



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

MAESTRÍA Y DOCTORADO EN CIENCIAS BIOQUÍMICAS

“ANÁLISIS DEL PAPEL EMERGENTE DE REDES DE INTERACCIÓN DE VARIANTES GENÉTICAS QUE EXPLICAN LA ASOCIACIÓN ENTRE LA ATEROESCLEROSIS Y EL DESARROLLO DE DEMENCIA VASCULAR”

TESIS

PARA OPTAR EL GRADO DE:

MAESTRO EN CIENCIAS BIOQUÍMICAS

PRESENTA:

ARMANDO OCAMPO DEL VALLE

DIRECTOR DE TESIS:

**DR. HERIBERTO MANUEL RIVERA
FACULTAD DE NUTRICIÓN, UAEM**

MIEMBROS DEL COMITÉ TUTOR:

**DR. ALBERTO DARSZON ISRAEL
INSTITUTO DE BIOTECNOLOGÍA, UNAM**

**DRA. PERLA DEYANIRA MALDONADO JIMÉNEZ
INSTITUTO NACIONAL DE NEUROLOGÍA Y NEUROCIROLOGÍA, PATOLOGÍA
VASCULAR CEREBRAL**

**DR. MAXIMINO ALDANA GONZÁLEZ
INSTITUTO DE CIENCIAS FÍSICAS, UANM**

**DRA. ALEJANDRA MEDINA RIVERA
LABORATORIO INTERNACIONAL DE INVESTIGACIÓN SOBRE EL GENOMA
HUMANO, UNAM**

**DRA. ROSA MARIA URIBE VILLEGAS
INSTITUTO DE BIOTECNOLOGÍA, UNAM**

CIUDAD DE MÉXICO, FEBRERO, 2022



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

“Sin importar lo mala que parezca ser la vida, siempre hay algo que puedes hacer y tener éxito. Mientras haya vida, hay esperanza”

Stephen Hawking

Agradecimientos

Este trabajo va dedicado a cada una de las personas que han creído en mí y me han apoyado. A mi familia, en especial a mi madre. Gracias por todo lo que me has dado, sin ti yo no podría tener nada de lo que he logrado. Al laboratorio, por los consejos brindados en lo académico, gracias doctor Rivera por hacerme participe de este proyecto. A Erika y su familia, por acompañarme en este viaje que hemos empezado a emprender y por su puesto a Scraps, Max y Vikingo, por darle alegría a la casa cuando se necesita. Gracias a mis amigos Edson, José Antonio, Karen, Karla, Chora, Rolando, Jona, Hugo por seguir acompañándome.

Este proyecto se realizó gracias al financiamiento del Consejo Nacional de Ciencia y Tecnología (CONACYT) y el apoyo de la máxima casa de estudios del país, la UNAM. Gracias brindar una educación pública y de calidad. Mi reconocimiento a la unidad de computo del Instituto de Biotecnología, por el esfuerzo realizado durante la pandemia. Gracias a la unidad académica, a Toño, Jalil, Gloria, por responder nuestras dudas y ayudarnos con cada trámite.

Estos últimos años han sido difíciles para la mayoría de nosotros. Y a pesar de todo, estuve rodeado de gente que me ayudó a crecer en lo académico y cómo persona. Como ya he mencionado, este trabajo está dedicado a todas las personas que están, y sobre todo a aquellos que espero algún día volver a ver. Infinitamente gracias.

Índice

	Página
Resumen	6
Introducción	7
Demencia vascular	7
Ateroesclerosis	7
Barrera hematoencefálica y unidad neurovascular	7
Fisiopatología de la demencia vascular y la ateroesclerosis	9
Epidemiología	11
Diagnóstico	11
Antecedentes	13
Medicina de precisión	13
Biomarcadores	13
Polimorfismos de un solo nucleótido	14
Estudio de asociación genómica completa	15
Desequilibrio de ligamiento	16
Estructura proteica y polimorfismos de un solo nucleótido	17
Definición de variantes asociadas a demencia vascular y ateroesclerosis	18
Pregunta de investigación	20
Hipótesis	20
Objetivo general	20
Objetivos particulares	20
Metodología	21
Asociación de variantes	21
Selección de un modelo celular en el desarrollo de demencia vascular	21
Redes de interacción	22
Desequilibrio de ligamiento de variantes propuestas	23
Asociación de polimorfismos por grupo poblacional y tipo de variante	24
Evaluación de estructura secundaria y efecto de los polimorfismos de un solo nucleótido	25
Procesamiento de datos y software	29
Resultados	31
Variantes presentes en demencia vascular y ateroesclerosis	31
Tipos de variantes	32
Localización cromosómica	33
Frecuencia poblacional	34
Asociación de polimorfismos con estirpes celulares	35
Vías metabólicas asociadas a demencia vascular y ateroesclerosis	36
Red de interacción	36
Desequilibrio de ligamiento y grupos poblacionales	38
Evaluación de estructura secundaria	46
Evaluación del efecto del polimorfismo en términos estructurales de la proteína ..	47
Discusión	49

	Página
Conclusión	56
Perspectivas	57
Limitaciones del proyecto	58
Consideraciones éticas	59
Anexo	60
Abreviaturas	74
Bibliografía	75

Resumen

El origen de la demencia vascular se relaciona con un conjunto de eventos que limitan el aporte de oxígeno y flujo sanguíneo hacia el sistema nervioso central. Dentro de estos factores se encuentra la aterosclerosis, la cual es una patología inflamatoria crónica de los vasos sanguíneos. De hecho, se han propuesto modelos fisiopatológicos que relacionan ambas enfermedades. No obstante, en este proyecto nos realizamos la siguiente pregunta: ¿es posible conocer la relación de las variantes genéticas asociadas a aterosclerosis y el desarrollo de demencia vascular? Para responder este cuestionamiento se identificaron los polimorfismos de un solo nucleótido derivados de estudios de asociación de genoma completo para ambas enfermedades. Posteriormente, mediante el uso de bases de datos y herramientas bioinformáticas se identificó su localización, su papel en el metabolismo, su relación con células de la unidad neurovascular y su desequilibrio de ligamiento. Los resultados obtenidos denotan a los oligodendrocitos y astrocitos como principales estirpes celulares en el desarrollo de demencia vascular. Asimismo, se denota a las variantes rs200133978 y rs143113210 presentes en los genes FAF1 y MMP20 como biomarcadores para el abordaje de esta enfermedad.

Introducción

Demencia vascular

La demencia vascular o deterioro cognitivo vascular es el segundo tipo de demencia en cuanto a casos reportados a nivel mundial, encontrándose solo por debajo de la demencia derivada a partir de la enfermedad de Alzheimer (1). Su origen se atribuye a un conjunto de factores cardiovasculares como disminución del flujo sanguíneo y reducción del aporte de oxígeno. Estos procesos derivan en la pérdida de conexiones entre neuronas, así como muerte de las células que conforman el tejido cerebral. De esta manera se da pauta a la disminución de las funciones cognitivas como lenguaje, memoria o interacción social (2). En virtud de que las alteraciones en la memoria pueden variar en cada persona durante la evolución de este padecimiento, se ha removido este criterio para el diagnóstico de la demencia vascular (3).

Ateroesclerosis

La ateroesclerosis por su parte hace referencia a la asociación entre la remodelación y posterior rigidez de arterias de mediano y gran calibre, así como la degradación y oxidación de lípidos que culminan en la formación de placas en las capas endoteliales de las arterias. Estas lesiones generan una serie de complicaciones cardiovasculares (4); por ejemplo la enfermedad isquémica coronaria y el evento cerebrovascular de tipo isquémico (5). Etiológicamente, se desconocen los factores que precipitan la acumulación de lípidos y formación de placas (6). No obstante, el riesgo de ruptura de la placa y el desarrollo de eventos trombóticos depende de la composición del núcleo lipídico de la misma, donde altos niveles de lípidos oxidados y aumento de células apoptóticas (células en eventos de muerte celular programada) debilitan la pared de la lesión ateroesclerótica. Lo que fomenta la exposición de moléculas capaces de activar al sistema plaquetario y el desarrollo de trombos, mismos que al viajar por el torrente sanguíneo y quedar atrapados en vasos de menor calibre disminuyen el paso de oxígeno (7). La culminación de estos eventos es la muerte de las células que componen el tejido blanco así como las complicaciones antes mencionadas (8).

Barrera Hematoencefálica (BHE) y unidad neurovascular

El cerebro humano es un órgano que consume el 20% del oxígeno y glucosa en estado basal (9). La mayor parte de la energía es utilizada por las neuronas para regular los gradientes iónicos y reestablecer los potenciales de membrana, posterior a la despolarización de las mismas (10). El resto se emplea en las vías de señalización celular, síntesis de proteínas y metabolismo celular (9). A la relación presente de las células que componen este tejido se le conoce como unidad neurovascular (UNV), la cual se compone por neuronas, astrocitos, oligodendrocitos, microglía, pericitos, células endoteliales y la membrana basal (11). Esta interacción está mediada por las uniones celulares y moléculas de adhesión que permiten la comunicación entre ellas. Cada elemento de la UNV tiene un papel

específico en la homeostasis de este tejido, por lo cual se hará una breve descripción de cada uno (Figura 1) (12).

Las neuronas, además de participar en procesos de plasticidad e integración de los estímulos periféricos se encargan de detectar la cantidad de oxígeno y nutrientes que llegan al sistema nervioso central a través de los vasos sanguíneos. De esta manera, transforman esta información en mensajes químicos y eléctricos que pueden generar dilatación o contracción de los capilares cerebrales. Estos estímulos llegan a las células endoteliales que recubren los vasos sanguíneos a través de interneuronas o las prolongaciones de los astrocitos, influyendo de esta manera en el tono vascular (12). Los astrocitos, como se ha mencionado poseen prolongaciones que le permiten ser punto de comunicación entre neuronas y vasos capilares. Asimismo, sintetizan y liberan sustancias como prostaglandinas, ATP u óxido nítrico, los cuales interactúan con las células endoteliales en eventos de vasodilatación o vasoconstricción (11). Los oligodendrocitos son responsables de la formación y mantenimiento de la vaina de mielina que envuelve los axones neuronales, este recubrimiento favorece que los potenciales de acción sean más rápidos y eficaces (13)(14). Además, liberan factores tróficos que permiten la supervivencia neuronal y el mantenimiento de la vaina de mielina (15). La microglía corresponde a células del sistema inmune en el cerebro, mismas que se activan durante eventos de daño. Cabe destacar que presenta dos fenotipos, uno de carácter proinflamatorio conocido como M1 y otro de tipo antiinflamatorio denominado M2 el cual participa en mecanismos de reparación durante procesos inflamatorios (1). Los pericitos durante la fase embrionaria liberan factores tróficos para el desarrollo de la red vascular. A lo largo de la etapa adulta se encuentran en la periferia de los vasos capilares, presentan proteínas contráctiles como alfa actina, tropomiosina y miosina que permiten su papel en la regulación del flujo sanguíneo y diámetro capilar (11). Las células endoteliales recubren los vasos sanguíneos, a partir de sus uniones celulares de tipo estrechas y adherentes, su interacción con los pericitos provee un mecanismo de permeabilidad selectiva que limita el paso de metabolitos presentes en el torrente sanguíneo. Este mecanismo se denomina barrera hematoencefálica (BHE) (12). La BHE permite el intercambio de moléculas entre el tejido cerebral y el torrente sanguíneo. Estos elementos se trasladan a través de tres tipos de transporte: A) un transporte pasivo mediante difusión de moléculas lipofílicas y moléculas pequeñas como oxígeno y dióxido de carbono. B) difusión facilitada de glucosa, aminoácidos y nucleósidos por medio de acarreadores y C) transporte de macromoléculas por vesículas (16)(1). Por último, La membrana o lámina basal se forma por proteínas secretadas por las células endoteliales y pericitos. Esta membrana rodea a los capilares separándola de los pericitos y las terminales de los astrocitos. De la misma manera, se encarga de generar un punto de soporte para los vasos sanguíneos y las células de la UNV mediante moléculas de adhesión y proteínas de unión celular (11). La comunicación entre estos componentes favorece la homeostasis de procesos metabólicos, desarrollo cerebral y reparación (1). Se ha descrito que los eventos con disminución del aporte de oxígeno y glucosa en el sistema nervioso central deriva en disfunción de la BHE y de la UNV, derivando en daño celular, muerte celular no programada y patogénesis de enfermedades neurocognitivas (17).

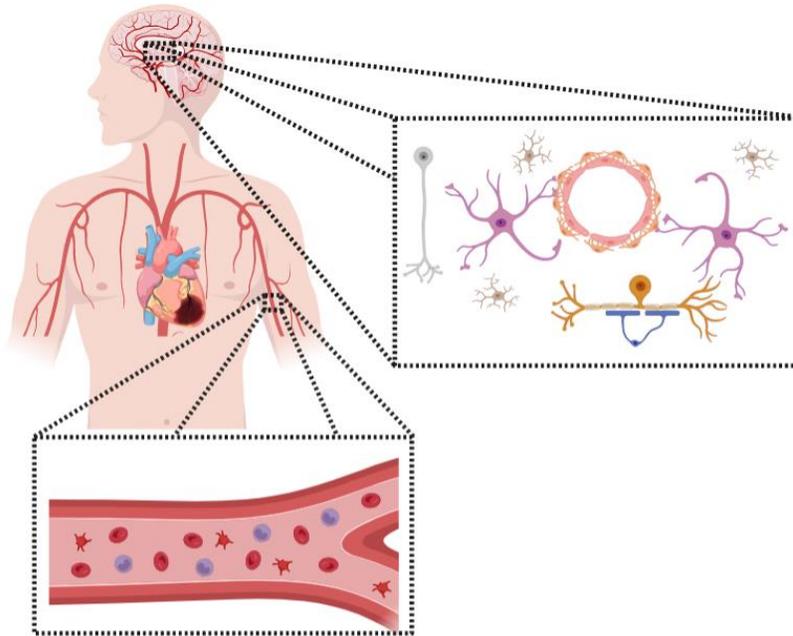


Figura 1. Fisiología de la unidad neurovascular y sistema circulatorio.

Fisiopatológica de la demencia vascular y la aterosclerosis

En condiciones fisiológicas donde existe un flujo sanguíneo constante que provee oxígeno y glucosa los miocitos, pericitos y células endoteliales mantienen una adecuada función de la BHE (12). No obstante, se ha descrito que cualquier perturbación que genere deficiencia en el flujo sanguíneo con una privación de oxígeno y glucosa por más de 90 segundos iniciará una cascada de procesos isquémicos. Es aquí donde se hace presente la relación entre las enfermedades cardiovasculares como la aterosclerosis y el desarrollo de demencia vascular, debido a que la formación de trombos posterior a la rotura de la placa aterosclerótica bloquean las arterias de menor calibre en el cerebro, limitando el flujo sanguíneo. Esto da como resultado una disminución del aporte de nutrientes hacia este tejido (17). De hecho, se ha detallado que los eventos tromboembólicos derivados de patologías cardiovasculares son responsables de al menos el 50% de los accidentes cerebrovasculares de tipo isquémico (18). Una vez presente la disminución del flujo sanguíneo se inicia una cadena de eventos inflamatorios, modulado tanto por las células del sistema inmune periférico como por la microglía (Figura 2). En este punto, la activación de la microglía permite la liberación de metaloproteasas y mediadores proinflamatorios, las cuales degradan la vaina de mielina de los axones neuronales, así como las uniones celulares que conforman la BHE (1). Este incidente da como resultado la pérdida de la permeabilidad selectiva, lo que favorece que los componentes del torrente sanguíneo puedan ingresar al tejido cerebral y tener contacto con las células que lo componen, generando daño y muerte de estas. Aunado a esto, se presentan los siguientes eventos en la fisiopatología de la demencia vascular: 1) Formación de radicales libres por

transporte de hemoglobina y hierro libre (Fe^{2+}), 2) degradación de la matriz extracelular e 3) inflamación y edema cerebral (17). A su vez, estos procesos patológicos permiten la activación de los astrocitos, con una posterior liberación de mediadores proinflamatorios, los cuales aumentan el daño; derivando en un mecanismo de retroalimentación positiva (2). Además del desequilibrio en la BHE, existe un desacoplamiento de la UNV, con ausencia de la autorregulación del flujo sanguíneo y limitación de liberación de factores tróficos. En tanto, como se mencionó previamente las metaloproteasas secretadas por la microglía generan daño en los oligodendrocitos y en la vaina de mielina que envuelve el axón neuronal, lo cual afecta la propagación del impulso nervioso, así como el aporte trófico hacia el axón resultando en degeneración axonal. Este evento da como resultado una disminución de la sinapsis neuronal, mismo que se considera como uno de los eventos principales que dan parte al declive cognitivo (1). Por último, Se denota que el sitio anatómico con mayor susceptibilidad en presentar este tipo de eventos es el hipocampo (17), con afección de las neuronas CA1 que se encuentran en esta región. Se ha relacionado a las neuronas de esta zona cerebral con la modulación de señales que permiten el aprendizaje, memoria y funciones cognitivas, por lo cual se sugiere que el daño y destrucción de estas se ve reflejado en el asentamiento del cuadro clínico presente en la demencia vascular (2).

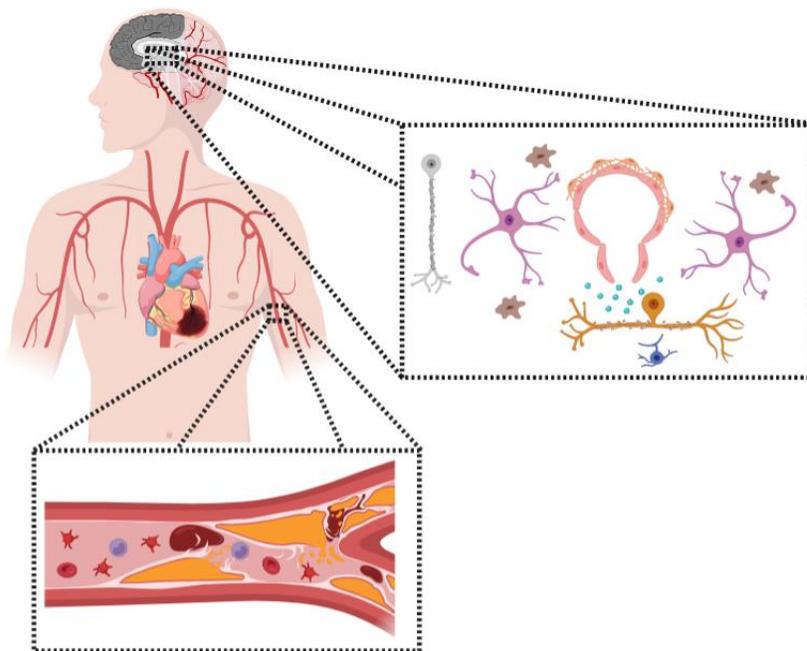


Figura 2. Fisiopatología de la demencia vascular. La formación de trombos derivado de la ruptura de la placa aterosclerótica bloquea la microcirculación de corazón y cerebro, dando parte a una disminución del aporte de oxígeno y glucosa. El resultado de este evento en el sistema nervioso central es la disfunción de la BHE y daños a las estirpes celulares de la UNV

Epidemiología

La demencia vascular es el segundo tipo de demencia con más casos reportados en el mundo (1). Actualmente, datos epidemiológicos describen un escenario mundial conformado por aproximadamente 17 millones de personas diagnosticadas con un tipo de demencia, de los cuales el 15% corresponde a demencia vascular, representando un costo anual de 200 mil millones de dólares (19). Asimismo, la Organización Mundial de la Salud (OMS) estima que para el año 2050 la población con demencia será mayor a 100 millones, de los cuales 15 millones corresponderán a demencia vascular (1). Se ha detallado que la demencia vascular se presenta en un 15 a 30% de personas que cuentan con antecedente de un evento cerebrovascular de tipo isquémico en los últimos 3 meses (3). A partir de esta asociación entre enfermedades cardiovasculares y alteraciones cognitivas, la Organización Mundial del Infarto sugiere que la prevención de los factores de riesgo de entidades cardiovasculares pueden disminuir el número de personas con probabilidad de presentar declive cognitivo vascular (17).

Una de las medidas estadísticas que nos permite conocer la relación de personas que presentan una enfermedad comparada con el resto de la población es la prevalencia. A partir de esta información se identifica el porcentaje de enfermos en proporción al total de habitantes. La prevalencia de demencia vascular en personas mayores de 65 años en comunidad europea es del 1.6%. Por su parte, el estudio de edad, demografía y memoria (ADAMS por sus siglas en inglés) realizado en Estados Unidos, describe una prevalencia del 14% en personas mayores de 70 años (20). En el caso de México, no existe una división concisa entre los tipos de demencia. No obstante, se estima que el 7.3% de la población mexicana presenta alguna forma de demencia. De hecho, se espera que para el año 2050 la cifra se aproxime a 3.5 millones de mexicanos con este padecimiento (21). En población asiática se presenta variación en cuanto a los casos epidemiológicos. Existiendo una prevalencia en mayores de 65 años de 6.3% y 11% en Corea y Japón, respectivamente. Asimismo, en China, solo se ha reportado esta entidad nosológica en el 0.9% de sus habitantes mayores de 60 años. Por último, para el continente africano a pesar de tener dificultades para la obtención y organización de la información; mediante un análisis sistemático se muestra una prevalencia de 2.4% en personas mayores de 50 años (20).

Diagnóstico

Durante el diagnóstico de la demencia vascular es necesario realizar una evaluación neuropsicológica del paciente (22) a partir de pruebas donde se explora la atención y la función motora. Actualmente, los exámenes más utilizados son: 1) Escala de evaluación cognitiva de Montreal y 2) Mini examinación del estado mental, cuyo objetivo es evaluar alteraciones cognitivas (3). Asimismo, para el apoyo del diagnóstico se utilizan imágenes clínicas (22), como resonancia magnética que permite conocer si se presentan infartos en alguna zona anatómica del tejido o si existe atrofia de la corteza cerebral (3). Aunado a los hallazgos clínicos y radiológicos es necesario establecer una relación con una enfermedad

cardiovascular que explique la etiología del cuadro clínico del paciente (23). Sin embargo, debido a la poca sensibilidad y especificidad de los métodos diagnósticos para la demencia vascular, en ocasiones los casos suelen tener un diagnóstico en etapas avanzadas de la enfermedad o *post-mortem* mediante autopsia (22). De igual modo, la detección de aterosclerosis se realiza en etapas avanzadas. Esta se basa en un estudio denominado angiografía, el cuál utiliza un catéter intravascular y medio de contraste para localizar placas de ateroma. Este proceso es invasivo y puede presentar complicaciones como hemorragia, efecto vaso-vagal o reacción anafiláctica al medio de contraste. Asimismo, este estudio no permite identificar qué placas son susceptibles a presentar roturas; siendo un método poco sensible (24). Por tal motivo, es crucial una evaluación temprana de estas entidades nosológicas, tomando en cuenta las características genéticas y ambientales, generando medidas de prevención y metodologías diagnósticas específicas que disminuyan la mortalidad o la probabilidad de desarrollar discapacidades (25). Por último, se ha reportado que el tratamiento en ambas patologías presenta un alto costo a través del tiempo y variabilidad en la efectividad. De esta manera, se refuerza el argumento de la búsqueda de metodologías preventivas y específicas (26).

Antecedentes

Medicina de precisión

En la actualidad, el abordaje clínico de las enfermedades complejas toma como metodología principal el análisis de un conjunto de signos y síntomas presentes en un periodo de tiempo corto, haciendo una relación de la etiología, los mecanismos fisiopatológicos y las estructuras anatómicas con el desarrollo de una enfermedad. De hecho, la elección de estrategias diagnósticas y terapéuticas se siguen de manera estadística. Se asume que todos los individuos comparten un genotipo y fenotipo en común (27). Como respuesta a este problema surge la medicina de precisión (MP), la cual tiene como objetivo generar un mejor entendimiento y estratificación de las enfermedades en conjunto con una detección y monitorización de la patología de interés (28). Esto provee al personal de salud de estrategias preventivas y terapéuticas que incluyen la variabilidad genética individual, ambiental (27), interacción con factores de riesgo (29) e historial familiar (30).

Dentro de los elementos que permiten a la MP cumplir este objetivo se incluyen la información de diversas bases de datos, el uso de técnicas computacionales y el conocimiento de las interacciones moleculares asociados a un fenotipo en un estado fisiológico o patológico (27). Por esta razón, es necesario un análisis multidimensional para convertir información heterogénea en datos útiles relacionados con un valor de riesgo en una enfermedad (28). Es aquí donde la ciencia de datos y las tecnologías ómicas permiten integrar la información (27), que incluye referencias a nivel celular y poblacional (28); lo que en su conjunto se conoce como biología de sistemas. El análisis “ómico” provisiona detalles de la naturaleza de la disfunción celular e identifica las vías y mecanismos moleculares asociados a enfermedades crónico degenerativas (31). En el contexto de la evidencia científica, la MP plantea la generación de modelos matemáticos que estimen la probabilidad de desarrollar una patología en particular en asociación a la variabilidad poblacional (32). La comprensión de estas metodologías da pauta a proponer nuevas herramientas preventivas, diagnósticas y terapéuticas (33), favoreciendo el uso de biomarcadores sensibles y específicos en el modelo de salud pública (34).

Biomarcadores

Como se ha mencionado previamente, uno de los puntos que pretende abarcar la MP es realizar medidas de prevención temprana, tomando en cuenta las particularidades biológicas y ambientales de cada persona; generando cuidados individuales (35). A partir de esta premisa, surge la búsqueda de marcadores biológicos, los cuales son moléculas o variaciones genéticas objetivas y evaluables (36) presentes tanto en procesos fisiológicos o patológicos (37), pudiendo ser un factor protector o de riesgo en el desarrollo de entidades nosológicas (38). En comparación a un síntoma, el cual suele ser subjetivo (35), estas herramientas deben ser objetivas, sensibles y específicas (37); además de proyectar un mejor entendimiento de las terapias individuales (39). La búsqueda de marcadores

biológicos presenta una combinación entre genotipo y fenotipo, integrando observaciones clínicas (36). De esta manera, al presentar evidencia suficiente pueden ser validados y utilizados en la clínica (35). Generando una visión funcional de la patogénesis, con un entendimiento del progreso de la enfermedad (36).

Actualmente, debido a la variabilidad de los procesos patológicos no se ha encontrado un biomarcador único relacionado con el desarrollo de una enfermedad. Sin embargo, basado en estas características heterogéneas se ha propuesto el uso de biopaneles, los cuales son un conjunto de biomarcadores (35) que toman en cuenta alteraciones de manera sistémica. A partir del uso de biomarcadores en el área de medicina de precisión se prevé un mejor entendimiento de la evolución de distintos procesos biológicos (36). En efecto, se espera que las nuevas medidas terapéuticas estén bajo el criterio de un grupo multidisciplinario. Asimismo, en la búsqueda de biomarcadores se remarca el uso de polimorfismos e un solo nucleótido, los cuales permiten el entendimiento de una enfermedad en relación a la información genética de cada persona (35).

Polimorfismos de un solo nucleótido

El genoma humano está conformado por aproximadamente 3 mil millones de pares de bases. Es esta información contenida en el núcleo celular la que le permite a cada individuo desarrollarse en su entorno. No obstante, cambios en esta serie de datos puede provocar alteraciones con un resultado negativo en la salud a través del tiempo (40). Las modificaciones más frecuentes que se han presentado a lo largo de la evolución son los polimorfismos de un solo nucleótido (SNPs por sus siglas en inglés) (41), teniendo una frecuencia de presentación del 90%, comparado con otras mutaciones que puedan establecerse (42)(43). Estas variantes se caracterizan por el cambio puntual de un nucleótido en el genoma en al menos el 1% de la población (44). Donde el canje del nucleótido puede o no generar repercusiones a nivel sistémico, asimismo puede derivar en un concepto de susceptibilidad o protección en el desarrollo de procesos nosológicos (45). El cambio de una base dependiendo de su posición puede resultar en alteraciones del marco abierto de lectura durante la traducción; reflejándose a nivel de proteína. Si el cambio del nucleótido tiene como consecuencia la inserción del aminoácido original, el polimorfismo se denomina sinónimo. Por lo general, este efecto es neutro y se genera cuando el canje se da en la tercera base del codón del marco de lectura (41). No obstante, los cambios del primer y segundo nucleótido tienen mayor probabilidad de generar sustituciones en el aminoácido insertado durante la traducción (46), estas variantes se denominan de sentido erróneo y su efecto sobre la estructura proteica dependerá de las características químicas del aminoácido insertado (47). Por su parte, las variantes sin sentido intercambian un codón de lectura por un codón de paro. De la misma manera, la alteración de un codón de paro que da parte a que continúe la traducción proteica, se conoce como SNP de sobre lectura (41). Igualmente, los polimorfismos no son propios de sitios exónicos, ya que pueden presentarse en regiones intrónicas con repercusiones en eventos de splicing. Asimismo, las regiones de inicio y fin de la traducción pueden verse modificadas, al igual que sitios potenciadoras o *enhancer* afectando la interacción

con las proteínas promotoras e iniciadoras responsables de la traducción (47). Además, los SNPs presentan variabilidad de frecuencia en cada grupo étnico (40), lo que se sugiere que el componente genético aditivo en el desarrollo de una enfermedad puede ser distinto en cada población (46)(48).

Estudio de asociación genómica completa

La relación de los SNPs con el desarrollo de entidades nosológicas se refleja mediante estudios de asociación genómica completa (GWAS por sus siglas en inglés) (49). Estos realizan la genotipificación y diversos análisis estadísticos para establecer una conexión entre los polimorfismos y fenotipos (50). Además de proponer variantes genéticas de riesgo, los GWAS incluyen información como frecuencia alélica, desequilibrio de ligamiento y tasa de recombinación (49)(51). A partir de la información derivada de los GWAS se ha propuesto la identificación de genes y vías con papeles específicos en la patogénesis de ciertas enfermedades (52), siendo un número limitado de cambios en el genotipo los responsables de un fenotipo en específico. Las variantes derivadas de estos genes presentan una alta prevalencia en las personas con la enfermedad de estudio (50). No obstante, debido a procesos de selección natural tienen una baja frecuencia en el resto de la población (52). Esto se ha definido como hipótesis omnigénica, y asume que existe una red de interacción de genes principales, que cuando presentan mutaciones o se pierde su función se aumenta el riesgo del desarrollo de enfermedades complejas (53). Como se ha mencionado previamente, los GWAS identifican polimorfismos asociados a una patología (49). Así como el conjunto de polimorfismos que se encuentran en desequilibrio de ligamiento con la variante de interés, siendo el punto de interés el identificar haplotipos de riesgo (54)(55)(56). Asimismo, mediante el proceso denominado imputación es posible estimar la participación de polimorfismos no genotipificados, sin que estén presentes en los estudios de secuenciación de la cohorte de estudio (51).

El catálogo público de GWAS (57) provee detalles de los estudios de asociación genómica completa de fenotipos en humanos. A la fecha se muestran 33 estudios para aterosclerosis (58)(59)(60)(61)(62)(57), y 3 para demencia vascular (63)(64)(65). En estos ensayos se exponen los polimorfismos relacionados a cada fenotipo, la localización de estas variantes en el genoma y la población en la que se realizó el estudio. No obstante, no se presenta una descripción del papel de las variantes genéticas en la fisiopatología de la enfermedad, así como el valor genético aditivo que representa cada polimorfismo en el sentido estricto de la heredabilidad para cada patología. Cabe destacar que no se ha publicado en este catálogo un estudio donde se haga referencia de manera conjunta a los polimorfismos presentes en demencia vascular y aterosclerosis. En cambio, se han puntualizado variantes en GWAS de aterosclerosis y evento cerebrovascular de tipo isquémico (66)(67). Empero, no enfatizan si los pacientes presentaban datos clínicos de demencia vascular o declive cognitivo.

Asimismo, se han identificado las estirpes celulares que participan en la patogénesis de estas enfermedades. Por ejemplo, ensayos de célula única detallan a las células

musculares lisas (68), linfocitos T CD4+ y macrófagos (69) como elementos principales en el desarrollo de placas ateroscleróticas no estables. Sin embargo, carecen de información en cuanto a los genes o polimorfismos que puedan tener un papel importante en el desarrollo de la enfermedad. De la misma manera, se han presentado artículos que muestran el papel de diferentes tipos celulares en el desarrollo de declive cognitivo (70)(71)(72). Por su parte, no se ha realizado un estudio con estas características en humanos con un enfoque que involucre la patogénesis de demencia vascular a partir de aterosclerosis. Hu et al (73) en 2021 refiere el papel de genes y vías metabólicas que participan en el inicio del declive cognitivo posterior a un evento isquémico. No obstante, este estudio se realizó utilizando modelos murinos. Es así, como se hace hincapié en la falta de estudios en humanos que describan el papel de los polimorfismos y tipos celulares en el desarrollo de demencia vascular a partir de aterosclerosis.

Desequilibrio de ligamiento

El desequilibrio de ligamiento (DL) describe la asociación no aleatoria de alelos en diferentes loci dentro de un cromosoma y la expectativa de heredarse en conjunto a la siguiente generación (74)(51)(75). De esta manera se denota un valor asociativo para cada conjunto de polimorfismos, cuya probabilidad tiende a ser distinta en cada grupo poblacional (76). Al conjunto de variantes presentes en una población se les ha denominado haplotipos (77). Esta agrupación de características genéticas dentro de una zona geográfica son el resultado de interacciones demográficas, selección natural y mutaciones (56)(78). Para definir el valor de DL entre dos SNPs y conocer si existe relación entre estas se utilizan las funciones D, D' y r². El valor de D para un par de polimorfismos se obtiene de la siguiente manera:

$$D_{AB} = p_{AB} - p_A p_B$$

Donde el DL para los alelos A y B (D_{AB}) está dado por la diferencia en la frecuencia con la que se presentan en conjunto ambas variantes (p_{AB}) y el producto de la frecuencia de estos alelos por separado ($p_A p_B$) (75). D_{AB} presenta un rango de -0.25 a 0.25, existiendo una relación proporcional con la probabilidad de que ambas variantes se presenten en conjunto en la siguiente generación. Sin embargo, este resultado puede verse influenciado por las frecuencias alélicas y el tamaño de la muestra de estudio (79). Para contrarrestar este sesgo se emplea la función de Lewontin o D' así como el valor de correlación r² (75). En el caso de D' se detalla la siguiente función utilizando la frecuencia de los alelos A y B:

$$D'_{AB} = \left\{ \begin{array}{l} \frac{D_{AB}}{\min(p_{AB}[1-p_B], [1-p_A]p_B)} \quad \forall D_{AB} > 0 \\ \frac{D_{AB}}{\max(p_A p_B, [1-p_A][1-p_B])} \quad \forall D_{AB} < 0 \end{array} \right\}$$

En primer lugar, se debe determinar si el valor de D_{AB} es menor o mayor a cero. Posteriormente se obtiene el valor mínimo o máximo, respectivamente para

normalizar el valor de D_{AB} con base en las frecuencias alélicas de los SNPs de estudio (p_A, p_B) (80). D' presenta un rango entre -1 a 1, siendo el valor más cercano a la unidad el que nos indica mayor relación entre las variantes (79). Por su parte, r^2 representa la correlación de las frecuencias alélicas (p_A, p_B) y se obtiene de la siguiente manera:

$$r^2 = \frac{D_{AB}^2}{p_A(1 - p_A)p_B(1 - p_B)}$$

Este coeficiente de correlación tiene un rango entre 0 y 1, siendo mayor la probabilidad para las variantes de heredarse en conjunto cuando el resultado es cercano a uno (75). Dentro de las ventajas de r^2 se describe la baja influencia del resultado ante la variabilidad de las frecuencias alélicas y el número de participantes en cada estudio, permitiendo realizar comparaciones de DL entre poblaciones (74)(79). De esta forma se desarrolla la identificación de patrones en el genoma, haciendo énfasis en la asociación presente entre cada haplotipo con un grupo étnico específico (81). En las figuras A1 y A2 en el apartado de anexo se muestran ejemplos de cómo el DL permite definir a dos variantes que se heredan en conjunto.

Estructura proteica y polimorfismos de un solo nucleótido

Las proteínas son macromoléculas conformadas por un número variable de aminoácidos unidos mediante un enlace peptídico. A nivel molecular se encuentran en la mayoría de los procesos biológicos que mantienen y replican la vida (82)(83). La organización proteica se describe en cuatro niveles: 1) la estructura primaria está conformada por el espacio de secuencia de aminoácidos que dan lugar a la proteína; 2) la formación de alfa hélices, hojas beta y plegamiento azaroso o sin estructura secundaria definida (*random coil*) se conoce como estructura secundaria; 3) la estructura terciaria comprende la organización tridimensional de las estructuras secundarias, formando un péptido y 4) la estructura cuaternaria comprende la asociación de múltiples péptidos (84)(85). Derivado de esta configuración tridimensional, las proteínas se ven inmersas en una variedad de funciones, por ejemplo el proporcionar soporte y estructura celular o favorecer la morfología tanto de sistemas unicelulares como en sistemas complejos. Asimismo, poseen capacidad catalítica, misma que da lugar a que los procesos biológicos se lleven a cabo (83)(85).

Dentro de un sistema biológico en específico se ha descrito que las proteínas no participan de manera aislada. De hecho, estas interactúan con otras proteínas, así como con membranas lipídicas, ácidos nucleicos, moléculas pequeñas o iones (86). A su vez, el conocimiento de la estructura tridimensional combinada con la información de la secuencia genera referencia sobre los sitios de actividad biológica (85). No obstante, esta variabilidad en los mecanismos donde la proteína se ve ensamblada no es un evento casual, sino el resultado de procesos de evolución y selección (84), dando parte a una conformación dinámica donde diferentes sitios pueden modificarse (83). Entre estos escenarios de regulación se incluyen cambios

alostéricos, interacciones intermoleculares y modificaciones postraduccionales (86). Es este mismo proceso de selección, el que favorece que tanto el genoma como el proteoma de los organismos sea activo (85). Igualmente, la serie de cambios en el genoma tienden a verse reflejados en la secuencia del marco abierto de lectura que dará parte a la formación de proteínas, generando cambios en la estructura primaria de estas. De esta manera, la función proteica se ve comprometida, llegando incluso a perderse (85). Es así como las variaciones alélicas como los polimorfismos de un solo nucleótido favorecen alteraciones en la estructura primaria, comprometiendo la estabilidad y función proteica (83). De esta forma, un mejor entendimiento sobre el plegado e interacción proteica en condiciones basales y patológicas proporcionará las bases de métodos más específicos para la detección oportuna de enfermedades complejas (87). En este punto la medicina de precisión sugiere utilizar una serie de clasificadores cuyo interés se base en delimitar si un polimorfismo en específico puede o no afectar el estado de salud de una persona (88)(83).

Definición de variantes asociadas a demencia vascular y aterosclerosis

Trabajos previos en el laboratorio de Biología de Sistemas y Medicina Traslacional adscrito a la Facultad de Nutrición de la Universidad Autónoma del Estado de Morelos, permitieron el desarrollo de un motor que realiza la búsqueda de polimorfismos de un solo nucleótido asociados a patologías mediante la minería de datos y minería de textos en bases de datos especializadas y curadas. Esta herramienta bioinformática se ha nombrado Selección de Variantes Automatizada (SVA v0.11) (Rivera *et al.* SVA v0.11 Núm. Reg. 03-2017-082911390800-01). La cual utiliza diagramas de flujo con palabras clave organizadas jerárquicamente de acuerdo con la descripción de un fenotipo de estudio. Incluyendo procesos patológicos y tipos celulares presentes en la construcción de un modelo nosológico. Estas palabras clave se encuentran bajo el vocabulario MeSH (Medical Subject Heading), lo cual regula que los términos empleados estén validados e indexados en el Centro Nacional para la Información Biotecnológica (National Center for Biotechnology Information, NCBI), obteniendo referencias biomédicas presente en el catálogo de la biblioteca nacional de medicina (U.S. NLM). Posteriormente, el motor se dirige a bases de datos especializadas y realiza la I) selección, II) preprocesamiento, III) transformación, IV) minería de datos y textos, V) interpretación y VI) evaluación de polimorfismos asociados a enfermedades de interés (89).

Una vez que se ha construido el fenotipo que describe a la patología y se han validado las palabras clave, el motor se dirige a la base de datos OMIM (Online Mendelian Inheritance in Man). Misma que es un compendio de genes humanos y fenotipos con actualización diaria y acceso libre. Esta base de datos divide sus entradas en cuatro grupos: 1) enfermedades mendelianas, 2) enfermedades complejas y susceptibilidad a cáncer, 3) variaciones benignas y 4) enfermedades somáticas celulares. Mediante el uso de OMIM y enfermedades complejas se obtienen los genes relacionados a las palabras clave del diagrama de flujo (90). A continuación, se discierne entre genes reguladores y aquellos que codifican para proteínas. Este paso se realiza al enlazar el resultado obtenido de OMIM con la

base de datos UniProt (Universal Protein Resource). La cual ofrece información de genes que codifican para proteínas. Mediante el apartado Swiss-Prot se recaba información proteómica curada y evaluada bajo métodos computacionales. Asimismo, se genera un código UniProt el cual permite identificar redundancia de genes. Además, se ofrece acceso a datos sobre topología, secuencia, función, estructura, y asociación de los genes con enfermedades (91). El siguiente paso consiste en obtener artículos de la literatura que respalden la relación de los genes encontrados con el desarrollo de la enfermedad de estudio. En este punto se utiliza el motor de búsqueda de PubMed, el cual usa la información generada previamente y de las palabras claves con el objetivo de encontrar artículos indexados en las diferentes revistas que forman parte del NCBI. De esta forma se extrae evidencia sobre la relación de la construcción del fenotipo patológico. A continuación, se identifica a los genes de mantenimiento. La función de estos se enfoca en la conservación de las funciones celulares basales. Por tal motivo, se espera que se expresen sin importar las condiciones de crecimiento, diferenciación o de mecanismos patológicos que pueda presentar la célula. Es decir, su presencia no se relaciona con la enfermedad de estudio (92). Una vez que se han identificado a los genes asociados a la patología, surge la pregunta sobre si estos permiten el progreso de la enfermedad, o si pueden prevenir los mecanismos patológicos. Este análisis se establece al utilizar ClinVar. Esta base de datos emplea los parámetros del Colegio Americano de Medicina Genética y Genómica (ACMG) para categorizar las variantes en los siguientes grupos: benigno, patogénico, de respuesta a fármacos, factor de riesgo, protector, fenotipos no patológicos y significado desconocido. Así, se conoce la naturaleza de las variantes genéticas relacionadas con el fenotipo (93). Posteriormente, el motor de búsqueda resalta los estudios GWAS que relacionan los polimorfismos con el genotipo y fenotipo de interés. También provee el valor de asociación (*valor p*) de cada SNP con la patología derivado del estudio de asociación genómica completa (57). A partir de estos datos la siguiente cuestión que responde el motor de búsqueda es el comportamiento de la frecuencia con la cual se presentan los SNPs en las diferentes zonas geográficas. Mediante el buscador de Ensembl se obtiene la frecuencia de este SNP en 26 grupos poblacionales, incluyendo información a nivel mundial o continental (94). Estos datos son obtenidos a través del estudio 1000 genomas fase 3 (95). Además de estas poblaciones se utilizan datos derivados del estudio Human Genome Diversity Project (HGDP) (96). Por último, el motor realiza la búsqueda de las vías metabólicas en las que se ven inmersas las variantes encontradas. En este punto hace uso de Reactome, la cual permite visualizar, interpretar y analizar las vías metabólicas humanas. Actualmente cuenta con referencias de 10,870 proteínas, 12,770 reacciones y 2,325 vías (97). Derivado de la información obtenida del SVA v0.11 se encontraron 77 SNPs asociados a demencia vascular (98). Por su parte, la búsqueda de variantes reportadas para aterosclerosis generó 665 resultados (89). Teniendo un total de 742 variantes en ambas patologías.

Pregunta de investigación

¿Es posible conocer la relación de las variantes genéticas asociadas a aterosclerosis y el desarrollo de demencia vascular?

Hipótesis

La progresión de la dinámica metabólica en estados fisiológicos bajo presiones de selección asociados a respuestas de daño por hipoxia dirigirá el umbral de la relación entre las variantes genéticas presentes durante el progreso de aterosclerosis que favorecen el desarrollo de demencia vascular.

Objetivo general

Estudiar el papel de variantes genéticas presentes en regiones codificantes relacionadas con el desarrollo de demencia vascular derivado de aterosclerosis en subpoblaciones de Europa, Asia, América y África.

Objetivos particulares

- 1) Definir variantes asociadas a ambas patologías.
- 2) Evaluar las variantes en sistemas celulares relacionados con demencia vascular.
- 3) Definir rutas de interacción de variantes mediante teoría de redes.
- 4) Evaluar desequilibrio de ligamiento de los polimorfismos asociados a demencia vascular y aterosclerosis.
- 5) Identificar la estructura tridimensional del producto de la traducción de variantes.

Metodología

Asociación de variantes

La demencia vascular es una enfermedad heterogénea ligada a distintos factores de riesgo (17). Se ha descrito que mecanismos que comprometen el flujo sanguíneo y el aporte de nutrientes como oxígeno y glucosa desencadenan el cuadro clínico (1). Asimismo, se propone que abordajes basados en el análisis de la información proveniente de bases de datos curadas pueden generar una propuesta para prevenir y detectar de manera oportuna esta enfermedad (17). Derivado del presente argumento, se utilizó la información obtenida de estudios previos en el laboratorio, sobre la búsqueda y selección de polimorfismos asociados a demencia vascular y aterosclerosis utilizando el motor de Selección de Variantes Aleatorias (SVA v0.11). Estos análisis están compuestos por 77 SNPs relacionados a demencia vascular (99) y 665 para aterosclerosis (89). A partir de esta información se establecieron los SNPs presentes en ambas patologías. Utilizando el modelo de intersección ($A \cap B$) se identificó a aquellos polimorfismos presentes en ambas enfermedades. A continuación con la información obtenida con los datos poblacionales del HGDP, se describió si las variantes reportadas en GWAS para aterosclerosis y demencia vascular están presentes en los grupos étnicos maya y pima; así como la frecuencia alélica de estos en cada zona geográfica. La visualización de la intersección de los datos estuvo dada mediante el uso de diagramas de Venn.

Selección de un modelo celular en el desarrollo de demencia vascular

Posterior a la identificación de los SNPs se evaluó la relación de éstos con las células que conforman la UNV. Esto mediante una herramienta desarrollada en el laboratorio, la cual se denomina SNP-Cell Classification and Risk Assessment of Dementia (SCRAD v0.9) (98). Ésta realiza la asociación entre el polimorfismo y el tipo celular en función de minería de textos. SCRAD se basa en el desarrollo de un modelo de predicción a partir de la búsqueda de artículos en PubMed para cada tipo celular de la UNV. Durante el análisis de esta información se eliminaron las palabras de unión, así como artículos determinados. A continuación, por medio del algoritmo XGBoost (Extreme Gradient Boosting) (100) el cual está conformado por un complejo de árboles de decisión, se generó un valor de asociación entre las palabras presentes en el texto con cada célula del sistema nervioso central. A través de estos datos se proveyó un valor de predicción entre una palabra al azar y un tipo celular de la UNV. Posterior al entrenamiento del modelo se ingresaron los resúmenes PMID de los genes con los cuales se deseaba realizar la asociación sobre la UNV. A continuación SCRAD calculó la relación de este término y el tipo celular mediante bosques aleatorios. Generando un valor denominado *cell score* con un rango entre 0 y 1. Este resultado fue proporcional a la relación del gen con una estirpe celular de la UNV. No obstante, este algoritmo se encuentra limitado por la información depositada en PubMed y el número de artículos destinados a cada tipo celular. Asimismo, solo funciona con los componentes de la UNV, desestimando

la probabilidad de presentarse en otro estirpe celular fuera del sistema nervioso central.

Redes de interacción

La teoría de redes o interactoma aplicada en la biología de sistemas genera un panorama sobre cómo los elementos presentes se comunican unos con otros, facilitando el entendimiento de procesos ecológicos, evolutivos y patológicos (101). Las redes de interacción son una pieza clave para abordar mecanismos biológicos (102), ofreciendo una posibilidad de comprender la organización de los procesos metabólicos que forman parte de los sistemas celulares (101). Existen dos componentes básicos en las redes de interacción: los nodos y los enlaces. Los nodos representan a los integrantes de un evento particular. Por ejemplo, genes, proteínas o tipos celulares. Por su parte, los enlaces muestran la interacción de los nodos unos con otros (103). Asimismo, basado en la naturaleza de los datos podemos dividir al interactoma en homogéneo o heterogéneo. Es decir, si se realiza una red compuesta solo por un tipo de variante se habla de una red homogénea. Por su parte, la red heterogénea está compuesta por elementos de distinta naturaleza. En otras palabras, una red conformada por información genética, traducción de proteínas, estirpes celulares, vías metabólicas, etc. De la misma manera, la teoría de redes permite relacionar diferentes interactomas entre sí en caso de que éstos tengan elementos en común. Esto favorece el desarrollo de redes bipartitas o tripartitas al interaccionar dos o tres conjuntos de redes respectivamente (104). La red de interacción que se generó en este proyecto utilizó información sobre los polimorfismos presentes en el desarrollo de la demencia vascular a partir de aterosclerosis. Siendo esta una red heterogénea conformada por genes, SNPs, vías metabólicas y estirpes celulares.

Para determinar la importancia relativa de cada nodo dentro de la red se utilizaron medidas globales y locales (105). Las medidas locales definen la influencia por parte de cada nodo con aquellos que se encuentran en su periferia. Asimismo las medidas globales indican el efecto de un nodo en relación a toda la red (106). El grado de centralidad, es una medida local, la cual identifica el número de enlaces que salen o entran en un nodo en específico (105). Esta medida está definida por la siguiente función:

$$Centralidad = \frac{\sum a(i,k)}{n}$$

Donde $a(i,k)$ corresponde al número de enlaces que inciden y salen del nodo de estudio. Por su parte, n compete al total de enlaces presentes en la red. A partir de esta medida, se reflejaron los elementos con mayor interacción dentro de la red (107). De hecho, aquellos nodos con una centralidad igual o mayor al valor de la media aritmética más dos veces la desviación estándar se denominaron nodos centros de red (101). Otra de las métricas utilizadas para evaluar el comportamiento de los nodos en el sistema fue la centralidad de intermediación (CI), la cual es una medida global que identifica a un elemento que funciona como conector entre dos

nodos (106). De esta forma se reconocen a los nodos que funcionan como puente de comunicación dentro de la red (105). Esta medida está descrita por la siguiente función:

$$CI = \sum \frac{\sigma_{st}(V)}{\sigma_{st}}$$

Donde σ_{st} es el número total de vías entre el nodo s y t . Mientras que $\sigma_{st}(V)$ es el total de rutas que existen a través de V (nodo de interés). Existe una variación de esta medida, que reconoce a aquellos nodos que al ser eliminados de la red tienen mayor probabilidad de generar colonias de nodos aislados. Estos se han denominado cuellos de botella y se caracterizan por tener un valor normalizado de centralidad de intermediación por encima de 0.2 (102). De esta manera, a partir de la metodología de redes de interacción se identificó aquellos elementos con importancia relativa en el desarrollo de demencia vascular.

Los valores de grado de centralidad y centralidad de intermediación para cada elemento de la red se determinaron mediante la paquetería *igraph* (v 1.2.7) (108), utilizando las funciones *degree(normalized = TRUE)* y *betweenness(normalized = TRUE)*. De esta manera se normalizaron los resultados, permitiendo identificar las métricas para nodos centros de red y cuellos de botella. Posteriormente, la visualización de los datos se realizó con la paquetería *gephi* (v0.9.2) (109). Para este gráfico el tamaño del nodo fue proporcional al valor de centralidad de intermediación, mientras que el tono de cada elemento fue distributivo acorde al grado de centralidad.

Desequilibrio de ligamiento de variantes propuestas

Las enfermedades crónico-degenerativas son el resultado de la interacción de factores ambientales y genéticos. A partir de los estudios de asociación de genoma completo es posible delimitar la participación de polimorfismos específicos relacionados a estas entidades nosológicas. No obstante, se ha descrito que ciertas variantes tienden a heredarse en conjunto, generando haplotipos de riesgo que dan parte a que determinados individuos tengan mayor o menor susceptibilidad hacia una enfermedad (77)(78)(79). Esta información se obtiene mediante el proceso denominado desequilibrio de ligamiento (DL) (51)(74)(75). Debido a que la mayoría de los GWAS se basan en microarreglos los cuales contemplan una fracción del genoma, el método de imputación amplía el número de variantes asociadas al describir el bloque de polimorfismos que tienden a heredarse en conjunto en una población (110). De esta manera se recupera información que se puede omitir por características metodológicas (111). Además, es necesario un panel de referencia para identificar estos haplotipos en cada población de estudio (112). Para definir la serie de polimorfismos que se encuentran en DL con los SNPs presentes en GWAS

de demencia y aterosclerosis se utilizó la paquetería de Rstudio seekerBio (98) y LDlink (<https://ldlink.nci.nih.gov/?tab=home>) del Instituto Nacional de Cáncer. Ambas paqueterías utilizan información de la base de datos Ensembl y el proyecto 1000 genomas fase 3 (98)(113). Asimismo, se empleó como genoma de referencia el ensamble 38 parche 13 del consorcio del genoma humano (GRCh38.p13) (114). Durante el desarrollo de este proyecto se evaluaron los grupos de polimorfismos presentes en 12 subpoblaciones. Los sitios geográficos se enlistan a continuación:

Nombre de la población	Abreviatura
Japoneses en Tokio, Japón	JPT
Chinos Han en Beijing, China	CHB
Punjabíes de Lahore, Pakistán	PJL
Finlandeses en Finlandia	FIN
Inglés en Inglaterra y Escocia	GBR
Ibéricos en España	IBS
Luyias en Webuye, Kenia	LWK
Yorubas en Ibadan, Nigeria	YRI
Americanos con ascendencia africana en suroeste de Estados Unidos de América	ASW
Ancestría mexicana en Los Ángeles, Estados Unidos de América	MXL
Peruanos en Lima, Perú	PEL
Colombianos en Medellín, Colombia	CLM

Tabla 1. Poblaciones 1000 genomas fase 3.

Para determinar los polimorfismos heredados en bloque se utilizaron los siguientes parámetros en seekerBio (98). Cada polimorfismo se estudió en una ventana de 500 mega bases, debido a que un rango mayor de bases aumenta la probabilidad de que los resultados obtenidos sean falsos positivos (115). Se seleccionaron aquellas variantes con un coeficiente de correlación igual o mayor a 0.6. Posteriormente, mediante información de la base de datos Ensembl y LDlink se describió la arquitectura del cromosoma en el que se encuentra el grupo de polimorfismos (94). De esta manera, se identificaron las variantes en DL con los polimorfismos asociados a las patologías de interés. Una vez construidos los haplotipos se identificaron aquellos polimorfismos cuyo efecto puede verse reflejado de manera directa en la estructura proteica. Mencionando como ejemplo las variantes de sentido erróneo (*missense*), sin sentido (*non-sense*), de sobre lectura (*stop lost*) o sinónimos (*synonymous*). Una vez que se reconocieron los SNPs con impacto estructural se utilizó la información depositada en las bases de datos de dbSNP (data base SNP) y del proyecto CCDS (<https://www.ncbi.nlm.nih.gov/projects/CCDS/CcidsBrowse.cgi>) para identificar el codón y residuo de afectado. De esta manera, se determinó la consecuencia de la variante sobre la estructura proteica primaria.

Asociación de polimorfismos por grupo poblacional y tipo de variante

Posterior al proceso de imputación mediante desequilibrio de ligamiento, se realizó la agrupación de los SNPs acorde al tipo de variante (sinónimo, de sentido erróneo, sin sentido, de sobre lectura, etc.) y población. El proceso de agrupación de los datos se efectuó mediante el coeficiente de similitud de Gower y la incrustación de vecinos estocásticos distribuidos en T (t-SNE). El coeficiente de similitud de Gower permite encontrar semejanzas dentro de un conjunto de datos conformado por elementos de tipo carácter. Esta métrica se observa como un promedio de las similitudes de las variables donde el resultado se encuentra en un rango entre 0 y 1, siendo proporcional este valor a la relación de similitud de las características (116)(117). El cálculo de similitud se realizó de la siguiente manera:

$$S_{ijk} = 0$$

Donde se otorga un valor de 0 si las variables i , j son idénticas para la característica k . Dicho de otra manera, si un polimorfismo se presentó en dos poblaciones diferentes, se asignó un valor de 0 a este conjunto. Por el contrario, cuando las variables i , j difieren se otorga un valor de 1 (116). Para obtener el valor de similitud entre grupos poblacionales y polimorfismos se utilizó la función *daisy* correspondiente a la paquetería *cluster* (versión 2.1.2) (Maechler M, 2021), dentro del lenguaje R. A continuación se dio parte a la construcción de los grupos derivado del valor de similitud entre variables. Este proceso se definió a partir del modelo de partición de medoides (PAM). En este punto se toma una variable como referencia para la construcción de cada conjunto, organizado desde los vecinos con menor diferencia. Esta variable central de la cual se construye cada grupo es conocido como medoide (118). De esta manera se buscó reconocer aquellas poblaciones con mayor similitud en cuanto a polimorfismos reportados. Posteriormente, para distribuir y graficar los datos mediante incrustación de vecinos estocástica distribuidos en T se utilizó la paquetería *Rtsne* (versión 0.15) (Krijthe J 2018). De esta forma se identificaron las poblaciones que tienden presentar los mismos polimorfismos.

Evaluación de estructura secundaria y efecto de los polimorfismos de un solo nucleótido

En apartados previos se describió el papel que presenta la secuencia de aminoácidos y la estructura tridimensional proteica en procesos biológicos que permiten el equilibrio en un organismo (85)(119). No obstante, cambios en el genoma como los polimorfismos de un solo nucleótido tienden a provocar alteraciones en el marco de lectura que da parte a la síntesis de proteínas (83). Estas variaciones presentan la probabilidad de generar un impacto en la estructura proteica; existiendo ganancia, disminución o pérdida de la función (88). De la misma manera, factores fisicoquímicos como solubilidad, estabilidad y conformación tridimensional se ven comprometidos durante estos sucesos (87). Esto da como resultado una serie de eventos cuya culminación son el desarrollo de diferentes enfermedades. Por tal motivo, es fundamental entender cómo los polimorfismos

alteran la homeostasis de las proteínas (119). A partir de este planeamiento se presenta la necesidad de evaluar la estabilidad de las estructuras proteicas ante la sustitución de residuos en la secuencia de aminoácidos. Este ensayo constó de dos partes, durante la primera fase se evaluó la estructura secundaria de las proteínas asociadas a los polimorfismos presentes en la demencia vascular y aterosclerosis identificando en que región se encontraba el residuo afectado y que tipo de estructura secundaria (alfa hélice, plegamiento azaroso) se veía comprometida. Dicha información se obtuvo de GOR IV (<https://prabi.ibcp.fr/htm/site/web/home>) y PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>). La parte complementaria consistió en establecer si el cambio puntual puede o no ser tolerado en término de conservación de residuos dentro de la estructura primaria, esto a partir de los datos generados en función de las herramientas SIFT (<https://sift.bii.a-star.edu.sg/>), PolyPhen2 (<http://genetics.bwh.harvard.edu/pph2/>) y PROVEAN (<http://provean.jcvi.org/index.php>). El conjunto de estos datos amplió el panorama sobre la patogénesis o protección que pudiesen representar los SNPs presentes en ambas entidades nosológicas. A continuación se describirá cada herramienta bioinformática, así como los métodos utilizados para clasificar los resultados.

Comenzaremos con la descripción de GOR IV, nombrado así por las iniciales de sus autores (Garnier, Osguthorpe, Robson. 1978). Este algoritmo utiliza un conjunto de datos conformado por información teórica de 267 ensayos de cristalografía de proteínas con resolución menor de 2.5 Å y datos sobre la interacción de los residuos que las constituyen, con el objetivo de generar un modelo basado en estadística Bayesiana que determine el tipo de estructura secundaria que se presentaría a partir de una secuencia de aminoácidos (120)(121)(122). En comparación a otros métodos matemáticos donde cada fenómeno se estudia de manera independiente, el teorema de Bayes toma en cuenta el comportamiento de la probabilidad en eventos pasados del conjunto de datos que conforman el experimento (123). A partir de esta hipótesis GOR IV emplea la siguiente función para calcular la probabilidad de encontrar una conformación en específico para cada residuo en la secuencia de estudio $I(S;R)$:

$$I(S; R) = \log \left[\frac{P(S|R)}{P(S)} \right]$$

Se calcula el logaritmo de la probabilidad de encontrarse la conformación S (la cual puede ser alfa hélice, hoja beta, plegamiento azaroso) dado el residuo R (siendo R uno de los 20 posibles aminoácidos). Y la probabilidad de $P(S)$, es decir la ocurrencia de la conformación S . Posteriormente, se evalúa a los residuos vecinos, basado en la premisa de que la conformación estructural no depende de un solo aminoácido, sino de los residuos próximos en la secuencia. En este aspecto se determinó la información de 8 residuos río arriba y río abajo. A partir de la información del segmento conformado por los 17 residuos se compararon los datos con el set de 267 proteínas y se predijo el estado conformacional de cada segmento en la secuencia. En función de este modelo el usuario coloca en el servidor de GOR IV la serie de aminoácidos que conforman la proteína de estudio como datos de

entrada, a continuación se procesa esta secuencia y el resultado consta en la estructura secundaria que derivaría en cada segmento (120).

Por su parte el modelo utilizado por PSIPRED se basa en un proceso de aprendizaje de máquina para determinar la estructura secundaria correspondiente de cada segmento de la secuencia de aminoácidos. Este algoritmo se conforma por una red neuronal de retro propagación y retroalimentación. Brevemente esta herramienta se basa en el uso de datos de entrada, los cuales son procesados por las capas que conforman la red, generando datos de salida. A continuación se calcula el error derivado de los resultados. Posteriormente, se corrigen los pesos de las capas neuronales y se ajustan estos mismos parámetros. Nuevamente, se evaluaron los datos de entrada con las correcciones en los pesos de las capas y se generaron los nuevos datos de salida. El conjunto de la información utilizada para entrenar y validar el modelo estuvo conformado por la información de 187 dominios proteicos con una resolución menor de 1.8 Å. Asimismo se agregan características de 340,000 secuencias no redundantes. Estos permitieron identificar el comportamiento de un conjunto de residuos en cada una de las estructuras secundarias que conforman cada proteína. Una vez entrenado el modelo, el usuario coloca la secuencia de aminoácidos de su interés. PSIPRED compara la información de esta secuencia con la derivada de su set de entrenamiento y define la estructura secundaria correspondiente a cada segmento de la secuencia. De manera similar a GOR IV, se valoró una ventana de 7 residuos río arriba y 7 río abajo, estimando un total de 15 residuos. A partir de estos datos, los resultados de salida presentaron la secuencia de aminoácidos y la conformación de la estructura secundaria correspondiente (124).

A continuación se describirán las herramientas bioinformáticas SIFT, PolyPhen2 y PROVEAN, las cuales proporcionaron información sobre si el cambio de residuos derivados de cada polimorfismo presenta o no un efecto sobre la proteína. Cabe resaltar que estas herramientas solo permiten la evaluación de variantes no sinónimas, por lo que el experimento se centró en polimorfismos que modifican el marco abierto de lectura. SIFT (*Sorting Intolerant from Tolerante*) predijo el efecto de la sustitución de un aminoácido en la secuencia de la proteína blanco, con base en alineamiento por homología (125). El antecedente de este algoritmo deriva en que residuos importantes para la estructura y función de una proteína se conservan a lo largo del tiempo y diversos procesos evolutivos. Para poder evaluar este proceso, SIFT necesita de una secuencia de aminoácidos así como la posición y el residuo modificado como datos de entrada, posteriormente se dirige a la base de datos UniProt y realiza la búsqueda de secuencias homólogas y realiza un alineamiento de estas. A continuación, calcula la probabilidad de que el residuo marcado sea sustituido por cada uno de los 20 aminoácidos existentes. Las probabilidades calculadas se normalizaron en relación con la probabilidad del aminoácido más frecuente (126). El resultado obtenido se denomina SIFT score y presenta un rango entre 0 y 1. Donde aquellos residuos con un valor menor a 0.05 se consideraron como cambios no tolerantes. Por su parte, los resultados iguales o mayores a 0.05 se determinaron como cambios tolerantes (125).

Por su parte PolyPhen2 (*polymorphism phenotyping v2*) (125) se basa en la probabilidad de que un residuo sea sustituido, derivado de la información de alineamiento por homología y comparación con un modelo entrenado a partir de características de secuencia y estructura. El fundamento de este algoritmo radica en la información de dos bases de datos HumDiv y HumVar. HumDiv se compone de 3,155 alelos dañinos con referencia en UniProt reportados en patologías y 6,321 alelos reportados como no dañinos. Por su parte HumVar se conforma de 13,032 mutaciones asociadas a enfermedad presentes en UniProt y 8,946 polimorfismos no sinónimos sin relación a una patología, los cuales se consideran como no dañinos. Sin embargo, este último aspecto genera un sesgo, ya que el hecho de no tener información reportada sobre una variante no convierte a la misma en un evento neutral. Debido al sesgo que genera HumVar, en este proyecto solo se utilizó la información derivada de HumDiv para predecir el efecto del polimorfismo sobre la proteína. Para clasificar los datos se utilizaron 8 características de secuencia y 3 estructurales (ver anexo A3 para mayor descripción). A partir de este proceso PolyPhen2 calculó el efecto del cambio de un residuo en una posición en específico en la secuencia proteica. En primer lugar, se generó una búsqueda de secuencias homólogas en relación con la secuencia de estudio, se realizó un alineamiento de éstas y se calculó la probabilidad de que el sitio marcado estuviera sustituido. Posteriormente y con base en un modelo de *naive* Bayes se calculó el impacto de esta sustitución a partir de la información del conjunto de datos de HumVar y HumDiv (127). Posterior al análisis de la información, PolyPhen2 se generaron dos resultados; uno calculado con los datos de HumDiv y otro derivado de HumVar. Ambos presentan un rango entre 0 y 1, haciendo tres puntos de corte para clasificar los resultados. Considerando a una variante con un valor entre 0.0-0.15 como benigna. Por su parte, resultados mayores a 0.15 y menores a 0.85 fueron tomados como probablemente dañinos. Por último, un puntaje de 0.85 a 1 describió a un cambio dañino (125).

De manera similar, PROVEAN (*Protein Variation Effect Analyzer*) (125) se basa en la comparación de secuencia homólogas y detección de residuos conservados. No obstante, a diferencia de los dos algoritmos previos permite identificar el efecto de cambios puntuales y mutaciones indel (inserciones/deleciones). El primer paso de esta herramienta es la búsqueda y alineamiento de secuencias homólogas a la secuencia de estudio. Posteriormente se determinó el grado de variabilidad de cada segmento, para este punto se definió el puntaje delta, mismo que se representó con la siguiente función:

$$\Delta(Q, v, S) = A(Q', S) - A(Q, S)$$

Donde el cambio en una secuencia en relación con un residuo específico dado una secuencia homóloga ($\Delta(Q, v, S)$), fue igual a la diferencia del puntaje de la secuencia de estudio ($A(Q', S)$) y la secuencia homóloga ($A(Q, S)$). El puntaje de cada secuencia derivó de la matriz de sustitución BLOSUM62 (Figura 3) (128), conformada por el logaritmo de la probabilidad de que un residuo sea sustituido en una secuencia. Por ejemplo, podemos definir que cuando una cisteína se sustituye por una alanina se provee un valor de 0. Por su parte, si la sustitución de esta cisteína hubiese sido por

versión de R 4.1.2 bajo el sistema operativo Linux (plataforma x86_64-redhat-linux-gnu). Por último, se remarca que todas las paqueterías fueron utilizadas bajo los argumentos originales.

Resultados

Variantes presentes en demencia vascular y aterosclerosis

En la figura 4 se observa que al presentarse los 665 polimorfismos de aterosclerosis y 77 de demencia vascular, existe una intersección de 12 variantes; lo que genera tres grupos de estudio. El sistema con mayor número de datos se conforma por polimorfismos propios para aterosclerosis, presentando 653 variantes. Enseguida, el conjunto de demencia vascular se estructura por 65 SNPs únicos. Esto ocasiona que el tercer grupo presente 12 polimorfismos. Es esta superposición la que indica las variantes presentes en ambas patologías. Por tal motivo, esta cohorte de polimorfismos se mostrará a lo largo del proyecto para validar la hipótesis planteada.

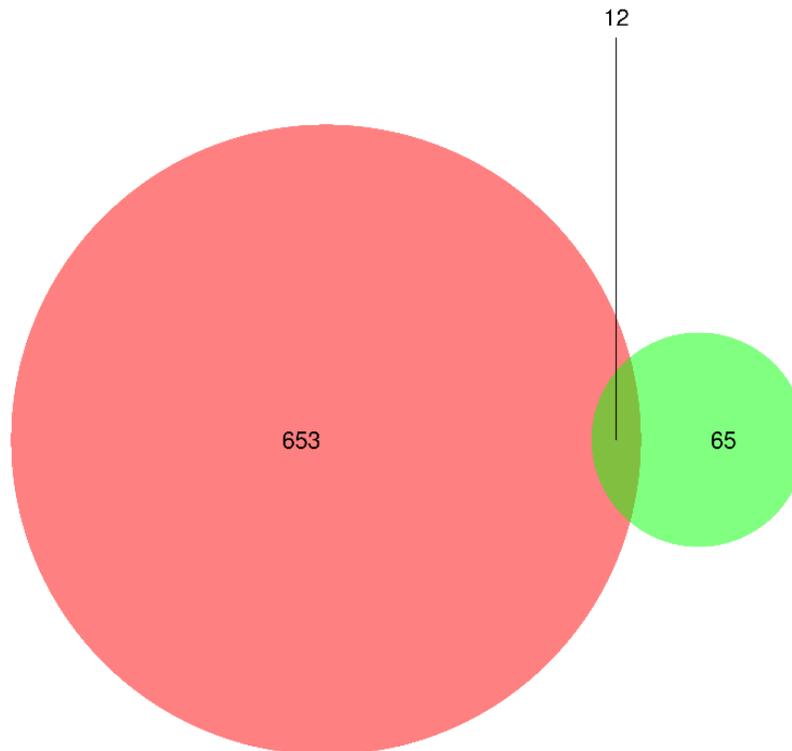


Figura 4. Diagrama de Venn conformado por polimorfismos de demencia vascular y aterosclerosis. En rojo, se muestran los SNPs para aterosclerosis. En verde se encuentran las variantes para demencia vascular. El área conjunta de ambos polígonos se conforma por 12 polimorfismos.

Asimismo, se agregaron dos sistemas dentro del estudio de conjuntos en la continuidad del análisis de la prevalencia de polimorfismos asociados a demencia vascular y aterosclerosis. Estos están conformados por datos de secuenciación por microarreglo de poblaciones pima y maya. Información obtenida de Human Genome diversity Project. En la figura 5 se observa la conjunción de los cuatro

sistemas, generando subgrupos con características en común. No obstante, manteniendo los objetivos del proyecto se destacan dos asociaciones. Existe un polimorfismo compartido dentro de los cuatro sistemas. Es decir, hay relación de un SNP presente en ambas enfermedades y los dos grupos étnicos. De la misma manera, se denota una variante para demencia vascular, aterosclerosis y población maya. Esta última no fue reportada en población pima.

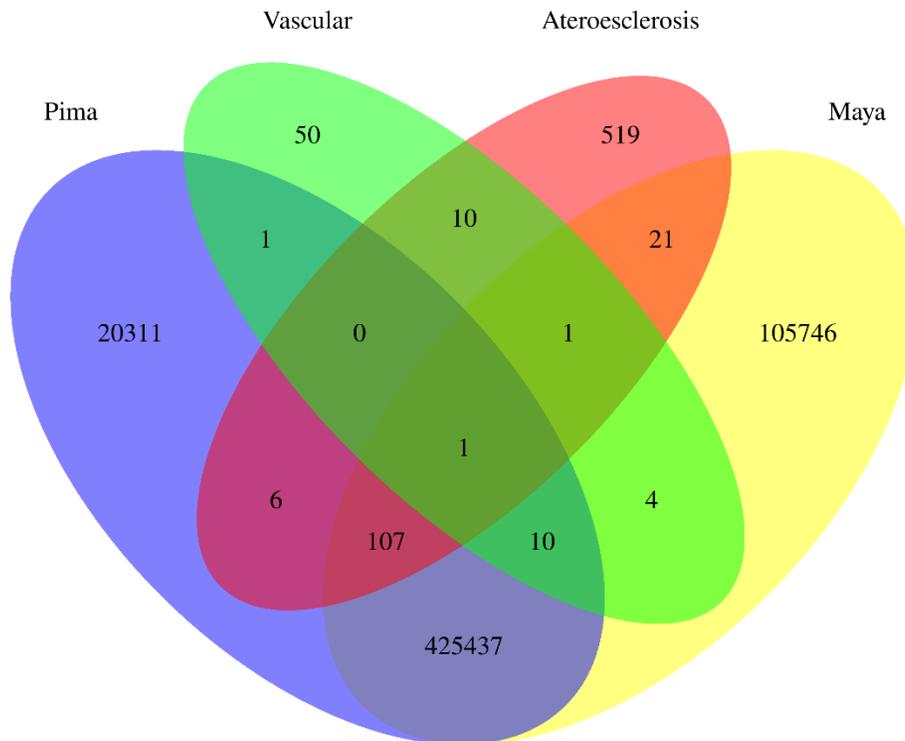


Figura 5. Diagrama de Venn conformado por polimorfismos derivados de los estudios de demencia vascular, aterosclerosis y Human Genome Diversity Project. En azul se representan los SNPs reportados en población pima, en verde las variantes para demencia vascular, en rojo los polimorfismos para aterosclerosis y en amarillo los SNPs de población maya. Se observa que los cuatro sistemas se interceptan en un solo polimorfismo

Tipo de variantes

La siguiente determinación consiste en identificar los genes y polimorfismos presentes en la intersección de los sistemas. La tabla 2 muestra en la primer columna el número de identificación para cada variante seguido del nombre del gen asociado (columna 2). La tercera serie de datos detalla la región donde se presentó la variante. Al continuar con la descripción de los hallazgos, se destacan los polimorfismos rs556621 y rs11984041. El primero se presenta tanto en poblaciones maya y pima y corresponde a una región intergénica entre los genes CC5L y

SUPT3H en el cromosoma 6. Por su parte, la segunda variante resaltada se asocia al gen HDAC9 y esta última, solamente se encuentra en población maya.

Polimorfismos	Gen	Localización
rs964184	ZPR1	Región 3' UTR
rs556621**		Región intergénica
rs3176471	CDKN2C	Región intrónica
rs17612742	EDNRA	Región intrónica
rs138364069	FAF1	Región intrónica
rs2107595	HDAC9	Región reguladora
rs11984041*		Región intrónica
rs72983521		Región intergénica
rs2417957	SLCO1B1	Región intrónica
rs7610618		Región intergénica
rs9982601		Región intergénica
rs12122341	TSPAN2	Región reguladora

Tabla 2. Tipo de polimorfismo de las variantes asociadas a demencia vascular y aterosclerosis.

**Polimorfismo presente en población maya y pima. *Polimorfismos presente en población maya.

Localización cromosómica

El determinar la localización de las variantes en el genoma favorece la identificación de características fenotípicas específicas o cuya herencia esté ligada al sexo. En la figura 6 se observa que los polimorfismos tienen una distribución variable en el genoma, presentándose en diferentes cromosomas. Siendo el cromosoma 1 el que cuenta con mayor número de variantes. Le siguen los cromosomas 7 y 11 con dos SNPs cada uno. Por último, los cromosomas 3, 4, 6, 12 y 21 contienen un polimorfismo cada uno. De esta forma se describe la participación de cromosomas somáticos en la patología, así como la ausencia de polimorfismos asociados a cromosomas sexuales. El eje vertical de este gráfico posee una escala basada en el logaritmo negativo del valor P o valor de asociación. La relación de esta notación es proporcional a la probabilidad de que la variable se presente en la enfermedad que fue reportada mediante su estudio de asociación genómica.

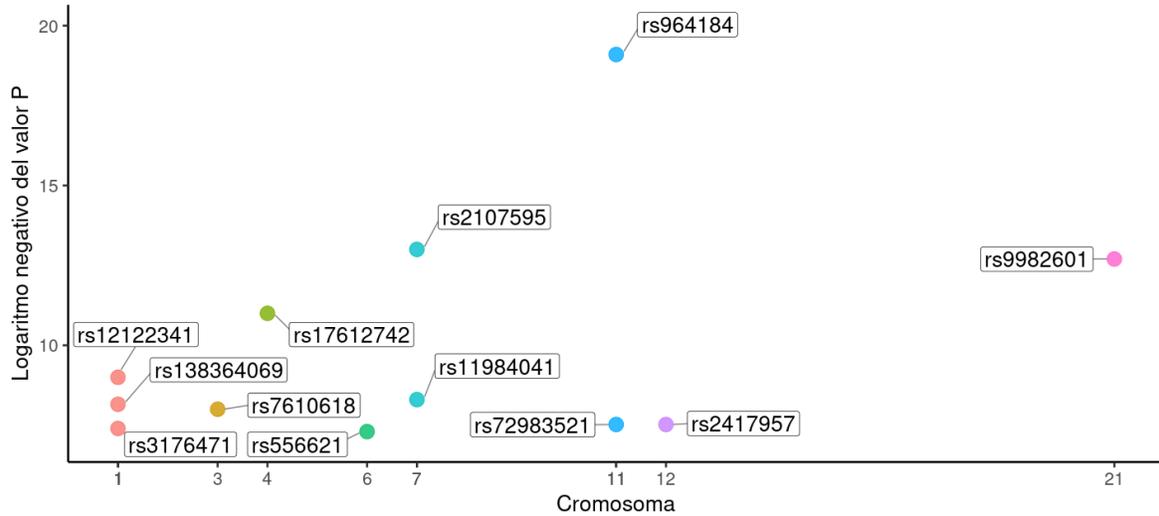


Figura 6. Localización cromosómica. El eje de las equis representa el número del cromosoma, mientras que el eje vertical muestra el logaritmo negativo del valor P. Cada polimorfismo está representado por un punto, donde se observa que las variaciones genéticas se encuentran en cromosomas somáticos.

Frecuencia poblacional

Para evaluar el comportamiento de la frecuencia del grupo de polimorfismos asociados a demencia vascular y aterosclerosis se empleó la frecuencia alélica menor o frecuencia de riesgo reportada en 12 subpoblaciones del proyecto 1000 genomas fase 3. Haciendo una excepción en las variantes rs556621 y rs11984041, donde se agregó información de población pima y maya del HGDP. La figura 7 muestra la variabilidad de la frecuencia en relación con la zona geográfica de estudio. Existen contrastes donde las frecuencias poblacionales tienden a agruparse de manera similar, como en los SNPs rs3176471 y rs138364069. Por otro lado, se ejemplifican eventos donde el mismo polimorfismo muestra variabilidad en la frecuencia de presentación de cada grupo étnico. Por ejemplo el SNP rs556621, el cual se encuentra ampliamente representado en poblaciones amerindias, así como la cohorte conformada por pobladores de Japón y China. Mientras que en grupos de ancestría africana, la frecuencia de esta variante se encuentra por debajo de 0.1. Es así como se denota que los polimorfismos presentan variaciones de frecuencia para cada región.

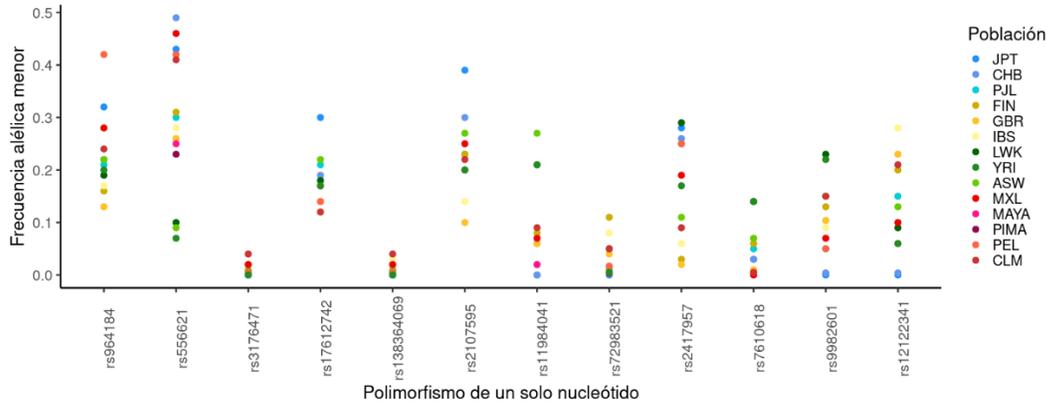


Figura 7. Frecuencia poblacional de polimorfismos asociados a demencia vascular y aterosclerosis. Cada punto representa una población, donde la distribución de los datos a través del gráfico está dada por la frecuencia alélica menor. JPT: japoneses en Tokio, Japón; CHB: chinos Han en Beijing, China; PHL: punyabíes en Lahore, Pakistán; FIN: finlandeses en Finlandia; GBR: ingleses en Inglaterra y Escocia; IBS: ibéricos en España; LWK: luyias en Webuye, Kenia; YRI: yorubas en Ibadan, Nigeria; ASW: americanos con ancestría africana en el suroeste de Estados Unidos de América; MXL: ancestría mexicana en Los Ángeles, Estados Unidos de América; PEL: peruanos en Lima, Perú; CLM: colombianos en Medellín, Colombia.

Asociación de polimorfismos con estirpes celulares

Posterior al determinar la distribución poblacional de cada variante, el siguiente paso es precisar la asociación del polimorfismo con un estirpe celular de la UNV a partir de artículos publicados en PubMed. SCRAD v0.9 responde esta cuestión mediante el uso de minería de textos, la figura 8 refleja la probabilidad de que cada variante se relacione con los componentes de la UNV. De la misma manera se detalla que este análisis solo puede realizarse con los SNPs presentes en regiones codificantes, por tal motivo se omiten las variantes sin asociación al marco abierto de lectura. A partir de estos datos se muestra que los polimorfismos tienden a asociarse con predominio hacia los oligodendrocitos y astrocitos.

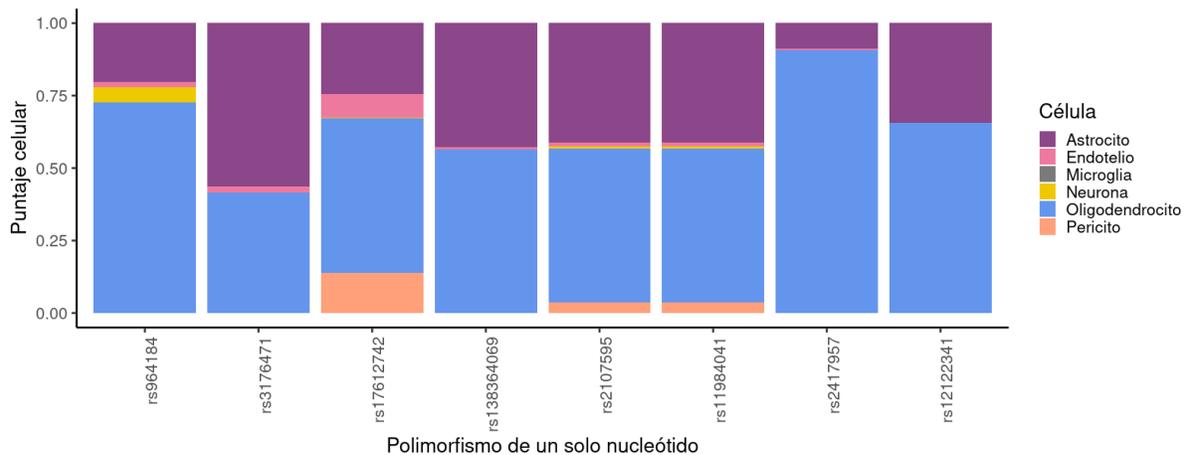


Figura 8. Polimorfismos y estirpes celulares. Se muestra la probabilidad de asociación de cada variante con un tipo celular de la unidad neurovascular. Los SNPs tienden a relacionarse con oligodendrocitos (azul) y astrocitos (morado). Existe una menor probabilidad de presentarse en microglía (gris), neuronas (amarillo) endotelio (rosa) o pericitos (salmón).

Vías metabólicas asociadas a demencia vascular y aterosclerosis

La determinación del evento nosológico se acompaña de las vías metabólicas que tienden a presentar alteraciones durante la disrupción de los procesos fisiológicos. Mediante la información de Reactome, se establecen las vías celulares asociadas a las variantes genéticas, desarrollando un panorama de las probables alteraciones durante el inicio de una enfermedad. En la figura 9 se observan las vías metabólicas asociadas a cada polimorfismo. En contraste, las variantes rs556621, rs72983521, rs7610618 y rs9982601 no cuentan con una vía descrita en Reactome al no estar presentes en un gen. En el caso de rs12122341 si bien se ha asociado al gen TSPAN2, es este último el que no cuenta con información presente en la base de datos de Reactome. Por tal motivo, estos polimorfismos se encuentran en la categoría “sin asociación”. No obstante, se continuará utilizando la información de estas variantes para los ensayos posteriores que se muestran en este escrito.

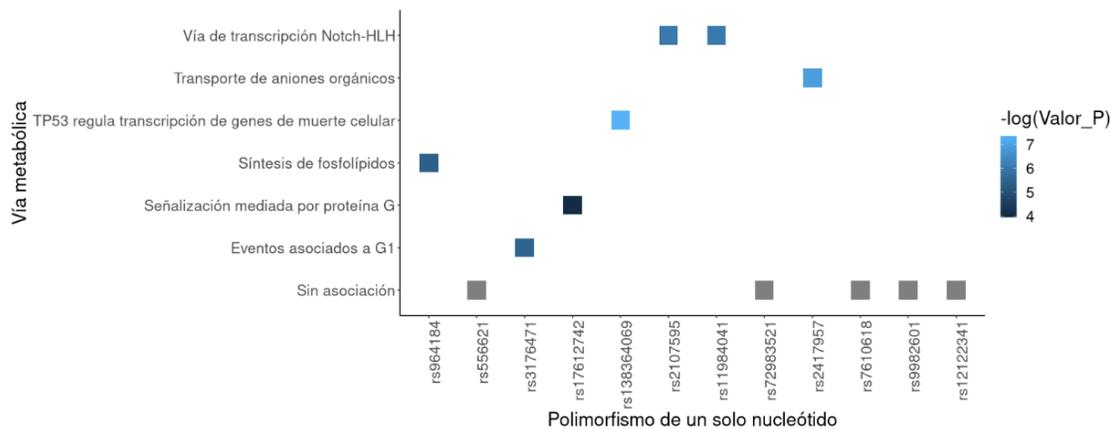


Figura 9. Vías metabólicas asociadas a los polimorfismos. El eje de las equis se conforma por los polimorfismos asociados a demencia vascular y aterosclerosis. Por su parte, en el eje vertical se encuentran las vías metabólicas descritas para cada variante en Reactome. El valor P muestra la significancia estadística de cada vía metabólica con los genes asociados a cada polimorfismo. El color del recuadro está dado por el logaritmo negativo del valor P. De esta manera, un tono más denso implica mayor asociación en esa vía.

Red de interacción

La figura 10 muestra la relación de los genes, polimorfismos, vías metabólicas y estirpes celulares mediante una red de interacción. En este gráfico el tamaño de cada nodo es proporcional al valor de centralidad de intermediación. Asimismo, la tonalidad de cada nodo corresponde al grado de centralidad. En otras palabras, se utilizó un color más denso para describir los nodos con mayor número de enlaces. De esta manera podemos observar que el nodo correspondiente a los oligodendrocitos se encuentra en el centro de la red, denotando su relevancia y el flujo de información que tiene a través del interactoma.

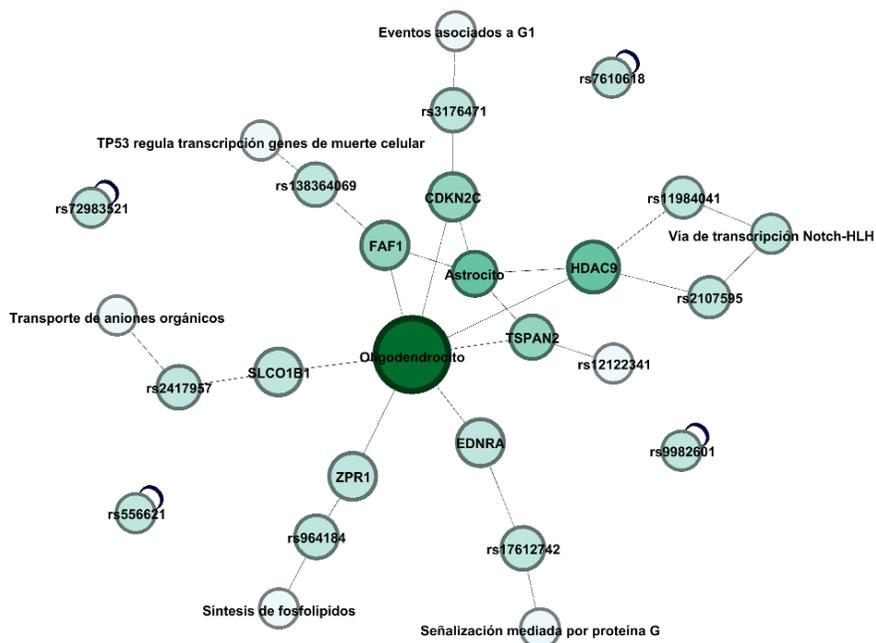


Figura 10. Interactoma conformado por genes, polimorfismos de un solo nucleótido, estirpes celulares y vías metabólicas asociadas al desarrollo de demencia vascular a partir de aterosclerosis. El tamaño del nodo se encuentra en relación con centralidad de intermediación mientras que el tono de estos es distributivo al grado de centralidad.

Las redes de interacción ofrecen una visión gráfica cualitativa sobre el comportamiento de los nodos en un sistema delimitado. Para complementar esta información en la figura 11 se observan los valores del grado de centralidad y centralidad de intermediación. A partir de estos datos se reconocen nodos destacados en cuanto al flujo de información de la red, tal es el caso de los nodos denominados “cuello de botella”. Previamente se describió que un nodo era considerado de esta manera cuando el valor de la métrica de centralidad de intermediación normalizada era mayor a 0.2. Por esta razón, se muestra una línea roja punteada que representa el límite que necesita un nodo para ser considerado dentro de esta categoría. Se muestra que solo los oligodendrocitos cumplen con esta característica, lo que indica que al ser eliminados de la red se generarían colonias de nodos aisladas. Asimismo, la designación de un nodo centro de red está dado por aquellos nodos con un valor de centralidad igual o mayor a la media aritmética más dos veces la desviación estándar. Las estimaciones presentes en esta red indican una media de 0.03 y una desviación estándar de 0.02. Es decir, aquellos nodos con un puntaje mayor a 0.08 en grado de centralidad normalizada serán considerados como centros de red. El límite para ser considerado en esta categoría está dado por la línea negra punteada. De igual forma, solo los oligodendrocitos cumplen con este criterio, lo que enfatiza su papel en relación con los elementos que dan parte al desarrollo de demencia vascular.

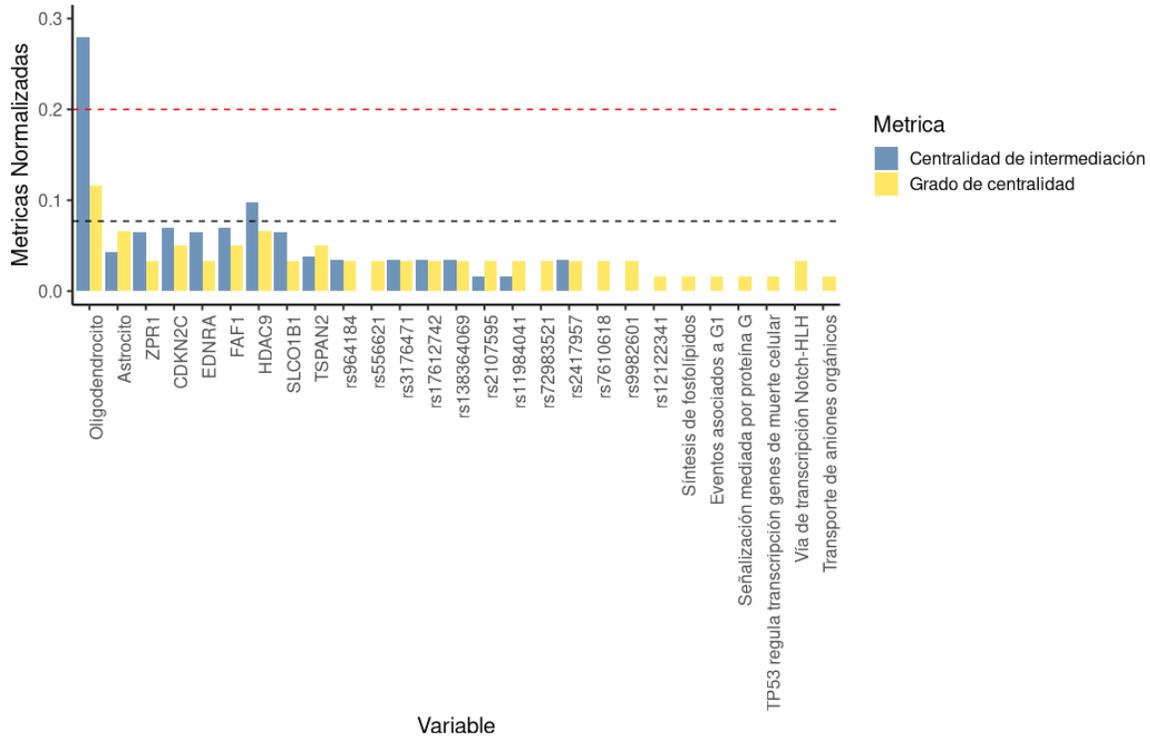
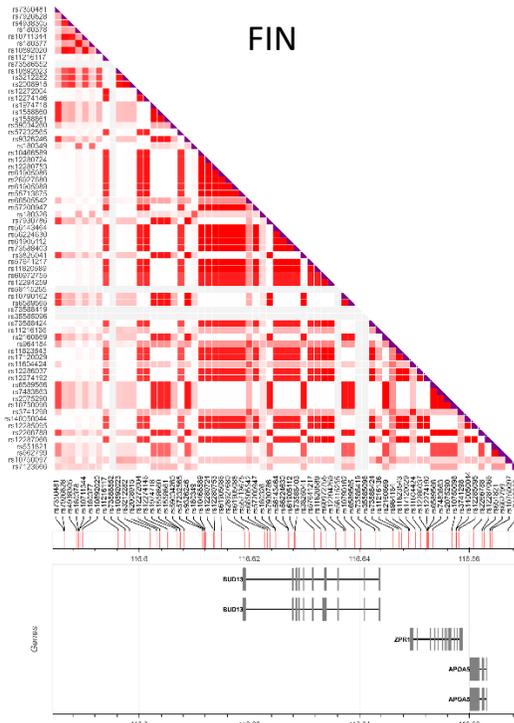
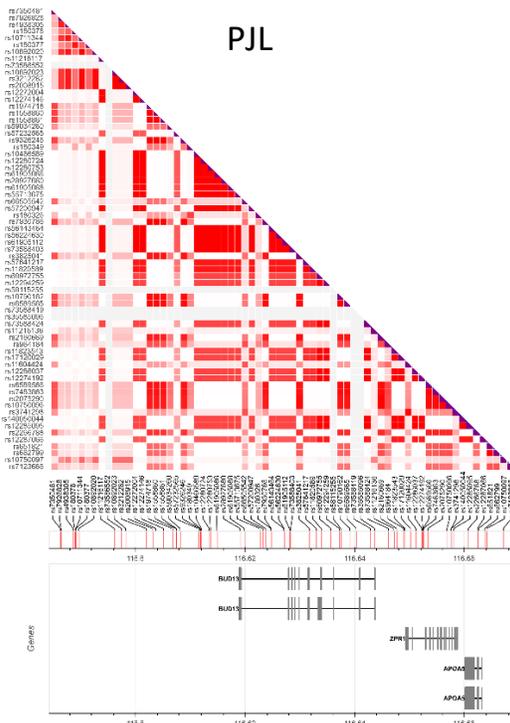
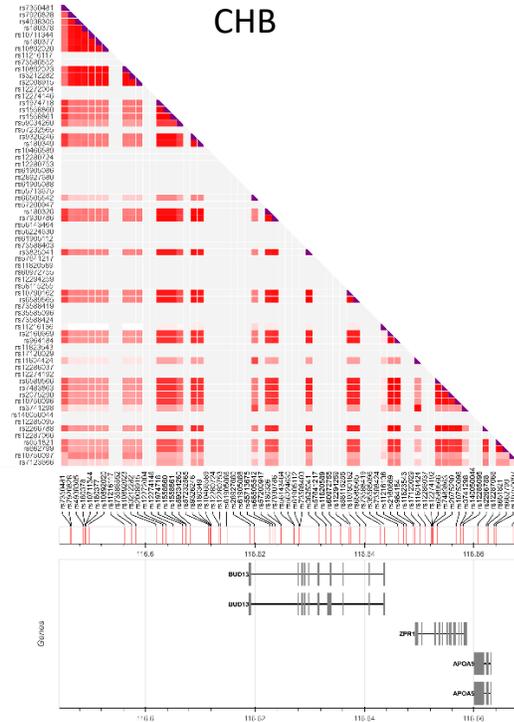
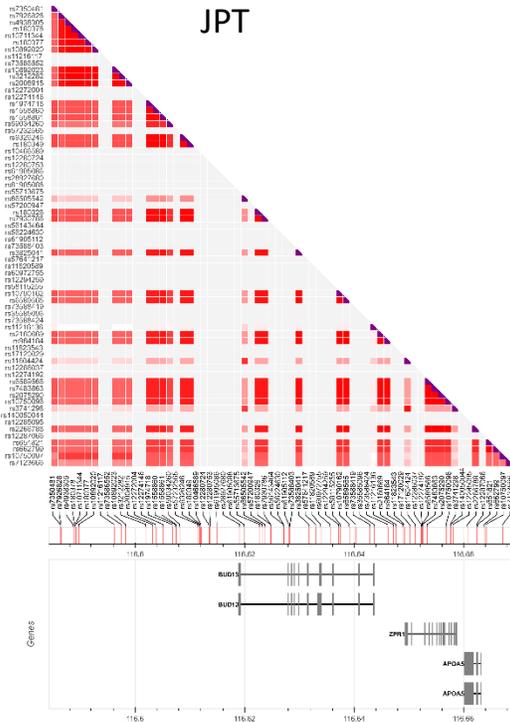


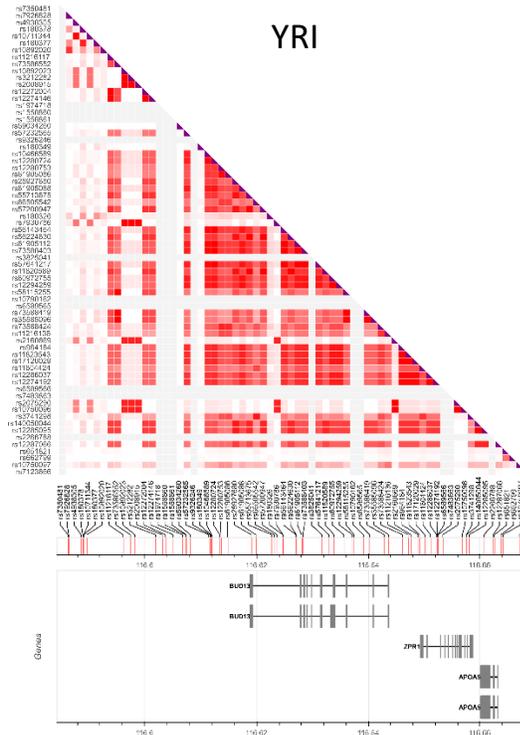
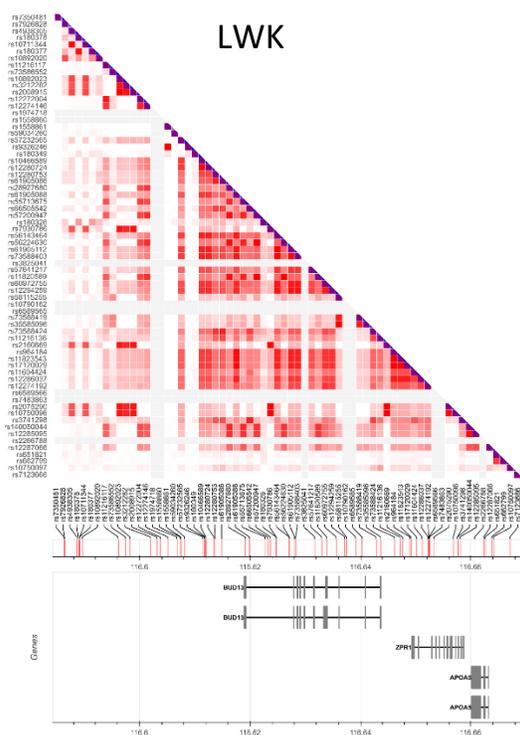
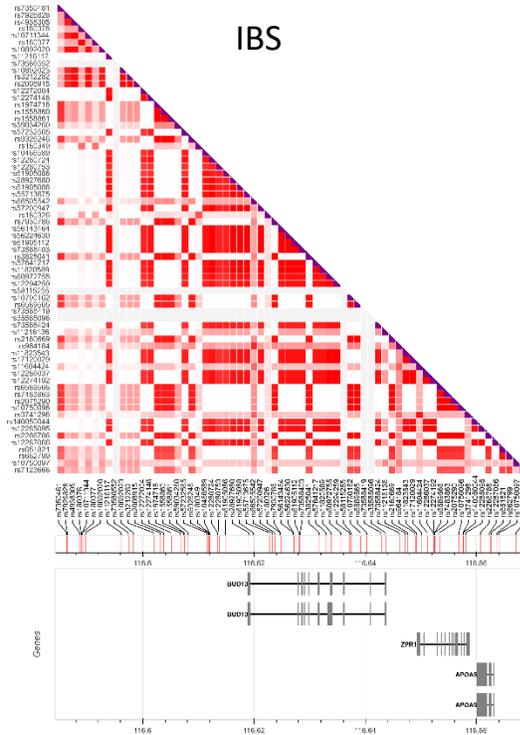
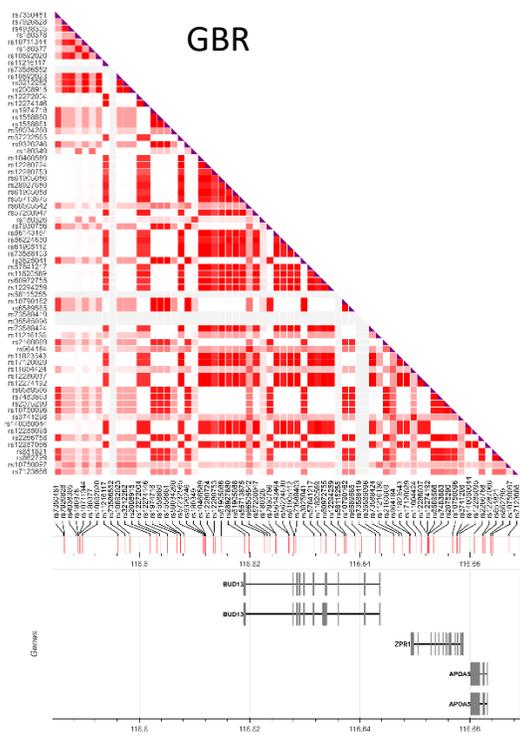
Figura 11. Grado de centralidad y centralidad de intermediación. Se muestran los valores de grado de centralidad normalizado (amarillo) y centralidad de intermediación normalizado (azul) para las variantes presentes en la red de interacción. Además, se agrega una línea horizontal roja que representa el límite para considerar un nodo como "cuello de botella" y una línea horizontal negra para identificar nodos centro de red. Donde solo los oligodendrocitos presentan estas características.

Desequilibrio de ligamiento y grupos poblacionales

El desequilibrio de ligamiento determina el conjunto de variantes que se heredan en bloque hacia la siguiente generación (76). A partir de esta información se define el grupo de polimorfismos o haplotipos que tienden a presentarse en una población específica (77). Evaluar los polimorfismos que se encuentran en DL con aquellos presentes en GWAS mediante el proceso de imputación permitirá realizar su asociación de manera indirecta en el desarrollo de las patologías de interés (115). Este método se realizó en 12 subpoblaciones del estudio de 1000 genomas fase 3 (95), en una ventana de 500 mega bases tomando las variantes con una r^2 mayor a 0.6. Posteriormente, se identificó a los polimorfismos estructurales con efecto en el marco de lectura. Cabe resaltar que esta metodología se desarrolló con los 12 SNPs asociados a demencia vascular y aterosclerosis. No obstante, los resultados presentes a continuación pertenecen a aquellos que responden los objetivos planteados en este proyecto (el resto de las imágenes se muestran en los anexos A4-A14). La figura 12 corresponde al desequilibrio de ligamiento del polimorfismo rs964184 presente en el gen ZPR1 en el cromosoma 11, en las 12 poblaciones de estudio. Esta figura se conforma por la asociación entre los SNPs encontrados en DL, donde el tono es distributivo al valor de r^2 . Es decir, entre más cercano a 1 sea

el valor de r^2 de dos variantes, se representará con un tono más oscuro. Asimismo, se presenta la arquitectura cromosómica con los genes que conforman esta región.





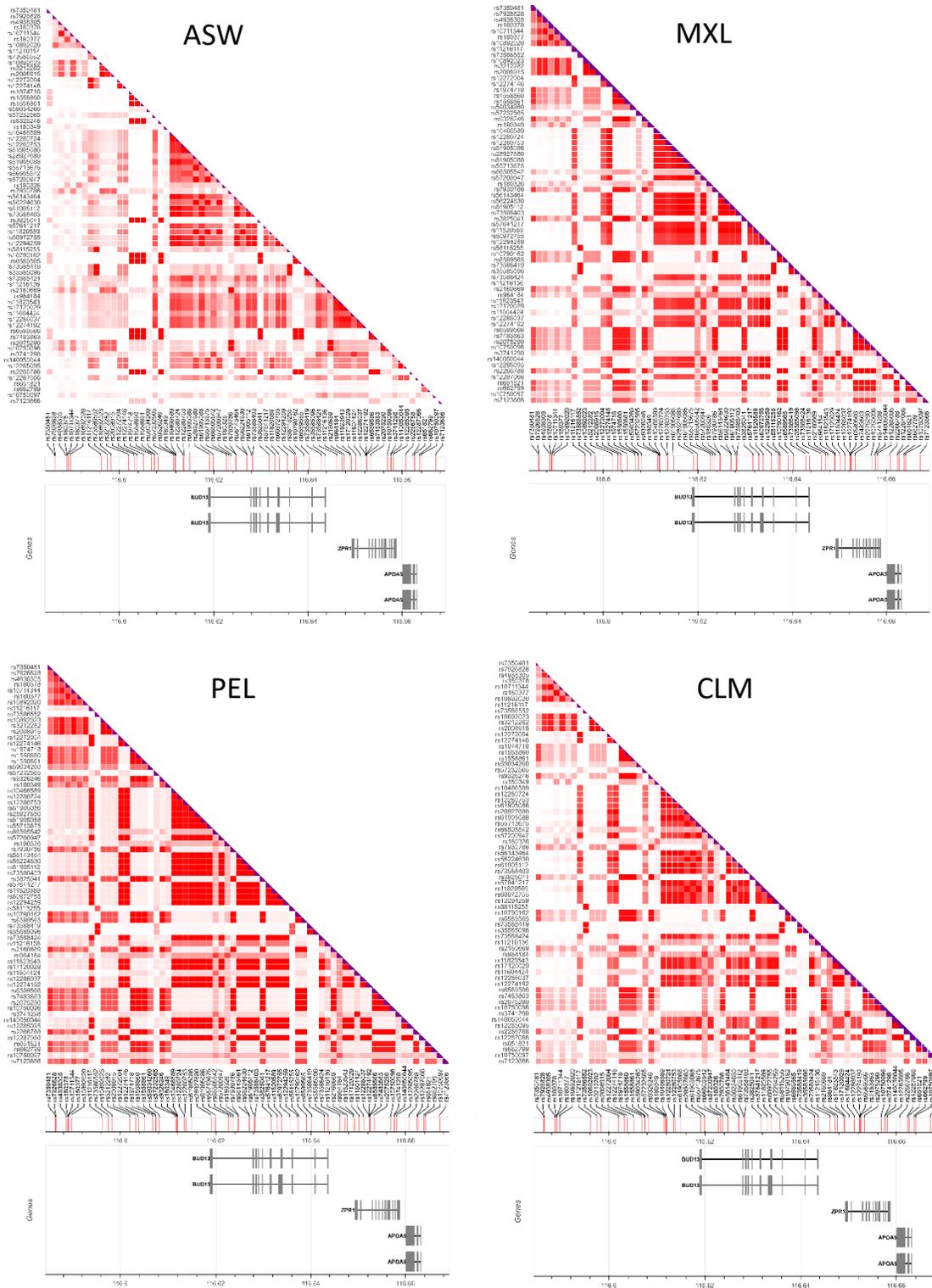


Figura 12. Desequilibrio de ligamiento de polimorfismos asociados a rs964184. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

Posteriormente, se identificaron aquellos SNPs que se comparten en cada grupo geográfico así como los que tienen un efecto en la estructura proteica. La siguiente imagen denota las variantes con estas características asociadas mediante el desequilibrio de ligamiento al polimorfismo rs964184. En la figura 13A, se observan los bloques de polimorfismos asociados a cada subpoblación. Los tonos azules corresponden a las poblaciones del continente asiático. Asimismo, la cohorte de subpoblaciones en verde describe ancestría africana. Los colores amarillos representan a la población europea y las notas rojizas detallan información de subpoblaciones amerindias. Se precisa que tanto las poblaciones asiáticas y africanas tienden a separarse, generando un grupo central conformado por población europea y amerindia. Asimismo, la figura 13B describe el tipo de variantes asociadas al polimorfismo rs964184. Se detallan tres SNPs. Los polimorfismos rs35385096 y rs11820589 se asocian al gen BUD13 y corresponden a variantes de sentido erróneo. Por su parte, la variante rs12287066 corresponde a una variante sinónima y se presenta en el gen APOA5.

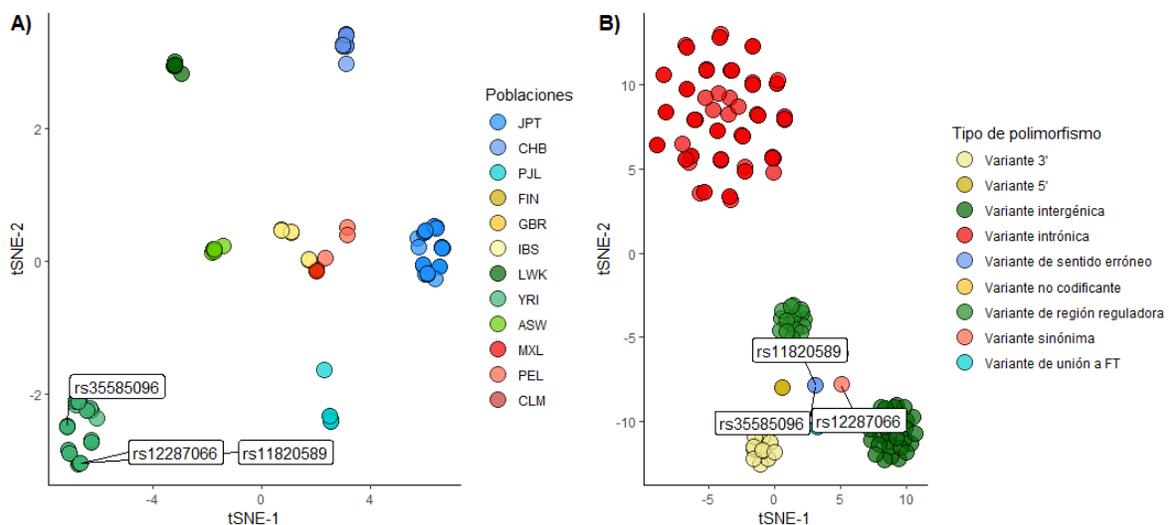


Figura 13. Desequilibrio de ligamiento de polimorfismos asociados a rs964184. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

A continuación se muestra la figura 14A conformada por información de polimorfismos en DL para el SNP rs3176471. Cabe resaltar que en las cohortes de estudio correspondientes a Japón, China y Nigeria dentro del proyecto 1000 genomas fase 3, no se detectó la variante rs3176471. Por tal motivo, estas subpoblaciones se omiten en el gráfico. No obstante, se exponen los resultados obtenidos en las 9 subpoblaciones restantes. Asimismo en la figura 14B se enfatizan a la variante rs200133978 de sentido erróneo, la cual está asociada al gen FAF1.

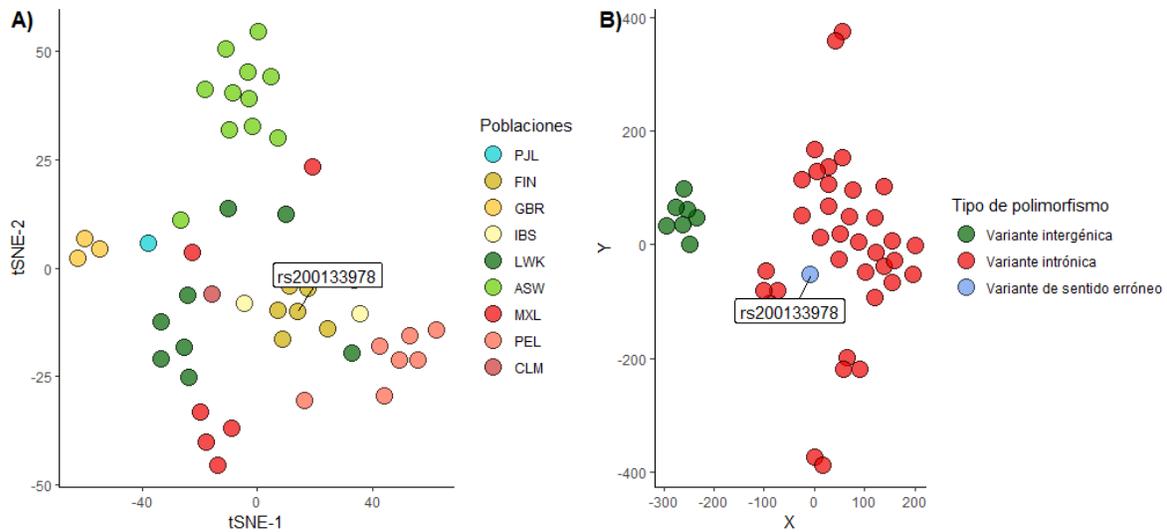


Figura 14. Tipo de polimorfismos asociados al SNP rs3176471. Se describe la característica de cada polimorfismo.

Continuando con la búsqueda de variantes con efecto estructural y su comportamiento a nivel poblacional se muestran los resultados del polimorfismo rs138364069 correspondiente a la región intrónica del gen FAF1. La figura 15A describe las variantes en DL para este polimorfismo y las poblaciones en las que se presentan. Asimismo, se observa la ausencia de información en población de Japón, China y Nigeria compartiendo la misma justificación que el apartado previo. En la figura 15B se señala al polimorfismo rs200133978 de sentido erróneo. Como se mencionó previamente, esta variante corresponde al gen FAF1. De esta manera se puede describir la interacción entre los polimorfismos rs3176471 y rs138363069. Ya que en conjunto se asocian con el SNP rs200133978 en el mismo grupo poblacional.

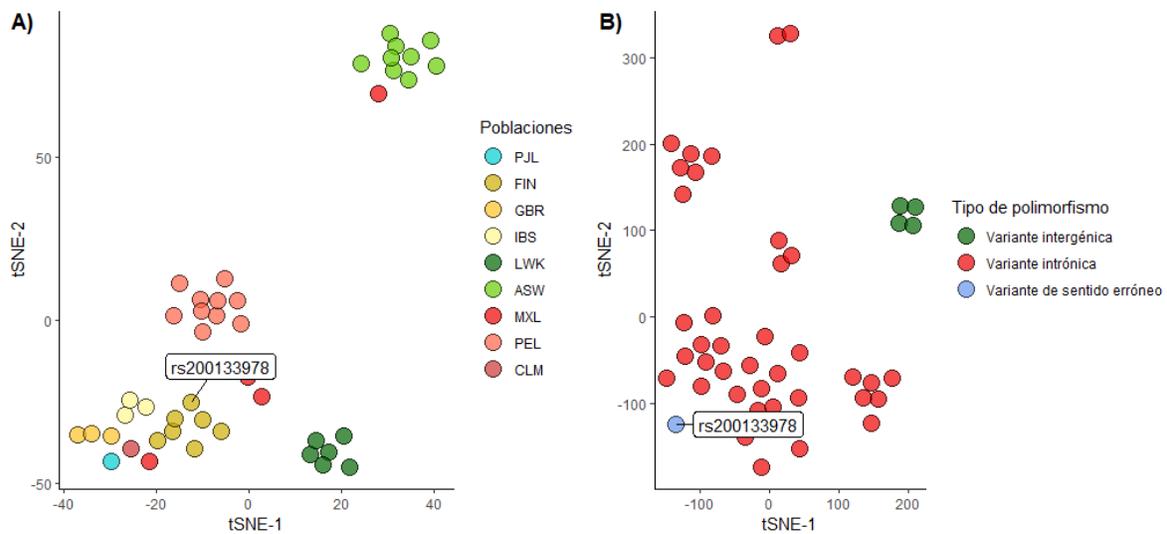


Figura 15. Tipo de polimorfismos asociados al SNP rs138364069. Se describe la característica de cada polimorfismo.

El SNP rs72983521 corresponde a una variante intergénica en el cromosoma 11. Cabe destacar que este polimorfismo no se encontró en la cohorte que conforma la subpoblación de Japón. No obstante, esta misma variante fue identificada en las subpoblaciones de China y Pakistán. Lo cual describe cambios en la frecuencia de presentación de ciertos polimorfismos a nivel continental. De esta manera, se hace énfasis en las diferencias genotípicas presentes entre cada grupo poblacional. La figura 16A muestra un conjunto central estructurado por variantes amerindias, europeas, asiáticas y en menor medida por polimorfismos de población con ascendencia africana en Estados Unidos. Asimismo, se detallan grupos conformados por variantes de población de Colombia y dos cohortes establecidas por SNPs de ancestría africana. De la misma forma, la figura 16B describe el tipo de polimorfismos que se encuentran en DL con el SNP rs72983521. Se resalta la presencia de dos variantes estructurales. En primer lugar se describirá al polimorfismo rs17368582 asociado al gen MMP12, el cual corresponde a una variante sinónima. Por su parte, el SNP rs143113210 presente en el gen MMP20 se asocia a las variantes de sentido erróneo.

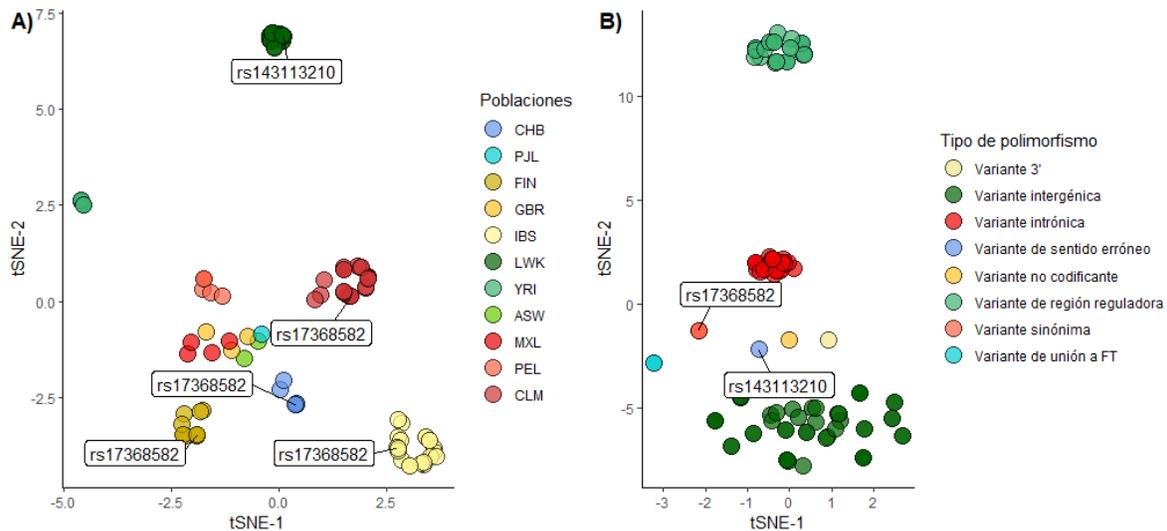


Figura 16. Tipo de polimorfismos asociados al SNP rs72983521. Se describe la característica de cada polimorfismo.

Por último se muestran los resultados de la evaluación poblacional del SNP rs7610618. En la imagen 17A se aprecia la ausencia de la población de Los Ángeles con ascendencia mexicana, ya que el polimorfismo rs7610618 no fue reportado en esta subpoblación. En esta imagen se observa un grupo establecido por variantes que corresponden a población de Medellín, Colombia. Así, como un conjunto dado por polimorfismos de habitantes de Europa y Asia. Por último se denota un grupo conformado por variantes de Perú y Nigeria. Posteriormente, la imagen 17B se compone de información sobre el tipo de variantes encontradas. Dentro de los polimorfismos presentes en desequilibrio de ligamiento con el SNP rs7610618 se encontró la variante rs34624984 de sentido erróneo asociada al gen CP.

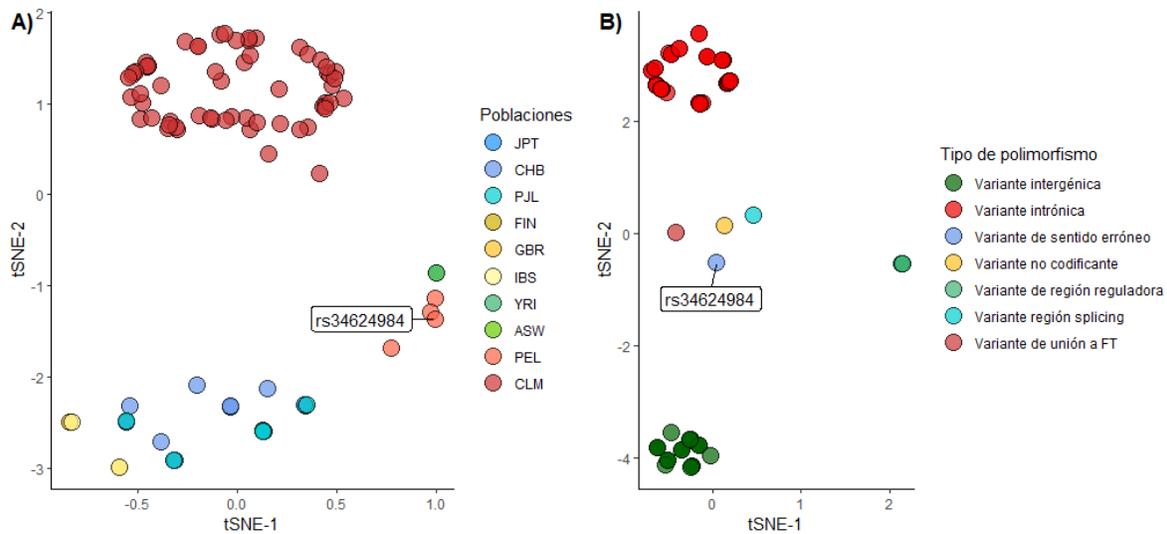


Figura 17. Tipo de polimorfismos asociados al SNP rs7610618. Se describe la característica de cada polimorfismo.

A partir de los datos derivados del desequilibrio de ligamiento y la comparación entre poblaciones del proyecto 1000 genomas fase 3 se observaron nuevas variantes genéticas asociadas a los polimorfismos presentes en demencia vascular y aterosclerosis. Estos SNPs tienden a presentar frecuencia diferente en cada grupo poblacional. Cabe resaltar que dentro de este conjunto de variantes se destacaron aquellas con un efecto en la estructura primaria proteica. La tabla 3 describe los polimorfismos encontrados mediante el método de imputación del desequilibrio de ligamiento, el valor del coeficiente de correlación, el gen asociado, el tipo de variante, el efecto sobre el codón, el residuo alterado por el polimorfismo y la población donde fue descrito.

SNP	SNP en DL	r ²	GEN	TIPO DE SNP	CODÓN	RESIDUO	POBLACIÓN
rs964184	rs35585096	0.61	BUD13	Sentido erróneo	GCC > TCC	A22S	YRI
rs964184	rs11820589	0.67	BUD13	Sentido erróneo	CCG > CTG	P148L	YRI
rs964184	rs12287066	0.67	APOA5	Sinónimo	ATC > ATA	I44	YRI
rs3176471; rs138364069	rs200133978	1.00	FAF1	Sentido erróneo	AAT > CAT	N34H	FIN
rs72983521	rs17368582	1.00	MMP12	Sinónimo	CCA > CCG	P279	IBS, FIN, CHB
rs72983521	rs143113210	1.00	MMP20	Sentido erróneo	GGC > GAC	G80D	LWK
rs7610618	rs34624984	1.00	CP	Sentido erróneo	CGT > TGT	R367C	PEL

Tabla 3. Polimorfismos estructurales asociados mediante método de imputación.

La información de estos cinco polimorfismos se utilizó en la herramienta SCRAD v0.9 para determinar su relación con las células de la UNV. La figura 18 muestra los resultados.

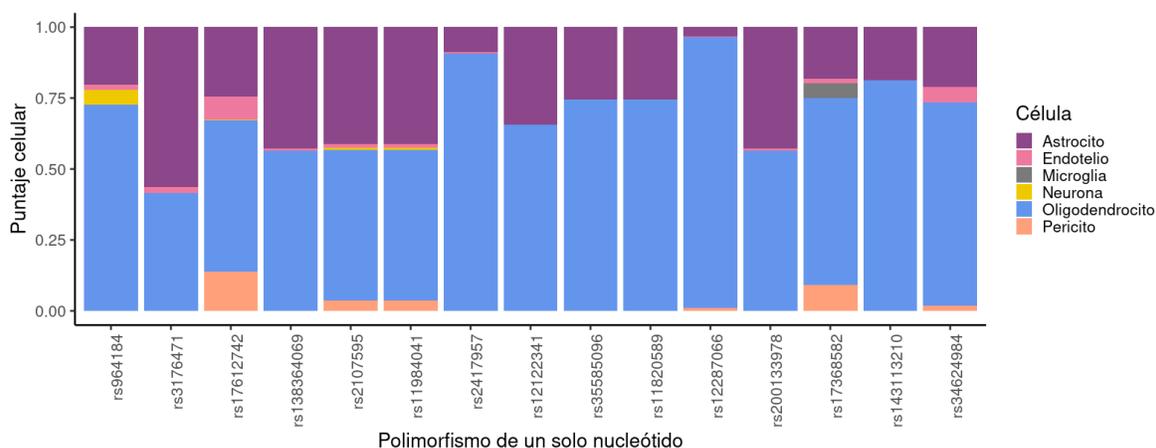


Figura 18. Polimorfismos y estirpes celulares. Se muestra la probabilidad de asociación de cada variante con un tipo celular de la unidad neurovascular. Los SNPs tienden a relacionarse con oligodendrocitos (azul) y astrocitos (morado). Existe una menor probabilidad de presentarse en microglía (gris), endotelio (rosa) o pericitos (salmón).

Esta imagen comprende información del primer conjunto de polimorfismos y se complementa con los datos de las variantes encontradas mediante el desequilibrio de ligamiento. Asimismo, se detalla la relación de estos polimorfismos con los oligodendrocitos y en menor proporción con los astrocitos. Estas observaciones sugieren que los oligodendrocitos son una pieza clave en la interacción de las variantes asociadas al desarrollo de demencia vascular.

Evaluación de estructura secundaria

Para determinar el efecto de los polimorfismos sobre la estructura y función proteica se utilizó GOR IV y PSIPRED. De esta manera, se puede obtener información sobre la estructura secundaria afectada cuando hay ausencia de información cristalográfica. En la tabla 4 se describen los resultados obtenidos en cada algoritmo. Para los polimorfismos rs35585096 y rs11820589 asociados al gen BUD13, ambas herramientas denotan que las variantes se presentan en sitios de plegamiento azaroso. Por su parte, el polimorfismo rs12287066 derivado del gen APOA5 se encuentra en una alfa hélice. Para el SNP rs200133978 del gen FAF1 se detalla en ambas herramientas que el residuo afectado se encuentra en un sitio de plegamiento azaroso. Asimismo los polimorfismos rs17368582 y rs143113210 asociados a los genes MMP12 y MMP20 respectivamente, se encuentran en sitios de plegamiento azaroso. Por último, la variante rs34624984 derivada del gen CP se describe en sitios de plegamiento azaroso.

SNP	GEN	PROTEÍNA	RESIDUO	GOR IV	PSIPRED
rs35585096	BUD13	Homólogo de BUD13	A22S	Plegamiento azaroso	Plegamiento azaroso
rs11820589	BUD13	Homólogo de BUD13	P148L	Plegamiento azaroso	Plegamiento azaroso
rs12287066	APOA5	Apolipoproteína 5	I44	Alfa hélice	Alfa hélice
rs200133978	FAF1	Factor 1 asociado a FAS	N34H	Plegamiento azaroso	Plegamiento azaroso
rs17368582	MMP12	Matriz metaloproteínasa 12	P279	Plegamiento azaroso	Plegamiento azaroso
rs143113210	MMP20	Matriz metaloproteínasa 20	G80D	Plegamiento azaroso	Plegamiento azaroso
rs34624984	CP	Ceruloplasmina	R367C	Plegamiento azaroso	Plegamiento azaroso

Tabla 4. Predicción de estructura secundaria mediante GOR IV y PSIPRED.

Evaluación del efecto del polimorfismo en términos estructurales de la proteína

Para analizar el efecto sobre la función proteica de los polimorfismos asociados a demencia vascular y aterosclerosis se utilizaron tres herramientas bioinformáticas (SIFT, PolyPhen2 y PROVEAN), cuya base radica en comparación de secuencias por homología. Este conjunto de algoritmos parte de la premisa que los residuos conservados tienden a presentar un papel relevante para la estructura y función proteica. Por tal motivo, al ser sustituidos comprometen estas características. No obstante, estas herramientas solo calculan el efecto de los polimorfismos no sinónimos. Es por esto por lo que las variantes rs12287066 y rs17368582 asociadas a los genes APOA5 y MMP12 no se incluyeron en el análisis. La tabla 5 describe los resultados obtenidos en cada algoritmo. Para el polimorfismo rs35585096 SIFT detalla que el cambio de la alanina 22 por una serina no es tolerable. Sin embargo, PolyPhen2 y PROVEAN muestran un cambio probablemente dañino y neutral, respectivamente. Por su parte, para el SNP rs11820589 SIFT y PROVEAN describen que el cambio del residuo 148 tiene un efecto negativo. En cambio, PolyPhen2 detalla a esta mutación como un evento benigno. Para los SNPs rs200133978 y rs143113210 se describe un efecto dañino sobre las proteínas a las cuales se encuentran asociadas. Cabe señalar que este resultado se presentó en los tres algoritmos. Por último, rs34624984 denota un efecto dañino en SIFT y PolyPhen2, mientras que PROVEAN lo destaca como una mutación neutral.

SNP	GEN	PROTEÍNA	RESIDUO	SIFT	PolyPhen2	PROVEAN
rs35585096	BUD13	Homólogo de BUD13	A22S	0.00 No tolerante	0.578 Probablemente dañino	-0.813 Neutral
rs11820589	BUD13	Homólogo de BUD13	P148L	0.02 No tolerante	0.047 Benigna	-4.166 Deletéreo
rs200133978	FAF1	Factor 1 asociado a FAS	N34H	0.00 No tolerante	1.00 Dañino	-2.993 Deletéreo
rs143113210	MMP20	Matriz metaloproteínasa 20	G80D	0.01 No tolerante	0.955 Dañino	-5.554 Deletéreo
rs34624984	CP	Ceruloplasmina	R367C	0.01 No tolerante	0.936 Dañino	-1.528 Neutral

Tabla 5. Predicción del efecto funcional de cada polimorfismo sobre la estructura proteica

Discusión

Una de las aplicaciones de los estudios de asociación de genoma completo es la conexión entre polimorfismos y entidades nosológicas (51) con el objetivo de identificar variantes específicas relacionadas al desarrollo de una patología (131). Actualmente se han reportado 328,226,200 polimorfismos (132), de los cuales 1,223,591 se asocian a un fenotipo (93). A su vez, solo 179,025 están en un estudio de asociación genómica (57). De estos, 665 se encuentran ligados a aterosclerosis (89) y 77 a demencia vascular (98). No obstante, en este proyecto se realizó un análisis mediante teoría de conjuntos de los polimorfismos presentes en ambos padecimientos. Se encontró una intersección de 12 variantes genéticas entre ambos sistemas. Además, se utilizaron los datos de población pima y maya derivados de Human Genome Diversity Project, donde se resalta la presencia del SNP rs556621 en mayas y pimas. Por su parte, la variante rs11984041 solo se encontró en población maya. Sin embargo, el resto de los polimorfismos presentes en el grupo de asociación entre demencia vascular y aterosclerosis no formaron parte del microarreglo utilizado en la genotipificación de los grupos étnicos del Human Genome Diversity Project. Por lo cual, la ausencia de estos en la población maya y pima es una cuestión metodológica y no genética. De igual manera, al realizar la comparación entre los dos grupos poblacionales se detalla que ambos presentan 425,555 polimorfismos en común, teniendo 105,772 SNPs únicos para población maya y 20,318 polimorfismos propios para población pima. Siendo estas diferencias poblacionales la clave para identificar características individuales de cada grupo étnico tanto en la construcción de un modelo evolutivo de ancestría como en el desarrollo de medidas diagnósticas y preventivas específicas.

Los cambios en el genoma pueden darse en cromosomas somáticos o sexuales. Por tal motivo, es preciso identificar la localización cromosómica del grupo de variantes presentes en nuestras dos patologías de estudio. En 2011 Schrijvers *et al* (64), mediante un GWAS, propone un loci asociado al desarrollo de demencia vascular presente en el cromosoma X. La variante descrita corresponde al SNP rs12007229 en el gen que codifica para el receptor de andrógenos. No obstante, no describe el mecanismo fisiopatológico que implica a este polimorfismo en el desarrollo de demencia vascular. Por su parte, los resultados de este proyecto indican que los SNPs asociados a ambas patologías se encuentran en cromosomas somáticos, siendo certeros en los cromosomas 1, 3, 4, 6, 7, 11, 12 y 21.

La información obtenida puntualiza 12 polimorfismos asociados a 7 genes, debido a que algunos SNPs se localizan en regiones no codificantes, mientras que las variantes rs2107595 y rs11984041 se asocian al gen HDAC9. Cabe resaltar que este gen codifica para la proteína histona deacetilasa 9 (91), que es responsable de la desacetilación de los residuos de lisina de las histonas H2A, H2B, H3 y H4; teniendo un papel en la regulación de la transcripción, ciclo celular y desarrollo (133). En 2013, Markus *et al* describen la presencia de ambos polimorfismos en la aceleración y formación de la placa aterosclerótica. Asimismo, asocia a estas variantes con la predisposición de ruptura de la placa y el desarrollo de complicaciones tromboembólicas (134). En contraste, en la figura 4 se observa un

polimorfismo con un valor de asociación mayor en comparación al resto. Este es el rs964184 correspondiente al gen ZPR1 que codifica para la proteína dedo de zinc 1. La variante reportada se encuentra en la región 3' UTR, cuya función se ha relacionado con la regulación de la localización del ARN mensajero, participando en la polaridad celular. De la misma manera, se describe que alteraciones en este gen culminan en modificaciones de la morfología de las dendritas neuronales, con disminución de la plasticidad y la sinapsis (135). El conjunto de estas características es un punto determinante en el desarrollo de demencia vascular a partir de complicaciones trombóticas y cambios en la morfología de las células que componen el tejido cerebral.

La evaluación de la frecuencia alélica determina la proporción de personas que tienen formas alternativas de un gen en específico con el total de la población de estudio (59). En este proyecto hace énfasis en la frecuencia alélica menor (MAF, por sus siglas en inglés), esta frecuencia remarca el hecho de que las variantes asociadas a patologías tienen menor representación dentro de la población (136) derivado de procesos de selección negativa a través del tiempo, mismos que tienden a ser diferentes dependiendo de las características ambientales y genéticas de cada grupo étnico, lo que se traduce en una frecuencia alélica distinta para cada variante de riesgo dependiendo de la población de estudio. De esta manera, la predisposición genética ante el desarrollo de una enfermedad compleja será diferente para cada población (137)(138). En este proyecto se muestra que la frecuencia alélica menor de los doce polimorfismos de estudio sigue esta premisa. Ejemplo de esto es el SNP rs556621, con una MAF superior a 0.4 para la población de Japón, China y amerindios. Por su parte la misma variable genética en ancestría africana tiene una frecuencia cercana a 0.1. Otro caso de la variabilidad de la distribución de los polimorfismos es el SNP rs9982601 donde ocurre un evento opuesto, ya que la presencia de este polimorfismo en población africana es superior a 0.2, pero en Asia su representación es nula. Dicho de otra manera, ninguna persona secuenciada en la cohorte de estudio presentó este polimorfismo.

Las enfermedades complejas son influenciadas por factores ambientales y genéticos (137). El análisis de las frecuencias muestra una distribución variable de los polimorfismos en cada zona geográfica. Lo anterior indica que la heterogeneidad implicada en el desarrollo de una patología a nivel genético puede ser diferente para cada grupo étnico. Por tal motivo, el evaluar la variabilidad en la frecuencia de presentación de un SNP asociado a una enfermedad generará medidas preventivas específicas acorde a las características poblacionales (139).

Conocer el efecto de los polimorfismos sobre el metabolismo genera información sobre los procesos que se ven afectados y cuya culminación son el asentamiento del cuadro clínico en la demencia vascular (140). No obstante, este proceso no depende de una sola vía metabólica. De hecho, se ha descrito que son el conjunto y la acumulación de estos cambios en el metabolismo los precursores de la neurodegeneración y posterior declive cognitivo (141). De esta manera en la figura 8 se muestran las principales vías metabólicas que se encuentran asociadas a los genes comprometidos por los polimorfismos. Con base en una revisión bibliográfica

se ejemplifica el papel de estas vías en un entorno fisiológico y patológico, permitiendo esclarecer los procesos que dan parte al desarrollo de demencia vascular. En principio, hablaremos sobre la homeostasis de los lípidos, cuyo papel en el sistema nervioso central deriva en la modulación de las membranas plasmáticas de los estirpes celulares, permitiéndoles ser parte de eventos como liberación de neurotransmisores, plasticidad sináptica, embebimiento de proteínas transmembranales e inflamación. Resaltando que el asentamiento del cuadro clínico puede tener origen en alteraciones de esta vía (142). Asimismo, los resultados presentes describen la asociación de los polimorfismos con las proteínas G. En este punto se denota que estas proteínas participan desde etapas tempranas del crecimiento y desarrollo hasta eventos como el mantenimiento de las sinapsis neuronales durante la etapa adulta. Por su parte, en el cerebro desarrollado se destaca su participación en la regulación de la cascada de la inflamación. Aún se desconoce su función, sin embargo se ha planteado que pueden tener un papel en la muerte neuronal durante eventos con disminución del aporte de oxígeno (143). Por último, se hace hincapié sobre la participación de dos polimorfismos asociados al gen HDAC9 y la vía Notch HLH. Este sistema de señalización se encuentra altamente conservado en los mamíferos, teniendo intervención en procesos de diferenciación celular, angiogénesis así como metabolismo de glucosa y lípidos. De hecho, participa en la progresión de la placa aterosclerótica y en la formación de placas inestables (144). Recordando, procesos dan parte al declive cognitivo presente en la demencia vascular.

El identificar las estirpes celulares presentes en el desarrollo de una patología, aunado a la asociación de genes afectados en cada variable permite el reconocimiento del mecanismo fisiopatológico. SCRAD v0.9 mediante una función probabilística, genera un puntaje de asociación entre los genes comprometidos por los polimorfismos y las células de la UNV. A partir de esta información se muestra que la mayoría de las variantes presentan un contexto celular con tendencia hacia los oligodendrocitos y astrocitos. Los oligodendrocitos son células especializadas del sistema nervioso central con la capacidad de formar la vaina de mielina que envuelve los axones neuronales. Este recubrimiento permite potenciales de acción más rápidos y eficientes (13)(14). La vaina de mielina es un complejo lipídico generado a partir de la prolongación del soma del oligodendrocito, el cual puede envolver varios axones neuronales a la vez. El proceso de mielinización sucede durante las etapas del desarrollo y la vida adulta. No obstante, su actividad es mayor durante el transcurso embrionario. Además, la mielinización se ha asociado con la plasticidad neuronal. Donde los oligodendrocitos sincronizan conexiones específicas, haciendo más efectiva la conducción nerviosa (13). Una vez realizado el proceso de mielinización, se inicia un evento de acoplamiento celular entre oligodendrocitos y neuronas, donde existe un flujo continuo de transferencia de información. Incluso, se presenta un mecanismo de daño reflejo, ya que alteraciones en los oligodendrocitos repercuten en el axón y viceversa. Por tal motivo, los periodos con disminución en el aporte de oxígeno generan daño neuronal y disfunción en el mecanismo de comunicación entre los oligodendrocitos y las neuronas. Es el conjunto de las alteraciones provocadas por la hipoxia lo que deriva

en enfermedades donde la mielolisis (destrucción de la vaina de mielina) tiene un papel clave (15).

Por su parte, los astrocitos son células estructurales complejas que participan en el equilibrio iónico, la regulación del volumen extracelular, la plasticidad, así como la formación, maduración y eliminación de las sinapsis. De hecho, se denota su participación en la creación de patrones de redes neuronales (72). No obstante, existe la propuesta que la diversidad de sus funciones radica en diferentes tipos de astrocitos con características morfológicas y estructurales, que le permiten realizar estas actividades (145). Asimismo, una de sus labores es la regulación de la permeabilidad selectiva de la barrera hematoencefálica, este proceso se genera en conjunto con las células endoteliales y los pericitos (17). De la misma manera, la afección de estos puede derivar en un mecanismo de daño a las células periféricas y aquellos con mecanismos asociados (72). En consecuencia, el deterioro a los astrocitos se refleja en disfunción de la barrera hematoencefálica, edema cerebral, neurodegeneración y daño axonal difuso (1). La integración de la información de la UNV permite entender como el deterioro de un subgrupo celular deriva en la amplificación del daño, involucrando una mayor proporción de estirpes celulares del sistema nervioso (145).

Como se ha mencionado, los oligodendrocitos presentan un acoplamiento metabólico con los axones neuronales posterior a la formación de la vaina de mielina (13). Por tal motivo, no resulta extraño que alteraciones en la arquitectura y metabolismo de un elemento se reflejan en el otro (15). Mediante la teoría de redes se visualiza que los oligodendrocitos son el nodo con mayor flujo de información. De hecho, basado en sus métricas se considera como un nodo centro de la red y cuello de botella. Debido al conjunto de elementos con los que se relacionan los oligodendrocitos, se convierten en una pieza clave para comenzar el análisis de la patogénesis de la demencia vascular. Estudios *post mortem* detallan una disminución de oligodendrocitos en relación con los axones que presentan daño en la vaina de mielina. Asimismo, mediante resonancia magnética se ha visualizado un decremento de axones mielinizados en pacientes con deterioro de las habilidades cognitivas (146). También, se detalla una conexión entre la alteración de las células endoteliales durante las enfermedades cardiovasculares como aterosclerosis y la disrupción de la barrera hematoencefálica, aunado a una pérdida del acoplamiento celular por parte de la UNV. De hecho, se describe que los oligodendrocitos presentan cambios tempranos reflejados en la unión con el axón neuronal durante procesos hipóxicos, incluso previo a la disfunción de la barrera hematoencefálica (147). Actualmente, no se tiene una explicación concreta de la degeneración de oligodendrocitos y disminución de la vaina de mielina en la demencia vascular. No obstante, observaciones en demencia asociada a Alzheimer denotan que la ruptura de la vaina de mielina y la disminución de oligodendrocitos se presentan en una etapa temprana de esta enfermedad. De hecho, estos cambios ocurren de manera precedente a la formación de placas amiloides y proteínas tau (148).

Se considera que el desarrollo de una enfermedad crónico-degenerativa involucra cuestiones sobre la susceptibilidad de una proporción de la población basada en

sus características genéticas (115). A partir de los estudios de asociación genómica completa es posible identificar variantes con relación estadística entre el grupo de estudio y la entidad nosológica (149)(79). No obstante, la búsqueda de polimorfismos se complementa con la evaluación del conjunto de variantes que tienden a heredarse en grupo dentro de una población (54)(150). Es precisamente este análisis indirecto mediante el método de imputación el que permite inferir la participación de variantes no genotipificadas en el desarrollo de estas enfermedades (151)(115). Como se ha descrito, al conjunto de alelos presentes en el genoma de una población en específico se le ha denominado haplotipo y es mediante la incorporación de la imputación por desequilibrio de ligamiento que se posibilita identificar segmentos cromosómicos con la capacidad de influir en el desarrollo de los diferentes estados nosológicos (54). Por tal motivo, se decidió evaluar las variantes en DL asociadas a los polimorfismos encontrados en demencia vascular y aterosclerosis, con el objetivo de describir los haplotipos presentes en cada una de las doce subpoblaciones de estudio del proyecto 1000 genomas fase 3 y detallar el mayor número de polimorfismos estructurales. En primer lugar se valoró el comportamiento y la frecuencia en cada población de los polimorfismos en desequilibrio de ligamiento. Denotando un modelo similar a los resultados de frecuencias alélicas, donde se presentan patrones característicos de polimorfismos que se heredan en conjunto para cada sitio geográfico. En el marco de la evaluación de haplotipos mediante el DL se destaca la participación de variantes con efecto en la estructura primaria proteica, encontrando polimorfismos sinónimos y de cambio de sentido. La tabla 3 muestra las variantes con efecto estructural encontrados mediante el método de imputación. Estos polimorfismos se asocian a los genes BUD13, APOA5, FAF1, MMP12, MMP20 y CP.

Derivado de la comprensión sobre la biología estructural, se ha hecho énfasis entre la relación de la estructura proteica y su función (83). Donde el papel de la proteína en diferentes procesos biológicos depende en gran parte de los residuos que la conforman y las modificaciones que estos últimos puedan presentar (119)(85). A partir de este hecho, se describe el papel de los polimorfismos estructurales en el desarrollo de diferentes patologías, ya que cambios en los residuos que la conforman pueden modificar sus características fisicoquímicas, con repercusión en la estructura tridimensional (82). Es este punto, donde entender cómo una alteración puntual a nivel genético se representa en un contexto de pérdida o ganancia de función proteica (83). Con el fin de determinar el papel de los polimorfismos sobre el producto del gen se utilizaron diferentes herramientas bioinformáticas. Las primeras dos GOR IV y PSIPRED predicen la organización secundaria a partir de la secuencia de aminoácidos. De esta manera, se definieron las estructuras comprometidas ante el cambio de un residuo, aun en ausencia de información cristalográfica (120)(152). Cabe destacar que se obtuvieron los mismos resultados en ambas herramientas, tomando en cuenta que estos algoritmos se basan en diferentes metodologías. Para los polimorfismos de BUD13, FAF1, MMP20, MMP12 y CP se describe que el residuo comprometido se presenta en sitios de plegamiento azaroso. Por su parte, el SNP sinónimo de APOA5 se encuentra en una alfa hélice. Una vez encontrado el sitio afectado por la variante genética, se complementó la información con SIFT, PolyPhen2 y PROVEAN. En el caso de PolyPhen2 se

omitieron los valores de HumVar, esto derivado del sesgo que pueden presentar los resultados. Asimismo, debido a que estas herramientas evalúan variantes no sinónimas, los SNPs rs12287066 y rs17368582 no se tomaron en cuenta para este ensayo. En primer lugar se destaca que los cinco SNPs obtuvieron un resultado no tolerante en SIFT, lo cual indica que los residuos comprometidos se localizan en sitios conservados. No obstante, HumDiv de PolyPhen2 describe a la variante rs35585096 asociada al gen BUD13 como probablemente dañina. Sin embargo, el SNP rs11820589 del mismo gen se representa como un cambio benigno. Por su parte las variantes rs200133978, rs143113210 y rs34624984 resultaron dañinas. Por último, PROVEAN detalla solamente a las variantes rs11820589, rs200133978 y rs143113210 con un efecto deletéreo. La variabilidad de los resultados radica en que estos algoritmos evalúan características diferentes además de la homología, basándose en modelos de funciones probabilísticas o comparación con un conjunto de datos. Sin embargo, de los cinco polimorfismos estudiados, solo dos presentaron un resultado con efecto dañino en las tres herramientas bioinformáticas. Estos son los polimorfismos rs200133978 y rs143113210 asociados a los genes FAF1 y MMP20, respectivamente. De esta manera al utilizar metodologías computacionales se genera información complementaria sobre la estructura y función proteica, aumentando el entendimiento del medio en el que participan (119). De igual manera, comprender la relación entre estructura y función en un estado fisiológico y la sucesión hacia una situación nosológica permite comprender la patogénesis de la misma (83). Cabe resaltar que mediante el uso de SCRAD v0.9 la variantes rs200133978 presentó asociación con oligodendrocitos y astrocitos. Por su parte el polimorfismo rs143113210 solo mostró valor significativo en relación con los oligodendrocitos.

Las metaloproteasas son una familia de aproximadamente 25 enzimas con tres dominios funcionales. Un dominio catalítico, un dominio C terminal y un dominio N terminal que se escinde durante la activación de la proteína. Las metaloproteasas se encuentran prácticamente en todos los tejidos, incluyendo el sistema nervioso central (70). Es aquí donde regulan de manera directa o indirecta mecanismos como crecimiento celular, apoptosis, migración y comunicación celular. No obstante, en eventos patológicos intervienen en procesos de degeneración axonal, inflamación, desmielinización, estrés oxidativo y reparación (153). Walker & Rosenberg (154), describen la participación de las metaloproteasas en la destrucción de la vaina de mielina, muerte de oligodendrocitos y disfunción de la barrera hematoencefálica durante hipoperfusión e hipoxia. También, se ha descrito que modificaciones en la estructura de las metaloproteasas pueden afectar su especificidad hacia los sustratos y por ende su actividad, dando parte a un evento patológico (70). De esta manera, el conjunto de alteraciones ambientales y genéticas tienen un impacto en la pérdida del equilibrio del sistema nervioso central (153) (71). El daño en la vaina de mielina da como resultado una disminución en la velocidad de transmisión del impulso nervioso (70), así como una pérdida en el acoplamiento metabólico presente entre los oligodendrocitos y axones neuronales, generando una posterior neurodegeneración, decremento de conexiones en la materia blanca y el asentamiento del cuadro clínico (71). En este proyecto se remarca el polimorfismo rs143113210 asociado al gen MMP20, el cual codifica para la metaloproteasa 20.

Actualmente no existe reporte del papel específico de este polimorfismo en el desarrollo de demencia vascular. No obstante, la información recabada en la literatura y los resultados de este trabajo sugieren su participación. Igualmente, este SNP de sentido erróneo modifica la glicina 80 en el dominio N terminal necesario para la activación de la metaloproteasa (70). Por lo cual, aunado a la información de PROVEAN, PolypHen2 y SIFT, sugerimos que el cambio de residuo tiene un efecto negativo sobre la estructura y función de la proteína, participando en el proceso de desmielinización y daño axonal durante la patogénesis de demencia vascular.

Por su parte el factor 1 asociado a FAS o FAF1 se compone por 5 dominios, de los cuales los dominios UBA y UBX participan en los mecanismos de degradación proteica mediada por el complejo proteasoma-ubiquitina (155)(156). Este factor participa en procesos como diferenciación celular (157), plegado de proteínas, mantenimiento neuronal, regulación del axón (158) y reparación del ADN. Esta proteína se expresa en testículos, musculo esquelético, corazón y cerebro (159). En la literatura se hace hincapié en la variabilidad de patologías que se asocian con alteraciones en esta proteína, como la degeneración de la corteza motora y oncogénesis (158). Kwangsik *et al* (160) describen que mutaciones en FAF1 derivan en el adelgazamiento de la corteza entorrinal del lóbulo temporal, generando declive cognitivo y formación de agregados TAU. Entretanto, la sobreexpresión de FAF1 en la corteza frontal se asocia con la enfermedad de Parkinson y demencia derivada por Alzheimer. Por su parte, García-Fuster *et al* (157) indican que la disminución en la expresión de FAF1 se relaciona con la formación de placas beta amiloide y pérdida de terminales presinápticas. De la misma manera, Anderson *et al* (161) vinculan a esta proteína con procesos de neurodegeneración, disfunción dendrítica y muerte neuronal en demencia generada por Alzheimer. Estos ejemplos, muestran que la pérdida de la homeostasis de FAF1 repercute en declive cognitivo. Por esta razón, se suscita la participación del polimorfismo rs200133978 en el progreso de demencia vascular.

El declive cognitivo no es solo el efecto de la pérdida neuronal. Sin embargo, es la consecuencia de cambios que se presentan a través del tiempo, incluyendo disfunción de la barrera hematoencefálica y daño a la UNV (157). De la misma manera, el inicio y desarrollo de enfermedades crónico-degenerativas depende de factores genéticos y ambientales. En cuanto a la información genética, esta resulta de la sinergia de variaciones en diferentes sitios del genoma (139). El conjunto de esta información así como ensayos posteriores permitirán construir el marco de sucesos presentes en la patogénesis del declive cognitivo, describiendo cómo el papel genético repercute a nivel celular, dando como resultado la persistencia del cuadro clínico que acompaña el desarrollo de demencia vascular.

Conclusión

La información previa muestra 12 polimorfismos presentes en demencia vascular y aterosclerosis. Las variantes reportadas se encontraron principalmente en cromosomas somáticos. Estas presentaron una frecuencia alélica fluctuante en cada una de las 12 subpoblaciones continentales. De manera interesante se identificaron dos SNPs en poblaciones maya y pima. No obstante, se presentó el inconveniente de las diferencias metodológicas entre el número de variantes analizadas en los microarreglos o biopaneles diseñados y usados para los proyectos 1000 genomas y Human Genome Diversity Project. Por lo anterior no fue posible realizar la comparación con el resto de los polimorfismos.

De la misma manera, la red de interacción conformada por vías metabólicas, estirpes celulares, polimorfismos y genes evoca a los oligodendrocitos como principal componente en el desarrollo de demencia vascular. De hecho, se refuerza esta premisa con información publicada, donde se describe que los oligodendrocitos son una estirpe celular que presenta degeneración durante etapas tempranas del desarrollo de la enfermedad y periodos prolongados de hipoxia. Asimismo, se proponen dos polimorfismos no sinónimos con efecto negativo sobre la estructura y función proteica. Por tal motivo, se propone realizar ensayos para evaluar el efecto de las variantes en los oligodendrocitos y su papel en el asentamiento del cuadro clínico y desarrollo de la demencia vascular.

Perspectivas

Anteriormente se mencionó que no fue posible comparar completamente los polimorfismos encontrados con las subpoblaciones del proyecto Human Genome Diversity Project, esto debido a problemas metodológicos. No obstante, mediante el desequilibrio de ligamiento es posible ampliar la información presente en Human Genome Diversity Project, haciendo una imputación de las variantes presentes en cada población. De esta manera, sería posible definir si existen bloques de polimorfismos similares en las poblaciones de Human Genome Diversity Project y 1000 genomas. Con la restricción de que este proceso deberá realizarse considerando el mismo genoma de referencia.

Se sugiere evaluar las estirpes celulares mediante ensayos de célula única. De esta forma se determinará si el mismo estirpe celular está presente en el desarrollo de la patología o existen variantes de la misma célula durante este evento.

Este proyecto plantea experimentos con la promesa de resolver disyuntivas encontradas que amplíen la información al definir el mecanismo fisiopatológico tales como el análisis de expresión de oligodendrocitos en estado basal y diferentes estadios de demencia vascular, determinación de regulación transcripcional, transcriptómica espacial, mutaciones sitio dirigidas para evaluar el efecto en la proteínas, entre otros.

Por último, se propone determinar el efecto de los polimorfismos sobre la función proteica mediante ensayos de acoplamiento molecular. Esto para puntualizar el papel de las variantes sobre la proteína. Lo cual generará información para determinar el papel de los polimorfismos en el desarrollo del cuadro clínico de demencia vascular.

Limitaciones del proyecto

Dentro de los límites de este proyecto se encuentra la ausencia de explicación de la heredabilidad de la demencia vascular a partir del uso de polimorfismos de un solo nucleótido. La heredabilidad en sentido amplio (H^2) mide la varianza fenotípica en una población (σ^2P) explicada por la varianza genética (σ^2G). De esta forma, podemos definir a la heredabilidad en sentido amplio con la siguiente función: $H^2 = \sigma^2G/\sigma^2P$. Por su parte, la heredabilidad en sentido estricto (h^2) se deriva de la relación de cocientes entre la varianza genotípica aditiva (σ^2A) sobre la varianza fenotípica (σ^2P). Desarrollando la siguiente función: $h^2 = \sigma^2A/\sigma^2P$ (162)(163)(164)(165)(138). La heredabilidad presenta un rango entre 0 y 1, si el valor se acerca a la unidad significa que la mayoría de variación observada de un fenotípico en específico está dada por cuestiones genéticas. De esta manera, se puede explicar si la diferencia entre individuos en el desarrollo de una patología depende de cuestiones ambientales o genéticas (48). Los efectos aditivos de los polimorfismos derivados de estudios de asociación genómica completa no explican de manera concisa la heredabilidad de enfermedades complejas, a este proceso se le ha denominado heredabilidad perdida (166)(138)(167). Esta discrepancia puede deberse a que la mayoría de los GWAS se basan en la detección de variantes mediante microarreglos, permitiendo estudiar solo una fracción del genoma (168). De esta manera se descartan aquellos polimorfismos con una frecuencia alélica menor al 5% o variantes raras cuya frecuencia alélica suele ser menor al 1% (164)(138). Por lo cual se sugiere que la evaluación del genoma completo permitirá identificar las variaciones genéticas que explican la heredabilidad de un fenotipo (163). Para la mayoría de los ensayos, la mayor parte de la heredabilidad permanece sin explicaciones (162). No obstante, el desarrollo de metodologías basadas en epigenética, epistasis, desequilibrio de ligamiento, búsqueda de variación en el número de copias o microsatélites (166)(167)(164) permitirán determinar esta variación genética y su papel dentro de la susceptibilidad para el desarrollo de enfermedades (165). Asimismo, este proyecto no describe la interacción del microbioma en el desarrollo de demencia vascular. Artículos previos denotan que existe una relación entre el microbioma humano y la presencia de un fenotipo en específico, permitiendo que el microbioma juegue un papel como factor de riesgo o protector en el desarrollo de enfermedades complejas (167)(169)(170)(171)(172). De hecho, se detalla que la diversidad del microbioma humano puede ser la clave para responder el problema de la heredabilidad perdida. Por lo cual, agregar la información de estudios de asociación metagenómica (MWAS, por sus siglas en inglés) en relación con un fenotipo en combinación a los estudios de asociación genómica completa puede generar un cálculo específico en el concepto de heredabilidad en diversos fenotipos (167). No obstante, los objetivos de este proyecto se limitan a identificar la relación de variantes genéticas en el desarrollo de demencia vascular a partir de aterosclerosis. Por lo cual no responden estas cuestiones. Sin embargo, es importante mencionarlas ya que la integración de esta información mejoraría el abordaje de enfermedades complejas.

Consideraciones éticas

Debido a que la mayoría de los estudios de asociación genómica completa se han realizado en poblaciones de Europa y Asia, existe una limitante en la información de variantes asociadas a una patología en poblaciones amerindias. Asimismo, se remarca que la información genética relacionada de población mexicana en el proyecto 1000 genomas fase 3 se encuentran definida a personas con ascendencia mexicana que radican en Los ángeles, EE. UU. Para contrarrestar esta carencia de información se agregaron las subpoblaciones pima (Chihuahua, México) y maya (Sur de México), con el objetivo de determinar si las variantes asociadas a demencia vascular y aterosclerosis derivadas de GWAS se presentaban en estas regiones. Bajo ningún sentido se desea discriminar a estas poblaciones originarias, ni generar argumentos sobre mayor susceptibilidad al desarrollo de estas patologías.

Anexo

Desequilibrio de ligamiento

En este segmento se ejemplificará cómo las funciones descritas en el apartado de desequilibrio de ligamiento permiten determinar si existe relación de dos polimorfismos y su probabilidad de presentarse en conjunto en la siguiente generación. En ambos casos se mostrará una cohorte conformada por información de diez individuos, donde el objetivo será definir el desequilibrio de ligamiento de los polimorfismos *A* y *B*. Los resultados de la primera población se muestran en la figura A1.

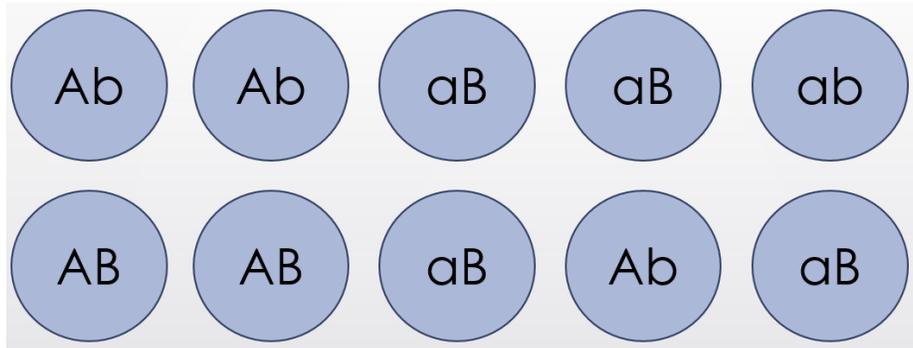


Figura A1. Genotipos de población 1.

La determinación del desequilibrio de ligamiento comienza por encontrar el valor de *D* de las variantes *A* y *B*, el cual se obtiene a través de la siguiente función:

$$D_{AB} = p_{AB} - p_A p_B$$

En primer lugar se calcula la frecuencia de ambos alelos en la misma persona. En este caso, solo se encontraron dos individuos que presentan *AB*. Posteriormente, se obtiene la frecuencia del alelo *A* y *B* por separado en la población. La función se modifica de la siguiente manera:

$$D_{AB} = 0.2 - (0.5 * 0.6)$$

De esta forma, se determina que el valor de *D* para la primer población es de -0.1. No obstante, este resultado varía en relación con el tamaño de la cohorte. Por lo que se obtendrá el valor de *D'* y *r*². Para *D'* se describe la siguiente función:

$$D'_{AB} = \left\{ \begin{array}{l} \frac{D_{AB}}{\min(p_{AB}[1-p_B], [1-p_A]p_B)} \quad \forall D_{AB} > 0 \\ \frac{D_{AB}}{\max(p_A p_B, [1-p_A][1-p_B])} \quad \forall D_{AB} < 0 \end{array} \right\}$$

Donde el valor de D determina la metodología a seguir. En este caso al tener un resultado de -0.1 se utilizará la función inferior. En este apartado se obtendrá la frecuencia de encontrar ambos alelos dividido entre el valor máximo de las frecuencias alélicas. Continuando con el ejemplo, el valor máximo de las frecuencias está determinado por p_A y p_B . La función de D' queda de la siguiente manera:

$$D'_{AB} = \frac{0.2}{0.5 * 0.6}$$

Obteniendo un valor de D' igual a 0.66 . Posteriormente se buscará el resultado del coeficiente de correlación, el cual se describe a continuación:

$$r^2 = \frac{D^2}{p_A(1 - p_A)p_B(1 - p_B)}$$

Se comienza con elevar al cuadrado el valor de D . Posteriormente, se colocan los valores de las frecuencias alélicas de A y B . Por último se divide el valor de D^2 entre el resultado derivado de las frecuencias de A y B . Se colocan los valores correspondientes, obteniendo lo siguiente:

$$r^2 = \frac{-0.1^2}{0.5(1 - 0.5) 0.6(1 - 0.6)}$$

Es así como se denota un coeficiente de correlación de 0.16 . En conjunto, los datos obtenidos en la población 1 indican poca relación entre las variantes A y B . Sugiriendo que no se heredan en conjunto. A continuación se muestra la figura A2 con los resultados de la segunda población a evaluar. En este ejemplo continuaremos con la descripción de las métricas de desequilibrio de ligamiento para los alelos A y B .

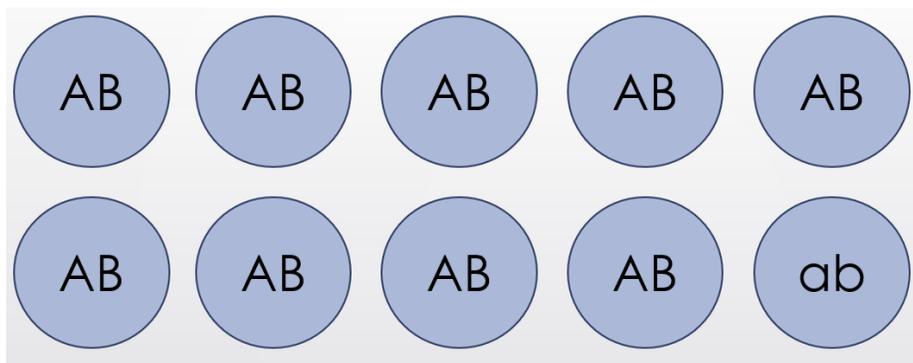


Figura A2. Genotipos de población 2.

De la misma manera, se comenzará con la búsqueda del valor de D :

$$D_{AB} = 0.9 - (0.9 * 0.9)$$

Se obtiene un valor de D igual 0.09. Por tal motivo, para determinar D' , se utilizará la función que corresponde cuando D es mayor a 0. Quedando la función como se describe a continuación:

$$D'_{AB} = \frac{0.09}{0.9 * 0.1}$$

El resultado es $D' = 1$. Por último se obtendrá el coeficiente de correlación. Al intercambiar los valores de las frecuencias alélicas se obtiene lo siguiente:

$$r^2 = \frac{-0.09^2}{0.9(1 - 0.9) 0.9(1 - 0.9)}$$

El valor del coeficiente de correlación es 1 al igual que D' . En conjunto esta información indica que en la población 2 los alelos A y B se heredan en bloque a la siguiente generación. A partir del desequilibrio de ligamiento se determinan los bloques de variantes presentes en cada sitio geográfico.

Polyphen2

En la tabla A3 se muestran las 8 características de secuencia y 3 estructurales que conforman el conjunto de datos en el cual se basa el algoritmo PolyPhen2 (173).

Característica	Definición	Tipo
Score1	Puntaje del alelo silvestre mediante matriz de sustitución	Secuencia
Score_delta	Diferencia de puntajes entre alelo silvestre y alelo con mutación mediante matriz de sustitución	Secuencia
Num_observ	Numero de residuos observados en la posición del alineamiento múltiple	Secuencia
Delta_volume	Cambio en el volumen de la cadena lateral del residuo	Secuencia
Pfam_hit	Posición de la mutación dentro/fuera de un dominio de proteína definido por Pfam	Secuencia
Id_p_max	Congruencia de la mutación en múltiples alineamientos	Secuencia
Id_q_min	Identidad de secuencia con el homólogo más cercano	Secuencia
Cpg_transition	Si la variante ocurre en el contexto de una transición CpG (regiones citosina-guanina)	Secuencia
Acc_normed	Área de superficie accesible del residuo	Estructural
B_fact	Factor beta de cristalografía	Estructural
Delta_prop_new	Cambio en el área de superficie accesible para residuos fuera de superficie	Estructural

Tabla A3. Características estructurales y de secuencia utilizadas en PolyPhen2.

Desequilibrio de ligamiento

En este apartado se muestra el resto de los resultados de disequilibrio de ligamiento de los polimorfismos asociados a demencia vascular y aterosclerosis.

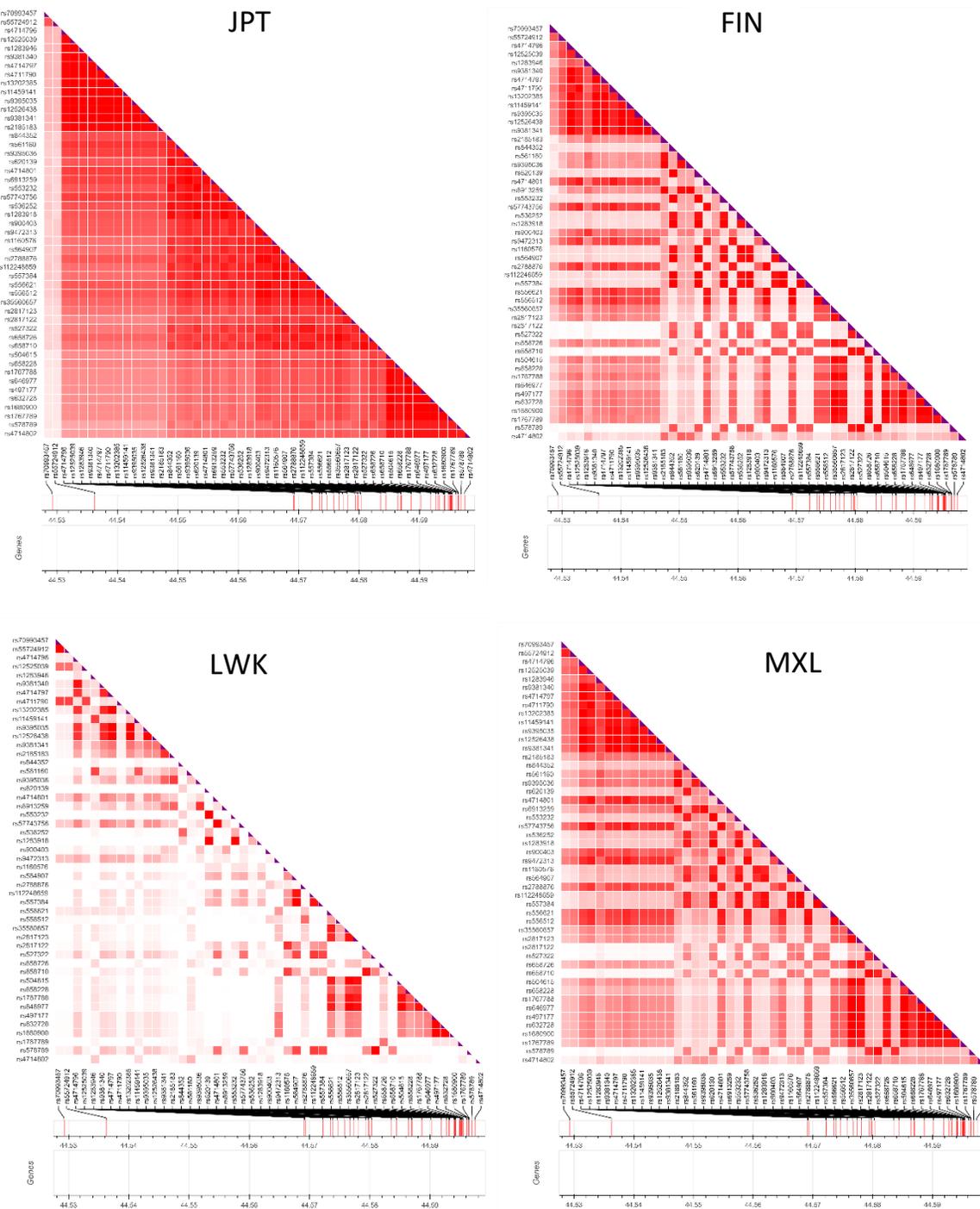


Figura A4. Disequilibrio de ligamiento de polimorfismos asociados a rs556621. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

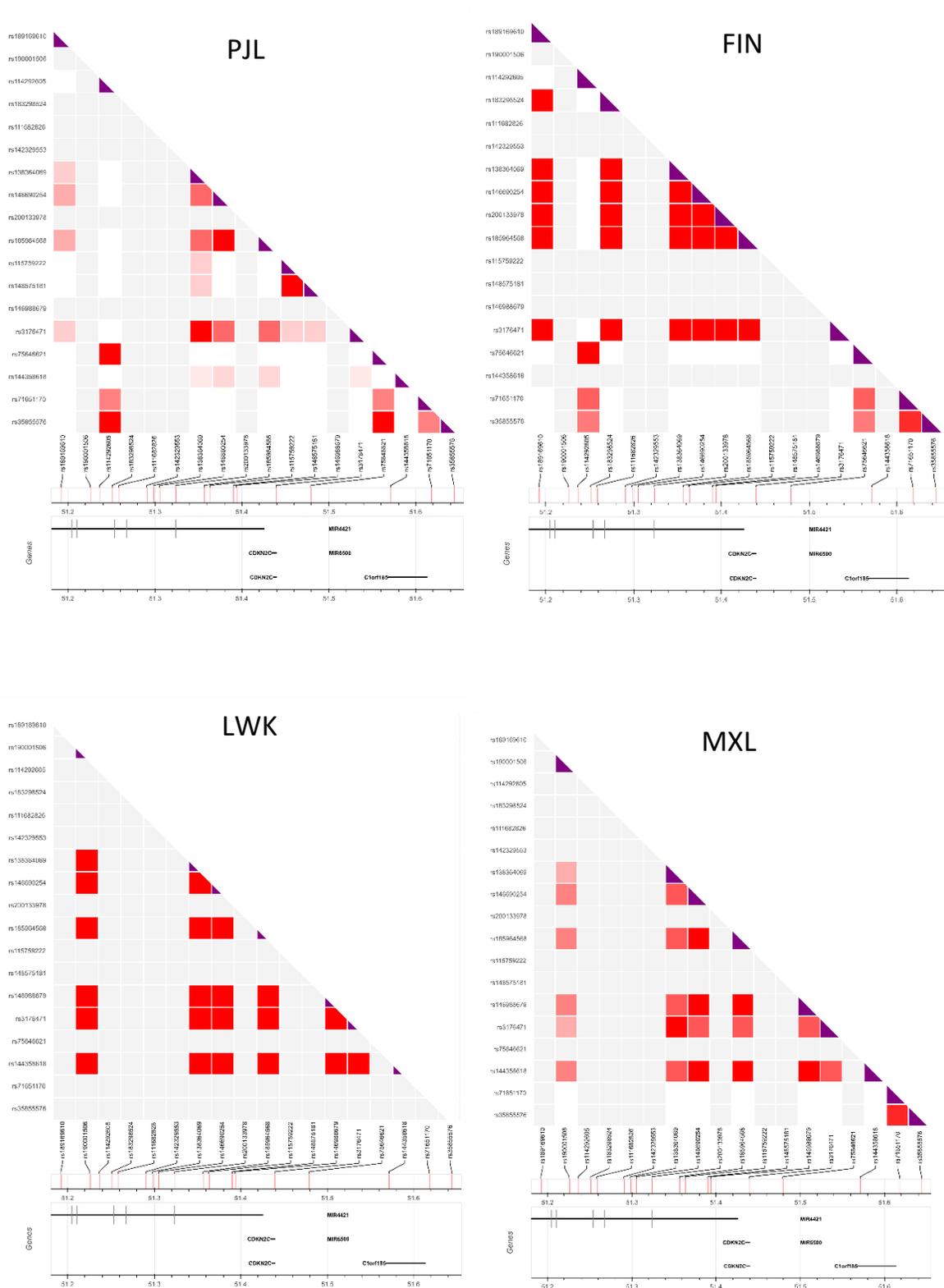


Figura A5. Desequilibrio de ligamiento de polimorfismos asociados a rs3176471. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

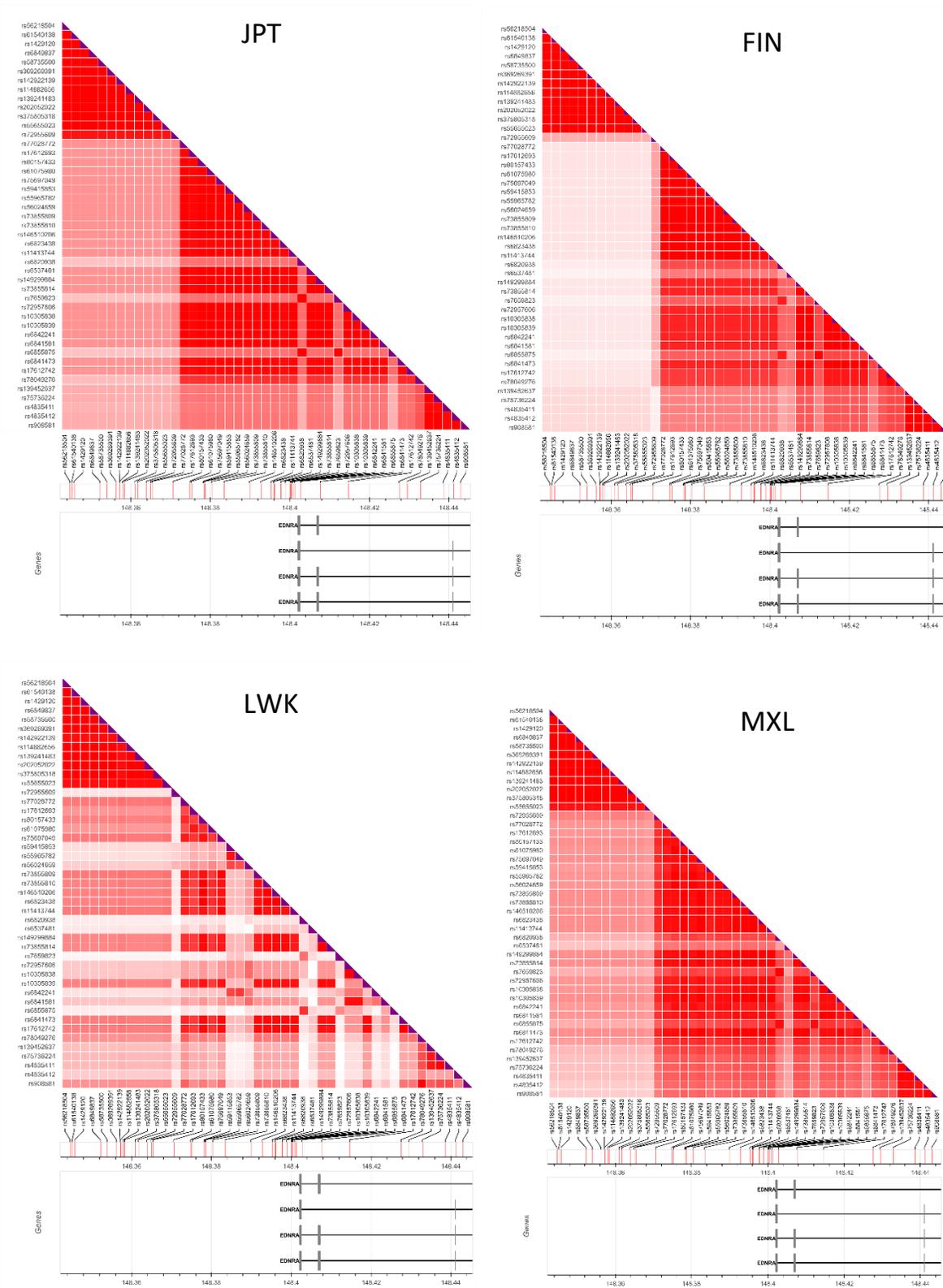


Figura A6. Desequilibrio de ligamiento de polimorfismos asociados a rs17612742. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

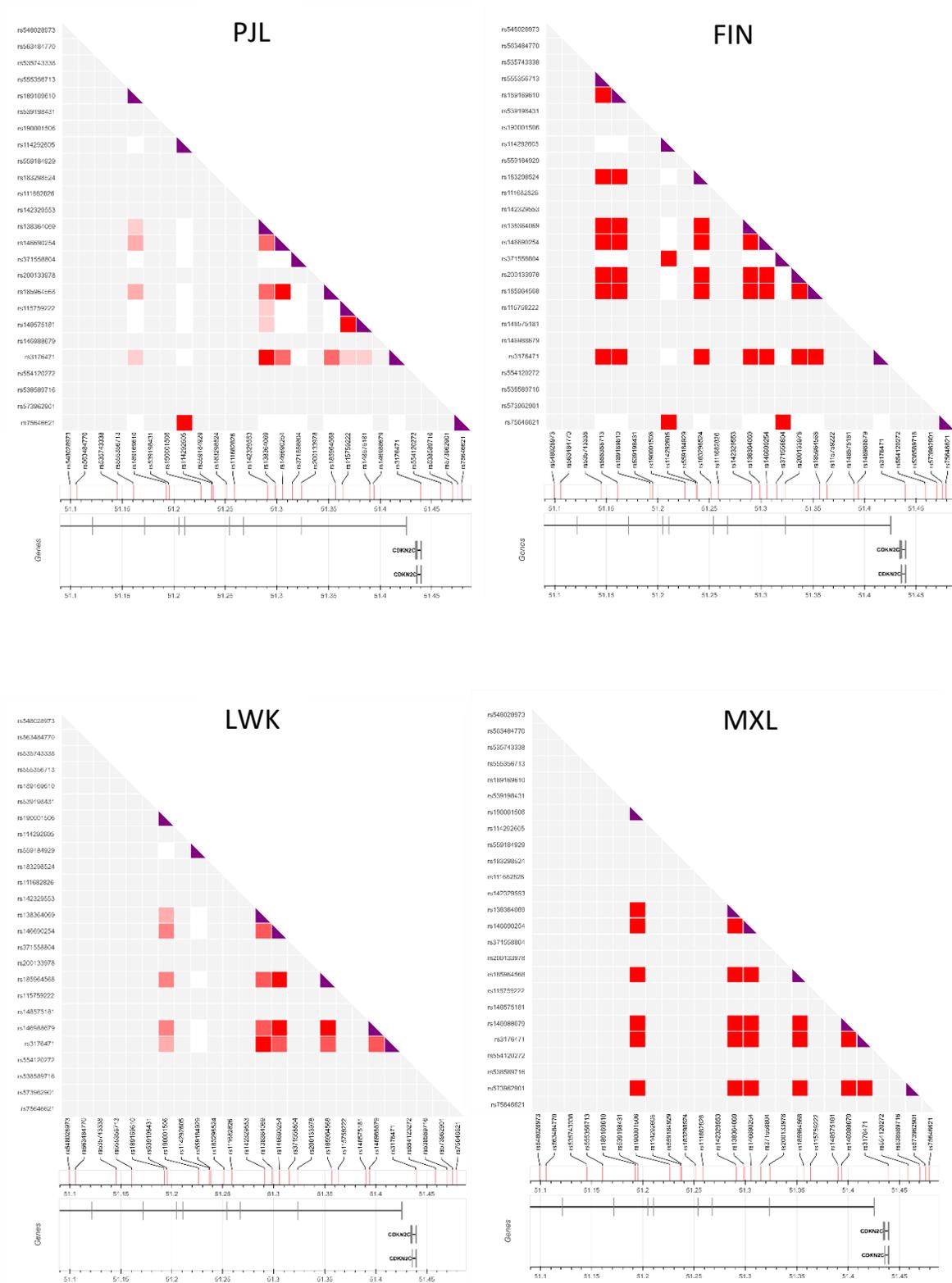


Figura A7. Desequilibrio de ligamiento de polimorfismos asociados a rs138364069. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

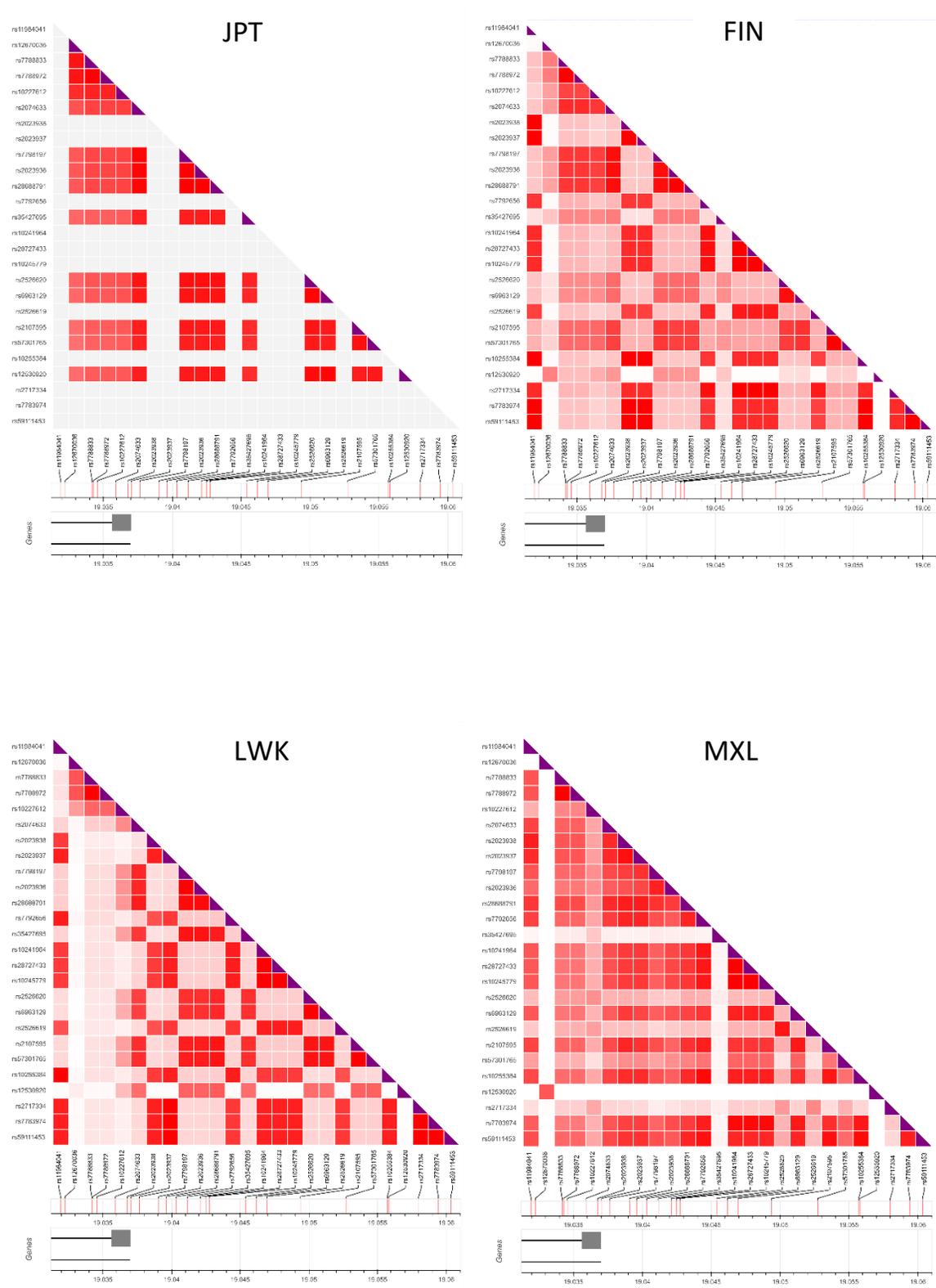


Figura A8. Desequilibrio de ligamiento de polimorfismos asociados a rs2107595. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

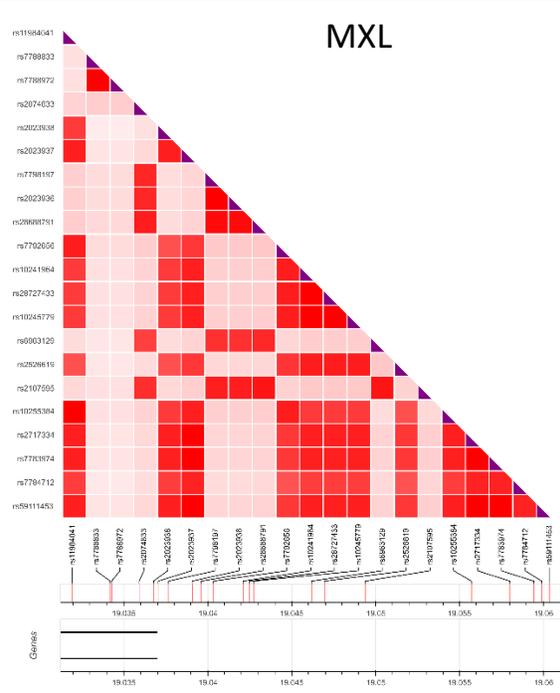
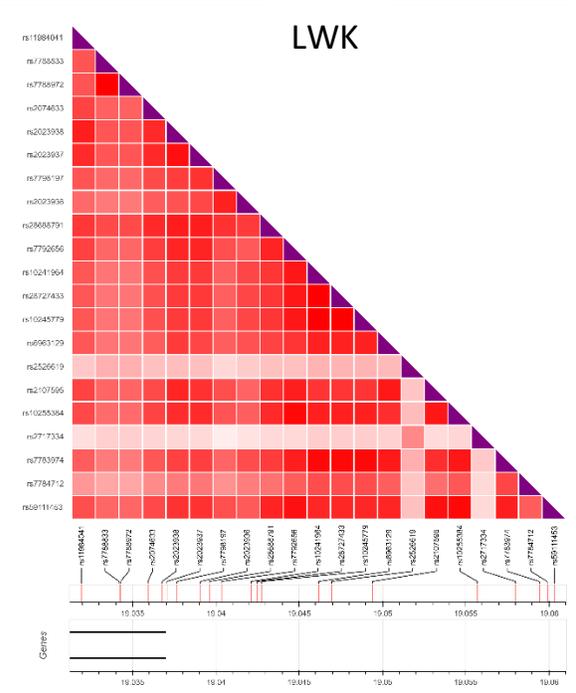
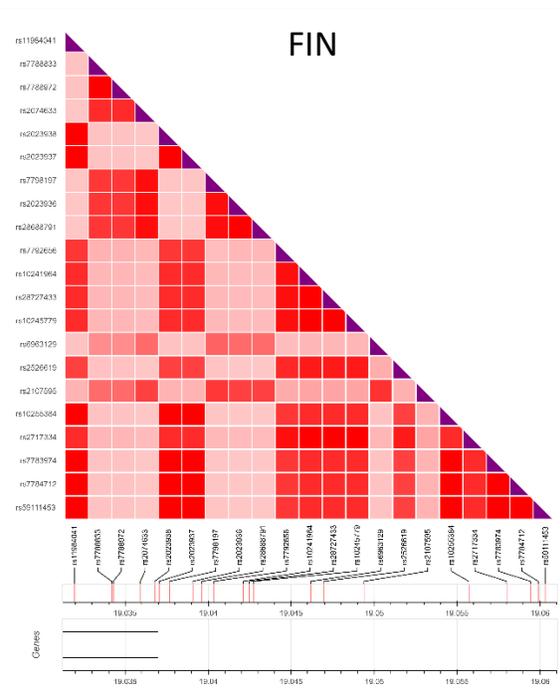
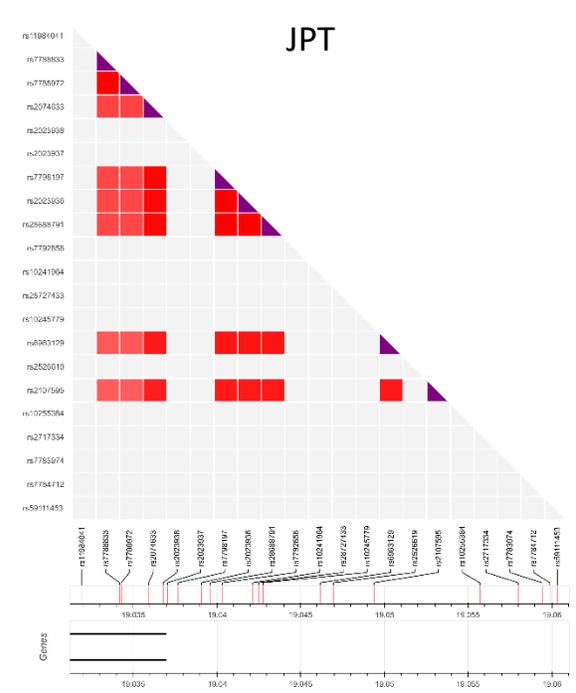


Figura A9. Desequilibrio de ligamiento de polimorfismos asociados a rs11984041. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

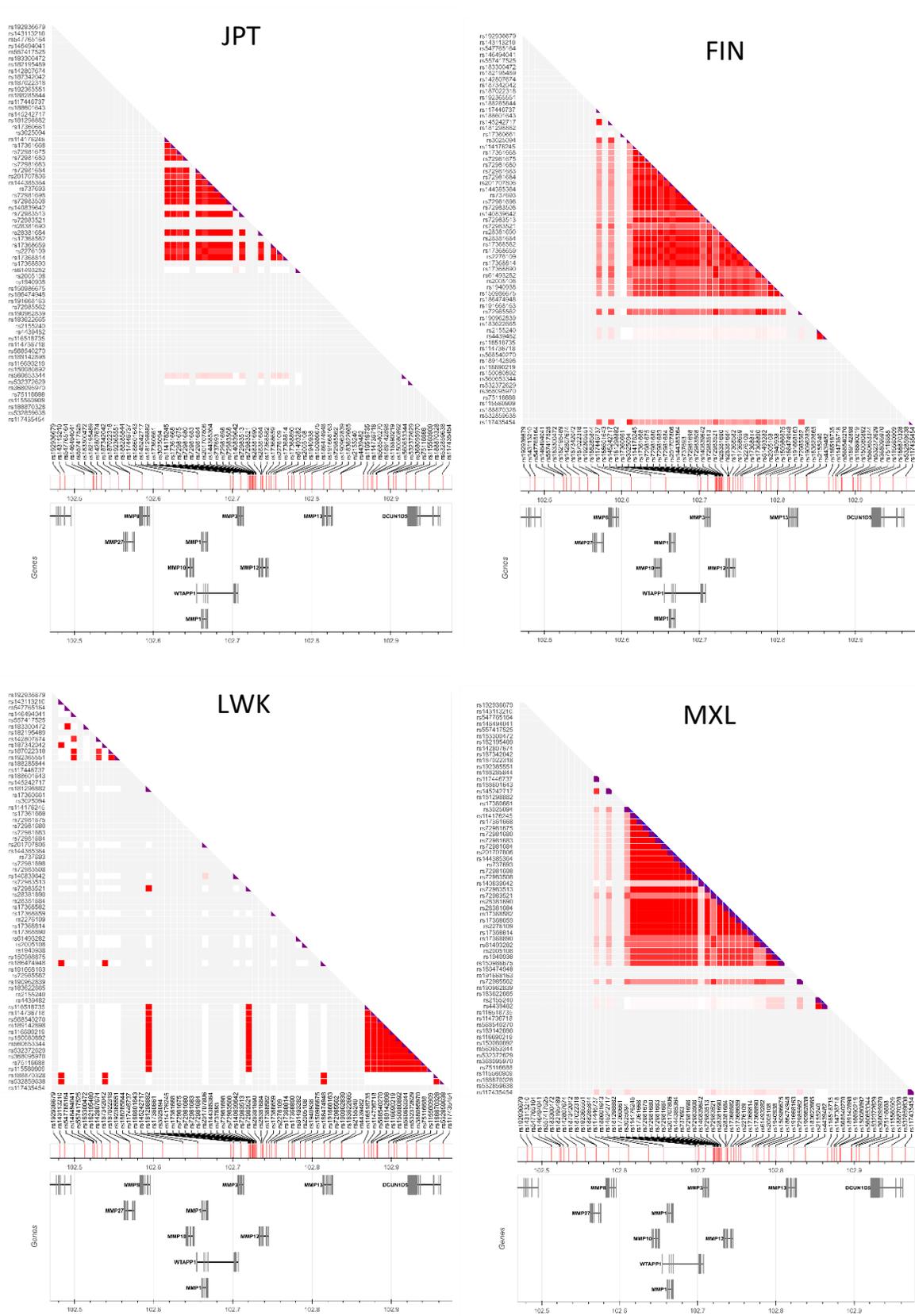


Figura A10. Desequilibrio de ligamiento de polimorfismos asociados a rs7298321. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

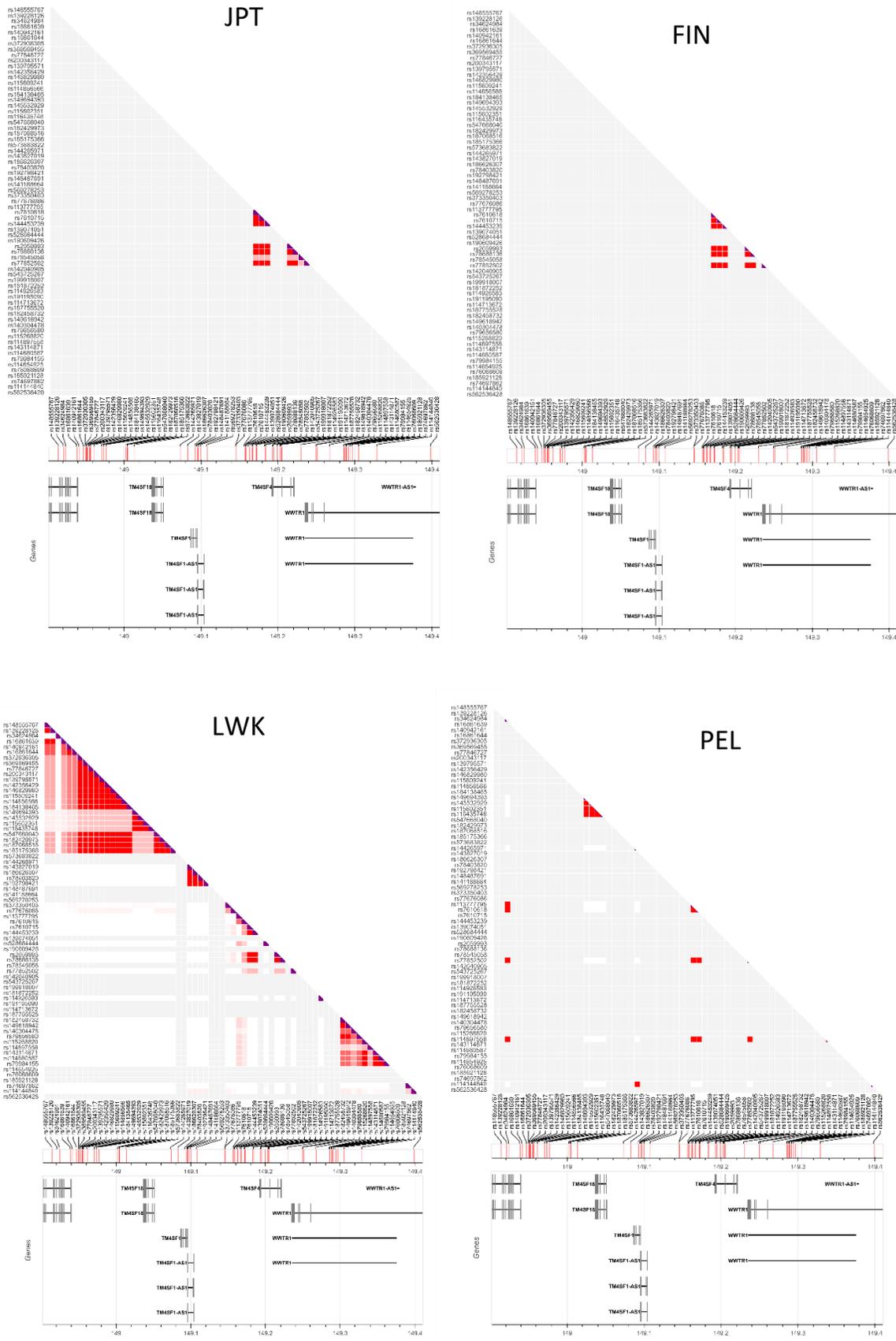


Figura A12. Desequilibrio de ligamiento de polimorfismos asociados a rs7610618. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

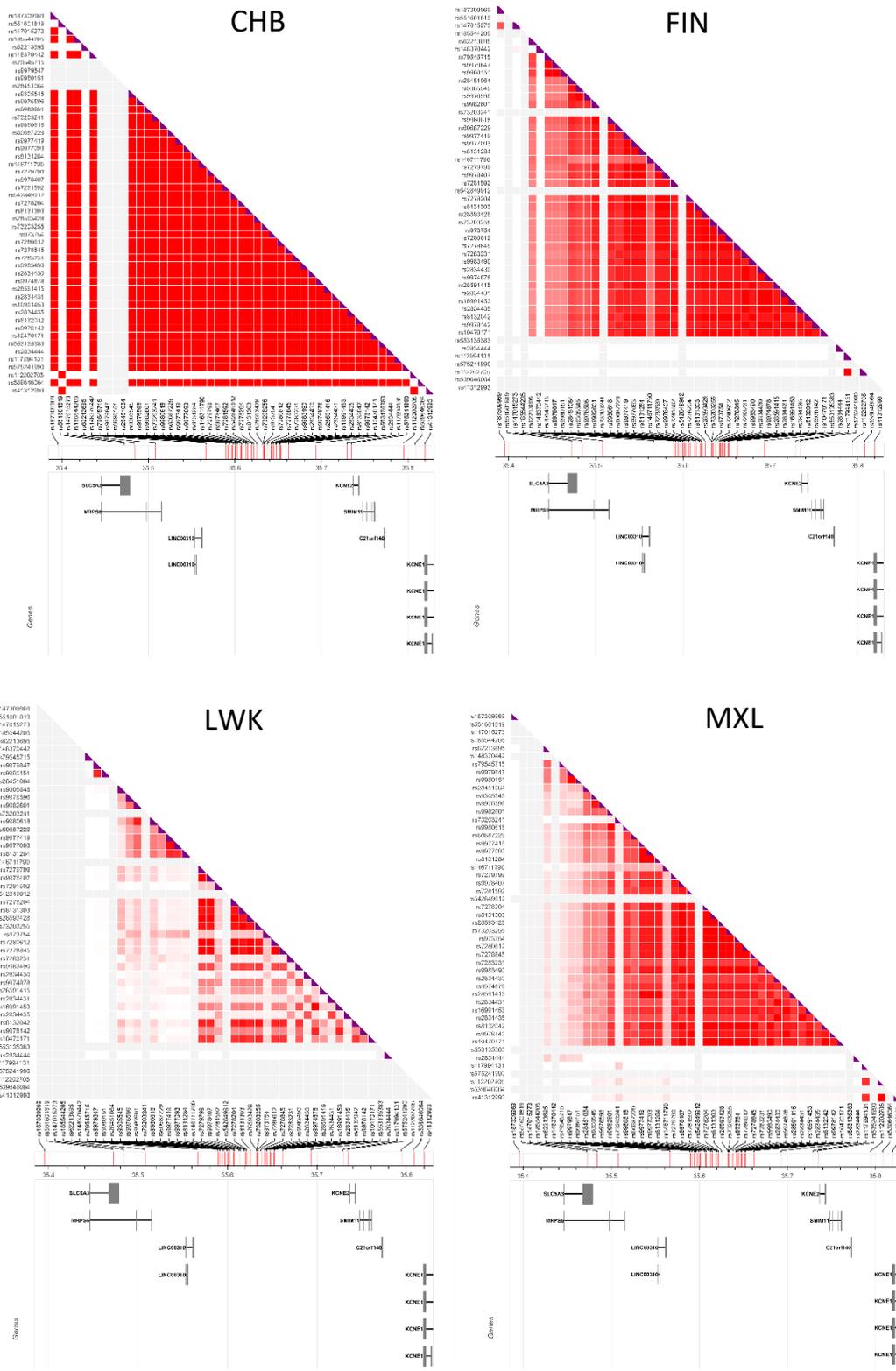


Figura A13. Desequilibrio de ligamiento de polimorfismos asociados a rs9982601. Se detallan los bloques de polimorfismos presentes en las 12 subpoblaciones. Igualmente se presenta la arquitectura cromosómica.

Abreviaturas

ACMG: Colegio Americano de Medicina Genética y Genómica (*American College of Medical Genetics and Genomics*)

BHE: barrera hematoencefálica

DL: desequilibrio de ligamiento

dnSNP: Base de Datos de SNP (data base SNP)

GWAS: estudio de asociación genómica completa (*Genome-Wide Association Studies*)

HGDP: Proyecto de Diversidad del Genoma Humano (*Human Genome Diversity Project*)

MAF: frecuencia alélica menor (*minor allele frequency*)

MeSH: encabezado de temas médicos (*medical subject heading*)

MP: medicina de precisión

NCBI: Centro Nacional para Información Biotecnológica (*National Center for Biotechnology Information*)

NLM: Biblioteca Nacional de Medicina (*National Library of Medicine*)

OMIM: Herencia Mendeliana en el Hombre en Línea (*Online Mendelian Inheritance in Man*)

OMS: Organización Mundial de la Salud

SCRAD: Clasificación de SNP-Células y evaluación del riesgo de demencia (*SNP-Cell Classification and Risk Assessment of Dementia*)

SNP: polimorfismo de un solo nucleótido (*Single Nucleotide Polymorphism*)

SVA: Selección de Variantes Automatizada

UniProt: Recurso Universal de Proteína (*Universal Protein Resource*)

UNV: unidad neurovascular

XGBoost: Aumento de Gradiente Extremo (*Extreme Gradient Boosting*)

Bibliografía

1. Wang F, Cao Y, Ma L, Pei H, Rausch WD, Li H. Dysfunction of Cerebrovascular Endothelial Cells: Prelude to Vascular Dementia. *Front Aging Neurosci.* 2018;10(November):1–23.
2. Kim JH, Ko PW, Lee HW, Jeong JY, Lee MG, Kim JH, et al. Astrocyte-derived lipocalin-2 mediates hippocampal damage and cognitive deficits in experimental models of vascular dementia. *Glia.* 2017;65(9):1471–90.
3. O'Brien JT, Thomas A. Non-Alzheimer's dementia 3 Vascular dementia. *Lancet* [Internet]. 2015;386(10004):1698–706. Available from: [http://dx.doi.org/10.1016/S0140-6736\(15\)00463-8](http://dx.doi.org/10.1016/S0140-6736(15)00463-8)
4. Aziz Director M. Pathogenesis of Atherosclerosis. *Med Clin Rev* [Internet]. 2016 [cited 2018 Mar 13]; Available from: <http://www.imedpub.com>
5. Herrington W, Lacey B, Sherliker P, Armitage J, Lewington S. Epidemiology of Atherosclerosis and the Potential to Reduce the Global Burden of Atherothrombotic Disease. *Circ Res.* 2016;118(4):535–46.
6. Silverstein RL. Inflammation, atherosclerosis, and arterial thrombosis: Role of the scavenger receptor CD36. *Natl Institutes Heal* [Internet]. 2009 [cited 2018 Mar 5]; Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2810530/pdf/nihms-168654.pdf>
7. Lee Y, Laxton V, Lin H, Chan Y, Fitzgerald-Smith S, To TLO, et al. Animal models of atherosclerosis (Review). *Biomed Reports.* 2017;
8. Badimon L, Padró T, Vilahur G. Atherosclerosis, platelets and thrombosis in acute ischaemic heart disease. *Eur Hear J Acute Cardiovasc Care* [Internet]. 2012;1(1):60–74. Available from: <http://journals.sagepub.com/doi/10.1177/2048872612441582>
9. McKenna MC, Dienel GA, Sonnewald U, Waagepetersen HS, Schousboe A. Energy Metabolism of the Brain. *Basic Neurochem.* 2012;200–31.
10. Iadecola C. The Pathobiology of Vascular Dementia. *Neuron* [Internet]. 2013;80(4):844–66. Available from: <http://dx.doi.org/10.1016/j.neuron.2013.10.008>
11. Bell AH, Miller SL, Castillo-Melendez M, Malhotra A. The Neurovascular Unit: Effects of Brain Insults During the Perinatal Period. *Front Neurosci.* 2020;13(January):1–19.
12. Muoio V, Persson PB, Sendeski MM. The neurovascular unit - concept review. *Acta Physiol.* 2014;210(4):790–8.
13. Emery B. Regulation of oligodendrocyte differentiation and myelination. *Science* (80-). 2010;330(6005):779–82.
14. Bergles DE, Richardson WD. Oligodendrocyte development and plasticity. *Cold Spring Harb Perspect Biol.* 2016;8(2):1–28.
15. Simons M, Nave KA. Oligodendrocytes: Myelination and axonal support. *Cold Spring Harb Perspect Biol.* 2016;8(1):1–16.
16. Toth P, Tarantini S, Csiszar A, Ungvari Z. Functional vascular contributions to cognitive impairment and dementia: Mechanisms and consequences of cerebral autoregulatory dysfunction, endothelial impairment, and neurovascular uncoupling in aging. *Am J Physiol - Hear Circ Physiol.* 2017;312(1):H1–20.
17. Hachinski V, Einhäupl K, Ganten D, Alladi S, Brayne C, Stephan BCM, et al. Preventing dementia by preventing stroke: The Berlin Manifesto. *Alzheimer's Dement.* 2019;15(7):961–84.

18. Kalaria RN. The pathology and pathophysiology of vascular dementia. *Neuropharmacology* [Internet]. 2018;134:226–39. Available from: <https://doi.org/10.1016/j.neuropharm.2017.12.030>
19. Smith EE. Clinical presentations and epidemiology of vascular dementia. *Clin Sci*. 2017;131(11):1059–68.
20. Rizzi L, Rosset I, Roriz-Cruz M. Global epidemiology of dementia: Alzheimer's and vascular types. *Biomed Res Int*. 2014;2014(Figure 1).
21. Gutiérrez-Robledo LM, Arrieta-Cruz I. Demencias en México: La necesidad de un Plan de Acción. *Gac Med Mex*. 2015;151(5):667–73.
22. Kalaria RN, Akinyemi R, Ihara M. Stroke injury, cognitive impairment and vascular dementia. *Biochim Biophys Acta - Mol Basis Dis* [Internet]. 2016;1862(5):915–25. Available from: <http://dx.doi.org/10.1016/j.bbadis.2016.01.015>
23. Biessels GJ. *Biochimica et Biophysica Acta* Diagnosis and treatment of vascular damage in dementia ☆. *BBA - Mol Basis Dis* [Internet]. 2016;1862(5):869–77. Available from: <http://dx.doi.org/10.1016/j.bbadis.2015.11.009>
24. Ibañez B, Badimon JJ, Garcia MJ. Diagnosis of Atherosclerosis by Imaging. *Am J Med* [Internet]. 2009;122(1 SUPPL.):S15–25. Available from: <http://dx.doi.org/10.1016/j.amjmed.2008.10.014>
25. Boudoulas KD, Triposciadis F, Geleris P, Boudoulas H. Coronary Atherosclerosis: Pathophysiologic Basis for Diagnosis and Management. *Prog Cardiovasc Dis*. 2016;58(6):676–92.
26. Paradise MB, Sachdev PS. Vascular Cognitive Disorder. *Semin Neurol*. 2019;39(2):241–50.
27. Antman EM, Loscalzo J. Precision medicine in cardiology. *Nat Rev Cardiol* [Internet]. 2016;13(10):591–602. Available from: <http://dx.doi.org/10.1038/nrcardio.2016.101>
28. Krönig. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med* [Internet]. 2016;15(6):894–7. Available from: <http://dx.doi.org/10.1186/s13073-016-0388-7>
29. Carrasco-Ramiro F, Peiró-Pastor R, Aguado B. Human genomics projects and precision medicine. *Gene Ther*. 2017;24(9):551–61.
30. Perkins BA, Caskey CT, Brar P, Dec E, Karow DS, Kahn AM, et al. Precision medicine screening using whole-genome sequencing and advanced imaging to identify disease risk in adults. *Proc Natl Acad Sci U S A*. 2018;115(14):3686–91.
31. Yun S-H, Sim E-H, Goh R-Y, Park J-I, Han J-Y. Platelet Activation: The Mechanisms and Potential Biomarkers. *Biomed Res Int* [Internet]. 2016 [cited 2018 Feb 18]; Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4925965/pdf/BMRI2016-9060143.pdf>
32. König IR, Fuchs O, Hansen G, von Mutius E, Kopp M V. What is precision medicine? *Eur Respir J* [Internet]. 2017;50(4):1–12. Available from: <http://dx.doi.org/10.1183/13993003.00391-2017>
33. Ramaswami R, Bayer R, Galea S. Precision Medicine from a Public Health Perspective. *Annu Rev Public Health*. 2018;39(1):153–68.
34. Lam SM, Shui G. Lipidomics as a Principal Tool for Advancing Biomedical Research. *J Genet Genomics* [Internet]. 2013;40(8):375–90. Available from: <http://dx.doi.org/10.1016/j.jgg.2013.06.007>
35. Weaver O, Leung JWT. Biomarkers and imaging of breast cancer. *Am J Roentgenol*.

2018;210(2):271–8.

36. Lin Y, Qian F, Shen L, Chen F, Chen J, Shen B. Computer-aided biomarker discovery for precision medicine: Data resources, models and applications. *Brief Bioinform.* 2017;20(3):952–75.
37. Sethi S, Brietzke E. Recent advances in lipidomics: Analytical and clinical perspectives. *Prostaglandins Other Lipid Mediat* [Internet]. 2017;128–129:8–16. Available from: <http://dx.doi.org/10.1016/j.prostaglandins.2016.12.002>
38. Rolim AEH, Henrique-Araújo R, Ferraz EG, de Araújo Alves Dultra FK, Fernandez LG. Lipidomics in the study of lipid metabolism: Current perspectives in the omic sciences. *Gene* [Internet]. 2015;554(2):131–9. Available from: <http://dx.doi.org/10.1016/j.gene.2014.10.039>
39. Katus HA, Giannitsis E. Biomarker in cardiology: DGK welcomes ESC Munich 2018. *Clin Res Cardiol* [Internet]. 2018;107(2):10–5. Available from: <http://dx.doi.org/10.1007/s00392-018-1300-9>
40. Chen X, Sullivan PF. Single nucleotide polymorphism genotyping: Biochemistry, protocol, cost and throughput. *Pharmacogenomics J.* 2003;3(2):77–96.
41. Tibaut M, Caprnda M, Kubatka P, Sinkovič A, Valentova V, Filipova S, et al. Markers of Atherosclerosis: Part 2 – Genetic and Imaging Markers. *Hear Lung Circ.* 2019;28(5):678–89.
42. Twyman RM. Single-Nucleotide Polymorphism (SNP) Analysis. *Encycl Neurosci.* 2009;871–5.
43. LaFramboise T. Single nucleotide polymorphism arrays: A decade of biological, computational and technological advances. *Nucleic Acids Res.* 2009;37(13):4181–93.
44. Dutt A, Beroukhim R. Single nucleotide polymorphism array analysis of cancer. *Curr Opin Oncol.* 2007;19(1):43–9.
45. Guest PC (Ed. . Multiplex Biomarker Techniques. *Methods Mol Biol* [Internet]. 2017;1546:317 pp. Available from: <http://www.springer.com/gb/book/9781493967292>
46. Chaudhary R, Singh B, Kumar M, Gakhar SK, Saini AK, Parmar VS, et al. Role of single nucleotide polymorphisms in pharmacogenomics and their association with human diseases. *Drug Metab Rev* [Internet]. 2015;47(3):281–90. Available from: <http://dx.doi.org/10.3109/03602532.2015.1047027>
47. Carlton VEH, Ireland JS, Useche F, Faham M. Functional single nucleotide polymorphism-based association studies. *Hum Genomics.* 2006;2(6):391–402.
48. Visscher PM, Hill WG, Wray NR. Heritability in the genomics era - Concepts and misconceptions. *Nat Rev Genet.* 2008;9(4):255–66.
49. Fareed M, Afzal M. Single nucleotide polymorphism in genome-wide association of human population: A tool for broad spectrum service. *Egypt J Med Hum Genet* [Internet]. 2013;14(2):123–34. Available from: <http://dx.doi.org/10.1016/j.ejmhg.2012.08.001>
50. Debette S. How to interpret a genome-wide association study (GWAS)? *Sang Thromb Vaiss.* 2012;24(5):240–7.
51. Dehghan A. Genome-Wide Association Studies. *Genet Epidemiol Methods Protoc Methods Mol Biol* [Internet]. 2018;1793(Gibson 2010):37–49. Available from: <papers2://publication/uuid/C1F14689-770E-4562-B283-11E273724734>
52. Wray NR, Wijmenga C, Sullivan PF, Yang J, Visscher PM. Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. *Cell* [Internet]. 2018;173(7):1573–80. Available from: <https://doi.org/10.1016/j.cell.2018.05.051>

53. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* [Internet]. 2017;169(7):1177–86. Available from: <http://dx.doi.org/10.1016/j.cell.2017.05.038>
54. Chen X, Li Z. Inference of haplotype effects in case-control studies using unphased genotype and environmental data. *Biometrical J.* 2008;50(2):270–82.
55. Evlin BD. Linkage Disequilibrium Measures for Fine-Scale Mapping : A Comparison A Comparison of Linkage Disequilibrium Measures for Fine-Scale Mapping. 2014;322(November 1997):311–22.
56. Schrodin SJ, Garcia VE, Rowland C, Jones HB. Pairwise linkage disequilibrium under disease models. *Eur J Hum Genet.* 2007;15(2):212–20.
57. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* 2019;47(D1):D1005–12.
58. Palmer ND, Langefeld CD, Ziegler JT, Hsu F, Haffner SM, Fingerlin T, et al. Candidate loci for insulin sensitivity and disposition index from a genome-wide association analysis of Hispanic participants in the Insulin Resistance Atherosclerosis (IRAS) Family Study. *Diabetologia.* 2010;53(2):281–9.
59. Donnell CJ, Cupples LA, D'Agostino RB, Fox CS, Hoffmann U, Hwang SJ, et al. Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. *BMC Med Genet.* 2007;8(SUPPL. 1):1–12.
60. Reilly MP, Li M, He J, Ferguson JF, Stylianou IM, Mehta NN, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: Two genome-wide association studies. *Lancet.* 2011;377(9763):383–92.
61. van Zuydam NR, Stiby A, Abdalla M, Austin E, Dahlström EH, McLachlan S, et al. Genome-Wide Association Study of Peripheral Artery Disease. *Circ Genomic Precis Med.* 2021;14(5):e002862.
62. Aherrahrou R, Guo L, Nagraj VP, Aguhob A, Hinkle J, Chen L, et al. Genetic Regulation of Atherosclerosis-Relevant Phenotypes in Human Vascular Smooth Muscle Cells. *Circ Res.* 2020;1552–65.
63. Moreno-Grau S, de Rojas I, Hernández I, Quintela I, Montreal L, Alegret M, et al. Genome-wide association analysis of dementia and its clinical endophenotypes reveal novel loci associated with Alzheimer's disease and three causality networks: The GR@ACE project. *Alzheimer's Dement.* 2019;15(10):1333–47.
64. Schrijvers EMC, Schürmann B, Koudstaal PJ, Van Den Bussche H, Van Duijn CM, Hentschel F, et al. Genome-wide association study of vascular dementia. *Stroke.* 2012;43(2):315–9.
65. Kim Y, Kong M, Lee C. Association of intronic sequence variant in the gene encoding spleen tyrosine kinase with susceptibility to vascular dementia. *World J Biol Psychiatry.* 2013;14(3):220–6.
66. Rosand J, Mitchell BD, Ay H, de Bakker PIW, Gwinn K, Kittner SJ, et al. Loci associated with ischaemic stroke and its subtypes (SiGN): A genome-wide association study. *Lancet Neurol* [Internet]. 2016;15(2):174–84. Available from: [http://dx.doi.org/10.1016/S1474-4422\(15\)00338-5](http://dx.doi.org/10.1016/S1474-4422(15)00338-5)
67. Rosand J. The NINDS Stroke Genetics Network: a genome-wide association study of ischemic stroke and its subtypes. *Lancet Neurol* [Internet]. 2016;15(2):174–84. Available

from:

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4912948/%0Ahttp://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC4912948&blobtype=pdf>

68. Wirka RC, Wagh D, Paik DT, Pjanic M, Nguyen T, Miller CL, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med* [Internet]. 2019;25(8):1280–9. Available from: <http://dx.doi.org/10.1038/s41591-019-0512-5>
69. Mundo-Sagardía JA, Figueroa Y, Altieri PI, Banchs HL, Escobales N, Crespo MJ. Single-cell immune landscape of human atherosclerotic plaques. *Nat Med*. 2019;25(3):1576–88.
70. Gorter RP, Baron W. Matrix metalloproteinases shape the oligodendrocyte (niche) during development and upon demyelination. *Neurosci Lett* [Internet]. 2020;729(April):134980. Available from: <https://doi.org/10.1016/j.neulet.2020.134980>
71. Miyamoto N, Pham LDD, Seo JH, Kim KW, Lo EH, Arai K. Crosstalk between cerebral endothelium and oligodendrocyte. *Cell Mol Life Sci*. 2014;71(6):1055–66.
72. Vasile F, Dossi E, Rouach N. Human astrocytes: structure and functions in the healthy brain. *Brain Struct Funct*. 2017;222(5):2017–29.
73. Hu Z, Hu K, Wang R, Gu Y, Ouyang W, Zhou J, et al. Differentially expressed genes accompanying neurobehavioral deficits in a modified rat model of vascular dementia. *Neurosci Lett*. 2021;750.
74. Hui TYJ, Burt A. Estimating linkage disequilibrium from genotypes under Hardy-Weinberg equilibrium. *BMC Genet*. 2020;21(1):1–11.
75. Slatkin M. Linkage disequilibrium - Understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*. 2008;9(6):477–85.
76. Slatkin M. Linkage disequilibrium understanding the evolutionary past and mapping the medical future. *Nat Rev Genet* [Internet]. 2016;9(6):477–85. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5124487/pdf/nihms-831771.pdf>
77. Tanaka T. International HapMap project. *Nippon rinsho Japanese J Clin Med*. 2005;63 Suppl 1:29–34.
78. Li, Na (Department of Biostatistics, University of Washington, Seattle W 98195), Stephens, Matthew (Department of Statistics, University of Washington, Seattle W 98195). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics*. 2003;165(December):2213–33.
79. Sved JA, Hill WG. One hundred years of linkage disequilibrium. *Genetics*. 2018;209(3):629–36.
80. Terwilliger JD, Hiekkalinna T. An utter refutation of the “Fundamental Theorem of the HapMap.” *Eur J Hum Genet*. 2006;14(4):426–37.
81. Vergara-Lope A, Jabalameli MR, Horscroft C, Ennis S, Collins A, Pengelly RJ. Linkage disequilibrium maps for European and African populations constructed from whole genome sequence data. *Sci Data* [Internet]. 2019;6(1):1–4. Available from: <http://dx.doi.org/10.1038/s41597-019-0227-y>
82. Qiao W, Akhter N, Fang X, Maximova T, Plaku E, Shehu A. From mutations to mechanisms and dysfunction via computation and mining of protein energy landscapes. *BMC Genomics* [Internet]. 2018;19(Suppl 7):1–13. Available from: <http://dx.doi.org/10.1186/s12864-018-5024-z>

83. Uversky VN. Protein intrinsic disorder and structure-function continuum [Internet]. 1st ed. Vol. 166, Progress in Molecular Biology and Translational Science. Elsevier Inc.; 2019. 1–17 p. Available from: <http://dx.doi.org/10.1016/bs.pmbts.2019.05.003>
84. Dey S, Levy ED. Inferring and using protein quaternary structure information from crystallographic data. *Methods Mol Biol.* 2018;1764:357–75.
85. Jez JM. Revisiting protein structure, function, and evolution in the genomic era. *J Invertebr Pathol* [Internet]. 2017;142:11–5. Available from: <http://dx.doi.org/10.1016/j.jip.2016.07.013>
86. Berezovsky IN, Guarnera E, Zheng Z, Eisenhaber B, Eisenhaber F. Protein function machinery: from basic structural units to modulation of activity. *Curr Opin Struct Biol* [Internet]. 2017;42:67–74. Available from: <http://dx.doi.org/10.1016/j.sbi.2016.10.021>
87. Alexey Strokach, Carles Corbi-Verge, Joan Teyra and PMK, Abstract. Predicting the Effect of Mutations on Protein Folding and Protein-Protein Interactions. *Comput Methods Protein Evol.* 2019;1851(January):2–6.
88. Laskowski RA, Stephenson JD, Sillitoe I, Orengo CA, Thornton JM. VarSite: Disease variants and protein structure. *Protein Sci.* 2020;29(1):111–9.
89. Loredó M, Rivera M. Disrupción de marcadores biológicos relacionados con aterosclerosis en células endoteliales humanas. Cuernavaca, Morelos; 2020.
90. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: Leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res.* 2019;47(D1):D1038–43.
91. Bateman A. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 2019;47(D1):D506–15.
92. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet.* 2013;29(10):569–74.
93. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 2018;46(D1):D1062–7.
94. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. *Nucleic Acids Res.* 2020;48(D1):D682–8.
95. Clarke L, Fairley S, Zheng-Bradley X, Streeter I, Perry E, Lowy E, et al. The international Genome sample resource (IGSR): A worldwide collection of genome variation incorporating the 1000 Genomes Project data. *Nucleic Acids Res.* 2017;45(D1):D854–9.
96. Mallick S, Li H, Lipson M, Mathieson I, Gymrek M, Racimo F, et al. The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. *Nature.* 2016;538(7624):201–6.
97. Jassal B, Matthews L, Viteri G, Gong C, Lorente P, Fabregat A, et al. The reactome pathway knowledgebase. *Nucleic Acids Res.* 2020;48(D1):D498–503.
98. Cuevas E, Rivera M. Identificación de Biomarcadores asociados a demencia para proponer un sistema biológico neuronal y circulante para su estudio. Cuernavaca, Morelos; 2020.
99. Cuevas E, Rivera M. Identificación de Biomarcadores asociados a demencia para proponer un sistema biológico neuronal y circulante para su estudio. 2020;
100. Sheridan RP, Wang WM, Liaw A, Ma J, Gifford EM. Extreme Gradient Boosting as a Method for Quantitative Structure-Activity Relationships. *J Chem Inf Model.* 2016;56(12):2353–60.
101. Layeghifard M, Hwang DM, Guttman DS. Constructing and Analyzing Microbiome Networks in R. *Methods Mol Biol.* 2018;1849:243–66.

102. Yu H, Kim PM, Sprecher E, Trifonov V, Gerstein M. The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLoS Comput Biol*. 2007;3(4):713–20.
103. Rai A, Pradhan P, Nagraj J, Lohitesh K, Chowdhury R, Jalan S. Understanding cancer complexome using networks, spectral graph theory and multilayer framework. *Sci Rep* [Internet]. 2017;7(December 2016):1–14. Available from: <http://dx.doi.org/10.1038/srep41676>
104. Hawe JS, Theis FJ, Heinig M. Inferring interaction networks from multi-omics data. *Front Genet*. 2019;10(JUN):1–13.
105. de Bernardi Schneider A, Ford CT, Hostager R, Williams J, Cioce M, Çatalyürek Ü V, et al. StrainHub: a phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*. 2019;(August):1–3.
106. Dopazo J, Erten C. Graph-theoretical comparison of normal and tumor networks in identifying BRCA genes. *BMC Syst Biol*. 2017;11(1):1–17.
107. Yu D, Lim J, Wang X, Liang F, Xiao G. Enhanced construction of gene regulatory networks using hub gene information. *BMC Bioinformatics*. 2017;18(1):1–20.
108. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal Complex Syst* [Internet]. 2006;Complex Sy(1695):1695. Available from: <http://igraph.sf.net>
109. Bastian M, Heymann S, Jacomy M. Gephi: An open source software for exploring and manipulating networks. *BT - International AAAI Conference on Weblogs and Social. Int AAAI Conf Weblogs Soc Media*. 2009;361–2.
110. Das S, Abecasis GR, Browning BL. Genotype imputation from large reference panels. *Annu Rev Genomics Hum Genet*. 2018;19(May):73–96.
111. Lert-itthiporn W, Suktitipat B, Grove H, Sakuntabhai A, Malasit P, Tangthawornchaikul N, et al. Validation of genotype imputation in Southeast Asian populations and the effect of single nucleotide polymorphism annotation on imputation outcome. *BMC Med Genet*. 2018;19(1):1–10.
112. Hui R, D'Atanasio E, Cassidy LM, Scheib CL, Kivisild T. Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. *Sci Rep* [Internet]. 2020;10(1):1–8. Available from: <https://doi.org/10.1038/s41598-020-75387-w>
113. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31(21):3555–7.
114. Schneider VA, Graves-Lindsay T, Howe K, Bouk N, Chen HC, Kitts PA, et al. Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res*. 2017;27(5):849–64.
115. Musunuru K, Kathiresan S. HapMap and mapping genes for cardiovascular disease. *Circ Cardiovasc Genet*. 2008;1(1):66–71.
116. Gower J. A GENERAL COEFFICIENT OF SIMILARITY AND SOME OF ITS PROPERTIES. *Biometrics*. 1971;27(4):857–74.
117. Marcello D'Orazio. Distances with Mixed-Type Variables, some Modified Gower's Coefficients. *J Chem Inf Model*. 2013;53(9):1689–99.
118. Farazdaghi MR, Razeghi M, Sobhani S, Raeisi Shahraki H, Motealleh A. A new clustering method for knee movement impairments using partitioning around medoids model. *Iran J*

Med Sci. 2020;45(6):451–62.

119. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL : homology modelling of protein structures and complexes. 2018;46(May):296–303.
120. Kouza M, Faraggi E, Kolinski A, Kloczkowski A. The GOR method of protein secondary structure prediction and its application as a protein aggregation prediction tool. *Methods Mol Biol.* 2017;1484:7–24.
121. Johnson KM. Using Bayes' rule in diagnostic testing: a graphical explanation. *Diagnosis (Berlin, Ger.)* 2017;4(3):159–67.
122. Igo RP, Kinzy TG, Cooke Bailey JN. Genetic Risk Scores. *Curr Protoc Hum Genet.* 2019;104(1):1–12.
123. Thompson DR, Martin CR. Bayes' theorem and its application to cardiovascular nursing. *Eur J Cardiovasc Nurs.* 2017;16(8):659–61.
124. Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol.* 1999;292(2):195–202.
125. Paniri A, Hosseini MM, Akhavan-Niaki H. First comprehensive computational analysis of functional consequences of Tmprss2 SNPs in susceptibility to SARS-CoV-2 among different populations. *J Biomol Struct Dyn [Internet].* 2020;0(0):1–18. Available from: <https://doi.org/10.1080/07391102.2020.1767690>
126. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc.* 2009;4(7):1073–82.
127. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. *Nat Methods [Internet].* 2010;7(4):248–9. Available from: <http://dx.doi.org/10.1038/nmeth0410-248>
128. Eddy SR. Where did the BLOSUM62 alignment score matrix come from? *Nat Biotechnol.* 2004;22(8):1035–6.
129. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the Functional Effect of Amino Acid Substitutions and Indels. *PLoS One.* 2012;7(10).
130. Booth DS, Szmidt-Middleton H, King N, Westbrook MJ, Young SL, Kuo A, et al. RStudio: Integrated Development for R. *Nature.* 2018;
131. Naj AC. Genotype Imputation in Genome-Wide Association Studies. *Curr Protoc Hum Genet.* 2019;102(1):1–15.
132. PubMed. The Single Nucleotide Polymorphism Database (dbSNP) of Nucleotide Sequence Variation. 2011;(Md). Available from: <http://www.ncbi.nlm.nih.gov/books/NBK21088/>
133. Koos JA, Bassett A. Genetics Home Reference: A Review. *Med Ref Serv Q [Internet].* 2018;37(3):292–9. Available from: <https://doi.org/10.1080/02763869.2018.1477716>
134. Markus HS, Mäkelä KM, Bevan S, Raitoharju E, Oksala N, Bis JC, et al. Evidence HDAC9 genetic variant associated with ischemic stroke increases risk via promoting carotid atherosclerosis. *Stroke.* 2013;44(5):1220–5.
135. Mayr C. Evolution and Biological Roles of Alternative 3'UTRs. *Trends Cell Biol [Internet].* 2016;26(3):227–37. Available from: <http://dx.doi.org/10.1016/j.tcb.2015.10.012>
136. Chouraki V, Seshadri S. Genetics of Alzheimer's disease [Internet]. Vol. 87, *Advances in Genetics.* Elsevier; 2014. 245–294 p. Available from: <http://dx.doi.org/10.1016/B978-0-12-800149-3.00005-6>

137. Kido T, Sikora-Wohlfeld W, Kawashima M, Kikuchi S, Kamatani N, Patwardhan A, et al. Are minor alleles more likely to be risk alleles? *BMC Med Genomics*. 2018;11(1):1–11.
138. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet* [Internet]. 2011;88(3):294–305. Available from: <http://dx.doi.org/10.1016/j.ajhg.2011.02.002>
139. Nikoghosyan M, Hakobyan S, Hovhannisyan A, Loeffler-Wirth H, Binder H, Arakelyan A. Population levels assessment of the distribution of disease-associated variants with emphasis on Armenians - A machine learning approach. *Front Genet*. 2019;10(APR):1–16.
140. Tan R, Traylor M, Rutten-Jacobs L, Markus H. New insights into mechanisms of small vessel disease stroke from genetics. *Clin Sci*. 2017;131(7):515–31.
141. Shen XN, Lu Y, Tan CTY, Liu LY, Yu JT, Feng L, et al. Identification of inflammatory and vascular markers associated with mild cognitive impairment. *Aging (Albany NY)*. 2019;11(8):2403–19.
142. Tsukahara T, Matsuda Y, Haniu H. Lysophospholipid-related diseases and PPAR γ signaling pathway. *Int J Mol Sci*. 2017;18(12).
143. Xu Y, Barnes AP, Alkayed NJ. Role of GPR39 in Neurovascular Homeostasis and Disease. 2021;1–24.
144. Asare Y, Campbell-James TA, Bokov Y, Yu LL, Prestel M, El Bounkari O, et al. Histone deacetylase 9 activates IKK to regulate atherosclerotic plaque vulnerability. *Circ Res*. 2020;811–23.
145. Sofroniew M V., Vinters H V. Astrocytes: Biology and pathology. *Acta Neuropathol*. 2010;119(1):7–35.
146. De Strooper B, Karran E. The Cellular Phase of Alzheimer's Disease. *Cell* [Internet]. 2016;164(4):603–15. Available from: <http://dx.doi.org/10.1016/j.cell.2015.12.056>
147. Hase Y, Horsburgh K, Ihara M, Kalara RN. White matter degeneration in vascular and other ageing-related dementias. *J Neurochem*. 2018;144(5):617–33.
148. Cai Z, Xiao M. Oligodendrocytes and Alzheimer's disease. *Int J Neurosci*. 2016;126(2):97–104.
149. Balliu B, Houwing-Duistermaat JJ, Böhringer S. Powerful testing via hierarchical linkage disequilibrium in haplotype association studies. *Biometrical J*. 2019;61(3):747–68.
150. Epidemiology G. Strain Variation in the Mycobacterium tuberculosis Complex: Its Role in Biology, Epidemiology and Control. 2017;1019:79–93. Available from: <http://link.springer.com/10.1007/978-3-319-64371-7>
151. Thomas DC, Witte JS. Point: Population stratification: A problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev*. 2002;11(6):505–12.
152. Buchan DWA, Minneci F, Nugent TCO, Bryson K, Jones DT. Scalable web services for the PSIPRED Protein Analysis Workbench. *Nucleic Acids Res*. 2013;41(Web Server issue):349–57.
153. Pu A, Stephenson EL, Yong VW. The extracellular matrix: Focus on oligodendrocyte biology and targeting CSPGs for remyelination therapies. *Glia*. 2018;66(9):1809–25.
154. Walker EJ, Rosenberg GA. Divergent role for MMP-2 in myelin breakdown and oligodendrocyte death following transient global ischemia. *J Neurosci Res*. 2010;88(4):764–73.
155. Kim KH, Kang W, Suh SW, Yang JK. Crystal structure of FAF1 UBX domain in complex with

p97/VCP N domain reveals a conformational change in the conserved FcISp touch-turn motif of UBX domain. *Proteins Struct Funct Bioinforma*. 2011;79(8):2583–7.

156. Hänzelmann P, Buchberger A, Schindelin H. Hierarchical binding of cofactors to the AAA ATPase p97. *Structure*. 2011;19(6):833–43.
157. Ramos-Miguel A, García-Sevilla JA, Barr AM, Bayer TA, Falkai P, Leurgans SE, et al. Decreased cortical FADD protein is associated with clinical dementia and cognitive decline in an elderly community sample. *Mol Neurodegener*. 2017;12(1):1–14.
158. Caballero-López MJ, Nieto-Díaz M, Yunta M, Reigada D, Muñoz-Galdeano T, del Águila Á, et al. XIAP Interacts with and Regulates the Activity of FAF1. *Biochim Biophys Acta - Mol Cell Res* [Internet]. 2017;1864(7):1335–48. Available from: <http://dx.doi.org/10.1016/j.bbamcr.2017.04.006>
159. Song JS, Joon KP, Lee JJ, Choi YS, Ryu KS, Kim JH, et al. Structure and interaction of ubiquitin-associated domain of human Fas-associated factor 1. *Protein Sci*. 2009;18(11):2265–76.
160. Kim D, Basile AO, Bang L, Horgusluoglu E, Lee S, Ritchie MD, et al. Knowledge-driven binning approach for rare variant association analysis: application to neuroimaging biomarkers in Alzheimer's disease. *BMC Med Inform Decis Mak*. 2017;17(Suppl 1):1–7.
161. Su JH, Anderson AJ, Cribbs DH, Tu C, Tong L, Kesslack P, et al. Fas and Fas Ligand are associated with neuritic degeneration in the AD brain and participate in β -amyloid-induced neuronal death. *Neurobiol Dis*. 2003;12(3):182–93.
162. Zuk O, Hechter E, Sunyaev SR, Lander ES. The mystery of missing heritability: Genetic interactions create phantom heritability. *Proc Natl Acad Sci U S A*. 2012;109(4):1193–8.
163. Maroilley T, Tarailo-Graovac M. Uncovering Missing Heritability in Rare Diseases. *Genes (Basel)*. 2019;10(4):1–18.
164. Génin E. Missing heritability of complex diseases: case solved? *Hum Genet* [Internet]. 2020;139(1):103–13. Available from: <https://doi.org/10.1007/s00439-019-02034-4>
165. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. *Nature* [Internet]. 2009;461(7265):747–53. Available from: <http://dx.doi.org/10.1038/nature08494>
166. Maher B. Personal genomes: The case of the missing heritability. *Nature*. 2008;456(7218):18–21.
167. Sandoval-Motta S, Aldana M, Martínez-Romero E, Frank A. The human microbiome and the missing heritability problem. *Front Genet*. 2017;8(JUN):1–12.
168. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet*. 2019;20(8):467–84.
169. Li S, Shao Y, Li K, Huangfu C, Wang W, Liu Z, et al. Vascular Cognitive Impairment and the Gut Microbiota. *J Alzheimer's Dis*. 2018;63(4):1209–22.
170. Mirzaei R, Bouzari B, Hosseini-Fard SR, Mazaheri M, Ahmadyousefi Y, Abdi M, et al. Role of microbiota-derived short-chain fatty acids in nervous system disorders. *Biomed Pharmacother* [Internet]. 2021;139(February):111661. Available from: <https://doi.org/10.1016/j.biopha.2021.111661>
171. Saji N, Murotani K, Hisada T, Kunihiro T, Tsuduki T, Sugimoto T, et al. Relationship between dementia and gut microbiome-associated metabolites: a cross-sectional study in Japan. *Sci Rep*. 2020;10(1):1–11.

172. Ling Y, Gong T, Zhang J, Gu Q, Gao X, Weng X, et al. Gut Microbiome Signatures Are Biomarkers for Cognitive Impairment in Patients With Ischemic Stroke. *Front Aging Neurosci.* 2020;12(October):1–13.
173. Adzhubei I a, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. Supprimentals: A method and server for predicting damaging missense mutations Ivan A Adzhubei , Steffen Schmidt , Leonid Peshkin , Vasily E Ramensky , Anna Gerasimova , Peer Bork , Alexey S Kondrashov & Shamil R Sunyaev Supplementary figures and text : *Nat Methods* [Internet]. 2010;7(4):248–9. Available from: <http://dx.doi.org/10.1038/nmeth0410-248>