ALZHEIMER+DISEASE+1&highlight=1+alzheimer+disease).
- [39] “Alzheimer disease, type 4 (Concept Id: C1847200) - MedGen - NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/medgen/376072>. [Accessed: 16-May-2020].
- [40] “La enfermedad de Alzheimer: bases moleculares y aproximaciones terapéuticas • Asociación Española de Científicos.” [Online]. Available: <https://www.aecientificos.es/la-enfermedad-de-alzheimer-bases-moleculares-y-aproximaciones-terapeuticas/>. [Accessed: 27-May-2020].
- [41] J. Rodrigo *et al.*, “Características neuropatológicas y moleculares de la enfermedad de Alzheimer,” *Rev. Esp. Geriatr. Gerontol.*, vol. 42, no. 2, pp. 103–110, 2007, doi: 10.1016/S0211-139X(07)73533-3.
- [42] “Enfermedad de Alzheimer: parte 2.” [Online]. Available: <https://www.hipocampo.org/alzheimer-2.asp>. [Accessed: 22-May-2020].
- [43] L. Gu and Z. Guo, “Alzheimer’s A β 42 and A β 40 peptides form interlaced amyloid fibrils,” *J. Neurochem.*, vol. 126, no. 3, pp. 305–311, Aug. 2013, doi: 10.1111/jnc.12202.
- [44] “Review status in ClinVar.” [Online]. Available: https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/#ac_def. [Accessed: 21-Dec-2019].
- [45] J. T. den Dunnen *et al.*, “HGVS Recommendations for the Description of Sequence Variants: 2016 Update,” *Hum. Mutat.*, vol. 37, no. 6, pp. 564–569, Jun. 2016, doi: 10.1002/humu.22981.
- [46] “How ClinVar validates submissions.” [Online]. Available: https://www.ncbi.nlm.nih.gov/clinvar/docs/validation/#var_condition. [Accessed: 21-Dec-2019].
- [47] “Sources of data in ClinVar.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/clinvar/docs/datasources/>. [Accessed: 21-Dec-2019].
- [48] S. Richards *et al.*, “Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics
-

- and the Association for Molecular Pathology,” *Genet. Med.*, vol. 17, no. 5, pp. 405–424, May 2015, doi: 10.1038/gim.2015.30.
- [49] “Representation of clinical significance in ClinVar and other variation resources at NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/>. [Accessed: 23-Dec-2019].
- [50] H. M. Lanoiselée *et al.*, “APP, PSEN1, and PSEN2 mutations in early-onset Alzheimer disease: A genetic screening study of familial and sporadic cases,” *PLoS Med.*, vol. 14, no. 3, Mar. 2017, doi: 10.1371/journal.pmed.1002270.
- [51] M. E. Conidi *et al.*, “Homozygous carriers of APP A713T mutation in an autosomal dominant Alzheimer disease family,” *Neurology*, vol. 84, no. 22, pp. 2266–2273, Jun. 2015, doi: 10.1212/WNL.0000000000001648.
- [52] E. S. Athan *et al.*, “A founder mutation in presenilin 1 causing early-onset Alzheimer disease in unrelated Caribbean Hispanic families,” *J. Am. Med. Assoc.*, vol. 286, no. 18, pp. 2257–2263, Nov. 2001, doi: 10.1001/jama.286.18.2257.
- [53] R. Queralt *et al.*, “A novel mutation (V89I) in the presenilin 1 gene in a family with early onset Alzheimer’s disease and marked behavioural disturbances,” *J. Neurol. Neurosurg. Psychiatry*, vol. 72, no. 2, pp. 266–269, 2002, doi: 10.1136/jnnp.72.2.266.
- [54] A. Lleó *et al.*, “Frequency of mutations in the presenilin and amyloid precursor protein genes in early-onset Alzheimer disease in Spain,” *Arch. Neurol.*, vol. 59, no. 11, pp. 1759–1763, Nov. 2002, doi: 10.1001/archneur.59.11.1759.
- [55] G. Raux *et al.*, “Molecular diagnosis of autosomal dominant early onset Alzheimer’s disease: An update,” *Journal of Medical Genetics*, vol. 42, no. 10. BMJ Publishing Group Ltd, pp. 793–795, 01-Oct-2005, doi: 10.1136/jmg.2005.033456.
- [56] D. Campion *et al.*, “Early-onset autosomal dominant Alzheimer disease: Prevalence, genetic heterogeneity, and mutation spectrum,” *Am. J. Hum. Genet.*, vol. 65, no. 3, pp. 664–670, 1999, doi: 10.1086/302553.
- [57] P. Poorkaj *et al.*, “Missense mutations in the chromosome 14 familial Alzheimer’s disease presenilin 1 gene,” *Hum. Mutat.*, vol. 11, no. 3, pp. 216–221, Jan. 1998, doi: 10.1002/(SICI)1098-1004(1998)11:3<216::AID-HUMU6>3.0.CO;2-F.
- [58] G. Devi *et al.*, “Novel presenilin 1 mutations associated with early onset of dementia in a family with both early-onset and late-onset Alzheimer disease,” *Arch. Neurol.*, vol. 57, no. 10, pp. 1454–1457, Oct. 2000, doi: 10.1001/archneur.57.10.1454.
-

- [59] S. Sutovsky *et al.*, “Neuropathology and biochemistry of early onset familial Alzheimer’s disease caused by presenilin-1 missense mutation Thr116Asn,” *J. Neural Transm.*, vol. 125, no. 6, pp. 965–976, Jun. 2018, doi: 10.1007/s00702-018-1850-z.
- [60] M. Gallo *et al.*, “The novel PSEN1 M84V mutation associated to frontal dysexecutive syndrome, spastic paraparesis, and cerebellar atrophy in a dominant Alzheimer’s disease family,” *Neurobiol. Aging*, vol. 56, pp. 213.e7-213.e12, Aug. 2017, doi: 10.1016/j.neurobiolaging.2017.04.017.
- [61] “Pedigree - MeSH Browser.” [Online]. Available: <https://meshb.nlm.nih.gov/record/ui?ui=D010375>. [Accessed: 03-Mar-2020].
- [62] “Twin Study - MeSH Browser.” [Online]. Available: <https://meshb.nlm.nih.gov/record/ui?ui=D018486>. [Accessed: 03-Mar-2020].
- [63] S. Richards *et al.*, “Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology,” *Genet. Med.*, vol. 17, no. 5, pp. 405–424, 2015, doi: 10.1038/gim.2015.30.
- [64] “Clinical Whole-Genome Sequencing Services.” [Online]. Available: https://www.illumina.com/clinical/illumina_clinical_laboratory.html. [Accessed: 26-Dec-2019].
- [65] “Frameshift Mutation | Talking Glossary of Genetic Terms | NHGRI.” [Online]. Available: <https://www.genome.gov/genetics-glossary/Frameshift-Mutation>. [Accessed: 18-May-2020].
- [66] “Definition of splice-site mutation - NCI Dictionary of Genetics Terms - National Cancer Institute.” [Online]. Available: <https://www.cancer.gov/publications/dictionaries/genetics-dictionary/def/splice-site-mutation>. [Accessed: 29-Dec-2019].
- [67] D. Bhartiya and V. Scaria, “Genomic variations in non-coding RNAs: Structure, function and regulation,” Academic Press Inc., Mar. 2016.
- [68] S. Chatterjee and J. K. Pal, “Role of 5'- and 3'-untranslated regions of mRNAs in human diseases,” *Biol. Cell*, vol. 101, no. 5, pp. 251–262, May 2009, doi: 10.1042/bc20080104.
- [69] C. Delay, F. Calon, P. Mathews, and S. S. Hébert, “Alzheimer-specific variants in the 3’UTR of Amyloid precursor protein affect microRNA function,” *Mol. Neurodegener.*, vol. 6, no. 1, p. 70, 2011, doi: 10.1186/1750-1326-6-70.
- [70] “What kinds of gene mutations are possible? - Genetics Home Reference - NIH.” [Online].
-

- Available: <https://ghr.nlm.nih.gov/primer/mutationsanddisorders/possiblemutations>. [Accessed: 19-May-2020].
- [71] “PRNP gene - Genetics Home Reference - NIH.” [Online]. Available: <https://ghr.nlm.nih.gov/gene/PRNP>. [Accessed: 23-Dec-2019].
- [72] Zhou, “Alzheimer’s Disease and Prion Protein,” *Intractable Rare Dis. Res.*, 2013, doi: 10.5582/irdr.2013.v2.2.35.
- [73] K. A. B. Kellett and N. M. Hooper, “Prion protein and Alzheimer disease,” in *Prion*, 2009, vol. 3, no. 4, doi: 10.4161/pri.3.4.9980.
- [74] Y. Zhao, P. Ho, Y. Yih, C. Chen, W. L. Lee, and E. K. Tan, “LRRK2 variant associated with Alzheimer’s disease,” *Neurobiol. Aging*, vol. 32, no. 11, pp. 1990–1993, Nov. 2011, doi: 10.1016/j.neurobiolaging.2009.11.019.
- [75] M. Toft *et al.*, “LRRK2 mutations are not common in Alzheimer’s disease,” *Mech. Ageing Dev.*, vol. 126, no. 11, pp. 1201–1205, Nov. 2005, doi: 10.1016/j.mad.2005.06.010.
- [76] S. Steinberg *et al.*, “Loss-of-function variants in ABCA7 confer risk of Alzheimer’s disease,” *Nat. Genet.*, vol. 47, no. 5, pp. 445–447, May 2015, doi: 10.1038/ng.3246.
- [77] Q. F. Zhao, J. T. Yu, M. S. Tan, and L. Tan, “ABCA7 in Alzheimer’s Disease,” *Molecular Neurobiology*, vol. 51, no. 3. Humana Press Inc., pp. 1008–1016, 01-Jun-2015, doi: 10.1007/s12035-014-8759-9.
- [78] D. Li *et al.*, “Mutations of presenilin genes in dilated cardiomyopathy and heart failure,” *Am. J. Hum. Genet.*, vol. 79, no. 6, pp. 1030–1039, 2006, doi: 10.1086/509900.
- [79] “Pick’s Disease - Common Symptoms and Causes.” [Online]. Available: <https://www.webmd.com/alzheimers/guide/picks-disease#1>. [Accessed: 29-Dec-2019].
- [80] X. Zhang, Y. Li, H. Xu, and Y. W. Zhang, “The γ -secretase complex: From structure to function,” *Frontiers in Cellular Neuroscience*, vol. 8, no. DEC. Frontiers Media S.A., 11-Dec-2014, doi: 10.3389/fncel.2014.00427.
- [81] T. D. Bird, *Alzheimer Disease Overview*. University of Washington, Seattle, 2018.
- [82] L. Sun, R. Zhou, G. Yang, and Y. Shi, “Analysis of 138 pathogenic mutations in presenilin-1 on the in vitro production of A β 42 and A β 40 peptides by γ -secretase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 114, no. 4, pp. E476–E485, Jan. 2017, doi: 10.1073/pnas.1618657114.
- [83] W. E. Klunk *et al.*, “Amyloid deposition begins in the striatum of presenilin-1 mutation carriers from two unrelated pedigrees,” *J. Neurosci.*, vol. 27, no. 23, pp. 6174–6184, Jun.
-

- 2007, doi: 10.1523/JNEUROSCI.0730-07.2007.
- [84] H. Wellington *et al.*, “Increased CSF neurogranin concentration is specific to Alzheimer disease,” *Neurology*, vol. 86, no. 9, pp. 829–835, Mar. 2016, doi: 10.1212/WNL.0000000000002423.
- [85] E. Portelius *et al.*, “Cerebrospinal fluid neurogranin concentration in neurodegeneration: relation to clinical phenotypes and neuropathology,” *Acta Neuropathol.*, vol. 136, no. 3, pp. 363–376, Sep. 2018, doi: 10.1007/s00401-018-1851-x.
- [86] F. Ali-Rahmani, C. L. Schengrund, and J. R. Connor, “HFE gene variants, iron, and lipids: A novel connection in Alzheimer’s disease,” *Frontiers in Pharmacology*, vol. 5 JUL. Frontiers Research Foundation, 2014, doi: 10.3389/fphar.2014.00165.
- [87] “Home - MeSH - NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/mesh>. [Accessed: 18-Feb-2020].
- [88] “Home - PubMed - NCBI.” [Online]. Available: <https://www.ncbi.nlm.nih.gov/pubmed>. [Accessed: 18-Feb-2020].
- [89] “Fundación Instituto Roche - Glosario de genética - penetrancia.” [Online]. Available: <https://www.institutoroche.es/recursos/glosario/penetrancia>. [Accessed: 24-May-2020].
- [90] A. Garrido-García *et al.*, “Neurogranin Expression Is Regulated by Synaptic Activity and Promotes Synaptogenesis in Cultured Hippocampal Neurons,” *Mol. Neurobiol.*, vol. 56, no. 11, pp. 7321–7337, Nov. 2019, doi: 10.1007/s12035-019-1593-3.

PRESUPUESTO

Diseño y aplicación de un método basado en SILE para la identificación de variaciones genéticas relevantes asociadas a la enfermedad de Alzheimer temprano

Documento II

1. NECESIDAD DEL PRESUPUESTO

El objetivo del presupuesto es valorar el coste económico del trabajo realizado en este Trabajo de Fin de Grado. Para el desarrollo de este presupuesto se van a tener en cuenta los siguientes costes: de personal, de software y de hardware.

Para el cálculo de la amortización tanto del software como del hardware se va a utilizar la siguiente fórmula:

$$\text{Coste imputable (sin IVA)} = t * \frac{C}{T}$$

Donde:

- **t.** Tiempo de uso del equipo medido en meses.
- **C.** Coste del equipo utilizado o de la licencia de software empleada.
- **T.** Tiempo de amortización medido en meses.

En la siguiente sección se va a presentar el presupuesto desglosado, donde se detallará el origen de los costes de personal, software y hardware.

2. PRESUPUESTO DESGLOSADO

2.1. Coste de Personal

En el cálculo del coste de personal se van a tener en cuenta tanto las horas dedicadas por el ingeniero que realiza el trabajo y por los tutores para la asesoría del trabajo, como el coste unitario por hora de cada profesional implicado.

Para el coste por hora del ingeniero biomédico se va a tener en cuenta el salario medio anual publicado por el Economic Research Institute (ERI), que en España es de 40.342 €. Teniendo en cuenta que ese salario se define para una jornada completa y que este año se tienen un total de 253 días laborables, esto deja un coste de 19,93 €/h.

En el caso de los tutores de este trabajo, ambos ingenieros informáticos, según la Universidad Europea el salario medio para un director de informática es de 53.396€. Calculando el coste por hora de la misma forma, se obtiene un coste de 26,38 €/h.

Con todos estos datos se puede calcular el coste total de personal, teniendo en cuenta el coste unitario definido para cada tipo de personal, y el número de horas empleadas en el desarrollo del proyecto. Toda esta información se resume en la **Tabla 4**.

COSTE DE PERSONAL			
Personal	Coste Unitario	Horas totales	Coste
Ingeniero biomédico	19.93 €/h	300 h	5.900,00 €
Ingeniero informático	26.38 €/h	20 h	527,60 €
Ingeniero informático	26.38 €/h	30 h	791,40 €
COSTE TOTAL DE PERSONAL			7.219,00 €

Tabla 4. Costes de personal

Por tanto, los costes de personal ascienden a 7.219,00 €.

2.2. Costes de Software

El objetivo de esta sección del presupuesto es, teniendo en cuenta la forma de cálculo para el coste imputable, determinar el coste que genera para el proyecto el software empleado. Concretamente, se ha utilizado el Sistema Operativo de Microsoft Windows 10 Home, Microsoft Office Hogar y Estudiantes 2019, así como la base de datos ClinVar. Estos costes se resumen en la **Tabla 5**:

COSTES DE SOFTWARE					
Programa	Coste de licencia	Número de licencias	Periodo de uso	Duración de la licencia	Coste imputable
Sistema Operativo de Microsoft Windows 10	145,00 €	1	6 meses	Indefinida (uso de 3 años)	24,17 €
Microsoft Office Hogar y Estudiantes 2019	149,00 €	1	6 meses	Indefinida (uso de 3 años)	24,83 €
ClinVar	0 €	Acceso libre	6 meses	Indefinidas	0 €
COSTE TOTAL DE SOFTWARE					49,00 €

Tabla 5. Costes de software

Por tanto, los costes de software ascienden a 49,00 €.

2.3. Costes de Hardware

En este apartado se detallan los costes del hardware utilizado, que en este caso es únicamente un ordenador portátil ASUS ZenBook 13, cuyo periodo de vida útil (periodo de amortización) se ha determinado en 3 años.

COSTES DE HARDWARE					
Programa	Coste de licencia	Unidades	Periodo de amortización	Periodo de uso	Coste imputable
ASUS ZenBook 13	995,83 €	1	3 años	6 meses	165,97 €
COSTE TOTAL DE HARDWARE					165,97 €

Tabla 6. Costes de hardware

Por tanto, los costes de software ascienden a 165,97 €.

3. PRESUPUESTO TOTAL

En este apartado se pretende calcular el importe total del desarrollo de este Trabajo Final de Grado. Para este cálculo, primeramente, hay que calcular el Presupuesto por Ejecución Material, que se corresponde con la suma de los costes totales de personal, software y hardware calculados en el apartado anterior. A partir de esto se pueden calcular los gastos generales y el beneficio industrial.

Los gastos generales se corresponden con un 13% del Presupuesto por Ejecución Material, y se corresponde con el capital necesario para llevar a cabo tareas en la empresa que no se encuentran directamente relacionadas con el producto o servicio ofrecido, y que por tanto no suponen un aumento en los beneficios de la empresa. El beneficio industrial supone un 6% del presupuesto por Ejecución Material y se corresponde con los beneficios reales que obtiene la empresa con la ejecución del proyecto.

La suma del valor del Presupuesto por Ejecución Material, los gastos generales y el beneficio industrial dan lugar al Presupuesto de Ejecución por Contrata. Sumando a este presupuesto el valor del IVA (21%) se obtendrá el presupuesto neto total necesario para llevar a cabo este Trabajo Final de Grado.

Según lo reflejado en la **Tabla 7** el presupuesto total requerido para el proyecto es de **diez mil setecientos cuatro con dieciocho euros (10.704,18 €)**.

CÁLCULO DEL PRESUPUESTO TOTAL	
Coste total de personal	7.219,00 €
Coste total de software	49,00 €
Coste total de hardware	165,97 €
Total Presupuesto de Ejecución Material	7.433,97 €
Gastos generales (13%)	966,42 €
Beneficio industrial (6%)	446,04 €
Total Presupuesto de Ejecución por Contrata	8.846,43 €
IVA (21%)	1.857,75 €
PRESUPUESTO TOTAL	10.704,18 €

Tabla 7. Cálculo del presupuesto total

ANEXOS

Diseño y aplicación de un método basado en SILE para la identificación de variaciones genéticas relevantes asociadas a la enfermedad de Alzheimer temprano

Documento III

Mireia Costa Sánchez
Grado en Ingeniería Biomédica
Curso académico 2019/2020

Anexo 1. Variaciones Clasificadas Como Aceptadas Tras La Aplicación Del Workflow Propuesto.

En este anexo se muestran las variaciones que han sido clasificadas como aceptadas tras la aplicación del workflow propuesto al caso del EOAD. A parte de proporcionar el identificador de la variación, se especifica el nivel de evidencia alcanzado y otra información relevante como el tipo de método, el gen en el que se localiza, el significado clínico, el *assertion criteria* utilizado, y la última fecha de revisión.

	Significado clínico	Gen	Condition	Fecha revisión	Assertion criteria	Clasificación
NM_000021.4(PSEN1): c.257T>G (p.Phe86Cys)	Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Jun 03, 2019	ACMG Guidelines, 2015	Moderate
NM_000021.4(PSEN1): c.626G>A (p.Gly209Glu)	Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 09, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
rs1555355250	Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Mar 29, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs63749836	Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Oct 31, 2018	ACMG Guidelines, 2015	Moderate
rs63751278	Pathogenic/ Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Dec 13, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
rs63751287	Pathogenic/ Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Aug 27, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
rs63751223	Likely Pathogenic	PSEN1	Alzheimer disease, type 3	Oct 31, 2018	ACMG Guidelines, 2015	Moderate
rs1553268799	Likely Pathogenic	PSEN2	Alzheimer disease, type 4	Sep 17, 2017	Nykamp K et al. (Genet Med 2017)	Limited
rs281865161	Pathogenic	APP	Alzheimer disease	Jan 07, 2019	Invitae Variant Classification Sherlock (09022015)	Limited

rs63750264 (p.Val717Phe)	Pathogenic	APP	Alzheimer disease	Nov 09, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
rs63750264 (p.Val717Ile)	Pathogenic	APP	Alzheimer disease	Aug 27, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
NM_000021.4(PSEN1): c.1254G>C (p.Leu418Phe)	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 25, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
NM_000021.4(PSEN1): c.617G>A (p.Gly206Asp)	Pathogenic	PSEN1	Alzheimer disease, type 3	Jun 01, 2018	Classification criteria August 2017	Limited
NM_000021.4(PSEN1): c.347C>A (p.Thr116Asn)	Pathogenic	PSEN1	Alzheimer disease, type 3	Aug 10, 2018	Invitae Variant Classification Sherlock (09022015)	Limited
rs1566650594	Pathogenic	PSEN1	Alzheimer disease, type 3	Apr 02, 2018	Nykamp K et al. (Genet Med 2017)	Limited
rs63750053	Pathogenic	PSEN1	Alzheimer disease, type 3	Oct 27, 2017	Nykamp K et al. (Genet Med 2017)	Limited
rs63750004	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 19, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs63750450	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 27, 2017	Nykamp K et al. (Genet Med 2017)	Limited
rs63750900	Pathogenic	PSEN1	Alzheimer disease, type 3	Dec 23, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs63749824	Pathogenic	PSEN1	Alzheimer disease, type 3	Dec 12, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs63750083	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 19, 2019	Invitae Variant Classification Sherlock (09022015)	Limited

rs63750265	Pathogenic	PSEN1	Alzheimer disease, type 3	May 28, 2019	Mendelics Assertion Criteria 2017	Limited
rs63750082	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 29, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs63750231	Pathogenic	PSEN1	Alzheimer disease, type 3	Sep 10, 2019	Invitae Variant Classification Sherlock (09022015)	Limited
rs661	Pathogenic	PSEN1	Alzheimer disease, type 3	Nov 16, 2017	Nykamp K et al. (Genet Med 2017)	Limited
rs63750526	Pathogenic	PSEN1	Alzheimer disease, type 3	Sep 11, 2017	Nykamp K et al. (Genet Med 2017)	Limited
NC_000001.10:g.(?_227069589)_(227083300_?)del	Pathogenic	PSEN2	Alzheimer disease, type 4	Sep 06, 2017	Nykamp K et al. (Genet Med 2017)	Limited
rs63750215	Pathogenic	PSEN2	Alzheimer disease, type 4	May 08, 2018	Nykamp K et al. (Genet Med 2017)	Limited
rs1799945	Pathogenic	HFE	Alzheimer disease	Oct 31, 2018	ACMG Guidelines, 2015	Moderate

Tabla 8. Variaciones clasificadas como aceptadas tras la aplicación del workflow propuesto