

《 인간 게놈 프로젝트 자료 》

하이에듀

주제	인간 게놈 프로젝트의 역사와 유전체 연구에 사용되는 컴퓨팅 기술
요약	인간 게놈 프로젝트의 역사, 최근 유전체 연구에 사용되는 컴퓨팅 기술에 대한 자료입니다. 자료1과 자료2를 참고해주세요. 추가 자료가 필요하시다면 언제든지 회신 부탁드립니다. :)

[자료1] 1. 인간 게놈 프로젝트의 역사

요약:

인간 게놈 프로젝트란 1980년대 말 선진국들이 결성한 프로젝트로, 2001년에 완료되었고 인간 게놈 99%의 지도를 완성하였으나 유용성의 문제 등 한계가 있었다. 기존 인간 게놈 프로젝트의 한계를 보완하기 위해 1000 유전체 프로젝트가 2008년에 시작되었고, 유전자변이를 기준으로 환자들과의 비교를 통해 다양한 질환특이적 유전자 변이를 밝혀낼 수 있게 되었다. 또한 미국 국립 인간 유전체 연구 기관에서는 ENCODE 등의 프로젝트를 진행중이기도 하다.

인간게놈프로젝트(Human Genome Project; HGP)는 1980년대 말 미국 국립보건원과 유럽, 일본 등 선진국들이 참여해 인체 게놈 사업기구(HUGO)라는 별도의 국제 학술회의가 결성되어 시작된 프로젝트이다. 2001년 2월 11일에 완료된 인간게놈프로젝트는 성인 남성 1명의 DNA 샘플을 이용해 인간 게놈 99%의 지도를 완성하였지만 샘플 수가 적어 유용성의 문제가 끊임없이 제기되어 왔다. 또한 인간 지도를 완성하게 되면 유전자 정보를 분석하여 질환 치료에 많은 도움을 줄 것이라는 기대와는 다르게 질병 치료에 실제적인 큰 도움을 주지는 못했다. 기존 인간게놈프로젝트의 한계를 보완하기 위하여, 1000명 을 대상으로 인간 유전체를 다시 자세히 분석해 보자는 취지의 1000 유전체 프로젝트(1000 Genomes Project)가 2008년에 시작되었다. 2001년도에 완성된 게놈 프로젝트의 DNA 염기 해독 기술을 그대로 사용하면 이번 프로젝트의 비용은 대략 5억 달러의 비용이 소요되지만 발전된 최신의 해독기술과 효율적인 관리방법으로 1000 유전체 프로젝트의 비용은 급속히 떨어졌다.

전 세계 75개 대학과 기업들의 연구자들이 결성한 컨소시엄인 1000 유전체 프로젝트의 성과는 '네이처'지에 'A map of human genome variation from population-scale sequencing'라는 제목으로 발표되었다. 이 프로젝트의 첫 단계에서는 서아프리카, 유럽, 중국과 일본에 살고 있는 179명의 개인 유전체의 염기서열을 완전히 분석하는 작업과, 부모와 자식 한 명으로 이루어진 표본 집단의 두 가족을 대상으로 유전체 염기 서열을 분석하는 작업, 그리고 697명의 엑손(exon)부분의 염기서열을 분석하는 것으로 이루어졌다. 연구자들은 이번 프로젝트에서 발굴된 유전자변이를 기준으로 하여 환자들과의 비교를 통해

다양한 질환특이적인 유전적 변이를 밝혀 낼 수 있을 것이다. 또한 서로 다른 인종들의 유전체 다양성을 비교함으로써 진화과정에서의 중요 역할을 하는 유전적 특징을 알아낼 수 있는 기반을 마련할 수 있게 되었다.

위와 같은 염기 서열과 연관된 인간유전체 연구의 진행과 더불어 미국 국립 인간 유전체 연구 기관 (National Human Genome Research Institute; NHGRI)에서는 유전체 지도의 완성뿐 아니라 그 기능을 밝히고 각 요소들의 역할을 이해하기 위해 ENCODE(ENCyclopedia Of DNA Elements) 프로젝트를 지원하고 있다.

[자료2] 인간 게놈 프로젝트에 사용되는 슈퍼 컴퓨팅

인간 게놈 프로젝트 자체에 사용되는 슈퍼 컴퓨팅보다는, 최근 연구 동향을 보았을 때 유전체 연구에 자주 사용되는 컴퓨팅 기술을 모아보았습니다.

1. 클라우드 컴퓨팅 기술

지난달 24일(현지 시간) 구글은 미국국립보건원(NIH)과 함께 새로운 이니셔티브를 결성하고, 글로벌 의생명 데이터 생태계를 구축하기로 했다고 밝혔다. 굴지의 정보기술(IT) 기업이 의생명 분야에 관심을 갖게 된 건 유전체(게놈) 해독부터 대사체 분석, 유전자 표현형 분석에 이르기까지 매일같이 연구실에서 쏟아져 나오는 데이터의 양이 어마어마하기 때문이다. 미국 국립생물공학정보센터(NCBI)에 따르면 7개월마다 그 양이 2배로 될 정도다.

이처럼 방대한 양의 데이터를 수용할 수 있는 대안으로 최근 인터넷 기반의 '클라우드 컴퓨팅'이 주목받고 있다. 클라우드 컴퓨팅은 정보를 사용자의 컴퓨터가 아닌 인터넷으로 연결된 다른 컴퓨터로 저장하거나 처리하는 기술이다. 클라우드 컴퓨팅을 활용하면 개인 컴퓨터의 한계를 넘어 아무리 큰 데이터도 안전하게 영구 저장이 가능할 뿐만 아니라 여러 명과 쉽게 공유할 수 있고 분석 시간도 3분의 1 이상 단축할 수 있다.

그동안 이미 구글은 유전체 데이터 분석 전용 클라우드 플랫폼인 '구글 지노믹스(Google Genomics)' 서비스를 운영해 왔다. PB(페타바이트·1PB는 약 100만 GB)급 유전체 데이터를 수초 내에 처리할 수 있다는 게 구글 측의 설명이다. 이를 바탕으로 NIH 연구 프로젝트에 참여하는 전 세계 2500여 개 기관 연구자들이 생성하는 연구 데이터를 한데 모아 구성원 누구나 자유롭게 내려받고 손쉽게 분석할 수 있는 시스템을 구축할 계획이다. 다이앤 그린 구글 클라우드 대표는 "연구자들이 클라우드 환경에서 무료로 활용할 수 있는 오픈 소스 툴도 개발 중이다. 앞으로는 인공지능(AI) 기술도 접목할 계획"이라고 밝혔다.

유전체 분석 분야에 뛰어드는 기업도 점차 늘고 있다. 아마존과 마이크로소프트 역시 '아마존웹서비스(AWS)'와 '애저(Azure)' 등 클라우드 컴퓨팅을 기반으로 비슷한 서비스를 제공하고 있다. 국내에서도 KT, 네이버 등이 제약사 등과 손을 잡고 유전체 빅데이터 사업을 벌이고 있다.

방대한 데이터 양도 양이지만 클라우드 컴퓨팅이 유전체 연구 분야에서 강력한 도구가 될 수 있는 이유는 전 세계에 흩어져 있는 여러 명의 연구자들이 연구자가 함께 이용할 수 있다는 장점 덕분이다. 보통 유전체 연구는 다양한 비교 데이터를 얻기 위해 여러 명의 연구자가 공동으로 수행한다. 2015년 '1000게놈 프로젝트'에 참여한 유럽생물정보학연구소, 미국 워싱턴대 등 국제 공동 연구진은 구글 지노믹스, AWS 등을 동원해 26개 인구집단 2504명의 유전정보를 총망라한 유전체 지도를 국제학술지 '네이처'에 발표했다.

미국과 프랑스, 캐나다, 한국 등 세계 17개국 2000여 명의 과학자가 위암, 폐암, 백혈병 등 50종에 이르는 암의 유전적 특성을 밝히기 위해 출범한 '국제 암 유전체 컨소시엄(ICGC)'의 연구 프로젝트에는 한국전자통신연구원(ETRI)이 개발한 클라우드 기반 유전체 분석용 고성능 슈퍼컴퓨팅 기술도 활용됐다. ICGC는 이미 세계 암 환자 2만5000명의 유전체(게놈)에 나타난 유전자 변이를 분석했으며, 2025년까지 20만 명의 임상 데이터를 확보한다는 목표를 갖고 있다. 최완 ETRI 책임연구원은 "클라우드 컴퓨팅 덕분에 연구자들이 공간적, 시간적 제약을 받지 않고 연구를 할 수 있는 셈"이라고 말했다.

(후략)

2. 양자컴퓨터의 활용

최근 다이애그노믹스가 세계 유일의 양자컴퓨터 제조사인 캐나다의 디웨이브(D-Wave)와 합작법인인 'DNA-Seq'라는 회사를 만들었다는 것. 인간게놈분석기술과 양자컴퓨터의 계산 능력을 결합해 새로운 차원의 개인게놈시대를 연다는 비전을 공유했다고 한다. 먼 미래의 이야기인줄로만 알았던 양자컴퓨터가 어느새 현실이 돼 우리 눈앞에 불쑥 모습을 드러내고 있는 셈이다.

이들이 추진하는 프로젝트를 요약하면 이렇다. 인간게놈에 500여개 있는 인산화효소(kinase) 유전자에 돌연변이가 생기는 게 암 발생의 주요 원인인데, 환자의 게놈분석으로 변이가 생긴 유전자를 확인한 뒤 양자컴퓨터로 변이 유전자가 만들어내는 변이 인산화효소의 구조를 예측한 뒤 이를 표적으로 한 약물을 찾는다는 것. 한마디로 개인게놈과 양자컴퓨터를 이용한 환자 맞춤형 항암치료인 셈이다.

(중략)

양자계산의 가능성은 이미 수십 년 전부터 이론물리학자들 사이에서 논의돼 왔고 알고리즘도 여럿 개발됐다. 문제는 이를 구현할 '실물' 양자컴퓨터가 없었던 것. 따라서 양자계산은 튜링머신처럼 관념의 영역에 속했다. 양자계산의 원리를 잠깐 소개하면(물론 필자도 이해하지는 못하지만), 기존의 디지털컴퓨터가 '0 또는 1'이라는 정보단위, 즉 비트(bit)를 바탕으로 작동하는데 반해 양자컴퓨터는 '0과 1'이라는 양자 비트, 즉 큐비트(qubit)를 단위로 한다. 직관에 반하는 양자역학의 세계에서는 양자상태의 중첩이라는 현상이 있고 따라서 큐비트가 10개면 2의 10승, 즉 1024가지 계산이 동시에 이뤄지고 100개면 2의 100승

이라는 어마어마한 경우의 수를 소화할 수 있다.

‘네이처’ 기사에 따르면 올해 41살인 디웨이브의 조르디 로즈 대표는 공대를 졸업하고 캐나다 브리티시컬럼비아대학에서 이론물리학으로 박사학위를 할 때 기술벤처투자전문가인 해이그 패리스의 강의를 듣고 양자컴퓨터를 만들어야겠다는 열정에 휩싸였다고 한다. 1999년 패리스가 꺾준 돈 4059.50 캐나다달러(약 400만 원)로 컴퓨터와 프린터를 산 로즈는 디웨이브란 회사를 차리고 투자자를 모았고, 이듬해 수백만 달러(수십억 원)를 확보해 연구팀 15곳에 투자해 양자컴퓨터의 가능성을 모색했다.

2001년 학술지 ‘사이언스’에는 미국 MIT 이론물리학센터 에드워드 패리 박사팀의 양자계산 알고리즘에 관한 논문이 실렸다. 연구자들은 **단열양자계산(adiabatic quantum computing)**이란 방법을 쓰면 **최적화문제(optimization problem)**를 푸는데 기존 디지털 컴퓨터보다 훨씬 뛰어나다는 걸 시뮬레이션으로 보인 것.

[참고문헌]

- [1] 인간 유전체 기능연구의 소개, 질병관리본부
https://www.kdca.go.kr/filepath/boardDownload.es?bid=0034&list_no=12618&seq=1
- [2] 수만명의 유전체 정보를 한곳에... ‘게놈 클라우드’가 뜬다, 동아사이언스
<https://www.dongascience.com/news.php?idx=23390>
- [3] 인간게놈이 양자컴퓨터를 만났을 때, 동아사이언스
<https://www.dongascience.com/news.php?idx=2054>