



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA
SUPERIOR INGENIEROS
INDUSTRIALES VALENCIA

TRABAJO FIN DE MASTER EN INGENIERÍA BIOMEDICA

IDENTIFICACIÓN DE VARIANTES GENÓMICAS MEDIANTE EL MÉTODO SILE EN LA MUERTE SÚBITA CARDIACA: APLICACIONES A LA MEDICINA DE PRECISIÓN.

AUTOR: NELVIA DEL CISNE GONZA AJILA

TUTOR: ÓSCAR PASTOR LÓPEZ

COTUTORA ANA LEÓN PALACIOS

Curso Académico: 2019-20

AGRADECIMIENTOS

En esta ocasión tan importante me gustaría agradecer el apoyo que he recibido durante mi etapa universitaria:

A mis tutores Óscar Pastor López y Ana León por dirigirme y apoyarme en el desarrollo de esta tesis.

A mi compañero de laboratorio Ignacio Pascual Hernández por su apoyo incondicional, su amistad y confianza durante estos últimos meses.

A mis amigos del máster, especialmente a Cristina Martín Vázquez y Laura González por su apoyo en los momentos difíciles.

Al Servicio Integrado de Empleo (SIE), especialmente a Ester, Susana, Ángela y Blanca por ser un amor de personas y darme el privilegio de trabajar con ellas y estudiar en la universidad.

A Maite Navarro, jefa de Ausolan en el Hospital Universitario Doctor Peset, por darme la facilidad para trabajar allí durante el grado y el máster. Y por supuesto, a mis compañeros de cocina por estar siempre pendientes de mí y hacerlo todo muy ameno.

A mis compañeros de baile por sacarme una sonrisa en los momentos difíciles de los últimos meses y hacer que siguiera adelante.

Y, por último, y más importante el apoyo de mis padres. Gracias a su valentía y coraje de viajar tan lejos para proporcionar a sus hijos un futuro prometedor.

RESUMEN

En los últimos años aumenta de forma continuada la necesidad de integrar la genómica y la medicina de precisión en el entorno clínico. La utilización correcta de los datos genómicos en la medicina personalizada es vital para mejorar la calidad, el potencial y la eficiencia de la atención médica. En este contexto, esta tesis de máster se enfoca en identificar las variantes genómicas relacionadas con la muerte súbita cardiaca a través del método SILE que permite convertir un problema de “Big Data” en una solución basada en “Smart Data”. La muerte súbita de origen cardiaco es uno de los principales problemas y es responsable de casi la mitad de todas las muertes por enfermedades del corazón. La detección de las variantes involucradas en este tipo de afección resulta pues crucial en el ámbito de la medicina de precisión. Tras la aplicación de las etapas de búsqueda e identificación del método SILE, se encontraron 38 variantes relevantes relacionados con 6 de los 9 fenotipos estudiados. Estas variantes son extraídas tras la aplicación de una serie de filtros y criterios de la ACMG en los datos extraídos de Clinvar, Ensembl y GWAS Catalog. Únicamente de los datos iniciales analizados el 0,18% de Clinvar, el 5,26 % de GWAS Catalog y 0,04% de Ensembl resultan relevantes.

Palabras Clave: variantes, muerte súbita cardiaca, SILE, medicina de precisión.

RESUM

En els últims anys augmenta de forma continuada la necessitat d'integrar la genòmica i la medicina de precisió en l'entorn clínic. La utilització correcta de les dades genòmiques en la medicina personalitzada és vital per a millorar la qualitat, el potencial i l'eficiència de l'atenció mèdica. En este context, esta tesi de màster s'enfoca a identificar les variants genòmiques relacionades amb la mort sobtada cardíaca a través del mètode SILE que permet convertir un problema de Big Data en una solució basada en Smart Data. La mort sobtada cardíaca és un dels principals problemes i és responsable de quasi la mitat de totes les morts per malalties del cor. La detecció de les variacions involucrades en la mort sobtada resulten perquè crucials en l'àmbit de la medicina de precisió. Després de l'aplicació de les etapes de busca i identificació del mètode SILE, es van trobar 38 variants rellevants relacionats amb 6 dels 9 fenotips estudiats. Estes variants són extretes després de l'aplicació d'una sèrie de filtres i criteris de l'ACMG en les dades extrets de Clinvar, Ensembl i GWAS Catalog. Únicament de les dades inicials analitzats el 0,18% de Clinvar, el 5,26 % de GWAS Catalog i 0,04% d'Ensembl resulten rellevants."

Paraules clau: variacions, mort sobtada cardíaca, SILE, medicina de precisió.

ABSTRACT

In recent years, the need to integrate genomics and precision medicine in the clinical environment has been constantly increasing. The correct use of genomic data in personalized medicine is vital to improve the quality, potential and efficiency of medical care. In this context, this master's thesis focuses on identifying genomic variants related to sudden cardiac death through the SILE method that makes it possible to convert a "Big Data" problem into a solution based on "Smart Data". Sudden cardiac death is one of the main problems and is responsible for almost half of all deaths from heart disease. The detection of the variations involved in sudden death are therefore crucial in the field of precision medicine. After the application of the search and identification stages of the SILE method, 38 relevant variants were found related to 6 of the 9 phenotypes studied. These variants are extracted after the application of filters propose by the SILE method and criteria of the ACMG in the data extracted from Clinvar, Ensembl and GWAS Catalog. Only of the initial data analyzed, 0.18% of Clinvar, 5.26% of GWAS Catalog and 0.04% of Ensembl are relevant.

Keywords: variations, sudden cardiac death, SILE, precision medicine.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

DOCUMENTOS CONTENIDOS EN LA MEMORIA

- DOCUMENTO I: MEMORIA
- DOCUMENTO II: PRESUPUESTO
- DOCUMENTO III: ANEXOS

ÍNDICE DE LA MEMORIA

CAPÍTULO 1. INTRODUCCIÓN	16
1.1 MOTIVACIÓN	16
1.2 OBJETIVOS DEL TRABAJO.....	17
1.3. ESTRUCTURA DEL DOCUMENTO	17
CAPÍTULO 2. ESTADO DEL ARTE	19
2.1. MEDICINA PRECISIÓN	19
2.2. MUERTE SÚBITA CARDÍACA.....	19
2.3. TRABAJOS RELACIONADOS.....	20
Capítulo 3. MÉTODO Y FENOTIPOS.....	21
3.1. MÉTODO SILE	21
3.2. FENOTIPOS ESTUDIADOS.....	22
3.2.1 Miocardiopatía dilatada familiar (<i>familial dilated cardiomyopathy</i>).....	23
3.2.2 Síndrome de QT corto (<i>Short QT syndrome</i>)	24
3.2.3 Miocardiopatía restrictiva familiar (<i>familial restrictive cardiomyopathy</i>).....	25
3.2.4 Síndrome de Loey-Dietz (<i>Loeys-Dietz syndrome</i>).....	26
3.2.5 Miocardiopatía hipertrófica familiar (<i>Primary familial hypertrophic cardiomyopathy</i>)	27
3.2.6 Síndrome de Brugada (<i>Brugada syndrome</i>)	27
3.2.7 Síndrome de Marfan (<i>Marfan syndrome</i>)	28
3.2.8 Taquicardia ventricular catecolaminérgica polimórfica (<i>Catecholaminergic polymorphic ventricular tachycardia (CPVT)</i>).....	29
3.2.9 Síndrome de QT largo (<i>Long QT syndrome</i>)	30
Capítulo 4. APLICACIÓN DEL MÉTODO SILE.....	32
4.1. BÚSQUEDA	32
4.1.1 Varsome	34
4.1.2 DbSNP	35
4.1.3 SNPedia	35
4.1.4 Clinvar.....	36
4.1.5 Ensembl	37
4.1.6 GWAS Catalog.....	39
4.2 IDENTIFICACIÓN	41
4.2.1 Guías y estándares para interpretación de variantes por el ACMG	52
4.2.2 CardioVai, Varsome e intervar	55

Capítulo 5. conclusiones y trabajo futuro	59
REFERENCIAS BIBLIOGRÁFICAS	61

ÍNDICE DEL PRESUPUESTO

1. NECESIDAD DEL PRESUPUESTO	69
2. PRESUPUESTO DESGLOSADO	69
2.1 COSTE DEL PERSONAL	69
2.2 COSTE DEL SOFTWARE	69
2.2 COSTE DEL HARDWARE	70
3. PRESUPUESTO TOTAL.....	70

ÍNDICE DE LOS ANEXOS

Anexo 1. Subcategorías de la cardiomiopatía dilatada familiar.	75
Anexo 2. Subcategorías de la cardiomiopatía hipertrófica familiar.	76
Anexo 3. Conjunto de variantes examinadas por un panel de expertos para la miocardiopatía hipertrófica.....	77
Anexo 4. Conjunto de variantes finales para la muerte súbita cardíaca (GWAS Catalog).....	78
Anexo 5. Variantes finales para la muerte súbita cardíaca.	79

LISTA DE FIGURAS

Figura 1. Patrón de herencia dominante [34]	24
Figura 2. Latidos cardíacos normal y anormal [46].	29
Figura 3. Varsome(Website).....	34
Figura 4. DbSNP (Website).	35
Figura 5. SNPedia (Website).....	35
Figura 6. Clinvar (Website).	36
Figura 7. Ejemplo de la búsqueda de variantes en Clinvar.....	37
Figura 8. Ensembl (Website).	38
Figura 9. Ejemplo de búsqueda en Ensembl.	38
Figura 10. GWAS Catalog (website).....	39
Figura 11. Ejemplo de búsqueda en GWAS Catalog.	40

Figura 12. Número de variantes totales relacionadas con la muerte súbita y los fenotipos que la provocan obtenidas en tres repositorios.	41
Figura 13. Flujo de trabajo para la clasificación de variantes [60].	43
Figura 14. Primeros resultados en Clinvar tras aplicar los filtros F1-F4 del método.	44
Figura 15. Primeros resultados en Clinvar tras aplicar los filtros F1-F4 del método SILE (parte 2).	44
Figura 16. Primeros resultados en ENSEMBL tras aplicar los filtros F1-F4 del método SILE (parte 1).	45
Figura 17. Primeros resultados en Ensembl tras aplicar los filtros F1-F4 del método SILE (parte 2).	46
Figura 18. Primeros resultados en GWAS Catalog tras aplicar el filtro F5.	47
Figura 19. Resumen del número de variantes obtenidas tras la clasificación de artículos.	50
Figura 20. Nuevo flujo de trabajo para la obtención de variantes genómicas.	51
Figura 21. CardioVAI (Website).	56
Figura 22. InterVAR (Website).	57
Figura 23. Variantes finales.	58

LISTA DE TABLAS

DOCUMENTO I. MEMORIA

Tabla 1. Descripción de las métricas utilizadas para la selección de las bases de datos.	33
Tabla 2. Criterios de calidad para la selección de variantes.	42
Tabla 3. Resultados de la clasificación de artículos.	49
Tabla 4. Aplicación de las reglas del ACMG.	54
Tabla 5. Conjunto de reglas que se cumplen en CardioVAI.	57
Tabla 6. Costes del personal del proyecto.	69

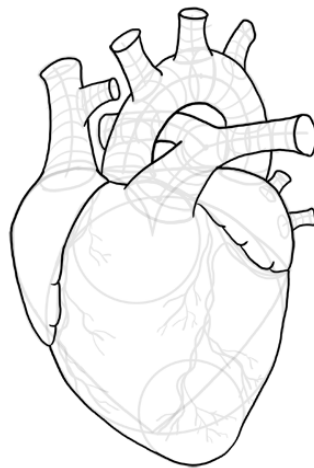
DOCUMENTO III. PRESUPUESTO

Tabla 7. Coste del software del proyecto.	70
Tabla 8. Coste del hardware del proyecto.	70
Tabla 9. Costes del presupuesto total.	71

MEMORIA

IDENTIFICACIÓN DE VARIANTES GENÓMICAS MEDIANTE EL MÉTODO SILE EN LA MUERTE SÚBITA CARDIACA: APLICACIONES A LA MEDICINA DE PRECISIÓN

DOCUMENTO I



Nelvia del Cisne Gonza Ajila

Curso académico 2019/2020

CAPÍTULO 1. INTRODUCCIÓN

Esta tesis de máster se enmarca en el ámbito de investigación de la ingeniería biomédica, concretamente en la calidad y análisis de datos genéticos y su uso en el diagnóstico clínico. En este capítulo se describe la motivación para realizar este proyecto, los objetivos y la estructura seguida en este trabajo fin de máster.

1.1 MOTIVACIÓN

En los últimos años, las tecnologías de secuenciación del ADN, por ejemplo, la conocida como “*Next Generation Sequencing*” (NGS), se van desarrollando cada vez más rápido y son a la vez mucho más económicas [1]. El proyecto genoma humano empezó oficialmente en 1990 culminándose en 2003 [2]. La combinación de las nuevas tecnologías de secuenciación y el conocimiento del genoma permiten la generación y compresión de información que resulta útil en el diagnóstico clínico, por ejemplo, las patologías genéticas. Actualmente, los investigadores pueden secuenciar millones de bases del ADN y compararlas generando una **gran cantidad de información** sobre la herencia de las enfermedades y la respuesta a factores ambientales. Además, la capacidad de secuenciar el genoma con mayor rapidez y eficacia crea un gran potencial para el **diagnóstico y tratamiento** de las enfermedades.

Cada vez está más cerca que un médico pueda realizar una secuenciación del ADN a un paciente en la consulta médica de forma rutinaria para detectar y tratar enfermedades de origen genético. Por ejemplo, el cáncer es una enfermedad cada vez más frecuente en la que los médicos pueden usar datos de secuenciación genética para identificar mediante estos el tipo específico de cáncer que tenga ese paciente. Esto le permite al médico tomar mejores decisiones en cuanto a los tratamientos y medicamentos más adecuados para el paciente.

Sin embargo, no es suficiente únicamente realizar la secuenciación, hay que entenderla y obtener los datos que resulten más útiles para que el médico pueda tomar las decisiones adecuadas. El **dominio genómico** es muy **complejo** a la hora de gestionar la información tanto por su heterogeneidad, la calidad y por la cantidad de datos que se generan cada día. Esto ocasiona que para un profesional de la salud realizar un diagnóstico genético sea una tarea difícil dada la complejidad de manejar grandes cantidades de información genética para dar un diagnóstico. Por eso, es muy importante desarrollar o aplicar técnicas y algoritmos de manera que se obtenga datos de calidad y de utilidad para la comunidad científica médica.

Tanto empresas como instituciones se enfocan en crear métodos para explotar la información genómica, como es el caso con el centro de investigación PROS en la UPV, donde se ha desarrollado este TFM. Este centro ha desarrollado un método basado en los **modelos conceptuales** que permite manejar e integrar la información genómica de una forma eficiente. Un modelo conceptual (MC) se define como una representación gráfica de un proceso o sistema con un objetivo específico. Éste permite la comprensión y comunicación de la información ayudando a la creación de sistemas de información, construcción de teorías, explicación de fenómenos o documentar un sistema existente [3].

Dada la cantidad de información sobre genómica disponible para ser explotada, se pretende mediante la utilización del método SILE organizar la información y extraer la que resulte

relevante para los profesionales de la salud, mejorando así el diagnóstico, el tratamiento, la detección y la prevención de patologías.

A través de este trabajo, se va a demostrar que con el uso del método desarrollado por esta institución se pueden identificar las variantes genómicas más relevantes asociadas con el riesgo de desarrollar una patología de entre la multitud de información disponible, en este caso las que provocan la muerte súbita cardíaca, con el objetivo de proporcionar un diagnóstico genético de calidad.

Una variante puede ser una mutación o polimorfismo. Una mutación se define como un cambio permanente en la secuencia de nucleótidos, mientras que un polimorfismo se define como una variante con una frecuencia superior al 1%. Sin embargo, los términos “mutación” y el “polimorfismo” generan confusión por lo que se recomienda que ambos términos sean reemplazados por el término “variante” con los siguientes modificadores: patogénico, probable patogénico, significado incierto, probable benigno, o benigno [4].

1.2 OBJETIVOS DEL TRABAJO

El objetivo general de esta tesis de máster es identificar un conjunto de variantes genómicas que causen la muerte súbita utilizando un método que se basa en el *Modelo Conceptual del Genoma Humano (MCGH)* [5] y que aborda la heterogeneidad y dispersión de la información, obteniendo así datos de calidad a través de la explotación de los repositorios genómicos disponibles.

Los objetivos fundamentales de esta tesis son los siguientes:

1. Analizar la necesidad de identificar las variantes genómicas que causan la muerte súbita cardíaca.
2. Identificar las variantes relevantes de ADN para la muerte súbita cardíaca mediante el método SILE para propósitos clínicos.
3. Validar las variantes genómicas encontradas.

Estos objetivos nos llevan al planteamiento de las siguientes preguntas de investigación.

- P1. ¿En qué ayudaría saber qué variantes genómicas causan la muerte súbita?
- P2. ¿Cuáles son las variantes identificadas utilizando el método SILE?
- P3. ¿De qué forma se puede validar las variantes genómicas que causan la muerte súbita?

1.3. ESTRUCTURA DEL DOCUMENTO

A continuación, una vez ya planteados los objetivos, se explica la estructura del trabajo fin de máster que consta de 5 capítulos.

- **CAPÍTULO 1. INTRODUCCIÓN.** En este capítulo se explica cuál es la motivación para realizar este trabajo final de master, los objetivos del documento y su estructura.
- **CAPÍTULO 2. ESTADO DEL ARTE.** Se presenta el estado del arte asociado con esta tesis. Se ha dividido en 2 partes: por un lado, el contexto de trabajo (la medicina de precisión)

y por otro el fenotipo/enfermedad objeto de estudio (la muerte súbita cardíaca). Además, se presentarán estudios previamente realizados utilizando la metodología empleada en este TFM.

- **CAPÍTULO 3. MÉTODO Y FENOTIPOS.** En este capítulo se profundiza en la descripción de los fenotipos relacionados con la muerte súbita cardíaca, los genes implicados, la frecuencia, el modo de herencia y las subcategorías de cada uno de ellos, lo que ayudará a llevar a cabo la aplicación del método. Además, se introduce el método utilizado en esta tesis de máster.
- **CAPÍTULO 4. APLICACIÓN DEL MÉTODO SILE.** Una vez conocidos los fenotipos a utilizar y el método, se procede a aplicar el método. Los pasos y resultados obtenidos durante su ejecución se exponen en este capítulo.
- **CAPÍTULO 5. CONCLUSIONES Y TRABAJOS FUTUROS.** Se presentan las conclusiones y contribuciones de este trabajo. Además, se plantean posibles líneas futuras.

CAPÍTULO 2. ESTADO DEL ARTE

En este capítulo se explica qué se entiende como medicina de precisión y una breve introducción sobre la muerte súbita cardíaca.

2.1. MEDICINA PRECISIÓN

El término "medicina de precisión" se puede definir como *"dar el tratamiento correcto, en el momento correcto, siempre, a la persona correcta"* [6]. Aunque su significado exacto no está aún claro, nos podemos referir a la medicina de precisión con otros términos como "medicina estratificada", "terapia dirigida" o "fenotipado profundo". La medicina de precisión incluye la información clínica, el estilo de vida, **la genética** y otros biomarcadores, por lo que va más allá de los "signos y síntomas" clásicos. La medicina de precisión usa información de los genes, proteínas y otras características de la patología de una persona de manera que se pueda determinar el diagnóstico o tratamiento más adecuado.

En la parte de la genética de la medicina de precisión es donde esta tesis juega un papel relevante. Este TFM, se centra en la **muerte súbita cardíaca** que causa un gran impacto social, además de ser una de las preocupaciones más importantes dentro del ámbito de la cardiología, dada la complejidad de ser detectada a tiempo. La inmensa mayoría de los pacientes que sufren una muerte súbita y no reciben atención médica fallecen en pocos minutos [7]. Trabajar en este ámbito, mejorando el entorno clínico preventivo, ayudaría a aumentar la tasa de supervivencia de las personas que la sufren dado que aún no se conoce una forma de prevenirla. La combinación de la genética y la tecnología para la prevención de cualquier enfermedad supone un gran impacto en el ahorro de los costes médicos.

A continuación, se comentarán de los antecedentes de la muerte súbita cardíaca.

2.2. MUERTE SÚBITA CARDÍACA

Se define la muerte súbita cardíaca (MS) como: *"Una muerte natural inesperada por causa cardíaca en un corto período de tiempo desde el inicio de los síntomas"* [8]. La MS se conoce desde hace miles de años [9]. En el antiguo Egipto, hace más de 4.000 años, la MS ya estaba asociada con la isquemia miocárdica [10]. En los papiros existentes de Ebers se describe que *"Si un paciente tiene dolor en el brazo y el lado izquierdo del pecho, existe una amenaza de muerte"* [11]. En China, hace 2.400 años, Chio ya asoció la muerte súbita con una arritmia cuando manifestó que *"Un pulso intermitente es un predictor de muerte inminente"* [12]. Casi al mismo tiempo, Hipócrates manifiesta: *"El dolor intenso en el pecho que se irradia a la clavícula y la espalda es un signo de mal pronóstico"* [13]. Además, fue el primero en introducir los factores de riesgo, como el de obesidad. Estos conceptos se mantienen actualmente. A lo largo de los siglos, la MS ha llamado la atención por la conmoción y el miedo que produce. En el siglo XIV, el conde Gastón de Foix murió repentinamente después de regresar de una cacería durante la cual tuvo contacto con agua helada. Sintió una opresión en el pecho y dijo: *"Soy un hombre muerto. Que Dios tenga misericordia de mí"* [14]. En el siglo XVIII, el Papa Clemente XI publicó un libro escrito por Lancisi sobre los casos frecuentes de MS que ocurrían en Roma [15]. En el siglo XIX, Von Bezold demostró que la oclusión experimental de las arterias coronarias causaba un paro cardíaco [10]. En el siglo XX, **Herrick** describió la descripción clínica del infarto de miocardio (IM) [16]. A lo largo de este siglo, el interés en la muerte súbita

aumentó a medida que su asociación con la enfermedad coronaria se hizo más evidente [17]. Ya en la segunda mitad del siglo XX, se demostró que, aunque la cardiopatía isquémica (CI) seguía siendo la causa de la MS en al menos el 80% de los casos [18], un grupo de enfermedades hereditarias con alteraciones estructurales (miocardiopatía) o sin causa orgánica aparente (canalopatías) explicaban muchos casos de muerte súbita en personas más jóvenes sin CI.

La muerte súbita probablemente sea uno de los desafíos más importantes de la cardiología moderna. Es probable que la incidencia anual de la muerte súbita, únicamente en los EEUU, esté entre 180 y 250.000 personas por año [19], aunque su incidencia en España y en países del Mediterráneo se ha visto disminuida [20]. La muerte súbita no es únicamente un reto por el **número de muertes** que se producen sino también por el **impacto social, económico y mediático** que esta produce. La incidencia de la muerte súbita aumenta con la edad, aproximadamente entre los 40 y 45 años en adelante. También es común cuando se tiene cualquier enfermedad cardíaca. La MS también se puede producir en bebés, pero estos son casos muy raros. En las primeras décadas de vida, la muerte súbita aparece durante actividades deportivas o por presencia de enfermedad cardíacas genéticas, por ejemplo, la cardiomiopatía hipertrófica. Es importante destacar que la muerte súbita no es una enfermedad, sino una consecuencia de una enfermedad cardíaca.

En el siguiente punto se menciona trabajos relacionados con el método utilizado en esta tesis.

2.3. TRABAJOS RELACIONADOS

El método SILE que se utiliza en esta tesis ha sido utilizado para determinar variantes relacionadas con enfermedades como el Alzheimer, el neuroblastoma o las cataratas:

- Análisis de datos genómicos para la identificación de variantes genómicas relevantes en la enfermedad de Alzheimer (TFM 2019) [21].
- Biotecnología para la medicina de precisión: Identificación de variantes genómicas para el diagnóstico de cáncer de mama (TFG 2018) [22].
- Un proceso para la identificación sistemática de variantes genómicas: aplicaciones a la medicina de precisión (TFM 2018) [23].
- Exploración de bases de datos genómicas dirigido por modelos conceptuales (TFM 2018) [24].
- Diseño de un sistema de información genómica para el diagnóstico del Neuroblastoma (TFG 2016) [25].

Este TFM es el primer estudio realizado en el ámbito de las cardiomiopatías, lo que supone la primera aplicación práctica del método para este fenotipo de interés, con el valor adicional de evaluar las particularidades que su aplicación en un dominio diferente puede presentar. Pretende demostrar además que la aplicación del método para diferentes fenotipos podría convertirse en la herramienta esencial para la generación de un biobanco de datos genómicos muy valioso para la realización de un diagnóstico genómico de calidad.

Esta tesis de máster se ha desarrollado en el centro de investigación PROS, perteneciente a la Universidad Politécnica de Valencia.

CAPÍTULO 3. MÉTODO Y FENOTIPOS

En este capítulo se explicará el método a utilizar, las enfermedades relacionadas con la muerte súbita cardíaca de las cuales se obtendrán las variantes genómicas causantes de la patología, los repositorios de datos implicados en la obtención de la información genética y los resultados obtenidos tras aplicar el método descrita en este capítulo.

3.1. MÉTODO SILE

En [5], se demuestra que aplicando los modelos conceptuales se pueden desarrollar sistemas de información genómica que permiten manejar las grandes cantidades de información genética (*Big Data*) disponible en la actualidad.

El método SILE se utiliza para identificar las variantes genómicas relevantes relacionadas con el estudio de enfermedades de origen genómico. Además, proporciona una forma de integrar y gestionar la información genómica. La información genómica gestionada mediante este método permite la obtención de variantes genómicas de calidad (*Smart Data*) que contribuyen a hacer de la medicina personalizada una realidad clínica que cambie la práctica médica convencional que se ha venido conociendo hasta ahora.

El método que se utiliza en esta tesis hace uso de los modelos conceptuales para poder gestionar la información genómica, en concreto el MC que utiliza es el Modelo Conceptual del Genoma Humano (MCGH) desarrollado por el centro PROS [26]. Este modelo conceptual consta de 5 vistas [5]:

- Vista Estructural: describe la estructura del genoma.
- Vista de Transcripción: muestra los conceptos y componentes relacionados con la síntesis de proteínas.
- Vista de Variaciones: modela el conocimiento relacionado con las diferencias en la secuencia de ADN de diversos individuos.
- Vista de Rutas Metabólicas: modela una serie de reacciones químicas que ocurren dentro de la célula.
- Vista de Fuentes de Datos y Bibliografía: proporcionada información sobre las fuentes de datos de cuales se ha extraído la información.

En concreto, la vista en la que se basa el método para saber qué información genómica manejar para identificar las variantes relevantes es la vista de variaciones.

El método SILE (*Search-Identification-Load-Exploitation*) tiene como objetivo sistematizar la búsqueda e identificación de información genética para ser cargada, analizada y explotada por un sistema de información genómica [26]. El método consta de 4 fases:

- **Búsqueda ("Search"):**
Se trata de realizar la búsqueda y selección de base de datos necesarias para extraer la información de interés.
- **Identificación ("Identification"):**
En esta parte del método se procede a filtrar la información obtenida en la fase anterior, identificando así las variantes relevantes para la realización de un diagnóstico genético.

- **Carga (“Load”):**
Una vez identificadas las variantes, se procede a su carga en una base de datos, para su posterior explotación [5]. La base de datos donde se almacena la información se denomina Human Genome Database (HGDB) y es el resultado de la implementación del modelo conceptual anteriormente mencionado. Esta base de datos fue desarrollada por el centro de investigación PROS [5].
- **Explotación (“Exploitation”):**
Durante esta fase se extrae nuevo conocimiento de la base de datos con herramientas especialmente diseñadas para analizar e interpretar la información genómica. Este es el último paso del método SILE, los datos obtenidos sobre las variantes genómicas, una vez cargados en el proceso de carga, se podrían utilizar para el diagnóstico genético por profesionales médicos. Una de las herramientas a utilizar es la aplicación GenesLove.Me [5]. Se trata de una aplicación web pensada en ofrecer Test Genéticos Directos al Consumidor (TGDC). Geneslove.me utiliza los datos del HGDB a través de la herramienta VarSearch [5]. Esta herramienta permite el análisis de variantes obtenidas de la secuenciación del ADN sobre muestras biológicas y que sean almacenados en formato FASTA o VCF. Obtener los TGDC permitiría a los seres humanos que se lo realicen un diagnóstico temprano de las enfermedades de origen genético.
Los datos utilizados en esta plataforma GenesLove.Me han pasado un proceso de estudio y validación realizado por profesionales médicos.

Al aplicar estas etapas se pasa de *Big Data* a *Smart Data*, permitiendo así obtener información precisa y fiable que nos acerca más al diagnóstico clínico en la medicina de precisión. El desarrollo de este trabajo se centra en la aplicación de las dos primeras etapas, ya que la carga y la explotación quedan fuera del ámbito de este trabajo fin de máster.

Una vez introducido el método se procede a explicar qué fenotipos van a formar parte del estudio en esta tesis.

3.2. FENOTIPOS ESTUDIADOS

Antes de empezar a trabajar con el método SILE, hay que familiarizarse con el fenotipo elegido dado que ayuda a que el proceso de identificación sea más simple. La predisposición genética a desarrollar una enfermedad depende de muchos factores y en ocasiones la alteración en un único gen no es suficiente para que la patología se manifieste.

Una enfermedad podría manifestarse por alteraciones en la secuencia del ADN de un solo gen (monogénicas) o por alteraciones en varios genes (poligénicas). Por ejemplo, la Diabetes Mellitus tipo 1 es poligénica [27] y el Síndrome de Down es monogénico [28]. Entender si una patología es monogénica o poligénica ayudar a determinar si presentar las alteraciones genéticas provoca o no la enfermedad. Además, es de gran utilidad a la hora de tratar la información relacionada con dicha enfermedad.

También es relevante conocer su penetrancia¹ en la población, así como el modo de herencia ya sea dominante o recesivo. Además, las enfermedades han ido adquiriendo distintos nombres y nomenclaturas a lo largo de los años, por lo que determinar esto es muy

¹ La penetrancia es la proporción de individuos que presentan el fenotipo en una población.

importante para la búsqueda de información sobre variantes genómicas asociadas a un fenotipo, evitando la pérdida de información y obtener así datos más precisos para empezar con la identificación de variantes.

El fenotipo tratado en este proyecto es la muerte súbita cardíaca que en sí no es una patología sino una causa de alguna enfermedad en la que el corazón está involucrado. Las enfermedades cardíacas asociadas a la MS se conocen como cardiomiopatías familiares (CF). Las CF son un grupo de enfermedades cardiovasculares que se clasifican como: miocardiopatías, canalopatías, algunas enfermedades aórticas y otros síndromes [29]. Todas estas tienen tres características en común:

- Se manifiestan en la familia.
- Son de origen genético.
- Pueden provocar la muerte súbita.

Puesto que las enfermedades que causan la muerte súbita tienen un origen genético, una vez ya identificadas las variantes que la causan se puede hacer un diagnóstico clínico. Según [29], las patologías genéticas involucradas con la muerte súbita y sobre las cuales nos vamos a centrar en este TFM son las siguientes:

- La miocardiopatía hipertrófica.
- La miocardiopatía dilatada.
- La miocardiopatía restrictiva.
- El síndrome de Brugada.
- La taquicardia ventricular catelominérgica polimórfica (TVCP).
- El síndrome de QT largo.
- El síndrome de QT corto.
- El síndrome de Marfan.
- El síndrome de Loeys-Dietz.

En los puntos siguientes se procede a describir los fenotipos nombrados anteriormente con el fin de conocer un poco más sobre ellos antes de comenzar con la identificación de las variantes.

3.2.1 Miocardiopatía dilatada familiar (*familial dilated cardiomyopathy*)

La miocardiopatía dilatada familiar (OMIM: PS115200) es una enfermedad genética que causa que el músculo cardíaco adelgace y se debilite en al menos una de las cuatro cámaras del corazón, provocando que el área abierta de la cámara se agrande (dilata). Como consecuencia, el corazón no es capaz de bombear sangre como de costumbre. A medida que pasa el tiempo, esta condición provoca insuficiencia cardíaca. Por lo general, los síntomas de la miocardiopatía dilatada familiar tardan muchos años en causar problemas de salud y típicamente empiezan en la edad adulta, pero pueden ocurrir en cualquier momento desde la infancia hasta la edad adulta tardía [30].

En cuanto a los sinónimos asociados al fenotipo se ha encontrado un único sinónimo, correspondiente a su acrónimo en MedGen²: FDCM.

² MedGen: <https://www.ncbi.nlm.nih.gov/medgen>.

Dentro de la miocardiopatía dilatada familiar se encuentran más de 30 subcategorías [31, 32, 33], por ejemplo, *Dilated cardiomyopathy 1A* o *Dilated cardiomyopathy 1BB*. Para más información ver Anexo 1.

Si se habla de los genes relacionados, se ha encontrado que las mutaciones que se producen en más de 30 genes son causantes de la miocardiopatía familiar. Estos genes son: RBM20, NEXN, LDB3, ABCC9, BAG3, TCAP, CSRP3, VCL, TTN, TPM1, TNNT2, TNNI3, TNNC1, TAZ, SGCD, SDHA, SCN5A, PSEN2, PLN, MYH7, MYH6, LMNA, LAMA4, FKTN, EYA4, DSG2, DES, CRYAB, ACTN2, ACTC1. Las mutaciones en el gen TTN representan aproximadamente el 20 por ciento de los casos de la miocardiopatía dilatada familiar. No está claro el porcentaje de las otras mutaciones en los otros genes que causan la miocardiopatía dilatada [30].

Se estima que 750.000 personas de los Estados Unidos tienen miocardiopatía dilatada, aproximadamente la mitad de estos casos son familiares [30].

La miocardiopatía a dilatada familiar tiene diferentes patrones de herencia dependiendo del gen involucrado. En el 80% de los casos, la miocardiopatía dilatada familiar se hereda siguiendo un patrón **autosómico dominante** (Figura 1), lo que significa que la persona afectada hereda la mutación de uno de los progenitores afectados. Sin embargo, algunas personas que heredan el gen alterado nunca desarrollan las características de la miocardiopatía familiar. Esta situación se conoce como penetrancia reducida. Otros casos resultan de nuevas mutaciones en el gen y ocurren en personas sin antecedentes de trastorno en su familia [30].

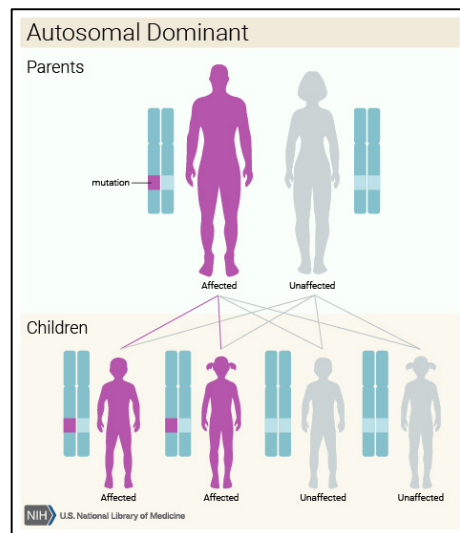


Figura 1. Patrón de herencia dominante [34]

En casos raros, esta condición se hereda siguiendo un patrón autosómico recesivo, lo que significa que dos copias del gen en cada célula tienen mutaciones. Los padres de un individuo con una condición autosómica recesiva tienen una copia del gen mutado, pero generalmente no muestran signos y síntomas de la condición.

3.2.2 Síndrome de QT corto (*Short QT syndrome*)

El síndrome de QT corto (OMIM: PS609620) es una condición que puede causar una disrupción del ritmo normal del corazón (arritmia). El término “QT corto” se refiere a un patrón específico de la actividad del corazón que es detectado mediante la realización de un electrocardiograma (ECG). En las personas que sufren esta condición, la parte del latido cardíaco conocida como

intervalo QT es anormalmente corta. Si no es tratada, la arritmia asociada con el síndrome de QT corto puede conducir a una variedad de signos y síntomas, desde mareos y desmayos hasta paro cardíaco y muerte súbita. Estos signos y síntomas pueden ocurrir en cualquier momento de la infancia hasta la vejez. Esta condición puede explicar algunos casos del síndrome de muerte súbita infantil (*Sudden infant death death syndrome (SIDS)*) el cuál es la mayor causa de muerte inexplicable en bebés menores de un año. No obstante, algunas personas con este síndrome nunca presentan problemas asociados con la condición [35].

Los sinónimos asociados a este fenotipo según MedGen son: *Familial short QT syndrome* y *SQT*.

En cuanto a las subcategorías de este fenotipo son [36]:

- **Short QT syndrome 1 (SQT1)**
- **Short QT syndrome 2 (SQT2)**
- **Short QT syndrome 3 (SQT3)**

En la búsqueda de información se encuentran asociados al síndrome descrito los siguientes genes: KCNH2, KCNJ2 y KCNQ1 [36].

El síndrome de QT corto se considera raro. Al menos 70 casos han sido identificados en todo el mundo desde que esta condición fue descubierta en el año 2.000. Sin embargo, en algunos casos puede estar subdiagnosticado porque algunas personas afectadas nunca experimentan síntomas [35].

El síndrome de QT corto parece tener un patrón de herencia **autosómico dominante**. Algunos individuos afectados tienen una historia familiar del síndrome de QT corto o problemas cardíacos relacionados y muerte súbita cardíaca. Otros casos del síndrome de QT corto son clasificados como esporádicos y ocurren en personas sin antecedentes familiares aparentes de problemas cardíacos relacionados [35].

3.2.3 Miocardiopatía restrictiva familiar (*familial restrictive cardiomyopathy*)

La miocardiopatía restrictiva familiar (OMIM: PS115210) es una enfermedad genética cardíaca que puede aparecer en cualquier momento desde la infancia hasta la edad adulta. Algunos niños con miocardiopatía restrictiva familiar no presentan síntomas o signos obvios, pero pueden morir repentinamente debido a un fallo cardíaco. Sin tratamiento la mayoría de niños que padecen este síndrome sobreviven solo unos años después de que son diagnosticados [37].

Los sinónimos asociados a este fenotipo según MedGen son: *Childhood Restrictive Cardiomyopathy* y *RCM*.

Los genes asociados que causan la miocardiopatía restrictiva familiar son: TNNT2, TNNI3. Las mutaciones en el gen TNNI3 son una de las mayores causas de esta condición y las mutaciones en otros genes asociados con la miocardiopatía restrictiva familiar representan un pequeño porcentaje de casos [37].

La prevalencia es desconocida. Aunque la miocardiopatía es una afección relativamente común, la miocardiopatía restrictiva, en la que se altera la relajación del músculo cardíaco es el tipo menos común. En los Estados Unidos y en Europa, la miocardiopatía restrictiva representa

menos del 5% de todas las miocardiopatías La proporción de la miocardiopatía restrictiva que se da en familias se desconoce [37].

Las subcategorías de esta condición y sus respectivos sinónimos según [38] son:

- **Familial restrictive cardiomyopathy 1 (RCM1):** *TNNI3-Related Familial Restrictive Cardiomyopathy*
- **Familial restrictive cardiomyopathy 2 (RCM2)**
- **Familial restrictive cardiomyopathy 3 (RCM3):** *TNNT2-Related Familial Restrictive Cardiomyopathy*

Esta condición se hereda en un patrón autosómico dominante [37].

3.2.4 Síndrome de Loeys-Dietz (*Loeys-Dietz syndrome*)

El síndrome de Loeys-Dietz (OMIM: PS609192) es un trastorno que afecta al tejido conectivo en muchas partes del cuerpo. Hay cinco tipos de este síndrome, etiquetados de tipo I a V, que se distinguen por su causa genética. Independientemente del tipo, los signos y los síntomas del síndrome descrito pueden manifestarse en cualquier momento de la vida del individuo y la gravedad es variable. Este síndrome se caracteriza por agrandamiento de la aorta. Las personas con este síndrome pueden tener aneurismas o disecciones en arterias de todo el cuerpo y tener arterias con giros y vueltas anormales (tortuosidad arterial) [39].

Se encontró un único sinónimo asociado al síndrome de Loeys-Dietz en MedGen correspondiente a su acrónimo: *LDS*.

Las subcategorías del síndrome de Loeys-Dietz y sus correspondientes sinónimos son las siguientes [40]:

- **Loeys-Dietz syndrome 1 (LDS1):** *Aortic aneurysm syndrome, familial thoracic 5; Furlong syndrome, Loeys-Dietz syndrome type 1A, Loeys-Dietz syndrome type 2A, TGFBR1-Related Loeys-Dietz Syndrome, TGFBR1-Related Thoracic Aortic Aneurysms and Aortic Dissections*
- **Loeys-Dietz syndrome 2 (LDS2):** *Aortic Aneurysm, Familial thoracic 3*
- **Loeys-Dietz syndrome 3 (LDS3):** *Aortic aneurysm, familial thoracic 3, Loeys-Dietz syndrome type 1B, Loeys-Dietz syndrome type 2B, TGFBR2-Related Loeys-Dietz Syndrome, TGFBR2-Related Thoracic Aortic Aneurysms and Aortic Dissections*
- **Loeys-Dietz syndrome 4 (LDS4):** *Aneurysm, aortic and cerebral, with arterial tortuosity and skeletal manifestations, TGFB2-Related Loeys-Dietz Syndrome*
- **Loeys-Dietz syndrome 5 (LDS5):** *Riehhoff syndrome (RNHF)*

Los cinco tipos del síndrome de Loeys-Dietz se distinguen por su causa genética: Las mutaciones en el gen TGFBR1 causan el tipo I, las mutaciones en TGFBR2 causan el tipo II, los cambios producidos en el gen SMAD3 causan el tipo III, mutaciones en el gen TGFB2 causan el tipo IV y las mutaciones en TGFB3 causan el tipo V. Estos cinco genes juegan un papel importante en una vía de señalización celular llamada *vía del factor de crecimiento transformante beta* (TGF- β), que dirige las funciones de las células del cuerpo durante el crecimiento y desarrollo [39].

Las mutaciones en el gen TGFBR1, TGFBR2, SMAD3, TGFB2 o TGFB3 dan como resultado la producción de una proteína con función reducida causando así la patología [40].

Se desconoce la prevalencia del síndrome de Loeys-Dietz, sin embargo, el síndrome de tipo I y II parecen ser las formas más comunes [39].

El síndrome de Loeys-Dietz tiene un patrón de herencia autosómico dominante. En aproximadamente el 75% de los casos, este trastorno es el resultado de una nueva mutación genética y ocurre sin antecedentes del trastorno en la familia. En otros casos, una persona afectada hereda la mutación de uno de los padres afectados [39].

3.2.5 Miocardiopatía hipertrófica familiar (Primary familial hypertrophic cardiomyopathy)

La miocardiopatía hipertrófica familiar (OMIM: PS192600) es una condición cardíaca caracterizada por engrosamiento (hipertrofia) del músculo cardíaco. A menudo esta afección comienza en la adolescencia o en la edad adulta, aunque puede desarrollarse en cualquier momento de la vida. Los síntomas son variables incluso entre la misma familia. Muchas personas afectadas no tienen síntomas. Un número pequeño de individuos afectados desarrollan insuficiencia cardíaca potencialmente mortal que puede requerir trasplante de corazón [41].

Los sinónimos encontrados en MedGen asociados a la miocardiopatía hipertrófica familiar son los siguientes:

- *Asymmetric septal hypertrophy.*
- *Familial Hypertrophic Cardiomyopathy.*
- *HCM.*
- *Hereditary ventricular hypertrophy.*
- *Hypertrophic cardiomyopathy.*
- *Idiopathic hypertrophic subaortic stenosis.*

Los genes más comúnmente involucrados son MYH7, MYBPC3, TNNT2 y TNNI3. Otros genes, algunos de los cuales no han sido identificados, también pueden estar involucrado en esta afección [41].

La miocardiopatía hipertrófica familiar afecta a aproximadamente 1 de cada 500 personas en todo el mundo. Es la enfermedad cardíaca genética más común en los Estados Unidos. Esta condición se hereda siguiendo un patrón autosómico dominante. En raras ocasiones, ambas copias del gen están alteradas, lo que conduce a signos y síntomas más graves. En la mayoría de los casos, una persona afectada tiene un padre con la afección [41].

Se encuentran más de 20 subcategorías [42] dentro de la miocardiopatía hipertrófica familiar, como la miocardiopatía hipertrófica tipo 1, tipo 2 y tipo3, entre otras (ver Anexo 2).

3.2.6 Síndrome de Brugada (*Brugada syndrome*)

El síndrome de Brugada (OMIM: 601144) es una condición que causa una interrupción del ritmo normal del corazón. Si no se trata, los latidos irregulares del corazón pueden causar desmayos, convulsiones, dificultad para respirar o muerte súbita. Estas complicaciones generalmente ocurren cuando una persona afectada está descansando o durmiendo. Este síndrome generalmente se manifiesta en la edad adulta, aunque puede desarrollarse en cualquier momento de la vida [43].

Los sinónimos asociados al síndrome de Brugada según MedGen son:

- *Sudden Unexplained Death Syndrome.*
- *Sudden unexplained nocturnal death syndrome.*

El síndrome de Brugada puede ser causado por mutaciones en uno de varios genes. El gen mutado más común es SCN5A, que se altera en aproximadamente el 30 por ciento de los casos. Las mutaciones en SCN3B, GPD1L, HCN4, KCNE3, SCN1B, CACNB2 o CACNA1C también pueden causar este síndrome [43].

Se desconoce la prevalencia exacta del síndrome de Brugada, aunque se estima que afecta a 5 de cada 10.000 personas a nivel mundial. Esta condición ocurre con más frecuencia en personas con ascendencia asiática, particularmente en poblaciones japonesas y del sudeste asiático. Aunque el síndrome de Brugada afecta tanto a hombres como a mujeres, la condición parece ser de 8 a 10 veces más común en hombres. Los investigadores sospechan que la testosterona, una hormona sexual presente en niveles mucho más altos en los hombres, puede explicar esta diferencia [43].

Las subcategorías según MedGen asociadas al síndrome de Brugada son:

- *Brugada syndrome 1 (BRGDA1).*
- *Brugada syndrome 2 (BRGDA2).*
- *Brugada syndrome 3 (BRGDA3).*
- *Brugada syndrome 4 (BRGDA4).*
- *Brugada syndrome 5 (BRGDA5).*
- *Brugada syndrome 6 (BRGDA6).*
- *Brugada syndrome 7 (BRGDA7).*
- *Brugada syndrome 8 (BRGDA8).*

Esta condición se hereda siguiendo un patrón autosómico dominante.

3.2.7 Síndrome de Marfan (*Marfan syndrome*)

El síndrome de Marfan (OMIM: 154700) es un trastorno sistémico del tejido conectivo con un alto grado de variabilidad clínica, comprende un amplio espectro fenotípico que varía desde leve (características del síndrome de Marfan en uno o pocos sistemas) hasta enfermedad multiorgánica neonatal grave y rápidamente progresiva. Las manifestaciones cardinales involucran los sistemas ocular, esquelético y cardiovascular. La mayor morbilidad³ y mortalidad⁴ [44] temprana en el síndrome de Marfan se relacionan con el sistema cardiovascular e incluyen la dilatación de la aorta. Con un manejo adecuado, la esperanza de vida de alguien con síndrome de Marfan se aproxima a la de la población general [45].

Los sinónimos del síndrome de Marfan según MedGen son:

- *FBN1-Related Thoracic Aortic Aneurysms and Aortic Dissections.*
- *Marfan syndrome classic.*
- *Marfan syndrome 1.*

³ La morbilidad se define como el número de personas que se enferman en una población y período determinados.

⁴ La mortalidad se define como el número de defunciones en una población y períodos determinados.

- *Marfanoid hypermobility syndrome.*

En los estudios que se han realizado, las mutaciones en el gen FBN1 causan el síndrome de Marfan. La incidencia del síndrome de Marfan es de aproximadamente 1 de cada 5.000 en todo el mundo [45]. El síndrome descrito se hereda en un patrón autosómico dominante. Al menos el 25% de los casos de síndrome de Marfan son el resultado de una nueva mutación en el gen FBN1. Estos casos ocurren en personas sin antecedentes del trastorno en su familia [45].

3.2.8 Taquicardia ventricular catecolaminérgica polimórfica (*Catecholaminergic polymorphic ventricular tachycardia (CPVT)*)

La taquicardia ventricular polimórfica catecolaminérgica (TVCP) (OMIM: 180902; 604772) es una condición que se caracteriza por un ritmo cardíaco anormal (arritmia). A medida que la frecuencia cardíaca aumenta en respuesta a la actividad física o al estrés emocional, puede desencadenar un latido cardíaco anormalmente rápido e irregular llamado taquicardia ventricular (Figura 2). Los episodios de taquicardia ventricular pueden causar mareos y desmayos. Las personas que padecen TVCP suelen comenzar con estos episodios en la infancia. Si no se reconoce ni se trata la TVCP, un episodio de taquicardia ventricular puede hacer que el corazón deje de latir (paro cardíaco) y provoque la muerte súbita. Los investigadores sospechan que la TVCP puede ser una causa importante de muerte súbita en niños y adultos jóvenes sin anomalías cardíacas reconocidas [46].

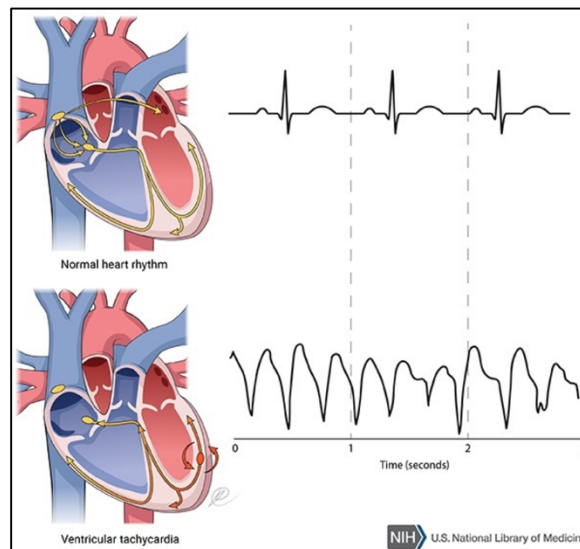


Figura 2. Latidos cardíacos normal y anormal [46].

Los sinónimos asociados según MedGen son:

- *Catecholamine-induced polymorphic ventricular tachycardia.*
- *Familial polymorphic ventricular tachycardia.*

La TVCP puede ser resultado de mutaciones en dos genes, el RYR2 y CASQ2. Las mutaciones en el gen RYR2 causan aproximadamente la mitad de todos los casos, mientras que las mutaciones en el gen CASQ2 representan del 1% al 2% de los casos.

Se estima que la prevalencia de TVCP es de aproximadamente 1 de cada 10.000 personas [46]. Sin embargo, la verdadera prevalencia de esta condición es desconocida.

Las subcategorías asociadas a TVCP son:

- *Ventricular tachycardia catecholaminergic polymorphic 1 (CPVT1).*
- *Ventricular tachycardia catecholaminergic polymorphic with or without atrial dysfunction and/or dilated cardiomyopathy.*

Cuando la TVCP resulta de mutaciones en el gen RYR2, tiene un patrón de herencia autosómico dominante, mientras que la TVCP que es causada por mutaciones en el gen CASQ2, tiene un patrón de herencia autosómico recesivo [46].

3.2.9 Síndrome de QT largo (*Long QT syndrome*)

Este síndrome (OMIM: PS192500) se caracteriza por episodios de desmayo (síncopes) y un grado variable de arritmia ventricular como lo indica el intervalo QT prolongado. Las formas heredadas son causadas por la mutación de genes que codifican las proteínas del canal de iones cardíacos. Este síndrome es una afección cardíaca que hace que el músculo cardíaco tarde más de lo normal en recargarse entre latidos. Con poca frecuencia, los latidos irregulares del corazón pueden causar desmayos (síncope), y aún más raramente, muerte súbita [32].

Se ha encontrado un único sinónimo asociado en MedGen: *LQTS (Long QT syndrome)*.

Los genes asociados al síndrome de QT largo según [47] son: KCNH2, KCNQ1, SCN5A, ANK2, AKAP9, CACNA1C, CALM1, CALM2, CAV3, KCNE1, KCNE2, KCNJ2, KCNJ5, SCN4B y SNTA1.

La prevalencia de LQTS ha sido estimada en 1 de 2.500 y generalmente se hereda de manera autosómica dominante [47].

Las subcategorías asociadas al LQTS son:

- *Long QT syndrome 1 (LQTS1).*
- *Long QT syndrome 1, acquired, susceptibility to.*
- *Long QT syndrome 2 (LQTS2).*
- *Long QT syndrome 2, acquired, susceptibility to.*
- *Long QT syndrome 3 (LQTS3).*
- *Long QT syndrome 4 (LQTS4).*
- *Long QT syndrome 5 (LQTS5).*
- *Long QT syndrome 6 (LQTS6).*
- *Long QT syndrome 7 (LQTS7).*
- *Long QT syndrome 8 (LQTS8).*
- *Long QT syndrome 9 (LQTS9).*
- *Long QT syndrome 10 (LQTS10).*
- *Long QT syndrome 11 (LQTS11).*
- *Long QT syndrome 12 (LQTS12).*
- *Long QT syndrome 13 (LQTS13).*
- *Long QT syndrome 14 (LQTS14).*
- *Long QT syndrome 15 (LQTS15).*
- *Long QT syndrome, acquired, reduced susceptibility to.*
- *Cardiac arrhythmia, ankyrin-B-related.*

- *Atrial fibrillation 17.*
- *Andersen syndrome.*

Tras describir los fenotipos a analizar en esta tesis de máster, se encuentra que la mayoría siguen un patrón de herencia **autosómico dominante**. En algunos casos raros en la miocardiopatía dilatada familiar y TVPC se produce un patrón de herencia recesivo. Además, todas las afecciones descritas son enfermedades **monogénicas**, es decir, únicamente sería necesaria una mutación en un solo gen para que se produzca la enfermedad.

Las patologías descritas anteriormente no tienen identificados todos los genes donde se producen las variantes que causan la enfermedad ni tampoco todas las variantes que causan la afección en los genes asociados conocidos.

Una vez definido el método SILE y los fenotipos asociados a la muerte súbita en el capítulo siguiente se procede a la aplicación del método.

CAPÍTULO 4. APLICACIÓN DEL MÉTODO SILE

En este capítulo se aplicará el método SILE con el objetivo de obtener las variantes genómicas más relevantes asociadas a la muerte súbita cardíaca.

4.1. BÚSQUEDA

Esta etapa del método SILE está pensada para la selección de repositorios genéticos que nos aporten información de la cual se obtenga datos de calidad que se puedan utilizar en el entorno clínico.

A la hora de buscar información se requiere seleccionar y analizar bases de datos para obtener las más adecuadas para adquirir las variantes a clasificar del fenotipo que se estudie. Dado que estamos trabajando con enfermedades cardíacas lo ideal sería utilizar bases de datos que se especialicen en la genética de las enfermedades del corazón. La búsqueda de repositorios se ha realizado en *Health Sciences Library Systems(HSLS)* [48], *Human Genome Variation Society* [49], *Nucleic Acids Research (NAR)* [50], y Google.

HSLS es una librería creada por la Universidad de Pittsburgh que ofrece una amplia gama de servicios de información y recursos en formato impreso y electrónico para profesores, estudiantes e investigadores de ciencias de la salud. Una de las tareas que hace es recolectar repositorios genómicos y herramientas de análisis clasificadas en:

1. *Comparative genomics* (66).
2. *General genomics databases and tools* (69).
3. *Genome annotation terms, ontologies, nomenclature, and classification* (49).
4. *Genome browsers, genome annotation, genomic sequence analysis* (47).
5. *Human genome databases, maps, and viewers* (41).
6. *Non-human vertebrates model organisms genomic databases* (53).
7. *Non-vertebrates model organisms genomic databases* (311).

En esta tesis resultan de interés los repositorios de las secciones *General genomics databases and tools* y *Human genome databases, maps, and viewers*.

HGVS gestiona un gran catálogo de bases de datos con el objetivo de promover la recopilación, documentación y distribución gratuita de información sobre variantes genómicas. Las bases biológicas están clasificadas en 12 secciones de las cuales las de interés son:

- *Disease Centered Central Mutation Databases*(10)
- *Central Mutation & SNP Databases*(21).

El NAR es una institución pública que difunde los resultados de investigaciones de vanguardia en aspectos físicos, químicos, bioquímicos y biológicos de ácidos nucleicos y proteínas involucradas en el metabolismo y/o interacciones de ácidos nucleicos. Además, esta institución se encarga de recolectar información de bases de datos biológicas que las clasifica en 15 secciones diferentes. La sección que resulta de interés en esta tesis es la que contiene bases de datos biológicas que albergan información de genes y enfermedades, esta sección a su vez se clasifica en:

- *General human genetics databases*(22).

- *General polymorphism databases(44).*
- *Cancer gene databases (51).*
- *Gene-, system- or disease-specific databases (58).*

De estas cuatro subsecciones son de interés todas menos la subsección que recoge información sobre genes que causan cáncer porque se está estudiando enfermedades cardíacas.

El primer paso para gestionar la calidad de la información es asegurarse de que las fuentes son fiables. En [51], se propone una metodología de calidad de datos específico para el dominio genómico a fin de asegurar la selección de repositorios de alta calidad que proporcionen a su vez datos valiosos. La metodología de calidad se divide en 5 fases:

- Descripción de dimensiones: describe las dimensiones de interés para ser medidas y su ámbito.
- Descripción métrica: describe la métrica asociada a cada dimensión.
- Selección de variables: selecciona las variables del MCHG donde las métricas van a ser aplicadas.
- Requisitos mínimos de DQ: establece los requisitos mínimos que las variables deben cumplir.
- Evaluación de DQ: Comparar la información de las bases de datos con los requisitos mínimos de DQ.

Todas estas fases están asociadas a una serie de métricas que se describen en la Tabla 1.

Credibilidad	M1	La información almacenada debe ser manualmente curada o revisada por expertos
	M2	Hay controles de calidad para asegurar la exactitud de la información presentada.
Relevancia	M3	La información contiene suficiente información y útil para para determinar los datos requeridos de acuerdo con los atributos del CSHG.
Reputación	M4	La base de datos debe ser mantenida por centros de investigación, instituciones o asociaciones de nivel internacional o muy conocidos.
Actualidad	M5	La base de datos tiene que estar activa y actualizada frecuentemente.
Accesibilidad	M6	La información deber ser pública y de libre acceso.
	M7	La base de datos debe proporcionar mecanismos de descarga de la información buscada.
	M8	Es muy recomendable que la base de datos permita el acceso a la información almacenada mediante herramientas informáticas.

Tabla 1. Descripción de las métricas utilizadas para la selección de las bases de datos.

Un análisis de las fuentes de datos disponibles ha permitido identificar como relevantes para este trabajo: Ensembl, GWAS Catalog, Varsome, DbSNP y SNPedia. El resto de opciones quedan descartadas por tratarse de repositorios de enfermedades específicas no cardíacas (cáncer, Alzheimer, Autismo...), de herramientas informáticas, urls no disponibles, etc.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca: aplicaciones a la medicina de precisión

De Clinvar, Ensembl y GWAS Catalog se sacará el conjunto de datos a analizar mientras que Varsome, dbSNP y SNPedia se utilizarán para completar la información que falte.

A continuación, se da una breve descripción de los repositorios que se utilizan.

4.1.1 Varsome

Varsome [52] es un conjunto de soluciones bioinformáticas potentes pero intuitivas para los datos de secuenciación de ADN de próxima generación, tanto para uso clínico como para investigación. El motor de búsqueda Varsome (Figura 3) es de libre acceso, con una base de conocimientos ampliamente reconocida y dirigida por la comunidad que permite consultas flexibles a través de más de 30 recursos de datos genéticos y genómicos.

Varsome Pro y *Varsome Clinical* son ediciones profesionales de Varsome con una potente funcionalidad y herramientas de análisis y extracción de datos más sofisticadas. Si bien *Varsome Pro* sirve principalmente a investigadores, *Varsome Clinical* es una plataforma clínicamente acreditada que permite el descubrimiento, anotación e interpretación de variantes de forma rápida y precisa de los datos de NGS para genomas completos, exomas y paneles genéticos, lo que ayuda a los médicos a realizar diagnósticos más rápidos y precisos y decisiones de tratamiento para condiciones genéticas.

Adquiere información de: Clinvar, dbSNP, gnomAD, HPO, Ensembl, RefSeq, GWAS, CGD, HGNC, UniGene, Orphanet, CIViC genes, GERP, dbNSFP, COSMIC, IARC TP53, ICGC, Kaviar, DANN scores, CIViC mutations, Uniprot variants, Uniprot domains, GHR, CPIC, DGV, DECIPHER, ExAC CNVs, ExAC genes, PanelApp, Mondo, PMKB.

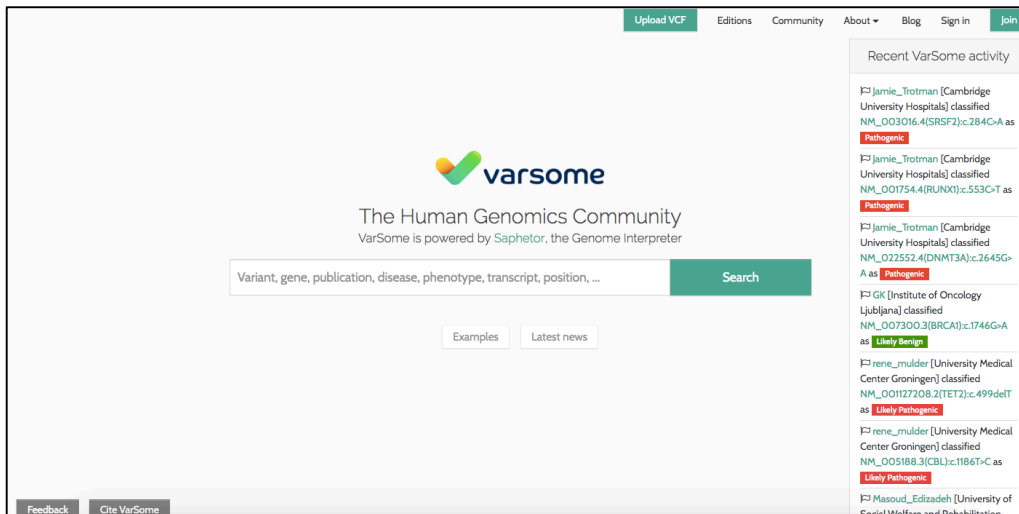


Figura 3. Varsome(Website).

Acceso a la base de datos:

- Usando la API de Varsome se puede obtener la información en formato JSON.
- A través de su buscador se puede buscar información usando el nombre de un fenotipo, enfermedad, variante, gen, publicación, identificador, región genómica o HGVS. Se puede consultar la información, pero no se puede descargar.

El acceso a los datos genómicos se hará utilizando el navegador de Varsome.

4.1.2 DbSNP

DbSNP [53] es un repositorio genómico de dominio público perteneciente al *National Center of Biotechnology Information (NCBI)*. Contiene variantes de un solo nucleótido, junto con la publicación, la frecuencia de población, consecuencia molecular e información de mapeo genómico y además variantes comunes y variantes clínicas (Figura 4).

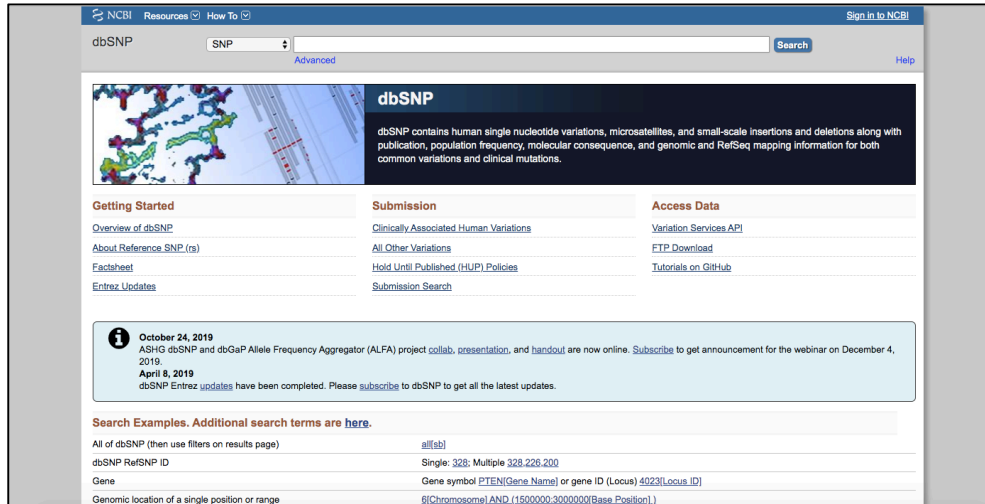


Figura 4. DbSNP (Website).

Acceso a la base de datos:

- A través del navegador mediante identificadores rs de la variante.
- Utilizando la API Entrez obteniendo los resultados en formato VCF y JSON.

En esta base de datos se consulta la información mediante los identificadores rs y se utilizará en esta tesis para completar la información de las variantes.

4.1.3 SNPedia

SNPedia [54] es un recurso wiki que se encarga de investigar la genética humana. Comparte información sobre las variantes citando la documentación científica revisada (Figura 5).

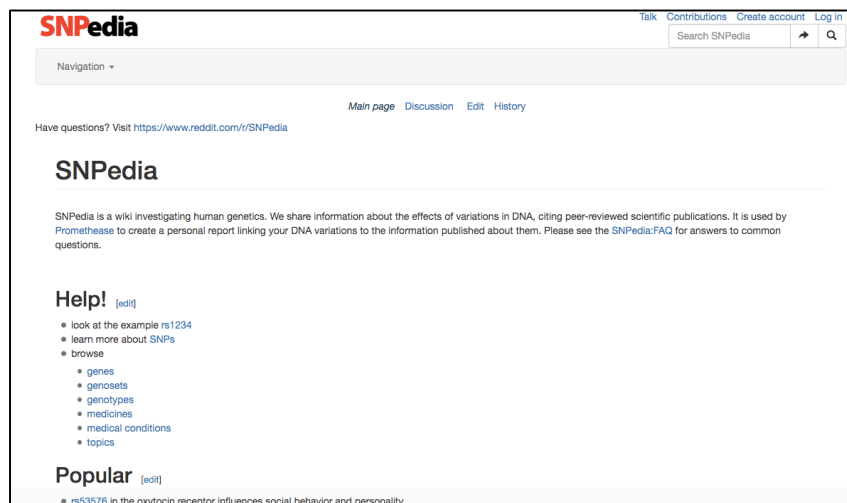


Figura 5. SNPedia (Website).

Acceso a la base de datos:

- Utilizando el navegador se puede obtener la información de la variante utilizando identificadores rs.
- Mediante bots⁵ programados en Perl, Python, Ruby o R/Bioconductor.

La consulta de la información en este TFM se hace a través del navegador.

4.1.4 Clinvar

Clinvar [55] es una base de datos de acceso gratuito de informes sobre las relaciones entre las variantes de los humanos y los fenotipos, con evidencia de apoyo (Figura 6). Esta base de datos facilita el acceso y la comunicación sobre las relaciones establecidas entre la variante humana y el estado de salud observado, y la historia de esa interpretación. Clinvar procesa las variantes encontradas en las muestras de pacientes, las afirmaciones hechas con respecto a su importancia clínica, la información sobre el remitente y otros datos de respaldo [56]. Esta base de datos pública es soportada por el *National Center of Biotechnology Information* (NCBI).

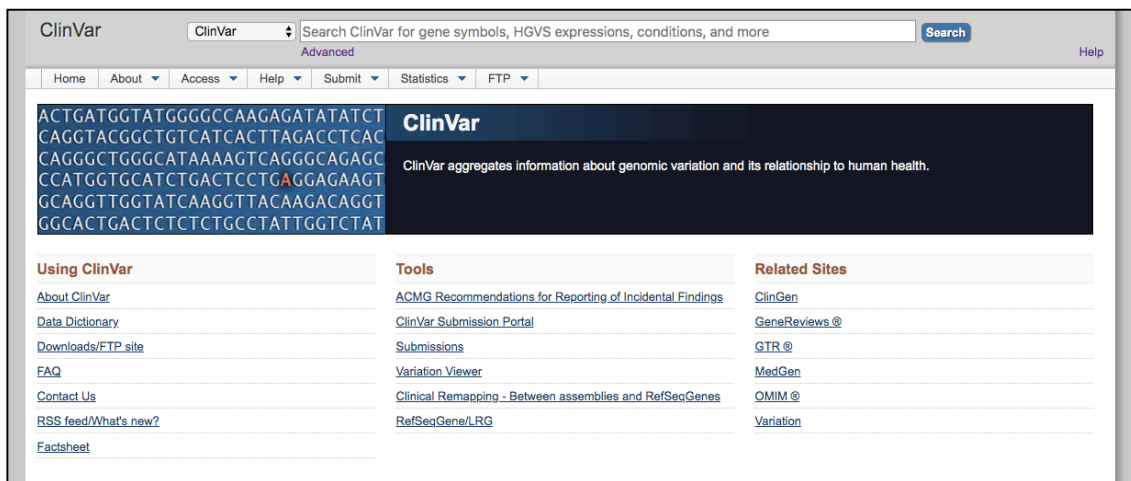


Figura 6. Clinvar (Website).

Acceso a la base de datos:

El acceso se puede realizar a través de las siguientes formas:

- A través del propio navegador (Figura 6) se puede obtener la información mediante consultas simples o más específicas usando su buscador avanzado. El buscador avanzado permite utilizar los operadores AND y OR, además de hacer búsquedas utilizando el nombre del fenotipo/enfermedad, identificador de la enfermedad, nombre de la variante, etc. Se obtiene los datos en formato tabular o una lista de indicadores de cada variante. Los datos pueden ser descargados del navegador en los siguientes formatos: XML, VCF y TSV.
- Utilizando API's. Una API es una interfaz de programación de aplicaciones, es decir, una interfaz que permite la conexión entre dos sistemas. Clinvar tiene la API E-utilities y Entrez. Los datos se obtienen con los formatos JSON o XML y es más completa que con el buscador propio de Clinvar.

⁵ Un bot es un programa informático que realiza tareas repetitivas a través de internet.

- Los datos además pueden ser accedidos a través de un servidor ftp disponible. Los formatos en el que se obtienen los datos son XML y VCF.

La información de este repositorio se obtiene utilizando el navegador. Para buscar las variantes de una enfermedad o fenotipo particular se ingresa el nombre de la enfermedad o fenotipo seguido de [dis], por ejemplo, “*familial restrictive cardiomyopathy*”[dis] (Figura 7).

Variation Location	Gene(s)	Protein change	Condition(s)	Clinical significance (Last reviewed)	Review status	Accession
1. TNNT2_3-BP DEL_285GGA	TNNT2		Familial restrictive cardiomyopathy 3	Pathogenic (Jan 25, 2008)	no assertion criteria provided	VCV000012418
2. NM_000364.4(TNNT2):c.565T>G (p.Ser189Ala) GRCh37: Chr1:201332459 GRCh38: Chr1:201363331	TNNT2	S179A, S189A, S149A, S174A	Familial hypertrophic cardiomyopathy 2, Familial restrictive cardiomyopathy 3, Left ventricular noncompaction 6	Uncertain significance (Apr 17, 2017)	criteria provided, single submitter	VCV000469526
3. NM_000364.4(TNNT2):c.*66G>A GRCh37: Chr1:201328272 GRCh38: Chr1:201359144	TNNT2		Familial restrictive cardiomyopathy, Left ventricular noncompaction cardiomyopathy, Hypertrophic cardiomyopathy, Dilated Cardiomyopathy, Dominant	Likely benign (Jun 14, 2016)	criteria provided, single submitter	VCV000294893
4. NM_000364.4(TNNT2):c.884A>C (p.Lys285Thr) GRCh37: Chr1:201328342 GRCh38: Chr1:201359214	TNNT2	K288T, K295T, K282T, K255T, K285T, K298T	Familial hypertrophic cardiomyopathy 2, Left ventricular noncompaction 6, Familial restrictive cardiomyopathy 3	Uncertain significance (Dec 8, 2016)	criteria provided, single submitter	VCV000404397

Figura 7. Ejemplo de la búsqueda de variantes en Clinvar.

El navegador da como resultado, en este caso, una tabla que contiene 296 variantes que han sido descargadas en un archivo con formato XML. La búsqueda proporciona para cada variante (*Variation*) el gen o genes donde se produce (*Gen*), el cambio proteico (*Protein change*), la patología asociada (*Condition*), el significado clínico (*Clinical significance*), si ha sido revisada o no por un experto para respaldar su importancia clínica (*Review Status*) y un identificador propio de Clinvar para identificar las variantes e ir añadiendo información a nivel interno (*Accession*). Las variantes vienen identificadas por un identificador HGVS, por ejemplo, NM_000059.3(BRCA2):c.5946del(p.Ser1982fs). La nomenclatura HGVS se utiliza como un estándar internacional para informar e intercambiar información respecto a las variantes encontradas en el ADN, ARN y secuencias de proteínas [57]. Si se accede a cada una de las variantes a través del navegador también viene con un identificador que empieza por rs seguido de una secuencia de números, por ejemplo, rs80359550. Tanto el identificador rs como el HGVS están conectados, es decir, se encuentra la misma información de una variante si se utiliza alguno de los dos identificadores. En algunos casos puede ser que únicamente esté disponible el identificador HGVS, por lo tanto, si no se encuentra una variante con el identificador rs hay que buscarla con el HGVS.

4.1.5 Ensembl

Ensembl [58] es un navegador web para genomas de vertebrados, entre ellos el del ser humano, que apoya la investigación en genética comparativa, evolución, variación de secuencias y regulación transcripcional. Ensembl anota genes, calcula múltiples alineaciones, predice la función reguladora y recopila datos de enfermedades. Las herramientas de Ensembl incluyen BLAST, BLAT, BioMart y el Variant Effect Predictor (VEP) para todas las especies

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca: aplicaciones a la medicina de precisión

compatibles. BioMart es una herramienta fácil de usar basada en la web que permite la extracción de datos sin ningún conocimiento de programación o comprensión de la estructura de la base de datos. El dataset, los filtros y los atributos de BioMART permiten hacer búsquedas más concretas según lo que nos interese (Figura 8).

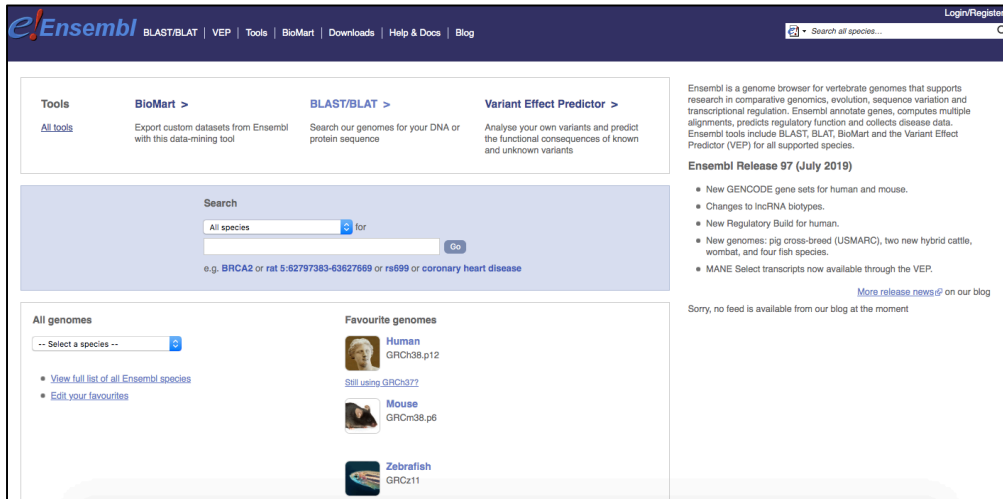


Figura 8. Ensembl (Website).

Acceso a la base de datos:

El acceso a los datos de la herramienta Biomart de Ensembl, que es la que se va a utilizar, se puede hacer a través de:

- *BioMart* Bioc R package: es una forma fácil y rápida de acceder a BioMart directamente desde un terminal de software R.
- *BioMart* Perl API permite integrar el Código de BioMart Perl en scripts personalizados de Perl.
- *BioMart* RESTful es una forma rápida y fácil de consultar los conjuntos de Ensembl usando Wget o Perl.
- El acceso a la información también se puede realizar a través de la aplicación web de forma manual, pudiendo enviarlos a través de un correo electrónico o descargándolos en un zip. Los datos se pueden obtener con los formatos: HTML, CSV, TSV, y XLS.

La búsqueda de variantes se va a realizar utilizando la aplicación web de *BioMart* (Figura 9).

Variant name	Variant source	Clinical significance	Phenotype name	Associated gene with phenotype	PubMed ID
rs137854606	dbSNP	not provided,pathogenic	SCN5A	SCN5A	22581833
rs137854608	dbSNP	not provided,pathogenic	SCN5A	SCN5A	11821428
rs137854609	dbSNP	not provided,pathogenic	SCN5A	SCN5A	11284011
rs16965924	dbSNP	benign,likely benign	SCN1B	SCN1B	25741868
rs16965930	dbSNP	benign,likely benign	SCN1B	SCN1B	17268157
rs16965933	dbSNP	benign,likely benign	SCN1B	SCN1B	25741868
rs66871189	dbSNP	benign,likely benign	SCN1B	SCN1B	20226894
rs72652027	dbSNP	benign,likely benign	SCN1B	SCN1B	25741868
rs12658361	dbSNP	benign	SCN1B	SCN1B	25741868
rs12658361	dbSNP	benign	SCN1B	SCN1B	25741868

Figura 9. Ejemplo de búsqueda en Ensembl.

Para obtener los datos se selecciona como base de datos *Ensembl Variation 96* y como *dataset Human Short Variants*. Además, *BioMart* permite especificar la información que va a devolver como resultado (*Attributes*) y los criterios de búsqueda (*Filters*).

En los criterios de búsqueda existe un apartado donde se selecciona el fenotipo del cual se requiere las variantes. Como ya se mencionó en el punto 3.2. FENOTIPOS ESTUDIADOS, cada fenotipo tiene varios sinónimos o subcategorías. Para no perder información cuando se realiza la consulta de variantes se tiene en cuenta todas estas formas de referirse a un mismo fenotipo. Los atributos permiten seleccionar la información que quiero obtener de la variante, por ejemplo, con que identificador se quiere tener los resultados (rs o HGVS), de que repositorio procede, el fenotipo al que pertenece, el significado clínico, el gen que está asociado, el identificador de la bibliografía asociada, entre otros.

4.1.6 GWAS Catalog

Esta base de datos (Figura 10) contiene un gran número de estudios de asociación de genoma publicados (GWAS). Estos estudios brindan una oportunidad para investigar el impacto de variantes comunes en enfermedades complejas. Proporciona una base de datos consistente, de búsqueda, visualizable, y disponible de forma gratuita de asociaciones genotipo/ fenotipo publicadas, que pueden integrarse fácilmente con otros recursos, y que es accesible por científicos, clínicos y otros usuarios de todo el mundo. Esta base de datos está respaldada por el *National Human Genome Research Institute* (NHGRI) y trabaja en conjunto con SPOT y Ensembl [59].

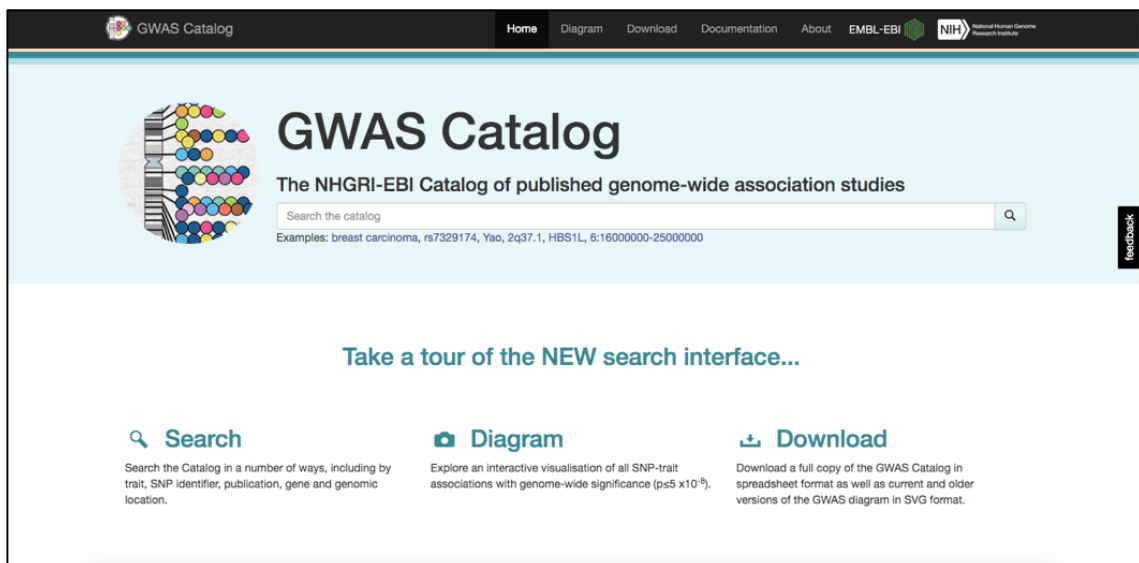


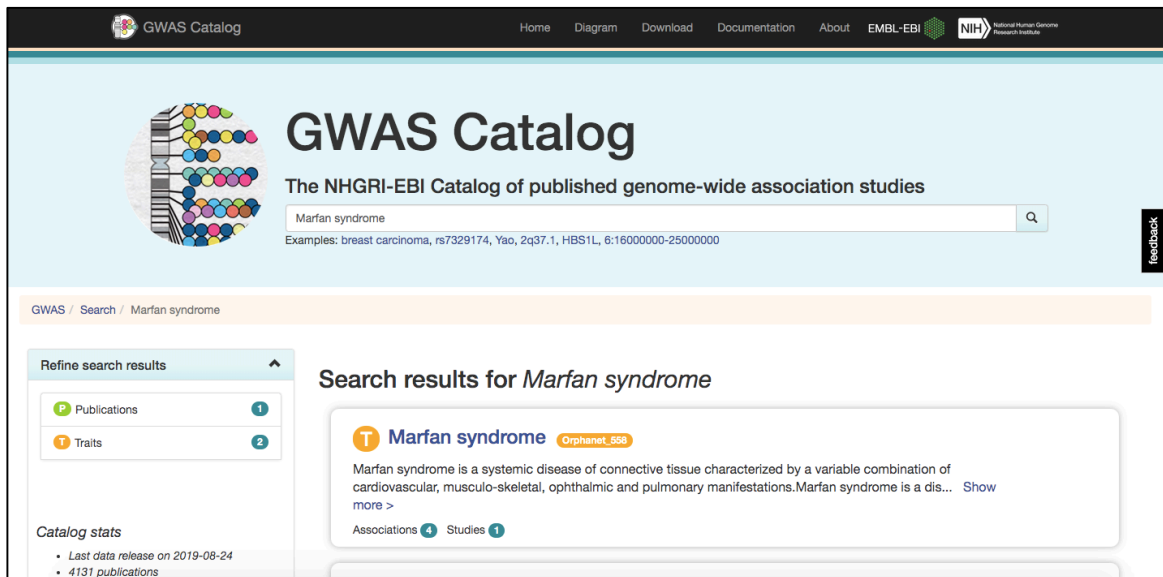
Figura 10. GWAS Catalog (website).

Acceso a la base de datos:

- Utilizando una API en GWAS Catalog se puede acceder a la documentación, estudios, variantes, asociaciones, datos estadísticos, etc.
- A través de un buscador básico que permite acceder a la información mediante: el fenotipo, los identificadores SNP, las publicaciones, el gen y la localización genómica. La información se obtiene en formato TSV, OWL/RDF.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca: aplicaciones a la medicina de precisión

La búsqueda de variantes en GWAS Catalog se realiza utilizando el nombre fenotipo y se obtiene como resultado una lista de variantes (*Associations*) y una lista de artículos (*Studies*). En la Figura 11 se puede observar un ejemplo de búsqueda en esta base de datos.



The screenshot displays the GWAS Catalog interface. At the top, there is a navigation bar with links for Home, Diagram, Download, Documentation, About, EMBL-EBI, and NIH. The main header features the GWAS Catalog logo and the text 'The NHGRI-EBI Catalog of published genome-wide association studies'. A search bar contains the text 'Marfan syndrome' and a search icon. Below the search bar, there are example search terms: 'breast carcinoma, rs7329174, Yao, 2q37.1, HBS1L, 6:16000000-25000000'. The search results section is titled 'Search results for Marfan syndrome' and includes a 'Refine search results' sidebar with 'Publications' (1) and 'Traits' (2). The main result for 'Marfan syndrome' is highlighted, with a description: 'Marfan syndrome is a systemic disease of connective tissue characterized by a variable combination of cardiovascular, musculo-skeletal, ophthalmic and pulmonary manifestations. Marfan syndrome is a dis... Show more >'. It also shows 'Associations 1' and 'Studies 1'. A 'Catalog stats' section at the bottom left indicates 'Last data release on 2019-08-24' and '4131 publications'.

Figura 11. Ejemplo de búsqueda en GWAS Catalog.

Una vez entendidos los fenotipos, el método y las bases de datos se procede a extraer toda la información almacenada en los bancos genómicos a partir de las cuales se iniciará la identificación que corresponde con la siguiente etapa del método.

En cada base de datos se realiza la búsqueda y obtención de los datos de las variantes según lo explicado en el punto anterior.

En la Figura 12 se muestra un resumen del número de variantes que se obtienen para cada una de las enfermedades cardíacas y para la muerte súbita en los 3 repositorios.

Se encontraron variantes para todos los fenotipos estudiados en Clinvar y Ensembl, pero en GWAS Catalog únicamente se encontraron variantes genómicas para miocardiopatía dilatada, miocardiopatía hipertrófica, síndrome de Brugada, síndrome de Marfan y para la muerte súbita.

Tras analizar los datos, algunas variantes que están siendo estudiadas y cuya información ha sido obtenido en Clinvar han sido reportadas o evaluadas por un panel de expertos:

- Una variante para la miocardiopatía dilatada: rs397516187 asociada al gen MYH7.
- Una variante para la miocardiopatía restrictiva: rs397516153 asociada al gen MYH7.
- Treinta y cinco variantes (ver Anexo 3) para la miocardiopatía hipertrófica asociadas a los genes RAF1 y MYH7.

La evaluación por un panel de expertos de las variantes da soporte al significado clínico proporcionado dado que ha sido evaluado por expertos en la materia.

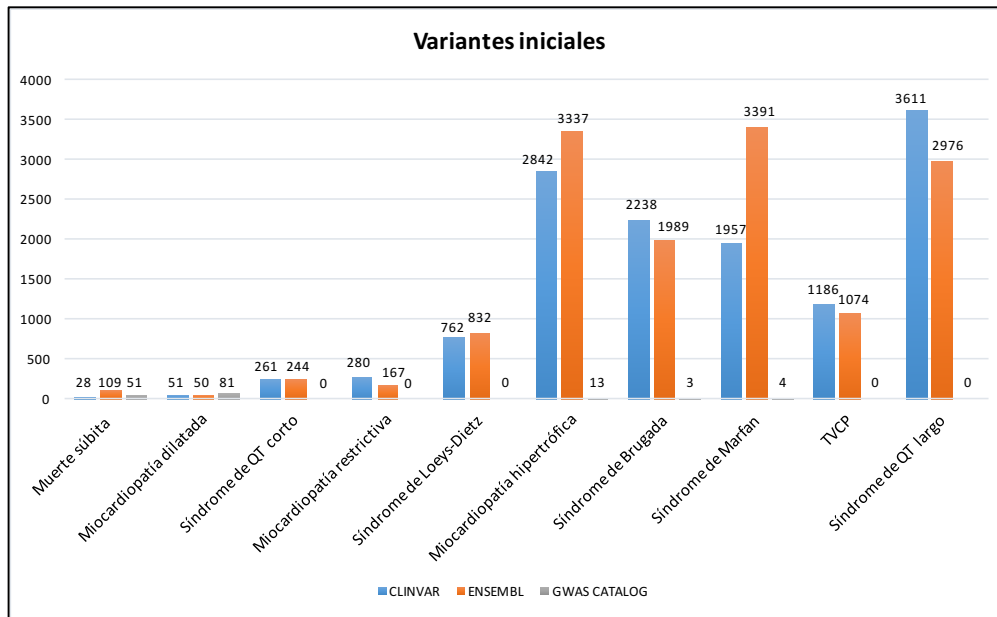


Figura 12. Número de variantes totales relacionadas con la muerte súbita y los fenotipos que la provocan obtenidas en tres repositorios.

Si se analiza la Figura 12, se observa que en general la miocardiopatía hipertrófica, el síndrome de Brugada, el síndrome de Marfan y el síndrome de QT largo son los fenotipos con el mayor número de variantes. Mientras que la miocardiopatía dilatada, el síndrome de QT corto y la miocardiopatía restrictiva y la muerte súbita son los que menos variantes a analizar presentan. Clinvar proporciona más variantes para el síndrome de QT largo, el TVPC, el síndrome de Brugada y el síndrome de QT corto en comparación con el resto de repositorios. Mientras que Ensembl da el mayor número de variantes para el síndrome de Marfan, la miocardiopatía hipertrófica, el síndrome de Loey-Dietz y la muerte súbita. Por último, GWAS Catalog ofrece el mayor número de variantes para la muerte súbita. Este repositorio no aporta información para 5 de los fenotipos estudiados: el síndrome de QT corto, la miocardiopatía restrictiva, el síndrome de Loey-Dietz, el TVCP y el síndrome que QT largo. En Clinvar y Ensembl, son los que mayor información proporcionan para el análisis de variantes.

Dada la heterogeneidad de los datos de los 3 repositorios, se decide analizar las variantes de forma manual por fenotipo y base de datos.

Tras la obtención de las variantes y un breve análisis de la información obtenida se procede a la siguiente etapa del método SILE: la identificación.

4.2 IDENTIFICACIÓN

Una vez seleccionados los repositorios y extraídos los datos de estos, empieza el análisis exhaustivo de la información para obtener un conjunto de variantes genómicas relevantes para su futura carga y explotación. Para ello se aplicará una serie de criterios de calidad establecidos por el método SILE [26] , ver Tabla 2.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

Precisión	M1	Revisar los atributos que puedan ser propensos a errores. Los errores sintácticos deben verificarse utilizando vocabularios controlados y diccionarios de datos específicos.
Complejidad	M2	La información requerida por el HGDB está presente.
Consistencia	M3	La información sobre las variantes es definida usando vocabularios estándar ontologías verificadas para determinar los atributos críticos como las expresiones HGVS, patogenicidad o efectos funcionales.
	M4	No debe haber conflicto en la interpretación clínica de cada variante.
	M5	No debe haber conflicto entre bases de datos relacionadas con las características estructurales de la variante.
Credibilidad	M6	Cada variante debe tener significado médico o consecuencias genealógicas y ser reproductibles.
	M7	La relación entre la variante y la enfermedad debe tener al menos un artículo publicado, un artículo revisado con datos estadísticos creíbles y de libre acceso.
Relevancia	M8	La frecuencia del alelo menor (MAF) de la variante debe ser menos que la frecuencia del fenotipo en la población.
	M9	El patrón de herencia, penetrancia y mecanismo de la variante debe ser consistente con la enfermedad.
	M10	Los estudios realizados en la bibliografía deben tener al menos 700 participantes y es deseable la replicación de estudios.
	M11	Para las variantes patogénicas el OR debe ser más grande de 1 y para las protectoras el OR debe ser menor que 1.
	M12	Para los estudios GWAS el p-valor debe ser menor que $5 \cdot 10^{-8}$.

Tabla 2. Criterios de calidad para la selección de variantes.

Los filtros a aplicar de forma secuencial para cumplir los criterios de calidad para la identificación de las variantes genómicas son los siguientes:

- F1: ¿Se especifica el significado clínico de la variante?
- F2: ¿Es el significado clínico distinto de “benign”, “likely benign”, “likely pathogenic”, “uncertain” o “conflicting interpretation”?
- F3: ¿Es el significado clínico de la variante igual a “conflicting interpretation”?
- F4: ¿Tiene la variante al menos un artículo asociado?
- F5: ¿Son relevantes los artículos asociados a la variante?
 - Estudio realizado en humanos.
 - Más de 700 participantes.
 - Evidencia estadística
 - Variantes patogénicas.
 - Variantes protectoras.
 - Estudio GWAS: p-valor menor que $5 \cdot 10^{-8}$.

- Deseable la replicación de estudios.
- F6: ¿Hay conflictos en la literatura asociada?
- F7: ¿Es el MAF de la variante menor o igual que la frecuencia del fenotipo de la variante?
- F8: ¿Es el patrón de herencia, penetrancia y mecanismo de la variante consistente con la enfermedad?

En la Figura 13 se puede ver el diagrama de flujo según el cual se han ido aplicando los filtros:

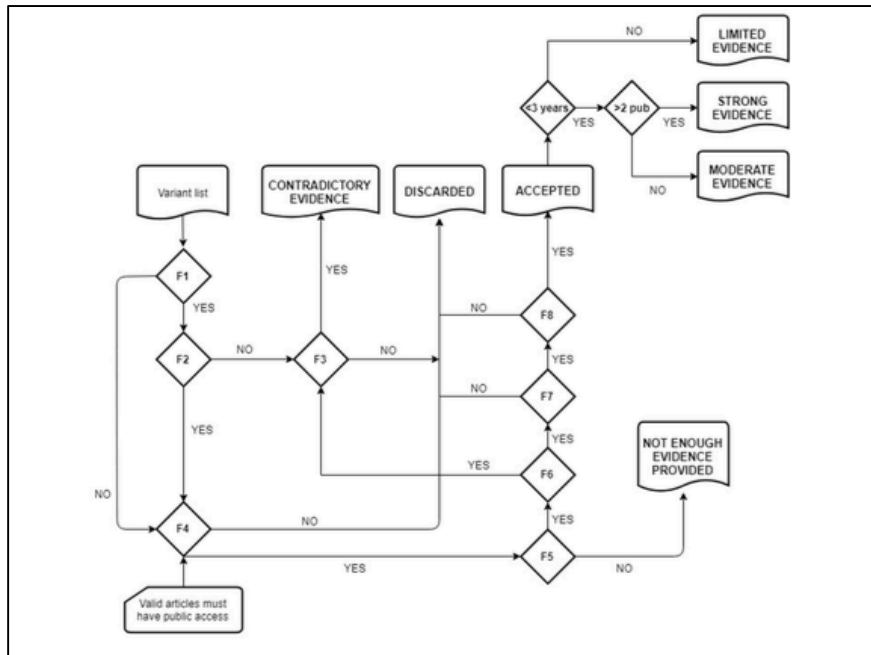


Figura 13. Flujo de trabajo para la clasificación de variantes [60].

A continuación, se analizan los resultados de aplicar los filtros descritos anteriormente por fenotipo y por repositorio genómico. Se empieza analizando los filtros F1-F4 dado que estos están relacionados con el significado clínico de la variante y se tiene que ver si tienen asociado como mínimo un artículo. Una vez acabado con estos se continuará con el resto de filtros.

En Clinvar aplicar los filtros F1- F3 es sencillo en comparación con el filtro F4. El filtro F4 es complejo de aplicar debido a que la documentación asociada no se encuentra en el archivo que se obtiene de forma manual del navegador de Clinvar. Para conseguir la documentación científica asociada a las variantes, 2.145 en el caso de Clinvar, se accede variante por variante en Clinvar para obtener los artículos relacionadas con estas.

En las Figura 14 y Figura 15 se muestra un resumen de los datos obtenidos tras aplicar el F4 en las variantes obtenidas de Clinvar. En las figuras, la columna azul representa el número de variantes que tienen al menos un artículo asociado y la columna naranja el número total de artículos asociados a las variantes.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca: aplicaciones a la medicina de precisión

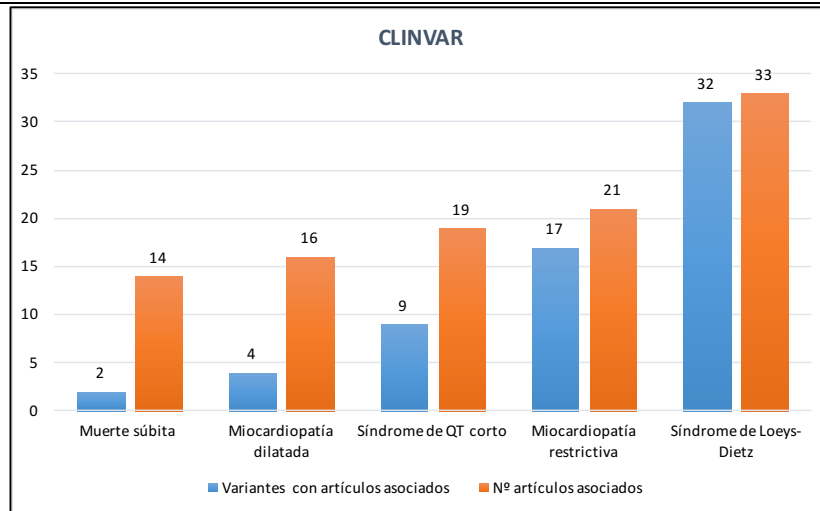


Figura 14. Primeros resultados en Clinvar tras aplicar los filtros F1-F4 del método SILE (parte 1).

En Clinvar tras aplicar los filtros F1, F2, F3 y F4 se obtiene 1.487 variantes y 1.835 artículos para continuar con su análisis siendo distribuidos de la siguiente forma:

- Muerte súbita: 2 variantes candidatas y 14 artículos.
- Miocardiomiopatía dilatada: 4 variantes candidatas y 16 artículos.
- Síndrome de QT corto: 9 variantes candidatas y 19 artículos.
- Miocardiomiopatía restrictiva: 17 variantes y 21 artículos.
- Síndrome de Loey-Dietz: 32 variantes y 33 artículos.
- Miocardiomiopatía hipertrófica: 234 variantes y 647 artículos.
- Síndrome de Brugada: 190 variantes y 132 artículos.
- Síndrome de Marfan: 205 variantes y 228 artículos.
- TVPC: 25 variantes y 55 artículos.
- Síndrome de QT largo: 769 variantes y 670 artículos.

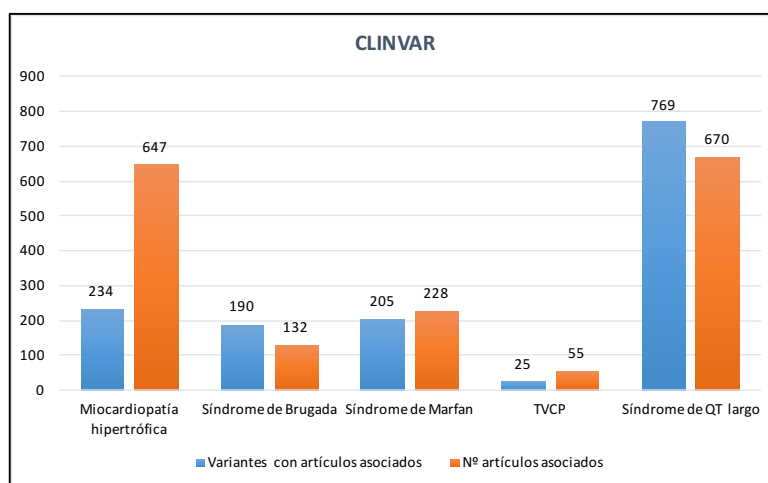


Figura 15. Primeros resultados en Clinvar tras aplicar los filtros F1-F4 del método SILE (parte 2).

De las variantes candidatas totales obtenidas de Clinvar tras superar el filtro F4, el 57,9% de las variantes tienen asignado como significado clínico "*pathogenic*", el 41,97 % variantes tienen significado clínico "*not provided*" correspondientes al síndrome de QT largo, síndrome de QT corto, síndrome de Brugada y a la muerte súbita. Además, el 0,13% tiene como significado clínico "*risk factor*" y pertenecen al síndrome de QT largo.

La mayor reducción de variantes se produce al aplicar los filtros F1-F3 (un 83,76%), lo que indica que la mayor parte de variantes genómicas estudiadas de este repositorio no tiene un significado clínico relevante. El filtro F4 ha reducido las variantes en un 69,32% del total de variantes obtenidas tras el filtro F1-F3, lo que significa que todas estas variantes no tienen documentación científica que les de soporte. Los filtros han permitido reducir el número de variantes iniciales en un 88,74%, obteniendo así un 11,26% de variantes con significado clínico relevante asociadas a artículos.

En la Figura 14 y Figura 15, se observa también que el número de artículos a tener en cuenta es elevado, por ejemplo, en el caso de la miocardiopatía hipertrófica o síndrome de QT largo se tiene 647 y 670 artículos asociados respectivamente. Para el síndrome de Marfan y síndrome de Brugada se tienen 228 y 132 artículos para analizar, el resto de fenotipos presenta un número mucho más reducido de artículos. Las variantes obtenidas en F4 son un conjunto de variantes candidatas sobre las que se necesita adquirir más datos y aplicar filtros para continuar con su clasificación.

En el caso de Ensembl, la información de qué artículos están asociados a las variantes se encuentra en el archivo que se obtiene a través de la aplicación web. Se parte de 14.169 variantes que tras aplicar los filtros F1-F3 se quedan en 2.008 variantes y que tras aplicar el filtro F4 se quedan en 873 variantes con significado clínico relevante y con artículos asociados. En la Figura 16 y Figura 17 de los datos obtenidos tras aplicar F1-F4.

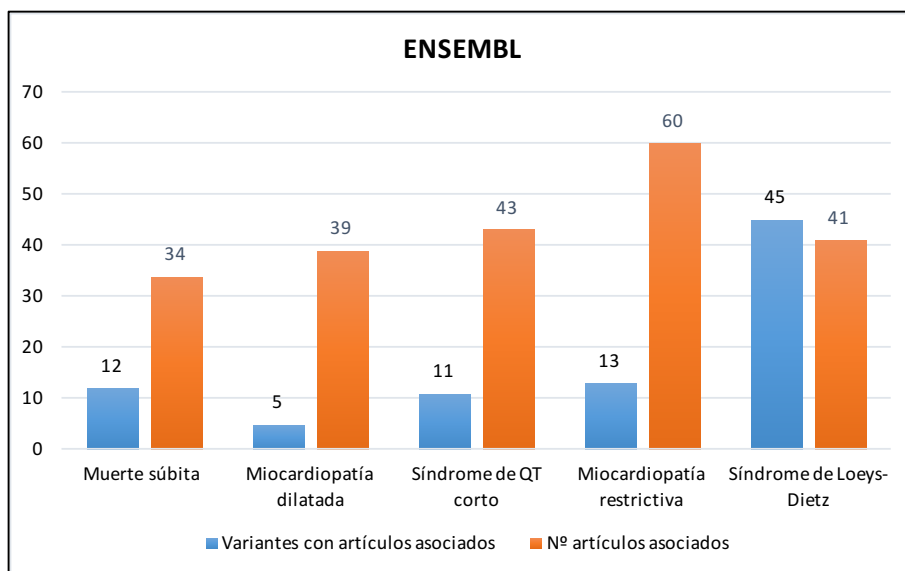


Figura 16. Primeros resultados en ENSEMBL tras aplicar los filtros F1-F4 del método SILE (parte 1).

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

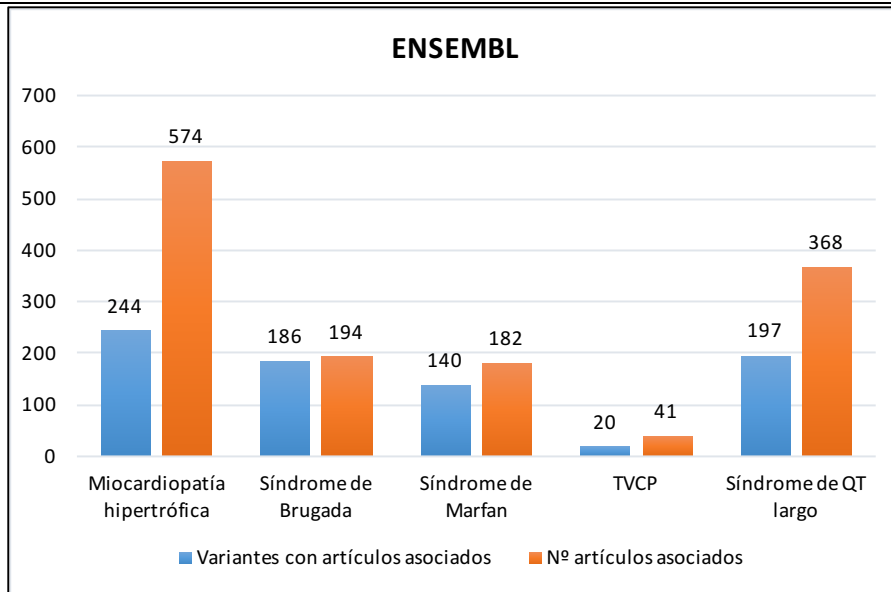


Figura 17. Primeros resultados en Ensembl tras aplicar los filtros F1-F4 del método SILE (parte 2).

Después de aplicar los filtros F1, F2, F3 y F4 a los datos iniciales obtenidos de Ensembl se consiguen 873 variantes y 1576 artículos distribuidos de la siguiente manera:

- Muerte súbita: 12 variantes candidatas y 34 artículos.
- Miocardiopatía dilatada: 5 variantes candidatas y 16 artículos.
- Síndrome de QT corto: 11 variantes candidatas y 43 artículos.
- Miocardiopatía restrictiva: variantes y 60 artículos.
- Síndrome de Loeys-Dietz: variantes y 41 artículos.
- Miocardiopatía hipertrófica: 244 variantes y 575 artículos.
- Síndrome de Brugada: 186 variantes y 194 artículos.
- Síndrome de Marfan: 140 variantes y 182 artículos.
- TVPC: 20 variantes y 41 artículos.
- Síndrome de QT largo: 197 variantes y 368 artículos.

El 79,38% de las variantes tiene asignado como significado clínico "*pathogenic*", el 19,93% de las variantes son "*not provided*" y corresponden a la muerte súbita, el síndrome de QT corto, síndrome de Brugada, síndrome de Marfan, y síndrome de QT largo. Y un 0,69% cuyo significado clínico es "*risk factor*" y pertenecen a la muerte súbita, miocardiopatía hipertrófica, miocardiopatía restrictiva, y síndrome de QT largo.

Al igual que en Clinvar, en Ensembl la mayor reducción de variantes se produce en los filtros F1-F3 (un 85,83%), lo que indica que la mayor parte de variantes genómicas estudiadas en este base de datos no tienen un significado clínico relevante. El filtro F4 ha reducido las variantes obtenidas en un 43,48%. En total, se ha reducido las variantes en un 93,83%, siendo el 6,26% relevantes hasta este punto.

Finalmente, la última base de datos a analizar es la de GWAS Catalog. Esta base de datos no presenta un campo manifestando el significado clínico por lo que se considera este campo como "vacío", por lo que tras aplicar F1 las variantes pasan directamente al filtro F4 donde se

analiza los artículos asociados. Todas las variantes iniciales pasan este filtro (152 variantes) por lo que el siguiente filtro a aplicar es el F5. Este filtro manifiesta lo siguiente:

1. El número de casos debe ser mayor o igual a 700.
2. El número de controles debe de ser mayor o igual que 700.
3. El p-valor debe ser menor que $5 \cdot 10^{-8}$.
4. Para variantes que aumenten el riesgo de sufrir la enfermedad, el OR debe ser mayor que 1 y el límite inferior del IC debe ser mayor que 1.
5. Para las variantes que disminuyan el riesgo de sufrir la enfermedad, el OR debe ser menor o igual que 0.9 y el límite superior del IC debe ser menor que 1.

El p-valor se puede definir como la probabilidad de error que se cometería en caso de rechazar una hipótesis con los datos que se tenga. En este caso se ha elegido $5 \cdot 10^{-8}$ para obtener una asociación alta entre la variante y el fenotipo.

El odds ratio (OR) es una medida de asociación entre un genotipo y un fenotipo. Un OR de 1.0 significa que la variante no afecta las probabilidades de tener la enfermedad, los valores superiores a 1.0 significan que existe una asociación entre la variante y el riesgo de enfermedad, y aquellos por debajo de 1.0 significan que existe una asociación negativa entre la variante y el riesgo de enfermedad. El intervalo de confianza (IC) alrededor del OR es tan importante como la medida de asociación en sí para que haya asociación de aumentar o disminuir el riesgo de la enfermedad.

La Figura 18 muestra un resumen de los resultados obtenidos.

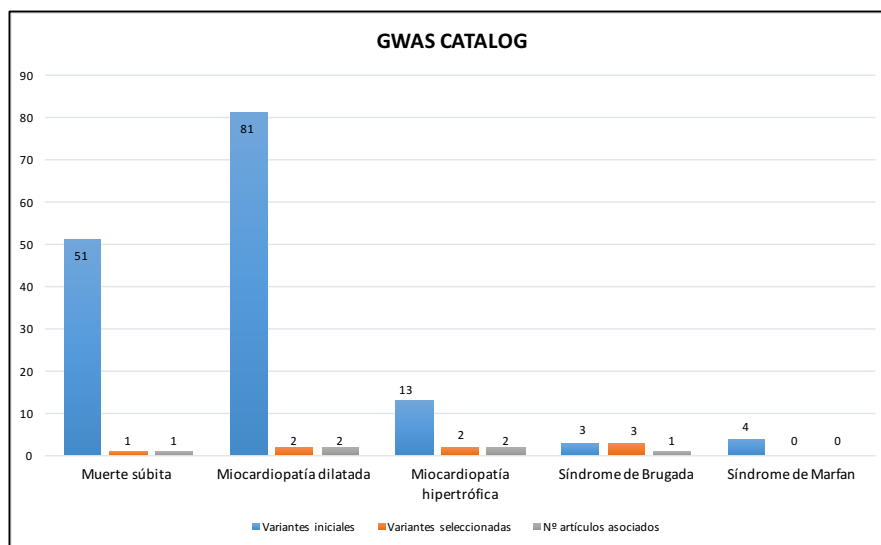


Figura 18. Primeros resultados en GWAS Catalog tras aplicar el filtro F5.

Se obtienen un total de 8 variantes tras aplicar el filtro F5. De todas estas, una de las variantes se relaciona a la muerte súbita (rs4665058). Esta variante se asocia con el infarto de miocardio, que puede ser causa o no de alguno de los fenotipos estudiados en esta tesis por lo que no se asigna a ninguno en específico. No obstante, esta variante no se descarta porque es relevante según los filtros aplicados y resulta de gran interés. El resto de las variantes afectan concretamente a la miocardopatía dilatada (2 variantes), la miocardopatía hipertrófica (2

variantes) y el síndrome de Brugada (3 variantes). Los detalles de estas variantes se encuentran en el Anexo 4.

El filtro F5 no se puede aplicar directamente a los datos de Clinvar y Ensembl dado que hay que extraer los datos estadísticos asociados a las variantes de la documentación científica. Como se ha visto, por ejemplo, en Ensembl la miocardiopatía hipertrófica tiene asociados 574 artículos únicos y diferentes. No es viable leerse esa cantidad de artículos con el tiempo disponible para la realización de este TFM y porque cada día aparecen nuevos artículos y sería una tarea interminable. Además, no estamos trabajando con un único fenotipo, sino con 9 fenotipos y 3 bases de datos por lo que la cantidad de artículos total es muy grande (1.898 únicos y diferentes) e inviable de leer por una persona. Por lo tanto, se debe buscar una alternativa para analizar estos artículos.

El primer problema que se encuentra es que hay que determinar de entre todos los artículos disponibles cuales son los más interesantes para obtener los datos que necesitamos. Existen los siguientes tipos de estudios:

- **Estudios de ligamento (*linkage*):** estos estudios permiten determinar la influencia que la genética tiene en el desarrollo de una determinada enfermedad, suelen realizarse en familias y pueden identificarse por los términos *Mesh*. Los *Medical Subject Headings (Mesh)* son descriptores controlados y jerárquicamente organizados que se utilizan para indexar, catalogar y buscar información biomédica y relacionada con la salud [61]. Los estudios de ligamento tienen asociados los siguientes términos *Mesh*: “*Linkage Disequilibrium*” o “*Genetic linkage*”.
- **Estudios de asociación (*association*):** estos estudios identifican las variantes de interés para una determinada población. Con este fin se compara la frecuencia relativa de las diferentes variantes de una serie de polimorfismos entre los individuos afectados y un grupo de controles. Estos artículos pueden identificarse por los términos *Mesh* “*Case-Control Studies*” o “*Genetic Association Studies*”. Estos estudios de asociación se pueden dividir en:
 - **Estudios GWAS:** estos estudios tienen como objetivo la identificación de variantes genómicas de interés para una determinada población a partir del estudio y comparación del genoma completo de los individuos. Estos artículos se pueden identificar por el término *Mesh* “*Genome-Wide Association Study*”.
 - **Estudios de cohortes (*cohort*):** el objetivo de estos estudios es identificar la relación entre los factores genéticos y los factores ambientales para el desarrollo de una enfermedad. Para ello es necesario comparar la frecuencia de aparición de un evento entre dos grupos, uno de los cuales está expuesto a un factor que no está presente en el otro grupo, a partir de un subconjunto de una población. Estos artículos se pueden encontrar utilizando el término *Mesh* “*Cohort Studies*”.

La automatización de la clasificación de artículos tiene en cuenta lo descrito anteriormente, además, el año de publicación, la revista donde se ha publicado, el título, el resumen y las

palabras clave. Dicha automatización está realizada por el centro de I+D PROS, por lo cual resultó muy sencillo obtenerla y para su aplicación se necesita únicamente los identificadores Pubmed de los artículos. Estos identificadores ya son conocidos cuando se aplicó el filtro F4 y se obtienen un total de 1.898 entre los repositorios. Una vez clasificados los estudios (ver resultados en

Tabla 3) se obtendrán los que más resulten de interés para obtener la información que se necesita, por lo que nos quedaremos con estudios que estén clasificados como ASSOCIATION, GWAS o COHORT o una combinación de estos.

TIPO	Nº ARTÍCULOS
ASSOCIATION	40
GWAS	52
COHORT	88
LINKAGE	55
ASSOCIATION AND COHORT	3
GWAS AND LINKAGE	9
GWAS AND COHORT	3
GWAS AND ASSOCIATION	3
COHORT AND LINKAGE	3
OTHER	1642
ARTICULOS TOTALES	1.898

Tabla 3. Resultados de la clasificación de artículos.

Los artículos clasificados como OTROS son documentos que no presenta algún indicativo de ser ASSOCIATION, GWAS o COHORT y describen características de las enfermedades, métodos informáticos para evaluar variantes, estudios de secuenciación, de biomoléculas, entre otros.

Tras la clasificación se obtienen **201 artículos en total** que son **interesantes** en la clasificación de variantes genómicas. Tras la clasificación se han reducido los artículos en un 89,4%. Del total de variantes que quedan por analizar, se obtienen aquellas que estén asociadas a algunos de estos artículos obtenidos en la clasificación. En la Figura 19 se muestra un resumen del número de variantes obtenidas.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca:
aplicaciones a la medicina de precisión

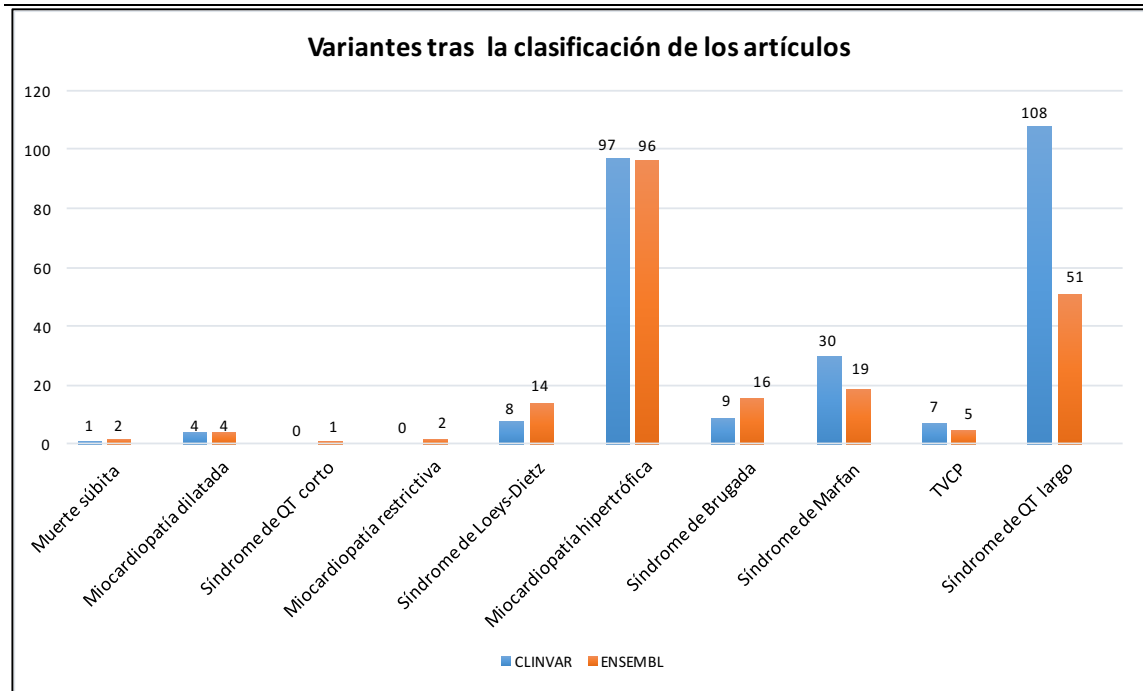


Figura 19. Resumen del número de variantes obtenidas tras la clasificación de artículos.

En Clinvar se encontraron un total de 264 variantes:

- Miocardiopatía dilatada: 4 variantes.
- Síndrome de Loey-Dietz: 8 variantes.
- Miocardiopatía hipertrófica: 97 variantes.
- Síndrome de Brugada: 9 variantes.
- Síndrome de Marfan: 30 variantes.
- TVCP: 7 variantes.
- Síndrome de QT largo: 108 variantes.

Además, se encontró una variante asociada (rs111033560) directamente con la muerte súbita. Según SNPedia, posiblemente esté relacionada con la miocardiopatía hipertrófica.

En cuanto a Ensembl se encontraron 210 variantes:

- Miocardiopatía dilatada: 4 variantes.
- Síndrome de QT corto: 1 variantes.
- Miocardiopatía restrictiva: 2 variantes.
- Síndrome de Loey-Dietz: 14 variantes.
- Miocardiopatía hipertrófica: 96 variantes.
- Síndrome de Brugada: 16 variantes.
- Síndrome de Marfan: 19 variantes.
- TVCP: 5 variantes.
- Síndrome de QT largo 51 variantes.

Asimismo, se encontraron 2 variantes relacionadas con la muerte súbita: rs203462 y rs111033560. La rs203462 se asocia a un defecto de conducción cardíaca mientras que la rs111033560 puede estar asociada a la miocardiopatía restrictiva o miocardiopatía dilatada.

Como ya se ha mencionado anteriormente, leer los artículos científicos y comprender la información en detalle llevaría mucho tiempo y no sería una tarea viable. Al estudiar la bibliografía asociada y quedarse con los estudios más relevantes, se obtienen 201 artículos importantes, pero aún es un número elevado de leer para obtener la información estadística que se requiere para aplicar el filtro F5. El siguiente filtro a aplicar (F6) también está relacionado con la literatura, y pide analizar si hay conflictos en la documentación científica. El resto de filtros no están relacionados con la literatura (F7 y F8), pero para llegar a estos antes se tiene que pasar los filtros de la literatura. Dada que leer cada artículo no es una opción, se plantea una alternativa a estos filtros (F5-F8) para llegar a las variantes relevantes finales.

La otra alternativa que se propone para continuar con la clasificación de las variantes genómicas son las reglas y criterios establecidos el por ACMG [4]. El flujo de trabajo que se plantea es el de la Figura 20.

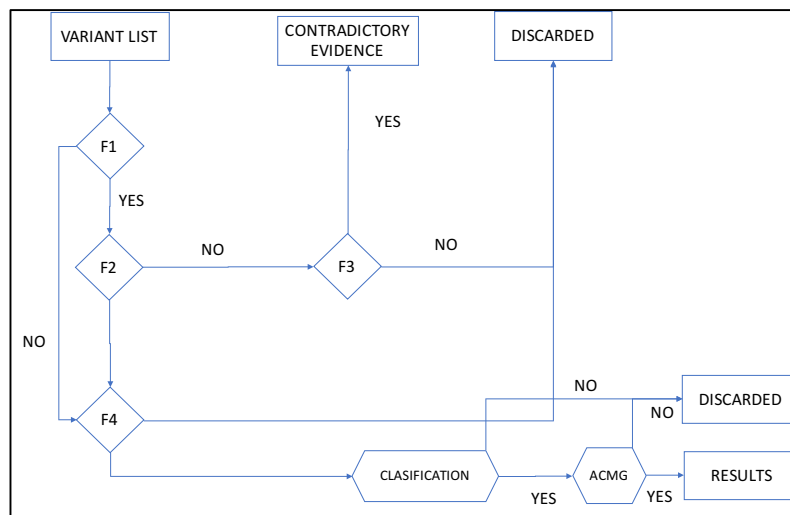


Figura 20. Nuevo flujo de trabajo para la obtención de variantes genómicas.

Los filtros F1-F4 se mantienen como en el primer flujo (Figura 13) mientras que el F5 se sustituye por un bloque de clasificación de la documentación científica. Por último, los filtros F6-F8 se sustituyen por los criterios propuestos por la ACMG para la obtención de variantes genómicas relevantes. Siendo:

- F1: ¿Se especifica el significado clínico de la variante?
- F2: ¿Es el significado clínico distinto de “benign”, “likely benign”, “likely pathogenic”, “uncertain” o “conflicting interpretation”?
- F3: ¿Es el significado clínico de la variante igual a “conflicting interpretation”?
- F4: ¿Tiene la variante al menos un artículo asociado?
- CLASIFICATION: ¿La variante tiene al menos un artículo asociado a los obtenidos en la clasificación?
- ACMG: ¿La variante cumple con al menos con un conjunto de reglas fiable establecidos por la ACMG?

En el siguiente punto se explican las reglas de la ACMG.

4.2.1 Guías y estándares para interpretación de variantes por el ACMG

El American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG) han desarrollado guías y pautas [4] para determinar el significado clínico de las variantes genómicas. Las pautas y guías establecidas se pueden aplicar únicamente a variantes de enfermedades que sean monogénicas. Sin embargo, estas guías deben aplicarse con cuidado, por ejemplo, cuando se interpretan variantes relacionadas con las enfermedades cardíacas que ocurren por variación en el gen MYH7. Este gen está involucrado con la miocardiopatía hipertrófica, miocardiopatía dilatada, miocardiopatía restrictiva, entre otras.

En la búsqueda de información sobre los fenotipos asociados a la muerte súbita cardíaca se concluyó que todas son monogénicas, es decir, que con un único gen afectado se podría desarrollar la enfermedad, por lo que las guías mencionadas anteriormente pueden ser aplicadas.

La ACMG consta de 28 criterios que se agrupan en bloques dependiendo si son:

- Evidencia muy fuerte de patogenicidad (PVS1).
- Evidencia fuerte de patogenicidad (PS1, PS2, PS3, PS4).
- Evidencia moderada de patogenicidad (PM1, PM2, PM3, PM4, PM5, PM6).
- Evidencia de apoyo de patogenicidad (PP1, PP2, PP3, PP4, PP5).
- Evidencia independiente de impacto benigno (BA1).
- Fuerte evidencia de impacto benigno (BS1, BS2, BS3, BS4).
- Evidencia de apoyo de impacto benigno (BP1, BP2, BP3, BP4, BP5, BP6, BP7).

Los criterios propuestos por la ACMG que se propone aplicar mediante herramientas informáticas que las implementen son: PS1, PS4, PM2, PP3, PP5, BA1, BS1, BP4 y BP6. Estos criterios son elegidos por los siguientes motivos:

- PS1 en el caso de que ya se conociera una variante relevante.
- PS4 está relacionado con el filtro F4.
- BA1, BS1 y PM2 están relacionados con el filtro F6.
- PP3 y BP4 se pueden analizar mediante herramientas informáticas.
- PP5 y BP6 están relacionadas con los paneles de expertos que se encargan de apoyar la significancia clínica de una variante.

Estos criterios son suficientes para apoyar el significado clínico patogénico de las variantes que es lo que se quiere conseguir en esta tesis.

Las reglas a aplicar afirman lo siguiente:

- **PS1:** "La variante produce el mismo cambio de aminoácidos en la misma posición que otra variante que se sabe que es patogénica". Por ejemplo, c.34G>C cambia a c.34G>T produciendo en ambos casos los aminoácidos Leucina y Valina. Si el primer cambio se considera patogénico, el segundo también.
- **PS4:** "La prevalencia de la variante en los individuos afectados aumenta significativamente en comparación con la prevalencia en los controles (OR > 5 y IC no cruza 1)". Las variantes genómicas con un OR muy elevado se consideran altamente

penetrantes. El intervalo de confianza (IC) alrededor del OR es tan importante como la medida de asociación en sí. Si el IC incluye 1.0 (por ejemplo, OR = 2.5, IC = 0.9– 7.4), hay poca confianza en la afirmación de asociación.

- **BA1:** "*La frecuencia alélica es superior al 5% en Exome Sequencing Project, 1.000 Genomes Project o el Exome Aggregation Consortium en poblaciones con más de 1.000 individuos*".
- **BS1:** "*La frecuencia de alelos en una población de control con más de 1.000 individuos es mayor de lo esperado para el trastorno*".
- **PM2:** "*La variante está ausente de los controles en Exome Sequencing Project, 1000 Genomes Project o el Exome Aggregation Consortium en poblaciones con más de 1.000 individuos*"

Evaluar la frecuencia de una variante genómica en una población control o general es útil para evaluar su posible patogenicidad. Esta información se puede encontrar en *Genomes Project, National Heart, Lung, and Blood Institute, Exome Sequencing Project, Exome Variant Server, Exome Aggregation Consortium*. En general, si la frecuencia alélica en una población de control de más de 1.000 individuos es mayor de lo esperado para el trastorno o si es superior al 5% se considera como fuerte apoyo para interpretar la variante como benigna de un trastorno mendeliano raro. Si la variante está ausente de una gran población general o una cohorte de control con más de 1.000 individuos, esta observación puede considerarse una evidencia moderada de patogenicidad.

- **PP3:** "*Múltiples líneas de evidencia computacional respaldan un efecto nocivo*".
- **BP4:** "*Múltiples líneas de evidencia computacional sugieren que no hay impacto en el gen o producto genético*".

Existen algoritmos que se encargan de predecir el significado clínico de las variantes utilizando los mismos datos o similares. Si todos los programas que se utilicen para realizar la predicción llegan al mismo resultado, esta evidencia se puede contar como apoyo. Si no están de acuerdo las predicciones, la evidencia no puede usarse para clasificar la variante.

- **PP5:** "*La fuente confiable informa que la variante es patogénica*".
- **BP6:** "*La fuente confiable informa que la variante es benigna*".

Cada vez existen más laboratorios clínicos, paneles de expertos, médicos, etc. que se pueden considerar como fuentes acreditadas. Estas fuentes acreditadas comparten la información de la clasificación de las variantes en repositorios que pueden o no tener fácil acceso. Esta información se puede utilizar como criterio para clasificar la variante.

Los criterios mencionados anteriormente se combinan en una serie de reglas de manera que se pueda proporcionar una clasificación de las variantes, dependiendo de la importancia de cada una (Tabla 4).

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

SIGNIFICADO	FÓRMULA
Fiables	PS1 AND PS4
	(PS1 OR PS4) AND (>=3 PM2 OR (2 PM2 AND ((PP3 AND PP5) OR >= 2 PP5))
Probablemente fiables	PS4 AND (1-2 PM2 OR ((PP3 AND PP5) OR >= 2 PP5))
	>= 3 PM2
	2 PM2 AND ((PP3 AND PP5) OR >= 2 PP5)
	PM2 AND ((PP3 AND >= 3 PP5) OR >=4 PP5)
Descartadas	BA1
	>= 2 BS1
	BS1 AND (BP4 OR BP6)
	BP4 AND >= 1 BP6
	>= 2 BP6

Tabla 4. Aplicación de las reglas del ACMG

Según la ACMG las variantes pueden ser **consideradas fiables** si cumplen alguna de las siguientes reglas:

- PS1 AND PS4: se deben cumplir las dos afirmaciones descritas.
- (PS1 OR PS4) y >=3 PM2: se debe cumplir PS1 o PS4 y que esté ausente en al menos 3 bases de datos de población.
- (PS1 OR PS4) AND 2 PM2 AND (PP3 AND PP5): que cumpla PS1 o PS4 y que esté ausente en al menos 2 bases de datos de población (2 PM2), soportada por al menos una línea computacional (PP3) y al menos esté reportada por una fuente fiable y con reputación como patogénica (PP5).
- (PS1 OR PS4) AND 2 PM2 AND >= 2PP5): se debe cumplir PS1 o PS4 y que esté ausente en al menos 2 bases de datos de población (2 PM2), y que al menos esté reportada como patogénica por dos fuentes fiables y con reputación como patogénica (>= 2PP5).
- (PS1 OR PS4) AND 1 PM2 AND (PP3 AND >= 3PP5): se debe cumplir PS1 o PS4 y que esté ausente en al menos 1 base de datos de población (1 PM2), soportada por al menos una línea computacional (PP3) y que al menos esté reportada como patogénica como mínimo por tres fuentes fiables y con reputación como patogénica (>= 3PP5).
- (PS1 OR PS4) AND 1 PM2 AND >= 4 PP5): se debe cumplir PS1 o PS4 y que esté ausente en al menos 1 base de datos de población (1 PM2), y que al menos esté reportada como patogénica como mínimo por cuatro fuentes fiables y con reputación (>= 4 PP5).

Las variantes pueden ser consideradas **posiblemente fiables** si cumplen alguna de las siguientes reglas (Tabla 4):

- PS4 AND 1-2 PM2: se debe cumplir PS4 y que esté ausente en 1 o 2 bases de datos de población (1-2 PM2).

- PS4 AND PP3 AND PP5: se debe cumplir PS4, que esté soportada por al menos una línea computacional (PP3) y reportada por una fuente fiable y con reputación como patogénica(PP5).
- PS4 AND ≥ 2 PP5: se debe cumplir PS4 y reportada por al menos dos fuentes fiables y con reputación como patogénica (≥ 2 PP5).
- ≥ 3 PM2: ausente como mínimo de 3 bases de datos de población.
- 2 PM2 AND PP3 AND PP5: debe estar ausente de 2 bases de datos de población, soportada por al menos una línea computacional (PP3) y reportada por una fuente fiable y con reputación como patogénica (PP5).
- 2 PM2 AND ≥ 2 PP5: debe estar ausente de 2 bases de datos de población y reportada por al menos dos fuentes fiables y con reputación como patogénica (≥ 2 PP5).
- 2 PM2 AND PP3 AND ≥ 3 PP5: debe estar ausente de 2 bases de datos de población, soportada por al menos una línea computacional (PP3) y reportada por al menos 3 fuentes fiables y con reputación como patogénica (≥ 3 PP5).
- 2 PM2 AND ≥ 4 PP5: debe estar ausente de 2 bases de datos de población, y reportada por al menos 4 fuentes fiables y con reputación como patogénica (≥ 4 PP5).

Las variantes pueden ser **descartadas** si cumple alguna de las siguientes combinaciones (Tabla 4):

- BA1.
- ≥ 2 BS1 (en al menos dos bases de datos de control).
- BS1 AND BP4: se cumpla BS1 y que al menos una línea computacional sugiere que no hay impacto en el gen (BP4).
- BS1 AND BP6: se cumpla BS1 y que al menos una fuente fiable y con reputación la hay reportada como benigna (BP6).
- BP4 AND ≥ 1 BP6: una línea computacional sugiere que no hay impacto en el gen (BP4) y como mínimo ha sido reportada por una fuente fiable y con reputación como Benigna(BP6).
- ≥ 2 BP6: ha sido reportada como mínimo 2 fuentes fiables y con reputación como benigna.

Una vez explicadas algunas de las reglas establecidas por el ACMG, se procede a su aplicación a las variantes relacionadas con los estudios que asocian las variantes genómicas con la enfermedad.

Juntado las variantes obtenidas en los tres repositorios con los que se ha trabajado hasta ahora (Clinvar: 210 variantes, Ensembl: 214 variantes y GWAS Catalog: 8 variantes), se obtienen un total de 356 variantes únicas para analizar con las reglas del ACMG.

Dado que no es un objetivo de esta tesis proporcionar una automatización a estas reglas se usarán los datos de otros repositorios que se encargan de implementarlas.

A continuación, se explica las herramientas que se encargan de implementar las reglas: PS1, PS4, PM2, PP3, PP5, BA1, BS1, BP4 y BP6.

4.2.2 CardioVai, Varsome e intervar

Las reglas PS1, PS4, PM2, PP3, PP5, BA1, BS1, BP4 y BP6 están implementadas o analizadas por diferentes herramientas como, por ejemplo, el *Laboratory for Bioinformatics, Mathematical Modelling and Synthetic Biology (BMS)* [62] que ha desarrollado CardioVAI, una herramienta

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca: aplicaciones a la medicina de precisión

especializada en patologías cardíacas. Las bases de datos Varsome, Intervar o ClinGen también analizan estas reglas, pero no están especializadas exclusivamente en enfermedades cardíacas. Se utilizará como referencia CardioVAI por su especificidad y se complementará con los datos obtenidos por Varsome e Intervar.

CardioVAI⁶ (Figura 21) interpreta variantes genómicas en 72 genes que están asociadas con enfermedades cardíacas. Para obtener la información se necesita el nombre de la variante representada por la nomenclatura HGVS. CardioVAI sigue las pautas del ACMG, implementando así 18 de 28 criterios proporcionados por ACMG y clasifica las variantes como: “Pathogenic”, “likely Pathogenic”, “benign”, “likely benign” y “uncertain significance”. Además, CardioVAI adapta los criterios y reglas del ACMG para las variantes de MYH7, de acuerdo con el panel de expertos en cardiomiopatía heredada de ClinGen. Esta base de datos tiene 3 formas de consultar la información con HGVS a nivel de proteína, ADN o con la posición de la variante. En esta tesis, se va a utilizar la consulta a nivel de ADN. Por ejemplo, c.4375C>T.

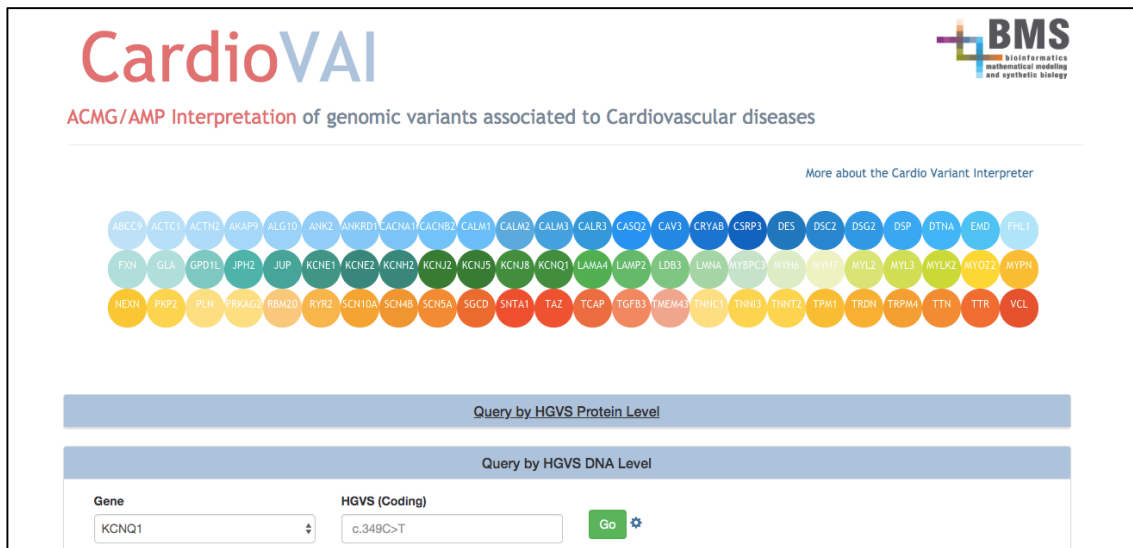


Figura 21. CardioVAI (Website).

El repositorio Varsome se describió previamente y además de utilizarse para completar información se emplea para analizar las reglas de la ACMG dado que elabora su propia implementación de 23 criterios. A diferencia de CardioVai no específica que realice una adaptación para las variantes del gen MYH7 ni tampoco está especializado en enfermedades cardíacas. La información de las reglas se puede obtener utilizando el identificador rs de la variante genómica.

⁶ CardioVAI: <http://cardiovai.engenome.com>

Figura 22. InterVAR (Website).

Por último, InterVAR⁷ (Figura 22), es una herramienta de software que implementa 18 criterios de la ACMG para la interpretación clínica de las variantes genómicas. Además, hace un reajuste manual para reinterpretar el significado clínico. Esta base de datos clasifica las variantes como “benign”, “likely benign”, “uncertain significance”, “likely pathogenic” y “pathogenic”. Las consultas en esta base de datos se pueden realizar de tres formas: utilizando la posición de la variante, el identificador rs o el identificador HGNC. En este caso, se utilizará la opción de consulta mediante el identificador rs. Tanto Varsome como Intervar no mencionan realizar adaptaciones de los criterios de la ACMG para las variantes del gen MYH7.

Tras utilizar las herramientas sobre un conjunto de variantes a analizar se obtuvieron los siguientes resultados: en CardioVAI se obtuvieron 34 variantes genómicas que cumplen con alguna de las combinaciones de los criterios de la ACMG descritos en Tabla 4. En la Tabla 5 se recogen las combinaciones de reglas que cumplen todas y cada una de las 34 de las variantes en CardioVAI. Como se puede observar las variantes cumplen tanto combinaciones fiables como posiblemente fiables.

SIGNIFICADO	FÓRMULA
Fiable	PS1 AND ≥ 3 PM2
	PS1 AND 2PM2 AND (PP3 AND PP5)
	PS1 AND 2PM2 AND ≥ 2 PP5
Posiblemente fiables	≥ 3 PM2
	2PM2 AND PP3 AND PP5
	2PM2 AND ≥ 2 PP5

Tabla 5. Conjunto de reglas que se cumplen en CardioVAI

⁷ Intervar: <http://wintervar.wglab.org/results.php>

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca: aplicaciones a la medicina de precisión

Las 34 variantes están relacionadas con los siguientes fenotipos: la miocardiopatía dilatada (3 variantes), miocardiopatía restrictiva (1 variante), miocardiopatía hipertrófica (19 variantes), síndrome de Loeys-Dietz (1 variante), síndrome de Brugada (9 variantes) y 1 variante asociada al infarto de miocardio. En Varsome e Intervar tienen únicamente implementadas para estas variantes las reglas PM2, PP3 y PP5, por lo tanto, las 34 variantes únicamente tienen el soporte de CardioVai.

El síndrome de Marfan no obtuvo variantes genómicas en CardioVAI dado que no se encontró una combinación de criterios establecidos por la ACMG para apoyar su significado clínico patogénico. No obstante, se obtuvo una combinación de criterios en Varsome para 4 variantes (rs397515848, rs140630, rs111231312, rs387906697) relacionados con este síndrome. Estas 4 variantes cumplen con todos los conjuntos de reglas considerados en Tabla 5.

Finalmente, como resultado final del proceso de identificación se han obtenido 38 variantes totales de las 3 bases de datos estudiadas para la muerte súbita (Figura 23). Los detalles de estas variantes se pueden ver en el Anexo 5. Se han obtenido variantes finales para 6 de los 9 fenotipos estudiados: miocardiopatía dilatada, miocardiopatía restrictiva, miocardiopatía hipertrófica, síndrome de Loeys-Dietz, síndrome de Marfan y síndrome de Brugada. De las 38 variantes genómicas, 24 proceden de Clinvar, 8 de GWAS y 6 de Ensembl.

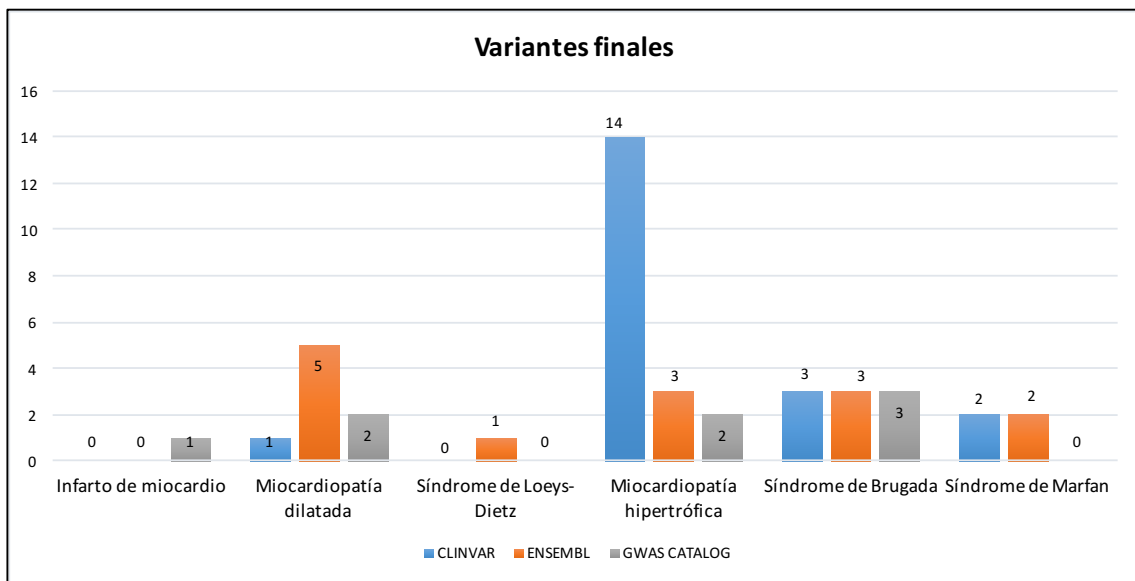


Figura 23. Variantes finales.

Con estos resultados se puede decir que de las variantes iniciales tras la aplicación de los filtros y de los criterios de la ACMG, únicamente el 0,18% de Clinvar, el 5,26 % de GWAS Catalog y 0,04% de Ensembl resultan relevantes del total de variantes.

Tras finalizar la etapa de identificación, las siguientes etapas que deben cursar las variantes genómicas obtenidas son la carga y la explotación. Estas etapas del método SILE no han sido implementadas porque se encuentran fuera de los objetivos y alcance de este TFM, por lo que se propone como líneas futuras.

CAPÍTULO 5. CONCLUSIONES Y TRABAJO FUTURO

La muerte súbita cardíaca es una de las preocupaciones más importantes dentro del ámbito de la cardiología por el número de muertes que se producen, el impacto social y por la dificultad de ser detectada a tiempo. Los médicos necesitan tener información de calidad sobre las variantes que la provocan para poder evitarla y dar el tratamiento más adecuado. El método SILE desarrollado por el centro de investigación PROS de la Universidad Politécnica de Valencia supone un avance en el manejo de la información genómica en el ámbito de la medicina de precisión, proporcionando así información genómica de calidad a los médicos.

A lo largo de esta tesis se ha explicado el conjunto de repositorios (Clinvar, Ensembl y GWAS Catalog) que se utiliza para obtener los datos genómicos a estudiar con el método SILE. También se ha descrito los fenotipos asociados a la muerte súbita cardíaca dado que es importante conocer todo lo que se pueda sobre el fenotipo. La información de los fenotipos, en concreto las subcategorías y sinónimos, ha sido de utilidad en el momento de la búsqueda y análisis de la información de las variantes genómicas relacionadas con los fenotipos.

Durante la ejecución del método SILE, en la etapa de identificación a la hora de aplicar el filtro F5 en Clinvar y Ensembl se encontró una gran cantidad de artículos para la extracción de información. La opción de analizar los artículos uno por uno se descarta por tratarse de 2.081 y se propone aplicar la clasificación de artículos realizado por el centro de I+D PROS y algunos de los criterios de la ACMG para obtener las variantes de interés.

A lo largo de este trabajo se han tratado 9 fenotipos que pueden causar la muerte súbita cardíaca, pero no se han obtenido variantes candidatas para todos los fenotipos estudiados en esta tesis. Esto se debe a que no toda la información analizada es relevante. Tras la aplicación de los filtros y los criterios de la ACMG se obtiene las variantes de calidad cumpliendo con el objetivo de la tesis, mientras que todas las demás son descartadas porque no son suficientemente fiables para realizar un diagnóstico genómico.

En resumen, lo que se hizo para cumplir el objetivo de esta tesis es:

- Selección de los repositorios más adecuados para identificar variantes relevantes de acuerdo al fenotipo de interés.
- Aplicación de una serie de criterios de calidad de acuerdo con el método SILE para la identificación de variantes relevantes.
- Para aquellas variantes sobre las que era necesario obtener más información se realizó una clasificación de la bibliografía asociada, eligiendo únicamente los artículos que se hayan clasificado como ASSOCIATION, GWAS, COHORT o una combinación de estos tipos de estudios. Esto redujo considerablemente el esfuerzo requerido para su análisis.
- Posteriormente, se propuso la aplicación de las normas ACMG a las variantes asociados a estos para obtener un conjunto de variantes relevantes. La clasificación se realizó utilizando las herramientas CardioVAI, Varsome e Intervar.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca: aplicaciones a la medicina de precisión

Tas completar el proceso se obtuvieron un total de 38 variantes: las 37 variantes asociados a los 6 fenotipos inicialmente bajo estudio y una al infarto de miocardio (rs4665058). Los fenotipos iniciales para los que se obtienen las 37 variantes son: la miocardiopatía dilatada, miocardiopatía restrictiva, el síndrome de Loeys-Dietz, síndrome de Brugada, síndrome de Marfan y miocardiopatía hipertrófica. Clinvar es la base de datos que ha aportado el mayor número de variantes a este trabajo (20 variantes). Del total de las variantes analizadas únicamente el 0,18% de Clinvar, el 5,26 % de GWAS Catalog y 0,04% de Ensembl resulta relevante.

Los resultados del método SILE han sido validados en otras enfermedades, pero los resultados obtenidos en este trabajo deben ser validados por profesionales médicos especializados en cardiología. Para ello se ha contactado con un profesional del equipo de cardiología del Hospital la Fe de Valencia, y los resultados del presente trabajo serán validados en el contexto de un trabajo en colaboración entre el centro PROS y el citado equipo médico.

En el desarrollo de esta tesis se han encontrado las siguientes limitaciones:

1. Heterogeneidad de las bases de datos: cada base de datos tiene su forma de mostrar su información.
2. No encontrar una base de datos específica que contenga información exclusivamente sobre variantes genómicas causantes de patologías del corazón.
3. GWAS proporciona información para menos de la mitad de los fenotipos propuestos a estudiar.
4. La cantidad de artículos a analizar de forma individual es enorme: en esta tesis 1.898.
5. Dificultad para obtener los datos estadísticos de Clinvar y Ensembl necesarios para la aplicación del filtro F5.

En cuanto a las líneas futuras de trabajo se propone lo siguiente:

- Para la limitación 5, se propone desarrollar herramientas informáticas basadas en machine learning que permitan obtener la información estadística de la documentación científica necesaria para clasificar las variantes genómicas según lo propuesto por el método SILE.
- Continuar con el análisis de los fenotipos estudiados aquí dado que cada día surge nueva información de interés que puede resultar relevante.
- Realización de las dos etapas del método SILE (carga y explotación) una vez los datos hayan sido validados.
- Por último y más importante, se propone como líneas futuras realizar la validación de las variantes obtenidas en este trabajo para que sean explotados dentro del ámbito de la medicina de precisión y la cardiología.

Esta tesis de master es la primera que se desarrolla para temas relacionados con la cardiología teniendo en cuenta el método SILE, las variantes genómicas y la medicina de precisión, contribuyendo y motivando a la realización de más trabajos que puedan debatir los resultados obtenidos en este TFM.

REFERENCIAS BIBLIOGRÁFICAS

- [1] HPE. "¿Qué es la secuenciación de nueva generación?" Accessed Oct. 2019, url: <https://www.hpe.com/mx/es/what-is/next-gen-sequencing.html>.
- [2] National Human Genome Project. "The Human Genome Project" Accessed Sep. 2019, url: <https://www.genome.gov/human-genome-project>.
- [3] Pastor, Óscar, "Conceptual Modelling of Life: Beyond the Homo Sapiens" class notes for Analysis of Genomic Data, Centro de Investigación en Métodos de Producción de Software, Universidad Politécnica de Valencia, Valencia, Spain, 2019..
- [4] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, 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, A. L. Q. A. Committee., Ed., May 2015.
- [5] J. F. Reyes, "Diseño y desarrollo de una sistema información genómica basado en un modelo conceptual holístico del genoma humano", Valencia, Valencia: PROS: Centro de investigación en métodos de Producción de software, Febrero, 2018 .
- [6] Sociedad Española de Bioquímica y Biología molecular, "La medicina de precisión como estrategia" Accessed Sep 2019..u. url: <https://www.sebbm.es/revista/articulo.php?id=346&url=la-medicina-de-precision-como-estrategia>, Ed.
- [7] Fundación Española del corazón, "Muerte Súbita" Accessed Sep 2019., url: <https://fundaciondelcorazon.com/informacion-para-pacientes/enfermedades-cardiovasculares/muerte-subita.html>.
- [8] MedGen, "Sudden Cardiac Death" Accessed Sep 2019., url: <https://www.ncbi.nlm.nih.gov/medgen/38841>.
- [9] Goldstein S, Bayés de Luna A, Guindo J. , "Sudden cardiac death", New York: Futura;1994.
- [10] Antonio Bayés de Luna, Roberto Elousa, Sudden Death, Institut Català de Ciències Cardiovasculars, Hospital de Sant Pau, Barcelona, Spain. Grupo de Epidemiología y Geriátrica Cardiovascular, IMIM y CIBER de Epidemiología y Salud Pública, Barcelona, Spain.

- [11] Diversus PS. , "De febre pestilente tractacus et curations quorumdam particularium morborum etc...", Frankfurt: Heiss of a Wechel; 1586. .
- [12] McCurdy E, editor., "The notebooks of Leonardo da Vinci", New York: Reynal and Hitchcock; 1938 p.125.
- [13] Littré E. Oeuvres complètes d'Hippocrate., Vol. 5. Paris: Bailière: 1846.
- [14] Froissart J, Trans JT., "Chronicles of Englad, France, Spain and adjoining countries", London: Henry G. Bohn; 1849..
- [15] Lancisi GM., "De Subitaneis Mortibus", Roma: Giovanni Francisco Buagni; 1706.
- [16] Herrick JB. , "Clinical features of sudden death obstruction of the coronary arteries", JAMA.1912;59:2015-22. .
- [17] Kuller L. Lilienfeld A. Fisher R. , "An epidemiological study of sudden and expected deaths in adults", Medicine.1967;46:341-61..
- [18] Myerburg R. Juntilla J. , "Sudden cardiac death caused by coronary heart disease", Circulation. 2012; 125:1043-52. .
- [19] Sumeet S. Chugh, Kyndaron Reinier, Carmen Teodorescu, Audrey Evanado, Elizabeth Kehr, Mershed Al Samara, Ronald Mariani, Karen Gunson, and Jonathan Jui, Epidemiology of Sudden Cardiac Death: Clinical and Research, From the Cardiac Arrhythmia Center, Division of Cardiovascular Medicine, Oregon Health and Science University, Portland, Oregon.
- [20] Jaume Marrugata, Roberto Elosuaa, Miguel Gil, Epidemiology of sudden cardiac death in Spain, Unitat de Lípids i Epidemiologia Cardiovascular. Institut Municipal d'Investigació Mèdica. Barcelona.
- [21] Pascual Fernández, Ignacio, "Análisis de datos genómicos para la identificación de variantes genómicas relevantes en la enfermedad del Alzheimer", Trabajo Fin de Master en Ingeniería Biomédica.
- [22] Alba Escalera Balsera, Biotecnología para la medicina de precisión: Identificación de variantes genómicas para el diagnóstico de cáncer de mama, Trabajo Fin de Grado en Ingeniería Biomédica , 2018.
- [23] Simranpreet Kahur, Un proceso para la identificación sistemática de variantes genómicas: aplicaciones a la medicina de precisión, Trabajo Fin de Master en Ingeniería Tecnología de Sistema de Software, 2018.

- [24] Vanessa Solís Cabrera, Exploración de bases de datos genómicas dirigido por modelos conceptuales, Trabajo Fin de Máster en Ingeniería y Tecnología de Sistemas de Software, 2018.
- [25] Clara Soler Pellicer, Diseño de un sistema de información genómica para el diagnóstico del Neuroblastoma, Trabajo Fin de Grado en Ingeniería Biomédica , 2016.
- [26] Ana León Palacio, Óscar Pastor López , Smart Data for Genomic Information Systems: the SILE Method, Research Center on Software Production Methods (PROS), December 2018/January 2019.
- [27] Nefrología, "Genética de la Diabetes Mellitus" Accesed Nov. 2019, url: <https://www.revistanefrologia.com/es-genetica-diabetes-mellitus-articulo-X2013757511002452>.
- [28] Fundación Iberoamericana Down21, "Enfermedades monogénicas del cromosoma 21" Accessed Nov. 2019, url: <https://www.downciclopedia.org/genetica/peculiaridades-en-el-sindrome-de-down/2927-enfermedades-monogenicas-del-cromosoma-21.html>.
- [29] Roberto Barriales-Villa, Juan Ramón Gimeno-Blanes, Esther Zorio-Grim a, Tomás Ripoll-Vera, Artur Evangelista-Masip, Ángel Moya-Mitjans, Luis Serratosa-Fernández, Dimpna C. Albert-Brotons, José Manuel García-Pinilla y Pablo García-Pavía, "Protocolo de actuación de las cardiomiopatías familiares: síntesis de recomendaciones y algoritmos de actuación", Rev Esp Cardiol. 2016; 69(3): 300-309.
- [30] Genetics Home Reference, "Familial Dilated Cardiomyopathy" Apr. 2019, <https://ghr.nlm.nih.gov/condition/familial-dilated-cardiomyopathy#statistics>.
- [31] MedGen, "Familial Dilated Cardiomyopathy" Accessed Apr. 2019, url: <https://www.ncbi.nlm.nih.gov/medgen/90951>.
- [32] MedGen, "Long QT syndrome" Accessed Apr. 2019, url: <https://www.ncbi.nlm.nih.gov/medgen/44193>.
- [33] MedGen, "Noncompaction cardiomyopathy". Accessed Apr. 2019., url: <https://www.ncbi.nlm.nih.gov/medgen/326592>.
- [34] National Library of Medicine (US), Genetics Home Reference, "Long QT syndrome" Apr. 2019, <https://ghr.nlm.nih.gov/>.
- [35] Genetics Home Rerefence, "Short QT syndrome" Accessed Apr. 2019, url:<https://ghr.nlm.nih.gov/condition/short-qt-syndrome>.
- [36] MedGen, "Short QT syndrome" Accessed Apr. 2019, url:

<https://www.ncbi.nlm.nih.gov/medgen/378835>.

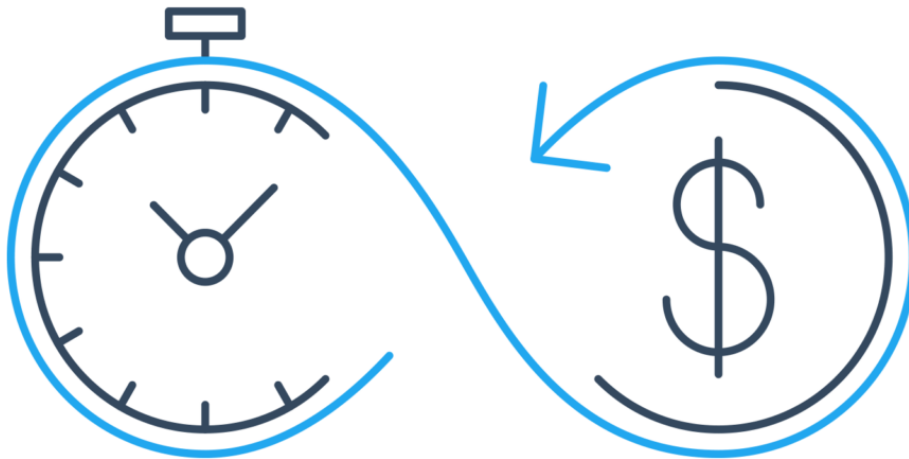
- [37] Genetics Home Reference, "Familial restrictive cardiomyopathy" Apr. 2019, <https://ghr.nlm.nih.gov/condition/familial-restrictive-cardiomyopathy#inheritance>.
- [38] MedGen, "Familial Restrictive Cardiomyopathy" Apr. 2019, <https://www.ncbi.nlm.nih.gov/medgen/382807>.
- [39] Genetics Home Reference, "Loeys-Dietz Syndrome" Accessed Apr. 2019, url: <https://ghr.nlm.nih.gov/condition/loeys-dietz-syndrome>.
- [40] MedGen, "Loeys-Dietz Syndrome" Accessed Apr. 2019, url: <https://www.ncbi.nlm.nih.gov/medgen/395827>.
- [41] Genetics Home Reference, "Familial hypertrophic cardiomyopathy" Accessed Apr. 2019, <https://ghr.nlm.nih.gov/condition/familial-hypertrophic-cardiomyopathy#synonyms>.
- [42] MedGen, "Primary familial hypertrophic cardiomyopathy" Accessed Apr. 2019, url: <https://www.ncbi.nlm.nih.gov/medgen/183649>.
- [43] Genetics Home Reference, "Brugada Syndrome". Apr. 2019, <https://ghr.nlm.nih.gov/condition/brugada-syndrome>.
- [44] Fundéu BBVA, "Mortalidad y Morbilidad" Accessed Nov. 2019., url: <https://www.fundeu.es/consulta/mortalidad-y-morbilidad-514/>.
- [45] Genetics Home Reference, "Marfan Syndrome" Apr. 2019, <https://ghr.nlm.nih.gov/condition/marfan-syndrome>.
- [46] Genetics Home Reference, "Catecholaminergic polymorphic ventricular tachycardia" Apr. 2019, url: <https://ghr.nlm.nih.gov/condition/catecholaminergic-polymorphic-ventricular-tachycardia#genes>.
- [47] National Center of Biotechnology Information, "Long QT Syndrome" Accessed Apr. 2019, url: <https://www.ncbi.nlm.nih.gov/books/NBK1129/>.
- [48] Health Sciences Library Systems, "General genomics databases and tools" Accessed Apr. 2019, url: https://www.hsls.pitt.edu/obrc/index.php?page=general_genomics.
- [49] Human genome variation Society, "Databases and tools" Accessed Apr. 2019, url: <http://www.hgvs.org/content/databases-tools>.
- [50] Nucleic Acid Research, "NAR Database summary paper category list" Accessed Apr. 2019, url: <https://www.oxfordjournals.org/nar/database/cat/8>.

- [51] Leon Palacio Ana , Pastor López Óscar, "Smart Data for Genomic Information Systems: the SILE Method", January 2019., Reseach Center on Software Production Methods (PROS), Polytechnic University of Valencia, Valencia, Spain.
- [52] Varsome, Varsome. Accesed Apr. 2019, url: <https://varsome.com/>.
- [53] National Center of Biotechnology Information, dbSNP Accessed Nov. 2019, url: <https://www.ncbi.nlm.nih.gov/snp/>.
- [54] SNPedia, SNPedia. Accessed Nov. 2019, url:<https://www.snpedia.com/index.php/SNPedia>.
- [55] National Center for Biotechnology Information. , Clinvar. Accessed Jun. 2019, url: <https://www.ncbi.nlm.nih.gov/clinvar/>.
- [56] Clinvar, "What is Clinvar?" Accessed Jun. 2019, url: <https://www.ncbi.nlm.nih.gov/clinvar/intro/>.
- [57] Sequence Variant Nomenclature, "What is the sequence variante nomenclature?" Accessed Nov. 2019, url: <https://varnomen.hgvs.org>.
- [58] Ensembl Genome Browser 98, Ensembl. Accessed Jun. 2019, url: <https://www.ensembl.org/index.html>.
- [59] GWAS Catalog, "GWAS Catalog" Accessed Apr.2019, url: <https://www.ebi.ac.uk/gwas/>.
- [60] Ana León, SILE for Smart Dummies: Instrucciones para la utilización de SILE en entornos de prácticas e introducción al manejo de información genómica, Centro de investigaxción y desarrollo PROS de la UPV, 2019.
- [61] U.S National Library of Medicine, "Welco to Medical Subject Headings". Accessed Oct 2019., url: <https://www.nlm.nih.gov/mesh/meshhome.html>.
- [62] Laboratory for Bioinformatics, Mathematical Modelling and Synthetic Biology, University of Pavia, Italy., url: <http://lab-bioinfo.unipv.it/index.php/en/>.
- [63] Ng D, Johnston JJ, Teer JK, Singh LN, Peller LC, Wynter JS, Lewis KL, Cooper DN, Stenson PD, Mullikin JC, Biesecker LG;, Interpreting secondary cardiac disease variants in an exome cohort., NIH Intramural Sequencing Center (NISC) Comparative Sequencing Program., Aug.

PRESUPUESTO

IDENTIFICACIÓN DE VARIANTES GENÓMICAS MEDIANTE EL MÉTODO SILE EN LA MUERTE SÚBITA CARDIACA: APLICACIONES A LA MEDICINA DE PRECISIÓN

DOCUMENTO II



Nelvia del Cisne Gonza Ajila

Curso académico 2019/2020

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

PRESUPUESTO

1. NECESIDAD DEL PRESUPUESTO

El presupuesto tiene como objetivo principal valorar el coste monetario de la tesis de master desarrollado en esta memoria. El presupuesto tiene en cuenta: el personal, el software y hardware.

La amortización de los costes del software y hardware se calculan utilizando la siguiente formula

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

Donde

- t es el tiempo de uso del equipo (en meses)
- C es el coste del equipo o licencia
- T es el tiempo de amortización (en meses)

2. PRESUPUESTO DESGLOSADO

2.1 Coste del personal

En el coste del personal se tienen en cuenta el coste del ingeniero que realiza el trabajo y las horas que han invertido los tutores (miembros del PROS) a la asesoría del trabajo teniendo en cuenta las horas invertidas.

COSTES DEL PERSONAL			
Perfil	Coste unitario	Horas totales	Coste
Ingeniero Biomédico	15€/hora	500 horas	7.500€
Tutor (Ingeniero informático)	25€/hora	30 horas	750€
Tutor (Ingeniero informático)	25€/hora	15 horas	375€
COSTE DEL PERSONAL TOTAL			8.600€

Tabla 6. Costes del personal del proyecto

2.2 Coste del software

En esta parte del proyecto se detallan los costes respecto de las licencias y las bases de datos. Se incluyen el coste del sistema operativo, del paquete de Microsoft Office y de las bases de datos utilizadas.

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca:
aplicaciones a la medicina de precisión

COSTE DEL SOFTWARE					
Programa	Coste de licencia	Número de licencias	Periodo de uso	Duración de la licencia	Coste imputable
Sistema Operativo MAC	30€	1	6 meses	Indefinida (dos años)	5€
Microsoft Office Hogar y Estudiantes	149€	1	6 meses	Indefinida (dos años)	24,83€
Bases de datos	0€	Acceso libre sin coste	6 meses	Indefinida	0€
COSTE DEL SOFTWARE TOTAL					29,83€

Tabla 7. Coste del software del proyecto

2.2 Coste del hardware

En el coste del hardware se tiene en cuenta el ordenador portátil utilizado MacBook Air.

COSTES DEL HARDWARE					
Equipo	Coste del Equipo (sin IVA)	Unidades	Periodo de amortización	Periodo de uso	Coste imputable (sin IVA)
MacBook Air	1.026 €	1	3 años	6 meses	171€
COSTE DEL HARDWARE TOTAL					171€

Tabla 8. Coste del hardware del proyecto

3. PRESUPUESTO TOTAL

Una vez calculados los presupuestos relacionados con el personal, software y hardware, se procede a calcular el Presupuesto de Ejecución Material. Este presupuesto corresponde a la suma del presupuesto del personal, software y hardware. A continuación, se calculan los gastos generales y los costes de beneficio industrial. Los gastos generales se refieren a los gastos necesarios para llevar a cabo la actividad de la empresa pero que no están relacionados con el producto o servicio. En cuanto al beneficio industrial, se alude a lo que se gana del proyecto. Los gastos generales y beneficio industrial corresponden al 13% y 6% respectivamente y se calcula a partir del Presupuesto de Ejecución Material. La suma del Presupuesto de Ejecución, los gastos generales y el beneficio industrial dan lugar al Presupuesto de ejecución por Contrata. Por último, se añade el IVA y se obtiene el presupuesto final

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca:
aplicaciones a la medicina de precisión

CÁLCULO DEL PRESUPUESTO TOTAL	
Coste del personal total	8.600€
Coste del software total	29.83€
Coste del hardware total	171€
Total del presupuesto de ejecución material	8.800,83€
Gastos generales (13%)	1.144,10€
Beneficio industrial (6%)	528,05€
Total del presupuesto de Ejecución por Contrata	10.472,98€
IVA (21%)	2.199,32
PRESUPUESTO TOTAL	12.672,3€

Tabla 9. Costes del presupuesto total

El presupuesto total del proyecto es de **12.672,**

ANEXOS

IDENTIFICACIÓN DE VARIANTES GENÓMICAS MEDIANTE EL MÉTODO SILE EN LA MUERTE SÚBITA CARDIACA: APLICACIONES A LA MEDICINA DE PRECISIÓN

DOCUMENTO III

Nelvia del Cisne Gonza Ajila
Curso académico 2019/2020

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

Anexo 1. Subcategorías de la cardiomiopatía dilatada familiar.

- **CARDIOMYOPATHY, DILATED, 1B**; CMD1B: *CARDIOMYOPATHY, FAMILIAL DILATED, 1; CMPD1CARDIOMYOPATHY, FAMILIAL DILATED; FDC.*
- **Dilated cardiomyopathy 1A**: *CARDIOMYOPATHY, CONGESTIVE; CARDIOMYOPATHY, DILATED, WITH CONDUCTION DEFECT 1; Cardiomyopathy, Familial Idiopathic; CMD1A; Dilated cardiomyopathy with conduction defect; Dilated Cardiomyopathy with Quadriceps Myopathy; Familial dilated cardiomyopathy with conduction defect due to LMNA mutation; Idiopathic dilated cardiomyopathy; LMNA-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1AA**: *ACTN2-Related Dilated Cardiomyopathy; CARDIOMYOPATHY, DILATED, 1AA, WITH OR WITHOUT LEFT VENTRICULAR NONCOMPACTION; CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 23, WITH OR WITHOUT LEFT VENTRICULAR NONCOMPACTION; CMD1AA.*
- **Dilated cardiomyopathy 1BB**: *CARDIOMYOPATHY, DILATED, 1BB, SUSCEPTIBILITY TO; CMD1BB; DSG2-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1C**: *CARDIOMYOPATHY, DILATED, 1C, WITH LEFT VENTRICULAR NONCOMPACTION; CARDIOMYOPATHY, DILATED, 1C, WITH OR WITHOUT LEFT VENTRICULA NONCOMPACTION; Cardiomyopathy, dilated, with left ventricular non-compaction; CMD1C; LDB3-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1CC**: *CMD1CC; NEXN-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1DD**: *CMD1DD; RBM20-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1E**: *CARDIOMYOPATHY, DILATED, WITH CONDUCTION DEFECT 2; CARDIOMYOPATHY, DILATED, WITH CONDUCTION DISORDER AND ARRHYTHMIA; CMD1E; SCN5A-Associated Dilated Cardiomyopathy; SCN5A-Related Dilated Cardiomyopathy; SCN5A-Related Disorders.*
- **Dilated cardiomyopathy 1EE**: *CMD1EE; MYH6-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1FF**: *Cardiomyopathy, dilated, 1N; CMD1FF.*
- **Dilated cardiomyopathy 1G**: *CMD1G; TTN-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1GG**: *CMD1GG.*
- **Dilated cardiomyopathy 1H**: *CARDIOMYOPATHY, DILATED, WITH CONDUCTION DEFECT; CMD1H.*
- **Dilated cardiomyopathy 1HH**: *BAG3-Related Dilated Cardiomyopathy; CMD1HH*
- **Dilated cardiomyopathy 1I** (UID 397998): *CMD1I; DES-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1II**: *CMD1II.*
- **Dilated cardiomyopathy 1J**: *CARDIOMYOPATHY, DILATED, WITH SENSORINEURAL HEARING LOSS, AUTOSOMAL DOMINANT; CMD1J; EYA4-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1JJ**: *CMD1JJ.*
- **Dilated cardiomyopathy 1K**: *CMD1K.*
- **Dilated cardiomyopathy 1L**: *CMD1L; SGCD-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1M**: *CMD1M; CSRP3-Related Dilated Cardiomyopathy.*

- **Dilated cardiomyopathy 1N:** *CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 25; CMH25; TCAP-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1O:** *CMD1P; PLN-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1P:** *CARDIOMYOPATHY, DILATED, WITH VENTRICULAR TACHYCARDIA; CMD1O.*
- **Dilated cardiomyopathy 1Q:** *CMD1Q.*
- **Dilated cardiomyopathy 1R:** *ACTC1-Related Dilated Cardiomyopathy; **CMD1R.***
- **Dilated cardiomyopathy 1S:** *CMD1S; Left ventricular non-compaction 5; MYH7-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1T:** *CMD1T; TMPO-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1V:** *CMD1V; PSEN2-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1W:** *CMD1W; VCL-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1X (UID 370583):** *CARDIOMYOPATHY, DILATED, WITH MILD OR NO PROXIMAL MUSCLE WEAKNESS; CMD1X; FKTN-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1Y:** *CMD1Y; TPM1-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 1Z:** *CMD1Z; TNNC1-Related Dilated Cardiomyopathy.*
- **Dilated cardiomyopathy 2A:** *CARDIOMYOPATHY, CONGESTIVE, AUTOSOMAL RECESSIVE; CARDIOMYOPATHY, DILATED, AUTOSOMAL RECESSIVE; CMD2A; TNNI3-Related Dilated Cardiomyopathy.*
- **Left ventricular non-compaction 6:** *CMD1D; Dilated cardiomyopathy 1D; TNNT2-Related Dilated Cardiomyopathy.*

Anexo 2. Subcategorías de la cardiomiopatía hipertrófica familiar.

- **CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 1;** *CMH1: CMH VENTRICULAR HYPERTROPHY, HEREDITARY ASYMMETRIC SEPTAL HYPERTROPHY; ASH HYPERTROPHIC SUBAORTIC STENOSIS, IDIOPATHIC.*
- **Cardiomyopathy, hypertrophic, midventricular, digenic (UID 864707):** *CARDIOMYOPATHY, HYPERTROPHIC, MIDVENTRICULAR, DIGENIC.*
- **Familial hypertrophic cardiomyopathy 1:** *Cardiomyopathy, familial hypertrophic; CAV3-Related Hypertrophic Cardiomyopathy; CMH1; MYH7-Related Familial Hypertrophic Cardiomyopathy.*
- **Familial hypertrophic cardiomyopathy 2:** *CMH2; TNNT2-Related Familial Hypertrophic Cardiomyopathy.*
- **Familial hypertrophic cardiomyopathy 3:** *CMH3; TPM1-Related Familial Hypertrophic Cardiomyopathy.*
- **Familial hypertrophic cardiomyopathy 4:** *CMH4.*
- **Familial hypertrophic cardiomyopathy 6:** *CMH6; Familial Hypertrophic Cardiomyopathy with Wolff-Parkinson-White Syndrome, PRKAG2-Related.*
- **Familial hypertrophic cardiomyopathy 7:** *CARDIOMYOPATHY, FAMILIAL HYPERTROPHIC, 7, MODIFIER OF; CMH7; TNNI3-Related Familial Hypertrophic Cardiomyopathy.*

- **Familial hypertrophic cardiomyopathy 8:** CARDIOMYOPATHY, HYPERTROPHIC, MID-LEFT VENTRICULAR CHAMBER TYPE, 1; CMH8; MYL3-Related Familial Hypertrophic Cardiomyopathy.
- **Familial hypertrophic cardiomyopathy 9:** CMH9.
- **Familial hypertrophic cardiomyopathy 10:** CARDIOMYOPATHY, HYPERTROPHIC, MID-LEFT VENTRICULAR CHAMBER TYPE, 2; CMH10; MYL2-Related Familial Hypertrophic Cardiomyopathy.
- **Familial hypertrophic cardiomyopathy 11:** ACTC1-Related Familial Hypertrophic Cardiomyopathy; CMH11.
- **Familial hypertrophic cardiomyopathy 12:** ACTC1-Related Familial Hypertrophic Cardiomyopathy; CMH11.
- **Familial hypertrophic cardiomyopathy 13:** CMH13; TNNC1-Related Familial Hypertrophic Cardiomyopathy.
- **Familial hypertrophic cardiomyopathy 14:** CMH14.
- **Familial hypertrophic cardiomyopathy 15:** CMH15.
- **Familial hypertrophic cardiomyopathy 16:** CMH16.
- **Familial hypertrophic cardiomyopathy 17:** CMH17.
- **Familial hypertrophic cardiomyopathy 18:** CMH18.
- **Familial hypertrophic cardiomyopathy 19:** CMH19.
- **Familial hypertrophic cardiomyopathy 20:** CMH20.
- **Familial hypertrophic cardiomyopathy 21:** CMH21.

Anexo 3. Conjunto de variantes examinadas por un panel de expertos para la miocardiopatía hipertrófica.

Variantes examinadas por un panel de expertos.				
rs121434594	rs397516201	rs267606908	rs727503261	rs121913624
rs200939753	rs727503246	rs397516161	rs121913630	rs397516269
rs369437262	rs141764279	rs36211715	rs121913641	rs397516209
rs193922390	rs145213771	rs138049878	rs121913638	rs186964570
rs141122361	rs45478699	rs727504310	rs371898076	rs121913631
rs3729823	rs587782962	rs397516153	rs121913626	rs121913632
rs201307101	rs45540831	rs376754645	rs397516101	rs397516101

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

Anexo 4. Conjunto de variantes finales para la muerte súbita cardíaca (GWAS Catalog)

Nº	IDENTIFICADOR	FENOTIPO O CAUSA	TAMAÑO DE LA MUESTRA INICIAL	TAMAÑO DE REPLICACIÓN DE LA MUESTRA	GEN	P VALOR	OR	CI	PUBMED ID
1	rs4665058	Muerte súbita	1.283 casos y 20.000 controles	3.119 casos y 11.146 controles	BAZ2B	2.00 E-10	1.92	[1.57-2.34]	21738491
2	rs2234962	Miocardiopatía dilatada	1.179 casos y 1.108 controles	1.165 casos y 1.302 controles	BAG3	4.00E-12	1.52	[1.22-1.89]	21459883
3	rs10927875	Miocardiopatía dilatada	1.179 casos y 1.108 controles	1.165 casos y 1.302 controles	ZBTB17	1.00E-09	1.32	[1.19-1.43]	21459883
4	rs11708996	Síndrome de Brugada	312 casos y 1.115 controles	594 casos y 806 controles, 208 casos japoneses, 1.1016 controles japoneses	SCN5A	1.00E-14	1.73	[1.51-1.99]	23872634
5	rs10428132	Síndrome de Brugada	312 casos y 1.115 controles	594 casos y 806 controles, 208 casos, 1.1016 controles japoneses	SCN10A	1.00E-68	2.55	[2.30-2.84]	23872634
6	rs9388451	Síndrome de Brugada	312 casos y 1.115 controles	594 casos y 806 controles, 208 casos japoneses y 1.1016 controles japoneses	HEY2	5.00E-17	1.58	[1.42-1.75]	23872634
7	rs4487645	Miocardiopatía hipertrófica	1.229 casos y	-	DNAH11	2.00E-	1.35	[1.22-	28025584

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardíaca:
aplicaciones a la medicina de precisión

			7.526 controles			09		1.49]	
8	rs79419269	Miocardopatía hipertrófica	1.229 casos y 7.526 controles	-	SMARCD3	5.00E- 08	1.41	[1.25- 1.60]	28025584

Anexo 5. Variantes finales para la muerte súbita cardíaca.

VARIANTES FINALES PARA LA MUERTA SÚBITA						
Nº	ID	Fenotipo o causa	Cambio	Gen	PMID	BASE DE DATOS
1	rs4665058	Infarto de miocardio	c.5689-1012T>G	BAZ2B	21738491	GWAS
2	rs267607004	Miocardopatía dilatada	c.1907G>A	RBM20	19712804 23861363 24503780	CLINVAR
3	rs2234962	Miocardopatía dilatada	c.451T>C	BAG3	21459883	GWAS
4	rs10927875	Miocardopatía dilatada	c.-3+222G>A	ZBTB17	21459883	GWAS
5	rs796051885	Síndrome de Loeys-Dietz	c.898C>T	TGFB3	25835445	ENSEMBL
6	rs199473180	Síndrome de Brugada	c.2893C>T	SCN5A	23321620	CLINVAR
7	rs199473153	Síndrome de Brugada	c.2254G>A	SCN5A	25650408	CLINVAR
8	rs121912776	Síndrome de Brugada	c.116C>T	CACNA1C	20817017	CLINVAR
9	rs72549410	Síndrome de Brugada	c.1231G>A	SCN5A	27041150	ENSEMBL
10	rs137854601	Síndrome de Brugada	c.5350G>A	SCN5A	24784157	ENSEMBL

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca:
aplicaciones a la medicina de precisión

11	rs137854611	Síndrome de Brugada	c.2204C>T /c.2204C>A	SCN5A	11076825	ENSEMBL
12	rs11708996	Síndrome de Brugada	c.2263-4859C>G	SCN5A	23872634	GWAS
13	rs10428132	Síndrome de Brugada	c.2280+3452A>C	SCN10A	23872634	GWAS
14	rs9388451	Síndrome de Brugada	g.125769231T>A	HEY2	23872634	GWAS
15	rs397515848	Síndrome de Marfan	c.7180C>T	FBN1	19863550 27724990	ENSEMBL
16	rs140630	Síndrome de Marfan	c.4930C>T	FBN1	15241795	CLINVAR
17	rs111231312	Síndrome de Marfan	c.4615C>T	FBN1	19863550 15241795 22772377	CLINVAR
18	rs387906697	Síndrome de Marfan	c.1240C>T	TGFBR1	21358634	ENSEMBL
19	rs397516457	Miocardiopatía hipertrófica	c.281G>A	TNNT2	20624503 20031602	CLINVAR
20	rs121964856	Miocardiopatía hipertrófica	c.275G>A	TNNT2	22144547 19150014 18533079	CLINVAR
21	rs727502886	Miocardiopatía hipertrófica	c.355G>A	ACTN2	20022194	CLINVAR
22	rs786204951	Miocardiopatía hipertrófica	c.1883A>G	ACTN2	20022194	CLINVAR
23	rs387906397	Miocardiopatía hipertrófica	c.3330+2T>C; c.3330+2T>G	MYBPC3	25356970 17937428 18467358	CLINVAR
24	rs187830361	Miocardiopatía hipertrófica	c.2374T>C	MYBPC3	24793961 23074333	CLINVAR

Identificación de variantes genómicas mediante el método SILE en la muerte súbita cardiaca:
aplicaciones a la medicina de precisión

25	rs397515907	Miocardiopatía hipertrófica	c.1505G>A	MYBPC3	20433692 22267749	CLINVAR
26	rs397515905	Miocardiopatía hipertrófica	c.1483C>G	MYBPC3	22267749 20433692 19659763 19150014 20624503	CLINVAR
27	rs397514752	Miocardiopatía hipertrófica	c.1469G>T	MYBPC3	23840593	CLINVAR
28	rs397516068	Miocardiopatía hipertrófica	c.655G>C	MYBPC3	20031618 23690394 20624503 23074333 20031602 25351510	CLINVAR
29	rs727504275	Miocardiopatía hipertrófica	c.611G>A	TNNI3	16199542	CLINVAR
30	rs397516357	Miocardiopatía hipertrófica	c.557G>A	TNNI3	26169204 20624503	CLINVAR
31	rs397516347	Miocardiopatía hipertrófica	c.422G>	TNNI3	12860912 22429680	CLINVAR
32	rs104894204	Miocardiopatía hipertrófica	c.172T>G	CSRP3	24705329	ENSEMBL
33	rs397514616	Miocardiopatía hipertrófica	c.91G>T	TNNC1	22815480	ENSEMBL
34	rs727504246	Miocardiopatía hipertrófica	c.566C>T	TNNT2	12473556	ENSEMBL
35	rs104894724	Miocardiopatía hipertrófica	c.433C>T	TNNI3	17599605 16020591	CLINVAR
36	rs4487645	Miocardiopatía hipertrófica	c.13050-714C>A	DNAH11	28025584	GWAS
37	rs79419269	Miocardiopatía hipertrófica	c.1173+122A>G	SMARCD3	28025584	GWAS
38	rs727503504	Miocardiopatía restrictiva	c.508C>T	TNNI3	20031618	CLINVAR