



지스트(광주과학기술원) 보도자료

<http://www.gist.ac.kr>

보도시점	배포 즉시 보도 부탁드립니다.	
배포일	2021.05.03.(월)	
보도자료 담당	홍보팀 조동선 팀장	062-715-2061
	홍보팀 이나영 선임행정원	062-715-2062
자료 문의	전기전자컴퓨터공학부 이현주 교수	062-715-2213

암 세포의 유전 변이 발굴 및 유전체 복원 알고리즘 개발

- 암 환자의 개인 맞춤 치료에 기여

- 지스트(광주과학기술원) 전기전자컴퓨터공학부 이현주 교수 연구팀은 전장유전체 데이터*를 분석하여 유전 변이를 발굴하고, 단일염기서열 수준으로 유전체 구조를 복원하는 그래프 기반의 새로운 알고리즘을 개발하였다.
- 연구팀은 변이 검출 오차가 많은 전장유전체 데이터에서 정확하게 유전 변이를 발굴하고, 암 환자에게서 기존의 방법으로 발견되지 않았던 재배열된 유전체 구조를 파악하는 데 성공했다.

*전장유전체 데이터(Whole Genome Sequencing data): 개별 개체의 전체 DNA의 염기 서열을 제공하는 데이터

- 인간의 유전체는 30억 개의 염기서열로 구성되어 있으며, 암 세포의 경우 정상 세포와는 다른 유전 변이가 존재한다. 개개인의 암에 따른 서로 다른 유전 변이를 정확히 파악해 내는 것은 개인 맞춤 치료에 있어 중요하다.
- 그러나 30억 개의 염기서열을 분석하여 암 세포가 갖고 있는 변이를

정확히 파악하는 것은 매우 어려운 과제이며, 특히 암 세포에서 과거에 현미경으로 관측되어 온 재배열된 염색체 구조는 아직 단일염기서열 수준으로 파악된 적 없다. 따라서 전장 유전체를 분석하여 이를 파악할 수 있는 알고리즘이 필요하다.

- 본 연구팀은 유전 변이 발굴 및 유전체 복원 알고리즘인 InfoGenomeR(Integrative Framework for Genome Reconstruction, 인포지노머)를 개발하여 구조 변이를 가진 염기 서열을 그래프 형태로 변환한 뒤, 구조 변이와 복제수 변이가 서로 일관성 있는 값을 가지도록 그래프를 재구성함으로써 검출 오차를 줄였다.
 - 이후, 이형접합 단일염기다형성(heterozygous single nucleotide polymorphism) 정보를 이용하여 하프로타입(haplotype) 그래프를 구성한 후, 최소 엔트로피(entropy)값을 갖는 오일러 경로(Eulerian path)를 찾음으로써 유전체 배열 형태를 복원하였다.
- 연구팀이 개발한 InfoGenomeR(인포지노머)는 유전 변이 검출 오차를 크게 줄였으며(InfoGenomeR의 구조 변이 검출 정밀도 98.1% 및 F-measure 94.9%), 암 세포주의 유전체 배열 형태를 단일염기서열 수준으로 복원하였다.
 - 유전 변이 검출 정확도는 국제적인 유전체 분석 기업인 일루미나(Illumina)의 알고리즘 만타(Manta)에 비해서 크게 향상된 수준을 확인하였다(manta의 구조 변이 검출 정밀도 94.2% 및 F-measure 90.4%).
- 연구팀은 개발한 알고리즘을 유방암 및 뇌암 환자 데이터에 적용하여 환자에게서 원형 유전체 구조(double minute)가 수십 배 증폭되어 있는 것을 파악하였으며, 이는 암 유형마다 특정 염색체가 재배열되는 과정을 동반함을 밝혀냈다. 또한, 재발하거나 전이된 암에서 기존에 존재한 암 부위에는 없었던 재배열된 유전체가 새롭게 나타남을 밝혀냈다.

- 이현주 교수는 “전장유전체 데이터만을 이용하여 암 세포의 유전체 배열 형태를 단일염기서열 수준의 복원하는 것은 도전적인 문제이고, 기존 알고리즘으로는 가능하지 않았지만, InfoGenomeR는 이를 성공적으로 수행한 첫 번째 알고리즘이다” 면서 “본 알고리즘의 결과로 나온 암 세포의 유전체 배열 형태를 기반으로 개인 맞춤 의료를 위한 암 관련 유전자의 발현 조절을 규명할 수 있기를 기대한다” 고 말했다.

- 지스트 이현주 교수와 석박사 통합과정의 이영훈 학생이 수행한 이번 연구는 한국연구재단의 개인연구지원사업(중견연구)과 정보통신기획평가원의 SW컴퓨팅산업원천기술개발사업의 지원으로 수행됐으며, 연구 성과는 세계적인 학술지인 ‘네이처 커뮤니케이션즈(Nature communications)’ 에 2021년 4월 29일에 온라인 게재되었다. <끝>

논문의 주요 내용

1. 논문명, 저자정보

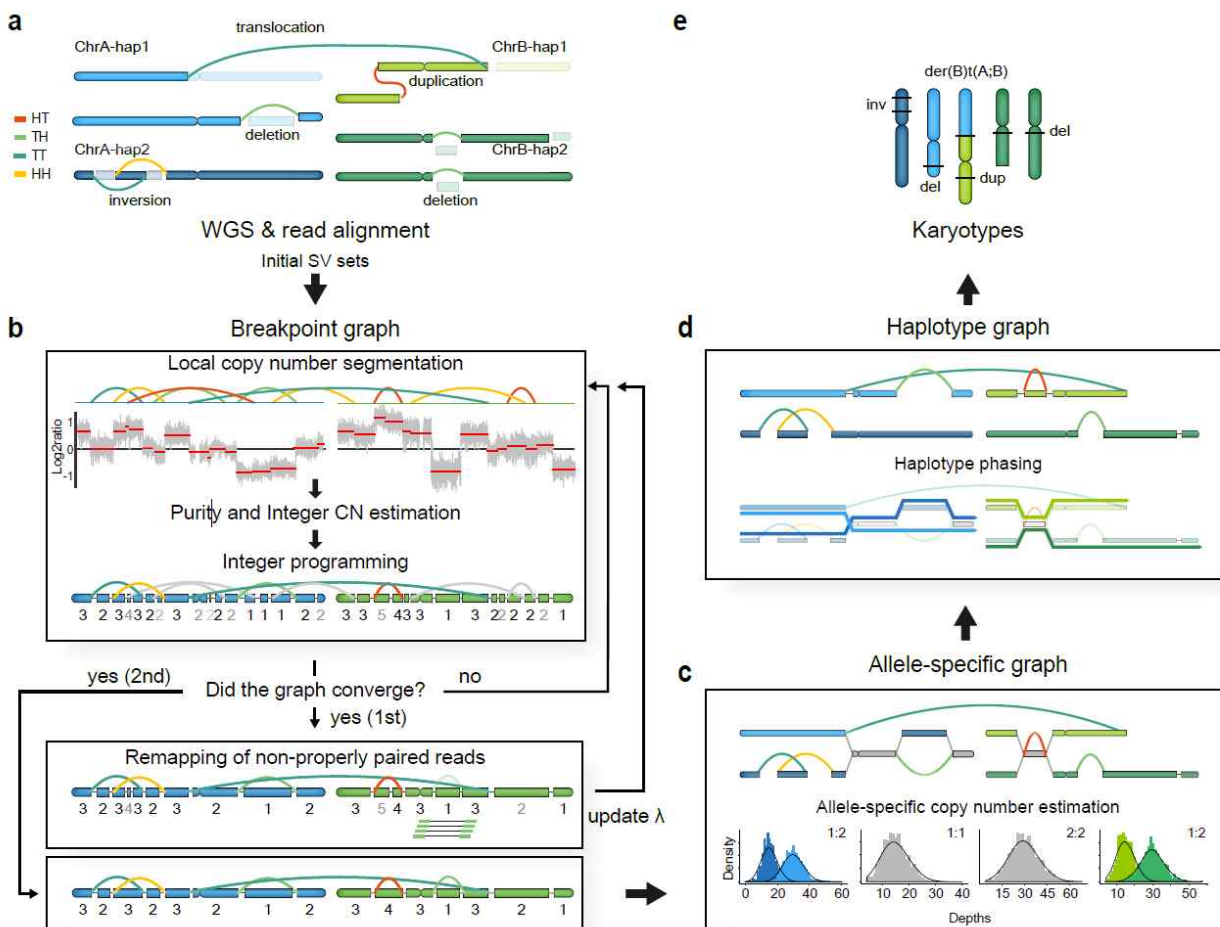
- 저널명: Nature Communications (IF : 12.121, 2020년 기준)
- 논문명 : Integrative reconstruction of cancer genome karyotypes using InfoGenomeR
- 저자 정보 : 이영훈 (제1저자, 전기전자컴퓨터공학부 석박사 통합과정생),
이현주 교수 (교신저자, 전기전자컴퓨터공학부)

용어 설명

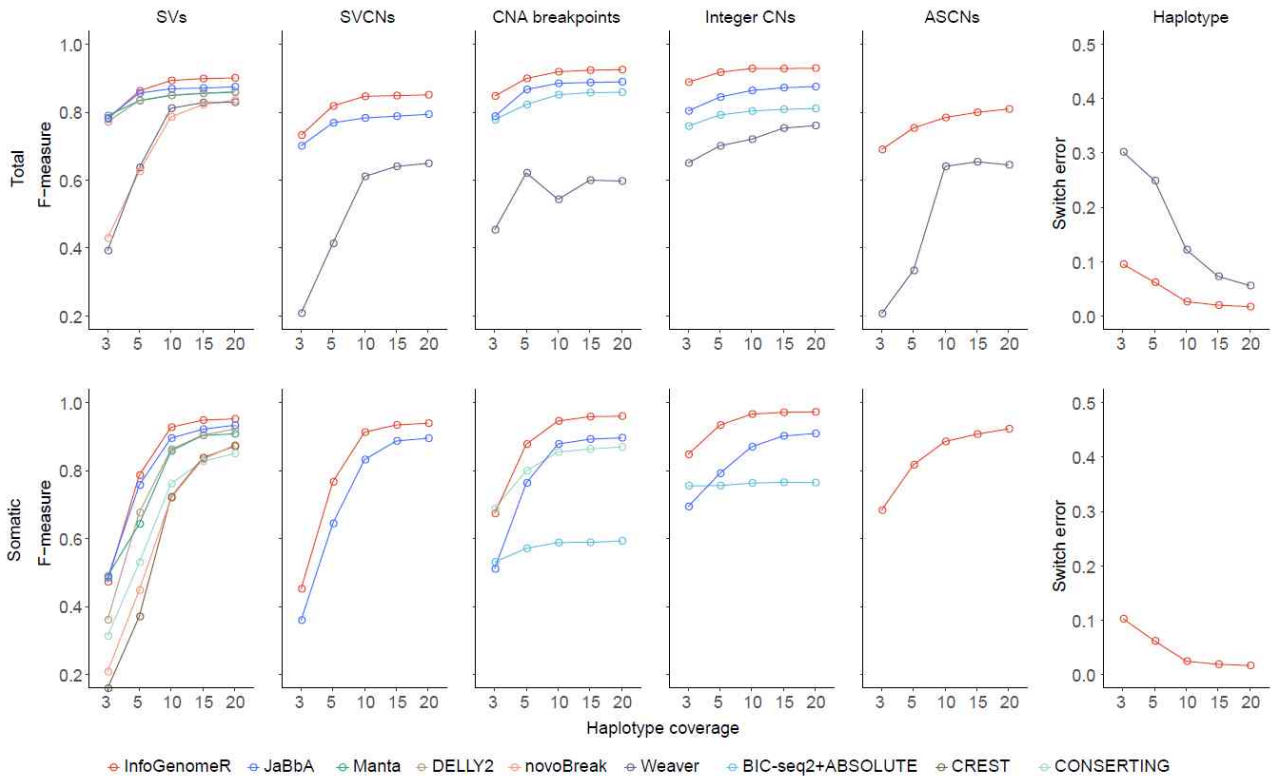
1. 염색체 구조 변이 (Structural Variation)

- 개체 염색체의 구조를 변화시키는 삽입(insertion), 삭제(deletion), 중복(duplication), 역위(inversion), 전좌(translocation) 등의 변이

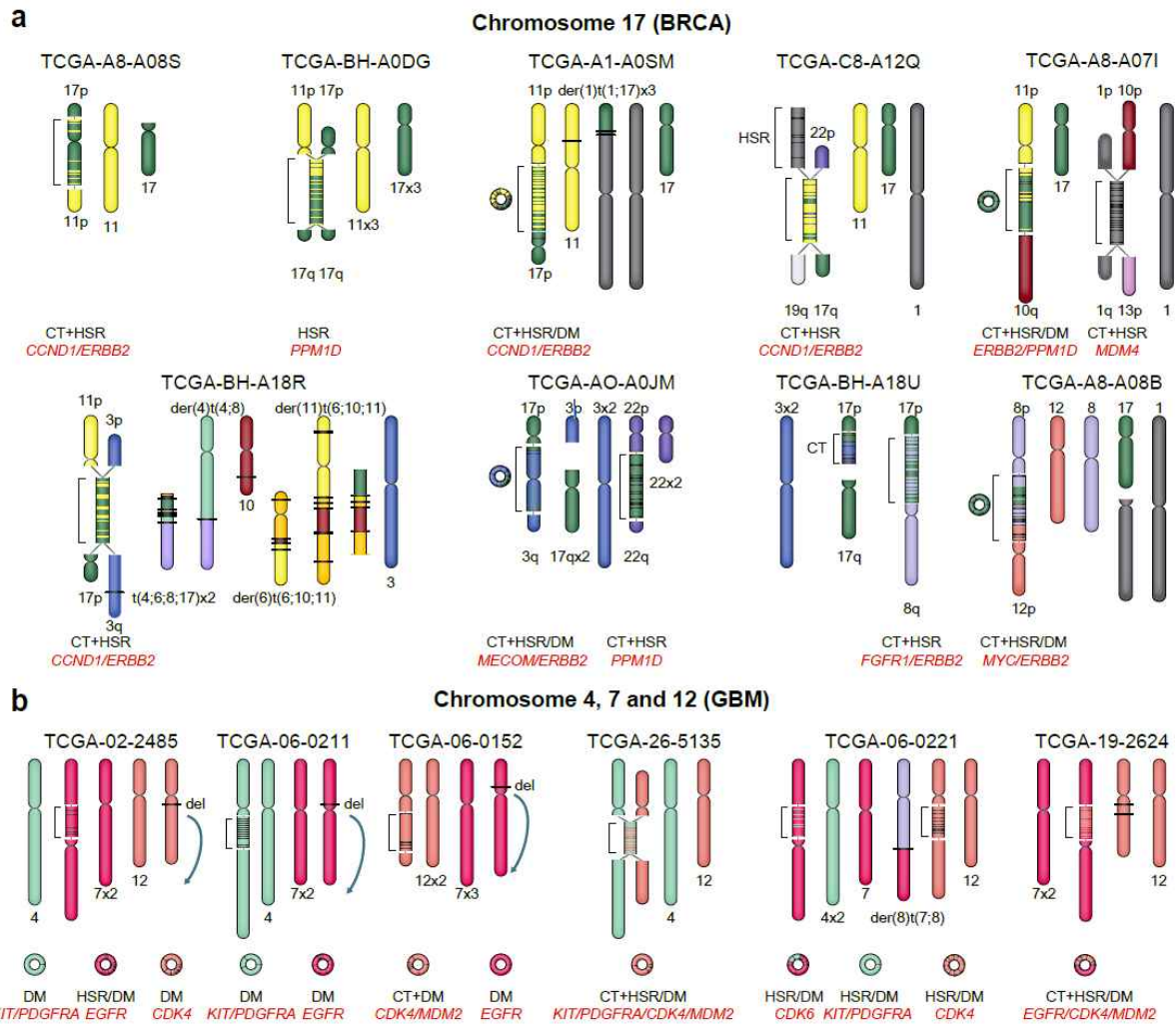
그림 설명



[그림 1] 유전 변이 발굴 및 유전체 복원 알고리즘 (InfoGenomeR)의 개략도. 해당 알고리즘은 전장유전체 데이터를 분석하여 변이를 발굴하고 최종 유전체 구조를 밝혀낸다.



[그림 2] 개발한 알고리즘 (InfoGenomeR)의 정확도. 해당 알고리즘은 다양한 유형의 유전 변이를 통합적으로 발굴하며 높은 정확도를 보인다.



[그림 3] 유방암 (BRCA) 및 뇌암 (GBM)에게서 파악한 유전체 구조. 암 별로 특정 염색체가 재배열되어 있다. 수십 배 증폭되어 있는 원형 염색체 구조가 반복적으로 발견되며 이 염색체는 주요 암 유발 유전자를 갖고 있다.