

유전체 빅데이터 공유로 가치 창출해야

이 병 욱

한국생명공학연구원
책임연구원



많은 동양인들은 술을 먹으면 얼굴이 빨개지는 반면 유럽의 서양인들은 이러한 경우가 매우 드물다. 또한 동양인은 오랜 기간 우유를 마셨던 서양인에 비해 우유를 잘 소화하지 못한다. 이는 동양인이 서양인과 비교해서 알코올 대사의 중간 산물인 아세트알데하이드를 분해하는 효소와 우유의 젖산 분해에 필요한 효소인 락타아제가 부족하기 때문이다.

동양인이 술과 우유 분해를 잘하지 못하는 근본적인 원인은 무엇일까? 그것은 동양인 세포에 존재하는 염색체에 그러한 정보가 없기 때문이다. 인간은 60조 개 이상의 세포들로 구성되어 있다. 세포는 부모로부터 물려받은 46개(23쌍)의 염색체를 가지고 있으며, 이들 염색체는 약 3.2억 개 DNA 염기(A, C, G, T)들로 구성되어 있다. 인간의 경우, 염색체에는 약 2만 개의 유전자들이 있으며, 유전자들이 가진 ‘유전정보’가 세포와 인간의 모든 것을 결정한다.

암은 유전자의 염기에 돌연변이가 생겨서 일어난다는 것을 알게 되었다. 따라서 이러한 돌연변이가 일어난 위치와 변이 정보를 알면 암을 예방하거나 치료를 할 수 있다는 것도 알게 되었다. 미국의 유명 여배우 안젤리나 졸리는 2013년 유전자 검사를 통해 자신이 대표적인 암 억제 유전자인 ‘BRCA1’에 돌연변이를 가지고 있어서 유방암에 걸릴 확률이 매우 높다는 것을 알게 되었다. 그녀는 유방암 예방을 위해 유방을 모두 절제했으며, 이를 통해 암 발생 확률을 상당히 줄일 수 있었다. 애플 창업자인 스티브 잡스 역시 췌장암 치료를 목적으로 약 1억원 이상 비용을 들여 자신의 개인 유전체 서열을 분석하여 암의 원인이 되는 돌연변이를 찾아서 치료하려고 시도하였다. 그의 시도는 아쉽게도 무위로 돌아갔고 2011년 운명을 하였다.

우리는 인간이 가진 모든 유전정보를 알기 위해서 다년간 노력한 결과, 2000년대 초 인간 DNA 염기서열을 정확히 해독한 유전체 지도 초



2000년대에 이르러 인간 DNA 염기서열을 해독한 유전체 지도 초안이 발표됐다.

(사진=픽사베이)

“

사람 한 명의

유전체 정보의 양은

약 250기가바이트다.

이러한 데이터를 분석하고

해석하는 국가적 기술이

필요하다.

”

안을 발표하였다. 이는 유전체 시대 서막을 알리는 역사적인 전환점이었다. 당시 한 명의 유전체 해독을 위해 10년이라는 기간과 3조원 비용이 소요됐으며 이는 유전체 대중화를 막는 장애물이었다. 하지만 2005년 새롭게 등장한 차세대 유전체 서열결정(NGS) 기술과 이후 급속한 기술발달은 우리를 ‘100달러 개인 유전체’ 시대로 접어들게 했다. MRI나 초음파 진료와 유사한 수준 비용으로 유전체 정보를 얻을 수 있게 됐다.

유전체 해독 비용의 급격한 감소로 인류는 유전체 빅데이터라는 새로운 시대를 맞이하게 됐다. 이처럼 바이오 분야에서 대량의 데이터들이 쏟아지면서, 대용량의 데이터로부터 새로운 지식을 생산하는 빅데이터 시대가 도래하였다. 이로 인해, 바이오는 실험과 이론 중심에서 데이터가 중심이 되는 과학으로 패러다임이 변화하고 있다. 더 나아가 바이오 빅데이터에 인공지능 기술이 접목되면서 신산업의 기회들이 창출되고 있다.

바이오 빅데이터 시대의 대표적인 신산업은 무엇보다도 개인의 유전체 정보와 건강정보를 토대로 최적화된 진단과 치료를 제공하는 정밀의료 산업이다. 기존 의학이 동일

진단을 받은 환자에게 동일한 치료법을 적용하는 평균적 환자에 맞춰진 방법인 반면, 정밀의료는 개인의 유전정보의 특징에 따른 맞춤형 치료법을 적용하는 방법이다.

또한 유전체 정보와 일상생활에서 수집할 수 있는 라이프로그 정보를 기반으로 건강관리를 위한 맞춤형 헬스케어 서비스도 가시화되고 있다. 여러 신생 벤처들이 바이오 빅데이터에 인공지능 기술을 적용함으로써, 신약 개발에 걸리는 시간과 비용을 혁신적으로 단축시키는 쾌거를 이루기도 했다. 인간에게 유익한 새로운 품종을 개발하는 동식물의 육종도 바이오 빅데이터를 기반으로 더욱 가속화 될 것이다.

이러한 새로운 흐름에 대비하기 위해 어떤 준비를 해야 하는지 생각해 보자. 우선 가장 시급한 것은 빅데이터를 저장하고 분석하고 해석하는 기술 개발이다. 사람 한 명의 유전체 정보량은 약 250기가바이트(GB)며, 1만 명이면 2.5페타바이트(PB) 데이터가 생산된다. 이러한 데이터를 개인 연구자 실험실에서 분석하는 것은 거의 불가능하다. 국가 차원의 유전체 분석 기술을 개발해 이를 클라우드 컴퓨팅으로 연구자들에게 제공해야 한다.

유전체 분야 기술력과 심각한 전문 인력 부족을 개선해야 한다. 유전체 정보 생산·공유, 분석, 인공지능(AI)을 이용한 응용, 정밀의료, 신약개발, 동식물 육종 등 각 분야별 독자 기술력을 갖춰야 한다. 그러기 위해서는 이러한 빅데이터를 다룰 수 있는 생명정보학(Bioinformatics) 연구자의 배출이 시급하다. 유전체 선진국 기술을 따라서는 영원히 유전체 선진국이 될 수 없다. 생명정보학자 양성을 위해 장기적인 대규모 투자가 이뤄져야 한다.

우리나라는 유전체 연구에 많은 연구비를 투입하지만, 사업이 종료되면 대부분 생산정보는 사장되고 있다. 데이터는 한곳에 모여야 비로소 빅데이터로 능력을 발휘하게 된다. 이를 위해 외국과 같이 연구자들이 자발적으로 자신의 데이터를 공유하는 문화가 확산되어야 한다.

유전체 분야는 블록체인, AI, 빅데이터 분석 등이 융합되고 바이오 빅데이터와 결합될 때 무궁무진한 활용성을 띤다. 복잡하고 이질적인 빅데이터를 어떻게 생산, 수집, 관리, 통합 분석해 새로운 가치를 창출할 것인가는 4차 산업혁명 시대를 맞이한 바이오 분야의 숙제이다.