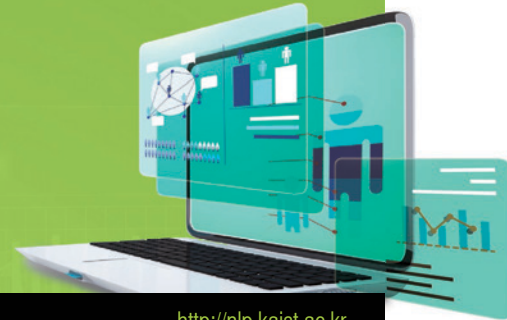


# 암과 관련있는 유전자를 찾아내는 검색툴, 온코서치

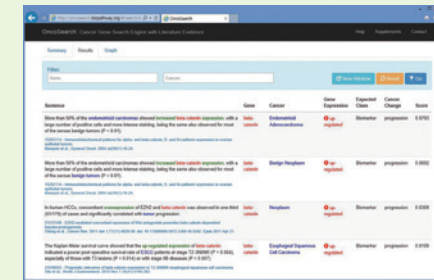


전산학부 박종철

<http://nlp.kaist.ac.kr>



빅데이터 기술 기반의 온코서치는 웹을 통한 공개적인 사용이 가능하여 전 세계의 의학/생물학/생물정보학 연구자들이 암을 포함한 다양한 복잡 질환 연구를 진행하는 과정에 대용량 문헌 정보를 효율적으로 통합하여 활용할 수 있도록 할 것이다. 이러한 통합 활용으로 질병 기작에 대한 연구자들의 이해도를 더욱 높이고, 새로운 진단 및 치료 기법 개발 등 복잡질환에 대한 연구가 급속하게 진전될 것이다. 또한, 수집된 대용량 문헌 데이터에 확률적 분석 및 이종적인 추론 기법을 적용한 유전자 네트워크 자동 추론 시스템을 토대로 복잡 질환 연구에 직접적으로 도움이 되는 새로운 지식 발견이 가능할 것이다.



YTN 사이언스투데이 언론 보도 자료(2014.05.22.)

암은 유전자의 이상 변화가 주요한 발생원인 중 하나로 알려져 있다. 따라서 암 연구에서는 암 발생에 관련된 유전자들을 파악하고 이들 유전자들이 암 발생 기작에 어떻게 기여하는 지를 잘 이해하는 것이 매우 중요하다. 이에 따라 각종 암 관련 유전자에 관하여 수많은 연구가 진행되었으며, 그 결과가 연구 문헌 데이터베이스로 축적되고 있다. 암 연구 진행을 위해서는 이러한 기존 연구 결과들을 효율적으로 수집 및 분석하는 것이 필수적으로 요구되지만, 관련 연구 문헌의 양이 너무 방대하기 때문에 빅데이터 환경의 고급 텍스트 마이닝 기술을 기반으로 하는 검색 엔진의 활용이 중요하다.

온코서치는 생물의로 텍스트 상의 유전자-암 간의 관계를 찾는 과정에 1) 유전자가 어떻게 변화하는지, 2) 암이 어떻게 변화하는지, 3) 유전자와 암 간의 인과 관계는 어떻게 정의되는지에 대한 정보를 고려하며 암에 대한 유전자의 역할을 추론하여 파악한다.

이러한 유전자의 역할 별 분류과정을 통하여 생물 데이터베이스에 이미 등록이 되어 있는 발암유전자(암을 일으키는 유전자), 암억제유전자(세포를 암으로부터 보호하는 유전자) 및 바이오마커(정상 및 암 상태를 구분하는데 사용할 수 있는 유전자)를 온코서치가 정확하게 구별할 수 있다는 것을 세계 최초로 보였다. 이와 관련하여 현재 표준으로 삼고 있는 유전자 데이터베이스의 하나인 UniProtKB 상에 나타나는 발암유전자의 6.87%, 암억제유전자의 3.76% 정도의 소규모 정보만 Vogelstein의 팀이 2013년 Science 학술지에 발표한 발암유전자 및 암억제유전자 리스트에 포함되어 있다는 것을 확인하였다.

이러한 사실은 새로운 발암유전자나 암억제유전자를 파악하는 과정이 아직 연구의 초기 단계에 머무르고 있거나 이들 유전자에 대한 정확한 정의가 생물 데이터베이스에 따라 달라진다는 것을 의미하며, 온코서치와 같은 특화된 웹툴을 통하여 현재 까지 알려진 전세계의 암 관련 연구 성과를 시기적절하게 조망할 수 있는 방법을 긴요하게 한다.

빅데이터 플랫폼에 기반한 온코서치는 빠르게 증가하는 방대한 문헌 자료로부터 새로운 생물학 지식을 추출 및 추론하여 중앙학을 비롯한 다양한 복잡 질환 연구에 중요한 촉매 역할을 할 것으로 예상된다.

**연구실적** | YTN, 동아일보를 비롯한 언론보도 25건 이상(2014년)  
| 국립암센터 발간 웹진 암연구동향, 한국과학창의재단 발간 월간 과학창의, 한국연구재단 발간 웹진 심층 취재 기사 보도(2014년)

**참고자료** | 논문: H.-J. Lee, T. C. Dang, H. Lee and J. C. Park, Nucleic Acids Research, 42(W1): W416-W421(2014). (Impact Factor = 8.278)

**연구지원** | 한국연구재단(중견연구자사업 핵심공동연구, 도약연구 성과확산 사업)