

- METAGENOME PROJECT -

프로젝트 안내

건강한 사람과 크론병(Crohn's disease) 환자의 장내미생물 군집 구성에 차이가 있는지 알아보고, 크론병 치료제(신약) 복용 전후에 따른 장내미생물 군집의 변화가 나타나는지 알아보기 위한 연구입니다. 분석을 위한 샘플의 정보는 아래를 참고하시기 바랍니다.

[프로젝트 샘플 정보]

Treatment Group (크론병 환자)		Normal Group (정상인)	
Before (복용전)	After (복용후)	Before (복용전)	After (복용후)
T01B	T01A	N06B	N06A
T02B	T02A	N07B	N07A
T03B	T03A	N08B	N08A
T04B	T04A	N09B	N09A
T05B	T05A	N10B	N10A

> Target region : 16S rDNA V3-V4 (약 440~465bp)

> Download 경로 : <https://data.macrogen.com/~macro3/BioEdu/202008/Project.zip>

연구 목적

1. **건강한 사람과 크론병 환자의 장내미생물 군집 다양성의 차이**가 있는지 확인합니다.
2. **치료제(신약) 복용 전후에 따른 장내미생물 군집 다양성의 변화**가 있는지 확인합니다.
 - 건강한 사람에게서 복용 전후에 따른 변화가 있는가?
 - 크론병 환자에게서 복용 전후에 따른 변화가 있는가?

> 각 샘플의 미생물 군집 다양성을 확인하고, 위 샘플 정보와 연구 목적을 참고하여 연구 결과를 기술하세요

분석 준비 단계

1. QIIME에서 요구하는 mapping file(샘플 정보 파일)을 작성하세요.

* Barcode, Linker Primer 서열 정보를 제외하고, 샘플의 정보만으로 작성할 것!

> [mapping file](#)을 아래에 표로 작성하세요.

Sample ID	Sample type	Administration	Treatment	Day
T01A	Crohn	Yes	Experiment	20200825
T02A	Crohn	Yes	Experiment	20200825
T03A	Crohn	Yes	Experiment	20200825
T04A	Crohn	Yes	Experiment	20200825
T05A	Crohn	Yes	Experiment	20200825
T01B	Crohn	No	Control	20200825
T02B	Crohn	No	Control	20200825
T03B	Crohn	No	Control	20200825
T04B	Crohn	No	Control	20200825
T05B	Crohn	No	Control	20200825
N06A	Normal	Yes	Experiment	20200825
N07A	Normal	Yes	Experiment	20200825
N08A	Normal	Yes	Experiment	20200825
N09A	Normal	Yes	Experiment	20200825
N10A	Normal	Yes	Experiment	20200825
N06B	Normal	No	Control	20200825
N07B	Normal	No	Control	20200825
N08B	Normal	No	Control	20200825
N09B	Normal	No	Control	20200825
N10B	Normal	No	Control	20200825

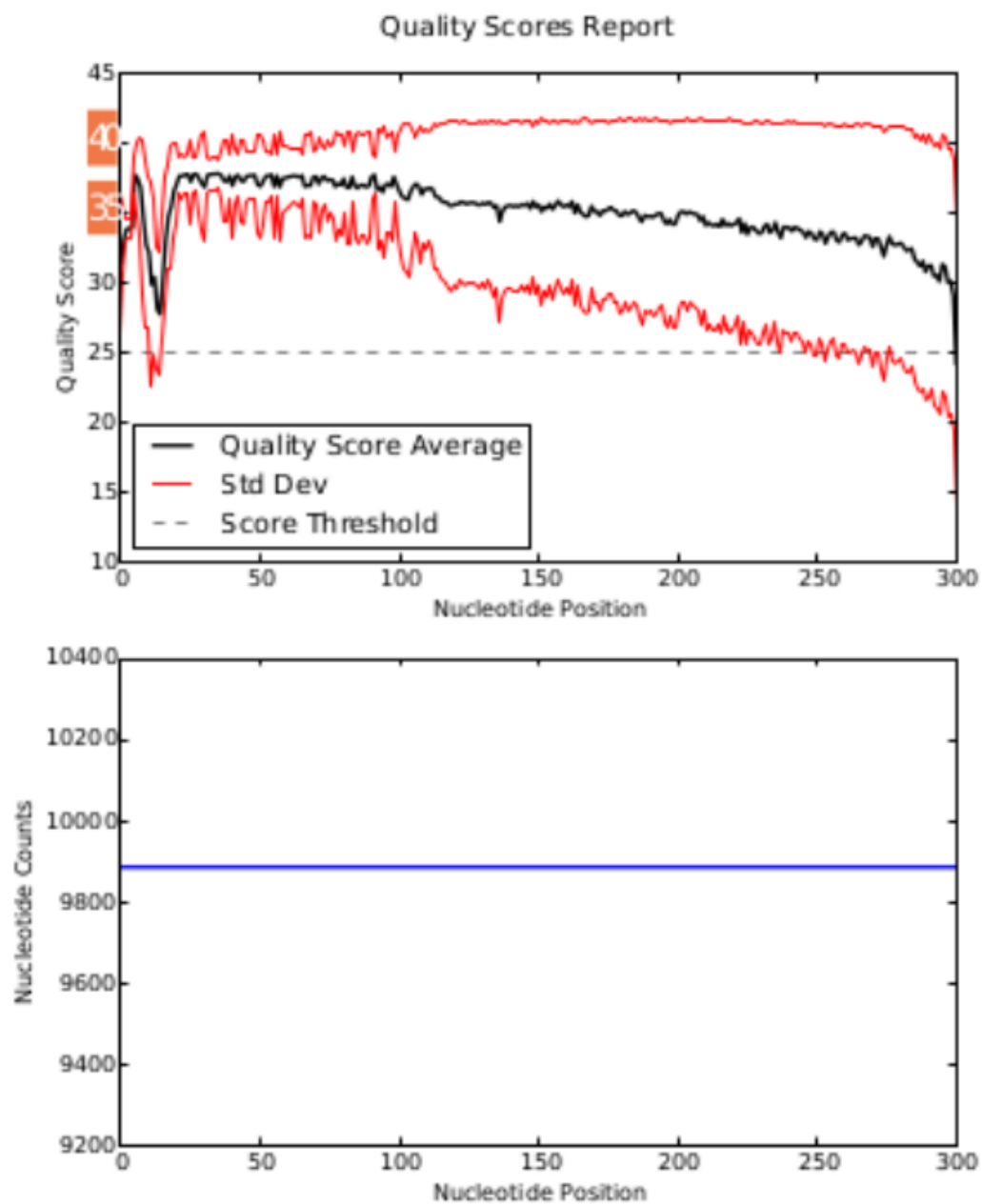
Preprocessing 단계

* [1.Raw]폴더 내 raw data 파일이 존재합니다.

2. T01B 샘플의 [quality score plot](#)을 그리고, [truncation position](#)을 기재하세요. (Read1, Read2 각각)

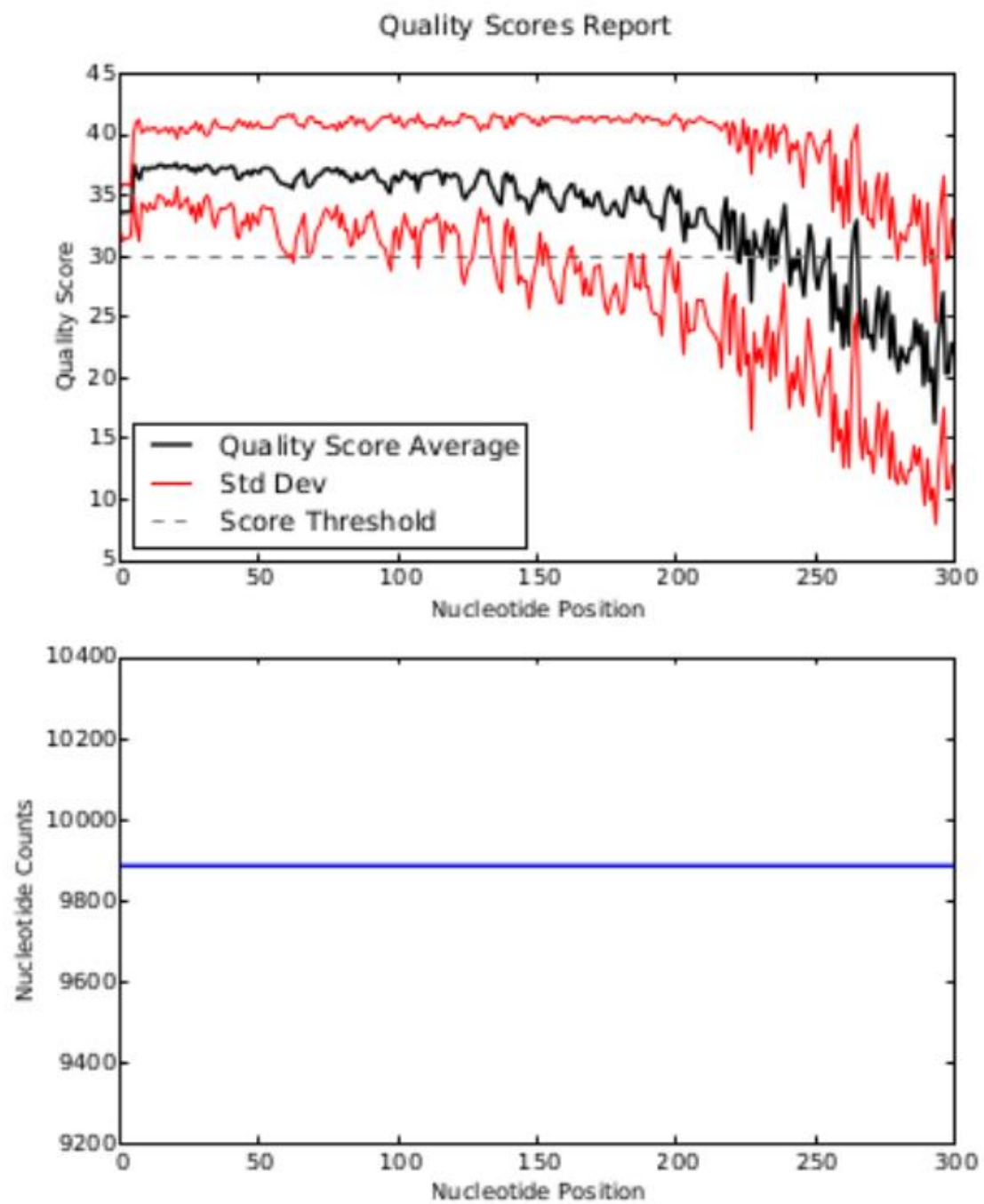
> Read1

- Suggested nucleotide truncation position : 300



> Read2

- Suggested nucleotide truncation position : 256

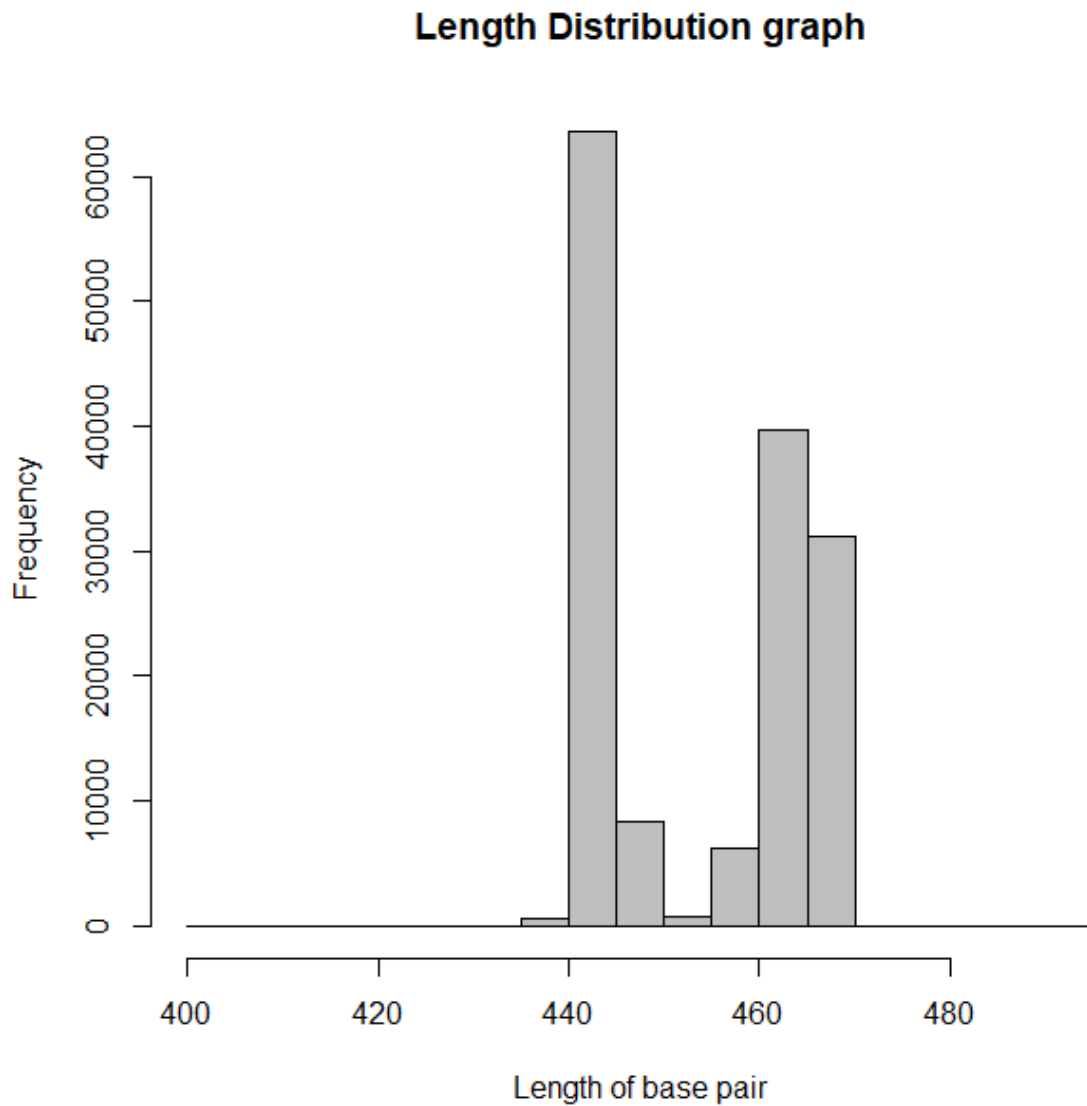


3. 모든 샘플에 대해 [read assembly](#)를 수행하고, 아래 [표에 STAT을 정리](#)하세요.

SampleID	Reads Count (Raw)	Reads count (Assembled)	Assembled Read 비율(%)	450bp미만의 Reads 개수	N base를 갖는 read(%)
T01A	9541	8150	85.42	3472	0.13%
T02A	9415	7992	84.89	3428	0.10%
T03A	8218	7204	87.66	3291	0.03%
T04A	8696	7503	86.28	3225	0.13%
T05A	9078	7829	86.24	3659	0.13%
T01B	9887	8456	85.53	4301	0.19%
T02B	9937	8369	84.22	3990	0.07%
T03B	8009	6881	85.916	3068	0.12%
T04B	9994	8505	85.10	3997	0.09%
T05B	9720	8453	86.97	4151	0.12%
N06A	8037	6824	84.91	3705	0.18%
N07A	9679	8049	83.16	3867	0.06 %
N08A	9138	7840	85.80	3687	0.05 %
N09A	8212	6878	83.76	3191	0.06%
N10A	8813	7776	88.23	4227	0.12%
N06B	9860	8414	85.33	4097	0.08 %
N07B	8153	7106	87.16	3571	0.07 %
N08B	8736	7648	87.55	3848	0.07%
N09B	8902	7563	84.96	3659	0.07 %
N10B	8815	7586	86.06	3664	0.12%

4. [2.Preprocessing] 폴더 내 'seqs.fna' 파일은 20개 샘플에 대해 read assembly 및 quality filter 등의 preprocessing 과정을 거친 결과 파일입니다.

'seqs.fna' 파일의 모든 reads에 대한, [length distribution graph](#)를 그리세요.



OTU picking 단계

5. OTU를 정의하는 두 가지 방법에 대한 차이점을 간략히 기술하세요.

Closed OTU picking	Denovo OTU picking
1.refrence 서열과 비교하는 방법. 2.새로운 otu를 찾기 힘들다. 3.시퀀싱 에러율이 낮다.	1. 가장 유사한 OTU끼리 서열비교(ref사용X). 2.새로운 otu를 찾을 수 있다. 3. reference에 의한 편향이 적다.

6. 'seqs.fna'에 대해, [Closed OTU picking 방식으로 OTU를 정의](#)하고, 총 OTU의 수와 OTU를 구성하는 전체 reads의 수를 기재하세요.

Gamma-diversity : 1552
Total read count : 63795

7. 앞서 생성한 biom 파일을 이용하여, QIIME에서 '[core_diversity_analysis.py](#)' command를 수행하고자 합니다. 이 때, [subsampling 기법의 normalization을 적용하려면, 어떤 옵션을 사용](#)해야 할까요? 옵션에 [적용할 값도 함께 기재](#)하세요.

-e 2731

Diversity analysis

** [3.DiversityAnalysis] 폴더 내 20샘플에 대한 core diversity 분석 결과가 존재합니다.

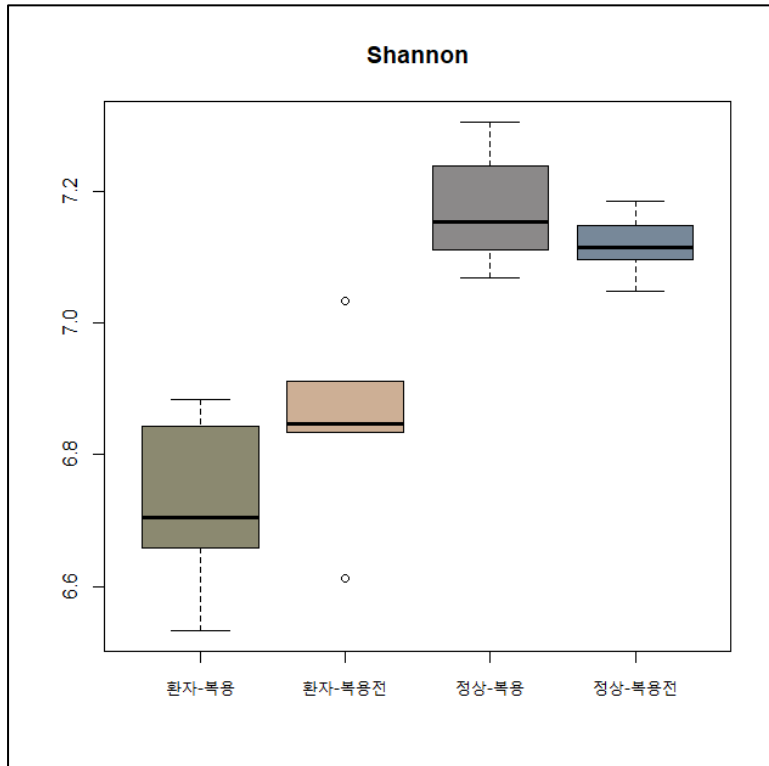
해당 결과물을 이용하여, 아래 문제에 답하세요.

8. 'otus_table.biom'파일을 이용하여, [alpha-diversity 분석을 수행한 결과](#)를 아래 표에 기재하세요.

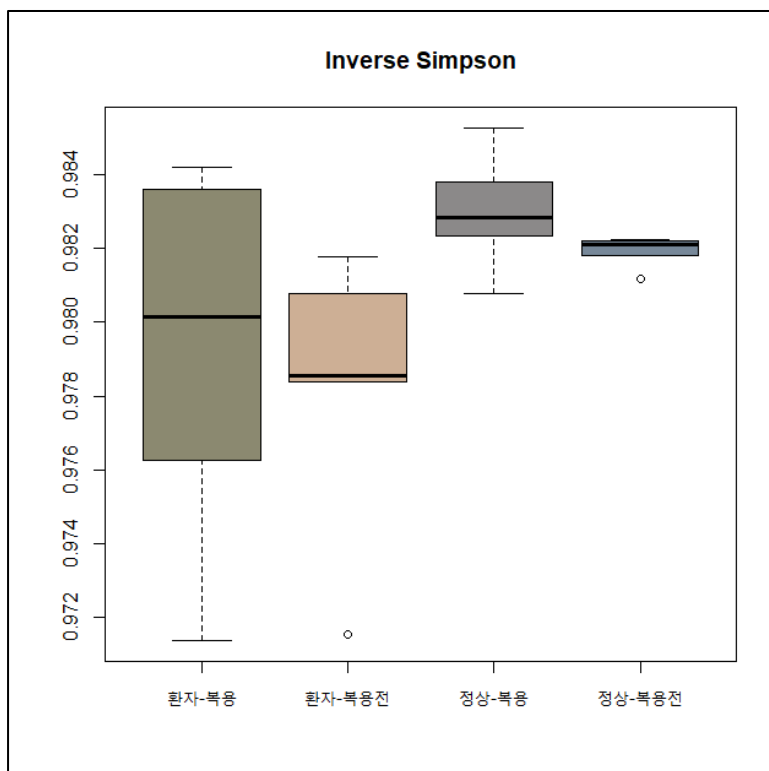
SampleID	OTUs	Chao1	Shannon	Inverse Simpson
T01A	319	461.3529	6.657892	0.976253
T02A	333	543.8491	6.533391	0.971387
T03A	318	481.75	6.704505	0.980134
T04A	337	526.06	6.883988	0.984186
T05A	334	488.3929	6.843044	0.983613
T01B	384	533.1429	7.033476	0.981775
T02B	364	552.1311	6.846905	0.978566
T03B	334	476.5965	6.613068	0.971542
T04B	362	540.1228	6.910792	0.980783
T05B	363	565.2321	6.833564	0.978375
N06A	424	621.5775	7.237545	0.985263
N07A	434	671.5405	7.152676	0.982833
N08A	438	704.2297	7.110781	0.982348
N09A	425	721.1	7.068139	0.980768
N10A	457	735.6143	7.304369	0.983786
N06B	439	678.4	7.148075	0.981798
N07B	429	660.2895	7.113774	0.98209
N08B	450	643.5172	7.184023	0.982204
N09B	414	565.1852	7.047773	0.981185
N10B	417	613.08	7.095155	0.982252

9. 위 표의 Shannon과 Inverse Simpson 값에 대해, [네 그룹의 box plot](#)을 그리세요.

> Shannon의 box plot



> Inverse Simpson의 box plot

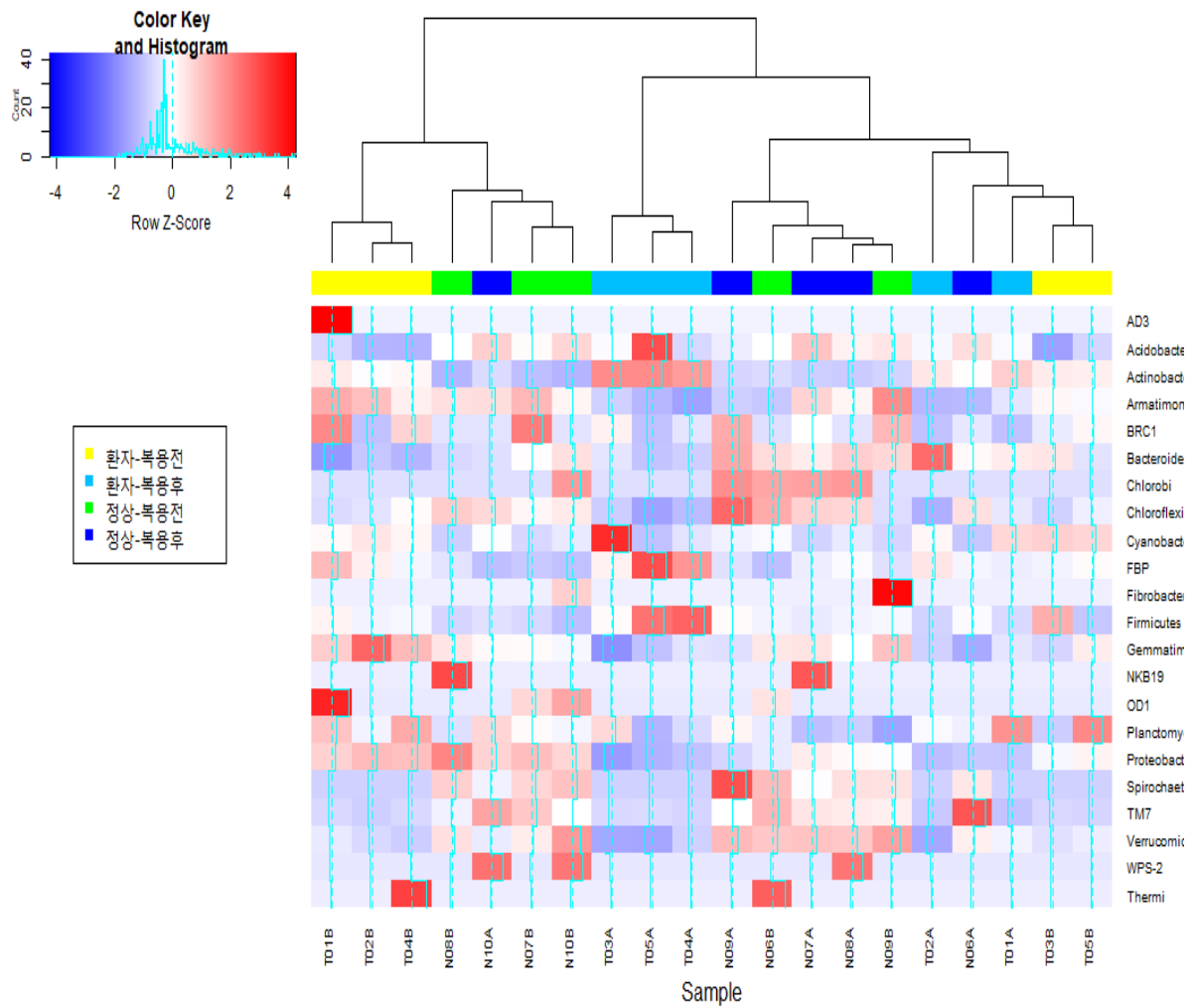


10. Taxonomy abundance(%)를 phylum(문) 수준에서 아래 3개의 graph로 표현하세요.

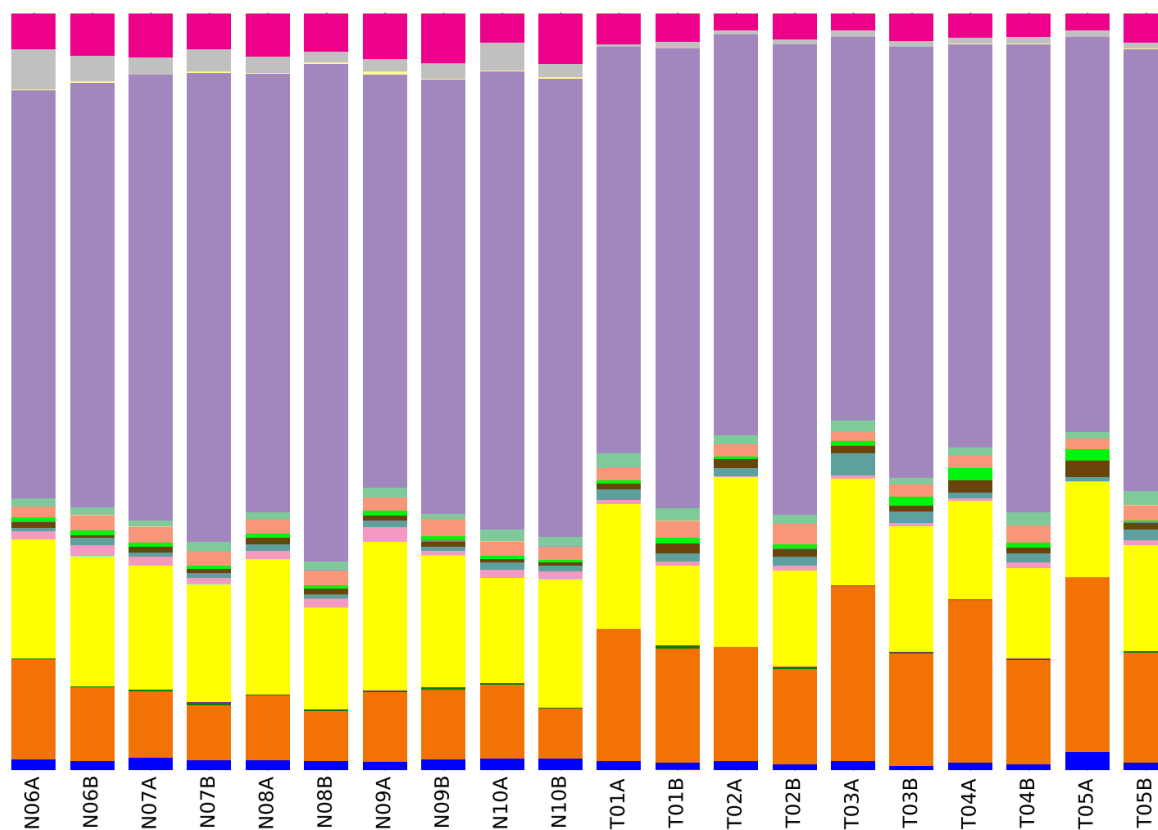
> Heatmap























!! 샘플의 그룹 정보를 color bar로 표현할 것.

!! X축(샘플)에 대해서만 Hierarchical clustering을 적용할 것.



> Bar graph – 20샘플의 분포 비율(%)

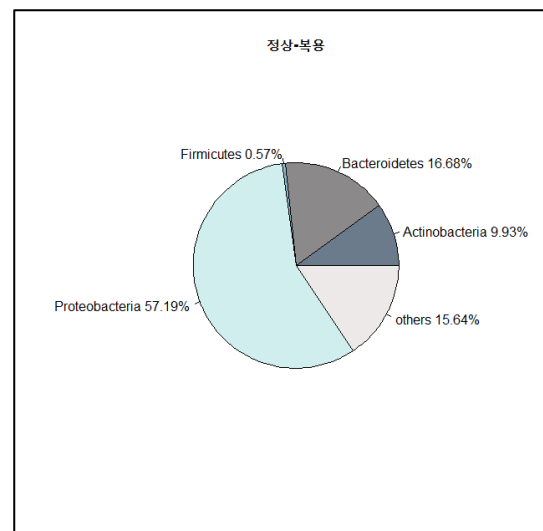
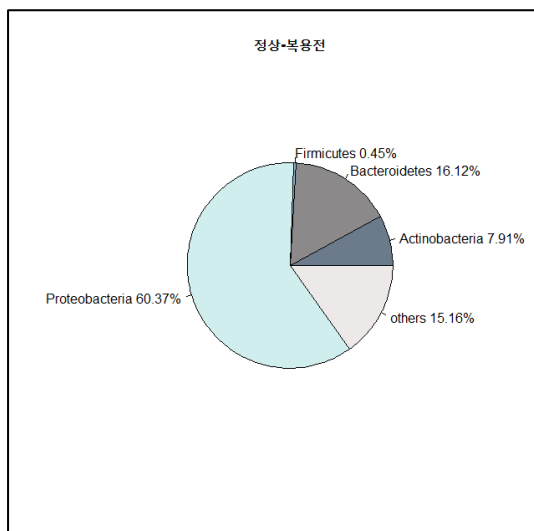
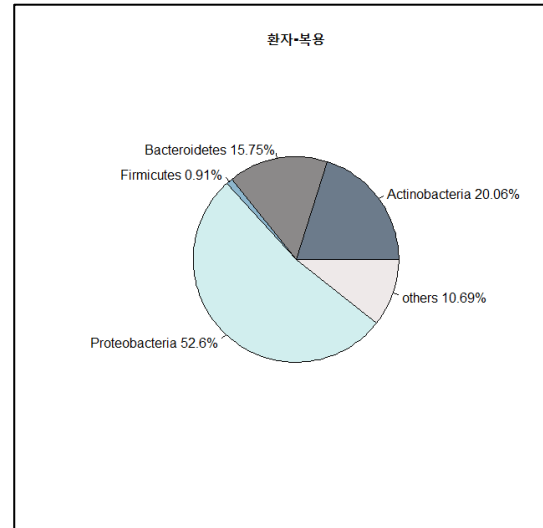
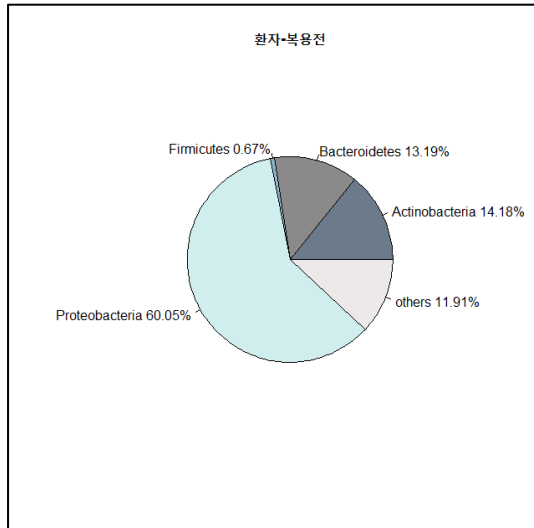


		Total	N06A	N06B	N07A	N07B	N08A	N08B	N09A	N09B	N10A	N10B	T01A	T01B	T02A	T02B	T03A	T03B	T04A	T04B	T05A	T05B
Legend	Taxonomy	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%	%
	k_Bacteria:p_AD3	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_Acidobacteria	1.2%	1.4%	1.2%	1.6%	1.3%	1.3%	1.2%	1.1%	1.4%	1.5%	1.5%	1.2%	1.0%	1.2%	0.7%	1.2%	0.6%	0.9%	0.7%	2.4%	0.9%
	k_Bacteria:p_Actinobacteria	13.0%	13.3%	9.8%	8.8%	7.4%	8.5%	6.6%	9.2%	9.2%	9.8%	6.6%	17.4%	15.1%	15.1%	12.7%	23.1%	14.8%	21.5%	13.8%	23.1%	14.6%
	k_Bacteria:p_Armatimonadetes	0.1%	0.0%	0.1%	0.2%	0.2%	0.1%	0.2%	0.1%	0.3%	0.2%	0.1%	0.1%	0.3%	0.0%	0.2%	0.1%	0.1%	0.0%	0.2%	0.0%	0.1%
	k_Bacteria:p_BRC1	0.1%	0.0%	0.0%	0.1%	0.2%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%	0.0%	0.2%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%	0.1%
	k_Bacteria:p_Bacteroidetes	15.4%	15.7%	17.2%	16.3%	15.6%	17.8%	13.5%	19.7%	17.4%	13.9%	17.0%	16.4%	10.5%	22.4%	12.8%	14.1%	16.7%	13.1%	11.9%	12.6%	14.0%
	k_Bacteria:p_Chlorobi	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_Chloroflexi	0.8%	1.0%	1.4%	1.1%	0.8%	1.1%	1.2%	1.9%	0.5%	1.1%	1.0%	0.6%	0.5%	0.2%	0.6%	0.4%	0.4%	0.3%	0.8%	0.1%	0.7%
	k_Bacteria:p_Cyanobacteria	1.0%	0.5%	0.9%	0.6%	0.6%	0.8%	0.6%	0.9%	0.6%	1.0%	0.8%	1.4%	1.1%	1.1%	1.2%	2.9%	1.4%	0.8%	1.1%	0.5%	1.4%
	k_Bacteria:p_FBP	0.9%	0.8%	0.4%	0.8%	0.5%	0.9%	0.7%	0.7%	0.7%	0.4%	0.4%	0.8%	1.4%	1.1%	1.0%	1.0%	0.8%	1.7%	0.8%	2.2%	0.9%
	k_Bacteria:p_Fibrobacteres	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_Firmicutes	0.6%	0.6%	0.6%	0.5%	0.4%	0.6%	0.4%	0.7%	0.6%	0.5%	0.3%	0.4%	0.7%	0.4%	0.6%	0.7%	1.1%	1.6%	0.6%	1.5%	0.3%
	k_Bacteria:p_Gemmatimonadetes	1.9%	1.4%	2.0%	2.0%	1.9%	1.9%	2.0%	1.7%	2.2%	1.9%	1.8%	1.7%	2.2%	1.6%	2.8%	1.2%	1.6%	1.7%	2.3%	1.5%	2.0%
	k_Bacteria:p_NKB19	0.0%	0.0%	0.0%	0.1%	0.0%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_OD1	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.1%	0.0%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_Plantomycetes	1.2%	1.1%	1.1%	0.8%	1.3%	0.9%	1.0%	1.3%	0.7%	1.5%	1.2%	1.9%	1.6%	1.2%	1.2%	1.5%	0.9%	1.0%	1.7%	0.8%	1.9%
	k_Bacteria:p_Proteobacteria	57.5%	53.9%	56.1%	58.9%	62.0%	58.0%	65.9%	54.6%	57.4%	60.7%	60.5%	53.7%	60.8%	53.0%	62.2%	50.7%	56.9%	53.4%	61.9%	52.3%	58.4%
	k_Bacteria:p_Spirochaetes	0.1%	0.1%	0.2%	0.1%	0.2%	0.1%	0.2%	0.4%	0.1%	0.1%	0.2%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_TM7	1.7%	5.3%	3.3%	2.3%	2.9%	2.2%	1.4%	1.7%	2.1%	3.7%	1.7%	0.4%	0.9%	0.5%	0.7%	0.8%	0.8%	0.8%	0.9%	0.9%	0.8%
	k_Bacteria:p_Verrucomicrobia	4.3%	4.7%	5.6%	5.8%	4.7%	5.7%	5.1%	6.0%	6.5%	3.8%	6.7%	4.0%	3.8%	2.2%	3.4%	2.2%	3.6%	3.2%	3.1%	2.2%	3.9%
	k_Bacteria:p_WPS-2	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
	k_Bacteria:p_Thermi	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

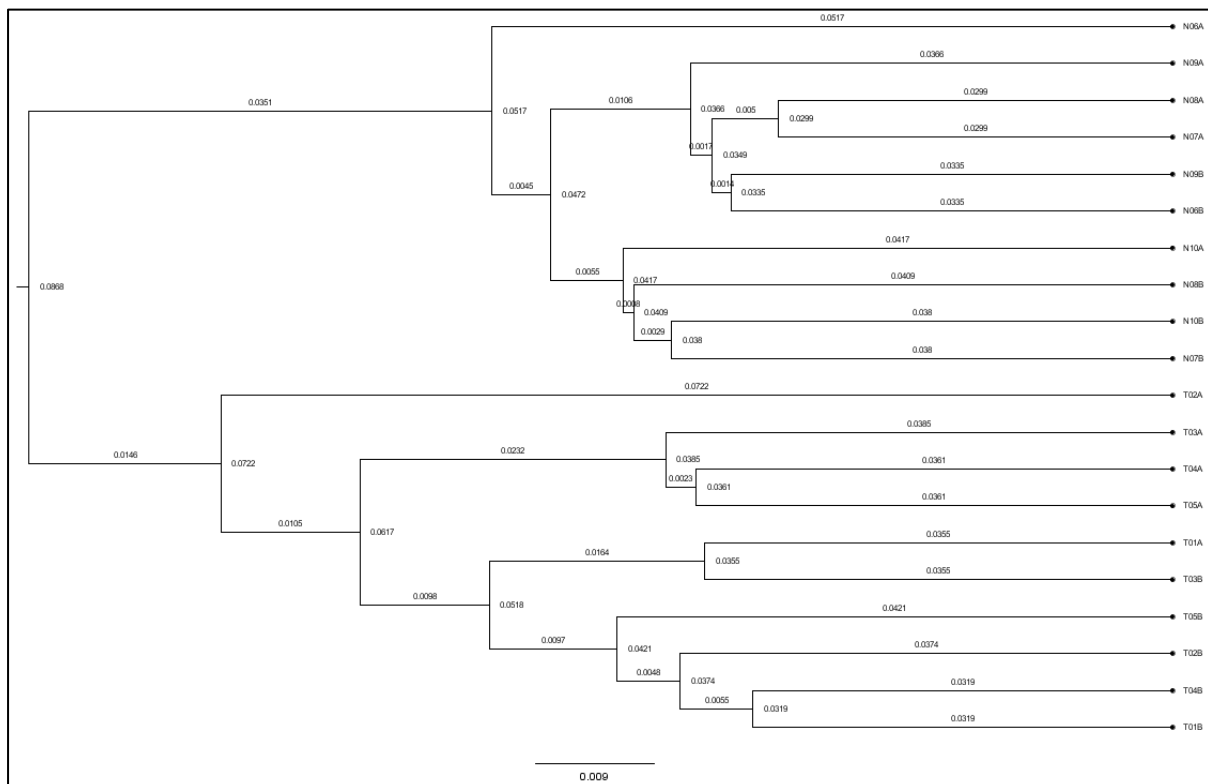
> Pie graph – 4그룹의 평균 분포 비율(%)

*단, 아래 4개 항목을 제외한 나머지는 Other로 합산하여 표기하세요.

Proteobacteria	Actinobacteria
Bacteroidetes	Firmicutes



11. [UPGMA tree \(weighted UniFrac\)](#)를 그리세요. (FigTree, TreeView X 등 활용)



12. 건강한 사람(Normal, Before)과 크론병 환자(Treatment, Before)의 [장내미생물 군집 구성에 차이가 '있다' 또는 '없다'로 주장](#)하고, 분석 결과를 근거로 제시하여 주장을 뒷받침하세요.

우선, 위의 alpha - diversity 표를 보며 건강한 사람(Normal, Before)과 크론병 환자(Treatment, Before)의 장내미생물 군집 구성에 차이가있음을 설명할 수 있다.

otu개수부터 살펴보았다. sample별로 신약을 복용하기 전 크론병 환자 sample의 otu수가 신약을 복용하기 전 건강한 사람의 otu수 보다 상대적으로 otu수가 적은 것을 확인 할 수 있다. 또한 각 otu에 대한 abundance를 phylum 수준에서 확인하여 장내미생물 군집 구성에 차이가 있음을 확인 할 수 있다. 이는 위의 pie graph를 통해 알 수 있다.

그래프를 보면, 5개의 phylum으로 나누어져있는 것을 알 수 있다.

"Actinobacteria", "Bacteroidetes", "Firmicutes", "Proteobacteria", "others" 로 구성 되어있으며, others 은 앞의 4개의 phylum을 제외한 나머지를 총칭한다. 건강한 사람(Normal, Before)과 크론병 환자 (Treatment, Before)의 pie graph를 보면, 크론병 환자에서는 Actinobacteria와 Firmicutes에서의 증가를 관찰 할 수 있다. 그리고 건강한 사람에서는 Proteobacteria, Bacteroidetes, others에서의 증가를 관찰 할 수 있다. 이를 통해 건강한 사람과 크론병 환자의 장내미생물 군집 조절에 차이가

있다 라고 주장할 수 있다.

13. 크론병 환자의 치료제 복용 전(Treatment, Before)과 복용 후(Treatment, After)의 장내미생물 군집 구성에 변화가 '있다' 또는 '없다'로 주장하고, 분석 결과를 근거로 제시하여 주장을 뒷받침 하세요.

크론병 환자의 치료제 복용 전과 복용 후 의 장내미생물 군집 구성에 변화를 관찰하기 위하여 alpha-diversity, heatmap, pie graph를 이용하였다.

크론병 환자가 신약을 복용하기 전과 후의 otu수를 보았다. 복용전, otu수가 복용후 otu수보다 상대적으로 큰 값을 가지고 있었다. 이를 통해 복용후 otu수가 감소함을 확인 하였다. 그리고 heatmap을 통해 color bar로 표시한 부분과, Hierarchical clustering된 부분을 확인 하였다. 만일 동일 인물에게서 얻은 sample이 약물 복용전과 복용후의 군집 조성의 변화가 없을 경우 heatmap에서 이루어지는 Hierarchical clustering의 결과가 가까운 거리로 표현 될 것이다. 하지만 heatmap을 확인해보았을 때, 약물 복용전과 복용후의 clustering 에따른 거리 결과가 다른 sample과 섞여 있는 것을 알 수 있다. 이를 통해 약물 복용 전과 후에 장내 미생물 군집 조성이 변화했을 것이라 생각 할 수 있다.

그리고 마지막으로 pie graph를 통해 otu abundance에 대한 비율을 보았을때, Proteobacteria, others는 감소하는 경향을 보였지만 ,나머지 Actinobacteria, Bacteroidetes, Firmicutes는 증가하는 경향을 보였다.

이 세가지 결과를 통해서 크론병 환자가 약물을 복용했을 때 장내미생물 군집 조성의 변화가 생겼다고 주장할 수 있다.

14. 본 분석 결과를 통해, 크론병 치료제(신약)의 약효와 부작용(건강한 사람이 복용했을 때)에 대해 의견을 기술하세요.

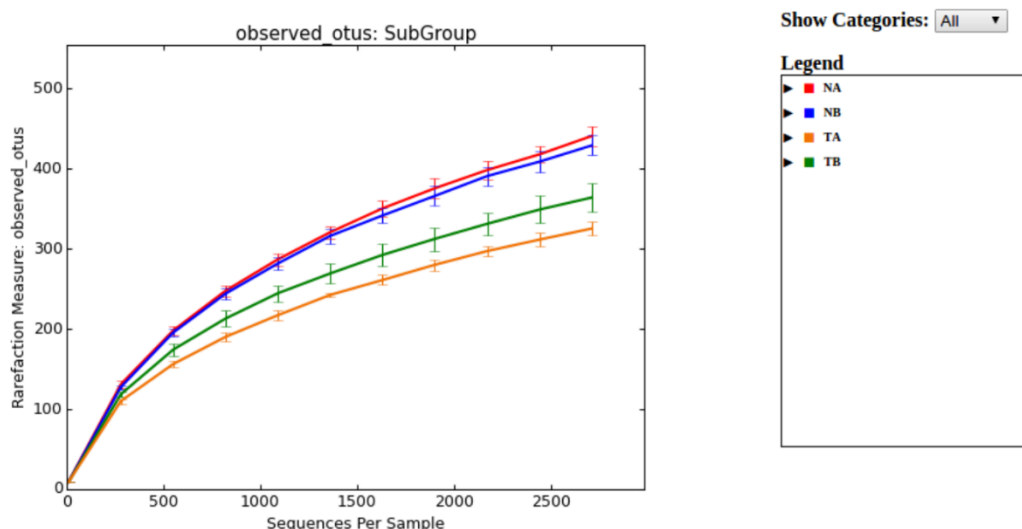
크론병 치료제의 약효를 증명하기 위해서는 약물을 복용하기 전에 환자에서 나타나는 미생물의 군집의 비율과 약물을 복용하기 전에 정상에서 나타나는 미생물 군집의 비율을 비교한뒤, 차이가 있는 군집을 병의 특성과 관련된 유의미한 군집으로 정의 해야한다. 따라서 약물이 효과가 있다면 약물을 복용하기 전 환자에서 많이 나타나는 군집은 약물을 복용하였을 때 감소되어야하고, 환자에서 적게 나타나는 군집은 정상과 비슷한 수치로 증가해야 약물의 효과가 의미가 있다고 판단 할 수 있다. Pie graph를 보면, 약물을 복용하기 전 정상인 그룹보다 약물을 복용하기전 환자 그룹에서 더 많이 발현되는 군집은 "Acidobacteria, Firmicutes"이다. 약물을 복용한뒤 환자군에서 나타나는 군집의 비율을 살펴보면, Acidobacteria, Firmicutes, bacteroidetes이 증가하였고,

proteobacteria는 감소하였다. 미생물의 군집의 비율을 통해 정상인 상태를 정의하자면, 전체에서 Acidobacteria가 7.91%, Firmicutes가 0.45%, Bacteroidetes가 16.12%만큼의 균형을 이루어야 정상이라고 할 수 있다. 하지만 약물 복용시 이러한 균형이 깨져 버리고 Actinobacteria의 비율이 크게 증가한다. 따라서 약물 복용시, 약물을 복용하기 전 정상인 상태로 변하는 것이 아니라 균형이 깨지기 때문에 오히려 부작용이 있다고 설명할 수 있다. Proteobacteria는 약물 복용전 정상 그룹과 환자그룹의 큰 차이가 없기 때문에 질병과 관련된 군집이라 정의하기 힘들다. Others는 환자끼리 비교하고 정상끼리 비교해보았을 때 약물 투여시 일정하게 증가하거나 감소하지 않기 때문에 질병과 관련되어있다고 정의하기 힘들다.

건강한 사람이 약물을 복용하였을때는 Acidobacteria, Firmicutes, bacteroidetes의 비율이 크게 변화하지 않기 때문에 약물에 대한 부작용이 나타나지 않는다고 설명할 수 있다.

결론적으로 크론병을 가진 환자와 정상인을 구분하는 기준은 Acidobacteria, Firmicutes, bacteroidetes이며, 이 군집의 비율에 의해서 정상인 사람의 특징과 질병을 가진 사람의 특징이 구분 된다고 할 수 있다. 약물 복용시 환자에게는 이러한 비율이 정상인 그룹처럼 변화하는 것이 아니라 오히려 비율이 더 크게 변동되어 약효가 없고 부작용이 의심된다고 설명할 수 있다. 하지만 정상인 그룹에 투여시 미생물 군집의 비율이 유지되기 때문에 부작용은 없다고 설명할 수 있다.

한가지 의심되는 부분을 추가하자면 이번 연구에서 얻은 데이터들을 가지고 약물에 대한 평가를 제대로 할 수 있는지에 대한 부분이다.



Rarefaction graph를 통해 보았을때, 시퀀싱을 더 많이 하게 된다면 새로운 otu가 나올 가능성이 있다. 따라서 이번 연구에서 얻은 데이터외에 약물이 작용할 때 반응하는 다른 otu들이 존재한다면, 실험 결과에 영향을 미칠 수 있지 않을까 생각된다. 약물의 효과에 대해서 검증을 하기 위해선, 발견할 수 있는 모든 otu에 대해서 약물의 효과를 검증해야 유의미한 결과라고 여길 수 있다고 생각하기 때문에 약간의 미심쩍은 부분이 있다.