



## Contents

<b>나의 첫 논문 쓰기</b>	<b>5</b>
이 책에 대하여 . . . . .	5
누구를 위한 책인가 . . . . .	5
주의사항 . . . . .	5
목차 . . . . .	5
문의 및 피드백 . . . . .	6
<b>1장. 나의 이야기를 만들기</b>	<b>6</b>
처음 논문을 썼던 시절 . . . . .	6
연구를 통해 이야기를 배우다 . . . . .	7
나를 사로잡은 질문 . . . . .	7
질문을 이어 쓰다 . . . . .	7
이 책을 통해 전하고 싶은 것 . . . . .	8
<b>2장. 논문 작성 시작하기</b>	<b>9</b>
논문 작성의 순서 . . . . .	9
논문 폴더 만들기 . . . . .	9
논문 폴더의 기본 구조 . . . . .	10
폴더 이름 정하기 . . . . .	10
드롭박스(Dropbox)에 폴더 만들기 . . . . .	10
기본 폴더 구조 만들기 . . . . .	10
Figures 폴더 . . . . .	11
Supplementary_Figures 폴더 . . . . .	11
Supplementary_Tables 폴더 . . . . .	12
Scripts 폴더 . . . . .	12
원고 MS 워드 파일 만들기 . . . . .	14
Main Manuscript 파일 . . . . .	14
버전 관리 규칙 . . . . .	15
EndNote 라이브러리 관리 . . . . .	15
EndNote 라이브러리 파일 위치 . . . . .	15
EndNote 사용 팁 . . . . .	16
공동저자와 원고 공유하기 . . . . .	16
언제 공유할까? . . . . .	16
워드 파일의 Track Changes 기능 . . . . .	16
본인(1저자)이 원고를 업데이트할 때 . . . . .	16
다른 사람의 원고를 수정할 때 (공동연구자, 공동저자의 경우) . . . . .	17
실전 예시: 버전 관리 워크플로우 . . . . .	17
실전 예시: 논문 폴더 완성본 . . . . .	18
이 부분에서 학생들이 하는 흔한 실수들 . . . . .	19
파일 관리 실수 . . . . .	19
초기 준비 실수 . . . . .	19
협업 실수 . . . . .	19

학생들이 자주 묻는 질문	20
언제 논문 폴더를 만들어야 하나?	20
어느 시점에 저널을 결정하는 것이 좋을까?	20
논문 전체 분량은 어떤 기준으로 정하나?	20
Figure를 먼저 그리고 본문을 쓰는 것이 효율적인가, 아니면 본문을 먼저 작성한 뒤 Figure를 그리는 것이 좋을까?	20
Figure 작업 시 사용하는 툴(예: Illustrator)은 새로 배워야 하나?	20
Main Figure와 Supplementary Figure를 구분할 때, 어떤 기준으로 배치 결정을 하나?	20
<b>3장. Figure 작성하기 — 시각화로 생각하기</b>	<b>20</b>
Figure는 논문의 설계도다	21
얼개부터 그려라: 초안 Figure의 힘	21
Figure의 구성요소	22
Figure를 그리기 전, 손으로 그려보면서, 스케치를 할 수 있다.	22
멀티패널은 꼭 Script로 그려야 한다.	24
예시: 기본 multi-panel Figure 작성	24
Figure는 자주, 여러번 수정할 수 있다.	26
Figure versioning 예시	26
Illustrator를 사용한 추가 편집	27
실전 팁	27
Figure 삽입하고 Figure Legend 작성하기	28
Figure 파일 삽입 형식	28
Figure Legend란?	28
Figure Legend 작성 형식	28
Figure Legend 작성 원칙	28
<b>4장. Results — 데이터가 말하게 하기</b>	<b>29</b>
Results의 핵심 원칙: 데이터가 주인공이다	29
섹션들을 구성하기	30
문단 구성의 원칙	31
문단 한개는 하나의 완결된 아이디어	31
각 문단의 첫 문장 = Topic Sentence	32
용어 사용의 정확성과 일관성	33
개념을 명확히 구분한다	33
용어 일관성이 중요한 이유: 독자의 인지적 색인	33
용어를 통일한다	34
Figure와 Table의 효과적 배치	34
본문에서 구체적으로 언급한다	34
순서대로 언급한다 (당연하게도..)	34
Figure 내용이 본문과 일치해야 한다	35
저널별 요구사항 확인	35
<b>5장. Results 섹션을 작성하기</b>	<b>35</b>
Results 섹션의 문단 구조 — 유기적 연결	35
AI를 활용하여 Results 섹션 작성하기	36
문장 작성할 때 우리가 자주하던 실수들	38
중요한 결과를 여러 문장에 분산시키지 말자	38
과도하게 세세하게 적는 학생들 (MBTI가 S 혹은 ST 인 경우들..)	38
문장을 작성하다가 안드로메다로 여행한 학생들 (MBTI가 NP인 경우들...)	39
문장의 길이: 너무 짧지도, 너무 길지도 않게	39
분석명이나 도구명을 주어로 쓰지 말자	39
처음 논문을 쓰는 학생들은 너무 자주 놀라고, 너무 자주 주목한다...	39
과정 중심 동사 줄이기	40
순환 구조 문장 피하기 (feat. 편쿨섹좌)	40

줄임말에 대하여 . . . . .	40
단어 선택의 정확성 . . . . .	41
수동태 vs 능동태 . . . . .	41
인용(Citation)의 적절한 사용 . . . . .	41
Results 문장을 작성할 때, 확인해야 하는 것 . . . . .	42
통계값은 필수 . . . . .	42
유효숫자 통일 . . . . .	43
샘플 수 명시 . . . . .	43
“higher”, “lower”, “enriched” – 정량적 근거 필요 . . . . .	43
사용한 통계 방법을 명시할 것 . . . . .	43
Non-significant 결과도 언급한다 . . . . .	44
비교 대상을 명확히 . . . . .	44
“Significant”라는 용어를 사용할 때 – 통계적 의미 . . . . .	44
<b>6장. Methods — 첫 논문 글쓰기의 튜토리얼 단계</b> . . . . .	<b>45</b>
연구 설계와 환자 모집 . . . . .	45
시료 수집 및 데이터 생산 . . . . .	45
데이터 품질관리(QC) . . . . .	46
데이터 처리 및 정렬 . . . . .	46
통계 분석 및 모델링 . . . . .	47
Gene Set Enrichment에 대한 작성 요령 . . . . .	47
재현성 및 데이터 코드 가용성 . . . . .	48
Methods를 쓸 때의 마음가짐 . . . . .	48
<b>7장. 참고문헌 조사 및 서지 관리</b> . . . . .	<b>48</b>
문헌 조사의 목적과 흐름: 연구의 계보 읽기 . . . . .	49
데이터베이스 탐색 전략 . . . . .	50
EndNote를 이용한 논문 관리 . . . . .	51
서지 관리의 습관 . . . . .	51
<b>8장. Introduction — 질문의 탄생</b> . . . . .	<b>51</b>
인트로란 무엇인가 . . . . .	52
인트로를 생각하기 vs 인트로를 작성하기 . . . . .	52
인트로의 역삼각형 구조: 세 개의 문단 . . . . .	53
1문단: 넓은 배경(Broad context) . . . . .	53
2문단: 공백(Gap) 제시 . . . . .	53
3문단: 목적(Objective) 제시 . . . . .	54
인트로에서 던진 질문(‘떡밥’)은 결과(Results)와 토의(Discussion) 섹션에서 회수되어야 한다 . . . . .	54
인트로 첫 문단, 첫 문장 . . . . .	55
인트로에서 주요 단어를 정의하여 독자의 머릿속에 색인시키는 방법 . . . . .	55
인트로 마지막 문단 — ‘무엇을 했는가’ vs ‘무엇을 할 것인가’ . . . . .	56
스타일 1: 연구 전체를 요약 (past tense) . . . . .	56
스타일 2: 연구 목적까지만 제시 (future tense or present tense) . . . . .	57
어떤 스타일을 선택할 것인가? . . . . .	57
Introduction 마지막 문단 vs Abstract: 어떻게 다른가? . . . . .	57
Abstract: 논문 전체의 압축본 . . . . .	57
Introduction 마지막 문단: 논문으로의 전환점 . . . . .	58
핵심 차이점 비교 . . . . .	58
실전 작성 전략 . . . . .	58
흔한 실수와 해결법 . . . . .	59
실전 연습 . . . . .	59
마무리: 두 섹션의 관계 . . . . .	60
<b>9장. Discussion과 Conclusion — 당신의 해석</b> . . . . .	<b>60</b>
논문 예시로 배우는 Discussion 구조 . . . . .	60

핵심 발견을 먼저 선언하기 . . . . .	60
Introduction의 떡밥 회수하기 . . . . .	61
사전 연구와의 비교를 통한 맥락 제공 . . . . .	61
주제별 흐름으로 서술하기 . . . . .	61
Limitation: 반성문이 아닌 발전적 비평 . . . . .	61
좋지 않은 Limitation . . . . .	61
발전적인 Limitation . . . . .	62
데이터의 양면성 설명하기 . . . . .	62
Future Direction: 계획서가 아닌 발전 방향 . . . . .	62
좋지 않은 Future Direction . . . . .	63
발전적인 Future Direction . . . . .	63
Conclusion의 역할: Introduction과 대칭을 이루는 마무리 . . . . .	63
Discussion 작성의 실전 전략 . . . . .	64
1. Introduction과 대조표 만들기 . . . . .	64
2. 결과의 위계를 정하기 . . . . .	64
3. 문단 단위로 주제 관리하기 . . . . .	64
4. “Therefore”, “However”, “Moreover”를 전략적으로 사용하기 . . . . .	64
5. 초안에서 과감하게 자르기 . . . . .	64
흔한 실수와 해결책 . . . . .	65
실수 1: Results의 반복 . . . . .	65
실수 2: 과도한 speculation . . . . .	65
실수 3: 방어적 태도 . . . . .	65
Discussion과 Conclusion의 조화 . . . . .	65
Discussion은 당신의 목소리다 . . . . .	66
<b>10장. 당신의 첫번째 논문을 위하여 . . . . .</b>	<b>66</b>
방향성의 설정 . . . . .	66
논문 작성의 타임라인 . . . . .	66
에디터와 리뷰어가 보는 것 . . . . .	67
추천 리뷰어의 선정 . . . . .	67
동료의 피드백 . . . . .	68
<b>부록 1. Daily Paper Summary Program — 요약하고 읽기 . . . . .</b>	<b>68</b>
1. 논문을 쇼핑하기 . . . . .	68
2. 검색 키워드 정하기 . . . . .	69
3. AI에 사용할 프롬프트 . . . . .	69
4. 나의 생각 / 의문 (My Thought) . . . . .	70
5. Notion에 기록하기 . . . . .	70
6. 시간의 분배: 절대로 30분을 넘기지 말것, 차라리 딸깍만 해도 된다 . . . . .	70
7. 마무리 . . . . .	70
<b>부록 2: 논문을 처음 읽는 학생들을 위한 가이드 . . . . .</b>	<b>71</b>
논문의 종류 . . . . .	71
연구 논문 읽기 . . . . .	71
연구 논문의 구조 . . . . .	71
효율적인 연구 논문 읽기 전략 . . . . .	71
연구 논문의 종류 . . . . .	72
리뷰 논문 읽기 . . . . .	72
리뷰 논문의 특징 . . . . .	72
효율적인 리뷰 논문 읽기 전략 . . . . .	73
일반적인 팁과 조언 . . . . .	73
1. 용어 정리 . . . . .	73
2. 꾸준함이 핵심 . . . . .	73
3. 반복 읽기의 힘 . . . . .	73

4. 완벽주의를 버린다 . . . . .	73
실전 연습 가이드 . . . . .	73
초보자를 위한 4주 실천 계획 . . . . .	73
중급자로 가는 길 . . . . .	74
마무리하며 . . . . .	74
더 읽어볼 자료 . . . . .	74

## 나의 첫 논문 쓰기

저자: 안준용 (고려대학교 바이오시스템의과학부)

Last Update: 2025년 11월 2일

### 이 책에 대하여

우리 연구실에 대학원 입학을 목표로 하는 학부연구생들, 그리고 첫 논문을 작성하는 대학원 신입생을 위하여, “첫 논문 작성”에 대한 실전 가이드를 적는다.

지난 6년간, 학생들과 여러 논문 작업을 했다. 학생들이 써온 논문 초고를 읽고, 많은 코멘트를 전달했다. 그러면서 학생들의 글쓰기 습관도 생각해보고, 글을 접근하는 사고방식, 데이터와 결과를 풀어내는 이야기들도 읽었다. 학생들마다 매우 다르지만, 첫 논문을 쓰면서 겪는 실수와 어려움에는 몇 가지 공통점들이 있었다. 그때마다 반복되는 피드백들을 연구실 노션 페이지에 하나씩 기록했다. 이 교재에서는 해당 자료들을 체계적으로 정리하고, 주제별로 묶어서 작성했다. 신입생들이나 인턴 학생들이 논문을 처음 쓸 때 겪는 혼란을 줄이고, 반복해서 하는 피드백을 미리 배울 수 있도록 하기 위해서다.

우리 연구실은 많은 시간을 논문 작성에 할애하고, 학생들은 학부연구생부터 박사과정까지 본인의 연구는 본인이 논문을 작성한다. 나는 한 연구자가 오롯하게 본인의 목소리로 글을 쓰기를 훈련시킨다. 이 과정에 느끼고 얻은 지식을 공유하고, 다음 세대의 우리 연구실 학생들의 교육을 위해 이 책을 작성한다.

### 누구를 위한 책인가

- An Lab의 대학원 신입생들, 인턴 학생들
- 논문 작성이 막막한 유전체 분야의 연구자
- 학생들의 논문 쓰기를 지도하는 지도교수나 연구자

### 주의사항

이 책의 내용은 우리 연구실에서 유전체 연구 논문을 작성하는 맥락에서 쓰였다. 실제 학생들과 논문을 작성하면서 남겼던 논문 원고의 코멘트들과 당시의 기억들을 바탕으로 작성했다. 지극히 개인적인 관점과 노하우를 담고 있다. 다른 분야의 논문 작성 방식과는 다를 수 있고, 다른 연구자들은 다른 접근법을 선호할 수 있다. 이 책의 내용을 절대적 규칙이 아닌 하나의 참고자료로 활용하고, 각자의 분야와 지도교수의 스타일에 맞게 조정하기를 권한다. 그리고 필자가 MBTI에 과몰입한 부분은 알아서 걸러 읽기 바란다.

### 목차

- 1장. 나의 이야기를 만들기
- 2장. 논문 작성 시작하기
- 3장. Figure 작성하기
- 4장. Results — 데이터가 말하게 하기
- 5장. Results 섹션을 작성하기
- 6장. Methods — 첫 논문 글쓰기의 튜토리얼 단계
- 7장. 참고문헌 조사 및 서지 관리
- 8장. Introduction 쓰기
- 9장. Discussion과 Conclusion — 당신의 해석

- 10장. 당신의 첫 논문을 위하여
- 부록1. Daily Paper Summary Program — 요약하고 읽기
- 부록2. 논문을 처음 읽는 학생들을 위한 가이드

## 문의 및 피드백

이 교재에 대한 문의사항이나 개선 제안은 GitHub Issues에 남겨주길 바란다.

## 1장. 나의 이야기를 만들기

이제 우리 연구실에서 나와 학생들이 함께 연구한 지도 어느덧 6년이 넘었다. 많은 학생들이 실험을 설계하고 데이터를 분석했으며, 자신만의 연구를 완성했다. 그들은 1저자 논문을 작성하면서, 자신의 생각과 첫 이야기를 세상에 내보냈다.

우리 연구실에서는 논문 초안을 작성할 때, 학생들이 처음부터 끝까지 직접 작성된다. 나는 그 거칠고 다듬어지지 않은 첫 원고를 열어보며, 문장 너머로 그 학생의 생각을 이해하기 위해 많은 시간을 들인다. 매일, 매주 반복하는 일이지만, 여전히 쉽지 않다. 글에는 한 개인의 생각의 방향이 담겨 있고, 초안에는 그 사람의 사고 구조가 고스란히 드러난다. 특히 첫 논문을 쓰는 일은 누구에게나 어렵다. 첫 논문을 쓴 학생에게 이것은 한 번도 가보지 않은 길이고, 처음 경험 해보는 과정이기 때문이다.

우리 연구실은 Notion을 활용해 연구 노트를 관리한다. 분석 과정에 쓰인 코드, 도출된 결과나 에러를 비롯해서 논문의 아이디어 등을 학생들이 직접 남긴다. 우리는 그 안에서 서로의 연구 노하우를 공유하고, 문제 해결의 과정을 함께 기록한다. 이런 기록들이 쌓여 나중에는 논문의 Methods 섹션이 되고, Results의 뼈대가 된다.

나는 그동안 학생들에게 “논문을 잘 작성하는 법”이라는 노션 페이지를 만들어 제공해 왔다. 수십개의 페이지로 이뤄진 이 문서는 수십 페이지에 이른다. 글의 구성부터 문단 전개, 문장의 구조, 그리고 피겨와 테이블 작성법까지 포함되어 있다. 연구의 큰 그림을 어떻게 제시할 것인지, 각 결과를 어떤 논리 순서로 배열할 것인지, 결론을 어떤 어조로 닫을 것인지까지 구체적으로 설명한다.

**이 책은 그 자료들을 정리하고 다시 쓴 것이다.** 또한 지난 몇년간 작성한 학생들의 “첫 논문 원고”를 다시 열어보면서, 내가 학생들에게 남긴 코멘트를 정리해서 글을 쓴다. 우리 연구실에 들어올 새로운 학생들, 그리고 생애 첫 논문을 쓰는 연구자들에게 도움이 되길 바란다. 특히 옹이크 데이터를 이용해 유전적·생물학적 현상을 탐구하는 학생들이 본인의 첫 논문을 작성할 때 필요한 기본 정보들을 제공한다. 이 책의 내용은 지극히 개인적인 경험을 바탕으로 한다. 지난 몇 년간 학생들을 지도하며 쌓은 실제 사례와 조언을 중심으로 구성했다.

1장에서는 내가 어떻게 논문 쓰기를 배웠는지, 그리고 연구자로서 어떤 질문들을 이어왔는지를 이야기한다. 여러분도 언젠가는 자신만의 연구 서사를 갖게 될 것이다. 이 장은 그 여정의 시작에 대한 이야기다.

## 처음 논문을 썼던 시절

2010년, 나는 호주에서 석사 과정을 시작했다. John Mattick 교수님의 연구실에서 로테이션 학생으로 있으면서 처음으로 논문이라는 것을 써야 했다. 주제는 **Long noncoding RNA**가 뇌 발달과 정신질환에 어떤 영향을 주는지를 다루는 리뷰 논문이었다. 7,500 단어 분량의 원고에 200편이 넘는 선행 연구를 인용했지만, 결국 출간되지는 못했다. 그러나 그 시기는 내게 과학 글쓰기를 처음 배운, 의미 있는 시절이었다. Mattick 교수님의 글을 자주 읽었다. 그의 문장은 명확했고, 논리적이면서도 부드러웠다. 데이터와 해석 사이의 거리를 짧게 유지하면서도, 문장 속에는 늘 사고의 흔적이 남아 있었다. 그의 글을 읽다 보면 연구자의 생각이 문장 구조 안에서 어떻게 흐르는지가 보였다. 나는 그런 글을 쓰고 싶었다.

대학원을 막 입학한 신입생이었고, 유학 생활도 처음이라 영어로 글을 쓰는 일이 쉽지 않았다. 글쓰기 수업을 체계적으로 받을 수도 없던 시기였다. 그래서 내가 선택한 방법은 단순했다. **잘 쓰여진 논문을 찾아 그 문법과 구조, 표현을 하나씩 따라 써보는** 일이었다. 예를 들어, Nature나 Cell에 실린 논문 중 특히 Introduction이 잘 쓰인 글을 골랐다. 그 문단의 첫 문장이 어떻게 주제를 연다음 두 번째 문장이 어떻게 배경을 확장하는지, 단락의 마지막 문장이 다음 단락으로 어떻게 연결되는지를 손으로 따라 썼다. 처음에는 문장 하나하나를 베껴 쓰는 수준이었지만, 점차 그 구조 안에서 나만의 문장을 만들 수 있게 되었다.

그렇게 6개월 동안 여러 연구자들의 글을 필사하며, **문장이 생각을 담는 그릇**이라는 것을 배워갔다. 단락을 어떻게 열고 닫는지, 주제를 어떻게 이어가는지, 결론에서 어떤 어조로 마무리하는지를 손끝으로 익혔다. 그때 썼던 리뷰 논문은

세상에 나가지 못했지만, 내게는 큰 전환점이었다. 논문을 쓴다는 것은 단순히 결과를 기록하는 일이 아니라, **자신의 생각을 정리하고 방향을 세우는 일**이라는 사실을 처음 깨달았기 때문이다.

---

## 연구를 통해 이야기를 배우다

2012년, 나는 박사과정을 시작했다. 직접 실험을 설계하고 데이터를 생산했으며, 그 결과를 바탕으로 첫 번째 1저자 논문을 쓰기 시작했다. 연구 주제는 호주 자폐 스펙트럼 장애 가족들의 유전체 데이터를 분석하는 일이었다. 샘플을 모으고, 실험을 수행하고, 데이터를 정리했다. 그러나 단순히 데이터를 분석하는 데서 그치지 않았다. **결과로부터 의미를 도출해야 했다.** 데이터의 흐름을 이해하고, 그 안에서 논리적 결론을 세우는 일이 필요했다. 이 과정이 바로 연구와 논문 쓰기의 핵심이다.

제1 지도교수 Charles 교수님, 제2 지도교수 Alex 박사님은 유전체 분석을 전공하지 않았다. 그래서 나는 데이터를 보여주는 대신, **그 분석이 어떤 의미를 가지는지를 글로 설득해야 했다.** 원고를 여러 번 수정하고 문장을 고쳐 쓰면서, 글이 단순히 결과의 요약이 아니라 사고의 구조라는 것을 배웠다.

그 시기에 학교에서 제공하는 아카데미 글쓰기 수업을 들었고, How to Write a Scientific Writing이라는 책을 도서관에서 빌려 읽었다. 그 책은 문법적 기초뿐 아니라, 글을 통해 사고를 정리하는 법을 알려주었다. 나는 아마존에서 책을 구입해 밑줄을 치며 읽었고, 문장의 형태와 논리의 흐름을 함께 익히면서 연구자의 언어가 어떻게 만들어지는지를 배웠다. **그때 배운 것은 글쓰기의 형식이 아니라 사고의 습관이었다.** 데이터를 보고 “이건 의미가 있다” 또는 “이건 예상과 다르다”라고 느끼는 것과, 그것을 논리적인 문장으로 풀어내는 것은 전혀 다른 능력이다. 후자는 훈련을 통해서만 키울 수 있다.

논문을 쓸 때 가장 어려운 부분은 “무엇을 발견했는가”가 아니라 “왜 이것이 중요한가”를 설명하는 일이다. 이 질문에 답하기 위해서는 자신의 연구를 더 큰 맥락 속에 위치시킬 수 있어야 한다. 관련 문헌을 많이 읽고, 자신의 결과가 기존 연구와 어떻게 대화하는지를 생각해 보라.

---

## 나를 사로잡은 질문

그 시절, 나는 자폐 연구를 하며 한 가지 질문에 몰두했다. **자폐는 하나의 유전자가 아니라 여러 유전자의 상호작용으로 설명될 수 있는 복잡한 현상**이라는 점이었다. 환자마다 다른 유전적 원인이 있었고, 그 유전자들이 서로 영향을 주고받으며 작동했다. 어떤 유전자는 다른 유전자의 발현을 조절했고, 그 작은 차이들이 누적되어 뇌의 발달 방향을 바꾸었다. 나는 이 복잡한 연결 구조를 이해하고 싶었다. 이 주제는 우리 연구실에서 이미 진행 중이던 연구였다. 두 번째 지도교수였던 Alex 박사님은 자폐 환자군의 유전 변이를 통한 분석하여, 유전자 네트워크의 연결성이 자폐의 중요한 단서가 될 수 있다는 연구를 진행한 적이 있었다. 나는 그 논문을 읽고 강한 인상을 받았다. **그가 던진 질문을 내 방식으로 이어가고 싶었다.**

새로운 데이터를 수집하고, 다른 분석 방법을 적용했다. 그 과정은 단순한 후속 연구가 아니라, 그가 열어둔 질문을 내 언어로 다시 써 내려가는 일이었다. 내가 만든 표와 그래프는 그의 연구 속 문장들과 대화를 나누는 듯했다. **여기서 중요한 점이 있다.** 좋은 연구는 완전히 새로운 것에서만 나오지 않는다. 오히려 기존의 질문을 더 깊이 파고들거나, 다른 각도에서 접근하거나, 새로운 데이터로 재검증하는 과정에서 의미 있는 발견이 나온다. 여러분의 첫 논문도 아마 지도교수나 선배 연구자의 질문에서 출발할 가능성이 크다. 그것은 전혀 부끄러운 일이 아니다. 중요한 것은 그 질문을 자신의 것으로 만들고, 자신만의 답을 찾아가는 과정이다.

---

## 질문을 이어 쓰다

박사과정을 마친 뒤에도 나는 같은 질문을 붙들고 있었다. 유전자의 조합과 발달의 복잡한 구조를 어떻게 해석할 수 있을까, 그리고 그 구조가 인간의 뇌에서 어떤 방식으로 작동하는가. 이 질문은 조금씩 형태를 바꾸며 내 연구의 축이 되었다. UCSF에서의 박사후 연구는 자폐 스펙트럼 장애의 유전적 구조를 해부하는 일로 이어졌다. 대규모 엑솜 데이터를 분석해 자폐의 원인이 되는 100여 개의 고신뢰 위험 유전자를 밝혔고, 신규 변이(de novo variant)가 신경 발달에 미치는 영향을 구체적으로 제시했다 (Satterstrom et al. 2020, Cell). 이때 전장 유전체(whole genome sequencing)를 본격적으로 하면서, 자폐의 다양한 유전적 조성을 연구할 수 있었다. 노코딩 영역에서 발생하는 regulatory variant가 자폐의 위험 요인으로 작용한다는 점을 보였고, 이를 검증하기 위해 ‘**Category-Wide Association Study (CWAS)**’라는

새로운 분석 체계를 만들었다 (An et al. 2018, Science; Werling et al. 2018, Nature Genetics). 이 연구는 논코딩 영역이 단순한 여분의 서열이 아니라 유전자 발현을 조절하는 핵심적 공간임을 드러냈다.

한국으로 돌아와, 내 연구실을 시작하면서 이 질문을 한국인 자폐 가족들에게 묻기 시작했다. 자폐 환자 가족의 유전체 데이터를 분석해 신규 변이가 임상적 표현형의 다양성에 어떤 영향을 미치는지를 정량화하고, 가족 내 성별에 따라 나타나는 유전적 패턴의 차이를 제시했다 (Kim et al. 2024, Genome Medicine; Kim et al. 2025, Genome Medicine). 또한 전두엽의 피질 층에 특이적 유전자에서 발견되는 짧은 텐덤 반복(expansion)이 자폐의 임상적 중증도와 적응력에 관여한다는 결과도 얻었다 (Kim et al. 2024, Psychiatry and Clinical Neurosciences).

이후에는 유전적 정보들을 세포 수준에서 해석하기 위해 단일세포 전사체나 다중 오믹스 데이터를 통합해 분석했다. 뇌 발달 과정에서 특정 세포 유형과 시기에 따라 자폐 관련 유전자들이 다르게 작동한다는 점을 정리하고, 이를 인간 뇌 발달 지도로 시각화했다 (Kim et al. 2024, Experimental & Molecular Medicine; Darbandi et al. 2024, Cell Reports). 이런 세포 유형별 기능 정보를 논코딩 변이 분석에 통합하기 위해 **CWAS-Plus**라는 확장 분석 방법을 개발했다 (Kim et al. 2024, Briefings in Bioinformatics). 이 방법은 기존 CWAS에 싱글셀 데이터를 추가해 드물지만 기능적으로 중요한 논코딩 변이를 탐색할 수 있도록 설계되었다.

여러분에게 보여주고 싶은 것은, 하나의 질문이 어떻게 10년 넘게 이어지며 진화할 수 있는지, 그리고 그 과정에서 논문들이 어떻게 서로 연결되어 하나의 큰 서사를 만드는지다. 첫 논문을 쓸 때는 이런 큰 그림이 보이지 않는다. 당장 눈앞의 데이터와 씨름하고, 한 문단을 쓰는 데도 며칠이 걸린다. 하지만 그 첫 논문이 다음 질문의 씨앗이 되고, 그 질문이 또 다른 연구로 이어진다. 지금 쓰고 있는 여러분의 첫 문장도 언젠가는 누군가의 출발점이 될 것이다.

---

## 이 책을 통해 전하고 싶은 것

지난 6년간 첫 세대의 학생들이 독립연구자로 성장했다. 논문을 스스로 구상하고, 작성하며, 연구한다. 많은 시행착오가 있었지만, 훌륭하게 해냈다. 올해 연말에 있을 Year-In-Review 모임(매년 연말에 연구실의 아젠다를 논의하는 자리)을 준비하는데, 한 선배 학생이 “몇 년 전 저희에게 해주신 조언들을 이번 신입생들에게도 해주시면 어떨까요?”라고 제안했다. 그래서 곰곰이 생각해보니 몇 년간 축적된 원고들, 그 안에 남겨진 코멘트들, 노션의 가이드라인 등을 모아서 책으로 만들면 좋겠다는 생각이 들었다.

이 책은 우리 연구실에 들어와 첫 논문을 쓰려는 학생들을 떠올리며 만들었다.

처음 논문을 쓰는 사람은 누구나 막막하다. 어디서부터 시작해야 할지, 어떤 순서로 써야 할지, 심지어 첫 문장을 어떻게 시작해야 할지조차 감이 오지 않는다. 데이터는 있는데 그것을 어떻게 문장으로 만들어야 할지 모르겠고, 레퍼런스를 찾아 읽어도 내 글과 어떻게 연결해야 할지 막막하다. Introduction은 너무 거창해 보이고, Methods는 지루하기만 하고, Discussion은 뭘 써야 할지 갈피가 잡히지 않는다. **그래도 괜찮다. 모두가 그렇게 시작한다.** 나도 처음 논문을 쓸 때는 한 문단을 쓰는 데 며칠이 걸렸다. 문장 하나를 고치고 또 고치면서, “이게 맞나?” 하고 수없이 의심했다. 지금 돌이켜보면 그 첫 논문은 완벽하지 않았다. 하지만 그 경험이 없었다면 지금의 나도 없었을 것이다.

대학에 연구실을 열고 학생들을 지도하고, 운이 좋게도 훌륭한 학생들을 만났다. 각자가 다른 모양으로 처음을 시작했다. 논문의 구조화가 어색했던 학생, 복문 형태의 문장을 즐겨해서 대표기도문처럼 논문을 쓰는 학생, 데이터를 정확하게 기술하지 못했던 학생, 군복무를 하며 싸지방에서 리뷰 논문을 적은 학생, 한문장을 짓기 위해, 밤새 고심하는 학생 등등. 지금은 그때보다 나은 것을 보며, 우리가 그간 해온 방식을 정리해본다.

이 책에서는 논문의 각 섹션을 어떻게 쓰는지 구체적으로 설명한다. Introduction에서 질문을 어떻게 제시하고, Methods에서 재현 가능한 서술을 어떻게 하며, Results에서 데이터를 어떻게 말하게 하는지, 그리고 Discussion에서 질문으로 어떻게 되돌아가는지를 다룬다. 막막하게 느껴지는 각 단계를 구체적인 예시와 함께 하나씩 풀어갈 것이다. 논문 쓰기는 특별한 재능이 아니라 배울 수 있는 기술이다. 처음에는 서툴러도, 한 문장씩 쓰다 보면 점점 나아진다. 이 책은 여러분이 그 첫 문장을 시작할 수 있도록, 그리고 마지막 문장까지 완주할 수 있도록 돕기 위해 쓰였다.

여러분의 첫 논문이 완벽할 필요는 없다. 중요한 것은 자신의 생각을 문장으로 만드는 연습을 시작하는 것이다. 데이터를 해석하고, 질문을 정리하고, 그것을 다른 사람이 이해할 수 있는 형태로 만드는 과정 자체가 여러분을 성장시킬 것이다. **논문은 한 번 쓰고 끝나는 결과물이 아니라, 질문이 자라나는 과정을 기록하는 일이다.** 여러분이 지금 쓰는 첫 논문은 다음 연구의 출발점이 되고, 그 질문은 또 다른 질문으로 이어진다. 내가 쓴 글이 다음 세대의 질문으로 이어지듯, 여러분의 첫 번째 문장도 언젠가 누군가의 출발점이 될 것이다.

자, 이제 시작해보자. 함께 첫 논문을 완성해보자. 그 여정이 이 책과 함께 시작되기를 바란다.

## 2장. 논문 작성 시작하기

### 논문 작성의 순서

논문을 처음 쓰는 학생들이 가장 먼저 묻는 질문은 “어디서부터 시작해야 하나요?”다. 서론(Introduction)부터 써야 한다고 생각하는 사람이 많은데, 실제로는 그렇지 않다.

추천하는 작성 순서:

1. **논문 폴더 세팅** - 드롭박스에 논문 작성을 위한 폴더를 만들고, 지도교수와 공동저자들에게 공유한다.
2. **Main Figure 준비와 Results 작성** - Main Figure 2개 이상이 준비되면 Results 섹션을 작성하기 시작한다. 또한 Table S1 (샘플 메타 정보)를 작성한다.
3. **전체 Figure 완성** - 모든 Main Figure를 완성한다. Figure 1개당 Results 서브섹션 1개를 작성하는 형태로 진행하며, Methods를 병렬적으로 작성한다.
4. **Introduction과 Discussion 작성** - 이 두 섹션은 서로 수미상관의 관계를 갖기 때문에 동시에 작성하는 것이 중요하다. Introduction에서 제기한 문제를 Discussion에서 해석하기 때문이다.

왜 이런 순서일까? 분석결과를 Figure로 만드는 것은 상대적으로 간단한 작업이다. Figure는 지도교수와 미팅을 하던 자료 (주로 슬라이드)에서 추출하여 작업을 할 수 있다. Figure를 바탕으로 각 Results 섹션을 쓴다. 몇개의 섹션을 쓰다보면, 자신이 발견한 것을 명확히 정리할 수 있다. 그 다음에 Introduction과 Discussion을 동시에 작성하면서 “왜 이 질문이 중요한가”를 설명하는 것이 더 쉽다. Figure가 먼저 완성되어야 Results를 쓸 수 있고, Results가 명확해야 Discussion을 논리적으로 전개할 수 있다.

### 논문 폴더 만들기

지금까지 실험을 하거나 데이터를 분석하면서 수많은 파일들을 만들어왔을 것이다. 그래프도 그려보고, 표도 만들어보고, 코드도 돌려보고... 그런데 막상 논문을 쓰려고 하면 어떤 파일이 최신 버전인지, 어떤 그림을 논문에 넣어야 할지 헷갈리기 시작한다.

“어? 이 그래프 어디 있더라?” “이게 최종 버전이었나? 아니면 저게 최종이었나?” “공동저자들에게 보낸 파일이 어느 거였지?”

이런 상황을 방지하기 위해, 논문 작성을 본격적으로 시작하기 전에 반드시 해야 할 일이 있다. 바로 **논문 전용 폴더를 만드는 것이다.**

지금까지 작업했던 폴더들을 한번 떠올려보자. 아마도 이런 식으로 구성되어 있을 것이다:

```
( )  
  
_ /  
  1_20240101/  
  1_ _20240215/  
    - - /  
    - - - /  
      - /  
      - /  
      - /
```

이런 구조는 **작업 폴더(working folder)**로는 괜찮지만, 논문을 쓰기에는 적합하지 않다. 너무 많은 파일들이 섞여 있고, 최신 버전을 찾기 어려우며, 협업이 어렵다. 지도교수나 공동연구자가 “지금 현재 논문에 들어갈 Figure 1을 보여주세요”라고 하면 바로 보여줄 수 있는가? 논문 제출 시에는 Figure, Table, Supplementary 파일들을 정확하게 제출해야 하는데, 어떤 파일을 올려야 할지 헷갈린다.

파일 크기 문제도 있다. 작업 폴더에는 bam, vcf와 같은 파일이 있거나, fastq와 같은 raw data가 들어있다. 이런 시퀀싱 데이터는 수백 GB에 달할 수 있다. 이런 거대한 파일들을 논문 폴더에 넣으면 드롭박스 용량도 부족하고, 동기화도

느려진다. 논문 작성과 제출에는 이런 raw data가 직접적으로 필요하지 않다. 대신 분석 결과를 정리한 **summary data**(요약 데이터)만 있으면 된다.

그래서 우리는 **논문 폴더**와 **작업 폴더**를 명확히 구분해야 한다.

**작업 폴더(Working Folder)**는 실험과 분석을 하는 공간이다. 시행착오의 흔적이 모두 남아있고, 수많은 버전의 파일들이 존재하며, 본인만 이해할 수 있는 구조로 되어 있다. 크기가 큰 raw data(원시 데이터)를 보관하는 곳이기도 하다. 시퀀싱 리드 파일, 이미지 원본 파일 등이 여기에 있으며, 로컬 서버나 클라우드에 저장한다.

반면 **논문 폴더(Paper Folder)**는 논문에 들어갈 최종 결과물만 정리하는 공간이다. 현재 시점의 “최신 버전”만 보관하고, 누구나 한눈에 이해할 수 있는 명확한 구조를 가지고 있다. 협업자들과 공유하는 공간이며, 작고 정리된 summary data(요약 데이터)만 보관한다. 통계 결과 테이블, 그래프용 데이터 등이 여기에 해당하며, 드롭박스에 저장하여 실시간 동기화한다.

논문을 쓰고 저널에 제출할 때는 500GB의 raw data를 직접 제출하지 않는다. 대신 Figure와 Table에 들어간 **summary data**를 제출하고(보통 수 MB), raw data는 GEO, SRA 같은 공개 데이터베이스에 업로드한 후 논문에는 **accession number**만 적는다. 그러니까 논문 폴더에는 “논문 작성과 제출에 실제로 필요한 가벼운 파일들”만 넣으면 된다.

생각해보자. 주방에서 요리를 하는 공간(작업 폴더)과 손님에게 내놓는 식탁(논문 폴더)은 다르다. 주방은 어질러져 있어도 괜찮지만, 식탁은 깔끔하게 정리되어 있어야 한다. 마찬가지로 작업 폴더에는 수백 GB의 재료(raw data)가 있지만, 논문 폴더에는 완성된 요리(summary data와 Figure/Table)만 담겨있다.

---

## 논문 폴더의 기본 구조

이제 본격적으로 논문 폴더를 만들어보자.

### 폴더 이름 정하기

논문 폴더의 이름은 paper\_ 형식으로 짓는다. 예를 들면 paper\_ASD\_WGS\_2024, paper\_SingleCell\_Organoid, paper\_RareVariant\_Analysis 같은 식이다. paper\_로 시작하면 여러 폴더들 중에서 논문 폴더를 바로 찾을 수 있고, 프로젝트명을 넣으면 나중에 여러 논문을 쓸 때 구분이 쉽다. 간결하면서도 내용을 알아볼 수 있어야 한다. 리뷰 논문의 경우엔 review\_로 폴더를 만든다.

### 드롭박스(Dropbox)에 폴더 만들기

논문 폴더는 반드시 **드롭박스**에 만들어야 한다. 개인적인 취향이지만, 구글 드라이브는 동기화가 느리다. 드롭박스의 유연성을 따라가기 어렵다. 자동 백업으로 컴퓨터가 고장나도 파일이 안전하고, 버전 관리로 실수로 파일을 지워도 복구할 수 있다. 협업이 용이해서 지도교수나 공동연구자와 실시간으로 파일을 공유할 수 있으며, 학교, 집, 외부에서도 같은 파일에 접근할 수 있다.

드롭박스의 가장 큰 장점은 실시간 협업이다. 여러 사람이 동시에 같은 폴더를 보면서 작업할 수 있다. Figure 1을 업데이트하면, 지도교수의 컴퓨터에도 즉시 반영된다. 교수가 원고에 코멘트를 달면, 바로 확인할 수 있다. 공동연구자가 Table을 수정하면, 모든 팀원이 최신 버전을 볼 수 있다. 이메일로 파일을 주고받으면 “어? 제가 받은 게 최신 버전인가요?” 같은 혼란이 생기지만, 드롭박스를 사용하면 **항상 모든 사람이 같은 최신 버전**을 보게 된다.

드롭박스 폴더를 만들었다면, 지도교수, 공동 1저자, 특정 부분에 기여하는 공동연구자들을 초대하자. 폴더를 공유하는 방법은 간단하다. 드롭박스에서 해당 폴더를 우클릭하고 “공유(Share)”를 선택한 후, 초대하고 싶은 사람의 이메일 주소를 입력하고 “편집 가능(Can edit)” 권한으로 초대하면 된다. 이렇게 하면 파일을 업데이트할 때마다 자동으로 동기화되고, 모든 팀원이 협업할 수 있다.

### 기본 폴더 구조 만들기

논문 폴더 안에는 다음과 같은 하위 폴더들을 만든다:

```
paper_ /
  Figures/
```

```
Supplementary_Figures/
Supplementary_Tables/
Scripts/
(      : Data/)
```

이제는 각 폴더의 역할을 자세히 알아보자.

## Figures 폴더

이 폴더는 논문 본문에 들어갈 Figure들의 최신 버전을 보관하는 곳이다. 중요한 것은, 이 폴더는 “논문에 들어갈 현재 최신 버전의 Figure”를 저장하는 곳이라는 점이다. 분석 과정에서 나온 수많은 그래프들을 저장하는 곳이 **아니다**. 예를 들어, 10가지 다른 방법으로 데이터를 시각화해봤다면, 그 10개의 그래프는 작업 폴더에 있어야 한다. 그 중에서 “이게 논문에 들어갈 거야!”라고 결정한 하나만 이 Figures 폴더에 넣는 것이다. 파일 이름은 Figure1\_v1.pdf, Figure2\_v1.pdf, Figure3\_v1.pdf 형식으로 버전 번호(v1, v2, v3...)를 붙여서 관리한다. 수정할 때마다 버전을 올리면, 나중에 “아, 이전 버전이 더 나았는데...”라고 생각될 때 돌아갈 수 있다.

실제 폴더 구조는 다음과 같다:

```
Figures/
  Figure1_v3.pdf (←      )
  Figure2_v2.pdf (←      )
  Figure3_v1.pdf (←      )
  old/
    Figure1_v1.pdf
    Figure1_v2.pdf
    Figure2_v1.pdf
```

현재 최신 버전은 Figures 폴더 바로 아래에 두고, 이전 버전들은 old 서브폴더에 보관한다. 이렇게 하면 폴더를 열었을 때 현재 사용 중인 Figure들만 바로 볼 수 있어서 명확하다. 누군가 “지금 시점에서 최신 버전의 Figure들을 달라”고 하면, Figures 폴더의 최상위에 있는 파일들을 주면 된다.

Figures 폴더에 들어가는 모든 Figure는 **Scripts 폴더에 저장된 코드를 실행하면 누구든지 똑같이 그릴 수 있어야 한다**. 이것이 재현 가능한 연구(reproducible research)의 핵심이다. 예를 들어 Figures/Figure1\_v3.pdf는 Scripts/Figure1.R을 실행하면 만들어져야 하고, Figures/Figure2\_v2.pdf는 Scripts/Figure2.R을 실행하면 만들어져야 한다. 이렇게 하면 리뷰어가 “Figure 1이 어떻게 만들어졌나요?”라고 물으면 코드를 보여줄 수 있고, 나중에 Figure를 수정해야 할 때 코드를 조금만 고치면 된다. 공동연구자가 Figure를 재생성할 수 있으며, 논문 출판 후 다른 연구자들이 방법론을 정확히 이해할 수 있다. Cell 이나 Nature 계열의 저널에서는 Code Reproducibility를 확인하기 위하여, 이와 같은 script 작성이 필수적이다.

## Supplementary\_Figures 폴더

본문에는 들어가지 않지만, Supplementary Information에 들어갈 Figure들을 보관한다. 본문에 들어가기엔 너무 자세하거나, 추가적인 분석 결과들이 여기에 들어간다. Supplementary Figure도 Main Figure와 **동일한 수준의 퀄리티**로 작성되어야 한다. “Supplementary”라고 해서 대충 만들면 안 된다. 리뷰어들은 Supplementary도 꼼꼼히 본다. Main Figure에는 논문의 핵심 메시지를 전달하는 Figure, 독자가 반드시 봐야 하는 결과, 논문의 스토리 전개에 필수적인 Figure가 들어간다. 반면 Supplementary Figure에는 Main Figure를 뒷받침하는 추가 증거, 통계적 검증 결과(예: QC plots, validation plots), 대체 분석 방법의 결과, 세부적인 기술적 디테일이 들어간다. 예를 들어 Main Figure 2가 “우리가 발견한 새로운 유전자 변이”라면, Supplementary Figure S2는 “해당 변이의 검증 실험 결과”가 될 수 있다. Supplementary Figure도 Scripts/FigureS1.R 같은 코드로 재현 가능해야 한다.

```
Supplementary_Figures/
  FigureS1_v1.pdf (←      )
  FigureS2_v2.pdf (←      )
  FigureS3_v1.pdf (←      )
  old/
    FigureS2_v1.pdf
```

## Supplementary\_Tables 폴더

본문에는 들어가지 않지만, Supplementary Information에 들어갈 Table들을 보관한다. 우리가 작성하는 주제의 논문들에서는 Table을 Main manuscript에 거의 넣지 않고, **거의 모든 Table을 Supplementary에 정리한다**. Table은 공간을 많이 차지하고, Main manuscript는 Figure 중심으로 스토리를 전개하는 것이 더 효과적이며, 상세한 데이터는 Supplementary에서 제공하는 것이 일반적이기 때문이다. Supplementary Table의 전형적인 구성을 살펴보면, **Table S1**은 샘플 메타데이터로 거의 모든 논문에 필수다. 보통 여러 개의 Sheet로 구성되는데, Sheet 1에는 각 Sheet에 대한 설명을, Sheet 2 (Table S1A)에는 샘플 기본 정보(나이, 성별, 진단 등)를, Sheet 3 (Table S1B)에는 샘플 QC 정보(시퀀싱 depth, mapping rate 등)를, Sheet 4 (Table S1C)에는 필요시 추가 메타데이터를 넣는다. **Table S2**에는 주요 분석 결과(예: 통계적으로 유의한 유전자 목록)를, **Table S3**에는 추가 분석 결과(예: pathway enrichment 결과)를 넣는다. Supplementary Table도 Scripts/TableS1.R 같은 코드로 자동 생성되어야 한다.

```
Supplementary_Tables/  
  TableS1_v1.xlsx (← )  
  TableS2_v2.xlsx (← )  
  TableS3_v1.xlsx (← )  
old/  
  TableS2_v1.xlsx
```

## Scripts 폴더

이 폴더는 Figure와 Table을 만드는 코드를 저장하는 곳으로, **논문의 재현성(reproducibility)을 보장하는 핵심 폴더다**. 많은 저널들이 요즘 재현성을 강조하면서 분석 코드 제출을 요구하며, 논문 심사에서도 매우 중요하게 다뤄진다. 주의할 점은, 여기에는 **모든 분석 코드**를 넣는 것이 아니라는 것이다. 실제 작업 폴더에서 데이터를 분석한 코드는 거기에 그대로 두면 된다. Scripts 폴더에 들어가는 코드는 Figure를 조합하고 레이아웃을 만드는 코드, Table을 최종 포맷으로 정리하고 엑셀로 저장하는 코드, 여러 분석 결과를 모아서 하나의 Figure/Table로 만드는 코드다.

```
Scripts/  
  Figure1.R      # Main Figure 1  
  Figure2.R      # Main Figure 2  
  Figure3.R      # Main Figure 3  
  FigureS1.R     # Supplementary Figure S1  
  FigureS2.R     # Supplementary Figure S2  
  TableS1.R      # Supplementary Table S1  
  TableS2.R      # Supplementary Table S2  
  utils.R        #
```

**언제부터 Scripts 폴더를 사용하나?** 답은 논문 작성 초기부터다. 많은 학생들이 “논문이 거의 다 끝나면 코드를 정리해야지”라고 생각하는데, 이것은 큰 실수다. Scripts 폴더의 코드는 **처음부터 작성하고 계속 업데이트**해야 한다. Figure를 수정할 때마다 코드만 조금 고치면 되고, 리뷰어가 수정을 요구하면 코드를 실행만 하면 되며, 나중에 “이 Figure를 어떻게 만들었더라?” 하고 헤맬 일이 없기 때문이다. 논문을 제출할 때는 Scripts 폴더의 코드를 GitHub에 업로드하거나 Zenodo에 업로드하여 DOI를 받는다. 그리고 논문의 “Data and Code Availability” 섹션에 다음과 같이 적는다:

“All code used to generate figures and tables is available at GitHub (<https://github.com/username/project>) and archived at Zenodo (DOI: 10.5281/zenodo.1234567).”

이것은 논문 심사에서 매우 중요한 요소다. 리뷰어들은 재현 가능한 연구를 선호하며, 코드가 잘 정리되어 있으면 긍정적인 평가를 받는다.

**Scripts 예시 1: Multi-panel Figure 만들기 (cowplot 사용)** 논문의 Figure는 보통 여러 개의 panel을 조합한 multi-panel plot이다. R의 cowplot 패키지를 사용하면 이를 쉽게 만들 수 있다. 다음은 Figure 1을 만드는 코드 예시다 (Scripts/Figure1.R):

```
# Figure1.R  
# Figure 1: Overview of study design and sample characteristics
```

```

# Author: [Your Name]
# Date: 2024-01-15

# Load libraries
library(ggplot2)
library(cowplot)
library(readr)

# Load data
sample_data <- read_csv("../Data/sample_metadata.csv")
seq_stats <- read_csv("../Data/sequencing_statistics.csv")

# Panel A: Study design schematic
# (
)
panel_a <- ggdraw() +
  draw_image("../Data/study_design.png")

# Panel B: Sample distribution by diagnosis
panel_b <- ggplot(sample_data, aes(x = diagnosis, fill = diagnosis)) +
  geom_bar() +
  theme_cowplot() +
  labs(x = "Diagnosis", y = "Number of samples")

# Panel C: Sequencing depth distribution
panel_c <- ggplot(seq_stats, aes(x = mean_depth)) +
  geom_histogram(bins = 30, fill = "steelblue") +
  theme_cowplot() +
  labs(x = "Mean sequencing depth", y = "Count")

# Combine panels
figure1 <- plot_grid(
  panel_a, panel_b, panel_c,
  labels = c("A", "B", "C"),
  ncol = 3,
  rel_widths = c(1, 1, 1)
)

# Save
ggsave("../Figures/Figure1_v1.pdf",
  figure1,
  width = 12,
  height = 4,
  dpi = 300)

```

이 코드를 실행하면 Figures/Figure1\_v1.pdf가 생성된다. 나중에 Figure를 수정하고 싶으면 코드를 고치고 다시 실행하면 된다.

**Scripts 예시 2: Table S1 만들기** Table S1 (샘플 메타데이터)는 거의 모든 논문에 필수다. 이것도 코드로 자동 생성해야 한다. 다음은 Table S1을 만드는 코드 예시다 (Scripts/TableS1.R):

```

# TableS1.R
# Supplementary Table S1: Sample metadata
# Author: [Your Name]
# Date: 2024-01-15

```

```

# Load libraries
library(openxlsx)
library(readr)
library(dplyr)

# Create a new workbook
wb <- createWorkbook()

# Sheet 1: Description
addWorksheet(wb, "Description")
description <- data.frame(
  Sheet = c("Table S1A", "Table S1B", "Table S1C"),
  Description = c(
    "Basic sample information (age, sex, diagnosis)",
    "Quality control metrics (sequencing depth, mapping rate)",
    "Additional metadata (ethnicity, medication)"
  )
)
writeData(wb, "Description", description)

# Sheet 2: Table S1A - Basic information
addWorksheet(wb, "Table S1A")
basic_info <- read_csv("../Data/sample_basic_info.csv") %>%
  select(sample_id, age, sex, diagnosis, batch)
writeData(wb, "Table S1A", basic_info)

# Sheet 3: Table S1B - QC metrics
addWorksheet(wb, "Table S1B")
qc_metrics <- read_csv("../Data/sample_qc_metrics.csv") %>%
  select(sample_id, mean_depth, mapping_rate, duplication_rate)
writeData(wb, "Table S1B", qc_metrics)

# Sheet 4: Table S1C - Additional metadata
addWorksheet(wb, "Table S1C")
additional_meta <- read_csv("../Data/sample_additional_metadata.csv") %>%
  select(sample_id, ethnicity, medication, family_history)
writeData(wb, "Table S1C", additional_meta)

# Save
saveWorkbook(wb, "../Supplementary_Tables/TableS1_v1.xlsx", overwrite = TRUE)

```

이 코드를 실행하면 Supplementary\_Tables/TableS1\_v1.xlsx가 생성된다.

---

## 원고 MS 워드 파일 만들기

논문 폴더를 만들었다면, 이제 실제로 글을 쓸 원고 파일을 만들어야 한다.

### Main Manuscript 파일

Main manuscript는 세 개의 분리된 워드 파일로 작성한다:

1. Manuscript\_[프로젝트명]\_Main\_v1.docx - Title, Abstract, Introduction, Results, Discussion, References를 포함한다.

2. Manuscript\_[프로젝트명]\_Methods\_v1.docx – Methods 섹션만 별도로 작성한다. Methods는 보통 매우 길고 기술적이어서, Main manuscript와 분리하는 것이 편리하다.
3. Supplementary\_Information\_v1.docx – Supplementary Figure legends, Supplementary Table legends, 추가적인 Methods 등을 포함한다. 왜 세 개로 분리할까? Main manuscript가 너무 길면 편집이 어렵고, Methods는 기술적 디테일이 많아서 분리하는 것이 깔끔하며, 저널 제출 시 보통 Main manuscript와 Supplementary를 별도 파일로 제출하기 때문이다. 예를 들어, paper\_ASD\_WGS\_2024 폴더에는 다음과 같은 파일들이 있어야 한다:

```
paper_ASD_WGS_2024/
  Manuscript_ASD_WGS_Main_v1.docx
  Manuscript_ASD_WGS_Methods_v1.docx
  Supplementary_Information_v1.docx
  ...
```

## 버전 관리 규칙

원고 파일도 버전 관리를 해야 한다. 파일명에 버전 번호를 붙이는 방식은 다음과 같다:

**기본 버전 형식:** v1.0, v1.1, v1.2, ... – 첫 번째 숫자(1)는 major version – 큰 변경이 있을 때 올린다 – 두 번째 숫자(0, 1, 2)는 minor version – 작은 수정이 있을 때 올린다

**실제 예시:**

```
Manuscript_Main_v1.0.docx ( )
Manuscript_Main_v1.1.docx ( )
Manuscript_Main_v1.2.docx ( )
Manuscript_Main_v2.0.docx ( )
Manuscript_Main_v2.1.docx ( )
```

**이전 버전은 old 폴더로:** 새 버전을 만들 때마다 이전 버전은 old/ 폴더로 옮긴다. 이렇게 하면 논문 폴더의 최상위에는 현재 작업 중인 최신 버전만 보인다.

## EndNote 라이브러리 관리

논문을 쓸 때 reference 관리는 매우 중요하다. 우리는 EndNote를 사용한다. 35만원 가량 되는 프로그램을 고려대에서는 무료로 제공한다...

### EndNote 라이브러리 파일 위치

EndNote 라이브러리는 두 개의 파일로 구성된다: – .enl 파일 (EndNote 라이브러리 파일) – .Data 폴더 (PDF 파일들이 저장된 폴더)

이 두 파일은 반드시 논문 폴더에 함께 있어야 한다. 예를 들면:

```
paper_ASD_WGS_2024/
  References_ASD_WGS.enl
  References_ASD_WGS.Data/
    PDF/
      paper1.pdf
      paper2.pdf
    ...
  ...
```

EndNote 라이브러리를 논문 폴더에 두면 여러 가지 장점이 있다. 다른 저자들도 같은 reference를 볼 수 있고, 드롭박스로 자동 백업되며, 논문 제출 시 필요한 모든 파일이 한 곳에 모여 있다.

## EndNote 사용 팁

**Cite While You Write:** 워드에서 글을 쓸 때 EndNote의 “Insert Citation” 기능을 사용한다. 이렇게 하면 reference가 자동으로 formatting되고, 나중에 저널을 바꿨을 때 reference style을 쉽게 변경할 수 있다.

**Reference 정리:** 논문을 쓰면서 계속 새로운 reference를 추가하게 된다. EndNote 라이브러리를 깔끔하게 유지하려면, 각 reference에 Keywords를 추가하고(예: “autism”, “WGS”, “de novo”), Rating을 매기고(중요한 논문은 5점), Notes에 간단한 메모를 추가한다(예: “이 논문은 우리 Discussion에서 인용”).

## 공동저자와 원고 공유하기

논문 폴더를 만들고 원고를 작성하기 시작했다면, 적절한 시점에 공동저자들과 공유해야 한다.

### 언제 공유할까?

**너무 이른 시점:** 초안도 없이 빈 폴더만 공유하면 공동저자들이 뭘 해야 할지 모른다. **적절한 시점:** Main Figure 2개 이상이 완성되고, Results 초안이 어느 정도 작성되었을 때다. 이 시점이면 논문의 방향이 명확하고, 공동저자들이 구체적인 피드백을 줄 수 있다. **너무 늦은 시점:** 논문을 거의 다 쓰고 제출 직전에 공유하면, 공동저자들의 피드백을 반영할 시간이 부족하다.

### 워드 파일의 Track Changes 기능

공동저자들과 원고를 주고받을 때는 Track Changes 기능을 반드시 사용해야 한다.

**Track Changes란?:** 워드의 기능으로, 문서의 모든 수정사항(추가, 삭제, 변경)을 자동으로 기록한다. 누가 언제 무엇을 바꿨는지 모두 보인다.

**Track Changes 켜는 방법:** - 워드 상단 메뉴에서 Review → Track Changes 클릭 - 이제 수정하는 모든 내용이 자동으로 tracking된다

**왜 필요할까?:** - 누가 무엇을 바꿨는지 명확하다 - 수정사항을 Accept 또는 Reject할 수 있다 - 수정 전 원본을 볼 수 있다 - 여러 사람의 피드백을 효율적으로 통합할 수 있다

### 본인(1저자)이 원고를 업데이트할 때

1저자로서 원고를 수정하고 새 버전을 만들 때의 규칙이다.

#### 버전 번호 올리기

수정할 때마다 버전을 올린다. 예를 들어:

Manuscript\_Main\_v2.6.docx ( )

↓ ( )

Manuscript\_Main\_v2.7.docx ( )

minor version (두 번째 숫자)을 하나씩 올린다. 작은 수정은 v2.6.1, v2.6.2 같은 식으로 세 번째 숫자를 사용할 수도 있다.

### Track Changes의 수정사항 Accept/Reject

공동저자가 Track Changes로 수정한 파일을 받으면, 각 수정사항을 검토하고 Accept(수락) 또는 Reject(거절)한다. 모든 수정사항을 처리했으면, **새 버전을 만들 때 Track Changes를 모두 Accept**한다. 그래야 다음 버전에서 깨끗한 상태로 시작할 수 있다. 새로운 버전을 공유할 때 이전 버전의 수정사항이 그대로 남아있으면 어떤 것이 새로운 수정인지 알 수 없고, 파일이 지저분해 보이며, 검토자가 혼란스러워한다.

### 해결된 코멘트 삭제하기

원고를 버전 업하여 회람할 때, 반드시 해결된 코멘트는 삭제한다. 예를 들어 다른 저자가 v2.0에 “이 문장 불명확함. 다시 써주세요”라는 코멘트를 달았고, 그 문장을 수정했다면, v2.1을 만들 때 그 코멘트를 삭제한다. 남겨두는 코멘트는 아직 해결하지 못한 문제, 논의가 더 필요한 부분, 다음 버전에서 다룰 예정인 사항이다. 이렇게 하면 검토자가 “아직 처리되지 않은 것”만 보게 되어 효율적이다.

### 파일명에서 이니셜 삭제

이전 버전에서 누군가 수정해준 파일명에 이니셜이 있다면, 새 버전을 만들 때 이니셜을 삭제한다. 예시:

Manuscript\_Main\_v2.6\_JYA.docx ( )

↓ ( )

Manuscript\_Main\_v2.7.docx (1 )

이렇게 하면 “이니셜 없는 파일 = 공식 최신 버전”이라는 것이 명확해진다.

### 다른 사람의 원고를 수정할 때 (공동연구자, 공동저자의 경우)

반대로, 누군가의 원고를 검토하고 수정할 때는 이렇게 한다.

### Track Changes 반드시 켜기

절대 규칙은 다른 사람의 원고를 수정할 때는 반드시 워드의 **Review** → **Track Changes**를 켜야 한다는 것이다. Track Changes를 켜지 않고 직접 수정하면 원저자가 무엇이 바뀌었는지 알 수 없고, 원저자가 동의하지 않는 수정도 강제로 들어가며, 논의의 여지가 없어진다.

### 파일명에 본인 이니셜 추가

원고를 수정한 후 저장할 때는 버전을 한 단계 올리고(예: v2.6 → v2.6.1), 파일명 맨 뒤에 **본인 이니셜**을 추가한다. 예시:

: Manuscript\_Main\_v2.6.docx

: Manuscript\_Main\_v2.6.1\_HJL.docx

이렇게 하면 누가 수정했는지 명확하고, 원본 파일은 보존되며, 1저자가 여러 사람의 피드백을 쉽게 통합할 수 있다.

### 코멘트 적극 활용

수정 외에 의견을 남길 때는 코멘트를 사용한다. “이 부분 데이터가 정확한가요?”, “이 문장을 이렇게 바꾸는 건 어떨까요?”, “Figure 2를 여기로 옮기면 어떨까요?” 같은 경우다. 코멘트는 제안이고, Track Changes는 수정이다. 둘을 적절히 섞어서 사용하자.

### 실전 예시: 버전 관리 워크플로우

현실적인 상황을 예로 들어보자.

**1단계:** 학생(1저자)이 초안을 작성한다

Manuscript\_Main\_v1.0.docx →

**2단계:** 다른 저자가 수정한다

Manuscript\_Main\_v1.0.docx ( )

→ (Track Changes )

→ Manuscript\_Main\_v1.0.1\_JYA.docx ( )

**3단계:** 학생이 수정사항을 반영한다

Manuscript\_Main\_v1.0.1\_JYA.docx ( )

→ ( → Accept/Reject → )

→ Manuscript\_Main\_v1.1.docx ( , )

**4단계:** 공동연구자도 검토를 요청한다

Manuscript\_Main\_v1.1.docx → A, B

5단계: 두 공동연구자가 각각 수정한다

A: Manuscript\_Main\_v1.1.1\_KSM.docx

B: Manuscript\_Main\_v1.1.2\_LHJ.docx

6단계: 학생이 모든 피드백을 통합한다

→ →

→ Manuscript\_Main\_v1.2.docx ( )

이런 식으로 계속 순환하면서 논문이 발전한다.

---

## 실전 예시: 논문 폴더 완성본

자, 이제 모든 것을 종합해서 완성된 논문 폴더가 어떻게 생겼는지 보자:

paper\_ASD\_WGS\_2024/

```
Manuscript_ASD_WGS_Main_v3.docx (← )
Manuscript_ASD_WGS_Methods_v2.docx (← )
Supplementary_Information_v1.docx (← )
References_ASD_WGS.enl
References_ASD_WGS.Data/
```

```
old/
  Manuscript_ASD_WGS_Main_v1.docx
  Manuscript_ASD_WGS_Main_v2.docx
  Manuscript_ASD_WGS_Methods_v1.docx
```

```
Figures/
  Figure1_v3.pdf (← )
  Figure2_v2.pdf (← )
  Figure3_v1.pdf (← )
  Figure4_v1.pdf (← )
old/
  Figure1_v1.pdf
  Figure1_v2.pdf
  Figure2_v1.pdf
```

```
Supplementary_Figures/
  FigureS1_v1.pdf (← )
  FigureS2_v2.pdf (← )
  FigureS3_v1.pdf (← )
old/
  FigureS2_v1.pdf
```

```
Supplementary_Tables/
  TableS1_v1.xlsx (← )
  TableS2_v2.xlsx (← )
  TableS3_v1.xlsx (← )
old/
  TableS2_v1.xlsx
```

```
Scripts/  
  Figure1.R  
  Figure2.R  
  Figure3.R  
  FigureS1.R  
  FigureS2.R  
  TableS1.R  
  TableS2.R  
  utils.R
```

이 폴더를 보면 어떤 파일이 최신인지 한눈에 알 수 있고, Figure와 Table이 몇 개인지 명확하며, 누구에게나 공유하기 쉽고, 저널 제출 준비가 간단하다.

---

## 이 부분에서 학생들이 하는 흔한 실수들

### 파일 관리 실수

**작업 폴더와 논문 폴더를 구분하지 않음:** 논문 폴더에 테스트 코드, 실패한 분석 결과, 중간 산물들을 모두 넣으면 폴더가 금방 지저분해진다. 작업 폴더와 논문 폴더를 엄격히 분리하자. 논문 폴더에는 “현재 논문에 들어갈 것”만 넣고, 작업 과정의 모든 것은 작업 폴더에 남겨두고 최종 결과물만 논문 폴더로 복사한다.

**“final”이라는 단어 사용:** 파일 이름을 Figure1\_final.pdf, Figure1\_final2.pdf, Figure1\_final\_final.pdf로 짓는 것도 문제다. 항상 \_v1, \_v2, \_v3 형식으로 버전을 관리하자. “final”이라는 단어는 절대 쓰지 말자. 논문에서 “final”은 없다. 항상 수정이 필요하니까.

### 초기 준비 실수

**Table S1을 나중에 만들려고 함:** 논문을 거의 다 쓰고 나서 Table S1 (샘플 메타데이터)을 만들려고 하면, 샘플 정보를 다시 정리해야 하고 때로는 분석을 다시 해야 할 수도 있다. 논문 작성을 시작하는 첫날 TableS1\_v1.xlsx를 만들자. 처음에는 불완전해도 괜찮다. 논문을 쓰면서 계속 업데이트하면 된다.

**드롭박스를 사용하지 않음:** 로컬 컴퓨터에만 파일을 저장하면, 다른 저자들에게 파일을 매번 이메일로 보내야 하고 버전 관리가 엉망이 된다. 논문 폴더를 만들자마자 드롭박스에 올리고 지도교수와 공유하자. 그러면 파일을 업데이트할 때마다 자동으로 동기화되고, 다른 저자들은 항상 최신 버전을 볼 수 있다.

**모든 것을 하나의 파일에 작성:** Title부터 References, Methods, Supplementary까지 모든 것을 하나의 파일에 쓰면 파일이 너무 길어지고 편집이 어렵다. Main manuscript, Methods, Supplementary를 세 개의 분리된 파일로 작성하자. 이렇게 하면 각 부분에 집중할 수 있고, 나중에 저널 제출 시에도 편리하다.

### 협업 실수

**Track Changes를 사용하지 않음:** 다른 저자가 Track Changes 없이 원고를 직접 고쳐서 보내면, 1저자는 무엇이 바뀌었는지 알 수 없다. 협업 초기에 모든 팀원에게 “반드시 Track Changes를 켜고 수정해주세요”라고 명확히 안내한다.

**동시에 같은 파일 편집:** 여러 저자가 동시에 같은 파일을 받아서 수정하면 버전 충돌이 발생할 수 있다. 한 사람씩 순차적으로 검토하도록 요청하거나, 서로 다른 부분(예: 한 사람은 Introduction, 다른 사람은 Discussion)을 맡기거나, 명확히 다른 버전 번호를 부여한다(v1.0.1\_A, v1.0.2\_B).

**해결된 코멘트를 삭제하지 않음:** 이미 수정한 부분의 코멘트를 삭제하지 않으면 파일에 수십 개의 코멘트가 쌓여서 혼란스럽다. 새 버전을 만들 때마다 해결된 코멘트는 과감히 삭제하자. 필요하면 “v1.1에서 수정 완료” 같은 메모를 남기고 삭제한다.

**“최종 버전” 여러 개:** 여러 사람이 각자 “최종 버전”이라고 생각하는 파일을 만들면, 어떤 것이 진짜인지 알 수 없다. 처음부터 “1저자가 공식 버전을 관리한다”는 규칙을 명확히 하고, 이니셜 없는 파일만 공식 버전으로 인정한다.

## 학생들이 자주 묻는 질문

### 언제 논문 폴더를 만들어야 하나?

지금 당장! 많은 학생들이 “분석이 다 끝나면 논문 폴더를 만들어야지”라고 생각한다. 하지만 이것은 잘못된 생각이다. 논문 폴더는 분석을 시작할 때 함께 만들어야 한다.

분석하면서 동시에 정리할 수 있다. 좋은 결과가 나올 때마다 바로 논문 폴더에 정리하면, 나중에 “어떤 결과가 어디 있더라?” 하고 헤맬 일이 없다. Table S1을 먼저 만들 수 있다. Supplementary Table S1 (샘플 메타데이터)은 초반에 만들어야 한다. Methods를 바로 작성할 수 있다. 실험이나 분석을 하자마자 Methods를 쓰면, 세부 사항을 잊어버리지 않는다. “음... 6개월 전에 이 분석을 어떻게 했더라?” 하고 기억을 더듬을 필요가 없다. 논문 쓰기가 덜 막막하다. 논문 폴더가 미리 준비되어 있고, Table S1과 Methods가 어느 정도 작성되어 있으면, 나중에 본격적으로 논문을 쓸 때 심리적 부담이 훨씬 줄어든다.

### 어느 시점에 저널을 결정하는 것이 좋을까?

연구를 시작할 때 지도교수와 논의하면 좋다. Figure가 2개 정도 준비되면, 지도교수가 “이제 논문을 쓸 때가 되었다”라고 할 텐데, 그때 타겟 저널을 지도교수와 상의하고 대략 정하는 것이 좋다. 타겟 저널이 정해지면 그 저널의 Author Guidelines를 확인하고, Figure 개수 제한, 단어 수 제한, 포맷 요구사항 등을 미리 파악할 수 있다.

### 논문 전체 분량은 어떤 기준으로 정하나?

분량은 투고하고자 하는 저널의 가이드라인에서 확인하고, 그 분량에 맞춰서 작성한다. 다만 분량을 맞춰서 작성한다고 생각하지 말고, 우선은 글을 쓰자. 이후에 문장을 교정하고 재배치하면서 분량을 조절하면 된다. 처음부터 “3,000단어 안에 써야 해”라고 생각하면 글쓰기가 위축된다. 먼저 하고 싶은 말을 다 쓰고, 나중에 다듬자.

### Figure를 먼저 그리고 본문을 쓰는 것이 효율적인가, 아니면 본문을 먼저 작성한 뒤 Figure를 그리는 것이 좋을까?

일반적으로는 Figure를 작성하고, 이후에 본문을 쓰는 것이 좋다. Figure가 있으면 Results를 쓸 때 “이 Figure는 이런 것을 보여준다”라고 설명할 수 있기 때문이다. 그러나 본문을 먼저 스토리처럼 작성하고, 그 흐름에 맞춰 결과를 배치하면서 Figure를 작성할 수도 있다. 그 과정에서 본문의 내용을 정확하게 기입하며 수정할 수 있다. 본인에게 맞는 방식을 연습해보는 것이 좋다.

### Figure 작업 시 사용하는 툴(예: Illustrator)은 새로 배워야 하나?

우리 연구실은 일러스트레이터로 Figure를 작업한다. 일러스트레이터 작업 능력은 필수다. R이나 Python으로 그래프를 그린 후, 일러스트레이터에서 레이아웃을 조정하고 레이블을 추가하는 방식으로 작업한다. 처음에는 어렵게 느껴지지만, 몇 번 연습하면 금방 익숙해진다.

### Main Figure와 Supplementary Figure를 구분할 때, 어떤 기준으로 배치 결정을 하나?

처음에는 Main Figure를 염두에 두고 Figure를 준비한다. 그렇게 그리다 보면, 스토리 전개상 우선순위가 높은 Figure들이 선별된다. 그렇게 주전과 후보 선수를 선별해 가며 Figure를 그리면 된다. 논문의 핵심 메시지를 전달하는 데 꼭 필요한 Figure는 Main Figure로, 추가적인 검증이나 세부 사항을 보여주는 Figure는 Supplementary Figure로 분류한다.

---

이제 논문 작성을 시작할 준비가 되었다. 논문 폴더를 만들고, 드롭박스에 올리고, 지도교수와 공유하자. 그리고 Table S1을 만드는 것부터 시작하자. 작은 한 걸음이 첫 논문으로 이어질 것이다.

## 3장. Figure 작성하기 — 시각화로 생각하기

논문을 쓸 때 가장 먼저 해야 할 일은 무엇일까? Introduction을 쓰는 것? 방법론을 정리하는 것? 아니다. Figure를 그리는 것이다.

보통 우리는 데이터 분석을 하고 결과가 나오면, 그 즉시 시각화를 한다. 생선된 플롯과 결과를 정리해서 미팅 슬라이드를 만든다. 이 슬라이드에서 Figure에 들어갈 내용을 구성해서, Figure 파일을 작성해본다.

처음 그리는 Figure는 완벽하지 않아도 된다. 오히려 완벽하지 않은 상태에서 Figure를 그려보는 것이 논문 작성의 핵심이다. Figure를 그리는 과정은 단순히 '그림을 만드는 것'이 아니라, **논문의 구조를 설계하고, 논리를 점검하고, 부족한 부분을 발견하는 과정**이다.

많은 학생들이 “결과가 완벽해지면 그때 Figure를 그리겠다”고 생각한다. 하지만 이것은 논문 쓰기를 10배 더 어렵게 만드는 접근이다. Figure는 논문의 골격이다. 골격 없이 살을 붙이려고 하면, 결국 무너진다.

---

## Figure는 논문의 설계도다

Figure를 먼저 그리면, 논문의 전체 구조가 눈에 보인다. 우리는 보통 하나의 Results 섹션에 하나의 Figure를 만든다. 보통 4개에서 7개 사이이다. 먼저 Figure 1부터 Figure 7까지 대략적인 흐름을 스케치해보자. 각 Figure가 어떤 질문에 답하고, 어떤 데이터를 보여줄지 정하는 것이다.

이 단계에서 모든 패널이 완성될 필요는 없다. 어떤 패널은 “아직 분석 중”이어도 괜찮다. 중요한 것은 **Figure의 전체 구조를 먼저 잡는 것**이다.

예를 들어, 자폐 스펙트럼장애(Autism Spectrum Disorder, ASD) 유전학 연구를 한다고 가정해보자 (Kim et al. 2025, Genome Medicine). 논문의 핵심 질문이 “de novo variant가 ASD 표현형에 미치는 영향은 가족 내 유전적 배경을 고려했을 때 어떻게 달라지는가?”라면, Figure 구성은 다음과 같이 설계할 수 있다:

- **Figure 1:** 가족 내 표현형 편차(WFSD)의 개념과 측정 방법
- **Figure 2:** WFSD를 사용한 유전자 발견 및 pathway 분석
- **Figure 3:** De novo variant의 신경발달 프로파일에 대한 효과 크기
- **Figure 4:** 가족 간 표현형 가변성이 큰 유전자 발굴

이렇게 Figure의 골격을 먼저 세우면, 논문의 논리가 명확해진다. 각 Figure가 하나의 주장을 담고, 그 주장들이 순서대로 연결되어 전체 스토리를 구성한다.

---

## 열개부터 그려라: 초안 Figure의 힘

처음 논문을 쓰는 학생들이 가장 많이 하는 실수는 **완벽한 결과가 나올 때까지 Figure를 그리지 않는 것**이다. “이 분석이 끝나면”, “저 데이터가 추가되면” 하면서 Figure 작성을 미룬다. 그러다 보면 논문 쓰기 자체가 막막해진다.

Figure는 **열개(scaffold)**부터 그려야 한다. 모든 패널이 완성되지 않아도 괜찮다. 빈 공간에 “Panel C: 분석 진행 중”이라고 적어두고, 나머지 패널부터 채워나가면 된다.

초안 Figure를 만들면, 논문 작성이 빠르게 진행된다. 그렇게 짜여진 구성에서 어떤 결과가 부족한지 즉시 보인다. 하나의 Figure 안에 존재하는 패널들 간 논리적 연결이 자연스러워지 점검할 수 있다. 지도교수나 공동저자와 논의할 구체적인 자료가 생긴다.

예를 들어, Figure 1의 초안을 그릴 때 Panel A (부모-자녀 SRS 상관관계), Panel B (WFSD 계산 방법)는 이미 완성했지만, Panel C와 D는 아직 분석 중일 수 있다. 그래도 괜찮다. 빈 패널에 placeholder를 넣어두고 전체 구조를 먼저 확인하는 것이 중요하다.

```
library(ggplot2)
library(cowplot)

# Panel A: plot
pA <- ggplot(family_data, aes(parent_SRS, child_SRS, color = group)) +
  geom_point() +
  theme_bw() +
  labs(title = "A. Correlation of SRS scores within family")
```

```

# Panel B: schematic
pB <- ggplot() +
  theme_void() +
  annotate("text", x = 0.5, y = 0.5, label = "B. WFSB calculation", size = 5)

# Panel C D: Placeholder ( )
placeholder_C <- ggplot() +
  theme_void() +
  annotate("text", x = 0.5, y = 0.5,
    label = "C. SRS deviation by genetic subgroup\\n( )",
    size = 4, color = "grey60")

placeholder_D <- ggplot() +
  theme_void() +
  annotate("text", x = 0.5, y = 0.5,
    label = "D. SRS deviation by gene category\\n( )",
    size = 4, color = "grey60")

# 4
fig1_draft <- plot_grid(pA, pB, placeholder_C, placeholder_D,
  ncol = 2, align = "hv")

ggsave("Figure1_draft.pdf", fig1_draft, width = 10, height = 8)

```

이렇게 초안을 만들면, 전체 Figure의 균형과 흐름을 미리 볼 수 있다. Panel C와 D의 분석이 완료되면, placeholder를 실제 plot으로 교체하기만 하면 된다.

## Figure의 구성요소

과학 논문의 Figure는 대부분 **multi-panel** 구조를 가진다. 예를 들어 Figure 1은 Figure 1A, 1B, 1C, 1D처럼 여러 패널로 구성된다. 각 Figure는 결과 섹션과 1:1로 대응하며, 섹션의 제목은 Figure의 결론을 요약한 문장으로 한다. 즉, Figure 1은 섹션 1의 결과를 시각화한 구조이고, 그 안의 A~D 패널은 스토리를 단계적으로 전개한다. 논문을 쓰기 전, 먼저 **4-6개의 Figure를 설계**한다. 이때 각 Figure가 어떤 질문에 답하고, 어떤 데이터를 보여줄지를 명확히 정리하면 논문의 전체 흐름이 자연스럽게 잡힌다.

Kim et al. 2025 논문의 Figure 1을 보자:

이 Figure는 “가족 내 표현형 편차를 통해 de novo variant의 효과를 더 정확하게 측정할 수 있다”는 하나의 메시지를 담고 있다:

- **Panel A:** 부모와 자녀의 SRS 점수 간 상관관계를 보여준다
- **Panel B:** WFSB (within-family standard deviation) 계산 방법을 도식화한다
- **Panel C:** 유전자형 subgroup별로 SRS deviation 분포를 비교한다
- **Panel D:** ASD/DD 관련 유전자 카테고리별로 효과 크기를 총화한다

각 패널이 독립적으로 존재하는 것이 아니라, A → B → C → D 순서로 논리가 전개된다. 이것이 좋은 multi-panel Figure의 특징이다.

**섹션 제목 예시:** – ✕ 나쁜 예: “Results 1: SRS scores and de novo variants” – □ 좋은 예: “Family-based analyses explain varying phenotype outcomes of a de novo mutation”

섹션 제목이 Figure의 결론과 정확히 일치한다. 독자는 제목만 봐도 무엇을 보여줄지 예상할 수 있다.

## Figure를 그리기 전, 손으로 그려보면서, 스케치를 할 수 있다.

나는 논문을 쓸 때, 종종 손으로 그림을 그려서 Figure에 들어갈 구성요소를 구상한다.

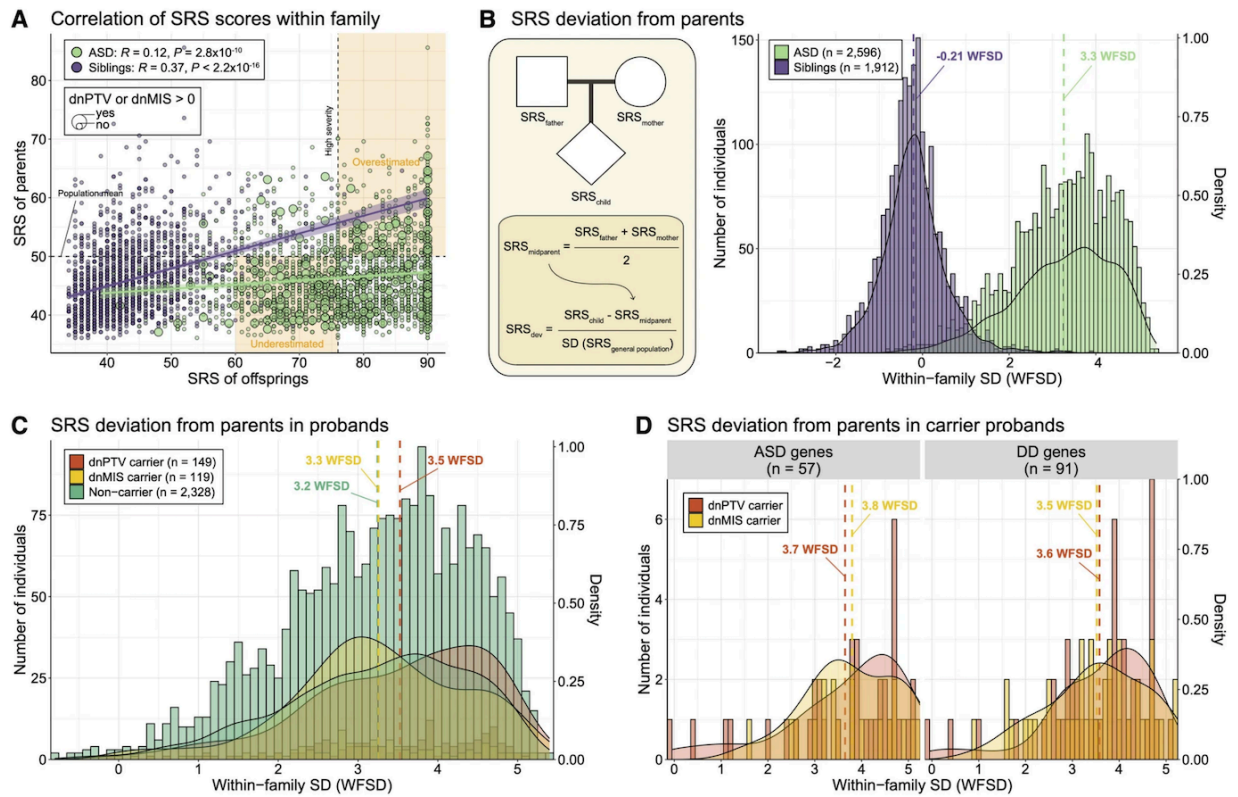


Figure 1: Figure 1

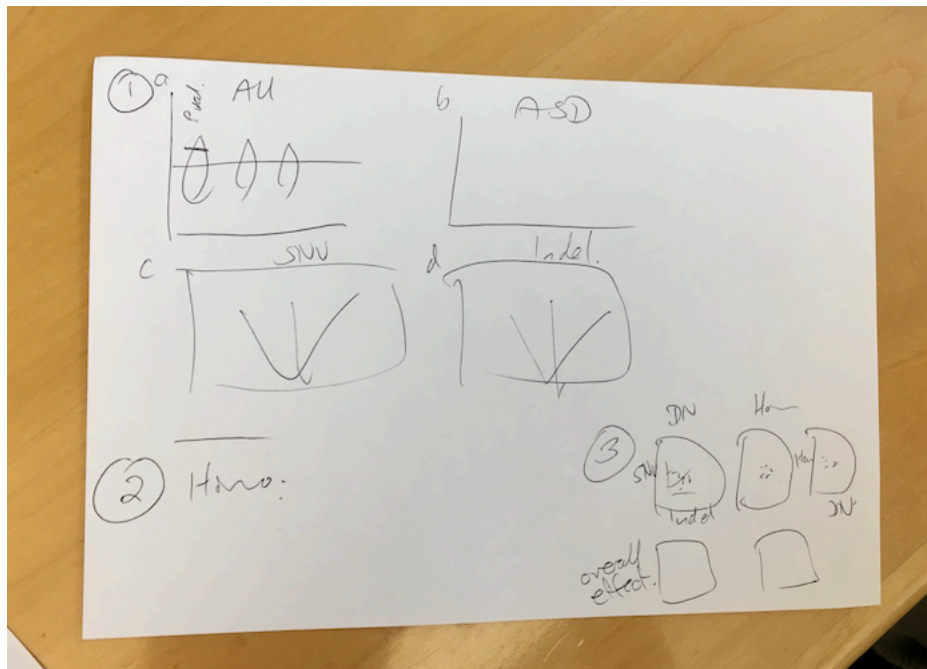


Figure 2: 손으로 Figure 그리기

낙서와 같은 이런 그림은 내용을 전개할 상상을 하게 만든다.

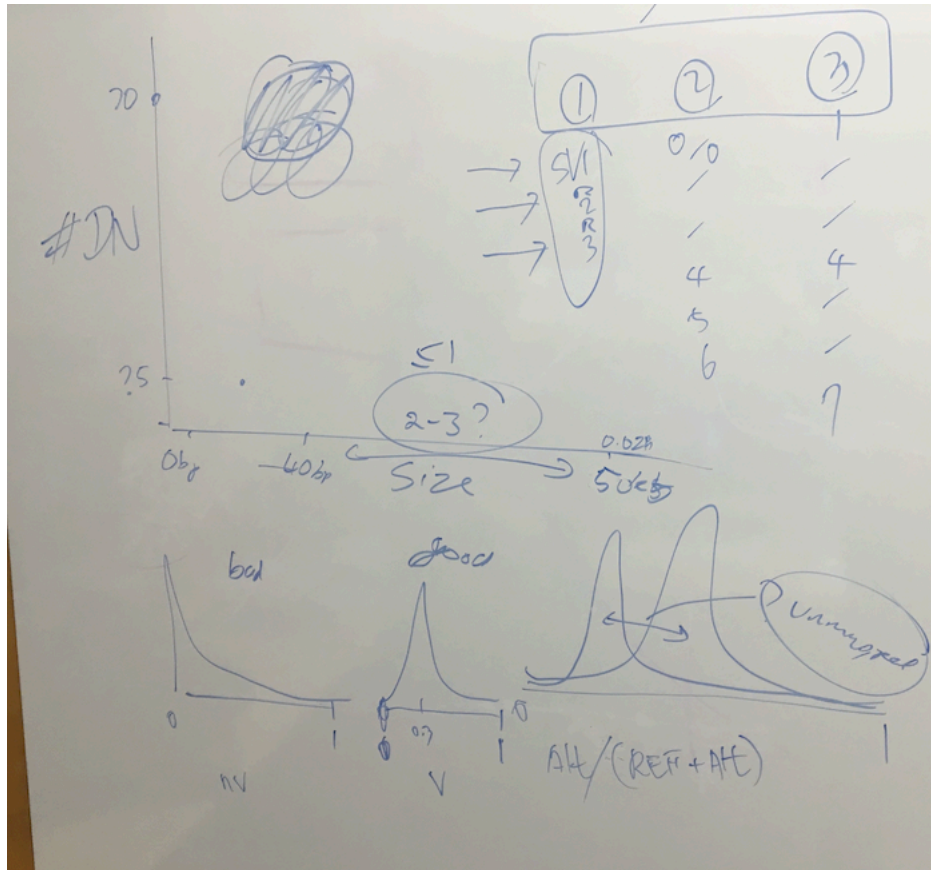


Figure 3: 손으로 Figure 그리기

MBTI가 극 N은 상상을 멈출 수 없어..

나는 학부수업에서 프로그래밍을 가르칠 때, descriptive stat 부분에서 꼭 손으로 그림을 그리는 방식을 가르친다. 이 과정은 데이터를 전달하는 방식을 생각해보게 만들고, plot을 만드는 접근을 쉽게 만든다. Script를 작성할 때도 도움이 된다. 요즘은 이런 그림을 GPT에게 보여주고, input 파일을 올려주면, 적절한 코드를 작성해주기도 한다. 세상이 참 좋아졌다.

## 멀티패널은 꼭 Script로 그려야 한다.

처음 논문을 쓰는 학생들이 가장 흔히 하는 실수는 분석 결과를 여러 장 출력해 PowerPoint나 Illustrator에서 조합하는 것이다. 제발, 절대로 이렇게 하지 말아달라. 이 방식은 당장은 편하지만, 여백·폰트·비율이 제각각이라 최종 완성까지 10배 이상 시간이 걸린다.

따라서 Figure는 처음부터 **script** → **plot** → **Illustrator** 편집 순서로 제작해야 한다. R의 ggplot2와 cowplot을 활용하면 모든 패널을 동일한 환경에서 출력할 수 있다. 여기서 중요한 것은 **재현성과 통일성**이다.

## 예시: 기본 multi-panel Figure 작성

```
library(ggplot2)
library(cowplot)
```

```

# theme ( )
theme_paper <- function() {
  theme_bw(base_size = 10) +
    theme(
      panel.grid.minor = element_blank(),
      plot.title = element_text(face = "bold", size = 11),
      axis.title = element_text(size = 9),
      legend.position = "right"
    )
}

# Panel A: Scatter plot
pA <- ggplot(family_data, aes(parent_SRS, child_SRS, color = dnv_status)) +
  geom_point(alpha = 0.6, size = 1.5) +
  geom_smooth(method = "lm", se = TRUE) +
  scale_color_manual(values = c("green4", "orange2", "purple3")) +
  labs(title = "A",
       x = "SRS of parents",
       y = "SRS of offsprings") +
  theme_paper()

# Panel B: Distribution histogram
pB <- ggplot(wfsd_data, aes(WFSD, fill = group)) +
  geom_histogram(alpha = 0.7, position = "identity", bins = 30) +
  scale_fill_manual(values = c("purple3", "green3")) +
  labs(title = "B",
       x = "Within-family SD (WFSD)",
       y = "Number of individuals") +
  theme_paper()

# Panel C: Boxplot by genetic subgroup
pC <- ggplot(genetic_subgroup, aes(group, WFSD, fill = group)) +
  geom_boxplot(outlier.size = 0.5) +
  scale_fill_manual(values = c("red3", "orange2", "green3")) +
  labs(title = "C",
       x = "",
       y = "Within-family SD (WFSD)") +
  theme_paper() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Panel D: Gene category stratification
pD <- ggplot(gene_category, aes(gene_type, WFSD, fill = variant_type)) +
  geom_bar(stat = "identity", position = "dodge") +
  scale_fill_manual(values = c("darkred", "orange2")) +
  labs(title = "D",
       x = "",
       y = "Mean WFSD") +
  theme_paper()

# 4 2x2 grid
fig1 <- plot_grid(pA, pB, pC, pD,
                  ncol = 2, nrow = 2,
                  align = "hv",
                  rel_widths = c(1, 1),

```

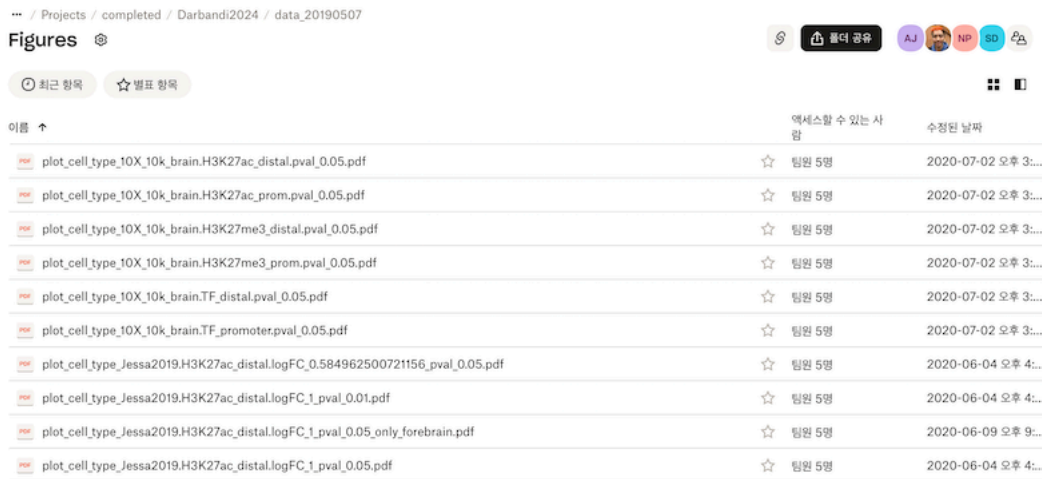
```
rel_heights = c(1, 1))
```

```
# PDF
```

```
ggsave("Figure1_WFSD.pdf", fig1, width = 8, height = 7, dpi = 300)
```

이 코드의 핵심은: 1. 공통 theme 함수 (theme\_paper())로 모든 패널의 스타일을 통일 2. 일관된 색상 팔레트 사용 3. script 한 번으로 전체 Figure 생성 — 수정이 필요하면 코드만 고치면 됨

믿거나 말거나, 아래는 우리가 사용한 작업 폴더인데, 여기엔 수천개의 plot이 들어있다. 손으로 작업할 수 없다. 컴퓨터가 하게 하라.



이름 ↑	액세스할 수 있는 사람	수정된 날짜
plot_cell_type_10X_10k_brain.H3K27ac_distal.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_10X_10k_brain.H3K27ac_prom.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_10X_10k_brain.H3K27me3_distal.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_10X_10k_brain.H3K27me3_prom.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_10X_10k_brain.TF_distal.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_10X_10k_brain.TF_promoter.pval_0.05.pdf	☆ 팀원 5명	2020-07-02 오후 3:...
plot_cell_type_Jessa2019.H3K27ac_distal.logFC_0.584962500721156_pval_0.05.pdf	☆ 팀원 5명	2020-06-04 오후 4:...
plot_cell_type_Jessa2019.H3K27ac_distal.logFC_1_pval_0.01.pdf	☆ 팀원 5명	2020-06-04 오후 4:...
plot_cell_type_Jessa2019.H3K27ac_distal.logFC_1_pval_0.05_only_forebrain.pdf	☆ 팀원 5명	2020-06-09 오후 9:...
plot_cell_type_Jessa2019.H3K27ac_distal.logFC_1_pval_0.05.pdf	☆ 팀원 5명	2020-06-04 오후 4:...

Figure 4: Many May plots

Figure는 자주, 여러번 수정할 수 있다.

Figure를 한 번 그리고 끝나는 것이 아니다. 논문을 쓰는 과정은 곧 Figure를 수정하는 과정이다.

분석 결과가 업데이트되면? → Figure를 다시 그린다.

패널 순서가 어색하면? → 재배열한다.

지도교수가 피드백을 주면? → 즉시 반영한다.

이 과정이 매일 반복된다. 그래서 script 기반으로 작업하는 것이 중요하다. PowerPoint에서 손으로 조합한 Figure는 수정할 때마다 처음부터 다시 만들어야 하지만, script로 만든 Figure는 코드 몇 줄만 바꾸면 된다.

## Figure versioning 예시

```
# Version 1: draft
```

```
ggsave("Figure1_v1_draft.pdf", fig1, width = 8, height = 7)
```

```
# Version 2:
```

```
pA <- pA + scale_color_manual(values = c("#2E8B57", "#FF8C00", "#9370DB"))
```

```
ggsave("Figure1_v2_color_revised.pdf", fig1, width = 8, height = 7)
```

```
# Version 3:
```

```
pA <- pA + coord_cartesian(xlim = c(30, 90), ylim = c(30, 90))
```

```
ggsave("Figure1_v3_axis_adjusted.pdf", fig1, width = 8, height = 7)
```

```
# Final version
```

```
ggsave("Figure1_final.pdf", fig1, width = 8, height = 7)
```

Figure를 매일 수정하면서, 논문의 논리가 점점 명확해진다. 처음에는 보이지 않던 문제점이 Figure를 통해 드러나고, 그것을 수정하면서 더 나은 스토리를 만들어간다.

## Illustrator를 사용한 추가 편집

R 스크립트로 Figure의 기본 구조를 완성했다면, 최종 단계는 **Adobe Illustrator**를 사용한 세밀한 편집이다. Illustrator는 벡터 기반 편집 도구로, 출판용 고품질 Figure를 만드는 데 필수적이다.

**1. 텍스트 통일** - Select > Object > Text Objects를 사용해 모든 텍스트를 선택한다 - 폰트 크기를 통일한다 (보통 8-10pt) - 저널 가이드라인에서 권장하는 폰트 크기가 명시되어 있다면 이를 따른다 - 폰트는 Arial 또는 Helvetica 사용 - 축 제목, 범례, panel label 모두 일관되게 유지한다

**텍스트 크기와 가독성**: 원고를 A4로 출력했을 때 글자가 충분히 읽을 수 있어야 한다. Figure를 작성하는 여러분은 대부분 20-30대의 좋은 시력을 가졌지만, 동료 평가를 하는 심사자는 노안이 있거나 작은 노트북 화면으로 Figure를 볼 수 있다. 따라서 불필요한 여백을 줄이고 공간을 최대한 활용하여 글자를 충분히 크게 표현하자.

**2. 색상 통일** - 같은 의미를 가진 요소는 같은 색상을 사용한다 - 예: dnPTV는 항상 빨간색, dnMIS는 주황색, non-carrier는 초록색 - Color picker를 사용해 RGB 값을 정확히 맞춘다 - 자주 사용하는 색상은 Swatch 패널에 저장하면 편리하다 - 같은 그룹을 표현할 때 미리 색상을 지정해두자

**3. 정렬과 여백** - 패널 간 여백을 균등하게 배치한다 - Align 패널을 사용해 객체들을 정확히 정렬한다 - 텍스트 정렬 시: 텍스트를 선택한 후 좌우 정렬 옵션을 선택하고, 여러 텍스트 객체를 동시에 선택하여 한 번에 정렬할 수 있다 - Panel label (A, B, C, D)의 위치를 일관되게 유지한다

**4. 파일 저장** - 작업 파일은 .ai 형식으로 저장한다 (수정 가능) - 최종 제출용은 .pdf 또는 .tiff 형식으로 export한다 - 해상도는 최소 300 dpi, 가능하면 600 dpi를 유지한다 - 색상 모드는 RGB (온라인용) 또는 CMYK (인쇄용)로 저널 규정에 따라 선택한다

## 실전 팁

**Panel label 배치**: 각 패널의 왼쪽 상단에 “A”, “B”, “C”, “D”를 배치한다. 폰트는 가능하면, Arial, 12pt. 제목은 Bold로. Text는 main, axis, label 간의 규칙이 있어야 한다. 색상은 당연히 검은색 (ggplot은 기본값으로 회색을 주기도 한다). 여백은 패널 테두리에서 5mm 안쪽. 단 여백은 완전히 감각적인 부분이라, 자주 그려봐야 한다.

**단일세포 분석 데이터 처리**: UMAP 플롯과 같은 단일세포 분석 결과는 수천, 수만 개의 점들이 벡터로 저장되어 있어 Illustrator에서 파일을 열고 작업하기 매우 어렵다. 이런 경우 UMAP을 래스터화(rasterize)하거나 고해상도 이미지 파일(PNG, TIFF)로 변환하여 삽입하는 것이 효율적이다.

**Figure 저장 방법**: R에서는 ggsave() 함수를 사용하여 저장한다. PDF로 저장할 때 주의: 글자가 개별 문자로 저장되지 않고 **단어 단위로 저장**되어야 한다. 단어/문장이 하나의 객체로 저장되어야 Illustrator에서 편집이 용이하다.

권장 저장 형식: ggsave("figure1.pdf", width = 8, height = 6, device = cairo\_pdf) -> “cairo\_pdf” 이 옵션이 단어를 객체로 만든다.

**색상 통일 체크리스트**: - [ ] 같은 그룹은 같은 색상 사용 - [ ] **색맹 친화적 색상 팔레트 필수 사용** (ColorBrewer, viridis 패키지 권장) - [ ] 흑백 인쇄 시에도 구분 가능한지 확인 - [ ] 색각 이상 시뮬레이터로 검증 (예: Coblis, Color Oracle)

**심미적 감각 기르기**: Figure 작성시 심미적인 부분은 매우 중요하다. 그러나 이것은 후천적으로 학습하기 어렵다. 우리 연구실에 미술관을 매주 가는 학생의 심미적 능력을 보고 나는 매일 놀린다... 이걸 학습으로 잘 되지 않구나... 그래서 논문을 읽다가 예쁜 그림이 나오면, 저장하고 수집하면 좋다.

**동료 피드백의 중요성**: 그림을 그리면 상대방에게 보여주고, 서로 감상회를 하는 것도 좋다. 종종 “나는 당신의 감성을 이해할 수 없어...”라는 경우가 있다.

**여러개의 객체를 일러스트레이터에서 편집하기**: 객체를 선택하고, Select>Same>Appearance를 선택한다. 그러면 같은 객체가 선택된다.

---

## Figure 삽입하고 Figure Legend 작성하기

논문을 Word로 작성할 때, Figure는 본문과 별도로 관리하는 것이 일반적이다. 하지만 우리 연구실에서는 각 섹션 뒤에 바로 Figure를 삽입한다. 논문 투고 시에도 저널 가이드라인에 “Figure를 본문 맨 뒤에 넣거나 별도 파일로 제출해야 한다”는 명시적 요구사항이 없다면, 본문의 각 섹션 바로 뒤에 Figure를 배치한다.

### Figure 파일 삽입 형식

Figure는 PNG 또는 PDF 형태로 본문에 삽입할 수 있다. 하지만 **PDF 파일은 삽입하지 않는 것을 권장한다**. PDF 파일은 벡터 정보를 포함하고 있어 원고 파일을 매우 느리게 만들 수 있다. 특히 복잡한 그래프가 여러 개 포함된 경우 Word 파일의 용량이 크게 증가하고 편집 속도가 현저히 느려진다.

따라서 Illustrator에서 **3000 픽셀 이상의 고해상도 PNG 파일로 export**한 후 해당 파일을 첨부한다. 이렇게 하면 출판 품질을 유지하면서도 원고 파일을 원활하게 관리할 수 있다.

### Figure Legend란?

Figure Legend는 Figure 아래에 위치하며, 해당 Figure가 무엇을 보여주는지 설명하는 텍스트이다. 좋은 Figure Legend는 독자가 본문을 읽지 않고도 Figure만으로 연구의 핵심 내용을 이해할 수 있도록 돕는다. Legend는 Figure의 전체 메시지를 전달하는 제목 문장과 각 패널(A, B, C, D 등)에 대한 상세 설명으로 구성된다.

### Figure Legend 작성 형식

Word 문서에서 Figure legend는 다음과 같이 작성한다:

Figure 1. Intrafamilial deviation of social responsive scale from parents.  
A) Correlation of social responsiveness scale (SRS) T-scores between offspring and parents within families. The effect of de novo variants (DNVs) may be underestimated when the offspring exhibit milder SRS severity but their parents have lower-than-average SRS. B) Calculation of within-family standard deviation (WFSD) for SRS in offspring from parents. WFSD is calculated by subtracting the mean parental SRS T-score from the offspring's score and dividing by the general population SD. C) SRS deviation distributions among genetic subgroups in ASD cases, showing greater deviations in carriers of de novo protein-truncating variant (dnPTV) and missense (dnMIS). D) SRS deviations stratified by gene category: ASD-/DD-associated genes.

### Figure Legend 작성 원칙

1. 첫 문장은 Figure 전체의 메시지 (Bold) – “**Figure 1. Intrafamilial deviation of social responsive scale from parents.**” – 이 문장은 Figure가 전달하는 핵심 메시지를 간결하게 요약한다 – Bold 처리하여 Figure 번호와 제목을 명확히 구분한다
2. 각 패널 설명은 알파벳 순서로 – 패널 순서는 좌에서 우로, 상에서 하로 읽는다 – (A) Panel A의 내용… – (B) Panel B의 내용… – 각 패널이 보여주는 데이터, 분석 방법, 주요 결과를 설명한다
3. 현재형 시제 사용 – Results 섹션은 과거형을 사용하지만, Figure Legend는 현재형을 사용한다 – □ “WFSD is calculated by…” – X “WFSD was calculated by…” – Figure는 현재 독자가 보고 있는 객관적 정보이므로 현재형으로 기술한다
4. 약어는 처음 나올 때 정의 – “social responsiveness scale (SRS)” – “within-family standard deviation (WFSD)” – 본문에서 이미 정의했다라도 Figure Legend에서 처음 나올 때는 다시 정의한다
5. 통계 정보 포함 (필요시) – “Error bars represent standard error of the mean.” – “ $P < 0.05$ ,  $P < 0.01$  (Student's t-test)” – “\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ ” – 사용된 통계 검정 방법과 유의수준 기호의 의미를 명시한다

6. 샘플 크기 명시 (중요한 경우) – “n = 50 per group” – “Error bars represent SEM from three independent experiments”

## 4장. Results — 데이터가 말하게 하기

Results 섹션을 쓸 때 가장 먼저 해야 할 일은 무엇일까? 많은 학생들이 첫 번째 분석 결과부터 순서대로 적기 시작한다. 하지만 좋은 Results는 **전체 구조를 먼저 설계**하고 시작해야 한다. 이번 장에서는 Results 섹션 전체의 뼈대를 어떻게 짜야 하는지, 각 subsection을 어떻게 배치하고 연결해야 하는지를 다룬다. 문장 레벨의 작성 기술은 다음 장에서 다룰 것이고, 여기서는 큰 그림에 집중한다.

### Results의 핵심 원칙: 데이터가 주인공이다

#### Results vs Methods vs Discussion

Results를 쓰기 전에 먼저 명확히 해야 할 것이 있다. Results, Methods, Discussion은 각각 다른 역할을 한다.

섹션	질문	예시
Methods	어떻게 했는가?	We performed NMF clustering using proteomic, phosphoproteomic, and acetylproteomic data
Results	무엇을 발견했는가?	We identified five multiomics subtypes characterized by distinct molecular pathways
Discussion	무엇을 의미하는가?	These subtypes have important implications for personalized treatment strategies

Results는 오직 “**무엇을 발견했는가**”에만 집중해야 한다.

#### 실제 논문에서의 예시

우리가 작성한 한국인 NSCLC 환자의 proteogenomics 연구 논문에서 가져온 예다 (Song et al. 2024, Nature Communications).

잘못된 Results (Methods가 섞임):

To establish comprehensive multiomics subtypes, we integrated proteomic, phosphoproteomic, and acetylproteomic data and conducted non-negative matrix factorization (NMF) clustering to identify multiomics subtypes.

이것은 거의 전부 Methods에 들어가야 할 내용이다. “To establish”, “we integrated”, “we conducted” – 이런 표현들은 모두 **과정**을 설명하는 것이다.

올바른 Results:

We identified five multiomics subtypes: metabolic (Subtype 1), alveolar-like (Subtype 2), proliferative (Subtype 3), hypoxic (Subtype 4), and immunogenic (Subtype 5), characterized based on genetic mutations, clinical phenotypes, and molecular pathways.

**발견**에만 집중한다. 어떻게 분석했는지는 Methods에 있다.

내가 학생들에게 자주 남기는 코멘트다:

“이런 내용은 쓰지 않아도 되고요. 서브타입을 정의한 기준은 Methods에 넣으면 됩니다.”

“바로 이전 문장들도 많이 Methods 같아요 ~ 수정을 해주세요. NMF clustering을 어떻게 수행했는지가 아니라, 무엇을 발견했는지를 써야 합니다.”

“사용했던 multiomics 데이터를 Table S1으로 정리하기. 현재 산발적으로 언급했는데, 이것 통합해서 넣으면 좋겠습니다. 그리고 Method에도 데이터 생성 과정을 기술해주세요.”

#### 원칙: Results에는 발견만, Methods에는 방법만

Results에 적합한 동사: – showed, demonstrated, revealed – observed, identified, found – consisted of, comprised, included – differed, increased, decreased

Results에 부적합한 동사 (Methods용): – performed, conducted, analyzed – established, constructed, developed – optimized, benchmarked, validated – collected, generated, processed

분석 도구의 사용법, parameter 설정, input/output 형식 등은 모두 Methods 섹션에 포함해야 한다.

## 섹션들을 구성하기

### 섹션 제목은 발견을 담아야 한다

각 subsection의 제목이 너무 일반적이면 독자의 관심을 끌기 어렵다.

평범한 제목: – “Subtype identification” – “Immune cell analysis” – “Survival analysis”

발견을 담은 제목 (Song et al. 2024, Nature Communications에서): – “A NSCLC subtype associated with poor prognosis and frequent metastasis” – “Proteogenomic features underlying whole-genome doubling in NSCLC subtypes” – “Multiomics profiling of neoantigens and immune clusters”

### 좋은 제목의 특징:

1. 구체적이다 – “Subtype identification”이 아니라 “A NSCLC subtype associated with poor prognosis”
2. 발견을 담는다 – 분석 방법이 아니라 무엇을 발견했는지
3. 관계를 보여준다 – “A affects B” 혹은 “A underlying B” 형태로 연관성 제시

제목만 읽어도 무엇을 발견했는지 알 수 있어야 한다.

### 섹션 간 연결성: 논문 제목과의 일관성

각 섹션이 자연스럽게 이어져야 한다. 독자가 “왜 갑자기 이 이야기?”라고 느끼면 안 된다.

내가 학생에게 남긴 코멘트:

“여기도 마찬가지로인데, 논문의 제목도 ‘NSCLC subtypes’이고, 첫 섹션도 ‘Identification of subtypes’로 시작했는데요. 갑자기 2번째, 3번째 섹션에서 뜬금없이 일반적인 immune cell 분석을 이야기합니다. 심지어 이 내용은 이미 다른 연구들에서 다 나온 이야기들이고요. 우리가 논문을 투고하면, 에디터가 2번째 혹은 3번째 페이지까지 보고 ‘이 연구 이미 다 한거네. 왜 또 해?’라고 할거고, 바로 리젝될겁니다. 다르게 없어요. 저라면, 2,3번을 합치고 바로 ‘Subtype-specific immune landscape’라는 주제로 두번째 섹션을 시작합니다. 그래야 1번째에서 웅장하게 시작한 Subtypes 이야기가 이어지죠.”

### Song et al., 2024 논문의 섹션 구조 분석:

논문 제목: “Proteogenomic analysis reveals non-small cell lung cancer **subtypes** predicting chromosome instability, and tumor microenvironment”

Results 섹션 구성: 1. Identification of **subtypes** in NSCLC patients by multiomics 2. A NSCLC **subtype** associated with poor prognosis and frequent metastasis 3. Cellular landscape of the five **subtypes** of NSCLC 4. Proteogenomic features underlying whole-genome doubling in NSCLC **subtypes** 5. Heterogeneous immune landscapes in NSCLC 6. Multiomics profiling of neoantigens and immune clusters

모든 섹션이 “subtypes”라는 중심 주제를 유지한다. 첫 섹션에서 5개 서브타입을 소개하고, 이후 섹션들이 각 서브타입의 특성을 다양한 관점(예후, 세포 구성, WGD, 면역 환경)에서 다룬다.

### 섹션의 첫 문장이 중요하다

각 섹션의 첫 문장은 해당 섹션의 방향을 제시한다.

내가 학생에게 남긴 코멘트:

“각 섹션명을 적었는데, 이 섹션들을 이어서 읽으면 전혀 내용이 이어지지 않습니다. 첫번째 섹션은 'NSCLC multiomics subtypes'라고 했는데요. 그러고 바로 뒤 섹션은 'Kinase activity in cancer'라고 되었어요. Kinase가 암에서 중요하다는걸 모르는 사람이 있나요? 전혀 연결이 안되요. 우리가 발견한 subtypes와 kinase의 관계를 설명해야죠. 저만 이해를 못하는건지요?”

### Song et al., 2024의 좋은 예시:

섹션 2의 첫 문장:

To replicate our subtype classification, we utilized multiomics or proteomics data from 462 patients with NSCLC obtained from previous studies.

섹션 3의 첫 문장:

Exploring the tumor microenvironment is crucial for understanding the mechanisms underlying cancer progression and for developing effective therapeutic strategies to target not only cancer cells but also the surrounding microenvironment.

두 번째 섹션이 첫 번째 섹션의 발견(5개 서브타입)을 다른 코호트에서 검증하겠다고 명확히 밝히고, 세 번째 섹션은 “왜 이제 중앙 미세환경을 봐야 하는가”를 설명하면서 자연스럽게 연결된다.

각 섹션의 첫 문장에서 “왜 이 내용이 지금 필요한가”를 명확히 해야 한다.

### 독자의 시선으로 읽기

논문을 다 쓴 후, 각 섹션의 제목만 순서대로 읽어보자.

잘못된 흐름:

1. Multiomics subtype identification
2. Kinase activity analysis
3. Cell type analysis
4. Survival analysis

독자 반응: “2번이 왜 갑자기? 1번에서 서브타입 이야기하다가 왜 갑자기 kinase?”

좋은 흐름 (Song et al. 2024, Nature Communications):

1. Identification of subtypes in NSCLC patients by multiomics
2. A NSCLC subtype associated with poor prognosis and frequent metastasis
3. Cellular landscape of the five subtypes of NSCLC
4. Proteogenomic features underlying whole-genome doubling in NSCLC subtypes
5. Heterogeneous immune landscapes in NSCLC
6. Multiomics profiling of neoantigens and immune clusters

독자 반응: “1번에서 5개 서브타입 소개, 2번에서 특히 예후가 나쁜 서브타입 집중 분석, 3번에서 서브타입별 세포 특성, 4번에서 증식성 서브타입의 WGD 특징, 5-6번에서 면역 환경 분석. 논리적이네!”

---

## 문단 구성의 원칙

### 문단 한개는 하나의 완결된 아이디어

너무 짧은 문단은 대부분 앞뒤 문단과 합칠 수 있다.

“한가지 궁금한게 (manuscript 전반적으로), 문단이 짧은데요. 여기도 문장 2개가 한 문단이라... 특별한 이유가 있을지?”

1-2문장으로 끝나는 문단은 피하자. 하나의 발견이나 개념을 설명하는 데 보통 3-5문장이 필요하다.

**Song et al., 2024의 문단 구조 예시:**

[ 1: Subtype 4 - 6 ]

To examine the molecular characteristics of Subtype 4, we extracted distinct features (proteins, phosphorylation, or acetylation) of the subtypes in our NMF clustering analysis. We found that the majority of NMF features present in Subtype 4 were phosphorylated sites (96%, 178/186), indicating that phospho-kinase interactions are a major signature. Therefore, we investigated the kinase activity of this subtype using phosphoproteomic data. We found significant enrichment of two kinases, CSNK2A1 ( $\text{FDR} = 2.3 \times 10^{-7}$ ) and GSK3B ( $\text{FDR} = 1.9 \times 10^{-2}$ ), which are known to phosphorylate various proteins in the PI3K-AKT signaling pathway, in Subtype 4 compared to other subtypes. Upon evaluating the relationship between the activity and expression levels of significant kinases ( $P < 0.05$ ), we observed a moderate correlation. Survival analysis based on feature expression showed that most of the unfavorable prognostic factors were phosphorylated sites differentially expressed in Subtype 4 (91%, 104/114).

[ 2: SLK - 5 ]

Notably, STE20-like serine/threonine-protein kinase at serine 347 (SLK (S347)), a protein phosphorylated by CSNK2A1, was significantly upregulated in Subtype 4 (adjusted  $P = 8.0 \times 10^{-3}$ ) and associated with unfavorable prognostic features ( $P = 3.0 \times 10^{-6}$ ). In the combined CPTAC dataset, we also found increased phosphorylation of SLK (S347) in Subtype 4 and poor survival outcomes. SLK mediates apoptosis downstream of the ErbB2 and PI3K pathways and is activated in a CSNK2A1-dependent manner. Recent studies have reported that high SLK expression is associated with reduced overall survival in HER2-positive patients and in glioma. Collectively, these results suggest that SLK participates in tumor progression and could be a key marker specific to Subtype 4.

각 문단이 하나의 완결된 발견을 담고 있다. 첫 번째 문단은 Subtype 4의 전반적인 인산화 특징을, 두 번째 문단은 특정 마커(SLK)에 대한 상세 분석을 다룬다.

**각 문단의 첫 문장 = Topic Sentence**

문단의 첫 문장(topic sentence)이 해당 문단의 핵심을 담아야 한다.

독자는 각 문단의 첫 문장만 읽어도 대략적인 흐름을 파악할 수 있어야 한다.

Song et al., 2024의 “A NSCLC subtype associated with poor prognosis” 섹션에서 각 문단의 첫 문장만 추출해보면:

Paragraph 1: To examine the molecular characteristics of Subtype 4, we extracted distinct features...

Paragraph 2: Notably, STE20-like serine/threonine-protein kinase at serine 347 (SLK (S347))...

Paragraph 3: Furthermore, Subtype 4 showed a significant upregulation of leucine-rich repeat...

Paragraph 4: Subtype 4 included numerous prognostic features in the HIF-1 and PI3K-AKT signaling pathway...

Paragraph 5: We evaluated the clinical significance of Subtype 4 according to the survival rates...

각 문단의 첫 문장만 읽어도 “Subtype 4의 분자적 특징 → SLK 마커 → LRRFIP1 마커 → 신호전달 경로 → 생존율 분석” 순서로 분석이 진행되었음을 알 수 있다.

## 용어 사용의 정확성과 일관성

### 개념을 명확히 구분한다

비슷해 보이는 용어들도 과학적으로는 다른 의미를 가진다. 정확하게 구분해서 사용해야 한다.

내가 학생들에게 남긴 코멘트:

“이건 불필요한 설명이에요. 단순히 proteomics를 보겠다는게 아니라 마치 phosphoproteomics만 보겠다는 의도 같구요. 그러면 이 섹션의 첫 부분에는 왜 phosphoproteomics가 중요한지 설명이 나와야하는거죠.”

“그리고 나서 global proteome을 보고, 마지막으로 acetylome을 보겠다고 하고, PTM 분석을 들어가면 됩니다. 각각의 역할이 다르니까요.”

### Song et al., 2024가 구분하는 용어들:

**Neoantigen vs Cryptic MAP** – Neoantigen: 체세포 돌연변이에서 유래한 종양 특이적 항원 – Cryptic MAP: 비코딩/비주석 전사체에서 유래한 MHC class I 결합 펩타이드

논문에서 이 둘을 명확히 구분하여 사용한다:

We inferred 85,430 neoantigen candidates and 775 cryptic MAPs and annotated the origin of the cryptic MAPs based on the matched transcripts. Interestingly, only confirmed cryptic MAPs showed a strong positive correlation with improved survival, although the number of cryptic MAPs was notably low across patients.

**HTE vs CTE (면역 클러스터)** – HTE (hot-tumor-enriched): 면역세포가 풍부한 “뜨거운” 종양 – CTE (cold-tumor-enriched): 면역세포가 적은 “차가운” 종양

### 용어 일관성이 중요한 이유: 독자의 인지적 색인

논문은 단순히 데이터를 나열하는 문서가 아니다. 독자의 머릿속에 **개념의 지도**를 그려주는 일이다.

논문을 읽는 독자는 마치 제품 브로셔를 보듯, 핵심 용어를 통해 전체 내용을 이해한다. 특정 용어가 반복될 때마다 독자는 이전에 읽었던 맥락을 다시 떠올린다. 이것을 **인지적 색인(cognitive index)**이라고 부를 수 있다.

### Song et al., 2024의 좋은 예:

논문 전체에서 “multiomics subtypes”라는 용어를 일관되게 사용한다: – Abstract: “...reveals five molecular subtypes” – Introduction: “...to define the molecular subtypes of NSCLC” – Results 섹션 1: “Identification of subtypes in NSCLC patients by multiomics” – Results 섹션 2: “A NSCLC subtype associated with poor prognosis” – Results 섹션 3: “Cellular landscape of the five subtypes” – Discussion: “...the five NSCLC subtypes enriched for WGD, oncogenes...”

독자는 “subtypes”라는 단어를 볼 때마다 “아, 저 5개 그룹을 말하는구나”라고 즉시 이해한다.

### 핵심 용어는 논문 초반에 확립한다

독자의 머릿속 색인을 만들려면, 핵심 용어를 Results 첫 섹션에서 명확히 정의해야 한다.

Song et al., 2024의 첫 Results 섹션:

We identified five multiomics subtypes: metabolic (Subtype 1), alveolar-like (Subtype 2), proliferative (Subtype 3), hypoxic (Subtype 4), and immunogenic (Subtype 5), characterized based on genetic mutations, clinical phenotypes, and molecular pathways.

이 한 문장이 논문 전체의 **인지적 프레임워크**를 설정한다. 독자는 이후 “Subtype 3”을 볼 때마다 “proliferative한 그 그룹”을 떠올린다.

### 일관성이 깨지면 독자는 길을 잃는다

만약 같은 개념을 다르게 부르면: – 첫 섹션: “multiomics subtypes” – 중간 섹션: “molecular clusters”  
– 마지막 섹션: “patient groups”

독자는 “이게 같은 건가? 다른 건가?”라며 계속 뒤로 돌아가서 확인해야 한다. 인지적 부담이 커진다.

#### 용어 일관성 체크리스트:

논문을 다 쓴 후, 다음을 확인하자: 1. 핵심 개념을 나타내는 용어가 전체 논문에서 동일한가? 2. 약어를 한번 정의한 후 계속 사용하는가? 3. Figure legend에서도 동일한 용어를 사용하는가? 4. Supplementary Material에서도 일관된 용어를 쓰는가?

#### 용어를 통일한다

한 논문 내에서 같은 개념을 다른 용어로 표현하면 독자가 혼란스러워한다.

#### Fig vs Figure:

“Fig도 쓰는 것도 있고 Figure로 쓰는 것도 있는데, 통일을 시켜주면 좋겠습니다.”

#### Cell subtype vs Molecular subtype vs Genomic subtype:

“바로 아래 문단에는 ‘molecular subtype’이라는 표현을 썼고, 여기는 그냥 ‘subtype’인데, 통일을 하면 좋겠습니다.”

“저는 molecular subtype은 아닌것 같아요. 왜냐면, RNA, protein expression을 측정하건 아니니까요. 그래서 genomic subtype이라는 표현을 쓰는게 좋을것 같아요.”

#### Song et al., 2024가 잘한 점:

논문 전체에서 용어를 일관되게 사용한다: – Multiomics subtypes (일관되게 Subtype 1-5) – WGD (whole genome doubling) – TMB (tumor mutational burden) – HTE/CTE (hot/cold-tumor-enriched)

한번 약어를 정의하면 논문 전체에서 동일하게 사용한다.

---

## Figure와 Table의 효과적 배치

### 본문에서 구체적으로 언급한다

Figure나 Table을 언급할 때는 구체적인 정보와 함께 제시해야 한다.

잘못된 예:

Results are shown in Figure 2.

올바른 예 (Song et al. 2024, Nature Communications):

We found significant enrichment of two kinases, CSNK2A1 (FDR =  $2.3 \times 10^{-7}$ ) and GSK3B (FDR =  $1.9 \times 10^{-2}$ ), which are known to phosphorylate various proteins in the PI3K-AKT signaling pathway, in Subtype 4 compared to other subtypes (Fig. 2d, Supplementary Fig. 2d and Supplementary Data 2c).

구체적인 숫자와 통계값을 본문에 제시하고, Figure는 시각적 확인을 위해 인용한다.

### 순서대로 언급한다 (당연하게도..)

Figure와 Table은 본문에 등장하는 순서대로 번호를 매겨야 한다. Figure 1A, 1B, 1C 순서로 언급하다가 갑자기 Figure 3을 언급하면 안 된다. 논문을 쓰다보면, 자주 순서가 바뀌는데, 부지런히 바꾸어야 한다. 동숲에서 꽃밭에 물을 주듯.. 매일 매일 모닝커피와 함께 부지런히 확인한다.

#### Song et al., 2024의 Figure 배치:

Results 1: Fig. 1a, 1b, 1c, 1d, 1e (Subtype identification)  
Results 2: Fig. 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2j (Subtype 4 )  
Results 3: Fig. 3a, 3b-m (Cellular landscape)  
Results 4: Fig. 4a-i (WGD features)  
Results 5: Fig. 5a-g (Immune landscape)  
Results 6: Fig. 6a-g (Neoantigens)

논리적인 순서로 배치되어 있다.

### Figure 내용이 본문과 일치해야 한다

내가 학생에게 남긴 코멘트:

“히트맵하고 이 부분의 서술이 매칭이 안됩니다. 여기서 GO 결과 같은게 들어간 것 같은데 (제가 이해한게 맞다면), 그런 부분이 히트맵에 표시가 되어야 할 것 같아요.”

본문에서 설명한 내용이 Figure에서 시각적으로 확인 가능해야 한다.

### Supplementary vs Main Figure

저널의 가이드라인을 확인해야 한다.

“투고하려는 논문 사이트에서 author guideline 확인해주세요. Supplementary Figure 라고 쓰는지 Figure S 라고 쓰는지 확인을 해주세요.”

일반적으로: – **Main Figure**: 핵심 발견 – **Supplementary Figure**: 보조 데이터, 검증 실험, 추가 분석

Song et al., 2024의 경우: – Main Figures (1–6): 서브타입 정의, 예후 분석, 세포 특성, WGD, 면역 환경 – Supplementary Figures: TMB 분포, 추가 kinase 분석, 검증 코호트 결과

---

## 저널별 요구사항 확인

Journal은 각자의 Author guideline이 있다. 투고 전에 반드시 저널의 가이드라인을 읽어야 한다.

확인할 사항: – Results 섹션의 길이 제한 – Figure와 Table 개수 제한 – Supplementary Material 형식 – 인용 스타일 – 약어 사용 규칙

## 5장. Results 섹션을 작성하기

하나의 Result 섹션 안에서, 심지어 한 문장 안에서 어떻게 효과적으로 쓸 것인가에 대해 알아보자. 좋은 구조를 가진 Results도 문장이 명확하지 않으면 독자에게 전달되지 않는다. 이 장에서는 실제 우리 학생들이 작성한 논문에 남겼던 코멘트를 바탕으로, 문장 레벨의 작성 기술을 다룬다.

---

### Results 섹션의 문단 구조 — 유기적 연결

Results의 하나의 섹션을 작성할 때, 보통 3–4개의 문단을 적는다. 그 문단들은 서로 유기적으로 연결된다.

우리가 작성한 뇌발달 싱글셀 논문 Kim et al. 2024, Exp Mol Med의 “Cellular landscape of neurodevelopmental disorder genes in early neuronal lineages” 섹션을 예로 살펴보자.

#### 문단 1: 세포 특이적 enrichment 패턴

Over the past decade, large-scale genomic studies have identified risk genes associated with neurological disorders... We prioritized 14 sets of genes previously reported to be associated with neurological disorders... Consequently, distinct patterns of gene enrichment were detected in a cell type-specific manner.

이 문단은 큰 그림을 제시한다. 어떤 disorder risk genes이 어떤 cell type에서 enriched 되어 있는지 보여준다.

## 문단 2: Neuronal lineage의 궤적 분석

We further investigated the cellular lineages underlying the neuronal clusters and the lineage-specific patterns of neurodevelopmental disorder risk genes. The pseudotime trajectory built on cellular identity reflected the clustering patterns of cells differentiating from radial glia to neuroblasts, eventually committing to mature neurons...

문단 1에서 “cell type-specific”이라고 했으니, 이제 더 깊게 들어간다. 단순히 어떤 cell type이 아니라, cell이 분화하는 궤적(trajectory) 상에서 어떻게 발현되는지를 본다.

## 문단 3-5: 각 Lineage의 상세 분석

Lineage 1 (excitatory neurons):

Lineage 1 represents a differentiation trajectory from radial glia to excitatory neurons... Among the autism risk genes, those involved in transcriptional regulation, such as RNASE1 and TCF7L2, are predominantly expressed in the early stages of this lineage...

Lineages 2 & 3 (inhibitory neurons):

Lineages 2 and 3 represented the differentiation trajectory from radial glia to inhibitory neurons, diverging into MGE-derived (C20) and CGE-derived (C23) neurons...

각 lineage마다 어떤 risk genes이 어느 시점에 발현되는지, 어떤 biological pathway가 enriched 되는지를 구체적으로 서술한다.

논리적 흐름: 1. 큰 그림: 어떤 cell type에서 risk genes이 enriched 되는가? 2. 중간 그림: 이 cell type들이 어떻게 분화하는가? (lineage) 3. 세부 그림: 각 lineage에서 언제, 어떤 genes이, 어떤 pathway와 함께 발현되는가?

각 문단의 첫 문장이 어떻게 연결되는지 주목하자: - 문단 1: “We prioritized 14 sets of genes...” - 문단 2: “We further investigated the cellular lineages...” - 문단 3: “Lineage 1 represents...” - 문단 4: “Like lineage 1, lineages 2 and 3 presented...”

“Further”, “Like” 같은 연결어가 없어도, 논리적 흐름이 자연스럽게 이어진다. 각 문단이 이전 문단에서 던진 질문에 답하거나, 더 깊은 분석을 제공하는 구조다.

---

## AI를 활용하여 Results 섹션 작성하기

불과 수년전만 하더라도, 논문을 쓴다는 것은 매우 어려운 일이었다. 영어로 글쓰기를 해야하기도 했고, 과학 글쓰기를 해야했으니까... 요즘은 챗지피티, 클로드와 같은 AI 툴이 나와서, 영작의 난이도를 많이 낮추었다. 그래서 처음 논문을 쓰는 학생들에게 AI 도구를 활용하는게 도움이 된다.

### 단계별 작성 프로세스

#### 1단계: 문단 구조 설계 (3-4개 문단)

먼저 한 섹션 안에 몇 개의 문단을 쓸지 구상한다. 보통 3-4개가 적당하다 (이와 같은 구상은 처음부터 완벽할 필요가 없다).

예시: - 문단 1: Risk genes의 cell type-specific enrichment - 문단 2: Pseudotime trajectory 전체 패턴 - 문단 3: Lineage 1 상세 분석 - 문단 4: Lineages 2 & 3 비교 분석

#### 2단계: 각 문단의 핵심 결과 나열 (4-5개)

각 문단에 들어갈 주요 결과들을 논리적으로 4-5개 적는다. 문장으로 여러분이 직접 써도 되지만, 꼭 완전한 문장일 필요는 없다. 또한 연결을 어떻게 해야할지 너무 고민할 필요가 없다. 처음 논문을 쓰는 학생들은 문장간의 연결에

과도하게 많은 시간을 보낸다. 문장간의 연결은 AI 툴이 해준다. 혹은 문장들을 던져주고 나서, “논리적으로 연결이 안되는게 있었어?”라고 AI에게 물어보면서, 점검해도 된다.

예시 (문단 3):

- Lineage 1 radial glia excitatory neuron
- NEUROD6 neuroblast
- Autism risk genes (RNASE1, TCF7L2) early stage
- Developmental delay genes peak (RAB14, MYCN later, FOXP2 postnatal)
- Wnt signaling pathway enriched (FDR =  $4.79 \times 10^{-2}$ )

이러면 대략 12개에서 20개의 핵심 포인트가 나온다.

### 3단계: AI로 문장 생성 및 검토

AI 툴에게 이 포인트들을 바탕으로 문단을 작성하도록 한다.

Prompt :

```
"           Results           :  
- Lineage 1 radial glia excitatory neuron  
- NEUROD6 neuroblast  
- Autism risk genes (RNASE1, TCF7L2) early stage  
..."
```

논리적으로 맞는지, 그리고 의도한대로 결과가 제시되었는지 여러 번 읽어보고 수정한다. 이 과정에 사람(당신)이 개입해야 한다. 어쩌면 이 작업이 가장 많은 시간이 걸릴지도 모른다. 처음 논문을 쓰는 학생에게 이 부분이 어렵다. 어떻게 쓰여진 것이 좋은 글인가? 마치 학생은 처음 농작물을 얻어본 농부와 같다. 그래서 본인이 수확한 것이 좋은 작물인지 아닌지 판별하는 능력을 키워야 한다. 이것에는 왕도가 없다. 내가 스스로 경험하고, 혹은 학생들의 결과물을 통해 경험한 것들을 적어본다.

먼저, AI는 결과를 설명하면서 해석을 종종 넣는다. 아래와 같은 경우이다.

#### 잘못된 예:

FOXP2 was highly expressed at the end of the lineage, particularly in the postnatal stage, **\*\*suggesting its critical role in late neuronal maturation and potentially contributing to autism pathogenesis through disruption of synaptic function\*\***.

#### 올바른 예:

FOXP2 was highly expressed at the end of the lineage, particularly in the postnatal stage.

해석(“suggesting its critical role...” )은 Discussion에 넣어야 한다. Results에서는 관찰한 사실만 적는다. 물론 논문에서 문장 뒤에 “,suggesting 무엇” 이렇게 풀어서 설명하는 경우가 종종 있다. 그러나 AI 툴은 과도하게 이런 작업을 한다. 그러다보니 Results의 문단들이 속도감 있게 전개되지 않고, 부수적인 해석들로 가득찬다.

또 한가지 GPT의 패턴이 있다. Results를 작성할 때, 마지막에 요약 문단을 생성한다. GPT가 자주 생성하는 마지막 문단:

Collectively, these results demonstrate that essential biological processes are dynamically regulated throughout neurodevelopment, highlighting the temporal specificity of differentiation.

또는:

Collectively, these results demonstrate that a foundation model trained and fine-tuned using the atlas enables the prediction of meaningful biological representations of gene perturbations, positioning it as a powerful resource for dissecting the functional consequences of genetic perturbations associated with NDDs.

이런 종합적 해석은 Discussion에 속한다. Results는 마지막 결과를 제시하고 끝내면 된다. 이와 같은 문장 형태가 종종 논문에 쓰인다. 개인적인 라이팅 방식이긴 하지만, 해석과 총론은 Discussion에 넣는 것을 권유한다. Results에서는 데이터가 말하게 하고, 데이터로 의미를 전달하게 문장이 쓰여야 한다.

또 한가지 주의 할 점은, AI의 라이팅은 자주 “과장”하거나 “모호”하게 문장을 작성한다. 그럴듯하게 보이는 작업물을 전달한다. 그래서 프롬프트를 입력할 때 가급적 정량적인 지표, 정확한 구성요소들을 넣어야 한다 (특히 Results 파트에서). 또한 나온 결과물을 읽어보고, 모호한 언어로 작성했는지를 점검해야 한다. 혹은 나온 결과물을 다른 AI에게 넣어서, “이거 재네가 한건데, 혹시 모호하게 쓰여진 문장이나, 그럴듯하게 보이는 문장이 있어? 구체적인 정보를 얻을수 있어? 혹시 그런 부분이 있으면 지적해봐” 라고 물어볼수도 있다.

---

## 문장 작성할 때 우리가 자주하던 실수들

### 중요한 결과를 여러 문장에 분산시키지 말자

한 문장에서 명확하게 전달하는 것이 훨씬 효과적이다.

내가 Kim et al., 2024 논문 작성 시 학생에게 남긴 코멘트다 (Kim et al. 2024, Genome Medicine):

“여기도 매번 반복되는 습관인데, 중요한 결과를 너무 여러 문장에 나눠서 씁니다. 너무 헛갈려요...”

“바로 직전 문장에서 attention issue와 genetic association 이야기를 하는데 왜 갑자기 여기서 ADHD-like phenotypes이 나오는거죠... 문장이 이해가 안됩니다..”

잘못된 예:

Previous studies showed attention deficits. Genetic associations were found.  
ADHD-like phenotypes were observed in siblings.

올바른 예:

Siblings of autistic individuals showed ADHD-like phenotypes, including attention deficits that were genetically associated with polygenic burden ( $r = 0.23$ ,  $p = 0.003$ ).

하나의 핵심 발견은 하나의 명확한 문장으로 표현하자. 단, 한 문장에 관련 없는 여러 내용을 우겨넣지는 말자.

### 과도하게 세세하게 적는 학생들 (MBTI가 S 혹은 ST 인 경우들..)

MBTI가 S인 학생들이 자주 하는 실수가 있다. 관찰한 모든 디테일을 다 적으려고 하는 것이다.

한 학생에게 남긴 코멘트다:

“처음 논문을 쓰다보면, 학생처럼 MBTI가 S인 친구들이 자주 하는 실수가, 분석하며 관찰한 모든 것을 다 적는 것인데요. 범죄 스릴러 드라마라고 생각해볼게요. 여기서 범행 → 사건 수사 → 전투씬 → 사건 해결 만 있으면 되겠죠? 사건 수사하면서 방문한 음식점의 역사를 소개한다든지, 메뉴를 소개한다든지 하는 이야기는 안 해도 됩니다. 이건 맛집 탐방 유튜브가 아니니까요. 논문도 마찬가지예요. 이 섹션에서 전하고 싶은 1가지 메시지가 무엇인가요? 그것만 생각하며 쓰면 됩니다.”

각 섹션, 각 문단의 핵심 메시지 하나에 집중하자. 그 메시지를 전달하는 데 필요한 최소한의 정보만 제공하면 된다.

Kim et al., 2024 (Exp Mol Med)의 집중된 서술:

“Cellular landscape of neurodevelopmental disorder genes in early neuronal lineages” 섹션: - 문단 1: Risk genes는 특정 cell type에서 enriched 된다 - 문단 2: Neuronal lineages는 발달 단계에 따라 diverge 한다 - 문단 3: Lineage 1 (excitatory)의 특징과 temporal pattern - 문단 4: Lineages 2 & 3 (inhibitory)의 특징과 비교

불필요한 관찰은 모두 생략했다.

## 문장을 작성하다가 안드로메다로 여행한 학생들 (MBTI가 NP인 경우들...)

일부 학생들은 문장을 작성하고 전개할 때, 우주 밖으로 나가버린다.

A > B > C > D 순서로 전개를 해야하는데, A > B > B' > B'' > Z (응??? ) > 가 (응?? 응??)

이런 친구들은 호기심이 많고 생각이 많다. 그래서 사실을 순차적으로 전달해야 하는 Results의 글쓰기에서 새로운 활로를 열어버린다. Z까지 가고 나면, 친구들을 다시 잡아서 돌아오기가 어렵다. 혹시 이글을 읽는 본인이 이런 경우라면 해결책이 있다.

1. 논문을 작성할 때, 작성하고자 하는 아이템 (피겨, 테이블, 데이터)를 마련하자.
2. 인터넷을 끈다. 인터넷을 끈다. 인터넷을 끈다 -> 호기심에 의한 검색을 원천 차단한다.
3. 논문을 작성한다. 완벽하지 않고, 마음에 들지 않을 것이다. 그러나 과거의 여러분보다 훨씬 빠른 속도로 글을 쓸 것이다. 확인한다.

## 문장의 길이: 너무 짧지도, 너무 길지도 않게

핵심은 하나의 주요 결과는 하나의 명확한 문장으로 표현하는 것이다.

In children with autism, de novo PTVs were observed more frequently in females than males in Korean cohort (RR = 2.0; P = 0.11), SSC (RR = 1.8; P =  $1.1 \times 10^{-2}$ ), and SPARK (RR = 1.2; P = 0.32) (Fig. 2C).

이 문장은: - 하나의 핵심 발견: 여성에서 de novo PTV가 더 많다 - 세 cohort의 증거: 한 문장에 효율적으로 제시 - 정량적 데이터: 모든 숫자 포함

## 분석명이나 도구명을 주어로 쓰지 말자

이것은 학생들이 자주 하는 실수다. 학부생들에게 글쓰기를 가르칠 때도 자주 보이는 실수이다.

- “Heatmap analysis revealed...”
- “The volcano plot showed...” -> Figure legend에서는 이렇게 쓸수 있으나, Results에서 써야 할 문장은 아니다.
- “GO analysis demonstrated...”
- “Our differential expression analysis found...”
- “The Seurat analysis showed...” (Seurat은 싱글셀 분석 툴)

대신 이렇게: - “Gene expression differed significantly between...” - “Among 500 genes tested, 45 showed...” - “Enriched pathways included synaptic transmission and...”

Seurat, DESeq2, GSEA 같은 툴 이름은 Methods에 적으면 된다. 만약 여러분이 Seurat 툴의 개발자라면, 써도 된다. 근데 그런 게 아니니까... Results에서는 발견한 사실만 간결하게 전달하자.

## 처음 논문을 쓰는 학생들은 너무 자주 놀라고, 너무 자주 주목한다...

학생들의 논문을 보면 GPT가 생성한 듯한 표현들이 많다. 특히 “Altogether”, “Together”, “Notably”, “Surprisingly” 같은 부사어를 과도하게 사용한다. 실제로 한 학생의 논문에서 내가 삭제 표시한 것들:

삭제한 표현들: - “Altogether, the Core and Extended Atlas constitute...” - “Together, these findings demonstrate...” - “Notably, both OligoEUR and OligoEAS exhibited...” - “Surprisingly, these models achieved...” - “In particular, telencephalic neurons exhibited...” - “Furthermore, the NOCAP contains...”

이런 표현들은 대부분 불필요하다. 없어도 문장의 의미는 그대로 전달된다.

## 수정 전:

Altogether, the Core and Extended Atlas constitute a comprehensive NOCAP annotated across hierarchical cell-type levels, designed to provide a robust and integrated atlas, providing the foundation for all subsequent analyses.

## 수정 후:

The Core and Extended Atlas constitute a comprehensive NOCAP annotated across hierarchical cell-type levels (Figure 1E).

훨씬 간결하고 명확하다. “Notably”나 “Surprisingly” 없이도 충분히 강력한 문장이다. 결과 자체가 “놀라움”을 전달한다. 한국 학생들은 부사를 많이 쓴다. 우리나라의 말투에도 그런 습관이 많다.

한 학생 논문에 내가 남긴 코멘트:

“[중요] 여기서부터 연결어와 부사, 불필요한 수사들을 삭제했습니다. 제가 삭제한 부분을 읽어보고, 의미가 제대로 전달되는지 확인해주세요. 왜 이 정도는 생략해도 되냐면, '결과(숫자)'를 보여주기 때문입니다. 만약 읽었을 때 속도감 있게 잘 읽혔는데, 뭔가 부족하다고 느껴지면, '결과(숫자)'를 더 보여주세요.”

불필요한 수사를 제거하고 나면, 정작 **정량적 결과(숫자)**가 부족하다는 것을 깨닫게 된다. 그때 숫자를 채워 넣으면 된다.

## 과정 중심 동사 줄이기

“established”, “optimized”, “benchmarked” 같은 과정 중심 동사도 Results에서는 최소화하자.

한 학생 논문의 예:

To optimize the Core Atlas integration, we benchmarked batch correction methods using established evaluation metrics, effectively minimizing batch effects while preserving biological variation.

이것은 Method에 들어갈 내용이다. Results에서는:

The Core Atlas resulted in approximately 1.4 million high-quality cells (Figure S1B).

결과만 간단히 제시하면 된다.

## 순환 구조 문장 피하기 (feat. 편쿨색좌)

학생들이 문장을 적을 때 자주 하는 실수가 있다. 잘못된 예 (순환 구조):

To find differentially expressed genes, we performed differential expression analysis. As a result, we observed differentially expressed genes between groups.

이 문장의 문제점: 1. “To find A”라고 시작 2. “we performed B” (B는 A를 찾기 위한 분석) 3. “we observed A” (결국 A를 찾았다)

이건 내용이 없는 문장이다. 당연히 differential expression analysis를 하면 differentially expressed genes를 찾는 거다. “To find A, we did B. We found A” 구조는 피하자. 바로 “We found A”로 시작하거나, “A differed/showed/exhibited” 같은 직접적 표현을 쓰자.

## 줄임말에 대하여

Academic writing에서는 축약형(contractions)을 쓰지 않는다.

내가 학생에게 남긴 코멘트:

“Academic writing에선 줄임말을 쓰지 않습니다...”

잘못된 예: - can't → cannot - don't → do not - it's → it is - isn't → is not - doesn't → does not

단, 표준 약어(acronyms)는 사용 가능하다: - DNA, RNA, PCA, GSEA, GO (이런 것들은 OK)

내가 말하는 “줄임말”은 isn't나 doesn't 같은 축약형(contractions)이다. 실제 전문 용어들은 당연히 줄여서 쓴다.

## 단어 선택의 정확성

Results와 Discussion에서 사용하는 동사가 다르다.

Results에 적합한 동사: - showed, demonstrated, revealed - indicated, suggested (약한 주장) - differed, increased, decreased - assessed, evaluated, examined, tested

Discussion에 적합한 동사: - speculate, hypothesize (추측) - implies, suggests (해석) - may, might, could (가능성)

내가 학생에게 남긴 코멘트:

“잘 쓰지 않는 말... 추측하는 것보다 assess, evaluate, investigate 등 이런 평가한 단어를 쓰는 게 좋습니다. 추측은 너무 강한 말이에요.”

## 수동태 vs 능동태

상황에 따라 적절히 선택해야 한다.

Results에서 능동태 (우리가 한 분석이나 찾은 것):

We examined 500 samples...

We performed differential expression analysis...

We identified three distinct lineages...

Results에서 수동태 (사전 연구에서 등장한 결과들):

Samples were clustered into three groups...

Genes were annotated using...

나는 Results에서 문장을 쓸 때, “We performed”, “We found”, “We observed”와 같은 능동태 문장을 쓴다. 꼭 수동태일 필요는 없다. 우리가 찾은 것, 우리가 실험한 것을 말하는 것이니까. 다만 다른 연구자들이 찾은 결과들을 언급할 때는 “Sanders et al. found” 이런 식의 능동태 문장은 쓰지 않는다. 다른 연구나 결과를 언급할 땐 수동태로 쓴다:

It was previously reported that... [1]

De novo variants have been shown to... [2,3]

이건 내가 주로 하는 라이팅 스타일이고, 이런 문장 방식은 사람마다 다를 수 있다.

능동태:

We conducted a quality control assessment for the WGS datasets...

We identified distinct neuronal lineages...

수동태:

Risk genes for neurodevelopmental disorders, including autism, developmental delay, and epilepsy, are predominantly expressed in neuronal cell types.

다른 연구 언급 (수동태):

De novo variants in autism have been shown to disrupt neuronal genes... [5]

## 인용(Citation)의 적절한 사용

### Previous studies 언급 시 반드시 인용

다른 연구의 결과를 언급할 때는 항상 인용이 필요하다. 내가 Kim et al., 2024 작성 시 자주 남기는 코멘트다:

“Previous studies 나 finding이 들어가면 citation이 있어야 해요.”

“Reference??”

잘못된 예:

As previously documented, DNV genes were associated with transcriptional regulation and synaptic transmission.

올바른 예:

As previously documented, DNV genes were associated with transcriptional regulation and synaptic transmission [2, 16, 21].

The autism and developmental delay risk genes were considerably enriched in both excitatory and inhibitory neuronal cell types, which is consistent with previous findings [24-26].

De novo variants in autism disrupt neuronal genes that are overexpressed in the prefrontal cortex during the mid-fetal period [5], indicating that risk perturbation at the molecular level precedes its clinical onset [6].

### 본인의 발견도 명확히 표시

본인의 새로운 발견인지, 기존 연구의 재확인인지 명확히 해야 한다.

### 새로운 발견:

We identified three distinct lineages diverged from radial glia and differentiated into excitatory and inhibitory neurons.

### 기존 연구의 확인:

Consistent with previous reports [23], we observed enrichment in synaptic genes.

### Citation 형식 통일

저널마다 인용 스타일이 다르다. 한 논문 내에서는 일관되게 유지해야 한다. 내가 학생에게 남긴 코멘트:

“투고하려는 논문 사이트에서 author guideline 확인해주세요. Supplementary Figure라고 쓰는지 Figure S라고 쓰는지 확인을 해주세요”

인용 스타일: – 위첨자: text<sup>2,3,4</sup> – 괄호: text [2, 3, 4] – Author-year: text (Smith et al., 2020)

Kim et al., 2024는 괄호 스타일 [숫자]를 일관되게 사용한다.

---

## Results 문장을 작성할 때, 확인해야 하는 것

### 통계값은 필수

“significantly enriched”라고만 쓰지 말자. 항상 구체적인 수치를 제공해야 한다.

잘못된 예:

Gene expression was significantly higher in cases than controls.

올바른 예:

Gene expression was significantly higher in cases than controls (fold change = 2.3,  $p = 0.001$ , t-test, two-sided,  $n = 500$ ).

Kim et al., 2024 (Genome Medicine)의 좋은 예:

In line with previous findings [58], we observed an average twofold higher proportion of individuals with a de novo PTV among autism cases with ID compared to those without ID in both females and males (Fig. 3A; Additional file 6: Table S5).

“twofold higher”라는 구체적 수치를 제시한다.

또 다른 예:

Male siblings had significantly higher scores for seven out of nine clinical phenotypes related to autism core symptoms (total symptom severity, social communication and restricted interest and repetitive behaviors) and significantly lower scores for five out of seven cognitive/adaptive behaviors than female siblings (Fig. 5C; Additional file 8: Table S7).

“seven out of nine”, “five out of seven”로 정량화했다.

“그리고 특정 pathway가 어느 시점에 x fold 만큼 올라갔고 내려갔는지 등등에 대한 정량적인 값이 있어야 할 것 같아요. 예를들어서, (Figure 1B). In particular, [pathway name] were significantly increased in [group] (XX fold, p=XXX, t-test, two-sided). 이런식으로 넣어야 합니다.”

## 유효숫자 통일

저널마다 유효숫자 규정이 다르다. 한 논문 내에서는 일관되게 유지해야 한다.

일반적으로: - p-value: 소숫점 2-3자리 ( $p = 0.001$  또는  $p = 1.1 \times 10^{-2}$ ) - effect size나 fold change: 소숫점 1-2자리 (RR = 2.0, fold change = 2.3) - 백분율: 정수 또는 소숫점 1자리 (65%, 5.6%)

## 샘플 수 명시

통계 분석 결과를 제시할 때는 항상 샘플 수를 명시해야 한다.

We analyzed samples from the top 5 brain regions by sample size (cortex: n=342, hippocampus: n=198, cerebellum: n=156, striatum: n=89, midbrain: n=67).

내가 학생 논문에 남긴 코멘트:

“샘플 수가 많은 순서대로 상위 5개 region과 샘플수 n=XX로 표시”

Kim et al., 2024 (Genome Medicine)의 예:

The Korean autism WGS data consists of 673 families of total 2,255 individuals, encompassing 21 multiplex families with more than two autistic children.

Of the 696 individuals with autism, 590 were males and 106 were females (5.6 male-to-female ratio). Siblings included 90 male and 123 female individuals (0.74 male-to-female ratio).

모든 그룹의 샘플 수가 명확하다.

## “higher”, “lower”, “enriched” - 정량적 근거 필요

비교 표현을 사용할 때는 반드시 정량적 근거를 제시해야 한다.

잘못된 예:

Oligogenic genes showed enrichment in synaptic transmission, comparable to DNV genes.

올바른 예:

Oligogenic genes showed enrichment in synaptic transmission (GO:0007268, 23 genes, adjusted p = 0.002), comparable to DNV genes (GO:0007268, 31 genes, adjusted p = 0.001).

## 사용한 통계 방법을 명시할 것

어떤 통계 검정을 사용했는지 항상 명시해야 한다. 저널마다 가이드라인이 있으니, 가이드라인을 확인하여 입력하길 바란다.

Groups differed significantly ( $p = 0.003$ , T-Test, Two-sided).

Proportions were compared using Fisher's exact test ( $p = 0.012$ ).  
We performed permutation testing (1,000 iterations) to assess significance.

### Non-significant 결과도 언급한다

통계적으로 유의하지 않은 결과도 값을 표기한다

Although we tested for differences in PRS between groups, no significant association was observed ( $p = 0.24$ ,  $n = 500$ ).

Sex ratios were similar across subtypes ( $p = 0.31$ , chi-square test).

### 비교 대상을 명확히

무엇과 무엇을 비교했는지 명확해야 한다.

애매한 표현:

De novo variants showed higher burden. (<- higher?)

명확한 표현:

De novo variants showed higher burden in cases (mean = 2.3) compared to controls (mean = 0.8,  $p < 0.001$ , t-test).

In children with autism, de novo PTVs were observed more frequently in females than males in Korean cohort (RR = 2.0; 0.066 variants per female and 0.033 per male;  $P = 0.11$ ).

누구와 누구를 비교했는지, 각각의 값이 무엇인지 모두 명확하다.

### “Significant”라는 용어를 사용할 때 - 통계적 의미

“Significant”는 과학 논문에서 특별한 의미를 가진다.

내가 Kim et al., 2024 작성 시 학생에게 남긴 코멘트다:

“한 문장에 너무 많은 내용이 있지 않나요? 두 개는 다른 표현형이고, 각기 따로 설명을 할애할 수 있을 것 같아요.”

“그리고 ‘significant’를 말하면 통계값이 제시되어야 하는데요. 이런 것 포함. 투고하고자 하는 저널에 나온 다른 논문들 보고 포맷을 파악해보세요. 어떤 어떤 내용들이 들어가야 할지..”

통계적 유의성을 말할 때:

Expression was significantly higher in cases ( $p < 0.05$ , t-test).

일반적 중요성을 말할 때:

This finding is particularly important because...

또는:

This represents a notable difference from previous studies.

“Significant”를 통계적 의미로만 사용하고, 일반적 중요성은 “important”, “notable”, “substantial” 등으로 표현하자.

통계적으로 유의할 때:

Male siblings had significantly higher scores for seven out of nine clinical phenotypes related to autism core symptoms ( $p < 0.05$ ).

일반적 중요성:

This finding is important for understanding sex-differential liability in autism.

## 6장. Methods — 첫 논문 글쓰기의 튜토리얼 단계

Methods 섹션은 논문에서 가장 '기계적'으로 쓸 수 있는 부분이다. Introduction이나 Discussion처럼 창의적인 서술이 필요한 것도 아니고, Results처럼 발견을 어떻게 표현할지 고민할 필요도 없다. 그저 **여러분이 무엇을 했는지를 정확하고 빠짐없이 기록**하면 된다.

그래서 나는 처음 논문을 쓰는 학생들에게 Methods부터 쓰라고 권한다. 다른 사람의 논문에서 비슷한 분석 방법을 찾아 문장 구조를 참고하고, 여러분의 연구에 맞게 숫자와 파라미터만 바꾸면 된다. 일종의 튜토리얼 단계라고 생각하면 된다.

하지만 '기계적'이라고 해서 아무렇게나 써도 되는 건 아니다. Methods가 부실하면 논문 전체의 신뢰도가 무너진다. 독자가 여러분의 결과를 재현할 수 없고, reviewer가 분석의 타당성을 검증할 수 없기 때문이다. 이 장에서는 Methods의 각 요소를 어떻게 구성하고 작성하는지 단계별로 정리한다.

---

### 연구 설계와 환자 모집

Methods의 첫 절은 보통 연구 설계와 대상자 모집으로 시작한다. 여기서 명확히 해야 할 것들이 있다.

**연구 목적, 코호트 구성, 윤리 승인**부터 적는다. IRB 승인 번호와 날짜, 동의서 취득 방식을 빠뜨리지 말자. 이건 단순한 형식이 아니라 연구의 윤리적 정당성을 보여주는 필수 요소다.

**포함·제외 기준**을 구체적으로 적어야 한다. “자폐 스펙트럼 장애 환자를 모집했다”고만 쓰면 안 된다. 어떤 진단 기준을 사용했는지, 연령 범위는 어떻게 되는지, 다른 공존 질환이 있는 경우는 포함했는지 제외했는지를 명시해야 한다.

**임상 변수 정의**도 중요하다. 분석에 직접 사용되는 변수(예: 나이, 성별, 진단, 병기, 인지 평가 점수)를 중심으로 정리한다. 이런 데이터는 임상팀으로부터 미리 표준화된 형태로 확보해두는 것이 좋다. 나중에 “이 환자 병기가 뭐였지?” 하고 찾아 헤매는 일이 없도록.

#### 예시 문장:

This study included 512 individuals diagnosed with autism spectrum disorder and 325 unaffected controls recruited under IRB approval (protocol #2020-0567). Written informed consent was obtained from all participants or legal guardians. Clinical characteristics such as age, sex, diagnostic category, and cognitive assessment scores were obtained from medical records (Table S1).

이 문장은: - 샘플 수 (512 cases, 325 controls) - IRB 번호 (#2020-0567) - 동의서 취득 방식 (written informed consent) - 수집한 임상 변수 (age, sex, diagnostic category, cognitive scores) - 데이터 출처 (medical records) - 상세 정보 위치 (Table S1)

모든 필수 정보를 담고 있다.

---

### 시료 수집 및 데이터 생산

대부분의 오믹스 연구에서는 혈액, 조직, 세포 등의 생물학적 시료로부터 DNA, RNA, 또는 단백질을 추출한다. 이 절에서는 실험 프로토콜의 모든 세부사항을 적을 필요는 없다. **데이터 종류, 생산 기관, 품질 지표**를 중심으로 간결하게 쓰면 된다.

**시료 채취 방법**을 간단히 적는다. 어떤 튜브에 담았는지(EDTA, heparin), 어떤 온도에서 보관했는지(-80°C, -20°C) 정도면 충분하다.

**데이터 생산 절차**도 핵심만 적는다. 시퀀싱 플랫폼(Illumina NovaSeq, PacBio 등), read 길이(2×150 bp, 2×100 bp), 평균 커버리지(30×, 50×)를 명시한다. 사용한 키트가 있다면 회사명과 제품명도 적는다.

**품질 지표**는 반드시 포함한다. Q30(quality score 30 이상인 read의 비율), read depth, mapping rate 같은 기본 지표를 보고한다.

#### 예시 문장:

Peripheral blood samples were collected in EDTA tubes and stored at  $-80^{\circ}\text{C}$ . Genomic DNA was extracted using QIAamp DNA Mini Kit (QIAGEN) and sequenced on Illumina NovaSeq 6000 (2×150 bp). Mean coverage was 30× with Q30 85%.

이렇게 몇 문장이면 충분하다. 실험 노트처럼 “DNA를 추출하고 농도를 측정하고 희석하고...”를 다 적을 필요는 없다.

---

## 데이터 품질관리(QC)

QC는 데이터 생성 직후 시행하는 가장 중요한 절차다. 오믹스 논문에서는 별도의 절로 다루는 것이 일반적이다. 왜냐하면 QC가 부실하면 그 뒤의 모든 분석이 무의미해지기 때문이다.

**주요 QC 항목을 데이터 종류별로 정리하면:**

**WGS/WES QC:** – Contamination (FREEMIX 값) – Call rate (성공적으로 genotype이 결정된 비율) – Depth (평균 커버리지) – Duplicates (중복 read 비율) – Sex concordance (유전적 성별과 보고된 성별 일치 여부)

**RNA-seq QC:** – RIN (RNA Integrity Number) – Mapping rate (reference에 align된 read 비율) – %rRNA (ribosomal RNA contamination) – Gene detection 수 (발현이 검출된 유전자 수)

**샘플 수준 QC:** – Ancestry (PCA로 확인한 유전적 ancestry) – Kinship (가족 관계 확인, 예상치 못한 친척 관계 제거) – Phenotype consistency (임상 데이터와 유전 데이터 일치 여부)

**QC 기준과 제외된 샘플 수를 구체적으로 기재해야 한다.** “품질이 낮은 샘플을 제거했다”고만 쓰면 안 된다. 어떤 기준으로, 몇 개를 제거했는지 명확히 해야 한다.

**예시 문장:**

Samples with mean coverage  $<20\times$  or contamination (FREEMIX)  $>0.05$  were excluded. RNA libraries with RIN  $<6.0$  or mapping rate  $<80\%$  were also removed. After QC, 174 WGS and 169 RNA-seq samples remained for downstream analysis.

이 문장은: – WGS QC 기준 (coverage  $<20\times$ , FREEMIX  $>0.05$ ) – RNA-seq QC 기준 (RIN  $<6.0$ , mapping rate  $<80\%$ ) – QC 후 남은 샘플 수 (174 WGS, 169 RNA-seq)

모든 결정이 투명하게 기록되어 있다.

---

## 데이터 처리 및 정렬

이 절은 정렬, 변이 호출, 정제의 전 과정을 기술하는 부분이다. 오믹스 분석을 하는 사람이라면 이미 익숙한 파이프라인일 것이다.

**Reference genome과 annotation 버전을 반드시 명시한다.** GRCh37인지 GRCh38인지, GENCODE v38인지 v105인지에 따라 결과가 달라질 수 있다.

**사용된 알고리즘 및 소프트웨어 버전도 정확히 적는다.** BWA-MEM v0.7.17, GATK v4.2.0처럼 버전까지 포함한다. 재현성을 위해 필수적이다.

**공통 문장은 표준 파이프라인을 참고해도 좋다.** GATK Best Practices나 nf-core 파이프라인 문서에서 Methods 문장을 가져와 여러분의 분석에 맞게 수정하면 된다. 이건 표절이 아니다. 표준 절차를 표준 방식으로 기술하는 것이다.

**변이·유전자 단위의 필터 기준도 포함한다.** MAF(minor allele frequency), call rate, Hardy-Weinberg equilibrium p-value 같은 기준을 명시한다.

**예시 문장:**

Reads were aligned to GRCh38 using BWA-MEM v0.7.17. Duplicates were marked and base quality was recalibrated using GATK v4.2.0. Joint genotyping was performed following GATK Best Practices. Variants with call rate  $<90\%$ , MAF  $<0.01$ , or Hardy-Weinberg  $p < 1 \times 10^{-6}$  were excluded. Variant annotation was conducted using

Ensembl VEP v105 with GENCODE v38.

이런 문장은 다른 논문에서 비슷한 구조를 찾아 참고하면 쉽게 쓸 수 있다. 중요한 건 여러분의 파라미터를 정확히 기록하는 것이다.

---

## 통계 분석 및 모델링

통계 분석 절에서는 **분석 단위**, **사용된 모델**, **공변량**을 명확히 해야 한다.

**분석 단위**가 무엇인지 적는다. Variant-level인지, gene-level인지, sample-level인지에 따라 분석 방법이 완전히 다르다.

**사용된 모델**을 명시한다. 선형 회귀(linear regression), 로지스틱 회귀(logistic regression), 혼합 모델(mixed-effects model) 등 어떤 모델을 사용했는지 적는다.

**공변량(covariates)**을 빠짐없이 나열한다. 나이, 성별, ancestry principal components, batch effects 등 모델에 포함된 모든 변수를 적어야 한다. 이게 빠지면 분석의 타당성을 검증할 수 없다.

**다중비교 보정 방식**도 명시한다. FDR(False Discovery Rate), Bonferroni correction, permutation-based correction 중 무엇을 사용했는지, 임계값은 얼마로 설정했는지( $FDR < 0.05$ ,  $p < 2.5 \times 10^{-6}$ ) 적는다.

**분석 소프트웨어와 버전**도 기재한다. R v4.3.0, PLINK v2.0, Python v3.9 등을 명시한다.

필요한 경우 **모델 검증 방법**도 적는다. Cross-validation, permutation testing, genomic control  $\lambda$  같은 것들이다.

### 예시 문장:

Association tests were performed using a linear mixed model adjusting for sex, age, and the first five ancestry principal components. Multiple testing was controlled using the Benjamini-Hochberg procedure ( $FDR < 0.05$ ). Statistical analyses were conducted in R v4.3.0 and PLINK v2.0.

이 문장은: - 모델 종류 (linear mixed model) - 공변량 (sex, age, 5 PCs) - 다중비교 보정 (Benjamini-Hochberg,  $FDR < 0.05$ ) - 사용 소프트웨어 (R v4.3.0, PLINK v2.0)

필요한 정보를 모두 담고 있다.

---

## Gene Set Enrichment에 대한 작성 요령

유전자 목록을 만들고 functional enrichment를 수행했다면, 이 과정도 Methods에 상세히 기술해야 한다.

**Gene list 구축 기준**을 명확히 한다. 통계 임계값( $FDR$ ,  $\log_2$  fold change 등)과 필터링 단계를 명시한다. “유의미한 유전자를 선택했다”고만 쓰면 안 된다.  $FDR \leq 0.05$ 인지,  $|\log_2FC| \geq 1$ 인지 구체적으로 적어야 한다.

**Gene ID 변환 규칙**도 적는다. Ensembl gene ID를 HGNC symbol로 어떻게 변환했는지, unmapped된 유전자는 몇 개인지, 어떤 버전의 annotation을 사용했는지 기재한다.

**Enrichment 분석의 세부사항**을 포함한다: - 사용한 데이터베이스 (GO, KEGG, Reactome) - 배경 유전자 집합 (전체 유전체? 발현된 유전자만?) - Gene set 크기 제한 (10-2000 genes) - 다중검정 보정 방식 ( $FDR$ , Bonferroni) - 사용한 패키지와 버전 (gProfiler, enrichR, clusterProfiler)

### 예시 문장:

Differentially expressed genes ( $FDR \leq 0.05$ ,  $|\log_2FC| \geq 1$ ) were mapped to Ensembl release 105 using biomaRt. Functional enrichment was tested using gProfiler (GO Biological Process, 10-2000 genes) with 10,000 permutations and Benjamini-Hochberg FDR correction.

이 문장은: - DEG 정의 ( $FDR \leq 0.05$ ,  $|\log_2 FC| \geq 1$ ) - ID 매핑 (Ensembl 105, biomaRt) - Enrichment tool (gProfiler) - DB 종류 (GO Biological Process) - Gene set 크기 (10-2000) - Permutation 횟수 (10,000) - 보정 방식 (Benjamini-Hochberg FDR)

모든 파라미터가 재현 가능하도록 기록되어 있다.

---

## 재현성 및 데이터·코드 가용성

현대 과학 논문에서는 재현성이 무엇보다 중요하다. 저널에 따라서는 이 절이 필수인 경우도 많다.

**분석 환경을 명시한다.** Nextflow, Snakemake 같은 워크플로우 관리 도구를 사용했다면 버전을 적는다. 컨테이너 이미지(Docker, Singularity)나 conda 환경을 사용했다면 그것도 기록한다.

**원시 데이터 접근 경로를 제공한다.** 데이터를 공개 저장소(dbGaP, EGA, GEO 등)에 기탁했다면 accession number를 명시한다. 개인정보 보호 때문에 공개할 수 없다면 “available upon reasonable request”라고 적을 수도 있지만, 가능하면 공개하는 것이 좋다.

**분석 코드 저장소도 링크한다.** GitHub, Zenodo DOI, Figshare 등에 코드를 올려두고 URL이나 DOI를 제공한다.

**부록 자료(Supplementary Tables)를 명확히 연결한다.** Table S1에 샘플 메타정보와 QC 지표가 있고, Table S2에 유전자별 통계 결과가 있다면 본문에서 언급해야 한다.

### 예시 문장:

All analyses were executed using Nextflow v22.10 with Singularity container (sha256:...). Source code and configuration files are archived at Zenodo (DOI:10.5281/zenodo.xxxxx). Raw and processed data are available through dbGaP accession phs003245.v1.p1. Sample metadata and QC metrics are summarized in Table S1.

이런 문장은: - 워크플로우 도구 (Nextflow v22.10) - 컨테이너 (Singularity, sha256) - 코드 저장소 (Zenodo DOI) - 데이터 저장소 (dbGaP accession) - 부록 위치 (Table S1)

투명성과 재현성을 모두 확보한다.

---

## Methods를 쓸 때의 마음가짐

처음 논문을 쓰는 학생들에게 하고 싶은 말이 있다. Methods는 여러분이 한 일을 정직하게 기록하는 공간이다. 결과를 예쁘게 포장할 필요도 없고, 거창하게 만들 필요도 없다. 그저 **다른 연구자가 똑같이 따라 할 수 있도록** 충분히 상세하게 쓰면 된다. 다른 논문의 Methods를 참고하는 것은 전혀 문제가 없다. 특히 표준 파이프라인(GATK Best Practices, DESeq2 workflow)을 사용했다면 해당 논문이나 공식 문서의 문장 구조를 가져와도 된다. 중요한 건 여러분의 파라미터와 버전 정보를 정확히 기록하는 것이다.

Methods를 쓰면서 “이 부분을 어떻게 했더라?”라는 생각이 든다면, 그건 실험 노트나 분석 로그를 다시 확인해야 한다는 신호다. 기억에 의존해서 쓰지 말고, 실제로 사용한 명령어, 파라미터, 버전을 확인하자. Methods가 탄탄하면 나머지 섹션을 쓰는 것도 훨씬 수월해진다. 여러분이 무엇을 했는지 명확히 정리되어 있으니, Results에서는 그 결과를 보고하면 되고, Discussion에서는 그 의미를 해석하면 된다. 첫 논문의 Methods를 잘 써두면, 다음 논문을 쓸 때도 그 템플릿을 활용할 수 있다. 같은 연구실에서 비슷한 분석을 하는 경우가 많으니, 한 번 제대로 만들어두면 계속 재사용할 수 있다.

자, 이제 여러분의 분석 과정을 정리해보자. 한 문장씩 차근차근 쓰다 보면 어느새 Methods 섹션이 완성되어 있을 것이다.

## 7장. 참고문헌 조사 및 서지 관리

논문은 어벤져스 시리즈와 같다. 각 영화가 독립적인 이야기를 다루지만 결국 하나의 거대한 서사로 연결되듯, 논문도 마찬가지다. 한 연구자가 던진 질문에 다른 연구자가 답하고, 그 답은 또 다른 질문을 낳는다. 누군가의 생각이 다른

누군가의 생각으로 전송되고 이어지며, 때로는 변증적으로 발전해온 지적 대화의 기록이다. 어떤 가설은 검증되고, 어떤 주장은 반박되며, 그 과정에서 지식은 조금씩 정교해진다. 어쩌면 당신이, 우리가 지금 던지려는 질문도 그럴 것이다. 그렇기에 논문을 쓰기 전에 가장 먼저 해야 할 일은 “이 대화가 어디에서 시작되었고, 지금 어디까지 왔는가”를 파악하는 것이다. 참고문헌 조사는 단순한 리스트 정리가 아니라, 이 거대한 지적 계보 속에서 내 연구가 어디에 위치하는지 찾아내는 작업이다. 그리고 그것이 바로 서론의 첫 문단과 두 번째 문단을 쓰기 위한 준비가 된다.

이 장에서는 참고문헌 조사와 서지 관리를 하나의 연구 과정으로 정리한다. 프로젝트별로 별도의 라이브러리 파일을 만들어 관리하는 방식을 사용하며, 초보자도 쉽게 따라 할 수 있도록 실제 화면 기준으로 구성한다. 여기서는 첫 논문 작성을 위한 방법을 제시하기에, 논문 읽기 자체가 어려운 학생들은 내가 미리 작성해 둔 “논문 읽기 방법”을 참고하길 바란다.

---

## 문헌 조사의 목적과 흐름: 연구의 계보 읽기

문헌 조사의 첫 단계는 ‘왜 조사하는가’를 분명히 하는 것이다. 서론의 첫 문단은 연구 분야의 큰 흐름을 보여주고, 두 번째 문단은 그 안의 공백을 설명해야 한다. 따라서 문헌 조사는 이 두 문단을 쓰기 위한 논리적 배경 조사라고 볼 수 있다. 리뷰 논문으로 큰 틀을 잡고, 원저 논문으로 세부 근거를 채운다. 논문을 단순히 요약하는 것이 아니라, ‘무엇이 알려졌고 무엇이 아직 남아 있는가’를 구분해 놓아야 한다. 이 작업이 잘 되면 인트로의 방향은 자연스럽게 결정된다.

논문은 저자들의 생각이 이어져온 타래다. 논문에는 저자들이 있고, 그 저자들의 생각이 이어져온 흐름이 있다. 각 논문은 독립적이지만 서로 연결되어 있다. 몇 가지 중요한 흐름과 생각들을 찾아서 모으는 것이 문헌 조사의 핵심이다.

### 사례: 두 개의 리뷰 논문에서 공통으로 인용되는 핵심 논문 찾기

최근 자폐 스펙트럼 장애(ASD) 유전학 분야의 두 리뷰 논문을 예로 들어보자:

1. “Genomics, convergent neuroscience and progress in understanding autism spectrum disorder” (Willsey et al., Nature Rev Neurosci, 2022)
2. “Advancing precision diagnosis in autism: Insights from large-scale genomic studies” (Kim et al., Mol Cells. , 2025)

이 두 리뷰는 3년 차이로 발표되었지만, 둘 다 Satterstrom et al., 2020 논문을 핵심 연구로 반복해서 인용한다.

### 첫 번째 리뷰에서 Satterstrom 2020이 언급되는 방식:

“Contemporary association analyses...have emerged, providing the field with robust and consistent results. For WES studies specifically, it has become commonplace to require a discovery P value of less than or equal to  $2.5 \times 10^{-6}$ ... the recent omnibus analysis by the Autism Sequencing Consortium is commonly cited as identifying 102 ASD risk genes, based on the Bayesian FDR threshold of 0.1”

→ 이 문장에서 Satterstrom 2020은 현재 ASD 유전학의 “표준 방법론”을 정립한 연구로 인정받고 있음을 알 수 있다. 통계적 기준( $FDR \leq 0.1$ )과 102개 유전자라는 구체적인 숫자가 분야의 공통 언어가 되었다.

“The initial progress in ASD gene discovery relied heavily on the identification of rare heterozygous de novo mutations... However, missense de novo mutations as well as rare inherited and somatic mutations have now also been demonstrated to carry risk”

→ Satterstrom 2020 이전에는 주로 de novo 변이만 주목받았으나, 이 연구가 missense 변이와 유전 변이까지 체계적으로 포함시킴으로써 연구 범위를 확장했다는 맥락을 보여준다.

“scRNA-seq has provided some additional granularity...ASD risk genes tend to be highly and/or specifically expressed in developing excitatory and inhibitory neurons”

→ Satterstrom 2020의 유전자 목록이 단세포 전사체 분석의 기준 데이터로 활용되고 있음을 보여준다.

### 두 번째 리뷰에서 Satterstrom 2020이 언급되는 방식:

“Subsequent efforts focused on integrating previously published WES datasets with SSC WES data, culminating in the identification of 65 ASD risk genes. These efforts continued over the years, resulting in a large-scale analysis of 35,584 WES samples in 2020. Expanding TADA to include rare inherited variants, this study identified 102 ASD-associated genes ( $FDR \leq 0.1$ )”

→ 여기서는 Satterstrom 2020을 Sanders 2015 (65개 유전자)에서 진화한 연구로 위치시킨다. 샘플 수 증가(35,584명)와 방법론 확장(TADA에 희귀 유전 변이 포함)이 핵심 기여로 강조된다.

“Early WES studies identified the critical role of de novo PTVs in ASD...providing the first large-scale genetic insights into the disorder. To further refine de novo association analyses and improve gene discovery, the Autism Sequencing Consortium was established”

→ Satterstrom 2020은 Autism Sequencing Consortium의 집단적 노력의 정점으로 묘사된다. 개별 연구가 아니라 컨소시엄 기반 협력 연구의 성과물이라는 점을 강조한다.

“Early genomic investigations of ASD identified recurrent de novo CNVs in synaptic genes...Subsequent large-scale exome sequencing revealed a significant excess of de novo PTVs in genes governing neuronal communication (NC) and gene expression regulation (GER)”

→ Satterstrom 2020이 발견한 유전자들을 기능별로 분류(NC vs. GER)하면서, 이 분류가 이후 연구의 해석 틀로 자리 잡았음을 보여준다.

### 두 리뷰를 비교하면 보이는 것:

1. **공통 인식:** 두 리뷰 모두 Satterstrom 2020을 ASD 유전자 발견의 “이정표”로 간주하며, 102개 유전자 목록과 통계적 기준을 강조한다.
2. **맥락의 차이:**
  - 2023년 리뷰는 이 연구를 “뇌 발달의 시공간적 수렴” 연구의 출발점으로 활용
  - 2025년 리뷰는 “임상 진단과 치료 개발”을 위한 기초 자료로 활용
3. **연구의 계보:**
  - **이전:** De Rubeis et al. 2014 (33개 유전자), Sanders et al. 2015 (65개 유전자)
  - **Satterstrom 2020:** 102개 유전자, TADA 방법론 확장
  - **이후:** Fu et al. 2022 (72개 유전자, CNV 포함), Trost et al. 2022 (53개 유전자, WGS 기반)

이처럼 하나의 핵심 논문이 두 개의 다른 리뷰에서 어떻게 인용되는지 추적하면, 그 논문이 분야에서 차지하는 위치와 이후 연구 방향을 파악할 수 있다. Satterstrom 2020은 “새로운 발견”이라기보다는 “체계적인 목록 구축”으로 기능했고, 이 목록이 이후 모든 연구의 기준점이 되었다.

---

## 데이터베이스 탐색 전략

PubMed, Web of Science, Scopus 같은 데이터베이스를 활용한다.

처음에는 넓게 검색하고, 그다음에는 키워드 조합을 줄여가며 핵심 논문만 남긴다. Boolean 연산자(AND, OR, NOT) 조합을 이용하면 검색 정확도를 높일 수 있다.

단계별 문헌 조사 깊이를 정리하면 다음과 같다.

1. **1차 조사:** 리뷰 논문 3-5편 (큰 그림 파악)
2. **2차 조사:** 핵심 원저 논문 10-15편 (방법론과 결과 이해)
3. **3차 조사:** 최신 논문 업데이트 (지난 1-2년)
4. **Citation chaining:** 핵심 논문의 참고문헌과 그 논문을 인용한 최신 논문 추적

논문 탐색은 하루 이틀에 끝나지 않는다. 한 프로젝트를 진행하는 동안에도 새로운 논문이 나오면 계속 업데이트해야 한다.

**Predatory Journal 논문은 피하라.** MDPI, Frontiers와 같은 일부 출판사의 논문은 동료 심사가 느슨하여 질이 낮은 경우가 많다. 특히 이런 저널들은 기존 연구를 중복적으로 정리한 리뷰 논문이 많아서, 논문 조사에 실질적인 도움이 되지 않는다.

고품질 리뷰를 찾으려면 Nature Reviews, Cell, Annual Reviews 시리즈처럼 impact factor가 높고 편집 기준이 엄격한 저널을 우선적으로 찾아야 한다.

---

## EndNote를 이용한 논문 관리

EndNote는 참조 수집, 검색, 그리고 MS Word 인용에 가장 효율적인 도구다. 나는 각 논문을 작성할 때마다 그 논문 전용 폴더를 만들고, 그 안에 독립된 .enl 파일을 생성한다. 논문마다 별도의 라이브러리를 유지하는 이유는, 프로젝트가 완결될 때까지 그 논문에 집중할 수 있고, 나중에 다시 참조할 때도 혼란이 없기 때문이다. 필요하다면 압축 백업(.enlx) 형태로 보관한다. 라이브러리를 만든 뒤에는 논문을 추가하고 정리한다. 데이터베이스에서 내보내기(Export), PDF 가져오기(Import), 또는 수동 입력 방식으로 참조를 추가한다. EndNote는 자동으로 저자, 제목, 연도 같은 메타데이터를 인식하지만, 정확하지 않을 때는 직접 수정한다.

그룹 기능을 이용하면 “읽은 논문(Read)” “리뷰 대상(To Review)” 처럼 상태별로 정리할 수 있다. EndNote 내부 검색 기능을 활용하면 라이브러리 전체를 빠르게 탐색할 수 있다. 특정 저널만 정렬해서 논문 수를 세거나, 조건에 맞는 논문을 자동 분류하는 Smart Group을 활용하면 문헌 정리가 훨씬 효율적이다.

---

## 서지 관리의 습관

서지 관리의 핵심은 도구가 아니라 습관이다. 참조 파일을 프로젝트 시작 단계부터 정리해두면, 서론을 쓸 때 시간을 절반으로 줄일 수 있다. 논문을 읽을 때 핵심 문장을 요약하고, 태그로 “결과 요약” “한계점” “적용 기술” 등을 남겨두면 나중에 서론을 쓸 때 매우 유용하다.

**읽으면서 메모하는 템플릿:** - 연구 질문: 이 논문이 답하려는 핵심 질문은? - 방법: 어떤 데이터와 분석을 사용했는가? - 핵심 결과: 가장 중요한 발견 1-2개 - 한계점: 저자가 인정한 약점 또는 내가 발견한 문제 - 내 연구와의 연결점: 이 논문이 내 서론/방법/고찰에 어떻게 쓰일까?

그리고 무엇보다 중복 참조나 출처 누락은 연구 신뢰도에 직결된다. 정확한 서지 관리 습관이 좋은 논문을 만든다.

**“안티 라이브러리” 효과가 있다.** 웹툰 《익명의 독서 중독자들》에서처럼, 책을 책장에 꽂아만 두어도 뭔가 아는 것 같은 기분이 드는 것처럼, EndNote 라이브러리도 비슷한 효과가 있다. 모든 논문을 당장 다 읽을 수는 없지만, 라이브러리에 정리해두면 “이 주제는 어디를 찾아봐야 한다”는 감이 생긴다.

이것은 단순한 착각이 아니다. 논문 제목과 저자명이 반복해서 눈에 들어오면서, 연구 분야의 지형도가 머릿속에 그려지기 때문이다. 서론을 쓸 때 “어디서 본 것 같은데...”라는 기억을 단 5초 만에 찾아낼 수 있는 것, 그것이 서지 관리의 진짜 힘이다.

주 단위 문헌 업데이트 루틴을 만들어두면, 최신 논문을 놓치지 않고 프로젝트 진행 중에도 지속적으로 지식을 업데이트할 수 있다. 프로젝트 종료 후에는 라이브러리를 .enlx 형태로 아카이빙하여 나중에 다시 활용할 수 있도록 보관한다.

## 8장. Introduction — 질문의 탄생

서론은 논문의 첫인상이다. 독자는 서론을 읽으며 “이 연구가 내 시간을 들일 가치가 있는가”를 판단한다. 리뷰어는 서론에서 연구의 방향과 깊이를 가늠하고, 에디터는 서론의 명료함으로 논문의 완성도를 예측한다. 그래서 서론은 단순히 배경을 나열하는 공간이 아니라, **연구의 필요성을 설득하는 장**이다.

좋은 서론은 독자를 자연스럽게 연구의 세계로 초대한다. 넓은 맥락에서 시작해 점차 구체적인 질문으로 좁혀가며, 마지막에는 “바로 이것이 우리가 답하고자 하는 질문”이라고 선언한다. 이 과정이 매끄럽게 흐를 때, 독자는 Methods로 넘어가기 전에 이미 연구의 의미를 이해하게 된다.

이번 장에서는 서론을 어떻게 구조화하고, 어떤 순서로 생각하며, 어떤 문장으로 시작하고 닫을 것인가를 다룬다. 특히 옉믹스 연구처럼 데이터가 복잡하고 방법론이 전문적인 분야에서, 서론은 독자가 연구의 큰 그림을 잃지 않도록 돕는 나침반 역할을 한다.

---

## 인트로란 무엇인가

서론은 논문의 '이야기'를 시작하는 첫 장면이다. 영화의 오프닝 시퀀스처럼, 서론은 독자에게 “왜 이 연구가 필요한가”를 설득하는 공간이다. 단순히 “이런 연구들이 있었다”를 나열하는 것이 아니라, **문제의식(problem statement)**과 **연구의 맥락(context)**을 설정하는 장이다.

서론의 핵심 목적은 세 가지다:

1. **맥락 제공(Context)**: 독자가 연구 주제의 중요성을 이해하도록 돕는다.
2. **공백 제시(Gap)**: 기존 연구에서 아직 답하지 못한 질문을 명확히 한다.
3. **목적 선언(Objective)**: 이 연구가 그 공백을 어떻게 채울 것인지 제시한다.

서론은 독자의 머릿속에 ‘이 논문이 어디로 가는가’를 심어주는 지도다. 독자가 Results를 읽으며 길을 잃지 않도록, 서론에서 명확한 방향을 제시해야 한다.

실제 논문에서 이 구조가 어떻게 작동하는지 살펴보자. 2018년 Science에 발표된 자폐 스펙트럼 장애의 비부호화 변이 연구 (An et al. 2018, Science)의 서론은 이 세 가지 요소를 명확히 보여준다.

**맥락 제공**: 서론은 자폐 스펙트럼 장애(ASD)가 높은 유전율을 보이는 신경발달장애이며, 수백 개의 유전자가 관여한다는 점에서 시작한다. 기존 연구들이 단백질 코딩 변이를 통해 상당한 진전을 이뤘음을 인정하며, 이 분야의 중요성을 확립한다.

**공백 제시**: 그러나 단백질 코딩 변이만으로는 ASD의 유전적 위험을 완전히 설명할 수 없다는 한계를 지적한다. 인간 유전체의 98%를 차지하는 비부호화 영역이 유전자 조절에 핵심적 역할을 한다는 것이 알려져 있지만, ASD에서 비부호화 변이의 기여도는 체계적으로 평가되지 않았다. 특히 기존의 genome-wide association study (GWAS) 접근법은 흔한 변이에 집중했고, 희귀한 비부호화 변이의 역할은 불분명했다.

**목적 선언**: 이 연구는 whole-genome sequencing (WGS) 데이터를 이용해, 비부호화 영역에서 드 노보 변이와 inherited 변이를 체계적으로 분석한다. 특히 Category-Wide Association Study (CWAS)라는 새로운 방법론을 개발해, 기능적으로 관련된 비부호화 영역들을 카테고리리 묶어 통계적 검정력을 높인다. 이를 통해 프로모터, 인핸서, 그리고 뇌 특이적 조절 영역에서 ASD와 연관된 변이를 식별한다.

이 세 가지 요소가 논리적으로 연결될 때, 독자는 “그래서 이 연구를 해야겠구나”라고 자연스럽게 공감하게 된다. 서론이 단순히 배경 지식을 나열하는 것이 아니라, **연구 질문의 필연성**을 구축하는 과정임을 보여주는 좋은 예다.

---

## 인트로를 생각하기 vs 인트로를 작성하기

많은 학생들이 서론을 쓸 때 막막함을 느낀다. 그 이유는 **생각하기(thinking)**와 **작성하기(writing)**를 구분하지 않기 때문이다. 이 두 과정은 서로 다른 단계이며, 각각 다른 방식으로 접근해야 한다.

**생각하기**는 연구의 질문을 찾는 과정이다. 이것은 책상 앞에 앉아서만 하는 일이 아니다. 오히려 캠퍼스를 걷다가, 커피를 마시다가, 혹은 집으로 가는 길에 문득 떠오르는 경우가 많다. 연구실 동료와 가벼운 대화를 나누면서 “이 데이터가 의미하는 게 뭘까?”라고 물어볼 수도 있고, 미팅에서 지도교수의 질문에 답하다가 새로운 방향을 발견할 수도 있다. 생각은 물과 같아서, 누구도 제약할 수 없고, 언제 어느 순간에나 할 수 있다. 손흥민이 홈런을 치고, 오타니가 골을 넣는 스포츠 중계를 볼 때도 할 수 있다.

이 단계에서는 “무엇이 아직 알려지지 않았는가?”, “왜 이 질문이 중요한가?”, “내 연구가 기존 연구와 어떻게 다른가?”를 자유롭게 고민한다. 완벽한 문장을 만들려고 하지 말고, 노트에 낙서하듯 생각의 흐름을 따라가는 것이 중요하다. 스케치하듯이 생각을 풀어내면 된다. 이 과정은 엄밀하지 않아도 되고, 때로는 이미 연구된 것들과 질문이 중복될 수도 있다. 그런 질문들은 나중에 문헌 조사를 통해 걸러질 것이다. 막상 쓰려고 하면, 특히 첫 논문을 쓸 때는 긴장하거나 너무 부담을 느껴서 생각이 굳어지기도 한다. 그래서 쓰기 전에 충분히 생각하는 시간을 갖는 것이 중요하다.

**작성하기**는 그 생각을 논리적 서사로 엮는 과정이다. 말 그대로 MS Word를 켜고 문서를 만드는 단계다. 이때부터는 그간 찾은 참고문헌이나 서론의 기본적인 양식들을 맞춰서 작성한다. 처음에는 개조식으로 bullet point로 구조를 잡아서 작성한다. 역삼각형 구조(아래에서 자세히 다룬다)를 고려해서, 첫 문단은 무엇을 다룰 것인지, 두 번째 문단은 어떤 공백을 제시할 것인지, 마지막 문단은 연구의 목적을 어떻게 선언할 것인지를 먼저 정리하는 것이다.

구조가 잡히면 그다음에는 문장으로 작성한다. 독자가 이해할 수 있는 순서로 정보를 배열하고, 각 문단이 다음 문단으로 자연스럽게 이어지도록 연결한다. 전문 용어를 적절히 정의하고, 인용을 배치한다. 첫 작성은 AI 도구를 사용해서 해봐도

된다. 생각을 정리한 bullet point를 AI에 넣어 초안을 만들고, 그것을 다듬어가는 방식도 효율적이다. 다만 AI가 만든 문장을 그대로 쓰는 것이 아니라, 자신의 생각이 제대로 담겼는지 확인하고 수정하는 과정이 반드시 필요하다.

좋은 서론은 생각의 흔적을 문장으로 정리한 결과물이다. 먼저 충분히 생각하고, 그 다음에 쓰기 시작하라. 생각이 정리되지 않은 상태에서 문장을 만들려고 하면, 글이 산만해지고 논리가 약해진다. 반대로 생각은 충분히 했는데 쓰기를 미루면, 그 생각들이 흩어져 버린다. 두 과정 모두 중요하고, 각각의 시간을 존중해야 한다.

---

## 인트로의 역삼각형 구조: 세 개의 문단

서론의 가장 효과적인 구조는 **역삼각형(inverted pyramid)**이다. 넓은 주제에서 시작해 점차 좁혀가며, 마지막에는 구체적인 연구 질문에 도달하는 형태다. 이 구조는 독자가 자연스럽게 “그래서 이 연구가 왜 필요한가”를 이해하게 만든다.

기본적인 세 문단 구조:

### 1문단: 넓은 배경(Broad context)

주제의 중요성과 학문적·사회적 의미를 제시한다. 독자가 “왜 이 주제가 중요한가”를 이해하도록 돕는다.

예시: > “Autism spectrum disorder (ASD) affects 1–2% of children worldwide and presents significant challenges for affected families and healthcare systems. Understanding the genetic architecture of ASD is crucial for early diagnosis, intervention, and the development of targeted therapeutics.”

이 문단에서는 연구 주제가 다루는 현상의 규모를 제시하고, 사회적·임상적 중요성을 언급한다. 독자에게 “이것은 중요한 문제다”라는 인식을 심어준다.

### 2문단: 공백(Gap) 제시

지금까지 알려진 것과 아직 모르는 것의 간극을 설명한다. 이 문단이 서론의 핵심이다.

예시: > “While exome sequencing studies have identified over 100 high-confidence ASD risk genes, these account for only 10–20% of cases. The majority of genetic liability remains unexplained. Recent studies suggest that noncoding regulatory variants may contribute substantially to ASD risk, but systematic analyses integrating cell-type-specific regulatory information are lacking.”

이 문단에서는 기존 연구의 성과를 인정하지만, 여전히 남은 질문을 명확히 한다. “무엇을 모르는가”를 구체적으로 제시하는 것이다.

공백을 제시할 때 주의할 점이 있다. 기존 연구를 폄하하지 않는다. “Previous studies were flawed”가 아니라 “remain incompletely understood”로 표현한다. 그리고 너무 넓은 공백이 아니라, 내 연구로 답할 수 있는 구체적인 공백을 제시해야 한다.

**때로는 공백이 두 개의 서로 다른 도메인에서 발생하기도 한다.** 예를 들어, 한 연구가 지식의 공백과 방법론적 공백을 동시에 다룰 수 있다. 기존 연구들이 특정 측면만 분석했다는 지식의 한계와, 적절한 분석 방법이나 데이터가 없었다는 방법론적 한계가 함께 존재하는 경우다. 이럴 때는 공백을 두 문단으로 나누는 것이 자연스럽다. 첫 번째 공백 문단에서는 “무엇을 모르는가”를, 두 번째 공백 문단에서는 “어떤 리소스나 방법이 부족했는가”를 다룬다. 그러면 서론이 세 문단이 아니라 네 문단이 된다. 이것은 전혀 문제가 없다. 중요한 것은 역삼각형의 논리적 흐름이 유지되는가이지, 문단의 개수가 아니다.

실제 논문에서 이 두 가지 방식을 비교해보자. 2018년 Science 논문 (An et al. 2018, Science)은 전형적인 세 문단 구조를 사용한다. 첫 문단에서 자폐의 유전적 복잡성과 비부호화 영역의 중요성을 제시하고, 두 번째 문단에서 기존 연구의 한계(비부호화 변이의 역할이 불분명함)를 하나의 통합된 공백으로 제시한다. 마지막 문단에서 연구의 목적과 주요 발견을 제시한다. 공백이 하나의 큰 주제(비부호화 영역의 미탐색)로 수렴하기 때문에, 한 문단으로 효과적으로 정리된다.

반면 2020년 Cell Reports 논문 (Werling et al. 2020, Cell Reports)은 네 문단 구조를 사용한다. 첫 문단에서 뇌 발달의 복잡성과 시공간적 조절의 중요성을 제시한다. 두 번째 문단에서는 발달 과정에서의 유전자 발현 변화가 신경정신질환과 연관된다는 지식의 공백을 제시한다. 세 번째 문단에서는 eQTL 연구의 중요성을 언급하면서, 특히 발달 중인 뇌에서의 eQTL 데이터가 부족하다는 리소스의 공백을 다룬다. 네 번째 문단에서 BrainVar 리소스를 소개하며 연구의 목적을

밝힌다. 여기서 공백이 두 차원으로 나뉜다. 발달 과정에서 유전자 조절이 질병에 어떻게 기여하는가라는 생물학적 지식의 공백과, 발달 중인 뇌의 eQTL 데이터가 부족하다는 리소스/방법론적 공백이다. 이 두 공백이 독립적이면서도 연결되어 있기 때문에, 두 문단으로 나누어 설명하는 것이 논리적으로 더 명확하다.

어느 방식이 더 나은가? 그것은 연구가 답하려는 질문의 구조에 달려 있다. 만약 하나의 명확한 공백을 다룬다면 세 문단이 깔끔하다. 하지만 두 개의 독립적인 공백을 다루거나, 공백의 총위가 복잡하다면 네 문단이 더 명료할 수 있다. 중요한 것은 각 문단이 논리적으로 연결되고, 독자가 자연스럽게 다음 질문으로 넘어갈 수 있는 것이다.

### 3문단: 목적(Objective) 제시

본 연구의 목표, 질문, 접근법을 명료하게 제시한다. 독자가 “이 논문이 무엇을 할 것인가”를 정확히 이해하도록 한다.

예시: > “In this study, we performed whole-genome sequencing of 2,000 ASD families to systematically evaluate the contribution of noncoding variants. We integrated cell-type-specific chromatin accessibility data from developing human brain to prioritize functional regulatory variants. Our analysis reveals that noncoding variants in cortical neuron enhancers are significantly enriched in ASD probands and contribute to phenotypic severity.”

이 문단에서는 연구 디자인의 핵심을 간단히 소개하고, 사용한 데이터와 방법의 독창성을 강조한다. 주요 발견을 미리 언급할 수도 있다(저널에 따라 다름).

**이 세 문단(혹은 네 문단)의 흐름이 매끄러울 때, 독자는 자연스럽게 연구의 논리를 따라오게 된다.** 각 문단이 다음 문단의 동기가 되도록 연결하는 것이 핵심이다. 첫 문단을 읽고 나면 “그래서 우리가 무엇을 알고 있나?”라는 질문이 생기고, 공백 문단을 읽고 나면 “그럼 어떻게 이 문제를 풀 것인가?”라는 질문이 생기고, 마지막 문단이 그 답을 제시하는 구조다. 이것이 역삼각형의 힘이다.

---

## 인트로에서 던진 질문(‘떡밥’)은 결과(Results)와 토의(Discussion) 섹션에서 회수되어야 한다

서론은 질문을 던지는 장이고, 결과와 토의는 그 질문에 대한 ‘답’을 주는 장이다. 좋은 논문은 서론에서 흘려버린 작은 단서(떡밥)를 후반부에서 자연스럽게 회수한다.

예를 들어, 서론에서 이렇게 썼다면: > “However, it remains unclear whether noncoding variants show cell-type-specific effects during brain development.”

결과 섹션에서는: > “We found that regulatory variants are enriched in cortical excitatory neurons specifically during mid-fetal development (Figure 3).”

토의 섹션에서는: > “Our findings demonstrate that noncoding ASD variants exhibit strong cell-type and developmental stage specificity, suggesting that...”

**“문제 제기 → 탐색 → 회수”의 구조를 통해 논문 전체의 리듬이 완성된다.** 서론에서 던진 질문이 결과와 토의에서 제대로 회수되지 않으면, 독자는 “그래서 이 연구가 원래 질문에 답했나?”라는 의문을 갖게 된다.

그런데 떡밥을 회수하지 않는 학생들이 있다. 특히 MBTI가 N(직관)과 P(인식) 성향이 높은 학생들은 서론에서 떡밥을 과도하게 뿌리는 경향이 있다. A, B, C 순서로 주제를 나열해야 하는데, 서론에서 세상 삼라만상, 모든 것을 다 설명하려고 한다. 유전자 조절 메커니즘부터 시작해서 진화적 관점, 질병의 역학, 치료법의 미래, 사회적 함의까지... 읽다 보면 눈알이 빠질 것 같다. 사실 그런 나열들은 정말 재미있고, 학생의 지적 호기심과 배경 지식을 보여준다. 하지만 서론은 리뷰 논문도 아니고 위키피디아가 아니다.

**서론은 글로 작성되지만, 사실 제품 브로셔와 같다.** 어떤 내용을 당신이 읽을 것이고, 어떤 경로로 생각의 흐름이 흐를지를 보여주는 것이다. 서론에서 언급한 모든 주제는 논문의 어딘가에서 다시 다뤄져야 한다. 만약 서론에서 “유전자 네트워크가 중요하다”고 강조했다면, Results에 네트워크 분석이 있어야 하고, Discussion에서 그 네트워크의 의미를 해석해야 한다. 만약 서론에서 “발달 시기별 차이”를 언급했다면, 결과에서 그 차이를 보여주는 Figure가 있어야 한다.

**떡밥은 반드시 회수되어야 한다.** 그렇지 않으면 독자는 “왜 이 이야기를 꺼냈지?”라고 혼란스러워한다. 리뷰어는 더 직설적이다. “서론에서 언급한 XX는 결과에서 다뤄지지 않았다. 이 부분을 삭제하거나, 결과를 추가하라”는 코멘트를 남긴다.

**학생 연습:** 서론을 쓴 후, 다음 질문에 답해보라. 서론의 각 문단을 읽으며, 그 문단에서 제기한 질문이나 언급한 개념이 논문의 어디에서 다시 등장하는지 추적해보는 것이다. 구체적으로는 이렇게 확인할 수 있다. 내가 서론에서 던진 핵심 질문은 무엇인가? 그 질문은 Results의 어느 Figure에서 답해지는가? Discussion에서 그 질문을 다시 언급하고 해석하는가? 만약 서론에서 언급했지만 Results나 Discussion에 나타나지 않는 내용이 있다면, 그것은 두 가지 중 하나다. 서론에서 불필요하게 언급했거나, 결과 분석이 미흡한 것이다.

이 연결고리가 명확하지 않다면, 서론을 다시 다듬거나 결과의 구성을 조정해야 한다. 때로는 서론을 줄이는 것이 답이고, 때로는 결과를 보강하는 것이 답이다. 중요한 것은 서론에서 던진 공과 Results-Discussion에서 받는 공이 정확히 일치해야 한다는 점이다.

**실제 예시:** 내가 지도한 학생 중 한 명은 서론에서 “유전자 네트워크의 허브 유전자가 자폐에 중요한가?”라는 질문을 던졌다. 서론의 2/3을 할애해서 네트워크 이론의 중요성을 설명했고, 허브 유전자의 역할을 강조했다. 그런데 결과에서는 개별 유전자의 변이 빈도만 분석하고, 네트워크 분석은 보조자료(Supplementary Figure)에만 포함시켰다. 메인 Figure에는 단 하나의 네트워크 그림도 없었다. 리뷰어는 이 불일치를 날카롭게 지적했다. “서론에서 네트워크를 강조했는데, 왜 메인 결과에는 없는가? 이것은 연구의 초점이 불분명하다는 것을 보여준다.” 우리는 네트워크 분석을 메인 Figure로 옮기고, Discussion에서도 네트워크 관점의 해석을 추가해서 서론의 질문에 직접 답하도록 수정했다. 논문은 그제야 리뷰를 통과했다.

서론을 쓸 때는 욕심을 부리고 싶은 유혹이 크다. 자신이 아는 모든 것을 보여주고 싶고, 연구 주제의 모든 측면을 다루고 싶다. 하지만 좋은 서론은 선택과 집중이다. 내 연구가 답할 수 있는 질문만 던지고, 그 질문에 대한 답을 Results와 Discussion에서 명확히 제시하는 것. 그것이 딱박을 제대로 회수하는 법이다.

---

## 인트로 첫 문단, 첫 문장

서론의 첫 문장은 독자의 ‘첫 3초’를 사로잡아야 한다. 이 문장이 독자를 연구의 세계로 초대하는 문이기 때문이다.

**좋은 첫 문장의 조건:** - 너무 기술적이지 않으면서도, 주제의 중요성을 명확히 전달한다. - 넓지만 모호하지 않은, 구체적인 맥락을 제공한다. - 독자가 “그래서 뭐?”라는 질문 없이 자연스럽게 다음 문장으로 넘어가도록 한다.

**좋은 예시:** > “Human genetic variation shapes individual differences in health, behavior, and disease susceptibility.”

“Autism spectrum disorder (ASD) is a highly heritable neurodevelopmental condition affecting 1 in 54 children.”

“The noncoding genome, comprising 98% of human DNA, harbors thousands of regulatory elements that control gene expression.”

**피해야 할 첫 문장:** - 너무 넓음: “Genetics is important.” (무의미하게 일반적) - 너무 좁음: “The rs12345 SNP has been associated with...” (독자가 맥락을 모름) - 역사적 서술: “In 1953, Watson and Crick discovered...” (특별한 이유가 없다면 불필요)

**첫 문장을 쓰는 실용적 방법:** 1. 내 연구가 다루는 현상이나 문제를 한 문장으로 정의한다. 2. 그 현상이 왜 중요한지 또는 흥미로운지를 암시한다. 3. 문장을 큰 소리로 읽어보며, “이 문장이 내 연구를 열기에 충분히 강력한가?”를 자문한다.

**‘왜 이 문장을 써야 하는가’를 스스로 설명할 수 있어야 한다.** 첫 문장은 습관적으로 쓰는 것이 아니라, 의도적으로 선택하는 것이다.

---

## 인트로에서 주요 단어를 정의하여 독자의 머릿속에 색인시키는 방법

서론은 단순히 연구를 소개하는 것을 넘어, 독자의 인지적 색인(cognitive index)을 설정하는 역할을 한다. 즉, 독자가 논문을 읽는 동안 특정 개념을 계속 떠올릴 수 있도록 핵심 용어를 초기 문단에 배치해야 한다. 이것은 독자의 머릿속에 개념의 지도를 그려주는 일이다.

예를 들어, 오믹스 연구에서 자주 등장하는 개념들이 있다. 희귀 변이(rare variant), 세포 유형 특이성(cell-type specificity), 유전자 조절(gene regulation), 드 노보 변이(de novo mutation), 인핸서(enhancer) 같은 용어들이다. 이런 용어들을 서론 초반에 정의하고 맥락 속에 위치시키면, 독자는 Results를 읽을 때 이 개념들을 자연스럽게 연결할 수 있다.

좋은 정의는 간결하면서도 맥락을 제공한다. 예를 들어, “Rare variants, defined as genetic variants with minor allele frequency <1%, are increasingly recognized as important contributors to complex disease risk”라고 쓰면, 독자는 희귀 변이가 무엇인지, 그리고 왜 중요한지를 동시에 이해한다. 마찬가지로 “Cell-type-specific regulatory elements, such as enhancers and promoters, control when and where genes are expressed during development”는 세포 유형 특이성이라는 개념을 예시와 함께 설명한다.

피해야 할 방식도 있다. 정의 없이 전문 용어를 남발하면 독자가 길을 잃는다. “We analyzed eQTLs in iPSC-derived neurons...”라고 쓰면, eQTL이 무엇인지 모르는 독자는 그 순간 멈춰 선다. 반대로 너무 교과서적인 정의도 문제다. “An enhancer is a regulatory DNA sequence that increases the transcription of target genes...”는 정확하지만 논문의 흐름을 끊는다. 논문은 교과서가 아니다. 정의는 연구의 맥락 안에서 자연스럽게 녹아들어야 한다.

그런데 학생들 중에는 서론에 같은 개념을 여러 용어로 설명하는 경우가 있다. 예를 들어, regulatory element, regulatory region, regulatory sequence, cis-regulatory element를 모두 다른 용어인 것처럼 사용한다. 혹은 de novo mutation과 spontaneous mutation을 구분 없이 섞어 쓴다. 그러면 독자나 리뷰어는 주요 개념을 인지하지 못한다. “이것들이 같은 것인가, 다른 것인가?”라는 혼란이 생긴다.

서론에서는 주요 개념을 빠르게 상대방의 머릿속에 남겨서, 뒤에 나올 내용을 쉽게 읽게 해야 한다. 그러기 위해서는 몇 가지 주요 용어들을 명확히 제시하고, 필요하다면 줄임말로 야무지게 포장해줘야 한다. “We refer to these as regulatory variants (RVs) throughout this paper”라고 선언하면, 이후 논문 전체에서 RV라는 약어를 일관되게 사용할 수 있다.

하지만 너무 많은 용어를 남발하거나, 너무 많은 줄임말을 등장시키면 안 된다. 서론 첫 페이지에 약어가 10개 이상 등장하면, 독자는 약어를 기억하는 데 에너지를 다 쓰고 내용은 머리에 들어오지 않는다. 그리고 줄임말 용어들을 너무 많이 등장시키면, 문장의 호흡이 길어져서 독자가 글을 읽기가 어렵다. “We performed CWAS on RVs in TF-binding CREs using scATAC-seq data from iPSC-derived ExNs”라는 문장은 기술적으로는 정확하지만, 약어 때문에 독자가 숨이 막힌다.

**글은 글이지만, 제품의 브로셔와 같다.** 글에는 공간적 흐름이 있다. 독자의 시선이 왼쪽에서 오른쪽으로, 위에서 아래로 움직이며 정보를 흡수한다. 그 흐름 속에서 독자가 멈춰야 하는 지점(새로운 개념의 정의)과 빠르게 넘어갈 수 있는 지점(이미 정의된 개념의 반복)을 설계해야 한다.

실용적인 전략을 제안한다. 서론 초안을 쓴 후, 핵심 용어 5-7개를 리스트로 정리한다. 각 용어가 서론 어디에서 처음 등장하는지 확인한다. 처음 등장할 때 한 문장으로 맥락적 정의를 제공한다. 그리고 이후 Results와 Discussion에서 그 용어가 일관되게 사용되는지 확인한다. 만약 서론에서 “regulatory variant”로 정의했다면, Results에서 갑자기 “regulatory mutation”으로 바꿔 쓰지 않는다. 용어의 일관성은 독자의 이해를 돕는 가장 기본적인 배려다.

용어를 정의하는 것은 독자를 존중하는 일이다. 전문가 독자에게는 확인의 기회를, 비전문가 독자에게는 이해의 발판을 제공하는 것이다. 그리고 무엇보다, 명확한 용어 정의는 저자 자신의 생각을 정리하는 데도 도움이 된다. “내가 정확히 무엇을 말하고 있는가?”를 용어를 통해 확인할 수 있기 때문이다.

---

## 인트로 마지막 문단 — ‘무엇을 했는가’ vs ‘무엇을 할 것인가’

서론의 마지막 문단은 논문 전체의 방향을 달는 자리이지만, **저널이나 분야에 따라 접근 방식이 다르다**. 크게 두 가지 스타일이 있다:

### 스타일 1: 연구 전체를 요약 (past tense)

일부 저널(특히 Cell, Science, Nature 등 고영향력 저널)은 서론 마지막에 **연구의 주요 발견까지 미리 언급**하는 방식을 선호한다.

예시: > “In this study, we integrated single-cell chromatin accessibility data with whole-genome sequencing from 2,000 ASD families. We identified 45 noncoding loci significantly enriched for rare variants in cortical neuron enhancers. These variants are associated with increased ASD severity and converge on gene regulatory networks critical for early brain development.”

**장점:** - 독자가 논문의 주요 발견을 미리 알고 Results를 읽을 수 있다. - 연구의 임팩트를 강조하기 좋다.

**단점:** - 결과를 너무 많이 공개하면 Results 섹션이 반복적으로 느껴질 수 있다.

## 스타일 2: 연구 목적까지만 제시 (future tense or present tense)

다른 저널(Nature Communications, 많은 의학 저널)은 서론에서 **연구의 목적과 접근법**까지만 밝히는 방식을 허용한다.

예시: > “Here, we aim to investigate how noncoding regulatory variants contribute to ASD risk. We integrate cell-type-specific epigenomic data with whole-genome sequencing to identify functional variants in neurodevelopmental regulatory elements.”

**장점:** - 서론이 간결하고, 결과에 대한 기대감을 유지할 수 있다. - 예상치 못한 발견이 있을 때 유연하게 서술할 수 있다.

**단점:** - 일부 저널에서는 “impact이 명확하지 않다”는 피드백을 받을 수 있다.

## 어떤 스타일을 선택할 것인가?

두 접근 모두 허용되며, 핵심은 **서론의 마지막 문장이 독자에게 이 논문이 어디로 향하는지 분명히 알려주는 것**이다.

**실용적 조언:** 1. **목표 저널의 최근 논문 5편을 읽어보라.** 서론 마지막 문단이 어떤 스타일인지 확인한다. 2. **저널의 author guidelines를 확인하라.** 일부 저널은 명시적으로 선호 스타일을 밝힌다. 3. **지도교수나 공동저자와 논의하라.** 분야의 관습이 중요하다.

**내 개인적 선호:** 나는 주로 스타일 1을 사용한다. 독자가 논문의 주요 메시지를 서론에서 이미 이해하고, Results를 읽으며 “어떻게 이 결론에 도달했는가”에 집중할 수 있기 때문이다. 특히 오믹스 연구처럼 방법론이 복잡한 경우, 서론에서 큰 그림을 먼저 보여주는 것이 독자에게 친절하다.

하지만 **정답은 없다.** 중요한 것은 일관성이다. 어떤 스타일을 선택하든, 서론 전체의 톤과 논리적 흐름이 자연스럽게 마지막 문단으로 이어져야 한다.

---

## Introduction 마지막 문단 vs Abstract: 어떻게 다른가?

많은 학생들이 헷갈려하는 질문이다. Introduction의 마지막 문단과 Abstract 모두 연구를 요약하는데, 왜 두 번 같은 내용을 쓰는 것처럼 느껴질까? 사실 두 섹션은 **목적, 독자, 그리고 서술 방식이 완전히 다르다.**

### Abstract: 논문 전체의 압축본

Abstract는 **논문 전체의 독립적인 요약**이다. 독자가 Abstract만 읽고도 연구의 배경, 방법, 결과, 결론을 모두 이해할 수 있어야 한다. 데이터베이스 검색에서 Abstract만 표시되는 경우가 많기 때문에, Abstract는 논문과 독립적으로 존재해야 한다.

### Abstract의 구조 (보통 200-250단어):

1. **배경 (1-2문장):** 연구 주제의 중요성
2. **목적 (1문장):** 무엇을 조사했는가
3. **방법 (2-3문장):** 어떤 데이터와 분석을 사용했는가
4. **결과 (3-4문장):** 주요 발견 (숫자 포함)
5. **결론 (1-2문장):** 연구의 의미와 함의

### Abstract 예시:

Autism spectrum disorder (ASD) is highly heritable, yet the genetic architecture remains incompletely understood. **[배경]** Here, we performed whole-genome sequencing of 2,000 ASD families to systematically evaluate noncoding regulatory variants. **[목적]** We integrated cell-type-specific chromatin accessibility data from developing human brain and applied a novel Category-Wide Association Study (CWAS) framework. **[방법]** We identified 45 noncoding loci significantly enriched for rare variants in cortical neuron enhancers (FDR < 0.1). These variants

were associated with increased symptom severity ( $\beta = 0.42$ ,  $P = 3.2 \times 10^{-5}$ ) and converged on 12 gene regulatory networks critical for early cortical development. **[결과]** Our findings demonstrate that noncoding regulatory variants contribute substantially to ASD risk and highlight the importance of cell-type-specific analyses in psychiatric genomics. **[결론]**

#### Introduction 마지막 문단: 논문으로의 전환점

반면 Introduction의 마지막 문단은 **Abstract**를 읽은 독자가 이제 본문으로 들어가는 준비 단계다. Abstract는 이미 읽었다고 가정하고, 더 구체적인 맥락과 연결고리를 제공한다.

#### Introduction 마지막 문단의 역할:

1. 서론에서 제기한 공백을 어떻게 채울 것인지 구체적으로 선언
2. 사용한 데이터와 접근법의 독창성 강조
3. 주요 발견의 의미를 서론의 맥락과 연결

#### 같은 연구의 Introduction 마지막 문단 예시:

To address these gaps, we performed whole-genome sequencing of 2,000 simplex ASD families and integrated single-cell chromatin accessibility profiles from human fetal cortex spanning 8–24 post-conception weeks. Unlike previous genome-wide association studies that focused on common variants, we developed a Category-Wide Association Study (CWAS) framework to systematically test functional noncoding categories for enrichment of rare de novo and inherited variants. This approach allowed us to identify cell-type- and developmental-stage-specific regulatory elements contributing to ASD risk. We discovered that rare variants in cortical excitatory neuron enhancers are significantly enriched in ASD probands and that variant burden in these regions correlates with clinical severity. These findings provide the first systematic evidence that noncoding regulatory variants, particularly those active during early cortical neurogenesis, play a substantial role in ASD etiology.

#### 핵심 차이점 비교

측면	Abstract	Introduction 마지막 문단
독자 가정	논문을 처음 보는 사람	서론 전체를 읽은 사람
맥락 의존성	독립적 (논문 없이도 이해 가능)	의존적 (서론의 공백 제시를 전제)
길이	200–250단어 (제한 있음)	1–2 문단 (더 자유로움)
방법 서술	간략 (“WGS 수행”)	구체적 (“2,000 simplex families”)
결과 서술	핵심 숫자 위주	결과의 의미 강조
시제	주로 과거형	과거형 + 현재형 혼합
톤	객관적, 중립적	설득적, 연결적

#### 실전 작성 전략

##### 1. Introduction 마지막 문단에서만 언급할 내용:

- “To address these gaps...” – 앞서 제시한 공백과의 명시적 연결
- 데이터의 독특한 특성: “spanning 8–24 post-conception weeks” (발달 시기별 데이터)
- 방법론의 혁신성: “Unlike previous GWAS...” (기존 연구와의 차별점)
- 결과가 공백을 어떻게 채우는지: “provide the first systematic evidence...”

##### 2. Abstract에서만 언급할 내용:

- 정량적 결과의 구체적 수치: “FDR < 0.1”, “ $\beta = 0.42$ ,  $P = 3.2 \times 10^{-5}$ ”
- 샘플 크기: “2,000 families”
- 발견한 요소의 개수: “45 noncoding loci”, “12 gene regulatory networks”

- **간결한 결론:** “highlight the importance of cell-type-specific analyses”

### 3. 두 곳 모두 언급하되 방식을 다르게:

**데이터 설명:** – Abstract: “whole-genome sequencing of 2,000 ASD families” – Introduction: “2,000 simplex ASD families… spanning 8–24 post-conception weeks”

**방법론:** – Abstract: “Category-Wide Association Study (CWAS) framework” – Introduction: “Unlike previous GWAS… we developed CWAS to systematically test functional categories…”

**주요 발견:** – Abstract: “45 loci… associated with symptom severity ( $\beta = 0.42$ )” – Introduction: “rare variants… are significantly enriched… and correlate with clinical severity”

## 흔한 실수와 해결법

### 실수 1: Introduction을 Abstract처럼 쓰기

× 나쁜 예: > “We performed WGS. We found 45 loci. These loci are important.”

□ 좋은 예: > “To systematically evaluate noncoding contributions to ASD, we performed WGS of 2,000 simplex families—the largest collection of family-based WGS data with deep phenotyping. This dataset, combined with developmental single-cell atlases, enabled us to discover that rare regulatory variants in early cortical neurons…”

**차이:** 나쁜 예는 Abstract의 압축 스타일을 그대로 가져왔다. 좋은 예는 서론의 공백(“noncoding contributions”)을 언급하고, 데이터의 독특함(“largest collection with deep phenotyping”)을 강조하며, 발견의 의미(“enabled us to discover”)를 연결한다.

### 실수 2: Abstract를 Introduction처럼 쓰기

× 나쁜 예: > “To address the gap in understanding noncoding variants, which has been a longstanding challenge in the field as noted by previous reviews, we…”

□ 좋은 예: > “Here, we performed WGS of 2,000 ASD families to evaluate noncoding regulatory variants.”

**차이:** Abstract는 배경 설명이나 문헌 검토 없이 바로 핵심으로 들어가야 한다. “To address the gap”이나 “longstanding challenge” 같은 서론 스타일의 표현은 불필요하다.

### 실수 3: 같은 문장을 복사-붙여넣기

일부 학생들은 Introduction 마지막 문단의 문장을 그대로 복사해서 Abstract에 붙여넣는다. 이것은 게으름의 표시이기도 하지만, 더 중요하게는 두 섹션의 목적을 이해하지 못했다는 신호다.

**해결법:** – Abstract를 먼저 쓰고, Introduction 마지막 문단은 나중에 작성한다 – 혹은 Introduction 전체를 완성한 후, Abstract를 독립적으로 다시 쓴다 – 두 섹션을 작성한 후, 의도적으로 **다른 단어**를 사용해 같은 내용을 표현한다

## 실전 연습

### 연습 1: 중복 체크

Introduction 마지막 문단과 Abstract를 나란히 놓고, 다음을 확인하라:

- 같은 문장이 있는가? → 있다면 하나를 다시 써라
- Introduction이 서론의 공백을 명시적으로 언급하는가?
- Abstract에는 구체적 수치가 있고, Introduction에는 의미가 강조되어 있는가?
- Introduction이 “To address…”, “Unlike previous…” 같은 연결 표현을 사용하는가?

### 연습 2: 독자 시점 전환

- **Abstract를 읽는 독자:** “이 논문이 읽을 가치가 있는가?”

- **Introduction** 마지막을 읽는 독자: “이제 Results로 넘어갈 준비가 되었는가?”

각 섹션을 읽고 위 질문에 “그렇다”고 답할 수 있는지 확인하라.

---

## 마무리: 두 섹션의 관계

Abstract와 Introduction 마지막 문단은 같은 연구를 다른 높이에서 바라보는 것이다. Abstract는 비행기에서 내려다본 전체 지형이고, Introduction 마지막 문단은 그 지형 속으로 걸어 들어가는 입구다.

Abstract는 “이 연구가 무엇을 했는가”를 독립적으로 전달하고, Introduction 마지막 문단은 “왜 이 연구가 필요했고, 어떻게 문제를 풀었는가”를 서론의 맥락 속에서 설명한다. 두 섹션 모두 연구를 요약하지만, **목적이 다르기 때문에 내용도 달라야 한다.**

좋은 논문은 Abstract를 읽고 “흥미롭다”고 생각한 독자가 Introduction을 읽으며 “이제 이해했다”고 느끼고, Results로 넘어갈 때 “어떻게 이 결과에 도달했는지 보고 싶다”는 기대를 갖게 만든다. 그 여정의 두 중요한 지점이 바로 Abstract와 Introduction 마지막 문단이다.

## 9장. Discussion과 Conclusion — 당신의 해석

Discussion의 본질은 데이터에서 통찰로 전환하는 것이다. Results가 데이터가 말하는 것이라면, Discussion은 당신이 독자에게 말하는 것이다. 실험 결과의 단순 나열이 아니라, 그 결과가 과학적 맥락에서 무엇을 의미하는지 해석하고 설명하는 공간이다. 좋은 Discussion은 Introduction에서 던진 떡밥을 하나씩 회수하며, “우리가 묻던 질문에 대한 답은 이것이다”라고 선언한다.

많은 학생들이 Discussion을 쓸 때 “First, Second, Third, Last…”로 결과를 순차적으로 나열한다. 이런 구조는 체크리스트를 완성하는 느낌을 주지만, 독자에게는 지루하고 기계적으로 다가온다. Discussion은 단순한 결과의 반복이 아니다. Introduction에서 제시한 연구 gap에 집중하며, 당신의 발견이 갖는 novelty와 implication을 설명해야 한다.

또 다른 함정은 Results의 모든 결과를 하나하나 해석하려는 시도다. 모든 데이터가 Discussion에 들어갈 필요는 없다. 주요한 발견(key findings)과 그것이 갖는 함의(implications)에 집중하라. 부차적인 관찰들은 과감히 생략하거나 간단히 언급하는 것이 오히려 메시지를 명확하게 만든다.

---

## 논문 예시로 배우는 Discussion 구조

실제 논문의 Discussion을 분석하며 효과적인 구조를 배워보자. 우리 연구실에서 자폐 가족의 유전변이 영향을 측정한 연구(Kim et al. 2025, Genome Medicine) 논문을 활용하여 설명하고자 한다.

### 핵심 발견을 먼저 선언하기

이 논문의 Discussion은 첫 문단에서 연구의 핵심 메시지를 명확히 전달한다:

“We measured the effect sizes of DNVs on behavioral symptom severity and adaptive functioning considering familial background leveraging large and ancestrally diverse ASD cohorts. This approach improved genotype–phenotype associations and led to the discovery of 18 novel ASD-associated genes. Furthermore, we identified 11 genes with high variability in intrafamilial effects…”

연구가 무엇을 했고, 무엇을 발견했는지 즉시 알 수 있다. Introduction에서 제기한 문제(“familial background를 고려하지 않으면 DNV 효과를 정확히 평가할 수 없다”)에 대한 답(“우리는 WFSD 방법으로 이를 해결했고, 18개의 새로운 유전자를 발견했다”)을 바로 제시한다. 이런 직설적인 시작은 독자가 Discussion의 방향을 바로 이해하도록 돕는다.

## Introduction의 떡밥 회수하기

이 논문의 Introduction은 “DNV의 표현형 효과가 가족 배경에 따라 크게 달라진다”는 문제를 제기했다. Discussion은 이 문제에 정확히 답한다:

“The use of WFSD provided less variable and more robust associations with ASD than raw phenotype scores, indicating improved biological interpretability.”

단순히 “우리 방법이 좋다”고 말하는 것이 아니라, Introduction에서 던진 specific한 질문에 대해 specific한 답을 제시한다. 독자는 “아, 이 연구가 정확히 그 문제를 해결했구나”라고 이해하게 된다.

## 사전 연구와의 비교를 통한 맥락 제공

Discussion에서는 자신의 결과를 기존 연구와 비교하며 맥락을 제공할 수 있다. 이 논문은 SFARI Gene database와 비교하며 검증한다:

“Comparison of gene-level effects of dnPTVs on SRS WFSD with external gene annotations (SFARI Gene) and effect size estimates demonstrated largely consistent enrichment patterns, supporting the validity of our approach.”

이런 비교는 두 가지 목적을 달성한다. 첫째, 자신의 결과가 기존 지식과 일치함을 보여 신뢰성을 높인다. 둘째, 새로운 발견(SPAG9, UNK, TNFRSF8 같은 novel genes)의 중요성을 부각시킨다.

오믹스 연구에서는 top candidate에 대해 관련 pathway나 disease relevance를 설명하는 것이 특히 중요하다. 예를 들어 이 논문은 PTEN 변이의 domain-specific effect를 설명하며:

“Missense variants within its phosphatase domain, particularly those affecting the substrate-binding and TI loop regions, exhibited significantly more severe SRS deviations... Consistent with these observations, prior multi-model functional assays demonstrated that p.Thr131Ile exhibits near-complete loss-of-function...”

단순히 “PTEN이 중요하다”가 아니라, “어떤 domain의 어떤 변이가 왜 severe한 효과를 보이는지” 기전과 함께 설명한다.

## 주제별 흐름으로 서술하기

First, Second, Third... 식 나열 대신, 이 논문은 주제별로 자연스럽게 흐른다:

1. **방법론의 우수성:** WFSD가 더 나은 association을 제공한다
2. **생물학적 통찰:** Domain-specific과 position-specific effect의 발견
3. **한계점과 미래 방향:** Sample size, ascertainment bias를 솔직하게 언급하되 발전적으로
4. **임상적 함의:** Precision intervention과 genetic counseling에 대한 응용

각 주제는 자연스럽게 다음 주제로 연결된다. “우리 방법이 좋다 → 그래서 새로운 생물학적 메커니즘을 발견했다 → 하지만 한계가 있어 더 연구가 필요하다 → 이는 임상에 이렇게 응용될 수 있다”는 논리적 흐름을 만든다.

---

## Limitation: 반성문이 아닌 발전적 비평

학생들이 자주 묻는 질문이 있다: “Limitation을 얼마나 솔직하게 써야 하나요?”

명확히 하자. Limitation 섹션은 반성문이 아니다. “우리 연구가 형편없어요”라고 자학하는 공간이 아니라, 연구의 맥락을 명확히 하고 미래 방향을 제시하는 전략적 공간이다.

## 좋지 않은 Limitation

너무 뻔한 것: “Sample size가 작았다”, “longitudinal data가 아니었다”처럼 모든 연구가 가질 수 있는 일반적 한계는 독자에게 새로운 정보를 주지 않는다.

근원적으로 해결할 수 없는 것: “우리는 생쥐 모델을 사용했기 때문에 인간과 다를 수 있다”는 생쥐 연구의 본질적 한계다. 이를 limitation으로 적는 것은 연구 설계 자체를 부정하는 것이다.

다음 연구 아이디어의 과도한 노출: “향후 연구에서는 X, Y, Z 변수를 추가로 분석할 것이다”처럼 구체적 연구 계획을 제시하면, 경쟁 연구자에게 아이디어를 주는 셈이다.

### 발전적인 Limitation

위 논문의 limitation은 발전적 비평의 좋은 예시다:

“SRS data were unavailable for the SPARK, which may reduce cohort diversity.”

단순히 “데이터가 없었다”로 끝나지 않는다. 바로 이어서:

“However, the adjustment for these covariates did not change the relative WFSD differences between dnPTV, dnMIS, and non-carriers.”

문제를 제기하되, 그것이 핵심 결론에 영향을 주지 않았음을 보인다. 이는 변증법적 접근이다. “데이터의 한계가 있다(정) → 하지만 우리는 이를 검증했다(반) → 따라서 결론은 여전히 robust하다(합)”.

Ascertainment bias에 대한 논의도 마찬가지다:

“Parental SRS T-scores were lower than the general population mean, and sibling VABS scores were higher than the general population mean, suggesting that comparisons with general population norms might underestimate the DNV effects. However, we cannot exclude the possibility of ascertainment bias being introduced, as participating families could have higher education levels and lower social impairments than the broader population.”

Bias 가능성을 솔직하게 인정한다. 하지만 바로 이어서:

“To evaluate whether such bias could systematically influence the estimated DNV effects, we conducted exploratory analyses testing interactions between parental or sibling phenotypes and DNV carrier status, which did not show any significant interaction effects.”

한계를 제시하되, 그것이 결과에 실제로 영향을 주었는지 검증한 것을 보여준다. 이것이 발전적 비평이다.

### 데이터의 양면성 설명하기

모든 데이터는 양면성을 갖는다. Within-family comparison은 familial background를 제어하는 장점이 있지만, 동시에 분석 가능한 샘플 수를 줄인다:

“While WFSD provides more specific estimates of DNV effects, the substantially smaller subset of probands with complete two-generation or sibling phenotyping imposes a further limitation. For example, while SRS scores were available in WFSD for a comparable number (Raw N = 2907; WFSD N = 2715), VABS data were not (Raw N = 10,232; WFSD N = 1955).”

Trade-off를 명확히 설명한다. “WFSD가 더 정확하지만, 샘플 수가 줄어든다”는 양면을 제시하고, “따라서 미래 연구에서는 family phenotyping을 더 많이 수집해야 한다”는 건설적 방향을 제시한다.

---

### Future Direction: 계획서가 아닌 발전 방향

또 다른 학생들의 질문: “Future Plan을 얼마나 구체적으로 써야 하나요?”

간단한 답: 연구 계획서를 쓰지 마라.

“향후 우리는 X cohort를 추가로 recruit하여 Y 분석을 수행할 것이다”는 grant proposal에 들어갈 내용이지, Discussion에 들어갈 내용이 아니다. 대신 학문적 발전 방향을 제시하라.

### 좋지 않은 Future Direction

- “We will recruit 5,000 additional samples from European cohorts”
- “Future studies will validate these findings using CRISPR”
- “We plan to investigate protein-protein interactions using IP-MS”

이런 문장들은 “우리가 이런 연구를 할 거예요”라고 선언한다. 너무 구체적이어서 경쟁자에게 아이디어를 주거나, 실행하지 못했을 때 비판받을 여지를 만든다.

### 발전적인 Future Direction

위 논문은 이렇게 제시한다:

“We suggest that future studies integrate intrafamilial deviations to account for familial background effects. WFSD may facilitate personalized variant interpretation or support genetic counseling frameworks. Investigations into the sources of familial influence, such as inherited variants, environmental factors, or their interplay, would deepen our understanding.”

“Future studies”라는 주어를 사용한다. “우리가”가 아니라 “학계가” 나아갈 방향이다. 구체적 방법론(CRISPR, IP-MS)을 제시하지 않고, 개념적 방향(integrate intrafamilial deviations, investigate sources of familial influence)을 제시한다.

또한 응용 가능성을 제시한다:

“Addressing these questions will advance our understanding of the neurobiological mechanisms underlying ASD variability and contribute to the development of precise support and interventions.”

“이런 연구가 왜 필요한가”에 대한 broader impact를 설명한다. 단순한 연구 계획이 아니라, 학문적·임상적 의의를 강조한다.

---

## Conclusion의 역할: Introduction과 대칭을 이루는 마무리

일부 저널은 Discussion과 별도로 Conclusion 섹션을 요구한다. 세 섹션(Introduction, Discussion, Conclusion)이 모두 있을 때, 각각의 역할은 무엇일까?

**Introduction:** 넓은 맥락 → 좁은 질문. “이 분야에는 이런 문제가 있고, 우리는 이 질문에 답하고자 한다.”

**Discussion:** 발견의 해석과 맥락화. “우리는 이런 것을 발견했고, 이는 기존 연구와 비교할 때 이런 의미를 갖는다. 한계는 이러하며, 미래 방향은 이렇다.”

**Conclusion:** 좁은 답 → 넓은 함의. “우리의 발견은 이것이고, 이는 더 큰 맥락에서 이런 영향을 미칠 것이다.”

Introduction과 Conclusion은 대칭을 이룬다. Introduction이 갈때기처럼 넓은 곳에서 좁은 곳으로 좁혀갔다면, Conclusion은 역갈때기처럼 구체적 발견에서 시작해 broader impact로 확장된다.

위 논문의 Conclusion을 보자:

“Accounting for familial background enables clinicians to predict the phenotypic outcomes of specific DNVs. This approach provides valuable insights into the biological mechanisms underlying ASD, enabling precise support and effective intervention strategies for individuals and families affected by ASD.”

핵심 메시지(familial background를 고려하면 DNV outcome을 예측할 수 있다)에서 시작해, 임상적 응용(precise support and intervention)으로 확장된다. 단순히 “우리는 WFSD를 개발했다”가 아니라, “이것이 환자와 가족에게 어떤 의미를 갖는가”로 마무리한다.

좋은 Conclusion은 독자에게 “So what?”에 대한 답을 준다. “당신이 발견한 것이 왜 중요한가? 이것이 세상을 어떻게 바꿀 것인가?” 이 질문에 답하며 논문을 마무리할 때, 독자는 연구의 의미를 온전히 이해하게 된다.

---

## Discussion 작성의 실전 전략

### 1. Introduction과 대조표 만들기

Discussion을 쓰기 전에, Introduction을 다시 읽으며 “던진 떡밥” 목록을 만들어라.

- Introduction에서 언급한 연구 gap은 무엇인가?
- 우리가 답하고자 한 specific question은 무엇인가?
- 기존 연구의 한계로 지적한 것은 무엇인가?

각 항목에 대해 Discussion에서 어떻게 답할지 메모하라. 모든 떡밥이 회수되었는지 확인하는 체크리스트가 된다.

### 2. 결과의 위계를 정하기

Results의 모든 결과가 같은 중요도를 갖지 않는다.

- **Primary findings:** Introduction의 main question에 직접 답하는 핵심 결과
- **Secondary findings:** 흥미롭지만 부수적인 관찰
- **Negative findings:** 유의하지 않았지만 의미 있는 결과

Discussion은 주로 primary findings에 집중한다. Secondary findings는 간단히 언급하거나 Supplementary Discussion으로 옮길 수 있다. Negative findings는 중요한 것만 선별적으로 논의한다.

### 3. 문단 단위로 주제 관리하기

각 문단은 하나의 명확한 주제를 다뤄야 한다.

**나쁜 예:** > “우리는 WFSD를 개발했다. 18개 유전자를 발견했다. PTEN이 흥미롭다. Sample size가 작았다. 임상적으로 중요하다.”

하나의 문단에 너무 많은 주제가 섞여 있다.

**좋은 예:** > **문단 1:** WFSD 방법론의 우수성과 검증 > **문단 2:** Domain-specific effect의 생물학적 메커니즘 (PTEN 예시) > **문단 3:** Position-specific effect의 의미 (NCKAP1, CHD2 예시) > **문단 4:** Limitation과 그에 대한 대응 > **문단 5:** 임상적·생물학적 함의

각 문단이 하나의 주제에 집중할 때, 논리적 흐름이 자연스럽게 만들어진다.

### 4. “Therefore”, “However”, “Moreover”를 전략적으로 사용하기

전환 단어는 문단 간 관계를 명확히 한다.

- **Therefore/Thus:** 인과 관계. “이런 결과가 나왔다, 따라서 이런 의미를 갖는다”
- **However/Nevertheless:** 대조. “이런 장점이 있다, 그러나 이런 한계도 있다”
- **Moreover/Furthermore:** 추가. “이것도 중요하다, 게다가 저것도 중요하다”

이런 전환어가 없으면 문단들이 단순 나열처럼 보인다. 전략적으로 사용하면 논리적 구조가 명확해진다.

### 5. 초안에서 과감하게 자르기

첫 초안은 보통 너무 길다. 모든 결과를 설명하려 하고, 모든 문헌을 인용하려 한다.

두 번째 draft를 쓸 때 자문하라: – “이 문단이 main message에 기여하는가?” – “이 문헌 인용이 꼭 필요한가, 아니면 습관적으로 넣은 것인가?” – “이 설명이 독자의 이해를 돕는가, 아니면 혼란을 주는가?”

Discussion은 짧고 강렬할 때 더 효과적이다. 3-4 페이지로 핵심 메시지를 전달하는 것이, 7-8 페이지로 모든 것을 설명하려는 것보다 낫다.

---

## 흔한 실수와 해결책

### 실수 1: Results의 반복

**나쁜 예:** > “Figure 2A shows that WFSD was 3.5 in dnPTV carriers. Figure 2B shows that it was 3.3 in dnMIS carriers. Figure 2C shows…”

이것은 Results의 반복이다. 독자는 이미 Results를 읽었다.

**좋은 예:** > “dnPTV carriers exhibited the largest deviation from their parents (3.5 WFSD), suggesting that protein-truncating variants have the most substantial impact on phenotypic divergence from familial baseline.”

같은 숫자를 언급하되, 해석(interpretation)을 더한다. “무엇을 발견했는가”가 아니라 “그것이 무엇을 의미하는가”에 집중한다.

### 실수 2: 과도한 speculation

**나쁜 예:** > “이 결과는 PTEN이 신경발달 과정에서 중추적 역할을 하며, 아마도 시냅스 형성을 조절하고, 수상돌기 가지치기에 영향을 주며, 신경전달물질 분비를 변화시킬 것이다.”

데이터가 뒷받침하지 않는 과도한 추론이다.

**좋은 예:** > “Functional assays demonstrated that p.Thr131Ile exhibits near-complete loss-of-function, consistent with the severe phenotypic impact we observed. This domain-specific effect aligns with the substrate-binding region’s known role in enzymatic activity.”

기존 문헌의 functional data와 자신의 phenotypic data를 연결한다. 추론이지만, 근거가 있는 추론이다.

### 실수 3: 방어적 태도

**나쁜 예:** > “우리 연구는 sample size가 작고, 다른 ethnicity를 포함하지 못했으며, longitudinal data가 없어서 인과관계를 확인할 수 없다. 또한…”

Limitation에서 너무 많은 것을 나열하며 방어적으로 변한다.

**좋은 예:** > “While our approach provides more accurate effect estimates, the substantially smaller subset with complete family phenotyping limits detection of rare variant effects. Future studies collecting standardized family-based assessments will enhance power for such analyses.”

한계를 인정하되, 그것을 미래 연구의 방향으로 전환한다. 방어가 아닌 건설적 제안이다.

---

## Discussion과 Conclusion의 조화

논문의 끝은 시작과 호응해야 한다. Introduction이 “우리는 이 질문에 답하고자 한다”로 시작했다면, Conclusion은 “우리는 이렇게 답했다”로 마무리한다.

효과적인 Conclusion의 구조:

1. **핵심 발견 재진술** (1-2문장): “We demonstrated that…”
2. **방법론적 기여** (1-2문장): “This approach enabled…”
3. **생물학적/임상적 함의** (2-3문장): “These findings provide…”
4. **미래 전망** (1-2문장): “Future integration of…”

전체를 5-6문장, 한 문단으로 간결하게 마무리한다. 새로운 정보를 추가하지 않고, 이미 논의한 내용을 bigger picture의 관점에서 재구성한다. 독자는 논문을 덮으며 생각한다. “이 연구는 무엇을 발견했는가? 왜 중요한가?” Conclusion은 이 두 질문에 대한 명확하고 기억에 남는 답을 제공해야 한다.

## Discussion은 당신의 목소리다

Results는 객관적이고 중립적이어야 한다. 하지만 Discussion은 다르다. 여기서는 당신의 해석, 당신의 통찰, 당신의 비전을 보여줄 수 있다. 좋은 Discussion을 쓰는 과학자는 단순히 데이터를 설명하는 것을 넘어선다. 그들은 데이터 뒤에 숨겨진 생물학적 의미를 보고, 그것을 독자가 이해할 수 있는 narrative로 만든다. Introduction에서 던진 질문에 답하되, 동시에 새로운 질문을 열어준다. 당신의 연구가 완벽할 필요는 없다. 모든 질문에 답할 필요도 없다. 하지만 당신이 답한 질문이 왜 중요한지, 그것이 과학을 어떻게 한 걸음 앞으로 나아가게 하는지 설명할 수 있어야 한다. Discussion을 쓸 때, 스스로에게 물어라: “1년 후 이 논문을 읽는 독자가 기억할 메시지는 무엇인가?” 그 하나의 메시지를 중심으로 모든 문단을 구성하라. 나머지는 과감히 버려라.

당신의 데이터는 말한다. 이제 당신이 말할 차례다.

## 10장. 당신의 첫번째 논문을 위하여

이 교재를 쓰기 전, 학생들이 여러가지 질문과 피드백을 주었다. 대부분을 챕터에 넣었는데, 몇가지 질문들이 남았다. 답을 남기기 위해, 이 장을 할애한다.

### 방향성의 설정

학생의 질문:

여러 논문을 작성할 때, 졸업 논문의 큰 방향성은 언제, 어떻게 설정해야 하나요?

단순한 질문이지만, 학위를 하는 학생에게 가장 중요한 질문이다. 대학원생은 두가지 논문을 적는다. 대학원 학위 졸업 논문과 연구 프로젝트에서 창발한 SCI 저널에 투고할 논문. 졸업 요건을 떠나서, 이공계 대학원생에게는 저널의 논문은 연구 커리어에 중요하다. 남궁석 선생님이 쓴 “과학자가 되는 방법 (이길 출판)”을 매년 신입생에게 선물한다. 책에 보면, 대학원생은 흡사 아이돌 가수 연습생과 같다고 비유한다. 나는 이 비유를 학생들에게 자주 전달한다. 연구실은 소속사와 같아서, 노래와 안무를 연습시켜주고, 아이돌이 되기 위한 연습생이 발표할 노래와 앨범 작업을 도와준다. 연구자의 논문은 가수의 앨범과도 같아서, 자신이 부르고 싶은 노래를 만들듯 연구하고 논문을 작성한다. 첫번째 싱글이 나오는 것처럼 첫번째 논문을 쓴다.

아직 데뷔하지 않은 가수이기에, 얼마나 노래를 잘할지, 춤을 잘 출지는 모른다. 그러나 그 연습생은 하고 싶은 음악이 있다. 대학원생도 마찬가지다. JYP엔터테인먼트 소속의 연습생이 버블검 팝 장르를 하고 싶기에 거기서 연습을 하고, 서정적인 멜로디를 하고 싶으면, 안테나에 들어가야 한다. 그리고 소속사의 음악과 작품을 공유하며 성장한다. 이공계 대학원에 진학한 학생은 그 연구실이 하던 오랜 기간의 생각과 작업들을 이어 받아, 자신의 새로운 영역을 창출한다. 일면 수동적이기도 하지만, 본질적으로 본인이 하고 싶은 연구를 찾고 해야한다. 대학원에 처음 들어온 학생들은 “A는 A다”라는 답을 맞추는 교육을 받다가, 어느 순간 세상에 “B를 발견했습니다”라고 말해야 하는 무대에 오른다. 이것은 과학적으로도 어렵지만, 그것을 전달하는 방향을 바꾸는 본질적 도전이기도 하다.

석사생과 박사생의 접근 방식이 다르다.

석사과정의 대학원생은 약 18개월 동안 논문을 작성할 시간을 갖는다. 그래서 1저자로 논문을 제출하고 심사 받기에 시간이 많이 부족하다. 우리 연구실은 입학전에 주제를 선정하기를 석사 학생들에게 권유한다. 이렇게 하면, 연구실에서 제공하는 데이터와 연구 방향을 충분히 공유할 수 있다.

박사과정의 대학원생은 약 3년동안 학위를 하게 되는데 (통합의 경우 5년), 이 시기동안 최소 2편의 논문을 작성할 수 있다. 보통 정부 연구과제가 3년의 연구기간을 갖고 있기에, 연구과제와 학위 연구가 잘 맞물려 갈 수 있다. 이 과정동안 어떠한 주제를 발전시켜서, 본인이 떠날 먼 연구의 여정의 종자돈으로 마련할지 고민해야 한다. 나는 이것이 큰 방향성을 결정하는데 중요한 부분이라 생각하고, 학생과 지도교수는 상시 고민하고 논의해야한다.

### 논문 작성의 타임라인

학생의 질문:

분석 단계, 작성 초반·중반·후반 등 단계별로 적절한 타임라인은 어떻게 잡는 게 좋을까요?

논문 작성에는 리듬이 있다. 처음 연구를 시작할 때는 막막하지만, 데이터를 만지고 분석하다 보면 어느 순간 “이야기가 보이는” 순간이 온다. 나는 학생들에게 피겨 2개가 나올 때까지는 분석에 집중하라고 조언한다.

피겨 2개라는 기준은 자의적이지만, 실용적이다. 보통 첫 번째 피겨는 메인 결과를 담는다. 두 번째 피겨는 그 결과를 뒷받침하거나 확장하는 내용이 된다. 이 두 개가 나오면, 연구의 뼈대가 서는 것이다. 그제야 “작성을 시작하지요”라고 말할 수 있다. 이 시점에 이르면, 학생들은 관성적으로 준비해야 할 작업들이 보이기 시작한다. 어떤 추가 분석이 필요한지, 어떤 레퍼런스를 더 찾아야 하는지 감이 온다.

분석 단계에서는 지도교수나 공동연구자들과 가능한 주기적인 미팅이 중요하다. 매주 혹은 격주로 만나서 진행 상황을 공유하고, 방향을 조율한다. 혼자 분석하다 보면 터널에 갇힐 수 있는데, 정기적인 미팅이 그것을 방지해준다.

중반 단계에서는 논문의 퀄리티를 높이는 데 집중한다. 그리고 공동연구자들의 파트를 담는 것에도 시간을 할애해야 한다. 특히 우리가 못하는 실험이나 추가 데이터 분석을 첨부하는 것이 이득이 있을지 생각하여 시간을 잘 조율하는 게 필요하다.

후반 단계는 다듬기의 시간이다. 글을 매끄럽게 하고, 피겨를 다시 보고, 레퍼런스를 체크한다. 이 과정에서 동료들의 피드백이 중요한데, 이것은 뒤에서 다시 이야기하겠다.

---

## 에디터와 리뷰어가 보는 것

논문을 투고하면, 에디터와 리뷰어라는 두 관문을 통과해야 한다. 이 둘은 서로 다른 것을 본다.

에디터는 우선 논문의 novelty를 본다. 이 연구가 해당 저널의 독자들에게 새로운 무엇을 줄 수 있는가? 저널의 스코프에 맞는가? 충분히 흥미로운가? 에디터는 보통 리뷰를 보낼 가치가 있는지를 빠르게 판단한다. 그래서 투고 후 며칠 내에 desk rejection을 받는 경우도 있다. 에디터는 수많은 논문을 처리해야 하기에, 첫인상이 중요하다. Abstract와 Introduction, 그리고 Main Figure들이 명확하고 설득력 있어야 한다.

리뷰어는 더 깊이 본다. 연구의 타당성, 방법론의 적절성, 결과의 해석, 그리고 한계점까지 꼼꼼히 본다. 리뷰어는 해당 분야의 전문가이기에, 작은 실수나 논리적 비약도 놓치지 않는다. 그래서 리뷰어 코멘트는 때로 날카롭고, 때로 건설적이며, 때로는(솔직히) 좀 억울하게 느껴지기도 한다.

두 관문 모두 novelty를 중요하게 본다는 점은 공통적이다. “이미 알려진 것”을 반복하는 연구는 통과하기 어렵다. 하지만 novelty가 꼭 혁명적인 필요는 없다. 기존 방법을 새로운 데이터에 적용하거나, 새로운 각도에서 접근하거나, 여러 조각을 잘 엮어서 새로운 insight를 주는 것도 충분히 의미 있다.

---

## 추천 리뷰어의 선정

학생의 질문:

추천 리뷰어는 어떤 기준으로 선정하는 것이 좋을까요?

논문을 작성하다 보면, 자주 인용하는 논문들이 나온다. 가만히 보면, 공통적인 연구자들이 보일 것이다. 그리고 당시에 1저자로 썼으나, 현재는 박사가 되었거나 독립연구자(교수)가 된 사람들도 있을 것이다. 그런 사람들의 리스트를 5-7명 정도 확보하는 게 좋다.

추천 리뷰어를 선정할 때는, 이 사람이 내 연구를 어느 측면에서 (재밌게) 읽어줄지 생각해보는 것이 중요하다. 예를 들어, 우리가 사용한 방법론을 개발한 사람, 우리가 연구한 질환이나 현상에 대해 오래 연구한 사람, 혹은 우리와 비슷한 데이터를 다룬 적이 있는 사람 등을 고려할 수 있다.

단, 너무 가까운 사람(같은 기관, 같은 연구 네트워크)은 피하는 게 좋다. 에디터가 conflict of interest를 우려할 수 있기 때문이다. 또한, 너무 유명한 big name을 추천하는 것도 양날의 검이다. 그들은 바쁘기에 리뷰를 안 할 수도 있고, 하더라도 매우 까다로울 수 있다.

## 동료의 피드백

연구실의 동료들에게 최소 2명 정도에게 작성 중인 원고(최종 버전이 아니어도 됨)를 보여주는 게 중요하다. 글은 결국 남에게 읽히면서 가치를 평가받는데, 내가 아닌 다른 사람이 내 글을 읽을 때 어떤 것에 집중하고 어떻게 받아들이는지를 확인해보는 게 중요하다.

나와 가까운 연구주제를 하는 학생 1명, 그리고 먼 주제를 하는 학생 1명, 혹은 연차를 다양하게 선정하여 동료들에게 보여주는 게 좋다. 가까운 주제를 하는 동료는 기술적인 디테일을 잡아줄 것이고, 먼 주제를 하는 동료는 큰 그림과 전달력을 피드백해줄 것이다. 혹은 같은 학과의 동기에게 보여주는 것도 방법일 수 있다.

그러기 위해선 평소에 서로 밥도 자주 먹고, 사우나도 다니고 하며 우정을 키워야 한다. 연구실 문화가 경쟁적이기만 하면, 이런 피드백을 주고받기 어렵다. 서로 돕는 문화를 만드는 것은 지도교수의 몫이기도 하지만, 학생들 스스로도 노력해야 한다. 내가 먼저 누군가의 초고를 읽어주면, 나중에 내가 도움이 필요할 때 그 사람이 기꺼이 읽어줄 것이다.

동료의 피드백을 받을 때는 방어적이 되지 말자. “이 부분이 이해가 안 돼요”라는 말을 들으면, 일단은 “아, 정말요?”라고 받아들이자. 그 사람이 이해 못 했다면, 리뷰어도 이해 못 할 가능성이 크다. 물론 모든 피드백을 다 받아들이는 필요는 없다. 하지만 최소한 경청하고, 왜 그런 피드백이 나왔는지 생각해보는 것은 가치가 있다.

## 부록 1. Daily Paper Summary Program — 요약하고 읽기

대학원 입학의 목표로 하는 학부연구생들을 위한 논문 읽기 습관 들이기 프로그램을 적는다. 대부분의 학생은 논문을 처음부터 완벽히 이해해야 한다고 생각한다. 하지만 그건 익숙해지는데까지 시간이 오래 걸린다. 논문을 이해하기 위해선, “논문의 구조”도 알아야하고, 해당 분야의 논문들에서 자주 쓰는 용어 (aka 과학 사투리)도 알아야 한다. 그리고 논문이 쓰여진 이전 연구들의 맥락도 이해해야 한다.

이런 이유 때문에 학생들이 처음 논문을 읽을때 모든 문장을 이해하려 하겠다고 생각하고 논문을 읽기 시작하면, 두세 페이지를 넘기기도 어렵다. 그래서 논문을 읽고 요약하는 형식이 아니라, 요약하고 읽는 형식을 추천한다. 우선 논문과 친숙해지기 위하여, 그리고 자신이 찾는 토픽이 어떠한 논문들에 등장하는지를 파악하는 것이다.

이번 장에서는 100일 동안 매일 100개의 논문을 요약하는 것을 목표로 진행하는 튜토리얼을 적는다.

### 1. 논문을 쇼핑하기

먼저 논문을 쇼핑하자. PubMed나 Google Scholar에 들어가서 읽고 싶은 키워드를 넣고 검색한다. 나는 Web of Science (WoS)를 학생들에게 추천한다. 학교 와이파이 혹은 학교 도서관 사이트를 통해 들어가면 접속할 수 있다. WoS에서는 보다 구조적인 검색이 가능하다. 키워드를 넣고 검색하면, 논문이 어디에 출간되었는지, 몇번 인용 (다른 연구에 의해 언급된 것)되었는지 확인할 수 있다. 특히 좋은 저널에 출간된 연구 논문을 선택해야 한다. 가능하면 아래 저널에 출간된 논문을 선택하자.

Nature, Nature Genetics, Nature Neuroscience,  
Nature Medicine, Nature Communications, Nature Machine Intelligence  
Cell, Neuron, Cell Genomics, Cell Reports Medicine  
Science, Science Translational Medicine, Science Advances  
Genome Biology, Genome Medicine, American Journal of Human Genetics  
JAMA, Lancet  
Molecular Psychiatry, Biological Psychiatry  
Cancer Discovery, Cancer Research

검색 결과로부터 논문을 선택하자. 다음과 같은 기준을 고려하여 선택한다.

1. 제목이 흥미로워 보이는 논문 (이거 읽고 싶은데?)
2. 최근 2년 이내에 출간된 논문 (2023, 2024, 2025년)
3. 위에 언급된 저널에 출간된 논문. Research Article 혹은 Resources 논문을 선택하자. Review나 Perspective를 선택하지 않는다.
4. 인용이 많이 된 논문 (단, 아주 최근에 출간된 논문은 인용수가 0일 수 있다)

5. (선택사항) 50일 쯤이 넘어가다보면, 자주 다운받은 저자들을 찾게 될 것이다. 저자 중에 “Joon-Yong An”가 있다면, “음 내가 최근 수집한 논문들 중, 이 저자의 논문을 10개나 모았구나.”. 그렇다면, 이런 고급진 저자는 믿고 이 사람의 논문을 선택하여 다운로드 하면 된다.

논문 링크에 들어가서 PDF 파일을 다운받는다. PDF 파일을 그대로 AI 툴 (ChatGPT, Gemini, Claude, NotebookLM 등)에 업로드한다. 즉, **읽기 전에 AI에게 먼저 요약**을 시킨다. 이렇게 100일동안 100편은 논문의 논문을 완벽히 이해하려고 하지 않고, 이 과정을 반복하며 습관을 들이는 것이다. **핵심은 '논문 구조와 언어에 익숙해지는 것'**이다. 요약은 읽기 전에 문턱을 낮추는 역할을 한다. 미리 길을 밝혀주면, 학생은 훨씬 가벼운 마음으로 본문을 볼 수 있다.

---

## 2. 검색 키워드 정하기

1번의 논문 검색시, 사용할 키워드가 막막하게 느껴질 수 있다. “나는 자폐 유전체가 관심이 있는데..”라고 하면, “Autism Genetics” 혹은 “Autism Genomics”로 검색하면 된다. 여기에 키워드를 보완해서 “Autism AND whole genome sequencing”으로 검색해도 된다.

이렇게 검색을 해도 검색 결과가 모호할 수 있다. Genetics라는 것은 단지 유전체 연구 뿐 아니라, 동물 모델이나 이미징 연구들까지 포함할 수도 있기 때문이다. 그래서 처음 10-20일 정도는 검색 결과에서 적절한 논문을 찾는게 어려울 수 있다. 이 과정에서 몇가지 요령을 터득해야하는데, 관련 키워드를 넓히는 방식이다. 가령 “Autism Genetics” 라고 처음 검색하다가, 몇가지 유전자들을 찾아서, 해당 유전자를 기반으로 검색어를 입력할 수도 있고, 단일세포 전사체나 AI와 같은 키워드를 넣어서 검색할 수도 있다. 혹은 완전히 새로운 키워드로 주제를 검색하여 논문을 요약하다가, 이전의 주제와 결합하여 생각해볼수도 있다. 그래서 가능한 검색 키워드는 넓고 자유롭게 접근하는게 좋다.

혹은 이 부분에서 너무 막연함을 느낀다면 (특히 MBTI가 ST, SF 학생들)은 지도교수와 상의하여, 검색 키워드를 조언 받는게 좋다.

---

## 3. AI에 사용할 프롬프트

아래 프롬프트를 그대로 복사해 사용한다.

```
Figure .
7 .
, .
:
1. (Background):
2. (Research Question):
3. (Data / Sample):
4. 1 (Result & Implication 1):
5. 2 (Result & Implication 2):
6. 3 (Result & Implication 3):
7. 4 (Result & Implication 4):
```

```
Figure :
"""
( PDF )
"""
```

AI의 결과는 논문을 읽기 전에 훑는 **사전 개요(pre-map)**로 사용한다. 처음에는 요약만 보고 논문을 담아도 좋다. 지금은 논문을 수집하고 빠르게 훑는 것만 연습한다.

(100일 후) 당신은 이보다 더 깊게 논문을 읽을 수 있다. AI로 요약하는 것 뿐 아니라, 논문의 구조와 주제를 연결하며 읽게 된다. 그리고 어느정도의 논문 읽기 근육이 생기면, 그때부터 정독을 하며 읽어야 한다.

---

## 4. 나의 생각 / 의문 (My Thought)

SI의 요약과 보고 난 뒤, 8번째 줄에 자신의 생각을 한 줄 남긴다. “왜 이런 접근을 썼을까?”, “이 결과가 의미하는 게 뭘까?” 같은 간단한 질문이면 충분하다. 이것은 정확하지 않아도 된다. 적는 것에 의미가 있다. 날이 지날 수록 더 나은 생각을 형성하게 될 것이다. 중요한 건 정확함이 아니라 ‘나도 생각할 수 있다’는 경험이다.

8.        /    : (                    )

처음엔 질문이 알을 수도 있다. 낱것의 생각일 수도 있다. 그러나 정말로 괜찮다. 하루 한 줄씩 적다 보면, 어느 날부터는 질문의 질이 달라진다. 그리고 그간 수집한 연구들 간의 연결점이 생각날 수도 있다. 그게 바로 **연구자로서 사고의 첫 신호**, 당신의 눈에 비늘이 벗겨지는 순간이다.

---

## 5. Notion에 기록하기

요약과 생각을 Notion에 기록한다. 논문을 100편 쌓는 것이 목표다. 한 편에 30분, 하루 한 편이면 3개월, 주 5회면 약 20주면 된다.

이 시기의 목표는 “깊이 이해”가 아니라 “익숙해지는 것”이다.

날짜	저널	논문 제목	링크	주요 키워드	요약(1-9번)	나의 생각(10번)
11/02	Nature Genetics	Rare coding variants shape synaptic regulation in autism	[링크]	autism, synapse, rare variant	(GPT 요약 내용)	“결과와 발달 시기의 관계를 더 찾아보고 싶다.”

---

## 6. 시간의 분배: 절대로 30분을 넘기지 말것, 차라리 딸깍만 해도 된다

GPT 요약을 본 뒤, 논문을 다시 열어보자. 이번에는 초록과 Figure를 천천히 읽는다. 이제 논문은 낯선 문서가 아니라, **이미 개요를 알고 있는 이야기**가 된다. 이 작은 차이가 ‘읽기’의 부담을 완전히 바꾼다. 하지만 100편을 모을때까지는 이 과정을 30분 이상 쓰지 않는다. 도리어, PDF 만 올리고 요약만 받아도 된다 (이러면 5분 이내로 할 수 있다). 대신 중요한 것은 **매일 반복해서 하는 것**이다.

나는 학생들에게 논문 읽기를 가르칠때 자주 언급하는 말이 있다. “초록과 Figure만” 읽어도 이해가 되는 논문이 있다. 논문을 구조적으로 잘 쓴 논문이다. 반대로 이해가 안되는 논문이 있다. 아무리 Nature, Science에 출간된 논문이어도 그렇다. 그래서 이와 같이 읽었을때, “아 잘 안읽힌다. 나는 못하네..”라고 생각하지 않아도 된다. 만약 당신의 실력이 부족한 것이라면, 100일의 연습이 지나고 다시 판단해도 된다. 처음부터 잘 안읽히는 몇가지 이유에 대해 너무 염려 할 필요가 없다.

---

## 7. 마무리

이 프로그램의 목적은 논문을 완벽히 읽는 것이 아니다. **논문을 두려워하지 않게 만드는 것, 매일의 습관을 만드는 것**이다. 처음 100편은 “요약하고 읽는” 시기다. 그 100편이 지나면, 그다음부터는 자연스럽게 “읽고 요약하는” 시기가 온다.

매일 30분, 가볍게 한 편. 이 루틴이 쌓이면, 당신은 어느새 논문 속에서 길을 잃지 않는 사람이 되어 있을 것이다.

## 부록 2: 논문을 처음 읽는 학생들을 위한 가이드

나는 2학년 1학기 수업에선 유전학을, 2학기에는 프로그래밍을 가르친다. 1학기 수업은 수업자료, 슬라이드를 위주로 진행된다, 그러나 2학기의 수업은 프로그래밍을 배우면서, 동시에 학생들은 본인의 포트폴리오를 작성한다. 포트폴리오는 출간된 논문을 선택해서, 해당 논문의 데이터를 활용해서, 분석하고 데이터를 시각화하는 것이다. 이 수업은 생물학과 학생들에게 제공되는데, 일생 처음으로 프로그래밍도 배우기도 하지만, 대부분의 학생들은 이 과정에서 처음으로 과학 논문을 읽어본다.

학부생들이 자주 하는 질문은 “교수님 학부생이 왜 논문을 읽어야 하나요?”이다. 나는 이에 대한 강한 관점과 의견을 갖고 있다. 과학은 매일 발전되며, 그 발전을 외우는게 아니라, 발전의 관점을 이해해야 한다는 것이다. 그래야 학생들은 과학의 주체자로 성장할 수 있다. 과학 논문은 학술지에 출간되며, 그 학술지는 전문지식을 다루지만, 본디 과학적 사실이 읽혀지기 위해 작성된다. 20세기보다 21세기, 그리고 시간이 지날 수록 과학 연구 논문은 점점 두꺼워지고, 어려워진다. 생물학이 발전하는 만큼 하나의 연구를 읽기위해 다양한 지식이 요구된다. 그럼에도 논문은 본디 몇가지의 지식을 전달하기 위해 작성되었으며, 주요 뼈대를 잘 찾아서 읽을 수 있다. 더욱이 요즘 시대에는 AI 기술을 이용하여 두꺼운 발견을 학부생들의 수준에서 읽을수 있다.

“교수님, 논문은 대학원에 가서나 읽는것 아닌가요?”라고 묻는 학생이 있다면, “학부 때 논문을 못읽으면, 대학원에 가도 논문을 읽을 수 없다”라고 답하곤 한다. 논문은 학부 때도 읽을수 있다. 그래서 내가 학부 2학년 수업에 사용하는 논문 읽는 법을 정리하여, 글을 적어본다.

### 논문의 종류

학술 논문은 크게 두 가지 유형으로 나뉜다: 연구 논문(research article)과 리뷰 논문(review article). 각각의 특성과 읽는 방법이 다르므로, 먼저 이 차이를 이해하는 것이 중요하다.

### 연구 논문 읽기

#### 연구 논문의 구조

연구 논문은 가설을 수립하고 실험을 수행한 뒤, 그 결과를 해석하여 발표하는 논문이다. 일반적으로 다음과 같은 구조를 따른다:

**Abstract (초록)** - 연구의 전체 내용을 압축한 요약본이다 - 넷플릭스나 영화 소개처럼, 해당 연구를 간단히 설명한다 - 서론, 본론(실험 및 결과), 결론을 각각 2-3문장으로 구성한다 - 일반적으로 영어 기준 200-300단어로 작성된다

**Introduction (서론)** - 연구 주제를 소개하고 배경 지식을 요약한다 - 연구의 필요성과 문제의식을 제시한다 - 경우에 따라 연구진의 사전 연구를 소개하기도 한다 - “우리는 이러한 연구를 수행하였다”는 식으로 마무리된다

**Materials & Methods (방법)** - 연구에 사용한 실험 방법을 상세히 기술한다 - 학술지마다 배치 위치가 다를 수 있다 (본문 중간 또는 참고문헌 뒤) - 내용이 길 경우 Supplementary Material(보충자료)에 별도로 제공된다

**Results (결과)** - 실험을 통해 관찰한 결과를 제시한다 - 간략한 해석을 포함하기도 한다 - “데이터가 연구자에게 말하고자 하는 것”에 해당한다

**Discussion (토의/결론)** - 결과에 대한 심층적인 해석을 제공한다 - “연구자가 데이터를 해석하는 방식”을 보여준다 - 연구자의 주관적 견해와 기존 연구와의 비교가 포함된다

**References (참고문헌)** - 논문 작성 시 인용한 다른 연구들의 목록이다

#### 효율적인 연구 논문 읽기 전략

##### 1단계: 전체 윤곽 파악하기

논문을 처음 접할 때는 다음 순서로 읽는다:

1. **제목 읽기:** 연구의 핵심 주제를 파악한다
2. **초록(Abstract) 읽기:** 연구의 전체 흐름을 이해한다
3. **Results 섹션의 소제목들 읽기:** 보통 5-7개의 섹션이 있으며, 각 섹션 제목은 넷플릭스 시리즈의 에피소드 제목처럼 연구의 흐름을 보여준다

여기까지 읽고 나면, 이 논문이 무엇을 말하려는지 상상해본다. 마치 영화나 드라마를 보기 전에 소개문을 읽으며 기대하는 것과 같은 과정이다. 물론 이런 방법을 모든 연구자들이 추천하거나 동의하지 않을 수 있다. 이 글을 적는 나는 극 N이기에, 보통 이렇게 하고 상상을 하며 기대한다.

## 2단계: 시각 자료 중심으로 읽기

텍스트를 줄줄이 읽기보다는, **Figure(그림)와 Table(표)를 중심으로 읽는 것**이 훨씬 효율적이다. 각 그림과 표는 연구의 핵심 데이터를 압축해서 보여준다.

각 Figure나 Table을 보면서: - 이 데이터가 무엇을 보여주는지 이해한다 - 2-3문장으로 다른 사람에게 설명할 수 있을 정도로 정리한다 - **반드시 본인의 언어로 정리한다** (논문의 문장을 그대로 베끼지 말고)

이 부분에서 AI 툴의 도움을 청할 수도 있다. 보통 논문을 처음 읽는 학생들은 논문의 PDF를 챗지피티에 올려두고 “요약해달라”고 할텐데, 그렇게 하고 나서 각 Figure를 하나씩 올리면서 각 부분을 설명해달라고 하자. 만약 AI의 설명이 너무 어렵다면, “대학교 1학년 학생의 수준에서 Figure 1A를 설명해줘”라고 구체적으로 요청을 해보자. 보통 이 정도의 요구를 하면 여러분이 이해할 수 있는 수준으로 설명을 해줄 것이다.

## 3단계: 선택적 읽기

**Methods는 처음엔 가볍게** - 논문 읽기 초보자는 Methods를 처음부터 세세하게 읽을 필요가 없다 - 같은 분야의 논문을 여러 편 읽다 보면, 공통적으로 사용되는 방법들이 보이기 시작한다 - 예를 들어, 엑스 시퀀싱 논문을 여러 편 읽다 보면 “어떤 캡처 기술을 사용했는가?”, “시퀀싱 깊이(depth)는 얼마인가?” 같은 공통 패턴이 보인다 - 처음에 이해가 안 되는 부분은 메모해두고, 비슷한 논문들을 더 읽으면서 공통점과 차이점을 파악한다 - 많은 학부생들이 통채로 논문을 읽으려다가 길을 잃거나 흥미를 잃는 경우를 정말 많이 봤다. 두가지가 어렵다. 유전체 연구와 같은 분야는 굉장히 다양한 첨단 기술을 사용한다. 그래서 기술의 용어가 충분히 통용되지 않은채로 논문이 작성될 수 있다. 교과서적인 용어가 등장하지 않을 수 있다. 정말 최신 기술이니까. 그래서 처음엔 실험 방법이나 기술명 자체에 집착하며 읽을 필요는 없다. 비슷한 몇편의 논문들을 읽다보면, 주요 기술이 눈에 보인다.

**Introduction은 나중에 깊이 있게** - 논문 읽기 초보자에게 Introduction은 우선순위가 낮다 - Abstract를 읽고 바로 Figure/Table로 가는 것이 더 효율적이다 - Abstract를 읽으며 상상한 내용이 실제 데이터와 부합하는지 확인해본다 - 처음엔 잘 안 맞을 수 있다. 이는 여러번의 실력 부족일 수도 있고, 논문이 잘못 쓰인 것일 수도 있다 - 하지만 이 과정을 반복하다 보면 패턴이 보이기 시작한다

**논문 읽기가 익숙해지면 Introduction이 보물창고가 된다** - 잘 쓴 Introduction은 해당 주제에 대한 미니 리뷰 역할을 한다 - 그런 논문들을 모아서 글의 구성 방식, 문단 구조 등을 분석한다 - 나는 좋은 Introduction 문단들을 복사해서 모아두고, 필사하듯 반복해서 읽는다 - 마치 그 문장들이 내 것이 될 때까지 익히는 것이다

## 4단계: 연구자 추적하기

논문의 **교신저자(corresponding author, 보통 마지막 저자)**의 이름을 기억하고 추적해본다. 그 연구자가 해당 분야의 전문가라면, 비슷한 주제의 논문을 여러 편 발표했을 것이다. 마치 영화 배트맨 트릴로지를 순서대로 보듯, 한 연구자의 여러 논문을 시간 순서대로 읽으면 그 분야의 논의가 어떻게 발전해왔는지 볼 수 있다.

## 연구 논문의 종류

연구 논문에도 길이에 따른 구분이 있다: - **Full-length article**: 일반적인 연구 논문 - **Short communication / Brief report**: 간단한 연구 결과를 빠르게 발표하는 짧은 형태의 논문 - 학술지마다 명칭이 약간씩 다를 수 있다

## 리뷰 논문 읽기

### 리뷰 논문의 특징

리뷰 논문(review article, 또는 종설 논문)은 특정 연구 주제나 최신 동향을 종합적으로 정리한 논문이다. 한 명 또는 여러 명의 전문가가 수십 편에서 수백 편의 연구 논문을 읽고 분석한 결과물이다.

리뷰 논문의 큰 장점은 **참고문헌(References)이 해당 분야의 중요 논문 목록**이라는 점이다. 특히 Nature Reviews, Trends in ... 시리즈 같은 고품질 리뷰 저널은 각 참고문헌마다 한두 줄로 핵심을 요약해주어 매우 유용하다. 우리나라 학회에서 운영하는 Experimental & Molecular Medicine 이나 Molecules and Cells 와 같은 저널에도 좋은 리뷰 논문이 많이 올라온다. 이런 학술지에서 리뷰를 찾는것도 도움이 된다.

## 효율적인 리뷰 논문 읽기 전략

### 섹션 단위로 나눠 읽기

리뷰 논문을 한 번에 통째로 읽으려 하지 말자. 보통 100편에 가까운 연구를 소개하기 때문에 부담이 크다. 대신 이렇게 접근한다:

1. 하나의 섹션을 선택한다 (보통 3-4개 문단으로 구성)
2. 그 섹션에서 언급된 연구 논문을 하루에 한두 편씩 읽는다
3. 각 논문은 위에서 설명한 방식으로 읽는다: Abstract → Figures/Tables
4. 본인의 문장으로 다시 정리한다

정리할 때의 팁: - 완벽한 문장이 아니어도 된다 - 한글, 영어, 혹은 혼용해서 작성해도 된다 - 중요한 것은 매일 조금씩, 몇 개월간 꾸준히 하는 것이다 - 특정 주제에 대한 연구들의 관계, 공통점, 차이점이 보일 때까지 계속한다

### 복수의 리뷰 논문 비교하기

리뷰 논문도 저자의 관점과 의도가 반영된다. 따라서: - 같은 주제의 리뷰 논문을 2-3편 선택한다 - 비슷한 섹션을 비교하며 읽는다 - 각 저자가 어떤 연구를 강조하고, 어떤 관점을 취하는지 파악한다

### 리뷰 논문의 패턴 파악하기

여러 리뷰 논문을 읽다 보면 흥미로운 점을 발견하게 된다. 같은 저자가 쓴 여러 리뷰에서 구성 방식, 내용 전개 등이 상당히 유사한 경우가 많다. 이는 해당 연구 분야에서 주로 논의되는 핵심 쟁점들이 있다는 의미다.

이런 공통 패턴을 파악하며 읽다 보면: - 특정 주제에 대한 “나만의 관점”이 생기기 시작한다 - 여러 리뷰의 좋은 부분들을 종합하는 능력이 생긴다 - 나는 첫 리뷰 논문을 쓸 때, 우수한 리뷰 논문 몇 편을 선정해서 그 구조와 토픽을 참고하며 작성했다

## 일반적인 팁과 조언

### 1. 용어 정리

전문 용어 단어장을 만든다 - 생소한 용어를 만날 때마다 기록한다 - 각 용어의 정의와 맥락을 함께 적어둔다 - Nature Reviews 같은 저널은 Box 형태로 용어 설명을 제공하기도 한다

### 2. 꾸준함이 핵심

하루 30분, 6개월의 마법 - 처음에는 정말 어렵고 막막하다 - 하지만 하루 30분씩 꾸준히 6개월만 하면 실력이 눈에 띄게 향상된다 - 고단한 과정을 지나면, 여러 개념들이 연결되기 시작하고 재미를 느끼게 된다 - 지름길은 없다. 꾸준함만이 답이다

### 3. 반복 읽기의 힘

어려운 논문은 덮어두고 다시 읽는다 - 한 번에 이해하려 하지 않는다 - 다음 날 다시 읽어본다 - 며칠에 걸쳐 반복해서 들여다본다 - 갑자기 “아하!” 하는 순간이 온다

### 4. 완벽주의를 버린다

논문 읽기 초보자가 흔히 하는 실수는 모든 것을 완벽하게 이해하려는 것이다. 하지만: - 처음부터 100% 이해할 수 없다 - 대략적인 흐름을 파악하는 것만으로도 충분하다 - 비슷한 논문을 더 읽으면서 이해도를 높여간다 - “이해의 깊이”는 시간과 경험에 비례한다

## 실전 연습 가이드

### 초보자를 위한 4주 실천 계획

1주차: 리뷰 논문 한 편 선택 - 관심 있는 주제의 최신 리뷰 논문을 찾는다 - 하루 30분씩 한 섹션씩 읽는다 - 이해가 안 되어도 괜찮다. 전체 흐름만 파악한다

**2주차: 연구 논문 3편 읽기** - 리뷰 논문에서 언급된 연구 논문 중 3편을 선택한다 - 각 논문의 Abstract와 Figures를 중심으로 읽는다 - 각 Figure를 2-3문장으로 요약하는 연습을 한다

**3주차: 같은 저자의 논문 추적** - 흥미로운 연구 논문의 교신저자를 찾는다 - 그 저자의 다른 논문 2-3편을 찾아 읽는다 - 연구가 어떻게 발전해왔는지 시간 순으로 정리한다

**4주차: 비교하며 읽기** - 같은 주제의 연구 논문 2편을 선택한다 - 두 논문의 접근법, 결과, 결론을 비교한다 - 공통점과 차이점을 정리한다

## 중급자로 가는 길

논문 읽기가 어느 정도 익숙해지면: - Introduction을 자세히 읽기 시작한다 - Methods의 세부 사항을 비교하기 시작한다 - Discussion에서 저자의 해석과 한계점을 비판적으로 평가한다 - 본인만의 견해를 형성하기 시작한다

## 마무리하며

논문 읽기는 과학적 사고의 기본이다. 처음에는 외국어를 배우는 것처럼 낯설고 어렵지만, 올바른 방법으로 꾸준히 연습하면 누구나 능숙해질 수 있다.

기억하자: - 완벽을 추구하지 말고 꾸준함을 추구한다 - 전체를 한 번에 이해하려 하지 말고 핵심부터 파악한다 - 남의 말을 그대로 받아들이지 말고 본인의 언어로 재정리한다 - 하루 30분, 6개월만 투자하면 확실한 변화를 경험할 것이다

논문 읽기는 단순히 정보를 습득하는 것이 아니다. 과학적 사고방식을 배우고, 비판적 분석력을 키우며, 결국 본인만의 연구 질문을 만들어가는 여정이다. 이 가이드가 그 여정의 출발점이 되기를 바란다.

## 더 읽어볼 자료

논문 읽기에 대해 더 알고 싶다면 다음 자료들을 참고하자:

**영상 자료** - <https://youtu.be/joY4CdU1TMA> - <https://youtu.be/Kllo6B14cNE> - <https://youtu.be/clIP0GIqxaE>

**블로그 글** - 페이퍼를 빨리/효율적으로 읽는 방법을 알려주마.

이 자료들은 다양한 관점에서 논문 읽기 전략을 제시하므로, 본인에게 맞는 방법을 찾는 데 도움이 될 것이다.