



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

POSGRADO EN CIENCIAS BIOLÓGICAS

Facultad de Medicina

**“IDENTIFICACIÓN DE NUEVOS
POLIMORFISMOS EN LOS GENES
ANTIOXIDANTES *SMALL MAF*”**

T E S I S

QUE PARA OBTENER EL GRADO ACADÉMICO DE

**MAESTRO EN CIENCIAS BIOLÓGICAS
(BIOLOGÍA EXPERIMENTAL)**

P R E S E N T A

HUMBERTO GUTIÉRREZ MALACATT

TUTOR PRINCIPAL DE TESIS: Dr. Emilio Joaquín Córdova Alarcón

COMITÉ TUTOR:

Dr. Emilio Rojas del Castillo
Dra. Lorena Sofía Orozco Orozco
Dr. Gonzalo Castillo Rojas
Dr. Miguel Ángel Cevallos Gaos

MÉXICO, D.F.

ABRIL, 2011



Universidad Nacional
Autónoma de México



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.

Dr. Isidro Ávila Martínez
 Director General de Administración Escolar, UNAM
 Presente

Me permito informar a usted que en la reunión ordinaria del Comité Académico del Posgrado en Ciencias Biológicas, celebrada el día 28 de junio de 2010, se aprobó el siguiente jurado para el examen de grado de **MAESTRO EN CIENCIAS BIOLÓGICAS (BIOLOGÍA EXPERIMENTAL)** del alumno **GUTIÉRREZ MALACAT HUMBERTO** con número de cuenta **509014955** con la tesis titulada **"IDENTIFICACIÓN DE NUEVOS POLIMORFISMOS EN LOS GENES ANTIOXIDANTES SMALL MAF."**, realizada bajo la dirección del : **DR. EMILIO JOAQUÍN CÓRDOVA ALARCÓN**

Presidente: DR. EMILIO ROJAS DEL CASTILLO
 Vocal: DR. GONZALO CASTILLO ROJAS
 Secretario: DR. EMILIO JOAQUÍN CÓRDOVA ALARCÓN
 Suplente: DR. MIGUEL ANGEL CEVALLOS GAOS
 Suplente: DRA. LORENA SOFÍA OROZCO OROZCO

Sin otro particular, me es grato enviarle un cordial saludo.

Atentamente
"POR MI RAZA HABLARA EL ESPIRITU"
 Cd. Universitaria, D.F., a 02 de febrero de 2011

M. del Coro Arizmendi
Dra. María del Coro Arizmendi Arriaga
 Coordinadora del Programa

Al posgrado en Ciencias Biológicas, UNAM.

A los apoyos recibidos por parte de beca CONACYT, con número de becario 220196.

Al Dr. Emilio J. Córdova Alarcón, director de la tesis, por el tiempo invertido en mi formación académica, por sus valiosos consejos y ayuda para la culminación de este trabajo, además de mostrar siempre su generosa amistad.

A la Dra. Lorena Orozco por su apoyo y consejos para la culminación de este trabajo en su Laboratorio de Inmunogenómica y Enfermedades Metabólicas, en el Instituto Nacional de Medicina Genómica (INMEGEN). Muchas Gracias!!!

A los miembros del jurado: Dr. Emilio Rojas del Castillo, Dr. Miguel Ángel Cevallos Gaos y Dr. Gonzalo Castillo Rojas por los valiosos comentarios, críticas y sugerencias en la revisión de la tesis.

Al Dr. Federico Centeno Cruz por compartir su experiencia, apoyo y amistad durante el desarrollo de este trabajo.

A todos los investigadores del laboratorio por su ayuda, consejos y amistad mostrada durante mi estancia en el laboratorio.

A todos mis compañeros de laboratorio y estudiantes que en gran medida han aportado su trabajo y amistad cotidianos. En especial a la M. en C. Mirna Morales, Biól. Montserrat Rojo de la Vega, Biól. Exp. Jennifer Sánchez, Juan Pablo Pánico y Biól. Angélica Méndez García.

A los integrantes de la Unidad de Secuenciación e Identificación de Polimorfismos del INMEGEN por su valiosa aportación en el desarrollo de este trabajo.

Este trabajo está dedicado a la memoria de mi madre.

A mi padre, por su paciencia y comprensión.

A mis hermanos, por los grandes momentos compartidos.

A mis sobrinos, Miguel Ángel, Verónica, David Ricardo y Jessica Alejandra.

I.	TABLAS.....	I
II.	FIGURAS.....	II
III.	RESUMEN.....	III
IV.	ABSTRACT.....	IV
1	INTRODUCCIÓN.....	1
1.1	Estrés oxidativo y daño a las macromoléculas biológicas.....	1
1.2	Enzimas inducibles de fase II.....	2
1.2.1	Secuencias ARE.....	3
1.3	Genes <i>MAF</i>	5
1.3.1	Estructura génica.....	5
1.3.2	Regiones regulatorias.....	7
1.3.3	Estructura proteica.....	8
1.3.4	Modificaciones post-traduccionales.....	12
1.3.5	Los genes <i>MAF</i> como activadores o represores.....	12
1.4	Regulación de la expresión.....	13
1.4.1	Funciones generales de los genes <i>MAF</i>	14
1.4.2	Diferenciación celular.....	15
1.4.3	Desarrollo embrionario.....	15
1.4.4	Redundancia funcional de las proteínas <i>MAF</i>	17
1.4.5	Asociación con enfermedades.....	17
1.5	Los polimorfismos y su importancia en el desarrollo de enfermedades.....	18
1.6	Haplotipos (<i>Haploid genotype</i>).....	19
2	ANTECEDENTES.....	21
2.1	SNPs en genes inducibles de fase II y asociación con enfermedades.....	21
3	JUSTIFICACIÓN.....	23

4	OBJETIVO GENERAL	24
4.1	Objetivos específicos	24
5	MATERIALES Y MÉTODOS	25
5.1	Población de estudio	25
5.2	Extracción de DNA genómico	25
5.3	La reacción en cadena de la polimerasa (PCR)	26
5.3.1	Diseño de oligonucleótidos para la secuenciación de los genes <i>MAF</i> ...	28
5.4	Electroforesis en geles de agarosa	32
5.5	Purificación de los productos de PCR	32
5.6	Secuenciación automatizada	33
5.7	Desequilibrio de ligamiento y análisis de haplotipos	34
6	RESULTADOS	35
6.1	Amplificación de los genes <i>MAF</i>	35
6.2	Polimorfismos en los genes <i>MAF</i>	41
6.3	Determinación de haplotipos en los genes <i>MAF</i>	50
6.4	Análisis <i>in silico</i>	54
7	DISCUSIÓN	57
8	CONCLUSIONES	65
9	PERSPECTIVAS	66
10	BIBLIOGRAFÍA	67

I. TABLAS.

Tabla 5-1. Oligonucleótidos utilizados en la amplificación del gen <i>MAFF</i>	29
Tabla 5-2. Oligonucleótidos utilizados en la amplificación del gen <i>MAFG</i>	30
Tabla 5-3. Oligonucleótidos utilizados en la amplificación del gen <i>MAFK</i>	30
Tabla 5-4. Condiciones óptimas de PCR para los amplicones <i>MAFF</i>	30
Tabla 5-5. Condiciones óptimas de PCR para los amplicones <i>MAFG</i>	31
Tabla 5-6. Condiciones óptimas de PCR para los amplicones <i>MAFK</i>	31
Tabla 6-1. Variantes encontradas en el gen <i>MAFF</i>	41
Tabla 6-2. Variantes encontradas en el gen <i>MAFG</i>	42
Tabla 6-3. Variantes encontradas en el gen <i>MAFK</i>	43
Tabla 10-6-4. Comparación de las frecuencias alélicas poblacionales para el gen <i>MAFF</i>	44
Tabla 10-6-5. Frecuencias alélicas y genotípicas de la nueva variante de <i>MAFG</i>	45
Tabla 10-6-6. Comparación de las frecuencias alélicas poblacionales para el gen <i>MAFK</i>	46
Tabla 6-7. Reporte de SNPs encontrados en los genes <i>MAF</i>	48
Tabla 6-8. Haplotipos del gen <i>MAFF</i>	53
Tabla 6-9. Haplotipos del gen <i>MAFK</i>	53

II. FIGURAS.

Figura 1-1. Ruta citoprotectora Nr2-Keap1.	4
Figura 1-2. Estructura de los tres genes <i>MAF</i>	7
Figura 1-3. Comparación de la estructura proteica de la familia <i>MAF</i>	9
Figura 1-4. Identidad de las proteínas <i>MAF</i> pequeñas en los mamíferos.....	10
Figura 1-5. Secuencias consenso de DNA reconocidas por los factores bZIP.....	11
Figura 6-1. Estandarización de las condiciones de amplificación de los genes <i>MAF</i>	36
Figura 6-2. Algunas variantes identificadas en los genes <i>MAF</i>	38
Figura 6-3. Doble inserción heteróciga (GCG y CGG) presente en la región reguladora de <i>MAFG</i>	40
Figura 6-4. Matriz de LD para <i>MAFF</i> según el parámetro r^2	51
Figura 6-5. Matriz de LD para <i>MAFK</i> según el parámetro r^2	51
Figura 6-6. Análisis <i>in silico</i> para sitios de unión de factores de transcripción..	56

III. RESUMEN.

Las proteínas MAF pequeñas, codificadas por los genes *MAFF*, *MAFG* y *MAFK* son factores reguladores de la transcripción que modulan la expresión de toda una batería de genes implicados en la respuesta al estrés oxidativo, la diferenciación y el desarrollo tisular. A pesar de su importancia fisiológica, todavía está incompleto el registro de polimorfismos funcionales en estos genes. El objetivo del presente estudio fue la caracterización de los polimorfismos de los genes *MAF* pequeños en la población mexicana. Para ello, se obtuvieron muestras de DNA a partir de sangre periférica de 112 individuos no relacionados entre sí. Posteriormente se secuenciaron las regiones regulatoria y codificante de estos genes, en cada una de las muestras. En los genes *MAFF*, *MAFG* y *MAFK* se encontraron 19, 7 y 16 polimorfismos, respectivamente, algunos de ellos descritos previamente. En ninguno de los genes *MAF* se identificaron variantes en la región codificante de la proteína, la mayor parte se encontró principalmente en el promotor y regiones UTRs. De acuerdo al número de polimorfismos encontrados, los genes *MAF* pequeños pueden ordenarse de la siguiente manera: *MAFF* > *MAFK* > *MAFG*. La menor variabilidad génica en las regiones regulatorias de *MAFG* concuerda con diversos modelos murinos que indican una mayor importancia fisiológica de *MAFG*. La frecuencia y el número de variantes encontradas en la población mexicana difiere de los datos reportados en la base dbSNP del NCBI (*National Center for Biotechnology Information*), obtenida a partir de la secuenciación del genoma humano. Estos resultados señalan la clara heterogeneidad genética presente en nuestra población y la necesidad de contar con estudios propios que permitan determinar, tanto la presencia, como la frecuencia de variantes génicas que puedan ser utilizados como posibles biomarcadores para estudios de susceptibilidad genética.

Palabras clave: SNPs, Nrf2, *MAFF*, *MAFG*, *MAFK*, *small MAFs*.

IV. ABSTRACT.

Small MAF transcription factors are coded by genes *MAFF*, *MAFG* and *MAFK*. These proteins regulate the expression of a whole battery of genes implicated in the response towards oxidative stress, cell differentiation and tissue development. In spite of its physiological importance, still unknown the presence of functional polymorphisms in these genes. The objective of the present study was the characterization of polymorphisms in small *MAF* genes in a sample of Mexican population. DNA collected from peripheral blood of 112 Mexican unrelated individuals was obtained. Regulatory and codifying regions were amplified by PCR and analyzed for the presence of polymorphisms in promoter and codifying regions. In *MAFF*, *MAFG* and *MAFK* genes we identified 19, 7 and 16 variants respectively, mainly in the promoter and UTRs regions, some of this variants were described previously. We do not found variants in the codifying regions. According to the number of founded variants, small *MAF* genes showed a different level of variability being the *MAFF* gene the most polymorphic, followed by *MAFK* and *MAFG* genes. The smaller genetic variability in *MAFG* agrees with diverse murine models that indicate a greater physiological importance of this gene. The frequency and number of variants found in Mexican population was different from data reported in dbSNP of NCBI (National Center for Biotechnology Information) for human genome. This study show clearly the genetic variability present in our population and the necessity to carry on studies that allow us determine the presence, as well as the frequency of genic variants that can be used like potential biomarkers for future studies of genetic susceptibility toward some diseases.

Keywords: polymorphisms, SNPs, *MAFF*, *MAFG*, *MAFK*, NRF2.

1 INTRODUCCIÓN.

1.1 Estrés oxidativo y daño a las macromoléculas biológicas.

Los organismos aerobios utilizan el oxígeno molecular (O_2) en numerosos procesos metabólicos esenciales, esto les confiere una gran ventaja evolutiva. Sin embargo, tienen que enfrentar la generación de especies reactivas del oxígeno (*reactive oxygen species*, ROS,) como subproductos normales del metabolismo celular. Entre las ROS generadas *in vivo* están el anión superóxido ($O_2^{\cdot-}$), los radicales peroxilo (RO_2^{\cdot}) y alcoxilo (RO^{\cdot}), así como el peróxido de hidrógeno (H_2O_2). La mitocondria es la principal fuente intracelular de ROS, sobre todo en los complejos I y III de la cadena respiratoria⁽¹⁾. También se producen ROS como intermediarios en otros procesos metabólicos celulares como la β -oxidación de los ácidos grasos, la actividad de los citocromos P450 y durante la fagocitosis de organismos patógenos por los neutrófilos y macrófagos⁽²⁾. Además de las fuentes endógenas, la generación de ROS puede deberse a la presencia de xenobióticos como fármacos y pesticidas así como por contaminantes ambientales.

A bajas concentraciones las ROS funcionan como mensajeros intracelulares, regulando múltiples rutas de transducción de señales en procesos como el control del crecimiento y la diferenciación celular, así como en la producción de citocinas. Sin embargo, cuando surge un desequilibrio entre los niveles intracelulares de las ROS y los sistemas antioxidantes celulares aparece el **estrés oxidativo**. De esta manera las ROS dañan a los ácidos nucleicos, las proteínas y los lípidos ocasionando mutaciones, pérdida de actividad enzimática, y alteraciones en la permeabilidad celular. El estrés oxidativo ha sido implicado en la etiología de varios procesos patológicos como el cáncer, las enfermedades cardiovasculares y autoinmunes, así como en padecimientos neurodegenerativos⁽³⁾.

1.2 Enzimas inducibles de fase II.

Dada la ubicuidad y la toxicidad del estrés oxidativo, las células cuentan con mecanismos que regulan de manera rigurosa la producción de las ROS. Este control se consigue principalmente mediante la actividad de una batería de moléculas antioxidantes y desintoxicantes que eliminan eficientemente las ROS. En el caso de las enzimas desintoxicantes, éstas pueden dividirse en dos tipos: a) de fase I y b) de fase II⁽⁴⁾. En el primer caso, los citocromos P450, llevan a cabo reacciones generadoras de grupos funcionales (-OH, -COOH, -NH₂, o -SH) que confieren mayor hidrofiliidad a las moléculas. Una consecuencia de estas reacciones es la producción de sustancias electrofílicas altamente nocivas para las moléculas biológicas. Por su parte, las enzimas de fase II intervienen de dos maneras: i) transforman a los xenobióticos en formas menos tóxicas, o b) llevan a cabo *reacciones de conjugación* (glucuronidación, sulfación, metilación, acetilación, etc.) para transformar a los metabolitos en sustancias hidrofílicas de fácil excreción. Finalmente, en la fase III intervienen varios tipos de bombas transportadoras que controlan la captura y eflujo de las sustancias a eliminar.

Las enzimas de fase II incluyen proteínas como NAD(P)H:quinona oxidorreductasa 1 (NQO1), hemo-oxigenasa 1 (HO-1), glutatión S-transferasas (GSTs) y UDP-glucuronosiltransferasas (UGTs), entre muchas otras. La regulación de las enzimas de fase II está coordinada por varios receptores nucleares responsivos a la interacción con sustancias endógenas y xenobióticos. En la región promotora de los genes de fase II se ha localizado secuencias responsivas a diversos factores de transcripción, una de las más importantes es el elemento de respuesta antioxidante (*antioxidant response element*, ARE)⁽⁵⁾.

1.2.1 Secuencias ARE.

Los elementos de respuesta antioxidante (AREs), ubicados en el promotor de numerosos genes de fase II presentan la secuencia consenso 5'-GTGAG/CnnnGC-3', en donde "n" representa a cualquier nucleótido y las bases GC en el extremo 3' son esenciales para la expresión de genes antioxidantes⁽⁶⁾. En los últimos años se ha demostrado que el heterodímero NRF2:MAF pequeñas^(7,8) es el principal regulador de genes que poseen sitios ARE⁽⁹⁾. Los ratones *Nrf2-null* se caracterizan por presentar niveles muy bajos, basales e inducibles, de diversas enzimas antioxidantes y desintoxicantes⁽¹⁰⁻¹²⁾. Por otro lado, cuando Nrf2 se induce genética o farmacológicamente aumentan los niveles de diversas enzimas de fase II⁽¹³⁻¹⁵⁾.

En condiciones basales el factor Nrf2 es retenido en el citoplasma por la proteína represora Keap1, de manera que continuamente es degradado por la ruta del proteosoma (**Figura 1-1**). Sin embargo, en condiciones de estrés oxidativo o en presencia de sustancias electrofílicas, se modifican ciertos residuos de cisteína sobre Keap1 (Cys 151, 273, 288), esto promueve un cambio conformacional en la proteína represora. Al mismo tiempo, algunas cinasas (como MAPK, PKC, PI3K) inducidas por los agentes estresores fosforilan a ciertos residuos de tirosina sobre Nrf2 (Ser40). Estos cambios permiten que Nrf2 se libere de Keap1 y se transloque al núcleo⁽¹⁶⁾. Nrf2 necesariamente debe formar heterodímeros para poder unirse a secuencias específicas en el DNA (AREs). Los complejos Nrf2:MAF pequeñas reclutan a la maquinaria de transcripción basal para la expresión de numerosos genes citoprotectores, entre ellos los de las enzimas de fase II.

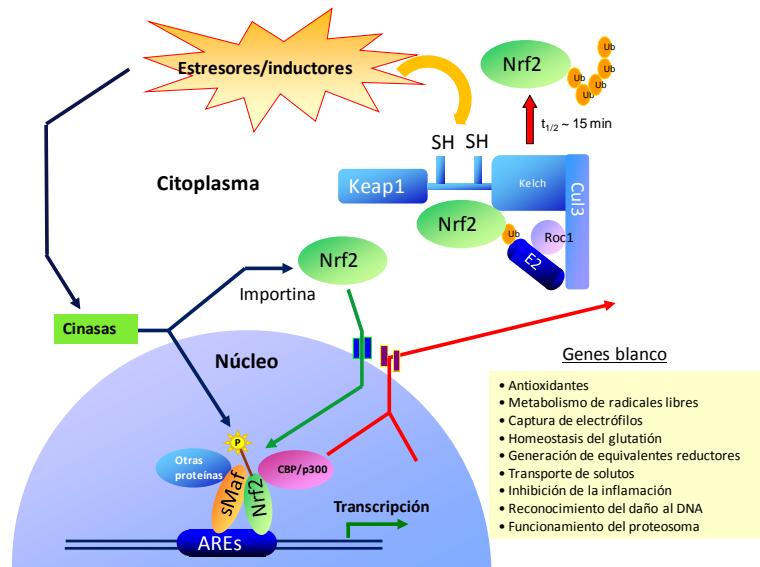


Figura 1-1. Ruta citoprotectora Nrf2-Keap1.

La cinasa GSK3 β controla el apagado de la expresión génica controlada por Nrf2. GSK3 β fosforila a Fyn en un residuo treonina desconocido y esto permite su translocación nuclear⁽¹⁷⁾. En el núcleo, Fyn fosforila a Nrf2 (Tyr568) y este cambio promueve la exportación de Nrf2 del núcleo, uniéndose nuevamente a Keap1 para ser degradado por el proteosoma⁽¹⁸⁾. Bach1 es una proteína nuclear que compete con Nrf2 para la unión a sitios ARE, de esta manera suprime la expresión de los genes controlados por Nrf2⁽¹⁹⁾. La regulación negativa de Nrf2 a través de Fyn y Bach1 es importante en la inactivación de los genes controlados por Nrf2. Al parecer, Nrf2 también puede formar un complejo con c-Myc para reprimir la expresión de genes de fase II⁽²⁰⁾. Por otro lado, Nrf2 puede unirse con c-Jun para la activación de genes con secuencias ARE⁽²¹⁾, mientras que con los integrantes de la familia Fos forma complejos represores de la transcripción⁽²²⁾.

La actividad transcripcional de Nrf2 depende de la formación de complejos con los integrantes de varias familias proteicas. Una de las más importantes es la familia de proteínas MAF pequeñas.

1.3 Genes *MAF*.

Los genes *MAF* pequeños (denominados en adelante como genes *MAF*) pertenecen a la familia *MAF* (*musculoaponeurotic fibrosarcoma oncogene homolog*) de factores de transcripción bZIP, junto con las proteínas *MAF* grandes (c-*MAF*, *MAFA*, *MAFB* y *NRL*)⁽²³⁾. Hasta el momento se han descrito tres genes *MAF* pequeños: *MAFF*, *MAFG* y *MAFK*. Los factores *MAF* son las únicas proteínas bZIP capaces de reconocer una secuencia de DNA excepcionalmente larga. En la **Figura 1-5** se muestran las secuencias de DNA contactadas por sus dominios proteicos.

1.3.1 Estructura génica.

Los genes *MAF* se localizaron en distintos cromosomas humanos; el gen *MAFF* se ubica en el brazo largo del cromosoma 22, banda citogenética 22q13.1, *MAFG* reside en el brazo largo del cromosoma 17, banda citogenética 17q25.3 y *MAFK* se localiza en el brazo corto del cromosoma 7, banda citogenética 7p22.3. La estructura general de estos genes consiste en tres exones separados por dos intrones (**Figura 1-2**); el primer exón es regulador y no se traduce. La región 3'-UTR de los genes *MAF* es extensa y posee numerosos residuos AT, además contiene varias señales de desestabilización (ATTTA)⁽²⁴⁻²⁶⁾, lo cual sugiere un rápido recambio proteico.

El gen *MAFF*, con un tamaño de 14, 630 pares de bases (NT_011520.12), está ubicado en la cadena positiva y posee dos intrones: el primer intrón tiene un tamaño aproximado de 11.6 Kb mientras que el segundo intrón es de 530 pb y se localiza dentro de la región codificante. Este gen se expresa abundantemente en la placenta y en menor proporción en los pulmones, el ojo, los testículos, la piel y

el tejido adiposo (www.genecards.org). *MAFF* da origen a 5 transcritos diferentes, todos ellos son codificantes y se producen en distintos tejidos.

El gen *MAFG* posee un tamaño de 9, 443 pb y está localizado en orientación reversa (hebra negativa). Este gen codifica a dos transcritos que producen la misma proteína. El segundo exón de *MAFG* contiene una secuencia 5'-UTR adicional y codifica a los primeros 12 aminoácidos de la proteína. El resto de la región codificante y al menos 1 Kb de la secuencia 3'-UTR están incluidos en el tercer exón. Ambos intrones son pequeños (4.3 Kb y 92 pb). El gen *MAFG* tiene una elevada expresión en el músculo esquelético y médula espinal, pero también se transcribe en el corazón y el cerebro.

El gen *MAFK* tiene un tamaño de 12, 330 pb y está orientado en la cadena positiva. El gen humano codifica para 3 transcritos que producen la misma proteína. Esta estructura está altamente conservada con respecto a los genes *MAF* del pollo, ratón y otros mamíferos⁽²⁶⁾.

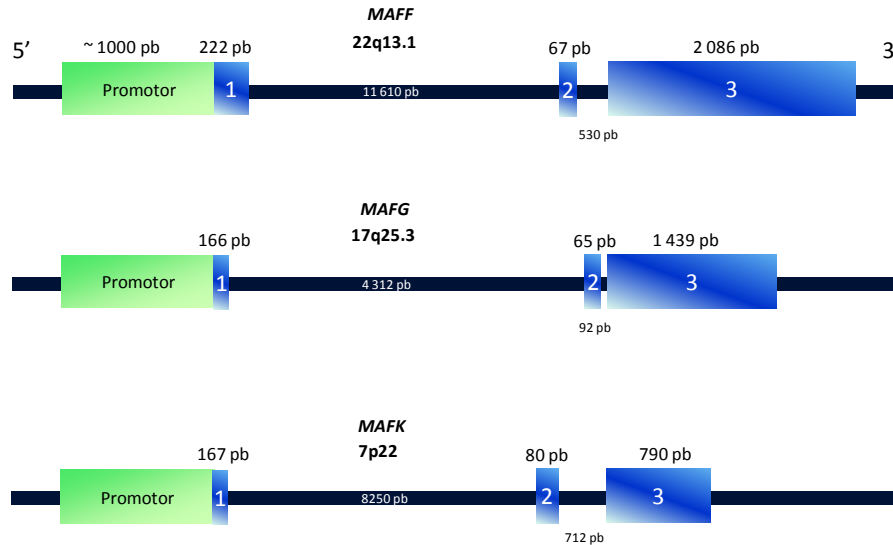


Figura 1-2. Estructura de los tres genes *MAF*. Cada uno de ellos se localiza en diferentes cromosomas, sin embargo comparten una estructura génica común, consistente en una región reguladora, tres exones y dos intrones. El exón 1 es regulador y la proteína es codificada por los exones 2 y parte del 3.

1.3.2 Regiones regulatorias.

Aunque se desconoce casi por completo el funcionamiento y las características de la región promotora de los genes *MAF* humanos, esta zona se distingue por un alto contenido GC en su región reguladora mínima¹, en donde residen islas CpG^{2(24,26)}. En los ratones, *MAFG* posee tres primeros exones alternativos responsivos a estímulos específicos, mientras que *MAFK* cuenta con dos promotores diferentes, cuya activación ocurre dependiendo si están en el tejido mesodérmico o en el cerebro⁽²⁷⁾. Un potenciador (*enhancer*) localizado 3'

¹ En la base de datos INGENUITY, la región promotora potencial de los genes *MAF* se considera 1 Kb río arriba del sitio de inicio de la transcripción (TSS) y 200 pb río abajo del TSS.

² Es una región genómica de 0.5 – 2.0 Kb y que posee un contenido > 50% de G+C. La p indica que las bases C y G están conectadas con un enlace fosfodiéster. Las islas CpG generalmente se localizan río arriba de genes *housekeeping* y también en algunos genes tejido-específicos. En las células de los mamíferos las islas CpG están hipometiladas con respecto al resto del genoma.

respecto al gen *MAFK* es esencial para la transcripción en los tejidos hematopoyéticos y células del músculo cardíaco⁽²⁸⁾. Finalmente, el gen *MAFF* murino opera con tres promotores diferentes, pero su función no se ha estudiado detalladamente⁽²⁹⁾.

1.3.3 Estructura proteica.

Entre las tres proteínas MAF existe un grado muy significativo de identidad en su secuencia de aminoácidos. El dominio básico de la familia MAF posee una equivalencia significativa (hasta el 40%) con otros factores bZIP, particularmente con las familias AP-1 y CREB/ATF^(23,30). El porcentaje de identidad de *MAFK*, de acuerdo a la secuencia de aminoácidos, es del 75% con *MAFG* y del 68% con *MAFF*. Las proteínas MAF forman heterodímeros con las familias CNC y Bach, así como con Fos (c-Fos, FosB, Fra1 y Fra2), c-MAF y p45³, pero no con Jun, v-MAF y *MAFB*⁽³¹⁾. En la **Figura 1-3** se presenta la estructura general de las proteínas MAF. La característica distintiva de la familia MAF es un segmento conservado en la porción N-terminal, conocido como la región de homología extendida (EHR, del inglés *extended homology region*)⁽³²⁾. Esta zona forma una hélice α anfipática, conformada por residuos hidrofóbicos y aminoácidos hidrofílicos, y es esencial para la unión específica con el DNA. La presencia de un dominio cierre de leucinas posibilita la dimerización entre sí mismas y con otros factores bZIP⁽³³⁾. Las tres proteínas MAF difieren en la región carboxilo terminal, por lo que se especula que esta zona podría ejercer una función particular a cada una de ellas⁽³²⁾, pero hasta el momento no se tienen evidencias de esto. Recientemente se demostró en la porción C-terminal de *MAFG* la existencia de una señal de localización en la matriz nuclear, esta señal sería indispensable para la ubicación de *MAFG* en un entorno cromatínico favorable para la transcripción⁽³⁴⁾.

³ p45 NF-E2 se expresa exclusivamente en las células hematopoyéticas. La pérdida de función de p45 ocasiona la muerte del embrión por defectos trombopoyéticos severos, y su sobreexpresión impulsa la liberación de plaquetas.

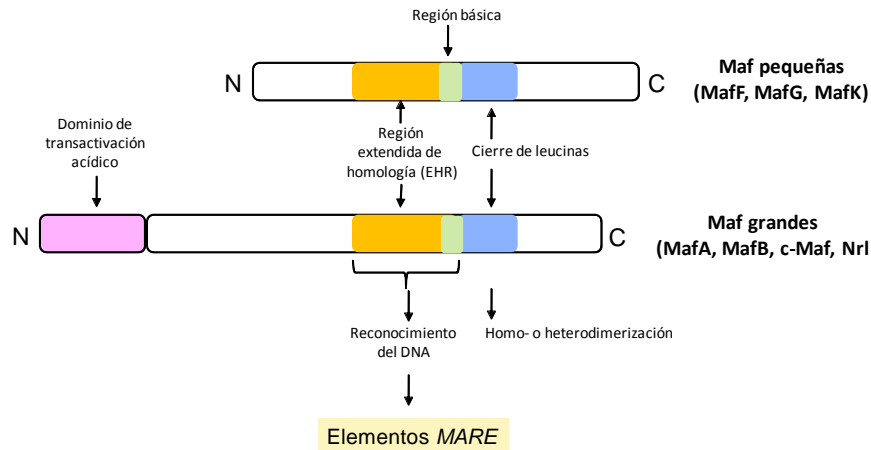


Figura 1-3. Comparación de la estructura proteica de la familia MAF. Los factores MAF están constituidos por una estructura modular en dominios funcionales. El dominio básico participa en la unión al DNA mientras que la dimerización sucede a través del cierre de leucinas. La región extendida de homología (EHR) es específica de las proteínas MAF y confiere especificidad de unión al DNA.

La comparación de las secuencias de todas las proteínas MAF en los mamíferos muestra que los dominios básico y cierre de leucina están altamente conservados, pero cada una contiene particularidades importantes (**Figura 1-4**)^(31,32). La proteína MAFF se compone de 164 residuos de aminoácidos (17.76 KDa), MAFG de 162 (17.85 KDa) y MAFK de 156 (17.52 KDa). El dominio cierre de leucinas en las tres proteínas MAF comprende los residuos 86-114; en tanto que el motivo básico de unión al DNA abarca los residuos 53 – 83 en las proteínas MAFF y MAFG, y los residuos 51 -76 en MAFK.

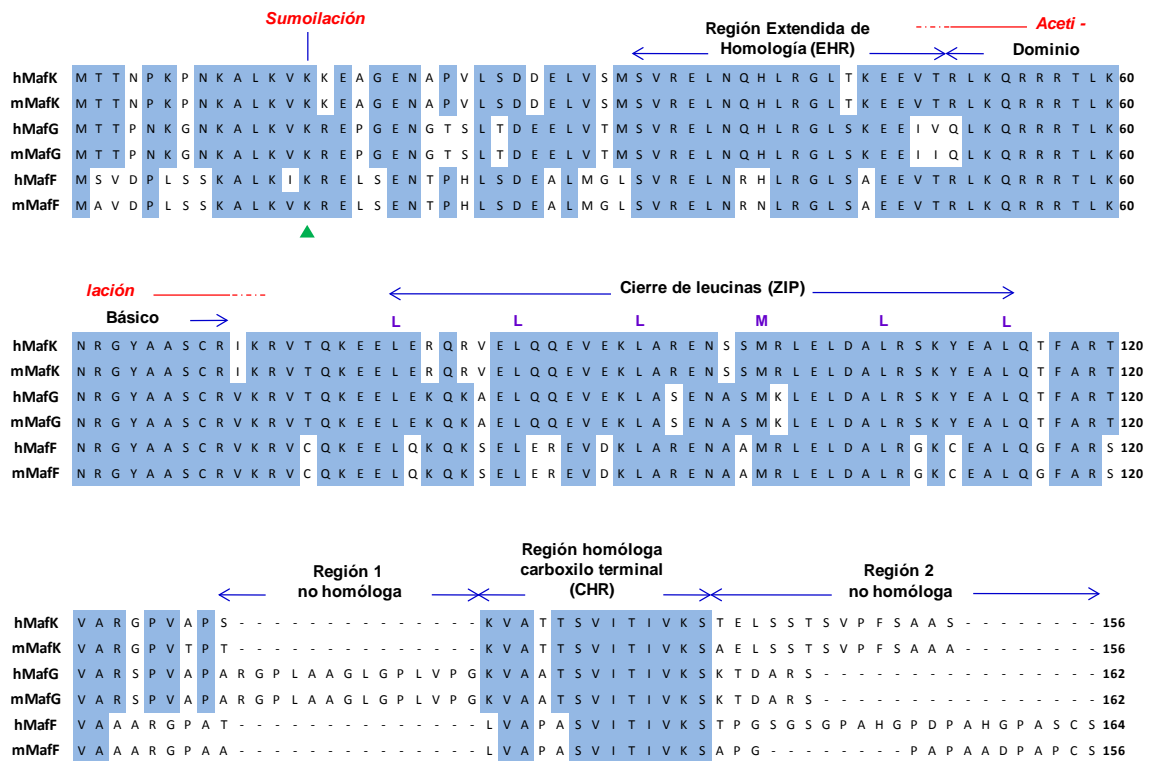


Figura 1-4. Identidad de las proteínas MAF pequeñas en los mamíferos. Se muestra la secuencia proteica de MAFF, MAFG y MAFK en el humano y ratón. En azul se indican los residuos conservados en por lo menos cuatro de las seis proteínas. También se resalta el dominio básico responsable de la unión al DNA y la región cierre de leucina utilizada en la dimerización. Existe una región conocida como la región homóloga C-terminal (VAXXSVITIVKS) cuya función se desconoce. Este motivo está flanqueado por dos zonas no homólogas características de cada proteína MAF. En rojo se marcan los residuos de aminoácidos modificados por sumoilación (consenso B-K-x-D/E, en donde B es un residuo hidrofóbico) y acetilación, caracterizado únicamente en MAFG. El triángulo verde marca la ubicación de un residuo insertado en la posición 27, presente en una variante de empalme de MAFG de la rata. Esquema modificado de la referencia ⁽³¹⁾.

Se tiene evidencia experimental de la interacción específica de las proteínas MAF con otros factores de transcripción, por ejemplo, MAFF interactúa con MIP⁽³⁵⁾ y NRF1⁽³⁶⁾; MAFG con PAX6⁽³⁷⁾ y MAFK con NRF3⁽³⁸⁾.

Las proteínas MAF pueden unirse a secuencias de DNA tipo MARE (del inglés *MAF-recognition element*) o ARE (Figura 1-5). La secuencia MARE se compone de un núcleo TRE de 7 pb ($\text{TGCTGA}^{\text{G/C}}\text{TCAGCA}$) o CRE de 8 pb ($\text{TGCTGA}^{\text{G/C/G}}\text{TCAGCA}$), que es contactado por el dominio básico de la proteína, y una secuencia TGC flanqueadora reconocida por el dominio EHR. Las secuencias core de MARE están flanqueadas por elementos GC. Esta secuencia de reconocimiento es peculiar de las proteínas MAF. Los genes blanco activados por los homo- y heterodímeros de la familia MAF difieren significativamente de los activados por las proteínas AP1 o CREB⁽³⁹⁾.

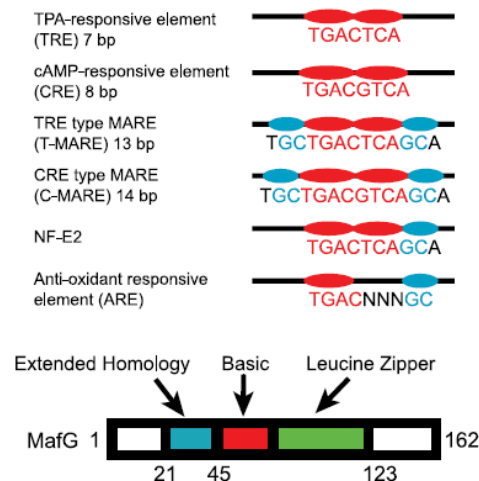


Figura 1-5. Secuencias consenso de DNA reconocidas por los factores bZIP. El núcleo TRE o CRE (en rojo) está implícito en las secuencias T-MARE y C-MARE, respectivamente. Las bases adyacentes a las secuencias centrales (en azul) son esenciales para el reconocimiento de las secuencias MARE por parte de las proteínas MAF, los sitios de unión para NRF2 y ARE son secuencias asimétricas compuestas por la mitad de un sitio TRE y una mitad de un sitio MARE. En la parte inferior se ilustran los dominios proteicos de MAFG que contactan al DNA. Esquema obtenido de la referencia⁽⁴⁰⁾.

1.3.4 Modificaciones post-traduccionales.

Aunque la actividad de las proteínas MAF está controlada principalmente a nivel transcripcional, algunas modificaciones post-traduccionales también rigen su funcionamiento. Por ejemplo, la actividad de MAFG puede ser modificada por acetilación ejercida por la proteína CBP (del inglés **CREB binding protein**)⁴⁽⁴¹⁾. Además, MAFG cuenta con un sitio consenso de sumoilación intervenido por SUMO2/3⁵, esto le permite reclutar a un complejo con actividad de histona desacetilasa represor de la transcripción⁽⁴²⁾. De esta forma, la acetilación y la sumoilación modulan de manera importante la actividad de las proteínas MAF.

1.3.5 Los genes *MAF* como activadores o represores.

Las proteínas MAF pueden activar o reprimir la transcripción dependiendo de su concentración proteica intracelular y el tipo de heterodímeros formados. Experimentos con ratones *knockout* señalan que las proteínas MAF promueven la transcripción de genes antioxidantes y desintoxicantes cuando forman heterodímeros con algunos factores de la familia CNC⁶ como Nrf1, Nrf2, Nrf3 y p45. Por otro lado, la interacción de las proteínas MAF con los factores Bach1⁷ y

⁴ CBP y p300 son coactivadores transcripcionales que interactúan con numerosos factores de transcripción para reclutar a la RNA polimerasa y otros reguladores transcripcionales. Además, CBP y p300 poseen actividad de histona acetiltransferasa, de tal manera que influyen en la actividad de la cromatina controlando a las histonas del nucleosoma.

⁵ SUMO (**s**mall **u**biquitin-**l**ike **m**odifier). Es una molécula similar a la ubiquitina; la modificación post-traduccionales de proteínas con SUMO regula una amplia variedad de procesos celulares, como translocación nuclear, estabilidad, reparación del DNA, modulación de factores de transcripción y formación de estructuras subnucleares.

⁶ Nombrados de esta manera debido a su semejanza estructural con la proteína cap 'n' collar (CNC) de *Drosophila*.

⁷ Bach1 compete con Nrf2 para unirse en los sitios ARE. Se ha demostrado que un residuo tirosina (Tyr486) de Bach1 tiene que ser fosforilado para permitir su salida del núcleo, de esta manera Nrf2 tiene acceso a ARE.

Bach2 reprime la expresión de estos genes. Finalmente, ensayos con genes reporteros han demostrado la función represora de los homodímeros MAF⁽⁴³⁾.

1.4 Regulación de la expresión.

La expresión de los genes *MAF* puede ser inducida de manera conjunta por diversas sustancias electrofílicas, ROS⁽⁴⁴⁾, así como por metales pesados como el cadmio y el zinc⁽⁴⁵⁾. Por otro lado, se ha descrito que cada uno de los genes *MAF* puede ser regulado de manera diferencial. En el caso de *MAFG*, su expresión se induce por agentes como el peróxido de hidrógeno y xenobióticos como la β -naftoflavona⁽⁴⁶⁾; la expresión de *MAFG* también puede inducirse por diversas condiciones estresantes como acidificación extracelular, estimulación hipercápica (7% CO₂) del bulbo raquídeo⁽⁴⁷⁾ y estímulo de barorreceptores con fenilefrina⁽⁴⁸⁾. Por el contrario, el interferón alfa (IFN- α 2b) es un regulador negativo de *MAFG*⁽⁴⁹⁾. Los transcritos de *MAFF* son inducidos por citocinas proinflamatorias como la interleucina 1 β (IL1 β)⁽⁵⁰⁾ y el factor de necrosis tumoral (TNF)⁽⁵⁰⁾. *MAFF* se expresa abundantemente en el miometrio alrededor de la culminación del parto. La inducción por arsénico en células del coriocarcinoma de la placenta, *in vitro*, aumentó ligeramente los niveles de expresión de *MAFF*, pero no los de *MAFG* y *MAFK*, los cuales intervienen en la diferenciación terminal de los megacariocitos, precursores de las plaquetas⁽⁵¹⁾. Por otra parte se ha demostrado que el factor Tal1/Scl controla la función de *MAFK* durante el desarrollo y diferenciación de células troncales hematopoyéticas⁽⁵²⁾.

A través del uso de microarreglos de expresión se han revelado alteraciones en la actividad de los genes *MAF* debido al tratamiento de diversas sustancias. Por ejemplo, el uso de angiotensina II en células adrenocorticales humanas indujo la expresión de *MAFF*⁽⁵³⁾. Por otro lado, la estimulación de células T47D (derivadas del cáncer de mama) con prolactina y estradiol promueve la expresión

de *MAFF*⁽⁵⁴⁾. En la línea celular PC12 de feocromocitoma⁸ el tratamiento con el factor de crecimiento nervioso, NGF (del inglés *nerve growth factor*) también induce la expresión de *MAFF*⁽⁵⁵⁾. La onconasa, una endorribonucleasa (RNasa) utilizada para tratar el mesotelioma maligno, promueve la expresión de *MAFF* además de otros genes implicados en la apoptosis, transcripción, inflamación y respuesta inmune⁽⁵⁶⁾. En el caso de *MAFG* diversos estudios han revelado cambios en su expresión debido a la exposición al humo del tabaco⁽⁵⁷⁾, interferón $\alpha 2b$ ⁽⁴⁹⁾, metales pesados⁽⁴⁵⁾, acidificación extracelular⁽⁵⁸⁾ e hipoxia⁽⁵⁹⁾. Finalmente, *MAFK* fue inducido por polifenoles presentes en extracto de té negro⁽⁶⁰⁾, exposición prenatal de ratas al perfluorooctano sulfonato (contaminante ambiental)⁽⁶¹⁾ y sulforafano⁽⁶²⁾.

1.4.1 Funciones generales de los genes *MAF*.

Las proteínas MAF regulan la transcripción de numerosos genes, especialmente los comprometidos en el sistema hematopoyético, pero también la de genes desintoxicantes de fase II, e interventores en el desarrollo y diferenciación. Los genes *MAF* se expresan ampliamente, sin embargo el nivel de sus transcritos varía considerablemente en diferentes tejidos. La disrupción de los genes *MAF* en los ratones ha puesto en relevancia las funciones de estas proteínas en la promoción de la producción de plaquetas, mantenimiento de la función neuronal, citoprotección contra el estrés xenobiótico y oxidativo, así como la regulación negativa del gen de la hemo-oxigenasa 1 (HO-1).

⁸ Los feocromocitomas son tumores adrenales productores de catecolaminas. Surgen debido a síndromes cancerosos hereditarios, o más frecuentemente, aparecen esporádicamente debido a una causa genética desconocida.

1.4.2 Diferenciación celular.

A nivel de la diferenciación celular se ha documentado que las tres proteínas MAF son fundamentales para el desarrollo de diferentes tipos celulares como los pertenecientes al tejido hematopoyético, el nervioso y el conectivo, entre otros. De esta forma, MAFG es una proteína importante en las últimas etapas de la megacariopoyesis⁽⁶³⁾; mientras que MAFK impide la conversión de células B en células plasmáticas e interviene en la diferenciación de células T⁽⁶⁴⁾. En el sistema nervioso, MAFK se ha vinculado al desarrollo de neuritas en células PC12⁹⁽⁶⁵⁾. Además, *MAFK* y *MAFG* se expresan en médula espinal, tronco cerebral, hipocampo y ganglios dorsales. *MAFF* y *MAFG* también contribuyen en la diferenciación de queratinocitos⁽⁶⁶⁾. Por su parte, *MAFF* colabora en la diferenciación de células precursoras de condrocitos de ratón bajo condiciones de hipoxia⁽⁶⁷⁾, la formación del cartílago *in vivo* acontece en un ambiente de hipoxia. También se ha vislumbrado la intervención de MAFK en la diferenciación de adipocitos⁽⁶⁸⁾.

1.4.3 Desarrollo embrionario.

Investigaciones efectuadas en modelos animales y seres humanos indican que los transcritos de estos genes varían significativamente entre los distintos tejidos, así como de manera temporal durante etapas particulares del desarrollo. Por ejemplo, en el ratón adulto *MAFF* sólo se expresa en abundancia en el pulmón⁽²⁹⁾; mientras que en humanos, la expresión de *MAFF* es significativa en el miometrio durante el parto debido a la inducción por parte de citocinas proinflamatorias como IL1B; en la placenta es menor la expresión⁽⁵⁰⁾. En los

⁹ Las células PC12 derivan de un feocromocitoma de rata y tienen la peculiaridad de diferenciarse hacia un fenotipo neuronal en respuesta al factor de crecimiento neuronal (NGF). Las células tratadas con NGF desarrollan neuritas y secretan neurotransmisores.

ratones recién nacidos, se observa una mayor expresión de *MAFF* en los queratinocitos, el cartílago y el periosteo.

En el hígado fetal murino, los transcritos *MAFK* proliferan en paralelo con la actividad hematopoyética^(69,70). Inicialmente, en el ratón, la expresión de *MAFK* se restringe a los tejidos mesodérmicos, luego se vuelve más prominente en las células del mesénquima. Posteriormente, esta expresión se debilita a la par que aumentan los niveles de *MAFK* en el hígado fetal, más tarde, *MAFK* es más abundante en el tubo neural. En los ratones, *MAFK* y *MAFG* presentan patrones de expresión casi perfectamente complementarios en los embriones, pero su patrón de expresión se vuelve complicado conforme progresa el desarrollo⁽⁷¹⁾. Los animales *MAFG*^{-/-} son viables y fértiles pero presentan una proliferación anormal de megacariocitos asociada con trombocitopenia, así como alteraciones neuronales. Por el contrario, no se detectan defectos aparentes en los animales *MAFK*^{-/(71)}. Esto demuestra que si bien las proteínas MAF pueden complementarse en las etapas iniciales de la formación del embrión, otras funciones podrían no ser compensadas adecuadamente. Poco después del nacimiento, los patrones de expresión de *MAFG* y *MAFK* se expanden y son más complejos, abarcando numerosos tejidos y órganos⁽⁷¹⁾. En particular, la expresión es mayor en los tejidos proliferantes, como la epidermis, los folículos pilosos, el tracto digestivo y respiratorio y la médula espinal. Esto señala la participación de las proteínas MAF en el crecimiento o división celular.

En los ratones, *MAFG* es mucho más abundante que *MAFK* en el cerebro anterior, mesencéfalo, cerebelo, capa celular del ganglio retiniano y el epitelio sensorial del oído interno⁽⁷²⁾.

1.4.4 Redundancia funcional de las proteínas MAF.

Dado el alto grado de identidad presente en las proteínas MAF, se ha propuesto que existe una elevada redundancia funcional entre ellas, hipótesis que se ha evaluado mediante el uso de modelos murinos. Los ratones *knockout* individuales *MAFF*^{-/-}, *MAFK*^{-/-} son aparentemente fértiles y normales. Por el contrario, los ratones *MAFG*^{-/-} se caracterizan por una proliferación anormal de megacariocitos, trombocitopenia¹⁰ y trastornos neurológicos moderados^(63,71,73). Los ratones doble *knockout* *MAFF*^{-/-}/*MAFK*^{-/-} también muestran un fenotipo en apariencia normal y son fértiles; mientras que los ratones *MAFF*^{-/-}/*MAFG*^{-/-} muestran una ligera exacerbación del fenotipo *MAFG*^{-/-}. Sin embargo, los ratones *MAFK*^{-/-}/*MAFG*^{-/-} presentan ataxia posterior crónica, trastornos en la diferenciación de células hematopoyéticas⁽⁷¹⁾ y mueren poco después del nacimiento. Los estudios anteriores son consistentes con una redundancia parcial de la función de las tres proteínas MAF.

1.4.5 Asociación con enfermedades.

Debido a su importancia como reguladores de genes citoprotectores y de la diferenciación celular, se han reportado alteraciones en la función de las proteínas MAF en diversas enfermedades humanas. En pacientes con diabetes tipo I se ha constatado una reducción en la expresión de *MAFF*. Por otra parte, se ha descubierto que el tratamiento de estos pacientes con ácido fólico para mejorar la función de las células progenitoras del endotelio aumenta los niveles de *MAFF*. Esta es la primera referencia de un vínculo entre *MAFF* y la diabetes mellitus de tipo 1⁽⁷⁴⁾. De manera similar, la expresión de *MAFG* está ligada en la respuesta hacia el humo del tabaco y el riesgo de cáncer en fumadores⁽⁵⁷⁾. Finalmente, la

¹⁰ Disminución persistente en el número de plaquetas.

amplificación génica de *MAFK* ha sido reportada en un cáncer pancreático familiar⁽⁷⁵⁾.

Uno de los aspectos que se ha asociado con la aparición de numerosas enfermedades es la presencia de cambios en el DNA que influyen en el funcionamiento de los genes. Los polimorfismos más comunes en el genoma humano son los polimorfismos de un solo nucleótido (SNPs), de manera que es necesaria una breve digresión para resaltar su importancia en los estudios genéticos.

1.5 Los polimorfismos y su importancia en el desarrollo de enfermedades.

Un polimorfismo genético es una variación en la secuencia del DNA que ocurre en al menos el 1% de la población general⁽⁷⁶⁾. Los polimorfismos de un solo nucleótido, SNPs (del inglés *single nucleotide polymorphisms*) son variantes generalmente bialélicas, que aparecen en una ubicación específica en el genoma. Los SNPs son las variaciones más comunes en el genoma humano ya que constituyen > 90% de la variación genética total⁽⁷⁷⁾. El tipo más común de SNP son las transiciones. Dado que la 5-metilcitosina es la base modificada más común y puede convertirse espontáneamente en timina, los SNPs C/T son los más numerosos en el genoma humano⁽⁷⁸⁾. Hasta el momento se han identificado más de 11 millones de SNPs en el genoma humano.

Los SNPs pueden ser codificantes o no codificantes dependiendo de su ubicación dentro de la región génica. Los SNPs codificantes se clasifican en dos grupos: i) los **sinónimos**, no generan un cambio de aminoácido, de manera que no alteran la secuencia primaria de la proteína y ii) los **no sinónimos**, que sí modifican la secuencia de aminoácidos de la proteína. El cambio de un aminoácido, ocasionado por un SNP, podría alterar la función proteica debido a un mal plegamiento, cambio de polaridad, fosforilación inadecuada y otras

repercusiones funcionales. Estos SNPs son los que se han asociado más frecuentemente con la susceptibilidad a desarrollar diversos padecimientos. Aunque los SNPs sinónimos no alteran la secuencia proteica, se ha propuesto que podrían ejercer cambios funcionales debido a la alteración de la cinética del plegamiento de las proteínas, causada por el empleo de codones no frecuentes⁽⁷⁹⁾.

Por otro lado, los SNPs situados en las regiones regulatorias de un gen (**rSNPs**) pueden afectar la transcripción, por consiguiente modifican la expresión de las proteínas correspondientes. En distintos genes se han reportado rSNPs que confieren susceptibilidad hacia patologías como la hipercolesterolemia, la hiperbilirrubinemia, el infarto al miocardio, el daño pulmonar agudo, y el asma⁽⁸⁰⁾.

1.6 Haplotipos (*Haploid genotype*).

En el año 2002 se inició el Proyecto Internacional del HapMap con el objetivo de catalogar las variaciones genéticas comunes aparecidas en el genoma humano, así como la distribución de estas variantes en la población mundial. Para lograr esto se genotipificaron 3.1 millones de SNPs utilizando 269 muestras de cuatro poblaciones¹¹. Un resultado importante de este proyecto fue la identificación de los bloques de haplotipos (regiones del genoma en las que los SNPs físicamente cercanos se heredan juntos). Esto permite trazar los patrones de herencia de grandes regiones del genoma analizando únicamente un subconjunto de SNPs (SNPs etiqueta) en cada bloque de haplotipos. La información derivada del proyecto HapMap, junto con el uso de ensayos de genotipificación en todo el genoma, ha permitido la identificación de genes asociados con enfermedades humanas como el infarto al miocardio, la diabetes y el asma⁽⁸¹⁾.

¹¹ **YRI** (Yoruba en Ibadan, Nigeria), **CHB** (chinos Han en Beijing, China), **JPT** (japoneses de Tokio, Japón) y **CEU** (residentes de Utah con ancestría del norte y occidente de Europa).

Los haplotipos constituyen un conjunto de SNPs estadísticamente asociados y presentes en el mismo cromosoma⁽⁷⁶⁾. Debido a que el genoma experimenta recombinación y esto puede involucrar a grandes segmentos de DNA, varios SNPs pueden permanecer ligados a través del tiempo. De esta manera, un grupo de SNPs es co-heredado de generación en generación en grandes bloques está en desequilibrio de ligamiento, LD (del inglés *linkage disequilibrium*). El LD entre variantes genéticas sucede como resultado de la no segregación independiente de las mismas en los cromosomas de una población, lo que ocasiona que dos o más variantes en posiciones físicas independientes se comporten como si fuera la misma. En general, el grado de LD se mide mediante los parámetros D' y r^2 , los cuales capturan diferentes aspectos de una asociación no aleatoria⁽⁸²⁾. De esta forma D' , una medida de la recombinación, adquiere valores desde 0 hasta 1, en donde 1 indica un LD completo (no recombinación); es decir, los dos SNPs siempre se heredan juntos. Un valor $D' \geq 0.8$ señala un LD significativo. Por su parte, r^2 , es una medida de correlación indicativa de la proporción en la que dos SNPs aparecen juntos.

El mapeo de genes responsables de la aparición de enfermedades complejas se basa en la identificación de una asociación entre los marcadores polimórficos, ya sea individualmente o como haplotipos, con el *loci* de susceptibilidad a una enfermedad.

2 ANTECEDENTES.

2.1 SNPs en genes inducibles de fase II y asociación con enfermedades.

Existen muy pocos reportes referentes a la presencia de polimorfismos en los genes *MAF*. Se ha encontrado una alteración de número de copia para *MAFK* en un caso familiar de carcinoma de páncreas⁽⁷⁵⁾. En fumadores con cáncer pulmonar los niveles de expresión de *MAFG* se han relacionado con un bajo nivel de enzimas antioxidantes (*GCLC*¹², *NQO1*, *TXNRD1*, *SLC7A11*¹³)⁽⁵⁷⁾. En ese estudio se confirmó una correlación entre la expresión de *MAFG* y la de varios genes poseedores de sitios ARE. Esto revela el alcance de identificar variantes en los genes *MAF* que potencialmente pudieran alterar la susceptibilidad hacia el desarrollo de algunas enfermedades.

Las proteínas *MAF* participan activamente en la regulación de diversas enzimas antioxidantes y desintoxicantes, principalmente a través de la ruta Nrf2-Keap1⁽⁷⁾. La presencia de polimorfismos en algunos de los genes antioxidantes y desintoxicantes regulados por los genes *MAF* se ha asociado con diferentes enfermedades. Por ejemplo, polimorfismos en *NQO1* se han vinculado a una predisposición a la leucemia linfoblástica aguda¹⁴⁽⁸³⁾, cáncer hepático⁽⁸⁴⁾ y de mama⁽⁸⁵⁾. Por otro lado, variantes en el gen que codifica para la glutatión S-transferasa se han asociado con la susceptibilidad a cáncer de esófago⁽⁸⁶⁾ y asma⁽⁸⁷⁾. Así también, el polimorfismo (GT)_n, un repetido localizado en el promotor

¹² Es la subunidad catalítica de la glutamato cisteína ligasa, implicada en la síntesis del glutatión. La expresión de enzimas productoras del glutatión está reducida en diversas condiciones, como en la diabetes mellitus, hepatitis alcohólica, envejecimiento, endotoxemia y colestasis crónica.

¹³ Un transportador de cistina. La cistina es un dímero de dos cisteínas unidas por un puente disulfuro.

¹⁴ La leucemia linfoblástica aguda (ALL) se caracteriza por un bloqueo en la diferenciación de los progenitores linfoides tempranos, esto se refleja en una acumulación de linfoblastos inmaduros.

del gen *HMOX1* es capaz de regular la expresión génica. Los repetidos (GT)_n más largos están asociados con un aumento de la susceptibilidad a enfermedades pulmonares como el enfisema inducido por el tabaquismo, neumonía y asma^(88,89). Este microsatélite también se ha relacionado con otras enfermedades como la artritis reumatoide⁽⁹⁰⁾, malaria cerebral⁽⁹¹⁾, trastornos por trasplante renal⁽⁹²⁾ y padecimientos de las arterias coronarias⁽⁹³⁾. Otros polimorfismos en el promotor de *HMOX1* también están asociados con el cáncer gástrico y tumores linfovasculares⁽⁹⁴⁾. Polimorfismos funcionales en el promotor de *NRF2* se han asociado con algunos padecimientos¹⁵ humanos como el daño pulmonar agudo⁽⁹⁵⁾, vitíligo⁽⁹⁶⁾, úlcera péptica⁽⁹⁷⁾, colitis ulcerante⁽⁹⁷⁾ y cáncer gástrico⁽⁹⁸⁾. Además, ciertos haplotipos conformados por SNPs reguladores de *NRF2* afectan la actividad del gen y están asociados con la gravedad de la enfermedad pulmonar obstructiva crónica (COPD)⁽⁹⁹⁾ y protección contra la enfermedad de Parkinson⁽¹⁰⁰⁾.

¹⁵ Consiste en comparar la morbilidad (ocurrencia de una enfermedad) observada en uno o varios grupos de individuos inicialmente indemnes a la enfermedad y definidos en función de su exposición a un factor de riesgo sospechoso de provocar esta enfermedad.

3 JUSTIFICACIÓN.

Numerosos polimorfismos presentes en genes antioxidantes se han vinculado con el desarrollo de diversas patologías. Sin embargo, a pesar de la relevancia de las proteínas MAF pequeñas como reguladores de la expresión génica citoprotectora, y de que su mal funcionamiento ocasiona serios trastornos, todavía no se tiene un registro completo de los SNPs presentes en estos genes y no se han emprendido estudios de secuenciación poblacionales. Por consiguiente, consideramos importante determinar la presencia de SNPs funcionales en estos genes.

4 OBJETIVO GENERAL.

Identificación de las variantes génicas en *MAFF*, *MAFG* y *MAFK* en una muestra de la población mexicana sana.

4.1 Objetivos específicos.

- Generación de un banco de muestras de DNA a partir de 112 individuos de la población mexicana no relacionados entre sí.
- Identificación de polimorfismos en las regiones promotora y codificantes.
- Determinación de haplotipos.
- Identificación de los SNPs etiqueta.
- Analizar *in silico* los posibles efectos funcionales de los polimorfismos encontrados.

5 MATERIALES Y MÉTODOS.

5.1 Población de estudio.

En esta investigación se incluyó un grupo de 112 individuos mexicanos, mayores de 18 años, obtenidos del banco de sangre del Instituto Nacional de Pediatría (72.3 % hombres; 27.7 % mujeres). Cada uno de los participantes firmó un consentimiento informado. Con el tamaño de muestra utilizado en este estudio se tiene una probabilidad >95% de identificar a los sitios polimórficos con una frecuencia alélica de al menos el 1%.⁽¹⁰¹⁾

5.2 Extracción de DNA genómico.

El DNA se extrajo a partir de 10 mL de sangre periférica utilizando el kit QIAamp Blood Maxi, de acuerdo al protocolo del proveedor (QUIAGEN 2005, California), el cual se describe brevemente a continuación. En un tubo de 50 mL se añadieron 10 mL de sangre periférica, centrifugándose a 2 500 rpm por 20 minutos. El botón de linfocitos obtenidos se transfirió a un tubo nuevo con 500 μ L de proteinasa y mezclado por inversión moderada. A continuación se vertieron 12 mL de amortiguador AL¹⁶ y el contenido se mezcló invirtiendo el tubo ~15 veces y después con agitación vigorosa durante al menos un minuto hasta conseguir una solución homogénea, para posteriormente incubar el lisado celular a 70 °C durante 10 min. La adhesión selectiva de DNA a la membrana de la columna se consiguió añadiendo 10 mL de etanol absoluto al tubo Falcon con la muestra y mezclando el tubo por inversión diez veces, seguido por una fuerte agitación. La solución del paso anterior se vertió en una columna QIAamp colocada en un tubo de centrífuga de 50 mL y se centrifugó a 3 000 rpm por 3 minutos para eliminar todas las impurezas al desechar el filtrado. El DNA adherido a la membrana se lavó vertiendo 5 mL de búfer AW1 a la columna y centrifugando después a 5 000 rpm

¹⁶ Contiene clorhidrato de guanidina y sales caotrópicas.

durante 1 minuto. A continuación se añadieron 5 mL de amortiguador AW2 a la columna y se centrifugó a 5 000 rpm por 15 minutos. La columna con el DNA se colocó en un tubo limpio, descartándose. Después se añadió 1 mL de búfer AE directamente sobre la membrana de la columna y se dejó incubar a temperatura ambiente por 5 minutos, después se centrifugó a 5 000 rpm por 2 minutos. El DNA obtenido fue cuantificado mediante espectrofotometría (ND 1000 UV-Vis v3.2, Thermo Fisher Scientific, EUA). Una vez cuantificadas las muestras, se efectuaron diluciones con agua, para obtener una concentración de 100 ng/ μ L de DNA. La pureza del templado debió cumplir con un cociente de absorbancia de 260/280 entre 1.7 a 1.9. Las diluciones fueron almacenadas en alícuotas a una concentración de 100 ng/ μ L a - 20 °C.

Integridad de DNA.

La integridad de DNA se evaluó mediante electroforesis en gel de agarosa al 1 % con amortiguador TBE 1x (Tris-Borato-EDTA) a 90 V por 35 minutos. El DNA fue visualizado por tinción con bromuro de etidio (agente intercalante) a una concentración de 0.5 μ g/mL. Para la toma de imágenes se empleó el transiluminador Gel Logic 1500 Digital Imaging System (KODAK).

5.3 La reacción en cadena de la polimerasa (PCR)

La PCR permite la amplificación exponencial de secuencias específicas de DNA. Una reacción de amplificación estándar incluye al DNA molde, una DNA polimerasa termoestable, dos oligonucleótidos, desoxinucleótidos trifosfato (dNTPs), búfer de reacción y magnesio. Una vez reunidos los componentes, la reacción se coloca en un termociclador, un instrumento necesario para someter la mezcla de reacción a distintas temperaturas durante un tiempo predeterminado. Esta serie de ajustes de temperatura y tiempo se conoce como un ciclo de amplificación. En teoría, cada ciclo de PCR duplica la cantidad de secuencia

blanco (amplicón) en la reacción. Diez ciclos tentativamente multiplican al amplicón alrededor de 1, 000 veces; 20 ciclos por un factor de un millón.

Cada ciclo de PCR incluye los pasos de desnaturalización del DNA molde, hibridación de los oligonucleótidos y síntesis de las nuevas cadenas (extensión).

a) La desnaturalización del DNA molde se consigue mediante calentamiento de la mezcla de reacción a 94 °C o más durante un minuto. En este paso las dos cadenas del DNA molde son separadas, produciendo el DNA molde de una sola cadena requerido para la replicación por parte de la DNA polimerasa termoestable. En el siguiente paso del ciclo, se reduce la temperatura hasta alrededor de 40 – 60 °C. A esta temperatura, los oligonucleótidos cebadores pueden formar asociaciones estables (hibridación) con el DNA blanco desnaturalizado y funcionan como cebadores para la DNA polimerasa. Este paso perdura durante aproximadamente 15 – 60 s. Finalmente, la síntesis del nuevo DNA comienza conforme se eleva la temperatura de reacción hasta la óptima de trabajo de la DNA polimerasa. Para la mayoría de las DNA polimerasas termoestables esta temperatura se ubica en el rango de 70 – 74 °C. El paso de extensión dura aproximadamente 1 – 2 minutos. El siguiente ciclo inicia con el regreso a la temperatura de 94 °C para la desnaturalización. Cada paso del ciclo debe optimizarse para cada combinación de DNA molde y oligonucleótidos. Después de 20 – 40 ciclos, el producto amplificado puede ser analizado para conocer su tamaño, cantidad, secuencia, etc.

Para amplificar la región promotora y codificante de los genes *MAF* pequeños, las muestras de DNA se desnaturalizaron a 95 °C por 60 segundos, posteriormente se disminuyó la temperatura a 60 °C durante 45 segundos, para permitir la hibridación de los oligonucleótidos cebadores con las cadenas de DNA molde. Después, se realizó el paso de extensión, para amplificar la región genómica deseada, elevando la temperatura a 72 °C, durante 1 minuto. El número de ciclos de amplificación utilizados fue de 30. Al concluir el último ciclo de

amplificación, las muestras se incubaron a 72 °C por 10 minutos para rellenar los extremos salientes de los productos recién sintetizados. Para las reacciones de amplificación se utilizó una enzima de alta fidelidad de tipo “*hot start*”: la JumpStart AccuTaq LA DNA polimerasa (D5809, Sigma-Aldrich, EUA).

Debido al alto contenido de GC en algunas de las regiones analizadas, se incluyeron diferentes aditivos como el dimetil sulfóxido (DMSO) o el GC-Melt (Clontech) para mejorar el rendimiento de la PCR y disminuir las amplificaciones inespecíficas.

5.3.1 Diseño de oligonucleótidos para la secuenciación de los genes MAF.

Las secuencias de los genes *MAF* humanos se obtuvieron del GenBank y sirvieron como referencia, tanto para el diseño de los oligonucleótidos como para la ubicación de los polimorfismos: *MAFF* geneID: 23764 (www.ncbi.nlm.nih.gov/gene/23764), secuencia genómica: NT_011520.12 (*Genome Reference Consortium Human Build 37*, GRCh37); *MAFG* geneID: 4097 (www.ncbi.nlm.nih.gov/gene/4097), secuencia genómica NT_010663.15 (GRCh37); *MAFK* geneID: 7975 (www.ncbi.nlm.nih.gov/gene/7975), secuencia genómica: NT_007819.17 (GRCh37).

Los oligonucleótidos para la amplificación de los genes *MAF* se diseñaron utilizando el programa *PrimerSelect* (Lasergene v7.1, DNASTar). Todos los oligonucleótidos tuvieron una longitud de entre 18 a 25 pares de bases; un contenido GC del 40 – 60% y ausencia de residuos G o C consecutivos en el extremo 3' para minimizar la hibridación inespecífica. Los oligonucleótidos utilizados presentaron una zona de traslape de al menos 100 pb entre amplicones contiguos, de esta forma los picos poco definidos, típicos de las porciones inicial y final de los electroferogramas, pudieron resolverse perfectamente en la secuencia del amplicón siguiente. Todos los oligonucleótidos utilizados se muestran en la

Tabla 5-1, Tabla 5-2 y Tabla 5-3. Las posiciones de los nucleótidos, basadas en la secuencia genómica de cada gen *MAF*, se numeraron tomando como referencia la primera base (+1) en el sitio de inicio de la transcripción (TSS). Para cada gen *MAF* se amplificó alrededor de 1 kb río arriba del TSS (a fin de incluir al promotor mínimo y sus elementos proximales), los tres exones y las regiones intrónicas inmediatamente adyacentes.

Para la amplificación de los genes *MAF* se utilizó una mezcla de reacción de 25 µL con 100 ng de DNA genómico, 0.62 U de la enzima JumpStart AccuTaq LA DNA polimerasa, búfer 10X con 2.5 mM de MgCl₂, dNTPs a 0.2 mM y 0.4 µM de cada oligonucleótido. La temperatura de alineamiento, así como las condiciones de PCR particulares para cada par de oligonucleótidos se muestran en la **Tabla 5-4, Tabla 5-5 y Tabla 5-6**. Las reacciones de PCR se corrieron en un termociclador GeneAmp PCR 9700 (Applied Biosystems, Foster City CAL, EUA).

Tabla 5-1. Oligonucleótidos utilizados en la amplificación del gen *MAFF*.

Amplicón	Oligonucleótido <i>directo</i> (5' a 3')	Mer	Oligonucleótido <i>reverso</i> (5' a 3')	Mer	Longitud amplicón (pb)
Prom1	TCCGGGCTGACGTGAAAGAT	20	CCCAAAAACCCGCTCCTGAT	20	484
Prom2	TTGTCCGGGTAAGCTGATGGTG	22	CCCCGCCCCCGTATGAC	17	573
Exón 1	AGGGGGCGGGGTCATACG	18	ATCCCGGGACGCAGAAACC	19	415
Exón 2	TAGGCGGGTTATGTGATGTGA	22	GGCCTGGTGCCTGCTGAG	18	513
Exón 3 ^a	GGCGGCTAAGGCGGTGAA	18	AGCCCGGGGTGGACTTGAC	19	490
Exón 3 ^b	GCGCCGCACACTCAAAAAC	19	TCTGCCCATCCCCAACCTAC	21	551
Exón 3 ^c	CTTCGTGGGCCCTGTCTTCCT	21	GGCCCTCCCACCACATTTCTG	21	543
Exón 3 ^d	TGCTGGCCTTCCGAACTAAA	22	TCAGCCCCCACCCACAGT	19	491
Exón 3 ^e	GGGGGCTGGCCAGTCAGAG	19	GCGCCCGGCTTAAACATTTTT	21	421
Exón 3 ^f	AAGCCGAGAGCAACAACAAA	22	CCGCCACCAGAAGCCCTAAG	20	530

Tabla 5-2. Oligonucleótidos utilizados en la amplificación del gen *MAFG*.

Amplicón	Oligonucleótido <i>directo</i> (5' a 3')	Mer	Oligonucleótido <i>reverso</i> (5' a 3')	Mer	Longitud amplicón (pb)
Prom1	GGCCCCGTGTGCTGTTTTGTA	21	GGGAGACTTCGCTGCTTGTG	22	498
Prom2	AGGCGGAGGGACGCTTCTGAG	21	GGCCCCGCGATTCCAACCTC	19	427
Prom3	CCCCCTACCGCTGGGCATCT	20	GCGGCGGGAGGGCAAACCT	18	525
Prom4	GGGCTGGGGTGGAGACCT	19	GCCGCAGCCGCACTCAC	17	432
Exón 1	CCGCGTCGCCGCTTCAT	17	GCGGCAGACCCGAAACCAG	19	401
Exón 2	GGCCAAGGGCTGCCTCCTATC	21	AGCGTAGCCGCGTTCTTGAG	21	478
Exón 3 ^a	AGCGGGAGCCGGTGAGAAT	20	GCCTAGTGGCCCCGAAAGAC	21	492
Exón 3 ^b	GACGCGCTGCGCTCCAAGTA	20	TCTAGTCCGTTCCACAGCACAA	24	484
Exón 3 ^c	CCGCCACCAGCGTCATCAC	19	CTTCAAGCCCGCCTACTCTCA	22	563
Exón 3 ^d	CGCGGGCTGGGAAGGATT	18	CAGCCACAAGAGGTGTCCGTAAG	24	557

Tabla 5-3. Oligonucleótidos utilizados en la amplificación del gen *MAFK*.

Amplicón	Oligonucleótido <i>directo</i> (5' a 3')	Mer	Oligonucleótido <i>reverso</i> (5' a 3')	Mer	Longitud amplicón (pb)
Prom1	CCCCGGGGATGTCTTTCTG	20	GAGCCCTTGCCACGCTTCT	20	514
Prom2	GTTGCCCAGGCCTGAACTCC	21	TGCCTGCGCCCCAGATA	18	461
Prom3	GCCGCTATCCCATTTTTACATT	24	TGGAGCCGGGAGTTCAAGA	20	420
Prom4	GGGCCGAGGCACATAACC	18	CCCGGCTCCCCAGCATC	17	457
Exón 1	TTCCAGCTGGCCTTAATGTTTT	24	GTGACTCAGCGCCAGGTAGC	21	560
Exón 2	GCTGGGCCCGATGGTG	17	AGGCGGAAGGCAGCAGAC	19	473
Exón 3 ^a	ATGGGCTCTGAGGCTGTCGGTAAC	24	GTAGCCGCGTTCTTGAGTGTGC	23	418
Exón 3 ^b	AGAACGCCCGGTGCTCA	18	GGGCTGCGCTTGGACTC	18	533
Exón 3 ^c	GACGCCCTGCGCTCCAAGTA	20	TTCCCGAGGCGCTTCCATTAT	22	473
Exón 3 ^d	GCTCCCCGAGGATGTGTC	19	AGGCCCTCCCCAGTGAAC	19	484

Tabla 5-4. Condiciones óptimas de PCR para los amplicones *MAFF*.

Oligonucleótidos	DNA polimerasa	Tm (°C/1 min)	Textensión	Ciclos	MgCl ₂ (mM)	DMSO (μL)	GC Melt (μL)
Prom1	TaqGold	62	72°C/30s	30	2.0	1	----
Prom2C-Exon1	JumpStart	63	68°C/30s	"	2.5	1	2.5
Exon1	TaqGold	62	72°C/30s	"	1.0	1	----
Exon2	" "	62	" "	"	1.0	1	----
Exon 2 ^b	" "	64	" "	"	2.0	1	----
Exon3 ^a	" "	64	" "	"	1.0	1	----
Exon3 ^b	" "	62	" "	"	1.0	1	----
Exon3 ^c	" "	64	" "	"	2.0	1	----
Exon3 ^d	" "	64	" "	"	1.0	1	----
Exon3 ^e	" "	64	" "	"	1.0	1	2.5
Exon 3 ^f	" "	62	" "	"	1.0	1	----
Exon 3 ^f INS	" "	60	" "	"	1.0	1	2.5

Tabla 5-5. Condiciones óptimas de PCR para los amplicones *MAFG*.

Oligonucleótidos	DNA polimerasa	Tm (°C/1min)	Textensión	Ciclos	MgCl ₂ (mM)	DMSO (μL)	GC Melt (μL)
Prom1	JumpStart	60	68°C/30s	30	2.5	1	----
Prom1	Taq gold	60	72°/30s	''	1	1	2
Prom2	JumpStart	62	68°C/30s	''	2.5	1	----
Prom2	Taq gold	62	72°/30s	''	1	1	2.5
Prom1F – Prom2R	Taq gold	60	72°/30s	''	1	1	2.5
Prom3	TaqGold	60	72°C/30s	''	1	1	2
Prom4	JumpStart	68	68°C/30s	''	2.5	1	2.5
Prom4	GC Genomic LA polimerasa	66	72°/30s	''	2.5	1	2.5
Exón 1	JumpStart	60	68°/30s	''	2.5	1	2.5
Prom4F-Ex1R	Jump Start	68	68°/60s	''	2.5	1	2.5
Exón 2	JumpStart	60	68°/30s	''	2.5	1	2.0
Exón 3 ^a	Taq gold	58	72°/30s	''	1.5	1	----
Exón 3 ^b	Taq gold	58	72°/30s	''	1.5	1	----
3 ^o F – 3 ^o R	Taq gold	58	72°/30s	''	1.5	1	----
Exón 3 ^c	Taq gold	60	72°/30s	''	1.0	0	----
Exón 3 ^d	Taq gold	60	72°/30s	''	1.0	0	----
3 ^o F – 3 ^o R	GC Genomic LA polimerasa	67	72°/30s	''	2.5	0	----
3 ^o F – 3 ^o R	Taq gold	67	72°/30s	''	2.0	1	2.5
Exón 3 ^e	Taq gold	60	72°/30s	''	1.0	0	----
3 ^o F – 3 ^o R	Taq gold	65	72°/30s	''	1.0	1	2.5
[3 ^o F – 3 ^o R]	GC Genomic LA polimerasa	67	72°/30s	''	2.5	0	----

Tabla 5-6. Condiciones óptimas de PCR para los amplicones *MAFK*.

Oligonucleótidos	DNA polimerasa	Tm (°C)/1min	Textensión	Ciclos	MgCl ₂ (mM)	DMSO (μL)	GC Melt (μL)
Prom1	Taq gold	60	72°/30s	30	2.0	0	----
Prom2	Taq gold	60	72°/30s	''	2.0	1	2.5
Prom1F-Prom2R	Taq gold	60	72°/30s	''	2.0	1	2.5
Prom3	TaqGold	62	72°C/30s	''	2.0	1	2.5
Prom4	JumpStart	60	68°C/30s	''	2.5	1	2.5
Prom3F-Prom4R	JumpStart	60	68°C/30s	''	2.5	1	2.5
Exón 1	JumpStart	60	68°/30s	''	2.5	1	2.5
Exón2-Exón3 ^a	TaqGold	62	72°/30s	''	2.0	1	2.5
Exón 2	JumpStart	60	72°/30s	''	2.5	1	2.0
Exón 3 ^a	Taq gold	58	72°/30s	''	1.5	1	----
Exón 3 ^b	JumpStart	60/45s	68°/45s	''	2.5	0	2.5
Exón 3 ^c	Advantage	64	72°/60s	''	2.5	0	----
Exón 3 ^d	Advantage	66	72°/60s	''	2.5	0	----
3 ^o F – 3 ^o R	GC Genomic LA polimerasa	67	72°/30s	''	2.5	0	----
3 ^o F – 3 ^o R	Taq gold	67	72°/30s	''	1.0	1	2.5
Exón 3 ^e	Taq gold	60	72°/30s	''	1.0	0	----
3 ^o F – 3 ^o R	Taq gold	65	72°/30s	''	1.0	1	2.5
[3 ^o F – 3 ^o R]	GC Genomic LA polimerasa	67	72°/30s	''	2.5	0	----

5.4 Electroforesis en geles de agarosa.

Los distintos amplicones *MAF* fueron sometidos a electroforesis en geles de agarosa (Invitrogen) al 2.5% teñidos con bromuro de etidio (0.5 mg/mL). Estos geles fueron preparados con amortiguador TBE 1X (Tris-base, ácido bórico y EDTA). Posteriormente, las bandas de DNA se observaron bajo luz UV empleando un fotodocumentador GEL Logic 1500 Imaging System (KODAK, EUA) y el programa Kodak Molecular Imagen v4.0.5. Para la determinación del tamaño de los fragmentos se utilizaron dos tipos de marcadores de peso molecular, la escalera de DNA de rango bajo (#SM1193) y la escalera *1 Kb Plus* (#SM1333), ambos de Fermentas (Thermo Fisher Scientific, EUA).

5.5 Purificación de los productos de PCR.

Los productos de PCR se purificaron utilizando el QIAquick PCR purification kit, siguiendo las indicaciones del proveedor¹⁷ (QIAGEN). En primer lugar, en un tubo eppendorf de 1.5 mL se agregaron 5 volúmenes de amortiguador PB (búfer de unión) por cada volumen de producto de PCR. La mezcla se añadió a una columna QIAquick colocada en un tubo colector de 2 mL y se centrifugó por 1 minuto, descartándose el sobrenadante. El producto de PCR unido a la membrana de la columna fue lavado con 750 μ L de amortiguador PE y centrifugado durante 1 minuto. El sobrenadante se removió y el etanol residual fue eliminado mediante centrifugación durante 1 minuto adicional. A continuación, la columna se colocó en un tubo nuevo de microcentrífuga de 1.5 mL y se aplicaron 30 μ L de amortiguador EB (amortiguador de elución; 10 mM Tris-Cl, pH 8.5) en el centro de la membrana de la columna. Después de dos minutos de incubación a temperatura ambiente, el DNA fue eluido por centrifugación durante un minuto y colectado en tubos eppendorf de 0.6 mL.

¹⁷ Nota: Se añade etanol absoluto al amortiguador PE antes de utilizarse. Todos los pasos de centrifugación son a 12 000 rpm en una microcentrífuga *Mini Spin* (Eppendorf).

5.6 Secuenciación automatizada.

La secuenciación de los genes *MAF* se llevó a cabo siguiendo el procedimiento de secuenciación BigDye Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, EUA) basado en el proceso automatizado de Sanger⁽¹⁰²⁾. Este método se fundamenta en la incorporación de dideoxynucleótidos en la cadena en crecimiento durante la reacción de amplificación, lo cual impide su posterior elongación debido a la ausencia de un grupo 3'-hidroxilo. El empleo de dideoxynucleótidos acoplados a distintos fluorocromos (A, C, G y T) permite detectar las 4 bases a través de la longitud de onda de excitación particular a cada fluorocromo. Además, mediante el uso de la electroforesis capilar es posible separar fragmentos con una resolución de una base.

Para la reacción de secuenciación se utilizaron 10 ng de cada producto de PCR sometidos a los siguientes parámetros de amplificación: a) desnaturalización inicial a 96°C por 1 minuto; b) 25 ciclos de desnaturalización a 96°C por 10 s, alineamiento a 50 °C por 5 s y extensión a 60 °C por 4 minutos. Una vez concluidos los ciclos de secuenciación, los amplicones se purificaron con el método BigDye XTerminator (Applied Biosystems) para eliminar sustancias que pudieran interferir con la inyección electrocinética, la separación por electroforesis o el análisis de los datos. Posteriormente, los productos purificados se sometieron directamente a una electroforesis capilar utilizando el polímero POP-7 para la separación de los fragmentos de DNA en un secuenciador modelo 3730xl (Applied Biosystems), que cuenta con un arreglo de 96 capilares (36.5 cm de longitud).

Las secuencias directa y reversa de los amplicones *MAF* se compararon con sus respectivos referentes en el GenBank utilizando los programas *Seqman* (Lasergene v7.1, DNASTar), *SeqScape* v2.7 (Applied Biosystems) y *Mutation Surveyor* (Softgenetics). Cada variante nueva se confirmó al amplificar y

secuenciar nuevamente a partir de la muestra original. Esto se realizó para descartar un posible error en la secuenciación o la introducción de mutaciones por parte de la polimerasa. Finalmente, se determinaron las frecuencias alélicas y genotípicas de las variantes encontradas.

5.7 Desequilibrio de ligamiento y análisis de haplotipos.

Los análisis sobre el equilibrio de Hardy-Weinberg y el desequilibrio de ligamiento (LD) se llevaron a cabo utilizando dos programas: *Haploview* v4.1 (Broad Institute) y *SNPAlyze* v7.0 (Dynacom, Co., Yokohama, Japón). Los análisis de pareamiento (*pairwise*) entre las distintas variantes se efectuaron obteniendo los coeficientes D' y r^2 . D' se utilizó para evaluar la probabilidad de recombinaciones pasadas en tanto que r^2 ($r^2 > 0.8$) sirvió como parámetro del grado de ligamiento entre un par de variantes. Algunos de los haplotipos se descubrieron sin ambigüedad identificando manualmente a los individuos con variantes homocigotas en todos los sitios, o que poseían una variante heterocigota en un solo sitio. Los haplotipos inferidos *in silico* con una muy baja frecuencia se agruparon bajo el rótulo de “otros”. Todos los SNPs etiqueta se analizaron para su desviación del equilibrio de Hardy-Weinberg empleando χ^2 .

6 RESULTADOS.

6.1 Amplificación de los genes *MAF*.

Una parte importante de este estudio consistió en la estandarización de las condiciones de amplificación; esto implicó la modificación de parámetros como la concentración de $MgCl_2$, oligonucleótidos, curvas de temperatura de hibridación y la incorporación de aditivos como DMSO (Sigma-Aldrich) y GC-Melt (Clontech). Las estandarizaciones se iniciaron con gradientes de concentración de magnesio (1.0 – 2.5 mM). En la **figura 6-1 (A)** se observa que la presencia de productos de PCR inespecíficos es inversamente proporcional a la concentración de $MgCl_2$. Los amplicones donde el gradiente de magnesio fue insuficiente para obtener una banda específica del tamaño esperado, se sometieron a un gradiente de temperatura de alineamiento. En este caso, el incremento de la temperatura de alineamiento disminuyó la presencia de los productos inespecíficos (**figura 6-1, B**). Debido al alto contenido de GC de las regiones regulatorias fue necesaria la incorporación de aditivos para facilitar la desnaturalización del DNA y obtener los productos de interés (**figura 6-1, C**).

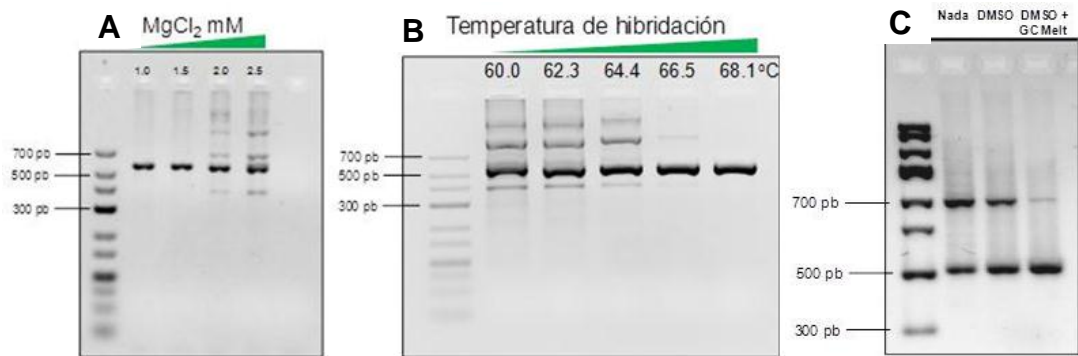


Figura 6-1. Estandarización de las condiciones de amplificación de los genes *MAF*. A. Efecto de la concentración de Mg en la PCR. B. Efecto de la temperatura sobre la PCR. C. Influencia de los reactivos especiales en la PCR.

Una vez obtenido un producto de PCR se realizó la secuenciación de las cadenas *forward* y *reverse* de cada muestra para confirmar en ambas secuencias la presencia de alguna variante. Las variantes más frecuentemente encontradas en nuestro estudio fueron los SNPs, aunque también se encontraron algunas deleciones e inserciones. Por ejemplo, en la **Figura 6-2 A** se muestra una nueva variante identificada en la región 3'-UTR de MAFG con una FAM = 0.126; hasta el momento este polimorfismo no se ha identificado en otros grupos humanos, por lo que tal vez podría ser exclusivo de nuestra población. En algunos casos fue posible advertir la presencia de algunas deleciones desde el corrimiento del producto de PCR en el gel de agarosa debido a la formación de DNA heterodúplex (**Figura 6-2, B**). Los heterodúplex se generan por un apareamiento incompleto de ambas cadenas de DNA debido a la deleción o inserción de bases. A la izquierda del gel se observa la banda de interés (484 pb), pero en algunas muestras se observa un DNA heterodúplex asociado a una deleción de 7 bases (SNP rs34308084, -/ATCTATC). En el panel derecho se observa el electroferograma correspondiente a un individuo normal y un homocigoto para la deleción de 7 pb, confirmada en las secuencias directa y reversa.

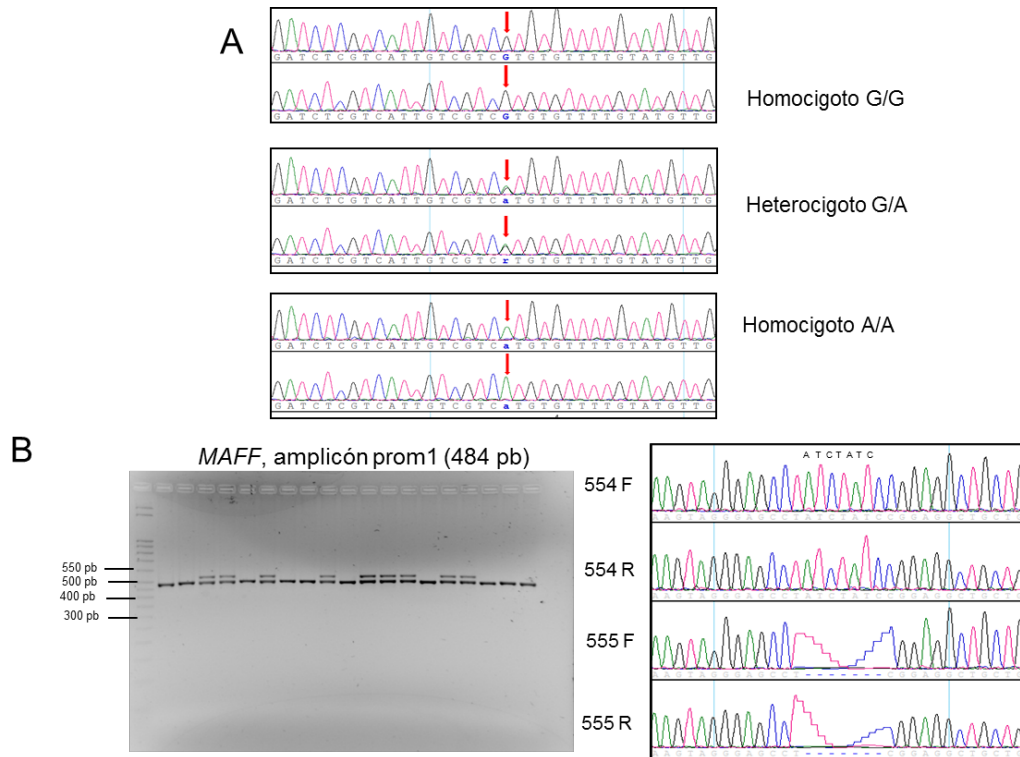


Figura 6-2. Algunas variantes identificadas en los genes *MAF*. **A.** Secuenciación de la región 3'-UTR de *MAFG*, que muestra un nuevo polimorfismo. **B.** Amplicón Prom1 de la región regulatoria de *MAFF* mostrando un DNA heterodúplex en algunas muestras debido a la delección mostrada en el electroferograma a la derecha.

En ocasiones, la presencia de dobles inserciones dificultó el análisis de las secuencias. Este obstáculo se superó utilizando el programa *Mutation Surveyor* v3.24 (Softgenetics, EUA) el cual genera un reporte de mutaciones resaltando las discrepancias en electroferogramas sintéticos separados, esto permitió visualizar fácilmente la doble inserción. En la **Figura 6-3** se muestran los cambios identificados en el amplicón Prom4 de *MAFG*. En **a)** se muestra el traslape de los picos (en amarillo) debido a la presencia de un alelo con dos inserciones de 3 pares de bases. **b)** Mediante el programa *Mutation Surveyor* v3.24 (Softgenetics) se analizó este cambio con respecto a la secuencia de referencia, las inserciones se indican con barras en rojo. Finalmente, en **c)** se reproducen los cambios en la secuencia, las inserciones se marcan con letras de distinto color (GCG y CGG). Las cadenas marcadas como F mut y R mut corresponden al alelo mutado, abajo se muestran las secuencias F y R del alelo normal. En amarillo se muestran las zonas no coincidentes con la secuencia de referencia.

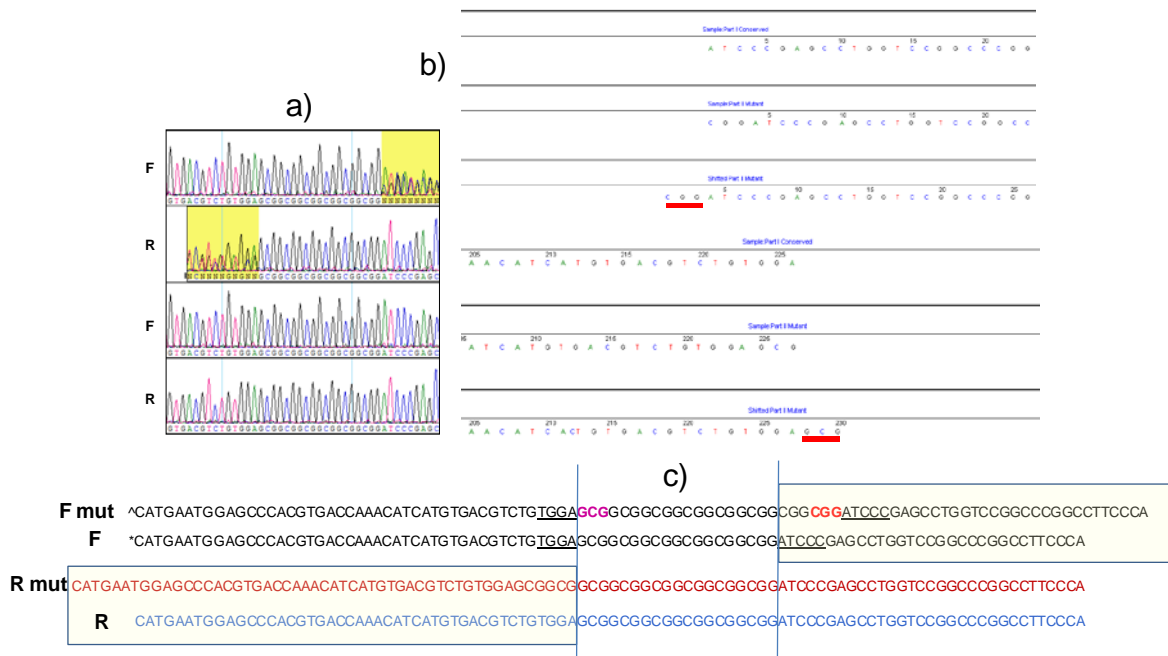


Figura 6-3. Doble inserción heterociga (GCG y CGG) presente en la región reguladora de MAFG.

6.2 Polimorfismos en los genes *MAF*.

El número de variantes identificadas fue distinto para cada gen. En el caso de *MAFF* se descubrieron 19 variantes, 6 de ellas nuevas (**Tabla 6-1**). De las variantes identificadas 7 se ubicaron en la región promotora, 4 en la región 5'-UTR, 2 en los intrones y 6 en la región 3'-UTR. Todas las variantes identificadas con una frecuencia > 30 % estuvieron en equilibrio de Hardy-Weinberg ($p \geq 0.05$)¹⁸.

Tabla 6-1. Variantes encontradas en el gen *MAFF*.

Este estudio	SNP ID		Posición		Cambio de nucleótido	Frecuencia del alelo menor
	dbSNP (NCBI)	Ubicación	NT_011520.12 Gen ID: 23764	Desde el TSS (NM_001161572.1)		
	rs2413508	Promotor	17987743	-765	TGTTTGGGGGACAG>CAAAGAAGGGCACT	0.467
	rs34308084	Promotor	17987882:17987888	-626	AGTAGGGAGCCTC-/ATCTATCGGAGGCTGCTGCC	0.446
	rs4374456	Promotor	17987946	-562	CATGACTGGAGGTG>CGTGAGAAATGGTC	0.446
	rs4608623	Promotor	17987947	-561	ATGACTGGAGGTG>TTGAGAAATGGTCC	0.446
	-550 ^a	N.V.	17987958	-550	TGGTGAGAAATGGT>CCCTATGCAGGAGA	0.004
	-514 ^a	N.V.	17987994	-514	ACCGCTAGCTTGA>ATCCTGAGGTGGGG	0.004
	-15 ^a	N.V.	17988493	-15	AAGCGCCGCTCCT>CGGCCGAGGGCGG	0.013
	36 ^a	N.V.	5'-UTR	36	ACTGGGTTTTTATG>TAATGAAAGGAATC	0.004
	rs11703601	5'-UTR	17988610	103	CGCGCGCCTCGG>TCCCCTGCGCCGCC	0.408
	rs909629	5'-UTR	17988663	156	GCTCCCGGGCGCC>GGGTTTCAGAGCGAC	0.004
	274 ^a	N.V.	5'-UTR	274	ACCAAGCGGGCCC>A>TGAAGCGGGTGAGG	0.004
	rs2267372	Intrón 1	17988803	296	TGAGGAACGCGGG>G>ACCAGGAGGACGGT	0.442
	rs2235264	Intrón 2	18000519	12012	CCCTGCCCATGCA>G>AAGCCTCCTCATCC	0.463
	rs76165778	3'-UTR	18001614	13107	CCTCTTGCAGCCC>ACCAAAGTGGGACC	0.008
	rs9610915	3'-UTR	18001649	13142	CTGGGAAGGGGAA>C>GTTGGGTAGGTTGG	0.483
	rs11544353	3'-UTR	18001867	13360	TGTGACTTACGGG>A>CCCGTCTGCTGAG	0.004
	13773 ^a	N.V.	3'-UTR	18002280	GCCAGTCAGAGAA>A>GGGGGGCCATGGA	0.004
	rs4821766	3'-UTR	18002780	14273	ATGGTGAAACCCC>G>ATCTCTACAAAAA	0.042
	rs5995550	3'-UTR	18002941	14434	GCAACATAGCAAA>A>CTCTGTCTCAAAA	0.408

^a Nuevas variantes detectadas en este estudio. N.V. = Nueva variante.

¹⁸ La ausencia de desviación implica que estas variantes no son nuevas, y que los eventos que los incorporaron en la población mexicana (mutación o migración), se han diluido en el tiempo.

Por otro lado, *MAFG* presentó 9 variantes, de las cuales 8 son nuevas. Cinco de las variantes se identificaron en la zona promotora del gen, una en la región 5'-UTR, una en el intrón 1 y dos en la región 3'-UTR. Con la excepción de una de las dos variantes detectadas en la región 3'-UTR ($f = 0.128$), todos los cambios encontrados en este gen presentaron frecuencias menores a 0.09 (**Tabla 6-2**).

Tabla 6-2. Variantes encontradas en el gen *MAFG*.

SNP ID		Ubicación	Posición		Cambio de nucleótido ^b	Frecuencia del alelo menor
Este estudio	dbSNP (NCBI) ^b		NT_010663.15 Gen ID:4097	Desde el TSS (NM_002359.2)		
	rs74006139	Promotor	127519	-980	GACTCGGGCTGTG T>C CCAGAGCAGGGGG	0.009
-806 ^a	N.V.	Promotor	127345	-806	GGCTTCGGGACTTT >C GGGGCCGGCCTGT	0.004
-458 ^a	N.V.	Promotor	126997	-458	CGGCCCGCGCC C>G TCCGGGACCCCG	0.004
-374 ^a	N.V.	Promotor	126913	-374	CAGGGCCCAGGCC C>A GCACTCGAGGGGG	0.004
-8 ^a	N.V.	Promotor	126547	-8	GTGACGTCTGTGG A>G CGCGCGCGCGCGCG	0.004
13 ^a	N.V.	5'-UTR	126527	13	GCGGCGCGCGCG A>C GGTCCCAGCCTGGT	0.004
4436-4444 ^a	N.V.	Intrón 1	122104 - 122095	4436-4444	CTGGGCCGGGG G>G GGGGGTGCGGCGA	0.009
5247 ^a	N.V.	3'-UTR	121292	5247	CAAAGCACAACT G>A TACCCAGGGCG	0.004
5296 ^a	N.V.	3'-UTR	121243	5296	CGTCATTGTCGT C>A TGTGTTTTGTATG	0.128

^a Nuevas variantes detectadas en este estudio; N.V. = Nueva variante

^b El gen *MAFG* tiene una orientación inversa. Aquí se presenta en orientación 5' → 3'.

En el gen *MAFK* se descubrieron 16 variantes, seis de ellas se localizaron en la zona reguladora del gen, nueve en los intrones, y una en la región 3'-UTR (**Tabla 6-3**). Todos los cambios encontrados estuvieron en equilibrio de Hardy-Weinberg ($p \geq 0.05$) y seis de ellas no habían sido reportadas previamente.

Tabla 6-3. Variantes encontradas en el gen *MAFK*.

SNP ID		Posición			Cambio de nucleótido	Frecuencia del alelo menor
Este estudio	dbSNP (NCBI)	Ubicación	NT_007819.17 Gen ID: 7975	Desde el TSS (NM_002360.3)		
	rs3829898	Promotor	1571619	-1157	AGAACCCGCAAGT G>A GGGGGGAGAGATG	0.454
	rs3814477	Promotor	1571826	-950	CTTCCCAGAACCC C>T GGGCAGATGCCAT	0.454
	rs3814478	Promotor	1571857	-919	GTCCTGGGGGCC C>T TCGGTTGCCAGG	0.454
	rs3814479	Promotor	1571928	-848	GGGACCCAGAAG T>C TGGCAAGGGGCT	0.454
	rs10254597	Promotor	1572501	-275	GTTTTTATGTT A>G TAGAGACAGAGAA	0.470
	rs11763285	Promotor	1572528	-248	TGGGGGAAGGGAG G>C TTTCGCTATGTTG	0.470
8380 ^a	N.V.	Intrón 1	1581155	8380	CGCTGCCGGCC TG>T ACCTGTGCTGGCT	0.013
8523 ^a	N.V.	Intrón 2	1581298	8523	CCAAGCAGGCC CG>A TCTCAAGCACAGG	0.013
8553 ^a	N.V.	Intrón 2	1581328	8553	GGGAAAGCCCC CG>A GCCCCACAGCCCC	0.017
8753 ^a	N.V.	Intrón 2	1581528	8753	TGGGCTGTGT CC>T GCCGGGCGCTACT	0.008
	rs34350497	Intrón 2	1581583:1581592	8809	GGCAGCGTGGCC: /AGCGTGGCC CTTCTGTGAACCGG	0.092
	rs892522	Intrón 2	1581681	8907	GTGCCTGTGGCC G>A GCTCAGTCCCAGC	0.458
	rs73042781	Intrón 2	1581756	8982	AACCAGGAGTGAC C>T GATGTCACCACTG	0.212
	rs3824071	Intrón 2	1581779	9004	CACTGGCTTTTCC C>T GTTCAGAAAGGGT	0.458
9030 ^a	N.V.	Intrón 2	1581805	9030	TCCTGTGGGGCCAC CA>A AGCTGTCACATGG	0.029
	rs3735655	Intrón 2	1581888	9113	TGAGAGCGCAGGC CG>A TGCAGGGGCACCG	0.202
9880 ^a	N.V.	3'-UTR	1582655	9880	GCCACCTGCTCC CC>T GCAGGATGTGTCT	0.004

Nuevas variantes detectadas en este estudio. N.V. = Nueva variante.

En la **Tabla 6-4**, hasta la **Tabla 6-7**, se comparan las frecuencias alélicas para algunos SNPs cuyos datos están disponibles en el NCBI para otras poblaciones. Debe señalarse que estos datos no provienen de estudios de resecuenciación, más bien se trata de SNPs comunes (FAM > 5%) identificados mediante microarreglos. Claramente se advierte diferencias que podrían ser características de nuestra población, como la presencia de alelos más frecuentes, SNPs raros o no detectados, e inclusive la presencia de nuevas variantes no detectadas en otras poblaciones.

Tabla 6-4. Comparación de las frecuencias alélicas poblacionales para el gen MAFF.

SNP, región	Población	Cambio, FAM	Figura																														
rs2413508, promotor	Pilot.1.CEU	G/C, 0.292	<p>MAFF rs2413508 (G/C)</p> <table border="1"> <thead> <tr> <th>Población</th> <th>C</th> <th>G</th> </tr> </thead> <tbody> <tr> <td>pilot.1.CEU</td> <td>0,292</td> <td>0,708</td> </tr> <tr> <td>pilot.1.YRI</td> <td>0,18</td> <td>0,82</td> </tr> <tr> <td>HapMap_CEU</td> <td>0,275</td> <td>0,725</td> </tr> <tr> <td>HapMap_HCB</td> <td>0,533</td> <td>0,467</td> </tr> <tr> <td>HapMap_JPT</td> <td>0,674</td> <td>0,326</td> </tr> <tr> <td>México</td> <td>0,482</td> <td>0,518</td> </tr> </tbody> </table>	Población	C	G	pilot.1.CEU	0,292	0,708	pilot.1.YRI	0,18	0,82	HapMap_CEU	0,275	0,725	HapMap_HCB	0,533	0,467	HapMap_JPT	0,674	0,326	México	0,482	0,518									
	Población	C		G																													
	pilot.1.CEU	0,292		0,708																													
	pilot.1.YRI	0,18		0,82																													
	HapMap_CEU	0,275		0,725																													
	HapMap_HCB	0,533		0,467																													
	HapMap_JPT	0,674		0,326																													
México	0,482	0,518																															
México	G/C, 0.482																																
rs4608623, promotor	Pilot.1.CEU	T/G, 0.458	<p>MAFF rs4608623 (T/G)</p> <table border="1"> <thead> <tr> <th>Población</th> <th>G</th> <th>T</th> </tr> </thead> <tbody> <tr> <td>pilot.1.CEU</td> <td>0,458</td> <td>0,542</td> </tr> <tr> <td>pilot.1.YRI</td> <td>0,52</td> <td>0,48</td> </tr> <tr> <td>HapMap_CEU</td> <td>0,483</td> <td>0,517</td> </tr> <tr> <td>HapMap_HCB</td> <td>0,467</td> <td>0,533</td> </tr> <tr> <td>HapMap_JPT</td> <td>0,443</td> <td>0,557</td> </tr> <tr> <td>HapMap_YRI</td> <td>0,661</td> <td>0,339</td> </tr> <tr> <td>México</td> <td>0,433</td> <td>0,567</td> </tr> </tbody> </table>	Población	G	T	pilot.1.CEU	0,458	0,542	pilot.1.YRI	0,52	0,48	HapMap_CEU	0,483	0,517	HapMap_HCB	0,467	0,533	HapMap_JPT	0,443	0,557	HapMap_YRI	0,661	0,339	México	0,433	0,567						
	Población	G		T																													
	pilot.1.CEU	0,458		0,542																													
	pilot.1.YRI	0,52		0,48																													
	HapMap_CEU	0,483		0,517																													
	HapMap_HCB	0,467		0,533																													
	HapMap_JPT	0,443		0,557																													
HapMap_YRI	0,661	0,339																															
México	0,433	0,567																															
México	T/G, 0.433																																
rs2267372, intrón	JBIC-allele	G/A, 0.259																															
	México	G/A, 0.429																															
SNP, región	Población	Cambio, FAM	Figura																														
rs2235264, intrón	Pilot.1.CEU	A/G, 0.500	<p>MAFF rs2235264 (A/G)</p> <table border="1"> <thead> <tr> <th>Población</th> <th>A</th> <th>G</th> </tr> </thead> <tbody> <tr> <td>Pilot.1.CEU</td> <td>0,5</td> <td>0,5</td> </tr> <tr> <td>Pilot.1.YRI</td> <td>0,18</td> <td>0,82</td> </tr> <tr> <td>JBIC-allele</td> <td>0,562</td> <td>0,438</td> </tr> <tr> <td>HapMap_HCB</td> <td>0,216</td> <td>0,784</td> </tr> <tr> <td>HapMap_JPT</td> <td>0,386</td> <td>0,614</td> </tr> <tr> <td>HapMap_YRI</td> <td>0,042</td> <td>0,958</td> </tr> <tr> <td>AoD Japanese KHP1</td> <td>0,56</td> <td>0,44</td> </tr> <tr> <td>México</td> <td>0,514</td> <td>0,486</td> </tr> <tr> <td>México</td> <td>0,558</td> <td>0,442</td> </tr> </tbody> </table>	Población	A	G	Pilot.1.CEU	0,5	0,5	Pilot.1.YRI	0,18	0,82	JBIC-allele	0,562	0,438	HapMap_HCB	0,216	0,784	HapMap_JPT	0,386	0,614	HapMap_YRI	0,042	0,958	AoD Japanese KHP1	0,56	0,44	México	0,514	0,486	México	0,558	0,442
	Población	A		G																													
	Pilot.1.CEU	0,5		0,5																													
	Pilot.1.YRI	0,18		0,82																													
	JBIC-allele	0,562		0,438																													
	HapMap_HCB	0,216		0,784																													
	HapMap_JPT	0,386		0,614																													
	HapMap_YRI	0,042		0,958																													
	AoD Japanese KHP1	0,56		0,44																													
México	0,514	0,486																															
México	0,558	0,442																															
México	A/G, 0.442																																

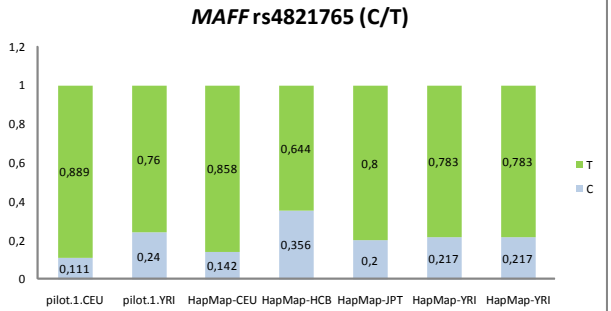
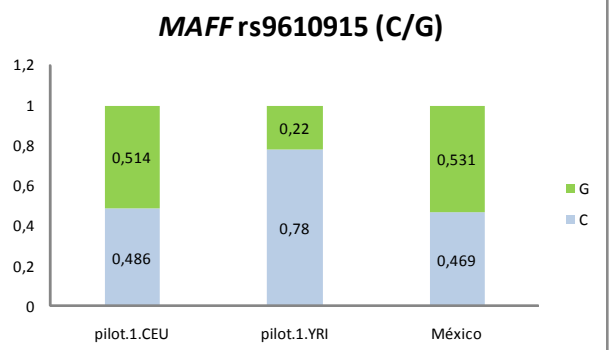
<p>rs4821765, intrón</p>	<p>Pilot.1.CEU T/C, 0.111 Pilot.1.YRI T/C, 0.240 HapMap-CEU T/C, 0.142 HapMap-HCB T/C, 0.356 HapMap-JPT T/C, 0.200 HapMap-YRI T/C, 0.217</p>	<p>México No identificado</p>	
<p>rs9610915, 3'-UTR</p>	<p>Pilot.1.CEU G/C, 0.486 Pilot.1.YRI C/G, 0.220</p>	<p>México G/C, 0.469</p>	
<p>rs5995550, 3'-UTR</p>	<p>México</p>	<p>A/C, 0.393</p>	

Tabla 6-5. Frecuencias alélicas y genotípicas de la nueva variante de *MAFG*.

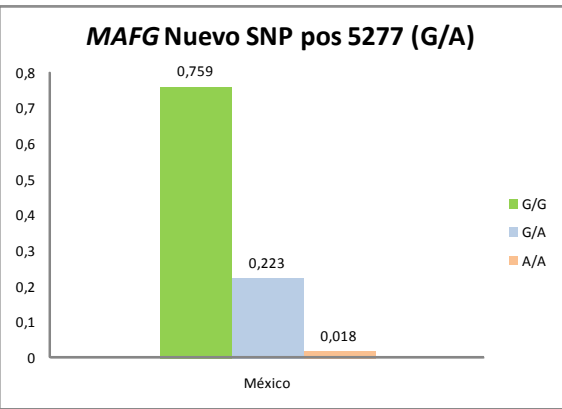
SNP, región	Población	Cambio, FAM	Figura
<p>Nueva variante, 3'-UTR</p>	<p>México</p>	<p>G/A, 0.129</p>	

Tabla 6-6. Comparación de las frecuencias alélicas poblacionales para el gen *MAFK*.

SNP, región	Población	Cambio, FAM	Figura
rs3829898, promotor	pilot.1.CHB+JPT pilot.1.YRI México	A/G, 0.261 A/G, 0.320 G/A, 0.454	<p>MAFK rs3829898 (G/A)</p>
rs3814477, promotor	HapMap-CEU México	T/C, 0.448 C/T, 0.454	<p>MAFK rs3814477 (T/C)</p>
rs10254597, promotor	HapMap-CEU HapMap-HCB HapMap-JPT HapMap-YRI México	G/G, 1.00 G/G, 1.00 G/G, 1.00 G/G, 1.00 A/G, 0.47 Alta frecuencia	<p>MAFK rs10254597 A/G</p>
SNP, región	Población	Cambio, FAM	Figura
rs892522, intrón	JBIC-allele AoD African American AoD Caucasian México	A/G, 0.335 G/A, 0.400 A/G, 0.440 G/A, 0.458	<p>MAFK rs892522 A/G</p>

<p>rs73042781, intrón</p>	<p>Pilot.1.YRI C/T, 0.08 México C/T, 0.212</p>	<p>MAFK rs73042781 (C/T)</p> <table border="1"> <caption>MAFK rs73042781 (C/T) Allele Frequencies</caption> <thead> <tr> <th>Group</th> <th>C</th> <th>T</th> </tr> </thead> <tbody> <tr> <td>pilot.1.YRI</td> <td>0.92</td> <td>0.08</td> </tr> <tr> <td>México</td> <td>0.788</td> <td>0.212</td> </tr> </tbody> </table>	Group	C	T	pilot.1.YRI	0.92	0.08	México	0.788	0.212			
Group	C	T												
pilot.1.YRI	0.92	0.08												
México	0.788	0.212												
<p>rs3824071</p>	<p>Pilot.1.CHB+JPT T/C, 0.250 JBIC-allele T/C, 0.340 México C/T, 0.458</p>	<p>MAFK rs3824071 (C/T)</p> <table border="1"> <caption>MAFK rs3824071 (C/T) Allele Frequencies</caption> <thead> <tr> <th>Group</th> <th>C</th> <th>T</th> </tr> </thead> <tbody> <tr> <td>pilot.1.CHB+JPT</td> <td>0.25</td> <td>0.75</td> </tr> <tr> <td>JBIC-allele</td> <td>0.34</td> <td>0.66</td> </tr> <tr> <td>México</td> <td>0.542</td> <td>0.458</td> </tr> </tbody> </table>	Group	C	T	pilot.1.CHB+JPT	0.25	0.75	JBIC-allele	0.34	0.66	México	0.542	0.458
Group	C	T												
pilot.1.CHB+JPT	0.25	0.75												
JBIC-allele	0.34	0.66												
México	0.542	0.458												

Tabla 6-7. Reporte de SNPs encontrados en los genes *MAF*.

MAFF						MAF
Número	Región	Pos	SNP	Validación		%
1	Promotor	-765	rs2413508 (G/C)			48,2
2	Promotor	-742	rs12165335 (G/A)			0
3	Promotor	-713	rs11912025 (G/A)	H		0
4	Promotor	-696	rs6001050 (C/T)			0
5	Promotor	-627	rs3041785 (-/ATC/ATCTATC)			0
6	Promotor	-627	rs79226988 (-/ATCTATC)			0
7	Promotor	-626	rs34308084 (-/ATCTATC)			43,3
8	Promotor	-626	rs76701240 (A/C)			0
9	Promotor	-625	rs74730019 (G/T)			0
10	Promotor	-562	rs4374456 (C/G)			43,3
11	Promotor	-562	rs71324897 (CT/GG)			0
12	Promotor	-561	rs4608623 (T/G)			43,3
13	Promotor	-550	Variante inédita (T/C)			0,4
14	Promotor	-514	Variante inédita (G/A)			0,4
15	Promotor	-437	rs5756971 (T/C)			0
16	Promotor	-142	rs62228645 (A/C)			0
17	Promotor	-15	Variante inédita (T/C)			1,3
18	5'-UTR	37	Variante inédita (G/T)			0,4
19	5'-UTR	103	rs11703601 (C/T)	H		39,3
20	5'-UTR	156	rs909629 (C/G)	H		0,4
21	5'-UTR	274	Variante inédita (A/T)			0,4
22	Intrón 1	296	rs2267372 (G/A)			42,9
23	Intrón 1	328	rs62228646 (A/G)			0
24	Intrón 2	12012	rs2235264 (A/G)			44,2
25	Intrón2	12319	rs4821765 (C/T)			0
26	Intrón2	12404	rs12166436 (C/T)			0
27	Intrón2	12413	rs75322268 (G/T)			0
28	3'-UTR	13095	rs73425641 (C/G)	H		0
29	3'-UTR	13107	rs76165778 (C/A)			0,9
30	3'-UTR	13142	rs9610915 (G/C)			46,9
31	3'-UTR	13360	rs11544353 (A/C)	H		0,4
32	3'-UTR	13773	Variante inédita (A/G)			0,4
33	3'-UTR	14064	rs75078936 (A/G)			0
34	3'-UTR	14091	rs1063237 (T/A)			0
35	3'-UTR	14273	rs4821766 (G/A)	H		4,1
36	3'-UTR	14434	rs5995550 (A/C)	H		39,3
37	3'-UTR	14569	rs3177639 (A/C)			0
38	3'-UTR	14574	rs3188073 (A/G)			0
39	3' cerca del gen	14581	rs36116754 (-/T)			0
40	3' cerca del gen	14619	rs67838059 (-/AAA)			0
41	3' cerca del gen	14622	rs60588032 (-/AAA)			0
42	3' cerca del gen	14623	rs71702045 (-/AAA)			0

Simbología:

Celdas verdes = SNPs encontrados en este estudio.

Celdas rojas = SNPs reportados en el NCBI, pero no identificados.

Celdas azules = Nuevas variantes encontradas.

La posición indicada es con respecto al sitio de inicio de la transcripción.

MAFG						MAF
Número	Región	Pos	SNP	Validación		%
1	Promotor	-980	rs74006139 (A/G)			0,9
2	Promotor	-806	Variante inédita (A/G)			0,4
3	Promotor	-482	rs4539653 (C/A)			0
4	Promotor	-458	Variante inédita (G/C)			0,4
5	Promotor	-374	Variante inédita (C/T)			0,4
6	Promotor	-185	rs35656863 (-/G)			0
7	Promotor	-8	Variante inédita (-/CGC)			0,4
8	5'-UTR	13	Variante inédita (-/GCC)			0,4
9	Intrón 1	4436-4444	Variante inédita (-/9C)			0,9
10	Exón 3 (sinónimo)	4822	rs1054499 (C/T)			0
11	Exón 3 (sinónimo)	5023	rs61752690 (C/T)			0
12	3'-UTR	5121	rs1054500 (C/A)			0
13	3'-UTR	5247	Variante inédita (C/T)			0,4
14	3'-UTR	5296	Variante inédita (C/T)			12,8
15	3'-UTR	5727	rs1879570 (C/A)			0
16	3'-UTR	6014	rs3087708 (C/T)			0

MAFK						% FAM
Número	Zona	Pos	SNP	Validación		%
1	Promotor	-1157	rs3829898 (G/A)			46,4
2	Promotor	-1155	rs57442949 (G/C)			0
3	Promotor	-950	rs3814477 (C/T)			46,4
4	Promotor	-919	rs3814478 (C/T)			46,4
5	Promotor	-874	rs73283896 (G/A)			0
6	Promotor	-848	rs3814479 (C/T)			46,4
7	Promotor	-640	rs12112155 (G/A)	H		0
8	Promotor	-275	rs10254597 (A/G)		H	47,3
9	Promotor	-248	rs11763285 (G/C)		H	47,3
10	Intrón 1	8354	rs10252127 (G/C)			0
11	Intrón 1	8380	Variante inédita (G/T)			1,3
12	Intrón 2	8523	Variante inédita (G/A)			1,3
13	Intrón 2	8553	Variante inédita (G/A)			1,8
14	Intrón 2	8734	rs35335648 (-/C)			0
15	Intrón 2	8753	Variante inédita (C/T)			0,9
16	Intrón 2	8798	rs71992214 (-/GCGTGGCCA)			0
17	Intrón 2	8799	rs3217410 (-/AGCGTGGCC)			0
18	Intrón 2	8806	rs72230386 (-/CCAGCGTGG)			0
19	Intrón 2	8808	rs34350497 (-/AGCGTGGCC)			8
20	Intrón 2	8907	rs892522 (G/A)			47,3
21	Intrón 2	8982	rs73042781 (C/T)			20,5
22	Intrón 2	9004	rs3824071 (C/T)			46,9
23	Intrón 2	9030	Variante inédita (C/A)			3,1
24	Intrón 2	9113	rs3735655 (G/A)			20,1
25	Exón 3 (No s	9537	rs1055475 (G/T)			0
26	Exón 3 (caml	9604	rs35857715 (-/A)			0
27	3'-UTR	9880	Variante inédita (C/T)			0,5
28	3'-UTR	10177	rs10714927 (-/T)			0
29	3'-UTR	10178	rs11313860 (-/T)			0
30	3'-UTR	10313	rs12671303 (G/T)			0

6.3 Determinación de haplotipos en los genes *MAF*.

Todos los polimorfismos cuya frecuencia del alelo menor (FAM) fue mayor a 0.03 se analizaron para el desequilibrio de ligamiento (LD), calculando los valores de D' y r^2 , así como la presencia de haplotipos utilizando los programas *SNPAlyze v7.0* (Dynacom, Japón) y *Haploview v4.2*. Debido al escaso número de variantes encontradas y la baja frecuencia de las mismas, el gen *MAFG* fue descartado de este análisis.

Para *MAFF* se consideraron 9 variantes con una FAM > 0.3, de las 36 combinaciones posibles, 24 de ellas mostraron un valor de $D' \geq 0.9$. En lo que atañe a los valores de r^2 (**Figura 6-4**) se observó un LD completo ($r^2 = 1$) entre los SNPs: rs34308084 (-/ATCTATC), rs4374456 (G/C), y rs4608623 (G/T). Estos tres SNPs situados en la región reguladora también conservan un fuerte ligamiento ($r^2 = 0.91$) con el cambio rs226732 (G/A) situado en el intrón 1.

Considerando las 10 variaciones más frecuentes en el gen *MAFK*, 25 de 45 combinaciones presentaron valores de $D' \geq 0.90$. Con respecto a los valores de r^2 (**Figura 6-5**) se observó un ligamiento completo ($r^2 = 1$) entre los SNPs rs3829898 (G/A), rs3814477 (C/T), rs3814478 (C/T) y rs3814479 (T/C). También se constató una $r^2 = 1$ entre los SNPs rs10254597 (A/G) y rs11763285 (G/C). Todos los ligamientos anteriores involucran a SNPs detectados en la región promotora. Por otro lado se advirtió un fuerte ligamiento ($r^2 = 0.946$) entre los pares de SNPs rs892522 (G/A), rs3824071(C/T) y rs73042781 (C/T), rs3735655 (G/A), estas últimas variantes se identificaron en el intrón 2.

	index	1	2	3	4	5	6	7	8	9
index		rs2413508	rs34308084	rs4374456	rs4608623	rs11703601	rs2267372	rs2235264	rs9610915	rs5995550
1	rs2413508									
2	rs34308084	0.6777								
3	rs4374456	0.6777	1							
4	rs4608623	0.6777	1	1						
5	rs11703601	0.4838	0.6558	0.6558	0.6558					
6	rs2267372	0.6649	0.9101	0.9101	0.9101	0.704				
7	rs2235264	0.6043	0.7882	0.7882	0.7882	0.4959	0.7711			
8	rs9610915	0.6532	0.7619	0.7619	0.7619	0.4818	0.7463	0.7246		
9	rs5995550	0.5375	0.7425	0.7425	0.7425	0.5947	0.7229	0.5474	0.5659	
index	index	1	2	3	4	5	6	7	8	9
1	1	rs2413508	rs34308084	rs4374456	rs4608623	rs11703601	rs2267372	rs2235264	rs9610915	rs5995550

Figura 6-4. Matriz de LD para *MAFF* según el parámetro r^2 . Las casillas en la parte superior e izquierda contienen SNPs analizados para su desequilibrio de ligamiento. Los distintos números indican el valor de correlación r^2 para cada par de SNPs. Un valor de $r^2 = 1$ significa un ligamiento perfecto, es decir, ambos SNPs siempre se heredan juntos. Un valor de $r^2 > 0.8$ señala un fuerte ligamiento. En la parte inferior se muestran los SNPs que están fuertemente ligados entre sí, uno de ellos captura el estado alélico de los otros.

	index	1	2	3	4	5	6	7	8	9	10
index		rs3829898	rs3814477	rs3814478	rs3814479	rs10254597	rs11763285	rs892522	rs73042781	rs3824071	rs3735655
1	rs3829898										
2	rs3814477	1									
3	rs3814478	1	1								
4	rs3814479	1	1	1							
5	rs10254597	0.5367	0.5367	0.5367	0.5367						
6	rs11763285	0.5367	0.5367	0.5367	0.5367	1					
7	rs892522	0.8601	0.8601	0.8601	0.8601	0.5073	0.5073				
8	rs73042781	0.1983	0.1983	0.1983	0.1983	0.1143	0.1143	0.2321			
9	rs3824071	0.8425	0.8425	0.8425	0.8425	0.4921	0.4921	0.9468	0.228		
10	rs3735655	0.1695	0.1695	0.1695	0.1695	0.0939	0.0939	0.2004	0.9188	0.1965	
index	index	1	2	3	4	5	6	7	8	9	10
1	1	rs3829898	rs3814477	rs3814478	rs3814479	rs10254597	rs11763285	rs892522	rs73042781	rs3824071	rs3735655
2	2										
3	3										
4	4										

Figura 6-5. Matriz de LD para *MAFK* según el parámetro r^2 . Las celdas en amarillo indican los SNPs fuertemente correlacionados ($r^2 \geq 0.9$), varios pares de SNPs están en desequilibrio de ligamiento perfecto ($r^2 = 1$). Las celdas azules señalan a los SNPs correlacionados entre $0.5 \leq r^2 < 0.9$. Por último, las celdas en blanco indican una escasa correlación $r^2 < 0.5$. En la parte inferior se observa, en distintos niveles y en amarillo, a los SNPs fuertemente ligados entre sí. Un SNP captura el estado alélico de los demás en su mismo nivel.

Debido a que se observaron ligamientos relativamente fuertes en *MAFF* y *MAFK* para las regiones analizadas, se llevó a cabo un análisis de haplotipos considerando un solo bloque de LD para cada gen. En el presente estudio, los haplotipos detectados con una frecuencia mayor a 0.03 se denominaron como haplotipos “comunes”. Mediante un análisis *in silico*, utilizando los programas *SNPAlyze v7.0* y *Haploview 4.2*, se estimó la configuración de haplotipos para *MAFF* con una probabilidad > 95% (**Tabla 6-4**). El análisis de genotipos permitió la identificación directa de 4 haplotipos comunes con variaciones homocigas en todos los loci (A, B, C y D) y 5 con una variación heterociga en un solo sitio (E, G, H, I y J). Los haplotipos comunes fueron el A (f= 0.422), B (f= 0.31), C (f= 0.04) y D (f= 0.032). Estos 4 haplotipos comunes explican el 81 % de toda la variabilidad observada.

Para *MAFK*, los haplotipos determinados e inferidos se muestran en la **Tabla 6-5**. Tres haplotipos se reconocieron sin ambigüedad debido a la presencia de variantes homocigas en todos los sitios (haplotipos I, II y III) o una variación heterociga en un solo sitio (IV y V). El haplotipo común predominante fue el I (f = 0.384), seguido por el II (f= 0,272), III (f= 0.164), IV (f= 0.052) y V (f= 0.041). Estos cinco haplotipos comunes constituyen el 91.3 % de toda la variabilidad observada.

Tabla 6-8. Haplotipos del gen MAFF.

Cambio de nucleótido ^a	rs2413508 ^d -824 G>C	rs34308084 ^d -685 -ATCTATC	rs4374456 ^d -621 C>G	rs4608623 ^d -620 I>G	rs11703601 ^d 44 C>T	rs2267372 ^d 237 G>A	rs2235264 ^d 11953 A>G	rs9610915 ^d 13083 G>C	rs59955 50 ^d 14375 A>C	Número de cromosomas	Frecuencia	
	REGIÓN REGULATORIA				5'-UTR	INTRÓN 1	INTRÓN 2	3'-UTR				
Haplotipos ^{b,c}	A	C	No inserción	C	T	C	G	A	G	A	95	0.423
	B	G	ATCTATC	G	G	T	A	G	C	C	69	0.310
	C	G	No inserción	C	T	C	G	A	G	A	10	0.047
	D	G	ATCTATC	G	G	C	A	G	C	C	7	0.032
	E	G	ATCTATC	G	G	T	A	A	C	C	5	0.023
	F	G	ATCTATC	G	G	C	A	G	C	A	4	0.018
	G	G	ATCTATC	G	G	T	A	G	C	A	4	0.017
	H	C	No inserción	C	T	T	G	A	G	A	3	0.016
	I	G	No inserción	C	T	C	G	A	C	A	3	0.015
	J	C	No inserción	C	T	C	G	A	C	A	3	0.013
Otros ^e										19	0.086	
										Total	224	1.000

^a Las posiciones están numeradas tal como se indica en la **Tabla 6-1**.

^b Alelo mayor en verde; alelo menor en azul.

^c Los haplotipos se asignan como letras del alfabeto.

^d Número asignado en el dbSNP en el National Center for Biotechnology Information.

^e Los haplotipos ambiguos que se infirieron en un solo individuo se agrupan como "otros". La predicción de haplotipos raros que se infirieron en solo un sujeto (f= 0.004) en ocasiones es inexacta.

Tabla 6-9. Haplotipos del gen MAFK.

Cambio de nucleótido ^a	rs3829898 -1157 G>A	rs3814477 -950 C>T	rs3814478 -919 C>T	rs3814479 -848 C>T	rs10254597 -275 A>G	rs11763285 -248 G>C	rs892522 8907 G>A	rs73042781 8982 C>T	rs3824071 9004 C>T	rs3735655 9113 G>A	Número de cromosomas	Frecuencia	
	REGION REGULATORIA				INTRÓN 2								
Haplotipos ^{b,c}	I	A	T	T	T	G	C	A	C	T	G	84	0.376
	II	G	C	C	C	A	G	G	C	C	G	63	0.281
	III	G	C	C	C	A	G	G	T	C	A	37	0.164
	IV	A	T	T	T	A	G	A	C	T	G	12	0.052
	V	G	C	C	C	G	C	G	C	C	G	9	0.040
	VI	G	C	C	C	G	C	G	T	C	A	6	0.028
	VII	G	C	C	C	A	G	A	C	T	G	5	0.021
	Otros ^d											8	0.038
										Total	224	1.000	

^a Las posiciones están numeradas tal como se indica en la **Tabla 6-3**.

^b Alelo mayor en verde; alelo menor en azul.

^c Los distintos haplotipos se designan en números romanos.

^d Los haplotipos ambiguos que se infirieron en un solo individuo se agrupan como "otros", no se muestran las variaciones en contradas en estos haplotipos.

Ninguna de las variantes identificadas en la región 3'-UTR de los genes *MAF* afectó la unión de miRNAs conocidos, según la búsqueda realizada en la base miRBase (<http://www.mirbase.org/search.shtml>).

6.4 Análisis *in silico*.

Dado que el mayor porcentaje de cambios encontrados en los genes *MAF* se descubrieron en regiones regulatorias (promotor y UTRs), *MAFF*: 89.4%, *MAFG*: 88.9% y *MAFK*: 40%; se decidió utilizar una herramienta bioinformática, rVISTA (<http://rvista.decode.org>), para indagar si los polimorfismos detectados en el promotor podrían alterar la actividad transcripcional del gen. Una de las principales características de rVISTA estriba en combinar la búsqueda de bases de datos como TRANSFAC con el análisis comparativo en sitios conservados, esto reduce el número de falsos positivos en ~ 95%, al mismo tiempo que mantiene una alta sensibilidad durante la búsqueda⁽¹⁰³⁾.

En primer lugar, se comparó la región promotora de los genes *MAF* humanos con su contraparte en el genoma del ratón (*Mus musculus*). La homología en la secuencia de nucleótidos de los genes *MAFF*, *MAFG* y *MAFK* humanos con sus contrapartes del ratón es del 91.24%, 90.55% y 89.74%, respectivamente (www.genecards.org). Dado que el genoma del ratón es uno de los más ampliamente estudiados, se decidió utilizarlo en el análisis de alineamiento y búsqueda de sitios conservados de unión para factores de transcripción (TFs) mediante rVISTA, en una región total de 1692 pb. De acuerdo con el reporte suministrado por rVISTA, se encontró que algunos de los polimorfismos presentes en la región promotora de los genes *MAF* podrían ser capaces de alterar la unión de algunos factores de transcripción.

Este análisis sugirió que el SNP rs11703601 (C/T) de *MAFF* está situado en el sitio de unión de los factores de transcripción PAX9 y NR2F2/COUP-TFII. En *MAFG* el SNP rs74006139 (T/C) reside en el sitio de unión del factor MTF1; mientras que las nuevas variantes situadas en las posiciones -806 (T/C), -458 (C/G) y +13 (-/CGG) del TSS alteran el sitio de unión del factor NUDR/DEAF-1. Además, el cambio situado en la posición -458 (C/G) del TSS podría afectar la unión de MINI19 y MINI20. En *MAFK*, los SNPs rs3814477 (C/T) y rs3814479 (T/C) tal vez afecten la unión del factor REST/NRSF; en tanto que los sitios de unión para ZNF153/STAF y ZNF423/ROAZ están alterados por los SNPs rs3814477 (C/T) y rs3814478 (C/T), respectivamente. Finalmente, el SNP rs10254597 (A/G) podría perturbar la unión del factor MEF2 y el rs11763285 (G/C) la del factor ZEC. En la **Figura 6-6** se ilustra los resultados de este análisis.

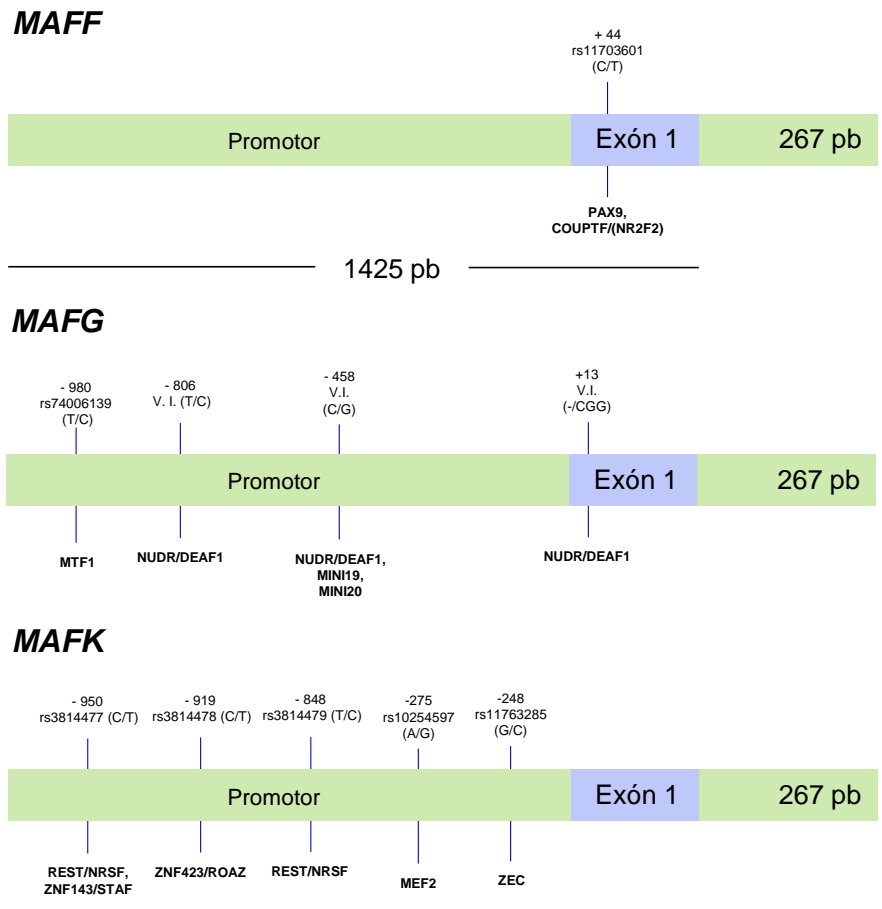


Figura 6-6. Análisis *in silico* para sitios de unión de factores de transcripción. En la parte superior se muestran los polimorfismos con su posición relativa al TSS. Abajo se muestran los TFs con sitios de unión alterados.

7 DISCUSIÓN.

Las enfermedades complejas surgen por la interacción entre los genes y el ambiente, es decir, la exposición de individuos genéticamente susceptibles a condiciones ambientales adversas favorece el desarrollo de enfermedades. Sin embargo, a diferencia de los padecimientos de tipo mendeliano, en las que la alteración de un solo gen propicia la patología, en las enfermedades complejas participan numerosos genes⁽¹⁰⁴⁾, y el efecto individual de cada uno de ellos es relativamente pequeño.

Los estudios de asociación de genoma completo (GWAS) han permitido la identificación de numerosos genes asociados a diversos padecimientos. Estos logros en buena medida se han obtenido porque en un estudio GWAS se analizan desde cientos de miles hasta millones de SNPs⁽¹⁰⁵⁾. Sin embargo, estos análisis principalmente se han orientado a poblaciones con descendencia europea y asiática, las cuales representan un subconjunto de la variación genética humana y es probable que este conocimiento no sea totalmente transferible a otras poblaciones⁽¹⁰⁶⁾.

La composición genética de la población mexicana es compleja, debido al alto mestizaje presente entre el componente amerindio, el europeo, y en menor proporción, el africano^(107,108). En general, el perfil de variantes génicas en la población mexicana está pobremente representado y se desconocen los patrones de desequilibrio de ligamiento específicos de la población. Dado lo anterior es necesario efectuar estudios de secuenciación en genes candidato o funcionalmente relevantes, que permitan identificar marcadores de susceptibilidad propios de nuestra población.

Los genes *MAF* son reguladores de la transcripción esenciales para el desarrollo embrionario y la respuesta hacia diversos tipos de estrés. Sin embargo,

a pesar de su importancia, se conoce poco sobre los polimorfismos que ocurren en sus secuencias. La información depositada en el *National Center for Biotechnology Information* (NCBI) para estos genes contiene únicamente los datos sobre las poblaciones estudiadas en el HapMap¹⁹ y para un gran número de polimorfismos falta validar y determinar su frecuencia alélica.

En nuestro estudio encontramos que de acuerdo al número de variantes encontradas, el orden de variabilidad de los genes *MAF* es *MAFF* (19)>*MAFK* (17)>*MAFG* (9). Este orden concuerda con el número de polimorfismos registrado en el NCBI para estas regiones analizadas: *MAFF*(36)>*MAFK*(24)>*MAFG*(8). Los niveles de variabilidad encontrados sugieren que *MAFG* podría ser el gen de mayor relevancia funcional de los tres, en tanto que, por su alta variabilidad *MAFF* podría ser el menos esencial de ellos. Esto es, *MAFG* es un gen participante en funciones de gran relevancia fisiológica (respuesta al estrés, megacariopoyesis, funcionamiento neuronal), esto ha impuesto restricciones para la aparición de SNPs que potencialmente pudieran alterar su actividad. En contraparte, *MAFF* es un gen con poca relevancia para la supervivencia (el ratón *knockout* es viable) y por ello puede acumular más cambios en su secuencia. El análisis de ratones *knockout* parece concordar con esta hipótesis, ya que la delección germinal de *MAFG* causa la aparición de fenotipos de moderada gravedad⁽⁶³⁾, en tanto que la supresión de *MAFF* y *MAFK*, e incluso con la doble delección *MAFG*^{-/-}:*MAFF*^{-/-} no presentan trastornos evidentes. Sin embargo, los ratones *MAFG*^{-/-}:*MAFK*^{-/-} son afectados por diversas anormalidades hematológicas y conductuales que culminan en la muerte peri o post-natal⁽⁷¹⁾.

¹⁹ 270 individuos provenientes de 4 poblaciones: YRI (Yoruba en Ibadan, Nigeria), CHB (chinos Han en Beijing, China), JPT (japoneses de Tokio, Japón) y CEU (residentes de Utah con ancestría del norte y occidente de Europa).

Interesantemente, la mayoría de las variantes identificadas en los genes *MAF* se ubicaron dentro de regiones con una alta plasticidad como la promotora y regiones UTRs, pero no se identificaron en las regiones codificantes, estos hallazgos son un reflejo de la importancia biológica de estas proteínas. Por ejemplo, los polimorfismos no sinónimos en la región codificante del gen podrían afectar negativamente la función de la proteína.

Los datos obtenidos en este trabajo señalan claramente grandes diferencias, tanto en el perfil como en las frecuencias de los polimorfismos en la población mexicana con respecto a otros grupos humanos (ver **Tabla 6-4** hasta **Tabla 6-7**). En principio, numerosos polimorfismos presentes en otras poblaciones no fueron identificados en este estudio. Cabe señalar que algunos de ellos, como el rs76701240 (A/C) de *MAFF*, no están validados a nivel poblacional y algunos otros (como el rs73283896 (G/A) de *MAFK*, frecuente en la población Yoruba) son exclusivos de ciertos grupos étnicos. Sin embargo, SNPs como el rs4821765 (T/C) de *MAFF*, que no fue identificado en nuestra muestra, presenta una alta frecuencia en otras poblaciones. Además, varios de los polimorfismos identificados en este estudio no han sido reportados en otros grupos poblacionales. Aunque la baja frecuencia presentada por la mayoría de estas nuevas variantes hace difícil clasificarlas como polimorfismos (FAM ~ 0.004), en la región 3'-UTR de *MAFG* se descubrió un nuevo polimorfismo, cuya elevada frecuencia (FAM= 0.128) podría situarlo como un gen marcador de ancestralidad en nuestra población (ver **Tabla 6-7**).

Por otro lado, a pesar de que diversos polimorfismos presentes en otras poblaciones también fueron identificados en este estudio, hubo importantes diferencias a nivel de su frecuencia. Por ejemplo, el SNP rs11703601 (C/T) de *MAFF* aparece en nuestra población con una FAM= 0.393, en el NCBI se reporta su descubrimiento en el genoma de un solo individuo. Finalmente, el SNP rs10254597 (A/G) de *MAFK* es bastante común en nuestra población (FAM=

0.47), sin embargo no se registran datos de frecuencia a pesar de estar validado mediante el HapMap, “1, 000 genomas” y otros estudios. Por el contrario, polimorfismos, como el rs4821766 (G/A) de *MAFF*, identificados a una baja frecuencia en nuestra población (FAM= 0.041) son más frecuentes en otros grupos humanos, como los coreanos (FAM= 0.091) o los Yoruba (FAM=0.5).

También se encontraron varios polimorfismos cuya frecuencia es semejante a la reportada en otras poblaciones. Por ejemplo, el SNP rs2235264 (A/G) de *MAFF* se identificó con una FAM de 0.442 en nuestra población, esta frecuencia es similar a la de las poblaciones asiática (KHP1, FAM= 0.514) y europea (pilot.1.CEU, FAM= 0.5). El cambio rs3829898 (G/A) de *MAFK* en nuestra población tiene una FAM= 0.464, mientras que en la muestra pilot.1.CHB + JPT es de 0.739 y de 0.680 en los Yoruba (pilot.1.YRI). Asimismo, se encontraron polimorfismos como el SNP rs4821766 (G/A) de *MAFF* encontrado con una FAM= 0.041 validado únicamente en coreanos con una FAM= 0.091.

La presencia de polimorfismos en las regiones 5'-UTR y 3'-UTR podría tener importantes efectos sobre la regulación del mRNA. Por un lado, la longitud y contenido GC, así como la ubicación de estructuras secundarias como los IRES (sitios internos de entrada al ribosoma) influyen en la tasa de traducción. Además, la región 5'-UTR contiene numerosos sitios de unión para proteínas que reprimen o promueven la traducción en respuesta a señales moleculares de transmisión. Por otro lado, la región 3'-UTR regula numerosos aspectos del metabolismo del mRNA como la exportación nuclear, ubicación citoplasmática, eficiencia de la traducción y estabilidad, así como la unión de microRNAs (miRNAs). Por ejemplo, el SNP rs61764370 situado en el gen *KRAS* altera un sitio complementario del miRNA Let-7. Este cambio fue encontrado en el 20% de los casos de cáncer pulmonar y solamente en el 5% de la población control; este SNP también se ha ligado con una reducción de la supervivencia de los pacientes con cáncer oral⁽¹⁰⁹⁾. Entre los 120 000 SNPs conocidos como residentes de la región 3'-UTR, alrededor

del 17% tentativamente alteran sitios de unión de miRNAs y un 8.6% crean nuevos sitios de unión⁽¹⁰⁹⁾.

Para los genes *MAF* solamente se conoce un estudio referente a la presencia de un SNP en la región 3'-UTR de *MAFG* y su nexa con alteraciones en la expresión o función génica. En un estudio efectuado en fumadores con riesgo de cáncer pulmonar, se demostró la influencia de mir-218 sobre la expresión de *MAFG* y alteración en su número de copias en este cáncer; se descubrió que la expresión de *MAFG* es modulada negativamente por mir-218⁽¹¹⁰⁾. Las tres variantes identificadas en nuestra población no afectan la unión de miRNAs conocidos, según la búsqueda realizada en la base miRBase (<http://www.mirbase.org/search.shtml>).

El etiquetado de SNPs se refiere a los métodos para seleccionar los SNPs representativos de la variación genética asociada a un conjunto de ellos. Esto reduce el número de variantes requerido para obtener una cobertura máxima de las regiones genómicas analizadas, reduciendo así tiempos y costos. El número de SNPs identificados con un fuerte LD ($r^2 \geq 0.9$) indica que *MAFF* es un gen que ha experimentado una mayor recombinación puesto que sólo 4 SNPs (4/9) estuvieron en fuerte LD y 4 haplotipos explicaron alrededor del 80% de la variación observada. Por el contrario, en *MAFK* un mayor número de SNPs está en fuerte LD y aproximadamente el 80% de la variabilidad está representada por sólo 3 haplotipos. La elección cuidadosa de algunos de estos SNPs permitirá predecir el estado alélico de las variantes en el resto de los SNPs. El conocimiento sobre los patrones de LD en los genes *MAF* permitirá el diseño de plataformas de genotipificación eficientes que permitan descartar SNPs redundantes o que no pueden ser genotipificados de manera confiable.

El análisis *in silico* para SNPs que potencialmente pudieran alterar los sitios de unión para factores de transcripción aportó datos interesantes. En el caso de *MAFF* el SNP rs11703601 podría afectar la unión de factores como PAX9 y NR2F2 (COUP-TF II). PAX9 es necesario para el desarrollo embrionario y se expresa en las células del mesénquima dental⁽¹¹¹⁾. Además este factor es importante en el desarrollo de labios y dientes, mutaciones en *PAX9* se han asociado con oligodontia (falta congénita de 6 o más dientes)⁽¹¹²⁾. En un reporte se indica la coexpresión de PAX9 y MAFF durante el desarrollo del primer arco faríngeo humano, una estructura embrionaria conservada que se desarrolla en el paladar y mandíbula⁽¹¹³⁾. El factor NR2F2²⁰ es un receptor nuclear huérfano que desempeña funciones esenciales en las células de Leydig, particularmente en la regulación de genes CYP450 esteroideogénicos⁽¹¹⁴⁾ y también es un regulador central de la angiogénesis⁽¹¹⁵⁾.

En la región regulatoria de *MAFG* algunas variantes se ubicaron en sitios de unión para factores de transcripción como MTF1, MUSCLE INI, MINI19 y MINI20. El factor MTF1 (*Metal-activated transcription factor 1*) es un mediador de la inducción de metalotioneínas I y II por el zinc, cadmio, arsénico y señales de estrés⁽¹¹⁶⁾, procesos en los que se ha visto implicada MAFG. Finalmente, se encontraron variantes en tres secuencias MINI (*muscle initiator sequences*), un dominio encontrado normalmente en genes específicos del músculo.

Las variantes situadas en el promotor de *MAFK* se ubicaron en sitios de unión para los factores REST/NRSF, ZNF423 (ROAZ), MEF2 y ZEC. El factor REST (*RE1-silencing transcription factor*) es un regulador transcripcional que reprime a toda una batería de genes de diferenciación neuronal en células que no son neuronas o células progenitoras de neuronas⁽¹¹⁷⁾. Esto es llamativo porque

²⁰ COUP-TF II es un receptor huérfano nuclear que pertenece a la superfamilia de receptores nucleares de hormonas tiroideas y reconoce secuencias repetidas de DNA de tipo RGG/TTCA.

MAFK opera en la neuritogénesis⁽⁶⁵⁾, en el estriato²¹ de la rata se ha demostrado la coexpresión de REST y *MAFK*⁽¹¹⁸⁾. Por otra parte, ZNF423 es un coactivador transcripcional de los receptores del ácido retinoico⁽¹¹⁹⁾. Se ha demostrado que cuando aumenta la expresión de p210BCR/ABL y ZNF423 se expresa de manera aberrante, se induce una crisis blástica en la leucemia mieloide crónica⁽¹²⁰⁾.

Las proteínas MEF2 (*myocyte enhancer factor 2*) regulan la expresión de varios genes que codifican proteínas contráctiles y otros genes asociados con el desempeño del músculo⁽¹²¹⁾. Por último, ZEC es una proteína dedo de zinc, descubierta en 2004, que se expresa abundantemente en los testículos, hígado y riñón, así como en las células troncales embrionarias⁽¹²²⁾.

Según el análisis efectuado con rVISTA, algunos de estos cambios podrían contribuir en el aspecto funcional o en la susceptibilidad a enfermedades. Dada las limitaciones inherentes de los estudios *in silico* esto requiere el respaldo experimental. Un primer paso en esta dirección serán los estudios de expresión *in vitro* de las variantes residentes en sitios de unión para factores de transcripción, esto aprovechando la composición de los haplotipos inferidos.

El análisis de las variantes encontradas en los genes *MAF* indica que nuestra población claramente difiere con respecto a otros grupos humanos. Para cada gen se identificaron SNPs ya reportados en el NCBI, otras variantes no pudieron ser detectadas y, finalmente, se descubrieron nuevas variantes no reportadas en el NCBI. Por otra parte, la identificación de diversos haplotipos en los genes *MAFF* y *MAFK* permitió identificar variantes con un fuerte desequilibrio de ligamiento que eventualmente pudieran utilizarse como SNPs marca en estudios futuros de casos-control. Dado que la mayor parte de los polimorfismos encontrados residen en las regiones promotora y UTRs, son factibles los estudios dirigidos a investigar

²¹ El estriato envía señales sobre el movimiento y pensamiento superior. Es un centro de mensajes y también recibe señales de varias regiones cerebrales más.

la presencia de polimorfismos funcionales que pudieran afectar la actividad de estos genes. Los haplotipos identificados pueden facilitar estos estudios dado que algunos de ellos están constituidos casi en su totalidad por alelos ancestrales y otros por alelos menores, de manera que es posible hacer estudios comparativos utilizando estos contrastes. Debe señalarse que hacen falta estudios de resecuenciación de estos genes en otras poblaciones para tener un registro más fino de la variación presente en estos genes e identificar a los polimorfismos comunes en las poblaciones humanas. Este trabajo da a conocer la variabilidad encontrada para estos genes en la población mexicana, los datos de frecuencia y estructura de haplotipos permitirán diseñar estrategias experimentales encaminadas a investigar el impacto de estos polimorfismos sobre la función y posible repercusión en la susceptibilidad a desarrollar diversos padecimientos en nuestra población.

8 CONCLUSIONES.

- En nuestra población, los tres genes *MAF* mostraron un distinto grado de variabilidad según el siguiente orden: *MAFF* > *MAFK* > *MAFG*. Esto sugiere la mayor importancia fisiológica de *MAFG* y concuerda con los estudios en líneas celulares y ratones *knockout*, en los que los fenotipos más graves aparecen en los animales *MAFG*^{-/-}, o en los doble mutantes *MAFG*^{-/-}:*MAFK*^{-/-}.
- La mayor parte de las variantes detectadas se ubicaron en las regiones promotoras y UTRs, ninguno de los cambios se detectaron en los exones codificantes. Lo anterior sugiere que las proteínas MAF participan en funciones críticas para el desarrollo y supervivencia de los organismos, de tal manera que las regiones codificantes de estos genes presentan poca plasticidad genética.
- La identificación de nuevas variantes, así como la ausencia de algunos polimorfismos frecuentes en otras poblaciones, indican un distinto perfil genético de nuestra población con respecto a otras poblaciones.
- Varios de los SNPs identificados se encontraron en fuerte desequilibrio de ligamiento ($r^2 > 0.8$). Algunos de los haplotipos inferidos para *MAFF* y *MAFK* están compuestos casi totalmente por los alelos ancestrales, y otros por los alelos menores. Además, el número y naturaleza de los haplotipos indica claramente que *MAFF* es el gen que ha experimentado una mayor recombinación.
- El análisis *in silico* efectuado con la herramienta bioinformática rVISTA sugiere que algunas de las variantes de la región promotora podrían afectar sitios de unión para TFs.

9 PERSPECTIVAS.

El número y la naturaleza de las variaciones identificadas en las regiones regulatorias de los genes *MAF*, ofrecen la posibilidad de emprender un análisis de expresión *in vitro* de estos genes. La detección de haplotipos, compuestos casi totalmente por alelos ancestrales y otros con alelos menores, permitiría efectuar el estudio de estos haplotipos extremos en construcciones promotor-gen reportero. Mediante este enfoque se tendría alguna oportunidad de descubrimiento de SNPs regulatorios funcionales. Inclusive, podría efectuarse un enfoque mediante mutagénesis dirigida para explorar esta posibilidad en el caso de las variantes nuevas.

10 BIBLIOGRAFÍA.

1. Zorov, D. B., Juhaszova, M. y Sollott, S. J.; Mitochondrial ROS-induced ROS release: an update and review. *Biochim Biophys Acta*; 2006. **1757** (5-6): 509-517.
2. Fialkow, L., Wang, Y. y Downey, G. P.; Reactive oxygen and nitrogen species as signaling molecules regulating neutrophil function. *Free Radic Biol Med*; 2007. **42** (2): 153-164.
3. Benz, C. C. y Yau, C.; Ageing, oxidative stress and cancer: paradigms in parallax. *Nat Rev Cancer*; 2008. **8** (11): 875-879.
4. Rushmore, T. H. y Kong, A. N.; Pharmacogenomics, regulation and signaling pathways of phase I and II drug metabolizing enzymes. *Curr Drug Metab*; 2002. **3** (5): 481-490.
5. Xu, C., Li, C. Y. y Kong, A. N.; Induction of phase I, II and III drug metabolism/transport by xenobiotics. *Arch Pharm Res*; 2005. **28** (3): 249-268.
6. Nguyen, T., Sherratt, P. J. y Pickett, C. B.; Regulatory mechanisms controlling gene expression mediated by the antioxidant response element. *Annu Rev Pharmacol Toxicol*; 2003. **43** 233-260.
7. Itoh, K., Chiba, T., Takahashi, S., Ishii, T., Igarashi, K., *et al.*; An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. *Biochem Biophys Res Commun*; 1997. **236** (2): 313-322.
8. Katsuoka, F., Motohashi, H., Ishii, T., Aburatani, H., Engel, J. D. y Yamamoto, M.; Genetic evidence that small maf proteins are essential for the activation of antioxidant response element-dependent genes. *Mol Cell Biol*; 2005. **25** (18): 8044-8051.
9. Pickett, C. B. P. D., Nguyen, T. y Nioi, P.; The Nrf2-ARE signaling pathway and its activation by oxidative stress. *J Biol Chem*; 2009.
10. Noda, S., Harada, N., Hida, A., Fujii-Kuriyama, Y., Motohashi, H. y Yamamoto, M.; Gene expression of detoxifying enzymes in AhR and Nrf2 compound null mutant mouse. *Biochem Biophys Res Commun*; 2003. **303** (1): 105-111.
11. Iizuka, T., Ishii, Y., Itoh, K., Kiwamoto, T., Kimura, T., *et al.*; Nrf2-deficient mice are highly susceptible to cigarette smoke-induced emphysema. *Genes Cells*; 2005. **10** (12): 1113-1125.
12. Reddy, N. M., Kleeberger, S. R., Kensler, T. W., Yamamoto, M., Hassoun, P. M. y Reddy, S. P.; Disruption of Nrf2 impairs the resolution of hyperoxia-induced acute lung injury and inflammation in mice. *J Immunol*; 2009. **182** (11): 7264-7271.
13. Chen, X. L. y Kunsch, C.; Induction of cytoprotective genes through Nrf2/antioxidant response element pathway: a new therapeutic approach for the treatment of inflammatory diseases. *Curr Pharm Des*; 2004. **10** (8): 879-891.
14. Surh, Y. J., Kundu, J. K. y Na, H. K.; Nrf2 as a master redox switch in turning on the cellular signaling involved in the induction of cytoprotective genes by some chemopreventive phytochemicals. *Planta Med*; 2008. **74** (13): 1526-1539.
15. Kundu, J. K. y Surh, Y. J.; Nrf2-Keap1 signaling as a potential target for chemoprevention of inflammation-associated carcinogenesis. *Pharm Res*; 2010. **27** (6): 999-1013.
16. Bloom, D. A. y Jaiswal, A. K.; Phosphorylation of Nrf2 at Ser40 by protein kinase C in response to antioxidants leads to the release of Nrf2 from I κ Nrf2, but is not required for Nrf2 stabilization/accumulation in the nucleus and transcriptional activation of antioxidant response element-mediated NAD(P)H:quinone oxidoreductase-1 gene expression. *J Biol Chem*; 2003. **278** (45): 44675-44682.

17. Kaspar, J. W., Niture, S. K. y Jaiswal, A. K.; Nrf2:Keap1 signaling in oxidative stress. *Free Radic Biol Med*; 2009. **47** (9): 1304-1309.
18. Jain, A. K. y Jaiswal, A. K.; Phosphorylation of tyrosine 568 controls nuclear export of Nrf2. *J Biol Chem*; 2006. **281** (17): 12132-12142.
19. Ishikawa, M., Numazawa, S. y Yoshida, T.; Redox regulation of the transcriptional repressor Bach1. *Free Radic Biol Med*; 2005. **38** (10): 1344-1352.
20. Levy, S. y Forman, H. J.; C-Myc is a Nrf2-interacting protein that negatively regulates phase II genes through their electrophile responsive elements. *IUBMB Life*; 2010. **62** (3): 237-246.
21. Levy, S., Jaiswal, A. K. y Forman, H. J.; The role of c-Jun phosphorylation in EpRE activation of phase II genes. *Free Radic Biol Med*; 2009. **47** (8): 1172-1179.
22. Venugopal, R. y Jaiswal, A. K.; Nrf1 and Nrf2 positively and c-Fos and Fra1 negatively regulate the human antioxidant response element-mediated expression of NAD(P)H:quinone oxidoreductase1 gene. *Proc Natl Acad Sci U S A*; 1996. **93** (25): 14960-14965.
23. Motohashi, H., Shavit, J. A., Igarashi, K., Yamamoto, M. y Engel, J. D.; The world according to Maf. *Nucleic Acids Res*; 1997. **25** (15): 2953-2959.
24. Kimura, T., Ivell, R., Rust, W., Mizumoto, Y., Ogita, K., Kusui, C., Matsumura, Y., Azuma, C. y Murata, Y.; Molecular cloning of a human MafF homologue, which specifically binds to the oxytocin receptor gene in term myometrium. *Biochem Biophys Res Commun*; 1999. **264** (1): 86-92.
25. Blank, V., Knoll, J. H. y Andrews, N. C.; Molecular characterization and localization of the human MAFG gene. *Genomics*; 1997. **44** (1): 147-149.
26. Iwata, T., Kogame, K., Toki, T., Yokoyama, M., Yamamoto, M. y Ito, E.; Structure and chromosome mapping of the human small maf-genes MAFG and MAFK. *Cytogenet Cell Genet*; 1998. **82** (1-2): 88-90.
27. Motohashi, H., Igarashi, K., Onodera, K., Takahashi, S., Ohtani, H., Nakafuku, M., Nishizawa, M., Engel, J. D. y Yamamoto, M.; Mesodermal- vs. neuronal-specific expression of MafK is elicited by different promoters. *Genes Cells*; 1996. **1** (2): 223-238.
28. Katsuoka, F., Motohashi, H., Onodera, K., Suwabe, N., Engel, J. D. y Yamamoto, M.; One enhancer mediates mafK transcriptional activation in both hematopoietic and cardiac muscle cells. *EMBO J*; 2000. **19** (12): 2980-2991.
29. Onodera, K., Shavit, J. A., Motohashi, H., Katsuoka, F., Akasaka, J. E., Engel, J. D. y Yamamoto, M.; Characterization of the murine mafF gene. *J Biol Chem*; 1999. **274** (30): 21162-21169.
30. Yang, Y. y Cvekl, A.; Large Maf Transcription Factors: Cousins of AP-1 Proteins and Important Regulators of Cellular Differentiation. *Einstein J Biol Med*; 2007. **23** (1): 2-11.
31. Blank, V.; Small Maf proteins in mammalian gene control: mere dimerization partners or dynamic transcriptional regulators? *J Mol Biol*; 2008. **376** (4): 913-925.
32. Blank, V. y Andrews, N. C.; The Maf transcription factors: regulators of differentiation. *Trends Biochem Sci*; 1997. **22** (11): 437-441.
33. Miller, M.; The importance of being flexible: the case of basic region leucine zipper transcriptional regulators. *Curr Protein Pept Sci*; 2009. **10** (3): 244-269.

34. Motohashi, H., Fujita, R., Takayama, M., Inoue, A., Katsuoka, F., Bresnick, E. H. y Yamamoto, M.; Molecular Determinants for Small Maf Protein Control of Platelet Production. *Mol Cell Biol*; 2010.
35. Ye, X., Li, Y., Huang, Q., Yu, Y., Yuan, H., *et al.*; The novel human gene MIP functions as a co-activator of hMafF. *Arch Biochem Biophys*; 2006. **449** (1-2): 87-93.
36. Murphy, P. y Kolsto, A.; Expression of the bZIP transcription factor TCF11 and its potential dimerization partners during development. *Mech Dev*; 2000. **97** (1-2): 141-148.
37. Kataoka, K., Yoshitomo-Nakagawa, K., Shioda, S. y Nishizawa, M.; A set of Hox proteins interact with the Maf oncoprotein to inhibit its DNA binding, transactivation, and transforming activities. *J Biol Chem*; 2001. **276** (1): 819-826.
38. Kobayashi, A., Ito, E., Toki, T., Kogame, K., Takahashi, S., Igarashi, K., Hayashi, N. y Yamamoto, M.; Molecular cloning and functional characterization of a new Cap'n' collar family transcription factor Nrf3. *J Biol Chem*; 1999. **274** (10): 6443-6452.
39. Yamamoto, T., Kyo, M., Kamiya, T., Tanaka, T., Engel, J. D., Motohashi, H. y Yamamoto, M.; Predictive base substitution rules that determine the binding and transcriptional specificity of Maf recognition elements. *Genes Cells*; 2006. **11** (6): 575-591.
40. Kurokawa, H., Motohashi, H., Sueno, S., Kimura, M., Takagawa, H., Kanno, Y., Yamamoto, M. y Tanaka, T.; Structural Basis of Alternative DNA Recognition by Maf Transcription Factors. *Mol Cell Biol*; 2009.
41. Blobel, G. A.; CBP and p300: versatile coregulators with important roles in hematopoietic gene expression. *J Leukoc Biol*; 2002. **71** (4): 545-556.
42. Motohashi, H., Katsuoka, F., Miyoshi, C., Uchimura, Y., Saitoh, H., Francastel, C., Engel, J. D. y Yamamoto, M.; MafG sumoylation is required for active transcriptional repression. *Mol Cell Biol*; 2006. **26** (12): 4652-4663.
43. Dhakshinamoorthy, S. y Jaiswal, A. K.; Small maf (MafG and MafK) proteins negatively regulate antioxidant response element-mediated expression and antioxidant induction of the NAD(P)H:Quinone oxidoreductase1 gene. *J Biol Chem*; 2000. **275** (51): 40134-40141.
44. Moran, J. A., Dahl, E. L. y Mulcahy, R. T.; Differential induction of mafF, mafG and mafK expression by electrophile-response-element activators. *Biochem J*; 2002. **361** (Pt 2): 371-377.
45. Suzuki, T., Blank, V., Sesay, J. S. y Crawford, D. R.; Maf genes are involved in multiple stress response in human. *Biochem Biophys Res Commun*; 2001. **280** (1): 4-8.
46. Wild, A. C., Moinova, H. R. y Mulcahy, R. T.; Regulation of gamma-glutamylcysteine synthetase subunit gene expression by the transcription factor Nrf2. *J Biol Chem*; 1999. **274** (47): 33627-33636.
47. Shimokawa, N., Okada, J. y Miura, M.; Cloning of MafG homologue from the rat brain by differential display and its expression after hypercapnic stimulation. *Mol Cell Biochem*; 2000. **203** (1-2): 135-141.
48. Kumaki, I., Yang, D., Koibuchi, N. y Takayama, K.; Neuronal expression of nuclear transcription factor MafG in the rat medulla oblongata after baroreceptor stimulation. *Life Sci*; 2006. **78** (15): 1760-1766.
49. Yamane, A., Nakamura, T., Suzuki, H., Ito, M., Ohnishi, Y., Ikeda, Y. y Miyakawa, Y.; Interferon-alpha 2b-induced thrombocytopenia is caused by inhibition of platelet production but not proliferation and endomitosis in human megakaryocytes. *Blood*; 2008. **112** (3): 542-550.
50. Massrieh, W., Derjuga, A., Doualla-Bell, F., Ku, C. Y., Sanborn, B. M. y Blank, V.; Regulation of the MAFF transcription factor by proinflammatory cytokines in myometrial cells. *Biol Reprod*; 2006. **74** (4): 699-705.

51. Massrieh, W., Derjuga, A. y Blank, V.; Induction of endogenous Nrf2/small maf heterodimers by arsenic-mediated stress in placental choriocarcinoma cells. *Antioxid Redox Signal*; 2006. **8** (1-2): 53-59.
52. Wilson, N. K., Miranda-Saavedra, D., Kinston, S., Bonadies, N., Foster, S. D., *et al.*; The transcriptional program controlled by the stem cell leukemia gene Scl/Tal1 during early embryonic hematopoietic development. *Blood*; 2009. **113** (22): 5456-5465.
53. Romero, D. G., Rilli, S., Plonczynski, M. W., Yanes, L. L., Zhou, M. Y., Gomez-Sanchez, E. P. y Gomez-Sanchez, C. E.; Adrenal transcription regulatory genes modulated by angiotensin II and their role in steroidogenesis. *Physiol Genomics*; 2007. **30** (1): 26-34.
54. Rasmussen, L. M., Frederiksen, K. S., Din, N., Galsgaard, E. D., Christensen, L., Berchtold, M. W. y Panina, S.; Prolactin and estrogen synergistically regulate gene expression and proliferation of breast cancer cells. *Endocr Relat Cancer*; 2010.
55. Dijkmans, T. F., van Hooijdonk, L. W., Schouten, T. G., Kamphorst, J. T., Fitzsimons, C. P. y Vreugdenhil, E.; Identification of new Nerve Growth Factor-responsive immediate-early genes. *Brain Res*; 2009. **1249** 19-33.
56. Altomare, D. A., Rybak, S. M., Pei, J., Maizel, J. V., Cheung, M., Testa, J. R. y Shogen, K.; Onconase responsive genes in human mesothelioma cells: implications for an RNA damaging therapeutic agent. *BMC Cancer*; 2010. **10** 34.
57. Wang, X., Chorley, B. N., Pittman, G. S., Kleeberger, S. R., Brothers, J., 2nd, Liu, G., Spira, A. y Bell, D. A.; Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer. *PLoS ONE*; 2010. **5** (8): e11934.
58. Shimokawa, N., Kumaki, I., Qiu, C. H., Ohmiya, Y., Takayama, K. y Koibuchi, N.; Extracellular acidification enhances DNA binding activity of MafG-FosB heterodimer. *J Cell Physiol*; 2005. **205** (1): 77-85.
59. Ueda, K., Xu, J., Morimoto, H., Kawabe, A. y Imaoka, S.; MafG controls the hypoxic response of cells by accumulating HIF-1alpha in the nuclei. *FEBS Lett*; 2008. **582** (16): 2357-2364.
60. Patel, R. y Maru, G.; Polymeric black tea polyphenols induce phase II enzymes via Nrf2 in mouse liver and lungs. *Free Radic Biol Med*; 2008. **44** (11): 1897-1911.
61. Wan, Y. J., Li, Y. Y., Xia, W., Chen, J., Lv, Z. Q., *et al.*; Alterations in tumor biomarker GSTP gene methylation patterns induced by prenatal exposure to PFOS. *Toxicology*; 2010. **274** (1-3): 57-64.
62. Tanito, M., Masutani, H., Kim, Y. C., Nishikawa, M., Ohira, A. y Yodoi, J.; Sulforaphane induces thioredoxin through the antioxidant-responsive element and attenuates retinal light damage in mice. *Invest Ophthalmol Vis Sci*; 2005. **46** (3): 979-987.
63. Shavit, J. A., Motohashi, H., Onodera, K., Akasaka, J., Yamamoto, M. y Engel, J. D.; Impaired megakaryopoiesis and behavioral defects in mafG-null mutant mice. *Genes Dev*; 1998. **12** (14): 2164-2174.
64. Igarashi, K., Ochiai, K. y Muto, A.; Architecture and dynamics of the transcription factor network that regulates B-to-plasma cell differentiation. *J Biochem*; 2007. **141** (6): 783-789.
65. Ro, Y. T., Jang, B. K., Shin, C. Y., Park, E. U., Kim, C. G. y Yang, S. I.; Akt regulates the expression of MafK, synaptotagmin I, and syntenin-1, which play roles in neuronal function. *J Biomed Sci*; 2010. **17** (1): 18.
66. Motohashi, H., Katsuoka, F., Engel, J. D. y Yamamoto, M.; Small Maf proteins serve as transcriptional cofactors for keratinocyte differentiation in the Keap1-Nrf2 regulatory pathway. *Proc Natl Acad Sci U S A*; 2004. **101** (17): 6379-6384.
67. Chen, L., Fink, T., Ebbesen, P. y Zachar, V.; Temporal transcriptome of mouse ATDC5 chondroprogenitors differentiating under hypoxic conditions. *Exp Cell Res*; 2006. **312** (10): 1727-1744.

68. Song, G., Park, K. y Wang, L.; Gene expression profiling reveals a diverse array of pathways inhibited by nuclear receptor SHP during adipogenesis. *Int J Clin Exp Pathol*; 2009. **2** (3): 275-285.
69. Martin, F., van Deursen, J. M., Shivdasani, R. A., Jackson, C. W., Troutman, A. G. y Ney, P. A.; Erythroid maturation and globin gene expression in mice with combined deficiency of NF-E2 and nrf-2. *Blood*; 1998. **91** (9): 3459-3466.
70. Lecine, P., Blank, V. y Shivdasani, R.; Characterization of the hematopoietic transcription factor NF-E2 in primary murine megakaryocytes. *J Biol Chem*; 1998. **273** (13): 7572-7578.
71. Onodera, K., Shavit, J. A., Motohashi, H., Yamamoto, M. y Engel, J. D.; Perinatal synthetic lethality and hematopoietic defects in compound mafG::mafK mutant mice. *EMBO J*; 2000. **19** (6): 1335-1345.
72. Lecoin, L., Sii-Felice, K., Pouponnot, C., Eychene, A. y Felder-Schmittbuhl, M. P.; Comparison of maf gene expression patterns during chick embryo development. *Gene Expr Patterns*; 2004. **4** (1): 35-46.
73. Katsuoka, F., Motohashi, H., Tamagawa, Y., Kure, S., Igarashi, K., Engel, J. D. y Yamamoto, M.; Small Maf compound mutants display central nervous system neuronal degeneration, aberrant transcription, and Bach protein mislocalization coincident with myoclonus and abnormal startle response. *Mol Cell Biol*; 2003. **23** (4): 1163-1174.
74. van Oostrom, O., de Kleijn, D. P., Fledderus, J. O., Pescatori, M., Stubbs, A., Tuinenburg, A., Lim, S. K. y Verhaar, M. C.; Folic acid supplementation normalizes the endothelial progenitor cell transcriptome of patients with type 1 diabetes: a case-control pilot study. *Cardiovasc Diabetol*; 2009. **8** 47.
75. Lucito, R., Suresh, S., Walter, K., Pandey, A., Lakshmi, B., et al.; Copy-number variants in patients with a strong family history of pancreatic cancer. *Cancer Biol Ther*; 2007. **6** (10): 1592-1599.
76. Crawford, D. C. y Nickerson, D. A.; Definition and clinical importance of haplotypes. *Annu Rev Med*; 2005. **56** 303-320.
77. Liao, P. Y. K. H. L.; From SNPs to functional polymorphism: The insight into biotechnology applications. *Biochemical Engineering Journal*; 2010. **En prensa** (XXX): 1-10.
78. Fryxell, K. J. y Moon, W. J.; CpG mutation rates in the human genome are highly dependent on local GC content. *Mol Biol Evol*; 2005. **22** (3): 650-658.
79. Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V. y Gottesman, M. M.; Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer. *Cancer Res*; 2007. **67** (20): 9609-9612.
80. Chorley, B. N., Wang, X., Campbell, M. R., Pittman, G. S., Noureddine, M. A. y Bell, D. A.; Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: current and developing technologies. *Mutat Res*; 2008. **659** (1-2): 147-157.
81. Manolio, T. A., Brooks, L. D. y Collins, F. S.; A HapMap harvest of insights into the genetics of common disease. *J Clin Invest*; 2008. **118** (5): 1590-1605.
82. Slatkin, M.; Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat Rev Genet*; 2008. **9** (6): 477-485.
83. Vijaykrishnan, J. y Houlston, R. S.; Candidate gene association studies and risk of childhood acute lymphoblastic leukemia: a systematic review and meta-analysis. *Haematologica*; 2010. **95** (8): 1405-1414.
84. Chiu, M. M., Ko, Y. J., Tsou, A. P., Chau, G. Y. y Chau, Y. P.; Analysis of NQO1 polymorphisms and p53 protein expression in patients with hepatocellular carcinoma. *Histol Histopathol*; 2009. **24** (10): 1223-1232.

85. Fagerholm, R., Hofstetter, B., Tommiska, J., Aaltonen, K., Vrtel, R., *et al.*; NAD(P)H:quinone oxidoreductase 1 NQO1*2 genotype (P187S) is a strong prognostic and predictive factor in breast cancer. *Nat Genet*; 2008. **40** (7): 844-853.
86. Li, D., Dandara, C. y Parker, M. I.; The 341C/T polymorphism in the GSTP1 gene is associated with increased risk of oesophageal cancer. *BMC Genet*; 2010. **11** 47.
87. Babusikova, E., Jesenak, M., Kirschnerova, R., Banovcin, P. y Dobrota, D.; Association of oxidative stress and GST-T1 gene with childhood bronchial asthma. *J Physiol Pharmacol*; 2009. **60 Suppl 5** 27-30.
88. Raval, C. M. y Lee, P. J.; Heme oxygenase-1 in Lung Disease. *Curr Drug Targets*; 2010.
89. Bai, C. H., Chen, J. R., Chiu, H. C., Chou, C. C., Chau, L. Y. y Pan, W. H.; Shorter GT repeat polymorphism in the heme oxygenase-1 gene promoter has protective effect on ischemic stroke in dyslipidemia patients. *J Biomed Sci*; 2010. **17** 12.
90. Wagener, F. A., Toonen, E. J., Wigman, L., Fransen, J., Creemers, M. C., *et al.*; HMOX1 promoter polymorphism modulates the relationship between disease activity and joint damage in rheumatoid arthritis. *Arthritis Rheum*; 2008. **58** (11): 3388-3393.
91. Takeda, M., Kikuchi, M., Ubalee, R., Na-Bangchang, K., Ruangweerayut, R., Shibahara, S., Imai, S. y Hirayama, K.; Microsatellite polymorphism in the heme oxygenase-1 gene promoter is associated with susceptibility to cerebral malaria in Myanmar. *Jpn J Infect Dis*; 2005. **58** (5): 268-271.
92. Courtney, A. E., McNamee, P. T., Middleton, D., Heggarty, S., Patterson, C. C. y Maxwell, A. P.; Association of functional heme oxygenase-1 gene promoter polymorphism with renal transplantation outcomes. *Am J Transplant*; 2007. **7** (4): 908-913.
93. Brydun, A., Watari, Y., Yamamoto, Y., Okuhara, K., Teragawa, H., Kono, F., Chayama, K., Oshima, T. y Ozono, R.; Reduced expression of heme oxygenase-1 in patients with coronary atherosclerosis. *Hypertens Res*; 2007. **30** (4): 341-348.
94. Lo, S. S., Lin, S. C., Wu, C. W., Chen, J. H., Yeh, W. I., Chung, M. Y. y Lui, W. Y.; Heme oxygenase-1 gene promoter polymorphism is associated with risk of gastric adenocarcinoma and lymphovascular tumor invasion. *Ann Surg Oncol*; 2007. **14** (8): 2250-2256.
95. Marzec, J. M., Christie, J. D., Reddy, S. P., Jedlicka, A. E., Vuong, H., *et al.*; Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury. *FASEB J*; 2007. **21** (9): 2237-2246.
96. Guan, C. P., Zhou, M. N., Xu, A. E., Kang, K. F., Liu, J. F., Wei, X. D., Li, Y. W., Zhao, D. K. y Hong, W. S.; The susceptibility to vitiligo is associated with NF-E2-related factor2 (Nrf2) gene polymorphisms: a study on Chinese Han population. *Exp Dermatol*; 2008. **17** (12): 1059-1062.
97. Arisawa, T., Tahara, T., Shibata, T., Nagasaka, M., Nakamura, M., *et al.*; Nrf2 gene promoter polymorphism is associated with ulcerative colitis in a Japanese population. *Hepatogastroenterology*; 2008. **55** (82-83): 394-397.
98. Arisawa, T., Tahara, T., Shibata, T., Nagasaka, M., Nakamura, M., *et al.*; The relationship between Helicobacter pylori infection and promoter polymorphism of the Nrf2 gene in chronic gastritis. *Int J Mol Med*; 2007. **19** (1): 143-148.
99. Hua, C. C., Chang, L. C., Tseng, J. C., Chu, C. M., Liu, Y. C. y Shieh, W. B.; Functional haplotypes in the promoter region of transcription factor Nrf2 in chronic obstructive pulmonary disease. *Dis Markers*; 2010. **28** (3): 185-193.
100. von Otter, M., Landgren, S., Nilsson, S., Celojevic, D., Bergstrom, P., *et al.*; Association of Nrf2-encoding NFE2L2 haplotypes with Parkinson's disease. *BMC Med Genet*; 2010. **11** 36.
101. Kruglyak, L. N., D. A.; Variation is the spice of life. *Nature Genetics*; 2001. **27** 234 - 236.

102. Sanger, F., Nicklen, S. y Coulson, A. R.; DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*; 1977. **74** (12): 5463-5467.
103. Dubchak, I. y Ryaboy, D. V.; VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes. *Methods Mol Biol*; 2006. **338** 69-89.
104. Schork, N. J., Murray, S. S., Frazer, K. A. y Topol, E. J.; Common vs. rare allele hypotheses for complex diseases. *Curr Opin Genet Dev*; 2009. **19** (3): 212-219.
105. Ku, C. S., Loy, E. Y., Pawitan, Y. y Chia, K. S.; The pursuit of genome-wide association studies: where are we now? *J Hum Genet*; 2010. **55** (4): 195-206.
106. Lins, T. C., Abreu, B. S. y Pereira, R. W.; TagSNP transferability and relative loss of variability prediction from HapMap to an admixed population. *J Biomed Sci*; 2009. **16** 73.
107. Silva-Zolezzi, I., Hidalgo-Miranda, A., Estrada-Gil, J., Fernandez-Lopez, J. C., Uribe-Figueroa, L., *et al.*; Analysis of genomic diversity in Mexican Mestizo populations to develop genomic medicine in Mexico. *Proc Natl Acad Sci U S A*; 2009. **106** (21): 8611-8616.
108. Salazar-Flores, J., Dondiego-Aldape, R., Rubi-Castellanos, R., Anaya-Palafox, M., Nuno-Arana, I., *et al.*; Population structure and paternal admixture landscape on present-day Mexican-Mestizos revealed by Y-STR haplotypes. *Am J Hum Biol*; 2010. **22** (3): 401-409.
109. Ryan, B. M., Robles, A. I. y Harris, C. C.; Genetic variation in microRNA networks: the implications for cancer research. *Nat Rev Cancer*; 2010. **10** (6): 389-402.
110. Schembri, F., Sridhar, S., Perdomo, C., Gustafson, A. M., Zhang, X., *et al.*; MicroRNAs as modulators of smoking-induced gene expression changes in human airway epithelium. *Proc Natl Acad Sci U S A*; 2009. **106** (7): 2319-2324.
111. Wang, Y., Groppe, J. C., Wu, J., Ogawa, T., Mues, G., D'Souza, R. N. y Kapadia, H.; Pathogenic mechanisms of tooth agenesis linked to paired domain mutations in human PAX9. *Hum Mol Genet*; 2009. **18** (15): 2863-2874.
112. Pawlowska, E., Janik-Papis, K., Poplawski, T., Blasiak, J. y Szczepanska, J.; Mutations in the PAX9 gene in sporadic oligodontia. *Orthod Craniofac Res*; 2010. **13** (3): 142-152.
113. Cai, J., Ash, D., Kotch, L. E., Jabs, E. W., Attie-Bitach, T., *et al.*; Gene expression in pharyngeal arch 1 during human embryonic development. *Hum Mol Genet*; 2005. **14** (7): 903-912.
114. Martin, L. J. y Tremblay, J. J.; Nuclear receptors in leydig cell gene expression and function. *Biol Reprod*; 2010. **83** (1): 3-14.
115. Lin, F. J., Chen, X., Qin, J., Hong, Y. K., Tsai, M. J. y Tsai, S. Y.; Direct transcriptional regulation of neuropilin-2 by COUP-TFII modulates multiple steps in murine lymphatic vessel development. *J Clin Invest*; 2010. **120** (5): 1694-1707.
116. He, X. y Ma, Q.; Induction of metallothionein I by arsenic via metal-activated transcription factor 1: critical role of C-terminal cysteine residues in arsenic sensing. *J Biol Chem*; 2009. **284** (19): 12609-12621.
117. Abrajano, J. J., Qureshi, I. A., Gokhan, S., Molero, A. E., Zheng, D., Bergman, A. y Mehler, M. F.; Corepressor for element-1-silencing transcription factor preferentially mediates gene networks underlying neural stem cell fate decisions. *Proc Natl Acad Sci U S A*; 2010. **107** (38): 16685-16690.
118. Diez, D., Grijota-Martinez, C., Agretti, P., De Marco, G., Tonacchera, M., Pinchera, A., de Escobar, G. M., Bernal, J. y Morte, B.; Thyroid hormone action in the adult brain: gene expression profiling of the effects of single and multiple doses of triiodo-L-thyronine in the rat striatum. *Endocrinology*; 2008. **149** (8): 3989-4000.

119. Huang, S., Laoukili, J., Epping, M. T., Koster, J., Holzel, M., *et al.*; ZNF423 is critically required for retinoic acid-induced differentiation and is a marker of neuroblastoma outcome. *Cancer Cell*; 2009. **15** (4): 328-340.
120. Miyazaki, K., Yamasaki, N., Oda, H., Kuwata, T., Kanno, Y., *et al.*; Enhanced expression of p210BCR/ABL and aberrant expression of Zfp423/ZNF423 induce blast crisis of chronic myelogenous leukemia. *Blood*; 2009. **113** (19): 4702-4710.
121. Potthoff, M. J. y Olson, E. N.; MEF2: a central regulator of diverse developmental programs. *Development*; 2007. **134** (23): 4131-4140.
122. Chen, G. Y., Muramatsu, H., Ichihara-Tanaka, K. y Muramatsu, T.; ZEC, a zinc finger protein with novel binding specificity and transcription regulatory activity. *Gene*; 2004. **340** (1): 71-81.