

BIO ECONOMY REPORT

February 2018, Issue 8

유전체 데이터 국제 표준화 동향

유전체 데이터 국제 표준화 동향

신수용 경희대 조교수
김지현 선임연구원
이계민 선임연구원

1. 서론

유전체 데이터를 임상 데이터에 통합하는 것은 정밀의료(precision medicine) 실현의 핵심적인 목표 중 하나로 여겨지고 있다. 미국은 정밀의료이니셔티브(Precision Medicine Initiative)의 All of Us 프로그램을 통해 국립보건원(NIH, National Institutes of Health)를 중심으로 백만 명 이상의 사람들의 라이프스타일 및 환경, 유전체 데이터를 수집함으로써 근거 기반 진료(evidence-based care), 나아가 정밀의료(precision medicine)를 실현하는 노력을 펼치고 있다. 그런데 수집된 유전체 데이터가 임상 및 진료 현장, 바이오헬스케어 산업에서 활발히 사용되기 위해서는 먼저 국제적인 표준화가 이루어져야 한다.

“유전체 검사를 통한 개인맞춤형 의료(personalized medicine) 혹은 정밀의료(precision medicine)를 실현하기 위해서는 유전체 검사결과가 기존의 다른 의료정보들과 서로 연동되어서 복합적인 분석을 할 수 있어야 하며 이를 위한 다양한 표준들이 개발되어야만 한다.”

국내에서는 차세대염기서열분석(NGS)기반 패널 검사가 올해 3월부터 수가화되어 보험 급여가 인정되었고, 현재까지 다수의 병원에서 검사를 수행하고 있다.¹⁾ 하지만, 패널 검사를 수행하는 모든 병원들이 기존 병원정보시스템에 텍스트 형태의 보고서를 생성하여 의료진에게 전달하고 있거나, 외부 수탁을 하는 경우에는 PDF파일로 결과를 전달받아서 의료진에게 제공하고 있다. 이러한 상황은 전세계적으로 동일한 상황이라 NGS를 이용한 임상유전체검사보고서가 병원정보시스템에 연동되어 있지 않거나, 연동되어 있더라도 비구조화된 형태로 의료진에게 전달하여 해당 정보를 이용한 임상 의사결정지원시스템(clinical decision support system) 등에서 전혀 활용을 하지 못하고 있는 실정이다.

유전체 검사를 통한 개인맞춤형 의료(personalized medicine) 혹은 정밀의료(precision medicine)를 실현하기 위해서는 유전체 검사결과가 기존의 다른 의료정보들과 서로 연동되어서 복합적인 분석을 할 수 있어야 한다. 또한, 유전체 검사 결과를 차후에 연구용으로 활용하기 위해서도 현재 비구조화된 형태의 정보로는 활용이 불가능하여 다시 정형화하는 과정을 거쳐야 한다. 현재 국내 병원들은 이런 문제를 극복하기 위해서 유전체 검사 보고서는 텍스트로 생성하여 의료진에게 제공하고, 자세한 검사 결과는 엑셀 파일 등으로 관리하고 있는 실정이다.

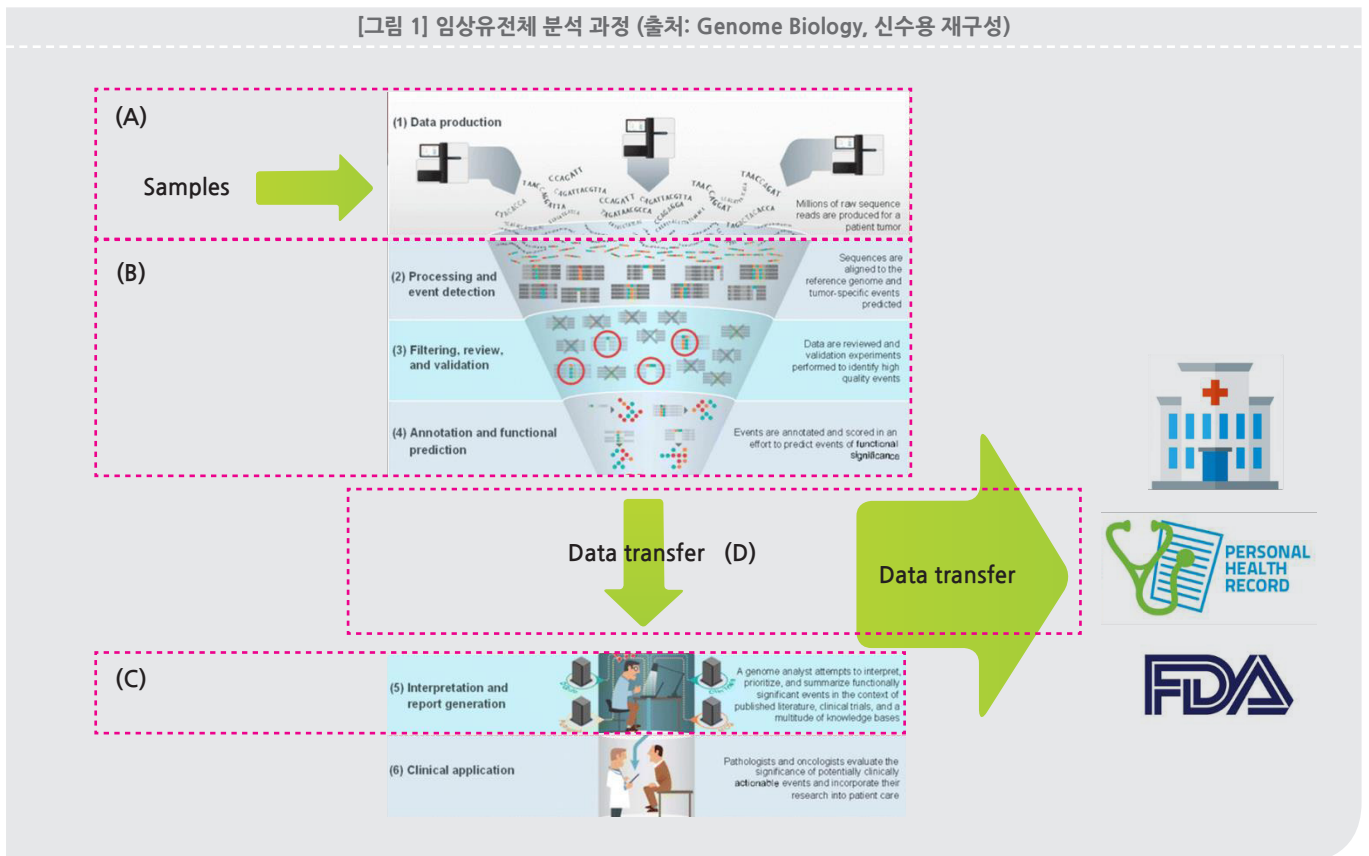
이러한 문제점을 극복하기 위해서는 의료 현장에서 유전체 정보를 활용하기 위한 다양한 표준들이 개발되어야만 한다. 본 보고서에서는 현재 국제적으로 진행되고 있는 임상 유전체 정보 표준을 소개하고, 향후 어떤 표준들이 더 개발되어야 하는지에 대해서 자세히 소개하고자 한다.

1) NGS 패널검사 실시 건수는 2017년 3분기(9월)까지 2,315건이고, 청구 건수는 1,453건이다. 2017년 3월 NGS 패널 검사기관으로 승인을 받은 총 22곳의 의료기관은 가톨릭대 인천성모병원, 가톨릭대 서울성모병원, 고려대 구로병원, 고려대 안암병원, 국립암센터, 길병원, 녹십자의료재단 녹십자의원, 랩지노믹스진단검사의학과의원, 부산대병원, 분당서울대병원, 분당차병원, 삼광의료재단 삼광의원, 삼성서울병원, 서울대병원, 서울아산병원, 서울의과학연구소부설 용인의원, 순천향대 서울병원, 아주대병원, 연세대 강남세브란스병원, 연세대 세브란스병원, 이원의료재단 이원의원, 인하대병원이다.

2. 유전체 분석 절차

그림 1은 진료 목적으로 유전체 검사를 해서 환자 치료에 활용하는 전체 과정을 보여주고 있다. 우선 유전체 서열분석기기(시퀀서, sequencer)를 이용해 유전체 데이터를 생성한다((1) Data production). 생성된 원시 데이터(raw data)를 분석 파이프라인으로 분석하여 유전변이를 찾는다((2) Processing and event detection ~ (4) Annotation and functional prediction). 발견한 유전변이를 해석하고 검사 보고서를 생성한다((5) Interpretation and report generation). 그리고 마지막으로 유전체검사 보고서를 확인하여 환자 치료에 활용한다 ((6) Clinical application). 중간에 명시된 데이터 전송(data transfer)은 외부수탁검사기관을 활용하는 경우를 가정하여 해당 기관이 유전변이결과와 원시 데이터 등을 의료기관으로 데이터를 전송하는 경우와 의료기관 사이에 유전체정보 교류 혹은 식품의약품국(FDA)에 보고하는 경우 등을 의미한다.

[그림 1] 임상유전체 분석 과정 (출처: Genome Biology, 신수용 재구성)

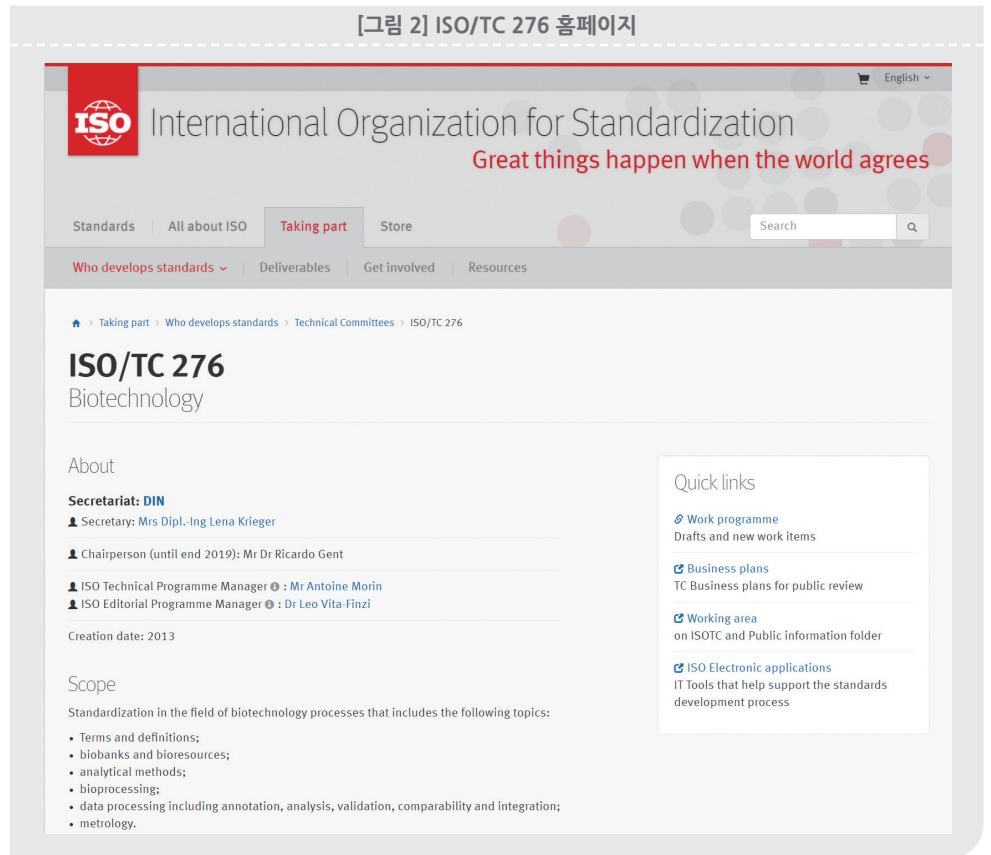


아주 많은 과정들이 관여하고 있다는 것을 확인할 수 있는데, 각 과정에서 필요한 표준들을 구분하기 위해서 크게 4개의 카테고리 ((A) ~ (D))로 구분할 수 있다. (그림 1의 빨간색 점선 박스). 다음 각 섹션에서 각 카테고리에 대해서 개발된 표준과 필요한 표준들에 대해서 소개해보고자 한다.

3. 유전체 서열분석 (시퀀싱) 과정

그림 1의 (A)에 해당하는 과정으로 환자 검체로부터 시퀀싱(sequencing)을 하는 과정이다. 이 과정은 현재 표준안이 없는 상황이다. 하지만 관련한 표준안들을 ISO/TC 276 Biotechnology (<https://www.iso.org/committee/4514241.html>)에서 개발을 하고 있다.

“시퀀싱 과정의 표준 개발은 현재 까지 초기 단계이기 때문에, 한국의 적극적인 참여를 통해 관련 표준안을 주도할 수 있는 상황이다.”



한국은 30개 공식 참여국의 일원으로 활동을 하고 있고, 2018년 2월 현재 1개의 표준이 출판되었으며, 18개의 표준이 개발 중에 있는데, 리스트는 ISO 홈페이지에서 확인할 수 있다.²⁾ 현재 까지 초기 단계이기 때문에, 한국의 적극적인 참여를 통해 관련 표준안을 주도할 수 있는 상황이다. 그리고, ISO/TC 215 Health informatics에서도 한국 주도로 관련 표준을 제안한 상황이다. 'ISO/PWI 22692 Quality control metrics for DNA sequencing'이라는 표준으로 현재는 개발 제안이 받아들여져서 예비작업항목 (Preliminary Work Item, PWI)로 등록이 되어 있고, 현재 정식 개발을 위한 신규 제안 (New Proposal, NP) 투표를 진행하고 있다.

다만 이 과정은 다른 규제를 통해 관리를 하고 있는데, 예를 들면, 한국의 경우 식품의약품안전처에서 2016년 7월에 발표한 “차세대염기서열분석(NGS) 임상검사실 인증 가이드라인”을 통해 임상검사실을 인증해 주고 있으며³⁾, 미국의 경우에는 FDA에서 관리하는 CLIA (Clinical Laboratory Improvement Amendments)를 통해 관리하고 있다.

2) <https://www.iso.org/committee/4514241/x/catalogue/p/0/u/1/w/0/d/0>
 3) 2017년 12월 기준 NGS 임상검사실 인증 제도를 통해 인증 받은 총 7개의 기관은 (주)테라젠이텍스 바이오연구소, 가톨릭대학교 서울성모병원, 삼성서울병원, (주)녹십자지능, (주)마크로젠, 연세대학교 세브란스병원, 서울아산병원이다.

시퀀서의 경우 현재 표준은 없으며, 기업체가 자체적으로 개발한 방식을 따르고 있는 상황이다. 다만 환자에게 적용하기 위해서는 시퀀서가 의료기기로 등록되어야 하기 때문에, 한국의 경우 식품의약품안전처에서 2016년 7월에 발표한 “차세대염기서열분석 (Next Generation Sequencing) 체외진단용 의료기기의 성능평가 가이드라인” 등을 통해 시퀀서를 관리하고 있다. 현재까지 미국 Illumina의 MiSeqDx와 Thermo Fisher Scientific의 Ion PGM Dx가 NGS 기반 유전자서열검사기로 허가를 받았으며 NGS 임상검사실의 경우 Illumina의 HiSeq 2500, NextSeq 500, NextSeq 550, Thermo Fisher Scientific의 Ion PGM Dx를 이용한 검사들을 제공하고 있다.

다만 생성된 파일들은 학계에서 공통적으로 사용하고 있는 FASTQ 파일 포맷 등이 사실상의 표준(de facto standard)로 사용되고 있다. 즉, 유전체 정보 저장의 단위는 현재 사실상의 표준으로 사용되는 파일 포맷들이 활용될 것으로 예상하고 있다.

4. 변이 분석 과정

“변이 분석 과정은 유전체 분석을 하는 파이프라인들로 구성되어 있으며 파이프라인의 성능을 검증하기 위한 노력들은 활발히 진행되고 있다.”

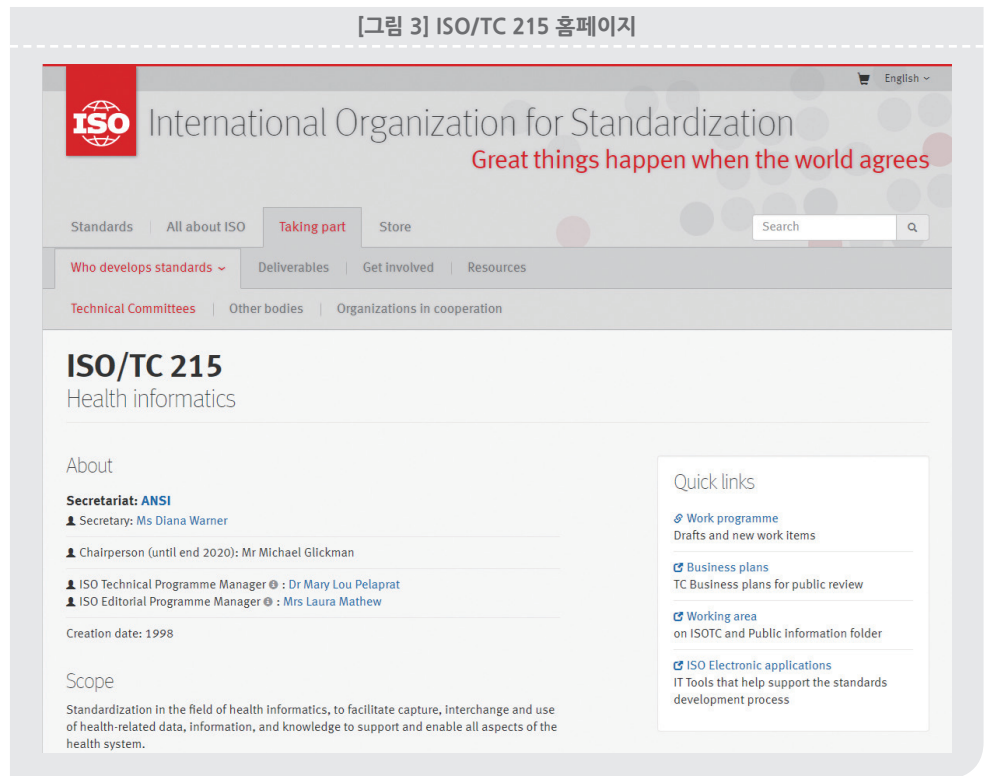
변이 분석 과정은 그림 1의 (B)에 해당하는 과정들로, 유전체 분석을 하는 파이프라인들로 구성되어 있다. 해당 과정들의 표준은 현재 시퀀싱 과정과 동일하게 발표된 표준은 없는 상황이며 역시 ISO/TC 276에 관련 표준들을 개발 진행 중이다.

다만, 파이프라인의 성능을 검증하기 위한 노력들은 활발히 진행되고 있다. 대표적인 것이 미국 정밀의료 프로젝트를 통해 시작된 PrecisionFDA (<https://precision.fda.gov/>)이다. PrecisionFDA에서는 Genome in a Bottle (GIAB) (<http://jimb.stanford.edu/giab/>) 컨소시엄과 미국 표준연구소인 NIST (National Institute of Standards and Technology)가 공동으로 개발한 Reference Material (RM 8398 - Human DNA for Whole-Genome Variant Assessment) (https://www-s.nist.gov/srmors/view_detail.cfm?srm=8398)을 이용하여 분석 프로그램들의 성능을 검증하는 노력을 진행하고 있다. Reference Material은 현존하는 모든 기술을 동원하여 해당 검체의 가능한 유전변이를 전부 찾아낸 변이 목록을 같이 제공하기 때문에, 프로그램 성능 검증을 위한 정답을 제공하는 역할을 하고 있다. 특히 참여를 독려하기 위해서 여러 챌린지를 만들어서 수상을 하고 있는데, Consistency Challenge, Truth Challenge, Hidden Treasures - Warm Up Challenge 등이 대표적이다. Consistency Challenge를 통해서 앞에서 소개한 RM 8398의 변이를 얼마나 정확히 찾아내는 지 평가하는 챌린지였고, Truth Challenge는 여기서 한걸음 더 나아가 정답을 공개하지 않은 별도의 데이터를 이용하여 성능을 다시 한번 평가한 챌린지였다. 2017년 7월에 있었던 Hidden Treasure 챌린지는 RM 8398에 변이 목록에 임의로 생성한 변이를 추가하여 분석 프로그램의 성능을 평가하였다.

5. 해석 및 검사 보고서 생성

그림 1의 (C)에 해당하는데, 유전 변이들을 발견한 이후 해당 변이들의 임상적 의미를 해석하고, 임상 유전체 검사 결과 보고서를 생성하는 과정이다. Health IT와 밀접한 관련이 있는 과정이라서 ISO/TC 215 Health informatics (<https://www.iso.org/committee/54960.html>)에서 활발히 표준이 개발되고 있다.

[그림 3] ISO/TC 215 홈페이지

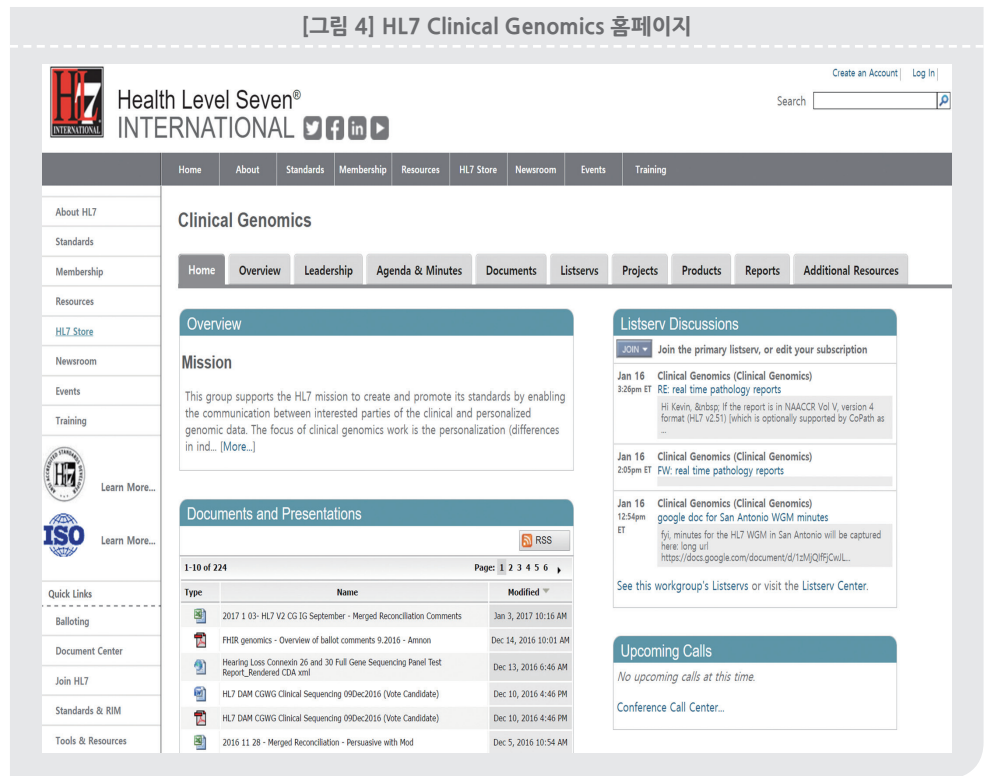


“해석 및 검사 보고서 생성 과정의 표준의 하나인 ISO/TS 20428 표준은 NGS 시퀀싱 결과를 전자의 무기록 (EMR, Electronic Medical Records)에 표현하기 위한 서식지 표준으로 한국에서 주도적으로 개발되었으며 임상유전체검사의 결과를 임상 현장에서 활용하기 위해서 가장 중요한 표준 중 하나라고 할 수 있다.”

다만, ISO/TC 215에서는 Health IT 전반에 대해서 표준을 개발하고 있어 현재 180개의 표준이 발표되어 있으나, 유전체와 관련된 표준은 현재 2개만 발표되어 있다. 2009년에 발표된 ISO 25720 Health informatics - Genomic Sequence Variation Markup Language (GSVML) (<https://www.iso.org/standard/43182.html>)과 2017년에 발표된 ISO/TS 20428 Health informatics - Data elements and their metadata for describing structured clinical genomic sequence information in electronic health records (<https://www.iso.org/standard/67981.html>)이다. ISO/TS 20428 표준은 한국에서 주도적으로 개발된 표준으로 NGS 기술을 이용하여 환자 검체를 NGS로 시퀀싱하는 유전체 검사의 결과를 전자의무기록 (EMR, Electronic Medical Records)에 표현하기 위한 서식지 표준을 제안하는 표준이다. 해당 표준의 경우 국내 여러 기업체 및 병원들에서 도입을 하기 위한 프로그램을 개발하고 있는 중이다. 임상유전체검사의 결과를 임상 현장에서 활용하기 위해서 가장 중요한 표준 중 하나라고 할 수 있다. ISO/TS 20428 발표 이후 한국에서 다시 ISO/PWI 22693 Health informatics - Structured clinical gene fusion report in EHRs를 제안하여 DNA 변이만을 표현하는 20428 표준을 확장하여 RNA 변이, 그 중에서도 융합유전자 (fusion gene) 검사 보고서의 표준화를 시작하였다.

ISO/TC 215 이외에도 미국의 표준개발기구 중 하나인 HL7 (Health Level Seven International)에서도 관련한 표준들을 활발히 개발하고 있다. 특히 HL7은 임상유전체 워킹그룹 (Clinical Genomics Working Group, <http://www.hl7.org/Special/committees/clingenomics/>)이 구성되어 있어서 미국의 관련 전문가들이 많이 참여하고 다양한 표준안들을 체계적으로 개발하고 있다.

[그림 4] HL7 Clinical Genomics 홈페이지

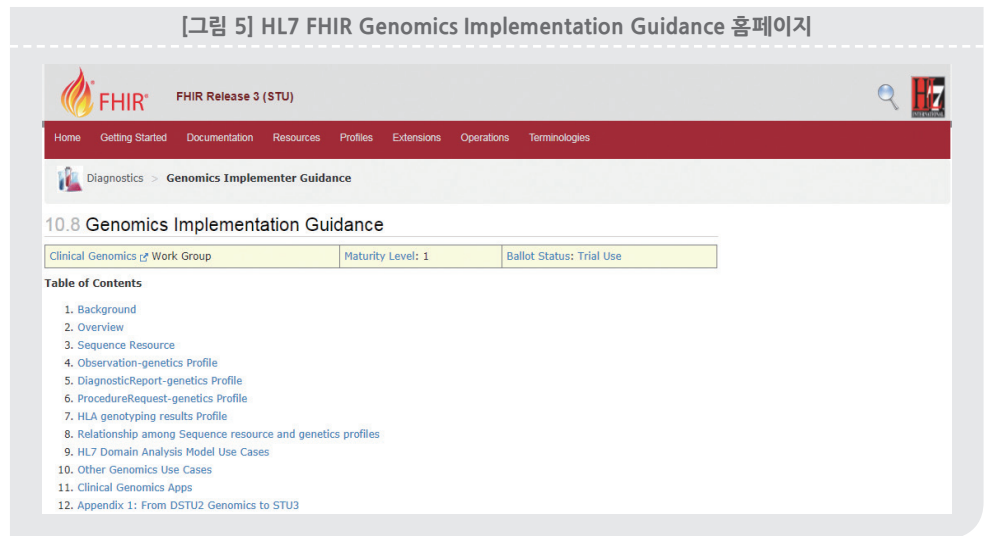


“미국의 표준개발기구 중 하나인 HL7 (Health Level Seven International)에 구성된 임상유전체 워킹그룹 (Clinical Genomics Working Group)은 IT가 관여하는 모든 영역에 대해서 표준화를 하는 노력을 진행하고 있다.”

특히, HL7 임상유전체 워킹그룹에서는 IT가 관여하는 모든 영역에 대해서 표준화를 하는 노력을 진행하고 있는데, 그림 1의 (A) ~ (D)에 해당하는 모든 영역에 대해서 관여를 하고 있다. 예를 들어 (A)의 경우에는 CLIA workflow support, Laboratory information management system 인터페이스 표준화 등이 있고, (C) 과정에서는 구조화된 서식지 (ISO/TS 20428과 유사)를 포함하여 가계도 등을 만드는 노력을 현재 활발히 진행하고 있다. 현재 5개 표준문건이 최종적으로 출판되어 있는데, 내부적으로 개발되는 문서들은 아주 많다. 가장 주목해야 할 것은 2017년 1월에 출판된 HL7 Domain Analysis Model: Clinical Sequencing 문서(http://www.hl7.org/implement/standards/product_brief.cfm?product_id=446)로 임상 유전체와 관련된 상세한 사용 사례(use case)를 잘 정리하고 있다.

그리고, HL7의 차세대 표준으로 개발되고 있는 FHIR (Fast Healthcare Interoperability Resources) (<https://www.hl7.org/fhir/>)에서는 HL7 FHIR Genomics Implementation Guidance(<https://www.hl7.org/fhir/genomics.html>)을 제공하고 있다. HL7 FHIR는 최근 각광을 받고 있는 웹프로그래밍 기법을 도입하여 모든 표준과 관련 자원(Resource)들을 JSON이나 XML으로 제공하고 있기 때문에 개발자들이 손쉽게 구현을 할 수 있다는 큰 장점을 가지고 있다. 해당 리소스를 이용한 구현 사례들은 <https://projects.iq.harvard.edu/fhirgenomics>에서 발견할 수 있다.

[그림 5] HL7 FHIR Genomics Implementation Guidance 홈페이지



미국 All of Us 프로그램이 기술적 측면의 문제를 해결하기 위해 운영 중인 파일럿 프로그램의 하나인 Sync for Genes는 FHIR Genomics를 구축함으로써 유전체 정보 공유를 표준화하는 것을 목표로 하고 있다.⁴⁾ Sync for Genes는 HL7의 FHIR을 데이터 교환의 기반으로 사용해 유전체 정보의 공유를 가능하게 하고 궁극적으로 임상유전체를 진료 현장(point-of-care)에 통합시키는 것을 목표로 하고 있다. 이를 위해 연구실, 제공자, 정부, 개발자, 환자가 참여하는 파일럿 프로그램을 통해 각 파일럿의 내부 포맷을 FHIR Genomics 포맷으로 변환해 공유 및 교환, 활용하는 방법을 개발하고 있다.

6. 데이터 전송 및 공유

마지막으로 그림 1 (D)에 해당되는 데이터 전송 및 공유를 위한 표준들도 개발되고 있다. 앞 장에서 소개한 GSVMML 표준이 대표적이다. GSVMML은 유전정보를 XML로 표현하여 HL7 version 3를 이용하여 정보 전송을 지원하는 표준이다. 다만 2009년에 발표되어 NGS 데이터는 취급하지 못한다는 한계가 있다. GSVMML은 일본에서 주도적으로 만든 표준으로 현재 이를 확장하여 ISO/NP 21393 Health informatics - Omics Markup Language (OML) (<https://www.iso.org/standard/70855.html>), ISO/AWI 25720 Health informatics - Whole genome sequence markup language (WGML) (<https://www.iso.org/standard/70856.html>)을 개발 시작하였다.

이 외에도 ISO/TC 215에서는 유전체 데이터 공유를 위한 표준도 개발하기 시작했다. 해당 표준은 한국에서 제안한 표준후보안으로, 영상정보(CT/MR 이미지)를 저장하고 전송하기 위해 DICOM (Digital Imaging and Communications in Medicine) 표준이 있는 것처럼 NGS 결과를 공유하기 위해서 어떤 데이터들을 어떤 방식으로 표현하여 저장하고 전송할 것인가에 대한 내용을 개발하고자 시작했다.

4) 미국 정밀의료이니셔티브에서 운영중인 두 개의 파일럿 프로그램에는 광범위한 의료기록 혹은 임상 데이터의 통합을 추구하는 Sync for Science와 유전체 데이터 공유를 추구하는 Sync for Genes이 있다.

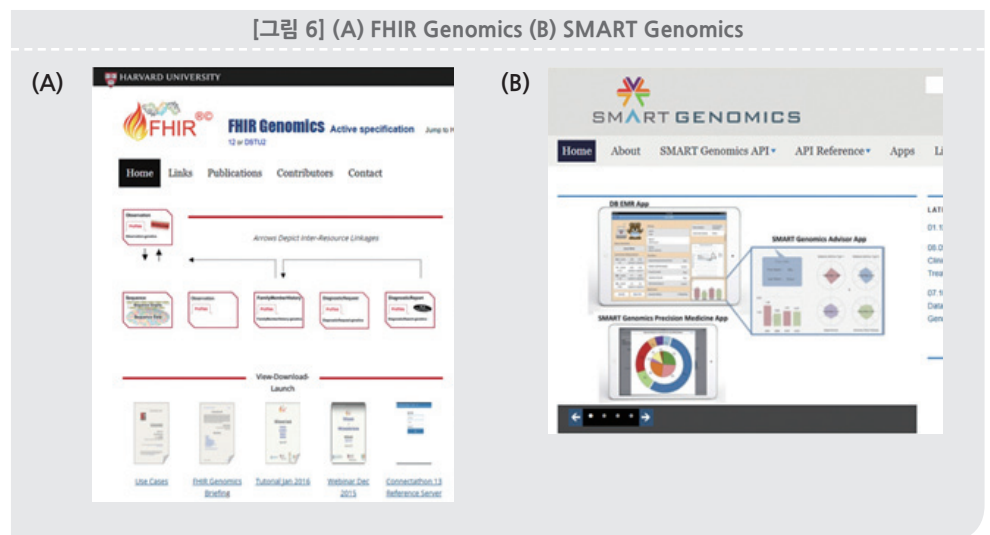
ISO/TC 215외에도 ISO/IEC JTC 1/SC 29/WG 11은 ISO/TC 276/WG 5와 협력하여 영상 정보를 압축하기 위한 mpeg 표준을 이용하여 NGS 결과를 압축하기 위한 표준안을 개발하고 있다(<http://genomsys.com/project/iso-standardization/>). 현재 ISO/IEC CD 23902-

1 Information technology -- Genomic Information Representation -- Part 1: Transport and Storage of Genomic Information (<https://www.iso.org/standard/57795.html>), ISO/IEC CD 23902-2 Information technology -- Genomic Information Representation -- Part 2: Coding of Genomic Information (<https://www.iso.org/standard/73536.html>) 등이 개발되고 있다.

그 외에도 임상시험 결과를 보고 하기 위한 전자문서 표준을 개발하는 CDISC (Clinical Data Interchange Standards Consortium) (<https://www.cdisc.org/>)에서도 pharmacogenomics/genetics 결과를 표준화하기 위한 CDISC Study Data Tabulation Model Implementation Guide: Pharmacogenomics/Genetics (<https://www.cdisc.org/standards/foundational/pharmacogenomicsgenetics-pgx/sdtmig-pgx-v10>)을 발표하였고, 2016년 12월부터 US FDA와 일본에서는 해당 표준을 이용하여 관련 임상시험 결과를 보고 하도록 의무화되었다.

7. 구현 사례

이런 표준안들의 실제 구현 사례로는 앞에서 소개한 하버드대학의 FHIR Genomics (<https://projects.iq.harvard.edu/fhirgenomics>)외에도 역시 하버드대학의 SMART Genomics (<https://projects.iq.harvard.edu/smartgenomics/home>)가 대표적이다. SMART Genomics는 SMART on FHIR 프로젝트의 일환으로 유전체 정보를 활용하기 위한 모바일 앱을 개발하고 있다. 현재 Genomics Advisor, DBEMR, Precision Medicine, Variant Mapper 등의 앱이 개발되고 있다.



8. 향후 로드맵

여러 표준 기구에서 유전체 표준들이 활발히 개발되기 시작하고 있는데, 좀 더 체계적인 유전체 표준 로드맵 개발에 대한 요구가 있어 ISO/TC 215 Working Group 2 주도로 유전체 표준 개발에 대해서 공동으로 협의하기로 하였다. ISO/TC 215/WG 2, HL7 임상유전체 워킹그룹, 대

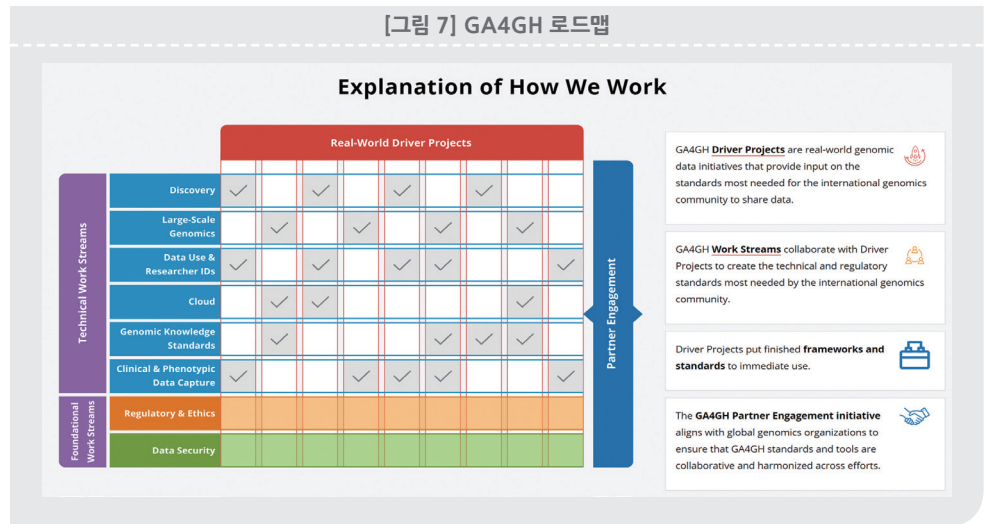
표적인 유전체 연구 컨소시엄인 Global Alliance for Genomics & Health (GA4GH) (<https://www.ga4gh.org/>), 보건의료용어 표준인 SNOMED (<https://www.snomed.org/>)와 협력을 하기로 하였다.

“좀 더 체계적인 유전체 표준 로드맵 개발에 대한 요구가 있어 ISO/TC 215 Working Group 2 주도로 HL7 임상유전체 워킹그룹, 대표적인 유전체 연구 컨소시엄인 Global Alliance for Genomics & Health (GA4GH), 보건의료용어 표준인 SNOMED는 유전체 표준 개발에 대해서 공동으로 협의하기로 하였다.”

특히 GA4GH에 주목할 필요가 있는데, GA4GH는 2013년 9개국 50명의 유전체 연구 및 의료계 관계자들이 모여 구성되어 현재 전세계 71개국의 500여 기관 및 2,000여명이 가입한 대형 기구다. 인권의 틀을 바탕으로 유전체 데이터의 책임 있는 공유를 실현하고 이를 통한 다양한 관련 연구 협력 활성화를 위해 정책을 고안하고 표준설정을 위해 개발 및 시험을 수행하는 단체다. GA4GH는 유전체 연구자들의 자발적인 참여를 바탕으로 다양한 프로젝트를 진행하고 있기 때문에, 일종의 사실상 표준이라고 할 수 있다. GA4GH의 활동체계는 실제 유전체 데이터를 생성하고 이용하는 이니셔티브인 Driver Projects와 이와 협력해 주요 부문 관련한 기술 및 규제 표준을 만들어내는 Work Streams로 구성된다. 2017년 기준 15개의 국제 이니셔티브가 운영 중이다. Work Streams은 다시 각 부문 혹은 프로젝트의 규제 및 윤리와 데이터 보안을 담당하는 Foundational Work Streams와 주요 기술 부문으로 나누어진 Technical Work Streams로 구성된다. 매년 GA4GH Plenary Meeting을 개최하는 GA4GH는 지난 10월 5번째 미팅에서 GA4GH Connect라는 향후 5년간의 전략계획을 발표하면서 각 Work Stream의 목표와 표준 현황 및 향후 활동 방안을 제시했다. 동 회의에서는 Driver Project의 하나인 EMBL-EBI archives: ENA / EVA / EGA에서 개발된 새로운 표준인 “htsget”이라는 데이터 추출 인터페이스(data-retrieval interface)를 발표했다.⁵⁾

5) 그동안 사용자들은 데이터 세트를 통째로 다운로드 해야 해서 작업속도도 느리고 자원이 많이 들어갔는데 새로운 표준인 “htsget”을 사용하면 게놈 중 관심있는 하위 부분만을 다운로드할 수 있다. <https://www.ga4gh.org/docs/GA4GH-Connect-A-5-year-Strategic-Plan.pdf>; <https://www.ga4gh.org/news/sAhZCeJjS96QHhVPIYwwWA.article>; <http://www.frontlinegenomics.com/news/15583/ga4gh-unveil-new-five-year-plan/>

[그림 7] GA4GH 로드맵



[표 1] GA4GH의 Technical Work Streams⁶⁾

부문	목표
발견(Discovery)	일원화된 데이터 발견 플랫폼을 개발해 전세계적으로 유전체 및 임상 데이터의 접근과 활용을 가속화함
대규모 유전체 (Large Scale Genomics)	대규모 유전체 데이터에 접근하는 표준화된 방법들을 개발함
데이터 사용과 연구자 신원 (Data Use & Researcher Identities, DURl)	연구자 및 휴대가능 전자신원을 제공할 수 있는 연구자 신원 제공 원을 정의함으로써 연구자 신원의 조화(harmonization)을 촉진 하고 가능하게 함
클라우드(Cloud)	GA5GH의 Driver Project들이 최신의 클라우드 환경을 최대한 이용할 수 있도록 도움
유전체 지식 표준(Genomic Knowledge Standards)	표준기반 요소들을 개발, 채택, 순응시켜 공통의 APIs(Application Programming Interfaces)를 통해 참고할 유전체정보(reference genomic information)의 교환을 가능하게 하고 유전체 데이터의 차후 분석을 가능하게 함
임상 및 표현형 데이터 수집 (Clinical & Phenotypic Data Capture)	유전체 의료와 연구에 사용 가능하도록 (예를 들어 전자의료시스템과 연구 간 정보를 수집 및 교환할 수 있도록) 임상 표현형을 기술(describe)하는 표준 및 정보 모델을 확립함으로써 유전체가 임상에서 채택되는 것을 지원함

이런 기관들과의 협력을 위해 잠정적으로 논의된 내용은 다음과 같다. 이런 기본적인 사항을 바탕으로 2018년 1월 HL7 미팅에서 좀 더 상세한 논의를 할 예정이고, 해당 결과를 2018년 5월 ISO/TC 215 미팅에서 발표할 계획이다.

[표 2] 유전체 데이터 표준 후보

Scope	- Clinical Sequencing using NGS
Use cases	- HL7 DAM: Clinical sequencing - GA4GH use cases
Action items	- Data QC - Data representation - Data exchange - Data model for research - Clinical template for EHRs - Data infrastructure (cloud) - Privacy & security - Regulation

9. 맺음말

정밀의료(precision medicine) 및 디지털 헬스케어가 각광을 받으면서 헬스케어 데이터를 분석, 통합, 관리하는 산업부문이 빠르게 성장하고 있어 유전체 데이터 표준화 기술의 적극적 개발을 통해 빠르게 시장에 진입하고 선점하는 노력이 필요하다. ICT 소프트웨어와 솔루션을 중심으로 한 헬스케어 IT 시장은 올 해 1,150억 달러에 이르고 그 중 아시아-태평양 시장이 약 10%

6) <https://www.ga4gh.org/howwework/workstreams/>

“국내 유전체, 임상데이터 등 많은 데이터가 생산되고 있는 가운데 이를 활용하여 시장을 개척하기 위해서는 국내의 많은 전문가들이 참여하여 표준 개발을 주도하는 것이 중요하고 이에 대한 관심과 지원이 필요할 것으로 생각된다.”

정도를 차지할 것으로 전망된다.⁷⁾ AI 활용 확대와 ICT 기업의 헬스케어 진출 확대와 함께 헬스 데이터 통합 (health data integration) 플랫폼의 확대가 주요 변화로 예측되고 있다. 헬스케어 데이터 분석 시장 또한 빠른 성장이 전망되고 있으며 그 중 의료 및 임상 데이터 분석을 통한 환자의 진료를 개선하는 임상 분석(clinical analytics) 부문의 중요성이 대두되고 있다.⁸⁾ 이는 헬스케어 데이터 분석 시장 중 의료비 지불자(payer) 및 병원(hospital)과 함께 진료(physician practice) 시장의 빠른 성장에서 엿볼 수 있다. 미국의 경우 2015년 전체 헬스케어 데이터 분석 시장이 38억7천8백만 달러였으며 그 중 진료 시장(약 3억2천8백만 달러, 전체 시장의 8.5%)은 재정 한계, 의료진 관심 부족, 낡은 IT 인프라 등의 요인으로 인해 병원 시장 (약 22억6천1백만 달러, 전체시장의 58.3%)와 의료비 지불자 시장 (약 12억8천9백만 달러, 전체시장의 33.2%)에 비해 규모가 작았다. 하지만 진료 시장은 20.2%의 가장 빠른 연평균성장률(병원 13.7%, 의료비 지불자 15.6%)로 확대되어 2020년에는 전체시장의 약 11%(병원 약 55%, 의료비 지불자 약 34%)를 차지할 것으로 전망되고 있다.

정밀의료 등의 현실화를 위해서 유전체 연구가 활발히 진행되고 있는 상황에서 연구 활성화 및 임상 적용을 위해서 유전체 데이터의 표준화는 반드시 필요한 상황이다. 지금까지 소개한 바와 같이 유전체 데이터의 표준은 이제 막 시작한 단계로, 한국이 여러 중요 표준들을 주도적으로 제안하고 개발하고 있는 것은 다행으로 생각된다. 국내 유전체, 임상데이터 등 많은 데이터가 생산되고 있는 가운데 이를 활용하여 시장을 개척하기 위해서는 국내의 많은 전문가들이 참여하여 표준 개발을 주도하는 것이 중요하고 이에 대한 관심과 지원이 필요할 것으로 생각된다.

7) Frost & Sullivan, 2018
 8) Frost & Sullivan, 2017. 헬스케어 데이터 분석 시장에는 임상 분석 외 재정 분석(financial analytics)와 운영 분석(operational analytics) 부문이 있다.

February 2018. Issue 8

저자소개

신수용

경희대학교 컴퓨터공학과 조교수
e-mail : sooyong.shin@gmail.com

김지현

한국바이오협회 한국바이오경제연구센터 선임연구원
전화 : 031-628-0013
e-mail : jkim@koreabio.org

이계민

한국바이오협회 한국바이오경제연구센터 선임연구원
e-mail : drkeimin@gmail.com

BIO ECONOMY REPORT

발행 | 2018년 2월

발행인 | 유승준

발행처 | 한국바이오협회 한국바이오경제연구센터
13488 경기도 성남시 분당구 대왕판교로 700
(삼평동, 코리아바이오파크) C동 1층
www.koreabio.or.kr



한국바이오경제연구센터
KOREA BIO-ECONOMY RESEARCH CENTER

Innovating Data Into Strategy & Business



9 772508 682002

ISSN 2508-6820