



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Industrial

Diseño y desarrollo de una fuente de datos sobre estudios
funcionales asociados al criterio PS3/BS3 de las guías
ACMG-AMP 2015 y aplicación en cardiopatías familiares

Trabajo Fin de Grado

Grado en Ingeniería Biomédica

AUTOR/A: Martínez Miguel, Marina

Tutor/a: Costa Sánchez, Mireia

Cotutor/a: Pastor López, Oscar

CURSO ACADÉMICO: 2022/2023

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

AGRADECIMIENTOS

Gracias, por supuesto, a mi tutora, Mireia Costa, que ha estado constantemente pendiente y me ha ayudado en todo lo que necesitaba. Gracias a mi cotutor, Óscar Pastor, que, junto a los compañeros del grupo PROS, me recibieron calurosamente el año pasado y han seguido apoyándome durante todo el recorrido.

Quiero agradecer a mis padres su apoyo y cuidado eterno, y que hayan aguantado mi poca comunicación y la distancia durante el curso. Os quiero mucho, siempre.

Por último, agradecer a mis compañeros y amigos, sean de hace años o recientes, que han estado conmigo en todas las experiencias de este curso y de la carrera. Sobre todo, gracias a Ana, por compartir conmigo el estrés del grupo de música, sacarme de casa cuando lo necesitaba y ayudarme siempre; y, por supuesto, gracias a Lledó y Lorena, mis dos ángeles de la guarda durante toda la carrera, porque somos y seguiremos siendo el Trío La La La.

RESUMEN

En 2015, el *American College of Medical Genetics and Genomics* y la *Association for Molecular Pathology* desarrollaron unas guías describiendo criterios para clasificar la patogenicidad de variaciones genómicas en enfermedades mendelianas. En el criterio PS3-BS3 se evalúan consecuencias genómicas, proteómicas y metabólicas de las variaciones, evidencia proporcionada por estudios funcionales. Sin embargo, esta evidencia únicamente está disponible en literatura científica, lo que dificulta la extracción de conocimiento y automatización de la evaluación del criterio.

Por este motivo, este Trabajo de Fin de Grado tiene como objetivo diseñar y desarrollar una fuente de información de estudios funcionales que permita automatizar la evaluación del criterio PS3-BS3 de las guías ACMG/AMP 2015. Como caso de estudio, este trabajo se centra en estudios asociados a cardiopatías familiares, enfermedades heredables de causa genética que afectan al corazón.

Primeramente, se analiza la información necesaria para evaluar el cumplimiento del criterio PS3-BS3 y se realiza un modelo conceptual para guiar la identificación de esta información en la literatura y estructurar el conocimiento extraído. A continuación, se realiza un curado manual de la literatura reportando estudios funcionales sobre 35 genes, seleccionados por expertos clínicos, asociados con las cardiopatías familiares. Finalmente, a partir de la información extraída, se genera la fuente de información. Como caso de uso se plantea una automatización de evaluación del criterio a partir de un VCF dado, validándola con VCF reales de pacientes de cardiología del Hospital General Universitario de Alicante.

De esta manera, en este Trabajo de Fin de Grado se proponen las herramientas necesarias para la evaluación del criterio PS3-BS3 de las guías ACMG/AMP 2015, de forma que pueda aplicarse como herramienta de ayuda al diagnóstico clínico.

Palabras clave: Estudios funcionales, guías ACMG-AMP, cardiopatías familiares, variaciones genómicas

RESUM

En 2015, el *American College of Medical Genetics and Genomics* i la *Association for Molecular Pathology* van desenvolupar unes guies descrivint criteris per a la classificació de la patogenicitat de variacions genòmiques en malalties mendelianes. En el criteri PS3-BS3 s'avaluen les conseqüències genòmiques, proteòmiques i metabòliques de les variacions, evidència proporcionada per estudis funcionals. Malgrat això, aquesta evidència està disponible únicament en literatura científica, el que dificulta l'extracció de coneixement i automatització de l'avaluació del criteri.

Per aquest motiu, aquest Treball de Fi de Grau té com objectiu dissenyar i desenvolupar una font d'informació d'estudis funcionals que permeten automatitzar l'avaluació del criteri PS3-BS3 de les guies ACMG/AMP 2015. Com a cas d'estudi, aquest treball es centrarà en estudis associats a cardiopaties familiars, malalties heretables de causa genètica que afecten al cor.

Primerament, s'analitza la informació necessària per a avaluar el compliment del criteri PS3-BS3 i es realitza un model conceptual per a guiar la identificació d'aquesta informació a la literatura i estructurar el coneixement extret. A continuació, es realitza un curat manual de la literatura, reportant estudis funcionals de 35 gens, seleccionats per experts clínics, associats amb cardiopaties familiars. Finalment, a partir de la informació extreta, es genera la font d'informació. Com a cas d'ús, es planteja una automatització d'avaluació del criteri partint d'un VCF, validant-la amb VCF reals de pacients de cardiologia de l'Hospital General Universitari d'Alacant.

D'aquesta manera, en aquest Treball de Fi de Grau es proposen les ferramentes necessàries per a l'avaluació del criteri PS3-BS3 de les guies ACMG/AMP 2015, de forma que poden aplicar-se com a ferramenta d'ajuda al diagnòstic clínic.

Paraules clau: Estudis funcionals, guies ACMG-AMP, cardiopaties familiars, variacions genòmiques

Títol: Disseny i desenvolupament d'una font de dades sobre estudis funcionals associats al criteri PS3/BS3 de les guies ACMG-AMP 2015 i aplicació en cardiopaties familiars.

ABSTRACT

In 2015, the *American College of Medical Genetics and Genomics* and the *Association for Molecular Pathology* defined a set of guidelines and criteria to classify the pathogenicity of genomic variants in mendelian diseases. The PS3-BS3 criterion evaluates the variants' genomic, proteomic, and metabolic consequences from evidence provided by functional studies. However, such evidence is only available in scientific literature, which complicates the knowledge extraction and the automatization of the criterion evaluation.

Therefore, the objective of this Final Degree Thesis is to design and develop an information resource of functional studies that allows the automatization of the PS3-BS3 criterion evaluation as defined by the ACMG/AMP 2015 Guidelines. As a case study, this work will focus on hereditary heart diseases, defined as hereditary genetically caused pathologies which affect the heart.

Firstly, the information required to evaluate the PS3-BS3 criterion fulfillment was analyzed, and a conceptual model was built to guide the identification of said information in literature resources and structure the extracted knowledge. Next, a manual curation of literature was conducted, reporting functional studies on 35 genes selected by experts which were associated with the diseases considered. Finally, with the extracted information, the information resource was generated. As a case of use, a proposal for the automatization of the criterion evaluation given an VCF was developed and then validated with real-life VCF from cardiology patients provided by Hospital General Universitario de Alicante.

By this process, the present Final Degree Thesis proposes the necessary tools for the evaluation of the ACMG/AMP 2015 Guidelines PS3-BS3 criterion in a way that could be implemented as a diagnostic decision support tool.

Keywords: Functional studies, ACMG-AMP guidelines, hereditary heart diseases, genomic variations

Title: Design and development of a functional studies data resource related to the PS3/BS3 criteria from the 2015 ACMG-AMP guidelines and its application for hereditary heart diseases.

ÍNDICE

DOCUMENTOS

Documento I. Memoria

Documento II. Presupuesto

Documento III. Anexos

ÍNDICE DE LA MEMORIA

1.	INTRODUCCIÓN.....	1
1.1.	Medicina de Precisión	1
1.2.	Guías ACMG/AMP	2
1.2.1.	<i>Criterios PS3 y BS3</i>	3
1.2.2.	<i>Introducción a los estudios funcionales</i>	3
1.3.	Cardiopatías Familiares	4
1.4.	Estructura del trabajo.....	5
2.	OBJETIVOS	6
3.	ESTADO DEL ARTE	7
3.1.	Bases de datos existentes	7
3.2.	Mejoras respecto al estado del arte	8
4.	METODOLOGÍA.....	9
5.	INVESTIGACIÓN DEL PROBLEMA.....	11
5.1.	Estudios funcionales: Caracterización y problemática.....	11
5.2.	Modelado conceptual	13
5.2.1.	<i>CSHG</i>	13
6.	DISEÑO Y DESARROLLO DE LA FUENTE DE INFORMACIÓN.....	15
6.1.	Diseño de la fuente de información	15
6.1.1.	<i>Ampliación del CSHG</i>	15

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

6.1.2. Estructura de la fuente de información.....	17
6.2. Desarrollo de la fuente de información	18
6.2.1. Primera búsqueda en repositorios	19
6.2.2. Segunda búsqueda y selección final	21
6.2.3. Extracción de información.....	22
6.3. Revisión y estandarización	25
7. RESULTADOS	28
7.1. Análisis de resultados.....	28
7.2. Caso de uso con VCF de pacientes reales	31
7.2.1. Implementación.....	31
7.2.2. Resultados caso de uso.....	32
7.2.3. Limitaciones.....	34
8. CONCLUSIONES Y TRABAJOS FUTUROS	36
8.1. Cumplimiento de objetivos iniciales	36
8.2. Trabajos futuros	38
9. BIBLIOGRAFÍA.....	40

ÍNDICE DEL PRESUPUESTO

1. OBJETIVOS DEL PRESUPUESTO.....	1
2. PRESUPUESTO DESGLOSADO	1
3. PRESUPUESTO TOTAL.....	3

ÍNDICE DE ANEXOS

ANEXO 1. FUENTE DE INFORMACIÓN DE ESTUDIOS FUNCIONALES ASOCIADA A CARDIOPATÍAS FAMILIARES.....	6
ANEXO 2. CÓDIGO DE PYTHON EMPLEADO EN EL CASO DE USO	20

ÍNDICE DE FIGURAS

Figura 1. Representación del Ciclo de Diseño, adaptado a partir de la metodología del Design Science, que se ha seguido en este Trabajo de Fin de Grado	10
Figura 2. Árbol de decisión para la evaluación de información funcional según las métricas de validación definidas por el grupo ClinGen (Brnich et al., 2020).	12

Figura 3. Diagrama de clases de la ampliación correspondiente a los estudios funcionales (marcados en azul) y sus conexiones directas al resto del CSHG.	16
Figura 4. Capturas de ejemplo de la estructura de la fuente de información.....	18
Figura 5. Flujo de trabajo completo del desarrollo de la fuente de información. Detalla la búsqueda en repositorios de artículos, los filtros y selección aplicados y la extracción de información relevante.	18
Figura 6. Ampliación del primer filtro, el cual se ha aplicado tanto en la primera como en la segunda búsqueda.	20
Figura 7. Ampliación del segundo filtro aplicado durante la lectura de los artículos.	22
Figura 8. Comparación de cantidad de artículos asociados a los genes y cantidad de variaciones que finalmente se han recogido en la extracción de información.	29
Figura 9. Cantidad de variaciones estudiadas de cada gen y las cardiopatías a las que se han asociado.	30

ÍNDICE DE TABLAS

Tabla 1. Reestructuración del campo “Variant Name” a tres columnas, con ejemplos de cada formato	26
Tabla 2. Reestructuración de los campos sobre controles y réplicas: “positiveControls” y “negativeControls” indican el número de controles, “PosControl” y “negControl” describen el tipo, “Biological replicates” y “Technical replicates” indican el número de réplicas y “replicates” indica el tipo.	26
Tabla 3. Resumen de resultados en la búsqueda inicial y la segunda búsqueda en bruto (apuntados directamente de los resultados de la consulta, sin comprobar los duplicados) y filtrados (tras eliminar duplicaciones).	28
Tabla 4. Cantidades finales de genes y artículos tras la lectura, donde se han descartados más artículos (ver Filtro 2, Figura 7).	29
Tabla 5. Número de variaciones asociadas a cada enfermedad.....	31
Tabla 6. Resumen de las 10 variaciones de la fuente de información identificadas en los pacientes reales.....	33

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Tabla 7. <i>Desglose de costes de personal asociados al trabajo realizado</i>	2
Tabla 8. <i>Desglose de costes de software asociados al trabajo realizado.</i>	2
Tabla 9. <i>Desglose de costes de hardware asociados al trabajo realizado.</i>	3
Tabla 10. <i>Presupuestos finales de este Trabajo de Fin de Grado.</i>	4

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

DOCUMENTO I. MEMORIA

Diseño y Desarrollo de una Fuente de Datos sobre Estudios Funcionales asociados al Criterio PS3/BS3 de las Guías ACMG-AMP 2015 y Aplicación en Cardiopatías Familiares

Marina Martínez Miguel

Grado en Ingeniería Biomédica

Curso académico 2022/2023

1. INTRODUCCIÓN

La medicina de precisión es un campo en crecimiento constante, pero de gran complejidad. Una de las actividades fundamentales es la interpretación de variaciones genéticas para determinar sus efectos y consecuencias clínicas. Con el propósito de facilitar esta interpretación y evaluación, en este Trabajo de Fin de Grado se va a diseñar y desarrollar una fuente de información de los estudios funcionales realizados sobre variaciones genéticas asociadas a cardiopatías familiares. Esta fuente recopilará los datos genómicos disponibles y los estructurará a partir de un modelo conceptual, de forma que se pueda automatizar la evaluación del denominado criterio PS3/BS3 de las Guías ACMG-AMP 2015.

Por tanto, en este capítulo se describirá, primero, el contexto de la medicina de precisión y la interpretación de variaciones genéticas. Seguidamente, se tratarán los conceptos clave del trabajo asociados a dicha interpretación; en concreto, las Guías ACMG/AMP, los criterios PS3/BS3, los estudios funcionales (y su relación con los criterios) y el contexto clínico seleccionado de cardiopatías familiares. Finalmente, se definirá la estructura que seguirá esta memoria.

1.1. Medicina de Precisión

El estudio del ADN humano está en progreso constante y, como resultado, se están generando cantidades ingentes de información en el ámbito genómico que cuentan con un gran potencial para la aplicación clínica. En el uso de esta información se basa la medicina de precisión, que se define como la prevención, el diagnóstico o el tratamiento de una enfermedad en función de la genética de cada paciente (Hurtado, 2022). De esta manera, la medicina de precisión busca, a partir del perfil genético de un individuo, conocer la presencia o predisposición a cualquier enfermedad.

Como es de esperar, este campo presenta una gran complejidad. Una de las causas principales es la variabilidad genética entre individuos. Esta variabilidad se asocia a uno de los conceptos genómicos fundamentales para la medicina de precisión, las variaciones genéticas. Las variaciones genéticas son cambios en la secuencia de elementos básicos (nucleótidos) del ADN y son los que nos diferencian como individuos, pues cada ser humano cuenta con, aproximadamente, entre 2100 y 2500 variaciones en su ADN (Auton et al., 2015). El ADN codifica todas las proteínas del cuerpo, por lo que estos cambios genéticos conllevan consecuencias de gravedads varias, ya que pueden ser totalmente inocuos, sin afectar a las proteínas resultantes, o ser causantes directos de enfermedad como consecuencia de provocar proteínas defectuosas. Este último caso es lo que se denomina una variación patogénica, y resulta, concretamente, en las enfermedades conocidas como mendelianas, patologías hereditarias causadas por una única variación genética. Por tanto, en el ámbito de la medicina de precisión, se crea una importante necesidad de investigación del efecto de estas alteraciones genéticas para poder explotar este conocimiento de cara a la aplicación clínica.

El proceso de determinar el efecto de una variación en el desarrollo de enfermedad se conoce como interpretación. Es un proceso complejo, que requiere analizar y comparar una gran variedad de información genómica. A esto, se suma la heterogeneidad de esta información, su dispersión en múltiples fuentes, junto con una considerable variabilidad de la calidad de datos y de terminologías empleadas. Dicha heterogeneidad hace evidente la necesidad de una estandarización de la calidad y terminología para facilitar el desarrollo e interpretación de las investigaciones.

Buscando dar soporte al proceso de interpretación, se han realizado esfuerzos por definir estándares como las Guías ACMG-AMP, que se explican en la siguiente sección, y que constituyen el punto de partida de este trabajo.

1.2. Guías ACMG/AMP

Para revisar las técnicas de interpretación de variaciones genéticas, en 2013 se creó un grupo de trabajo formado por miembros del *American College of Medical Genomics and Genomics* y de la *Association for Molecular Pathology*. Tras su investigación, publicaron en 2015 las Guías ACMG-AMP (Richards et al., 2015). Estas recomendaciones son un conjunto de criterios para evaluar la patogenicidad, es decir, la capacidad de causar enfermedad, de variaciones asociadas a enfermedades mendelianas en función de la evidencia disponible en diferentes bases de datos y en literatura científica. De esta manera, estas guías tratan de proporcionar a todos los expertos una base común para determinar si una variación es potencialmente la causa de una enfermedad específica.

Sin embargo, es importante destacar que el papel de las variaciones genéticas no es una cuestión binaria. Una variación puede considerarse patogénica o benigna con mayor o menor confianza según la gravedad de su efecto y las condiciones en las que se ha dado. Por ejemplo, una sustitución de un nucleótido puede no alterar el aminoácido (elementos unidad que forman la proteína) resultante y, por tanto, no afectar al funcionamiento de la proteína, o, en cambio, puede provocar la aparición de un codón de parada¹ que provoca una proteína incompleta y defectuosa. Para permitir una interpretación común de esta escala de patogenicidad, las Guías ACMG-AMP 2015 proponen un estándar de clasificación de las variaciones en cinco categorías de significancia clínica:

- **Patogénica.** Causante de enfermedad con probabilidad > 95%.
- **Probablemente patogénica.** Causante de enfermedad con probabilidad > 90%.
- **Benigna.** No causante de enfermedad con probabilidad superior > 95%.
- **Probablemente benigna.** No causante de enfermedad con probabilidad > 90%.
- **Significado incierto.** Evidencia conflictiva o insuficiente.

A la hora de realizar esta clasificación, se evalúan veintiocho criterios, 16 de patogenicidad y 12 de benignidad, cada uno de los cuales evalúa distintos tipos de información sobre las variaciones. Según el tipo de información, cada criterio tendrá un peso “muy fuerte”, “moderado” o “de apoyo” en la clasificación de la variación. Finalmente, según las combinaciones de la evaluación de estos criterios, se podrá clasificar la variación siguiendo unas normas específicas dictadas en las guías.

Concretamente, este Trabajo de Fin de Grado se focalizará en el criterio PS3 y su contraparte, el criterio BS3. Estos dos criterios evalúan un tipo de análisis concreto: los denominados estudios funcionales. Por tanto, las siguientes subsecciones se dedicarán a definir los dos criterios y los estudios funcionales.

¹ Codón de parada: conjunto de nucleótidos que, en lugar de dar lugar a un aminoácido, detiene el proceso de transformación del ARN a proteína.

1.2.1. Criterios PS3 y BS3

Los criterios PS3 y BS3 son criterios complementarios que se centran en la evidencia proveniente de estudios funcionales (detallados en la subsección 1.2.2) y que forman parte del conjunto de criterios que proporcionan una fuerte evidencia de patogenicidad (PS3) y benignidad (BS3) de las Guías ACMG/AMP.

Por una parte, el criterio PS3 abarca los estudios funcionales bien establecidos, de tipo *in vitro* (realizado en un laboratorio) o *in vivo* (realizado en el interior de un organismo vivo), que apoyan el efecto dañino de un gen o de un producto de un gen. Análogamente, el criterio BS3 concierne a los estudios funcionales bien establecidos, *in vitro* o *in vivo*, que no muestran ningún efecto dañino en la funcionalidad de las proteínas o el *splicing* – corte y empalme de las zonas codificantes del ARN para formar el precursor de la proteína – del ARN (Brnich et al., 2020). Adicionalmente, las Guías ACMG/AMP notan que el término de “estudios bien establecidos” hace referencia a aquellos estudios que han sido validados y cuya reproducibilidad y robustez en un laboratorio de diagnóstico clínico han sido demostradas.

La gran influencia de estos criterios se debe a la importancia de la información funcional. Conocer con certeza el efecto de una variación genética en los mecanismos biológicos marca una diferencia considerable al clasificar dicha variación. Es decir, una demostración experimental de que una variación es definitivamente benigna o patogénica es un resultado irrefutable, teniendo en cuenta la complejidad asociada. En consecuencia, estos criterios también son una herramienta importante en la reclasificación de variaciones de significado incierto. Este hecho se muestra claramente en un estudio realizado para cuantificar el potencial de la información funcional para la reclasificación de variaciones (Brnich et al., 2018), donde se calculó que al incluir datos funcionales se podían reclasificar el 94% de las variaciones según las normas teóricas establecidas por las Guías ACMG/AMP 2015, o el 76% si se aplicaba la ponderación Bayesiana de Tavtigian et al., 2018 que permite modelar cuántas pruebas son necesarias para superar criterios conflictivos (Brnich et al., 2018).

El impacto de ambos criterios es evidente, y su evaluación depende de la información funcional disponible, la cual se extrae de los estudios funcionales que definiremos en la siguiente sección y que caracterizaremos en la sección 5.1.

1.2.2. Introducción a los estudios funcionales

La genómica funcional tiene como objetivo desarrollar y aplicar enfoques experimentales para evaluar la función de los genes combinando metodologías experimentales con el análisis estadístico y computacional de los datos (Hieter & Boguski, 1997). En esta rama, surgen los denominados estudios funcionales, un tipo de estudios que se realizan para determinar las consecuencias funcionales de las variaciones genéticas (Rodenburg, 2018). Engloba los experimentos realizados para determinar cómo los genes contribuyen a procesos biológicos y son una forma de definir el mecanismo molecular por el que una variación tiene las consecuencias que tiene, sean benignas (inocua) o patogénicas (causante de enfermedad).

Este tipo de análisis puede realizarse de múltiples formas. Por una parte, hay distintos niveles moleculares en los que se puede realizar el experimento, ya sea genómico (estudia la estructuración del genoma y cromosomas), ADN (estudia secuencias de nucleótidos), ARN (el producto intermedio

entre ADN y proteína), proteínas (estudia la proteína resultante) o metabolitos (sustancias de procesos biológicos). Por otro lado, el enfoque es distinto según el método experimental empleado. Al nivel más general, podemos distinguir entre experimentos *in vivo* (realizado en el interior de un organismo vivo), *ex vivo* (realizado en una muestra proveniente de un organismo vivo) o *in vitro* (realizado en un laboratorio), incluso *in silico* (modelo o simulación computacional). A nivel más específico, existe una mayor variedad de tipos de experimento. Los enfoques más comunes son experimentos de rescate (introducir un gen o proteína funcional para comprobar si puede restablecer la función normal), edición genética mediante la herramienta CRISPR/Cas9, biomarcadores (aspectos o elementos biológicos que se estudian para detectar anomalías en sistemas biológicos), iPSCs (células madre pluripotentes inducidas, es decir, células que han sido reconvertidas a célula madre), o modelos con microorganismos o animales (Rodenburg, 2018).

La existencia de diferentes formas de realizar los estudios funcionales causa una gran heterogeneidad en la información generada por estos estudios. Esto dificulta la evaluación de la información proporcionada por estos estudios durante el proceso de interpretación, dada la diferencia en características como el tipo de variación genética, el contexto del experimento, el tipo de experimento y varios aspectos de su realización. Dada la complejidad que supone la variedad de factores, se dedicará la sección 5.1 para detallar estas características y la problemática asociada.

Al trabajar en un contexto de medicina de precisión, el conocimiento extraído de los estudios funcionales se destina al ámbito clínico. Sin embargo, hay incontables enfermedades para las que se pueden realizar estudios funcionales, por lo que la cantidad de artículos que pueden existir, cada uno con sus características individuales, es inmensa. En consecuencia, para este Trabajo de Fin de Grado se ha realizado una selección de enfermedades para acotar el rango de estudios a analizar, como describiremos en la siguiente sección.

1.3. Cardiopatías Familiares

Las guías ACMG-AMP se centran en enfermedades mendelianas, según lo comentado en la sección 1.2. Las enfermedades mendelianas comprenden un conjunto de patologías muy amplio. Tomando como referencia la información registrada en OMIM² (*Online Mendelian Inheritance in Man*[®], Johns Hopkins University), un conocido compendio público de genes humanos y fenotipos³, a día 29 de mayo 2023 se han identificado 7375 enfermedades mendelianas asociadas a una base molecular conocida.

Abarcar el estudio de un número tan elevado de enfermedades en este trabajo es inviable. Por este motivo, este Trabajo de Fin de Grado se focaliza en el estudio de las cardiopatías familiares que comprenden patologías hereditarias del corazón y vasos arteriales. Estas condiciones se caracterizan por estar asociadas a la muerte súbita, además de presentar una complejidad de efectos y síntomas que dificulta un diagnóstico preciso. Existen varios tipos de cardiopatías familiares, como miocardiopatías que afectan al músculo cardíaco, canalopatías que alteran las propiedades eléctricas

² URL de referencia en OMIM: <https://www.omim.org/statistics/geneMap>

³ Fenotipo: característica de un organismo, que es anormal a nivel clínico, y que está asociada a una determinada enfermedad (Scheuermann et al., 2009).

del corazón, aortopatías, y enfermedades sistémicas y enfermedades raras con afectación cardíaca (Sociedad Española de Cardiología, n.d.).

Se ha seleccionado este grupo de enfermedades debido a que son objeto de estudio en los proyectos CardioVal (AP INBIO, 2021)⁴ y OGMIOS (PROS UPV, 2021)⁵, los cuales se están llevando a cabo en el grupo de investigación en el que se está desarrollando este Trabajo de Fin de Grado. De esta manera, el listado completo de las patologías estudiadas es el siguiente: Miocardiopatía Hipertrófica (HCM), Miocardiopatía Dilatada (DCM), Miocardiopatía Restrictiva (RCM), Miocardiopatía No Compactada, Miocardiopatía Arritmogénica (ARVC), Síndrome de QT Largo (LQTS), Síndrome de QT Corto (SQTS), Síndrome de Brugada (BrS), Taquicardia Ventricular Polimórfica Catecolaminérgica (CPVT), Síndrome de Marfan (MS) y Síndrome de Loeys-Dietz (LDS).

1.4. Estructura del trabajo

En esta sección se sintetiza la estructura de la presente memoria de Trabajo de Fin de Grado, con una breve descripción de los nueve capítulos:

- **Capítulo 1. Introducción.** Presentación del marco contextual del Trabajo de Fin de Grado y la motivación que lo impulsa.
- **Capítulo 2. Objetivos.** Definición del objetivo principal y subobjetivos del Trabajo de Fin de Grado, junto a las preguntas de investigación asociadas.
- **Capítulo 3. Estado del Arte.** Revisión de herramientas existentes para la interpretación de estudios funcionales sobre variaciones genéticas y posibles mejoras.
- **Capítulo 4. Metodología.** Descripción de la metodología *Design Science* en la que se ha basado el desarrollo de este Trabajo de Fin de Grado.
- **Capítulo 5. Investigación del problema.** Primera etapa del *Design Science* dedicada a la caracterización y problemática de los estudios funcionales, seguido de la propuesta de una solución mediante modelado conceptual.
- **Capítulo 6. Diseño y desarrollo de la fuente de información.** Segunda etapa del *Design Science* que abarca el diseño de la estructura de la fuente de información, la selección de fuentes, la extracción de la información y su introducción en la fuente de información.
- **Capítulo 7. Resultados.** Se ha analizado la fuente de información resultante y se ha llevado a cabo un caso de uso con datos de pacientes reales.
- **Capítulo 8. Conclusiones y trabajos futuros.** Presentación de las conclusiones del Trabajo de Fin de Grado acorde a los objetivos y preguntas de investigación planteados en el Capítulo 2 y las potenciales investigaciones futuras en base a este Trabajo de Fin de Grado.
- **Capítulo 9. Bibliografía.** Listado de referencias y bibliografía del Trabajo de Fin de Grado.

⁴ Proyecto del Grupo PROS en colaboración con IIS La Fe de Valencia enfocado al diseño y desarrollo de un prototipo basado en Inteligencia Artificial Explicable para la gestión de información genética relacionada con el riesgo de sufrir muerte súbita de origen cardíaco.

⁵ Proyecto del Grupo PROS en colaboración con IIS La Fe de Valencia, INCLIVA, ISABIAL y BIONOS para desarrollar un sistema inteligente de apoyo a la toma de decisiones clínicas en medicina de precisión.

2. OBJETIVOS

El objetivo de este Trabajo de Fin de Grado es el diseño y desarrollo de una fuente de información de estudios funcionales sobre variaciones genómicas asociadas a cardiopatías familiares. Este objetivo principal se ha dividido en tres subobjetivos más específicos, cada uno de ellos guiado por una serie de preguntas de investigación (PI):

Objetivo 1. Investigación del problema

PI1- ¿Qué es un estudio funcional?

PI2 - ¿Cuáles son las problemáticas asociadas al uso de los estudios funcionales?

PI3- ¿Existe algún modelo que nos permita representar la información asociada a estudios funcionales?

PI4- ¿Es necesario introducir mejoras a este modelo conceptual?

Objetivo 2. Diseño y desarrollo de la fuente de información

PI1- ¿Cómo se estructurará la fuente de información a partir de la información relevante de los estudios funcionales?

PI2- ¿Cuáles son los estudios funcionales de interés para las cardiopatías familiares?

PI3- ¿Cómo se va a obtener la información de los estudios funcionales?

PI4- ¿Se puede estandarizar la información extraída?

Objetivo 3. Caso de uso

PI1- ¿Qué resultados se han obtenido en el diseño y desarrollo de la fuente de información?

PI2- ¿Cómo se puede comprobar la utilidad de esta fuente de información?

PI3- ¿La fuente de información cumple los objetivos esperados?

PI4- ¿Cuáles son los fallos, limitaciones, y mejoras identificadas?

3. ESTADO DEL ARTE

Antes de abordar los tres objetivos detallados en el capítulo previo, se ha realizado una revisión de los proyectos existentes con relación a los estudios funcionales y las Guías ACMG/AMP. Se han revisado el tipo de herramientas y recursos disponibles que incluyen datos funcionales para determinar si hay proyectos similares a este Trabajo de Fin de Grado.

3.1. Bases de datos existentes

Solo se han descubierto dos herramientas que recopilen información funcional: el Repositorio de Datos Funcionales (FDR, en adelante), y un conjunto de artículos de oncología llamado *Hi Set*.

Una de estas es el Repositorio de Datos Funcionales en la que está trabajando actualmente el grupo *Clinical Genome Resource*, conocido como ClinGen y dedicado a ser un recurso que defina la relevancia clínica de genes y variaciones para la investigación y medicina de precisión⁶. En concreto, el FDR se ha creado para integrar la significancia clínica (clasificación descrita en la sección 1.2), datos estadísticos y el impacto funcional de variaciones genéticas. Para los datos estructurales (p. ej. tipo de variación, gen al que pertenece o localización) y los datos estadísticos de variaciones, emplea la información disponible en bases de datos públicas, referenciando dichas plataformas. Para la significancia clínica y el impacto funcional, la información se incorpora a medida que se extrae manualmente por usuarios externos o por los Paneles de Expertos en Curación de Variaciones (VCEP) del propio grupo a partir de estudios funcionales. Una vez introducida la información sobre una variación, los usuarios pueden consultarla a través de una interfaz web que está actualmente en desarrollo. El repositorio es parte de un proyecto del grupo para crear una Interfaz de Curación de Variaciones (VCI) que permita aplicar los criterios de las Guías ACMG/AMP y clasificar las variaciones de forma acorde (Preston et al., 2022).

Cabe destacar, que el FDR está en desarrollo, por lo que no constituye un recurso totalmente estable. No existe ninguna publicación sobre el repositorio, por lo que se cuenta con información limitada sobre el mismo y las conclusiones expuestas se han definido a partir de la información disponible sobre su estructuración y contenido en otras publicaciones (Preston et al., 2022) y su primera versión web de interfaz de usuario⁷.

La otra herramienta que se ha encontrado es una recopilación de estudios funcionales denominada *Hi Set*, un conjunto de 22 artículos que se utilizó en un estudio para la reclasificación de variaciones de significado incierto del gen BRCA1, el cual está asociado a cáncer de mama y cáncer de ovario. Los 22 artículos fueron considerados “bien establecidos” por los investigadores, y fueron seleccionados por dos factores: el estudio de más de 10 variaciones, y una sensibilidad y especificidad mayor del 80%, calculada mediante la comparación con un panel de variaciones de referencia proveniente de las bases de datos ENIGMA y ClinVar (Lyra et al., 2021).

⁶ URL del grupo ClinGen: <https://www.clinicalgenome.org/>

⁷ Functional Data Repository UI: <https://ldh.genome.network/fdr/ui/>

3.2. Mejoras respecto al estado del arte

Las dos herramientas descritas en la sección previa muestran diferencias marcadas con la idea de este Trabajo de Fin de Grado. Por ello, se hablará de estas diferencias y se describirá la novedad que supone la propuesta original.

En primer lugar, ambas herramientas presentan una diferencia en cuanto al enfoque clínico. El FDR pretende abarcar cualquier enfermedad que tenga un panel de expertos asignado en ClinGen. La cantidad de variaciones que ClinGen contiene con información funcional asociada es reducida, al igual que la cantidad de enfermedades asociadas. En cuanto al *Hi Set*, este conjunto se centra en oncología, concretamente en cáncer de mama y cáncer de ovario. Ante la falta de información funcional asociada a cardiopatías familiares, este trabajo aportará nuevos datos curados para varias patologías y genes.

En segundo lugar, la accesibilidad y presentación de la información es muy diferente en ambos. En el caso del FDR son dos aspectos incompletos. Por ejemplo, para buscar por variación hay que emplear un identificador creado por el grupo ClinGen, no está disponible la búsqueda por genes y los resultados solo están disponibles en formato **.json*, por lo que no pueden visualizarse los datos directamente, sino que dependerá del usuario y el visualizador que este tenga disponible. En cuanto al *Hi Set*, la información sobre los artículos está proporcionada en un documento **.xlsx*, donde la información está partida en varias hojas, con los datos referentes al contenido de los artículos separados de la variación a la que están asociados. Dada la complejidad que ha supuesto poder encontrar la información en ambos recursos, se pretende que la fuente de información resultante de este trabajo sea fácil de acceder y consultar en caso necesario.

En tercer lugar, se han comparado los procesos de selección y validación de los artículos. Del FDR no hay información suficiente para determinar su metodología. Para el *Hi Set* solo se consideran estudios que analizan más de 10 variaciones, por lo que estudios de un número menor de variaciones que sean válidos y estén bien establecidos quedan excluidos del conjunto. Para el desarrollo de este trabajo, se pretende incluir cualquier artículo que describa un estudio funcional, independientemente de la cantidad de variaciones que estudie. De esta manera, al evitar este tipo de limitación, la información resultante de este trabajo se puede considerar completa y relevante.

Finalmente, ninguna de las herramientas cuenta con una representación conceptual del dominio en la que basar la estructuración de su información. En contraste, este Trabajo de Fin de Grado pretende abordar la asociación de la información que se extraiga de artículos empleando un modelo conceptual. Este tipo de modelo proporcionan una perspectiva comprensiva y fácilmente expandible del dominio, puesto que se basa en definir los conceptos clave de un campo y representar las relaciones entre ellos. De esta manera, la representación de la información se ajusta mejor a la realidad del dominio, asegurando su precisión y utilidad.

4. METODOLOGÍA

El presente Trabajo de Fin de Grado se ha realizado siguiendo la metodología de investigación *Design Science* tal y como fue planteada por Roel Wieringa para los sistemas de información (Wieringa, 2014). Así, el procedimiento general de la metodología consiste en el diseño e investigación de artefactos de un contexto determinado en respuesta a su problemática asociada.

En este caso, se puede definir el artefacto como *el diseño y desarrollo de una fuente de información de estudios funcionales* y el contexto correspondiente es la *interpretación de variaciones genómicas*, en este caso ciñéndonos a las recomendaciones de las guías ACMG-AMP y al dominio de las cardiopatías familiares.

La propuesta de Wieringa distingue dos tipos de procedimientos a seguir según el tipo de problema. Por un lado, para los problemas prácticos plantea un ciclo de diseño, o ciclo de ingeniería, que considera el impacto en el mundo real tomando en cuenta las necesidades de los usuarios. Por otro lado, para problemas de conocimiento, define un ciclo empírico, o ciclo experimental, que busca obtener un mayor entendimiento del mundo real sin necesariamente crear una nueva solución. En base a los objetivos definidos previamente de este Trabajo de Fin de Grado, se reconoce que es un problema práctico. Como tal, se siguió el ciclo de diseño, el cual consta de cinco etapas:

1. **Investigación del problema:** caracterización del problema mediante el planteamiento y respuesta de preguntas en relación con el marco conceptual del problema, las necesidades de los usuarios, las mejoras que se pueden realizar y cómo realizarlas.
2. **Diseño de la solución:** especificación de requisitos y elementos de la solución propuesta.
3. **Validación de la solución:** evaluación de la solución diseñada para comprobar si abarca correctamente la problemática y cumple lo establecido en la etapa de investigación.
4. **Implementación de la solución:** aplicación de la solución a la problemática real.
5. **Evaluación de la implementación:** análisis de la implementación para comprobar si se ha resuelto el problema de manera satisfactoria.

Con este Trabajo de Fin de Grado se han abordado las primeras dos etapas del ciclo, puesto que son las etapas realmente dedicadas al diseño de la solución. Para la validación necesitaríamos contar con la revisión y aprobación por parte de un experto externo, por lo que, como alternativa, se ha desarrollado un caso de uso para comprobar la utilidad de la solución diseñada con datos reales. Las últimas dos etapas, en cambio, serían más apropiadas para entornos industriales donde se desarrolla un producto a gran escala que debe ser evaluado para asegurar un correcto funcionamiento. Teniendo esto en cuenta, la aplicación del procedimiento queda reflejado en los subobjetivos del capítulo 2 de esta memoria y en la Figura 1.



Figura 1. Representación del Ciclo de Diseño, adaptado a partir de la metodología del Design Science, que se ha seguido en este Trabajo de Fin de Grado

En adelante, se reflejarán las tres etapas en el desarrollo del trabajo mediante capítulos dedicados a cada una. La primera parte, la investigación, cubrirá el contexto de los estudios funcionales, definiendo la problemática de heterogeneidad de información y cómo interpretarla y sintetizarla de forma precisa. Seguidamente, para la segunda etapa se definirá la estructura de la fuente de información, y se realizará la búsqueda, extracción y presentación de dicha información. Finalmente, se hará uso de la fuente de información resultante en el caso de uso, empleando datos VCF anotados de pacientes reales, y se revisará el cumplimiento de los objetivos y las limitaciones encontradas.

5. INVESTIGACIÓN DEL PROBLEMA

Acorde a la primera etapa de la metodología del *Design Science*, este capítulo describirá el problema práctico que se va a tratar en este Trabajo de Fin de Grado. Para ello, se proporcionará una descripción general del concepto de estudio funcional para determinar, a continuación, la problemática asociada a estos estudios y, por otra parte, cuál es la información relevante de los mismos. Una vez caracterizados los estudios funcionales, se describirá la solución de modelado conceptual propuesta que conformará la base para la siguiente etapa del ciclo de diseño.

5.1. Estudios funcionales: Caracterización y problemática

En este Trabajo de Fin de Grado, se ha definido un estudio funcional como un experimento realizado para determinar cómo los genes contribuyen a procesos biológicos, buscando definir el mecanismo molecular por el que un cambio en la secuencia del ADN provoca la presencia o predisposición a una enfermedad.

Tal y como se describió en la introducción a los estudios funcionales, existe una amplia variedad de estudios funcionales, cada uno con sus propias particularidades y aspectos a evaluar para determinar su precisión y relevancia. Por ello, el criterio PS3/BS3 de las guías ACMG/AMP busca guiar este análisis de relevancia de los estudios, definiendo una serie de criterios para determinar su fiabilidad; sin embargo, se proporciona poco detalle en las guías sobre cómo aplicar estos criterios de forma precisa.

Lo que proponen las guías ACMG/AMP es la evaluación de los estudios funcionales en base al procedimiento experimental, considerando tres características principales que describen brevemente como se expone a continuación: la **validez** del estudio según lo bien que refleje el ambiente biológico, la **reproducibilidad** del experimento realizado, que incluye la información que permita la reproducción del experimento a compañeros de la comunidad científica, y la **robustez** del estudio, que se asocia a la integridad de la muestra e incluye datos como el método y momento de adquisición, el almacenaje y transporte (Richards et al., 2015).

Ante la falta de detalles, surge un problema de variabilidad, puesto que cada experto que evalúa un estudio lo hace de forma distinta. Por tanto, para interpretar la información de los estudios de forma precisa, es necesario un análisis más profundo. A raíz de esto, en 2019, el grupo ClinGen realizó un estudio comparativo para revisar el proceso de evaluación de los estudios, para comprobar las diferencias de interpretación entre expertos (Kanavy et al., 2019). Se estudió la evaluación realizada por seis Paneles de Expertos (VCEP), cada uno asignado a una enfermedad distinta. Los resultados reafirmaron la necesidad de una estandarización de la evaluación. En base a esto, el subgrupo de Interpretación de Variaciones Secuenciales de ClinGen desarrolló en 2020 una serie de recomendaciones y un modelo provisional (Figura 2) para la aplicación del criterio PS3/BS3 (Brnich et al., 2020). Se revisaron los puntos relevantes en la interpretación como las consecuencias moleculares, la terminología y el uso de controles. A partir de estos se plantearon las cinco recomendaciones finales descritas a continuación:

- i) Las muestras no artificiales provenientes de un paciente deben ser evaluadas con mayor rigor dada la variabilidad de la genética individual.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

- ii) El peso de los resultados obtenidos de muestras generadas en organismos modelo estará ajustado en función del rigor y reproducibilidad de los métodos empleados.
- iii) Aunque es preferible la mínima edición o manipulación que pueda causar situaciones no naturales, como la sobreexpresión de proteínas, no es excluyente de ser un estudio bien establecido. Se deberá considerar si la situación está justificada y controlada correctamente en el estudio.
- iv) Para unificar la terminología de interpretación, se deberían emplear los términos “funcionalmente normal” y “funcionalmente anormal” para describir los resultados.
- v) La entidad que ha realizado el estudio no debe influir en el cumplimiento del criterio. La validación se hará puramente según los parámetros que se consideren.

Adicionalmente, el grupo planteó un proceso, mostrado en el árbol de decisión de la Figura 2, para comprobar si el criterio es aplicable, y en caso afirmativo, qué peso deberá tener la evidencia a la hora de interpretar la información. Así, este proceso consta de cuatro pasos siendo 1) definir el mecanismo biológico de la enfermedad, 2) comprobar que el experimento se considera un estudio funcional y modela el mecanismo de enfermedad, 3) validar el experimento realizado en base a las métricas de validación, 4) utilizar los resultados para la interpretación de variaciones.

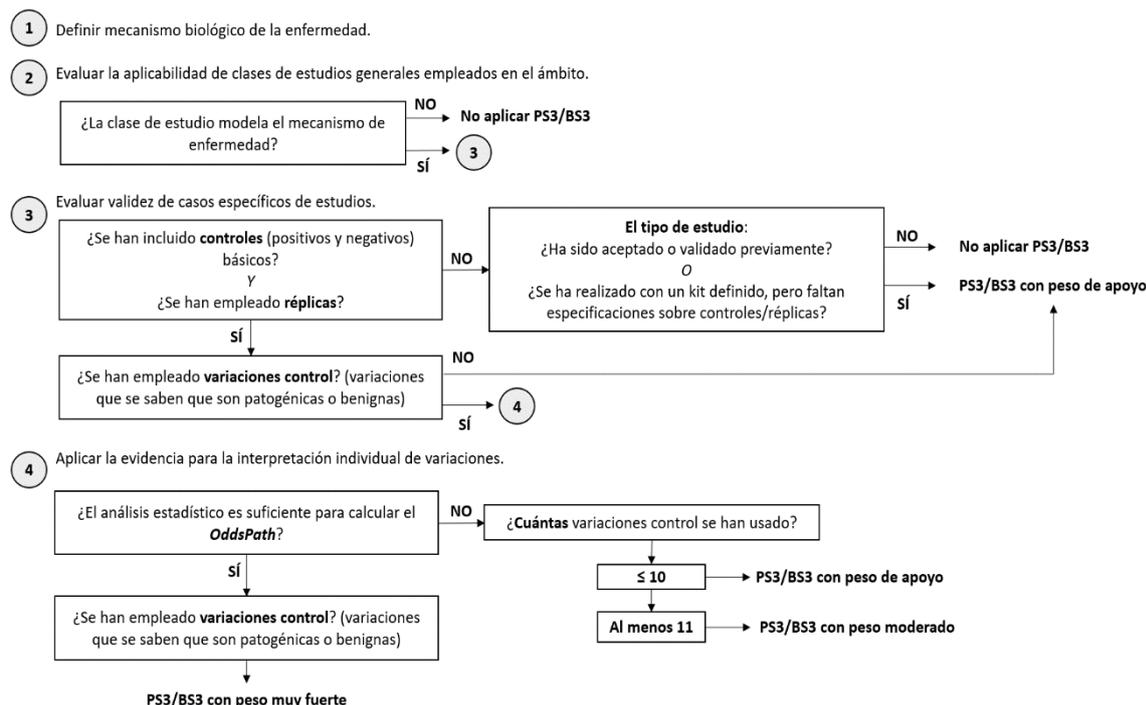


Figura 2. Árbol de decisión para la evaluación de información funcional según las métricas de validación definidas por el grupo ClinGen (Brnich et al., 2020).

Con estos estudios, se demuestra la complejidad de la interpretación de variaciones. Existe una variabilidad de numerosos factores, lo cual influye en el enfoque de las investigaciones y en la interpretación de resultados. Más gravemente, la problemática principal es que la información funcional se encuentra dispersa en distintas fuentes y bases de datos de forma heterogénea, como consecuencia de la variabilidad mencionada y debido a que los esfuerzos de estandarización son un

suceso reciente. Es por esto por lo que el objetivo de este Trabajo de Fin de Grado es realizar una recopilación de la información funcional disponible sobre variaciones, de forma que se genere un recurso que sirva de referencia y permita una clasificación de variaciones eficiente.

Empleando las tentativas del grupo ClinGen como referencia, se pueden identificar los campos de información relevantes que se necesitan para evaluar un estudio funcional. La siguiente cuestión para desarrollar una fuente de información es poder estructurar la caracterización de los estudios funcionales de forma adecuada. Se ha considerado que la herramienta más indicada es el modelado conceptual, por lo que, en las siguientes secciones, se va a definir este sistema, justificando su uso en este Trabajo de Fin de Grado, y se describirá su aplicación para el diseño y desarrollo de la fuente de información.

5.2. Modelado conceptual

El dominio genómico en el cual se enfoca este Trabajo de Fin de Grado es un perfecto ejemplo de un dominio de gran complejidad, por lo que para representarlo de forma precisa es importante utilizar un sistema apropiado. A la hora de trabajar en un problema complejo, el método más efectivo es descomponerlo en conceptos simples y comprender cómo estos están conectados. Es justamente esto lo que permite el modelado conceptual, por lo que supone la herramienta perfecta para caracterizar el campo genómico.

Por definición, un modelo conceptual es una representación de un sistema mediante el cual se definen la estructuración y relaciones de los datos de un dominio determinado (Bornberg-Bauer & Paton, 2002). Esto permite crear una visión holística del campo donde los conceptos fundamentales y sus relaciones quedan claramente reflejados.

Ha habido varias aproximaciones en la comunidad científica de desarrollar modelos conceptuales que describan diferentes aspectos del genoma humano, como la relación genoma-fenotipo a nivel genético, proteico y metabólico (Paton et al., 2000), las estructuras e interacciones de las proteínas (Ram & Wei, 2004) y metadatos de datos genómicos (Bernasconi et al., 2017). Sin embargo, todos estos proyectos están acotados en áreas específicas de la genómica, por lo que no se representa el ámbito genómico en toda su complejidad. Con el propósito de enfrentar esta problemática, el grupo PROS, perteneciente al Instituto VRAIN de la UPV, ha desarrollado un Modelo Conceptual del Genoma Humano (CSHG). Iniciado en 2016 (Reyes Román et al., 2016), abarca los conceptos claves en relación con la medicina de precisión. Al igual que el conocimiento genómico, está en constante evolución (García S. et al., 2020; Reyes Román, 2018), adaptando las secciones existentes e integrando nuevas secciones según los avances y descubrimientos científicos.

5.2.1. CSHG

El objetivo de este modelo conceptual es proporcionar una perspectiva holística del genoma humano para facilitar la comprensión del genoma humano (García S. et al., 2020). En este Trabajo de Fin de Grado, será la base donde se definirán inicialmente los parámetros a evaluar en los estudios funcionales y servirá de referencia para el diseño de la fuente de información.

Actualmente, el CSHG abarca varias dimensiones del contexto genómico mediante cinco secciones o vistas principales:

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

1. Vista Estructural, donde se representa los componentes fundamentales del ADN, mostrando su organización y estructura.
2. Vista de Transcripción, en la que se modela el proceso y elementos de transcripción del ADN a proteínas.
3. Vista de Variación, donde se detallan las posibles variaciones que puedan darse en la secuencia de ADN.
4. Vista de Bibliografía, que proporciona información sobre las fuentes y referencias bibliográficas asociadas a las entidades incluidas en el CSHG.
5. Vista de Rutas metabólicas, en la que se representan los elementos de los procesos biológicos celulares.

Estas vistas son el esqueleto del CSHG, y cuentan con numerosas conexiones entre ellas que permiten la descripción completa de datos del genoma humano. Tomando esta base, en este Trabajo de Fin de Grado se ha ampliado el modelo conceptual existente para incluir e integrar toda la información relativa a los estudios funcionales y conocer el impacto funcional de una variación genética. Los estudios funcionales supondrán una nueva entidad que irá embebida en esas conexiones, sobre todo entre las vistas de variación y bibliografía. De esta manera, en el siguiente capítulo, relativa a la etapa de diseño y desarrollo de la fuente de información, se describirá la adición de la nueva sección, cuya estructura será la inspiración para el diseño de la fuente de información de estudios funcionales.

6. DISEÑO Y DESARROLLO DE LA FUENTE DE INFORMACIÓN

Una vez realizada la investigación del problema, se ha avanzado a la segunda etapa del *Design Science*. Este capítulo abordará el diseño de la fuente de información, explicando la ampliación del CSHG para estructurar la información relevante, y cómo se ha adaptado dicha estructura para la fuente de información. Se continuará con el desarrollo de la fuente de información, detallando la búsqueda de los artículos, los criterios de selección y la extracción de información de estos. Por último, se describirá la revisión y estandarización de la información extraída.

6.1. Diseño de la fuente de información

A lo largo de la presente memoria, se ha estudiado la complejidad que presenta interpretar variaciones a partir de estudios funcionales. La problemática principal descrita – la falta de un recurso de estudios funcionales – se va a abordar en este Trabajo de Fin de Grado generando una fuente de información funcional para variaciones asociadas a cardiopatías familiares. De esta manera, la evaluación de los criterios PS3/BS3, que requieren esta información funcional de las variaciones, se podrá realizar de forma eficiente.

El primer paso para diseñar la fuente de información ha sido identificar la información relevante de los estudios, lo cual se ha realizado a partir de la caracterización realizada previamente, tomando como base las métricas y parámetros destacados en las publicaciones del grupo ClinGen (Brnich et al., 2020; Kanavy et al., 2019). A continuación, se ha estructurado la información relevante mediante el modelado conceptual, realizando una ampliación del CSHG para incluir esta nueva entidad. En las secciones siguientes se explicará esta ampliación y, seguidamente, se describirá cómo se adaptará esa estructura para construir la fuente de información.

6.1.1. Ampliación del CSHG

Para determinar la información relevante, se han considerado dos referencias. Por una parte, se han analizado los ejemplos de interpretación proporcionados en el estudio comparativo de ClinGen (Kanavy et al., 2019), realizados por sus Paneles de Expertos en Curación de Variaciones (VCEP). La mayor influencia ha sido la labor realizada por el panel asignado a evaluar estudios sobre miocardiopatía hereditaria asociada al gen MYH7, dado que esta enfermedad se incluye en las estudiadas en este trabajo. Por otra parte, se han incluido las métricas del proceso de validación propuesto por Brnich et al., 2020, (Figura 2), las cuales fueron seleccionadas por ClinGen porque permiten valorar las diferencias entre un mismo experimento realizado en distintos laboratorios, permitiendo evaluar la validez de los estudios. Se ha prestado especial atención a los parámetros del tercer paso – evaluación de validez de estudios – puesto que el cuarto paso está enfocado a aplicar la interpretación, tarea que queda en mano de los expertos de interpretación de variaciones.

Una vez definidos los parámetros necesarios, se tiene la información necesaria para evaluar la información funcional de los estudios sobre las variaciones investigadas. Esta descomposición de los estudios funcionales en elementos fundamentales se ha usado para representar estos estudios con modelado conceptual. Puesto que no existía ninguna entidad correspondiente a los estudios funcionales, se ha realizado una ampliación del CSHG para introducir el nuevo concepto tal y como se representa en la Figura 3. De esta manera, la información funcional se ha integrado en la descripción del dominio genómico del CSHG, ampliando su visión holística.

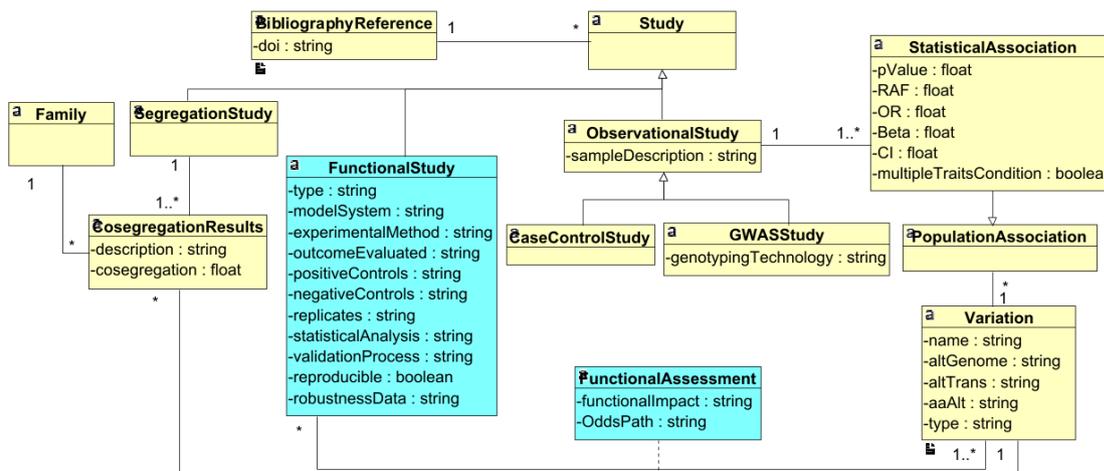


Figura 3. Diagrama de clases de la ampliación correspondiente a los estudios funcionales (marcados en azul) y sus conexiones directas al resto del CSHG.

La adición se ha llevado a cabo en la vista de Bibliografía, donde encontramos el concepto “estudio” (*Study*). Este concepto es una entidad general al que se introducen varias clases de estudios. Con esta estructura, cada clase puede tener su entidad y descripción propia, pero todos los estudios, independientemente del tipo, se identifican y referencian de forma homogénea.

La ampliación realizada ha consistido en la creación de una nueva clase de estudio. Para ello, se ha creado una entidad llamada *FunctionalStudy*, que contiene varios campos, correspondientes a la información relevante de los estudios funcionales. Como se puede ver en la Figura 3, la entidad de un estudio funcional (*FunctionalStudy*) está relacionada directamente con la entidad variación (*Variation*), de manera que cada estudio puede describir el impacto de una o varias variaciones. Esta descripción la representamos con el campo de *FunctionalAssessment*, como resultado de la asociación de ambas entidades mencionadas, y mostrará el impacto funcional de la variación según la información del estudio correspondiente.

Para definir la estructura interna de los estudios funcionales, la información relevante se ha organizado en los once conceptos definidos a continuación:

- 1- **type**: Tipo de estudio realizado.
- 2- **modelSystem**: Tipo de sistema empleado como modelo, pudiendo ser in vitro, in vivo o in silico.
- 3- **experimentalMethod**: Método experimental empleado para el estudio.
- 4- **outcomeEvaluated**: Aspecto o parámetro que se ha evaluado a través del experimento.
- 5- **positiveControls**: Cantidad de controles positivos empleados.
- 6- **negativeControls**: Cantidad de controles negativos empleados.
- 7- **replicates**: Cantidad de réplicas o repeticiones realizadas en el experimento.
- 8- **statisticalAnalysis**: Datos del análisis estadístico descrito en el estudio.
- 9- **validationProcess**: Proceso de validación del estudio.
- 10- **reproducible**: Posibilidad de reproducción de los experimentos.

11- **robustnessData**: Datos que garantizan la robustez del estudio, o la integridad de la muestra, como definida en la sección 5.1.

En cuanto al campo intermedio *FunctionalAssessment*, se han incluido dos conceptos en su estructura: el impacto funcional (es decir, si el cambio ha afectado a la función normal del gen) y el *OddsPath*, un parámetro probabilístico creado por el grupo ClinGen. Dado que este dato solo se aplica en el cuarto paso, el cual deben llevarlo a cabo los expertos en interpretación de variaciones, no se tendrá en cuenta para la fuente de información.

6.1.2. Estructura de la fuente de información

Gracias a la ampliación del CSHG, contamos con la representación holística y completa que buscábamos obtener. Puesto que la información está estructurada en muchos campos, se ha decidido que el formato de la fuente de información sea una hoja de cálculo de Microsoft Excel (*.xlsx), lo cual garantiza una amplia capacidad y facilidad a la hora de consultarla y modificarla.

La información provendrá de los artículos científicos sobre estudios funcionales de variaciones investigadas con relación a cardiopatías familiares. Como un estudio funcional puede tratar más de una variación, asignaremos una variación por fila, y los diferentes campos de información se corresponderán a las columnas. Dado que el núcleo del trabajo son los estudios funcionales, las primeras columnas consistirán en los datos necesarios para identificar el artículo. Seguidamente, irán la enfermedad a la que se ha asociado la variación en el artículo, y los datos de la variación. Así, los títulos y contenido de estas primeras columnas serán los siguientes:

- **Artículo**: título original del artículo.
- **DOI**: enlace DOI correspondiente.
- **Disease**: cardiopatía familiar a la que se ha asociado la variación.
- **Gene**: nombre del gen donde ocurre la variación.
- **Variant**: nombre de la variación.

Tras estos datos, las siguientes once columnas serán los conceptos de definición del *FunctionalStudy* en el modelo conceptual, con el mismo nombre y en el mismo orden que se han listado previamente en la sección 6.1.1. Finalmente, la última columna será el impacto funcional (*functionalImpact* en la entidad *FunctionalAssessment* de la Figura 3), y prescindiremos del parámetro *OddsPath* como hemos explicado en la sección previa.

Contamos, en total, con 17 campos de información, y cada uno constituye una columna individual en la fuente de información. El aspecto completo de la fuente de información puede verse en la Figura 4, con ejemplos de la información respectiva. El contenido de las columnas se explicará con mayor detalle en la sección 6.2, correspondiente al desarrollo de la fuente de información.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

ARTÍCULO	DOI	Disease	Gene	Variant	type	model System	experimental Method	outcome Evaluated
FLNC Gene Splice	10.1016	Dilated Cardi	FLNC	c.7251D1G>	OBI:00016	In vivo	Zebrafish embryos injec	Morphology/structu
FLNC Gene Splice	10.1016	Dilated Cardi	FLNC	c.5669-1del	OBI:00016	In vivo	Zebrafish embryos injec	Morphology/structu

positive Controls	negative Controls	replicates	statistical Analysis	validation Process	reproducible	robustness Data	functional Impact
2	1	4	In experiments det	Organism model, th	Yes	Male and female	Functionally abnormal. Tre
2	1	4	In experiments det	Organism model, th	Yes	Male and female	Functionally abnormal. Spl

Figura 4. Capturas de ejemplo de la estructura de la fuente de información

6.2. Desarrollo de la fuente de información

El desarrollo constituye la segunda parte de esta etapa del ciclo de diseño. En este Trabajo de Fin de Grado, aquí se abarca la búsqueda, selección y extracción de la información de estudios funcionales a partir de artículos disponibles en repositorios públicos de literatura científica. Esta sección describirá este proceso detallando las fuentes empleadas, la búsqueda y selección de artículos (representada en la Figura 5), cómo se ha enfocado la lectura para la extracción de la información, y, finalmente, la revisión y estandarización de la fuente de información una vez completada la lectura de artículos.

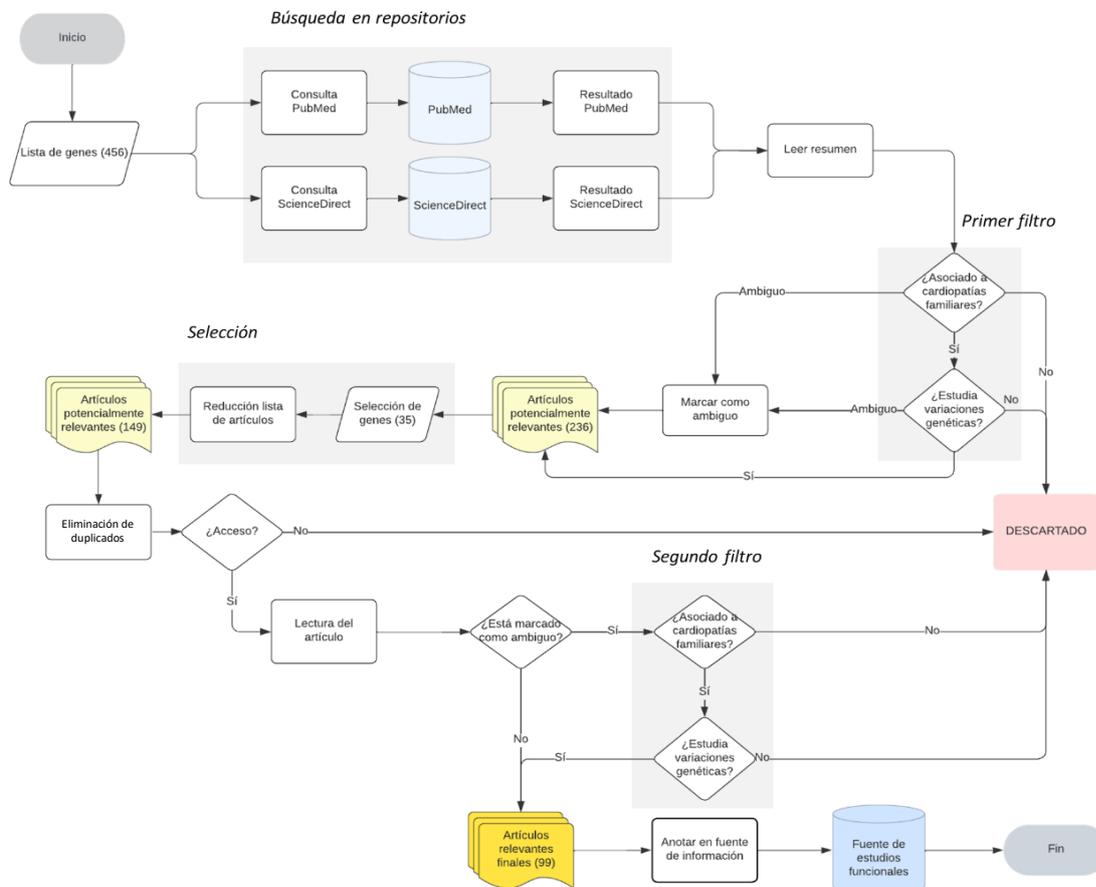


Figura 5. Flujo de trabajo completo del desarrollo de la fuente de información. Detalla la búsqueda en repositorios de artículos, los filtros y selección aplicados y la extracción de información relevante.

6.2.1. Primera búsqueda en repositorios

Los estudios funcionales se basan en distintos tipos de experimentos, como se ha expuesto en la sección 1.2.2., por lo que la información está presente en los artículos científicos que relatan la investigación realizada. Como se ha expuesto en el capítulo 3 sobre el estado del arte, no existe actualmente una fuente que integre datos de estudios funcionales de variaciones genéticas. Como consecuencia, el objetivo de este Trabajo es recopilar la información de estudios funcionales a partir de la lectura de artículos, concretamente aquellos sobre variaciones asociadas a cardiopatías familiares, y guardarla en una fuente de información. Ante la inmensa cantidad de artículos disponibles, se ha realizado una búsqueda y selección específica para poder determinar los artículos relevantes de entre todos los documentos disponibles.

Se ha comenzado por seleccionar los repositorios de artículos donde realizar la búsqueda. Para ello, se han realizado búsquedas de artículos sobre estudios funcionales en varios navegadores. De esta manera, se han revisado los repositorios o bases de datos existentes y se han valorado según la accesibilidad y la cantidad de publicaciones que recogen. Se han determinado dos repositorios candidatos, puesto que son los que más artículos científicos revisados recogen: *ScienceDirect*, cuyo alcance es muy amplio, dado que incluye todo tipo de documentos sobre cualquier sector científico, y *PubMed*, el cual se centra en publicaciones biomédicas. Independientemente de sus diferencias, ambas bases de datos tienen el reconocimiento de la comunidad científica internacional. Se ha observado que ambos contienen una gran cantidad de artículos comunes, pero también se han detectado varios que solo han aparecido en uno de los dos repositorios. Por tanto, se ha decidido utilizar ambos para asegurar la mayor cobertura de literatura posible.

El siguiente paso ha sido realizar la búsqueda en cada uno de los repositorios. Ambas bases de datos permiten una búsqueda avanzada mediante palabras claves, y comprenden el uso de operadores booleanos como AND (permite unir varios términos de búsqueda), OR (permite incluir términos alternativos). Adicionalmente, PubMed reconoce el uso del asterisco como delimitador. Es decir, actúa como un comodín que permite buscar palabras derivadas. Por ejemplo, la consulta "variaci*" retornará tanto los resultados con la palabra "variación" como con la palabra "variaciones".

Como palabras clave, se han empleado los términos en inglés, por lo que estudio funcional equivale a *functional assay* o *functional study*, y cardiopatías familiares equivale a *hereditary heart diseases*. Se han realizado diversas combinaciones de estas palabras, haciendo uso de los operadores, para valorar la cantidad de resultados que retorna cada combinación. Tras comparar los resultados, se han ejecutado finalmente las siguientes consultas para cada repositorio:

PubMed:

((*functional assay*[Title/Abstract]) OR *functional stud**[Title/Abstract]) AND (*cardio**) AND (**GEN**)

Cómo resultado, esta búsqueda retornará aquellos artículos de *PubMed* que contengan las palabras *functional assay* o *functional study/studies* en el título y resumen, que estén asociados o contengan cualquier palabra que empiece por *cardio* y que mencionen el gen correspondiente.

ScienceDirect:

((*functional study*) OR (*functional assay*)) AND (*cardiomyopathy* OR *heart*) AND **GEN**

Retornará aquellos artículos en *ScienceDirect* que contengan los términos *functional assay* o *functional study* junto con *cardiomyopathy* o *heart* en el título, resumen o palabras claves especificadas por el autor. En la consulta de *PubMed*, gracias al marcador *, se ha empleado solo el término “cardio*”, puesto que abarca una variedad de términos comunes en el ámbito de enfermedades de corazón (*cardiomyopathy, cardiovascular, cardiologist/s, cardiopathy...*). Sin embargo, dado que *ScienceDirect* no reconoce este marcado, para aumentar el alcance de la búsqueda se ha incluido la palabra *heart*, una palabra con altas probabilidades de ser mencionada cuando se trata cualquier cardiopatía.

Se puede observar que ambas consultas terminan con la palabra **GEN**. Este término será el nombre del gen para el que queremos encontrar estudios. Dada la naturaleza de las enfermedades mendelianas – causadas por una variación en un gen en concreto – esto permite realizar las búsquedas de forma estructurada y enfocada a las cardiopatías familiares que nos interesan. Para garantizar la mayor precisión posible, se ha empleado una lista, aportada por expertos clínicos en cardiología. Esta lista comprende 456 genes de interés que se han asociado a las cardiopatías familiares seleccionadas.

Para cada uno de estos genes, se han realizado dos consultas, una en cada repositorio, lo que supone un total de 912 consultas para recopilar los artículos de interés. Además, para cada una de esas consultas, no todos los resultados que se retornaban eran publicaciones que se ajustaran al perfil de estudio funcional. Por tanto, a medida que se han realizado las búsquedas, se ha aplicado un primer filtrado de resultados (Figura 6): para cada artículo retornado por las consultas, se ha leído el resumen y se ha valorado si la investigación descrita se ajusta a la definición de estudio funcional (si se estudia el efecto funcional de variaciones genéticas) y si estaba asociado a las cardiopatías familiares seleccionadas. Algunos casos presentaban cierta ambigüedad dado que mencionaban el estudio de variaciones sin especificar datos de la variación, o mencionaban una de las cardiopatías familiares como enfermedad potencialmente asociada sin detallar la evidencia tras estas suposiciones. Para poder descartar con mayor certeza, estos artículos se han marcado como ambiguos y se han incluido en la lista provisional de artículos relevantes para completar su lectura y juzgar definitivamente en base al contenido.

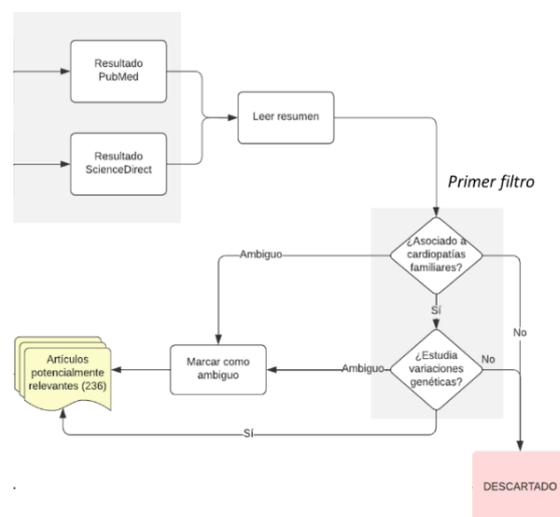


Figura 6. Ampliación del primer filtro, el cual se ha aplicado tanto en la primera como en la segunda búsqueda.

6.2.2. Segunda búsqueda y selección final

Una vez seleccionados los artículos, se han tomado dos artículos aleatorios de la selección para anotar la información según la estructura definida en la sección 6.1.1 y comprobar si hace falta modificar algún aspecto de la búsqueda. Al leer por completo uno de los primeros artículos, se ha visto que no detallaba el estudio funcional, sino que era una revisión de la investigación, por lo que se ha sustituido por el artículo del estudio original. Esta publicación sí contiene los datos del estudio funcional; sin embargo, no estaba presente en la recopilación realizado, a pesar de constar en PubMed, una de las bases de datos que se habían revisado. Comparando el contenido del artículo con la consulta realizada, se ha determinado que la ausencia se debe a que el artículo nuevo se describe como un análisis funcional (*functional analysis*), un término diferente que no se incluía en la consulta inicial dado que se había considerado un nombre demasiado ambiguo para la búsqueda.

Antes de continuar la lectura, se ha repetido la búsqueda para algunos de los genes más estudiados en cardiopatías familiares (p. ej. RYR2, SCN5A, MYH7) añadiendo el nuevo término. Viendo que en múltiples ocasiones las nuevas consultas han añadido publicaciones relevantes para los genes comprobados, se ha decidido realizar una segunda búsqueda completa para asegurar que no quedaran excluidos posibles artículos de interés. Por tanto, el nuevo formato de consulta queda como se muestra a continuación:

PubMed:
(("functional assay"[Title/Abstract]) OR "functional stud*" [Title/Abstract] OR "functional analysis"[Title/Abstract]) AND (cardio*) AND (GEN)

ScienceDirect:
((functional study) OR (functional assay) OR (functional analysis) AND (cardiomyopathy OR heart) AND GEN

Aplicando el mismo proceso de selección y descarte descritos para la primera búsqueda (Figura 6), se han ido añadiendo nuevos artículos a la lista de publicaciones para la lectura. Finalmente, al terminar las consultas, se ha realizado un filtrado de artículos duplicados. Es decir, artículos asociados a más de un gen de la lista y que, por tanto, constan más de una vez en la recopilación.

El número final de artículos seleccionados para la lectura ha sido 236 publicaciones, asociadas a 84 genes. Dada esta cantidad de artículos, el tiempo necesario para leer y apuntar manualmente la información relevante para todos los estudios es muy elevado. Para ajustar el trabajo al tiempo disponible, se ha acotado la selección. Para ello, los expertos clínicos han realizado una revisión de la lista y han indicado los genes más relevantes para los fenotipos de cardiopatías familiares que se quieren estudiar. Tras esta selección de genes, se ha reducido la lista a 35 genes y, en consecuencia, se ha reducido la cantidad de artículos asociados a 149 publicaciones (Selección, Figura 5). Una vez recopilados los artículos de interés, se puede avanzar a la extracción de los parámetros relevantes.

6.2.3. Extracción de información

De forma resumida, para obtener esta información de los artículos, el proceso seguido ha sido leer cada artículo, aplicar un segundo proceso de filtrado y, si sigue considerándose un artículo válido – que describe un estudio funcional – anotar las características del estudio según la estructura definida en la sección 6.1.2. Este proceso está representado en la Figura 7.

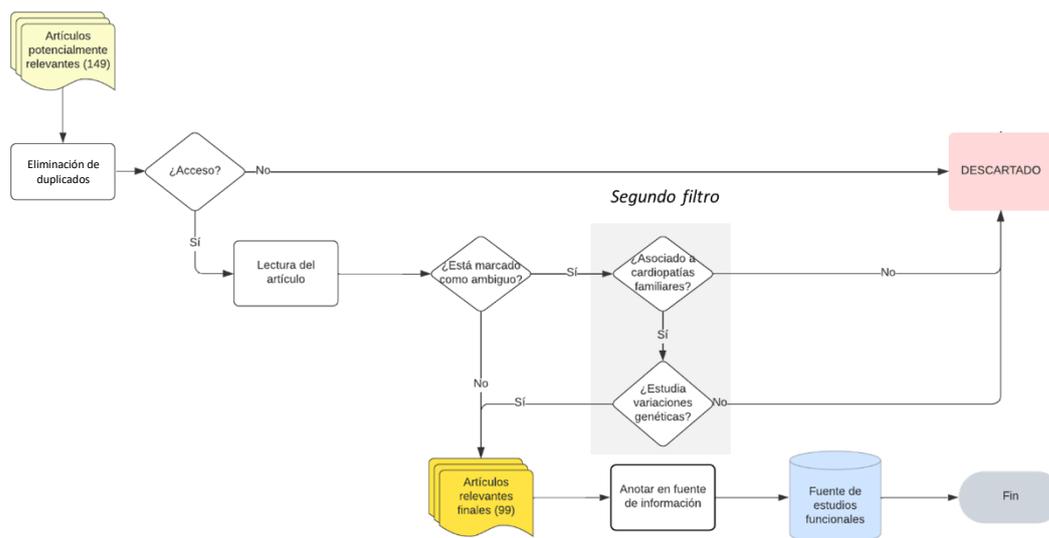


Figura 7. Ampliación del segundo filtro aplicado durante la lectura de los artículos.

El segundo filtrado se ha aplicado a lo largo del proceso de lectura. Por una parte, si no se podía acceder al texto completo de alguna publicación, se ha descartado el estudio. Luego, de los artículos que se habían marcado como ambiguos durante las consultas (descrito en la sección 6.2.1.), se ha leído el artículo completo. La mayoría de los artículos marcados como ambiguos han resultado ser una revisión de literatura y conocimiento, por lo que no detallaban estudios realizados, sino que resumían los datos disponibles sobre las variaciones. Por tanto, no se consideran un estudio funcional y han sido descartados.

En cuanto a los artículos aceptados, se ha procedido a la extracción de información. A la hora de incorporar los datos en la fuente de información, se han tomado varias medidas durante la lectura para cumplir esta visión. Para explicarlas, recorreremos los campos de información en orden, proporcionando ejemplos de cada uno. Cualquier dato que no estuviera disponible en un artículo se ha marcado como NA (*not available*).

Columna 1. Artículo. Se presenta el nombre del artículo tal y como se presenta en las publicaciones. Dado que dentro de un mismo artículo pueden estudiarse varias variaciones, se repetirá el título para cada variación.

Ejemplo: *Mutations in Filamin C causes a new form of familial hypertrophic cardiomyopathy*

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Columna 2. DOI. Contiene el DOI de referencia del artículo en formato de hipervínculo para proporcionar un acceso simple a la publicación.

Ejemplo: [10.1038/ncomms6326](https://doi.org/10.1038/ncomms6326)

Columna 3. Disease. Se indica el nombre completo de la enfermedad a la que está asociada la variación, sin siglas.

Ejemplo: *Hypertrophic Cardiomyopathy*

Columna 4. Gene. Se ha utilizado el nombre del gen que aparece en el artículo. Aunque hay casos en los que los genes se identifican por más de un nombre, al estar tratando con los genes más relevantes, la heterogeneidad de nombres es mucho menor, dado que son genes altamente conocidos y estudiados.

Ejemplo: *FLNC*

Columna 5. Variant name. El nombre de la variación se ha anotado tal y como aparece en el artículo.

Ejemplo: *p.V123A*

Columna 6. Type (tipo de experimento). Para asegurar una consistencia a lo largo del desarrollo de la fuente de información y facilitar su comprensión e interoperabilidad para cualquier uso, se han empleado términos ontológicos para definir el tipo de estudio, al igual que en los ejemplos de Kanavy et al., 2019. La ventaja de utilizar estos términos es que forman parte de un lenguaje estándar (ontología) donde se definen los conceptos de un ámbito y se les asigna un código identificativo único. Se ha empleado el buscador de ontologías OSL (<https://www.ebi.ac.uk/ols4>) del grupo EMBL-EBI (<https://www.ebi.ac.uk/>). Este buscador integra numerosas ontologías, por lo que se ha tratado de emplear, mayoritariamente, términos de tres de las más reconocidas: las ontologías ECO (evidencia y conclusiones, (Giglio et al., 2019)), BAO (estudios biológicos, (Abeyruwan et al., 2014)) y OBI (investigaciones biomédicas, (Brinkman et al., 2010)).

Ejemplo: *ECO:0000007. Immunofluorescence evidence, OBI:0000854. Western Blot assay, BAO:BAO_0000129. Immunoassay, BAO:BAO_0010252. Protein expression assay*

Columna 7. ModelSystem. Originalmente, las definiciones de los criterios PS3 y BS3 hablan de estudios *in vitro* y *ex vivo*, pero para este trabajo se han incluido los estudios *in silico* para poder valorar la información de este tipo de estudios.

Ejemplo: *In vitro*

Columna 8. ExperimentalMethod. Tras leer la sección de métodos, se ha realizado un resumen del procedimiento experimental descrito.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Ejemplo: *H9C2 rat cardiac myoblasts, neonatal rat cardiac myocytes and C2C12 cell myotubes nucleofected with the cDNA clones containing mutations or wild-type FLNC were studied by Western Blot, immunofluorescence and confocal microscopy.*

Columna 9. OutcomeEvaluated. En base a la descripción de los resultados, se determinó qué característica se ha evaluado para analizar la funcionalidad de la variación.

Ejemplo: *Formation of protein aggregates*

Columna 10. PositiveControls. Los controles positivos consisten en incluir en el experimento una muestra cuyo resultado al someterse a ese experimento sea conocido. Esto permite asegurar que se ha realizado correctamente el experimento. Por tanto, se ha anotado el tipo de control.

Ejemplo: *Healthy heart samples*

Columna 11. NegativeControls. Los controles negativos son muestras que no se espera que produzca un resultado, utilizados para comparar la influencia de factores externos sobre la muestra. En caso de haberse usado, se ha anotado el tipo de control.

Ejemplo: *Wild-type*

Columna 12. Replicates. Se han considerado las veces que se ha repetido el experimento y para cuántas unidades de células, cultivos u organismos.

Ejemplo: *Biological replicates: wild type/mutants. H9C2 rat cardiac myoblasts , neonatal rat cardiac myocytes and C2C12 cell myotubes*

Columna 13. StatisticalAnalysis. La descripción, tal y como se han descrito en el artículo, de los métodos estadísticos empleados para determinar parámetros de significancia estadística, la significancia de los resultados y comparaciones de los resultados de distintos experimentos.

Ejemplo: *Data are presented as mean \pm SEM. Differences among groups were evaluated by one-way ANOVA and Student's t-test. Statistical analysis was performed using SPSS 20.0. A two-sided $p < 0.05$ was considered statistically significant.*

Columna 14. ValidationProcess. Para determinar la validez, se han indicado los parámetros que presenta el artículo de entre los controles positivos y negativos, las réplicas, el análisis estadístico, y predicciones *in silico*.

Ejemplo: *Comparison with other functional studies, positive and negative controls, replicates, polyphen & SIFT predictions.*

Columna 15. Reproducible. Según el nivel de detalle de la sección de métodos, se ha considerado si el artículo es reproducible o no.

Ejemplo: Yes

Columna 16. RobustnessData. Se ha realizado un resumen de los datos disponibles sobre el origen, transporte, tratamiento y almacenaje de las muestras.

Ejemplo: *All participants had provided written informed consent (...), genomic DNA was extracted from peripheral blood leukocytes with a Qiagen kit according to the manufacturer's instructions, Rat cardiac myoblast cell line H9C2 was kindly provided by (...). Cells were grown in DMEM supplemented with (...).*

Columna 17. FunctionallImpact. Se ha clasificado como “funcionalmente normal” o “funcionalmente anormal” según las recomendaciones del grupo ClinGen (Brnich et al., 2020), y se ha añadido una categoría de “desconocido” ante la ambigüedad de algunos estudios. Se ha incluido también una descripción de las consecuencias moleculares para mayor contexto.

Ejemplo: *Functionally abnormal. Significant formation of filamin C aggregates which might impair sarcomere function in all types of cells studied.*

Durante esta etapa, siguiendo el proceso descrito, se han leído los 149 artículos, de los cuales se han descartado 50, tal y como se ha descrito al inicio de esta sección. De las publicaciones aceptadas, se ha extraído, para cada una de ellas, la información relevante. De esta manera, se ha rellenado el Excel y se ha caracterizado la información funcional para todas las variaciones que han sido sujeto de estudio en estos estudios funcionales. Este proceso de extracción ha ocupado la mayor parte de la investigación, dado que, dependiendo de la complejidad del artículo, la lectura y extracción de la información de un artículo puede durar desde 15 minutos hasta 3 horas. Al realizar este Trabajo de Fin de Grado a lo largo del curso lectivo, en su totalidad, la extracción de información ha durado 21 semanas.

6.3. Revisión y estandarización

La última parte del desarrollo ha sido la revisión y estandarización de la fuente de información una vez completada la extracción de información para la lista de artículos. A partir de observaciones y problemas identificados durante la lectura, se han modificado algunos campos para estandarizarlos y ordenar la información de forma comprensible y acorde al modelo CSHG.

El primer punto que se ha abordado es el nombre de las variaciones. El formato de este difiere en función de si la variación se ha investigado a nivel genómico, de ADN o proteico, y según las preferencias de los investigadores. Se distinguen cuatro formatos distintos en la fuente de información generada: el identificador de dbSNP⁸, variación en secuencia de ADN codificante (prefijo

⁸ dbSNP: Base de datos de variaciones del NHI, National Health Institute de EEUU (<https://www.ncbi.nlm.nih.gov/snp/>).

“c.” según la nomenclatura HGVS⁹), variación en secuencia de proteína (prefijo “p.”, según la nomenclatura HGVS) y variación en secuencia genómica (prefijo “g.”, según la nomenclatura HGVS).

En las variaciones a nivel de proteína se han realizado dos cambios para estandarizar el formato: añadir el prefijo “p.” para casos en los que no se indicaba explícitamente en el artículo, y homogeneizar la nomenclatura de los aminoácidos. Para nombrar los aminoácidos, existen dos tipos de abreviaciones estándar. Una consiste en un código de tres letras (p. ej. Glu, Gly, His) y la otra consiste en un código de una letra (p. ek. E, G, H, respectivamente a los ejemplos anteriores). Dado que cada artículo empleaba de forma aleatoria un código u otro, se ha optado por ajustar todos los nombres a la abreviación de una letra, dado que suponían la menor cantidad de cambios.

Por otra parte, aunque todos los formatos que aparecen son conocidos en el ámbito de investigación de variaciones, la nomenclatura de la base de datos dbSNP es más limitante, dado que es poco intuitivo al no proporcionar ningún dato sobre la variación y requiere conocer o consultar su base de datos para identificar la variación. Por tanto, se ha decidido reducir a los tres nombres aceptados por la HGVS. Para ello, se ha consultado la base de datos y se ha sustituido manualmente los nombres dbSNP con el nombre de la nomenclatura HGVS correspondiente. Los tres tipos de nombres resultantes se han separado por columnas (Tabla 1) para mayor claridad.

Genomic position	Variant in transcript (c.)	Variant in protein (p.)
Chr7:g.128854898_128854915del	c.7251D1G>A	p.A2430V

Tabla 1. Reestructuración del campo “Variant Name” a tres columnas, con ejemplos de cada formato

Para los controles, para cumplir con el formato *integer* (valor numérico) definido en el CSHG, se han separado en dos columnas: el número de tipos distintos de controles usados y el tipo. De forma similar, para las réplicas se han dividido los datos en los dos géneros de réplicas existentes: biológicas (cuántas versiones biológicas se han usado en el estudio) y técnicas (repeticiones para comprobar el estado de los aparatos empleados). El resultado es el mostrado en la Tabla 2.

positiveControls	PosControl	negativeControls	negControl	Biological replicates	Technical replicates	replicates
NA	Healthy heart samples	1	WT	4	NA	Biological replicates: wild-type/mutants H9C2 rat cardiac myoblasts , neonatal rat cardiac myocytes and C2C12 cell myotubes

Tabla 2. Reestructuración de los campos sobre controles y réplicas: “positiveControls” y “negativeControls” indican el número de controles, “PosControl” y “negControl” describen el tipo, “Biological replicates” y “Technical replicates” indican el número de réplicas y “replicates” indica el tipo.

⁹ HGVS: *Human Genome Variation Society*. Este grupo ha desarrollado una nomenclatura que se ha convertido en el estándar para las variaciones (<https://varnomen.hgvs.org/>).

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Una vez realizados estos cambios, se ha dado por finalizado el desarrollo de la fuente de información, la cual ha sido concluida tal y como se presenta en el Anexo 1 de este Trabajo de Fin de Grado.

7. RESULTADOS

Este capítulo describirá la última etapa del ciclo de diseño seguido en este Trabajo de Fin de Grado. Se hará una vista general de los resultados de la extracción de la información y se explicará el caso de uso aplicado a la fuente de información resultante, el cual se ha realizado con datos de pacientes reales.

7.1. Análisis de resultados

El proceso de desarrollo de la fuente de información ha constado de varias actividades. Primero, se han determinado dos repositorios de referencia. Seguidamente, se han seleccionado, en dos búsquedas, los artículos de estudios funcionales relevantes a las cardiopatías familiares. Finalmente, se ha extraído la información de los artículos. En esta sección, revisaremos estos pasos, mostrando los resultados obtenidos tras cada uno.

Primera búsqueda. Usando las búsquedas avanzadas descritas en la sección 6.2.1., se han realizado un total de 912 consultas, de las cuales solo 152 han retornado artículos. En total ha habido 593 resultados, asociados a 100 genes. Tras el primer filtrado (Figura 6) se han descartado 268 publicaciones y se ha observado que gran parte de estos rechazados hacían alusión a estudios funcionales, pero no hacían un estudio de una variación. La aparición de tantos artículos no relacionados directamente con variaciones seguramente haya ocurrido porque el término “variación” no está incluido explícitamente en las búsquedas definidas. La razón tras esta decisión es que se consideró que la combinación del término “*functional assay*” y el nombre del gen generaba un buen equilibrio en la consulta sin restringir ni generalizar excesivamente. Tras aplicar el filtrado de artículos duplicados (asociados a más de un gen), se ha reducido a 180 artículos.

Segunda búsqueda. Como se ha explicado en la sección 6.2.2, se ha repetido el proceso de búsqueda y filtrado con un nuevo término añadido. La cantidad de resultados ha aumentado en un 120% con respecto a la primera búsqueda, retornando 1306 artículos (Tabla 3). Sin embargo, la cantidad de artículos seleccionados ha sido similar a la primera búsqueda y ha incrementado solo en un 14,77% (373 artículos). Estos resultados refuerzan la precisión del criterio personal de selección de artículos e indican, como se esperaba, que el término “*functional analysis*” es demasiado general. Tras eliminar los duplicados, la lista se ha reducido a 236 artículos, lo cual supone un aumento del 30% respecto a la primera búsqueda.

	<i>Primera búsqueda</i>	<i>Segunda búsqueda</i>
Resultados	593	1306
Artículos seleccionados	En bruto: 325 Filtrado: 180	En bruto: 373 Filtrado: 236
% Seleccionado (en bruto)	54,81%	29,21%
% Descartado (en bruto)	45,19%	70,79%

Tabla 3. Resumen de resultados en la búsqueda inicial y la segunda búsqueda en bruto (apuntados directamente de los resultados de la consulta, sin comprobar los duplicados) y filtrados (tras eliminar duplicaciones).

Extracción de información. Finalmente, se han leído los artículos asociados a la lista reducida de 35 genes relevantes, lo cual corresponde a 149 artículos (ver último párrafo de la sección 6.2.2.). Tras el segundo filtrado (Figura 7), la cantidad final es de 99 artículos (Tabla 4). De estos, se ha extraído la información funcional para cada variación estudiada en estas publicaciones. Como resultado, la fuente de información resultante, la cual puede consultarse en el Anexo 1 de esta memoria, describe un total de 285 variaciones genéticas que han sido asociadas a la lista de cardiopatías familiares seleccionadas.

	ANTES DE LA LECTURA	TRAS LA LECTURA
Genes	35	40
Artículos	149	99

Tabla 4. Cantidades finales de genes y artículos tras la lectura, donde se han descartados más artículos (ver Filtro 2, Figura 7).

En la Figura 8 podemos ver una comparación, para los 35 genes seleccionados, de los artículos que había asociados a cada gen y el número de variaciones que se ha apuntado finalmente de cada uno. De esta forma, se puede ver cómo ha afectado el segundo filtrado de artículos, habiendo genes de los que finalmente no ha habido variaciones que anotar. En contrapartida, las variaciones de los artículos se repartían a lo largo de 40 genes, incluyendo algunos que no estaban en la lista seleccionada, dado que muchos de los artículos estudiaban más de un gen. Dada la pequeña cantidad que suponían (20 variaciones) y para no excluir la información, se han incluido en la fuente de información las variaciones de genes que no estaban en la lista.

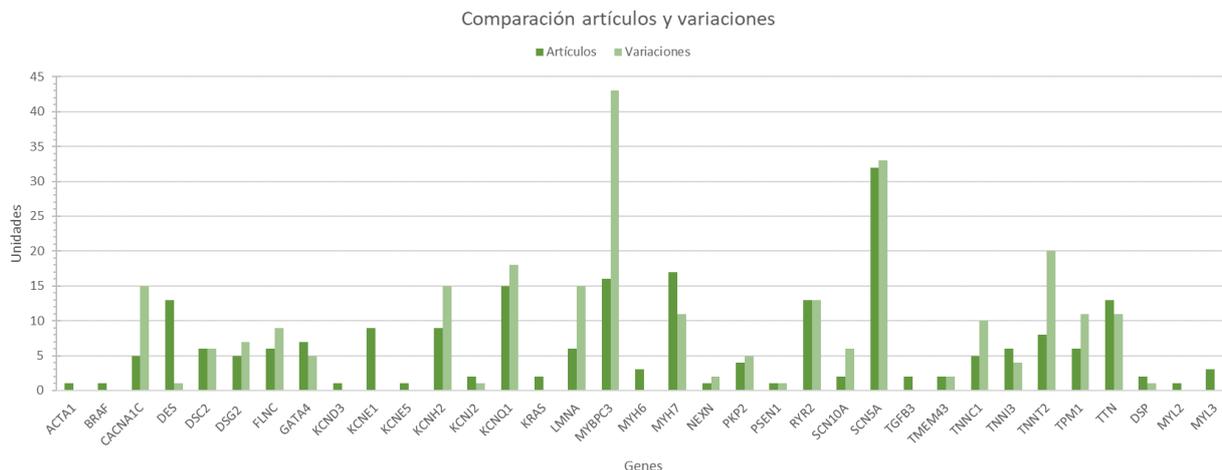


Figura 8. Comparación de cantidad de artículos asociados a los genes y cantidad de variaciones que finalmente se han recogido en la extracción de información.

Por otra parte, los genes pueden estar asociados a varias cardiopatías a la vez, lo cual puede apreciarse en la Figura 9 donde se muestra la cantidad de variaciones recogidas, en la fuente de información, de cada gen y la enfermedad a las que están relacionadas. Se puede observar que el gen con más variaciones estudiadas es el MYBPC3, y la mayoría están asociadas a la Miocardiopatía Hipertrófica (HCM). Le sigue el gen SCN5A, donde la mayoría de las variaciones son con relación al

Síndrome de Brugada (BrS). De hecho, este gen, junto con el RYR2, son los que están asociadas a más cardiopatías a la vez (cuatro en ambos casos).

La enfermedad con más variaciones detectadas ha sido la HCM, con 101 variaciones asociadas, seguida de la Miocardiopatía Dilatada, una enfermedad bastante similar. Las enfermedades para las que menos información se ha extraído han sido el Síndrome de Loeys-Dietz y la Miocardiopatía Restrictiva (Tabla 5), y el síndrome de Marfan, que ha sido la única patología para la que no se han encontrado estudios funcionales.

Variaciones por gen para cada enfermedad

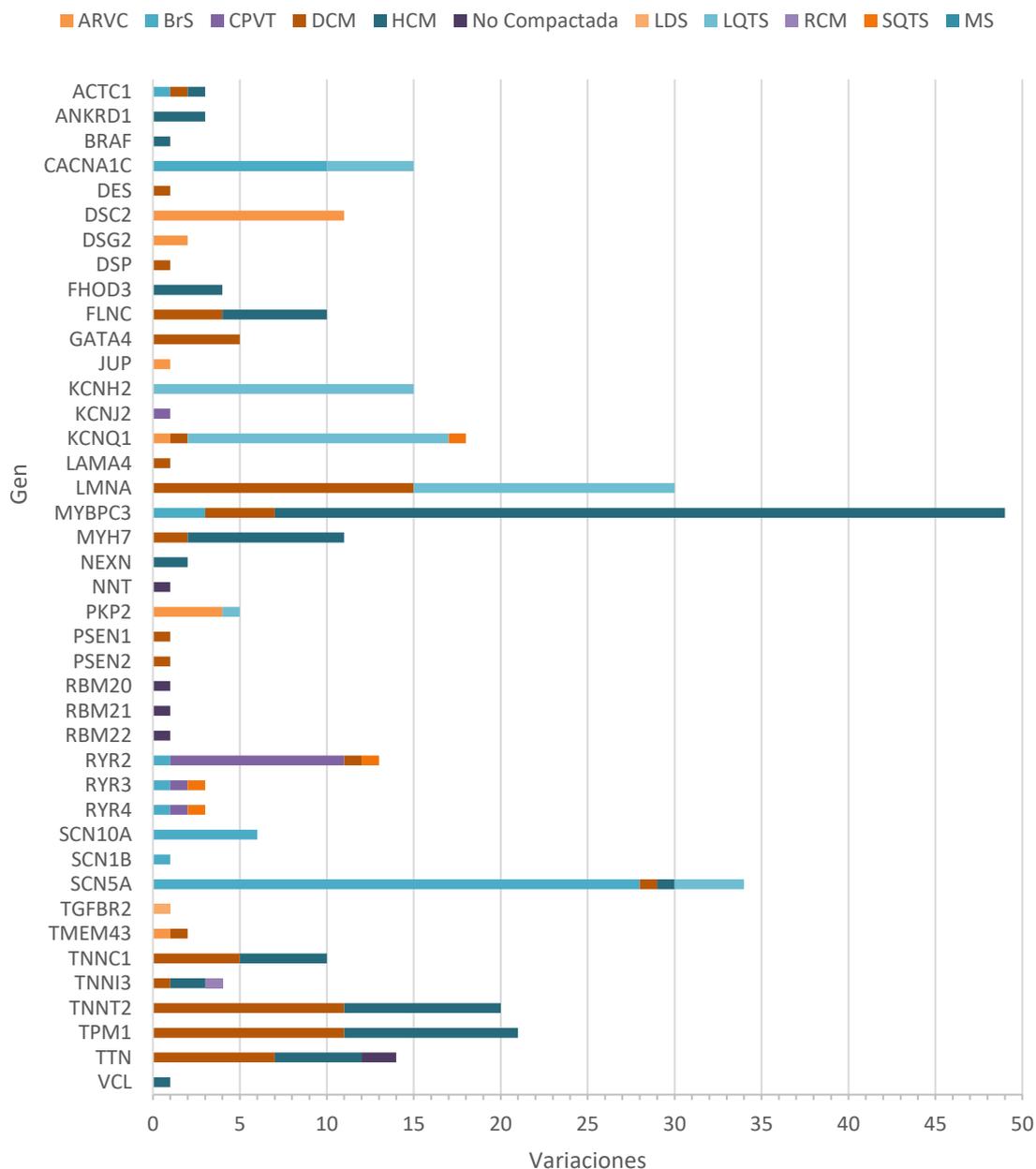


Figura 9. Cantidad de variaciones estudiadas de cada gen y las cardiopatías a las que se han asociado.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

	ARVC	BrS	CPVT	DCM	HCM	No Compactada	LDS	LQTS	RCM	SQTS	MS
Variaciones	20	52	13	75	101	6	1	55	1	4	0

Tabla 5. Número de variaciones asociadas a cada enfermedad.

Se ha analizado también la reproducibilidad, donde la mayoría de los estudios se han considerado reproducibles (94,74%) dado que se detallaban claramente los procedimientos de laboratorio. Finalmente, se ha comprobado el impacto funcional de las variaciones (ver la descripción de la columna 17, Sección 6.2.3.) y se ha observado que la mayoría de las variaciones han sido consideradas “funcionalmente anormal” (79,65%). Luego un 18,60% han sido consideradas “funcionalmente normales” y un 5% se han marcado como “Desconocido”. Por tanto, podemos considerar que los estudios recogidos son relevantes, dado que la mayoría determinan consecuencias moleculares definidas, y una gran parte seguramente pueda justificar una clasificación de patogenicidad.

Una de las innovaciones de este trabajo ha sido incluir los estudios de tipo *in silico* (p.ej. simulaciones computacionales). Esta clase de experimentos han constituido, aproximadamente, un 28% de los estudios analizados. La utilidad de la fuente de información, así como la relevancia de los análisis *in silico*, se evaluará mediante un caso de uso, que detallaremos en la siguiente sección.

7.2. Caso de uso con VCF de pacientes reales

La fuente de información desarrollada recopila la información funcional de un gran número variaciones genéticas, y constituye una herramienta con gran potencial para dar soporte a la interpretación de variaciones y su clasificación. Al tener una fuente de datos como esta, se puede hacer una evaluación funcional de variaciones de forma automática, facilitando la evaluación del criterio PS3/BS3 en el ámbito clínico o analítico.

Con esta aplicación práctica en mente, se ha desarrollado un caso de uso con datos de pacientes reales para evaluar la utilidad de la fuente de información que se ha generado. Esta experiencia consistirá en comprobar si las variaciones recopiladas en la fuente de información se encuentran en los datos de pacientes reales que han sido diagnosticados alguna cardiopatía familiar. A continuación, se describirá cómo se ha llevado a cabo la comprobación y los resultados obtenidos.

7.2.1. Implementación

Los datos de pacientes han sido proporcionados por el Hospital General Universitario de Alicante, en formato VCF (Formato de Llamada de Variación), un tipo de archivo empleado ampliamente en bioinformática para almacenar secuencias o variaciones genéticas. Se han recibido los datos de 42 pacientes que han sido diagnosticados alguna cardiopatía familiar y se ha empleado una versión anotada de los VCF, con formato *.csv. Al estar anotados, estos archivos contienen información ampliada de las variaciones, incluyendo no solo datos estructurales e identificativos de la variación (p. ej. cromosoma, gen, posición), sino también interpretaciones clínicas y predicciones de patogenicidad.

El objetivo de este experimento es detectar las variaciones de la fuente de información que estén presentes en alguno de los 42 ficheros de pacientes. Por tanto, para realizar la comparación de las variaciones de forma rápida y efectiva se ha optado por desarrollar un código de Python, el cual se presenta en el Anexo 2 de esta memoria. De forma resumida, el proceso seguido ha constado de los siguientes pasos:

- 1- Crear una versión reducida del VCF anotado para cada paciente que solo incluya las variaciones dadas en los mismos genes que están presentes en la fuente de información.
- 2- Comparar los nombres de las variaciones de la fuente de información con los nombres de la versión reducida del VCF anotado de cada paciente.
- 3- Devolver para cada paciente las variaciones que coincidan, mostrando el nombre de la variación y su impacto funcional descrito en la fuente de información.

7.2.2. Resultados caso de uso

Tras ejecutar el código resumido en la sección previa, se han obtenido las coincidencias de variaciones presentes en los pacientes. En total, se han detectado 9 variaciones a lo largo de los 42 pacientes. Aunque es un número reducido, se puede considerar un resultado coherente, dado que, frente a los millones de posibles variaciones que pueden presentar los pacientes, la fuente de información generada abarca una cantidad muy concreta. Ha habido 6 pacientes en los que no se ha detectado ninguna variación de las descritas en la fuente de información. Los otros 36 pacientes han mostrado entre 1 – 4 coincidencias, siendo el promedio de coincidencia de 2 variaciones por paciente.

Muchas de las variaciones detectadas se repetían entre los pacientes, lo cual indica que son variaciones comunes. Esto concuerda con el conjunto de enfermedades con el que estamos trabajando, dado que comparten características, siendo esta una de las razones que dificultan su diagnóstico. Para los pacientes con coincidencias, al comparar los diagnósticos de los médicos con la enfermedad a la que se han asociado las variaciones en la fuente de información, se ha comprobado que muchas no coinciden. Por ejemplo, el Paciente 1 presenta las variaciones p.I962V, p.L1092P y p.R1973Q, asociadas a CPTV y Síndrome de Brugada en la fuente de información, pero el diagnóstico clínico ha sido Miocardiopatía Hipertrófica (HCM). Dada la similitud entre cardiopatías mencionada, una misma variación pueda estar asociada a distintas enfermedades. Por tanto, estos casos en los que no coinciden las enfermedades relacionadas no implican que, o bien el diagnóstico del hospital, o bien la fuente de información, sean erróneos. Simplemente, no existe actualmente evidencia funcional que asocie las variaciones con las enfermedades diagnosticadas.

Las 9 variaciones de la fuente de información detectadas en los pacientes se resumen en la Tabla 6. Analizando la información funcional y la clasificación del VCF, podemos ver que 5 de las variaciones se consideran coherentes dado que el carácter de la anotación clínica coincide con el impacto funcional determinado en la fuente de información desarrollada. En concreto, todas se clasifican como carácter benigno en los VCF anotados y en la fuente de información se determina que son funcionalmente normales. Por el contrario, los resultados de las otras 4 variaciones son conflictivos, puesto que presentan resultados discordes en el VCF y en la fuente de información.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

VARIACIÓN	GEN	FREC.	PATOLOGÍA	IMPACTO FUNCIONAL	CLASIFICACIÓN VCF	COHERENTE	NOTAS
p.I962V	SCN10A	20	Taquicardia Ventricular Polimórfica Catecolaminérgica (CPVT)	Funcionalmente normal	Benigna	Sí	-
p.L1092P	SCN10A	18	Taquicardia Ventricular Polimórfica Catecolaminérgica (CPVT)	Funcionalmente normal	Benigna	Sí	-
p.R1973Q	CACNA1C	1	Síndrome de Brugada (BrS)	Funcionalmente normal	Benigna/Probablemente benigna	Sí	-
p.K263R	TNNT2	7	Miocardiopatía Hipertrófica (HCM)	Funcionalmente anormal	Benigna/Probablemente benigna	No	Método de estudio <i>in silico</i> .
p.S236G	MYBPC3	11	Miocardiopatía Hipertrófica (HCM)	Funcionalmente normal	Benigna/Probablemente benigna	Sí	-
p.V158M	MYBPC3	8	Miocardiopatía Hipertrófica (HCM)	Funcionalmente anormal	Benigna	No	Método de estudio <i>in silico</i> .
p.R278C	TNNT2	1	Miocardiopatía Hipertrófica (HCM)	Funcionalmente anormal	Interpretaciones de patogenicidad conflictivas	No	Estudio incluido en perfil ClinVar
p.G1885E	RYR2	1	Taquicardia Ventricular Polimórfica Catecolaminérgica (CPVT)	Funcionalmente anormal	Benigna	No*	*Efecto benigno en solitario. Efecto patogénico combinada con otra mutación.
p.S509P	SCN10A	1	Síndrome de Brugada (BrS)	Funcionalmente normal	Benigna	Sí	-

Tabla 6. Resumen de las 10 variaciones de la fuente de información identificadas en los pacientes reales.

Dos de las variaciones coherentes, p.R1973Q y p.S236G se presentan como “benignas/ posiblemente benignas”, por lo que la existencia de un estudio funcional que reitere la benignidad de la variación aumentaría la confianza de esta clasificación. Se han revisado los artículos de la fuente de información y se ha comprobado que ninguno de los dos se incluye en el perfil de las variaciones en ClinVar^{10,11}, la base de datos empleada para la anotación del VCF. Habiendo revisado los artículos correspondientes, se considera que la coincidencia de la variación p.R1973Q es un refuerzo a la clasificación como benigna. En cuanto a la variación p.S236G, no se puede asegurar con la misma confianza que se pudiera añadir al perfil de la variación, dado que el artículo realiza un estudio de tipo *in silico*, el cual no se considera actualmente en los criterios PS3/BS3. Estos análisis consisten en

¹⁰ Perfil de variación p.R1973Q: National Center for Biotechnology Information. ClinVar; [VCV000093419.21], <https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000093419.21> (accessed July 2, 2023).

¹¹ Perfil de variación p.S236G: National Center for Biotechnology Information. ClinVar; [VCV000042786.47], <https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000042786.47> (accessed July 2, 2023).

métodos computacionales que emplean predicciones, simulaciones y modelos de las moléculas. Si bien tienen ventajas como mayor velocidad y el factor coste-efectividad, se ven limitados por aspectos como modelos fijos y una falta de comprensión y flexibilidad para simular el ambiente biológico natural.

Por otra parte, se ha revisado la información funcional extraída sobre las variaciones conflictivas para buscar posibles causas de la diferencia de carácter. Se ha observado que dos de las cuatro variaciones, p.K263R y p.V158M, se corresponden también a estudios de sistemas *in silico*, por lo que podría considerarse que esta discordia supone un argumento a favor de no incluir los experimentos *in silico* a la hora de evaluar los criterios PS3/BS3.

Otra de las variaciones que difieren entre los datos funcionales y los datos clínicos es la variación p.R278C, cuya interpretación en el VCF es conflictiva. Se ha comprobado la variación en ClinVar¹² y se ha visto que el artículo correspondiente de la fuente de información está presente en su perfil, asociada a tres interpretaciones: dos concluidas como “no especificada”, por lo que no se indicó ninguna enfermedad concreta, y una como “miocardiopatía”, lo cual engloba aquellas enfermedades donde la función muscular del corazón se ve afectado. Por tanto, teniendo la información funcional como justificación, se plantea que de estas tres interpretaciones se podría considerar dominante la asociación a miocardiopatías, aunque no se especifique una en concreto. Afinar la condición concreta dependerá de futuras investigaciones funcionales.

Finalmente, la última de las variaciones conflictivas, p.G1885E, es un caso peculiar. En solitario altera la actividad normal de la proteína afectada, pero no causa enfermedad, sino que aumenta de forma controlada la actividad del proceso biológico en el que participa la proteína, por lo que se consideraría una variación benigna. Sin embargo, en combinación con otra variación (p.G1886S) el efecto es totalmente contrario y sí podría considerarse patogénica. Dado que no se presenta esta otra variación, se puede considerar que la información es coherente. Se ha comprobado también el perfil de ClinVar¹³ para esta variación y se ha confirmado que el artículo de la fuente de información está incluido entre sus referencias.

7.2.3. Limitaciones

El caso de uso realizado ha dado unos resultados satisfactorios, dado que se han identificado un número de variaciones genéticas coherente a partir de la fuente de información que se ha generado a lo largo de este Trabajo de Fin de Grado. Sin embargo, hay algunas limitaciones y posibles mejoras que quedan pendientes.

Primero, la inclusión de estudios *in silico* debe reevaluarse, puesto que, aunque se ha extraído información de estos estudios, al aplicarla al caso de uso se han obtenido resultados cuestionables. Revisando los datos de estudios *in silico* que se han incluido en la fuente de información, se ha visto que la mayor parte de estos estudios presentan un número escaso de controles y repeticiones. Esta falta de variedad podría ser comprensible dado que, al realizarse computacionalmente, hay mayor

¹² Perfil de variación p.R278C: National Center for Biotechnology Information. ClinVar; [VCV000012411.56], <https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000012411.56> (accessed July 2, 2023).

¹³ Perfil de variación p.G1885E: National Center for Biotechnology Information. ClinVar; [VCV000036746.45], <https://www.ncbi.nlm.nih.gov/clinvar/variation/VCV000036746.45> (accessed July 2, 2023).

confianza en que los resultados sean óptimos. Sin embargo, esta falta de repeticiones y controles hace dudar la calidad de la evidencia que proporciona el estudio. Por tanto, dada la gran diferencia que supone un experimento físico frente a un experimento virtual, una solución posible sería definir parámetros y procesos de validación específicos para los estudios *in silico*. Para ello, haría falta realizar una caracterización de los estudios *in silico*, para conocer sus características, la viabilidad y cantidad óptima de controles y repeticiones, y demás consideraciones a tener en cuenta.

En segundo lugar, una limitación de la fuente de información son los filtros que se han tenido que aplicar. La lista de genes era muy extensa, pero para ajustarnos al tiempo disponible, se ha tenido que reducir considerablemente. Esto ha causado que, a pesar de que los genes seleccionados son los más relevantes, muchos otros que podrían ser interesantes han tenido que excluirse. Consecuentemente, podría ser la razón por la que no se han obtenido resultados con relación al Síndrome de Marfan y por la que solo se han podido identificar unas pocas variaciones para enfermedades menos comunes como el Síndrome de Loeys-Dietz y la Miocardiopatía Restrictiva.

Con esta última sección se da por concluido el caso de estudio, el cual nos ha permitido comprobar la utilidad de la fuente de información y aplicarla de forma práctica en datos de pacientes reales. Se ha mostrado también cómo el análisis de la información funcional puede influir de manera significativa en la interpretación de variaciones y la complejidad que presenta.

8. CONCLUSIONES Y TRABAJOS FUTUROS

En este Trabajo de Fin de Grado se ha diseñado y desarrollado una fuente de información sobre estudios funcionales que permite caracterizar variaciones genéticas asociadas a cardiopatías familiares y facilita la evaluación del criterio PS3/BS3 de las Guías ACMG/AMP, uno de los criterios con más peso a la hora de interpretar variaciones.

En base a la metodología del *Design Science*, se definieron inicialmente unos subobjetivos (capítulo 2), los cuales se plantearon como preguntas de investigación para guiar el trabajo que se pretendía realizar a lo largo de las etapas del ciclo de diseño. De esta manera, se ha llevado a cabo la investigación del problema, en la que hemos caracterizado los estudios funcionales, su problemática, y el uso de modelado conceptual. Seguidamente, se ha realizado el diseño y desarrollo de la fuente de información, donde se ha estructurado la información relevante de los estudios funcionales, se han seleccionado los recursos y se ha extraído la información necesaria para completar la fuente de información. Adicionalmente, se ha realizado un caso de uso para aplicar la fuente de información de forma práctica y comprobar su utilidad en base a datos clínicos reales.

Habiendo completado todas las etapas, en este capítulo se explicará brevemente cómo se ha resuelto cada pregunta de investigación y se comentarán las futuras líneas de trabajo que surgen a partir de este Trabajo de Fin de Grado.

8.1. Cumplimiento de objetivos iniciales

En esta sección resumiremos las respuestas a las preguntas de investigación de los objetivos planteados previamente en el segundo capítulo de esta memoria de Trabajo de Fin de Grado.

Objetivo 1. Investigación del problema

Las respuestas a las preguntas de este primer objetivo se han respondido, principalmente, a lo largo del capítulo 5, con soporte de otras secciones para una mejor contextualización.

PI1- ¿Qué es un estudio funcional?

Un estudio funcional evalúa el efecto de una variación genética sobre los procesos biológicos. Es decir, estos estudios permiten definir las consecuencias moleculares causadas por un cambio concreto en un gen y entender cómo se asocia a una enfermedad, tal y como se ha detallado en la introducción (sección 1.2.2.) y en la caracterización de los estudios (sección 5.1.).

PI2 - ¿Cuáles son las problemáticas asociadas al uso de los estudios funcionales?

La problemática principal es la falta de un recurso que recopile la información funcional determinada mediante estudios funcionales. En consecuencia, la información sufre una gran dispersión, además de una gran heterogeneidad. Esta última está causada por la variabilidad de estudios y de interpretaciones, y por la falta de estándares de evaluación hasta hace poco, como definido en la sección 5.1.

PI3- ¿Existe algún modelo que nos permita representar la información asociada a estudios funcionales?

Para la representación de información funcional se ha considerado que el mejor método es el modelado conceptual dado que permite representar campos complejos como el genómico como se ha detallado en la sección 5.2. Concretamente, se ha empleado el Modelo Conceptual del Genoma Humano (CSHG), el cual se ha descrito en la sección 5.2.1, para poder integrar la información funcional con el resto de los datos genómicos de las variaciones, proporcionando una perspectiva holística y precisa.

PI4- ¿Es necesario introducir mejoras a este modelo conceptual?

El modelo CSHG no incluía ningún concepto relacionado a la información funcional, por lo que se ha realizado una ampliación a la vista de bibliografía, descrita en la sección 5.2.1., para incluir dos conceptos nuevos: los estudios funcionales y el impacto funcional resultante de una variación. Esta ampliación se ha detallado en la sección 6.1.1., dado que formaba parte del proceso de diseño.

Objetivo 2. Diseño y desarrollo de la fuente de información

Esta etapa se ha descrito en el capítulo 6 de la memoria, por lo que las respuestas a este segundo objetivo han sido resueltas detalladamente en dicho capítulo.

PI1- ¿Cómo se estructurará la fuente de información a partir de la información relevante de los estudios funcionales?

A partir de los estudios realizados por ClinGen, se ha identificado la información necesaria para evaluar los estudios funcionales. Estos conceptos fundamentales han sido la referencia para la ampliación del CSHG explicada en la sección 6.1.1., la cual, a su vez, ha servido de base para la estructura de la fuente de información, descrita en la sección 6.1.2.

PI2- ¿Cuáles son los estudios funcionales de interés para las cardiopatías familiares?

La información de estudios funcionales se encuentra dispersa en la literatura científica, por lo que se han elegido dos repositorios de publicaciones, *PubMed* y *ScienceDirect*, de los que se seleccionarán los artículos sobre estudios funcionales relevantes mediante consultas de búsqueda avanzada detalladas en las secciones 6.2.1. y 6.2.2.

PI3- ¿Cómo se va a obtener la información de los estudios funcionales?

El proceso de extracción de información se ha descrito en la sección 6.2.3., y ha consistido en la lectura de los artículos seleccionados a partir de las búsquedas avanzadas en los repositorios, y la anotación simultánea de la información relevante en la fuente de información. De esta manera, siguiendo las consideraciones descritas en la sección correspondiente, se ha poblado la fuente de información con los datos correspondientes.

PI4- ¿Se puede estandarizar la información extraída?

Ante la heterogeneidad de la información, se ha revisado a medida que se iba extrayendo para determinar las modificaciones que fueran necesarias y, tras la lectura y recopilación de los datos, se han unificado las nomenclaturas y se ha reestructurado la fuente de información, como se narra en la sección 6.3.

Objetivo 3. Caso de uso

Las respuestas a las preguntas de investigación de este objetivo se encuentran en el capítulo 7, el correspondiente al caso de uso que se ha llevado a cabo.

PI1- ¿Qué resultados se han obtenido en el diseño y desarrollo de la fuente de información?

Los resultados de las subetapas del diseño y desarrollo se han analizado en la sección 7.1., y el resultado final ha sido una fuente de información que recopila, a partir de estudios funcionales, la información funcional relevante para casi 300 variaciones asociadas a una lista de cardiopatías familiares y que se dan a lo largo de unos 40 genes.

PI2- ¿Cómo se puede comprobar la utilidad de esta fuente de información?

Para validar la utilidad de la fuente de información, se ha desarrollado un caso de uso que muestra la posible aplicación clínica del recurso generado. Esta prueba, descrita en la sección 7.2.1, ha consistido en comprobar, mediante un código de Python, si las variaciones que se han caracterizado funcionalmente en la fuente de información están presentes en pacientes reales de cardiopatías familiares, empleando los datos VCF proporcionados por los clínicos.

PI3- ¿La fuente de información cumple los objetivos esperados?

Los resultados del caso de uso, detallados en la sección 7.2.2., han sido satisfactorios dado que se ha identificado un número coherente de variaciones en los pacientes reales a partir de la fuente de información que hemos generado con este trabajo, y se han obtenido resultados realistas que han dado pie a un análisis crítico de la información.

PI4- ¿Cuáles son los fallos, limitaciones, y mejoras identificadas?

Tal y como se explica en la sección 7.2.3., la inclusión de los estudios funcionales realizados *in silico* no parece ser conveniente, dado que ha dado lugar a información que no se ha podido evaluar de forma adecuada al estilo de experimento. Otra clara limitación es que este trabajo se ha realizado para una cantidad de datos acotada según unas enfermedades y unos genes seleccionados, por lo que hay una gran posibilidad de ampliación.

8.2. Trabajos futuros

En base a la labor realizada, se han identificado posibles líneas de trabajo que podrían extenderse para ampliar este Trabajo de Fin de Grado y su alcance.

En primer lugar, una de las líneas de trabajo más claras es la ampliación de la fuente de información. Se podrían incluir, por ejemplo, los artículos y genes relacionados a cardiopatías familiares que se han tenido que excluir de este Trabajo de Fin de Grado. Incluso, a lo largo de un extenso periodo de tiempo, se podría ampliar o replicar para otras patologías, aplicando lo que se ha aprendido con el desarrollo de la fuente actual, y siguiendo el diseño y desarrollo metodológico propuesto en el capítulo 4.

Además, como se ha comentado en la sección 7.2.3., se podría realizar una caracterización de los estudios *in silico* para determinar parámetros relevantes y desarrollar un sistema de evaluación mejor planteado de cara a este tipo de estudios para poder incluirlos como información funcional de las variaciones. Análogamente, para el campo de la robustez del estudio (ver listado de conceptos de

la sección 6.1.1.), también podría realizarse una caracterización que permita evaluar esta característica de forma simple. Actualmente, muchos de los campos de la fuente de datos son un párrafo de texto muy extenso dado que la única descripción de las guías era que tratara la integridad de las muestras según el origen, manipulación y almacenaje de las muestras. Este campo podría condensarse si se definen de forma concisa y correcta los parámetros relevantes de la robustez.

Otra posible ampliación es diseñar y crear una interfaz de visualización de los datos funcionales. Como se ha visto en el caso de uso, la información tiene aplicaciones clínicas inmediatas, por lo que hacerla accesible a los usuarios de forma estructurada y estética sería un posible enfoque. Se buscaría que los usuarios puedan consultar la fuente de forma simple y eficiente, y podría estar dirigida tanto a clínicos como a analistas o demás investigadores. Sin embargo, para ello, habría que realizar primero una revisión de los datos que debe recoger y su presentación, para asegurar que la información que se publique sea lo más completa y correcta posible.

Por otra parte, en cuanto al desarrollo de la fuente de información, todo el proceso de selección y extracción de información se ha realizado manualmente. Por tanto, podría realizarse una tentativa para automatizar estos procesos empleando inteligencia artificial o *machine learning* para la comprensión de textos, aprovechando el proceso definido en este trabajo y el conocimiento adquirido durante el desarrollo de este Trabajo para el entrenamiento del sistema.

9. BIBLIOGRAFÍA

- Abeyruwan, S., Vempati, U. D., Küçük-McGinty, H., Visser, U., Koleti, A., Mir, A., Sakurai, K., Chung, C., Bittker, J. A., Clemons, P. A., Brudz, S., Siripala, A., Morales, A. J., Romacker, M., Twomey, D., Bureeva, S., Lemmon, V., & Schürer, S. C. (2014). Evolving BioAssay Ontology (BAO): modularization, integration and applications. *Journal of Biomedical Semantics*, 5(S1), S5. <https://doi.org/10.1186/2041-1480-5-S1-S5>
- AP INBIO. (2021). *PROGRAMA INBIO 2021 SUBPROGRAMA DE FOMENTO DE ACCIONES PREPARATORIAS (AP) DE APOYO A LA EXPLORACIÓN Y FORMULACIÓN DE FUTUROS PROYECTOS DE INVESTIGACIÓN /INNOVACIÓN MEMORIA CIENTÍFICO-TÉCNICA AP TÍTULO DE LA ACCIÓN PREPARATORIA*. www.inbio.es
- Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korb, J. O., Lander, E. S., Lee, C., Lehrach, H., ... Schloss, J. A. (2015). A global reference for human genetic variation. In *Nature* (Vol. 526, Issue 7571, pp. 68–74). Nature Publishing Group. <https://doi.org/10.1038/nature15393>
- Bernasconi, A., Ceri, S., Campi, A., & Masseroli, M. (2017). *Conceptual Modeling for Genomics: Building an Integrated Repository of Open Data* (pp. 325–339). https://doi.org/10.1007/978-3-319-69904-2_26
- Bornberg-Bauer, E., & Paton, N. W. (2002). *Keywords: conceptual data model (CDM), database, entity-relationship (ER), unified modelling language (UML), biological data Conceptual data modelling for bioinformatics*. <https://academic.oup.com/bib/article/3/2/166/181814>
- Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Soldatova, L. N., Stoeckert, C. J., Turner, J. A., & Zheng, J. (2010). Modeling biomedical experimental processes with OBI. *Journal of Biomedical Semantics*, 1(S1), S7. <https://doi.org/10.1186/2041-1480-1-S1-S7>
- Brnich, S. E., Abou Tayoun, A. N., Couch, F. J., Cutting, G. R., Greenblatt, M. S., Heinen, C. D., Kanavy, D. M., Luo, X., McNulty, S. M., Starita, L. M., Tavtigian, S. V., Wright, M. W., Harrison, S. M., Biesecker, L. G., Berg, J. S., Abou Tayoun, A. N., Berg, J. S., Biesecker, L. G., Brenner, S. E., ... Topper, S. (2020). Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. In *Genome Medicine* (Vol. 12, Issue 3). BioMed Central Ltd. <https://doi.org/10.1186/s13073-019-0690-2>
- Brnich, S. E., Rivera-Muñoz, E. A., & Berg, J. S. (2018). Quantifying the potential of functional evidence to reclassify variants of uncertain significance in the categorical and Bayesian interpretation frameworks. *Human Mutation*, 39(11), 1531–1541. <https://doi.org/10.1002/humu.23609>
- García S., A., Palacio, A. L., Román, J. F. R., Casamayor, J. C., & Pastor, O. (2020). Towards the understanding of the human genome: A holistic conceptual modeling approach. *IEEE Access*, 8, 197111–197123. <https://doi.org/10.1109/ACCESS.2020.3034793>
- Giglio, M., Tauber, R., Nadendla, S., Munro, J., Olley, D., Ball, S., Mitraka, E., Schriml, L. M., Gaudet, P., Hobbs, E. T., Erill, I., Siegele, D. A., Hu, J. C., Mungall, C., & Chibucos, M. C. (2019). ECO, the

- Evidence & Conclusion Ontology: community standard for evidence information. *Nucleic Acids Research*, 47(D1), D1186–D1194. <https://doi.org/10.1093/nar/gky1036>
- Hieter, P., & Boguski, M. (1997). Functional genomics: It's all how you read it. In *Science* (Vol. 278, Issue 5338, pp. 601–602). <https://doi.org/10.1126/science.278.5338.601>
- Hurtado, C. (2022). Precision medicine: concepts applications and projections. In *Revista Medica Clinica Las Condes* (Vol. 33, Issue 1, pp. 7–16). Ediciones Doyma, S.L. <https://doi.org/10.1016/j.rmcl.2022.01.002>
- Kanavy, D. M., McNulty, S. M., Jairath, M. K., Brnich, S. E., Bizon, C., Powell, B. C., & Berg, J. S. (2019). Comparative analysis of functional assay evidence use by ClinGen Variant Curation Expert Panels. *Genome Medicine*, 11(1). <https://doi.org/10.1186/s13073-019-0683-1>
- Lyra, P. C. M., Nepomuceno, T. C., De Souza, M. L. M., Machado, G. F., Veloso, M. F., Henriques, T. B., Dos Santos, D. Z., Ribeiro, I. G., Ribeiro, R. S., Rangel, L. B. A., Richardson, M., Iversen, E. S., Goldgar, D., Couch, F. J., Carvalho, M. A., & Monteiro, A. N. A. (2021). Integration of functional assay data results provides strong evidence for classification of hundreds of BRCA1 variants of uncertain significance. *GENETICS in MEDICINE*. <https://doi.org/10.1038/s41436>
- Online Mendelian Inheritance in Man® from the Johns Hopkins University. (2023, May 26). *OMIM Gene Map Statistics*. OMIM Morbid Map Scorecard (Updated May 26th, 2023). <https://www.omim.org/statistics/geneMap>
- Paton, N. W., Khan, S. A., Hayes, A., Moussouni, F., Brass, A., Eilbeck, K., Goble, C. A., Hubbard, S. J., & Oliver, S. G. (2000). Conceptual modelling of genomic information. In *BIOINFORMATICS* (Vol. 16, Issue 6).
- Preston, C. G., Wright, M. W., Madhavrao, R., Harrison, S. M., Goldstein, J. L., Luo, X., Wand, H., Wulf, B., Cheung, G., Mandell, M. E., Tong, H., Cheng, S., Iacocca, M. A., Pineda, A. L., Popejoy, A. B., Dalton, K., Zhen, J., Dwight, S. S., Babb, L., ... Costa, H. A. (2022). ClinGen Variant Curation Interface: a variant classification platform for the application of evidence criteria from ACMG/AMP guidelines. *Genome Medicine*, 14(1). <https://doi.org/10.1186/s13073-021-01004-8>
- PROS UPV. (2021). *Proyectos estratégicos en cooperación. Convocatoria 2021*.
- Ram, S., & Wei, W. (2004). *Modeling the Semantics of 3D Protein Structures* (pp. 696–708). https://doi.org/10.1007/978-3-540-30464-7_52
- Reyes Román, J. F. (2018). *DISEÑO Y DESARROLLO DE UN SISTEMA DE INFORMACIÓN GENÓMICA BASADO EN UN MODELO CONCEPTUAL HOLÍSTICO DEL GENOMA HUMANO* [Universitat Politècnica de València]. <https://doi.org/10.4995/Thesis/10251/99565>
- Reyes Román, J. F., Pastor, Ó., Casamayor, J. C., & Valverde, F. (2016). Applying conceptual modeling to better understand the human genome. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9974 LNCS, 404–412. https://doi.org/10.1007/978-3-319-46397-1_31
- Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), 405–424. <https://doi.org/10.1038/gim.2015.30>

Rodenburg, R. J. (2018). The functional genomics laboratory: functional validation of genetic variants. In *Journal of Inherited Metabolic Disease* (Vol. 41, Issue 3, pp. 297–307). Springer Netherlands. <https://doi.org/10.1007/s10545-018-0146-7>

Scheuermann, R. H., Ceusters, W., & Smith, B. (2009). Toward an ontological treatment of disease and diagnosis. *Summit on Translational Bioinformatics, 2009*, 116–120.

Sociedad Española de Cardiología. (n.d.). *Cardiopatías Familiares y Genética Cardiovascular*. La Sección. Retrieved May 29, 2023, from <https://secardiologia.es/cardiopatias-familiares/la-seccion>

Tavtigian, S. V., Greenblatt, M. S., Harrison, S. M., Nussbaum, R. L., Prabhu, S. A., Boucher, K. M., & Biesecker, L. G. (2018). Modeling the ACMG/AMP variant classification guidelines as a Bayesian classification framework. *Genetics in Medicine*, 20(9), 1054–1060. <https://doi.org/10.1038/gim.2017.210>

Wieringa, R. J. (2014). *Design Science Methodology for Information Systems and Software Engineering* (1st ed.). Springer Berlin, Heidelberg.

DOCUMENTO II. PRESUPUESTO

Diseño y Desarrollo de una Fuente de Datos sobre Estudios Funcionales asociados al Criterio PS3/BS3 de las Guías ACMG-AMP 2015 y Aplicación en Cardiopatías Familiares

Marina Martínez Miguel

Grado en Ingeniería Biomédica

Curso académico 2022/2023

1. OBJETIVOS DEL PRESUPUESTO

El objetivo del presupuesto es determinar el coste económico del conjunto de actividades que se han llevado a cabo en este Trabajo de Fin de Grado. En este caso, para determinar el coste total se han de considerar tres partes: el coste del personal, los costes de software utilizados y los costes de hardware empleados.

Para calcular el coste imputable al software y hardware empleados se aplicará la siguiente ecuación:

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

Siendo t el tiempo de uso del equipo (meses), C el coste del equipo o de la licencia del software (euros), y T el tiempo de amortización (meses).

Con esta base, en la siguiente sección se aporta el presupuesto desglosado en las tres partes mencionadas previamente.

2. PRESUPUESTO DESGLOSADO

Para desglosar el presupuesto se mostrará el resumen de costes para cada una de las tres partes que se aplican en este Trabajo de Fin de Grado: coste personal (Tabla 7), coste de software (Tabla 8) y coste de hardware (Tabla 9).

El coste de personal se determinará según las horas de trabajo invertidas por el Ingeniero asignado al trabajo, junto con las horas dedicadas por los tutores, puesto que no ha habido otro asesoramiento. Como coste unitario por hora de cada profesional se han tomado como referencia los criterios para elaborar presupuestos de I+D de la Universitat Politècnica de València¹⁴, que describe cómo calcular el presupuesto del personal en función de su posición. Así, el coste unitario para un Ingeniero Biomédico es 24,99 €/h. En cuanto a los tutores, se establecen, para un catedrático 51,40 €/h, y para un investigador predoctoral 24,99 €/h. Considerando estos valores y las horas de trabajo de cada individuo, se calculan los costes de personal tal y como se muestra en la Tabla 7.

¹⁴ <https://wiki.upv.es> › Guia_Presupuesto_2018

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

COSTES DE PERSONAL			
Personal	Coste unitario (€/h)	Horas Totales (h)	Costes Total (€)
Ingeniero Biomédico: Titulado Superior	24,99	300	7.497
Tutor: Investigador Predoctoral	51,40	80	4.112
Cotutor: Catedrático	31,45	45	1.415,25
Coste Total:			13.042,25

Tabla 7. Desglose de costes de personal asociados al trabajo realizado

En cuanto al software, en este trabajo se han empleado cuatro programas: el sistema operativo de Microsoft Windows versión 11 Pro, una licencia del paquete Microsoft Office para el Hogar y Estudiantes, varias bases de datos genómicas y biomédicas (*ClinVar*, *dbSNP*, *PubMed*, *ScienceDirect* y el buscador de ontologías *EMBL-EBI*), y el entorno JupyterLab, mediante el programa Anaconda, para programar en Python. Combinando los costes de licencias de cada uno de estos, y empleando la ecuación definida en la sección de objetivos de presupuesto, se han calculado los costes de software.

COSTES DE SOFTWARE					
Programa	Coste de Licencia (€)	Número de licencias	Periodo de uso (meses)	Duración de la Licencia	Coste imputable (€)
Sistema Operativo Microsoft Windows 11 Pro	145	1	8	Indefinida (Uso 3 años)	32,22
Microsoft Office Hogar y Estudiantes	149	1	8	Indefinida (Uso 3 años)	33,11
Bases de datos	0	Acceso Libre	8	-	0
Anaconda: Jupyter Lab	0	1	8	Indefinida (Uso 3 años)	0
Coste Total:					65,33

Tabla 8. Desglose de costes de software asociados al trabajo realizado.

Se puede ver en la Tabla 8 que el coste total de software asciende a 65,33 € para las actividades de este Trabajo de Fin de Grado.

Finalmente, los costes de hardware se han calculado de forma análoga a los costes de software. Por tanto, se ha empleado la misma ecuación aplicándola al único componente de hardware utilizado, el cual ha sido un ordenador portátil Dell Inspiron 14 Plus (7420), con un procesador Intel® Core™ i7-12700H (velocidad hasta 4,7 GHz, 14 núcleos), una memoria de 16 GB y capacidad de almacenamiento de 512 GB.

COSTES DE HARDWARE					
Programa	Coste unitario (€)	Unidades	Periodo de amortización (meses)	Periodo de uso (meses)	Coste imputable (€)
Ordenador portátil: DELL Inspiron 14 Plus 7420	908,26	1	36 (3 años)	8	201,84
Coste Total:					201,84

Tabla 9. Desglose de costes de hardware asociados al trabajo realizado.

Como se puede ver en la Tabla 9, se tiene un coste total de 201,84 € asociados al hardware utilizado a lo largo del trabajo. Con esto, en la siguiente sección se detallará el presupuesto total de este Trabajo de Fin de Grado.

3. PRESUPUESTO TOTAL

A partir de la información calculada en las secciones previas, en este apartado se muestra el presupuesto total requerido para llevar a cabo este Trabajo de Fin de Grado.

Con el Presupuesto por Ejecución Material – la suma de los costes de personal, software y hardware de la sección previa – se tiene la base a la que se añadirán los Gastos Generales y el Beneficio Industrial. Los primeros se corresponden al capital invertido en tareas sin beneficio directo para la empresa y se calcula como el 13% del Presupuesto por Ejecución Material. El segundo representa los beneficios que obtiene la empresa a partir del proyecto en cuestión, y supondrá una cantidad del 6% del Presupuesto por Ejecución Material.

Seguidamente, se calcula el Presupuesto de Ejecución por Contrata como la suma de los Gastos Generales y el Beneficio Industrial. Añadiendo a este el valor del IVA, del 21%, se obtiene el Presupuesto Total Neto del presente Trabajo de Fin de Grado como se resume en la Tabla 10.

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

PRESUPUESTO TOTAL	
Coste Total De Personal	13.042,25 €
Coste Total De Software	65,33 €
Coste Total De Hardware	201,84 €
Presupuesto De Ejecución Material	13.309,42 €
Gastos Generales (13%)	1.730,22 €
Beneficio Industrial (6%)	798,57 €
Presupuesto Por Ejecución Por Contrata	15.838,21 €
IVA	3.326,02 €
Presupuesto Neto Total	19.164,23 €

Tabla 10. Presupuestos finales de este Trabajo de Fin de Grado.

Como Presupuesto Total requerido para el desarrollo de este Trabajo de Fin de Grado, siguiendo las condiciones descritas en la sección del presupuesto desglosado, se obtiene una cantidad total de **19.164,23 €**.

DOCUMENTO III. ANEXOS

Diseño y Desarrollo de una Fuente de Datos sobre Estudios Funcionales asociados al Criterio PS3/BS3 de las Guías ACMG-AMP 2015 y Aplicación en Cardiopatías Familiares

Marina Martínez Miguel
Grado en Ingeniería Biomédica
Curso académico 2022/2023

ANEXO 1. FUENTE DE INFORMACIÓN DE ESTUDIOS FUNCIONALES ASOCIADA A CARDIOPATÍAS FAMILIARES.

ARTÍCULO	DOI	Disease	Gene	Genomic position	Variant in transcript (c.)	Variant in protein (p.)	type	model/system	experimentalMethod	outcome/evaluated	positiveControls	PosControl	negativeControls	negControl	Biological replicates	technical replicates	replicates	statisticalAnalysis	validationProcess	reproducible	robustnessData	Functional Impact		
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	F L N C	N A	c.7591-1_7591dup	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	Contrc	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, pathogenic, hetero	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c.11214-516>A	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, pathogenic, hetero	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	N A	p. G148R	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, pathogenic, hetero	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	N A	p. G148R	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Unknown, identified in three hypertrophic	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	V C L	N A	N A	p. K983fs	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, pathogenic, hetero	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	F H O D 3	N A	N A	p. T502fs	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, recessive, in fibro	
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	F H O D 3	N A	N A	c.1646+16>C	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, recessive, splice
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	F H O D 3	N A	N A	c.1646+16>T	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, recessive, splice
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	F H O D 3	N A	N A	c.1646+16>A	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, recessive, splice
Whole genome sequencing biobank study	10.1093/bioinformatics/btt102	Hypertrophic Cardiomyopathy	B R A F	N A	N A	Chr13:14163142136>A	N A	ECM0029	endonucleolytic site	In vitro	RNA sequencing (10 mycardial samples)	functional assessment of regulatory variant	3	For massivel parallel reporter assay (MP	RNA seq	RNA seq	3	RNA seq	Biological replicates	Frequency, ChIP-qPCR, confocal microscopy	Controls, data, replicates, postdocs	Yes	A bioarray control, 10000 genome-Polj	Functionally altered, associated with
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. T320M	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. Q428E	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. E850D	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. N1255S	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. A1648T	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. A1717G	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. R1880Q	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. R1973Q	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	
Role of CACNA1C in Brugada syndrome	10.1093/eurheartj/ehp102	Brugada's Syndrome	C A C N A 1 C	N A	N A	p. G2084E	OM00216	Electroporology assay	In vitro	Cellular	Current-voltage relationship	N A	N A	1	W T	2	6 cells	6 biological replicates (3 technical)	Amiga, ROC analysis, biographic/Simon	Biological replicates, negative control, post	Yes	Overall, the study cohort included 705 patients	Functionally altered, recessive, Shorter C	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Role of CACNA1C in long QT syndrome Pre	10.1093/eurheartj/eha063	Brugada Syndrome	CACNA1C	N	A	N	A	p. V1707I	DB000276	Electrophysiology assay	In vitro	Cellular	Current-voltage relationship	N	A	N	A	1	W	T	2	6 cells	Biological	Genetic	Yes	Overall the study cohort included 705 patients	Functionally abnormal	Molecular	Flow cytometry
Long QT syndrome type 3 and CACNA1C	10.1093/eurheartj/eha063	Long QT Syndrome	CACNA1C	N	A	N	A	p. P381S	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	2	2 cells	Biological	Genetic	Yes	The study cohort included 270 subjects	Functionally abnormal	FEKCTCD3/DMAC	
Long QT syndrome type 3 and CACNA1C	10.1093/eurheartj/eha063	Long QT Syndrome	CACNA1C	N	A	N	A	p. M456I	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	2	13 cells	Biological	Genetic	Yes	The study cohort included 270 subjects	Functionally abnormal	FEKCTCD3/BEVIG	
Long QT syndrome type 3 and CACNA1C	10.1093/eurheartj/eha063	Long QT Syndrome	CACNA1C	N	A	N	A	p. A582D	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	2	18 cells	Biological	Genetic	Yes	The study cohort included 270 subjects	Functionally abnormal	FEKCTCD3/BEVIG	
Long QT syndrome type 3 and CACNA1C	10.1093/eurheartj/eha063	Long QT Syndrome	CACNA1C	N	A	N	A	p. R858H	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	2	22 cells	Biological	Genetic	Yes	The study cohort included 270 subjects	Functionally abnormal	FEKCTCD3/BEVIG	
Long QT syndrome type 3 and CACNA1C	10.1093/eurheartj/eha063	Long QT Syndrome	CACNA1C	N	A	N	A	p. G1783C	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	2	14 cells	Biological	Genetic	Yes	The study cohort included 270 subjects	Functionally abnormal	FEKCTCD3/BEVIG	
Missense mutation in Desmoglein-2	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.E102K	EC000026	Genetic	In vitro	Whole	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	9	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Missense mutation in Desmoglein-2	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L347C	EC000026	Genetic	In vitro	Whole	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	9	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
The ABCG2/MDR1 gene mutation in the	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A	N	A	p.A831Tfs34	EC000026	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	3	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
A novel mutation in the	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.E102K	EC000026	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
A novel mutation in the	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L347C	EC000026	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Genetic mutation in the	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A	N	A	p.G780del	EC000026	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	PKP2	N	A			p.L378G>A	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	PKP2	N	A			p.L511-3A>G	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	PKP2	N	A			p.T146-1G>C	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	PKP2	N	A			p.T578-3A>G	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	PKP2	N	A			p.L493-9L493-11del	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L378-2T>G	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L783T>A	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L759T>G	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Functional assessment of potential splice sites	10.1093/eurheartj/eha063	Arrhythmogenic Right Ventricular Cardiomyopathy	DSG2	N	A			p.L350A>G	EC000594	In vitro	In vitro	HL-1 cells	Interaction with plakophilin-2 and plakoglobin	N	A	N	A	1	W	T	2	2	Biological	Genetic	Yes	All patients	Functionally abnormal	Adaptive	Western blot
Molecular autopsy and subsequent function	10.1093/eurheartj/eha063	Sudden cardiac death, possible arrhythmogenic	KCNQ1	N	A			p.L449C>A	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	3	5 WT c	Biological	Genetic	Yes	The study was approved by the Ethical Committee	Functionally abnormal	Adaptive	Western blot
Molecular autopsy and subsequent function	10.1093/eurheartj/eha063	Sudden cardiac death, possible arrhythmogenic	DSG2	N	A			p.L993H	DB000276	Whole-cell patch clamp assay	In vitro	Whole	Depolarization currents, inactivation and	N	A	N	A	1	W	T	3	5 WT c	Biological	Genetic	Yes	The study was approved by the Ethical Committee	Functionally abnormal	Adaptive	Western blot

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Common polymorphism of the Des	https://doi.org/10.1051/hvnt/201536	Sublet/Caricac/Leah/posible/Arhythm	D S G 2	N A	c. *1113_ *1115-	N A	ECO:00	Computational structure	In silico	Genotype-phenotype analysis was perform	Genotype-phenotype analysis was perform	N A	Control samples	2	WT	gpl3/conv	4	3 6	Biologi	Chaperone was coedited using PolyP	Stats, neg control	Yes	At least 12 Sublet/Caricac/Leah/posible	Functionally abnormal. Protein analysis
Deaf myopathy associated with myofibrillar	https://doi.org/10.1093/cvr/cvz017	Arhythmogenic Right Ventricular Cardiom	D S G 2	N A	c. 710T>A	p. L237Ter	ECO:00	In vitro	Pathogenicity analysis of the novel variant	Expression of DSG2, Location, Zebrafish	N A	N A	1	WT	3	AC16	model n=3	Biologi	Data generated using a custom pipeline	Repeat in another cohort	Yes	The ABC pathway was recruited in Drog	Functionally abnormal. The functional analysis	
Mutations in Flamin1 Cause a new form	https://doi.org/10.1093/cvr/cvz017	Hypertrophic Cardiomyopathy	F L N C	N A	N A	p. V123A	ECO:00	In vitro	HS-Cat cardiac myoblasts, neuronal data	Formation of protein aggregates	N A	Healthy heart samples	1	WT	4	N A	Biologi	N A	Comparison with other functional studies	Yes	All participants had provided written inform	Functionally abnormal. Significant functional		
Mutations in Flamin1 Cause a new form	https://doi.org/10.1093/cvr/cvz017	Hypertrophic Cardiomyopathy	F L N C	N A	N A	p. A1539T	ECO:00	In vitro	HS-Cat cardiac myoblasts, neuronal data	Formation of protein aggregates	N A	Healthy heart samples	1	WT	4	N A	Biologi	N A	Comparison with other functional studies	Yes	All participants had provided written inform	Functionally abnormal. Mutation of Flamin		
Mutations in Flamin1 Cause a new form	https://doi.org/10.1093/cvr/cvz017	Hypertrophic Cardiomyopathy	F L N C	N A	N A	p. R2133H	ECO:00	In vitro	HS-Cat cardiac myoblasts, neuronal data	Formation of protein aggregates	N A	Healthy heart samples	1	WT	4	N A	Biologi	N A	Comparison with other functional studies	Yes	All participants had provided written inform	Functionally abnormal. Mutation of Flamin		
Mutations in Flamin1 Cause a new form	https://doi.org/10.1093/cvr/cvz017	Hypertrophic Cardiomyopathy	F L N C	N A	N A	p. A2430V	ECO:00	In vitro	HS-Cat cardiac myoblasts, neuronal data	Formation of protein aggregates	N A	Healthy heart samples	1	WT	4	N A	Biologi	N A	Comparison with other functional studies	Yes	All participants had provided written inform	Functionally abnormal. Mutation of Flamin		
Functional alteration of myosin heavy chain	https://doi.org/10.1051/hvnt/201536	Dilated Cardiomyopathy	F L N C	N A	N A	N A	ECO:00	In vitro	Asiatic Myosin Heavy Chain 2	Protein structure analysis	N A	N A	N A	N A	2	N A	Biologi	Protein structure analysis	N A	N A	N o	N A	Functionally abnormal. The mutation affect	
PKC δ Gene Silencing Mutations Cause Dilate	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	F L N C	N A	N A	p. 7151016>A	N A	DB:000167	See knockdown assay	In vivo	Zebrafish	Workup of structures of zebrafish hearts	2	Unspliced zebrafish, c53MO	1	Scrambled c53-MO	4	2 4 1	Biologi	Immunoprecipitation and Western blot	Yes	Male and female with zebrafish (Zebrafish)	Functionally abnormal. Gene silencing	
PKC δ Gene Silencing Mutations Cause Dilate	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	F L N C	N A	N A	p. 5569-1616	N A	DB:000167	See knockdown assay	In vivo	Zebrafish	Workup of structures of zebrafish hearts	2	Unspliced zebrafish, c53MO	1	Scrambled c53-MO	4	1 6 4	Biologi	Immunoprecipitation and Western blot	Yes	Male and female with zebrafish (Zebrafish)	Functionally abnormal. Gene silencing	
Functional Validation of a Pathogenic Muta	https://doi.org/10.1051/hvnt/201536	Flamin1 Cardiomyopathy	F L N C	N A	N A	p. G2151S	N A	DB:000167	See knockdown assay	In vivo (animal model)	Zebrafish model (Danio rerio), microinject	Embryonic lethality, survival after birth, and	N A	N A	1	Control oligomer	3	5	SNAP25C Control oligomer (p=0.01)	p < 0.05	Controls	N o	N A	Functionally abnormal. Decrease survival
GM4 Gene Function Mutation Variants	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	G A T A	N A	N A	p. C271S	N A	ECO:00005	Surface expression gene	In vitro	The re	Activation of AMP-activated protein kinase	N A	N A	1	WT	2	3	Biologi	Data are expressed as mean \pm SD. Contin	Alignment of 244 protein sequences	Yes	Peptide mass spectrometry	Functionally abnormal. Activated AMP-prot
Prevalence and spectrum of GM4 mutat	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	G A T A	N A	N A	p. V39L	N A	ECO:00005	Loss-of-function mutation	In vitro	Hela c	Total protein accumulation	N A	N A	1	WT	2	3	Biologi	Western blot, Substrate competition test, ch	State- prediction Mutation Taster and Poly	Yes	Peptide mass spectrometry	Functionally abnormal. Activated AMP-prot
Prevalence and spectrum of GM4 mutat	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	G A T A	N A	N A	p. P226Q	N A	ECO:00005	Loss-of-function mutation	In vitro	Hela c	Total protein accumulation	N A	N A	1	WT	2	3	Biologi	Western blot, Substrate competition test, ch	State- prediction Mutation Taster and Poly	Yes	Peptide mass spectrometry	Functionally abnormal. Activated AMP-prot
Prevalence and spectrum of GM4 mutat	https://doi.org/10.1093/cvr/cvz017	Dilated Cardiomyopathy	G A T A	N A	N A	p. T279S	N A	ECO:00005	Loss-of-function mutation	In vitro	Hela c	Total protein accumulation	N A	N A	1	WT	2	3	Biologi	Western blot, Substrate competition test, ch	State- prediction Mutation Taster and Poly	Yes	Peptide mass spectrometry	Functionally abnormal. Activated AMP-prot
A novel GM4 Gene Function Mutation	https://doi.org/10.1051/hvnt/201536	Dilated Cardiomyopathy	G A T A	N A	N A	p. 8716>C	p. V291L	ECO:00005	Loss-of-function mutation	In vitro	For ge	Western blotting, protein structure	N A	N A	1	WT	2	3	Biologi	Data are expressed as mean \pm SD. Control	Stats, variant was validated by resequenc	Yes	The clinical studies were performed with in	Functionally abnormal. Absent in 400 cont
Characterization of Sublet/Caricac/Leah/pos	https://doi.org/10.1051/hvnt/201536	Long QT Syndrome	K C N Q 1	N A	N A	p. T587M	N A	DB:000126	Electrophysiology assay	In vitro	COS7 c	COS7 c	N A	N A	1	WT	2	6	Biologi	Western blotting, electrophysiology	Stats, control	Yes	Genetic data was obtained from biobank	Functionally abnormal. Resolved from tree
Unexplained cardiac arrest: a role of con	https://doi.org/10.1093/cvr/cvz017	Unexp	P K P 2	N A	N A	p. 2431C>A	p. R811S	DB:000126	Electrophysiology assay	In vitro	Site di	Molecular biology and expression of KCNQ	N A	N A	1	WT	6	3	Biologi	Western blotting, electrophysiology	Prediction: PolyPhen2, Mutation Taster, S	Yes	This study was conducted in full agreement	Functionally abnormal. B1T (Ule) being a
Unexplained cardiac arrest: a role of con	https://doi.org/10.1093/cvr/cvz017	Unexp	K C N Q 1	N A	N A	c. 191_210 del	p. 784>P434A*	DB:000126	Electrophysiology assay	In vitro	Site di	Molecular biology and expression of KCNQ	N A	N A	1	WT	6	3	Biologi	Western blotting, electrophysiology	Prediction: PolyPhen2, Mutation Taster, S	Yes	This study was conducted in full agreement	Functionally abnormal. Pathogenicity rep
Using the genome aggregation database, a	https://doi.org/10.1093/cvr/cvz017	Long QT Syndrome 1	K C N Q 1	N A	N A	p. F127L	ECO:00	In silico	In silico	Expression, I/Os current	N A	N A	1	WT	2	1 4	Biologi	Substrate competition test, Western blot	Controls, stats, prediction tools	Yes	All of the "pathogenic" KCNQ cases were	Functionally normal. None of the KCNQ ca		
Using the genome aggregation database, a	https://doi.org/10.1093/cvr/cvz017	Long QT Syndrome 2	K C N Q 1	N A	N A	p. P477L	ECO:00	In silico	In silico	Expression, I/Os current	N A	N A	1	WT	2	1 2	Biologi	Substrate competition test, Western blot	Controls, stats, prediction tools	Yes	All of the "pathogenic" KCNQ cases were	Functionally normal. None of the KCNQ ca		
Using the genome aggregation database, a	https://doi.org/10.1093/cvr/cvz017	Long QT Syndrome 3	K C N Q 1	N A	N A	p. L619M	ECO:00	In silico	In silico	Expression, I/Os current	N A	N A	1	WT	2	1 1	Biologi	Substrate competition test, Western blot	Controls, stats, prediction tools	Yes	All of the "pathogenic" KCNQ cases were	Functionally normal. None of the KCNQ ca		
Molecular characterization of a novel unde	https://doi.org/10.1093/cvr/cvz017	Long QT Syndrome	K C N Q 1	N A	N A	p. N1524>C (KCNQ1-F6)	N A	DB:000126	Electrophysiology assay	In vitro	550 pr	RT-PCR	3	FF433/ha2 control/NA samples	1	WT	6	N A	Biologi	At least was confirmed by Sanger	Controls, stats	Yes	This study	Functionally abnormal. KCNQ1 V324G
Molecular characterization of a novel unde	https://doi.org/10.1093/cvr/cvz017	Long QT Syndrome	K C N H 2	N A	N A	p. D170N (H316-F6)	N A	DB:000126	Electrophysiology assay	In vitro	551 pr	RT-PCR	3	FF433/ha2 control/NA samples	1	WT	3	N A	Biologi	At least was confirmed by Sanger	Controls, stats	Yes	This study	Functionally abnormal. KCNQ1 H316N

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Anthropometric Cardiomorphologic criteria	10.1016/j.ymgme.2014.11.013	Vertical Tachycardia Syndrome of Dilé	K C N Q 1	N A N A	p.R397Q	BM000276	Electropathology assay	In vitro	10 pat	Interaction between experimental groups	1	Critical fluorescent microscopy	1	W T 2	20 cells	Biologi	Values were expressed as mean standard error	Protein degradation, sites, post controls	Y e s	The institutional review board of the Span	Functionally abnormal. Affected a deep	
Functional assessment of compound heterozygosity	10.1016/j.ymgme.2015.01.005	Long QT Syndrome	K C N Q 1	N A N A	p. R 5 9 1 H	BM000276	Electropathology assay	In vitro	Case detection and functional studies	Tend to be expressed in the same tissue	1	Control group	1	W T 5	WT n=	Biologi	Data analysis and drawings were performed	Values from biochemical, sites, post controls	Y e s	100 patients were selected from patients	Functionally abnormal. Excluded a severe	
Functional assessment of compound heterozygosity	10.1016/j.ymgme.2015.01.005	Long QT Syndrome	K C N H 2	N A N A	p. R 3 2 8 C	BM000276	Electropathology assay	In vitro	Case detection and functional studies	Tend to be expressed in the same tissue	1	W T 5	HERG1	HERG: n=15	Biologi	Data analysis and drawings were performed	Values from biochemical, sites, post controls	Y e s	100 patients were selected from patients	Functionally normal. Mutation of not known		
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c. 4719 C>T	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Short truncating pr
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c. 10910 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Extension of intron 1
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c. 4299 G>C	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally normal. Intron deletion
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c. 4299+6T>C	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally normal. Functionally normal
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N H 2	N A	c. 11228 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Missense change
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N H 2	N A	c. 2145 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Leu to Ile change
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N H 2	N A	c. 2391+9 G>T	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Stop skipping gene
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c.683+5G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Complex change
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c. 1032+9 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Multiple position
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c. 1032 G>T	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. 3 reads in control
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c.477+5G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Near complete de
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c. 1032 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Affects splicing in
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Long QT Syndrome	K C N Q 1	N A	c.386+6T>G	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. Not compatible with
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c.393-5C>T	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal? Periodic short
Functional Assay: Reassembly Splicing Site	https://doi.org/10.1038/ng.2824	Brugada Syndrome	SCN5A	N A	c. 4437+9 G>A	EO000594	Splicing modification (reporter)	In vitro	Minigene	Splicing outcomes (RT-PCR)	2	Contrc	2	WTNo	2	3	Biologi	Statistical analyses were performed with a	Sites, regional, post controls, Splicing site	Y e s	Putativ	Functionally abnormal. RT-PCR and NTRN
Anti-Flag detection (ODQ) mutation T5	10.1016/j.ymgme.2015.01.011	Long QT Syndrome	K C N Q 1	N A	p. R 2 5 9 C	BM000270	Whole cell patch clamp assay	In vitro	HEK293 cells transfected with cDNA for WT	Expression of hERG and effect of ODQ on	2	GFP (g 1	W T 2	3	1	Biologi	All data were expressed as mean ± SEM. All	Post controls, regional, sites	Y e s	The complementary domain (nucleic acid)	Functionally normal. No significant effect	
Anti-Flag detection (ODQ) mutation T5	10.1016/j.ymgme.2015.01.011	Long QT Syndrome	K C N Q 1	N A	p.T587M	BM000270	Whole cell patch clamp assay	In vitro	HEK293 cells transfected with cDNA for WT	Expression of hERG and effect of ODQ on	2	GFP (g 1	W T 2	3	1	Biologi	All data were expressed as mean ± SEM. All	Post controls, regional, sites	Y e s	The complementary domain (nucleic acid)	Functionally abnormal. When co-express	
A new ODQ mutation at the S5 segment	10.1016/j.ymgme.2015.01.011	Short QT Syndrome	K C N Q 1	N A	p.F279I	BM000270	Whole cell patch clamp assay	In vitro	ODQ and FRET assays were conducted in WT	Correlation between electrophysiological	1	6x279M (MUT) co-localization (MUT)	2	2 WT & 6x279M (MUT) co-localization	4	COS7:	Data were expressed as mean ± SEM. All	Competition studies, sites	Y e s	The institutional review board at the Inst	Functionally abnormal. Affects hERG WT	
Missense mutation variants aggregate	10.1016/j.ymgme.2015.01.011	Dilated LMNA	N A	N A	14 variants	EO000593	Aggregation assay	In vitro	Functionally	AGGR	1	6	In silico	2	≥ 3	Biologi	All aggregation data were expressed as mean ± SEM. All	Controls	Y e s	All missense variants were made in single	Functionally abnormal. All of the tested	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. R190W	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. A246V	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.R249Q	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. D257Y	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.E358K	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. R377H	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.L435P	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. T453I	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.R482L	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.E527K	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. R527C	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p.L530P	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. D540V	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Most pathogenic variants aggregate	1139841435-01-0025-1	Dilatec	LMNA	N	A	N	A	p. R541C	EO000691	Edgpath zap videre	In vitro	Función	AGGR	1	6	15 pathogenic controls (Labs) in OIra	1	6	In silico	2	≥	3	Biologi	Aggregation data presented in	Stats, controls	Yes	All disease variants were not in silico	Functionally dominant, strong aggregation
Clinical genetic outcome of the event	1139316767-01-0465-1	LVNC	RBM20	N	A	N	A	p.R634L	940340_V000001	NA Edgpath zap	In vitro	RNAi	Knockdown	77k, p.63	Splicing activity, titration protein weight control	1	1: P 6 4 1 L	2	WT	Lef 2	PSI	ass	Biologi	Package version 1.1.2 and GspPath	Predictions with PROVEAN, SIFT and PolyPhen	Yes	Pathogenicity in vitro confirmed in cancer	Functionally dominant, likely not pathogenic
Clinical genetic outcome of the event	1139316767-01-0465-1	LVNC	RBM21	N	A	N	A	p.P798L	940340_V000001	NA Edgpath zap	In vitro	RNAi	Knockdown	77k, p.63	Splicing activity, titration protein weight control	1	1: P 6 4 1 L	2	WT	Lef 2	PSI	ass	Biologi	Package version 1.1.2 and GspPath	Predictions with PROVEAN, SIFT and PolyPhen	Yes	Pathogenicity in vitro confirmed in cancer	Functionally dominant, did not show disease
Clinical genetic outcome of the event	1139316767-01-0465-1	LVNC	RBM22	N	A	N	A	p. P1105L	940340_V000001	NA Edgpath zap	In vitro	RNAi	Knockdown	77k, p.63	Splicing activity, titration protein weight control	1	1: P 6 4 1 L	2	WT	Lef 2	PSI	ass	Biologi	Package version 1.1.2 and GspPath	Predictions with PROVEAN, SIFT and PolyPhen	Yes	Pathogenicity in vitro confirmed in cancer	Functionally dominant, did not show disease
Functional study in a model system	1131111111-01-1571-1	Long QT Syndrome 2	KCNH2	N	A	N	A	c.3017delG	08100270	Whole-exon path zap	In vitro	Homo	KCNH2	1	Control cells in solution (path zap)	1	W T	2	HEK29	Biologi	GoPath Path 2.0, GspPath	Stats, neg. control, neg. control	Yes	Whole-exon confirmed via whole-exon	Functionally dominant, strong aggregation			
Utilization of the gene: aggregation data	1139516767-01-0465-1	Long QT Syndrome 1	KCNH2	N	A	N	A	p.196V	08100270	Whole-exon path zap	In silico	30 LogCS	Syndrome associated with	In silico	1	NA	loading control	GAPDH	1	W T	2	9	Biologi	Statist	Stats, neg. control, n=10	Yes	Prediction analysis with pathogenicity	Functionally dominant, statistically significant
Utilization of the gene: aggregation data	1139516767-01-0465-1	Long QT Syndrome 1	KCNH2	N	A	N	A	p. P241L	OBI:OC	In silico	30 LogCS	Syndrome associated with	In silico	1	NA	loading control	GAPDH	1	W T	2	1 2	Biologi	Statist	Stats, neg. control, n=12	Yes	Prediction analysis with pathogenicity	Functionally dominant, All Web help	
Utilization of the gene: aggregation data	1139516767-01-0465-1	Long QT Syndrome 1	KCNH2	N	A	N	A	p.H254Q	OBI:OC	In silico	30 LogCS	Syndrome associated with	In silico	1	NA	loading control	GAPDH	1	W T	2	1 4	Biologi	Statist	Stats, neg. control, n=14	Yes	Prediction analysis with pathogenicity	Functionally dominant, strong aggregation	
Utilization of the gene: aggregation data	1139516767-01-0465-1	Long QT Syndrome 1	KCNH2	N	A	N	A	p. G314S	OBI:OC	In silico	30 LogCS	Syndrome associated with	In silico	1	NA	loading control	GAPDH	1	W T	2	1 1	Biologi	Statist	Stats, neg. control, n=11	Yes	Prediction analysis with pathogenicity	Functionally dominant, statistically significant	
Utilization of the gene: aggregation data	1139516767-01-0465-1	Long QT Syndrome 1	KCNH2	N	A	N	A	p.P935S	OBI:OC	In silico	30 LogCS	Syndrome associated with	In silico	1	NA	loading control	GAPDH	1	W T	2	1 2	Biologi	Statist	Stats, neg. control, n=12	Yes	Prediction analysis with pathogenicity	Functionally dominant, statistically significant	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Utilization of the genome aggregation data	10.1093/hmg/ddz004	Long QT Syndrome 2	K C N H 2	N A N A	p . P 9 6 3 T	OBI:OC	In silico	34 Long QT Syndrome associated WNVs	In silico	1	WT	control	GAP2	1	W T 2	1 1	p . P 9 6 3 T - n = 11	Suberf test was performed in	Stats, seg control, in silico	Yes	Prediction analysis with pathogenic variant	Functionally normal. All Web pages		
Utilization of the genome aggregation data	10.1093/hmg/ddz004	Long QT Syndrome 2	K C N H 2	N A N A	p . G 1 8 7 S	OBI:OC	In silico	35 Long QT Syndrome associated WNVs	In silico	1	WT	control	GAP2	1	W T 2	1 2	p . G 1 8 7 S - n = 11	Suberf test was performed in	Stats, seg control, in silico	Yes	Prediction analysis with pathogenic variant	Functionally normal. All Web pages		
Utilization of the genome aggregation data	10.1093/hmg/ddz004	Long QT Syndrome 2	K C N H 2	N A N A	p . A 2 0 3 T	OBI:OC	In silico	34 Long QT Syndrome associated WNVs	In silico	1	WT	control	GAP2	1	W T 2	1 1	p . A 2 0 3 T - n = 11	Suberf test was performed in	Stats, seg control, in silico	Yes	Prediction analysis with pathogenic variant	Functionally normal. Abnormal at course		
Functional analysis of CYP2E1 gene	10.1093/hmg/ddz004	Long QT Syndrome 2	K C N H 2	N A N A	p . E 6 3 7 K	OBI:OC	In vivo	Cardiac arrest with ventricular tachycardia	Ability of E637K mutant to rescue	N A	N A	N A	2	WTFiv	5	3 time	Biological	Statistical analysis was performed in	Stats, pos control, neg control	Yes	This research was approved by the	Functionally normal. In the Web page		
Mutations in ERM1 and ERM2 genes	https://doi.org/10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	N E X N	N A	c . 3 9 1 C > G	p . Q 1 3 1 E	940340_000129	Immunoassay	In vitro	Both v	Location both variants present in	N A	N A	2	WT	Common population assays	3	N A	Biological wild type variants	N A	Neg controls, in silico prediction	N o	N A	Functionally abnormal. Misense, in silico
Mutations in ERM1 and ERM2 genes	https://doi.org/10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	N E X N	N A	c . 8 3 5 C > T	p . R 2 7 9 C	940340_000129	Immunoassay	In vitro	Both v	Location both variants present in	N A	N A	2	WT	Common population assays	3	N A	Biological wild type variants	N A	Neg controls, in silico prediction	N o	N A	Functionally abnormal. Misense, in silico
Functional Studies and in Silico Analysis	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 5 0 6 - 2 4 > C	N A	EO000594	Splice site mutation study	In vitro	We found 25 different mutations of which	Predicted role of variant and their effect on	N A	N A	1	W T 3	3	N A	Biological wild type variants	N A	In silico prediction	Yes	Genetic DNA was obtained from peripheral	Functionally abnormal. Loss of function	
Functional Studies and in Silico Analysis	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 9 0 6 - 1 6 > T	N A	EO000594	Splice site mutation study	In vitro	We found 25 different mutations of which	Predicted role of variant and their effect on	N A	N A	1	W T 3	3	N A	Biological wild type variants	N A	In silico prediction	Yes	Genetic DNA was obtained from peripheral	Functionally normal. No deleterious effect	
Functional Studies and in Silico Analysis	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 3 0 8 + 3 6 > C	N A	EO000594	Splice site mutation study	In vitro	We found 25 different mutations of which	Predicted role of variant and their effect on	N A	N A	1	W T 3	3	N A	Biological wild type variants	N A	In silico prediction	Yes	Genetic DNA was obtained from peripheral	Functionally abnormal. WT generated from	
Functional Studies and in Silico Analysis	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 3 9 3 - 5 C > A	N A	EO000594	Splice site mutation study	In vitro	We found 25 different mutations of which	Predicted role of variant and their effect on	N A	N A	1	W T 3	3	N A	Biological wild type variants	N A	In silico prediction	Yes	Genetic DNA was obtained from peripheral	Functionally abnormal. Deletion of base	
Functional Studies and in Silico Analysis	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 6 1 7 - 7 T > C	N A	EO000594	Splice site mutation study	In vitro	We found 25 different mutations of which	Predicted role of variant and their effect on	N A	N A	1	W T 3	3	N A	Biological wild type variants	N A	In silico prediction	Yes	Genetic DNA was obtained from peripheral	Functionally normal. No deleterious effect	
Protein and Functional Studies Revealed	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 2 3 7 3 i n s G	N A	EO000019	amino acid mutation study	In vitro	A protein screen was performed in	Protein	N A	N A	1	W T 2	2	W T 7, W T 8	Biological	Recombinant antibodies, pGL3, quantiferon	Stats, seg control, pos control	Yes	Tissue of the interventricular septum (V1)	Functionally abnormal. This mutation intro	
Protein and Functional Studies Revealed	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 7 7 2 G > A	N A	EO000019	amino acid mutation study	In vitro	A protein screen was performed in	Protein	N A	N A	1	W T 2	2	W T 6, W T 7	Biological	Recombinant antibodies, pGL3, quantiferon	Stats, seg control, pos control	Yes	Tissue of the interventricular septum (V1)	Functionally abnormal. Homozygous NBP	
Protein and Functional Studies Revealed	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 8 0 9 T > G	p . I 6 0 3 M	940340_000129	RNA splice assay	In vitro	BADGE1 calibration expressed in vitro	educated after 2005-6 to 2006-7	1	Non-carrier control	1	W T 2	2	N A	Biological wild type variants	R M S D	Neg control, pos control	Yes	Homologous protein genetic analysis	Functionally abnormal. Results show that f	
A feilin	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y H 7	N A	c . 5 6 4 7 G > A	p . E 1 8 3 K	EO000019	amino acid mutation study	In vitro	Immunoblotting analysis was performed	Morphology of cat ventricles with	N A	N A	1	W T 2	2	N A	Comparisons of homozygous wild type	N A	Neg control, in silico prediction	Yes	Cat sar	Functionally abnormal. Effects the canal	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 7 7 2 G > A	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	N A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 9 1 7 - 9 G > A	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 6 1 4 + 4 4 T	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 9 1 8 - 1 4 > G	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 3 0 9 + 1 6 > A	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 2 3 7 3 i n s G	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This variant is in	
WBC Strain mutation enhancement	10.1093/hmg/ddz004	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c . 1 9 0 5 + 1 6 > A	N A	EO000019	amino acid mutation study	In vitro	Myocardial tissue from patients with hyper	• Fluorescence	M A	N A	1	W T 2	2	4	Biological	Base sequenced in mean 450M. The data	Neg control, pos control	Yes	Myocardial tissue was obtained from the in	Functionally abnormal. This is a splice site	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3100delG	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This variant is pathogenic	
WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3193insC	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This variant is pathogenic	
WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3228_3227insT	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This variant is pathogenic	
WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3331+96>C	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This is a pathogenic variant	
WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3331-16>C	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This is a pathogenic variant	
WPC3 truncation mutation	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.3372C>A	N A	EO00019: amino acid system study	In vitro	Myocardial tissue from patients with hyperlipidemia	• Fluorescence microscopy	M A	N A	1	W T	2	4	Biological	Database: Ensembl, ESM, The data	Weg controls, pos control	Yes	Myocardial tissue was obtained from the in vitro study	Functionally abnormal. This variant results in a truncated protein	
Increase in plectin C&D sensitivity	https://doi.org/10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	N A	p.E264K	DB00019: gene knock in	In vitro	Myocardial tissue from patients with hyperlipidemia	Immunoblot analysis	The functional analysis	N A	N A	1	W T	3	n = 4 - 4	Biological	Database: Ensembl, ESM, CompBio, Controls, stats	Yes	The investigation confirms the guide for the variant	Functionally abnormal. A truncated protein	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.916-136>A	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally abnormal. (Pathogenic)	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.916-136>C	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally normal. No RNA effect	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.916-16>C	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally abnormal. (Pathogenic)	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.918+136>A	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally normal. No RNA effect	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.1214-136>A	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally abnormal. (Pathogenic)	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.1214-14>G	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally abnormal. (Pathogenic)	
Missense of the cardiac myosin binding	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.1216+49C>T	N A	EO00052: Transient mutation	In silico	Genotype	Transcription consequences, phylogenetics	N A	N A	N A	N A	N A	N A	N A	N A	In silico prediction (Splice-Sensitivity, RefSeq)	No	Two hundred and fifty patients diagnosed with hyperlipidemia	Functionally normal. Known polymorphism	
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.472G>A	p.V158M	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally abnormal. Individually variant
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.223G>A	p.D75N	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally normal. (Highly pathogenic)
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.446C>T	p.A149V	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Unknown. Predicted (VUS according to ACMG)
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.530G>A	p.R177H	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally normal. Predicted (Benign)
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.649A>G	p.S217G	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally normal. (Highly pathogenic)
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.706A>G	p.S236G	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally normal. Predicted (Benign)
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.772G>A	p.E258K	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally abnormal. Previously reported
Association of variants in MYH7, MYBPC3	10.1016/j.jmcp.2018.11.001	Hyperlipidemia	M Y B P C 3	N A	c.1564G>A	p.A1511H	EO00094: Computational structure	In silico	DM-Extraction and analysis, non-targeted	In silico prediction tools: PROVEAN, SIFT, PolyPhen	N A	N A	N A	N A	N A	N A	N A	N A	Statistical analysis was performed using SPSS	Stats	Yes	The study protocol was approved by the Ethical Committee	Functionally normal. Predicted (Likely Benign)

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y B P C 3	N A	c. 23756G>A	p. W 7 9 2 X	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Pathology reported		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y H 7	N A	c. 7466G>A	p. R 2 4 9 Q	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally normal. Pathological likely benign		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y H 7	N A	c. 7887T>C	p. I 2 6 3 T	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Associated in the Eth		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y H 7	N A	c. 9586G>A	p. V 3 2 0 M	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Described in the Eth		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y H 7	N A	c. 1357C>T	p. A r g 4 5 3 C y s	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Described in the Eth		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	M Y H 7	N A	c. 4954G>T	p. D 1 6 5 2 Y	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Described in the Eth		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	T N N T 2	N A	c. *1136G>A	N A	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally normal. Pathological likely benign		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	T N N T 2	N A	c. 304C>T	p. R 1 0 2 W	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Likely pathogenic in		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	T N N T 2	N A	c. 788A>G	p. K 2 6 3 R	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Functionally abnormal. Associated with the		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	T N N T 2	N A	c. 815A>G	p. N 2 7 2 S	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Unknown. Predicted VUS according to ACMG		
Association of variants in MYO10 [MPP3]a	10.1016/j.ymgme.2021.100474	Hypertrophic Cardiomyopathy	T N N T 2	N A	c. 890G>A	p. W 2 9 7 X	ECO00694 Computational structure	In silico	DN-Extraction and analysis, non-targeted	In silico predictions: PROVEAN, SIFT, P	N A	N A	N A	N A	N A	N A	Sees prediction tools: polyphen2, vesin	N A	Sees prediction tools: polyphen2, vesin	Statistical analysis was performed using PP	Stats	Yes	The study protocol was approved by the Eth	Unknown. Predicted VUS according to ACMG		
Novel mutation in sarcomeric protein gene	https://doi.org/10.1016/j.ymgme.2021.100474	Dilated Cardiomyopathy	M Y H 7	N A	N A	p. A 2 2 3 T	ECO00694 Computational structure	In silico	After selecting 45 unrelated patients, meta	Type of mutation: p.S144GfsTer10	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	Yes	This study included 45 unrelated patients	Functionally abnormal. The exchange point
Novel mutation in sarcomeric protein gene	https://doi.org/10.1016/j.ymgme.2021.100474	Dilated Cardiomyopathy	M Y H 7	N A	N A	p. S 6 4 2 L	ECO00694 Computational structure	In silico	After selecting 45 unrelated patients, meta	Type of mutation: p.S144GfsTer10	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	Yes	This study included 45 unrelated patients	Functionally abnormal. This variation is not
Novel mutation in sarcomeric protein gene	https://doi.org/10.1016/j.ymgme.2021.100474	Dilated Cardiomyopathy	M Y B P C 3	N A	N A	p. N 9 4 8 T	ECO00694 Computational structure	In silico	After selecting 45 unrelated patients, meta	Type of mutation: p.S144GfsTer10	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	N A	Yes	This study included 45 unrelated patients	Unknown. This mutation in MPP3 could
Autosomal recessive cardiomyopathy	10.1016/j.ymgme.2021.100474	Hypertrophic	M Y B P C 3	N A	c. 170C>A	p. A 5 7 D	ECO000093 amino acid modification study	In vivo	Morph	Heart functionality, ability to resuscitate	2	Control DNA for exon 11 deletion in mice	1	W T 3			ML2 zebrafish modeling (M1+10) ML0	Biological	Database: zebrafish, mouse, ERM, Statist	In silico predictors	Yes	Clinical	Functionally abnormal. Homozygous recessive			
Autosomal recessive cardiomyopathy	10.1016/j.ymgme.2021.100474	Hypertrophic	M Y B P C 3	N A	c. 106G>T	p. E 3 6 T e r	ECO000093 amino acid modification study	In vivo	Morph	Heart functionality, ability to resuscitate	2	Control DNA for exon 11 deletion in mice	1	W T 3			ML2 zebrafish modeling (M1+10) ML0	Biological	Database: zebrafish, mouse, ERM, Statist	In silico predictors	Yes	Clinical	Functionally abnormal. Homozygous recessive			
Autosomal recessive cardiomyopathy	10.1016/j.ymgme.2021.100474	Hypertrophic	M Y B P C 3	N A	c. 412-16>A	N A	ECO:001001	In vivo	Morph	Heart functionality, ability to resuscitate, Mitochondrial localization of the MLC2 protein	2	Control DNA for exon 11 deletion in mice	1	W T 3			N A	Biological	Database: zebrafish, mouse, ERM, Statist	In silico predictors	Yes	Clinical	Functionally abnormal. Presumably homozygous			
Mutation of Presenilin Gene in Dilated C	10.1016/j.ymgme.2021.100474	Dilated Cardiomyopathy	P S E N 1	N A	c. 998A>G	p. D 3 3 3 G	940344_000023 Protein expression analysis	In vitro	Tissue studies, fibroblast culture studies, p	Measurement of [Ca ²⁺] _i response to ligand	3	To identify mutations: DNA samples, exons	N A	N A	N A	4	control patient (3 patients, 45 cells) control	Biological	Database: ERM, qPCR, Biochem, Immun	Stats, controls	Yes	DNA was prepared from fresh whole blood	Functionally abnormal. Novel missense variant			
Mutation of Presenilin Gene in Dilated C	10.1016/j.ymgme.2021.100474	Dilated Cardiomyopathy	P S E N 2	N A	c. 389C>T	p. S 1 3 0 L	940344_000023 Protein expression analysis	In vitro	Tissue studies, fibroblast culture studies, p	Measurement of [Ca ²⁺] _i response to ligand	3	To identify mutations: DNA samples, exons	N A	N A	N A	4	control patient (3 patients, 45 cells) control	Biological	Database: ERM, qPCR, Biochem, Immun	Stats, controls	Yes	DNA was prepared from fresh whole blood	Functionally abnormal. Missense variant of			

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Novel role of CACNA1C in familial AF	10.1093/eur.era144.152	Caechonheg polymorphi	R Y R 2	N A N A	p . 4 2 0 Q	ECM0020. Genet	In vitro	The Ry	Ca ²⁺ current concentration Ca ²⁺ release	N A N A 1	W T 2	Caffeir Biologi	Continuous variables expressed	Stats, neg control	Yes	Deletion of R1610 and previously approved	Functionally abnormal. The R1610Q	
Clonal and molecular diversity of	10.1093/eur.era144.153	Caechonheg polymorphi	R Y R 2	N A N A	p . 6 3 5 7 S	BAO:B	In vitro	Geneti	Ca ²⁺ release, gene expression	N A N A 1	W T 2	Caffeir Biologi	Median, standard error of the mean (SEM)	In silico predictions, stats, neg control	Yes	Mutated	Functionally abnormal. Pathogenicity not	
A novel P/Q type calcium channel	10.1093/eur.era144.154	Caechonheg polymorphi	R Y R 2	N A N A	p . 1 4 8 5 5 M	MM0020. In vitro	In vitro	Homol	Ca ²⁺ release, Ca ²⁺ release	N A N A 1	W T 2	8 Biologi	Mean ± SEM, p < 0.05	In silico predictions	Yes	The family was assessed for enrichment	Functionally abnormal. In silico predictions	
A type 3 polymorphic variant causes	10.1093/eur.era144.155	scTdPC	R Y R 2	N A N A	p . 5 4 9 3 8 F	OB0020. In vitro	In vitro	scTdPC	In silico analysis of Ca ²⁺ release	N A N A 1	W T 2	Cytopl Biologi	Data are expressed as mean ± SD	In silico analysis, stats, control	Yes	Missed R1610Q was P3 amplified in	Functionally abnormal. P3 amplified was	
A type 3 polymorphic variant causes	10.1093/eur.era144.156	scTdPC	R Y R 2	N A N A	p . V 1 0 2 4 I	OB0020. In vitro	In vitro	scTdPC	In silico analysis of Ca ²⁺ release	N A N A 1	W T 2	Cytopl Biologi	Data are expressed as mean ± SD	In silico analysis, stats, control	Yes	Missed R1610Q was P3 amplified in	Functionally abnormal. Pathogenicity not	
A type 3 polymorphic variant causes	10.1093/eur.era144.157	scTdPC	R Y R 2	N A N A	p . A 2 6 7 3 V	OB0020. In vitro	In vitro	scTdPC	In silico analysis of Ca ²⁺ release	N A N A 1	W T 2	Cytopl Biologi	Data are expressed as mean ± SD	In silico analysis, stats, control	Yes	Missed R1610Q was P3 amplified in	Functionally abnormal. Pathogenicity not	
Identification of a novel source of	10.1111/ajmh.11713	Caechonheg polymorphi	K C N J 2	N A N A	p . E 1 1 8 D	OB0020. Whole-exome	In vitro	Geneti	Whole-exome sequencing, Ca ²⁺ release	N A N A 1	W T 2	3 0 Biologi	All data are expressed as mean ± SD	Stats, neg control	Yes	After obtaining appropriate approval from	Functionally abnormal. NO/250C/100	
Identification of a novel source of	10.1111/ajmh.11714	Caechonheg polymorphi	R Y R 2	N A N A	c.169-351_170-657del	OB0020. Whole-exome	In vitro	Geneti	Whole-exome sequencing, Ca ²⁺ release	N A N A 1	W T 2	N A Biologi	All data are expressed as mean ± SD	Stats, neg control	Yes	After obtaining appropriate approval from	Functionally abnormal. Pathogenicity not	
Subclinical atrial fibrillation in	10.1093/eur.era144.158	Caechonheg polymorphi	R Y R 2	N A N A	p . Q 9 3 6 K	OBI:OC	In vitro	Expression of WT and mutant cAC1C2	Ca ²⁺ release, Ca ²⁺ release	N A N A 1	W T 2	N A Biologi	Data are expressed as mean ± SD	In silico analysis, stats, neg control	Yes	We collected the clinical data and blood	Functionally abnormal. Ca ²⁺ release	
AFK-related Mutations in Drosophila	10.1523/JNEUROSCI.4339-12.2013	Caechonheg polymorphi	R Y R 2	N A N A	p . G 1 8 8 6 S	OBI:OC	In vitro	HEK293T expression of cAC1C2	In silico analysis of Ca ²⁺ release	1	W T 3	4 Biologi	Mean values are presented as mean ± SE	Paired control, neg control, simple stats	Yes	To generate the AFK mutants, the corresp	Functionally abnormal. The substitution of	
AFK-related Mutations in Drosophila	10.1523/JNEUROSCI.4339-12.2013	Caechonheg polymorphi	R Y R 2	N A N A	p . G 1 8 8 5 E	OBI:OC	In vitro	HEK293T expression of cAC1C2	In silico analysis of Ca ²⁺ release	1	W T 3	4 Biologi	Mean values are presented as mean ± SE	Paired control, neg control, simple stats	Yes	To generate the AFK mutants, the corresp	Functionally abnormal. It was even more	
AFK-related Mutations in Drosophila	10.1523/JNEUROSCI.4339-12.2013	Caechonheg polymorphi	R Y R 2	N A N A	p . G 1 8 8 5 D	OBI:OC	In vitro	HEK293T expression of cAC1C2	In silico analysis of Ca ²⁺ release	1	W T 2	4 Biologi	Mean values are presented as mean ± SE	Paired control, neg control, simple stats	Yes	To generate the AFK mutants, the corresp	Functionally abnormal. It was even more	
Pathway mechanism of catecholamine	10.1093/eur.era144.159	Caechonheg polymorphi	R Y R 2	N A N A	p . A 1 6 5 D	ECM0020. Targeted	In vivo	Anima	ECG, e	N A N A 1	W T 2	3 - 7 Biologi	Data are presented as mean ± SD	Stats, control	Yes	All clinical studies were approved by the	Functionally abnormal. Clinical data	
Common variants in CACNA1C gene	10.1093/eur.era144.160	Brugada Syndrome	S C N 1 0 A	N A N A	p . S 5 0 9 P	OB0020. Whole-exome	In vitro	Cellular electrophysiology study with	In current and potential, binding effect	N A N A 1	W T 2	1 3 - 2 9 Biologi	Continuous variables with normal distribution	Stats, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. No significant associa	
Common variants in CACNA1C gene	10.1093/eur.era144.161	Brugada Syndrome	S C N 1 0 A	N A N A	p . 1 9 6 2 V	OB0020. Whole-exome	In vitro	Cellular electrophysiology study with	In current and potential, binding effect	N A N A 1	W T 2	1 3 - 2 9 Biologi	Continuous variables with normal distribution	Stats, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. No significant associa	
Common variants in CACNA1C gene	10.1093/eur.era144.162	Brugada Syndrome	S C N 1 0 A	N A N A	p . A 1 0 7 3 V	OB0020. Whole-exome	In vitro	Cellular electrophysiology study with	In current and potential, binding effect	N A N A 1	W T 2	1 3 - 2 9 Biologi	Continuous variables with normal distribution	Stats, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. Labels: 4073 and	
Common variants in CACNA1C gene	10.1093/eur.era144.163	Brugada Syndrome	S C N 1 0 A	N A N A	p . L 1 0 9 2 P	OB0020. Whole-exome	In vitro	Cellular electrophysiology study with	In current and potential, binding effect	N A N A 1	W T 2	1 3 - 2 9 Biologi	Continuous variables with normal distribution	Stats, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. No significant associa	
Mutations in CACNA1C are responsible for	10.1093/eur.era144.164	Brugada Syndrome	S C N 1 0 A	N A N A	p . R 1 4 L	OBI:OC	In vitro	Site di	Sodium channel current capability of Nat	1	W T 3	N A Biologi	Data are presented as mean ± SD	In silico pred, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. This mutation reduce	
Mutations in CACNA1C are responsible for	10.1093/eur.era144.165	Brugada Syndrome	S C N 1 0 A	N A N A	p . R 1 2 6 8 Q	OBI:OC	In vitro	Site di	Sodium channel current capability of Nat	1	W T 3	N A Biologi	Data are presented as mean ± SD	In silico pred, neg control	Yes	Informed consent was obtained from all	Functionally abnormal. The R1268Q substitution	
Novel CACNA1C p.R1610Q mutation	10.1093/eur.era144.166	Brugada Syndrome	S C N 5 A	N A N A	c.5000T>A	OB0020. Whole-exome	In vitro	Pathologic examination of cardiac	Electrophysiological properties, current	N A N A 1	W T 3	≥ 3 exp Biologi	Functional data are presented as mean ± SD	Stats, neg control	Yes	The patient and his relatives were evaluate	Functionally abnormal. We were first	
Brugada syndrome related by compound	10.1111/ajmh.11715	Brugada Syndrome	S C N 5 A	N A N A	p . 1 4 4 9 Y > H	OB0020. Whole-exome	In vitro	Pathologic examination of cardiac	Current density, location of mutation effects	N A N A 1	W T 2	N A Biologi	The data that support the findings of this s	The data that support the findings of this s	In silico predictions, inpatient observation	No	The data that support the findings of this s	Functionally abnormal. Use of functional
Causing mutations in polymorphisms	10.1093/eur.era144.167	Long C	S C N 5 A	N A N A	p . M 6 4 5 R	OB0020. Whole-exome	In vitro	Xenop	Electrophysiological properties of, MGS	N A N A 1	W T 2	N A Biologi	Mean ± SD, non-parametric Mann-Whitney	Stats, neg control	Yes	Functionally abnormal. Outcomes with	Functionally abnormal. Outcomes with	
Novel heterozygous mutation in CACNA1C	10.1093/eur.era144.168	Brugada Syndrome	S C N 5 A	N A N A	c.4282G>T	OB0020. Whole-exome	In vitro	Whole-exome sequencing of WT and mutant	Electrophysiological properties, current	N A N A 1	W T 2	1 0 - 1 6 Biologi	Mean ± SD, student's t, p < 0.05	Stats, neg control	Yes	Functionally abnormal. Current density	Functionally abnormal. Current density	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

High Risk for Brugada Syndrome	https://doi.org/10.1051/med/201530149	Brugada Syndrome	S C N 5 A	N A N A	p . T 1 8 7 I	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Electrophysiology (Current density, voltage)	N A N A	1	WT + p.R1750G	2	N A	Biological: wild-type/heterozygous	N A	Neg control	Yes	Subjects were Japanese Brugada patients	Functionally abnormal. No functional data	
High Risk for Brugada Syndrome	https://doi.org/10.1051/med/201530149	Brugada Syndrome	S C N 5 A	N A N A	p . D 3 5 6 N	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Electrophysiology (Current density, voltage)	N A N A	1	WT + p.R1750G	2	N A	Biological: wild-type/heterozygous	N A	Neg control	Yes	Subjects were Japanese Brugada patients	Functionally abnormal. No functional data	
High Risk for Brugada Syndrome	https://doi.org/10.1051/med/201530149	Brugada Syndrome	S C N 5 A	N A N A	p . K 1 5 7 0 F 1	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Electrophysiology (Current density, voltage)	N A N A	1	WT + p.R1750G	2	N A	Biological: wild-type/heterozygous	N A	Neg control	Yes	Subjects were Japanese Brugada patients	Functionally abnormal. No functional data	
High Risk for Brugada Syndrome	https://doi.org/10.1051/med/201530149	Brugada Syndrome	S C N 5 A	N A N A	p . R 1 6 2 3 X	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Electrophysiology (Current density, voltage)	N A N A	1	WT + p.R1750G	2	N A	Biological: wild-type/heterozygous	N A	Neg control	Yes	Subjects were Japanese Brugada patients	Functionally abnormal. No functional data	
The pathophysiological basis of the BrS	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . R 1 5 1 2 W	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Membrane voltage-clamp (subcurrents)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	near wild-type (subcurrents, no voltage)	Stats, neg control	Yes	R1520W was created in the COSM-4/2010	Functionally abnormal. No baseline control	
A Novel Mutation in the Cx36 Gene	https://doi.org/10.1007/s10046-013-0265-6	Long QT Syndrome 3	S C N 5 A	N A N A	p . L 6 1 9 F	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Membrane current (I _{NaP} , depolarization rate)	N A N A	1	WT	2	7 - 1 0	Biological: wild-type/heterozygous	near wild-type (subcurrents, no voltage)	Stats, neg control	Yes	This study was performed according to a protocol	Functionally abnormal. Current density was	
Dysgenic Changes of Cx36 and Connexin 43	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . R 1 7 9 X	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Whole-cell current (at various membrane potentials)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	N A	neg control	Yes	The patient gave informed consent for the study	Functionally abnormal. No baseline control	
Association Between Brugada Syndrome and Connexin 36	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . R 3 6 7 S	08/2002/0. Whole-exon capture	In vitro	1-14 kb (No-Exon Capture) (p.539-612)	Membrane current (voltage-clamp)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Continuous variables were compared using	neg control	Yes	Participants were recruited from the BrS registry	Functionally abnormal. No functional data	
Association Between Brugada Syndrome and Connexin 36	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . R 5 3 5	08/2002/0. Whole-exon capture	In vitro	1-14 kb (No-Exon Capture) (p.539-612)	Membrane current (voltage-clamp)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Continuous variables were compared using	neg control	Yes	Participants were recruited from the BrS registry	Functionally abnormal. Loss of function	
Association Between Brugada Syndrome and Connexin 36	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . R 8 9 3 C	08/2002/0. Whole-exon capture	In vitro	1-14 kb (No-Exon Capture) (p.539-612)	Membrane current (voltage-clamp)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Continuous variables were compared using	neg control	Yes	Participants were recruited from the BrS registry	Functionally abnormal. Missense. Loss of function	
Association Between Brugada Syndrome and Connexin 36	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . W 1 3 4 5 C	08/2002/0. Whole-exon capture	In vitro	1-14 kb (No-Exon Capture) (p.539-612)	Membrane current (voltage-clamp)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Continuous variables were compared using	neg control	Yes	Participants were recruited from the BrS registry	Functionally abnormal. Missense. Loss of function	
Association Between Brugada Syndrome and Connexin 36	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . G 1 7 4 3 R	08/2002/0. Whole-exon capture	In vitro	1-14 kb (No-Exon Capture) (p.539-612)	Membrane current (voltage-clamp)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Continuous variables were compared using	neg control	Yes	Participants were recruited from the BrS registry	Functionally abnormal. Missense. Loss of function	
Connexin 36 Mutation is Associated With Brugada Syndrome	https://doi.org/10.1007/s10046-013-0265-6	Biventricular Cardiomyopathy	S C N 5 A	N A N A	p . R 2 2 2 Q	08/2002/0. Whole-exon capture	In vitro	Patch clamp on Drosophila ommatidia	Membrane current density, activation rate	N A N A	1	WT	2	3	Biological: wild-type/heterozygous	N A	Neg control. Number of samples: 1 model	Yes	Membrane current was obtained from patch clamp	Functionally abnormal. Activating effect on	
New Connexin 36 Mutation in a Brugada Syndrome Patient	https://doi.org/10.1007/s10046-013-0265-6	Ventricular Pre-excitation	S C N 5 A	N A N A	p . R 2 2 5 P	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (at various membrane potentials)	N A N A	1	WT	2	7-16	Biological: wild-type/heterozygous	Substrate current (at various membrane potentials)	Neg control, 3 samples	Yes	The informed consent procedure was approved	Functionally abnormal. Loss of expression of Cx36	
Functional Analysis of Connexin 36 Gene in Brugada Syndrome	https://doi.org/10.1159/000318957	Brugada Syndrome	S C N 5 A	N A N A	p . V 2 8 1 A	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (density and conductance)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Data are presented as mean values ± standard deviation	Neg control, molecular model	Yes	This study was performed according to the protocol	Functionally abnormal. No functional data	
Functional Analysis of Connexin 36 Gene in Brugada Syndrome	https://doi.org/10.1159/000318957	Brugada Syndrome	S C N 5 A	N A N A	p . L 1 5 8 2 P	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (density and conductance)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Data are presented as mean values ± standard deviation	Neg control, molecular model	Yes	This study was performed according to the protocol	Functionally abnormal. No functional data	
Functional Analysis of Connexin 36 Gene in Brugada Syndrome	https://doi.org/10.1159/000318957	Brugada Syndrome	S C N 5 A	N A N A	p . F 5 4 3 L	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (density and conductance)	N A N A	1	WT	2	8-51	Biological: wild-type/heterozygous	Data are presented as mean values ± standard deviation	Neg control, molecular model	Yes	This study was performed according to the protocol	Functionally abnormal. Loss of function in the	
Functional Analysis of Connexin 36 Gene in Brugada Syndrome	https://doi.org/10.1159/000318957	Brugada Syndrome	S C N 5 A	N A N A	p . K 1 4 1 9 E	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (density and conductance)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	Data are presented as mean values ± standard deviation	Neg control, molecular model	Yes	This study was performed according to the protocol	Functionally abnormal. No substrate current	
Connexin 36 Mutation in a Brugada Syndrome Patient	https://doi.org/10.1007/s10046-013-0265-6	Long QT Syndrome	S C N 5 A	N A N A	p . N 1 3 2 5 S	01/2002/0. amino acid sequencing	In vivo	Examination of the tissue of adult myocardium	Histology and morphological parameters	1	Non-specific (WT) control tissue	3	microarray (WT COSM-154102) anti	4	3	Biological: wild-type/heterozygous	All values are presented as mean ± SEM. Statistics: control, samples, stats	Yes	Experimental procedures performed in this study	Functionally abnormal. Structural abnormality	
Compound Heterozygous Connexin 36 Mutations in Brugada Syndrome	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . F 1 5 7 1 L	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Substrate current (step depolarization protocol)	N A	Cells in control conditions	1	WT	3	7 - 9	Biological: wild-type/heterozygous	Data are reported as mean ± standard deviation	neg control, samples, stats	Yes	Variant introduced with QuickChange Site-Directed Mutagenesis	Functionally abnormal. The variant causes
Functional Analysis of a Novel Connexin 36 Mutation in Brugada Syndrome	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . G 1 7 1 2 C	08/2002/0. Whole-exon capture	In vitro	Conductance measurement in oocytes	Membrane current (stable baseline cell line)	N A N A	1	WT	2	1 0	Biological: wild-type/heterozygous	SPR-100 in the presence of near wild-type	neg control, stats	Yes	Plasmids pCMV40 containing the Cx36 mutation	Functionally abnormal. Loss of protein	
Multiple arrhythmic and conduction system abnormalities in a patient with Brugada Syndrome	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 5 A	N A N A	p . A 7 3 5 E	08/2002/0. Whole-exon capture	In vitro	Patch clamp on HEK293 cells with WT or mutant Cx36	Whole-cell sodium currents	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	N A	neg control	Yes	This study was approved by the institutional review board	Functionally abnormal. Loss of function. No	
Brugada Syndrome Caused by Solenon Connexin 36 Mutation	https://doi.org/10.1007/s10046-013-0265-6	Brugada Syndrome	S C N 1 B	N A N A	p . A 1 9 7 V	08/2002/0. Whole-exon capture	In vitro	Whole-exon capture on HEK293 cells	Na ⁺ channel gating (at various membrane potentials)	N A N A	1	WT	2	N A	Biological: wild-type/heterozygous	The electrophysiological data were analyzed	neg control, stats	Yes	This study was conducted in accordance with the protocol	Functionally abnormal. Missense, loss of function	

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Non-coding cardiac troponin I gene	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N I 3	N A N A	p. A 2 V	EN000085: Surrogate gene ex	In vitro	Two-h'	The effect of the mutation on troponin I k	N A N A 1	W T 4	Biologi	t test,	Stats, neg controls	Y e s	cell lines were established from	Functionally abnormal. No rescue was
Genetic Studies of Hypertrophic Cardiomyopathy	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N T 2	N A N A	p. R 2 3 6 H	EN000085: Gdpmwt assay evidence	In vitro	PP2C/D	size and structure function. The	1	W T 2	Biologi	4 0 8	Stats, neg controls	Y e s	CFP-tagged PP2C/D were expressed into	Functionally abnormal. No rescue was
D38V mutation in Desmin in Cardiac myocytes	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N I 3	N A	c. G 5 6 8 T p. D 1 9 0 Y	EN000085: Molecular mechanism study	In silico	5 ml of	Pathogenicity, conservation and effect on	N A	N A N A N A N A N A N A N A	Biologi	Control for identification of variants	Stats, neg controls	N o	Thin plates developed with oligonucle	Functionally abnormal. No rescue was
Altered Regulation Properties of Human Ca	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N I 3	N A N A	p. R 1 4 5 G	BAO:B	In vitro	The Tit constructs were overexpressed in E	Reconstitution of Troponin Complex, Pri	N A N A 1	W T 3	Biologi	9	Stats, neg controls	Y e s	Human cDNAs (TL) obtained as a gift fr	Functionally abnormal. Both mutations ca
Altered Regulation Properties of Human Ca	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N I 3	N A N A	p. R 1 6 2 W	BAO:B	In vitro	The Tit constructs were overexpressed in E	Reconstitution of Troponin Complex, Pri	N A N A 1	W T 3	Biologi	9	Stats, neg controls	Y e s	Human cDNAs (TL) obtained as a gift fr	Functionally abnormal. Both mutations ca
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. R 1 3 4 G	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber force development was evaluated wit	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. R 1 5 1 C	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber f	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. G 1 5 9 Q	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber f	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. R 2 0 5 W	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber f	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. K 2 1 0 del	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber f	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Clinical and Functional Characterization of	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. E 2 4 4 D	OB:000145: Binding assay	In vitro	To identify the mutations. Biochemical exp	Fiber f	N A N A 1	W T 2	Biologi	N A	Stats, neg controls	Y e s	Human cardiac wild type troponin T (TC	Functionally abnormal. No rescue was
Loss of One Copy of Troponin C in Cardiac myocytes	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T N N T 2	N A N A	p. R 1 3 9 H	OB:000145: Binding assay	In vitro	Functional studies in cardiac muscle slices	The Ca ²⁺ -dependent force development	N A N A 1	W T 2	Biologi	5	Stats, neg controls	N o	Written informed consent was obtained fr	Functionally abnormal. No rescue was
Mutation Location of Hypertrophic Cardiomyopathy	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N T 2	N A N A	p.179N	OB:000145: Binding assay	In vitro	Hypertrophic Cardiomyopathy (HCM) case comp	Myofibrillar function after troponin reconst	5	W T 1	Biologi	0 2	Stats, pos control, neg control	Y e s	Exchange experiments were performed in	Functionally abnormal. A small rescue fr
Mutation Location of Hypertrophic Cardiomyopathy	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N T 2	N A N A	p.R94C	OB:000145: Binding assay	In vitro	Hypertrophic Cardiomyopathy (HCM) case comp	Myofibrillar function after troponin reconst	5	W T 1	Biologi	0 2	Stats, pos control, neg control	Y e s	Exchange experiments were performed in	Functionally abnormal. A small rescue fr
Mutation Location of Hypertrophic Cardiomyopathy	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T N N T 2	N A N A	p. R 2 7 8 C	OB:000145: Binding assay	In vitro	Hypertrophic Cardiomyopathy (HCM) case comp	Myofibrillar function after troponin reconst	5	W T 1	Biologi	0 2	Stats, pos control, neg control	Y e s	Exchange experiments were performed in	Functionally abnormal. No rescue was
Loss of Function Mutations in MYH6 in Cardiac myocytes	https://doi.org/10.1016/j.ymgme.2014.04.004	Left Ventricular Non-Compaction	N N T	N A	c. 6310_6339 insT p.R213fs	EN000085: amino acid substitution study	in vivo	Complementation strategy in transgenic	MO dc 1	GFP control 1	W T 2	Biologi	Mutat	Stats, pos control, neg control	Y e s	RNAi approach for this study was established fr	Functionally abnormal. No rescue was
Troponin gamma phosphorylation	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T P M 1	N A N A	p.S283D	BAO:B	In vitro	Differen	DSC to	N A N A 1	W T 2	Biologi	3	Stats, neg controls	Y e s	The S283D and S210L in HCM mutant con	Functionally abnormal. No rescue was
Troponin gamma phosphorylation	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T P M 1	N A N A	p.S61D	BAO:B	In vitro	Differen	DSC to	N A N A 1	W T 2	Biologi	3	Stats, neg controls	Y e s	The S283D and S210L in HCM mutant con	Functionally abnormal. No rescue was
Genetic Studies of Hypertrophic Cardiomyopathy	https://doi.org/10.1016/j.ymgme.2014.04.004	Dilated Cardiomyopathy	T P M 1	N A N A	p. D 2 3 0 N	MMO:	In vitro	In vitro	ATPase	N A N A 1	W T 2	Biologi	3	Stats, neg controls	Y e s	All subjects provided informed consent an	Functionally abnormal. No rescue was
Mechanistic Investigation of Cardiac myocytes	https://doi.org/10.1016/j.ymgme.2014.04.004	Hypertrophic Cardiomyopathy	T P M 1	N A N A	p. E 6 2 Q	BAO:B	In vitro	DNA, p	Measurement of Ca ²⁺ sensitivity of huma	1	W T 2	Biologi	3 (each experiment)	Stats, neg controls	Y e s	Plasmids encoding human cardiac alpha	Functionally abnormal. No rescue was

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

Identificación de Mutaciones Prónicas <https://doi.org/10.1002/ajmg.12914> Hipertrofia Cardiaca M Y H 7 N A N A p. N 2 3 2 S 08/2004 Sequence assay In vitro Three high polymorphic chromosome 9p1 Electroporation in cardiac cells N A N A N A N A N A N A N A N A N A N A N A N A N A N A N o Genomic DNA as a source of DNA for functional studies Functional genomics in the heart

ANEXO 2. CÓDIGO DE PYTHON EMPLEADO EN EL CASO DE USO

```
# PART1 REDUCE VCF
import os
import pandas as pd
# List of gene names to match
gene_names = ['ACTC1', 'ANKRD1', 'BRAF', 'CACNA1C', 'DES', 'DSC2', 'DSG2', 'DSP', 'FHOD3', 'FLNC', 'GATA4', 'JUP', 'KCNH2', 'KCNH2', 'KCNJ2', 'KCNQ1', 'KCNQ1', 'LAMA4', 'LMNA',
'MYBPC3', 'MYH7', 'NEXN', 'NNT', 'PKP2', 'PSEN1', 'PSEN2', 'RBM20', 'RBM21', 'RBM22', 'RYR2', 'SCN10A', 'SCN1B', 'SCN5A', 'SCN5A', 'TGF', 'TMEM43', 'TNNC1', 'TNNI3', 'TNNT2',
'TPM1', 'TTN', 'TTNT2', 'VCL']
parent_directory = r'C:\Users\olymp\Documents\TFG\INFORMES_Cardio_Alicante_Abril'
folder_names = ['OGM-04-{:03d}'.format(i) for i in range(1, 43)]
# Loop through the folders
for folder_name in folder_names:
    # Construct the file path
    csv_file = os.path.join(parent_directory, folder_name, 'annotated.hg19_multianno.csv')
    # Load the CSV file
    csv_data = pd.read_csv(csv_file, delimiter='\t')
    # Filter rows based on matching gene names
    filtered_data = csv_data[csv_data['Gene.refGene'].isin(gene_names)]
    # Check if any rows matched the gene names
    if len(filtered_data) == 0:
        print(f"No matching rows found in folder: {folder_name}")
    else:
        # Create a new file with the filtered data
        output_file = os.path.join(parent_directory, folder_name, 'reduced_annotated.hg19_multianno.csv')
        filtered_data.to_csv(output_file, index=False)
        print(f"Filtered data saved to: {output_file}")
    print() # Add an empty line between each folder's results
# PART2 FIND VARIANTS, RETURN FUNCTIONAL IMPACT
import re
csv_file2 = r'C:\Users\olymp\Documents\TFG\infoCSHG_csv.csv'
parent_directory = r'C:\Users\olymp\Documents\TFG\INFORMES_Cardio_Alicante_Abril'
folder_names = ['OGM-04-{:03d}'.format(i) for i in range(1, 43)]
# Loop through the folders
for folder_name in folder_names:
```

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

```
# Construct the file path
csv_file1 = os.path.join(parent_directory, folder_name, 'reduced_annotated.hg19_multianno.csv')
# Load the CSV files
csv_data1 = pd.read_csv(csv_file1)
csv_data2 = pd.read_csv(csv_file2)
# Extract the desired columns from both datasets
varan_data = csv_data1[['AChange.refGene', 'GeneDetail.refGene', 'CLNHGVS']]
varfunp_data = csv_data2['Variant in protein (p.)'].astype(str)
varfunc_data = csv_data2['Variant in transcript (c.)'].astype(str)
varfung_data = csv_data2['Genomic position'].astype(str)
# Extract the varfunp values using regular expressions
varfunp_values = varfunp_data.str.findall(r"p\[A-Za-z\d*_>\(\)\+=]+\").explode().tolist()
varfunc_values = varfunc_data.str.findall(r"c\[A-Za-z\d*_>\(\)\+=]+\").explode().tolist()
varfung_values = varfung_data.str.findall(r"g\[A-Za-z\d*_>\(\)\+=]+\").explode().tolist()
# Extract the varan values using regular expressions
varanp_values = varan_data['AChange.refGene'].apply(lambda x: re.findall(r"p\[A-Za-z\d*_>\(\)\+=]+\\"", str(x))).explode().tolist()
varanc_values1 = varan_data['AChange.refGene'].apply(lambda x: re.findall(r"c\[A-Za-z\d*_>\(\)\+=]+\\"", str(x))).explode().tolist()
varanc_values2 = varan_data['GeneDetail.refGene'].apply(lambda x: re.findall(r"c\[A-Za-z\d*_>\(\)\+=]+\\"", str(x))).explode().tolist()
varang_values = varan_data['CLNHGVS'].apply(lambda x: re.findall(r"g\[A-Za-z\d*_>\(\)\+=]+\\"", str(x))).explode().tolist()
# Find the common elements
common_elementsp = set(varanp_values).intersection(varfunp_values)
common_elementsc = set(varanc_values1).intersection(varfunc_values).union(set(varanc_values2).intersection(varfunc_values))
common_elementsg = set(varang_values).intersection(varfung_values)
# Print the folder name
print(f"Folder: {folder_name}")
common_elements = set()
common_elements |= common_elementsp
common_elements |= common_elementsc
common_elements |= common_elementsg
for element in common_elements:
    matched_row_p = csv_data2[csv_data2['Variant in protein (p.)'] == element]
    matched_row_c = csv_data2[csv_data2['Variant in transcript (c.)'] == element]
    matched_row_g = csv_data2[csv_data2['Genomic position'] == element]
    if len(matched_row_p) > 0:
```

Diseño y desarrollo de una fuente de datos sobre estudios funcionales asociados al criterio PS3/BS3 de las guías ACMG-AMP 2015 y aplicación en cardiopatías familiares

```
    impact_text_p = matched_row_p['functionalImpact'].values[0]
    print(f"Variant: {element} \nImpact (Protein): {impact_text_p}")
    print()
if len(matched_row_c) > 0:
    impact_text_c = matched_row_c['functionalImpact'].values[0]
    print(f"Variant: {element} \nImpact (Transcript): {impact_text_c}")
    print()
if len(matched_row_g) > 0:
    impact_text_g = matched_row_g['functionalImpact'].values[0]
    print(f"Variant: {element} \nImpact (Genomic): {impact_text_g}")
    print()
print()
```