

(12) 특허협력조약에 의하여 공개된 국제출원

(19) 세계지식재산권기구
국제사무국

(43) 국제공개일
2020년 5월 7일 (07.05.2020)

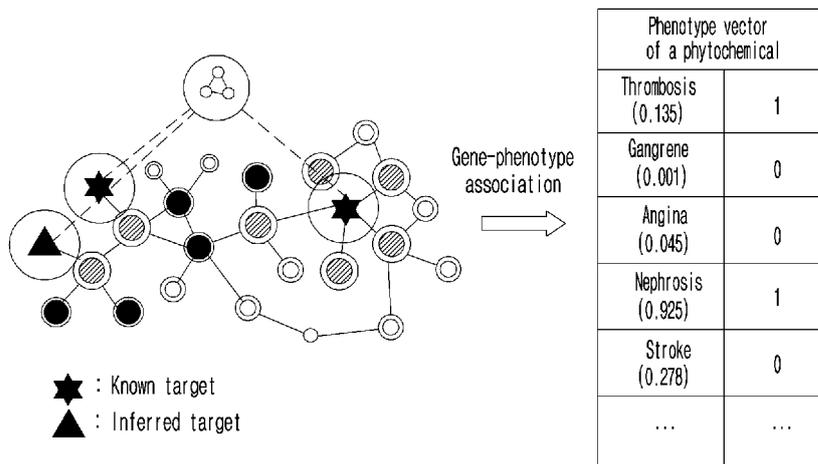


(10) 국제공개번호
WO 2020/091185 A1

- (51) 국제특허분류: *G16H 50/30* (2018.01) *G16H 50/70* (2018.01) *G16H 20/90* (2018.01) *G16C 20/50* (2019.01)
- (21) 국제출원번호: PCT/KR2019/008244
- (22) 국제출원일: 2019년 7월 4일 (04.07.2019)
- (25) 출원언어: 한국어
- (26) 공개언어: 한국어
- (30) 우선권정보: 10-2018-0130202 2018년 10월 29일 (29.10.2018)KR
- (71) 출원인: 재단법인 전통천연물기반 유전자동의보감 사업단 (BIO-SYNERGY RESEARCH CENTER) [KR/KR]; 34141 대전시 유성구 대학로 291 정문술빌딩 513호, Daejeon (KR). 한국과학기술원 (KOREA ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY) [KR/KR]; 34141 대전시 유성구 대학로 291, Daejeon (KR).
- (72) 발명자: 이도현 (LEE, Doheon); 34141 대전시 유성구 대학로 291, Daejeon (KR). 유선용 (YOO, Sunyong); 34141 대전시 유성구 대학로 291, Daejeon (KR).
- (74) 대리인: 이원희 (LEE, Won Hee); 06132 서울시 강남구 테헤란로 147 성지하이츠 2차 8층, Seoul (KR).
- (81) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 국내 권리의 보호를 위하여): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 지정국 (별도의 표시가 없는 한, 가능한 모든 종류의 역내 권리의 보호를 위하여): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM,

(54) Title: METHOD FOR PREDICTING HEATH EFFECT OF PHYTOCHEMICAL, USING INTEGRATED ANALYSIS BASED ON MOLECULAR NETWORK, CHEMICAL PROPERTY, AND ETHNOPHARMACOLOGICAL EVIDENCE, AND SYSTEM THEREFOR

(54) 발명의 명칭: 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 및 이를 위한 시스템



(57) Abstract: The present invention relates to a method for predicting health effects of phytochemicals by using integrated analysis based on molecular networks, chemical properties, and ethnopharmacological evidences, and a system therefor. Specifically, a method for predicting health effects of phytochemicals, using integrated analysis based on molecular networks, chemical properties, and ethnopharmacological evidences according to the present invention does not utilizes individual information separately, but analyzes the information in an integrated manner and as such, can analyze and predict health effects of phytochemicals on a mass scale with high reliability. Therefore, the method and a system therefor can be advantageously used to develop drugs that utilize phytochemicals having useful functions.

(57) 요약서: 본 발명은 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 및 이를 위한 시스템에 관한 것이다. 구체적으로, 본 발명에 따른 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법은 각각의 정보를 개별적으로 이용하지 않고 통합적으로 분석함으로써, 신뢰도 높게 파이토케미컬의 건강효과를 대규모로 분석하여 예측할 수 있으므로, 상기 방법 및 이를 위한 시스템은 유용한 기능을 갖는 파이토케미컬을 이용한 약물 개발에 유용하게 사용될 수 있다.

[다음 쪽 계속]



WO 2020/091185 A1

ZW), 유라시아 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 유럽 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

규칙 4.17에 의한 선언서:

- 신규성을 헤치지 아니하는 개시 또는 신규성 상실의 예외에 관한 선언 (규칙 4.17(v))

공개:

- 국제조사보고서와 함께 (조약 제21조(3))

명세서

발명의 명칭: 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 및 이를 위한 시스템

기술분야

- [1] 본 발명은 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 및 이를 위한 시스템에 관한 것이다.

배경기술

- [2] 파이토케미컬(phytochemical)은 식물의 화학적 경로에 의해 생산되는 화합물을 말하며, 종종 2차 대사물질(secondary metabolite)이라고도 불린다. 최근 연구들은 많은 파이토케미컬들이 인간 세포의 기능에 이로운 역할을 하는 것으로 보고되고 있고, 이러한 파이토케미컬들이 풍부한 식품들이 건강을 증진시킬 수 있음이 보고되어 왔다(Liu, R. H. *et al.*, *Am. J. Clin. Nutr.* 78, 517S-520S, 2003; 및 Mursu, J. *et al.*, *Am. J. Clin. Nutr.* 99, 328-333, 2013).
- [3] 그러나, 파이토케미컬의 효능에 대한 연구는 대부분 *in vitro* 스크리닝 방법을 통해 이루어지고 있어, 다수의 파이토케미컬에 대한 분석을 위해서는 대량 실험이 필요하며, 이는 생산성이 낮고 시간과 비용이 크게 소요된다는 문제점이 있다.
- [4] 이에, 최근에는 주로 분자적 또는 민족약학적(ethnopharmacological) 정보에 기반한 *in silico* 접근 방법이 제안되어 왔다. 분자-기반 접근 방법은 파이토케미컬과 승인된 의약품 간의 분자 구조, 분자적 작용기전, 또는 표적 단백질과 같은 분자적 정보의 유사성에 집중하여 파이토케미컬의 잠재적 효과를 예측하는 방법이다. 그러나, 이러한 접근 방법은 특정 파이토케미컬의 특정 표현형에 대한 효과를 예측하는데 적합하며, 파이토케미컬이 인체에 미치는 전신적인 효과를 예측하기는 어려운 방법이다. 또한, 몇 가지 민족약학적 정보에 기반한 접근 방법이 제안되었으나, 이러한 방법들은 특정 질환 치료 목적의 식물들 또는 파이토케미컬을 선별하기 위한 분자적 분석 또는 *in vitro* 평가 전의 예비적인 수단으로써만 민족약학적 정보를 사용하는데 집중되어 있다. 따라서, 상기와 같은 방법은 대량의 후보물질들 중에서 파이토케미컬들을 선별하는데 유용할 수 있으나, 식물들은 수백 종의 파이토케미컬들을 함유하고 있기 때문에 여전히 생산성이 낮은 문제점이 있다. 따라서, 파이토케미컬의 분자적 정보, 화학적 특성 및 민족약학적 증거를 통합적으로 분석함으로써, 파이토케미컬이 인체에 전신적으로 미치는 건강효과를 예측할 수 있는 새로운 *in silico* 분석 방법의 발명이 필요하다.

발명의 상세한 설명

기술적 과제

- [5] 본 발명의 목적은 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 및 이를 위한 시스템을 제공하는 것이다.

과제 해결 수단

- [6] 상기 목적을 달성하기 위하여, 본 발명은 분자 네트워크를 이용하여 파이토케미컬(phytochemical)의 건강효과(health effect)를 추론하는 단계(단계 1); 파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은 파이토케미컬을 도출하는 단계(단계 2); 및 민족약학적(ethnopharmacological) 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적 증거와 의미 유사도(semantic similarity)가 높은 파이토케미컬의 건강효과를 도출하는 단계(단계 3)를 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법을 제공한다.
- [7] 또한, 본 발명은 상기 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법을 실행시키기 위한 명령들을 포함하는 프로그램이 저장된 컴퓨터로 판독가능한 매체를 제공한다.
- [8] 또한, 본 발명은 분자 네트워크를 이용하여 파이토케미컬의 건강효과를 추론하는 모듈; 파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은 파이토케미컬을 도출하는 모듈; 및 민족약학적 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적 증거와 의미 유사도가 높은 파이토케미컬의 건강효과를 도출하는 모듈을 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 시스템을 제공한다.

발명의 효과

- [9] 본 발명에 따른 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법은 각각의 정보를 개별적으로 이용하지 않고 통합적으로 분석함으로써, 신뢰도 높게 파이토케미컬의 건강효과를 대규모로 분석하여 예측할 수 있으므로, 상기 방법 및 이를 위한 시스템은 건강증진 기능을 갖는 파이토케미컬을 이용한 약물 개발에 유용하게 사용될 수 있다.

도면의 간단한 설명

- [10] 도 1a 내지 도 1c는 본 발명의 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법 과정을 나타낸 도면이다.
- [11] 도 2a 내지 도 2c는 파이토케미컬의 민족약학적 용도를 탐색하는 과정을 나타낸 도면으로 도 2a는 서술적인 표현들로부터 표현형 관련 용어들의 추출하는 과정; 도 2b는 특정 파이토케미컬을 포함하는 식물 도출하는 과정; 및

도 2c는 파이토케미컬의 표현형 네트워크로부터 의미 유사도 분석하는 과정을 나타내는 도면이다.

- [12] 도 3은 분자 네트워크를 기반으로 예측된 파이토케미컬의 건강효과 분포(왼쪽) 및 분자 네트워크 및 민족약학적 증거를 기반으로 예측된 파이토케미컬의 건강효과 분포(오른쪽)를 나타내는 그래프이다.

발명의 실시를 위한 형태

- [13] 이하 본 발명을 상세히 설명한다.

[14]

- [15] 본 발명은 분자 네트워크를 이용하여 파이토케미컬(phytochemical)의 건강효과(health effect)를 추론하는 단계(단계 1); 파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은 파이토케미컬을 도출하는 단계(단계 2); 및 민족약학적(ethnopharmacological) 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적 증거와 의미 유사도(semantic similarity)가 높은 파이토케미컬의 건강효과를 도출하는 단계(단계 3)를 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법을 제공한다.

- [16] 상기 단계 1은 분자 네트워크에서 RWR(Random Walk with Restart) 알고리즘을 수행하여 파이토케미컬의 표현형 벡터를 제작하는 단계(단계 1-1); 고정된 수의 표적 단백질들로부터 파이토케미컬의 표적들을 무작위로 선택하여 랜덤 표현형 벡터를 제작하는 단계(단계 1-2); 및 상기 랜덤 표현형 벡터에서 통계적으로 유의한 파이토케미컬의 표현형을 도출함으로써 파이토케미컬의 건강효과를 추론하는 단계(단계 1-3)를 포함할 수 있다.

- [17] 상기 단계 1-1은 하기의 단계를 포함하는 것일 수 있다:

- [18] (a) 파이토케미컬들의 분자 표적 정보에 기반하여 분자 네트워크 상의 시드 노드(seed nodes)에 초기값을 할당하는 단계 및

- [19] (b) 하나의 노드에서 이웃한 노드로의 전이확률(transition probability)을 계산하는 단계,

- [20] 여기서 각 노드의 전이확률은 시간 단계가 $t+1$ 일 때 하기 수학적 식 1로 정의 됨;

- [21] [수학적 식 1]

- [22]
$$p_{(t+1)} = (1 - r)W^T p_t + r p_0$$

- [23] (r: 각 시간 단계의 랜덤 워커(random walker)의 재시작 확률(restarting probability);

- [24] W: 분자 네트워크의 정규화된 인접 행렬(normalized adjacency matrix); 및

- [25] P_t 및 P_0 : 시간단계가 t 일 때 및 초기 각 노드의 확률 벡터).

- [26] 상기 단계 1-3은 하기 수학적 식 2로 표현되는 수식에 의해 계산되는 p 값이 0.01보다 낮은 표현형들을 통계적으로 유의한 파이토케미컬의 표현형으로 판단하는 단계를 포함하는 것일 수 있다:

- [27] [수학식 2]
- [28] $p = (r + 1)/(n + 1)$
- [29] (r: 표현형 값보다 큰 값을 갖는 파이토케미컬의 랜덤 표현형 벡터의 수; 및
- [30] n: 파이토케미컬의 랜덤 표현형 벡터의 수).
- [31] 상기 단계 2의 화학적 특성은 분자량, 옥탄올(octanol)-물 분배 계수(partition coefficient)의 로그 값(AlogP), 수소결합 공여자(hydrogen-bond donors) 수, 수소결합 수용자(hydrogen-bond acceptor) 수, 회전가능한 결합(rotatable bond)의 수, 인간 장관 흡수(human intestinal absorption, HIA), Caco-2 투과성(permeability), 뇌-혈관 장벽(blood-brain barrier, BBB) 투과성 및 리핀스키의 5 규칙(Lipinski's rule of five, RO5)으로 이루어진 군으로부터 선택되는 어느 하나 이상일 수 있다.
- [32] 상기 단계 3의 의미 유사도는 하기 수학식 3으로 표현되는 수식에 의해 계산되는 것일 수 있다:
- [33] [수학식 3]
- [34]
$$\text{sim}(c_1, c_2) = \frac{2 \times \text{depth}(lcs(c_1, c_2))}{\text{path}(c_1, lcs(c_1, c_2)) + \text{path}(c_2, lcs(c_1, c_2)) + 2 \times \text{depth}(lcs(c_1, c_2))}$$
- [35] (sim: 의미 유사도;
- [36] depth: 근원 표현형(root UMLS)으로부터 해당 표현형까지의 깊이;
- [37] path: 각 표현형 간의 거리; 및
- [38] lcs(c_1, c_2): c_1 및 c_2 개념의 공통적인 가장 낮은 계층의(가장 구체적인) 개념(the lowest common subsumer)).
- [39] 상기 단계 3 이후에는 시각화 수단을 통해 도출된 정보에 대한 시각화 데이터를 생성하는 단계 및 이를 출력 수단을 통해 시각화 데이터로 출력하는 단계를 추가적으로 포함할 수 있다. 상기 출력 수단은 모니터, 프린터 및 플로터로 이루어진 군으로부터 선택되는 어느 하나일 수 있으며, 결과물을 출력할 수 있는 어떠한 수단이어도 무방하다.
- [40]
- [41] 또한, 본 발명은 상기 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법을 실행시키기 위한 명령들을 포함하는 프로그램이 저장된 컴퓨터로 판독가능한 매체를 제공한다. 컴퓨터로 판독가능한 매체는 컴퓨터 시스템에 의하여 읽혀질 수 있는 데이터가 저장되는 모든 종류의 기록장치를 포함한다. 컴퓨터로 판독가능한 기록매체의 예로는 ROM, RAM, CD-ROM, 자기 테이프, 플로피디스크, 광데이터 저장장치 등이 있으며, 캐리어 웨이브(예를 들어 인터넷을 통한 전송)의 형태로 구현되는 것도 포함한다. 또한 컴퓨터로 판독가능한 기록매체에는 네트워크로 연결된 컴퓨터 시스템에 분산되어 분산방식으로 컴퓨터가 읽을 수 있는 코드가 저장되고 실행될 수 있다.
- [42]

- [43] 또한, 본 발명은 분자 네트워크를 이용하여 파이토케미컬의 건강효과를 추론하는 모듈; 파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은 파이토케미컬을 도출하는 모듈; 및 민족약학적 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적 증거와 의미 유사도가 높은 파이토케미컬의 건강효과를 도출하는 모듈을 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 시스템을 제공한다.
- [44] 상기 분자 네트워크를 이용하여 파이토케미컬의 건강효과를 추론하는 모듈은 분자 네트워크에서 **RWR** 알고리즘을 수행하여 파이토케미컬의 표현형 벡터를 제작하는 모듈; 고정된 수의 표적 단백질들로부터 파이토케미컬의 표적들을 무작위로 선택하여 랜덤 표현형 벡터를 제작하는 모듈; 및 상기 랜덤 표현형 벡터에서 통계적으로 유의한 파이토케미컬의 표현형을 도출함으로써 파이토케미컬의 건강효과를 추론하는 모듈을 포함하는 것일 수 있다.
- [45] 상기 시스템은 상기 모듈들에 의해 도출된 정보에 대한 시각화 데이터를 생성하는 모듈 및 이를 시각화 데이터로 출력하는 출력 모듈을 추가적으로 포함할 수 있다.
- [46] 상기 출력 모듈은 모니터, 프린터 및 플로터로 이루어진 군으로부터 선택되는 어느 하나일 수 있으며, 결과물을 출력할 수 있는 어떠한 수단이어도 무방하다.
- [47]
- [48] 본 발명의 구체적인 실시예에서, 본 발명자들은 **RWR** 알고리즘을 수행하여 파이토케미컬의 표현형 값 목록을 제작하고, 파이토케미컬에 대한 랜덤 표현형 벡터를 제작한 후, 통계적으로 유의한 표현형들을 선별하여, 파이토케미컬의 건강효과를 추론하였다(도 1a 참조).
- [49] 또한 본 발명자들은 화학 구조 정보가 알려진 파이토케미컬들을 대상으로 **RO5**, **HIA**, **Caco-2** 투과성 및 **BBB** 투과성을 계산하여 파이토케미컬의 화학적 특성에 대한 정보를 획득하였다(도 1b 참조).
- [50] 또한 본 발명자들은 파이토케미컬의 민족약학적 용도를 탐색하여, 분자 네트워크 분석에 따라 예측된 파이토케미컬의 건강효과들 중 민족약학적 증거를 갖는 건강효과들을 도출하고, 그 분포를 비교하였다(도 1c 내지 도 3 참조).
- [51] 또한 본 발명자들은 본 발명의 방법에 따른 파이토케미컬의 건강효과 예측 방법의 정밀도 및 민감도를 분석하여 본 발명의 예측 방법이 우수한 성능을 가짐을 확인하였고, 본 발명의 방법에 따라 예측된 파이토케미컬의 건강효과가 외부 문헌에서 확인됨에 따라, 본 발명의 예측 방법이 신뢰성 있는 방법임을 확인하였다.
- [52] 따라서, 본 발명에 따른 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법, 이를 실행시키기 위한 명령들을 포함하는 프로그램이 저장된 컴퓨터로 판독가능한

매체, 및 상기 방법을 실행하기 위한 시스템은 파이토케미컬을 이용한 약물 개발에 유용하게 사용될 수 있다.

[53]

[54] 이하 본 발명을 실시예에 의해 상세히 설명한다.

[55]

[56] 단, 하기 실시예는 본 발명을 예시하는 것일 뿐, 본 발명의 내용이 하기 실시예에 의해서 한정되는 것은 아니다.

[57]

[58] [실시예 1]

[59] 데이터 수집

[60] 파이토케미컬 및 식물의 구성 화합물에 대한 정보는

KTKP(<http://www.koreantk.com/>), TCMID(Xue, R. *et al.*, *Nucleic Acids Res.* 1089-1095, 2012) 및 FooDB(<http://foodb.ca/>)로부터 수집하였다. 식물의 민족약학적 용도는 KTKP, TCMID 및 Kampo(<http://kampo.ca/>)로부터 수집하였고, 파이토케미컬의 분자 표적은 DrugBank, DCDB(Drug Combination Database, v. 2.0, Liu, Y. *et al.*, *Database*, bau124, 2014), CTD(Comparative Toxicogenomics Database, Davis, A. P. *et al.*, *Nucleic Acids Res.* 39, D1067-1072, 2011), MATADOR(Gunther, S. *et al.*, *Nucleic Acids Res.* 36, D919-922, 2008), STITCH(Kuhn, M. *et al.*, *Nucleic Acids Res.* 42, D401-407, 2013) 및 TTD(Zhu, F. *et al.*, *Nucleic Acids Res.* 40, D1128-1136, 2011)로부터 수집하였다. 유전자-표현형 관계는 CTD로부터 수집하였고, 19,093 개의 노드 및 270,970 개의 엣지(edge)를 포함하는 단백질-단백질 상호작용 네트워크는 BioGrid(v. 3.4.136, Chatr-Aryamontri, A. *et al.*, *Nucleic Acids Res.* 43, D470-478, 2015) 및 CODA(Context-Oriented Directed Associations, Hwang, W. *et al.*, *BMC Med. Inf. Decis. Making* 13, S4, 2013)에서 수집하였다. 표현형 네트워크는 생물의학에 관한 다양한 용어들에 대한 통합 정보를 제공하는 UMLS(unified medical language system; Bodenreider, O. *Nucleic Acids Res.* 32, D267-270, 2004)의 2017AA 버전으로부터 얻었다. UMLS에서 각각의 구별되는 생물학적인 개념에는 개념 고유 식별자(concept unique identifier, CUI)가 할당되는데, 관련된 개념들의 목록인 MRREL 목록으로부터 UMLS 관계들의 11 개 유형 중 상위어 관계(broader relationships, RB), 하위어 관계(narrower relationships, RN) 및, 동의어, 상위어 및 하위어 이외의 관계(other-related relationships, RO)를 갖는 CUI들을 수집하여, 총 220,104 개의 CUI 및 663,018 개의 관계를 수집하였다.

[61] 골드-스탠다드 세트(gold-standard set)를 위하여 약물 유래 파이토케미컬을 DrugBank(v. 4.0, Law, V. *et al.*, *Nucleic Acids Res.* 42, D1091-1097, 2014)에서 수집하였다. 표현형-관련 용어들을 추출하기 위하여 DrugBank, CTD, ClinicalTrials.gov(Zarin, D. A. *et al.*, *New Engl. J. Med.* 364, 852-860, 2011) 및 DCDB에서 MetaMap 도구(Aronson, A. R. *et al.*, *J. Am. Med. Inf. Assoc.* 17,

229-236, 2010)를 이용하여 약물-표현형 관계를 수집하였고, 135 개의 의미 유형(semantic types) 중 "질환 또는 신드롬"과 같이 표현형에 관련된 20 개의 의미 유형에 할당된 메타시소러스(Metathesaurus) 개념을 이용하였다(표 1).

[62] [표1]

Semantic type	
Acquired Abnormality	Mental Process
Anatomical Abnormality	Mental or Behavioral Dysfunction
Biologic Function	Neoplastic Process
Congenital Abnormality	Pathologic Function
Cell or Molecular Dysfunction	Physiologic Function
Disease of Syndrome	Sign or Symptom
Experimental Model of Disease	Clinical attribute
Finding	Hazardous or Poisonous Substance
Injury or Poisoning	Body, Part, Organ, or Organ Component
Laboratory or Test Result	Tissue

[63] [실시에 2]

[64] 분자 네트워크를 이용한 파이토케미컬의 건강효과 추론

[65] **2-1. RWR(Random Walk with Restart) 알고리즘 수행을 통한 파이토케미컬의 표현형 값 목록 제작**

[66] 파이토케미컬의 퍼져있는(propagated) 효과들을 조사하기 위하여, 단백질-단백질 상호작용 정보에 기반하여 분자 네트워크를 제작하고, RWR 알고리즘을 수행하였다.

[67] 구체적으로, 파이토케미컬들의 표적 정보에 기반하여 분자 네트워크 상의 시드 노드(seed nodes)에 초기값을 할당하였다. 파이토케미컬들의 표적 정보 중 직접적 관계는 파이토케미컬과 표적 단백질 사이의 결합 정보를 포함하는 반면, 간접적 관계는 단백질 발현, 화합물 유도 인산화 또는 파이토케미컬의 활성 대사체에 의한 영향에 따른 변화를 야기시키는 상호작용을 포함한다. 분자 네트워크에서 파이토케미컬의 생물학적 활성은 복잡한 상호작용들로부터 변화될 수 있고, 파이토케미컬의 결합 표적 정보는 합성 의약품들에 비하여 많이 알려져 있지 않기 때문에, 직접적 및 간접적 관계에 대한 정보를 모두 사용하였고, 직접적 및 간접적 관계에 대한 초기값은 각각 1 및 0.3으로 할당하였다.

[68] 이후, 한 노드에서 이웃한 노드로의 전이확률(transition probability)을 계산하였다. 상기 전이확률은 분자 네트워크 상에 퍼져있는 약물 효과를 나타내는 것으로 가정하였고, 시간 단계가 $t + 1$ 일 때 각 노드의 전이확률 벡터는

하기 수학식 1에 의해 정의되었다:

[69] [수학식 1]

$$[70] \quad p_{(t+1)} = (1-r)W^T p_t + r p_0$$

[71] (r: 각 시간 단계의 랜덤 워커(random walker)의 재시작 확률(restarting probability)로 본 발명에서는 0.7로 설정함;

[72] W: 분자 네트워크의 정규화된 인접 행렬(normalized adjacency matrix); 및

[73] P_t 및 P₀: 시간단계가 t일 때 및 초기 각 노드의 확률 벡터).

[74]

[75] RWR 알고리즘을 통하여 모든 노드가 정상 상태(steady state, P_{t+1} - P_t < 10⁻⁸)에 도달할 때까지 랜덤 워커를 시뮬레이션하였다. 특정 표현형에 대한 치료 표적 또는 바이오마커인 모든 유전자들에 대하여 RWR 결과에서 얻은 유전자 값들을 유전자-표현형 관계에 기반하여 해당하는 표현형에 연결시켜, 최종적으로 각 파이토케미컬에 대한 정량화된 표현형 값 목록을 제작하였다.

[76]

[77] 2-2. 랜덤 표현형 벡터로부터 유의한 표현형들의 선별

[78] 상기 실시예 2-1에서 제작한 표현형 값 목록의 표현형 값들은 파이토케미컬과 표현형 사이의 관계에 있어서 반드시 그 연관성의 크기를 나타내지는 않기 때문에, 파이토케미컬의 랜덤 표현형 벡터를 제작하고, 이를 추론된 표현형 값들의 목록과 비교함으로써, 유의한 값을 가지는 표현형들을 선별하고자 하였다. 랜덤 표현형 벡터는 고정된 수의 표적 단백질들로부터 파이토케미컬의 표적들을 무작위로 선택함으로써 제작하였다. 각 파이토케미컬에 대하여 1,000 개의 랜덤 표현형 벡터들을 제작하였고, 하기 수학식 2에 의해 p 값을 계산하여, 계산된 p 값이 0.01보다 낮은 표현형들을 통계적으로 유의한 표현형들로 선별하였다:

[79] [수학식 2]

$$[80] \quad p = (r + 1)/(n + 1)$$

[81] (r: 표현형 값보다 큰 값을 갖는 파이토케미컬의 랜덤 표현형 벡터의 수; 및

[82] n: 파이토케미컬의 랜덤 표현형 벡터의 수).

[83]

[84] 2-3. 추론된 파이토케미컬의 건강효과들의 분포

[85] 상기 실시예 2-1 내지 2-2에 의해, 총 591 개의 파이토케미컬에 대한 잠재적인 건강효과들을 추론하였고, 그 결과 각 파이토케미컬에 대하여 평균적으로 415.6 ± 27.3 개의 건강효과들이 예측되었다(신뢰구간(confidence interval): 0.95, 도 3의 왼쪽 그래프).

[86]

[87] [실시예 3]

[88] 파이토케미컬의 화학적 특성의 계산

- [89] 파이토케미컬의 화학적 특성을 분석하여 생체이용률이 높은 파이토케미컬을 도출하기 위하여, 파이토케미컬의 물리화학적 특성 및 생리학적 영향을 포함하는 화학적 특성을 계산하였으며, 화학 구조들에 대한 정보가 알려진 총 512 개의 파이토케미컬을 대상으로 분석하였다(도 1B).
- [90] 구체적으로, 물리화학적 특성은 분자량, 옥탄올(octanol)-물 분배 계수(partition coefficient)의 로그 값(AlogP), 수소결합 공여자(hydrogen-bond donors) 수, 수소결합 수용자(hydrogen-bond acceptor) 수 및 회전가능한 결합(rotatable bond)의 수를 포함한다. 생리학적 영향은 인간 장관 흡수(human intestinal absorption, HIA), Caco-2 투과성(permeability), 뇌-혈관 장벽(blood-brain barrier, BBB) 투과성 및 리핀스키의 5 규칙(Lipinski's rule of five, RO5)을 포함한다. HIA 및 BBB 값은 Shen's work(Shen, J. *et al.*, J. Chem. Inf. Model, 50, 1034-1041, 2010)를 이용하여 계산하고, Caco-2 투과성은 정량적 구조-활성 관계(quantitative structure-activity relationship, QSAR) 모델(Pham The, H. *et al.*, Mol. Inform. 30, 376-385, 2011)을 통하여 계산하였다. RO5 및 다른 물리화학적 특성들은 CDK(Chemistry Development Kit) Descriptor Calculator(Guha, R. <http://www.rguha.net/code/java/cdkdesc.html>)를 이용하여 계산하였다.

- [91] [표2]

	RO5	HIA	Caco-2 투과성	BBB 투과성
RO5	446	401	280	365
HIA		482	330	407
Caco-2 투과성			335	303
BBB 투과성				428

- [92] (가로축과 세로축이 교차하는 칸의 수치는 각 항목을 동시에 만족하는 파이토케미컬의 수를 의미함)
- [93] 파이토케미컬의 RO5, HIA, Caco-2 투과성 및 BBB 투과성을 조사한 결과, 446, 482, 335 및 428 개의 파이토케미컬들이 각각 RO5, HIA, Caco-2 투과성 및 BBB 투과성을 만족하는 것으로 나타났다(표 2).
- [94]
- [95] [실시예 4]
- [96] 파이토케미컬의 민족약학적 용도 탐색
- [97] 파이토케미컬의 예측된 효과들의 민족약학적 증거를 갖는 식물들을 추출하기 위하여, 서술적인 표현들로부터 표현형 용어들은 추출되고 구조화되어야 하고, 표현형 간의 복잡한 관계는 정량화되어야 한다. 이를 위하여, 먼저, MetaMap 도구를 적용하여 서술적인 문서들로부터 표현형 관련 용어들을 추출하였다(도 2a).
- [98] 다음으로, 외부 데이터베이스 정보에 기반하여 알고 싶은 파이토케미컬을

포함하는 식물들을 탐색하였다(도 2b). 다음으로, UMLS의 계층적 관계에
기반하여 표현형 네트워크를 제작하였고, 네트워크 상의 표현형 간의 거리와
깊이를 고려하여 표현형 간의 의미 유사도(semantic similarity)를 Wu, Z. 및
Palmer, M.이 제안하고, 하기 수학적 식 3으로 표현되는 의미 유사도 측정 방법(Wu,
Z. & Palmer, M. 133-138 (Association for Computational Linguistics))에 의해
계산하였다(도 2c).

[99] [수학적 식 3]

$$[100] \quad \text{sim}(c_1, c_2) = \frac{2 \times \text{depth}(\text{lcs}(c_1, c_2))}{\text{path}(c_1, \text{lcs}(c_1, c_2)) + \text{path}(c_2, \text{lcs}(c_1, c_2)) + 2 \times \text{depth}(\text{lcs}(c_1, c_2))}$$

[101] (sim: 의미 유사도;

[102] depth: 근원 표현형(root UMLS)으로부터 해당 표현형까지의 깊이;

[103] path: 각 표현형 간의 거리; 및

[104] lcs(c_1, c_2): c_1 및 c_2 개념의 공통적인 가장 낮은 계층의(가장 구체적인) 개념(the lowest common subsumer)).

[105]

[106] 표현형 간의 의미 유사도 점수가 0.8보다 큰 경우, 표현형 간의 관련성이 높은 것으로 가정하였다. 따라서, 파이토케미컬의 예측된 건강효과들에 대한 가능한 모든 쌍(pairs)과 식물의 민족약학적 효과 사이의 의미 유사도를 계산하고, 그 점수가 0.8보다 높은 식물들을 선별함으로써, 최종적으로 파이토케미컬의 민족약학적 용도를 탐색하였다.

[107] 분석 결과, 실시예 2-3에서 분자 네트워크를 기반으로 예측된 건강효과들 중 평균적으로 약 129.1 ± 11.4 개의 건강효과들(약 31%)이 민족약학적 증거를 갖는 것으로 나타났다(도 3의 오른쪽 그래프).

[108]

[109] [실시예 5]

[110] 본 발명에 따른 파이토케미컬의 건강효과 예측 방법의 성능 평가

[111] **5-1. 데이터 수집**

[112] 실험적으로 검증된 정보를 골드-스탠다드 양성 세트로 수집하였다.

DrugBank의 정보를 치료 효과를 평가하기 위한 세트로 사용하고, SIDER(Side Effect Resource)로부터 얻은 정보를 부작용 효과를 평가하기 위한 세트로 사용하였다. 보다 많은 수의 파이토케미컬을 분석하기 위하여 추가적으로 CTD로부터 잠재적 후보효과들을 수집하여 실버-스탠다드(silver-standard) 양성 세트로 사용하였다.

[113]

[114] **5-2. 분자 네트워크 분석을 이용한 방법의 정밀도 및 민감도 분석**

[115] 먼저, 분자 네트워크 분석을 이용한 파이토케미컬의 효과 예측 방법의 정밀도(precision, p) 및 민감도(recall, r) 값을 하기 수학적 식 4 및 5에 의해

계산하였다:

[116] [수학식 4]

[117] 정밀도 = $TP/(TP + FP)$; 및

[118] [수학식 5]

[119] 민감도 = $TP/(TP + FN)$

[120] (TP: 실제 양성인 데이터가 양성으로 검출된 경우(true positive);

[121] FP: 실제 음성인 데이터가 양성으로 검출된 경우(false positive); 및

[122] FN: 실제 양성인 데이터가 음성으로 검출된 경우(false negative)).

[123]

[124] [표3]

분자 네트워크 분석을 통한 치료 효과, 부작용 및 잠재적 후보효과 예측 방법의 정밀도 및 민감도 성능

	편포도	치료 효과	부작용	잠재적 후보효과
정밀도	1:1	0.921 ± 0.032	0.922 ± 0.021	0.942 ± 0.005
	1:10	0.518 ± 0.059	0.432 ± 0.040	0.706 ± 0.013
	All	0.006 ± 0.001	0.049 ± 0.010	0.522 ± 0.022
민감도	All	0.738 ± 0.062	0.576 ± 0.061	0.909 ± 0.011

[125] 각 파이토케미컬마다 평균 415.6 개의 잠재적 건강효과를 예측하기 때문에, 정밀도가 매우 낮게 나타나므로, 양성 세트 및 음성 세트 간의 편포도(skewness)를 조절하여 정밀도 성능을 평가하고자 하였다. 모든 가능한 건강효과들 중 골드-스탠다드 양성 세트를 제외한 나머지를 골드-스탠다드 음성 세트로 사용하고, 양성 세트 및 음성 세트의 비율을 1:1 또는 1:10으로 조절하여 정밀도를 분석한 결과, 편포도를 조절한 경우 정밀도 성능이 높게 나타났다(표 3).

[126] 또한, 민감도 성능을 분석한 결과, 61 개 파이토케미컬의 270 개 치료 효과들 중 191 개의 표현형이 예측되었고($r=0.738 \pm 0.062$), 부작용 예측에 있어서 60 개 파이토케미컬의 1,784 개 표현형 중 1,059 개의 표현형이 예측되었으며($r=0.576 \pm 0.061$), 잠재적 후보효과 예측에 있어서 453 개 파이토케미컬의 136,862 개 표현형 중 119,233 개의 표현형이 예측되었다($r=0.909 \pm 0.011$)(표 3).

[127] 전반적으로, 분자 네트워크 분석을 통한 건강효과들의 예측은 우수한 성능을 나타내었다.

[128]

[129] **5-3. 분자 네트워크 분석 및 민족약학적 증거를 이용한 방법의 정밀도 분석**

[130] 민족약학적 증거를 고려하거나 고려하지 않은 경우의 예측 성능을 비교하기 위하여, 상기 실시예 5-2에서 예측된 파이토케미컬의 건강효과들을 민족약학적 용도의 존재 여부에 따라 분류하여 정밀도를 분석하였다.

[131]

[132] [표4]

분자 네트워크 분석 및 민족약학적 증거를 이용한 치료 효과, 부작용 및 잠재적 후보효과 예측 방법의 정밀도 성능

	편포도	치료 효과	부작용	잠재적 후보효과
정밀도	1:1	0.941 ± 0.035	0.761 ± 0.033	0.948 ± 0.014
	1:10	0.541 ± 0.069	0.319 ± 0.055	0.732 ± 0.037
	All	0.014 ± 0.003	0.025 ± 0.005	0.563 ± 0.059

[133] 그 결과, 민족약학적 증거를 고려하는 경우, 그렇지 않은 경우보다 치료 효과 및 잠재적 후보효과 예측 면에서 정밀도 성능이 증가하였다. 다만, 부작용에 대한 예측에서는 정밀도가 감소하였는데, 이는 치료적 용도에 대한 민족약학적 증거만을 사용했기 때문에 나타난 결과로 해석된다(표 4).

[134]

[135] **5-4. 화학적 특성 분석을 이용한 방법의 정밀도 및 민감도 분석**

[136] 화학적 특성을 이용하여 파이토케미컬의 건강효과 예측 방법에 있어서 화학적 특성 분석을 이용하는 경우 예측 성능이 개선되는지 여부를 확인하기 위하여 BBB 투과성에 따른 신경질환 치료 효능 예측에 대한 정밀도 및 민감도를 분석하였다.

[137] 구체적으로, 신경활성 기능을 조절하기 위해서는 파이토케미컬들이 BBB를 통과해야 하므로, BBB 투과성 여부에 따라 파이토케미컬을 2 개의 독립적인 세트로 분류하고, 신경질환 치료 효능에 대한 예측 성능을 비교하였다. 그 결과, BBB를 통과하는 세트의 정밀도 및 민감도 값($p = 0.611 \pm 0.046$, $r = 0.725 \pm 0.033$)이 BBB를 통과하지 못하는 세트의 정밀도 및 민감도 값($p = 0.312 \pm 0.052$, $r = 0.558 \pm 0.042$)보다 높았다.

[138]

[139] 상기 실시예 5-1 내지 5-4를 통한 분석 결과, 전반적으로 분자 네트워크 분석, 민족약학적 증거 및 화학적 특성을 개별적으로 이용하는 경우보다 이를 통합적으로 이용하는 본 발명의 방법이 파이토케미컬의 건강효과들을 예측하는 성능이 우수함을 알 수 있었다.

[140]

[141] [실시예 6]

[142] **본 발명에 따른 파이토케미컬의 건강효과 예측 방법의 신뢰성 검증**

[143] 본 발명에 따른 파이토케미컬의 건강효과 예측 방법의 신뢰성을 검증하기 위하여, 예측된 건강효과들이 외부 문헌에서 확인되는지 조사하였다.

[144]

[145] **6-1. 데이터 세트 추출**

[146] 파이토케미컬의 예측된 건강효과들에 기반하여 2 개의 독립적인 세트를 제작하였다. 먼저, 분자 네트워크, 경구-이용가능성(oral-availability) 및 민족약학적 증거에서 양성으로 예측된 파이토케미컬-표현형 관계 세트를 만들고, 대조군으로, 이와 동일한 수의 무작위 관계를 추출하여 무작위 관계 세트를 만들었다.

[147]

[148] **6-2. 공동발생(co-occurrence) 분석**

[149] 1950 내지 2013 년 동안 개시된 13,200,786 개의 PubMed 초록들을 대상으로 동일한 문장 안에 파이토케미컬 및 그에 해당하는 표현형이 모두 포함된 초록의 수를 공동발생 수(n_c)로 계산하였다.

[150]

[151] [표5]

	공동발생	자카드 인덱스	피셔 검정 ^a	FDR 검정 ^b
예측된 관계 세트	1.25	1.8×10^{-4}	2984	1341
무작위 관계 세트	0.09	9.5×10^{-6}	612	274
Mann-Whitney U test의 p 값	< 0.001	< 0.001	< 0.001	< 0.001

[152] ^a피셔 정확 검정에서 p 값이 0.001보다 낮은 파이토케미컬-건강효과 관계의 수^b FDR 검정에서 q 값이 0.05보다 낮은 파이토케미컬-건강효과 관계의 수

[153]

[154] 그 결과, 예측된 관계 세트의 공동발생 수는 무작위 관계 세트의 공동발생 수에 비하여 13.8 배 더 크게 나타났다(표 5).

[155]

[156] **6-3. 자카드 인덱스(Jaccard index) 분석**

[157] 파이토케미컬 및 표현형들의 빈도수에서의 차이를 보정하기 위하여 공동발생 수를 자카드 인덱스로 표준화하였다. 먼저, 추가적으로 파이토케미컬 및 표현형을 하나 이상 포함하는 PubMed 초록의 수를 세고(n_0), 각 파이토케미컬-표현형 관계의 공동발생 수인 n_c 를 n_0 로 나누어 자카드 인덱스를 계산하였다. 예를 들면, '퀘르세틴(quercetin)-뇌졸중'의 파이토케미컬-표현형 쌍에 대한 자카드 인덱스를 계산하는 경우, 퀘르세틴 및 뇌졸중을 모두 포함하는 초록이 50 개가 있고, 이 둘 중 하나 이상을 포함하는 초록이 200 개가 있다면, 자카드 인덱스는 0.25가 된다.

[158] 분석 결과, 예측된 관계 세트의 평균 자카드 인덱스는 무작위 관계 세트의 평균 자카드 인덱스보다 18.9 배 더 높게 나타났다(표 5).

[159]

[160] **6-4. 피셔 정확 검정(Fisher's exact test) 분석**

[161] 파이토케미컬 및 표적 건강효과를 포함하는 PubMed 초록의 수를 세어 피셔 정확 검정을 수행하고, p 값이 0.001보다 낮은 통계적으로 유의한 관계 수를 계산하였다.

[162] 그 결과, 예측된 관계 세트의 유의한 관계 수는 무작위 관계 세트의 유의한 관계 수보다 4.8 배 더 높게 나타났다(표 5).

[163]

[164] **6-5. FDR 검정 분석**

[165] 많은 수의 관계를 분석하는 경우에는 피셔 정확 검정 시 다중 검정에 따른 문제가 발생하며, 많은 위 양성(false positive) 결과를 초래하므로, 이를 보완하기 위하여 추가적으로 FDR 검정 분석을 수행하여, q 값이 0.05보다 낮은 유의한 관계를 조사하였다.

[166] 그 결과, q 값이 0.05보다 낮은 관계의 수는 예측된 관계 세트에서 무작위 관계 세트에 비하여 4.9 배 높게 나타났다(표 5).

[167]

[168] **6-6. Mann-Whitney U test 분석**

[169] 마지막으로, Mann-Whitney U test를 수행하고 해당하는 p 값을 계산하여 문헌적 증거와 예측된 관계 세트 및 무작위 관계 세트 간의 통계적 차이를 조사하였다. 이 때, p 값은 0.05보다 낮은 경우를 통계적으로 유의한 것으로 판단하였다.

[170] 그 결과, 모든 항목에서 예측된 관계 세트 및 무작위 관계 세트 간의 p 값이 0.05보다 낮게 나타났으며, 이는 두 세트에 대한 문헌 증거가 통계적으로 유의한 차이를 나타냄을 의미한다(표 5).

[171]

[172] 상기 실시예 6-1 내지 6-6을 통한 분석 결과, 본 발명의 방법이 파이토케미컬의 건강효과를 확인하는데 유용하게 사용될 수 있음을 알 수 있었다.

[173]

[174] [실시예 7]

[175] 본 발명에 따른 방법을 이용한 파이토케미컬의 잠재적 건강효과 예측

[176] 본 발명에 따라 몇 가지 파이토케미컬에 대한 잠재적 건강효과를 예측하고, ClinicalTrials.gov에서 해당 효과에 대한 임상시험 여부를 조사하였다.

[177]

[178] [표6]

Phytochemical	Potential Health Effects	Rank	n_e	Clinical Trials (Phase)	Exemplary Studies (PMID)
Choline	Neurological disorder	2	48	-	-
	Alzheimer's disease	81	31	NCT02648906 (Phase 4)	12787861, 15647594, 12637119, 7913981
	Cognitive impairment	419	3	NCT01363648 (Phase 4)	21195433, 19304299, 26366063, 23403474
	Pain	6	44	NCT00720343 (Phase 4)	15780465, 19372354, 16942753, 29082318
Isoquercitrin	Hypertension	2	64	NCT01691404 (Not Applicable)	20134098, 25460361, 17951477, 16636461
	Thrombosis	10	56	NCT02195232 (Phase 2, 3)	12854360, 15234778, 20148891, 20626032
Niacin	Neurological Disease	2	128	-	-
	Parkinson's Disease	59	65	NCT03462680 (Not Applicable)	26273459, 25455298, 26988916, 18381761
	Heart condition	7	117	NCT00120289 (Phase 3) NCT00000599 (Phase 3)	28057839, 23916935, 28927896, 10924076
	Vascular disease	9	113	-	-
	Cardiovascular disease	20	105	NCT00715273 (Phase 4) NCT02003638 (Not Applicable)	3295315, 19159436, 15258194, 24641964

[179] (n_e : 민족약학적 증거의 수; 및

[180] Rank: 민족약학적 증거의 수로 분류한 표현형 순서 중 해당 표현형의 순위)

[181]

[182] 그 결과, 전반적으로 파이토케미컬과 잠재적 관계가 깊은 표현형일수록 관련된 임상시험이 진행 중인 경우가 많았고, 해당 임상시험들은 대체로 제3상 또는 제4상 임상시험인 경우가 많았다(표 6). 상기 결과는 본 발명에 따른 파이토케미컬의 건강효과 예측 방법이 파이토케미컬의 특정 질병에 대한 잠재적 치료 효과를 예측하는데 유용하게 사용될 수 있음을 제시한다.

[183]

[184]

청구범위

- [청구항 1] 분자 네트워크를 이용하여 파이토케미컬(phytochemical)의 건강효과(health effect)를 추론하는 단계(단계 1); 파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은 파이토케미컬을 도출하는 단계(단계 2); 및 민족약학적(ethnopharmacological) 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적 증거와 의미 유사도(semantic similarity)가 높은 파이토케미컬의 건강효과를 도출하는 단계(단계 3)를 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법.
- [청구항 2] 제1항에 있어서, 상기 단계 1은 분자 네트워크에서 RWR(Random Walk with Restart) 알고리즘을 수행하여 파이토케미컬의 표현형 벡터를 제작하는 단계(단계 1-1); 고정된 수의 표적 단백질들로부터 파이토케미컬의 표적들을 무작위로 선택하여 랜덤 표현형 벡터를 제작하는 단계(단계 1-2); 및 상기 랜덤 표현형 벡터에서 통계적으로 유의한 파이토케미컬의 표현형을 도출함으로써 파이토케미컬의 건강효과를 추론하는 단계(단계 1-3)를 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법.
- [청구항 3] 제2항에 있어서, 상기 단계 1-1은 하기의 단계를 포함하는 것인, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 방법:
 (a) 파이토케미컬들의 분자 표적 정보에 기반하여 분자 네트워크 상의 시드 노드(seed nodes)에 초기값을 할당하는 단계 및
 (b) 하나의 노드에서 이웃한 노드로의 전이확률(transition probability)을 계산하는 단계,
 여기서 각 노드의 전이확률은 시간 단계가 t+1일 때 하기 수학적 식 1로 정의 됨;
 [수학적 식 1]

$$p_{(t+1)} = (1 - r)W^T p_t + r p_0$$
 (r: 각 시간 단계의 랜덤 워커(random walker)의 재시작 확률(restarting probability);
 W: 분자 네트워크의 정규화된 인접 행렬(normalized adjacency matrix); 및 p_t 및 p_0 : 시간단계가 t일 때 및 초기 각 노드의 확률 벡터).
- [청구항 4] 제2항에 있어서, 상기 단계 1-3은 하기 수학적 식 2로 표현되는 수식에 의해 계산되는 p 값이 0.01보다 낮은 표현형들을 통계적으로 유의한 파이토케미컬의 표현형으로 판단하는 단계를 포함하는, 분자 네트워크,

화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한
파이토케미컬의 건강효과 예측 방법:

[수학식 2]

$$p = (r + 1)/(n + 1)$$

(r: 표현형 값보다 큰 값을 갖는 파이토케미컬의 랜덤 표현형 벡터의 수;
및

n: 파이토케미컬의 랜덤 표현형 벡터의 수).

[청구항 5] 제1항에 있어서, 상기 단계 2의 화학적 특성은 분자량, 옥탄올(octanol)-물
분배 계수(partition coefficient)의 로그 값(AlogP), 수소결합
공여자(hydrogen-bond donors) 수, 수소결합 수용자(hydrogen-bond
acceptor) 수, 회전가능한 결합(rotatable bond)의 수, 인간 장관 흡수(human
intestinal absorption, HIA), Caco-2 투과성(permeability), 뇌-혈관
장벽(blood-brain barrier, BBB) 투과성 및 리핀스키의 5 규칙(Lipinski's rule
of five, RO5)으로 이루어진 군으로부터 선택되는 어느 하나 이상인, 분자
네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한
파이토케미컬의 건강효과 예측 방법.

[청구항 6] 제1항에 있어서, 상기 단계 3의 의미 유사도는 하기 수학식 3으로
표현되는 수식에 의해 계산되는 것인, 분자 네트워크, 화학적 특성 및
민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과
예측 방법:

[수학식 3]

$$\text{sim}(c_1, c_2) = \frac{2 \times \text{depth}(\text{lcs}(c_1, c_2))}{\text{path}(c_1, \text{lcs}(c_1, c_2)) + \text{path}(c_2, \text{lcs}(c_1, c_2)) + 2 \times \text{depth}(\text{lcs}(c_1, c_2))}$$

(sim: 의미 유사도;

depth: 근원 표현형(root UMLS)으로부터 해당 표현형까지의 깊이;

path: 각 표현형 간의 거리; 및

lcs(c₁, c₂): c₁ 및 c₂ 개념의 공통적인 가장 낮은 계층의(가장 구체적인)
개념(the lowest common subsumer)).

[청구항 7] 제1항 내지 제5항 중 어느 한 항에 따른 분자 네트워크, 화학적 특성 및
민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과
예측 방법을 실행시키기 위한 명령들을 포함하는 프로그램이 저장된
컴퓨터로 판독가능한 매체.

[청구항 8] 분자 네트워크를 이용하여 파이토케미컬의 건강효과를 추론하는 모듈;
파이토케미컬의 화학적 특성을 조사하여 생체이용률이 높은
파이토케미컬을 도출하는 모듈; 및
민족약학적 증거를 탐색하여 상기 추론된 건강효과 중 민족약학적
증거와 의미 유사도가 높은 파이토케미컬의 건강효과를 도출하는 모듈을

포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 시스템.

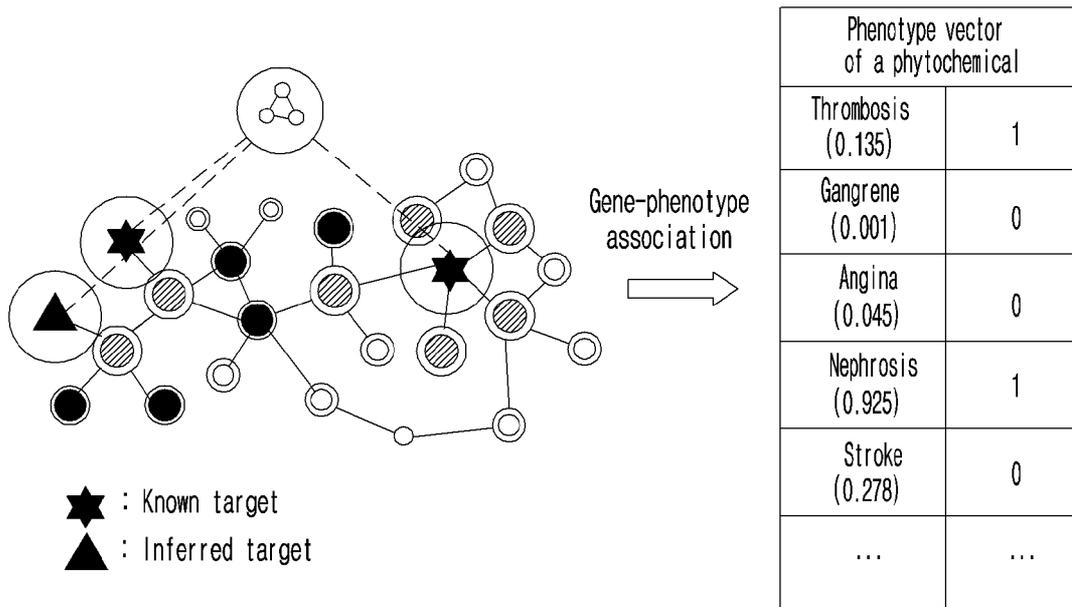
[청구항 9]

제8항에 있어서, 상기 분자 네트워크를 이용하여 파이토케미컬의 건강효과를 추론하는 모듈은 분자 네트워크에서 RWR 알고리즘을 수행하여 파이토케미컬의 표현형 벡터를 제작하는 모듈;

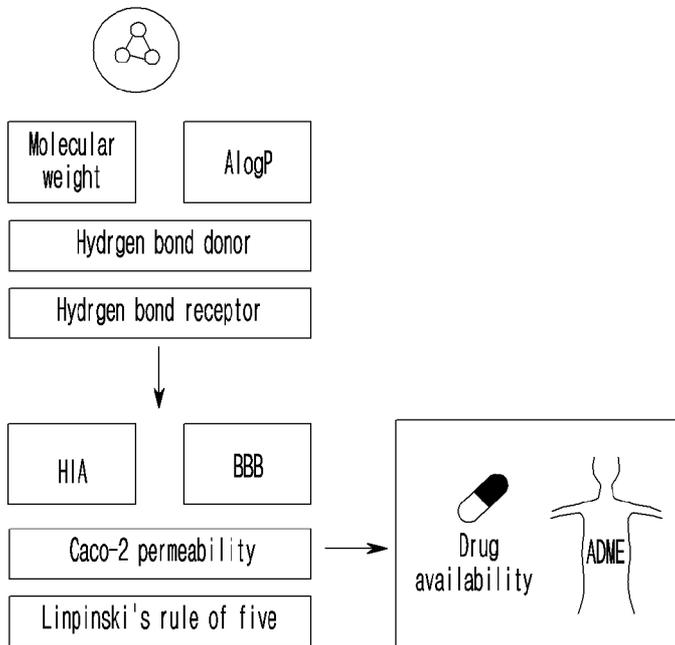
고정된 수의 표적 단백질들로부터 파이토케미컬의 표적들을 무작위로 선택하여 랜덤 표현형 벡터를 제작하는 모듈; 및

상기 랜덤 표현형 벡터에서 통계적으로 유의한 파이토케미컬의 표현형을 도출함으로써 파이토케미컬의 건강효과를 추론하는 모듈을 포함하는, 분자 네트워크, 화학적 특성 및 민족약학적 증거에 기반한 통합 분석을 이용한 파이토케미컬의 건강효과 예측 시스템.

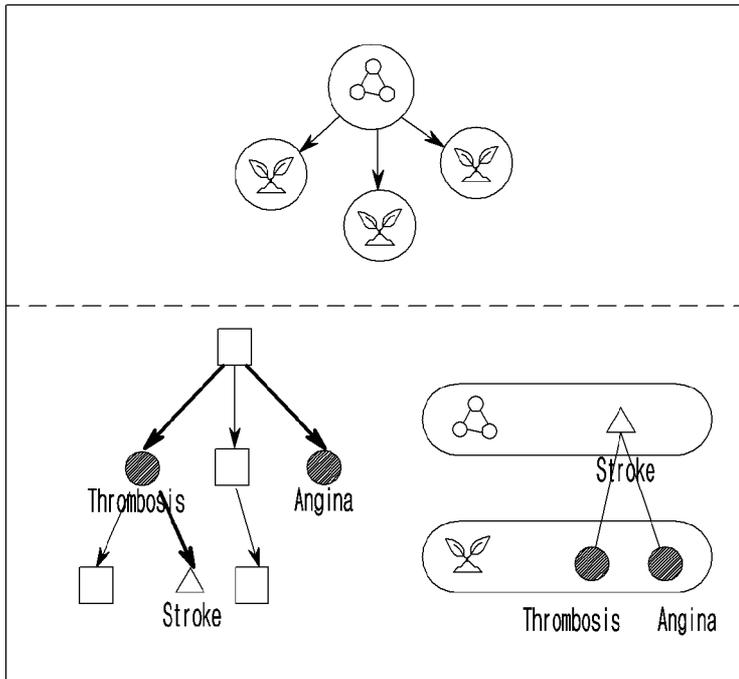
[도 1a]



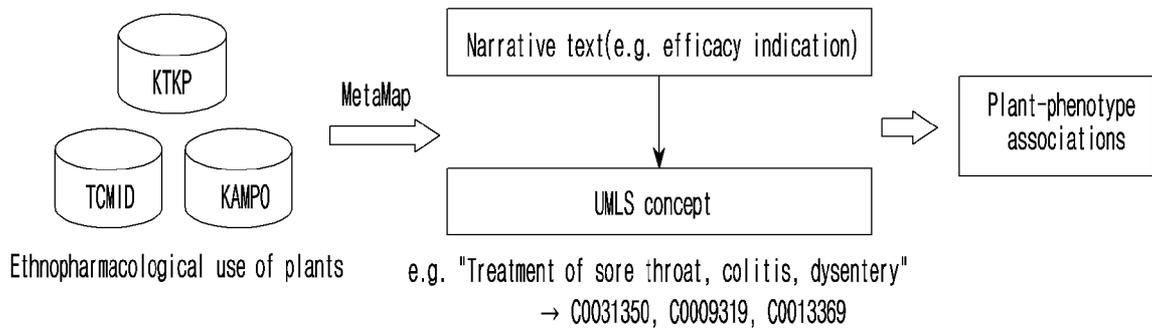
[도 1b]



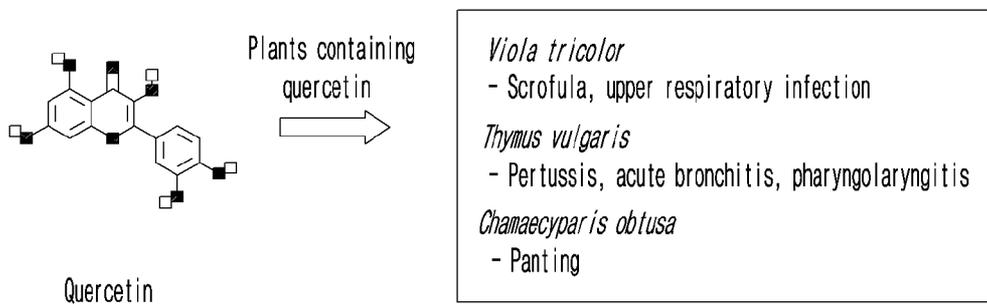
[도1c]



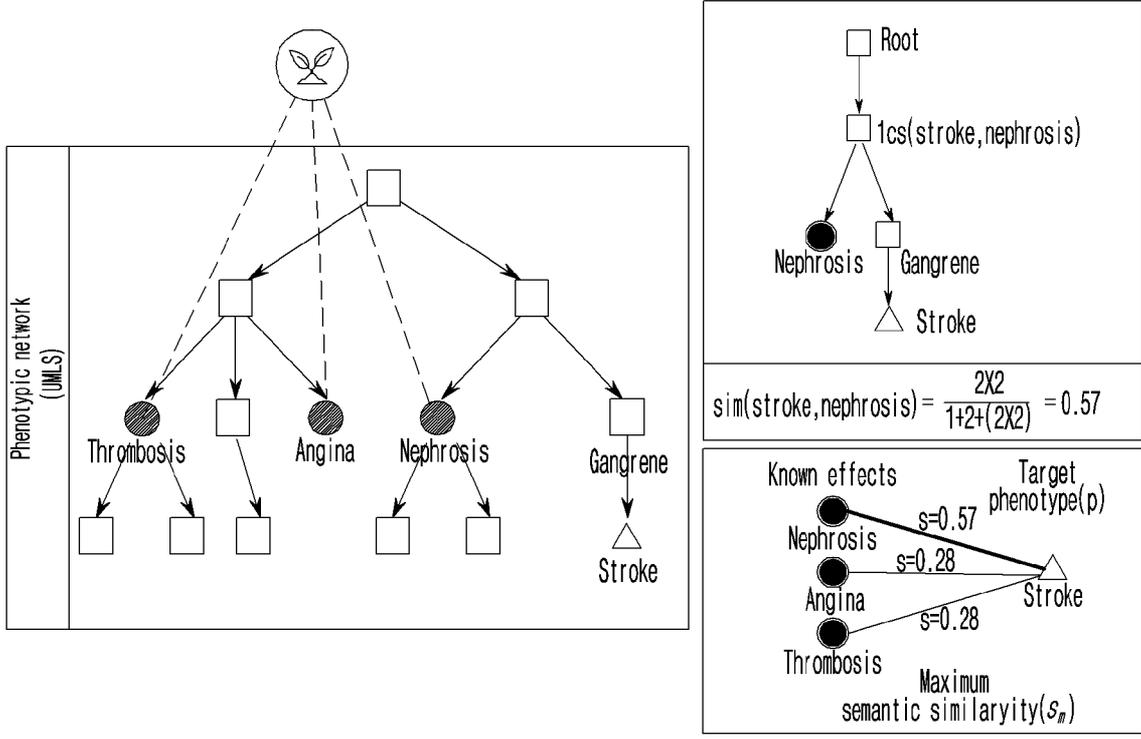
[도2a]



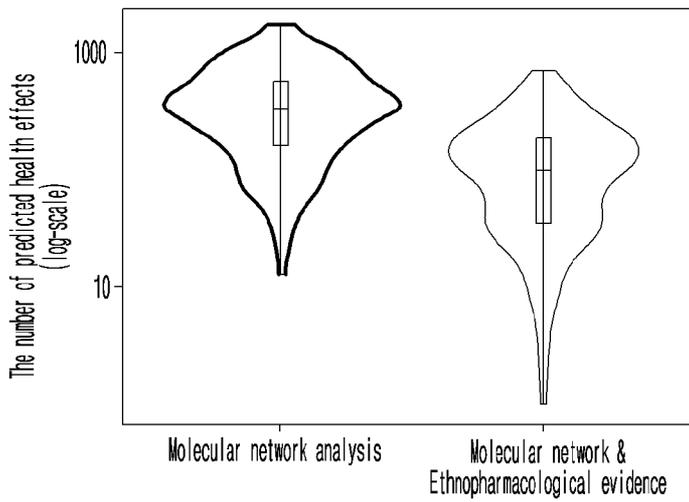
[도2b]



[도2c]



[도3]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/008244

A. CLASSIFICATION OF SUBJECT MATTER

G16H 50/30(2018.01)i, G16H 20/90(2018.01)i, G16H 50/70(2018.01)i, G16C 20/50(2019.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G16H 50/30; G01N 30/00; G01N 30/02; G16H 20/90; G16H 50/70; G16C 20/50

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Korean utility models and applications for utility models: IPC as above

Japanese utility models and applications for utility models: IPC as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

eKOMPASS (KIPO internal) & Keywords: phytochemical, ethnopharmacological evidence, health, benefit

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LIU, Rui Hai. Whole grain phytochemicals and health. <i>Journal of Cereal Science</i> . November 2007, vol. 46, no. 3, pages 207-219. See pages 207-217.	1-9
A	SAID, O. et al. Ethnopharmacological survey of medicinal herbs in Israel, the Golan Heights and the West Bank region. <i>Journal of Ethnopharmacology</i> . December 2002, vol. 83, no. 3, pages 251-265. See pages 251-264.	1-9
A	TARIQ, Akash et al. Review on ethnomedicinal, phytochemical and pharmacological evidence of Himalayan anticancer plants. <i>Journal of Ethnopharmacology</i> . 10 February 2015, vol. 164, pages 96-119. See pages 96-117.	1-9
A	ALBUQUERQUE, Tania Gocalves et al. Nutritional and phytochemical composition of <i>Annona cherimola</i> Mill. fruits and by-products: Potential health benefits. <i>Food Chemistry</i> . 15 February 2016, vol. 193, pages 187-195. See pages 187-194.	1-9
A	KR 10-2006-0132707 A (COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH) 21 December 2006. See claim 1.	1-9



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

17 OCTOBER 2019 (17.10.2019)

Date of mailing of the international search report

17 OCTOBER 2019 (17.10.2019)

Name and mailing address of the ISA/KR

Korean Intellectual Property Office
Government Complex Daejeon Building 4, 189, Cheongsa-ro, Seo-gu,
Daejeon, 35208, Republic of Korea

Facsimile No. +82-42-481-8578

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/KR2019/008244

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>YOO, Sunyong et al. Discovering Health Benefits of Phytochemicals with Integrated Analysis of the Molecular Network, Chemical Properties and Ethnopharmacological Evidence. <i>Nutrients</i>. 08 August 2018, vol. 10, no. 8 See pages 1-17.</p> <p>* The above document is a known document declaring exceptions to lack of novelty by the applicant.</p>	1-9

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/KR2019/008244

Patent document cited in search report	Publication date	Patent family member	Publication date		
KR 10-2006-0132707 A	21/12/2006	AU 2005-208148 A1	11/08/2005		
		CN 1938586 A	28/03/2007		
		CN 1938586 B	13/05/2015		
		EP 1709440 A2	11/10/2006		
		JP 05103021 B2	19/12/2012		
		JP 05743940 B2	01/07/2015		
		JP 2007-519921 A	19/07/2007		
		JP 2012-154937 A	16/08/2012		
		US 2007-0288217 A1	13/12/2007		
		US 2011-0040486 A1	17/02/2011		
		WO 2005-073713 A2	11/08/2005		
		WO 2005-073713 A3	13/10/2005		

A. 발명이 속하는 기술분류(국제특허분류(IPC)) G16H 50/30(2018.01)i, G16H 20/90(2018.01)i, G16H 50/70(2018.01)i, G16C 20/50(2019.01)i		
B. 조사된 분야 조사된 최소문헌(국제특허분류를 기재) G16H 50/30; G01N 30/00; G01N 30/02; G16H 20/90; G16H 50/70; G16C 20/50 조사된 기술분야에 속하는 최소문헌 이외의 문헌 한국등록실용신안공보 및 한국공개실용신안공보: 조사된 최소문헌란에 기재된 IPC 일본등록실용신안공보 및 일본공개실용신안공보: 조사된 최소문헌란에 기재된 IPC		
국제조사에 이용된 전산 데이터베이스(데이터베이스의 명칭 및 검색어(해당하는 경우)) eKOMPASS(특허청 내부 검색시스템) & 키워드: 파이토케미컬(phytochemical), 민족약학적 증거(ethnopharmacological evidence), 건강(health), 효과(benefit)		
C. 관련 문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
A	RUI HAI LIU. `Whole grain phytochemicals and health`. Journal of Cereal Science. 2007년 11월, 제46권, 제3호, 페이지 207-219. 페이지 207-217 참조.	1-9
A	O. SAID 등. `Ethnopharmacological survey of medicinal herbs in Israel, the Golan Heights and the West Bank region` Journal of Ethnopharmacology. 2002년 12월, 제83권, 제3호, 페이지 251-265. 페이지 251-264 참조.	1-9
A	AKASH TARIQ 등. `Review on ethnomedicinal, phytochemical and pharmacological evidence of Himalayan anticancer plants`. Journal of Ethnopharmacology. 2015년 2월 10일, 제164권, 페이지 96-119. 페이지 96-117 참조.	1-9
A	TANIA GOCALVES ALBUQUERQUE 등. `Nutritional and phytochemical composition of Annona cherimola Mill. fruits and by-products: Potential health benefits`. Food Chemistry. 2016년 2월 15일, 제193권, 페이지 187-195. 페이지 187-194 참조.	1-9
A	KR 10-2006-0132707 A (카운실 오브 사이언티픽 엔드 인더스트리얼 리서치) 2006. 12.21 청구항 1 참조.	1-9
<input checked="" type="checkbox"/> 추가 문헌이 C(계속)에 기재되어 있습니다. <input checked="" type="checkbox"/> 대응특허에 관한 별지를 참조하십시오.		
* 인용된 문헌의 특별 카테고리: “A” 특별히 관련이 없는 것으로 보이는 일반적인 기술수준을 정의한 문헌 “D” 본 국제출원에서 출원인이 인용한 문헌 “E” 국제출원일보다 빠른 출원일 또는 우선일을 가지나 국제출원일 이후 “X”에 공개된 선출원 또는 특허 문헌 “L” 우선권 주장에 의문을 제기하는 문헌 또는 다른 인용문헌의 공개일 또는 다른 특별한 이유(이유를 명시)를 밝히기 위하여 인용된 문헌 “O” 구두 개시, 사용, 전시 또는 기타 수단을 언급하고 있는 문헌 “P” 우선일 이후에 공개되었으나 국제출원일 이전에 공개된 문헌 “T” 국제출원일 또는 우선일 후에 공개된 문헌으로, 출원과 상충하지 않으며 발명의 기초가 되는 원리나 이론을 이해하기 위해 인용된 문헌 “X” 특별한 관련이 있는 문헌. 해당 문헌 하나만으로 청구된 발명의 신규성 또는 진보성이 없는 것으로 본다. “Y” 특별한 관련이 있는 문헌. 해당 문헌이 하나 이상의 다른 문헌과 조합하는 경우로 그 조합이 당업자에게 자명한 경우 청구된 발명은 진보성이 없는 것으로 본다. “&” 동일한 대응특허문헌에 속하는 문헌		
국제조사의 실제 완료일 2019년 10월 17일 (17.10.2019)	국제조사보고서 발송일 2019년 10월 17일 (17.10.2019)	
ISA/KR의 명칭 및 우편주소  대한민국 특허청 (35208) 대전광역시 서구 청사로 189, 4동 (둔산동, 정부대전청사) 팩스 번호 +82-42-481-8578	심사관 강민정 전화번호 +82-42-481-8131	

C(계속). 관련문헌		
카테고리*	인용문헌명 및 관련 구절(해당하는 경우)의 기재	관련 청구항
X	<p>SUNYONG YOO 등. `Discovering Health Benefits of Phytochemicals with Integrated Analysis of the Molecular Network, Chemical Properties and Ethnopharmacological Evidence`. Nutrients. 2018년 8월 8일, 제10권, 제8호. 페이지 1-17 참조. * 위 문헌은 출원인이 신규성 상실의 예외로서 선언한 공지 문헌임.</p>	1-9

국제조사보고서에서 인용된 특허문헌	공개일	대응특허문헌	공개일
KR 10-2006-0132707 A	2006/12/21	AU 2005-208148 A1 CN 1938586 A CN 1938586 B EP 1709440 A2 JP 05103021 B2 JP 05743940 B2 JP 2007-519921 A JP 2012-154937 A US 2007-0288217 A1 US 2011-0040486 A1 WO 2005-073713 A2 WO 2005-073713 A3	2005/08/11 2007/03/28 2015/05/13 2006/10/11 2012/12/19 2015/07/01 2007/07/19 2012/08/16 2007/12/13 2011/02/17 2005/08/11 2005/10/13