

The Cancer Genome Atlas (TCGA)/The Cancer Imaging Archive (TCIA)와 유방암 영상 표현형 연구

송성은 · 김원화 · 문우경

서울대학교병원 영상의학과

서론

2012년 유방암을 비롯한 암유전체지도(The Cancer Genome Atlas, TCGA) 프로젝트의 연구 결과가 네이처 등에 발표되었다(1). TCGA 프로젝트를 통해 유방암은 한 종류의 질환 군이 아니며 Luminal A, Luminal B, HER2-enriched, Basal-like 아형으로 나누어진다는 것을 확인했으며 유방암 아형에서 흔히 발견되는 돌연변이 유전자들이 존재함을 알게 되었다. 유전체 지도에서 발견된 새로운 지식을 유방암 환자의 예방, 진단 및 치료 방법 개발에 연결시키려는 노력들이 현재 활발히 진행 중이다. TCGA는 임상적, 유전학적, 병리학적 자료와 함께 영상 데이터도 함께 수집하였고 미국 국립암연구소(National Cancer Institute, NCI)의 종양영상아카이브(The Cancer Imaging Archive, TCIA)에 저장 관리되고 있다. 뇌종양, 폐암, 신장암, 전립선암, 유방암 데이터를 이용한 영상표현형 분석과 유전체 및 예후와의 연관성을 찾고자 하는 영상유전체(Imaging Genomics or Radiogenomics) 연구들이 진행 중이며 일부 결과들이 Radiology와 RSNA에서 발표되었다(2, 3).

영상 표현형(Imaging Phenotype) 연구의 가설은 영상에서 보이는 특징(Features)과 암의 생물학적 특성은 서로 상관관계가 있다는 것이다(4). 양성과 악성 종양의 소견, 관상피내암과 침윤암의 소견, 저등급암과 고등급암 또는 림프절 전이 암의 영상소견을 영상의학과 의사 또는

컴퓨터로 분석하고 암의 진단, 치료반응과 예후 등과 관련성이 높은 파라미터 또는 파라미터들의 조합과 예측 모델을 찾아내려는 연구들이 지난 20년간 이루어졌다. 최근에는 항암제 임상시험과 관련해서 분자아형별 영상 특징 및 치료반응과 관련된 영상 바이오마커를 발굴하고자 하는 연구가 진행 중이다(5, 6). 하지만 대규모 유전체 정보의 접근이 개별병원에서는 쉽지 않았기 때문에 현재까지 연구는 병리조직의 단순염색이나 면역염색에 근거한 분류와 비교하는 것이 대부분이었다. TCGA/TCIA 이용한 데이터의 사용이 가능해지면서 영상유전체 연구가 본격화될 것에 대한 기대와 관심이 높다.

본 종설에서는 유방암 영상 표현형 연구를 위해 진행되고 있는 TCGA/TCIA 데이터 수집과 자기공명영상(MRI) 영상분석 기준 및 최근 발표된 영상유전체 연구 결과를 소개하고 영상획득의 표준화 등 향후 해결해야 될 과제에 대해 논하고자 한다.

1. TCGA/TCIA 데이터 수집

TCGA는 대규모 유전체 시퀀싱을 포함한 게놈 분석 기술을 통해 암의 분자생물학적 기초에 대한 이해를 촉진하기 위해 2006년 시작되었으며 표준화를 통해 여러 기관의 데이터를 수집하였다. TCGA데이터에는 임상 진단, 치료 과정, 병리학적 진단, 종양 발생 부위, 수술 방법, 유전자 발현, RNA 시퀀스, 염색체 숫자, 종양 이질성, DNA 메틸레이션 패턴, miRNA 발현, DNA 시퀀스, 단백질과 인단백질 발현 및 CT, MRI와 PET 등의 영상 정보가 포함되어 있다(1, 2). 다형성 아교모세포종(Glioblastoma Multiforme), 난소암 및 폐암에 초점을 맞춘 TCGA 시범사업은 유전체 변화가 특정 종양 유형에 따라 구성 될 수 있음을 확인했으며, 시범사업의 성공은 유방암 등 20개 이상의 추가 종양 유형을 수집하고 특성을 저장하는 방향으로 확산되었다. 유방암에 대한

이 논문은 2014년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No.2012R1A2A1A01010846)

통신저자 : 문우경

(110-744) 서울시 종로구 연건동 28,

서울대학교 의과대학 영상의학교실

Tel. (02)2072-3928, Fax. (02)743-6385

E-mail: moonwk@snu.ac.kr

데이터(TCGA-BREAST)는 4가지 기본 분자 아형을 고려하여 저장되었다.

TCGA 프로젝트 환자의 고유번호와 연결된 영상 데이터는 TCIA에 저장되고 있으며 유전자형에 따른 표현형 분석에 초점을 맞추고 TCGA 참여기관에서 영상 데이터가 수집되고 있다. 메이요 클리닉, 로즈웰파크 암연구소, 메모리얼 슬로안케터링 암병원, 피츠버그대학과 마이애미대학 등이 유방암 영상소스사이트(Image Source Site)에 참여하고 있으며 2014년 현재 122명 유방암환자의 유방촬영술과 MRI 영상 데이터가 NCI의 중앙영상프로그램(Cancer Imaging Program, CIP)에 의해 연구자들에게 제공되고 있다(3). TCIA에는 TCGA이외에도 Quantitative Imaging Network(QIN)의

Breast DCE-MRI collection 등 다수의 공개 유방영상 데이터가 있으며 사용의 편의성을 보장하기 위해 영상정보학 관련 분석 소프트웨어 등도 제공하고 있다.

2. TCGA/TCIA 영상표현형 분석

영상 표현형 연구를 위해 TCGA내에 유방 표현형 연구자 그룹(TCGA Breast Phenotype Research Group)이 조직되었으며 이 연구자 그룹은 TCIA에 소속되어 있는 영상의학과 의사와 정량적 분석을 위한 컴퓨터 분석 전문 연구자로 구성되어있고 유전체, 병리, 임상 정보 등 다른 TCGA데이터와 비교하기 위한 유방영상특징 스코어링 프로젝트(Breast Imaging Feature Scoring Project)를 진행 중이다(7).

Table 1. Breast Feature Key by TCGA Breast Phenotype Research Group

Question Group	Feature	Definition
Breast Observation	1.1 Extent	Multifocal/ Multicentric/ Unifocal/ No lesions found/ Flag for review
	1.2 Flag for extensive heterogeneous disease	Yes/ No
	1.3 Fibroglandular Tissue (FGT)	Almost entirely fat/ Scattered FGT/ Heterogeneous FGT/ Extreme FGT/ Not applicable
	1.4 Associated Features	Nipple retraction/ Nipple invasion/ Skin retraction/ Skin invasion/ Edema/ Axillary lymphadenopathy/ Pectoralis invasion/ Chest wall invasion/ None apply
	1.5 Select the Dominant findings	Mass/ Nonmass/ Not applicable
Mass Observation	2.1 Breast Side	Left/ Right/ Not applicable
	2.2 Shape	Round-Oval/ Irregular/ Not applicable
	2.3 Internal enhancement	Homogeneous/ Heterogeneous/ Not applicable
	2.4 Clip in index lesion	Present/ Absent
	2.5 Dark internal septations	Present/ Absent/ Not applicable
	2.6 Rim enhancement	Present/ Absent/ Not applicable
	2.7 Margin	Circumscribed/ Not circumscribed-irregular/ Not circumscribed-spiculated
	2.8 T2 pattern	Homogeneous/ Heterogeneous/ Not applicable
	2.9 Predominant T2 signal intensity	Low/ Iso/ High/ Not applicable
	2.10 T2 peritumoral edema	Present/ Absent/ Not applicable
	2.11 Associate Nonmass Enhancement	Present/ Absent/ Not applicable
Nonmass Observation	3.1 Internal Enhancement	Homogeneous/ Heterogeneous/ Clumped/ Clustered Ring/ Not applicable
	3.2 Distribution	Focal/ Linear/ Segmental/ Regional/ Multiple regions/ Diffuse/ Not applicable
Measure diameter	4.1 Measurement	Longest diameter of the dominant finding within the ruler tool

유방 표현형 연구자 그룹은 미국방사선의학회(ACR) 영상보고데이터시스템(BI-RADS)에 기초한 분석 용어(Breast Feature Key)를 사용하여 유방 MRI 영상 표현형을 분류하고 있으며 3명의 의사가 시맨틱 표현형(Semantic Phenotype)을 점수화하여 TCGA 데이터에 케이스 매칭 하여 저장하고 있다. TCGA/TCIA 영상 분석에는 유방, 종괴, 비종괴 관찰과 직경 측정 등 4가지 목표 템플릿이 있다(Table 1). MRI BI-RADS와 비교하여 차이가 있는 것은 시간-신호 곡선 분류 등 역동학(Kinetics) 분석이 포함되지 않은 것과 T2강조영상에서 특징적인 소견 기술을 추가한 것으로 T2패턴(Homogeneous/ Heterogeneous/ Not applicable), T2신호강도(Low/ Iso/ High/ Not applicable), 종양 주위부종(Present/ Absent/ Not applicable)을 기록해야 한다. 또한 영상과 유전자 분석 대상간의 일치여부 확인을 위해 병변내부에 클립 유무를 기록하도록 했다. TCGA/TCIA에서는 영상 표현형 연구의 표준화를 위한 증례 품질 컨트롤 서식(Breast Case Quality Control Form), 등록 서식(Breast Enrollment Form), 추적 서식(Breast Follow-Up Form) 등을 임상적 특징, 유전학적 특징, 병리학적 특징에 대한 데이터와 함께 규격화하여 저장하고 있다.

정량적 표현형(Quantitative Phenotype)은 시카고 대학에서 만든 컴퓨터 비전(Computer Vision)이라는 프로그램을 사용하여 분석이 진행되고 있다. 이 컴퓨터 보조진단(Computer-aided Diagnosis, CAD) 프로그램은 유방과 병변을 자동 분획(Segmentation) 하고 유방 종양 병변과 종양 배경의 역동적 특성, 크기, 형태학적 특성 및 텍스처 특성을 얻은 후 그 정보를 종합해서 종양의 분자아형을 분류한다(그림 1).

3. TCGA/TCIA 데이터를 이용한 연구 사례

(1) 뇌종양 연구

에모리 대학의 연구진은 2013년 Radiology에 다형성 아교모세포종 데이터를 이용한 영상유전체 연구 결과를 발표하였다. TCGA에 유전 정보를 가지고 있는 75명 환자의 수술 전 MRI를 세 명의 신경전공 영상의학 의사가 크기, 위치, 종양 형태에 대하여 표준화된 특징 세트를 이용하여 판독하였으며, 생존율과 종양 크기, 종양 형태 사이의 연관성에 대하여 분석하였다. 종양의 조영 증강, 비 조영되는 부분, 괴사, 부종, 크기의 다양성과 생존율을 평가한 결과, 조영 증강되는 종양 부피와 장축 크기가 나쁜 생존율과 관련이 있었다. 전신경(Proneural) 아형의 아교모세포종이 다른 아형에서 보다 조영 증강이 적

게 되었고 중간엽(Mesenchymal) 아형은 비조영증강 종양의 비중이 낮았다(8). 또한, TP53 변이 종양은 T2 강조 영상과 FLAIR 영상에서 평균적으로 작은 크기로 보이며, EGFR 변이 종양은 TP53 변이 종양보다 통계적으로 유의하게 크며 5%이상의 조영 증강이나 종양 괴사와 관련성이 있었다.

Lee Moffitt 암센터의 연구진은 TCGA에 유전 정보를 가지고 있는 32명 아교모세포종 환자의 수술 전 MRI에서 공간 이질성을 조영증강영상과 T2강조 또는 FLAIR 영상의 히스토그램으로 분석하고 환자의 생존율과 비교하였다. 그 결과 낮은 조영증강을 보이면서 세포 밀도가 높은, 즉 완전 괴사가 아닌 영역(Habitat)을 많이 갖는 종양이 유의하게 나쁜 생존율 그룹과 연관성이 있었다(9).

(2) 유방암 연구

듀크 대학의 Mazurowski 등은 48명 유방암 환자의 TCGA/TCIA MRI 영상과 유전자 데이터를 이용하여 4가지 분자아형과 컴퓨터 비전 알고리즘으로 추출한 23개 영상특징 간에 연관성이 있는지를 분석하였다. 그 결과 Luminal B 아형과 조영증강관련 특징에 유의한 상관관계가 있었다($P=.0015$). 즉, 배경질질에 비해 병변 조영증강이 강한 암은 Luminal B 아형일 가능성이 높았다(10).

시카고대학의 Giger 등은 유방 종양의 크기, 형태, 이질성의 정량적 표현형 특징 등을 영상의학과 의사의 시맨틱 표현형과 비교한 결과, 0.8-0.9 정도의 일치도(Interclass Correlation)를 보고하였다. 또한 TCGA/TCIA 데이터를 이용해서 정량적 표현형 특징과 유방암 수용체 예측에 관한 연구결과를 발표하였는데 98명 환자에서 ER, PR, HER2 양성 예측의 AUC 값은

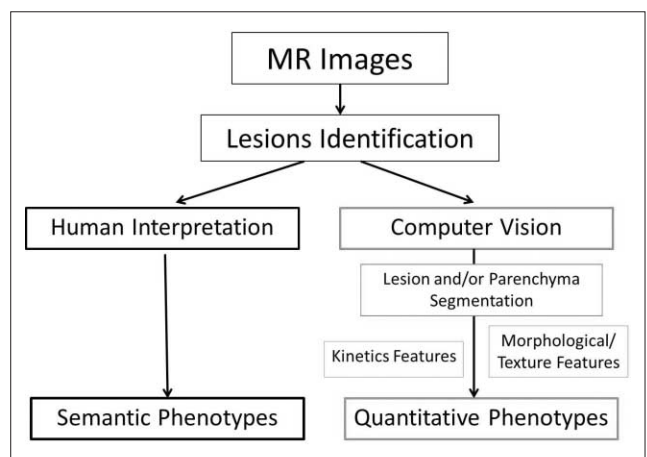


그림 1. MRI 영상에서 시맨틱 표현형과 정량적 표현형의 도출 과정

각각 0.79, 0.68, 0.61이었고 종양 볼륨, 텍스처와 역동학 관련 특징들이 분류에 유용하였다(11).

4. 향후 과제

영상표현형 연구를 위한 대규모 데이터 수집과 관리 및 영상 분석 등을 위해서는 몇가지 해결해야 될 과제들이 있다. 첫번째, 각 기관마다 검사 기종이나 영상 획득과 영상 후처리 방식이 다르다는 점이다. 조영증강 MRI 검사 방법의 차이는 영상 특징 분석 결과의 차이를 초래할 수 있다. 이와 관련하여 QIN과 Quantitative Imaging Biomarker Alliance(QIBA) 등 영상획득과 분석 및 계측(Metrology)의 표준화운동이 RSNA 등 영상관련 학회에 의해 진행되고 있다(12). 두번째, 표현형 연구를 위한 영상 분석 방법의 문제점이다. 시멘틱 표현형은 BI-RADS 기준에 따라 3명의 유방영상 전문가들이 분석, 증례마다 메모를 달고 있으나 판독자간 편차 및 주관성을 배제하기 힘들다. 한편 정량적 표현형 데이터 분석에는 시카고 대학에서 만든 컴퓨터 비전 프로그램을 사용하고 있으나 다른 기관에서는 접근이 어렵다. 대규모 데이터의 재현성 높은 분석을 위해서는 국제적으로 널리 사용할 수 있는 CAD 프로그램의 개발과 보급이 진행되어야 할 것이다. 세번째, 영상의학과 의사에게는 익숙하지 않은 빅데이터 처리 등 영상정보학에 대한 이해와 암 유전체학 및 생물학에 대한 지식이 필요하다(13). 해당 분야 전문가와의 협력체계 구축과 영상유전체학 연구를 수행할 영상의학과 의사의 훈련과 전문가 육성이 진행되어야 한다. 마지막으로 한국인 여성으로 부터 획득한 유방암 영상 데이터베이스 구축이 필요하다. 젊은 여성, 치밀 유방 여성의 표현형 연구는 우리나라 여성의 유방암 조기진단 및 치료, 예방법 개발에 도움을 줄 것이며 수집된 데이터는 오픈소스 형태로 해외 연구진과 공유 가능할 것이다.

결론

TCGA/TCIA가 제공하는 표준화된 영상 데이터와 분

석 기준은 유방암 표현형 연구를 촉진하기 위한 포괄적인 노력의 일환으로 영상의학과 의사들에게 새로운 도전과 기회를 제공하고 있다. 유방암의 영상 표현형과 분자생물학적 상관관계에 대한 이해는 향상된 유방암 진단과 예후 예측평가를 가능케 하고, 좀 더 효과적인 치료 및 예방 계획을 수립하는데 이용할 수 있을 것이다.

참고문헌

1. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012;490(7418):61-70
2. TCGA: cancergenome.nih.gov
3. TCIA: cancerimagingarchive.net
4. Colen R, Foster I, Gatenby R, et al. NCI workshop report: clinical and computational requirements for correlating imaging phenotypes with genomics signatures. *Transl Oncol* 2014;7(5):556-569
5. Barker AD, Sigman CC, Kelloff GJ, Hylton NM, Berry DA, Esserman LJ. I-SPY 2: an adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clin Pharmacol Ther* 2009;86(1):97-100
6. Mukhtar RA, Yau C, Rosen M, et al. Clinically meaningful tumor reduction rates vary by prechemotherapy MRI phenotype and tumor subtype in the I-SPY 1 TRIAL (CALGB 150007/150012; ACRIN 6657). *Ann Surg Oncol* 2013;20(12):3823-3830
7. <https://wiki.cancerimagingarchive.net/display/Public/TCGA+Breast+Phenotype+Research+Group>
8. Gutman DA, Cooper LA, Hwang SN, et al. MR imaging predictors of molecular profile and survival: multi-institutional study of the TCGA glioblastoma data set. *Radiology* 2013;267(2):560-569
9. Zhou M, Hall L, Goldgof D, et al. Radiologically defined ecological dynamics and clinical outcomes in glioblastoma multiforme: preliminary results. *Transl Oncol* 2014;7(1):5-13
10. Mazurowski MA, Zhang J, Grimm LJ, et al. Radiogenomic analysis of breast cancer: luminal B molecular subtype is associated with enhancement dynamics at MR imaging. *Radiology* 2014;273(2):365-372
11. Giger ML. Relationship of quantitative MRI-based phenotypes and the molecular classifications of breast cancers in the TCGA/TCIA dataset. SSJ01-06. 2014 RSNA.
12. http://qibawiki.rsna.org/index.php?title=Main_Page
13. Kuo MD, Jamshidi N. Behind the numbers: decoding molecular phenotypes with radiogenomics--guiding principles and technical considerations. *Radiology* 2014;270(2):320-325