



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2010-0028985
(43) 공개일자 2010년03월15일

(51) Int. Cl.

C12Q 1/68 (2006.01)

(21) 출원번호 10-2008-0087980

(22) 출원일자 2008년09월05일

심사청구일자 2008년09월05일

(71) 출원인

고려대학교 산학협력단

서울 성북구 안암동5가1 고려대학교 내

(72) 발명자

김열홍

서울특별시 송파구 방이동 89 올림픽선수촌아파트
257-305

정해윤

서울특별시 중랑구 묵1동 180-3 현대하이츠 A동
402호

(뒷면에 계속)

(74) 대리인

이문섭

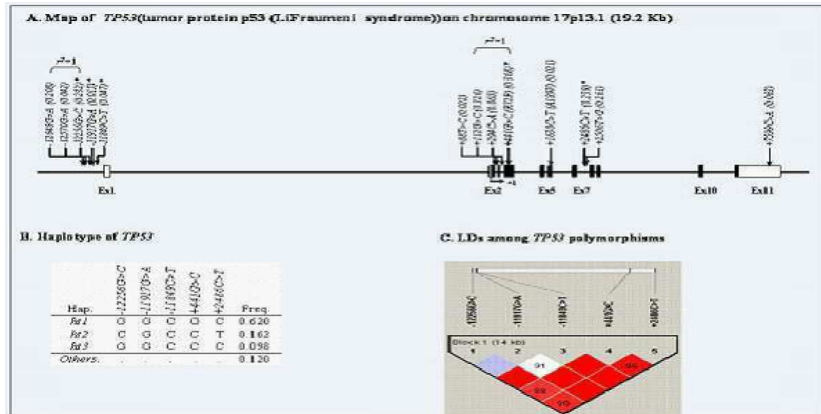
전체 청구항 수 : 총 10 항

(54) 한국인의 폐암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트

(57) 요약

본 발명은 한국인의 폐암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트에 관한 것으로, 보다 구체적으로 본 발명은 TP53(Tumor Protein 53) 유전자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 다형성의 빈도를 측정하여 한국인의 폐암 발생 위험도를 예측하는 방법 및 이를 이용하는 진단 키트에 관한 것이다. 본 발명에 의하면, 폐암 억제 유전자 TP53의 단일염기다형성 및 이의 일배체형과 한국인 폐암 발생과의 연관성을 확인하여 TP53이 폐암 위험 표지 인자로 사용 가능한지 여부를 확인하는데 유용하게 사용될 수 있다.

대표도 - 도1



(72) 발명자

황영미

서울특별시 종로구 창신3동 쌍용아파트 202동 807호

성재숙

서울특별시 구로구 개봉2동 333-8 스카이빌 203호

신형두

서울시 금천구 가산동 371-28 우림라이온스밸리 B동 14층1407호

박병래

서울시 금천구 가산동 371-28 우림라이온스밸리 B동 14층1407호

김준석

서울특별시 서초구 방배동 88-53 엘지방배자이아파트 101-304

신상원

서울특별시 양천구 목동 907-14 벽산미라지타워 1704호

서희연

서울특별시 성동구 금호동3가 두산아파트 115-1510

서재홍

미국 마이애미 01002 압허스트 그린드라이브 롤링 103

이 발명을 지원한 국가연구개발사업

과제고유번호 A010250A7100406A600020A

부처명

연구사업명 보건의료기술개발사업

연구과제명 폐암 및 유방암/난소암 유전체 연구센터(정부)

주관기관 고려대학교

연구기간 2006년 04월 01일 ~ 2007년 03월31일

특허청구의 범위

청구항 1

1) 핵산시료로부터 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 (Tumor Protein 53) 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 빈도를 측정하는 단계; 및

2) 단계 1)에서 측정된 실험군의 다형성 빈도를 대조군과 비교하여 상관관계를 구하는 단계를 포함하는, 한국인에서 폐암 발생 위험도를 예측하는 방법.

청구항 2

제1항에 있어서, 1) 단계의 다형성 빈도가 DNA 서열분석 (sequencing), PCR-SSCP (single stranded conformation polymorphism) 분석, RFLP (restriction fragment length polymorphism) 분석, 대립형질 (allele) 특이적 혼성화 방법, DNA 칩 방법, ELISA (enzyme-linked immunosorbent assay) 방법 및 면역블롯 (immunoblot) 방법, 스냅샷(sNaPShot) 분석 방법, TDGS(Two-dimensional Gene Scanning), Taq-Man[®] 및 TOGATM으로 구성된 군으로부터 선택된 하나 이상의 방법에 의해 측정되는 것을 특징으로 하는 방법.

청구항 3

제1항에 있어서, 2) 단계에서 상관관계가 스튜던트 t-검정 (Student's t-test), 카이-스퀘어 검정 (Chi-square test) 및 다중 로지스틱 회귀 분석 (multiple logistic regression analysis)으로 구성된 군으로부터 선택된 하나 이상의 방법에 의해 얻어지는 것을 특징으로 하는 방법.

청구항 4

제1항의 방법을 이용하여 한국인에서 폐암의 유전적 위험 인자를 검사하는 진단 키트.

청구항 5

서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 단일염기다형성 부위를 증폭시킬 수 있는 프라이머쌍.

청구항 6

제5항에 있어서, 프라이머쌍이 서열번호:1 및 2로 기재되는 염기서열을 갖는 올리고뉴클레오티드, 서열번호:3 및 4로 기재되는 염기서열을 갖는 올리고뉴클레오티드, 서열번호:5 및 6로 기재되는 염기서열을 갖는 올리고뉴클레오티드, 서열번호:7 및 8로 기재되는 염기서열을 갖는 올리고뉴클레오티드 및 서열번호:9 및 10로 기재되는 염기서열을 갖는 올리고뉴클레오티드로 이루어진 군에서 선택된 어느 하나인 것을 특징으로 하는 프라이머쌍.

청구항 7

서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역

시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 단일염기다형성 부위에 혼성화할 수 있는 프로브쌍.

청구항 8

제7항에 있어서, 프로브쌍이 서열번호:11 및 12로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:13 및 14로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:15 및 16로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:17 및 18으로 기재되는 염기서열을 갖는 올리고뉴클레오타이드 및 서열번호:19 및 20으로 기재되는 염기서열을 갖는 올리고뉴클레오타이드로 이루어진 군에서 선택된 어느 하나인 것을 특징으로 하는 프로브쌍.

청구항 9

서열번호 21로 이루어지는 폴리뉴클레오타이드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되고 상기 750번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오타이드, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되고 상기 1089번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오타이드, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되고 상기 1157번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오타이드, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되고 상기 13446번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오타이드, 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되고 상기 15491번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오타이드 및 이들의 상보적인 폴리뉴클레오타이드로 이루어진 군 중에서 선택된 하나 이상의 폴리뉴클레오타이드로 이루어지는 폐암 감수성 진단용 마커.

청구항 10

서열번호 21로 이루어지는 폴리뉴클레오타이드에서, 서열번호 21의 750번째(TP53 유전자의 번역시작점으로부터 -12256번째) 염기가 G이고, 서열번호 21의 1089번째(TP53 유전자의 번역시작점으로부터 -11917번째) 염기가 G이고, 서열번호 21의 1157번째(TP53 유전자의 번역시작점으로부터 -11849번째) 염기가 C이고, 서열번호 21의 13446번째(TP53 유전자의 번역시작점으로부터 +441번째) 염기가 C이고, 서열번호 21의 15491번째(TP53 유전자의 번역시작점으로부터 +2486번째) 염기가 C로 일배체형(haplotype)을 이루는 것을 특징으로 하는 감수성 진단용 마커.

명세서

발명의 상세한 설명

기술 분야

[0001] 본 발명은 한국인의 폐암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트에 관한 것으로, 보다 구체적으로 본 발명은 TP53 유전자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 단일염기다형성의 빈도를 측정하여 한국인의 발암 발생 위험도를 예측하는 방법 및 이를 이용하는 진단 키트에 관한 것이다.

배경 기술

[0002] 폐암은 폐에 발생하는 암이다. 흡연, 공해 등이 가장 큰 원인인 선진국형 암으로, 20세기에 들어서면서 구미 각국에서 급격히 증가하기 시작하여 전 세계적으로 매년 130만 명 이상이 폐암으로 사망하며, 암으로 인한 사망에서 가장 높은 비중을 차지하고 있다. 한국의 경우에도, 매년 10여 만 명의 암 환자가 새로 발생하고 5만 여명의 암 환자가 사망하고 있는 것으로 보고되었다. 더욱이 암의 발생 빈도는 최근 들어 더욱 증가하는 추세로 현재 암은 우리나라 성인 사망 원인의 2위를 차지하고 있다. 특히, 폐암은 한국 성인에서 발생하는 암중에서 약 12%를 차지하며 위암, 간암에 이어 제3위의 발생률을 보이며 매년 남녀 모두에서 발생율이 증가하고 있다. 폐암의 발생율은 여성보다 남성에서 현저히 높으며 상대적으로 45세 미만의 젊은 환자의 비율이 높은 것으로 보고되었다. 더구나 폐암은 진단 당시 이미 다른 장기로 전이를 하였거나 전이가 없는 경우에도 국소적으로 진행되어 근치적절제술, 함암화학요법, 방사선치료 등의 다양한 치료법에도 불구하고 치료 후 재발과 전이에 의해 5년 생존

율이 5% 정도에 머무르는 완치율이 매우 낮은 종양으로 암에 의한 사망율 1위를 차지하고 있다. 따라서 치료 성적을 높이기 위해서는 폐암의 발생 및 그의 원인 기전, 암의 성장, 그리고 전이에 관여하는 분자생물학적 조절 기전 등을 규명하여 각 단계별로 대응하는 새로운 진단 및 치료 방법의 개발이 절실히 요구된다.

[0003] 1970년대부터 분자 생물학의 발달과 함께 종양 유전자 및 종양억제 유전자가 발굴되면서 암은 다단계 유전자 변이 질환임이 규명되었다. 즉, 암은 단일 유전자 결손 질환과 달리 여러 개의 유전자가 유전자 돌연변이, 증폭, 결손, 치환 등의 단계적 변화 결과 악성으로 이행한다는 발암 기전이 제시되었다. 지금까지 '종양 유전학' 분야에서는 이러한 분자생물학적 다단계 모델 하에 미리 예측한 특정 1-2개의 유전자를 대상으로 그 발현이나 또는 돌연변이 유무를 조사해왔으나 종양의 다단계 모델을 완벽하게 밝히거나 발암 기전 규명에 효과적이지 못했다.

[0004] 생명과학 분야의 발달로 한꺼번에 수많은 유전자들의 발현을 측정할 수 있는 기법들이 개발되었거나 개발 중이기 때문에 21세기 생명 공학 분야는 엄청나게 축적된 유전정보들을 분석해서 그 유전자의 기능과 상호작용을 밝히는 functional genomics 시대로 이행해 가고 있다. 유전정보를 집중화하고 컴퓨터를 이용하여 비교 분석하는 생물정보학과 더불어 초정밀 집적 마이크로어레이를 이용한 다량의 유전자 발현을 고속 검색할 수 있는 기술은 분자생물학적 지식과 현대 기계 및 전자공학의 기술을 접목시킨 것으로 유전자 발현 연구에 획기적인 전지를 마련하였을 뿐만 아니라 functional genomics 라는 새로운 학문의 장을 개척하는 등 다양한 용도로 활용될 중요한 기반 기술이다.

[0005] 최근 동일한 발암요인에 대해서도 개인이 지닌 유전적 감수성에 따라 위험도가 달라지며 이러한 개인 및 인종간의 감수성을 구분 짓는 가장 중요한 표지자로 SNP(single nucleotide polymorphism)가 제시되고 있다.

[0006] SNP란 개인과 개인 간의 DNA에 존재하는 한 염기쌍(single base-pair variation)의 차이를 말하는 것으로 DNA sequence 다형성(polymorphism)중에서 가장 많이 존재하는 형태이다 (약 1개/1kb).

[0007] SNP가 주목을 끄는 이유는 가장 높은 빈도를 보이는 변이 형태이며, 상당히 안정적인 변이이고 낮은 변이율을 보인다는 점이다. 특히, 단백질을 만드는 유전자 부위에 위치한 core SNP의 경우는 기능적 중요성을 갖기 때문에 연구가 집중되고 있으며, 중요한 유전자들의 발현을 조절하는 부위에 대한 해당 유전자의 검색도 수행되고 있다. 또한, SNP는 DNA chip 기술과 접목되어 질병 발생에 대한 개인의 감수성 차이를 규명하여 21세기 맞춤의학 혹은 예측의학이라는 새로운 시대로 나가는데 있어 가장 핵심적인 역할을 할 것으로 예상된다. 암 진단 및 치료의 초기 단계에서 이와 같은 약제학적으로 중요한 단백질의 유전적 다양성에 대한 효과를 고려하는 것은 항암치료의 실패에 대한 위험성을 낮출 수 있는 중요한 요소로 작용하여개체간의 정확하고 신뢰할 수 있는 진단 및 치료 정보를 제공할 것이다.

[0008] 이와 같은 SNP 연구는 세계적으로 활발히 진행되고 있지만 각 나라마다 인종적인 차이 때문에 한계가 있다. 따라서 국내 SNP 연구는 국제적인 경쟁력이 충분히 있으며, 호흡기암에 대한 cSNPs에 대한 연구는 많지 않으므로 SNP의 체계적인 발굴과 기능 연구 등에 대한 효율적인 투자가 필요한 실정이다.

[0009] 본 발명은 폐암 발생을 예측할 수 있는 방법, 특히, 유전자 다형 분석을 사용하여 여자, 남성, 흡연자와 비흡연자에서 폐암 발생 위험을 평가하는 방법에 관한 것이다. 보다 구체적으로, 본 발명은 한국인에서 암 억제 유전자인 TP53 유전자의 다형성, haplotype과 폐암의 위험성과의 연관성을 예측하는 것이다.

[0010] 폐암은 다양한 원인에 의해서 다단계적으로 발생하는 암으로서, 특히, 암유발유전자(oncogenes)과 암 억제 유전자의 (tumor suppressor genes)의 유전적 변이가 주요한 원인 인자 중 하나로 보고되고 있다. 대표적인 암 억제 유전자인 TP53은 염색체의 17p13.1에 위치하고 DNA의 회복, 세포주기와 세포사멸에 중요한 역할을 한다. 또한 미스센스돌연변이와 같은 유전적 변이에 의한 TP53의 기능 상실이 폐암의 암 발생과정에서 중요한 기능을 하는 것으로 보고되었다. 특히, TP53 변이는 90%이상의 소세포성폐암(small cell lung cancer, SCLC)와 50%이상의 비소세포성폐암 (non-small cell lung cancer, NSCLC)에서 나타난다.

[0011] 최근 들어서 폐암 관련 유전자의 단일염기다형성 (single nucleotide polymorphisms; SNP)과 폐암 발생 간의 연관성에 대한 연구가 활발히 이루어지고 있다. 이를 통해 폐암 발생과 예후에 대한 예측으로 폐암의 치료 효율을 높이거나 하는 폐암진단 및 평가에 유용한 다수의 바이오마커 (biomarker)가 개발 중이다. Tp53의 단일염기 다형성과 폐암 발생과의 연관성에 대해서도 많은 연구가 이뤄지고 있으나 대부분은 Arg72 또는 Pro72 단백질로 번역되는 엑손4 코돈 72의 두번째 염기의 G/C 변이에 관한 것이다(Petitjean et al.2007). 그러나 이에 대한 지금까지의 연구들은 코돈 72의 다형성과 폐암 발병 간의 연관성에 대해 서로 모순되는 결과만을 내놓고 있으며 서로 일치점을 찾지 못하고 있는 상태이다. 일부 연구에서는 코돈 72의 다형성은 폐암 발병과 뚜렷한 연관성이 있다고 보고하나 (Nadji et al. 2007; Popanda et al. 2007) 또 다른 일부에서는 이들 사이에 전혀 연관성이

없다고 보고하고 있다(Khadang et al. 2007; Rajaraman et al. 2007). 뿐만 아니라 코돈 72 이외에 다른 TP53 유전자의 다형성에 대해서는 연구가 거의 이뤄지지 않고 있으며 한국인에게 있어 TP53 유전자의 다형성과 폐암 발병과의 연관성에 대해서는 본 발명 이전에는 전혀 연구된 바 없다. 이에 본 발명자들은 한국인에게서 암 억제 유전자인 TP53의 단일염기다형성 및 일배체형과 폐암발생과의 연관성 존부를 확인할 수 있는 방법 및 이를 이용한 진단 키트를 찾아낸 바 본 발명은 이에 관한 것이다.

발명의 내용

해결 하고자하는 과제

- [0012] 본 발명의 목적은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성의 빈도를 측정하여 한국인의 발암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트를 제공하는 것이다.
- [0013] 또한, 본 발명의 목적은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위를 증폭시킬 수 있는 프라이머쌍과 이를 포함하는 폐암 발생 위험도 진단용 키트를 제공하는 것이다.
- [0014] 또한, 본 발명의 목적은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되고 상기 750번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되고 상기 1089번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되고 상기 1157번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되고 상기 13446번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되고 상기 15491번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드 및 이들의 상보적인 폴리뉴클레오티드로 이루어진 군 중에서 선택된 하나 이상의 폴리뉴클레오티드로 이루어지는 폐암 감수성 진단용 마커를 제공하는 것이다.
- [0015] 본 발명의 목적은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째(TP53 유전자의 번역시작점으로부터 -12256번째) 염기가 G이고, 서열번호 21의 1089번째(TP53 유전자의 번역시작점으로부터 -11917번째) 염기가 G이고, 서열번호 21의 1157번째(TP53 유전자의 번역시작점으로부터 -11849번째) 염기가 C이고, 서열번호 21의 13446번째(TP53 유전자의 번역시작점으로부터 +441번째) 염기가 C이고, 서열번호 21의 15491번째(TP53 유전자의 번역시작점으로부터 +2486번째) 염기가 C로 일배체형(haplotype)을 이루는 것을 특징으로 하는 감수성 진단용 마커를 제공하는 것이다.
- [0016] 본 발명의 목적은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위에 혼성화할 수 있는 프로브쌍을 제공하는 것이다.

과제 해결수단

- [0017] 상기 목적을 달성하기 위하여, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성의 빈도를 측정하여 한국인의 발암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트를 제공한다.
- [0018] 또한, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위를 증폭시킬 수 있는 프라이머쌍과 이를 포함하는 폐암 발생 위험도 진단용 키트를 제공한다.
- [0019] 또한, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되고 상기 750번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되고 상기 1089번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되고 상기 1157번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되고 상기 13446번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되고 상기 15491번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드 및 이들의 상보적인 폴리뉴클레오티드로 이루어진 군 중에서 선택된 하나 이상의 폴리뉴클레오티드로 이루어지는 폐암 감수성 진단용 마커를 제공한다.
- [0020] 또한, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째(TP53 유전자의 번역시작점으로부터 -12256번째) 염기가 G이고, 서열번호 21의 1089번째(TP53 유전자의 번역시작점으로부터 -11917번째) 염기가 G이고, 서열번호 21의 1157번째(TP53 유전자의 번역시작점으로부터 -11849번째) 염기가 C이고, 서열번호 21의 13446번째(TP53 유전자의 번역시작점으로부터 +441번째) 염기가 C이고, 서열번호 21의 15491번째(TP53 유전자의 번역시작점으로부터 +2486번째) 염기가 C로 일배체형(haplotype)을 이루는 것을 특징으로 하는 감수성 진단용 마커를 제공한다.
- [0021] 또한, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위에 혼성화할 수 있는 프로브쌍을 제공한다.
- [0022] 이하, 본 발명을 상세히 설명한다.
- [0023] 본 발명의 한 양태에 따르면, 한국인의 발암 발생 위험도를 예측하는 방법은
- [0024] 1) 핵산시료로부터 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기

(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 빈도를 측정하는 단계; 및

- [0025] 2) 단계 1)에서 측정된 실험군의 다형성 빈도를 대조군과 비교하여 상관관계를 구하는 단계를 포함한다.
- [0026] 본 발명에 있어서,
- [0027] 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성은 -12256G>C로 표시한다.
- [0028] 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -11917번째) G가 A로 치환되는 단일염기다형성은 -11917G>A로 표시한다.
- [0029] 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성은 -11849C>T로 표시한다.
- [0030] 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성은 +441G>C로 표시한다.
- [0031] 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성은 +2486C>T로 표시한다.
- [0032] 본 발명에서 폐암 발생 위험도를 예측하기 위한 표지 인자로 선택된 인간 염색체 17p13.1에 위치하는 TP53 유전자는 Gene Bank 등재번호 NT_010718로 공지되어 있으며, 본 발명에서는 서열 중 -7512445~ +7531642 서열을 서열번호 21에 나타내었다.
- [0033] 염색체 17p13.1 상의 TP53 유전자 다형성은 단일 염기 다형성이 위치하는 염색체 번호, 염색체 상의 동원체를 중심으로 한 단완 (short arm: p)부분 및 밴드 번호를 알 수 있는바, 이는 17번 염색체 상의 단완 부분의 13.1 밴드 상에 존재함을 의미한다.
- [0034] 본 발명에서의 용어 "다형성(polymorphism)"은 유전학적으로 결정된 집단 내에서 2이상의 대체적 서열 또는 대체적 대립형질의 발생을 의미한다. 바람직한 다형성 마커는 선택된 집단에서 1% 이상, 더욱 바람직하기로는 10% 또는 20% 이상의 발생빈도를 나타내는 2개의 대립형질을 의미한다. 다형성 중에서, 사람에 따라 단일 염기만이 다른 것을 단일 염기 다형성(SNP)이라 한다.
- [0035] 본 발명에서의 용어 "핵산시료"는 DNA(gDNA, cDNA) 그리고 RNA 분자를 포괄적으로 포함하는 의미를 갖는다.
- [0036] 본 발명의 방법에서 출발물질이 gDNA인 경우, gDNA의 분리는 당업계에 공지된 통상의 방법에 따라 실시될 수 있다(참조: Rogers & Bendich (1994)).출발물질이 mRNA인 경우에는, 당업계에 공지된 통상의 방법에 총 RNA를 분리하여 실시된다(참조: Sambrook, J. et al., Molecular Cloning. A Laboratory Manual, 3rd ed. Cold Spring Harbor Press(2001); Tesniere, C.et al., Plant Mol. Biol. Rep., 9:242(1991); Ausubel, F.M. et al., Current Protocols in Molecular Biology, John Willey & Sons(1987); 및 Chomczynski, P. et al., Anal. Biochem. 162:156(1987)). 분리된 총 RNA는 역전사효소를 이용하여 cDNA로 합성한다.
- [0037] 본 발명의 방법에 있어서, 피검체로부터 DNA를 분리하는 방법은 당업계에 알려진 방법을 통하여 이루어질 수 있다. 예를 들면, 조직 또는 세포로부터 DNA를 직접적으로 정제하거나 PCR과 같은 증폭 방법을 사용하여 특정한 영역을 특이적으로 증폭하고 이를 분리함으로써 이루어질 수 있다. 본 발명에 있어서, DNA란 DNA 뿐만 아니라 mRNA로부터 합성되는 cDNA도 포함하는 의미이다. 피검체로부터의 핵산 수득은 예를 들면, PCR 증폭법, 리가제 연쇄 반응(LCR)(Wu 및 Wallace, Genomics 4, 560(1989), Landegren 등, Science 241, 1077(1988)), 전사증폭(transcription amplification)(Kwoh 등, Proc. Natl. Acad. Sci. USA 86, 1173(1989)) 및 자가유지 서열 복제(Guatelli 등, Proc. Natl. Acad. Sci.USA 87, 1874(1990)) 및 핵산에 근거한 서열 증폭(NASBA)이 사용될 수 있다.
- [0038] 본 발명의 방법 1) 단계의 다형성 빈도를 측정하는 단계는 당업계에 공지된 다양한 방법을 응용하여 수행될 수 있다. 예를 들어, DNA 서열분석(sequencing), PCR-SSCP(single stranded conformation polymorphism) 분석, RFLP(restriction fragment length polymorphism) 분석, 대립형질(allele) 특이적 혼성화 방법, DNA 칩 방법, ELISA(enzyme-linked immunosorbent assay) 방법 및 면역블롯(immunoblot) 방법, 스냅샷(sNaPShot) 분석 방법, TDGS(Two-dimensional Gene Scanning), Taq-Man[®] 및 TOGATM으로 구성된 군으로부터 선택된 하나 이상

의 방법에 의해 측정될 수 있으나, 이에 한정되는 것은 아니다.

- [0039] 다형성 부위의 증폭은 당업계에 공지된 다양한 방법을 응용하여 실시될 수 있다. 예를 들어, 본 발명에 응용될 수 있는 기술은, 본 발명의 방법에 이용될 수 있는 증폭 기술은 PCR증폭 (참조: Miller, H. I. (WO 89/06700) 및 Davey, C. et al. (EP 329,822)), 리가아제 체인 반응 (LCR, Wu, D.Y. et al., Genomics 4:560 (1989)), 중합효소 리가아제 체인 반응 (Barany, PCR Methods and Applic., 1:5-16(1991)), Gap-LCR (WO 90/01069), 리 페어 체인 반응 (EP 439,182), 3SR (Kwoh et al., PNAS, USA, 86:1173(1989)) 및 NASBA (U.S. Pat. No. 5,130,238)을 포함하나, 이에 한정되는 것은 아니다. 가장 바람직하게는 PCR 증폭에 따라 증폭한다.
- [0040] 본 발명에서의 용어 "다형성 빈도"는 피검체(정상군과 환자군을 포함한다)에서 분리된 핵산시료를 분석하였을 때 피검체의 단일염기다형성 부위에서 치환된 염기가 검출되는 횟수를 말한다.
- [0041] 분리된 DNA의 염기서열의 분석은 당업계에 알려진 다양한 방법에 의하여 이루어질 수 있다. 예를 들면, 디데옥 시 법에 의하여 직접적인 핵산의 뉴클레오타이드 서열의 결정을 통하여 이루어지거나 단일 염기 다형성 부위의 서 열을 포함하는 프로브 또는 그에 상보적인 프로브를 상기 DNA와 혼성화시키고 그로부터 얻어지는 혼성화 정도를 측정함으로써 다형성 부위의 뉴클레오타이드 서열을 결정하는 방법 등이 이용될 수 있다. 혼성화의 정도는 예를 들면, 검출 가능한 표지를 표적 DNA에 표지하여, 혼성화된 표적 DNA 만을 특이적으로 검출함으로써 이루어질 수 있으나, 그외 전기적 신호 검출방법 등이 사용될 수 있다.
- [0042] 본 발명의 방법 2) 단계에서 상관관계는 당업계에 공지된 다양한 방법을 응용하여 수득할 수 있다. 예를 들어, 스튜던트 t -검정 (Student's t-test), 카이-스퀘어 검정 (Chi-square test) 및 다중 기호 회귀 분석 (multiple logistic regression analysis)으로 구성된 군으로부터 선택된 하나 이상의 방법에 의해 수득할 수 있으나, 이에 한정되는 것은 아니다.
- [0043] 예를 들어, 폐암 발생의 위험도 분석에 있어서 로지스틱회귀분석을 통해 구한 오즈비와 95% 신뢰구간의 결과가 폐암 발생의 위험도를 분석하는 기준이 된다.
- [0044] 상기 오즈비는 대조군에서 해당 단일 염기 다형성/일배체형을 가질 확률에 대한 환자군 중에서 해당 단일 염기 다형성/일배체형을 가질 확률의 비율을 나타낸다. 오즈비가 1을 초과하는 경우 해당 단일 염기 다형성/일배체형 이 환자군과 폐암 발생 여부와 연관성이 있음을 나타낸다. 오즈비가 클수록 상기 연관성도 증가한다.
- [0045] 상기 95% 신뢰구간은 해당 오즈비가 존재할 확률이 95%인 구간을 나타내는 것이다. 즉, 신뢰구간의 하계가 1 이 하이고 상계가 1 이상인 경우에는 질병과의 연관성을 유의하다고 판단할 수 없다.
- [0046] $p\text{-value} < 0.05$ 일 때 폐암과의 연관성이 유의하다고 판단할 수 있다. 즉, 위험도 오즈비가 유의한 것으로 판단할 수 있다.
- [0047] 24명의 한국인을 대상으로 TP53의 단일염기다형성을 분석한 결과 13개의 단일염기다형성을 발견하였으며, 이중 5개인 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 에 대해서만 환자군과 정상군을 포함한 총 595 명의 대상군에서 유전자형분석을 수행하였다.
- [0048] 본 발명자들은 로지스틱 회귀 분석을 통하여 연령, 성별 및 흡연 여부를 보정한 환자군 및 정상군의 TP53 유전 자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 다형성 빈도를 구하고, 폐암 발생 위험도 간의 상관관계를 분석하였다.
- [0049] TP53의 단일 염기 다형성(-12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T)의 분포와 폐암 발생 위험 도 간의 상관관계를 알아본 결과, 폐암 발생 위험도에 있어서 환자군과 정상군 간에 거의 차이를 나타내지 않았 다.
- [0050] 또한, 본 발명에서는 상기 5개의 다형성 간에 높은 연관 비평형(Linkage disequilibrium, LD) 을 나타나기 때문 에, 일배체형 3(-12256G, -11917G, -11849C, +441C 및 +2486C), 일배체형 1(-12256G, -11917G, -11849C, +441G, +2486C) 및 일배체형 2 (-12256C, -11917G, -11849C, +441C, +2486T) 와 폐암 발생 위험도 간의 상관관 계도 분석하였다. 구체적으로, 각 다형성의 일배체형에 대한 오즈비(ORs) 및 이들의 95% 신뢰구간(CI)을 일배체 형을 비교함으로써 폐암 위험도와의 연관성을 평가하였다.
- [0051] 일배체형 1(-12256G, -11917G, -11849C, +441G, +2486C) 및 일배체형 2 (-12256C, -11917G, -11849C, +441C, +2486T) 와는 달리 일배체형 3(-12256G, -11917G, -11849C, +441C 및 +2486C)은 다른 일배체형들에 비하여 폐 암 발생 위험도를 증가시켰다. codominant의 경우 오즈비=1.63, 95% CI= 1.11-2.40, P=0.01을 나타냈고,

dominant의 경우 오즈비= 1.80, 95% CI= 1.17-2.77, P=0.007을 나타냈다.

- [0052] 이에, 대상군을 보다 하위그룹으로 세분화하여 하위그룹에서의 일배체형 3의 빈도를 조사하여 일배체형 3과 폐암 발생 위험도 간의 상관관계를 분석하였다. 하위그룹은 성별, 흡연여부 및 세포 타입에 따라 분류하였다.
- [0053] 특히, 하위그룹 중 남성, 흡연자, 선암 세포 환자군에게서 일배체형 3과 폐암 발생 위험도 간의 상관관계가 유의하게 나타났다. 일배체형 3을 가진 남성, 흡연자, 선암 세포 환자군은 폐암 발생 위험도가 증가하였다.(P-값<0.05)
- [0054] 남성에게서 일배체형 3과 폐암 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, -/ht3(일배체형 3)의 유전자형의 경우 analyzing models referent에서는 오즈비=1.93, 95% 신뢰구간=1.12-3.32, P-값=0.02 및 dominant에서는 오즈비=1.79, 95% 신뢰구간=1.06-3.03, P-값=0.03을 나타냈다.
- [0055] 또한, 흡연자에게서 일배체형 3과 흡연 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, 흡연자에서 -/ht3(일배체형3)의 유전자형의 경우 analyzing models referent에서는 오즈비=2.03, 95% 신뢰구간=1.15-3.57, P-값=0.01 및 dominant에서는 오즈비=1.86, 95% 신뢰구간=1.08-3.22, P-값=0.03을 나타냈다.
- [0056] 또한, 세포 타입 중 선암에게서 일배체형 3과 흡연 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, -/ht3(일배체형3)의 유전자형의 경우 analyzing models referent에서는 오즈비=2.44, 95% 신뢰구간=1.21-4.92, P-값=0.01, codominant에서는 오즈비=2.18, 95% 신뢰구간=1.18-4.00, P-값=0.01 및 dominant에서는 오즈비=2.46, 95% 신뢰구간=1.24-4.87, P-값=0.01을 나타냈다.(표4 참조)
- [0057] 본 발명의 다른 양태에 따르면, 본 발명은 상기 방법을 이용하여 한국인에서 폐암의 유전적 위험 인자를 검사하는 진단 키트를 제공한다.
- [0058] 본 발명의 키트는 하기 기술할 프라이머를 필수 구성요소로 하는바, 프라이머에 대한 상세한 설명은 본 발명의 키트에 이용되는 프라이머에도 적용된다. 따라서, 반복적인 기재에 따른 본 명세서의 과도한 복잡성을 피하기 위해 그 기재를 생략한다.
- [0059] 본 발명의 키트가 PCR 증폭반응에 적용되는 경우, 예컨대 본 발명의 키트는 선택적으로, PCR 증폭에 필요한 시약, 예컨대, 완충액, DNA 중합효소 (예컨대, *Thermus aquaticus* (Taq), *Thermus thermophilus* (Tth), *Thermusfiliformis*, *Thermis flavus*, *Thermococcus literalis* 또는 *Pyrococcus furiosus* (Pfu)로부터 수득한 열 안정성 DNA 중합효소), DNA 중합 효소 조인자 및 dNTPs를 포함할 수 있다. 본 발명의 키트는 상기한 시약 성분을 포함하는 다수의 별도 패키징 또는 컴파트먼트로 제작될 수 있다.
- [0060] 또한, 본 발명의 다른 양태에 따르면 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -1197번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위를 증폭시킬 수 있는 프라이머쌍을 제공한다.
- [0061] 본 발명에서 TP53 유전자의 폐암 발암 관련 SNP를 탐색하기 위해서는 프라이머 디자인이 중요하다. 당업자가 디자인할 수 있는 모든 프라이머가 본 발명의 프라이머 범위에 포함될 수 있다.
- [0062] 바람직하게는, 상기 프라이머쌍 중에서 전방향 프라이머 역방향 프라이머는 SNPs염기에 인접한 서열과 어닐링할 수 있는 뉴클레오티드 서열을 포함하여, 상기 전방향 또는 역방향 프라이머 3' 말단은 상기 SNPs염기에 매칭되는 염기를 갖는다.
- [0063] 예를 들어, 750번째 뉴클레오티드에 있는 SNP인 "C"를 검출하여 TP53 유전자 다형성을 동정하는 경우, 상기 SNP의 업스트림에 상보적인 서열을 전방향 프라이머로서 이용할 수 있으며, 역방향 프라이머 3' 말단은 상기 SNP에 매칭되는 염기 즉, "C" 염기를 가지며, 나머지 부분은 서열번호 21에서 750번째 뉴클레오티드의 다운스트림에 위치한 서열과 실질적으로 대응되는 서열을 갖는다. 이러한 프라이머쌍을 이용하여 750번째 SNP를 검출하는 증폭 반응을 실시할 수 있다.
- [0064] 본 발명의 바람직한 구현예에 따르면, 본 발명의 프라이머쌍은 서열번호:1 및 2로 기재되는 염기서열을 갖는 올리고뉴클레오티드, 서열번호:3 및 4로 기재되는 염기서열을 갖는 올리고뉴클레오티드, 서열번호:5 및 6로 기재

되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:7 및 8로 기재되는 염기서열을 갖는 올리고뉴클레오타이드 및 서열번호:9 및 10로 기재되는 염기서열을 갖는 올리고뉴클레오타이드로 이루어진 군에서 선택된 어느 하나일 수 있다.

[0065]

Loci	primer	Sequence
-12256G>C	Forward	CGATTCGAGGACTCAACTCACTAG ; 서열번호 1
	Reverse	GGCTGGCTCAGCTGTGT ; 서열번호 2
-11917G>A	Forward	CTGTGCTTCCCTCCTCTTCTG ; 서열번호 3
	Reverse	GCTAAGGAACACAGTGCTTTCAAAA ; 서열번호 4
-11849C>T	Forward	TGAAAGCACTGTGTTCTTAGCA ; 서열번호 5
	Reverse	CGGAAGGTGGACCGAAATCC ; 서열번호 6
+441G>C	Forward	CCAGATGAAGCTCCAGAATGC ; 서열번호 7
	Reverse	GCCGCCGGTGTAGGA ; 서열번호 8
+2486C>T	Forward	GCCTCTGCTTGCCTCTGA ; 서열번호 9
	Reverse	GGTGGATGGGTAGTAGTATGGAAGA ; 서열번호 10

[0066]

본 발명에서의 용어 "프라이머"란 적절한 버퍼 중의 적절한 조건(예를 들면, 4개의 다른 뉴클레오타이드 트리포스페이트 및 DNA, RNA 폴리머라제 또는 역전사 효소와 같은 중합체) 및 적당한 온도 하에서 주형-지시 DNA 합성의 시작점으로서 작용할 수 있는 단일가닥 올리고뉴클레오타이드를 말한다.

[0067]

상기 프라이머의 적절한 길이는 사용목적에 따라 달라질 수 있으나, 통상 15 내지 30 뉴클레오타이드이다 짧은 프라이머 분자는 일반적으로 주형과 안정한 혼성체를 형성하기 위해서는 더 낮은 온도를 필요로 한다. 프라이머 서열은 그 3'말단이 TP53의 SNP부위와 정렬하는 것이 바람직하다. 상기 프라이머는 SNP 부위를 포함하는 표적 DNA에 혼성화하고, 상기 프라이머가 완전한 상동성을 보이는 대립형질 형태의 증폭을 개시한다. 이 프라이머는 반대편에 혼성화하는 제2프라이머와 쌍을 이루어 사용된다. 증폭에 의하여 두개의 프라이머로부터 산물이 증폭되고, 이는 특정 대립형질 행태가 존재한다는 것을 의미한다.

[0068]

또한, 본 발명의 다른 양태에 따르면 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오타이드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되는 단일염기다형성, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -1197번째) G가 A로 치환되는 단일염기다형성, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되는 단일염기다형성, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되는 단일염기다형성 및 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되는 단일염기다형성으로 이루어진 군에서 선택된 하나 이상의 다형성 부위에 혼성화할 수 있는 프로브쌍을 제공한다. 당업자가 디자

인할 수 있는 모든 프로브가 본 발명의 프로브 범위에 포함될 수 있다.

[0069] 본 발명에서의 용어 "프로브"는 특정 뉴클레오타이드 서열에 혼성화될 수 있는 디옥시뉴클레오타이드 및 리보뉴클레오타이드를 포함하는 자연 또는 변형되는 모노머 또는 결합을 갖는 성형의 올리고머를 의미한다. 바람직하게는, 프로브는 혼성화에서의 최대 효율을 위하여 단일가닥일 수 있다. 프로브는 바람직하게는 디옥시리보뉴클레오타이드일 수 있다. 본 발명에 이용되는 프로브로서, 본 발명의 SNP에 완전하게(perfectly) 상보적인 서열이 이용될 수 있으나, 특이적 혼성화를 방해하지 않는 범위 내에서 실질적으로(substantially) 상보적인 서열이 될 수도 있다. 본 발명에 이용되는 프로브는 TP53 유전자 염기서열 중 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T을 적어도 하나 이상 포함하는 DNA 단편에 혼성화될 수 있는 서열을 포함한다. 보다 바람직하게는, 상기 프로브의 3' 말단, 5' 말단은 상기 SNPs 염기에 상보적인 염기를 갖는다. 일반적으로, 혼성화에 의해 형성되는 듀플렉스(deuplex)의 안정성은 말단의 서열의 일치에 의해 결정되는 경향이 있기 때문에, 3' 말단, 5' 말단에 SNP 염기에 상보적인 염기를 갖는 프로브에서 말단 부분이 혼성화되지 않으면, 이러한 듀플렉스는 해체될 수도 있다.

[0070] 혼성화에 적합한 조건은 Joseph Sambrook, et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.(2001) 및 Haymes, B. D., et al., Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985)에 개시된 사항을 참조하여 결정할 수 있다. 혼성화에 이용되는 조건은 온도, 이온세기 (완충액 농도) 및 유기 용매와 같은 화합물의 존재 등을 조절하여 결정될 수 있다. 이러한 조건은 혼성화되는 서열에 의존하여 다르게 결정될 수 있다.

[0071] 본원발명의 바람직한 구현예에 따르면, 폐암 발생 위험도를 측정하는 방법에 있어서, 다형성 빈도를 측정하는 단계에서 바람직하게는 TaqMan[®] 사용하여 다형성의 빈도를 측정하도록 할 수 있다. TaqMan[®] probe는 5' 말단은 형광물질(FAM등)로 3' 말단은 quencher 물질(TAMRA등)로 수식한 oligonucleotide(TaqMan[™] probe)를 말한다. TaqMan[™] probe를 PCR반응용액에 첨가하면 annealing step에서 template DNA에 특이적으로 혼성화하지만 프로브 상에 quencher에 의해 형광 발색은 억제된다. extension step에서 Taq DNA polymerase가 같은 5'→3' exonuclease활성으로 template에 hybridize한 TaqMan[™] probe가 분해되어 형광색소가 프로브에 유리되면서 quencher에 의한 억제가 해제되어 형광을 띄게 된다.

[0072] 본 발명의 바람직한 구현예에 따르면, 본 발명의 프로브쌍은 서열번호:11 및 12로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:13 및 14로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:15 및 16으로 기재되는 염기서열을 갖는 올리고뉴클레오타이드, 서열번호:17 및 18로 기재되는 염기서열을 갖는 올리고뉴클레오타이드 및 서열번호:19 및 20으로 기재되는 염기서열을 갖는 올리고뉴클레오타이드로 이루어진 군에서 선택된 하나일 수 있다.

[0073]

Loci		Sequence
-12256G>C	Forward	CGATTCGAGGACTCAACTCACTAG ; 서열번호 1
	Reverse	GGCTGGCTCAGCTGTGT ; 서열번호 2
	VIC	CCTCCCGTAGCTCC ; 서열번호 11
	FAM	CCTCCCTAGCTCC ; 서열번호 12
-11917G>A	Forward	CTGTGCTTCCCTCTCTCTG ; 서열번호 3
	Reverse	GCTAAGGAACACAGTGCTTTCAAAA ; 서열번호 4
	VIC	CTGTTGCGCTCTCCT ; 서열번호 13
	FAM	CTGTTGCTTCTCCT ; 서열번호 14
-11849C>T	Forward	TGAAAGCACTGTGTTCTTAGCA ; 서열번호 5
	Reverse	CGGAAGGTGGACCGAAATCC ; 서열번호 6

	VIC	CAGCAAGAGGCCCCG ; 서열번호 15
	FAM	CAGCAAAAGGCCCCG ; 서열번호 16
+441G>C	Forward	CCAGATGAAGCTCCCAGAATGC ; 서열번호 7
	Reverse	GCCGCCGGTGTAGGA ; 서열번호 8
	VIC	TCCCCGCGTGGCC ; 서열번호 17
	FAM	CTCCCCCGTGGCC ; 서열번호 18
+2486C>T	Forward	GCCTCTGCTGCCTCTGA ; 서열번호 9
	Reverse	GGTGGATGGGTAGTAGTATGAAGA ; 서열번호 10
	VIC	CCCTGGGCCACCT ; 서열번호 19
	FAM	CCCCTGGGTCCACCT ; 서열번호 20

[0074] 또한, 본 발명의 다른 양태에 따르면, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째 염기(TP53 유전자의 번역시작점으로부터 -12256번째) G가 C로 치환되고 상기 750번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1089번째 염기(TP53 유전자의 번역시작점으로부터 -1197번째) G가 A로 치환되고 상기 1089번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 1157번째 염기(TP53 유전자의 번역시작점으로부터 -11849번째) C가 T로 치환되고 상기 1157번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 13446번째 염기(TP53 유전자의 번역시작점으로부터 +441번째) G가 C로 치환되고 상기 13446번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드, 서열번호 21의 15491번째 염기(TP53 유전자의 번역시작점으로부터 +2486번째) C가 T로 치환되고 상기 15491번째 염기를 포함하는 10-700개의 연속적인 DNA단편으로 구성되는 폴리뉴클레오티드 및 이들의 상보적인 폴리뉴클레오티드로 이루어진 군 중에서 선택된 하나 이상의 폴리뉴클레오티드로 이루어지는 폐암 감수성 진단용 마커를 제공한다.

[0075] 본 발명에서 용어, "마커"는 게놈에 있는 특정 서열로 특정 형질 다형성(polymorphism)을 가리키는 표지로서 이용될 수 있는 것을 의미한다.

[0076] 본 발명에서 용어, "뉴클레오티드"는 단일 가닥 또는 이중 가닥 형태로 존재하는 디옥시리보뉴클레오티드 또는 리보뉴클레오티드이며, 다르게 특별하게 언급되어 있지 않은 한 자연의 뉴클레오티드의 유사체를 포함한다.(Scheit, Nucleotide Analogs, John Wiley, New York(1980); Uhlman 및 Peyman, Chemical Reviews, 90:543-584(1990))

[0077] DNA 단편의 크기는 전체 유전자의 full length가 아닌 한 다형성 부위의 염기를 포함하는 어떤 크기의 단편이라도 무방하나, 바람직하게는 10 내지 수백 염기인 것을 특징으로 하며, 더욱 바람직하게는 10-700개 염기인 것을 특징으로 한다.

[0078] 10-30개 염기의 경우는 상기 SNP를 탐지하기 위한 프로브 내지 프라이머로 이용할 수 있으며, 그 이상의 염기의 경우는 PCR-RFLP등에 이용될 수 있다.

[0079] 또한, 본 발명의 다른 양태에 따르면, 본 발명은 서열번호 21로 이루어지는 폴리뉴클레오티드에서, 서열번호 21의 750번째(TP53 유전자의 번역시작점으로부터 -12256번째) 염기가 G이고, 서열번호 21의 1089번째(TP53 유전자의 번역시작점으로부터 -11917번째) 염기가 G이고, 서열번호 21의 1157번째(TP53 유전자의 번역시작점으로부터 -11849번째) 염기가 C이고, 서열번호 21의 13446번째(TP53 유전자의 번역시작점으로부터 +441번째) 염기가 C이고, 서열번호 21의 15491번째(TP53 유전자의 번역시작점으로부터 +2486번째) 염기가 C로 일배체형(haplotype)을 이루는 것을 특징으로 하는 감수성 진단용 마커를 제공한다.

[0080] 본 발명에서 용어 "일배체형(haplotype)"은 개체로부터 단일 유전자에 대한 복수의 다형성 및 개체마다 상이한 다수의 다형성 부위에서 변이의 특정 조합을 말한다. 상기 일배체형은 각각의 단일염기다형성 부위를 포함하는 폴리뉴클레오티드 단편들의 조합 또는 복수의 단일염기다형성 부위를 포함하는 폴리뉴클레오티드 단편으로 이루어질 수 있다. 바람직하게는, 전자의 경우에는 각 단편이 10-700개의 염기를 포함할 수 있고 후자의 경우에는 수백 개 이상의 염기를 포함할 수 있다.

[0081] 최근 연구들에서는 질병과의 연관성을 예측하는데 있어서 단일 염기 다형성 분석보다 일배체형 분석을 보다 상위에 두고 있다. (Sun, T., et al., J. Natl. Cancer Inst., 96, 1030-1036, 2004; Zhang, X., et al., J. Med. Genet., 42, 479-484, 2005; Judson, R., et al., Pharmacogenomics, 1, 15-26, 2000; 및 Judson, R. 및 Stephens, J. C., Pharmacogenomics, 2, 7-10, 2001).

[0082] 이하, 실시예를 통하여 본 발명을 더욱 상세히 설명하고자 한다. 이들 실시예는 오로지 본 발명을 보다 구체적으로 설명하기 위한 것으로, 본 발명의 요지에 따라 본 발명의 범위가 이들 실시예에 의해 제한되지 않는다는 것은 당업계에서 통상의 지식을 가진 자에 있어서 자명할 것이다.

효 과

[0083] 이상, 상기에서 살펴본 바와 같이 본 발명 TP53 유전자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 다형성의 빈도를 측정하여 한국인의 발암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트는 암 억제 유전자 TP53의 단일염기다형성 및 이의 일배체형과 한국인 폐암 발생과의 연관성을 확인하여 TP53이 폐암의 유전적인 위험 인자로서, 폐암의 치료 진단 표지 인자로 사용 가능한지를 확인하는데 사용될 수 있다.

발명의 실시를 위한 구체적인 내용

[0084] 실시예 1. 피검자의 선정 및 측정

[0085] 본 연구는 2001년 10월부터 2004년 4월까지 폐암 및 유방암, 난소암 유전체센터/고려대 병원 (안암, 안산, 구로)에서 폐암 진단받은 환자군 299명과 심혈관 유전체센터, 알리지/호흡계질환 유전체센터와 계명대 동산 병원에서 모집된 정상군 296명에서 동의서를 받아 진행되었다. 정상군은 소변검사, 혈액검사, 가슴 방사선 사진 등을 포함하는 물리적 검사, 과거 병력, 실험실 검사를 거쳐 선택하였으며 식이, 흡연, 음주와 과거병력등을 인터뷰를 통해 확인하였다. 본 실험의 목적 달성을 위하여 이전에 적어도 하루에 한번 흡연을 한 사람을 흡연자로 보았다. 본 연구는 해당기관의 생명윤리 심의위원회 심의를 통과하여 진행되었다. 이러한 환자군과 정상군의 임상학적 특징은 표1에 정리되어 있다.

[0086]

[0087] <표 1> 폐암 환자 및 정상 대조군의 임상학적 특징

[0088]

특징	대조군	실험군	p-value
집단의 크기	n = 296	n = 299	
평균 연령 (평균±표준편차)	51.8±3.2	62.3±0.3	<0.0001
성별, n			
남성	283	219	<0.0001
여성	16	80	
흡연, n			
비흡연자	90	76	0.009
흡연자	204	204	
불분명한 자	5	19	

[0089] 실시예 2. 염기서열 분석(sequence analysis)

[0090] 24명의 한국인 DNA샘플을 대상으로 ABI prism 3700 DNA analyzer(Foster City, CA)를 이용하여 TP53의 프로모터 부위, 엑손-인트론 경계 부위 및 모든 엑손(exon)을 포함하는 모든 부위의 서열 분석을 하였다. 증폭 및 서열 분석을 위한 모든 프라이머는 GenBank sequence(OMIM No. 191170)를 기초로 하여 디자인하였다. 서열변이는 크로마토그램에 의해 확인하였다.

[0091] 그 결과 24명의 한국인을 대상으로 TP53의 단일염기다형성을 분석한 결과 13개의 단일염기다형성을 발견하였으며, 위치, 빈도(7.5%초과), LD, 일배체 태깅 상태를 기초로 하여 13개 중에서 5개인 -12256G>C, -11917G>A,

-11849C>T, +441G>C 및 +2486C>T 만 환자군과 정상군을 포함한 총 595명의 대상군에서의 유전자형분석 대상으로 선택하였다. 정상군과 환자군 모두의 유전자형 빈도는 HWE상태에 있었다. 결과는 하기 표2에서 나타난 바와 같다.

Table 2. Frequencies of TP53 polymorphisms in Korean population

Loci	Position	Amino acid change	rs#	Allele Frequency	Heterozygosity	HWE ^a	
						Case	Control
-12648G>A	promoter	.	rs2287498	0.208	0.330		0.999
-12570G>A	promoter	.	novel	0.042	0.080		0.978
-12256G>C	Promoter	.	rs2287499	0.232	0.356	0.864	0.994
-11917G>A	Promoter	.	novel	0.015	0.030	0.985	0.939
-11849C>T	Promoter	.	rs17806770	0.047	0.090	0.987	0.743
+88T>C	intron	.	novel	0.022	0.043		0.994
+112G>C	intron	.	rs1642785	0.326	0.440		0.914
+294C>A	intron	.	novel	0.065	0.122		0.946
+441G>C	Exon4	Arg72Pro	rs1042522	0.368	0.465	0.805	0.994
+1630C>T	Exon5	Ala189Val	novel	0.021	0.041		0.995
+2486C>T	Intron7	.	rs12947788	0.270	0.394	0.643	1.000
+2506T>G	intron	.	rs12951053	0.261	0.386		0.895
+7399C>A	3'utr	.	novel	0.063	0.117		0.948

^a p-values of deviation from lung cancer and normal subjects

Bold face means SNPs genotyped in a larger-population (n=595) and plain faces were based on the sequencing data (n=24)

진한 글씨체는 표본 집단이 환자군과 정상군을 포함한 595명을 대상으로 한 단일 염기 다형성을 분석한 결과이고, 연한 글씨체는 이보다 작은 크기의 표본 집단인 환자군과 정상군을 포함한 24명을 대상으로 단일 염기 다형성을 분석한 결과이다.

상기 표에서 UTR은 비번역 구간을 의미한다.

실시예 3. 유전자형 분석

3-1. 피검자의 genomic DNA 준비

실험 수행을 위해 동의를 받은 환자군과 정상군을 포함한 총 595명의 대상군에서 채취한 혈액에서 Puregene DNA purification system을 이용해 게놈 DNA (genomic DNA; gDNA)를 추출하였다.

3-2. 표적 DNA의 증폭 및 SNP 분석

유전자형 분석을 통한 다형성 빈도 측정은 Taq-Man[®]으로 수행하였다.

TP53의 다형성 부위의 유전자형분석을 위한 모든 프라이머(primer)와 프로브(probe)는 TaqMan system 분석용으로 Primer Express (Applied Biosystems)을 이용해 디자인하였으며 프로브는 형광물질, FAM과 VIC으로 표시하였다. PCR은 900nM PCR 프라이머와 200nM의 TaqMan MGB 프로브 존재 하에서 UNG (Applied Biosystems)없이 TaqMan Universal Master Mix를 이용하여 실시하였다. 이 반응은 20ng의 genomic DNA를 전체 5ul의 volume으로 384 well에서 수행하였다. PCR이 끝난 후에 Prism 7900HT instrument로 옮겨져서 fluoresce intensity를 측정하여 유전자형분석을 수행하였다. The reactions were carried out in a 384-well format, with a total reaction volume of 5 ul using 20 ng of genomic DNA. The plates를 thermal cycler (PE 9700, Applied Biosystems) 안에 넣고 and heated at 50 22분 동안 50℃로 가열하고 10분 동안 95℃로 가열한 후 95℃에서 15초간 회전 후 다시 60?에서 1분간 회전시켰다. The TaqMan assay plates를 각 웰의 형광도를 분석하기 위해 a Prism 7900HT instrument (Applied Biosystems)로 옮겼다. automated software (SDS 2.1)를 이용해 형광도 데이터를 분석하였다.

<표 3> Sequences of amplifying, Taqman probe for TP53 polymorphisms genotyping

Loci		Sequence
-12256G>C	Forward	CGATTCGAGGACTCAACTCACTAG
	Reverse	GGCTGGCTCAGCTGTGT
	VIC	CCTCCCGTAGCTCC
	FAM	CCTCCCGTAGCTCC
-11917G>A	Forward	CTGTGCTTCCCTCCTCTTCTG
	Reverse	GCTAAGGAACACAGTGCTTCAAAA
	VIC	CTGTTCGCCTCTCCT
	FAM	CTGTTCGCCTCTCCT
-11849C>T	Forward	TGAAAGCACTGTGTTCTTAGCA
	Reverse	CGGAAGGTGGACCGAAATCC
	VIC	CAGCAAGAGGCCCG
	FAM	CAGCAAAAGGCCCG
+441G>C	Forward	CCAGATGAAGTCCCAGAATGC
	Reverse	GCCGCCGGTGTAGGA
	VIC	TCCCCGCGTGGCC
	FAM	CTCCCCCGTGGCC
+2486C>T	Forward	GCCTCTGCTTGCTCTGA
	Reverse	GGTGGATGGGTAGTAGTATGAAGA
	VIC	CCCTGGGCCACCT
	FAM	CCCCTGGGTCCACCT

3-3. 통계분석 (statistical analysis)

카이제곱 분석(chi-square test)을 통해 개별 변이가 군집 내 각 좌위에서 HWE 상태에 있는지를 결정했다. HWE 상태는 하아디-와인버그 평형 (Hardy- Weinberg equilibrium)의 상태를 나타내는 것으로 정상군과 환자군의 하아디-와인버그 평형 여부를 나타내는 것이다. 두 개의 대립 형질이 환자군과 정상군에서 나타나는 빈도가 통계적으로 유의한지를 분석하기 위해 사용하였다. 로지스틱 회귀 분석 (logistic regressions)을 통해 나이, 성별 및 흡연 유무를 공변수로서 보정하여 p-값을 산출했다. 각 locus의 대립형질 빈도 (p 및 q=1-p)에 대한 heterozygosity(H)을 구하기 위한 계산식은 $H=1-p_2-q_2=2p(1-p)$ 이다. 또한 모든 대립형질사이에 연관 불균형 (linkage disequilibrium), 르윈턴 D' (leuontin's D'(|D'|)) 및 r^2 을 측정하여 단일염기다형성간의 일배체형 (haplotype)과 빈도는 stephen에 의해서 제시된 알고리즘 이용하여 분석하였다. 로지스틱 회귀 분석은 연령, 성별, 흡연 상태를 공변수로 보정된 오즈비, 95% 신뢰구간, 대응 P-값을 계산하기 위해 사용하였다. 서로 연관비평형 상태에 있는 단일 염기 다형성의 다중 검정에 따라 생기는 오차를 적절히 보정 하기 위해 TP53의 독립 마커 좌위의 개수를 the spectral decomposition (SpD)에 기초하여 the software SNPSpD (<http://genepi.qimr.edu.au/general/daleN/SNPSpD/>)을 사용하여 계산하고 보정된 P^{corr*} -값을 산출하였다. Statistical power는 Statistical Power Calculator(http://www.dssresearch.com/toolkit/spc및/power_p2.asp)를 통해서 산출하였다.

실시예 4. TP53 다형성 및 일배체형과 폐암 발생 위험도 간의 연관성 판단

본 발명자들은 로지스틱 회귀 분석을 통하여 연령, 성별 및 흡연 여부를 보정한 전, 후 환자군 및 정상군의 TP53 유전자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 다형성 빈도를 구하고, 폐암 발생 위험도 간의 상관관계를 분석하였다. 또한, 본 발명에서는 상기 5개의 다형성 간에 높은 연관 비평형(Linkage disequilibrium, LD)을 나타내기 때문에, 일배체형 3(-12256G, -11917G, -11849C, +441C 및 +2486C), 일배체형 1 및 일배체형2 와 폐암 발생 위험도 간의 상관관계도 분석하였다. 구체적으로, 각 다형성의 일배체형에 대한 오즈비(ORs) 및 이들의 95% 신뢰구간(CI)을 일배체형을 비교함으로써 폐암 위험도와의 연관성을 평가하였다. 하기 표 4에 그 결과를 나타내었다.

[0107] <표 4>

Loci	Genotype Case	Control	Codominant				Dominant				Recessive				Statistical power** (%)				
			Crude OR (95% CI)	Crude P	Adjusted OR (95% CI)	P	P ^{***}	Crude OR (95% CI)	Crude P	Adjusted OR (95% CI)	P	P ^{***}	Crude OR (95% CI)	Crude P		Adjusted OR (95% CI)	P	P ^{***}	
-12256G > C	G	199(0.61)	169(0.58)	0.92 (0.70 - 1.21)	0.56	0.87 (0.63 - 1.21)	0.42	NS	0.89 (0.64 - 1.24)	0.48	0.85 (0.57 - 1.27)	0.42	NS	0.99 (0.49 - 2.00)	0.97	0.84 (0.37 - 1.92)	0.68	NS	9.1
	CG	94(0.34)	105(0.37)	1.21		1.21													
	C	16(0.06)	17(0.06)																
-11917G > A	G	274(0.98)	280(0.96)	0.51 (0.19 - 1.38)	0.19	0.57 (0.18 - 1.82)	0.34	NS	0.51 (0.19 - 1.38)	0.19	0.57 (0.18 - 1.82)	0.34	NS	-	-	-	-	-	26.8
	AG	6(0.02)	12(0.04)	1.38		1.38													
	A	0(0.00)	0(0.00)																
-11849C > T	C	249(0.89)	271(0.92)	1.56 (0.89 - 2.73)	0.12	1.85 (0.93 - 3.68)	0.08	NS	1.53 (0.87 - 2.72)	0.14	1.85 (0.93 - 3.69)	0.08	NS	-	-	-	-	-	34.7
	CT	30(0.11)	22(0.08)	2.73		3.68													
	T	1(0.00)	0(0.00)																
+441G > C	G	108(0.39)	120(0.41)	1.11 (0.87 - 1.41)	0.41	1.11 (0.83 - 1.49)	0.47	NS	1.11 (0.79 - 1.54)	0.56	1.10 (0.73 - 1.65)	0.65	NS	1.22 (0.76 - 1.97)	0.41	1.25 (0.71 - 2.21)	0.44	NS	13.4
	CG	130(0.46)	136(0.46)	1.41		1.49													
	C	42(0.15)	37(0.13)																
+2486C > T	C	156(0.57)	146(0.58)	0.83 (0.64 - 1.08)	0.17	0.85 (0.63 - 1.16)	0.32	NS	0.76 (0.55 - 1.06)	0.11	0.81 (0.54 - 1.21)	0.30	NS	0.92 (0.59 - 1.70)	0.80	0.85 (0.41 - 1.75)	0.65	NS	28.6
	CT	96(0.35)	120(0.41)	1.08		1.16													
	T	21(0.08)	24(0.08)																
ht1	-/-	46(0.17)	46(0.16)	0.98 (0.78 - 1.24)	0.88	1.02 (0.77 - 1.35)	0.91	NS	0.93 (0.60 - 1.46)	0.77	0.96 (0.57 - 1.64)	0.89	NS	1.00 (0.71 - 1.41)	0.99	1.06 (0.70 - 1.60)	0.79	NS	8.1
	-/ht3	122(0.45)	132(0.46)	1.24		1.35													
	ht1/ht1	195(0.38)	111(0.38)																
ht2	-/-	195(0.71)	195(0.67)	0.84 (0.62 - 1.15)	0.28	0.79 (0.55 - 1.15)	0.22	NS	0.83 (0.58 - 1.19)	0.31	0.80 (0.52 - 1.23)	0.30	NS	0.73 (0.28 - 1.96)	0.54	0.54 (0.17 - 1.75)	0.31	NS	19.1
	-/ht2	71(0.26)	84(0.29)	1.15		1.15													
	ht2/ht2	7(0.03)	10(0.03)																
ht3	-/-	209(0.77)	225(0.85)	1.63 (1.11 - 2.40)	0.01	1.55 (0.98 - 2.46)	0.06	NS	1.80 (1.17 - 2.77)	0.007	1.74 (1.05 - 2.89)	0.03	NS	1.33 (0.35 - 5.00)	0.67	0.70 (0.11 - 4.58)	0.71	NS	59.8
	-/ht3	59(0.22)	38(0.13)	2.40		2.46													
	ht3/ht3	5(0.02)	4(0.01)																

Logistic regression models were used for calculating odds ratios (OR) (95% confidential interval (CI)) and corresponding P values for each single nucleotide polymorphism (SNP) site and haplotype. Results of codominant, dominant, and recessive models are also given. Bold data indicate P values < 0.05

[0108] Logistic regression models were used for calculating odds ratios (OR) [95% confidence interval (CI)] and corresponding P values for each single nucleotide polymorphism (SNP) site and haplotype. Results of codominant, dominant, and recessive models are also given. Bold data indicate P values < 0.05.

[0109] TP53의 단일 염기 다형성(-12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T의)의 분포와 폐암 발생 위험도 간의 상관관계를 알아본 결과, 폐암 발생 위험도에 있어서 환자군과 정상군 간에 거의 차이를 나타내지 않았다(표 4 참조).

[0110] 그러나 일배체형 3(-12256G, -11917G, -11849C, +441C 및 +2486C)은 다른 일배체형들에 비하여 폐암 발생 위험도를 증가시켰다. codominant의 경우 오즈비=1.63, 95% CI= 1.11-2.40, P=0.01을 나타냈고, dominant의 경우 오즈비= 1.80, 95% CI= 1.17-2.77, P=0.007을 나타냈다.

[0111] 대상군을 보다 하위그룹으로 세분화하여 하위그룹에서의 일배체형 3의 빈도를 조사하여 일배체형 3과 폐암 발생 위험도 간의 상관관계를 분석하였다. 하위그룹은 성별, 흡연여부 및 세포 타입에 따라 분류하였다. 하기 표 5는 이에 대한 결과를 나타내었다.

[0112] <표 5>

Subgroups	Genotype	Case	Control	Analyzing models referent			Codominant			Dominant			Recessive		
				OR (95% CI)	P	p ^{corr}	OR (95% CI)	P	p ^{corr}	OR (95% CI)	P	p ^{corr}	OR (95% CI)	P	p ^{corr}
Male	-/-	153(0.77%)	236(0.86%)	1											
	-het3	45(0.23%)	33(0.12%)	1.93(1.12-3.32)	0.02	NS	1.55(0.96-2.49)	0.07	NS	1.79(1.06-3.03)	0.03	NS	0.44(0.05-4.24)	0.48	NS
	het3/het3	1(0.01%)	4(0.01%)	0.71(0.23-2.22)	0.56	NS									
Female	-/-	56(0.76%)	11(0.69%)	1											
	-het3	14(0.19%)	5(0.31%)	8.45(0.12-619.82)	0.33	NS	8.45(0.12-619.82)	0.33	NS	8.45(0.12-619.82)	0.33	NS	-	-	-
	het3/het3	4(0.05%)	0(0.00%)	-	-										
Smoker	-/-	151(0.76%)	169(0.85%)	1											
	-het3	44(0.22%)	26(0.13%)	2.03(1.15-3.57)	0.01	NS	1.59(0.98-2.60)	0.06	NS	1.86(1.08-3.22)	0.03	NS	0.45(0.05-4.32)	0.49	NS
	het3/het3	3(0.02%)	4(0.02%)	0.72(0.23-2.23)	0.57	NS									
Nonsmoker	-/-	58(0.77%)	78(0.87%)	1											
	-het3	15(0.20%)	12(0.13%)	1.16(0.26-5.13)	0.85	NS	1.20(0.28-5.10)	0.81	NS	1.18(0.27-5.18)	0.83	NS	-	-	-
	het3/het3	2(0.03%)	0(0.00%)	-	0.99	NS									
Cell type															
Adenocarcinoma	-/-	81(72.3%)	247(85.5%)	1											
	-het3	28(25.0%)	38(13.2%)	2.44(1.21-4.92)	0.01	0.05	2.18(1.18-4.00)	0.01	0.05	2.46(1.24-4.87)	0.01	0.04	2.06(0.21-19.9)	0.53	NS
	het3/het3	3(2.7%)	4(1.4%)	1.13(1.09-1.18)	0.41	NS									
Squamous-carcinoma	-/-	65(81.3%)	247(85.5%)	1											
	-het3	15(18.8%)	38(13.2%)	1.12(0.65-2.98)	0.403	NS	1.18(0.58-2.39)	0.64	NS	1.30(0.61-2.76)	0.50	NS	-	-	-
	het3/het3	0(0.0%)	4(1.4%)	-	-	NS									

[0113]

[0114] 특히, 하위그룹 중 남성, 흡연자, 선암 세포 환자군에게서 일배체형 3과 폐암 발생 위험도 간의 상관관계가 유의하게 나타났다. 일배체형 3을 가진 남성, 흡연자, 선암 세포 환자군은 폐암 발생 위험도가 증가하였다.(P-값 <0.05)

[0115] 남성에게서 일배체형3과 폐암 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, -/ht3(일배체형3)의 유전자형의 경우 analyzing models referent에서는 오즈비=1.93, 95% 신뢰구간=1.12-3.32, P-값=0.02 및 dominant에서는 오즈비=1.79, 95% 신뢰구간=1.06-3.03, P-값=0.03을 나타냈다.

[0116] 또한, 흡연자에게서 일배체형 3과 흡연 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, 흡연자에

서 -/ht3(일배체형3)의 유전자형의 경우 analyzing models referent에서는 오즈비=2.03, 95% 신뢰구간=1.15-3.57, P-값=0.01 및 dominant에서는 오즈비=1.86, 95% 신뢰구간=1.08-3.22, P-값=0.03을 나타냈다.

[0117] 또한, 세포 타입 중 선암에게서 일배체형 3과 흡연 발생 위험도 간의 상관관계가 유의하게 나타났다. 구체적으로, -/ht3(일배체형3)의 유전자형의 경우 analyzing models referent에서는 오즈비=2.44, 95% 신뢰구간=1.21-4.92, P-값=0.01, codominant에서는 오즈비=2.18, 95% 신뢰구간=1.18-4.00, P-값=0.01 및 dominant에서는 오즈비=2.46, 95% 신뢰구간=1.24-4.87, P-값=0.01을 나타냈다.(표 5 참조)

[0118] 그러나 환자군과 정상군 비교 실험에서는 검사의 대상(N)이 증가할수록 (예; 100K, 1M chip, expression chip) 임의 오차 (random error)가 증가하기 때문에 다중검정 (multiple testing)의 문제점을 해결하기 위해서 보정이 필요하다. 따라서 서로 연관비평형 상태에 있는 단일 염기 다형성의 다중 검정에 따라 생기는 오차를 적절히 보정 하기 위해 TP53의 독립 마커 좌위의 개수를 the spectral decomposition (SpD)에 기초하여 the software SNPSpD (<http://genepi.qimr.edu.au/general/daleN/SNPSpD/>)을 사용하여 계산하고 보정된 P^{corr*} -값을 산출하였다. 그 결과 하위그룹은 물론 전체 대상군에서 일배체형 3과 폐암 발생 위험도 간에는 연관성이 없는 것으로 밝혀졌다.

[0119] 결과적으로 한국인에 있어서, TP53의 단일 염기 다형성(-12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T)과 폐암 발생 위험도 간에는 연관성이 없으며, 일배체형 1 및 2와 폐암 발생 위험도 간에도 전혀 연관성이 없었다. 다만, 일배체형 3(-12256G, -11917G, -11849C, +441C 및 +2486C)의 빈도가 한국인 특히, 남성, 흡연자, 선암 세포 환자군에서 높게 나타나 일배체형 3과 폐암 발생 위험도간에 연관성 여부가 의심되었으나 오차를 보정하여 다시 P^{corr*} -값을 산출해 본 결과 일배체형 3 역시 폐암 발생 위험도와 무관하였다.

[0120] 종래 이뤄진 발명들은 TP53의 단일염기다형성 중 Arg72 또는 Pro72 단백질로 번역되는 엑손4 코돈 72의 두번째 염기의 G/C 변이에 관한 것이었다(Petitjean et al.2007). 뿐만 아니라 코돈 72의 다형성과 폐암 발병 간의 연관성에 대해 서로 모순되는 결과만을 내놓고 있으며 서로 일치점을 찾지 못하고 있었고 한국인의 TP53 다형성과 폐암 발생 위험도 간의 연관성에 대해서는 밝혀낸 바가 전혀 없다.

[0121] 본 발명자들이 비로소 한국인에서는 TP53의 단일 염기 다형성(-12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T) 및 일배체형 1,2,3과 폐암 발생과는 무관함을 밝혀낸 것이다.

산업이용 가능성

[0122] 이상, 상기에서 살펴본 바와 같이 본 발명 TP53 유전자의 -12256G>C, -11917G>A, -11849C>T, +441G>C 및 +2486C>T 다형성의 빈도를 측정하여 한국인의 발암 발생 위험도를 예측하는 방법 및 이를 이용한 진단 키트는 암 억제 유전자 TP53의 단일염기다형성 및 이의 일배체형과 한국인 폐암 발생과의 연관성을 확인하여 TP53이 폐암의 유전적인 위험 인자로서, 폐암의 치료 진단 표지 인자로 사용 가능한지를 확인하는데 사용될 수 있다.

도면의 간단한 설명

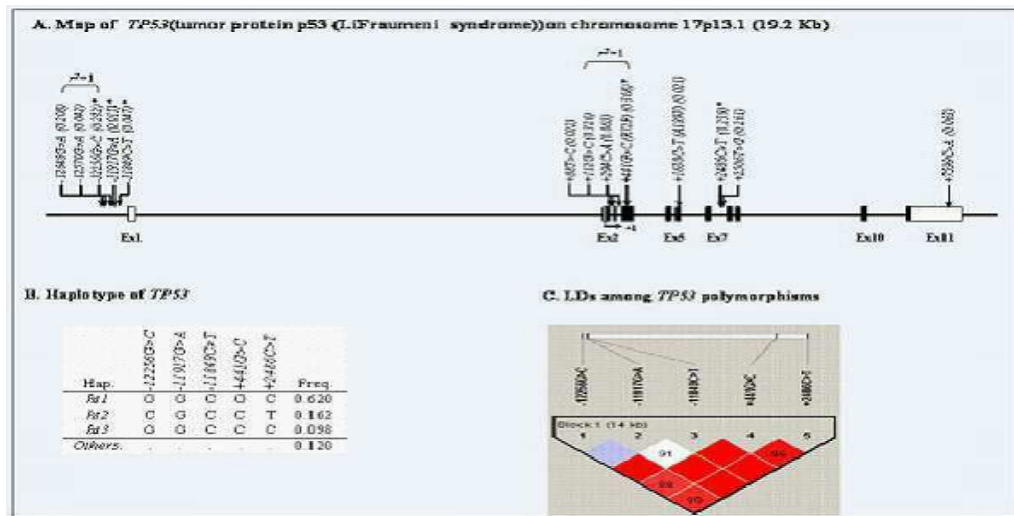
[0123] 도1의 A는 유전자 TP53의 유전자 지도 및 SNPs를 나타낸 것이다. 엑손은 검은색 블록으로 표시되어 있고 5', 3'UTR은 흰색 블록으로 표시되어 있다. 화살표는 SNP를 가르킨다. 번역시점의 첫 번째 염기는 핵산 +1로 표시되어 있다.

[0124] 도1의 B는 TP53의 일배체형을 나타낸다. 빈도수 > 0.03인 일배체형만 나타냈다.

[0125] 도1의 C는 TP53의 5개 SNPs간의 Linkage disequilibrium coefficient ($|D'|$)을 나타낸 것이다.

도면

도면1



서 열 목 록

<110>	korea university industry and academy collaboration foundation
<120>	Method for predicting lung cancer risk by measuring TP53 gene polymorphism and kit for diagnosing lung cancer risk using the same

<160> 21

<170> Kopatent In 1.71

$\langle 210 \rangle$	1
-----------------------	---

<211> 24

<212> DNA

<213> Homo sapiens

 $\langle 400 \rangle$ 1

cgattcgagg actcaactca ctag

24

 $\langle 210 \rangle$ 2

<211> 17

<212> DNA

<213> Homo sapiens

 $\langle 400 \rangle$ 2

ggctgggtca gctgtgt 17

<210> 3
<211> 21
<212> DNA
<213> Homo sapiens

<400> 3
ctgtgcttcc ctctcttct g 21

<210> 4
<211> 25
<212> DNA
<213> Homo sapiens

<400> 4
gctaaggaac acagtgttt caaaa 25

<210> 5
<211> 23
<212> DNA
<213> Homo sapiens

<400> 5
tgaaagcact gtgttcctta gca 23

<210> 6
<211> 20
<212> DNA
<213> Homo sapiens

<400> 6
cggaaggtgg accgaaatcc 20

<210> 7
<211> 22
<212> DNA
<213> Homo sapiens

<400> 7

ccagatgaag ctcccagaat gc 22

<210> 8
<211> 15
<212> DNA
<213> Homo sapiens

<400> 8
gccgccggtg tagga 15

<210> 9
<211> 18
<212> DNA
<213> Homo sapiens

<400> 9
gcctctgctt gcctctga 18

<210> 10
<211> 25
<212> DNA
<213> Homo sapiens

<400> 10
ggtggatggg tagtagtatg gaaga 25

<210> 11
<211> 14
<212> DNA
<213> Homo sapiens

<400> 11
cctcccgtag ctcc 14

<210> 12
<211> 14
<212> DNA
<213> Homo sapiens

<400> 12

cctcccctag ctcc 14

<210> 13
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 13
 ctgttcgcct ctct 15

<210> 14
 <211> 15
 <212> DNA
 <213> Homo sapiens

<400> 14
 ctgttcgctt ctct 15

<210> 15
 <211> 14
 <212> DNA
 <213> Homo sapiens

<400> 15
 cagcaagagg cccg 14

<210> 16
 <211> 14
 <212> DNA
 <213> Homo sapiens

<400> 16
 cagcaaaagg cccg 14

<210> 17
 <211> 13
 <212> DNA
 <213> Homo sapiens

<400> 17

tccccgcgtg gcc	13
<210> 18	
<211> 14	
<212> DNA	
<213> Homo sapiens	
<400> 18	
ctccccccgt ggcc	14
<210> 19	
<211> 14	
<212> DNA	
<213> Homo sapiens	
<400> 19	
ccctgggccc acct	14
<210> 20	
<211> 15	
<212> DNA	
<213> Homo sapiens	
<400> 20	
cccctgggtc cacct	15
<210> 21	
<211> 21198	
<212> DNA	
<213> Homo sapiens	
<400> 21	
gtcaggagcc ctagaaacag gggagagtta gaaagctggc cagacctatg cttttcaagt	60
gtagggctag ggctgagcct gcctctgggg taggtaagcc cccctgaatc cttgagggaa	120
gtagaagaca caaactgcta gataaaatgt aagctcagtc taaaagggt acgtgccgt	180
tctccagct ctggggcatc cctctcctag aaaactggac tgttttacag tgaaaatctc	240

gggggtggtc agtcacctgc cccgttgta tccttaccac ttacagcctt tcaagaagtt	300
ctcaggttgg gtgctgaact ctgaccagga accactgaga aatcgaggca gctgggagaa	360
gctgtagttc caagcgctga aaggaagatg ggggacaata aacctgggtc gccaagcaaa	420
gggggcagag gcctggagaa gtgggtctca ggaccagagg acagatcgac ctcacacttc	480
atctccccag actccacact ccactgccat caccacttac gtgtctccct cgtcctctgc	540
agcgggttcc ccagaggtat cttccatggc ttttcagac cccaactctg gcccgttcgc	600
ttcttttca gaaaggctcc cgtttgcttc ttctgcagga aggcctgtat ttccagaaag	660
ttcttgctcc tcgattcgag gactcaactc actaggggaa ccaaactctg ttccagggg	720
agtggagaga gaaactgggt cccctcccg tagctcctgg gacacagctg agccagccac	780
aggatctggg gacaaccggg gcggatcccc ctttcggga ggcggtggca tcagttcaga	840
gtccgcattt ttattcatcg gggaagcgtg gggagaagga tgggctggag ctgggtcctg	900
gtctgaagga cagcagtcgg gagctaacgg ttgagtctcc aaagtcttca tactgcagag	960
gaagcacagc ggagattagc ctacgccagg atggcttcga agttctcagg gatccgacgc	1020
agagctaaag aaaccacct gtgcttcct cctcttctgg gagiaggcag aagactcccg	1080
ggaggagagg cgaacagcgg acgccaattc tttgaaagc actgtgttcc ttagcaccgc	1140
gggtcgtac gggcctcttg ctgtcgcggg atttcggtcc accttcgat tgggccgccg	1200
catcccgat cagatttcgc gggcgacca cggaaccgc ggagccggga cgtgaaaggt	1260
tagaagttt cccgttccca tcaagcccta gggtcctcg tggctgctgg gagttgtagt	1320
ctgaacgctt ctatcttggc gagaagcgcc tacgtcccc ctaccgagtc ccgcgtaat	1380
tcttaaagca cctgcaccgc cccccgccg cctgcagagg gcgcagcagg tcttgcacct	1440

cttctgcate tcattctcca ggcttcagac ctgtctccct cattcaaaaa atatttatta	1500
tcgagctctt acctgctacc cagcactgat ataggcactc aggaatacaa caatgaataa	1560
gatagtagaa aaattctata tcctcataag gcttacgttt ccatgtactg aaagcaatga	1620
acaaataaat cttatcagag tgataagggt tgtgaaggag attaaataag atgggtgat	1680
ataaagtatc tgggagaaaa cgttagggtg tgatattacg gaaagccttc ctaaaaaatg	1740
acattttaac tgatgagaag aaaggatcca gctgagagca aacgcaaaag ctttcttct	1800
tccacccttc atatttgaca caatgcagga ttcttccaaa atgatttcca ccaattctgc	1860
cctcacagct ctggcttgca gaattttcca ccccaaatg ttagtatcta cggcaccagg	1920
tcggcgagaa tctgactct gcacctctt ccccaactec atttctttg cttectccgg	1980
caggcggatt acctgccctt acctgtcatg gcgactgtcc agctttgtgc caggagcctc	2040
gcaggggttg atgggattgg ggttttcccc tcccatgtgc tcaagactgg cgctaaaagt	2100
tttgagcttc tcaaaagtct agagccaccg tccagggagc aggtagctgc tgggctccgg	2160
ggacactttg cgttcgggtt gggagcgtgc ttccacgac ggtgacacgc ttccctggat	2220
tgggtaagct cctgactgaa cttgatgagt cctctctgag tcacgggctc tcggctccgt	2280
glattttcag ctcggaataa tcgttggggc tgggggtggg gcagtgggga cttagcgagt	2340
ttgggggtga gtgggatgga agcttggcta gagggatcat cataggagt gcattgttgg	2400
gagacctggg tgtagatgat ggggatgta ggaccatccg aactcaaagt tgaacgccta	2460
ggcagaggag tggagctttg gggaaccttg agccggccta aagcgtactt ctttgacat	2520
ccaccgggtg ctgggcgtag ggaatccctg aaataaaaga tgcacaaagc attgaggtct	2580

gagacttttg gatctcgaaa cattgagaac tcatagctgt atattttaga gcccatggca	2640
tcctagttaa aactggggct ccattccgaa atgatcattt gggggtgatc cggggagccc	2700
aagctgctaa ggicccacaa ctcccgacc ttgtctctc ctggagcgat ctttcaggc	2760
agccccggc tccgctagat ggagaaaatc caattgaagg ctgtcagtcg tggaaagttag	2820
aagtgctaaa ccaggggttt gcccgccagg ccgaggagga ccgtcgcaat ctgagaggcc	2880
cggcagccct gttattgttt ggctccacat ttacatttct gcctcttgca gcagcatttc	2940
cggtttcttt ttgccggagc agctcactat tcacccgatg agaggggagg agagagagag	3000
aaaatgtcct ttaggccggt tctcttact tggcagaggg aggtctctat tctccgctg	3060
catttctttt tctggattac ttagttatgg cctttgcaaa ggcaggggta ttgttttga	3120
tgcaaacctc aatccctccc ctcttttgaa tgggtgtccc caccgccgg gtgcctgca	3180
acctaggcgg acgtacat ggctgagac agggaggga agaagtgtgc agaaggcaag	3240
cccgaggta tttcaagaa tgagtatac tcatttccc ggaggaaaa aaaaaagaat	3300
gggtacgtct gagaatcaaa ttttgaaaga gtcaatgat gggtcgttg ataattgtc	3360
ggaaaaacaa tctacctgtt atctagcttt gggctaggcc attccagttc cagacgcagg	3420
ctgaacgtcg tgaagcgga gggcgggcc cgcaggcgtc cgtgtggtcc tccgtgcagc	3480
cctccggccc gagccggttc ttctggttag gagcggaac tcgaattcat ttctccgct	3540
gccccatctc ttagctcgcg gttgtttcat tccgcagttt cttcccatgc acctgccgcg	3600
taccggccac tttgtccgt acttacgtca tcttttctt aaatcgaggt ggcatttaca	3660
cacagcgcca gtgcacacag caagtgcaca ggaagatgag ttttgcccc taaccgtcc	3720
gtgatgccta ccaagtcaca gaccttttc atcgtcccag aaacgtttca tcacgtctct	3780

tcccagtcga ttcccgaccc cacctttatt ttgatctcca taaccatttt gcctgttgga	3840
gaacttcata tagaatggaa tcaggctggg cgtgtggct cagcctgca ctttgggagg	3900
ccgaggcggg cggattactt gaggatagga gttccagacc agcgtggcca acgtggtgaa	3960
tccccgtctc tactaaaaa taaaaaatt agctgggcgt ggtgggtgcc tgtaatcca	4020
gctattcggg agggtaggc aggagaatcg cttgaacccg ggaggcagag gttgcagtga	4080
gccaa gatcg tgccactaca ctccagcctg ggcgacaaga acgaaactcc gtctcaaaaa	4140
aaagggggga atcatacatt atgtgctcat tttgtcggg cttctgtcct tcaatgtact	4200
gtctgacatt cgttcatgtt gtatatatca gtattttgt ccttttcatt tagtatagtc	4260
catcgattgt ataccgtcc ttttgatggc cttttgagtt gtttccatt tgcggttatg	4320
aaataaagct gctataaaca ttctgttaca attctttttg tgatcatatg ttttcgtgtt	4380
ttttggagaa ataccttagga ggggaattgc gagtttggaa gtaaaaagta gctgtatttt	4440
gaactttttc agaagctctg agttttccag agcggttgta ccattttaca ctccaactag	4500
caaggtatgg gatttattat ggttgtgcca cagccttccg gacattaggt attgtcagtc	4560
tttctaattg ggtatatact tgtggttgta atttacagtt ctctattgac taaggatgtt	4620
cagcattttt tcatgtgcct attggccatt cgtattttgt ttgtaaagta gctcttcgag	4680
tcttttacct gttattttgg tttttgttt gtttttattg ttcagttgtg ggactgcttt	4740
atacattctg gatacaagtc ctttatcaga tccatgtgtc gtgaatgttt tcttctgac	4800
tgttgcttgc ctattttgtt gctttacaga gtttacagta tcttaagagg agtggattta	4860
tcttttttat gttcagtatt tgccctgtcc tgtttaggac atcttttttt ttttttttaa	4920

ccccagggtc atgaagatat tatcttacat tttcttttag gacctttatg gtgtgaagtt	4980
ttacagtaag gtccttgagc cattaattaa tttcttaaat taattgttta tgggtgagg	5040
tgtaggagtc agtctctggg atctttcccg tatggaaatc cagttattct gtctccactt	5100
gttgaatatag gcttcccttc tctactgaat gcttttaatt ttaattatct tacagttgga	5160
glatagggct accatttttag tgctattttc tttttttctt tgttaatttt tgagacaggg	5220
actcacactg ttgccaggc tagagtacaa tggcacaatc aaggcttact gcagcctcga	5280
acccctgggc tcaagcagtc ctctagcagc ctccagagta gctgggatta ctccaccaca	5340
cccagctaac tattttatct tttgtattg acaggatctc actatgttgc ccaggctggt	5400
ctcaaatgc tggcctcaag ctttcatccc atctcgccct cccaaagtgc tgggattaca	5460
ggtgtgagcc accatgcctg acctcttagt gctattttct atttatctcc tctgttctct	5520
gtctctctta aacgttggag gaagaaacag taccatctt acacaaactc ttcagaaaac	5580
agaggaacag actgggcgcg gtggctcata cctgtaatct cagcactttg gtacgtgag	5640
gcaggggatc atttgagtc gggagtctga gaccagcctg gccaacacgg cgaaacccca	5700
tctctactaa aaatacaaaa agtagctagg cgtggtgaca catacctgta atgccagtta	5760
ctcaggaggc tgaggcaca gaatcccttg aacctgggaa gcggaggttg cagtgagccg	5820
agattgcgcc actgcactcc agcctgggca acagagttag accctgtctc agaaaaaaaa	5880
agaaagaaag aaaaaataga ggaatatttc ccaacttggt ttcgaagcca gcataatcct	5940
ggtaccaaaa ccaacaagg acattataag aaaagaaaat atagaccaat attcctgtta	6000
gcatagacat gcaacagcta accaatctta gcaaaccaaa cctggtaata tagaaaaaag	6060
gataaatagg ccagtcgagg tggctcacgc ctgtaatccc agcactttgg gaggtgagg	6120

caggcagatc acttgaggtc aggagtttga gaccagcctg accaacaatgg tgaaaccccg	6180
tttctaataa aaatacaaaa atcaggctgg gcacgggtggc tcacgcctgt aatcccagca	6240
ctttgggagg ccgagggtggg cagatcacga ggtcaggagt tcaagaccag cctgaccaat	6300
gtggtgaaac gccatctcta ctaaaaatac aaaaatcagc cgggtgtggtg gcacctgcct	6360
glaatccag ctactcagga ggctgaggca gaattgcttg aaccggggag gcagaggttg	6420
cagttagcca agatcgtgcc actgcactcc agcctgggcg acagagcaag acttcatctc	6480
aaaaaaaaa aaaaattagc tgggcatggt ggtgggcacc tgaaatcca gctactcggg	6540
agtctgaggc aggagaatcg cttgaacca ggaggcagaa gttgcactga gctgggatca	6600
caccattgca ctccagcctg ggcaacagag tgagactcca tctcaaaaaa agaaaaagaa	6660
aaaggataaa tacattctaa ccaataatg tttatctcat gattgtagct gattcaacat	6720
tcaaaaattg gcctggtgca gtagctcagg cctgtaatcc caacatttta ggaggctgag	6780
gcaggaagat ctcttgagcc caggatttca agaccagcct gggcaacata gtcagactgg	6840
tctttactgg ggggaaaaaa atcagtctgt gtaattcacc acattaacaa agggaacat	6900
aaaaacccta tgatcatttc aacagatgta gcaaaagcag ttaatgatat tcaacacata	6960
tgcatgatta caaaccaacc aacctcctag caaactaggg aaaggaaact taacctagtt	7020
tgataacagg gcgtccacag tcggagtcc actagcagca tacataatgg tagaaaactc	7080
agtgctgccg ggcgcggtgg ctcacgcctg taatgccagc actttgggag gcctaggcgg	7140
gcggatcacg aggtcaggag atcgagactg tcctgactag catgctgaaa ccccgctctt	7200
actaaaaata caaaaacaaa aaattagccg ggcatggtgg cgggcgccta tagtccagc	7260

tactcgggag gctgaggcga gagaatggcg tgaacccggg aggcggagct tgcagagcct	7320
agatcgtgcc actgcactcc agcctgggtg acagagtggag acttcgtctc aaaaaaaaaa	7380
aaaaaaaaa aagaaaagaa aactcaacgc tttttcctct aagatcagga actagaaaag	7440
gatttgactc tcacaacgtt gataccatac tggaggtttt aaccaggcaa gaaaaagaaa	7500
taatgagggc cgggtgcggt ggctcaggcc tgtaatccca gcactttggg aagccgagac	7560
gggtggatca cgaggtcagg agatcgagac catcctggct aacacggtga aaccctgtct	7620
ctactaaata tacaaaaat tagccgggcg tagtggcggg cgctttagt cccagctact	7680
ggggaggctg aggcaggaga atggcgtgaa ctcagggggc ggagcttgca gtgagctgag	7740
atcgagccac tgcactccag cctgggcgac agagcaagac tgtgtctcaa aaaaaaaaaa	7800
agaaaaagaa ataattgatta gtggcccgat gtctcacgcc tataatccca gcactttggg	7860
aggccgaggt gggcagatca cctgaggtct ggagttggag accagcctga caaatggt	7920
gaaacctcgt ctctattaaa atattaaaaa aatagccagg cgttggccgg gtacagtggc	7980
tcatgcctgt aatcccagca ctttgggagg ccgaggtggg tggatcacct gaggtcagga	8040
gttcaacacc agcctggcca acatggtgaa acccatctc tactaaaaat acaaaaatta	8100
gccgggcgta gtggcgggcg cctgtaatcc cagctacttg ggaggcttag gcaggagaat	8160
cgcttgaacc tgggaggcgg aggtttagt gagccgagat tgcaccattg cactccagcc	8220
tgggtgacaa aagcaaaaac tccgtctcaa aaaaaaaga attagccagg gtagtggtg	8280
aacgcctgta gtcccagcta ctcaggagc agaggcagga gaatcacttg aaccaggag	8340
gcagaggttg cagtgagccg agattgtccc attgcactcc agcctaggcg acaagagcaa	8400
aattccatgt caaaaaaaaa aaaaaaaaag gaaagaaaaa aaataacgat tagaaaggaa	8460

gaaataaaac acattcacag ccagtatgat tctatacata catgtcctaa tggggccagg	8520
cgtaggtggct catgcctgta atcctagcac ttttaggagg ctgaggcagg tggcttcct	8580
gggaccagcc tggccaacat ggtgaaaccc caactctaataaaaaatacaa aaaatcagcc	8640
aggcgtgggtg acgggcacct ctaatcccag ctactcagga ggctgaggca ggagaattgc	8700
ttggacctgg gaggcagagg ttgcagtga ccgagatcgc gctattgcac tccagcctgg	8760
gcaacaagag tgaactccg gcagggtgtg gtggcttacg cctgtaatcc cagcacttcg	8820
ggaggctgag gcaggccgat cacctgaggt caggagtttg agaccaacct aacatggtga	8880
aaccccgctct ctactaaaaa tacaagaatt agctgggtgt agtgggtggc gcctgtaatc	8940
ccagctactt gggaggctga gacagaagaa ttgcttgaac ccaggagggtg gaggttgtag	9000
tgagctgaga tcatgccatt gcacaccacg ccgggcaaca gagcgagatt ccgtctcaaa	9060
aaaaaaaaa aagagtgaat ctctatctca aaaaaaaaaa aaagtcctaa tggaaaaatcc	9120
ataaaaagct accaaaacta ataaataaat atagcagggt tgcagggttac agggcaatat	9180
agttatccct ctatctgtag gggcttggtt ctgggactcc tcacacacca aaccacaga	9240
tgtctaagtc ccatatataa gacgtatag tatttgatt taacctacac atatcctccc	9300
atatagttaa aattatctct agattactta cattaccccc atacaatgaa aatgctaag	9360
tacatgcaag tatgtatgta agtacttgta ctatattgtt tagggaatca ctggacatat	9420
aggccttcaa gactgatacc agcagccact gttgaagattc tggtcaggcc tgcccctgtt	9480
tggggtctca gttgatctca ttgccttccc acccagccaa gggcacctgc atttctcttg	9540
gtccctggc catttggaag gcctagttca gcctggcaca tttgtatcct ggcccactga	9600

tgctgtgacc cctgggaagg tcctgctctg aaaaacacgg agattttagt tgctactgaa	9660
gatttgagag ataaagacag ggagacctgt ctgtagacct gtgtccctcc aagtgggatt	9720
gagactttgg gccccccatt tcaggacagc acctcctggc ctgttgactg aatagatccc	9780
tgaaggaggt gtacttgcatt taatggagtg ggggtgggag cagtagcaca gatccgcact	9840
aacaatcaca cagttctctc tagaataata atatagaaca agtgaaatag aacaattgca	9900
gaaagagcta acctttgttg agctcttact gtgtgcccag cactttcctc aactctacat	9960
ttcccataat acacagagta ctaggtaggc caggcttggg ggctcacgcc tgtaatccca	10020
gcactttagg aggccaaggg ggggtggatca cctgaggtcg ggagtccaag accagcctga	10080
ccaacatggt gaaaccccggt ctctactaga agtacaaaat tagccaggtg tgggtgcaca	10140
tgctttagt cctagctact cagcaggctg aggcaggaga atcatttgaa tccgggagga	10200
ggttgtagta agcggagata gtgccactgt actccagcct gggcaataag agctgagact	10260
ccgtctcaaa ataaaataaa ataaaataaa ataaaataaa ataaaataaa aaaagaaaag	10320
agcctgccat taaaggagct gtttggtagg ggatgttttg tcagtgcaa caacagaaaa	10380
gtgggctggg cacagtgggt catgcctgta atcccagcac tttgggaggc caaggcgggc	10440
ggatcacctg aagtgggag ttcaagacca gcctgaccaa tatggagaaa cccgtctct	10500
actaaaaata caaaattagc cgggcgcagt ggcgcatgcc tgtaatccca gctactcggg	10560
aggctgaggc aggagaatcg cttgaacctg ggaggcagag gttgcggtga gccgagatcg	10620
caccattgca ctccagcctg gacgagagca aaactctgtc tcaaaaaaaaa aaaaaaacag	10680
aaaagtgtaa caaacactta cagtaggcat gtttcttagc aaatctgatg acaaatttgg	10740
cataaagaaa gagagcatcc ctgaaaaaaaa aaaaaagaaa aagaaagaga gcatcctgcc	10800

tgggcaacat agtgaaaccc tgcctctaca aaaaaactca aaaattggcc ggggtgcagtg	10860
gtcacacct gtaatcccag cactttggga gtcggaggcg ggaggatcac ctgaggtcag	10920
gagttcgaaa ccagcctggc caacatggca aaaccccatc tctactaaaa atacaaaaaa	10980
ttaatcaggc gcatttgggtg gcgcctgtaa tcccagctac tcaggaagtt gaggcaagag	11040
gatcgcttga atctgggagg tggaggttac agtgagtca gatcacacca ctgcactcta	11100
gcctgggtga cagggcgaga ctccgtctcc aaaaaaaaaa agaaaaagaa aaagactaaa	11160
aaattagcca ggcaggcctc tgtgttccca gctacttggg aggctgaggc aggagaatca	11220
ctgagcccag gagtccgagg ctgtagttag ccatgattgc accactgtac cctagcttgg	11280
gcaacaaagc aagaccctgc ctcaaaagaa aaaagaaaga aagaaagaac atggcggggc	11340
aggcacagtg gtcacacct gtaatcccag cgttttgaga ggccgaggca ggtggatcac	11400
aaggtcagga gtccacacc agcctggcca acatggtgaa accctgtctc tactaaaaat	11460
acaaaaaatc agccaggcat ggtggcaggg gcctgtaatc ccagctactc gggaggctga	11520
ggcaggagaa ttgcttgaaa ccagaaggca gaggttgcag tgagcctaga ctgcaccact	11580
gcactccagc ctgggcgaaa agagccaaac tccatctcaa aaaacaaaca aaaaaacaaa	11640
acaaaagaaa acatggcaaa gcctttgaaa gcttgtctgg gagaaggtgc gatgatagtt	11700
gcataacttc gtgcaagatg ctggtccaca caggggtgc ccttgtctct ttctcgtct	11760
cttaacctct catataacag gcttgtgtgt tattcacatt tattgagccc aagcaggtgc	11820
aaggcattgt gatctaatac tttgttcagc aagacaacaa gatagatcac tgccctgccc	11880
ttaggaagtg tatatgctat tagaggaaac agataaaata aacaaggaaa agtatcagac	11940

aatgtaagtg ctatgagaat gcaaatgagg tgatgtgaat taaaatagga tgacttaaag	12000
tctgcacggg aaggagccta ccccatgtt cctggctagc caaggaacca ccagttgatt	12060
agcagagaag ggcagccagt ctagctagag cttttgggga agagggagtg gttgttaaga	12120
gatgagatta aagaagccga gacgggcat tctgtagggg tttgtaatgc agggctgagg	12180
agtgtccgaa gagaatgggc aggtgagcgg tgagacagtt gttcttcag aagctttgca	12240
gtgaaaggaa tcaaagaaat ggagccgtgt atcaggtggg gaagggtggg ggccaagggg	12300
gtgtccttcc ccatacagag attgcaggct gagaatgact atatccttgt taacaggagg	12360
tgggagcagg gcacggtagc tcacacctgt aatcttggca ctttaggagg ctgagcgagg	12420
ccgatcacct gaagtaagga gttcagacc agcctggcca acatgcaaag ccctgtctct	12480
actaaaaata caaaaattag ctgggtgtgg tggtagctgc ctgtaatccc agctactcgg	12540
gagactgagg caggagaatg gcttgaacc ggaaggtaga ggttcagtg agctgagatc	12600
atgccactgt gctccagcct aggtgacaga gagagactcc atctcaaaaa aaaaaaaaaa	12660
atacaggaag ggagtggga atagggtgca catttaggaa gtcttgggga tttagtgtg	12720
ggaaggttg aagtcctct ctgattgtct tttctcaaa gaagtgcag gctggtgagg	12780
ggtggggcag gagtgcctgg gttgtggtga aacattggaa gagagaatgt gaagcagcca	12840
ttcttttct gctccacagg aagccgagct gtctcagaca ctggcatggt gttgggggag	12900
ggggttctt cctgcaggc ccaggtgacc cagggttga agtgctcat gctggatccc	12960
cacttttct cttgcagcag ccagactgcc ttccgggtca ctgcatgga ggagccgag	13020
tcagatccta gcgtcagacc cctctgagt caggaaacat tttcagacct atggaaactg	13080
tgagtggatc cattggaagg gcagccccc cccccacc ccaacccag cccctagca	13140

gagacctgtg ggaagcgaaa attccatggg actgactttc tgctcttgtc tttcagactt	13200
cttgaaaaca acgttcttgtt aaggacaagg gttgggctgg ggacctggag ggctggggac	13260
ctggagggtt ggggggctgg ggggctgagg acctggctct ctgactgtct tttcaccca	13320
tctacagtcc cccttgccgt cccaagcaat ggatgatttg atgtgtccc cggacgatat	13380
tgaacaatgg ttactgaag acccaggtcc agatgaagct cccagaatgc cagaggctgc	13440
tccccgctg gccctgcac cagcagctcc tacaccggcg gccctgcac cagccccctc	13500
ctggccccctg tcatcttctg tcccttcca gaaaacctac cagggcagct acggtttccg	13560
tctgggcttc ttgcattctg ggacagccaa gtctgtgact tgcacgttca gtgcccctga	13620
ggggctggct tccatgagac ttcaatgcct ggccgtatcc cctgcattt cttttgtttg	13680
gaactttggg attcctcttc accctttggc ttctgtcag tgtttttta tagtttacc	13740
acttaatgtg tgatctctga ctctgtccc aaagttaga attccccct tgaatttggg	13800
cttttatcca tccatcaca cctcagcat ctctctggg gatgcagaac tttcttttt	13860
cttcatccac gtgtattcct tggcttttga aaataagctc ctgaccaggc ttggtggctc	13920
acacctgcaa tcccagcact ctcaaagagg ccaaggcagg cagatcacct gagcccagga	13980
gttcaagacc agcctgggta acatgatgaa acctcgtctc tacaaaaaa tacaaaaat	14040
tagccaggca tgggtgtgca cacctatagt cccagccact taggaggctg aggtgggaag	14100
atcacttgag gccaggagat ggaggctgca gtgagctgtg atcacaccac tgtgtccag	14160
cctgagtgac agagcaagac cctatctcaa aaaaaaaaa aaaaaagaaa agctcctgag	14220
gtgtagacgc caactctctc tagctcgcta gtgggttgca ggagggtctt acgcatgttt	14280

gtttctttgc tgccgtcttc cagttgcttt atctgttcac ttgtgccctg actttcaact	14340
ctgtctcctt cctcttccca cagtactccc ctgccctcaa caagatgttt tgccaactgg	14400
ccaagacctg ccctgtgcag ctgtgggttg attccacacc cccgcccggc acccgcgcc	14460
gcgccatggc catctacaag cagtcacagc acatgacgga ggttgtgagg cgctgcccc	14520
accatgagcg ctgctcagat agcgatggtg agcagctggg gctggagaga cgacagggt	14580
ggttgcccag ggtccccagg cctctgattc ctactgatt gctcttaggt ctggccctc	14640
ctcagcatct tatccgagtg gaaggaaatt tgcgtgtgga gtatttgat gacagaaaca	14700
cttttcgaca tagtgtgtg gtgccctatg agccgcctga ggtctggttt gcaactgggg	14760
tctctgggag gaggggttaa ggggtgttgt cagtggcct ccaggtgagc agtagggggg	14820
ctttctctg ctgcttattt gacctcccta taaccccatg agatgtgcaa agtaaatggg	14880
tttaactatt gcacagtga aaaaactgaa gcttacagag gctaagggcc tccccgtctt	14940
ggctggggcg agtggctcat gcctgtaatc ccagcacttt gggaggccaa ggcaggcgga	15000
tcacgaggtt gggagatcga gaccatcctg gctaacggtg aaaccccgtc tctactgaaa	15060
aatacaaaaa aaaattagcc gggcgtgtgt ctgggcacct gtagtcccag ctactcgga	15120
ggctgaggaa ggagaatggc gtgaacctgg gcggtggagc ttgcagtgag ctgagatcac	15180
gccactgcac tccagcctgg gcgacagagc gagattccat ctcaaaaaa aaaaaaaag	15240
gcctccctg cttgccacag gtctcccaa ggcgcactgg cctcatcttg ggcctgtgtt	15300
atctcctagg ttggctctga ctgtaccacc atccactaca actacatgtg taacagtcc	15360
tgcattggcg gcatgaaccg gaggcccatc ctccatca tcacactgga agactccagg	15420
tcaggagcca cttgccacc tgcacactgg cctgctgtgc cccagcctct gcttgcctct	15480

gacccttggg cccacctctt accgatttct tccatactac taccatcca cctctcatca	15540
catccccggc ggggaatctc ctactgtct ccactcagtt tctttttctc tggctttggg	15600
acctcttaac ctgtggcttc tctccacct acctggagct ggagcttagg ctccagaaag	15660
gacaaggggtg gtggggagta gatggagcct ggttttttaa atgggacagg taggacctga	15720
tttcttact gcctcttct tctcttttcc tctctgagt agtggtaatc tactgggacg	15780
gaacagcttt gaggtgcgtg tttgtcctg tcttgggaga gaccggcgca cagaggaaga	15840
gaatctccgc aaaaaagggg agcctcacca cgagctgccc ccaggagca ctaagcgagg	15900
taagcaagca ggacaagaag cgggtggagga gaccaagggt gcagttatgc ctcagattca	15960
cttttatcac ctttcttgc ctctttcta gcactgcca acaacaccag ctctctccc	16020
cagccaaaga agaaaccact ggatggagaa tatttcaccc ttcaggtact aagtcttggg	16080
acctcttacc aagtggaaa tttccagtct aacctcaaa atgccgtttt cttcttgact	16140
gttttacctg caattggggc atttgccatc agggggcagt gatgcctcaa agacaatggc	16200
tcttggttgt agtaactaa cttcagaaca ccaacttata ccataatata tattttaaag	16260
gaccagacca gccttcaaaa agaaaattgt taaagagagc atgaaaatgg ttctatgact	16320
ttgcctgata cagatgtctc ttgacttacg atggtgttac ttctgataa actcgtcgtc	16380
agttgaaaat attgtaagtt gaaaatggat ttaatacacc taatctaagg aacatcatag	16440
cttagcctag cctgcttttt tttttttttt ttttggagac agagtctcac tctgtcaccc	16500
aggctggagt gcagtggcgg gatctcggct cactgcaacc tccgccttct gggttcaagc	16560
gattctctg cctcagccca ctgagtagct gggattacag gcacctgccc cgacgcccag	16620

ctaatttttt gttattttatt tatttttttt tttagtagag atgaggtttc accatgttgg	16680
ccaggctagt ctggaactcc tgaccttggtg atctgcctgc cttggcctcc caaagtgtctg	16740
ggattacagg cgtagaccac cgcacccggc ctgcctagcc tactttttatt ttatttttaa	16800
tggagacagc atcttgctct gttgccaggc ctggattaca gtgatgtgat catagctcat	16860
tataccctcc tgggctcaag caatccccct aactctgcct cccagtagc taggaccaca	16920
ggcatacacc accataccca gctaattttt aaaatttttt gtagatagat agagtctcac	16980
tatgttgccc aggttggtct ctagectact tttttgagac aaggctctgc tctgtcaccc	17040
aggctggata gagtgcagta gtgcagtcac agctcactgc agcctccacc tcccaggctc	17100
catccatcct cccagctcag cctcccaagt tgcctcaact acaggcctgc accacatgc	17160
ctggctaatt tttattttatt tattttttatt ttattttatt ttattttttt gagactcagt	17220
ctcactctgt cgcccaggct ggagtgcagt ggcatgatct cggtcactg caacctctgc	17280
ctcctgggtt caagtgatc tcctgcctca gcctccgaa tagctaggac tacaagcgcc	17340
tgctaccacg cccagctaatt ttttgatttt ttagtagaga cagggtttca ccatgttggc	17400
caggctggtc tcgaacttct gaccatgtga tccgcccgcc tcggcctccc aaagtgtctgg	17460
gattacaggt gtgagccacc acgcccggct aattttttatt tattatttta aagacagagt	17520
ctcactctgt cactcaggct agagtgcagt ggacacatct cagctcactg cagccttgac	17580
ctccctgggc tccggtgatt tcacctccc aagtagctag gactacagc acatgccacg	17640
acaccagct aattttttat tttctgtgaa gtcaaggtct tgctacgttg cccatgctgg	17700
tatcaaacc ctgggtcaa tcaatccttc cacctcagcc tcccaagta ttggggttac	17760
aggcatgagc taccacactc agccctagcc tacttgaaac gtgttcagag catttaagtt	17820

accctacagt tgggcaaagt catctaacac aaagcccttt ttatagtaat aaaatgttgt	17880
atatctcatg tgatttatgt aatatgttta ctgaaagtga gaaacagcat ggttgcatga	17940
aaggaggcac agtcgagcca ggcacagcct gggcgagag cgagactcaa aaaaagaaaa	18000
ggccaggcgc actggctcac gcctgtaatc ccagcatttc gggaggctga ggcgggtgga	18060
tcacctgagg tcaggagttc aagaccagcc tagccaacat ggtgaaaccc cgtctctact	18120
aaaatacaaa aattaaccgg gcgtgatggc aggtgcctgt aatcccagct acttgggagg	18180
ctgaggcagg agaatcgctt gaaccaggag gcggaggttg caggagacca agatggcgcc	18240
actgcactcc agcctgggag atagagttag actccgtctc agaaaaaaaa gaaaagaaac	18300
gaggcacagt cgcatgcaca tgtagtccca gttacttgag aggctaaggc aggaggatct	18360
cttgagccca agagtttgag tccagcctga acaacatagc aagacatcat ctctaaaatt	18420
taaaaaaggg ccgggcacag tggctcacac ctgtaatccc agcactttgg gaggtggagg	18480
tgggtagatc acctgacgtc aggagttaga aaccagcctg gctaacatgg tgaagcccca	18540
tctctactaa aaacacaaaa attagccagg tgtggtagca cacgcctgta gtcccagcta	18600
ctcgggaggc tgaggcaca gaatcacttg aacccagag gcggagattg caatcagcca	18660
agattgcacc attgcactcc cgcctgggca acagagttag accccatctc aaaataaata	18720
aataaatatt tttaaaagtc agctgtatag gtacttgaag tgcagtttct actaaatgca	18780
tgttgctttt gtaccgtcat aaagtcaaac aattgtaact tgaacctct tttactcag	18840
gtactgtgta tatacttact tctcccccct cctgtttgct gcagatccgt gggcgtgagc	18900
gcttcgagat gttccgagag ctgaatgagg ccttggaact caaggatgcc caggctggga	18960

aggagccagg ggggagcagg gctcactcca ggtgagtac ctcagccct tcttgccct	19020
actccctgc ctcttaggt tggaaagcca taggattcca ttctatcct gccttcattg	19080
tcaaaggcag ctgaccccat ctcatgggt cccagccctg cacagacatt ttttagtct	19140
tctccggtt gaatcctata accacattct tgcctcagt tatccacaga acatccaaac	19200
ccaggagcga gtgtggatac ttctttgcca ttctccgaa ctcccagccc agagctggag	19260
ggtctcaagg aggggcctaa taattgtga atactgaata cagccagagt ttcagtcac	19320
atactcagcc ctgcatgca ccggcaggtc ctaggtagc cccgtcaaac tcagtttct	19380
tatatataaa atgggtaag ggggccgggc gcagtggctc acgaatcca cactctggga	19440
ggccaaggcg agtggatcac ctgaggtcgg gagtttgagc ccagcctgac caacatggag	19500
aaacccatc tctactaaa atacaaaagt agccgggctt ggtgatgcat gcctgtaac	19560
ccagctacct actcgggagg ctgaggcagg agaatcgctt gaacccggga ggcagaggtt	19620
gcggtgagct gagatctcac cattacactc cagcctgggc aacaagagt aaactccgtc	19680
tcaaaaaaga taaataaagt aaaatggggt aagggaagat tacgagacta atacacacta	19740
atactctgag gtgctcagta aacatatttg catggggtgt ggccaccatc ttgatttgaa	19800
ttccgttgt cccagccta ggcccttcaa agcattggtc agggaaaagg ggcacagacc	19860
ctctactca tgtgatgca tctctctcc ctgcttctgt ctctacagc cacctgaagt	19920
ccaaaaagg ttagtctacc tcccgcata aaaaactcat gttcaagaca gaaggcctg	19980
actcagactg acattctcca ctcttgttc cccactgaca gctcccacc cccatctctc	20040
cctccctgc cttttgggt ttgggtctt tgaaccttg ctgcaatag gtgtgcgtca	20100
gaagcaccca ggacttccat ttgcttctc cggggctcc actgaacaag ttggcctgca	20160

ctggtgtttt gttgtgggga ggaggatggg gagtaggaca taccagctta gattttaagg	20220
tttttactgt gagggatggt tgggagatgt aagaaatgtt cttgcagtta agggttagtt	20280
tacaatcagc cacattctag gtaggggccc acttcaccgt actaaccagg gaagctgtcc	20340
ctcactgttg aattttctct aacttcaagg cccatatctg tgaaatgctg gcatttgac	20400
ctacctcaca gattgcattg tgagggttaa tgaaataatg tacatctggc cttgaaacca	20460
ccttttatta catggggctc agaacttgac ccccttgagg gtgcttgctt cctctccctg	20520
ttggtcggtg ggttggtagt ttctacagtt gggcagctgg ttaggtagag ggagttgtca	20580
agtctctgct ggcccagcca aacctgtct gacaacctct tggatgaacct tagtacctaa	20640
aaggaaatct caccceatcc cacacctgg aggatttcac ctcttgata tgatgatctg	20700
gateccacaa gacttgtttt atgctcaggg tcaatttctt tttctttttt tttttttttt	20760
tttctttttc ttgagactg ggtctcgctt tgttgcccag gctggagtgg agtggcgtga	20820
tcttgctta ctgcagcctt tgcctccccg gctcgagcag tcctgcctca gcctccggag	20880
tagctgggac cacaggttca tgccacatg gccagccaac ttttgcattg ttgttagaga	20940
tggggtctca cagtgttgcc caggctggct tcaaactcct gggctcaggc gatccacctg	21000
tctcagcctc ccagagtgtt gggattacaa ttgtgagcca ccacgtccag ctggaagggt	21060
caacatcttt tacattctgc aagcacatct gcattttcac cccaccttc cctccttct	21120
ccctttttat atcccatctt tatatcgatc tcttatttta caataaaact ttgtgccac	21180
ctgtgtgtct gaggggtg	21198