(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0031446
(43) 공개일자 2023년03월07일

(51) 국제특허분류(Int. Cl.)
C12Q 1/6883 (2018.01)
(52) CPC특허분류
C12Q 1/6883 (2022.01)
C12Q 2600/156 (2013.01)
(21) 출원번호 10-2021-0113683
(22) 출원일자 2021년08월27일
심사청구일자 2021년08월27일

(71) 출원인
한림대학교 산학협력단
강원도 춘천시 한림대학길 1, 한림대학교(옥천동)
(72) 발명자
전진평
경기도 용인시 수지구 동천로 75(동천동)
김봉준
강원도 춘천시 교동길 50
(뒷면에 계속)
(74) 대리인
김남혁

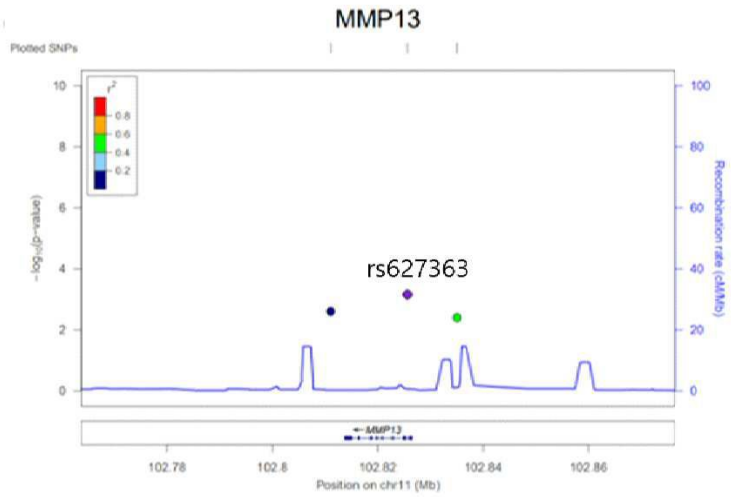
전체 청구항 수 : 총 14 항

(54) 발명의 명칭 **매트릭스 메탈로프로테이나제 13 유전자의 단일염기다형성을 포함하는 뇌동맥류 진단 마커**

(57) 요 약

본 발명은 매트릭스 메탈로프로테이나제 13 유전자의 단일염기다형성을 검출 또는 증폭할 수 있는 제제를 포함하는 뇌동맥류의 예측 또는 진단용 조성물, 키트, 정보 제공 방법 및 진단 장치에 관한 것이다.

대 표 도 - 도1



(72) 발명자
윤동혁
충청북도 단양군 어상천면 어상천로 1589-8

홍은표
서울특별시 노원구 노원로1길 103, 3동 205호(공릉동, 대명아파트)

이 발명을 지원한 국가연구개발사업

과제고유번호	1465034383
과제번호	HR21C0198040021
부처명	보건복지부
과제관리(전문)기관명	한국보건산업진흥원
연구사업명	연구중심병원육성R&D
연구과제명	유전체·인공지능 기반 환자맞춤형 신치료 개발 및 관리 기술 개발
기 여 율	1/1
과제수행기관명	한림대학교 산학협력단
연구기간	2021.07.01 ~ 2029.12.31

명 세 서

청구범위

청구항 1

MMP13 유전자에 대한 단일염기다형성(SNP)을 검출 또는 증폭할 수 있는 제제를 포함하는, 뇌동맥류의 예측 또는 진단용 조성물.

청구항 2

제1항에 있어서,

상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것인, 조성물.

청구항 3

제2항에 있어서,

상기 rs627363의 염기는 T 또는 C인 것인, 조성물.

청구항 4

제2항에 있어서,

상기 rs150333315의 염기는 C 또는 T인 것인, 조성물.

청구항 5

제2항에 있어서,

상기 rs72987535의 염기는 T 또는 C인 것인, 조성물.

청구항 6

제1항에 있어서,

MMP1, MMP2, MMP17, MMP24 및 MMP25에서 선택된 어느 하나 이상의 유전자 서열을 검출 또는 증폭할 수 있는 제제를 더 포함하는, 조성물.

청구항 7

제1항 내지 제6항 중 어느 한 항의 조성물을 포함하는, 뇌동맥류의 예측 또는 진단용 키트.

청구항 8

목적하는 개체로부터 분리된 생물학적 시료로부터, MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 확인하는 단계를 포함하는, 뇌동맥류의 예측 또는 진단용 정보 제공 방법.

청구항 9

제8항에 있어서,

상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것인, 방법.

청구항 10

제9항에 있어서,

상기 rs627363의 염기가 C일 경우, T인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것인, 방법.

청구항 11

제9항에 있어서,

상기 rs150333315의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것인, 방법.

청구항 12

제9항에 있어서,

상기 rs72987535의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것인, 방법.

청구항 13

제8항에 있어서,

상기 개체는 아시아인인 것인, 방법.

청구항 14

목적하는 개체에 대하여 MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 수신하는 수신부; 및

상기 정보를 입력하여 분석하는 연산부;를 포함하는, 뇌동맥류의 예측 또는 진단 장치.

발명의 설명

기 술 분 야

[0001] 본 발명은 매트릭스 메탈로프로테이나제 13 유전자의 단일염기다형성을 검출 또는 증폭할 수 있는 제제를 포함하는 뇌동맥류의 예측 또는 진단용 조성물, 키트, 정보 제공 방법 및 진단 장치에 관한 것이다.

배 경 기 술

[0002] 매트릭스 메탈로프로테이나제 13(matrix metallopeptidase 13; MMP13)은 매트릭스 메탈로프로테이나제(MMP) 유전자군의 일부로, 콜라케나아제 3(Collagenase 3) 효소를 암호화하는 유전자이다. 상기 콜라케나아제 3는 대부

분의 MMP와 마찬가지로 비활성 전구체로 분비하는데, 프로 도메인(pro-domain)이 절단되면 활성화되어 PDB:1PEX로 구성된 활성 효소가 잔존한다. 상기와 같이 전사 및 번역된 단백질은 배아 발달, 생식 및 조직 재형 성과 같은 정상적인 생리적 과정뿐만 아니라 관절염 및 전이와 같은 질병 과정에서 세포 외 기질 분해(extracellular matrix; ECM)에 관여하는데, 다른 MMP 유전자들은 암과의 연관성이 주로 연구되었으나, MMP13 유전자의 단일염기다형성(single nucleotide polymorphism; SNP)에 따른 뇌동맥류와의 관계는 밝혀진 바가 없다.

[0003] 뇌동맥류란 뇌동맥의 일부에 결손이 생겨 그 부분이 돌출된 것을 말하며, 즉, 뇌혈관의 일부가 약한 경우에는 혈관벽이 늘어나 꽈리모양으로 불거져 나오는 것을 의미한다. 주로 뇌동맥류 파열은 자발성 지주막하출혈이 원인 중 대부분을 차지하고 동맥 가지나 근위부에 주로 발생한다. 파열된 뇌동맥류 진단 방법으로는 뇌컴퓨터단층촬영(CT 촬영), 뇌혈관조영술이 있고, 파열되지 않은 뇌동맥류의 진단 방법으로는 자기공명혈관조영술(MRA)이 있다. 상기 뇌동맥류의 경우 대동맥류(aortic aneurysm)와 같은 다른 동맥류와 원인, 진단 방법, 치료 방법 및 예후가 상이할 수 있어, 다른 동맥류 진단 및 치료 방법을 뇌동맥류의 진단 및 치료 방법에 적용하기 번거로울 뿐만 아니라, 뇌동맥류가 발병할지 여부를 미리 예측하거나 뇌동맥류 발병 후 예후를 예측하기에는 어렵다는 문제점이 있다.

[0004] 한편 전장유전체 관련분석(Genome-wide association study; GWAS)은 질환 및 약물 반응성에 대한 유전적 요인을 총체적으로 탐색하는 연구 방법을 말하며, 일본 이화학연구소의 Ozaki(2002) 그룹에서 최초로 시도된 연구 방법이다. HapMap 프로젝트의 완성 및 마이크로어레이 기술의 급속한 발전으로, 유전자 다형성에 관한 전장유전체 스캔이 일상적인 작업이 되었고, 수백의 전장유전체 연관 분석 GWAS이 성공적으로 수행되어 심혈관 질환, 당뇨 등의 통상적인 질병과 관련된 다수의 공지되거나 신규한 유전자 변이체의 발견을 이끌고 있다. 또한, 이러한 기술은 다양한 질환의 진단과 관련된 유전자 다형성 식별에 대한 강력한 수단을 제공하고 있다.

[0005] 이에 본 발명자들은 뇌동맥류 발병 진단, 발병 위험 예측 및 예후 예측을 효율적이고 손쉽게 할 수 있는 방법을 개발하고자 노력한 결과, MMP13 유전자 상의 단일염기다형성(SNP) 대립 형질로부터 뇌동맥류를 예측 또는 진단할 수 있는 진단용 조성물을 개발하여 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0006] 본 발명은 MMP13 유전자에 대한 단일염기다형성(SNP)을 검출 또는 증폭할 수 있는 제제를 포함하는, 뇌동맥류의 예측 또는 진단용 조성물, 키트, 정보 제공 방법 및 진단 장치를 제공하는 것을 목적으로 한다.

[0007] 또한, 본 발명이 해결하고자 하는 과제들은 이상에서 언급된 과제로 제한되지 않으며, 언급되지 않은 또 다른 과제들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

과제의 해결 수단

[0008] 상기와 같은 목적을 달성하기 위하여 본 발명의 일 실시예에 따르면, MMP13 유전자에 대한 단일염기다형성(SNP)을 검출 또는 증폭할 수 있는 제제를 포함하는, 뇌동맥류의 예측 또는 진단용 조성물을 제공할 수 있다.

[0009] 또한 본 발명에서, 상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것일 수 있다.

[0010] 또한 본 발명에서, 상기 rs627363의 염기는 T 또는 C인 것일 수 있다.

[0011] 또한 본 발명에서, 상기 rs150333315의 염기는 C 또는 T인 것일 수 있다.

[0012] 또한 본 발명에서, 상기 rs72987535의 염기는 T 또는 C인 것일 수 있다.

[0013] 또한 본 발명에서, MMP1, MMP2, MMP17, MMP24 및 MMP25에서 선택된 어느 하나 이상의 유전자 서열을 검출 또는 증폭할 수 있는 제제를 더 포함하는 것일 수 있다.

[0014] 본 발명의 다른 일 실시예에 따르면, 본 발명의 조성물을 포함하는, 뇌동맥류의 발병 위험 예측 또는 진단용 키트를 제공할 수 있다.

[0015] 본 발명의 또 다른 일 실시예에 따르면, 개체로부터 분리된 생물학적 시료로부터, MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 확인하는 단계를 포함하는, 뇌동맥류의 예측 또는 진단용 정보 제공 방법을 제공하는 것일 수 있다.

[0016] 또한 본 발명에서, 상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것일 수 있다.

[0017] 또한 본 발명에서, 상기 rs627363의 염기가 C일 경우, T인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0018] 또한 본 발명에서, 상기 rs150333315의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0019] 또한 본 발명에서, 상기 rs72987535의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0020] 또한 본 발명에서, 상기 개체는 아시아인인 것일 수 있다.

[0021] 본 발명의 다른 일 실시예에 따르면, 목적하는 개체에 대하여 MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 수신하는 수신부; 및 상기 정보를 입력하여 분석하는 연산부;를 포함하는, 뇌동맥류의 예측 또는 진단 장치를 제공하는 것일 수 있다.

발명의 효과

[0022] 본 발명에 따라 MMP13 유전자에 대한 단일염기다형성(SNP)을 확인할 경우, 높은 정확도로 뇌동맥류를 예측 또는 진단할 수 있다.

[0023] 본 발명의 효과들은 이상에서 언급된 효과로 제한되지 않으며, 언급되지 않은 또 다른 효과들은 아래의 기재로부터 통상의 기술자에게 명확하게 이해될 수 있을 것이다.

도면의 간단한 설명

[0024] 도 1은 MMP13 유전자의 단일염기다형성(SNP)을 확인하여 뇌동맥류 발생을 예측 또는 진단하는 능력을 확인한 결과를 나타낸 것이다.

도 2는 MMP 유전자 분석, SNP 분석 및 MMP 유전자의 기능 분석을 수행하고 그 결과를 통합하는 과정을 나타낸 것이다.

도 3은 MMP 유전자 마커와 뇌동맥류의 연관성을 확인하여 그 결과를 나타낸 것이다.

도 4는 MMP 유전자 마커와 뇌동맥류의 연관성을 확인하여 그 결과를 나타낸 것이다.

도 5는 MMP 유전자 마커와 뇌동맥류의 연관성을 확인하여 그 결과를 나타낸 것이다.

도 6은 MAGMA를 사용한 MMP 유전자 변이체에 대한 분석을 수행한 결과를 나타낸 것이다.

도 7은 MAGMA를 사용한 MMP 유전자 변이체에 대한 분석을 수행한 결과를 나타낸 것이다.

도 8은 GWAS를 사용하여 MMP 유전자의 SNP를 분석한 결과와 MAGMA를 사용하여 MMP 유전자 기반 분석을 한 결과를 조합하여 나타낸 것이다.

도 9는 뇌동맥류와 관련 있을 것으로 예측되는 5개의 MMP 유전자들에 대한 단백질-단백질 상호 작용(protein-protein interaction; PPI) 네트워크를 확인한 결과를 나타낸 것이다.

도 10은 TIMP1, TIMP2 및 TIMP3의 세 가지 조절자 및 MMP 유전자들 간의 허브 네트워크를 확인한 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0025] 이하, 본 발명에 대하여 상세히 설명하기로 한다. 이에 앞서, 본 명세서 및 특허청구범위에 사용된 용어 또는 단어는 통상적이거나 사전적인 의미로 한정해서 해석되어서는 아니되며, 발명자는 그 자신의 발명을 가장 최선의 방법으로 설명하기 위해 용어의 개념을 적절하게 정의할 수 있다는 원칙에 입각하여 본 발명의 기술적 사상에 부합하는 의미와 개념으로 해석되어야만 한다. 따라서, 본 명세서에 기재된 실시예에 기재된 구성은 본 발명의 가장 바람직한 일 실시예에 불과할 뿐이고 본 발명의 기술적 사상을 모두 대변하는 것은 아니므로, 본 출원 시점에 있어서 이들을 대체할 수 있는 다양한 균등물과 변형예들이 있을 수 있음을 이해하여야 한다.

[0026]

[0027] 본 발명에서 용어 "다형성(polymorphism)"이란 하나의 유전자 좌위(locus)에 두 가지 이상의 대립 유전자(allele)가 존재하는 경우를 말하며, 다형성 부위 중에서, 사람에 따라 단일 염기만이 다른 것을 단일염기다형성(single nucleotide polymorphism; SNP)이라 한다. 진단용 바이오 마커로 사용되기에 바람직한 단일염기다형성(SNP)은 선택된 집단에서 1% 이상, 더욱 바람직하게는 5% 또는 10% 이상의 발생 빈도를 나타내는 두 가지 이상의 대립 유전자를 가진다.

[0028] 본 발명에서 용어 "대립 유전자"는 상동 염색체의 동일한 유전자 좌위에 존재하는 한 유전자의 여러 타입을 의미한다. 대립 유전자는 다형성을 나타내는데 사용되기도 하며, 예컨대, 단일염기다형성(SNP)은 두 종류의 대립인자(biallele)를 갖는다.

[0029] 본 발명에서 사용되는 용어, "rsID"란 1998년부터 단일염기다형성(SNP) 정보를 축적하기 시작한 NCBI가 초기에 등록되는 모든 단일염기다형성(SNP)에 대하여 부여한 독립된 표지자를 의미한다. 본 발명에서는 rs627363과 같은 형태로 기재하였다. 이와 같은 표에 기재된 rsID는 본 발명에서 바이오 마커로 사용되는 단일염기다형성(SNP)을 의미한다. 당업자라면 상기 rsID를 이용하여 단일염기다형성(SNP)의 위치 및 서열을 용이하게 확인할 수 있을 것이다. NCBI의 dbSNP(The Single Nucleotide Polymorphism Database) 번호인 rsID에 해당하는 구체적인 서열은 시간이 지남에 따라 약간 변경될 수 있으나, 본 발명의 범위가 상기 변경된 서열에도 미치는 것은 당업자에게 자명할 것이다.

[0030] 본 발명에서 용어 "뇌동맥류(cerebral aneurysm 또는 brain aneurysm)"는 두개내 동맥류(Intracranial aneurysm; IA)로도 지칭되며, 뇌동맥 또는 뇌정맥의 벽의 약화가 혈관의 국소적 확장 또는 풍선확장(ballooning)을 유발하는 뇌혈관 질환이다. 뇌동맥류의 일반적인 위치는 윌리스환(Circle of Willis)이라고 알려진, 뇌의 기부에 있는 동맥이다. 뇌동맥류의 약 85%가 윌리스환의 앞부분에서 발생하고, 뇌의 앞부분과 중간부분에 혈액을 공급하는 내부 경동맥 및 그들의 주요 분지를 포함한다.

[0031] 본 발명에서 용어 "예측"은, 특정 개체에 대하여 뇌동맥류가 발병할 가능성이 있는지, 발병할 가능성이 있다면 뇌동맥류가 발병할 가능성이 불특정 다수인에 비하여 상대적으로 높은 지 여부를 판별하는 것을 의미한다.

[0032] 본 발명에서 용어 "진단"은, 특정 개체에 대하여 뇌동맥류가 이미 발병하였는지 여부를 판별하는 것을 의미한다.

[0033] 본 발명에서 용어, "개체"는 뇌동맥류가 발병되거나 발병될 가능성이 있는 모든 생물체를 의미하며, 구체적인 예로, 마우스, 원숭이, 소, 돼지, 미니돼지, 가축, 인간 등을 포함하는 포유동물, 양식어류 등을 포함할 수 있으며, 이에 제한되는 것은 아니다.

[0034] 본 발명에서 용어, "시료" 내지 "생물학적 시료"는 뇌동맥류가 발병되거나 발병될 가능성이 있는 개체로부터 유래된 임의의 물질, 생물학적 체액, 조직 또는 세포를 의미하는 것으로, 예를 들면, 전혈(whole blood), 백혈구(leukocytes), 말초혈액 단핵 세포(peripheral blood mononuclear cells), 백혈구 연층(buffy coat), 혈장(plasma), 혈청(serum), 객담(sputum), 눈물(tears), 점액(mucus), 세비액(nasal washes), 비강 흡인물(nasal aspirate), 호흡(breath), 소변(urine), 정액(semen), 침(saliva), 복강 세척액(peritoneal washings), 복수(ascites), 낭종액(cystic fluid), 뇌척수막 액(meningeal fluid), 양수(amniotic fluid), 선액(glandular fluid), 췌장액(pancreatic fluid), 림프액(lymph fluid), 흉수(pleural fluid), 유두 흡인물(nipple aspirate), 기관지 흡인물(bronchial aspirate), 활액(synovial fluid), 관절 흡인물(joint aspirate), 기관 분비물(organ secretions), 세포(cell), 세포 추출물(cell extract) 또는 뇌척수액(cerebrospinal fluid)을 포함할 수 있다.

[0035] 또한, 이들 시료로부터 유전자 시료를 얻을 수 있으며, 유전자 시료는 핵산, 예를 들어, DNA, mRNA, 또는 mRNA로부터 합성되는 cDNA 등을 포함할 수 있으며, 이로부터 단일염기다형성(SNP)의 염기 정보를 확인할 수 있는 한, 그 종류는 이에 제한되는 것은 아니다.

[0036] 본 발명에서, 단일염기다형성(SNP)을 검출 또는 증폭하는 제제는, RNA 시퀀싱(RNAseq) 또는 상기 유전자에 특이적으로 결합하는 프라이머, 프로브 및 안티센스 뉴클레오티드로 이루어진 군에서 선택된 1종 이상을 포함할 수 있다.

[0037] 본 발명에서 상기 "프라이머"는 표적 유전자 서열을 인지하는 단편으로서, 정방향 및 역방향의 프라이머 쌍을 포함하나, 바람직하게는, 특이성 및 민감성을 가지는 분석 결과를 제공하는 프라이머 쌍이다. 프라이머의 핵산

서열이 시료 내 존재하는 비-표적 서열과 불일치하는 서열이어서, 상보적인 프라이머 결합 부위를 함유하는 표적 유전자 서열만 증폭하고 비특이적 증폭을 유발하지 않는 프라이머일 때, 높은 특이성이 부여될 수 있다.

[0038] 본 발명에서, 상기 단일염기다형성(SNP)의 증폭에 사용되는 프라이머는, 적절한 버퍼 중의 적절한 조건(예를 들면, 4개의 다른 뉴클레오시드 트리포스페이트 및 DNA, RNA 폴리머라제 또는 역전사 효소와 같은 중합제) 및 적당한 온도 하에서 주형-지시 DNA 합성의 시작점으로서 작용할 수 있는 단일가닥 올리고뉴클레오티드를 말한다. 상기 프라이머의 적절한 길이는 사용 목적에 따라 달라질 수 있으나, 통상 15 bp 내지 30 bp 길이의 뉴클레오티드 일 수 있다. 짧은 프라이머 분자는 일반적으로 주형과 안정한 혼성체를 형성하기 위해서는 더 낮은 온도를 필요로 한다. 프라이머 서열은 주형과 완전하게 상보적일 필요는 없으나, 주형과 혼성화 할 정도로 충분히 상보적이어야 한다.

[0039] 본 발명에서 상기 "프로브"란 시료 내의 검출하고자 하는 표적 물질과 특이적으로 결합할 수 있는 물질을 의미하며, 상기 결합을 통하여 특이적으로 시료 내의 표적 물질의 존재를 확인할 수 있는 물질을 의미한다. 프로브의 종류는 당 업계에서 통상적으로 사용되는 물질로서 제한은 없으나, 바람직하게는 PNA(peptide nucleic acid), LNA(locked nucleic acid), 펩타이드, 폴리펩타이드, 단백질, RNA 또는 DNA일 수 있으며, 가장 바람직하게는 PNA이다. 보다 구체적으로, 상기 프로브는 바이오 물질로서 생물에서 유래되거나 이와 유사한 것 또는 생체 외에서 제조된 것을 포함하는 것으로, 예를 들어, 효소, 단백질, 항체, 미생물, 동식물 세포 및 기관, 신경세포, DNA, 및 RNA일 수 있으며, DNA는 cDNA, 게놈 DNA, 올리고뉴클레오타이드를 포함하며, RNA는 게놈 RNA, mRNA, 올리고뉴클레오타이드를 포함하며, 단백질의 예로는 항체, 항원, 효소, 펩타이드 등을 포함할 수 있다.

[0040] 본 발명에서 상기 “LNA(Locked nucleic acids)”란, 2'-O, 4'-C 메틸렌 브릿지를 포함하는 핵산 아날로그를 의미한다 [J Weiler, J Hunziker and J Hall Gene Therapy (2006) 13, 496.502]. LNA 뉴클레오사이드는 DNA와 RNA의 일반적 핵산 염기를 포함하며, Watson-Crick 염기 쌍 규칙에 따라 염기 쌍을 형성할 수 있다. 하지만, 메틸렌 브릿지로 인한 분자의 'locking'으로 인해, LNA는 Watson-Crick 결합에서 이상적 형상을 형성하지 못하게 된다. LNA가 DNA 또는 RNA 올리고뉴클레오티드에 포함되면, LNA는 보다 빠르게 상보적 뉴클레오티드 사슬과 쌍을 이루어 이중 나선의 안정성을 높일 수 있다. 본 발명에서 상기 “안티센스”는 안티센스 올리고머가 왓슨-크릭 염기쌍 형성에 의해 RNA 내의 표적 서열과 혼성화되어, 표적서열 내에서 전형적으로 mRNA와 RNA: 올리고머 헤테로이중체의 형성을 허용하는, 뉴클레오티드 염기의 서열 및 서브유닛간 백본을 갖는 올리고머를 의미한다. 올리고머는 표적 서열에 대한 정확한 서열 상보성 또는 근사 상보성을 가질 수 있다.

[0041] 본 발명에서, 상기 프로브는 대립 유전자 특이적 프로브로서, 같은 종의 두 구성원으로부터 유래한 핵산 단편 중에 단일염기다형성(SNP)이 존재하여, 한 구성원으로부터 유래한 DNA 단편에는 혼성화하나, 다른 구성원으로부터 유래한 단편에는 혼성화하지 않는다. 이 경우 혼성화 조건은 대립 유전자간의 혼성화 강도에 있어서 유의한 차이를 보여, 대립 유전자 중 하나에만 혼성화 하도록 충분히 엄격해야 한다. 이렇게 함으로써 다른 대립 유전자 형태 간에 좋은 혼성화 차이를 유발할 수 있다.

[0042] 본 발명에서, 상기 단일염기다형성(SNP)의 유전자형을 확인하는 것, 시퀀싱 분석, 예를 들어, 자동염기서열분석기를 사용한 시퀀싱 분석, 파이로시퀀싱(pyrosequencing), 마이크로어레이에 의한 혼성화, PCR-RELP(restriction fragment length polymorphism)법, PCR-SSCP(single strand conformation polymorphism)법, PCR-SSO(specific sequence oligonucleotide)법, PCR-SSO법과 도트 하이브리드화법을 조합한 ASO(allele specific oligonucleotide) 하이브리드화법, TaqMan-PCR법, MALDI-TOF/MS법, RCA(rolling circle amplification)법, HRM(high resolution melting)법, 프라이머 신장법, 서던 블롯 하이브리드화법, 도트 하이브리드화법 등의 방법에 의하여 수행될 수 있다. 구체적으로,

[0043] 본 발명에서, 상기 단일염기다형성(SNP)의 염기를 결정하는 단계는, 예를 들어, 대상으로부터 얻은 시료로부터 분리한 핵산으로부터 단일염기다형성(SNP)을 포함하는 염기서열을 시퀀싱(sequencing)하고, 단일염기다형성(SNP)의 대립형질을 직접적으로 확인함으로써 수행될 수 있다. 염기서열의 시퀀싱은 당 업계에 공지된 방법을 통하여 수행될 수 있다. 예를들어, 염기서열의 시퀀싱은 개체로부터 얻은 시료로부터 분리한 핵산을 주형으로 하고, 각 단일염기다형성(SNP)을 포함하는 유전자 부위를 증폭할 수 있는 프라이머를 사용하여 PCR을 수행하는 단계, 및 상기 PCR 반응물을 제한효소 절단방법으로 그 서열을 확인하는 단계를 순차적으로 수행함으로써 이뤄질 수 있다.

[0044] 본 발명에서, 상기 시퀀싱 분석은 차세대 염기서열 분석에 의할 수 있다. 상기 차세대 염기서열 분석(Next Generation Sequencing; NGS)은 주형 DNA를 대상으로 짧은 길이의 염기서열을 대용량으로 빠르게 생성시킬 수 있는 방식으로, 저렴한 비용과 빠른 데이터 생산으로 인해 전통적인 생거(Sanger) 시퀀싱 방식을 빠르게 대체하

고 있는 분석 방법이다.

[0045] 본 발명에 따른 유전자의 서열 정보는 알려져 있으므로, 당업자라면 이를 바탕으로 상기 유전자에 특이적으로 결합하는 프라이머, 프로브 또는 안티센스 뉴클레오티드를 용이하게 디자인할 수 있을 것이며 또한 특정 디자인 없이도 일반적인 RNA 시퀀싱 방법을 통해 정성 또는 정량 분석이 가능하다.

[0046] 본 발명에서 용어 "키트"는 바이오 마커 성분에 특이적으로 결합하는 프로브 또는 항체를 검출 가능한 표지로 표지하여 바이오 마커의 발현 수준을 평가할 수 있는 도구를 말한다. 프로브 또는 항체 관련하여 검출 가능한 물질을 기질과의 반응에 의해서 직접적으로 표지하는 것뿐만 아니라, 직접적으로 표지된 다른 시약과의 반응성에 의한 발색하는 표지체가 접합된 간접적 표지도 포함한다. 상기 표지체와 발색 반응할 발색 기질 용액, 세척액 및 기타 다른 용액 등을 포함할 수 있으며, 사용되는 시약 성분을 포함하여 제작될 수 있다. 본 발명에서 키트는 RT-PCR을 수행하기 위해 필요한 필수 요소를 포함하는 키트일 수 있으며, 마커 유전자에 대한 특이적인 각각의 프라이머 쌍 외에도 테스트 튜브, 반응 완충액, 데옥시뉴클레오티드(dNTPs), Taq-중합효소, 역전사효소, DNase, RNase 억제제, 멸균수 등을 포함할 수 있다. 또한, 키트는 DNA 칩을 수행하기 위해 필요한 필수 요소를 포함하는 예후 예측용 유전자를 검출하기 위한 키트일 수 있다. DNA 칩 키트는 유전자 또는 그의 단편에 해당하는 cDNA가 프로브로 부착되어 있는 기판을 포함하고 기판은 정량 대조군 유전자 또는 그의 단편에 해당하는 cDNA를 포함할 수 있다. 본 발명의 키트는 당 업계에 공지되어 있는 것이라면, 이에 제한되지 않는다.

[0047] 본 발명에서 상기 키트는 RT-PCR 키트 또는 DNA 칩 키트일 수 있으나, DNA를 분석하기 위한 것이라면 제한없이 포함될 수 있다.

[0048] 본 발명의 상기 키트는 분석 방법에 적합한 한 종류 또는 그 이상의 다른 구성 성분 조성물, 용액 또는 장치를 더 포함할 수 있다. 예를 들면, 본 발명에서 상기 키트는 역전사 중합효소반응을 수행하기 위해 필요한 필수 요소를 더 포함할 수 있다. 역전사 중합효소반응 키트는 마커 단백질을 암호화하는 유전자에 대해 특이적인 프라이머 쌍을 포함한다. 프라이머는 상기 유전자의 핵산 서열에 특이적인 서열을 가지는 뉴클레오티드로써, 약 7 bp 내지 50 bp의 길이, 보다 바람직하게는 약 10 bp 내지 30 bp의 길이를 가질 수 있다. 또한 대조군 유전자의 핵산 서열에 특이적인 프라이머를 포함할 수 있다. 그 외 역전사 중합효소반응 키트는 테스트 튜브 또는 다른 적절한 용기, 반응 완충액(pH 및 마그네슘 농도는 다양), 데옥시뉴클레오타이드(dNTPs), Taq-폴리머라아제 및 역전사효소와 같은 효소, DNase, RNase 억제제 DEPC-수(DEPC-water), 멸균수 등을 포함할 수 있다.

[0049] 또한, 본 발명의 상기 키트는 DNA 칩을 수행하기 위해 필요한 필수 요소를 포함할 수 있다. DNA 칩 키트는 유전자 또는 그의 단편에 해당하는 cDNA 또는 올리고뉴클레오티드(oligonucleotide)가 부착되어 있는 기판, 및 형광 표지 프로브를 제작하기 위한 시약, 제제, 효소 등을 포함할 수 있다. 또한 기판은 대조군 유전자 또는 그의 단편에 해당하는 cDNA 또는 올리고뉴클레오티드를 포함할 수 있다.

[0050] 본 발명의 상기 키트는 고정체를 더 포함할 수 있고, 상기 고정체로는 니트로셀룰로오즈 막, PVDF 막, 폴리비닐(polyvinyl) 수지 또는 폴리스티렌(polystyrene) 수지로 합성된 웰 플레이트(Well plate), 유리로 된 슬라이드 글래스 등이 사용될 수 있으나, 이에 제한되는 것은 아니다.

[0051] 또한, 본 발명의 상기 키트에서 2차 항체의 표지체는 발색 반응을 하는 통상의 발색제가 바람직하며, HRP(horseradish peroxidase), 염기성 탈인산화효소(alkaline phosphatase), 콜로이드 골드(coloid gold), FITC(폴리 L-라이신-플루오르세인 아이소티오시아네이트), RITC(로다민-B-아이소티오시아네이트) 등의 형광물질(fluorescein) 및 색소(dye) 등의 표지체가 사용될 수 있으나, 이에 제한되는 것은 아니다.

[0052] 또한, 본 발명의 상기 키트에서 발색을 유도하기 위한 발색 기질은 발색 반응을 하는 표지체에 따라 사용하는 것이 바람직하며, TMB(3,3',5,5'-테트라메틸 베지딘), ABTS[2,2'-아지노-비스(3-에틸벤조티아졸린-6-설폰산)], OPD(o-페닐렌다이아민) 등을 사용할 수 있다. 이때, 발색 기질은 완충 용액(0.1 M NaAc, pH 5.5)에 용해된 상태로 제공되는 것이 더욱 바람직하다. TMB와 같은 발색기질은 이차 항체 접합체의 표지체로 사용된 HRP에 의해 분해되어 발색 침적체를 생성하고, 이 발색 침적체의 침적 정도를 육안으로 확인함으로써 상기 마커 단백질들의 존재 유무를 검출한다.

[0053] 본 발명의 상기 키트에서 세척액은 인산염 완충 용액, NaCl 및 트윈 20을 포함하는 것이 바람직하며, 0.02 M 인산염 완충용액, 0.13 M NaCl, 및 0.05% 트윈 20으로 구성된 완충 용액(PBST)이 더욱 바람직하다. 세척액은 항원-항체 결합 반응 후 항원-항체 결합체에 2차 항체를 반응시킨 다음 적당량을 고정체에 첨가하여 3 내지 6회 세척한다. 반응 정지 용액은 황산 용액(H2SO4)이 바람직하게 사용될 수 있다.

[0054] 본 발명에서, 단일염기다형성(SNP)이 위치하는 매트릭스 메탈로프로테이나제(Matrix metalloproteinases; MMPs)

유전자들은 혈역학 적 스트레스에 대한 반응으로 세포 외 기질(ECM) 리모델링에서 중요한 역할을 할 것으로 예상되나, 뇌동맥류 형성에 대한 유전적 영향은 논란의 여지가 있는 유전자들이다. 상기 MMP 유전자는 MMP1, MMP2, MMP3, MMP7, MMP8, MMP9, MMP10, MMP11, MMP12, MMP13, MMP14, MMP15, MMP16, MMP17, MMP18, MMP19, MMP20, MMP21, MMP23A, MMP23B, MMP24, MMP25, MMP26, MMP27 및 MMP28 등이 알려져 있다.

[0055]

[0056] 본 발명의 일 구현예에 따르면, MMP13 유전자에 대한 단일염기다형성(SNP)을 검출 또는 증폭할 수 있는 제제를 포함하는, 뇌동맥류의 예측 또는 진단용 조성물을 제공하는 것이다.

[0057] 본 발명에서, 상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것일 수 있다.

[0058] 본 발명에서, 상기 MMP13의 유전자 서열은 서열번호 1로 표시될 수 있고, 상기 MMP13의 유전자 서열 중 상기 단일염기다형성(SNP)의 예시 및 이의 위치에 대한 정보는 하기 표 1에 나타낸 바와 같다.

[0059]

표 1

[0060]

Chr	BP	M	m	rsID	gene
11	102825635	T	C	rs627363	MMP13
11	102811090	C	T	rs150333315	MMP13
11	102835040	T	C	rs72987535	MMP13

[0061] 본 발명에서, 상기 rs627363의 염기는 T 또는 C인 것일 수 있고, 일 예로 11번 염색체의 102825635번째 염기가 T 또는 C인 것일 수 있다.

[0062] 본 발명에서, 상기 rs150333315의 염기는 C 또는 T인 것일 수 있고, 일 예로 11번 염색체의 102811090번째 염기가 C 또는 T인 것일 수 있다.

[0063] 본 발명에서, 상기 rs72987535의 염기는 T 또는 C인 것일 수 있고, 일 예로 11번 염색체의 102825635번째 염기가 T 또는 C인 것일 수 있다.

[0064] 본 발명에서, rs1555322, rs2425024, rs6119593, rs11642206 및 rs16938619을 검출 또는 증폭할 수 있는 제제를 더 포함하는 뇌동맥류의 예측 또는 진단용 조성물을 제공하는 것일 수 있다. 상기 단일염기다형성(SNP)을 추가적으로 확인할 경우, 뇌동맥류에 대한 예측 또는 진단의 정확도를 높일 수 있으며, 상기 단일염기다형성(SNP)의 정보는 하기 표 2와 같다.

[0065]

표 2

[0066]

Chr	BP	M	m	rsID	gene	서열번호
20	33849179	G	A	rs1555322	MMP24	2
20	33844938	A	C	rs2425024	MMP24	3
20	33833729	C	A	rs6119593	MMP24	4
16	3091695	A	G	rs11642206	MMP25	5
11	4807548	A	G	rs16938619	MMP26	6

[0067] 본 발명에서, MMP1, MMP2, MMP17, MMP24 및 MMP25에서 선택된 어느 하나 이상의 유전자 서열을 검출 또는 증폭할 수 있는 제제를 더 포함하는 것일 수 있다. 분석이 되는 상기 MMP 유전자들의 정보는 하기 표 3과 같다.

[0068]

표 3

[0069]

유전자	Entrez 유전자 ID	염색체	시작	끝
MMP1	4312	11	102660651	102668891
MMP2	4313	16	55423612	55540603

MMP17	4326	12	132312938	132336328
MMP24	10893	20	33814457	33864801
MMP25	64386	16	3096682	3110727

[0070] 본 발명의 다른 일 구현예에 따르면, 본 발명의 조성물을 포함하는, 뇌동맥류의 발병 위험 예측 또는 진단용 키트를 제공하는 것일 수 있다.

[0071] 본 발명에서, 생략된 나머지 기재들은 앞서 기재된 진단용 조성물의 기재와 마찬가지로 해석될 수 있다.

[0072]

[0073] 본 발명의 또 다른 일 구현예에 따르면, 목적하는 개체로부터 분리된 생물학적 시료로부터, MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 확인하는 단계를 포함하는, 뇌동맥류의 예측 또는 진단용 정보 제공 방법을 제공하는 것일 수 있다.

[0074] 본 발명에서, 상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것일 수 있다.

[0075] 본 발명에서, 상기 rs627363의 염기가 C일 경우, T인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다. 일례로, 상기 rs627363의 염기가 C인 동형접합(homogygous)일 경우, 나머지 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0076] 본 발명에서, 상기 rs150333315의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다. 일례로, 상기 rs150333315의 염기가 T인 동형접합(homogygous)일 경우, 이형접합(heterozygous)인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0077] 본 발명에서, 상기 rs72987535의 염기가 T일 경우, C인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다. 일례로, 상기 rs72987535의 염기가 T인 동형접합(homogygous)일 경우, 나머지 경우, 특히 이형접합(heterozygous)인 경우에 비하여 뇌동맥류의 발병 위험이 높을 것으로 예측하거나, 뇌동맥류가 발병한 것으로 진단하는 것일 수 있다.

[0078] 본 발명에서, 상기 개체는 인간일 수 있으며, 구체적으로 한국인 또는 아시아인일 수 있으나, 이에 제한되는 것은 아니다.

[0079] 본 발명에서, 상기 개체의 위험 예측 또는 진단을 수행하기 위해, 비유전적(non-genetic) 정보를 분석하는 단계를 더 포함하고, 상기 비유전적 정보는 성별, 연령, 고혈압, 당뇨, 고지혈증, 흡연, 동맥류의 가족력, 생화학적 척도 및 임상적 척도로 구성되는 군으로부터 선택되는 하나 이상일 수 있다.

[0080] 본 발명에서, 상기 단일염기다형성(SNP)을 확인한 결과를 분석하는 단계를 더 포함할 수 있다. 상기 분석은 당업계에서 일반적으로 사용되는 통계학적 분석 방법을 이용하여 통계처리 할 수 있으며, 예를 들어, 스튜던트 t-검정(Student's t-test), 카이-스퀘어 검정(Chi-square test), 선형 회귀선 분석(linear regression line analysis), 다변량 로지스틱 회귀분석(multiple logistic regression analysis), 메타 분석(meta-analysis) 등을 통해 얻은 연속 변수(continuous variables), 절대 변수 (categorical variables), 대응비(odds ratio, OR) 및 95% 신뢰구간(confidence interval, CI) 등의 변수를 이용하여 분석할 수 있다.

[0081] 본 발명에서, 생략된 나머지 기재들은 앞서 기재된 진단용 조성물 및 키트의 기재와 마찬가지로 해석될 수 있다.

[0082]

[0083] 본 발명의 또 다른 일 구현예에 따르면, 목적하는 개체에 대하여 MMP13 유전자에 대한 단일염기다형성(SNP)의 대립형질을 수신하는 수신부; 및 상기 정보를 입력하여 분석하는 연산부;를 포함하는, 뇌동맥류의 예측 또는 진단 장치를 제공하는 것일 수 있다.

[0084] 본 발명에서, 상기 MMP13 유전자에 대한 단일염기다형성(SNP)의 위치는 rs627363, rs150333315 및 rs72987535에서 선택된 어느 하나 이상인 것일 수 있다.

[0085] 본 발명에서, 상기 수신부는 MMP1, MMP2, MMP17, MMP24 및 MMP25에서 선택된 어느 하나 이상의 유전자 서열에 대한 정보를 더 수신하는 것일 수 있다.

[0086] 본 발명에서, 상기 연산부는 머신 러닝 또는 통계학적 분석 방법을 이용하여 통계처리 하는 것일 수 있다. 상기 머신 러닝 모델은, 지도 학습, 비지도 학습, 준지도 학습, 강화 학습 등 다양한 방식에 기반할 수 있고, 상기 학습은 배치 학습(batch learning), 온라인 학습(online learning) 등에 의할 수 있다. 일 예로, 상기 머신 러닝 모델은 최근접 중심 분류기(Nearest Centroid), 서포트 벡터 머신(Support Vector Machine; SVM), 베이지안 분류기(Bayesian Classifier), 랜덤 포레스트(random forest), 의사 결정 트리(decision tree) 및 그라데이션 부스트 머신(gradient boost machine; GBM)에서 선택된 어느 하나 이상일 수 있다.

[0087] 본 발명에서, 생략된 나머지 기재들은 앞서 기재된 진단용 조성물, 키트 및 정보 제공 방법의 기재와 마찬가지로 해석될 수 있다.

[0088]

[0089] 이하 본 발명을 실시예에 의하여 더욱 상세하게 설명한다. 이들 실시예는 단지 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 범위가 이들 실시예에 국한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에게 있어서 자명할 것이다.

[0090]

[0091] 본 발명에서, 도 1과 같이 MMP13 유전자의 단일염기다형성(SNP)을 확인할 경우, 뇌동맥류 진단 마커, 상기 마커를 검출 또는 증폭할 수 있는 제제를 포함하는 뇌동맥류 진단용 조성물, 뇌동맥류 진단용 키트, 및 상기 조성물 및 키트를 이용한 뇌동맥류 진단을 위한 정보 제공 방법에 활용할 수 있음을 하기 실시예를 통하여 확인하였다.

[0092]

[0093] 실시예 1. GWAS 데이터 획득

[0094] GWAS 데이터 세트에는 성인 뇌동맥류(IA) 환자 250명과 대조군 296명이 포함되었다. 파열된 동맥류(ruptured aneurysm)와 전방 순환계(anterior circulation) 순환의 수는 각각 151명(60.4%)과 221명(88.4%)이었다. Axiom™ Asia Precision Medicine Research Array(PMRA) 키트(미국 매사추세츠 주 Thermo Fisher Scientific)는 인간 게놈 버전 19(빌드 37)를 기반으로 한 동남아시아 인구를 포괄하는 750,000개 이상의 SNP를 포함하는 원시 유전 데이터를 생성하는 데 사용되었다. 원시 데이터는 (1) 유전형 호출률 ≥ 95%, (2) 마이너 대립 유전자 빈도(MAF) ≥ 0.01, (3) 하디-바인버그(Hardy-Weinberg equilibrium; HWE) 평형 p-값 ≥ $1x10^{-6.11}$ 조건으로 필터링 되었다. 본 발명의 실시예는 참여 병원 기관 심사위원회(2017-9, 2018-6, 2019-6)의 승인을 받았다.

[0095] 또한, 성별, 연령, 비파열 뇌동맥류 또는 지주막하 출혈과 같은 임상 증상, 고혈압, 당뇨, 고지혈증, 흡연, 동맥류의 가족력 등 여러 변수를 고려하여 의료 기록을 검토하였으며, 동맥류의 크기, 위치(전방 순환 또는 후방 순환), 및 수(단수 또는 복수)에 대한 혈관조영 변수를 검토하였다.

[0096]

[0097] 실시예 2. 생물 정보학 분석

[0098] MMP 유전자의 SNP 분석 및 MMP 유전자의 기능 분석을 수행하고 그 결과를 통합하는 과정을 도 2와 같이 나타내었다. GWAS 데이터에 주석을 달기 위해 MMP 유전자군에서 SNP/INDEL의 변형 호출을 수행하였다. 맨해튼 플롯(Manhattan Plot) 및 분위수-분위수(quantile-quantile; Q-Q) 플롯은 게놈 전체 연관 연구의 기능 매핑 및 주석(Functional Mapping and Annotation of Genome-Wide Association Studies. FUMA GWAS; https://fuma.ctglab.nl/)에서 구현된 MAGMA 1.6을 사용하여 수행되었다. 기본 설정 및 유의 수준은 유전자 기반 테스트에 의해 p= 0.05 / 23 = $2.17x10^{-3}$으로 정의되었다. Locuszoom 플롯(http://csg.sph.umich.edu/locuszoom)을 사용하여 +/- 50Kb로 초점을 맞춘 지역 플롯을 생성하였다. 유전자 온톨로지(GO) 분석은 생물학적 과정(BP), 세포 성분(CC) 및 분자 기능(MF) 측면에서 분자 서명 데이터베이스(MSigDB C5)를 사용하여 수행되었다(http://www.broadinstitute.org/msigdb). 단백질-단백질 상호 작용(protein-protein interaction; PPI) 네트워크는 상호 작용하는 유전자/단백질 검색 도구(Search Tool for the Retrieval of Interacting Genes/Proteins; STRING) v11(http://string-db.org)을 사용하여 수행되었다. 최소 필수 상호 작용 점수는 중간 신뢰도 0.7로 설정되었다.

[0099]

[0100] 실시예 3. SNP 기반 분석(1)

[0101] MMP 유전자 마커와 뇌동맥류의 연관성을 확인하여 그 결과를 하기 도 3, 도 4 및 도 5에 나타내었다.

[0102] 도 3에서 나타난 것과 같이, MMP 유전자의 343개 유전자 마커 중 총 49개의 SNP/INDEL이 뇌동맥류와 암시적인 연관성을 나타냈다($p < 0.05$). 또한 도 4 및 도 5에서와 같이 SNP 기반 분석 모델의 Q-Q 플롯은 관찰된 p-값의 부풀려진 분포를 보여주었다.

[0103]

[0104] 실시예 4. SNP 기반 분석(2)

[0105] 뇌동맥류와 관련된 MMP 유전자의 상위 20개 SNP/INDEL 마커를 확인하여 염색체(Chr), 염기위치(BP), SNP(rsID) 및 유전자(gene)에 따른 뇌동맥류와의 관계를 확인하여 오즈비(OR), p-값, 유전자형(mm/Mm/MM), 뇌동맥류가 발생한 경우(Case) 및 정상 대조군(Control)을 하기 표 4에 나타내었다.

[0106]

표 4

[0107]

No	Chr	BP	M	m	OR	p-값	Case (mm/Mm/MM)	Control (mm/Mm/MM)	rsID	gene
1	20	33844938	A	C	0.43	2.40E-06	7/47/196	15/113/168	rs2425024	MMP24
2	20	33833729	C	A	0.39	5.90E-05	1/30/218	6/71/219	rs6119593	MMP24
3	11	4807548	A	G	3.12	8.80E-05	0/48/200	0/23/273	rs16938619	MMP26
4	20	33849179	G	A	0.37	1.60E-04	1/23/226	5/58/233	rs1555322	MMP24
5	16	3091695	A	G	0.43	2.60E-04	0/43/207	2/81/212	rs11642206	MMP25
6	16	58105677	T	C	2.05	2.70E-04	3/90/157	3/66/226	rs6499933	MMP15
7	11	102825635	T	C	1.55	7.00E-04	71/123/55	59/137/99	rs627363	MMP13
8	20	33859013	C	A	0.44	8.80E-04	2/23/223	5/58/233	rs2425041	MMP24
9	16	55534078	G	A	0.38	9.50E-04	1/15/232	5/47/244	rs80345658	MMP2
10	11	102811090	C	T	0.1	2.50E-03	0/2/247	0/19/277	rs150333315	MMP13
11	11	4835458	G	A	1.74	3.70E-03	6/73/171	6/53/235	rs7945484	MMP26
12	16	55427433	G	A	0.05	3.70E-03	0/1/249	1/18/276	rs74019045	MMP2
13	11	4909000	G	A	2.01	3.80E-03	0/58/192	2/35/259	rs2595982	MMP26
14	11	102835040	T	C	0.68	4.00E-03	34/104/112	53/150/92	rs72987535	MMP13
15	20	33834708	G	A	0.49	4.20E-03	0/33/217	0/69/227	rs3764733	MMP24
16	12	132333682	A	G	0.68	6.50E-03	26/88/133	29/153/112	rs79572159	MMP17
17	16	55502376	T	G	0.27	6.50E-03	0/6/243	1/21/274	rs11859163	MMP2
18	16	55436868	G	A	0.21	6.80E-03	0/4/246	1/17/278	rs142441049	MMP2
19	20	33828825	G	A	1.44	6.90E-03	57/132/61	53/139/104	rs6142338	MMP24
20	11	4961003	G	A	0.33	8.80E-03	0/9/239	0/27/269	rs36013257	MMP26

[0108] 상기 표 4에 나타난 것처럼, MMP24 유전자에 위치한 rs2425024의 인트론 변이체는 뇌동맥류와 가장 중요한 연관성을 보이는 바, 이는 상기 rs2425024 SNP 변이가 뇌동맥류 형성을 방지할 수 있는 것을 확인하였다(OR = 0.43, 95% CI : 0.30-0.61; $p = 2.4 \times 10^{-6}$). 또한 상기 MMP24 유전자에 위치한 SNP 중 rs6119593 및 rs1555322 역시 뇌동맥류 형성을 방지할 수 있는 효과를 확인하였다. 반면, 상기 MMP24 유전자에 위치한 SNP 중 rs16938619는 뇌동맥류의 위험을 크게 증가시키는 것을 확인할 수 있었다(OR = 3.12, 95% CI: 1.76-5.50; $p = 8.85 \times 10^{-5}$). 상기 변이체의 G 대립 유전자는 뇌동맥류 그룹에 비해 대조군에서 드물게 검출된 것을 확인하였다.

[0109]

[0110] 실시예 5. MMP 유전자 변이체 분석

[0111] MMP 유전자군 중 뇌동맥류 예측과 관련되는 MMP 유전자를 확인하기 위하여, MAGMA를 사용한 MMP 유전자 변이체에 대한 분석을 수행한 결과를 하기 표 5, 도 6 및 도 7에 나타내었다.

[0112]

표 5

[0113]

Gene	Chr	Start	Stop	참조된 SNP	매개변수의 수	p-값
MMP24	20	33814457	33864801	7	3	7.96E-07
MMP13	11	102813724	102826463	3	2	1.79E-03
MMP2	16	55423612	55540603	23	13	0.01
MMP17	12	132312938	132336328	11	5	0.02
MMP1	11	102660651	102668891	2	1	0.02
MMP26	11	4726157	5013659	46	23	0.07
MMP3	11	102706532	102714534	6	1	0.08
MMP16	8	89044237	89340254	30	15	0.12
MMP7	11	102391239	102401484	5	3	0.12
MMP19	12	56229217	56236750	3	2	0.14
MMP25	16	3096682	3110727	5	2	0.14
MMP10	11	102641234	102651359	4	2	0.14
MMP11	22	24110413	24126503	5	2	0.14
MMP14	14	23305766	23318236	9	5	0.2
MMP20	11	102447566	102496063	13	5	0.39
MMP27	11	102562218	102576537	7	3	0.55
MMP28	17	34083268	34122711	6	2	0.56
MMP9	20	44637547	44645200	6	3	0.72
MMP15	16	58059470	58080805	4	1	0.77
MMP21	10	127455022	127464390	2	1	0.87
MMP8	11	102582526	102597781	5	2	0.98

[0114] 상기 표 5에 나타난 것처럼, MMP 유전자군 중 MMP24 유전자는 가장 많은 뇌동맥류 연관 유전자로 나타난 것으로 확인되었다($p = 7.96x10^{-6}$). 또한 도 6 및 도 7에 나타난 것처럼, MAGMA를 사용하여 뇌동맥류 예측을 할 경우 MMP24 및 MMP13 유전자 변이체의 경우 게놈 전체의 유의성 임계 값($p = 0.0024$)을 통과한 바, 본원발명과 같이 MMP24 및 MMP13 유전자를 확인할 경우, 뇌동맥류의 발생 여부를 예측할 수 있음을 확인하였다.

[0115]

[0116] 실시예 6. SNP 분석 및 변이체 분석의 조합

[0117] GWAS를 사용하여 MMP 유전자의 SNP를 분석한 결과와 MAGMA를 사용하여 MMP 유전자 기반 분석을 한 결과를 조합하여 도 8에 나타내었다.

[0118] 도 8에 나타난 것과 같이, 뇌동맥류와 연관성이 높은 MMP 유전자의 SNP는 MMP24에서 rs2425024, MMP13에서 rs627363, MMP2에서 rs80345658, MMP1에서 rs79572159 및 MMP17에서 rs2071232인 것을 확인할 수 있었고, MAGMA 결과 뇌동맥류와 연관성이 높은 MMP 유전자는 MMP24, MMP13, MMP2, MMP17 및 MMP1의 5개로 확인되었다.

[0119]

[0120] 실시예 7. In silico에서 MMP 유전자군의 기능 분석

[0121] MMP 유전자들 간 생물학적 과정(biological process; BP)을 분석하였다. 구체적으로는, 뇌동맥류와 관련 있을 것으로 예측되는 5개의 MMP 유전자들에 대한 단백질-단백질 상호 작용(protein-protein interaction; PPI) 네트워크를 확인하여 도 9에 나타내었고, 이를 통하여 TIMP1, TIMP2 및 TIMP3의 세 가지 조절자 및 상기 MMP 유전자들 간의 허브 네트워크를 확인하여 도 10에 나타내었다.

[0122] 도 9에 나타난 것처럼, MMP 유전자군들 간의 상기 분석을 통하여 가장 풍부한 생물학적 과정기능이 콜라겐 이화 과정이었고, 다음으로는 세포 외 구조 조직(extracellular structure organization)과 세포 외 기질(extracellular matrix; ECM) 분해 기능을 하는 것을 확인하였다. 또한 세포 성분의 경우, 메탈론도펩티다아제(metalloendopeptidase) 활성이 뇌동맥류 형성에서 가장 풍부한 것으로 나타났으며, MF 클러스터에서 ECM은 5개의 뇌동맥류 관련 MMP 유전자들이 중첩되어 가장 풍부한 것으로 확인되었다.

[0123] 또한 도 10에 나타난 것처럼, 뇌동맥류 형성과 관련한 암시적인 허브 네트워크를 예측한 결과, MMP 유전자군 중

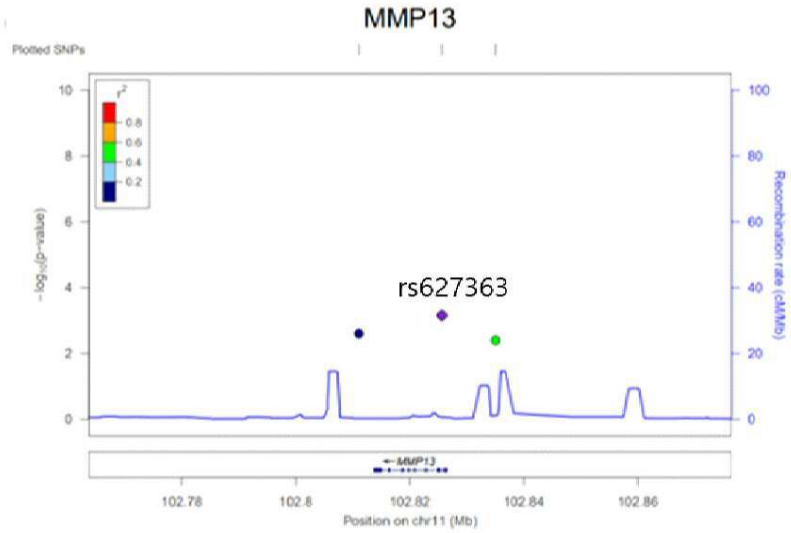
MMP2 유전자가 TIMP2 및 TIMP3 조절자와 관련이 있는 것으로 나타난 바, 뇌동맥류 형성에 관련이 있음을 확인할 수 있었다.

[0124]

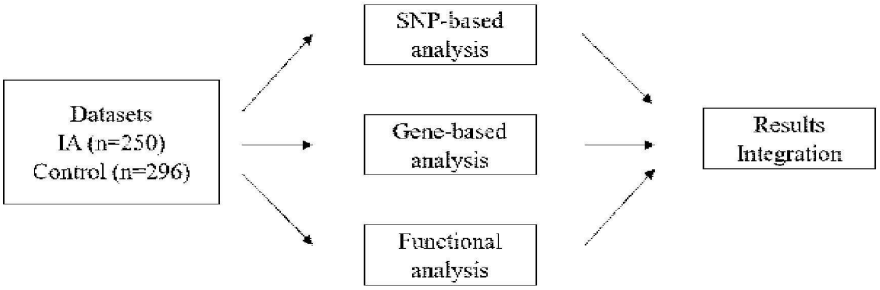
[0125] 이상으로 본 발명의 특정한 부분을 상세히 기술하였는 바, 당업계의 통상의 지식을 가진 자에게 있어서, 이러한 구체적 기술은 단지 바람직한 실시예일 뿐이며, 이에 의해 본 발명의 범위가 제한되는 것이 아닌 점은 명백할 것이다. 따라서, 본 발명의 실질적인 범위는 첨부된 청구항들과 그것들의 등가물에 의하여 정의된다고 할 것이다.

도면

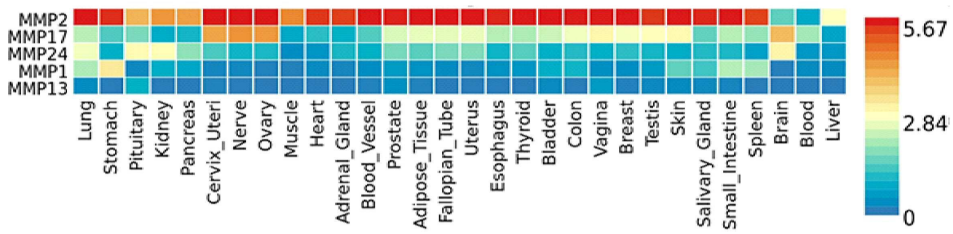
도면1



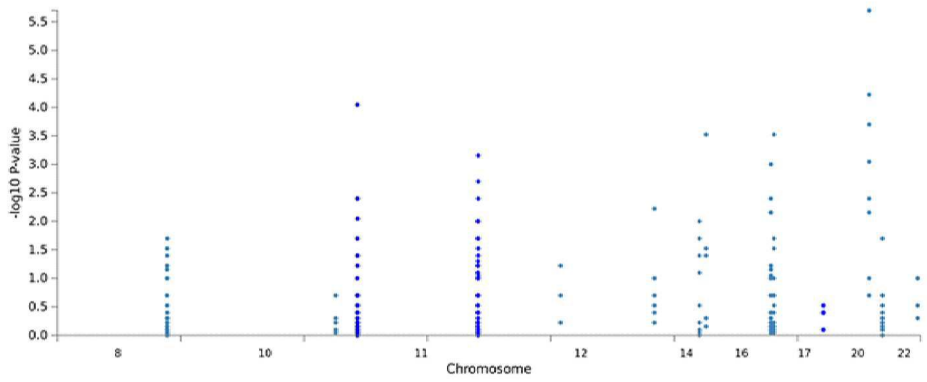
도면2



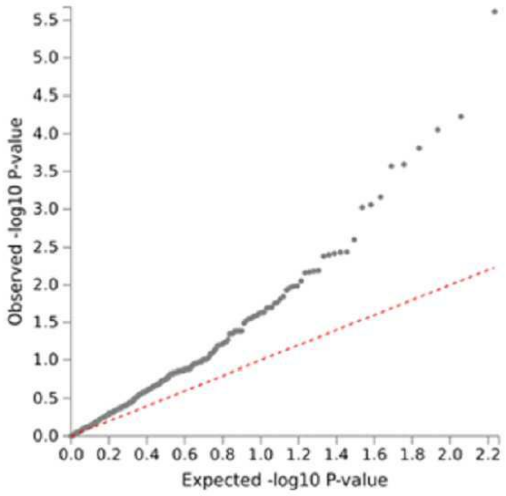
도면3



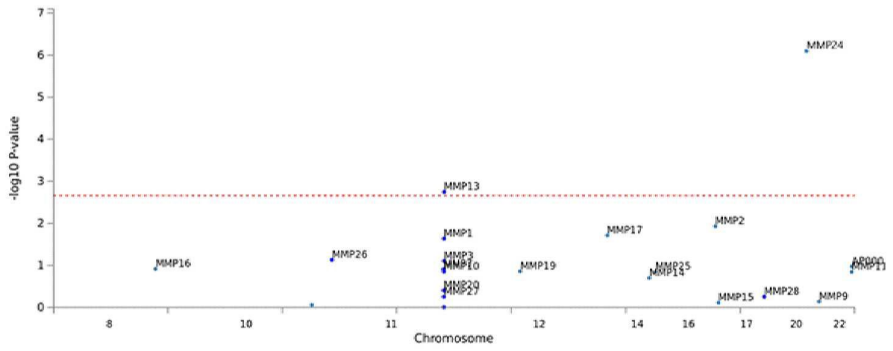
도면4


-log10 P-value
Chromosome

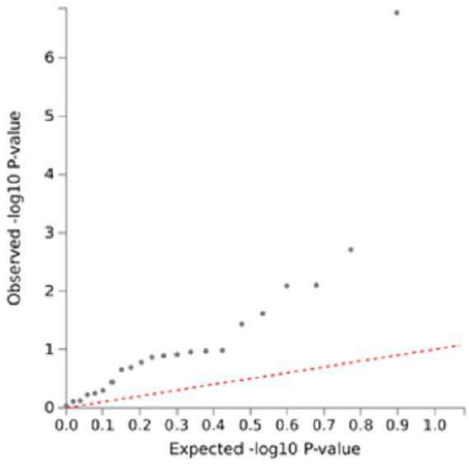
도면5


Observed -log10 P-value
Expected -log10 P-value

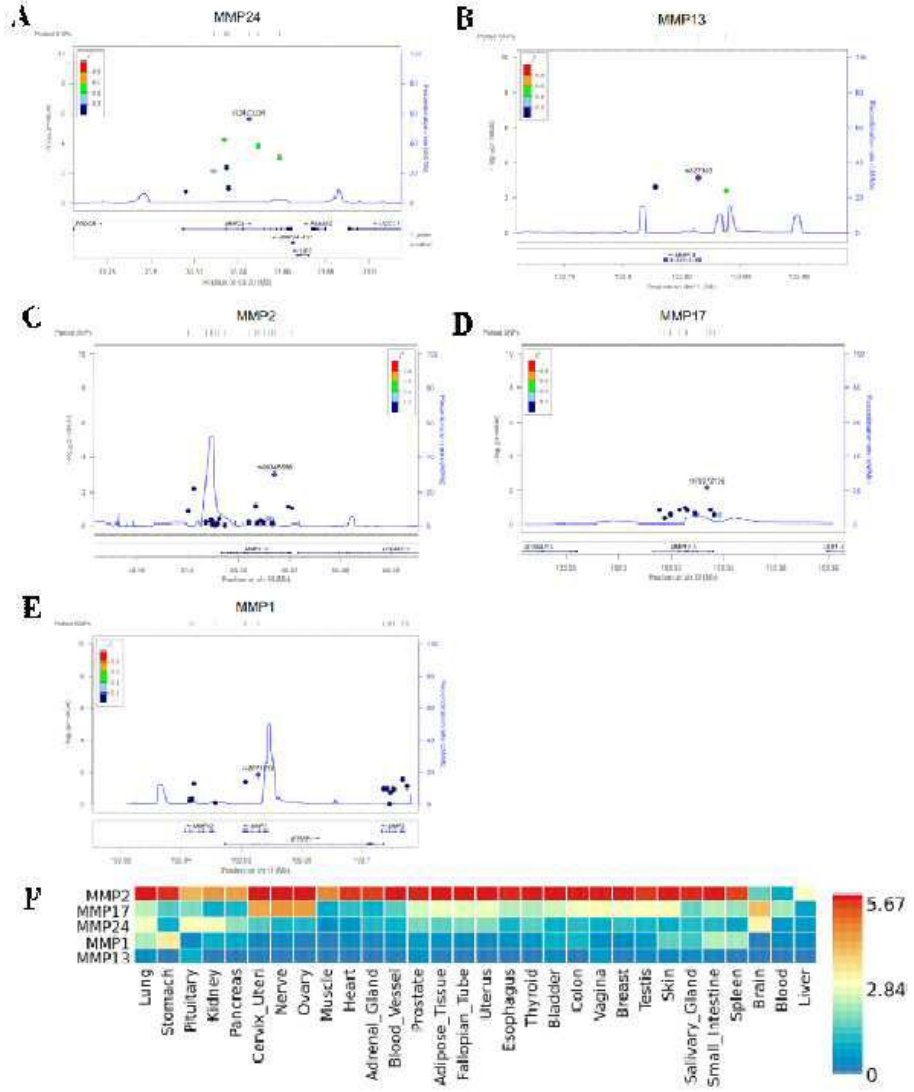
도면6


-log10 P-value
Chromosome
MMP24
MMP13
MMP1
MMP17
MMP2
MMP16
MMP26
MMP19
MMP15
MMP28
MMP9

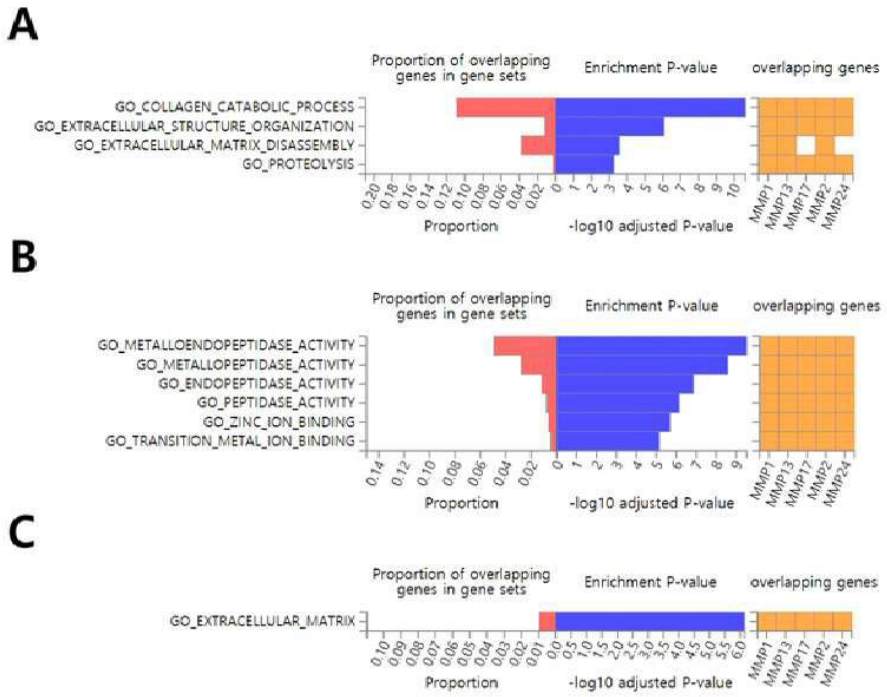
도면7


Observed -log10 P-value
Expected -log10 P-value
0
1
2
3
4
5
6
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

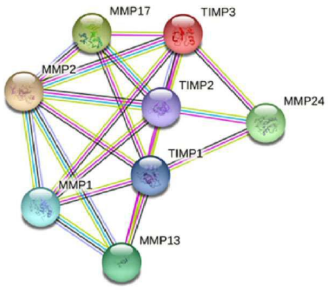
도면8


A
MMP24
B
MMP13
C
MMP2
D
MMP17
E
MMP1
F
MMP2
MMP17
MMP24
MMP1
MMP13
Lung
Stomach
Pituitary
Kidney
Pancreas
Cervix_Uter
Nerve
Ovary
Muscle
Heart
Adrenal_Gland
Blood_Vessel
Prostate
Adipose_Tissue
Fallopian_Tube
Uterus
Esophagus
Thyroid
Bladder
Colon
Vagina
Breast
Testis
Skin
Salivary_Gland
Small_Intestine
Spleen
Brain
Blood
Liver
5.67
2.84
0

도면9



도면10



서 열 목 록

<110> INDUSTRY ACADEMIC COOPERATION FOUNDATION, HALLYM UNIVERSITY

<120> A intracranial aneurysm diagnostic marker comprising a single nucleotide polymorphism in the matrix metalloproteinase 13 gene

<130> P212020

<160> 6

<170> KoPatentIn 3.0

<210> 1

<211> 12740

<212> DNA

<213> Homo sapiens

<400> 1

tctgttgaaa atatattttt attataaaca acataaacac ttcctaatac actttgcaaa 60

tatgcatttc attttctttg cctgagtttt tatcaacagt gtctctgagc acaatagttc 120

ttcccttgat ggccgatcat atattcaata agtgccaagc accctcccca agtatcaata 180

ggcactgtgg gaagtgctgg gggatttttt taaatgcaga gaaaaaaagg catccttcca 240

tatctcaagg gatcacaatc tagtagactg ctttaaaaaa aaatttggat gtctcttttt 300

taatatacat acttccactt tataagatat attttatttc atgctataca agtctttaga 360

aaatagccta tatatccaag acaaactgga caattaaata attagctagg taaaataggg 420

ttaaaatgca acttttagtc ttaaataatc aaataattta tgaaaaaggg atgccatggc 480

cttagactat tttgataaca attttaactt tgtaatgtgt gacttctgaa gaacctgaaa 540

tattttaata aaaagacatc atccttcttt atcctttata atatgcaaag gaatatttgt 600

cttatatgag agcagacttt gagtcattgc caaaagtgaa aattaaacag aattcaaagg 660

ccacatctac tattcttacc actgctcttt tgtctcctgt ctttaatgac tttgatattt 720

cgttagtgtc tgtgtgaaga agggcacata catctgaata tcctcagtgc aaaatgaaaa 780

ttatttattt ggaagttaca gtctattatt acagacattt tcctttggtt tatttctttt 840

atgtatggct ctcccaaatt tccattttca tactatttta tacttttag taagaatgat 900

ttattcatta tatgcatgta tttactttat gcccttgtaa aatttcctta taaatatatt 960

tttaaattat gctctcattg acagaccatg tgtcccattt gtggtgtggg aagtatcatc 1020

aaccattgct tttgtacagc aacaagaatc agatgctctt tagagatcct ccatttcatt 1080

actctaacat tcttcaatgt ggttccagcc acgcatagtc atatagatac tacatattca 1140

aagataactt actgaagctt gttcacagaa ccaagctttc tcctgatagc tcttcttccc 1200

ctaccccgca cttctggaag tattacccca aatgctcttc aggatttaaa taacaatttt 1260

taaaaagaca cttaacacca caaaatggaa tttgctggca tgacgcgaac aatacggtta 1320

ctccagatgc tgtattcaaa ctgtatgggt ccgttgaaaa aatagatata acctataaga 1380

aaaagcataa agacaatttc agagtttgaa aataataagg ttttaatcca gttaatatta 1440

atggttcaat ctcttagatc ccatttgtag gaatttcaaa ctttttcttt ttaaactcta 1500

ttagttcttt aatagttgtt cctttttaaa ttttttattt ttttattttt atttttttagt 1560

tgttctttta aggtgtgaat cctcctacac agttttgcta cttggattat cgatttttgt 1620

gtgcgtatat gagacagagt ctcactctgt catccaggct agagtgcagt ggcagtgatc 1680

ttggctcact gcaacctctg actcctaagt tcaagtgatt ctcctgcctc agcctcctga 1740

ctagctggga ttacaggtgc atgccaccac tcccagcttt tttttttttt tttttttttgg 1800

attttagta gagatggggt ttcgccatgt tggtcaggct ggtcttgaac tcctgacctc 1860
aagtgatccg cccacctcag cttcccaaag tgctcagatt gcaggcatga gccaccgcac 1920
ccagcctatt tgttataagt tatcaataat atctcttttc atagttaaag gaggtcggtg 1980
aggtcaggcc tgtaatctga gcactttggg aggccgaggc gggtgaatcg cttgaggcca 2040
ggagttcaag accagcctag ccaacatggc gaaactccat ctctactaaa aatgcaaaaa 2100
tcagccaggc atggtggcag gcacctgtaa tcccagctac tccagaggct gaggcaggag 2160
aatctcttga acccaggagg cggaggttgc agcgagctca gattgcgaca atgaactcca 2220

gcctgggtga cagagtgaga ctctgtcaaa aaaaaaaaaa aaggttgaga acataattaa 2280
agtaaaaagt cactatatag ttaaggtagt tatgagtctt tatatataaa ttacttccca 2340
ctccacagcc agtttgaatt ctggcaatat ccttgcattt tggcatcttt aagcatagat 2400
taaggtaatg ttgaaggttt agcatctcgt atttatagaa ttagcaatca ttaagtaaaa 2460
ttcttattta ttctaaccat gtttttttg gtttgctgtc tttgtaggta gtcacacagg 2520
aacaattaat tccttaactg ccatacattt attaacacat ctaggtccta tataaaaatg 2580
ctttgtacag aaaagagttt tgaggggcta cagaagctcc tctttaaagt cagtgcaatg 2640

taactactta ccatttttct catagacagc atctacttta tcaccaattc ctgggaagtc 2700
ttcttctatt agtctcggat agtctttatc cataatatgg ttagtatcat catatctatt 2760
taaagaaaaa aaactcctaa gtcagttata tttcttagaa agtaccttag aaataataca 2820
atagatacaa tggcatcttt caaaattgac aagaatttta aaattttcag tccatttaaa 2880
gcatctggta gctttttaaa attaggaatg agaatactga gtgtacagtg gatgcttgac 2940
gagtatttgt tgacttgccc aaaggccatc tttttaaaga gagttacaag ctcatttcca 3000
tttcacaaat gatatatctg ttataaaagt atgtcattag gaacatgttc ttaagcatga 3060

aataaatgca tgtttgtgca tagaacaatt agctgttgat ctctcactat gtgtaagaca 3120
acttgcactt acttgcactt acttagtgta cagtaatggc tacaaaaata tttaaggttt 3180
aatttctgta ctaaaaagtt gtaggggcat atggtaatac taaaggtagg aagggaagca 3240
tagctttccc cctccataac tgcatatttc aagctccttc tcccctgccc acagccacac 3300
cctaaccagg caaacatatc ctaaaattaa cctaggaaaa caggctccca taaggtgcca 3360
ctactgttac tgatgagagc tttttgaata tcttggaggt atgttccaac aagtaaaaga 3420
taagaaccac taaactacca ataaagctaa ggcatgtgta tacacataac tgcaacgcca 3480

ggcaatagaa gtcaagcatg agagagtaat catgaaatgt catgggagta caggaagccg 3540
ggcaataacc cctgaatggg gaaatcaaaa gaggctttgt agatgaaaag atatttcgaa 3600
tgaactttaa acaatggata gaatttcaag agggagagat gcagggaagg atcttctggc 3660

aaaaaaaaat tgcctaagag aagcttaagg ggatataaaa aacagcatat ctgagatgca 3720
gtaagcaatt aatcctgact gaacaatata cttcttatgg aggtagggaa tgaaatgcgg 3780
tacctggtag aagctggcag tgatttctag aatgttgtaa atgtcattgt cagaaataag 3840
tcaaaaggca aagatgatat atacaggtat gtgcgtgtgt gcgggtatgc atgtgtgtgt 3900

gtgagactgt gtgtatgtag agatggagga aagtttccac tttaaacatt tagaatgtag 3960
ggtactgatg agatattccc attcccaaat atagacttga atctggactc ttaaatgttg 4020
ttgatgtcaa tgctgcctaa atctttacat gtgttttaaa cattccagct tcactcgccc 4080
ttcagaaagt ttcactttat aaactatggg cctgtcagct tcctggtcaa cctctccagt 4140
ggatcagcta actctgcaga taggagggcc ttctctaaac caatcttttc ctgtactcat 4200
gctggtatcc tctttggcca cggaagatgg agctacatgt aaattaggag aaagtgacaa 4260
gaaagagaat acaggagaaa gattagaaag agaaggtgcc aattgggtgt ggtggctcac 4320

gactgtaatt ccaccatttt gggaggctga ggcaggtgga ttgctggaag ccaggagcct 4380
gaccagtgtg gcaaaacccc gtctctacta aaaatacaaa aattagttgg gtgtagtggc 4440
acatgcctgt aatcccagct tcttgggaga ctgagacaca agacatgaga attgcttcaa 4500
tacaggaggc agaggttaca gtgagccaag atcatgccac tacactcaag cctggataac 4560
agagcaagac tcagtctcaa aaaaaaagaa agaaagaggg ggtgccaaag atggttgaga 4620
ggcacaacaa atggctaaga aaagaatgtt tgagaagagg aatgtttgag tattgtgtgt 4680
gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gttcaagaga gcttgagact tttaaataca 4740

tgtctgaggg tctgcctgaa agagctgaca tggaatgaaa atgaatgact agtatgacat 4800
ttacaatagt gtgtaggaat aatgtatagt agtgacagat gtttgaggca ctttgcggct 4860
atgacacaga tgggtcagta cctactgctg ccatacctcc agacctggtt tcctgagaac 4920
aggagagtct tgcctgtatc ctcaaagtga acagctgcac ttatcttctt aacttctttt 4980
ggaagaccca gttcagatat ttttttggga taaccttcca gaatgtcata accattaaga 5040
gcccaaaatt ttctacctgg aatgagtgaa ataacagttc tattatccaa gggaatctat 5100
taggacttac atatatagac atgtgcttaa agacaggccc acttttacat tgagaaatat 5160

taataaaata tttagtattt acagtcatca tgtttatgaa acataaaaat ataagcattc 5220
ttgaaacaaa aaaataaaat ttaggatctt aacttcccaa ttatgaatat agctaatcaa 5280
atatcacagt taacttggca atttattagc tctgaatatt aagtttaaag ctaggtggaa 5340
aaaatacatt aaaatttgta catgaacaaa caacacagta ggaagagtaa attagactga 5400
aaactcaaac gttacaccag tttgaaagtt attttattct ctaaaaagta gctttattat 5460
aacattaata ttataaatag ttgaacaata caaacagtgt cagcaggaag ttatgtaatc 5520

attttaaatt atggtcataa agataaagta gcaatttgga aattggtcat aattttatgt 5580

caaggaaagc agtagaatat aaaaatgtat ctatactatg aaaataaata tgtaaaaatt 5640
tgtattacat aatgaaaatg agcaaatgaa ataaactgag ttttcagaag gaagaatgat 5700
tcataatatt ttttccactt tttcagaata tattggtata atgttttctc aataaacaat 5760
ttttaagcag aactaagaga aaaaactgac atgattcata tgaaagaaag agtggatatg 5820
acaagactaa taatgtagac cttttatatg ttcaatttgg agattcatgg tagaatggtt 5880
tcaggtactt tctggcttcc atttttttaa tacttggtct ttaatttagg tatattttca 5940
gtagtgatca ttttactgaa tctctagtgg catcaaattc agcactgcct cccctggtct 6000

tgtgtgagga ctcctctgtg gaaagcttac ctctgaagat gaagatgagg tcatgagaag 6060
ggtgctcata tgcagcatca atacggttgg gaagttctgg ccaaaatgat ttcgttaaaa 6120
acagctccgc atcaacctgc tgaggatgca ggcgccagaa gaatctaaca caaaagtaaa 6180
atggagtttt ctgtcacatt tttaaatgca atatttctat cctgctagtc acctctctcc 6240
acatcaacac agacaagtta gatacaacca atacctccca cctgtgaagg gaaaaaaaat 6300
tccattaccc tttaagcgcc agtgacaagt tgtgctatcc taactagcgt aaggtattca 6360
actttgaaat ctcacagagc ctgtgtgagg tctctgagtt ctttgcaggt tgaaacagtc 6420

ttgctcaatt ttgggtatct tgccaggggc aagcaaatgt gccaggcatt gtgttggcct 6480
ccagagagaa gatatgagtt atatgaacta tgtcttccat ggtctcatgg cctgggaaaa 6540
catacatgag aaacaccccc cgccccacga gtacaatgta attagagtta taacagaggg 6600
gataaacatg ctattggatc acagatgagg gatcaatgaa tgaatgaata aacaagcata 6660
ctggaacagt gtagaaatta aaattttata ttgcatttct gggttcattt catttcatcc 6720
ctcaaattcc aacatgtatc tagaatcaaa gtctttccat accaaaacga atccgtgtgg 6780
gcctaacgag taaccattta ttaatacagc agttactttc aatgcttgtc tgtttttaaa 6840

acaaatggga tttctagatt ttcttggtta tatttataag aaagactgta actagagatg 6900
taaaaggaag ctacacaatg tcaaggttat acgttgaaaa gtaaaacatt tatccatcag 6960
tcatttattt gatctgaaat atgtaaataa actgcctgcc catttttact gctaacttcg 7020
cctttggaag tctgacagga tgttttggca atatgcagaa gcagcttcac agatgtttgt 7080
cgcatacaga ctttatgaaa gaatctcaag agtacctgtc tttaaagatc attgtttctc 7140
ctcggagact ggtaatggca tcaagggata aggaagggtc acatttgtct ggcgtttttg 7200
gatgtttagg gttggggtct tcatctcctg gacctgtagt cgtgaaaggc aagagagaag 7260

ttatgagtgt gacattactg ataacctgag tcacagtaca acaggctgat ctgctgatgg 7320
acagtgggag agggtttctc caggaaaagc tgtaggtcct ggggctccta catcatggtg 7380
caagtaccgt gtaatcttac ttccttcctc ccctccactt ccatccttca agacctctcg 7440
agattcttta gaagtataaa cttaggtttc ttgaggttga tgataaaaag gagggaaaaa 7500
agcaacagaa tttgagaaat tttccctaca ccttgcacaa tctattgaca gattgctctt 7560
gaaaactaga ttccccttaa agccattttt gatggcgtat ttttcctgca aaagccctca 7620
tcttgcctcc tcctcttgac tattttacca gctgtcaaca tcagctaatg ttgcttcacc 7680

agtttgccat ggtgcagctg tatctcagct gtgacaaacc aggccttcag cacttatgtg 7740
acttgcgaca tgttggtaaa tttccaaccg tctcccatga gtgcagacag gatgggggtg 7800
cagtatataa taataataat cctacactgt ttttctttct gtctagtttt tattagttct 7860
ttcatcagtc cctgagtcat aacaatgtag aaataataat tgcaaaataa atttgcaaat 7920
tccctttgcc atcttcaagt tcatgcccat ggctatgcta gctactgtga tagattgaaa 7980
tctagagggt aaaagaatgg gttggatgct atgagcaagt tttctgagct atgtgctcta 8040
gaagtgttta cactgatgac tgtgtactgt ggctttaact ctataacagg tcatctataa 8100

ataaatactt gggttgaagt ttaataattt tgtaaatatt attgaaaact tgggggggga 8160
agtgaaaagg aagaattttt tttttacaat ggttactgta gatgttgatt aaaaactacc 8220
acaggccact tgagaggttt ctactgtgcc agattggact ggttgaaaga gtacttggaa 8280
tttctaaaag ccttaggatc tcactttcac caccaacctt gttactatag gtatgctact 8340
taaactctac agggctttgt ttcttcattg gttaaatggg gaacaaaata tttttctttc 8400
ttattgcaga ggtgtttgga ggatatcaaa tgcacattgc aatgtataaa gtacaaatac 8460
gaggagttat agtctgcaaa aatttggaga atgaaatctt ttctatttga gtgggtagtt 8520

agatttggaa agaaaaaatg ggtatgtttt cctaggtttt aacaacaaaa ccaagaaagc 8580
agcattggca gtggtgaatt caattagtag agtcatgtgt tgcttaacaa cagggataca 8640
ttctgaaaaa tgtgttgtta ggagattttg tcattgggtg aatattgtag cgtgtactta 8700
cacaaaccta gatatataac ctactacaga tataggctat atggaacagc ccatgtctcc 8760
taggctagaa acctgtccag catattactg ttctgagtac tgcaggcaca tgtaacacaa 8820
tggtgagtat taagtgtatc taaatataag aaaggtacat taaaaatact gtattacagt 8880
attatgaaac ctttgtcatg catgcgcttc actgtaccaa aacattatta tgcagcatga 8940

ctgcaattta gaaaataaat gcatgggttt cttctgcttt tataactgat catattctat 9000
ttctataacc gagttaccat agagagactg gatcccttgt acatcgtcat caggaagcat 9060
aaagtggctt ttgccggtgt aggtgtagat aggaaacatg agtgctccag ggtccttgga 9120

gtggtcaaga cctaaggagt ggccgaactc atgcgcagca acaagaaaca agttgtagcc 9180
tgtaagaaaa caaagaaaca atgagaaaaa aaggaattcc agtgaccagg taatgaaagg 9240
aatatcattt tatctatctg gtatgtttct ttcttggcta tatactttct tttctctttc 9300
ttcagtctcc tactttttaa aatcaatatt cagttgcttg aaggaagatt caaagtattg 9360

atttaatcct tataaataat ttagtgttct tttgattata gctgcctaca caattagaat 9420
gtcaacagat tactctcatg taattagtaa ttaataagga tttactgtga atacattaga 9480
actcctgatt tgcgcttagt attctcaggt caggtgaggg aagtggacag agattacctt 9540
aaacgaggag tggggaaatc acatagggtg gtacctagct agggtgctgc ccagaacatc 9600
ctccaatcag gaggcagcat ccaaagggat ggtaccaagg atggaggatg cactaaatag 9660
gttctctgtg tttactcttg ggagcaaatt ccaaccaaat gatttttcc agaatgtctc 9720
tagactctgg gttctggaga aacagaaact ctagacatga ttagccatct tttctctact 9780

ctagaagact actatgtcac tgtatttaca ttgttccctc taagcattct gtctgccaat 9840
tttgagcctc ctaatgcagg tatttcaaac tggaaaaaag aaaataaaaa actaaaagaa 9900
aaataagaag tgcaaggcaa taggattgtt gttacctgtg cttagattac ctgaaaacta 9960
gagccagtgg attccatttg accttaaata ccctctgcta ttgtagcctg tgcatgaaga 10020
aaatgaacca tcctagcaaa ccttgcagac catgagagca attagaggtc acataatctg 10080
agtgttaact aattcttcac tgaatgacta gtaatggata cccctcttgg ccttccactg 10140
aaaaatgaat ccagacttgt ttgtggatgt ttttattcac ttcaattaaa tatttcaaac 10200

ccagaattta taaaggaacc taaatttgag gcttactgtg attctacctt aacaatgctg 10260
gatattattt cttcttcctg aaaccattct tacctgtgaa atacccagaa cataaatatg 10320
taaattttca cttttccctc catagacaga ttggaggtat tagatttgag acactgatca 10380
tgtaaggccc agttttattc tttggagaaa taatcttttt tatttaaatg aaaaataaaa 10440
ctcagctgtc tatgaagaca acatctactt gtcataacta agtcctagtt agaacaatct 10500
atggcttttc ttcctaagtt aacattccca attattttga gacttgaaga tttaactact 10560
aactgatctc attcaacatt cagtttcatt ttgctcttgt aaatacaacc atcatcttgt 10620

ctgtttcatg ctgtttggag taagagtctg tctaggtttt taagaatagt gtttacaata 10680
tatagacaag actgaatcaa gcaactcata taatgggaaa cagaaatttg attgaacata 10740
tttctaccaa gcaaaaaact atttcaattc acatctaatt catgtatgac ccatagatcc 10800
agccctggta gtttttagtt agatatttct aataaactgt acttcctcaa catgccaaaa 10860
ggtaatttgc agtaataaca tgttggctga cagacataaa cattttggtc ctacttcaaa 10920
gttcattagt atccctcaag ccacagtgac agattaagta gattttgttc atttttaat 10980

gttgatatca atttgtagaa atgtcacttg tagagacact aatacatcta agataaactt 11040

tatttaagaa aatatagtca aataccaaat atatttaact tttatgtgtt ttaacttcat 11100
actagtgtgt tcactgtcta atctaataac acataaatga tatctttata agaaacatta 11160
cctttggaac tacttgtcca ggtttcatca tcatcaaaat gggcatctcc tccataattt 11220
ggcccaggag gaaaagcatg agccagcagg ccagagggcc catcaaatgg gtagaagtcg 11280
ccatgctcta cacacaaaag caagggttag gagtcttatc acatccagga gtacttagca 11340
caggtgtttg gtaaattagt gttgaattct ttaaaaagta atgaatgact gaaataaata 11400
attttaaaat ggaaccaaga attagaatat aaaaatgttt gtaaaataaa tgtgcatcat 11460

taccttaat tccaaaagag atcatgatgt cagcaatgcc atcgtgaagt ctggtaaaat 11520
tcagaggagt tacatcggac caaactttga aggctttttt gaatgccttt tcgacttcag 11580
aatgagtcat atcaggggtg taattcacaa ttctatttaa cagagaaagt ttcaatgaac 11640
tttttggaaa gtgacagcta ttttagaata cattttcttg gtatattgct tttcaattca 11700
aaatgcaata atttctttga aagttctact agaggtttat tgtttcaaat aatgtattaa 11760
ttatgatgcc atacaagaat aaaaagtcaa aaattcagcc acatttgagt atctccaaat 11820
atctgaccat gtggcatctt aaatttaaaa tattccaatg tttaatataa tttattttgc 11880

atgctaaaag tttagattac attgataaag ctcaaataaa tgaaatagct ctgtaaagtt 11940
tctagaaaga gtggaactct tcatcttgag cactgttgta ataaatacac ctaatgtact 12000
atgtttgcaa attttcttta acctactttt atttctcttt tgatttcatc tctaaaagaa 12060
atactttgag atgcccacta aaaaataatg ctgagttaat gtaacaataa tagatgttat 12120
gaggtattct cggcaaccat attaaagcta ttctggaatt taactgccaa ttaaataata 12180
aggcctactt aatattagac atttaatact acaagaaaag gctaacaaat atgaaagaaa 12240
gatagcctat gatttacctg taggttaaat tcattttgga ccatttaaga gttcgaggga 12300

aaacattgta ttcacccaca tcaggaaccc cgcatcttgg ctttttcatg acatctaagg 12360
tgttatcgtc aagtttgcca gtcacctcta agccgaagaa agactgcatt tctcggagcc 12420
tctcagtcat ggagcttgct gcattctcct tcaggattcc cgcgagattt gtaggatggt 12480
agtatgatct caggtagcgc tagaaaagac accaaaatga actgcgttta aaagagaagg 12540
acatttctga gatgtaccaa ccgcatcatc aggattggca agatactcta cctctgcaaa 12600
ctggaggtct tcctcagaca aatcatcttc atcaccacca ctgggaaggg gcagggcccg 12660
acaatgagtc cagctcaaga agaggaaggc agccaggacc cctggatgca tcttgaatgg 12720

tgatgcctgg ggactgttgt 12740

<210> 2

<211> 30

<212> DNA

<213> Homo sapiens

<220><221> variation

<222> (15)

<223> r = g or a

<400> 2

ttaggggtag ccctrctggc catattcaca 30

<210> 3

<211> 30

<212> DNA

<213> Homo sapiens

<220><221> variation

<222> (15)

<223> m = a or c

<400> 3

ggctcagcct aggcmaagtc atgaacctgt 30

<210> 4

<211> 30

<212> DNA

<213> Homo sapiens

<220><221> variation

<222> (15)

<223> m = c, a or t

<400> 4

tttatagtgc cttgmttttt tgggggggtag 30

<210> 5

<211> 30

<212> DNA

<213> Homo sapiens

<220><221> variation

<222> (15)

<223> r = a, g or t

<400> 5

ttcacactga attcrttgtt ctgcaacttg 30

<210> 6

<211> 30

<212> DNA

<213> Homo sapiens

<220><221> variation

<222> (15)

<223> r = a or g

<400> 6

agaatttaat gccartcctc cctccaaaat 30