

유전체정보기반 정밀의료 발전방향

Development of Genetic Information-Based Precision Medical Care



»»» 위원장

방영주 서울대학교 의과대학 명예교수(한국과학기술한림원 의약학부 정회원)

김열홍 고려대학교 의과대학 교수

»»» 위 원

박용양 성균관대학교 의과대학 교수

신수용 성균관대학교 삼성융합의과학원 교수

심효섭 연세대학교 의과대학 부교수

이범희 울산대학교 의과대학 부교수

이원복 이화여자대학교 법학전문대학원 교수

최종일 고려대학교 의과대학 교수

목차



I. 현황 및 미래기술

1. 서론	2
2. 정밀의료의 장점	5
3. 정밀의료의 실현을 위해 극복해야 할 과제	7
4. 산업적 가치에 대한 고찰	8

II. 유전체 염기서열 분석의 원리와 미래기술

1. 인간 유전체 염기서열 정보 분석이 왜 중요한가? (서론)	16
2. 유전체 분석기술의 핵심은 무엇인가?	17
3. 유전체정보분석의 미래는?	25
4. 유전체정보분석의 시대에 무엇을 할 것인가? (결론)	29

III. 유전체 염기서열 분석 데이터의 임상응용

1. 종양질환	34
가. 폐암	34
나. 폐암 이외의 고형암	49
2. 유전성 질환	65
가. 희귀 유전질환	65
나. 유전성 심장질환	79

IV. 빅데이터 활용의 신기술

1. 인공지능 기반의 유전체 분석을 통한 신약 후보물질 발굴: 딥지노믹스	102
2. IBM 왓슨을 이용한 유전체 변이 해석 연구	104
3. 개인 유전 정보 빅데이터를 이용한 신약 개발 활용 사례	107
4. 유전체 및 진료기록, 라이프로그 등을 통합한 빅데이터 연구	110

V. 개정 개인정보 보호법의 가명정보 제도와 유전체 연구

1. 개인정보 보호법에 도입된 “가명정보”란?	126
2. 유전체 연구에 미치는 영향	128
3. 유전체 연구와 관련하여 남아 있는 쟁점	132
4. 유럽연합과의 비교	137
5. 결론	141

목차



표목차

〈표 2.1〉 NGS 장비에 대한 분석성능 비교	19
〈표 2.2〉 유전체정보별 분석기준과 데이터양	20
〈표 2.3〉 국가별 대규모 유전체 빅데이터 프로젝트	27
〈표 3.1〉 폐암에서 발견되는 유전자 변이 및 해당 표적 항암제	36
〈표 3.2〉 미국 FDA와 유럽 EMA 승인을 받은 표적항암제와 타겟 유전자 변이 진단법	50
〈표 3.3〉 대표적인 암억제유전자와 가족성 암증후군	52
〈표 3.4〉 대표적인 암 관련 BRAF 돌연변이	54
〈표 3.5〉 각종 암에서의 RAS isoform 돌연변이의 빈도	57
〈표 3.6〉 선택적 KRAS 억제제 개발 현황	58
〈표 4.1〉 인간 전문가와 왓슨의 암 유전체 분석 결과 비교	105
〈표 4.2〉 All-of-Us에서 수집하는 데이터의 종류	112
〈표 4.3〉 All-of-Us의 세부적인 목표 및 타임라인	114
〈표 4.4〉 코치의 지도를 받으면서 유의미하게 개선된 수치들	120

그림목차

[그림 1.1] 폐선암의 돌연변이	4
[그림 2.1] 유전체 염기서열 정보분석 워크플로우	18
[그림 2.2] NGS 데이터로부터 원인 돌연변이를 찾기 위해 필요한 생물정보분석 워크플로우	22
[그림 2.3] 싱글셀 유전체정보 분석 단계	25
[그림 2.4] 임상-유전체 통합 빅데이터 분석 워크플로우	30
[그림 3.1] 폐암의 진단 및 유전체 분석 모식도	37
[그림 3.2] 유전체 분석에 근거한 표적 항암제 치료 반응 1의 예	38
[그림 3.3] 유전체 염기서열 분석 데이터의 적용 분야	46
[그림 3.4] 유전체정보기반 정밀의료에서의 의료서비스	65
[그림 3.5] 유전체 분석 방법	67
[그림 3.6] 심방세동 환자 GWAS 연구에서 발견된 9개 유전자 위치 메타분석 Manhattan plot (좌); 단일염기다형성에 따른 발작성 심방세동 전극도자절제술 재발 예측 (우)	80
[그림 3.7] 돌연심장마비 또는 급사의 주요 원인	81
[그림 3.8] 우리나라 유전성 부정맥 연령별 발병률(건강보험 표준 코호트)(좌); 원인 질환 다기관 레지스트리 연구 결과(우)	83
[그림 3.9] 이온채널병증의 유전자 변이 및 변이 위치에 따른 질환 중증도	84
[그림 3.10] 유전성 심장질환에서의 유전적 변이와 질병 발현	85
[그림 3.11] 유전성 부정맥에서의 이온통로 기능 이상 분자생물학적 기전	87
[그림 3.12] 유전성 부정맥 표적 유전자 검사의 진단율과 표현형과의 일치율	88
[그림 3.13] 심혈관질환 NGS 기반 유전자 검사 패널 예(The TruSight Cardio Sequencing Kit, Illumina)	90
[그림 3.14] 유전성 심장질환 클리닉(Cardiovascular genetics clinic)에서의 유전성 부정맥 진단 임상 프로세스	91
[그림 3.15] 이온채널병증에서의 arrhythmogenicity 측정을 통한 pathogenicity 평가를 위한 세포패치클램프 기능평가(functional test)	92
[그림 3.16] 유전체/유전자 데이터의 임상적 정보로의 적용 플랫폼	94
[그림 4.1] 딥러닝에 기반한 방법으로 RNA스플라이싱에 관여하는 변이 찾기	103
[그림 4.2] UNCSeg의 결과와 왓슨 포 지노믹스의 결과 비교	106
[그림 4.3] 23andMe 고객의 폭발적인 증가	107
[그림 4.4] All-of-Us 프로젝트의 전반적 구조	113
[그림 4.5] 구글 베이스라인 프로젝트 앱	116
[그림 4.6] 수천 가지의 데이터를 9개월 동안, 3번에 걸쳐 측정한 연구	117
[그림 4.7] 유전체를 포함하여 방대한 데이터를 분석하여 발견한 상관관계	119

I. 현황 및 미래기술





현황 및 미래기술

1 서론

2015년 1월 당시 미국 대통령 오바마는 연두교서에서 정밀의료(precision medicine)에 관한 대규모 연구를 추진할 것을 천명하였는데 이는 세계적으로 커다란 반향을 불러왔고, 이후 정밀의료는 의료가 지향하는 궁극적인 목표로 자리 잡게 되었다. 정밀의료란 각 환자의 유전체(genomics), 환경 및 생활습관의 차이에 따라 서로 달라지는 질병의 특성에 따라 질병을 예방하거나 치료함을 목적으로 하는 것이다. 특히, 유전체적 차이는 개인마다 서로 다른 질병의 발생이나 진행, 치료에 대한 서로 다른 치료효과 등을 설명하는데 필수적인 요인임을 알게 되었기에, 보다 효율적인 유전체분석과 유전체정보의 광범위한 활용은 학문적으로나 산업적으로 매우 중요한 과제가 되었다.

전통적으로 우리는 임상적인 특징, 실험실 검사결과, 병리조직학적 소견 등에 근거하여 특정 질환을 정의하고 있다. 예컨대, 고혈압은 수축기 혈압이 140mmHg이 넘거나 이완기 혈압이 90mmHg 이상인 경우로 정의되고, 당뇨병은 공복 시 혈당이나 당화 혈색소(HbA1C) 수치에 의해 진단되며, 암은 조직검사에 이은 병리조직학적 소견에 따라 진단된다. 과거에는 환자들이 이러한 방법으로 특정 질환으로 진단되면, 모든 환자들을 동일하게 접근하고 치료하였다. 그러나 각각의 질환은 서로 다른(heterogenous) 여러 질환의 집단으로 이루어졌고, 환자들마다 서로 다른 접근과 치료가 필요하다는 사실을 알게 되었고, 이러한 인식에서 정밀의료의 개념이 싹트게 되었다. 이러한 이질성(heterogeneity)에 대한 인식과 정밀의료의 적용은 암의 진료에서 출발하였지만, 이제는 거의 모든 질환에도 보편적으로 적용된다. 가령 당뇨병의 경우에도 환자마다 특정 약제에 대한 효과나 부작용이 동일하지 않고, 또한 합병증의 발생 시기나 종류도 서로 다른 것이다. 요컨대 우리가 현재 하나의 질환으로 부르는 것들이 실제로는 서로 다른 여러 아형(type)의 질환이 섞여 있는 것이며, 보다 나은 환자의 진료를 위해서는 이러한 아형으로 진단되고 치료되어야만 한다는

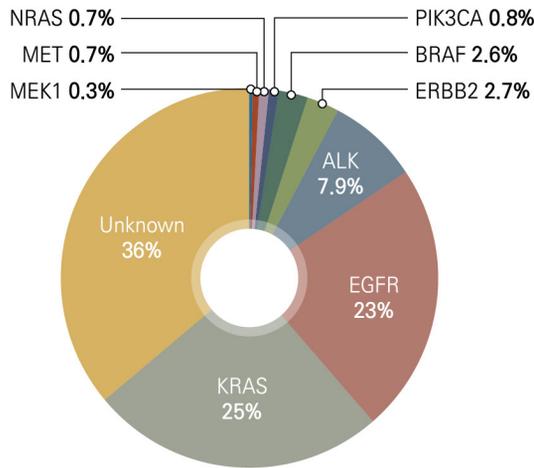
것이며, 이러한 관점에서 정밀의료는 질환의 세분화(segmentation)로 정의할 수도 있는 것이다. 과거에는 비슷한 개념으로 사용했던 맞춤형의료(personalized medicine)는 환자 개개인마다 서로 다른 치료를 한다는 의미를 내포하고 있어서, 최근에는 잘 사용되지는 않게 되었으나, 내용적으로는 매우 비슷하다.

정밀의료의 핵심은 각개인의 유전체정보가 된다. 지난 30여년에 걸친 유전체분석의 엄청난 발전이 정밀의료의 개념을 싹틔우고, 정밀의료의 구현을 이끌게 한 것이다. 1990년에 시작된 인간게놈프로젝트(Human Genome Project)가 2003년에 완료됨으로써 30억 쌍으로 이루어진 인간게놈의 모든 염기서열이 밝혀졌다. 이후 개발된 차세대염기서열분석(next generation sequencing, NGS) 기술덕분에 각개인의 유전체분석이 아주 빠르게, 비교적 저렴한 가격에 가능하게 되었다. 현재 많이 사용되고 있는 NGS 방법이나 그 기술의 미래발전 전망은 이 보고서의 다른 장에서 상세히 기술된다. 둘째로, 단백질학(proteomics), 대사체학(metabolomics), 마이크로바이옴(microbiomics) 등의 여러 연구는 다양한 생체정보를 제공하고, 이를 통해 각 개인이 노출되는 서로 다른 환경에 대한 생체 반응을 분석할 수 있게 되었고 이러한 생체반응이 질환의 발생과 진행에 어떠한 작용을 하는지를 이해하게 해주었다. 셋째로, 모바일 건강정보기술(mobile health technology)의 발전으로 각 개인의 생활환경이나 생체반응 등을 실시간 모니터링 할 수 있게 된 것도 생활환경이나 활동 등이 질환의 발생에 미치는 영향을 분석하는 데에 크게 도움이 될 것이다. 이와 같이 각각의 개인에 대해 아주 다양한 정보와 장기간에 걸친 의료정보를 같이 처리하고 분석해야만 우리가 원하는 정밀의료에 관한 정보와 지식을 얻을 수 있다. 컴퓨터기술의 발전이 없었다면 상상할 수 없을 만큼 많은 양의 정보를 보관, 처리하고 분석하는 것은 불가능한 일이다. 여기에 더해 이러한 정보를 해석하고 활용하기 위해서는 생물정보 분석 기술이나 빅데이터 분석기술 등의 발전이 필연적으로 요구되었고, 실제로 엄청나게 빠른 속도로 발전되었다. 빅데이터의 활용은 다른 분야 이상으로 의료분야에서 매우 중요하기에 다른 장에서 상세히 다룬다.

정밀의료는 언젠가는 인간의 모든 질병을 예방하고 치료하는데 활용될 것으로 기대되지만, 현재까지는 주로 암 환자의 진료에 활용되어 왔다. 암은 여러 유전자 변이가 축적되어 생기는 질환이고, 일부 암에서는 특정 유전자의 변이가 발암과정에 핵심적인 역할을 하기에, 이러한 드라이버 암유전자(driver oncogene)의 단백질의 기능을 억제하는 항체 또는 화학물질(small molecule)들이 표적치료제(molecularly targeted agent)로 개발되어

널리 사용되고 있다. 대표적으로 비소세포폐암(non-small cell lung cancer, NSCLC)을 들 수 있으며, <그림 1.1>에서 보는 바와 같이 폐선암(lung adenocarcinoma)은 서로 다른 드라이버 암유전자가 유전자융합(gene fusion), 돌연변이(mutation) 및 유전자증폭(gene amplification) 등에 의해 활성화되어 발생하게 되는데, 이들의 임상적 특성이나 진행속도 등은 서로 다르며, 서로 다른 표적치료제가 사용되게 된다. 최근에 가장 큰 주목을 받고 있는 면역항암제(immuno-oncology drugs)의 경우에는 암조직의 MSI(microsatellite instability) 여부나 PD-L1 발현과 같은 바이오마커(biomarker)에 따라 치료율이 크게 달라지기에 결국 바이오마커의 유무에 따라 암을 세분화할 수 있고, 이 또한 정밀의료가 지향하는 목표가 되고 있다. 때로 유전자정보는 암의 예방에도 활용되고 있다. BRCA(breast cancer susceptibility gene)이나 TP53 등의 암감수성유전자(cancer susceptibility gene)의 생식돌연변이(germline mutation)를 지닌 사람들은 특정 암들의 발생빈도가 매우 높기에 암의 예방이나 조기진단을 위해 정기적인 검진을 하거나 표적장기의 예방적 수술 등을 하고 있다.

그림 1.1 폐선암의 돌연변이



자료: Sholl, L. M. *et al.*, 2015

2 정밀의료의 장점

전통적 의료에 비교하여 정밀의료의 장점은 세 가지로 나누어볼 수 있다. 우선 환자들에게 보다 나은 치료를 제공할 수가 있다. “The right drug for the right patient at the right time”이 정밀의료를 가장 잘 표현한 어구로 각각의 환자에 있어서 가장 적절한 시점에, 가장 효과가 있을 약제를 선택하여 치료한다는 개념이다. 이러한 목표가 가장 먼저 현실화 된 것이 앞에서 기술하였듯이 암 환자에 대한 표적치료제의 사용이다. 폐선암의 경우 <그림 1.1>과 같이 서로 다른 드라이버 암유전자의 발현에 따라 여러 아형으로 나뉘어지며, 그 치료가 서로 달라진다. 폐암세포가 EGFR(epidermal growth factor receptor) 유전자의 돌연변이를 지녔다면, gefitinib이나 erlotinib과 같은 EGFR-TKI(tyrosine kinase inhibitor)가 특효약이 되며, ALK(anaplastic lymphoma receptor tyrosine kinase) 유전자의 융합이 있다면 crizotinib, ceritinib 등의 ALK-TKI로 치료하게 된다. 이러한 표적치료제는 대상환자들의 70% 이상에서 치료효과를 보일 뿐만 아니라, 그 부작용이 세포독성항암제에 비해 훨씬 적기에 환자의 삶은 질도 훨씬 좋다. 반면에, 이러한 드라이버 암유전자의 발현이 없는 환자에서는 효과를 전혀 기대할 수 없는 표적치료제의 사용으로 인한 부작용이나 비싼 비용을 피할 수 있고, 불필요한 시간의 허비도 막을 수가 있는 것이다. 미래에는 정밀의료가 암 외에 여러 만성질환의 진단과 치료에도 폭넓게 활용될 것으로 기대되고 있다. 당뇨병의 발병에는 여러 개의 유전자가 관여하고 식이 등의 환경요인 또한 중요하기에 현재의 정보와 지식만으로는 정밀의료의 활용이 많이 제한되지만, 앞으로 다양한 생체정보와 생활환경 등에 관한 정보가 더해지면, 환자의 치료에 아주 유용하게 활용될 것으로 기대되는 것이다. 현재의 의료에서는 모든 환자들에게 동일한 용량의 약제를 사용하는데, 이에 따라 일부 환자는 과도한 부작용을 겪기도 하고, 다른 환자들에서는 치료효과가 떨어지기도 한다. 약물유전체학(pharmacogenomics)에서는 유전체 차이에 따른 약물대사의 차이를 밝힘으로써 환자마다 서로 다른 약제나 약제용량을 선정하고자 하는 노력들을 하고 있으며, 이 역시 보다 나은 치료를 하는 데에 매우 중요한 정밀의료의 분야가 된다. 정밀의료가 암질환이나 희귀유전성질환의 진료에 어떻게 활용되고 있는지는 다른 장에서 보다 상세히 다룰 것이다.

두 번째로 정밀의료의 실현은 사회나 보건의료제도에 커다란 이익을 가져오게 된다. 여러 만성질환의 예방이나 조기진단은 사회구성원들의 건강증진에 기여하고, 결과적으로 사회의 의료비용을 줄이게 된다. 예컨대 가족성 고콜레스테롤혈증(familial cholesterolemia) 환자들은 혈중 콜레스테롤수치가 비정상적으로 높으며, 50세 이전에 심근경색 등 여러 심장질환을 일으킬 수 있는 유전성질환이다. 이 질환은 인구 250명당 1명꼴로 발생하지만, 이들 중 1/6 정도에서만 진단이 이루어지고 있다. NGS를 활용하면 보다 많은 환자를 조기에 진단해 낼 수 있고, 적절한 약물치로나 식이 등을 통해 심장질환의 발생을 방지할 수 있기에, 사회적으로 의료비용 등 많은 비용의 절감을 기대할 수가 있다. 만성질환 환자의 경우에도 개인마다 합병증의 발생위험도를 미리 예측함으로써 그 발생을 막거나 지연할 수가 있고, 또한 검사주기 등을 다르게 하여 삶의 질을 높이고 진료비용을 줄일 수 있게 된다. 정밀의료의 활용으로 환자들에게 가장 유효한 치료를 제공하고, 불필요한 약제 부작용을 피할 수 있기에 진료비용을 줄이고, 입원 수요를 줄여서 의료시설의 보다 효율적인 활용을 가능케 한다. 먼 훗날에는 사람이 태어날 때 유전체분석을 통해 그 사람이 일생에 가지게 될 주요 질환들의 위험도를 미리 예측하는 일이 가능해지고, 이에 따라 사람마다 식이, 생활습관, 또는 생활환경을 다르게 하여 각종 질환의 발생을 억제하는 것이 가능해질 것이다. 유전체정보를 중심으로 한 정밀의료의 발전은 궁극적으로는 치료중심의 의료에서 예방중심의 의료로 패러다임의 대전환을 일으킬 것으로 기대되는 것이다.

마지막으로, 정밀의료의 활용으로 신약개발 등 새로운 의료창출이 보다 효율적으로 이루어지고 있다. 임상시험에서 바이오마커를 활용하여 신약의 치료효과가 높은 환자군을 선정하거나, 반대로 치료효과가 전혀 없거나 과도한 부작용이 발생할 환자군을 제외함으로써 임상시험의 성공 확률을 높이고, 비용과 시간을 줄여서 임상시험의 효율성을 개선할 수 있음이 지난 20여 년간에 걸쳐 이미 증명이 되었다. 예컨대 위암 환자에서 anti-HER2 항체인 trastuzumab의 치료효과를 확인하기 위한 3상 임상시험에서는 HER2 양성 위암 환자만을 등록함으로써 임상시험의 성공을 가져왔다. 다른 한편, 이 약제의 치료효과를 전혀 없을 HER2 음성 위암 환자들을 임상시험에서 제외함으로써 임상시험의 윤리성을 더 높일 수가 있었던 것이다. 이러한 이유로 최근에는 신약개발의 초기단계부터 보다 효과적인 환자선정을 가능하게 할 바이오마커의 개발에 많은 노력을 쏟게 되었다.

3 정밀의료의 실현을 위해 극복해야 할 과제

정밀의료는 이미 암이나 희귀 유전성질환의 진단과 치료에 널리 활용되고 있으며, 미래에는 대부분의 질환의 진료에 있어서 핵심적인 개념이나 플랫폼으로 자리 잡게 될 것임에 틀림없다. 그러나 이러한 정밀의료의 실현되기 위해서는 반드시 극복해야만 할 여러 과제나 우려가 남아 있다. 정밀의료의 구현을 위해서는 아주 많은 사람들의 유전체정보 등 여러 생체정보가 필요하고, 이들의 생활환경과 습관에 관한 정보와 함께 진료기록이나 병력들이 장기간에 걸쳐 수집되고 분석되어야만 한다. 이렇듯이 방대한 자료를 만들고, 장기간 저장하고, 분석하기 위해서는 막대한 인프라(infrastructure)와 비용이 필요하다. 핵심이 되는 유전체정보를 얻기 위한 비용과 시간이 NGS의 발달로 많이 용이해졌다고는 하지만, 수백만 명의 유전체정보를 얻기 위해서는 더 많이 저렴해져야만 한다. 사실 유전체 정보 외에도 각종 생체정보나 라이프로그 정보를 얻기 위해서는 아직도 개발되어야만 할 기술들이 많이 남아 있고, 그 비용 또한 천문학적 수준이 된다. 정밀의료의 발전은 인류의 건강과 삶의 질을 엄청나게 개선시킬 수 있기에, 여러 나라가 힘을 합쳐서 연구하고 자료를 공유한다면 보다 이른 시기에 정밀의료의 완전한 구현이 가능할 것이다. 또한, 이러한 국가 간 협력은 인종 간 차이, 생활환경이나 문화의 차이 등이 질병의 발생에 미치는 어떻게 영향도 동시에 규명하는 데에 큰 도움이 될 것임에 틀림없다. 그러나 정밀의료연구로부터 얻어지는 자료의 엄청난 지적재산권이나 산업적 이익 때문에 이러한 국제협력은 쉽게 이루어지지 않고 있는 것이 현실이다.

정밀의료의 연구에 이용되는 정보들은 각 개인에게는 아주 민감한 정보들이다. 이러한 정보의 사용에 따른 윤리적, 법적 문제가 적지 않고, 또한 정보의 노출 위험이 도사리고 있는 것도 큰 문제이다. 아주 먼 훗날에는 태어날 때 얻어지는 유전체정보에 따라 질병 위험도나 직업 적합성 등을 미리 예측하는 일이 가능해질 수도 있다. 1997년에 제작된 영화 가타카(Gattaca)에서는 유전체정보에 따라 사람의 우열을 평가하고, 소위 열등한 유전체를 지닌 주인공이 우주비행사가 되는 것을 결코 용납하지 않는다. 유전체정보 외에도 라이프로그 정보, 진료기록 등도 모두 민감한 정보이므로 개인의 사생활이나 인권 침해 가능성에 대한 대책이 필요하다. 다른 한편, 정밀의료 연구로부터 얻어지는 새로운 지식이나 기술에 대한 지적재산권의 소유권을 둘러싸고 적지 않은 법적 분쟁이 야기될 수 있다. 이러한 법적 이슈에 대해서는 다른 장에서 논의할 것이다.

4 산업적 가치에 대한 고찰

정밀의료의 구현을 위해서는 많은 산업적 발전이 필요하고, 동시에 정밀医료를 활용하는 여러 유관 산업의 발전이 기대된다. 실제로 2015년에는 세계적인 정밀의료 관련 시장의 규모가 40조 원 정도였는데 2019년에는 60조 원에 이를 것으로 예측되어 연평균 성장률이 10%를 웃돌고 있다. 이러한 추세에 힘입어 2026년에는 시장규모가 120조 원을 넘을 것으로 예측되고 있다. 정밀의료와 관련된 산업들을 열거하면 아래와 같다.

가. 유전자염기분석서열분석(gene sequencing)

정밀의료의 핵심은 유전체정보라고 할 수 있다. 암 환자의 암조직의 유전자분석을 통한 표적치료제의 개발은 정밀의료에 대한 확신을 가지게 했고, NGS 기술을 통하여 보다 빠르고, 보다 저렴하게 많은 사람들의 유전체분석을 할 수 있게 되었기에 빠른 속도로 정밀의료 분야가 발전하게 되었다. 따라서 NGS 기술 관련 산업이 크게 성장하였고, 지금도 보다 적은 시료를 가지고, 보다 정확하게, 보다 빠르게, 보다 저렴하게 유전체분석을 수행하기 위한 새로운 기술개발을 위해 많은 기업들이 엄청난 노력을 경주하고 있다.

암 환자에서 정밀의료의 발전은 암조직의 유전체분석에 근거한다. 그런데 암조직의 채취 또는 조직생검에는 출혈, 감염 등 합병증의 발생위험을 지니고 있으며, 특히 병소가 심부 장기에 있는 경우에는 그 위험도가 더욱 높아진다. 병소가 아주 작거나, 종양 주위에 큰 혈관 등이 있으면 그 위험도가 너무 커져서 조직검사를 아예 할 수가 없다. 한편, 표적 치료제로 치료를 진행하면서 암세포에 새로운 유전자변이가 올 수 있고, 이러한 변이는 약제 저항성의 원인이 되는 경우가 많다. 따라서 치료를 진행하면서 이러한 유전자변이의 출현을 알기 위해서는 반복된 암조직생검이 필요하다. 그러나 조직생검에 따른 위험 때문에 실제로 쉽지 않은 일이다.

이러한 문제점을 해결하기 위하여 근래에는 혈액을 채취하여 혈액 속에 미량 존재하는 암세포나 그 DNA를 이용하여 유전체분석을 하는 액체생검(liquid biopsy)에 대한 연구가 활발히 진행되었고, 이미 일부 임상진료에 활용되고 있다. 액체생검은 조직생검에 따른 위험이 없고, 반복적으로 암세포의 유전체분석을 시행할 수 있다는 장점을 지닌다. 또한

수술이나 방사선치료 등의 치료를 하고 나서 체내에 남아 있을 수 있는 미세잔존질환 (minimal residual disease)을 방사선학적 검사보다 훨씬 일찍 진단해낼 수 있다는 것도 큰 강점이 된다. 현재 액체생검은 그 검사특이도에는 문제가 없으나 아직 검사예민도가 85% 수준이기에 보다 적은 양의 종양세포나 DNA를 검출하여 유전체분석을 할 수 있는 기술개발을 위한 노력이 경주되고 있다.

나. 생물정보학(bioinformatics)

정밀의료의 구현을 위해서는 특정질환을 지닌 아주 많은 수의 환자들의 유전체정보와 함께 각종 omics 정보가 필수적이며, 거기에 이들의 생활환경이나 생활습관에 관한 자료, 그리고 장기간의 의학적 자료 등 방대하고 복잡한 생물학적 자료를 분석하고 해석해야만 하기에 생물학, 컴퓨터과학, 정보공학, 수학, 통계학 등 여러 학문을 이해하고 활용할 수 있는 생물정보학의 발전이 필요하게 되었다. 가령 암 환자에서 특정 유전자변이를 찾아낸 경우 그 임상적 의미를 해석하여 임상 의에게 필요한 정보를 제공해야만 환자에게 도움이 될 것이다. 따라서 이러한 생물정보 분석을 위한 여러 원천기술이나 소프트웨어의 개발과 실행에 관련된 산업이 급속도로 발전하게 되었다. 2017년에는 그 시장규모가 7조 원 정도였는데, 2025년에는 20조 원에 이를 것으로 예측되고 있다.

다. 빅데이터분석(big data analytics)

앞서 언급되었듯이 엄청나게 방대한 자료, 즉 빅데이터의 저장과 분석, 그리고 해석을 위한 자료산업의 발전과 성장은 필연적으로 요구된다. 특히, 의료에서는 CT scan과 같은 방사선학적 검사나 병리조직학적 검사와 같은 이미지나 unstructured data를 다량 분석해야만 하기에 빅데이터의 처리와 분석이 더 까다롭다. 앞으로 이 분야 기술과 산업의 발전이 시급하다고 할 수 있으며, 이 분야에 대해서는 다른 장에서 상세히 기술하고자 한다.

라. 신약개발

정밀의료를 통한 새로운 치료제 개발의 표적을 발견하고 그 표적의 기능을 억제하거나 향진하는 물질을 개발하는 전략은 신약개발의 가장 중요한 전략이 되었다. 암은 유전자 변이가 축적되어 발생하며, 특정유전자가 발암과정을 주도하는 경우에 이를 드라이버 암 유전자라 칭하는데, 이 유전자의 단백산물을 표적으로 한 표적치료제의 개발은 이미 새로운 항암제 개발의 가장 중요한 전략이 되었다. 암 이외의 다른 질환에서도 유전체분석은 신약 개발의 중요한 출발점이 되고 있다. 예컨대, 가족성 고콜레스테롤혈증을 지닌 한 가족에서는 PCSK9(proprotein convertase subtilisin-kexin type 9) 유전자의 변이가 그 원인임이 밝혀졌는데, 이들 가족의 90%는 관상동맥질환을 발병하고 60%의 가족이 조기사망을 하게 된다. 결국 이 유전자를 표적으로 한 항체들이 개발되었고 이들 약제는 이들 환자들의 치료에 유용하게 쓰이게 되었다.

정밀의료는 신약개발을 위한 표적을 찾아내는 데에 이용될 뿐만 아니라, 그 효과를 검증하기 위한 임상시험에서도 아주 유용하게 활용된다. 즉, 신약개발의 표적은 그 약제의 치료효과를 볼 확률이 높은 환자군을 선별하거나 효과가 없을 환자를 제외하는 바이오마커로 활용될 수가 있는 것이다. 이렇듯이 치료효과를 예측하게 하는 예측성 바이오마커(predictive biomarker)를 지닌 환자만을 등록하는 것을 enrichment trial이라고 하는데, 그 유용성은 이미 널리 입증되었다. 이러한 전략은 임상시험의 비용과 기간을 줄일 뿐 아니라 임상시험의 성공 확률을 높이게 하는 것이다. 최근에는 암의 종류에 상관없이 특정 바이오마커를 지닌 암 환자를 모두 등록하는 바구니 임상시험(basket trial)이 많은 관심을 끌고 있는데, 이미 MSI-H 환자군에서 면역항암제인 pembrolizumab이, 그리고 NTRK fusion이 있는 암 환자에서 TRK 억제제인 larotrectinib이나 entrectinib 등이 이러한 임상시험을 통해 암종류에 상관없이 바이오마커에만 근거하여 시판허가가 가능해졌다.

마. 동반진단기술(companion diagnostics)

동반진단기술은 특정질환에서 특정 약제나 생물학적 제제를 안전하고 효과적으로 사용하기 위하여 필수적인 정보를 제공하는 진단기술로 정의되며, 이를 활용하여 특정

약제의 치료효과를 기대할 수 있는 환자군을 찾아내거나, 특정약제의 사용 시 심각한 부작용이 일어날 수 있는 환자군을 선별해 내거나, 특정약제를 사용함에 있어서 환자의 안전성과 치료효과를 개선할 수 있도록 치료효과를 모니터링하는 데에 활용될 수 있다. 암 환자에서 사용되는 표적치료제들의 경우 환자선정에 활용되는 예측성 바이오마커를 지닌 환자를 찾아내는 데에 이용되는 진단기술이 그 대표적인 예이다. 예컨대 ALK 양성 비소세포폐암에서 ALK 유전자의 융합이 있는 경우 ALK 억제제가 효과가 있음이 밝혀졌는데, 실제로 암 환자에서 ALK 유전자의 융합 여부를 진단할 수 있는 FISH 검사나 면역조직화학검사가 치료제의 승인과 동시에 동반진단기술로 승인받은 것이다. 이러한 동반진단기술의 개발은 신약개발과 함께 평행(parallel)하게 이루어져서 그 가치가 검증되어야만 한다. 동반진단기술의 활용은 치료효과가 높게 기대되는 환자에서만 특정 약제를 처방할 있고, 또한 과도한 부작용이 우려되는 환자들은 회피할 수 있기에, FDA나 EMEA 같은 규제기관이나 보험회사에서의 요구가 상당히 커졌기에 이제는 신약개발의 아주 중요한 부분으로 자리 잡게 되었다. 따라서 동반진단기술과 관련된 진단키트를 개발하고 제조하는 관련 산업들이 매우 빠르게 성장하고 있다.

바. IT 기술

정밀의료를 위해서는 많은 사람들이 실시간에 접하는 환경의 변화나 사람들의 반응을 실시간으로 장기간 모니터링해야만 한다. 이러한 자료의 획득과 저장과 분석을 위해서는 많은 IT 기술이 필요하다. 혈압이나 심박수 같은 생체징후를 센서(sensor)를 통해 실시간 감지해내면 이러한 자료를 바로 전자건강기록(electronic health record)으로 저장하고 이를 분석하고 해석하여 임상의에게 전달하기 위해서는 사물인터넷(internet of things)이나 딥러닝, 인공지능(artificial intelligence) 등의 첨단기술들이 필요하고, 동시에 민감한 개인정보를 지키기 위해서는 블록체인(blockchain) 기술이 필요하다. 정밀의료의 완벽한 구현을 위해서는 보다 많은 IT 기술이 개발되고 활용되어야만 한다.

참고문헌

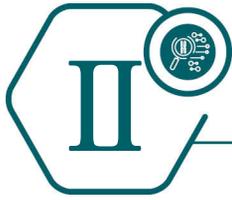
- Afza, M., Islam, S. M. R., Hussain, M. *et al.*(2020). “Precision medicine informatics: Principles, prospects, and challenges”. *IEEE Access*, (9).
- Charles River Associates(2018). “The benefits of personalised medicine to patients, society and healthcare systems”.
- Collins, F. S., Varmus, H.(2015). “new initiative on precision medicine”. *New England journal of medicine*, 372(9), pp.793-795. (DOI: 10.1056/NEJMp1500523).
- Hess, C. N., Low Wang, C. C., Hiatt, W. R.(2018). “PCSK9 Inhibitors: Mechanisms of Action, Metabolic Effects, and Clinical Outcomes”. *Annual Review Of Medicine*, 69(1), pp.133-145. (DOI: 10.1146/annurev-med-042716-091351).
- Ho, D., Quake, S. R., McCabe, E. R. B. *et al.*(2020). “Enabling technologies for personalized and precision medicine”. *Trends in Biotechnology*, 38(5), pp.497-518. (DOI: 10.1016/j.tibtech.2019.12.021).
- Jameson, J. L., Longo, D. L.(2015). “Precision medicine – Personalized, problematic, and promising”, *New England journal of medicine*, 372(23), pp.2229-2234. (DOI: 10.1056/NEJMs1503104)
- Kohane, I. S.(2015). “Ten things we have to do to achieve precision medicine”. *Science*, 349(6243), pp.37-38. (DOI: 10.1126/science.aab1328).
- Manrai, A. K., Kohane, I. S.(2017). “Bioinformatics and precision medicine”. In Wright A, Cresswell K, Bates D. W, Sheikh A(Eds). “Key Advances in Clinical Informatics”. Academic Press, pp.145-160.
- Pantel, K., Alix-Panabières, C.(2019). “Liquid biopsy and minimal residual disease – latest advances and implications for cure”. *Nature Reviews Clinical Oncology*, 16(7), pp.409-424. (DOI: 10.1038/s41571-019-0187-3).
- Sholl, L. M., Aisner, D. L., Varella-Garcia, M. *et al.*(2015). “Multi-institutional oncogenic driver mutation analysis in lung adenocarcinoma: The Lung Cancer

Mutation Consortium Experience”. *Journal of Thoracic Oncology*, 10(5), pp.768-777. (DOI: 10.1097/JTO.0000000000000516).

Ugalmugle, S., Swain, R.(2020). “Precision medicine market size by technology, by application, by end-use, industry analysis report, regional outlook, application potential, competitive market share & forecast, 2020–2026”. Global Market Insights.

II. 유전체 염기서열 분석의 원리와 미래기술





유전체 염기서열 분석의 원리와 미래기술

1 인간 유전체 염기서열 정보 분석이 왜 중요한가? (서론)

사람의 세포에는 부모로부터 물려받은 23쌍의 상동염색체(homologous chromosome)가 있다. 부모에서 온 염색체가 유사한 크기와 구조를 가진 염색체끼리 쌍을 이루고 유전 정보를 전달한다. 염색체를 구성하는 분자는 DNA로 아데닌(adenine, A), 타이민(thymine, T), 구아닌(guanine, G), 싸이토신(cytosine, C) 등 4개의 염기(base)로 구성된 뉴클레오타이드(nucleotide)가 서열을 이루면서 한 가닥의 사슬을 이루고 있다. 23개 염색체마다 길이가 다르나 전체는 30억 개 뉴클레오타이드의 순서로 표시할 수 있다. 하나의 세포에 있는 DNA의 물리적인 길이는 3미터이며, 이를 A, T, G, C 염기 글자로 표시하면 대략 700megabyte 크기가 된다.

1953년에 J. Watson과 F. Crick이 발표한 “이중나선(double helix)” 구조에 의하면 DNA는 두 사슬에 있는 A와 T, 그리고 G와 C 염기가 서로 마주보면서 수소결합에 의한 쌍을 이루고 있다. DNA 이중나선 구조를 통해 DNA 복제가 어떻게 정확하게 일어날 수 있는지 설명할 수 있게 되었고, 부모의 DNA에 있는 30억 개 염기서열의 정보가 정확하게 자식들에게 전달될 수 있는지 알게 되었다. 또한 30만 년 전에 아프리카에서 출현한 Homo sapiens종이 어떻게 현재 인류를 구성하였는지, 그리고 인종마다 생김새가 다르고 성격이 다른지 설명할 수 있게 된 것이다.

사람의 30억 개 염기서열을 규명하는 것은 생물종으로서 인간의 특성을 설명할 수 있다. 인간과 가장 가까운 대형유인원(great apes) 중에 침팬지는 전체 유전체 염기서열 중에 1.3%가 사람과 다르다. 단백질마다 2개 정도의 아미노산이 다른 셈이다. 이러한 차이가 침팬지와 사람의 생물학적, 사회문화적, 행동학적 특성을 구별한다. 사람은 개인별로 3.3백만 개의 single nucleotide variation(SNV)과 50만 개의 insertion/deletion(Indel)이

있고, 이 중에서 250~300개 정도는 유전자의 기능이 소실되는 loss-of-function(LoF) 변이를 갖고 있다. 개인 간의 차이는 이러한 염기서열의 차이에 의해 발생한다. 따라서 개인의 유전체 염기서열을 분석하면 각 개인의 질병 감수성, 약물반응, 약물치료 효과와 같은 유전적 특성을 설명할 수 있다.

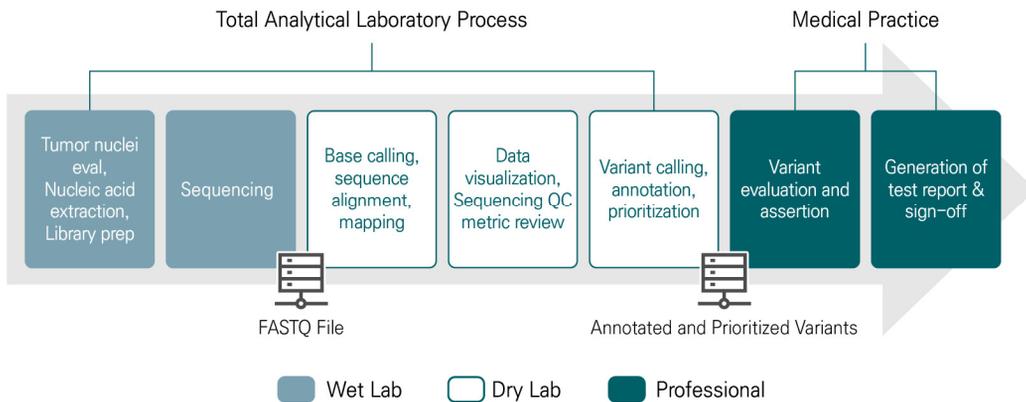
사람의 전체 유전체 염기서열에 대한 정보는 2003년에 발표된 Human Genome Project(HGP)을 통해 완성되었는데, 이 과정에서 유전체 염기서열을 효율적으로 분석할 수 있는 기술들이 개발되었다. 인간 표준 유전체 염기서열을 얻기 위해서 HGP에서 핵심적으로 사용한 초고속 시퀀싱 기술은 capillary sequencing 기술이었다. 이후 개발된 유전체정보분석 기술들은 개인별로 유전체 염기서열을 분석하기 위해서 분석량을 획기적으로 늘리고, 시간을 단축하기 위해 노력해야 했다. 이제까지 개발된 유전체정보 분석 기술의 핵심은 많은 DNA 조각을 동시에 분석하는 shotgun sequencing 방법과, 인간 표준 유전체 염기서열을 이용해서 생물정보학적 분석으로 해독하는 단계로 축약할 수 있다. HGP에서 3조 원의 비용이 소요된 것에 비해 이제 next generation sequencing(NGS)으로 통칭하는 대용량 염기서열 분석기술(high-throughput multiple parallel sequencing)로 개인의 30억 개 염기서열을 수십만 원 수준에서 분석할 수 있게 되었다. HGP의 목표는 사람의 표준 염기서열을 분석하는 것이지만, 이제 NGS를 통해 환자나 개인의 염기서열을 분석해서 개인별로 진단하고 정밀의료를 통해 개인맞춤 치료를 하는 것이 중요한 시대가 된 것이다.

2 유전체 분석기술의 핵심은 무엇인가?

2005년에 454 Life Sciences사가 400Mb정도를 10시간에 분석할 수 있는 GS20라는 대용량 유전체분석기술을 처음으로 시장에 소개한 이후로 현재 Illumina사와 Thermo Fisher Scientific사, 그리고 BGI사의 시퀀서가 각각 미국 및 중국 식약처의 허가를 받아 임상적으로 분자진단에 사용되고 있다. 최대 분석용량은 하루에 60명의 whole genome sequence를 동시에 분석할 수 있는 수준이다. NGS는 생화학적으로 DNA 조각에 있는 염기의 순서를 정하는 첫 번째 단계인 시퀀싱(sequencing) 단계와 이후 두 번째 단계로

동시에 수십만 개 또는 수억 개의 DNA 조각으로 나누어진 정보를 분석하는 생물정보분석 (bioinformatics) 단계로 나눌 수 있다 <그림 2.1>. 병원에서는 환자의 유전체 염기서열을 분석한 후에는 각 변이의 임상적 의미를 해석하고 보고하는 단계가 추가된다. 많은 국가에서 이미 임상적으로 NGS를 분자진단에 활용하고 있으며, 이에 대한 기술표준과 허가에 대한 기준이 정해져 있다. 국내에서도 2017년부터 희귀질환과 암 환자를 대상으로 보험급여를 통한 서비스가 시작되었고, 식품의약품안전처가 NGS에 대한 기술적 표준과 가이드라인을 제공하고 있다.

그림 2.1 유전체 염기서열 정보분석 워크플로우



가. 시퀀싱(sequencing) 기술

2020년 현재 국내에서 보건복지부로부터 임상진단에 사용가능한 장비로 허가받아 NGS 임상검사에서 운영 중인 시퀀서를 기준으로 하면, Illumina사와 Thermo Fisher Scientific사의 기술이 보편적으로 병원 정밀의료를 위해 활용되고 있다. 최근에는 중국 BGI사가 개발한 MGISEQ 계열의 시퀀서에 대한 기술 및 성능평가분석 논문이 발표되었고, 이는 앞으로 임상에 사용될 수 있다. 이들 세 가지 시퀀싱 장비에 대한 기술을 비교분석하면 <표 2.1>과 같다.

표 2.1 NGS 장비에 대한 분석성능 비교

Features	Ultra-high Throughput		Medium Throughput	
	~60 germline whole genome sequencing per run in average 30X read depth		~48 sequencing of 2 M-sized cancer panel per run in average 1,000X read depth	
Product Model	NovaSeq 6000	MGISEQ-T7	NextSeq 550	MGISEQ-200
Max. Throughput / RUN	6,000 Gb	6,000 Gb	120 Gb	150 Gb
Effective Reads	2,0000 M	5,000 M	400 M	500 M
Average run time	~13-38 hours	~24 hours	12-30 hours	10-66 hours

대용량 시퀀싱 기술은 하나의 반응을 얼마나 작은 규모로 동시에 할 수 있는지가 관건이다. Emulsion PCR 기술개발을 시작으로 많은 논문과 특허를 통해서 현재 NGS 기술이 지속적으로 발전하고 있다. 시퀀싱 분석과정을 다시 단계별로 나누면 첫 번째는 DNA를 작은 조각으로 만든 후에 시퀀싱이 가능하도록 library를 제작하는 전처리 과정이다. 유전체 전장염기서열을 분석할 때에는 세포나 조직, 혈액에서 추출한 genomic DNA를 사용할 수 있지만, 단백질을 합성하는 coding region인 exon 부분만을 분석하게 되면 해당 영역을 PCR 또는 probe hybridization을 통해 선택적으로 골라서 library를 제작한다. 임상적으로 진단에 필요한 유전자 영역만을 선택적으로 분석하기 위해서 만든 targeted gene panel sequencing의 경우에도 일부 유전자가 포함된 영역만을 분석하게 된다 <표 2.2>, <그림 3.5>. Illumina사와 Thermo Fisher Scientific사, 그리고 BGI사의 경우 각각 고유한 library 제작방법을 사용하는데, 임상진단을 목적으로 생검조직이나 혈액을 이용할 경우 적게는 10~50ng DNA를 사용하여 진단할 수 있어야 한다.

표 2.2 유전체정보별 분석기준과 데이터양

Applications	Whole genome sequencing (WGS)	Whole exome sequencing (WES)	Targeted gene panel sequencing	RNA sequencing (RNA-seq)	Whole genome- bisulfite sequencing (WGBS)
Detectable variations	SNV, Indel, CNV, translocation and all structural variations	SNV, Indel, CNV and exonic gene fusion	SNV, Indel, CNV and gene fusion of selected 50~500 genes	Gene expression, splicing and fusion	DNA methylation
Target region	3,000 M	30 - 60M	0.5 - 3 M	~30 M	3,000 M
Average read numbers	30X (germline) 60 ~ 90X (somatic)	50X (germline) 200X (somatic)	200X (germline) ~1,000X (somatic)	5~200 M	5~15X
FASTQ file size	~300 Gb (germline)	20Gb (germline)	10 Gb (germline)	5~10 Gb	~150 Gb

시퀀싱 분석과정의 두 번째 단계는 시퀀서 장비를 이용하여 염기서열을 해독하는 단계인데, 시퀀싱 장비마다 고유한 생화학적 원리를 적용하고 있다. Illumina사의 경우 bridge amplification을 통해 좁은 공간에서 효율적으로 DNA 조각을 분리하고 분석할 수 있게 하였다. 고집적 분석결과를 고해상도 이미지 분석을 통해 100~200bp 정도 길이의 DNA 염기서열을 해석한다. 2017년에 출시된 NovaSeq 6000의 경우 동시에 48명의 전장유전체염기서열에 해당하는 3terabase 크기의 정보를 해석할 수 있는 대용량 분석장비이다. 국내에서 보건복지부가 허가하는 NGS 임상검사실에서 사용할 수 있는 Illumina사의 장비는 MiSeqDx와 NextSeqDx로 각각 15Gb 및 200Gb 정보의 데이터를 생산할 수 있어서 임상에서 진단에 필요한 50~200개 유전자를 분석하는 panel sequencing에 적합하다.

Thermo Fisher Scientific사의 시퀀싱 장비의 경우 형광표지나 이미지를 사용하지 않으며 DNA polymerase가 합성해 나가는 과정에서 뉴클레오타이드마다 유리되는 이온의 전하량이 다르다는 것을 이용해서 수소이온농도를 측정하는 방식이다. 국내에서는 Ion PGM Dx와 Ion Proton S5가 식품의약품안전처의 허가를 받아 NGS 임상검사실에서 사용할 수 있으며, 각각 12Gb 및 500Gb 정도의 데이터를 생산할 수 있다. Thermo Fisher Scientific사의 시퀀싱 장비의 특이점은 특정 hotspot mutation에 대한 정보는 정확도가 높으나 나머지 영역에 대한 변이는 정확도가 낮다는 것이다. 따라서 아직 의미가 명확하지 않은 variant of unknown significance(VUS)에 대한 정보를 제공하는데

한계가 있다. 따라서 Ion Proton S5 및 PGM Dx 장비의 데이터로부터 새로운 의미 있는 변이를 찾는 것은 불가능하다. 이를 반증하는 것은 우리가 많이 접하고 사용하는 전장유전체 염기서열 빅데이터 중에는 Ion Proton 장비를 이용한 유전체 데이터를 보기 어렵다. 따라서 임상적으로나 연구목적으로 필요한 데이터를 축적하고 생산하기 위해서는 데이터의 일반적인 품질을 보장할 수 있는 방법을 선택하는 것이 필요하다.

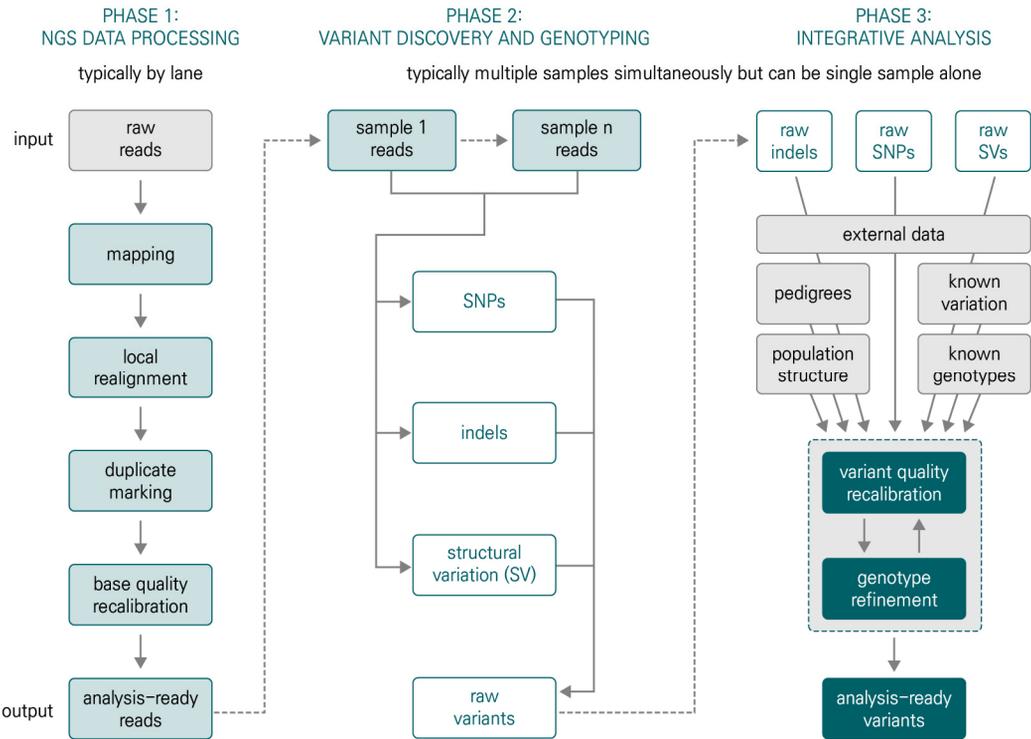
2018년에 시장에 출시된 중국 BGI사의 MGISEQ은 기존 Complete Genomics사가 개발한 DNA nanoball sequencing 방법을 사용하였다. 일종의 rolling circle 방식으로 DNA 조각을 반복적으로 시퀀싱해서 정확도를 높이는 방법이다. 현재 MGISEQ-2000과 같은 대용량 시퀀서의 성능 분석결과가 보고되고 있는데 Illumina사의 분석성능과 비교하여 큰 차이가 없는 것으로 보인다. 시장에서의 경쟁을 위해 비용이 상대적으로 저렴하며, 현재 중국 식약처에서는 의료장비로 허가를 받은 상태이다.

나. 생명정보분석(bioinformatics) 기술

시퀀싱 단계에서 얻은 데이터로부터 임상 분자진단에 필요한 최종 리포트를 생산하는 과정은 다시 크게 3단계로 구분할 수 있다 <그림 2.2>. 첫 번째는 시퀀싱 장비에서 생산된 초벌 데이터를 분석 가능한 형태로 가공하는 단계이다. 시퀀서에서 생산하는 1차 데이터는 각 bead나 cell에서 생산된 이미지 형태인데, 이를 A, T, G, C와 같이 뉴클레오티드 정보로 전환하여 text 형태의 FASTQ로 저장한다. <표 2.2>에 정리한 바와 같이 FASTQ 파일의 크기는 분석한 유전체의 종류에 따라 달라지는데 혈액에서 분리한 DNA로 germline variation을 분석하는 WGS을 기준으로 하면 정보의 분량은 대략 300gigabyte 정도 크기가 된다. 생명정보분석을 위한 두 번째 단계에서는 FASTQ file 형태로 저장된 짧은 염기서열 정보를 표준 염기서열과 비교하여 유전체상의 위치정보가 연계된 BAM file로 전환된다. 이 과정에서 사용하는 표준 염기서열에 따라 결과가 달라질 수 있다. 특히 지속적으로 업데이트되는 인간 유전체 염기서열 정보를 정확하게 파악해야 한다. 또한 시료 내에 존재하는 병원체를 분석하기 위해서는 사람 외에 바이러스나 세균에 대한 유전체 정보를 template로 하면 세포 또는 검체 내에 존재하는 다른 미생물을 동시에 분석할 수 있다. 마지막 세 번째 단계에서는 표준 염기서열과 다른 부분을 골라 차이가 있는 부분만

VCF file의 형태로 전환한다. 이를 통해 전장유전체 염기서열 분석의 경우에도 30억 개 염기서열 중에 개인에서 차이가 있는 3백만 개 SNV, Indel 등이 1.5gigabyte 정도의 크기 파일로 줄어들게 된다.

그림 2.2 NGS 데이터로부터 원인 돌연변이를 찾기 위해 필요한 생물정보분석 워크플로우



자료: DePristo, M. A. *et al.*, 2011. Nature Genetics, 43, pp.491-498.

생물정보분석 과정에서 데이터 품질관리와 옵션에 따라 측정되는 변이가 달라질 수 있다. 따라서 생물정보분석 과정에 대한 엄격한 품질관리가 필요하다. 따라서 임상에서 진단목적의 생물정보분석은 파이프라인이나 패키지 형태로 개발된 하나의 소프트웨어를 사용하고 있다. WGS과 같이 대용량 데이터를 분석하기 위해서는 각 단계에서 최적의 옵션을 정하고 품질관리와 함께 제공하는 Dragen과 같은 상업적 분석파이프라인을 사용할 수 있다.

정밀의료 클리닉에서 사용하는 임상 분자진단 유전체정보분석을 위해서는 각 변이에 대한 임상적 의미를 추가한 주석달기(annotation) 단계가 추가로 더 필요하다. 염기서열의

변이가 유전성 암이나 희귀질환의 진단에 중요한 원인인지, 그리고 표적항암제 치료반응에 해당하는 타겟 돌연변이인지 확인이 필요하다. 질환과 관련된 유전체 변이에 대한 정보는 PubMed를 비롯한 ClinVar, OMIM, CIPIC, CIVIC와 같은 공공 데이터베이스에 축적되고 있다. 유전체정보의 해독단계에서 주석달기를 위해서는 환자의 다양한 돌연변이 중에서 어떤 변이가 임상적으로 검증된 변이인지 정리하여 환자의 진단과 치료에 활용해야 한다. 공공 데이터로부터 추출된 지식정보의 성격상 유전체 변이에 대한 주석달기는 병원이나 진단 솔루션에 따라 동일한 변이에 대한 해석이 서로 다를 수 있다. 따라서 각 병원의 정밀의료 경쟁력이 이러한 정보의 해석단계에서 발생하게 된다. 미국 대형병원의 경우에는 강력한 bioinformatics 및 IT 인프라를 기반으로 하여 자체적인 시스템을 운영하고 있다. 상업적으로는 Foundation Medicine이나 Guardant Health의 경우 자체적인 데이터베이스에 의존하여 주석달기 단계에서 수월성을 유지하고 있다.

다. 새로운 유전체 염기서열 분석기술 소개

대규모 유전체정보분석 프로젝트에도 불구하고 암 환자에 대한 치료가 제한적인 것은 아직도 암에 대한 이해가 부족하기 때문이다. 암유전체 수준에서는 구조변이(structural variation) 분석이나 변이 간 관계에 대한 phasing, transposon insertion, single cell과 같은 새로운 분야와 기술에 대한 수요가 늘어나고 있다. 유전체 구조변이 분석을 위해서는 긴 염기서열 정보분석기술이 필요하다. PacBio사는 single molecule real time sequencing(SMRT) 기술을 기반으로 30,000bp 크기의 긴 DNA 조각을 분석할 수 있는 RS 장비를 2011년 초에 출시하였다. 긴 DNA 조각에 adaptor를 붙여서 circle의 형태로 만들어준 후에, 고정된 DNA polymerase에서 형광표지된 뉴클레오티드가 붙을 때 일어나는 신호를 ZMW(zero-mode waveguide)에서 반복적으로 정보를 읽도록 되어 있다. 특히 circular DNA가 원형으로 계속 돌면서 반복적으로 읽어 들이기 때문에 에러를 줄일 수 있다는 장점이 있다. 최근에는 Roche Diagnostics사와 공동으로 분자진단 제품을 개발하고 있다. 다른 시퀀싱 장비가 100~200bp 정도의 짧은 DNA 조각을 분석하는 것에 비해 긴 DNA 조각을 분석하는 것은 유전체 구조변이, 특히 translocation이나 transposon과 같은 변이를 분석한다는 장점이 있다. 또한 de novo assembly를 통해

새로운 종의 유전체를 분석하거나 표준 염기서열 없이 분석하여 새로운 유전적 특성을 분석하는 데에 활용할 수 있다. 그 사례로 2016년에 발표된 한국인 유전적 특성 분석에서 long-read sequencing 방법을 활용한 바가 있다 (<https://www.pacb.com>).

Nanopore sequencing 방법은 nanometer 크기의 좁은 공간으로 DNA 가닥이 이동하면서 발생하는 뉴클레오타이드별 전류의 차이를 이용하여 염기서열을 결정하는 기술이다. 이 경우 DNA 증폭이 필요하지 않고, 화학적 표지가 필요하지 않으며, 10~40kbp 크기의 긴 하나의 DNA 또는 RNA 조각에 대한 정보를 얻을 수 있다는 장점이 있다. Oxford Nanopore Technologies사의 PromethION의 경우 최대 15Tb 크기의 데이터를 48시간 내에 생산할 수 있다. 또한 MinION의 경우에는 USB 크기의 100그램 미만의 초경량 시퀀싱 장비로 우주비행사나 밀림, 오지에서 유전체분석이 필요한 경우에 point-of-care에서 쉽게 활용할 수 있다는 장점이 있다. 최근에는 COVID-19의 원인 바이러스인 SARS-CoV 염기서열의 신속한 분석에 활용되기도 하였으며, PacBio의 기술과 마찬가지로 haplotyping이나 de novo assembly에 유리하다. 하지만 SNV나 Indel과 같은 변이에 대한 분석 정확도가 90% 정도로 장비자체의 오류가 높기 때문에 염기서열 분석을 위한 활용에는 제한점이 많다 (<https://nanoporetech.com>).

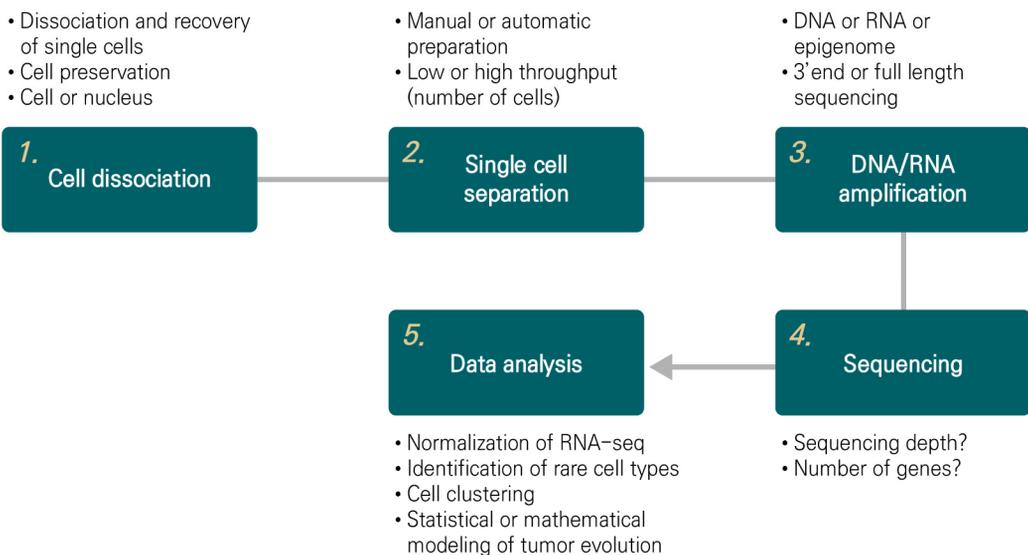
정밀의료 클리닉에 필요한 유전체정보분석기술에 대한 선택과정에서 고려해야 하는 것은 분석정확도와 함께 소요시간과 비용이다. 임상진단을 위해 새로 소개된 기술로는 GenapSys사가 개발한 Gene Electronic Nano-Integrated Ultra-Sensitive(GENIUS) technology 기술이 주목을 받고 있다. 기존 시퀀싱 방법과 달리 DNA 합성과정에서 염기가 추가될 때 생기는 임피던스의 차이를 COMS sequencing chip 위에 있는 센서를 이용해서 측정하는 것이다. 현재 최대 144만 개에 센서를 이용한 chip으로 하루 만에 대용량 데이터를 생산할 수 있는데 형광물질이나 나노장비를 사용하지 않기 때문에 장비가 작고 분석비용이 저렴하다. 임상에서 작은 영역의 유전자에 대한 분자진단 패널 시퀀싱에 적합한 기술로 임상적 유용성이 높을 것으로 예상된다 (<https://www.genapsys.com>).

3 유전체정보분석의 미래는?

가. 단일세포 유전체분석 기술(single cell genome analysis)

대용량 유전체분석기술의 발전에 따라 개인별 유전체정보분석의 비용과 접근성의 한계를 극복하였지만 임상적으로 종양조직의 이질성(tumor heterogeneity)나 암 미세환경(tumor microenvironment)의 분석과 같은 새로운 영역에 대한 도전이 필요하다. Microfluidics 기술을 기반으로 수만 개의 세포를 동시에 oil phase 내에 물방울로 capture를 해서 bead에 붙은 RNA를 동시에 분석하는 drop-seq 방법이 개발되면서 단일세포에 대한 유전체 분석이 활발하게 시작되었다. 단일세포 유전체분석은 조직에서 세포를 단일세포 수준으로 분리하는 단계로부터 시작해서 분리된 공간에서 RNA를 추출하고 증폭하는 단계, 그리고 이어서 시퀀싱한 후에 생물정보분석하는 단계로 구분할 수 있다 <그림 2.3>. 전반적으로 각각의 세포에 대한 시퀀싱 과정으로 설명할 수 있으나 환자의 조직으로부터 세포를 잘 수집하고, 하나의 세포에 존재하는 10pg 이내의 극소량의 RNA를 사용해서 고품질의 데이터를 생산하는 것, 그리고 수천 개의 세포로부터 얻은 RNA-seq 데이터를 동시에 분석해야 하는 생물정보분석기술이 핵심이다.

그림 2.3 싱글셀 유전체정보 분석 단계



단일세포 RNA-seq 데이터를 분석해서 세포의 특성을 분석하기 위해서는 각 세포를 정확하게 정의할 수 있어야 하고 기존에 알려진 조직이나 장기에 대한 정보는 불충분하기 때문에, 인간의 다양한 세포에 대한 표준 데이터를 생산하는 프로젝트가 시작되었다. Human Cell Atlas (<https://www.humancellatlas.org>)는 연구자가 중심이 된 국제 공동연구 협의체로 인체 장기별로 여러 연구자들이 데이터를 모아서 정확한 단일세포 유전체정보를 제공하는 노력을 하고 있다. 예를 들어 말초혈액에 존재하는 면역세포의 다양성을 정의하기 위해 전 세계 다양한 인종을 대상으로 만여 명의 단일세포 유전체 데이터를 생산하고 있다. 최근에 발표된 인간의 주요 장기별 심층 단일세포 유전체 분석 결과는 향후 인체 생물학뿐만 아니라 질환 연구에 기준이 될 것으로 예상된다. 건강한 개인의 단일세포 유전체정보를 바탕으로 암 환자의 조직이나 말초혈액에 대한 병태생학적 해석이 가능할 것이다.

종양이질성과 암미세환경에 대한 분석도 매우 중요한 단일세포 유전체 분석 분야다. 최근 면역항암제에 대한 임상시험에서 치료반응에 대한 새로운 바이오마커 발굴을 위해서 단일세포 분석을 활용하고 있다. 면역항암제 치료에 따라 변하는 면역세포 타입이나 유전자를 특정함으로써 새로운 면역항암제의 치료반응을 이해할 수 있다. 암미세환경 분석을 통해 새로운 면역항암치료 표적을 찾거나 기존 치료제의 작용기전을 기반으로 동반진단 기술을 개발할 수 있다.

나. 유전체 빅데이터 연구

유전체 염기서열 정보 생산기술이 보편화 되고, 실제 임상검사의 하나로 자리매김하면서 동시에 데이터의 축적이 빠르게 진행되고 있다. 특히 국가별로 유전체정보가 국가정보 자원과 헬스케어산업의 인프라로서 가치가 높을 것으로 판단하여 대규모 시퀀싱 프로젝트를 통해 빅데이터를 구축하고 있다 <표 2.3>.

표 2.3 국가별 대규모 유전체 빅데이터 프로젝트

Project	Country	Contents
All of US	USA	Genome-wide SNP genotyping with clinical and life style data in one million volunteers
Genomics England	UK	Whole genome sequencing in 100K patients with rare disease or cancer
UK BioBank	UK	Genome-wide SNP genotyping with clinical data in 500K volunteers in NHS
Estonian Genome Project	Estonia	Whole genome sequencing in 52K volunteers with regular follow-up on health status
China Precision Medicine Initiative	China	Whole genome sequencing in 100M for precision medicine by 2030
Japan Genomic Medicine Program	Japan	Integrated database of clinical and genomic information and Tohoku Medical Megabank Project in 150K Japanese
Genomic Medicine France 2025	France	Whole genome sequencing in 60K patients per year for cancer, rare disease and diabetes by 2025
National Genome Strategy	Finland	Using genomic effectively in healthcare and in the promotion of health and wellbeing by 2020

다. 한국인 유전체 빅데이터 연구

1999년 과학기술부에서 국가 경쟁력 및 국민의 삶의 질 향상을 목표로 유전체 연구 지원 및 육성방향에 관한 기본정책을 수립하였다. 이에 따라 인간유전체기능연구사업단을 발족하여 10년간 한국인에 다발하는 위암 및 간암의 발병을 예측하여 이를 조기에 진단 하거나 치료 효율을 증대시켜 암환자의 생존율을 향상시키기 위한 기술 개발을 추진하였다. 2014년에는 과학기술정보통신부, 보건복지부, 산업통상자원부, 해양수산부, 농림축산식품부, 농촌진흥청 및 산림청 등 정부의 7개 부 및 청이 참여하는 포스트게놈 다부처 유전체사업단을 출범시켜 2021년까지 총 5,800억 원을 투자하여 국내 유전체 연구가 활성화 되었다. 그동안 질병관리본부 인체자원은행에 40여 만 명의 인체유래물 자원과 유전체정보가 확보되었고, 한국인유전체역학조사사업을 통해 수백편의 논문이 발표되었다.

NGS 기술발전에 따라 전장유전체 정보에 대한 보건의료 및 산업계에서 활용도가 증가하면서 국내에서도 2020년부터 미래 헬스케어 산업과 서비스를 위한 유전체정보의 활용을 염두에 두고 한국인 백만 명에 대한 유전체 및 임상정보, 생활정보를 망라하는 ‘국가 바이오 빅데이터 구축사업’을 시작하였다. 시범사업을 통해 희귀질환자 모집 및 기존 대규모 국가 유전체 연구사업과 연계를 통해 총 2만 명 이상의 임상정보와 유전체 데이터를 구축할 예정이다.

한국인의 유전체 특성은 동아시아 인종에서 중국인 및 일본인의 변이형과 유사하며 한국인 유전체 빅데이터 분석을 통해 동아시아인종의 특성을 분석할 수 있다. 한국인 유전체 데이터는 우리 주변 환경 및 전통적인 관습에 대한 데이터를 포함하는 보건의료 빅데이터를 인공지능 분석에 활용할 수 있기 때문에 필요하다. 고유한 우리의 특성을 반영하면 새로운 신약개발 타겟을 발굴할 가능성이 높기 때문에 충분한 가치가 있다고 생각한다. 또한 한국인에서만 나타나는 인구집단 내 0.1% 이하의 희귀변이들이 서로 다르며, 이런 희귀 변이에 질환발생 및 표현형질의 차이를 설명할 수 있을 것으로 판단된다.

라. 유전체 빅데이터 구축 및 활용전략

NGS 기술에 기반한 유전체분석은 DNA 수준에서 유전자 변이가 질병의 발생이나 치료 반응, 예후예측에 미치는 영향을 파악하고 있다. 생물학적으로 DNA methylation 및 chromatin regulation, 단백질 및 대사체와 같이 다양한 수준에서 조절되고 있기 때문에 생리학적 현상과 질병에 대한 정확한 분석을 위해서는 다중오믹스(multi-omics) 분석기술의 접목이 필요하다. 이러한 다중오믹스 분석은 같은 시료를 대상으로 동시에 분석하는 것이 필수적이나 현재 유전체 분석을 위해 사용되고 있는 인체유래물의 양이 적고, 유전체에 비해 단백질 및 대사체등에 대한 다중오믹스 분석 비용 때문에 제한적으로 데이터가 수집되고 있다. 글로벌 컨소시엄에서도 일부 시료에 대해 특수한 목적으로 다중오믹스 데이터를 수집하고 있다. 아직은 다중오믹스 연구는 생물학적인 조절기전 연구와 같은 기초 기반연구를 목적으로 하고 있으며 보건의료 분야 및 산업적 활용을 위한 유용성에 대한 검증이 필요하다.

국내에서는 유전체정보를 기반으로 한 글로벌 임상시험이나 보험급여를 통한 임상검사와 같이 NGS 기술이 정밀의료에 적용되고 있으며, 이러한 바이오의료 빅데이터의 유용성은 해외 네트워크를 통한 글로벌 빅데이터 사업과 연계되는 것이 필요하다. 특히 의료 및 유전체 데이터 표준에 참여하여 데이터 활용도를 높여야 하며, 동시에 해외 데이터를 통합 분석할 수 있는 기반을 만드는 것이 필요하다. 현재 International Cancer Genome Consortium(ICGC) 및 Accelerating Research in Genomic Oncology(ARGO)에 국내 병원과 연구자가 참여하여 임상-유전체 통합데이터 구축과 분석에 기여하고 있다.

4 유전체정보분석의 시대에 무엇을 할 것인가? (결론)

바이오헬스케어 유전체 분야에서 '빅데이터'의 중요성은 NGS 기술의 광범위한 사용에 따라 대두되기 시작했다. 최근 10여 년에 걸쳐 대규모 유전체분석 컨소시엄의 성공적인 결과에 따라 유용한 정보를 가진 유전체 데이터베이스의 수와 규모가 증가하고 있다. 대표적으로 TCGA (<https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>)와 ICGC (<https://icgc.org>)에서 제공하는 수만 명의 암유전체에 대한 체계적인 데이터는 암유전체 연구뿐만 아니라 표적항암제 및 면역항암제 개발에 유용한 정보를 제공하고 있다. NGS 장비의 보편적인 사용이 증가하고 데이터 생성 비용이 줄어들고 있기 때문에 향후 지속적인 데이터 증가를 쉽게 예상할 수 있다.

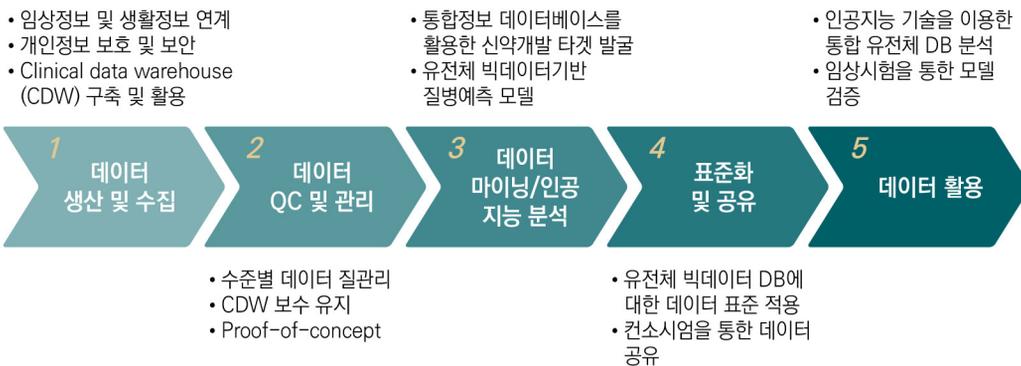
최근에는 미국 Memorial Sloan Kettering Cancer Center의 IMPACT 임상연구를 통해 얻은 암유전체정보 및 임상시험 데이터 (<https://www.mskcc.org/msk-impact>)를 비롯하여 영국의 UK BioBank (<https://www.ukbiobank.ac.uk/>)의 마이크로어레이 데이터와 Genomics England (<https://www.genomicsengland.co.uk/>)의 전장유전체 분석데이터가 임상정보와 함께 제공되고 있다. 이는 단순한 시퀀싱 데이터뿐만 아니라 표현형 정보가 통합된 데이터의 활용이 활발하게 일어나고 있다는 것을 의미한다. 앞으로 RNA 시퀀싱(RNA-seq) 및 메틸레이션 정보에 대한(BIS-seq)와 같은 다중오믹스 분석결과와 3차원 크로마틴 입체 염기서열 분석과 같은 훨씬 더 복잡한 후성유전체분석 결과가 포함될 수 있다. 이러한 새로운 시퀀싱 기술은 암에 대한 생물학적 특성을 다각적으로 파악하는 분석 패러다임의 진화를 유도할 것으로 예상된다.

성공적인 유전체-임상 통합정보 빅데이터 가치구현을 위해서 단계별로 고려해야 할 사항이 있다 <그림 2.4>. 첫 단계는 유전체 데이터와 임상정보의 생산 및 수집에 대한 것으로 실제 보건의료 빅데이터로서의 가치가 있는 수준의 질과 양이 보장되는 디자인이 필요하다. 이를 위해서는 유전체 데이터 생산방법에 대한 면밀한 검토와 질 관리 시스템이 필요하며, 연구자 및 사용자 편의성이 확보된 data warehouse 구축이 전제되어야 한다. 개인정보의 보호와 데이터 보안에 대해서는 윤리적 및 법적 근거가 확실한 형태의 데이터 수집 전략이 필요하다. 두 번째 단계는 데이터의 관리로 수집된 데이터의 가치를 높이기 위한 정리 및 검증이 필요하다. 지속적으로 데이터의 보수유지 및 유전학적 방법

과 생명정보분석기술을 이용한 검증이 수반되어야 한다. 세 번째 단계는 데이터 유용성에 대한 근거를 만들기 위한 수준의 데이터 마이닝과 인공지능 기술의 적용이다. 기존 패러다임에 적합한 수준의 데이터인지, 딥러닝을 포함한 분석기술의 적용이 가능한 정도의 데이터가 확보되었는지에 대한 체계적인 분석이 필요하다.

이러한 보건의료 유전체 빅데이터는 글로벌 표준에 따라 국내외 연구자들과 공유되고 호환될 수 있어야 한다. 글로벌 유전체 빅데이터 컨소시엄 참여와 데이터 공유를 통해 확보된 데이터의 질적 수준을 검증할 수 있다. 이러한 단계를 거친 후에 비로소 우리는 유전체-임상정보 통합 빅데이터를 신약개발과 인공지능 질환예측모델을 구축하는 데에 활용할 수 있다. 빅데이터를 기반으로 개발된 신약과 예후예측모델은 임상시험을 통해 검증되어야 하며, 이를 통해 빅데이터의 임상적 유용성을 증명할 수 있게 될 것이다.

그림 2.4 임상-유전체 통합 빅데이터 분석 워크플로우



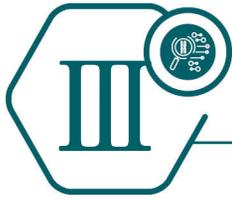
새로운 데이터 세트가 점점 더 빠른 속도로 생산됨에 따라 NIH(National Institutes of Health)가 후원하는 ‘Big Data to Knowledge’(BD2K, <https://datascience.nih.gov/bd2k>)와 같은 이니셔티브가 개발되고 있다. 앞으로 멀지 않은 장래에 대규모 데이터베이스를 기반으로 암 생물학을 비롯한 생물학 전반에 걸쳐 새로운 지식 추구를 위한 빅데이터의 가치가 구현될 것으로 믿는다. 국내에서도 2020년에 데이터3법으로 통칭되는 데이터 활용을 촉진하기 위한 법적 근거가 구축되었다. 유전체정보를 포함한 병원과 정부에 축적되어 있는 보건의료 빅데이터의 활용이 촉진될 수 있는 계기가 되길 기대한다.

참고문헌

- Abecasis, G., Altshuler, D., Auton, A. *et al.*(2010). “A map of human genome variation from population-scale sequencing”. *Nature*, 467, pp.1061-1073. (DOI: 10.1038/nature09534).
- Han, X., Zhou, Z., Fei, L. *et al.*(2020). “Construction of a human cell landscape at single-cell level”. *Nature*, 25. (DOI: 10.1038/s41586-020-2157-4).
- Hutter, C., Zenklusen, J. C.(2018). “The Cancer Genome Atlas: Creating Lasting Value Beyond Its Data”. *Cell*, 173(2), pp.283-285. (DOI: 10.1016/j.cell.2018.03.042).
- Kim, D. W., Lee, J. Y., Yang, J. S. *et al.*(2020). “The Architecture of SARS-CoV-2 Transcriptome”. *Cell*, 181, pp.1-8. (DOI: 10.1016/j.cell.2020.04.011).
- Macosko, E. Z., Basu, A., Satija, R. *et al.*(2015). “Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets”. *Cell*, 161(5), pp.1202-1214. (DOI: 10.1016/j.cell.2015.05.002).
- Seo, J., Rhie, A., Kim, J. *et al.*(2016). “De novo assembly and phasing of a Korean human genome”. *Nature*, 538, pp.243-247. (DOI: 10.1038/nature20098).
- Zhang, L., Li, Z., Skrzypczynska, K. M. *et al.*(2020). “Single-Cell Analyses Inform Mechanisms of Myeloid-Targeted Therapies in Colon Cancer”. *Cell*, 181(2), pp.442-459. (DOI: 10.1016/j.cell.2020.03.048).

Ⅲ. 유전체 염기서열 분석 데이터의 임상응용





유전체 염기서열 분석 데이터의 임상응용

1 중앙질환

가. 폐암

1) 폐암의 현황

가) 폐암의 역학

(1) 폐암의 발생률 및 생존율

대한민국 국가암정보센터 (<https://www.cancer.go.kr/>) 최신 통계 자료에 의하면 2017년 새로 발생한 암 환자 수는 남자 122,292명, 여자 109,963명으로 총 232,255명으로 집계되었다. 이중 폐암 환자는 남자 18,657명(발생 순위 2위), 여자 8,328명(발생 순위 5위)로 총 26,985명이었다. 이는 전체 암 환자에서 위암, 대장암에 이어 세 번째에 해당하였다.

최근 5년간(2013~2017년) 발생한 전체 암 환자의 5년 상대생존율은 70.4%로, 10명 중 7명 이상은 5년 이상 생존하는 것으로 보고되었다. 이중 폐암의 상대생존율은 30.2%로 10명 중 3명만이 생존하여 상대적으로 매우 낮은 생존율을 보였다.

(2) 폐암의 사망률

2018년 암으로 사망한 대한민국 국민은 총 79,153명으로 전체 사망자의 26.5%가 암으로 사망하였다. 이중 사망률이 가장 높은 암종은 폐암이었으며, 전체 암사망자의 22.5%인 17,852명이 폐암으로 사망하였다.

이와 같은 통계 자료는 폐암이 국민 건강에 미치는 영향이 매우 크다는 것을 보여 주며, 조기진단 및 치료 혁신을 위해 계속 노력해야 함을 시사한다.

(가) 폐암의 분류

① 폐암의 조직병리학적 분류

㉠ 폐암의 기원 조직 및 세포에 따른 종류

폐는 타 장기와 마찬가지로 여러 종류의 조직으로 구성되어 있으며 해당 조직에서 기원하는 다양한 종류의 암이 모두 발생할 수 있다. 즉, 암종(carcinoma), 육종(sarcoma), 림프종(lymphoma) 등이 원발적으로 발생할 수 있다. 그러나 이중 상피 기원의 악성 종양에 해당하는 암종이 95% 이상으로 가장 높은 빈도로 발생한다. 암종은 크게 소세포폐암(small cell carcinoma)과 비소세포폐암(non-small cell carcinoma)으로 분류하고, 전체 폐암에서 각각 15%와 85%를 차지한다. 소세포폐암과 비소세포폐암은 조직학적 소견, 유전학적 특성, 임상 양상이 모두 상이함이 잘 알려져 있다. 비소세포폐암은 크게 선암종(adenocarcinoma)과 편평세포암종(squamous cell carcinoma)로 분류할 수 있으며, 전체 폐암에서 선암종이 약 60%로 현재 가장 높은 비율을 차지한다.

㉡ 조직병리학적 분류의 의의

조직병리학적 분류는 진행성 폐암에서 항암제 선택 및 분자병리 검사의 범위를 결정하는 중요한 기준으로 제시되고 있다. 암 환자 치료의 중요한 지침 중 하나인 National Comprehensive Cancer Network(NCCN, www.nccn.org) 가이드라인에 의하면 비소세포-비편평세포(non-small cell non-squamous) 폐암의 경우 EGFR, ALK 등의 유전자 검사를 반드시 시행할 것을 권고하고 있다.

② 폐암의 분자병리학적 분류

폐암은 조직학적 분류가 가장 기본적인 토대이며 분화가 불량할 경우 면역조직화학 염색을 통해 특정 단백질 발현 여부에 따라 추가적으로 분류할 수 있다. 앞서 기술한 폐암들은 각각 TTF-1, p40, CD56 등의 독특한 단백질을 발현하게 되므로, 이러한 단백질들을 진단 표지자로 사용하고 있다. 최근 분자생물학 및 유전체 분석기술의 발전으로 폐암의 발생 및 진행 과정에서 중요한 역할을 하는 유전자 변이들이 다수 발견되었다 <표 3.1>. 대표적인 예로 EGFR, BRAF 돌연변이, ALK, ROS1 및 RET 융합 유전자 등이다. 이는 표적항암제의 개발로 이어지면서, 해당 항암제의 안전성과 유효성을 예측하는 중요한

바이오마커로 사용되고 있다. 또한 이러한 유전자 변이를 단일 유전자 검사(single gene assay)로 검출하고 있으나, 최근 차세대 염기서열 분석법(next generation sequencing, NGS)이 임상에 활발히 사용되고 있다 <그림 3.1>.

표 3.1 폐암에서 발견되는 유전자 변이 및 해당 표적 항암제

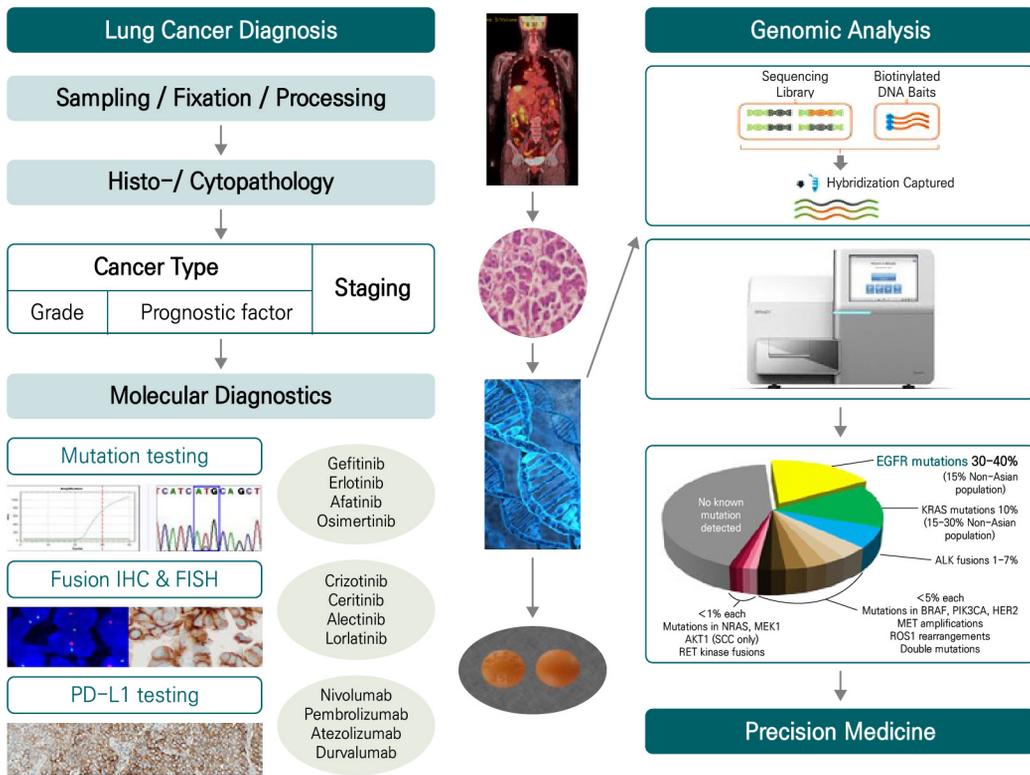
Gene	Representative subtypes or variants	Frequency	Targeted agents
Mutations			
<i>EGFR</i>	Exon 19 deletion, Exon 21 L858R, Exon 20 T790M	40–50% in ADCs ¹ 10–20% in ADCs ²	Gefitinib, Erlotinib, Afatinib, Osimertinib
<i>KRAS</i>	G12X, G13X, G61X	5–10% in ADCs ¹ 20–30% in ADCs ²	(G12C) AMG–510
<i>BRAF</i>	V600E	1–4% in ADCs	Dabrafenib+Trametinib
<i>HER2</i>	p.A775 G776insYVMA in exon 20	1–2% in ADCs	Trastuzumab, Neratinib
<i>MET</i>	Splice site mutations around or in exon 14	3–4% in ADCs	Crizotinib, Cabozantinib, Capmatinib
Gene fusions			
<i>ALK</i>	<i>EML4-ALK</i> , <i>TGF-ALK</i> , <i>KIF5B-ALK</i>	5% in ADCs	Crizotinib, Ceritinib, Alectinib, Lorlatinib
<i>ROS1</i>	<i>CD74-ROS1</i> , <i>EZR-ROS1</i> , <i>SLC34A2-ROS1</i> , <i>SDC4-ROS1</i>	1% in ADCs	Crizotinib, Ceritinib, Entrectinib
<i>RET</i>	<i>KIF5B-RET</i> , <i>CCDC6-RET</i>	1% in ADCs	Cabozantinib, Vandetanib, LOXO–292, BLU–667
<i>NTRK1</i>	<i>MPRIP-NTRK1</i> , <i>CD74-NTRK1</i> , <i>TPM3-NTRK1</i>	<1% in ADCs	Larotrectinib, Entrectinib
<i>FGFR1/3</i>	<i>FGFR3-TACC3</i> , <i>BAG4-FGFR1</i>	1% in NSCLCs	Erdafitinib
<i>NRG1</i>	<i>CD74-NRG1</i> , <i>SLC3A2-NRG1</i> , <i>VAMP2-NRG1</i>	7% in mucinous ADCs	ERBB3 inhibitor
Amplifications			
<i>FGFR1</i>	Gene amplification	13–22% in SQCs	FGFR inhibitor
<i>EGFR</i>	Gene amplification	8–9% in SQCs,	EGFR inhibitor
<i>MET</i>	Gene amplification	2–4% in ADCs	Crizotinib
<i>HER2</i>	Gene amplification	1–2% in ADCs	Trastuzumab, Neratinib

ADC, adenocarcinoma; SQC, squamous cell carcinoma; NSCLC, non-small cell lung carcinoma; NA, not available;

¹Asian populations, ²Western populations

자료: Shim, H. S. *et al.*, 2017. Molecular Testing of Lung Cancers 재구성

그림 3.1 폐암의 진단 및 유전체 분석 모식도



자료: Shim, H. S. *et al.*, 2017

2) 폐암의 유전체 염기서열 분석 데이터 적용

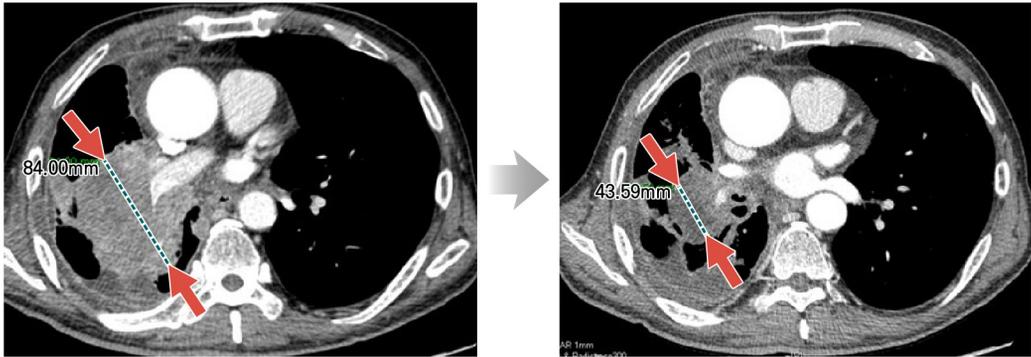
앞서 기술한 폐암 현황에 대한 서론에 이어 폐암 환자의 진단 및 치료에 유전체 염기서열 분석 데이터를 어떻게 적용할 수 있을지 살펴보려고 한다. 폐암을 병기에 따라 구분하여 진행성 폐암에서의 적용을 먼저 다루고, 이어 조기 폐암에서의 적용, 그리고 마지막으로 폐암의 조기진단을 위한 적용 부분으로 나누어 살펴보려고 한다.

가) 진행성 폐암에 대한 적용

진행성 폐암은 전신적인 항암 치료가 필요하므로, 안정성과 유효성이 검증된 약제를 투약하는 것이 필요하다. 각각의 항암제는 작용하는 기전이 다르고 모든 환자에서 동일한 효능을 보이지는 않으므로, 안정성과 유효성을 예측할 수 있는 바이오마커에 따라 투약해야

한다 <그림 3.2>. 현재 사용되는 바이오마커는 암의 발생 및 진행에 중요한 역할을 하는 드라이버 돌연변이(driver mutation)에 근거하므로, 유전체 염기서열 분석 데이터는 폐암 환자의 항암제를 선택하는 기준이 되고 있다.

그림 3.2 유전체 분석에 근거한 표적 항암제 치료 반응 1의 예(좌: 사용 전, 우: 사용 후)



- ① 표적항암제 선택을 위한 유전자 변이 검출
- ② 진행성 폐암 환자의 일차 치료제 선택

- EGFR 돌연변이 폐암

2004년 EGFR 억제제(당시 제피티닙)의 반응을 예측할 수 있는 EGFR 돌연변이의 발견은 폐암 환자를 위한 개인맞춤-정밀의학의 첫 출발이 되었다. 또한 유전체 염기서열 분석 데이터가 암 환자의 치료 및 삶의 질에 가져올 거대한 변화를 알리는 서막이었다. 이후 진행된 고식적 항암화학요법과의 비교 임상시험을 통하여 EGFR 억제제는 EGFR 돌연변이가 있는 진행성 폐암 환자의 일차치료제로 승인받게 되었다. 이어 2세대, 3세대 표적항암제가 개발되어 약제의 유효성 및 안정성이 더욱 향상되고 있다. EGFR 돌연변이 검출에 근거하여 EGFR 억제제를 선택하는 치료방법은 이후 추가로 발견된 유전자 변이와 해당 표적항암제 개발로 이어지면서 정밀의료의 기본 모델이 되었다.

- ALK 융합 유전자 폐암

2007년 폐암에서 ALK 융합 유전자가 발견되면서 폐암 치료는 또 다른 국면을 맞게 된다. ALK 융합 유전자 발견 이후 불과 3년 만에 ALK 억제제 크리조티닙에 대한 임상시험

결과가 발표되면서, 표적의 발견에서부터 약제 승인에 이르기까지 가장 단기간 내에 안정성과 유효성을 입증한 대표적인 항암제로 알려져 있다. 또한, ALK 융합 유전자를 실제 임상 샘플(포르말린-고정, 파라핀-포매 조직)에서 검출하는 분자 진단법이 항암제와 함께 승인을 받음으로써 ‘동반 진단(companion diagnostics)’이라는 개념을 도입하는 계기가 되었다.

- ROS1 융합 유전자 폐암

ALK 융합 유전자에 이어 ROS1 융합 유전자의 발견 및 이후 진행된 ROS1 억제제의 임상 효능 증명은 폐암 환자에 대한 치료의 폭을 넓히는 계기가 되었다. 특히, ROS1은 ALK와 키나제 부위의 단백질 구조가 매우 유사하여 크리조티닙을 포함한 ALK 억제제들이 ROS1 융합 유전자 폐암 환자에게서도 유사한 유효성과 안정성을 보이게 됨이 알려졌다.

- BRAF 돌연변이 폐암

BRAF V600E 돌연변이는 흑색종을 포함한 다양한 암에서 발견되는 대표적인 종양 유전자 변이이다. 진행성 BRAF V600E 폐암은 다브라페닙 및 트라메티닙 병용요법이 안정성과 유효성을 인정받아 일차 치료 약제로 승인받았다.

- 기타 표적 가능한 유전자 변이들

상기 기술된 유전자 변이 이외에도 RET 융합 유전자, NTRK 융합 유전자, NRG1 융합 유전자, ERBB2 돌연변이, MET 엑손 14번 결손 돌연변이 등이 발견되었고 현재 활발하게 임상 시험 중이다 <표 3.1>. 또한, 그동안 표적항암제의 대상이 되지 못했던 KRAS 돌연변이 폐암에 대하여 신약 효능 결과가 발표되면서, KRAS 돌연변이 폐암 치료의 새로운 장이 열릴 것으로 기대를 모으고 있다. 특히 주목할만한 점은 상기 기술된 종양 유전자 변이들은 같은 종양 내에서 상호 배타적(mutually exclusive)으로 나타난다는 점이다. 즉, EGFR 돌연변이 폐암에서 ALK 융합 유전자 또는 KRAS 돌연변이를 동시에 동반하는 경우는 매우 드물다. 그렇기 때문에, 폐암 세포는 하나의 종양 유전자 및 해당 신호전달 체계에 의존을 하게 되고, 이는 소위 ‘종양 유전자 중독(oncogene addiction)’을 일으켜 표적항암제에 대한 감수성이 증가하게 된다.

- 동반 유전자 변이(co-occurring mutation) 검출

상기 기술된 드라이버 종양 유전자 및 종양 유전자 증독에 의하여 표적항암제에 좋은 반응을 보이지만, 실제 동일한 드라이버 유전자 변이를 가진 환자군에서도 세부적인 약제 반응은 다양하게 나타난다. 이렇게 항암제 반응 및 향후 무진행 생존율에 영향을 미치는 인자로 드라이버 종양 유전자와 동반되어있는 유전자 변이가 영향을 미치는 것으로 나타났다. 종양 유전자는 상호 배타적인 성격으로 서로 동반되지 않으나, TP53, STK11, RB1 등의 종양 억제 유전자 또는 CDK4, CDK6 등의 세포 주기와 관련된 유전자 변이는 동반될 수 있으며, 이러한 동반 유전자 변이는 표적항암제 반응 및 예후에 영향을 미치는 것으로 보고되고 있다. 따라서 단일 유전자 검사뿐만이 아니라 암 유전체를 총체적으로 분석하는 포괄적인 진단 방법이 중요하다고 하겠다.

㉔ 표적항암제 내성과 관련된 유전자 변이 검출

- 내성 기전의 종류

해당 표적을 가지고 있는 폐암 환자는 초기에 표적 항암제에 좋은 반응을 보이지만, 일반적으로 약 12개월 후 약물 내성을 보이고 암은 다시 진행하게 된다. 이러한 약물 내성을 극복하고자 내성 기전을 탐구하는 연구가 진행이 되었다. 대표적인 내성 기전은 최초 진단 당시에는 발견되지 않았던 추가적인 유전자 변이가 발생하는 것이다. 추가적인 돌연변이는 동일 종양 유전자내에 추가적으로 일어나거나(on-target), 다른 종양 유전자 또는 신호전달경로가 활성화되는 것(off-target)이다.

- 온 타겟 기전

같은 종양 유전자내(on-target)에 생기는 대표적인 획득 내성의 예로 EGFR T790M 돌연변이를 들 수 있다. 1,2세대 EGFR 억제제를 사용한 후 약제 내성이 생긴 EGFR 돌연변이 폐암의 약 50%에서 최초 진단 시에 관찰되지 않았던 EGFR T790M 돌연변이가 발견된다. 획득 돌연변이는 ALK 융합 및 ROS1 융합 유전자 폐암에서도 유사하게 발견된다. 대표적인 예로 ALK L1196M, ROS1 G2032R 돌연변이가 있다. 이외에도 매우 다양한 유형의 내성 관련 획득 유전자 변이가 보고되고 있다. 이러한 획득 돌연변이를 억제할 수 있는 3세대 표적 치료제(오시머티닙, 레이저티닙 등)가 개발되어 승인을 받거나 임상 시험 중이다.

- 오프 타겟 기전

획득 내성 기전으로 다른 종양 유전자 또는 신호전달경로가 활성화되는(off-target) 대표적인 예는 EGFR 돌연변이 폐암에서 발생하는 MET 유전자 증폭이다. 이러한 폐암에 대해서 EGFR 억제제 및 MET 억제제 병용요법이 임상 시험 중이다. 이 외에도 ERBB2, AXL, IGF1R 등의 다양한 유전자 활성화 및 융합 유전자 변이의 발생 등이 계속해서 보고되고 있다.

- 포괄적인 유전체 분석의 중요성

앞서 기술한 바와 같이 표적 치료제의 대상이 되는 종양 유발 드라이버 돌연변이, 동반 유전자 변이, 내성 기전과 관계된 유전자 변이는 그 종류 및 변이 발생 위치가 매우 다양하다. 따라서 점 돌연변이, 유전자 복제수 변이, 융합 유전자 등을 동시에 검출할 수 있는 포괄적인 패널 구성 및 유전체 분석이 폐암 환자의 치료에 필수적임을 알 수 있다.

② 면역 항암제 반응을 예측하는 유전체 데이터 도출

㉑ 면역 항암제의 대표적인 예측 표지자

- 면역 관문 억제제(immune checkpoint inhibitor)

항암화학요법, 표적항암요법에 이어 면역항암요법이 개발되어 현재 다양한 암종에서 치료제로 사용되고 있다. 특히, 폐암은 흑색종과 더불어 면역 관문 억제제를 이용한 면역 치료가 가장 먼저 적용되기 시작한 암이다. 면역 항암제는 표적항암제와 비교하여 약물 효과의 지속 기간이 길고 전체 생존율을 향상시키는 것으로 알려져 있다. 그러나 전체 환자에서 면역 항암제에 효과를 보이는 환자는 약 20~30% 정도로 보고되고 있다. 따라서 면역 항암제의 반응을 예측할 수 있는 바이오마커의 중요성은 약물 개발의 초기 단계에서부터 제기되었다.

- PD-L1 단백질 발현

이러한 바이오마커 중 PD-L1 단백질 발현은 면역 항암제의 예측 표지자로 가장 먼저 사용된 표지자이다. 폐암 환자의 진단에 사용하는 포르말린-고정, 파라핀-포매 조직을 이용하여 면역조직화학 염색을 시행하고 발현 정도를 정량적으로 표시한 후, 기준치를

정하여 최종 양성과 음성을 판정한다. PD-L1 단백질 발현과 면역관문 억제제 반응과의 상관관계 분석을 통해서, PD-L1 염색은 동반진단 또는 동반보조진단으로 함께 승인을 받았다. 그러나 PD-L1 발현만으로는 면역 항암제에 반응을 보이는 환자를 정확하게 가려낼 수 없다는 한계가 있다.

㉔ 종양 돌연변이 수(tumor mutation burden, TMB)

- 신생항원(neoantigen)과 TMB

암에 대한 면역 반응은 여러 가지 인자가 관여를 하는데, 그중 하나는 암이 만들어내는 신생항원(neoantigen)이다. 신생항원을 많이 만들어낼수록 암에 대한 면역 반응은 더 강하게 일어난다. 신생항원은 보통 돌연변이 유전자에서 만들어지므로, 돌연변이의 수가 많을수록 면역 반응을 일으키는 신생항원의 수는 증가하게 될 것이다. 이에 근거하여 여러 연구에서 TMB는 면역 항암제의 반응을 예측할 수 있는 표지자로 제시되고 있다.

- TMB의 정의 및 표준화 필요성

TMB의 정의는 전체 엑솜(whole exome)에서의 돌연변이 수인데 보통 메가베이스(megabase)당 돌연변이의 수로 표현한다. TMB는 전체 엑솜 염기서열 분석을 기반으로 도출해야 하나, 300개 이상의 유전자 패널 염기서열 분석결과와 전체 엑솜 분석결과가 서로 상관관계를 보이면서, 다양한 NGS 기반 패널 검사가 사용되고 있다. 또한 각 연구마다 돌연변이에 포함시키는 유전자 변이 유형이 조금씩 상이하여 이에 대한 표준화 필요성이 강하게 대두되고 있다.

- 흡연과의 연관성

폐암은 일반적으로 흡연과의 관련성이 높으며, 흡연자 폐암에서 TMB는 높게 관찰된다. 흡연력이 높은 폐암 환자에서 면역 항암제의 반응이 높게 나타나는 것도 이러한 배경과 관련 있는 것으로 보고되었다.

㉔ 기타 면역 항암제 반응을 예측하는 유전체 표지자

TMB와 함께 면역 항암제의 반응을 예측하는 주요 표지자는 부적합 결합 DNA 교정 (mismatch repair, MMR) 유전자의 기능 상실 또는 현미부수체 불안정성(microsatellite instability, MSI) 여부이다. 이 표지자는 암의 원발 장기와 상관없이 항암제를 사용할 수 있는 최초의 분자 표지자가 되었다. 물론, 폐암에서는 MMR 유전자 결핍이나 MSI 유형의 암종은 매우 드문 것으로 보고되고 있고, 주로 대장암, 위암, 자궁내막암에서 관찰된다. 그러나 이러한 유전체 변화는 모든 장기에서 관찰될 수 있으므로, 암 환자의 진단과 치료에서 암 유전체 분석의 중요성은 더욱 증가하였다고 결론지을 수 있다. TMB, MSI 이외에도 사람 백혈구 항원(Human Leukocyte Antigen, HLA)의 변이, T 세포 수용체 (T-cell receptor, TCR) 변이 등이 관여하는 것으로 알려지고 있으며, 해당 분야의 연구가 활발히 진행 중이다.

나) 조기(early stage) 폐암에 대한 적용

(1) 조기 폐암 환자를 위한 유전체 분석의 필요성

최근 영상 기술의 발전 및 건강 검진의 증가로 조기 폐암의 빈도가 증가하고 있다. 조기 폐암의 표준 치료 지침은 완전한 수술적 절제이다. 그러나 완전한 수술적 절제를 받은 1~3기의 조기 폐암 환자 중에서 약 20~50%의 환자는 암 재발을 경험하고 있다. 또한 수술 후 재발한 환자의 예후는 매우 불량한 것으로 보고되었다. 이러한 조기 폐암 환자의 예후 예측과 치료방법 결정에 있어서 기존의 임상-병리학적 인자만으로는 잠재적인 전이, 재발 등을 예측하기 어려우므로 유전체 데이터 분석에 근거한 활발한 연구가 진행 중이다.

(2) 수술 후 재발 관련 인자 및 추가 치료 여부 결정

(가) 수술 검체를 이용한 유전체 분석

치료 목적의 수술적 완전 절제를 시행하고, 그 이후 추적 관찰에서 재발 및 생존율을 예측하기 위한 표지자 및 그에 근거한 추가 치료 여부 결정은 조기 폐암 환자에게 정밀 의료를 실현하기 위해 매우 중요한 사항이라 할 수 있다. 이러한 표지자를 발굴하기 위해 일차적으로 수술로 절제된 폐암에 대한 유전체 분석이 진행되고 있다.

(나) DNA 염기서열 분석

수술 검체의 유전체 분석은 먼저 DNA 염기서열을 통하여 앞서 기술한 바 있는 드라이버 종양 유전자의 돌연변이뿐만이 아니라 동반 유전자 변이도 검출하는 것이다. 이러한 분석을 통하여 세포 주기, Hippo, TGF, p53과 관련된 경로에 돌연변이가 동반되어 활성화되어 있을 때, 예후가 불량함이 보고되었다. 또한 이러한 변이로 활성화된 신호전달 경로의 개수가 많을수록 재발이 빈번함이 보고되었다.

(다) RNA 염기서열 분석

RNA 염기서열 분석이나 mRNA 발현 분석을 통하여 재발 관련 인자를 발굴할 수 있다. 이러한 유전자 발현 검사는 최소 네 개 이상의 유전자 패널을 구성하여 조사한다. 대표적으로 보고된 패널로는 4개 유전자 패널(RHOV, CD109, FRRS1, and LINC00941), 10개 유전자 패널(NUDCD1, E2F1, HOXB7, MCM6, SERPINB5, E2F4, HSPG2, SF3B1, RRM2, and SCGB3A1) 등이 있다.

(2) 수술 후 미세 잔류 질환(minimal or molecular residual disease, MRD) 검출 및 암 재발 모니터링

(가) MRD의 의의

수술로 완전 절제를 받은 폐암 환자들에서 재발이 빈번히 관찰되는 것은 미세 잔류 암세포들이 인체에 남아 있기 때문이며 이를 MRD이라 지칭한다. 이러한 MRD는 수술 후 환자의 예후와 관련이 높은 것으로 알려져 있어 이를 검출하기 위한 연구가 진행되고 있다. 또한, 현재 수술받은 폐암 환자는 정기적인 영상 검사를 통해 재발 여부에 대한 추적 관찰을 받는다. 그러나 영상 단계에서 암 재발이 발견되었을 경우 이미 4기에 해당하는 경우가 많고, 그 이후 전신적인 항암 치료를 받더라도 예후가 불량하다.

(나) 순환 종양 DNA(circulating tumor DNA, ctDNA)를 이용한 MRD 검출

MRD를 검출하기 위해 ctDNA를 이용할 수 있음이 보고되었다. ctDNA는 혈액 내에 존재하는 세포 유리 DNA(cell free DNA)로써 암세포에서 도래한 DNA 가닥들이다. ctDNA 분석은 암 조직을 이용한 유전체 분석 방법보다 기술적인 측면과 생물학적 측면에서 더

많은 주의가 필요하다. 기술적인 측면은 혈액에 매우 미량으로 존재하는 ctDNA를 어떻게 검출해 낼 수 있느냐의 문제이며, 생물학적 측면은 발견된 체세포 유전자 변이가 해당 암세포에서 도래하였는지 확인할 수 있느냐의 문제이다. 왜냐하면 연령이 올라갈수록 혈액 내에는 클론성 조혈(clonal hematopoiesis)에 의한 체세포 유전자 변이가 빈번하게 관찰될 수 있기 때문이다. 따라서 ctDNA를 이용한 유전체 분석은 이 두 가지 측면에 대한 고려 및 제한점을 극복하기 위한 보완이 뒷받침되어야 한다.

(다) ctDNA 분석용 개인 맞춤 패널

폐암 수술을 받은 환자의 경우, 해당 수술 검체를 이용하여 해당 폐암의 유전체 분석을 시행하고 이에 근거하여 추적 관찰용 ctDNA 패널을 제작할 수 있다. 환자별로 폐암 조직 검체의 유전체 분석에 근거하여 ctDNA 용 패널을 제작할 경우, 심층 염기서열(deep sequencing)이 가능하여 검출율이 더 높아지는 것으로 보고되었다. 이는 영국에서 진행되고 있는 TRACERx(TRACKing non-small cell lung Cancer Evolution through therapy [Rx]) 연구에서 검증되었다.

다) 폐암 조기 진단을 위한 유전체 데이터 적용

(1) 기존 선별 검사(screening test)의 한계

미국에서 진행된 국가 폐 선별 시험(the National Lung Screening Trial)에서 저선량 흉부 컴퓨터 단층촬영(low-dose CT) 검사는 단순 흉부 촬영에 비하여 폐암에 의한 사망률을 20% 정도 낮추었다. 그러나 이러한 영상 검사는 폐암이 아닌 단순 폐 결절을 다수 발견하는 문제점이 있었다. 또한, 실제 흡연자로 구성된 고위험군들이 이러한 영상 검사를 정기적으로 받기가 쉽지 않았다. 객담 검사나 기존에 사용되는 혈액 표지자 검사는 민감도가 낮다는 문제점이 있었다.

(2) ctDNA 유전체 분석을 이용한 폐암의 조기 진단

앞서 기술한 ctDNA 염기서열 분석을 통해서 폐암을 조기에 진단하고자 하는 연구들이 계속 진행 중이다. 최신 연구결과 중 하나는 스탠포드 암연구소에서 개발한 Cancer Personalized Profiling by deep sequencing(CAPP-Seq)으로 NGS 표적 심층 염기서열을 이용한 검사법이다. 이 방법을 이용하여 연구자들은 1기, 2기, 3기 폐암 환자의 수술 전

채취한 혈액에서 각각 42%, 67%, 88% 비율의 폐암 ctDNA를 발견하였다고 보고하였다. 또한, ctDNA에 대한 기계 학습 알고리즘을 통해서 고위험 환자군에서 실제 폐암 환자를 더 정확하게 구별해 낼 수 있었다. 향후 ctDNA 분석을 이용한 암 조기진단 연구는 더욱 심화될 것으로 보인다.

3) 결론 및 향후 전망

이상, 폐암 환자의 진단과 치료에 있어 유전체 염기서열 분석 데이터의 임상적용을 살펴보았다. 폐암을 병기에 따라 구분하여 진행성 폐암에서의 적용, 조기 폐암에서의 적용, 그리고 마지막으로 폐암의 조기 진단을 위한 적용 부분으로 나누어 살펴보았다. 현재, 폐암 조직을 이용하여 암 유전체를 분석하고 그 결과에 따라 항암제를 선택하는 개인맞춤-정밀의료가 활발히 진행 중이다. 또한 ctDNA 분석을 통하여 치료 반응을 모니터링하고, 미세 잔류 질환 여부를 평가하고, 폐암을 조기에 진단하고자 하는 노력이 더 심화될 것으로 보인다. 따라서 유전체 염기서열 분석 데이터는 암 확진 이전 단계인 선별 검사에서부터 진단, 병기 결정 및 예후 예측, 치료 선택, 이후 추적 관찰에 이르기까지 다양한 측면에서 광범위하게 적용될 것으로 전망된다 <그림 3.3>. 이에 따라, 궁극적으로, 유전체 염기서열 분석 데이터는 폐암으로 인한 사망률을 낮추고 폐암 환자의 삶의 질 향상에 기여할 것이다.

그림 3.3 유전체 염기서열 분석 데이터의 적용 분야



참고문헌

- Abbosh, C., Birkbak, N. J., Swanton, C.(2018). “Early stage NSCLC - challenges to implementing ctDNA-based screening and MRD detection”. *Nature Reviews Clinical Oncology*, 15(9), pp.577-586. (DOI: 10.1038/s41571-018-0058-3).
- Abbosh, C., Birkbak, N. J., Wilson, G. A. *et al.*(2017). “Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution”. *Nature*, 545, pp.446-451. (DOI: 10.1038/nature22364).
- Blakely, C. M., Watkins, T. B. K., Wu, W. *et al.*(2017). “Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers”. *Nature Genetics*, 49(12), pp.1693-1704. (DOI: 10.1038/ng.3990).
- Canon, J., Rex, K., Saiki, A. Y. *et al.*(2019). “The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity”. *Nature*, 575, pp.217-223. (DOI: 10.1038/s41586-019-1694-1).
- Chabon, J. J., Hamilton, E. G., Kurtz, D. M. *et al.*(2020). “Integrating genomic features for non-invasive early lung cancer detection”. *Nature*, 580, pp.245-251. (DOI: 10.1038/s41586-020-2140-0).
- Havel, J. J., Chowell, D., Chan, T. A.(2019). “The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy”. *Nature Reviews Cancer*, 19(3), pp.133-150. (DOI: 10.1038/s41568-019-0116-x).
- NCCN(2020). National Comprehensive Cancer Network(NCCN) Guidelines, Non-Small Cell Lung Cancer, Version 3. 2020, www.nccn.org
- Park, E., Shim, H. S.(2020). “Detection of Targetable Genetic Alterations in Korean Lung Cancer Patients: A Comparison Study of Single-Gene Assays and Targeted Next-Generation Sequencing”. *Cancer Research and Treatment*, 52(2), pp.543-551. (DOI: 10.4143/crt.2019.305).

- Reck, M., Rabe, K. F.(2017). “Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer”. *New England journal of medicine*, 377(9), pp.849-861. (DOI: 10.1056/NEJMra1703413).
- Rotow, J., Bivona, T. G.(2017). “Understanding and targeting resistance mechanisms in NSCLC”. *Nature Reviews Cancer*, 17(11), pp.637-658. (DOI: 10.1038/nrc.2017.84).
- Shim, H. S., Choi, Y. L., Kim, L. *et al.*(2017). “Molecular Testing of Lung Cancers”. *Journal of pathology and translational medicine*, 51(3), pp.242-254. (DOI: 10.4132/jptm.2017.04.10.).
- Shukla, S., Evans, J. R., Malik, R. *et al.*(2017). “Development of a RNA-Seq Based Prognostic Signature in Lung Adenocarcinoma”. *Journal of the National Cancer Institute*, 109(1). (DOI: 10.1093/jnci/djw200).
- Skoulidis, F., Heymach, J. V.(2019). “Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy”. *Nature Reviews Cancer*, 19(9), pp.495-509. (DOI: 10.1038/s41568-019-0179-8).
- Zhou, J., Sanchez-Vega, F., Caso, R. *et al.*(2019). “Analysis of Tumor Genomic Pathway Alterations Using Broad-Panel Next-Generation Sequencing in Surgically Resected Lung Adenocarcinoma”. *Clinical Cancer Research*, 25, pp.7475-7484. (DOI: 10.1158/1078-0432.CCR-19-1651).

나. 폐암 이외의 고형암

1) 서론

정밀의료는 각 환자 개인별 유전체 분석결과를 기반으로 하여 의료진이 치료법 선택에 도움을 받는 진료 접근을 말한다. 맞춤의료라고 불리기도 하는 이러한 의료개념은 새로운 것이 아니라 최근 과학과 기술의 발전으로 인하여 가능해진 면이 있다.

정밀의료는 환자, 진료의사, 의과학연구자, 국가 등 참여 주체 모두가 이득을 얻을 수 있는 최선의 의료 시스템이다. 환자는 자신에게 맞는 약이나 치료방법을 처음부터 시도할 수가 있어 부작용을 최소화하면서 최선의 치료결과를 얻을 수 있다. 진료하는 의사도 자신이 치료하는 환자의 치료결과가 좋아야 의료인으로서 긍지와 보람을 느끼게 되고, 새로운 치료 및 진단법을 개발하는 연구자들은 효과적인 신약개발을 적은 비용과 짧은 기간 안에 이룰 수 있다. 국가는 국민의 난치성 질환의 효과적인 치료를 한정된 건강보험 재정으로 이룰 수 있는 최선의 방법이 맞춤치료이다.

개인별 맞춤치료의 가장 앞선 분야는 항암치료 분야이다. 과거 세포독성 항암치료제는 부작용은 심하면서, 치료효과는 투여 받은 암 환자의 10~40%만이 일시적인 암병변의 호전을 기대할 수 있는 정도였다. 세포독성 항암치료제는 암의 원발 부위에 따라 (예 : 위암, 대장암, 간암, 폐암, 유방암 등) 약제를 병용하여 투여하였으나, 암세포의 분자생물학적 특이 변이를 찾아 공략하는 표적항암제가 개발되면서 개인별 맞춤치료의 시대가 열렸다. 최근 암정밀의료는 암종별 치료 약제 개발이 아닌 tissue-agnostic(조직과 무관한) 신약개발 및 허가가 대세를 이루고 있다. 암의 발생 원인이 선천적 유전성 배경을 바탕으로 후천적 발암환경에 노출되어 다양한 장기에 개인별 위험성이 더해져서 암이 발생한다는 원리를 이해한다면 당연한 현상이다. 이러한 개념은 과거 조직학적 진단과 분류에 집착하는 신약개발 프레임을 뒤집으며 새로운 암정밀의료 생태계 구축을 추구하고 있다.

2) 암정밀의료의 다양한 발전

가) 드라이버 암유전자의 발견 및 표적치료

많은 연구자들이 암세포의 유전자 변이 특성을 분석한 결과 각 환자 개인별로, 동일 환자에서 암세포의 위치별로 다양한 양상을 보이며 동일 환자의 체내에서 암세포가 시간의

경과에 따라 진화를 보이며 유전자 변이를 새로이 획득해 가는 현상을 확인하게 되었다. 이러한 암세포 유전자 변이의 다양성과 진화 개념은 더욱 각 개인별 정밀의료의 필요성을 확인시켜 준 계기가 되었다.

모든 암의 발생과 진행은 정상 조직과 달리 암조직에만 나타난 특정 유전자 변이(암 유전자)에 의하여 나타난다. 암정밀의료는 암조직의 DNA 분석을 통하여 해당 암 환자의 암세포에서 주된 역할을 하는 유전자 변이(드라이버 암유전자)를 찾아내어 이러한 유전자의 작용을 선택적으로 차단하는 약제(표적치료제)를 사용하여 매우 효과적인 치료 효과를 얻고 있다. 대표적인 드라이버 암유전자와 표적치료제는 <표 3.2>과 같다.

다른 암발생 관련 유전자 변이는 암의 발생을 억제하는 유전자(암억제유전자)가 손실되는 현상으로 선천적으로 부모로부터 물려받은 암억제유전자의 이상과 더불어 후천적 추가 손상으로 인하여 암발생을 막지 못하는 경우이다. 이러한 가족성 암증후군 환자들은 여러 종류의 암이 다발성으로 발생하는 경향을 보인다. 대표적인 암억제유전자와 가족성 암증후군은 <표 3.3>과 같다.

표 3.2 미국 FDA와 유럽 EMA 승인을 받은 표적항암제와 타겟 유전자 변이 진단법

Gene/protein	Anticancer agent	Indications	Biomarker	Routine testing
ALK	Crizotinib, ceritinib, alectinib, lorlatinib, brigatinib	NSCLC	ALK translocation	FISH, IHC
Androgen receptor (AR)	Abiraterone, enzalutamide, dalurotamide, apalutamide	Prostate cancer	AR expression	IHC
BCL-2	Venetoclax	Chronic myeloid leukemia	BCL-2 protein expression, BCL-2 amplification/translocation	IHC (FISH, DNA/RNA sequencing), PCR
BCR/ABL	Imatinib, dasatinib, nilotinib, bosutinib, ponatinib	Chronic myeloid leukemia	BCR/ABL1 fusion	IHC, PCR, DNA sequencing
BRAF	Dabrafenib+trametinib, vemurafenib+cobimetinib, encorafenib+binimetinib	Melanoma, NSCLC, anaplastic thyroid cancer, hairy cell leukemia	BRAF V600E/K mutations	IHC, PCR, DNA sequencing
C-KIT, PDGFR	Imatinib	Gastrointestinal stromal tumor	c-KIT Exon 9 and 11 mutations, PDGFR mutations	IHC, DNA sequencing
PDGFRB	Imatinib	Myelodysplastic/myeloproliferative syndromes	PDGFRB rearrangement	FISH

Gene/protein	Anticancer agent	Indications	Biomarker	Routine testing
BRCA	Olaparib, talazoparib, rucaparib	Breast cancer, ovarian cancer, prostate cancer	Germline/somatic BRCA 1/2 mutations	DNA sequencing
Estrogen/progesterone receptors (ER/PR)	Tamoxifen, raloxifene, fulvestrant, toremifene	Breast cancer	ER/PR expression	IHC
erBB2/HER-2	Trastuzumab, pertuzumab, ado-trastuzumab, emtansine, neratinib	Breast cancer, gastric cancer	HER-2 protein expression, HER-2 amplification	IHC, FISH
EGFR	Gefitinib, erlotinib, afatinib, dacomitinib	NSCLC	EGFR exon 19 deletion, exon 21 L858R mutation	DNA sequencing, PCR
	Osimertinib		EGFR T790M mutation	
FGFR2/3	Erdafitinib	Bladder cancer	FGFR3 mutations, FGFR2/3 fusions	DNA sequencing, FISH
FLT3	Midostaurin, gilteritinib	Acute myeloid leukemia	FLT3 mutations	DNA sequencing, PCR
IDH1/2	Ivosidenib, enasidenib	Acute myeloid leukemia	IDH1/2 mutations	IHC, DNA sequencing
MET	Crizotinib	NSCLC	MET amplification, MET exon 14 alterations	FISH, DNA/RNA sequencing
MSI-H or dMMR	Pembrolizumab	MSI-H or dMMR solid tumors	MLH1, MSH2, MSH6, PMS2 protein expression, MSI high	IHC, DNA sequencing, PCR
	Nivolumab and ipilimumab	Colorectal cancer		
NTRK	Larotrectinib, entrectinib	Solid tumors with NTRK fusions	NTRK protein expression, NTRK fusion	IHC, FISH, DNA/RNA sequencing
PI3KCA	Alpelisib	Breast cancer	PI3KCA mutation	DNA sequencing
PI3KCA (alpha and delta)	Copanlisib	Follicular lymphoma	PI3K mutation	DNA sequencing
PI3K (delta and gamma)	Duvelisib	Chronic lymphocytic leukemia, small lymphocytic lymphoma	PI3K mutation	DNA sequencing
RAS (negative predictor)	Cetuximab, panitumumab	Colorectal cancer	KRAS/NRAS wildtype	DNA sequencing
RET	LOXO-292	NSCLC, medullary thyroid cancer	RET fusion, RET mutation	FISH, DNA/RNA sequencing
ROS1	Crizotinib, entrectinib	NSCLC	ROS translocation	FISH, DNA/RNA sequencing

AR androgen receptor, *dMMR* deficient mismatch repair, *ER* estrogen receptor, *FISH* fluorescence in situ hybridization, *IHC* immunohistochemistry, *MSI-H* high levels of microsatellite instability, *NSCLC* non-small cell lung cancer, *PR* progesterone receptor

자료: Malone, E. R. *et al.*, 2020

표 3.3 대표적인 암억제유전자와 가족성 암증후군

Syndrome	Gene	Major Phenotypic feature	Mode of inheritance
Hereditary Breast/Ovarian Cancer (HBOC) 1	BRCA1	Female breast, ovarian cancer	Dominant
Hereditary Breast/Ovarian Cancer (HBOC) 2	BRCA2	Male and female breast, ovarian cancer	Dominant
Moderate Risk Breast/Ovarian	CHEK2, ATM, NBS1, RAD50, BRIP1, PALB2, FGFR2, TNRC9, MAP3K1, LSP1, CASP8	Breast and ovarian cancer	
Familial Adenomatous Polyposis (FAP)	APC	Multiple colorectal and upper GI adenomatous polyps	Dominant
Hereditary Non-Polyposis Colorectal Cancer (HNPCC)	MLH1, MSH2, MSH6, PMS2, PMS1	Colorectal, endometrial cancer	Dominant
MUTYH Associated Polyposis	MYH	Colorectal polyps and colorectal cancer	Recessive
Multiple Endocrine Neoplasia (MEN)1	MEN1	Parathyroid adenomas, entero-pancreatic tumors, pituitary tumors	Dominant
Multiple Endocrine Neoplasia (MEN)2	RET	Medullary thyroid carcinomas, pheochromocytomas	Dominant
Li-Fraumeni	TP53	Breast cancer, sarcomas, leukemia and brain tumors	Dominant
Turcot	APC, MLH1, MSH2	Colorectal polyps, medulloblastomas	Dominant
Juvenile polyposis	SMAD4, BMPR1A	Hamartomatous intestinal polyps	Dominant
Muir-Torres	MLH1, MSH2	Adenomatous colorectal polyps	Dominant
Von-Hippel-Lindau	VHL	Renal cell carcinoma	Dominant
Peutz-Jeghers	STK11	Hamartomatous intestinal polyps	Dominant
Hereditary Diffuse Gastric Cancer	CDH1	Diffuse gastric cancer	Dominant
Cowden	PTEN	Breast, gastrointestinal cancer	Dominant
Bannayan-Ruvalcaba-Riley	PTEN	Hamartomatous intestinal polyps	Dominant
Familial Retinoblastoma	RB1	Primary eye cancer	Recessive
Wilm's tumor	WT1	Kidney cancer	Recessive
Familial melanoma	CDKN2A	Melanoma	Dominant
Axia Telangiectas	ATM	Lymphoma, breast cancer	Recessive
Bloom syndrome	BLM	Solid tumors	Recessive
Xeroderma Pigmentosum	XPB, XPD, XPA	Skin cancer	Recessive
Fanconi's anemia	FANC (A,B,C,D1,D2,E,F,G,I,J, L,M,N) FANCD1/BRCA2, FANCI/BACH1/BRIP1, FANCG/PALB2	Acute myeloid leukaemia	Recessive

자료: Fostira, F. *et al.*, 2007

(1) HER2 유전자 변이

HER2는 EGFR family의 receptor tyrosine kinase로써 HER2 유전자의 증폭, 혹은 과발현은 침샘암(17~44%), 유방암(20%), 위암(20%), 난소암(27%), 자궁암(18~80%), 자궁경부암(21%), 폐암(2.5%), 담도암(20%), 췌장암(26%), 직결장암(5%), 방광암(12.4%), 전립선암(10%) 등 다양한 암종에서 확인되며, HER2 유전자의 돌연변이는 이들 암종의 1~9%에서 검출된다. HER2를 억제하는 치료법의 개발은 1990년도에 HER2 유전자의 증폭, 혹은 과발현이 있는 유방암 환자를 대상으로 trastuzumab이라는 anti-HER2 단클론 항체가 개발되어 효능을 입증하며 시작되었다. 이후 pertuzumab이라는 anti-HER2 IgG1 단클론항체, EGFR과 HER2 tyrosine kinase를 모두 차단하는 lapatinib, trastuzumab과 tubulin-binding 약제인 DM1을 결합시킨 T-DM1, trastuzumab과 tubulin-binding 약제인 deruxtecan을 결합시킨 Enhertu 등 약제가 HER2 유전자의 증폭, 혹은 과발현이 있는 유방암 환자에 보조요법, 신보조요법, 전이성 1차, 전이성 2차 사용이 허가되었다. 유방암에 이어 HER2 유전자의 증폭, 혹은 과발현이 있는 전이성 위암 환자를 대상으로 trastuzumab과 fluoropyrimidine + platinum 병용이 1차 치료제로 허가되어 사용되고 있으나 다른 anti-HER2 약제들의 전이성 2차, 보조요법으로의 사용은 아직 임상시험 중이거나 임상시험에 실패하였다. 담도암, 폐암, 직결장암, 방광암 환자 중 HER2 유전자의 증폭, 혹은 과발현이 있거나 pathogenic 변이가 있는 환자를 대상으로 basket trial, 혹은 소규모의 2상 임상시험에서 10~40%의 반응률을 보여 anti-HER2 치료법의 가능성을 보였다. 그러나 위암 등 타 고형암에서는 HER2 유전자의 증폭, 혹은 과발현이 있다 하여도 유방암과 달리 anti-HER2 약제들의 효과가 뚜렷하지 않는데, 이는 HER2 발현의 패턴이 암종별로 차이가 있고 종양 내 HER2 발현 암세포의 다양성, trastuzumab 치료 후 HER2 발현 암세포의 소실 정도에 차이가 있기 때문으로 설명하고 있다.

(2) BRAF 유전자 변이

BRAF 단백질은 활성화되면 dimer를 형성하고 serine/threonine-specific protein kinase로 기능한다. 이 dimers들이 RAS GTPases에 의해 활성화되고, RAF가 세포막에 결합된다.

BRAF 유전자의 돌연변이는 다양한 암의 주요 드라이버 암유전자로 흑색종의 약 60%에서 BRAF 돌연변이가 확인된다. BRAF 돌연변이는 비호지킨 림프종, 직결장암, 유두상 갑상선암, 비소세포폐암, 교모세포종 등에서도 보고되어 있다.

BRAF 유전자는 현재까지 약 30종류의 돌연변이가 세포의 기능 변화를 초래하는 것으로 밝혀졌으며 3그룹으로 분류된다 <표 3.4>. Class 1 BRAF 돌연변이는 RAS-independent 하게 monomer로 작용하는 반면 Class 2 BRAF 돌연변이는 dimer로 작용을 하며, Class 1과 Class 2 BRAF 돌연변이 단백질 모두 상위 조절단백인 RAS GTPases의 영향 없이 지속적으로 암세포의 성장을 초래한다. 이에 반해 Class 3 BRAF 돌연변이 단백질은 적절한 활성화를 위해서는 RAS signaling이 필요하다. 따라서 Class 1 또는 Class 2 BRAF 돌연변이와 달리 Class 3 BRAF-돌연변이 암은 상위 RAS signaling을 차단하는 치료적 접근이 가능하다.

표 3.4 대표적인 암 관련 BRAF 돌연변이

Class	RAS Dependent	Missense Mutations
Class 1	No	V600E, V600K, V600D, V600R, V600M, etc.
Class 2	No	K601E, K601N, K601T, L597Q, L597V, G469A, G469V, G469R, G464V, etc.
Class 3	Yes	D287H, V459L, G466V, G466E, G466A, S467L, G469E, N581S, N581I, D594N, D594G, D594A, D594H, F595L, G596D, etc.

자료: Zaman, A. *et al.*, 2019

Vemurafenib은 최초의 BRAF V600E 특이 억제제로써 흑색종에 FDA 사용허가를 받은 약제이다. 이어 dabrafenib이 BRAF 돌연변이 양성 흑색종 환자를 대상으로 한 3상 임상시험에서 뚜렷한 반응률과 생존기간의 향상을 입증하여 2013년에 허가되었다. BRAF V600E 돌연변이를 가진 hairy cell carcinoma에서도 비슷한 치료효과가 나타난 반면 직결장암과 갑상선암은 BRAF V600E 돌연변이를 가지고 있음에도 치료효과가 뚜렷하지 않았다. 오히려 일부의 BRAF 돌연변이로 인한 암들은 BRAF를 차단하였을 때 역설적으로 MAPK pathway가 활성화되어 피부의 편평상피세포암이 발생하는 환자들이 있다. 이들은 이차적으로 HRAS 돌연변이가 발생하고 이로 인해 MAPK pathway가 역설적으로 활성화되어 2차 암이 발생하는 것으로 설명된다. 또한 BRAF 억제제는 단독 투여 시 내성 발생으로 반응유지기간이 짧다는 단점이 있다. 이러한 단점을 극복하고자 흑색종에서

MEK 억제제인 trametinib과 병용 투여한 결과, 암의 진행기간 및 생존기간이 뚜렷이 향상되었으며 MEK 억제제 투여시 나타났던 MAPK pathway의 역설적 활성화도 나타나지 않았다. 현재까지 BRAF와 MEK 억제제의 병용투여로는 trametinib + dabrafenib, cobimetinib + vemurafenib 병용투여가 BRAF V600E 흑색종 치료요법으로 FDA 승인을 받았다. BRAF 돌연변이 전이성 대장암에서는 dabrafenib과 trametinib 병용 투여로 12% 정도의 반응률을 볼 수 있었으나 BRAF, MEK, 그리고 EGFR을 동시에 함께 차단하면 반응률이 상승한다는 관찰을 근거로 3상 임상시험을 진행하여 encorafenib과 cetuximab 병용투여가 FDA로부터 BRAF V600E 양성 대장암의 치료로 허가되었다. 이러한 임상시험 결과는 EGFR을 통한 MAPK 재활성화를 차단하는 것이 BRAF 억제 과정에서 필요하다는 것을 반증한다.

(3) NTRK 융합 유전자 변이

Neurotrophic tropomyosin receptor kinase(NTRK) 융합유전자는 NTRK1, NTRK2, 또는 NTRK3 유전자가 다른 유전자와 융합이 된 경우로 많은 성인 고형암과 소아암에서 드라이버 암유전자로 작용한다. NTRK 융합유전자가 생성되면 수용체에 ligand의 결합이 없어도 TRK kinase가 지속적으로 활성화되며, 직결장암과 유두상 갑상선암에서 처음 보고되었다. NTRK 융합유전자가 발견되는 암종은 크게 두 가지로 나뉘는데, 첫째는 NTRK 융합유전자가 흔하게 검출되는 희귀암으로 분비성 유방암, 침샘의 유선 유사 분비성 암종, 선천중배엽성신장종, 유아형 섬유육종 등 희귀암의 90%에서 발견된다. 다른 종류는 다빈도 암에서 아주 낮은 확률로 검출되거나(<5%), 유두상 갑상선암, Spitzoid 종양, KIT 음성 소화관기질성종양, 소아성 교종 등에서는 5~25%의 빈도로 나타난다. NTRK 융합유전자는 폐선암, 췌선암, 두경부편평상피세포암, 유방암, 직결장암, 담도암, 신장암, 흑색종, 성인의 뇌종양 GIST를 제외한 연조직육종 등에서는 1% 미만의 낮은 빈도를 보인다. 드물게 급성 림프구성 백혈병과 급성 골수성 백혈병 등 혈액암에서도 발견된다. 가장 대표적인 NTRK 특정억제제는 larotrectinib으로 세 가지 TRK 단백을 강력히 선택적으로 억제한다. 2018년 11월 미국 FDA는 larotrectinib을 절제 불가능하거나 전이성 NTRK 융합 양성 성인 및 소아고형암에 허가하였다. Larotrectinib의 허가는 LOXO-TRK-14001(NCT02122913), SCOUT(NCT02637687), 그리고 NAVIGATE(NCT02576431)라는 3개의 다기관 공개 임상시험의 결과를 근거로 하였으며,

총 55명의 NTRK 융합 양성 성인 및 소아고형암 환자 중 반응률 75%, 완전반응률 22%의 뛰어난 효과를 보였다. 반응유지기간도 73%의 환자는 6개월 이상, 39%에서는 1년 이상을 나타내었다. 2019년 8월 entrectinib이 두 번째로 허가되었으며 54명의 대상 환자 중 반응률 57%, 6개월 이상 반응유지기간 68%, 1년 이상 반응유지기간 45%를 보였다. Crizotinib, cabozantinib, lestaurotinib, altiratinib, foretinib, ponatinib, nintedanib, merestinib, MGCD516, PLX7486, DS-6051b 그리고 TSR-011 등 멀티키나제 억제제들도 TRK 억제 효과가 있어 임상 개발 중이다.

(4) KRAS 억제제

KRAS와 다른 RAS의 동형 단백질들은 사람의 암에서 가장 흔하게 나타나는 암유전자이지만 많은 노력에도 불구하고 RAS를 억제할 수 있는 약제의 개발에는 아직까지 성공하지 못하였다. 사람에서는 HRAS, NRAS 그리고 KRAS 등 3가지 RAS 단백질이 알려져 있으며 KRAS는 RAS superfamily 또는 RAS-like GTPases라고 알려져 있는 small guanosine triphosphate(GTP) binding 단백질 그룹에 속한다. KRAS는 세포 표면에서 핵으로 신호전달 분자들의 활성을 초기에 시작하는 센서 역할을 하며, 세포의 분화, 성장, 주화성, 세포자멸사 등 광범위한 기능에 관련된 신호전달 역할을 한다. 세포 외부의 자극에 의하여 불활성 RAS-GDP가 RAS-GTP로 활성화되면 mitogen-activated protein kinase(MAPK) pathway, phosphoinositide 3-kinase(PI3K) pathway, 그리고 Ral-GEFs pathway 등으로 활성이 전달된다.

RAS 유전자의 돌연변이는 암종별로 차이가 있는데, KRAS가 돌연변이가 가장 많은 isoform으로 RAS 돌연변이의 86%를 차지한다. KRAS 돌연변이는 췌장암의 90%, 직결장암 환자의 30~40%, 비소세포폐암 환자의 15~20%에서 발견된다. 이외에도 담도암, 자궁내막암, 자궁경부암, 방광암, 간암, 골수성백혈병과 유방암에서도 발견된다 <표 3.5>.

표 3.5 각종 암에서의 RAS isoform 돌연변이의 빈도

Primary tissue	KRAS(%)	HRS(%)	NRAS(%)	Total (%)
Pancreas	90	0	<1	90
Colon	30-50	1	6	42
Small intestine	35	0	<1	35
Biliary tract	26	0	2	28
Endometrium	17	<1	5	22
Lung	19	<1	1	20
Skin (melanoma)	1	1	18	20
Cervix	8	9	2	19
Urinary tract	5	10	1	16

자료: Liu, P. *et al.*, 2019

KRAS 유전자의 돌연변이는 주로 코돈 12, 13, 또는 61에서 검출된다. 낮은 빈도로 코돈 63, 117, 119, 그리고 146에서 나타나기도 한다. RAS 유전자의 동일 부위 돌연변이라 하여도 치환되는 아미노산에 따라 생물학적 기능에 차이가 있어 암의 경과나 치료에 대한 반응이 다를 수 있다. 예를 들어, 직결장암과 폐암에서 KRAS 코돈 12의 glycine이 valine으로 치환된 G12V 돌연변이는 glycine이 aspartic acid으로 치환된 G12D 돌연변이에 비하여 예후가 나쁘다고 알려져 있다. KRAS 유전자의 돌연변이는 암발생에 기여할 뿐만 아니라 종양의 미세환경에도 관여하여 종양 주변 간질세포로부터 각종 사이토카인과 성장인자를 분비토록 함으로써 암의 진행 및 유지에 기여한다.

돌연변이로 활성화된 KRAS의 기능을 억제하려는 많은 노력이 대부분 실패하여 KRAS를 타겟으로 하는 약제 개발은 불가능한 것으로 생각되어 왔다. 대신에 RAS의 하위 신호 전달체계인 MAPK, 혹은 PI3KA를 억제하는 약물개발에 집중하였다. BRAF 억제제인 vemurafenib과 dabrafenib, MEK1/MEK2 억제제인 trametinib과 combimetinib, 그리고 하위 transcription factors인 Fos-like antigen 1(FOSL1) 억제제 등이 KRAS 돌연변이형 폐암과 췌장암을 대상으로 시도된 바 있다. 최근에는 KRAS 표면의 수용체를 직접 타겟으로 하는 약물개발이 새로운 가능성으로 떠오르고 있다. 특히 새로운 약물개발 전략은 KRAS-G12C 암단백질을 직접 억제하는 약물개발로써 KRAS-G12C는 전체 KRAS 코돈 12 돌연변이의 10~20%를 차지하며 KRAS 돌연변이 양성 폐선암의 50%에서 발견된다. KRAS-G12C 억제제로 개발 중인 약제는 ARS-853과 ARS-1620이 대표적이다. 최소 5가지 이상의 KRAS 억제제가 임상개발 중이며 이중 3개는 G12C 돌연변이만을 타겟으로 하고 있다 <표 3.6>.

표 3.6 선택적 KRAS 억제제 개발 현황

Drug	Company	Properties	Status
AMG 510	Amgen	G12C inhibitor	Phase I/II, monotherapy and with PD1 blocker
MRTX849	Mirati Therapeutics	G12C inhibitor	Phase I/II
JNJ-74699157/ ARS-3248	J&J and Wellspring Biosciences	G12C inhibitor	Phase I
BI 1701963	Boehringer Ingelheim	KRAS-SOS1 inhibitor	Phase I, monotherapy and with MEK inhibitor trametinib
mRNA-5671	Moderna Therapeutics	Cancer vaccine for G12C, G12D, G13D, G12V	Phase I, monotherapy and with PD1-blocker pembrolizumab
G12D inhibitor	Mirati Therapeutics	G12D inhibitor	IND- enabling studies in 2020
RAS(ON) inhibitors	Revolution Medicines	Tri- complex inhibitors of mutated GTP- bound KRAS	Preclinical
NA	Bayer	KRAS-SOS1 inhibitor	Preclinical
NA	Sanofi/X-Chem	G12C inhibitor	Preclinical
NA	X- Chem	G12C inhibitor, for active and inactive KRAS	Preclinical
BBP-454	BridgeBio Pharma	Pan- KRAS inhibitors	Preclinical

IND, investigational new drug. aAll three lead G12C inhibitors bind only the inactive GDP- bound KRAS.

자료: Mullard, A., 2019

나) 암유전자 분석을 활용한 고위험군 선별 및 치료전략

광범위한 유전자 변이를 분석하여 동일 암종을 몇 가지 그룹으로 분류를 함으로써 각 그룹별 예후에 따른 치료전략을 수립하는 노력이 이루어지고 있다. 예후가 나쁜 그룹의 환자들에게는 보조항암치료를 추가하거나 효과가 좋은 항암제를 권장하는 전략도 개발되고 있다. 대표적인 분자생물학적 암종별 하위그룹에 따른 치료법 선택은 유방암의 luminal A, luminal B, 삼중음성, HER2 양성 및 정상 유사그룹 등 분류에 따라 선택되며, 위암은 Epstein-Barr virus(EBV)군, microinstability(MSI)군, genomically stable(GS)군 및 chromosomal instability(CIN)군으로 분류하여 일부 하위군(MSI와 EBV군)에서는 면역관문억제제가 권장된다. 직결장암은 4개의 consensus molecular 하위군(CMSs)로 분류하여 각각 CMS1(microsatellite instability immune), CMS2(canonical), CMS3(metabolic) CMS4(mesenchymal) 등으로 분류하지만 아직까지 하위군에 따라 치료법을 다르게 권장하지는 않고 있다. 췌장암과 전립선암 등 여러 암종에서도 분자

생물학적 암종별 하위그룹을 분류하는 연구가 지속적으로 진행되고 있다. 암종별 하위 그룹 진단에는 다양한 유전자 발현을 종합적으로 적용하는 방법이 많이 사용되나, 일부 하위그룹은 아직까지 선택적 치료법이 확립되지 않은 경우도 있다. 분명히 암종별로 하위 그룹 분류가 가능한 분자생물학적 공통점을 가진 환자그룹이 존재하나 각 그룹별 치료 전략이 없으면 굳이 다양한 진단법으로 분류를 하는 것이 임상적 의미가 없으므로 향후 이에 대한 연구가 지속적으로 필요하다.

다) 재발 모니터링 및 조기 진단

최근 혈액내 순환암세포(circulating tumor cells, CTCs), 암세포 유래 DNA(circulating tumor DNA, ctDNA)의 연구가 진행되면서 암의 조기진단과 완치적 절제 후 재발 모니터링 목적으로 많은 연구가 이루어지고 있다. 특히 비침습적으로 혈액만 채취하여 검사할 수 있고, 중복암 환자에서 특정 암의 재발여부를 진단할 수 있다는 장점을 가지고 있다. 특히 초기 병기의 직결장암과 유방암의 수술적 절제 후 미세잔존암의 존재 여부를 검사하여 예후 판정이 가능하다는 보고가 연속적으로 발표되고 있다. CTC 또는 ctDNA를 사용하는 액체생검 진단법은 암 환자의 원발암뿐만 아니라 전이암 조직에서 흘러나온 모든 암세포 유래 DNA를 종합적으로 분석할 수 있고, 시간의 경과에 따라 순차적으로 유전자 변이를 비교함으로써 치료에 대한 내성의 발생을 추적할 수 있다는 장점도 있다.

정상인에서 암을 조기에 선별하는 암검진 진단법은 매우 높은 정확성과 특이성을 필요로 한다. 위양성과 위음성 결과는 많은 사람에게 불필요한 검사와 이로 인한 비용 및 부작용을 초래할 수 있기 때문이다. 정상세포와 다른 암세포 특이 유전자 변이를 혈액에서 찾아내는 CTC 혹은 ctDNA 사용진단법은 암검진 진단법으로 매우 유망하며, ctDNA 검사로 초기 직결장암, 위식도암, 췌장암, 유방암에서 각각 73%, 57%, 48%, 그리고 50%에서 진단이 가능하여 전체적으로 1기 암에서는 47%, 2기와 3기 암에서는 55%와 69%로 증가하는 결과를 보였다. 그러나 CTC 혹은 ctDNA 단독진단법보다는 기존의 암검진 검사와 보완적으로 병용하여 활용하는 접근이 필요하며, 최종적 임상에 도입을 위해서는 확증적 대규모 전향적 임상시험이 필요하다.

라) 암정밀의료 시대의 임상시험

암정밀의료의 처음 도입되는 시기에는 대형 글로벌제약회사의 신약개발 필요성에 따라 약제의 표적이 되는 유전자 변이를 가진 환자만을 골라서 진행되었으며, 특정 유전자변이가 없는 환자는 참여가 제한되거나 기존 치료를 받는 대조군에 포함되었다. 그러나 이러한 접근으로는 대규모 환자 유전자 정보 파악도 어렵고 환자들의 수요도 충족할 수 없으며 신약을 개발하는 특정 제약회사가 모든 정보를 독점하게 됨을 알게 되었다. 특히 암 환자들은 다양한 유전자변이를 가지고 있는 반면 이러한 수백 가지 유전자 변이를 한꺼번에 검사할 수 있는 인프라도 부족하고, 사용허가된 신약 항암제와 임상시험 중인 신약이 많지 않아 실제로 암 환자에게 도움이 되는 효과적인 정밀의료 맞춤형 치료가 불가능하였다.

이러한 한계점을 극복하기 위한 전략이 마스터프로토콜(플랫폼임상시험)이라는 프로젝트로 이는 국가나 메머드급 대형 암전문기관이 주체가 되어 진행되는 것이 일반적이다. 대표적인 임상시험으로 미국의 Signature, I-SPY, National Cancer Institute(NCI)-sponsored NCI-MATCH, NCI-MPACT, ALCHEMIST, Lung-MAP, Pediatric MATCH, Exceptional Responders, Lung Cancer Mutation Consortium(LCMC) 임상시험, 그리고 미국임상암학회(ASCO)가 주도하는 TAPUR(Targeted Agent and Profiling Utilization Registry) 임상시험 등이 있다. 다른 국가 차원의 암정밀의료 플랫폼임상시험은 일본의 SCRUM-Japan, LC-SCRUM-Asia, MONSTAR-SCREEN 등이 있고, 우리나라의 K-MASTER 프로그램이 있다. 본 프로그램의 장점은 참여하는 환자들의 유전자 정보를 동일한 대량 유전자검사법으로 최대한 확보하고, 각각의 유전자 변이를 타겟으로 하는 신약으로 치료받을 수 있는 임상시험을 체계적으로 오픈하여 환자들에게 많은 치료기회를 제공함과 동시에 빅데이터 분석 시스템을 구축하여 이를 활용한 신약개발의 시너지를 추구하는 프로젝트라는 점이다. 특히 이 프로젝트에 참여한 환자는 본인의 특정 유전자 변이를 타겟으로 하는 신약 임상시험이 없다 하여도 추후 새로운 임상시험이 시작되면 치료를 받을 수 있는 기회를 가질 수 있으며, 새로운 임상시험 참여가 불가능하여도 다른 치료법의 효과 여부에 대한 실제 의료현장 자료 수집이 가능하다. 이 프로젝트가 많은 장점을 가지고 있으나 성공하기 위해서는 참여 환자들이 신뢰할 수 있는 공정성, 투명성, 전문성이 확보되어야 하며, 새로운 신약의 치료기회를 제공하는 다수의 연구자 주도 임상시험이 가능하여야 한다. 또한 동일한 유전자분석 플랫폼을 사용하여 빅데이터 구축이 가능하여야 한다. 이를 위해서는 지속적인 국가연구비를 지원하여 연구자 주도 임상시험들이

다수 수행되어야 하며 빅데이터 구축을 위한 투자가 선행되어야 하므로 국가적인 지원이 반드시 필요하며, 전 국민이 암정밀의료의 당위성과 필요성에 공감대를 이루어야 하겠다.

암과 같은 난치병의 극복을 위해서는 과거의 비효율적인 신약개발 프레임에서 벗어나 환자, 진료의사, 의과학연구자, 국가 등 참여 주체 모두에게 도움이 되는 전략을 수립하고 함께 협력하여야 전 인류의 질병으로 인한 고통을 줄일 수 있을 것이다.

마) 유전성 암증후군

현재까지 300개 이상의 유전성 암증후군이 알려져 있다. 가장 대표적인 증후군은 유전성 유방암 난소암 증후군(Hereditary breast-ovarian cancer syndrome, HBOC)과 린치 증후군(Lynch syndrome)이 있다. 대부분의 HBOC 환자들은 BRCA1 혹은 BRCA2 유전자의 생식세포 돌연변이를 가지고 있으며, BRCA1 돌연변이 여자환자는 70세 이전에 47~66%에서 유방암이, 35~46%에서는 난소암이 발생한다. BRCA2 돌연변이 보유자는 유방암과 난소암 이외에도 췌장암과 흑색종의 발생 위험도 높다. Mismatch repair 유전자의 생식세포 돌연변이를 가지고 있는 린치 증후군 환자들은 직결장암 발생 위험이 80%에 달하며, 자궁내막암 등 다른 암 발생의 위험도 높다.

암조직에서 추출한 DNA를 사용하여 NGS 패널 시퀀싱 분석을 하면 체세포 돌연변이 뿐 아니라 일부 환자에서 pathogenic 생식세포 돌연변이가 동시에 확인된다(~7%). 최근 여러 임상시험을 통해 BRCA1와 BRCA2 유전자의 생식세포 혹은 체세포 돌연변이를 가진 환자는 백금계 항암제에 좋은 반응을 보이며 PARP 억제제 투여로 암의 진행을 억제한다는 보고가 나오고 있다. 또한 Mismatch repair 유전자의 생식세포 돌연변이를 가지고 있는 암 환자들은 면역관문억제제에 좋은 반응을 보이는 것으로 알려져 유전성 암증후군 관련 생식세포 돌연변이 보유자들을 대상으로 정밀 암검진 및 암치료전략 수립이 필요하다.

2) 암정밀의료의 실현을 위해 극복해야 할 과제

암정밀의료 치료법이 환자에게 사용되기 위해서는 특정 유전자 및 변이가 어떤 임상적 결과를 초래하는지가 확인되어야 한다. 이러한 과정은 많은 시간과 노력이 필요하며, 수많은 유전자분석에 이어 대량의 데이터를 분석할 수 있는 생물정보학의 발전이 뒤따라야 한다.

또한, 특정 환자에게 적절한 검사가 시행되고 검사결과에 따라 적절한 약이 투약되기 위해서는 의료진들의 더 많은 전문성과 지식을 필요로 한다.

방대한 양의 분자생물학적 분석 데이터의 저장과 전문가들의 접근성 확보, 데이터 플랫폼 구축, 임상적 의미를 덧붙인 데이터베이스 작업 등 노력과 함께 새로이 밝혀진 지식이 자동적으로 업데이트되어 사용자가 최선의 진료에 활용될 수 있어야 한다. 또한 의료 빅데이터를 효율적으로 분석하고 경향성을 파악하여 치료법 결정을 지원할 수 있는 수준까지 발전시켜야 한다.

새로운 검사법과 치료법이 개발되면서 이러한 상품들의 혁신성을 감안하면서 안전성, 유효성을 평가할 수 있는 인허가 규제기관의 동반되어야 한다. 미국 FDA는 발빠른 정밀 의료 관련 인허가 프로그램을 도입함으로써 2005년 신약 허가의 단 5%만이 정밀의료 관련 신약이었으나 2017년에는 30%에 달하게 되었다.

3) 암정밀의료의 미래

CTC와 ctDNA를 사용한 액체생검 진단법은 비침습적이고 반복적으로 분석이 가능하여 암조기진단 및 선별에 유용할 것으로 기대된다. 아직 기술적 한계점을 극복하여야 하는 문제는 있으나 분명히 미래 암정밀의료 분야에서 핵심적 역할을 담당할 것이다. Organoid는 환자의 암조직을 체외에서 배양하고 약물반응 평가가 가능하여 암 환자 개인별 특성과 효능을 매우 효과적으로 분석할 수 있다. 여러 가지 표적치료 약물 중 단클론 항체를 이용하는 전략은 개인별 맞춤 표적 차단뿐 아니라 면역반응을 향상시키고 다른 항암제를 함께 암세포에만 전달해 주는 장점을 활용할 수 있다. 최근 암치료 성적을 가장 많이 향상시키는데 기여한 약물은 면역관문억제제로써 효능을 예측할 수 있는 바이오마커만 좀 더 확실해진다면 단독, 혹은 병용 투여로 획기적 치료효과 향상을 기대할 수 있을 것이다. 개인별 맞춤 암치료의 궁극적 완성은 암백신과 CAR T-cell 치료법이 될 것이다. 각 환자별 neoantigen을 확보하고 면역세포를 활성화하여 특정 암세포만을 공격하는 전략은 진정한 암정밀의료의 미래가 분명하다. 추가적으로 이러한 의학적 발전뿐 아니라 암정밀의료의 혜택을 볼 수 있는 모든 환자들이 새로운 치료법에 접근가능하도록 사회적 체계를 구축하는 노력이 함께 이루어져야 하며 전 국민의 정밀의료에 대한 인식 제고가 선행되어야 할 것이다.

참고문헌

- Cocco, E., Scaltriti, M., Drilon, A.(2018). “NTRK fusion-positive cancers and TRK inhibitor therapy”. *Nature Reviews Clinical Oncology*, 15(12), pp.731-747. (DOI: 10.1038/s41571-018-0113-0).
- Fostira, F., Thodi, G., Konstantopoulou, I. *et al.*(2007). “Hereditary cancer syndromes”. *Journal of Balkan Union of Oncology*, 12, pp.13-22.
- Krzyszczuk, P., Acevedo, A., Davidoff, E. J.(2018). “The growing role of precision and personalized medicine for cancer treatment”. *Technology*, 6(03&04), pp.79-100. (DOI: 10.1142/S2339547818300020).
- Liu, P., Wang, Y., Li, X.(2019). “Targeting the untargetable KRAS in cancer therapy”. *Acta Pharmaceutica Sinica B*, 9(5), pp.871-879. (DOI: 10.1016/j.apsb.2019.03.002).
- Malone, E. R., Oliva, M., Sabatini, P. J. B. *et al.*(2020). “Molecular profiling for precision cancer therapies”. *Genome Medicine*, 12(1). (DOI: 10.1186/s13073-019-0703-1).
- Meric-Bernstam, F., Mills, G. B.(2012). “Overcoming implementation challenges of personalized cancer therapy”. *Nature Reviews Clinical Oncology*, 9(9), pp.542-548. (DOI: 10.1038/nrclinonc.2012.127).
- Mullard, A.(2019). “Cracking KRAS”. *Nature Reviews Drug Discovery*, 18(12), pp.887-891. (DOI: 10.1038/d41573-019-00195-5).
- Ngeow, J., Eng, C.(2016). “Precision medicine in heritable cancer: when somatic tumour testing and germline mutations meet”. *npj Genomic Medicine*, 1(1). (DOI: 10.1038/npjgenmed.2015.6).
- Oh, D., Bang, Y.(2020). “HER2-targeted therapies - a role beyond breast cancer”. *Nature Reviews Clinical Oncology*, 17(1), pp.33-48. (DOI: 10.1038/s41571-019-0268-3).
- Scheerens, H., Malong, A., Bassett, K. *et al.*(2017). “Current status of companion and complementary diagnostics: strategic considerations for development and

launch”. *Clinical and Translational Science*, 10(2), pp.84-92. (DOI: 10.1111/cts.12455).

Tsimberidou, A. M., Eggermont, A. M., Schilsky, R. L.(2014). “Precision cancer medicine: the future is now, only better”. *American Society of Clinical Oncology educational book*, 34, pp.61-69. (DOI: 10.14694/EdBook_AM.2014.34.61).

Zaman, A., Wu, W., Bivona, T. G.(2019). “Targeting Oncogenic BRAF: Past, Present, and Future”. *Cancers*, 11(8). (DOI: 10.3390/cancers11081197).

2 유전성 질환

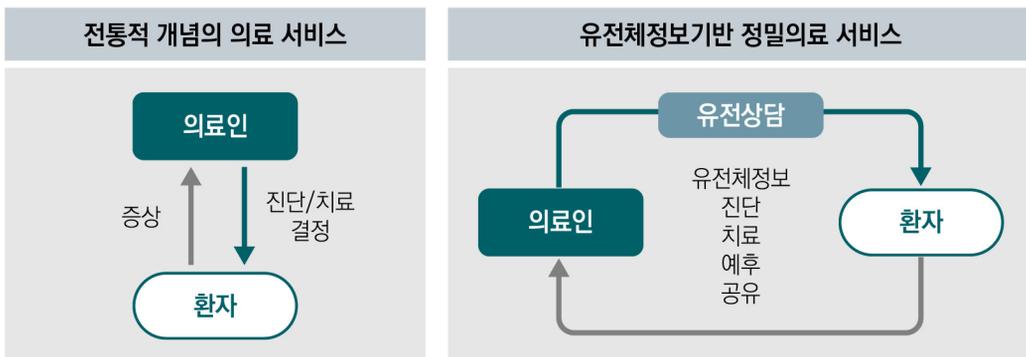
가. 희귀 유전질환

1) 희귀질환과 정밀의료

정밀의료는 질환의 중등도에 따라 획일적으로 관리 및 치료를 해왔던 과거의 의료 행태에서, 환자 개인의 유전체정보, 환경, 생물학적 정보에 따라 개개인에게 맞춤 의료 서비스를 제공하는 방향으로 변화를 가져오고 있다.

정밀의료에서는 동일 질환으로 진단된 환자라 하더라도 환자 개개인이 가진 유전체 정보, 생물학적 정보, 사회경제학적 및 생태학적 환경에 따라 질환의 예방, 치료 및 관리의 의료 서비스를 차별화하여 받게 된다. 정밀의료에서는 이전의 종속적이었던 의료인과 환자의 관계는 평형화되고, 의료정보는 이전의 의료인 독점 방식에서 환자와 의료인이 서로 정보를 공유하게 되며, 환자가 의료 서비스에 있어서 의사 결정의 중심 역할을 맡게 된다 <그림 3.4>.

그림 3.4 유전체정보기반 정밀의료에서의 의료서비스



이러한 정밀의료 서비스가 가장 활발히 이루어지고 있는 질환군이 희귀질환이다. 희귀질환은 국가마다 정의가 다양하나 우리나라는 희귀질환관리법 제2조에 의거하여 유병 인구가 2만 명 이하이거나 진단이 어려워 유병인구를 알 수 없는 질환으로 보건복지부령으로

정한 절차와 기준에 따라 정한 질환을 칭한다. 전 세계적으로는 5,000~8,000개 이상의 희귀질환이 있으며, 우리나라에는 1,000여 종이 넘는 희귀질환이 알려져 있다.

희귀질환군에 속한 각각의 질환은 적게는 10명 이내에서 많게는 수천 명의 환자가 있어, 개별 질환의 유병률은 매우 낮으나 전체 질환은 천여 개 이상의 질환에 이르고, 우리나라 총 인구의 약 10%가 희귀질환자이다.

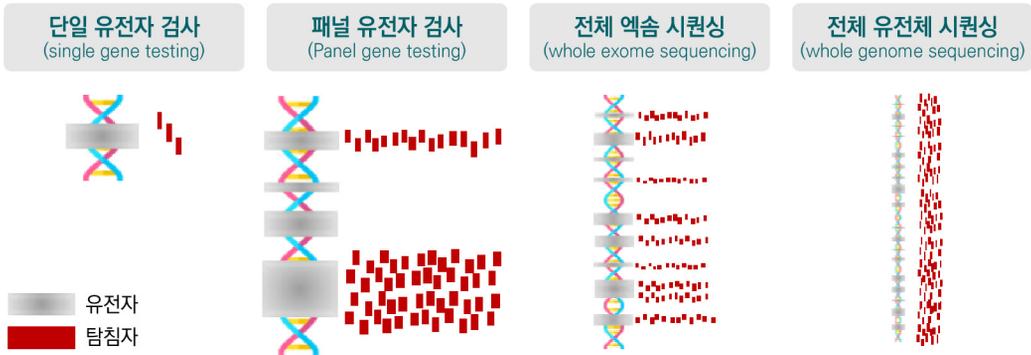
희귀질환은 80% 이상이 유전적 원인으로 생각되어지고 있다. 하지만 질환의 종류는 매우 많은데 비해, 질환별 환자 수는 매우 드물어서 진단을 받는데도 많은 시간이 걸리고, 진단을 받지 못하고 지내는 경우도 많다. 이로 인해 많은 희귀질환자들이 정신적, 신체적, 사회적, 경제적 고통을 받고 있다. 희귀질환은 진단이 된다고 하여도 치료법이 없는 경우가 많았고, 질환의 자연 경과를 잘 몰라서 관리도 받지 못하는 경우가 대부분이었다. 그러나 유전진단 기술의 발전에 따라, 질환에 대한 이해도가 높아지게 되었고, 질환 별 관리법이 가능해졌으며, 치료제도 하나 둘 개발이 되고 있다. 또한, 희귀질환자들은 유전진단을 통해 향후 가족계획에도 도움을 받을 수 있게 되었다.

2) 유전체 분석 기술의 발전과 희귀질환의 진단법의 발전

가) 유전체 분석 기술의 발전

희귀질환에서 정밀의료를 현실화하는데 반드시 필요한 요건이 개개인의 유전체정보 확인이다. 2000년대 이전에는 개개인의 유전체정보는 단일 유전자분석이나 인간에 존재하는 단일염기 다형성(Single nucleotide polymorphism)을 단일적으로 혹은 복합적으로 분석하는 수준이었다. 그러나 2000년대 초반에 들어와서 인간게놈지도의 완성 및 차세대염기서열 분석법의 도입과 더불어 인간게놈 전체를 짧은 시간에 분석하는 것이 가능하게 되었다 <그림 3.5>, <표 2.2>.

그림 3.5 유전체 분석 방법



(1) 단일 유전자분석

단일 유전자분석은 1988년 에블리히 등에 의해 중합효소연쇄반응(Polymerase Chain Reaction, PCR) 기법과 프래드릭 생어 등이 개발한 생어 염기서열 분석법(Sanger sequencing analysis, Sanger dideoxy procedure)이 소개되면서 보편화 되었다. 이는 사람 유전체에서 특정 하나의 유전자만 증폭하여 원하는 부위의 DNA를 분리하고, 이를 생어 분석을 통하여 염기서열을 분석하는 방법이다. 단일 유전자분석은 특정 유전질환이 의심되고, 원인 유전자가 1개이거나 소수일 경우에 사용할 수 있는 방법으로, 현재까지도 유전진단을 위해 가장 보편적으로 사용되는 방법이다 <그림 3.5>. 가령, 윌슨병, 파브리 병, 고셔병, 신경섬유종증 1형 등 1개의 유전자가 원인으로 알려진 희귀질환의 진단에 유용한 검사이다. 그러나 1개의 유전자를 분석하는데 일반적인 상업적 검사실에서는 3주 이상의 시간이 소요되고, 한 질환에서 원인 유전자가 여러 개인 경우 수개월 이상의 시간이 소요될 수 있다는 단점이 있다. 또한 환자를 진료한 의사가 원인 희귀유전질환을 의심하지 못하면 검사가 불가능하다.

(2) 차세대 염기서열 분석법(Next generation sequencing, NGS)

1983년 미국 정부에 의해 계획된 인간게놈프로젝트는 2003년에 마무리되어서 인간의 게놈 지도가 완성되었다. 차세대 염기서열 분석법이 개발되면서, 인간 유전체의 분석 속도가 획기적으로 단축되었다. 차세대 염기서열 분석법은 High-throughput sequencing, Massive parallel sequencing 또는 Second-generation sequencing이라고도 불리는

기법으로, 다수의 유전체를 작은 절편의 DNA로 나누어서 동시에 분석할 수 있는 기법을 지칭한다. 이 획기적 기술의 도입으로 인간 유전체(게놈)의 신속한 분석이 가능하게 되었다. 차세대 염기서열 분석법을 이용한 유전체 분석은 크게 3가지 방법이 존재한다. 패널 유전자 검사(panel gene sequencing), 전체 엑솜 시퀀싱(whole exome sequencing), 전체 유전체 시퀀싱(whole genome sequencing)이 이에 해당한다 <그림 3.5>.

나) 차세대 염기서열 분석법을 이용한 유전 진단법

(1) 패널 유전자 검사

패널 유전자 검사는 차세대 염기서열 분석법을 이용하여 특정 희귀유전질환과 관련이 있는 유전자들을 한 번에 분석하는 방법이다 <그림 3.5>. 가령 유전성 망막변성의 경우 관련 원인 유전자가 80개 이상이 존재한다. 반면 대부분의 원인 유전자들은 극히 일부의 환자에서만 돌연변이를 발견할 수 있다. 이러한 경우 단일 유전자 검사로는 현실적으로 원인 유전자를 찾는 데 필요 이상의 시간이 소요될 수 있다. 따라서 80개 이상의 원인 유전자를 한 번에 검사할 수 있는 패널 유전자 검사가 유용하게 사용될 수 있다.

패널 유전자 검사는 인간 유전체에서 보고자 하는 유전자 수십에서 수백여 개를 표적하는 탐침을 디자인하여 차세대염기서열 분석법을 이용하여 동시에 시퀀싱하는 방법이다. 표적 부위는 500회 이상의 탐침이 반복되어 정확한 분석이 가능하다. 유전성 망막 질환, 유전성 심근병증, 유전성 난청 등 원인 유전자가 다수인 희귀질환 군에 사용하는데 적절하다. 그러나 디자인된 표적에 포함 안 된 부위에 원인 유전자가 존재하는 경우 패널 검사로는 원인 유전자의 돌연변이를 찾지 못한다는 단점이 있다. 또한 진료 의사가 해당 질환 군을 의심하지 못하면 검사가 불가능하다.

(2) 전체 엑솜 시퀀싱

엑솜(exome)은 set-of-exons를 뜻하는 것으로 약 30억 개의 인간 게놈 염기쌍 중 약 1~2%에 해당하는 단백질을 전사하는 부위를 종합하여 이르는 용어로, 사람의 유전질환의 돌연변이의 약 80%가 여기에 존재한다. 전체 엑솜 시퀀싱은 단백질을 전사하는 약 20,000개 이상의 유전자를 차세대 염기서열을 이용하여 동시에 분석하는 방법이다 <그림 3.5>.

전체 엑솜 시퀀싱은 유전질환으로 의심되나 패널 유전자 검사로 발견이 되지 않은 경우에 시행해 볼 수 있으며, 또한 유전질환이 의심되나 의료인이 특정 질환으로 진단할 수 없는 경우에도 시행해 볼 수 있다. 최근에는 전체 엑솜 시퀀싱의 비용이 절감되어 일부에서는 패널 검사를 대체하여 시행되기도 한다. 전체 엑솜 시퀀싱은 패널 유전자 검사보다 검사할 수 있는 유전자의 개수는 월등히 높으나, 각각 유전자에 평균 50~100여 개의 탐침이 반복되어 패널 검사보다는 표적 부위에 대한 집중도는 떨어진다고 볼 수 있다. 또한 전체 엑솜 시퀀싱은 단백을 전사하지 않는 유전체 부위(non-coding genome)와 50bp 이상의 염기서열의 구조적 변이(structural variation)는 발견하지 못한다는 단점이 있다.

전체 엑솜 시퀀싱은 특히 출생 후 다발성 선천성 기형이나 발달 지연/인지 장애가 있는 소아 희귀환자에서 임상적인 진찰만으로 진단이 어려울 경우 많이 사용되는데, 이를 이용한 희귀유전질환의 진단율은 30-40% 정도에 이르는 것으로 알려져 있다. 또한 진료 의사가 질환을 정확하게 의심하지 못하는 경우 환자의 증상과 유전정보를 매칭하여 희귀유전 질환을 진단하게 되고, 이를 통하여 진료 의사는 알지 못하였던 새로운 질환에 대한 정보를 습득하여 환자에게 의료 서비스를 제공하게 된다.

(3) 전체 유전체 시퀀싱

전체 유전체 시퀀싱(Whole genome sequencing)은 말 그대로 인간의 모든 유전체 정보를 분석하는 방법이다 <그림 3.5>. 전체 유전체 시퀀싱은 전체 엑솜 시퀀싱에서 발견하지 못하는 단백을 전사하지 않는 유전체 부위(non-coding genome)의 돌연변이와 50bp 이상의 염기서열의 구조적 변이(structural variation)를 발견할 수 있다는 장점이 있다. 한 사람의 전체 엑솜 시퀀싱으로 발생하는 데이터양이 8GB임에 비하여 전체 유전체 시퀀싱은 한 사람당 150GB로 훨씬 더 방대한 양의 데이터가 생산된다.

전체 유전체 시퀀싱은 전체 엑솜 시퀀싱보다 상대적으로 높은 유전질환 진단율을 기대할 수 있다. 일례로 희귀유전질환이 의심되는 소아를 대상으로 전체 유전체 시퀀싱은 30~57%의 진단율이 보고되고 있다. 이러한 진단율은 전체 엑솜 시퀀싱에 비해 유의하게 높은 수치는 아니지만, 보고마다 대상이 된 환자들의 특성에 따라 진단율의 차이가 있어 직접 비교는 무리가 있다. 그러나 전체 유전체 시퀀싱은 전체 엑솜 시퀀싱에 비해 전사

부위의 돌연변이 발견률, 단백질의 기능에 유의한 영향을 줄 수 있는 변이를 좀 더 높은 빈도로 찾아낼 수 있어서 좀 더 명확하고 정확한 유전체정보를 제공한다.

3) 희귀질환에서 유전체정보기반 정밀의료의 현실화

가) 유전질환의 진단 및 치료

유전체 분석 기술의 눈부신 발전은 많은 희귀유전질환의 빠르고 정확한 진단에 큰 기여를 하였다. 특히 희귀유전질환 중 조기 진단이 중요한 태아, 신생아, 소아기의 환자군에서 획기적인 진단법으로서의 역할이 증대되고 있다.

출생아의 약 2%는 선천성 기형을 가지고 있고, 이 중 약 20%는 태아기 혹은 출생 후 사망하게 되며, 이들 기형의 상당수가 유전적 원인에 의한 희귀유전질환이다. 선천성 기형 외에도 인구 100명당 0.5~1명은 발달 장애를 가진다. 발달장애는 발달지연, 인지장애, 지체장애, 행동장애, 발달의 퇴행 등의 다양한 임상 상을 보이는 질환 군으로서, 질환으로 고통 받고 있는 환자 50% 이상이 초등학교 입학 전에 원인 진단이 되지 못하고 있다. 우리나라에서는 지적발달장애가 17세 미만 청소년과 어린이들에게서 가장 높은 비중을 차지하는 장애유형이다. 현재까지 발달장애 환자의 약 50%에서 유전적 원인을 차지하고 있고, 유전적 원인의 발굴에 전체 엑솜 시퀀싱이나 전체 엑솜 시퀀싱의 도입이 큰 역할을 해왔다.

성인기에서도 유전체 분석은 다양한 희귀유전질환의 진단에 중요한 역할을 하는데, 가족성 종양 증후군(familial cancer syndrome), 망막 색소 변성증(retinitis pigmentosa), 심근병증(cardiomyopathy), 신경퇴행성질환(neurodegenerative disorder) 등 다양한 희귀유전질환 중 성인기에 발현하는 질환들이 이에 해당한다.

유전적 진단은 단순한 진단 이외에 정밀의료의 관점에서 많은 역할을 한다. 유전적 진단은 단순한 진단적 가치 외에 질환의 예방, 관리 및 치료에 중요한 정보를 제공한다.

(1) 진단적 가치

유전체 분석을 이용하여 진단되는 유전질환은 대부분 희귀유전질환이다. 희귀질환은 각 질환에 대한 의료 정보가 부족하여, 환자의 진료를 담당하는 의사에게도 생소한 질환인 경우가 대부분이다. 이러한 경우 각 희귀질환 환자의 임상정보를 공유하고 유전체 변이

정보를 공유하여 우리나라 내에서 뿐만 아니라 전 세계적으로 각 질환에 대한 지식을 축적하게 된다.

유전체 분석을 통하여 진단된 희귀질환자 및 그의 가족은 진단된 질환에 대하여 환자가 주의해야 할 증상은 무엇인지, 장기적으로 어떠한 예후를 가지게 될지 매우 궁금하게 된다. 위에서 설명한 국내적, 국제적으로 공유된 희귀질환에 대한 유전적, 임상적 정보 지식은 환자에서 질환의 양상 및 경과를 이해하고 예측하는 주요한 자료로 활용되어진다. 특히 우리나라와 같이 인터넷 보급이 널리 보편화된 나라에서는 희귀질환자와 가족이 실시간으로 이환된 질환에 대한 정보를 찾아보게 되며, 각종 온라인 혹은 오프라인의 환자자조모임을 통해 정보를 공유함으로써 질환의 치료 및 관리에 적극적으로 참여하게 된다.

유전체 분석은 정밀의료의 관점에서 환자의 장기 예후 및 증상의 중증도를 예측하는데 중요한 정보를 제공한다. 상당수의 희귀유전질환은 유전자의 돌연변이와 환자 증상 간의 상관성(유전형-표현형의 상관성)이 보고되어 있어, 유전체 분석을 통한 유전 진단이 환자의 증상을 파악하는데 도움이 된다. 또한, 사람에서 발견되는 유전자의 돌연변이는 프레임 시프트돌연변이(frameshift mutation), 스플라이싱돌연변이(splicing mutation), 종결 돌연변이(nonsense mutation), 엑손결실/중복돌연변이(exonic deletion/duplication) 등처럼 단백질 기능을 완전 소실시키는 돌연변이들이 있고, 반면 인프레임돌연변이(inframe mutation), 과오돌연변이(missense mutation)는 부분적으로 단백질의 기능을 보유할 수 있는 경우가 있다. 따라서 후자의 돌연변이를 가진 환자의 경우에는 원인 유전자의 부분적 기능을 보유함으로써 전자의 돌연변이를 가진 환자보다 질환의 중증도가 낮을 가능성이 있다. 그러나 같은 돌연변이를 가진 환자 간에도 증상의 중증도 차이가 심한 경우들이 있으므로, 돌연변이의 종류와 질환의 중증도의 상관성을 일반화할 수는 없으며, 다만 환자의 질환 경과에 대한 예측 자료 중 하나로 활용될 수는 있다.

(2) 질환의 관리, 예방 및 치료적 가치

유전체 분석을 통한 희귀유전질환의 진단은 관리 및 치료의 관점에서 정밀医료를 시행하는 데에 중요한 역할을 한다. 현재까지 대부분의 유전질환은 근본적인 치료가 되지 않고 있다. 그럼에도 불구하고 보존적 치료와 통한 효과적 관리를 통해 환자의 생존 기간을 연장시키고 삶의 질을 향상시킬 수 있다. 따라서 다양한 의료정보를 통하여 진단된 환자들의

증상을 파악하고, 질환별로 주의해야 하는 증상에 대한 모니터링 계획을 세우는 것이 중요하다.

관리적 차원에서 중요한 희귀유전질환군 중 대표적인 예가 가족성 종양 증후군이다. 가족성 종양 증후군의 경우 종양이 발생하지 않은, 즉 증상 발현 이전 단계의 질환을 가진 환자를 발견할 수 있으며, 유전적 진단 후 해당 질환에 대한 정기적인 탐색 검사를 통하여 종양을 발생 초기에 치료를 함으로써, 생존율을 높이고 종양 발생에 의한 합병증의 발생을 최소화할 수 있다.

또 다른 예는 유전적 원인에 의하여 발생하는 심장의 부정맥 질환이다. 이 질환군의 환자는 증상을 모르고 살다가 갑작스런 빈맥의 발생하여 갑자기 사망하는 위험이 있다. 따라서 무증상기의 대상자에서 부정맥 위험이 알려진 유전자의 이상을 발견함으로써, 생활 방식의 조절 및 심장 전문의 진료를 통하여 빈맥 발생에 의한 급사의 위험을 최소화할 수 있다.

위에서 언급한 대로 대부분의 유전질환은 근본적인 치료가 어렵다. 그럼에도 불구하고 일부 희귀질환에서는 획기적인 치료제가 개발되어 환자들의 질환 치료에 많은 기여를 하고 있다. 이의 대표적인 예가 유전자의 이상에 의해 생성을 못하거나 부족해진 단백질을 유전자 재조합 기술을 이용하여 체외에서 부족한 단백질을 생성하여 주입하는 치료법이다. 라이소좀 축적 질환(lysosomal storage disease)에서 행해지는 효소대체요법이 이에 해당한다.

유전자 돌연변이 유형에 따른 특이적 치료법들도 개발되고 있는데, 이의 대표적인 예시가 샤프론 제제를 이용한 유전질환의 치료이다. 샤프론 제제는 과오돌연변이에 의하여 비정상적 3차 구조를 가지고 있는 유전자 전사 단백질에 특이적으로 결합하여 3차 구조를 개선시켜 단백질의 기능을 향상시키는 물질이다. 샤프론 제제는 유전자 특이적이며, 돌연변이 특이적이어서 유전자 검사로 특정 유전자에 샤프론 제제에 반응하는 돌연변이가 발견된 경우에 사용이 가능하다. 근육질환에서 종결돌연변이를 읽고 지나감 전사(read-through)를 일으키는 약제를 사용하여 돌연변이 단백을 기능이 회복된 정상 단백질로 전환시킴으로써 치료 효과를 유도하는 치료제도 개발되고 있다. 스프라이싱돌연변이에 의해 발생하는 유전질환에서 안티센스 올리고 핵산(antisense oligonucleotide, ASO)을 이용해서 잘못된 전사를 교정하는 치료법들이 제시되고 있다.

환자의 유전질환에서 원인이 되는 유전자의 이상을 직접 교정하는 유전자 치료법도 활발히 진행되고 있다. 특히 이중 안과 질환인 망막색소변성증 20번과 천추근육위축(spinal muscular atrophy)에서의 유전자 치료는 최근 미국 식품의약국(U.S. FOOD & Drug administration)에서 승인된 바 있다 (<https://www.fda.gov>). 향후 좀 더 많은 유전질환에서 유전자 치료법이 도입될 것으로 예측된다.

지금껏 언급한 새로운 치료법들은 환자의 유전적 진단 및 원인 유전자의 돌연변이를 기반으로 하여 개발된 치료법들로서 희귀질환에서 유전체정보기반 정밀의료의 대표적 성공 사례라 할 수 있다.

나) 임신 및 산전 유전 진단

유전체 분석을 이용한 정밀의료의 중요한 의료 분야가 태아 및 신생아의 희귀유전 질환의 진단이다. 우리나라의 대통령 직속 저출산·고령사회위원회의 자료에 의하면 연간 40~50만 명에 이르던 신생아 출산 수가 30만 명이 안 되는 저출산 시대에 처해 있다. 이로 인해 장기적 국가 인적 경쟁력 약화의 위험 및 고령화 사회로 인한 부작용이 염려되고 있다.

그러나 한편으로는 이러한 저출산 시대에서는 어린이 한명 한명에 대한 양질의 의료 서비스의 필요성은 더욱 증대되고 있으며, 선진국형 의료 서비스인 정밀의료 서비스를 통하여 출산아 개개인이 건강한 미래의 인적자원으로 성장하도록 하고자 부단히 노력이 필요하다.

유전질환이 의심되어 출생 후 신생아기에 사망한 아이의 부모의 경우 다음번 임신 시 같은 질환으로 이환된 태아를 임신할 수 있는 위험성으로 인한 두려움으로 차기 임신을 계획하지 못하고, 최악의 경우 가정이 붕괴될 수도 있다. 이러한 경우 유전체 분석을 통한 정밀의료는 매우 중요한 역할을 한다. 출생 후 얼마 안 되어 수개월 내에 사망하는 경우 임상 진찰만으로 원인적 진단을 하기가 힘들고, 원인 질환이 의심된다고 하더라도 과거 단일 유전자 검사로 유전 진단을 하던 시기에는 유전자 하나씩 검사하는 데만 1~2개월의 시간이 소요되어 결국 환자 사망 전에 유전자 진단을 하지 못하는 경우가 허다하였고, 적게는 수개월에서 수년의 시간 후에 유전진단을 하거나, 이러한 검사 후에도 진단을 못하는 경우가 많았다. 그러나 전체 엑솜 시퀀싱, 전체 유전체 시퀀싱과 같은 유전체 분석법은 위에서

언급한 것처럼 이러한 진단 기간 및 정확도를 향상시켜 사망 전에 진단되는 환자들의 비율이 증가되었으며, 수 주 내에 유전 진단을 하여, 부부에게 다음번 임신 시 태아가 같은 유전 질환으로 이환될 가능성을 예측하고 이를 예방하는 데에 활용하게 되었다.

본인이 유전질환에 이환된 성인의 경우에도 자녀가 같은 질환으로 이환될 가능성으로 인한 두려움으로 임신을 꺼려하는 경우가 있는데, 이러한 경우에도 유전체 분석을 통한 정밀의료가 중요한 역할을 하게 된다.

원인 유전자의 돌연변이를 알고 있는 경우, 현재 태아에서 확인 가능한 유전진단 법으로는 용모막, 양수, 제대혈 채혈 등이 있으며, 유전질환의 이환율이 높은 경우는 착상 전 유전 진단을 수행할 수 있다. 착상 전 유전 진단은 부모의 정자와 난자를 체외 수정 후 배아를 수일 배양 후 일부 세포를 떼어내어 유전자 검사를 시행하여 유전자 돌연변이가 없는 배아를 엄마 자궁에 착상시키는 방법으로 유전질환에 이환되지 않은 태아를 임신하는 효과적인 방법이다.

다) 유전체 분석 이차 소견의 의미

한 사람의 유전체를 분석하면 개인당 약 10만 개의 유전자 변이가 발견이 된다. 이러한 변이들은 환자의 증상과 관계가 없는 변이들이 대부분인데, 유전체 분석으로 인해 발생하는 이러한 부차적 혹은 이차적으로 발생하는 소견을 어느 범위까지 검사한 당사자에게 공개해야 하는지에 대해 논란이 있다. 이의 중요한 윤리적 근거 중 하나가 유전자 검사를 받는 당사자는 유전적 원인에 대하여 알 권리도 있으나, 알고 싶지 않을 권리도 있기 때문이다. 또한, 정보 제공의 정당성에 대한 충분한 고려 없이 환자의 증상과 무관한 이차 소견의 정보를 제공할 경우 환자에게 심각한 정신적 고통과 이로 인한 부작용을 일으킬 수 있다.

그럼에도 불구하고 일부 질환은 무증상기 혹은 질환의 조기에 발견하여 관리한다면, 해당 환자의 생존 기간 및 삶의 질 향상에 기여할 수 있다. 이러한 관점에서 미국의학유전 학회에서는 유전체 분석을 한 환자에서 환자의 증상과는 무관하더라도 가족성종양증후군, 선천성 심장병, 윌슨병, 고지혈증, 고암모니아혈증, 악성고열증 등에 관련된 60여개의 유전자들에 변이가 있을 경우 이를 검사 당사자에게 알리고 이에 따른 의료를 시행하도록 권고하고 있다. 가령 망막색소변성으로 전체 엑스 시퀀싱을 한 30세 여성에게 경우

망막색소변성의 원인 유전자 중 하나인 EYS 유전자의 돌연변이 외에 가족성종양증후군의 일종인 BRCA 유전자에 돌연변이가 이차 소견으로 발견된다면, 이를 환자에게 알려서 조기 유방암 검사를 시행 받음으로 유방암의 발생에 의한 사망의 위험을 줄일 수 있다.

유전체 분석에 따르는 이차 소견의 공개 범위는 국가마다 윤리적 사회적 법적 관점이 다르고, 또 시대에 따라 변화하기 때문에 공개 범위는 향후 지속적으로 변경될 가능성이 높다.

라) 유전체 분석에서 유전 상담의 중요성

유전체정보기반의 정밀의료 시스템에서 주요한 의사 결정 과정은 환자와 의사의 논의에 따라 이루어지게 되는데, 이런 과정에서 중요한 역할을 하는 것이 유전질환에 대한 전문적 지식 및 경험에 기반을 둔 유전상담 서비스이다. 유전상담은 유전질환을 유추하기 위한 가계도 작성에서부터, 유전 검사의 과정, 검사의 필요성, 결과가 양성일 경우 혹은 음성일 경우의 해석, 결과에 따른 질환의 관리 및 치료법, 가족 구성원 검사의 필요성, 임신 계획 상담 등에 이르는 포괄적 의료 서비스이며, 유전 진단 후 환자가 겪게 되는 정신적, 정서적 문제에 대한 상담의 과정까지도 포함하게 된다 <그림 3.4>.

전문적인 유전상담은 유전학의 지식을 가진 전문의, 환자의 주치의 혹은 간호사와 같은 의료인이 될 수도 있고, 유전상담학을 전공한 전문 유전상담사가 맡을 수도 있다. 우리나라와 같이 짧은 진료 시간에 많은 환자를 상담해야 하는 의료 시스템에서는 담당 의료인이 상세한 유전상담 서비스를 수행하기는 현실적으로 매우 어렵다. 따라서 향후 유전체정보기반 정밀의료의가 점점 현실화되고 보편화됨에 따라 유전상담사의 필요성 및 역할이 증대될 것으로 보인다. 현재 우리나라의 경우 유전상담사가 국가 자격증이 부여되는 직업군으로 인정받지는 못하고 있으나, 의학유전학 관련 학회 및 기관에서 유전상담사 교육 프로그램을 시행하고 있고, 일부 대학에서 유전상담사 과정의 정규대학원 제도를 실시하여, 이를 통해 유전상담사 인증 사업을 하고 있으며, 머지않은 미래에 유전상담사가 정식 직업군으로 인정될 것으로 예측된다.

4) 향후 발전방향

유전체정보 분석 기술은 눈부신 속도로 발전하고 있어서, 적은 비용과 적은 시간에 많은 사람의 유전체 검사를 할 수 시대에 도달하게 되었다. 이러한 검사를 통하여 생성되는 개개인의 유전체정보는 매우 방대하여, 어느 한 단체 혹은 기관에서 관리할 수 없을 수준이 될 것이며, 따라서 이 거대한 양의 바이오 데이터는 범국가적 차원에서 관리하고 연구 및 의료에 활용하도록 하여, 전체 국민이 정밀의료의 혜택을 누리도록 해야 한다.

희귀유전질환은 그 희귀성으로 인해 유전체분석 기술의 발전에도 불구하고, 개개인에서 원인 유전자의 이상을 찾아내는 데에 많은 시간과 노력이 필요하다. 분석과정에서도 희귀유전질환의 유전체 이상에 대한 지식을 지속적으로 업데이트해야 하며, 이러한 연유로 분석하는 과정에서 오류가 발생할 위험이 있다. 희귀유전질환은 조기 진단을 통하여 빠른 관리와 치료가 환자의 삶의 질에 매우 중요하기 때문에 유전체 분석 시간을 최대한 단축하고, 분석 오류를 최소화하기 위한 노력이 필요하다. 따라서 향후 유전체 분석에 있어서 인공지능을 이용한 유전체 분석 기술이 큰 역할을 할 것으로 본다. 인공지능을 이용하면 분석 시간의 최소화 및 희귀유전질환 관련 유전체정보의 실시간 업데이트를 통하여 진단 오류의 위험성을 줄일 수 있어 유전체 분석가들에게 유용한 도구로 활용될 수 있을 것으로 예상된다.

개인의 유전체정보는 개인 식별, 다양한 유전질환의 위험 인자 등 민감한 개인정보가 포함되어 있어서, 이를 관리하고 유용함에 있어서 철저한 보완 및 개인정보 보호가 필요하다. 그러나 과도한 법률적 제한은 바이오 데이터를 이용한 첨단 연구를 수행하는 데에 오히려 걸림돌이 될 수 있어서, 윤리적 사회적 법률적 공감대가 허용되는 범위에서 국민의 건강관리 및 의료 서비스 향상을 위해 연구의 충분한 자율성 및 창의성을 보장해 주어야 한다. 바이오 데이터를 사용한 의료 사업은 경쟁이 매우 치열한 분야로서, 첨단 의료 관련 기술 및 치료법을 선도하는 것이 매우 중요한바 법률적 제도가 이의 걸림돌이 되어 각종 선도 연구의 시기를 놓치는 일은 없어야 하겠다.

참고문헌

- Atwal, P. S., Brennan, M. L., Cox, R., Niaki, M., Platt, J., Homeyer, M. *et al.*(2014). “Clinical whole-exome sequencing: are we there yet?”, *Genetics in Medicine*, 16(9), pp.717-719. (DOI: 10.1038/gim.2014.10).
- Clark, M. M., Stark, Z., Farnaes, L., Tan, T. Y., White, S. M., Dimmock, D. *et al.*(2018). “Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases”. *npj Genomic Medicine*, 3(1), p.16. (DOI: 10.1038/s41525-018-0053-8).
- Desnick, R. J., Schuchman, E. H.(2002). “Enzyme replacement and enhancement therapies: lessons from lysosomal disorders”. *Nature Reviews Genetics*, 3(12), pp.954-966. (DOI: 10.1038/nrg963).
- Dixon-Salazar, T. J., Silhavy, J. L., Udpa, N., Schroth, J., Bielas, S., Schaffer, A. E. *et al.* (2012). “Exome sequencing can improve diagnosis and alter patient management” . *Science Translational Medicine*, 4(138). (DOI: 10.1126/scitranslmed.3003544).
- Feriozzi, S., Hughes, D. A.(2020). “New drugs for the treatment of Anderson-Fabry disease”. *JN: Journal of Nephrology*, [Epub ahead of print]PMID: 32193835. (DOI: 10.1007/s40620-020-00721-4).
- Kalia, S. S., Adelmanm K., Bale, S. J., Chung, W. K., Eng, C., Evans, J. P. *et al.*(2017). “Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics”. *Genetics In Medicine*, 19(2), pp.249-255. (DOI: 10.1038/gim.2016.190).
- Kim, Y. M., Yum, M. S., Heo, S. H., Kim, T., Jin, H. K., Bae, J. S. *et al.*(2020). “Pharmacologic properties of high-dose ambroxol in four patients with Gaucher disease and myoclonic epilepsy”. *Journal of Medical Genetics*, 57(2), pp.124-131. (DOI: 10.1136/jmedgenet-2019-106132).

- Kingsmore, S. F., Cakici, J. A., Clark, M. M., Gaughran, M., Feddock, M., Batalov, S. *et al.*(2019). “A Randomized, Controlled Trial of the Analytic and Diagnostic Performance of Singleton and Trio, Rapid Genome and Exome Sequencing in Ill Infants”. *The American Journal of Human Genetics*, 105(4), pp.719-733.
- Mercuri, E., Muntoni, F., Osorio, A. N., Tulinius, M., Buccella, F., Morgenroth, L. P. *et al.*(2020), “Safety and effectiveness of ataluren: comparison of results from the STRIDE Registry and CINRG DMD Natural History Study”. *Journal of Comparative Effectiveness Research*, 9(5), pp.341-360. (DOI: 10.2217/ce-2019-0171).
- National Human Genome Research Institute (NHGRI). <https://www.genome.gov>
- Ng, S. B., Turner, E. H., Robertson, P. D., Flygare, S. D., Bigham, A. W., Lee, C. *et al.*(2009). “Targeted capture and massively parallel sequencing of 12 human exomes”, *Nature*, 461, pp.272-276. (DOI: 10.1038/nature08250).
- Schuster, S. C.(2007). “Next-generation sequencing transforms today's biology.”, *Nature Chemical Biology*, 5(1) pp.16-18. (DOI: 10.1038/nmeth1156).
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. *et al.*(2001). "The sequence of the human genome", *Science*, 291, pp.1304-1351. (DOI: 10.1126/science.291.5507.1155d).
- Yang, Y., Muzny, D. M., Reid, J. G., Bainbridge, M. N., Willis, A., Ward, P. A. *et al.*(2013). “Clinical whole-exome sequencing for the diagnosis of mendelian disorders”. *New England journal of medicine*, 369(16), pp.1502-1511. (DOI: 10.1056/NEJMoa1306555).
- Yau, T. O.(2019). “Precision treatment in colorectal cancer: Now and the future”, *JGH Open*, 3(5), pp.361-369. (DOI: 10.1002/jgh3.12153).

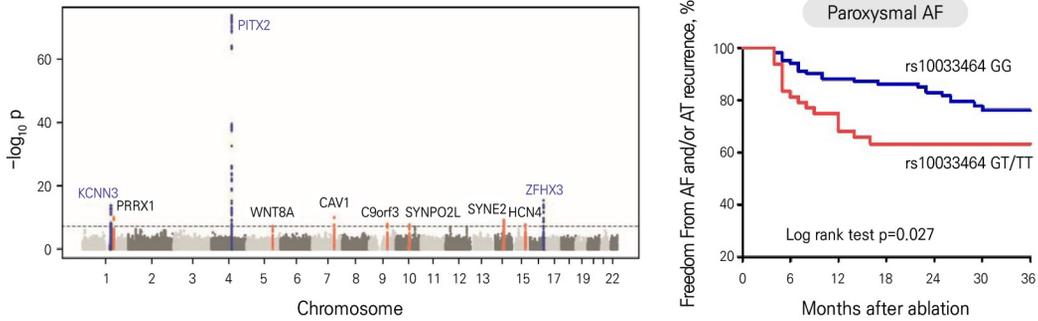
나. 유전성 심장질환

1) 심혈관질환 유전체 연구

인간게놈 서열해독 프로젝트 완료 이후 심혈관 의료 영역에서도 유전체정보를 이용한 많은 연구들이 진행되었는데, 그 중 단일염기다형성(single nucleotide polymorphism, SNP)을 이용한 전장유전체 상관성분석(Genom-wide association study, GWAS)이 대표적이라 할 수 있다. 미국 국립 인간유전체 연구소(National Human Genome Research Institute, NHGRI)의 보고에 따르면 GWAS를 이용한 여러 형질 연구 중 심혈관 질환 또는 위험인자가 상당수를 차지하고 있으며, 일반 인구 집단에서의 GWAS 연구를 통해 심전도 소견과 여러 유전자 위치(loci)의 연관성을 보고하였다 (예: QT interval과 *NOS1AP*). 또한, 약제 효능 또는 부작용과의 연관성도 주요 연구 분야인데, *SLCO1B1* 유전체 연구에서 그 변이와 스타틴 유발 근육병증이 강한 연관성이 있고, 유전체분석(genotyping)을 통해 안전하고 효과적인 스타틴 치료를 가능하게 할 수 있음을 제시한 바 있다.

심혈관 질환중 뇌경색, 심부전 등 심각한 합병증을 일으키는 부정맥인 심방세동은 여러 후천적 요인(고령, 음주, 고혈압, 판막성 심장질환 등)이 발병의 주 위험 인자인 것으로 알려져 있는데, 대규모 GWAS 메타분석 연구에서 폐정맥 발달 단계에서의 대칭성 및 심근 소매(myocardial sleeve) 형성에 관여하는 전사 인자를 코딩하는 염색체 4q25에 위치하는 *PITX2*가 심방세동 발병과 밀접한 연관이 있는 것을 확인하였으며, 이후 4q25 유전자 단일염기다형성을 약물, 전기제세동, 및 전극도자절제술 등의 치료 후 재발 여부 예측에 이용한 연구들을 통해 유전적 소인이 심방세동에서의 발병 요인 중의 하나임을 제시하였다 <그림 3.6>.

그림 3.6 심방세동 환자 GWAS 연구에서 발견된 9개 유전자 위치 메타분석 Manhattan plot (좌); 단일염기다형성에 따른 발작성 심방세동 전극도자절제술 재발 예측 (우)



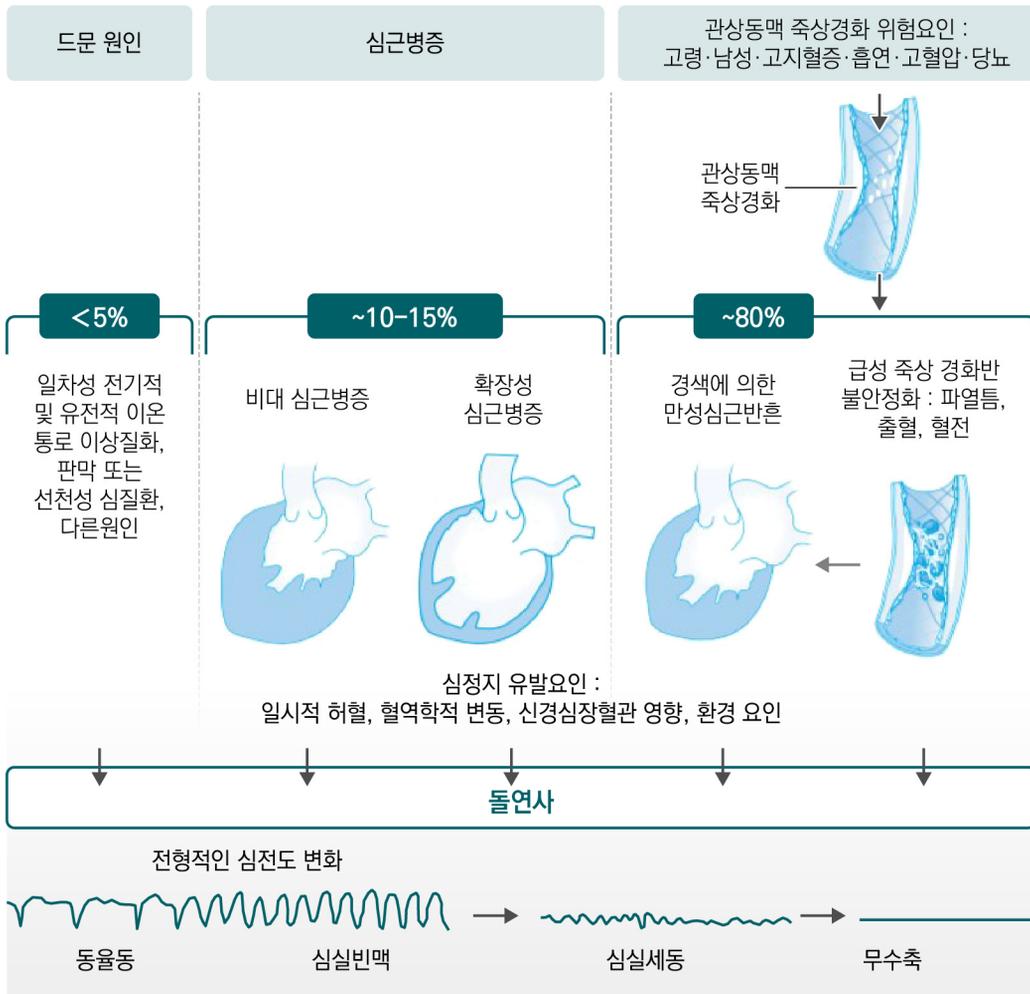
자료: Ellinor, P. T. *et al.*, 2012
Choi, J. I *et al.*, 2019

또한, 약물 유전체학(pharmacogenetics) 등을 통해 실제 임상에서 심혈관질환에 적용되고 있다. 뇌졸중과 혈전증 예방으로 사용되는 항응고제인 와파린 대사에 관여하는 *CYP2C9*와 *VKORC1*의 유전체 변이 여부에 따라 와파린 용량 조절을 하기도 한다.

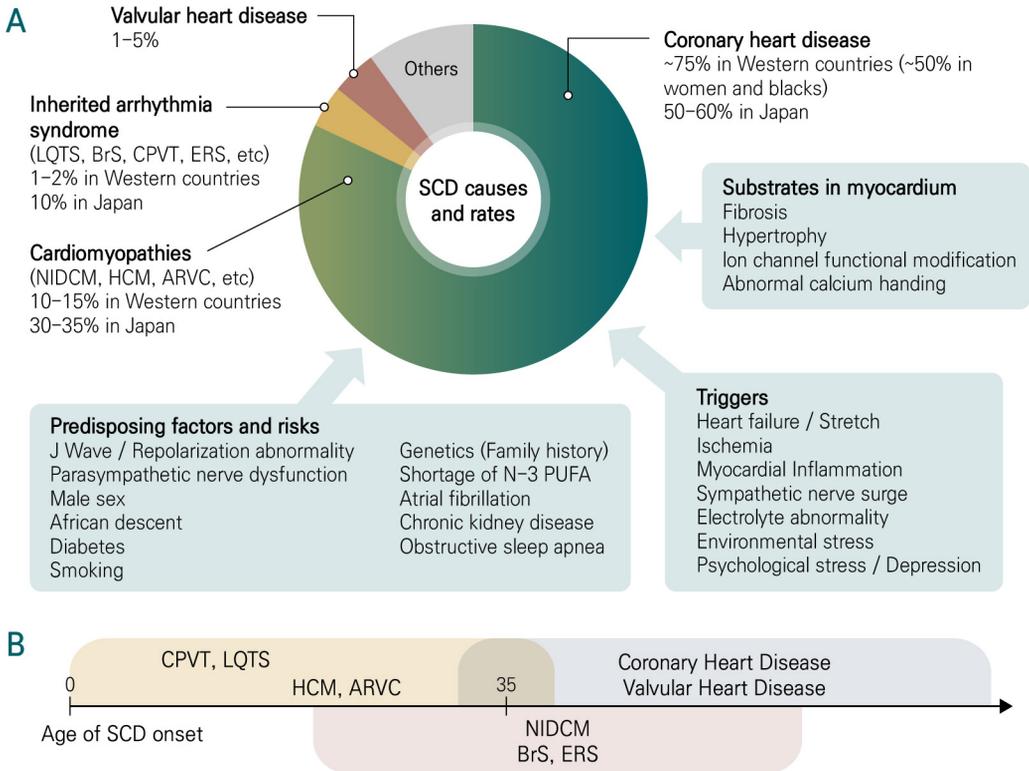
2) 유전성 심장질환 임상적 특징

유전자 변이에 의한 직접적인 이상으로 발생하는 심혈관질환으로는 이온채널병증(channelopathy), 심근병증(cardiomyopathy), 대동맥혈관병증(aortopathy) 등이 있다. 이중 유전성 부정맥(inherited arrhythmia) 또는 일차성 전기적 질환(primary electrical disorder)이라고도 알려져 있는 이온채널병증은 구조적인 심장 이상 소견 없이 유전적 요인에 의해 심실성 부정맥이 유발되어 돌연심장마비 또는 급사를 일으키는 질환이다. 급사의 원인으로는 관상동맥질환, 심근병증, 이온채널병증 등이 있는데, 서양의 보고에 의하면 주요 원인인 관상동맥질환이 전체의 60~80%, 유전성 부정맥은 5% 이하를 차지한다고 알려져 있었으나, 최근 일본에서의 보고에 의하면 유전성 부정맥이 약 10%를 차지한다고 알려졌다 <그림 3.7>. 우리나라 건강보험공단 데이터를 이용한 최근 연구에서도 유전성 부정맥이 전체 급사의 원인 중 약 14%까지 차지하는 것으로 보고되었는데, 이를 통해 아시아인, 특히 한국인에서 급사 원인으로서의 유전성 부정맥 유병률이 높은 것으로 추정되고 있다.

그림 3.7 돌연심장마비 또는 급사의 주요 원인



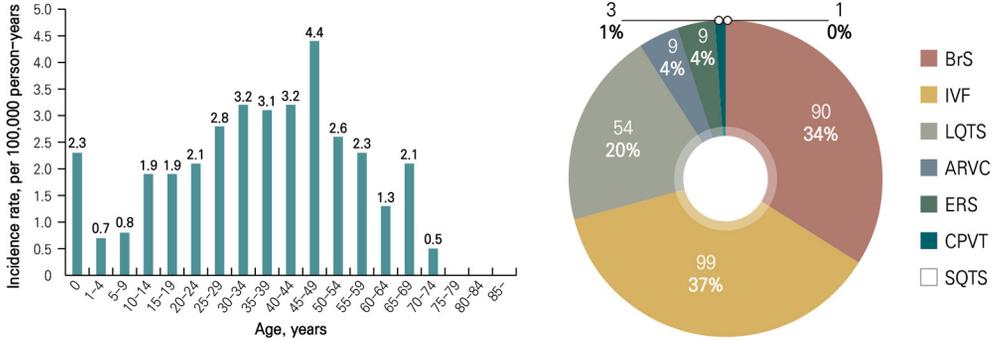
자료: 최종일, 2012; Huikuri, H. V. *et al.*, 2018



자료: Hayashi, M., Shimizu, W., Albert, C. M., 2015

유전성 부정맥은 이온채널 단백을 부호화(encoding)하는 유전자 변이에 의해 발생하는 질환으로, 긴 QT 증후군(long QT syndrome, LQTS), 브루가다 증후군(Brugada syndrome), 카테콜아민성 다형심실빈맥(catecholaminergic polymorphic ventricular tachycardia), 짧은 QT 증후군(short QT syndrome), 재분극 증후군(early repolarization syndrome), 특발성 심실빈맥(idiopathic ventricular fibrillation), 부정맥 유발 우심실 이형성 심근병증(arrhythmogenic right ventricular cardiomyopathy, ARVC) 등이 있다. 우리나라 급사 예방을 위한 제세동기 적응증 원인 질환으로서 유전성 부정맥이 차지하는 비율은 서양에 비해 높은 편이며, 국내 15개 대학병원에서 진행된 다기관 코호트 연구에 의하면 브루가다 증후군, 특발성 심실세동, 긴 QT 증후군의 순으로 보고되었다 <그림 3.8>. 그 중 긴 QT 증후군, 브루가다 증후군 및 부정맥 유발 우심실 이형성 심근병증에 대해 기술하고자 한다.

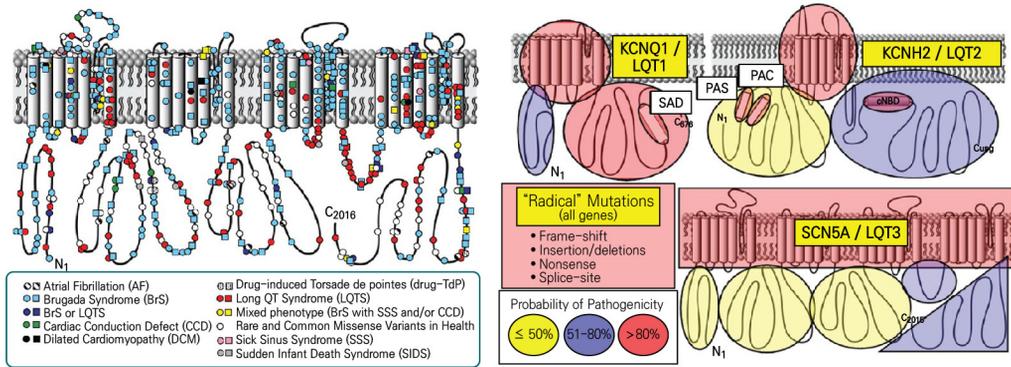
그림 3.8 우리나라 유전성 부정맥 연령별 발병률(건강보험 표준 코호트(좌));
원인 질환 대기관 레지스트리 연구 결과(우)



자료: 노승영·최종일 등, 2018, 미국심장학회 연례학술대회.; 오석규·김윤기·최종일 등, 2018, 대한심장학회 추계학술대회.

긴 QT 증후군은 심전도상 QTc 간격이 증가되어 있을 때 진단이 되며(QTc >500ms 또는 LQTS risk score >3.5 또는 원인 불명의 실신 동반시 QTc 480-499 ms), 경험이 많은 기관에서 시행한 유전자 검사상 병적인(pathologic) LQTS 변이 확진 시 QT 간격에 상관없이 긴 QT 증후군으로 진단할 수 있다. 유전자 검사상 LQTS 유전자 변이 이상으로 확인된 긴 QT 증후군 환자의 약 20~25%에서는 QTc 간격이 정상 범위인 것으로 알려져 있다. 현재까지 긴 QT 증후군과 연관된 17개의 유전자 변이가 발견되어 알려져 있으며, 그 진단 양성률은 약 60~70%이다. 그중 3가지 유전자 *KCNQ1*(loss-of-function of I_{Ks} -conducting K^+ channel, $K_v7.1$), *KCNH2*(loss-of-function-of I_{Kr} -conducting K^+ channel, $K_v11.1$), *SCN5A*(gain-of-function of I_{Na} -conducting Na^+ channel, $Na_v1.5$)가 양성으로 판명된 환자가 전체의 약 90%를 차지하며, 그 유전자 변이형에 따라 임상 예후에 대한 차이를 보이기도 한다 <그림 3.9>. 또한, 긴 QT 증후군의 유전자형에 따른 베타차단제 약물치료 효과의 차이가 있음이 메타분석에서 보고되기도 하였다. 그러나 유전자 검사 이외에도 임상 양상, 심전도 소견 등을 종합적으로 고려한 환자 개개인에 대한 위험도 평가가 이루어져야 한다.

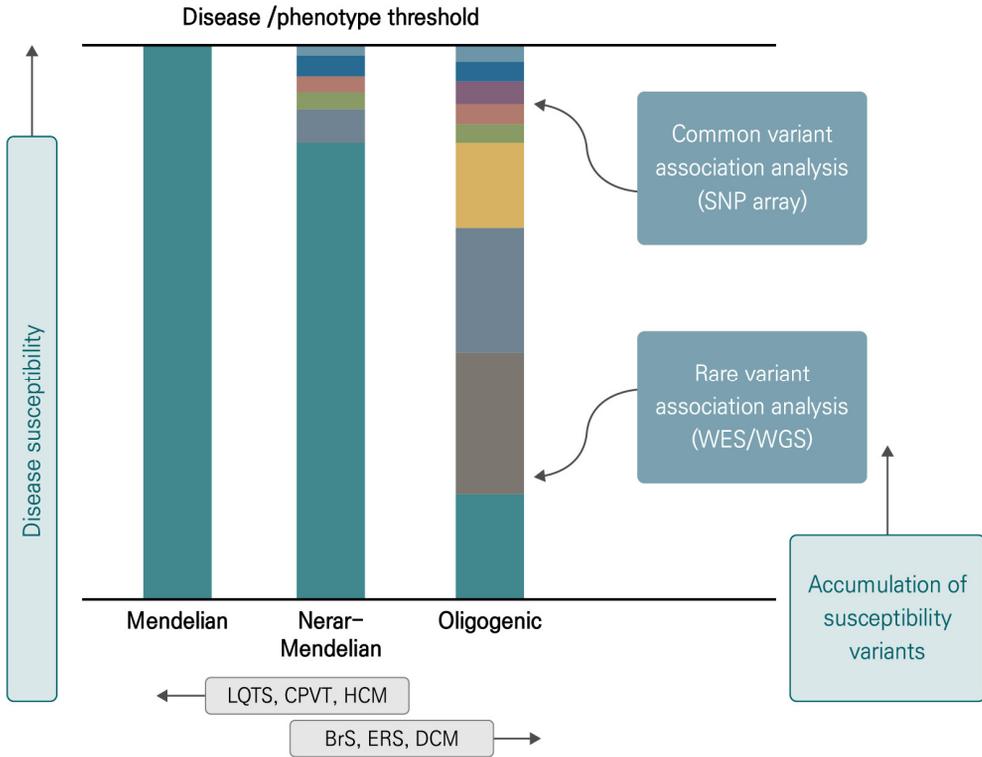
그림 3.9 이온채널병증의 유전자 변이 및 변이 위치에 따른 질환 증증도



자료: Tester, D. J., Ackerman, M. J., 2011

브루가다 증후군은 우리나라를 포함한 아시아에서 발병률이 높는데, 특징적인 심전도 소견(type I, coved ECG)만으로 진단이 가능하나, 일반적인 위치보다 상향 유도 또는 flecainide 등 소듐 통로 차단제 약물 유발 검사가 진단에 도움이 되기도 한다. 상염색체 우성 유전 형태를 보이는 것으로 알려져 있는데 심장 근육 세포에 존재하는 이온채널에 관련된 유전자 변이에 의해 발생되며 현재까지 21개의 유전자에서 변이가 발견되었다. *SCN5A* loss-of-function 변이가 대표적인 유전자 이상이며 ClinGen에서의 브루가다 증후군 관련 undisputed 유전자인 것으로 알려져 있다. 최근 일본에서 *SCN5A* 변이 유무에 따라 임상적 예후의 유의한 차이를 보이는 것이 보고되기도 하였다. 유전성 질환임에도 40대 남성에서 발병이 가장 흔한 것으로 보고되는데, 이는 유전성 심장질환의 유전 이환 특성에 따른 질환 발병 감수성 차이에 의한 것으로 설명되기도 한다 <그림 3.10>.

그림 3.10 유전성 심장질환에서의 유전적 변이와 질병 발현



자료: Bezzina, C. R., Lahrouchi, N., Priori, S. G., 2015

부정맥 유발 우심실 이형성 심근병증은 심근세포 사이의 결합을 담당을 구조물인 테스모솜(desmosome)과 관련된 유전자의 변이가 원인으로, plakophilin-2(*PKP2*), desmoplakin(*DSP*), desmocollin-2(*DSC2*), desmogelin-2(*DSG2*), plakoglobin(*JUP*) 등이 알려져 있다. 심근 세포 간 결합력이 약해지고 심근세포가 과사되면서 지방 및 섬유 조직으로 대체되고 우심실이 변형되어 회귀 기전(reentry)에 의해 심실성 부정맥이 발생하는 것으로 알려져 있다. 유전자 변이 검사가 양성인 환자에서의 임상 표현형은 다양하지만 점차 진행하여 20세 전에는 10% 정도이나 60세까지 50%의 환자에서 임상적 질환으로 발현된다. 심장초음파 또는 심장 MR 등의 영상 검사 및 심근 조직 검사가 진단에 중요하며, 최근 유전자 검사가 진단에 이용되기도 하는데, 특히 가족력이 있는 경우에는 유전자 검사가 권고된다.

유전성 부정맥은 ① 첫 증상이 심실빈맥이나 심실세동 등 치사성 부정맥에 의한 급사나 실신(syncope)으로 나타나며, ② 특징적인 심전도 소견이나 정밀 검사 없이는 조기 진

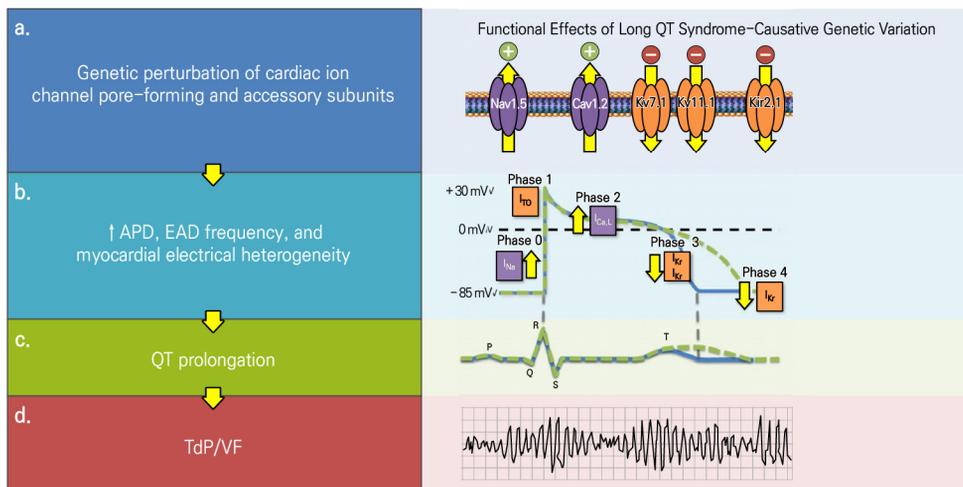
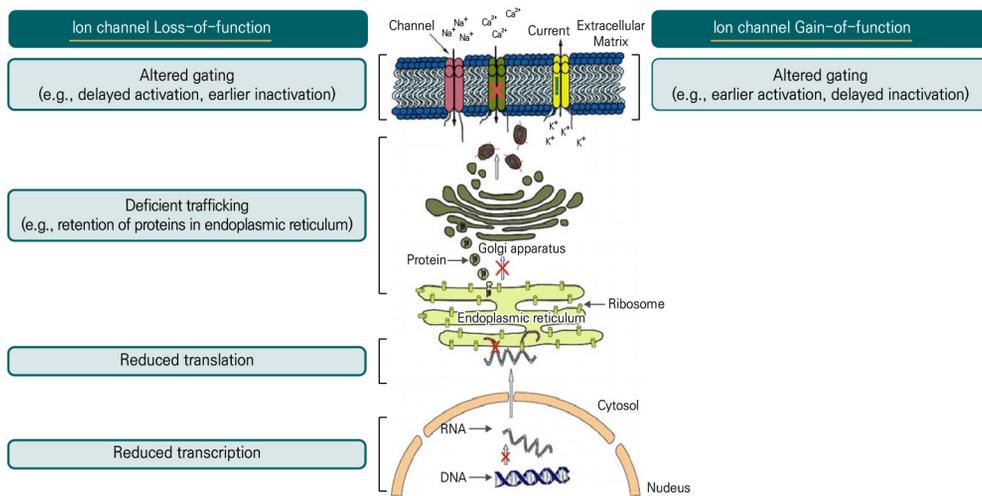
단이나 예측이 어렵고, ③ 유전성 원인, 가족성 이환으로 완치가 불가능하며, ④ 대부분 젊은 나이에 발병하므로 사회경제적인 파급 효과가 크고, ⑤ 급사 예방을 위해서는 제세동기 삽입 치료를 받아야 한다는 점이 임상적 특징이다. 그럼에도 불구하고, 유전성 부정맥 환자들은 의료 혜택의 사각지대에 놓인 경우가 드물지 않다. 따라서 보건학적으로 다음과 같은 사항들이 고려되어야 한다. ① 급사(돌연사, 심정지)라는 중대한 결과를 초래할 수 있는 질환임에도 불구하고 현재 긴 QT 증후군, 카테콜아민 다형성 심실성빈맥만 희귀질환으로 지정되어 있어 브루가다 증후군등 대표적인 유전성 부정맥들의 추가 희귀질환 지정이 필요하다. ② 통계청 코드상 각각에 대한 세분화 없이 I42.8, I49.8 두 가지 KCD 질병코드에만 등록되어 있는 실정이므로, 각각 병명에 따라 질병 코드를 세분화하는 것이 환자의 등록, 관리 및 치료에 필요하다 (제8차 한국표준질병사인분류 개정안 발표 예정으로 대한심장학회/대한부정맥학회를 통해 관련 의견서가 제출되었다). ③ 현재는 심율동전환 제세동기 거치술을 시행 받을 때만 중증질환 본인부담금 산정특례 고시 대상에 해당되어 혜택을 받고 있으나 이에 대한 확대 지정이 필요하다. ④ 급사 예방을 위해 심율동전환 제세동기 거치술을 받은 경우 심장장애 등급대상으로 선정될 수 있게 할 필요가 있다. 일본의 경우 인공심장박동기를 삽입한 경우에도 장애등급으로 되어 있으나, 우리나라는 현재 3점의 점수만 부과되어 심장 장애등급 대상이 되지 못하는 실정이다. ⑤ 2017년부터 차세대염기서열분석(next generation sequencing, NGS) 기반 유전자 패널 검사가 요양 급여 대상이 되었으나, 가족에서의 유전자 검사 등 급여 기준 확대를 통해 희귀질환의 보다 정확한 진단이 가능해질 것으로 기대된다.

3) 유전성 부정맥에서의 유전자 이상

심근세포 활동 전위(action potential)의 발생으로 심장의 전기적 활동이 나타나는데, 심근의 sarcolemma에 발현된 이온통로의 개폐에 의해 생성되며, 이온통로 단백질은 유전적 특성에 의해 심장내 위치에 따라 다양한 정도로 발현된다. DNA의 아미노산 서열에서 뉴클레오타이드의 치환과 결실에 따라 나타날 수 있는 silent mutation, missense mutation, nonsense mutation, frameshift mutation 또는 in-frame mutation의 유전적 변이에 의해 이온채널 단백질 생성에 이상이 발현될 수 있다. 다양한 유전성 또는 후천적 부정맥 질환들은 이온채널의 개폐 과정 기전의 변화와 연관이 있는 각 이온채널 또는 그 부속 단백질을 코딩하는 유전자들의 변이로 인해 유발된다고 알려져 있다. 세포 내

transcription 감소, translation 감소 또는 trafficking 결핍 등 유전 정보를 통한 단백질 생성 과정의 여러 단계에서의 이상으로 인해 gain-of-function 또는 loss-of-function의 이온 채널 기능 이상이 초래될 수 있다. 유전적 변이에 의해 이온통로 전기생리에 관여하는 단백질 생성 이상이 발생되면 심근세포 활동전위 각 단계에서의 이온통로 조절 및 분자생물학적 특성이 변화함으로써 세포내 칼슘이온의 대사 이상이 유발되고 결과적으로 부정맥이 발생한다 (그림 3.11).

그림 3.11 유전성 부정맥에서의 이온통로 기능 이상 분자생물학적 기전

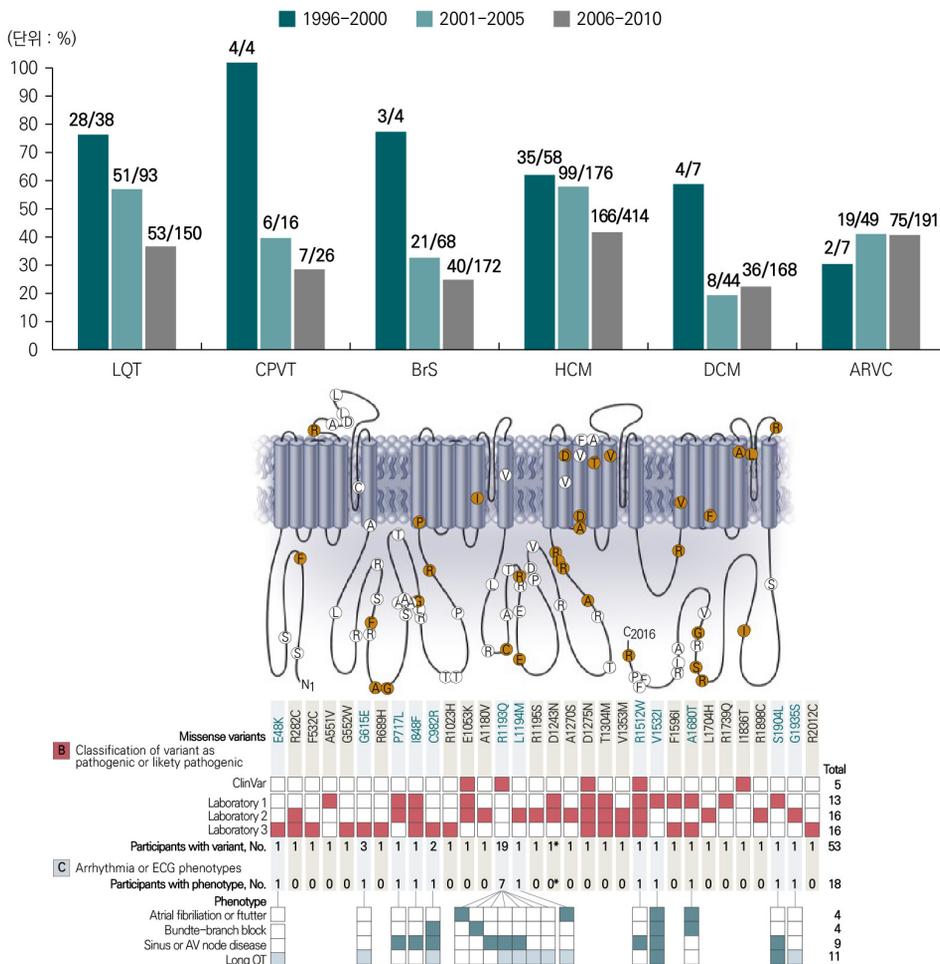


자료: Amin, A. S., Tan, H. L., Wilde, A. A., 2010
 Giudicessi, J. R., Wilde, A. A. M., Ackerman, M. J., 2018

4) 유전성 심장질환에서의 NGS 기반 유전자 검사

1990년대 후반부터 시행된 표적 유전자 검사(targeted genetic testing)의 유전상 부정맥 진단율은 보다 많은 환자에서 검사가 이루어지게 되면서 2000년대 후반부터는 대부분 질환에서 오히려 감소되는 양상을 보였으며, 경험이 많은 기관에서조차 병적인 유전자 변이의 소견을 보이더라도 임상적 진단 및 표현형과의 일치율에 있어 연관이 없는 경우도 많았음을 다기관 연구를 통해 보고하였다 <그림 3.12>.

그림 3.12 유전성 부정맥 표적 유전자 검사의 진단율과 표현형과의 일치율



자료: Amin, A. S., Tan, H. L., Wilde, A. A., 2010
 Giudicessi, J. R., Wilde, A. A. M., Ackerman, M. J., 2018

또한, 유전성 심장질환의 스크리닝을 위한 유전자 검사의 비용이 초기에는 너무 높아 비용 대비 진단 효과에 있어 장점이 없었다. 그러나 NGS를 통해 전체 엑솜 시퀀싱 등 광범위한 유전자 이상을 신속하게 확인할 수 있는 새로운 패러다임의 유전자 검사가 유전성 심장질환에서도 적용할 수 있게 되었다.

NGS 과정을 보면, 인체 genomic DNA가 대략 200~500 base pairs의 작은 파편들로 분할되면 합성된 adaptor sequence들이 파편들 말단부위에 결합하게 되고, adaptor와 연결된 DNA 파편들은 상보적 핵산 염기 결합에 의해 모든 코딩 엑손들과 합체되도록 고안된 biotinylated RNA bait와 혼합되어 배양되고, 표적 DNA 부위가 streptavidin 코팅 자장 구슬에 의해 포획되고 biotinylated bait 가닥에 선택적으로 결합되어 결합된 DNA 파편들을 움직이지 못하게 고정시키면, 세척을 통해 결합하지 못한 DNA와 남은 bait를 제거한 후 시퀀싱을 위한 포획 DNA library가 생성되어, 최종적으로 next-generation sequencer를 통해 모든 포획 DNA들이 배열되면 데이터 분석이 가능해지는 원리이다.

NGS 검사 이전에 임상에서의 유전성 부정맥에 대한 유전자 검사는 긴 QT 증후군 (*KCNQ1*, *KCNH2*, *SCN5A*), 카테콜아민성 다형심실빈맥(*RYR2*), 브루가다 증후군 (*SCN5A*)에 국한되어 시행됐었으나, NGS가 보편화되면서 패널을 이용한 검사들이 시행되었다. <그림 3.13>은 미국 Illumina사에서 제작한 174개의 유전자에 대한 심혈관 질환 패널의 예이다. 그러나 이 패널은 서양, 특히 백인에서의 유전자 데이터를 근거로 유전자 종류가 선택되어 있으므로 인종 간 차이 가능성으로 인해 한국인에서 일률적으로 적용하는데 적절치 않을 수 있다. 또한, 서양에서는 긴 QT 증후군, 아시아에서는 브루가다 증후군이 가장 흔한 유전성 부정맥이라는 각 질환별 인종 간 발병률 차이가 있으므로 인종 간 차이를 고려한 유전자 패널 구성에 대한 고려가 요구된다고 하겠다.

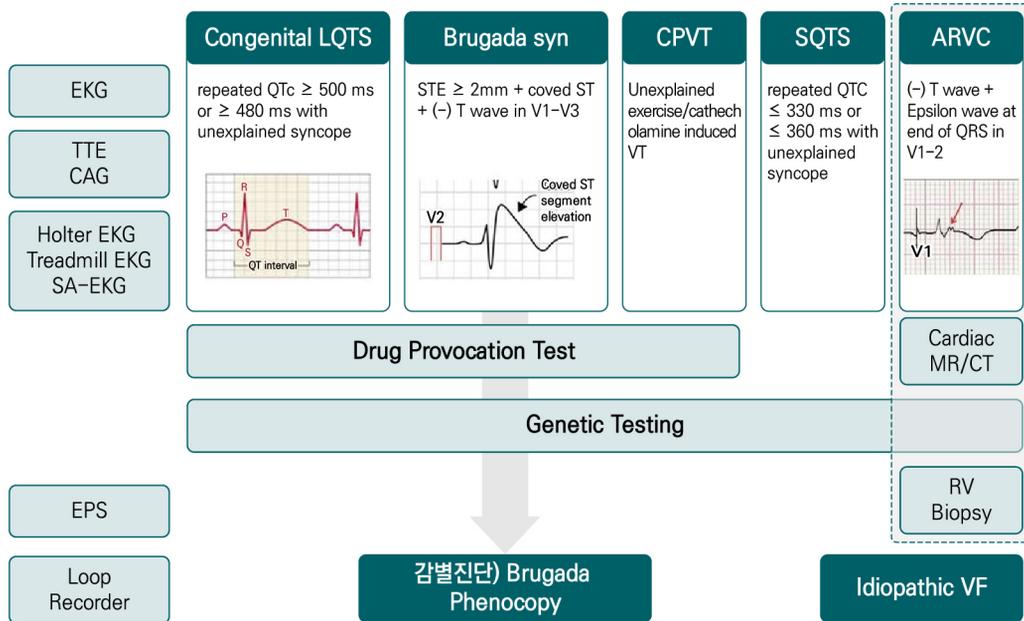
그림 3.13 심혈관질환 NGS 기반 유전자 검사 패널 예(The TruSight Cardio Sequencing Kit, Illumina)

Disease group	Disease	Gene
Inherited Arrhythmia	Long QT syndrome	AKAP9, ANK2, CACNA1C, CALM1, CAV3, KCNE1, KCNE2, KCNE3, KCNH2, KCNJ2, KCNJ5, KCNQ1, RYR2, SCN4B, SCN5A, SNTA1
	Short QT syndrome	CACNA2D1, KCNH2, KCNJ2, KCNQ1
	Brugada syndrome	ABCC9, CACNA1C, CACNA2D1, CACNB2, GPD1L, HCN4, KCND3, KCNE3, KCNH2, KCNJ5, PKP2, RANGRF, SCN2B, SCN3B, SCN5A, TRPM4
	Catecholaminergic polymorphic Ventricular Tachycardia	CALM2, CASO2, KCNE1, KCNJ2, RYR2, TRDN
Cardio-myopathy	Hypertrophic cardiomyopathy	ACTA1, ACTC1, ACTN2, ANKRD1, BRAF, CACNA1C, CALR3, CASQ2, CAV3, COX15, CRYAB, CSRP3, DES, FHL1, FHL2, FXN, GAA, GLA, HRAS, JPH2, KCNQ1, KLF10, LAMP2, MAP2K1, MAP2K2, MYBPC3, MYH6, MYHT, MYL2, MYL3, MYLK, MYO6, MYOZ2, MYPN, NEXN, PDLIM3, PRKAG2, PTPN11, RAF1, SLC25A4, SOS1, TCAP, TNNC1, TNNI3, TNNT2, TPM1, TRIM63, TTN, VCL
	Dilated cardiomyopathy	ABCC9, ACTA1, ACTC1, ACTN2, ANKRD1, BAG3, CRYAB, CSRP3, CTF1, DES, DMD, DNAJC19, DOLK, DSC2, DSG2, DSP, EMD, EYA4, FKBP, FKTN, GATAD1, ILK, JUP, LAMA2, LAMA4, LAMP2, LDB3, LMNA, MURC, MYBPC3, MYH6, MYH7, MYPN, NEXN, NPPA, POLIM3, PKP2, PLN, PRDM16, RBM20, SCN5A, SGC8, SGC8, TAZ, TCAP, TMPO, TNNC1, TNNI3, TNNT2, TPM1, TTN, TXNRD2, VCL, ZBTB17
	Arrhythmogenic right ventricular Cardiodomyopathy	DES, DSC2, DSG2, DSP, JUP, LMNA, PKP2, RYR2, SCN5A, RGF3, TMEM43, TTN
	Restrictive cardiomyopathy	ACTC1, DES, HFE, MYH7, MYL2, MYL3, MTPN, TNNI3, TNNT2, TPM1
	Left ventricular non-compaction	ACTC1, CASQ2, DTNA, MIB1, MYBPC3, MYH7, PRDM16, TAZ, TNNT2, TPM1
	Noonan syndrome	BRAF, CBL, KRAS, MAP2K1, NRAS, PTPN11, RAF1, SHOC2, SOS1
Aortopathies	Marfan syndrome	CBS, FBN1, LTBP2, TGFBR1, TGFBR2
	Loeys-Dietz syndrome	FBN1, TGFBR1, TGFBR2
	Familial aortic aneurysm	ACTA2, COL3A1, EFEMP2, FBN1, MYH11, MYLK, NOTCH1, SLC2A10, SMAD3, TGF2, TGFBR1, TGFBR2
	Aortic valve disease	ELN, FBN1, NOTCH1
Other cardiac disease	Familial hypercholesterolemia	ABCG5, ABCG8, APOA4, APOA5, APOB, APOC2, APOE, CETP, CREB3L3, GCKR, GPIIIBP1, LDLR, LDLRAP1, LMF1, LPL, PCSK9, SREBF2, ZHX3
	Others	ALMS1, COL5A1, COL5A2, CRELD1, DPP6, FBN2, GJA5, HADHA, HSPB8, JAG1, KCNA5, NKX2-5, NODAL, PRKAR1A, RYR1, SALL4, SCN1B, SCD2, SDHA, SEPN1, SMAD4, TBX3, TBX5, TBX20, TTR, ZIC3

우리나라에서는 2017년부터 NGS 기반 유전자 검사에 대해 보험 급여 적용이 되면서, 이온채널병증과 심근병증 등 유전성 심장질환 환자에서도 임상 진료 목적으로 NGS 기반 유전자 검사가 시행되게 되었다. 급여 대상 질환은 유전성 질환으로 분류되어 있으며, 필수 유전자에 대한 규정은 없는 상태이다. Level I(유전자 수가 2~30개이거나 유전자 길이가 150kb 이하인 경우)으로 수가 산정을 받게 되어 대개 30개 이내의 유전자들로 구성된 패널 검사로 진행된다. 그러나 패널에 포함되지 않는 드문 유전자에 의한 경우는 진단이 불가능하며, 환자에서 유전자 변이가 확인되더라도 가족에서의 스크리닝 유전자 검사는 아직 급여 대상에 해당되지 않는 점 등은 향후 해결되어야 할 임상에서의 제한점이라 하겠다.

유전성 심장질환은 임상 진료, 검사, 유전자 검사 및 치료 이외에 유전자 결과의 가족 상담뿐 아니라 신체적/정신적/직업적 지지요법과 치료 등 복잡한 접근법이 요구된다. 임상 현장에서 효율적으로 이를 운영하기 위해서는 유전성 심장질환 클리닉 진료 시스템이 필요한데, 미국이나 유럽과 달리 국내에서는 일부 병원에서만 운영되고 있는 실정이며 <그림 3.14>, 유전자 검사 비용, 판독 전문가 인력 부족 및 유전성 질환에 대한 사회적 편견 등의 여러 요인이 복합적으로 작용된 것으로 생각된다.

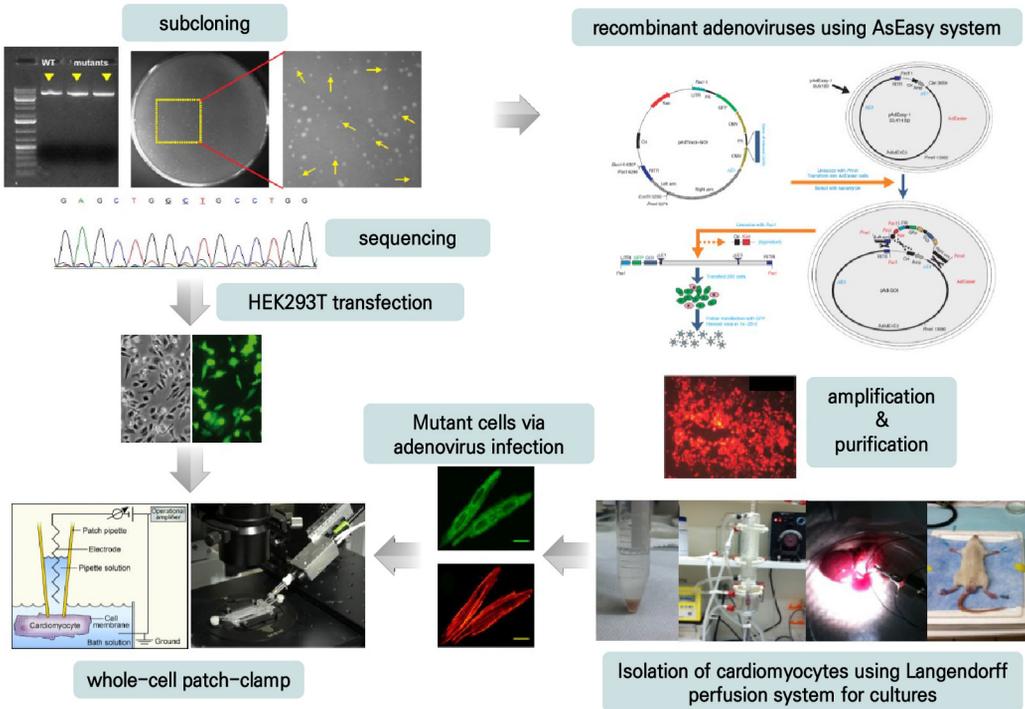
그림 3.14 유전성 심장질환 클리닉(Cardiovascular genetics clinic)에서의 유전성 부정맥 진단 임상 프로세스



자료: 정주희·최종일 등, 2019, 대한심장학회 추계학술대회.

유전성 심장질환에서의 유전자 검사는 진단뿐 아니라 질환에 따라서는 예후 예측 및 치료에 대해서도 임상적 유용성이 있다고 알려져 있다. 특히, 긴 QT 증후군, 카테콜아민성 다형심실빈맥, 비후성 심근병증, 확장성 심근병증 환자에서의 유용성이 높은 편이다. 일반적인 임상 검사로 진단이 어려운 경우에는 환자뿐 아니라 가족 스크리닝(cascade screening) 및 부검 진단(autopsy) 시에도 NGS 기반 유전자 검사를 이용한 분자생물학적 진단(molecular diagnosis)이 도움이 될 수 있어 향후 그 유용성이 확대될 것으로 기대된다. 치료적 측면에서의 적용은 아직까지는 제한적이다. 유전자형 변이에 따른 맞춤형 약물치료(긴 QT 증후군 3형에서의 mexiletine, flecainide, ranolazine)와 유전자 치료(CSQ2-R33Q 유전자 변이 카테콜아민성 다형심실빈맥에서의 flecainide)에 대한 연구결과들이 발표되었으나 실제 임상 치료 적용에 대해선 추가적인 데이터가 필요하다고 볼 수 있다. 유전자 변이 검사에서 양성인 경우 급사 예방을 위한 제세동기 삽입에 대해서도 아직까지는 근거가 부족한 실정이다(recommendation class IIb, level of evidence C). Pathogenicity에 대한 보다 정확한 판정을 위해 이온채널 패치클램프 등 세포기능 평가를 통해 이온채널 기능 이외에 arrhythmogenicity를 파악하는 것이 도움이 되기도 한다(그림 3.15).

그림 3.15 이온채널병증에서의 arrhythmogenicity 측정을 통한 pathogenicity 평가를 위한 세포패치클램프 기능평가(functional test)



자료: Choi, J. I., Wang, C., Pitt, G. S., 2016. PLOS ONE, 11(3). 재구성
 Luo, J. I., Deng, Z. L., Luo, X. A. *et al.*, 2007. nature protocols, 2(5), pp.1236-1247. 재구성

5) 유전자 검사의 현재와 문제점

단일 유전자에 대한 표적 검사에서 NGS 기반 검사로 변화됨에 따라 비용과 시간이 감소되었을 뿐 아니라 여러 유전자들을 포함한 패널 검사가 가능해지면서 진단율이 향상되었다. 그러나 아직까지 임상적으로 이용되는 유전자 검사는 주로 coding region인 exon에 국한된 검사로 intron에서의 이상이나 splicing variant 등은 밝혀내기가 어려울 수 있으며, dominant negative 돌연변이나 미토콘드리아 유전 형태 등의 특이 기전에 의해 발생하는 유전성 심장질환에서도 제한이 있을 수 있다. 따라서 전체 엑솜뿐 아니라 광범위한 전체 유전체(whole genome)에 대한 검사가 향후 임상에서도 적용될 것으로 예상된다.

기술의 발전으로 신속한 검사는 가능해졌으나 대용량 데이터의 filtering을 통해 해석하고,

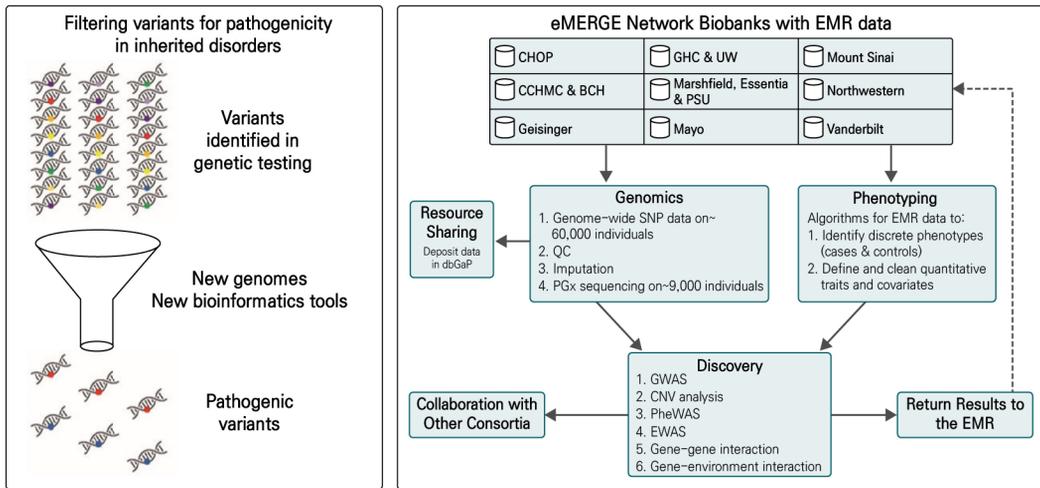
그 결과를 진단, 예후 예측 등 임상에 적용하는 문제가 더 중요한 문제로 대두되었다. 고형암의 경우, 영상 검사에서 종양이 확인되고 치료를 통해 크기 및 종양 표지자 변화 등을 확인할 수 있다. 이와는 달리, 유전성 심장질환에서 임상적 주요 판단 목표는 진단 이외에도 급사 또는 돌연사 등 주요 사건의 발생 여부, 즉 일종의 ‘all or nothing’, 예측이라는 어려움이 있다. 이를 보완하기 위해 미국 의학유전자/유전체 학회(American College of Medical Genetics and Genomics, ACMG)는 데이터 종류와 근거 수준을 종합적으로 고려한 유전자 결과 해석 가이드라인을 제시하였는데, ‘benign’, ‘likely benign’, ‘variants unknown significance(VUS)’, ‘likely pathogenic’, ‘pathogenic’의 5단계로 구분하여 판정하는 것을 권고하고 있다. 이중 VUS는 ‘가능성이 있는 검사 결과 (maybe test result)’로 해석되어지는데, 특히 임상적 표현형(phenotype)이 없거나 약한 경우에는 VUS만을 근거로 진단이나 치료 등의 임상적 결정을 내리기에 어려운 점이 있어 ‘유전학적 연옥(genetic purgatory)’으로 표현되기도 한다. 또한, 가족에서의 유전자 검사는 결과 해석에 주의를 요하는데, 급사의 가족력이 있더라도 무증상 환자에서 병적 양성 결과를 보이는 경우 제세동기 삽입 등의 치료에 대해서는 아직까지 권고하지 않다. 그러나 기능적 검사를 기반으로 한 위험도 평가를 통해 환자 맞춤형 의료로의 임상적 적용이 향후 유전성 심장질환 환자를 대상으로도 가능해질 것으로 기대된다. NGS 기반 유전자 패널 검사 시 우연히 발견되는 다양한 유전자 결과도 문제점 중의 하나이다. 예를 들어, 유전성 심장질환이 의심되는 소아 환자에서 유전자 검사상 대표적 유전성 암유전자인 *BRCA1*의 병적 변이가 우연히 발견된 경우에 그 임상적 판단과 해결에 대해서는 논란의 여지가 있을 수 있다.

유전자 검사 결과로 인해 보험 가입 또는 혜택에 있어 환자와 그 가족들이 차별을 받을 가능성도 제기되어왔다. 2008년 미국에서는 건강보험과 고용에 있어 유전정보로 인해 차별받는 것을 금지하는 유전자 정보 비차별 법안(Genetic Information Nondiscrimination Act, GINA)을 통과시켰다. 그럼에도 불구하고, 급사를 유발하는 유전성 부정맥으로 진단된 200명 환자의 단면 설문 연구(cross-sectional, survey)에 따르면, 답변자(54%) 중 60%에서 보험 가입이 거부되었고, 그 이유의 57%는 이미 진단된 유전성 부정맥 때문이었으며, 보험 신청자 37%에서는 보험료의 증가가 있었다고 보고하였다. 즉, 유전자 검사 건수의 증가가 GINA 법안에도 불구하고 유전성 부정맥 환자와 그 가족들에 대한 보험 가입 및 조건에 부정적인 영향을 끼치고 있다는 것이 확인된 것이다.

우리나라에서도 이에 대한 법률적 보완에 대한 논의가 있어야 할 것이다.

우리나라는 유전자 검사 시행 자격에 대한 기준을 아직까지는 엄격하게 부과하고 있으나, 미국의 23andMe와 같이 소비자를 직접 대상으로 하는 상업 목적의 유전자 검사(direct to consumer, DTC genetic testing) 서비스에 대한 논란이 증가될 것으로 예상된다. 현재는 위증하지 않은 의학적 상태에 대해 일부 기관에서 유전자 검사를 소비자로부터 직접 의뢰받아 시행하고 있으며 이에 대한 확대 적용의 목소리가 높아지고 있는 상황이다. 그러나 급사와 같은 심각한 의학적 상황이 발생할 수도 있는 유전성 심장질환 환자나 가족에서 진료나 검사 등 임상적 평가에 기반한 의학적 판단 없이 bioinformatics 데이터를 기반으로 한 유전자 검사 결과만을 제시하는 것에 대해서는 신중하게 접근할 필요가 있다.

그림 3.16 유전체/유전자 데이터의 임상적 정보로의 적용 플랫폼



자료: Pitt, G. S., 2018
Crawford, D. C., Crosslin, D. R., Tromp, G. *et al.*, 2014

5) 심혈관 정밀의료(Cardiovascular precision medicine)

전체 유전자 시퀀싱 등 대용량 데이터 검사 분석이 보편화될수록 인공 지능 등 데이터 처리 기술의 고도화로 유전체 또는 유전자 정보 해석 등 bioinformatics 분야에서의 발전과 맞물려 의료에서의 획기적인 패러다임 변화가 예상된다 <그림 3.16-좌>. 그러나 심각한 의학적 상태가 발생할 수 있는 유전성 심장질환 환자들에서 단순히 유전자 검사 결과만으로 임상적 판단을 하는 것은 위험성이 있을 수 있다. 즉, 증상, 가족력, 기저 질환, 복용 약물, 검사 소견 등 환자의 다양한 임상적 상태를 종합적으로 고려하여 환자 개개인에 적용시키는 맞춤형 진료가 의료의 주된 형태가 될 것이다. 따라서 이러한 미래 정밀의학의 구현을 위해서는 유전체 또는 유전자 분석 데이터와 함께 의료기관의 환자 임상 정보에 대한 의무 기록을 통합하는 시스템 플랫폼 구축이 요구되는 시점이다. 그 예가 미국 국립 인간 유전체 연구소에서 주도하는 eMERGE(the Electronic Medical Records and Genomics) Network으로 <그림 3.16-우>, 대용량 유전체 데이터와 환자 정보를 기반으로 한 예측을 통해 임상에서 위험군에 대한 선제적인 조치를 가능하게 하고자 하는 것이다. '안젤리나 졸리 효과'의 예처럼 유전성 심장질환 환자에서 심정지 등 위험한 합병증에 대한 예방이 가능하게 하고자 기대하는 것이다.

원인을 밝히지 못한 돌연사 환자의 경우 드문 유전자 이상이 그 원인일 가능성이 높다. 긴 QT 증후군이나 카테콜아민 다형심실빈맥 환자들에 발견되는 calmodulin 단백을 부호화하는 *CALM* 유전자 변이는 매우 드물지만 돌연사 등 심각한 임상 양상을 보이는 것으로 알려져 있으며, 국제 협력 레지스트리 연구를 통해 74명의 환자가 등록되어 그 데이터가 최근 발표되었다. 따라서 유전성 심장질환에서는 희귀질병에 대한 다기관 코호트나 레지스트리 구축의 국제 협력도 반드시 필요하다. 이러한 국제 협력 연구에 대한 우리나라의 참여를 위해서는, 드물지만 임상적으로는 심각할 수 있는 희귀 유전성 심장질환 유전자에 대해서도 NGS 기반 유전자 검사 지원이 요구된다고 하겠다

최근 우리나라의 기관의 주도(the GenomeAsia 100K project) 아시아 지역에서의 유전체 분석결과를 보고하였는데, 이를 통해 백인과는 다른 유전자 정보를 이용한 다양한 연구들이 가능해질 것으로 예측하고 있다. 예를 들면, 아시아에서 출혈 합병증 빈도가 많은 항응고제 치료 시 와파린 대사체에 대한 인종 간 유전체 연구를 통해 환자에게 맞는 용량 조절에 적용하는 것이다. 이는 서양과는 다른 인종, 문화 및 생활 습관을 가진 우리나라에서

유전성 심장질환을 포함한 심혈관질환에서의 유전체 또는 유전자 정보 기반 연구에도 적용시킬 수 있다는 점을 시사한다고 하겠다.

유전체 또는 유전자 정보를 임상에 적용시키고자 하는 의학 흐름의 변화는 궁극적으로 질병의 예방과 함께 환자의 치료 성적을 향상시키고자 하는 개인 맞춤형 정밀의료를 구현시키는 초석이 될 것이다. 더불어, 유전체 기반 정밀의료는 유전적인 원인이 발병 기전이면서 급사와 같이 사회경제적으로 중요하고 의학적으로 위중한 합병증을 일으키는 유전성 심장질환의 연구와 진료에서도 핵심 방향이 될 것으로 확신하는 바이다.

참고문헌

- 최종일(2012). “돌연 심장사 회복 환자의 치료 및 관리” 부정맥 (Arrhythmia), 대한심장학회 부정맥연구회, Part 05, Chapter 02. 군자출판사.
- 최종일(2016). “Genetics and Genetic Therapy for Cardiac Arrhythmia” 심혈관질환 기초 연구 가이드북(Basic Research in Cardiovascular Diseases), Part 4 Arrhythmia: 5. 대한심장학회 기초과학연구회.
- Ahn, J., Kim, H. J, Choi, J. I. *et al.*(2017). “Effectiveness of beta-blockers depending on the genotype of congenital long-QT syndrome: A meta-analysis”. *PLOS ONE*, 12(10). (DOI: 10.1371/journal.pone.0185680).
- Amin, A. S., Tan, H. L., Wilde, A. A.(2010). “Cardiac ion channels in health and disease”. *Heart Rhythm*, 7(1), pp.117-126. (DOI: 10.1016/j.hrthm.2009.08.005).
- Bezzina, C. R., Lahrouchi, N., Priori, S. G.(2015). “Genetics of Sudden Cardiac Death”. *Circulation Research*, 116(12), pp.1919-1936. (DOI: 10.1161/CIRCRESAHA.116.304030).
- Choi, J. I., Baek, Y. S., Roh, S. Y. *et al.*(2019). “Chromosome 4q25 variants and biomarkers of myocardial fibrosis in patients with atrial fibrillation”. *Journal Of Cardiovascular Electrophysiology*, 30(10), pp.1904-1913. (DOI: 10.1111/jce.14104).
- Crawford, D. C., Crosslin, D. R., Tromp, G. *et al.*(2014). “eMERGEing progress in genomics—the first seven years”. *Frontiers in Genetics*, 5. (DOI: 10.3389/fgene.2014.00184).
- Ellinor, P. T., Lunetta, K. L., Albert, C .M. *et al.*(2012). “Meta-analysis identifies six new susceptibility loci for atrial fibrillation”. *Nature Publishing Group*, 44(6), pp.670-675. (DOI: 10.1038/ng.2261).
- Giudicessi, J. R., Wilde, A. A. M., Ackerman, M. J.(2018). “The genetic architecture of long QT syndrome: A critical reappraisal”. *Trends In Cardiovascular Medicine*, 28(7), pp.453-464. (DOI: 10.1016/j.tcm.2018.03.003).

- Hayashi, M., Shimizu, W., Albert, C. M.(2015). "The Spectrum of Epidemiology Underlying Sudden Cardiac Death". *Circulation Research*, 116(12), pp.1887-1906. (DOI: 10.1161/CIRCRESAHA.116.304521).
- Hofman, N., Tan, H. L., Alders, M. *et al.*(2013). "Yield of Molecular and Clinical Testing for Arrhythmia Syndromes Report of 15 Years'Experience". *Circulation*, 128(14), pp.1513-1521. (DOI: 10.1161/CIRCULATIONAHA.112.000091).
- Huikuri, H. V., Castellanos, A., Myerburg, R. J.(2001). "Sudden death due to cardiac arrhythmias", *N Engl J Med*, 345(20), pp.1473-1482. (DOI: 10.1056/NEJMra000650).
- Moss, A. J., Zareba, W., Benhorin, J. *et al.*(1995). "ECG T-wave patterns in genetically distinct forms of the hereditary long Q T syndrome" *Circulation*, 92(10), pp.2929-2934. (DOI: 10.1161/01.cir.92.10.2929).
- Pitt, G. S.(2018). "An update on the journey towards precision medicine in cardiology". *European Heart Journal*, 39(40), pp.3627-3628. (DOI: 10.1093/eurheartj/ehy637).
- Rho, S. Y., Choi, J. I. *et al.*(2018). "Inherited Cardiac Arrhythmias as a Cause of Sudden Cardiac Arrest in Korea: A Cohort From National Health Insurance Service Database". *Circulation*, 136, Issue suppl_1.
- Richards, S., Aziz, N., Bale, S. *et al.*(2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology". *Genetics In Medicine*, 17(5), pp.405-424. (DOI: 10.1038/gim.2015.30).
- Roh, S.Y., Choi, J.I., Kim, M.S., *et al.*(2019). "Trends in the use of implantable cardioverter-defibrillators for prevention of sudden cardiac arrest: A South Korean nationwide population-based study". *Pacing and Clinical Electrophysiology*, 42(8), pp.1086-1094. (DOI: 10.1111/pace.13741).

- Tester, D. J., Ackerman, M. J.(2011). “Genetic testing for potentially lethal, highly treatable inherited cardiomyopathies/channelopathies in clinical practice”. *Circulation*, 123(9), pp.1021-1037. (DOI: 10.1161/CIRCULATIONAHA.109.914838).
- The SEARCH Collaborative Group(2008). “*SLCO1B1* variants and statin-induced myopathy-a genomewide study”. *New England journal of medicine*, 359(8), pp.789-799. (DOI: 10.1056/NEJMoa0801936).
- Van Driest, S. L., Wells Q. S., Stallings, S. *et al.*(2016). “Association of Arrhythmia-Related Genetic Variants With Phenotypes Documented in Electronic Medical Records”. *The Journal of the American Medical Association*, 315(1), pp.47-57. (DOI: 10.1001/jama.2015.17701).
- Yamagata, K., Horie, M., Aiba, T. *et al.*(2017). “Genotype-Phenotype Correlation of SCN5A Mutation for the Clinical and Electrocardiographic Characteristics of Probands With Brugada Syndrome: A Japanese Multicenter Registry”. *Circulation*, 135(23), pp.2255-2270. (DOI: 10.1161/CIRCULATIONAHA.117.027983).

IV. 빅데이터 활용의 신기술





빅데이터 활용의 신기술

1 인공지능 기반의 유전체 분석을 통한 신약 후보물질 발굴: 딥지노믹스

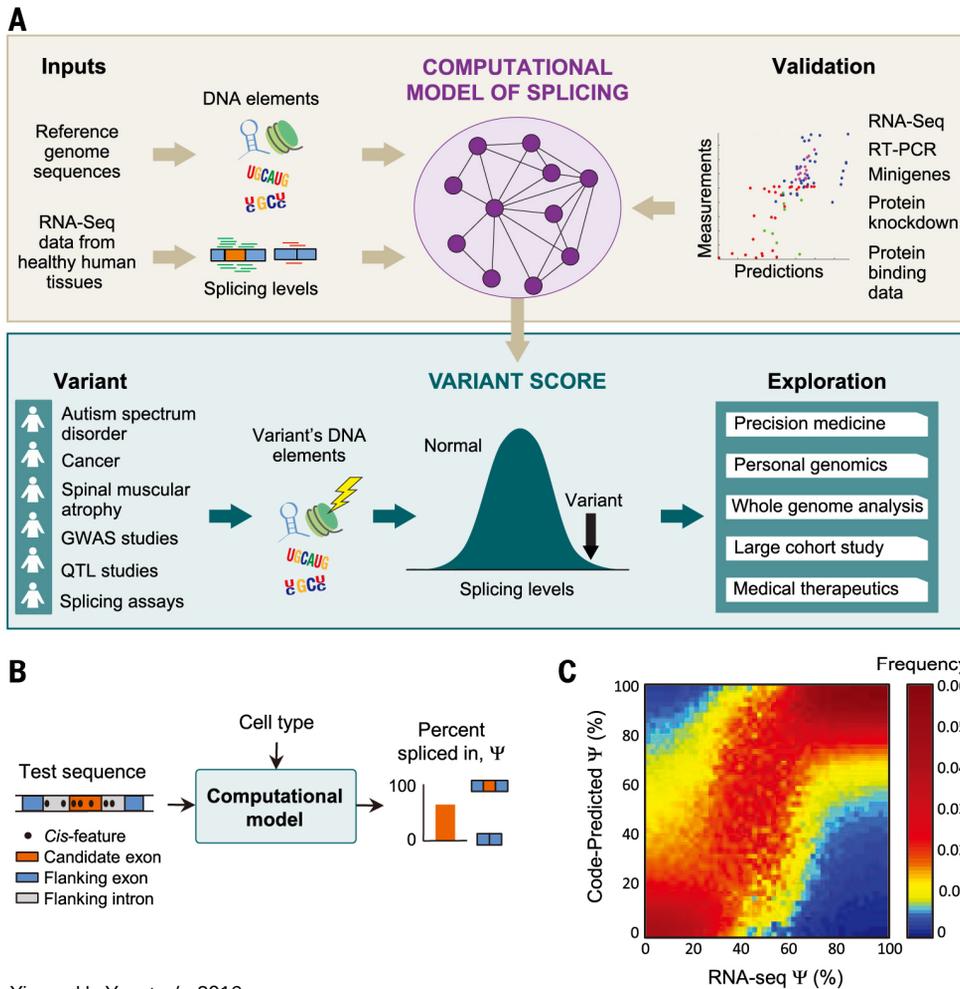
최근 유전체 의학 분야에서도 적용 가능한 강력한 딥러닝 및 인공지능 알고리즘들이 개발되고 있다. 이들 알고리즘은 기존에는 1,000개 이상의 노드(16,000개 이상의 CPU)를 가진 분산형 병렬 처리 컴퓨터에서만 가동 가능했으나 지금은 대규모 기계학습에 적합한 GPU 클러스터가 나와 빠르고 경제적으로 대규모 기계학습이 가능해졌다.

이로 인해 AI 기술을 이용한 유전체 빅데이터 분석 관련 연구 및 비즈니스는 급속히 발전하고 있으며 유전체 빅데이터와 딥러닝 기술이 탑재된 인공지능이 개발되고 있다. ‘유전체 + AI’에 있어 선도적으로 대응하고 있는 대표적인 회사가 바로 딥지노믹스(DeepGenomics)이다. 딥지노믹스는 기계학습과 유전체 빅데이터를 결합해 맞춤형 의료 서비스를 제공하는 것을 목표로 하고 있다.

기계학습을 사용하여 통합된 AI 기술을 이용해 질병을 진단하고 개인 맞춤 치료를 위해 사람의 능력보다 더 뛰어나고 빠르게 환자를 위해 필요한 정보를 임상 의사들에게 제공하고자 하는 것이다. 이를 위해 유전체와 의료 데이터 분석을 통해 작고 미묘한 신호를 찾아 정밀 의학, 유전체 검사 그리고 새로운 신약을 개발하는 것을 사업의 방향으로 잡고 있다.

이 회사는 중추신경계, 눈, 간과 관련된 멘델리안 질병에 집중해 신약을 개발할 계획을 발표했으며, 그 시작으로 멘델리안 질환 치료제의 초기 단계 개발에 집중하고 있다. 특히 단일 유전자 변이로 인해 유전되는 유전질환(monogenic disease)을 집중적으로 다루고 있으며, 특정 질병과 인과 관계가 있지만 검출하기 매우 어려운 유전질환 돌연변이의 분석에 주력하고 있다.

그림 4.1 딥러닝에 기반한 방법으로 RNA스플라이싱에 관여하는 변이 찾기



자료: Xiong, H. Y. *et al.*, 2016

2015년 척수 근육 위축증(Spinal muscular atrophy)과 비용종성 대장암(nonpolyposis colorectal cancer)과 같은 질병들에 관여하는 DNA 패턴을 동정하는 데 있어 딥러닝이 도움이 될 수 있음을 사이언스 지에 논문으로 출판하며 증명하였다. 딥러닝을 통해서 WGS 데이터 중에서 RNA 스플라이싱에 영향을 미치는 intronic/exonic의 유전 변이를 가려내는 연구를 주로 진행했다. RNA 스플라이싱은 유전자 발현에 큰 영향을 미치므로, RNA 스플라이싱에 영향을 주는 유전 변이는 질병을 일으킬 가능성이 크다는 논리이다.

이 2015년 Science 논문을 보면, 딥지노믹스에서 개발한 딥 러닝 알고리즘으로 spinal muscular atrophy, hereditary nonpolyposis colorectal cancer, autism 등에 관여하는

유전 변이를 성공적으로 점수를 매겨 가려낼 수 있다고 주장하고 있다. 이 방법은 기존에 GWAS, population 기반의 계산, QTL, functional annotation 등과 모두 독립적이어서, 상호보완적으로 사용할 수 있을 것이라고 자평하고 있다. 또한 이러한 유전 변이가 엑손뿐만 아니라 인트론에도 많이 존재한다는 것을 발견했기 때문에, 기존에 WES으로 나온 결과만을 본다면 인트론에 있는 중요 변이들을 모두 놓칠 것이라는 언급도 있다.

신약 개발은 임상시험 단계에서 대부분 실패하고 평균 약 3조 원 이상의 신약개발 비용이 들어가지만, 인공지능은 제약회사들의 실패 수를 대폭 줄이고 성공적인 약물을 더 빨리 찾을 수 있게 도움을 줄 수 있다. 이에 따라 최근 기계학습을 이용해 신약개발을 하는 회사들이 부쩍 늘어나고 있다. 딥지노믹스 외에도 대표적인 회사로 영국의 인공지능 회사인 베네볼런트 에이아이(BenevolentAI)와 구글의 알파벳의 칼리코(Calico) 등이 있다.

2 IBM 왓슨을 이용한 유전체 변이 해석 연구

IBM의 인공지능 왓슨은 암 환자의 진료를 보조하는 ‘왓슨 포 온콜로지’가 잘 알려져 있지만, 그 이외에도 임상 시험에서 환자를 매칭해주는 ‘왓슨 포 클리니컬 트라이얼 매칭’, 그리고 유전체 분석에 왓슨을 이용하는 ‘왓슨 포 지노믹스’도 있다.

2017년 Neurol Genet에는 ‘왓슨 포 지노믹스’를 유전체 분석에 접목한 소규모 연구가 소개되었다. Glioblastoma tumor specimen에서 얻은 WGS 데이터를 해석하여 치료에 사용할 수 있는 유전 변이(actionable mutation)를 찾는 것에 대하여 세 가지 방식을 비교하였다. 첫 번째는 인간 전문가 그룹이었다. 생물정보학 및 종양내과 전문의들로 구성된 그룹으로, WGS 및 RNA-seq을 해석하였다. 두 번째는 IBM 왓슨 포 지노믹스로, WGS 및 RNA-seq을 해석하였다. 세 번째는 파운데이션 메디슨의 상용 암 유전자 타겟 패널인 파운데이션 원(Foundation One)을 활용하였다. 그 결과 인간 전문가 그룹이 가장 많은 ‘치료에 사용할 수 있는 유전 변이’를 찾아내었으며, 왓슨 그룹은 상대적으로 더 적은 유전 변이를 찾아내었다.

하지만, VCF 파일에서 유전 변이를 찾아내는 데 걸린 시간은 인간 전문가 그룹에 비해서 왓슨 그룹이 훨씬 적게 걸렸다. 인간 전문가 그룹은 (한 사람이 분석하는 것으로 환산하여) 160시간이 걸렸던 반면, 왓슨 포 지노믹스는 단 10분 만에 분석을 완료하였다. 즉, 왓슨 포 지노믹스는 인간 전문가에 비해 놓치는 유전 변이들이 존재하지만, 그럼에도 불구하고 빠른 시간 내에 분석을 마치므로 의료 현장에서는 의학적인 가치를 지닐 수 있다는 결론이다.

표 4.1 인간 전문가와 왓슨의 암 유전체 분석 결과 비교

Gene	Variant	Identified variant			Identified associated drugs		
		NYGC	WGA	FO	NYGC	WGA	FO
CDKN2A	Deletion	Yes	Yes	Yes	Palbociclib, LY2835219 LEE001	Palbociclib, LY2835219	Clinical trial
CDKN2B	Deletion	Yes	Yes	Yes	Palbociclib, LY2835219 LEE002	Palbociclib, LY2835219	Clinical trial
EGFR	Gain (whole arm)	Yes	-	-	Cetuximab	-	-
ERG	Missense P114Q	Yes	Yes	-	RI-EIP	RI-EIP	-
FGFR3	Missense L49V	Yes	VUS	-	TK-1258	-	-
MET	Amplification	Yes	Yes	Yes	INC280	Crizotinib, cabozantinib	Crizotinib, cabozantinib
MET	Frame shife R755fs	Yes	-	-	INC280	-	-
MET	Exon skipping	Yes	-	-	INC280	-	-
NF1	Deletion	Yes	-	-	MEK162	-	-
NF1	Nonsense R461*	Yes	Yes	Yes	MEK162	MEK162, cobimetinib, trametinib, GDC-0994	Everolimus, temsirolimus, trametinib
PIK3R1	Insertion R562_M563insl	Yes	Yes	-	BKM120	BKM120, LY3023414	-
PTEN	Loss (whole arm)	Yes	-	-	Everolimus, AZD2014	-	-
STAG2	Frame shife R1012 fs	Yes	Yes	Yes	Veliparib, clinical trial	Olaparib	-
DNMT3A	Splice site 2083-1G>C	-	-	Yes	-	-	-
TERT	Promoter-146C>T	Yes	-	Yes	-	-	-
ABL2	Missense D716N	Germline	NA	VUS			
mTOR	Missense H1687R	Germline	NA	VUS			
NPM1	Missense E169D	Germline	NA	VUS			
NTRK1	Missense G18E	Germline	NA	VUS			
PTCH1	Missense P1250R	Germline	NA	VUS			
TSC1	Missense G1035S	Germline	NA	VUS			

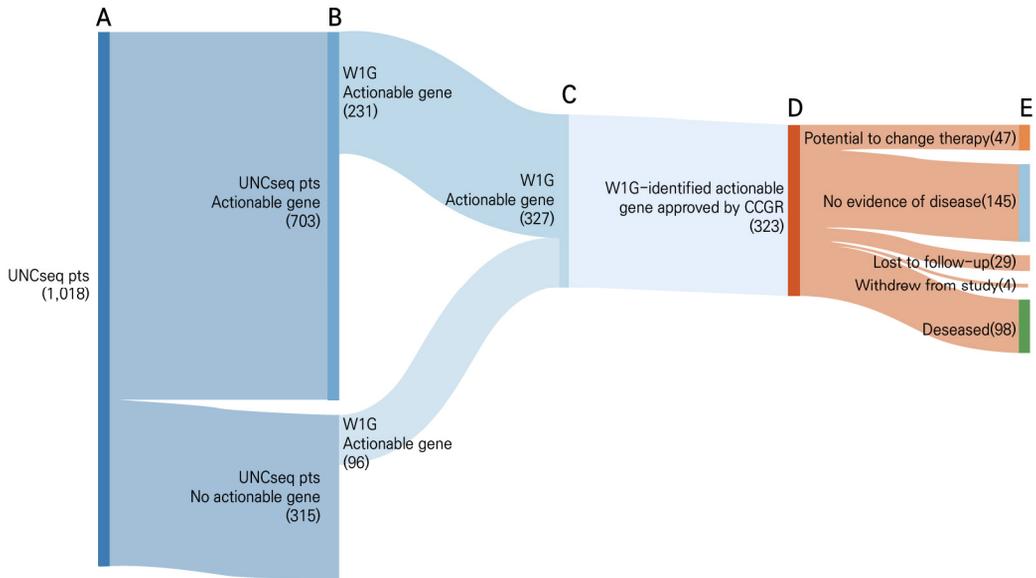
Abbreviations: FO = FoundationOne; NYGC = New York Genome Center; RNA-seq = RNA sequencing; WGA = Watson Genomic Analytics; WGS = whole-genome sequencing.

Genes, variant description, and, where appropriate, candidate clinically relevant drugs are listed. Variants identified by the FO as variants of uncertain significance (VUS) were identified by the NYGC as germline variants.

자료: Wrzeszczynski, K. O. *et al.*, 2017

또한 2018년 온콜로지스트(Oncologist)에는 왓슨 포 지노믹스가 기존의 수작업 파이프라인 및 Molecular Tumor Board에 비해서 더 많은 유전 변이를 찾아낸다는 논문이 발표되었다. 1,018명의 targeted exon sequencing 데이터로 기존에 UNCseq informatics pipeline 및 Molecular Tumor Board의 분석을 수행하였다. 이 데이터로 왓슨 포 지노믹스의 분석을 추가적으로 진행한 결과 총 323명(32%)의 환자에서 치료 가능한 유전 변이(actional genetic alteration)이 발견되었다. ‘치료 가능하다’고 새롭게 발견된 대부분의 유전 변이는 최근 새롭게 시작된 임상 시험 때문이었으며, 이러한 분석을 위해서는 한 케이스당 3분 이하의 시간밖에 소요되지 않았다.

그림 4.2 UNCSeq의 결과와 왓슨 포 지노믹스의 결과 비교



자료: Patel, N. M. *et al.*, 2018

3 개인 유전 정보 빅데이터를 이용한 신약 개발 활용 사례

실리콘밸리의 개인 유전 정보 분석 회사인 23andMe는 병원을 거치지 않고 개인 고객의 타액을 직접 우편으로 받아 건강, 질병 발병 가능성, 조상 분석 등에 대한 유전 정보를 분석해주는 서비스를 제공한다. 2019년 3월을 기준으로 23andMe의 고객 숫자는 1,000만 명을 돌파했다.

- 1) 사실 유전 정보를 분석한 고객 숫자로 따지면 23andMe보다 엔세스트리닷컴(Ancestry.com)이 더 많다. 엔세스트리닷컴은 조상 분석이나 친척 찾기를 목적으로 한 유전 정보 분석을 제공하는 회사이다.
- 2) 하지만 23andMe의 위력은 고객의 유전 정보, 즉 유전형(genotype)뿐만 아니라, 병력, 약물 부작용 등의 건강에 대한 특징, 즉 표현형(phenotype)까지 보유하고 있다는 점이다.

그림 4.3 23andMe 고객의 폭발적인 증가



23andMe는 고객이 타액을 보내고 유전 정보 분석 결과를 볼 때, 자신의 건강에 대한 특징을 인류를 위한 연구용으로 '기부'할 것을 요청받는다. 여기에는 과거에 어떤 질병을 앓은 적이 있다거나, 약에 대한 부작용이 있다거나, 카페인에 민감한지, 대머리인지 등의 일반적인 건강상태에 대한 질문도 있다. 이러한 데이터 기부 요청에 대해서, 자신의 데이터를 기꺼이 기부하는 사람이 80%에 달한다.

즉, 23andMe는 (추정해보면) 약 800만 명 내외의 유전형-표현형 데이터베이스를 보유하고 있다. 이는 전 세계 최대의 규모의 데이터베이스로 어느 회사, 대학, 연구기관보다도 크다. 더 나아가 이 정보는 비식별화되어 있기는 하지만, 회사에서는 개별 소비자들의 메일 주소까지 알고 있으므로, 해당 유전형을 가진 특정 고객군에게 (고객의 동의하에) 접근할 수 있을 것이다. 이러한 의미에서 23andMe는 데이터 회사라고 할 수도 있다.

23andMe의 대규모 유전형-표현형 데이터는 신약 개발과 임상 시험에 큰 도움을 줄 수 있다. 글로벌 제약사는 이미 2015년부터 이 데이터의 활용에 관심을 가지기 시작했다. 2015년 제네틱과 화이자는 이 데이터를 신약 개발 목적으로 구매하기도 했다. 제네틱의 경우 이 데이터를 무려 6,000만 달러에 구매했다. 23andMe가 상장기업이 아닌지라 정확한 비교는 어렵지만, 이렇게 제약사에 B2B로 데이터를 판매해서 벌어들인 수익이 B2C로 유전정보를 분석한 것 못지 않게 크다는 분석도 있었다.

더 나아가 23andMe는 이런 데이터를 기반으로 하여 자체적으로 신약을 개발하겠다는 계획을 발표했다. 개인 유전 정보 분석 회사에서 출발하여, 축적한 데이터를 바탕으로 새로운 방식의 제약사가 되겠다는 포부를 밝힌 것이다. 2015년 23andMe는 제네틱의 최고 과학 책임자이던 스타 과학자, 리차드 셸러(Richard Scheller) 등 제약사 출신 전문가들을 영입하여 자체적 신약 개발을 본격적으로 시작했다. 2019년 1월, 23andMe는 자체적으로 개발한 monoclonal antibody를 제약사에 라이선싱했다. 싸이토카인 IL-36를 저해하는 항체이며, 이를 스페인의 신약개발회사인 알미랄(Almirall)에 라이선싱-아웃했다. 적응증은 주로 건선, 루푸스나 다른 염증성 질환 등이 될 것으로 알려졌다.

2018년에는 23andMe는 GSK에게 향후 4년 동안 유전 정보 DB에 대한 독점적 접근권을 주고, 3억 달러의 투자를 받았다. 23andMe와 GSK는 유전형-표현형 DB를 크게 세 가지 측면에서 신약개발에 활용할 수 있을 것이라고 언급하고 있다. 신약 표적 발굴을 더 안전하고 효과적으로 할 수 있다. 유전 정보는 질병의 메커니즘이나 발병 경로에

대한 이해를 높여주기 때문이다. 유전 정보를 기반으로 한 표적 물질의 선택은 신약 개발의 성공률을 높여줌과 동시에, 리스크도 줄여줄 수 있다. 표적 치료제에 효능을 보일만한 특정 환자 집단을 파악하는데 도움을 준다. 더 나아가, 이러한 데이터를 바탕으로 특정 질병에 대한 새로운 환자군을 정의할 수도 있다. 임상 시험의 환자 모집에 활용할 수 있다. 23andMe는 특정 질병에 걸린 환자군 및 환자군의 유전형까지 파악하고 있기 때문에, 임상 시험에 참여할 수 있는 환자를 쉽고 빠르게 모집할 수 있다.

특히, GSK는 파킨슨 병의 신약 개발에 23andMe의 데이터를 활용할 계획이다. GSK는 파킨슨 병의 발병에 관여한다고 알려진 LRRK2 유전자의 변이에 관련돼 신약을 개발하는 것으로 알려져 있다. 그런데 현재 미국의 파킨슨병 환자 100만 명 중, LRRK2 변이를 가진 사람은 10,000명 정도로 추정된다. 즉, GSK에서는 이 신약 후보 물질의 임상시험을 하기 위해서는 파킨슨 환자 중에 LRRK2 변이를 가진 사람을 골라내어야 하는데, 100명의 파킨슨 환자의 유전 정보를 검사해야만 1명의 LRRK2 변이를 보유한 환자를 찾을 수 있는 것이다. 하지만 23andMe는 기존에 축적해둔 데이터베이스에서, 이미 임상 시험에 참여하기로 동의한 250여 명의 LRRK2 변이를 보유한 파킨슨 환자의 정보를 가지고 있다. 이러한 23andMe의 데이터가 없었다면 GSK는 최소한 파킨슨 환자 25,000명의 유전 정보를 검사해야 했을 것이다. GSK는 23andMe의 데이터 덕분에 신약 개발을 더욱 가속화 할 수 있게 되었다.

더 나아가, 23andMe는 2020년 4월, 유전형에 따라 COVID-19의 중증도가 달라지는지를 파악하기 위해서 연구를 시작한다고 밝혔다. 기존 감염질환의 경우 유전형에 따라 감수성(susceptibility)이 달라진다는 것은 알려져 있지만, 중증도가 유전형에 영향을 받는지는 아직 명확하지 않다. 지금까지 23andMe는 1,000만 명 이상의 유전형 데이터를 분석하였고, 그 중 80% 정도는 표현형 데이터까지 보유하고 있으므로, 기존 고객들 중에 본인의 동의 하에 COVID-19에 관련된 데이터를 얻을 수 있다면, 대규모의 연구를 그동안 23andMe가 항상 진행해왔던 방식으로 진행해볼 수 있다.

미국의 23andMe 고객 중, 수십만 명 규모가 참여할 것을 기대하고 있으며, 여기에는 COVID-19에 양성/음성을 받은 경우가 모두 포함될 것이다. 데이터의 취합은 온라인으로만 이뤄지고, 데이터는 longitudinal하게 얻어진다. 처음에 베이스라인 설문 조사를 한 번 하고, 매달 정기적인 설문 조사를 통해서 데이터를 얻게 된다. 이렇게 참가자 중에서 신규 확진자가 나올 경우 그 전후 데이터를 얻을 수 있으며, 전체 과정은 IRB의 감독을 받는다.

23andMe는 이런 연구를 통해서 코로나 바이러스의 감염 메커니즘에 대해서 더 이해할 수 있을 것이라고 밝혔다.

4 유전체 및 진료기록, 라이프로그 등을 통합한 빅데이터 연구

일반적인 의학 연구는 특정한 가설이나 주제를 가지고 시작한다. 어떤 원리를 밝히거나 질병 치료에 관한 발견을 하려고 한다면, 구체적인 가설에서 시작하여 그 가설의 검증에 필요한 데이터를 얻는다. 이렇게 꼭 필요한 사람으로부터, 꼭 필요한 데이터만 모으는 것이 지금까지의 일반적인 연구의 방법이다. 이를 ‘탑-다운(top-down)’ 방식의 연구라고 부를 수 있을 것 같다.

하지만 최근에는 유전체 등의 대규모 데이터를 바탕으로, 이와 정반대의 연구가 등장하고 있다. 즉, 가설을 먼저 세우고 필요한 데이터만 모으는 것이 아니라, 일단 ‘모든 데이터’를 먼저 모으고 거기에서 의미를 찾으려고 하는 것이다. 많은 사람에 대해서, 측정할 수 있는 모든 데이터를 일단 모아놓은 이후에, 방대한 데이터에서 건강관리와 질병 치료, 예방에 대한 가치 있는 ‘무엇인가’를 얻으려고 하는 것이다. 이는 기존 연구와는 반대 방향인, ‘바텀-업(bottom-up)’ 방식의 연구라고 할 수 있겠다.

이 데이터를 통해 유의미한 결론을 내리기 위해서는, 다양한, 양질의 데이터를 가능한 많은 사람에게서, 최대한 자주, 최대한 오랜 기간 측정해야 한다. 그리고 이러한 데이터를 저장할 공간과 분석할 방법도 필요하다. 여기에는 여러 기술적인 장벽뿐만이 아니라, 막대한 비용과 자원, 시간, 저장 공간이 소요된다.

하지만 디지털 기술의 발전에 따라, 데이터의 종류와 양과 질의 폭발적인 개선, 데이터를 측정할 수 있는 각종 기술의 발전, 이러한 데이터를 모두 통합할 수 있는 플랫폼, 통신 기술, 클라우드 컴퓨팅 등의 발전은 결국 이렇게 ‘모든 데이터를 일단 모아놓고 보는’ 방식의 연구를 가능하게 한다. 더 나아가 이러한 막대한 데이터를 분석할 수 있는 인공지능 기술도 발전했다.

이렇게 방대한 숫자의 참여자들을 대상으로 유전체 데이터뿐만 아니라 진료 기록, 라이프로그 등을 시계열로 측정하여 새로운 의학적 인사이트를 얻으려는 연구를 몇 가지 소개하려 한다.

[미국 정부의 All-of-Us 프로젝트]

미국 정부는 정밀 의료를 구현하기 위해 2015년부터 ‘정밀 의료 이니셔티브(Precision Medicine Initiative)’를 진행하고 있다. 그 일환으로 미국 국립 보건원(NIH)에서 진행하고 있는 대규모 프로젝트 중의 하나가 바로 ‘All-of-Us’라고 하는, 말 그대로 ‘모든 사람의, 모든 데이터’를 모으겠다는 도전적인 프로젝트이다.

정밀 의료는 각 개인의 유전 정보, 환경, 생활 습관 등의 차이를 종합적으로 고려하여 질병의 치료 및 예방에 적용하겠다는 것이다. 과거의 의료가 ‘평균적인 사람’을 대상으로 개발된 치료를 모든 환자에게 적용하는, 소위 ‘하나의 크기가 모든 사람에게 맞는다(one-size-fits-all)’는 방식과 비교하면 완전히 새로운 패러다임이라고 할 수 있다. 그런데 이렇게 각 개인의 유전 정보, 환경, 생활 습관 등의 차이를 고려해서 새로운 의료를 제공하기 위해서는 일단 이러한 데이터 사이의 관계를 파악하고, 환자의 유형을 더 세부적으로 분류하며, 그러한 유형에 맞는 치료가 무엇인지를 새롭게 연구해야 한다.

이를 위해서 진행하는 것이 바로 ‘정밀 의료 이니셔티브 코호트 프로그램’이라고도 부르는, ‘All-of-Us’ 프로젝트이다. 이 프로젝트는 2018년 5월에 미국 전체를 대상으로 출범했는데, 적어도 100만 명 이상의 미국인을 자발적으로 모아서, 진료 기록, 가족력, 유전 정보, 혈액 및 소변 검사 결과, MRI 등의 영상 의료 데이터, 그리고 핏빗 등 웨어러블 디바이스를 통한 데이터를 모두 수집하고 있다. 여기에 더해, 향후에는 보험사의 청구 데이터(claims data)나, 날씨와 대기질과 같은 환경 데이터, 그리고 트위터 피드와 같은 SNS 데이터도 수집할 것을 고려하고 있다.

표 4.2 All-of-Us에서 수집하는 데이터의 종류

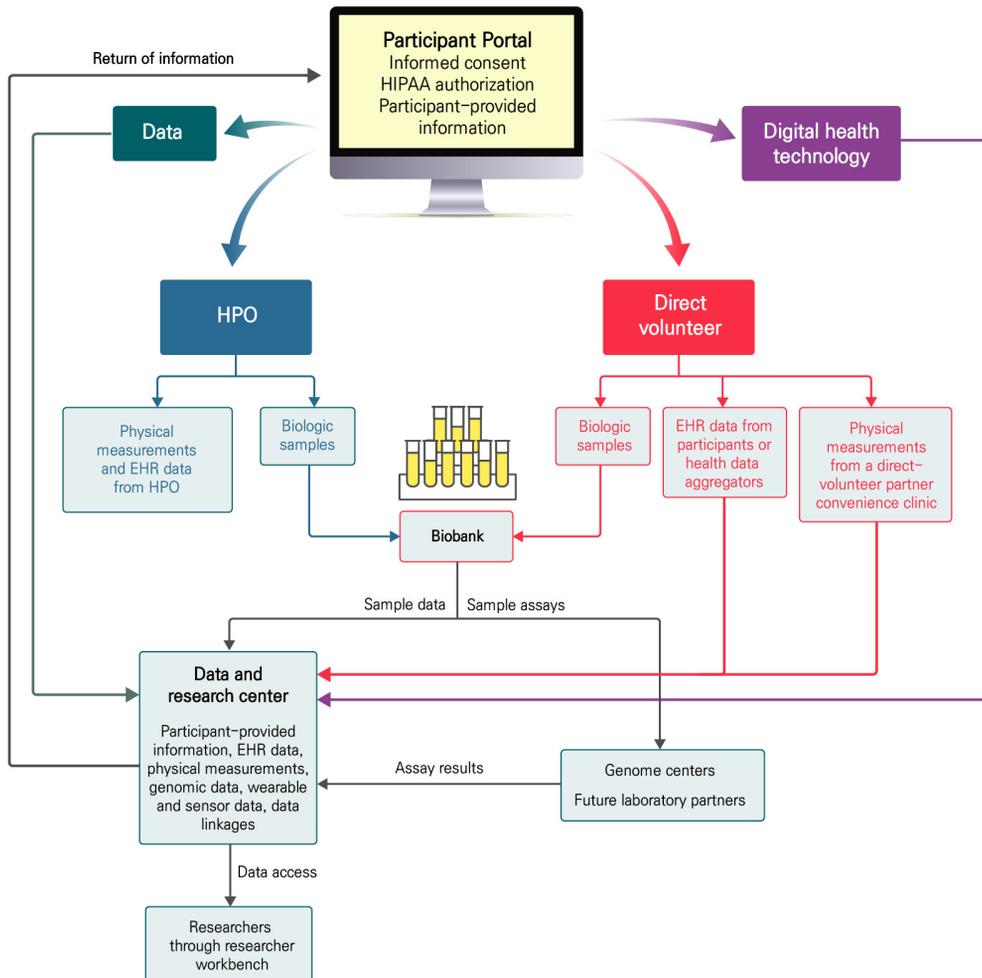
Data Source	Details
Current sources	
Health surveys	Initial surveys include information on sociodemographic characteristics, overall health, lifestyle, and substance use, with subsequent modules covering personal and family medical history and access to health care.
Physical measurements	Per-protocol measurements include blood pressure, heart rate, weight, height, body-mass index, and hip and waist circumferences.
Biospecimens †	Blood and urine samples are tested for DNA, RNA, cell-free DNA, serum, and plasma. If blood specimens cannot be obtained, saliva specimens are obtained.
Electronic health records	Initial capture of structured data includes billing codes, medication history, laboratory results, vital signs, and encounter records from health care provider organizations. Records will be expanded to include narrative documents. Pilot studies are testing data collection through Sync for Science and other health data aggregators.
Digital health information	Data can be captured from compatible participant-owned devices such as Fitbit. Pilot studies of other devices and linkage to health apps are being explored.
Future sources	
Health surveys	Additional modules, including surveys regarding social behavioral determinants of health, are under development.
Bioassays	Pilot studies for genotyping and whole-genome sequencing are expected to begin by early 2020. Additional pilot studies of bioassays are planned.
Health care claims data	Systems for the use of claims data, including billing codes and medication data, are under development.
Geospatial and environmental data	These data include geospatial linkage to measures such as weather, air quality, pollutant levels, and census data. Assays and sensor-based measurements of exposure are under consideration.
Other sources	Voluntary contributions of data from social networks (e.g., Twitter feeds) and additional biospecimen collections are under consideration.

자료: Comstock, J., 2015

이렇게 100만 명이나 되는 사람들로 구성된 연구는 인류 역사상 최대 규모라고 해도 과언이 아닐 것이며, 수집하는 데이터의 종류도 전례 없이 다양하다. 미국 정부는 이러한 연구를 위해서 총 2억 1,500만 달러(한화 약 2,600억 원)의 막대한 예산을 투입하면서 의료의 패러다임을 전환하기 위해 노력하고 있다.

이러한 연구 역시 특정 질병을 대상으로 구체적인 가설을 먼저 세우고, 이를 검증하기 위해서 시작한다기보다는, 다양한 사람의 다차원적인 데이터를 수집, 분석함으로써 기존의 ‘소규모’ 연구에서 파악하기 어려웠던 새로운 의학적인 통찰을 얻기를 기대한다고 볼 수 있다.

그림 4.4 All-of-Us 프로젝트의 전반적 구조



Genome centers in the All of Us program generate genomic data from biosamples. The researcher workbench is the platform for data analysis in the program. EHR denotes electronic health record, HIPAA Health Insurance Portability and Accountability Act, and HPO health care provider organization.

자료: Comstock, J., 2015

특히 이렇게 대규모 참여자와 대량의 데이터를 수집하는 것은 연구의 편파성(bias)을 없애기 위해서 매우 중요하다. 기존에도 구축된 환자 코호트가 없었던 것은 결코 아니지만, 코호트에 포함되는 사람의 숫자가 제한적이거나, 다양성이 충분히 확보되지 않았던 경우가 대부분이었기 때문이다.

예를 들어, 인종, 성별, 나이, 지역, 수입 및 경제력, 의료기관 접근성, 교육 수준에 따라 이 연구에 포함되는 구성원에 편중 현상이 있는 경우가 많았다. 이러한 데이터를 기반으로 도출된 연구 결과 역시 이러한 데이터 편중이 반영되게 된다. 이 때문에 All-of-Us 프로그램에서는 인상적이게도, 기존의 의학 연구에서 상대적으로 소외당했던 소수 인종 등의 참가자들을 특정 비중 이상 포함시키는 것을 목표로 하고 있다.

이렇게 진행되고 있는 All-of-Us 프로그램에 대한 최신 업데이트가 2019년 8월 뉴 잉글랜드 저널 오브 메디슨(NEJM)에 발표되었다. 이 프로젝트는 2017년 5월부터 시범사업 단계를 거쳐, 2018년 5월에 미국 전역으로 확대하여 정식으로 출범했다. 2019년 7월 기준으로 미국 전역의 무려 340군데 이상의 병원 등에서 참가자를 모집하고 있다. 출범 1년 정도 만에 이 프로그램에 참가하여 자신의 데이터를 제공한 사람은 23만여 명에 달하여, 100만 명을 달성하겠다는 목표의 1/5 정도를 이미 달성하였다. 특히 이 중에 생체 시료 및 진료 기록 등의 핵심 데이터를 제공하기로 한 ‘핵심 참가자(core participants)’는 175,000명이었다. 연구자들에 따르면, 매주 3,100명의 ‘핵심 참가자’가 리크루팅되고 있는 추세를 보면 100만 명의 ‘핵심 참가자’를 모으는 목표는 2024년에 달성할 수 있을 것으로 예측하고 있다.

All-of-Us 프로그램에서는 이렇게 구성된 코호트를 기반으로 2027년과 그 이후까지 질병 바이오마커의 발견, 새로운 질병 분류법의 개발, 인공지능 개발 등의 세부적인 목표까지도 단계별로 세워놓고 있다. ‘All-of-Us’ 프로젝트의 책임자인 에릭 디쉬먼(Eric Dishman)은 “이러한 다양한 데이터의 통합은 생활 습관과 환경이 건강에 미치는 영향에 대한 새로운 이해를 가능하게 하며, 이는 궁극적으로 사람들이 아주 정밀하고, 개별적인 방식으로 건강을 유지할 수 있게 해줄 것이다”고 밝히고 있다.

표 4.3 All-of-Us의 세부적인 목표 및 타임라인

Goal	Years				
	End of 2018 (N=94,000)	End of 2019 (N=>200,000)	2020-2022 (N=<650,000)	2023-2027 (N=>1 million)	After 2027 (N=>1 million)
Return data to participants	+	+	+++	++-	++++
Establish demonstration projects †		+	+++	+	+
Discover genetic and environmental correlates with disease			--	++-	++++
Improve predictions of therapeutic safety and efficacy					

Goal	Years				
	End of 2018 (N=94,000)	End of 2019 (N=>200,000)	2020-2022 (N=<650,000)	2023-2027 (N=>1 million)	After 2027 (N=>1 million)
Discover disease biomarkers			--	++-	+++
Connect mobile health, digital health, and sensor data with clinical outcomes			--	++-	+++
Develop new disease classifications			+	++-	++++
Support clinical trials			+	++-	+++
Enable machine-learning applications			--	++-	++++
Improve understanding of health disparities			--	++-	+++
Develop and test new therapeutic agents					++

자료: Monegain, B., 2018

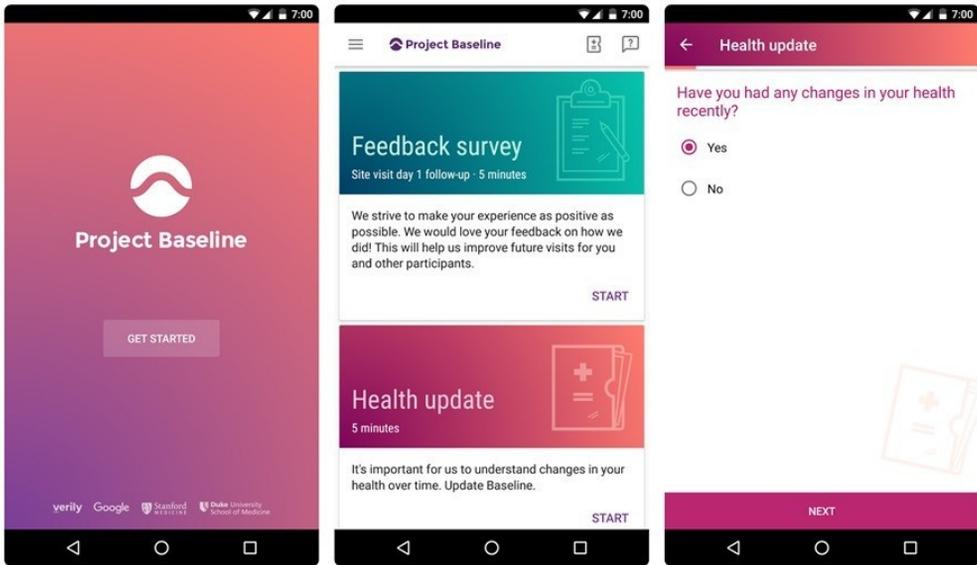
[구글 베이스라인 프로젝트]

구글도 유사한 유형의 연구를 진행하고 있다. 구글은 여러 문샷 프로젝트를 진행하고 있는데, 그 중의 하나가 바로 인간의 건강이라는 것을 제대로 정의하겠다는 베이스라인 프로젝트(baseline project)이다.

대부분의 의학 연구는 질병에 걸려 있는 것과 같은 ‘정상이 아닌’ 상태를 대상으로 한다. 그런데 ‘정상이 아닌’ 상태를 파악하기 위해서는 먼저 ‘무엇이 정상인지’를 정의해야만 한다. 하지만 지금까지 이러한 ‘정상’, 혹은 ‘완전히 건강한 상태’의 인간에 대해서는 연구가 충분히 이뤄지지 않았다는 것이 구글의 주장이다. 그러므로 모든 의학 연구는 이러한 건강한 상태라는 ‘기준’, 즉 베이스라인을 재정의하는 것에서 시작해야 한다는 것이다.

이를 위해서 구글(알파벳)의 생명과학 분야 자회사인 베릴리(Verily)에서는 ‘베이스라인 프로젝트’를 진행해오고 있다. 이 연구는 2014년 출범 당시에는 175명 정도로 규모가 그리 크지 않았으나, 2017년부터 규모를 대폭 확대했다. 이 프로젝트는 4년 동안 10,000명에 달하는 개인의 건강 상태를 면밀하게 추적하여 데이터를 축적하는 것을 골자로 한다. 수집하는 데이터는 두 가지 종류의 기기를 통한 심박수, 수면 패턴 및 유전체 데이터, 감정 상태, 진료 기록, 가족력, 정기적인 소변, 타액, 혈액 검사 등의 다양한 데이터를 포괄한다. 개별적인 검사는 스탠퍼드 대학병원과 듀크 대학병원에서 진행한다.

그림 4.5 구글 베이스라인 프로젝트 앱



구글에서는 이러한 베이스라인 스타디를 진행하기 위해서, 데이터 측정을 위한 스마트 워치인, ‘스타디 워치(Study Watch)’를 2017년 공개했다. 이 시계는 심전도, 심박, 피부 전기 활성도, 관성 움직임(inertial movement) 등의 데이터를 측정한다고 한다. 특히, 이 기기에는 장기간의 데이터 측정을 위한 세심한 배려도 들어 있다. 저전력 디스플레이를 통해서 일주일에 한 번만 충전하면 되고, 데이터 저장 공간 및 데이터 압축 기능을 통해서 연구 참여자들이 일주일에 한 번만 데이터를 동기화하면 된다. 또한 침대 매트리스 아래에 까는 형식의 센서를 통해서 수면 모니터링도 한다.

구글이 이렇게 다양한 환자에 대한, 다양한 종류의 데이터를, 장기간 측정하는 것에서 과연 무엇을 얻을 수 있을지는 아직 확실하지 않다. 다만, 이 프로젝트를 이끄는 제시카 메가 박사는 베이스라인 프로젝트를 ‘건강에 대한 구글 지도를 만드는 것’에 비유했다. 우리가 현재 어느 위치에 있으며, 주위에 지형지물이 어떠한지, 어느 도로에 가장 차가 많이 막히고, 목적지에 가기 위해서 어디가 최단 경로인지 알기 위해서는 일단 지도가 필요하다. 이처럼 만약 다양한 건강 상태, 유전형, 생활 습관을 지닌 10,000명의 다양한 데이터를 수년 동안 추적 관찰하게 되면, 우리의 건강에 대한 ‘지도’를 가질 수도 있을 것이라는 점이다.

이러한 지도를 통해서 우리는 건강한 상태라는 것이 무엇인지, 혹은 질병이 발병하기까지 건강한 상태에서부터 오랜 기간 동안 어떠한 단계를 거치는지에 대해서 파악할 수 있다. 만약

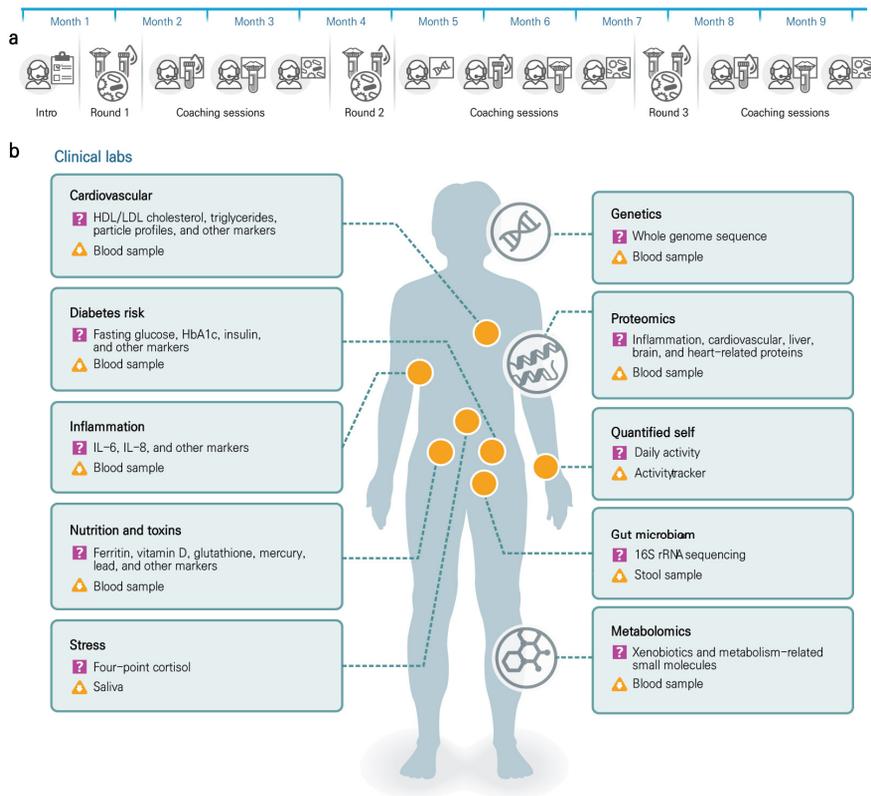
이를 이해하게 되면 개개인의 환자에 대해서 건강을 유지하고, 질병을 예방, 예측, 치료하기 위한 중요한 통찰을 얻을 가능성이 있다는 것이 구글의 생각이다.

[100만 명 웰니스 프로젝트]

앞서 설명한 구글의 베이스라인 프로젝트와 미국 정부의 'All-of-Us' 프로젝트는 한창 진행 중으로 아직 구체적인 결과가 발표되지는 않았으며, 실질적인 성과가 나올지를 판단하기 위해서는 시일이 더 소요될 것이다.

하지만 이와 비슷한 방향성의 연구에 대한 결과가 2017년 네이처 바이오테크놀러지에 발표된 바가 있다. 바로 '10만 명 웰니스 프로젝트'라고 하는 연구의 초기 결과에 대한 논문이다.

그림 4.6 수천 가지의 데이터를 9개월 동안, 3번에 걸쳐 측정한 연구



자료: Price, N. D. *et al.*, 2017

이 연구에서는 총 9개월 동안, 모든 참여자에 대해서, 활동량, 심혈관, 당뇨, 염증, 영양, 스트레스, 유전체, 단백질, 미생물체 등 측정할 수 있는 모든 데이터를 측정한다. 이렇게 측정한 데이터의 종류는 수천 가지에 달한다. 활동량, 수면, 심박 등의 데이터는 핏빗으로 9개월 동안 매일 측정하였으며, 단백질, 대사체, 미생물체 및 혈액 검사는 9개월의 연구 기간 동안, 3개월 간격으로 총 3번씩 측정했다.

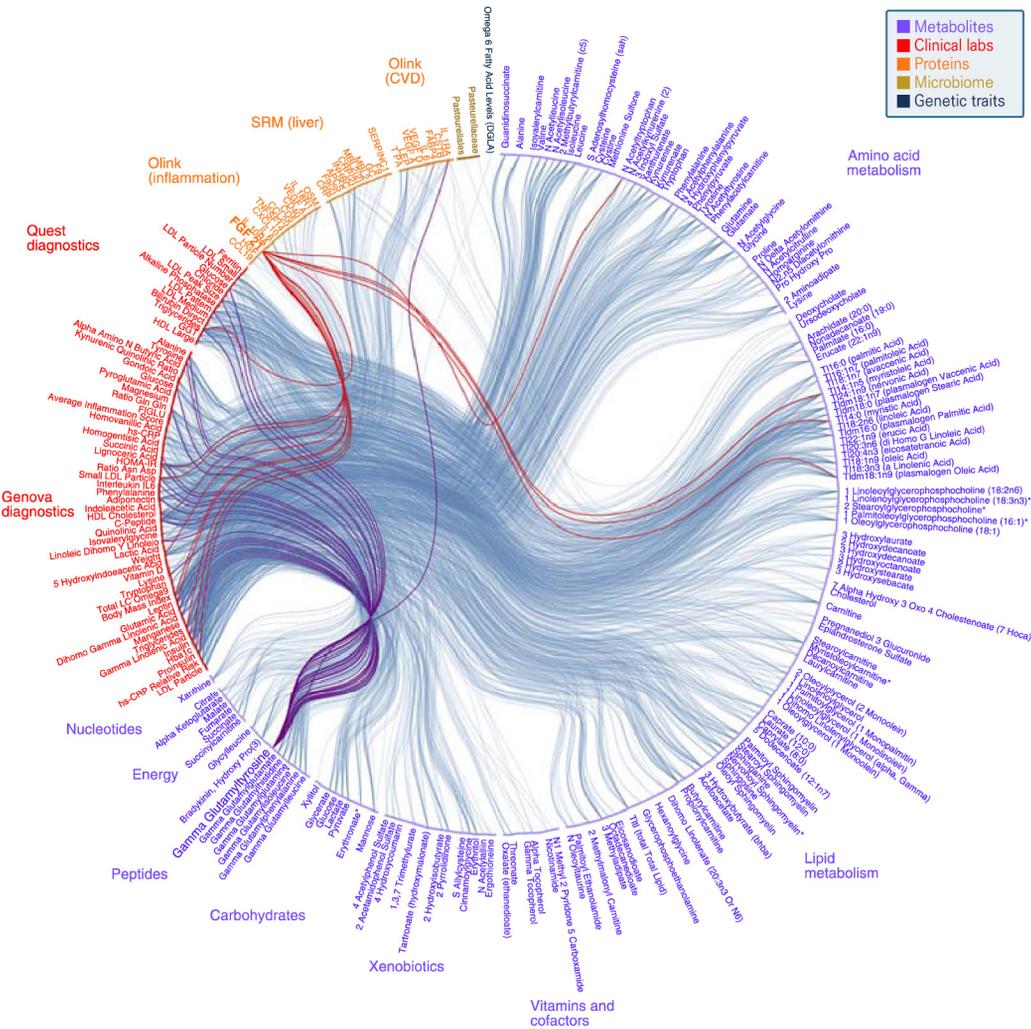
연구자들은 이렇게 모은 데이터를 ‘개인의, 밀도 높은, 다이나믹 데이터 클라우드 (personal, dense, dynamic data cloud)’라고 이름 붙였다. 개별 참여자의 측정 가능한 모든 데이터를 장기간, 밀도 높게 측정하게 되면, 그 사람의 건강 변화와 질병의 발병 징후 등을 이 통합된 데이터 내에서 파악할 수 있다는 이야기이다.

최종적으로 10만 명 분석을 목표로 하는 연구에, 이 논문에서는 초창기 참여한 108명만을 분석한 결과만을 발표한 것이다. 사실 지금까지 논문으로 결과가 보고된 연구 중에는 가장 다양한 종류의 데이터를, 장기간, 여러 사람에 대해서 측정한 것이라고 볼 수 있다.

논문에서는 단순히 데이터의 측정에 그치지 않고, 이 데이터를 다양하게 분석하여 건강 관리에 대한 새로운 통찰을 찾아내려고 하였다. 더 나아가, 개별 참가자들에게 맞춤형 개인 코칭을 제공함으로써 실제로 건강이 개선되는지를 살펴보았다.

먼저 연구자들은 108명에 대해서 9개월 동안 측정한 수천 가지의 데이터들 사이에 서로 상관관계가 높은 것이 있는지를 분석했다. 그 결과 3,470 가지의 유의미한 상관관계와, 2,406 가지의 유의미한 데이터 변화를 발견했다. 상관관계라는 것은 두 종류의 데이터가 같은 추세로 변화한다는 것으로, 두 데이터가 서로 직간접적으로 연결되어 있을 가능성을 고려해볼 수 있다. 이 중에는 이미 의학적으로 잘 알려진 관계도 있었던 반면, 완전히 새롭게 발견한 관계도 있었다. 다만, 이 수치 사이에 ‘왜’ 관계가 높은지, 두 값이 원인과 결과의 관계가 있는 것인지, 혹은 그저 우연히 비슷하게 바뀌었는지를 증명하는 것은 향후 더 많은 연구를 통해 밝혀져야 할 것이다.

그림 4.7 유전체를 포함하여 방대한 데이터를 분석하여 발견한 상관관계



자료: Price, N. D. et al., 2017

이러한 데이터를 측정하는 것이 실제로 참가자의 건강을 개선할 수 있는지를 보기 위해서, 세 번의 측정 결과를 바탕으로 개인 코치가 개별 참가자에게 맞춤형 조언을 제공했다. 참가자들은 유전 정보, 임상 검사 결과 등에 따라서 식습관 개선, 운동 필요, 스트레스 관리, 영양 관리 필요, 의사 진료 권고 등의 몇 가지 그룹으로 나누어, 그에 맞는 코치를 받았다.

특히 특정 수치가 정상 범위를 벗어난 것이 발견되면, 코치가 이 수치를 정상으로 돌릴 수 있도록 의학적으로 증명되어 있는 생활 습관 변화나 교육 프로그램 이수를 권했다.

예를 들어, 공복 혈당 수치가 상승한 경우라면 당뇨 전단계라는 의미이므로, 당뇨 예방 프로그램(Diabetes Prevention Program)에 근거한 조언을 제공하였다. 그 결과 비타민D, 당화혈색소 등의 수치에 큰 개선 효과가 있었으며, 콜레스테롤 수치, 당뇨 위험도, 염증 수치 등에도 지속적인 개선이 있었다.

표 4.4 코치의 지도를 받으면서 유의미하게 개선된 수치들

Clinical laboratory test		Changes in labs in participants out-of-range at baseline		
Health area	Name	N	Δ per round	P-value
Nutrition	Vitamin D	95	+7.2 ng/mL/round	7.1×10^{-25}
Nutrition	Mercury	81	-0.002 mcg/g/round	8.9×10^{-9}
Diabetes	HbA1c	53	-0.085%/round	9.2×10^{-6}
Cardiovascular	LDL particle number (Quest)	30	+130 nmol/L/round	9.3×10^{-5}
Nutrition	Methylmalonic acid (Genova)	3	-0.49 mmol/mol creatinine/round	2.1×10^{-4}
Cardiovascular	LDL pattern (A or B)	28	-0.16 /round	4.8×10^{-4}
Inflammation	Interleukin-8	10	-6.1 pg/mL/round	5.9×10^{-4}
Cardiovascular	Total cholesterol (Quest)	48	-6.4 mg/dL/round	7.2×10^{-4}
Cardiovascular	LDL cholesterol	57	-4.8 mg/dL/round	8.8×10^{-4}
Cardiovascular	LDL particle number (Genova)	70	-69 nmol/L/round	1.2×10^{-3}
Cardiovascular	Small LDL particle number (Genova)	73	-56 nmol/L/round	3.5×10^{-3}
Diabetes	Fasting glucose (Quest)	45	-1.9 mg/dL/round	8.2×10^{-3}
Cardiovascular	Total cholesterol (Genova)	43	-5.4 mg/dL/round	1.2×10^{-2}
Diabetes	Insulin	16	-2.3 IU/mL/round	1.5×10^{-2}
Inflammation	TNF-alpha	4	-6.6 pg/mL/round	1.8×10^{-2}
Diabetes	HOMA-IR	19	-0.56 /round	2.0×10^{-2}
Cardiovascular	HDL cholesterol	5	+4.5 mg/dL/round	2.2×10^{-2}
Nutrition	Methylmalonic acid (Quest)	7	-42 nmol/L/round	5.2×10^{-2}
Cardiovascular	Triglycerides (Genova)	14	-18 mg/dL/round	1.4×10^{-1}
Diabetes	Fasting glucose (Genova)	47	-0.98 mg/dL/round	1.5×10^{-1}
Nutrition	Arachidonic acid	35	+0.24 wt%/round	1.9×10^{-1}
Inflammation	hs-CRP	51	-0.47 mcg/mL/round	2.1×10^{-1}
Cardiovascular	Triglycerides (Quest)	17	-14 mg/dL/round	2.4×10^{-1}
Nutrition	Glutathione	6	+11 micromol/L/round	2.5×10^{-1}
Nutrition	Zinc	4	-0.82 mcg/g/round	3.0×10^{-1}
Nutrition	Ferritin	10	-1.4 ng/mL/round	3.1×10^{-1}

Clinical laboratory test		Changes in labs in participants out-of-range at baseline		
Health area	Name	N	Δ per round	P-value
Inflammation	Interleukin-6	4	-1.1 pg/mL/round	3.8×10^{-1}
Cardiovascular	HDL large particle number	8	+210 nmol/L/round	4.9×10^{-1}
Nutrition	Copper	10	+0.006 mcg/g/round	6.0×10^{-1}
Nutrition	Selenium	6	+0.035 mcg/g/round	6.2×10^{-1}
Cardiovascular	Medium LDL particle number	20	+2.8 nmol/L/round	8.5×10^{-1}
Cardiovascular	Small LDL particle number (Quest)	14	-2.3 nmol/L/round	8.8×10^{-1}
Nutrition	Manganese	0	N/A	N/A
Nutrition	EPA	0	N/A	N/A
Nutrition	DHA	0	N/A	N/A

자료: Price, N. D. *et al.*, 2017

이렇게 방대한 종류의 데이터를 분석한 연구 중에, 논문으로 출판된 것을 기준으로 108명 규모는 지금까지 가장 크다고 할 수 있지만, 그래도 절대적으로 본다면 그리 크다고 할 수는 없다. 기간도 9개월이면 개인의 건강 개선 효과를 관찰하기에 그리 긴 기간이라고 하기는 어렵다. 연구자들은 이 연구를 2020년까지 10만 명에 대해서 진행하는 것을 목표로 후속 연구를 진행하고 있다.

특히, 이러한 규모의 연구에서는 데이터를 일일이 사람이 분석하기가 매우 어렵다. 데이터의 종류만 수천 가지가 되고, 참여자의 수도 10만 명으로 늘어나고, 기간도 길어진다면 결국 인공지능 등 자동화된 분석의 힘을 빌려야 할 것이다. 더 나아가 리로이 후드 박사와 네이튼 프라이스 박사는 이 연구 결과를 기반으로, 시애틀 기반의 애리베일(Arivale)이라는 회사를 창업했다. 애리베일에서는 일반 고객들을 대상으로 이 논문에서 나온 것과 동일한 방식의 데이터 측정 및 분석을 통한 건강관리 서비스를 제공하는 것을 목표로 하고 있다.

참고문헌

- 김태형(2017). “딥지노믹스 통해 본 ‘유전체+AI의 미래’”. 바이오스펙테이터.
(http://www.biospectator.com/view/news_view.php?varAtcId=3694).
- 문상훈·김영진·김봉조(2016). “미국 정밀의료 프로젝트 소개”. 질병관리본부 주간건강과질병.
(2016.01.21.).
- 최윤섭(2014). “구글의 새로운 X 프로젝트: 인간 신체의 비밀을 밝히겠다!”. 최윤섭의 Healthcare Innovation. (2014.08.05.).
- 최윤섭(2017). “프로젝트 베이스라인: 미래 의료를 향한 구글의 야심”. 최윤섭의 Healthcare Innovation. (2017.07.26.).
- Chen, C.(2015). “23andMe Turns Spit Into Dollars in Deal With Pfizer”. Bloomberg.
(2015.01.13.).
- Clark, K.(2018). “Scoop: 23andMe is raising up to \$300M”. PitchBook. (2018.07.24.).
- Comstock, J.(2015). “NIH seeks feedback on how to collect clinical data via smartphones, wearables”. Mobihealthnews. (2015.07.07.).
- Garde, D.(2015). “23andMe dives into biotech with Genentech R&D star Richard Scheller on board”. FierceBiotech. (2015.03.12.).
- Genetic Home Reference(2020). “What is the Precision Medicine Initiative?”.
precision medicine, p.3. (2020.08.17.).
- GSK(2018). “GSK and 23andMe sign agreement to leverage genetic insights for the development of novel medicines”. (2018.07.25.).
- Hale, C.(2019). “Fitbit launches bring-your-own-device program for the NIH’s All of Us megastudy”. FierceBiotech. (2019.01.16.).
- Herper, M.(2015). “Surprise! With \$60 Million Genentech Deal, 23andMe Has A Business Plan”. Forbes. (2015.01.06.).

- Mack, H.(2017). “Verily introduces health-tracking Study Watch for use in clinical research”. *Mobihealthnews*. (2017.04.17.).
- Marr, A.(2014). “Google's New Moonshot Project: the Human Body”. *The Wall Street Journal* (2014.07.27.).
- Maxmen, A.(2017). “Google spin-off deploys wearable electronics for huge health study”. *Nature*, 547(7661), pp.13-14. (DOI: 10.1038/547013a).
- Monegain, B.(2018). “23andMe lands \$300M investment from GSK”. *Mobihealthnews*. (2018.07.25.).
- NIH(2018). “All of Us Research Program Background”. NIH(National Institutes of Health) News & Events. (2018.05.01.).
- Patel, N. M. *et al.*(2018). “Enhancing Next-Generation Sequencing-Guided Cancer Care Through Cognitive Computing”. *Oncologist*, 23(2), pp.179-185. (DOI: 10.1634/theoncologist.2017-0170).
- Petrone, J.(2017). “23andMe wades further into drug discovery”. *Nature Bio Technology*, 35(10), p.897. (DOI: 10.1038/nbt1017-897).
- Price, N. D. *et al.*(2017). “A wellness study of 108 individuals using personal, dense, dynamic data clouds”. *Nature Biotechnology*, 35(8), pp.747-756. (DOI: 10.1038/nbt.3870).
- Stanford Medicine(2018). “The Project Baseline Study”. Stanford Medicine Department of Medicine.
- The All of Us Research Program Investigators(2019). “The “All of Us” Research Program”. *The New England Journal of Medicine*. (DOI: DOI: 10.1056/NEJMr1809937).
- TheDNAGeek(2019). “23andMe Has More Than 10 Million Customers”. (2019.04.08.).
- Winslow, R.(2015). “23andMe to Mine Genetic Database for Drug Discovery”. *The Wall Street Journal*. (2015.03.12.).

- Wrzeszczynski, K. O. *et al.*(2017). “Comparing sequencing assays and human-machine analyses in actionable genomics for glioblastoma”. *Neurology Genetics*, 3(4). (DOI: 10.1212/NXG.000000000000164).
- Xiong, H. Y. *et al.*(2016). “The human splicing code reveals new insights into the genetic determinants of disease”. *Science*, 347(6218). (DOI: 10.1126/science.1254806).
- Zhang, S.(2015). “Of Course 23andMe's Plan Has Been to Sell Your Genetic Data All Along”. Gizmodo. (2015.06.01.).
- 23andMe. “23andMe Research Innovation Collaborations Program”. (<https://research.23andme.com/research-innovation-collaborations/>).
- 23andMe(2020). “23andMe Customers Participate in COVID-19 Genetic Study”. 23andMe Blog. (2020.04.06.).

V. 개정 개인정보 보호법의 가명정보 제도와 유전체 연구





개정 개인정보 보호법의 가명정보 제도와 유전체 연구

1 개인정보 보호법에 도입된 “가명정보”란?

지난 2020년 2월 4일 『개인정보 보호법』의 개정법률이 공포되었고 2020년 8월 5일부터 시행될 예정이다. 『개인정보 보호법』은 같은 날에 개정 공포된 『정보통신망 이용촉진 및 정보보호 등에 관한 법률』 그리고 『신용정보의 이용 및 보호에 관한 법률』과 함께 소위 “데이터 3법”으로 불리면서 우리나라의 개인정보 보호를 위한 규제의 골격을 이루고 있다. 개인정보 보호법의 개정 내용 가운데 새로 도입된 “가명정보”는 특히 유전체정보의 분석과 이용에 있어 상당한 영향을 미칠 수 있다.

개정 개인정보 보호법상 가명정보의 정의에 관한 조문은 다음과 같다.¹⁾

개인정보 보호법 제2조

1. “개인정보”란 살아 있는 개인에 관한 정보로서 다음 각 목의 어느 하나에 해당하는 정보를 말한다.
 - 가. 성명, 주민등록번호 및 영상 등을 통하여 개인을 알아볼 수 있는 정보
 - 나. 해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보. 이 경우 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.
 - 다. 가목 또는 나목을 제1호의2에 따라 가명처리함으로써 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보(이하 “가명정보”라 한다)
- 1의2. “가명처리”란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것을 말한다.

1) 개정 개인정보 보호법뿐만 아니라 같은 날 공포된 『신용정보의 이용 및 보호에 관한 법률』 또한 “가명정보”의 개념을 새로이 도입하고 있다. 이 글은 유전체정보 연구에 관한 글이므로 직접적인 관련이 있는 개인정보 보호법만을 분석 대상으로 한다.

“개인정보”의 정의는 개인정보 보호법이 적용될 것인지 아닌지를 결정하는 핵심적인 개념이다. 개정된 정의 조항은 가), 나), 다) 3개의 부분집합으로 구성되어 있는데, 가)목과 나)목은 개정전의 “개인정보” 정의를 거의 그대로 옮겨온 것이고,²⁾ 다)목은 이번 개정을 통하여 처음 도입된 “가명정보”의 개념을 “개인정보”의 범위에 추가한 것이다. 조금 복잡해 보이지만, 제1항 다)목과 이어지는 제1의 2항을 종합하면, 가명정보란 “**그 자체로 또는 다른 정보와 쉽게 결합하여 특정 개인을 알아볼 수 있는 정보를 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 처리함으로써, 원래의 상태로 복원하기 위한 추가 정보의 사용·결합 없이는 특정 개인을 알아볼 수 없는 정보**”로 정의가 될 것이다. 이 정의 규정만 놓고 보면 가명정보가 무엇을 말하는지 바로 머리에 떠오르지 않는 측면도 있다. 통상적으로는 성명, 주소, 전화번호, 주민등록번호 등 정보주체와 바로 연결이 되는 직접 식별자(direct identifiers)를 삭제하거나 대체하는 대신 정보주체와 바로 연결이 되지 않는 간접 식별자(indirect identifiers)는 그대로 유지하는 방식의 정보처리를 가명처리라고 부르는데, 개정 개인정보 보호법의 가명처리란 이와 같이 일반적으로 이해되는 가명화(pseudonymization)를 채택한 입법으로 보면 된다.

우리나라의 개인정보 보호법제에 “가명정보” 또는 “가명처리”라는 개념이 등장한 것인 이번이 개정된 개인정보 보호법이 처음은 아니다. 지난 2016년 6월 국무조정실, 행정자치부, 방송통신위원회, 금융위원회, 미래창조과학부, 보건복지부 등 6개 부처는 “개인정보 비식별 조치 가이드라인”이라는 지침(이하 “비식별 조치 가이드라인”)을 발표한 바 있다. 이 비식별 조치 가이드라인은 개인정보 보호법이 적용되는 개인정보의 범위가 지나치게 넓고 예측하기 어렵다는 비판에 대응하여 “개인을 알아볼 수 없는” 처리 방법의 기준을 제시한 가이드라인이다. 이 가이드라인은 비식별 조치 방법으로 가명처리, 총계 처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등을 제시하고 있다. 비식별 조치 가이드라인은 가명처리를 “개인 식별이 가능한 데이터를 직접적으로 식별할 수 없는 다른 값으로 대체하는 기법”이라고 설명하면서, 데이터의 변형 또는 변질 수준이 적은 것이 장점인 반면, 대체 값 부여 시에도 식별 가능한 고유 속성이 계속 유지된다는 것을 단점으로 기술하고 있었다.³⁾

2) 다만 나)목의 후문, 즉 “다른 정보와 쉽게 결합할 수 있는지 여부는 다른 정보의 입수 가능성 등 개인을 알아보는 데 소요되는 시간, 비용, 기술 등을 합리적으로 고려하여야 한다.”는 요소는 새로 삽입되었다.

3) 국무조정실 등 6개 관계부처(2016). “개인정보 비식별 조치 가이드라인-비식별 조치 기준 및 지원·관리체계 안내”. p.31.

위 비식별 조치 가이드라인이 정리하고 있듯이, 가명처리는 총계처리, 데이터 삭제, 데이터 범주화, 데이터 마스킹 등 다른 종류의 비식별조치와는 달리 데이터의 변형 또는 변질 수준이 적기 때문에 비식별화 이후의 분석에 야기되는 왜곡이 적다. 다만, 역시 비식별 조치 가이드라인이 지적하고 있듯이 대체 값, 즉 가명 자체가 원래의 정보 주체와 1:1로 고유하게 대응하는 값을 이용하는 것이므로 가명과 정보주체를 연결해주는 함수 또는 키 맵만 확보한다면 정보주체를 역추적하는 것이 얼마든지 가능하다. 그렇기 때문에 비식별 조치 가이드라인은 가명처리를 개인정보의 비식별 조치 방법의 한 가지로 소개하고 있으면서도, “가명처리 기법만 단독 활용된 경우는 충분한 비식별 조치로 보기 어렵다”라는 단서를 달고 있다.⁴⁾ 개정 개인정보 보호법도 가명처리의 이러한 기술적인 한계를 감안하여 가명처리 된 정보를 여전히 “개인정보”의 한 범주로 포함시키고 있다. 하지만 아래에서 보는 바와 같이 개인정보에 포함되기는 하더라도 가명정보는 별도의 특칙을 통하여 과학적 연구에 있어 다른 종류의 개인정보에 비하여 상대적으로 쉽게 활용할 수 있는 여지가 생겼으므로 그 의미는 상당하다.

2 유전체 연구에 미치는 영향

가. 가명정보 처리에 대한 특례

이처럼 가명처리를 거친 가명정보는 정의상으로는 여전히 “개인정보”의 한 종류를 이루지만, 개정 개인정보 보호법은 가명정보에 대하여는 제3장의 일부로 “제3절 가명정보의 처리에 관한 특례”로 신설된 섹션에서 다양한 특례를 인정하고 있기 때문에 정보처리자의 입장에서는 같은 개인정보라고 하더라도 가명처리가 된 가명정보이나 아니냐가 큰 차이가 있다. 개정 개인정보 보호법상 가명정보에 인정되는 특칙들은 다음과 같다.⁵⁾

4) 국무조정실 등 6개 관계부처(2016). "개인정보 비식별 조치 가이드라인-비식별 조치 기준 및 지원·관리체계 안내". p.7.

5) 아래 2개 조문 사이에 위치한 제28조의3 부터 제28조의6은 가명정보의 보호를 위한 규정들이다.

개인정보 보호법

제28조의2(가명정보의 처리 등) ① 개인정보처리자는 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있다.

② 개인정보처리자는 제1항에 따라 가명정보를 제3자에게 제공하는 경우에는 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함해서는 아니 된다.

제28조의7(적용범위) 가명정보는 제20조, 제21조, 제27조, 제34조 제1항, 제35조부터 제37조까지, 제39조의3, 제39조의4, 제39조의6부터 제39조의8까지의 규정을 적용하지 아니한다. 6)

제2조(정의)

8. “과학적 연구”란 기술의 개발과 실증, 기초연구, 응용연구 및 민간 투자 연구 등 과학적 방법을 적용하는 연구를 말한다.

연구자라면 무엇보다도 위 조문 가운데 “과학적 연구”에 해당하는 경우 정보주체의 동의 없이도 가명정보를 처리할 수 있다는 조문에 눈길이 끌릴 것이다. 그런데 기존 개인정보 보호법 체제하에서의 비식별화는 개인을 알아볼 수 없도록 해야 했고, 그러려면 유전체 정보의 염기서열을 그대로 존치할 경우 개인을 알아볼 수 없는 것이냐 하는 의문이 남아 있었기 때문에, 유전체정보 그대로는 개인정보로서 이용이 여의치 않은 커다란 문제가 있었다. 특히 유전체정보의 경우에는 연구자료로서의 가치를 보존하면서 비식별조치를 하는 방법은 현실적으로 가명처리 밖에 없다. 이론상으로는 예컨대 식별력이 높은 구간의 SNP를 임의로 바꾸는 방식으로 유전체정보 자체에 노이즈를 추가하는 방법을 생각할 수도 있지만, 그렇게 되면 정확한 SNP의 분석이 불가능해지므로 유전체정보를 연구하는 의미가 상실된다. 그런데 가명처리 자체만으로는 기존 비식별 조치 가이드라인에서도 충분한 비식별화는 아니라는 단서를 달고 있었고, 법원에서도 이를 충분한 비식별화로 인정하지 않는 태도를 보이고 있었다.7) 이에 반하여 새롭게 도입되는 가명정보 제도 하에서는 직접 식별자만을 대체하고 염기서열 정보는 그대로 유지하더라도 가명정보에 대한 특칙을 적용받을 수 있으므로 유전체 연구에 있어 획기적인 전환점이 되는 셈이다. 예컨대 질병의 진단을 위하여 유전체 분석 검사를 받은 환자의 유전체정보에 대하여 가명처리를 하게 되면 환자 본인의 동의가 없더라도 이를 연구에 이용하거나 과학적 연구를 하려는 제3자에게 제공할 수도 있게 되는 것이다.

6) 제20조(정보주체 이외로부터 수집한 개인정보의 수집 출처 등 고지), 제21조(개인정보의 파기), 제27조(영업양도 등에 따른 개인정보의 이전 제한), 제34조(개인정보 유출 통지 등), 제35조(개인정보의 열람), 제36조(개인정보의 정정·삭제), 제37조(개인정보의 처리정지 등). 제39조의3 부터 제38조의8은 정보통신서비스 제공자에게 적용되는 규정들이다.

7) 아래에서 설명할 약학정보원 사건 참조

“약학정보원” 사건

의약품 구매를 위하여 환자가 전문의약품의 처방전을 약국에 제출하면, 이 처방전은 약국 경영프로그램을 통하여 전산상으로 약학정보원이라는 기관에 집결되게 된다. 그런데 이 처방전 정보가 집결되는 약학정보원이 처방전에 담긴 처방내역을 다국적 데이터 마이닝 업체인 “I”사와 이용계약을 맺고 제공한 것이 개인정보 보호법 위반이 아닌가가 민사적·형사적으로 문제가 된 사건이다.⁸⁾ 이 사건에서 약학정보원은 최초에는 환자의 13자리의 주민등록번호 중 홀수 자리와 짝수 자리를 각 다른 암호화규칙에 따라 영문으로 치환한 다음 양끝 2자리에 임의의 영문자를 추가로 입력하는 방식으로 비식별화를 하였다가 (“1기 암호화 방식”), 이후 환자의 주민등록번호를 SHA 함수를 이용하여 일방향 암호화 방식으로 변경하였다 (“2기 암호화 방식”), 최종적으로는 환자의 주민등록번호가 아니라 성명, 생년월일, 성별을 SHA 함수를 이용하여 암호화 하였다 (“3기 암호화 방식”).

민사사건 항소심을 다룬 서울고등법원은 “개인정보에 암호화 등 적절한 비식별화 조치를 취함으로써 특정 개인을 식별할 수 없는 상태에 이른다면 이는 식별성을 요건으로 하는 개인정보에 해당한다고 볼 수 없고, 따라서 정보주체의 동의 없이 통계작성 등의 용도로 이용되거나 제3자에게 제공되더라도 정보통신망법이나 개인정보 보호법을 위반한 것으로 볼 수 없다. 다만, 비식별화 조치가 이루어졌다고 하더라도 재식별 가능성이 합리적으로 존재한다면 적절한 비식별화 조치가 이루어지지 않은 것”이라고 전제하였다. 이어서 서울고등법원은, “1기 암호화 방식은 그 자체로 식별하지 않으면서 특정만 하는 방법으로는 충분하지 아니하고 특정한 값에 대한 다양한 추론을 통해 쉽게 복호화할 수 있어 개인이 식별될 우려가 클 뿐만 아니라 비식별화 알고리즘이 약학정보원과 I사 사이에 공유되고 있어 적절한 비식별화 조치가 이루어졌다고 보기 어렵고, 2, 3기 암호화 방식도 1기 암호화 방식과 결합하여 개인을 식별할 수 있다고 봄이 타당하므로, 결국 1,2,3기 암호화 방식으로 처리가 된 정보는 여전히 개인정보에 해당한다”고 판단하였다.

이 사건에서 약학정보원과 “I”사가 의존한 비식별 조치 방법은 가명처리에 해당하는데, 기존의 개인정보 보호법 하에서는 가명처리 후에도 “개인을 식별할 수 있는지 여부”가 개인정보 보호법의 적용 또는 통계나 학술연구 목적의 특칙 적용 여부를 좌우하였으나, 만약 같은 사건이 개정 개인정보 보호법의 테두리 안에서 일어난다면 개인을 식별할 수 있는지 여부가 아니라 “개인정보처리자가 가명정보를 제3자에게 제공하면서 특정 개인을 알아보기 위하여 사용될 수 있는 정보를 포함”시켰는지의 여부, 즉 맵핑 테이블이나 복호화 함수를 제공했는지 등에 따라 결정될 것이다

나. 기존 개인정보 보호법상 연구 특칙과의 비교

이번에 개인정보 보호법이 개정되기 이전에도 연구를 목적으로 한 개인정보의 이용에 대한 특례가 없었던 것은 아니다. 제18조 제2항 제4호에 통계작성 및 학술연구의 목적으로 정보주체의 동의 없이도 이미 수집한 개인정보를 이용하거나 이를 제3자에게 제공하는 것이 허용되는 다음과 같은 예외 규정이 있었다.

8) 이 사건에서 “I”사는 약학정보원 이외의 다른 경로를 통해서도 처방전 데이터를 수집했으나, 여기서는 간결한 설명을 위하여 약학정보원만을 대표로 언급한다.

기존 개인정보 보호법

- 제18조(개인정보의 목적 외 이용·제공 제한) ① 개인정보처리자는 개인정보를 제15조제1항에 따른 범위를 초과하여 이용하거나 제17조 제1항 및 제3항에 따른 범위를 초과하여 제3자에게 제공하여서는 아니 된다.
- ② 제1항에도 불구하고 개인정보처리자는 다음 각 호의 어느 하나에 해당하는 경우에는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 있을 때를 제외하고는 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있다. 다만, 제5호부터 제9호까지의 경우는 공공기관의 경우로 한정한다.
4. 통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우

기존 개인정보 보호법상의 연구 목적 이용 특칙은 상당한 비판의 대상이었다. 우선 학술 연구를 목적으로 하는 경우 인정되는 특칙은 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없다는 전제하에 최초 수집 목적 외의 용도, 즉 학술연구 목적으로 이용하거나 제3자에게 제공할 수 있다는 것인데, 이는 우리 개인정보 보호법상 언론, 종교단체, 정당이 각각 취재·보도, 선교, 선거 입후보자 추천 등 고유 목적을 달성하기 위하여 수집·이용하는 개인정보에 대하여는 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없을 것을 요구하지도 않으면서 개인정보 보호법 제3장부터 제7장까지를 아예 적용에서 제외하는 특칙보다도⁹⁾ 훨씬 엄격하다. 또한 이웃 일본의 個人情報の保護に関する法律(개인정보의 보호에 관한 법률)에서 대학 등 연구기관이나 그에 속한 사람이 학술연구 목적으로 개인정보를 이용할 경우 보다 넓은 범위에서 법 적용을 제외하는 것과도 대비된다.¹⁰⁾

나아가 학술연구 등의 목적으로 동의 없이 이용하거나 제3자에게 제공하기 위하여는 “특정 개인을 알아볼 수 없는 형태”로 정보를 제공하여야만 했다. 여기서 “특정 개인을 알아볼 수 없는 형태”가 어떤 의미인지는 논란이 많았는데, 대체로 위에서 언급한 6개 부처의 비식별 조치 가이드라인에서 제시된 비식별조치를 취해야 하는 것으로 이해가 되었다. 앞에서 설명한 것처럼 유전체정보는 정보로서의 가치를 훼손하지 않고는 비식별화가 용이하지 않기 때문에, 이는 바꿔 말하면 기존 개인정보 보호법의 학술연구 목적의 특칙은 유전체 정보의 연구에 있어서 그다지 의미있는 역할을 할 수 없었다.

따라서 이번 개인정보 보호법의 개정을 통하여 가명정보에 대한 특칙이 인정됨으로써 기존의 개인정보 보호법 체제하에서 유전체정보를 연구 목적으로 활용하거나 제3자에게 제공하는데 어려움이 있었던 점이 개선되었다고 할 수 있다.

9) 개인정보 보호법 제58조 제1항.

10) 일본 個人情報の保護に関する法律 제76조

끝으로 외국처럼 개인들이 자발적으로 DTC 유전자 검사 등을 통하여 획득한 유전 정보를 혈족 찾기 사이트 등의 공중 영역에서 활발하게 공유하는 바람에 난수표 같은 염기서열만을 갖고 그 본인을 추적하는 것이 가능해진 경우에는 유전체정보 자체를 비식별화할 필요성이 있지만, 아직 비교 대상으로 삼을 공개된 유전체정보가 방대한 양으로 존재하지 않는 우리나라에서는 염기서열에 불과한 유전체정보를 비식별화 할 필요가 없는 것은 아닌가 하는 반론을 생각해 본다. 물론 그러한 반론에 논리적인 타당성이 없는 것은 아니다. 그러나 유전체정보라는 것은 일생동안 불변하는데다가, 우리나라에서도 DTC 유전자 검사의 규제 완화 등을 통하여 일반인들의 유전체정보가 자발적으로 공중 영역에서 접근가능해지는 날이 얼마든지 올 수도 있다. 따라서 유전체정보는 유전체정보 이용 패턴이 바뀔 수도 있는 미래를 염두에 두고 보호 체제를 논해야 할 것이므로, 보호의 정도는 외국처럼 공중 영역에 유전체 정보가 이미 다량 존재하고 있는지에 따라 달라질 것은 아니라고 본다.¹¹⁾

3 유전체 연구와 관련하여 남아 있는 쟁점

가. 가명정보의 특칙이 유전정보에도 적용되는가

개인정보 보호법은 사상·신념, 노동조합·정당의 가입·탈퇴, 정치적 견해, 건강, 성생활 등에 관한 정보, 유전자검사 등의 결과로 얻어진 유전정보, 범죄경력자료를 “민감정보”로 따로 분류한다.¹²⁾ 민감정보는 그 처리를 법령에서 요구하거나 허용하는 경우가 아니라면 정보주체에게 일반 개인정보의 처리와는 별도의 동의를 받아야 하고, 이러한 별도의 동의가 없는 경우에는 원칙적으로 처리가 금지된다.¹³⁾ 여기서의 “유전정보”란 개인정보 보호법이

11) 다만 외국에서 혈족 찾기 사이트 등에 유전(체) 정보가 대량으로 누적되는 것은 정보주체 본인들의 자발적인 행위에 의한 것이므로 “타인”의 개인정보를 보호하기 위한 개인정보 보호법제에 의한 규제가 어려운 것은 외국이나 우리나라나 마찬가지이다. 하지만 유전(체) 정보는 가까운 혈족끼리 상당 부분을 공유하므로 나의 정보가 일부 타인의 정보를 구성하기 때문에 자기 혼자 모든 결정을 하는 것이 적절하지 않을 수도 있고, 그런 점을 지적한 논문으로는 이원복(2019). “내 DNA 정보는 내 마음대로 사용해도 되는가—DNA 정보의 특수성과 자기정보 통제권의 제한”. 정보법학, 제23권 제1호, pp.182-215.

12) 개인정보 보호법 제23조 제1항, 개인정보 보호법 시행령 제18조. 최근 개정된 이 시행령은 추가로 개인의 신체적, 생리적, 행동적 특징에 관한 정보로서 특정 개인을 알아볼 목적으로 일정한 기술적 수단을 통해 생성한 정보, 즉 바이오메트릭 정보와 인종이나 민족에 관한 정보를 민감정보에 추가하였다.

13) 개인정보 보호법 제23조 제1항.

별도의 정의를 두고 있지는 않지만, 유전체정보도 당연히 유전정보에 속할 것이다. 그런데 정보주체의 동의 없이 최초 수집목적과 다른 목적으로 개인정보를 이용하거나 또는 제3자에게 개인정보를 제공하는 것이 허용되는 예외사유인 개인정보 보호법 제18조이 민감정보에 대하여도 적용되는지와 관련하여 논란이 있었다.

개인정보는 최초 정보주체에게 동의를 받은 범위 내에서만 이용을 하고 제3자에게 제공하는 것이 원칙이지만,¹⁴⁾ 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없다는 전제하에 개인정보를 목적 외의 용도로 이용하거나 이를 제3자에게 제공할 수 있는 몇 가지 예외 사유를 개인정보 보호법이 인정하고 있다.¹⁵⁾ 그리고 앞에서 본 바와 같이 금번의 개인정보 보호법 개정 전까지는 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우”를 그 예외사유의 하나로 인정하고 있었다. 유전정보와 같은 민감정보도 이렇게 목적 외의 용도로 이용하거나 이를 제3자에게 제공하는 것이 허용될 것인가? 일부 학자들은 민감정보를 더 두텁게 보호하는 조문의 위치가 동의 없이 개인정보를 이용할 수 있는 예외적인 사유를 나열한 조문의 위치보다 나중에 있고, 민감정보를 더 두텁게 보호하는 취지를 감안하면 동의 없이 사용하는 예외적인 사유는 민감정보에 대하여 인정되어서는 안 된다는 태도를 취한다. 반면, 정보주체 본인의 동의 없이 목적 외의 용도로 이용하거나 이를 제3자에게 제공하는 것이 허용되는 예외 사유들이 극히 제한적이고, 정보주체 또는 제3자의 이익을 부당하게 침해할 우려가 없어야 하며, 일부 예외사유는 공공기관에게만 허용되고, 일부 예외사유는 오히려 정보주체 본인의 이익에 부합하며,¹⁶⁾ 통계작성 및 학술연구 등의 목적을 위한 경우는 특정 개인을 알아볼 수 없는 형태로 제공해야 한다는 제약까지 있기 때문에, 민감정보라고 하더라도 법이 정한 예외 사유를 충족하는 이상 본인의 동의 없이 최초 수집 목적 외의 용도로 이용하거나 이를 제3자에게 제공하는 것이 허용된다는 것이 필자의 견해이다.

개인정보 보호법이 개정되면서 유일하게 “통계작성 및 학술연구 등의 목적을 위하여 필요한 경우로서 특정 개인을 알아볼 수 없는 형태로 개인정보를 제공하는 경우”가 본인의 동의 없이 목적 외의 용도로 이용하거나 이를 제3자에게 제공하는 것이 허용되는 예외

14) 개인정보 보호법 제18조 제1항.

15) 개인정보 보호법 제18조 제2항.

16) 개인정보 보호법 제18조 제2항 제3호는 “정보주체 또는 그 법정대리인이 의사표시를 할 수 없는 상태에 있거나 주소불명 등으로 사전 동의를 받을 수 없는 경우로서 명백히 정보주체 또는 제3자의 급박한 생명, 신체, 재산의 이익을 위하여 필요하다고 인정되는 경우”를 예외사유로 인정하고 있다.

사유에서 삭제가 되고, 대신 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보 주체의 동의 없이 가명정보를 처리할 수 있는 것으로 대체가 되었다. 하지만 본인 동의 없는 이용이나 제3자 제공이 허용되는 예외사유가 민감정보에 대하여도 적용이 될 것인가 하는 이슈가 새로 만들어진 가명정보 특칙에도 계속하여 제기될 여지가 있다. 국회에서 개인정보 보호법을 개정하는 과정에서 민감정보도 가명처리 시 특칙이 인정되는 것으로 볼 것인지에 대하여 논의를 거쳤더라면 “입법자의 의도”를 판단할 수 있는 자료가 있으므로 법 해석에 결정적인 도움이 되었을 것이라는 아쉬움이 있다.

나. 『생명윤리 및 안전에 관한 법률』과의 관계

『생명윤리 및 안전에 관한 법률』(이하 “생명윤리법”)은 인간대상 연구의 맥락에서 연구 대상자의 개인정보를 보호하기 위한 규정들을 두고 있다. 연구대상자로부터 유래하는 개인정보에 대하여 생명윤리법과 개인정보 보호법이 각각 어떤 범위에서 적용될 것인가가 문제될 수 있다. 생명윤리법은 “생명윤리 및 안전에 관하여는 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에 따른다.”¹⁷⁾고 하고 있고, 인간을 대상으로 하는 연구에 대하여 적용되는 기본법이 생명윤리법이므로, 연구의 맥락에서 개인정보에 대한 보호는 일단 생명윤리법이 적용이 되고, 생명윤리법에서 특별히 규정하고 있지 않는 내용은 생명 윤리법과 충돌이 없는 범위에서 개인정보 보호법 역시 적용이 된다고 본다.

예를 들어 연구를 하는 과정에서 연구자가 구체적으로 동의를 받지 않는 정보까지 수집하는 상황을 가정해보자. 생명윤리법과 개인정보 보호법 모두 정보를 수집할 때는 정보주체 본인으로부터 동의를 얻도록 하고 있다.¹⁸⁾ 그런데 개인정보 보호법은 동의를 받지 않는 정보 수집 및 이용에 대하여 형사처벌이 가능하도록 하고 있는 반면,¹⁹⁾ 생명윤리법은 그에 상응하는 처벌 조항을 두고 있지 않다. 이 경우 비록 생명윤리법은 처벌 조항을 두고 있지 않지만 개인정보 보호법이 처벌 조항을 두고 있으므로 여전히 개인정보 보호법에 따라 형사처벌이 가능할 것인가가 문제될 수 있다. 법 이론적으로 살펴볼 요소가 많은 쟁점이지만,

17) 생명윤리법 제4조 제1항.

18) 각각 생명윤리법 제16조 제1항과 개인정보 보호법 제15조 제1항 제1호.

19) 개인정보 보호법 제18조 제1항은 동의를 받은 범위를 넘어서는 개인정보의 이용을 금지하고 있고, 제71조 제2호는 이를 위반할 경우 5년 이하의 징역 또는 5천만 원 이하의 벌금으로 처벌할 수 있도록 한다.

필자의 견해로는 생명윤리법이 2012년 개정 당시 인간대상 연구와 관련한 조항들을 대거 새로 입법하면서 해당 조항을 위반하는 데에 대한 형사처벌은 입법하지 않은 것은 입법자가 이를 형사처벌의 대상으로 삼지 않겠다는 의지를 표명한 것이므로, 개인정보 보호법을 대신 적용하여 처벌할 수는 없다고 본다. 하지만 생명윤리법과 개인정보 보호법이 모두 적용될 수 있을 것 같은 상황에서 일률적으로 생명윤리법만이 적용된다는 의미는 아니고, 그 상황을 개별적으로 분석하는 수밖에 없다.

이 글이 다루고 있는 가명정보와 관련하여서도 생명윤리법은 개인정보 보호법과는 차별화된 독자적인 규정 체계를 두고 있다. 생명윤리법에서는 “가명정보” 또는 “가명처리”라는 개념이 명문으로 등장하지는 않는다. 하지만 “익명화”를 “개인식별정보를 영구적으로 삭제하거나, 개인식별정보의 전부 또는 일부를 해당 기관의 고유식별기호로 대체하는 것”이라 정의하고 있고,²⁰⁾ 연구자가 연구대상자로부터 동의를 받아 개인정보를 제3자에게 제공할 때 익명화하도록 하고 있는데,²¹⁾ “개인식별정보의 전부 또는 일부를 해당 기관의 고유식별기호로 대체하는 것”은 개인정보 보호법상의 가명처리와 실질적으로 다르지 않은 개념이다. 또한 보건복지부 산하 국가생명윤리정책원이 운영하는 기관생명윤리위원회의 정보포털에 의하면 생명윤리법에 따라 연구기관의 기관윤리위원회가 개인식별정보를 “익명화”하는 방법의 하나로 “기증자의 개인식별정보를 은행의 고유식별번호로 대체하고 해당 연결 파일에 대한 암호키를 보안책임자가 가지고 있는 방식”을 소개하고 있는데, 이는 가명처리(pseudonymization)에 해당함에 의문이 없다.²²⁾

그렇다면 유전체정보의 연구에 있어서는 생명윤리법과 개인정보 보호법이 각각 어떤 역할을 할 것인가? 우선 유전체정보가 수집되는 경위가 처음부터 연구 목적이라면 생명윤리법이 적용이 될 것이고, 그렇지 않고 희귀병의 진단을 위하여 수집된 경우라면 연구 목적이 아니므로 개인정보 보호법이 적용되는 것이 원칙이다. 진단 목적으로 수집된 유전체 정보를 이후에 연구 목적으로 이용하거나 다른 연구 기관과 같은 제3자에게 제공하고자

20) 생명윤리법 제2조 제19호.

21) 생명윤리법 제18조.

22) 국가생명윤리정책원(2020). “기관생명윤리위원회 정보포털: 정보 관리 및 개인정보보호”. <https://www.irb.or.kr/UserMenu06/PreservationInformationManagement.aspx> (접속일자: 2020.05.10.) 이 웹사이트가 익명화 방법으로 소개하는 다른 방법은 “기증자의 개인식별정보를 영구적으로 삭제하는 방식”인데, 이는 anonymization에 해당한다. 즉, 기관생명윤리위원회 정보포털은 “익명화”의 방법으로 pseudonymization과 anonymization을 제안하는 셈이다. “익명화”는 anonymization의 번역이므로, 생명윤리법이 개인정보 보호법처럼 “익명화” 대신 “비식별화”라는 용어를 사용했다라면 지금처럼 “익명화” 방법에 다시 “가명화”와 “익명화”가 포함되어버리는 순환론적인 구분을 피할 수 있어 덜 혼란스러웠을 것이라는 아쉬움이 있다.

할 경우에는 지금까지 다른 개인정보 보호법상의 가명처리가 적용된다.²³⁾ 그게 아니라 최초 유전체정보의 수집부터 생명윤리법에 따라 연구의 일환으로 이루어졌다면 그때는 개인정보 보호법 대신 생명윤리법이 정한 최초 목적외 이용 및 제3자 제공에 관한 규정들이 적용이 될 것이다.²⁴⁾

하지만 유전체정보를 기반으로 하는 정밀의료의 맥락에서는 “진단/치료 vs. 연구”의 이분법적인 접근이 무너지는 점에 주목할 필요가 있다. 다수의 과학자들이 제안한 소위 “N of 1” 임상시험, 즉 연구대상자가 한 명인 임상시험이 등장하고 있기 때문이다.²⁵⁾ 이러한 연구대상자 1명의 임상시험은 연구의 결과에 따라 직접 연구대상자에게 적용할 수 있는 치료방법의 발견으로 이어질 수도 있기 때문에, 환자 개개인의 유전적 그리고 환경적 특성을 면밀하게 분석하여 진단과 치료를 결정하는 정밀의료의 채택해야 할 방법론으로 제시되고 있기도 하다. 그렇다면 이러한 “N of 1”의 임상시험을 연구로 보아서 생명윤리법의 적용 대상으로 삼을 것인지, 아니면 외형은 임상시험의 형식을 취하고 있지만 그 실질은 통상의 진단이나 치료와 다를 바 없으므로 개인정보 보호법의 적용 대상으로 삼을 것인지가 이슈가 될 수 있다. 생명윤리법으로 대표되는 연구대상자 보호제도란 인류의 지식 발전에는 커다란 공헌을 하지만 연구를 통하여 직접적인 혜택을 받기는 어렵고 오히려 해를 입을 수도 있는 연구대상자를 보호하기 위한 윤리적인 필요성에서 출발했다는 점을 감안한다면, 외형은 임상시험이지만 그 시험의 혜택이 임상시험 대상자 본인에게 직접적으로 귀속될 것이 처음부터 예정된 “N of 1” 임상시험의 경우에는 진단과 치료에 준하여 개인정보 보호법의 적용 대상으로 보는 것이 실질에 더 부합한다는 것이 필자의 견해이다. 따라서 “N of 1” 임상시험을 통하여 유전체정보를 수집하고 분석하는 경우에는 생명윤리법이 아닌 개인정보 보호법에 따라 환자의 동의를 얻고, 이를 향후 연구목적으로 제3자에게 제공하고자 하는 경우에도 개인정보 보호법에 따른 가명처리를 거치면 가능한 것으로 해석할 것이다. 그렇지 않고 생명윤리법을 따르도록 한다면 환자의 동의를 구해야 한다는 점 자체는 큰 차이가 없으나, 개인정보 보호법에서는 요구되지 않지만 생명윤리법에서 요구하

23) 물론 유전정보에 대하여도 가명처리에 따른 특칙이 허용될 것인가 하는 이슈가 있을 수는 있다는 점은 앞에서 언급한 바와 같다.

24) 생명윤리법 제16조 제3항 및 제18조. 이에 대한 자세한 설명은 이원복(2018). “유전체 시대의 유전정보 보호와 공유를 위한 개인정보 보호법제의 고찰”. 법조, 제67권 제3호, pp.597-644.

25) 해외에서 진행 중인 “N-of-1” 임상시험을 찾아보는 것은 어렵지 않다. 예를 들어 미국 컬럼비아 대학 병원에서는 다수의 암에 대하여 “N-of-1” 임상시험을 진행하고 있다. Columbia University Medical Center Systems Biology Lab, “N-of-1 Clinical Trials and Precision Medicine for Cancer”, <http://califano.c2b2.columbia.edu/n-of-1-clinical-trials-precision-medicine-cancer> (접속일자: 2020.05.10)

는 기관윤리위원회의 심의를 거쳐야 한다는 결론이 되는데,³¹⁾ 기관윤리위원회의 심의를 거치는데 필요한 소요기간을 감안하면 환자 본인의 복리에 과연 도움이 되는 절차인가 하는 점에 큰 의문이 있다.²⁷⁾

4 유럽연합과의 비교

우리나라의 기존 개인정보 보호법은 유럽의 영향을 받았다.²⁸⁾ 유럽 연합은 과거에는 1995년에 채택한 Directive 95/46/EC³⁾을 통하여 각 회원국이 개인정보 보호를 위하여 입법해야 하는 공통된 원칙을 규정하고, 실질적인 입법은 유럽연합 각 회원국이 위 원칙을 준수하는 범위 내에서 각자 도입하는 형태를 취하였다.²⁹⁾ 그러다가 유럽 연합 내에서 개인정보 보호를 통일된 방법으로 취해야 한다는 목소리가 높아지자, 지난 2016년 각 회원국에게 직접 구속력을 갖고 적용되는 General Data Protection Regulation(GDPR)이라는 제목의 규범을 채택한 바 있고, 이는 지난 2018년부터 시행되기 시작하였다. 그동안 우리나라의 개인정보 보호법이 유럽의 영향을 많이 받아 온 만큼, 2018년의 GDPR이 우리나라 개인정보 보호법의 2020년 개정에도 영향을 미쳤을 것을 짐작하는 것은 어렵지 않다.

실제로 GDPR이 담고 있는 가명처리 조항은 우리 개인정보 보호법의 그것과 매우 흡사하다. GDPR의 가명처리 조항은 1995년의 Directive 95/46/EC³⁾에는 포함되어 있지 않던 것이 2016년에 GDPR이 새로이 제정되면서 도입된 것이기도 하다. 우선 GDPR의 가명처리 정의 관련 조항을 본다 (밑줄은 모두 필자가 부기한 것임). 아래와 같이 GDPR의 전문에서 가명처리된 개인정보도 여전히 개인정보로 취급해야 함을 천명한 점이라든가 정의 규정에서 가명처리가 어떤 과정을 거친 정보를 말하는지 정한 내용은 우리 개정 개인정보

26) 생명윤리법 제15조.

27) 외국에서도 N-of-1 임상시험이 기관윤리위원회의 심의를 거쳐야 하는가를 놓고 의견이 나뉘고 있다. Cen, R., Hussain, A., Pak, K. J., Mitchell, G., Nikles, J., Gaudreau, S., Bazzano, L. A. and Breault, J. L.(2016). "Do N-of-1 Trials Need IRB Review?". *Journal of Empirical Research on Human Research Ethics*, 11(3), pp.250-255. (DOI: 10.1177/1556264616662560).

28) 그에 반하여 아래에서 상술할 연구대상자 보호법제는 미국 Common Rule의 영향을 받았다.

29) 이 지침이 우리나라의 개인정보 보호법 제정에도 직접적인 영향을 미친 것으로 알려져 있다. 함인선(2013). "EU의 '1995년 개인정보지침'에 관한 법적 고찰". 법학논총, 33권 1호.

보호법과 큰 차이가 없다.

Recital 26

... Personal data which have undergone pseudonymization, which could be attributed to a natural person by the use of additional information, should be considered to be information on an identifiable natural person. ...

Article 4 – Definitions

...

5. 'pseudonymisation' means the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person;

...

그러나 GDPR에서 가명처리가 갖는 위상은 우리 개정 개인정보 보호법과 상당한 차이가 있다. GDPR에서 가명처리를 언급한 나머지 조문들을 아래에서 본다.

Article 6 – Lawfulness of processing

...

4. Where the processing for a purpose other than that for which the personal data have been collected is not based on the data subject's consent or on a Union or Member State law which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected, take into account, inter alia:

...

e) the existence of appropriate safeguards, which may include encryption or pseudonymisation.

...

Article 25 – Data protection by design and by default

1. Taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing, the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organisational measures, such as pseudonymisation, which are designed to implement data-protection principles, such as data minimisation, in an

effective manner and to integrate the necessary safeguards into the processing in order to meet the requirements of this Regulation and protect the rights of data subjects.

...

Article 32 – Security of processing

1. Taking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk, including inter alia as appropriate:

- a) the pseudonymisation and encryption of personal data;

...

Article 89 – Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes

1. Processing for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, shall be subject to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. Where those purposes can be fulfilled by further processing which does not permit or no longer permits the identification of data subjects, those purposes shall be fulfilled in that manner. ...

즉, GDPR에서 가명처리란 개인정보를 비식별화 하는 방법의 하나로서, 개인정보를 가명처리한 경우에는 그만큼 개인정보 보호를 위한 조치를 취한 것으로 인정이 되는 것이다. 위의 제6조 제4항이라든가 제32조 제1항을 보면 가명처리와 암호화(encryption)를 나란히 언급한 것에서도 그런 취지는 명확하게 드러나고 있다. GDPR에 따른 과학적 연구 특례를 인정받아 개인정보를 취급하는 경우라고 하더라도 적절한 보호조치(safeguards)를 게을리 하면 안 되는데, 적절한 보호조치의 하나로서 가명처리를 언급하고 있는 제89조 제1항을 보더라도 그 점은 명백하다. 한마디로 GDPR하에서는 가명처리가 과학적 연구의 특례를 인정받기 위한 전제 조건으로서 요구되는 것이 아니고, 과학적 연구의 맥락이든 아니든 개인정보의 식별력을 줄임으로써 개인정보를 보호하는 방법의 하나로 인정되고 있는 것이다.

그렇다면 GDPR에서 과학적 연구에 인정되는 특례는 무엇인가? 그 특례는 개정된 우리 개인정보 보호법상 가명정보를 과학적 연구에 이용할 경우 인정되는 특례와 이제는 크게

다르지 않다. 우리나라가 2020년 초 개인정보 보호법을 개정하는 과정에서 GDPR을 모델 삼아 참고했기 때문에 당연한 결과이다. 예를 들어 GDPR은 이미 적법하게 수집한 개인정보를 과학적 연구에 이용함에 있어서는 추가 동의를 요구하지 않도록 하고 있고,³⁰⁾ 정보주체 본인에게 정보처리를 알려야 하는 고지 의무라든가 소위 잊혀질 권리로 불리는 정보 삭제권에 대하여도 일부 예외를 인정하며, 정보주체가 본인의 정보처리에 대하여 반대할 수 있는 반대권에 대하여도 면제가 인정될 수 있음은 앞에서 본 바와 같이 이제는 우리 개인정보 보호법도 가명처리 된 개인정보를 과학적 연구에 이용할 때 인정하는 특례와 유사하다. 다만 위에서 우리 개인정보 보호법에서는 가명정보의 특칙이 유전(체)정보와 같은 민감정보에 대하여도 적용이 되는지 의문이 있을 수 있다고 했고 가명처리가 과학적 연구 특례를 인정받기 위하여 전제 조건이므로 결국 이는 유전(체)정보의 과학적 연구 이용 방법에 직접적으로 영향을 미칠 수 있음을 살펴보았다. 반면 유럽 GDPR은 유전정보를 포함하는 민감정보의 처리를 원칙적으로 금지하면서 특별히 두텁게 보호하는 것은 우리나라와 동일한 반면,³¹⁾ 과학적 연구에 대하여는 명문으로 이러한 원칙에 대한 예외를 두고 있고, 조문의 위치 또한 민감정보 처리에 관한 조문 내부에 두고 있어서 우리나라와 같이 조문 위치로 인한 논란이 생길 여지도 없다.³²⁾

결론적으로 우리 개인정보 보호법제의 모델이 되어 온 유럽연합의 개인정보 보호 규범인 GDPR과 우리 개인정보 보호법을 비교한다면, GDPR에서는 과학적 연구를 인정받기 위한 요건에 특별히 가명처리와 같은 전제조건이 없고 민감정보에 대하여도 과학적 연구의 특례가 인정되는지에 관하여 의문의 여지가 없다는 점에서 차이가 있지만, 일단 과학적 연구에 해당할 경우 인정되는 특례들의 범위는 2020년 개인정보 보호법의 개정을 통하여 이제 우리나라도 GDPR과 대동소이해졌다. 또한, 가명처리만 놓고 보면, GDPR에서 가명 처리는 과학적 연구가 준수해야 하는 개인정보 보호방법의 하나로 인정되어 있는 반면, 우리 개인정보 보호법에서는 가명처리가 과학적 연구 특례를 인정받기 위한 전제조건이므로 다소 차이가 있다.

30) 과학적 연구라면 정보를 최초로 수집하는 단계에서부터 정보주체의 동의가 면제되는 것인가 하는 점에 있어서는 명문의 규정이 없으므로 논란이 있을 수 있다.

31) GDPR은 “민감정보”라는 표현 대신 “special categories of information”, 즉 “특별한 부류의 정보”라는 표현을 사용하고 있다.

32) GDPR Article 9 – Processing of special categories of personal data. 다만 GDPR Article 6 (4)에 의하면 본인의 명시적 동의가 없는 목적으로 개인정보를 사용하고자 할 경우 최초의 정보 수집 목적과 너무 동떨어진 것은 아닌지를 따질 때 정보처리가 고려해야 할 사항으로서 해당 정보가 민감정보에 해당하는지를 살펴야 한다는 내용은 있다.

5 결론

이번 개인정보 보호법 개정은 우리나라 개인정보 보호법제가 너무 정보주체의 자기 결정권만을 보호한 나머지 2차적인 연구와 상업화를 지나치게 억제했다는 사회적인 공감대가 반영된 것이다. 개정을 통하여 도입된 내용 가운데 가명정보 및 가명정보 처리의 특례는 특히 유전체정보의 수집과 분석이 핵심이 되는 정밀의료 연구자들에게 상당히 중요한 의미가 있다. 물론 유전체정보의 수집이나 분석이 진료의 맥락이 아닌 연구의 맥락으로 이루어지는 경우에는 어차피 개인정보 보호법이 아니라 생명윤리 및 안전에 관한 법률이 적용되는 것이고, 생명윤리 및 안전에 관한 법률은 기존에 가명처리에 관한 독자적인 규율을 갖고 있었기 때문에 개인정보 보호법 개정이 가져다주는 변화가 없는 것은 아닌가 하는 반문을 할 수도 있지만, 진료 목적으로 수집되는 유전체정보에는 개인정보 보호법이 직접 적용이 될 것이고, 정밀의료의 일부로 활발히 논의가 되고 실제로 외국에서는 시행이 되고 있는 “연구대상자 1인 임상시험”, 즉 “N-of-1 clinical trial”은 그 실질을 연구가 아니라 진료로 보아 생명윤리 및 안전에 관한 법률의 규율 대상이 아니라 개인정보 보호법의 규율 대상으로 본다면, 이번 개인정보 보호법의 개정이 가져다주는 의미는 그만큼 더 크다고 할 것이다.

다만, 가명처리가 과학적 연구 특례를 인정받기 위한 전제 조건인 우리나라와 달리 유럽 연합은 유전체 연구와 같은 과학적 연구에서 가명처리가 강제되는 것이 아니라 개인정보 보호를 위한 하나의 방법으로 인정되고 있다는 점과 유전(체) 정보와 같은 민감정보 역시 과학적 특례 인정 대상인지 여부에 대한 의문이 적다는 점은 앞으로 우리 개인정보 보호법제가 나아갈 추가 개선 방향을 시사하는 것이라고 하겠다.

참고문헌

- 국무조정실 등 6개 관계부처(2016). “개인정보 비식별 조치 가이드라인-비식별 조치 기준 및 지원·관리체계 안내”.
- 김현숙(2020). “과학적 연구목적을 위한 개인정보 처리에 관한 비교법적 연구”. 정보법학, 24권 1호, pp.115-144.
- 박태신(2018). “인간대상연구에서 개인정보처리에 대한 고찰”. 저스티스, 제167호, pp.280-305.
- 이동진(2017). “개인정보 보호법 제18조 제2항 제4호, 비식별화, 비재산적 손해-이른바 약학정보원 사건을 계기로”. 정보법학, 제21권 제3호, pp.259-260.
- 이원복(2020). “유전체 연구와 개정 개인정보 보호법의 가명처리 제도”. 법학논집, 제25권 제1호.
- 이원복(2019). “내 DNA 정보는 내 마음대로 사용해도 되는가-DNA 정보의 특수성과 자기정보 통제권의 제한”. 정보법학, 제23권 제1호, pp.182-215.
- 이원복(2018). “유전체 시대의 유전정보 보호와 공유를 위한 개인정보 보호법제의 고찰”. 법조, 제67권 제3호, pp.597-644.
- 함인선(2013). “EU의 ‘1995년 개인정보지침’에 관한 법적 고찰”. 법학논총, 33권 1호. pp.253-294.
- Cen, R., Hussain, A., Pak, K. J., Mitchell, G., Nikles, J., Gaudreau, S., Bazzano, L. A. and Breault, J. L.(2016). “Do N-of-1 Trials Need IRB Review?”. *Journal of Empirical Research on Human Research Ethics*, 11(3), pp.250-255. (DOI: 10.1177/1556264616662560).
- El Emam, K. & Hintze, M. “Comparing the Benefits of Pseudonymization and Anonymization”. Under the GDPR, p.3.
- Schork, N. J.(2015). “Personalized Medicine: Time for One-Person Trials”. *Nature*, 520, p.609.
- William Lowrance, W. W.(2012). “Privacy, Confidentiality, and Health Research”. Cambridge University Press. p.118.

한림연구보고서 133

유전체정보기반 정밀의료 발전방향

Development of Genetic Information-Based Precision Medical Care

발행일 2020년 10월

발행처 한국과학기술한림원

발행인 한민구

전화 031) 726-7900

팩스 031) 726-7909

홈페이지 <http://www.kast.or.kr>

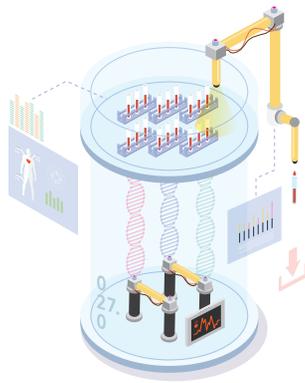
E-mail kast@kast.or.kr

편집/인쇄 경성문화사 02) 786-2999

I S B N 979-11-86795-54-5 94080

978-89-88706-06-0 (세트)

- 이 책의 저작권은 한국과학기술한림원에 있습니다.
- 한국과학기술한림원의 동의 없이 내용의 일부를 인용하거나 발췌하는 것을 금합니다.



이 사업은 복권기금 및 과학기술진흥기금 지원을 통한 사업으로
우리나라의 사회적 가치 증진에 기여하고 있습니다

