



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0062224
(43) 공개일자 2020년06월03일

(51) 국제특허분류(Int. Cl.)
C12N 15/63 (2006.01) C07K 14/705 (2006.01)
C12Q 1/68 (2018.01)
(52) CPC특허분류
C12N 15/63 (2013.01)
C07K 14/705 (2013.01)
(21) 출원번호 10-2020-7009932
(22) 출원일자(국제) 2018년09월06일
심사청구일자 없음
(85) 번역문제출일자 2020년04월06일
(86) 국제출원번호 PCT/US2018/049674
(87) 국제공개번호 WO 2019/051033
국제공개일자 2019년03월14일
(30) 우선권주장
62/555,440 2017년09월07일 미국(US)

(71) 출원인
리제너론 파마슈티칼스 인코포레이티드
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리
버 로드 777
(72) 발명자
테슬로비취 도스탈, 탄야
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리
버 로드 777
배크맨, 조슈아
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리
버 로드 777
(74) 대리인
김진환, 박지하, 김민철

전체 청구항 수 : 총 124 항

(54) 발명의 명칭 SLC14A1 변이체 및 그 용도

(57) 요약

본 개시내용은 관상 동맥 질환(CAD)으로부터의 보호와 연관된 변이체 인간 용질 운반체 패밀리 14 구성원 1(SLC14A1) 단백질을 암호화하는 변이를 포함하는, cDNA를 포함하는 핵산 분자를 제공한다. 본 개시내용은 또한 이러한 변이의 확인에 기초하여 응고 병태를 발생시킬 위험에 있는 대상체를 분류하는 방법을 제공한다.

(52) CPC특허분류
C12Q 1/68 (2018.05)

명세서

청구범위

청구항 1

인간 용질 운반체 패밀리 14 구성원 1(Solute Carrier Family 14 Member 1: SLC14A1) 단백질을 암호화(encoding)하는 cDNA로서,

서열번호 9와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는, cDNA.

청구항 2

제1항에 있어서, 상기 핵산 서열은 서열번호 9를 포함하는, cDNA.

청구항 3

인간 용질 운반체 패밀리 14 구성원 1(SLC14A1) 단백질을 암호화하는 cDNA로서,

서열번호 10과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 상기 핵산 서열은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는, cDNA.

청구항 4

제3항에 있어서, 상기 핵산 서열은 서열번호 10을 포함하는, cDNA.

청구항 5

제1항 내지 제4항 중 어느 한 항에 따른 cDNA를 포함하는 벡터.

청구항 6

제5항에 있어서, 외인성 도너 서열(exogenous donor sequence)을 추가로 포함하는, 벡터.

청구항 7

제5항 또는 제6항에 있어서, 상기 벡터는 플라스미드를 포함하는, 벡터.

청구항 8

제5항 또는 제6항에 있어서, 상기 벡터는 바이러스를 포함하는, 벡터.

청구항 9

조성물로서,

제1항 내지 제4항 중 어느 한 항에 따른 cDNA 및 담체를 포함하는 조성물.

청구항 10

조성물로서,

제5항 내지 제8항 중 어느 한 항에 따른 벡터 및 담체를 포함하는 조성물.

청구항 11

숙주 세포로서,

제1항 내지 제4항 중 어느 한 항에 따른 cDNA를 포함하는 숙주 세포.

청구항 12

숙주 세포로서,

제5항 내지 제8항 중 어느 한 항에 따른 벡터를 포함하는 숙주 세포.

청구항 13

제11항 또는 제12항에 있어서, 상기 cDNA가 상기 숙주 세포에서 활성인 프로모터에 작동 가능하게 연결된, 숙주 세포.

청구항 14

제13항에 있어서, 상기 프로모터가 유도성 프로모터인, 숙주 세포.

청구항 15

제11항 내지 제14항 중 어느 한 항에 있어서, 상기 숙주 세포가 박테리아 세포, 효모 세포 또는 곤충 세포인, 숙주 세포.

청구항 16

제11항 내지 제14항 중 어느 한 항에 있어서, 상기 숙주 세포가 포유류 세포인, 숙주 세포.

청구항 17

단리된 변경-특이적(alteration-specific) 프로브 또는 프라이머로서,

적어도 약 15개의 뉴클레오타이드를 포함하고, SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화되, 상기 변경-특이적 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 SLC14A1 암호화 핵산 서열의 일부에 상보적인 핵산 서열 또는 그의 보체를 포함하는, 단리된 변경-특이적 프로브 또는 프라이머.

청구항 18

단리된 변경-특이적 프로브 또는 프라이머로서,

적어도 약 15개의 뉴클레오타이드를 포함하고, SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화되, 상기 변경-특이적 프로브 또는 프라이머는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 SLC14A1 암호화 핵산 서열의 일부에 상보적인 핵산 서열 또는 그의 보체를 포함하는, 단리된 변경-특이적 프로브 또는 프라이머.

청구항 19

단리된 변경-특이적 프로브 또는 프라이머로서,

서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 갖는 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적이고/이거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 갖는 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적인 핵산 서열을 포함하되, 상기 변경-특이적 프로브 또는 프라이머는 서열번호 2에 따른 6963번 내지 6965번 위치 또는 그의 보체; 서열번호 5에 따른 226번 내지 228번 위치 또는 그의 보체; 서열번호 6에 따른 394번 내지 396번 위치 또는 그의 보체; 서열번호 9에 따른 226번 내지 228번 위치 또는 그의 보체; 서열번호 10에 따른 394번 내지 396번 위치 또는 그의 보체에 상응하는 위치를 포함하는 핵산서열의 일부에 상보적인 핵산 서열을 포함하는, 단리된 변경-특이적 프로브 또는 프라이머.

청구항 20

인간 대상체에서 인간 대상체가 SLC14A1 변이체 핵산 분자를 보유하는지의 여부를 결정하는 방법으로서,

상기 대상체로부터 수득된 샘플을 분석(assay)하여, 상기 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부 및/또는 상기 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는

SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 21

제20항에 있어서, 상기 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우 및/또는 상기 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우, 상기 인간 대상체는 응고 병태 또는 관상 동맥 질환(coronary artery disease: CAD)을 발생시킬 위험(risk)이 감소된 것으로 분류되는, 방법.

청구항 22

제20항 또는 제21항에 있어서, 상기 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우 및/또는 상기 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우, 상기 인간 대상체는 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류되는, 방법.

청구항 23

제21항 또는 제22항에 있어서, 상기 응고 병태는 혈전증, 폐 색전증, 심근경색(myocardial infarction: MI), 정맥 혈전색전증(venous thromboembolism: VTE), 심부 정맥 혈전증(deep vein thrombosis: DVT), 뇌동맥류 및 뇌졸중으로부터 선택되는, 방법.

청구항 24

제20항 내지 제23항 중 어느 한 항에 있어서, 상기 분석법은,

상기 샘플에서 SLC14A1 게놈 핵산 서열의 일부를 서열분석(sequencing)하되, 상기 서열분석된 일부는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치를 포함하는, 서열분석하는 단계;

상기 샘플에서 SLC14A1 mRNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계;

상기 샘플에서 SLC14A1 mRNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계;

상기 샘플에서 mRNA 핵산 분자로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 및/또는

상기 샘플에서 mRNA 핵산 분자로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계를 포함하는, 방법.

청구항 25

제20항 내지 제23항 중 어느 한 항에 있어서, 상기 분석은,

a) 상기 샘플을, i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 서열의 위치에 가장 가까운(proximate) SLC14A1 게놈 핵산 서열의 일부; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA의 위치에 가장 가까운 SLC14A1 mRNA 핵산 서열의 일부; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA의 위치에 가장 가까운 mRNA로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부에 혼성화하는 프라이머와 접촉시키는 단계;

b) 적어도 i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 핵산 서열의 위치; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA 핵산 서열의 위치; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA 핵산 서열의 위치에 걸쳐 상기 프라이머를 연장시키는 단계; 및

c) 상기 프라이머의 연장 산물이, i) 서열번호 2에 따른 SLC14A1 게놈 핵산 서열의 6963번 내지 6965번 위치에 상응하거나; ii) 서열번호 5에 따른 SLC14A1 mRNA 핵산 서열의 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 SLC14A1 mRNA 핵산 서열의 394번 내지 396번 위치에 상응하거나; 또는 iii) 서열번호 9에 따른 SLC14A1 cDNA 핵산 서열의 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 SLC14A1 cDNA 핵산 서열의 394번 내지 396번 위치에 상응하는 위치에서, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 26

제20항 내지 제23항 중 어느 한 항에 있어서, 상기 분석은,

상기 샘플을, 엄격한 조건하에 SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열에 특이적으로 혼성화하고 상응하는 야생형 SLC14A1 핵산 서열에 혼성화하지 않는 프라이머 또는 프로브와 접촉시키되, 상기 SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는 것인 단계, 및

혼성화가 발생하였는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 27

제20항 내지 제26항 중 어느 한 항에 있어서, 상기 방법은 시험관내 방법인, 방법.

청구항 28

인간 대상체가 SLC14A1 Val76Ile 단백질 및/또는 SLC14A1 Val132Ile 단백질을 보유하는지의 여부를 결정하는 방법으로서,

상기 인간 대상체로부터 수득된 샘플에 분석법을 수행하여서, 상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 29

제28항에 있어서, 상기 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질이 확인되는 경우 및/또는 상기 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질이 확인되는 경우, 상기 인간 대상체는 응고 병태 또는 관상 동맥 질환(CAD)을 발생시킬 위험이 감소된 것으로 분류되는, 방법.

청구항 30

제28항 또는 제29항에 있어서, 상기 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이 확인되는 경우 및/또는 상기 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이 확인되는 경우, 상기 인간 대상체는 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류되는, 방법.

청구항 31

제29항 또는 제30항에 있어서, 상기 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류 및 뇌졸중으로부터 선택되는, 방법.

청구항 32

제28항 내지 제31항 중 어느 한 항에 있어서, 효소-결합 면역흡착 분석법(enzyme-linked immunosorbent assay: ELISA)은 상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에

서 이소류신을 포함하는지의 여부를 결정하기 위해 이용되는, 방법.

청구항 33

제28항 내지 제32항 중 어느 한 항에 있어서, 상기 방법은 시험관내 방법인, 방법.

청구항 34

응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성(susceptibility)을 결정하는 방법으로서,

a) 상기 인간 대상체로부터 수득된 샘플을 분석하여, 상기 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부 및/또는 상기 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계; 및

b) 상기 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 경우 및/또는 상기 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 경우, 상기 인간 대상체를 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로 분류하거나, 또는

상기 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 경우 및/또는 상기 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 경우, 상기 인간 대상체를 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류하는 단계를 포함하는, 방법.

청구항 35

제34항에 있어서, 상기 분석은,

상기 샘플에서 SLC14A1 게놈 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치를 포함하는 것인 단계;

상기 샘플에서 SLC14A1 mRNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계;

상기 샘플에서 SLC14A1 mRNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계;

상기 샘플에서 mRNA 핵산 분자로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 및/또는

상기 샘플에서 mRNA 핵산 분자로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부를 서열분석하되, 상기 서열분석된 일부는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계를 포함하는, 방법.

청구항 36

제34항에 있어서, 상기 분석은,

a) 상기 샘플을, i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 게놈 핵산 서열의 일부; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA 핵산의 위치에 가장 가까운 SLC14A1 mRNA 핵산 서열의 일부; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA의 위치에 가장 가까운 mRNA로부터 수득된 SLC14A1 cDNA 핵산 서열의 일부에 혼성화하는 프라이머와 접촉시키는 단계;

b) 적어도 i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 핵산 서열의 위치; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA 핵산 서열의 위치; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA 핵산 서열의 위치에 걸쳐 상기 프라이머를

연장시키는 단계; 및

c) 상기 프라이머의 연장 산물이, i) 서열번호 2에 따른 SLC14A1 게놈 핵산 서열의 6963번 내지 6965번 위치에 상응하거나; ii) 서열번호 5에 따른 SLC14A1 mRNA 핵산 서열의 226번 내지 228번 위치에 상응하거나 또는 서열번호 6에 따른 SLC14A1 mRNA 핵산 서열의 394번 내지 396번 위치에 상응하거나; 또는 iii) 서열번호 9에 따른 SLC14A1 cDNA 핵산 서열의 226번 내지 228번 위치에 상응하거나 또는 서열번호 10에 따른 SLC14A1 cDNA 핵산 서열의 394번 내지 396번 위치에 상응하는 위치에서, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 37

제34항에 있어서, 상기 분석은,

상기 샘플을, 엄격한 조건하에 SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열에 특이적으로 혼성화하고 상응하는 야생형 SLC14A1 핵산 서열에 혼성화하지 않는 프라이머 또는 프로브와 접촉시키는 단계로서, 상기 SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는, 상기 접촉시키는 단계, 및

혼성화가 발생하였는지의 여부를 결정하는 단계를 포함하는, 방법.

청구항 38

제34항 내지 제37항 중 어느 한 항에 있어서, 상기 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류 및 뇌졸중으로부터 선택되는, 방법.

청구항 39

제34항 내지 제38항 중 어느 한 항에 있어서, 응고 병태를 발생시킬 위험이 증가된 대상체에 대해서, 상기 응고 병태를 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함하는, 방법.

청구항 40

제34항 내지 제39항 중 어느 한 항에 있어서, CAD를 발생시킬 위험이 증가된 대상체에 대해서, CAD를 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함하는, 방법.

청구항 41

제34항 내지 제40항 중 어느 한 항에 있어서, 상기 방법은 시험관내 방법인, 방법.

청구항 42

응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법으로서,

a) 상기 인간 대상체로부터 수득된 샘플을 분석하여, 상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하는 단계; 및

b) 상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 경우 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 경우, 상기 인간 대상체를 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로 분류하거나, 또는

상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 경우 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 경우, 상기 인간 대상체를 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류하는 단계를 포함하는, 방법.

청구항 43

제42항에 있어서, 상기 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류 및 뇌졸중으로부터 선택되는, 방법.

청구항 44

제42항 또는 제43항에 있어서, 효소-결합 면역흡착 분석법(ELISA)은 상기 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 상기 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하기 위해 사용되는, 방법.

청구항 45

제42항 내지 제44항 중 어느 한 항에 있어서, 상기 방법은 시험관내 방법인, 방법.

청구항 46

제42항 내지 제45항 중 어느 한 항에 있어서, 응고 병태를 발생시킬 위험이 증가된 대상체에 대해서, 상기 응고 병태를 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함하는, 방법.

청구항 47

제42항 내지 제45항 중 어느 한 항에 있어서, CAD를 발생시킬 위험이 증가된 대상체에 대해서, CAD를 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함하는, 방법.

청구항 48

세포를 변형시키는 방법으로서,

상기 세포로 발현 벡터를 도입하는 단계를 포함하고, 상기 발현 벡터는 이소류신을 암호화하는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 코돈을 포함하는 뉴클레오타이드 서열을 포함하는 재조합 SLC14A1 유전자를 포함하는, 세포를 변형시키는 방법.

청구항 49

제48항에 있어서, 상기 방법은 시험관내 방법인, 세포를 변형시키는 방법.

청구항 50

세포를 변형시키는 방법으로서,

상기 세포로 발현 벡터를 도입하는 단계를 포함하고, 상기 발현 벡터는 서열번호 13과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자를 포함하는, 세포를 변형시키는 방법.

청구항 51

제50항에 있어서, 상기 방법은 시험관내 방법인, 세포를 변형시키는 방법.

청구항 52

세포를 변형시키는 방법으로서,

상기 세포로 발현 벡터를 도입하는 단계를 포함하고, 상기 발현 벡터는 서열번호 14와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자를 포함하는, 세포를 변형시키는 방법.

청구항 53

제52항에 있어서, 상기 방법은 시험관내 방법인, 세포를 변형시키는 방법.

청구항 54

세포를 변형시키는 방법으로서, 상기 세포로 SLC14A1 폴리펩타이드 또는 그의 단편을 도입하는 단계를 포함하고, 상기 SLC14A1 폴리펩타이드는 서열번호 13과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는, 세포를 변형시키는 방법.

청구항 55

제54항에 있어서, 상기 방법은 시험관내 방법인, 세포를 변형시키는 방법.

청구항 56

세포를 변형시키는 방법으로서, 상기 세포로 SLC14A1 폴리펩타이드 또는 그의 단편을 도입하는 단계를 포함하고, 상기 SLC14A1 폴리펩타이드는 서열번호 14와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는, 세포를 변형시키는 방법.

청구항 57

제56항에 있어서, 상기 방법은 시험관내 방법인, 세포를 변형시키는 방법.

청구항 58

단리된 핵산 분자로서,

서열번호 2와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 상기 핵산 서열은 이소류신을 암호화하는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 코돈을 포함하는, 단리된 핵산 분자.

청구항 59

제58항에 있어서, 상기 핵산 서열은 이소류신을 암호화하는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 코돈을 포함하는, 단리된 핵산 분자.

청구항 60

제58항 또는 제59항에 있어서, 상기 핵산 서열은 서열번호 2를 포함하는, 단리된 핵산 분자.

청구항 61

제58항 내지 제60항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 벡터.

청구항 62

제61항에 있어서, 외인성 도너 서열을 추가로 포함하는, 벡터.

청구항 63

제61항 또는 제62항에 있어서, 상기 벡터는 플라스미드를 포함하는, 벡터.

청구항 64

제61항 또는 제62항에 있어서, 상기 벡터는 바이러스를 포함하는, 벡터.

청구항 65

제58항 내지 제60항 중 어느 한 항에 따른 단리된 핵산 분자 및 담체를 포함하는 조성물.

청구항 66

제61항 내지 제64항 중 어느 한 항에 따른 벡터 및 담체를 포함하는 조성물.

청구항 67

제58항 내지 제60항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 숙주 세포.

청구항 68

제61항 내지 제64항 중 어느 한 항에 따른 벡터를 포함하는 숙주 세포.

청구항 69

제67항 또는 제68항에 있어서, 상기 단리된 핵산 분자는 상기 숙주 세포에서 활성화된 프로모터에 작동 가능하게 연결된, 숙주 세포.

청구항 70

제69항에 있어서, 상기 프로모터는 유도성 프로모터인, 숙주 세포.

청구항 71

제67항 내지 제70항 중 어느 한 항에 있어서, 상기 숙주 세포는 박테리아 세포, 효모 세포 또는 곤충 세포인, 숙주 세포.

청구항 72

제67항 내지 제70항 중 어느 한 항에 있어서, 상기 숙주 세포는 포유류 세포인, 숙주 세포.

청구항 73

단리된 핵산 분자로서,

서열번호 5와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는, 단리된 핵산 분자.

청구항 74

제73항에 있어서, 상기 핵산 서열은 서열번호 5의 서열을 포함하는, 단리된 핵산 분자.

청구항 75

단리된 핵산 분자로서,

서열번호 6과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 상기 핵산 서열은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는, 단리된 핵산 분자.

청구항 76

제75항에 있어서, 상기 핵산 서열은 서열번호 6의 서열을 포함하는, 단리된 핵산 분자.

청구항 77

벡터로서,

제73항 내지 제76항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 벡터.

청구항 78

제77항에 있어서, 외인성 도너 서열을 추가로 포함하는, 벡터.

청구항 79

제77항 또는 제78항에 있어서, 상기 벡터는 플라스미드를 포함하는, 벡터.

청구항 80

제77항 또는 제78항에 있어서, 상기 벡터는 바이러스를 포함하는, 벡터.

청구항 81

조성물로서,

제73항 내지 제76항 중 어느 한 항에 따른 단리된 핵산 분자 및 담체를 포함하는 조성물.

청구항 82

조성물로서,

제77항 내지 제80항 중 어느 한 항에 따른 벡터 및 담체를 포함하는 조성물.

청구항 83

숙주 세포로서,

제73항 내지 제76항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 숙주 세포.

청구항 84

숙주 세포로서,

제77항 내지 제80항 중 어느 한 항에 따른 벡터를 포함하는 숙주 세포.

청구항 85

제83항 또는 제84항에 있어서, 상기 단리된 핵산 분자는 상기 숙주 세포에서 활성인 프로모터에 작동 가능하게 연결된, 숙주 세포.

청구항 86

제85항에 있어서, 상기 프로모터는 유도성 프로모터인, 숙주 세포.

청구항 87

제83항 내지 제86항 중 어느 한 항에 있어서, 상기 숙주 세포는 박테리아 세포, 효모 세포 또는 곤충 세포인, 숙주 세포.

청구항 88

제83항 내지 제86항 중 어느 한 항에 있어서, 상기 숙주 세포는 포유류 세포인, 숙주 세포.

청구항 89

단리된 핵산 분자로서,

서열번호 13과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 폴리펩타이드를 암호화하는 핵산 서열 또는 그의 보체를 포함하되, 상기 폴리펩타이드는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는, 단리된 핵산 분자.

청구항 90

제89항에 있어서, 상기 핵산 서열은 서열번호 13에 따른 폴리펩타이드 서열을 암호화하는, 단리된 핵산 분자.

청구항 91

단리된 핵산 분자로서,

서열번호 14와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 폴리펩타이드를 암호화하는 핵산 서열 또는 그의 보체를 포함하되, 상기 폴리펩타이드는 서열번호

14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는, 단리된 핵산 분자.

청구항 92

제91항에 있어서, 상기 핵산 서열은 서열번호 14에 따른 폴리펩타이드 서열을 암호화하는, 단리된 핵산 분자.

청구항 93

백터로서,

제89항 내지 제92항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 백터.

청구항 94

제93항에 있어서, 외인성 도너 서열을 추가로 포함하는 백터.

청구항 95

제93항 또는 제94항에 있어서, 상기 백터는 플라스미드를 포함하는 백터.

청구항 96

제93항 또는 제94항에 있어서, 상기 백터는 바이러스를 포함하는 백터.

청구항 97

조성물로서,

제89항 내지 제92항 중 어느 한 항에 따른 단리된 핵산 분자 및 담체를 포함하는 조성물.

청구항 98

조성물로서,

제93항 내지 제96항 중 어느 한 항에 따른 백터 및 담체를 포함하는 조성물.

청구항 99

숙주 세포로서,

제89항 내지 제92항 중 어느 한 항에 따른 단리된 핵산 분자를 포함하는 숙주 세포.

청구항 100

숙주 세포로서,

제93항 내지 제96항 중 어느 한 항에 따른 백터를 포함하는 숙주 세포.

청구항 101

제99항 또는 제100항에 있어서, 상기 단리된 핵산 분자는 상기 숙주 세포에서 활성인 프로모터에 작동 가능하게 연결된, 숙주 세포.

청구항 102

제101항에 있어서, 상기 프로모터는 유도성 프로모터인, 숙주 세포.

청구항 103

제99항 내지 제102항 중 어느 한 항에 있어서, 상기 숙주 세포는 박테리아 세포, 효모 세포 또는 곤충 세포인, 숙주 세포.

청구항 104

제99항 내지 제102항 중 어느 한 항에 있어서, 상기 숙주 세포는 포유류 세포인, 숙주 세포.

청구항 105

단리된 프로브 또는 프라이머로서,

적어도 약 15개의 뉴클레오타이드를 포함하는 핵산 서열을 포함하되, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 갖는 인간 SLC14A1 단백질을 암호화하는 핵산 서열을 갖는 핵산 분자에 특이적으로 혼성화하고/하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 갖는 인간 SLC14A1 단백질을 암호화하는 핵산 서열을 갖는 핵산 분자에 특이적으로 혼성화하거나, 또는 이들 핵산 분자 중 적어도 하나의 보체에 특이적으로 혼성화하는, 단리된 프로브 또는 프라이머.

청구항 106

제105항에 있어서, 상기 프로브 또는 프라이머는 DNA를 포함하는, 프로브 또는 프라이머.

청구항 107

제105항에 있어서, 상기 프로브 또는 프라이머는 RNA를 포함하는, 프로브 또는 프라이머.

청구항 108

제105항 내지 제107항 중 어느 한 항에 있어서, 상기 프로브 또는 프라이머는 엄격한 조건하에 상기 SLC14A1 단백질을 암호화하는 핵산 서열 또는 이의 보체에 특이적으로 혼성화하는, 프로브 또는 프라이머.

청구항 109

제105항 내지 제108항 중 어느 한 항에 있어서, 상기 프로브 또는 프라이머가 표지를 포함하는, 프로브 또는 프라이머.

청구항 110

제109항에 있어서, 상기 표지가 형광 표지, 방사선표지 또는 바이오틴인, 프로브 또는 프라이머.

청구항 111

지지체로서,

제105항 내지 제110항 중 어느 한 항에 따른 프로브가 부착된 기질을 포함하는 지지체.

청구항 112

제111항에 있어서, 상기 지지체는 마이크로어레이인, 지지체.

청구항 113

응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성을 결정하기 위한, 제105항 내지 제110항 중 어느 한 항에 따른 단리된 프로브 또는 프라이머, 또는 제17항 내지 제19항 중 어느 한 항에 따른 단리된 변형-특이적 프로브 또는 프라이머의 용도.

청구항 114

단리된 폴리펩타이드로서,

서열번호 13의 아미노산 서열을 갖는 SLC14A1 변이체 폴리펩타이드와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 아미노산 서열을 포함하되, 상기 폴리펩타이드는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는, 단리된 폴리펩타이드.

청구항 115

제114항에 있어서, 상기 SLC14A1 변이체 폴리펩타이드는 서열번호 13의 아미노산 서열을 포함하는, 폴리펩타이드.

청구항 116

단리된 폴리펩타이드로서,

서열번호 14의 아미노산 서열을 갖는 SLC14A1 변이체 폴리펩타이드와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 아미노산 서열을 포함하되, 상기 폴리펩타이드는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는, 단리된 폴리펩타이드.

청구항 117

제116항에 있어서, 상기 SLC14A1 변이체 폴리펩타이드는 서열번호 14의 아미노산 서열을 포함하는, 폴리펩타이드.

청구항 118

제114항 내지 제117항 중 어느 한 항에 있어서, 상기 폴리펩타이드는 이종성 펩타이드(heterologous)에 추가로 융합된, 폴리펩타이드.

청구항 119

제118항에 있어서, 상기 이종성 분자는 면역글로불린 Fc 도메인, 펩타이드 정제 태그, 형광 단백질 또는 형질도입 도메인을 포함하는, 폴리펩타이드.

청구항 120

제114항 내지 제117항 중 어느 한 항에 있어서, 상기 폴리펩타이드는 표지에 추가로 연결된, 폴리펩타이드.

청구항 121

제120항에 있어서, 상기 표지는 폴리에틸렌 글리콜, 폴리시알산 또는 글리콜산을 포함하는, 폴리펩타이드.

청구항 122

제120항에 있어서, 상기 표지는 검출 가능한 형광 표지 또는 방사선표지를 포함하는, 폴리펩타이드.

청구항 123

조성물로서,

제114항 내지 제122항 중 어느 한 항에 따른 폴리펩타이드 및 담체 또는 부형제를 포함하는 조성물.

청구항 124

숙주 세포로서,

제114항 내지 제122항 중 어느 한 항에 따른 폴리펩타이드를 발현하는 숙주 세포.

발명의 설명

기술 분야

[0001] 서열 목록에 대한 참조

[0002] 본 출원은 101 킬로바이트의 크기로 2018년 9월 6일에 생성된 18923800902SEQ 명칭의 텍스트 파일로 전자로 제출된 서열 목록을 포함한다. 서열 목록은 본원에 참고로 인용된다.

[0003] 기술분야

[0004] 본 개시내용은 일반적으로 유전학의 분야에 관한 것이다. 더 특히, 본 개시내용은, 예를 들어 관상 동맥 질환(coronary artery disease: CAD)으로부터의 보호와 연관된 용질 운반체 패밀리 14 구성원 1(Solute Carrier Family 14 Member 1: SLC14A1)에서의 유전자 변경 및 폴리펩타이드 변이체에 관한 것이다.

배경 기술

- [0005] 다양한 참고문헌, 예를 들어 특허, 특허 출원, 수탁 번호, 기술 저널 및 전문 저널은 본 명세서에 걸쳐 인용된다. 각각의 참고문헌은 모든 목적을 위해 본원에 그 전문이 참고로 인용된다.
- [0006] 관상 동맥 질환(CAD)은 심장에 혈액, 산소 및 영양소를 제공하는 관상 동맥이 손상되거나 이환될 때 발생한다. CAD의 흔한 원인은 콜레스테롤 함유 침착물(플라크) 및 염증이다. 플라크 빌드업은 관상 동맥이 좁아지게 하여서, 심장으로의 혈류를 감소시킨다. 몇몇 경우에, 혈류의 감소는 흉통(협심증), 숨참 또는 다른 관상 동맥 질환 징후 및 증상을 초래할 수 있다. 완전한 봉쇄는 심근경색을 초래할 수 있다.
- [0007] 심부 정맥 혈전증(deep venous thrombosis: DVT) 및 폐 색전증으로 이루어진 정맥 혈전색전증(venous thromboembolism: VTE)은 정맥에서의 혈병의 형성을 특징으로 하는 재발하고 쇠약하게 만드는 질환이다. 가족 기반 연구는 유전 변이가 VTE 위험(risk)에 대한 주요 기여자임을 시사한다. 그러나, VTE는 복잡한 병인을 갖고, GWAS를 통해 확인된 다형은 VTE의 유전 가능한 성분의 약 5%를 차지하여서, 질환의 유전적 토대에 제한된 통찰력을 제공한다. VTE 위험에 영향을 미치는 신규한 유전자 변이체의 확인은 새로운 치료학적 표적을 분명히 보여주고, VTE 예방 및 치료에 대해 현재의 치료제에 대한 더 안전하고 더 효과적인 대안으로 길을 안내할 수 있다.

발명의 내용

- [0008] 본 개시내용은 SLC14A1의 생물학을 이해하는 것을 돕고, 응고 병태 및 CAD의 진단 및 치료가 수월하게 하는 SLC14A1 변이체를 제공한다. 본 개시내용은, 응고 장애 및 CAD로부터의 보호와 연관된 것으로 본원에서 입증된, SLC14A1 변이체 폴리펩타이드를 암호화하는 핵산 분자(즉, 게놈 DNA, mRNA 및 cDNA) 및 SLC14A1 변이체 폴리펩타이드를 제공한다.
- [0009] 본 개시내용은 또한 인간 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 단리된 핵산 분자를 제공하되, 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그의 보체를 포함하거나, 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그의 보체를 포함한다.
- [0010] 본 개시내용은 또한 인간 SLC14A1 단백질의 적어도 일부를 암호화하는 핵산 서열을 포함하는 게놈 DNA 분자를 제공하되, 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그 핵산 서열의 보체를 포함하거나, 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그 핵산 서열의 보체를 포함한다.
- [0011] 본 개시내용은 또한 인간 SLC14A1 단백질의 적어도 일부를 암호화하는 핵산 서열을 포함하는 mRNA 분자를 제공하되, 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그 핵산 서열의 보체를 포함하거나, 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열 또는 그 핵산 서열의 보체를 포함한다.
- [0012] 본 개시내용은 또한 인간 SLC14A1 단백질의 적어도 일부를 암호화하는 핵산 서열을 포함하는 cDNA 분자를 제공하되, 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그 핵산 서열의 보체를 포함하거나, 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 것인 핵산 서열 또는 그 핵산 서열의 보체를 포함한다.
- [0013] 본 개시내용은 또한 본원에 개시된 임의의 단리된 핵산 분자를 포함하는 벡터를 제공한다.
- [0014] 본 개시내용은 또한 본원에 개시된 임의의 단리된 핵산 분자 또는 벡터 및 담체를 포함하는 조성물을 제공한다.
- [0015] 본 개시내용은 또한 본원에 개시된 임의의 단리된 핵산 분자 또는 벡터를 포함하는 숙주 세포를 제공한다.
- [0016] 본 개시내용은 또한 인간 SLC14A1 단백질의 적어도 일부를 포함하는 단리된 또는 재조합 폴리펩타이드를 제공하되, 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신 또는 그 핵산 서열의 보체를 포함하거나, 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신 또는 그 핵산 서열의 보체를 포함한다.
- [0017] 본 개시내용은 또한 본원에 개시된 임의의 단리된 또는 재조합 폴리펩타이드 및 담체를 포함하는 조성물을 제공한다.

- [0018] 본 개시내용은 또한 적어도 약 5개의 뉴클레오타이드를 포함하는 핵산 서열을 포함하는 프로브 또는 프라이머를 제공하되, 이들은 인간 SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화하되, 여기서 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하거나, 또는 인간 SLC14A1 단백질을 암호화하는 핵산 서열의 보체에 혼성화하되, 여기서 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다.
- [0019] 본 개시내용은 또한 본원에 개시된 임의의 프로브가 혼성화하는 기질을 포함하는 지지체를 제공한다.
- [0020] 본 개시내용은 또한 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적인 핵산 서열을 포함하는 변경-특이적(alteration-specific) 프로브 또는 프라이머를 제공하고, 여기서 상기 변경-특이적 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치를 암호화하거나 또는 서열번호 14에 따른 132번 위치를 암호화하는 핵산 분자의 일부에 상보적인 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치를 암호화하는 핵산 분자의 일부에 특이적으로 혼성화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치를 암호화하는 핵산 분자의 일부에 특이적으로 혼성화하거나 또는 이들 핵산 분자의 적어도 하나의 보체에 특이적으로 혼성화한다. 변경-특이적 프로브 또는 프라이머는 야생형 SLC14A1 단백질을 암호화하는 핵산 서열을 갖는 핵산 분자에 혼성화하지 않는다.
- [0021] 본 개시내용은 또한 응고 병태 또는 응고 병태를 발생시킬 위험, 또는 관상 동맥 질환 또는 관상 동맥 질환을 발생시킬 위험을 갖는 인간 대상체를 확인하는 방법을 제공하고, 여기서 상기 방법은 상기 대상체로부터 수득된 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을; 및/또는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 핵산 분자의 존재 또는 부재를 검출하는 단계를 포함하고; 상기 변이체 SLC14A1 단백질 및/또는 상기 변이체 SLC14A1 단백질을 암호화하는 핵산 분자의 부재는 상기 대상체가 응고 병태 또는 응고 병태를 발생시킬 위험, 또는 관상 동맥 질환 또는 관상 동맥 질환을 발생시킬 위험을 갖는다는 것을 나타낸다.
- [0022] 본 개시내용은 또한 인간 대상체에서 응고 병태를 진단, 응고 병태를 발생시킬 위험, 관상 동맥 질환 또는 관상 동맥 질환을 발생시킬 위험을 검출하기 위한 방법으로서, 인간 대상체로부터 수득된 SLC14A1 단백질을 암호화하는 핵산 분자에서 변경의 존재 또는 부재를 검출하되, 상기 변경은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 것인 단계; 및 상기 대상체가 상기 변경이 결여되고 응고 병태 또는 관상 동맥 질환의 하나 이상의 증상을 갖는 경우에 응고 병태 또는 관상 동맥 질환을 갖는 것으로 인간 대상체를 진단하거나, 또는 상기 대상체가 상기 변경이 결여되고 응고 병태 또는 관상 동맥 질환의 하나 이상의 증상을 갖지 않는 경우에 응고 병태 또는 관상 동맥 질환을 발생시킬 위험이 있는 것으로 인간 대상체를 진단하는 단계를 포함하는 방법을 제공한다.
- [0023] 본 개시내용은 또한 응고 병태를 예방하거나 치료하거나 저해하는 치료제로 응고 병태 환자를 치료하는 방법으로서, 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 DNA 샘플에서 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 응고 병태와 연관된 유전자 변이체 중 하나 이상을 가질 때, 응고 병태를 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는 방법을 제공한다.
- [0024] 본 개시내용은 또한 응고 병태를 예방하거나 치료하거나 저해하는 치료제로 응고 병태 환자를 치료하는 방법으로서, 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 단백질 샘플에서 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 응고 병태와 연관된 유전자 변이체 중 하나 이상을 가질 때, 응고 병태를 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는 방법을 제공한다.
- [0025] 본 개시내용은 또한 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로 관상 동맥 질환(CAD) 환자를 치료하는 방법으로서, 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 DNA 샘플에서 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관

된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 관상 동맥 질환과 연관된 유전자 변이체 중 하나 이상을 가질 때, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는 방법을 제공한다.

[0026] 본 개시내용은 또한 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로 관상 동맥 질환(CAD) 환자를 치료하는 방법으로서, 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 단백질 샘플에서 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 관상 동맥 질환과 연관된 유전자 변이체 중 하나 이상을 가질 때, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는 방법을 제공한다.

[0027] 본 개시내용은 또한 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 갖는 인간 대상체에서 응고 병태의 치료에서 사용하기 위한 응고의 저해제를 제공한다.

[0028] 본 개시내용은 또한 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 갖는 인간 대상체에서 CAD의 치료에서 사용하기 위한 제제를 제공한다.

도면의 간단한 설명

[0029] 본 명세서의 부분에 통합되고 이를 구성하는 동반된 도면은 몇몇 양상을 예시하고, 설명과 함께 본 개시내용의 원칙을 설명하도록 작용한다.

도 1은 활성화된 부분 트롬보플라스틴 시간(activated partial thromboplastin time: aPTT)에 대한 유전 연관 연구의 그래프 결과를 보여준다.

도 2는 분석에서 aPTT와의 새로운 연관을 보여준다.

도 3은 SLC14A1 Val76Ile에 대한 aPTT 메타-분석의 포레스트 선도(Forest plot)를 보여준다.

도 4는 aPTT와의 SLC14A1 Val76Ile 메타-분석 연관에 대한 구역적 선도를 보여준다.

도 5는 SLC14A1 V76I에 대한 CAD 메타-분석의 포레스트 선도를 보여준다.

도 6은 분석에서 aPTT와의 새로운 연관을 보여준다.

본 개시내용의 추가적인 이점은 부분적으로는 하기 설명에 기재될 것이고 부분적으로는 그 설명으로부터 자명해 지거나 또는 본원에 개시된 실시형태의 실행에 의해 학습될 수 있다. 본 개시내용의 이점은 첨부된 청구범위에 특히 제시된 요소 및 조합에 의해 실현되고 획득될 것이다. 상기 일반 설명 및 하기 상세한 설명 둘 다가 오직 예시적이고 설명적이고, 청구된 바대로 실시형태를 제한하지 않는다고 이해되어야 한다.

발명을 실시하기 위한 구체적인 내용

[0030] 본 개시내용의 양상에 관한 다양한 용어는 본 명세서 및 청구범위에 걸쳐 사용된다. 이러한 용어는, 달리 표시되지 않는 한, 당해 분야에서 이의 보통의 의미가 주어져야 한다. 다른 구체적으로 정의된 용어는 본원에 제공된 정의와 일치하는 방식으로 해석되어야 한다.

[0031] 달리 명확히 기술되지 않는 한, 본원에 기재된 임의의 방법 또는 양상이 특정한 순서로 이의 단계가 수행될 것을 요하는 것으로 결코 의도되지 않는다. 따라서, 청구범위 또는 명세서에서 특정한 순서로 단계들이 제한되지 않는다고 방법 청구항이 구체적으로 기재하지 않는 경우, 어떤 점에서는 순서가 추론되는 것으로 결코 의도되지 않는다. 이는 단계 또는 조작 흐름의 배열과 관련한 논리 문제, 문법 조직화 또는 구두점으로부터 유래된 명백한 의미, 또는 본 명세서에 기재된 양상의 수 또는 유형을 포함하여, 해석을 위한, 임의의 가능한 비표현 기준을 유지한다.

[0032] 본원에서 사용될 때, 단수 형태의 표현은, 문맥이 명확히 달리 기술하지 않는 한, 복수의 지시대상을 포함한다.

[0033] 본원에서 사용될 때, "대상체" 및 "환자"의 용어는 상호 교환 가능하게 사용된다. 대상체는 포유류를 포함한 임의의 동물을 포함할 수 있다. 포유류는, 제한 없이, 농장 동물(예를 들어, 말, 소, 돼지), 반려 동물(예를 들어, 개, 고양이), 실험실 동물(예를 들어, 마우스, 래트, 토끼) 및 비인간 영장류를 포함한다. 몇몇 실시형태에

서, 상기 대상체는 인간이다.

- [0034] 본원에서 사용될 때, "핵산", "핵산 분자", "핵산 서열", "폴리뉴클레오타이드" 또는 "올리뉴클레오타이드"는 임의의 길이의 뉴클레오타이드의 중합체 형태를 포함할 수 있고, DNA 및/또는 RNA를 포함할 수 있고, 단일-가닥, 이중-가닥 또는 다중 가닥일 수 있다. 핵산의 하나의 가닥은 또한 이의 보체를 지칭한다.
- [0035] 본원에서 사용될 때, "에 상응하는"의 구절 또는 이의 문법 파생어는, 주어진 아미노산 또는 핵산 서열 또는 위치의 넘버링의 맥락에서 사용될 때, 주어진 아미노산 또는 핵산 서열이 기준 서열(예를 들어, (야생형 또는 전장) SLC14A1의 핵산 분자 또는 폴리펩타이드인 본원에서의 기준 서열)과 비교될 때의 기재된 기준 서열의 넘버링을 지칭한다. 다른 말로, 주어진 중합체의 잔기(예를 들어, 아미노산 또는 뉴클레오타이드) 수 또는 잔기(예를 들어, 아미노산 또는 뉴클레오타이드) 위치는 주어진 아미노산 또는 핵산 서열 내의 잔기의 실제 숫자 위치에 의하기보다는 기준 서열과 관련하여 지칭된다. 예를 들어, 주어진 아미노산 서열은 2개의 서열 사이의 잔기 일치율을 최적화시키도록 갭을 도입함으로써 기준 서열로 정렬될 수 있다. 이 경우에, 갭이 존재하지만, 주어진 아미노산 또는 핵산 서열에서의 잔기의 넘버링은 이것이 정렬되는 기준 서열과 관련하여 이루어진다.
- [0036] 예를 들어, "단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 인간 SLC14A1 단백질"의 구절(및 유사한 구절)은, SLC14A1 단백질의 아미노산 서열이 서열번호 13의 서열로 정렬되는 경우, SLC14A1 단백질이 서열번호 13의 76번 위치에 상응하는 위치에서 이소류신을 보유한다는 것을 의미한다. 본원에서, 이러한 단백질은 또한 "변이체 SLC14A1 단백질" 또는 "SLC14A1 Val76Ile"라 칭해진다.
- [0037] 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질은 주어진 SLC14A1 단백질과 서열번호 13의 아미노산 서열 사이의 서열 정렬을 수행함으로써 쉽게 확인될 수 있다. 마찬가지로, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질은 주어진 SLC14A1 단백질과 서열번호 14의 아미노산 서열 사이에 서열 정렬을 수행함으로써 쉽게 확인될 수 있다. 서열번호 13에서 76번 위치에 상응하는 위치에서 이소류신을 확인하기 위해, 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 확인하기 위해 서열 정렬을 수행하기 위해 이용될 수 있는 다양한 컴퓨터 알고리즘이 존재한다. 예를 들어, NCBI BLAST 알고리즘(Altschul *et al.*, 1997, Nuc. Acids Res., 25, 3389-3402) 또는 CLUSTALW 소프트웨어(Sievers *et al.*, 2014, Methods Mol. Biol., 1079, 105-116)를 사용함으로써 서열 정렬이 수행될 수 있다. 그러나, 서열은 또한 수동으로 정렬될 수 있다.
- [0038] 본 개시내용에 따르면 SLC14A1에서의 특정한 변이가 출혈 시간의 연장(예를 들어, 혈액 응고의 감소)과 연관할 수 있거나, 관상 동맥 질환에 대해 보호하도록 작용할 수 있다고 관찰되었다. SLC14A1에서의 이 변이가 응고 병태에 대해 보호를 추가로 제공할 수 있다고 믿어진다. SLC14A1 유전자 또는 단백질의 변이체가 인간에서 관상 동맥 질환과 관련한 이러한 보호 기능과 어떠한 이전의 공지된 연관을 갖지 않는다고 믿어진다. 이환된 가족 구성원에서 관상 동맥 질환에 대한 보호의 표현형으로 분리하는 SLC14A1 유전자에서의 희귀 변이체는 본 개시내용에 따라 확인되었다. SLC14A1 핵산에서의 이러한 보호 변형은 기능 소실을 갖는 SLC14A1 단백질 또는 SLC14A1 저차형태(예를 들어, 기능의 부분 소실) 단백질을 생성시킨다. 예를 들어, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신에 의한 발린의 대체를 발생시키는 유전자 변형은 이러한 변형을 갖는 인간이 관상 동맥 질환을 발생시키는 것에 대한 보호를 보유할 수 있거나, 관상 동맥 질환을 발생시킬 위험의 감소를 가질 수 있다는 것을 나타낸다고 관찰되었다.
- [0039] 종합하면, 본원에 기재된 유전자 분석은 놀랍게도 기능 소실 또는 기능 부분 소실을 갖는 SLC14A1 단백질을 생성시키는 SLC14A1 유전자에서의 변이체가 관상 동맥 질환에 대한 감수성의 감소와 연관된다는 것을 나타내고, 신체에서 응고 기반 증상에 대한 감수성의 감소와 연관된다고 믿어진다. 따라서, 응고 병태 또는 관상 동맥 질환에 대한 보호와 연관된 SLC14A1 변형을 보유하지 않는 인간 대상체는 응고 병태 또는 관상 동맥 질환이 저해되고/되거나, 이의 증상이 감소되고/되거나, 증상의 발생이 억제되도록 치료될 수 있다. 따라서, 본 개시내용은 단리된 또는 재조합 SLC14A1 변이체 핵산 분자, 예컨대 유전자, mRNA 및 cDNA, 및 단리된 또는 재조합 SLC14A1 변이체 폴리펩타이드를 제공한다. 추가적으로, 본 개시내용은, 위험에 있는 대상체 또는 활성 질환을 갖는 대상체가 치료될 수 있도록, 이러한 대상체에서의 응고 병태 또는 관상 동맥 질환을 발생시킬 위험을 확인하거나 계층화하기 위해, 또는 응고 병태 또는 관상 동맥 질환을 갖는 것으로 대상체를 진단하기 위해 대상체에서의 이러한 변이체의 확인을 발휘하는 방법을 제공한다.
- [0040] 2개의 야생형 SLC14A1 단백질에 대한 아미노산 서열은 서열번호 11 및 서열번호 12에 기재된다. 서열번호 11을 갖는 야생형 SLC14A1 단백질은 389개의 아미노산의 길이인 한편, 서열번호 12를 갖는 야생형 SLC14A1 단백질은 445개의 아미노산의 길이이다. 서열번호 11은 76번 위치에서 발린을 포함하고, 서열번호 12는 132번 위치에서

발린을 포함한다.

- [0041] 본 개시내용은 응고 병태 또는 관상 동맥 질환으로부터의 보호와 연관된 SLC14A1 변이체 단백질을 암호화하는 핵산 분자를 제공한다. 예를 들어, 본 개시내용은 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 단리된 핵산 분자를 제공하고, 여기서 변이체 SLC14A1 단백질은 기능 소실 단백질 또는 기능 부분 소실 단백질이다. 특히, 본 개시내용은 인간 SLC14A1 단백질을 암호화하는 핵산 서열 또는 상기 핵산 서열의 보체를 포함하는 단리된 핵산 분자를 제공하고, 여기서 상기 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함한다.
- [0042] 몇몇 실시형태에서, 핵산 분자는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 아미노산 서열을 갖는 인간 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열 또는 상기 핵산 서열의 보체를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 핵산 분자는 서열번호 13을 암호화하지 않는다. 본원에서, 퍼센트 서열 동일성을 언급하는 경우, 서열 동일성의 더 높은 백분율이 더 낮은 것에 비해 바람직하다.
- [0043] 몇몇 실시형태에서, 본 개시내용은 인간 SLC14A1 단백질을 암호화하는 핵산 서열 또는 상기 핵산 서열의 보체를 포함하는 단리된 핵산 분자를 제공하고, 여기서 상기 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다.
- [0044] 몇몇 실시형태에서, 핵산 분자는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 아미노산 서열을 갖는 인간 SLC14A1 단백질을 암호화하고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열 또는 상기 핵산 서열의 보체를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 핵산 분자는 서열번호 14를 암호화하지 않는다. 본원에서, 퍼센트 서열 동일성을 언급하는 경우, 서열 동일성의 더 높은 백분율이 더 낮은 것에 비해 바람직하다.
- [0045] 야생형 SLC14A1 게놈 DNA의 핵산 서열은 서열번호 1에 기재된다. 서열번호 1을 포함하는 야생형 SLC14A1 게놈 DNA는 28,394개의 뉴클레오타이드의 길이이다. 서열번호 1을 살펴보면, 야생형 SLC14A1 게놈 DNA의 6963번 위치는 구아닌이다.
- [0046] 본 개시내용은 변이체 SLC14A1 단백질을 암호화하는 게놈 DNA 분자를 제공한다. 몇몇 실시형태에서, 게놈 DNA 분자는 기능 소실 단백질 또는 기능 부분 소실 단백질인 변이체 SLC14A1 단백질을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다.
- [0047] 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13을 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 게놈 DNA는 서열번호 13을 암호화하는 핵산 서열을 포함하지 않거나 이들로 이루어지지 않는다.
- [0048] 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 14에 따른 132번 위치에 상응하는 위

치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 14를 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76 번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 게놈 DNA는 서열번호 14를 암호화하는 핵산 서열을 포함하지 않거나 이들로 이루어지지 않는다.

[0049] 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 게놈 DNA는 서열번호 1에 따른 6963번 위치에 상응하는 위치에서 구아닌을 포함한다. 몇몇 실시형태에서, 게놈 DNA는 서열번호 2와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 게놈 DNA는 서열번호 2에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 게놈 DNA는 서열번호 2와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 게놈 DNA는 서열번호 2에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0050] 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 2와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 단, 핵산 서열은 이소류신을 암호화하는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 코돈을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 뉴클레오타이드를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 2를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 게놈 DNA는 서열번호 2와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열을 포함하되, 단, 핵산 서열은 이소류신을 암호화하는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 코돈을 포함하고, 그리고, 단, 변이체 SLC14A1 게놈 DNA는 서열번호 2 또는 그 보체를 포함하지 않는다.

[0051] 몇몇 실시형태에서, 단리된 핵산 분자는 전체보다 적은 게놈 DNA 서열을 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 2의 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 2000개, 적어도 약 3000개, 적어도 약 4000개, 적어도 약 5000개, 적어도 약 6000개, 적어도 약 7000개, 적어도 약 8000개, 적어도 약 9000개, 적어도 약 10000개, 적어도 약 11000개, 적어도 약 12000개, 적어도 약 13000개, 적어도 약 14000개, 적어도 약 15000개, 적어도 약 16000개, 적어도 약 17000개, 적어도 약 18000개, 적어도 약 19000개, 적어도 약 20000개, 적어도 약 21000개, 적어도 약 22000개, 적어도 약 23000개, 적어도 약 24000개, 적어도 약 25000개, 적어도 약 26000개, 적어도 약 27000개 또는 적어도 약 28000개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 2의 적어도 약 1000 내지 적어도 약 2000개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다.

[0052] 몇몇 실시형태에서, 단리된 핵산 분자는 전체보다 적은 게놈 DNA 서열을 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 2의 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 2000개 또는 적어도 약 3000개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 이러한 인접 뉴클레오타이드는 본원에 기재된 cDNA 분자를 제조하도록 인접 뉴클레오타이드의 다른 핵산 분자와 조합될 수 있다.

[0053] 이러한 단리된 핵산 분자는, 예를 들어 변이체 SLC14A1 mRNA 및 단백질을 발현하도록 사용될 수 있거나 또는 외인성 도너 서열로서 사용될 수 있다. 집단 내에 유전자 서열이 다형, 예컨대 SNP로 인해 변할 수 있다고 이해된

다. 본원에 제공된 예는 오직 예시적인 서열이고, 다른 서열이 또한 가능하다.

- [0054] 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 13 또는 서열번호 14를 암호화하는 하나 이상의 비필수 분절(nonessential segments)이 상응하는 야생형 SLC14A1 게놈 DNA와 관련하여 결실된 변이체 SLC14A1 미니유전자(minigene)를 포함한다. 몇몇 실시형태에서, 결실된 비필수 분절(들)은 하나 이상의 인트론 서열을 포함한다. 몇몇 실시형태에서, SLC14A1 미니유전자는 서열번호 13 또는 서열번호 14의 일부와 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 적어도 약 99% 또는 100%의 서열 동일성을 갖고, 여기서 미니유전자는 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아테닌을 갖는 핵산 서열을 포함한다.
- [0055] 2개의 야생형 SLC14A1 mRNA의 핵산 서열은 서열번호 3 및 서열번호 4에 기재된다. 서열번호 3을 포함하는 야생형 SLC14A1 mRNA는 1170개의 뉴클레오타이드의 길이이다. 서열번호 3을 살펴보면, 야생형 SLC14A1 mRNA의 226번 위치는 구아닌이다. 서열번호 4를 포함하는 야생형 SLC14A1 mRNA는 1338개의 뉴클레오타이드의 길이이다. 서열번호 4를 살펴보면, 야생형 SLC14A1 mRNA의 394번 위치는 구아닌이다.
- [0056] 본 개시내용은 또한 변이체 SLC14A1 단백질을 암호화하는 mRNA 분자를 제공한다. 몇몇 실시형태에서, mRNA 분자는 기능 소실 단백질 또는 기능 부분 소실 단백질인 변이체 SLC14A1 단백질을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다.
- [0057] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13을 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 mRNA는 서열번호 13을 암호화하는 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0058] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 14를 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 mRNA는 서열번호 14를 암호화하는 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0059] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 226번 위치에 상응하는 위치에서 아테닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 mRNA는 서열번호 5에 따른 226번 위치에 상응하는 위치에서 구아닌을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 mRNA는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 GUC를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0060] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도

도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 5에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 5에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0061] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 5와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하고, 그리고, 단, 변이체 SLC14A1 mRNA는 서열번호 5에 따른 핵산 서열 또는 그의 보체를 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0062] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 mRNA는 서열번호 6에 따른 394번 위치에 상응하는 위치에서 구아닌을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 mRNA는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 GUC를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0063] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 6에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 6에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0064] 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 단, 핵산 서열은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 mRNA는 서열번호 6과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하되, 단, 핵산 서열은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하고, 그리고, 단, 변이체 SLC14A1 mRNA는 서열번호 6에 따른 핵산 서열을 포함하지 않는다.

[0065] 몇몇 실시형태에서, 단리된 핵산 분자는 전체 SLC14A1 mRNA 서열보다 적은 뉴클레오타이드를 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 5의 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 12개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 1100개 또는 적어도 약 1200개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 5의 적어도 약 200개 내지 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 mRNA 분자는 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 5의 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도

도 약 200개, 적어도 약 300개, 적어도 약 400개 또는 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 mRNA 분자가 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈을 포함한다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 5에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함한다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 AUC를 포함한다.

[0066] 몇몇 실시형태에서, 단리된 핵산 분자는 전체 SLC14A1 mRNA 서열보다 적은 뉴클레오타이드를 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 6의 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 12개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 1100개, 적어도 약 1200개 또는 적어도 약 1300개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 6의 적어도 약 200개 내지 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 mRNA 분자는 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 6의 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개 또는 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 mRNA 분자는 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈을 포함한다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 6에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함한다. 몇몇 실시형태에서, 이러한 mRNA 분자는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 AUC를 포함한다.

[0067] 2개의 야생형 SLC14A1 cDNA의 핵산 서열은 서열번호 7 및 서열번호 8에 기재된다. 서열번호 7을 포함하는 야생형 SLC14A1 cDNA는 종결 코돈을 포함하여 1173개의 뉴클레오타이드의 길이이다. 서열번호 7을 살펴보면, 야생형 SLC14A1 cDNA의 226번 위치는 구아닌이다. 서열번호 8을 포함하는 야생형 SLC14A1 cDNA는 종결 코돈을 포함하여 1341개의 뉴클레오타이드의 길이이다. 서열번호 8을 살펴보면, 야생형 SLC14A1 cDNA의 394번 위치는 구아닌이다.

[0068] 본 개시내용은 또한 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 cDNA 분자를 제공한다. 몇몇 실시형태에서, 변이체 cDNA 분자는 기능 소실 단백질 또는 기능 부분 소실 단백질인 변이체 SLC14A1 단백질을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13 또는 서열번호 14에 따른 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0069] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13을 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 cDNA는 서열번호 13에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0070] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%

의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 14를 갖는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 cDNA는 서열번호 14에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0071] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 cDNA는 서열번호 9에 따른 226번 위치에 상응하는 위치에서 구아닌을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 cDNA는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 GUC를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0072] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0073] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 9와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하고, 그리고, 단, 변이체 SLC14A1 cDNA는 서열번호 9에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0074] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 cDNA는 서열번호 10에 따른 394번 위치에 상응하는 위치에서 구아닌을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 이에 반해, 야생형 SLC14A1 cDNA는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 GUC를 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0075] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어지되, 단, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0076] 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도

도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 10에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화한다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 cDNA는 서열번호 10과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 핵산 서열 또는 그의 보체를 포함하거나 또는 이들로 이루어지되, 단, 상기 핵산 서열은 서열번호 10에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하고, 그리고, 단, 변이체 SLC14A1 cDNA는 서열번호 10에 따른 핵산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.

[0077] 몇몇 실시형태에서, 단리된 핵산 분자는 전체 SLC14A1 cDNA 서열보다 적은 뉴클레오타이드를 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 9의 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 12개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 1100개 또는 적어도 약 1200개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 9의 적어도 약 200개 내지 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 cDNA 분자가 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 9의 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개 또는 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 cDNA 분자가 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈을 포함한다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함한다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 AUC를 포함한다.

[0078] 몇몇 실시형태에서, 단리된 핵산 분자는 전체 SLC14A1 cDNA 서열보다 적은 뉴클레오타이드를 포함한다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 10의 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 12개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 1100개, 적어도 약 1200개 또는 적어도 약 1300개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 10의 적어도 약 200개 내지 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 cDNA 분자가 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단리된 핵산 분자는 서열번호 10의 적어도 약 50개, 적어도 약 60개, 적어도 약 70개, 적어도 약 80개, 적어도 약 90개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개 또는 적어도 약 500개의 인접 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 cDNA 분자가 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈을 포함한다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함한다. 몇몇 실시형태에서, 이러한 cDNA 분자는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 AUC를 포함한다.

[0079] 본 개시내용은 또한 변이체 SLC14A1 게놈 DNA(예컨대, 서열번호 2), 변이체 SLC14A1 미니유전자, 변이체 SLC14A1 mRNA(예컨대, 서열번호 5 및/또는 서열번호 6), 및/또는 변이체 SLC14A1 cDNA(예컨대, 서열번호 9 및/또는 서열번호 10)에 혼성화하는 단리된 핵산 분자를 제공한다. 몇몇 실시형태에서, 이러한 단리된 핵산 분자는 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 11개, 적어도 약 12개, 적어도 약 13개, 적어도 약 14개, 적어도 약 15개, 적어도 약 16개, 적어도 약 17개, 적어도 약 18개, 적어도 약 19개, 적어도 약 20개, 적어도 약 21개, 적어도 약 22개, 적어도 약 23개, 적어도 약 24개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 55개, 적어도 약 60개, 적어도 약 65개, 적어도 약 70개, 적어도 약 75개, 적어도 약 80개, 적어도 약 85개, 적어도 약 90개, 적어도 약 95개, 적어도 약 100개, 적어도 약 200개, 적어도 약 300개, 적어도 약 400개, 적어도 약 500개, 적어도 약 600개, 적어도 약 700개, 적어도 약 800개, 적어도 약 900개, 적어도 약 1000개, 적어도 약 2000개, 적어도 약 3000개, 적어도 약

4000개, 적어도 약 5000개, 적어도 약 6000개, 적어도 약 7000개, 적어도 약 8000개, 적어도 약 9000개, 적어도 약 10000개, 적어도 약 11000개 또는 적어도 약 1200개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 적어도 15개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 적어도 15개의 뉴클레오타이드 내지 적어도 약 35개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 이러한 단리된 핵산 분자는 엄격한 조건하에 변이체 SLC14A1 게놈 DNA(예컨대, 서열번호 2), 변이체 SLC14A1 미니유전자, 변이체 SLC14A1 mRNA(예컨대, 서열번호 5 및/또는 서열번호 6), 및/또는 변이체 SLC14A1 cDNA(예컨대, 서열번호 9 및/또는 서열번호 10)에 혼성화한다. 이러한 핵산 분자는, 본원에 기재된 또는 예시된 바와 같은, 예를 들어 프로브로서, 프라이머로서 또는 변경-특이적 프로브 또는 프라이머로서 사용될 수 있다.

[0080] 몇몇 실시형태에서, 단리된 핵산 분자는 변이체 SLC14A1 게놈 DNA(예컨대, 서열번호 2), 변이체 SLC14A1 미니유전자, 변이체 SLC14A1 mRNA(예컨대, 서열번호 5 및/또는 서열번호 6), 및/또는 변이체 SLC14A1 cDNA(예컨대, 서열번호 9 및/또는 서열번호 10)와 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 적어도 약 99% 또는 100% 동일한 핵산 분자의 적어도 약 15개의 인접 뉴클레오타이드에 혼성화한다. 몇몇 실시형태에서, 단리된 핵산 분자는 약 15개 내지 약 100개의 뉴클레오타이드, 또는 약 15개 내지 약 35개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 약 15개 내지 약 100개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단리된 핵산 분자는 약 15개 내지 약 35개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다.

[0081] 몇몇 실시형태에서, 본원에 개시된 임의의 핵산 분자, 게놈 DNA 분자, cDNA 분자 또는 mRNA 분자는 정제될 수 있고, 예를 들어 적어도 약 90% 순수하다. 몇몇 실시형태에서, 본원에 개시된 임의의 핵산 분자, 게놈 DNA 분자, cDNA 분자 또는 mRNA 분자는 정제될 수 있고, 예를 들어 적어도 약 95% 순수하다. 몇몇 실시형태에서, 본원에 개시된 임의의 핵산 분자, 게놈 DNA 분자, cDNA 분자 또는 mRNA 분자는 정제될 수 있고, 예를 들어 적어도 약 99% 순수하다. 정제는 인간 제조 정제 기법에 의해 인간 손에 따른다.

[0082] 본 개시내용은 또한 본원에 개시된 임의의 단리된 핵산 분자, 게놈 DNA 분자, cDNA 분자 또는 mRNA 분자의 단편을 제공한다. 몇몇 실시형태에서, 단편은 본원에 개시된 임의의 핵산 서열, 또는 그의 임의의 보체의 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 11개, 적어도 약 12개, 적어도 약 13개, 적어도 약 14개, 적어도 약 15개, 적어도 약 16개, 적어도 약 17개, 적어도 약 18개, 적어도 약 19개, 적어도 약 20개, 적어도 약 21개, 적어도 약 22개, 적어도 약 23개, 적어도 약 24개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 55개, 적어도 약 60개, 적어도 약 65개, 적어도 약 70개, 적어도 약 75개, 적어도 약 80개, 적어도 약 85개, 적어도 약 90개, 적어도 약 95개 또는 적어도 약 100개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 단편이 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단편은 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 11개, 적어도 약 12개, 적어도 약 13개, 적어도 약 14개, 적어도 약 15개, 적어도 약 16개, 적어도 약 17개, 적어도 약 18개, 적어도 약 19개, 적어도 약 20개, 적어도 약 21개, 적어도 약 22개, 적어도 약 23개, 적어도 약 24개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개 또는 적어도 약 50개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 이와 관련하여, 더 긴 단편이 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단편은 적어도 약 20개, 적어도 약 25개, 적어도 약 30개 또는 적어도 약 35개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단편은 적어도 약 20개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단편은 적어도 약 25개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단편은 적어도 약 30개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 단편은 적어도 약 35개의 인접 잔기를 포함하거나 또는 이들로 이루어진다. 단편이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 분자의 일부를 포함하거나 또는 이들로 이루어지는 것으로 고안된다. 이러한 단편은, 예를 들어 본원에 기재된 또는 예시된 바와 같은, 프로브로서, 프라이머로서 또는 대립유전자-특이적 프라이머로서 사용될 수 있다.

[0083] 본 개시내용은 또한 프로브 및 프라이머를 제공한다. 본 개시내용의 프로브 또는 프라이머는 본원에 개시된 임의의 핵산 분자 또는 그의 보체에 특이적으로 혼성화하는 핵산 서열을 갖는다. 몇몇 실시형태에서, 프로브 또는 프라이머는 엄격한 조건하에 본원에 개시된 임의의 핵산 분자에 특이적으로 혼성화한다. 본 개시내용은 또한 보통의 조건(moderate conditions)하에 본원에 개시된 임의의 핵산 분자 또는 그의 보체에 혼성화하는 핵산 서열

을 갖는 핵산 분자를 제공한다. 본 개시내용에 따른 프로브 또는 프라이머는 바람직하게는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 코돈 또는 그의 보체를 포함한다. 본 개시내용에 따른 프로브 또는 프라이머는 바람직하게는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 코돈 또는 그의 보체를 포함한다. 이와 같이, 바람직한 실시형태에서, 본 개시내용은 본원 상에서 정의되어 있고 하기에 더 자세히 정의된 변경-특이적 프라이머를 제공한다.

[0084] 본 개시내용에 따른 프로브는 (예를 들어, 서열번호 13 및/또는 서열번호 14에 따른) 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 핵산 분자(예를 들어, 게놈 DNA, mRNA, 및/또는 cDNA)를 검출하도록 사용될 수 있다. 게다가, 본 개시내용에 따른 프라이머는 변이체 SLC14A1 단백질을 또는 이의 단편을 암호화하는 핵산 분자를 증폭시키도록 사용될 수 있다. 본 개시내용은 또한 상기에 기재된 프라이머 중 하나를 포함하는 프라이머의 쌍을 제공한다.

[0085] 본원에 개시된 핵산 분자는 자연 발생적 SLC14A1 게놈 DNA, cDNA 또는 mRNA 전사체의 핵산 서열을 포함할 수 있거나, 또는 비자연 발생적 서열을 포함할 수 있다. 몇몇 실시형태에서, 자연 발생적 서열은 동의 돌연변이 또는 암호화된 SLC14A1 폴리펩타이드에 영향을 갖지 않는 돌연변이로 인해 비자연 발생적 서열과 상이할 수 있다. 예를 들어, 서열은 동의 돌연변이 또는 암호화된 SLC14A1 폴리펩타이드에 영향을 갖지 않는 돌연변이를 제외하고 동일할 수 있다. 동의 돌연변이 또는 치환은 제조된 아미노산 서열이 변형되지 않도록 단백질을 암호화하는 유전자의 엑손에서 또 다른 뉴클레오타이드에 대한 하나의 뉴클레오타이드의 치환이다. 이는 하나 초과 3개 염기 쌍 코돈에 의해 몇몇 아미노산이 암호화되면서 유전 암호의 축퇴성 때문에 가능하다. 동의 치환은, 예를 들어 코돈 최적화의 과정에서 사용된다. 본원에 개시된 핵산 분자는 코돈 최적화될 수 있다.

[0086] 개시된 핵산 분자와 상호작용할 수 있는 기능적 폴리뉴클레오타이드가 본원에 또한 제공된다. 기능적 폴리뉴클레오타이드는 특이한 기능, 예컨대 표적 분자의 결합 또는 특이적 반응의 촉매화(catalyzing)를 갖는 핵산 분자이다. 기능적 폴리뉴클레오타이드의 예는 안티센스 분자, 압타머, 리보자임, 트리플렉스 형성 분자 및 외부 가이드 서열을 포함하지만, 이들로 제한되지는 않는다. 기능적 폴리뉴클레오타이드는 표적 분자가 보유하는 특이적 활성의 이펙터, 저해제, 조절제 및 자극제로서 작용할 수 있거나, 또는 기능적 폴리뉴클레오타이드는 임의의 다른 분자와 독립적인 신생 활성을 보유할 수 있다.

[0087] 안티센스 분자는 정준 또는 비정준 염기 짝짓기를 통해 표적 핵산 분자와 상호작용하도록 설계된다. 안티센스 분자 및 표적 분자의 상호작용은, 예를 들어 RNase-H 매개된 RNA-DNA 하이브리드 분해를 통해 표적 분자의 파괴를 촉진하도록 설계된다. 대안적으로, 안티센스 분자는 표적 분자에서 보통 발생하는 처리 기능, 예컨대 전사 또는 복제를 차단하도록 설계된다. 안티센스 분자는 표적 분자의 서열에 기초하여 설계될 수 있다. 표적 분자의 가장 접근 가능한 영역을 확인하는 것에 의해 안티센스 효율을 최적화하기 위한 많은 방법이 존재한다. 예시적인 방법은 시험관내 선택 실험 및 DMS 및 DEPC를 사용한 DNA 변형 연구를 포함하지만, 이들로 제한되지는 않는다. 안티센스 분자는 일반적으로 약 10^{-6} 이하, 약 10^{-8} 이하, 약 10^{-10} 이하 또는 약 10^{-12} 이하의 해리 상수(k_d)로 표적 분자에 결합한다. 안티센스 분자의 설계 및 사용을 돕는 방법 및 기법의 대표적인 샘플은 미국 특허의 하기 비제한적인 목록에서 발견될 수 있다: 제5,135,917호; 제5,294,533호; 제5,627,158호; 제5,641,754호; 제5,691,317호; 제5,780,607호; 제5,786,138호; 제5,849,903호; 제5,856,103호; 제5,919,772호; 제5,955,590호; 제5,990,088호; 제5,994,320호; 제5,998,602호; 제6,005,095호; 제6,007,995호; 제6,013,522호; 제6,017,898호; 제6,018,042호; 제6,025,198호; 제6,033,910호; 제6,040,296호; 제6,046,004호; 제6,046,319호; 및 제6,057,437호. 안티센스 분자의 예는 안티센스 RNA, 소간섭 RNA(small interfering RNA: siRNA) 및 짧은 헤어핀 RNA(short hairpin RNA: shRNA)를 포함하지만, 이들로 제한되지는 않는다.

[0088] 본원에 개시된 단리된 핵산 분자는 RNA, DNA, 또는 RNA 및 DNA 둘 다를 포함할 수 있다. 단리된 핵산 분자는 또한, 예컨대 벡터에서의 이중성 핵산 서열 또는 이중성 표지에 연결되거나 융합될 수 있다. 예를 들어, 본원에 개시된 단리된 핵산 분자는 단리된 핵산 분자를 포함하는 벡터 또는 외인성 도너 서열 및 이중성 핵산 서열에 있을 수 있다. 단리된 핵산 분자는 또한 이중성 표지, 예컨대 형광 표지에 연결되거나 또는 융합될 수 있다. 표지의 다른 예는 본원에서 다른 곳에 개시되어 있다.

[0089] 표지는 직접적으로 검출 가능(예를 들어, 형광단) 또는 간접적으로 검출 가능(예를 들어, 합텐, 효소 또는 형광단 켄처)할 수 있다. 이러한 표지는 분광학적, 광화학적, 생화학적, 면역화학적 또는 화학적 수단에 의해 검출 가능할 수 있다. 이러한 표지는, 예를 들어 방사선 계수 장치로 측정될 수 있는 방사선표지; 가시적으로 관찰될 수 있거나 또는 분광광도계로 측정될 수 있는 안료, 염료 또는 다른 발색체; 스핀 표지 분석기로 측정될 수 있는 스핀 표지; 및 출력 신호가 적합한 분자 부가물의 여기에 의해 생성되고, 염료에 의해 흡수된 광에 의한 여

기에 의해 가시화될 수 있거나 또는 표준 형광측정계 또는 영상화 시스템으로 측정될 수 있는 형광 표지(예를 들어, 형판단)를 포함한다. 표지는 또한, 예를 들어 출력 신호가 신호 화합물의 화학 변형에 의해 생성된 화학 발광 물질; 금속 함유 물질; 또는 효소-의존적인 신호의 2차 생성, 예컨대 무색 기질로부터의 유색 생성물의 형성이 발생하는 효소일 수 있다. 용어 "표지"는 접합된 분자가, 후속하여 기질과 함께 첨가될 때, 검출 가능한 신호를 생성하도록 사용되도록 접합된 분자에 선택적으로 결합할 수 있는 "태그" 또는 합텐을 또한 지칭할 수 있다. 예를 들어, 태그로서 바이오틴을 사용하고 이어서 태그에 결합하는 겨자무과산화효소(horseradish peroxidase: HRP)의 아비딘 또는 스트렙타비딘 접합체를 사용하고, 이어서 HRP의 존재를 검출하는 열량측정 기질(예를 들어, 테트라메틸벤지딘(tetramethylbenzidine: TMB)) 또는 형광원성 기질을 사용할 수 있다. 정제를 수월하게 하도록 태그로서 사용될 수 있는 예시적인 표지는 myc, HA, FLAG 또는 3XFLAG, 6XHis 또는 폴리히스티딘, 글루타티온-S-전환효소(GST), 말토스 결합 단백질, 에피토프 태그 또는 면역글로불린의 Fc 부분을 포함하지만, 이들로 제한되지는 않는다. 많은 표지는 공지되어 있고, 예를 들어 입자, 형판단, 합텐, 효소 및 이의 열량 측정, 형광원성 및 화학발광 기질 및 다른 표지를 포함한다.

[0090] 개시된 핵산 분자는, 예를 들어 뉴클레오타이드 또는 비천연(non-natural) 또는 변형된 뉴클레오타이드, 예컨대 뉴클레오타이드 유사체 또는 뉴클레오타이드 치환체를 포함할 수 있다. 이러한 뉴클레오타이드는 변형된 염기, 당 또는 포스페이트 기를 함유하거나 또는 그 구조에 비천연 모이어티가 혼입된 뉴클레오타이드를 포함한다. 비천연 뉴클레오타이드의 예는 디데옥시뉴클레오타이드, 바이오티닐화, 아미노화, 탈아미노화, 알킬화, 벤질화 및 형판단 표지된 뉴클레오타이드를 포함하지만, 이들로 제한되지는 않는다.

[0091] 본원에 개시된 핵산 분자는 또한 하나 이상의 뉴클레오타이드 유사체 또는 치환체를 포함할 수 있다. 뉴클레오타이드 유사체는 염기, 당 또는 포스페이트 모이어티에 대한 변형을 함유하는 뉴클레오타이드이다. 염기 모이어티에 대한 변형은 A, C, G 및 T/U의 천연 및 합성 변형, 및 상이한 퓨린 또는 피리미딘 염기, 예를 들어 슈도우리딘, 우라실-5-일, 하이포잔틴-9-일(I) 및 2-아미노아데닌-9-일 등을 포함하지만, 이들로 제한되지는 않는다. 변형된 염기는 5-메틸사이토신(5-me-C), 5-하이드록시메틸 사이토신, 잔틴, 하이포잔틴, 2-아미노아데닌, 아데닌 및 구아닌의 6-메틸 및 다른 알킬 유도체, 아데닌 및 구아닌의 2-프로필 및 다른 알킬 유도체, 2-티오우라실, 2-티오티민 및 2-티오사이토신, 5-할로우라실 및 사이토신, 5-프로필 우라실 및 사이토신, 6-아조 우라실, 사이토신 및 티민, 5-우라실(슈도우라실), 4-티오우라실, 8-할로, 8-아미노, 8-티올, 8-티오알킬, 8-하이드록실 및 다른 8-치환된 아데닌 및 구아닌, 5-할로, 특히 5-브로모, 5-트리플루오로메틸 및 다른 5-치환된 우라실 및 사이토신, 7-메틸구아닌 및 7-메틸아데닌, 8-아자구아닌 및 8-아자아데닌, 7-테아자구아닌 및 7-테아자아데닌 및 3-테아자구아닌 및 3-테아자아데닌을 포함하지만, 이들로 제한되지는 않는다. 소정의 뉴클레오타이드 유사체 예를 들어, 5-치환된 피리미딘, 6-아자피리미딘, 및 2-아미노프로필아데닌, 5-프로필우라실, 5-프로필사이토신 및 5-메틸사이토신을 포함하지만, 이에 제한되는 것은 아닌 N-2, N-6 및 O-6 치환된 퓨린은 듀플렉스 형성의 안정성을 증가시킬 수 있다. 대개, 염기 변형은 증가된 듀플렉스 안정성과 같은 고유한 특성을 달성하도록, 예를 들어 당 변형, 예컨대 2'-O-메톡시에틸과 조합될 수 있다.

[0092] 뉴클레오타이드 유사체는 또한 당 모이어티의 변형을 포함할 수 있다. 당 모이어티에 대한 변형은 리보스 및 데옥시 리보스의 천연 변형, 및 합성 변형을 포함하지만, 이들로 제한되지는 않는다. 당 변형은 2' 위치에서의 하기 변형을 포함하지만, 이들로 제한되지는 않는다: OH; F; O-, S- 또는 N-알킬; O-, S- 또는 N-알케닐; O-, S- 또는 N-알킬닐; 또는 O-알킬-O-알킬로, 상기 알킬, 알케닐 및 알킬닐은 치환된 또는 비치환된 C₁₋₁₀알킬 또는 C₂₋₁₀알케닐 및 C₂₋₁₀알킬닐일 수 있다. 예시적인 2' 당 변형은 또한 -O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-ONH₂, 및 -O(CH₂)_nON[(CH₂)_nCH₃]₂를 포함하지만, 이들로 제한되지는 않으며, 상기 n 및 m은 1 내지 약 10이다.

[0093] 2' 위치에서 다른 변형은 C₁₋₁₀알킬, 치환된 저급 알킬, 알크아릴, 아르알킬, O-알크아릴 또는 O-아르알킬, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, 헤테로사이클로알킬, 헤테로사이클로알크아릴, 아미노알킬아미노, 폴리알킬아미노, 치환된 실릴, RNA 절단 기, 리포터 기, 삽입제(intercalator), 올리고뉴클레오타이드의 약동학적 특성을 개선하기 위한 기, 또는 올리고뉴클레오타이드의 약력학적 특성을 개선하기 위한 기, 및 유사한 특성을 갖는 다른 치환기를 포함하지만, 이들로 제한되지는 않는다. 유사한 변형은 또한 당 위의 다른 위치에서, 특히 3' 말단 뉴클레오타이드 위의 또는 2'-5' 연결된 올리고뉴클레오타이드에서의 당의 3' 위치 및 5' 말단 뉴클레오타이드의 5' 위치에서 이루어질 수 있다. 변형된 당은 또한 CH₂ 및 S와 같은 브릿징(bridging) 고리 산소에서 변형을 함유하는 것을 포함할 수 있다. 뉴클레오타이드 당 유사체는 또한 펜토프

라노실 당 대신에 당 모방체, 예컨대 사이클로부틸 모이어티를 가질 수 있다.

- [0094] 뉴클레오타이드 유사체는 또한 포스페이트 모이어티에서 변형될 수 있다. 변형된 포스페이트 모이어티는 2개의 뉴클레오타이드 사이의 연결이 포스포로티오에이트, 키랄 포스포로티오에이트, 포스포로디티오에이트, 포스포트리에스테르, 아미노알킬포스포트리에스테르, 메틸 및 다른 알킬 포스포네이트, 포스포네이트, 포스포르아미데이트, 티오노포스포르아미데이트, 티오노알킬포스포네이트, 티오노알킬포스포트리에스테르 및 보라노포스페이트를 함유하도록 변형될 수 있는 것을 포함하지만, 이들로 제한되지는 않는다. 메틸 및 다른 알킬 포스포네이트는 3'-알킬렌 포스포네이트 및 키랄 포스포네이트를 포함하고, 포스포르아미데이트는 3'-아미노 포스포르아미데이트 및 아미노알킬포스포르아미데이트를 포함한다.
- [0095] 2개의 뉴클레오타이드 사이의 이 포스페이트 또는 변형된 포스페이트 연결은 3'-5' 연결 또는 2'-5' 연결을 통해서일 수 있고, 상기 연결은 3'-5'에서 5'-3'로 또는 2'-5'에서 5'-2'로와 같은 반대 극성을 함유할 수 있다. 다양한 염, 혼합 염 및 유리 산 형태가 또한 포함된다.
- [0096] 뉴클레오타이드 치환체는 뉴클레오타이드와 유사한 기능 특성을 갖지만, 포스페이트 모이어티, 예컨대 펩타이드 핵산(peptide nucleic acid: PNA)을 함유하지 않는 분자를 포함한다. 뉴클레오타이드 치환체는 왓슨-크릭(Watson-Crick) 또는 후그스틴(Hoogsteen) 방식으로 핵산을 인식하지만 포스페이트 모이어티가 아닌 모이어티를 통해 함께 연결되는 분자를 포함한다. 뉴클레오타이드 치환체는 적절한 표적 핵산과 상호작용할 때 이중 나선 유형 구조에 부합할 수 있다.
- [0097] 뉴클레오타이드 치환체는 또한 포스페이트 모이어티 또는 당 모이어티가 대체된 뉴클레오타이드 또는 뉴클레오타이드 유사체를 포함한다. 몇몇 실시형태에서, 뉴클레오타이드 치환체는 표준 인 원자를 함유하지 않을 수 있다. 포스페이트에 대한 치환체는, 예를 들어 짧은 사슬 알킬 또는 사이클로알킬 뉴클레오사이드간 연결, 혼합된 이중원자 및 알킬 또는 사이클로알킬 뉴클레오사이드간 연결, 또는 하나 이상의 짧은 사슬 이중원자성 또는 복소환식 뉴클레오사이드간 연결일 수 있다. 이들은 (부분적으로 뉴클레오사이드의 당 부분으로부터 형성된) 모르폴리노 연결을 갖는 것; 실록산 골격; 설페이드, 설페사이드 및 설펜 골격; 포름아세틸 및 티오포름아세틸 골격; 메틸렌 포름아세틸 및 티오포름아세틸 골격; 알켄 함유 골격; 설펜메이트 골격; 메틸렌이미노 및 메틸렌 하이드라지노 골격; 설포네이트 및 설포아미드 골격; 아미드 골격; 및 혼합된 N, O, S 및 CH₂ 성분 부분을 갖는 기타를 포함한다.
- [0098] 뉴클레오타이드 치환체에서 뉴클레오타이드의 당 및 포스페이트 모이어티 둘 다가, 예를 들어 아미드 유형 연결(아미노에틸글리신)(PNA)에 의해 대체될 수 있다고 또한 이해된다.
- [0099] 예를 들어, 세포 흡수를 증대시키도록 뉴클레오타이드 또는 뉴클레오타이드 유사체에 분자(접합체)의 다른 유형을 또한 연결시킬 수 있다. 접합체는 뉴클레오타이드 또는 뉴클레오타이드 유사체에 화학적으로 연결될 수 있다. 이러한 접합체는, 예를 들어 지질 모이어티, 예컨대 콜레스테롤 모이어티, 콜산, 티오에테르, 예컨대 핵실-S-트리틸티올, 티오킨올레스테롤, 지방족 사슬, 예컨대 도데칸디올 또는 운데실 잔기, 인지질, 예컨대 디-핵사데실-rac-글리세롤 또는 트리에틸암모늄 1,2-디-O-핵사데실-rac-글리세롤-3-H-포스포네이트, 폴리아민 또는 폴리에틸렌 글리콜 사슬, 아다만탄 아세트산, 팔미틸 모이어티, 또는 옥타데실아민 또는 핵실아미노-카보닐-옥시콜레스테롤 모이어티를 포함한다.
- [0100] 본 개시내용은 또한 본원에 개시된 핵산 분자의 임의의 하나 이상을 포함하는 벡터를 제공한다. 몇몇 실시형태에서, 벡터는 본원에 개시된 핵산 분자의 임의의 하나 이상 및 이중성 핵산을 포함한다. 벡터는 핵산 분자를 운반할 수 있는 바이러스성 또는 비바이러스성 벡터일 수 있다. 몇몇 실시형태에서, 벡터는 플라스미드 또는 코스미드(예를 들어, 추가 DNA 분절이 결합될 수 있는 원형 이중-가닥 DNA)이다. 몇몇 실시형태에서, 벡터는 추가 DNA 분절이 바이러스 게놈으로 결합될 수 있는 바이러스성 벡터이다. 몇몇 실시형태에서, 벡터는 이것이 도입되는 숙주 세포에서 자율적으로 복제할 수 있다(예를 들어, 박테리아 복제 기원을 갖는 박테리아 벡터 및 에피솜 포유류 벡터). 몇몇 실시형태에서, 벡터(예를 들어, 비-에피솜 포유류 벡터)는 숙주 세포로의 도입 시 숙주 세포의 게놈으로 통합될 수 있고, 이로써 숙주 게놈과 함께 복제된다. 더욱이, 특정 벡터는 이들이 작동 가능하게 연결된 유전자의 발현을 지시할 수 있다. 이러한 벡터는 본원에서 "재조합 발현 벡터" 또는 "발현 벡터"라 칭해진다. 이러한 벡터는 또한 표적화 벡터(즉, 외인성 도너 서열)일 수 있다.
- [0101] 몇몇 실시형태에서, 본원에 개시된 다양한 유전자 변이체에 의해 암호화된 단백질을 개시된 유전자 변이체를 암호화하는 핵산 분자를 발현 벡터로 삽입하는 것에 의해 발현되어서, 유전자는 발현 제어 서열, 예컨대 전사 및 번역 제어 서열에 작동 가능하게 연결된다. 발현 벡터는 플라스미드, 코스미드, 레트로바이러스, 아데노바이러스

스, 아데노-연관된 바이러스(adeno-associated virus: AAV), 식물 바이러스, 예컨대 꽃양배추 모자이크 바이러스 및 담배 모자이크 바이러스, 효모 인공 염색체(yeast artificial chromosome: YAC), 엡스타인-바(Epstein-Barr: EBV)-유래 에피솜 및 당해 분야에 공지된 다른 발현 벡터를 포함하지만, 이들로 제한되지는 않는다. 몇몇 실시형태에서, 개시된 유전자 변이체를 포함하는 핵산 분자는 벡터에 결합될 수 있어서, 벡터 내의 전사 및 번역 제어 서열은 유전자 변이체의 전사 및 번역을 조절하는 의도된 기능을 제공한다. 발현 벡터 및 발현 제어 서열은 사용되는 발현 숙주 세포와 맞도록 선택된다. 개시된 유전자 변이체를 포함하는 핵산 서열은 별개의 벡터 내로 또는 변이체 유전 정보와 동일한 발현 벡터 내로 삽입될 수 있다. 개시된 유전자 변이체를 포함하는 핵산 서열은 표준 방법(예를 들어, 개시된 유전자 변이체를 포함하는 핵산 및 벡터에서의 상보적 제한 부위에서의 결합, 또는 제한 부위가 존재하지 않는 경우에는 평활 말단 결합(blunt end ligation)에 의해 발현 벡터로 삽입될 수 있다.

- [0102] 개시된 유전자 변이체를 포함하는 핵산 서열 이외에, 재조합 발현 벡터는 숙주 세포에서 유전자 변이체의 발현을 제어하는 조절 서열을 보유할 수 있다. 조절 서열의 선택을 포함하여 발현 벡터의 설계는 형질전환되는 숙주 세포의 선택, 원하는 단백질의 발현의 수준 등과 같은 인자에 따라 달라질 수 있다. 포유류 숙주 세포 발현을 위한 원하는 조절 서열은, 예를 들어 포유류 세포에서 단백질 발현의 높은 수준을 지시하는 바이러스 요소, 예컨대 레트로바이러스 LTR, 사이토메갈로바이러스(cytomegalovirus: CMV)(예컨대, CMV 프로모터/인핸서), 유인원 바이러스 40(Simian Virus 40: SV40)(예컨대, SV40 프로모터/인핸서), 아데노바이러스(예를 들어, 아데노바이러스 주요 후기 프로모터(adenovirus major late promoter: AdMLP))로부터 유래된 프로모터 및/또는 인핸서, 폴리오마 및 강한 포유류 프로모터, 예컨대 천연 면역글로불린 및 액틴 프로모터를 포함할 수 있다. 박테리아 세포 또는 진균 세포(예를 들어, 효모 세포)에서 폴리펩타이드를 발현시키는 방법은 또한 널리 공지되어 있다.
- [0103] 프로모터는, 예를 들어 구성적 활성 프로모터, 조건적 프로모터, 유도성 프로모터, 시간적으로 제한된 프로모터(예를 들어, 발생 조절된 프로모터), 또는 공간적으로 제한된 프로모터(예를 들어, 세포-특이적 또는 조직-특이적 프로모터)일 수 있다. 프로모터의 예는, 예를 들어 WO 제2013/176772호에서 발견될 수 있다.
- [0104] 유도성 프로모터의 예는, 예를 들어 화학적으로 조절된 프로모터 및 물리적으로 조절된 프로모터를 포함한다. 화학적으로 조절된 프로모터는, 예를 들어 알코올-조절된 프로모터(예를 들어, 알코올 탈수소효소(alcohol dehydrogenase)(alcA) 유전자 프로모터), 테트라사이클린-조절된 프로모터(예를 들어, 테트라사이클린 반응성 프로모터, 테트라사이클린 조작자 서열(tetO), tet-On 프로모터 또는 tet-Off 프로모터), 스테로이드 조절된 프로모터(예를 들어, 랫트 글루코코르티코이드 리셉터, 에스트로겐 리셉터의 프로모터 또는 에크디손 리셉터의 프로모터) 또는 금속-조절된 프로모터(예를 들어, 금속단백질 프로모터)를 포함한다. 물리적으로 조절된 프로모터는, 예를 들어 온도-조절된 프로모터(예를 들어, 열 충격 프로모터) 및 광-조절된 프로모터(예를 들어, 광-유도성 프로모터 또는 광-억제성 프로모터)를 포함한다.
- [0105] 조직-특이적 프로모터는, 예를 들어 뉴런-특이적 프로모터, 교세포-특이적 프로모터, 근육 세포-특이적 프로모터, 심장 세포-특이적 프로모터, 신장 세포-특이적 프로모터, 골 세포-특이적 프로모터, 내피 세포-특이적 프로모터, 또는 면역 세포-특이적 프로모터(예를 들어, B 세포 프로모터 또는 T 세포 프로모터)일 수 있다.
- [0106] 발생 조절된 프로모터는, 예를 들어 오직 배아 발생 단계 동안에만 또는 오직 성체 세포에서만 활성인 프로모터를 포함한다.
- [0107] 개시된 유전자 변이체를 포함하는 핵산 서열 및 조절 서열 이외에, 재조합 발현 벡터는 추가 서열, 예컨대 숙주 세포에서 벡터의 복제를 조절하는 서열(예를 들어, 복제 기원) 및 선택 가능한 마커 유전자를 보유할 수 있다. 선택 가능한 마커 유전자는 벡터가 도입되는 숙주 세포의 선택을 수월하게 할 수 있다(예를 들어, 미국 특허 제 4,399,216호; 제4,634,665호; 및 제5,179,017호 참조). 예를 들어, 선택 가능한 마커 유전자는 벡터가 도입된 숙주 세포에서 약물, 예컨대 G418, 하이그로마이신 또는 메토평렉세이트에 내성을 부여할 수 있다. 예시적인 선택 가능한 마커 유전자는 (dhfr-숙주 세포에서 메토평렉세이트 선택/증폭과 사용하기 위한) 디히드로엽산 환원 효소(dihydrofolate reductase: DHFR) 유전자, (G418 선택을 위한) neo 유전자, 및 글루타메이트 합성효소(glutamate synthetase: GS) 유전자를 포함하지만, 이들로 제한되지는 않는다.
- [0108] 추가 벡터는, 예를 들어 2016년 7월 28일에 제출된 미국 가출원 제62/367,973호에 기재되어 있으며, 상기 문헌은 본원에 그 전문이 참고로 인용된다.
- [0109] 본 개시내용은 또한 본원에 개시된 단리된 핵산 분자, 게놈 DNA 분자, cDNA 분자 또는 mRNA 분자 중 임의의 하나 이상을 포함하는 조성물을 제공한다. 몇몇 실시형태에서, 조성물은 약제학적 조성물이다.

- [0110] 본 개시내용은 또한 변이체 SLC14A1 폴리펩타이드를 제공한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 기능 소실 폴리펩타이드 또는 기능 부분 소실 폴리펩타이드이다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13 또는 서열번호 14를 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0111] 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 아미노산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하되, 단, 변이체 SLC14A1 폴리펩타이드는 서열번호 13에 따른 아미노산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0112] 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 14에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 14에 따른 아미노산 서열을 포함하거나 또는 이들로 이루어진다. 몇몇 실시형태에서, 변이체 SLC14A1 폴리펩타이드는 서열번호 14에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하되, 단, 변이체 SLC14A1 폴리펩타이드는 서열번호 14에 따른 아미노산 서열을 포함하지 않거나 또는 이들로 이루어지지 않는다.
- [0113] 본 개시내용은 또한 본원에 개시된 임의의 폴리펩타이드의 단편을 제공한다. 몇몇 실시형태에서, 단편은 암호화된 폴리펩타이드(예컨대, 서열번호 13 및/또는 서열번호 14의 아미노산 서열을 갖는 폴리펩타이드)의 적어도 약 10개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 55개, 적어도 약 60개, 적어도 약 65개, 적어도 약 70개, 적어도 약 75개, 적어도 약 80개, 적어도 약 85개, 적어도 약 90개, 적어도 약 95개, 적어도 약 100개, 적어도 약 150개, 적어도 약 200개, 적어도 약 250개, 적어도 약 300개 또는 적어도 약 350개의 인접 아미노산 잔기를 포함한다. 이와 관련하여, 더 긴 단편이 더 짧은 것에 비해 바람직하다. 몇몇 실시형태에서, 단편은 암호화된 폴리펩타이드의 적어도 약 10개, 적어도 약 15개, 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 55개, 적어도 약 60개, 적어도 약 65개, 적어도 약 70개, 적어도 약 75개, 적어도 약 80개, 적어도 약 85개, 적어도 약 90개, 적어도 약 95개 또는 적어도 약 100개의 인접 아미노산 잔기를 포함한다. 이와 관련하여, 더 긴 단편이 더 짧은 것에 비해 바람직하다.
- [0114] 본 개시내용은 또한 변이체 SLC14A1 폴리펩타이드를 포함하는 단리된 폴리펩타이드를 포함하는 이합체를 제공하고, 여기서 폴리펩타이드는 본원에 개시된 임의의 폴리펩타이드로부터 선택된다.
- [0115] 몇몇 실시형태에서, 본원에 개시된 단리된 폴리펩타이드는 이중성 폴리펩타이드 또는 이중성 분자 또는 표지에 연결되거나 또는 융합되고, 이의 많은 예는 본원 어딘가에 개시되어 있다. 예를 들어, 단백질은 이중성 폴리펩타이드에 융합될 수 있어서 증가된 또는 감소된 안정성을 제공할 수 있다. 융합된 도메인 또는 이중성 폴리펩타이드는 폴리펩타이드 내에서 N-말단에, C-말단에 또는 내부에 위치할 수 있다. 융합 파트너는, 예를 들어 T 헬퍼 에피토프를 제공하는 것을 도울 수 있거나(면역학적 융합 파트너), 또는 천연 재조합 폴리펩타이드보다 더 높은 수율로 단백질을 발현하는 것을 도울 수 있다(발현 인핸서). 소정의 융합 파트너는 면역학적 융합 파트너 및 발현 증대 융합 파트너 둘 다이다. 다른 융합 파트너는 폴리펩타이드의 용해도를 증가시키도록 또는 원하는 세포내 구획에 폴리펩타이드를 표적화하는 것을 수월하게 하도록 선택될 수 있다. 몇몇 융합 파트너는 폴리펩타이드의 정제를 수월하게 하는 친화도 태그를 포함한다.

- [0116] 몇몇 실시형태에서, 융합 단백질은 이중성 분자에 직접 융합되거나, 링커, 예컨대 펩타이드 링커를 통해 이중성 분자에 연결된다. 적합한 펩타이드 링커 서열은, 예를 들어 하기 인자에 기초하여 선택될 수 있다: 1) 가요성 연장된 입체형태(conformation)를 채택하는 능력; 2) 제1 및 제2 폴리펩타이드에서 기능적 에피토프와 상호작용하는 2차 구조의 채택에 대한 저항; 및 3) 폴리펩타이드 기능적 에피토프와 반응하는 소수성 또는 하전된 잔기의 결여. 예를 들어, 펩타이드 링커 서열은 Gly, Asn 및 Ser 잔기를 함유할 수 있다. 거의 중성인 다른 아미노산, 예컨대 Thr 및 Ala는 또한 링커 서열에서 사용될 수 있다. 링커로서 유용하게 사용될 수 있는 아미노산 서열은, 예를 들어 문헌[Maratea *et al.*, *Gene*, 1985, 40, 39-46; Murphy *et al.*, *Proc. Natl. Acad. Sci. USA*, 1986, 83, 8258-8262]; 및 미국 특허 제4,935,233호 및 제4,751,180호에 개시된 것을 포함한다. 링커 서열은 일반적으로, 예를 들어 1개 내지 약 50개의 아미노산의 길이일 수 있다. 제1 및 제2 폴리펩타이드가 기능적 도메인을 분리시키고 입체 장애를 방지하도록 사용될 수 있는 비필수 N 말단 아미노산 영역을 가질 때 링커 서열은 일반적으로 필요하지 않다.
- [0117] 몇몇 실시형태에서, 폴리펩타이드는 세포-침투 도메인에 작동 가능하게 연결된다. 예를 들어, 세포-침투 도메인은 HIV-1 TAT 단백질, 인간 B형 간염 바이러스로부터의 TLM 세포-침투 모티프, MPG, Pep-1, VP22, 단순 포진 바이러스로부터의 세포-침투 펩타이드 또는 폴리아르기닌 펩타이드 서열로부터 유래될 수 있다. 예를 들어, WO 제2014/089290호를 참조한다. 세포-침투 도메인은 N 말단에, C 말단에 또는 단백질 내의 임의의 곳에 위치할 수 있다.
- [0118] 몇몇 실시형태에서, 폴리펩타이드는 추적 또는 정제의 용이를 위해 이중성 폴리펩타이드, 예컨대 형광 단백질, 정제 태그 또는 에피토프 태그에 작동 가능하게 연결된다. 형광 단백질의 예는 녹색 형광 단백질(예를 들어, GFP, GFP-2, tagGFP, turboGFP, eGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질(예를 들어, YFP, eYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), 청색 형광 단백질(예를 들어, eBFP, eBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire), 청록색(cyan) 형광 단백질(예를 들어, eCFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), 적색 형광 단백질(예를 들어, mKate, mKate2, mPlum, DsRed monomer, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 오렌지색 형광 단백질(예를 들어, mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato), 및 임의의 다른 적합한 형광 단백질을 포함하지만, 이들로 제한되지는 않는다. 태그의 예는 글루타티온-S-전환효소(glutathione-S-transferase: GST), 키틴 결합 단백질(chitin binding protein: CBP), 말토스 결합 단백질, 티오레독신(thioredoxin: TRX), 폴리(NANP), 탠덤 친화도 정제(tandem affinity purification: TAP) 태그, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, 혈구응집소(hemagglutinin: HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 히스티딘(histidine: His), 바이오틴 카복실 운반 단백질(biotin carboxyl carrier protein: BCCP) 및 칼모둘린을 포함하지만, 이들로 제한되지는 않는다. 몇몇 실시형태에서, 이중성 분자는 면역글로불린 Fc 도메인, 펩타이드 정제 태그, 형질도입 도메인, 폴리(에틸렌 글리콜), 폴리시알산 또는 글리콜산이다.
- [0119] 몇몇 실시형태에서, 단리된 폴리펩타이드는 비천연 또는 변형된 아미노산 또는 펩타이드 유사체를 포함한다. 예를 들어, 많은 D-아미노산 또는 자연 발생적 아미노산과 상이한 관능성 치환기를 갖는 아미노산이 있다. 자연 발생적 펩타이드의 반대의 입체이성질체, 및 펩타이드 유사체의 입체이성질체가 개시되어 있다. 이들 아미노산은, 부위-특이적 방식으로 유사체 아미노산을 펩타이드 사슬 내로 삽입하도록, tRNA 분자를 선택된 아미노산으로 하전시키고, 예를 들어 앰버 코돈을 사용하는 유전자 작제물을 조작함으로써, 폴리펩타이드 사슬 내로 용이하게 혼입될 수 있다.
- [0120] 몇몇 실시형태에서, 단리된 폴리펩타이드는 펩타이드 모방체이며, 이들은 펩타이드를 닮도록 제조될 수 있지만 천연 펩타이드 연결을 통해 연결되어 있지 않다. 예를 들어, 아미노산 또는 아미노산 유사체를 위한 연결은 $-CH_2NH-$, $-CH_2S-$, $-CH_2-$, $-CH=CH-$ (시스 및 트랜스), $-COCH_2-$, $-CH(OH)CH_2-$ 및 $-CHHSO-$ 를 포함하지만, 이들로 제한되지는 않는다. 펩타이드 유사체, 예컨대 b-알라닌, g아미노부티르산 등은 결합 원자 사이에 1개보다 많은 원자를 가질 수 있다. 아미노산 유사체 및 펩타이드 유사체는 대개 증대된 또는 바람직한 특성, 예컨대 더 경제적인 생산, 더 큰 화학 안정성, 증대된 약물학적 특성(반감기, 흡수, 역가, 효능 등), 변경된 특이성(예를 들어, 생물학적 활성의 광역 스펙트럼), 감소된 항원성 및 다른 바람직한 특성을 갖는다.
- [0121] 몇몇 실시형태에서, 단리된 폴리펩타이드는 D-아미노산을 포함하는데, D 아미노산은 펩티다제에 의해 인식되지 않기 때문에 더 안정한 펩타이드를 생성하도록 사용될 수 있다. 동일한 유형의 D-아미노산에 의한 공통 서열의 하나 이상의 아미노산의 시스템 치환(예를 들어, L-리신 대신에 D-리신)은 더 안정한 펩타이드를 생성하도록 사

용될 수 있다. 시스템인 잔기는 2개 이상의 펩타이드를 함께 고리화하거나 또는 부착시키도록 사용될 수 있다. 이것은 특정한 입체형태로 펩타이드를 구축시키기 위해 유리할 수 있다(예를 들어, Rizo and Gierasch, *Ann. Rev. Biochem.*, 1992, 61, 387 참조).

- [0122] 본 개시내용은 또한 본원에 개시된 임의의 폴리펩타이드를 암호화하는 핵산 분자를 제공한다. 이것은 특이적 폴리펩타이드 서열과 관련된 모든 축퇴성 서열(하나의 특정한 폴리펩타이드 서열을 암호화하는 서열을 갖는 모든 핵산, 및 단백질 서열의 개시된 변이체 및 유도체를 암호화하는 축퇴성 핵산을 포함하는 핵산)을 포함한다. 이와 같이, 각각의 특정한 핵산 서열이 본원에 세세히 기재되지 않을 수 있지만, 각각의 및 모든 서열은 사실 개시된 폴리펩타이드 서열에 걸쳐 본원에 개시되고 기재된다.
- [0123] 핵산 내의 핵산 서열 또는 폴리펩타이드 내의 아미노산 서열의 특정한 스트레치 사이의 퍼센트 동일성(또는 퍼센트 상보성)은 스미스(Smith) 및 워터만(Waterman)(*Adv. Appl. Math.*, 1981, 2, 482-489)의 알고리즘을 이용하는 디폴트 설정을 이용하여 BLAST 프로그램(기본 국소 정렬 조사 도구) 및 PowerBLAST 프로그램(Altschul *et al.*, *J. Mol. Biol.*, 1990, 215, 403-410; Zhang and Madden, *Genome Res.*, 1997, 7, 649-656)을 사용하여 또는 Gap 프로그램(Wisconsin Sequence Analysis Package, 유닉스용 버전 8, Genetics Computer Group(위스콘신 주 매디슨 유니버시티 리서치 파크))을 사용함으로써 일상적으로 결정될 수 있다. 본원에서, 퍼센트 서열 동일성을 언급하는 경우, 서열 동일성의 더 높은 백분율이 더 낮은 것에 비해 바람직하다.
- [0124] 본 개시내용은 또한 본원에 개시된 임의의 하나 이상의 핵산 분자 및/또는 임의의 하나 이상의 폴리펩타이드 및 담체 및/또는 부형제를 포함하는 조성물을 제공한다. 몇몇 실시형태에서, 담체는 핵산 분자 및/또는 폴리펩타이드의 안정성을 증가(예를 들어, 분해 생성물이 출발 핵산 또는 단백질의 0.5 중량% 미만과 같은 한계치 아래에 있는 주어진 저장 조건(예를 들어, -20°C, 4°C 또는 주위 온도) 하에 기간의 연장; 또는 생체내 안정성의 증가)시킨다. 담체의 예는 폴리(락트산)(PLA) 마이크로구, 폴리(D,L-락트산-코-글리콜산)(PLGA) 마이크로구, 리포솜, 마이셀, 인버스 마이셀, 지질 코클레이트 및 지질 미세소관을 포함하지만, 이들로 제한되지는 않는다. 담체는 완충 염 용액, 예컨대 PBS, HBSS 등을 포함할 수 있다.
- [0125] 본 개시내용은 또한 본원에 개시된 임의의 폴리펩타이드 또는 이의 단편을 제조하는 방법을 제공한다. 이러한 폴리펩타이드 또는 이의 단편은 임의의 적합한 방법에 의해 제조될 수 있다. 예를 들어, 폴리펩타이드 또는 이의 단편은 이러한 폴리펩타이드 또는 이의 단편을 암호화하는 핵산 분자(예를 들어, 재조합 발현 벡터)를 포함하는 숙주 세포로부터 제조될 수 있다. 이러한 방법은 폴리펩타이드 또는 이의 단편을 제조하기에 충분한 조건 하에 폴리펩타이드 또는 이의 단편을 암호화하는 핵산 분자(예를 들어, 재조합 발현 벡터)를 포함하는 숙주 세포를 배양하여서 폴리펩타이드 또는 이의 단편을 제조하는 단계를 포함할 수 있다. 핵산은 숙주 세포에서 활성인 프로모터에 작동 가능하게 연결될 수 있고, 배양은 핵산이 발현되는 조건하에 수행될 수 있다. 이러한 방법은 발현된 폴리펩타이드 또는 이의 단편을 회수하는 단계를 추가로 포함할 수 있다. 회수는 폴리펩타이드 또는 이의 단편을 정제하는 것을 추가로 포함할 수 있다.
- [0126] 단백질 발현에 적합한 시스템의 예는 숙주 세포, 예를 들어 박테리아 세포 발현 시스템(예를 들어, 에스체리치아 콜라이(*Escherichia coli*)), 락토코커스 락티스(*Lactococcus lactis*)), 효모 세포 발현 시스템(예를 들어, 사카로마이세스 세레비시아에(*Saccharomyces cerevisiae*)), 피치아 파스토리스(*Pichia pastoris*)), 곤충 세포 발현 시스템(예를 들어, 바콜로바이러스 매개된 단백질 발현) 및 포유류 세포 발현 시스템을 포함한다.
- [0127] 폴리펩타이드 또는 이의 단편을 암호화하는 핵산 분자의 예는 본원에 다른 곳에 더 자세히 개시되어 있다. 몇몇 실시형태에서, 핵산 분자는 숙주 세포에서 발현에 코돈 최적화된다. 몇몇 실시형태에서, 핵산 분자는 숙주 세포에서 활성인 프로모터에 작동 가능하게 연결된다. 프로모터는 이중성 프로모터(예를 들어, 자연 발생적 프로모터가 아닌 프로모터)일 수 있다. 에스체리치아 콜라이에 적합한 프로모터의 예는 아라비노스, *lac*, *tac* 및 T7 프로모터를 포함하지만, 이들로 제한되지는 않는다. 락토코커스 락티스에 적합한 프로모터의 예는 P170 및 니신 프로모터를 포함하지만, 이들로 제한되지는 않는다. 사카로마이세스 세레비시아에 적합한 프로모터의 예는 구성적 프로모터, 예컨대 알코올 탈수소효소(alcohol dehydrogenase: ADHI) 또는 엔올라제(enolase: ENO) 프로모터 또는 유도성 프로모터, 예컨대 PHO, CUP1, GAL1 및 G10을 포함하지만, 이들로 제한되지는 않는다. 피치아 파스토리스에 적합한 프로모터의 예는 알코올 옥시다제 I(AOX I) 프로모터, 글리세르알데하이드 3 포스포이트 탈수소효소(GAP) 프로모터 및 글루타티온 의존적 포름알데하이드 탈수소효소(FLDI) 프로모터를 포함하지만, 이들로 제한되지는 않는다. 바콜로바이러스-매개된 시스템에 적합한 프로모터의 예는 후기 바이러스 강한 폴리헤드린 프로모터이다.
- [0128] 몇몇 실시형태에서, 핵산 분자는 단백질 정제를 수월하게 하도록 폴리펩타이드 또는 이의 단편과 인프레임으로

태그를 암호화한다. 태그의 예는 본원의 다른 곳에 개시되어 있다. 이러한 태그는, 예를 들어 파트너 리간드에 결합(예를 들어, 수지에 부동화)할 수 있어서, 태그화된 단백질은 모든 다른 단백질(예를 들어, 숙주 세포 단백질)로부터 분리될 수 있다. 친화도 크로마토그래피, 고성능 액체 크로마토그래피(high performance liquid chromatography: HPLC) 및 크기 배제 크로마토그래피(size exclusion chromatography: SEC)는 발현된 단백질의 순도를 개선하기 위해 이용될 수 있는 방법의 예이다.

[0129] 다른 방법은 또한 폴리펩타이드 또는 이의 단편을 제조하도록 이용될 수 있다. 예를 들어, 2개 이상의 펩타이드 또는 폴리펩타이드는 단백질 화학 기법에 의해 함께 연결될 수 있다. 예를 들어, 펩타이드 또는 폴리펩타이드는 Fmoc(9-플루오렌메틸옥시카보닐) 또는 Boc(*tert*-부틸옥시카르보닐) 화학물질을 사용하여 화학적으로 합성될 수 있다. 이러한 펩타이드 또는 폴리펩타이드는 표준 화학 반응에 의해 합성될 수 있다. 예를 들어, 펩타이드 또는 폴리펩타이드는 합성되고 이의 합성 수지로부터 절단되지 않을 수 있는 한편, 펩타이드 또는 단백질의 다른 단편은 합성되고 후속하여 수지로부터 절단될 수 있어서, 다른 단편에서 기능적으로 차단된 말단 기를 노출시킨다. 펩타이드 축합 반응에 의해, 이들 2개의 단편은 각각 이의 카복실 및 아미노 말단에서 펩타이드 결합을 통해 공유 연결될 수 있다. 대안적으로, 펩타이드 또는 폴리펩타이드는 본원에 기재된 바대로 생체내 독립적으로 합성될 수 있다. 일단 분리되면, 이 독립적인 펩타이드 또는 폴리펩타이드는 유사한 펩타이드 축합 반응을 통해 펩타이드 또는 이의 단편을 형성하도록 연결될 수 있다.

[0130] 몇몇 실시형태에서, 클로닝된 또는 합성 펩타이드 분질의 효소 결합은 비교적 짧은 펩타이드 단편이 더 큰 펩타이드 단편, 폴리펩타이드 또는 전체 단백질 도메인을 생성하도록 연결되게 한다(Abrahmsen *et al.*, *Biochemistry*, 1991, 30, 4151). 대안적으로, 합성 펩타이드의 천연 화학 결합은 더 짧은 펩타이드 단편으로부터 큰 펩타이드 또는 폴리펩타이드를 합성으로 작제하도록 이용될 수 있다. 이 방법은 2-단계 화학 반응으로 구성될 수 있다(Dawson *et al.*, *Science*, 1994, 266, 776-779). 제1 단계는 초기 공유 생성물로서 티오에스테르 연결된 중간체를 생성하도록 아미노 말단 Cys 잔기를 함유하는 또 다른 비보호된 펩타이드 분질과의 비보호된 합성 펩타이드-티오에스테르의 화학선택적 반응일 수 있다. 반응 조건의 변화 없이는, 이 중간체는 결합 부위에서 천연 펩타이드 결합을 형성하도록 자발적인 빠른 분자내 반응을 겪을 수 있다.

[0131] 몇몇 실시형태에서, 화학 결합의 결과가 비자연(비펩타이드) 결합이므로, 비보호된 펩타이드 분질은 화학적으로 연결될 수 있고, 여기서 펩타이드 분질 사이에 결합이 형성된다(Schnolzer *et al.*, *Science*, 1992, 256, 221).

[0132] 몇몇 실시형태에서, 폴리펩타이드는 발현 후 변형, 예를 들어, 글라이코실화, 아세틸화 및 포스포릴화 등, 및 자연 발생적 및 비자연 발생적 둘 다인 당해 분야에 공지된 다른 변형을 보유할 수 있다. 폴리펩타이드는 전체 단백질 또는 이의 하위서열일 수 있다.

[0133] 본 개시내용은 또한, 본원에 개시된 하나 이상의 폴리펩타이드를 암호화할 수 있는 폴리뉴클레오타이드를 포함하는 핵산 분자 또는 이의 보체를 포함하는 재조합 발현 벡터를 포함하는 숙주 세포를 배양하여서 폴리펩타이드를 제조하는 단계를 포함하는, 본원에 개시된 임의의 폴리펩타이드를 제조하는 방법을 제공한다.

[0134] 본 개시내용은 또한, 핵산 분자를 포함하는 벡터를 포함하여, 본원에 개시된 임의의 하나 이상의 핵산 분자, 및/또는 임의의 하나 이상의 폴리펩타이드를 포함하는 세포(예를 들어, 재조합 숙주 세포)를 제공한다. 세포는 시험관내, 생체외 또는 생체내일 수 있다. 핵산 분자는 프로모터 및 다른 조절 서열에 연결될 수 있어서, 이들은 암호화된 단백질을 제조하도록 발현된다. 이러한 세포의 세포주는 추가로 제공된다.

[0135] 몇몇 실시형태에서, 세포는 전능성 세포 또는 다능성 세포(예를 들어, 배아 줄기(embryonic stem: ES) 세포, 예컨대 설치류 ES 세포, 마우스 ES 세포 또는 래트 ES 세포)이다. 전능성 세포는 임의의 세포 유형을 생성시킬 수 있는 미분화 세포를 포함하고, 다능성 세포는 하나 초과의 분화 세포 유형으로 발생할 능력을 보유하는 미분화 세포를 포함한다. 이러한 다능성 및/또는 전능성 세포는, 예를 들어 ES 세포 또는 ES 유사 세포, 예컨대 유도 다능성 줄기(induced pluripotent stem: iPS) 세포일 수 있다. ES 세포는 배아로의 도입 시 배아를 발생시키는 임의의 조직에 기여할 수 있는 배아 유래 전능성 또는 다능성 세포를 포함한다. ES 세포는 배반포의 내부 세포 덩어리로부터 유래될 수 있고, 임의의 3개의 척추동물 배엽(내배엽, 외배엽 및 중배엽)의 세포로 분화할 수 있다. 본 개시내용에 따르면, 배아 줄기 세포는 비인간 배아 줄기 세포일 수 있다.

[0136] 몇몇 실시형태에서, 세포는 1차 체세포 또는 1차 체세포가 아닌 세포이다. 체세포는 배우자, 생식 세포, 생식모 세포 또는 미분화 줄기 세포가 아닌 임의의 세포를 포함할 수 있다. 몇몇 실시형태에서, 세포는 또한 1차 세포일 수 있다. 1차 세포는 유기체, 장기 또는 조직으로부터 직접적으로 분리된 세포 또는 세포의 배양물을 포함한다. 1차 세포는 형질전환되지 않고 불멸이 아닌 세포를 포함한다. 1차 세포는 조직 배양에서 이전에 계대배양되

지 않거나 조직 배양에서 이전에 계대배양되었지만 조직 배양에서 무기한으로 계대배양될 수 없는 유기체, 장기 또는 조직으로부터 수득된 임의의 세포를 포함한다. 이러한 세포는 종래의 기법에 의해 단리될 수 있고, 예를 들어 체세포, 조혈 세포, 내피 세포, 상피 세포, 섬유아세포, 간충직 세포, 각질세포, 멜라닌세포, 단핵구, 단핵 세포, 지방세포, 지방선구세포, 뉴런, 신경교 세포, 간세포, 골격 근원세포 및 평활근 세포를 포함한다. 예를 들어, 1차 세포는 연결 조직, 근육 조직, 신경계 조직 또는 상피 조직으로부터 유래될 수 있다.

[0137] 몇몇 실시형태에서, 세포는 보통 무기한으로 증식하지 않을 수 있지만, 돌연변이 또는 변경으로 인해, 정상 세포 노화가 회피되고, 대신에 분열을 계속 겪을 수 있다. 이러한 돌연변이 또는 변경은 천연 발생할 수 있거나, 의도적으로 유도될 수 있다. 불활화된 세포의 예는 중국 햄스터 난소(Chinese hamster ovary: CHO) 세포, 인간 배아 신장 세포(예를 들어, HEK 293 세포) 및 마우스 배아 섬유아세포 세포(예를 들어, 3T3 세포)를 포함하지만, 이들로 제한되지는 않는다. 불활화된 세포의 많은 유형은 널리 공지되어 있다. 불활화된 또는 1차 세포는 재조합 유전자 또는 단백질을 배양하기 위해 또는 발현시키기 위해 전형적으로 사용되는 세포를 포함한다. 몇몇 실시형태에서, 세포는 분화 세포, 예컨대 간 세포(예를 들어, 인간 간 세포)이다.

[0138] 세포는 임의의 소스 유래될 수 있다. 예를 들어, 세포는 진핵생물 세포, 동물 세포, 식물 세포 또는 진균(예를 들어, 효모) 세포일 수 있다. 이러한 세포는 어류 세포 또는 조류 세포일 수 있거나, 이러한 세포는 포유류 세포, 예컨대 인간 세포, 비인간 포유류 세포, 설치류 세포, 마우스 세포 또는 래트 세포일 수 있다. 포유류는 인간, 비인간 영장류, 원숭이, 유인원, 고양이, 개, 말, 황소, 사슴, 들소, 양, 설치류(예를 들어, 마우스, 래트, 햄스터, 기니아 피그), 가축(예를 들어, 소, 돼지, 양, 수송아지 등; 양, 예컨대 양, 염소 등; 및 돼지, 예컨대 새끼돼지 및 수퇘지)를 포함하지만, 이들로 제한되지는 않는다. 조류는 닭, 칠면조, 타조, 거위, 오리 등을 포함하지만, 이들로 제한되지는 않는다. 길들인 동물 및 농업 동물이 또한 포함된다. "비인간 동물"의 용어는 인간을 배제한다.

[0139] 추가 숙주 세포는, 예를 들어 2016년 7월 28일자에 출원된 미국 가출원 제 62/367,973호에 기재되어 있으며, 이는 본원에 그 전문이 참고로 인용된다.

[0140] 본원에 개시된 핵산 분자 및 폴리펩타이드는 임의의 수단에 의해 세포로 도입될 수 있다. 형질주입 프로토콜, 및 세포로 핵산 또는 단백질을 도입하기 위한 프로토콜은 변할 수 있다. 비제한적인 형질주입 방법은 리포솜, 나노입자, 칼슘, 덴드리머, 및 양이온성 중합체, 예컨대 DEAE-덱스트란 또는 폴리에틸렌이민을 사용한 화학-기반 형질주입 방법을 포함한다. 비화학 방법은 전기천공, 초음파천공(sono-poration) 및 광학 형질주입을 포함한다. 입자 기반 형질주입은 유전자 총 또는 자석 보조 형질주입의 이용을 포함한다. 바이러스 방법은 형질주입에 또한 이용될 수 있다.

[0141] 세포로의 핵산 또는 단백질의 도입은 또한 전기천공에 의해, 세포질내 주사에 의해, 아데노바이러스에 의한, 아데노 연관된 바이러스에 의한, 렌티바이러스에 의한, 레트로바이러스에 의한, 바이러스 감염에 의해, 형질주입에 의해, 지질 매개된 형질주입에 의해 또는 뉴클레오펙션(nucleofection)에 의해 매개될 수 있다. 뉴클레오펙션은 핵산 기질이 세포질로뿐만 아니라 핵막을 통해 핵으로 전달되게 하는 개선된 전기천공 기술이다. 게다가, 본원에 개시된 방법에서의 뉴클레오펙션의 이용은 전형적으로 정기적인 전기천공보다 훨씬 더 적은 세포(예를 들어, 정기적인 전기천공에 의한 700 만개와 비교하여 불과 약 200 만개)를 요한다. 몇몇 실시형태에서, 뉴클레오펙션은 LONZA[®] NUCLEOFECTOR[™] 시스템을 이용하여 수행된다.

[0142] 세포로의 핵산 또는 단백질의 도입은 또한 미량주사에 의해 달성될 수 있다. mRNA의 미량주사는 보통 (예를 들어, 번역 기계로 mRNA를 직접 전달하기 위해) 세포질로 되지만, 단백질 또는 DNA의 미량주사는 보통 핵으로 된다. 대안적으로, 미량주사는 핵 및 세포질 둘 다로의 주사에 의해 수행될 수 있지만: 침은 처음에 핵으로 도입될 수 있고, 제1 양은 주사될 수 있고, 세포로부터 침을 제거하면서 제2 양이 세포질로 주사될 수 있다. 뉴클레아제 제제 단백질이 세포질로 주사되면, 단백질은 핵/전핵으로의 전달을 보장하도록 핵 국제화 신호를 포함할 수 있다.

[0143] 세포로 핵산 또는 단백질을 도입하기 위한 다른 방법은, 예를 들어 벡터 전달, 입자-매개된 전달, 엑소좀-매개된 전달, 지질-나노입자-매개된 전달, 세포-침투-펩타이드-매개된 전달, 또는 이식형-장치-매개된 전달을 포함할 수 있다. 생체내 세포를 변형시키기 위해 대상체에게 핵산 또는 단백질을 투여하는 방법은 본원 다른 곳에 개시되어 있다. 세포로의 핵산 및 단백질의 도입은 유체역학 전달(hydrodynamic delivery: HDD)에 의해 또한 달성될 수 있다.

[0144] 세포로 핵산 또는 단백질을 도입하기 위한 다른 방법은, 예를 들어 벡터 전달, 입자-매개된 전달, 엑소좀-매개

된 전달, 지질-나노입자-매개된 전달, 세포-침투-펩타이드-매개된 전달 또는 이식형-장치-매개된 전달을 포함할 수 있다. 몇몇 실시형태에서, 핵산 또는 단백질은 담체, 예컨대 폴리(락트산)(PLA) 마이크로구, 폴리(D,L-락트산-코-글리콜산)(PLGA) 마이크로구, 리포솜, 마이셀, 인버스 마이셀, 지질 코클레이트 또는 지질 미세소관에서 세포로 도입될 수 있다.

[0145] 본 개시내용은 또한 프로브 및 프라이머를 제공한다. 프로브 및 프라이머의 예는, 예를 들어 상기에 개시되어 있다. 본 개시내용은 본원에 개시된 임의의 핵산 분자에 특이적으로 혼성화하는 핵산 서열을 포함하는 프로브 및 프라이머를 제공한다. 예를 들어, 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 본원에 기재된 임의의 핵산 분자에 혼성화하거나, 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함할 수 있다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 13 또는 서열번호 14에 따른 변이체 SLC14A1 단백질을 암호화하는 핵산 분자에 혼성화하거나, 이들 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 본원에 기재된 임의의 핵산 분자에 혼성화하거나, 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함할 수 있다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 13에 따른 변이체 SLC14A1 단백질을 암호화하는 핵산 분자에 혼성화하거나, 이들 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 본원에 기재된 임의의 핵산 분자에 혼성화하거나, 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함할 수 있다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 14에 따른 변이체 SLC14A1 단백질을 암호화하는 핵산 분자에 혼성화하거나, 이들 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다.

[0146] 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 13에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자에 혼성화하거나, 이 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 13에 따른 아미노산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자에 혼성화하거나, 이 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다.

[0147] 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 14에 따른 아미노산 서열과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자에 혼성화하거나, 이 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머는 서열번호 14에 따른 아미노산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자에 혼성화하거나, 이 핵산 분자의 보체에 혼성화하는 핵산 서열을 포함한다.

[0148] 프로브 또는 프라이머는 임의의 적합한 길이를 포함할 수 있고, 이의 비제한적인 예는 적어도 약 5개, 적어도 약 8개, 적어도 약 10개, 적어도 약 11개, 적어도 약 12개, 적어도 약 13개, 적어도 약 14개, 적어도 약 15개, 적어도 약 16개, 적어도 약 17개, 적어도 약 18개, 적어도 약 19개, 적어도 약 20개, 적어도 약 21개, 적어도 약 22개, 적어도 약 23개, 적어도 약 24개 또는 적어도 약 25개의 뉴클레오타이드의 길이를 포함한다. 바람직한 실시형태에서, 프로브 또는 프라이머는 적어도 약 18개의 뉴클레오타이드의 길이를 포함한다. 프로브 또는 프라이머는 약 10개 내지 약 35개, 약 10개 내지 약 30개, 약 10개 내지 약 25개, 약 12개 내지 약 30개, 약 12개 내지 약 28개, 약 12개 내지 약 24개, 약 15개 내지 약 30개, 약 15개 내지 약 25개, 약 18개 내지 약 30개, 약 18개 내지 약 25개, 약 18개 내지 약 24개, 또는 약 18개 내지 약 22개의 뉴클레오타이드의 길이를 포함할 수 있다. 바람직한 실시형태에서, 프로브 또는 프라이머는 약 18개 내지 약 30개의 뉴클레오타이드의 길이이다.

[0149] 본 개시내용은 또한 변경-특이적 프로브 및 변경-특이적 프라이머를 제공한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열 또는 또는 이의 보체를 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적이고/이거나 혼성화하거나 또는 특이적으로 혼성화하는 핵산 서열 또는 이의 보체

를 포함한다.

- [0150] 본 개시내용의 맥락에서 "특이적으로 혼성화한다"는 프로브 또는 프라이머(예를 들어, 변경-특이적 프로브 또는 변경-특이적 프라이머)가 야생형 SLC14A1 단백질을 암호화하는 핵산 분자에 혼성화하지 않는다는 것을 의미한다. 몇몇 실시형태에서, 변경-특이적 프로브는 서열번호 13에 따른 76번 위치 또는 이의 보체에 상응하는 위치에서 이소류신을 암호화하는 핵산 코돈에 특이적으로 혼성화한다. 몇몇 실시형태에서, 변경-특이적 프라이머 또는 프라이머 쌍은 변이체 SLC14A1 단백질을 암호화하는 핵산 분자의 영역(들)에 특이적으로 혼성화하여서, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈이 이로부터 제조된 임의의 전사체 내에 포함된다. 몇몇 실시형태에서, 변경-특이적 프로브는 서열번호 14에 따른 132번 위치 또는 이의 보체에 상응하는 위치에서 이소류신을 암호화하는 핵산 코돈에 특이적으로 혼성화한다. 몇몇 실시형태에서, 변경-특이적 프라이머 또는 프라이머 쌍은 변이체 SLC14A1 단백질을 암호화하는 핵산 분자의 영역(들)에 특이적으로 혼성화하여서, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈은 이로부터 제조된 임의의 전사체 내에 포함된다.
- [0151] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열 또는 이의 보체를 포함하고, 여기서 그 단백질은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열 또는 이의 보체를 포함하고, 여기서 그 단백질은 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다.
- [0152] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 13과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하는 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 13을 갖는 변이체 SLC14A1 단백질을 암호화하는 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.
- [0153] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하는 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14를 갖는 변이체 SLC14A1 단백질을 암호화하는 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.
- [0154] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 2와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 2에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.
- [0155] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 mRNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 mRNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.

- [illegible]

자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 13을 갖는 변이체 SLC14A1 단백질을 암호화하는 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.

[0162] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 14를 갖는 변이체 SLC14A1 단백질을 암호화하는 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.

[0163] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 9와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 9에 따른 226번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 9에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.

[0164] 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치에서 코돈 AUC를 포함하는 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 10과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하고, 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함한다. 몇몇 실시형태에서, 변경-특이적 프로브 또는 변경-특이적 프라이머는 서열번호 10에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자에 상보적이고/이거나 혼성화하거나, 또는 특이적으로 혼성화하는 핵산 서열을 포함한다.

[0165] 본 개시내용은 또한 적어도 약 15개의 뉴클레오타이드를 포함하고 SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화하는 단리된 변경-특이적 프로브 또는 프라이머를 제공하고, 여기서 변경-특이적 프로브 또는 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 SLC14A1 암호화 핵산 서열의 일부에 상보적인 핵산 서열 또는 이의 보체를 포함한다.

[0166] 본 개시내용은 또한 적어도 약 15개의 뉴클레오타이드를 포함하고, SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화하는 단리된 변경-특이적 프로브 또는 프라이머를 제공하고, 여기서 변경-특이적 프로브 또는 프라이머는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 SLC14A1 암호화 핵산 서열의 일부에 상보적인 핵산 서열 또는 이의 보체를 포함한다.

[0167] 본 개시내용은 또한 서열번호 13의 아미노산 서열을 갖는 SLC14A1 변이체 폴리펩타이드와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 아미노산 서열을 포함하는 단리된 폴리펩타이드를 제공하고, 단, 폴리펩타이드는 서열번호 13에 따른 76번 위치에 상응하는 위치에서

이소류신을 포함한다. 몇몇 실시형태에서, SLC14A1 변이체 폴리펩타이드는 서열번호 13의 아미노산 서열을 포함한다.

- [0168] 본 개시내용은 또한 서열번호 14의 아미노산 서열을 갖는 SLC14A1 변이체 폴리펩타이드와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일한 아미노산 서열을 포함하는 단리된 폴리펩타이드를 제공하고, 단 폴리펩타이드는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, SLC14A1 변이체 폴리펩타이드는 서열번호 14의 아미노산 서열을 포함한다.
- [0169] 본 개시내용은 또한 응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성을 결정하기 위한 본원에 기재된 임의의 단리된 프로브 또는 프라이머 또는 본원에 기재된 단리된 변경-특이적 프로브 또는 프라이머의 용도를 제공한다.
- [0170] 본 개시내용의 프로브 또는 프라이머와 관련하여 상기에 기재된 길이는 또한 본 개시내용의 변경-특이적 프로브 또는 변경-특이적 프라이머에 대해 필요한 부분만 약간 수정하여 적용된다.
- [0171] 본 개시내용은 또한 상기 기재된 바와 같은 변경-특이적 프라이머 중 2개를 포함하는 변경-특이적 프라이머의 쌍을 제공한다.
- [0172] 몇몇 실시형태에서, 프로브 또는 프라이머(예를 들어, 변경-특이적 프로브 또는 변경-특이적 프라이머)는 DNA를 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머(예를 들어, 변경-특이적 프로브 또는 변경-특이적 프라이머)는 RNA를 포함한다. 몇몇 실시형태에서, 프로브 또는 프라이머(예를 들어, 변경-특이적 프로브 또는 변경-특이적 프라이머)는 엄격한 조건, 예컨대 높은 엄격한 조건하에 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 혼성화한다.
- [0173] 몇몇 실시형태에서, 프로브는 표지를 포함한다. 몇몇 실시형태에서, 표지는 형광 표지, 방사선표지 또는 바이오틴이다. 몇몇 실시형태에서, 프로브의 길이는 상기에 기재되어 있다. 대안적으로, 몇몇 실시형태에서, 프로브는 적어도 약 20개, 적어도 약 25개, 적어도 약 30개, 적어도 약 35개, 적어도 약 40개, 적어도 약 45개, 적어도 약 50개, 적어도 약 55개, 적어도 약 60개, 적어도 약 65개, 적어도 약 70개, 적어도 약 75개, 적어도 약 80개, 적어도 약 85개, 적어도 약 90개, 적어도 약 95개 또는 적어도 약 100개의 뉴클레오타이드를 포함하거나 또는 이들로 이루어진다. 프로브(예를 들어, 대립유전자 특이적 프로브)는, 예를 들어 본원에 개시된 임의의 핵산 분자를 검출하기 위해 사용될 수 있다. 바람직한 실시형태에서, 프로브는 적어도 약 18개의 뉴클레오타이드의 길이를 포함한다. 프로브는 약 10개 내지 약 35개, 약 10개 내지 약 30개, 약 10개 내지 약 25개, 약 12개 내지 약 30개, 약 12개 내지 약 28개, 약 12개 내지 약 24개, 약 15개 내지 약 30개, 약 15개 내지 약 25개, 약 18개 내지 약 30개, 약 18개 내지 약 25개, 약 18개 내지 약 24개, 또는 약 18개 내지 약 22개의 뉴클레오타이드의 길이를 포함할 수 있다. 바람직한 실시형태에서, 프로브는 약 18개 내지 약 30개의 뉴클레오타이드의 길이이다.
- [0174] 본 개시내용은 또한 본원에 개시된 프로브의 임의의 하나 이상이 부착된 기질을 포함하는 지지체를 제공한다. 고체 지지체는 본원에 개시된 분자, 예컨대 임의의 프로브가 회합될 수 있는 고체 상태 기질 또는 지지체이다. 고체 지지체의 형태는 어레이이다. 고체 지지체의 또 다른 형태는 어레이 검출기이다. 어레이 검출기는 다수의 상이한 프로브가 어레이, 그리드 또는 다른 조직화된 패턴으로 커플링된 고체 지지체이다.
- [0175] 고체 지지체에서 사용하기 위한 고체-상태 기질은 분자가 커플링될 수 있는 임의의 고체 재료를 포함할 수 있다. 이것은 아크릴아미드, 아가로스, 셀룰로스, 니트로셀룰로스, 유리, 폴리스티렌, 폴리에틸렌 비닐 아세테이트, 폴리프로필렌, 폴리메타크릴레이트, 폴리에틸렌, 폴리에틸렌 옥사이드, 폴리실리케이트, 폴리카보네이트, 테플론, 플루오로카본, 나일론, 실리콘 고무, 폴리언하이드라이드, 폴리글리콜산, 폴리락트산, 폴리오르토에스테르, 폴리프로필푸마레이트, 콜라겐, 글리코사미노글리칸 및 폴리아미노산과 같은 재료를 포함한다. 고체 상태 기질은 박막, 막, 병, 접시, 섬유, 직조 섬유, 성형 중합체, 입자, 비드, 마이크로입자 또는 조합을 포함하는 임의의 유용한 형태를 가질 수 있다. 고체 상태 기질 및 고체 지지체는 다공성 또는 비다공성일 수 있다. 고체 상태 기질에 대한 형태는 미량정량 접시, 예컨대 표준 96웰 유형이다. 몇몇 실시형태에서, 웰마다 하나의 어레이를 보통 함유하는 멀티웰 유리 슬라이드가 사용될 수 있다. 이 특징은 분석법 재현성의 더 높은 제어, 증가된 쓰루풋 및 샘플 취급, 및 자동화의 용이를 허용한다. 몇몇 실시형태에서, 지지체는 마이크로어레이이다.
- [0176] 본원에 개시된 임의의 폴리펩타이드는 하나 이상의 치환(예컨대, 보존적 아미노산 치환), 삽입 또는 결실을 추가로 가질 수 있다. 삽입은, 예를 들어 아미노 또는 카복실 말단 융합, 및 단일 또는 다수의 아미노산 잔기의

서열내 삽입을 포함한다. 공지된 서열을 갖는 DNA에서 미리 결정된 부위에서 치환을 만드는 기법, 예를 들어 M13 프라이머 돌연변이유발 및 PCR 돌연변이유발은 널리 공지되어 있다. 아미노산 치환은 전형적으로 단일 잔기이지만, 한번에 다수의 상이한 위치에서 발생할 수 있고; 삽입은 보통 약 1개 내지 10개의 아미노산 잔기의 차수일 수 있고; 결실은 약 1개 내지 30개의 잔기의 범위일 것이다. 결실 또는 삽입은 인접한 쌍, 즉 2개의 잔기의 결실 또는 2개의 잔기의 삽입에서 이루어질 수 있다. 치환, 결실, 삽입 또는 임의의 이들의 조합은 최종 작제물에 도달하도록 조합될 수 있다. 몇몇 실시형태에서, 돌연변이는 리딩 프레임 밖에 서열을 위치시키지 않고, 2차 mRNA 구조를 생성하는 상보적 영역을 생성하지 않는다.

[0177] 본 개시내용은 또한 조성물을 제조하고 본원에 기재된 방법을 이용하기 위한 키트를 제공한다. 본원에 기재된 키트는 대상체의 샘플에서 하나 이상의 유전자 변이체를 검출하기 위한 분석법 또는 분석법들을 포함한다.

[0178] 몇몇 실시형태에서, 인간 SLC14A1 변이체의 확인을 위한 키트는 상기에 기재된 조성물 및 방법을 사용한다. 몇몇 실시형태에서, 기본 키트는 본원에 개시된 임의의 핵산 분자(예를 들어, 서열번호 2, 서열번호 5, 서열번호 6, 서열번호 9, 및/또는 서열번호 10 등)에서의 유전좌위에 대한 올리고뉴클레오타이드 프라이머 또는 프로브, 예컨대 변경-특이적 프로브 또는 변경-특이적 프라이머의 적어도 하나의 쌍을 갖는 용기를 포함할 수 있다. 키트는 또한 사용 설명서를 선택적으로 포함할 수 있다. 키트는, 예를 들어 증폭된 유전좌위의 각각에 지향된 대립유전자 래더, 증폭을 위한 충분한 분량의 효소, 증폭을 수월하게 하는 증폭 완충제, 효소 활성을 수월하게 하는 2가 양이온 용액, 증폭 동안의 가닥 연장을 위한 dNTP, 전기영동법을 위한 증폭된 재료의 제조를 위한 로딩 용액, 주형 대조군으로서의 게놈 DNA, 재료가 분리 매질에서 기대된 바대로 이동하도록 보장하는 크기 마커, 및 사용자를 가르치고 사용 시 오류를 제한하기 위한 프로토콜 및 매뉴얼 중 하나 이상과 같은 다른 선택적인 키트 성분을 또한 포함할 수 있다. 키트에서의 다양한 시약의 양은 또한 다수의 인자, 예컨대 공정의 최적 감수성에 따라 변할 수 있다. 매뉴얼 적용에서 사용하기 위한 시험 키트 또는 자동화 샘플 준비, 반응 셋업, 검출기 또는 분석기와 사용하기 위한 시험 키트를 제공하는 것은 이 교시내용의 범주 내에 있다.

[0179] 몇몇 실시형태에서, 키트는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 게놈 DNA 분자 또는 이의 보체의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 13 또는 서열번호 14와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는, 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 게놈 DNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 2를 갖는 변이체 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 게놈 DNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다.

[0180] 몇몇 실시형태에서, 키트는 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 2와 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고 서열번호 2에 따른 6963번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 2에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 게놈 DNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다.

[0181] 몇몇 실시형태에서, 키트는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체

리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 10과 적어도 약 90%, 적어도 약 91%, 적어도 약 92%, 적어도 약 93%, 적어도 약 94%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99%의 서열 동일성을 갖고, 서열번호 10에 따른 394번 위치에 상응하는 위치에서 아데닌을 포함하는 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다. 몇몇 실시형태에서, 키트는 서열번호 10에 따른 핵산 서열을 포함하거나 또는 이들로 이루어진 변이체 SLC14A1 cDNA 분자의, 증폭을 위한 올리고뉴클레오타이드 프라이머(예를 들어, 변경-특이적 프라이머), 또는 검출을 위한 적어도 하나의 표지된 올리고뉴클레오타이드 프로브(예를 들어, 변경-특이적 프로브)의 적어도 하나의 쌍을 포함한다.

- [0187] 몇몇 실시형태에서, 본원에 개시된 임의의 키트는 뉴클레오타이드 래더, 프로토콜, 효소(예컨대, 증폭, 예컨대 중합효소 연쇄 반응(polymerase chain reaction: PCR)에 사용되는 효소), dNTP, 완충제, 염 또는 염들, 및 조절 핵산 샘플 중 임의의 하나 이상을 추가로 포함할 수 있다. 몇몇 실시형태에서, 본원에 개시된 임의의 키트는 검출 가능한 표지, 어닐링 반응을 수행하는 데 필요한 제품 및 시약, 및 설명서 중 임의의 하나 이상을 추가로 포함할 수 있다.
- [0188] 몇몇 실시형태에서, 본원에 개시된 키트는 서열번호 2의 6963번 위치에 상응하는 위치에서, 서열번호 5 및/또는 서열번호 9의 226번 위치에 상응하는 위치에서, 또는 서열번호 6 및/또는 서열번호 10의 394번 위치에 상응하는 위치에서 아데닌에 직접적으로 혼성화하는 3' 말단 뉴클레오타이드를 포함하는 프라이머 또는 프로브 또는 변경-특이적 프라이머 또는 변경-특이적 프로브를 포함할 수 있다.
- [0189] 당업자는 사용된 검출 기법이 일반적으로 제한이 아니라는 것을 이해한다. 오히려, 넓은 다양한 검출 수단은 개시된 방법 및 키트의 범주 내에 있고, 단 이들은 앰플리콘의 존재 또는 부재가 결정되게 한다.
- [0190] 몇몇 양상에서, 키트는 본원에 개시된 프라이머 또는 프로브 중 하나 이상을 포함할 수 있다. 예를 들어, 키트는 개시된 유전자 변이체 중 하나 이상에 혼성화하는 하나 이상의 프로브를 포함할 수 있다.
- [0191] 몇몇 양상에서, 키트는 개시된 세포 또는 세포주 중 하나를 포함할 수 있다. 몇몇 양상에서, 키트는 형질전환 세포 또는 세포주를 생성하는 데 필요하는 재료를 포함할 수 있다. 예를 들어, 몇몇 양상에서, 키트는 개시된 유전자 변이체 중 하나 이상을 포함하는 핵산 서열을 포함하는 세포 및 벡터를 포함할 수 있다. 키트는 세포 배양에 대한 배지를 추가로 포함할 수 있다.
- [0192] 본 개시내용은 또한 대상체 인간으로부터의 생물학적 샘플에서 SLC14A1 변이체 게놈 DNA, mRNA, cDNA, 및/또는 폴리펩타이드의 존재를 검출하는 방법을 제공한다. 몇몇 실시형태에서, SLC14A1 변이체 게놈 DNA, mRNA, 및/또는 cDNA는 기능 상실 또는 기능 부분 상실을 갖는 변이체 SLC14A1 폴리펩타이드를 생성시킨다. 집단 내의 유전자 서열 및 이러한 유전자에 의해 암호화된 mRNA 및 단백질이 다형, 예컨대 단일 뉴클레오타이드 다형으로 인해 변할 수 있다고 이해된다. SLC14A1 게놈 DNA, mRNA, cDNA 및 폴리펩타이드에 대해 본원에 제공된 서열은 오직 예시적인 서열이다. SLC14A1 게놈 DNA, mRNA, cDNA, 및 폴리펩타이드에 대한 다른 서열이 또한 가능하다.
- [0193] 본 개시내용은 또한, 대상체로부터 수득된 샘플을 분석하여서, 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부 및/또는 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계를 포함하는, 인간 대상체가 SLC14A1 변이체 핵산 분자를 보유하는지의 여부를 결정하는 방법을 제공한다. 몇몇 실시형태에서, 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우 및/또는 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자가 확인되는 경우, 인간 대상체는 응고 병태 또는 관상 동맥 질환(CAD)을 발생시킬 위험이 감소된 것으로 분류된다. 몇몇 실시형태에서, 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우 및/또는 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 핵산 분자가 확인되는 경우, 인간 대상체는 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류된다. 몇몇 실시형태에서, 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류

및 뇌졸중으로부터 선택된다.

- [0194] 본 개시내용은 또한, 인간 대상체로부터 수득된 샘플에서 분석법을 수행하여서, 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하는 단계를 포함하는, 인간 대상체가 SLC14A1 Val76Ile 단백질 및/또는 SLC14A1 Val132Ile 단백질을 보유하는지의 여부를 결정하는 방법을 제공한다. 몇몇 실시형태에서, 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질이 확인되는 경우 및/또는 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질이 확인되는 경우, 인간 대상체는 응고 병태 또는 관상동맥 질환(CAD)을 발생시킬 위험이 감소된 것으로 분류된다. 몇몇 실시형태에서, 샘플에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이 확인되는 경우 및/또는 샘플에서 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이 확인되는 경우, 인간 대상체는 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로 분류된다. 몇몇 실시형태에서, 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류 및 뇌졸중으로부터 선택된다. 몇몇 실시형태에서, 효소-결합 면역흡착 분석법(enzyme-linked immunosorbent assay: ELISA)은 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하기 위해 사용된다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0195] 생물학적 샘플은 대상체로부터의 임의의 세포, 조직 또는 생체액으로부터 유래될 수 있다. 샘플은 임의의 임상적으로 관련된 조직, 예컨대 골수 샘플, 종양 생검, 세침 흡입액, 또는 체액, 예컨대 혈액, 치은열구액, 혈장, 혈청, 림프, 복수액, 낭종액 또는 소변의 샘플을 포함할 수 있다. 몇몇 경우에, 샘플은 협착 면봉을 포함한다. 본원에 개시된 방법에서 사용된 샘플은 분석법 포맷, 검출 방법의 성질, 및 샘플로서 사용되는 조직, 세포 또는 추출물을 기초하여 변할 수 있다. 생물학적 샘플은 이용되는 분석법에 따라 다르게 처리될 수 있다. 예를 들어, 변이체 SLC14A1 핵산 분자를 검출할 때, 게놈 DNA에 대해 샘플을 단리하거나 농후화하도록 설계된 예비 처리가 이용될 수 있다. 다양한 공지된 기법은 이 목적에 이용될 수 있다. 변이체 SLC14A1 mRNA의 수준을 검출할 때, 상이한 기법은 생물학적 샘플을 mRNA로 농후화시키도록 이용될 수 있다. mRNA의 존재 또는 수준 또는 특정한 변이체 게놈 DNA 유전자좌위의 존재를 검출하기 위한 다양한 방법이 이용될 수 있다.
- [0196] 본 개시내용은 또한 인간 대상체에서의 SLC14A1 변이체 핵산 분자를 검출하는 방법을 제공하고, 여기서 SLC14A1 변이체 핵산 분자는 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화한다. 몇몇 실시형태에서, 인간 대상체에서의 SLC14A1 변이체 핵산 분자를 검출하는 방법은 대상체로부터 수득된 샘플을 분석하여서, 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열을 포함하는지의 여부 또는 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계를 포함한다.
- [0197] 본 개시내용은 또한 인간 대상체에서 변이체 SLC14A1 단백질의 존재 또는 부재를 검출하는 방법을 제공하고, 여기서 SLC14A1 변이체 단백질은 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질이다. 몇몇 실시형태에서, 변이체 SLC14A1 단백질의 존재 또는 부재를 검출하는 방법은 생물학적 샘플에서의 단백질의 적어도 일부를 서열분석하여서, 그 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 아미노산 서열을 포함하는지의 여부를 결정하는 단계를 포함한다.
- [0198] 몇몇 실시형태에서, 본 개시내용은, 생물학적 샘플에서의 핵산의 적어도 일부를 서열분석하여서, 그 핵산이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계를 포함하는, 변이체 SLC14A1 핵산 분자의 존재 또는 부재를 검출하는 방법을 제공한다. 본원에 개시된 임의의 변이체 핵산 분자는 본원에 기재된 임의의 프로브 및 프라이머를 사용하여 검출될 수 있다.
- [0199] 몇몇 실시형태에서, 대상체에서 응고 병태-연관된 변이체 SLC14A1 핵산 분자 또는 CAD-연관된 변이체 SLC14A1 핵산 분자(예를 들어, 게놈 DNA, mRNA 또는 cDNA)의 존재 또는 부재를 검출하는 방법은 대상체로부터 수득된 생물학적 샘플에서 분석법을 수행하는 단계를 포함하고, 그 분석법은 생물학적 샘플에서의 핵산 분자가 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화하는 변이체 SLC14A1 핵산 분자를 포함하는지의 여부를 결정한다.

- [0200] 몇몇 실시형태에서, 대상체에서 응고 병태-연관된 변이체 SLC14A1 핵산 분자 또는 CAD-연관된 변이체 SLC14A1 핵산 분자(예를 들어, 게놈 DNA, mRNA 또는 cDNA)의 존재 또는 부재를 검출하는 방법은 대상체로부터 수득된 생물학적 샘플에서 분석법을 수행하는 단계를 포함하고, 그 분석법은 생물학적 샘플에서의 핵산 분자가 본원에 개시된 임의의 변이체 SLC14A1 핵산 서열(예를 들어, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자)을 포함하는지의 여부를 결정한다. 몇몇 실시형태에서, 생물학적 샘플은 세포 또는 세포 용해물을 포함한다. 이러한 방법은, 예를 들어 SLC14A1 게놈 DNA 또는 mRNA를 포함하는 대상체로부터의 생물학적 샘플을 수득하는 단계, 및 mRNA의 경우에 선택적으로 mRNA를 cDNA로 역전사시키는 단계, 및 SLC14A1 게놈 DNA, mRNA 또는 cDNA의 위치가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는지의 여부를 결정하는 생물학적 샘플에서 분석법을 수행하는 단계를 추가로 포함할 수 있다. 이러한 분석법은, 예를 들어 특정한 SLC14A1 핵산 분자의 이들 위치의 식별을 결정하는 단계를 포함할 수 있다. 몇몇 실시형태에서, 대상체는 인간이다.
- [0201] 몇몇 실시형태에서, 분석법은 대상체로부터의 생물학적 샘플에서 핵산 분자의 SLC14A1 게놈 DNA 서열의 적어도 일부를 서열분석하되, 서열분석된 일부는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하거나, 서열분석된 일부는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하는 것인 단계; 대상체로부터의 생물학적 샘플에서 핵산 분자의 SLC14A1 mRNA 서열의 적어도 일부를 서열분석하되, 서열분석된 일부는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하거나, 서열분석된 일부는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치를 포함하는 것인 단계; 또는 대상체로부터의 생물학적 샘플에서 핵산 분자의 SLC14A1 cDNA 서열의 적어도 일부를 서열분석하되, 서열분석된 일부는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하거나, 서열분석된 일부는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하는 것인 단계를 포함한다.
- [0202] 몇몇 실시형태에서, 분석법은 a) 생물학적 샘플을, i) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 게놈 DNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 게놈 DNA 서열의 일부; ii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 mRNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 mRNA 서열의 일부; 또는 iii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 cDNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 상응하는 위치에서 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 cDNA 서열의 일부에 혼성화하는 프라이머와 접촉시키는 단계; b) 적어도 i) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 게놈 DNA 서열의 위치 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 게놈 DNA 서열의 위치; ii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 mRNA 서열의 위치 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 mRNA 서열의 위치; 또는 iii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 cDNA 서열의 위치 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈에 멀리 있는 뉴클레오타이드 위치에 상응하는 SLC14A1 cDNA 서열의 위치에 걸쳐 프라이머를 연장시키는 단계; 및 c) 프라이머의 연장 산물이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는지의 여부를 결정하거나, 프라이머의 연장 산물이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는지의 여부를 결정하는 단계를 포함한다. 몇몇 실시형태에서, 오직 SLC14A1 게놈 DNA가 분석된다. 몇몇 실시형태에서, 오직 SLC14A1 mRNA가 분석된다. 몇몇 실시형태에서, SLC14A1 mRNA로부터 수득된 SLC14A1 cDNA가 오직 분석된다.

- [0203] 몇몇 실시형태에서, 분석법은 a) 생물학적 샘플을, i) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 게놈 DNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 게놈 DNA 서열의 일부; ii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 mRNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 mRNA 서열의 일부; 또는 iii) 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 cDNA 서열의 일부 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는 SLC14A1 cDNA 서열의 일부에 혼성화하는 변경-특이적 프라이머와 접촉시키는 단계; b) 변경-특이적 중합효소 연쇄 반응 기법을 이용하여 프라이머를 연장시키는 단계; 및 c) 연장이 발생하는지의 여부를 결정하는 단계를 포함한다. 변경-특이적 중합효소 연쇄 반응 기법은 핵산 서열에서의 결실과 같은 돌연변이를 검출하도록 이용될 수 있다. DNA 중합효소는 주형과의 미스매치가 존재할 때 연장되지 않으므로, 변경-특이적 프라이머가 사용된다. 기본적인 변경-특이적 중합효소 연쇄 반응 기법의 다수의 변형은 당업자의 재량에 따른다.
- [0204] 변경-특이적 프라이머는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열 또는 핵산 서열에 대한 보체에 상보적인 핵산 서열을 포함할 수 있다. 예를 들어, 변경-특이적 프라이머는 서열번호 13을 암호화하는 핵산 서열에 상보적인 핵산 서열 또는 이 핵산 서열에 대한 보체를 포함할 수 있다. 대안적으로, 변경-특이적 프라이머는 서열번호 14를 암호화하는 핵산 서열에 상보적인 핵산 서열 또는 이 핵산 서열에 대한 보체를 포함할 수 있다. 변경-특이적 프라이머는 바람직하게는, 핵산 서열이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화할 때, 변이체 SLC14A1 단백질을 암호화하는 핵산 서열에 특이적으로 혼성화한다.
- [0205] 몇몇 실시형태에서, 분석법은 샘플에서 핵산 분자의 SLC14A1 게놈 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치를 포함하는 것인 단계; 샘플에서 핵산 분자의 SLC14A1 mRNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 샘플에서 핵산 분자의 SLC14A1 mRNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계; 샘플에서 핵산 분자의 SLC14A1 cDNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 및/또는 샘플에서 핵산 분자의 SLC14A1 cDNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계를 포함한다.
- [0206] 몇몇 실시형태에서, 분석법은 a) 샘플을, i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 게놈 서열의 일부; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA의 위치에 가장 가까운 SLC14A1 mRNA 서열의 일부; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA의 위치에 가장 가까운 SLC14A1 cDNA 서열의 일부에 혼성화하는 프라이머와 접촉시키는 단계; b) 적어도 i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 핵산 서열의 위치; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA 핵산 서열의 위치; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA 핵산 서열의 위치에 걸쳐 프라이머를 연장시키는 단계; 및 c) 프라이머의 연장 산물이 i) 서열번호 2에 따른 SLC14A1 게놈 핵산 서열의 6963번 내지 6965번 위치에 상응하는 위치에서 이소류신을 암호화하거나; ii) 서열번호 5에 따른 SLC14A1 mRNA의 226번 내지 228번 위치에 상응하거나 서열번호 6에 따른 SLC14A1 mRNA 핵산 서열의 394번 내지 396번 위치에 상응하는 위치에서 이소류신을 암호화하거나; iii) 서열번호 9에 따른 SLC14A1 cDNA 핵산 서열의 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 SLC14A1 cDNA 핵산 서열의 394번 내지 396번 위치에 상응하는 위치에서 이소류신을 암호화하거나, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는, 코돈을 포함하는지의 여부를 결정하는 단계를 포함한다.
- [0207] 몇몇 실시형태에서, 분석법은 생물학적 샘플을 엄격한 조건하에 변이체 SLC14A1 게놈 DNA 서열, mRNA 서열 또는 cDNA 서열에 특이적으로 혼성화하고 상응하는 야생형 SLC14A1 서열에 혼성화하지 않는 프라이머 또는 프로브와

접촉시키는 단계, 및 혼성화가 발생하는지의 여부를 결정하는 단계를 포함한다.

- [0208] 몇몇 실시형태에서, 분석법은 RNA 서열분석(RNA sequencing: RNA-Seq)을 포함한다. 몇몇 실시형태에서, 분석법은 또한 역전사효소 중합효소 연쇄 반응(reverse transcriptase polymerase chain reaction: RT-PCR)을 통해 mRNA를 cDNA로 역전사시키는 단계를 포함한다.
- [0209] 몇몇 실시형태에서, 상기 방법은 표적 핵산 서열에 결합하고, 변이체 SLC14A1 게놈 DNA, mRNA 또는 cDNA를 포함하는 폴리뉴클레오타이드를 특이적으로 검출하고/하거나 확인하기에 충분한 뉴클레오타이드 길이의 프로브 및 프라이머를 사용한다. 혼성화 조건 또는 반응 조건은 이 결과를 달성하기 위해 조작자에 의해 결정될 수 있다. 이 뉴클레오타이드 길이는 본원에 기재되거나 예시된 임의의 분석법을 포함하는 선택의 검출 방법에서 사용하기에 충분한 임의의 길이일 수 있다. 일반적으로, 예를 들어 약 8개, 약 10개, 약 11개, 약 12개, 약 14개, 약 15개, 약 16개, 약 18개, 약 20개, 약 22개, 약 24개, 약 26개, 약 28개, 약 30개, 약 40개, 약 50개, 약 75개, 약 100개, 약 200개, 약 300개, 약 400개, 약 500개, 약 600개 또는 약 700개, 또는 초과 뉴클레오타이드, 또는 약 11개 내지 약 20개, 약 20개 내지 약 30개, 약 30개 내지 약 40개, 약 40개 내지 약 50개, 약 50개 내지 약 100개, 약 100개 내지 약 200개, 약 200개 내지 약 300개, 약 300개 내지 약 400개, 약 400개 내지 약 500개, 약 500개 내지 약 600개, 약 600개 내지 약 700개, 또는 약 700개 내지 약 800개, 또는 초과 뉴클레오타이드의 길이를 갖는 프라이머 또는 프로브를 사용한다. 바람직한 실시형태에서, 프로브 또는 프라이머는 적어도 약 18개의 뉴클레오타이드의 길이를 포함한다. 프로브 또는 프라이머는 약 10개 내지 약 35개, 약 10개 내지 약 30개, 약 10개 내지 약 25개, 약 12개 내지 약 30개, 약 12개 내지 약 28개, 약 12개 내지 약 24개, 약 15개 내지 약 30개, 약 15개 내지 약 25개, 약 18개 내지 약 30개, 약 18개 내지 약 25개, 약 18개 내지 약 24개, 또는 약 18개 내지 약 22개의 뉴클레오타이드의 길이를 포함할 수 있다. 바람직한 실시형태에서, 프로브 또는 프라이머는 약 18개 내지 약 30개의 뉴클레오타이드의 길이이다.
- [0210] 이러한 프로브 및 프라이머는 높은 엄격성 혼성화 조건하에 표적 서열에 특이적으로 혼성화할 수 있다. 프로브 및 프라이머는 표적 서열과의 인접 뉴클레오타이드의 완전한 핵산 서열 동일성을 가질 수 있지만, 표적 핵산 서열과 상이하고, 표적 핵산 서열을 특이적으로 검출하고/하거나 확인하는 능력을 보유하는 프로브는 종래의 방법에 의해 설계될 수 있다. 따라서, 프로브 및 프라이머는 표적 핵산 분자와 약 80%, 약 85%, 약 90%, 약 91%, 약 92%, 약 93%, 약 94%, 약 95%, 약 96%, 약 97%, 약 98%, 약 99% 또는 100%의 서열 동일성 또는 상보성을 공유할 수 있다.
- [0211] 몇몇 실시형태에서, 특이적 프라이머는 생물학적 샘플에서 변이체 SLC14A1 유전좌위를 확인하기 위해 또는 특이적 SLC14A1 mRNA 또는 cDNA의 수준을 결정하기 위해 특이적 프로브로서 사용될 수 있거나 자체가 검출될 수 있는 앰플리콘을 제조하도록 변이체 SLC14A1 유전좌위 및/또는 SLC14A1 변이체 mRNA 또는 cDNA를 증폭시키도록 사용될 수 있다. SLC14A1 변이체 유전좌위는 서열번호 13에 따른 76번 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에서 이소류신을 암호화하는 위치에 상응하는 위치를 포함하는 게놈 핵산 서열을 나타내도록 사용될 수 있다. 프로브가 핵산 분자에 대한 프로브의 결합을 허용하는 조건 하에 생물학적 샘플에서 핵산 분자와 혼성화될 때, 이 결합은 생물학적 샘플에서 검출될 수 있고, 변이체 SLC14A1 유전좌위의 존재 또는 변이체 SLC14A1 mRNA 또는 cDNA의 존재 또는 수준의 표시를 허용한다. 결합된 프로브의 이러한 확인이 기재되어 있다. 특이적 프로브는 변이체 SLC14A1 유전자의 특이적 영역과 적어도 약 80%, 약 80% 내지 약 85%, 약 85% 내지 약 90%, 약 90% 내지 약 95%, 및 약 95% 내지 약 100% 동일한(또는 상보적인) 서열을 포함할 수 있다. 특이적 프로브는 변이체 SLC14A1 mRNA의 특이적 영역과 적어도 약 80%, 약 80% 내지 약 85%, 약 85% 내지 약 90%, 약 90% 내지 약 95%, 및 약 95% 내지 약 100% 동일한(또는 상보적인) 서열을 포함할 수 있다. 특이적 프로브는 변이체 SLC14A1 cDNA의 특이적 영역과 적어도 약 80%, 약 80% 내지 약 85%, 약 85% 내지 약 90%, 약 90% 내지 약 95%, 및 약 95% 내지 약 100% 동일한(또는 상보적인) 서열을 포함할 수 있다.
- [0212] 몇몇 실시형태에서, 생물학적 샘플의 핵산 보체가 변이체 SLC14A1 단백질을 암호화하는(예를 들어, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는) 핵산 서열을 포함하는지의 여부를 결정하기 위해, 생물학적 샘플은, 서열번호 9에 따른 186번 위치에 상응하는 위치에서 세린을 암호화하는 위치에서 뉴클레오타이드의 존재에 대해 진단적인 앰플리콘을 제조하도록, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치에 인접한 5' 플랭킹 서열로부터 유래된 제1 프라이머, 및 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치와 인접한 3' 플랭킹 서열로부터 유래된 제2 프라이머를 포함하는 프라이머 쌍을 사용하여 핵산 증폭 방법으로 처리될 수 있다. 몇몇

실시형태에서, 앰플리콘은 프라이머 쌍과 하나의 뉴클레오타이드 염기 쌍의 합산 길이로부터 DNA 증폭 프로토콜에 의해 제조 가능한 앰플리콘의 임의의 길이까지의 길이의 범위일 수 있다. 이 거리는 1개의 뉴클레오타이드 염기 쌍으로부터 증폭 반응의 한계, 또는 약 20000개의 뉴클레오타이드 염기 쌍의 범위일 수 있다. 선택적으로, 프라이머 쌍은 서열번호 13에 따른 76번 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치를 포함하는 영역 및 서열번호 13에 따른 76번 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 위치의 각각의 측에서 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 또는 초과 뉴클레오타이드를 축적시킨다. 유사한 앰플리콘은 mRNA 및/또는 cDNA 서열로부터 생성될 수 있다.

[0213] 프로브 및 프라이머를 제조하고 사용하기 위한 대표적인 방법은, 예를 들어 *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Vol. 1-3, ed. Sambrook *et al.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1989(이하, "Sambrook *et al.*, 1989"); *Current Protocols in Molecular Biology*, ed. Ausubel *et al.*, Greene Publishing and Wiley-Interscience, New York, 1992(정기적으로 업데이트됨)(이하, "Ausubel *et al.*, 1992"); 및 Innis *et al.*, PCR Protocols: A Guide to Methods and Applications, Academic Press: San Diego, 1990)에 기재되어 있다. PCR 프라이머 쌍은, 예를 들어 그 목적을 위해 의도된 컴퓨터 프로그램, 예컨대 Vector NTI 버전 10(Informax Inc.(메릴랜드주 베데스다)); PrimerSelect(DNASTAR Inc.(위스콘신주 매디슨)); 및 Primer3(Version 0.4.0.COPYRGT., 1991, Whitehead Institute for Biomedical Research(메사추세츠주 캠브리지))에서의 PCR 프라이머 분석 도구를 이용함으로써 공지된 서열로부터 유래될 수 있다. 추가적으로, 서열은 가시적으로 스캐닝될 수 있고, 프라이머는 공지된 가이드라인을 이용하여 수동으로 확인될 수 있다.

[0214] 임의의 핵산 혼성화 또는 증폭 또는 서열분석 방법은 변이체 SLC14A1 유전자 유전좌위의 존재 및/또는 변이체 SLC14A1 mRNA 또는 mRNA로부터 제조된 cDNA의 수준을 특이적으로 검출하도록 이용될 수 있다. 몇몇 실시형태에서, 핵산 분자는 SLC14A1 핵산의 영역을 증폭시키도록 프라이머로서 사용될 수 있거나, 핵산 분자는, 예를 들어 엄격한 조건하에, 변이체 SLC14A1 유전자 유전좌위를 포함하는 핵산 분자 또는 변이체 SLC14A1 mRNA 또는 mRNA로부터 제조된 cDNA를 포함하는 핵산 분자에 특이적으로 혼성화하는 프로브로서 사용될 수 있다.

[0215] 다양한 기법은, 예를 들어 핵산 서열분석, 핵산 혼성화 및 핵산 증폭을 포함하여 당해 분야에서 이용 가능하다. 핵산 서열분석 기법의 예시적인 예는 사슬 종결자 (Sanger) 서열분석 및 염료 종결자 서열분석을 포함하지만, 이들로 제한되지는 않는다.

[0216] 다른 방법은 정제된 DNA, 증폭된 DNA 및 고정된 세포 준비에 지향된 표지된 프라이머 또는 프로브를 사용하는 것을 포함하는 서열분석 이외의 핵산 혼성화 방법(형광 동소 혼성화(*fluorescence in situ* hybridization)(FISH))을 수반한다. 몇몇 방법에서, 표적 핵산은 검출 전에 또는 검출과 동시에 증폭될 수 있다. 핵산 증폭 기법의 예시적인 예는 중합효소 연쇄 반응(PCR), 리가제 연쇄 반응(ligase chain reaction: LCR), 가닥 대체 증폭(strand displacement amplification: SDA) 및 핵산 서열 기반 증폭(nucleic acid sequence based amplification: NASBA)을 포함하지만, 이들로 제한되지는 않는다. 다른 방법은 리가제 연쇄 반응, 가닥 대체 증폭 및 호열성 SDA(thermophilic SDA)(tSDA)를 포함하지만, 이들로 제한되지는 않는다.

[0217] 예를 들어, 혼성화 보호 분석법(Hybridization Protection Assay: HPA), 실시간의 증폭 과정의 정량적 평가 및 샘플에 초기에 존재하지만 실시간 증폭에 기초하지 않은 표적 서열의 분량의 결정을 포함하는, 임의의 방법은 비증폭된 또는 증폭된 폴리뉴클레오타이드를 검출하기 위해 이용될 수 있다.

[0218] 반드시 서열 증폭을 요하지 않고, 예를 들어 염색체 물질의 썬던(DNA:DNA) 블롯 혼성화, 동소 혼성화(*in situ* hybridization: ISH) 및 형광 동소 혼성화(FISH)의 공지된 방법에 기초한, 핵산을 확인하는 방법이 또한 제공된다. 썬던 블로팅은 특이적 핵산 서열을 검출하도록 이용될 수 있다. 이러한 방법에서, 샘플로부터 추출된 핵산은 단편화되고, 매트릭스 겔에서 전기영동으로 분리되고, 막 필터로 이송된다. 필터 결합된 핵산은 관심 대상의 서열에 상보적인 표지된 프로브와의 혼성화로 처리된다. 필터에 결합된 혼성화된 프로브는 검출된다. 임의의 이러한 방법에서, 그 공정은 본원에 기재되거나 예시된 임의의 프로브를 사용한 혼성화를 포함할 수 있다.

[0219] 혼성화 기법에서, 엄격한 조건은 프로브 또는 프라이머가 이의 표적에 특이적으로 혼성화하도록 이용될 수 있다. 몇몇 실시형태에서, 폴리뉴클레오타이드 프라이머 또는 프로브는 엄격한 조건하에 다른 서열(예를 들어, 상응하는 야생형 SLC14A1 유전좌위, 야생형 mRNA 또는 야생형 cDNA)보다 검출 가능하게 더 높은 정도로, 배경에 비해 10배 초과를 포함하여, 예컨대 배경에 비해 적어도 2배, 적어도 3배, 적어도 4배, 또는 초과로 이의 표적 서열(예를 들어, 변이체 SLC14A1 유전자 유전좌위, 변이체 SLC14A1 mRNA 또는 변이체 SLC14A1 cDNA)에 혼성화할 것이다. 몇몇 실시형태에서, 폴리뉴클레오타이드 프라이머 또는 프로브는 엄격한 조건하에 다른 서열보다 검

출 가능하게 더 높은 정도로 적어도 2배로 이의 표적 서열에 혼성화할 것이다. 몇몇 실시형태에서, 폴리뉴클레오타이드 프라이머 또는 프로브는 엄격한 조건하에 다른 서열보다 검출 가능하게 더 높은 정도로 적어도 3배로 이의 표적 서열에 혼성화할 것이다. 몇몇 실시형태에서, 폴리뉴클레오타이드 프라이머 또는 프로브는 엄격한 조건하에 다른 서열보다 검출 가능하게 더 높은 정도로 적어도 4배로 이의 표적 서열에 혼성화할 것이다. 몇몇 실시형태에서, 폴리뉴클레오타이드 프라이머 또는 프로브는 엄격한 조건하에 다른 서열보다 검출 가능하게 더 높은 정도로 배경에 비해 10배 초과로 이의 표적 서열에 혼성화할 것이다. 엄격한 조건은 서열 의존적이고, 상이한 환경에서 상이할 것이다. 혼성화의 엄격성 및/또는 세척 조건을 제어함으로써, 프로브와 100% 상보적인 표적 서열이 확인될 수 있다(동종성 프로빙). 대안적으로, 엄격성 조건은 더 낮은 정도의 동일성이 검출되도록(이종성 프로빙) 서열에서 약간의 미스매칭을 허용하도록 조정될 수 있다.

[0220]

DNA 혼성화, 예를 들어 약 45°C에서의 6X 염화나트륨/시트르산나트륨(SSC), 이어서 50°C에서의 2X SSC의 세척을 촉진하는 적절한 엄격성 조건은 공지되거나, *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6에서 발견될 수 있다. 전형적으로, 혼성화 및 검출에 대한 엄격한 조건은 염 농도가 pH 7.0 내지 8.3에서 약 1.5 M 미만의 Na 이온, 전형적으로 약 0.01 내지 1.0 M Na 이온 농도(또는 다른 염)이고, 온도가 짧은 프로브(예를 들어, 10개 내지 50개의 뉴클레오타이드)에 대해 적어도 약 30°C 및 더 긴 프로브(예를 들어, 50개 초과 뉴클레오타이드)에 대해 적어도 약 60°C인 것일 것이다. 엄격한 조건은 탈안정화제, 예컨대 포름아미드의 첨가에 의해 또한 달성될 수 있다. 예시적인 낮은 엄격성 조건은 37°C에서의 30 내지 35% 포름아미드, 1 M NaCl, 1% SDS(황산 나트륨 도데실)의 완충제 용액에 의한 혼성화, 및 50 내지 55°C에서의 1X 내지 2X SSC(20X SSC = 3.0 M NaCl/0.3 M 시트르산삼나트륨) 중의 세척을 포함한다. 예시적인 보통의 엄격성 조건은 37°C에서의 40 내지 45% 포름아미드, 1.0 M NaCl, 1% SDS 중의 혼성화, 및 55 내지 60°C에서의 0.5X 내지 1X SSC 중의 세척을 포함한다. 예시적인 높은 엄격성 조건은 37°C에서의 50% 포름아미드, 1 M NaCl, 1% SDS 중의 혼성화, 및 60 내지 65°C에서의 0.1X SSC 중의 세척을 포함한다. 선택적으로, 세척 완충제는 약 0.1% 내지 약 1% SDS를 포함할 수 있다. 혼성화의 기간은 일반적으로 약 24시간 미만, 보통 약 4 내지 약 12시간이다. 세척 시간의 기간은 적어도 평형에 도달하기에 충분한 시간의 길이일 것이다.

[0221]

혼성화 반응에서, 특이성은 전형적으로 혼성화 후 세척의 함수이고, 중요한 인자는 최종 세척 용액의 이온 농도 및 온도이다. DNA-DNA 하이브리드에 대해, T_m 은 Meinkoth and Wahl, *Anal. Biochem.*, 1984, 138, 267-284의 식으로부터 근사치화될 수 있다: $T_m = 81.5^\circ\text{C} + 16.6(\log M) + 0.41(\% \text{ GC}) - 0.61(\% \text{ form}) - 500/L$; 여기서 M은 1가 양이온의 몰농도이고, %GC는 DNA에서의 구아노신 및 사이토신 뉴클레오타이드의 백분율이고, % form은 혼성화 용액 중의 포름아미드의 백분율이고, L은 염기 쌍에서의 하이브리드의 길이이다. T_m 은 상보적 표적 서열의 50%가 (한정된 이온 농도 및 pH 하에) 완벽하게 일치된 프로브에 혼성화하는 온도이다. T_m 은 미스매칭의 각각의 1%에 대해 약 1°C만큼 감소하고; 이에 의해, T_m , 혼성화, 및/또는 세척 조건은 원하는 동일성의 서열에 혼성화하도록 조정될 수 있다. 예를 들어, 90% 이상의 동일성을 갖는 서열이 추구되는 경우, T_m 은 10°C 감소할 수 있다. 일반적으로, 엄격한 조건은 한정된 이온 농도 및 pH에서 특이적 서열 및 이의 보체에 대한 열 융점(T_m)보다 약 5°C 더 낮도록 선택된다. 그러나, 심하게 엄격한 조건은 열 융점(T_m)보다 1°C, 2°C, 3°C 또는 4°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있고; 보통의 엄격한 조건은 열 융점(T_m)보다 6°C, 7°C, 8°C, 9°C 또는 10°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있고; 낮은 엄격성 조건은 열 융점(T_m)보다 11°C, 12°C, 13°C, 14°C, 15°C 또는 20°C 낮은 온도에서의 혼성화 및/또는 세척을 이용할 수 있다. 식, 혼성화 및 세척 조성물, 및 원하는 T_m 을 이용하여, 보통의 기술자는 혼성화의 엄격성 및/또는 세척 용액의 변동이 본질적으로 기재되어 있다는 것을 이해한다. 원하는 정도의 미스매칭이 45°C(수성 용액) 또는 32°C(포름아미드 용액) 미만의 T_m 을 발생시키는 경우, 더 높은 온도가 이용될 수 있도록 SSC 농도를 증가시키는 것이 최적이다.

[0222]

예를 들어, 단백질 서열분석 및 면역검정을 포함하여 생물학적 샘플에서 변이체 SLC14A1 폴리펩타이드의 존재를 검출하거나 수준을 정량화하는 방법이 또한 제공된다. 몇몇 실시형태에서, 인간 대상체에서 변이체 SLC14A1 단백질(예를 들어, 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질)의 존재를 검출하는 방법은 생물학적 샘플에서 변이체 SLC14A1 단백질(예를 들어, 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질)의 존재를 검출하는 인간 대상체로부터의 생물학적 샘플에서 분석법을 수행하는 단계를 포함한다. 몇몇 실시형태에서, 인간 대상체에서 변이체 SLC14A1 단백질(예를 들어, 서열번호 13 및/또는 서열번호 14)의 존재를 검출하는 방법은 생물학적 샘플에서 변이체 SLC14A1 단백질(예를 들어, 서열번호 13 및/또는 서열번호 14)의 존

재를 검출하는 인간 대상체로부터의 생물학적 샘플에서 분석법을 수행하는 단계를 포함한다.

- [0223] 단백질 서열분석 기법의 예시적인 비제한적인 예는 질량 분광법 및 Edman 분해를 포함하지만, 이들로 제한되지는 않는다. 면역검정의 예시적인 예는 면역침강, 웨스턴 블롯, 면역조직화학, ELISA, 면역세포화학, 유세포분석법 및 미유노-PCR을 포함하지만, 이들로 제한되지는 않는다. 다양한 공지된 기법(예를 들어, 열량측정, 형광, 화학발광 또는 방사성)을 이용하여 검출 가능하게 표지된 다중클론 또는 단일클론 항체는 면역검정에서 사용하기에 적합하다.
- [0224] 본 개시내용은 또한 세포로 발현 벡터를 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 발현 벡터는 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화하는 뉴클레오타이드 서열을 포함하는 변이체 SLC14A1 유전자를 포함한다.
- [0225] 본 개시내용은 또한 세포로 발현 벡터를 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 발현 벡터는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드 서열을 포함하는 변이체 SLC14A1 유전자를 포함한다. 몇몇 실시형태에서, 발현 벡터는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치에서 이소류신을 암호화하는 코돈을 포함하는 뉴클레오타이드 서열을 포함하는 재조합 SLC14A1 유전자를 포함한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0226] 본 개시내용은 또한 세포로 발현 벡터를 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 발현 벡터는 서열번호 13과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자를 포함한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0227] 본 개시내용은 또한 세포로 발현 벡터를 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 발현 벡터는 서열번호 14와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 폴리펩타이드를 암호화하는 핵산 분자를 포함한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0228] 본 개시내용은 또한 세포로 변이체 SLC14A1 폴리펩타이드 또는 이의 단편을 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 SLC14A1 폴리펩타이드는 서열번호 13과 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0229] 본 개시내용은 또한 세포로 변이체 SLC14A1 폴리펩타이드 또는 이의 단편을 도입하는 단계를 포함하는 세포를 변형시키는 방법을 제공하고, 여기서 SLC14A1 폴리펩타이드는 서열번호 14와 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98% 또는 적어도 약 99% 동일하고, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.
- [0230] 본 개시내용은 또한 응고 병태 또는 CAD를 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법을 제공한다. 몇몇 실시형태에서, 상기 방법은 변이체 SLC14A1 게놈 DNA, mRNA 또는 mRNA로부터 수득된 cDNA의 존재를 검출하는 단계를 포함하고, 여기서 변이체 SLC14A1 게놈 DNA, mRNA 또는 mRNA로부터 수득된 cDNA는 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화한다.
- [0231] 몇몇 실시형태에서, 상기 방법은 대상체로부터 수득된 생물학적 샘플로부터 수득된 변이체 SLC14A1 게놈 DNA, mRNA, 또는 mRNA로부터 수득된 cDNA의 존재를 검출하는 단계를 포함한다. 집단 내의 유전자 서열 및 이러한 유전자에 의해 암호화된 mRNA가 다형, 예컨대 단일 뉴클레오타이드 다형(single nucleotide polymorphism)(SNP)으로 인해 변할 수 있다고 이해된다. 변이체 SLC14A1 게놈 DNA, mRNA, cDNA 및 폴리펩타이드에 대한 본원에 제공된 서열은 오직 예시적인 서열이고, 추가적인 SLC14A1 대립유전자를 포함하는 다른 이러한 서열이 또한 가능하다.
- [0232] 몇몇 실시형태에서, 상기 방법은 a) 대상체로부터 수득된 샘플을 분석하여서 샘플에서의 핵산 분자가 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계; 및 b) 핵산 분자가 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, 핵산 분자가 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하지 않는 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함한다.

- [0233] 몇몇 실시형태에서, 상기 방법은 a) 대상체로부터 수득된 샘플을 분석하여서 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계; 및 b) 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열을 포함하지 않는 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함한다.
- [0234] 몇몇 실시형태에서, 분석법은 샘플에서 핵산 분자의 SLC14A1 게놈 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 위치를 포함하는 것인 단계; 샘플에서 핵산 분자의 SLC14A1 mRNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 5에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 샘플에서 핵산 분자의 SLC14A1 mRNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것이 단계; 샘플에서 핵산 분자의 SLC14A1 cDNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 9에 따른 226번 내지 228번 위치에 상응하는 위치를 포함하는 것인 단계; 및/또는 샘플에서 핵산 분자의 SLC14A1 cDNA 서열의 일부를 서열분석하되, 서열분석된 일부는 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 위치를 포함하는 것인 단계를 포함한다. 본원에 개시된 임의의 핵산 분자(예를 들어, 게놈 DNA, mRNA 또는 cDNA)는 서열분석될 수 있다. 몇몇 실시형태에서, 검출 단계는 전체 핵산 분자를 서열분석하는 단계를 포함한다.
- [0235] 몇몇 실시형태에서, 검출 단계는 SLC14A1 단백질을 암호화하는 핵산 분자의 적어도 일부를 증폭시키되, 증폭된 핵산 분자는 서열번호 13에 따른 76번 위치에 상응하는 위치를 포함하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치를 포함하는 아미노산 서열을 암호화하는 것인 단계; 핵산 분자를 검출 가능한 표지로 표시하는 단계; 표시된 핵산을 프로브를 포함하는 지지체와 접촉시키되, 프로브는 엄격한 조건하에 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열에 혼성화하는 핵산 서열을 포함하는 것인 단계; 및 검출 가능한 표지를 검출하는 단계를 포함한다. 본원에 개시된 임의의 핵산 분자는 증폭될 수 있다. 예를 들어, 본원에 개시된 임의의 게놈 DNA, cDNA 또는 mRNA 분자는 증폭될 수 있다. 몇몇 실시형태에서, 핵산 분자는 mRNA이고, 상기 방법은 증폭 단계 전에 mRNA를 cDNA로 역전사시키는 단계를 추가로 포함한다.
- [0236] 몇몇 실시형태에서, 분석법은 a) 샘플을, i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 서열의 위치에 가장 가까운 SLC14A1 게놈 서열의 일부; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA의 위치에 가장 가까운 SLC14A1 mRNA 서열의 일부; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA의 위치에 가장 가까운 SLC14A1 cDNA 서열의 일부에 혼성화하는 프라이머와 접촉시키는 단계; b) 적어도 i) 서열번호 2에 따른 6963번 내지 6965번 위치에 상응하는 SLC14A1 게놈 핵산 서열의 위치; ii) 서열번호 5에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 6에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 mRNA 핵산 서열의 위치; 또는 iii) 서열번호 9에 따른 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 394번 내지 396번 위치에 상응하는 SLC14A1 cDNA 핵산 서열의 위치에 걸쳐 프라이머를 연장시키는 단계; 및 c) 프라이머의 연장 산물이 i) 서열번호 2에 따른 SLC14A1 게놈 핵산 서열의 6963번 내지 6965번 위치에 상응하거나; ii) 서열번호 5에 따른 SLC14A1 mRNA 핵산 서열의 226번 내지 228번 위치에 상응하거나, 서열번호 6에 따른 SLC14A1 mRNA 핵산 서열의 394번 내지 396번 위치에 상응하거나; iii) 서열번호 9에 따른 SLC14A1 cDNA 핵산 서열의 226번 내지 228번 위치에 상응하거나, 서열번호 10에 따른 SLC14A1 cDNA 핵산 서열의 394번 내지 396번 위치에 상응하는 위치에서 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 뉴클레오타이드를 포함하는지의 여부를 결정하는 단계를 포함한다.
- [0237] 몇몇 실시형태에서, 분석법은 샘플을 엄격한 조건하에 SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열에 특이적으로 혼성화하지만 상응하는 야생형 SLC14A1 핵산 서열에 혼성화하지 않는 프라이머 또는 프로브와 접촉시키는 단계, 및 혼성화가 발생하는지의 여부를 결정하는 단계를 포함한다. 몇몇 실시형태에서, SLC14A1 변이체 게놈 핵산 서열, SLC14A1 변이체 mRNA 핵산 서열 또는 SLC14A1 변이체 cDNA 핵산 서열은 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을

암호화하거나 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화한다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.

[0238] 본 개시내용은 또한, a) 인간 대상체로부터 수득된 샘플을 분석하여서 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하는 단계; 및 b) 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 경우 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 경우 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함하는, 응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법을 제공한다. 몇몇 실시형태에서, 효소-결합 면역흡착 분석법(ELISA)은 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부 및/또는 샘플에서의 SLC14A1 단백질이 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하도록 이용된다. 몇몇 실시형태에서, 그 방법은 시험관내 방법이다.

[0239] 상기 방법의 몇몇 실시형태에서, 검출 단계는 SLC14A1 단백질을 암호화하는 핵산 분자의 적어도 일부를 서열분석하는 단계를 포함한다. 서열분석된 핵산 분자는 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질을 암호화할 수 있다. 몇몇 실시형태에서, 서열분석된 핵산 분자는 서열번호 13에 따른 76번 위치에 상응하는 위치를 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치를 포함하는 아미노산 서열을 암호화할 수 있다. 서열번호 2에 따른 6963번 위치에 상응하는 위치(예를 들어, 게놈 DNA)에서의, 또는 서열번호 5 또는 서열번호 9에 따른 226번 위치에 상응하는 위치(예를 들어, mRNA)에서의, 또는 서열번호 6 또는 서열번호 10에 따른 394번 위치에 상응하는 위치(예를 들어, cDNA)에서의 아데닌의 존재는 각각 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 변이체 SLC14A1 단백질을 발생시킨다. 검출 단계는 전체 SLC14A1 단백질을 암호화하는 핵산 분자를 서열분석하는 단계를 포함할 수 있다.

[0240] 상기 방법의 몇몇 실시형태에서, 검출 단계는 SLC14A1 단백질을 암호화하는 핵산 분자의 적어도 일부를 증폭시키는 단계, 핵산 분자를 검출 가능한 표지로 표지하는 단계, 표지된 핵산을 프로브를 포함하는 지지체와 접촉시키되, 프로브는, 예를 들어 엄격한 조건하에를 포함하여, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열, 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 암호화하는 핵산 서열에 특이적으로 혼성화하는 핵산 서열(또는 서열번호 2에 따른 6963번 위치에 상응하는 위치(예를 들어, 게놈 DNA)에서, 또는 서열번호 5 또는 서열번호 9에 따른 226번 위치에 상응하는 위치(예를 들어, mRNA)에서, 또는 서열번호 6 또는 서열번호 10에 따른 394번 위치에 상응하는 위치(예를 들어, cDNA)에서 아데닌을 갖는 핵산 서열)을 포함하는 것인 단계, 및 검출 가능한 표지를 검출하는 단계를 포함한다. 증폭된 핵산 분자는 바람직하게는 서열번호 13에 따른 76번 위치에 상응하는 위치를 포함하는 아미노산 서열을 암호화하거나, 바람직하게는 서열번호 14에 따른 132번 위치에 상응하는 위치를 포함하는 아미노산 서열을 암호화한다. 핵산이 mRNA를 포함하는 경우, 상기 방법은 증폭 단계 전에 mRNA를 cDNA로 역전사시키는 단계를 추가로 포함할 수 있다. 몇몇 실시형태에서, 결정 단계는 핵산 분자를 검출 가능한 표지를 포함하는 프로브와 접촉시키는 단계 및 검출 가능한 표지를 검출하는 단계를 포함한다. 프로브는 바람직하게는, 예를 들어 엄격한 조건하를 포함하여, 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는 핵산 서열 또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 아미노산 서열을 암호화하는 핵산 서열에 특이적으로 혼성화하는 핵산 서열(또는 서열번호 2에 따른 6963번 위치에 상응하는 위치(예를 들어, 게놈 DNA)에서, 또는 서열번호 5 또는 서열번호 9에 따른 226번 위치에 상응하는 위치(예를 들어, mRNA)에서, 또는 서열번호 6 또는 서열번호 10에 따른 394번 위치에 상응하는 위치(예를 들어, cDNA)에서 아데닌을 갖는 핵산 서열)을 포함한다. 핵산 분자는 인간 대상체로부터 수득된 세포 내에 존재할 수 있다.

[0241] 본원에 개시된 방법에서 이용될 수 있는 다른 분석법은, 예를 들어 역전사 중합효소 연쇄 반응(RT-PCR) 또는 정량적 RT-PCR(qRT-PCR)을 포함한다. 본원에 개시된 방법에서 이용될 수 있는 더 다른 분석법은, 예를 들어 RNA 서열분석(RNA-Seq), 이어서 생물학적 샘플에서의 변이체 mRNA 또는 cDNA의 존재 및 분량의 검출을 포함한다.

- [0242] 본원에 기재된 방법은 시험관내, 인시츄 또는 생체내 수행될 수 있다.
- [0243] 본 개시내용은 또한, a) 인간 대상체로부터 수득된 샘플에서 분석법을 수행하여서 샘플에서의 SLC14A1 단백질이 기능 소실 단백질 또는 기능 부분 소실 단백질인지의 여부를 결정하는 단계; 및 b) SLC14A1 폴리펩타이드가 기능 소실 단백질 또는 기능 부분 소실 단백질인 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, SLC14A1 폴리펩타이드가 기능 소실 단백질 또는 기능 부분 소실 단백질이 아닌 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함하는, 응고 병태 또는 CAD를 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법을 제공한다.
- [0244] 본 개시내용은 또한, a) 인간 대상체로부터 수득된 샘플에서 분석법을 수행하여서 샘플에서의 SLC14A1 단백질이 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는지의 여부를 결정하는 단계; 및 b) SLC14A1 폴리펩타이드가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, SLC14A1 폴리펩타이드가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함하는, 응고 병태 또는 CAD를 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법을 제공한다. 몇몇 실시형태에서, 인간 대상체는 이러한 결정을 필요로 한다. 몇몇 실시형태에서, 인간 대상체는 응고 병태 또는 CAD를 갖는 친척을 가질 수 있다.
- [0245] 본 개시내용은 또한, a) 인간 대상체로부터 수득된 샘플에서 분석법을 수행하여서 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부 및/또는 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는지의 여부를 결정하는 단계; 및 b) 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 경우 및/또는 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 서열을 포함하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 감소된 것으로서 인간 대상체를 분류하거나, 샘플에서의 핵산 분자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 경우 및/또는 샘플에서의 핵산 분자가 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 경우 응고 병태 또는 CAD를 발생시킬 위험이 증가된 것으로서 인간 대상체를 분류하는 단계를 포함하는, 응고 병태 또는 관상 동맥 질환(CAD)을 발생시키는 것에 대한 인간 대상체의 감수성을 결정하는 방법을 제공한다.
- [0246] 본원에 기재된 임의의 방법은 응고 병태를 갖거나 응고 병태를 발생시킬 위험이 증가된 대상체에 대해 응고 병태를 (부분적으로 또는 완전히) 예방하거나 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함할 수 있다. 몇몇 실시형태에서, 항응고제는 헤파린, 와파린(COUMADIN[®]) 및 JANTOVEN[®]), 리바룩사반(XARELTO[®]), 다비가트란(PRADAXA[®]), 에픽사반(ELIQUIS[®]), 에독사반(SAVAYSA[®]), 에녹사파린(LOVENOX[®]), 폰다파리눅스(ARIXTRA[®]), 달테파린(FRAGMIN[®]), 비발리루딘(ANGIOMAX[®]), 아가트로반(ACOVA[®]) 또는 안티트롬빈 III(THROMBATE III[®])이다. 몇몇 실시형태에서, 항응고제는 본원에 기재된 임의의 변이체 SLC14A1 폴리펩타이드이다.
- [0247] 본원에 기재된 임의의 방법은 CAD를 갖거나 CAD를 발생시킬 위험이 증가된 대상체에 대해 CAD를 (부분적으로 또는 완전히) 예방하거나 치료하거나 저해하는 치료제를 투여하는 단계를 추가로 포함할 수 있다. 몇몇 실시형태에서, 치료제는 콜레스테롤 변형 약제(예를 들어, 스타틴, 니아신, 피브레이트 또는 담즙산 봉쇄제 등), 아스피린, 베타 차단제, 니트로글리세린, 안지오텐신 전환 효소(angiotensin-converting enzyme: ACE) 저해제, 및/또는 안지오텐신 II 리셉터 차단제(angiotensin II receptor blocker: ARB)이다.
- [0248] 본 개시내용은 또한, 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 DNA 샘플에서 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 응고 병태와 연관된 유전자 변이체 중 하나 이상을 가질 때, 응고 병태를 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는, 응고 병태를 예방하거나 치료하거나 저해하는 치료제로 응고 병태 환자를 치료하는 방법을 제공한다. 응고 병태와 연관된 유전자 변이체는 이러한 활성을 갖는 본원에 개시된 임의의 변이체일 수 있다. 몇몇 실시형태에서,

응고 병태와 연관된 하나 이상의 유전자 변이체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자이다. 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 것은 본원에 기재된 임의의 방법을 포함할 수 있다. 몇몇 실시형태에서, 유전자형 분석법이 응고 병태 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자를 포함한다는 것을 나타낼 때, 응고 병태 환자는 응고 병태를 예방하거나 치료하거나 저해하는 치료제로, 그러나 응고 병태 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자를 포함하는 경우보다 낮거나 덜 빈번한(예를 들어, 약 10% 낮거나 덜 빈번하거나, 약 20% 낮거나 덜 빈번하거나, 약 30% 낮거나 덜 빈번하거나, 약 40% 낮거나 덜 빈번하거나, 약 50% 낮거나 덜 빈번하거나, 약 60% 낮거나 덜 빈번하거나, 약 70% 낮거나 덜 빈번한) 용량으로 치료된다. 몇몇 실시형태에서, 응고 병태를 예방하거나 치료하거나 저해하는 치료제는 헤파린, 와파린(COUMADIN® 및 JANTOVEN®), 리바록사반(XARELTO®), 다비가트란(PRADAXA®), 에픽사반(ELIQUIS®), 에독사반(SAVAYSA®), 에녹사파린(LOVENOX®), 폰다파리누스(ARIXTRA®), 달테파린(FRAGMIN®), 비발리루딘(ANGIOMAX®), 아가트로반(ACOVA®) 또는 안티트롬빈 III(THROMBATE III®)이다.

[0249]

본 개시내용은 또한, 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 단백질 샘플에서 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 응고 병태와 연관된 유전자 변이체 중 하나 이상을 가질 때, 응고 병태를 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는, 응고 병태를 예방하거나 치료하거나 저해하는 치료제로 응고 병태 환자를 치료하는 방법을 제공한다. 응고 병태와 연관된 유전자 변이체는 이러한 활성을 갖는 본원에 개시된 임의의 변이체일 수 있다. 몇몇 실시형태에서, 응고 병태와 연관된 하나 이상의 유전자 변이체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이다. 분석법을 수행 중이거나 수행함으로써 환자가 응고 병태와 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 것은 본원에 기재된 임의의 방법을 포함할 수 있다. 몇몇 실시형태에서, 분석법이 응고 병태 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 포함한다는 것을 나타낼 때, 응고 병태 환자는 응고 병태를 예방하거나 치료하거나 저해하는 치료제로, 그러나 응고 병태 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 포함하는 경우보다 낮거나 덜 빈번한(예를 들어, 약 10% 낮거나 덜 빈번하거나, 약 20% 낮거나 덜 빈번하거나, 약 30% 낮거나 덜 빈번하거나, 약 40% 낮거나 덜 빈번하거나, 약 50% 낮거나 덜 빈번하거나, 약 60% 낮거나 덜 빈번하거나, 약 70% 낮거나 덜 빈번한) 용량으로 치료된다. 몇몇 실시형태에서, 응고 병태를 예방하거나 치료하거나 저해하는 치료제는 헤파린, 와파린(COUMADIN® 및 JANTOVEN®), 리바록사반(XARELTO®), 다비가트란(PRADAXA®), 에픽사반(ELIQUIS®), 에독사반(SAVAYSA®), 에녹사파린(LOVENOX®), 폰다파리누스(ARIXTRA®), 달테파린(FRAGMIN®), 비발리루딘(ANGIOMAX®), 아가트로반(ACOVA®) 또는 안티트롬빈 III(THROMBATE III®)이다.

[0250]

본 개시내용은 또한, 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 DNA 샘플에서 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 관상 동맥 질환과 연관된 유전자 변이체 중 하나 이상을 가질 때, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로 관상 동맥 질환(CAD) 환자를 치료하는 방법을 제공한다. 관상 동맥 질환과 연관된 유전자 변이체는 이러한 활성을 갖는 본원에 개시된 임의의 변이체일 수 있다. 몇몇 실시형태에서, 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자이다. 유전자형 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관된 하나 이상의

유전자 변이체를 갖는지의 여부를 결정하는 것은 본원에 기재된 임의의 방법을 포함할 수 있다. 몇몇 실시형태에서, 유전자형 분석법이 관상 동맥 질환 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 암호화하는 핵산 분자를 포함한다는 것을 나타낼 때, 관상 동맥 질환 환자는 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로, 그러나 관상 동맥 질환 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자를 포함하는 경우보다 낮거나 덜 빈번한(예를 들어, 약 10% 낮거나 덜 빈번하거나, 약 20% 낮거나 덜 빈번하거나, 약 30% 낮거나 덜 빈번하거나, 약 40% 낮거나 덜 빈번하거나, 약 50% 낮거나 덜 빈번하거나, 약 60% 낮거나 덜 빈번하거나, 약 70% 낮거나 덜 빈번한) 용량으로 치료된다. 몇몇 실시형태에서, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제는 콜레스테롤 개질 약제, 아스피린, 베타 차단제, 니트로글리세린, 안지오텐신 전환 효소(ACE) 저해제, 및/또는 안지오텐신 II 리셉터 차단제(ARB)이다. 몇몇 실시형태에서, 콜레스테롤 개질 약제는 스타틴, 니아신, 피브레이트 또는 담즙산 봉쇄제이다.

[0251] 본 개시내용은 또한, 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지를 결정하기 위해 환자로부터 수득된 단백질 샘플에서 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 단계; 및 환자가 관상 동맥 질환과 연관된 유전자 변이체 중 하나 이상을 가질 때, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제를 환자에게 투여하는 단계를 포함하는, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로 관상 동맥 질환(CAD) 환자를 치료하는 방법을 제공한다. 관상 동맥 질환과 연관된 유전자 변이체는 이러한 활성을 갖는 본원에 개시된 임의의 변이체일 수 있다. 몇몇 실시형태에서, 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질이다. 분석법을 수행 중이거나 수행함으로써 환자가 관상 동맥 질환과 연관된 하나 이상의 유전자 변이체를 갖는지의 여부를 결정하는 것은 본원에 기재된 임의의 방법을 포함할 수 있다. 몇몇 실시형태에서, 분석법이 관상 동맥 질환 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하는 SLC14A1 단백질을 포함한다는 것을 나타낼 때, 관상 동맥 질환 환자는 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제로, 그러나 관상 동맥 질환 환자가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질 및/또는 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 포함하는 경우보다 낮거나 덜 빈번한(예를 들어, 약 10% 낮거나 덜 빈번하거나, 약 20% 낮거나 덜 빈번하거나, 약 30% 낮거나 덜 빈번하거나, 약 40% 낮거나 덜 빈번하거나, 약 50% 낮거나 덜 빈번하거나, 약 60% 낮거나 덜 빈번하거나, 약 70% 낮거나 덜 빈번한) 용량으로 치료된다. 몇몇 실시형태에서, 관상 동맥 질환을 예방하거나 치료하거나 저해하는 치료제는 콜레스테롤 개질 약제, 아스피린, 베타 차단제, 니트로글리세린, 안지오텐신 전환 효소(ACE) 저해제, 및/또는 안지오텐신 II 리셉터 차단제(ARB)이다. 몇몇 실시형태에서, 콜레스테롤 개질 약제는 스타틴, 니아신, 피브레이트 또는 담즙산 봉쇄제이다.

[0252] 치료 물질의 투여는 비경구, 정맥내, 경구, 피하, 동맥내, 두개내, 척추강내, 복강내, 국소, 비강내 또는 근육내(이들로 제한되지는 않음)를 포함하는 임의의 적합한 경로에 의할 수 있다. 투여를 위한 약제학적 조성물은 바람직하게는 무균이고 실질적으로 등장성이고, GMP 조건하에 제조된다. 약제학적 조성물은 단위 투여형(즉, 단일 투여를 위한 투약량)에서 제공될 수 있다. 약제학적 조성물은 하나 이상의 생리학적으로 및 약제학적으로 허용 가능한 담체, 희석제, 부형제 또는 보조제를 사용하여 제형화될 수 있다. 제형은 선택된 투여 경로에 따라 달라진다. "약제학적으로 허용 가능한"의 용어는 담체, 희석제, 부형제 또는 보조제가 제형의 다른 성분과 상용성이고, 이의 수혜자에게 실질적으로 해롭지 않다는 것을 의미한다.

[0253] 본원에 기재된 임의의 실시형태에서, 상기 방법은 응고 병태 및/또는 CAD를 갖거나 가질 위험에 있는 대상체의 검출, 진단, 확인 및/또는 치료에 이용될 수 있다. 본원에 기재된 임의의 실시형태에서, 상기 방법은 응고 병태를 갖거나 가질 위험에 있는 대상체의 검출, 진단, 확인 및/또는 치료에 이용될 수 있다. 본원에 기재된 임의의 실시형태에서, 상기 방법은 CAD를 갖거나 가질 위험에 있는 대상체의 검출, 진단, 확인 및/또는 치료에 이용될 수 있다. 몇몇 실시형태에서, 응고 병태는 혈전증, 폐 색전증, 심근경색(MI), 정맥 혈전색전증(VTE), 심부 정맥 혈전증(DVT), 뇌동맥류 및 뇌졸중으로부터 선택된다. 몇몇 실시형태에서, 상기 방법은 조혈 병태를 갖거나 가질 위험에 있거나 이를 요하는 대상체의 검출, 진단, 확인 및/또는 치료에 이용되지 않는다.

- [0254] 본 개시내용은 또한 변이체 SLC14A1 단백질을 갖는 인간 대상체에서 응고 병태의 치료에서 사용하기 위한 항응고제를 제공하고, 여기서 변이체 SLC14A1 단백질은 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질이다. 몇몇 실시형태에서, 항응고제는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 변이체 SLC14A1 단백질을 갖는 인간 대상체에서의 응고 병태의 치료에서 사용하기 위한 것이다. 몇몇 실시형태에서, 인간 대상체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질 및/또는 SLC14A1 단백질을 암호화하는 핵산 분자에 대해 양성 시험되었다. 몇몇 실시형태에서, 치료는 인간 대상체가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자를 갖는지 또는 갖지 않는지의 여부를 결정하는 단계를 포함한다. 몇몇 실시형태에서, 인간 대상체는 본원에 기재된 임의의 방법을 이용함으로써 응고 병태를 갖는 것으로서 또는 응고 병태를 발생시킬 위험을 갖는 것으로서 확인되었다. 몇몇 실시형태에서, 항응고제는 헤파린, 와파린(COUMADIN[®] 및 JANTOVEN[®]), 리바록사반(XARELTO[®]), 다비가트란(PRADAXA[®]), 에픽사반(ELIQUIS[®]), 에독사반(SAVAYSA[®]), 에녹사파린(LOVENOX[®]), 폰다파리누스(ARIXTRA[®]), 달테파린(FRAGMIN[®]), 비발리루딘(ANGIOMAX[®]), 아가트로반(ACOVA[®]) 또는 안티트롬빈 III(THROMBATE III[®]). 몇몇 실시형태에서, 항응고제는 본원에 기재된 임의의 변이체 SLC14A1 폴리펩타이드이다.
- [0255] 본 개시내용은 또한 응고 병태를 발생시키는 대상체의 감수성을 결정하기 위한 본원에 개시된 임의의 변이체 SLC14A1 게놈 DNA, mRNA, cDNA, 폴리펩타이드 및 혼성화 핵산 분자의 용도를 제공한다.
- [0256] 본 개시내용은 또한 변이체 SLC14A1 단백질을 갖는 인간 대상체에서 CAD의 치료에서 사용하기 위한 제제를 제공하고, 여기서 변이체 SLC14A1 단백질은 기능 소실 SLC14A1 단백질 또는 기능 부분 소실 SLC14A1 단백질이다. 몇몇 실시형태에서, 항-CAD 제제는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 변이체 SLC14A1 단백질을 갖는 인간 대상체에서의 CAD의 치료에서 사용하기 위한 것이다. 몇몇 실시형태에서, 인간 대상체는 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자에 대해 양성 시험되었다. 몇몇 실시형태에서, 치료는 인간 대상체가 서열번호 13에 따른 76번 위치에 상응하는 위치에서 이소류신을 포함하지 않거나, 서열번호 14에 따른 132번 위치에 상응하는 위치에서 이소류신을 포함하지 않는 SLC14A1 단백질을 암호화하는 핵산 분자를 갖는지 또는 갖지 않는지의 여부를 결정하는 단계를 포함한다. 몇몇 실시형태에서, 인간 대상체는 본원에 기재된 임의의 방법을 이용함으로써 CAD를 갖는 것으로서 또는 CAD를 발생시킬 위험을 갖는 것으로서 확인되었다. 몇몇 실시형태에서, 제제는 콜레스테롤 개질 약제(예를 들어, 스타틴, 니아신, 피브레이트 또는 담즙산 봉쇄제 등), 아스피린, 베타 차단제, 니트로글리세린, 안지오텐신 전환 효소(ACE) 저해제, 및/또는 안지오텐신 II 리셉터 차단제(ARB)이다. 몇몇 실시형태에서, 제제는 본원에 기재된 임의의 변이체 SLC14A1 폴리펩타이드이다.
- [0257] 본 개시내용은 또한 응고 병태를 발생시키는 대상체의 감수성을 결정하기 위한 본원에 개시된 임의의 변이체 SLC14A1 게놈 DNA, mRNA, cDNA, 폴리펩타이드 및 혼성화 핵산 분자의 용도를 제공한다.
- [0258] 상기에 또는 하기에 인용된 모든 특허 문헌, 웹사이트, 다른 공보, 수탁 번호 등은, 각각의 개별 항목이 참고로 그렇게 인용된다고 구체적으로 및 개별적으로 표시된 것과 동일한 정도로, 모든 목적을 위해 그 전문이 참고로 인용된다. 서열의 상이한 버전이 다른 시간에서의 수탁 번호와 연관되는 경우, 본 출원의 유효 출원일에 수탁 번호와 연관된 버전이 의도된다. 유효 출원일은 이용 가능한 경우 실제 출원일 또는 수탁 번호에 관한 선행 출원의 출원일 중 이른 것을 의미한다. 마찬가지로, 공보, 웹사이트 또는 기타의 상이한 버전이 다른 시간에 공개되는 경우, 그 출원의 유효 출원일에 가장 최근에 공개된 버전이 달리 표시되지 않는 한 의도된다. 본 개시내용의 임의의 특징, 단계, 요소, 실시형태 또는 양상은, 구체적으로 달리 표시되지 않는 한, 임의의 다른 특징, 단계, 요소, 실시형태 또는 양상과 조합되어 사용될 수 있다. 본 개시내용이 명확성 및 이해의 목적을 위해 예시 및 예에 의해 약간 자세하게 기재되어 있지만, 소정의 변화 및 변형이 첨부된 청구범위의 범주 내에 실행될 수 있다는 것이 명확할 것이다.
- [0259] 본원에서 인용된 뉴클레오타이드 및 아미노산 서열은 뉴클레오타이드 염기에 대한 표준 철자 약어 및 아미노산에 대한 1철자 코드를 이용하여 표시된다. 뉴클레오타이드 서열은 서열의 5' 말단에서 시작하여 3' 말단으로 정

방향으로(즉, 각 선에서 왼쪽으로부터 오른쪽으로) 진행하는 표준 관례를 따른다. 각각의 뉴클레오타이드 서열의 오직 하나의 가닥이 표시되지만, 상보적 가닥은 디스플레이된 가닥을 약간 참조하여 포함되는 것으로 이해된다. 아미노산 서열은 서열의 아미노 말단에서 시작하여 카복시 말단으로 정방향으로(즉, 각 선에서 왼쪽으로부터 오른쪽으로) 진행하는 표준 관례를 따른다.

[0260] 하기 실시예는 더 자세히 실시형태를 기재하도록 제공된다. 이들은 청구된 실시형태를 제한하지 않고 예시하도록 의도된다.

[0261] 실시예

[0262] 하기 실시예는 본원에 청구된 화합물, 조성물, 물품, 장치 및/또는 방법이 어떻게 제조되고 평가되는지의 완전한 개시내용 및 설명을 당해 분야의 보통의 기술자에게 제공하도록 제시되고, 순수히 예시적인 것으로 의도되고, 본 발명자들이 이 발명을 어떻게 여기는지의 범주를 제한하도록 의도되지 않는다. 수(예를 들어, 양, 온도 등)와 관련하여 정확성을 보장하려고 노력이 되지만, 약간의 오류 및 편차가 차지할 수 있다. 달리 나타내지 않는 한, 부는 중량부이고, 온도는 °C 단위이거나 상온이고, 압력은 대기압이거나 대기압에 가깝다.

[0263] 실시예 1: 환자 모집 및 표현형분석

[0264] MyCode Community Health Initiative는 조사 목적을 위해 정보제거된 전자 건강 기록(electronic health records: EHR) 및 게놈 정보에 대한 접근을 제공하도록 동의한 125,000명 초과 Geisinger Health System(GHS) 환자의 코호트이다. Regeneron Genetics Center와 GHS 사이의 DiscovEHR 협력의 일부로서, 전장 엑스 서열분석은 대부분 유럽 혈통의 90,000명 초과 GHS 참여자에서 완료되었다. 이 응고 연구의 제1 상에서, 내인성 응고 경로의 생체의 측정치인 활성화된 부분 트롬보플라스틴 시간에 대한 유전 연관 연구는 17,630명의 유럽 혈통 개체에서 완료되었다(도 1 참조). 많은 환자가 기록된 다수의 aPTT 측정치를 가지므로, 각각의 환자에 대한 aPTT의 최소 수명 측정은 (항응고제 용법의 잠재적인 영향을 최소화하도록) 선택되고, 정맥 혈전색전증의 병력을 갖는 모든 개체는 분석으로부터 배제되었다. 이 발견 분석으로부터의 발견을 계획하기 위해, aPTT는 추가 5,892명의 유럽 혈통 GHS 참여자에서 분석되었다. 응고항진성이 정맥 및 동맥 혈전증에 대한 잠재적인 위험 인자이므로, 본 발명자들은 또한 96,180명의 개체에서의 관상 동맥 질환(CAD) 위험에 대한 *SLC14A1* V76I의 기여(GHS로부터 뽑힌 아프리카계 미국인 및 유럽 혈통 개체 및 Regeneron Genetics Center에서 서열분석된 2개의 추가 연구), 및 Regeneron Genetics Center에서 또한 서열분석된 13,963명의 대만인 개체에서의 CAD 위험에 대한 *SLC14A1* 예상된 기능 소실 변이체(c.510-1G>A)의 기여를 평가하였다.

[0265] 실시예 2: 게놈 샘플

[0266] 게놈 DNA는 말초 혈액 샘플로부터 추출되고, 전장 엑스 서열분석을 위해 리제네론 유전학 센터(Regeneron Genetics Center: RGC)로 옮겨지고, -80°C에서 자동화 바이오뱅크에 저장되었다. 형광 기반 정량화는 서열분석 목적을 위해 적절한 DNA 분량 및 품질을 보장하도록 수행된다.

[0267] 1μg의 DNA는 150개의 염기 쌍의 평균 단편 길이(Covaris LE220)로 전단되고, Kapa Biosystems로부터 커스텀 시약 키트로 엑스 포획을 위해 제조되었다. 샘플은 NimbleGen SeqCap VCRome 2.1 또는 Integrated DNA Technologies xGen 엑스 표적 설계를 이용하여 포획되었다. 샘플은 바코드화되고, 혼주되고, v4 화학물질을 갖는 Illumina HiSeq 2500에서 75 bp 짝짓기-말단 서열분석을 이용하여 서열분석에 대해 다중화되었다. 포획된 단편은 20x 또는 더 높은 커버리지에서 커버된 표적 염기의 최소 85%를 달성하도록 서열분석되었다. 서열분석 이후에, 데이터는 샘플-수준 데이터 제조 및 분석을 위한 표준 도구를 실행하기 위해 DNAnexus 및 AWS를 이용하는 RGC에서 개발된 클라우드 기반 파이프라인을 사용하여 처리되었다. 간단히, 서열 데이터는 Illumina의 CASAVA 소프트웨어를 이용하여 생성되고 탈다중화되었다. 서열 판독은 맵핑되고, BWA-mem을 이용하여 GRCh38 인간 게놈 기준 어셈블리로 정렬되었다. 정렬 후, 이중 판독은 마킹되고, Picard 도구를 사용하여 플래깅되고, 인텔은 GATK를 이용하여 재정렬되어서 변이체 콜 품질을 개선하였다. SNP 및 INDEL 변이체 및 유전자형은 GATK의 HaplotypeCaller를 이용하여 호출되고, GATK로부터의 변이체 품질 점수 재보정(Variant Quality Score Recalibration: VQSR)은 전체 변이체 품질 점수를 주석표시되도록 적용되었다. 서열분석 및 데이터 품질 미터 통계학은 포획 성능, 정렬 성능 및 변이체 호출을 평가하도록 각각의 샘플에 대해 포획되었다.

[0268] 실시예 3: 게놈 데이터 분석

[0269] 최소 판독 깊이(10 초과), 유전자형 품질(30 초과) 및 대립유전자 균형(15% 초과)에 대한 표준 품질-관리 필터는 호출된 변이체에 적용되었다. 통과 변이체는 RGC 개발된 주석표시 및 분석 파이프라인을 이용하여 이의 가능한 기능 효과(동의, 비동의, 스플라이싱, 프레임쉬프트 또는 비프레임쉬프트 변이체이든)에 기초하여 분류되고

주석표시되었다. 가족 관련성은 PRIMUS(Staples *et al.*, Amer. J. Human Genet., 2014, 95, 553-564)를 이용하여 코호트에서의 관련성 및 관계를 추론하도록 유전자 데이터로부터의 혈통에 의한 동일성(identity by descent: IBD) 유래 미터법 및 이 가족에 대한 보고된 가계와의 상호참조에 의해 동일성을 통해 검증되었다.

[0270] 전장 엑솜 연관 분석(exome-wide association analysis: exWAS)은 유전의 상가 모델(위험 대립유전자의 0개, 1개 또는 2개의 카피)을 가정하여 본 발명자들의 발견 코호트에서 aPTT에 대해 수행되었다. 본 발명자들은 연령, 연령 제곱, 성별 및 집단 계층화를 설명하기 위한 처음의 4개 주요 성분에 대한 공변량 조정으로 8 이상의 소수 대립유전자 수를 갖는 모든 변이체에 대해 선형 복합 모델을 수행하도록 가계 복합 모델 분석(Mixed Models Analysis in Pedigrees: MMAP)을 이용하였다. 제1회차의 분석에 대해, 신호는 $P \leq 1 \times 10^{-6}$ 을 갖는 경우 추적 관찰에 선택되었다. aPTT에 대한 몇몇 널리 확립된 연관 신호를 재생성하는 것 이외에, 신규한 연관($P = 8.4 \times 10^{-7}$)은 유럽(MAF = 0.002)에서 희귀하지만, 아프리카계 미국인(MAF = 0.07)에서 더 흔히 발견되는, *SLC14A1* 미스센스 변이체(V76I)로 확인되었다(도 1 및 도 2).

[0271] 이 발견에 대한 추가 지지를 제공하기 위해, 본 발명자들은 5,892명의 유럽 혈통 GHS 참여자의 독립 하위집단에서 분석을 수행하였고, PLINK v1.9를 이용한 고정 효과 역분산 가중을 이용한 발견 및 재생성 코호트에 대한 연관 통계학의 메타-분석을 수행하였다. 본 발명자들은 재생성 코호트에서의 명목상 유의한 연관($P = 0.035$) 및 전체 메타-분석에서의 응고 시간의 증가와의 연관에 대한 강한 증거($P = 1.1 \times 10^{-7}$)를 관찰하였다(도 3 및 도 4).

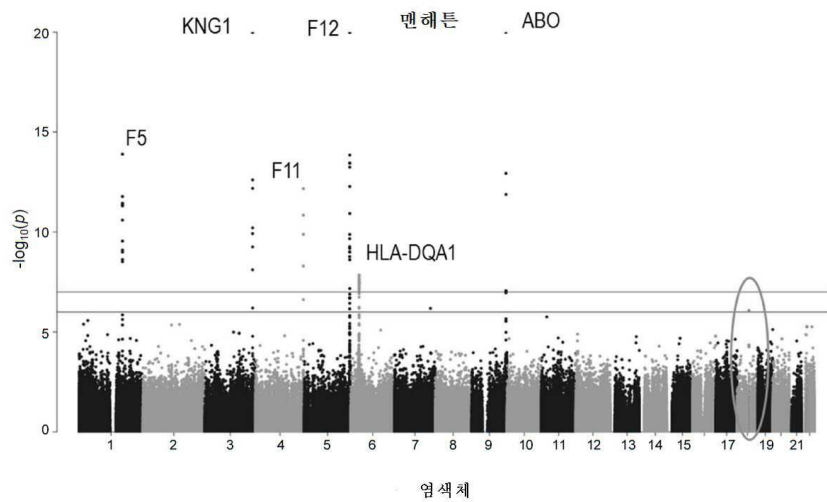
[0272] *SLC14A1* V76I의 임상 관련성을 평가하기 위해, 본 발명자들은 유전자형 및 표현형 데이터를 갖는 96,180명의 다민족 개체에서 혈전증(CAD)의 측정과의 연관에 대해 피셔 정확 검증(Fisher's Exact Test)을 수행하였다. CAD와의 *SLC14A1* V76I 연관은 7개의 상이한 데이터세트에서 독립적으로 평가되었고, (1: GHS 데이터세트로부터의 2,178/24,407명의 유럽 혈통 CAD 사례/대조군; 2: GHS 데이터세트로부터의 13,713/38,005명의 추가 유럽 혈통 CAD 사례/대조군; 3: GHS 데이터세트로부터의 18/765명의 아프리카계 미국인 CAD 사례/대조군; 4: 3,896/3,575명의 독립 유럽 혈통 사례/대조군; 5: 887/1,142명의 독립 아프리카계 미국인 사례/대조군; 6: 4,620/1,496명의 독립 유럽 혈통 사례/대조군; 7: 925/553명의 독립 아프리카계 미국인 사례/대조군) 요약 통계학은 PLINK v1.9로 고정 효과 역분산 가중을 이용하여 메타-분석되었다. 종합적으로, *SLC14A1* V76I는 이들 7개의 코호트에 걸쳐 CAD에 대한 보호 효과를 입증하였다($P = 0.016$, $B = 0.81$)(도 5). 추가적으로, 본 발명자들은 대만인 코호트(c.510-1G>A, 374개의 동형접합자, 1개의 소수 대립유전자 동형접합자)에서 CAD와 *SLC14A1* 예측된 기능 소실 변이체 사이의 연관을 평가하도록 로지스틱 회귀를 이용하였다. 본 발명자들은 비보균자와 비교하여 CAD의 감소된 위험을 갖는 *SLC14A1* c.510-1G>A 보균자를 주목하였다($P = 0.02$, $OR = 0.71$)(도 6).

[0273] 실시예 4: 검출

[0274] 대상체에서의 소정의 유전자 변이체의 존재는 대상체가 응고병증 또는 관상 동맥 질환을 갖거나 발생시킬 위험이 증가된다는 것을 나타낼 수 있다. 혈액 샘플과 같은 샘플은 대상체로부터 획득될 수 있다. 핵산은 흔한 핵산 추출 키트를 사용하여 샘플로부터 분리될 수 있다. 대상체로부터 획득된 샘플로부터 핵산을 분리시킨 후, 핵산은 유전자 변이체가 존재하는지의 여부를 결정하도록 서열분석된다. 핵산의 서열은 제어 서열(야생형 서열)과 비교될 수 있다. 대상체로부터 획득된 샘플로부터 획득된 핵산과 제어 서열 사이의 차이의 발견은 유전자 변이체의 존재를 나타낸다. 이 단계는 상기 실시예에서 및 본 개시내용에 걸쳐 기재된 바대로 수행될 수 있다. 하나 이상의 유전자 변이체의 존재는 혈전증 사건 또는 관상 동맥 질환을 갖거나 발생시킬 대상체의 증가된 위험을 나타낸다.

도면

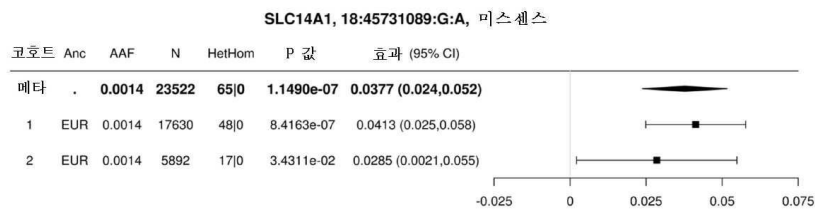
도면1



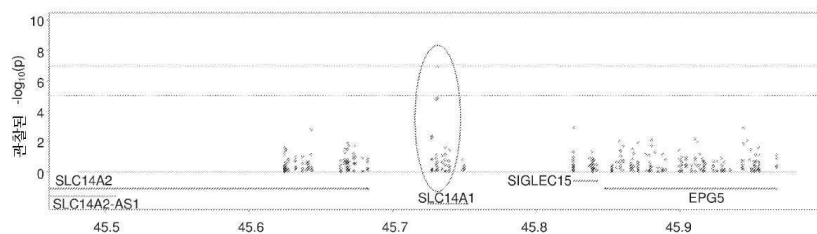
도면2

SNP	CHR:BP (hg38)	유전자	효과	AAF	예측	효과 (%Δ)	P 값
rs113029149	18:45731089	SLC14A1	Val76Ile	0.002	1/5 손상	4.3	1.63E-07

도면3



도면4

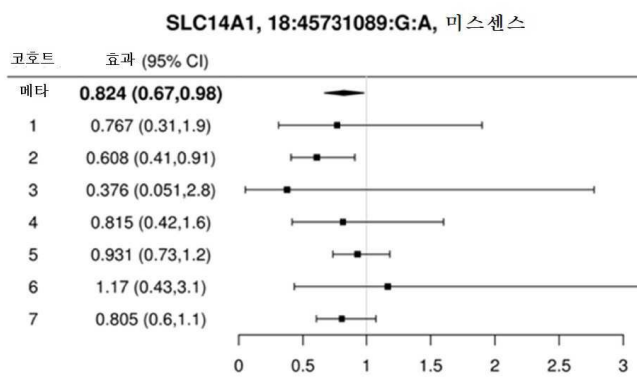


도면5a

SLC14A1, 18:45731089:G:A, 미스센스

코호트	Anc	AAF	N_cs	HetHom_cs	N_ct	HetHom_ct	P 값
메타	.	0.0047	26237	307 3	69943	565 15	1.604e-02
1	EUR	0.0015	2178	5 0	24407	73 0	7.589e-01
2	EUR	0.0016	13713	29 0	38005	132 0	1.391e-02
3	AFR	0.07	18	1 0	765	92 8	4.101e-01
4	EUR	0.0023	3896	16 0	3575	18 0	5.511e-01
5	AFR	0.073	887	123 1	1142	164 4	5.642e-01
6	EUR	0.0019	4620	18 0	1496	5 0	9.054e-01
7	AFR	0.07	925	115 2	553	81 3	1.461e-01

도면5b



도면6

SNP	CHR:POS	영향	사례 WT:HET:HOM	대조군 WT:HET:HOM	OR (95% CI)	P 값
rs78937798	18:45734273	c.342-1G>A (스플라이스 수용체)	3884:82:1	7723:236:0	0.73 (0.55, 0.96)	0.022

서열 목록

- <110> Regeneron Pharmaceuticals, Inc.
TESLOVICH DOSTAL, Tanya
BACKMAN, Joshua
- <120> Solute Carrier Family 14 Member 1 (SLC14A1) Variants And Uses
Thereof
- <130> 189238.00902
- <150> 62/555,440
- <151> 2017-09-07
- <160> 14
- <170> PatentIn version 3.5
- <210> 1

<211> 28394

<212> DNA

<213> Homo Sapien

<400> 1

acacagagca gagtggggct ctgagtatat aactgttagg tgcctccctc cagcaccatc	60
tcttgagaag cactctccct tgtcgtggag gtgggcaaat ctttatcagc cactgccttc	120
tgctgccagg aagccagcta gagtgggtgta agtactcatc cttatttcta ttcatttcca	180
actattcatc atttggggct tgtcttcaca gttctaagtt ttgctctttt tcttaatgaa	240
gaaaatgttt tatatcaccg gaattgatca gaagtagcaa aatcagagtt ctggtagact	300
agaaagcaat ttaccaaagc cacaggtctc ttcttgaag ctcaaaggca tgcctttatt	360
cgtgatttct gaagcaaggt gcatgcagca cctgagctga tgtggaagag gttttgcagg	420
gaggtgtcca ccaatgtgc tcaatgattc tgggttaatc aacactatta ggagtttcag	480
gttgtgttct tgaataata atttgggctg tgttcttgaa ataagttcga ggcgagtgtc	540
tacaagactc aaaagaaaa agtgggccac tgggaatggc ctttccagt gatggattta	600
tggactctc tgtgtgtgct gtcagtctga agggaatgtt cttgtgcacc catcgggaga	660
acaagtcagt cacaactgaa gccacgaatt tggcagcttc cttgcagctg cactctctgg	720
agtctggaat caagacttct gggagtagtg tttccaagg agggaagtgt ttaaccagg	780
acacaggaat atctgacagc attttctttg tttccaatta cagctttaaa gaaaactggg	840
catctctgc tacttaaaa caaaaactac ctaaaataaa gattatagta agtaccaaat	900
aagtgtaaat gctgaaagtc tctttattat gctagacatc gagtgtttaa atgctttctt	960
ctatatccat atccaacact tcatattatt tttaaaagta atagctgaag catggaaaat	1020
tgaagacttc aggtctctcc aattgcacaa atttctaata catgctggca atagaatata	1080
ttttatttcg tgtaataaaa tagaggatat tagttgacct gaaatcttga tattgccttg	1140
tattaaaatg ctaagcactg cttcatctta ctagtgatct ggggtatgaa aagtgccttt	1200
tgacttctgc tggaaagctc ttcaggtgca gcttccagga tattcttggg atgttaactt	1260
cagcacacat aagccttgct gtagatgtgt cagctttgag gcacaggag acatttgttt	1320
gtcagagagt aactgcttct ggcaagggca tagggtgaaa ctggggatag cagagctctt	1380
tctttgtggt tgitcaacc ccacccaag attagttaa agtgaccgtg aagatagtct	1440
gtgccaccg catcgctaag tctagacct cctgcatac tccagcacac agaaactgct	1500
gcttcacttg tttgttgact tgaaccgaac cttgggtggc attaatgtgc ctggccaag	1560
actgaaaaat taagaaccac cagagctgac ctattccata agaccagtc tgcctgccac	1620

gtactgagtg aaictggatg atgccacac tgatccttgg ttttctcttc tataaaatga	1680
aggcttgaac tacgtggctc ctaaaatcct acctagctct caaatctctc ttggttctag	1740
gaaaatattg atgttgagct caaggaagg gttctccaag gtgtgtgatt ttggtggtag	1800
aggaaaggcc ggtgccaggc aggggcagaa ggagacgtg tctacactga gaaaatgtga	1860
caaccctgc ttgtctcttt ttctattctt cattgtttct tatttctttg tttttagctt	1920
tatataacat gagagcccta ccaactgggtt tcttaacctt ttgttcttta tcaaataaaa	1980
atattcataa tgcaacatgc aggcacatca gtgtggtaca gaactagcca gctagtttac	2040
tataggtaaa tatacacaca tgcattgcaca cacacaattt ttacctgaga catgtcagaa	2100
gtgtttccta aaattgtgga tttttctgag tcattctggt aaagggtagg ttttcaggtt	2160
ttaggccaag ccagaagaag aaagtataaa cagaataaac aacaggggga gaaaaagaga	2220
aataccacac acacaactgg aacttctggt aaaagagtga tattcttggga tgcaatggaa	2280
gttttaaaaa ggaaaaagaa aatttataaa aagctgccac atttgtgga ttcaactaaa	2340
aactgtttat tattaacaaa gtgatgttca aaatttaaga gttcttggcc tggcatgatg	2400
acttatgcct gtaateccag tgttttggga ggctaagggt ggaggatcac ttgaggccag	2460
gaattcaaaa ccagcctgga caatacaatg agactttgtc tctaaaaaaa aataaaataa	2520
attaaataa acacagctgg atgtggtggc acaggaaaaa aaaataccat ttaggagtct	2580
cttaaaggca gcttgtgaat gcttacaaag cgtggctagt atcttattac agaaaacaga	2640
gccacatca tgcattcttc ttctcacatt tcataaacia ggccaaggga aactgctgtg	2700
gggcaacctg ttgcttttgt gttggtcccc aagatgcagc cctcacaatc tgccccaaa	2760
cgtgtcagaa catgaacccc ctctcccc tctggaagaa gcaacctcag atccaacagc	2820
agagacagc agcagaacia aatctgggca ttggtccctg tgtaggatgg ctccccgta	2880
ttttttttt aagcaaagta aatgaacatc aaatttccat agtcagctgc tgtctttctg	2940
cccactgaga gctcttttgt gaaggcaaag tctccttct tcattagcgg tctccatgt	3000
ggggccacat ctccccac caggaaacca gtgggcgcgc tccagcccc ctcagcttgc	3060
cttttgcgtg gtcattagag ctagggcaca cgtcatgctg attcacatat ttttgcctt	3120
tgtcatgtat tgagaaaaag taaggatgaa tggacgtct ttgattggcg gcgctggtga	3180
cgcccgctat ggtcctgttt ggaaggaccc ttttggaaact aaagctggtg acgcagcgcg	3240
cagaggcatc gcccggttaa gcttggccct ggcagatggg tcgcaggaaac aggtatgctt	3300
ccttcgtgca gcctctggct cggggaacct gggagcctgc tccaaactct ggtgtatctt	3360
ttccgggcag agcctgggaa gtgggggttg gctgtgagct aagccaaagg cacagggatc	3420

ttggtccaaa aagcccatg gcgctcacct tggtttagag gctagacat tgagctgaga	3480
agttttgaca gccatggaag agctggggat aagtcacctg gggttttacg tttaccctgt	3540
gtctatttta ttagagtgc ttttacttat tgcctttct tcttagttga aattaatggc	3600
ctgcttcact ggggctaaga tgtttgaaca ttagcagaag gtcctggctg catagccttg	3660
ccttgcttc ccagtttaga tgtaaggact cttaaagttc cctaagaaat gcaaatat	3720
tagcatggca aaattctagg ccaactacaa ctgtaagttt cgtatcttc ctaagtgtt	3780
ctcatgcctg acttctggag caaggagtca ggtctccag gggctctaga agggctcagc	3840
tgttcagaat aaatggttc tggggactct aaaatagcag caactgtctg cccaggtcat	3900
gagaagacc cttctgcag gacatcctag ccctacaacc catccaatt atgttgaaat	3960
tagattcaca aatggcaata agtcttctat atgttgggct gtcgatttgg agaaaactag	4020
tttaactttt acttaacttt ggggtggctca acaggagact cgggccgctc aggtctctca	4080
tcacgtctgg ccagttctat tatcaggttt cgaatctgta tctccaaat cctgaggtg	4140
atgggatatt tcaagccctc taaaataaat aaatatatgc tgggaatttt gagaacatga	4200
atttgtttat tctgaaatgg tccatgttcc tgttttggga gttgatggaa aatgccactt	4260
gagtgttttc attgatgct gccaccttag ggttttatag attcagttcc agaaactcaa	4320
ggcatttatt tctttgggct gcttgcctt gcctgagctg aagcctgatg cctccataa	4380
gttgggtatg ctttgaaaat gggtcactac agcagaggca tgggcttatt aagcaatatg	4440
ttcagctatg aaatttgaag agggagataa tctgaaaata aatgacagcc accacttaga	4500
ttatgaaata gaagtacttt ttcataagtg cttaattatt catacgttt tttatcttta	4560
actatggagc caactcagct ccatatggac ttaattttgg ttcctgacct ccaagattca	4620
ttgcaagtca cacagatgtt ggtatctaac attgttttac cgagataaaa tgaccttggt	4680
ctggaatgca ttgtataaaa agctgctttt ttgtgtaaag attaatagtt tggcattgtt	4740
taaaaagcag aatggttagt tgggcagtga ggtaatacaa ttgaaatgta attgctacca	4800
ataaatcagt tacccatatt gatttcttta ctgggattaa tagaagccaa agctagagtt	4860
caactttttt taataggtat aacttagtat ctgttcattg ctatttgta gctatggtaa	4920
atggaacaat gatggggcca gaaatatcca ttaggacat ttgatcacag cctggcaaca	4980
cagagaagac aggtgtgtt ctctatgtgg gctttcagtg tttctttggt agtgtcttat	5040
gtggctgtgg cttaacatt ccacaattat gccttcagg gtcgatgat tttggcgttt	5100
ccctgcttc caattgacct ggctgtgctg ttggctgttc ttgcacactc aagtggtttt	5160

tgccattggc ttctccctc agcctgcctc tgggattatg ccactgctat tcttttttat	5220
ctaccatcag cacaatgaaa tcatcatttt tgtcttcaag gtaccaaatt ctggtgatat	5280
tgggtgctttc ttgcagctac ttatcatgag aagtgaatgg tctcatagtg aacacagtca	5340
tggttatagt gticatacgt tccagagaca tgtttcctat aattatgccc tgcacatttt	5400
tctatcatac aatccttaga ttacagctct tgggttttca acagctttgt ccaattccat	5460
ctttcccagt ttctctacct tgatgaaata tccttcttgc ctggtttttac atattttaat	5520
aacaaattcc aaaagtaaag agtatctgag gcagtcacat gacataagga caaatccaag	5580
ccatcttgga cttgcagagg gtggggagac cgtgtcaaca cacacaattt taaaaatttc	5640
ttccctttca atctttttaa aacaaaactt ttataaaat aaaaatgtaa tttaaaaagg	5700
ctacctgtct tggcaagtag ctgatcagcc tgcattggtg agcaggccat tccataacct	5760
ggtttcttgc tcttaattg acagcatgga gtaacgtac ttaatttcag ctctttctac	5820
gtgatttgac tcattctgtt aacattaact gtttttcagt cttctcaact agactgaact	5880
ccttaagtgc aagaaatata cgcttagtaa atgtttgttg gaccagacac tgcaccttat	5940
gaaattaaag accagaacat tctcatggta gcattacaga cactgatggc aaaggtactg	6000
tgggatttgg gtttggctaa taagctctgt ggtgggtgtt cagaaggaaa atggtgctct	6060
cttagttcta tggaacatag tggtcagat cttctactgt aaccaggccc aaagctggct	6120
aatctggagg gctctgcctt agggatactt ataagctctg tcttccctc aaggagccag	6180
aggaagagat agccatggag gacagcccca ctatggttag agtggacagc cccactatgg	6240
ttaggggtga aaaccaggtt tgcctatgtc aagggagaag gtgcttcccc aaagctcttg	6300
gctatgtcac cggtagcatg aaagaacttg ccaaccagct taaaggtatt tatcctttca	6360
cattttggag agacaggaga agtagctttg ggggaaatgg tttcttgta cttctactta	6420
taccttagt tatattctcc aactttttat agatctcttt actcaccatt tttctacttt	6480
tatcttttaa cctgcaaacc tctccatttt tttttcttat ggagacagta gccagggccc	6540
agctcatatt agaaggcacc tggcttcac ctgtagtttc agtacttaaa acttaaat	6600
attcctttgg cttcagaatt tgtacctata agcatgaaaa taagtgcatt agatgctttc	6660
aggagcttag attctaggag gggcagtggt ggttgagcat acagtagata gaggctttca	6720
gggatctggg tgccactaat gcaacaatgg gttgagagag aaatattaaa gaaatatcaa	6780
aaatgtttca ctccaggag gttttgctga ttttgctcag ggtgggcctg tggttgaaga	6840
gtatcacttg gcagcttcct tagctctgct ttacctcacc ccttcagac aaaccctgg	6900
tgctccagtt cattgaactg attctccggg gcataatcca agtgggtgtc gtcaacaacc	6960
ccgtcagtgg aatcctgatt ctggtaggac ttcttgttca gaaccctgg tgggctctca	7020

ctggctggct gggaacagtg gtctccactc tgatggccct cttgtcagc caggacaggt	7080
aggtgtacc tttcaagcct tctcagctcc cttctgagac acaggggctg accagttact	7140
gtgggcaaca gtgataaaac cacatccttc ccaggataaa caacatttag tccacagaac	7200
tgtttatatt tgtttttagt cagaggtcag ggaatcagtt acagtctctt gctcttgata	7260
tctgaataaa tggtctgtct aaatgatgcc agattcttgt ggcattacgt gctaaccaga	7320
actaagctac aagtatttcc ctggagaggt tctgaaggga tcttctttaa tgattgataa	7380
aattatttgt cgtcagcatt ctatttggga aaaagtgc atgaattcag aaaaagtttt	7440
agtggcttaa taacccccgt tatacttgt tgctatgatg agtttaggaa actcattctt	7500
catagacagt gcaaaggta gctcagctcc tggagaaaag aataacctg aattccaatt	7560
gagtggattc tgacttaaga agccttagtg agtcttctga tatattgatt agattaaaaa	7620
tagcacacac ttataaatt gatctgtcat tgaagaagt atgagctgac tctcaccagg	7680
gcagtagata gctccccact agccagttcc tttaggagg gaaccagtat tccaggtgtc	7740
tgagatcaac gcataatccc aatccccagt gtggtcatta cacaactaag ctcttgtaac	7800
actggctgca aattgcctaa agaggtccgt ggggagagag ttagcaaag ctccactttt	7860
ctatcaattt caaggagtct gatttgctcc ctgtagaagg ggattttata gcttaggtta	7920
aactctattc caatgcagc caagaaaagg tctcctcagt ttggggatgg agtctataat	7980
tgtgccatac tgaatattcc ttatgattt tgctctgatg aaacatgatc aactcattt	8040
ttgtcagata ttatttagaa gacaagtc ttatatgtgt tagtttcaa tgttttactt	8100
tccttggctt gaaaagactg cattaaaatg gaaattctct gttttaagta aatatatgtc	8160
ttctgtggc tttaaciatg gcattccaca attttagat gtggcatta atttccact	8220
gatcaaactc aagcattaac atctccaagt cagttgtga gaggacaagt ctgcatggct	8280
ctctactgtc atgtgtagtc ccagtctctg agttgtacct ttgcaaattg tatcacctec	8340
catttgccct caaggattat ttaagggaac caaagaactt ttgaataggg aacccacat	8400
ttaatgttca tctggattaa tgtacgtgac atcatcttg ctgttgcaat ggtgcctcct	8460
ggcccagtta gaaacaagcc aagaagcagc tgtcacacta tcccttacca gcccctgcag	8520
tgtggctcac tggtatagc acctcctgct cgagcccagc attaggcctc acctactcac	8580
ttcaccatct ttactcccc atccccctac agacatcacc cttgagtgac aggcccttgg	8640
gaagtggatc ctgtgccttt cacggtgcca gacgttgcca actctcagag ctgtgggaat	8700
cctgccttgt caggtaac aatctagggt cccatcaatg gtggattata taaagaatat	8760
gtggtgcata tacaacacga actactacat agccataaaa aggatigaaa tcaagtcctt	8820
tgcagcagca tggatgtatc tggagaccaa tatcctaagt gaattaatgt agtaacagaa	8880

aatcaaatc cacacgtttt cacttacaat taggagctaa acactgggta aacacggaca	8940
tggaatagt agacaactgg gactccaaaa gaggagagga agggaaacaa gtgttgaaaa	9000
cctacctatc aggtactttg ttactatatt gggtgacgag ttcaatagaa gcccacacct	9060
cagtcagcat catgcaatc atctatgtaa caaacctgca catgtacccc ctcaatctaa	9120
agaaggagaa gaagacgggg aagaaatgag attgaatact aagcaaaaag taacctcaga	9180
aagaactggg tgctcaacat gcacataatt aaatgggata cttctccaag taagagaaaa	9240
gcaattgttc ttctttgcaa taactttgaa atgtgcgttt ggagacaaca aaatagaagc	9300
atcaggacac aaaaatgtat actaacctgg aagattaatg ttgataagat caaagacact	9360
gtgaaagtga atttacattt caggaatctt atatctctca ccaagaaatc aaacttaagc	9420
aacagtttca tatgctaaaa gcgctcttca agtcagaggc tcttgattta aaagaataac	9480
tttccaaagg aaaggtctaa agaaaacaga gcagattgcc ttactaaact cccctttcct	9540
ctcagccact gtagacctgt ctttagccgt gacacctgta gagggagtca ttctctatca	9600
ggggtcccca acccctgcac tggagacagg tacctgtctg tggcctgttg ggaactgggc	9660
cgcacagcag gaggtgagcg gtgggcgagt gagcatttcc acctgagctc cgcctcctgt	9720
cagatcagca gaagcattag cttctcataa gattgcgaac ccattatga actgggcatg	9780
tgagggatct aggttgcttg ctecttatga gaatctaatg cctgataatc tgaggtgga	9840
cagtttcatc ccgaaatcat ccccatctc ccatccatgg aaaattgtct tccatgaaac	9900
ctgtccctgg ggccaaaaa gctggggacc actgatctaa atgcacattt atattttat	9960
ctatgtatat ttactttcat gtctttatta gttttgtac gatgtttacg tagactttga	10020
aatacatttc caaataaat ctcatTTTTT aatatgaata tgatctggaa gttactagt	10080
ttatttatgt gcaagtcaa ccaaagctca ccaggaat gtccgtgctg tgtctcttgc	10140
cccacaggtc attaatagca tctgggctct atggctacaa tgccaccctg gtgggagtac	10200
tcatggctgt ctttcggac aaggagact atttctggtg gctgttactc cctgtatgtg	10260
ctatgtccat gacttggtaa gttacaattg gttttcaaaa tgcctttttg aaaaaaaaa	10320
catggcagaa ggagggaatg ggagttgtta tatggcagag tttcagtttt gcaagatgaa	10380
atatgttctc tgaatgtata gtggtgatgg ttgtacaaca atgtgattgt ccttaatgtc	10440
attgagctgc acacttaaaa atggttagcc gggtgcggtg gttcttgttt gtagtccaaa	10500
ctattcagaa ggctgagggg gaaggatcac ttgagcccag gagttagggg ctgcagtgag	10560
ctatgattgc gtcaccgcac tccagttctc cgaacctcct tgcttgggct aagtgaggag	10620

gaggaggagg aggagaagga tggaaaggag gaggagtagc aggaggagca ggagggcaag 10680
gagaaggagg aagaggagca ggaggaggac aaacagttaa aatggtaaat ttaaaattgg 10740
attccagtag attctgtcta ttggaacag aaacaacat tttaaaagat gtatatattcc 10800
ttacaaccag ttatttggcc ttttgtctga tctggctaca catccactaa tacctctcaa 10860
ccagaggtgg ctgcacattg acatttccat ggggaaggga aacagtgtg caatgaagat 10920
acgagtgcag gtgtcttttt ggtagaaaca cactgatgca cgtggccccc acatacatt 10980
gactctccc tccaagact ctactgtcat tggctctcgg tagcgcctgg gctttgggag 11040

tttctaagc ttccagatg actctaaagt atagccaaag ttgagacca ctctctccat 11100
cattgcctct caaacttgag caatatgaga atcacctgca gggtttgta caccacaggc 11160
atctgtctcc cgccccagg gtttctgatg cagtctatct ggggtggggc ccgagaattt 11220
gcgtttctaa cgcattccca catgatgtg ggagaaccac tgtgcctacg tgaattcccc 11280
cttaccacc tgccccccag gtctccctta gaaaaattt ttttctgaa ttcctttttt 11340
ttcaaacca aatccttcaa actagttttt atgttgacaa tgtcttacat ctttttctg 11400
gaaacaaaga tttccttctt tctatattgt agttaatat aaaatactaa tatgcacata 11460

aataagcaca gcctgctgtg ggcagtgtct gcagaaggga tgcccaccct tactgtaccc 11520
acgggtgtgt ggacgaggac ctacctgtag agctaaactc ttcaggaagt aatttgggcc 11580
ctgctctgaa gaataggttc gtgggaaggga ggcctagcct gtaagtgtc accacgtcc 11640
cttcacaat ccaggaaaa gggagtctg gtctttaagt gatggctctt tgattgggcc 11700
aacaagttag agcctatgag ggacctcggg accatgcagc ccagccccac agtttatggg 11760
ctctgaggct aaggagatgc gccttgccca ggtcatgcaa tttatcaaca gctcaaggac 11820
acacactctg cccaccaac tgtgatatca tttctcca gctcacacta cctgcatcct 11880

tgaacgattg tttctctttt caaaaatag gtatatataa gaaataatat ctgccaatc 11940
agaatcaggg ttgcctctag tggggaggga gggacataag agcaagtga gggacaaagg 12000
ggactttaac tatgtagata atattttatt ttgtatgtca taagtacttc aaaaatattt 12060
ttaaatctc aatatatagc tcactctgag caacccaga gtagaatttt tcaaaagcca 12120
aataagctga gatttgattt ttactttat gtaatattha ctgcctctat aataggattt 12180
atcccaagtt tttttctgt ggcaaatgtg ccaacacaac acgtaagggg cctgttgga 12240
ggtgaaacaa agcccccca gagtatagcg attccgtgtg tcagcctgct ttgtcacatg 12300

cacattcttt tgctctgttc ttttttagc ccaattttct caagtgcatt gaattccatg 12360
ctcagcaaat gggacctccc cgtcttcacc ctccctttca acatggcggt gtcaatgtac 12420
ctttcagcca caggacatta caatccattc tttccagcca aactggatcat acctataact 12480

acagctccaa atatctcctg gtctgacctc agtgccttgg aggttaagaga cactggcttc	12540
tcacattcgc cctggctctg caagatacgc aatggcctcc tggtaactg tccacgggtg	12600
tcagagtctc ctagatgctc aggactatgg tggcctttct gccttcatct tgccatttaa	12660
agcatttgtt ctactccaga gcattagggt ctaagggtt ttttaaaatt actatttagt	12720
caagctgatt tttctgcctt tccccctaaa catctacagt gctaaccceca gactacagtt	12780
ccactgggag tcaactctatc gtaagcttgg ggggtgggggt gatgggagcc agcccttaag	12840
gcatgtggcc tccagcctgg ttttaaatct tccatagtct actccctcca atcaaaaaac	12900
tggatgctta ctcttagagc ttctgacaga acctctctat tctgcttttc cttatggcat	12960
agctcataga acatctacaa taatttaggg ttcccaagct ttggtaggca tcagaatcac	13020
ctggggagct ttaaatatcc aaacaggctt catctcagac cctctaaatc acaatctcta	13080
agggtggggc ctggaacctg ttttaacaaa ctcccaaat tgtgatgcgg gccagagttt	13140
gagaaccact gtatcaagggt gtgaatccta tgtatctctt taaagatggc tataaagaga	13200
ttctgtatct tttaaaacct ggttaaccceca aatcaaatc cagctcttcc tgttgggtgtg	13260
taataaatat gttaaggtt tctggattat caagaacaag agaacacctg aaattagaag	13320
aaaaccaaag aaaccttacc ttttaaatgt gctctccac tgtcagggtt tgaaacgccc	13380
ttttgtcttc ttgtttgagt gatcaaaaca cagaggagc tcaagtcacc ttctccctag	13440
cttcttgcca gaaaactaaa gggagcacct ggaataaatt cagaaggaaa aaatcaaaga	13500
ttcattagaa ctacctatga aaaataacag tataaaatag cattaatcga tctagaactg	13560
cactaacaca ggagcctcta gcccctatgt gctatataaa tttagatgta gattagttaa	13620
aaatttaggtt cctcaacctc tctagccaca tctcaggtgc ttgatagcca cacgtggcta	13680
ggaccactgt tattagacag cacagatata gactattcca tcatctcgga aagttatcct	13740
gcacagtgtt gatctggggc aggggaagcc ttgtccttct cactctgaat gaacagccca	13800
tcctcagcac caacccaac cctatggcta cctgagagag agttctgcag ccaagtccaa	13860
aaacaaacaa acaaacaaaa aaagcatatg ccacttttgc caagttccct ggtctagaaa	13920
tagcaaatg tctagacatg aagactcagc atgggctgga agaatttaga gtccatctta	13980
gggtagagtc aaactcacac tatggtcttg tgcccttagc caatgttaga ctcagcctaa	14040
tataagaggg gagaagacac tccccctgt gccaaagctg gggctccctc tggtagagtc	14100
actgcctcca gaaggtcttt ggtacataca cgacctagca atggtaggaga gggcaagatg	14160
ggaactgagg aaaacatctt tcagtaaatg gccttgctca aaaggacat gctatggcta	14220
attatgccta tctagccct accagaagtt cagctgtaaa gaatgatcac ttgttaggtt	14280
cagttaaacc ttgttactc ctgagaactg caattctgtg aacagaataa ctaaatcag	14340

gcctcagcca gaaagtagaa ttatgacatt tccatgtatt tttgtgttt gagacctgct	14400
tgacagtgtg tcataactag aataagctaa aaatatcttt gtttaaatga atacatgttc	14460
cacttaatga cagaaaagta aattcacaaa cttgctaaaa attacttcta aattgtggac	14520
aagataacct ggctttgggt ctctggcttt agtgtaagca tccaaattgc atagtataa	14580
taatctctat tgaacatagg gatgcatgga tagattaaat caccctcaac actgatggac	14640
atttgaaagc aaaagaagtg tcagctgtgg tccttgccat ccccagtagg aggcaaggca	14700
gacccata gccaggagca gtgagtggca ccaagctggg agcttaacag tgaccaaggc	14760
caagtgtcag tgcaagcagg agagcacagg gggagctttg agaaggcatg tgttgcacgc	14820
accaggaag ggctgggtga tctctgggga taaagctgaa ggatgactgg gatTTTTctg	14880
taatcaaaga gagagaattt taaatggtat taacactgtt cttgaaagag gtaaggtatg	14940
tccaatctaa aattacattg taggagtttg tgggtgtcct gtgggtttct gttcagtgt	15000
tttggtagcc tcatttttct taaatttctt ttgcagtgtg tgaaatctat accagtggga	15060
gttggtcaga tctatggctg tgataatcca tggacagggg gcattttcct gggagccatc	15120
ctactctct cccactcat gtgctgcat gctgcatag gatcattgct gggcatagca	15180
gcgggtgagc acaagagccc ttaccaaata ttgagcacct cctccatccc atgcattgcc	15240
tcaggcatct tctgtctcc agatcttct tgagatcttg gcttcctagg gaccaatggg	15300
agttcccggtg atgttctctg ctaactttca atcccacct cagtttctt ccagaacatc	15360
ctgcttttag tctgagttc tgaccctcc tgtcttaaca ggactcagtc tttcagcccc	15420
atttgaggac atctacttg gactctgggg tttaacagc tctctggcct gcattgcaat	15480
gggaggaatg ttcatggcgc tcacctggca aaccacctc ctggctcttg gctgtgtga	15540
gtctcccag cccctggggg agggctgtc atgactacag gatctcaatc aaggataagc	15600
agtaaaaacg gactgcatga aaaatcaggg ccagggttct ggcttgagcc cacttgctgt	15660
ctaagtgtg gaacaggaca agtgacgtcc cctctctgag agcattaaaa tcacctctgc	15720
ctacctctct gatgatttg aaggcaggag cctattgagt catattaata tcctaaaaca	15780
tggatgtttg ggaggataga aaaagaaaaa tcccagttat tcttcagctt tatccccaga	15840
gataaccag cccttcctg gtgcatgcca cacatgcct ctcaaagctc cccctgtgct	15900
cacgggctct ccagcttgca ctgacacttg gccttgcca ccaataagct cctagaatgg	15960
tggcactcac tgctcctggc tgtgaggatc tgccatgcct cccactgggg atggcaagga	16020
cctcagctga cactcctttt gctttcaact gacttgtctt gcgttcttca aactagtgt	16080

ttgaccaac aaactaaacg ggaataaactc cagctaaata cagagcaatg tccccctgta	16140
aatcagggtt gattacattt acccctttga gtgagcatca cagtaacca gccattctaa	16200
aaattcagaa tgcatacagaa tcacctgaaa gacttggttaa aacacaaatc gctgggcccc	16260
ctcctcagtc tgattcagcg tcagagataa ggggaagaat atttcttttt ttatttttct	16320
aaaaaacagt ctattctga gccaaagatcg cgccactgca cttcagcctg ggcaacagag	16380
caagacttca tctcaaaaaa aaaaaaaaaa gagaaaagaa aaaaaaagaa aaagggtctc	16440
attctgttgc ccaggctgga gtgcggtggt gtgaacacag ctactgcag cctcaacctc	16500
ctgggctcaa gcaatcctgc agcctcagcc tcccaagtaa agtagctagg accacaggcg	16560
tgccaccatg cctgggttaat tttttatttt ttatagagat ggggtctccc tatgttacct	16620
aggctgatct tgaattcccg ggctcaagca atcctccgc ctcacacctc caaagtctg	16680
ggattacagg cataagccac catgccgca gaatttcac ttctaacaag ttctcagggg	16740
gtgctgatgc tgttgcctc aggatcacat ttcaagaact gctgtattaa tcctttctga	16800
ctcccagtgt tctagccaga ctacgctgt cagagcgaga aggcacctg agacctctac	16860
tccatccttc ttactttact gttggggtcc tgaggccaga gaggctaagg gatgtgccgc	16920
agggaatctg gacagcaatg ggtaaatcca cccccgaac ccacacttac catccacctc	16980
cagagttatc ccaccgact cctctgcttc ctttttatag cattcaggcc ctcacggcaa	17040
cctcttaggt gaaaacagac tgcattgat ttggatctga aaagctaata gatcccaggt	17100
ggatttttag tggaggctca ttcaccata gcctctggca tgcctaattc aatcaaagta	17160
taagcattta agataatatt ctagagtga gagaatgaga tttgcttggg aacaaaaagg	17220
aggagggata gtgtaatgtg gagaaattat gtctaactta gtggaaatat atgtctagaa	17280
tcagtttatt accagattaa tcaagccaag gtatctaac agttatgaaa acagtgggcc	17340
atgtatcagg cgggtttaga atagatttct gactggcag aaaatgggat ggtaccaacg	17400
gtttctaaag acccattcca ttttgattcg atgctatagc aagggttaaca taactcaggt	17460
tgctgtgatg tagccatgta gatgtcattt tgcataatc ttactatta ctacgtatt	17520
tcacctagct gttctgtga aatgttgaac tccttctcca tattcgttca caaggataaa	17580
ggagaggatt acagacaggt gctgtagcca cctgagttca gctgggttgg aatgtttatc	17640
ctacaacctt tcagctttat tctgagattg gttaggggtt tccacctgag ttcagctggg	17700
ttagaatgtt tatcctacaa cctttcagct ttattctgag attggttagg ggtttcaaac	17760
ctttatttgg gatgcatacc tttatttttc tggaggaagt agccacaaat atgtattaaa	17820
cacacatgat acaaaagaca gtaccaggaa gagcaagggg tttagaagct ttaggtccca	17880
tgcagttcct gcacagagtg ttacaataga gggcagaagc caggcaaggg agtgagccca	17940

agaggacat gcaatctttg tgggagaaga agaagtccat agtacaggat tctccagggg	18000
gccatttcca ctcagaatta tcacaaagta cctccaggaa gaagggggct tttccataaa	18060
tgctagaaaa taagaggagg aattctgttt ggtggaaagt gtggtgcagg ccagcatggg	18120
gacagcctga gcatgtcctt caagatcaag gagaaggcat tttgagcaca ggagatggcg	18180
acgaggtttt tgtttttctg ggttttttgt tgtttttgt tttttggttt tttttttttt	18240
ttttttgaca gagtcttgct ctgttgccag gctggaatgc agtggcacag tggcacgac	18300
ttggctcact gcaacctccg actccctggf tcaagcggtt ctctgcctc agcctcccaa	18360
gtagctgggc ttacaggcac gcaccatcac gcctagctaa tttttgtatt tttagtagag	18420
acggggtttc accatgttgg ccaggatggt ctcaatcttc tgacctcatg atctgtccac	18480
cccggcctcc caaagtgtg ggattacaag tatgagccac cgcacctggc ggggtgtgag	18540
ttttttgttt tatgtttgtg ttgtttgttg agatggactc ttgctctgta gctcaggctg	18600
gcatgcagtg gcacgatctc agctcactgc aacctctgcc tcccgggtcc cggttcaagc	18660
aattcttctg cctcagcctc ccagtagct gggattacag gcatgtgcca ccatgcccag	18720
ctaatttttt tttgtatttt tagtagagat ggggtttcac catgttggcc aggctggtct	18780
tgaactcctg acctcgtgat ccacctgcct tggcctccca aagtgtctggg attacaggcg	18840
tgagccacag tgcccagcta gtgatgaggt ttgacagac catggagaag aatgaagtcg	18900
aagctcttga catgttgttt ccccaaagt ggaatctttg atattttctc aattatagaa	18960
gcagcacaga tttattglat aaaacaaaac aaaaatgtaa tctglataga aatgtatgaa	19020
acagaaagt gaaatactcc atcttactcc ctagagaggg cttttttgcc cccttcttat	19080
aaggatcctt gtgattacat tgggtccatt caatagtcta ggaaattctc tccatctcaa	19140
ggtctttaac ttaatcacag ctgctgctaa ttcccttttg ccatgtgagg tcacatattc	19200
tcaagttctg aggtttaaga tgtagacgtc tttggagacc attattcttc ctaccacact	19260
caccttctt tggatagatt ttttttttt ttaactggtg tagcataatg gttgaggcag	19320
tcaactgagc taaagagctc agactctggt gccagacagc ctggattcaa ttccagcagg	19380
tctgctactt actagegtat ttgcttatga atgtaagcaa attacttaac ctttctatgc	19440
ctcagtttcc ccatcttaga aaatggaagt taccatattt aattcataca gttgttctga	19500
tgattaagtt agttaatgca tgtctgaaac tcatagaaca aatagtgtct agcactcgct	19560
cagcactatt taaaagtctg gaaaaacagt ttttctggtg gatttgcata acttattaag	19620
aatcaagctt gtttatttct tctctcaat tgcttaagtt tatcaacatc tgtatcttct	19680
ccccaaatat gactgatacc caagcctgcc ttacttctct ctgagaaggc ccaccctga	19740
tgactactaa aaccattgat actgtataga atttttattt tggatttgtc gtaagtataa	19800

gtttttgttt tgggtacttg cttatttagg caactgtaaa ctttattaac ttgcttattc	19860
actctgactt agttcatatt aaccttctgt actttttttt ttttgagaca gagtctcact	19920
ctgttcccca ggctggagtg cagtggcaca atctcagctc actgcagcct ccacctcctg	19980
ggttcaagcg attcctatgc ctcagactcc caagtagctg ggattacaga catgcaccac	20040
catgcccagc taattttttg tactttttgt agagacaggg ttttgccatg ttggccaggc	20100
tggctcmeta ctctgacct caagtgatcc acctgcctcg gcctccmeta gtgctaggat	20160
tactggtgga ttactttttc aaagaggggt tgcaaagaga gttttgtttt cttcaaagag	20220
ggtttgmeta gagaccttgt atgctggaga atatcttcat ttacattca tttaaatttt	20280
agtttagcta gctacmeta tcaagattta acattttttt ctcaatattt tgaaagttgt	20340
cctcaaagac tactccattg tcttcttata cccaaattg ctattaagat gtctgmeta	20400
aaactaattc ttgtmeta ttgattttatt tttctctctg gactctctga atttctctt	20460
tgcataatag atatataat ggttttattt cactattatc tgtctagatg taactttttt	20520
ttctatgcta gtaggtactc aagtcctctc aacatgagcc ctcatatctt cttttaattc	20580
tggaaacatc atcagttttt actttgtcaa atcttttcaa ttttccct ctcctctgt	20640
gatttctagt atttgaglac aatactttat gctaagtttt tcataactct tgactttttc	20700
ttaatatattt ccactatct tttctgagg cccttcagtt cagctgattg gcccgatcat	20760
tctttggctc tgtccattgc accgatcaca ttatctgttg agttctccat ttctggttca	20820
ttaattaaat ttactggct ggggtgcagt atcacacct gtaaacccag cactttggga	20880
ggccaaggcg ggtgatcac aaggtaaga gattgagacc atctggcta acacgtgaa	20940
accccatctc tactaaaaat acaaaaattt gctggcgctg gtagcacgcg cctatagtcc	21000
cagctactca ggaggtgag gcaggagaac cacttgaact cggaaggcag agctgcagt	21060
agctgagatc atgccactgc actccagcct gggtgacaga gagagactct gtctcaaaga	21120
aaaaattatc gactgtaggt tgttcagttt gttgtccttc ttttatgta tttgtctcc	21180
tgggatgtcc ctttcttg tctgggagc tcacgtttcc ctcgggatac cagctgtttg	21240
ggtgagtctc tgggcagaga tggaagccca ggttgagct gcatttttcc tggatcatct	21300
aagmetaag ggttccctg ccacaggtg tagaacctcc attgtcaag gctgtggaga	21360
tggtagctgt gtagacattt tatatgataa gtccctttt gctgggggaa gttcagattg	21420
cttctagttt gaaatcatta caaagagtcc tgaaatgaat atttttgta caaatgtcct	21480
tgtgtacttt gtacaagcat ttctgtaaga aagaagattc acctctttt caagaagcta	21540

aattgatggg ttaaaggga tgccaatttt gatttcagtg gatgccaact tcatctccaa	21600
aagagccata ccagtttcca ctgctgccag cagtgtgtga gaggcccccac tgggccccca	21660
caaggtacaa tcagactttt aaatctctgt gcatggattt ttgagacaga tctccagccc	21720
cccttggaag gcaaatctca catgtaaaat gccacagcaa gtttcagctt gtccacatca	21780
ccctgatact gccaaacaaa agaccaaccc tcttagccaa cataaataag tgacagacat	21840
ttattacaga gctgtttttt tatcagtcct cagtggcttt atcaggaagt ggactcagga	21900
aactctgaca gaacctggca ctgctgtctt tctggcctct aagccagagc aactgcgtgg	21960
ccagagaaca tctcaatgtt gttgttttac cagtggagag tgtaaacata ttgtgtatct	22020
cttcccaatg gttgggttat cgcagtggga ctcacctgtg gcagtcattt ggaagggaca	22080
ctatccagga ggagctgaaa tccagtttcc ccttcagtac tcaagggcct tttcttcct	22140
cagctaccaa gaatgctgtc agggctcattg cctacaaact gatgatgctg tgcagaattg	22200
cgctctact gtaaggcttt cccggtccta ctggcgagt cttaatgac atacctacca	22260
ttaaataatc tatcattgt actatggaga gaaaagcaac tttgaattgg agatcacttc	22320
acagcagcat aacagtatga gacgtaaacg tgccaaaagt gaccttaga agtgtaatgg	22380
atattttaaa aagagagaaa gcaacaaggc ctcagtgtct caggggtggt gttgtgtag	22440
agggggcact caagagatca gggacagagg gccccagtgc ttggcagagg gccaatgaat	22500
agttgttaaa ttaattgatt aaatttcaac aatgaatgaa attggtgtaa ccaaggagag	22560
aaacccttct aagccaagcc atgagcacc tctgtctcag agcagtagct cagtcccatg	22620
gtgaaagaga tgcatttaca gctgtgttta tggaaataca agctctcatt tgagattctt	22680
cacctccag taaggcagat cttcaagggt cttttttaca gatgatgaaa ctagattcca	22740
agacagtgat ttgttataca acaataaaa tggcagagct gggatttgaa accagtactg	22800
tttccaaaga ccagccttcc cactagtgt gagacaattc atacgtgaaa gaatttgata	22860
tactattgaa taagaaacac caggataaaa agacaaaata ttggtlaaaag gacagaagtc	22920
tatggttaaag taaatgagga tcacagagcc tctccacca tgtctgccac atccccacac	22980
accaagatag ctgacgtacc agacatgaag acgagatggt gagtgtgtct cacggtgagc	23040
tccggtggcc caagtggctg tgtggccatt atatgaaggc cattcttcag gctgtcccca	23100
tgaacctga gggcttccct gagcctctgt gagccttctc ttcaacaaa actgaggaat	23160
agataattag ctggttgaga tctttgcttt tgtgttttta cactgaaagt cacccatata	23220
ctcgaattac tgattctaca attttttggc cactcaaagc aaataaaaac ataagacgtt	23280
ggctggggcg ggtggctcat gcctgtaatc ccagcacttt gggaggccga gacgggcaga	23340
tgacaaggtc aggagattga gaccatcctg gttacatgg tgaaaccccg tctctactaa	23400

caatacaaaa aaaaaaaaaat tagctgggcg tagtgggtggg cacctgtagt cccagctact	23460
cgggaggctg aggcaggaga atggcgtgaa cccaggaggc ggagcttgca gtgagcagag	23520
atcacgccag tgcctccag cctgggagac tgagttagac tccatctcca aaaaaataa	23580
aaaataaaaa aaaagacgtt tattcattga ttttaatggg attggagaag atgttatcaa	23640
ggggaggaat ctcaagtttg tgttcagttc ctgctgttct ctgagttctt tccttcttat	23700
tttgtaaaca tggttttgtt ttggttttta gtacacaggc tgccaaagca agcactatga	23760
ttttttgtag ctgtgaattc aattcattaa tatgagaatc ctagatgcta tctcaagaaa	23820
cattcatagg ttccatttta attcagctat gcttggataa aacatcagag aaattttatt	23880
gccatggaag gcctttccct taagtattag caataacaac aaaatagtaa ccataaaaaa	23940
actaccttta ttgagcactt actgtgtgct aaacacatgc attatttcct ttcacctca	24000
caccaacacc atgaaaaata tattcctctt acttccattg tacaggtagg gaaatggagg	24060
cttaaacag agcccatgga gctcctaagt gatggagcca ggatttgaac ccaggactgc	24120
tgactttagg ctcatgcttg taatcagggc actgtgcatt ccaggtagatt tatattggaa	24180
ggcagccttt cctgtgatta aaagtgcac tacgaagcat tgttctttcc ctcctttttt	24240
tttctgtagc ccgtttcacg gcctatcttg gagtcggcat ggcaaacctt atggctgagg	24300
tgagtttgc ttagtctcac ttttcattag cgtaattgac cagcttaca ctatatggga	24360
aatgctcctg aagtcactg ggctggcatc cagtggcagg atccatgacc atgagaagca	24420
ctgctctccc ttctcctgga gctccctggc ctttctttca gcatcacagc aaactttagt	24480
ccaaaccaca atcacccagt tgttacaagt atcagattgc ttggtttaaa aaaaaatgaa	24540
acgtaggttg tataacatat tatcaagttc agagtctaac tctaagtgat aagaagtaga	24600
ctttaggata tcttttactt aaacagaaaag ccagatattc cattgcaggt gatgcagggc	24660
cggtttctga tagcttagtc catgttgatg tggcatggc tgctaaggag tcaaggcagt	24720
atctagccct ttgggcagca gcatggagat ttatctggg agggtcctta aggagacaca	24780
gtgtctttct ggtggaaagc caaagtcca ttacacacat gcatgatgga gagtacatca	24840
gagcacatgg ggcccttcac atgtcaacaa agaagattca caggcatcag tcccaggacc	24900
caaatgggca agctgcacac cagagtcagc taggaagaca gaaaaatatg gaccttagg	24960
ccctgtcctt tggatattct gatagagtag gtcttgatg atgcttgaac atctgtgttt	25020
ttttttaact cccccagatg attctgatgt gcagtcagat tagggtaccc ctacactcca	25080
tcacacccca gggagggtcca tgcatcaggt cagagctaac caatgggtga tgctcagaat	25140
tgtgtgagtt tccatgagca gcacaaagag gacctacct caaggaactt agagtctatt	25200
tgggagacag aatggaaaga aacaaagcaa gtcaagtcta agatctagac caggcagaag	25260

tcaaggtcag agaggtcact gtgggctgga ctaatcagag aaggccttgt ggacatgaag	25320
actggtcagg ggccatttgc agtttgcaag tgcattctct gtcaaatgtt ctcttggcac	25380
atctggtgca ggaagtctga atatatgaga gggagagaaa gacatacaag atagagacat	25440
aagtggctgc cctaaagaat ggatgtcaac attccaacaa ctcaatgccc tgagattgta	25500
aattcagtct ccacgagcat gcacagaatc cagagcaatg cccccagtgg ttcattcccc	25560
tgggctgaat gcaagtagag ggggatgcct tgtgcagctc agctgtcaga tgggatctga	25620
aaggagcgtg tggctttctc ttcttcccca ggttggattg ccagcttgta cctggccctt	25680
ctgtttggcc acgctattgt tctcatcat gaccacaaaa aattccaaca tctacaagat	25740
gcccctcagt aaagtactt atcctgaaga aaaccgcac ttctacctgc aagccaagaa	25800
aagaatggtg gaaagccctt tgtgagaaca agccccattt gcagccatgg tcacgagtca	25860
tttctgcctg actgctccag ctaacttcca gggcttcagc aaactgctgt ttttcacgag	25920
tatcaacttt catactgacg cgtctgtaat ctgttcttat gctcatTTtg tattttcctt	25980
tcaactccag gaatatcctt gagcatatga gagtacacac caggatgatg gctctggtat	26040
ggaatttgaa accccaatgg ggccttggca ctaagactgg aatgtatata aagtcaaagt	26100
gctccaacag aaggaggaag tgaaaacaaa ctattagtat ttattgatat tcttggtgtt	26160
tagctggctc gatgatgta acagtattaa aaattaaacc ccataaacca actaagcctt	26220
atggaattca cagtcacaaa atcgaagtta atccagaatt ctgtgataag cagcttggct	26280
ttttttttaa atcaatgcaa gttacacatt atagccagaa tctgtatcac agaggtgcaa	26340
gctgacagca gagctcagtc cccacttctt gcaaacaatg gcctgcaccc tatcccttgt	26400
gtgtgtgaca ttctctcatg ggacaatgtt ggggtttttc agactgacag gactgcaaga	26460
gggagaaagg aattttgca atcaaaatta ttctgtattg caacttttct cagagattgc	26520
aaaggatttt ttaggtagag attatttttc cttatgaaaa atgatctgtt ttaaatgaga	26580
taaaatagga gaagtccctg gcttaacctg ttcttacata ttaaagaaaa gttacttact	26640
gtatttatga aatactcagc ttaggcattt ttactttaac ccctaaattg attttgtaaa	26700
tgccacaaat gcatagaatt gttaccaacc tccaaagggc tctttaaaat catatttttt	26760
attcatttga ggatgtctta taaagactga aggcaaaggc cagattgctt acgggtgtta	26820
tttttataag ttgttgaatt ccttaattta aaaaagctca ttattttttg cacactcaca	26880
atattctctc tcagaaatca atggcatttg aaccaccaa aagaaataaa gggctgagtg	26940
cgggtggtca cgcctgtaat cccagcactt tggggagccc aggcgggcag attgcttgaa	27000

cccaggagtt caagaccagc ctgggcagca tggtgaaacc ctgtatctac aaaaaataca	27060
aaaattagcc aggcattggtg gtgggtgcct gtagttccag ctacttggga ggctgaggtg	27120
ggaaaatgac ttgagcccag gaggaggagg ctgcagttag ctaagattgc accactgcac	27180
tccaacctgg gcgacaagag tgaaactgtg tctctcaaaa aaaaaaaaaa acaaacaaaa	27240
acaaaaacaa aacaaaaacaa aacaaaaacaa aacaggtgag gattccccctg ttttcctctc	27300
tttaatttta aagttaacag ttccgttaaag tctctgtaac caaacatact gaagacagca	27360
acagaagtca cgttcaggga ctggctcaca cctgtaatcc cagcactttg ggagatggag	27420
gtaaaaggat ctcttgagcc caggagttca agaccagctt gggcaacata gcaagactcc	27480
atctcttaaa aaataaaaat agtaacattt gccagggtgta gcagcacaca tctgcagcag	27540
ctactcagga ggctgaggtg gaaagatcgc ttgtgcacag aagttcgagg ctgcagttag	27600
ctatatgatac atgtcactgc actccagcct gtgtgaccga gcaagaccct atctcaaaaa	27660
aattaattaa ttaattaatt aattaattta aaaaggaagt catgttcatt tactttccac	27720
ttcagtggtg atcgtgtagt attttggagg ttggaaagt aaacgtagga atcctgaaga	27780
ttttttccac ttctagtttg cagtgtcag tgcacaatat acattttgct gaatgaataa	27840
acagaaatag ggaagtaaac ctacaaatat tttagggaga agctcacttc ttccttttct	27900
caggaaacca agcaagcaaa catatcgttc caattttaaa acccagtgac caaagccttt	27960
ggaactatga atttgcaact gtcatagggt tatggatatt gctgtggaga agctcaattt	28020
tcagtgtttg aactgaaccc tttcttgta gggaacgtgt gaaagaagaa ttgtggggaa	28080
aaaaaagcaa gcataaccaa agatcatcag cagtgaagaa tctaggctgt ggctgagaga	28140
accagaggcc tctaaaatgg acccgagtcg atcttcagaa cagggatcta ccatgcagga	28200
gcttctgtg ctacacaaa tctgtaaatg ggaacattgt acattgtcga atttaaatga	28260
tattaatttt ctcaagctat ttttgttact attttcctaa aattgaatat ttgcaggag	28320
cacttatact ttttcctaat gtctgtataa caaatttcta tgcaagtaca tgaataaatt	28380
atgtcacag ctca	28394
<210> 2	
<211> 28394	
<212> DNA	
<213> Homo Sapien	
<400> 2	
acacagagca gagtggggct ctgagtatat aactgttagg tgccctccctc cagcaccatc	60
tcctgagaag cactctccct tgtctggag gtgggcaaat ctttatcagc cactgccttc	120

tgctgccagg aagccagcta gagggtgta agtactcatc cttatttcta ttcatttcca	180
actattcatc atttggggct tgtcttcaca gttctaagtt ttgctctttt tcttaatgaa	240
gaaaatgttt tatatcaccc gaattgatca gaagtagcaa aatcagagtt ctggtagact	300
agaaagcaat ttaccaaagc cacaggcttc ttcttgaag ctcaaaggca tgcctttatt	360
cgtgatttct gaagcaaggt gcatgcagca cctgagctga tgtggaagag ggtttgcagg	420
gaggtgtcca cccaatgtgc tcaatgattc tgggttaate aacactatta ggagtttcag	480
gttgtgttct tgaataata atttgggctg tgttcttgaa ataagttcga ggcgagtgct	540
tacaagactc aaaagaaaa agtgggccac tgggaatggc cctttccagt gatggattta	600
tggactcctc tgtgtgtgct gtcattgctga agggaatgtt cttgtgcacc catcgggaga	660
acaagtcagt cacaactgaa gccacgaatt tggcagcttc cttgcagctg cactctctgg	720
agtctggaat caagacttct gggagtagtg ttttccaagg agggaagtgt tttaccagg	780
acacaggaat atctgacagc attttctttg tttccaatta cagctttaaa gaaaactggg	840
catctcctgc tacttaaaat caaaaactac ctaaaataaa gattatagta agtaccaat	900
aagtgtcaat gctgaaagtc tctttattat gctagaccat gagtgtttaa atgctttctt	960
ctatatccat atccaacact tcatattatt tttaaaagta atagctgaag catggaaaat	1020
tgaagacttc aggtctctcc aattgcacaa atttctaata catgctggca atagaatata	1080
ttttatttcg tgtaataaaa tagaggatat tagttgacct gaaatcttga tattgccttg	1140
tattaaaatg ctaagcactg cttcatttta ctagtgatct ggggtatgaa aagtgccttt	1200
tgacttctgc tggaaagctc ttcagggtgca gcttccagga tattcttggg atgttaactt	1260
cagcacacat aagccttgct gtagatgtgt cagctttgag gcacaggag acatttgctt	1320
gtcagagagt aactgcttct ggcaagggca tagggtgaaa ctggggatag cagagctctt	1380
tctttgtggt tgttcaacc ccccccaag attagttcaa agtgaccgtg aagatagtct	1440
gtgccaccg catcgctaag tcttagccct ctctgcatac tccagcacac agaaactgct	1500
gcttcacttg tttgttgact tgaaccgaac cttgggtggc attaatgtgc ctggcccaag	1560
actgaaaaat taagaaccac cagagctgac ctattccata agaccagctc tgcctgccac	1620
gtactgagtg aatctggatg atgcccactc tgatccttgg ttttctcttc tataaaatga	1680
aggtctgaac tacgtggctc ctaaaaatct acctagctct caaatctctc ttggttctag	1740
gaaaatattg atgttgagct caaggaagg gttctccaag gtgtgtgatt ttggtggtag	1800
aggaaaggcc ggtgccaggc aggggcagaa ggagacgctg tctacactga gaaaatgtga	1860

caacccctgc ttgtctcttt tttcattctt cattgtttct tatttctttg ttttagctt	1920
tataatacat gagagcccta ccactgggtt tcttaacat ttgttcttta tcaaataaaa	1980
atattcataa tgcaacatgc aggcacatca gtgtggtaca gaactagcca gctagtttac	2040
tataggtaaa tatacacaca tgcatgcaca cacacaattt ttacctgaga catgtcagaa	2100
gtgtttccta aaattgtgga tttttctgag tcattctggt aaagggtagg ttttcaggtt	2160
ttaggccaag ccagaagaag aaagtaaaaa cagaataaac aacaggggga gaaaaagaga	2220
aataccacac acacaactgg aacttctggt aaaagagtga tattcttgga tgcaatggaa	2280
gttttaaaaa ggaaaaagaa aatttataaa aagctgccac atttgtggaa ttcaactaaa	2340
aactgtttat tattaacaaa gtgatgttca aaatttaaga gttcttggcc tggcatgatg	2400
acttatgcct gtaatccag tgttttggga ggctaagggt ggaggatcac ttgaggccag	2460
gaattcaaaa ccagcctgga caatacaatg agactttgtc tctaaaaaa aataaaataa	2520
attaaaaata acacagctgg atgtggtggc acaggaaaaa aaaataccat ttaggagtct	2580
cttaaaggca gcttgtgaat gcttacaag cgtggctagt atcttattac agaaaacaga	2640
gcccacatca tgcatcttc ttctcacatt tcataaaca ggccaaggga aactgctgtg	2700
gggcaacctg ttgctttggt gttggtcccc aagatgcagc ctcacaatc tgccccaaa	2760
cgtgtcagaa catgaacccc ctctccccc tctggaagaa gcaacctcag atccaacagc	2820
agagacacgc agcagaacaa aatctgggca ttggtccctg tgtaggatgg ctcccgtaa	2880
ttttttttt aagcaaagta aatgaacatc aaatttccat agtcagctgc tgtctttctg	2940
cccactgaga gctctttggt gaaggcaaag tcctcttct tcattagcgg tctcccatgt	3000
ggggccacat ctccctcac caggaacca gtggcgcgcc tccagcccc ctgagcttgc	3060
cttttgcgtg gtcattagag ctagggcaca cgtcatctg attcacatat ttttgcctt	3120
tgtcatgtat tgagaaaaag taaggatgaa tggacggtct ttgattggcg gcgctggtga	3180
cggcgtcat ggtcctgttt ggaaggacct ttttggact aaagctggtg acgcagcgcg	3240
cagaggcatc gcccggctaa gcttggccct ggcagatggg tcgcaggaa aggtatgctt	3300
ccttcgtgca gcctctggct cggggaacct gggagcctgc tccaaactct ggtgtatctt	3360
ttccgggcag agcctgggaa gtgggggttg gctgtgagct aagccaaagg cacagggatc	3420
ttggtccaaa aagccccatg gcgctcacct tggtttagag gctagaccat tgagctgaga	3480
agttttgaca gccatggaaa agctgggat aagtcacctg gggttttacg tttacctgt	3540
gtctatttta ttagagtgcc ttttacttat gtcccttct tcttagttga aattaatggc	3600
ctgcttcaact ggggctaaga tgtttgaaca ttagcagaag gtctggctg catagccttg	3660
ccttgctctc ccagttagga tgtaaggact cttaaagttc cctaagaaat gcaaatattt	3720

tagcatggca aaattctagg ccaactacaa ctgtaagttt cgtattttct ctaagtgggt	3780
ctcatgcctg acttctggag caaggagtca ggtctcccag gggctctaga aggggtcagc	3840
tgttcagaat aaatggttcc tggggactct aaaatagcag caactgtctg cccaggtcat	3900
gagaagaccc ctctctgcag gacatcctag ccctacaacc catcccaatt atgttgaaat	3960
tagattcaca aatggcaata agtcttctat atgttgggct gtcgatttgg agaaaactag	4020
tttaatcttt acttaacttt ggggtggctca acaggagact cgggccgctc aggcctctca	4080
tcacgtctgg ccagttctat tatcagggtt cgaatctgta tctccaaaat ctctgagggtg	4140
atgggatatt tcaagccctc taaaataaat aaatatatgc tgggaatttt gagaacatga	4200
atttgtttat tctgaaatgg tccatgttcc tgccttggga gttgatggaa aatgccactt	4260
gagtgttttc atttgatgct gccaccttag ggttttatag attcagttcc agaaactcaa	4320
ggcattttatc tcttgggctt gcttgtcctt gcctgagctg aagcctgatg cctcccataa	4380
gttggatagg ctttgaanaa gggtcactac agcagaggca tgggctttatc aagcaatatg	4440
ttcagctatg aaatttgaag agggagataa tctgaaaata aatgacagcc accacttaga	4500
ttatgaaata gaagtacttt ttcataagtg ctttaattatt catagcgttt ttatcttta	4560
actatggagc caactcagct ccatatggac ttaatttttg ttcctgacct ccaagattca	4620
ttgcaagtca cacagatgtt ggtatctaac attgttttac cgagataaaa tgaccttgggt	4680
ctggaatgca ttgtataaaa agctgctttt ttgtgtaaag attaatagtt tggcattgtt	4740
taaaaagcag aatgggttagt tgggcagtga ggtaatacaa ttgaaatgta attgctacca	4800
ataaatcagt taccatatt gatttcttta ctgggattaa tagaagccaa agctagagtt	4860
caactttttt taataggtat aacttagtat ctgttcattg ctatttgta gctatggtaa	4920
atggaacaat gatggggcca gaaatatcca tgaggacct ttgatcacag cctggcaaca	4980
cagagaagac aggcctggtt ctctatgtgg gctttcagtg tttctttggt agtgtcttat	5040
gtggctgtgg cttaacatt ccacaattat gccttcagg gtcgatgat tttggcgttt	5100
ccctgcttcc caattgacct ggctgtgctg ttggctgttc ttgcacactc aagggtggtt	5160
tgccattggc ttctccctc agcctgcctc tgggattatg ccactgctat tcttttttat	5220
ctaccatcag cacaatgaaa tcatcatttt tgccttcaag gtaccaaatt ctggtgatat	5280
tgggtgcttc ttgcagctac ttatcatgag aagtgaatgg tctcatagt aacacagtca	5340
tggttatagt gtacatacgt tccagagaca tgtttcctat aattatgccc tgcacatttt	5400
tctatcatac aatccttaga ttacagctct ttggttttca acagctttgt ccaattccat	5460
ctttccagct ttctctacct tgatgaaata tcttcttgc ctggttttac atatttaaat	5520
aacaaattcc aaaagtaaag agtatctgag gcagtcacat gacataagga caaattcaag	5580

ccatcttggga cttgcagagg gtggggagac cgtgtcaaca cacacaattt taaaaatttc	5640
ttccctttca atctttttaa aacaaaactt tttataaaat aaaaatgtaa tttaaaaagg	5700
ctacctgtct tggcaagtag ctgatcagcc tgcattgggtg agcaggccat tccataacct	5760
ggtttcttgc tccttaattg acagcatgga gctaacgtac ttaatttcag ctctttctac	5820
gtgatttgac tcattctgtt aacattaact gtttttcagt cttctcaact agactgaact	5880
ccttaagtgc aagaaataca cgcttagtaa atgtttgttg gaccagacac tgcaccttat	5940
gaaattaaag accagaacat tctcatggta gcattacaga cactgatggc aaaggtactg	6000
tgggatttgg gtttggctaa taagctctgt ggtggtgttt cagaaggaaa atggtgctct	6060
cttagttcta tggaacatag tggccagat cttctactgt aaccaggccc aaagctggct	6120
aatctggagg gctctgcctt agggatactt ataagctctg tccttcctc aaggagccag	6180
aggaagagat agccatggag gacagcccca ctatggttag agtggacagc cccactatgg	6240
ttaggggtga aaaccaggtt tcgccatgtc aagggaagaag gtgcttcccc aaagctcttg	6300
gctatgtcac cggtagcatg aaagaacttg ccaaccagct taaaggtatt taccctttca	6360
cattttggag agacaggaga agtagctttg ggggaaatgg tttcctggta cttctactta	6420
taccttagt tatattctcc aactttttat agatctcttt actcaccatt tttctacttt	6480
tatcttttaa cctgcaaacc tctccatttt tttttcttat ggagacagta gccagggccc	6540
agctcatatt agaaggcacc tggcttcac ctagtatttc agtacttaaa acttaaat	6600
attccttttg cttcagaatt tgtacctata agcatgaaaa taagtgcatt agatgctttc	6660
aggagcttag attctaggag gggcagtggt ggttgagcat acagtagata gaggctttca	6720
gggatctggg tgccactaat gcaacaatgg gttgagagag aaatattaaa gaaatatcaa	6780
aaatgtttca cttccaggag gttttgctga ttttgctcag ggtgggcctg tggttgaaga	6840
gtatcacttg gcagcttct tagctctgct ttacctcacc cttccagac aaaccctgg	6900
tgtccagtt cattgactgg attctccggg gcatacceca agtgggtgtc gtcaacaacc	6960
ccatcagtgg aatcctgatt ctggtaggac ttcttggtca gaacctctgg tgggctctca	7020
ctggctggct gggaacagtg gtctccactc tgatggccct cttgtcagc caggacaggt	7080
aggtgtacce tttcaagcct tctcagctcc cttctgagac acaggggctg accagttact	7140
gtgggcaaca gtgataaac cacatcttc ccaggataaa caacatttag tccacagaac	7200
tgtttatatt tgtttttagt cagaggctcag ggaatcagtt acagtctctt gctcttgata	7260
tctgaataaa tggtctggtc aaatgatgcc agattcttgt ggcattacgt gctaaccaga	7320

actaagctac aagtatttcc ctggagaggt tctgaaggga tcttctttaa tgattgataa	7380
aattatttgt cgtcagcatt ctatttggga aaaagtgc atgaattcag aaaaagtttt	7440
agtggcttaa taacccccgt tatacttctgt tgctatgatg agtttaggaa actcattctt	7500
catagacagt gcaaaggcca gctcagctcc tggagaaaag aataacctg aattccaatt	7560
gagtggattc tgacttaaga agccttagtg agtcttctga tatattgatt agattaaaaa	7620
tagcacacac ttataaaatt gatctgtcat tgaagaagt atgagctgac tctcaccagg	7680
gcagtagata gctccccact agccagttcc tttagggagg gaaccagtat tccaggtgtc	7740
tgagatcaac gcataatccc aatccccagt gtggtcatta cacaactaag ctcttgtaac	7800
actggctgca aattgcctaa agaggtccgt ggggagagag ttagcaaatg ctccactttt	7860
ctatcaattt caaggagtct gatttctcc ctgtagaagg ggattttata gcttaggtta	7920
aactctattc caatgcacg caagaaaagg tctctcagt ttggggatgg agtctataat	7980
tgtgccatac tgaatattcc ttatgattt tgctctgatg aaacatgatc aactcatttt	8040
ttgtcagata ttatttagaa gacaagtc tttatgtgt tagtttcaa tgttttactt	8100
tccttggctc gaaaagactg cattaaaatg gaaattctct gttttaagta aatatatgtc	8160
ttcctgtggc tttaactatg gcattccaca attttagat gtgcccatta attttccact	8220
gatcaaacac aagcattaac atctccaagt cagtgttga gaggacaagt ctgcatggct	8280
ctctactgtc atgtgtagtc ccagtccttg agttgtacct ttgcaaatg tatcacctcc	8340
catttgccct caaggattat ttaaggga caaagaactt ttgaataggg aacccacat	8400
ttaatgttca tctggattaa tgtacgtgac atcatcttgc ctgttgcaat ggtgcctcct	8460
ggcccagtta gaaacaagcc aagaagcagc tgtcacacta tcccttacca gccctgcag	8520
tgtggctcac tggctatagc acctctgct cgagcccagc attaggcctc acctactcac	8580
ttcaccatct ttactcccc atccccctac agacatcatc cttgagtgac aggcccttgg	8640
gaagtggatc ctgtgccttt cacggtgcca gacgttgcca actctcagag ctgtgggaat	8700
cctgccttgt cagggtcaatc aatctagggt cccatcaatg gtggattata taaagaatat	8760
gtggtgcata tacaacacga actactacat agccataaaa aggattgaaa tcaagtcctt	8820
tgcagcagca tggatgtatc tggagaccaa taccctaagt gaattaatgt agtaacagaa	8880
aatcaaatc cacacgtttt cacttacaat taggagctaa aactgggta aacacggaca	8940
tggaaatagt agacaactgg gactccaaaa gaggagagga agggaaacaa gtgttgaaaa	9000
cctacctatc aggtactttg ttactattt gggtgacgag ttcaatagaa gcccaaacct	9060
cagtgcagc catgcaatc atctatgtaa caaacctgca catgtacccc ctcaatctaa	9120
agaaggagaa gaagacgggg aagaatagag attgaatact aagcaaaaag taacctcaga	9180

aagaactggg tgctcaacat gcacataatt aaatgggata cttctccaag taagagaaaa	9240
gcaattgttc ttctttgcaa taactttgaa atgtgcgttt ggagacaaca aaatagaagc	9300
atcaggacac aaaaatgtat actaacctgg aagattaatg ttgataagat caaagacact	9360
gtgaaagtga atttacattt caggaatctt atatctctca ccaagaaatc aaacttaagc	9420
aacagtttca tatgctaaaa gcgctcttca agtcagaggc tcttgattta aaagaataac	9480
tttccaaagg aaaggctaaa agaaaacaga gcagattgcc ttactaaact cccctttcct	9540
ctcagccact gtagacctgt ctttagccgt gacacctgta gagggagtca ttctctatca	9600
ggggtcccca acccctgcac tggagacagg tacctgtctg tggcctgttg ggaactgggc	9660
cgcacagcag gaggtgagcg gtggcgaggt gagcatttcc acctgagctc cgcctcctgt	9720
cagatcagca gaagcattag cttctcataa gagtgcgaac ccattatga actgggcatg	9780
tgagggatct aggttgcttg ctcttatga gaatctaata cctgataatc tgagggtgaa	9840
cagtttcatc ccgaaatcat cccccattcc ccatccatgg aaaattgtct tccatgaaac	9900
ctgtccctgg ggccaaaaag gctggggacc actgatctaa atgcacattt atatttttat	9960
ctatgtatat ttcaattcat gtctttatta gttttgtac gatgcttacg tagactttga	10020
aatacatttc caaatataat ctcatTTTTT aatatgaata tgatctggaa gttactagt	10080
ttattttatg gcaagtgcaa ccaaagctca ccaggaat gtccgtgctg tgtctcttgc	10140
cccacaggtc attaatagca tctgggctct atggctacaa tgccaccctg gtgggagtac	10200
tcatggctgt ctttccggac aaggagact atttctggtg gctgttactc cctgtatgtg	10260
ctatgtccat gacttggtaa gttacaattg gttttcaaaa tgcctttttg aaaaaaaaa	10320
catggcagaa ggagggaatg ggagttgtta tatggcagag ttccagtttt gcaagatgaa	10380
atatgttctc tgaatgtata gtggtgatgg ttgtacaaca atgtgattgt ccttaatgtc	10440
attgagctgc acacttaaaa atggttagcc ggtgctggtg gttcttgttt gtagtccaaa	10500
ctattcagaa ggctgagggg gaaggatcac ttgagcccag gagttagggg ctgcagtgag	10560
ctatgattgc gtcaccgcac tccagtctc cgaacctct tgcctgggct aagtgaggag	10620
gaggaggagg aggagaagga tggaaaggag gaggagtgc aggaggagca ggagggcaag	10680
gagaaggagg aagaggagca ggaggaggac aaacagttaa aatggtaaat ttaaaattgg	10740
attccagtag attctgtcta ttggaacag aaacaacct tttaaaagat gtatatctcc	10800
ttacaaccag ttatttggcc ttttgtctga tctggctaca catccactaa tacctctcaa	10860
ccagaggtgg ctgcacattg acatttccat gggaaggga aacagtgtg caatgaagat	10920
acgagtgcag gtgtcttttt ggtagaaaca cactgatgca cgtggccccc acatacatt	10980
gactctccc tccaagact ctactgtcat tggctctcgg tagcgcctgg gctttgggag	11040

tttctaaagc ttcccagatg actctaaagt atagccaaag ttgagacca cttcttccat	11100
cattgcctct caaacttgag caatatgaga atcacctgca gggtttgta caccacaggc	11160
atctgtctcc cgccccagg gtttctgatg cagtctatct ggggtggggc cggagaattt	11220
gcgtttctaa cgcatccca catgatgctg ggagaaccac tgtgcctacg tgaattcccc	11280
cttaccaccc tgccccagg gtctccctta gaaaaaattt ttttctgaa ttcctttttt	11340
ttcaaaccce aatccttcaa actagttttt atgttgacaa tgtcttacat cctttttctg	11400
gaaacaaaga tttccttctt tctatattgt agttaaatat aaaatactaa tatgcacata	11460
aataagcaca gcctgctgtg ggcagtgtct gcagaaggga tgccaccct tactgtacct	11520
acgggtgtgt ggacgaggac ctacctgtag agctaaactc ttcaggaagt aatttgggcc	11580
ctgtctgaa gaatagggtc gtgggaagga ggcctagcct gtaagtgtc accacgtcc	11640
cttcacaat ccaggaaaaa gggagtcttg gtctttaagt gatggctctt tgattgggcc	11700
aacaagttag agcctatgag ggacctcggg accatgcagc ccagccccc agtttatggg	11760
ctctgaggct aaggagatgc gccttgccca ggtcatgcaa tttatcaaca gctcaaggac	11820
acacactctg ccccaacca tgtgatatca ttttctcca gctcacacta cctgcatcct	11880
tgaacgattg tttctctttt caaaaaatag gtatattaaa gaaataatat ctgccaatc	11940
agaatcaggg ttgcctctag tggggaggga gggacataag agcaagtgga gggacaaagg	12000
ggactttaac tatgtagata atattttatt ttgtatgtca taagtacttc aaaaatattt	12060
ttaaaatctc aatatatagc tcaactgag caaccccaga gtagaatttt tcaaagcca	12120
aataagctga gatttgattt ttactttat gtaatattta ctgcctctat aataggattt	12180
atcccaagtt ttttttctgt ggcaaatgtg ccaacacaac acgtaagggg cctgttgga	12240
ggtgaacaaa agccctcca gagtatagcg attccgtgtg tcagcctgct ttgtcacatg	12300
cacattcttt tgcctgttc ttttttagc ccaattttct caagtgcatt gaattccatg	12360
ctcagcaaat gggacctccc cgtcttcacc ctccctttca acatggcgtt gtcaatgtac	12420
ctttcagcca caggacatta caatccattc tttccagcca aactggcat acctataact	12480
acagctccaa atatctctg gtctgacctc agtgcctgg aggttaagaga cactggcttc	12540
tcacattcgc cctggctctg caagatagc aatggcctcc tggtaactg tccacgggtg	12600
tcagagtctc ctatgctc aggactatgg tggcctttct gccttcactc tgccatttaa	12660
agcatttggt ctactccaga gcattagggt ctaagggtt ttttaaaatt actatttagt	12720
caagctgatt tttctgcctt tttccctaaa catctacagt gctaaccce gagtacagtt	12780

ccactgggag tcactctatc gtaagcttgg gggtaggggt gatgggagcc agcccttaag	12840
gcatgtggcc tccagcctgg ttttaaatct tccatagtct actccctcca atcaaaaaac	12900
tggatgctta ctcttagagc ttctgacaga acctctctat tctgcttttc cttatggcat	12960
agctcataga acatctacaa taatttaggg ttcccaagct ttggtaggca tcagaatcac	13020
ctggggagct ttaataaccc aaacaggctt catctcagac cctctaaatc acaatctcta	13080
agggtggggc ctggaacctg ttttaacaaa ctcccaaat tgtgatgcgg gccagagttt	13140
gagaaccact gtatcaaggg gtgaatccia tgtatctctt taaagatggc tataaagaga	13200
ttctgtattt tttaaaacct ggtaacca aatcaaattc cagctcttcc tgttggtgtg	13260
taataaatat gttaaaggtt tctggattat caagaacaag agaacacctg aaattagaag	13320
aaaaccaaag aaaccttacc tttttaatgt gctctccac tgtcaggta tgaaacgccc	13380
ttttgtcttc ttgtttgagt gatcaaaaca cagcaggagc tcaagtcacc ttctccctag	13440
cttcttgcca gaaaactaaa gggagcacct ggaaataatt cagaaggaaa aaatcaaaga	13500
ttcattagaa ctacctatga aaaataacag tataaaatag cattaatcga tctagaactg	13560
cactaacaca ggagcctcta gccccatgtg gctatataaa tttagatgta gattagttaa	13620
aaattgagtt cctcaacctc tctagccaca tctcagggtc ttgatagcca cacgtggcta	13680
ggaccactg tattagacag cacagataca gactattcca tcatctcgga aagttatcct	13740
gcacagtgt gatctggggc aggggaagcc ttgtccttct cactctgaat gaacagccca	13800
tcctcagcac caacccaac cctatggcta cctgagagag agttctgcag ccaagtccaa	13860
aaacaacaa acaaacaaaa aaagcatatg ccatcttgc caagttccct ggtctagaaa	13920
tagcaaatg tctagacatg aagactcagc atgggctgga agaatttaga gtccatctta	13980
gggtagatgc aaactcacac tatggtctgg tgcccttagc caatgttaga ctcagcctaa	14040
tataagaggg gagaagacac ttcccttgt gccaaagctg gggtccctc tggtagagtc	14100
actgcctcca gaaggtcttt ggtacataca cgacctagca atggtagaga gggcaagatg	14160
ggaactgagg aaaacatctt tcaagtaaatg gccttgctca aaaggacat gctatggcta	14220
attatgccta tctagccct accagaagtt cagctgtaaa gaatgatcac ttgttaggtt	14280
cagttaaacc ttgttactc ctgagaactg caattctgtg aacagaataa ctaaattcag	14340
gcctcagcca gaaagtagaa ttatgacatt tccatgtatt tttgtgttt gagacctgct	14400
tgacagttgt tcataactag aataagctaa aaatatctt gttaaataa atacatgttc	14460
cacttaatga cagaaaagta aattcacaaa ctgctaaaa attacttcta aattgtggac	14520
aagataacct ggctttgggt ctctggcttt agtgtaagca tccaaattgc atagtataa	14580
taatctctat tgaacatagg gatgcattga tagattaaat caccctcaac actgatggac	14640

atttgaaagc aaaagaagtgc tcagctgtgg tccttgccat cccagtagg aggcaaggca	14700
gatactcata gccaggagca gtgagtggca ccaagctggg agcttaacag tgaccaaggc	14760
caagtgtcag tgcaagcagg agagcacagg gggagctttg agaaggcatg tgttgcatgc	14820
accagggaag ggctgggtga tctctgggga taaagctgaa ggatgactgg gatttttctg	14880
taatcaaaga gagagaattt taaatggtat taacactgtt ctgaaagag gtaaggtatg	14940
tccaatctaa aattacattg taggagtttg tgggtgtcct gtgggtttct gttcagttgt	15000
tttggtagcc tcatttttct taaatttctt ttgcagttgt tgaaatctat accagtggga	15060
gttggtcaga tctatggctg tgataatcca tggacagggg gcattttcct gggagccatc	15120
ctactctctt cccactcat gtgcctgcat gctgccatag gatcattgct gggcatagca	15180
gcgggtgagc acaagagccc ttaccaaata ttgagcacct cctccatccc atgcattgcc	15240
tcaggcatct tctgtgctcc agatcttctt tgagatcttg gcttcctagg gaccaatggg	15300
agttcccggtg atgcttctctg ctaactttca atcccacct cagtttctt ccagaacatc	15360
ctgcctttag tctgagttc tgacctctc tctcttaaca ggactcagtc tttcagcccc	15420
atttgaggac atctactttg gactctgggg tttcaacagc tctctggcct gcattgcaat	15480
gggaggaatg ttcatggcgc tcacctggca aaccacctc ctggctcttg gctgtggtga	15540
gtctcccacg cccctggggg agggctgctc atgactacag gatctcaatc aaggataagc	15600
agtaaaaacg gactgcatga aaaatcaggg ccagggttct ggcttgagcc cacttgctgt	15660
ctaagtgtgt gaacaggaca agtgacgtcc cctctctgag agcattaaaa tcacctctgc	15720
ctacctctct gatgattgtg aaggcaggag cctattgagt catattaata tcctaaaaca	15780
tggatgtttg ggaggataga aaaagaaaaa tccagttat tcttcagctt tatccccaga	15840
gatacaccag ccttccctg gtgcatgcca cacatgcctt ctcaaagctc cccctgtgct	15900
cacgggctct ccagcttgca ctgacacttg gccttgcca ccaataagct cctagaatgg	15960
tggcactcac tgctcctggc tgtgaggatc tgccatgcct cccactgggg atggcaagga	16020
cctcagctga cactcctttt gctttcaact gacttgctt gcgttcttca aactagtgt	16080
ttgacccaac aaactaaacg ggaataactc cagctaaata cagagcaatg tcccctggtg	16140
aatcagggtt gattacattt acccctttga gtgagcatca cagtaacca gccattctaa	16200
aacttcagaa tgcatcagaa tcacctgaaa gacttggtta aacacaaatc gctgggcccc	16260
ctcctcagtc tgattcagcg tcagagataa ggggaagaat atttctttt ttatttttct	16320
aaaaaacagt ctattctga gccaaagatc cgccactgca cttcagcctg ggcaacagag	16380
caagacttca tctcaaaaaa aaaaaaaaaa gagaaaagaa aaaaaagaa aaagggtctc	16440
attctgttgc ccaggctgga gtgcggtggt gtgaacacag ctactgcag cctcaacctc	16500

ctgggctcaa gcaatcctgc agcctcagcc tcccaagtaa agtagctagg accacaggcg	16560
tgccaccatg cctggttaat tttttatatt ttatagagat ggggtctccc tatgttaccc	16620
aggctgatct tgaattcccg ggctcaagca atcctcccgc ctcacacctc caaagtgtctg	16680
ggattacagg cataagccac catgccggca gaatttccac ttctaacaag ttctcagggg	16740
gtgctgatgc tgttgctctc aggatcacat ttcaagaact gctgtattaa tcctttctga	16800
ctcccagtgt tctagccaga ctcagcctgt cagagcgaga aggcatcctg agacctctac	16860
tccatccttc ttactttact gttggggctc ttaggccaga gaggctaagg gatgtgccgc	16920
agggaatctg gacagcaatg ggtaaatcca cccccgaac ccacattac catccacctc	16980
cagagttatc ccaccgcaat cctctgcttc ccttttatag cattcaggcc ctcacggcaa	17040
cctcttaggt gaaaacagac tgcattgat ttggatctga aaagctaata gatcccaggt	17100
ggatttttag tggaggtcca ttaccata gccctctggca tgcctaattc aatcaaagta	17160
taagcattta agataatatt ctagagtga gagaatgaga ttgtcttggg aacaaaaagg	17220
aggagggata gtgtaattgt gagaaattat gtctaattca gtggaaatat atgtctagaa	17280
tcagtttate accagattaa tcaagccaag gtatctaaac agttatgaaa acagtgggcc	17340
atgtatcagg cgggtttaga atagatttct gactggcag aaaatgggat ggtaccaacg	17400
gtttctaaag acccattcca ttttgattcg atgtatagc aagggtaca taactcaggt	17460
tgctgtgatg tagccatgta gatgtcattt tgcataattc ttactatta ctcagctatt	17520
tcacctagct gttctgtga aatgttgaac tccttctcca tattcgttca caaggataaa	17580
ggagaggatt acagacaggt gctgtagcca cctgagttca gctgggttgg aatgtttatc	17640
ctacaacctt tcagctttat tctgagattg gttaggggtt tccacctgag ttcagctggg	17700
ttagaatgtt tatcctacaa cctttcagct ttattctgag attggttagg ggtttcaaac	17760
ctttatttgg gatgcatacc tttatttttc tggaggaagt agccacaaat atgtattaaa	17820
cacacatgat acaaaagaca gtaccaggaa gagcaagggg tttagaagct ttaggtccca	17880
tgcatctct gcacagagtg ttacaataga gggcagaagc caggcaaggg agtgagccca	17940
agaggacat gcaatctttg tgggagaaga agaagtccat agtacaggat tctccagggg	18000
gccatttcca ctcagaatta tcacaaagta cctccaggaa gaagggggct tttcataaaa	18060
tgctagaaaa taagaggagg aattctgttt ggtggaaagt gtggtgcagg ccagcatggg	18120
gacagcctga gcatgtcctt caagatcaag gagaaggcat tttgagcaca ggagatggcg	18180
acgaggtttt tgtttttctg ggttttttgt tgttttttgt tttttggttt tttttttttt	18240

tttttgaca gagtcttgct ctgttgccag gctggaatgc agtggcacag tggcacgac	18300
ttggctcact gcaacctccg actccctggt tcaagcggtt ctctgcctc agcctccaa	18360
gtagctgggc ttacaggcac gcaccatcac gcctagctaa tttttgtatt tttagtagag	18420
acggggtttc accatgttgg ccaggatggt ctcaatcttc tgacctcatg atctgtccac	18480
cccggcctcc caaagtgtcg ggattacaag tatgagccac cgcacctggc ggggtctgag	18540
ttttttgttt tatgtttgtg ttgtttgttg agatggactc ttgctctgta gctcaggctg	18600
gcatgcagtg gcacgatctc agctcactgc aacctctgcc tcccgggtcc cggttcaagc	18660
aattcttctg cctcagcctc cccagtagct gggattacag gcatgtgcca ccatgcccag	18720
ctaatttttt ttgtatttt tagtagagat ggggtttcac catgttgcc aggctggtct	18780
tgaactctg acctcgtgat ccacctgcct tggcctcca aagtgtggg attacaggcg	18840
tgagccacag tgcccagcta gtgatgaggt ttgacagac catggagaag aatgaagtcg	18900
aagctcttga catgtttgtt ccccaaagtg ggaatctttg atattttctc aattatagaa	18960
gcagcacaga tttattglat aaaacaaaac aaaaatgtaa tctgtataga aatgtatgaa	19020
acagaaagtg gaaatactcc atcttactcc ctagagaggg cttttttgcc cccttcttat	19080
aaggatcctt gtgattacat tgggtccatt caatagtcta ggaaattctc tccatctcaa	19140
ggtctttaac ttaatcacag ctgctgctaa ttcccttttg ccatgtgagg tcacatatc	19200
tcaagttctg aggtttaaga tgtagacgtc ttggagacc attattcttc ctaccacact	19260
caccttctt tggatagatt ttttttttt ttaactggtg tagcataatg gttgaggcag	19320
tcaactgagc taaagagctc agactctggt gccagacagc ctggattcaa ttccagcagg	19380
tctgctactt actagcgtat ttgcttatga atgtaagcaa attacttaac ctttctatgc	19440
ctcagtttcc ccatcttaga aaatggaagt taccatattt aattcataca gttgttctga	19500
tgattaagtt agttaatgca tgtctgaaac tcatagaaca aatagtgtct agcactcgt	19560
cagcactatt taaaagtctg gaaaaacagt ttttctggtg gatttgcata acttattaag	19620
aatcaagctt gtttatcttc tctctcaat tgcttaagtt tatcaacatc tgtatcttct	19680
ccccaaatat gactgatacc caagcctgcc ttacttctc ctgagaaggc ccaccctga	19740
tgactactaa aaccattgat actgtataga atttttatct tggatttctc gtaagtataa	19800
gtttttgttt tgggtacttg cttatttagg caactgtaaa ctttattaac ttgcttattc	19860
actctgactt agttcatatt aaccttctgt acttttttt tttttagaca gactctcact	19920
ctgttcccca ggctggagtg cagtggcaca atctcagctc actgcagcct ccacctcctg	19980
ggttcaagcg attcctatgc ctgagactcc caagtagctg ggattacaga catgcaccac	20040
catgcccagc taattttttg tactttttgt agagacaggg ttttgccatg ttggccaggc	20100

tggtctcaaa ctctgacct caagtgatcc acctgcctcg gcctcccaaa gtgctaggat	20160
tactggtgga ttactttttt aaagaggggt tgcaaagaga gttttgtttt cttcaaagag	20220
ggtttgcaaa gagaccttgt atgctggaga atatcttcat tttaacattca tttaaatttt	20280
agtttagcta gctaccaaac tcaagattta acattttttt ctcaatattt tgaaagtgt	20340
cctcaaagac tactccattg tcttcttata cccaaaattg ctattaagat gtctgaaaag	20400
aaactaatte ttgttaaaat tgattttatt tttctctctg gactctctga attttctctt	20460
tgcatatgag atatatatat ggttttattt cactattatc tgtctagatg taactttttt	20520
ttctatgcta gtaggtactc aagtcctctc aacatgagcc ctcatatctt cctttaattc	20580
tggaacatc atcagttttt actttgtcaa atcttttcaa tttttccct ctccttctgt	20640
gatttctagt atttgagtac aatactttat gctaagtttt tcataactct tgactttttc	20700
ttaatatattt ccactatctt tttctgagg cccttcagtt cagctgattg gcccgatcat	20760
tctttggctc tgtccattgc accgatcaca ttatctgttg agttctccat ttctggttca	20820
ttaattaaat ttactggct gggtgcagtg actcacacct gtaaaccag cactttggga	20880
ggccaaggcg ggtgatcac aaggtaaga gattgagacc atctggcta acacggtgaa	20940
accccatctc tactaaaaat acaaaaatta gctgggcgtg gtagcacgcg cctatagtcc	21000
cagctactca ggaggctgag gcaggagaac cacttgaact cggaaggcag agctgcagtg	21060
agctgagatc atgccactgc actccagcct gggtgacaga gagagactct gtctcaaaga	21120
aaaaattatc gactgtaggt tgttcagttt gttgtccttc ttttatggta tttgtctcc	21180
tgggatgtcc cttttccttg tcttgggagc tcacgtttcc ctcgggatac cagctgtttg	21240
ggtgagtcct tgggcagaga tggaagccca ggttgagct gcatttttcc tgggtgatct	21300
aaggaaaaag gggtccctg ccacaggggt tagaacctcc attgtcaag gctgtggaga	21360
tgggtgactgt gtagacattt tatatgataa gtgccctttt gctgggggaa gttcagattg	21420
cttctagttt gaaatcatta caaagagtcc tgaaatgaat atttttggta caaatgtcct	21480
tgtgtacttt gtacaagcat ttctgtaaga aagaagattc accttctttt caagaagcta	21540
aattgatggg ttaaagggaa tgccaatttt gatttcagtg gatgccaact tcatttccaa	21600
aagagccata ccagtttcca ctgctgccag cagtgtgtga gagtcccccac tgggccccca	21660
caaggtacaa tcagactttt aaatctctgt gcatggattt ttgagacaga tctccagccc	21720
cccttgaaa gcaaatctca catgtaaaat gccacagcaa gtttcagctt gtccacatca	21780
ccctgatact gccaaacaaa agaccaacc tcttagccaa cataaataag tgacagacat	21840
ttattacaga gctgtttttt tatcagtcct cagtggcttt atcaggaagt ggactcagga	21900
aactctgaca gaacctggca ctgtgtctt tctggcctct aagccagagc aactgcgtgg	21960

ccagagaaca tcicaatgtt gttgttttac cagtggagag tgtaaacata ttgtgtatct 22020

cttccaatg gttgggttat cgcagtggga ctcacctgtg gcagtcatt ggaagggaca 22080

ctatccagga ggagctgaaa tccagtttcc cttcagtag tcaagggcct tttcttcct 22140

cagctaccaa gaatgctgtc agggtcattg cctacaaact gatgatgctg tgcagaattg 22200

cgcctctact gtaaggcttt cccggtccta ctggcgagt ctttaattgac atacctacca 22260

ttaaataatc tatcacttgt actatggaga gaaaagcaac tttgaattgg agatcacttc 22320

acagcagcat aacagtatga gacgtaaagc tgccaaaagt gagccttaga agtgtaatgg 22380

atattttaaa aagagagaaa gcaacaaggc ctcatgtgct caggggtggt gttgtggtag 22440

agggggcact caagagatca gggacagagg gccccagtgc ttggcagagg gccaatgaat 22500

agttgttaaa ttaattgatt aaatttcaac aatgaatgaa attggtgtaa ccaaggagag 22560

aaaccttctt aagccaagcc atgagcacc tctgtctcag agcagtagct cagtcccatg 22620

gtgaaagaga tgcatttaca gctgtgttta tggaaataga agctctcatt tgagattctt 22680

cacctcccag taaggcagat cttcaagggt cttttttaca gatgatgaaa ctagattcca 22740

agacagtgat ttgttataga acaataaaaa tggcagagct gggatttgaa accagtactg 22800

tttccaaaga ccagcctttc ccactagtgt gagacaattc atacgtgaaa gaatttgata 22860

tactattgaa taagaaacac caggataaaa agacaaaata ttggtaaaag gacagaagtc 22920

tatggtaaag taaatgagga tcacagagcc tctcccacca tgtctgccac atccccacac 22980

accaagatag ctgacgtacc agacatgaag acgagatggt gagtgtgtct cacggtgagc 23040

tccggtggcc caagtggctg tgtggccatt atatgaaggc cattcttcag gctgtcccca 23100

tgaacctga gggcttccct gagcctctgt gagccttctc ttcaacaaa actgaggaat 23160

agataattag ctggttgaga tctttgcttt tgtgtttta cactgaaagt cacccatata 23220

ctcgaattac tgattctaca attttttggc cactcaaagc aaataaaaac ataagacgtt 23280

ggctgggcgc ggtggctcat gcctgtaatc ccagcacttt gggaggccga gacgggcaga 23340

tgacaaggtc aggagattga gaccatctg gttacatgg tgaaacccg tctctactaa 23400

caatacaaaa aaaaaaaaaat tagctgggcg tagtgggtggg cacctgtagt ccagctact 23460

cgggaggctg aggcaggaga atggcgtgaa ccaggaggc ggagcttgca gtgagcagag 23520

atcacgccag tgcctccag cctgggcgac tgagttagac tccatctcca aaaaaataa 23580

aaaaataaaa aaaagacgtt tattcattga ttttaatggt attggagaag atgttatcaa 23640

ggggaggaat ctcaagtttg tgttcagttc ctgctgttct ctgagttctt tccttcttat 23700

tttgtaaaca tggttttgtt ttggttttta gtacacaggc tgccaaagca agcactatga	23760
tttttttag ctgtgaattc aattcattaa tatgagaatc ctatagtgcta tctcaagaaa	23820
cattcatagg ttccatttta attcagctat gcttggataa aacatcagag aaattttatt	23880
gccatggaag gcctttccct taagtattag caataacaac aaaatagtaa ccataaaaaa	23940
actaccttta ttgagcactt actgtgtgct aaacacatgc attatttcct ttcacctca	24000
caccaacacc atgaaaaata tattcctctt acttccattg tacagggtgag gaaatggagg	24060
cttaaaacag agcccatgga gctcctaagt gatggagcca ggatttgaac ccaggactgc	24120
tgactttagg ctcatgcttg taatcagggc actgtgcatt ccagggtgatt tatattggaa	24180
ggcagccttt ccgtgatta aaagtgcac tacgaagcat tgttctttcc ctcccttttt	24240
tttctgtagc ccgtttcacg gcctatcttg gagtcggcat ggcaaacttt atggctgagg	24300
tgagtttgc ttagtctcac ttttcattag cgtaatgac cagcttaca ctatatggga	24360
aatgtcctg aagtccactg ggctggcatc cagtggcagg atccatgacc atgagaagca	24420
ctgctctccc ttctcctgga gctccctggc ctttctttca gcatcacagc aaactttagt	24480
ccaaaccaca atcacccagt tgttacaagt atcagattgc ttggtttaaa aaaaaatgaa	24540
acgtaggttg tataacatat tatcaagttc agagtctaac tctaagtgat aagaagtaga	24600
ctttaggata tcttttactt aaacagaaaag ccagatattc cattgcaggt gatgcagggc	24660
cggtttctga tagcttagtc catgttgatg tggatcatggc tgctaaggag tcaaggcagt	24720
atctagccct ttgggcagca gcatggagat ttatctggg agggctccta aggagacaca	24780
gtgtctttct ggtggaaagc caaagtccta ttacacacat gcatgatgga gagtacatca	24840
gagcacatgg ggcccttcac atgtcaacaa agaagattca caggcatcag tcccaggacc	24900
caaatgggca agctgcacac cagagtcagc taggaagaca gaaaaatatg gaggcctagg	24960
ccctgtcctt tggatattct gatagagtag gtcttgtatg atgcttgaac atctgtgttt	25020
ttttttaact ccccgatg attctgatgt gcagtcagat tagggtacct ctacactcca	25080
tcacacccca gggagggtcca tgcacaggt cagagctaac caatggtgta tgctcagaat	25140
tgtgtgagtt tccatgagca gcacaaagag gacctacct caaggaactt agagtctatt	25200
tgggagacag aatggaaaga aacaaagca gtcaagtcta agatctagac caggcagaag	25260
tcaaggtcag agaggtcact gtgggctgga ctaatcagag aaggccttgt ggacatgaag	25320
actggtcagg ggccatttgc agtttgcaag tgctatctct gtcaaatgtt ctcttggcac	25380
atctggtgca ggaagtctga atatatgaga gggagagaaa gacatacaag atagagacat	25440
aagtggctgc cctaagaat ggatgtcaac attccaacaa ctcaatgccc tgagattgta	25500
aattcagtct ccacgagcat gcacagaatc cagagcaatg ccccgaggg ttcaccccc	25560

tgggctgaat gcaagtagag ggggatgcct tgtgcagctc agctgtcaga tgggatctga	25620
aaggagcgtg tggctttctc ttcttcccca gggtggattg ccagcttgta cctggccctt	25680
ctgtttggcc acgctattgt tctcatcat gaccacaaaa aattccaaca tctacaagat	25740
gcccctcagt aaagttactt atcctgaaga aaaccgcatc ttctacctgc aagccaagaa	25800
aagaatggtg gaaagccctt tgtgagaaca agccccattt gcagccatgg tcacgagtca	25860
tttctgcctg actgctccag ctaacttcca gggtctcagc aaactgctgt ttttcacgag	25920
tatcaacttt catactgacg cgtctgtaat ctgttcttat gctcattttg tattttcctt	25980
tcaactccag gaatatcctt gagcatatga gagtacatc caggatgatgt gctctggtat	26040
ggaatttgaa accccaatgg ggccttggca ctaagactgg aatgtatata aagtcaaagt	26100
gctccaacag aaggaggaag tgaacaaca ctattagtat ttattgatat tcttgggtgt	26160
tagctggctc gatgatgta acagtattaa aaattaaacc ccataaacca actaagcctt	26220
atggaattca cagtcacaaa atcgaagtta atccagaatt ctgtgataag cagcttggct	26280
ttttttttaa atcaatgcaa gttacacatt atagccagaa tctgtatcac agaggtgcaa	26340
gctgacagca gagctcagtc cccacttctt gcaacaatg gcctgcaccc tatcccttgt	26400
gtgtgtgaca ttctctcatg ggacaatgtt ggggtttttc agactgacag gactgcaaga	26460
gggagaaagg aattttgtca atcaaaatta ttctgtattg caacttttct cagagattgc	26520
aaaggatttt ttaggtagag attatttttc cttatgaaaa atgatctgtt ttaaatgaga	26580
taaaatagga gaagttcctg gcttaacctg ttcttacata ttaaagaaaa gttacttact	26640
gtatttatga aatactcagc ttaggcattt ttactttaac ccctaaattg attttgtaaa	26700
tgccacaaat gcatagaatt gttaccaacc tccaaagggc tctttaaat catattttt	26760
attcatttga ggatgtctta taaagactga aggcaaaggc cagattgctt acgggtgtta	26820
tttttataag ttgttgaatt ccttaattta aaaaagctca ttattttttg cacactcaca	26880
atattctctc tcagaaatca atggcatttg aaccacaaa aagaaataaa gggtgagtg	26940
cgggtgctca cgcctgtaat ccagcactt tggggagccc aggcgggcag attgcttgaa	27000
cccaggagtt caagaccagc ctgggcagca tggtgaaacc ctgtatctac aaaaaataca	27060
aaaattagcc aggcattggtg gtgggtgcct gtagttccag ctacttggga ggctgaggtg	27120
ggaaaatgac ttgagcccag gaggaggagg ctgcagttag ctaagattgc accactgcac	27180
tccaacctgg gcgacaagag tgaactgtg tctctcaaaa aaaaaaaaa acaaacaaaa	27240
acaaaaacaa aacaaaacaa aacaaaacaa aacaggtgag gattcccctg ttttctctc	27300
tttaatttta aagtatcag ttccgtaaag tctctgtaac caaacatact gaagacagca	27360
acagaagtca cgttcaggga ctggctcaca cctgtaatcc cagcactttg ggagatggag	27420

gtaaaaggat ctcttgagcc caggagttca agaccagctt gggcaacata gcaagactcc	27480
atctcttaaa aaataaaaaat agtaacatta gccagggtga gcagcacaca tctgcagcag	27540
ctactcagga ggctgagggtg gaaagatcgc ttgtgcacag aagttcgagg ctgcagttag	27600
ctatatgatac atgtcactgc actccagcct gtgtgaccga gcaagaccct atctcaaaaa	27660
aattaattaa ttaattaatt aattaattta aaaaggaagt catgttcatt tactttccac	27720
ttcagtgtgt atcgtgtagt attttggagg ttggaaagt aaacgtagga atcctgaaga	27780
ttttttccac ttctagttag cagtgtctag tgcacaatat acattttgct gaatgaataa	27840
acagaaatag ggaagtaaac ctacaaatat tttagggaga agctcacttc ttccttttct	27900
caggaaacca agcaagcaaa catatcggtc caatttttaa acccagtgac caaagccttt	27960
ggaactatga atttgcaact gtcatagggt tatggatatt gctgtggaga agctcaattt	28020
tcagtgtttt aactgaaccc tttcttggtta gggaacgtgt gaaagaagaa ttgtggggaa	28080
aaaaaagcaa gcataaccaa agatcatcag cagtgaagaa tctaggctgt ggctgagaga	28140
accagaggcc tctaaaatgg acccgagtcg atcttcagaa cagggatcta ccatgcagga	28200
gtttcttgtg ctcacacaaa tctgtaaatg ggaacattgt acattgtcga atttaaatga	28260
tattaatttt ctcaagctat tttgttact attttcctaa aattgaatat ttgcaggag	28320
catttatact ttttctaact gtctgtataa caaatttcta tgcaagtaca tgaataaatt	28380
atgctcacag ctca	28394
<210> 3	
<211> 1170	
<212> DNA	
<213> Homo Sapien	
<400> 3	
auggaggaca gccccacuau gguuagagug gacagcccca cuauuguuag gggugaaaac	60
cagguuucgc caugucaagg gagaaggugc uuccccaaag cucuuggcua ugucaccggu	120
gacaugaaag aacuugccaa ccagcuuaaa gacaaacccg uggugcucca guucauugac	180
uggauucucc ggggcuauc ccaaguggug uucgucaaca accccgucag uggaaucug	240
auucugguag gacuucuugu ucagaacccc uggugggcuc ucacuggcug gcugggaaca	300
guggucucca cucugauggc ccucuugcuc agccaggaca ggucuuuuu agcaucuggg	360
cucuauggcu acaaugccac ccugguggga guacucaugg cugucuuuuc ggacaaggga	420
gacuuuuuc gguggcuguu acucccugua ugugcuauu ccaugacuug cccauuuuc	480
ucaaugucau ugaaauccau gcucagcaaa ugggaccucc ccgucuucac ccuccuuuc	540

aacauggcgu ugucaaugua ccuuucagcc acaggacauu acaauccguu cuuuccagcc	600
aaacugguca uaccuauaac uacagcucca aaauaucuccu ggucugaccu cagugcccug	660
gaguuguuga aaucuauacc agugggaguu ggucagauu auggcuguga uaauccaugg	720
acagggggca uuuuccuggg agccauccua cucuccuccc cacucaugug ccugcaugcu	780
gccauaggau cauugcuggg cauagcagcg ggacucaguc uuucagcccc auuugagaac	840
aucuacuug gacucugggg uuucaacagc ucucuggccu gcauugcaau gggaggaug	900
uucauggcgc ucaccuggca aaccaccuc cuggcucuug gcugugcccu guucacggcc	960
uaucuuggag ucggcauggc aaacuuuug gcugagguug gauugccagc uuguaccugg	1020
cccuucuguu ugcccacgc auuguuccuc aucaugacca caaaaaauuc caacaucuac	1080
aagaugcccc ucaguaaagu uacuuauccu gaagaaaacc gcaucuucua ccugcaagcc	1140
aagaaaagaa ugguggaaag ccuuuuguga	1170
<210> 4	
<211> 1338	
<212> DNA	
<213> Homo Sapien	
<400> 4	
augaauggac ggucuuugau uggcggcgcu ggugacgccc gucauggucc uguuuggaag	60
gacccuuuug gaacuaaagc uggugacgca gcgcgcagag gcaucgcccg gcuaagcuug	120
gcccuggcag augggucgca ggaacaggag ccagaggaag agauagccau ggaggacagc	180
cccacuaugg uuagagugga cagccccacu augguuaggg gugaaaacca gguuucgcca	240
ugucaaggga gaaggugcuu ccccaaagcu cuuggcuau ucaccgguga caugaaagaa	300
cuugccaacc agcuuaaaga caaacccgug gucuccagu ucauugacug gauucuccgg	360
ggcauauccc aagugguguu cguacaac cccgucagug gaauccaau ucugguagga	420
cuucuuguuc agaaccuccg gugggcucuc acuggcuggc ugggaacagu ggucuccacu	480
cugauggccc ucuugcucag ccaggacagg ucauuauag caucugggcu cuauggcua	540
aaugccacc uggugggagu acucauggcu gucuuuucgg acaagggaga cuauuucugg	600
uggcuguuac ucccuguau ugcuauugc augacuugcc cauuuuucuc aagugcauug	660
aaauccaugc ucagcaaaug ggaccucccc gucuuacccc ucccuucaa cauggcguug	720
ucaauguacc uuucagccac aggacauuac aauccauuc uuccagcaa acuggucaua	780
ccuaaacia cagcucaaa uaucuccugg ucugaccuca gugcccugga guuguugaaa	840
ucuauaccag ugggaguugg ucagaucua ggucugaua auccauggac agggggcauu	900

uuccugggag ccauccuacu cuccucccca cucaugugcc ugcaugcugc cauaggauca	960
uugcugggca uagcagcggg acucagucuu ucagccccaau uugaggacau cuacuuugga	1020
cucugggguu ucaacagcuc ucuggccugc auugcaaugg gaggaauuu cauggcgcuc	1080
accuggcaaa cccaccuccu ggcucuuggc ugugcccugu ucacggccua ucuuggaguc	1140
ggcauggcaa acuuuauuggc ugagguugga uugccagcuu guaccuggcc cuucuguuug	1200
gccacgcua uguuccuau caugaccaca aaaaauucca acaucuacaa gaugcccccuc	1260
aguaaaguua cuuauccuga agaaaaccgc aucuucuacc ugcaagccaa gaaaagaug	1320
guggaaagcc cuuuguga	1338
<210> 5	
<211> 1170	
<212> DNA	
<213> Homo Sapien	
<400> 5	
auggaggaca gccccacuau gguuagagug gacagcccca cuaugguuag gggugaaaac	60
cagguuucgc caugucaagg gagaaggugc uuccccaaag cucuuggcua ugucaccggu	120
gacaugaaag aacuugccaa ccagcuuaaa gacaaaccgc uggugcucca guucauugac	180
uggauucucc ggggcauau ccaaguggug uucgucaaca accccaucag uggaauccug	240
auucugguag gacuucuugu ucagaacccc uggugggcuc ucacuggcug gcugggaaca	300
guggucucca cucugauggc ccucuugcuc agccaggaca ggucuuuau agcaucuggg	360
cucuauggcu acaaugccac ccugguggga guacucaugg cugucuuuuc ggacaaggga	420
gacuuuuuc gguugcuguu acucccugua ugugcuauu ccaugacuug cccauuuuuc	480
ucaaugcrau ugaauuccau gcucagcaaa ugaggaccucc ccgucuucac ccucccuuuc	540
aacauggcgu ugucaaugua ccuuucagcc acaggacauu acaauccguu cuuuccagcc	600
aaacugguca uaccuauaac uacagcucca aaauucuccu ggucugaccu cagugcccug	660
gaguuguuga aaucuauacc agugggaguu ggucagauu auggcuguga uaauccaugg	720
acagggggca uuuccuggg agccauccua cucuccucc cacucaugug ccugcaugcu	780
gccauaggau cauugcuggg cauagcagcg ggacucaguc uuucagcccc auuugagaac	840
aucuacuug gacucugggg uucaacagc ucucuggccu gcauugcaau gggagggaug	900
uucauggcgc ucaccuggca aaccaccuc cuggcucuug gcugugcccu guucacggcc	960
uaucuuagg ucggcauggc aaacuuuau gcugagguug gauugccagc uuguaccugg	1020
ccuucuguu ugccacgc auuguuccuc aucaugacca caaaaaauuc caacauuac	1080

aagaugcccc ucaguaaagu uacuuauccu gaagaaaacc gcaucuucua ccugcaagcc	1140
aagaaaagaa ugguggaaag cccuuuguga	1170
<210> 6	
<211> 1338	
<212> DNA	
<213> Homo Sapien	
<400> 6	
augaauaggac ggucuuugau uggcggcgcu ggugacgccc gucauggucc uguuuggaag	60
gacccuuuug gaacuaaagc uggugacgca gcgcgcagag gcaucgcccg gcuaagcuug	120
gcccuggcag augggucgca ggaacaggag ccagaggaag agauagccau ggaggacagc	180
cccacuaugg uuagagugga cagccccacu augguuaggg gugaaaacca gguuucgcca	240
ugucaaggga gaaggugcuu ccccaaagcu cuuggcuauug ucaccgguga caugaaagaa	300
cuugccaacc agcuuaaaga caaacccgug gugcuccagu ucauugacug gauucuccgg	360
ggcauauccc aagugguguu cgucaacaac cccaucagug gaauccuaau ucugguagga	420
cuucuuguuc agaaccuccug gugggcucuc acuggcuggc ugghaacagu ggucuccacu	480
cugaugggcc ucuugcucag ccaggacagg ucauuauuag caucugggcu cuauggcuauc	540
aaugccaccc uggugggagu acucauggcu gucuuuucgg acaagggaga cuauuucugg	600
uggcuguuac ucccuguaug ugcuaugucc augacuugcc caauuuucuc aagugcauug	660
aaauccaugc ucagcaaaug ggaccucccc gucuucaccc ucccuucaa cauggcguug	720
ucaauguacc uuucagccac aggacauuac aauccauucu uuccagccaa acuggucaua	780
ccuaaacia cagcuccaaa uaucuccugg ucugaccuca gugcccugga guuguugaaa	840
ucuauaccag ugggaguugg ucagaucuaug ggcugugaua auccauggac agggggcauu	900
uuccugggag ccauccuacu cuccucccca cucaugugcc ugcaugcugc cauaggauca	960
uugcugggca uagcagcggg acucagucuu ucagccccau uugaggacau cuacuuugga	1020
cucugggguu ucaacagcuc ucuggccugc auugcaaugg gaggaauuu cauggcguc	1080
accuggcaaa cccaccuccu ggcucuuggc ugugcccugu ucacggccua ucuuggaguc	1140
ggcauggcaa acuuuauaggc uagguuugga uugccagcuu guaccuggcc cuucuguuug	1200
gccacgcuaug uguuccuau caugaccaca aaaaauucca acaucuacia gaugccccuc	1260
aguaaaguua cuuauccuga agaaaaccgc aucuucuacc ugcaagccaa gaaaagaug	1320
guggaaagcc cuuuguga	1338
<210> 7	

<211> 1170

<212> DNA

<213> Artificial Sequence

<220><223> Wild-type SLC14A1 cDNA 1

<400> 7

atggaggaca gccccactat ggtagagtg gacagcccca ctatggtag gggtgaaaac 60

caggtttcgc catgtcaagg gagaagggtc tccccaaag ctcttggtta gtccaccgt 120

gacatgaaag aacttgccaa ccagcttaaa gacaaaccg tgggtgtcca gtccattgac 180

tggattctcc ggggcatatc ccaagtgtg ttcgtcaaca acccgtcag tggaatcctg 240

attctggtag gacttctgt tcagaacccc tgggtgggtc tctctggctg gctgggaaca 300

gtgtgttcca ctctgatgc cctctgtc agccaggaca ggtcattaat agcatctggg 360

ctctatggct acaatgccac cctgtggga gtactcatgg ctgtctttc ggacaaggga 420

gactatttct ggtggctgtt actccctgta tgtgtatgt ccatgacttg cccaatttc 480

tcaagtgc atgaattccat gtcagcaaa tgggacctc ccgtcttcac cctcccttc 540

aacatggcgt tgcaatgta ccttcagcc acaggacatt acaatccgtt cttccagcc 600

aaactgtca tacctataac tacagctcca aatattcct ggtctgacct cagtgcctg 660

gagttgtga aatctatacc agtgggagtt ggtcagatct atggctgtga taatccatgg 720

acagggggca ttttctggg agccatccta ctctctccc cactcatgtg cctgcatgt 780

gccataggat cattgtctgg catagcagcg ggactcagtc tttcagcccc atttgagaac 840

atctactttg gactctgggg tttcaacagc tctctggcct gcattgcaat gggaggaatg 900

ttcatggcgc tcacctggca aaccacctc ctggctcttg gctgtgccct gttcacggcc 960

tatcttgag tcggcatggc aaactttatg gctgaggttg gattgccagc ttgtacctgg 1020

cccttctgtt tggccacgct attgttctc atcatgacca caaaaaattc caacatctac 1080

aagatgcccc tcagtaaagt tacttattcct gaagaaaacc gcattctcta cctgcaagcc 1140

aagaaaagaa tgggtgaaag ccctttgtga 1170

<210> 8

<211> 1338

<212> DNA

<213> Artificial Sequence

<220><223> Wild-type SLC14A1 cDNA 2

<400> 8

atgaatggac ggtctttgat tggcggcgct ggtgacgcc gtcattgtcc tgtttggaag 60

gacccttttg gaactaaagc tggtagcgca gcgcgcagag gcatcgcccg gctaagcttg 120
gccctggcag atgggtcgca ggaacaggag ccagaggaag agatagccat ggaggacagc 180

cccactatgg ttagagtga cagccccact atggttaggg gtgaaaacca ggtttcgcca 240
tgtcaaggga gaaggtgctt ccccaaagct ctggctatg tcaccggtga catgaaagaa 300
cttgccaacc agcttaaaga caaacccgtg gtgtccagt tcattgactg gattctccgg 360
ggcatatccc aagtgggtgt cgtaacaac ccgctcagt gaatccta tctggtagga 420
cttcttggtc agaaccctg gtgggtctc actggctggc tgggaacagt ggtctccact 480
ctgatggccc tctgtctcag ccaggacagg tcattaatag catctgggct ctatggctac 540
aatgccaccc tgggtggagt actcatggct gtcttttcgg acaagggaga ctatttctgg 600

tggctgttac tccctgtatg tgctatgtcc atgacttgc caattttctc aagtgcattg 660
aattccatgc tcagcaaatg ggacctcccc gtcttcccc tccctttcaa catggcgttg 720
tcaatgtacc tttcagccac aggacattac aatccattct ttccagcaa actggtcata 780
cctataacta cagctccaaa tatctcctgg tctgacctca gtgccctgga gttgttgaaa 840
tctataccag tgggagtgg tcagatctat ggctgtgata atccatggac agggggcatt 900
ttcctgggag ccatcctact ctctcccca ctcatgtgcc tgcatgctgc cataggatca 960
ttgtctggca tagcagcggg actcagtctt tcagcccat ttgaggacat ctactttgga 1020

ctctgggggt tcaacagctc tctggcctgc attgcaatgg gaggaatgtt catggcgctc 1080
acctggcaaa cccacctct ggctcttggc tgtgccctgt tcacggccta tcttgagtc 1140
ggcatggcaa actttatggc tgaggttga ttgccagctt gtacctggcc ctctgtttg 1200
gccacgctat tgttctcat catgaccaca aaaaattcca acatctaca gatgccctc 1260
agtaaagtta ctatctga agaaaaccgc atcttctacc tgcaagcaa gaaaagaatg 1320
gtggaaagcc ctttgtga 1338

<210> 9
<211> 1170
<212> DNA

<213> Artificial Sequence
<220><223> Variant SLC14A1 (Val76Ile) cDNA
<400> 9

atggaggaca gccccactat ggtagagtg gacagcccca ctatggttag gggtgaaaac 60
caggtttcgc catgtcaagg gagaaggtgc tccccaaag ctcttggcta tgtaccggt 120
gacatgaaag aacttgccaa ccagcttaa gacaaaccg tgggtctcca gttcattgac 180

tggattctcc ggggcatatc ccaagtgggtg ttcgtcaaca accccatcag tggaaatcctg 240
attctggttag gacttcttgt tcagaacccc tgggtgggtc tcactggctg gctgggaaca 300
gtggtctcca ctctgatggc cctcttgtc agccaggaca ggtcattaat agcatctggg 360

ctctatggct acaatgccac cctgggtggga gtactcatgg ctgtcttttc ggacaaggga 420
gactatttct ggtggctgtt actccctgta tgtctatgt ccatgacttg cccaattttc 480
tcaagtgcac tgaattccat gctcagcaaa tgggacctcc ccgtcttcac cctccctttc 540
aacatggcgt tgtaaatgta cctttcagcc acaggacatt acaatccgtt ctttcagcc 600
aaactggtea tacctataac tacagctcca aatatctcct ggtctgacct cagtgcctg 660
gagttgttga aatctatacc agtgggagtt ggtcagatct atggctgtga taatccatgg 720
acagggggca ttttctggg agccatccta ctctctccc cactcatgtg cctgcatgct 780

gccataggat cattgctggg catagcagcg ggactcagtc tttcagcccc atttgagaac 840
atctactttg gactctgggg tttcaacagc tctctggcct gcattgcaat gggagggaatg 900
ttcatggcgc tcacctggca aaccacctc ctggctcttg gctgtgccct gttcacggcc 960
tatcttggag tcggcatggc aaactttatg gctgaggttg gattgccagc ttgtacctgg 1020
cccttctgtt tggccacgct attgttctc atcatgacca caaaaaattc caacatctac 1080
aagatgcccc tcagtaaagt tacttatect gaagaaaacc gcactttcta cctgcaagcc 1140
aagaaaagaa tgggtggaag ccctttgtga 1170

<210> 10

<211> 1338

<212> DNA

<213> Artificial Sequence

<220><223> Variant SLC14A1 (Val132Ile) cDNA

<400> 10

atgaatggac ggtctttgat tggcggcgt ggtgacgcc gtcattgtcc tgtttggaag 60
gaccttttg gaactaaagc tgggtgacgca gcgcgcagag gcacgcccc gctaagcttg 120
gccctggcag atgggtcgca ggaacaggag ccagaggaag agatagccat ggaggacagc 180
cccactatgg ttagagtga cagccccact atggttaggg gtgaaaacca ggtttcgcca 240
tgtcaaggga gaaggtgctt ccccaaagct cttggctatg tcaccggtga catgaaagaa 300

cttgccaacc agcttaaaga caaacccgtg gtgtccagt tcattgactg gattctccgg 360
ggcatatccc aagtgggtgtt cgtcaacaac cccatcagtg gaatcctaata tctggtagga 420
cttctgttgc agaaccctg gtgggtctc actggctggc tgggaacagt ggtctccact 480

ctgatggccc tcttgctcag ccaggacagg tcattaatag catctgggct ctatggctac 540
aatgccaccc tgggtgggagt actcatggct gtcttttcgg acaagggaga ctatttctgg 600
tggctgttac tccctgtatg tgctatgtcc atgacttgcc caattttctc aagtgcattg 660
aattccatgc tcagcaaatg ggacctcccc gtcttcaccc tccctttcaa catggcgttg 720

tcaatgtacc tttcagccac aggacattac aatccattct ttccagccaa actggtcata 780
cctataacta cagctccaaa tatctcctgg tctgacctca gtgccctgga gttgttgaaa 840
tctataccag tgggagtggg tcagatctat ggctgtgata atccatggac agggggcatt 900
ttctggggag ccactctact ctctcccca ctcatgtgcc tgcattgtgc cataggatca 960
ttgctgggca tagcagcggg actcagtctt tcagcccat ttgaggacat ctactttgga 1020
ctctgggggt tcaacagctc tctggcctgc attgcaatgg gaggaatgtt catggcgctc 1080
acctggcaaa cccacctct ggctcttggc tgtgcctgt tcacggccta tcttgagtc 1140

ggcatggcaa acittatggc tgaggttggg ttgccagctt gtacctggcc cttctgtttg 1200
gccacgctat tgttctcat catgaccaca aaaaattcca acatctacaa gatgccctc 1260
agtaaagtta cttatectga agaaaaccgc atcttctacc tgcaagccaa gaaaagaatg 1320
gtggaaagcc ctttgtga 1338

<210> 11
<211> 389
<212> PRT
<213> Homo Sapien
<400> 11

Met Glu Asp Ser Pro Thr Met Val Arg Val Asp Ser Pro Thr Met Val

1 5 10 15

Arg Gly Glu Asn Gln Val Ser Pro Cys Gln Gly Arg Arg Cys Phe Pro

20 25 30

Lys Ala Leu Gly Tyr Val Thr Gly Asp Met Lys Glu Leu Ala Asn Gln

35 40 45

Leu Lys Asp Lys Pro Val Val Leu Gln Phe Ile Asp Trp Ile Leu Arg

50 55 60

Gly Ile Ser Gln Val Val Phe Val Asn Asn Pro Val Ser Gly Ile Leu

65 70 75 80

Ile Leu Val Gly Leu Leu Val Gln Asn Pro Trp Trp Ala Leu Thr Gly

	85	90	95
Trp Leu Gly Thr Val Val Ser Thr Leu Met Ala Leu Leu Leu Ser Gln			
	100	105	110
Asp Arg Ser Leu Ile Ala Ser Gly Leu Tyr Gly Tyr Asn Ala Thr Leu			
	115	120	125
Val Gly Val Leu Met Ala Val Phe Ser Asp Lys Gly Asp Tyr Phe Trp			
	130	135	140
Trp Leu Leu Leu Pro Val Cys Ala Met Ser Met Thr Cys Pro Ile Phe			
145	150	155	160
Ser Ser Ala Leu Asn Ser Met Leu Ser Lys Trp Asp Leu Pro Val Phe			
	165	170	175
Thr Leu Pro Phe Asn Met Ala Leu Ser Met Tyr Leu Ser Ala Thr Gly			
	180	185	190
His Tyr Asn Pro Phe Phe Pro Ala Lys Leu Val Ile Pro Ile Thr Thr			
	195	200	205
Ala Pro Asn Ile Ser Trp Ser Asp Leu Ser Ala Leu Glu Leu Leu Lys			
	210	215	220
Ser Ile Pro Val Gly Val Gly Gln Ile Tyr Gly Cys Asp Asn Pro Trp			
225	230	235	240
Thr Gly Gly Ile Phe Leu Gly Ala Ile Leu Leu Ser Ser Pro Leu Met			
	245	250	255
Cys Leu His Ala Ala Ile Gly Ser Leu Leu Gly Ile Ala Ala Gly Leu			
	260	265	270
Ser Leu Ser Ala Pro Phe Glu Asp Ile Tyr Phe Gly Leu Trp Gly Phe			
	275	280	285
Asn Ser Ser Leu Ala Cys Ile Ala Met Gly Gly Met Phe Met Ala Leu			
	290	295	300
Thr Trp Gln Thr His Leu Leu Ala Leu Gly Cys Ala Leu Phe Thr Ala			
305	310	315	320
Tyr Leu Gly Val Gly Met Ala Asn Phe Met Ala Glu Val Gly Leu Pro			
	325	330	335
Ala Cys Thr Trp Pro Phe Cys Leu Ala Thr Leu Leu Phe Leu Ile Met			

340 345 350
 Thr Thr Lys Asn Ser Asn Ile Tyr Lys Met Pro Leu Ser Lys Val Thr
 355 360 365
 Tyr Pro Glu Glu Asn Arg Ile Phe Tyr Leu Gln Ala Lys Lys Arg Met

 370 375 380
 Val Glu Ser Pro Leu
 385
 <210> 12
 <211> 445
 <212> PRT
 <213> Homo Sapien
 <400> 12
 Met Asn Gly Arg Ser Leu Ile Gly Gly Ala Gly Asp Ala Arg His Gly
 1 5 10 15
 Pro Val Trp Lys Asp Pro Phe Gly Thr Lys Ala Gly Asp Ala Ala Arg
 20 25 30
 Arg Gly Ile Ala Arg Leu Ser Leu Ala Leu Ala Asp Gly Ser Gln Glu
 35 40 45

 Gln Glu Pro Glu Glu Glu Ile Ala Met Glu Asp Ser Pro Thr Met Val
 50 55 60
 Arg Val Asp Ser Pro Thr Met Val Arg Gly Glu Asn Gln Val Ser Pro
 65 70 75 80
 Cys Gln Gly Arg Arg Cys Phe Pro Lys Ala Leu Gly Tyr Val Thr Gly
 85 90 95
 Asp Met Lys Glu Leu Ala Asn Gln Leu Lys Asp Lys Pro Val Val Leu
 100 105 110
 Gln Phe Ile Asp Trp Ile Leu Arg Gly Ile Ser Gln Val Val Phe Val

 115 120 125
 Asn Asn Pro Val Ser Gly Ile Leu Ile Leu Val Gly Leu Leu Val Gln
 130 135 140
 Asn Pro Trp Trp Ala Leu Thr Gly Trp Leu Gly Thr Val Val Ser Thr
 145 150 155 160

Leu Met Ala Leu Leu Leu Ser Gln Asp Arg Ser Leu Ile Ala Ser Gly
 165 170 175
 Leu Tyr Gly Tyr Asn Ala Thr Leu Val Gly Val Leu Met Ala Val Phe
 180 185 190

 Ser Asp Lys Gly Asp Tyr Phe Trp Trp Leu Leu Leu Pro Val Cys Ala
 195 200 205
 Met Ser Met Thr Cys Pro Ile Phe Ser Ser Ala Leu Asn Ser Met Leu
 210 215 220
 Ser Lys Trp Asp Leu Pro Val Phe Thr Leu Pro Phe Asn Met Ala Leu
 225 230 235 240
 Ser Met Tyr Leu Ser Ala Thr Gly His Tyr Asn Pro Phe Phe Pro Ala
 245 250 255
 Lys Leu Val Ile Pro Ile Thr Thr Ala Pro Asn Ile Ser Trp Ser Asp

 260 265 270
 Leu Ser Ala Leu Glu Leu Leu Lys Ser Ile Pro Val Gly Val Gly Gln
 275 280 285
 Ile Tyr Gly Cys Asp Asn Pro Trp Thr Gly Gly Ile Phe Leu Gly Ala
 290 295 300
 Ile Leu Leu Ser Ser Pro Leu Met Cys Leu His Ala Ala Ile Gly Ser
 305 310 315 320
 Leu Leu Gly Ile Ala Ala Gly Leu Ser Leu Ser Ala Pro Phe Glu Asp
 325 330 335

 Ile Tyr Phe Gly Leu Trp Gly Phe Asn Ser Ser Leu Ala Cys Ile Ala
 340 345 350
 Met Gly Gly Met Phe Met Ala Leu Thr Trp Gln Thr His Leu Leu Ala
 355 360 365
 Leu Gly Cys Ala Leu Phe Thr Ala Tyr Leu Gly Val Gly Met Ala Asn
 370 375 380
 Phe Met Ala Glu Val Gly Leu Pro Ala Cys Thr Trp Pro Phe Cys Leu
 385 390 395 400
 Ala Thr Leu Leu Phe Leu Ile Met Thr Thr Lys Asn Ser Asn Ile Tyr

405 410 415
 Lys Met Pro Leu Ser Lys Val Thr Tyr Pro Glu Glu Asn Arg Ile Phe
 420 425 430
 Tyr Leu Gln Ala Lys Lys Arg Met Val Glu Ser Pro Leu
 435 440 445
 <210> 13
 <211> 389
 <212> PRT
 <213> Homo Sapien
 <400> 13
 Met Glu Asp Ser Pro Thr Met Val Arg Val Asp Ser Pro Thr Met Val
 1 5 10 15
 Arg Gly Glu Asn Gln Val Ser Pro Cys Gln Gly Arg Arg Cys Phe Pro
 20 25 30
 Lys Ala Leu Gly Tyr Val Thr Gly Asp Met Lys Glu Leu Ala Asn Gln
 35 40 45
 Leu Lys Asp Lys Pro Val Val Leu Gln Phe Ile Asp Trp Ile Leu Arg
 50 55 60
 Gly Ile Ser Gln Val Val Phe Val Asn Asn Pro Ile Ser Gly Ile Leu
 65 70 75 80
 Ile Leu Val Gly Leu Leu Val Gln Asn Pro Trp Trp Ala Leu Thr Gly
 85 90 95
 Trp Leu Gly Thr Val Val Ser Thr Leu Met Ala Leu Leu Leu Ser Gln
 100 105 110
 Asp Arg Ser Leu Ile Ala Ser Gly Leu Tyr Gly Tyr Asn Ala Thr Leu
 115 120 125
 Val Gly Val Leu Met Ala Val Phe Ser Asp Lys Gly Asp Tyr Phe Trp
 130 135 140
 Trp Leu Leu Leu Pro Val Cys Ala Met Ser Met Thr Cys Pro Ile Phe
 145 150 155 160
 Ser Ser Ala Leu Asn Ser Met Leu Ser Lys Trp Asp Leu Pro Val Phe
 165 170 175

Thr Leu Pro Phe Asn Met Ala Leu Ser Met Tyr Leu Ser Ala Thr Gly
180 185 190
His Tyr Asn Pro Phe Phe Pro Ala Lys Leu Val Ile Pro Ile Thr Thr
195 200 205
Ala Pro Asn Ile Ser Trp Ser Asp Leu Ser Ala Leu Glu Leu Leu Lys
210 215 220
Ser Ile Pro Val Gly Val Gly Gln Ile Tyr Gly Cys Asp Asn Pro Trp
225 230 235 240

Thr Gly Gly Ile Phe Leu Gly Ala Ile Leu Leu Ser Ser Pro Leu Met
245 250 255
Cys Leu His Ala Ala Ile Gly Ser Leu Leu Gly Ile Ala Ala Gly Leu
260 265 270
Ser Leu Ser Ala Pro Phe Glu Asp Ile Tyr Phe Gly Leu Trp Gly Phe
275 280 285
Asn Ser Ser Leu Ala Cys Ile Ala Met Gly Gly Met Phe Met Ala Leu
290 295 300
Thr Trp Gln Thr His Leu Leu Ala Leu Gly Cys Ala Leu Phe Thr Ala

305 310 315 320
Tyr Leu Gly Val Gly Met Ala Asn Phe Met Ala Glu Val Gly Leu Pro
325 330 335
Ala Cys Thr Trp Pro Phe Cys Leu Ala Thr Leu Leu Phe Leu Ile Met
340 345 350
Thr Thr Lys Asn Ser Asn Ile Tyr Lys Met Pro Leu Ser Lys Val Thr
355 360 365
Tyr Pro Glu Glu Asn Arg Ile Phe Tyr Leu Gln Ala Lys Lys Arg Met
370 375 380

Val Glu Ser Pro Leu

385

<210> 14

<211> 445

<212> PRT

<213> Homo Sapien

<400> 14

Met Asn Gly Arg Ser Leu Ile Gly Gly Ala Gly Asp Ala Arg His Gly

1 5 10 15

Pro Val Trp Lys Asp Pro Phe Gly Thr Lys Ala Gly Asp Ala Ala Arg

20 25 30

Arg Gly Ile Ala Arg Leu Ser Leu Ala Leu Ala Asp Gly Ser Gln Glu

35 40 45

Gln Glu Pro Glu Glu Glu Ile Ala Met Glu Asp Ser Pro Thr Met Val

50 55 60

Arg Val Asp Ser Pro Thr Met Val Arg Gly Glu Asn Gln Val Ser Pro

65 70 75 80

Cys Gln Gly Arg Arg Cys Phe Pro Lys Ala Leu Gly Tyr Val Thr Gly

85 90 95

Asp Met Lys Glu Leu Ala Asn Gln Leu Lys Asp Lys Pro Val Val Leu

100 105 110

Gln Phe Ile Asp Trp Ile Leu Arg Gly Ile Ser Gln Val Val Phe Val

115 120 125

Asn Asn Pro Ile Ser Gly Ile Leu Ile Leu Val Gly Leu Leu Val Gln

130 135 140

Asn Pro Trp Trp Ala Leu Thr Gly Trp Leu Gly Thr Val Val Ser Thr

145 150 155 160

Leu Met Ala Leu Leu Leu Ser Gln Asp Arg Ser Leu Ile Ala Ser Gly

165 170 175

Leu Tyr Gly Tyr Asn Ala Thr Leu Val Gly Val Leu Met Ala Val Phe

180 185 190

Ser Asp Lys Gly Asp Tyr Phe Trp Trp Leu Leu Leu Pro Val Cys Ala

195 200 205

Met Ser Met Thr Cys Pro Ile Phe Ser Ser Ala Leu Asn Ser Met Leu

210 215 220

Ser Lys Trp Asp Leu Pro Val Phe Thr Leu Pro Phe Asn Met Ala Leu

225 230 235 240

Ser Met Tyr Leu Ser Ala Thr Gly His Tyr Asn Pro Phe Phe Pro Ala

	245		250		255
Lys	Leu	Val	Ile	Pro	Ile
Thr	Thr	Ala	Pro	Asn	Ile
Ser	Trp	Ser	Asp		
	260		265		270
Leu	Ser	Ala	Leu	Glu	Leu
Lys	Ser	Ile	Pro	Val	Gly
Val	Gly	Gln			
	275		280		285
Ile	Tyr	Gly	Cys	Asp	Asn
Pro	Trp	Thr	Gly	Gly	Ile
Phe	Leu	Gly	Ala		
	290		295		300
Ile	Leu	Leu	Ser	Ser	Pro
Leu	Met	Cys	Leu	His	Ala
Ala	Ile	Gly	Ser		
	305		310		315
					320
Leu	Leu	Gly	Ile	Ala	Ala
Gly	Leu	Ser	Leu	Ser	Ala
Pro	Phe	Glu	Asp		
	325		330		335
Ile	Tyr	Phe	Gly	Leu	Trp
Gly	Phe	Asn	Ser	Ser	Leu
Ala	Cys	Ile	Ala		
	340		345		350
Met	Gly	Gly	Met	Phe	Met
Ala	Leu	Thr	Trp	Gln	Thr
His	Leu	Leu	Ala		
	355		360		365
Leu	Gly	Cys	Ala	Leu	Phe
Thr	Ala	Tyr	Leu	Gly	Val
Gly	Met	Ala	Asn		
	370		375		380
Phe	Met	Ala	Glu	Val	Gly
Leu	Pro	Ala	Cys	Thr	Trp
Pro	Phe	Cys	Leu		
	385		390		395
					400
Ala	Thr	Leu	Leu	Phe	Leu
Ile	Met	Thr	Thr	Lys	Asn
Ser	Asn	Ile	Tyr		
	405		410		415
Lys	Met	Pro	Leu	Ser	Lys
Val	Thr	Tyr	Pro	Glu	Glu
Asn	Arg	Ile	Phe		
	420		425		430
Tyr	Leu	Gln	Ala	Lys	Lys
Arg	Met	Val	Glu	Ser	Pro
Leu					
	435		440		445