

DNA 시퀀싱과 AI

I. 들어가며

II. 유전체학(genomics)

1. 유전체(genome)
2. 염색체(chromosome)
3. DNA (deoxyribonucleic acid)
4. 유전자(gene)
5. DNA 염기서열(DNA 시퀀스, DNA sequence)
6. 변이(variants, mutation)

III. 유전체학과 AI

1. 유전체 정보와 유전정보
2. 유전체 정보 분석
3. AI 기술 적용사례

IV. 의료에의 적용

V. 유전체 정보와 관련된 법적 쟁점

1. 유전체 정보의 특성
2. 유전체 정보와 GDPR
3. 유전체 정보와 개인정보 보호법
4. 유전정보와 생명윤리법

VI. 결론



진태환
서울대학교
법학전문대학원 12기

I. 들어가며

지금까지 의학은 ‘평균적 인간’을 전제로 표준화된 치료 지침과 진료 방식을 이용하는 ‘평균적 의학’의 형태로 발전해왔다. 그 결과 어떤 약물의 처방이 특정인에게는 효과를 내지 못하고, 특정인에게는 과한 효과를 내어 독이 되기도 하였다. 하지만 2003년 인간게놈프로젝트가 완성되고 AI 기술이 급속도로 성장하면서 개인의 고유한 유전적 특성을 고려한 ‘맞춤의학’이 발전하고 있다. 맞춤의학은 개인의 유전적 정보의 총합인 유전체에서 DNA 시퀀싱을 통해 분석가능한 유전체 정보(genomic data)를 추출하고, 이 정보를 AI 기술 등을 이용해 분석하는 것을 기초로 한다. 이 때 유전체 정보는 개인의 질병, 가족력 등 건강 정보뿐만 아니라 개인의 성격, 기호 등 민감한 사생활 정보까지 포함하고 있다. 이에 맞춤 의학을 비롯한 미래의학에서 유전체 정보를 활용하는 데 앞서 여러 법적 쟁점에 대한 논의가 필요하다. 이하에서는 DNA 시퀀싱을 통해 추출된 유전체 정보가 무엇인지, AI 기술과 결합하여 어떻게 활용될 수 있는지, 현행법제 상 유전체 정보는 어떻게 정의되고 규제되고 있는지 살펴보고자 한다.

II. 유전체학(genomics)

1. 유전체(genome)

유전체란 생명체가 가지고 있는 모든 유전적 정보의 총합으로, ‘게놈’이라고도 부른다. 모든 생명체는 DNA를 유전물질로 사용하므로 세포내 DNA 전체, 즉 한 생명체가 가지는 모든 DNA의 염기서열의 총합을 유전체라고 한다. 유전체는 단백질을 암호화하는 유전자 서열(coding sequence)과 이들 유전자 사이에 존재하는 비암호화 서열(non-coding sequence)로 이루어져 있다. 대부분의 비암호화 서열이 반복서열(repetitive sequence, repeats, DNA repeats, repetitive DNA)의 형태로서 존재한다. 유전체는 전사라고 불리는 과정에 의해 유전자로부터 유래된 RNA 분자의 집합인 전사체로 발현된다. 전사체는 이후 번역¹⁾되어 세포의 단백질 집합체인 단백질체로 발현되어 생명의 기초를 이룬다. 즉, 유전체는 생명의 기초를 형성하는 정보를 품고 있는 것이다.

1) mRNA의 염기서열을 3개의 단위로 끊은 트리플렛 코드(Triplet code)로 해석하여, 각 코돈(Codon)마다 지정된 아미노산을 순서대로 종합하는 과정이다.

2. 염색체(chromosome)

염색체는 생물체(엄밀히는 진핵생물²⁾)의 세포분열과정에서만 관찰되며, 세포분열기에 들어가면서 유전물질을 안전하게 보호하고 딸세포들에게 형평성 있게 유전정보를 분배하기위해 염색사가 꼬이고 뭉쳐져 응축되어 X자형태를 이룬 구조물이다. 진핵세포의 유전물질(DNA 또는 RNA)은 히스톤 단백질의 도움을 받아 꼬여 염색사 형태로 핵에 흩어져 있다가, 분열기(M기)에 접어들면 염색사가 응축하여 염색체를 형성한다. 염색시약에 의해 염색된 분열 중인 세포를 현미경으로 보고 발견하여 ‘염색체’라고 이름 붙였다. 생물체들의 염색체는 모양과

2) 진핵세포로 이루어진 생물로, 일반적으로 동물이나 식물 등 대다수의 다세포생물을 포함한다.

크기가 같은 염색체가 보통 2개씩 존재하는데, 이는 어버이로부터 각각 하나씩 받았기 때문이다. 이러한 2개의 염색체를 '상동 염색체'라고 한다.

3. DNA (deoxyribonucleic acid)

DNA는 유전물질(genetic material)을 담고 있는, 세포의 핵 속에서 발견되는 두 종류의 핵산(nucleic acid) 중 하나이다. 핵산은 뉴클레오티드라는 단위체가 반복적으로 되어 있는 중합체이며, 각 뉴클레오티드는 당, 인산 및 질소염기로 구성되어 있다. DNA상의 질소 염기는 4개의 형태(아데닌(A), 구아닌(G), 사이토신(C), 타이민(T))로 구성되어 있으며, 이 염기의 서열이 유전정보를 암호화한다. DNA는 2개의 상보적인 뉴클레오티드 사슬로 구성되어 있다. 생명체의 종류에 따라 예외가 있으나 보통 DNA는 유전정보를 보관 및 보존하는 데 이용되며, RNA는 유전정보를 발현하여 단백질을 만드는 과정 등에서 이용된다.

4. 유전자(gene)

유전자에 대한 정의는 이것을 이용하는 상황에 따라 달라지지만, 통상 하나의 유전적인 특징을 암호화하고 있는 정보의 단위로서 사용된다. 유전자는 구조 단백질이나 효소 같은 것을 만드는 정보와 같이 세포를 만드는 데 필요한 정보를 갖고 있는 DNA의 조각들이다. 유전자는 대립유전자라고 하는 다양한 형태로 나타난다. 하나의 특징을 결정하는 유전자는 대립유전자라고 하는 여러 형태로 존재할 수 있다. 각 생명체가 갖는 대립유전자 세트를 유전자형(genotype)이라 한다.

유전학에서 가장 중요한 개념 중의 하나는 표현형(또는 형질, phenotype or trait)과 유전자형 사이를 구별하는 것이다. 형질은 어떤 특징의 외적인 발현으로 환경과 유전자의 상호작용의 결과이며, 직접적으로 유전되지 않는다. 반면 유전자가 유전되어 환경요인과 함께 형질의 발현을 결정한다. 즉 한 개체가 소유하고 있는 유전정보는 유전자형이며, 형질은 이 유전자형의 표현형이다. 예를 들어 혈액형에서 A형은 표현형이며, A형 항원을 암호화하고 있는 유전정보는 유전자형이다.

5. DNA 염기서열(DNA 시퀀스, DNA sequence)

DNA 염기서열은 DNA의 기본단위 뉴클레오티드의 구성성분 중 하나인 염기 A (adenine), T (thymine), G (guanine), C (cytokine)의 4가지 화학물질들이 나열되어 있는 것을 말하며, 이 4가지의 물질이 어떻게 조합되었느냐에 따라 그 세포의 기능 및 성질이 결정된다. 지구상의 모든 생명체들은 염기서열을 통해 생물의 유전형질을 결정하는 단백질을 지정하게 된다. 즉 핵 속의 유전자, 유전자 속의 염기서열은 세포의 기능과 성질을 규정해 주는 프로그램이다. 인간의 경우가 4가지 종류의 염기가 30억쌍으로 개개인마다 다르게 일정한 순서로 늘어서 있

3) 김용성/배영안, "기능 없는 97% DNA의 비밀 - 에이즈바이러스의 조상은 인간?", 과학동아 2000년 7월호(2000.7), 동아사이언스, 82면.

4) 2002년 햅맵 프로젝트에서 0.3%로 발표되기도 하였다.

5) 오정환/윤병욱, "인간 게놈의 단일염기변형 (Single Nucleotide Polymorphism; SNP)에 대한 이해", 대한구강악안면외과학회지 제 34권 제4호(2008), 대한구강악안면외과학회, 450-451면.

다. 이렇게 알파벳 A, T, G, C로 나열된 염기서열(시퀀스, sequence)을 읽는 것을 염기서열분석(시퀀싱, sequencing)이라고 부른다. 인간의 경우 총 30억쌍의 염기 중에서 9천만쌍(3%)만이 실제로 단백질을 합성하는 '유전자'(엑손)이다. 나머지 97%는 별다른 역할 없이 동일한 염기서열이 반복되는 '반복 DNA'(repetitive DNA)이다. 결국 진핵생물에서 게놈 크기의 폭발적인 증가는 유전자와는 별 상관없는 반복 DNA와 같은 비유전자 DNA (non-coding DNA)가 늘어났기 때문인 것으로 추정된다. 결국 개구리가 인간보다 25배 큰 게놈을 가지고 있다고 해서 인간보다 더 많은 유전자를 가진 것은 아닌 것이다.³⁾

6. 변이(variants, mutation)

DNA를 구성하는 염기 하나, 혹은 다수가 달라지는 것을 변이(variant, mutation)라고 한다. 이러한 변이는 또한 유전되는 선천적 변이와 환경 기타 영향으로 인한 후천적 변이로 나누어지는데, 선천적 변이는 생식세포 돌연변이(germ-line mutation), 후천적 변이는 체세포 변이(somatic mutation)로 구분된다.

인간의 변이 중 대표적인 변이가 단일염기다형성(single nucleotide polymorphism; "SNP")이다. 햅맵 프로젝트(International HapMap Project)로 인해 인간의 유전자가 서로 99.9% 유사하고, 나머지의 0.1%⁴⁾가 특성의 질환에 관한 감수성이나 약제에 대한 부작용 등의 개체 차이에 관여하고 있다고 추정되고 있다. 이 0.1%의 부분에 포함되는 유전자 다형의 대부분을 차지하는 것이 SNP이다.⁵⁾ SNP는 평균 300쌍의 염기에 하나 꼴로 일어나는 염기의 변이 중 그 인구 사회의 1% 이상에서 일어나는 흔한 변이를 가리킨다. 이 흔한 변이가 사람마다 다른 특성을 나타낸다. 예를 들면, 누구는 커피를 마시면 잠을 잘 못 자고, 누구는 술을 마시면 얼굴이 빨개지고, 누구는 태어날 때부터 쌍꺼풀이 있으며, 누구는 특정 질환이 더 잘 생기는 것이다. 약 300쌍의 염기 당 하나의 SNP가 존재하는 것으로 추정되므로 유전체 전체 염기서열 30억 쌍에 존재하는 SNPs는 약 1,000만 개가 있을 것으로 추정되고 있다. 이 중에서 단백질의 생물학적 기능을 변화시킬 가능성이 높은 약 25%의 SNP가 연구 대상으로 주목받고 있으며, 수개에서 수십개의 SNPs는 인간의 형질, 특히 건강에 유의한 영향을 미치는 것으로 알려져 있다. 따라서 30억쌍의 서열을 모두 분석하는 것을 SNPs를 분석하는 것으로 대체할 수 있다. 이것이 DNA 시퀀싱, 나아가 유전체치료의 핵심이다. 또한 이러한 개인별 염기의 차이인 SNP는 개인별로 고유하므로 DNA 지문 등에 활용되어 개인 식별에도 활용되고 있다.

정리하면, 인간의 유전체는 총 22개의 상염색체와 1개의 성염색체로 이루어져 있다. 이를 22권의 전집과 1권의 별책부록의 백과사전으로 비유한다면 한 권의 책이 염색체, 각 페이지는 염색체 위치, 한 문장은 유전자, 한 글자는 염기(base; A, T, G, C)에 해당한다. 즉, 인간은 30억쌍의 글자와 2-3만개의 문장으로 이루어진 23권의 책으로 암호화되어 있는 것이다.

III. 유전체학과 AI

1. 유전체 정보와 유전정보

유전체 정보(genomic data)는 DNA 염기서열분석을 통해 추출된 시퀀싱된(서열화된) DNA와 개별 DNA 변이(variant)를 통칭한다. 시퀀서를 통해 추출된 유전체 정보는 FASTA, FASTQ 포맷 등과 같은 텍스트 기반 포맷으로 저장된 후, 생물정보학(bioinformatics) 파이프라인들을 통해 유전체 변이 정보를 담고 있는 VCF 포맷 파일로 변환된다.

유전정보(genetic data)는 이 시퀀싱된 DNA 중 특히 유전과 관련된 특정 유전자에 대한 정보이다. 다시 말해 생물이 종 보존을 위해 부모로부터 자식에 게로, 또는 세포분열 시 모세포에서 자세포로 자신의 모든 형질을 전달하기 위한 정보를 의미한다. 이렇게 자신이 가진 모든 형질을 다음세대로 전달하는 과정을 유전(heredity)라고 하며, 이때 전달되는 정보가 유전정보이다. 유전정보는 단백질을 합성하는 유전자 정보 외에도 자신의 형질을 다음세대에 전달하기 위한 정보의 해독 및 발현 등의 모든 과정에 필요한 정보를 포함한다. 유전체 정보는 이러한 유전정보를 포함하는 넓은 개념으로 연구 실무와 달리 현행법상 이들을 엄밀하게 구분하고 있지는 않다.

2. 유전체 정보 분석

가. 전(前)염기서열분석(pre-sequencing)

분석을 위한 표본 수집 및 전처리(preprocessing) 과정이다.

나. 염기서열분석(sequencing)

DNA 시퀀싱이란 DNA 분자들을 구성하고 있는 A, T, G, C 뉴클레오타이드의 순서를 결정하는 것을 말한다. DNA 염기서열 분석기(시퀀서, DNA Sequencer)를 이용하면 표본에서 DNA 염기서열 자료(리드, reads)를 읽어낼 수 있다. DNA 염기서열을 눈으로 읽을 수는 없기 때문에 기계장치의 힘을 빌리는데 이렇게 DNA 샘플에서 서열을 읽어내어 사람이 읽을 수 있는, 또는 컴퓨터 전용 프로그램이 분석할 수 있는 알파벳 서열로 출력해주는 장비를 시퀀서라고 한다. 수십 년 동안 생거 시퀀싱이란 방법을 사용하는 시퀀서들이 활약했었는데(캐필러리 시퀀서), 최근 몇 년 전부터 생거 시퀀싱이 아닌 새로운 방법을 사용하는 시퀀서들이 등장해서 함께 활약하기 시작했다. 이렇게 새로 등장한 시퀀서들은 차세대 시퀀서(Next Generation Sequencer; “NGS”)라고 구분한다. 경우에 따라서 제2세대 시퀀서라고도 하는데 그 이유는 또 다른 방식의 시퀀서, 즉 제3세대, 제4세대 시퀀서라고 할 수 있는 것들이 논의되거나 등장하고 있기 때문이다.

다. 데이터 처리(data processing)

생물정보학 파이프라인들이 염기서열을 재구성하고 유전체의 변이(variants)를 식별하는데 사용된다.

라. 분석과 해석(analysis and interpretation)

머신러닝과 딥러닝 기술의 적용이 가장 활발한 분야이다. 표현형분석(phenotyping), DNA 시퀀싱, GGAS (Genome-Wide Association Study) 등의 유전체 정보 분석 기술에 딥러닝 기술이 결합되고 있다.

3. AI 기술 적용사례

유전체 자료 분석을 위한 AI 기술은 주로 SNP 식별단계와 분석단계에서 사용되고 있으며, 특히 SNP 식별단계의 연구가 늘어나고 있다. 기존에는 최장공통부문서열(LCS)과 같은 서열비교 알고리즘이 주로 사용되었으나, 최근 SNP를 빠르고 정확하게 찾는 대회에서 우승한 팀이 이미지 분석에 자주 사용되는 합성곱신경망(CNN) 알고리즘을 사용하여 주목받기도 하였다.

CNN 알고리즘의 작동원리는 다음과 같다. 우선 참조값(reference) 행에 참조하고자 한 기존의 유전체 정보를 넣고, 시퀀싱을 통해 얻은 염기서열을 위치에 맞춰 데이터를 만든다. CNN 알고리즘의 특징은 이러한 테이블을 하나의 이미지로 만들어, ATGC 각각의 이미지색상을 RGB값으로 저장하여 분석하는 것이다. 즉 염기서열 텍스트 대신 이미지 자체로 분석을 하는 것이다. 구체적으로 CNN 알고리즘은 일반적인 인공신경망처럼 역전파 알고리즘을 이용하여 필터를 학습한다. 필터는 그 특징이 데이터에 있는지 없는지를 검출해주는 함수로, 보통 3*3, 4*4 행렬로 정의되며 입력 받은 데이터에서 그 특성을 가지고 있으면 결과 값이 크게 나오고, 특성을 가지고 있지 않으면 0에 가까운 값이 나오게 돼서 데이터가 그 특성을 가지고 있는지 없는지 여부를 알 수 있게 해준다. 필터를 적용해서 얻어낸 결과는 정량적인 값으로 나오기 때문에 그 특징이 “있다 없다”의 비선형 값으로 바뀌주는 과정이 필요하다. 이 과정을 담당하는 것이 활성화(activation) 함수이며, 주로 레일루(ReLU) 함수가 사용된다. 역전파 알고리즘을 통해 적합된 모델에 탐색하고자 하는 시퀀싱 데이터를 입력하면 해당하는 변이 부분마다 평가하여 참조값과 비교한 결과값을 알려준다.

IV. 의료에의 적용

현대 의료를 넘어 다가올 미래 의료를 4P 의료시대라 부른다. 이는 맞춤형/정밀의료(Personalized/Precision Medicine), 예방의료(Preventive Medicine), 예측의료(Predictive Medicine), 참여의료(Participatory Medicine)를 지칭하며, 시스템 생물학의 선구자인 리로이 후드(Leroy Hood) 등이 2000년대 중반에 처음 소개한 개념이다. 당시에는 추상적인 개념이었으나 AI 기술이 발전하면서

4P 의료시대가 현실화되고 있다. 맞춤형료⁶⁾ 나 정밀의료⁷⁾가 발전하고 있을 뿐 아니라, 유전자 분석과 관련하여 전장유전체 연관분석(Genome-Wide Association Study; “GWAS”)에 딥러닝을 접목하는 방식으로 발전하고 있다.⁸⁾ GWAS는 특정 질병군과 건강한 대조군의 전장유전체(모든 유전체의 위치)를 분석한 후 의미 있는 유전자 변이를 역으로 찾아내는 방식(case v. control)으로서, (종래의 방식인 후보유전자 연관분석(Candidate Gene Association Study; “CGAS”)과 달리) 특정 질병과 관련된 유전적 변이를 최대한 간단하게 특정하는 장점이 있다.⁹⁾ 대규모 플랫폼의 개발(DNA 칩, Roche사의 454, Sequenom사의 iPLEX, Illumina사의 Solexa 등)로 수백만 개의 SNP를 동시에 분석하는 것이 비교적 저렴해졌기 때문에 가능해진 방식이나, 최초 연구된 코호트 외에 다른 대상으로 질병-대조군 연구를 했을 때는 재현성이 떨어지는 문제가 있고, 암이나 심혈관 질환 같은 다중 인자가 원인이 되는 경우에 소수 유전자만으로 질병을 예측하는 것은 논리가 약하다는 비판이 존재한다.¹⁰⁾ 이에 딥러닝 기술을 기반으로 수백, 수천개의 유전자 마커를 대상으로 하는 질병예측 모형이 등장하고 있다.¹¹⁾ 딥러닝 기술을 통해 재현성 뿐만 아니라 모델의 정확성이 증가하였다. 질병과 형질에 영향을 미치는 변이는 극히 일부분이고, 대부분의 변이는 유해하지 않기 때문에 두 변이 집단간에 양적 불균형이 존재한다. 이를 해결하기 위한 딥러닝 기반의 새로운 방법론인 hyper-SMURF가 등장하는 등 새로운 모델들이 등장하면서 모델의 성능평가값인 Area Under the Precision-Recall Curve (AUPRC) 값이 지속적으로 개선되고 있다.¹²⁾

V. 유전체 정보와 관련된 법적 쟁점

유전체 정보에 적용되는 법으로는 「개인정보 보호법」, 「생명윤리 및 안전에 관한 법률」(이하「생명윤리법」), 「의료법」 등이 있다. 이 중 2020. 8.부터 시행된 개정 개인정보 보호법과 생명윤리법을 중심으로 관련 법적 쟁점을 검토하고자 한다. 이에 앞서 프라이버시와 관련된 유전체 정보의 특성과 EU 일반정보보호법(General Data Protection Regulation; “GDPR”)을 검토하고자 한다.

1. 유전체 정보의 특성

프라이버시와 관련하여 유전정보의 예외성(genetic exceptionalism)을 바탕으로 유전체 정보가 다른 종류의 자료와 본질적으로 다르다는 견해와 유전체 정보는 전통적인 일반건강정보와 큰 차이가 없다는 견해가 대립하고 있다.¹³⁾ 유전정보의 예외성을 지지하는 견해는 일반적인 의료정보에 비해 유전체 정보가 가지는 특성을 강조하고, 그에 대한 특별한 보호를 주장한다. 유전체 정보는 예측성, 특정성, 불변성, 가계성 등의 특성으로 인해 전통적인 일반건강정보와 상이하다는 것이다.¹⁴⁾ 나아가 우생학의 이름 아래 저질러진 인종차별의 전례가 반복될 것을 우려하여, 유전자 차별의 문제나 유전자 프라이버시 보호의 문제에 대처하기 위하여 기존의 정보보호법제의 틀에 만족하지 아니하고 유전자정보에 대하

6) 유전체에 따른 개인의 차이를 임상에 적용. 와 파린이라는 항응고제의 농도를 VKORC1, CYP2C9라는 효소의 유전적 변이에 관한 개인차에 따라 결정하는 것이 대표적이다.

7) 유전체 정보 외 단백질체(Proteomics), 전사체(Transcriptomics), 대사체(Metabolomics) 등 각종 인체유래물에서 얻은 정보(토탈오믹스, Totalomics)를 바탕으로 진단과 치료를 개선한다.

8) 김경철, “유전체, 다가온 미래 의학”, 미디어이트뉴스, 2020, 72-74면.

9) 김경철, 앞의 책, 67면.

10) 김경철, 앞의 책, 68면.

11) Arloth J, Eraslan G, Andlauer TFM, Martins J, Iurato S, Kühnel B, et al. (2020). DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning. PLoS Comput Biol 16(2), 1-28.

12) Schubach, M., Re, M., Robinson, P. N., & Valentini, G. (2017). Imbalance-Aware Machine Learning for Predicting Rare and Common Disease-Associated Non-Coding Variants. Scientific reports, 7(1), 1-12.

13) McGuire AL et al(2008), 이상용(2002) 등은 유전정보의 예외성을 지지하고 있으며 A PHG Foundation(2020: 2017)과 Lawrence O et al(1999)는 이를 반박하는 입장이다.

14) 이상용, “유전자정보의 보호와 이용통제에 관한 연구”, 형사정책연구원 연구총서 02-21(2002.12), 한국형사정책연구원, 26면에 서재인용.

15) 정성기, “유전자정보에 대한 법적 보호의 특수성”, 과학기술법연구 제6권 제1호(2000), 한남대학교 과학기술법연구원, 194면.

여 특별한 법적 보호를 주장하거나 이른바 유전자기본권의 문제를 제기한다.¹⁵⁾ 유전정보의 예외성에 반대하는 견해는 기존의 일반건강정보도 유전체 정보와 마찬가지로 예측성, 특정성, 불변성, 가계성 등의 특징을 지니고 있고 그 차이는 정도의 문제에 지나지 않는다고 주장한다. A PHG Foundation (2017) 보고서는 개인정보 보호법과 정책 수립에 영향을 미치는 유전체 정보의 특성으로 ① 강한 식별성(strongly identifying), ② 가계성(familial), ③ 정태성(static), ④ 특수성에 대한 오인(commonly understood to be special) 등을 제시하며 유전정보의 예외성을 부정하였다. 주목할 특성은 강한 식별성으로, 유전체 정보가 다른 정보와 결합될 경우 강한 식별성을 나타내지만 유전정보 단독으로는 특정 개인을 식별하는 것이 불가능하다고 한다. 따라서 유전체 정보를 다른 개인정보와 달리 취급할 이유가 없다는 것이다. 오히려 문제는 유전체 정보를 다른 개인정보와 마찬가지로 익명화(anonymization)하는 데에 현실적인 제약이 존재한다는 것이다. 이에 관하여 너필드생명윤리위원회(Nuffield Council on Bioethics)는 “비가역적인 익명화는 현실적으로 불가능하다”고 선언하기도 하였다.

A PHG Foundation(2017)은 유전체 정보 보호의 어려움으로 ① 동의의 어려움, ② 비식별화의 어려움, ③ 접근 보안의 어려움, ④ 내재적으로 식별가능(intrinsically identifying)하다는 오해로 인한 어려움 4가지를 제시하였다.

첫째, 유전체 정보는 정보제공자의 동의를 통한 처리가 어렵다. 유전체 정보의 가계성으로 인해 동의가 필요한 정보제공자의 범위가 모호하고 대규모 연구의 경우 현실적으로 수많은 사람의 구체적인 동의를 얻는 것이 어렵기 때문이다. 따라서 유전체 정보 보호를 위해 주로 비식별화 기술이 사용된다.

둘째, 자료의 비식별화가 어렵다. 자료의 비식별화는 자료의 가공(manipulation)과 접근 제한(restricting access)으로 구분되는데, 그 중 자료의 가공이 일반적으로 사용된다. 하지만 유전체 정보는 다른 개인정보와 쉽게 결합하기 때문에 몇 개의 식별자를 삭제하는 수준으로는 비식별화가 불가능하다. 이에 자료를 지나치게 가공하면 오히려 자료가 쓸모 없어진다.

셋째, 비식별화 방법으로 데이터에 접근을 제한하는 것도 어렵다. 유전체 정보는 개인의 일생에 걸친 자료일 뿐만 아니라 세대를 이어지는 자료로서 수명이 길기 때문에 접근 제한은 한계를 지닌다. 유전체 정보를 오랫동안 보관하면서 기존 보안기술이 낙후되기 때문이다. 또한 접근 제한 역시 보안이 엄격해질수록 자료의 유용성이 크게 떨어진다.

넷째, 유전체 정보가 내재적 식별성을 지녀서 그 자체로서 특정인을 식별할 수 있다는 오해이다. 유전정보의 예외성과 유전체 정보의 내재적 식별성을 부정하는 견해에 의하면 이는 사실이 아니다. 식별화(identification)와 개별화(individuation)는 별개의 개념이다. 개별화는 오직 어떤 개인에게만 속하는 특성을 지칭하는데, 유전체 정보는 개별화 특성과 강한 식별성을 모두 지니기 때문에 내재적 식별성을 지닌다고 오해되고 있다. 자료의 식별성은 자료의 내재적 요인에만 기인하는 것이 아니라 다른 자료와의 연결성에 기인하는 것이다.

2. 유전체 정보와 GDPR

가. GDPR

EU는 2016. 5. 27. GDPR을 채택하였고, 동법은 2년의 유예기간을 거쳐 2018. 5. 25.부터 시행되고 있다. GDPR은 EU의 개인정보 보호법령으로 본문 11장(Chapter), 173개 전문(Recital), 99개 본문(Article)으로 구성되어 있으며, 구속력 있는 법률(regulation)로서 EU 모든 회원국에 직접적으로 적용된다. 기존 개인정보보호 지침이었던 정보보호지침(Data Protection Directive, 95/46 EC; “DPD”)에 개인정보책임자 지정, 영향평가 등을 추가하여 기업의 책임을 강화하였고, 정보이동권, 처리제한권 등 새로운 권리를 추가하고 반대권, 삭제권 등을 강화하여 정보주체의 권리를 강화하였다. 동법 적용대상은 EU에 사업장을 운영하는 기업, EU에 거주하는 주민에게 물품·서비스를 제공하는 기업, EU에 거주하는 주민의 행동을 모니터링하는 기업 중 하나에 해당하는 기업으로써 EU 주민의 개인정보를 처리하는 경우 한국 기업도 적용 대상이 된다. 중대한 위반의 경우 2천만 유로 또는 직전 회계연도의 글로벌 총 매출의 4%에 해당하는 금액을 과징금(administrative fine)으로 부과할 수 있고, 일반적 위반의 경우 1천만 유로 또는 직전 회계연도의 글로벌 총 매출의 2%에 해당하는 금액 가운데 높은 금액을 과징금으로 부과할 수 있으며(Article 83), 벌금 외에도 경고(warning), 견책(reprimand), 개인정보 처리정지(suspension) 등의 제재가 가능하여 강력한 제재를 채택하고 있다. EU내 각국은 GDPR를 준수하면서 입법재량조항에 대해서는 자국 사정에 맞게 대응하고 있다.

나. GDPR상 유전체 정보의 규제

GDPR 4(13)의 유전정보(genetic data)는 “자연인의 생리나 건강에 관해 고유한 정보를 제공하는 해당인의 선천적 또는 후천적인 유전자 특성과 관련한 개인정보로서, 특히 해당 자연인의 생물학적 샘플 분석을 통해 획득하게 된다”라고 정의되어 있는데, 유전체 정보(genomic information) 중 가족 유전체 정보(familial genomic information)와 개인을 식별시키지 않는 유전체 정보를 제외한 정보를 의미한다. 이러한 유전정보는 GDPR 4(15)의 건강정보(data concerning health)에 속하고, 더 넓게는 GDPR 4(1)의 개인정보(personal data)에 해당한다.

유전체 정보에 대한 별도의 정의는 되어있지 않으나, 유전정보를 넓은 의미에서 정의하여 그 보호의 범위를 넓히고 있다. 또한 제9조제1항에서 한국 개인정보 보호법의 “민감정보”에 해당하는 “특별 범주의 개인정보”(special categories of personal data)에 유전정보(genetic data), 고유식별정보인 생체인증정보(biometric data for the purpose of uniquely identifying a natural person), 건강정보를 포함시키고, 그 처리를 원칙적으로 금지하여 두텁게 보호하고 있다.

16) PHG Foundation, “The GDPR and genomic data”, 2020. 5.

17) 이원복, “유전체 연구와 개정 개인정보 보호법의 가명처리 제도”, 이화여자대학교 법학논집 제25권 제1호 통권 71호(2020.9), 이화여자대학교 법학연구소, 202년.

하지만 예외적으로 제9조제2항에 해당하는 경우에 한하여 민감정보를 처리할 수 있으며, 민감정보 중에서 유전정보, 생체정보 또는 건강에 관한 정보는 제9조4항에 따라 회원국 내에서 데이터 활용 제한을 위해 추가 조건을 유지하거나 도입할 수 있다. 유전체 정보에 대해서는 제9조 제2항의 조건 중 정보주체의 명시적 동의(a), 의료 목적(h), 공중보건 목적(i), 과학적 연구 목적(j), 명백히 일반에게 공개된 정보(e), 중대한 공익 목적(g) 등이 관련이 있다.¹⁶⁾ 특히 과학적 연구 목적(j)과 관련하여서는 제89조 제1항에 따라 각국이 이러한 원칙에 대한 예외를 개별 입법을 통하여 둘 수 있는데, 그 경우 민감정보의 처리가 추구하는 목적에 비례하여 이루어져야 하고, 민감정보를 보호하는 근본원칙을 존중해야 하며, 정보주체의 기본권과 이익을 보호할 수 있는 적절하고도 구체적인 안정장치를 두어야 한다. 민감정보의 과학적 연구 목적 이용 시 인정되는 예외 근거를 이처럼 민감정보 처리에 관한 조문 내부에 두고 있으므로, 후술한 국내 개인정보 보호법과 달리 논란이 생길 여지가 없다.¹⁷⁾

3. 유전체 정보와 개인정보 보호법

가. 개인정보와 유전체 정보

한국의 개인정보 보호법은 제2조제1호에서 “개인정보”를 살아 있는 개인에 관한 정보로서 직접식별자, 간접식별자 및 가명정보로 정의하고 있다. 유전체 정보의 경우 「개인정보 보호법」 제2조제1호 나목의 “해당 정보만으로는 특정 개인을 알아볼 수 없더라도 다른 정보와 쉽게 결합하여 알아볼 수 있는 정보”에 해당하는지 명확하지 않다. 유전체 정보는 그 특수성으로 인해 개개인과 1대1로 대응하지만 그것이 필연적으로 개인식별을 의미하는 것은 아니며, 성명, 주민등록번호 등의 식별자 없이 유전체 정보 자체만으로 개인을 직접 식별할 수는 없기 때문이다. 이에 개인식별자와 유전체 정보가 매칭된 자료에 쉽게 접근할 수 없는 경우, 유전체 정보가 개인정보 보호법상 “개인정보”에 해당되는지 문제된다. A PHG Foundation (2017) 보고서 등이 주장하듯 유전체 정보는 그 자체로는 특정인을 식별할 수 없지만, 다른 정보와 결합하면 강한 식별성을 지니게 된다. 유전체 정보를 추출할 때 개인식별자와 매칭된 자료로 추출되어 연구진들에게만 접근 가능하다고 하여도, 유전체 정보가 애초에 연구목적으로 사용된다는 점을 고려하였을 때 “쉽게 결합할 수 있”다고 보는 것이 합리적이다. 이에 유전체 정보는 개인정보 보호법상 “개인정보”에 해당함이 타당하다.

한편 유전정보와 관련해서 한국의 동법 시행령 제18조제1호는 동법 제23조제1항의 “민감정보”에 속하는 “대통령령으로 정하는 정보”로 “유전자검사 등의 결과로 얻어진 유전정보”를 민감정보로 분류하며 GDPR과 마찬가지로 유전정보를 두텁게 보호하고 있다.

나. 민감정보 처리 규제

검토한 바와 같이 유전정보는 개인정보 보호법상 개인정보 중 민감정보에 해당하여 동법 규제의 대상이다. 이에 유전체 정보는 원칙적으로 처리가 금지되어 있으나, 동법 제23조제1항 각 호에 해당하는 경우 예외적으로 처리가 허용된다. 그러나 각 호에 따라 민감정보를 처리하는 경우에도 동조 제2항에 의해 그 민감정보가 분실·도난·유출·위조·변조 또는 훼손되지 아니하도록 동법 제29조에 따른 안전성 확보에 필요한 조치를 하여야 한다. 또한, 일정한 예외사유에 해당하는 경우를 제외하면, 동법 제23조제1항제1호에 따른 법정고지사항을 정보주체에게 알리고 다른 개인정보의 처리에 대한 동의와 별도로 동의를 받아야 한다.

개인정보보호위원회가 고시한 「개인정보의 안전성 확보조치기준」 제7조제2항, 제5항 및 정보통신서비스 제공자등에 적용되는 특칙인 「개인정보의 기술적·관리적 보호조치 기준」 제6조 제2항 제7호에 따라 바이오정보¹⁸⁾는 안전한 알고리즘으로 암호화하여 저장하여야 한다. 구 행정자치부(現 행정안전부)와 한국인터넷진흥원의 「개인정보의 암호화 조치 안내서」에 의하면 “암호화”란 일상적인 문자로 쓰이는 평문을 암호키를 소유하지 않은 사람이 알아볼 수 없도록 기호 또는 다른 문자 등의 암호문으로 변환하는 것을 말한다. 개인정보가 비인가자에게 유출 또는 노출되더라도 그 내용을 확인할 수 없거나 어렵게 하는 보안기술이다. 암호화의 종류에는 대칭키 암호화, 공개키 암호화, 일방향(해쉬함수) 암호화가 있다. 한편 “안전한 암호 알고리즘”이란 국내 및 미국, 일본, 유럽 등의 국외 암호 연구 관련 기관에서 권고하는 암호 알고리즘을 의미한다. 그 예로 대칭키 암호 알고리즘, 공개키 암호 알고리즘, 일방향 암호 알고리즘 등이 있으며, 세부 사항은 「개인정보의 암호화 조치 안내서」 참고자료에 소개되어 있다. 문제는 이 같은 개인정보 보호법상 민감정보 처리에 대한 규정들이 지나치게 엄격하여 연구 등 공익 목적에 필요한 정보의 활용을 저해한다는 것이다. 민감정보 처리가 가능한 예외적인 경우(법령에서 요구하거나 허용하는 경우, 혹은 정보주체의 동의를 받는 경우)가 존재하나 전자는 그 범위가 협소하고, 후자는 유전체 정보의 특성상 정보의 활용가능성이 떨어진다. 유전체 정보의 분석을 위해 필요한 자료의 규모가 방대하고, 유전체 정보의 가계성으로 인해 동의의 범위가 모호하기 때문에 수많은 정보주체의 구체적인 동의를 구하는 것이 현실적으로 불가능하기 때문이다. 또한 민감정보의 처리가 가능하여도 정보처리와 관련된 관리계획을 수립하여야 하고, 암호화 설비를 갖추어야 하는 등 현실적인 비용문제가 존재한다.

이에 2020. 8. 시행된 개정 개인정보 보호법은 가명정보의 처리에 관한 특례를 신설하여 통계작성, 과학적 연구, 공익적 기록보존 등을 위하여 정보주체의 동의 없이 가명정보를 처리할 수 있도록 하였다. 해외에서는 이미 유전체 정보 처리를 위하여 주로 가명화(pseudonymization) 기술이 사용되고 있다.¹⁹⁾ 이에 유전체 정보의 가명화 방안이 문제된다.

18)

“바이오정보”란 지문, 얼굴, 홍채, 정맥, 음성, 필적 등 개인을 식별할 수 있는 신체적 또는 행동적 특징에 관한 정보로서 그로부터 가공되거나 생성된 정보를 포함하며 유전정보를 포함하는 것으로 해석한다(개인정보의 안전성 확보조치기준 제2조 제16호, 개인정보의 기술적·관리적 보호조치 기준 제2조 제8호).

20)

이원복, 앞의 논문(주17), 213면.

19)

PHG Foundation, “Identification and genomic data”, 2017. 12.

21)

이원복, 앞의 논문(주17), 213면.

다. 유전체 정보의 가명화 방안

1) 기술적 논의

개정 개인정보 보호법상 개인정보의 “가명처리”는 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가정보가 없이는 특정 개인을 알아볼 수 없도록 처리하는 것으로 정의된다. 개인에 관한 정보 가운데는 성명, 주소, 전화번호, 주민등록번호 등 특정 정보주체로 바로 연결 지을 수 있는 직접식별자(direct identifiers)가 있고, 성별이나 국적과 같이 개인에 관한 정보이기는 하나 특정 정보주체와 바로 연결을 지을 수 없는 간접식별자(indirect identifiers)가 있다. 그 가운데 직접 식별자를 모두 삭제하거나 그 자체로 특정 정보주체를 알아볼 수 없는 랜덤 문자로 대체하는 방식의 비식별 처리를 가명화(pseudonymization)라고 부르는데, 개정 개인정보 보호법은 바로 이 가명화라는 개념을 입법화하고 있다.

가명처리의 대상이 되는 유전체 정보는 이름이나 주소와 같은 직접식별자와, 연령, 질병명 등 간접식별자, 유전체 정보 그 자체로 이루어진 자료 테이블이다. 이러한 유전체 정보의 가명처리 방안에 대하여 국내의 일원화된 논의는 존재하지 않으나, 일부에서 의견을 제시하고 있다.²⁰⁾ 이 견해에 의하면 유전체 정보는 가공을 거치면 연구 자료로서의 가치가 크게 떨어지기 때문에 그대로 유지하면서 연구에 이용할 수 있도록 하고, 직접식별자를 전면적으로 가명처리하고 간접식별자의 경우 재식별의 위험이 존재하므로 연구에 필요한 범위를 제외하고는 삭제를 하거나 비식별화 작업을 취하여야 한다. 또한 유전체 정보 그 자체도 다형성(polymorphism)을 보이는 구간이라서 잠재적으로 식별력은 높지만 연구 목적과는 무관할 것으로 예상되는 염기서열 부분은 마치 흑색 마커로 덧칠하는 것처럼 “X”와 같은 무의미한 문자로 “암막처리(redaction)” 할 것을 제안하였다. 그리고 장기적으로는 유전체 정보 자체를 염기서열의 형태로 제공하는 것이 아니라 암호화하는 방식으로 염기서열에 대한 분석만 허용하는 기술적 방식의 도입을 주장하였다.

2) 법적 논의

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, 대체 값(가명 자체가 원래의 정보 주체와 1:1로 고유하게 대응하는 값)을 이용하는 것이므로 가명과 정보주체를 연결해주는 복호화 함수 또는 매핑 테이블만 확보한다면 정보주체를 역추적하는 것이 얼마든지 가능하다. 그렇기 때문에 가명처리를 거친 가명정보는 정의상으로는 여전히 “개인정보”의 한 종류를 이루지만, 개정 개인정보 보호법은 가명정보에 대하여는 제3장의 일부로 신설된 “제3절 가명정보의 처리에 관한 특례” 섹션에서 다양한 특례를 인정하고 있다.²¹⁾

기존 개인정보 보호법 체제하에서의 비식별화는 개인을 알아볼 수 없도록 해야 했고, 그러려면 유전체 정보의 염기서열을 그대로 존치할 경우 개인을 알아볼 수 없는 것이나 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 있었다. 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.

그러나 2020. 9. 보건복지부와 개인정보위원회에서 공동 발행한 “보건의료 데이터 활용 가이드라인”에 따르면, 유전체 정보는 그 안에 담긴 정보의 내용을 모두 해석해내지 못하고 있고, 부모·조상·친척 등의 제3자 정보를 담고 있을 수 있으므로 적절한 가명처리 방법이 개발될 때까지는 가명처리 가능여부를 유보하며, 예외적 경우를 제외하고는 본인 동의 기반으로만 사용할 수 있다고 한다.²²⁾ 해당 부분을 전제하면 다음과 같다.

「보건의료 데이터 활용 가이드라인」상 예외적 경우

- ① 널리 알려진 질병에 관한 유전자 변이 유무 또는 변이 유형 :
구체적 변이정보(예: Loci)가 아닌 큰 단위의 유전자(gene)
단위 정보로 제공함으로써 개인 재식별 가능성을 크게 낮춤
- ② 생식세포 변이 정보를 제거한 신생물 고유(neoplasm)의
신규변이 정보 : 생식세포 변이(정상조직 변이)를 제거한
신규 생성 변이 정보는 암의 원인이 되는 변이 정보만
포함하게 되므로 개인 식별 가능성 없음

동 가이드라인은 유전체 정보 자체에 대한 가명처리가 필요하다는 입장에 선 것으로 읽히는데, 이는 개인정보 보호법의 정의 규정에 엄밀하게 부합하는지도 의문이 있다. 개인정보 보호법에서 “가명처리”란 개인정보의 일부를 삭제하거나 일부 또는 전부를 대체하는 등의 방법으로 추가 정보가 없이는 특정 개인을

22) 개인정보 보호위원회·보건복지부(2020), “보건의료 데이터 활용 가이드라인”

23) 이원복, 앞의 논문(주17), 216면.

24) 이원복, 앞의 논문(주17), 216면.

알아볼 수 없도록 처리하는 것인데, 유전체 정보는 고유식별성을 지니지 않는다는 견해에 의하면 대조할 수 있는 유전체 데이터베이스가 없이는 그 자체로 특정 개인을 알아볼 수는 없으므로, 유전체 정보 자체를 일부 변형시키는 방식을 취하지 않더라도 추가정보 없이 특정 개인을 알아볼 수 없다는 요건을 충족할 수 있을 것이다. 따라서 “적절한 가명처리 방법이 개발될 때까지” 유보한다는 것은 적절한 암호화 방식이 개발될 때까지는 유전체 정보를 가명처리를 통하여 특칙을 적용받는 것은 불가능하다는 입장인 것으로 보인다.²³⁾

생각건대 유전체 정보에 대하여 앞서 검토한 바와 같이 직접식별자는 그 전부를, 간접식별자는 연구에 필요한 범위를 제외하고는 삭제를 하거나 비식별화 작업을 취한다면 가명처리된 유전체 정보는 특칙의 적용대상으로 고려되어야 한다. 유전정보의 예외성과 유전체 정보의 고유식별성을 부정하는 견해에 의하면 유전체 정보가 다른 건강정보에 비해 별개로 취급되어야 할 합리적인 이유가 크게 존재하지 않는다. 최소한 과학적 연구의 경우라도 가명처리 특칙을 인정하는 것이 일견 타당해 보인다.²⁴⁾

4. 유전정보와 생명윤리법

가. 생명윤리법상 유전정보의 개념

생명윤리법은 유전체 정보를 별도로 규정하는 대신 동법 제2조 제14호에서 “유전정보”란 인체유래물을 분석하여 얻은 개인의 유전적 특징에 관한 정보라고 정의하고 있다. 또한 동조 제11호는 “인체유래물”(人體由來物)이란 인체로부터 수집하거나 채취한 조직·세포·혈액·체액 등 인체 구성물 또는 이들로부터 분리된 혈청, 혈장, 염색체, DNA (deoxyribonucleic acid), RNA (ribonucleic acid), 단백질 등으로 정의하고 있다. 생명윤리법상 유전정보는 GDPR의 유전정보(genetic data)에 대응하고, 인체유래물은 유전체 정보(genomic data)를 포함하는 개념으로 해석된다.

나. 생명윤리법상 유전정보의 보호

생명윤리법은 인간대상연구 및 연구대상자 보호를 다루고 있는 제3장에서 유전정보의 보호를 규정하고 있다. 동법 제16조는 인간대상연구의 동의를 규정하고 있으며, 구체적으로 제1항과 제2항은 연구대상자로부터 각 호의 사항이 포함된 서면동의(전자문서를 포함한다. 이와 같다) 의무를, 제3항은 서면동의 면제사유를, 제4항은 연구자의 연구대상자에 대한 설명의무를 규정하고 있다. 동법 제18조는 개인정보의 제3자 제공을 규정하고 있다. 동조 제1항은 기관위원회의 심의를 규정하고 있으며, 제2항은 연구대상자의 동의 없이는 개인정보의 익명화를 의무화하고 있다

한편 동법 제4조는 생명윤리 및 안전에 관하여는 다른 법률에 특별한 규정이 있는 경우를 제외하고는 이 법에 따른다고 규정하여, 개정 개인정보 보호법

의 가명정보의 처리에 관한 특례가 유전정보에도 적용되는지 문제가 되었다. 이에 2020년 8월 보건복지부 생명윤리과는 유권해석으로 개정 개인정보 보호법의 “가명처리”는 생명윤리법의 “익명화”에 포함되는 것으로 해결하였다. 또한 의료기관에서 진료목적으로 수집된 의료데이터 등을 개인정보 보호법 상의 가명처리를 통해 연구목적 등으로 이용하려는 경우 “연구대상자등에 대한 기존의 자료나 문서를 이용하는 연구”로 간주하고 기관 차원에서 가명처리가 확인된 경우 IRB 심의 및 동의를 면제할 수 있다고 해석하였다. 그러나 IRB 심의 및 동의 면제의 경우, “연구대상자등에 대한 기존의 자료나 문서를 이용하는 연구”는 기존 생명윤리법 시행규칙 제13조제1항제3호에도 규정되어 있던 내용으로 개정 개인정보 보호법의 가명처리 특례에도 불구하고 실무상 IRB에 심의 의뢰 없이 가명정보를 처리하는 것은 어려울 것으로 생각된다.

다. 생명윤리법과 차별 금지

차세대 시퀀서들이 등장하여 유전체 정보 추출이 저렴해지고 빨라지면 서 유전체 정보에 대한 접근성은 더욱 높아지고 있다. 이에 일각에서는 유전체 정보에 의한 차별에 대한 우려가 제기되었고, 실제로 보험회사에서 가입심사 시 유전체 정보를 이용하기 시작하였다.²⁵⁾ 이에 세계 각국이 유전체 정보에 의한 차별을 금지하는 법률을 제정하였고, 한국도 2012년 생명윤리법 전부개정시 제46조를 신설하여 유전정보에 의한 차별을 금지하고 있다.

VI. 결론

유전체 정보는 DNA 시퀀싱을 통해 유전체에서 추출한 DNA 서열(A, T, G, C 조합) 자료이다. 유전체 정보는 강한 식별성, 가계성, 정태성 등의 특성을 지니고 있지만 그 자체로는 특정인을 식별할 수 없어 고유 식별성을 지닌다고 할 수 없다. 하지만 표현형 자료나 각종 개인신상 자료와 결합하여 쉽게 특정인을 식별할 수 있기 때문에 법적 규제가 필수적이다. 현재 국내 개인정보 보호법, 생명윤리법, 그리고 의료법에서 이를 규율하고 있다.

미래의학은 개인의 유전체 정보를 바탕으로 하여 더욱 정밀하고 개인지향적으로 발전할 것이다. 유전체 정보를 활용한 치료는 공상영화에나 나올 법한 얘기에서 어느새 진지한 논의의 대상이 되었다. 인간의 존엄성과 직접 관련된 민감한 영역인 만큼, 학계 및 일반 대중의 논의가 반드시 선행되어야 할 것이다. 특히 AI 기술의 발전으로 관련 산업의 발전 속도가 그 어느 때보다 빨라진 지금, 법이 기술의 발전속도에 맞추어 나가는 것은 더욱 어려워졌다. 다행히 2018년 시행된 EU의 GDPR에 영향을 받아 한국도 2020년 개인정보 보호법이 개정되고 관련 논문이 게재되는 등 유전체 정보에 둘러싼 법적 논의가 본격적으로 시작되고 있는 모습이다. 기존 개인정보 보호법은 개인정보의 활용을 지나치게 제한하고 있

다는 비판을 받았고, 이에 개정 개인정보 보호법은 가명화 처리 특례를 신설하여 연구 목적의 활용에 대한 제한을 완화하였다. 그러나 검토하였듯이 “보건의료 데이터 활용 가이드라인”은 여전히 유전체 정보의 활용에 대해 보수적으로 접근하고 있는 것으로 보인다. 조금 더 다양한 논의를 통해 합리적인 방향으로 발전할 수 있기를 기대한다.

25)

최창희, “유전자 정보 차별 금지법 도입 검토 필요”, KIRI 리포트 제446호(2018.06), 보험연구원, 31-32면.