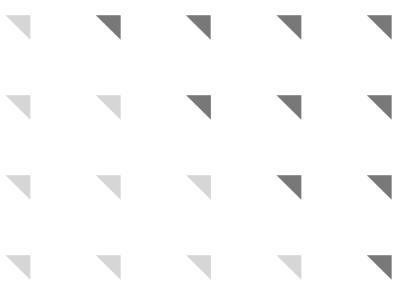


제129회 한림원탁토론회

‘데이터 사이언스와 바이오 강국 코리아의 길’

2018년 9월 18일(화), 10:00
엘타워 메리골드 B홀(5F)





초대의 말씀

인공지능(AI), 사물인터넷(IoT), 로봇산업 등 소위 4차 산업혁명으로 일컬어지는 여러 혁신적인 과학기술의 근간에는 ‘빅데이터’로 통칭되는 데이터 사이언스가 주도적인 역할을 해오고 있으며 향후에 그 중요성은 더 확대될 것이라는 것은 이제 주지의 사실입니다.

특히, 4차 산업혁명과 관련된 여러 산업 중 바이오 산업은 우리나라의 주요한 미래의 성장 동력이자 빅데이터와 아주 밀접한 관련이 있는 산업입니다. 또한, 데이터 사이언스 학문과 빅데이터 기술을 통한 바이오정보의 분석 활용 기술은 바이오 산업 및 보건, 의학, 약학, 치의학 등의 다양한 기초연구 분야의 핵심기술로 자리매김하고 있으며, 앞으로 ‘정밀 의학’ 시대를 열기 위한 필수적인 요소가 되고 있습니다.

바이오산업분야의 국가 경쟁력 제고를 위해 아직 우리나라에서는 도입 초기에 있는 데이터 사이언스를 바이오 산업분야에서 집중 조명해보고, 이를 통해 데이터 사이언스 인력 양성 및 데이터 사이언스를 바이오 산업, 혹은 의료 산업에 어떻게 연결해야 되는지에 대한 정책제안은 이제 더 늦출 수 없는 시기가 되었다고 생각합니다.

이에, 우리 한림원은 빅데이터 및 의·약학 관련 전문가들을 모시고 ‘데이터 사이언스와 바이오 강국 코리아의 길’을 주제로 제129회 한림원탁토론회를 개최하고자 합니다. 관련 과학기술계와 기업체 종사자들이 많이 참석하시어 우리나라가 데이터사이언스 및 바이오 강국으로 갈 수 있도록 고견을 내어 주시기 바랍니다.

감사합니다.

2018년 9월
한국과학기술한림원 원장 이명철

한림원탁토론회는 국가 과학기술의 장기적인 비전과 발전전략을 마련하고 국가사회 현안문제에 대한 과학기술적 접근 및 해결방안을 도출하기 위해 개최되고 있습니다.

PROGRAM

제129회 한림원탁토론회 '데이터 사이언스와 바이오 강국 코리아의 길'

사회: 유욱준 한림원 총괄부원장

09:30~10:00 (30') 등록

10:00~10:10 (10') 개회사

이명철 한국과학기술한림원 원장

10:10~10:30 (20') 주제발표 1

‘데이터 사이언스와 정밀 의학’

박태성 한림원 이학부 정회원(서울대학교)

10:30~10:50 (20') 주제발표 2

‘데이터 사이언스와 의료 민주화’

윤형진 서울대학교병원 의생명연구원 정보실장

10:50~11:10 (20') 주제발표 3

‘브레인 이미징(Brain Imaging)과 데이터 사이언스’

이동수 한림원 의약학부 정회원(서울대학교)

PROGRAM

제129회 한림원탁토론회 '데이터 사이언스와 바이오 강국 코리아의 길'

11:10~11:40 (30') **지정토론**

- 좌 장: **박성현** 한림원 이학부 종신회원(사회적책임경영품질원)
- 토론자: **김호** 한국보건정보통신학회 회장
이영조 한림원 이학부 정회원(서울대학교)
이태림 방송통신대학교 교수
정진호 한림원 의약학부장(서울대학교)
한현욱 차의과대학교 부교수

11:40~12:10 (30') **종합토론**

12:10 **폐회**

CONTENTS

제129회 한림원탁토론회 '데이터 사이언스와 바이오 강국 코리아의 길'

I. 주제발표 1 '데이터 사이언스와 정밀 의학'	1
• 박태성 한림원 이학부 정회원(서울대학교)	
II. 주제발표 2 '데이터 사이언스와 의료 민주화'	35
• 윤형진 서울대학교병원 의생명연구원 정보실장	
III. 주제발표 3 '브레인 이미징(Brain Imaging)과 데이터 사이언스'	57
• 이동수 한림원 의약학부 정회원(서울대학교)	
IV. 지정토론 (좌장: 박성현 한림원 이학부 종신회원(사회적책임경영품질원))	73
• 김 호 한국보건정보통계학회 회장	77
• 이영조 한림원 이학부 정회원(서울대학교)	83
• 이태림 방송통신대학교 교수	85
• 정진호 한림원 의약학부장(서울대학교)	89
• 한현욱 차의과대학교 부교수	93

주제발표 1

I

데이터 사이언스와 정밀 의학

발제자 약력

성명	박 태 성	
소속	서울대학교 자연과학대학 통계학과	
1. 학력		
기간	학교명	전공 및 학위
1980~1984	서울대학교	계산통계학
1984~1986	서울대학교	통계학
1986~1990	University of Michigan	생물통계학
2. 주요 경력		
기간	기관명	직위, 직책
2012~2016	고차원 생물정보통계연구단	창의연구단장
2007~2009	서울대학교 통계학과	학과장
2005~2008	서울대학교 협동과정생물정보학전공	전공주임
2009~2010	미국 워싱턴대학교 생물통계학과	방문교수
2002~2003	미국 피츠버그 대학교 생물통계학과	방문교수
2001~현 재	서울대학교 통계학과	교수
1999~2001	서울대학교 통계학과	부교수
1992~1996	한국외국어대학교 통계학과	부교수
1991~1992	National Institute of Child Health and Human Development National Institutes of Health, USA	방문연구원
1990~1991	미국 아이오와 대학교 예방의학과	방문연구원

발제 1

데이터 사이언스와 정밀 의학

• • •

박 태 성
한림원 이학부 정회원 (서울대학교)

129회 한림원 토론회
2018년 9월 18일

데이터 사이언스와 정밀의학

- » 통계학 = 데이터 사이언스?
- » 정밀의학
- » 사례들
 - 유방암: OncotypeDX
 - 간암
 - 췌장암
 - IPMN Nomogram
- » Conclusion

서울대학교 통계학과 박태성



Data Science

Prediction Medicine

Examples

Conclusion

데이터 사이언스

- **Data science** is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data in various forms, both structured and unstructured, similar to data mining.
- **Data science** is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, information science, and computer science.

2
Wikipedia

Data Science

Prediction Medicine

Examples

Conclusion

Statistics = Data Science?

- In November 1997, C.F. **Jeff Wu** gave the inaugural lecture entitled "Statistics = Data Science?" for his appointment to the H. C. Carver Professorship at the University of Michigan.
- Statistical work as a trilogy of
 1. data collection,
 2. data modeling and analysis, and
 3. decision making.
- **Statistics = data science**
- **Statisticians = data scientists**

3
Wikipedia

Data Science Prediction Medicine Examples Conclusion

데이터 사이언스

Harvard Business Review

ARTWORK: TAMAR COHEN, ANDREW J. BUBOLTZ, '2011, SILK SCREEN ON A PAGE FROM A HIGH SCHOOL YEARBOOK, 8.5" X 12"

DATA

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

WHAT TO READ NEXT

What Data Scientists Really Do, According to 35 Data Scientists

Data Science Prediction Medicine Examples Conclusion

데이터 사이언스

Nate Silver: What I need from statisticians

FEATURES

Author: Statistics Views

Date: 23 8월 2013

Copyright: Image copyright of the American Statistical Association

Exclusive interview: Nate Silver was the guest speaker at this year's [Joint Statistical Meetings](#) in Montreal, Canada where he gave the President's Invited Address. He is currently the editor-in-chief of [FiveThirtyEight](#) blog which will shortly move to [ESPN](#) and a Special Correspondent for ABC News but he became a household name when he correctly predicted 49 out of 50 states in the US Presidential election in 2008 by using Bayesian statistics. Four years later, he correctly predicted the winner of all 50 states and the District of Columbia during the 2012 US Presidential Elections between President Barack Obama and Senator Mitt Romney.

ASA

Data Science Precision Medicine Examples Conclusion

데이터 사이언스

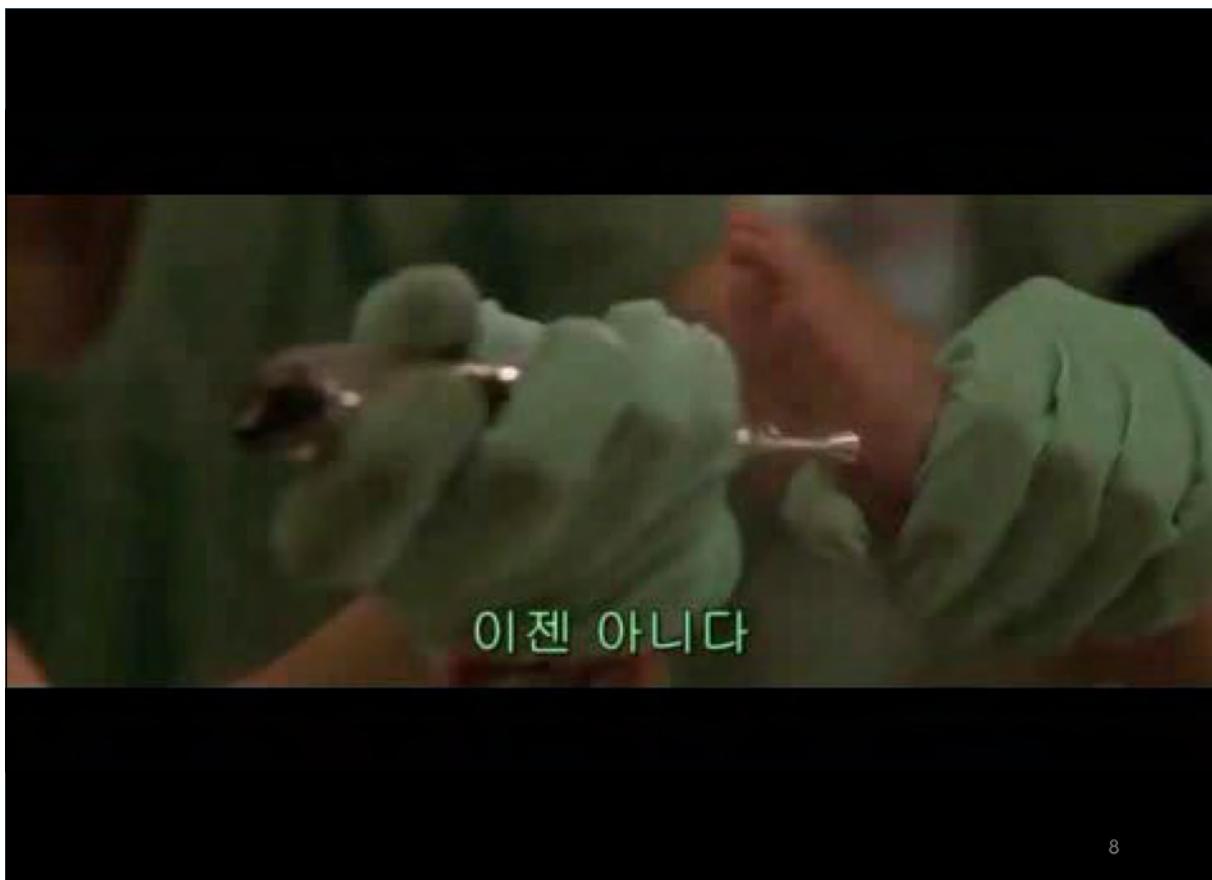
Methods can be abused and that is true with any approach, including Bayesian but it offers a more coherent, philosophical way in which to look at the world and...would be useful for journalists to use as well.

"I think data scientist is a sexed-up term for a statistician."

ASA

Precision Medicine

» Precision Medicine



8

Data Science

Prediction Medicine

Examples

Conclusion

Precision Medicine

Jan 30: Obama Unveils “Precision Medicine” Initiative



“Most medical treatments have been designed for the ‘average patient’ ... treatments can be very successful for some patients but not for others.”



<http://www.msnbc.com/msnbc/obama-seeks-215-million-personalized-medicine>

© 2014 MapR Technologies. **MAPR** 12

9

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

The diagram shows a progression from a large, diverse group of people (labeled 'One-size fits-all medicine') to a smaller, more homogeneous group (labeled 'Stratified medicine'), and finally to a single, highly personalized individual (labeled 'Precision medicine').

Stratification: Patients are grouped by: Disease Subtypes Demographics Clinical features Biomarkers

Personalisation: Patient individual: Preferences, Clinical features Medication history Environment Behaviours & habits Biomarker

(Original Graphic Via CPGR)

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

WITHOUT PRECISION MEDICINE: A single patient receives the same therapy. Some benefit, others do not, and there are adverse effects.

WITH PRECISION MEDICINE: Each patient benefits from tailored therapy based on DNA tests.

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

Why Precision Medicine?



Increases survival rates



Targets tumors with greater accuracy



Mitigates unnecessary treatments



Reduces prescription errors

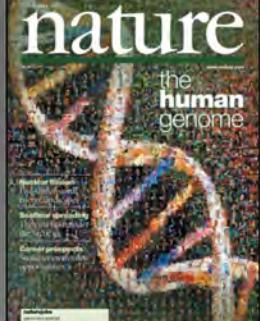
Impact of Precision Medicine in Personal Healthcare - Source Weillcornell.org

12

Data Science Prediction Medicine Examples Conclusion

Human Genome Project

- The project with the goal of determining the sequence of nucleotide base pairs that make up human DNA.
- It remains the world's largest collaborative biological project. the project
- Formally launched in 1990 and was declared complete in 2003.
- Most of the government- sponsored sequencing was performed in twenty universities and research centers in the United States, the United Kingdom, Japan, France, Germany, Canada, and China.





International Human Genome Sequencing Consortium

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

- 유전체서열분석기술의 발전
 - 초고속 (13년 → 5일 → 1일)
 - 저비용 (3조원 → 300만원 → 100만원)
- 생물정보 자료의 샘플 증가
- 개인별 유전체 분석 프로젝트 가능

\$1000 Genomes Era

Data Science Prediction Medicine Examples Conclusion

Single Nucleotide Polymorphism, SNP

- Single Nucleotide Variation (SNV)
 - Single base (A,C,G,T) insertion or deletion, a transition or transversion
 - Occupy 1% of entire genome

Data Science Prediction Medicine Examples Conclusion

Phenotypic Difference

Fat



Not Between Fats and Thins

Skin colors



Height



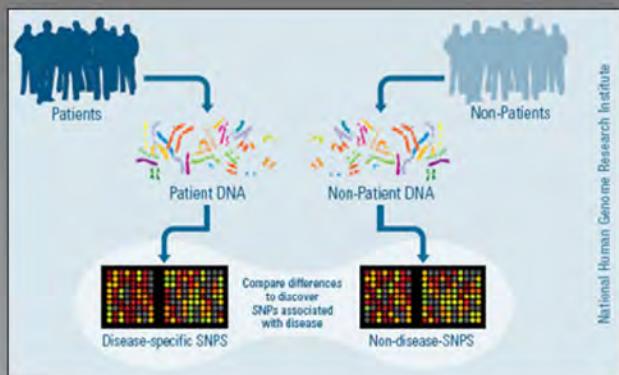
Case/ Control



Data Science Prediction Medicine Examples Conclusion

Genome Wide Association Studies (GWAS)

- Studies of genetic variation across the entire genome
- Designed to identify associations
 - ✓ genetic markers & observable traits
 - ✓ genetic markers & the presence/ absence of a disease
- Rely on research tools and technologies



National Human Genome Research Institute

Disease-specific SNPs

Non-disease-SNPs

Compare differences to discover SNPs associated with disease

Data Science Prediction Medicine Examples Conclusion

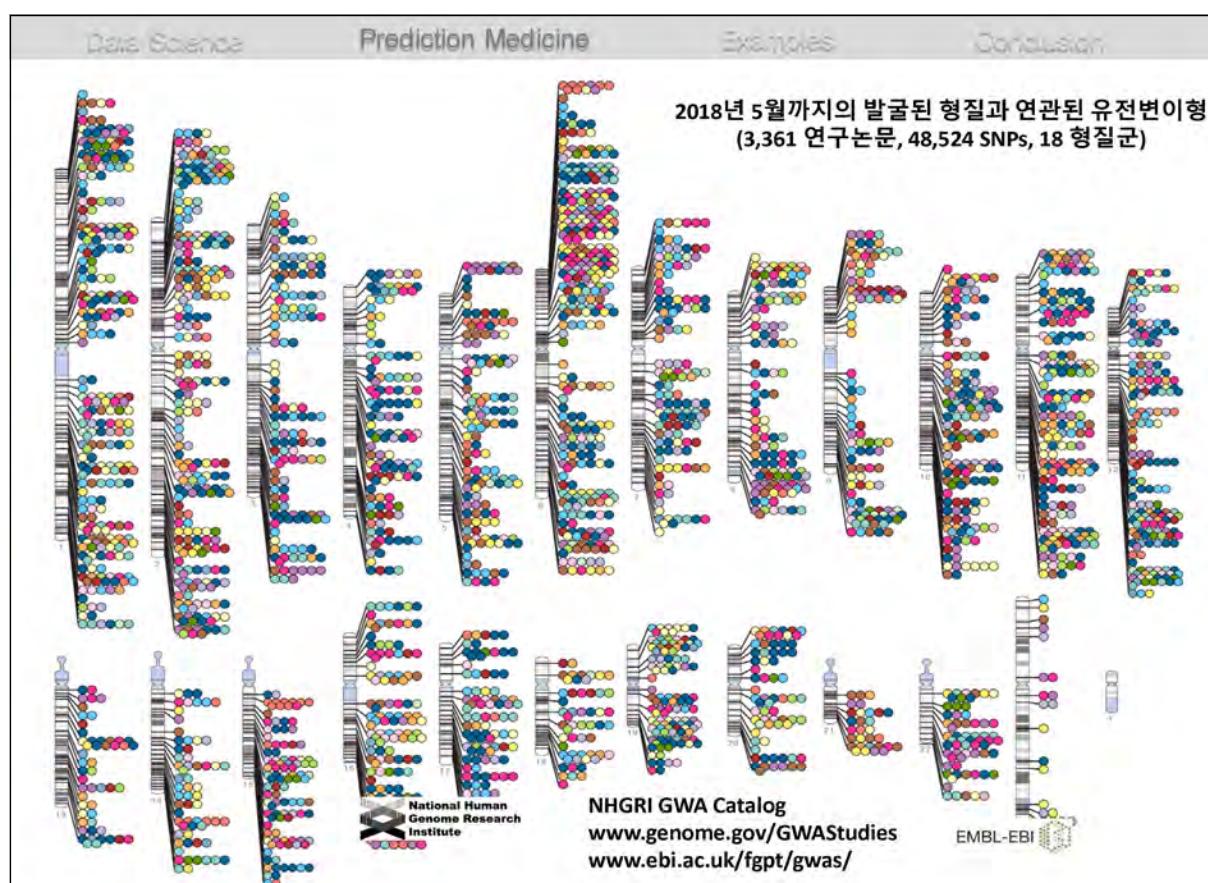
질환과 연관 있는 SNP 찾기: Association Test

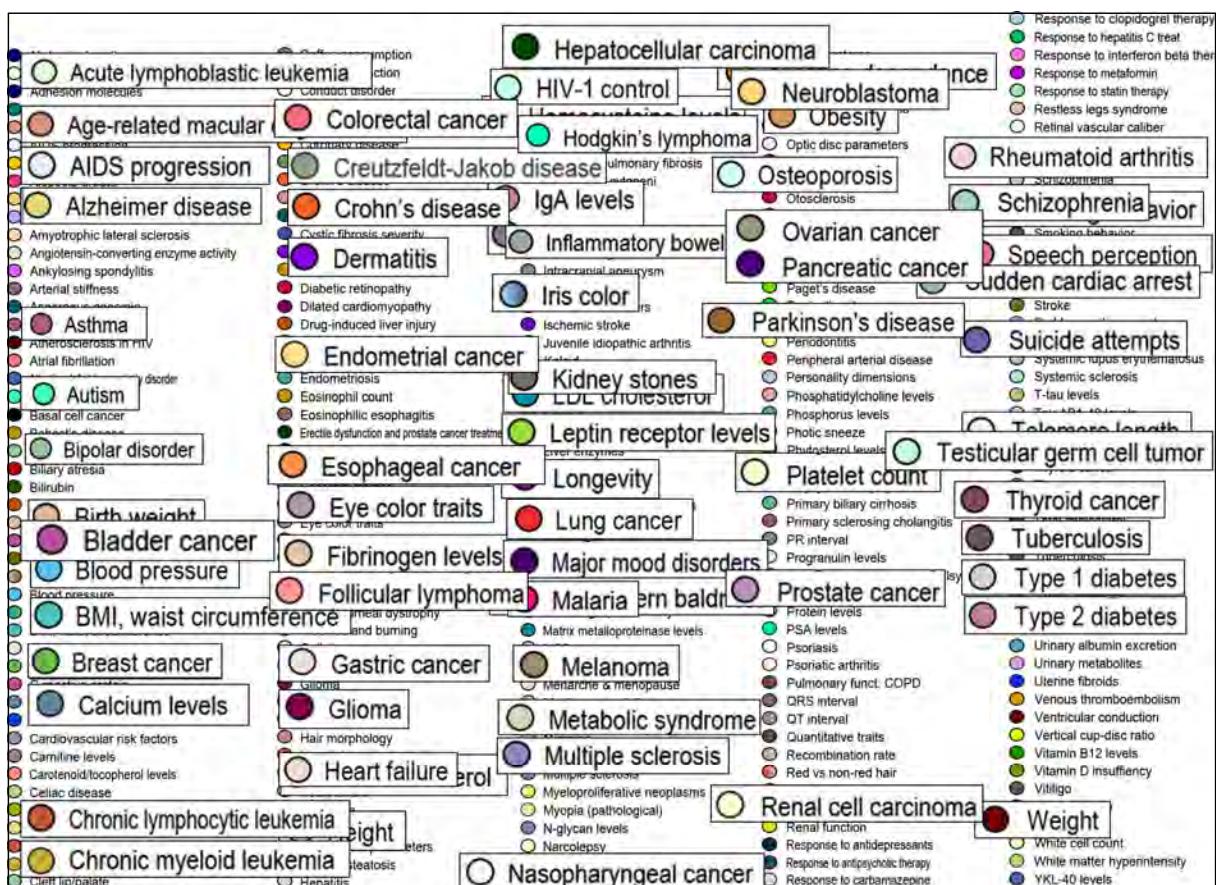
정상인 (대조군)

환자 (실험군)

질환여부	S_1	S_2	S_3	...	S_p
X	AA	BB	Cc	...	Xx
X	Aa	BB	CC	...	XX
X	AA	Bb	Cc	...	xx
X	aa	Bb	cc	...	xx
O	aa	bb	Cc	...	XX
O	Aa	bb	CC	...	XX
O	aa	Bb	cc	...	Xx
O	aa	ab	cc	...	XX

18





Data Science Prediction Medicine Examples Conclusion

Precision Medicine

- 적용 사례 (해외)
 - 유전형검사를 통해 **87% 유방암, 50% 자궁암** 확률이 예측되었음
 - **유방절제, 자궁적출수술**을 받아서 관련 질병을 예방하였음
 - 2013년 안젤리나 줄리의 수술 사실 공개 후, **유방암 관련 유전형 검사 수요 40% 증가**



“It gave me an estimated 87 percent risk of breast cancer and a 50 percent risk of ovarian cancer.”
Angelina Jolie-Pitt, The New York Times

Angelin Jolie-Pitt removes ovaries, fallopian tubes
REDUCES RISK OF OVARIAN CANCER BY 99%

abc ACTION NEWS

BIBS Department of Statistics Seoul National University

Data Science Prediction Medicine Examples Conclusion



The New York Times

“I can tell my children that they don't need to fear they will lose me to breast cancer.”

Angelina Jolie in “My Medical Choice”
Published on May 14, 2013

23

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

- 정밀의학을 통한 질환 예측 서비스 과정

샘플수집 유전체 자료생성 유전체 자료분석 분석결과 열람 의료지원 연계

SNP Chip

- 가족 관계 분석
- 질환 별 위험도
- 약물 부작용 위험도

- 유전자 검사에 대한 건강보험 적용 (2017년 3월 1일)
 - 고령암 10종: 위암, 폐암, 대장암, 유방암, 난소암, 악성 뇌종양
 - 혈액암 6종: 급성 골수성, 림프구성 백혈병 등
 - 유전성 질환 4개: 난청, 망막색소변형 등

24

Data Science Prediction Medicine Examples Conclusion

Precision Medicine

- 질환 예측 서비스 예시 (해외) – 23andMe

TIME Specials Best Inventions of 2008

23andMe genetics just got personal.

Start filling in the gaps with your DNA

\$99* Our new low price for all! Was: \$199

Order Now >

Discover Health & Ancestry Keep Your Doctor Informed Participate In Research

Data Science
Prediction Medicine
Examples
Conclusion

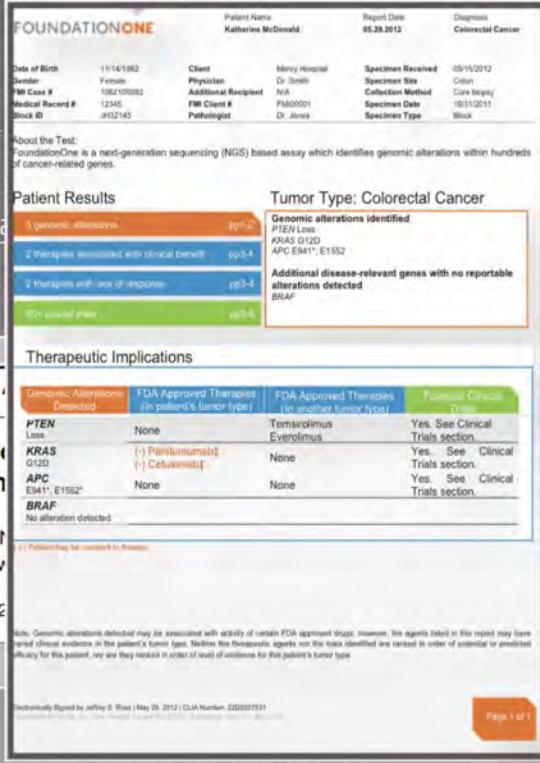
Precision Medicine

- 질환 예측 서비스 예시 (해외)
 - 스티브잡스도 Foundation One 변이는 찾았지만 치료제의 부재로
 - Google: 직원에게 Foundation



A View from...
Steve Jobs and Precision Medicine
A type of DNA sequencing is becoming very personal
September 2012

출처: MIT Technology Review



Patient Name: Katherine McDonald **Report Date:** 05.29.2012 **Diagnosis:** Colorectal Cancer

Specimen Received: 05/15/2012 **Specimen Site:** Colon **Collection Method:** Core biopsy **Specimen Date:** 10/11/2011 **Specimen Type:** Block

About the Test: FoundationOne is a next-generation sequencing (NGS) based assay which identifies genomic alterations within hundreds of cancer-related genes.

Patient Results:

- 3 germline alterations (0.0%)
- 2 alterations associated with clinical benefit (0.0%)
- 2 alterations with risk of response (0.0%)
- 100% normal areas (0.0%)

Tumor Type: Colorectal Cancer

Genomic alterations identified:

- KRAS G12D
- APC E141*, E1552*

Additional disease-relevant genes with no reportable alterations detected:

- BRAF

Therapeutic Implications:

Genomic Alterations Detected	FDA Approved Therapies (in patient's tumor type)	FDA Approved Therapies (in another tumor type)	Further Clinical Trials
PTEN Loss	None	Temozolomus, Everolimus	Yes. See Clinical Trials section.
KRAS G12D	(+) Panitumumab [†] , (+) Cetuximab [†]	None	Yes. See Clinical Trials section.
APC E141*, E1552*	None	None	Yes. See Clinical Trials section.
BRAF No alteration detected			

[†] Patient may be resistant to these.

Notes: Genomic alterations detected may be associated with activity of certain FDA approved drugs. However, the agents listed in this report may have limited clinical evidence in the patient's tumor type. Neither the therapeutic agents nor the sites identified are ranked in order of potential or predicted utility for this patient, nor are they ranked in order of level of evidence for that patient's tumor type.

Electronically Signed by Jeffrey S. Ross | May 29, 2012 | CLIA Number: 22D327631

© 2012 Foundation Medicine, Inc. All rights reserved. Foundation Medicine and the Foundation Medicine logo are registered trademarks of Foundation Medicine, Inc.

Page 1 of 1

통계 예측모형 사례

- 유방암: OncotypeDX
- 간암
- 췌장암
- IPMN Nomogram

Introduction Prediction Models Examples Considerations

OncoType DX

- » OncotypeDx 는 특정유방암(lymph node negative, estrogen receptor-positive tumors receiving tamoxifen) 환자에 대해 향후 10년 내에 재발 위험성에 대해 예측함 (distant recurrence)하여 화학요법 치료를 결정할 수 있는 근거를 제시함
- » 2018년 5월 90개 이상의 나라에서 90만명 이상의 환자들이 혜택을 받았음
- » 가격은 \$4,175
- » *New England Journal of Medicine* 2004, 현재까지 4,996회 피인용 되었음

예측마커 선택 > 예측모형 > 모형평가 > 임상적용

250 후보 유전자
RT-PCR
독립 표본에 대해
고/중/저 위험군 예측

16개 유방암 유전자
5개 참고 유전자
재발지수계산
고/중/저 위험군

oncotype DX
Uncover the Unexpected

Introduction Prediction Models Examples Considerations

OncoType DX

16 cancer related genes
5 reference genes

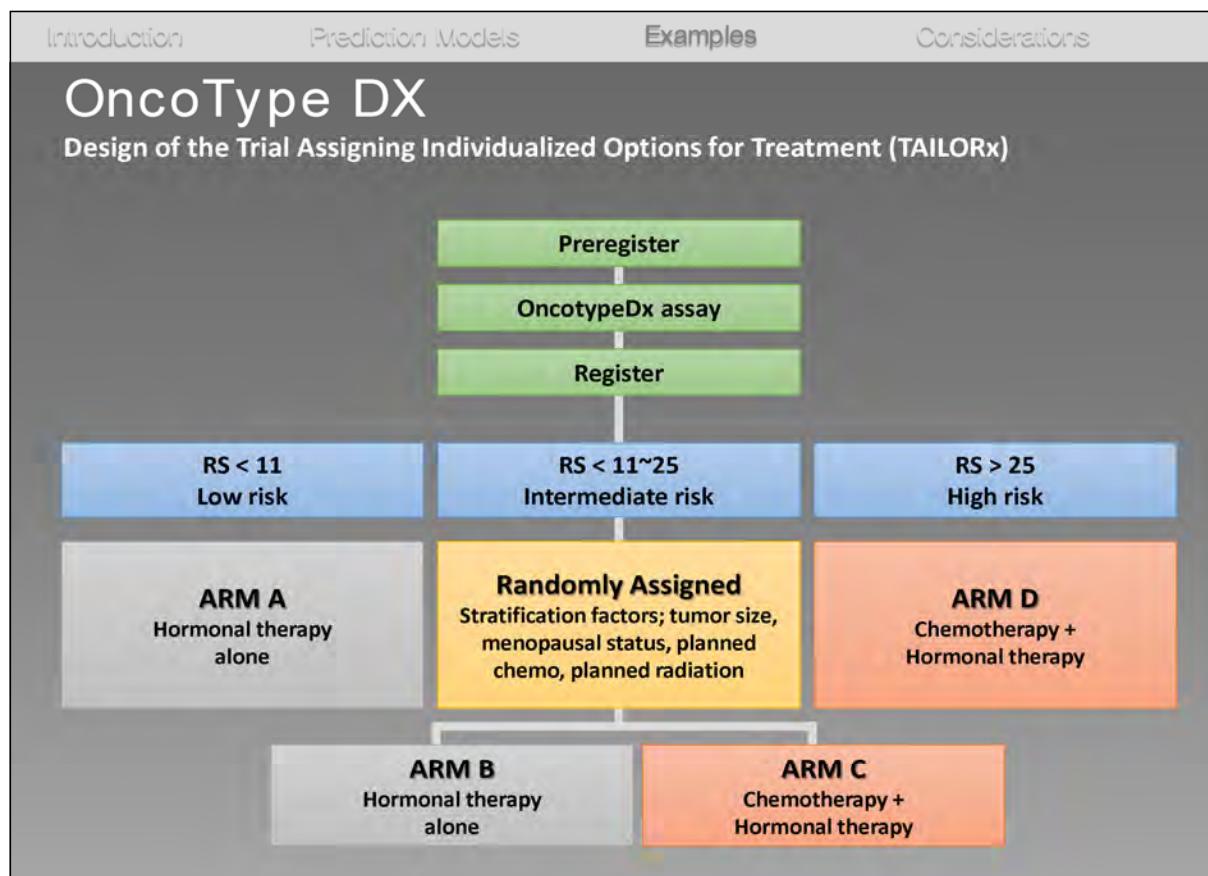
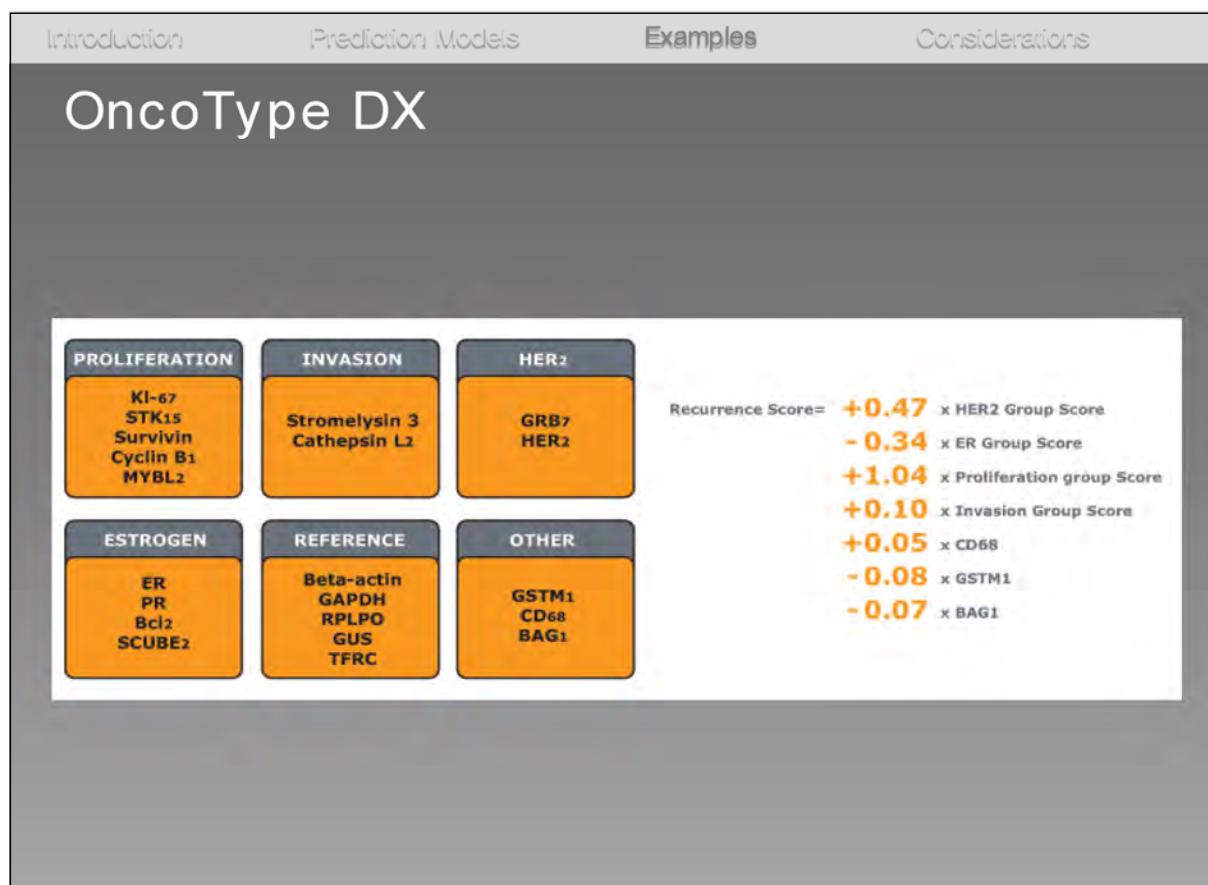
RT-PCR

Recurrence score (RS) algorithm

Low-Risk Group
Intermediate-Risk Group
High-Risk Group

Rate of Distant Recurrence at 10 Yr (% of patients)

Recurrence Score



통계 예측모형 사례

- 유방암: OncotypeDX
- 간암
- 췌장암
- IPMN Nomogram

32

Introduction

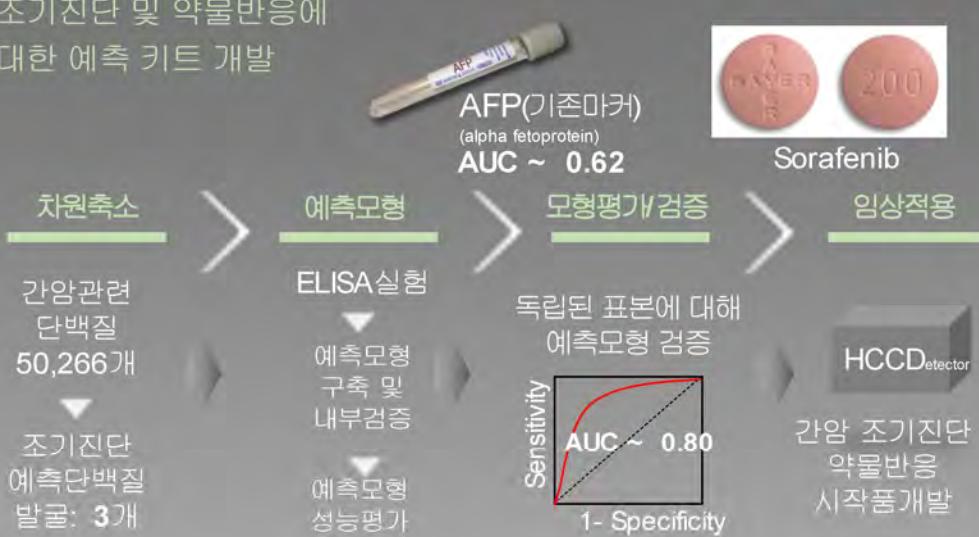
Prediction Models

Examples

Considerations

간암 조기진단/ 약물반응 예측모형 개발

- » 서울대학교병원 소화기내과, 의공학과 및 (주)진매트릭스 협력으로 간암 조기 진단/약물반응에 대한 예측모형을 구축 (산자부 지원 과제)
- » 조기진단 및 약물반응에 대한 예측 키트 개발



Introduction Prediction Models Examples Considerations

조기진단/ 약물반응 예측마커 선택

	수집 샘플수	차원축소		예측모형	임상적용
		유전체	단백질체 (MRM)		
조기진단 control	458		120	40	251
조기진단 case	1,548		60	20	251
약물반응	282	80	115	80	193

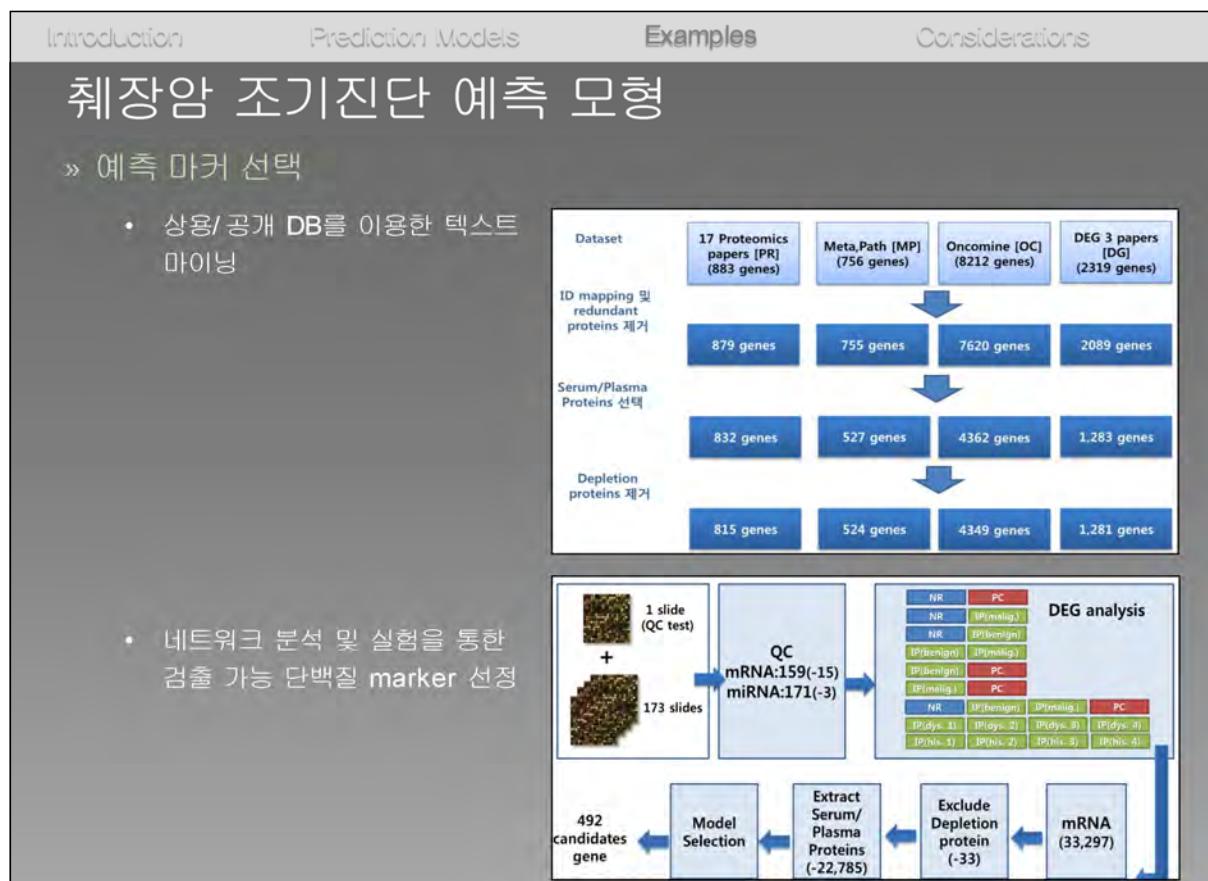
» 대규모 조기진단/약물반응 코호트를 이용한 마커 발굴

- 차원축소: 조기진단 180명, 약물반응 115명을 이용한 마커 선택
- 예측모형: ELISA 기술 이용, 총 140명을 이용한 예측모형 구축 및 모형평가/검증 수행
- 임상적용: 약 700명의 독립된 조기진단/약물반응 코호트에 기반한 예측모형 적용 및 성능평가 수행

34

통계 예측모형 사례

- 유방암: OncotypeDX
- 간암
- 췌장암
- IPMN Nomogram



Introduction Prediction Models Examples Considerations

췌장암 조기진단 예측 모형

» 모형 평가

- 대규모 코호트 ($n > 1,000$)를 이용한 다중마커 조기진단 모형 평가
- 기존 췌장암 진단 마커로 사용된 CA19-9보다 매우 향상된 결과를 검증함.

실선: 발굴한 최적 모형을 이용하였을 때의 ROC curve
점선: CA19-9을 이용한 예측 모형의 ROC curve

Introduction Prediction Models Examples Considerations

췌장암 조기진단 예측 모형

» 임상적용예정

- MRM- MS은 현재 의료용으로 사용이 불가능하고 대량 상용화가 어려움.
- 실제 임상적용을 위해서 Immunoassay에 기반한 췌장암 진단 키트를 개발함.
- Immunoassay의 재현성 확인 결과 소규모 검사 결과에서 매우 좋은 성능을 보여줌.
- 현재 대규모 코호트에 대한 임상 시험을 계획중임.

Data set of AUC	CA19- 9	MRM- MS	Immunoassay
(A) Training Set	0.844	0.941*	0.941*
(B) Test Set	0.804	0.892*	0.907*
(C) Test Set of Stage I/II	0.777	0.864	0.875*
(D) Test Set of Stage III/IV	0.861	0.951	0.975

❖ p- values < 0.05 (compared with CA19- 9, DeLong test)

Introduction Prediction Models Examples Considerations

췌장암 조기진단 예측 모형

리아프
건강의학

췌장암 완치율 높인 조기진단분석기술 개발
■ 암이 되는 건강정보

온수진 의학전문기자 | yes05@hanmail.net

2017.11.01 13:56:28

서울대 병원-SK텔레콤 공동 연구 개발
미국·중국·일본·유럽연합 등 주요국가에 특히 출원

전단 어려워 가장 위험한 암 10명 중 9명 사망
췌장암을 조기에 지금보다 더 정확히 발견할 수 있는 새로운 진단법이 국내 연구진에 의해 개발됐다. 췌장은 복부 안쪽에 위치해 있어 진단과 치료가 쉽지 않아 암 중에서도 완치율이 가장 낮은 위험한 암이다. 따라서 정확한 조기 진단이 특히 중요하다.

서울대와 서울대병원, SK텔레콤 공동연구팀(의공 학교실 김영수, 외과 장진영, 통계학과 박태성 교수, SK텔레콤 IVD 사업본부)은 질량분석기의 다중반응검지법과 호소면역 측정법을 융합하여 췌장의 새로운 진단법인 '3-마커파널'을 개발했다.

새로 개발된 진단방법은 물역시료에서 기준 췌장 암 검사법인 'CA19-9'보다 진단 정확도를 10~30% 이상 높일 수 있다. 연구팀은 검사시약을 비롯한 이 진단 시스템을 국내를 비롯해 미국, 중국, 일본과 유럽 연합 등 주요국가에 특히 출원 신청했다.



MEDICAL TRIBUNE

전체기사 뉴스 병원 행정 제약 학회 IT/의료기기 의약경제 저널서치 검색

질량분석기로 췌장암 조기진단법 개발

입력 / 2017.10.23 10:49

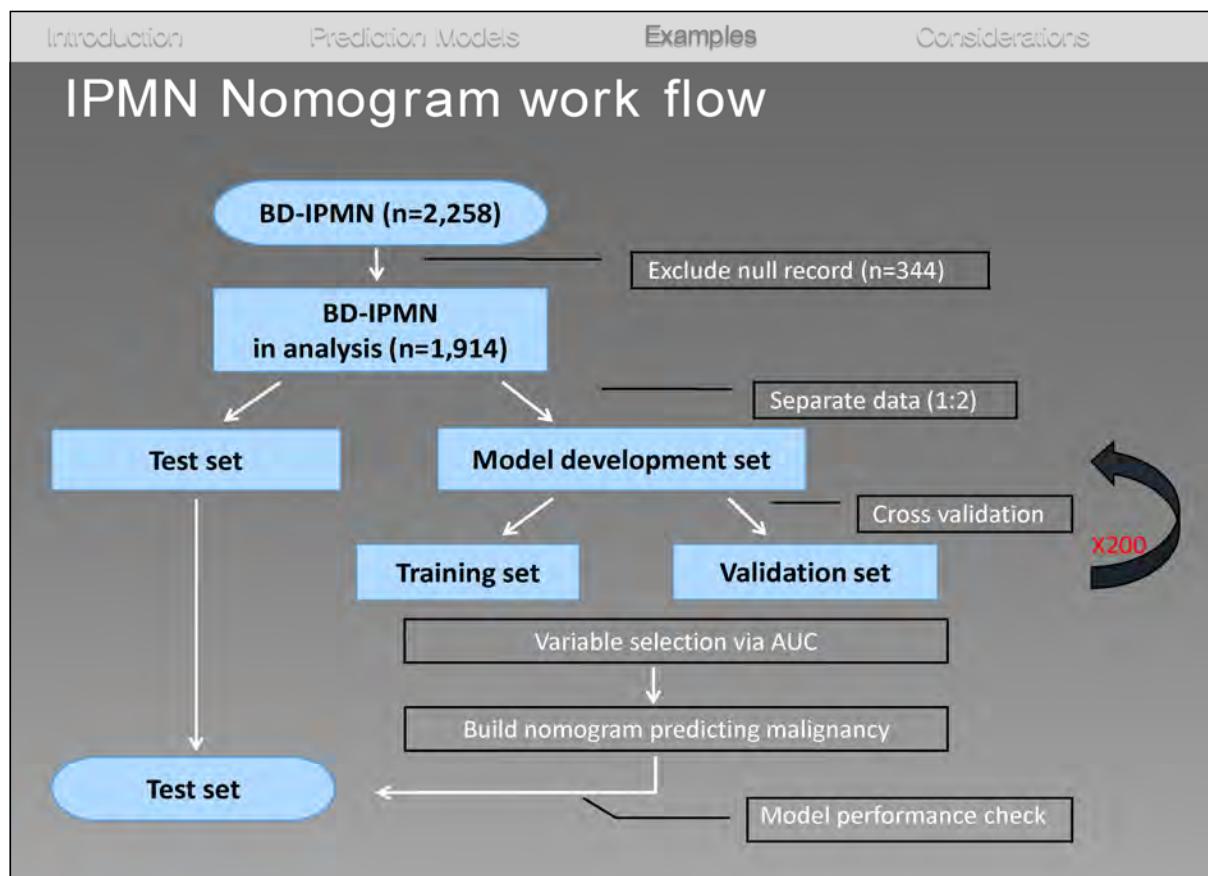
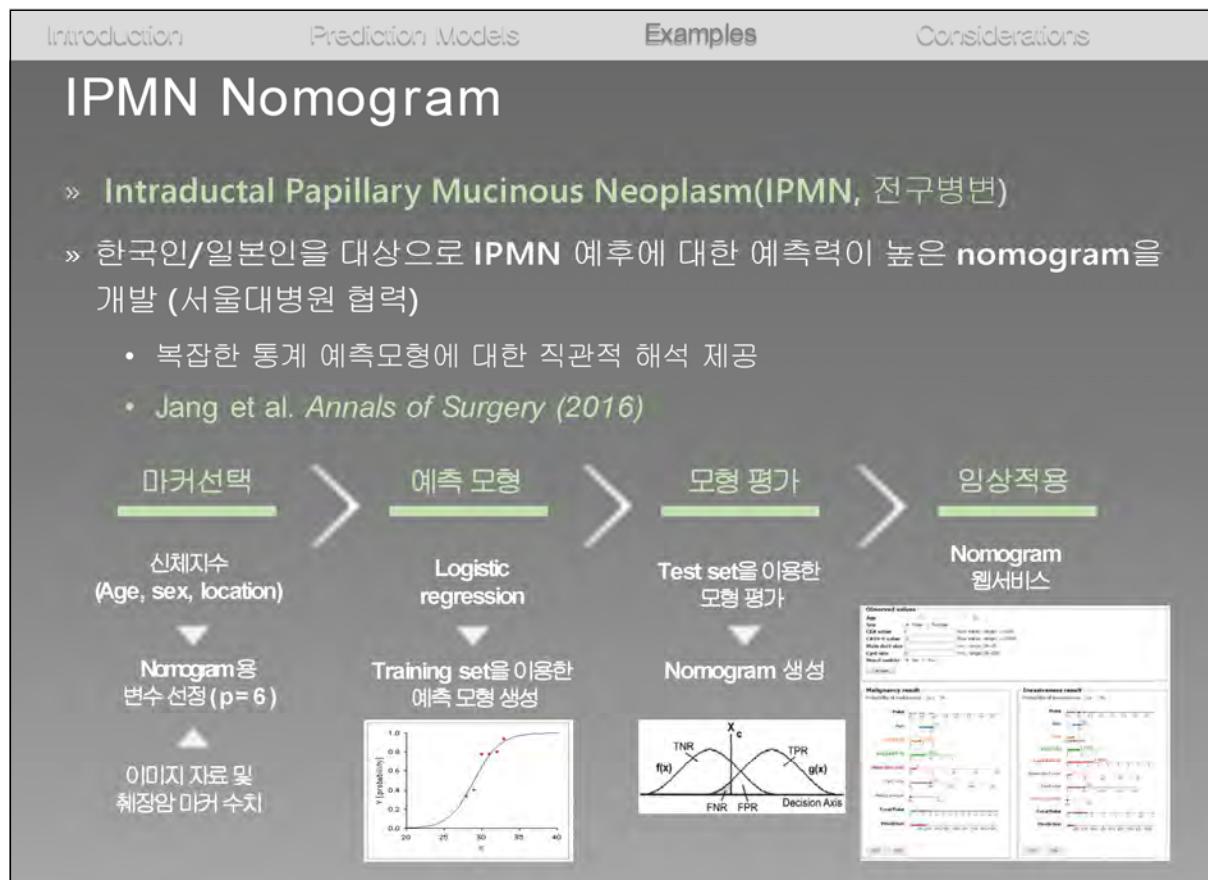
[메디칼트리뷴 김준호 기자] 질량분석기로 췌장암을 조기 진단하는 방법이 개발됐다.

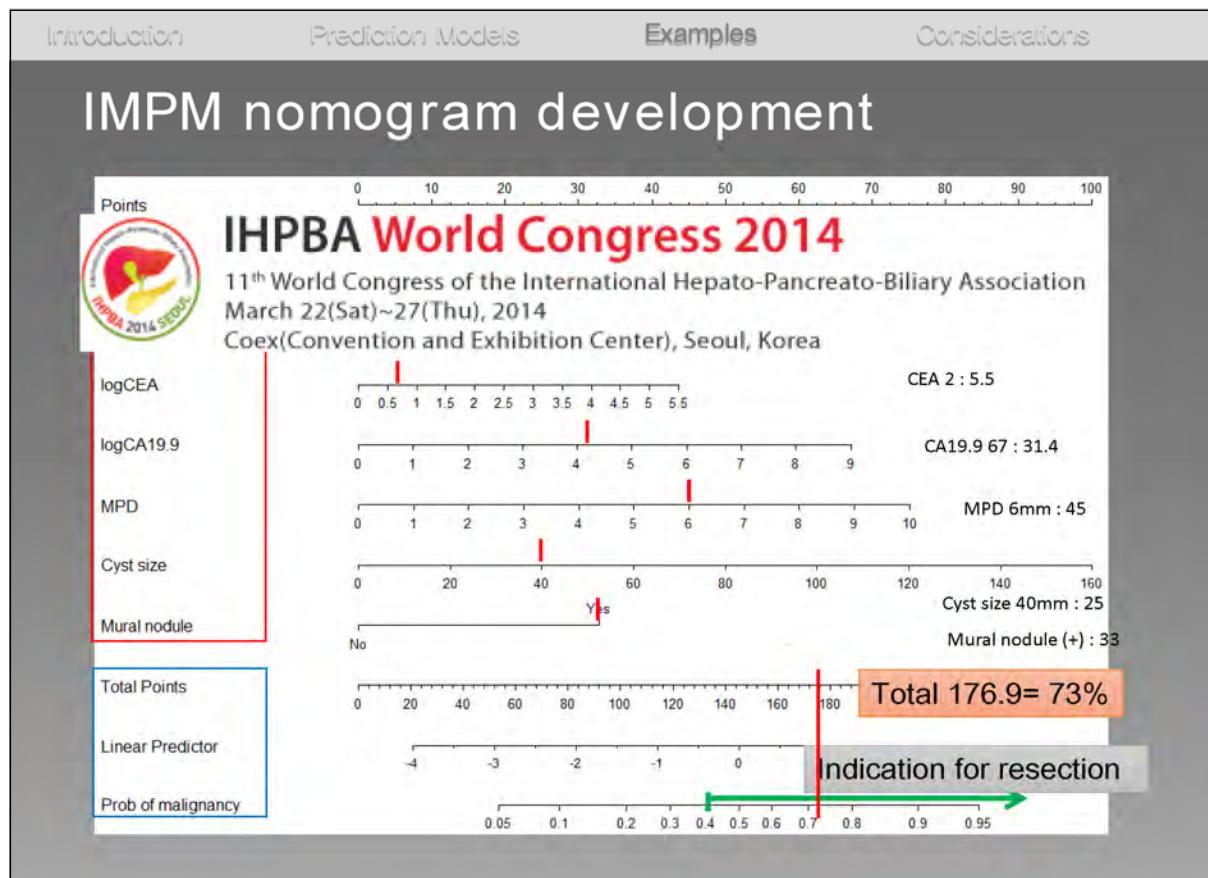
서울대·서울대병원-SK텔레콤 공동연구팀(의공학교실 김영수, 외과 장진영, 통계학과 박태성 교수, SK텔레콤 IVD 사업본부)은 질량분석기의 다중반응검지법과 호소면역 측정법을 융합해 기존 췌장암 검사법 CA19-9보다 정확한 3-마커파널을 개발했다고 미국암학술지 온코타

것에 발표했다.

통계 예측모형 사례

- 유방암: OncotypeDX
- 간암
- 췌장암
- IPMN Nomogram





Where are we now?
 Where are we heading for?

Introduction Prediction Models Examples Considerations

Precision Medicine

Precision Medicine - Paradigm Shift

The diagram shows a progression from 'Yesterday' to 'Today' to 'Tomorrow'. In 'Yesterday', the data is 'Symptom-based' and the action is 'Intuition Medicine'. In 'Today', the data is 'Cohort-based' and the action is 'Evidence-based Medicine'. In 'Tomorrow', the data is 'Algorithm-based' and the action is 'Precision Medicine'. Below this, a central text states: 'Application of rules, algorithms and reference databases enables ACTIONABLE clinical decision support & PRECISE/EFFICIENT care'. To the left, there are icons for 'Personal Molecular Profiles (PMP, multi-omics)' and 'Molecular Imaging'. To the right, there are icons for 'Big Data Analytics', 'Reference Databases (Rules & Algorithms)', 'Mobile Information Communication Technologies', and 'Precision Medicine (prevention, diagnosis, treatment)'. Arrows indicate the flow from data to actions and from data to precision medicine.

InnVentis

Yesterday Today Tomorrow

DATA

ACTIONS

Symptom-based Cohort-based Algorithm-based

Intuition Medicine Evidence-based Medicine Precision Medicine

Application of rules, algorithms and reference databases enables ACTIONABLE clinical decision support & PRECISE/EFFICIENT care

Personal Molecular Profiles (PMP, multi-omics)

Molecular Imaging

Non-molecular content (waveforms)

Big Data Analytics

Reference Databases (Rules & Algorithms)

Mobile Information Communication Technologies

Precision Medicine (prevention, diagnosis, treatment)

Precision Medicine © Dr. Thomas Whittemore, InnVentis

Introduction Prediction Models Examples Considerations

Precision Medicine

A research technician at the University of Pittsburgh Medical Center prepares to collect a blood sample from a participant in the 'All of Us' research program.

DAKE KANG / AP

Introduction Prediction Models Examples Considerations

Precision Medicine

HEALTH NEWS MARCH 7, 2015 / 10:20 AM / 4 YEARS AGO

NIH director sees solving data puzzle as key to U.S. precision medicine

Julie Steenhuysen 3 MIN READ [Twitter](#) [Facebook](#)

LA JOLLA, Calif. (Reuters) - One of the earliest tasks in President Barack Obama's \$215 million "precision medicine" program will be finding a way to piece together data from a hodgepodge of existing studies, said Dr. Francis Collins, director of the National Institutes of Health.

Introduction Prediction Models Examples Considerations

BC Platforms: Software Platform for Clinical Genomics

Primary use (healthcare)

- Clinical genomics
- Preventive medicine

Secondary use (research)

- Biobanks & Pharma industry collaboration
- Academic research

BC Data Integration and Management Platform

- Genome data
- Pedigree data
- Clinical data
- omics, biomarkers
- Annotations

Genotyping / NGS devices
Pedigree data
Hospital databases, EHR, Data lakes
Various devices, files (-omics, wearables etc)
Public & private annotation data

Introduction Prediction Models Examples Considerations

Collaborations

Our solutions are used in 22 countries at over 70 institutions.

Partners and collaborators:

- illumina*
- Microsoft
- Sentieon
- Google Cloud

Introduction Prediction Models Examples Considerations

Codigo46 The largest translational biobank in Latin America

BC|GENOME

BC|PIPE GT

BC Genome Data Warehouse

Azure Cloud Blob storage

Azure Batch for Linux

Microsoft

Azure Pack/Stack

Multiple servers

Premium storage

Introduction Prediction Models Examples Considerations

- Data vs. Analysis



- Single Omics Data Analysis



- High- Dimensional Omics Data

Data Generation < Analysis < Interpretation

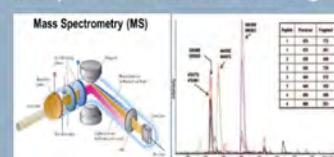
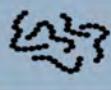
\$1,000 < \$100,000 < \$1,000,000

Courtesy of E Wijzman

Introduction Prediction Models Examples Considerations

개인의 특성을 반영한 정밀의학

바이오 빅데이터 = 다중오믹스 + EMR

Microarray Next Generation Sequencing	Mass spectrometry Multiple reaction monitoring	Anthropometry investigation Blood, Urine test ...
		
 유전체 전사체 EV 메타지놈	 단백체	 EMR 혈액, 소변검사, CT 신체계측 등

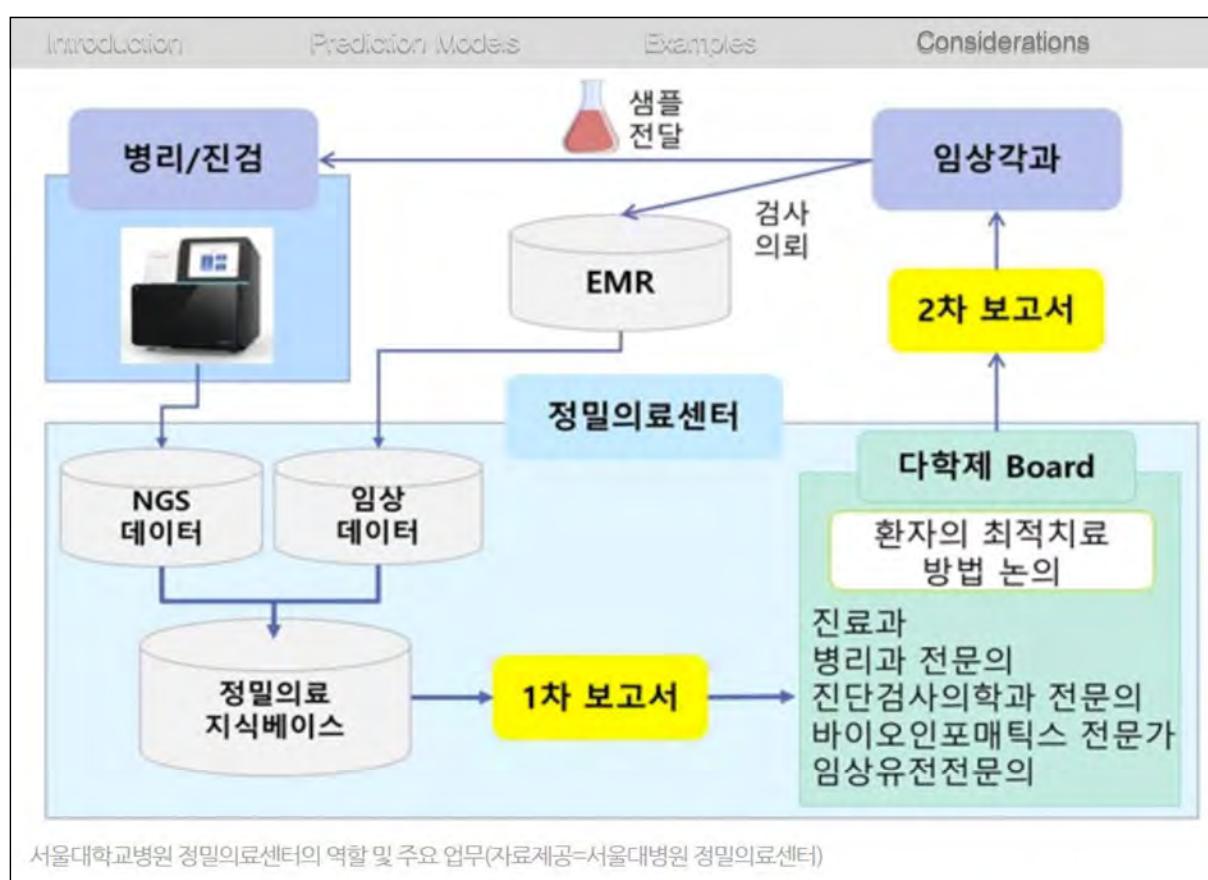
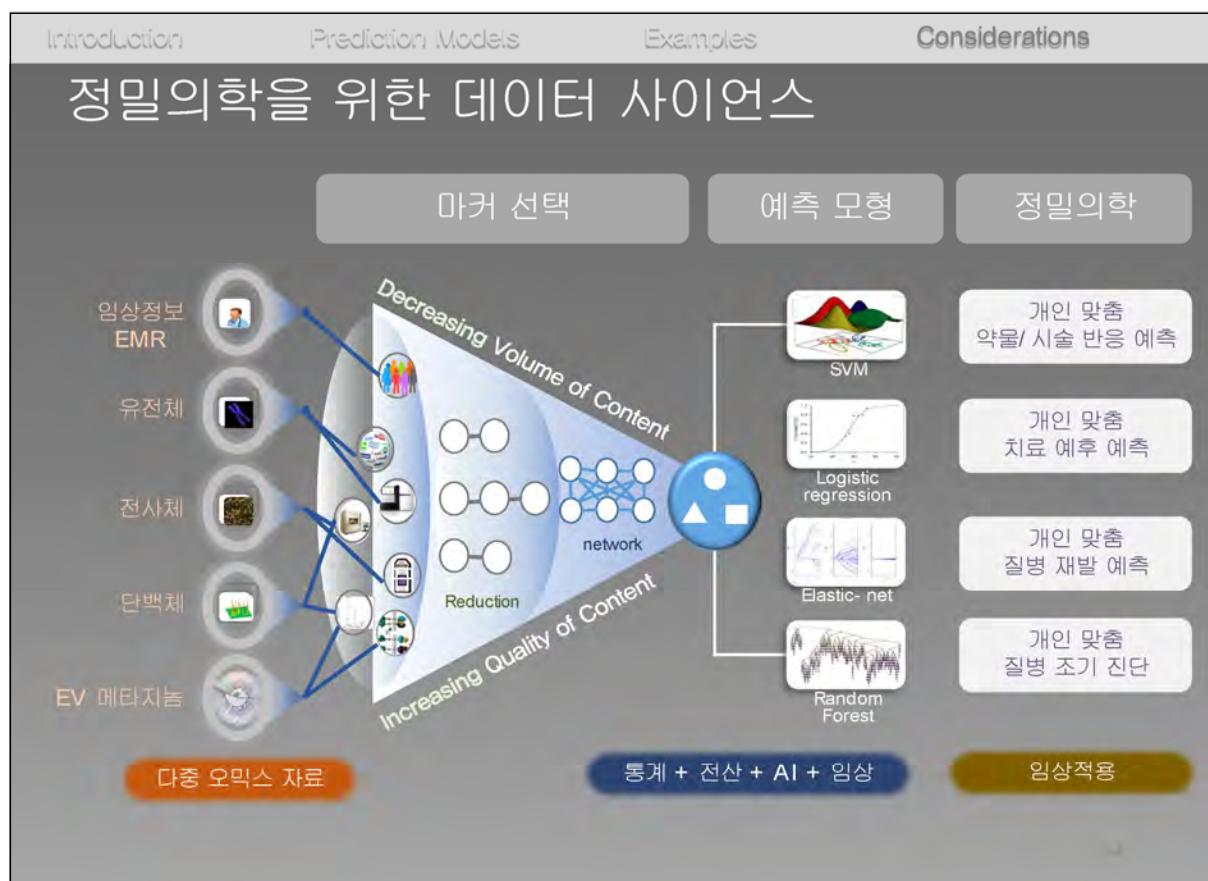
개인 특성을 반영한 의료

개인 맞춤 약물/ 시술 반응 예측

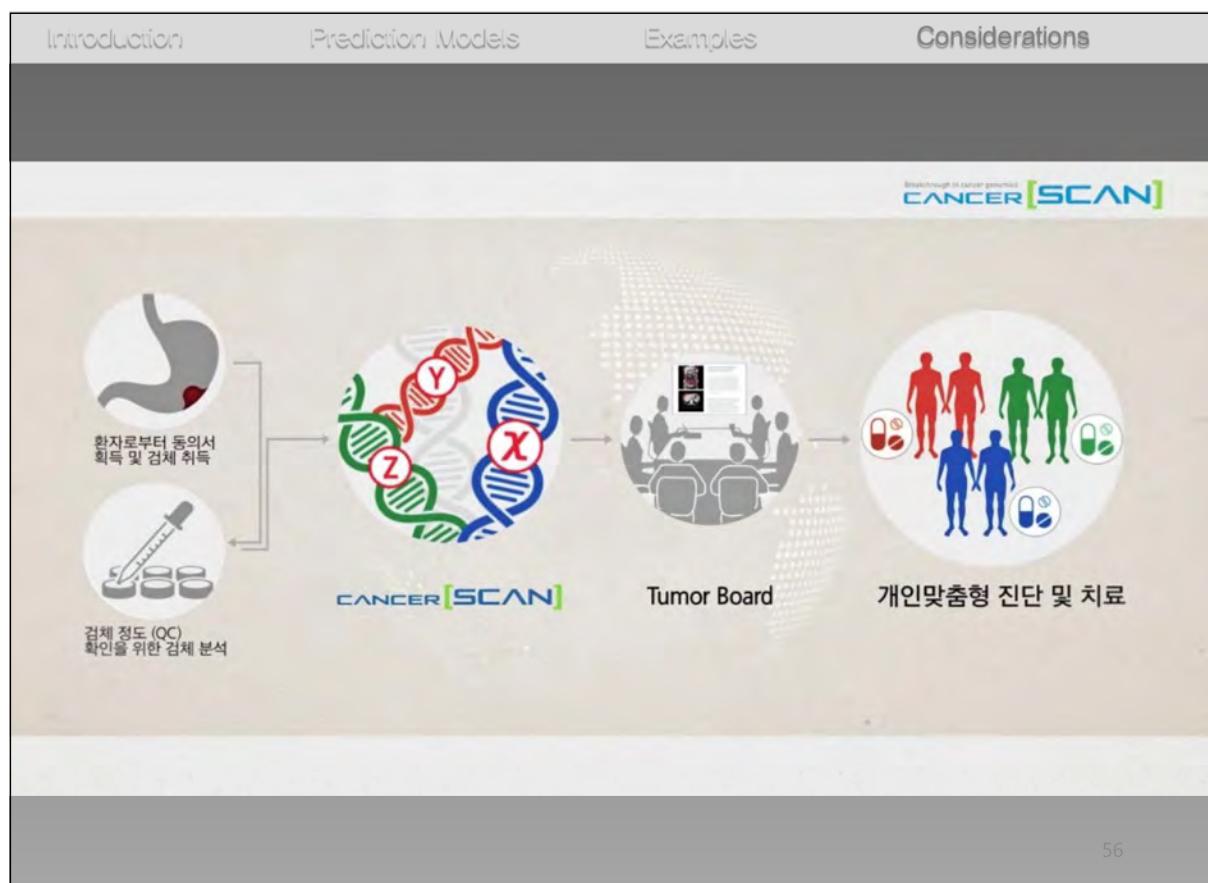
개인 맞춤 치료 예후 예측

개인 맞춤 질병 재발 예측

개인 맞춤 질병 조기 진단



서울대학교병원 정밀의료센터의 역할 및 주요 업무(자료제공=서울대병원 정밀의료센터)



56

주제발표 2

II

데이터 사이언스와 의료 민주화

발제자 약력

성명	윤형진	
소속	서울의대	
1. 학력		
기간	학교명	전공 및 학위
1980~1986	서울의대	의학과/의학사
1988~1993	서울대학교대학원	내과학전공/의학석사
1996~2001	서울대학교대학원	내과학전공/의학박사
2. 주요 경력		
기간	기관명	직위, 직책
2018~현재	서울의대 의료빅데이터연구센터	부센터장
2018~현재	의료인공지능학회추진위원회	부회장
2018~현재	대한역학회	부회장
2015~현재	서울대학교병원 의생명연구원	임상의과학정보실장
2014~2015	서울대학교병원 의생명연구원	융합연구협력부장
2010~현재	서울의대 의공학교실	교수
2007~2010	서울대학교병원	연구교수
2002~2012	(주)서울클리니칼지노믹스	부설연구소장
1997~2000	Harvard Medical School Massachusetts General Hospital	Research Fellow
1995~1997	서울의대 내과학교실	전임강사

발제2

데이터 사이언스와 의료 민주화

•••

윤형진

서울대학교병원 의생명연구원 정보실장

최근 유행하고 있는 4차산업혁명이라는 개념은 전문가들 사이에서도 명확하게 합의가 이루 어지지 않은 상태이지만, 4차산업혁명을 주도하고 있다고 거론되는 여러 첨단 기술의 성과가 실생활에 적용되어 우리의 삶을 바꿀 것이라는 전망에 대해서는 이견이 없는 것으로 보인다. 특히, 4차산업혁명의 상당 부분이 보건의료와 관련이 되어 있다는 점이 흥미롭다.

현재 선진국은 물론 우리나라에서도 보건의료의 비용은 급등하는 반면 성과는 그에 맞추어 개선되지 않고 있어서 큰 사회문제가 되고 있다. 보건의료의 저효율성에 대한 해결책으로 임상부문과 의학 연구 부문 사이에 데이터가 원활하게 흐를 수 있도록 하고 이를 통하여 보건의료 시스템의 효율을 개선하고자 하는 움직임이 시작되었는데, 이 시도의 근간에는 바로 데이터 사이언스가 있고, 4차산업혁명의 지향점과 상당부분 맥을 같이 한다고 할 수 있다. 이와 함께 지난 수십 년 간 집중적으로 투자를 하였던 의료기관에서 행해지는 의료나 유전체로 대별되는 개인의 생물학적 특성보다도 오히려 건강에 더 많은 영향을 미치는 것으로 알려져 있는 개인의 생활 습관이나 환경 등을 개선하는 것에 관심을 갖고 투자하여야 한다는 지적이 설득력을 얻고 있다.

한편으로, 의료기관의 의료 정보를 포함하여 자신의 건강과 관련된 데이터 일체를 자신이 저장하고 관리하면서 자신의 건강관리에 활용하고자 하는 소위 ‘quantified self’ 운동도 이미 오래전에 시작된 바 있다. ‘Quantified self’ 운동은 개인 단위로 건강과 직간접적으로 관련된 데이터 일체를 수집하는 것이 필요하며 이는 의료기관 단위로 의료 정보를 수집하고 관리하던 전통적인 의료 시스템의 파괴적 변혁을 수반하게 될 것으로 보인다. 혹자는 성직자를 통하여만 접하였던 성경을 인쇄술의 발명으로 일반 신도들이 직접 접할 수 있게 되고 결국 종교혁명으로 이어졌던 것에 비교하기도 하고, 전통적인 가부장적인 의료 시스템에 대비하여 ‘의료 민주화’라고 말하기도 한다. ‘의료 민주화’의 개념은 아직 명확하지 않으나, 건강 정보에 대한 의료 소비자의 자기결정권과 의사와 환자간의 ‘shared decision’이 핵심이라고 할 수 있다.

그러나 이러한 시도는 처음이 아니고 그리 성공적이지 못하였다. 그 원인을 여러 가지로 말할 수 있으나, 가장 중요한 것은 사용자가 자신의 건강과 관련된 데이터를 모두 수집하고 관리할 때 기대할 수 있는 혜택이 명확하지 않았던 것이 아닌가 생각된다. 자신의 데이터를 스스로

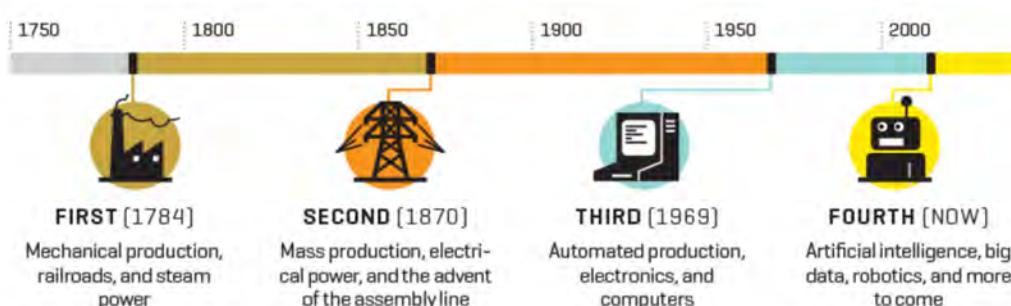
관리한다고 하여도 사용자가 자신의 데이터를 이해하지도 못하고 더 좋은 방법으로 활용할 길도 없다면 굳이 그러한 수고를 들일 필요가 없기 때문이다. 성경을 새로 발명된 인쇄술로 대량으로 쌍 값에 인쇄한다고 해도 일반인들이 이해할 수 없는 라틴어로 성경이 쓰여져 있다면 무용지물이나 마찬가지라고 할 수 있다. 종교혁명은 대량으로 인쇄된 라틴어 성경이 아니라 이를 각국의 언어로 번역한 번역본에 의하여 촉발되었다. 즉 지금 진행되고 있는 ‘의료 민주화’가 우리 생활을 변화시키고 더 나아가 우리의 건강을 개선하고 사회전체적인 보건의료 비용을 절감하기 위해서는 자신의 건강 관련 데이터를 일반인들이 이해할 수 있도록 번역해주는 작업이 필요하며, 여기에 데이터 사이언스가 핵심적인 역할을 하여야 한다. 따라서 ‘의료 민주화’의 성패는 데이터 사이언스에 달려 있고, 보건의료 데이터 사이언스가 앞으로 추구해야하는 방향 중의 하나가 일반인이 이해할 수 있는 의료 정보 제공이라고 할 수 있다.

데이터 사이언스와 의료 민주화

서울대학교 의과대학
윤형진

4차 산업혁명

THE FOUR INDUSTRIAL REVOLUTIONS



(<http://fortune.com/2016/03/08/davos-new-industrial-revolution/>)

Drivers of the 4th Industrial Revolution

- Leveraging the pervasive power of digitization and information technology

Physical	Digital	Biological
Autonomous vehicles	Internet of things	Genome sequencing
3D printing	Sensors	Synthetic biology
Advanced robotics	Remote monitoring	Personalized medicine
New materials	Blockchain	Gene editing
	Digital platforms	Bioprinting
		Neurotechnology

(Klaus Schwab, 2016, The Fourth Industrial Revolution)

Tipping points expected to occur by 2025

Table 1: Tipping points expected to occur by 2025

	%	
10% of people wearing clothes connected to the internet	91.2	90% of the population using smartphones
90% of people having unlimited and free (advertising-supported) storage	91.0	90% of the population with regular access to the internet
1 trillion sensors connected to the internet	89.2	Driverless cars equaling 10% of all cars on US roads
The first robotic pharmacist in the US	86.5	The first transplant of a 3D-printed liver
10% of reading glasses connected to the internet	85.5	30% of corporate audits performed by AI
80% of people with a digital presence on the internet	84.4	Tax collected for the first time by a government via a blockchain
The first 3D-printed car in production	84.1	Over 50% of internet traffic to homes for appliances and devices
The first government to replace its census with big-data sources	82.9	Globally more trips/journeys via car sharing than in private cars
The first implantable mobile phone available commercially	81.7	The first city with more than 50,000 people and no traffic lights
5% of consumer products printed in 3D	81.1	10% of global gross domestic product stored on blockchain technology
		The first AI machine on a corporate board of directors

Gaps in Pt Outcome and Costs of Care

Table 1. Crude and Predicted Mortality Rates in Study Cohorts according to Level of Medicare Spending in Hospital Referral Region of Residence*

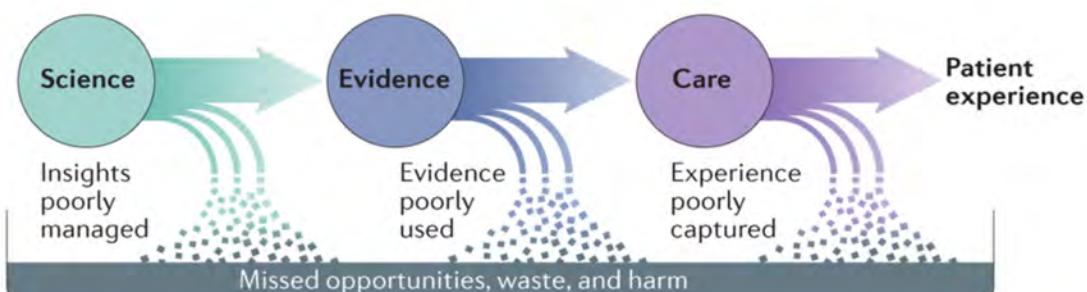
Variable	Quintile of EOL-EI					Test for Trend†
	1 (Lowest)	2	3	4	5 (Highest)	
←—————%—————→						
Hip fracture cohort						
Observed 30-day mortality rate	7.8	7.2	6.9	6.9	6.6	↓
Observed 1-year mortality rate	24.4	23.9	23.9	24.3	24.2	>0.05
Predicted 1-year mortality rate	24.5	24.1	24.1	24.1	23.9	↓
Colorectal cancer cohort						
Observed 30-day mortality rate	4.5	4.6	4.8	4.8	4.4	>0.05
Observed 1-year mortality rate	20.6	20.7	21.7	21.1	20.9	>0.05
Predicted 1-year mortality rate	21.1	20.8	21.2	20.8	20.9	>0.05
Acute MI cohort						
Observed 30-day mortality rate	18.5	18.4	19.2	18.2	18.5	>0.05
Observed 1-year mortality rate	30.7	31.3	32.6	31.6	33.3	↑
Predicted 1-year mortality rate	31.2	31.5	31.8	32.0	33.2	↑
MCBS cohort						
Observed 30-day mortality rate	0.2	0.4	0.4	0.3	0.2	>0.05
Observed 1-year mortality rate	4.6	4.8	5.0	5.1	5.3	>0.05
Predicted 1-year mortality rate	4.9	5.1	5.3	5.0	5.1	>0.05

* Crude mortality rates were based on 30-day and 1-year follow-up for all cohort members with no censoring (follow-up for mortality was complete at 1 year for all). Predicted mortality rates were based on logistic regression equations that included individual- and ZIP code-level variables only. EOL-EI = End-of-Life Expenditure Index; MCBS = Medicare Current Beneficiary Survey; MI = myocardial infarction.

† Arrows show the direction of any statistically significant association ($P \leq 0.05$) between the mortality rate and regional EOL-EI differences. An arrow pointing upward indicates that as spending increases across regions, the mortality rate increases. A P value greater than 0.05 was considered not significant.

(Fisher ES, 2003)

현재의 헬스케어 -2012 US Institute of Medicine 보고서



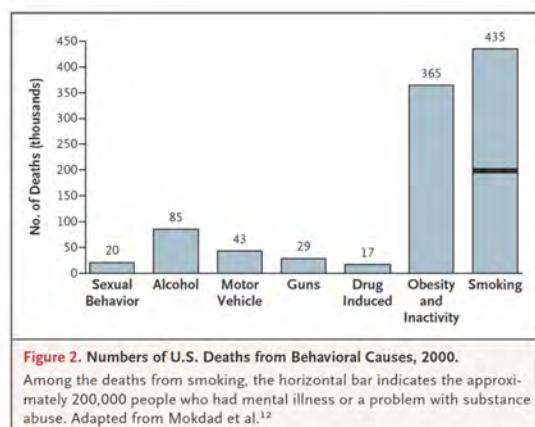
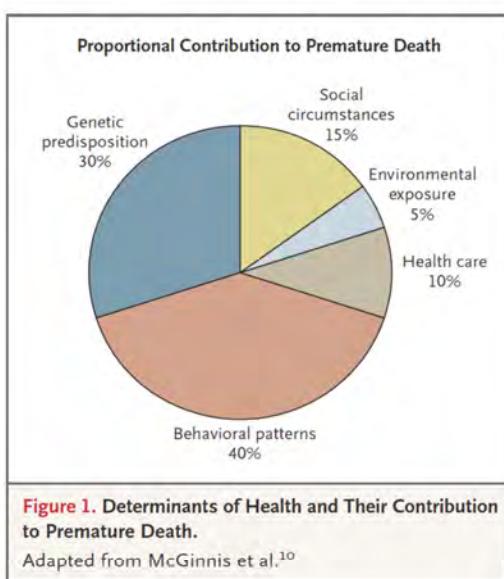
Nature Reviews | Cardiology

(Rumsfeld JS; Nat Rev Cardiol, 2016)

빅데이터 시대의 헬스케어

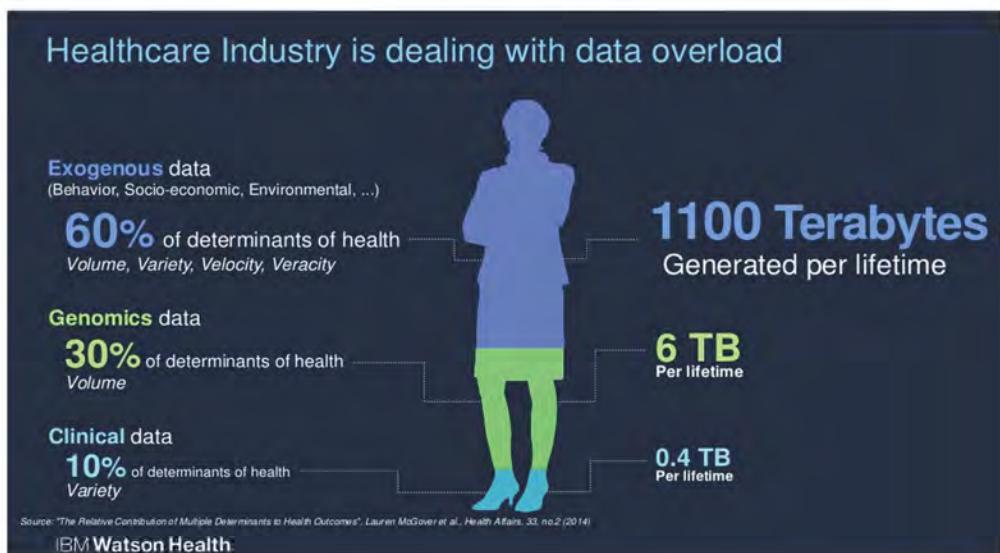
- 의료의 질 향상과 치료 효율 제고에 대한 요구
- 분석 가능한 데이터의 증가
- 데이터 분석 기술의 발전

생활습관과 건강



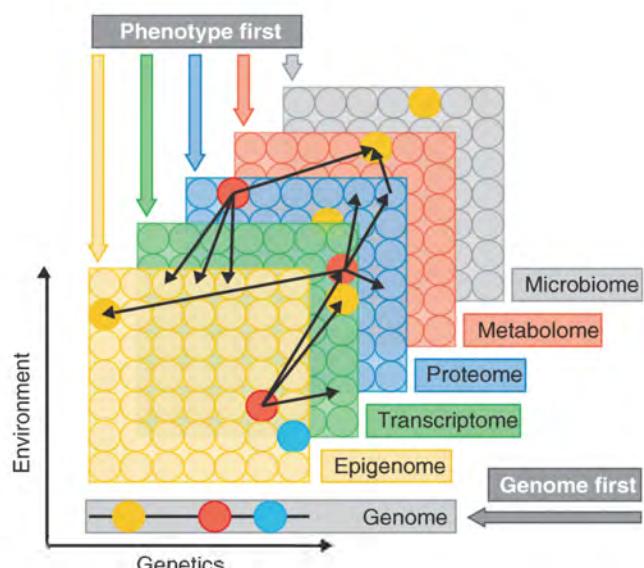
(Schroeder SA, N Engl J Med 2007)

보건의료 분야의 데이터 생산



(<http://www.slideshare.net/levshapiro/mhealth-israelibm-watson-for-healthcare-startups>)

Multi-omics Approaches to Disease



(Hasin Y et al; Genome Biol, 2017)

Activity Trackers & Smart Phones



(<https://www.gymrat-fitness.com/wp-content/uploads/2016/12/THE-BEST-Activity-Tracker.jpg>;
<https://www.arnnet.com.au/article/630841/samsung-dominates-global-smartphone-market/>)

Precision Medicine Initiative

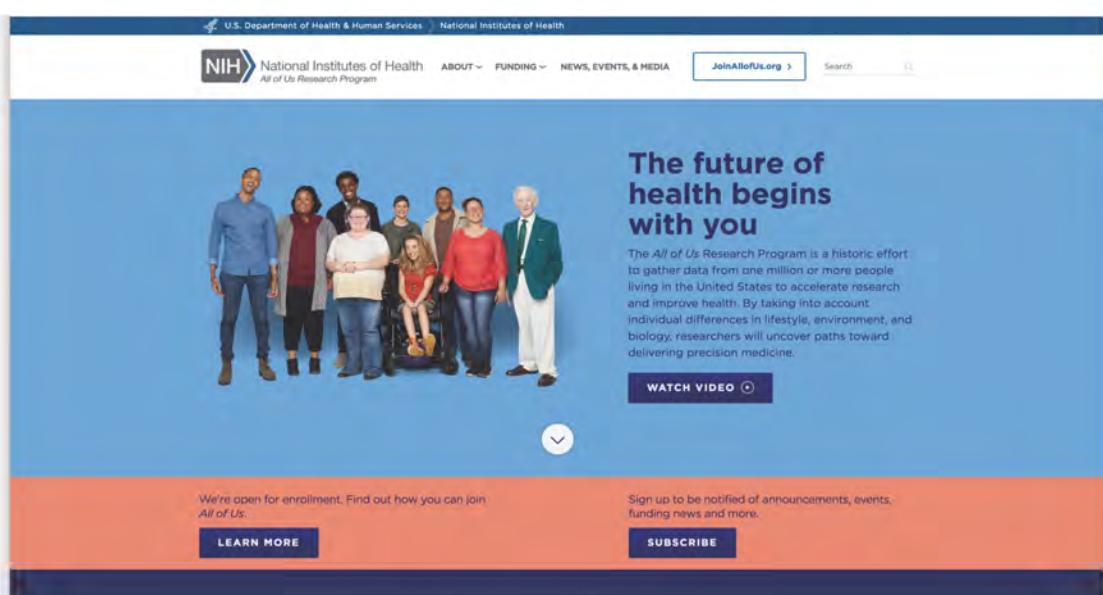


(<http://www.wsj.com/articles/obama-to-lay-out-215-million-precision-medicine-plan-1422615602>)

Precision Medicine Initiative

- \$215M in 2016
- \$130M for Research Cohort
 - 미국인 100만명 이상
 - 의무기록
 - 대상자의 유전자, 단백체, 대사체, 미생물 등 분석
 - 환경과 생활습관
 - 대상자 생성 데이터
 - 개인 디지털 장비 및 웨어러블 센서 데이터

All of Us Research Program



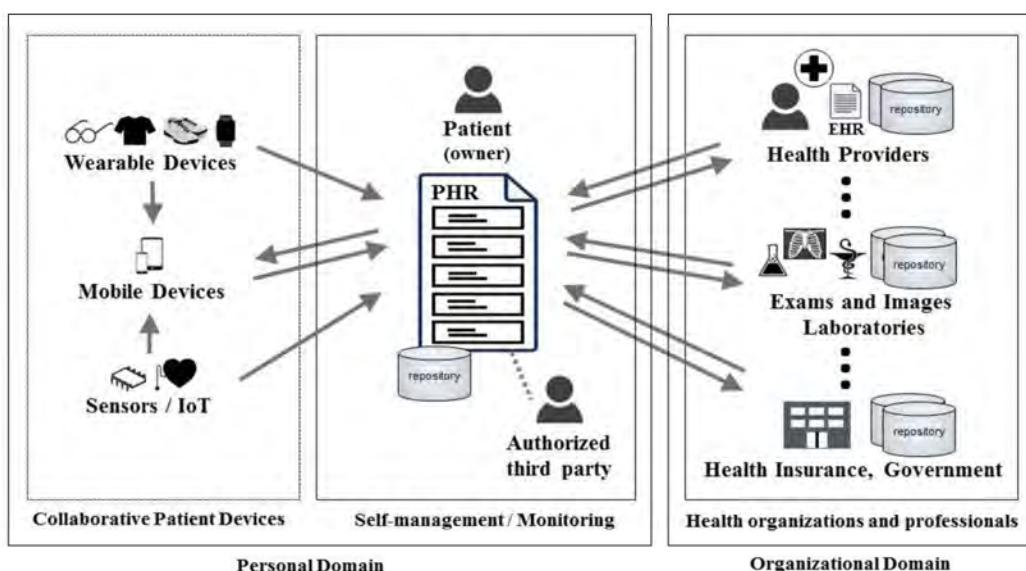
(<https://allofus.nih.gov/>)

Quantified Self



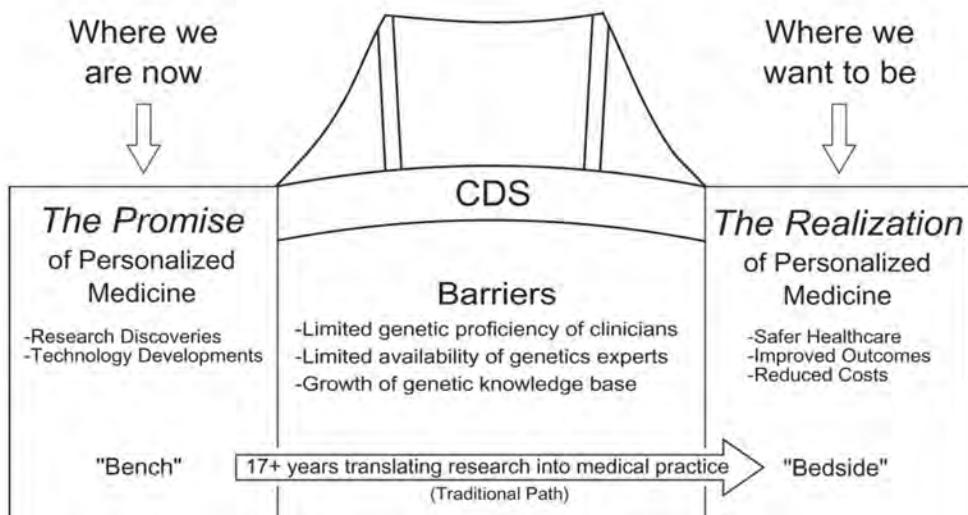
(<https://venngage.com/gallery/post/quantify-your-health-with-the-quantified-self-movement-2/>)
https://robsymestheoutsideview.files.wordpress.com/2014/03/small_sparkrpt.png)

Personal Health Records



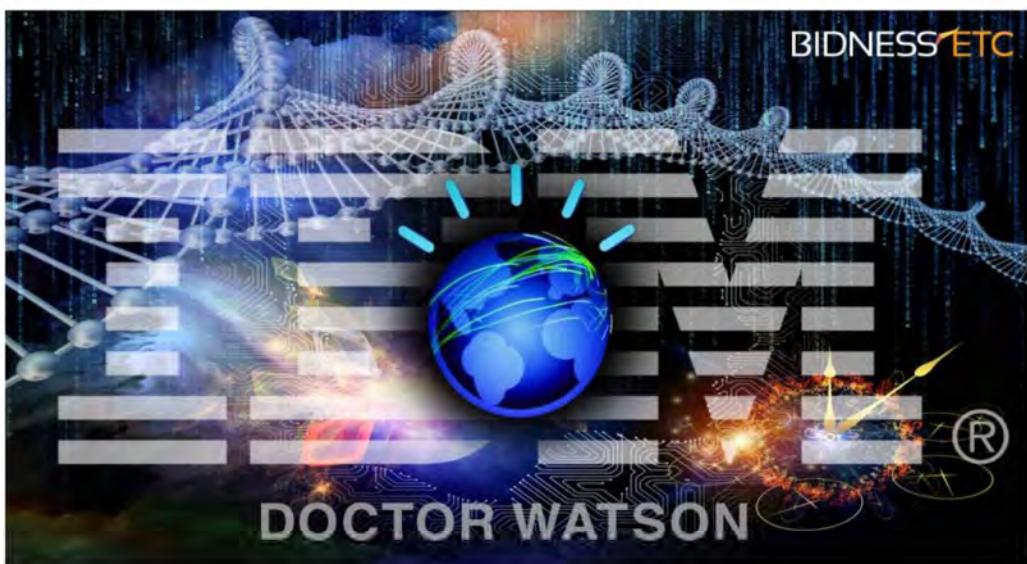
(Roehrs A et al, J Med Internet Res 2017)

임상의사결정지원시스템



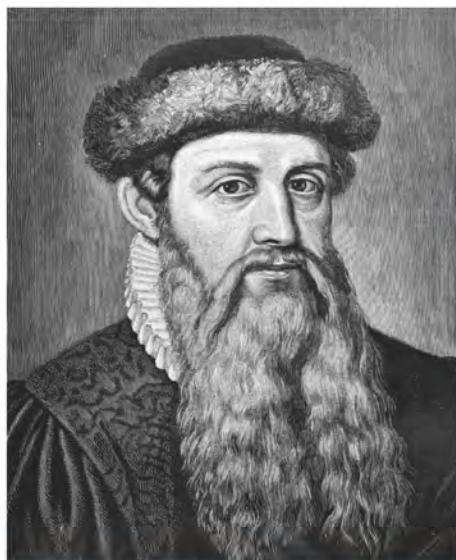
(Welch BM & Kawmoto K, J Am Med Inform Assoc, 20:383, 2013)

빅데이터, 인공지능, 헬스케어



(<http://www.bidnessetc.com/business/ibm-to-further-develop-watson-healthcare-its-primary-target/>)

Gutenberg Bible and the Protestant Reformation



https://en.wikipedia.org/wiki/Johannes_Gutenberg#/media/File:Gutenberg.jpg
https://ko.wikipedia.org/wiki/%EA%B5%AC%ED%85%90%EB%B2%A0%EB%A5%B4%ED%81%AC_%EC%84%B1%EA%B2%BD

Democratization of Health Care

- A transition that will be driven by continued consumerization, data sharing and most notably, the growth of disruptive technologies
(<https://channels.theinnovationenterprise.com/articles/democratising-healthcare-the-next-medical-revolution>)
 - Wearables, IoT, AI
- Democratization of health information
 - Hospital or Caregiver → Patient or Person

Traditional vs Democratized Health Care

Traditional Health Care

- Asymmetry of medical information
- Paternalism
- Delegating ownership of data utilization
- Caregiver-centric
- Fragmented medical information
- Medical data only

Democratized Health Care

- Sharing of medical information
- Shared decision making
- Retaining ownership of data utilization
- Consumer-centric
- Integrated medical information
- Wholistic approach

Shared Decision Making

Opinion | OP-ED CONTRIBUTOR

My Medical Choice

By ANGELINA JOLIE MAY 14, 2013



LOS ANGELES

MY MOTHER fought cancer for almost a decade and died at 56. She held out long enough to meet the first of her grandchildren and to hold them in her arms. But my other children will never have the chance to know her and experience how loving and gracious she was.

We often speak of "Mommy's mommy," and I find myself trying to explain the illness that took her away from us. They

RELATED COVERAGE



Angelina Jolie's Disclosure Highlights a Breast Cancer Dilemma MAY 14, 2013

Opinion | Letters
Angelina Jolie's Preventive Surgery
MAY 14, 2013

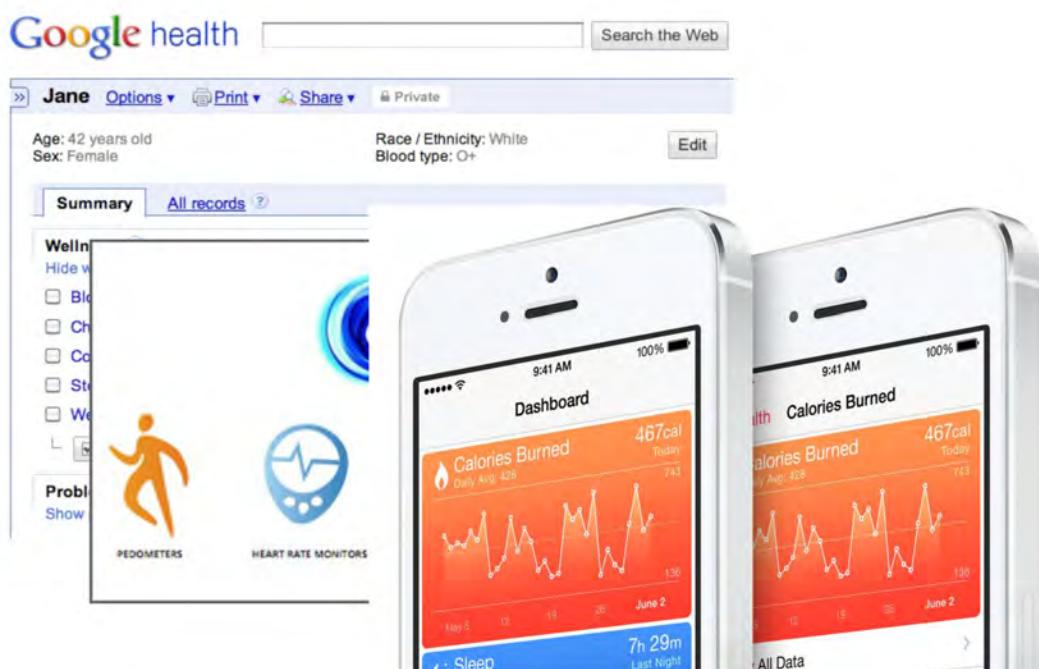
(<https://www.nytimes.com/2013/05/14/opinion/my-medical-choice.html>)

Myriad Genetics



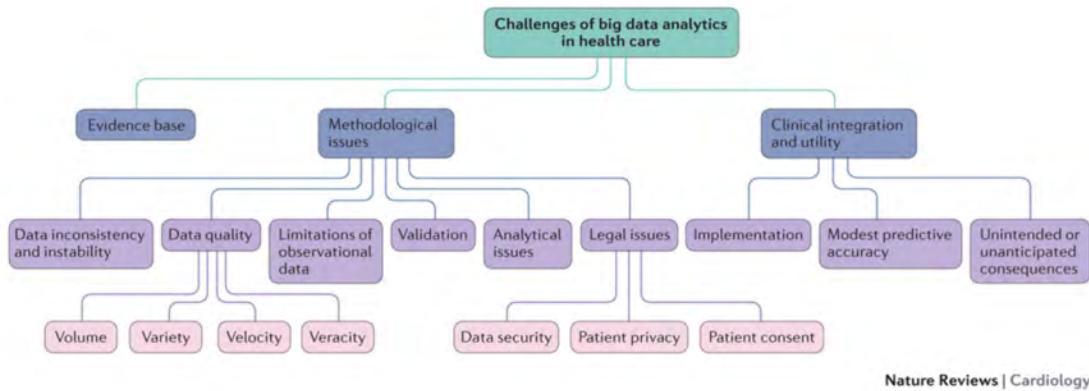
(<https://myriad.com/>)

(<http://www.synbiowatch.org/2013/06/supreme-court-strikes-down-gene-patents-in-breast-cancer-gene-case/>)

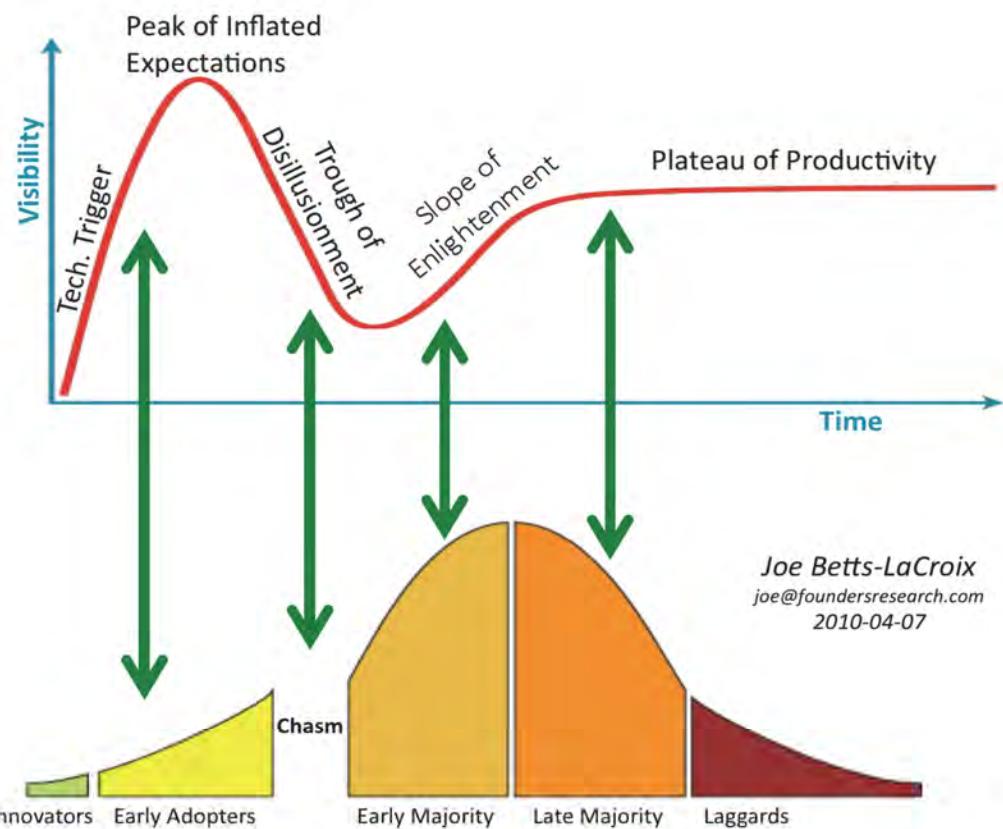


(<https://www.mobihealthnews.com/8904/google-health-redesigns-adds-mobile-app-partners>;
<https://manipalblog.com/keep-track-of-your-health-online-the-easy-way-with-microsoft-healthvault/>;
<https://www.forbes.com/sites/gordonkelly/2014/08/21/iphone-6-doctors-raise-concerns-over-apple-healthkit/#3f0a5e053469>)

Challenges of Data Science in Health Care



(Rumsfeld JS; Nat Rev Cardiol, 13:350, 2016)



Health Literacy

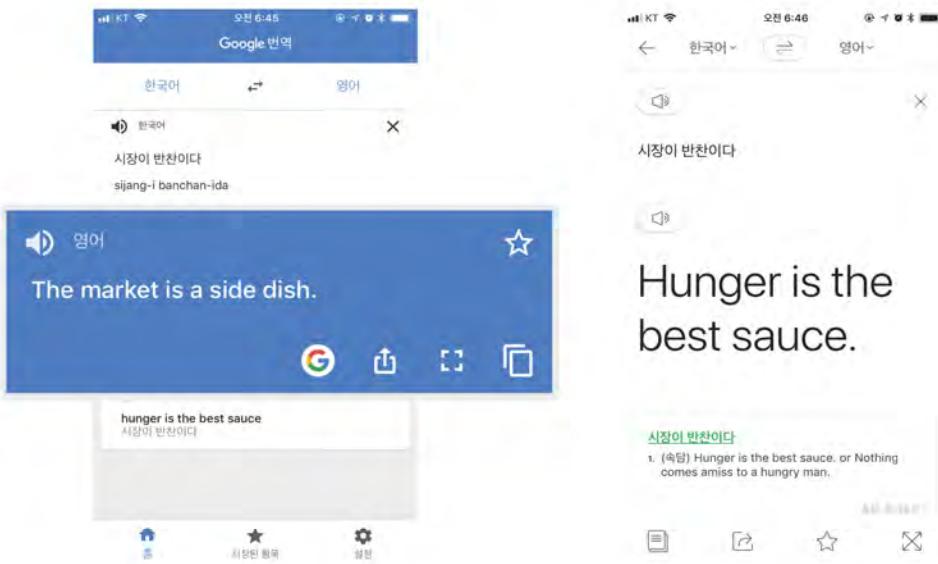


(<https://www.safetyandquality.gov.au/publications/health-literacy-infographics/>)

Data Science and Democratization of Health Care

- Data science for health literacy of laypeople
 - Visual EMR base-on AI & big data

Google Translate vs Papago



주제발표 3

III

브레인 이미징(Brain Imaging)과 데이터 사이언스

발제자 약력

성명	이동수	
소속	서울대학교	
1. 학력		
기간	학교명	전공 및 학위
1976~1982	서울대학교	의학사
1983~1990	서울대학교	의학박사
2. 주요 경력		
기간	기관명	직위, 직책
2016~현재	International Society for Nanomedicine	Vice President
2016~현재	World Federation of Nuclear Medicine and Biology	President-elect
2015~현재	Nuclear Medicine and Molecular Imaging	Editor-in-Chief
2009~현재	서울대학교 융합과학기술대학원 분자의학바이오제약학과	교수
2006~현재	과학기술한림원	정회원
1990~현재	서울대학교/서울대학교병원 핵의학과	임상교수-기금교수-교수

발제3

브레인 이미징(Brain Imaging)과 데이터 사이언스

•••

이동수

한림원 의약학부 정회원 (서울대학교)

현재 임상에 응용되는 뇌영상은 CT, MRI, PET, SPECT인데, 주로 구조와 대사변화를 조사하여 외상, 종양과 혈관질환(CT, MRI, SPECT)과 치매, 파킨슨병, 뇌전증(PET)을 조사하는 쓰인다. 뇌발달장애, 정신질환을 비롯한 뇌기능 이상 장애는 아직 이 뇌영상 도구에 이상소견이 나타나지 않아, 현재의 단순 해석만으로는 진단, 예후예측, 질병경과를 바꾸는 약제(disease-modifying drug)의 효과 평가에 사용되지 않는다. 유전체 데이터가 일부 종양연관유전자를 찾아 암진단보조와 예후예측에 겨우 쓰이고, 전장분석 유전체 결과 대부분 또는 어렵게 얻은 단일세포의 전사체 분석 결과는 아직 어떻게 다루어야 임상에 필요한 자료로 바꿀 수 있을지 몰라 버려두고 있는 것과 같다.

뇌영상을 분석하는 기술은 지난 30여년 동안, 특히 통계파라미터매핑(SPM)과 합리적 다중비교(FDR) 분석법이 확립되면서 객관화, 정량화한 신뢰성 높은 결과를 생산하게 되어, 임상에서도 이 기술로 뇌 형태/대사의 분포 즉 토포그라피를 비교하여, 질병군과 정상인의 차이나, 질병 환자 개인의 질병증증도, 향후 진전예측에 쓰고 있다. 그러나 대부분의 정신질환은 이 토포그라피 분석에도 아무런 이상소견이 나타나지 않아, 지난 10여 년 동안 뇌 각 부분의 연결성을 분석하면 이상 소견을 찾을 수 있을 거라는 예측이 공감을 얻었으며, 최근 분석방법과 통계처리에 크고 작은 진전이 있었다.

뇌 지역 연결성 분석에 필요한 토대를 위상수학으로부터 얻어, 우리(이혜경, 이동수, 2012~)는 뇌영상으로 질병군의 브레인그래프를 제작하는데 문제였던 연결강도 역치 설정의 임의성을 제거할 수 있었고, 이렇게 생산한 그래프를 순열통계 방법으로 처리하여 질병군의 특성을 밝혔다. 이제는 이 통계방법을 고도화하고 개별 환자의 연결성을 해석하기 위해 PET/휴식상태뇌기능MRI(PET/rs-fMRI) 영상데이터를 분석하고 있다. 여기 사용되는 위상수학적데이터해석(Topological data analysis)은 오믹스 데이터를 해석하는 데에도 최근 사용되었다.

뇌영상과 오믹스 결과는 대표적 빅데이터로서, 분석의 수학적프레임과 비교의 통계프레임을 개발하고 세련되게 하여 맞춤적용을 적절히 함으로써, 임상에 이용할 수 있게 된다. 수학과 통계의 토대는 SRC 데이터과학과 지식창출연구센터(2012-2018)와 ERC 산업수학연구센터(2017-현재)에서 집중적으로 다루었거나 다루고 있다. 적절한 수학과 통계 프레임을 적용하여 해석할 뇌영상 빅데이터는 개별 환자에 대한 해석이 가능해져서 결국 임상에 적용하려면 여러 가지 난제를 풀어야 하는데, 그 중 어떤 부분은 기계학습의 도움을 받을 수 있다. 딥러닝으로 진화한 인공지능/기계학습은 결국 missing data imputation이 가능하고, 희소(scanty)데이터로부터 추론하는 문제가 풀리고, 정상데이터로만 학습하여 비정상을 정의하게 되고, 이어 어느 부분이 비정상인지 찾는(XAI; explainable AI) 문제를 다루는데 까지 발전하였다. 우리(최홍윤, 이동수)는 최근 생성모델, 특히 적대적생성모델로 이런 작업을 수행할 수 있음을 보고하였다. 또 우리(이재성, 최홍윤, 이동수)는 표준적 합성곱신경망(CNN)을 이용하여 PET-CT에서 CT 없이도 PET 영상을 얻고, MRI 없이도 표준판을 만들고 이를 이용하여 여러 종류의 PET을 정량할 수 있으며, PET과 함께 촬영한 저화질 CT를 초고화질 CT로 바꾸고, 수백 명 PET 데이터로 50대부터 90대까지 뇌포도당대사의 노화진전을 빈틈없이(missing data imputation) 묘사하는 등 이 부분 도약 발전의 토대를 최근 제공하였다.

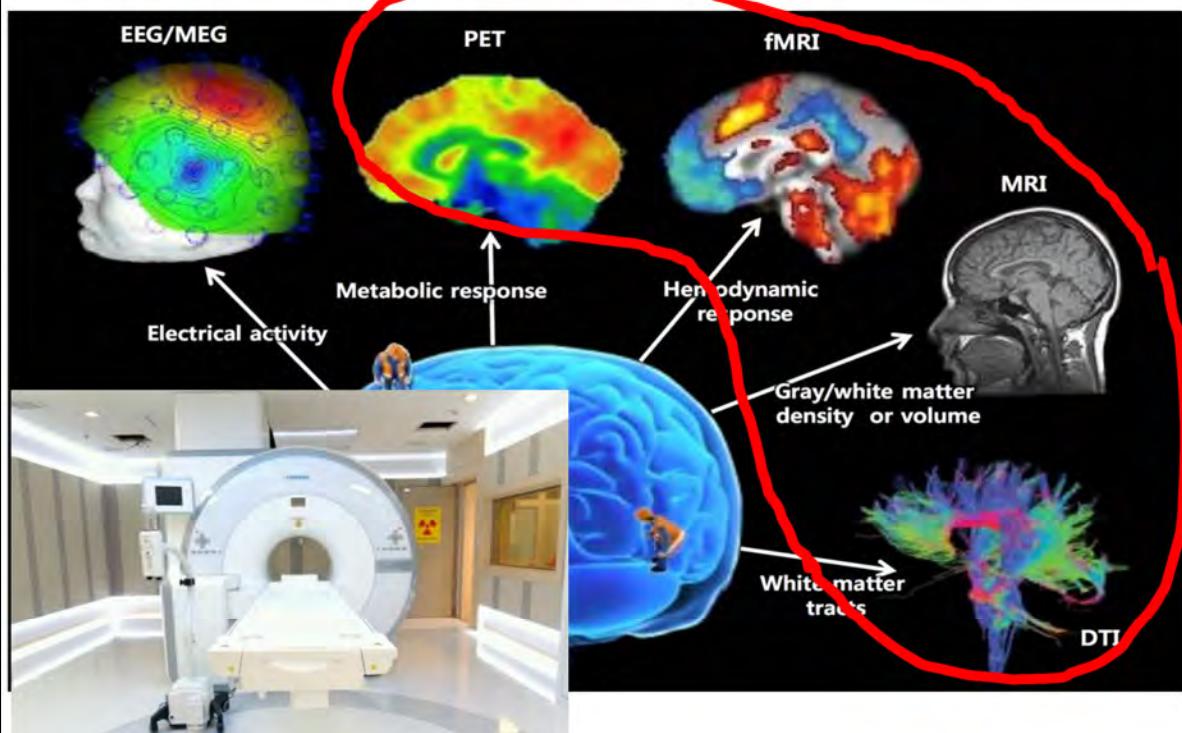
딥러닝을 적절히 순차하여 응용하여 보니, 의료 데이터를 간략한 표지(Label)로 만드는 작업부터, 오믹스 빅데이터로 예후를 쉽게 예측하게 되어서, 적절한 수학/통계 원리를 이용하여 뇌 영상 토포그라피/연결성을 분석하되 객관화/정량화를 쉽게 하여 결국 임상에 널리 이용 가능하게 될 것으로 예상한다. 그런데 이를 더 빨리 구현하려면, 오믹스와 뇌영상 부문의 개방데이터가 필요하고, 여기 더하여 각 의료기관/병원에서 생산한 맞춤데이터와 연계하여 분석하도 가능하여야 하고, 이를 위한 인적/물적 플랫폼을 구성하도록 지원할 필요가 있고, 무엇보다도 사람의 지능으로는 종잡을 수 없고, 거대하기만 한 데이터를 보는 관점을 제공할 수학과 통계의 틀을 혁신하는 시도가 긴요하다. 이 시도를 가능하게 할 연구개발 지원 구조, 병원(산업체)과 연구자의 연구통합, 팀의 국제경쟁력을 확보하도록 격려하고, 확보한 경우 지속적으로 리더십을 발휘할 수 있도록 지원하는 정책을 수립하여야 할 때다.

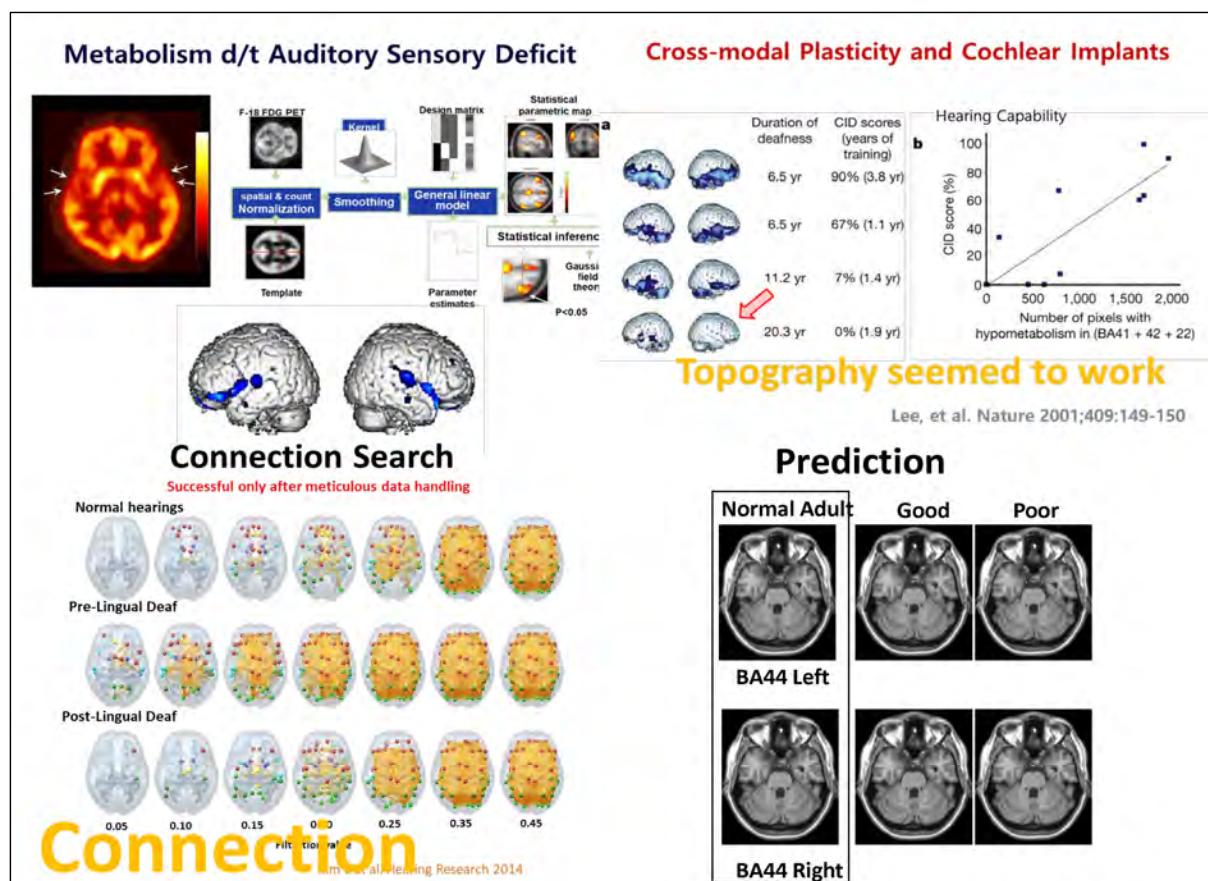
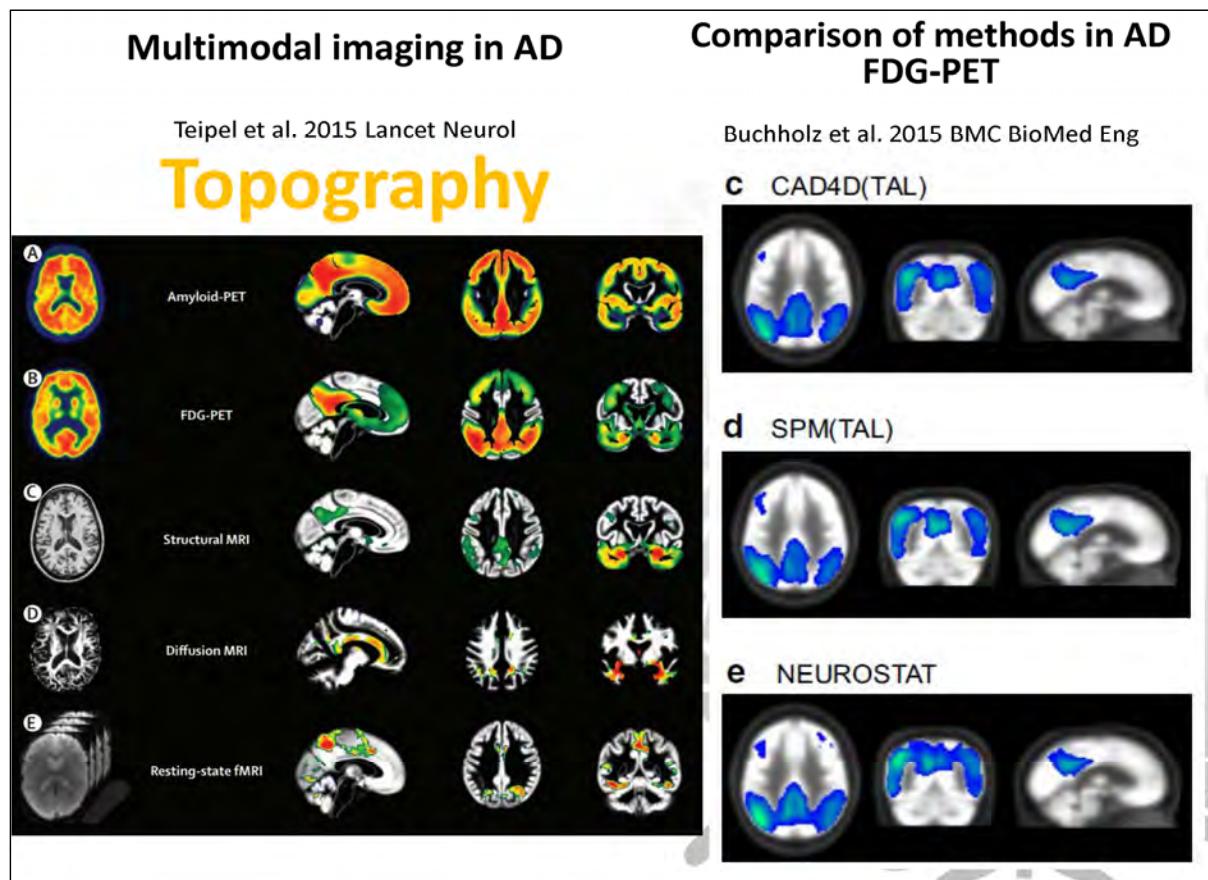
브레인 이미징에서 데이터사이언스로

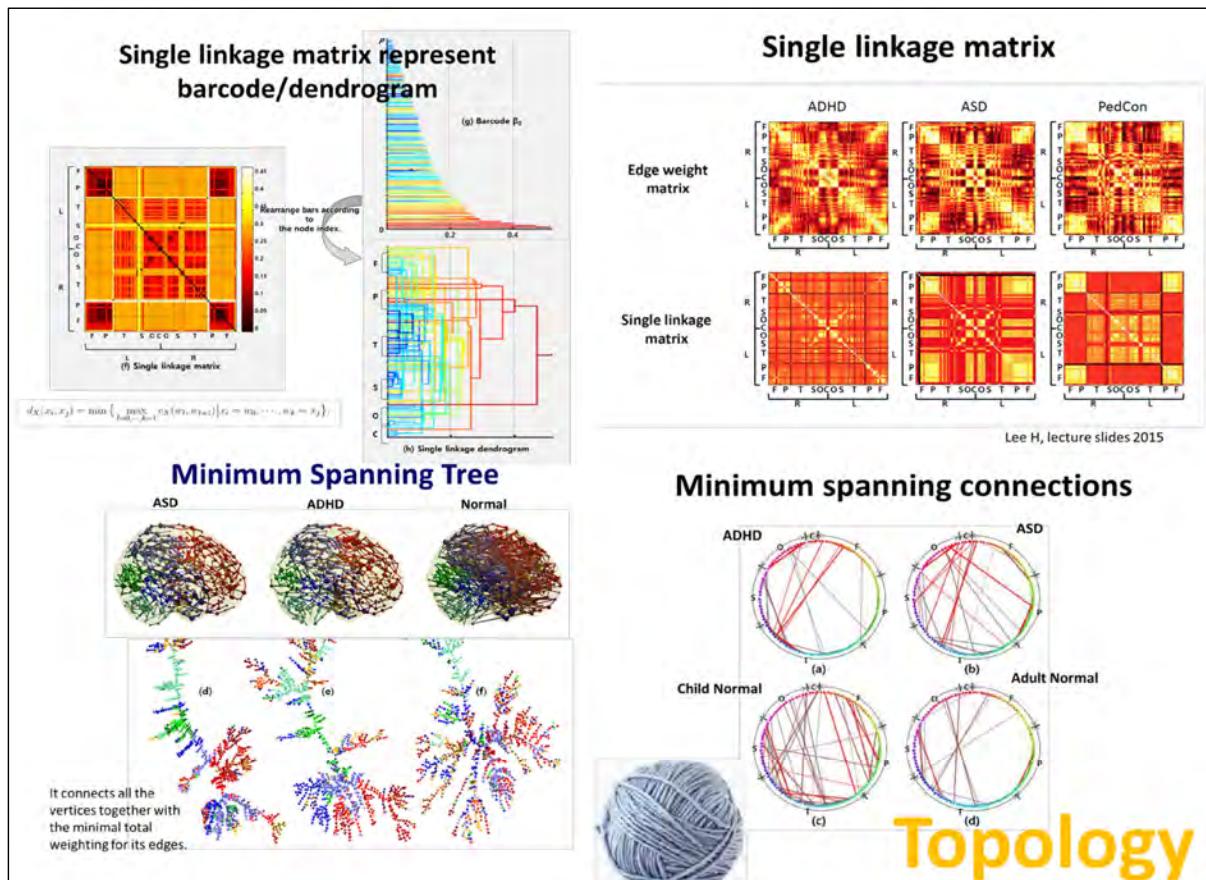
서울대학교 이동수



Brain Imaging

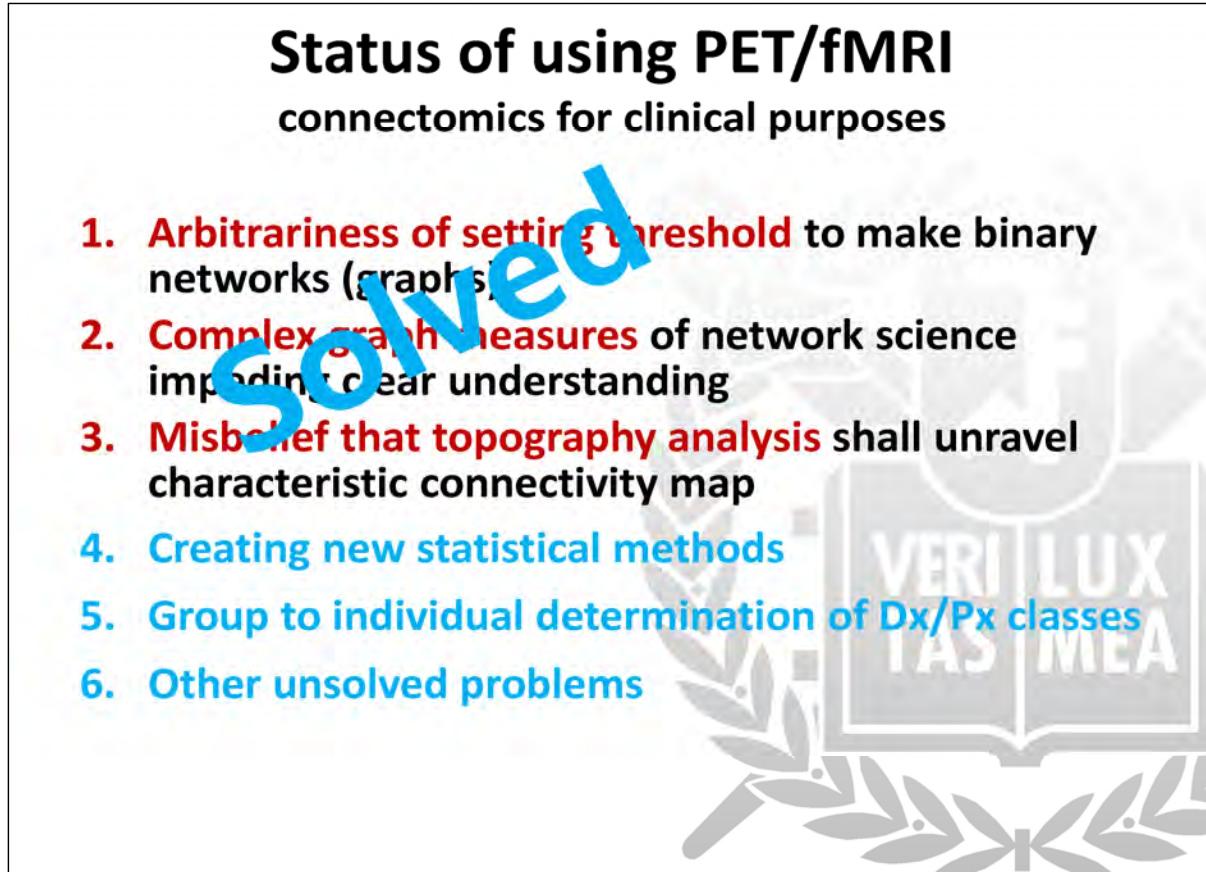






Status of using PET/fMRI connectomics for clinical purposes

1. Arbitrariness of setting threshold to make binary networks (graphs)
2. Complex graph measures of network science impeding clear understanding
3. Misbelief that topography analysis shall unravel characteristic connectivity map
4. Creating new statistical methods
5. Group to individual determination of Dx/Px classes
6. Other unsolved problems



Successful Application of Topological Approach (Persistent Homology)

- PET 2012 IEEE TMI
- T1 MRI 2014 Hearing Research
- Activation fMRI 2015 Brain Connectivity
- Rat PET 2015 Neuroimage
- Mouse PET 2016 Sci Rep
- PET MRI 2017 Hum Brain Mapping
- MEG 2017 Sci Rep
- Resting state fMRI: ongoing



Public Database

Deep Learning using Conv. NN, VAE or GAN etc.

ADHD-200
ADHD

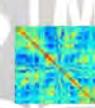
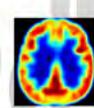
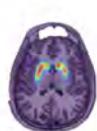
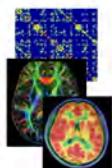
MCIC
Schizophrenia

ABIDE
Autism spectrum disorder

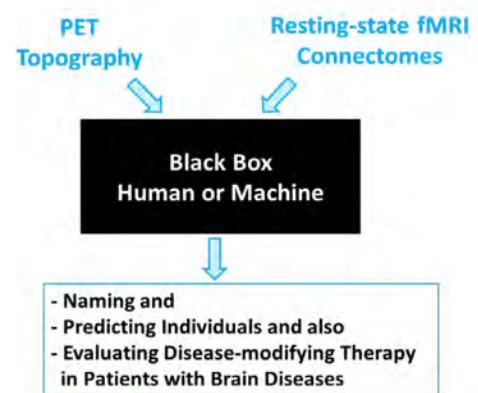
HCP
Normal fMRI

ADNI
Alzheimer's disease

PPMI
Parkinson's disease



Single Individual
Individual Analysis



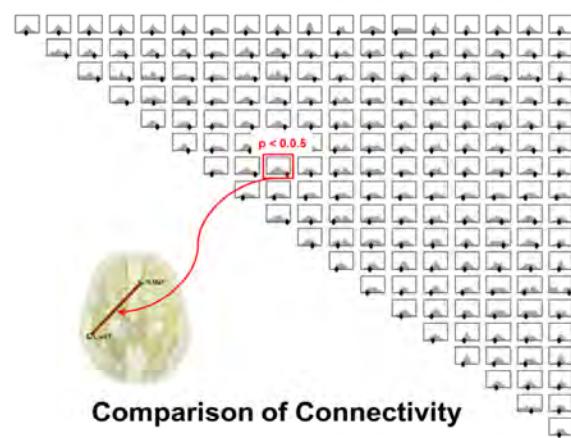
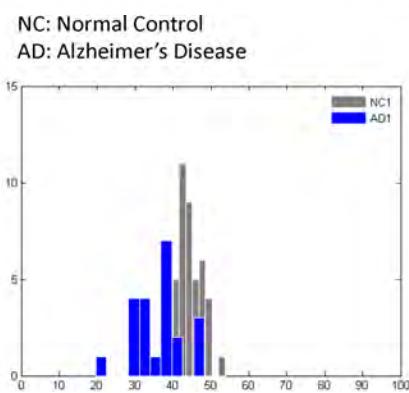
Brain Imaging 데이터를 다루는 데 중요한 두 가지

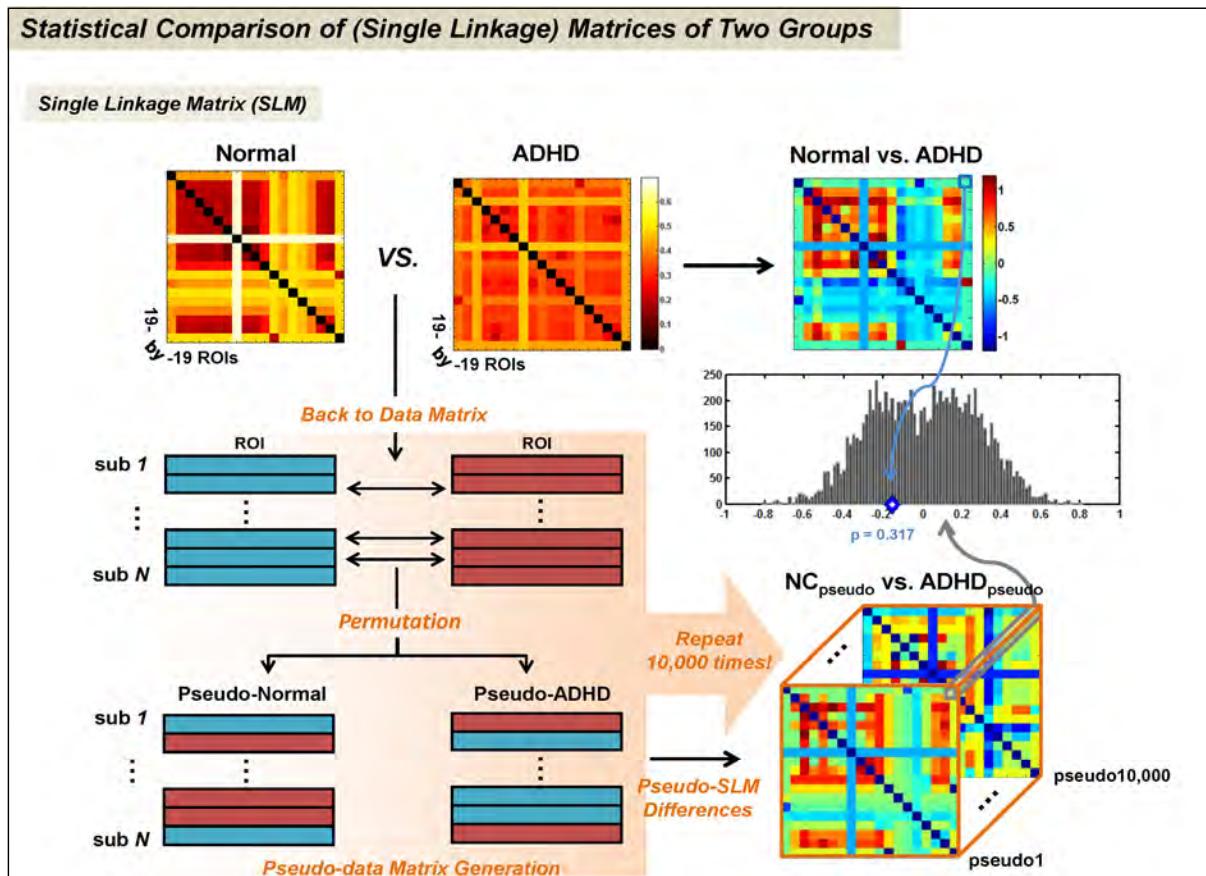
통계와 위상수학데이터해석

Creating new statistical methods

Multiple comparison: control FDR

huge number of voxels and dozens of subjects



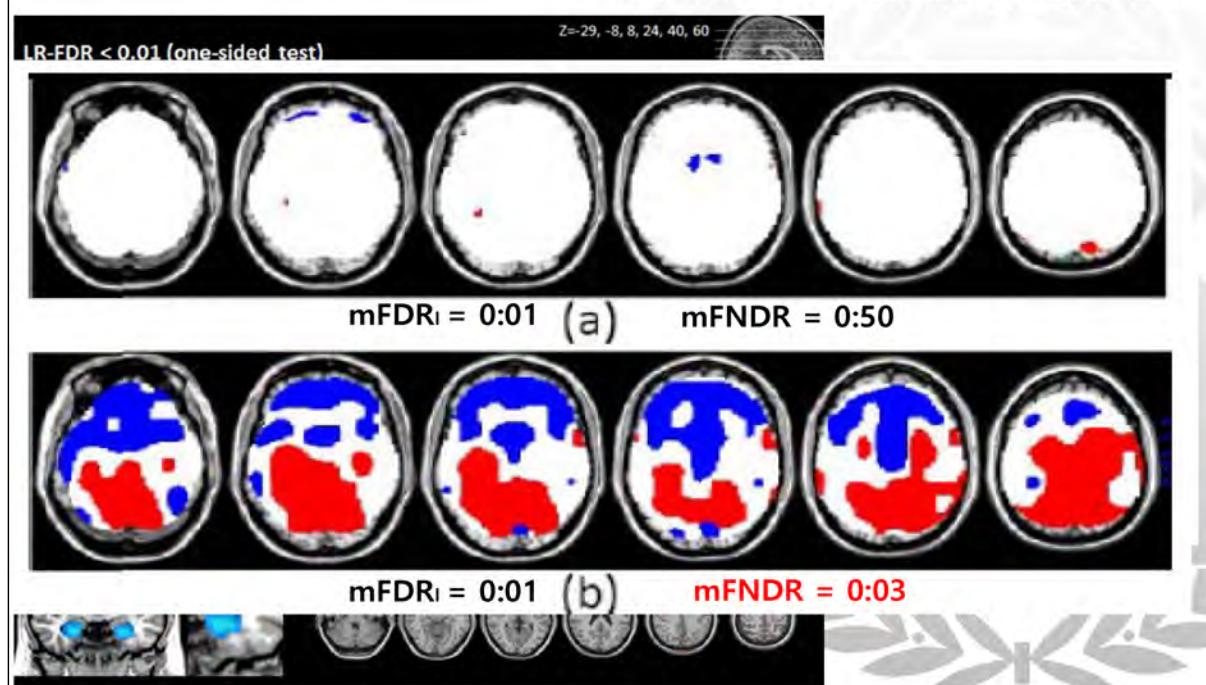


Optimal likelihood-ratio multiple testing

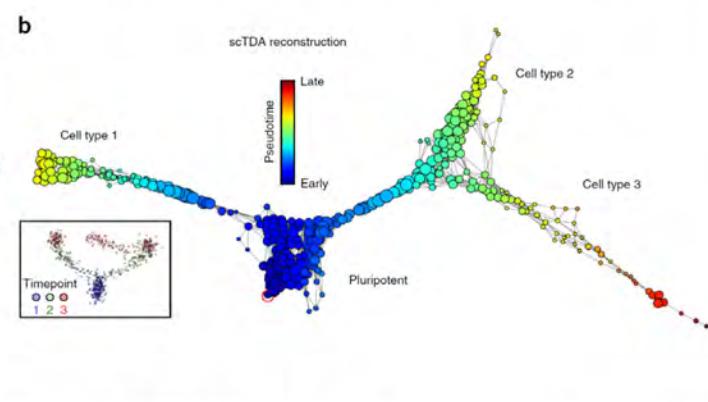
Lee, Kang et al. BMC Med Res Method 2015

Extended likelihood to multiple test under a hidden Markov random field model

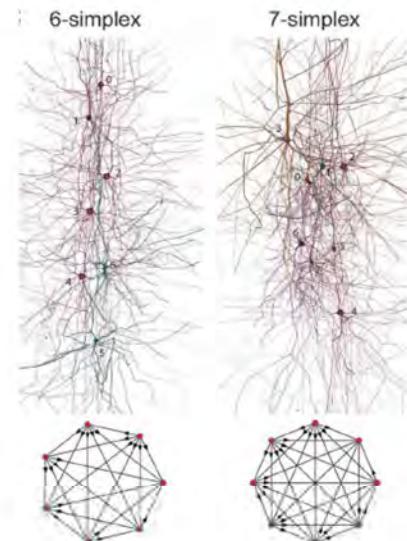
Lee D and Lee Y J. Multivariate Analysis 2016



Industrial & Mathematical Data Analysis and Research Center



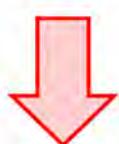
Rizvi et al 2017 Nat Biotech



Reimann et al 2017 Front Comp Neurosc

Current deep learning in medicine

Current DL models for medical fields



Diagnostic classification

- ✓ Differentiate diabetic retinopathy from normal
- ✓ Differentiating dermatologic malignancy
- ✓ **Differentiating images of Alzheimer's dementia from controls**

Supervised learning from specialized datasets

Limited Situation & Supervised

For images (Plain, CT, MRI, PET)

- ✓ Image segmentation
- ✓ Lesion localization
- ✓ Enhancing image reconstruction

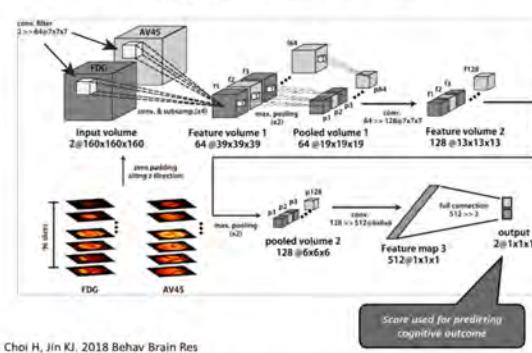
Unsolved Problems

Black Box
Human or Machine

CNN, LSTM, VAE, GAN, Graph CNN, Relational CNN, whatever helpful...

딥러닝으로 빅데이터 처리하기

Predicting Outcome Using Deep Learning-based Biomarker



Choi H, Jin KJ. 2018 Behav Brain Res

MIT
Technology
Review

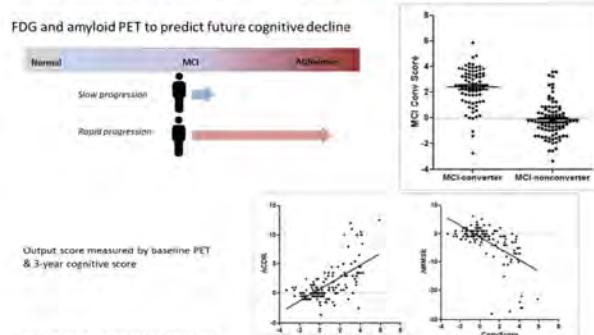
MIT Technology Review
Global Panel

- Reversing Life
- Software Predicts Cognitive Decline Using Brain Images
- Spotting people most at risk of developing Alzheimer is a difficult task. Now a new algorithm can identify those likely to be diagnosed in the next three years.
- by Emerging Technologies from the arXiv April 2017

Today, Hongyoon Choi at the Cheonan Public Health Center and Kyong Hwan Jin at the Korea Advanced Institute of Science and Technology, both in South Korea, say they have used deep learning to develop just such a technique. These guys say their process can accurately identify people likely to develop Alzheimer's in the next three years.

MIT Technology Review

Predicting Mild Cognitive Impairment Outcome Using Deep Learning-based Biomarker



Choi H, Jin KJ. 2018 Behav Brain Res

Deep Learning making Prediction: just Utility

Online Demo



Future cognitive decline prediction as well as current diagnosis

at <https://elhfdpet.appspot.com>

Making Population Distribution along Aging

Individual is compared with population of certain ages
Aging of individual brain using generative model

Missing data imputation

VAE model for generation of brain PET images of each age epoch
→ Aging movie

Choi H, Kang H, Lee DS. 2018 Front Aging Neurosci

Comparing Individual with 'Population of his/her age'

Individual aging brain: Comparing with **virtual population**

Population distribution of brain metabolism of each age by iterative generation using the VAE model

Evaluating individual brain's aging compared with general population
Define 'pathologic aging'

Choi H, Kang H, Lee DS. 2018 Front Aging Neurosci

Unsupervised learning

Age - 50.0

Generating Absent (or Missing) Images de novo

MR-less amyloid PET quantification by generating MR image

ADNI Alzheimer's disease

GAN (Generative Adversarial Networks) for MR generation with only PET

Unsupervised learning

Choi H, Lee DS, J Nucl Med 2018

PET/CT without CT

- PET routinely needs additional CT scan for attenuation correction
- Deep Learning-based attenuation map generation without additional CT scan

Hwang D, ..., Lee JS, JNM 2018 Epub [Improving accuracy of simultaneously reconstructed activity and attenuation maps using deep learning](#)

Computed tomography super-resolution using deep convolutional neural network

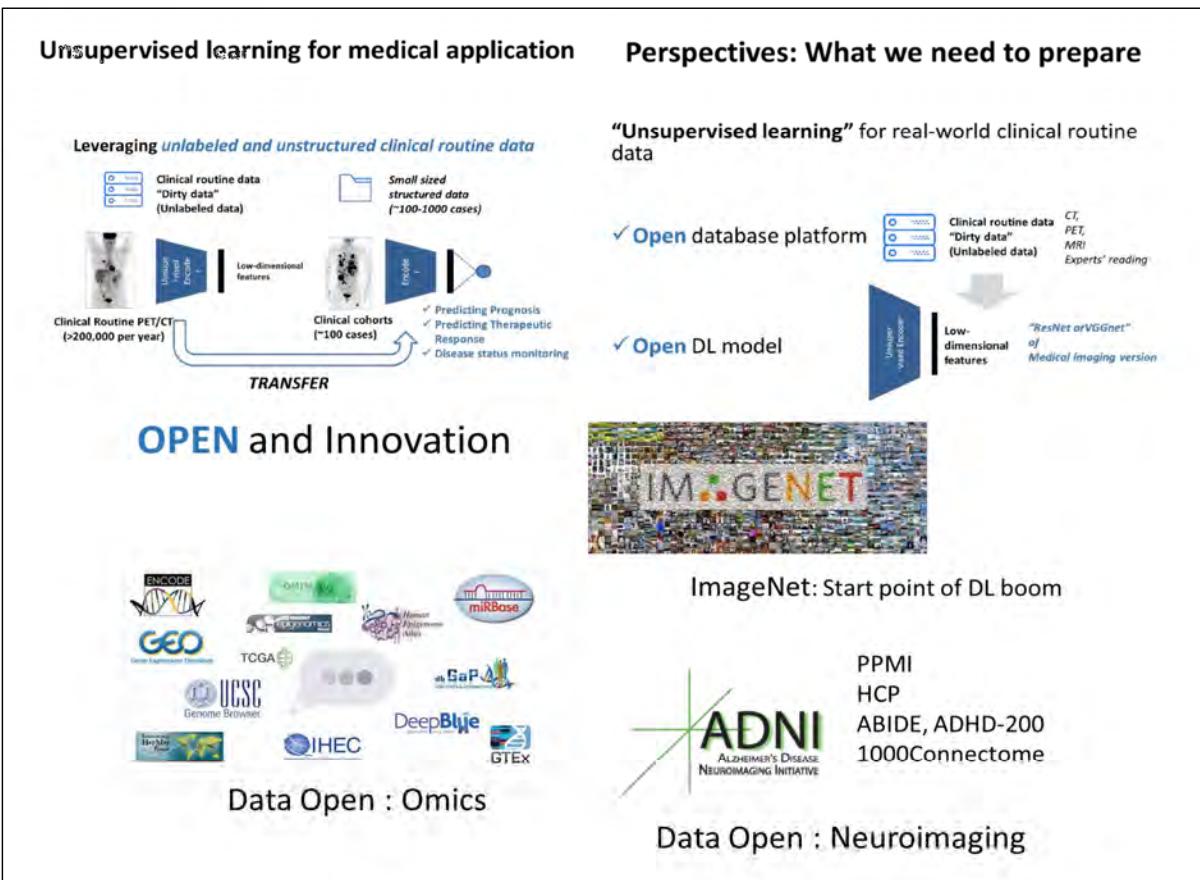
Park, ..., Lee JS, Phys Med Biol 2018

Deep CNN for recovering 1 mm slices from 5 mm ones.
(a) input image (5 mm slice), (b) input image focused by applying RL deblurring, (c) CNN output, and (d) ground truth (1 mm slice).

3D inpainting of brain MRI

Kang SK, ..., Lee JS, 2018 submitted

	Coronal	Sagittal	Coronal GM	Sagittal GM
Label				
Linear interpolation				
Neural network				



Suggestions

- **Statistics:** Hierarchical General Linear Model
 - Connectivity Data N^2 Dimension
- **Open Data:** Omics and Brain Imaging Data
 - Public and KBASE
- **Mathematical Frame of Analysis**
 - Persistent Homology over Hyperbolic Space
 - Topological Data Analysis

IV

지정토론

좌장 약력

성명	박 성 현	
소속	서울대학교 자연과학대학 통계학과 명예교수	
1. 학력		
기간	학교명	전공 및 학위
1964~1968	서울대학교 공과대학	화학공학과 공학사
1970~1972	미국 North Carolina State University	산업공학과 공학석사
1972~1975	미국 North Carolina State University	통계학과 이학박사
2. 주요 경력		
기간	기관명	직위, 직책
2016~현 재	사회적책임경영품질원	회장
2013~2015	국가과학기술자문회의 과학기술기반분과	의장
2013~2016	한국과학기술한림원	원장
2012~2014	건국대학교 경영대학 기술경영학과	석좌교수
2010~2012	한국연구재단 기초연구본부	본부장
2010~현 재	서울대학교 통계학과	명예교수
2000~2002	서울대학교 자연과학대학	학장
1995~1997	한국통계학회	회장
1977~2010	서울대학교 자연과학대학 통계학과	조교수, 부교수, 교수
1975~1977	미국 Mississippi State University 경영대학	조교수

토론자 약력

성명	김호	
소속	서울대학교 보건대학원	
1. 학력		
기간	학교명	전공 및 학위
1988	서울대학교	자연과학대학 계산통계학과 이학사
1990	서울대학교	자연과학대학 계산통계학과 통계학 석사
1996	미국 노쓰캐롤라이나 대학	보건통계학 박사
2. 주요 경력		
기간	기관명	직위, 직책
현재	서울대학교 보건대학원	원장
현재	한국보건통계정보학회	회장
현재	대한민국의학한림원	정회원
현재	질병관리본부 조사연구사업	평가위원
현재	예방의학회	이사 및 편집위원
현재	기후변화학회	이사
현재	서울대학교 아시아에너지환경지속발전연구소(AIEE S) 환경연구센터장	환경연구센터장

토론문

빅 데이터 시대의 보건학

•••

김 호

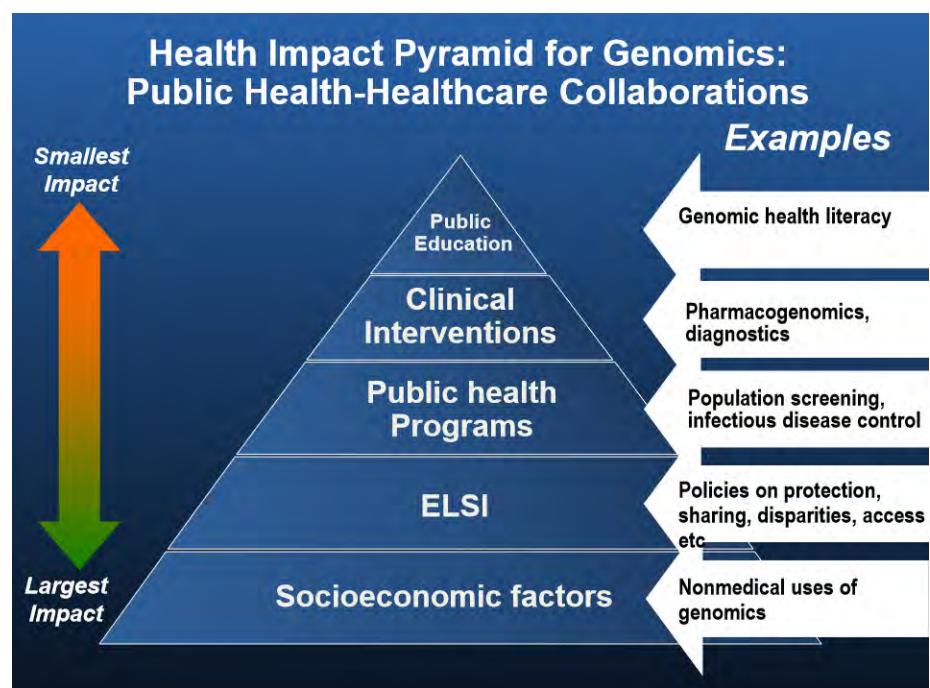
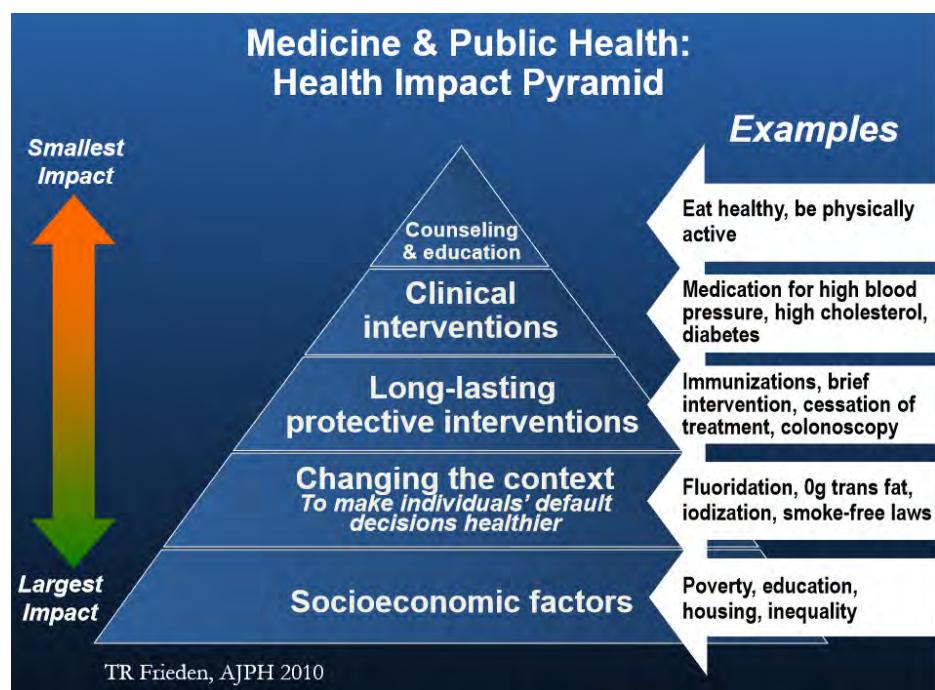
한국보건정보통계학회 회장

1. 정밀보건학 (Precision Public Health)의 특성

- 인구 수준의 건강을 연구하는 다학제 학문인 보건학은 정밀의학의 성공여부를 판단할 수 있는 이론적 근거를 제시함
- 치료에서 예방으로의 패러다임의 변화: 보건학적 개념
- 감염병 유행여부를 조기 진단할 수 있는 기반을 제공
- 감시체계, 평가체계, 보건정보 시스템의 획기적 발전에 기여

2. 의료 빅데이터 활용에 대한 논란

- 자유로운 의료빅데이터 활용의 순작용이 부작용보다 훨씬 클 것인가? 건강형평성 측면에서 많은 우려가 있음
- 공공성을 확보한 연구에서 의료빅데이터의 활용은 장려되어야 하고 많은 학회에서 이를 위한 노력을 경주하고 있음 (우리나라의 건강보험 자료는 잘 활용될 경우 그 가치가 대단히 큼. 이론적으로는 모든 국민의 건강상태를 실시간으로 알 수도 있음)
- 대중과의 소통 및 사회적 동의를 얻는 과정이 매우 중요함. 대중을 설득할 수 있는 좋은 예제가 필요함. 이러한 과정을 통하여 윤리적 정당성을 확보하려는 노력과 동시에 법제화 등 제도적 노력을 기울여야 함.



Muin J. Khoury MD, PhD

CDC Office of Public Health Genomics NCI Division of Cancer Control and Population Sciences



런던의 Wellcome Trust 본사에서 찍은 사진, 의료빅데이터 활용의 순작용을 대중에게 설명하려는 노력을 보여줌. “How patient data saves lives; Understanding Patient Data have launched a series of animations to explain how data saves lives. This video shows how data from other women with diabetes helps to improve Mandeep’s care and data about her will help the care of others too. Find out how #datasaveslives”

토론자 약력

성명	이영조	
소속	서울대학교 통계학과 교수	
1. 학력		
기간	학교명	전공 및 학위
1975~1979	서울대학교	학사
1979~1981	Iowa State Univ.	석사
1981~1983	Iowa State Univ.	박사
2. 주요 경력		
기간	기관명	직위, 직책
2011~현재	한국연구재단선정 데이터과학과지식창출 연구센터 (SRC)	센터장
2006~2007	한국데이터마이닝학회	회장
1994~현재	서울대학교 통계학과	부교수, 교수
1986~1994	한림대학교 통계학과	조교수, 부교수
1985~1986	Oklahoma State Univ. 통계학과	조교수
1983~1985	Oregon State Univ. 통계학과	조교수

토론자 약력

성명	이태림	
소속	한국방송통신대학 정보통계학과	
1. 학력		
기간	학교명	전공 및 학위
1973~1977	서울대 간호학과	학사
1977~1979	서울대 보건대학원	보건통계학 석사(PMPH)
1979~1981	서울대 자연대 통계학과	통계학 학사
1981~1983	서울대 자연대 통계학과	통계학 석사
1984~1989	중앙대 통계학과	통계학 박사
2. 주요 경력		
기간	기관명	직위, 직책
2018~현재	International Prize in Statistical Foundation	Organizing Committee
2015~2018	국제생물통계학회(IBS)	Executive Board Director
2014~2015	한국통계학회	부회장
2014~2015	방송대 자연대	학장
2011~2013	국제통계교육협회(IASE)	부회장
2009~2011	한국보건정보통계학회	회장
2009~2011	식약처	국가보건통계위원
2008~2010	통계청	국가통계위원
2006~현재	복지부	국가보건복지통계위원
2003~2007	IBS 한국지부 한국통계학회 생물통계연구회	회장

토론문

•••

이태림

방송통신대학교 교수

바이오 산업이 IT기술과 데이터 사이언스를 통해 지금까지와는 다른 모형으로 변화된 4차 산업혁명이 급속히 우리에게 다가오고 있다. 산업구조의 대변혁 과정에서 데이터가 바이오산업 혁신의 핵심자원으로 등장하여 지금까지의 공급중심에서 생성된 데이터를 온라인상에서 수집하고 분석하여 실시간 맞춤형 생산 공급이 이루어지는 수요자 중심으로 변하고 있다. 최근 헬스케어가 예측 예방 중심으로 변화하고 있어 축적된 빅데이터의 활용이 핵심이 되고 개인 맞춤형의료에 의한 의료비용 절감 측면에서 보건의료분야에 새로운 패러다임을 제공할 수 있어서 바이오분야에서의 빅데이터의 관리 및 활용을 위한 데이터사이언스 대한 요구도가 급증하고 있다.

이에 바이오데이터의 특수성과 다양성을 확인하고 융합의 필요성과 이를 위한 표준화 플랫폼을 찾고 강조하여 데이터 사이언스를 활용한다면 바이오강국으로 가는 길의 실마리를 찾을 수 있을 것이다.

스탠포드의 Biomedical Data Science Initiatives의 비전에 대규모의 계산과 데이터분석으로 인간의 건강과 질병을 향상시키고자하여 유전체자료를 환자 진료에 임상자료와 오믹스 데이터까지 통합 적용하여 환자를 깊이 이해하고자 한다는 것으로 여기에 모바일헬스자료도 건강을 도모하는 데 이용할 수 있다고 정의하고 있다.

최근 의료정보의 패러다임의 변화로 사람중심에서 정보자원 중심으로 바뀌고 이에 따른 자료 분석을 통한 정보산출이 중요하여 자료분석을 위한 데이터사이언스의 활용과 역할이 필수적이고 이를 위한 데이터사이언스 교육이 중요한 시대이다. 또한 요즈음과 같은 분야의 융합시대에 기술의 융합 즉 IT, BT, NT, DT(Data Technology) 등의 신기술 융복합이 주도하는 차세대 기술혁명이 성공을 좌우하므로 이 또한 교육과정에 반영되어야 한다.

의료서비스의 패러다임의 변화는 맞춤의학 예측의학 시대가 도래하여 바이오-정보-의학의 기술은 이 모든 계층을 꿰뚫을 핵심기술로서의 의료-건강 분야의 민주화를 위해 응용가능성이 가장 높은 기술로 현 4차 산업혁명시대의 사회에서 요구되는 분야이며 이에 대학 교육의 교과 과정에 이러한 통합분야가 필수적으로 포함되어야 한다.

국가차원에서도 현실적으로 요구되는 바이오정보 관련 전문인력의 부족을 보완하기 위해 BITEC(Biomedical Information Training & Education Center)등의 프로젝트를 복지부 지원으로 진행하여 5년간의 BITEC 프로젝트 결과인 교육콘텐츠를 방송대 프라임칼리지 OER 싸이트에 탑재하여 의학전공자 자연과학 전공자들의 바이오정보학, 의료정보학, 데이터사이언스 무료교육과 수련을 도모하고 있다.

브레인 이미징 자료의 분석과 모형화는 데이터사이언스를 응용하는 전형적인 예로 객관화와 정량화한 신뢰성이 높은 결과 생산을 위해 질병군과 정상인의 차이를 분류모형으로 적합하여 질병환자 개인의 질병중증도 향후 진전 예측에 의한 예후진단 등을 Clinical Decision Support System을 통해 적용할 수 있고 머신러닝의 분류모형을 통한 시스템 구축이 가능하다.

Deep Learning을 응용하여 의료데이터를 이해하기 쉬운 시각화 그래프로 작성하고 오믹스 데이터로 예후를 쉽게 예측할 수 있고 적절한 데이터사이언스 원리를 이용하여 객관화/정량화를 쉽게 하여 임상에 적용 가능하다. Deep Learning과 Minimal Spanning Tree 등의 모델링을 공공의 데이터베이스에 적용하여 얻은 결과를 알츠하이머, 파킨스병, 우울증 등의 진단, 예후 예측에 응용할 수 있다. 이러한 모형을 통하여 임상의학자료의 숨은 자료 구조를 계량적으로 찾아주고 시각화하여 일반인들의 개인 의료정보를 이해하기 쉽도록 도모하는 것이 필요하다.

의료자료에 분류모형, Deep Learning, 머신러닝을 적용하여 감별진단, 예후진단에 적용하여 수리적인 진단이 가능하고 이러한 분류모형을 통하여 질병의 유사군인 집약화를 통한 감별 진단 세부진단의 기준을 구할 수 있고 비슷한 증상들의 군을 만들어 질병의 증상, Biomarker 등을 규명할 수 있다.

정밀의학을 위한 데이터사이언스 적용은 마커의 선택- 예측모형 구축- 정밀의학으로 진행되는 데 이 과정의 해결을 위해서는 통계, 전산, AI, 임상이 통합된 인력이 필요한데 이를 위해서는 각 대학의 교육 커리큘럼에 반영하고 관련된 교육기관과 연구소 정부기관이 협업하는 BITEC와 같은 정부지원의 비정규 교육 프로그램 컨소시움의 구성도 또 하나의 해결방법이 될 것이다.

토론자 약력

성명	정진호	
소속	서울대학교 약학과 (한림원 의약학부 학부장)	
1. 학력		
기간	학교명	전공 및 학위
1974~1978	서울대학교	제약학과, 학사
1978~1980	서울대학교	생명약학, 석사
1981~1987	미국 Johns Hopkins 대	독성학, 박사
2. 주요 경력		
기간	기관명	직위, 직책
2016~현재	과기부 국민생활과학자문단	단장
2016~현재	한국과학기술한림원	의약학부 학부장
2013~2014	(사)한국약학교육협의회	이사장
2011~2016	국조실 식품안전정책위원회	심의위원
2011~2013	서울대학교 약학대학	학장
2009~2010	한국독성학회	회장
2008~2011	서울대학교 환경안전원	원장
2007~2008	한국식품위생안전성학회	회장
2007~2010	식품의약품안전청	연구심의위원회, 위원장
2004~현재	한국과학기술한림원	정회원
1987~1988	미국 NIH	방문연구원

토론문

•••

정진호

한림원 의약학부장(서울대학교)

현재 AI와 빅 데이터를 활용하여 정밀의학 여러 분야에서 혁명이 일어나고 있지만, 본 토론자는 맞춤형 신약개발의 응용분야에 초점을 맞추고자 합니다.

신약개발은 많은 비용이 소모되며, 다양한 연구 인프라가 필요하며, 또한 정부의 허가를 받기가 쉽지 않습니다. 또한 최근 연구보고서에 따르면 신약이 시판이 된다고 해도 모든 사람, 또는 인종에 따라 동일한 효과를 보이지 않습니다. 이러한 이유 때문에 신약개발을 주도하는 다국적 제약회사들은 사람들에게 좋은 효과를 갖는 약을 개발하기 보다는 많은 이익을 창출할 확률이 높은 약의 개발하고자 하는 압박을 받고 있는 것이 사실입니다.

그러나 다국적 제약회사의 이러한 패러다임은 최근 발전하는 AI와 빅 데이터를 활용하여 큰 변화가 일어나고 있습니다. 제약회사는 AI와 빅 데이터를 활용하여 보다 효과적인 임상 시험을 거쳐 신약개발에 중대한 변화를 가져올 수 있는 새로운 방법을 제시하고 있습니다.

데이터 사이언스가 가져온 이점을 살펴보기 전에 과거에 개발된 약의 갖고 있는 문제점을 먼저 지적하고자 합니다. 매일 수백만 명의 사람들은 별로 효과가 없는 약을 복용하고 있습니다. *Nature*에 실린 통계에 따르면 미국에서 가장 높은 매출을 올리는 상위 10 위이내의 처방약 중에서 효과가 좋은 것은 25% 정도 환자에게만 적절한 치료효과를 보입니다. 일부 처방약은 25명 중 1명에게만 효과를 보이고 있습니다. 그리고 콜레스테롤을 낮추기 위해 사용하는 일부 스타틴 (statins) 제제는 50명 중에 1명에게만 유익한 효과가 있을 뿐입니다. 또한 서양인을 대상으로 임상시험을 거친 약들은 다른 인종에게는 해로운 약이 될 수도 있습니다.

이처럼 허가받은 신약들이 인체 효과가 매우 다르게 나타나는 중요한 이유는 사람마다 유전체 구성 (계놈) 차이, 생체 단백질 (프로테옴) 차이, 그리고 인체가 갖고 있는 미생물 차이가 있기 때문입니다. 과거에 선진국의 신약들은 주로 백인을 대상으로 임상시험을 거쳐 개발되었기

때문에 타 국가, 인종에는 잘 맞지 않는 경향을 보여 왔습니다. 이러한 문제점을 극복하기 위해 서는 대상 환자의 다양한 풀 구성, 유전체에 맞춘 대상환자 확보, 등을 통하여 많은 사람들에게 효과가 있는 약을 찾아내려고 했지만 시간과 비용이 천문학적으로 증가되고 있습니다. 한 조사에 따르면 신약개발 비용이 매년 8.5%씩 증가되는 다국적 제약회사가 견디기 어려운 상황에 몰리고 있습니다.

발제 자료에 언급된 바처럼 2015년 오바마 정부는 2.15억불 '국가 정밀의학 이니셔티브' 프로그램을 출범시켰습니다. 즉 100만명 이상의 국민들이 어떤 종류의 약물에 반응하며, 특정 질병에 걸릴 위험이 있는지, 그리고 건강과 운동을 통해 나이를 먹고 사망하는 지에 관한 건강기록 Data를 축적하고 있습니다. 목표는 앞으로 수년 내에 코호트에 1 백만명 이상의 미국인을 등록하는 것이며 모든 개인의 데이터 (개인정보 비공개)는 관심있는 연구원에게 제공 될 예정입니다.

코호트의 모든 구성원은 자신의 유전자 정보를 등록하고, 자신의 건강 기록, 생활 습관 및 환경 노출을 추적합니다. 이런 과정을 거친 빅 데이터의 플랫폼은 다양한 용도로 활용될 수 있으며 신약개발 과정에서는 소규모 효율적인 임상시험을 위한 플랫폼 역할을 할 수도 있습니다.

과거의 임상시험은 환자의 유전정보, 질병 양상 등이 개인에 따라 차이가 있는데도 불구하고 이런 문제를 고려하지 않은 채 통계적 유의성을 밝히는 게 쉽지 않았습니다. 따라서 소수 사람들에게는 효과가 있어도 훨씬 많은 사람들이 효과를 못 볼 경우 약의 개발은 중단되곤 했습니다.

그러나 데이터 플랫폼이 구축되면 임상시험 과정에서 약의 효과를 볼 것으로 예상되는 사람을 미리 골라내서 임상시험을 진행할 수 있습니다. 때문에 맞춤형 신약개발을 통하여 성공률도 높아지며 비용과 시간이 절감됩니다. 또 인종이나 민족별, 더욱 나아가 개인별로 약을 세분화시킬 수 있습니다.

그리고 과거에 개발과정에서 독성과 낮은 효과 등으로 중도에 탈락했던 후보물질들이 재평가가 가능합니다. 또한 A라는 질병에 쓰던 약을 B·C 등 다른 질병 치료제로 범위를 확대하는 등 약의 재배치(repositioning)나 재목적화(repurposing)가 가능해지는 등 신약개발과 응용에 다양한 방식으로 활용되게 됩니다.

이처럼 데이터 플랫폼을 활용하여 임상시험을 설계하고 관리하는 방식으로 변경된다면 미래의 '맞춤형 정밀 의약품' 개발 비용이 현저히 절감될 수 있으며, 환자의 특성에 맞춘 임상시험 결과의 개선을 통하여 올바른 약제를 환자들에게 신속하게 처방할 수 있을 것입니다. 궁극적으로 제약회사는 신약개발 속도가 빨라지고 비용이 절감되며 환자들에는 뛰어난 효과가 있는 약을 복용하게 됨으로써 제약회사와 환자 모두에게 만족할 수 있는 전략이 될 것입니다.

토론자 약력

성명	한현욱	
소속	차의과학대학교 의학전문대학원 정보의학교실 교수	
1. 학력		
기간	학교명	전공 및 학위
1994~2001	한양대학교	공학사 (B.S)
2002~2004	서울대학교	공학석사 (M.S)
2007~2015	차의과학대학교	의학 석사 및 박사 (M.D, Ph.D)
2. 주요 경력		
기간	기관명	직위, 직책
2018~현재	차의과학대학교 의학전문대학원 정보의학교실	주임교수 / 조교수
2018~현재	대한의료정보학회	이사
2017~현재	세계중계생명정보학회	위원
2017~현재	대한의료정보학회 유전체데이터분석	강사
2017~2018	아주대학교 의과대학교 의료정보학과	연구조교수
2015~2017	차의과학대학교 의학전문대학원 예방의학교실	임상조교수
2010~2015	서울대학교 시스템바이오정보의학연구센터	선임연구원
2004~2006	LG전자기술원	주임연구원

토론문

헬스케어 분야 데이터 사이언티시트 양성을 위한 제언

•••

한현욱

차의과대학교 부교수

빅데이터, 인공지능, 사물인터넷 및 클라우드로 대변되는 4차 산업혁명 시대의 막이 올랐습니다. 이에 의료계에서도 이러한 변화의 움직임이 어떠한 형태로 의료에 영향을 미칠지에 관해 진지하게 논의가 이루어지고 있습니다. 최근 대한민국의 의사들을 상대로 4차 산업혁명과 미래 의료 변화에 관한 설문조사를 한 결과 대부분의 의사들이 이러한 새로운 기술들이 미래 의료 환경에 어떻게든 영향을 미칠 것이란 것에 동의하였고, 그 중심에 인공지능, 빅데이터, 로봇과 같은 최첨단 기술이 있을 것으로 생각하였습니다. 이러한 이유로 의학교육에서도 지능정보사회를 대비하기 위한 새로운 교육과정을 도입해야 한다는 논의도 본격적으로 이루어지고 있습니다.

미래의료는 과거, 경험 중심의 전통적인 의료 행위를 지양하고 데이터기반의 근거중심의료로 변화될 것입니다. 이는 미래의료가 추구하는 정밀의학이란 키워드로 대변된다고 할 수 있는데, 정밀의학(Precision Medicine)란 진료정보, 유전정보 및 생활습관 등 개인의 건강정보를 기반으로 개별사람들에게 최적화된 진단 및 치료를 적용하는 새로운 헬스케어 패러다임을 말합니다.

이를 실현하기 위해 가장 중요한 것은 앞서 언급한 개인의 건강정보를 어떻게 체계적으로 수집하고 분석 할 수 있느냐 하는 것입니다. 의료는 그 양이 방대할 뿐만 아니라 고도로 전문성이 요구되는 분야입니다. 그리고, 데이터분석의 기반이 되는 통계학 또한 다양한 전문적 지식을 요구합니다. 더군다나, 최근에는 과거의 정형적인 구조를 갖는 소규모 데이터, 일명 스몰데이터를 넘어 구조와 형태가 다양하고, 그 양을 짐작하기도 어려울 정도로 방대한 빅데이터란 개념이 등장을 하였습니다. 여기에 전통적인 임상적 판단기준으로 삼아 왔던 평균의 개념보다는 데이터 자체의 통찰을 통해 네트워크 이론 기반의 인공지능의 개념이 등장하였고, 방대한 양의 데이터를 분석할 수 있는 하드웨어와 소프트웨어 ICT 인프라인 클라우드 또한 빅데이터를 분석하는데 활용되고 있습니다. 즉, 전통적인 통계학을 뛰어 넘어 대규모 데이터를 다루기 위한 컴퓨터 공학의 중요성이 지속적으로 부각되고 있다는 것입니다.

여기에 여러분들이 주목해 할 것이 바로 의료 분야에서 다루는 빅데이터인 “헬스케어 빅데이터”에 관한 것입니다. 그리고 이를 학문적으로 체계적으로 수립한 학문분야가 의료정보학과 생물정보학이란 연구 분야입니다. 하지만, 두 분야 또한 엄밀하게 다루는 영역이 다릅니다. 의료 정보학에서는 병원정보시스템을 기반으로 한 진료정보를 중심으로 임상데이터와 생체신호 데이터를 주로 다루고 있고, 생물정보학 분야에서는 주로 세포수준의 DNA 및 RNA등 미시적 분자 생물학 데이터를 다루고 있습니다. 최근에는 중계의학(Translational Medicine)이란 용어가 등장하며 의료정보학과 생물정보학의 융합의 중요성이 강조되고 있는데, 이 두 분야를 함께 통찰 할 수 있는 연구가 필요합니다. 현재, 이런 접점에서 수많은 새로운 지식의 체계가 만들어지고 있습니다. 예를 들어, 의학연구의 기반이 되는 컨트롤드 보케블러리로 UMLS나 SNOME CT등이 있고, 의료데이터 표준안으로 거론되는 것으로 HL7 FHIR나 공통데이터 모델(CDM)등이 있습니다. 유전자 수준의 연구를 위해서는 Gene Ontology, 차세대유전체 분석 기술(NGS)과 같은 생물정보학 분야에서만 다루는 기술적 이해가 필요합니다. 이는 의학이나 생물학 분야의 학생들에게 절대로 가르치고 있지도 않을 뿐만 아니라, 전통적인 통계학이나 컴퓨터공학에서 언급조차 안되고 있습니다. 이러한 접점에 있는 다양한 지식의 체계를 제대로 이해하지 못한 상태로 “헬스케어 빅데이터” 분야의 데이터 사이언티스는 그 역할을 제대로 수행해 내기 어렵습니다.

어떤 분야이건 그들만의 리그가 있습니다. 그 리그의 특성을 제대로 이해하지 못한 상태에서 단순한 데이터 분석 기술만을 가르치게 된다면, 사실상 데이터 사이언티스트 양성이 아니라 데이터 테크니션만 양산할 뿐입니다. 따라서, 진정한 의미의 정밀의료 분야의 “데이터 사이언티스”를 양산하기 위해서는 의학, 생물학, 통계학 및 컴퓨터공학을 아우르는 새로운 개념의 학과 개설을 검토할 필요가 있습니다. 또한, 국가차원에서 이런 분야 인재 양성의 중요성을 인식한다면, 이 분야에 도전하는 대학원생들에게 합리적인 인센트브를 제공해줘야 합니다. 왜냐하면, 이 모든 분야를 온전히 섭렵하는데 까지는 상당한 시간과 금전이 필요하기 때문입니다.

현재 IBM, 구글, 아마존과 같은 글로벌 IT기업 뿐만 아니라, 국내 네이버, 삼성, SK등 굴지의 국내 IT기업들도 모두 헬스케어 빅데이터 산업에 뛰어들고 있습니다. 하지만, 현재 우리의 교육현장에서 이런 분야를 모두 아무르는 인재를 찾기는 너무나도 어려운 게 현실입니다. 더군다나 지금 이 분야에서 현재 활동하는 여러 사람들이.. 많은 경우 이론가의 수준을 넘지 못하고 있다는 것도 문제입니다. 의료의 현안 문제를 다각도로 이해하고 통계학이나 컴퓨터공학의 루우 레벨의까지 기술적인 이해를 통해 이를 통찰할 수 있는 의료분야 데이터 사이언티스트 리더 양성이 시급한 이유입니다.

한림원탁토론회는...

•••

한림원탁토론회는 국가 과학기술의 장기적인 비전과 발전전략을 세우고, 동시에 과학기술 현안 문제에 대한 해결방안을 모색하기 위한 목적으로 개최되고 있는 한림원의 대표적인 정책토론행사입니다.

지난 1996년 처음 개최된 이래 지금까지 100여회에 걸쳐 초중등 과학교육, 문·이과 통합문제, 국가발전에 미치는 기초과학 등 과학기술분야의 기본문제는 물론 정부출연연구소의 발전방안, 광우병의 진실, 방사능, 안전 방제 등 국민생활에 직접 영향을 미치는 문제에 이르기까지 광범위한 주제를 다루고 있습니다.

한림원은 과학기술 선진화에 걸림돌이 되는 각종 현안문제 중 중요도와 시급성에 따라 주제를 선정하고, 과학기술 유관기관의 최고책임자들을 발제자로 초빙하여, 한림원 석학들을 비롯해 산·학·연·정의 전문가들이 심도 깊게 토론을 진행하고 있습니다.

토론후에는 책자로 발간, 정부, 국회와 관련기관에 배포함으로써 정책 개선방안을 제시하고 정책 입안자료를 제공하여 여론 형성에 기여하도록 힘쓰고 있습니다.

■ 한림원탁토론회 개최실적 (1996년 ~ 2018년) ■

회수	일자	주제	발제자
1	1996. 2. 22.	초중등 과학교육의 문제점	박승재
2	1996. 3. 20.	과학기술분야 고급인력의 수급문제	서정현
3	1996. 4. 30.	산업계의 연구개발 걸림돌은 무엇인가?	임효빈
4	1996. 5. 28.	과학기술 행정과 제도, 무엇이 문제인가?	박우희
5	1996. 7. 9.	연구개발 평가제도, 무엇이 문제인가?	강계원

회수	일자	주제	발제자
6	1996. 10. 1.	정부출연연구소의 역할과 기능에 대하여	김훈철
7	1996. 11. 4.	21세기 과학기술비전의 실현과 정치권의 역할	김인수
8	1997. 2. 25.	Made in Korea, 무엇이 문제인가?	채영복
9	1997. 4. 2.	산업기술정책, 무엇이 문제인가?	이진주
10	1997. 6. 13.	대학교육, 무엇이 문제인가?	장수영
11	1997. 7. 22.	대학원 과학기술교육, 무엇이 문제인가?	김정욱
12	1997. 10. 7.	과학기술 행정체제, 무엇이 문제인가?	김광웅
13	1998. 1. 22.	IMF, 경제위기 과학기술로 극복한다.	채영복
14	1998. 3. 13.	벤처기업의 활성화 방안	김호기, 김영대, 이인규, 박금일
15	1998. 5. 29.	국민의 정부의 과학기술정책	강창희
16	1998. 6. 26.	정보화시대의 미래와 전망	배순훈
17	1998. 9. 25.	과학기술정책과 평가제도의 문제	박익수
18	1998. 10. 28.	경제발전 원동력으로서의 과학기술의 역할	김상하
19	1999. 2. 12.	21세기 농정개혁의 방향과 정책과제	김성훈
20	1999. 3. 26.	지식기반 경제로의 이행을 위한 경제정책 방향	이규성
21	1999. 5. 28.	과학기술의 새천년	서정욱
22	1999. 9. 10.	신 해양시대의 해양수산정책 발전방향	정상천
23	2000. 2. 10.	21세기 환경기술발전 정책방향	김명자
24	2000. 4. 14.	경제발전을 위한 대기업과 벤처기업의 역할	김각중

회수	일자	주제	발제자
25	2000. 6. 16.	과학·기술방전 장기 비전	임관
26	2000. 9. 15.	국가 표준제도의 확립	김재관
27	2000. 12. 1.	국가 정보경쟁력의 잣대: 전자정부	이상희
28	2001. 5. 4	환경위기 극복과 지속가능 경제발전을 위한 과학기술개발전략	박원훈, 류순호, 문길주, 오종기, 한무영, 한정상
29	2001. 7. 18	국가 과학기술발전에 미치는 기초과학의 영향	임관, 명효철, 장수영
30	2001. 9. 21	산업계에서 원하는 인재상과 공학교육의 방향	임관, 한송엽
31	2001. 10. 31	적조의 현황과 앞으로의 대책	홍승룡, 김학균
32	2001. 12. 5	광우병과 대책	김용선, 한홍율
33	2002. 7. 19	첨단기술 (BT,ET,IT,NT)의 실현을 위한 산업화 대책	한문희, 이석한, 한송엽
34	2002. 9. 13	우리나라 쌀 산업의 위기와 대응	이정환, 김동철
35	2002. 11. 1	생명윤리 – 과학 그리고 법: 발전이냐 규제냐?	문신용, 이신영
36	2003. 3. 14	과학기술분야 졸업생의 전공과 직업의 연관성	조황희, 이만기
37	2003. 6. 18	국내 농축산물 경역현황과 발전방안	배상호
38	2003. 6. 27	대학과 출연연구소간 연구협력 및 분담	정명세
39	2003. 9. 26	그린에너지 기술과 발전 방향	손재익, 이재영, 홍성안
40	2004. 2. 20	미래 고령사회 대비 국가 과학기술 전략	오종남
41	2004. 10. 27	고유가시대의 원자력 이용	정근모
42	2004. 12. 7	농산물 개방화에 따른 국내 고추산업의 현황과 발전전략	박재복
43	2005. 9. 30	과학기술윤리	송상용, 황경식, 김환석

회수	일자	주제	발제자
44	2005. 11. 25	과학기술용어의 표준화 방안	지제근
45	2005. 12. 1	융합과학시대의 수학의 역할 및 수학교육의 방향	정근모, 최형인, 장준근
46	2005. 12. 15	해양바이오산업, 왜 중요한가?	김세권, 김동수
47	2006. 11. 7	첨단과학시대의 교과과정 개편방안	박승재
48	2006. 12. 22	과학기술인 복지 증진을 위한 종합 대책	설성수
49	2007. 6. 29	선진과학기술국가 가능한가? - Blue Ocean을 중심으로	김호기
50	2007. 11. 9	우리나라 수학 및 과학교육의 문제점과 개선방향	김도한, 이덕환
51	2008. 5. 9	태안반도 유류사고의 원인과 교훈	하재주
52	2008. 5. 8	광우병과 쇠고기의 안전성	이영순
53	2008. 6. 4	고병원성조류인플루엔자(AI)의 국내외 발생양상과 우리의 대응방안	김재홍
54	2008. 10. 8	High Risk, High Return R&D, 어떻게 해야 하는가?	김호기
55	2008. 11. 11	식량위기 무엇이 문제인가?	이정환
56	2008. 12. 11	초중고 수학 과학교육 개선방안	홍국선
57	2008. 12. 17	우리나라 지진재해 저감 및 관리대책의 현황과 개선방안	윤정방
58	2009. 2. 19	21세기 지식재산 비전과 실행 전략	김영민
59	2009. 3. 31	세계주요국의 나노관련 R&D 정책 및 전략분석과 우리의 대응전략	김대만
60	2009. 7. 20	국가 수자원 관리와 4대강	심명필
61	2009. 8. 28	사용후핵연료 처리 기술 및 정책 방향	송기찬, 전봉근
62	2009. 12. 16	세종시와 국제과학비즈니스밸트	이현구

회수	일자	주제	발제자
63	2010. 3. 18	과학도시와 기초과학 진흥	김중현
64	2010. 6. 11	지방과학기술진흥의 현황과 과제	정선양
65	2011. 2. 28	국제과학비지니스벨트와 기초과학진흥	민동필, 이충희
66	2011. 4. 1	방사능 공포, 오해와 진실	기자회견
67	2012. 11. 30	융합과학/융합기술의 본질 및 연구방향과 국가의 지원시스템	이은규, 여인국
68	2013. 4. 17	한미원자력협정 개정협상에 거는 기대와 희망	문정인
69	2013. 6. 11	통일을 대비한 우리의 식량정책 이대로 좋은가?	이철호
70	2013. 7. 9	과학기술중심사회를 위한 과학기술원로의 역할과 의무	이원근
71	2013. 7. 22	대학입시 문·이과 통합, 핵심쟁점과 향후 과제는?	박재현
72	2014. 1. 17	국가안보 현안과제와 첨단과학기술	송대성
73	2014. 3. 4	융합과학기술의 미래 – 인재교육이 시작이다	강남준, 이진수
74	2014. 5. 9	과학기술연구의 새 지평 젠더혁신	이혜숙, 조경숙, 이숙경
75	2014. 5. 14	남북한 산림협력을 통한 한반도 생태통일 방안은?	김호진, 이돈구
76	2014. 5. 22	창조경제와 과학기술	이공래, 정선양
77	2014. 5. 29	재해·재난의 예방과 극복을 위한 과학기술의 역할은?	이원호, 윤정방
78	2014. 6. 10	벼랑 끝에 선 과학·수학 교육	정진수, 배영찬
79	2014. 6. 14	문학과 과학, 그리고 창조경제	정종명, 최진호
80	2014. 6. 25	‘DMZ세계평화공원’과 남북과학기술협력	정선양, 이영순, 강동완
81	2014. 7. 24	국내 전통 발효식품산업 육성을 위한 정책 대안은?	신동화

회수	일자	주제	발제자
82	2014. 9. 17	'과학기술입국의 꿈'을 살리는 길은?	손경한, 안화용
83	2014. 9. 30	한국 산업의 위기와 혁신체제의 전환	이근
84	2014. 11. 14	경제, 사회, 문화, 산업 인프라로서의 사물인터넷(IoT): 그 생태계의 실현 및 보안방안은?	김대영, 김용대
85	2014. 11. 28	공유가치창출을 위한 과학기술의 나아갈 길은? 미래식품과 건강	권대영
86	2014. 12. 5	창발적 사고와 융합과학기술을 통한 글로벌 벤처 생태계 조성 방안	허석준, 이기원
87	2015. 2. 24	구제역·AI의 상재화: 정부는 이대로 방지할 것인가?	김재홍
88	2015. 4. 7	문·이과 통합 교육과정에 따른 과학·수학 수능개혁	이덕환, 권오현
89	2015. 6. 10	이공계 전문가 활용 및 제도의 현황과 문제점	이건우, 정영화
90	2015. 6. 25	남북 보건의료 협정과 통일 준비	신희영, 윤석준
91	2015. 7. 1	메르스 현황 및 종합대책	이종구
92	2015. 7. 3	'정부 R&D 혁신방안'의 현황과 과제	윤현주
93	2015. 9. 14	정부 R&D예산 감축과 과학기술계의 과제	문길주
94	2015. 10. 23	사회통합을 위한 과학기술 혁신	정선양, 송위진
95	2015. 11. 4	생명공학기술을 활용한 우리나라 농업 발전방안	이향기, 박수철, 곽상수
96	2015. 11. 9	유전자기위 기술의 명과 암	김진수
97	2015. 11. 27	고령화사회와 건강한 삶	박상철
98	2015. 12. 23	따뜻한 사회건설을 위한 과학기술의 역할: 국내외 적정기술을 중심으로	박원훈, 윤제용
99	2016. 2. 29	빅데이터를 활용한 의료산업 혁신방안은?	이동수, 송일열, 유희준
100	2016. 4. 18	대한민국 과학기술; 미래 50년의 도전과 대응	김도연

회수	일자	주제	발제자
101	2016. 5. 19	미세먼지 저감 및 피해방지를 위한 과학기술의 역할	김동술, 박기홍
102	2016. 6. 22	과학기술강국, 지역 혁신에서 답을 찾다	남경필, 송종국
103	2016. 7. 6	100세 건강과 장내 미생물 과학! 어디까지 왔나?	김건수, 배진우, 성문희
104	2016. 7. 22	로봇 기술과 미래	오준호
105	2016. 8. 29	융합, 융합교육 그리고 창의적 사고	김유신
106	2016. 9. 6	분노조절장애, 우리는 얼마나 제대로 알고 있나?	김재원, 허태균
107	2016. 10. 13	과학기술과 미래인류	이광형, 백종현, 전경수
108	2016. 10. 25	4차 산업혁명시대에서 젠더혁신의 역할	이우일, 이혜숙
109	2016. 11. 9	과학기술과 청년(부제: 청년 일자리의 현재와 미래)	이영무, 오세정
110	2017. 3. 8	반복되는 구제역과 고병원성 조류인플루엔자, 정부는 이대로 방치할 것인가?	류영수, 박최규
111	2017. 4. 26	지속가능한 과학기술 혁신체계	김승조, 민경찬
112	2017. 8. 3	유전자교정 기술도입 및 활용을 위한 법·제도 개선방향	김정훈
113	2017. 8. 8	탈원전 논란에 대한 과학자들의 토론	김경만, 이은철, 박홍준
114	2017. 8. 11	새롭게 도입되는 과학기술혁신본부에 바란다	정선양, 안준모
115	2017. 8. 18	ICT 패러다임을 바꿀 양자통신, 양자컴퓨터의 부상	허준, 최병수, 김태현, 문성욱
116	2017. 8. 22	4차 산업혁명을 다시 생각한다	홍성욱, 이태억
117	2017. 9. 8	살충제 계란 사태로 본 식품안전관리 진단 및 대책	이향기, 김병훈
118	2017. 11. 17	미래 과학기술을 위한 정책입법 및 교육, 어떻게 해야 하나?	박형욱, 양승우, 최윤희

회수	일자	주제	발제자
119	2017. 11. 28	여성과기인 정책 업그레이드	민경찬, 김소영
120	2017. 12. 8	치매국가책임제, 과학기술이 어떻게 기여할 것인가?	김기웅, 뮤인희
121	2018. 1. 23	항생제내성 수퍼박테리아! 어떻게 잡을 것인가?	정석훈, 윤장원, 김홍빈
122	2018. 2. 6	신생아 중환자실 집단감염의 발생원인과 환자안전 확보방안	최병민, 이재갑, 임채만, 천병철, 박은철
123	2018. 2. 27	에너지전환정책, 과학기술자 입장에서 본 성공여건	최기련, 이은철
124	2018. 4. 5	과학과 인권	조효제, 민동필, 이중원, 송세련
125	2018. 5. 2	4차 산업혁명시대 대한민국의 수학교육, 이대로 좋은가?	권오남, 박형주, 박규환
126	2018. 6. 5	국가 R&D 혁신 전략 – 국가 R&D 정책 고도화를 위한 과학기술계 의견 –	류광준, 유욱준
127	2018. 6. 12.	건강 100세를 위한 맞춤 식품 필요성과 개발 방향	박상철, 이미숙, 김경철
128	2018. 7. 4.	제1회 세종과학기술포럼	성창모, 박찬모, 이공래

MEMO

MEMO

MEMO

MEMO

www.kast.or.kr

본 사업은 과학기술진흥기금 및 복권기금의 지원으로 시행되고 있습니다.