Journal of Biotechnology and Bioindustry
Vol.8, December, 2020. 38~47
https://doi.org/10.37503/jbb.2020.8.38

간추린 세포 내 생명 정보 해석

조 성 국*

충청북도 증평군 대학로 61 한국교통대학교 보건생명대학 식품생명학부 생명공학전공 27909

Interpretation of Biological Data at a Glance

Sung-Gook Cho*

Department of Biotechnology, Korea National University of Transportation, Jeungpyeong 27909, Korea

ABSTRACT

Biological data are accumulated in various concepts and approaches, making multi-dimensional features. As data from omics including genomics, epigenomics, transcriptomics, proteomics, metabolomics and phenomics are very complex and heterogeneous, statistical approaches are required to make comprehensive one-fit-all data. This review introduces recent approaches for combining and interpreting biological data in multi-dimension structure, and considerations on combining bioinformatic data with view of genomic, epigenomic and transcriptomic levels will be introduced. Moreover, limtations of computational biology based on artificial intelligence will be discussed.

Key words : biological data, omics, bioinformatics, systems biology

Ⅰ. 서 론

생명이란 무엇인가에 대한 단순한 물음은 근본적이면서도 도전적인 질문 중 하나이다. 답하기 어려운 이유는 복잡한 생명체의 체계로 인해 발생하며, 최근의 연구 기법들은 복잡한 생명체에 대한 더욱 복잡한 자료들을 내놓기 시작했다. 이러한 복잡한 자료들은 생명체의 체계를 다차원으로 축적할 수 있는 가능성을 열었으며, 반대로 다차원 접근법이 이러한 복잡한 체계에 대한 이해를 위해 적합한 방법임이 밝혀지고 있다. 차원을 이루는 자료들은 DNA 염기서열부터 표현형에 대한 정보들까지 다양하다. 이러한 다양한 정보들은 생물정보학 분야에서 자료 분석 방법에 대한 획기적 발전을 이뤄냈다. 이러한 자료 분석을 위해 하나의 통합된 자료 유형으로 분류되어야 한다. 이는 생명체에 대한 복잡한 정보 구조를 단순화 또는 획일화 하기 위한 방법이다. 그러나, 질병에 대한 진단 등에 대한 정보들은 하나의 자료 유형으로 획일화 하기가 어렵다. 이러한 문제는 생물학이나 의약학 분야의 자료에 있어서도 유의미한 정보를 추출하는데 있어 마찬가지의 극복할 점이다(1, 2).

다양한 자료들을 전지적 시점으로 보기 위해 통합하고 정보를 만들어내는 것은 상당한 의미를 지닌다. 이러한 노력들은 세포, 조직, 장기, 개체 수준에서 또는 질병 수준에서 이뤄지고 있다(1, 2). 예를 들어, 집단 수준의 DNA 염기서열을 통한 분석에 있어, 분자 수준의 정립된 신호 기전 정보와 통합되어 특정 질병에 대한 예측이 가능해야 한다. 그러나, 신뢰도를 높일수록 자료들은 방대해지며, 이에 따라 정보들 또한 넘쳐나게 된다. 이에 따라 목적으로 하는 자료와 정보를 통계적으로 유의미한 수준에서 가공하고 제시할 필요성이 높아진다. 따라서, 자료와 정보는 다차원 구조를 형성할 수밖에 없으며, 이러한 구조는 세포에서 개체 그리고 생태 수준에서 다차원 축적 구조의 기하학적 정보 모형을 생각할 수 있는 계기가 된다.

생물학적 흐름에 맞춰 유전체에서 표현형에 이르는 각 단계는 연결되어 있으나, 자료들은 이질적인데다가 양적으로 방대하다. 여기에선 자료 통합에 대한 원리와 방법에 대해 간단히 기술하고자 한다. 나아가서, 여러 자료들에 대한 통합과 이에 대한 적합한 설명에 대해 간략히 논하고자 한다. 세포생물학적 관점에서 자료의 통합을 위한 인공지능 학습 방법과 통계학적 접근법에 초점을 둘 것이며, 최근 생명과학 및 의생명과학에 주로 사용되는 인공지능 기술에 관심

* chosg@ut.ac.kr

있거나 컴퓨터를 이용하여 이질적인 방대한 양의 자료들을 가지고 연구하고자 하는 이들에게는 매우 간략한 요약이 될 것이다.

II. 생명 정보에 대해 고려할 점

생명 정보를 통합하기 위한 기계학습을 위한 몇 가지 고려할 점들이 있다. 의생명과학 정보를 포함한 생명에 대한 정보들은 양적으로 상당하지만, 이들은 매우 복잡하다는 점이 공통점일 것이다. 따라서, 정보들은 공간적으로 다면체를 형성할 수 있으나, 정보들은 각각 분산되어 있다고 할 수 있다. 이는 양적으로 비슷한 사회관계망이나 자연어, 시각정보들과 같은 정보들과는 다른 성격을 지닌다. 전형적인 전장 유전체 연관 연구(genone-wide association study, GWAS)로부터 얻어지는 단일염기 다형성(single nucleotide polymorphism) 정보와 각 개체에 대한 표현형에 대한 정보의 간격을 고려하면 그 정보들은 생명체 정보 전체의 공간 내 상당한 거리를 두고 분산되어 있다고 할 수 있다. GWAS로부터 얻은 정보로는 관심 있는 표현형과 관련된 유의미한 양태를 발견하기란 상당히 어렵다. 따라서, 생물학적 정보 및 임상 정보들에 대한 유의미한 정보들을 통합하여 추출하는 방법은 상당한 노력이 필요하다(3, 4).

또한, 생명 정보들은 의도하지 않더라도, 대체로 편향되어 있다. 이는 기술적인 한계로부터 야기된 것이거나, 물리적인 또는 자연적인 제약이나 자료 수집의 편향성에 기인한다. 예를 들어, 약물 표적에 대한 자료를 수집할 때 하나의 약물이 하나의 유전자 산물에만 결합하는 것은 아니다. 이를 확장하여, 약물들에 대한 표적 단백질들에 대한 정보를 구축할 때, 정보들이 고를 것이란 것은 상상하기 어렵다. 나아가서, 알려지지 않은 약물의 표적에 대한 기전 정보 및 대상 질병, 장기, 세포 및 시간 등을 고려하면, 특정 약물에 대한 정보는 상당히 복잡해지며, 이는 자료 수집에 대한 근본적인 한계를 극복해야만 하는 어려움이 있다. 따라서, 이러한 한계 내에서 유의미한 정보를 추출하는 것 또한 상당한 어려움이 존재한다(5-7).

나아가서, 생명과학은 근본적으로 기존 알려진 지식의 영역에서 새로운 지식을 탐구 및 발견하는 것이라 할 수 있다. 예를 들면, 신약 개발에 있어 동물실험 결과들로부터 임상결과를 예측하는 것이 이에 속할 것인데, 이러한 학습 모델은 종종 새로운 자료에 있어 가설에 부합하지 않는 결과로 인식하여 탈락시킬 수 있다(8). 따라서, 단절 없는 완벽한 생명 정보가 구축되지 않은 이상, 유의미한 정보를 만족스럽게 추출하기란 어려운 일이다. 따라서, 생명 정보의 복잡성과 불완전성에 기인하여, 특정 자료들로부터 학습된 모델들은 생명에 대한 일부 지식만을 전달할 뿐이다. 따라서, 생명 정보에 대한 직관적인 자료 수집 및 통합이 의생명과학을 비롯한 생명과학에 있어 매우 중요하다.

자료들에 대한 통합은 크게 두 가지로 나눌 수 있다. 간략하게, 하나는 수직적인 것이고, 다른 하나는 수평적인 것이라 할 수 있다. 수직적인 자료 통합은 생물학적으로 정해진 관점에서 세포 및 조직 그리고 개체와 집단에 대한 자료들을 모으는 것이라 할 수 있다(9). 반면, 수평적 자료 통합은 후성유전체나 단백질체 수준 등 생물학적으로 정해진 수준에서 자료를 모으는 것이다(10). 자료 통합은 기술적으로 세 가지 방법 중 하나를 따르게 된다(1, 11). 인공지능을 통한 방법으로 유의미한 정보를 추출하기 전에 자료들을 하나의 통일된 자료로 통합하는 것이다. 이는 인공지능을 통한 분석 이전에 자료가 통합되기 때문에 이론적으로는 상당히 그럴듯하다. 이는 다면체 감소 및 대표 학습과 같은 자료에 대한 특성을 학습하는 방법으로 가능하다(12, 13). 반면, 후반부에 통합하는 방법도 가능하다. 수집되는 정보들에 대해 단계별로 통합하여 가중치를 부여하여 통합하는 방법이다(14, 15). 또한, 다중커널학습(multiple kernel learning)이나 집단행렬분해(collective matrix factorization) 또는 심층 신경망 학습(deep neural network learning)과 같은 중개 통합도 가능하다. 이는 정보의 다양성을 설명하기에 충분한 알고리즘을 기초로 한다. 이는 앞서 두 방법과 달리 자료들을 통합하지도 않으며, 각각에 대한 비교 가중치를 부여하지도 않는다. 대신, 자료들에 대한 구조를 유지한 채로 분석 단계 동안 자료들을 통합한다. 이는 수행 능력에 있어 상당히 탁월하나, 알고리즘 자체의 한계를 지니고 있다(11, 16-19).

나아가서, 자료 통합 방법들은 다양한 예측 결과들을 만든다. 예를 들어, 생물학적으로 특정한 시점에서 수집된 유전자 발현 수준에 대한 자료는 유전자들에 대한 기능에 따라 집단별로 나뉘어 새로운 결과들을 예측하게 하며, 단백질들 사이의 결합에 대한 정보들을 또한 예측하게 한다. 자료가 많아질수록 다양한 결과들이 또한 생산되며, 이는 신약에 대한 표적 정보에 대한 예측뿐만 아니라 표적 유전자 정보를 통한 질병에 대한 예측 또한 가능케 한다(20-22).

III. 세포 내 생명 정보 이해

다세포생물을 떠올려 보자. 거의 같은 DNA 정보서열을 가지고 있음에도 각각의 세포들은 상당한 개성들을 지니고

있다. 이러한 세포들의 특성들은 물리적인 요인 등 다양한 세포 내외의 복잡한 요인들이 결정한다고 할 수 있겠지만, 간단하게는 유전자 발현 수준 차이로 설명하기에 충분하다. 이는 후성유전학적인 관점에서 설명 가능하다고도 할 수 있다(23, 24). DNA 염기나 히스톤 단백질의 화학적 변형은 결과적으로 염색질의 물리적 그리고 화학적 변화를 야기하며, 이는 유전자의 발현을 결정하기에 충분하다. 연구자들은 여러 실험적 방법들로 후성유전체에 대한 자료들을 수집할 것이다. 히스톤 단백질의 상태 변화에 대한 자료는 염색질 면역침전 서열분석(chromatin immunoprecipitation seqeuncing, ChIP-seq)으로 파악 가능하다(25). 또한, 뉴클레오좀의 상태 변화 자료를 통해서도 가능하다. 이는 deoxyribonuclease sequencing(DNase-seq)이나 assay for transposase-accesible chromatin(ATAC-seq) 등으로 파악이 가능하다(26, 27). 각각에 대한 서열 정보들은 지도화를 통해 분석되는데, 인공지능을 통한 방법은 이러한 정보들을 통해 설명 가능한 유의미한 생명 정보를 추출해야 한다(28, 29). 후성유전학적인 관점에서 보자면, 정확한 정보를 추출하기 위해 연구자는 여러 후성유전학적 연구 방법들을 통해 얻은 자료들을 통합해야만 한다. 방대한 후성유전학적 자료들은 반자동 유전자 주석(semi-automated genomic annotation, SAGA) 방법에 의해 유전자들을 새로이 군집화 할 수 있다. 이는 세분화된 유전자들에 대한 상태 변화를 직관적으로 보여줄 수 있는 장점을 지니며, 유전체들에 대한 주석화가 이 방법을 통해 이뤄졌으며, 은닉마코프 모델(hidden Markov model)을 기초로 한 HMMSeg(https://noble.gs.washington.edu/proj/hmmseg/)이나 ChromHMM(http://compbio.mit.edu/ChromHMM/) 등 자율학습들이 이용되었다(30-37).

나아가서, 염색질에 결합하는 단백질들의 결합 상태가 유전자 발현 수준을 결정할 것이며, 이는 세포의 특성을 결정할 것이다. 전사인자들은 염색질에 결합하여 유전자 발현을 조절하는 단백질들을 의미한다. 약 1,600여 개의 전사인자들이 여러 세포들에서 특징적인 결합 양상을 나타낸다(38-40). 따라서, 이들의 결합이 언제, 어디에서 일어나는지에 대한 자료들은 유전자 발현 조절에 대한 이해를 돕는다. ChIP-seq은 살아있는 세포에서 전사인자의 염색질 결합을 이해하기 위해 흔히 이용된다(25). 이를 응용한 여러 방법들이 또한 이용되는데, 이들 방법들 모두 여러 생물학적 조건들에서는 전사인자의 근겁한 DNA 결합 부위 동정에 한계가 있다(41, 42). 이들은 모두 특정 표적에 대한 항체가 필요하다는 것이며, 이는 생산에 어려움을 겪는 항체들 또한 존재하기에, 특정 표적에 대한 한계를 나타낸다. 또한, CRISPR eitope tagging ChIP-seq(CETCh-seq)과 같은 방법들은 유전체 편집 과정으로 인해 예상치 못한 결과를 야기할 수 있다(43). 컴퓨터를 이용한 방법은 여러 전사인자들에 대한 결합을 항체나 세포 없이 예측할 수 있다(44-47). 이는 전사인자가 결합하는 염기서열에 대한 특징적인 영역에 대한 이해로부터 출발한다. 이 영역들은 위치 가중치 값의 적용에 중요하게 작용한다. 따라서, 기존의 ChIP-seq 자료로부터 확보된 특정 전사인자에 대한 결합 영역들에 대한 정보들이 활용된다. 예를 들어, 밈(multiple expectation maximization for motif eliciation, MEME)과 같은 전통적인 방법들은 BLAST(https://blast.ncbi.nlm.nih.gov/Blast.cgi)에서 쉽게 볼 수 있는 부분영역에 대한 최대 기대치 알고리즘을 이용하여 전사인자의 결합 영역에 대한 결과를 도출한다. 그럼에도, 이러한 예측 방법은 실험을 통해 도출된 결과와 일치하지 않을 수 있다. 따라서, 전통적인 실험 및 컴퓨터 예측, 두 방법을 통해 도출된 결과를 면밀히 고려해야 한다(48-50).

따라서, 실험을 통한 자료들과 컴퓨터를 통해 분석된 자료들은 재차 통합된다. 이는 각각의 자료들이 가지는 불확실성, 역으로 말하면 신뢰도에 기인한다. 따라서, 후성유전체 정보는 전사인자의 결합에 대한 정보와 통합되어 더욱 정확한 염색질 상태 정보를 제공하게 된다(45-52). HINT(https://www.regulatory-genomics.org/hint/introduction/)는 염색질이 히스톤 단백질로부터 풀려있는지를 탐색 후 은닉마코프 모델을 통해 전사인자 결합 정보를 도출한다(51). 이는 전사인자 결합 영역에 대한 정보 없이도 가능하다. CENTIPEDE (http://centipede.uchicago.edu/)의 경우, 전사인자의 위치 가중치 값과 진정염색질 또는 히스톤 단백질 변형에 대한 정보를 통합하여 전사인자 결합 부위를 결정한다(52). 따라서, 사후 확률값이 중요하게 작용한다고 할 수 있다. 무작위 의사결정 방법인 random forest를 통해 메틸화된 DNA 부위를 탐색하여 전사인자의 결합 여부를 결정하는 방법도 있다(53). Virtual ChIP-seq(https://hoffmanlab.org/proj/virchip/)의 경우엔 RNA-seq 자료를 추가로 활용한다(54). 따라서, DNA 염기 서열 자료와 후성유전학적 정보와 전사인자들에 대한 결합 정보들이 통합되어 특정 전사인자의 결합부위를 예측하게 된다. 따라서, 개별 전사인자에 대한 가중치와 자동변수에 대한 연구자의 이해를 요구한다. 이에 더해, 유전체는 염기서열을 늘어뜨린 선에 불과해 보이지만, 실제 이것은 공간을 형성한다. 이 공간적 구성이 유전자 발현 및 세포의 특성을 결정함에 있어 상당히 중요하다. 염색제 입체 형태 포착 방법(chromosome conformation capture assay)은 특정 유전체 영역에서 공간적 근접도를 양적으로 표현한다(55-57). 이는 위상 영역(topological domain)을 도출하지만, 삼각행렬로 정보가 생산되기 때문에 그 자체의 한계를 지닌다. 몬

테카를로 방법은 자료 도출 과정에서 정보의 손실을 낮추며, MEGABASE(https://ndb.rice.edu/MEGABASE-Documentation) 등과 같은 기술은 마코프 연쇄 몬테카를로(Markov chain Monte Carlo) 방법 등을 이용하여 ChIP-seq 자료로부터 위상 정보를 상세히 한다(55-58).

또한, 세포의 특성에 대한 이해를 위해 우리는 비변역 부위에 대한 이해가 필요하다. 단백질로 번역되는 서열에 대한 자료들을 통해 생물학적 정보를 설명하는 것보다 훨씬 어렵다고 할 수 있다. 수많은 비번역 서열들의 변이는 개체의 표현형이나 질병과 연관되어 있다. 생물학적으로 더욱 중요한 것은 이 비번역 서열들의 변이가 유전체의 상태에 영향을 미치고 유전자 발현 그리고 표현형에 이르는 연쇄적 작용을 한다는 것이다(59-61). 결국 간단한 방법은 무시해도 되는 비번역서열 변이를 이해하고 제거하는 것이다(62, 63). 이는 자료의 양을 줄이고 의미있는 정보를 도출하는 방법으로, 유전자 조절 부위에서 종종 일어나는 현상임이 생물학적 지식으로 쌓여 있기에, 이에 기반한 방법들을 이용할 수 있다. 예를 들어, gkm-SVM(http://www.beerlab.org/gkmsvm/)은 enhancer 부위의 짧은 서열들(k-mers)들을 찾아내고, 이들의 중첩도를 파악한다(64). 학습된 자료는 전사인자 결합 정보를 포함하고 있기 때문에, 알고자 하는 서열들에 대한 비교 분석을 하는 것이 이 방법의 핵심이라 할 수 있다. 또한, 자료 통합을 통한 유전체의 보전 영역에 대한 이해로 돌연변이 부위를 제거할 수도 있다. 진화의 과정 동안 거의 변하지 않은 이 보전 부위는 집단 그리고 종의 경계를 넘어 존재하기 때문에, 충분히 일어날 수 있는 이 영역의 돌연변이들은 질병 등 심각한 상황을 야기할 것이다. 결합된 주석 의존적 제거(combined annotation depedent depletion, CADD; https://cadd.gs.washington.edu/)는 63개의 특징들을 통합 후 제거 가능한 돌연변이 서열을 결정한다(65). 이는 선형 커널 서포트 벡터 머신을 사용하여 후성유전체 자료와 보전 서열에 대한 주석을 통해 결정되며, Eigen(http://www.columbia.edu/~ii2135/eigen.html)의 경우 비지도적 방법을 통해 보전 점수, 단백질 점수, 대립형질 발현 빈도 등을 이용하여 행렬을 블록화하여 정확도를 높인다(66). 한편, 자연선택 압력을 적용하는 경우도 있겠으나, 이들은 앞서 언급한 SAGA 방법과 크게 다르지 않다(67, 68).

최근 연구 동향은 다세포 개체 내 단일세포에 대한 양적 그리고 동적 이해를 위한 연구가 주를 이룬다. 다세포 생명체에서 세포들은 조직을 형성하여 기능에 대한 이해를 쉽게 할 수 있겠으나, 조직 내 세포 이질성에 대한 이해 및 동질 세포 사이에서 비교 가능한 세포 간 상태 변화 또한 연구 대상이다(69-71). 따라서, 자료들은 방대해지고 있으며, 이들을 통합하여 의미 있는 정보들을 추출해 내는 각각의 기술적 절차마다 그 자체로 커다란 벽이라 할 수 있다. 전통적인 방법은 혼합 방법(pooled assay)으로 조직 내 세포들이 혼재한 상태에서 세포 간 변이를 인정하고 분석하는 방법이었다. 그러나, 최근의 연구 기법은 단일 세포 수준에서 생물학적 자료들을 도출해내며, 이는 현재 유전체, 전사체 그리고 단백질체에 이르는 자료들을 제공한다(69-74). 따라서, 단일 세포 수준에서 도출된 유전체 자료들을 통합하여 조직 내 동질 세포 집단 및 이질세포집단들의 특성 규명 및 조직 간 차이 등 개체의 표현형 및 질병에 대한 이해를 위해서는, 인공지능을 통한 자료 통합 및 정보 추출이 상당히 중요하다고 할 수 있다(69, 71, 73). 단일 세포 RNA 서열 분석(single cell RNA sequencing, scRNA-seq)은 개별 세포의 유전자 발현 수준 및 세포 집단의 기능적 다양성과 이질성을 파악하기에 상당히 유용하다. 이 방법은 하나의 표본 내 이질적인 세포들이 어떻게 생물학적으로 조화를 이루는지에 대한, 전통적인 방법으로는 어려운 질문에 답을 한다. 또한, 표본 내 혼재된 이질 세포들로부터 얻은 자료들은 세포 구성에 따라 편향성을 보일 수 있으나, 이러한 문제를 동시에 해결한다(69, 71, 75-77). 그러나, 기술적 문제가 없는 것은 아니다. 특정 세포에서 유전자 발현이 전혀 없다고 파악되는 경우, 또는 자료 통합 시 왜곡을 일으키는 경우 등을 고려해야 한다(69, 71, 75-79). 이러한 문제들을 해결하기 위해 비지도적 학습이 다양하게 연구되었다(80-86). 이들 연구들은 차원 축소를 고려한 알고리즘들이거나 다집단 형성을 위한 알고리즘에 대한 연구들이 주를 이룬다. 대표적으로, 영 과잉 요인 분석(zero-inflated factor analysis, ZIFA; https://github.com/epierson9/ZIFA) 등은 이러한 문제들을 해결하기 위한 방법들이나, 방대한 자료들로부터 도출된 다면체 정보들에 대한 강력한 통계적 가설을 요구한다는 것이 한계라 할 수 있다(81, 82). 즉, 제한된 scRNA-seq 플랫폼이나 기술을 항상 요구하지는 않더라도 가설 설정에서 통계적 난관을 겪는다. 이와 달리, 앙상블 방법은 세포들로부터 자료들을 유사도를 기반으로 우선 군집화한 후 통합하는 방법이다(80, 83). 이는 볼록함수가 아닌 경우의 최적화 방법(non-convex optimization)을 이용하여 군집을 정리하고 통합한다. 따라서, 이를 확대해서 해석해 보면, 여러 scRNA-seq 자료들을 합쳐서 분석하고자 할 때에는 다중작업학습(multi-task learning)을 이용하여 유사도 측정의 정확성을 높일 수 있다(84-90).

이에 더해서, scRNA-seq 자료와 염색체 구조나 단백질체, 대사체 등 자료들을 통합하여 단일 세포에 대한 다중 오믹스 분석을 통한 정보를 획득하고자 하는 연구가 최근 추세이다(90-94). 이는 단일 세포에 대한 이해 그리고 이질적 세

포들로 구성된 조직이나 개체 그리고 집단에 대한 이해를 하기에 상당히 강력한 정보를 제공하나, 이는 현재 인공지능을 통한 계산이 없이는 불가하다. 이에 대한 최근 연구들은 상관관계나 군집에 대한 자료 이해를 통해 접근하는 방법이 주를 이룬다. 그러나, 상관관계를 통한 방법의 경우, 앞서 scRNA-seq에서 언급한 것과 같이, 특정 정보가 유실되는 경우를 고려하지 않을 수 없다. 이러한 자료 통합에 있어 도출되는 문제는 실제로 존재하는 생물학적 현상일 수도 있고, 계산상 문제일 수도 있다. 이를 해결하기 위해 확산모델을 사용하거나, 군집에 대한 대표 값을 설정하여 상관 분석을 하는 방법들이 이용된다. 또는, 각각의 자료들을 각기 수집 및 통합 후에 다시 비교통합 하는 방법도 있다. 이는 Gaussian process latent variable model과 같은 방법 등을 통해 해결하는 경우가 되겠다. 그러나, 어느 경우라 해도 통계 분석 및 정보 도출에 있어 계산의 복잡함이 늘 따른다(89, 90, 93-95). 한편, 단일 세포 내에서 각기 다른 상태들에 대한 자료들을 통합하는 것에서 시간 또는 외부의 여러 조건 등을 고려해보자. 즉, 실험적 조건 또는 연구하고자 하는 목적에 부합한 조건에 일치하는 단일 세포 내 다중 오믹스 정보를 생각해 볼 수 있다. 최근의 연구들은 근사추론, 빠른 소프트웨어 구현 또는 심화학습 등을 이용하여 이를 해결한다(96-100). 예를 들어, SCANPY(https://github.com/theislab/Scanpy)의 경우 Tensorflow와 함께 Python 기반으로 작동한다(97). 한편, 다중신경망 학습을 이용하는 방법도 있다. 이는 메모리 최적 일부 자료에 대한 확률적 기울기를 계산한다. 이러한 자동 논리회로(autoencoder)를 활용한 예 중 하나가 SAUCIE(https://github.com/KrishnaswamyLab/SAUCIE)인데, 이는 자료 간 표준화를 통해 다중오믹스 정보를 추출하는 방법이라 할 수 있다(100).

IV. 결 론

인공지능을 통한 생명 정보 통합은 상당한 의미를 지님을 직관적으로 알 수 있다. 그럼에도, 그 의미를 도출하기란 상당히 어렵다는 것 또한 쉽게 이해된다. 세포 내 현상들에 대한 무수한 자료들에 대해 유의미한 정보를 도출하는 것부터, 개체와 집단에 대한 또는 질병을 포함한 생명 현상들에 대한 자료 수집 및 통합 그리고 정보 도출 및 해석에 이르는 기술적 단계들은 상당한 어려움이 존재한다. 방대한 자료를 생산하기 위해 현재 사용되는 기술들 및 이것들을 통합하여 정보를 도출하기 위한 알고리즘들은 대체로 축약의 방법을 사용하며, 축약의 방법이 옳은지를 통계적으로 판단하는 방법을 또한 이용한다. 그러나, 이는 생물학적 실재를 예측하기엔 적합할 수 있으나, 결과적으로 실험적으로 증명해야만 하는 이중 오류를 조심해야 하는 문제가 이미 존재한다. 이러한 기술적 문제들은 신약을 개발함에 있어서도 마찬가지이며, 특정 질병에 대한 진단 또는 예후 인자를 도출하는 과정에서도 마찬가지이다(101-103).

최근 연구들은 시간을 고려하거나 생물학적 환경 요인을 적용하여 방대한 자료들을 생산하고 통합하여 동적인 자료로 보이게끔 한다. 그러나, 진정한 동적 자료인 물리적 이동을 고려한 자료들이 적용되는 경우 이는 더욱 복잡해진다. 생산적 적대 신경망(generative adversarial network, GAN)과 같은 방법은 이에 대한 하나의 해결책처럼 보인다. 이는 임상에서 영상 자료들에 대해 상당히 유용하게 이용될 것으로 판단되며, 나아가서 조직 내 단일세포 수준의 현미경 상 수준의 자료들에 대한 통합 및 정보 추출에 유의미하게 적용될 것으로 보인다(104-107).

방대한 자료들을 통합하여 도출된 유의미한 생명 정보로부터 숨겨진 의미를 도출하는 것 또한 하나의 도전 과제라 할 수 있다. 기술적 흐름에 따라 도출된 생명 정보가 편향될 가능성이 있으며, 그것이 기술적으로 문제가 아니라 하더라도, 방대한 생물학적 자료들로부터 잃게 되는 해석되지 않은 결과들은 의미가 파악되지 않은 채로 묻히게 된다. 따라서, 실험적 방법의 한계에 대한 이해를 통해 여러 방법들을 사용하여 자료들을 수집할 필요가 있으며, 자료 통합 및 정보 도출에 있어서도 여러 방법들을 사용하고 비교 분석해야 하는 어려움이 있다(107-110).

최근의 인공지능 기술은 생명과학에서 의과학에 집중되어 자료들이 통합되고 있다. 이는 당장은 신약 개발이나 질병에 대한 진단에 활용되겠으나, 추후 생명에 대한 이해로 귀결될 것으로 보인다. 특이점은 분명 생명의 내재적 복잡성에 대한 이해에 있을 것이다. 자료 통합에 있어 요구되는 방법들 또한 통합되어야 할 필요가 있으며, 이는 컴퓨터 하드웨어적인 성능 개선 또한 요구한다. 기술적으로 개선된 양자컴퓨터가 미래의 도전을 넘어설 것으로 예상하나, 하드웨어에 대한 내재적 문제가 해결된 것은 아니다. 즉, 소프트웨어뿐만 아닌 하드웨어의 통합적 자세가 미래 지식 특이점이 될 것으로 예측할 수 있다(111-114).

종합하여, 세포 내 유전체 수준에서 일반적인 생물 정보 통합에 대한 예들과 한계에 대해 간략히 소개하였다. 수학적 지식이 최근 연구 흐름을 위해 상당히 요구되는 것으로 보여질 수 있고, 생명과학 분야에서 이러한 정보 이해에 대한 연구가 활발하지만, 유의미한 생명 정보 도출과 해석을 위해서는 궁극적으로 생물학적 지식을 요구한다. 따라서, 컴퓨터를 활용한 연구는 생명체를 이용한 실험 연구와 함께

방향성이 혼재된 상태로 발전할 것으로 예측할 수 있다.

참고문헌

1. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. (2015) Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet. 16, 85~97.
2. Karczewski KJ, Snyder MP. (2018) Integrative omics for health and disease. Nat Rev Genet. 19, 299.
3. Linghu B, Snitkin ES, Hu Z, Xia Y, DeLisi C. (2009) Genome-wide prioritization of disease genes and identification of disease-disease associations from an integrated human functional linkage network. Genome Biol. 10, R91.
4. Lundby A, Rossin EJ, Steffensen AB, et al. (2014) Annotation of loci from genome-wide association studies using tissue-specific quantitative interaction proteomics. Nat Methods. 11, 868~74.
5. Zitnik M, Zupan B. (2015) Data imputation in epistatic MAPs by network-guided matrix completion. J Comput Biol. 22, 595~608.
6. Zitnik M, Leskovec J. (2017) Predicting multicellular function through multi-layer tissue networks. Bioinformatics. 33, i190~i198.
7. Greene CS, Krishnan A, Wong AK, et al. (2015) Understanding multicellular function and disease with human tissue-specific networks. Nat Genet. 47, 569~79.
8. Mullainathan S, Obermeyer Z. (2017) Does machine learning automate moral hazard and error? Am Econ Rev. 107, 476~80.
9. Zitnik M, Zupan B. (2016) Jumping across biomedical contexts using compressive data fusion. Bioinformatics. 32, i90~i100.
10. Libbrecht MW, Ay F, Hoffman MM, et al. (2015) Joint annotation of chromatin state and chromatin conformation reveals relationships among domain types and identifies domains of cell-type-specific expression. Genome Res. 25, 544~57.
11. Zitnik M, Zupan B. (2015) Data fusion by matrix factorization. IEEE Trans Pattern Anal Mach Intell. 37, 41~53.
12. Zitnik M, Zupan B. (2012) Nimfa: A python library for nonnegative matrix factorization. J Mach Learn Res. 13, 849~53.
13. Sarajlić A, Malod-Dognin N, Yaveroğlu ÖN, Pržulj N. (2016) Graphlet-based characterization of directed networks. Sci Rep. 6, 35098.
14. Yang P, Hwa Yang Y, Zhou BB, Zomaya AY. (2010) A review of ensemble methods in bioinformatics. Curr Bioinform. 5, 296~308.
15. Wu CC, Asgharzadeh S, Triche TJ, D'argenio DZ. (2010) Prediction of human functional genetic networks from heterogeneous data using RVM-based ensemble learning. Bioinformatics. 26, 807~13.
16. Brayet J, Zehraoui F, Jeanson-Leh L, Israeli D, Tahi F. (2014) Towards a piRNA prediction using multiple kernel fusion and support vector machine. Bioinformatics. 30, i364~i370.
17. Mariette J, Villa-Vialaneix N. (2017) Unsupervised multiple kernel learning for heterogeneous data integration. Bioinformatics. 34, 1009~15.
18. Singh A, Shannon CP, Gautier B, et al. (2019) DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics. 35, 3055~62.
19. Zitnik M, Agrawal M, Leskovec J. (2018) Modeling polypharmacy side effects with graph convolutional networks. Bioinformatics. 34, i457~466.
20. Carreras-Puigvert J, Zitnik M, Jemth AS, et al. (2017) A comprehensive structural, biochemical and biological profiling of the human NUDIX hydrolase family. Nat Commun. 8, 1541.
21. Cowen L, Ideker T, Raphael BJ, Sharan R. (2017) Network propagation: A universal amplifier of genetic associations. Nat Rev Genet. 18, 551~62.
22. Zitnik M, Zupan B. (2015) Gene network inference by fusing data from diverse distributions. Bioinformatics. 31, i230~i239.
23. Lappalainen T, Greally JM. (2017) Associating cellular epigenetic models with human phenotypes. Nat Rev Genet. 18, 441~51.
24. Li M, Zou D, Li Z, et al. (2019) EWAS Atlas: A curated

knowledgebase of epigenome-wide association studies. Nucleic Acids Res. 47, D983~8.

25. Johnson DS, Mortazavi A, Myers RM, Wold B. (2017) Genome-wide mapping of *in vivo* protein-DNA interactions. Science. 316, 1497~502.
26. Song L, Crawford GE. (2010) DNase-seq: A high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. Cold Spring Harb Protoc. 2010, pdb.prot5384.
27. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 10, 1213~8.
28. Holder LB, Haque MM, Skinner MK. (2017) Machine learning for epigenetics and future medical applications. Epigenetics. 12, 505~14.
29. Arora I, Tollefsbol TO. (2020) Computational methods and next-generation sequencing approaches to analyze epigenetics data: Profiling of methods and applications. Methods. S1046~2023(20). 30203~6.
30. The ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) project. Science. 306, 636~40.
31. Bujold D, Morais D.A.d.L, Gauthier C, et al. (2016) The international human epigenome consortium data portal. Cell Syst. 3, 496~9.
32. Libbrecht MW, Ay F, Hoffman MM, et al. (2015) Joint annotation of chromatin state and chromatin conformation reveals relationships among domain types and identifies domains of cell-type-specific expression. Genome Res. 25, 544~57.
33. Hoffman MM, Ernst J, Wilder SP, et al. (2013) Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 41, 827~41.
34. Ernst J, Kellis M. (2012) ChromHMM: Automating chromatin-state discovery and characterization. Nat Methods. 9, 215~6.
35. Baum LE, Petrie T. (1966) Statistical inference for probabilistic functions of finite state markov chains. The Annals of Mathematical Statistics. 37, 1554~63.
36. Baum LE, Petrie T, Soules G, Weiss N. (1970) A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. The Annals of Mathematical Statistics. 41, 164~71.
37. Day N, Hemmaplardh A, Thurman RE, Stamatoyannopoulos JA, Noble WS. (2007) Unsupervised segmentation of continuous genomic data. Bioinformatics. 23, 1424~6.
38. Lambert SA, Jolma A, Campitelli LF, et al. (2018) The human transcription factors. Cell. 172, 650~65.
39. Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. (2009) A census of human transcription factors: Function, expression and evolution. Nat Rev Genet. 10, 252~3.
40. Andersson R, Sandelin A. (2020) Determinants of enhancer and promoter activities of regulatory elements. Nat Rev Genet. 21, 71~87.
41. He Q, Johnston J, Zeitlinger J. (2015) ChIP-nexus enables improved detection of *in vivo* transcription factor binding footprints. Nat Biotechnol. 33, 395~401.
42. Skene PJ, Henikoff S. (2017) An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. Elife. 6, e21856.
43. Savic D, Partridge EC, Newberry KM, et al. (2015) CETCh-seq: CRISPR epitope tagging ChIP-seq of DNA-binding proteins. Genome Res. 25, 1581~9.
44. D'haeseleer P. (2006) What are DNA sequence motifs? Nat Biotechnol. 24, 423~5.
45. Cao Y, Kitanovski S, Hoffmann D. (2020) intePareto: An R package for integrative analyses of RNA-Seq and ChIP-Seq data. BMC Genomics. 21(Suppl 11), 802.
46. Subkhankulova T, Naumenko F, Tolmachov OE, Orlov YL. (2020) Novel ChIP-seq simulating program with superior versatility: IsChIP. Brief Bioinform. bbaa352.
47. Qin Q, Fan J, Zheng R, et al. (2020) Lisa: Inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. Genome Biol. 21, 32.
48. Bailey TL, Elkan C. (1995) Unsupervised learning of

multiple motifs in biopolymers using expectation maximization. Mach Learn. 21, 51~80.

49. Jayaram N, Usvyat D, Martin ACR. (2016) Evaluating tools for transcription factor binding site prediction. BMC Bioinformatics. 17, 547.
50. Wasserman WW, Sandelin A. (2004) Applied bioinformatics for the identification of regulatory elements. Nat Rev Genet. 5, 276~87.
51. Gusmao EG, Dieterich C, Zenke M, Costa IG. (2014) Detection of active transcription factor binding sites with the combination of DNase hypersensitivity and histone modifications. Bioinformatics. 30, 3143~51.
52. Pique-Regi R, Degner JF, Pai AA, et al. (2011) Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 21, 447~55.
53. Xu T, Li B, Zhao M, Szulwach KE, Street RC, et al. (2015) Base-resolution methylation patterns accurately predict transcription factor bindings *in vivo*. Nucleic Acids Res. 43, 2757~66.
54. Karimzadeh M, Hoffman MM. (2018) Virtual ChIP-seq: Predicting transcription factor binding by learning from the transcriptome. bioRxiv, 168419~168436.
55. Lieberman-Aiden E, van Berkum NL, Williams L, et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 26, 289~93.
56. Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, et al. (2016) HiChIP: Efficient and sensitive analysis of protein-directed genome architecture. Nat Methods. 13, 919~22.
57. Rao SSP, Huntley MH, Durand NC, et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. 159, 1665~80.
58. Di Pierro M, Cheng RR, Lieberman Aiden E, Wolynes PG, Onuchic JN. (2017) *De novo* prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. Proc Natl Acad Sci. 114, 12126~31.
59. Hindorff LA, Sethupathy P, Junkins HA, et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci. 106, 9362~7.
60. Manolio TA, Collins FS, Cox NJ, et al. (2009) Finding the missing heritability of complex diseases. Nautre. 461, 747~53.
61. Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards B. (2018) Genetic architecture: The shape of the genetic contribution to human traits and disease. Nat Rev Genetics. 19, 110~24.
62. Shlyueva D, Stampfel G, Stark A. (2014) Transcriptional enhancers: From properties to genome-wide predictions. Nat Rev Genet. 15, 272~86.
63. Chen L, Capra JA. (2020) Learning and interpreting the gene regulatory grammar in a deep learning framework. PLos Comput Biol. 16, e1008334.
64. Ghandi M, Mohammad-Noori M, Ghareghani N, et al. (2016) gkmSVM: An R package for gapped-kmer SVM. Bioinformatics. 32, 2205~7.
65. Kircher M, Witten DM, Jain P, et al. (2014) A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 46, 310~5.
66. Ionita-Laza I, McCallum K, Xu B, Buxbaum JD. (2016) A spectral approach integrating functional genomic annotations for coding and noncoding variants. Nat Genet. 48, 214~20.
67. Gronau I, Arbiza L, Mohammed J, Siepel A. (2013) Inference of natural selection from interspersed genomic elements based on polymorphism and divergence. Mol Biol Evol. 30, 1159~71.
68. Gulko B, Hubisz MJ, Gronau I, Siepel A. (2015) A method for calculating probabilities of fitness consequences for point mutations across the human genome. Nat Genet. 47, 276~83.
69. Regev A, Teichmann SA, Lander ES, et al. (2017) The human cell atlas. Elife. 6, e27041.
70. Clevers H, Rafelski S, Elowitz M, et al. (2017) What is your conceptual definition of “cell type” in the context of a mature organism? Cell Syst. 4, 255~9.
71. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al.

(2020) The human tumor atlas network: Charting tumor transitions across space and time at single-cell resolution. Cell. 181, 236~49.

72. Kelsey G, Stegle O, Reik W. (2017) Single-cell epigenomics: Recording the past and predicting the future. Science. 358, 69~75.
73. Tian H, Liu H, Zhu Y, Xing D, Wang B. (2020) The trends of single-cell analysis: A global study. Biomed Res Int. 2020, 7425397.
74. Kaya-Okur HS, Jassens DH, Henikoff JG, Ahmad K, Henikoff S. (2020) Efficient low-cost chromatin profiling with CUT&Tag. Nat Protoc. 15, 3264~83.
75. Dai H, Li L, Zeng T, Chen L. (2019) Cell-specific network constructed by single-cell RNA sequencing data. Nucleic Acids Res. 47, e62.
76. Chen G, Ning B, Shi T. (2019) Single-cell RNA-Seq technologies and related computational data analysis. Front Genet. 10, 317.
77. Li Y, Ma A, Mathé EA, et al. (2020) Elucidation of biological networks across complex diseases using single-cell omics. Trends Genet. 36, 951~66.
78. Zheng GX, Terry JM, Belgrader P, et al. (2017) Massively parallel digital transcriptional profiling of single cells. Nat Commun. 8, 14049.
79. Yuan GC, Cai L, Elowitz M, Enver T, Fan G, et al. (2017) Challenges and emerging directions in single-cell analysis. Genome Biol. 18, 84.
80. Wang B, Zhu J, Pierson E, Ramazzotti D, Batzoglou S. (2017) Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. Nat Methods. 14, 414~6.
81. Pierson E, Yau C. (2015) ZIFA: Dimensionality reduction for zero-inflated single-cell gene expression analysis. Genome Biol. 16, 241.
82. Cleary B, Cong L, Cheung A, Lander ES, Regev A. (2017) Efficient generation of transcriptomic profiles by random composite measurements. Cell. 171, 1424~36.
83. Kiselev VY, Kirschner K, Schaub MT, Andrews T, Yiu A, et al. (2017) Sc3: Consensus clustering of single-cell RNA-seq data. Nat Methods. 14, 483~6.
84. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. (2018) Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 36, 411~20.
85. Sun Z, Chen L, Xin H, et al. (2019) A Bayesian mixture model for clustering droplet-based single-cell transcriptomic data from population studies. Nat Commun. 10, 1649.
86. Haghverdi L, Lun ATL, Morgan MD, Marioni JC. (2018) Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat Biotechnol. 36, 421~7.
87. Soneson C, Robinson MD. (2018) Bias, robustness and scalability in single-cell differential expression analysis. Nat Methods. 15, 255~61.
88. Andrews TS, Kiselev VY, McCarthy D, Hemberg M. (2021) Tutorial: Guidelines for the computational analysis of single-cell RNA sequencing data. Nat Proc. 16, 1~9.
89. Luecken MD, Theis FJ. (2019) Current best practices in single-cell RNA-seq analysis: A tutorial. Mol Syst Biol. 15, e8746.
90. Nguyen ND, Wang D. (2020) Multiview learning for understanding functional multiomics. PLoS Comput Biol. 16, e1007677.
91. Cai M, Li L. (2017) Subtype identification from heterogeneous TCGA datasets on a genomic scale by multi-view clustering with enhanced consensus. BMC Med Genomics. 10(Suppl 4), 75.
92. Koh HWL, Fermin D, Vogel C, et al. (2019) iOmicsPASS: Network-based integration of multiomics data for predictive subnetwork discovery. NPJ Syst Biol Appl. 5, 22.
93. Chierici M, Bussola N, Marcolini A, et al. (2020) Integrative network fusion: A multi-omics approach in molecular profiling. Front Oncol. 10, 1065.
94. Macaulay IC, Ponting CP, Voet T. (2017) Single-cell multiomics: Multiple measurements from single cells. Trends Genet. 33, 155~68.
95. Welch JD, Hartemink AJ, Prins JF. (2017) MATCHER: Manifold alignment reveals correspondence between single cell transcriptome and epigenome dynamics.

Genome Biol. 18, 138.

96. Iacono G, Mereu E, Guillaumet-Adkins A, Corominas R, et al. (2018) Bigscale: An analytical framework for big-scale single-cell data. Genome Res. 28, 878~90.
97. Wolf FA, Angerer P, Theis FJ. (2018) SCANPY: Large-scale single-cell gene expression data analysis. Genome Biol. 19, 15.
98. Campbell KR, Yau C. (2019) A descriptive marker gene approach to single-cell pseudotime inference. Bioinformatics. 35, 28~35.
99. Lin C, Jain S, Kim H, Bar-Joseph Z. (2017) Using neural networks for reducing the dimensions of single-cell RNA-Seq data. Nucleic Acids Res. 45, e156.
100. Amodio M, Srinivasan K, van Dijk D, et al. (2019) Exploring single-celldata with deep multitasking neural networks. Nat Methods. 16, 1139~45.
101. Kasif S, Roberts RJ. (2020) We need to keep a reproducible trace of facts, predictions, and hypotheses from gene to function in the era of big data. PLos Biol. 18, e3000999.
102. Biswas N, Chakrabarti S. (2020) Artificial intelligence (AI)-based systems biology approaches in multi-omics data analysis of cancer. Front Oncol. 10, 588221.
103. Koromina M, Pandi MT, Patrinos GP. (2019) Rethinking drug repositioning and development with artificial intelligence, machine learning, and omics. OMICS. 23, 539~48.
104. Hiranuma N, Lundberg SM, Lee SI. (2019) AIControl: Replacing matched control experiments with machine learning improves ChIP-seq peak identification. Nucleic Acids Res. 47, e58.
105. Alzubaidi A, Tepper J, Lofti A. (2020) A novel deep mining model for effective knowledge discovery from omics data. Artif Intell Med. 104, 101821.
106. Xu Y, Zhang Z, You L, et al. (2020) scIGANs: single-cell RNA-seq imputation using generative adversarial networks. Nucleic Acids Res. 48, e85.
107. Barigye SJ, de la Vega JMG, Perez-Caastillo Y. (2020) Generative adversarial networks (GANs) based synthetic sampling for predictive modeling. Mol Inform. 39, e2000086.
108. Wu F, Lopatkin AJ, Needs DA, et al. (2019) A unifying framework for interpreting and predicting mutualistic systems. Nat Commun. 10, 242.
109. Hassall KL, Mead A. (2018) Beyond the one-way ANOVA for ‘omics data. BMC Bioinformatics. 19(Suppl 7), 199.
110. Le V, Quinn TP, Tran T, Venkatesh S. (2020) Deep in the bowel: Highly interpretable neural encoder-decoder networks predict gut metabolites from gut microbiome. BMC Genomics. 21(Suppl 4), 256.
111. Wang F, Preininger A. (2019) AI in health: State of the art, challenges, and future directions. Yearb Med Inform. 28, 16~26.
112. Isse K, Lesniak A, Grama K, et al. (2012) Digital transplantation pathology: Combining whole slide imaging, multiplex staining and automated image analysis. Am J Transplant. 12, 27~37.
113. Walker EA, Pallathadka SA. (2020) How a quantum computer could solve a microkinetic model. J Phys Chem Lett. 12, 592~7.
114. Zhang H, Liu DE, Wimmer M, Kouwenhoven LP. (2019) Next steps of quantum transport in Majorana nanowire devices. Nat Commun. 10, 5128.

Received Dec. 6, 2020, Revised Dec. 12, 2020, Accepted Dec. 15, 2020