

# 커뮤니티에 의한, 유전자 기반 신피질 셀타입의 분류 및 명명법

김민환

Allen Institute for Brain Science

E-mail: meanhwank@alleninstitute.org, meanhwan@gmail.com

## 요약문

우리 뇌의 피질 회로(cortical circuits)의 기능을 이해하기 위해서는, 그 안의 각 요소를 구성하는 다양한 세포들을 체계적으로 분류하는 것이 필요하다. 과거의 피질 세포의 해부학적, 생리학적, 분자적인 특징을 가지고 한 분류 시도는 데이터의 부족함으로, 통합된 신경 세포 및 아교 세포의 분류를 잘 이끌어 내지 못했고, Single-cell transcriptomics 방법을 통해서, 처음으로 많은 양의 피질 세포를 체계적으로, 완전하고(complete), 정확하고(accurate), 영구적인(permanent) 형태의 데이터를 얻을 수 있게 되었다. 그리고, 이 데이터의 통계적인 분석 방법을 통해서, 기존의 해부학적, 생리학적 요소를 바탕으로 정의된 셀타입(cell-type)과의 종 종 일치함을 보았으며, 다른 신피질 영역 간의 비교 및 종(species) 간에 보존되는 것 또한 볼 수 있었다. 이 새로운 방법을 통해서, 포유류 신피질의 유전자 기반 셀타입을 근간으로 한 분류의 채택을 여기서 건의하고자 한다. 이 분류는 계층 구조(hierarchical)를 가지고 있으며, 표준화된 명명법을 사용한다. 여기서 셀타입은 확률적인 정의를 기반으로 하며, 발달 과정(developmental stages) 및 종(species)에 따라서 다른 방법에 의해서 얻은 데이터 셀타입을 통합할 수 있을 것으로 생각한다. 이 커뮤니티를 기반으로 한 분류 및 지식 그래프(knowledge graph)와 같은 방법으로 데이터를 통합하는 것은 아마도 피질 회로의 연구 영역의 기초를 제공할 수 있을 것으로 여겨진다. 마지막으로, 이 커뮤니티를 기반으로 한 분류, 명명법, 및 데이터 통합은 뇌조직뿐만 아닌, 다른 몸의 셀타입 지도를 제공할 가능성을 제공한다고 조망한다.

**Key Words:** transcriptomics, genome, taxonomy, cell types, modality, neocortex, ontology, classification, nomenclature, knowledge graph

본 자료는 A community-based transcriptomics classification and nomenclature of neocortical cell types. *Nat. Neurosci.*, 23, 1456-1468 (2020)의 논문을 한글로 번역, 요약한 자료입니다.

## 목 차

1. 지금까지의 대뇌 피질 셀타입(cell type) 분류의 역사 개요 및 셀타입(cell type) 분류의 중요성
2. 새로운 셀타입(cell type) 분류의 방법론, 전사체 유전자 기반(transcriptomics)
3. 다른 양식(modalities)들에 의한 셀타입(cell type) 분류와의 일치 문제
4. 대뇌 피질 셀타입(cell type) 분류의 어려운 점
5. 확률 및 계층 구조로 정의된 대뇌 피질 셀타입(cell type) 정의
6. 대뇌 피질 셀타입(cell type)의 통합된 분류(ontology) 및 명명법(nomenclature)
7. 커뮤니티 데이터를 통합할 수 있는 셀타입(cell type) 지식 그래프(knowledge graph)
8. 분류(classification)의 유지 및 업데이트
9. 커뮤니티에 의한, 전사체 유전자를 기반 대뇌 피질 셀타입(cell type)의 분류 및 명명법

### 1. 지금까지의 대뇌 피질 셀타입(cell type) 분류의 역사 개요 및 셀타입(cell type) 분류의 중요성

현대 생물학의 기초 개념은 독일의 병리학자였던, 피르호(Rudolf Virchow, 1821-1902)의 세포 이론에서 출발했다고 볼 수 있으며, 이는 생물체의 구조, 번식, 그리고 병리학적 설명의 기본 단위를 세포로 보는 관점이다. 이 아이디어는 생물 조직을 구성하고 있는 세포들의 목록(catalogs)을 만들기 위한 첫 과정으로 그때 사용한 현미경을 통해서 이루어졌는데, Leeuwenhoek, Hooke, Schleiden와 Schwann 등의 과학자들에 의해서 이루어졌다. 각 생물 종의 이러한 카탈로그 분류 및 그들의 지도는 세포 분류(cell taxonomies)를 보다 조직적으로 나눌 수 있었고, 그들의 공통된 특징을 바탕으로 그룹을 만들어 랭킹(ranks)과 계층구조(hierarchy)를 만들 수 있다. 분류법(taxonomies)은 중요하며, 그 이유는 이를 통해서 그 분야의 기초 개념을 확립하고, 조직적으로 지식을 쌓을 수 있기 때문이다. 이러한 노력의 핵심은, 공유되는 표현형 특징(phenotypic characteristics)을 잘 이해할 수 있도록, 정확하게 셀타입(cell type)을 정의하는 것이다.

피르호의 이론(Virchow's cell theory)은 신경계의 기본 구조 단위는 각 신경세포(neuron)에 있다고 주장한 뇌과학자 카잘(Cajal)의 'neuron doctrine'에 의해서 뇌신경 기관에 확대되었고, 그 이후, 많은 과학자들에 의해서 여러 생물 종의 신경계의 많은 셀타입(cell type)이 기술되었다. 하지만, 이 노력은 높은 차원의 인지 기능을 담당하고 있는 것으로 알려져 있는, 포유류 뇌의 가장 큰 부분을 차지하고 있는 대뇌 피질(cerebral cortex) 혹은 신피질(neocortex)에서 특히 어려움을 겪고 있었고, 그 이유 중 하나는 얇은 층으로 이루어져 있는 포유류의 신피질은 흥분(excitatory) 및 억제(inhibitory) 신경 세포로 구성되어 있는데, 카잘(Cajal)에 의해서 표현된, "뚫고 들어갈 수 없는 정글"이라고 불리워졌을 만큼 복잡한 회로를 가지고 있기 때문이다. 이 대뇌 피질의 구조는 다른 대뇌 피질 지역과 다른 포유류 종과 비교할 때 매우 유사하며, 이는 피질의 진화 과정 동안 반복 및 근간이 되는 기능이 같은 표준적 대뇌 피질 회로('canonical' cortical microcircuit)를 설명하는 것으로

여겨진다.

100년이 지난 지금까지 서서히 이루어진 연구 결과를 통해서, 신피질의 신경세포(neurons), 아교세포(glial cells)들은 많은 셀타입(cell type)을 가지고 있음을 알 수 있었다. 그리고, 이 다른 셀타입(cell types)들은 대뇌 피질 내에서, 그들의 절대적 그리고 상대적 숫자를 기반으로, 기능 및 계산의 각 역할을 수행할 것으로 여겨진다. 카잘(Cajal)을 비롯한 많은 다른 연구자들의 조직 염색 기법(histological stain)을 통해서 해부학적 분류(anatomical classification)가 이루어졌으며 (그림 1a-c), 이러한 해부학적 분류는 수십 가지의 피라미드 모양 세포(pyramidal neurons), 짧은 축색 돌기 세포(short-axon cells), 아교 세포(glial cells) 등이 분류되었고, 이는 다른 연구자들에 의해서 더 보충되었다. 하지만, 대뇌 피질 세포의 셀타입(cell type)을 어떻게 정의할 수 있을지, 그리고 각각의 셀타입(cell type)의 뇌 조직 속에서의 숫자에 대한 예측 등에 있어서 연구자들 사이에서 명확한 일치를 보지 못해 왔다.

최근 몇십 년 동안, 새로운 세포 모양, 미세구조(ultrastructural), 면역 조직학적 염색(immunohistochemical), 전기생리학적인(electrophysiological) 방법 등을 통해서, 다른 셀타입(cell type)들의 발달 과정 중의 기원을 알 수 있게 되었고 (그림 1d-h), 감시 하의(supervised) 혹은 비감시 하의(unsupervised) 분류 방법을 통해서 좀 더 세밀한 대뇌 피질 세포의 표현형 측정(phenotypic measurements)이 가능하게 되었다. 카잘(Cajal)의 고향 스페인에서 열린 2005 페틸라 회의(Petilla Convention)를 통해서 이루어진, 신피질의 억제 세포(inhibitory cell)의 분류 노력은 해부학적, 생리학적, 분자적인 특징을 설명할 수 있는 표준화된 용어를 만들 수 있었다. 이것은 유용했지만, 연구자들이 그들의 연구에 적용하는 데에는 부족했고, 그 이유 중의 하나는 대뇌 피질 신경 세포의 표현형의 특징을 설명하는 데이터 양의 부족을 들 수 있다. 즉 초기의 많은 연구는 수십 개, 많으면 몇백 개 정도의 샘플(sample)을 기반으로 하였지만, 이는 거의 200억 개의 세포를 가진 인간 신피질에 비하면 많이 모자란 숫자이다.

페틸라 회의(Petilla convention)의 성과라고 볼 수 있는 것은 아마도 아직 표준적으로 여러 다른 양식(modality)의 표현형 특징을 가진 셀타입(cell type)을 통합할 수 있는 한 가지의 분류 방법이 없다는 것을 깨닫는 것이다. 많은 연구자들이 여러 가지 다른 방법을 통해서 측정할 수 있는 독립적으로 정의될 수 있는 셀타입(cell type)의 존재에 대해서는 동의해 오고 있지만, 분류에 있어서 가장 바람직한 근거를 찾는 것에는 동의를 얻지 못해 왔다. 원리적으로 보면 여러 기준을 사용할 수 있는데, 1) 해부학적, 세포 간의 연결, 2) 각 세포의 세포막의 전기생리학적 성질, 3) 구조 및 생리학적 특징, 4) 분자적 표지(molecular markers), 5) 발생학적 근원(developmental origins), 6) 후생학적 상태(epigenetic attractor states), 7) 종 간의 유사점을 찾는 진화적 접근(evolutionary approaches identifying homology across species) 등의 특징을 기반으로 하는 것이다. 이상적으로는, 이 분류는 서로 모여져서 동의할 수 있으며, 근본적으로 겹칠 수 있어야 한다. 즉, 해부학적, 분자적, 생리학적 특징을 바탕으로 한 분류 간에 근본적인 일치가 있어야 한다. 하지만 모두를 합친 분류 방법을 찾는 것은 쉽지 않으며, 문헌을 살펴보면, 특정 셀타입(cell type)으로 신경 세포를 배정함에 있어서 연구자들 간의 근본적 차이를 보여 왔다. 그리고, 심지어 전문가들 사이에서도 무엇이 실상(ground truth)을 구성하는지에 대해서 합의점을 찾지 못하기도 한다. 예를 들어, 많은 논문들에서 무엇이 상들리에 세포(chandelier cell)인지에 대해서는 동의를 보이지만, 억제 세포들 중

중요한 부분을 차지하고 있는 바스켓 모양의 세포(basket cell)에 대한 개념은 불분명하다.

이와 같은 불완전성은 기술적인 어려움에 더해져 더욱 심해졌다. 전통적인 방법은 노동 집약적이고, 데이터를 얻는 데 많은 시간이 걸리며, 많은 경우 정량적이지 않으며, 일반적으로 표준화되고, 조직적인 방법으로 샘플을 얻지 못함에 더욱 악화되었다. 그러므로, 분리된 셀타입(cell type)의 존재나 다양한 분류의 중요성을 차지한 후에도, 셀타입(cell type) 문제가 얼마나 어려운지는 놀라운 일이 아니다.

## 2. 새로운 셀타입(cell type) 분류의 방법론, 전사체 유전자 기반(transcriptomics)

최근의 이루어진, 단세포 유전자 발현 분석 방법(high-throughput single-cell transcriptomics; scRNAseq)은 새로운 정량적인 유전 정보를 제공함으로써, 셀타입(cell type) 분류의 패러다임이 바뀌고 있다. 이 접근 방법은 다량의 유전자 정보를 상대적으로 빠른 속도, 적은 비용으로 얻을 수 있으며, 연관된 후생학(epigenomics) 방법으로 세포의 기능과 상태에 중요한 메틸레이션(methylation) 부분과, 예측되는 전사 조절 부분(putative gene transcriptional regulation)을 찾을 수 있다. 이는 휴먼 게놈 프로젝트(Human Genome Project)를 통해서 개발된 방법적, 개념적, 경제적으로 새로운 기법을 제시했고, 미국의 뇌 연구(BRAIN Initiative) 프로그램을 통해서 더 발전되었다. 게놈(Genomes)을 가지고, 각 세포의 RNA를 조직에서 증폭시키는 방법을 통해서, 모든 전사체의 정보(entire transcriptomes which include the sequence and structure of transcripts) 생성이 가능해졌다. 초기에는 실험 당 수백 개의 세포에서의 측정이 가능했지만, 지금은 한 번에 수천 개의 세포 혹은 세포핵의 실험이 가능하며, 동시에 컴퓨터 계산 능력의 발전을 통한 분석은, 대뇌 피질을 포함한, 어떤 뇌조직의 신경 세포의 다양성을 분류, 분석을 용이하게 만들었다 (그림 2).

개념적으로, 게놈(genome)은 각 종들에 대한 전체의 유전자 발현을 보여주는(as the complete set of genes being expressed) 종 안의 총체적 유전자 기술(internal genetic description)을 의미하는 반면, 전사체(transcriptomics)는 시공간적 상태에서 각 생물 기관 내에서의 각 세포에 발현된 전체 유전자의 정보를 통해, 그 내부의 코드를 통해서 각 세포를 기술할 수 있다. 기술적으로, scRNAseq의 규모는 대뇌 피질과 같은 복잡한 뇌 영역의 다양한 세포의 완전하고 정량적인 기술을 처음으로 이루어 낸 것이다. 하지만, 이 확률적 방법으로 기술된 유전자 셀타입(cell type)은 조직 내에서 각 세포의 높은 차원에서의 조망도를 제공하는 것이지, 특정 작은 정도의 셀 표지(cellular markers) 혹은 다른 특징을 기반으로 한 정의와는 근본적으로 다르다.

현재 방법의 규모와 정확성, 그리고 정보량은 기존의 뇌 과학에서 사용된 세포 표현형을 찾는 것보다 훨씬 앞서고 있으며, 브레너(Brenner)에 의해서 주창된 생명과학의 표준(gold standard)으로서 완전하고(complete), 정확하고(accurate), 영구적인(permanent) [CAP] 기준(criteria)에 가깝다고 할 수 있다. 즉, 이런 주된 노력은 분자적 방법을 통한 셀타입(cell type)의 완전한 기술을 여러 다른 수준에서 여러 다른 기관에 의해서 이루어지고 있다. 예를 들면, 대뇌 피질은 앨런 연구소(Allen Institute for Brain Science), 뇌 전체를 목표로 한 미국 국립 보건원 프로젝트(the National Institute of Health (NIH) BRAIN Initiative Cell Census Network), 몸 전체를 목표로 한 휴먼 셀

아틀라스 프로젝트(the Human Cell Atlas) 등이 있다. 그리고, 휴먼 게놈 프로젝트(the Human Genome Project)는 다른 종들과 유전자 비교를 통해서 같은 조상에서 물려받은 유전 정보(orthologous genes)을 밝힐 수 있는 수단을 제공해 주며, 이 노력은 인간 및 모델 종들의 모든 혹은 거의 모든 셀타입(cell type)을 정의하고, 나아가 진화 과정 중 셀타입(cell type)의 다양성을 이해할 수 있는 기회를 제공할 것으로 여겨진다. 이 큰 규모의 투자는 뇌 과학 분야에 큰 영향을 줄 수 있을 것으로 예측되며, 특정 셀타입(cell type)을 조절 혹은 표적(target)할 수 있는 기술들에 의해서 신경망의 기능에 대해서, 질문하고 밝힐 수 있는 방법을 제공할 것이다.

전사체(transcriptomics) 기반 분류는 세포의 다양성의 문제를 묶을 수 있는 다음과 같은 장점을 제공한다.

1. 빠르고 많은 양의 전사체 정보 획득은 복잡한 조직 내에서의 다양한 세포의 완전하고, 조직적인 분석을 가능하게 한다.
2. 발달 과정 동안 세포에 발현된 유전자는 궁극적으로 그 세포의 구조와 기능의 기반이 되며, 이 전사체 정보는 유전자의 세포 수준에서의 기능을 해석할 수 있는 정보를 제공한다.
3. 세포의 분자적 정의는 셀타입 표지의 발견(identification of cell-type markers) 및 표적할 수 있는 유전적 도구(creation of genetic tools to target)를 만들어 내고, 궁극적으로 다른 연구자들에게 표준화된 데이터를 얻을 수 있는 수단을 제공해 줄 수 있다.
4. 전사체 정보는 또한, 유전자와 관련된 질병과 그들의 세포 수준에서의 작동 위치 간의 잠재적 연결을 찾을 수 있는 방법을 제시하며, 나아가 인간 질병에 관한 정보를 제공할 수 있다. 게놈 연구(Genome-wide association studies; GWAS)와의 연결을 통해서, 병과 관련된 유전자를 찾고 그 중심이 되는 셀타입(cell type)의 유전자 발현 변화를 찾아낼 수 있다.
5. 발현 분포도는 진화적, 혹은 발달 과정의 시간 동안 셀타입(cell type)의 정량적 비교를 가능케 한다. 이는 종 간의 셀타입(cell type) 및 발달 단계에 따른 셀타입(cell type)의 정렬, 비교 분석을 가능하게 한다.
6. 전사체 유전 정보는, 다른 몸 기관에서 비슷한 유전자를 사용하는 예처럼, 몸의 기관들 사이의 셀타입(cell type)을 서로 비교할 수 있으며, 이는 우리 몸의 모든 세포를 같은 방법에 의해서 분류할 수 있는 가능성을 보여 준다.

즉, 지금 이루어지고 있는 대뇌 피질의 전사체 연구 초기 데이터를 보면, 이미 많은 생물학적 통찰을 제공하고 있다. 예를 들면, 쥐와 인간의 대뇌 피질 조직의 scRNAseq 분석을 통해서 복잡하지만, 한정된 분자적으로 정의된 각 대뇌 피질 부분에 따라서 ~100 셀타입(cell type)을 찾을 수 있었고, 이는 이전의 다른 양식을 통한 분류에 의한 문헌들과 잘 일치한다. 더해서, 유전자 기반 셀타입(cell type)의 응집체(agglomerative) 형태의 분류에 의한 그룹들 간의 유사성은 구조가 만들어지는 원리를 반영하기도 한다. 나무 혹은 dendrogram의 초기 가지치기(branch)하는 부분은 주요 셀 클래스(신경세포와 비 신경세포, 혹은 활성 세포와 억제 세포)를 반영하며, 세분화된 그룹 분류는 다른 발달 프로그램(developmental program)을 반영한다. 예를 들면, 대뇌 피질 세포의

활성 및 억제 세포 분류를 보면 그들의 발달 기원이 다름을 알 수 있고, 억제 세포 내의 분류를 보면 발달 과정 중의 차이(medial and caudal subdivisions of ganglionic eminence and the preoptic area)를 이미 볼 수 있다 (그림 2a). 그리고 많은 문헌들이 보여주는 세포 운명 특화(cell fate specification) (그림 2b) 및 전사체의 정량적인 발달 과정에서의 특화 및 성숙 과정 (그림 2c), 신경 세포 간의 연결 및 교류 (그림 2d), 그리고, 포유류 및 파충류에서 공통된 분류가 가능함을 알 수 있다 (그림 2e). 이는 분자 단위에서 특화된 작동 원리가 진화적으로 보존되고 있음을 말해 준다고 할 수 있다.

### 3. 다른 양식(modalities)들에 의한 셀타입(cell type) 분류와의 일치 문제

사실 세포의 모양, 생리학적 특징, 그리고 어떻게 신경 세포가 연결되어 있는지를 주로 바탕으로 셀타입(cell type)을 분류해오던 전통적인 연구자들에게 전사체를 기반으로 한 분류를 제안하는 것은, 만약 이 분류가 다른 형태의 특징들로 분류된 셀타입(cell type)과의 일치가 보이지 않는다면 많은 어려움을 가질 것이다. 하지만, 최근의 망막 신경 세포의 연구를 보면, 이 두 가지 다른 형태를 기반으로 한 분류 방법이 많은 유사점을 가지고 있음을 보여 준다.

다른 양식을 기반으로 한 셀타입(cell type) 분류와 전사체 기반 대뇌 피질 셀타입(cell type)의 일치함의 증거들도 조금씩 쌓이고 있다 (그림 3). 실험 방법의 발전은 유전자를 기반으로 한, 특정 크리 라인(Cre lines), 바이러스, 공간 내에서의 전사체 발현 사진 기술(spatial transcriptomics method) 등을 통해서, 모델 동물, 및 인간 조직 내에서 셀타입(cell type)의 표현형을 찾는 연구가 활발히 이루어지고 있다. 그리고, 가장 효과적인 방법은 위원회(consortium)에 의해서 만들어진 유전자 기반 체제에서 커뮤니티의 다른 실험 방법의 연결로 이루어지는 노력이 합쳐지는 것이라고 생각한다.

### 4. 대뇌 피질 셀타입(cell type) 분류의 어려운 점

비록, 지금까지 주요 셀 서브 클래스(subclass) 혹은 높은 차원에서의 다른 양식을 통해 측정된 셀타입(cell type)의 일치를 보여 주었지만, 세밀한 셀타입(cell type)에 관한 것들은 아직도 많은 근거들을 찾아야 한다. 또 다른 어려움으로 들 수 있는 것은, 주어진 셀타입(cell type)의 표현형의 변동(variation)이 존재한다는 것이다. 즉, 이 변화를 일으키는 요인 중 하나는, 세포의 상태에 따라서 유전자의 발현이 변할 수 있기 때문이다. 어떤 연구에서는, 분리된 존재로서의 셀타입(cell type)의 정의 자체가 가능하지 않으며, 아마도, 복잡한 상태의 요소로서 기술되어야 한다는 제안도 있다. 그리고, 이 연속적인 변화들은 그들의 생리학적 역할과 관련 있을 것으로 여겨진다. 예를 들어, 피라미드 모양 세포의 수상 돌기가 쥐의 대뇌 피질 내에서 위치하는 공간에 따라서 모양이 점차적으로 변화하고 있음을 알고 있다 (그림 4a). 그리고, 이는 인간 뇌에서 역시 발달 과정에서 전사 조절 인자의 발현이 점차적으로 변하고 있음을 알고 있다 (그림 4b). 이와 같은 표현형, 그리고 공간적 변화는 현실적으로 셀타입(cell type)을 정의하는 데 있어서, 어디까지를 한계로 같은 셀타입(cell type)으로 혹은 다른 셀타입(cell type)으로 구분할 것인지, 즉 어느 선에서

셀타입(cell type)을 나누는 것이 적절한지는 아직 논쟁이 이루어지고 있는 상태이다.

전사체를 기반으로 한 분류의 특별한 이점 중의 하나는 정량적으로 다른 종에서의 같은 유전자의 공통된 변화를 살펴봄으로써 'Ur-classification'이 가능하다는 것이며, 예를 들어, 최근의 인간 대뇌 피질에 관한 연구를 보면, 기본적인 세포의 구성도는 쥐의 뇌와 비교할 때 많이 보존됨을 알 수 있었고, 그로 인해서, 유사한 셀타입(cell type)을 찾을 수 있었다 (그림 4c). 하지만, 이 연구에서 볼 수 있듯이, 많은 경우에 세부적 셀타입(cell type)을 비교할 때 많은 부분에서 유사하게 정렬하기 쉽지 않은 부분도 가지고 있었다. 마지막으로, 우리는 비 신경세포(non-neuronal)의 셀타입(cell type) 분류에서도 많은 차이점을 발견할 수 있었다. 즉, 쥐와 인간과의 차이점을 찾을 수 있었으나, 인간은 다른 유인원과 비슷함을 발견할 수 있었다. 즉, 종 간의 셀타입(cell type)의 차이점과 유사점은 아마도 확률적이고, 계층 구조로 정의된 셀타입(cell type) 분류법으로 기술할 때 가장 잘 표현될 수 있을 것이라고 생각한다.

## 5. 확률 및 계층 구조로 정의된 대뇌 피질 셀타입(cell type) 정의

현재의 전사체 기반 증거를 분석해 볼 때, 샹들리에 세포(chandelier cells) 혹은 로즈힙(rosehip cells)과 같은 유전자 정보를 통한, 혹은 다른 방법으로 측정된 특정 세포의 특징을 보이는, 아주 뚜렷한 셀타입(cell type)들이 있다. 한편, 세포 상태의 존재, 세포 위치에 따른 표현형의 변형, 종간의 유사성 및 차이점을 함께 가지고 있다는 것은, 각각의 분리된 셀타입(cell type) 분류 및, 그 분류의 종 간의 통합하려는 노력의 난점을 지적하는 것이다. 그래서 성숙하지 못한 상태에서 유동적이지 않은 셀타입(cell type) 분류 방법 채택은, 관찰된 표현형의 변화 정도의 중요성 및 생물학적 해석을 어렵게 할 것이다. 그래서, 타당한 앞으로의 접근은 실제적으로 작동 가능한 정량적으로 정의하는 것이라고 생각한다.

클러스터 분석 방법(Cluster analysis)은 대뇌 피질 세포를 그들의 구조, 생리학적 표현형 혹은 유전자 발현, 최근의 전사체 정보를 포함해서 널리 사용되어 왔다. 다양한 비 감시 하의(unsupervised) 혹은 감시 하의(supervised) 방법이 사용되었으며, 그들의 예(multilayer perceptrons, logistic regression, k-nearest neighbors, affinity propagation, Bayesian classifiers, naïve Bayes, topic modelling, t-distributed stochastic neighbor embedding (t-SNE), graph theory, and autoencoders)는 다양하다. 이러한 방법들은 측정 가능한 각각의 특징을 바탕으로, 통계적으로 정의된 그룹 및 클러스터(cluster)로 나누면서 증거를 기반으로 한 확률적인 셀타입(cell type)의 정의를 자연스럽게 나타낸다.

확률적 셀타입(cell type)의 정의는 특히 전사체 정보에 적용 가능하다. 그 이유는, 높은 공간의 차원을 기반으로 하며, 상대적으로 변동이 크며, 다른 접근 방법으로 비슷한 결과를 가져다 주기 때문이다. 하지만, 이것은 아마도 확률적으로 정의된 유전자 셀타입(cell type)과 그 각각 내의 그리고 상호 간의 변화에 대한 커뮤니티의 동의를 수반한다. 가장 이상적으로 생각할 수 있는 것은 정량적인 셀타입(cell type) 정의가 통계적 방법과는 무관하며, 다른 특질(resolution, complexity, variability, uniqueness)들과 변수와의 관계와 같은 정량적 측정의 기술을 포함하는 것이다. 이 클러스터 분석(cluster analysis)의 타당성을 찾는 방법은 두 가지가 있다. 첫째는, 명확히 정의된 그룹(cluster)

간의 경계가 확실히 정의된 딱딱한 분류('hard' clustering), 그리고, 둘째는 확률적으로 각 그룹의 결정되는 연약한 분류('soft' ('fuzzy') clustering)이다. 이 분류 자체가 확률적 본성을 가지고 있지만, 각 그룹 내의 각각의 요소와의 거리와 그 그룹 사이의 거리를 측정함으로써, 분류 결과의 타당성 정도를 측정할 수 있다.

전사체 분류(transcriptomic taxonomy)를 표현하는 가장 자연스러운 접근 방법은 계층 구조 형태(hierarchical framework)를 채택하는 것이다. 그리고, 이 클러스터 분석(cluster analysis)은, 이 각각의 그룹을 나뉘어가지 모양으로 표현하는, 계층, 계급적 구조를 채택하는 것이 자연스러운 접근이다. 하지만, 계층 구조적 유전자의 관계도를 간단한 나뉘어가지 모양(tree-like structure)으로 표현 가능하지 않을 수도 있다. 아마도 더 복잡하고 높은 차원의 관계도를 필요로 하는 경우가 있으며, 다른 표현법의 사용이 요구된다.

마지막 질문은 주어진 분류(classification or taxonomy)가 어떻게 타당하게 이루어졌다는 근거를 가질 수 있는지 하는 질문이다. 목적은 분류 시스템 자체의 정의가 아니라 어떻게 대뇌 피질의 다양한 세포들을 종합적으로 기술할 수 있는냐이다. 실험적 방법이, 모든 셀타입(cell type)이 존재하는 상태에서 얻어진 것인지, 그리고 그에 따른 분류가 정확하고 올바르게 이루어졌는지가 중요하다. 즉, 찾아진 그룹이 통계적으로 확고한지를 여러 다른 통계적 방법으로 확인해야 한다. 그리고 통계적으로 독립적인, 다른 축의 데이터를 가지고 분류한 셀타입(cell type)과 비교하는 것이 중요하다. 다른 양식을 통해서 얻어진 데이터가 특히 이런 면에서 유용하며, 예를 들면, 분자적, 생리학적, 그리고 세포 모양을 함께 볼 수 있는 패치 식(patch-seq), 혹은 공간 전사체 분포 측정 방법(spatial transcriptomics methods) 등을 들 수 있다 (그림 3a-c). 마지막으로 베이시언 형식(Bayesian framework)을 가진 확률적 정의는 셀타입(cell type)의 컴퓨터 모델 설정, 그리고 실험 데이터 기반 분류를 통해서 알고리즘의 수행 측정의 기준을 제공할 수 있다.

## 6. 대뇌 피질 셀타입(cell type)의 통합된 분류(ontology) 및 명명법(nomenclature)

공동체에 동의를 얻어 채택되기 위해서는, 데이터를 기반, 전사체 분류가 다른 시스템에도 적용 가능한, 공식적으로 합쳐진 셀타입(cell type) 분류 및 이름을 가질 수 있어야 한다. 이름은 중요하며, 옛날 바스크(Basque) 사람들의 격언을 보면, "이름을 가진 것만이 존재 가능하다."라는 말이 전해지며, 유사하게 중국 속담에 "지혜의 시작은 각각에게 각각에 맞는 이름을 불러줄 때부터 시작한다."라는 말이 있다. 이 분류의 목표는, 여러 그룹에서 모아진 다양한 데이터를 서로가 동의한 방법으로 모아서, 위원회의 전문가에 의해서 정리된 모습으로 다시 일반에게 공개하는 것이다. 이 이름 짓는 것은, 데이터 분류에 있어서 중요한 프로젝트이며, 지난 세포 명명법(ontology) 노력에 더해질 수 있으며, 이 명명법(ontology)을 만드는 공동체에게도 좋은 훈련이 될 것이다 (see Open Biomedical Ontology Foundry, <http://www.obofoundry.org>).

실제로 데이터를 기반으로 한 전사체 분류의 이름 명명은, 기존의 세포 명명법(ontology) 커뮤니티에서 가지고 있지 않았던, 다른 여러 난관을 가지고 있지만, 극복 가능하다고 생각한다. 첫 번째로, 전사체를 기반으로 한 유사한 셀타입(cell type)이 다른 여러 해부학적 위치에 존재할 수 있다. 그러므로, 유전자를 기반으로 한 셀타입(cell type)은 해부학적 구조와의 적당한 레벨의 관계를



가지고 있어야 한다. 그리고, 대뇌 피질의 영역에 따른 현격히 그리고 서서히 변하는 것 역시 셀타입(cell type)을 정의하기 힘든 난관이다. 즉, 어느 주어진 피질 영역에 어느 숫자만큼의 유전자 기반 셀타입(cell type)을 가지고 있으며, 이 중 대부분의 셀타입(cell type)은 주변의 다른 대뇌 피질 영역, 혹은 다른 종에서 다소 연속적인 방법으로 변하는 것으로 보인다 (그림 4a,b). 유사하게, 이 분류 체계는 발달 과정의 전 시간적 요소를 포함해야 한다. 세포는 그들의 발달 과정, 혹은 상대적 위치를 고려한 그들의 위치를 기반으로, 정량적으로 정의될 수 있다. 마지막으로, 종 간의 가능한 한 정량적으로 배치가 가능하지만, 이 배치는 진화적 거리의 증가를 가진 다른 수준의 데이터 확보가 있을 때 이루어질 수 있을 것이다. 여러 생물학적 축을 고려한, 참고할 수 있는 명명법을 만드는 것은 커뮤니티에 큰 도움이 될 것으로 보지만, 이는 커뮤니티의 참여 노력이 요구된다.

여기서 제안한 전사체 기반으로 한 분류에 따르면, 유전학(genomics)에서 많은 배울 점을 볼 수 있다. 예를 들면, 참조할 수 있는 분류를 데이터가 쌓임에 따라서, 업데이트와 정교화 과정을 거칠 수 있다. 예를 들어, 유전자(genome) 데이터 구축의 초기 버전에서 시작해, 데이터 집합을 통한 변화를 통해서, 점점 안정적으로 형태를 갖추게 되었다. 지금의 유전자 명명법처럼, 여러 개의 알리아스(multiple aliases)을 가진 공식적 심볼은 세포의 해부학적 위치 및 다른 표현형과 연관된 용어와 연결될 수 있다. 이 명명법은 같은 조상에서 보전된 유전자(orthologous)가 같은 이름을 쓰는 것처럼, 비슷하게 같은 조상에서 보전된 셀타입(cell type)은 다른 종에서 쓸 수 있어야 한다. 유전자 정보에서 사용한 블라스트 정렬 도구(BLAST alignment tools)처럼 셀타입(cell type) 분류에도 비슷한 컴퓨터를 이용한 도구를 통해서, 모든 연구자들이 그들의 데이터를 참고 분류표에 대입해 볼 수 있으면 훨씬 유용할 것이다. 마지막으로, 유전자 분류의 유사성 관점에서, 셀타입(cell type) 분류 역시 어디서 나누고, 어떻게 통합하며, 시공간적, 그리고 진화적 요소를 고려한(with varying levels of splitting or lumping; spatial, temporal or evolutionary criteria), 여러 다른 버전의 셀타입(cell type) 분류를 생각할 수 있다.

명명법은 또한 다른 여러 난제들을 가지고 있다. 현재는, 표준화된 명명법의 부재로 각각의 연구에 사용된 셀타입(cell type)을 따라가는 것에 어려움이 있다. 자연스러운 아이디어 하나는, 현재 가장 많이 쓰이고 있는 각각의 셀타입(cell type)의 가장 잘 정의된 유전자를 기반으로 명명하는 것이다. 하지만, 이 특정 유전자는 각각의 소집단 내 모든 세포에 항상 발견되는 것은 아니다. 그리고, 한 종에서 셀타입(cell type)을 가장 잘 정의할 수 있는 유전자가 종종 다른 종에서는 보존되지 않는 경우가 있다. 전통적으로는 그들의 해부학적 세포 특징을 기반으로 한 이름을 붙여왔고(예를 들면, chandelier, double-bouquet, basket, Martinotti, pyramidal cells), 이 많은 사람들이 써온 이름을 명명법에 함께 넣어 결합하는 것은 유용해 보인다. 하지만, 예를 들어 말꼬리 모양의 축삭 돌기(horsetail axons)와 같은 특징은 종에 따라서 변할 수 있다. 그리고, 새롭게 발견된 셀타입(cell type)은 이와 같은 해부학적 특징을 알 수 없는 경우도 있으므로, 특정 유전자 표지(marker genes)을 기반한 이름 짓는 방법이 더 유용할 것으로 여겨진다.

세포의 형태나 위치, 혹은 각 세포의 특정 유전자를 기반으로 하기보다는, 좀 더 추상적인 명명법을 채택하는 것이 보다 유동적이고 다른 종이나 대뇌 피질을 벗어난 다른 뇌 부분의 적용이 용이할 것이다. 전사체를 기반으로 한 명명법(treating transcriptomic cell clusters as sequence data

(partially implemented for Allen Institute datasets: <https://portal.brain-map.org/explore/classes/nomenclature>))이 그 예이다. 분석을 통한 데이터 셋의 각각의 세포 소 집단(cluster)은 그들의 고유한 접근 ID를 갖게 된다. 확고하게 재현이 잘 되는 소 집단의 경우는 기존에 알려진 명명법이나 역사적으로 불러 온 이름과 다른 알리아스(alias) 숫자뿐만 아니라 공식적인 이름 혹은 상징을 가질 수도 있다. 셀타입(cell type)과 더불어, 높은 차원에서의 클래스(higher-order classes; for example, caudal ganglionic eminence (CGE)-driven GABAergic interneurons, GABAergic interneurons, neurons)는 이름이 붙여질 수 있으며, 이 셀타입(cell type)과 클래스(class)는 다른 종의 여러 레벨(type, class)과 정렬 가능할 것이다.

## 7. 커뮤니티 데이터를 통합할 수 있는 셀타입(cell type) 지식 그래프(knowledge graph)

셀타입(cell type)의 정의는 그들의 기능을 포괄할 수 있는 토대를 기반으로 해야 한다. 게놈(genome)과 유사하게, 유전자의 정의도 다양한 정보를 바탕으로, 그들의 사용, 기능, 병과의 연관성을 수반한다. 반면, 확률적으로 정의된 셀타입(cell type)은, 게놈(genome)의 단백질 코딩 부분과 같이 정해진 정의를 내릴 수 있는 것과는 다르게, 더 많은 정보가 쌓이게 되면, 그 셀타입(cell type)의 이해 정도와 그 기능이 바뀔 수 있음을 예측할 수 있다. 그러므로, 우리의 지식과 셀타입(cell type)에 대한 이해를 보다 유동적으로 조합할 수 있는 방법은, 살아 있고, 업데이트가 가능한, 개개인의 조회와 추론이 가능한 형태로 이루어져야 한다. 또한, 온라인 기반으로 데이터를 모을 수 있는 플랫폼이, 뇌 과학자들 사회에서의 협력적인 참여도를 높여, 큰 효과를 발휘할 수 있을 것으로 본다.

그 하나의 예로 기술 기업과 컴퓨터 공학에서 데이터를 통합하는 커뮤니티 플랫폼으로 널리 쓰이고 있는 방법인 지식 그래프(knowledge graph)를 들 수 있다 (그림 5). 지식 그래프(Knowledge graph)는 데이터의 관계도라고 할 수 있으며, 노드(nodes)는 데이터의 각 요소를 말하며(예를 들면, 셀타입(cell type)과 그들의 속성), 그 각 요소를 잇는 링크(links)나 에지(edges)는 그들의 관계 혹은 서로의 통계학적 연결을 나타낸다. 그리고, 노드(node) 간의 그래프에서 정의된(graph-theoretic) 거리의 확률적 연결 정도, 혹은 알려진 지식을 기반으로 측정 가능하며, 대뇌 피질의 셀타입(cell type)의 지식 그래프(knowledge graph)는 표준화된 전사체 데이터로 시작할 수 있을 것이다. 그리고, 해부학적, 전기 생리학적, 발달 과정과 같은 다른 양식(modality)에서 얻은 데이터를 합칠 수 있을 것으로 본다. 즉, 예를 들면, 세포의 정체성을 결정짓는 여러 요소들(cellular interactions, splicing, local translation, protein phosphorylation) 등은 현재의 scRNA-seq 데이터에는 포함되어 있지 않지만, 미래의 CAP 데이터를 얻게 되면, 이 지식 그래프(knowledge graph)에 첨가될 수 있다. 이와 같은 지식 그래프(knowledge graph)는 새로운 데이터가 이용 가능하게 될 때 두 가지 이용 방법을 생각할 수 있다. 첫째는, 알려진 셀타입(cell type)과 그 셀타입(cell type)의 성질을 새로운 데이터 셋에서 찾고, 각각의 새로운 세포 데이터를 확률적이고, 베이시언(Bayesian) 정의에 따라서 어떤 특정 셀타입(cell type)에 속해 있는지 배정할 수 있다. 둘째는, 새로운 데이터의 첨가로 인해서 바뀐 노드(node)의 정체성(nodes identities)와 그들의 거리는, 이 지식 그래프(knowledge graph)의 전통적인 최적화 알고리즘(optimization algorithms)을 통해서 수동, 혹은 자동으로 업데이트될 수

있다.

여기서 제안된 셀타입(cell type) 지식 구성은, 현재 진행 중인 명명법(ontology) 노력에 도움이 되는, 살아있고 업데이트가 가능한 대뇌 피질의 셀타입(cell type) 분류가 될 수 있다. 이 표준화된 데이터베이스는 데이터베이스 관리자에 의한 관리 및 공개된 알고리즘(open-source algorithms)을 통해서 이끌어 나갈 수 있다. 이처럼 역동적인 데이터베이스이지만, 새로운 데이터를 합칠 경우, 동료들의 검토를 통한 표준화된 방법으로 출판된 데이터만 받아들일 것이며, 이는 검토가 가능한, 정량적, 정성적 셀타입(cell type) 분류를 통해서 통합하는, 연구 영역의 공통분모를 제공할 것이다. 컴퓨터를 기반으로 해서 사용자는 새로운 데이터를 현재의 셀타입(cell type)에 관한 이해 상태에 적용해, 그 속성의 유사성을 비교할 수 있을 것이다. 이는 단순한 참고 문헌 제공 뿐 아닌, 과학 포럼, 및 교육적 역할, 나아가 연구자, 의사, 교육자들이 모두 사용하고 쓸 수 있는 역동적이고, 살아있는 자원이 될 수도 있을 것이다.

## 8. 분류(classification)의 유지 및 업데이트

분류(classification), 이름 명명(nomenclature)과 그와 연동된 지식 그래프(knowledge graph)는 이 분야의 전문가 그룹에 의해서 관리될 수 있다고 생각한다. 이 전문가 그룹은 기본 분류 구조를 유지하면서 통계적 분류 모델을 디자인하고, 주변 동료들의 검토 및 공개적 업데이트를 통해서 유지되고 보완될 것이다. 예를 들어, 다음과 같은 전문가 그룹(NIH BRAIN Initiative Cell Census Network (BICCN), The NeuroLex-International Neuroinformatics Coordinating Facility (INCF), The Neuroscience Information Framework (NIF), The Human Brain Project (HBP), The Human BioMolecular Atlas Program (HuBMAP), The Human Cell Atlas (HCA)) 등이 이를 담당할 수 있을 것이고, 이 중요한 형태(key infrastructure)의 지식을 누가 유지하고, 지원하는지와는 상관없이, 가장 중요한 것은 커뮤니티 안에서 열린 소통을 통해서 관리될 수 있는 노력일 것이다.

## 9. 커뮤니티에 의한, 전사체 유전자를 기반 대뇌 피질 셀타입(cell type)의 분류 및 명명법

결론적으로 말하면, 피질 연구 분야는 대뇌 피질 셀타입(cell type)의 분류가 단세포 전사체 데이터와 그들의 통계적, 확률적 그룹을 기반으로 정의된 셀타입(cell type)을 정량적인 기준으로 합류시킬 준비가 되었다고 생각한다. 물론, 유전자를 기반으로 한 시작단계이지만, 추가적인 CAP 데이터 셋이 유효하면, 이 분류는 교정 및 변경될 수 있고, 점점 더 확고한 여러 양식을 포괄한 분류(true multimodal classification)가 될 것이다. 우리의 이 분류법을 통해서 모든 포유류 종들에서, 그리고 종 마다의 특성을 가지면서, 진화 과정 속에서 보존된 넓은 틀에서 요약할 수 있는 유사한 구조를 가진 척추 동물에 확장 가능하다고 전망한다. 즉, 이와 같이 생물 종 간의 셀타입(cell type)을 체계화된 접근 방법으로 비교하는 것만이 뇌 피질의 셀타입(cell type)이 어떻게 진화해 왔는지 이해를 가능케 할 것으로 본다.

이 분류법은 공동체에서 채택할 경우만 유용하고 성공 가능하다고 생각하며, 명명법뿐 아니라, 공동체에서 다양한 실험 방법의 개발을 통해서 여러 방면의 연구자들의 유연한 접근이

필요하다. 우리는 각각의 항체나 전사 인자 측정법(RNA probes)를 사용해서 셀타입(cell type)을 나타낼 수 있는 키 분자를 표지해주는 분자 유전적 실험 기술, 혹은 특정 셀타입(cell type)이 표지된 쥐들을 커뮤니티에 공유하는 것에 큰 기대를 하고 있다. 그리고, 직접 데이터 셋 안에서 비교하고, 새로운 데이터가 첨가되었을 때, 기존 데이터에 투영할 수 있는 통계적 방법이 요구된다. 즉, 분류법의 중요한 한 부분으로써, 과학자들이 자유롭게 데이터를 살펴보고, 지식을 더하고, 새로운 지식을 얻을 수 있도록, 그래프 및 분석 도구를 가진 열린 정보학 근간이 함께 발전되어야 한다.

추가적으로, 열린 방식 형태 내에서, 점점 더 공동체가 주도하는 데이터 과학 쪽의 경향과 발맞추어, 이 분류법도 공동체의 의견에 따라서 진화되기를 기대한다. 기술 기업들에서 활발히 쓰고 있는 이 데이터 그래프를 통한 지식의 통합은 지식의 전파를 가속화하고, 데이터는 다른 연구 분야 사람들과 연결되지 않은 채, 저널 논문 출판과 동시에 사라지는 이른바 출판 묘지(publication graveyard)를 피할 수 있을 것이다. 여러 생물 종에서 쓸 수 있는 일관되고 통합된 새로운 명명법을 찾는 노력이 필요하며, 이 공통의 명명법의 확립이 의미 있는 데이터 통합과 공유의 표준화된 언어를 제공할 것으로 기대해 본다.

이 공동체를 기반으로 한, 공통의 명명법 및 지식 그래프(knowledge graph)로 강화된, 분류 노력이 만약에 성공한다면, 뇌의 다른 부분, 혹은 다른 몸의 일부로 확장 가능할 것으로 전망한다. 그런 면에서 이 대뇌 피질 셀타입(cell type)의 분류, 즉 여러 각도에서 접근한 오랜 전통의 뇌과학의 중심 문제를 시험대에 올려 놓을 수 있는 좋은 기회라 본다.

### 요약 번역자의 글:

*읽어 주셔서 고맙습니다. 좀 더 관심이 있으시면 아래의 논문들을 더 읽어보시길 추천 드립니다.*

- [1] Cell types, circuits, computation. Azeredo, da Silveira R., Roska, B. Curr Opin Neurobiol. 21, 664-671 (2011).
- [2] Neuronal cell-type classification: challenges, opportunities and the path forward. Zeng, H., Sanes, J.R. Nat Rev Neurosci 18, 530-546 (2017).
- [3] Neuron Names: A Gene- and property-based name format, with special reference to cortical neurons. Shepherd, G.M. et al., Front Neuroanat 13, 25 doi: 10.3389/fnana.2019.00025 (2019).
- [4] Common cell type nomenclature for the mammalian brain. Miller, J.A. et al., Elife, 9, e59928 (2020).

The views and opinions expressed by its writers do not necessarily reflect those of the Biological Research Information Center.

김민환(2021). 커뮤니티에 의한, 유전자 기반 신피질 셀타입의 분류 및 명명법. BRIC View 2021-R38  
Available from <https://www.ibric.org/myboard/read.php?Board=report&id=3942> (Dec 02, 2021)

Email: [view@ibric.org](mailto:view@ibric.org)