

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0004060

(43) 공개일자 2022년01월11일

- (51) 국제특허분류(Int. Cl.)
C12N 15/10 (2017.01) *C07K 14/725* (2006.01)
C40B 40/08 (2006.01)
- (52) CPC특허분류
C12N 15/1093 (2013.01)
C07K 14/7051 (2013.01)
- (21) 출원번호 10-2021-7035399
- (22) 출원일자(국제) 2020년04월03일
 심사청구일자 없음
- (85) 번역문제출일자 2021년10월29일
- (86) 국제출원번호 PCT/US2020/026558
- (87) 국제공개번호 WO 2020/206238
 국제공개일자 2020년10월08일
- (30) 우선권주장
 62/829,813 2019년04월05일 미국(US)
 (뒷면에 계속)

- (71) 출원인
 루트페이스 제노믹스, 인크.
 미국 02472 매사추세츠주 워터타운 65 그로브 스트리트 스위트 203
- (72) 발명자
 천 시
 미국 02472 매사추세츠주 워터타운 그로브 스트리트 65 스위트 203
- 포터 엘리
 미국 02472 매사추세츠주 워터타운 그로브 스트리트 65 스위트 203
- (74) 대리인
 김진희, 김태홍

전체 청구항 수 : 총 58 항

(54) 발명의 명칭 T-세포 수용체 유전자 어셈블리를 위한 조성물 및 방법

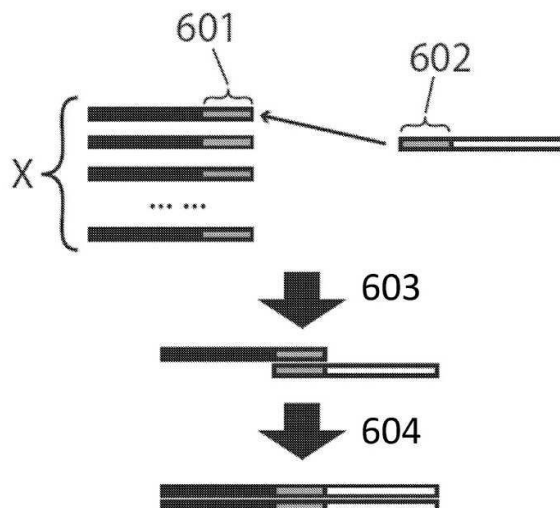
(57) 요약

T-세포 수용체 유전자 어셈블리를 위한 조성물 및 방법

본 명세서의 요약서

T-세포 수용체를 인코딩하는 핵산 서열을 어셈블링하기 위한 조성물들과 방법들이 본 명세서에서 제공된다.

대표도 - 도6



(52) CPC특허분류

C40B 40/08 (2013.01)

(30) 우선권주장

62/838,465 2019년04월25일 미국(US)

62/898,053 2019년09월10일 미국(US)

62/972,231 2020년02월10일 미국(US)

명세서

청구범위

청구항 1

핵산 분자 T-세포 수용체를 인코딩하는 (TCR)쇄 또는 이의 일부분을 생성시키는 방법에 있어서, 다음을 포함하는 방법:

- (a) TCR쇄의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 제공하고;
- (b) 핵산 분자들 다수를 제공하고, 이들 다수의 핵산 분자중에서 각 핵산 분자 TCR V 유전자로부터 유래된 서열을 포함하며, 이때 상기 다수의 핵산 분자는 적어도 두 개의 상이한 TCR V 유전자들로부터 유래된 적어도 두 개의 상이한 서열을 포함하고; 그리고
- (c) 동일한 구획 안에서 (a)의 적어도 하나의 핵산 분자에 (b)의 다수의 핵산 분자를 접촉시키고, 이때 (a)의 적어도 하나의 핵산 분자는 상기 다수 핵산 분자들중 핵산 분자에 연계시켜 CDR3을 인코딩하는 서열 및 상기 적어도 두 개의 상이한 TCR V 유전자들중 하나로부터 유래된 서열을 포함하는 제 3 핵산 분자를 만들 수 있고, 그렇게 함으로써 TCR쇄 또는 이의 일부분을 인코딩하는 핵산 분자가 생성된다.

청구항 2

청구항 1에 있어서, 이때 적어도 하나의 핵산 분자 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR쇄의 CDR3을 인코딩하는 서열을 포함하는, 방법.

청구항 3

청구항 1 또는 2에 있어서, 이때 (a)의 적어도 하나의 핵산 분자는 상기 적어도 두 개의 상이한 TCR V 유전자들중 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있는, 방법.

청구항 4

청구항 1에 있어서, 이때 상기 적어도 하나의 핵산 분자는 TCR쇄의 J 영역을 더 포함하는, 방법.

청구항 5

청구항 2에 있어서, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR쇄의 J 영역을 더 포함하는, 방법.

청구항 6

청구항 1-5중 임의의 한 항에 있어서, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TCR V 유전자 또는 마우스 TCR V 유전자인, 방법.

청구항 7

청구항 1-6중 임의의 한 항에 있어서, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 및 TRAV41로 구성된 군에서 선택되는, 방법.

청구항 8

청구항 1-6중 임의의 한 항에 있어서, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5,

TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 및 TRBV30로 구성된 군에서 선택되는, 방법.

청구항 9

청구항 1-8중 임의의 한 항에 있어서, 이때 적어도 두 개의 상이한 TCR V 유전자로부터 유래된 서열 다수에서 각 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함하는, 방법.

청구항 10

청구항 1-9중 임의의 한 항에 있어서, 이때 상기 TCR쇄는 TCR알파쇄, TCR베타쇄, TCR감마쇄, 또는 TCR델타쇄인, 방법.

청구항 11

청구항 1-10중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR쇄의 추가 CDR3을 인코딩하는 추가 서열을 더 포함하는, 방법.

청구항 12

청구항 11에 있어서, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR쇄의 추가 J 영역을 포함하는, 방법.

청구항 13

청구항 11 또는 12에 있어서, 이때 CDR3을 인코딩하는 서열과 추가 CDR3을 인코딩하는 추가 서열은 최대한 100개 뉴클레오타이드에 의해 분리되어 있는, 방법.

청구항 14

청구항 11-13중 임의의 한 항에 있어서, 이때 상기 TCR쇄와 상기 추가 TCR쇄는 TCR쇄의 동계 쌍인, 방법.

청구항 15

청구항 1-14중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자 커넥터 서열을 포함하고, 이 커넥터 서열은 상기 적어도 하나의 핵산 분자를 상기 다수 핵산 분자들중 핵산 분자에 연계시켜 제 3 핵산 분자를 생성시킬 수 있는, 방법.

청구항 16

청구항 15에 있어서, 이때 상기 적어도 하나의 핵산 분자와 상기 다수 핵산 분자들중 핵산 분자는 기능성 TCR쇄 또는 이의 일부분을 인코딩하는, 방법.

청구항 17

청구항 15 또는 16에 있어서, 이때 상기 다수 핵산 분자들중 핵산 분자 안티-커넥터 서열을 포함하고, 이 안티-커넥터 서열은 (a)의 적어도 하나의 핵산 분자의 커넥터 서열에 상보적인, 방법.

청구항 18

청구항 1-17중 임의의 한 항에 있어서, (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 연계시키는 것을 더 포함하는, 방법.

청구항 19

청구항 18에 있어서, 이때 연계는 (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 혼성화시키는 것을 포함하는, 방법.

청구항 20

청구항 19에 있어서, 이때 혼성화는 (a)의 적어도 하나의 핵산 분자의 커넥터 서열을 (b)의 다수의 핵산 분자의 핵산 분자의 안티-커넥터 서열에 혼성화시키는 것을 포함하는, 방법.

청구항 21

청구항 18-20중 임의의 한 항에 있어서, (i) (a)의 적어도 하나의 핵산 분자를 주형으로 삼아 상기 다수 핵산 분자들중 핵산 분자의 자유 3' 단부를 연장시키고, 및/또는 (ii) 상기 다수 핵산 분자들중 핵산 분자를 주형으로 삼아 (a)의 적어도 하나의 핵산 분자의 자유 3' 단부를 연장시켜, 제 3 핵산 분자를 만드는 것을 더 포함하는, 방법.

청구항 22

청구항 1-21중 임의의 한 항에 있어서, (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 결합시키는 것을 더 포함하는, 방법.

청구항 23

청구항 1-22중 임의의 한 항에 있어서, 제 3 핵산 분자에 제한절단 효소를 접촉시켜, 스티커 단부를 만드는 것을 더 포함하는, 방법.

청구항 24

청구항 1-23중 임의의 한 항에 있어서, 제 3 핵산 분자에 추가 핵산 분자를 접촉시키는 것을 더 포함하는, 방법.

청구항 25

청구항 24에 있어서, 이때 상기 추가 핵산 분자는 TCR 채의 불변 영역 또는 이의 일부분을 인코딩하는, 방법.

청구항 26

청구항 24 또는 25에 있어서, 제 3 핵산 분자와 상기 추가 핵산 분자를 결합시키는 것을 더 포함하는, 방법.

청구항 27

청구항 1-26중 임의의 한 항에 있어서, 이때 핵산 분자들 다수 (이때, 각 핵산 분자는 상이한 TCR 채 또는 이의 일부분을 인코딩하는)가 동일한 구획 안에서 생성되는, 방법.

청구항 28

청구항 27에 있어서, 이때 상기 핵산 분자들 다수에서 적어도 5가지 상이한 핵산 분자가 상기 동일한 구획 안에서 생성되는, 방법.

청구항 29

청구항 1-26중 임의의 한 항에 있어서, 이때 상기 다수의 핵산 분자중 적어도 10개 상이한 핵산 분자가 상기 동일한 구획 안에서 생성되는, 방법.

청구항 30

청구항 1-29중 임의의 한 항에 있어서, 이때 상기 동일한 구획은 웰, 튜브, 또는 소적인, 방법.

청구항 31

청구항 1-30중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자 특유 바코드를 포함하는, 방법.

청구항 32

청구항 31에 있어서, 이때 상기 특유 바코드는 프라이머 결합 부위인, 방법.

청구항 33

청구항 15-30중 임의의 한 항에 있어서, 이때 상기 커넥터 서열은 특유 바코드를 포함하는 방법.

청구항 34

청구항 33에 있어서, 이때 상기 특유 바코드는 프라이머 결합 부위인, 방법.

청구항 35

(a) 핵산 분자들 다수를 포함하는 조성물에 있어서, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함하고, CDR3 서열을 포함하지 않고, 이때 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 안티-커넥터 서열을 포함하고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 안티-커넥터 서열을 포함하고, 이때 제 1 안티-커넥터 서열은 제 2 안티-커넥터 서열과는 상이하며, 이때 제 1 핵산 분자의 TCR V 유전자로부터 유래된 서열과 제 2 핵산 분자는 상이한 TCR V 유전자로부터 유래되고; 그리고

(b) TCR 쉘의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자, 이때 상기 적어도 하나의 핵산 분자는 제 1 안티-커넥터 서열에 상보적인 제 1 커넥터 서열을 더 포함하는, 조성물.

청구항 36

청구항 35에 있어서, 이때 상기 조성물은 액체 조성물인 조성물.

청구항 37

청구항 35 또는 36에 있어서, 이때 상기 (a)의 핵산 분자 다수와 상기 (b)의 적어도 하나의 핵산 분자는 동일한 구획 안에 있는, 조성물.

청구항 38

청구항 35-37중 임의의 한 항에 있어서, 이때 상기 TCR V 유전자로부터 유래된 서열은 TCR V 유전자의 적어도 10개 뉴클레오티드를 포함하는, 조성물.

청구항 39

청구항 35-38중 임의의 한 항에 있어서, 이때 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자인, 조성물.

청구항 40

청구항 35-39중 임의의 한 항에 있어서, 이때 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함하는, 조성물.

청구항 41

청구항 35-40중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자는 TCR 쉘의 J 영역을 더 포함하는, 조성물.

청구항 42

청구항 35-41중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR 쉘의 추가 CDR3을 인코딩하는 추가 서열을 더 포함하는, 조성물.

청구항 43

청구항 42에 있어서, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR 쉘의 추가 J 영역을 더 포함하는, 조성물.

청구항 44

청구항 42 또는 43에 있어서, 이때 CDR3을 인코딩하는 서열과 CDR3을 인코딩하는 추가 서열은 최대한 100개 뉴클레오티드에 의해 분리되어 있는, 조성물.

청구항 45

청구항 42-44중 임의의 한 항에 있어서, 이때 상기 TCR 쉘과 상기 추가 TCR 쉘은 TCR 쉘의 동계 쌍인, 조성물.

청구항 46

청구항 35-45중 임의의 한 항에 있어서, 이때 상기 (b)의 적어도 하나의 핵산 분자 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 체계의 CDR3를 인코딩하는 서열을 포함하는, 조성물.

청구항 47

청구항 46에 있어서, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 TCR 체계의 상이한 CDR3를 인코딩하는, 조성물.

청구항 48

청구항 46 또는 47에 있어서, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자 상이한 커넥터 서열을 포함하고, 이때 상이한 커넥터 서열은 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있는, 조성물.

청구항 49

청구항 35-48중 임의의 한 항에 있어서, 이때 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 TCR V 유전자 서열을 포함하는, 조성물.

청구항 50

청구항 49에 있어서, 이때 상기 TCR V 유전자 서열은 재배열된 유전자에서 상기 CDR3를 인코딩하는 서열에 인접한 TCR V 유전자에서 적어도 3개 뉴클레오타이드를 포함하는, 조성물.

청구항 51

청구항 35-50중 임의의 한 항에 있어서, 이때 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 사전-결정된 서열을 포함하는, 조성물.

청구항 52

청구항 35-51중 임의의 한 항에 있어서, 이때 제 1 커넥터 서열은 제 1 안티-커넥터 서열에 혼성화되는, 조성물.

청구항 53

청구항 35-52중 임의의 한 항에 있어서, 이때 상기 (b)의 적어도 하나의 핵산 분자는 특유 바코드를 포함하는, 조성물.

청구항 54

청구항 53에 있어서, 이때 상기 특유 바코드는 프라이머 결합 부위인, 조성물.

청구항 55

청구항 35-52중 임의의 한 항에 있어서, 이때 상기 적어도 하나의 핵산 분자의 제 1 커넥터 서열은 특유 바코드를 포함하는, 조성물.

청구항 56

청구항 55에 있어서, 이때 상기 특유 바코드는 프라이머 결합 부위인, 조성물.

청구항 57

핵산 분자들 다수를 생성시키는, 다음을 포함하는 방법:

(a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 체계의 제 1 CDR3 및 제 2 TCR 체계의 제 2 CDR3를 인코딩하는 서열을 포함하고, 이때 제 1 CDR3과 제 2

CDR3은 TCR 쉐의 동계 쌍으로부터 유래되며;

(b) 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고

(c) 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자에 연계되어, 제 1 CDR3과 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 인코딩하는 서열을 포함하는 선형의 핵산 분자를 형성하고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 상기 TCR 쉐의 동계 쌍으로부터 유래된다.

청구항 58

핵산 분자들 다수를 생성시키는, 다음을 포함하는 방법:

(a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 (i) 제 1 T-세포 수용체 (TCR) 쉐의 제 1 CDR3 및 제 2 TCR 쉐의 제 2 CDR3을 인코딩하는 합성 서열, 그리고 (ii) 제 3 T-세포 수용체 (TCR) 쉐의 제 3의 CDR3 및 제 4의 TCR 쉐의 제 4의 CDR3을 인코딩하는 합성 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 제 1 TCR 쉐의 동계 쌍으로부터 유래되며, 제 3의 CDR3 및 제 4의 CDR3은 제 2 TCR 쉐의 동계 쌍으로부터 유래되며;

(b) 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고

(c) 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자에 연계되어, 제 1 CDR3과 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 인코딩하는 서열을 포함하는 핵산 분자를 형성하고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 상기 TCR 쉐의 동계 쌍으로부터 유래된다.

발명의 설명

기술 분야

상호-참조

본 출원은 2019년 4월 5일자로 제출된 미국 가특허 출원 번호 62/829,813, 2019년 4월 25일자로 제출된 미국 가특허 출원 번호 62/838,465, 2019년 9월 10일자로 제출된 미국 가특허 출원 번호 62/898,053 및 2020년 2월 10일자로 제출된 미국 가특허 출원 번호 62/972,231에 대해 우선권을 주장하며, 이들 출원은 이들 전문이 여기에 참고자료로 편입된다.

서열 목록

본 출원에는 ASCII 형식으로 전자적으로 제출된 서열 목록의 전문을 본원의 참고자료에 편입되어 있다. 2020년 5월 15일자로 만들어진 전술한 ASCII 사본의 이름은 53563-705_601_SL.txt이며, 크기는 69,187 바이트이다.

배경 기술

발명의 배경

T-세포 수용체 (TCR)는 항원-주요 조직적합성 복합체의 인지를 담당하여, 염증 반응의 개시를 유도할 수 있다. 세포독성 T 세포들과 헬퍼 T 세포들을 비롯한, 많은 T 세포 하위집단이 존재한다. 세포독성 T 세포들 (CD8+ T 세포로 또한 알려짐)은 비정상적 세포들, 예를 들면, 바이러스-감염된 세포들 또는 종양 세포들을 죽인다. 헬퍼 T 세포들 (CD4+ T 세포로 또한 알려짐)은 다른 면역 세포들의 활성화 및 성숙을 지원한다. 세포독성 세포들과 헬퍼 T 세포들은 모두 각각의 반응을 촉발하는 특이적 표적 항원을 인지한 후 기능을 수행한다. T 세포의 항원 특이성은 T 세포 표면에 발현된 TCR에 의해 특정될 수 있다. T 세포 수용체는 2개의 폴리펩티드 쉐 (가장 일반적으로 알파 쉐와 베타 쉐이지만, 소수의 T 세포는 감마 쉐 및 델타 쉐를 발현시킬 수 있다)로 구성된 이종이량체 단백질이다. TCR의 특정 아미노산 서열과 생성된 3차원 구조는 TCR 항원 특이성과 친화력을 특정한다. 모든 개별 T 세포에 대한 TCR 쉐들의 아미노산 서열 및 코딩 DNA 서열은 거의 항상 독특하거나, 또는 유기체의 전체

TCR 레퍼토리에서 매우 적은 풍도로 존재하는데, 그 이유는 가능한 TCR 서열이 매우 많기 때문이다. 이러한 방대한 서열 다양성은 여러 세포 메커니즘을 통해 T 세포 발달 동안 달성될 수 있으며, 다양한 잠재적 항원에 반응하는 면역계 능력의 중요한 측면이 될 수 있다.

[0007]

TCR 레퍼토리를 분석하면 면역 체계의 속성과 질병의 병인 및 진행, 특히 알려지지 않은 항원이 촉발시키는 질환을 더 잘 이해하는 데 도움이 될 수 있다. TCR 레퍼토리의 극단적인 다양성과 TCR의 이분법적 속성은 분석에서 주된 극복과제가 될 수 있다. 고-처리량 시퀀싱은 비록 단편적 길이 분석 (spectratyping)보다 더 많은 비용이 들지만 더 큰 시퀀싱 깊이와 TCR 클론형 풍도의 훨씬 더 정확한 정량화가 가능할 수 있다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0008]

본 발명의 요약

[0009]

태생적으로 쌍을 이룬 T-세포 수용체들 (TCRs) (또는 동계 TCR 쌍)을 인코딩하는 핵산 서열을 어셈블리하기 위한 조성물 및 방법들이 본원에서 제공된다. 예를 들면, TCR은 TCR 알파쇄 및 TCR 베타쇄를 포함할 수 있거나, 또는 TCR은 TCR 감마쇄 및 TCR 델타쇄를 포함할 수 있다. 단일 세포 바코딩 및 시퀀싱 기술을 이용하여, 이에 국한되지 않는 다양한 방법들을 이용하여 태생적으로 쌍을 이룬 TCRs를 인코딩하는 서열을 식별해낼 수 있다. 상기 태생적으로 쌍을 이룬 TCRs를 인코딩하는 서열을 획득한 후, 본원에 기술된 조성물들 및 방법들을 이용하여 임의의 주어진 숙주 세포(들)에서 상기 태생적으로 쌍을 이룬 TCRs를 발현시키는 하나 또는 그 이상의 핵산 서열을 신속하고, 높은 처리량으로, 비용-효율적인 방식으로 구축하거나, 또는 어셈블리할 수 있다. 상기 하나 또는 그 이상의 핵산 서열은 약 1, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 TCRs를 인코딩하는 상이한 서열을 포함할 수 있다.

[0010]

한 측면에서, 본 명세서는 T-세포 수용체 (TCR)쇄 또는 이의 일부분을 인코딩하는 핵산 분자를 만드는 방법을 제공하며, 이 방법은 다음을 포함한다: (a) TCR쇄의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 제공하고; (b) 다수의 핵산 분자들을 제공하고, 이들 다수의 핵산 분자 각각은 TCR V 유전자에서 유래된 서열을 포함하며, 이때 상기 다수 핵산 분자는 적어도 두 개의 상이한 TCR V 유전자들로부터 유래된 적어도 두 개의 상이한 서열을 포함하고; 그리고 (c) 동일한 구획(compartment)에서 (a)의 적어도 하나의 핵산 분자를 (b)의 다수의 핵산 분자에 접촉시키고, 이때 (a)의 적어도 하나의 핵산 분자는 상기 다수 핵산 분자들중 핵산 분자에 연계되어, CDR3을 인코딩하는 서열 및 상기 적어도 두 개의 상이한 TCR V 유전자들중 하나로부터 유래된 서열을 포함하는 제 3 핵산 분자를 만들 수 있고, 그렇게 함으로써 상기 TCR쇄 또는 이의 일부분을 인코딩하는 핵산 분자가 생성된다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 적어도 약 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열을 포함한다. 일부 구체예들에서, 상기 다수의 핵산 분자 (이들 다수의 핵산 분자들중에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고)는 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 상이한 TCR V 유전자들로부터 유래된 상이한 서열을 포함한다.

[0011]

일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR쇄의 CDR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, (a)의 적어도 하나의 핵산 분자는 상기 적어도 두 개의 상이한 TCR V 유전자들중 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 상기 TCR쇄의 J 영역을 더 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR쇄의 J 영역을 더 포함한다. 일부 구체예들에서, 상기 적어도 두 개의 TCR V 유전자는 인간 TCR V 유전자들 또는 마우스 TCR V 유전자들이다. 일부 구체예들에서, 상기 적어도 두 개의 TCR V

유전자는 인간 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 및 TRAV41로 구성된 군에서 선택된다. 일부 구체예들에서, 상기 적어도 두 개의 TCR V 유전자는 인간 TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 및 TRBV30으로 구성된 군에서 선택된다. 일부 구체예들에서, 상기 적어도 두 개의 상이한 TCR V 유전자로부터 유래된 서열 다수에서 각 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 TCR쇄는 TCR알파쇄, TCR베타쇄, TCR감마쇄, 또는 TCR델타쇄다. 일부 구체예들에서, 상기 적어도 하나의 핵산분자는 추가 TCR쇄의 추가 CDR3을 인코딩하는 추가 서열을 더 포함한다. 일부 구체예들에서, 상기 적어도 하나의 핵산분자는 상기 추가 TCR쇄의 추가 J영역을 포함한다. 일부 구체예들에서, CDR3을 인코딩하는 서열 및 추가 CDR3을 인코딩하는 추가 서열은 최대한 100개 뉴클레오타이드에 의해 분리된다. 일부 구체예들에서, 상기 TCR쇄 및 추가 TCR쇄는 TCR쇄의 동계쌍이다. 일부 구체예들에서, 상기 적어도 하나의 핵산분자는 커넥터(connector) 서열을 포함하며, 이 커넥터 서열은 상기 적어도 하나의 핵산분자를 상기 다수 핵산분자들중 핵산분자에 연계시켜 제 3 핵산분자를 생성시킬 수 있다. 일부 구체예들에서, 상기 적어도 하나의 핵산분자 및 상기 다수 핵산분자들중 핵산분자는 기능성 TCR쇄 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 다수 핵산분자들중 핵산분자는 안티-커넥터(anti-connector) 서열을 포함하며, 이 안티-커넥터 서열은 (a)의 적어도 하나의 핵산분자의 커넥터 서열에 상보적이다. 일부 구체예들에서, 상기 방법은 (a)의 적어도 하나의 핵산분자와 (b)의 다수의 핵산분자의 핵산분자를 연계시키는 것을 더 포함한다. 일부 구체예들에서, 연계(linking)는 (a)의 적어도 하나의 핵산분자와 (b)의 다수의 핵산분자의 핵산분자를 혼성화시키는 것을 포함한다. 일부 구체예들에서, 혼성화(hybridizing)는 (a)의 적어도 하나의 핵산분자의 커넥터 서열을 (b)의 다수의 핵산분자의 핵산분자의 안티-커넥터 서열에 혼성화시키는 것을 포함한다. 일부 구체예들에서, 상기 방법은 (i) (a)의 적어도 하나의 핵산분자를 주형으로 삼아 상기 다수 핵산분자들중 핵산분자의 자유 3' 단부를 연장시키고, 및/또는 (ii) 상기 다수 핵산분자들중 핵산분자를 주형으로 삼아 (a)의 적어도 하나의 핵산분자의 자유 3' 단부를 연장시켜, 제 3 핵산분자를 만드는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 (a)의 적어도 하나의 핵산분자와 상기 다수 핵산분자들중 핵산분자 (b)에 결합시키는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 3 핵산분자에 제한절단 효소를 접촉시켜, 스틱키(sticky) 단부를 만드는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 3 핵산분자에 추가 핵산분자를 접촉시키는 것을 더 포함한다. 일부 구체예들에서, 상기 추가 핵산분자는 TCR쇄의 불변 영역 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 방법은 제 3 핵산분자와 상기 추가 핵산분자를 결합시키는 것을 더 포함한다. 일부 구체예들에서, 핵산분자들 다수(이때, 각 핵산분자는 상이한 TCR쇄 또는 이의 일부분을 인코딩하는)가 동일한 구획 안에서 생성된다. 일부 구체예들에서, 상기 핵산분자들 다수에서 적어도 5가지 상이한 핵산분자가 동일한 구획 안에서 생성된다. 일부 구체예들에서, 상기 핵산분자들 다수에서 적어도 10가지 상이한 핵산분자가 동일한 구획 안에서 생성된다. 일부 구체예들에서, 상기 동일한 구획 안에 상기 핵산분자 다수에서 적어도 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 핵산분자들이 생성된다. 일부 구체예들에서, 상기 동일한 구획은 웰, 튜브, 또는 소적이다. 일부 구체예들에서, 상기 적어도 하나의 핵산분자는 특유 바코드를 포함한다. 일부 구체예들에서, 상기 특유 바코드는 프라이머 결합 부위다. 일부 구체예들에서, 상기 커넥터 서열은 특유 바코드를 포함한다. 일부 구체예들에서, 상기 특유 바코드는 프라이머 결합 부위다.

[0012]

또다른 측면에서, 본 명세서는 다음을 포함하는 조성물을 제공한다: (a) 핵산분자들 다수, 이때 상기 다수 핵산분자들중 각 핵산분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함하고, CDR3 서열을 포함하지 않고, 이때 상기 다수 핵산분자들중 제 1 핵산분자는 제 1 안티-커넥터 서열을 포함하고, 상기 다수 핵산분자들중 제 2 핵산분자는 제 2 안티-커넥터 서열을 포함하고, 이때 제 1 안티-커넥터 서열은 제 2 안티-커넥터 서열과는 상이하며, 이때 제 1 핵산분자의 TCR V 유전자로부터 유래된 서열과 제 2 핵산분자는 상이한 TCR V 유전자로부터 유래되고; 그리고 (b) TCR쇄의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산분자, 이때 상기 적어도 하나의 핵산분자는 제 1 커넥터 서열을 더 포함하며, 이 커넥터 서열은 제 1 안티-커넥터 서

열에 상보적이다.

[0013]

일부 구체예들에서, 상기 조성물은 액체 조성물이다. 일부 구체예들에서, 상기 (a)의 핵산 분자 다수와 상기 (b)의 적어도 하나의 핵산 분자는 동일한 구획 안에 있다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 이 TCR V 유전자의 적어도 10개 뉴클레오타이드를 포함한다. 일부 구체예들에서, 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 상기 TCR 쉼의 J 영역을 더 포함한다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 추가 TCR 쉼의 추가 CDR3을 인코딩하는 추가 서열을 더 포함한다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자는 상기 추가 TCR 쉼의 추가 J 영역을 더 포함한다. 일부 구체예들에서, CDR3을 인코딩하는 서열과 CDR3을 인코딩하는 추가 서열은 최대한 100개 뉴클레오타이드에 의해 분리된다. 일부 구체예들에서, 상기 TCR 쉼 및 추가 TCR 쉼은 TCR 쉼의 동계 쌍이다. 일부 구체예들에서, 상기 (b)의 적어도 하나의 핵산 분자 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 쉼의 CDR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 TCR 쉼의 상이한 CDR3을 인코딩한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 커넥터 서열을 포함하고, 이때 상이한 커넥터 서열은 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있다. 일부 구체예들에서, 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 TCR V 유전자 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자 서열은 재배열된(rearranged) 유전자에서 상기 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자에서 적어도 3개 뉴클레오타이드를 포함한다. 일부 구체예들에서, 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 사전-결정된(pre-determined) 서열을 포함한다. 일부 구체예들에서, 제 1 커넥터 서열은 제 1 안티-커넥터 서열에 혼성화된다. 일부 구체예들에서, 상기 (b)의 적어도 하나의 핵산 분자는 특유 바코드를 포함한다. 일부 구체예들에서, 상기 특유 바코드는 프라이머 결합 부위다. 일부 구체예들에서, 상기 적어도 하나의 핵산 분자의 제 1 커넥터 서열은 특유 바코드를 포함한다. 일부 구체예들에서, 상기 특유 바코드는 프라이머 결합 부위다.

[0014]

또다른 측면에서, 본 명세서는 핵산 분자들 다수를 생성시키는 방법을 제공하는데, 이 방법은 다음을 포함한다: 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래되고; 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고, 이때 핵산 분자는 불변 도메인을 인코딩하는 서열을 포함하지 않고; 그리고 제 1 다수 핵산분자들에서 핵산 분자와 제 2 다수 핵산 분자들에서 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 형성하게 되고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 TCR V 유전자는 TCR 쉼의 동계 쌍으로부터 유래된다.

[0015]

일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 TCR 쉼의 상이한 제 1 CDR3 및/또는 제 2 TCR 쉼의 상이한 CDR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자는 적어도 약 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들은 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 상이한 TCR V 유전자들을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들은 동일한 구획 안에서 접촉된다. 일부 구체예들에서, 제 1 다수의 핵산 분자들에서 핵산 분자는 커넥터 서열을 더 포함하고, 이때 상기 커넥터 서열은 제 1 다수의 핵산 분자들에서 핵산 분자와 제 2 다수의 핵산 분자들에서 핵산 분자를 연계한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 핵산 분자는 안티-커넥터 서열을 더 포함하고, 안티-커넥터 서열은 상기 커넥터 서열에 상보적이다. 일부 구체예들에서, 상기 커넥터 서열은 안티-커넥터 서열에 혼성화되어, 제 1 다수의 핵산 분자들에서 핵산 분자와 제 2 다수의 핵산 분자들에서 핵산 분자를 연계시킨다. 일부 구체예들에서, 상기 커넥터 서열은 코돈-다각화되어, 제 1 다수의 핵산 분자들에서 핵산 분자의 커넥터 서열은 제 1 다수의 핵산 분자들에서 다른 핵산 분자들의 다른 커넥터 서열과는 상이하다. 일부 구체예들에서, 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 TCR 쉼의 제 1 J 영역 및/또는 제 2

TCR 쉼의 제 2 J 영역을 더 포함한다. 일부 구체예들에서, (i) 제 1 TCR 쉼은 TCR 알파 쉼이고, 제 2 TCR 쉼은 TCR 베타 쉼이거나, 또는 (ii) 제 1 TCR 쉼은 TCR 감마 쉼이고, 제 2 TCR 쉼은 TCR 델타 쉼이다. 일부 구체예들에서, 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 핵산 분자는 이중-가닥으로 된 핵산 분자다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 핵산 분자는 자가-절단 캡티드의 일부분을 인코딩하는 서열을 더 포함한다. 일부 구체예들에서, 상기 안티-커넥터 서열은 제 2 다수의 핵산 분자들에서 핵산 분자의 오버행 (overhang)이다. 일부 구체예들에서, 상기 커넥터 서열 또는 상기 안티-커넥터 서열의 길이는 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 뉴클레오티드이다. 일부 구체예들에서, 상기 방법은 (i) 제 2 다수의 핵산 분자들에서 핵산 분자에 혼성화된 제 1 다수의 핵산 분자들에서 핵산 분자의 3' 단부를 연장시키거나, 및/또는 (ii) 제 1 다수의 핵산 분자들에서 핵산 분자에 혼성화된 제 2 다수의 핵산 분자들에서 핵산 분자의 3' 단부를 연장시키는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 1 다수의 핵산 분자들에서 핵산 분자에 제 2 다수의 핵산 분자들에서 핵산 분자를 결합시키는 것을 더 포함한다.

[0016] 일부 구체예들에서, 상기 방법은 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자에 제한절단 효소를 접촉시켜, 스티커 단부를 만드는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자에 불변 영역 또는 이의 일부분을 인코딩하는 서열을 인코딩하는 추가 핵산 분자를 접촉시키는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 상기 추가 핵산 분자에 상기 스티커 단부를 통하여 결합시키는 것을 더 포함한다. 일부 구체예들에서, 제 1 CDR3을 인코딩하는 서열과 제 2 CDR3을 인코딩하는 제 2 서열은 최대한 약 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 또는 5개 뉴클레오티드에 의해 분리된다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.

[0017] 또다른 측면에서, 본 명세서는 다음을 포함하는 조성물을 제공한다: 제 1 다수의 핵산 분자, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3과 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래되며; 그리고 제 2 다수의 핵산 분자들, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열을 포함하지 않고; 이때 (i) 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 제 1 CDR3 및/또는 제 2 CDR3을 인코딩하는 서열을 포함하고, 및/또는 (ii) 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자는 적어도 약 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들은 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 상이한 TCR V 유전자들을 포함한다.

[0018] 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자 커넥터 서열을 더 포함하고, 이때 주어진 커넥터 서열은 제 1 다수의 핵산 분자들에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 연결시키는데 이용가능하다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 각 핵산 분자 안티-커넥터 서열을 더 포함하고, 이 안티-커넥터 서열은 상기 커넥터 서열에 상보적이다. 일부 구체예들에서, 상기 커넥터 서열은 코돈-다각화되어, 제 1 다수의 핵산 분자들에서 주어진 핵산 분자의 주어진 커넥터 서열은 제 1 다수의 핵산 분자들에서 다른 핵산 분자들의 다른 커넥터 서열들과는 상이하다. 일부 구체예들에서, 상기 커넥터 서열은 아미노산 서열을 인코딩한다. 일부 구체예들에서, 상기 커넥터 서열은 제 1 TCR 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열과 프레임 내(in frame)에 있다. 일부 구체예들에서, 상기 커넥터 서열은 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 뉴클레오티드를 포함한다. 일부 구체예들에서, 상기 커넥터 서열은 제 1 TCR 쉼의 제 1 CDR3 또는 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열에 인접한 상기 TCR V 유전자의 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13,

14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 뉴클레오티드를 포함한다. 일부 구체예들에서, 상기 주어진 커넥터 서열에 의해 인코딩된 주어진 아미노산 서열은 적어도 하나의 다른 커넥터 서열에 의해 인코딩된 적어도 하나의 다른 아미노산 서열과 동일하거나, 또는 실질적으로 동일하다. 일부 구체예들에서, 상기 주어진 커넥터 서열에 의해 인코딩된 주어진 아미노산 서열은 다른 커넥터 서열에 의해 인코딩된 다른 아미노산 서열과 상이하다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 TCR 쉼의 제 1 J 영역 및/또는 제 2 TCR 쉼의 제 2 J 영역을 더 포함한다. 일부 구체예들에서, 상기 조성물은 액체 조성물이다. 일부 구체예들에서, 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들은 동일한 구획 안에 있다. 일부 구체예들에서, 상기 제 1 다수의 핵산 분자들에서 주어진 핵산 분자는 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 주어진 커넥터 서열을 통하여 연계된다. 일부 구체예들에서, 상기 제 1 다수의 핵산 분자에서 주어진 핵산 분자는 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 주어진 안티-커넥터 서열에 혼성화된 주어진 커넥터 서열을 통하여 혼성화된다. 일부 구체예들에서, 제 1 CDR3을 인코딩하는 서열과 제 2 CDR3을 인코딩하는 서열은 최대한 100개 뉴클레오티드에 의해 분리된다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자들에서 또는 제 2 다수의 분자들에서 각 핵산 분자는 화학적으로 합성된다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자의 길이는 최대한 약 250, 240, 230, 220, 210, 200, 190, 180, 170, 160, 150, 140, 130, 120, 110, 100, 또는 50개 뉴클레오티드이다.

[0019] 또다른 측면에서, 본 명세서는 핵산 분자들 다수를 포함하는 조성물을 제공하며, 이때 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열 서열을 포함하며, 이때 상기 다수의 핵산 분자는 제 1 커넥터 서열을 갖는 제 1 핵산 분자와 제 2 커넥터 서열을 갖는 제 2 핵산 분자를 포함하며, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하다.

[0020] 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 커넥터 서열을 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하지 않는다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 불변 도메인을 인코딩하는 서열을 포함하지 않는다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 TCR V 유전자의 적어도 10, 20, 30, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000개, 또는 그 이상의 뉴클레오티드를 포함한다. 일부 구체예들에서, 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자다.

[0021] 또다른 측면에서, 본 명세서는 핵산 분자들 다수를 포함하는 조성물을 제공하며, 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) 쉼의 CDR3 또는 이의 일부분을 인코딩하고, 이때 상기 다수의 핵산 분자 제 1 커넥터 서열을 갖는 제 1 핵산 분자와 제 2 커넥터 서열을 갖는 제 2 핵산 분자를 포함하고, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하다.

[0022] 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 J 영역을 더 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쉼의 제 1 CDR3 또는 이의 일부분과 제 2 TCR 쉼의 제 2 CDR3 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 TCR 쉼의 상이한 CDR3 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 커넥터 서열을 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR V 유전자의 200, 150, 100, 80, 50, 40, 30, 20, 또는 10개 이상의 뉴클레오티드를 포함하지 않는다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 불변 도메인을 인코딩하는 서열을 포함하지 않는다. 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 재배열된 유전자에서 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자의 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 뉴클레오티드를 포함한다. 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 사전-결정된 서열을 포함한다. 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자 서열에 상보적인 서열을 포함한다. 일부 구체예들에서, 상기 조성물 제 2

다수의 핵산 분자들을 더 포함하며, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 제 2 다수 핵산 분자들에서 제 1 핵산 분자는 제 1 안티-커넥터 서열을 포함하고, 이때 제 1 안티-커넥터 서열은 제 1 커넥터 서열에 상보적이다. 일부 구체예들에서, 제 2 다수 핵산 분자들에서 제 2 핵산 분자는 제 2 안티-커넥터 서열을 포함하고, 이때 제 2 안티-커넥터 서열은 제 2 커넥터 서열에 상보적이다. 일부 구체예들에서, 제 2 다수 핵산 분자들에서 제 1 핵산 분자의 제 1 안티-커넥터 서열은 제 1 다수 핵산 분자들에서 제 1 핵산 분자의 제 1 커넥터 서열에 연계된다. 일부 구체예들에서, 제 2 다수 핵산 분자들에서 제 2 핵산 분자의 제 2 안티-커넥터 서열은 제 1 다수 핵산 분자들에서 제 2 핵산 분자의 제 2 커넥터 서열에 연계된다.

[0023] 또다른 측면에서, 본 명세서는 핵산 분자들 다수를 포함하는 조성물을 제공하며, 각 분자들은 T-세포 수용체 (TCR) 쇄의 적어도 10개 아미노산 (가령, 일부 경우들에서, 적어도 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200개, 또는 그 이상의 아미노산을 인코딩하는)을 포함하고, 이때 상기 다수의 핵산 분자 제 1 커넥터 서열을 갖는 제 1 핵산 분자와 제 2 커넥터 서열을 갖는 제 2 핵산 분자를 포함하고, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하며, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쇄의 일부분을 인코딩하고, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쇄의 적어도 10개 (가령, 일부 경우들에서, 적어도 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 100, 200개, 또는 그 이상의 아미노산) 아미노산을 인코딩하는 서열과 프레임 안에 있다.

[0024] 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쇄 유전자의 적어도 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 인접 뉴클레오타이드를 포함하고, TCR 쇄의 적어도 10개 아미노산을 인코딩하는 서열과 프레임 안에 있다. 일부 구체예들에서, 제 1 커넥터 서열과 제 2 커넥터 서열은 TCR 쇄의 적어도 두 개의 인접 아미노산을 인코딩한다. 일부 구체예들에서, 상기 TCR 쇄의 일부분의 TCR 쇄와 적어도 10개 아미노산을 인코딩하는 서열에 의해 인코딩되는 TCR 쇄는 동일하다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상기 TCR 쇄의 CDR3 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 상기 TCR 쇄의 J 영역을 더 포함한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쇄의 제 1 CDR3 또는 이의 일부분과 제 2 TCR 쇄의 제 2 CDR3 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쇄의 제 1 J 영역과 제 2 TCR 쇄의 제 2 J 영역을 더 포함한다. 일부 구체예들에서, 제 1 CDR3 또는 이의 일부분을 인코딩하는 서열과 제 2 CDR3 또는 이의 일부분을 인코딩하는 서열은 최대한 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 또는 5개 뉴클레오타이드에 의해 분리된다. 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 제 1 커넥터 서열 또는 제 2 커넥터 서열은 사전-결정된 서열을 포함한다. 일부 구체예들에서, 제 1 커넥터 서열은 제 2 커넥터 서열의 뉴클레오타이드와는 상이한 적어도 하나의 뉴클레오타이드를 포함한다. 일부 구체예들에서, 제 1 커넥터 서열은 제 2 커넥터 서열과 동일한 아미노산 서열을 인코딩한다. 일부 구체예들에서, 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이한 아미노산 서열을 인코딩한다.

[0025] 또다른 측면에서, 본 명세서는 핵산 분자들 다수를 생성시키는 방법을 제공하며, 이들 다수의 핵산 분자들중에서 각 핵산 분자는 T-세포 수용체 (TCR) 쇄 또는 이의 영역을 인코딩하고, 상기 방법은 다음을 포함한다: 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들을 접촉시켜, 적어도 두 개의 (가령, 적어도 약 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000, 또는 그 이상의) 상이한 핵산 분자를 포함하는 제 3의 다수 핵산 분자를 생성시키고, 이때 상기 적어도 두 개의 상이한 핵산 분자 각각은 상이한 TCR 쇄 또는 이의 영역을 인코딩하는 상이한 서열을 갖고, 이때 상기 적어도 두 개의 상이한 핵산 분자는 동일한 구획 안에서 생성된다.

[0026] 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 상기 TCR 쇄의 CDR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 상기 TCR 쇄의 J 영역을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상기 TCR 쇄의 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자는 인간 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14,

TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 또는 TRAV41이다. 일부 구체예들에서, 상기 TCR V 유전자는 인간 TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 또는 TRBV30이다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다. 일부 구체예들에서, 상기 TCR쇄는 TCR알파쇄, TCR베타쇄, TCR감마쇄, 또는 TCR델타쇄다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 추가 TCR쇄의 추가 CDR3을 인코딩하는 추가 서열을 더 포함한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 각 핵산 분자는 추가 TCR쇄의 추가 J영역을 포함한다. 일부 구체예들에서, 상기 TCR쇄 및 추가 TCR쇄는 TCR쇄의 동계 쌍이다. 일부 구체예들에서, 상기 다수 핵산 분자들중 핵산 분자는 상이한 TCR 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 주어진 핵산 분자는 커넥터 서열을 포함하고, 이때 커넥터 서열은 제 1 다수의 핵산 분자에서 주어진 핵산 분자를 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 연계시키는데 이용 가능하다. 일부 구체예들에서, 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자는 기능성 TCR쇄 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들에서 주어진 핵산 분자 안티-커넥터 서열을 포함하고, 이때 안티-커넥터 서열은 제 1 다수의 핵산 분자들에서 주어진 핵산 분자의 커넥터 서열에 상보적이다. 일부 구체예들에서, 상기 방법은 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 연계시키는 것을 더 포함한다. 일부 구체예들에서, 연계는 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자의 혼성화를 포함한다. 일부 구체예들에서, 혼성화는 제 1 다수의 핵산 분자에서 주어진 핵산 분자의 커넥터 서열을 제 2 다수의 핵산 분자들에서 주어진 핵산 분자의 상기 안티-커넥터 서열에 혼성화시키는 것을 포함한다. 일부 구체예들에서, 상기 방법은 (i) 제 1 다수의 핵산 분자에서 주어진 핵산 분자를 주형으로 하여 제 2 다수의 핵산 분자들에서 주어진 핵산 분자의 자유 3' 단부를 연장시키거나, 및/또는 (ii) 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 주형으로 하여 제 1 다수의 핵산 분자들에서 핵산 분자의 자유 3' 단부를 연장시켜, 제 3 다수의 핵산 분자들의 핵산 분자를 생성시키는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 결합시키는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 상기 제 3 다수의 핵산 분자들의 핵산 분자에 제한절단 효소를 접촉시켜 스티커 단부를 만드는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 상기 제 3 다수의 핵산 분자들의 핵산 분자에 추가 핵산 분자를 접촉시키는 것을 더 포함한다. 일부 구체예들에서, 상기 추가 핵산 분자는 TCR쇄의 불변 영역 또는 이의 일부분을 인코딩한다. 일부 구체예들에서, 상기 방법은 상기 제 3 다수의 핵산 분자들의 핵산 분자와 상기 추가 핵산 분자를 결합시키는 것을 더 포함한다. 일부 구체예들에서, 제 3 다수의 핵산 분자들중 적어도 5개 (가령, 일부 경우들에서, 적어도 약 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 1,000, 2,000, 3,000, 4,000, 5,000, 10,000, 20,000, 30,000, 40,000 개, 또는 그 이상의) 상이한 핵산 분자가 동일한 구획 안에서 생성된다. 일부 구체예들에서, 제 3 다수의 핵산 분자들에서 적어도 10개 상이한 핵산 분자가 동일한 구획 안에서 생성된다. 일부 구체예들에서, 상기 동일한 구획은 웰, 튜브, 또는 소적이다.

[0027]

또다른 측면에서, 본 명세서는 핵산 분자들 다수를 생성시키는 방법을 제공하는데, 이 방법은 다음을 포함한다: (a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR)쇄의 제 1 CDR3 및 제 2 TCR쇄의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR쇄의 동계 쌍으로부터 유래되고; (b) 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고 (c) 제 1 다수 핵산분자들에서 핵산 분자와 제 2 다수 핵산 분자들에서 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 선형 핵산 분자를 형성하게 되고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 TCR V 유전자는 TCR쇄의 동계 쌍으로부터 유래된다. 일부 구체예들에서, 제 1 다수의 핵산 분자는 적어도 약 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개,

또는 그 이상의 상이한 서열을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들은 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 상이한 TCR V 유전자들을 포함한다.

[0028] 또다른 측면에서, 본 명세서는 핵산 분자들 다수를 생성시키는 방법을 제공하며, 이 방법은 다수를 포함한다: (a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 (i) 제 1 T-세포 수용체 (TCR) 쉘의 제 1 CDR3과 제 2 TCR 쉘의 제 2 CDR3을 인코딩하는 합성 서열, (ii) 제 3 T-세포 수용체 (TCR) 쉘의 제 3의 CDR3과 제 4의 TCR 쉘의 제 4의 CDR3을 인코딩하는 합성 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 제 1 TCR 쉘의 동계 쌍으로부터 유래되며, 이때 제 3의 CDR3 및 제 4의 CDR3은 제 2 TCR 쉘의 동계 쌍으로부터 유래되며; (b) 제 2 다수의 핵산 분자들을 제공하고, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고 (c) 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들을 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 형성하고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 TCR 쉘의 동계 쌍으로부터 유래된다. 일부 구체예들에서, 제 1 다수의 핵산 분자는 적어도 약 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열을 포함한다. 일부 구체예들에서, 제 2 다수의 핵산 분자들은 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 상이한 TCR V 유전자들을 포함한다.

[0029] 또다른 측면에서, 본 명세서는 대상체로부터 취한 조직 샘플 안에 태생적으로 쌍을 이룬 T-세포 수용체 (TCR)의 서열을 식별해내는 방법을 제공하며, 이 방법은 다음을 포함한다: (a) 해당 대상체로부터 획득한 다수의 말초 T 세포들을 함유하는 샘플에서 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 하나 또는 그 이상의 쌍을 이룬 서열을 식별해내고, 이때 하나 또는 그 이상의 쌍을 이룬 서열 각각은 CDR3 서열을 포함하고; 그리고 (b) 태생적으로 쌍을 이룬 다른 TCR 쉘이 알려지지 않은 조직 샘플에서 TCR의 TCR 쉘의 조직 CDR3 서열을 식별해내고, 이때 이 조직 CDR3 서열은 상기 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 하나 또는 그 이상의 쌍을 이룬 서열에서 적어도 하나의 쌍을 이룬 서열의 CDR3 서열과 일치하며, 그렇게 함으로써 상기 적어도 하나의 쌍을 이룬 서열은 해당 조직 샘플에서 태생적으로 쌍을 이룬 TCR의 서열로 확인된다. 일부 구체예들에서, (a)에서 식별이란 다수의 말초 T 세포들을 함유하는 샘플에서 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs을 시퀀싱하는 것을 포함한다. 일부 구체예들에서, 상기 시퀀싱은 단일 세포 시퀀싱을 포함한다. 일부 구체예들에서, 상기 단일 세포 시퀀싱은 다수의 말초 T 세포들을 다수의 구획으로 분배하는 것을 포함하고, 각 구획은 다수의 말초 T 세포들에서 개별 말초 T 세포를 포함한다. 일부 구체예들에서, 해당 조직 샘플은 체액 샘플이 아니다. 일부 구체예들에서, 해당 조직 샘플은 고형 종양 샘플이다. 일부 구체예들에서, 해당 조직 샘플은 고정된 또는 냉동된 샘플이다. 일부 구체예들에서, 다수의 말초 T 세포들을 함유하는 샘플은 말초 혈액 단핵 세포 (PBMC) 샘플이다. 일부 구체예들에서, 상기 방법은 (a)의 실시예 앞서, 해당 대상체로부터 혈액 샘플을 획득하는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 (a)의 실시예 앞서, 해당 혈액 샘플로부터 말초 혈액 단핵 세포들을 분리시키는 것을 더 포함한다. 일부 구체예들에서, 해당 조직 샘플은 종양-침윤성 T 세포를 포함한다.

[0030] 또다른 측면에서, 본 명세서는 표적-반응성 T-세포 수용체 (TCR)를 식별해내는 방법을 제공하며, 이 방법은 다음을 포함한다: (a) 본원에 기술된 방법들을 이용하여 식별된 TCR을 포함하는 세포를 제공하고; 그리고 (b) 상기 세포를 항원-제시 세포 (APC)에 의해 제시된 표적 항원에 접촉시키고, 이때 해당 세포는 TCR을 통하여 해당 APC에 의해 제공되는 표적 항원에 결합하고, 그렇게 함으로써 상기 TCR은 표적-반응성 TCR로 확인된다. 일부 구체예들에서, 상기 표적 항원은 종양 항원 (가령, 종양-연관된 항원들 또는 종양-특이적 항원들)이다. 일부 구체예들에서, 상기 방법은 표적-반응성 TCR을 인코딩하는 서열을 숙주 세포로 운반하는 것을 더 포함한다. 일부 구체예들에서, 상기 방법은 해당 숙주 세포를 해당 대상체에게 투여하는 것을 더 포함한다. 일부 구체예들에서, 상기 숙주 세포는 T 세포이다. 일부 구체예들에서, 상기 T 세포는 자가유래(autologous) T 세포이다. 일부 구체예들에서, 상기 T 세포는 동종이계(allogeneic) T 세포이다. 일부 구체예들에서, 상기 세포는 리포터 세포 계통이며, 이때 리포터 세포 계통은 APC에 의해 제시되는 표적 항원에 해당 세포가 결합할 때 발현되는 리

포터 유전자를 포함한다.

[0031] 본 발명의 또 다른 양상들 및 이점들은 아래 상세한 설명으로부터 해당 분야의 숙련된 기술자들에게 용이하게 자명해질 것이며, 상세한 설명에는 본 발명의 예시적인 구체예들만이 제시되고 설명되어 있다. 알 수 있는 바와 같이, 본 발명은 그 외 상이한 구체예들 또한 가능하며, 이의 몇 가지 세부사항은 다양한 명백한 관점들에서 변형될 수 있으며, 이들 모두 본 발명에서 벗어나지 않는다. 따라서, 도면 및 상세한 설명은 본질적으로 제한적인 것이 아니라 예시적인 것으로 간주된다.

[0032] **참고문헌으로 포함**

[0033] 본 명세서에 언급된 모든 간행물 및 특허 출원은 각각의 개별 간행물 또는 특허 출원이 구체적으로 및 개별적으로 참조로 포함되도록 지시된 것과 동일한 정도로 본 명세서에 참조 문헌으로 인용된다. 참고문헌으로 포함된 간행물 그리고 특허 또는 특허 출원이 본 명세서에 포함된 내용과 모순되는 경우, 본 명세서의 내용이 우선하거나 및/또는 이러한 임의의 모순되는 자료에 선행하는 것으로 한다.

도면의 간단한 설명

[0034] **도면의 간단한 설명**

본 발명의 새로운 특징들은 첨부된 청구범위에 구체적으로 제시된다. 본 발명의 특징 및 이점들은 본 발명의 원리를 이용하는 예시적 구체예들을 제시하는 하기 상세한 설명 그리고 다음과 같은 첨부도면 (본원에서 또한 "도면" 및 "도")을 참고하여 보다 잘 이해될 것이다:

도 1A-1C는 T-세포 수용체를 인코딩하는 핵산 구조체를 만드는 예시적인 계획을 도시한다 .

도 2A는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 2B는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 3A는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 3B는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 4A는 TCR V 유전자의 생식계열 게놈 DNA를 도해한 것이다.

도 4B는 TCR V-J 유전자의 재배열된 게놈 DNA를 도해한 것이다.

도 4C는 TCR V-D-J 유전자의 재배열된 게놈 DNA를 도해한 것이다.

도 5는 CDR3-J 폴리뉴클레오타이드를 정확한 V 유전자 생식계열 폴리뉴클레오타이드에 연계시키는 것과 연관된 잠재적인 과제를 도시한 것이다. 점선 화살표는 CDR3-J 폴리뉴클레오타이드와 맞지 않는 V 유전자 생식계열 폴리뉴클레오타이드 사이에 일어날 수 있음을 나타낸다.

도 6은 CDR3-J 폴리뉴클레오타이드 (백색 상자에 연결된 회색 상자)가 명시된, 사전-합성된다. V 유전자 생식계열 폴리뉴클레오타이드 (가는 화살표로 나타난 회색 상자에 연결된 흑색 상자)에 프라이머 연장부의 중첩에 의해 연계된 도해를 나타낸다. 상부 두꺼운 화살표 (603)는 사전-합성된 V 유전자 생식계열 폴리뉴클레오타이드 (601) 상의 커넥터 서열과 CDR3-J 폴리뉴클레오타이드 (602) 상의 커넥터 서열 간의 혼성화를 나타낸다. 바닥의 두꺼운 화살표 (604)는 프라이머 연장을 나타낸다. 601은 커넥터 서열로 지칭될 수 있고, 602는 안티-커넥터 서열로 지칭될 수 있다 (또는 이 역도 성립한다).

도 7은 임의적인(arbitrary) 커넥터 (701) 및 안티-커넥터 (702) 서열을 이용한, CDR3-J 폴리뉴클레오타이드와 명시된 V 유전자 생식계열 폴리뉴클레오타이드의 연계를 나타낸다.

도 8은 TCR 유전자 자가-어셈블리의 전반적인 원리를 나타낸다. 801: 사전-합성된다 V 유전자 생식계열 폴리뉴클레오타이드. 802: CDR3-J 서열 (가령, CDR3-J 폴리뉴클레오타이드)을 포함하는 폴리뉴클레오타이드. 803: V 유전자 생식계열 폴리뉴클레오타이드 서열 및 CDR3-J 서열을 포함하는 핵산 서열. X는 폴리뉴클레오타이드의 수를 나타내며, 각 폴리뉴클레오타이드는 상이한 V 유전자 생식계열 폴리뉴클레오타이드의 일부분이다. Y는 CDR3-J 폴리뉴클레오타이드의 숫자다. Y는 X보다 더 클 수 있다. 화살표는 벌크 반응을 나타내며, 이때 각 CDR3-J 폴리뉴클레오타이드는 명시된, 사전-합성된다 V 유전자 생식계열 폴리뉴클레오타이드에 연계된다.

도 9A는 중앙 샘플 안에 중앙-침윤성 TCRs를 식별해내기 위해 혈액 샘플을 이용한 예시적인 작업흐름을 도시한

다.

도 9B는 도 9A에서 나타난 방법을 이용하여 식별된 TCRs의 예시적인 적용을 나타낸다.

도 9C는 도 9A에서 나타난 방법을 이용하여 식별된 TCRs의 예시적인 적용을 나타낸다.

도 10A는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 10B는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 11A는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 11B는 본원에 기술된 방법들을 이용하여 예시적인 시뮬레이션 결과를 도시한다.

도 12는 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 13은 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 14는 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 15는 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 16은 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 17은 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

도 18은 본원에 기술된 유전자 어셈블리 방법을 평가하는 예시적인 차세대 시퀀싱 데이터를 도시한다.

발명을 실시하기 위한 구체적인 내용

[0035] 발명의 상세한 설명

[0036] 본 명세서에서, 단수의 용도는 다른 명시적 언급이 없는 한, 복수를 포함한다. 또한, "또는"이란 명시적으로 다른 언급이 없는 한, "및/또는"을 말한다. 유사하게, "포함하다(단수형, 복수형: comprise, comprises)", "포함하는", "내포하다(단수형, 복수형: include, includes), 및 "내포하는"이란 제한적인 의미를 갖는 의도는 아니다.

[0037] 용어 "약(about)" 또는 "대략(approximately)"은 당업자에 의해 결정된 특정 값에 대한 허용가능한 오차를 의미하며, 이는 어떻게 이러한 값이 측정되는 또는 결정되는 방법에 따라 부분적으로 의존적일 것이며, 예를 들면, 측정 시스템의 제약이 있을 것이다. 예를 들면, "약(about)"이란 당업계의 관행에 따라 1 또는 1 이상의 표준 편차 이내를 의미할 수 있다. 대안으로, "약"이란 주어진 값의 최대 20%, 최대 10%, 최대 5%, 또는 최대 1%의 범위를 의미할 수 있다. 대안적으로, 특히 생물학적 시스템 또는 과정과 관련하여, 이 용어는 값의 배수 이내, 예를 들어 5배 이내, 또는 2배 이내를 의미할 수 있다. 특정 값이 출원 및 청구범위에 기술된 경우, 달리 언급되지 않는 한, "약"이라는 용어는 특정 값에 대해 허용 가능한 오차 범위 내를 의미하는 것으로 추정해야 한다.

[0038] 2개 이상의 일련의 숫자 값에서 첫 번째 숫자 값 앞에 "적어도", "~보다 큰" 또는 "~보다 크거나 또는 같은"이라는 용어가 붙을 경우, 이 용어 "적어도", "~보다 큰" 또는 "~보다 크거나 또는 같은"이란 일련의 숫자 값에 있는 각 숫자 값에 모두 적용된다. 예를 들면, 1, 2 또는 3보다 크거나 또는 같다는 표현은 1보다 크거나 또는 같고, 2보다 크거나 또는 같고, 또는 3보다 크거나 같다는 의미와 대등하다.

[0039] 2개 이상의 일련의 숫자 값에서 첫 번째 숫자 값 앞에 "단지(no more than)", "~보다 적은" 또는 "~보다 적거나 또는 같은"이라는 용어가 붙을 경우, 이 용어 "단지", "~보다 적은" 또는 "~보다 적거나 또는 같은"이란 일련의 숫자 값에 있는 각 숫자 값에 모두 적용된다. 예를 들면, 3, 2 또는 1보다 적거나 또는 같다는 표현은 3보다 적거나 또는 같고, 2보다 적거나 또는 같고, 또는 1보다 적거나 같다는 의미와 대등하다.

[0040] 용어 "폴리뉴클레오타이드", "핵산" 및 "올리고뉴클레오타이드"는 본 명세서에서 호환사용된다. 이들은 다양한 길이의 뉴클레오타이드의 고분자 형태를 나타낼 수 있다. 이들은 데옥시리보뉴클레오타이드 및/또는 리보뉴클레오타이드, 또는 이의 유사체들을 포함할 수 있다. 폴리뉴클레오타이드에는 아데노신 (A), 시토신 (C), 구아닌 (G), 티민 (T) 및 우라실 (U) 또는 이의 변이체에서 선택된 하나 또는 그 이상의 뉴클레오타이드가 내포될 수 있다. 뉴클레오타이드에는 뉴클레오시드와 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 그 이상의 포스페이트 (PO₃)기들이 내포될 수 있다. 뉴클레오타이드에는 핵염기, 오탄당 (리보스 또는 데옥시리보스), 및 하나 또는 그 이상의 포스페이트

트 기들이 내포될 수 있다. 폴리뉴클레오티드는 임의의 3-차원 구조를 가질 수 있으며, 다양한 기능을 수행할 수 있다. 폴리뉴클레오티드는 선형, 원형, 스템-루프(stem-loop) 및 분기-형과 같은 다양한 배치를 가질 수 있다. 다음은 폴리뉴클레오티드의 비-제한적인 예시들이다: 유전자 또는 유전자 단편의 코딩 영역 또는 넌-코딩 영역, 링키지 분석으로 특정된 좌(좌들), 엑손,인트론, 메신저 RNA (mRNA), 전달 RNA (tRNA), 리보솜 RNA (rRNA), 짧은 간섭 RNA (siRNA), 짧은-헤어핀 RNA (shRNA), 마이크로-RNA (miRNA), 원형 RNA, 리보자임, cDNA, 재조합 폴리뉴클레오티드, 분기된(branched) 폴리뉴클레오티드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 프로브, 그리고 프라이머. 폴리뉴클레오티드는 하나 또는 그 이상의 변형된 뉴클레오티드, 이를 테면, 메틸화된 뉴클레오티드와 뉴클레오티드 유사체들을 포함할 수 있다. 존재하는 경우, 뉴클레오티드 구조에 대한 변형은 중합체의 조립 전 또는 후에 부여될 수 있다. 뉴클레오티드 서열은 비-뉴클레오티드 구성요소들에 의해 중단될 수 있다. 올리고뉴클레오티드는 중합화 후 추가 변형될 수 있는데, 이를 테면, 라벨링(labeling) 구성요소의 콘주게이션에 의해 변형될 수 있다. 폴리뉴클레오티드에는 비-표준 뉴클레오티드(들), 비-천연 뉴클레오티드(들), 뉴클레오티드 유사체(들) 및/또는 변형된 뉴클레오티드를 비롯한, 하나 또는 그 이상의 뉴클레오티드 변이체들이 내포될 수 있다.

[0041] 본원에 사용된 용어 "서열"은 핵산 분자에서 뉴클레오티드의 순서, 또는 펩티드의 아미노산 잔기의 순서를 지칭한다. 핵산 서열은 데옥시리보핵산 (DNA) 서열 또는 리보핵산 (RNA) 서열일 수 있고; 선형, 원형 또는 분지형일 수 있고, 단일-가닥 또는 이중-가닥일 수 있다. 서열이 돌연변이되어, 참조 서열 (가령, 야생형 서열)과 상이할 수 있다. 서열은 임의의 길이, 예를 들면, 2 ~ 1,000,000개 또는 그 이상 길이의 아미노산 또는 뉴클레오티드 (또는 이들 사이의 임의의 정수 값, 또는 이들 이상의 임의의 정수 값), 가령, 약 100 ~ 약 10,000개 뉴클레오티드 또는 약 200 ~ 약 500개 아미노산 또는 뉴클레오티드일 수 있다. 일부 경우들에서, 주어진 핵산 서열은 주어진 핵산 서열의 서열 정보 및 주어진 핵산 서열의 역-보체 서열을 포괄할 수 있다. 일부 경우들에서, DNA 서열은 이 DNA로부터 전사되는 상응하는 RNA 서열의 서열 정보를 포함할 수 있다. 상기 서열은 폴리뉴클레오티드 또는 폴리펩티드 분자의 알파벳 표시일 수 있다. 상기 서열은 컴퓨터 프로세서에서 사용할 수 있는 정보 단편일 수 있다. 일부 경우들에서, 핵산 서열은 물리적 핵산 분자 자체를 나타내는 데 사용될 수 있다.

[0042] 본원에서 사용된 바와 같이, 용어 "블런트(blunt) 단부"는 이중-가닥으로 된 핵산 분자의 단부를 지칭하는데, 이때 해당 핵산 분자의 한 가닥의 단부에 있는 실질적으로 모든 뉴클레오티드가 상기 동일한 핵산 분자의 다른 가닥에 있는 마주하는 뉴클레오티드와 염기쌍을 이룬다. 만일 해당 핵산 분자가 소위 "오버행" 또는 "스티커 단부"로 본원에서 불리는 길이가 적어도 하나의 뉴클레오티드를 갖는 단일-가닥 부분을 내포한다면, 이 핵산 분자는 블런트 단부가 아니다.

[0043] 본원에서 사용된 바와 같이, "TCR V 유전자"는 생식계열 입체배열에서 T-세포 수용체 가변 (V) 유전자의 게놈 핵산 서열을 지칭하는데, 이는 리더 펩티드의 제 1부 (가령, IMGT으로 정의될 때, L-PART1), 인트론 (가령, IMGT으로 정의될 때, V-INTRON) 그리고 엑손 (가령, IMGT으로 정의될 때, V-EXON)을 인코딩하는 서열과 함께, 5'UTR 및 3'UTR (재조합 신호 서열 내포)을 포함한다. 상기 재조합 신호 서열은 헵타머 (가령, IMGT으로 정의될 때, V-HEPTAMER)와 나노머 (가령, IMGT으로 정의될 때, V-NONAMER)를 포함하며, 이들은 스페이서 요소 (가령, IMGT으로 정의될 때, V-SPACER)에 의해 구별되어 있을 수 있다. V-EXON은 리더 펩티드의 제 2 부 (L-PART2)을 인코딩하는 서열 및 V-REGION을 포괄된다. TCR V 유전자의 예로는 TCR 알파 가변 (TRAV) 유전자, TCR 베타 가변 (TRBV) 유전자, TCR 감마 가변 (TRGV) 유전자, 및 TCR 델타 가변 (TRDV) 유전자가 내포된다. 본원에 기술된 핵산은 상기 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. "~로부터 유래된"이란 참조 서열과 적어도 약 40%, 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 90%, 적어도 약 95%, 또는 100% 서열 동일성을 갖는 서열을 의미한다. TCR V 유전자로부터 유래된 서열은 상기에서 기술된 바와 같이, TCR V 유전자의 게놈 핵산 서열의 전체 길이 서열일 수 있다. TCR V 유전자로부터 유래된 서열은 해당 TCR V 유전자의 적어도 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500개, 또는 그 이상의 뉴클레오티드를 포함하는, TCR V 유전자의 일부분일 수 있다. TCR V 유전자로부터 유래된 서열은 코돈-최적화된 (또는 코돈-다각화되어) 핵산 서열일 수 있다. 주어진 핵산 서열의 코돈-최적화된 서열은 해당 서열의 단백질-코딩 영역이 주어진 핵산의 단백질-코딩 영역과 동일한 아미노산 서열을 인코딩하는 변형된 핵산 서열을 지칭한다. 상기 변형된 핵산 서열은 주어진 핵산 서열과 상이한 서열일 수도 있고, 또는 주어진 핵산으로부터 유래될 수도 있다. 제한(restriction) 부위를 제거하기 위해, 폴리뉴클레오티드 서열에서 원치 않는 이차 구조를 제거하기 위해, CDR3-J 폴리뉴클레오티드와 TCR V 유전자의 명시된 사전-합성된 부분의 정확한 연계를 촉진시키기 위해, 또는 다른 목적으로 코돈-최적화가 이행될 수 있다. 코돈-최적화 또는 코돈 다양화(codon diversification)는 주어진 핵산 서열에서 하나 또는 그 이상의 뉴클레오티드를 변경시킴으로써 이루어질 수 있다. 예를 들면, 코돈-최적화 또는 코돈 다양화는 계산 방법으로 이루어질 수 있다. 본 명세서에서 코돈-최적

화와 코돈 다양화는 호환이될 수 있다.

- [0044] 본원에서 사용된 바와 같이, 용어 "V-REGION"는 생식계열 게놈 DNA 또는 cDNA에서 TCR V 유전자의 코딩 영역 (V-HEPTAMER 앞에 뉴클레오티드가 (존재하는 경우), 1 또는 2개 뉴클레오티드가 내포됨), 또는 재배열된 게놈 DNA 또는 cDNA에서 V-(D)-J 재배열에 의해 보통 3' 말단에 정리되어 있는 가변 (V) 영역을 지칭한다.
- [0045] 용어 "D-REGION"이란 본원에서 사용된 바와 같이, 생식계열 게놈 DNA 또는 cDNA에서 TCR D 유전자의 코딩 영역 (5' D-HEPTAMER 다음에 및/또는 3' D-HEPTAMER의 앞에 뉴클레오티드가 (존재하는 경우), 1 또는 2개 뉴클레오티드가 내포됨), 또는 부분적으로-재배열된 또는 재배열된 게놈 DNA 또는 cDNA에서 D-J 또는 V-D-J 재배열에 의해 보통 5' 및/또는 3'에 정리되어 있는 다양성 (D) 영역을 지칭한다.
- [0046] 용어 "J-REGION"이란 본원에서 사용된 바와 같이, 생식계열 게놈 DNA 또는 cDNA에서 TCR J 유전자의 코딩 영역 (J-HEPTAMER 다음에 뉴클레오티드가 (존재하는 경우), 1 또는 2개 뉴클레오티드가 내포됨), 또는 재배열된 게놈 DNA 또는 cDNA에서 V-(D)-J 재배열에 의해 보통 5'에 정리되어 있는 이음(J) 영역을 지칭한다.
- [0047] 용어 "V-J-REGION"이란 본원에서 사용된 바와 같이, 재배열된 게놈 DNA 또는 cDNA에서 V-REGION 및 J-REGION을 포함하는 TCR 쉼의 코딩 영역을 지칭한다.
- [0048] 용어 "V-D-J-REGION"이란 본원에서 사용된 바와 같이, 재배열된 게놈 DNA 또는 cDNA에서 V-REGION, D-REGION, 및 J-REGION을 포함하는 TCR 쉼의 코딩 영역을 지칭한다.
- [0049] 용어 "연계하다(link)" 또는 "연결하다(connect)"는 본 명세서에서 호환사용된다. 이들은 두 개 또는 그 이상의 핵산 분자를 물리적으로 연계시키는 것을 말한다. 두 개 또는 그 이상의 핵산 분자는 이들 두 개 또는 그 이상의 핵산 분자가 연속된 핵산 분자를 이루도록 연계될 수 있다. 상기 두 개 또는 그 이상의 핵산 분자는 공유적으로 연계될 수 있거나, 또는 비-공유적으로 연계될 수 있다. 연계는 수소 결합, 이온 및 공유 결합, 또는 반 데르 발스(van der Waals) 힘의 형성을 비롯한, 다양한 방식으로 달성될 수 있다.
- [0050] 참조 핵산 서열 (또는 펩티드 서열)에 대한 서열 동일성 퍼센트(%)는 해당 서열들을 정렬시키고, 그리고 최대 서열 동일성 퍼센트를 얻기 위해 필요하다면 갭을 도입시켜, 참조 핵산 서열 (또는 펩티드 서열)에 있는 뉴클레오티드 (또는 펩티드 서열인 경우 아미노산 잔기)와 후보 서열에 있는 동일한 뉴클레오티드 (또는 아미노산 잔기)의 백분율이며, 임의의 보존 치환은 서열 동일성 부분으로 간주되지 않는다. 동일성 백분율을 결정하기 위한 정렬은 예를 들어 BLAST, BLAST-2, CLUSTALW, ALIGN 또는 Megalign (DNASTAR) 소프트웨어와 같은 공개적으로 이용가능한 컴퓨터 소프트웨어를 사용하여 당업계의 기술 범위 내에 있는 다양한 방법으로 성취될 수 있다. 당업자는 서열 정렬을 위하여 적절한 매개변수를 결정할 수 있는데, 비교되는 서열의 전장에 걸쳐 최대 정렬을 얻기 위하여 필요한 임의의 알고리즘이 포함된다.
- [0051] 핵산 서열 또는 아미노산 서열에 적용되는 경우 용어 "실질적으로 동일한" 및 이의 문법적 등가표현은 핵산 또는 아미노산 서열은 상기에서 기술된 프로그램, 가령, BLAST에서 표준 매개변수를 이용하여, 참조 서열과 비교하였을 때, 적어도 90% 서열 정체 또는 그 이상의, 적어도 95%, 적어도 98% 또는 적어도 99% 동일성을 갖는 서열을 포함한다는 의미다. 예를 들면, BLASTN 프로그램 (뉴클레오티드 서열의 경우)은 디폴트로 단어 길이 (W) 11, 기대치 (E) 또는 10, M = 5, N = -4가 이용되며, 그리고 두 가닥의 비교이다. 아미노산 서열의 경우, BLASTP 프로그램은 디폴트로 단어 길이 (W) 3, 기대치 (E) 10, 및 BLOSUM62 스코어링 매트릭스를 이용한다 (Henikoff & Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 (1992)참고).
- [0052] **개요**
- [0053] 서열 T-세포 수용체 (TCR)에 고-처리량의 쌍을 이룬 시퀀싱이 이용될 수 있다. 예를 들면, 단일-세포 기술의 발달로, 개별 T 세포들은 절연 구획으로 분할될 수 있고, 여기에서 동일한 T 세포로부터 TCR 알파 쉼과 베타 쉼 mRNAs는 동일한 특유 바코드에 부착될 수 있다. 이러한 시스템 중 일부는 상업적으로 이용 가능하다 (가령, 10X Genomics). T-세포 수용체 알파 가변 (TRAV) 유전자 정체(identity), CDR3 알파 서열, T-세포 수용체 알파 이음 (TRAJ) 유전자 정체, T-세포 수용체 베타 가변 (TRBV) 유전자 정체, CDR3 베타 서열, 및 T-세포 수용체 베타 이음 (TRBJ) 유전자 정체를 기록하는 쌍을 이룬 서열 정보로 전장의, 발현가능한 TCR의 구축이 가능할 수 있다. 그러나, 기능 연구 또는 스크리닝을 위해, 세포에 도입될 수 있는 DNA 또는 RNA 형태로 이러한 TCR 서열을 합성하는 기술은 저-처리량 기술일 있다. 본 명세서는 TCR 서열 (가령, 일부 경우들에서, 쌍을 이룬, 전장의, 발현가능한 TCR 서열)을 인코딩하는 폴리뉴클레오티드의 초고-처리량 구축을 허용하는 다중 방법 및 조성물을 제공한다.

[0054] **T-세포 수용체 (TCR)**

[0055] TCR은 다양한 암 또는 감염성 유기체와 연합된 항원을 인지하는 T 세포의 능력을 부여하는데 이용될 수 있다. 상기 TCR은 두 개의쇄, 가령, 알파 (α)쇄와 베타 (β)쇄, 또는 감마 (γ)쇄와 델타 (δ)쇄로 구성된다. 이러한쇄들을 구성하는 단백질은 TCR의 엄청난 다양성을 생성하는 독특한 메커니즘을 사용하는 DNA에 의해 인코딩된다. 이러한 다중-소단위 면역 인지 수용체는 CD3 복합체와 연합되고, 항원-제시 세포들 (APCs)의 표면에 MHC 클래스 I 단백질과 II 단백질에 의해 제시되는 펩티드에 결합한다. APC 상에서 항원성 펩티드에 대한 TCR의 결합은 T 세포와 APC 사이의 접촉 지점에서 면역학적 시냅스에서 발생하는 T 세포 활성화의 중심 사건일 수 있다.

[0056] 상기 TCR은 MHC 클래스 I 분자의 맥락에서 T 세포 에피토프를 인지할 수 있다. MHC 클래스 I 단백질들은 고등 척추동물의 모든 유핵(nucleated) 세포들에서 발현될 수 있다. MHC 클래스 I 분자는 12-kDa 경쇄 β -2 마이크로 글로불린과 비-공유적으로 연합된 46-kDa 중쇄로 구성된 이중이량체다. 인간의 경우, 몇 가지 MHC 대립유전자들이 있는데, 이를테면, 예를 들면, HLA-A2, HLA-A1, HLA-A3, HLA-A24, HLA-A28, HLA-A31, HLA-A33, HLA-A34, HLA-B7, HLA-B45 및 HLA-Cw8들이다. 일부 구체예들에서, MHC 클래스 I 대립유전자는 HLA-A2 대립유전자이며, 일부 집단의 경우 해당 집단의 대략적으로 50%에서 발현된다. 일부 구체예들에서, 상기 HLA-A2 대립유전자는 HLA-A*0201, *0202, *0203, *0206, 또는 *0207 유전자 산물일 수 있다. 일부 경우들에서, 상이한 집단 간에 하위유형 (subtypes)의 빈도에서 차이가 있을 수 있다. 예를 들면, 일부 구체예들에서, HLA-A2 양성 백인 (Caucasian) 집단의 95% 이상은 HLA-A*0201인 반면, 중국인 (Chinese) 집단에서 빈도는 대략적으로 다음과 같이 보고되었다: 대략적으로 23% HLA-A*0201, 45% HLA-A*0207, 8% HLA-A*0206 및 23% HLA-A*0203.

[0057] 일부 구체예들에서, 상기 TCR은 MHC 클래스 II 분자의 맥락에서 T 세포 에피토프를 인지할 수 있다. MHC 클래스 II 단백질들은 APCs 하위집단에서 발현될 수 있다. 인간의 경우, 몇 가지 MHC 클래스 II 대립유전자가 있는데, 이를테면, 예를 들면, DR1, DR3, DR4, DR7, DR52, DQ1, DQ2, DQ4, DQ8 및 DPI들이다. 일부 구체예들에서, 상기 MHC 클래스 II 대립유전자는 HLA-DRB1*0101, HLA-DRB*0301, HLA-DRB*0701, HLA-DRB*0401 또는 HLA-DQB1*0201 유전자 산물이다.

[0058] 상기 TCR쇄는 가변 도메인 (또는 가변 영역) 및 불변 도메인 (또는 불변 영역)을 포함할 수 있다. 상기 가변 도메인은 IMGT 특유 번호매김 시스템으로 정의될 때, V-도메인일 수 있다. 상기 가변 도메인은 TCR쇄의 V-J-REGION 또는 V-D-J-REGION에 상응할 수 있다. 상기 불변 도메인은 IMGT 특유 번호매김 시스템으로 정의될 때, C-도메인일 수 있다. 일부 경우들에서, 상기 불변 도메인은 상기 불변 영역의 부분일 수 있다. 예를 들면, 전장의 불변 영역은 불변 도메인 (세포외 영역), 연결 영역, 막경유 영역, 및 세포질 영역을 포함할 수 있다.

[0059] TCR α 쇄 또는 TCR δ 쇄의 가변 도메인은 생식계열에서 다수의 가변 (V) 및 이음 (J) 유전자 세그먼트에 의해 인코딩될 수 있고, TCR β 쇄 또는 TCR γ 쇄의 가변 도메인은 다양성 (D) 유전자 세그먼트에 의해 추가로 인코딩된다. 각 유전자 세그먼트는 재조합 신호 서열의 측면에 있을 수 있다. 재조합 신호는 스페이스 요소의해 구별되어 있는 헵타머 및 나노머를 포함할 수 있다. 상기 스페이스 요소의 길이는 12 또는 23 bp일 수 있다. V(D)J 재조합하는 동안, 각 유전자 세그먼트의 하나의 무작위 대립유전자는 다른 것들과 재조합되어, 기능성 가변 도메인을 형성한다. 가변 도메인과 불변 (C) 유전자 세그먼트의 재조합으로 기능성 TCR쇄 전사체가 만들어질 수 있다. 추가적으로, 무작위 뉴클레오티드는 및/또는 유전자 세그먼트 사이의 접합(junction) 부위 사이의 정선 부위에서 추가되거나, 및/또는 결손될 수 있다. 이 과정은 강력한 조합 (재결합할 유전자 영역에 따라 다름) 및 접합 다양성(추가/삭제된 뉴클레오티드 및 이의 수에 따라)으로 이어질 수 있고, 결과적으로 과량의, 가변성이 매우 큰 TCR 레퍼토리가 생성되고, 이는 과다 항원의 식별을 담보할 수 있다. 기능적 TCR을 형성하기 위해, α 쇄와 β 쇄, 또는 γ 쇄와 δ 사슬의 쌍형성("어셈블리"이라고도 함)에 의해 추가적인 다양성이 달성될 수 있다. 재조합, 무작위 삽입, 결손 및 치환에 의해, T 세포 수용체를 인코딩하는 유전자 소집단은 $10^{15} \sim 10^{20}$ 개의 TCR 클론유형을 만들어낼 잠재력을 갖는다. 본원에서 사용된 바와 같이, "클론유형(clonotype)"이란 동일한 면역수용체를 나르는 면역 세포 집단을 지칭한다. 예를 들면, 클론유형은 동일한 TCR을 나르는 T 세포 집단, 또는 동일한 BCR (또는 항체)을 나르는 B-세포 집단을 지칭한다. 면역수용체 다양성 맥락에서 "다양성"이란 집단에서 면역수용체 (가령, TCR, BCR 및 항체) 클론유형의 수를 지칭한다. 본원에서 사용된 바와 같이, "동계 쌍 조합"이란 T 세포로부터 두 개의쇄의 TCR 고유 조합 (가령, TCR α 와 TCR β , 또는 TCR γ 와 TCR δ)을 지칭한다. 동일한 TCR에서 두 개의 동일한 동계 쌍 조합이 일어날 수 있다. 예를 들면, 동일한 클론유형을 갖는 T 세포들은 TCR α 쇄와 TCR β 쇄의 동일한 동계 쌍 조합을 갖는다. 클론유형에서 다양성이 높을 수록 동계 쌍 조합에서 다양성이 높음을 나타낼 수 있다.

[0060] 각 TCR쇄는 구조에 상보성 결정 영역(CDR1-3)이라고 명명된, 3개의 추가변 루프를 함유할 수 있다. V 유전자들에 의해 CDR1 및 CDR2가 인코딩될 수 있고, TCR과 MHC 복합체의 상호작용에 요구될 수 있다. 그러나, CDR3은 (1) V 유전자와 J 유전자 사이의 접합 영역 (TCR α 또는 TCR γ 의 경우), 또는 (2) V 유전자와 D 유전자 사이의 접합 영역 및 D 유전자와 J 유전자 사이의 접합 영역 (TCR β 또는 TCR δ 의 경우)에 의해 부분적으로 인코딩되며, 따라서 매우 가변적일 수 있다. CDR3은 상기 TCR에서 펩티드 항원과 바로 접촉하는 영역일 수 있다. CDR3은 T 세포 클론유형을 결정하기 위한 관심대상 영역으로 이용될 수 있다. 한 개체의 T 세포에 의한 모든 TCRs의 합은 TCR 레퍼토리 TCR 프로파일이라고 한다. 상기 TCR 레퍼토리는 질환의 개시 및 진행에 따라 변화될 수 있다. 따라서, 상이한 질환 상태, 이를 테면 암, 자가면역, 염증 및 감염성 질환 하에서 면역 레퍼토리 상태의 결정이 질환 진단 및 예후에 유용할 수 있다.

[0061] TCR는 전장의 TCRs, 뿐만 아니라 항원-결합 부분 또는 이의 항원-결합 단편 (또는 MHC-펩티드 결합 단편으로 지칭됨)을 포괄하는 것으로 이해되어야 한다. 일부 구체예들에서, 상기 TCR은 무손상(intact) 또는 전장의 TCR이다. 일부 구체예들에서, 상기 TCR은 전장 TCR보다 작지만 MHC-펩티드 복합체에 결합하는 것과 같이, 가령, MHC 분자에 결합된 특정 항원성 펩티드에 결합하는 항원-결합 부분이다. 일부 경우에, TCR의 항원-결합 부분 또는 단편은 전장 또는 무손상 TCR의 구조적 도메인의 일부만을 함유할 수 있지만, 전체 TCR이 결합하는 에피토프 (가령, MHC-펩티드 복합체)에 결합할 수 있다. 일부 경우들에서, TCR의 항원-결합 부분 또는 단편은 TCR의 가변 도메인, 이를 테면, 특이적 MHC-펩티드 복합체에 결합을 하기 위한 결합 부위를 형성하는데 충분한 TCR의 가변 α 쇄 및 가변 β 쇄를 함유하며, 이때 이를 테면, 일반적으로 각쇄는 세 개의 상보성 결정 영역을 함유한다. 항원-결합 도메인인 결합 도메인을 갖거나, 또는 항원-결합 도메인에 상동성인 결합 도메인을 갖는 폴리펩티드 또는 단백질들이 내포된다.

[0062] TCR 분자는 알파쇄 (α 쇄 또는 TCR α 쇄, TRA 유전자/서열에 의해 인코딩됨)와 베타쇄 (β 쇄 또는 TCR β 쇄, TRB 유전자/서열에 의해 인코딩됨), 또는 감마쇄 (γ 쇄 또는 TCR γ 쇄, TRG 유전자/서열에 의해 인코딩됨)와 델타쇄 (δ 쇄 또는 TCR δ 쇄, TRD 유전자/서열에 의해 인코딩됨)에 의해 형성될 수 있다. 이들 면역수용체쇄는 가변 도메인 (가령, 재배열된 VDJ 영역 또는 VJ 영역에 의해 인코딩됨)을 가질 수 있다. 상기 가변 도메인의 일부분은 추가변 도메인일 수 있다. 상기 추가변 영역에는 상보성 결정 영역 (CDRs), 예를 들면, CDR1, CDR2 및 CDR3이 내포될 수 있다. 일부 경우들에서, 하나의 T 세포 안에서, 오직 하나의 기능성 α 쇄 서열과 하나의 기능성 β 쇄 서열만이 발현될 수 있다. 일부 경우들에서, 하나의 T 세포 안에서, 오직 하나의 기능성 γ 쇄 서열과 하나의 기능성 δ 쇄 서열만이 발현될 수 있다.

[0063] 칩-기반 올리고뉴클레오티드 합성: 기회와 도전

[0064] 칩-기반 고-처리량 올리고뉴클레오티드 합성 기술은 임의의 서열을 가진 수십만 또는 수백만 개의 올리고뉴클레오티드를 한 번에 합성할 수 있을 정도로 발전했을 수 있지만, 이러한 방식에서 합성된 올리고뉴클레오티드의 길이는 약 200 ~ 300개 염기로 제한적일 수 있다. 대조적으로, 전장의 TCR 구조체의 길이는 거의 2kb가 될 수 있다. 언뜻 보면, 칩-기반 합성은 TCR 유전자 합성 문제를 해결하기에는 불충분해 보일 수 있다. 그러나, TCR의 구조를 조사하면 기회가 드러날 수 있다. 첫째, TCR 알파쇄와 베타쇄의 불변 영역(가령, TRAC 및 TRBC)은 일정할 수 있다. 따라서, TCR쇄의 불변 영역을 인코딩하는 폴리뉴클레오티드 서열은 상기 TCR 서열의 나머지에 붙어있을 수 있다. 둘째, BCR/항체 서열과 달리, TCR은 체세포 과돌연변이를 겪지 않을 수 있으며, 이것은 CDR3 영역 외부의 서열이 생식계열 기원일 수 있음을 의미한다. 따라서, 폴리뉴클레오티드들 (각각 TCR V 유전자 또는 이의 일부분으로부터 유래된 서열을 포함하는)은 사전-합성될 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 이 TCR V 유전자의 일부분일 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 코돈-최적화된 서열이거나, 또는 하나 또는 그 이상의 변형된 뉴클레오티드를 포함할 수 있다. 예를 들면, L-PART1 (리더 펩티드의 제 1부), L-PART2 (리더 펩티드의 제 2부), FR1, CDR1, FR2, CDR2 및 FR3 (L-V-REGION으로 지칭됨)에 대한 코딩 서열을 포함하는 TCR V 유전자로부터 유래된 서열이 사전-합성될 수 있다. 또다른 예로써, FR1, CDR1, FR2, CDR2 및 FR3 (V-REGION으로 지칭됨)에 대한 코딩 서열을 포함하는 TCR V 유전자로부터 유래된 서열이 사전-합성될 수 있다. L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 또는 FR3의 핵산 서열 세그먼트는 IMGT 특유 번호매김 시스템 (<http://www.imgt.org>)에 따라 정의될 수 있다. 일부 경우들에서, 상기 TCR V 유전자로부터 유래된 서열은 L-PART1을 인코딩하는 서열에서 시작하고, 제 2 보존된 시스테인 (가령, IMGT로 정의될 때, 제 2-CYS은 V-도메인의 위치 104에서 보존된 시스테인에 대한 코드에 상응함)을 인코딩하는 코돈에서 종료되는 서열을 포함할 수 있다. 인간 게놈에는 약 80개 또는 그 이상의 TCR V 유전자들 (가령, TRAV 및 TRBV 유전자들)이 있기 때문에, 이러한 "V 유전자 생식계열 폴리뉴클레오티드 라이브러리" (도 8에서 801로 나타냄, 괄호 X)의 합성이 실현가능할 수 있다. 일부 경우들에서, 종 (가령, 인간 또는 마우스)에서 TCR V 유전자들의 하위집단을

합성하여 "V 유전자 생식계열 폴리뉴클레오티드 라이브러리"가 생성된다. TCR V 유전자들의 모든 식별된 또는 하위집단이 합성될 수 있다. 예를 들면, 해당 종의 TCR V 유전자들에서 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 TCR V 유전자들을 합성하여 이 라이브러리를 만들 수 있다. 일부 경우들에서, 종에서 모든 식별된 TCR V 유전자들을 합성하여 이 라이브러리를 만든다. 상기 TCR V 유전자는 TRAV, TRBV, TRGV, 또는 TRDV일 수 있다. 본원에서 기술된 바와 같이, 일부 경우들에서, "V 유전자 생식계열 폴리뉴클레오티드"는 TCR V 유전자의 게놈의 일부 또는 코돈-최적화된 폴리뉴클레오티드를 지칭한다. 상기 TCR V 유전자로부터 유래된 서열은 V 유전자 생식계열 폴리뉴클레오티드일 수 있다. FR3과 불변 영역 (가령, CDR3과 J 영역의 나머지 부분, 본원에서 "CDR3-J"로 지칭됨) 사이의 서열 길이는 적어도 약 10, 20, 30, 40, 50, 60, 70, 80, 90개 또는 그 이상의 뉴클레오티드일 수 있고, 또는 일부 경우들에서, 최대 약 90개 뉴클레오티드 길이일 수 있다. TCR의 알파 쇄와 베타 쇄의 CDR3-J 서열 길이는 적어도 약 50, 60, 70, 80, 90, 100, 120, 150, 180개 또는 그 이상의 뉴클레오티드이다. TCR의 알파 쇄와 베타 쇄의 CDR3-J 서열 (일부 경우들에서, 전부 최대 약 180개 뉴클레오티드 길이)은 올리고 뉴클레오티드 ("쌍을 이룬 CDR3-J 올리고", "쌍을 이룬 CDR3-J 올리고뉴클레오티드" 또는 "쌍을 이룬 CDR3-J 폴리뉴클레오티드", 이들은 호환사용될 수 있다)에 내포될 수 있으며, 이들은 칩-기반 합성에 적합할 수 있다 (도 8에서, 802 및 802를 포괄하는 괄호 Y로 나타냄). 일부 경우들에서, 상기 쌍을 이룬 CDR3-J 폴리뉴클레오티드는 TCR 감마 쇄의 CDR3-J 서열과 TCR 델타 쇄의 CDR3-J 폴리뉴클레오티드를 포함할 수 있다. 본원에서 사용된 바와 같이, 용어 "CDR3-J 폴리뉴클레오티드", "CDR3-J 올리고뉴클레오티드", 및 "CDR3-J 올리고" (이들은 호환사용될 수 있다)는 하나 또는 그 이상의 CDR3-J 서열을 포함하는 폴리뉴클레오티드 서열을 지칭한다. CDR3-J 폴리뉴클레오티드는 쌍을 이룬 CDR3-J 폴리뉴클레오티드 (가령, 쌍을 이룬 TCR 쇄의 CDR3-J 서열을 포함)일 수 있다. CDR3-J 폴리뉴클레오티드 (가령, 쌍을 이루지 않는)는 상기 쌍을 이룬 TCR 쇄중 하나로부터 CDR3-J 서열 만을 함유할 수 있다. 예를 들면, CDR3-J 폴리뉴클레오티드는 TCR 알파 쇄, TCR 베타 쇄, TCR 감마 쇄, 또는 TCR 델타 쇄로부터 오로지 CDR3-J 서열만을 함유할 수 있다. 남은 과제는 이러한 쌍을 이루는 CDR3-J 올리고뉴클레오티드를 고-처리량으로 발현 가능한 TCR 구성물로 전환하는 것 (가령, 한 번의 배지에서 >1,000개의 TCRs을 작제)이다. 본원에 기술된 방법들을 이용하여, 벌크 반응에서 쌍을 이룬 CDR3-J 올리고뉴클레오티드를 이들의 대응하는 V 유전자 생식계열 폴리뉴클레오티드에 연계시킬 수 있다 (가령, 도 8, 803). 일부 경우들에서, 상기 CDR3-J 폴리뉴클레오티드 풀(pool) (가령, 쌍을 이룬 또는 쌍을 이루지 않는)은 적어도 약 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열을 포함할 수 있다. V 유전자 생식계열 폴리뉴클레오티드 라이브러리는 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80개 또는 그 이상의 TCR V 유전자들을 포함할 수 있다. 본원에 기술된 방법들을 이용할 때, 상이한 태생적으로 쌍을 이룬 TCRs을 인코딩하는 적어도 2, 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열 다수가 생성될 수 있다. 상기 태생적으로 쌍을 이룬 TCRs이 단일 구역 안에서 벌크로 생성될 수 있다.

[0065] TCR V 유전자 서열을 포함하는 핵산 구조의 생식계열 또는 재배열된 유전자 구조체 예시들은 도 4A-4C에서 보여준다. 예를 들면, 도 4A에서는 TCR V 유전자의 생식계열 게놈 DNA를 나타내는데, 이들은 L-PART1, V-INTRON, V-EXON, 및 재조합 신호 서열 (V-HEPTAMER, V-SPACER, 및 V-NONAMER)을 포함한다. 두 개의 보존된 시스템인 또한 도 4A에서 보여준다. V-(D)-J 재조합 후, 재배열된 게놈 DNA의 예시적인 구조체는 도 4B 또는 도 4C에서 보여준다. CDR3은 (i) V-REGION과 J-REGION 사이의 접합 (또는 접합 영역) 또는 (ii) V-REGION과 D-REGION 사이의 접합, 그리고 D-REGION과 J-REGION 사이의 접합에 의해 인코딩될 수 있다.

[0066] 상기 TCR V 유전자들은 매우 다변할 수 있다. 인간의 경우, TRA에 대해 40개 이상의 기능성 V 유전자들이 식별되었으며, 예를 들면, TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 및

TRAV41가 내포된다. 이들 V 유전자 중에서, 일부는 동일한 하위그룹으로 분류될 수 있으며, "TRAV" 바로 뒤에는 동일한 하위그룹 번호로 표시되지만, "-" 기호 뒤에는 다른 번호로 표시된다. 예를 들면, TRAV1-1과 TRAV1-2는 동일한 하위그룹에 속한다. 본원에서 사용된 바와 같이, "그룹"은 동일한 "유전자 유형" (가령, V, D, J 또는 C 유형)을 공유하고, 이들 동일한 "쇄 유형"의 폴리펩티드 합성에 잠재적으로 참여하는 유전자들 집단이다. 더 나아가, 그룹에는 허위유전자들(pseudogenes)과 고아유전자(orphanes)가 내포된다. "하위그룹"이란 주어진 종에서 동일한 그룹에 속하고, 뉴클레오티드 수준 (V, D, 및 J의 경우 생식계열 입체배열에서)에서 75% 이상의 동일성을 공유하는 유전자 세트를 의미한다.

[0067]

인간의 경우, TRB에 대해 40개 이상의 기능성 V 유전자들이 식별되었으며, 예를 들면, TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 및 TRBV30이 내포된다. 다른 종, 가령, 마우스의 경우, V 유전자들은 IMGT 데이터베이스에서 찾아볼 수 있다.

[0068]

커넥터 서열을 다양화시킨다

[0069]

V 유전자 생식계열 폴리뉴클레오티드와 CDR3-J 폴리뉴클레오티드의 연계는 분자 생물학 기술, 이를 테면 결찰 (ligation) 및 프라이머 연장의 중첩에 의해 이루어질 수 있다 (도 6). 그러나, 칩-기반 올리고뉴클레오티드 합성력을 최대한으로 이용하기 위해, 하나의 벌크 반응에서 CDR3-J 올리고뉴클레오티드 수천개 또는 그 이상을 이들의 대응하는 V 유전자 생식계열 폴리뉴클레오티드에 연결시킬 수 있다 (도 8, 803 화살표로 나타냄). 이를 수행하는 주요 과제는 V 유전자 생식계열 폴리뉴클레오티드 (도 6, 601)와 CDR3-J (도 6, 602)의 사이의 커넥터 영역은 보존된 FR3 영역일 수 있다. 따라서, 벌크 반응에서, 어떤 V 유전자 생식계열 폴리뉴클레오티드가 어떤 CDR3-J에 연결되어 있는지 제어하기 어려울 수 있다 (도 5에서 나타낸 바와 같이, 실선 화살표는 정확한 V 유전자 생식계열 폴리뉴클레오티드에 대한 연결을 나타내고, 점선 화살표는 맞지 않는 V 유전자 생식계열 폴리뉴클레오티드에 대한 연결을 나타낸다). 예를 들면, TCR 서열은 특정 CDR3-J 베타 서열에 연결된 TRBV4-1에 의해 형성될 수 있다. 상기 벌크 반응에서, TRBV4-1 및 TRBV4-2 모두에 대한 V 유전자 생식계열 폴리뉴클레오티드가 존재할 수 있고, 이들 TRBV 유전자에 대한 FR3 영역은 매우 유사할 수 있다. 따라서, 이 TCR에 대한 CDR3-J 올리고뉴클레오티드가 TRBV4-2 생식계열 폴리뉴클레오티드에 잘 못 연결될 수 있다. 이 문제를 경감시키기 위해, 코돈 다양화를 사용하여 서로 상이한 FR3 서열 간에 유사성을 만들 수 있다. 예를 들면, 상기 커넥터 서열이 코돈-다각화될 수 있어서, 상이한 핵산 서열을 가질 수 있고, 심지어 이들이 동일한 아미노산 서열을 인코딩할 수 있다. 코돈 다양화는 계산 방법, 실시예 2에 나타낸 방법에 의해 이루어질 수 있다. 무작위로 또는 임의의 규칙에 따라 아미노산에 코돈을 할당함으로써, 핵산 서열 다수가 생성될 수 있으며, 여기서 각각의 핵산 서열은 동일한 아미노산 서열을 코딩할 수 있다. 다음으로, 임의적인 규칙에 따라 점수를 할당하기 위해, 핵산 서열 다수가 계산적으로 평가될 수 있다. 상기 임의적인 규칙은 여러 인자들, 이를 테면, 제한절단 부위, 원하지 않는 서열과 혼성화하는 경향, 주어진 서열과 혼성화하는 경향 또는 서열의 원하지 않는 2차 구조와 같은 인자들을 고려할 수 있다. 다음으로, 상기 점수에 기초하여, 핵산 서열은 코돈-다각화된 커넥터 서열로서 핵산 서열 다수부터 선택될 수 있다. 상기 코돈-다각화된 커넥터 서열을 이용하여 CDR3-J 폴리뉴클레오티드와 TCR V 유전자의 명시된 사전-합성된 부분의 정확한 연계를 이룰 수 있다. 알려진 TCR V 유전자들의 일부 또는 전부를 포함하는 "V 유전자 생식계열 폴리뉴클레오티드 라이브러리" 에서, 예를 들면, 각 상이한 TCR V 유전자는 상이한 커넥터 서열을 보유할 수 있고, 이를 이용하여 대응하는 CDR3-J 올리고뉴클레오티드에 정확하게 연결시켜, 참조 서열에 따른 TCR 쇄를 만들 수 있다. 상기 참조 서열은 TCR 쇄들의 동계 쌍을 시퀀싱함으로써 생성될 수 있다. 그러나, 일부 경우들에서, 커넥터 서열이 어느 정보 다양화할 수 있는 지 불분명할 수 있고, 벌크 반응에서 V 유전자 생식계열 폴리뉴클레오티드와 CDR3-J 올리고뉴클레오티드 간의 연결이 어느 정도 정확한지 불분명할 수 있다. 실시예 2에서 나타낸 바와 같이, 인간 TCR V 유전자들의 FR3 영역을 다양화시켜, 임의의 주어진 CDR3-J 서열에 대해 '연결-오류 확률'은 실제로 감지할 수 없게될 수 있다. 실시예 2 c에서 제시된 알고리즘을 이용하여 '코돈-다각화된 V 유전자 생식계열 폴리뉴클레오티드'와 이에 대응하는 CDR3-J 서열을 만들 수 있다.

[0070]

다양한 커넥터 서열 세트가 발견되면, 분자 생물학 기술 (가령, 결찰, 제한 절단, 원형화)을 사용하여 CDR3-J 올리고뉴클레오티드 풀을 발현 가능한 전장의, 발현가능한 TCR 풀로 전환하는 많은 방법을 사용할 수 있다. 실시예 1은 예시적인 작업흐름을 제공한다. 본원에서 제공되는 방법을 벌크 반응에서 개별 TCR 쇄 (가령, 쌍을

이루지 않는쇄)의 풀을 또한 만들 수 있다. 예를 들면, TCR 알파쇄의 풀을 만들기 위해, 각 개별 CDR3-J 올리고뉴클레오타이드는 CDR3과 TCR 알파쇄의 J 영역을 포함할 수 있지만, 또다른 CDR3과 TCR 베타쇄의 J 영역은 포함하지 않을 수 있고, 그 다음 TCR 알파쇄를 만들기 위해 CDR3-J 올리고뉴클레오타이드를 대응하는 TRAV 유전자에 연계시키는데 이용할 수 있다.

[0071] TCRs을 인코딩하는 핵산 분자를 구축하는 방법

[0072] 본원에서 기술하는 TCRs을 인코딩하는 핵산 분자는 두 개 또는 그 이상의 핵산 단편으로부터 구축될 수 있다. 일부 구체예들에서, 상기 두 개 또는 그 이상의 핵산 단편은 제 1 핵산 분자, 제 2 핵산 분자, 제 3 핵산 분자, 제 4 핵산 분자, 등등으로 불릴 수 있다. 핵산 분자를 구축할 때, 혼성화, 연장, 절찰 및 효소적 절단/분열(cleavage)을 비롯한, 그러나 이에 국한되지 않는 표준 분자 생물학 기술이 사용될 수 있다.

[0073] 본원에서 기술한 핵산 단편은 TCR쇄 또는 이의 일부분을 인코딩할 수 있다. 예를 들면, 상기 핵산 단편에 의해 인코딩되는 TCR쇄의 일부분은 약 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 200, 250개, 또는 그 이상의 아미노산을 포함할 수 있다. 상기 핵산 단편은 기능성 TCR쇄를 인코딩하는 서열을 포함할 수 있다. 상기 기능성 TCR쇄는 전장의 TCR쇄일 수도 있고, 또는 아닐 수도 있다. 상기 기능성 TCR쇄는 하나 또는 그 이상의 돌연변이 또는 변형을 포함할 수 있다. 일부 경우들에서, 기능성 TCR쇄는 숙주 세포에서 발현될 때, TCR복합체(가령, TCR α , TCR β , CD3 γ , CD3 δ , CD3 ϵ , 및 ζ 쇄를 갖는 복합체)로 통합될 수 있다. 일부 경우들에서, 기능성 TCR은 이의 표적 리간드에 결합할 수 있다. 일부 경우들에서, 기능성 TCR은 숙주 세포에서 발현될 때, 세포 막으로 통합될 수 있다. 일부 경우들에서, 기능성 TCR은 숙주 세포에서 발현될 수 있다.

[0074] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 CDR3을 인코딩하는 서열을 포함할 수 있다.

[0075] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 제 1 TCR쇄의 제 1 CDR3과 제 2 TCR쇄의 제 2 CDR3을 인코딩하는 서열을 포함할 수 있고, 이때 제 1 CDR3과 제 2 CDR3은 TCR쇄의 동계쌍으로부터 유래될 수 있다. 일부 구체예들에서, 제 1 CDR3을 인코딩하는 서열과 제 2 CDR3을 인코딩하는 서열은 최대 약 100, 90, 80, 70, 60, 50, 40, 30, 20, 10, 또는 5개의 뉴클레오타이드에 의해 분리되어 있다.

[0076] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 TCR V 유전자 서열 또는 이의 일부분을 포함할 수 있다. TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 TCR V 유전자로부터 유래된 서열 서열을 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 V-REGION 핵산 서열을 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2 및/또는 FR3 핵산 서열을 인코딩하는 서열을 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 리더 펩티드를 인코딩하는 서열을 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2 및/또는 FR3 핵산 서열을 인코딩하는 서열을 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 TCR V 유전자의 일부분을 포함하거나, 또는 이의 일부분일 수 있다. 상기 TCR V 유전자의 부분 길이는 적어도 10개 뉴클레오타이드일 수 있다. 예를 들면, 상기 TCR V 유전자의 부분의 길이는 약 10, 20, 30, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000개, 또는 그 이상의 뉴클레오타이드일 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 하나 또는 그 이상의 변형된 뉴클레오타이드를 포함할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 코돈-최적화되고(또는 코돈-다각화되어), 이 TCR V 유전자 또는 이의 일부분과는 상이한 서열을 갖지만, 그러나 동일한 아미노산 서열을 인코딩할 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 CDR3의 일부분을 인코딩하는 서열을 포함하지 않을 수 있다. 상기 TCR V 유전자로부터 유래된 서열은 재배열된 유전자의 접합 영역의 서열을 포함하지 않을 수 있다.

[0077] 상기 TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 불변 도메인을 인코딩하는 서열 또는 이의 일부분을 포함할 수 있다. 상기 TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 불변 영역 또는 이의 일부분을 인코딩하는 서열을 포함할 수 있다. 일부 경우들에서, 상기 불변 도메인 또는 불변 영역은 TCR알파 불변 도메인 또는 불변 영역, TCR베타 불변 도메인 또는 불변 영역, TCR감마 불변 도메인 또는 불변 영역, 또는 TCR델타 불변 도메인 또는 불변 영역이다. 일부 경우들에서, 상기 불변 영역은 불변 도메인을 포함한다. 일부 경우들에서, 상기 불변 영역은 막경유 영역, 연결 영역, 세포질 영역, 또는 이의 조합을 더 포함한다.

[0078] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 커넥터 서열을 포함할 수 있

다. 상기 커넥터 서열을 이용하여 하나의 핵산 분자를 또다른 핵산 분자에 연계시킬 수 있다. 상기 하나의 핵산 분자의 커넥터 서열은 또다른 핵산 분자의 안티-커넥터 서열과 혼성화될 수 있다(가령, 염기쌍 형성 또는 염기쌍들을 형성) 상기 안티-커넥터 서열은 상기 커넥터 서열과 상보적(가령, 완전하게 상보적 또는 실질적으로 상보적)일 수 있다. 상기 안티-커넥터 서열은 특정 조건(가령, 온도, 완충 조건, pH, 등) 하에서 상기 커넥터 서열과 혼성화가능할 수 있다. 상기 안티-커넥터 서열은 상기 커넥터 서열의 역-보체 서열(또는 상보적 서열)일 수 있다. 상기 커넥터 서열과 상기 안티-커넥터 서열이 혼성화될 때, 형성된 염기쌍(들)은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100개, 또는 그 이상의 염기쌍일 수 있다. 상기 커넥터 서열과 상기 안티-커넥터 서열 사이에 형성된 염기쌍은 인접되거나 또는 비-인접될 수 있다. 예를 들면, 비-인접 염기쌍이 형성되는 경우, 쌍을 이루지 않는 영역이 있을 수 있거나, 또는 쌍을 이룬 영역을 구별짓는 영역이 있을 수 있다. 제 1 핵산 분자 커넥터 서열을 포함한다면, 그 다음 제 2 핵산 분자 상의 커넥터 서열의 상보적 서열을 안티-커넥터 서열로 지칭할 수 있다. 상기 커넥터 서열(또는 안티-커넥터 서열)의 길이는 다양할 수 있다. 예를 들면, 상기 커넥터 서열(또는 안티-커넥터 서열)의 길이는 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 100개, 또는 그 이상의 뉴클레오타이드일 수 있다. 상기 커넥터 서열(또는 안티-커넥터 서열)의 길이는 약 300, 250, 200, 150, 100, 90, 85, 80, 75, 70, 65, 60, 55, 50, 45, 40, 35, 30, 25, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3 또는 2개 뉴클레오타이드일 수 있다. 상기 커넥터 서열(또는 안티-커넥터 서열)은 핵산 분자의 5' 단부 또는 3' 단부에 있을 수 있다. 상기 커넥터 서열(또는 안티-커넥터 서열)은 또한 핵산 분자의 내부 서열일 수도 있다. 예를 들면, 상기 커넥터 서열이 내부 커넥터 서열일 수 있고, 해당 핵산 분자의 내부 서열(가령, 해당 내부 커넥터 서열의 옆에 있는 서열)을 절단함으로써, 5' 단부 또는 3' 단부에 노출될 수 있다. 상기 내부 커넥터 서열의 예는 **실시예 1**, 내부-쇄 커넥터(ICC)로 제공된다. 일부 경우들에서, 커넥터 서열과 안티-커넥터 서열을 이용하여 TCR 쇄의 CDR3 또는 이의 일부분을 인코딩하는 핵산 분자와 TCR V 유전자 또는 이의 일부분을 포함하는 또다른 핵산 분자를 연계시킨다. 일부 경우들에서, 커넥터 서열과 안티-커넥터 서열을 이용하여 TCR의 J 영역을 인코딩하는 핵산 분자와 TCR V 유전자 또는 이의 일부분을 포함하는 또다른 핵산 분자를 연계시킨다. 일부 경우들에서, 커넥터 서열과 안티-커넥터 서열을 이용하여 CDR3 또는 이의 일부분 및 TCR의 J 영역을 인코딩하는 핵산 분자와 TCR V 유전자 또는 일부분을 포함하는 또다른 핵산 분자를 연계시킨다. 일부 경우들에서, 커넥터 서열과 안티-커넥터 서열을 이용하여 CDR3 또는 이의 일부분, J 영역, 및 TCR V 유전자 또는 이의 일부분을 인코딩하는 서열을 포함하는 핵산 분자와 TCR의 불변 도메인 또는 이의 일부분을 인코딩하는 또다른 핵산 분자를 연계시킨다.

[0079] 상기 커넥터 서열(또는 상기 안티-커넥터 서열)은 TCR V 유전자의 부분(가령, 재배열된 유전자에서 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자의 부분)을 인코딩하는 서열일 수 있다. 그리고 이러한 경우에서, 커넥터 서열 풀에서 커넥터 서열과 하나 또는 그 이상의 다른 커넥터 서열은 동일한 아미노산 서열(가령, CDR3에 인접한 TCR V 유전자의 보존된 부분)을 인코딩할 수 있다. 상기 커넥터 서열이 TCR V 유전자의 보존된 부분을 인코딩할 때, 상기 커넥터 서열은 코돈-다각화될 수 있고, 해당 커넥터 서열을 이용하여 핵산 분자를 또다른 핵산 분자에 특이적으로 연계시켜, TCR 동계 쌍을 인코딩하는 핵산 분자가 구축된다. 일부 구체예들에서, 상기 커넥터 서열(또는 안티-커넥터 서열)은 재배열된 유전자에서 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자의 적어도 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 50, 60, 70, 80, 90, 100, 150, 200개, 또는 그 이상의 뉴클레오타이드를 포함한다. 상기 커넥터 서열과 안티-커넥터 서열의 특이성 때문에, 벌크 반응(가령, 동일한 구획내)에서 상이한 TCRs를 인코딩하는 상이한 서열을 갖는 핵산 분자의 풀이 구축될 수 있다. 상기 커넥터 서열은 CDR3를 인코딩하는 서열이 참조 서열(가령, 시퀀싱에 의해 결정된 TCR 쇄의 고유 서열)에 따라 연결되어야 하는 TCR V 유전자를 규정할 수 있다. 상기 커넥터 서열(또는 상기 안티-커넥터 서열)은 TCR V 유전자의 일부를 인코딩하지 않을 수 있는 임의적인(가령, 사전-결정된) 서열일 수 있다. 그리고 이러한 경우, 두 핵산 단편을 함께 연계한 후, 상기 임의적인 서열을 제거할 수 있다.

[0080] **도 7**은 임의적인 커넥터(701) 및 안티-커넥터(702) 서열을 이용하여 CDR3-J 폴리뉴클레오타이드를 명시된 V 유전자 생식계열 폴리뉴클레오타이드에 연계시키는(가는 화살표) 예시를 나타낸다. 여기에서, 각 V 유전자 생식계열 폴리뉴클레오타이드는 부분적으로 이중-가닥으로 된 구조를 갖는다. 이 도면에서 상부 가닥은 이의 우측에 3' 단부를 갖고, 이때 이의 3' 단부에 단일-가닥으로 된 영역을 보유한다. 상기 커넥터 서열과 안티-커넥터 서열은 단일 가닥일 수 있고, 서로 혼성화될 수 있다. 상기 커넥터 서열과 안티-커넥터 서열은 특정 혼성화 목적만을 제공하며, TCR과 전혀 무관할 수 있으며, 따라서 임의적이다. 상기 커넥터와 안티-커넥터 간에 혼성화 이후, V 유전자 생식계열 폴리뉴클레오타이드의 상부 가닥의 3' 단부는 CDR3-J 폴리뉴클레오타이드에 혼성화될 수 있고,

DNA 중합효소에 의해 연장될 수 있다. CDR3-J 폴리뉴클레오타이드에 혼성화되는 V 유전자 생식계열 폴리뉴클레오타이드의 상부 가닥 3' 단부 상에 있는 뉴클레오타이드의 수는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 또는 15개 일 수 있고, 또는 최대 20개가 될 수 있다.

[0081] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 자가-절단 펩티드를 포함할 수 있다. 상기 자가-절단 펩티드는 2A 펩티드, intein 펩티드, 또는 hedgehog 펩티드일 수 있다. 2A 펩티드의 예로는 P2A (가령, 서열: ATNFSLLKQAGDVEENPGP(서열 식별 번호:284)), E2A (가령, 서열 QCTNYALLKLAGDVESNPGP)(서열 식별 번호:285), F2A (가령, 서열 VKQTLNFDLLKLAGDVESNPGP)(서열 식별 번호:286), 및 T2A (가령, 서열 EGRGSLLTCTGDVEENPGP)(서열 식별 번호:287)펩티드가 내포되나, 이에 국한되지 않는다.

[0082] TCR 또는 이의 일부분을 인코딩하는 핵산 분자를 구축하는데 이용되는 핵산 단편은 제한절단 효소 인지 부위를 포함할 수 있다. 예를 들면, 상기 제한절단 효소 인지 부위는 유형 IIS 제한절단 효소에 대한 인지 부위일 수 있다. 본 명세서에서 유용할 수 있는 유형-IIS 제한절단 효소들의 예로는 EarI, MnlI, PleI, AlwI, BbsI, BbvI, BcoDI, BsaI, BseRI, BsmAI, BsmBI, BspMI, Esp3I, HgaI, SapI, SfaNI, BbvI, BsmFI, BsrDI, BtsI, FokI, BseRI, HphI, MlyI 및 MboII이 내포되나, 이에 국한되지 않는다. 일부 경우들에서, 두 개 또는 그 이상의 상이한 제한절단 효소를 핵산 구축 공정 동안 이용할 수 있다. 일부 경우들에서, 4-bp 5' 오버행을 만드는 제한 절단 효소 (예를 들면, BbsI, BbvI, BcoDI, BsaI, BsmBI, FokI, 등)를 이용할 수 있다. 일부 경우들에서, 블런트 단부 또는 3' 오버행을 만드는 제한 절단 효소 (예를 들면, BseRI, BsrDI, BtsI, MlyI, 등)를 이용할 수 있다.

[0083] TCR 또는 이의 일부분을 인코딩하는 핵산 분자 구축에 이용되는 핵산 단편은 고리화(circularized)될 수 있다. 예를 들면, 상기 핵산 단편은 해당 핵산 단편의 두 개 단부를 결합을 통하여 이음으로써 고리화될 수 있다. 상기 결합은 블런트 단부 결합일 수 있다. 5'-에서-3'로의 엑소뉴클레아제 (가령, Gibson 어셈블리), 3'-에서-5'로의 엑소뉴클레아제 (가령, 서열 및 리가제 독립적 클로닝 또는 SLIC), 또는 USER 효소 믹스 (가령, USER 친화적 DNA 재조합 또는 USERec)를 이용하여 스틱커 단부를 만든 후에 결합을 수행할 수 있다 고리화 방법의 추가 예로는 원형 중합효소 연장 클로닝 (CPEC) 및 심리스(seamless) 결합 클로닝 추출 (SLiCE) 어셈블리가 내포되나, 이에 국한되지 않는다. 대안적으로, 중첩 PCR에 의해 이들 두 단부를 이을 수 있다. 예를 들면, T4 DNA 리가제, T4 RNA 리가제, 대장균(E. coli) DNA 리가제를 비롯한, 그러나 이에 국한되지 않는, 다양한 리가제들이 결합에 이용될 수 있다.

[0084] 상기 TCR쇄 또는 이의 일부분을 인코딩하는 핵산 분자 구축에 이용되는 핵산 단편은 화학적으로 합성될 수 있다. 예를 들면, 상기 핵산 단편을 칩-기반 합성에 의해 사전-합성할 수 있다. 일부 경우들에서, 상기 합성된 핵산 단편의 길이는 약 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400개, 또는 그 이상의 뉴클레오타이드일 수 있다. 일부 경우들에서, 상기 합성된 핵산 단편의 길이는 약 500, 450, 400, 350, 300, 250, 200, 150, 100, 90, 80, 70, 60, 50, 40, 30, 20, 또는 10개 뉴클레오타이드이거나, 또는 이보다 적을 수 있다.

[0085] TCR의 두 펩티드쇄를 인코딩하는 두개의 핵산 서열은 몇 가지 방향, 예를 들면, 머리에서-머리, 머리에서-꼬리, 및 꼬리에서-꼬리 방향으로 구축될 수 있다. 본원에서 기술된 바와 같이, "머리"란 센스 핵산 가닥의 "5' 단부"를 지칭하고, "꼬리"는 센스 핵산 가닥의 "3' 단부"를 지칭한다. 일부 경우들에서, 상기 방향은 머리에서-꼬리 방향이며, TCR을 인코딩하는 쌍을 이룬 핵산 서열 (가령, TRA 다음 TRB, 또는 TRB 다음 TRA)이 조절될 수 있다.

[0086] 본원에서 기술된 임의의 핵산 분자는 이중-가닥으로 된 핵산 분자 또는 단일-가닥으로 된 핵산 분자일 수 있다. 일부 경우들에서, 핵산 분자는 이중-가닥으로 된 영역과 단일-가닥으로 된 영역을 포함할 수 있다. 예를 들면, 커넥터 서열 또는 안티-커넥터 서열을 갖는 핵산 분자는 단일-가닥으로 된 영역 (가령, 오버행 또는 스틱커 단부)으로 커넥터 서열 영역 또는 안티-커넥터 서열 영역을 갖는 이중-가닥으로 된 핵산 분자일 수 있다. 상기 오버행의 길이는 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 그 이상의 뉴클레오타이드일 수 있다. 상기 오버행은 핵산 분자의 5' 단부 또는 3' 단부에 있을 수 있다.

[0087] 본원에서 기술된 임의의 핵산 분자는 하나 또는 그 이상의 변형된 뉴클레오타이드를 포함할 수 있다. 변형된 뉴클레오타이드의 예들에는, 다이아미노퓨린, 5-플루오로우라실, 5-브로모우라실, 5-클로로우라실, 5-아이오도우라실, 하이포잔틴, 잔틴, 4-아세틸시토신, 5-(카르복시하이드록실메틸)우라실, 5-카르복시메틸아미노메틸-2-티오우리딘, 5-카르복시메틸아미노메틸우라실, 다이하이드로우라실, 베타-D-갈락토실큐에오신, 이노신, N6-아이소펜텐일아데닌, 1-메틸구아닌, 1-메틸이노신, 2,2-다이메틸구아닌, 2-메틸아데닌, 2-메틸구아닌, 3-메틸시토신, 5-

메틸시토신, N6-아데닌, 7-메틸구아닌, 5-메틸아미노메틸우라실, 5-메톡시아미노메틸-2-티오우라실, 베타-D-만노실쿠에오신, 5'-메톡시카르복시메틸우라실, 5-메톡시우라실, 2-메틸티오-D46-아이스오펜일아데닌, 우라실-5-옥시아세트산(v), 위부톡소신, 수도우라실, 쿠에오신, 2-티오시토신, 5-메틸-2-티오우라실, 2-티오우라실, 4-티오우라실, 5-메틸우라실, 우라실-5-옥시아세트산 메틸에스테르, 우라실-5-옥시아세트산 (v), 5-메틸-2-티오우라실, 3-(3-아미노-3-N-2-카르복시프로필)우라실, (acp3)w, 2,6-다이아미노퓨린 등이 포함되나, 이에 제한되는 것은 아니다. 일부 경우들에서, 뉴클레오티드는 트리포스페이트 모이어티에 대한 변형을 비롯하여, 이들의 포스페이트 모이어티에서 변형을 내포할 수 있다. 이러한 변형의 비-제한적인 예시에는 더 긴 길이의 포스페이트 쇠(가령, 4, 5, 6, 7, 8, 9, 10개 또는 그 이상의 포스페이트 모이어티를 갖는 포스페이트 쇠) 및 티올 모이어티(가령, 알파-티오텐리포스페이트 및 베타-티오텐리포스페이트)를 갖는 변형이 내포된다. 핵산 분자들은 또한 염기 모이어티(예컨대, 통상적으로 상보적 뉴클레오티드와 수소 결합을 형성할 수 있는 하나 이상의 원자들 및/또는 전형적으로 상보적 뉴클레오티드와 수소 결합을 형성할 수 없는 하나 이상의 원자들), 당 모이어티 또는 포스페이트 골격에서 변형될 수도 있다. 핵산 분자는 아민-변형된 그룹, 가령, 아미노알릴 1-dUTP (aa-dUTP) 및 아미노핵심아크릴아미드-dCTP (aha-dCTP)을 또한 함유하여, 아민 반응성 모이어티, 가령, N-하이드록시 숙신이미드 에스터 (NHS)의 공유 부착을 가능하게 할 수 있다. 본 명세서의 올리고뉴클레오티드에서 표준 DNA 염기쌍 또는 RNA 염기쌍에 대한 대체는 입방 mm당 비트 수로 더 높은 밀도, 더 높은 안전성(우발적이거나 또는 의도적인 천연 독소 합성에 대한 내성), 광-프로그래밍된 중합효소에서 더 용이한 분별, 또는 더 낮은 2차 구조를 제공할 수 있다. 이러한 대체 염기쌍은 새로운 및/또는 증폭 합성을 위한 천연 및 돌연변이 중합효소와 양립할 수 있다.

[0088] TCRs를 인코딩하는 핵산 분자 구축의 예시적인 작업흐름은 도 1A-1C에 나타낸다. TCRs를 인코딩하는 핵산 분자를 구축하기 위해 본원에 기술된 방법을 사용하기에 앞서, 다양한 기존 방법(가령, 단일 세포 바코딩 및 시퀀싱)을 이용하여 TCRs의 동계 쌍 다수를 사전-결정할 수 있다. 다양한 시퀀싱 방법을 이용하여 쌍을 이룬 TCR 쌍들의 서열을 결정할 수 있는데, 예를 들면, Sanger 시퀀싱, 고 처리량 시퀀싱, 합성단위 시퀀싱, 단일 분자 시퀀싱, 연결단위 시퀀싱, RNA-Seq (Illumina), 차세대 시퀀싱, 디지털 유전자 발현 (Helicos), Clonal Single MicroArray (Solexa), 샷건 시퀀싱, Maxim-Gilbert 시퀀싱 또는 대규모 병렬 시퀀싱이 사용될 수 있다. 상기 시퀀싱 라이브러리로부터 쌍을 이룬 서열은 TCR 쌍들에 대한 동계 쌍에 대한 참조 서열로 삼을 수 있고, 커넥터 서열과 안티-커넥터 서열 사이의 특이적 상호작용을 통해 어떤 CDR3가 어떤 V 유전자와 쌍을 이루는지 알 수 있다. 본원에 기술된 방법들을 이용하여 상이한 TCRs를 인코딩하는 핵산 분자 다수를 벌크로 구축할 수 있는데, 한 가지 분자의 구축을 도 1A-1C에서 예로써 나타낸다. 제 1 CDR3 (가령, CDR3 α) 및 제 2 CDR3 (가령, CDR3 β)를 인코딩하는 서열을 포함하는 제 1 핵산 분자를 제 1 TCR V 유전자 (가령, TRAV)로부터 유래된 서열을 포함하는 제 2 핵산 분자에 접촉시킬 수 있다. 제 1 핵산 분자의 커넥터 서열 (가령, ConA#*)은 제 2 핵산 분자의 안티-커넥터 서열 (ConA#)에 혼성화되어, 두 핵산 분자를 연계시킬 수 있다. 연장 및 절찰을 실행하여 제 1 TCR V 유전자로부터 유래된 서열 및 제 1 CDR3 및 제 2 CDR3를 인코딩하는 서열을 포함하는 제 3 핵산 분자를 생성시킬 수 있다. 그 다음, 제한절단 효소 (가령, 도 1A의 TIISRE1)를 이용하여 제 3 핵산 분자의 오버행 (또는 스틱커 단부)을 만들 수 있다. 그 다음, 제 3 핵산 분자를 제 1 불변 영역 또는 불변 도메인 (가령, TRBC)을 인코딩하는 서열을 포함하는 제 4 핵산 분자에 접촉시킬 수 있다. 그 다음, 제 3 핵산 분자를 오버행을 통하여 제 4 핵산 분자에 절찰시켜, 제 1 TCR V 유전자로부터 유래된 서열, 제 1 CDR3 및 제 2 CDR3를 인코딩하는 서열, 및 제 1 불변 영역을 인코딩하는 서열을 포함하는 제 5 핵산 분자를 만들 수 있다. 상기 제 5 핵산 분자를 고리화시키고, 제한절단 효소 (가령, TIISRE3)로 절단하여, 내부 커넥터 서열 (가령, ICC)을 노출시킬 수 있다. 그 다음, 상기 제 5 핵산 분자를 제 2 TCR V 유전자 (가령, TRBV)로부터 유래된 서열을 포함하는 제 6 핵산 분자에 접촉시킬 수 있다. 상기 제 5 핵산 분자를 커넥터 서열과 안티-커넥터 서열의 상호작용을 통하여 제 6 핵산 분자에 절찰시킬 수 있다. 그 다음, 상기 제 6 핵산 분자를 제한절단 효소 (가령, TIISRE2)를 이용하여 절단하여, 오버행을 만들 수 있다. 그 다음, 상기 제 6 핵산 분자를 제 2 불변 영역 또는 불변 도메인 (가령, TRAC)을 인코딩하는 서열을 포함하는 제 7 핵산 분자에 접촉시킬 수 있다. 상기 제 6 핵산 분자와 제 7 핵산 분자를 절찰시켜 쌍을 이룬 TCR 쌍들을 인코딩하는 모든 영역을 포함하는 제 8 핵산 분자를 만들 수 있다. 숙주 세포에서 TCR 쌍 발현을 위한 발현 벡터로 상기 제 8 핵산 분자를 더 구축시킬 수 있다. TCR V 유전자로부터 유래된 서열을 포함하는 핵산 단편은 단일-가닥일 수 있으며, 그러한 경우, CDR3를 인코딩하는 핵산 단편의 커넥터 서열의 3' 단부는 안티-커넥터 서열과의 혼성화 시점에 연장될 수 있다.

[0089] 본원에 기술된 방법들을 이용하여 개별 TCR 쌍들의 풀, 예를 들면, TCR 알파 쌍들 또는 TCR 베타 쌍들의 풀을 만들 수 있다.

[0090] 본원에서 기술된 핵산 분자들 다수를 만드는 방법들은 제 1 다수의 핵산 분자 (또는 핵산 단편)를 제공하는 것

을 포함한다. 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함할 수 있다. 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래될 수 있다. 그 다음, 제 2 다수의 핵산 분자들이 제공될 수 있다. 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. 상기 핵산 분자는 불변 도메인을 인코딩하는 서열을 포함하지 않을 수 있다. 그 다음, 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시킬 수 있다. 제 1 다수의 핵산 분자들에서 핵산 분자를 제 2 다수의 핵산 분자들에서 핵산 분자에 연계시켜, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 만들 수 있다. 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 TCR 쉼의 동계 쌍으로부터 유래될 수 있다.

[0091] 핵산 분자들 다수를 만드는 방법 (이때 이들 다수의 핵산 분자중에서 각 핵산 분자 T-세포 수용체 (TCR) 쉼 또는 이의 영역을 인코딩함)은 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시켜, 적어도 두 개의 상이한 핵산 분자를 포함하는 제 3 다수의 핵산 분자를 만드는 것을 포함한다. 상기 적어도 두 개의 상이한 핵산 분자에서 각 핵산 분자는 상이한 TCR 쉼 또는 이의 영역을 인코딩하는 상이한 서열을 보유할 수 있다. 상기 적어도 두 개의 상이한 핵산 분자는 동일한 구획 안에서 생성될 수 있다. 일부 경우들에서, 상이한 TCRs을 인코딩하는 적어도 약 5, 10, 20, 50, 100, 200, 300, 400, 500, 1,000, 1,500, 2,000, 2,500, 3,000, 3,500, 4,000, 5,000, 6,000, 7,000, 8,000, 9,000, 10,000, 12,000, 15,000, 20,000, 100,000, 1,000,000, 10,000,000개, 또는 그 이상의 상이한 서열이 동일한 구획 안에서 생성될 수 있다.

[0092] 본원에서 기술된 핵산 분자들 다수를 만드는 방법은 제 1 다수의 핵산 분자의 제공을 포함할 수 있다. 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함할 수 있다. 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래될 수 있다. 그 다음, 제 2 다수의 핵산 분자들이 제공될 수 있다. 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. 그 다음, 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시킬 수 있다. 제 1 다수의 핵산 분자들에서 핵산 분자를 제 2 다수의 핵산 분자들에서 핵산 분자에 연계시켜, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 선형 핵산 분자를 만들 수 있다. 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 TCR 쉼의 동계 쌍으로부터 유래될 수 있다.

[0093] 핵산 분자들 다수를 만드는 방법은 제 1 다수의 핵산 분자의 제공을 포함할 수 있다. 제 1 다수의 핵산 분자들에서 핵산 분자는 (i) 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 합성 서열, 그리고 (ii) 제 3 T-세포 수용체 (TCR) 쉼의 제 3의 CDR3 및 제 4의 TCR 쉼의 제 4의 CDR3을 인코딩하는 합성 서열을 포함할 수 있다. 제 1 CDR3 및 제 2 CDR3은 제 1 TCR 쉼의 동계 쌍으로부터 유래될 수 있고, 제 3 CDR3 및 제 4 CDR3은 제 2 TCR 쉼의 동계 쌍으로부터 유래될 수 있다. 그 다음, 제 2 다수의 핵산 분자들이 제공될 수 있다. 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. 그 다음, 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자를 접촉시킬 수 있다. 제 1 다수의 핵산 분자들에서 핵산 분자를 제 2 다수의 핵산 분자들에서 핵산 분자에 연계시켜, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 만들 수 있다. 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 TCR 쉼의 동계 쌍으로부터 유래될 수 있다.

[0094] T-세포 수용체를 인코딩하는 (TCR) 쉼 또는 이의 일부분을 인코딩하는 핵산 분자를 만드는 방법은 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자의 제공을 포함할 수 있다. 그 다음, 핵산 분자들 다수가 제공될 수 있다. 이들 다수의 핵산 분자중에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. 상기 다수의 핵산 분자는 적어도 두 개의 상이한 TCR V 유전자들로부터 유래된 적어도 두 개의 상이한 서열을 포함할 수 있다. 일부 경우들에서, 상기 다수의 핵산 분자는 적어도 2, 5, 10, 15, 20, 25, 30, 35, 40개 또는 그 이상의 상이한 TCR V 유전자들로부터 유래된 적어도 2, 5, 10, 15, 20, 25, 30, 35, 40개 또는 그 이상의 상이한 서열을 포함할 수 있다. 그 다음, 상기 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 동일한 구획 내에서 다수의 핵산 분자에 접촉시킬 수 있다(이때 각 핵산 분자는 각 TCR V 유전자로부터 유래된 서열을 포함한다). 상기 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 상기 다수 핵산 분자들중 핵산 분자에 연계시켜, CDR3을 인코딩하는 서열 및 상기 적어도 두 개의 상이한 TCR V 유전자들중 하나로부터 유래된 서열을 포함하는 제 3 핵산 분자를 만들 수 있고, 그렇게 함으로써 상기 TCR 쉼 또는 이의 일부분을 인코딩하는 핵산 분자가 생성될 수 있다.

[0095] 본원에 기술된 방법들에 이용될 수 있는 본원에서 기술된 조성물은 제 1 다수의 핵산 분자를 포함할 수 있다. 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2

CDR3을 인코딩하는 서열을 포함할 수 있다. 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래될 수 있다. 상기 조성물은 제 2 다수의 핵산 분자들을 더 포함할 수 있다. 제 2 다수의 핵산 분자들에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함할 수 있다. 제 2 다수의 핵산 분자들에서 각 핵산 분자는 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열을 포함하지 않을 수 있다. 이 조성물에서, (i) 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 제 1 CDR3 및/또는 제 2 CDR3을 인코딩하는 서열을 포함할 수 있고, 및/또는 (ii) 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함할 수 있다.

[0096] 본원에 기술된 방법들에 이용될 수 있는 본원에서 기술된 조성물은 다수의 핵산 분자를 포함할 수 있다. 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함할 수 있다. 상기 다수의 핵산 분자는 제 1 커넥터 서열을 갖는 제 1 핵산 분자와 제 2 커넥터 서열을 갖는 제 2 핵산 분자를 포함할 수 있다. 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이할 수 있다.

[0097] 본원에 기술된 방법들에 이용될 수 있는 본원에서 기술된 조성물은 다수의 핵산 분자를 포함할 수 있다. 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) 쉼의 CDR3을 인코딩할 수 있다. 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 커넥터 서열을 포함할 수 있고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 커넥터 서열을 포함할 수 있다. 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이할 수 있다.

[0098] 본원에 기술된 방법들에 이용될 수 있는 본원에서 기술된 조성물은 다수의 핵산 분자를 포함할 수 있다. 이들 다수의 핵산 분자들중에서 각 핵산 분자는 T-세포 수용체 (TCR) 쉼의 적어도 10개 아미노산을 인코딩하는 서열을 포함할 수 있다. 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 커넥터 서열을 포함할 수 있고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 커넥터 서열을 포함할 수 있다. 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이할 수 있다. 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쉼의 일부분을 인코딩할 수 있다. 제 1 커넥터 서열 또는 제 2 커넥터 서열은 상기 TCR 쉼의 적어도 10개 아미노산을 인코딩하는 서열과 프레임 안에 있을 수 있다.

[0099] 본원에 기술된 방법들에 이용될 수 있는 본원에서 기술된 조성물은 다수의 핵산 분자를 포함할 수 있다. 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함할 수 있고, CDR3 서열을 포함하지 않을 수 있다. 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 안티-커넥터 서열을 포함할 수 있고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 안티-커넥터 서열을 포함할 수 있다. 제 1 안티-커넥터 서열은 제 2 안티-커넥터 서열과는 상이할 수 있다. 제 1 핵산 분자와 제 2 핵산 분자의 TCR V 유전자로부터 유래된 서열은 상이한 TCR V 유전자로부터 유래될 수 있다. 상기 조성물은 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 더 포함할 수 있다. 상기 적어도 하나의 핵산 분자는 제 1 안티-커넥터 서열에 상보적인 제 1 커넥터 서열을 더 포함할 수 있다.

[0100] 본 명세서는 다수의 TCRs를 포함하는 TCR 라이브러리를 어셈블리 또는 합성하기 위한 조성물 및 방법을 제공한다. 일부 경우들에서, 추가 특징화 또는 조작을 위해, 상기 TCR 라이브러리로부터 특정 TCR 서열 (가령, 관심 대상-TCR)을 분리 또는 정제하는데 유용할 수 있다. 이를 위해, TCR 또는 이의 일부분을 인코딩하는 서열을 구축하는데 이용되는 핵산 분자 또는 단편 안에 바코드가 내포될 수 있다. 일부 경우들에서, CDR3을 인코딩하는 서열을 포함하는 핵산 단편은 바코드를 포함한다. 일부 경우들에서, 제 1 TCR 쉼의 제 1 CDR3과 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함하는 핵산 단편은 바코드를 포함한다. 예를 들면, CDR3-J 올리고 또는 쌍을 이룬 CDR3-J 올리고는 바코드를 포함할 수 있다. 상기 커넥터 서열 (또는 일부 경우들에서, 상기 안티-커넥터 서열)은 바코드를 포함할 수 있다. CDR3-J 올리고의 쉼-간(inter-chain) 커넥터 (또는 ICC)는 바코드를 포함할 수 있다. 상기 바코드는 프라이머 결합 부위, 가령, TCR-특이적 프라이머-결합 부위 또는 DOPBS일 수 있다.

[0101] 예를 들면, 쌍을 이룬 CDR3-J 올리고 풀에서 특유 쌍을 이룬 CDR3-J를 인코딩하는 각 서열 (가령, 도 1A)은 특유 바코드 (또는 특유 DOPBS)를 포함할 수 있다. DOPBSes의 서열을 임의적으로 기획할 수 있다. 일반적인 곤란함, 이를 테면, 원치않는 이차 구조, 제한 부위, TCR 유전자들에서 다른 서열과의 유사성, 또는 프라이머-결합 부위 간의 유사성을 회피하도록, DOPBSes의 서열을 기획할 수 있다. 바코드 (또는 DOPBS)의 길이는 적어도 약 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40개, 또는 그 이상의 뉴클레오타이드일 수 있다. DOPBS는 쌍을 이룬 CDR3-J 풀의 각 서열 안에 내포된 추가 서열일 수 있다. DOPBS는 쌍을 이룬 CDR3-J 풀의 각 서열 안에 이미 내포된 서열일 수 있다. 예를 들면, 커넥터 서열 또는 이의 일부분을 DOPBS로 이용할 수 있다. 표 3에 열거된 서열을 DOPBSes로 이용할 수 있다. T2A-3에 대응하는 포워드 프라이머, 그리고 관심대상-TCR과 연합된

DOPBS에 대응하는 리버스 프라이머를 이용하는 다이알-아웃 PCR에서 도 1C에서 단계(9)의 산물을 주형으로 이용할 수 있다. 해당 PCR 산물은 도 1C의 단계(10) 및 단계 (11)을 거칠 수 있다. 최종 산물은 관심대상-TCR을 주로 함유할 수 있다.

[0102] TCRs의 발현

[0103] 본원에서 제공되는 방법을 이용하여, 핵산 분자들의 풀 (이때, 각 핵산 분자는 TCR 또는 이의 일부분을 인코딩한다)은 발현을 위해 숙주 세포로 추가 운반될 수 있다. 상기 구축된 핵산 분자를 숙주 세포 안에서 발현되도록 벡터 안에 삽입시킬 수 있다. 상기 구축된 핵산 분자를 선형 또는 원형 핵산 가닥으로 수령 세포로 운반할 수 있다. 일부 경우들에서, 상기 구축된 핵산 또는 구축된 핵산을 포함하는 벡터를 전기 천공에 의해 수령 세포로 운반할 수 있다. 일부 경우들에서, 상기 구축된 핵산 또는 구축된 핵산을 포함하는 벡터를 운반체, 이를 테면, 양이온 중합체를 이용하여 운반할 수 있다.

[0104] 상기 벡터는 플라스미드, 트랜스포존 (가령, Sleeping Beauty, Piggy Bac), 아데노바이러스 벡터, AAV 벡터, 레트로바이러스 벡터 또는 렌티바이러스 벡터일 수 있다. 벡터의 비제한적인 예로는 플라스미드, 셔플 벡터, 파지미드, 코스미드, 비리온, 레트로바이러스 벡터, 아데노바이러스 벡터 또는 유전자 치료에 일반적으로 사용되는 입자 및/또는 벡터가 내포된다. 적합한 플라스미드 벡터의 비-제한적 실시예는 pUC, pBR322, pET, pBluescript, 그리고 이의 변이체들을 포함한다. 더욱이, 상기 벡터는 추가 발현 조절 서열 (예로써, 인핸서 서열, Kozak 서열, 폴리아데닐화 서열, 전사 종료 서열, 등등), 선별가능한 표지 서열 (예로써, 항생제 저항성 유전자), 복제 원점, 및 이와 유사한 것을 포함할 수 있다. 벡터는 숙주 세포에서 복제를 허용하는 핵산 서열, 이를 테면 복제 원점을 포함할 수 있다. 벡터는 하나 또는 그 이상의 선택가능한 마커 유전자 및 다른 유전적 요소를 포함할 수 있다. 벡터는 TCR의 발현을 허용하는 서열에 작동가능하게 연결된 본 명세서에 따른 TCR 또는 이의 일부분을 인코딩하는 구축된 핵산 서열이 내포된 발현 벡터일 수 있다. 벡터의 추가 예시에는 다음이 내포되나, 이에 국한되지 않는다: 바이러스 벡터 및 비-바이러스 벡터, 이를 테면 레트로바이러스 벡터 (렌티바이러스 벡터 포함), 복제 컴피턴트 및 복제 결함 및 이의 무기력 형태(gutless)를 비롯한 아데노바이러스 벡터, 아데노-연합된 바이러스 (AAV) 벡터, 원숭이 바이러스 40 (SV-40) 벡터, 소의 파필로마 벡터, Epstein-Barr 벡터, 헤르페스 벡터, 백시니아 벡터, Moloney 뮤린 백혈병 벡터, Harvey 뮤린 육종 바이러스 벡터, 뮤린 유방 종양 바이러스 벡터, Rous 육종 바이러스 벡터 및 비-바이러스성 플라스미드. 곤충 세포에서 발현을 시키기 위해 벡트로바이러스 벡터가 적합할 수 있다. 비-바이러스성 벡터는 나노입자, 양이온성 지질, 양이온성 고분자, 금속성 나노고분자, 나노막대, 리포솜, 미셀, 마이크로버블, 세포-투과성 펩티드, 또는 지방구(liposphere)로 제형화될 수 있다.

[0105] 일부 구체예들에서, 상기 벡터는 자가-증폭성 RNA 레플리콘(자가-복제성 (m)RNA로 또한 불림), 자가-복제 (m)RNA, 자가-증폭성 (m)RNA, 또는 RNA 레플리콘이다. 상기 자가-증폭성 RNA 레플리콘은 자체 복제되는 RNA이다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘은 세포의 내부에서 자체 복제될 수 있다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘은 RNA 중합효소 및 관심 분자를 인코딩한다. RNA 중합효소는 RNA-의존성 RNA 중합효소 (RDRP 또는 RdRp)일 수 있다. 상기 자가-증폭성 RNA 레플리콘은 프로테아제 또는 RNA 캡핑(capping) 효소를 또한 인코딩할 수 있다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘 벡터는 알파바이러스로 알려진 토가비리데(Togaviridae) 패밀리의 바이러스이거나, 또는 이로부터 유래되는데, 여기에는 Eastern 말의 뇌염 바이러스 (EEE), Venezuelan 말의 뇌염 바이러스 (VEE), Everglades 바이러스, Mucambo 바이러스, Pixuna 바이러스, Western 말의 뇌염 바이러스 (WEE), Sindbis 바이러스, South African 아르보바이러스 번호 86, Semliki Forest 바이러스, Middelburg 바이러스, Chikungunya 바이러스, Onyong-nyong 바이러스, Ross River 바이러스, Barmah Forest 바이러스, Getah 바이러스, Sagiyama 바이러스, Bebaru 바이러스, Mayaro 바이러스, Una 바이러스, Aura 바이러스, Whataroa 바이러스, Babanki 바이러스, Kyzylagach 바이러스, Highlands J 바이러스, Fort Morgan 바이러스, Ndumu 바이러스, Buggy Creek 바이러스, 그리고 International Committee on Taxonomy of Viruses (ICTV)에서 알파바이러스로 분류된 기타 모든 바이러스가 내포될 수 있다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘은 알파바이러스의 감쇠된 형태, 이를 테면, VEE TC-83 백신 균주의 일부이거나 또는 이를 포함한다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘 벡터는 숙주 세포, 표적 세포, 또는 기관에서 특정 기능 (가령, 연장되거나 증가된 이분 면역수용체 발현)을 위해 시험관내, 생체내, 생체외, 또는 가상에서 공작되거나, 또는 선별되었다. 예를 들면, 상기 자가-증폭성 RNA 레플리콘의 상이한 변이체들을 품고 있는 숙주 세포 집단은 상이한 시점에서 관심대상의 하나 또는 그 이상의 분자 (자가-증폭성 RNA 레플리콘에 인코딩되거나, 또는 숙주 게놈에 인코딩된)의 발현 수준을 기반으로 선별될 수 있다. 일부 구체예들에서, 상기 선별된 또는 공작된 자가-증폭성 RNA 레플리콘을 변형시켜 숙주 세포 또는 유기체로부터 유형

I 인터페론 반응, 선천적 항바이러스 반응, 또는 조정된 면역 반응을 감소시켜, 해당 숙주 세포, 표적 세포, 또는 유기체에서 RNA 레플리콘의 단백질 발현을 더 지속시키거나, 또는 더 높은 수준으로 발현되도록 한다. 일부 구체예들에서, 이러한 최적화된 자가-증폭성 RNA 레플리콘 서열은 원하는 표현형 형질 (가령, 야생형 균주 또는 백신 균주와 비교하였을 때, 관심대상 분자의 더 높은 발현 또는 그 이상의 지속된 발현, 또는 벡터에 대항하는 선천적 항바이러스 감소된 면역 반응)을 갖는 개별 세포 또는 세포 집단으로부터 획득된다. 일부 구체예들에서, 원하는 또는 선별된 자가-증폭성 RNA 레플리콘 서열을 품고 있는 세포는 자가-증폭성 RNA 레플리콘을 포함하는 치료제로 대상체 (가령, 인간 또는 동물)을 처리한 후, 유익한 반응 특징 (가령, 엘리트 응답자 또는 완전한 차도를 보인 상태의 대상체)를 갖는 대상체로부터 얻었다. 일부 구체예들에서, 상기 자가-증폭성 RNA 레플리콘은 하나 또는 그 이상의 하위-게놈 폴리뉴클레오티드(들)을 만들기 위해, 하나 또는 그 이상의 하위-게놈 서열(들)을 함유할 수 있다. 일부 구체예들에서, 상기 하위-게놈 폴리뉴클레오티드는 세포의 해독 기전에 의한 해독을 위한 기능성 mRNA 분자로 작용한다. 자가-증폭성 RNA 레플리콘 상의 특정 서열 요소 (가령, 하위-게놈 프로모터 또는 SGP)는 중합효소가 하위-게놈 서열로부터 하위-게놈 폴리뉴클레오티드를 만들도록 지시하는 기능을 통하여 하위-게놈 폴리뉴클레오티드가 생성될 수 있다. 일부 구체예들에서, 상기 SGP는 RNA-의존적 RNA 중합효소 (RDRP 또는 RdRp)에 의해 인지된다. 일부 구체예들에서, 단일 자가-증폭성 RNA 레플리콘 상에 다중 SGP 서열이 존재하며, 이분(bipartite) 면역수용체, 이분(bipartite) 면역수용체의 구성성분, 또는 추가 제제를 인코딩하는 하위-게놈 서열의 상류에 위치할 수 있다. 일부 구체예들에서, SGP 서열의 뉴클레오티드 길이 또는 조성물이 변형되어, 하위-게놈 폴리뉴클레오티드의 발현 특징이 변경될 수 있다. 일부 구체예들에서, 비-동일한 SGP 서열은 자가-증폭성 RNA 레플리콘 상에 위치하여, 대응하는 하위-게놈 폴리뉴클레오티드의 비율은 해당 SGP 서열이 동일한 경우와는 상이하다. 일부 구체예들에서, 비-동일한 SGP 서열은 TCR 및 추가 작용제 (가령, 사이토킨)의 생산을 지시하여, TCR 발현의 증가, 표적 세포 또는 숙주에 대한 세포독성 효과 없이, 표적 세포의 증가 또는 더 빠른 확장을 유도하거나, 또는 RNA 레플리콘에 대한 선천성 또는 적응 면역 반응을 약화를 유도하는 서로에 대한 비율로 생성된다. 일부 구체예들에서, 서로에 대한 하위-게놈 서열 및 SGP 서열 그리고 게놈 서열 자체를 이용하여 서로에 대해 하위-게놈 폴리뉴클레오티드의 비율을 변경시킬 수 있다. 일부 구체예들에서, TCR을 인코딩하는 SGP 및 하위-게놈 서열은 TCR의 발현이 추가 작용제에 비해 실질적으로 증가되도록, 추가 작용제를 인코딩하는 SGP 및 하위-게놈 영역의 하류에 위치할 수 있다. 일부 구체예들에서, RNA 레플리콘 또는 SGP는 최적량의 사이토킨이 발현되도록 선별되거나 또는 공작되어, 해당 사이토킨은 T 세포의 확장을 촉진시키거나, TCR의 치료효과를 증가시키지만, 그러나 심각한 부작용, 이를 테면 사이토킨 방출 증후군, 사이토킨 폭풍, 또는 신경학적 독성을 일으키지 않는다.

[0106] 이들 두 쇠의 발현은 두 개 프로모터 또는 하나의 프로모터에 의해 구동될 수 있다. 일부 경우들에서, 두 개 프로모터가 이용된다. 일부 경우들에서, 두 개 프로모터는 이들 두 쇠에 대한 각 단백질-코딩 서열과 함께 머리에 서-머리로, 머리에서-꼬리로, 또는 꼬리에서-꼬리 방향으로 배열될 수 있다. 일부 경우들에서, 한 개 프로모터가 이용된다. 두 개의 단백질-코딩 서열을 프레임 안에서 연계시켜, 하나의 프로모터를 이용하여 두 쇠를 모두 발현시킬 수 있다. 그리고 이와 같은 경우, 두 개 단백질-코딩 서열은 머리에서-꼬리 방향으로 정렬되며, 리보솜 결합 부위 (가령, 내부 리보솜 결합 부위 또는 IRES), 프로테아제 분열 부위, 또는 자가-공정화 분열 부위 (이를 테면, 2A 펩티드를 인코딩하는 서열)에 연결되어, 바이시스트론성(bicistronic) 발현을 용이하게 한다. 일부 경우들에서, 상기 두 쇠들을 펩티드 링커로 연계시켜, 이들 두 쇠가 단일-쇄 폴리펩티드로 발현될 수 있다. 각 발현된 쇠는 재배열된 V(D)J 유전자를 비롯한, 완전한 가변 도메인 서열을 함유할 수 있다. 각 발현된 쇠는 CDR1, CDR2, 및 CDR3을 비롯한, 완전한 가변 도메인 서열을 함유할 수 있다. 각 발현된 쇠는 FR1, CDR1, FR2, CDR2, FR3, 및 CDR3을 비롯한, 완전한 가변 도메인 서열을 함유할 수 있다. 일부 경우들에서, 각 발현된 쇠는 불변 도메인 서열을 더 함유할 수 있다.

[0107] 발현 벡터를 창출하기 위해, 해당 구축된 핵산 분자에 추가 서열이 부가될 수 있다. 이들 추가 서열에는 벡터 골격 (가령, 표적 세포 또는 일시적 숙주, 이를 테면, 대장균에서 해당 벡터의 복제에 필요한 요소들), 프로모터, IRES, 자가-절단 펩티드를 인코딩하는 서열, 종료자, 부속 유전자들 (이를 테면, 페이로드), 뿐만 아니라 면역수용체 폴리뉴클레오티드의 부분 서열 (이를 테면, 불변 도메인을 인코딩하는 서열의 일부분)이 내포된다.

[0108] 프로테아제 분열 부위에는 다음이 내포되나, 이에 국한되지 않는다; 엔테로키나제 분열 부위: (Asp)4Lys (서열 식별 번호:288); 인자 Xa 분열 부위: Ile-Glu-Gly-Arg(서열 식별 번호:289); 트롬빈 분열 부위, 가령, Leu-Val-Pro-Arg-Gly-Ser(서열 식별 번호:290); 레닌 분열 부위, 가령, His-Pro-Phe-His-Leu-Val-Ile-His(서열 식별 번호:291); 콜라게나제 분열 부위, 가령, X-Gly-Pro (이때 X는 임의의 아미노산임); 트립신 분열 부위, 가령, Arg-Lys; 바이러스 프로테아제 분열 부위, 이를 테면, 피코르나바이러스의 프로테아제 2A 분열 부위, 간염 A 바이러스 3C 분열 부위, 인간 리노바이러스 2A 프로테아제 분열 부위, 피코르나바이러스 3 프로테아제 분

열 부위를 비롯한, 그러나 이에 국한되지 않는 바이러스 2A 또는 3C 프로테아제 분열 부위; 그리고 카스파제 프로테아제 분열 부위, 가령, 활성화된 카스파제-3에 의해 인지되고, 분열된 DEVD (서열 식별 번호:292)(이때 분열은 제 2 아스파르트산 다음에 일어난다). 일부 구체예들에서, 본 명세서는 프로테아제 분열 부위를 포함하는 발현 벡터를 제공하며, 이때 프로테아제 분열 부위는 세포의 프로테아제 분열 부위 또는 바이러스 프로테아제 분열 부위를 포함한다. 일부 구체예들에서, 제 1 단백질 분열 부위는 퓨린에 의해 인지되는 부위; IPNV의 VP4; 담배 에치(etch) 바이러스 (TEV) 프로테아제; 리노바이러스의 3C 프로테아제; PC5/6 프로테아제; PACE 프로테아제, LPC/PC7 프로테아제; 엔테로키나제; Factor Xa 프로테아제; 트롬빈; 제네나제 I; MMP 프로테아제; 순무 (turnip) 모자이크 포티바이러스의 핵 함포 단백질 a(N1a); Dengue 유형 4 플라비바이러스의 NS2B/NS3, 황열 바이러스의 NS3 프로테아제; 콜리플라워 모자이크 바이러스의 ORF V; KEX2 프로테아제; CB2; 또는 2A를 포함한다. 일부 구체예들에서, 상기 단백질 분열 부위는 바이러스 내부적으로 분열가능한 신호 펩티드 분열 부위이다. 일부 구체예들에서, 상기 바이러스 내부적으로 분열가능한 신호 펩티드 분열 부위는 C형 인플루엔자 바이러스, C형 간염 바이러스, 한타바이러스, 플라비바이러스, 또는 루벨라 바이러스로부터 유래된 부위를 포함한다.

[0109] 본 명세서의 벡터 안에 내포되는 적합한 IRES 요소는 진핵성 리보솜과 관계를 맺을 수 있는 RNA 서열을 포함할 수 있다. 일부 구체예들에서, 본 명세서의 IRES 요소는 적어도 약 250개 염기쌍, 적어도 약 350개 염기쌍, 또는 적어도 약 500개의 염기쌍이다. 본 명세서의 IRES 요소는 바이러스, 포유류, 및 초파리를 비롯한, 그러나 이에 국한되지 않은 유기체의 DNA로부터 유래될 수 있다. 일부 경우들에서, IRES 요소가 유래될 바이러스 DNA에는 피코르나바이러스 상보적 DNA (cDNA), 뇌척수심근염 바이러스 (EMCV) cDNA 및 폴리오바이러스 cDNA가 내포되나, 이에 국한되지 않는다. IRES 요소가 유래될 포유류 DNA의 예시에는 면역글로불린 중쇄 결합 단백질 (BiP)를 인코딩하는 DNA 및 염기성 섬유모세포 성장인자 (bFGF)를 인코딩하는 DNA가 내포되나, 이에 국한되지 않는다. IRES 요소가 유래될 초파리 DNA의 예시에는 드로소필라 멜라노가스터(*Drosophila melanogaster*)로부터 유래된 안테나페디아(*Antennapedia*) 유전자가 내포되나, 이에 국한되지 않는다. 폴리오바이러스 IRES 요소의 추가 예시에는 예를 들면, 폴리오바이러스 IRES, 뇌척수심근염 바이러스 IRES, 또는 A형 간염 바이러스 IRES가 내포된다. 플라비바이러스 IRES 요소의 예시에는 C형 간염 바이러스 IRES, GB 바이러스 B IRES, 또는 소의 바이러스 설사 바이러스 IRES 또는 전통적인 돼지 열 바이러스 IRES를 비롯한, 그러나 이에 국한되지 않은 페스티바이러스 IRES가 내포된다.

[0110] 자가-공정 분열 부위의 예시에는 intein 서열; 변형된 intein; hedgehog 서열; 다른 hog-패밀리 서열; 2A 서열, 가령, 구제역(Foot and Mouth) 질환 바이러스 (FMDV)에서 유래된 2A 서열; 그리고 이들 각각의 변이가 내포되나, 이에 국한되지 않는다.

[0111] 재조합 면역글로불린 또는 다른 단백질 발현을 위한 벡터에는 임의의 갯수의 프로모터가 내포될 수 있으며, 이때 상기 프로모터는 구성적, 조절가능한 또는 유도성, 세포-유형 특이적, 조직-특이적, 또는 종(species)-특이적이다. 추가 예시에는 테트라사이클린-반응성 프로모터가 내포된다. 상기 벡터는 재조합적으로 구축된 유전자가 발현될 숙주 세포에 맞게 개작된 레플리콘일 수 있고, 이 벡터는 박테리아 세포, 예를 들어, 대장균에서 기능하는 레플리콘을 포함할 수 있다. 상기 프로모터는 구성적이거나 또는 유도적일 수 있으며, 이때 유도는 예를 들어, 특정 세포 유형 또는 특정 성숙 수준과 연관된다. 대안적으로, 다수의 바이러스 프로모터가 적합할 수 있다. 프로모터의 예시에는 β -액틴 프로모터, SV40 초기(early) 및 후기(late) 프로모터, 면역글로불린 프로모터, 인간 사이토메갈로바이러스 프로모터, 레트로바이러스 프로모터, 신장 인자 1A (EF-1A) 프로모터, 포스포글리세레이트 키나제 (PGK) 프로모터, 및 Friend 비장 병소-형성 바이러스 프로모터가 내포된다. 상기 프로모터는 인헨서와 연합되거나, 또는 연합되지 않을 수 있으며, 이때 인헨서는 자연적으로 특정 프로모터와 연합되거나, 또는 상이한 프로모터와 연합될 수 있다.

[0112] **적용**

[0113] 본원에 기술된 조성물들 및 방법들을 다양한 응용에 적용될 수 있다. 예시적인 적용은 시퀀싱 데이터, 가령, 단일 세포 시퀀싱 데이터로부터 태생적으로 쌍을 이룬 TCRs를 인코딩하는 서열을 재-구축하는 것일 수 있다. 일부 적용에서, 종양-침윤성 T 세포들에서 식별된 태생적으로 쌍을 이룬 TCRs를 인코딩하는 서열의 재-구축을 원할 수 있다. 이들 적용에서, 대상체로부터 새로운 조직 샘플 (가령, 새로운 고형 종양 샘플)을 이용하여 태생적으로 쌍을 이룬 TCRs의 두 TCR 쌍 모두의 서열 정보를 얻기 위한 단일 세포 시퀀싱할 수 있다. 그러나, 종양-침윤성 세포들을 함유하는 조직 샘플 (가령, 체액 샘플이 아닌 고형 물질 샘플)이 냉동된 샘플 또는 고정된 샘플 (가령, FFPE 샘플)인 경우, 단일 세포 현탁액을 얻기 위해 세포를 분리하는 것이 극복과제가 될 수 있다. 이런 경우들에서, 상기 동일한 대상체로부터 말초 T 세포들을 함유하는 혈액 샘플을 이용하여 태생적으로 쌍을 이룬 TCRs의 서열 정보를 얻기 위한 단일 세포 시퀀싱할 수 있다. 상기 혈액 샘플은 해당 조직 샘플로부터 혈

류로 방출된 종양-침윤성 T 세포들을 함유할 수 있기 때문에, 해당 혈액 샘플에서 획득된 서열은 이들 종양-침윤성 T 세포의 서열을 함유할 수 있다. 그 다음, 동일한 대상체의 조직 샘플을 벌크 시퀀싱에 이용할 수 있다. 해당 조직 샘플의 벌크 시퀀싱이 태생적으로 쌍을 이룬 TCRs의 쌍을 이룬 서열을 제공하지 않을 수 있지만, 개별 TCR 쌍들에 대한 CDR3 서열을 제공할 수 있다. 해당 조직 샘플의 벌크 시퀀싱으로부터 획득된 CDR3 서열(본원에서 "조직 CDR3 서열"로 지칭됨)을 이용하여 혈액 샘플의 단일 세포 시퀀싱에서 획득한 쌍을 이룬 서열과 함께 정렬할 수 있다. 상기 쌍을 이룬 서열의 CDR3 서열이 조직 CDR3 서열과 일치되는 경우, 해당 쌍을 이룬 서열이 식별되고, 임의의 하류-응용 부분에 이용될 수 있다.

[0114] 단일 세포 시퀀싱이란 개별 세포들로부터 서열 정보를 얻는 것을 지칭한다. 단일 세포 시퀀싱에서, 세포 집단이 단일 세포 현탁액으로 만들어지고, 개별 부분으로 구획화될 수 있다. 각 구획 내에, 단일 세포로부터 방출된 서열을 바코드화시키고, 나중에 서열화시킬 수 있다. 다양한 단일 세포 시퀀싱 방법들을 TCR 재구성에 이용할 수 있다 (De Simone M, Rossetti G and Pagani M (2018) Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. Front. Immunol. 9:1638 참고). 벌크 시퀀싱이란 세포 집단으로부터 서열 정보를 얻는 것을 지칭한다. 벌크 시퀀싱에서, 세포 혼합물로부터 핵산 분자를 단리시키고, 함께 시퀀싱할 수 있다.

[0115] 도 9A는 종양 샘플 안에 종양-침윤성 TCRs를 식별해내기 위해 혈액 샘플을 이용한 예시적인 작업흐름을 보여준다. 우선, 환자로부터 혈액 샘플을 빼낼 수 있다. 그 다음, 말초 혈액 단핵 세포들을 함유하는 PBMC 샘플을 해당 혈액 샘플로부터 단리시킬 수 있다. 예를 들면, 이 세포는 혈액 층을 분리하는 친수성 다당류인 피콜(ficoll)과 구배 원심분리를 사용하여, 전혈로부터 추출할 수 있다. 그 다음, PBMC 샘플로부터 T 세포를 단리시킬 수 있다. T 세포는 예를 들어 PERCOLL™ 구배를 통한 원심분리 또는 역류 원심 분리에 의해 적혈구를 용해하고, 단핵구를 고갈시킴으로써, PBMC로부터 단리해낼 수 있다. 임의선택적으로, T 세포의 하위집단은 마커-기반 분류에 의해 추가로 농축될 수 있다. 상기 마커는 세포 표면 마커일 수 있다. 세포 표면 마커의 예에는 CD39, CD69, CD103, CD25, PD-1, TIM-3, OX-40, 4-1BB, CD137, CD3, CD28, CD4, CD8, CD45RA, CD45RO, GITR 및 FoxP3이 내포되나, 이에 국한되지 않는다. 상기 마커는 사이토킨일 수 있다. 사이토킨 마커의 예에는 IFN- γ , TNF-알파, IL-17A, IL-2, IL-3, IL-4, GM-CSF, IL-10, IL-13, 그랜자임 B 및 퍼포린이 내포되나, 이에 국한되지 않는다. 상기 T 세포 또는 T 세포들의 하위집단은 그 다음 태생적으로 쌍을 이룬 TCRs의 획득된 쌍을 이룬 서열(가령, 도 9A에서 정보적으로 쌍을 이룬 TCR 서열)에 대해 단일 세포 시퀀싱을 거칠 수 있다. 상기 동일한 환자로부터 또한 종양 샘플을 획득할 수 있다. 상기 종양 샘플은 고정된 샘플일 수도 있고, 또는 냉동된 샘플일 수도 있다. 예를 들면, 상기 종양 샘플은 고정화 제제, 이를 테면, 포르말데히드에 의해 고정될 수 있다. 상기 종양 샘플은 포르말린-고정된 파라핀-매립된 (FFPE) 조직 샘플일 수 있다. 그 다음, 상기 종양 샘플은 TCR 쌍들의 CDR3를 획득하기 위해 벌크 시퀀싱을 받을 수 있다. 그 다음, 종양 샘플로부터 획득된 CDR3 서열을 이용하여 쌍을 이룬 서열의 CDR3 서열과 비교하여, 종양-침윤성 TCRs를 식별해낼 수 있다. 상기 종양-침윤성 TCRs는 정상적인 T 세포들 또는 세포 계통에서 발현될 수 있고, 이들은 도 9A에서 "가상(virtual) TILs"로 나타낸다.

[0116] 도 9B는 가상 TILs의 예시적인 적용을 나타낸다. 상기 가상 TILs는 리포터 시스템을 포함할 수 있고, 이것은 표적-반응성 TCRs에 대한 리포터-기반된 T 세포 선별에 이용될 수 있다. 예를 들면, 상기 가상 TILs는 리포터 유전자를 포함하는 리포터 세포일 수 있고, 이의 리포터 유전자는 세포의 TCR이 표적 항원에 결합할 때 신호를 보내도록 조절된다. 이들 가상 TILs는 항원-로딩된 항원-제시 세포들 (APCs) 또는 인위적인 APCs과 접촉함으로써 활성화될 수 있다. 그 다음, 상기 리포터 시스템 또는 다른 선별 기전 (가령, 세포 표면 마커 또는 사이토킨 마커)에 의해 생성된 신호를 기반으로 하여, 예를 들면, FACS에 의해 표적-반응성 T 세포들을 선별해낼 수 있다. 상기 선별은 해당 세포들을 MHC-결합된 항원에 접촉한 후, 가상 TILs에서 세포 표면 마커 발현에 기반될 수 있다. 상기 세포 표면 마커는 CD25, CD69, CD39, CD103, CD137, 뿐만 아니라 다른 T 세포 활성화 마커, 또는 임의의 이의 조합일 수 있다. 상기 선별은 칼슘 유입을 기반으로 할 수 있다. 상기 선별은 또한 리포터 유전자 발현을 기반으로 할 수 있다. 상기 리포터 유전자는 형광 단백질 (이를 테면, GFP 및 mCherry)일 수 있다. 상기 리포터 유전자는 TCR 신호전달에 의해 조절되는 전사 인자의 제어 하에 있을 수 있다. 이들 전사 인자의 예에는 AP-1, NFAT, NF- κ B, Runx1, Runx3, 등이 내포되나, 이에 국한되지 않는다. 상기 선별은 이를 테면, ICS 및 사이토킨 포획 분석 방법을 이용하여 활성화된 가상 TILs로부터 방출된 사이토킨에 기반을 둘 수 있다. 도 9C는 가상 TILs의 또다른 응용 적용을 보여준다. 표적-반응성 TCRs를 식별한 후, 상기 표적-반응성 TCRs를 숙주 세포, 이를 테면 자가유래 T 세포들 (해당 조직 샘플 및 혈액 샘플이 획득된, 동일한 환자로부터 단리된 T 세포)로 운반되고, 발현될 수 있다. 상기 표적-반응성 TCRs는 동종이계 T 세포로 운반되고, 발현될 수 있다. 그 다음, 상기 표적-반응성 TCRs를 발현시키는 T 세포를 질환 이를 테면, 암을 치료하기 위해 동일한

환자에게 투여할 수 있다.

[0117] 대상체의 조직 샘플 (가령, 고형 샘플)에서 태생적으로 쌍을 이룬 T-세포 수용체 (TCR)의 서열을 식별해내는 방법은 다수의 말초 T 세포들을 함유하는 해당 대상체로부터 획득한 샘플 안에 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 하나 또는 그 이상의 쌍을 이룬 서열을 식별해내는 것을 포함할 수 있다. 상기 하나 또는 그 이상의 쌍을 이룬 서열 각각은 CDR3 서열을 포함할 수 있다. 그 다음, 해당 조직 샘플 안에 TCR의 TCR쇄의 조직 CDR3 서열을 식별해내고, 태생적으로 쌍을 이룬 다른 TCR쇄는 알려지지 않을 수 있다. 상기 조직 CDR3 서열은 상기 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 상기 하나 또는 그 이상의 쌍을 이룬 서열의 적어도 하나의 쌍을 이룬 서열의 CDR3 서열과 합치될 수 있고, 그렇게 함으로써 해당 조직 샘플에서 태생적으로 쌍을 이룬 TCR의 서열로써 상기 적어도 하나의 쌍을 이룬 서열을 식별해낼 수 있다. 표적-반응성 T-세포 수용체 (TCR)을 식별해내는 방법이 또한 본원에서 제공된다. 상기 방법은 본원에 기술된 방법들을 이용하여 식별된 TCR을 포함하는 세포의 제공을 포함할 수 있다. 그 다음, 상기 세포는 항원-제시 세포 (APC)에 의해 제시되는 표적 항원과 접촉될 수 있다. 상기 세포는 상기 TCR을 통하여 APC에 의해 제시되는 표적 항원에 결합할 수 있고, 그렇게 함으로써 상기 TCR은 표적-반응성 TCR로 식별될 수 있다.

[0118] 본원에서 기술된 APC는 전문적인 APC, 이를 테면, 수지상 세포, 대식세포, 또는 B 세포일 수 있다. 상기 APC는 단핵구, 또는 단핵구-유래된 수지상 세포일 수 있다. aAPC는 T 세포 수용체용 리간드 및 공동-자극 분자를 발현시킬 수 있고, T 세포를 활성화시키고, 전달을 위해 확장시킬 수 있으며, 일부 경우에서 이들의 잠재력 및 기능이 개선된다. aAPC는 T 세포 활성화를 위한 임의의 유전자를 발현시키도록 공작될 수 있다. aAPC는 T 세포 확장을 위한 임의의 유전자를 발현시키도록 공작될 수 있다. aAPC는 비드, 세포, 단백질, 항체, 사이토킨, 또는 임의의 조합일 수 있다. aAPC는 계능 이식을 겪는 세포 집단으로 신호를 전달할 수 있다. 예를 들면, aAPC는 신호 1, 신호 2, 신호 3 또는 임의의 조합을 전달할 수 있다. 신호 1은 항원 인지 신호일 수 있다. 예를 들면, 신호 1은 펩티드-MHC 복합체 또는 CD3을 지향하는 작용성 항체의 결합에 의한 TCR의 결합이며, 이것은 CD3 신호-형질도입 복합체의 활성화로 이어질 수 있다. 신호 2는 공동-자극성 신호일 수 있다. 예를 들면, 공동-자극성 신호는 항-CD28, 유도성 공동-자극자 (ICOS), CD27, 및 4-1BB (CD137)일 수 있다(이들은 차례로 ICOS-L, CD70, 및 4-1BBL에 결합한다). 신호 3은 사이토킨 신호일 수 있다. 사이토킨은 임의의 사이토킨일 수 있다. 사이토킨은 IL-2, IL-7, IL-12, IL-15, IL-21, 또는 임의의 이의 조합일 수 있다.

[0119] 일부 경우들에서, aAPC를 이용하여 세포 집단을 활성화시키고 및/또는 확장시킬 수 있다. 일부 경우들에서, 인공적인 것은 동종특이성(allospecificity)을 유도하지 못할 수 있다. 일부 경우들에서, aAPC는 HLA를 발현시키지 못할 수 있다. aAPC는 활성화 및/또는 자극에 이용될 수 있는 유전자들을 안정적으로 발현시키도록 유전적으로 변형될 수 있다. 일부 경우들에서, 활성화에 K562 세포가 이용될 수 있다. 확장에 K562 세포가 또한 이용될 수 있다. K562 세포는 인간 적백혈병 세포 계통일 수 있다. K562 세포는 관심대상의 유전자를 발현시키도록 공작될 수 있다. K562 세포들은 내인적으로(endogenously) HLA 클래스 I, II, 또는 CD1d 분자를 발현시키지 않을 수 있지만, 그러나 ICAM-1 (CD54) 및 LFA-3 (CD58)을 발현시킬 수 있다. K562를 공작하여 신호 1을 T 세포들에게 전달할 수 있다. 예를 들면, K562 세포들은 HLA 클래스 I를 발현시키도록 공작될 수 있다. 일부 경우들에서, K562 세포들은 추가 분자, 이를 테면 B7, CD80, CD83, CD86, CD32, CD64, 4-1BBL, 항-CD3, 항-CD3 mAb, 항-CD28, 항-CD28mAb, CD1d, 항-CD2, 막-결합된 IL-15, 막-결합된 IL-17, 막-결합된 IL-21, 막-결합된 IL-2, 절두된(truncated) CD19, 또는 임의의 조합을 발현시키도록 공작될 수 있다. 일부 경우들에서, 공작된 K562 세포는 CD80 및 CD83에 추가하여, 항-CD3 mAb의 막 형태, 클론 OKT3를 발현시킬 수 있다. 일부 경우들에서, 공작된 K562 세포는 CD80 및 CD83에 추가하여, 항-CD3 mAb의 막 형태, 클론 OKT3, 항-CD28 mAb의 막 형태를 발현시킬 수 있다.

[0120] **키트**

[0121] 본원에서 기술된 조성물은 키트로 제공될 수 있다. 예를 들면, 상기 키트는 폴리뉴클레오티드 분자 다수를 구축하는데 이용될 수 있는 핵산 분자 풀을 갖는 용기를 포함할 수 있고, 이때 각 폴리뉴클레오티드는 TCR쇄 또는 이의 일부분, 또는 TCR쇄의 동계 쌍을 인코딩한다. 일부 경우들에서, 핵산 분자 풀의 각 핵산 분자는 TCR쇄의 CDR3를 인코딩한다. 일부 경우들에서, 핵산 분자 풀의 각 핵산 분자는 TCR쇄의 동계 쌍의 제 1 CDR3 및 제 2 CDR3를 인코딩한다. 일부 경우들에서, 핵산 분자 풀의 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함한다. 일부 경우들에서, 핵산 분자 풀의 각 핵산 분자는 본원에서 기술된 바와 같이, 커넥터 서열을 포함한다. 상기 커넥터 서열은 핵산 분자의 동일한 풀 안에 다른 커넥터 서열과 상이한 서열을 보유할 수 있다. 상기 키트는 하나 또는 그 이상의 용기를 포함할 수 있고, 각 용기는 핵산 분자 풀을 함유한다. 상기 키트에서

제공되는 핵산 분자는 액체 형태 또는 건조된 형태 (가령, 동결건조된 형태)일 수 있다.

[0122] 상기 키트는 사용자가 TCRs을 인코딩하는 폴리뉴클레오타이드 분자 다수를 구축하는데 핵산 분자 풀의 사용에 관한 지침이 되는 지침 자료를 더 포함할 수 있다.

[0123] 상기 키트는 핵산 구축 반응에 이용될 수 있는 적어도 하나의 시약 (가령, 완충, 효소, 첨가제, 등)을 더 포함할 수 있다.

[0124] **실시예**

[0125] **실시예 1. CDR3-J 올리고뉴클레오타이드 풀을 전장의, 발현가능한 TCR 풀로 전환**

[0126] 이 실시예는 3가지 유형 IIS 제한절단 효소를 이용하여 스티커 단부를 생성시켰다. 이러한 효소들은 시판되고 있다. 이 실시예에서, 4-bp 5' 오버행 (예를 들면, BbsI, BbvI, BcoDI, BsaI, BsmBI, FokI, 등)을 생성시키는 두 개 효소, 그리고 블런트 단부 또는 3' 오버행 (예를 들면, BseRI, BsrDI, BtsI, MlyI, 등)을 생성시키는 한 개의 제한절단 효소가 이용된다. 사용할 최적의 효소 세트는 실용적인 요소 (가령, 국소 가용성, 절단 효율성, 스타 활성)에 따라 달라질 수 있으며, 실험적으로 쉽게 선택할 수 있다. 여기에서, 처음 두 개 제한절단 효소는 TIISRE1, TIISRE2로 지칭되며, 마지막 제한절단 효소는 TIISRE3이라고 한다.

[0127] 이 실시예에서, 쌍을 이룬 CDR3-J 올리고뉴클레오타이드는 알파 및 베타 CDR3-J의 코딩 서열에 대해 '머리에서-꼬리' 방향으로 합성된다. 환언하면, 알파 CDR-3J 및 베타 CDR-3J는 동일한 5'에서 3' 방향으로 합성된다. 결과적으로 생성된 전장의, 발현가능한 TCR 폴리뉴클레오타이드 또한 머리에서-꼬리 방향이다. 쌍을 이룬 CDR3-J 올리고뉴클레오타이드는 다른 방향, 예를 들면, 머리에서-머리 방향과 꼬리에서-꼬리 방향으로 합성될 수 있다. 본원에서 기술된 방법들은 U.S. 특허 출원 번호 62/718,227, 62/725,842, 62/732,898, 62/818,355 및 62/823,831 (이들 각각은 본원 참고자료에 전문이 편입됨)에서 기술된 방법과 조합되어, 쌍을 이룬 CDR-3J 올리고뉴클레오타이드를 기획하고, 다른 방향으로 된, 전장의, 발현가능한 TCR 폴리뉴클레오타이드를 획득할 수 있다.

[0128] **도 1A-1C**에 나타난 바와 같이, 쌍을 이룬 CDR3-J 올리고는 하기에서 기술된 다른 개재(intervening) 도메인을 갖는, TRBJ, CDR3베타, TRAJ, 및 CDR3알파의 역-보체 서열 (5'에서 3' 순서로)을 함유한다. 이 문서를 통하여, 기호 '*'는 상보성을 나타낸다. 예를 들면, 만약 P가 폴리뉴클레오타이드 서열을 지칭한다면, P*는 P의 역-보체를 지칭한다. 또한, 적합한 경우, 문자 X를 이용하여 A 또는 B를 지칭할 수 있다. 예를 들면, TRXV는 TRAV 및 TRBV를 집합적으로 지칭하는데 이용될 수 있다. 명확하게 하기 위해, 본 실시예 및 **도 1A-1C**에서, TRAJ 도메인 및 TRBJ 도메인은 차례로 TRAJ 영역 및 TRBJ 영역을 인코딩하는 폴리뉴클레오타이드 서열을 지칭하며, 이들은 CDR3에 내포되지 않는다.

[0129] BCC는 "베타 불변 커넥터(beta constant connector)"의 머릿글자 약어이며, 이의 기능은 TRBC 서열과 연결되는 것이다. ConB#는 특이적 TRBV 서열의 커넥터이며, 기호 #는 TRBV 유전자의 수치적 ID를 나타낸다. 유사하게, ConA#는 특이적 TRAV 서열의 커넥터이다. ICC는 "쇄-간 커넥터(inter-chain connector)"의 머릿글자 약어이며, 이를 이용하여 TRBV를 ConB#에 연결시키고, 뿐만 아니라 TRAC를 TRAJ에 연결시키는데 이용할 것이다.

[0130] ConB# 및 ConA# 도메인은 코돈-다각화되어 (**실시예 2** 참고), 상이한 TRBV 유전자들에 대한 ConX#는 뉴클레오타이드 수준에서 충분히 상이하고, ConX#와 ConX#*는 해당 ConX#와 ConX#*의 수치적인 ID가 동일한 경우에만 높은 수율로 혼성화될 수 있다.

[0131] 48개의 부분적으로 이중-가닥으로 된 TRAV#_GL 폴리뉴클레오타이드 (IMGT에서 각 TRAV 유전자에 대해 기능성으로 주석이 붙은 하나)의 라이브러리는 전통적인 방법들을 이용하여 준비할 수 있다. 모든 TRAV#_GL 폴리뉴클레오타이드를 혼합하여 TRAV#_GL 풀을 창출해낼 수 있다. GL은 생식계열(germline)을 나타내는 약어다. 각 TRAV#_GL 폴리뉴클레오타이드의 상부 가닥은 (1) P2A-3 도메인을 함유하며, 이는 자가-절단 P2A 펩티드의 3' 부분을 인코딩하고, (2) TRAV#_GL5 도메인, 이는 TRAV#의 생식계열 서열 (L, FR1, CDR1, FR2, CDR2이 내포됨), 그리고 ConA#의 FR3 상류 부분을 이러한 순서로 인코딩하며, 그리고 (3) FR3의 최종 스트레치를 인코딩하고, 코돈-다각화된 ConA#를 함유한다. 각 TRAV#_GL 폴리뉴클레오타이드의 하부 가닥은 TRAV#_GL5* 및 P2A-3*를 함유한다. 따라서, TRAV#_GL 폴리뉴클레오타이드는 서열 ConA#과 함께 3' 오버행을 갖는다. 48개 TRBV#_GL 폴리뉴클레오타이드의 라이브러리와 풀을 유사하게 준비할 수 있다. TRAV#_GL에서 P2A-3 도메인은 TRBV#_GL에서 T2A-3로 대체될 수 있다. T2A는 또다른 자가-절단 펩티드이다.

[0132] 1,000 ~ 500,000개의 쌍을 이룬 CDR3-J 올리고뉴클레오타이드 풀은 칩-기반 합성에 의해 준비될 수 있다.

- [0133] 단계 (1)에서, ConA#와 ConA## 간의 특이적 혼성화를 허용하는 온도에서 TRAV#_GL 풀을 쌍을 이룬 CDR3-J 풀과 혼합시킬 수 있다. 그 다음, 단계 (2)에서, DNA 중합효소를 이용하여 TRAV#_GL의 상부 가닥을 연장시킬 수 있고, 리가제를 이용하여 쌍을 이룬 CDR3-J 올리고와 TRAV#_GL의 하부 가닥을 결합시킬 수 있다.
- [0134] BCC는 TIISRE1의 인지 부위를 함유한다. 단계 (3)에서, TIISRE1을 이용하여 BCC를 절단하고, 하부 가닥에 4-염기 5' 오버행만 남긴다. 이 실시예에서, 상기 4 염기는 TRBC1의 첫 4개 염기의 안티센스다. 단계 (4)에서, 이 분열 산물을 사전-준비된 TRBC_P2A-5_SE (이것은 온전체 TRBC1 서열과 P2A-5 도메인을 함유하고, 이 TRBC1 서열의 시작 지점에 4-염기로 된 5' 오버행을 갖음)에 결합시킬 수 있다. 상기 P2A-5 도메인은 P2A 코딩 서열의 5' 단부 부분이다. SE는 스틱커 단부(sticky end)를 나타내는 약어다. 이러한 결합 생성은 단계 (5)에서 PCR-증폭된 것일 수 있다.
- [0135] 단계 (6)에서, 이러한 증폭 산물을 U.S. 특허 출원 번호 62/718,227, 62/725,842, 62/732,898, 62/818,355 및 62/823,831에서 기술된 방법을 이용하여 P2A-5와 P2A-3 사이를 결합시킴으로써, 원형화시킬 수 있다. 결합 후, P2A-5와 P2A-3은 P2A를 형성한다. 이 실시예에서, 상기 ICC는 TIISRE3의 인지 부위를 함유하며, 단계 (7)에서 하부 가닥 상의 ConB##의 바로 3'를 제거하는데 이용할 수 있다. 상부 가닥 상의 절단 부위의 중요성은 적다. 단계 (8)에서, 이러한 분열 산물은 상부 가닥과 하부 가닥이 분리될 때까지 가열시킬 수 있다. TRBC1의 처음 염기 ~ 20개 염기를 함유하는 프라이머를 이용하여 하부 가닥을 연장시켜, 하부 가닥의 3' 단부 지점에서 단일-가닥으로 된 영역을 남겨 둔다. 이 가닥의 3' 단부 끝은 ConB## 도메인이다. 단계 (9)에서, TRBV#_GL 풀을 추가하여 TRBV#_GL의 상부 가닥의 ConB#가 대응하는 ConB##과 혼성화될 수 있다. DNA 중합효소와 리가제를 추가하여 이 혼성화 산물을 완전히 이중-가닥으로 된 DNA로 전환시킬 수 있다.
- [0136] ICC의 나머지 부분 또한 TIISRE2의 인지 부위를 함유하고, 단계 (10)에서 이를 이용하여 ICC를 제거하여, TRAC의 첫 4개 염기의 안티센스 서열인 4-염기 5' 오버행이 남아있게 된다. 단계 (11)에서, 사전-준비된 TRAC_SE를 상기 5' 오버행에 결합시켜, 완전체 TRAC 서열을 형성시킬 수 있는데, 이는 상기 기술된 단계 (4)와 유사하다.
- [0137] 최종 산물을 CRISPR/TALEN/ZFN-기반의 녹-인(knock-in)에 이용된 적절한 "상동" 서열 또는 렌티바이러스 골격에 결합시킬 수 있다.
- [0138] **실시예 2. 인간 TRAV 서열 및 TRBV 서열을 이용한 코돈 다양화 테스트**
- [0139] 이 실시예에서 코돈-다각화된 ConA# 서열과 ConB# 서열을 기획하기 위해 열역학-기반 알고리즘이 제공된다. 이 알고리즘은 MATLAB 언어로 기재되어 있다. 이 알고리즘에 사용된 일부 변수 및 사용자 정의 기능은 아래의 '주석' 섹션에 설명되어 있으며, 나머지는 코드에 대한 코멘트에 설명되어 있거나, 또는 숙련자들에게 자명하다. 일부 사용자 정의 기능은 공개적으로 사용 가능한 열역학 매개변수(예: 염기쌍 스택의 경우 ΔH 및 ΔS) 및 모델(가령, 루프 크기의 함수로서의 ΔS)을 사용하는 DNA 혼성화의 열역학-기반 시뮬레이션에 의존한다. 이러한 매개변수와 모델은 John SantaLucia Jr에 의해 광범위하게 게시되었다. 숙련된 당업자는 처음부터 또는 NUPACK과 같이 공개적으로 사용 가능한 소프트웨어 패키지의 도움으로 이러한 기능을 쉽게 작성할 수 있다. 이 알고리즘은 초기 설계와 코돈 다양화의 두 단계가 함유하며, 이는 각각 Script1 및 Script2에 설명되어 있다. 초기 기획에서 ConA# 서열과 ConB# 서열은 원래의 TRAV 서열 또는 TRBV 서열에 따라 기획된다. 모든 ConX#의 모든 ConX##에 대한 혼성화 수율은 기선 역할을 하도록 산출된다 (도 2A 및 도 2B). 도 2A는 코돈 다양화 없이, 원래의 TRAV 서열에 따라 기획된 커넥터 서열의 혼성화 수율을 보여준다 (ConA# 에서 ConA##로). 도 2B는 코돈 다양화 없이, 원래의 TRBV 서열에 따라 기획된 커넥터 서열의 혼성화 수율을 보여준다 (ConB# 에서 ConB##로). 코돈 다양화 동안, 일부 TRXV#_GL의 마지막 ~60개 염기에 대한 코돈 선택이 무작위화되고, 특정 혼성화를 허용하는 ConX# 서열이 선택된다. 그 다음, 코돈-다각화된 서열 세트를 사용하여, 모든 ConX##에 대한 모든 ConX#의 혼성화 수율을 계산하여, 코돈 다양화가 성공적인지를 확인한다 (도 3A 및 도 3B). 도 3A는 코돈-다각화된 커넥터 서열의 혼성화 수율 (ConA#에서 ConA##로)을 보여준다. 도 3B는 코돈-다각화된 커넥터 서열의 의 혼성화 수율 (ConB#에서 ConB##로)을 보여준다.
- [0140] **Script1: 최초 기획**
- [0141] clear
- [0142] fHybTemp = 60; %혼성화 온도 (단위: 섭씨 C)
- [0143] fConcNa = 125; %나트륨 이온 농도 (단위: mM)
- [0144] fConcMg = 3; % 마그네슘 이온 농도 (단위: mM)

```

[0145] fConcQB = 5; % ConA# 또는 ConB#의 농도
[0146] %struPara에서 상기 매개변수들의 저장.
[0147] struPara.fHybTemp = fHybTemp;
[0148] struPara.fConcNa = fConcNa;
[0149] struPara.fConcMg = fConcMg;
[0150] struPara.fConcQB = fConcQB;
[0151] cChain = 'B'; % 값은 이들은 차례로 초기 ConA 서열과 ConB 서열을 기획하기 위한 A 또는 B일 수 있다.
[0152] cFileGeneSeq = sprintf('hsTR%sV_UTR200-L-V_Sorted_FOnly.txt',cChain); % 주석 참고
[0153] % V 유전자 서열들의 판독
[0154] fidGene = fopen(cFileGeneSeq);
[0155] raGeneInfo = textscan(fidGene,'%s\t%d\t%d\t%s','Headerlines',1);
[0156] fclose(fidGene);
[0157] %최초로 기획된 ConA 서열 또는 ConB 서열을 보관하는 셀 어레이 (ralon1)를 초기화함
[0158] iTotalNumOfGene = size(raGeneInfo{1},1);
[0159] ralon1s{1} = -ones(iTotalNumOfGene,2);
[0160] ralon1s{2} = cell(iTotalNumOfGene,4);
[0161] for iGeneNum = 1:iTotalNumOfGene
[0162] fprintf('Designing lon1 for gene #%u.\n',iGeneNum);
[0163] cGeneName = raGeneInfo{1}{iGeneNum};
[0164] cGeneSeq = raGeneInfo{4}{iGeneNum};
[0165] iLStart = raGeneInfo{2}(iGeneNum);
[0166] cCDS = cGeneSeq(iLStart:end);
[0167] cAA = nt2aa(cCDS);
[0168] iPosAAConservedC = find(cAA=='C',1,'last');
[0169] viPosNTConservedC = iPosAAConservedC*3-2:iPosAAConservedC*3;
[0170] vPosOfTinCys(iGeneNum) = viPosNTConservedC(1); % CDR3의 N 말단에서 보존된 Cys에 대한 코돈의 첫 뉴클레오타이드 위치
[0171] cSA60 = cCDS(viPosNTConservedC(1)-59:viPosNTConservedC(1));
[0172] ralon1ofThisGene = fun_Designlon1(cSA60,struPara); % 주석 참고
[0173] disp(ralon1ofThisGene{1})
[0174] ralon1s{1}(iGeneNum,:) = ralon1ofThisGene{1};
[0175] ralon1s{2}(iGeneNum,:) = ralon1ofThisGene{2};
[0176] end
[0177] %%
[0178] cTime = datestr(now);
[0179] cTime(cTime==' ')='_';

```

```

[0180] cTime(cTime==':')='_';
[0181] save(['IniDesignlonl_',cChain,num2str(fHybTemp),'_',cTime,'.mat'],'vPosOfTinCys','cChain','raGeneInfo','ralonls','struPara');
[0182] %% 교차 혼성화 수율 계산
[0183] xFracBoundHyb_THyb = -ones(iTotalNumOfGene,iTotalNumOfGene);
[0184] for iSimQB = 1:iTotalNumOfGene
[0185]     for iDE = 1:iTotalNumOfGene
[0186]         if xFracBoundHyb_THyb(iSimQB,iDE) >= 0
[0187]             continue;
[0188]         end
[0189]         cSimQB = ralonls{2}{iSimQB,3}; % cSimQB = ConA/B의 서열
[0190]         cDE = ralonls{2}{iDE,4}; % cDE = ConA/B*의 서열
[0191]         fThisFracBound_THyb = NP_GetBoundFrac(cSimQB,fConcQB,cDE,fConcQB/100,...
[0192]             fHybTemp,'Na',fConcNa,'Mg',fConcMg); % 주석 참고
[0193]         xFracBoundHyb_THyb(iSimQB,iDE) = fThisFracBound_THyb;
[0194]     end
[0195] end
[0196] %%
[0197] save('IniDesign.mat','xFracBoundHyb_THyb','xFracBoundHyb_THybMinus50C');
[0198] figure;
[0199] colormap(gray);
[0200] imagesc(1-xFracBoundHyb_THyb);
[0201] Script1에 대한 주석:
[0202] 파일 "hsTRAV_UTR200-L-V_Sorted_FOnly.txt" 및 "hsTRBV_UTR200-L-V_Sorted_FOnly.txt"는 IMGT 데이터베이스에서 '기능성'으로 주석이 달란 모든 TCR V 유전자들을 기록한 TSV 파일이다. 각 파일에는 4개의 열이 있다: 제 1 열은 V 유전자의 이름이고, 제 4 열은 시작 코돈(L-PART1의) 상류 ~200 nt에서 시작하는 V 유전자 cDNA 시퀀싱 서열이고, 제 2 열은 시작 코돈의 첫 번째 뉴클레오티드의 위치다. 제 3 컬럼은 V 유전자 (가령, L-PART2 다음)의 첫 뉴클레오티드 위치다.
[0203] "fun_Designlonl" 함수는 두 개의 입력을 사용하여 ConA 또는 ConB 서열을 되돌린다. (1) TRXV#_GL의 마지막 60개 염기를 기록하는 변수 cSA60 및 (2) 변수 struPara에 저장된 열역학적 모델링을 위한 매개변수. 간단히 말해서, 함수는 다음 명령문을 만족하는 cSA60(ConX로 표시됨)의 3' 단부에서 끝나는 cSA60의 가장 짧은 연속 하위서열을 찾는다: 서열 ConX를 갖는 제 1 DNA 올리고뉴클레오티드 5 nM와 서열 ConX*를 갖는 제 2 DNA 올리고뉴클레오티드 0.05 nM가 혼합될 때, struPara.fHybTemp, struPara.fConcNa 및 struPara.fConcMg에 의해 각각 특정된 온도, 나트륨 이온 농도 및 마그네슘 이온 농도에서 제 2 올리고뉴클레오티드의 97% 이상이 제 1 올리고뉴클레오티드에 결합될 것으로 예측된다. 이 함수(ralonlofThisGene)의 출력은 두 개 쉼을 갖는 셀 어레이이며, 여기에서 제 1 셀 ralonlofThisGene{1}은 1x2 벡터, 여기에서 ralonlofThisGene{1}(1)은 본 실시예에서 이용되지 않는 출력이며, ralonlofThisGene{1}(2)는 cSA60 상에서 ConX의 제 1 염기 위치다. ralonlofThisGene{2}는 1x4 셀 어레이이며, 여기에서 ralonlofThisGene{2}{1} 및 ralonlofThisGene{2}{2}은 본 실시예에서 이용되지 않으며, ralonlofThisGene{2}{3}은 ConX의 서열이고, ralonlofThisGene{2}{4}은 ConX*의 서열이다. 당업자는 상기에서 설명한 대로 이 함수를 작성할 수 있다.
[0204] 함수 "NP_GetBoundFrac"는 5 nM (fConcQB에 의해 기록됨)의 제 2 뉴클레오티드와 0.05 nM (fConcQB/100에 의해

```

기록됨)의 제 1 뉴클레오티드가 60°C (fHybTemp에 의해 기록됨)에서 혼합되고, 125 mM Na⁺ (fConcNa에 의해 기록됨) 및 5 mM Mg⁺⁺ (fConcMg에 의해 기록됨)에서 평형을 이룰 때, 서열 ConX을 갖는 제 2 DNA 올리고뉴클레오티드에 결합된 서열 ConX*을 갖는 제 1 DNA 올리고뉴클레오티드 분해를 되돌리한다.

[0205] 이 스크립트에 의해 생성된 이미지는 위에서 설명한 조건에서 0.05nM의 ConX#*가 5nM의 ConX#와 혼합될 때, ConX#*의 어느 부분이 얼마나 ConX#에 결합될 것인지를 예측되는지에 대한 그레이 스케일 히트 맵을 보여준다. 도 2A 및 도 2B에 나타난 것과 같이, 특히 TRBV의 경우 실질적인 교차-결합 (가령, 잘못된-연결)이 존재한다 (도 2B).

[0206] Script 2. 코돈 다양화

```
[0207] clear;
[0208] load('IniDesign.mat');
[0209] fConcQB = struPara.fConcQB;
[0210] fConcNa = struPara.fConcNa;
[0211] fConcMg = struPara.fConcMg;
[0212] fHybTemp = struPara.fHybTemp;
[0213] fCodonFreqThreshold = 0.15; % 최저 허용된 코돈 빈도
[0214] fSSThreshold = 0.6;
[0215] fCrossAssemblyThreshold = 0.02; % 최저 허용된 잘못-연결된 수준
[0216] iMaxLengthDE = 35; % ConX의 최대 허용 길이
[0217] struParaDiv.fCodonFreqThreshold = fCodonFreqThreshold;
[0218] struParaDiv.fSSThreshold = fSSThreshold;
[0219] struParaDiv.fCrossAssemblyThreshold = fCrossAssemblyThreshold;
[0220] struParaDiv.iMaxLengthDE = iMaxLengthDE;
[0221] %% 시작
[0222] ralonlNewDesign = cell(1,2);
[0223] iTotNumOfGene = size(ralonls{1},1);
[0224] xCrossPrimeAlreadyDesigned_Hyb_THyb = -ones(iTotNumOfGene);
[0225] %%
[0226] for iGeneToModify = 1:iTotNumOfGene
[0227]     cSA60 = ralonls{2}{iGeneToModify,1};
[0228]     cAAInFrame = nt2aa(cSA60(3:59));
[0229]     fprintf('%s\n',cSA60);
[0230]
[0231]     bHaveYouTriedInitialDesign = false;
[0232]     while(1)
[0233]         if ~bHaveYouTriedInitialDesign
[0234]             ralonlofThisGene{1} = ralonls{1}(iGeneToModify,:);
[0235]             ralonlofThisGene{2} = ralonls{2}(iGeneToModify,:);
```

```

[0236]         bHaveYouTriedInitialDesign = true;
[0237]     else %Randomize
[0238]         cTrialMiddle57 = fun_aa2nt(cAAInFrame,raCodonTable,fCodonFreqThreshold); % 주석 참고
[0239]         cAAToMakeSure = nt2aa(cTrialMiddle57,'AlternativeStartCodons','false');
[0240]         if ~strcmpi(cAAInFrame,cAAToMakeSure)
[0241]             error('something is wrong');
[0242]         end
[0243]         cTrialSA60 = [cSA60(1:2),cTrialMiddle57,'T'];
[0244]         fprintf('%s\n',cTrialSA60);
[0245]         ralonlofThisGene = fun_Designlonl(cTrialSA60,STRUPara);
[0246]     end
[0247]     iStartPosBM = ralonlofThisGene{1}(2);
[0248]     cTrialSimQB = ralonlofThisGene{2}{3};
[0249]     cTrialDE = ralonlofThisGene{2}{4};
[0250]
[0251]     % Check DE length
[0252]     iLefttestStartPosBM = 60 - iMaxLengthDE + 1;
[0253]     if iStartPosBM < iLefttestStartPosBM
[0254]         fprintf('DE too long\n');
[0255]         continue;
[0256]     end
[0257]
[0258]     %%
[0259]     vRow_THyb = -ones(1,iGeneToModify);
[0260]     vColumn_THyb = -ones(iGeneToModify,1);
[0261]
[0262]     for iDE = 1:iGeneToModify
[0263]         if iDE ~= iGeneToModify
[0264]             cDE = ralonlNewDesign{2}{iDE,4};
[0265]         else
[0266]             cDE = cTrialDE;
[0267]         end
[0268]         fFracHyb_THyb = NP_GetBoundFrac(cTrialSimQB,fConcQB,cDE,fConcQB/100,...
[0269]             fHybTemp,'Na',fConcNa,'Mg',fConcMg);
[0270]         vRow_THyb(iDE) = fFracHyb_THyb;
[0271]     end

```

```

[0272]         for iQB = 1:iGeneToModify
[0273]             if iQB ~= iGeneToModify
[0274]                 cExtSimQB = ralon1NewDesign{2}{iQB,2};
[0275]             else
[0276]                 cExtSimQB = cTrialExtSimQB;
[0277]             end
[0278]             fFracHyb_THyb = NP_GetBoundFrac(cExtSimQB,fConcQB,cTrialDE,fConcQB/100,...
[0279]                 fHybTemp,'Na',fConcNa,'Mg',fConcMg);
[0280]             vColumn_THyb(iQB) = fFracHyb_THyb;
[0281]
[0282]         end
[0283]
[0284]         vRowToUse = vRow_THyb;
[0285]         vColumnToUse = vColumn_THyb;
[0286]
[0287]         if vRowToUse(end) >= 0.5 && sum(vRowToUse(1:end-1))<fCrossAssemblyThreshold && ...
[0288]             vColumnToUse(end) >= 0.5 && sum(vColumnToUse(1:end-1))<fCrossAssemblyThreshold
[0289]             ralon1NewDesign{1}(iGeneToModify,:) = ralon1ofThisGene{1};
[0290]             ralon1NewDesign{2}(iGeneToModify,:) = ralon1ofThisGene{2};
[0291]
[0292]             xCrossPrimeAlreadyDesigned_Hyb_THyb(iGeneToModify,1:iGeneToModify) = vRow_THyb;
[0293]             xCrossPrimeAlreadyDesigned_Hyb_THyb(1:iGeneToModify,iGeneToModify) = vColumn_THyb;
[0294]
[0295]             break;
[0296]         end
[0297]
[0298]     end
[0299] end
[0300] %%
[0301] figure;
[0302] colormap(gray)
[0303] imagesc(1-xCrossPrimeAlreadyDesigned_Hyb_THyb);
[0304] Script2에 대한 주석:
[0305] "fun_aa2nt" 함수는 입력 raCodonTable에서 제공하는 코돈 테이블 정보, 그리고 입력 fCodonFreqThreshold에서
    제공하는 최저 허용 코돈 빈도를 사용하여, 입력 서열 cAAInFrame과 동일한 폴리펩티드를 인코딩하는 폴리뉴클
    레오티드 시퀀스를 반환한다.

```

[0306] 이 스크립트에 의해 생성된 이미지는 코돈 다양화 이후, 위에서 설명한 조건에서 0.05nM의 ConX#*가 5nM의 ConX#와 혼합될 때, ConX#*의 어느 부분이 ConX#에 결합될 것인지를 예측되는지에 대한 그레이 스케일 히트 맵을 보여준다. 도 3A 및 도 3B에 나타난 것과 같이, 오로지 특이적 혼성화만이 눈에 띄게 일어날 것으로 예상된다. 따라서, 이 실시예는 코돈 다양화 계획이 가능함을 보여주고, 코돈 다각화된 ConA 서열과 ConB 서열을 얻는 방법을 보여준다.

[0307] 실시예 3. 마우스 TRAV 유전자 및 TRBV 유전자로부터 유래된 커넥터 서열들

[0308] 이 실시예는 마우스 TRAV 유전자 및 TRBV 유전자로부터 유래된 코돈-다각화된 커넥터 서열을 제공한다. 상기 실시예들과 유사하게, ConA는 특이적 TRAV 서열의 커넥터이며, ConB는 특이적 TRBV 서열의 커넥터이다. 코돈 다양화는 실시예 2에서 기술된 것과 동일한 방법들을 이용하여 실행되었다. 표 1에서는 마우스 TRAV 유전자들로부터 유래된 코돈-다각화된 커넥터 서열을 보여준다. 표 2에서는 마우스 TRBV 유전자들로부터 유래된 코돈-다각화된 커넥터 서열을 보여준다. 표 1 및 2에서, 각 V 유전자에 대한 유전자 이름 및 기탁 번호는 제 1 열에 나타내고, 대응하는 커넥터 서열은 제 2 열에 나타낸다.

[0309]

표 1. 마우스 TRAV 유전자들로부터 유래된 커넥터 서열

서 열 식 별 번 호	마우스 TRAV 유전자들의 커넥터	커넥터 서열
1	Trav1 ENSMUST00000103567.5 의 > ConA	TCAAAGACTCTGCCTCATACCTCT
2	Trav2 ENSMUST00000196939.1 의 > ConA	CTGAGAGACGCAGCTGTGTATTACT
3	Trav3- 1 ENSMUST00000103569.2dml 의 > ConA	GGGGACTCAGCCGTGTACTTCT
4	Trav3- 3 ENSMUST00000181768.2 의 > ConA	GGTGACTCCGCAGCCTATTTCT
5	Trav3- 4 ENSMUST00000103670.3 의 > ConA	TGGGGATAGCGCAGTCTATTTCT
6	Trav3d- 3 ENSMUST00000196023.1 의 > ConA	CCGGAGACAGCGCAGTTTATTTTT
7	Trav3n- 3 ENSMUST00000197557.1 의 > ConA	GGTGACAGCGCCGTCTATTTTT
8	Trav4- 2 ENSMUST00000103637.5 의 > ConA	TGGAGGACTCAGGCACTTACTTCT
9	Trav4- 3 ENSMUST00000103655.2 의 > ConA	TGGAGGACTCTGGGACATACTTTT
10	Trav4-4-	TGGAGGACTCTGGCACCTATTTTT

[0310]

	dv10 ENSMUST00000103663.5 의 > ConA	
11	Trav4d- 3 ENSMUST00000103592.1 의 > ConA	ACTCGAGGATTCCGGTACTTATTCT
12	Trav4d- 4 ENSMUST00000103600.2 의 > ConA	GAAGACTCCGGGACCTACTTTT
13	Trav4n- 3 ENSMUST00000103618.1 의 > ConA	GCTGGAGGATTCCGGAACCTATTCT
14	Trav4n- 4 ENSMUST00000103627.2 의 > ConA	CTCGAAGATAGCGGCACATATTTT
15	Trav5- 1 ENSMUST00000103570.1 의 > ConA	GCCTGGTGATAGCGCAATATACTTCT
16	Trav5d- 4 ENSMUST00000179701.1 의 > ConA	TGGCGACTCTGCAATGTACTTCT
17	Trav5n- 4 ENSMUST00000179997.1 의 > ConA	CCCGGAGACTCTGCTATGTATTTT
18	Trav6- 1 ENSMUST00000103571.1 의 > ConA	GGAATCCGATAGCGCAGTCTATTACT
19	Trav6- 2 ENSMUST00000198058.1 의 > ConA	TCCGACAGCGCTGTCTACTACT
20	Trav6- 3 ENSMUST00000180549.2 의 > ConA	AAGAGATTGATAGCGCTGTTACTACT

[0311]

21	Trav6- 4 ENSMUST00000184650.1 의 > ConA	AGGAATCTGATTCCGCAGTCTATTTT
22	Trav6- 5 ENSMUST00000181210.2 의 > ConA	AGAATCTGATAGCGCCGTTTATTATT
23	Trav6- 6 ENSMUST00000103584.3 의 > ConA	GTCCGACTCCGCAGTCTACTACT
24	Trav6-7- dv9 ENSMUST00000103638.5 의 > ConA	AGGAGTCTGATTCTGCAGTCTACTATT
25	Trav6d- 3 ENSMUST00000181483.2 의 > ConA	CCAAGAAATAGATTCCGCAGTCTACTAC T
26	Trav6d- 4 ENSMUST00000180717.2 의 > ConA	GTCTGACAGCGCAGTCTACTTCT
27	Trav6d- 5 ENSMUST00000180687.2 의 > ConA	AGGAAAGCGATTCTGCAGTCTATTACT
28	Trav6d- 6 ENSMUST00000197754.1 의 > ConA	AAGAGTCTGACTCCGCAGTTTATTATT
29	Trav6d- 7 ENSMUST00000178650.2 의 > ConA	AGAATCCGACTCTGCAGTTTACTATT
30	Trav6n- 5 ENSMUST00000103611.1 의 > ConA	GAGTCTGATAGCGCTGTGTACTACT
31	Trav6n- 6 ENSMUST00000181793.2 의 >	GAATCTGACTCTGCCGTTTACTATT

[0312]

	ConA	
32	Trav6n- 7 ENSMUST00000179607.2 의 > ConA	AGTCCGACTCTGCTGTGTACTACT
33	Trav7- 1 ENSMUST00000198019.1 의 > ConA	CCATCTGATTCCGCACTGTATTCT
34	TRAV7-2 AC004407 의 > ConA	ACCTTCTGATAGCGCTCTCTATTTT
35	Trav7- 3 ENSMUST00000177622.3 의 > ConA	CCTTCTGATTCTGCACTGTACCTGT
36	Trav7- 4 ENSMUST00000181728.2 의 > ConA	CCAAGCGATTCTGCACTGTATTTT
37	Trav7- 5 ENSMUST00000200609.1 의 > ConA	CCTCTGACTCTGCAGTCTACCTCT
38	Trav7- 6 ENSMUST00000103641.5 의 > ConA	CCCAGCGACTCTGCAGTTTATCTCT
39	Trav7d- 2 ENSMUST00000200127.1 의 > ConA	CCGACAGCGCACTCTACCTGT
40	Trav7d- 3 ENSMUST00000179789.3 의 > ConA	TTCCGACTCTGCACTGTATCTGT
41	Trav7d- 4 ENSMUST00000178768.3 의 > ConA	TCCGATAGCGCCCTGTATTCT
42	Trav7d- 5 ENSMUST00000197128.1 의 > ConA	CTCCGATTCCGCACTCTATCTCT

[0313]

43	Trav7D- 6 ENSMUST00000196756.1 의 > ConA	CCTCCGATAGCGCTGTTTATCTCT
44	Trav7n- 4 ENSMUST00000103609.1 의 > ConA	GCGACAGCGCCCTGTACTTTT
45	Trav7n- 5 ENSMUST00000199753.1 의 > ConA	CCCTCTGATAGCGCACTGTATCTCT
46	Trav7n- 6 ENSMUST00000178100.2 의 > ConA	CTTCTGACAGCGCTGTGTATCTGT
47	Trav8- 1 ENSMUST00000103643.3 의 > ConA	GCGAGGACACAGCTGTTTACTTTT
48	TRAV8-2 AC004096 의 > ConA	AGTGCGAAGATACAGCAGTTTACTTCT
49	Trav8d- 1 ENSMUST00000103580.3 의 > ConA	CGGTGTGAGGATACTGCTGTTTATTCT
50	Trav8d- 2 ENSMUST00000198439.4 의 > ConA	CGAAGATACCGCCGTCTACTTTT
51	Trav8n- 2 ENSMUST00000103632.3 의 > ConA	GGCAACTGACACAGCAGTCTACTTTT
52	Trav9- 1 ENSMUST00000103581.5 의 > ConA	GAGCGATTCTGCCGTTTACTTCT
53	Trav9- 2 ENSMUST00000103654.2 의 > ConA	TCCGATTCCGCCGTGTATTTT
54	Trav9-	TTGGTCTGATTCTGCAGTTTACTTTT

[0314]

	4 ENSMUST00000103662.5 의 > ConA	
55	Trav9d- 1 ENSMUST00000178426.3 의 > ConA	GGTCTGATTCCGCTGTCTACTTTT
56	Trav9d- 2 ENSMUST00000199746.4 의 > ConA	CTGGTCTGACTCTGCTGTTTATTTT
57	Trav9d- 3 ENSMUST00000178252.2 의 > ConA	GGTCCGACTGGGCAGTCTATTTT
58	Trav9d- 4 ENSMUST00000200548.1 의 > ConA	TGGTCTGATTCTGCCGTCTATTCT
59	Trav9n- 2 ENSMUST00000198913.4 의 > ConA	AGCGACTCTGCCGTGTATTCT
60	Trav9n- 3 ENSMUST00000177705.2 의 > ConA	GAGCGATTGGGCAGTCTACTTTT
61	Trav9n- 4 ENSMUST00000103626.2 의 > ConA	TGGTCCGATTCTGCTGTCTATTTT
62	Trav10 ENSMUST00000103583.4 의 > ConA	AGCCTGAAGATTCAGCCATCTACTTCT
63	Trav10d ENSMUST00000103646.4 의 > ConA	CCCGAGGACTCTGCTATTACTTCT
64	Trav10n ENSMUST00000103612.1 의 > ConA	ACAGCCAGAAGATTCTGCAATATACTTC T
65	Trav11 ENSMUST00000103585.3 의 > ConA	GCTCGATGACACAGCTACATACATCT
66	TRAV11D AC004101 의 > ConA	TCCTGGATGATACTGCAACATACATAT

[0315]

67	Trav12- 1 ENSMUST00000200115.1 의 > ConA	CTCTCTGACTCTGCACTGTACTACT
68	Trav12- 2 ENSMUST00000180972.2 의 > ConA	ACTGTCTGACTCTGCACTCTATTACT
69	Trav12- 3 ENSMUST00000103657.5 의 > ConA	CTGTCCGATTCTGCACTCTACTACT
70	Trav12d- 1 ENSMUST00000181360.2 의 > ConA	AACTGTCTGATTCTGCTCTGTACTATT
71	Trav12d- 2 ENSMUST00000103593.2 의 > ConA	TCCGACTCCGCTCTGTATTTTT
72	Trav12d- 3 ENSMUST00000177703.2 의 > ConA	AGCGACTCTGCCCTCTACTACT
73	Trav12n- 1 ENSMUST00000198682.1 의 > ConA	TCTCTGACTCCGCTCTCTACTACT
74	Trav12n- 2 ENSMUST00000103619.2 의 > ConA	TCTCTGATTCTGCCCTCTACTTTT
75	Trav12n- 3 ENSMUST00000179583.2 의 > ConA	GCTCTCCGATTCTGCTCTGTATTATT
76	Trav13- 1 ENSMUST00000103651.3 의 > ConA	GACAACAGACTCAGGCACTTATCTCT
77	Trav13- 2 ENSMUST00000103658.3 의 >	AACAACACTGACTCTGGCACATATTTTT

[0316]

	ConA	
78	TRAV13-3 AC003995 의 > ConA	CACTGATAGCGGAACCTACCTCT
79	Trav13-4- dv7 ENSMUST00000180380.2 의 > ConA	TGACAGCGGCACCTACCTGT
80	Trav13- 5 ENSMUST00000103671.3 의 > ConA	ACTACAGATTCCGGCACTTACTTCT
81	Trav13d- 1 ENSMUST00000103588.3 의 > ConA	GCCAGATAACTGATTCTGGTACTTACCT GT
82	Trav13d- 2 ENSMUST00000197954.1 의 > ConA	ACAACTGACAGCGGAACATATCTCT
83	Trav13d- 3 ENSMUST00000179512.2 의 > ConA	AAATAACAGATAGCGGTACATACCTGT
84	Trav13d- 4 ENSMUST00000196079.1 의 > ConA	CCACAGATTCTGGCACCTACTTCT
85	Trav13n- 1 ENSMUST00000198359.1 의 > ConA	ACTGACTCCGGAACCTACCTCT
86	Trav13n- 2 ENSMUST00000196941.1 의 > ConA	ACCGACTCTGGCACTTACCTGT
87	Trav13n- 3 ENSMUST00000179580.2 의 > ConA	AATCACAGACTCTGGAACCTATCTGT
88	Trav13n- 4 ENSMUST00000196105.1 의 > ConA	CCAAATTACCGATTCTGGTACATACCTCT

[0317]

89	Trav14- 1 ENSMUST00000198297.1 의 > ConA	CGGAGATAGCGCCACATACTTTT
90	Trav14- 2 ENSMUST00000179267.3 의 > ConA	AACCTGGAGATTCTGCAACATATTCT
91	Trav14- 3 ENSMUST00000103589.5 의 > ConA	CTGGGGACTCTGCAACTTACTTCT
92	Trav14d- 1 ENSMUST00000181038.2 의 > ConA	CCTGGGAGACTCAGCTACCTACTTCT
93	Trav14d- 2 ENSMUST00000196802.1 의 > ConA	CCGGGGATAGCGCTACTTATTTTT
94	Trav14d-3- dv8 ENSMUST00000103608.3 의 > ConA	CCTGGGAGATTCCGCAACTTACTTTT
95	Trav14n- 1 ENSMUST00000177578.1 의 > ConA	CCAGGGGATTCTGCTACCTATTTTT
96	Trav14n- 2 ENSMUST00000197614.1 의 > ConA	CCCGGAGATTCTGCCACTTATTCT
97	Trav14n- 3 ENSMUST00000103652.4 의 > ConA	CTGGCGACAGCGCTACTTATTCT
98	Trav15-1-dv6- 1 ENSMUST00000103653.2 의 > ConA	AACCAGACGATTCGGGAAAGTATTCT
99	Trav15-2-dv6- 2 ENSMUST00000103660.3 의 >	CAGAGGATTCAGGGACGTACTTCT

[0318]

	ConA	
100	Trav15d-1-dv6d- 1 ENSMUST00000103616.4 의 > ConA	CCAGACGACTCCGGAAAAGTACTTTT
101	Trav15d-2-dv6d- 2 ENSMUST00000199800.1 의 > ConA	CCGAGGACTCCGGTACATACTTCT
102	Trav15n- 1 ENSMUST00000103590.3 의 > ConA	AACCCGATGACTCTGGTAAGTATTTT
103	Trav15n- 2 ENSMUST00000199112.1 의 > ConA	GCCAGAAGACTCCGGTACATATTTT
104	Trav16 ENSMUST00000103667.5 의 > ConA	TCAAATTGAAGATTCTGCAGTCTACTTTT
105	Trav16d- dv11 ENSMUST00000103606.1 의 > ConA	GATTGAGGACTCGGCAGTATATTCT
106	Trav16n ENSMUST00000199280.1 의 > ConA	AAATCGAAGACTCTGCAGTTTACTTTT
107	Trav17 ENSMUST00000103672.8 의 > ConA	GAGCGACTCAGCCAAGTACTTCT
108	Trav18 ENSMUST00000103673.10 의 > ConA	AGGGGATGCTGGGATCTACTTTT
109	Trav19 ENSMUST00000103674.5 의 > ConA	CCCGAAGATACAGCTGTCTACCTGT
110	Trav21- dv12 ENSMUST00000180938.2 의 > ConA	AGGGACGCAGCAGTCTATCATT
111	Trav23 ENSMUST00000199137.1 의 > ConA	GCCACTCTGCCATCTACTTCTGT

[0319]

[0320] 표 2. 마우스 TRBV 유전자들로부터 유래된 커넥터 서열

서열 식별 번호	마우스 TRBV 유전자들의 커넥터	커넥터 서열
112	Trbv1 ENSMUST00000103262.2 의 > ConB	GGCGCACACTGTACTGCACAT
113	Trbv2 ENSMUST00000103263.2 의 > ConB	TGATGACTCGGCCACATACTTCT
114	Trbv3 AE000663 의 > ConB	TGGAGGACTCAGCTGTGTACTTCT
115	Trbv4 ENSMUST00000103265.4 의 > ConB	ACCAGAAGATAGCGCAGTTTATCTGT
116	Trbv5 ENSMUST00000103266.2 의 > ConB	ATCCAGAAGACTCAGCTGTCTATTTTT
117	Trbv12-1 M15614 의 > ConB	TCGAAGATAGCGCCATGTACTTTT
118	Trbv12-2 M15613 의 > ConB	ACTGGAAGATAGCGCTGTGTATTCT
119	Trbv13- 1 ENSMUST00000194399.1 의 > ConB	AAGCCAGACCAGCCTCTATTTTT
120	Trbv13- 2 ENSMUST00000103270.3 의 > ConB	CCCCTCTCAGACATCAGTGTACTTCT
121	Trbv13- 3 ENSMUST00000103271.1 의 > ConB	GCCAGACCGCCGTGTATTCT
122	Trbv14 ENSMUST00000103272.3 의 > ConB	GGCGACACAGCCACCTATCTCT
123	Trbv15 ENSMUST00000103273.2 의 > ConB	GCCTAAAGACAGCGCTGTTTATCTCT
124	Trbv16 ENSMUST00000103274.3 의 > ConB	CCAGGACTCAGCGGTGTATCTTT
125	Trbv17 ENSMUST00000103275.3 의 > ConB	GCCTAGAGTATTCTGCCATGTACCTCT

[0321]

126	Trbv19 ENSMUST00000103276.2 의 > ConB	AAAAATGAGATGGCAGTCTTCCTCT
127	Trbv20 ENSMUST00000103277.1 의 > ConB	CGAGGATAGGGGCCTGTATCTCT
128	Trbv23 ENSMUST00000193997.5 의 > ConB	GCAGAAGACTCAGCACTGTACTTGT
129	Trbv24 IMGT 의 > ConB	GACGACTCAGCACTGTACCTCT
130	Trbv26 ENSMUST00000193064.1 의 > ConB	GGGGACTCCGCACTCTATCTCT
131	Trbv29 ENSMUST00000103281.2 의 > ConB	AAACAAACCAGACATCTGTGTACTTCT
132	Trbv30 ENSMUST00000103282.2 의 > ConB	GGCCTGGAGACAGCAGTATCTATTCT
133	Trbv31 ENSMUST00000193003.1 의 > ConB	TCAGCCATAGCGGTTTTTACCTCT

[0322]

[0323]

초기 기획에서 ConA# 서열과 ConB# 서열은 원래의 TRAV 서열 또는 TRBV 서열에 따라 기획된다. 본원에서 사용된 바와 같이, 기호 #는 TRAV 유전자 또는 TRBV 유전자의 수치적 ID를 나타낸다. 모든 ConX#의 모든 ConX#*에 대한 혼성화 수율은 기선 역할을 하도록 산출된다 (도 10A 및 도 10B). 도 10A는 코돈 다양화 없이, 원래의 TRAV 서열에 따라 기획된 커넥터 서열의 혼성화 수율을 보여준다 (ConA# 에서 ConA#*로). 도 10B는 코돈 다양화 없이, 원래의 TRBV 서열에 따라 기획된 커넥터 서열의 혼성화 수율을 보여준다 (ConB# 에서 ConB#*로). 코돈 다양화 동안, 일부 TRXV#_GL의 마지막 ~60개 염기에 대한 코돈 선택이 무작위화되고, 특정 혼성화를 허용하는 ConX# 서열이 선택된다. 그 다음, 코돈-다각화된 서열 세트를 사용하여, 모든 ConX#*에 대한 모든 ConX#의 혼성화 수율을 계산하여, 코돈 다양화가 성공적인지를 확인한다 (도 11A 및 도 11B). 도 11A는 코돈-다각화된 커넥터 서열의 혼성화 수율 (ConA#에서 ConA#*로)을 보여준다. 도 11B는 코돈-다각화된 커넥터 서열의 의 혼성화 수율 (ConB#에서 ConB#*로)을 보여준다.

[0324]

실시예 4. 임의적인 서열을 갖는 커넥터 서열

[0325]

표 3은 도 7에 기술된 계획에 따라, CDR3-J 폴리뉴클레오타이드 및 명시된 V 유전자 생식계열 폴리뉴클레오타이드를 연계하기 위해 커넥터 서열로서 사용될 수 있는 임의적인 서열을 제공한다.

[0326] 표 3. 임의적인 서열을 갖는 커넥터 서열

서열 식별 번호	커넥터 서열	서열 식별 번호	커넥터 서열
134	CCGGGATTTTGTGACTCATC	209	CTTGTCCTACTAAACGCAACG
135	GAGGATCGTATGTTTCGCAC	210	CGGGTATCACTGGGTAATGA
136	CTTGTTGTGCACTTACCGTAC	211	GGAACAGAGACCAATCCAGT
137	TGATGCATCTCCAGTACAGG	212	GTGGGCATCCGAAATTCAG
138	CTTCTGTGTGTACCTCGACA	213	CGCGACGACATTACCAATAG
139	CTTGCAATCCTTACCGTGC	214	CGTCGGAATATGCTCTCAGA
140	GCAAGTGTGAAAATGACCC	215	GATGCAGATCAATGAGTGGC
141	GAGTCTAGTCTCACAACCCA	216	ACTGCTTACAAGTGTCCACG
142	GAAATGTTGAGGACTCCACG	217	ACTGTATGCAAGCTAGTCCC
143	CCTAACAGATGCTACGTGGA	218	AGATCTCCCAAAAGTGTCCG
144	GTAGGTCCACACAGATTCCA	219	TTCCAGAACCATGTGATCCC
145	GCCAGTCACAGCAAATACAC	220	GCCTTGTCTTTCAACCTCTG
146	CCGCTACCAGTATGTACCTT	221	GATACGGATCTTCACATGCG
147	ACTGTGTTCTTGTCTTCCG	222	CGCTCATCTAGGTTGGACTA
148	TGAATGCATCTACGGTACCG	223	CGCGTTCAGATTCCAAACAG
149	GCGCTTATCAATCTTGCTCG	224	GCCTGGTTACACATGCTATC
150	CGGTCAATTCAGTAGCCACT	225	GCAAAGGTCCTACAGGTTTC
151	GGACACATGTACACTAGCCA	226	GGCTTCCATGTCTATGCTC
152	TGGGAGCTCTACGAAAATCC	227	GCGACATTAGCAGAGTAGGT
153	GTTCTCGAGATCGTCACACA	228	CTCGCCATACTATCTGCATG
154	CTTCTGCATTGATCCTTGC	229	CTACTGAACACTTGGCAAGC
155	GCAGAGTTGTGTGATTGGAG	230	CTGTTCAATTCCTGTGCGAG
156	CCCATCAATTCGGAACCATC	231	CACTGAGATGGAATTTGGCG
157	ATCGTAACCCAAGTCTGTGG	232	GTCCATCACAACTTCCACTG
158	GGCGAAATGATCCCTGAATG	233	GGCTCAGTCTACTTTGCTTC
159	AGTGCTCAGAACTTTCAGGC	234	GTTAGCTCCGACACAATGG
160	GATCGTTAACTCTTTGGGCG	235	CCGTGACACACTTTCATCTC
161	ACACGAGGATTGCTGTAGAG	236	CCGGGATGTCATTATGAGCA
162	TTCTACCACATTGTCTCGGG	237	GAGTGTCTACGAGATCAGA

[0327]

163	GAATGGCTAAACTGTGTCCC	238	TAACGTCTCTCTGAGTGTGG
164	CGGACTGTACGAGAACTGA	239	ACCCTAGACAAGAGACACCT
165	CTTGCGACAACTACTCCTG	240	TGCTCAGTACTCTTCATCGC
166	CCGTTTACTTTGTGCGCCAC	241	AGCTCAATCATGGCTATCGG
167	TGGATGATCACTTCGGCG	242	CTACACATTGCATCCAACCC
168	CCAACCTCTATATGTGCCCA	243	ACTTGTCTGAATAGCTCAGGC
169	TGAACAGGTATGCTCCAGAG	244	GTCTACCCTGAGAACCAGTT
170	TTGTGGATATCGTCTGGTCC	245	CAGCAACAACCTACCTTAGC
171	CTGTGGAACTCGACTCTTGT	246	CGATTGTGGTACGTGTCC
172	CTCCAGGATGCACAAATTCG	247	GCCATTTCCTTTGTACCTGC
173	GGCTCATGACAAAACACAGG	248	GGCCAATAGAGAGACCACAA
174	CCGAATCCGAAAACAACACG	249	CGGAGTCACATGGGTAGAAT
175	AGACCTAACACTGTGATGCG	250	CCCAGTACATTTGTCGGTTG
176	CCTGGGTGAGCATAAACTTC	251	CCCACTAGCTGCTACTCAAA
177	GAGTCTTGACGAACAAAGG	252	GGTGTGCGTCAAAGTAGAC
178	CACGTACCCATCATGTTCTG	253	TACTCCAGCTCTTACTGTGC
179	CCGTGTTAGTCAAGTGTGTG	254	GGATGAGCAGTCAACAGTTC
180	CTGGTGGCATAAATGGAACG	255	TCAGGATCGATCAGTTGTCC
181	TGGATGTGGGTATCAATGGG	256	CCTCTCTTTTGTGCGGAAAC
182	TGTGGCTAACGTAGGACAAG	257	TGCCTAGGATTCGAGAACG
183	CCCTCGTTGTGAAAATGTGC	258	GGCATTGTCCTTAACCTTCGC
184	TCGTCATAGGTCAGCTTACG	259	TGCATCTAACTACGATGGGC
185	CCTGATGACCTCTATGCCAA	260	CCCTAGTAGCCACACAACAT
186	CGGCAAGAATGAATAGGGTG	261	GTGCCATGAATCATCGTCTC
187	GTGCTATTGGTGGGAAATGG	262	CGCTCTGATGAAAGCTCCAT
188	GCCATGTTTGCTTACTGACG	263	CCAGCCATAGTGCATATCCT
189	CGTTGTGGCATTATTAGCG	264	GAGATTGTCATGTGGTCGAC
190	GCGGTAGGATTGGATCTCAT	265	CCGAGTCTAACAGGAAATC
191	CCTCGCAAAGCTGTTATGAC	266	CGCTTCGACTGAACCTTATG
192	GCCTTCATGTTATTGGACGC	267	CGATGCGACCAATAGAAGTG
193	AGCTGTAGTGTTCTTGAGGC	268	GCCCTTGGTACGACATATTG
194	GGTAGTGTTCTGTGACATG	269	CAGTGATTAGGTGACGCAG

[0328]

195	CGCGGCATATGTTTCATATCC	270	GGCATGGAAGAGGTAGTTTC
196	GAGACTGGATCATGCAACAG	271	CCGATCGTATTCTGTGTCCA
197	CACAACCTCTCTGGACTCCA	272	CTAAGTCAAGCACATGGGAC
198	CGACCATGATCTGTATGCGT	273	GATCCACACTCAATCTCCTG
199	GGTGTGACTCTTGTTCCTG	274	CCTTGTACATGCTGGTATC
200	ACGTACATACAAGTCTGGCG	275	CGCGATTGTGGTTAATAGGC
201	CCTCAAGGATTCACCTGCAA	276	GTAGGCAAAGTTCACCACAC
202	CTGTATAGGATGTCCACGCA	277	GCCACGAATCGAACAAGTAC
203	GCCTGTGATTGGTAAATGCG	278	TTGAGATCTCGATGAGCACG
204	CGCACTCGTAGCATCTAGAA	279	GGGCCAAGATCTATTTCGTCA
205	CGATTGTGTGCCCTAGCTG	280	GTGGCTATAGGTATGTCCGA
206	CCCACCTTCATCTGACTCTGA	281	CCACACTTCTGCATTTCGAC
207	CGGCATTGTACAGGTGTAC	282	CGGCATCTCAAAGCACATAC
208	TCTCCTATTCCTGAACGG	283	CGTCCACAAATTTACTGCCC

[0329]

[0330]

실시예 5. 차세대 시퀀싱을 이용하여 어셈블리된 TCR 유전자들의 특징화

[0331]

쌍을 이룬 TCRs를 인코딩하는 핵산 서열의 풀은 본원에 기술된 방법들을 이용하여 (가령, 실시예 1에 방법에서 약간의 변형을 가한) 준비하였다. 태생적으로 쌍을 이룬 TCRs를 인코딩하는 참조 서열은 공개적으로 이용가능한 라이브러리로부터 구하였다. 본 실시예에서 입증하기 위해 553개 참조 서열이 선별되었다. 이 실시예에서, CDR3-J α (또는 CDR3-J α 단편)를 인코딩하는 핵산 서열과 CDR3-J β (또는 CDR3-J β 단편)를 인코딩하는 핵산 서열을 별도로 합성하였다. 대안적으로, 쌍을 이룬 CDR3-J α 및 CDR3-J β 는 한개 단편 상에 함께 합성될 수 있다.

[0332]

553개의 CDR3-J α 단편과 553개의 CDR3-J β 단편이 합성되었고, 함께 연결되어 (가령, 절찰, 중첩 PCR, 등에 의해), 쌍을 이룬 CDR3-J α -CDR3-J β 단편 풀을 만들었다. 확실하게 CDR3-J α 가 태생적으로 쌍을 이룬 CDR3-J β 에 결합되도록, 임의적인 커넥터 서열을 각 CDR3-J α 상에 합성하고, CDR3-J α 단편 풀에서 다른 임의적인 커넥터 서열과 교차-혼성화를 최소화시킬 수 있도록 임의적인 커넥터 서열을 기획하였다. 상기 임의적인 커넥터 서열의 상보적 서열은 태생적으로 쌍을 이룬 CDR3-J β 상에 합성하였다. 그 다음, TRAV 단편 풀 (참조 서열에 따른 사전-합성됨)을 쌍을 이룬 CDR3-J α -CDR3-J β 단편에 연결시켜, TRAV-CDR3-J α -CDR3-J β 단편 풀을 만들었으며, 이때 각 단편은 이의 동계 CDR3-J α 에 연결된 TRAV 서열을 포함한다. 그 다음, TRBC1 서열을 TRAV-CDR3-J α -CDR3-J β 단편의 하류에 부속시켜, TRAV-CDR3-J α -CDR3-J β -TRBC1 단편을 형성하였다. 이들 단편을 고리화시키고, CDR3-J β 의 바로 상류를 절단하여 재-선형화시켜, CDR3-J β -TRBC1-TRAV-CDR3-J α 단편을 만들었다. TRBC1과 TRAV 단편은 프레임-내 자가-절단 P2A 서열이 TRBC1과 TRAV를 연결시키도록 기획되었다. 그 다음, TRBV 단편 (참조 서열에 따른 사전-합성됨) 풀을 CDR3-J β -TRBC1-TRAV-CDR3-J α 단편에 연결시켜 TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 를 만들었으며, 이것은 클론의 풍도 및 이들 클론의 연결 정확도를 평가하기 위해 차세대 시퀀싱 (NGS)을 겪게되었다. 여기에서, NGS 데이터에서 각 클론은 특유 서열을 지칭한다. 553개 참조 서열이 본 실시예에서 이용되었기 때문에, NGS 데이터에는 총 553개 클론이 있었다. 본원에 기재된 데이터 분석을 위해, CDR3-J α 서열을 사용하여 클론을 제시하였다.

[0333]

도 12는 쌍을 이룬 CDR3-J α -CDR3-J β 단편을 생성시킨 후, 각 클론의 정확도 및 풍도를 나타낸다. 각 데이터 점은 CDR3-J α -CDR3-J β 단편의 클론에 대응한다. 정확도는 동족 CDR3-J β 단편에 연결된 CDR3-J α 단편의 분획을 지칭한다. 각 CDR-J α 있어서, 정확도는 올바르게 연결된 CDR3-J β 단편의 수를 연결된 CDR3-J β 단편의 총 수로 나누어 계산할 수 있다. 풍도(abundance)는 전체 클론 풀에서 각 클론의 분획(fraction)을 나타내며, 해당 클론의 총 관독 수를 모든 클론의 총 관독 수로 나누어 계산할 수 있다. 데이터에 따르면 553개 클론 중 497개가 상자에 표시된 바와 같이, 95% 이상의 정확도와 0.1/553 이상의 풍도를 갖는다.

[0334]

도 13은 TRAV-CDR3-J α -CDR3-J β 단편을 생성시킨 후, 각 클론의 정확도 및 풍도를 나타낸다. 각 데이터 점은 TRAV-CDR3-J α -CDR3-J β 단편의 클론에 대응한다. 정확도는 동족 TRAV 단편에 연결된 CDR3-J α -CDR3-J β 단편의 분획을 지칭한다. 각 CDR3-J α -CDR3-J β 에 있어서, 정확도는 올바르게 연결된 TRAV 단편의 수를 연결된 TRAV 단편의 총 수로 나누어 계산할 수 있다. 풍도는 전체 클론 풀에서 각 클론의 분획을 나타내며, 해당 클론

의 총 판독 수를 모든 클론의 총 판독 수로 나누어 계산할 수 있다. 데이터에 따르면 553개 클론 중 523개가 상자에 표시된 바와 같이, 95% 이상의 정확도와 0.1/553 이상의 풍도를 갖는다.

[0335] 도 14는 상기 풀에서 각 클론에 각 TRAV를 매핑시킨 히트맵(heatmap)을 나타낸다. 클론 번호는 이의 동계 TRAV 유전자 이름에 따라 등급화된다. 데이터는 각 클론에 있어서, 판독의 대다수가 정확한 TRAV 서열을 갖는다는 것을 보여주며, 이것은 CDR3-J α -CDR3-J β 단편을 이의 동계 TRAV 단편에 연결할 때 높은 정확도를 나타낸다.

[0336] 도 15는 CDR3-J α -CDR3-J β 단편을 생성시킨 후 풍도 대비 TRAV-CDR3-J α -CDR3-J β 단편을 생성시킨 후, 각 클론의 풍도 (가령, 도 15에서 TRAV 추가)를 보여준다. 데이터는 전반적인 편향이 CDR3-J α 및 CDR3-J β 단편의 결합 동안 편향에 의해 지배된다는 것을 보여준다. 이러한 편향은 쌍을 이룬 CDR3-J α -CDR3-J β 단편을 직접 합성함으로써 감소시키거나, 또는 회피할 수 있다.

[0337] 도 16은 TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 단편을 생성시킨 후, 각 클론의 정확도 및 풍도를 나타낸다. 각 데이터 점은 TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 단편의 클론에 대응한다. 정확도는 동계 TRBV 단편에 연결된 CDR3-J β -TRBC1-TRAV-CDR3-J α 단편의 분획을 지칭한다. 각 CDR3-J β -TRBC1-TRAV-CDR3-J α 에 있어서, 정확도는 올바르게 연결된 TRBV 단편의 수를 연결된 TRBV 단편의 총 수로 나누어 계산할 수 있다. 풍도는 전체 클론 풀에서 각 클론의 분획을 나타내며, 해당 클론의 총 판독 수를 모든 클론의 총 판독 수로 나누어 계산할 수 있다. 데이터에 따르면 553개 클론 중 514개가 상자에 표시된 바와 같이, 95% 이상의 정확도와 0.1/553 이상의 풍도를 갖는다.

[0338] 도 17은 상기 풀에서 각 클론에 각 TRBV를 매핑시킨 히트맵을 나타낸다. 클론 번호는 이의 동계 TRBV 유전자 이름에 따라 등급화된다. 데이터는 각 클론에 있어서, 판독의 대다수가 정확한 TRBV 서열을 갖는다는 것을 보여주며, 이것은 CDR3-J β -TRBC1-TRAV-CDR3-J α 단편을 이의 동계 TRBV 단편에 연결할 때 높은 정확도를 나타낸다.

[0339] 도 18은 TRBV-CDR3-J β -TRBC1-TRAV-CDR3-J α 단편을 생성시킨 후, 각 클론의 전체적인 정확도 및 풍도를 나타낸다. 각 클론에 대한 전체 정확도는 도 12, 13 및 16에 나타난 각 단계의 정확도를 곱하여 계산되었다. 풍도는 이 클론의 총 판독 수를 모든 클론의 총 판독 수로 나누어 계산하였다.

[0340] 본 발명의 다양한 구체예들을 본 출원에 나타내고 설명하였으나, 해당 분야의 숙련된 기술자들에게 이러한 구체예들이 단지 예로서 제공되는 것임은 자명할 것이다. 본 발명의 범위에서 벗어나지 않는 수많은 변형들, 변화들, 및 치환들이 해당 분야의 숙련된 기술자들에 의해 이루어질 수 있다. 본 출원에 기재된 본 발명의 구체예들에 대한 다양한 대체예들이 사용될 수 있음을 이해하여야 한다.

[0341] 구체예 단락

[0342] [1] T-세포 수용체 (TCR) 쇠 또는 이의 일부분을 인코딩하는 핵산 분자를 만드는, 다음을 포함하는 방법: (a) TCR 쇠의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자를 제공하고; (b) 다수의 핵산 분자들을 제공하고, 이들 다수의 핵산 분자 각각은 TCR V 유전자에서 유래된 서열을 포함하며, 이때 상기 다수 핵산 분자는 적어도 두 개의 상이한 TCR V 유전자들로부터 유래된 적어도 두 개의 상이한 서열을 포함하고; 그리고 (c) 동일한 구획(compartment)에서 (a)의 적어도 하나의 핵산 분자를 (b)의 다수의 핵산 분자에 접촉시키고, 이때 (a)의 적어도 하나의 핵산 분자는 상기 다수 핵산 분자들중 핵산 분자에 연계되어, CDR3을 인코딩하는 서열 및 상기 적어도 두 개의 상이한 TCR V 유전자들중 하나로부터 유래된 서열을 포함하는 제 3 핵산 분자를 만들 수 있고, 그렇게 함으로써 상기 TCR 쇠 또는 이의 일부분을 인코딩하는 핵산 분자가 생성된다.

[0343] [2] 단락 [1]의 방법, 이때 적어도 하나의 핵산 분자는 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 쇠의 CDR3을 인코딩하는 서열을 포함한다.

[0344] [3] 단락 [1] 또는 [2]의 방법, 이때 (a)의 적어도 하나의 핵산 분자는 상기 적어도 두 개의 상이한 TCR V 유전자들중 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있다.

[0345] [4] 단락 [1]의 방법, 이때 상기 적어도 하나의 핵산 분자는 TCR 쇠의 J 영역을 더 포함한다.

[0346] [5] 단락 [2]의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 쇠의 J 영역을 더 포함한다.

[0347] [6] 단락 [1]-[5]중 임의의 하나의 방법, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TCR V 유전자 또는 마우스 TCR V 유전자이다.

- [0348] [7] 단락 [1]-[6]중 임의의 하나의 방법, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 및 TRAV41로 구성된 군에서 선택된다.
- [0349] [8] 단락 [1]-[6]중 임의의 하나의 방법, 이때 상기 적어도 두 개의 TCR V 유전자는 인간 TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 및 TRBV30로 구성된 군에서 선택된다.
- [0350] [9] 단락 [1]-[8]중 임의의 하나의 방법, 이때 적어도 두 개의 상이한 TCR V 유전자로부터 유래된 서열 다수에서 각 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함한다.
- [0351] [10] 단락 [1]-[9]중 임의의 하나의 방법, 이때 상기 TCR쇄는 TCR알파쇄, TCR베타쇄, TCR감마쇄, 또는 TCR델타쇄이다.
- [0352] [11] 단락 [1]-[10]중 임의의 하나의 방법, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR쇄의 추가 CDR3를 인코딩하는 추가 서열을 더 포함한다.
- [0353] [12] 단락 [11]의 방법, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR쇄의 추가 J영역을 포함한다.
- [0354] [13] 단락 [11] 또는 [12]의 방법, 이때 CDR3를 인코딩하는 서열 및 추가 CDR3를 인코딩하는 추가 서열은 최대한 100개 뉴클레오타이드에 의해 분리되어 있다.
- [0355] [14] 단락 [11]-[13]중 임의의 하나의 방법, 이때 상기 TCR쇄 및 추가 TCR쇄는 TCR쇄의 동계쌍이다.
- [0356] [15] 단락 [1]-[14]중 임의의 하나의 방법, 이때 상기 적어도 하나의 핵산 분자는 커넥터 서열을 포함하고, 이 커넥터 서열은 상기 적어도 하나의 핵산 분자를 상기 다수 핵산 분자들중 핵산 분자에 연계시켜 제 3 핵산 분자를 생성시킬 수 있다.
- [0357] [16] 단락 [15]의 방법, 이때 상기 적어도 하나의 핵산 분자 및 상기 다수 핵산 분자들중 핵산 분자는 기능성 TCR쇄 또는 이의 일부분을 인코딩한다.
- [0358] [17] 단락 [15] 또는 [16]의 방법, 이때 상기 다수 핵산 분자들중 핵산 분자 안티-커넥터 서열을 포함하고, 이 안티-커넥터 서열은 (a)의 적어도 하나의 핵산 분자의 커넥터 서열에 상보적이다.
- [0359] [18] 단락 [1]-[17]중 임의의 하나의 방법, (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 연계시키는 것을 더 포함한다.
- [0360] [19] 단락 [18]의 방법, 이때 연계는 (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 혼성화시키는 것을 포함한다.
- [0361] [20] 단락 [19]의 방법, 이때 혼성화는 (a)의 적어도 하나의 핵산 분자의 커넥터 서열을 (b)의 다수의 핵산 분자의 핵산 분자의 안티-커넥터 서열에 혼성화시키는 것을 포함한다.
- [0362] [21] 단락 [18]-[20]중 임의의 하나의 방법, (i) (a)의 적어도 하나의 핵산 분자를 주형으로 삼아 상기 다수 핵산 분자들중 핵산 분자의 자유 3' 단부를 연장시키고, 및/또는 (ii) 상기 다수 핵산 분자들중 핵산 분자를 주형으로 삼아 (a)의 적어도 하나의 핵산 분자의 자유 3' 단부를 연장시켜, 제 3 핵산 분자를 만드는 것을 더 포함한다.
- [0363] [22] 단락 [1]-[21]중 임의의 하나의 방법, (a)의 적어도 하나의 핵산 분자와 (b)의 다수의 핵산 분자의 핵산 분자를 결합시키는 것을 더 포함한다.
- [0364] [23] 단락 [1]-[22]중 임의의 하나의 방법, 제 3 핵산 분자에 제한절단 효소를 접촉시켜, 스티커 단부를 만드는 것을 더 포함한다.
- [0365] [24] 단락 [1]-[23]중 임의의 하나의 방법, 제 3 핵산 분자에 추가 핵산 분자를 접촉시키는 것을 더 포함한다.

- [0366] [25] 단락 [24]의 방법, 이때 상기 추가 핵산 분자는 TCR 쉼의 불변 영역 또는 이의 일부분을 인코딩한다.
- [0367] [26] 단락 [24] 또는 [25]의 방법, 제 3 핵산 분자와 상기 추가 핵산 분자를 결합시키는 것을 더 포함한다.
- [0368] [27] 단락 [1]-[26]중 임의의 하나의 방법, 핵산 분자들 다수 (이때, 각 핵산 분자는 상이한 TCR 쉼 또는 이의 일부분을 인코딩하는)가 동일한 구획 안에서 생성된다.
- [0369] [28] 단락 [27]의 방법, 이때 상기 핵산 분자들 다수에서 적어도 5가지 상이한 핵산 분자가 상기 동일한 구획 안에서 생성된다.
- [0370] [29] 단락 [1]-[26]중 임의의 하나의 방법, 이때 적어도 10개 상이한 핵산 분자 of 상기 다수의 핵산 분자가 상기 동일한 구획 안에서 생성된다.
- [0371] [30] 단락 [1]-[29]중 임의의 하나의 방법, 이때 상기 동일한 구획은 웰, 튜브, 또는 소적이다.
- [0372] [31] 단락 [1]-[30]중 임의의 하나의 방법, 이때 상기 적어도 하나의 핵산 분자 특유 바코드를 포함한다.
- [0373] [32] 단락 [31]의 방법, 이때 상기 특유 바코드는 프라이머 결합 부위다.
- [0374] [33] 단락 [15]-[30]중 임의의 하나의 방법, 이때 상기 커넥터 서열은 특유 바코드를 포함한다.
- [0375] [34] 단락 [33]의 방법, 이때 상기 특유 바코드는 프라이머 결합 부위다.
- [0376] [35] 다음을 포함하는 조성물:
- [0377] (a) 핵산 분자들 다수, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함하고, CDR3 서열을 포함하지 않고, 이때 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 안티-커넥터 서열을 포함하고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 안티-커넥터 서열을 포함하고, 이때 제 1 안티-커넥터 서열은 제 2 안티-커넥터 서열과는 상이하며, 이때 제 1 핵산 분자의 TCR V 유전자로부터 유래된 서열과 제 2 핵산 분자는 상이한 TCR V 유전자로부터 유래되고; 그리고
- [0378] (b) TCR 쉼의 CDR3을 인코딩하는 서열을 포함하는 적어도 하나의 핵산 분자, 이때 상기 적어도 하나의 핵산 분자는 제 1 안티-커넥터 서열에 상보적인 제 1 커넥터 서열을 더 포함할 수 있다.
- [0379] [36] 단락 [35]의 조성물, 이때 상기 조성물은 액체 조성물이다.
- [0380] [37] 단락 [35] 또는 [36]의 조성물, 이때 상기 (a)의 핵산 분자 다수와 상기 (b)의 적어도 하나의 핵산 분자는 동일한 구획 안에 있다.
- [0381] [38] 단락 [35]-[37]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 이 TCR V 유전자의 적어도 10개 뉴클레오티드를 포함한다.
- [0382] [39] 단락 [35]-[38]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자이다.
- [0383] [40] 단락 [35]-[39]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및/또는 FR3을 인코딩하는 서열을 포함한다.
- [0384] [41] 단락 [35]-[40]중 임의의 하나의 조성물, 이때 상기 적어도 하나의 핵산 분자는 TCR 쉼의 J 영역을 더 포함한다.
- [0385] [42] 단락 [35]-[41]중 임의의 하나의 조성물, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR 쉼의 추가 CDR3을 인코딩하는 추가 서열을 더 포함한다.
- [0386] [43] 단락 [42]의 조성물, 이때 상기 적어도 하나의 핵산 분자는 추가 TCR 쉼의 추가 J 영역을 더 포함한다.
- [0387] [44] 단락 [42] 또는 [43]의 조성물, 이때 CDR3을 인코딩하는 서열과 CDR3을 인코딩하는 추가 서열은 최대한 100개 뉴클레오티드에 의해 분리되어 있다.
- [0388] [45] 단락 [42]-[44]중 임의의 하나의 조성물, 이때 상기 TCR 쉼과 추가 TCR 쉼은 TCR 쉼의 동계 쌍이다.
- [0389] [46] 단락 [35]-[45]중 임의의 하나의 조성물, 이때 상기 (b)의 적어도 하나의 핵산 분자는 제 1 다수의 핵산 분자를 포함하고, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 쉼의 CDR3을 인코딩하는 서열을 포함한다.

- [0390] [47] 단락 [46]의 조성물, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 TCR 쉼의 상이한 CDR3을 인코딩한다.
- [0391] [48] 단락 [46] 또는 [47]의 조성물, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자 상이한 커넥터 서열을 포함하고, 이때 상이한 커넥터 서열은 임의의 주어진 단일 TCR V 유전자로부터 유래된 서열을 포함하는 상기 다수 핵산 분자들중 핵산 분자에 특이적으로 연계될 수 있다.
- [0392] [49] 단락 [35]-[48]중 임의의 하나의 조성물, 이때 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 TCR V 유전자 서열을 포함한다.
- [0393] [50] 단락 [49]의 조성물, 이때 상기 TCR V 유전자 서열은 재배열된 유전자에서 상기 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자에서 적어도 3개 뉴클레오타이드를 포함한다.
- [0394] [51] 단락 [35]-[50]중 임의의 하나의 조성물, 이때 제 1 안티-커넥터 서열 또는 제 2 안티-커넥터 서열은 사전-결정된 서열을 포함한다.
- [0395] [52] 단락 [35]-[51]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열은 제 1 안티-커넥터 서열에 혼성화된다.
- [0396] [53] 단락 [35]-[52]중 임의의 하나의 조성물, 이때 상기 (b)의 적어도 하나의 핵산 분자는 특유 바코드를 포함한다.
- [0397] [54] 단락 [53]의 조성물, 이때 상기 특유 바코드는 프라이머 결합 부위다.
- [0398] [55] 단락 [35]-[52]중 임의의 하나의 조성물, 이때 상기 적어도 하나의 핵산 분자의 제 1 커넥터 서열은 특유 바코드를 포함한다.
- [0399] [56] 단락 [55]의 조성물, 이때 상기 특유 바코드는 프라이머 결합 부위다.
- [0400] [57] 핵산 분자들 다수를 생성시키는, 다음을 포함하는 방법: (a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3 및 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래되고; (b) 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고, 이때 핵산 분자는 불변 도메인을 인코딩하는 서열을 포함하지 않고; 그리고 (c) 제 1 다수 핵산분자들에서 핵산 분자와 제 2 다수 핵산 분자들에서 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 형성하게 되고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 TCR V 유전자는 TCR 쉼의 동계 쌍으로부터 유래된다.
- [0401] [58] 단락 [57]의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 TCR 쉼의 상이한 제 1 CDR3 및/또는 제 2 TCR 쉼의 상이한 CDR3을 인코딩하는 서열을 포함한다.
- [0402] [59] 단락 [57] 또는 [58]의 방법, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0403] [60] 단락 [57]-[59]중 임의의 하나의 방법, 이때 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들은 동일한 구획 안에서 접촉된다.
- [0404] [61] 단락 [57]-[60]중 임의의 하나의 방법, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 커넥터 서열을 더 포함하고, 이때 상기 커넥터 서열은 제 1 다수의 핵산 분자들에서 핵산 분자와 제 2 다수의 핵산 분자들에서 핵산 분자를 연계한다.
- [0405] [62] 단락 [61]의 방법, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 안티-커넥터 서열을 더 포함하고, 안티-커넥터 서열은 상기 커넥터 서열에 상보적이다.
- [0406] [63] 단락 [62]의 방법, 이때 상기 커넥터 서열은 안티-커넥터 서열에 혼성화되어, 제 1 다수의 핵산 분자들에서 핵산 분자와 제 2 다수의 핵산 분자들에서 핵산 분자를 연계시킨다.
- [0407] [64] 단락 [58]-[63]중 임의의 하나의 방법, 이때 상기 커넥터 서열은 코돈-다각화되어, 제 1 다수의 핵산 분자들에서 핵산 분자의 커넥터 서열은 제 1 다수의 핵산 분자들에서 다른 핵산 분자들의 다른 커넥터 서열과는 상이하다.

- [0408] [65] 단락 [57]-[64]중 임의의 하나의 방법, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 TCR 쉼의 제 1 J 영역 및/또는 제 2 TCR 쉼의 제 2 J 영역을 더 포함한다.
- [0409] [66] 단락 [57]-[65]중 임의의 하나의 방법, 이때 (i) 제 1 TCR 쉼은 TCR 알파 쉼이고, 제 2 TCR 쉼은 TCR 베타 쉼이거나, 또는 (ii) 제 1 TCR 쉼은 TCR 감마 쉼이고, 제 2 TCR 쉼은 TCR 델타 쉼이다.
- [0410] [67] 단락 [57]-[66]중 임의의 하나의 방법, 이때 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자이다.
- [0411] [68] 단락 [57]-[67]중 임의의 하나의 방법, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 이중-가닥으로 된 핵산 분자이다.
- [0412] [69] 단락 [57]-[68]중 임의의 하나의 방법, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 자가-절단 펩티드의 일부분을 인코딩하는 서열을 더 포함한다.
- [0413] [70] 단락 [62]-[69]중 임의의 하나의 방법, 이때 상기 안티-커넥터 서열은 제 2 다수의 핵산 분자들에서 핵산 분자의 오버행이다.
- [0414] [71] 단락 [62]-[70]중 임의의 하나의 방법, 이때 상기 커넥터 서열 또는 상기 안티-커넥터 서열의 길이는 적어도 3개 뉴클레오티드이다.
- [0415] [72] 단락 [63]-[71]중 임의의 하나의 방법, (i) 제 2 다수의 핵산 분자들에서 핵산 분자에 혼성화된 제 1 다수의 핵산 분자들에서 핵산 분자의 3' 단부를 연결시키거나, 및/또는 (ii) 제 1 다수의 핵산 분자들에서 핵산 분자에 혼성화된 제 2 다수의 핵산 분자들에서 핵산 분자의 3' 단부를 연결시키는 것을 더 포함한다.
- [0416] [73] 단락 [57]-[72]중 임의의 하나의 방법, 제 1 다수의 핵산 분자들에서 핵산 분자에 제 2 다수의 핵산 분자들에서 핵산 분자를 결합시키는 것을 더 포함한다.
- [0417] [74] 단락 [57]-[73]중 임의의 하나의 방법, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자에 제한절단 효소를 접촉시켜, 스티커 단부를 만드는 것을 더 포함한다.
- [0418] [75] 단락 [57]-[74]중 임의의 하나의 방법, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자에 불변 영역 또는 이의 일부분을 인코딩하는 서열을 인코딩하는 추가 핵산 분자를 접촉시킨다.
- [0419] [76] 단락 [74] 또는 [75]의 방법, 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 상기 추가 핵산 분자에 상기 스티커 단부를 통하여 결합시키는 것을 더 포함한다.
- [0420] [77] 단락 [57]-[76]중 임의의 하나의 방법, 이때 제 1 CDR3을 인코딩하는 서열과 제 2 CDR3을 인코딩하는 제 2 서열은 최대한 100개 뉴클레오티드에 의해 분리되어 있다.
- [0421] [78] 단락 [57]-[77]중 임의의 하나의 방법, 이때 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.
- [0422] [79] 단락 [57]-[77]중 임의의 하나의 방법, 이때 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 더 포함한다.
- [0423] [80] 다음을 포함하는 조성물: (a) 제 1 다수의 핵산 분자, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉼의 제 1 CDR3과 제 2 TCR 쉼의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR 쉼의 동계 쌍으로부터 유래되며; 그리고 (b) 제 2 다수의 핵산 분자들, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열을 포함하지 않고;
- [0424] 이때 (i) 제 1 다수의 핵산 분자에서 각 핵산 분자는 상이한 제 1 CDR3 및/또는 제 2 CDR3을 인코딩하는 서열을 포함하고, 및/또는 (ii) 제 2 다수의 핵산 분자들에서 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0425] [81] 단락 [80]의 조성물, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자 커넥터 서열을 더 포함하고, 이때 주

어진 커넥터 서열은 제 1 다수의 핵산 분자들에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 연계시키는데 이용가능하다.

- [0426] [82] 단락 [80] 또는 [81]의 조성물, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자 안티-커넥터 서열을 더 포함하고, 안티-커넥터 서열은 상기 커넥터 서열에 상보적이다.
- [0427] [83] 단락 [81] 또는 [82]의 조성물, 이때 상기 커넥터 서열은 코돈-다각화되어, 제 1 다수의 핵산 분자들에서 주어진 핵산 분자의 주어진 커넥터 서열은 제 1 다수의 핵산 분자들에서 다른 핵산 분자들의 다른 커넥터 서열들과는 상이하다.
- [0428] [84] 단락 [81]-[83]중 임의의 하나의 조성물, 이때 상기 커넥터 서열은 아미노산 서열을 인코딩한다.
- [0429] [85] 단락 [84]의 조성물, 이때 상기 커넥터 서열은 제 1 TCR 쉘의 제 1 CDR3 및 제 2 TCR 쉘의 제 2 CDR3을 인코딩하는 서열과 프레임 내에 있다.
- [0430] [86] 단락 [81]-[85]중 임의의 하나의 조성물, 이때 상기 커넥터 서열은 적어도 3개 뉴클레오티드를 포함한다.
- [0431] [87] 단락 [86]의 조성물, 이때 상기 커넥터 서열은 커넥터 서열은 재배열된 유전자에서 제 1 TCR 쉘의 제 1 CDR3 또는 제 2 TCR 쉘의 제 2 CDR3을 인코딩하는 서열에 인접한 상기 TCR V 유전자의 적어도 3개 뉴클레오티드를 포함한다.
- [0432] [88] 단락 [84]-[87]중 임의의 하나의 조성물, 이때 상기 주어진 커넥터 서열에 의해 인코딩된 주어진 아미노산 서열은 적어도 하나의 다른 커넥터 서열에 의해 인코딩된 적어도 하나의 다른 아미노산 서열과 동일하거나, 또는 실질적으로 동일하다.
- [0433] [89] 단락 [84]-[87]중 임의의 하나의 조성물, 이때 상기 주어진 커넥터 서열에 의해 인코딩된 주어진 아미노산 서열은 다른 커넥터 서열에 의해 인코딩된 다른 아미노산 서열과 상이하다.
- [0434] [90] 단락 [80]-[89]중 임의의 하나의 조성물, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 제 1 TCR 쉘의 제 1 J 영역 및/또는 제 2 TCR 쉘의 제 2 J 영역을 더 포함한다.
- [0435] [91] 단락 [80]-[90]중 임의의 하나의 조성물, 이때 상기 조성물은 액체 조성물이다.
- [0436] [92] 단락 [80]-[91]중 임의의 하나의 조성물, 이때 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자는 동일한 구획 안에 있다.
- [0437] [93] 단락 [81]-[92]중 임의의 하나의 조성물, 이때 상기 제 1 다수의 핵산 분자들에서 주어진 핵산 분자는 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 주어진 커넥터 서열을 통하여 연계된다.
- [0438] [94] 단락 [93]의 조성물, 이때 상기 제 1 다수의 핵산 분자에서 주어진 핵산 분자는 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 주어진 안티-커넥터 서열에 혼성화된 주어진 커넥터 서열을 통하여 혼성화된다.
- [0439] [95] 단락 [80]-[94]중 임의의 하나의 조성물, 이때 제 1 CDR3을 인코딩하는 서열과 제 2 CDR3을 인코딩하는 서열은 최대한 100개 뉴클레오티드에 의해 분리되어 있다.
- [0440] [96] 단락 [80]-[95]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.
- [0441] [97] 단락 [80]-[95]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.
- [0442] [98] 단락 [80]-[97]중 임의의 하나의 조성물, 이때 제 1 다수의 핵산 분자들에서 또는 제 2 다수의 분자들에서 각 핵산 분자는 화학적으로 합성된다.
- [0443] [99] 단락 [80]-[98]중 임의의 하나의 조성물, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자의 길이는 최대한 약 250 뉴클레오티드이다.
- [0444] [100] 핵산 분자들 다수를 포함하는 조성물, 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) V 유전자로부터 유래된 서열을 포함하며, 이때 상기 다수의 핵산 분자 제 1 커넥터 서열을 갖는 제 1 핵산 분자와 제 2 커넥터 서열을 갖는 제 2 핵산 분자를 포함하고, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하다.

- [0445] [101] 단락 [100]의 조성물, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0446] [102] 단락 [100] 또는 [101]의 조성물, 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 커넥터 서열을 포함한다.
- [0447] [103] 단락 [100]-[102]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 CDR3을 인코딩하는 서열을 포함하지 않는다.
- [0448] [104] 단락 [100]-[103]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 불변 도메인을 인코딩하는 서열을 포함하지 않는다.
- [0449] [105] 단락 [100]-[104]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 상기 TCR V 유전자의 적어도 10개 뉴클레오티드를 포함한다.
- [0450] [106] 단락 [100]-[105]중 임의의 하나의 조성물, 이때 상기 TCR V 유전자는 TRAV 유전자, TRBV 유전자, TRGV 유전자, 또는 TRDV 유전자이다.
- [0451] [107] 핵산 분자들 다수를 포함하는 조성물, 상기 다수 핵산 분자들중 각 핵산 분자는 T-세포 수용체 (TCR) 쉼의 CDR3을 인코딩하고, 이때 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 커넥터 서열을 포함하고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 커넥터 서열을 포함하고, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하다.
- [0452] [108] 단락 [107]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 J 영역을 더 포함한다.
- [0453] [109] 단락 [107]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쉼의 제 1 CDR3과 제 2 TCR 쉼의 제 2 CDR3을 인코딩한다.
- [0454] [110] 단락 [109]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쉼의 제 1 J 영역과 제 2 TCR 쉼의 제 2 J 영역을 더 포함한다.
- [0455] [111] 단락 [107]-[110]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 TCR 쉼의 상이한 CDR3을 인코딩한다.
- [0456] [112] 단락 [107]-[111]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 상이한 커넥터 서열을 포함한다.
- [0457] [113] 단락 [107]-[112]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR V 유전자의 200개를 넘는 뉴클레오티드를 포함하지 않는다.
- [0458] [114] 단락 [107]-[113]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 불변 도메인을 인코딩하는 서열을 포함하지 않는다.
- [0459] [115] 단락 [100]-[114]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0460] [116] 단락 [115]의 조성물, 이때 상기 TCR V 유전자로부터 유래된 서열은 재배열된 유전자에서 상기 CDR3을 인코딩하는 서열에 인접한 TCR V 유전자에서 적어도 3개 뉴클레오티드를 포함한다.
- [0461] [117] 단락 [100]-[116]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 사전-결정된 서열을 포함한다.
- [0462] [118] 단락 [107]-[114]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자 서열에 상보적인 서열을 포함한다.
- [0463] [119] 단락 [107]-[114] 및 [118]중 임의의 하나의 조성물, 이때 상기 조성물은 제 2 다수의 핵산 분자들을 더 포함하며, 제 2 다수의 핵산 분자들에서 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0464] [120] 단락 [119]의 조성물, 이때 제 2 다수 핵산 분자들에서 제 1 핵산 분자는 제 1 안티-커넥터 서열을 포함하고, 이때 제 1 안티-커넥터 서열은 제 1 커넥터 서열에 상보적이다.
- [0465] [121] 단락 [119] 또는 [120]의 조성물, 이때 제 2 다수 핵산 분자들에서 제 2 핵산 분자는 제 2 안티-커넥터

서열을 포함하고, 이때 제 2 안티-커넥터 서열은 제 2 커넥터 서열에 상보적이다.

- [0466] [122] 단락 [120] 또는 [121]의 조성물, 이때 제 2 다수 핵산 분자들에서 제 1 핵산 분자의 제 1 안티-커넥터 서열은 제 1 다수 핵산 분자들에서 제 1 핵산 분자의 제 1 커넥터 서열에 연계된다.
- [0467] [123] 단락 [121] 또는 [122]의 조성물, 이때 제 2 다수 핵산 분자들에서 제 2 핵산 분자의 제 2 안티-커넥터 서열은 제 1 다수 핵산 분자들에서 제 2 핵산 분자의 제 2 커넥터 서열에 연계된다.
- [0468] [124] 핵산 분자들 다수를 포함하는 조성물, 각 핵산 분자는 T-세포 수용체 (TCR) 쉼의 적어도 10개 아미노산을 인코딩하는 서열을 포함하고, 이때 상기 다수 핵산 분자들중 제 1 핵산 분자는 제 1 커넥터 서열을 포함하고, 상기 다수 핵산 분자들중 제 2 핵산 분자는 제 2 커넥터 서열을 포함하며, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이하고, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쉼의 일부분을 인코딩하고, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 상기 TCR 쉼의 적어도 10개 아미노산을 인코딩하는 서열과 프레임 안에 있다.
- [0469] [125] 단락 [124]의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR 쉼 유전자의 적어도 4개 인접 뉴클레오티드를 포함하고, 상기 TCR 쉼의 적어도 10개 아미노산을 인코딩하는 서열과 프레임 안에 있다.
- [0470] [126] 단락 [124] 또는 [125]의 조성물, 이때 제 1 커넥터 서열과 제 2 커넥터 서열은 TCR 쉼의 적어도 두 개의 인접 아미노산을 인코딩한다.
- [0471] [127] 단락 [124]-[126]중 임의의 하나의 조성물, 이때 상기 TCR 쉼의 일부분의 TCR 쉼과 적어도 10개 아미노산을 인코딩하는 서열에 의해 인코딩되는 TCR 쉼은 동일하다.
- [0472] [128] 단락 [124]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0473] [129] 단락 [124]-[128]중 임의의 하나의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 상기 TCR 쉼의 CDR3을 인코딩한다.
- [0474] [130] 단락 [129]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 TCR 쉼의 J 영역을 더 포함한다.
- [0475] [131] 단락 [129]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쉼의 제 1 CDR3과 제 2 TCR 쉼의 제 2 CDR3을 인코딩한다.
- [0476] [132] 단락 [131]의 조성물, 이때 상기 다수 핵산 분자들중 각 핵산 분자는 제 1 TCR 쉼의 제 1 J 영역과 제 2 TCR 쉼의 제 2 J 영역을 더 포함한다.
- [0477] [133] 단락 [131] 또는 [132]의 조성물, 이때 제 1 CDR3를 인코딩하는 서열과 제 2 CDR3을 인코딩하는 서열은 최대 100개 뉴클레오티드에 의해 분리되어 있다.
- [0478] [134] 단락 [124]-[133]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0479] [135] 단락 [124]-[134]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열 또는 제 2 커넥터 서열은 사전-결정된 서열을 포함한다.
- [0480] [136] 단락 [100]-[135]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열은 제 2 커넥터 서열의 뉴클레오티드와는 상이한 적어도 하나의 뉴클레오티드를 포함한다.
- [0481] [137] 단락 [100]-[136]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과 동일한 아미노산 서열을 인코딩한다.
- [0482] [138] 단락 [100]-[136]중 임의의 하나의 조성물, 이때 제 1 커넥터 서열은 제 2 커넥터 서열과는 상이한 아미노산 서열을 인코딩한다.
- [0483] [139] 핵산 분자들 다수를 생성시키는 방법, 이들 다수의 핵산 분자들중에서 각 핵산 분자는 T-세포 수용체 (TCR) 쉼 또는 이의 영역을 인코딩하고, 상기 방법은 다음을 포함한다: 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들을 접촉시켜, 적어도 두 개의 상이한 핵산 분자를 포함하는 제 3의 다수 핵산 분자를 생성시키고, 이때 상기 적어도 두 개의 상이한 핵산 분자 각각은 상이한 TCR 쉼 또는 이의 영역을 인코딩하는 상이한 서열을 갖고, 이때 상기 적어도 두 개의 상이한 핵산 분자는 동일한 구획 안에서 생성된다.

- [0484] [140] 단락 [139]의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 상기 TCR 쉼의 CDR3을 인코딩하는 서열을 포함한다.
- [0485] [141] 단락 [140]의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 TCR 쉼의 J 영역을 포함한다.
- [0486] [142] 단락 [139]-[141]중 임의의 하나의 방법, 이때 제 2 다수의 핵산 분자들에서 각 핵산 분자는 TCR 쉼의 TCR V 유전자로부터 유래된 서열을 포함한다.
- [0487] [143] 단락 [142]의 방법, 이때 상기 TCR V 유전자는 인간 TCR V 유전자이다.
- [0488] [144] 단락 [142] 또는 [143]의 방법, 이때 상기 TCR V 유전자는 인간 TRAV1-1, TRAV1-2, TRAV2, TRAV3, TRAV4, TRAV5, TRAV6, TRAV7, TRAV8-1, TRAV8-2, TRAV8-3, TRAV8-4, TRAV8-6, TRAV9-1, TRAV9-2, TRAV10, TRAV12-1, TRAV12-2, TRAV12-3, TRAV13-1, TRAV13-2, TRAV14, TRAV16, TRAV17, TRAV18, TRAV19, TRAV20, TRAV21, TRAV22, TRAV23, TRAV24, TRAV25, TRAV26-1, TRAV26-2, TRAV27, TRAV29, TRAV30, TRAV34, TRAV35, TRAV36, TRAV38-1, TRAV38-2, TRAV39, TRAV40, 또는 TRAV41이다.
- [0489] [145] 단락 [142] 또는 [143]의 방법, 이때 상기 TCR V 유전자는 인간 TRBV2, TRBV3-1, TRBV4-1, TRBV4-2, TRBV4-3, TRBV5-1, TRBV5-4, TRBV5-5, TRBV5-6, TRBV5-8, TRBV6-1, TRBV6-2, TRBV6-3, TRBV6-4, TRBV6-5, TRBV6-6, TRBV6-8, TRBV6-9, TRBV7-2, TRBV7-3, TRBV7-4, TRBV7-6, TRBV7-7, TRBV7-8, TRBV7-9, TRBV9, TRBV10-1, TRBV10-2, TRBV10-3, TRBV11-1, TRBV11-2, TRBV11-3, TRBV12-3, TRBV12-4, TRBV12-5, TRBV13, TRBV14, TRBV15, TRBV16, TRBV18, TRBV19, TRBV20-1, TRBV24-1, TRBV25-1, TRBV27, TRBV28, TRBV29-1, 또는 TRBV30이다.
- [0490] [146] 단락 [139]-[145]중 임의의 하나의 방법, 이때 상기 TCR V 유전자로부터 유래된 서열은 FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.
- [0491] [147] 단락 [139]-[145]중 임의의 하나의 방법, 이때 상기 TCR V 유전자로부터 유래된 서열은 L-PART1, L-PART2, FR1, CDR1, FR2, CDR2, 및 FR3을 인코딩하는 서열을 포함한다.
- [0492] [148] 단락 [139]-[147]중 임의의 하나의 방법, 이때 상기 TCR 쉼은 TCR 알파 쉼, TCR 베타 쉼, TCR 감마 쉼, 또는 TCR 델타 쉼이다.
- [0493] [149] 단락 [140]-[148]중 임의의 하나의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 추가 TCR 쉼의 추가 CDR3을 인코딩하는 추가 서열을 더 포함한다.
- [0494] [150] 단락 [149]의 방법, 이때 제 1 다수의 핵산 분자에서 각 핵산 분자는 가 TCR 쉼의 추가 J 영역을 포함한다.
- [0495] [151] 단락 [149] 또는 [150]의 방법, 이때 상기 TCR 쉼과 상기 추가 TCR 쉼은 TCR 쉼의 동계 쌍이다.
- [0496] [152] 단락 [139]-[151]중 임의의 하나의 방법, 이때 상기 다수 핵산 분자들중 핵산 분자는 상이한 TCR 또는 이의 영역을 인코드한다.
- [0497] [153] 단락 [139]-[152]중 임의의 하나의 방법, 이때 제 1 다수의 핵산 분자에서 주어진 핵산 분자 커넥터 서열을 포함하고, 이때 커넥터 서열은 제 1 다수의 핵산 분자에서 주어진 핵산 분자를 제 2 다수의 핵산 분자들에서 주어진 핵산 분자에 연계시키는데 이용가능하다.
- [0498] [154] 단락 [153]의 방법, 이때 상기 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 상기 제 2 다수의 핵산 분자들에서 주어진 핵산 분자는 기능성 TCR 쉼 또는 이의 영역을 인코드한다.
- [0499] [155] 단락 [153] 또는 [154]의 방법, 이때 상기 제 2 다수의 핵산 분자들에서 주어진 핵산 분자 안티-커넥터 서열을 포함하고, 이때 안티-커넥터 서열은 제 1 다수의 핵산 분자들에서 주어진 핵산 분자의 커넥터 서열에 상보적이다.
- [0500] [156] 단락 [153]-[155]중 임의의 하나의 방법, 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 연계시키는 것을 더 포함한다.
- [0501] [157] 단락 [156]의 방법, 이때 연계는 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자의 혼성화를 포함한다.
- [0502] [158] 단락 [157]의 방법, 이때 혼성화는 제 1 다수의 핵산 분자에서 주어진 핵산 분자의 커넥터 서열을 제 2

다수의 핵산 분자들에서 주어진 핵산 분자의 상기 안티-커넥터 서열에 혼성화시키는 것을 포함한다.

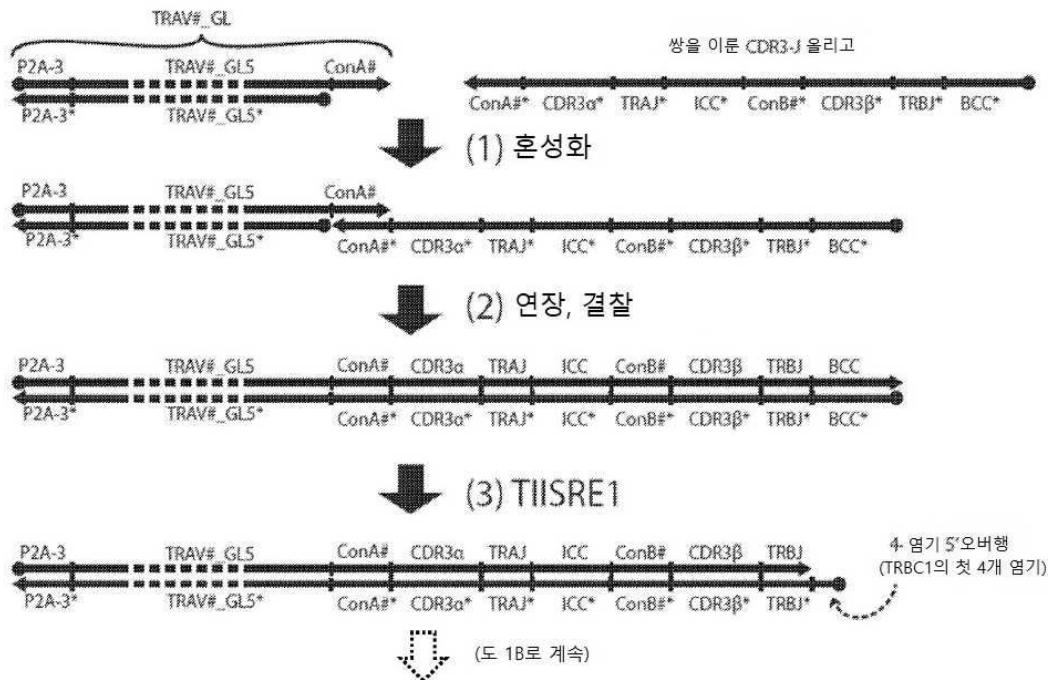
- [0503] [159] 단락 [156]-[158]중 임의의 하나의 방법, (i) 제 1 다수의 핵산 분자에서 주어진 핵산 분자를 주형으로 하여 제 2 다수의 핵산 분자들에서 주어진 핵산 분자의 자유 3' 단부를 연장시키거나, 및/또는 (ii) 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 주형으로 하여 제 1 다수의 핵산 분자들에서 핵산 분자의 자유 3' 단부를 연장시켜, 제 3 다수의 핵산 분자들의 핵산 분자를 생성시키는 것을 더 포함한다.
- [0504] [160] 단락 [139]-[159]중 임의의 하나의 방법, 제 1 다수의 핵산 분자에서 주어진 핵산 분자와 제 2 다수의 핵산 분자들에서 주어진 핵산 분자를 결합시키는 것을 더 포함한다.
- [0505] [161] 단락 [139]-[160]중 임의의 하나의 방법, 상기 제 3 다수의 핵산 분자들의 핵산 분자에 제한절단 효소를 접촉시켜 스티커 단부를 만드는 것을 더 포함한다.
- [0506] [162] 단락 [139]-[161]중 임의의 하나의 방법, 상기 제 3 다수의 핵산 분자들의 핵산 분자에 추가 핵산 분자를 접촉시키는 것을 더 포함한다.
- [0507] [163] 단락 [162]의 방법, 이때 상기 추가 핵산 분자는 TCR 쉘의 불변 영역 또는 이의 일부분을 인코딩한다.
- [0508] [164] 단락 [162] 또는 [163]의 방법, 상기 제 3 다수의 핵산 분자들의 핵산 분자와 상기 추가 핵산 분자를 결합시키는 것을 더 포함한다.
- [0509] [165] 단락 [139]-[164]중 임의의 하나의 방법, 이때 상기 제 3 다수의 핵산 분자의 적어도 5개 상이한 핵산 분자가 상기 동일한 구획 안에서 생성된다.
- [0510] [166] 단락 [139]-[165]중 임의의 하나의 방법, 이때 상기 제 3 다수의 핵산 분자의 적어도 10개 상이한 핵산 분자가 상기 동일한 구획 안에서 생성된다.
- [0511] [167] 단락 [139]-[166]중 임의의 하나의 방법, 이때 상기 동일한 구획은 웰, 튜브, 또는 소적이다.
- [0512] [168] 핵산 분자들 다수를 생성시키는, 다음을 포함하는 방법: (a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 1 T-세포 수용체 (TCR) 쉘의 제 1 CDR3 및 제 2 TCR 쉘의 제 2 CDR3을 인코딩하는 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 TCR 쉘의 동계 쌍으로부터 유래되고; (b) 제 2 다수의 핵산 분자를 제공하고, 이때 제 2 다수 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고 (c) 제 1 다수 핵산분자들에서 핵산 분자와 제 2 다수 핵산 분자들에서 핵산 분자를 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 선형 핵산 분자를 형성하게 되고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 TCR V 유전자는 TCR 쉘의 동계 쌍으로부터 유래된다.
- [0513] [169] 핵산 분자들 다수를 생성시키는, 다음을 포함하는 방법: (a) 제 1 다수의 핵산 분자를 제공하고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 (i) 제 1 T-세포 수용체 (TCR) 쉘의 제 1 CDR3과 제 2 TCR 쉘의 제 2 CDR3을 인코딩하는 합성 서열, (ii) 제 3 T-세포 수용체 (TCR) 쉘의 제 3의 CDR3과 제 4의 TCR 쉘의 제 4의 CDR3을 인코딩하는 합성 서열을 포함하고, 이때 제 1 CDR3 및 제 2 CDR3은 제 1 TCR 쉘의 동계 쌍으로부터 유래되며, 이때 제 3의 CDR3 및 제 4의 CDR3은 제 2 TCR 쉘의 동계 쌍으로부터 유래되며; (b) 제 2 다수의 핵산 분자들을 제공하고, 이때 제 2 다수의 핵산 분자들에서 핵산 분자는 TCR V 유전자로부터 유래된 서열을 포함하고; 그리고 (c) 제 1 다수의 핵산 분자와 제 2 다수의 핵산 분자들을 접촉시키고, 이때 제 1 다수의 핵산 분자들에서 핵산 분자는 제 2 다수의 핵산 분자들에서 핵산 분자와 연계되어 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자로부터 유래된 서열을 포함하는 핵산 분자를 형성하고, 이때 제 1 CDR3 및 제 2 CDR3을 인코딩하는 서열과 상기 TCR V 유전자는 TCR 쉘의 동계 쌍으로부터 유래된다.
- [0514] [170] 대상체로부터 취한 조직 샘플 안에 태생적으로 쌍을 이룬 T-세포 수용체 (TCR)의 서열을 식별해내는, 다음을 포함하는 방법: (a) 해당 대상체로부터 획득한 다수의 말초 T 세포들을 함유하는 샘플에서 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 하나 또는 그 이상의 쌍을 이룬 서열을 식별해내고, 이때 하나 또는 그 이상의 쌍을 이룬 서열 각각은 CDR3 서열을 포함하고; 그리고 (b) 태생적으로 쌍을 이룬 다른 TCR 쉘가 알려지지 않은 조직 샘플에서 TCR의 TCR 쉘의 조직 CDR3 서열을 식별해내고, 이때 이 조직 CDR3 서열은 상기 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs의 하나 또는 그 이상의 쌍을 이룬 서열에서 적어도 하나의 쌍을 이룬 서열의 CDR3 서열과 일치하며, 그렇게 함으로써 상기 적어도 하나의 쌍을 이룬 서열은 해당 조직 샘플에서 태생적

으로 쌍을 이룬 TCR의 서열로 확인된다.

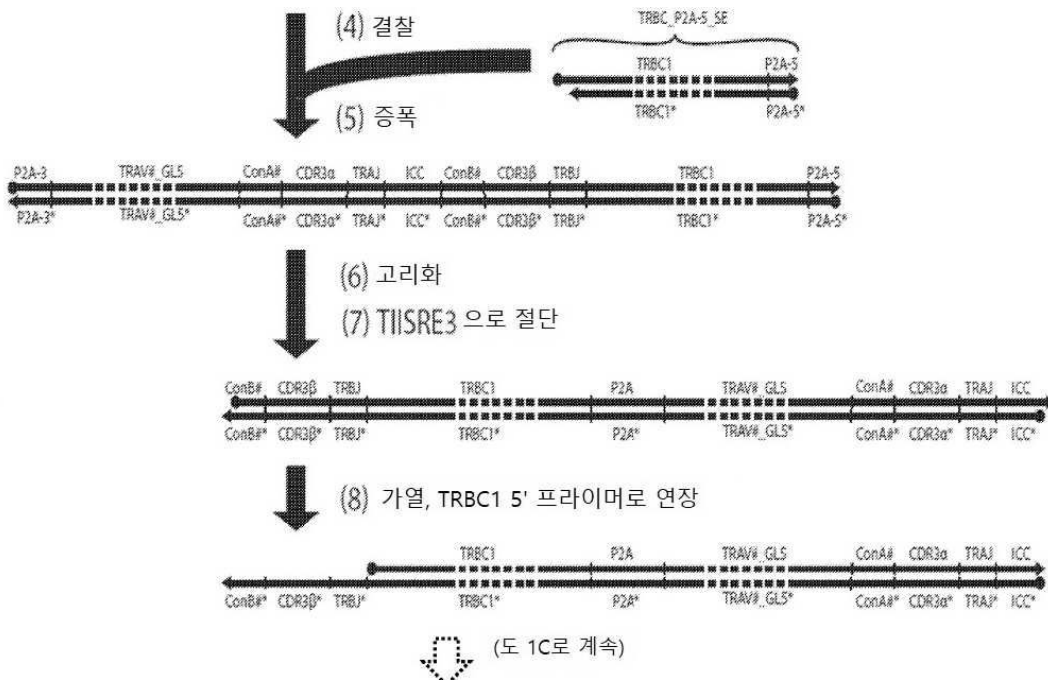
- [0515] [171] 단락 [170]의 방법, 이때 (a)에서 식별은 다수의 말초 T 세포들을 함유하는 샘플에서 하나 또는 그 이상의 태생적으로 쌍을 이룬 TCRs을 시퀀싱하는 것을 포함한다.
- [0516] [172] 단락 [171]의 방법, 이때 상기 시퀀싱은 단일 세포 시퀀싱을 포함한다.
- [0517] [173] 단락 [172]의 방법, 이때 상기 단일 세포 시퀀싱은 다수의 말초 T 세포들을 다수의 구획으로 분배하는 것을 포함하고, 각 구획은 다수의 말초 T 세포들에서 개별 말초 T 세포를 포함한다.
- [0518] [174] 단락 [170]-[173]중 임의의 하나의 방법, 이때 해당 조직 샘플은 체액 샘플이 아니다.
- [0519] [175] 단락 [170]-[174]중 임의의 하나의 방법, 이때 해당 조직 샘플은 고형 종양 샘플이다.
- [0520] [176] 단락 [170]-[175]중 임의의 하나의 방법, 이때 해당 조직 샘플은 고정된 샘플 또는 냉동된 샘플이다.
- [0521] [177] 단락 [170]-[176]중 임의의 하나의 방법, 이때 다수의 말초 T 세포들을 함유하는 샘플은 말초 혈액 단핵 세포 (PBMC) 샘플이다.
- [0522] [178] 단락 [170]-[177]중 임의의 하나의 방법, (a)의 실시예 앞서, 해당 대상체로부터 혈액 샘플을 획득하는 것을 더 포함한다.
- [0523] [179] 단락 [178]의 방법, (a)의 실시예 앞서, 해당 혈액 샘플로부터 말초 혈액 단핵 세포들을 분리시키는 것을 더 포함한다.
- [0524] [180] 단락 [170]-[179]중 임의의 하나의 방법, 이때 해당 조직 샘플은 종양-침윤성 T 세포를 포함한다.
- [0525] [181] 표적-반응성 T-세포 수용체 (TCR)를 식별해내는, 다음을 포함하는 방법: (a) 단락 [170]-[180]중 임의의 하나에서 식별된 TCR을 포함하는 세포를 제공하고; 그리고 (b) 상기 세포를 항원-제시 세포 (APC)에 의해 제시된 표적 항원에 접촉시키고, 이때 해당 세포는 TCR을 통하여 해당 APC에 의해 제공되는 표적 항원에 결합하고, 그렇게 함으로써 상기 TCR은 표적-반응성 TCR로 확인된다.
- [0526] [182] 단락 [181]의 방법, 이때 상기 표적 항원은 종양 항원이다.
- [0527] [183] 단락 [181] 또는 [182]의 방법, 표적-반응성 TCR을 인코딩하는 서열을 숙주 세포로 운반하는 것을 더 포함한다.
- [0528] [184] 단락 [183]의 방법, 해당 숙주 세포를 해당 대상체에게 투여하는 것을 더 포함한다.
- [0529] [185] 단락 [183] 또는 [184]의 방법, 이때 상기 숙주 세포는 T 세포이다.
- [0530] [186] 단락 [185]의 방법, 이때 상기 T 세포는 자가유래 T 세포이다.
- [0531] [187] 단락 [185]의 방법, 이때 상기 T 세포는 동종이계 T 세포이다.
- [0532] [188] 단락 [181]-[187]중 임의의 하나의 방법, 이때 상기 세포는 리포터 세포 계통이며, 이때 리포터 세포 계통은 APC에 의해 제시되는 표적 항원에 해당 세포가 결합할 때 발현되는 리포터 유전자를 포함한다.

도면

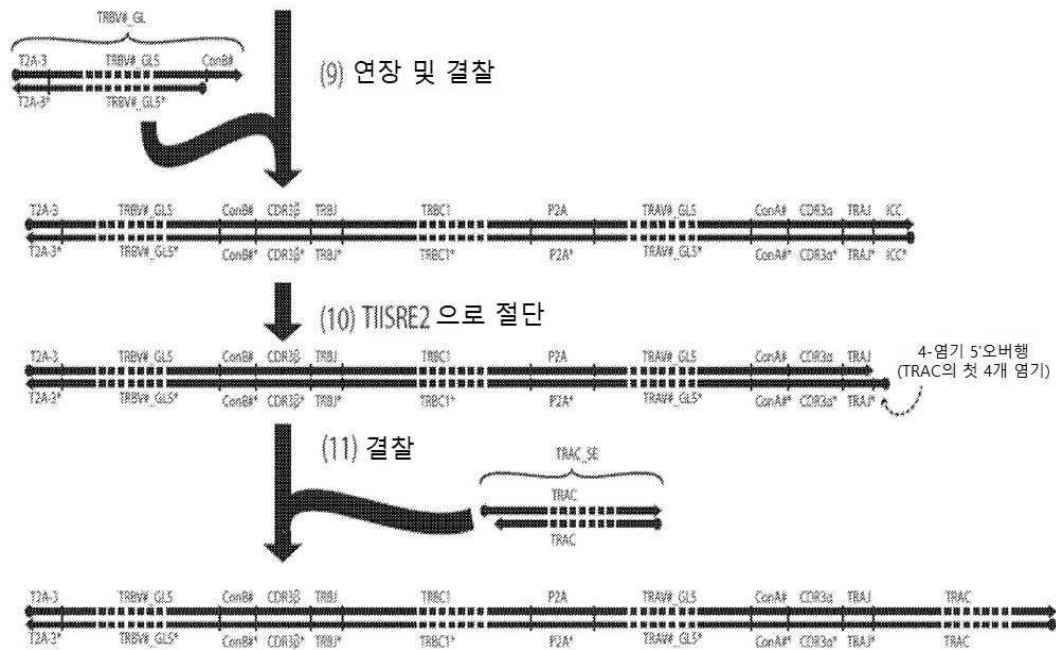
도면1a



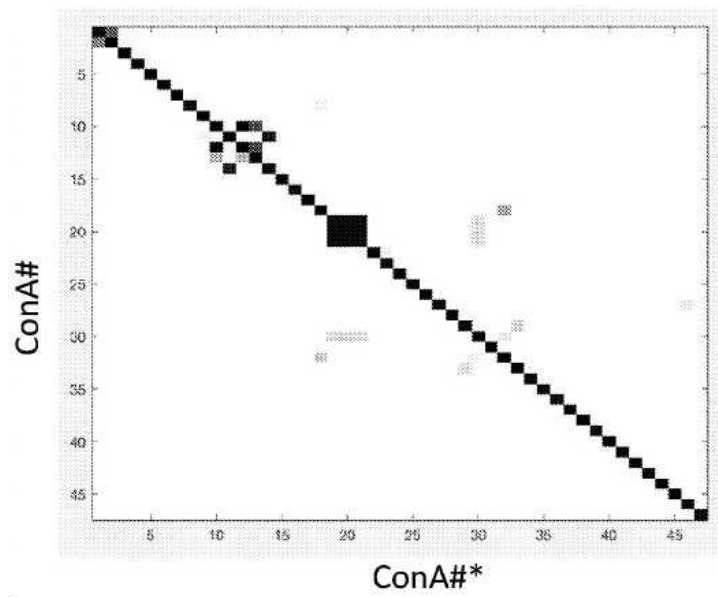
도면1b



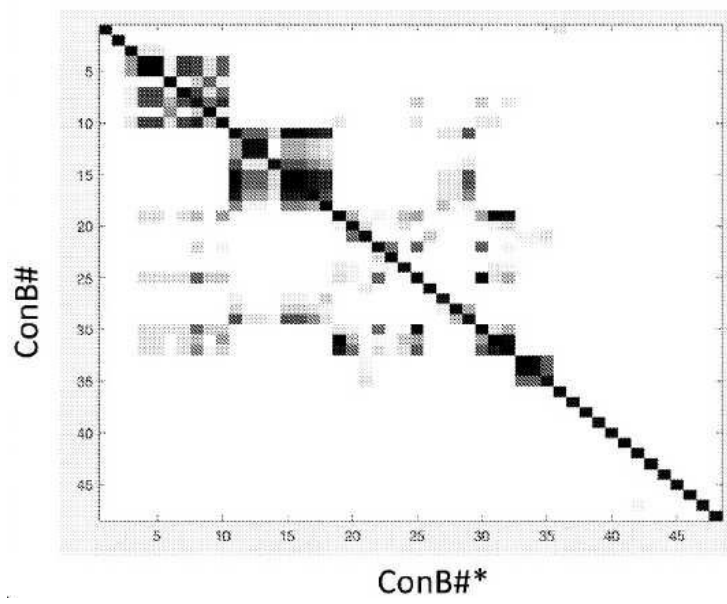
도면1c



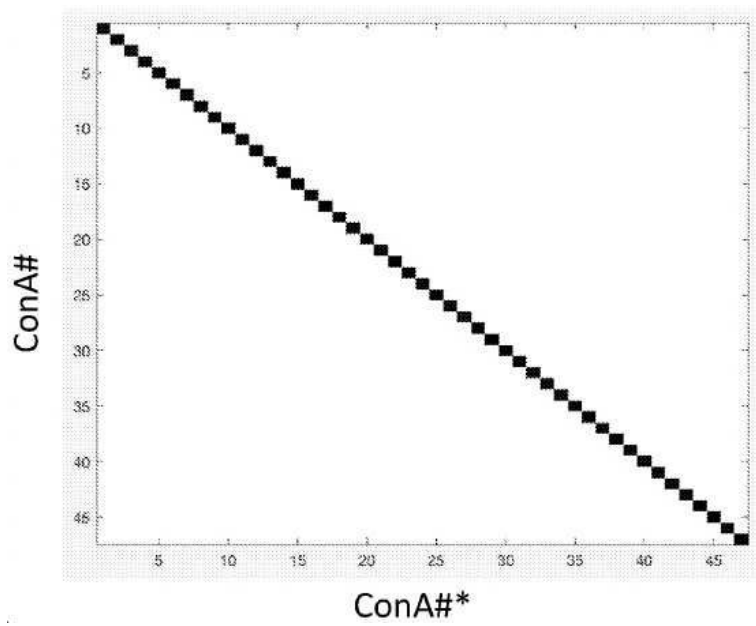
도면2a



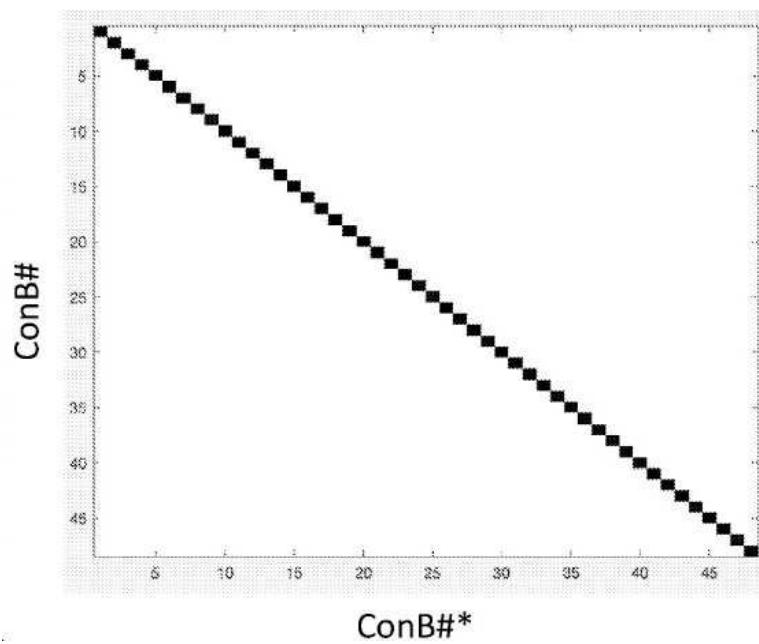
도면2b



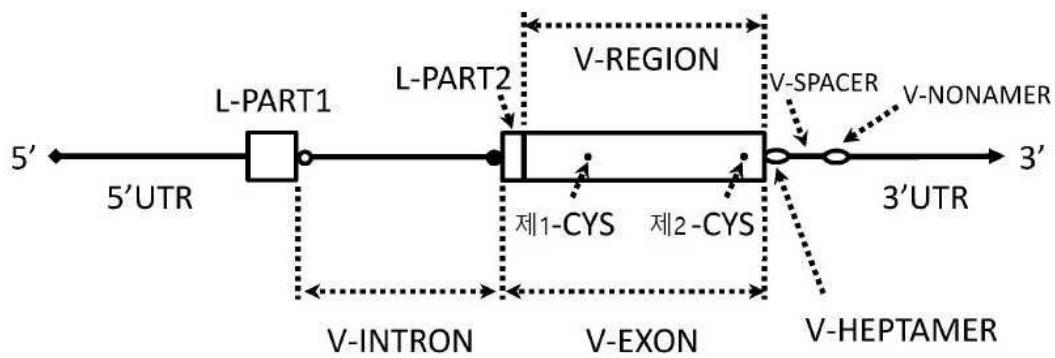
도면3a



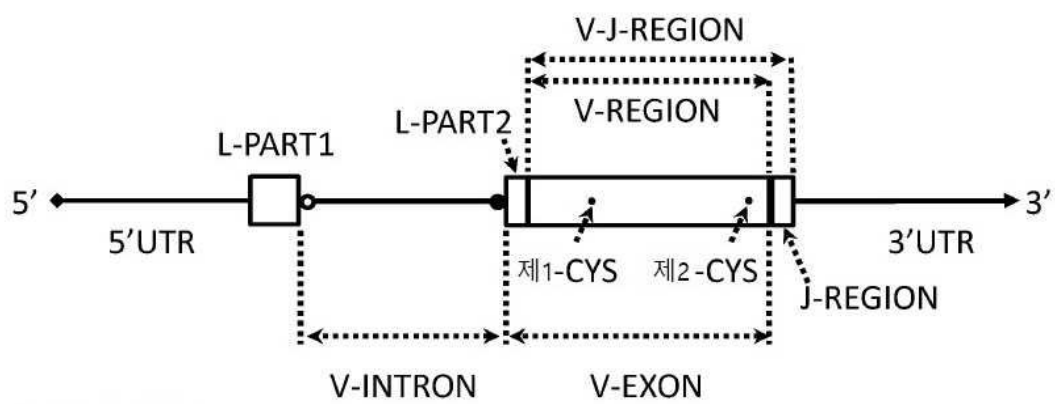
도면3b



도면4a



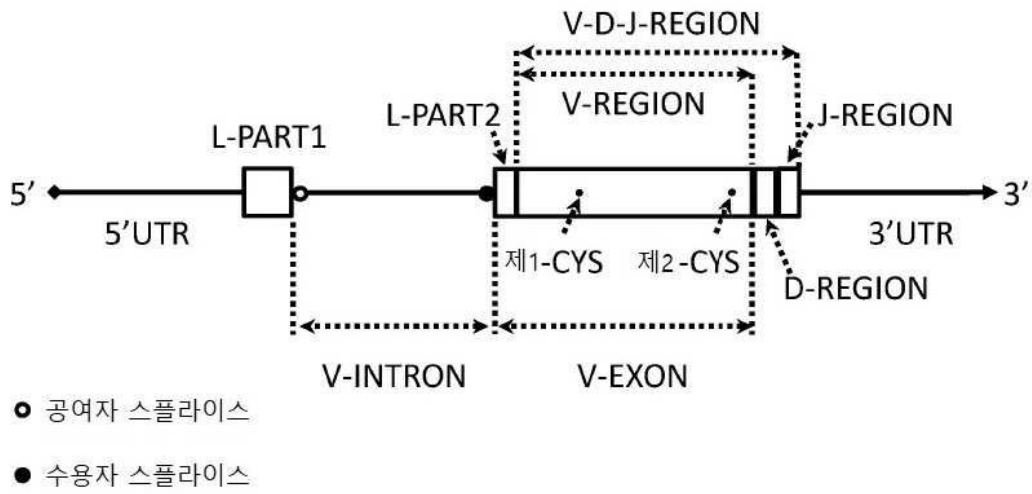
도면4b



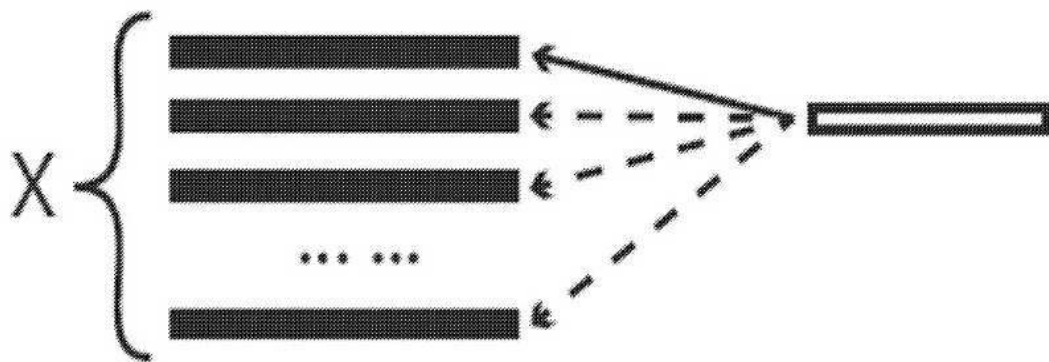
○ 공여자 스플라이스

● 수용자 스플라이스

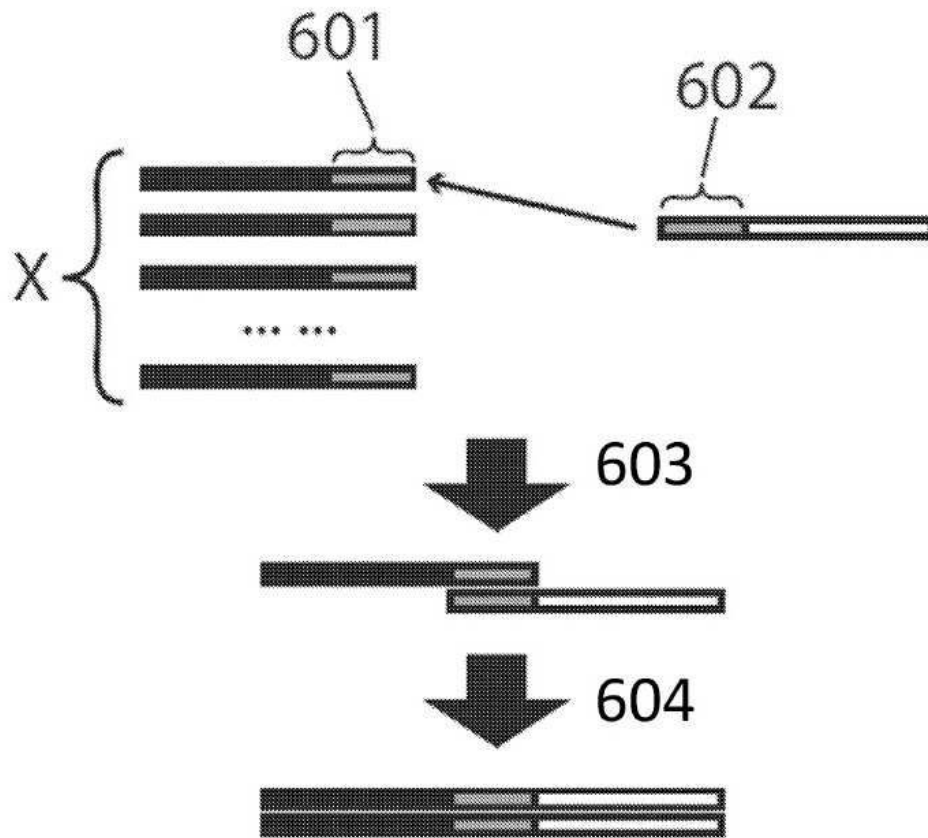
도면4c



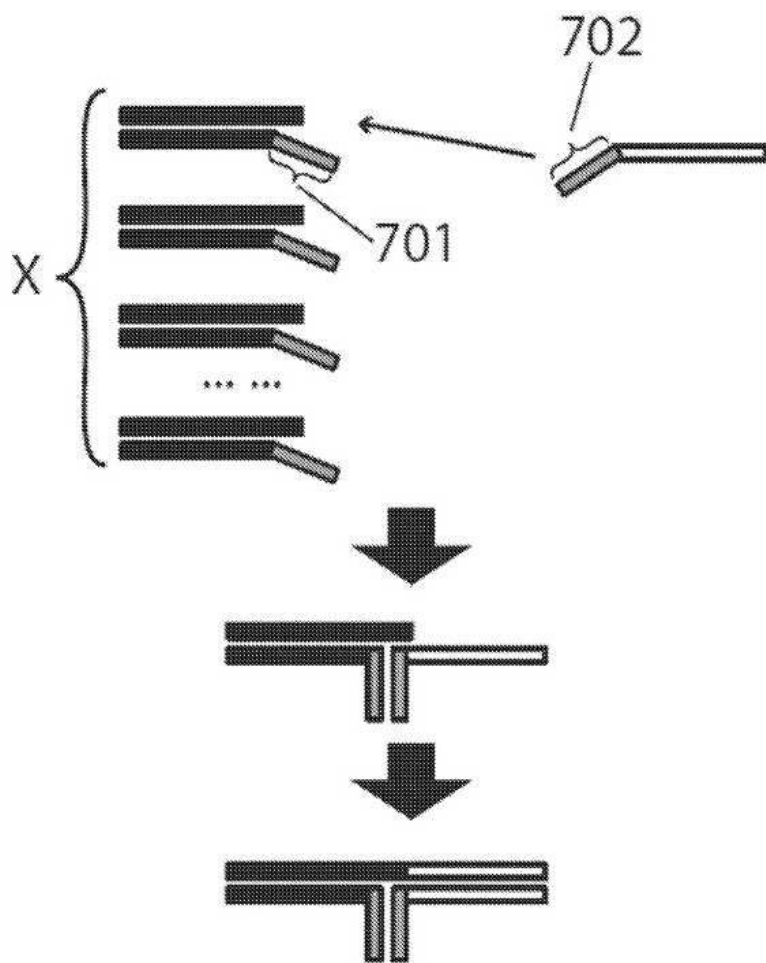
도면5



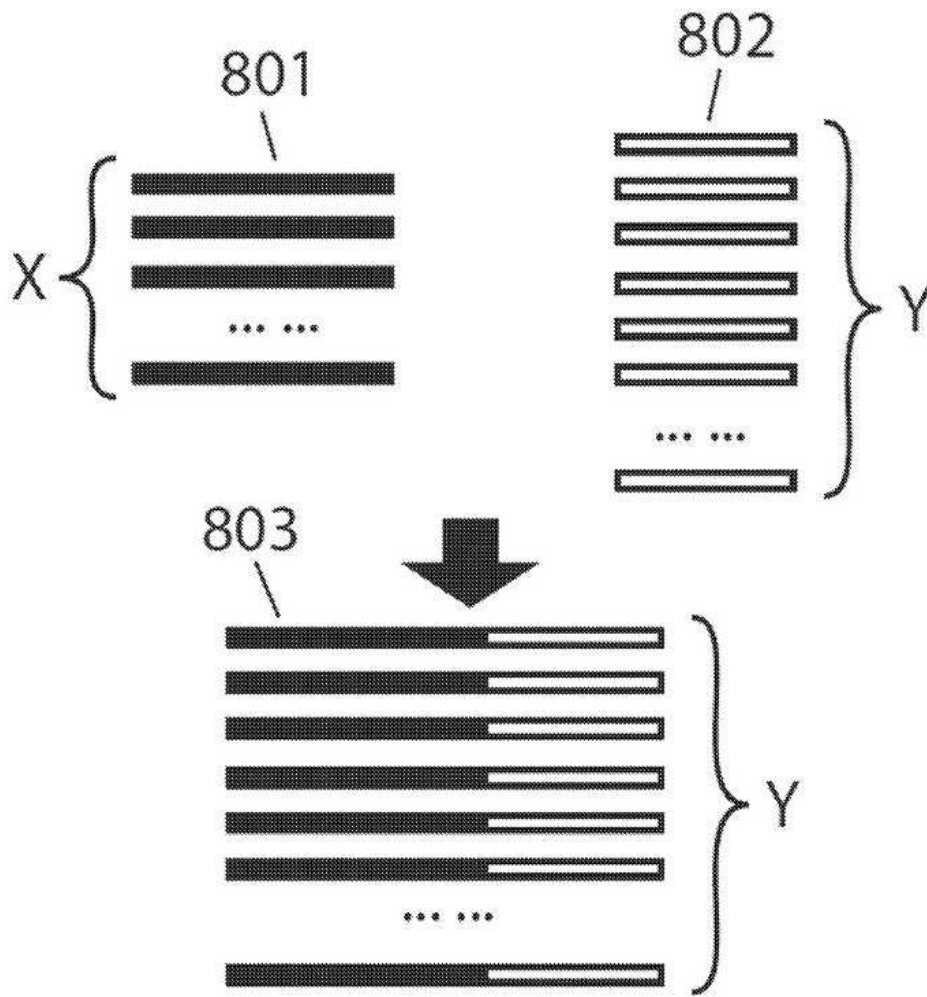
도면6



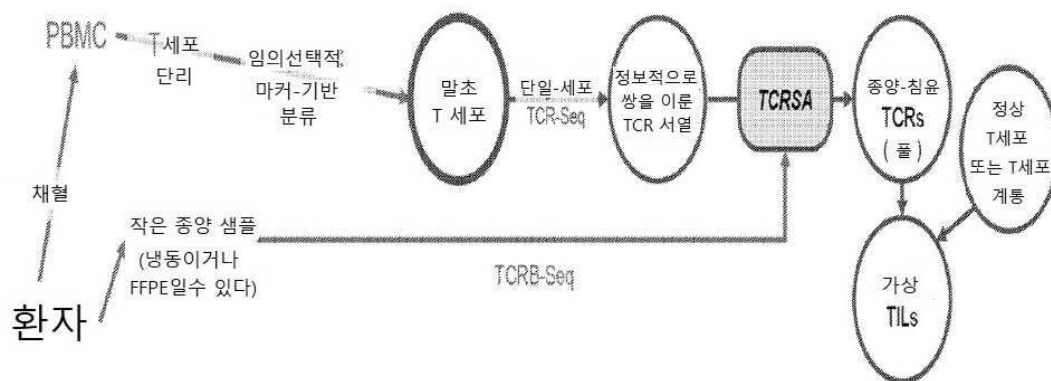
도면7



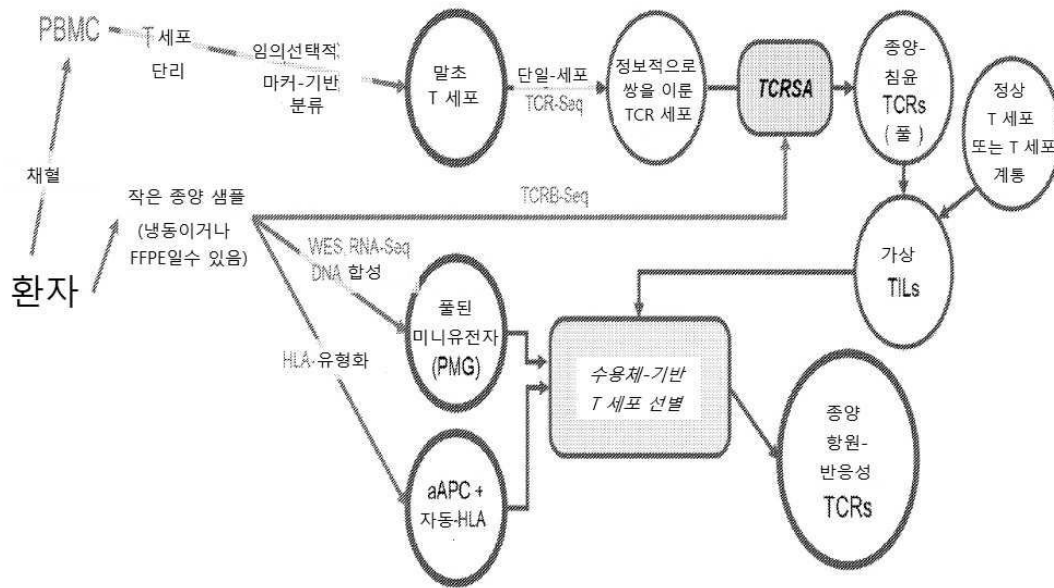
도면8



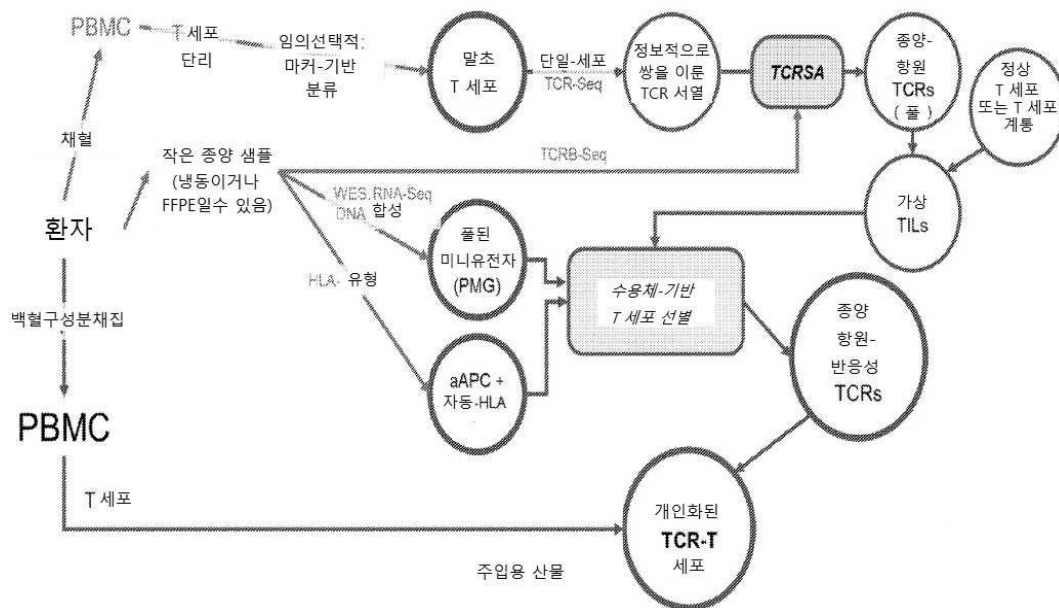
도면9a



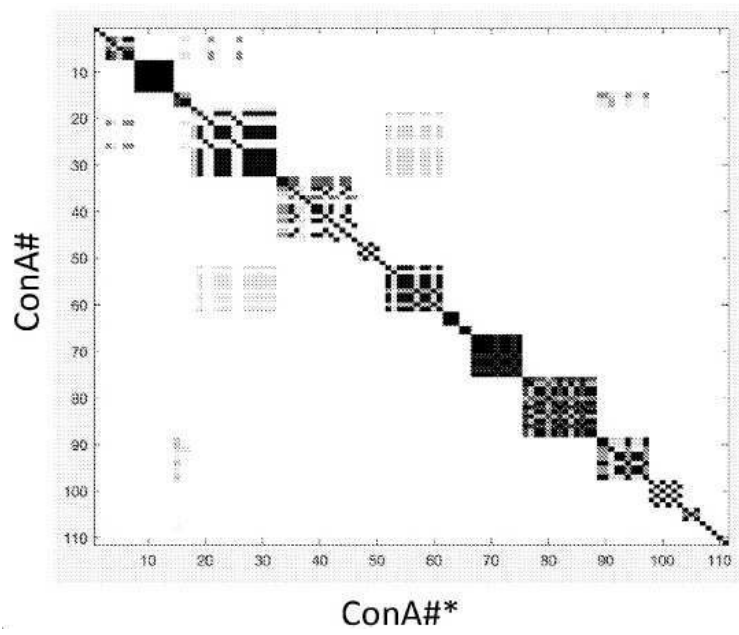
도면9b



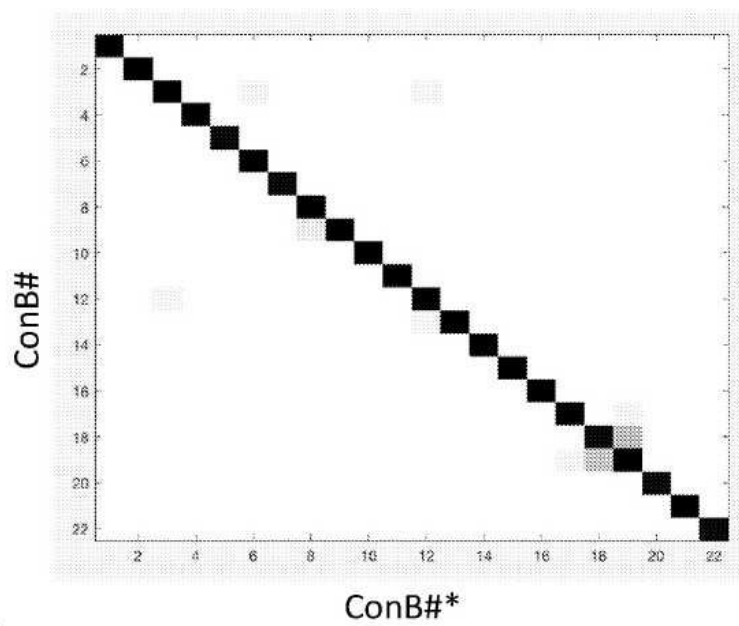
도면9c



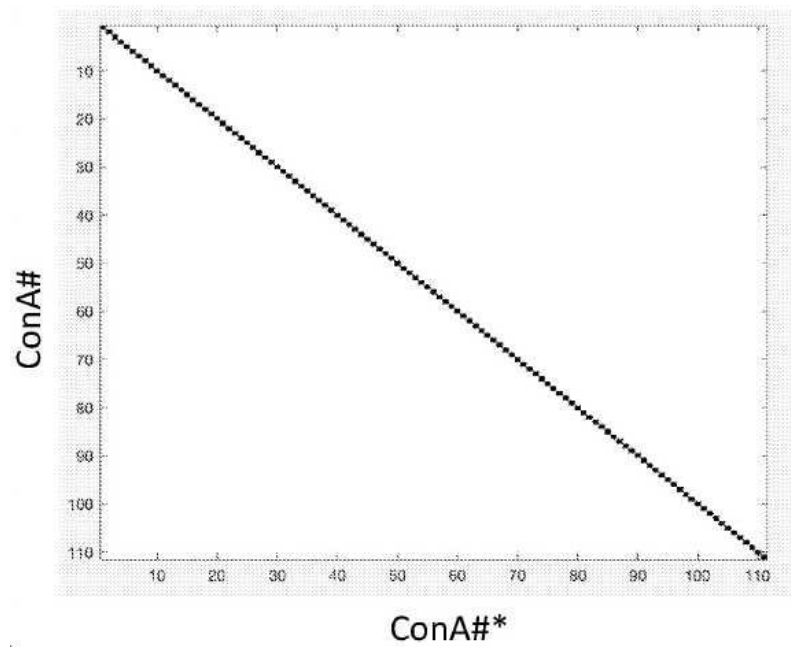
도면10a



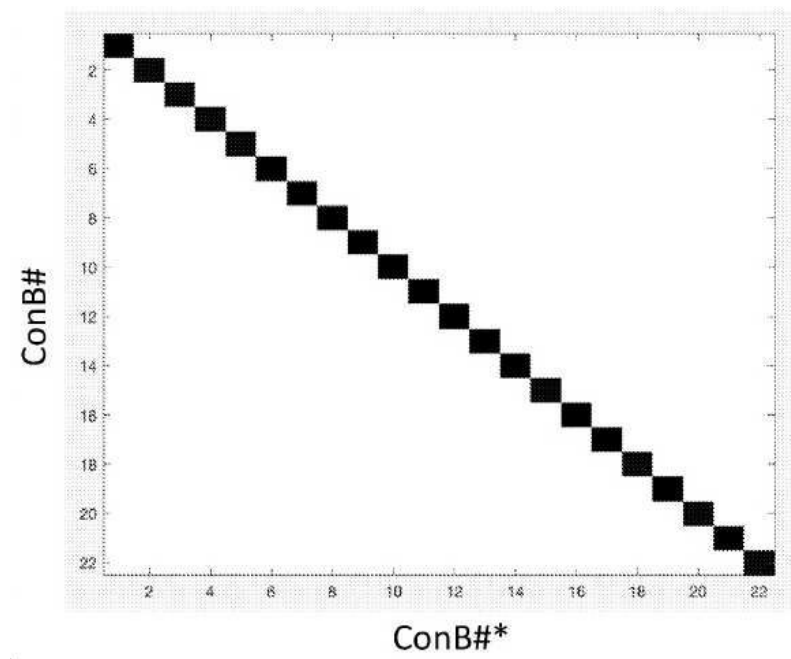
도면10b



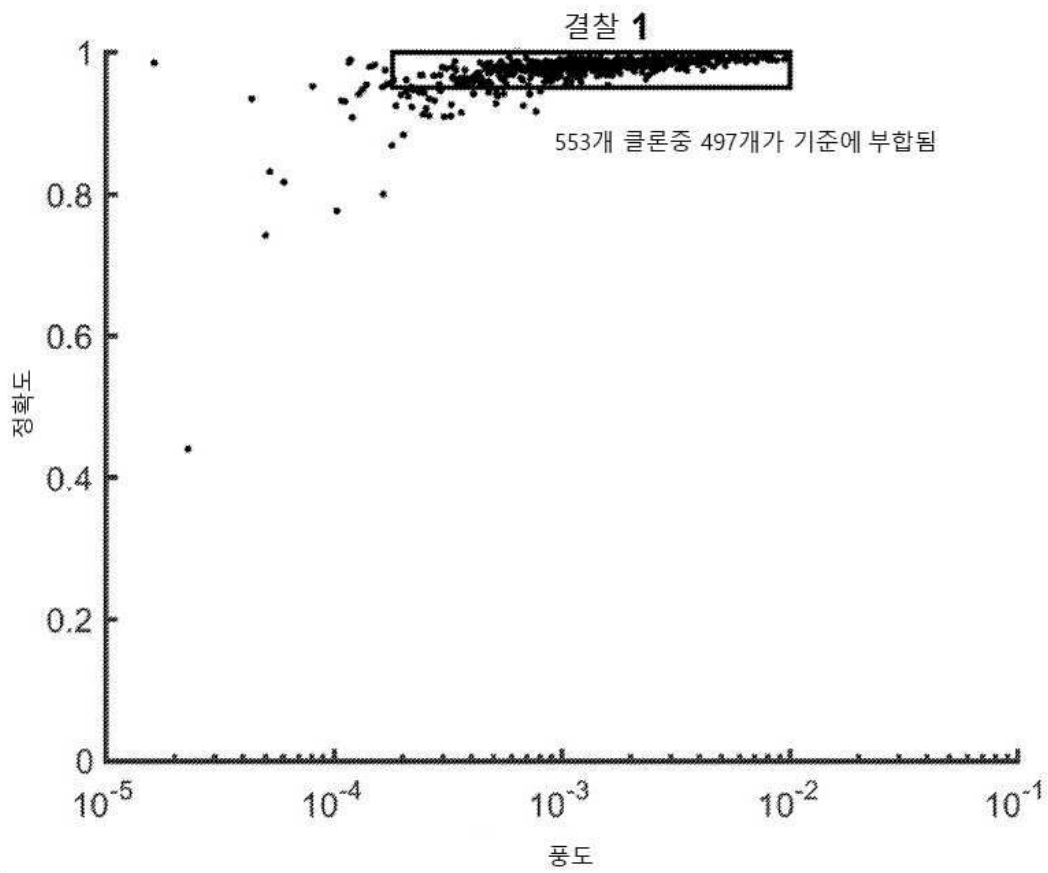
도면11a



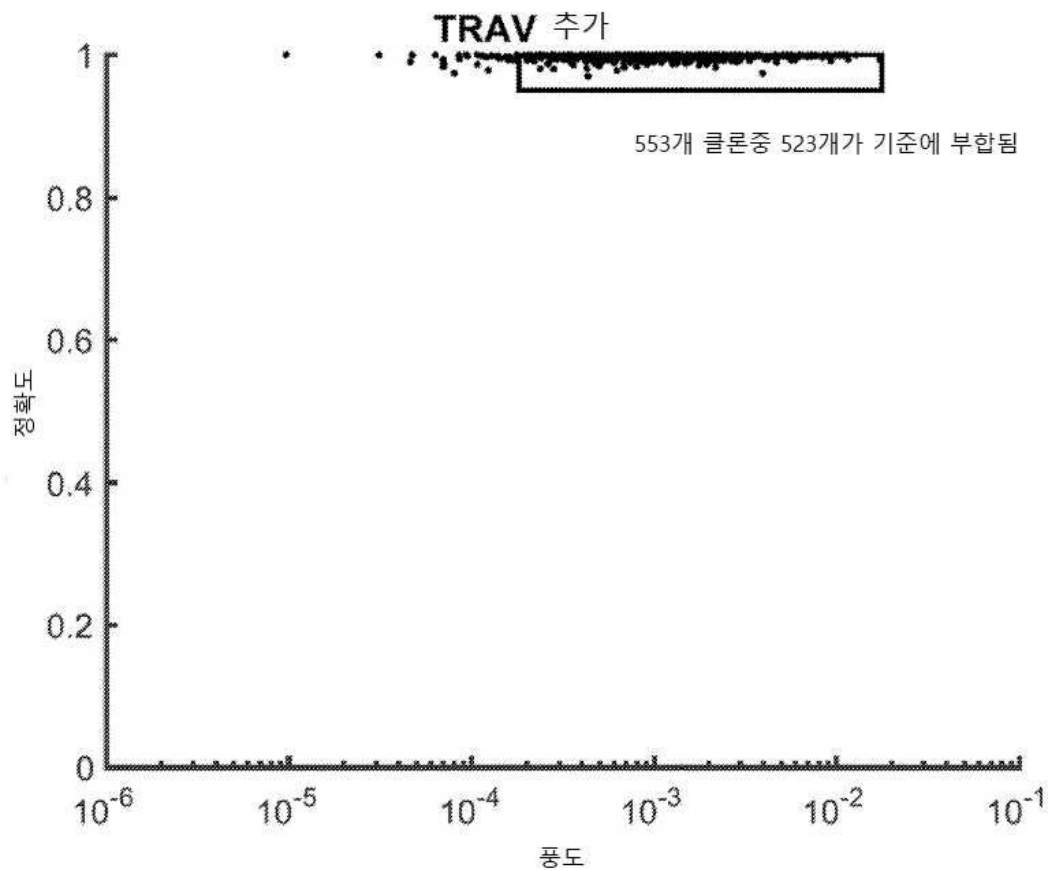
도면11b



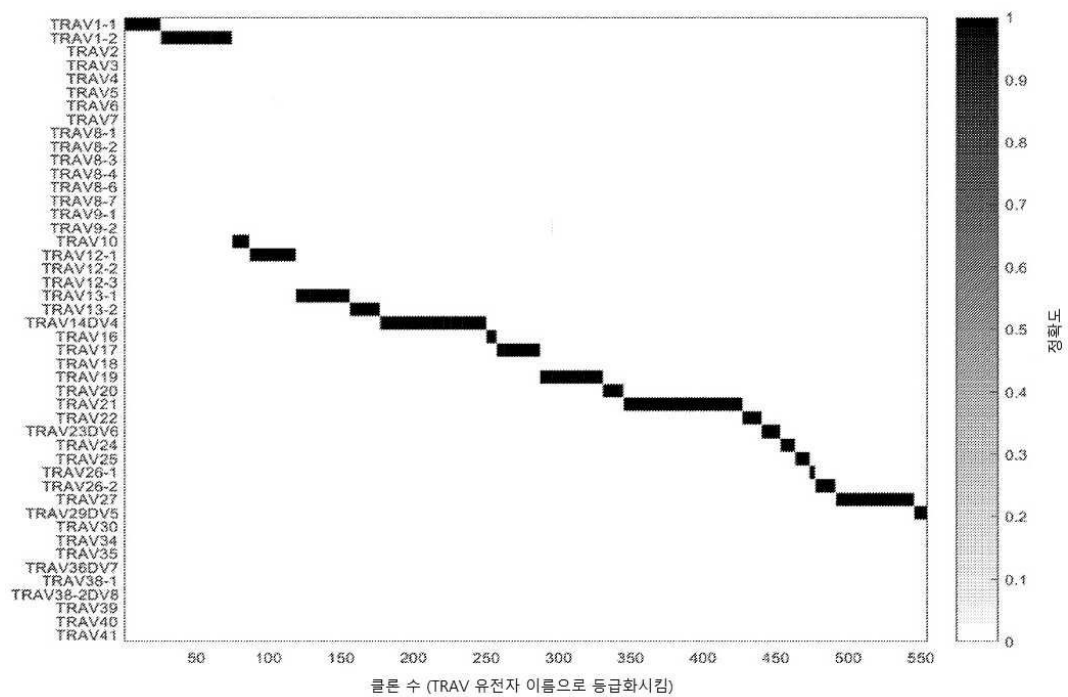
도면12



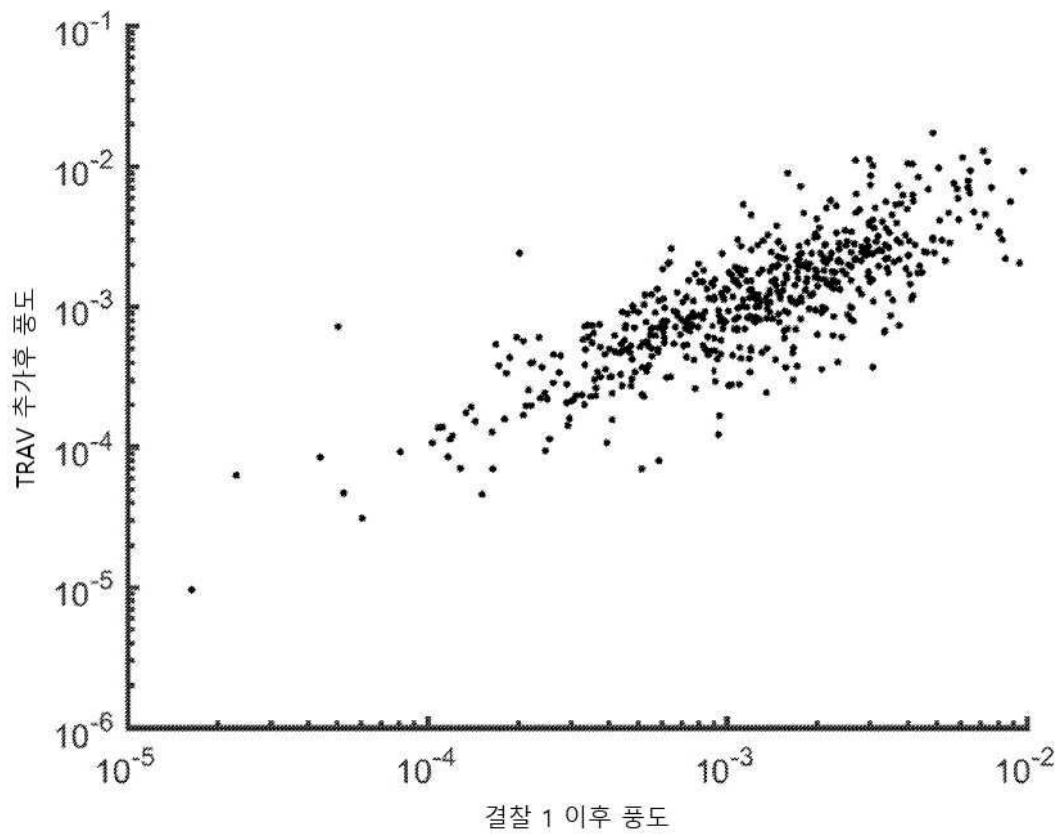
도면 13



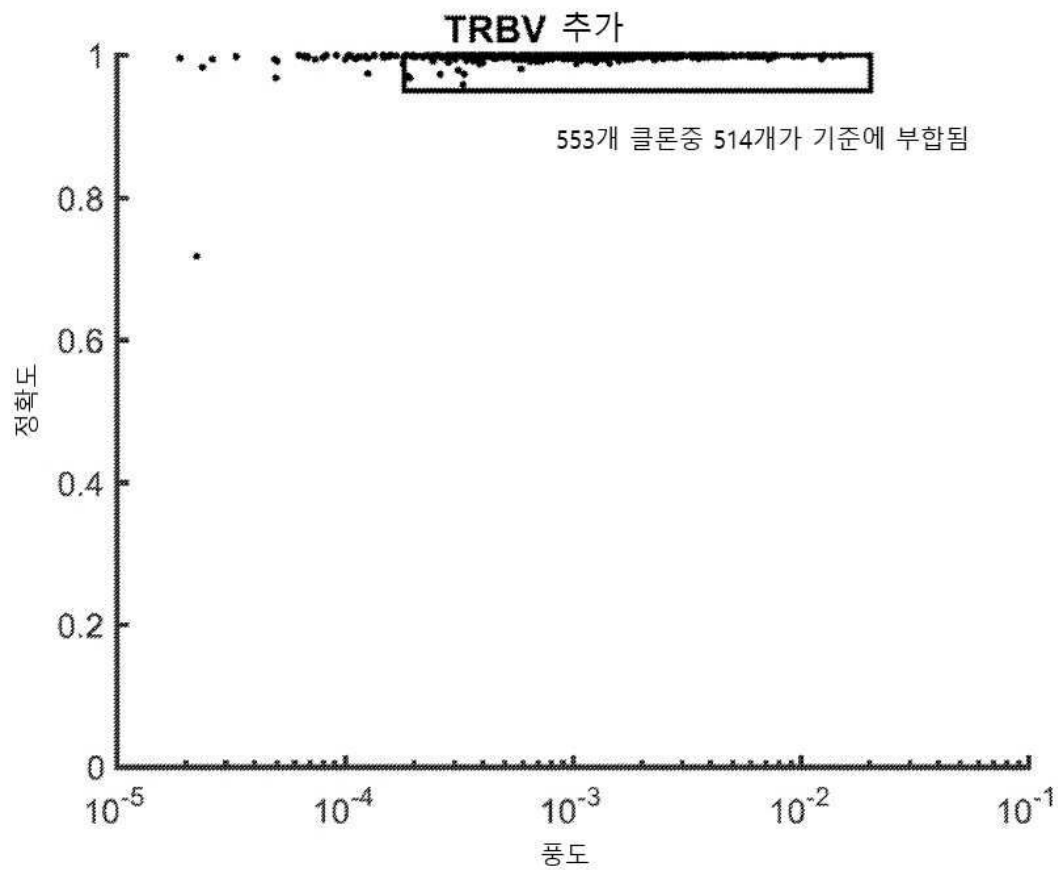
도면14



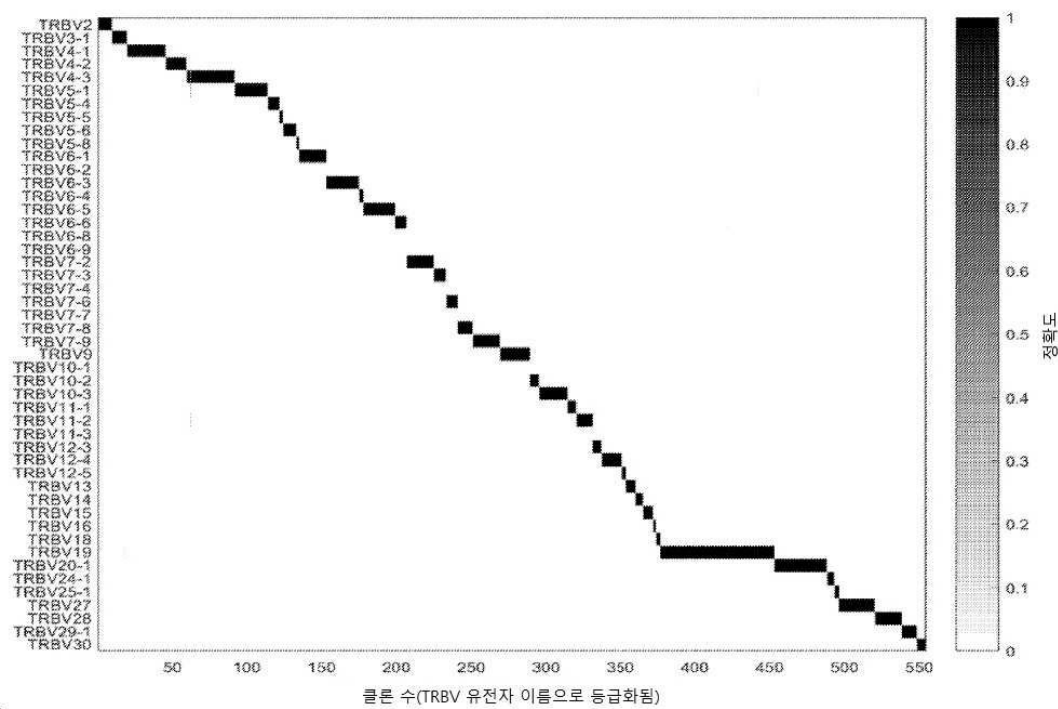
도면15



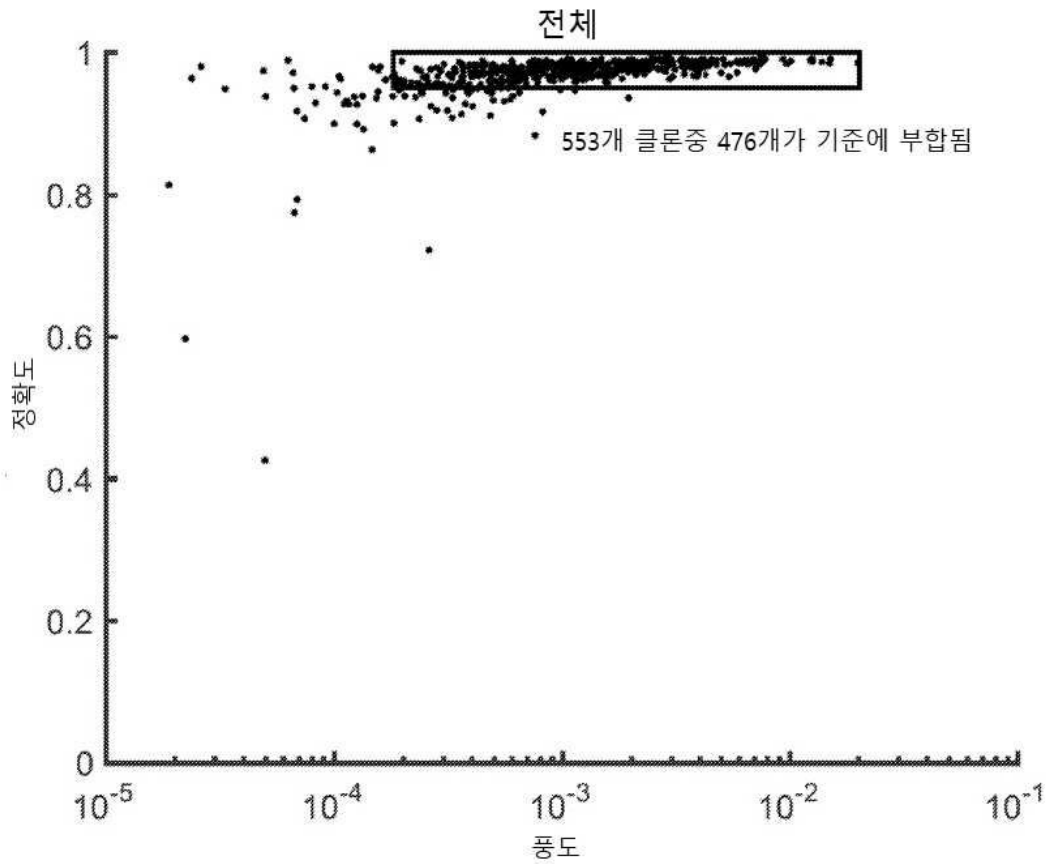
도면16



도면17



도면18



서열 목록

SEQUENCE LISTING

<110> ROOTPATH GENOMICS, INC.

<120> COMPOSITIONS AND METHODS FOR T-CELL RECEPTOR GENE ASSEMBLY

<130> 53563-705.601

<140> PCT/US2020/026558

<141> 2020-04-03

<150> 62/972,231

<151> 2020-02-10

<150> 62/898,053

<151> 2019-09-10

<150> 62/838,465

<151> 2019-04-25

<150> 62/829,813

<151> 2019-04-05

<160> 292

<170> PatentIn version 3.5

<210> 1

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 1

tcaaagactc tgcctcatatc ctct 24

<210> 2

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 2

ctgagagacg cagctgtgta ttact 25

<210> 3

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 3

ggggactcag ccgtgtactt ct 22

<210> 4

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 4

ggtgactccg cagcctatct ct 22

<210> 5

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 5

tggggatagc gcagtctatt tct

23

<210> 6

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 6

ccggagacag cgcagtttat tttt

24

<210> 7

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 7

ggtgacagcg ccgtctatatt tt

22

<210> 8

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 8

tggaggactc aggcacttac ttct

24

<210> 9

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 9

tggaggactc tgggacatac tttt 24

<210> 10

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 10

tggaggactc tggcacatat tttt 24

<210> 11

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 11

actcgaggat tccgtactt atttct 26

<210> 12

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 12

gaagactccg ggacctactt tt 22

<210> 13

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 13

gctggaggat tccggaacct atttct

26

<210> 14

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 14

ctcgaagata gcggcacata ttttt

25

<210> 15

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 15

gcctggtgat agcgcaatat acttct

26

<210> 16

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 16

tggcgactct gcaatgtact tct

23

<210> 17

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 17
cccggagact ctgctatgta ttttt 25

<210> 18
<211> 26
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 18
ggaatccgat agcgagctct attact 26

<210> 19
<211> 22
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 19
tccgacagcg ctgtctacta ct 22

<210> 20
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 20
aagagattga tagcgctgtt tactact 27

<210> 21
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 21

aggaatctga ttccgcagtc tattttt 27

<210> 22

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 22

agaatctgat agcgccgttt attatt 26

<210> 23

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 23

gtccgactcc gcagtctact act 23

<210> 24

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 24

aggagtctga ttctgcagtc tactatt 27

<210> 25

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 25
ccaagaaata gattccgcag tctactact 29

<210> 26
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 26
gtctgacagc gcagtctact tct 23

<210> 27
<211> 27
<212> DNA
<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 27
aggaaagcga ttctgcagtc tattact 27

<210> 28
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 28
aagagtctga ctccgcagtt tattatt 27

<210> 29
<211> 26
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 29

agaatccgac tctgcagttt actatt 26

<210> 30

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 30

gagtctgata gcgctgtgta ctact 25

<210> 31

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 31

gaatctgact ctgccgttta ctatt 25

<210> 32

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 32

agtccgactc tgctgtgtac tact 24

<210> 33

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 33

ccatctgatt ccgcactgta tttct 25

<210> 34

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 34

accttctgat agcgctctct attttt

26

<210> 35

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 35

ccttctgatt ctgcactgta cctgt

25

<210> 36

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 36

ccaagcgatt ctgcactgta ttttt

25

<210> 37

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 37

cctctgactc tgcagtctac ctct

24

<210> 38

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 38

cccagcgcact ctgcagttta tctct 25

<210> 39

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 39

ccgacagcgc actctacctg t 21

<210> 40

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 40

ttccgactct gcactgtatc tgt 23

<210> 41

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 41

tccgatagcg ccctgtattt ct 22

<210> 42

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 42

ctccgattcc gcactctatc tct 23

<210> 43

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 43

cctccgatag cgctgtttat ctct 24

<210> 44

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 44

gcgacagcgc cctgtacttt t 21

<210> 45

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 45

ccctctgata gcgcactgta tctct 25

<210> 46

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 46

cttctgacag cgctgtgtat ctgt

24

<210> 47

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 47

gcgaggacac agctgtttac tttt

24

<210> 48

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 48

agtgccaaga tacagcagtt tacttct

27

<210> 49

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 49

cgggtgtgagg atactgctgt ttatttct

28

<210> 50

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 50

cgaagatacc gccgtctact ttt 23

<210> 51

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 51

ggcaactgac acagcagtct actttt 26

<210> 52

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 52

gagcgattct gccgtttact tct 23

<210> 53

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 53

tccgattccg ccgtgtatTT tt 22

<210> 54

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 54

ttggtctgat tctgcagttt actttt	26
<210> 55	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 55	
ggtctgattc cgtgtctac tttt	24
<210> 56	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 56	
ctggtctgac tctgtgttt atttt	26
<210> 57	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 57	
ggtccgactg ggcagtctat tttt	24
<210> 58	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 58	
tggtctgatt ctgccgtcta ttct	25

<210> 59
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 59
 agcgactctg ccgtgtatct ct 22

<210> 60
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 60
 gagcgattgg gcagtctact ttt 23

<210> 61
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 61
 tgggtccgatt ctgctgtcta ttttt 25

<210> 62
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 62
 agcctgaaga ttcagccatc tacttct 27

<210> 63

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 63

cccgaggact ctgctattta cttct 25

<210> 64

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 64

acagccagaa gattctgcaa tatacttct 29

<210> 65

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 65

gctcgatgac acagctacat acatct 26

<210> 66

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 66

tcctggatga tactgcaaca tacatat 27

<210> 67

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 67

ctctctgact ctgcactgta ctact 25

<210> 68

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 68

actgtctgac tctgcactct attact 26

<210> 69

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 69

ctgtccgatt ctgcactcta ctact 25

<210> 70

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 70

aactgtctga ttctgctctg tactatt 27

<210> 71

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 71

tccgactccg ctctgtat tt 22

<210> 72

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 72

agcgactctg ccctctacta ct 22

<210> 73

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 73

tctctgactc cgtctctac tact 24

<210> 74

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 74

tctctgattc tgccctctac tttt 24

<210> 75

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 75
 gctctccgat tctgctctgt attatt 26
 <210> 76
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 76
 gacaacagac tcaggcactt atctct 26
 <210> 77
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 77
 aacaactgac tctggcacat attttt 26
 <210> 78
 <211> 23
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 78
 cactgatagc ggaacctacc tct 23
 <210> 79
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 79
 tgacagcggc acctacctgt 20
 <210> 80
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 80
 actacagatt ccggcactta cttct 25
 <210> 81
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 81
 gccagataac tgattctggc acttacctgt 30
 <210> 82
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 82
 acaactgaca gcggaacata tctct 25
 <210> 83
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 83

aaataacaga tagcgggtaca tacctgt	27
<210> 84	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 84	
ccacagattc tggcacctac ttct	24
<210> 85	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 85	
actgactccg gaacctacct ct	22
<210> 86	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 86	
accgactctg gcacttacct gt	22
<210> 87	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 87	
aatcacagac tctggaacct atctgt	26

<210> 88
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 88
 ccaaattacc gattctggta catacctct 29
 <210> 89
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 89
 cggagatagc gccacatact ttt 23
 <210> 90
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 90
 aacctggaga ttctgcaaca tatttct 27
 <210> 91
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 91
 ctggggactc tgcaacttac ttct 24
 <210> 92
 <211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 92

cctggagact cagctaccta cttct

25

<210> 93

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 93

ccggggatag cgctacttat tttt

24

<210> 94

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 94

cctggagatt ccgcaactta ctttt

25

<210> 95

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 95

ccaggggatt ctgctaccta ttttt

25

<210> 96

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 96

cccggagatt ctgccactta ttct 25

<210> 97

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 97

ctggcgacag cgctacttat ttct 24

<210> 98

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 98

aaccagacga ttcgggaaag tatttct 27

<210> 99

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 99

cagaggattc agggacgtac ttct 24

<210> 100

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 100
ccagacgact ccggaaagta ctttt 25
<210> 101
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 101
ccgaggactc cggtacatac ttct 24
<210> 102
<211> 27
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 102
aaccgatga ctctggttaag tatTTTT 27
<210> 103
<211> 26
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 103
gccagaagac tccgtacat attttt 26
<210> 104
<211> 29
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 104

tcaaattgaa gattctgcag tctactttt 29

<210> 105

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 105

gattgaggac tcggcagtat atttct 26

<210> 106

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 106

aaatcgaaga ctctgcagtt tactttt 27

<210> 107

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 107

gagcgactca gccagttact tct 23

<210> 108

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 108

aggggatgct gggatctact ttt	23
<210> 109	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 109	
cccgaagata cagctgtcta cctgt	25
<210> 110	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 110	
agggacgcag cagtctatca tt	22
<210> 111	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 111	
gccactctgc catctacttc tgt	23
<210> 112	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 112	
ggcgcacact gtactgcaca t	21

<210> 113

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 113

tgatgactcg gccacatact tct 23

<210> 114

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 114

tggaggactc agctgtgtac ttct 24

<210> 115

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 115

accagaagat agcgagttt atctgt 26

<210> 116

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 116

atccagaaga ctcagctgtc tatTTTT 27

<210> 117

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 117

tcgaagatag cgccatgtac tttt 24

<210> 118

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 118

actggaagat agcgtgtgt atttct 26

<210> 119

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 119

aagccagacc agcctctatt ttt 23

<210> 120

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 120

cccctctcag acatcagtg acttct 26

<210> 121

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 121

gccagaccgc cgtgtatttc t 21

<210> 122

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 122

ggcgacacag ccacctatct ct 22

<210> 123

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 123

gcctaaagac agcgctgttt atctct 26

<210> 124

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 124

ccaggactca gcggtgtatc ttt 23

<210> 125

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 125

gcctagagta ttctgccatg tacctct

27

<210> 126

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 126

aaaaatgaga tggcagtcctt cctct

25

<210> 127

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 127

cgaggatagg ggccgtatc tct

23

<210> 128

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 128

gcagaagact cagcactgta cttgt

25

<210> 129

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 129

gacgactcag cactgtacct ct

22

<210> 130

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 130

ggggactccg cactctatct ct

22

<210> 131

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 131

aaacaaacca gacatctgtg tacttct

27

<210> 132

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 132

ggcctggaga cagcagtatc tatttct

27

<210> 133

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 133

tcagccatag cggtttttac ctct 24

<210> 134

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 134

ccgggatttt gtgactcatc 20

<210> 135

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 135

gaggatcgta tgtttcgac 20

<210> 136

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 136

cttgtgtgca cttaccgtac 20

<210> 137

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 137
 tgatgcatct ccagtacagg 20

<210> 138
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 138
 cttctgtgtg tacctcgaca 20

<210> 139
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 139
 cttgcaatcc ttaccgtgc 20

<210> 140
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 140
 gcaagtgtgg aaaatgaccc 20

<210> 141
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 141

gagtctagtc tcacaacca 20

<210> 142

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 142

gaaatgttga ggactccacg 20

<210> 143

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 143

cctaacagat gctacgtgga 20

<210> 144

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 144

gtaggtccac acagattcca 20

<210> 145

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 145

gccagtcaca gcaaatacac 20

<210> 146

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 146

ccgctaccag tatgtacctt

20

<210> 147

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 147

actgtgttcc ttgtcttccg

20

<210> 148

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 148

tgaatgcatc tacggtaccg

20

<210> 149

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 149

gcgcttatca atcttgctcg

20

<210> 150

<211> 20

<212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 150
 cggatcaattc agtagccact 20

<210> 151
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 151
 ggacacatgt acactagcca 20

<210> 152
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 152
 tgggagctct acgaaaatcc 20

<210> 153
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 153
 gttctcgaga tcgtcacaca 20

<210> 154
 <211> 20
 <212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 154

cttctgcatt cgatccttgc 20

<210> 155

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 155

gcagagttgt gtgattggag 20

<210> 156

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 156

cccatcaatt cggaaccatc 20

<210> 157

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 157

atcgtaaccc aagtctgtgg 20

<210> 158

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 158

ggcgaaatga tccctgaatg

20

<210> 159

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 159

agtgcacaga actttcaggc

20

<210> 160

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 160

gatcgtaac tctttgggag

20

<210> 161

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 161

acacgaggat tgctgtagag

20

<210> 162

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 162

ttctaccaca ttgtctcggg 20

<210> 163

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 163

gaatggctaa actgtgtccc 20

<210> 164

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 164

cggactgtac gagaaactga 20

<210> 165

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 165

cttgcgacaa actactcctg 20

<210> 166

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 166

ccgttttact ttgtcgccac	20
<210> 167	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 167	
tgatgatat cacttcggcg	20
<210> 168	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 168	
ccaacctcta tatgtgcca	20
<210> 169	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 169	
tgaacaggta tgctccagag	20
<210> 170	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 170	
ttgtggatat cgtctgtcc	20

<210> 171
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 171
 ctgtggaact cgactcttgt 20

<210> 172
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 172
 ctccaggatg cacaattcg 20

<210> 173
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 173
 ggctcatgac aaaacacagg 20

<210> 174
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 174
 ccgaatccga aaacaacacg 20

<210> 175

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 175

agacctaaca ctgtgatgcg 20

<210> 176

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 176

cctgggtgag cataaacttc 20

<210> 177

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 177

gagtcttgga cgaacaaagg 20

<210> 178

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 178

cacgtaccca tcatgttctg 20

<210> 179

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 179

ccgtgttagt caagtgtgtg 20

<210> 180

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 180

ctggtggcat aaatggaacg 20

<210> 181

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 181

tggatgtggg tatcaatggg 20

<210> 182

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 182

tgtggctaac gtaggacaag 20

<210> 183

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 183

ccctcgttgt gaaaatgtgc

20

<210> 184

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 184

tcgtcatagg tcagcttacg

20

<210> 185

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 185

cctgatgacc tctatgcca

20

<210> 186

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 186

cggcaagaat gaatagggtg

20

<210> 187

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 187
 gtgctattgg tgggaaatgg 20
 <210> 188
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 188
 gccatgtttg cttactgacg 20
 <210> 189
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 189
 cgttggtggca ttcattagcg 20
 <210> 190
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 190
 gcggtaggat tggatctcat 20
 <210> 191
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 191
cctcgcaaag ctgttatgac 20
<210> 192
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 192
gccttcatgt tattggacgc 20
<210> 193
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 193
agctgtagtg ttcttgaggc 20
<210> 194
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 194
gtagtggttc gtgtgacatg 20
<210> 195
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 195

cgcggcataat gttcataatcc 20

<210> 196

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 196

gagactggat catgcaacag 20

<210> 197

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 197

cacaacttct ctggactcca 20

<210> 198

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 198

cgaccatgat ctgtatgcgt 20

<210> 199

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 199

ggtgtgactc ttgtttccgt 20

<210> 200
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 200
 acgtacatac aagtcctggcg 20
 <210> 201
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 201
 cctcaaggat tcactcgcaa 20
 <210> 202
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 202
 ctgtatagga tgtccacgca 20
 <210> 203
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 203
 gcctgtgatt ggtaaagcg 20
 <210> 204
 <211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 204

cgactcgta gcatctagaa

20

<210> 205

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 205

cgatttggtg tccctagctg

20

<210> 206

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 206

cccacttcat ctgactctga

20

<210> 207

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 207

cggcattgta caggtgttac

20

<210> 208

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 208

tctcctat t t c c c t g a a c g g 20

<210> 209

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 209

c t t g t c c a c t a a a c g c a a c g 20

<210> 210

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 210

c g g g t a t c a c t g g g t a a t g a 20

<210> 211

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 211

g g a a c a g a g a c c a a t c c a g t 20

<210> 212

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 212

gtgggcatcc gaaatttcag

20

<210> 213

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 213

cgcgacgaca ttaccaatag

20

<210> 214

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 214

cgtcggaata tgctctcaga

20

<210> 215

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 215

gatgcagatc aatgagtggc

20

<210> 216

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 216

actgcttaca agtgtccacg 20

<210> 217

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 217

actgtatgca agctagtccc 20

<210> 218

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 218

agatctccca aaagtgtccg 20

<210> 219

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 219

ttccagaacc atgtgatccc 20

<210> 220

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 220

gccttgcttt tcaacctctg	20
<210> 221	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 221	
gatacggatc ttcacatgcg	20
<210> 222	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 222	
cgctcatcta ggttgacta	20
<210> 223	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 223	
cgcggttcaga ttccaaacag	20
<210> 224	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 224	
gcctggttac acatgctatc	20

<210> 225

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 225

gcaaaggtcc tacagtttc 20

<210> 226

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 226

ggctttccat gtctatgctc 20

<210> 227

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 227

gcgacattag cagagtaggt 20

<210> 228

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 228

ctcgccatac tatctgcatg 20

<210> 229

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 229

ctactgaaca ctggcaagc 20

<210> 230

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 230

ctgttcaatt cctgtgcgag 20

<210> 231

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 231

cactgagatg gaatttggcg 20

<210> 232

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 232

gtccatcaca acttccactg 20

<210> 233

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 233

ggctcagtct actttgcttc 20

<210> 234

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 234

gtagcttcc gacacaatgg 20

<210> 235

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 235

ccgtgacaca ctttcatttc 20

<210> 236

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 236

ccgggatgtc attatgagca 20

<210> 237

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 237

gagtgtccta cgagatcaga 20

<210> 238

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 238

taacgtctct ctgagtgtgg 20

<210> 239

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 239

accctagaca agagacacct 20

<210> 240

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 240

tgctcagtac tcttcacgc 20

<210> 241

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 241
agctcaatca tggctatcgg 20

<210> 242
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 242
ctacacattg catccaaccc 20

<210> 243
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 243
acttgtcgaa tagctcaggc 20

<210> 244
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 244
gtctaccctg agaaccagtt 20

<210> 245
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 245
 cagcaacaac ctaccttagc 20
 <210> 246
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 246
 cgatttggtg gtacgtgtcc 20
 <210> 247
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 247
 gccatttcct ttgtacctgc 20
 <210> 248
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide
 <400> 248
 ggccaataga gagaccacaa 20
 <210> 249
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 249
cggagtcaca tgggtagaat 20
<210> 250
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 250
cccagtagat ttgtcggttg 20
<210> 251
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 251
cccactagct gctactcaaa 20
<210> 252
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 252
ggtgttcgct caaagtagac 20
<210> 253
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 253

tactccagct cttactgtgc	20
<210> 254	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 254	
ggatgagcag tcaacagttc	20
<210> 255	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 255	
tcaggatcga tcagttgtcc	20
<210> 256	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 256	
cctctctttt gtgcggaac	20
<210> 257	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 257	
tgccataggat ttcgagaacg	20
<210> 258	

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 258

ggcattgtcc ttaacttcgc

20

<210> 259

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 259

tgcataaac tacgatgggc

20

<210> 260

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 260

ccctagtagc cacacaacat

20

<210> 261

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 261

gtgccatgaa tcatcgtctc

20

<210> 262

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 262

cgctctgatg aaagctccat 20

<210> 263

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 263

ccagccatag tgcatacct 20

<210> 264

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 264

gagattgtca tgtggtcgac 20

<210> 265

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 265

ccgcagtcta acaggaaatc 20

<210> 266

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 266

cgcttcgact gaaccttatg

20

<210> 267

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 267

cgatgcgacc aatagaagtg

20

<210> 268

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 268

gcccttggta cgacatatg

20

<210> 269

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 269

cagtgattta ggtgacgcag

20

<210> 270

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 270

ggcatggaag aggtagtttc

20

<210> 271

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 271

ccgatcgat tctgtgtcca

20

<210> 272

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 272

ctaagtcaag cacatgggac

20

<210> 273

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 273

gatccacact caatctcctg

20

<210> 274

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 274

ccttgtcaca tgctggtatc 20

<210> 275

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 275

cgcgattgtg gttaataggc 20

<210> 276

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 276

gtaggcaaag ttcaccacac 20

<210> 277

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 277

gccacgaatc gaacaagtac 20

<210> 278

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 278

ttgagatctc gatgagcacg	20
<210> 279	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 279	
gggccaagat ctattcgta	20
<210> 280	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 280	
gtggctatag gtatgtccga	20
<210> 281	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 281	
ccacactttc tgcattcgac	20
<210> 282	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 282	
cggcatctca aagcacatac	20

<210> 283

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 283

cgtccacaaa ttactgccc

20

<210> 284

<211> 19

<212> PRT

<213> Porcine teschovirus

<400> 284

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

Pro Gly Pro

<210> 285

<211> 20

<212> PRT

<213> Equine rhinitis A virus

<400> 285

Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser

1 5 10 15

Asn Pro Gly Pro

20

<210> 286

<211> 22

<212>

> PRT

<213> Foot-and-mouth disease virus

<400> 286

Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val

1 5 10 15

Glu Ser Asn Pro Gly Pro

20

<210> 287

<211> 18

<212> PRT

<213> Thosea asigna virus

<400> 287

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro

1

5

10

15

Gly Pro

<210> 288

<211> 5

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Enterokinase cleavage site

<400> 288

Asp Asp Asp Asp Lys

1

5

<210> 289

<211> 4

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Factor Xa cleavage site

<400> 289

Ile Glu Gly Arg

1

<210> 290

<211> 6

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Thrombin cleavage site

<400> 290

Leu Val Pro Arg Gly Ser

1 5

<210> 291

<211> 8

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Renin cleavage site

<400> 291

His Pro Phe His Leu Val Ile His

1 5

<210> 292

<211> 4

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

Caspase protease cleavage site

<400> 292

Asp Glu Val Asp

1