



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0051739
(43) 공개일자 2024년04월22일

(51) 국제특허분류(Int. Cl.) <i>G16B 30/00</i> (2019.01) <i>G16B 20/20</i> (2019.01) <i>G16B 40/00</i> (2019.01) <i>G16B 50/00</i> (2019.01) (52) CPC특허분류 <i>G16B 30/00</i> (2019.02) <i>G16B 20/20</i> (2019.02) (21) 출원번호 10-2022-0131806 (22) 출원일자 2022년10월13일 심사청구일자 2022년10월13일	(71) 출원인 인하대학교 산학협력단 인천광역시 미추홀구 인하로 100(용현동, 인하대학교) (72) 발명자 신현태 인천광역시 중구 서해대로 366, A동 771호(신흥동3가) (74) 대리인 특허법인정진
---	---

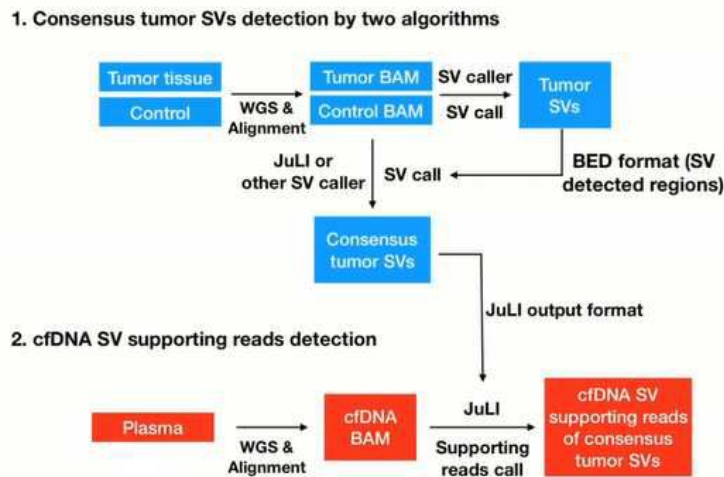
전체 청구항 수 : 총 11 항

(54) 발명의 명칭 cfDNA의 구조 변이 서열 검출을 통한 미세잔존질환 진단 방법

(57) 요약

본 발명은 암 환자 유래 cfDNA 시료를 이용하여 높은 민감도로 구조 변이 서열을 검출하고, 검출된 구조 변이 데이터를 이용하여 암 환자의 미세잔존질환 진단에 사용하기 위한 새로운 WGS (whole genome sequence) 분석 시스템 및 방법에 관한 것이다. 본 발명에서는 암 세포주와 NA12878을 1:12,800 비율로 섞은 경우에도 구조 변이 서열을 검출하였으며, 전이성 폐암환자의 cfDNA 에서도 구조 변이 서열을 민감하게 확인하였다. 따라서 본 발명에 따른 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법을 이용하면, 암 유래 cfDNA 가 낮은 순도로 혈액 등의 존재하더라도, 높은 민감도로 환자 특이적 구조 변이 서열을 검출할 수 있고, 암 치료 후 환자에 남아 있는 암 세포를 적은 시료로도 검출할 수 있어, 암 종과 상관없이 범용적으로 미세 잔존 질환 진단에 활용할 수 있다.

대표도 - 도1



(52) CPC특허분류

G16B 40/00 (2019.02)

G16B 50/00 (2019.02)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711118302
과제번호	2020R1F1A1072214
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	개인기초연구(과기정통부)(R&D)
연구과제명	혈장 세포유리 디옥시리보핵산의 구조변이 분석을 이용한 악성흑색종 재발 모니터링
방법 개발 및 임상 적용	
기 여 율	40/100
과제수행기관명	인하대학교의과대학부속병원
연구기간	2020.06.01 ~ 2022.02.28

이 발명을 지원한 국가연구개발사업

과제고유번호	1711137725
과제번호	2021R1A5A2031612
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	집단연구지원(R&D)
연구과제명	세포교신 제어 연구센터
기 여 율	40/100
과제수행기관명	인하대학교
연구기간	2021.06.01 ~ 2028.02.29

이 발명을 지원한 국가연구개발사업

과제고유번호	1711132675
과제번호	2021M3E5D1A02015185
부처명	과학기술정보통신부, 복지부
과제관리(전문)기관명	한국연구재단
연구사업명	혁신형의사과학자공동연구(R&D)(과기정통부, 복지부)
연구과제명	환자 주도 온라인 유전성 피부 질환 등록 시스템 구축과 구축된 데이터베이스를 활
용한 관련 연구 활성화	
기 여 율	20/100
과제수행기관명	인하대학교의과대학부속병원
연구기간	2021.01.01 ~ 2022.12.31

명세서

청구범위

청구항 1

- 1) 환자 유래 암 조직의 WGS (whole genome sequence)를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 및
- 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자의 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드 (supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 2

제1항에 있어서, 상기 방법은 WGS의 시퀀싱 심도(depth)를 조절하여 검출 민감도를 조절하는 것을 특징으로 하는, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 3

제1항에 있어서, 상기 WGS 분석 소프트웨어는 DELLY, BRASS, SvABA, dRanger, Pindel, BreakDancer, GASV, Hydra, CNVnator 및 JuLI 로 이루어진 군에서 선택되는 2종 이상인, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 4

제1항에 있어서,

상기 1) 단계는

- 1-1) DELLY, BRASS, SvABA, dRanger, Pindel, BreakDancer, GASV, Hydra 및 CNVnator 로 이루어진 군에서 선택된 1종의 구조 변이 분석 소프트웨어로 원발암 시료의 제 1 구조 변이 위치 데이터를 얻는 단계; 및
- 1-2) 상기 1-1) 단계에서 획득된 제 1 구조 변이 위치 데이터를 JuLI 에 입력하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 를 순차적으로 수행하는 것을 포함하는, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 5

제1항에 있어서, 상기 cfDNA는 암 환자의 뇌척수액, 흉수, 심낭액, 복수, 소변, 전혈, 혈장 또는 혈청 유래인, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 6

제1항에 있어서, 상기 2) 단계의 암 환자의 cfDNA 는 암 치료 중 또는 암 치료 후의 환자에서 획득되는 것인, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 7

제1항의 구조 변이는 유전자의 복제, 결실, 전좌 및 삽입으로 이루어진 군에서 선택된 1종 이상인, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 8

제1항에 있어서, 상기 암은 위암, 폐암, 비소세포성 폐암, 유방암, 난소암, 간암, 기관지암, 비인두암, 후두암, 췌장암, 방광암, 대장암, 결장암, 자궁경부암, 골암, 비소세포성 골암, 혈액암, 피부암(흑색종 등), 두부 또는 경부 암, 자궁암, 직장암, 항문 부근암, 결장암, 나팔관암, 자궁내막암, 질암, 음문암, 호지킨병(Hodgkin's disease), 식도암, 소장암, 내분비선암, 갑상선암, 부갑상선암, 부신암, 연조직 육종, 요도암, 음경암, 전립선암, 만성 또는 급성 백혈병, 림프구 림프종, 신장 또는 수뇨관암, 신장세포 암종, 신장골반암종, 배수성암(polyploid carcinoma), 침샘암, 육종암, 가성점액종, 간모세포종, 고환암, 교모세포종, 구순암, 난소생식세포종양, 기저세포암, 다발성골수종, 담낭암, 맥락막흑색종, 바터팩대부암, 복막암, 부신암, 설암, 소세포암, 소아림프종, 신경모세포종, 십이지장암, 요관암, 성상세포종, 수막종, 신우암, 외음부암, 흉선암, 중추신경계(central nervous system, CNS) 종양, 1차 중추신경계 림프종, 척수종양, 뇌간 신경교종 또는 뇌하수체 선종인, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법.

청구항 9

- 1) 환자 유래 암 조직의 WGS (whole genome sequence)를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 및
- 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자에서 암 치료 후 수득된 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드 (supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 미세잔존질환(Minimal Residual Disease; MRD)에 대한 정보를 제공하는 방법.

청구항 10

제9항에 있어서, 상기 미세잔존질환에 대한 정보는 치료 중 또는 치료 후에 잔존하는 암세포 유무, 암 재발 가능성 또는 암 치료 예후에 대한 정보인, 방법.

청구항 11

제9항에 있어서, 상기 2) 단계의 치료는 방사선 요법, 면역요법, 호르몬 요법, 화학요법 치료 또는 외과적 절제 치료인, 방법.

발명의 설명

기술 분야

[0001] 본 발명은 암 환자 유래 cfDNA (Cell-free DNA) 시료에서 높은 민감도로 구조 변이 서열을 검출하고, 이를 이용하여 암 환자의 미세잔존질환(Minimal Residual Disease; MRD)을 진단하기 위한 새로운 전장 유전체 서열 분석(Whole-Genome Sequencing, WGS) 시스템 및 방법에 관한 것이다.

배경 기술

[0003] 차세대 염기서열 분석(Next Generation Sequencing)은 유전체를 무수히 많은 조각으로 나눠서 읽고, 얻어진 염기서열 조각을 정렬하여 유전체의 서열을 분석하는 유전체 고속 분석 방법이다. 차세대 염기서열 분석 기술을

통한 전장 유전체 서열 분석(Whole-Genome Sequencing, WGS)은 거의 모든 유형의 체세포 돌연변이(somatic variant) 검출에 유용하고, 이러한 유용성 덕분에 여러분야에서 광범위하게 사용되고 있으며, 특히 암 유전체학에서 매우 중요한 역할을 하고 있다.

[0004] 전 세계적으로는 유전체 분석 사업은 빠르게 발전하고 있으며, 이러한 차세대 염기서열 분석 기법은 임상유전체학(clinical genomics), 약물유전체학(pharmaco - genomics) 및 중개 임상(translational medicine) 분야에서도 활발히 활용되고 있다.

[0005] 구조 변이(Structural Variants, SV)는 암 발생 과정에 중요한 역할을 하므로, 암 유전체에서 체세포 구조변이를 검출하기 위해 많은 생물 정보학 알고리즘과 툴이 개발되었다. 유전체의 특정 위치가 잘려서 서로 결합하여 발생하는 구조 변이는 결합 부위에 융합(fusion) DNA 서열을 생성하게 되고 이는 정상 조직에서는 확인되지 않는 특이적인 고유 서열이다. 암 세포에서는 apoptosis로 인해 혈중으로 cfDNA가 유입되는 것으로 알려져 있고, cfDNA에서 원발암의 특이적인 구조 변이에 의한 융합 DNA 서열이 검출된다면, 이를 이용하여 암 치료 후 체내에 남아있는 미세잔존질환(Minimal Residual Disease; MRD)을 진단하는데 사용할 수 있을 것으로 기대된다.

[0006] cfDNA에서 돌연변이를 찾기 위한 방법으로 targeted high-depth 시퀀싱이 널리 이용되고 있다. 상기 방법은 돌연변이를 민감하게 찾을 수 있다는 장점이 있으나, high-depth를 구현하기 위해 개별 암 종에 적합한 표적화된 패널을 준비해야 하고, 노이즈를 줄이기 위한 실험적인 전처리 공정이 필요하다.

[0007] 따라서 암 미세잔존질환을 진단하기 위해 범용적으로 활용할 수 있는, cfDNA 시료의 WGS를 이용하여 개별 암 특이적인 구조 변이 서열을 효과적으로 검출할 수 있는 새로운 방법에 대한 필요성이 있다.

발명의 내용

해결하려는 과제

[0009] 본 발명자들은 cfDNA 시료에서 구조 변이 서열을 민감하게 검출하기 위해서, 원발암의 WGS를 활용하여 구조 변이를 찾고, 이 구조 변이의 서열을 cfDNA의 WGS에서 매칭하는 방법을 통해 적은 수의 구조 변이 서열 정보를 가지고 있는 리드들을 cfDNA에서 검출할 수 있고, 이를 환자의 미세잔존질환 진단에 활용할 수 있음을 확인하고 본 발명을 완성하였다.

과제의 해결 수단

[0011] 상기 목적을 달성하기 위하여, 본 발명은 1) 환자 유래 암 조직의 WGS (whole genome sequence)를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 및 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자의 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드(supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법을 제공한다.

[0012] 또한 본 발명은 1) 환자 유래 암 조직의 WGS (whole genome sequence)를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 및 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자에서 암 치료 후 수득된 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드(supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 미세잔존질환(Minimal Residual Disease; MRD)에 대한 정보를 제공하는 방법을 제공한다.

발명의 효과

[0014] 본 발명에서는 암 세포주와 NA12878을 1:12,800 비율로 섞은 경우에도 구조 변이 서열을 검출하였으며, 전이성 폐암환자의 cfDNA에서도 구조 변이 서열을 민감하게 확인하였다. 따라서 본 발명에 따른 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법을 이용하면, 암 유래 cfDNA가 낮은 순도로 혈액 등의 존재하더라도,

높은 민감도로 환자 특이적 구조 변이 서열을 검출할 수 있고, 암 치료 후 환자에 남아있는 암 세포를 적은 시료로도 검출할 수 있어, 암 종과 상관없이 범용적으로 미세 잔존 질환 진단에 활용할 수 있다.

도면의 간단한 설명

- [0016] 도 1은 본 발명의 cfDNA 내 구조 변이 서열 검출 방법의 흐름도를 나타낸 도이다.
- 도 2는 본 발명 검출 방법 중 2단계에서 수행되는 대응 참조 서열과 각 지지 리드를 비교하는 구조 변이 서열 확인 공정의 모식도이다.
- 도 3은 세포주 혼합 실험을 이용한 분석 검증 결과를 요약하여 나타낸 도이다.
- 도 4 및 도 5는 임상 검증 실험 중 case 1환자 유래 원발암 조직에서 확인된 구조 변이가 cfDNA-1 및 cfDNA-2 시료상에서 동일하게 확인되고 normal control에서는 관찰되지 않음을 보여주는 결과이다.
- 도 6은 case 1 환자의 원발암 조직에서 KRAS 변이가 검출 (36%)되고, cfDNA-1, cfDNA-2 시료에서도 같은 변이가 19%, 6% 검출되는 결과를 확인한 도이다.
- 도 7 및 도 8은 임상 검증 실험 중 case 2환자 유래 원발암 조직에서 확인된 구조 변이가 cfDNA-1 및 cfDNA-2 시료상에서도 동일하게 확인되고 normal control에서는 관찰되지 않음을 보여주는 결과이다.
- 도 9은 case 2 환자의 원발암 조직에서 TP53 변이가 검출 (45%)되고, cfDNA-1, cfDNA-2 시료에서도 같은 변이가 2%, 0% 검출되는 결과를 확인한 도이다.
- 도 10은 case 2 환자의 원발암 조직에서 SMARCA4 변이가 검출 (27%)되고, cfDNA-1, cfDNA-2 시료에서도 같은 변이가 5%, 2% 검출되는 결과를 확인한 도이다.

발명을 실시하기 위한 구체적인 내용

- [0017] 본 발명은 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법 및 이를 이용한 미세잔존질환(Minimal Residual Disease; MRD)에 대한 정보를 제공하는 방법에 관한 것이다.
- [0018] 본 발명에 따르면, 암 환자의 원발암 조직에서 미리 확인된 구조 변이를 이용하여 치료 후 암 환자의 cfDNA에 존재하는 구조 변이 서열을 높은 민감도로 빠르게 비교 분석할 수 있어, 치료 후 환자에 잔존하는 암 세포를 모니터링하고 암 치료 경과 및 예후, 재발 가능성 등에 대한 정보를 제공할 수 있다.
- [0020] 이하, 본 발명에 대하여 상세히 설명한다.
- [0022] 본 발명은 1) 환자 유래 암 조직의 WGS를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 데이터를 얻는 단계; 및 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자의 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드 (supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법에 관한 것이다.
- [0023] 본 발명의 1) 단계는 원발암 시료의 공통 (consensus) 구조 변이 데이터를 얻는 단계이며, 환자 유래 암 조직의 WGS 및 정상 대조군의 WGS를 2가지 이상의 구조 변이 분석 데이터로 분석한다.
- [0024] 상기 단계에서는 보다 정확한 환자의 구조 변이 데이터를 얻기 위하여 2가지 이상의 구조 변이 분석 소프트웨어를 이용하는 것을 특징으로 한다. 해당 단계에서는 암 환자에서 수득된 암 조직, 즉 원발암 시료 및 정상 대조군을 이용하여 WGS 및 정렬 (alignment)을 수행하고, 이를 통해 시퀀싱 리드들이 어떤 염색체의 어느 위치에 있는 DNA 인지 표준 유전체 (reference genome) 상에서 위치를 확인하는 매핑 (Mapping) 이 수행된다. 매핑을 완료하면 각 시퀀싱 리드에 대하여 표준 유전체 상의 염색체 번호 및 위치 정보가 제공되며, 이를 포함하는 정렬된 염기 조각인 BAM(binary alignment map) 포맷을 얻을 수 있다. 제1 구조 변이 분석 소프트웨어를 이용하여, 수득된 종양 조직 유래 '종양 BAM' 파일과 '대조군 BAM' 파일에서 각 시퀀싱 리드를 비교 분석하여 특정 위치에

표준 유전체 서열과 다른 구조 변이가 일어난 변이체의 존재를 확인하기 위한 '구조 변이 호출 (Structural variant call, SV call)' 이 수행되며, 이를 통해 '제 1 구조 변이 위치 데이터'를 획득할 수 있다.

- [0025] 이어서 본 발명의 '공통 (consensus) 구조 변이 위치 데이터'를 획득하기 위하여, 제 2 구조변이 분석 소프트웨어를 이용하여 구조 변이 호출을 수행한다. 이때 공통 구조변이를 빠르게 확인하기 위해, 제1 구조 변이 분석 소프트웨어를 이용하여 획득된 '제1 구조 변이 위치 데이터'를 제2 구조 변이 분석 소프트웨어에 입력하여 구조 변이 호출을 수행할 수 있다.
- [0026] 즉, 본 발명에 있어서, '공통(consensus) 구조 변이 위치 데이터'란 복수 개의 구조 변이 분석 소프트웨어를 통한 SV call 에서 공통적으로 존재하는 것으로 확인된 구조 변이 위치에 대한 정보를 의미한다.
- [0027] 상기 1) 단계는 암환자의 원발암 시료에 존재하는 구조 변이에 대한 DB 를 구축하기 위한 목적으로 이용되는 것일 수 있으며, 이와 같이 한번 획득된 공통 구조 변이 위치 데이터는 향후 다양한 시점에서 획득한 동일 암 환자의 cfDNA 구조 변이와의 비교 분석에 모니터링 목적으로 반복 활용될 수 있다.
- [0028] 본 발명에 사용되는 구조 변이 분석 소프트웨어는 당 분야에서 전장 유전체 분석을 통해 체세포 구조 변이를 검출할 수 있는 것으로 알려진 다양한 종류의 소프트웨어를 제한없이 사용할 수 있으며, 예컨대 DELLY, BRASS, SvABA, dRanger, Pindel, BreakDancer, GASV, Hydra, CNVnator 및 JuLI 로 이루어진 군에서 선택되는 2종 이상을 사용할 수 있다. 상기 2종 이상의 분석 소프트웨어는 순차적으로 사용될 수 있고, 제1 WGS 분석 소프트웨어는 SV call 을 통해 각 시퀀싱 리드를 분석하여 특정 위치에 표준 유전체 서열과 다른 구조 변이가 일어난 변이체의 존재를 확인하고, '제1 구조 변이 위치 데이터'를 제공한다. 이후 제2 WGS 분석 소프트웨어에 제1 WGS 분석 소프트웨어에서 획득된 제1 구조 변이 위치 데이터를 VCF, BED format 등 으로 입력하고, 분석 대상인 종양 BAM 파일을 입력하면, 미리 확인된 제1 구조 변이 위치 데이터를 기반으로 신속하게 2종의 WGS 분석 소프트웨어에서 공통적으로 존재하는 '공통 구조 변이 위치 데이터'를 획득할 수 있다. WGS 분석 소프트웨어의 선택은 이에 제한되는 것은 아니나, 제1 WGS 분석 소프트웨어는 많은 정보량을 신속하게 처리할 수 있는 소프트웨어를 선택하는 것이 바람직하고, 본 발명에서는 바람직한 일 예로 DELLY (Version: 0.8.7, <https://github.com/dellytools/delly>)를 이용하였다. 또한 DELLY 에서 제공된 제 1 구조 변이 위치 데이터를 기반으로 신속하게 공통 구조 변이를 탐색하는 제2 WGS 분석 소프트웨어로 공개 소프트웨어인 JuLI(<https://github.com/sgilab/JuLI>, J Mol Diagn. 2020 Mar;22(3):304-318) 를 이용하였다.
- [0029] 따라서 본 발명의 상기 1) 단계는 1-1) DELLY, BRASS, SvABA, dRanger, Pindel, BreakDancer, GASV, Hydra 및 CNVnator 로 이루어진 군에서 선택된 1종의 분석 소프트웨어로 원발암 시료의 제 1 구조 변이 위치 데이터를 얻는 단계; 및 1-2) 상기 1-1) 단계에서 획득된 제 1 구조 변이 위치 데이터를 JuLI 에 입력하여 공통적으로 확인되는 원발암 시료의 공통(consensus) 구조 변이 위치 데이터를 얻는 단계; 를 순차적으로 수행하는 것을 포함하는, 암 환자 유래 cfDNA 시료에서 구조 변이를 검출하는 방법일 수 있다.
- [0030] 또한 본 발명은 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자의 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드 (supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계;를 포함한다.
- [0031] 본 발명은 암종에 상관없이 범용적으로 cfDNA 시료에 존재하는 적은 수의 구조 변이를 높은 민감도로 검출할 수 있음을 특징으로 한다. 본 발명에서 상기 cfDNA는 암 환자의 뇌척수액, 흉수, 심낭액, 복수, 소변, 전혈, 혈장 또는 혈청 유래의 cfDNA 일 수 있으나, 이에 제한되지 않는다.
- [0032] 본 발명은 통상적인 변이 검출 방법들과 마찬가지로, 검출 민감도는 sequencing depth에 따라 조절할 수 있다.
- [0033] 한편 상기 2) 단계의 분석 대상이 되는 암 환자의 cfDNA 는 암 치료 중 또는 암 치료 후의 환자에서 획득되는 것일 수 있다. 본 발명에서는 상기 1) 단계에서 암 환자의 원발암 조직 내 존재하는 공통 구조 변이 위치에 대한 DB를 구축하고, 이후 환자가 암 치료를 받는 중 또는 받은 후, 환자의 cfDNA 내 구조 변이 위치 서열과 비교함으로써, 치료 후 환자에 남아있는 미세한 수준의 잔존 암 세포에 대한 정보를 제공할 수 있다.
- [0034] 보다 구체적으로 상기 2) 단계는 다음과 같다: 먼저 암 환자의 cfDNA (Cell-free DNA) 를 WGS 및 정렬하여 획득된 각 시퀀싱 리드에 대하여 표준 유전체 상의 염색체 번호 및 위치 정보가 기록된 cfDNA BAM(binary alignment map) 포맷, 즉 cfDNA 매핑(mapping) 데이터를 얻는다. 이후 획득된 cfDNA 매핑 데이터인 cfDNA BAM 파일을 1) 단계에서 획득된 공통 구조 변이 위치 데이터와 함께 구조 변이 분석 소프트웨어에 입력하고, 지지 리드 콜 (supporting read call) 을 수행하여, 1) 단계에서 확인된 공통 구조 변이 위치에서 대응 참조 서열 (counter reference sequence) 과 함께 상기 지지 리드들을 정렬함으로써, cfDNA 시료 내 존재하는 구조 변이

에 의해 발생한 특이 서열을 매칭하는 단계이다.

- [0035] 상기 대응 참조 서열은 표준 유전체 데이터와 동일한 의미로 사용될 수 있으며, hg19 또는 hg38 등 일 수 있다.
- [0036] 본 발명의 방법은 암 종과 상관없이 모든 암에 범용적으로 사용할 수 있으며, 기존 영상이나 혈액 검사, 표적화 고속 시퀀싱 (targeted high-depth sequencing) 방법과 달리, 암 종류에 따라 검사 항목이나 사전 디자인을 해야 하는 별도의 공정을 수반하지 않는다. 따라서 본 발명의 대상이 되는 암은 당 분야에 공지된 암 종을 제한없이 포함할 수 있고, 예컨대 위암, 폐암, 비소세포성 폐암, 유방암, 난소암, 간암, 기관지암, 비인두암, 후두암, 췌장암, 방광암, 대장암, 결장암, 자궁경부암, 골암, 비소세포성 골암, 혈액암, 피부암(흑색종 등), 두부 또는 경부 암, 자궁암, 직장암, 항문 부근암, 결장암, 나팔관암, 자궁내막암, 질암, 음문암, 호지킨병 (Hodgkin's disease), 식도암, 소장암, 내분비선암, 갑상선암, 부갑상선암, 부신암, 연조직 육종, 요도암, 음경암, 전립선암, 만성 또는 급성 백혈병, 림프구 림프종, 신장 또는 수뇨관암, 신장세포 암종, 신장골반암종, 배수성암(polyploid carcinoma), 침샘암, 육종암, 가성점액종, 간모세포종, 고환암, 교모세포종, 구순암, 난소생식세포종양, 기저세포암, 다발성골수종, 담낭암, 맥락막흑색종, 바터팡대부암, 복막암, 부신암, 설암, 소세포암, 소아림프종, 신경모세포종, 십이지장암, 요관암, 성상세포종, 수막종, 신우암, 외음부암, 흉선암, 중추신경계(central nervous system, CNS) 종양, 1차 중추신경계 림프종, 척수종양, 뇌간 신경교종 또는 뇌하수체 선종일 수 있다.
- [0038] 본 발명에 있어서, 검출 대상이 되는 구조 변이는 체세포 구조 변이일 수 있고, 유전자의 복제, 결실, 전좌 및 삽입으로 이루어진 군에서 선택된 1종 이상일 수 있으며, 비교 대상인 대응 참조 서열과 비교하여 서열상 특이적인 차이를 보이는 것을 말한다. 대부분의 암은 암 발생 초기에 수십개에서 수백개의 구조 변이가 발생하며, 암이 진행되는 동안 이러한 구조 변이가 유지된다. 암 세포에서는 세포자멸사에 의하여 혈중에 cfDNA 가 흘러나 오게 되며, 혈중에 존재하는 cfDNA에서 암 세포에서 발견된 구조 변이 서열이 확인된다면, 환자의 몸 속에 암이 잔존하고 있고, 미세잔존질환을 유발할 가능성이 있음을 진단할 수 있다. 따라서, 본 발명의 검출 대상이 되는 구조 변이 서열은, 환자의 암세포에 존재하는 것으로 확인된 구조 변이 중 cfDNA에 존재하는 구조 변이의 서열이다.
- [0040] 본 발명의 또다른 양태로, 본 발명은 1) 환자 유래 암 조직의 WGS (whole genome sequence)를 2가지 이상의 구조 변이 분석 소프트웨어로 분석하여 공통적으로 확인되는 원발암 시료의 공통 (consensus) 구조 변이 위치 데이터를 얻는 단계; 및 2) 상기 1) 단계에서 획득한 구조 변이 위치의 참조 서열(reference sequence)과 해당 환자에서 암 치료 후 수득된 cfDNA WGS를 매칭하여, 공통 구조 변이 위치에 존재하는 구조 변이 서열 지지 리드 (supporting read)를 획득함으로써, cfDNA 시료 내 존재하는 구조 변이에 의해 발생한 특이 서열을 매칭하는 단계; 를 포함하는, 미세잔존질환(Minimal Residual Disease; MRD)에 대한 정보를 제공하는 방법을 제공한다.
- [0042] 본 발명에 있어, 미세잔존질환이란, 최소 잔류 질병이라고도 하며 치료 중이나 치료를 완료한 환자에 남아있는 소수의 악성 세포가 분자적으로 검출되는 상태를 의미한다. 미세잔존질환은 다양한 혈액암과 고형암의 추적 관찰 대상이 되며, 미세잔존질환을 확인하여 환자의 치료에 대한 반응성을 확인하고 재발의 위험을 예측할 수 있다. 따라서 본 발명의 미세잔존질환에 대한 정보는 치료 중 또는 치료 후에 잔존하는 암세포 유무, 암 재발 가능성 또는 암 치료 예후에 대한 정보일 수 있다. 만약 본 발명의 방법을 수행하여, 환자의 cfDNA 시료에서 원발암 시료에서 확인된 구조 변이 서열이 확인되는 경우, 이 환자는 암 세포가 잔존하거나, 치료에 대한 반응성이 낮거나, 재발의 위험이 높거나, 또는 암 치료 후 예후가 나쁠 것으로 예측할 수 있다.
- [0043] 본 발명에서 암 환자의 치료는 방사선 요법, 면역요법, 호르몬 요법, 화학요법 치료 또는 외과적 절제 치료 등 당 분야에 공지된 치료법을 모두 제한없이 포함한다.
- [0044] 또한 본 발명은 환자 자신의 원발암 공통 구조 변이와 치료 이후 cfDNA 구조 변이 서열을 비교하여 미세잔존질환에 대한 정보를 제공하므로, 환자 맞춤형 정보 제공 방법으로 활용될 수 있다.
- [0045] 본 발명을 이용하면, 매우 낮은 순도의 cfDNA를 포함하는 치료 후 환자의 혈액, 혈장 또는 혈청 유래 시료에서도 높은 민감도로 구조 변이 서열을 검출할 수 있다. 본 발명의 일 구현예에서는 1:12,800까지 희석된 시료에서도 충분한 수의 지지 리드를 확인할 수 있고 이를 통해 환자 cfDNA 시료의 구조 변이 서열을 확인하여 미세잔존

질환에 대한 정보를 제공할 수 있음을 확인하였다.

[0046] 본 발명의 암 환자 유래 cfDNA 시료에서 구조 변이 서열을 검출하는 방법 및 미세잔존질환에 대한 정보를 제공하는 방법은 모두 컴퓨터 시스템에 의해 *in silico* 로 수행되는 것인 방법일 수 있다. 따라서 염기서열 변이 정보는 컴퓨터 시스템을 통하여 접수/수득될 수 있으며, 이런 측면에서 본 발명의 방법은 유전자 변이 정보를 컴퓨터 시스템으로 접수하는 단계를 추가로 포함할 수 있다.

[0048] 상술한 본 발명의 내용은 상호 모순되지 않는 한, 서로 동일하게 적용되며, 당해 기술분야의 통상의 기술자가 적절한 변경을 가해 실시하는 것 또한 본 발명의 범주에 포함된다.

[0050] 이하 본 발명을 실시예를 통해 상세하게 설명하나 본 발명의 범위가 하기 실시예로만 한정되는 것은 아니다.

[0052] **실시예 1. 혈장 cfDNA 내 구조 변이 특이 서열 검출 방법 구축**

[0053] 암환자의 혈장 cfDNA 내 구조 변이 특이 서열을 확인하기 위하여, 다음의 2 단계로 이루어진 프로세스를 구축하였다:

[0055] **1. 환자의 공통 (consensus) 구조 변이 확인 단계**

[0056] 환자가 가진 공통 (consensus) 구조 변이를 찾기 위하여, 2가지 종류의 분석툴을 이용하였다. 환자 유래의 암 조직과 대조군 시료를 이용하여 WGS을 수행한 후 BWA 의 default setting으로 정렬(alignment)을 수행하였다. 분석툴로는 DNA 체세포 변이, DNA 융합을 검출할 수 있는 구조 변이 검출 소프트웨어인 JuLI(<https://github.com/sgilab/JuLI>)와 다른 구조변이 검출 소프트웨어인 DELLY (DELly Version: 0.8.7, <https://github.com/dellytools/delly>) 를 활용하였다. 1차적으로 DELLY를 통해 종양 BAM 및 대조군 BAM 파일을 이용한 구조 변이 호출을 수행하여 환자의 원발암 시료에 존재하는 구조 변이 위치(제 1구조 변이 위치 데이터)를 확인하였다. 확인된 환자 구조 변이 위치 결과를 BED 포맷으로 JuLI 에 입력하고, 종양 BAM 파일을 입력한 후 구조 변이 호출을 수행하여 DELLY 와 JuLI 에서 공통적으로 구조 변이가 의심되는 위치를 확인하여 공통 (consensus) 구조 변이 위치 데이터를 얻었다. 확인된 공통 구조 변이 위치 데이터를 이후 2단계의 JuLI output 포맷 형태로 활용하였다.

[0058] **2. 지지 리드 검출을 통한 cfDNA 구조 변이 서열 검출 단계**

[0059] 암 발생 이후 환자에서 혈장 시료를 수득한 후, 해당 시료에서 DNA를 추출하여 WGS 분석을 수행한 후 BWA 의 default setting으로 정렬(alignment)을 수행하여 cfDNA BAM 파일을 수득하였다. 상기 1 단계를 통해 확보한 공통 구조 변이 위치와 동일한 위치에 구조 변이 특이 서열을 의미하는 지지 리드(supporting read) 가 있는지 cfDNA BAM 파일에서 확인하기 위하여 JuLI 프로그램의 'callread' 함수를 적용하였다. 'callread' 함수를 이용하면, 공통(consensus) 구조 변이 위치에서 hg19 또는 hg38과 같은 대응 참조 서열(counter reference sequence)과 각 지지 리드의 비교를 통해 구조 변이에 의해 발생한 특이 서열을 매칭할 수 있다. 이때, JuLI 파라미터 중 splitratio의 default 값을 조정하면 특이도를 조절할 수 있으며 0.95 로 설정하여 특이도를 높였다.

[0060] 이러한 방법을 이용하면, 적은 수의 리드 존재만으로도 확인하고자 하는 fusion site 에서 공통 구조 변이를 서포팅하는 특이 서열을 가진 cfDNA 구조 변이 지지 리드가 있는지 높은 민감도로 검출할 수 있다.

[0062] 상기 2단계로 이루어진 본 발명의 혈장 cfDNA 내 구조 변이 서열 검출 방법의 흐름도를 도 1에 나타내었고, 2 단계의 대응 참조서열과 지지 리드의 비교를 도 2에 모식화하여 나타내었다.

[0064] **실험예 1. 분석 검증**

[0065] 이하에서는 실제 환자의 시료와 세포주를 이용한 분석 검증 결과를 설명한다.

[0067] **1.1 분석 시료의 준비**

[0068] cfDNA 에 존재하는 DNA 구조 변이를 확인하기 위하여, 다음과 같은 시료를 준비하였다. 분석 검증 실험을 위하여 암세포 및 표준 물질을 시료로 이용하였다. 변이 정의를 수행함에 있어, 표준 물질로 NA12878를 지정하였고, 이를 기반으로 구조 변이를 비교 검출하였다. NA 12878는 Coriell institute에서 구매하여 사용하였다. 5개의 암세포 주 (WM2664, A375, SNU16, HCC1954, HCC95)는 한국 세포주은행에서 구입하여 사용하였다.

[0069] NA12878의 DNA와 5개의 암세포 주 (WM2664, A375, SNU16, HCC1954, HCC95)에서 채취한 DNA를 일반적인 cfDNA 절편 사이즈로 알려져 있는 150 내지 170bp 로 sharing 하였다. 5종 암세포주를 각각 10ng/ul 로 희석하고 동일 부피 50 ul로 혼합하여 최종 10ng/ul, 250ul 의 혼합 시료를 준비하여 초기 시료로 준비하였다. NA 12878 의 초기 시료량은 50 ng/ul 이다. 이후 이들 5종의 암세포 혼합 시료와 NA 12878 혼합 비율을 cell line: NA12878 = 1:100, 1:200, 1:400, 1:800, 1:1600, 1:3200, 1:6400 및 1:12800 로 달리하여 제조하고 본 발명의 민감도를 확인하기 위한 실험에 이용하였고, 모든 실험은 3회 반복 수행하였다.

[0070] 환자의 임상 시료는 인하대병원에 내원한 전이성 폐암 환자로부터 각각 FFPE (Formalin Fixed Paraffin Embedded) 암 조직, 혈액 buffy coat (control tissue) 를 얻었으며, 다른 시점에서 채취한 혈장 샘플을 준비하여 cfDNA 를 채취하였다.

[0072] **1.2 세포주 혼합 실험을 이용한 분석 검증 (Analytical validation)**

[0073] 실시예 1의 방법을 이용한 분석 검증을 수행하였다. 먼저 NA 12878 와 5개의 암세포 주 (WM2664, A375, SNU16, HCC1954, HCC95), 및 이들의 혼합 시료를, cell line: NA12878 = 1:100, 1:200, 1:400, 1:800, 1:1600, 1:3200, 1:6400 및 1:12800 로 혼합 비율을 달리하여 3회 반복 분석한 결과를 표 1 내지 표 8에 나타내었다. 각 표는 5종의 암 세포주에서 구조 변이 검출을 3회 반복한 값을 나타내며, 동일한 Sample ID 는 하나의 BAM 파일을 의미한다. CD_21_16908 내지 CD_21_16912는 다음과 같이 각 암 세포주에 대응되는 데이터이다: CD_21_16908 (A375SM), CD_21_16909 (HCC95), CD_21_16910 (HCC1954), CD_21_16911 (SNU16), CD_21_16912 (WM2664). 암 세포주는 평균 22X 로 데이터가 생산되었고, 혼합 시료는 평균 70X로 데이터를 생산하였다. 각 세포주의 변이가 레퍼런스로만 된 WGS에서 검출되는지 확인한 데이터는 표 9에 나타내었다.

표 1

[0075]

Cell-line	5MIX+NA12878 혼합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16908	1:100	CD_21_17603_CL_D_SRG_1	WGS_60X - 1st	171	89	162
CD_21_16909	1:100	CD_21_17603_CL_D_SRG_1	WGS_60X - 1st	292	130	239
CD_21_16910	1:100	CD_21_17603_CL_D_SRG_1	WGS_60X - 1st	475	277	847
CD_21_16911	1:100	CD_21_17603_CL_D_SRG_1	WGS_60X - 1st	194	115	689
CD_21_16912	1:100	CD_21_17603_CL_D_SRG_1	WGS_60X - 1st	247	131	251
CD_21_16908	1:100	CD_22_06205_CL_D_SRG_1	WGS_60X - 2nd	171	51	87
CD_21_16909	1:100	CD_22_06205_CL_D_SRG_1	WGS_60X - 2nd	292	84	158
CD_21_16910	1:100	CD_22_06205_CL_D_SRG_1	WGS_60X - 2nd	475	181	407
CD_21_16911	1:100	CD_22_06205_CL_D_SRG_1	WGS_60X - 2nd	194	80	382
CD_21_16912	1:100	CD_22_06205_CL_D_SRG_1	WGS_60X - 2nd	247	79	118

CD_21_16 908	1:100	CD_22_10086_CL_D _SRG_1	WGS_60X - 3rd	171	64	110
CD_21_16 909	1:100	CD_22_10086_CL_D _SRG_1	WGS_60X - 3rd	292	83	137
CD_21_16 910	1:100	CD_22_10086_CL_D _SRG_1	WGS_60X - 3rd	475	170	442
CD_21_16 911	1:100	CD_22_10086_CL_D _SRG_1	WGS_60X - 3rd	194	76	396
CD_21_16 912	1:100	CD_22_10086_CL_D _SRG_1	WGS_60X - 3rd	247	79	109

표 2

[0076]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16 908	1:200	CD_21_17605_CL_D _SRG_1	WGS_60X - 1st	171	18	23
CD_21_16 909	1:200	CD_21_17605_CL_D _SRG_1	WGS_60X - 1st	292	49	71
CD_21_16 910	1:200	CD_21_17605_CL_D _SRG_1	WGS_60X - 1st	475	100	197
CD_21_16 911	1:200	CD_21_17605_CL_D _SRG_1	WGS_60X - 1st	194	45	194
CD_21_16 912	1:200	CD_21_17605_CL_D _SRG_1	WGS_60X - 1st	247	46	71
CD_21_16 908	1:200	CD_22_06207_CL_D _SRG_1	WGS_60X - 2nd	171	23	38
CD_21_16 909	1:200	CD_22_06207_CL_D _SRG_1	WGS_60X - 2nd	292	45	60
CD_21_16 910	1:200	CD_22_06207_CL_D _SRG_1	WGS_60X - 2nd	475	94	192
CD_21_16 911	1:200	CD_22_06207_CL_D _SRG_1	WGS_60X - 2nd	194	44	182
CD_21_16 912	1:200	CD_22_06207_CL_D _SRG_1	WGS_60X - 2nd	247	38	56
CD_21_16 908	1:200	CD_22_10087_CL_D _SRG_1	WGS_60X - 3rd	171	42	64
CD_21_16 909	1:200	CD_22_10087_CL_D _SRG_1	WGS_60X - 3rd	292	60	86
CD_21_16 910	1:200	CD_22_10087_CL_D _SRG_1	WGS_60X - 3rd	475	139	290
CD_21_16 911	1:200	CD_22_10087_CL_D _SRG_1	WGS_60X - 3rd	194	67	299
CD_21_16 912	1:200	CD_22_10087_CL_D _SRG_1	WGS_60X - 3rd	247	69	108

표 3

[0077]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_169 08	1:400	CD_21_17606_CL_D _SRG_1	WGS_60 X - 1st	171	7	11
CD_21_169 09	1:400	CD_21_17606_CL_D _SRG_1	WGS_60 X - 1st	292	34	41
CD_21_169 10	1:400	CD_21_17606_CL_D _SRG_1	WGS_60 X - 1st	475	60	108
CD_21_169 11	1:400	CD_21_17606_CL_D _SRG_1	WGS_60 X - 1st	194	30	92

CD_21_169 12	1:400	CD_21_17606_CL_D _SRG_1	WGS_60 X - 1st	247	25	32
CD_21_169 08	1:400	CD_22_06208_CL_D _SRG_1	WGS_60 X - 2nd	171	14	18
CD_21_169 09	1:400	CD_22_06208_CL_D _SRG_1	WGS_60 X - 2nd	292	24	27
CD_21_169 10	1:400	CD_22_06208_CL_D _SRG_1	WGS_60 X - 2nd	475	68	113
CD_21_169 11	1:400	CD_22_06208_CL_D _SRG_1	WGS_60 X - 2nd	194	30	91
CD_21_169 12	1:400	CD_22_06208_CL_D _SRG_1	WGS_60 X - 2nd	247	25	35
CD_21_169 08	1:400	CD_22_10088_CL_D _SRG_1	WGS_60 X - 3rd	171	16	23
CD_21_169 09	1:400	CD_22_10088_CL_D _SRG_1	WGS_60 X - 3rd	292	26	36
CD_21_169 10	1:400	CD_22_10088_CL_D _SRG_1	WGS_60 X - 3rd	475	70	130
CD_21_169 11	1:400	CD_22_10088_CL_D _SRG_1	WGS_60 X - 3rd	194	26	114
CD_21_169 12	1:400	CD_22_10088_CL_D _SRG_1	WGS_60 X - 3rd	247	28	38

표 4

[0079]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_169 08	1:800	CD_21_17607_CL_D _SRG_1	WGS_60 X - 1st	171	5	11
CD_21_169 09	1:800	CD_21_17607_CL_D _SRG_1	WGS_60 X - 1st	292	7	9
CD_21_169 10	1:800	CD_21_17607_CL_D _SRG_1	WGS_60 X - 1st	475	20	35
CD_21_169 11	1:800	CD_21_17607_CL_D _SRG_1	WGS_60 X - 1st	194	11	33
CD_21_169 12	1:800	CD_21_17607_CL_D _SRG_1	WGS_60 X - 1st	247	9	10
CD_21_169 08	1:800	CD_22_06209_CL_D _SRG_1	WGS_60 X - 2nd	171	6	8
CD_21_169 09	1:800	CD_22_06209_CL_D _SRG_1	WGS_60 X - 2nd	292	13	18
CD_21_169 10	1:800	CD_22_06209_CL_D _SRG_1	WGS_60 X - 2nd	475	27	49

CD_21_169 11	1:800	CD_22_06209_CL_D _SRG_1	WGS_60 X - 2nd	194	21	68
CD_21_169 12	1:800	CD_22_06209_CL_D _SRG_1	WGS_60 X - 2nd	247	12	15
CD_21_169 08	1:800	CD_22_10089_CL_D _SRG_1	WGS_60 X - 3rd	171	5	9
CD_21_169 09	1:800	CD_22_10089_CL_D _SRG_1	WGS_60 X - 3rd	292	11	17
CD_21_169 10	1:800	CD_22_10089_CL_D _SRG_1	WGS_60 X - 3rd	475	27	49
CD_21_169 11	1:800	CD_22_10089_CL_D _SRG_1	WGS_60 X - 3rd	194	21	65
CD_21_169 12	1:800	CD_22_10089_CL_D _SRG_1	WGS_60 X - 3rd	247	9	9

표 5

[0081]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_169 08	1:1600	CD_21_17604_CL_D _SRG_1	WGS_60 X - 1st	171	4	4
CD_21_169 09	1:1600	CD_21_17604_CL_D _SRG_1	WGS_60 X - 1st	292	8	10
CD_21_169 10	1:1600	CD_21_17604_CL_D _SRG_1	WGS_60 X - 1st	475	11	23
CD_21_169 11	1:1600	CD_21_17604_CL_D _SRG_1	WGS_60 X - 1st	194	8	29
CD_21_169 12	1:1600	CD_21_17604_CL_D _SRG_1	WGS_60 X - 1st	247	6	9
CD_21_169 08	1:1600	CD_22_06206_CL_D _SRG_1	WGS_60 X - 2nd	171	5	5
CD_21_169 09	1:1600	CD_22_06206_CL_D _SRG_1	WGS_60 X - 2nd	292	2	2
CD_21_169 10	1:1600	CD_22_06206_CL_D _SRG_1	WGS_60 X - 2nd	475	15	23
CD_21_169 11	1:1600	CD_22_06206_CL_D _SRG_1	WGS_60 X - 2nd	194	12	31
CD_21_169 12	1:1600	CD_22_06206_CL_D _SRG_1	WGS_60 X - 2nd	247	10	11
CD_21_169 08	1:1600	CD_22_10090_CL_D _SRG_1	WGS_60 X - 3rd	171	3	3
CD_21_169 09	1:1600	CD_22_10090_CL_D _SRG_1	WGS_60 X - 3rd	292	5	6

CD_21_169 10	1:1600	CD_22_10090_CL_D _SRG_1	WGS_60 X - 3rd	475	17	26
CD_21_169 11	1:1600	CD_22_10090_CL_D _SRG_1	WGS_60 X - 3rd	194	12	27
CD_21_169 12	1:1600	CD_22_10090_CL_D _SRG_1	WGS_60 X - 3rd	247	10	11

표 6

[0083]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16 908	1:3200	CD_22_10091_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	171	5	5
CD_21_16 909	1:3200	CD_22_10091_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	292	2	2
CD_21_16 910	1:3200	CD_22_10091_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	475	9	12
CD_21_16 911	1:3200	CD_22_10091_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	194	10	14
CD_21_16 912	1:3200	CD_22_10091_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	247	3	5
CD_21_16 908	1:3200	CD_22_10094_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	171	2	3
CD_21_16 909	1:3200	CD_22_10094_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	292	7	9
CD_21_16 910	1:3200	CD_22_10094_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	475	14	19
CD_21_16 911	1:3200	CD_22_10094_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	194	11	19
CD_21_16 912	1:3200	CD_22_10094_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	247	4	5
CD_21_16 908	1:3200	CD_22_10097_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	171	1	1
CD_21_16 909	1:3200	CD_22_10097_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	292	2	2

CD_21_16 910	1:3200	CD_22_10097_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	475	13	17
CD_21_16 911	1:3200	CD_22_10097_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	194	5	7
CD_21_16 912	1:3200	CD_22_10097_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	247	1	1

표 7

[0084]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16 908	1:6400	CD_22_10092_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	171	2	2
CD_21_16 909	1:6400	CD_22_10092_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	292	0	0
CD_21_16 910	1:6400	CD_22_10092_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	475	3	4
CD_21_16 911	1:6400	CD_22_10092_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	194	9	12
CD_21_16 912	1:6400	CD_22_10092_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	247	1	1
CD_21_16 908	1:6400	CD_22_10095_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	171	1	2
CD_21_16 909	1:6400	CD_22_10095_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	292	2	2
CD_21_16 910	1:6400	CD_22_10095_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	475	8	13
CD_21_16 911	1:6400	CD_22_10095_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	194	2	2
CD_21_16 912	1:6400	CD_22_10095_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	247	1	1
CD_21_16 908	1:6400	CD_22_10098_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	171	2	3
CD_21_16 909	1:6400	CD_22_10098_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	292	4	5

CD_21_16 910	1:6400	CD_22_10098_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	475	4	8
CD_21_16 911	1:6400	CD_22_10098_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	194	5	12
CD_21_16 912	1:6400	CD_22_10098_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	247	4	4

표 8

[0085]

Cell-line	5MIX+NA12878 혼 합비율	Sample ID	Remark	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16 908	1:12800	CD_22_10093_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	171	2	2
CD_21_16 909	1:12800	CD_22_10093_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	292	0	0
CD_21_16 910	1:12800	CD_22_10093_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	475	5	8
CD_21_16 911	1:12800	CD_22_10093_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	194	3	3
CD_21_16 912	1:12800	CD_22_10093_CL_D _SRG_1	WGS_60_hi gh dilution - 1st	247	0	0
CD_21_16 908	1:12800	CD_22_10096_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	171	2	2
CD_21_16 909	1:12800	CD_22_10096_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	292	1	1
CD_21_16 910	1:12800	CD_22_10096_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	475	1	2
CD_21_16 911	1:12800	CD_22_10096_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	194	3	4
CD_21_16 912	1:12800	CD_22_10096_CL_D _SRG_1	WGS_60_hi gh dilution - 2nd	247	1	1
CD_21_16 908	1:12800	CD_22_10099_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	171	0	0
CD_21_16 909	1:12800	CD_22_10099_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	292	1	2

CD_21_16 910	1:12800	CD_22_10099_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	475	5	5
CD_21_16 911	1:12800	CD_22_10099_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	194	0	0
CD_21_16 912	1:12800	CD_22_10099_CL_D _SRG_1	WGS_60_hi gh dilution - 3rd	247	1	1

표 9

[0086]

Cell-line	reference	Sample ID	Total fusion number of each cell-line	cfDNA Fusion Number	Reads Count
CD_21_16908	NA12878	CD_22_06210_ET_D_SRG_1	171	0	0
CD_21_16909	NA12878	CD_22_06210_ET_D_SRG_1	292	0	0
CD_21_16910	NA12878	CD_22_06210_ET_D_SRG_1	475	0	0
CD_21_16911	NA12878	CD_22_06210_ET_D_SRG_1	194	0	0
CD_21_16912	NA12878	CD_22_06210_ET_D_SRG_1	247	0	0

[0088]

암 세포주에 존재하는 구조 변이수를 의미하는 'Total fusion number of each cell-line (실시에 1의 원발암 시료 구조 변이에 대응)'은 각 암종에서 다른 세포주에 중복으로 존재하는 germline구조 변이를 제거한 후 각각 171, 292, 475, 194, 247 개였으며, 표 1 내지 표 8에 나타난 바와 같이, 각 암세포주에 존재하는 구조 변이가 cfDNA BAM파일에서도 검출(cfDNA Fusion Number)되고, 이를 서포팅하는 지지 리드가 확인되었다 (reads count). 또한, 세포주가 섞이지 않은 레퍼런스 WGS에서는 각 세포주의 구조 변이 서열이 검출되지 않음을 확인하였다 (표9).

[0089]

세포주를 이용한 실험의 민감도와 특이도를 정리한 결과를 표 10 및 도 3에 나타내었다.

표 10

[0090]

	Cancer cell purity	Detected/Total	Sensitivity
NA12878 + 5 cancer cell lines (1:100): 3 replicates	1.0000%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:200): 3 replicates	0.5000%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:400): 3 replicates	0.2500%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:800): 3 replicates	0.1250%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:1600): 3 replicates	0.0625%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:3200): 3 replicates	0.0313%	15/15	100.0%
NA12878 + 5 cancer cell lines (1:6400): 3 replicates	0.0156%	14/15	93.3%
NA12878 + 5 cancer cell lines (1:12800): 3 replicates	0.0078%	11/15	73.3%
	Cancer cell purity	Not detected/Total	Specificity
NA12878	0.0000%	5/5	100.0%

[0092]

표 10에 정리된 바와 같이, 희석 배율을 높일수록 지지 리드의 수는 점차 감소하는 경향을 나타냈지만, 1:12,800까지 희석하더라도 구조 변이 서열을 확인할 수 있는 서포트 서열이 확인되고 약 73.3%의 높은 민감도

로 검출할 수 있음을 확인하였다. 특히, 구조변이가 475개를 가지고 있는HCC1954의 경우 1:12,800으로 희석해도 100% 검출 할 수 있었다. 이는 본 발명의 방법을 이용하면 cfDNA 시료에 존재하는 구조 변이 서열을 매우 효과적으로 검출하여 암 환자의 미세잔존질환에 대한 정보를 제공할 수 있음을 의미한다.

1.3 임상 시료를 이용한 임상 검증 (Clinical validation)

본 발명의 방법을 이용하여 실제 환자의 혈액, cfDNA 에서도 변이를 효과적으로 검출가능한지 임상 검증 실험을 수행하였다.

시료를 환자의 원발암 조직, 혈액, 2가지 이상의 다른 시점에서 수득한 cfDNA1, cfDNA2 로 하여 실시예 1과 같은 분석 방법을 이용하였다.

각 암환자는 전이성 폐암환자이며, 각각 case1, case 2,로 기재하였다. 환자로부터 수득한 FFPE 암조직을 이용하여 약30X로 WGS 데이터가 생성되었고, 나머지 cfDNA 시료들로부터 평균 48X 정도의 WGS 데이터가 생성되었다.

구체적으로 환자의 FFPE 암조직을 시료로 하여 DELLY를 이용한 WGS 분석을 수행하고 BWA 의 default setting으로 alignment 하였으며, JuLI를 이용하여 2가지 분석툴에서 공통적으로 확인되는 환자의 공통 (consensus) 구조 변이를 확인하였다. 이를 JuLI output 포맷으로 활용하였다. 이후 서로 다른 시점에서 수득한 혈장 cfDNA샘플의 WGS에 의해 수득된BAM 파일과 상기에서 확인된 공통 (consensus) 구조 변이 데이터를 대응 참조 서열(counter reference sequence)과 비교하는 2단계 supporting read call 을 통해 cfDNA 구조 변이 서열을 확인하였다. 이를 통해 서로 다른 시점에서 수득한 환자의 혈장 내 존재하는 cfDNA 구조 변이를 보다 민감하게 검출할 수 있으므로, 환자에 남아있는 암 미세잔존질환에 대한 진단이 가능하다.

환자 Case 1

도 4 및 도 5는 Case 1의 분석결과이며, 각기 다른 시점인 2006.11.07, 2007.04.12 시점에 수득한cfDNA를 시료로 하여 구조 변이 서열을 확인한 결과를 각각 나타낸다. 도 4 및 도 5에서는 환자 유래 원발암 조직에서 확인된 구조 변이가 cfDNA-1 및 cfDNA-2 시료상에서도 동일하게 확인되는 것을 확인하였다.

본 발명의 효과를 검증하기 위하여 환자의 원발암을 패널 시퀀싱을 수행하여 암 돌연변이를 확인하고, 이를 cfDNA에서 확인하였다. 패널 시퀀싱을 통해 case 1 환자가 KRAS 변이를 약 30% 나타내는 환자임을 확인하였고, 동일 변이가 원발암, cfDNA-1 및cfDNA-2 의 WGS에서도 실제로 확인되었다. WGS에서 KRAS 변이 위치를 확인한 결과는 도6 에 나타내었다.

도 6에서는 원발암 조직에서 패널 시퀀싱과 비슷한 정도의 KRAS 변이가 검출 (36%)되는 것을 확인하였고, cfDNA-1, cfDNA-2 시료에서도 KRAS 변이가 19%, 6% 검출되어, 본 발명의 방법을 이용하면, cfDNA 시료의 구조 변이 서열 분석을 통해 미세잔존암을 확인할 수 있음을 검증하였다. 해당 환자에서 검출된 Read count 수는 하기 표 11과 같다.

표 11

cfDNA samples	Tumor Fusion Number	cfDNA Fusion Number	ReadsCount
cfDNA-1	131	115	4122
cfDNA-2	131	112	1926

표 11에 따르면, case1 환자는 원발암 시료에서 분석한 결과 총 131개의 구조 변이 서열인 융합(fusion) 서열을 가지고 있는 환자로, 본 발명의 방법을 통해 cfDNA 에서도 구조 변이가 115 또는 112개 확인되며, 19% KRAS변이를 나타낸 cfDNA-1 시료에서 4122 개의 지지 리드, 6% KRAS 변이를 나타낸 cfDNA-2 시료에서 1926 개의 지지 리드가 확인되었다.

이는 혈장 cfDNA를 WGS 로 분석하는 종래 방법이 5% 이하의 변이는 검출하기 어려운 낮은 민감도를 나타내는 것과 비교하여, 본 발명은 6% 정도의 변이에서도 1900 개 이상의 리드를 나타내므로, 훨씬 우수한 검출능이 있음을 보여주는 결과이다.

[0109] **환자 Case 2**

[0110] 도 7 및 도 8 는 Case 2의 분석결과이며, 각기 다른 시점인 2010.10.29, 2011.01.04 시점에 수득한cfDNA 를 시료로 하여 구조 변이 서열을 확인한 결과를 나타낸다. 도 7 및 도 8에서는 환자 유래 원발암 조직에서 확인된 구조 변이가 cfDNA-1 및 cfDNA-2 시료상에서도 확인됨을 나타낸다.

[0111] 패널 시퀀싱을 통해 case 2 환자가 가지고 있는 돌연변이들을 원발암, cfDNA-1 및cfDNA-2 의 WGS에서도 실제로 확인하였고 이는 도9 및 도10에 나타내었다.

[0112] 도 9에서 cfDNA-1, cfDNA-2에서 TP53변이가 각각2%, 0%, 도 10에는 SMARCA4 변이가 각각 5%, 2% 변이가 확인되었으나, 표 12에 나타낸 바와 같이, 2-5% 변이를 나타낸 cfDNA-1에서 128 개의 리드, 0-2% 변이에서 118 개의 리드가 확인되어, 2 내지 5% 이하의 낮은 대립유전자 빈도를 나타내는 시료에서도 구조 변이 존재를 나타내는 지지 리드가 충분히 검출됨을 확인하였다.

표 12

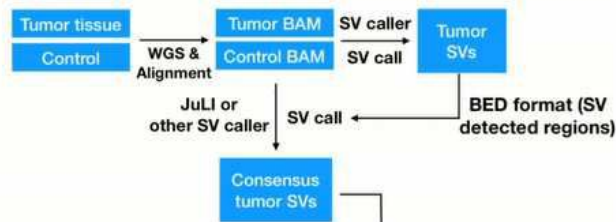
[0113]

cfDNA samples	Tumor Fusion Number	cfDNA Fusion Number	Reads Count
cfDNA-1	62	23	128
cfDNA-2	62	22	118

도면

도면1

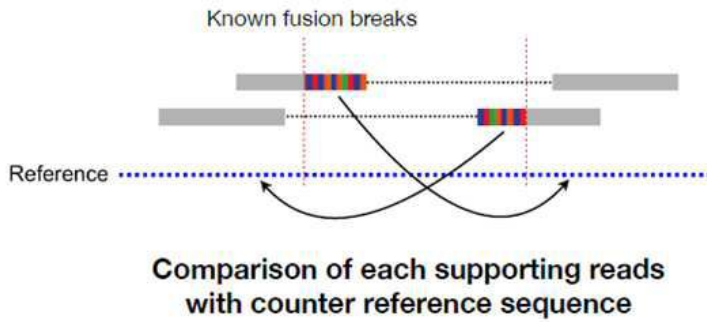
1. Consensus tumor SVs detection by two algorithms



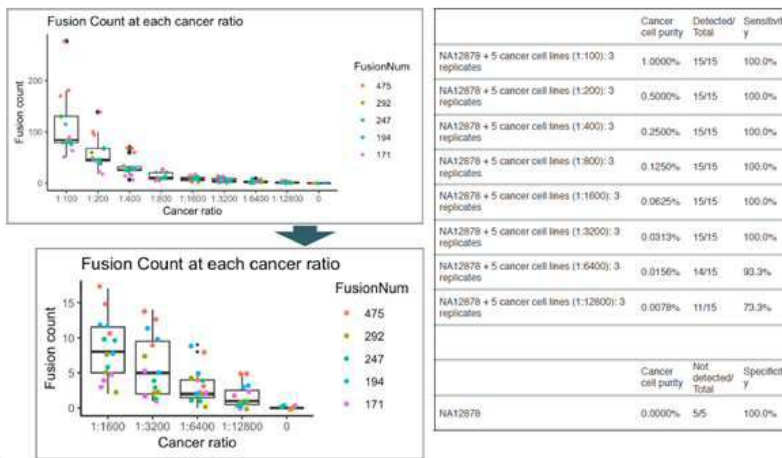
2. cfDNA SV supporting reads detection



도면2



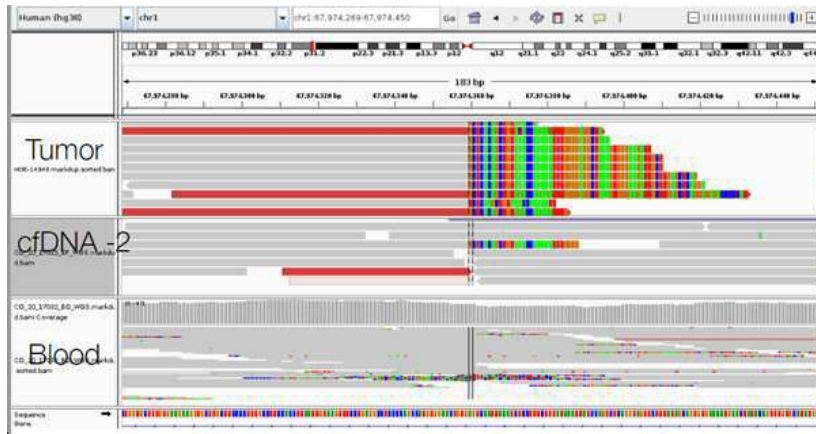
도면3



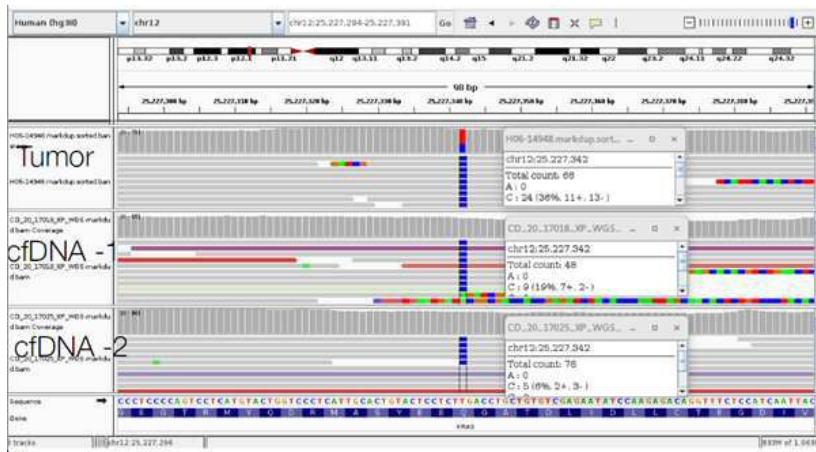
도면4



도면5



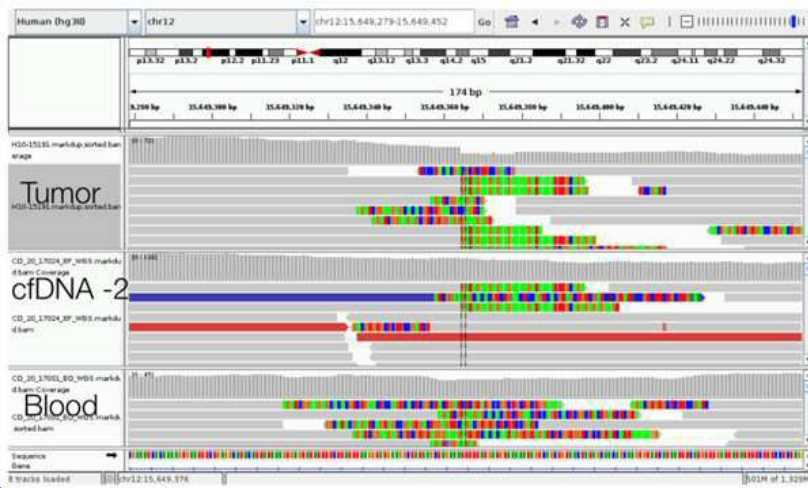
도면6



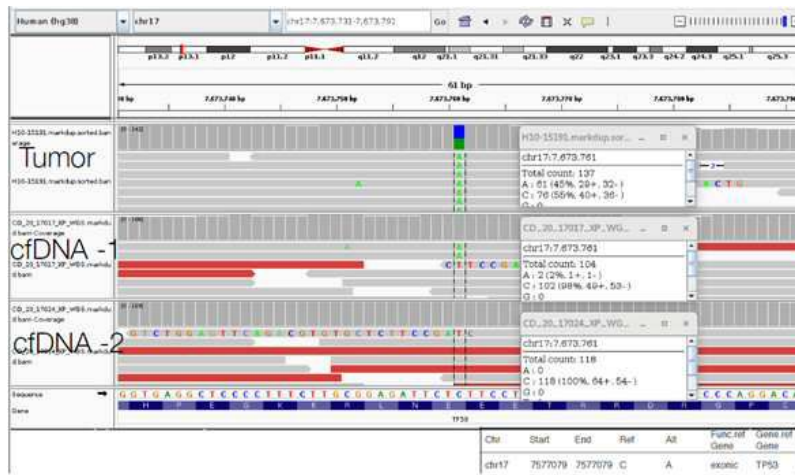
도면7



도면8



도면9



도면10

