

통합 데이터 분석 기반 뇌신경계 유전질환 진단 기술 개발

정성원, Ph.D

가천대학교 의예과 유전체외과학전공

가천대 길병원 유전체외과학연구소

2019. 11. 01. @한국지능시스템학회 추계학술대회

뇌신경계 유전질환 ~ 희귀질환

◆ 희귀질환의 정의

국내 유병인구 20,000명 이하거나
진단이 어려워 유병인구를 알 수 없는 질환

미국: 유병인구 200,000명 미만인 질환
일본: 유병인구 50,000명 미만인 질환
유럽: 1/2,000명(0.05%) 미만의 유병률을 가진
life-threatening or chronically debilitating diseases

희귀질환의 특성

- ◆ 유전성 (genetic)
 - 80% 가 유전성 질환
 - 소아가 50% 이상
 - 특정 유전자의 이상으로, 해당 유전자가 맡고 있는 생체 기능의 이상에 의한 질환

- ◆ 높은 사망률
 - 생후 1년 내 사망 원인의 35%
 - 5세 안에 30% 사망

- ◆ 다양성
 - 7,000종 이상의 희귀질환 (어떤 유전자에, 어떤 이상이 생겼는지에 따라)
 - 임상양상 및 유전형의 다양성

- ◆ 개별 질환은 희귀하나, 전체적으로는 희귀하지 않음
 - 일반적으로 전체 인구의 약 5~10%

- 12살
- 다양한 근육 무력증 연관 증상 (걷기 힘들, 대화 어려움, 머리 세우기 힘들, 음식 삼키기 어렵고 호흡곤란)
- 10년 이상 원인을 몰랐음



- 환자의 전체 유전자 sequence 분석
- DDC 유전자에 변이(mutation)를 발견
... 신경세포 사이 신호 전달 물질인 도파민과 세로토닌의 생성에 관여
- 환자에게 도파민작용제를 투여
- 수개월 뒤, 휠체어가 필요 없을 정도로 증상 개선





선천성 근무력증 11살 보경이가
태어나 처음 걷게 된 과정

2003년 희귀 유전성 질환인
선천성 근무력증으로 출생

두 살 때 근육 조직 검사 결과
근무력증 진단

지난해 열 살 때까지 줄곧 눕거나
누군가 앉혀주면 앉아 생활해옴

서울대병원, 미국 워싱턴대 의대
유전체 분석 연구소에 검사 의뢰



근육질환 관련 DOK7 유전자
변이 발견

올해 초부터 신경과 근육 연결
작용 활성 신경 물질(아세틸콜린)
강화제 투여

6월부터 조금씩 일어서기 시작

10월 현재 화장실 걸어서 가고
계단 오르기 시작

자료: 서울대병원 소아신경과

- 11세
- 선천성 근무력증
- 희귀질환 연관 유전자 분석에서 DOK7 유전자 변이 발견
... 신경세포와 근육세포 연결 형성에 관여
- 신경~근육세포 연결 작용 활성 물질 투여
- 약 5개월 뒤 증상 개선 확인

희귀질환 진단의 가치

- ◆ 환자 및 보호자에게 병 원인에 대한 답을 제시
 - 원인을 찾을 때 까지 여러 병원을 돌지 않아도 됨
- ◆ 치료에 대한 가능성 제시
 - 원인 제거에 적합한 치료제가 존재하는 경우 증상 완화/개선 가능성이 있음
 - 향후 맞는 치료제의 개발을 기다릴 수 있음
- ◆ 부모에게 유전질환에 대한 상담 제공
 - 자녀가 질환의 원인 유전자 변이를 부모로부터 물려받은 경우

희귀질환 진단의 가치

◆ 희귀질환 치료제 개발에 연결

- 원인 유전자의 오작동을 상쇄하는 방향의 신약 개발

◆ 희귀질환 신약의 경제성

- 환자의 전 생애에 걸친 처방 + 고가의 약값
 - 2018년 기준 미국 희귀질환 환자 약제 비용 평균 15만 달러 (일반 환자 3만달러)
 - 노바티스사의 척수성근위축증 치료제 졸겐스마 가격 25억원
- 2022년까지 전세계 처방 의약품 매출 20% 이상을 희귀의약품이 차지할 것으로 전망
- 일본 제약회사 다케다는 희귀질환 치료제 개발 기업 샐리어를 70조원에 인수 (2019년 1월)

희귀질환의 진단



선천성 근육력중 11살 보경이가 태어나 처음 걷게 된 과정

2009년 희귀 유전성 질환인 선천성 근육력중으로 출생

두 살 때 근육조직 검사 결과 근육력중 진단

지난해 열 살 때까지 중궁 놓거나 누군가 없었으면 아이 생활해움

서울대병원, 미국 워싱턴대 외대 유전체 분석 연구소에 검사 의뢰

근육질환 관련 DDK1 유전자 변이 발견

올해 초부터 신경과 근육 연결 작용 활성 신경 물질(아세틸콜린) 강화제 투여

6월부터 조금씩 일어서기 시작

10월 현재 화장실 걸어서 가고 계단 오르기 시작

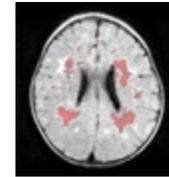
자료: 서울대병원 소아신과과



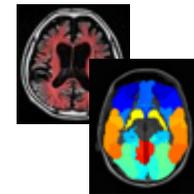
이상 증상



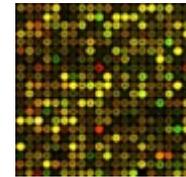
기존 알려진
희귀질환의
증상과 비교



뇌 영상의 이상 확인



구조적 이상
병변 존재 여부



환자가 가진 유전자 변이 확인

Gene X

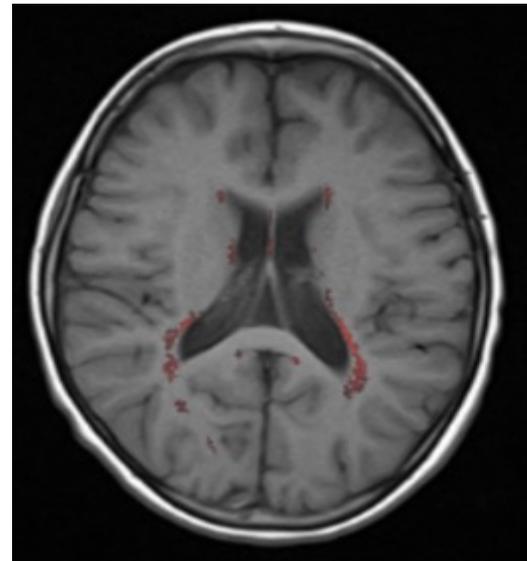
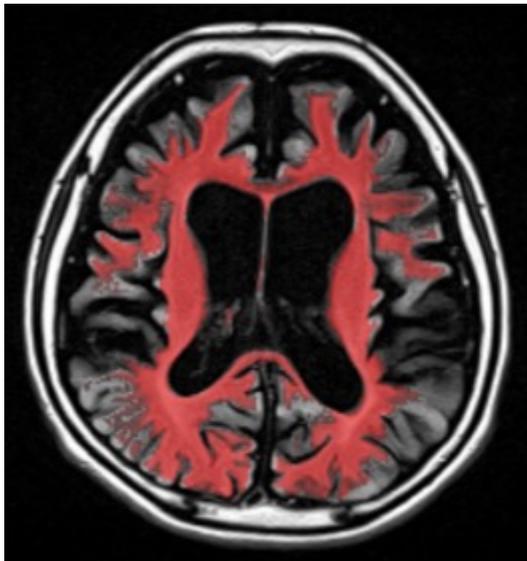
질환 유발
가능성이 높은
변이 1~2개
식별

7,000 종 이상의 희귀질환 중 하나를 찾는 것

진단

희귀질환 진단의 어려움

- ◆ 임상 증상의 다양성 및 경계의 불분명성
 - 현재 환자의 증상이 7,000여개 이상 알려진 희귀질환 중 어느 것과 유사한가?
- ◆ 동일 질환에서도 증상에 차이가 있을 수 있음

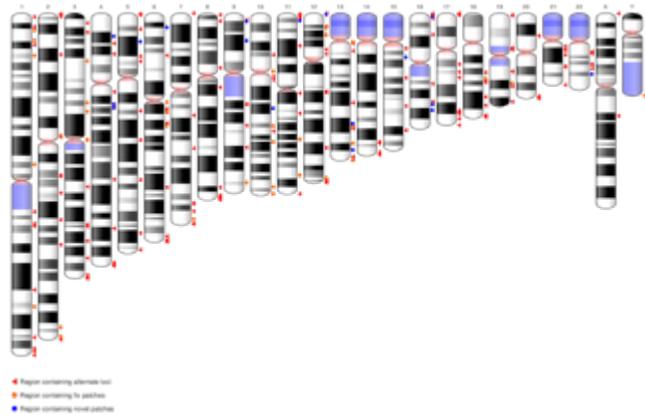


Ceroid lipofuscinosis, neuronal, 6 (CLN6)

동일 질환 두 환자의 MRI 에 나타난 병변 영역에 차이가 큼

희귀질환 진단의 어려움

- ◆ 질환 유발 가능성이 높은 유전자 변이 식별의 어려움



인간 표준 참조 DNA sequence

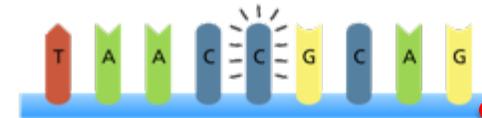


환자의 혈액에서 DNA 를 채취하여
표준 참조 DNA sequence 대비
다른 곳을 확인

Original sequence



Point mutation



	A	B	C	D	E
184543	chrY	59023917	59023917	T	C
184544	chrY	59023958	59023958	G	A
184545	chrY	59024836	59024836	G	T
184546	chrY	59024845	59024845	G	A
184547	chrY	59024852	59024852	G	A
184548	chrY	59025553	59025553	C	A
184549	chrY	59025998	59025998	T	C
184550	chrY	59026513	59026513	A	G
184551	chrY	59027058	59027058	A	G
184552	chrY	59027652	59027652	T	G
184553	chrY	59027700	59027700	A	C
184554	chrY	59028885	59028885	T	A
184555	chrY	59028889	59028889	A	C
184556	chrY	59028959	59028959	C	T
184557	chrY	59028981	59028981	C	T
184558	chrY	59029265	59029265	C	T
184559	chrY	59029294	59029294	T	C
184560	chrY	59030625	59030625	C	T
184561	chrY	59030673	59030673	G	A
184562	chrY	59031861	59031861	T	G
184563	chrY	59032128	59032128	A	T
184564	chrY	59032136	59032136	A	G
184565	chrY	59032147	59032147	C	G
184566	chrY	59032457	59032457	T	C
184567	chrY	59033287	59033287	T	-

> 18만개

알려진 유전자만 확인해도, 한 사람이 갖는
유전자 변이는 굉장히 많음

원인 변이 1~2개 를 찾아야 함

희귀질환의 일반적인 진단 방법

- ◆ Step 1. 환자가 가진 유전자 변이 중 희귀한 변이만 추려냄 (대규모 유전체 정보 DB 활용)

ExAC Browser Beta

Interested in working on the development of this resource? [Apply here.](#)

6만여명

ExAC Browser (Beta) | Exome Aggregation Consortium

Search for a gene or variant or region

Examples - Gene: PCSK9, Transcript: ENST00000407236, Variant: 22-46615880-T-C, Multi-allelic variant: rs18

About ExAC

The Exome Aggregation Consortium (ExAC) is a coalition of investigators seeking to aggregate and harmonize exome sequencing data from a wide variety of large-scale sequencing projects, and to make summary data available for the wider scientific community.

gnomAD browser

gnomAD v2.1.1 | Search

gnomAD v3 released: 71,702 genomes aligned on GRCh38.

12만여명

gnomAD

genome aggregation database

gnomAD v2.1.1 | Search by gene, region, or variant

Please note that gnomAD v2.1.1 and v3 contain largely non-overlapping samples and both datasets must be used to capture the full set of variation across gnomAD. For more information, see the FAQ "Should I switch to the latest version of gnomAD?"



유전자A 의 101번째 염기서열 이상 25% of population



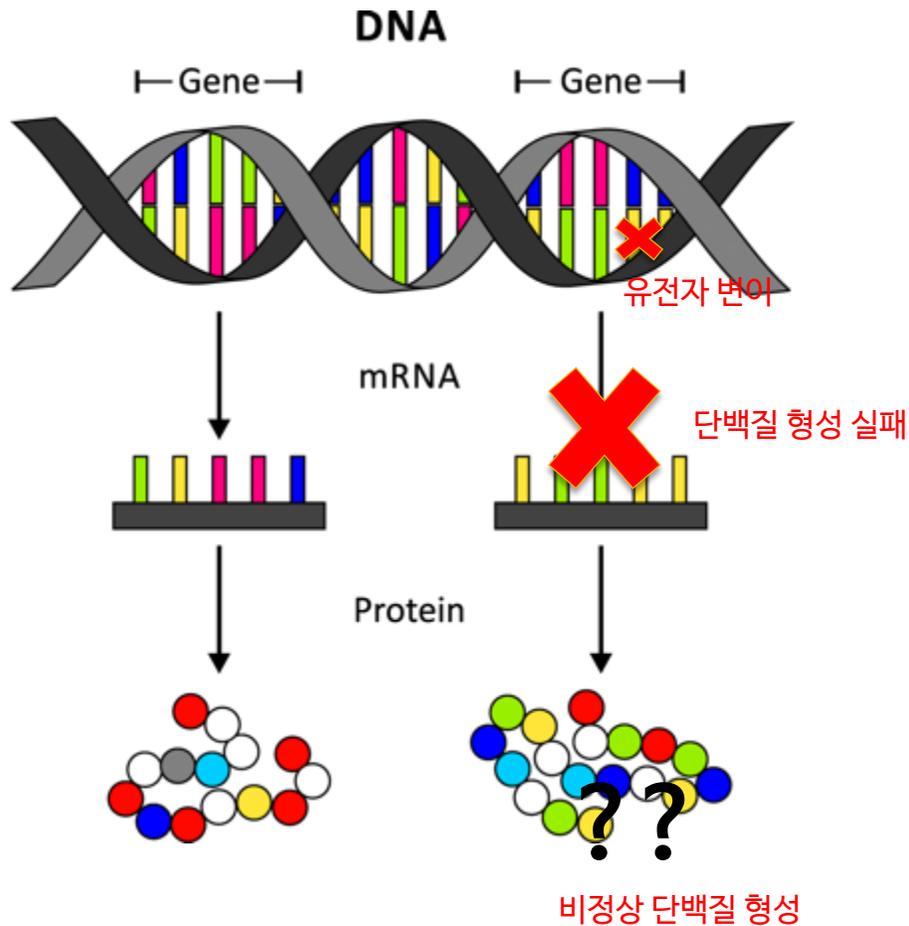
유전자B 의 42번째 염기서열 이상 0.002% of population



여전히 수십개 이상의 유전자 변이가 추려짐

희귀질환의 일반적인 진단 방법

- ◆ Step 2. 희귀한 변이들 중 유전자 기능에 손상을 끼칠 수 있을 것으로 예상되는 변이를 찾음



	A	D	E	F	G	H	AJ	AK	AL	AM	AN	AO
1	Gene.refGene	Chr	Start	End	Ref	Alt	SIFT_score	SIFT_pred	Polyphen2_H	Polyphen2_H	Polyphen2_H	Polyphen2_H
2	GJB4	chr1	35226964	35226964	G	A	0.06	T	0.993	D	0.706	P
3	FPGT-TNNI3	chr1	74716436	74716436	C	G	0.01	D	1	D	0.996	D
4	PPOX	chr1	161138854	161138854	C	G	0.04	D	0.997	D	0.944	D
5	FCGR2A	chr1	161483723	161483723	G	A	-999	.	-999	.	-999	.

각 유전자 변이의 나쁜 정도를 예측하는 소프트웨어 도구들이 존재

각 도구들의 정확도는 완벽하지 않음
(믿는다고보다, 참고하면 좋은 정도의 수준)

희귀질환의 일반적인 진단 방법

- ◆ Step 3. 현재까지 추려진 유전자 변이들, 유전자 손상 가능성, 환자 증상을 종합적으로 고려하여 "match" 되는 질환을 조사



여러가지 질환 유전자 데이터베이스 조사

256000

LEIGH SYNDROME; LS 질환명

Alternative titles; symbols
NECROTIZING ENCEPHALOPATHY, INFANTILE SUBACUTE, OF LEIGH; SNE

Other entities represented in this entry:
LEIGH SYNDROME DUE TO MITOCHONDRIAL COMPLEX I DEFICIENCY, INCLUDED
 LEIGH SYNDROME DUE TO MITOCHONDRIAL COMPLEX II DEFICIENCY, INCLUDED
 LEIGH SYNDROME DUE TO MITOCHONDRIAL COMPLEX III DEFICIENCY, INCLUDED
 LEIGH SYNDROME DUE TO MITOCHONDRIAL COMPLEX IV DEFICIENCY, INCLUDED
 LEIGH SYNDROME DUE TO MITOCHONDRIAL COMPLEX V DEFICIENCY, INCLUDED

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
2q35	Leigh syndrome	256000	Mi, AR	3	BCS1L	603647
5p15.33	Leigh syndrome	256000	Mi, AR	3	SDHA	600857
9q34.2	Leigh syndrome, due to COX IV deficiency	256000	Mi, AR	3	SURF1	185620
10q24.2	Leigh syndrome due to cytochrome c oxidase deficiency	256000	Mi, AR	3	COX15	603646
17p12	Leigh syndrome due to mitochondrial COX4 deficiency	256000	Mi, AR	3	COX10	602125

연관 유전자

INHERITANCE

- Autosomal recessive
- Mitochondrial

GROWTH

Other

- Failure to thrive

HEAD & NECK 증상

Eyes

- Ophthalmoplegia
- Optic atrophy
- Nystagmus
- Strabismus
- Ptosis
- Pigmentary retinopathy

RESPIRATORY

- Abnormal respiratory patterns
- Respiratory failure

SKIN, NAILS, & HAIR

Hair

- Hypertrichosis

MUSCLE, SOFT TISSUES

- Hypotonia

NEUROLOGIC

Central Nervous System

- Psychomotor retardation
- Hypotonia
- Ataxia
- Dystonia
- Dysarthria

데이터 분석 기반 접근 방법의 적용 가능성

- ◆ Step 1. 환자가 가진 유전자 변이 중 희귀한 변이만 추려냄 (대규모 유전체 정보 DB 활용)
 - ⇒ 대규모 유전체 정보 DB 를 기계적으로 검색
- ◆ Step 2. 희귀한 변이들 중 유전자 기능에 손상을 끼칠 수 있을 것으로 예상되는 변이를 찾음
 - ⇒ 각 변이가 “질환을 일으킨다 / 일으키지 않는다” 의 classification problem

데이터 분석 기반 접근 방법의 적용 가능성

- ◆ 환자의 임상 증상을 기존 알려진 질환 증상과 어떻게 정량적으로 비교? Ontology 기반 비교



Human Phenotype Ontology: 13,000 개 이상의 환자 증상 용어에 대한 hierarchical ontology structure

환자 증상을 표준 ontology 로 표현

Ontology 기반 semantic similarity 계산 기법을 활용

데이터 분석 기반 접근 방법의 적용 가능성

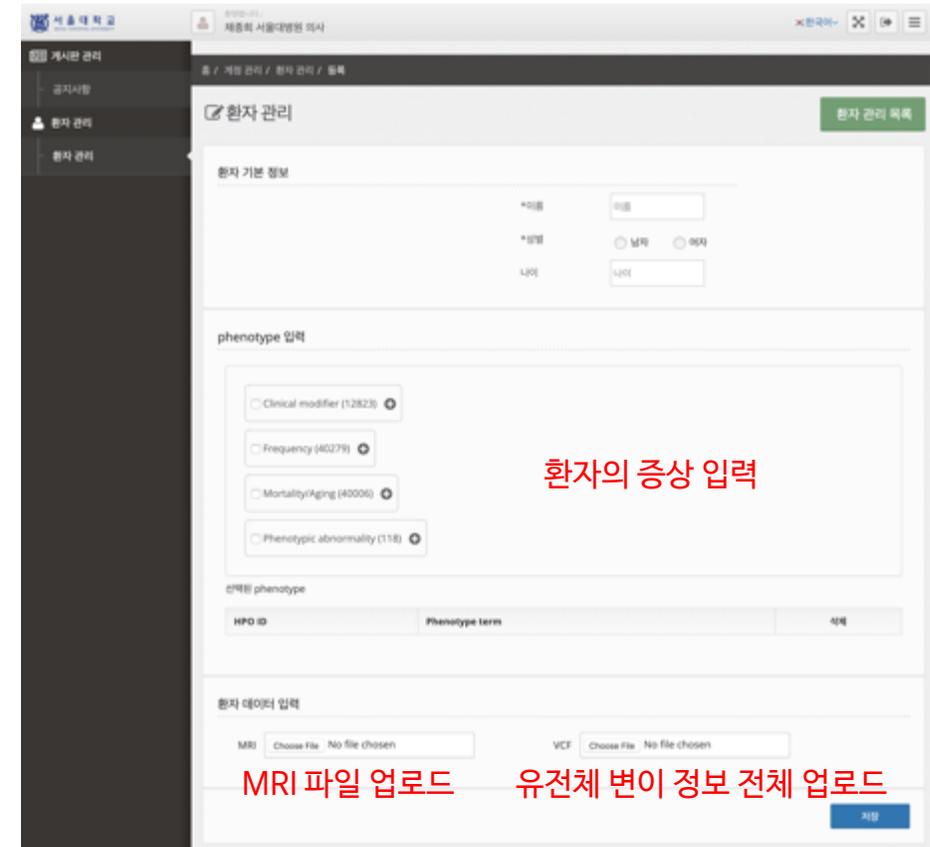
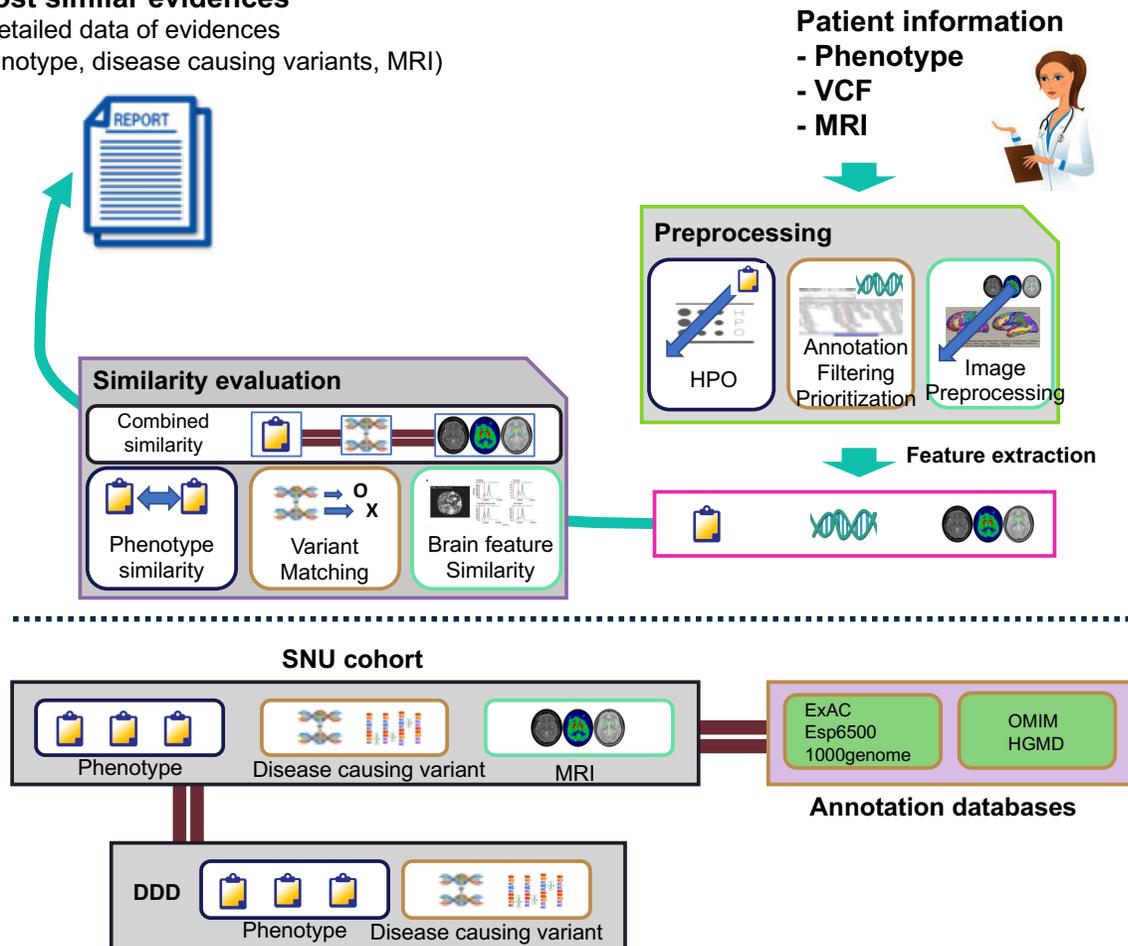
- ◆ Step 3. 현재까지 추려진 유전자 변이들, 유전자 손상 가능성, 환자 증상을 종합적으로 고려하여 "match" 되는 질환을 조사
 - 환자 유전자 변이에 기반한 질환 유사도 평가
 - 환자 증상에 기반한 질환 유사도 평가
 - MRI 데이터를 사용 가능한 경우 MRI 상의 유사도에 기반한 질환 유사도 평가
 - 상기 세 가지 유사도를 종합한 유사도를 기반으로 최종 질환 예측 결과 제시

본 연구의 접근 방법

◆ 데이터에 기반한 질환 예측 결과 제시

- Most similar evidences

... Detailed data of evidences
(phenotype, disease causing variants, MRI)



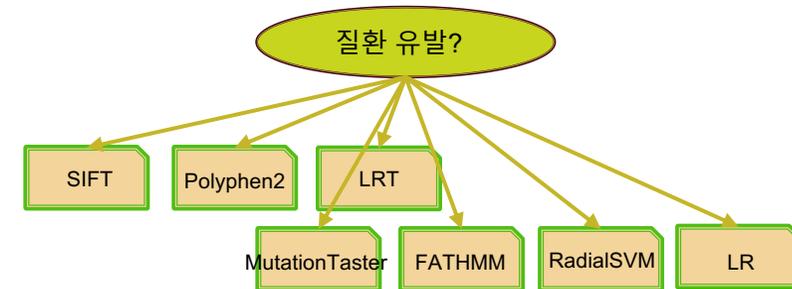
임상의는 환자 데이터 업로드만 하고 결과 확인

유전체 데이터에 기반한 질환 유사도 평가

- ◆ I. 대규모 유전체 데이터 기반 희귀 유전자 변이 판별
- ◆ II. 희귀 유전자 변이들에 대한 질환 유발 가능성 평가

	A	D	E	F	G	H	AJ	AK	AL	AM	AN	AO
1	Gene.refGene	Chr	Start	End	Ref	Alt	SIFT_score	SIFT_pred	Polyphen2_H	Polyphen2_H	Polyphen2_H	Polyphen2_H
2	GJB4	chr1	35226964	35226964	G	A	0.06	T	0.993	D	0.706	P
3	FPGT-TNNI3	chr1	74716436	74716436	C	G	0.01	D	1	D	0.996	D
4	PPOX	chr1	161138854	161138854	C	G	0.04	D	0.997	D	0.944	D
5	FCGR2A	chr1	161483723	161483723	G	A	-999	.	-999	.	-999	.

각 유전자 변이의 나쁜 정도를 예측하는 소프트웨어 도구들이 존재



질환 유발 가능성이 있는 것으로 알려진 약 56,000개의 유전자 변이
+
질환 유발 가능성이 없는 무작위로 선택된 약 56,000개의 유전자 변이
-> 112,000 개의 유전자 변이에 대한, 유전자 변이의 나쁜 정도를
예측하는 7개 소프트웨어의 평가 값 획득

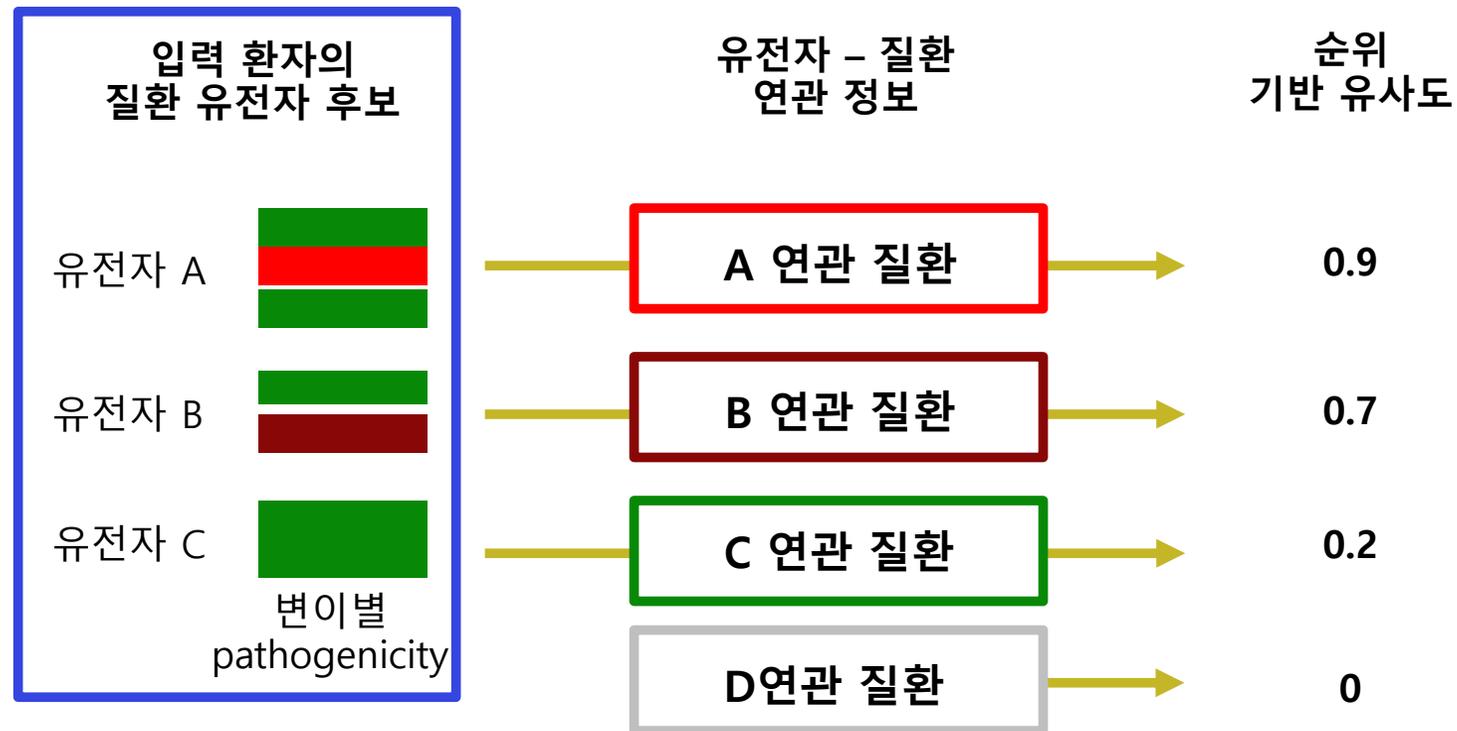
입력 데이터: 112,000 개의 유전자 변이
Feature: 7개 예측 소프트웨어의 예측 값

Naïve Bayes classifier 구축
(여러 질환 유발 가능성 예측 도구의 ensemble)

Before publication

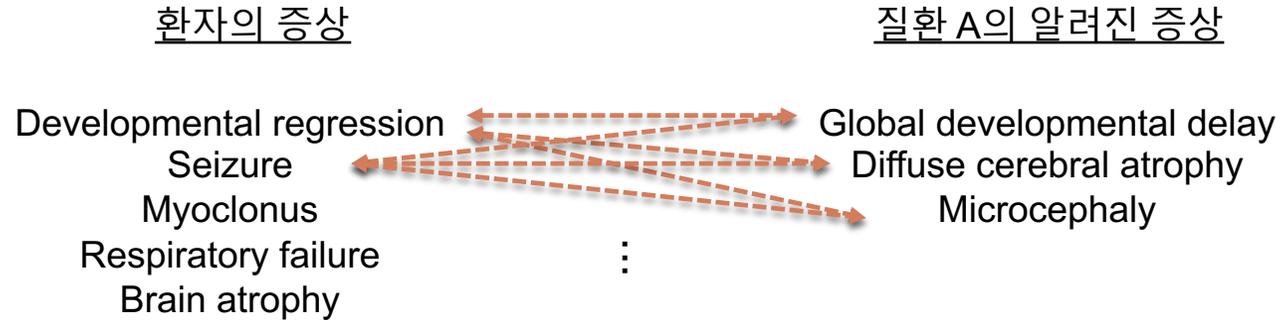
유전체 데이터에 기반한 질환 유사도 평가

- ◆ III. 질환 유발 가능성이 높은 정도에 따라 해당 유전자와 연관된 질환의 유사도를 결정



Before publication

임상 증상에 기반한 질환 유사도 평가

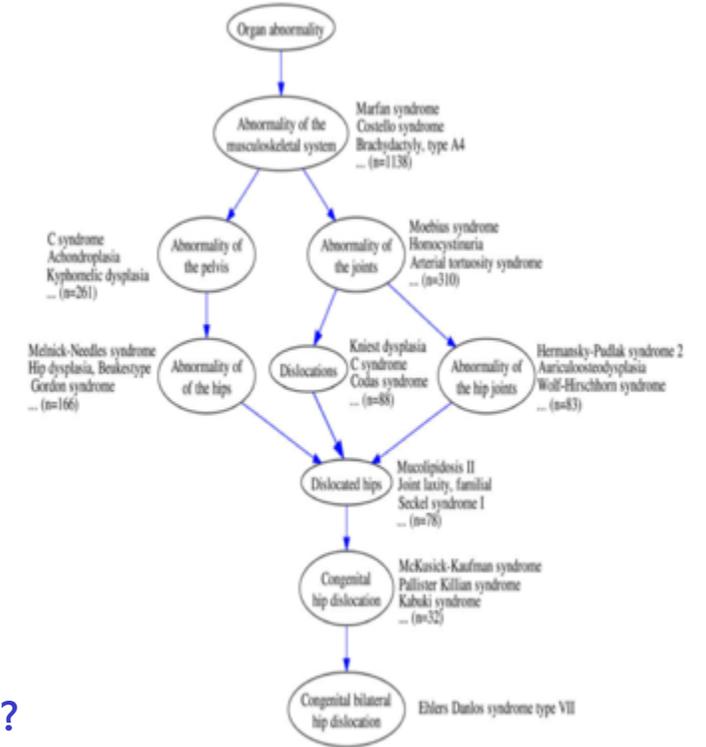


Ontology-based term-to-term similarity:

Information coefficient, Jiang-Conrath, Graph IC, Relevance, Wang, Lin, Resnik, ...

Aggregating multiple term-to-term similarities:

Max, Mean, FunSimAvg, FunSimMax, BMA, ...



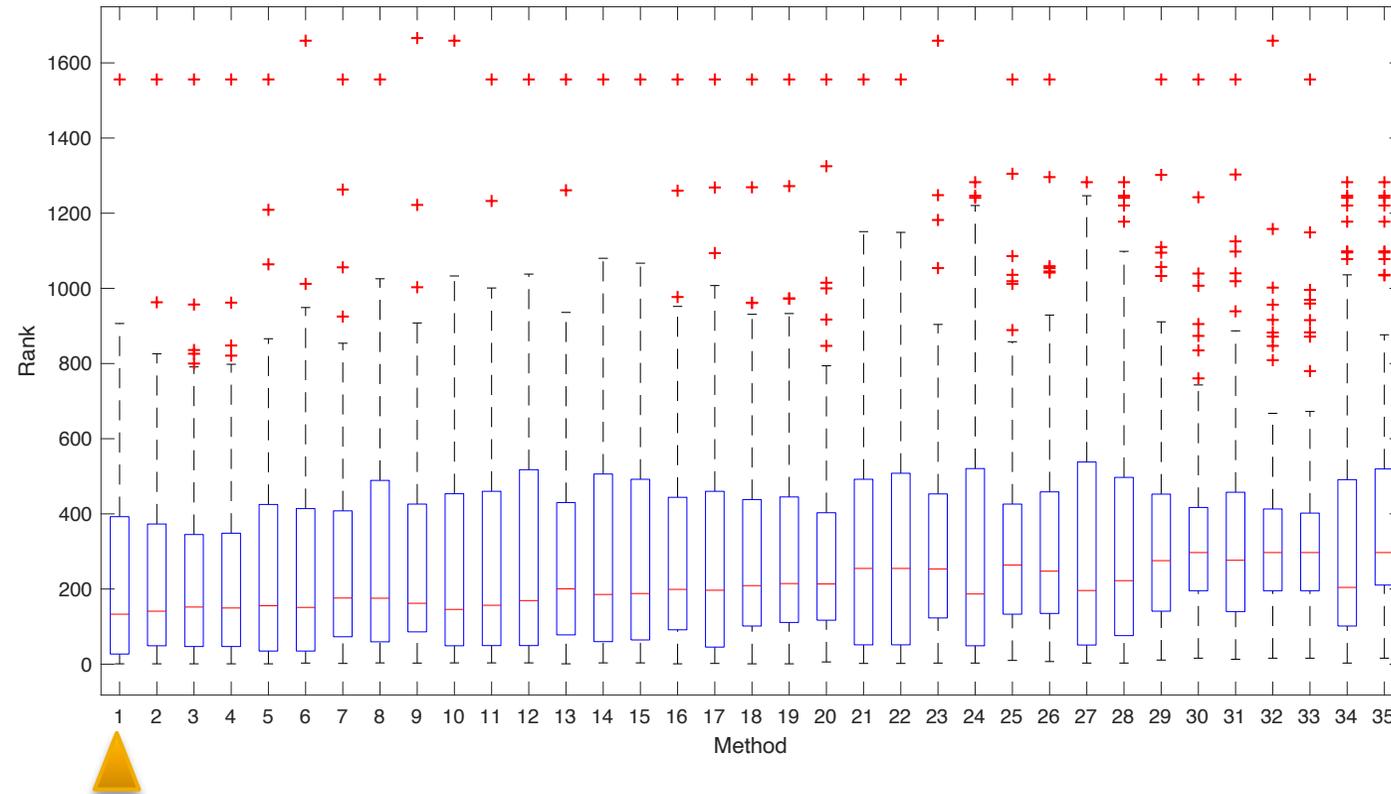
어떤 계산 방법이 최적인가?

Before publication

임상 증상에 기반한 질환 유사도 평가

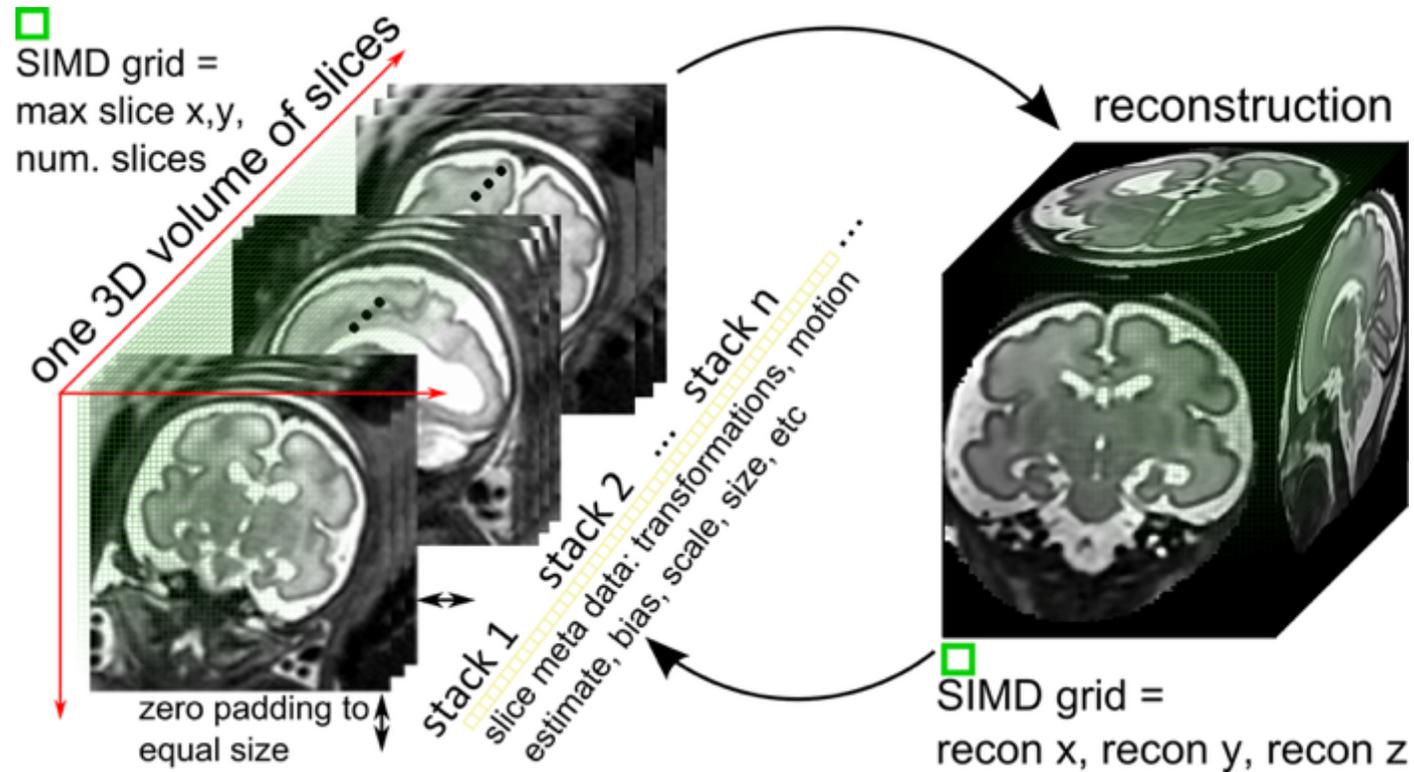
- ◆ 서울대병원의 진단 확정 희귀질환 환자 110례를 이용하여 평가
 - 35가지 방법을 이용하여, 각 케이스에 대해 정답 질환을 몇 순위로 예측하는지 확인
 - 정답 질환의 순위를 가장 높게 주는 방법을 선택

Before publication



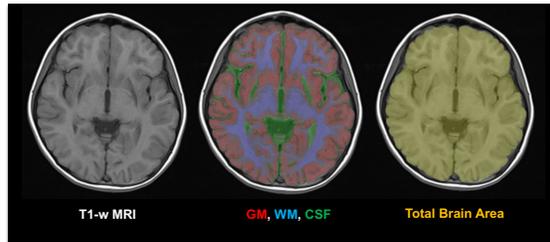
MRI 영상으로부터 Feature 추출 (시도)

- ◆ 전처리: 2D slice 로부터 3D 볼륨 재구성

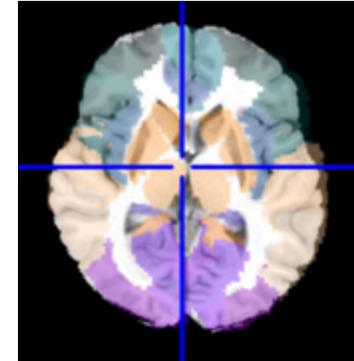


MRI 영상으로부터 Feature 추출 (시도)

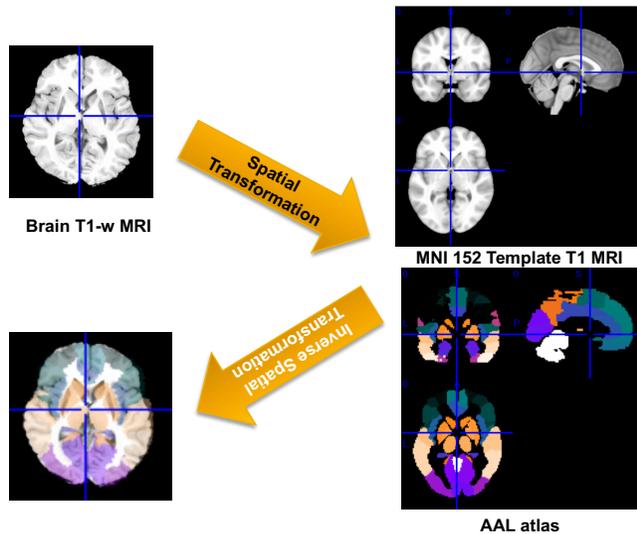
◆ 해부학적 영역의 3D 볼륨 확인



뇌 영역 식별



해부학적 영역의 볼륨 계산



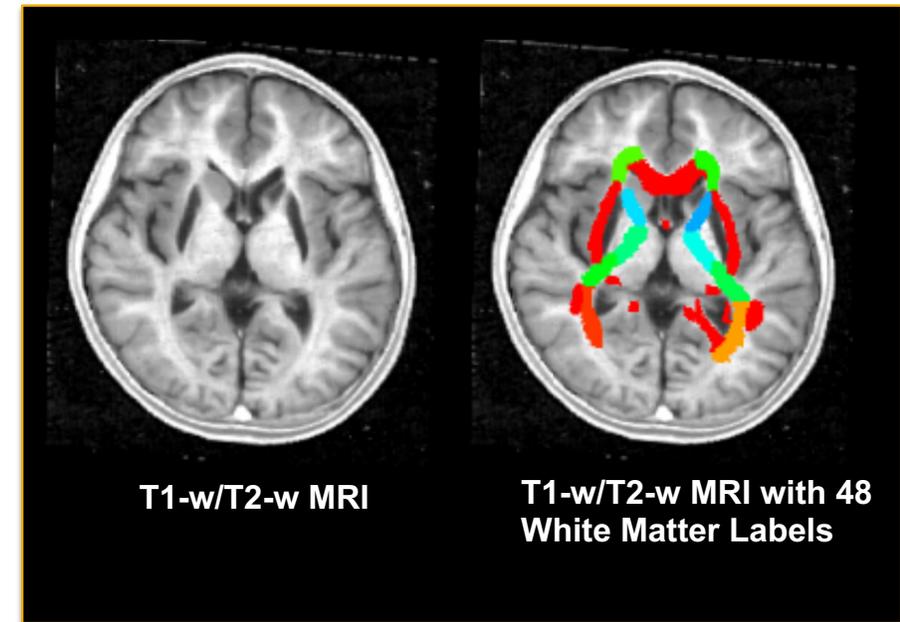
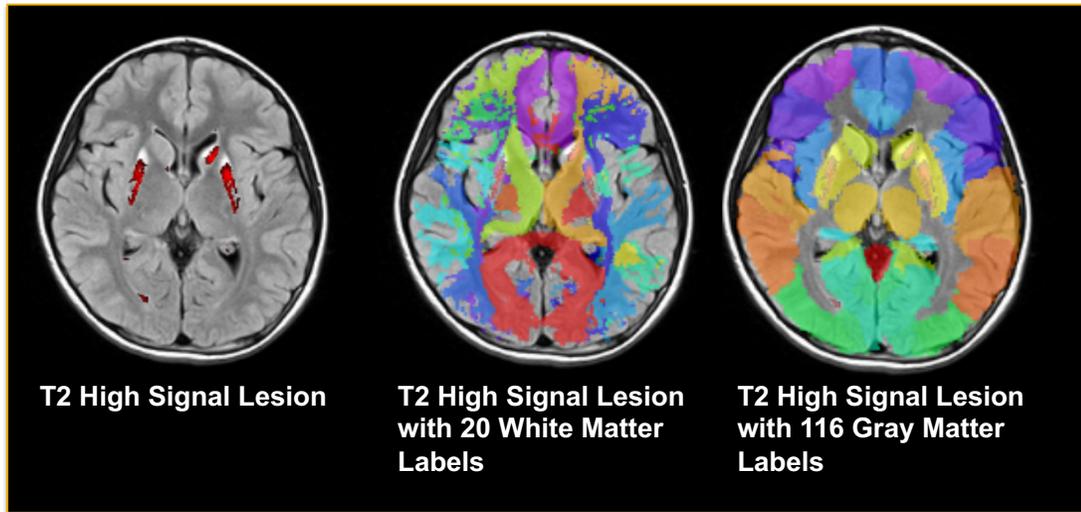
표준 atlas 에 mapping 을 통한 해부학적 영역 식별

74645118_10Y	Volume in mm3
Total Brain Volume	1363827.985
Total Gray matter Volume	808412.4138
Total White matter Volume	448472.3592
Total CSF Volume	311366.1129
Volume of Frontal Lobe	240720.0677
Volume of Temporal Lobe	186647.2427
Volume of Parietal Lobe	134996.3086
Volume of Occipital Lobe	99955.46179
Volume of Basal Ganglia	18362.6969
Volume of Thalamus	10266.09464
Volume of Cerebellum	114038.7286
Volume of left precentral gyrus	17565.7984
Volume of right precentral gyrus	15104.20443
Volume of left superior frontal gyrus (dorsolateral)	18629.27505
Volume of right superior frontal gyrus (dorsolateral)	22267.78318
Volume of left superior frontal gyrus, orbital	4631.086339
Volume of right superior frontal gyrus, orbital	5039.461374

진단을 알고 있는 다른 환자의 MRI 볼륨과 비교하여 유사도 평가

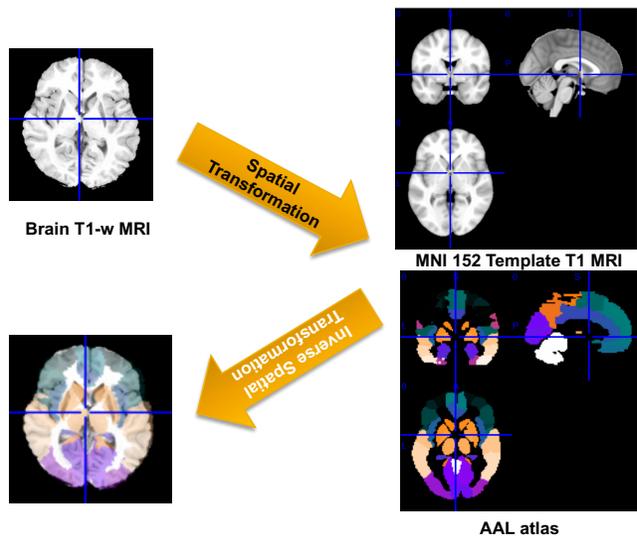
MRI 영상으로부터 Feature 추출 (시도)

- ◆ 병변 부위 식별 및 볼륨을 계산하여 비교에 사용
- ◆ 뇌 백질 발달 정도의 평가



MRI 영상 활용의 현실적 제약

- ◆ 3D 볼륨 재구성을 위해 필요한 수준의 데이터가 부족
 - Slice 사이의 간격이 촘촘해야 신뢰성 있는 3D 볼륨 재구성이 가능
 - 그러나 임상 현장에서는 일반적으로 간격을 비교적 넓게 잡음 (촬영 시간 단축 등)
- ◆ 소아 뇌 영역에 대한 표준 atlas 의 부재
 - 많은 희귀질환 환자가 2~3세 미만의 영유아



표준 atlas 에 mapping 을 통한 해부학적 영역 식별

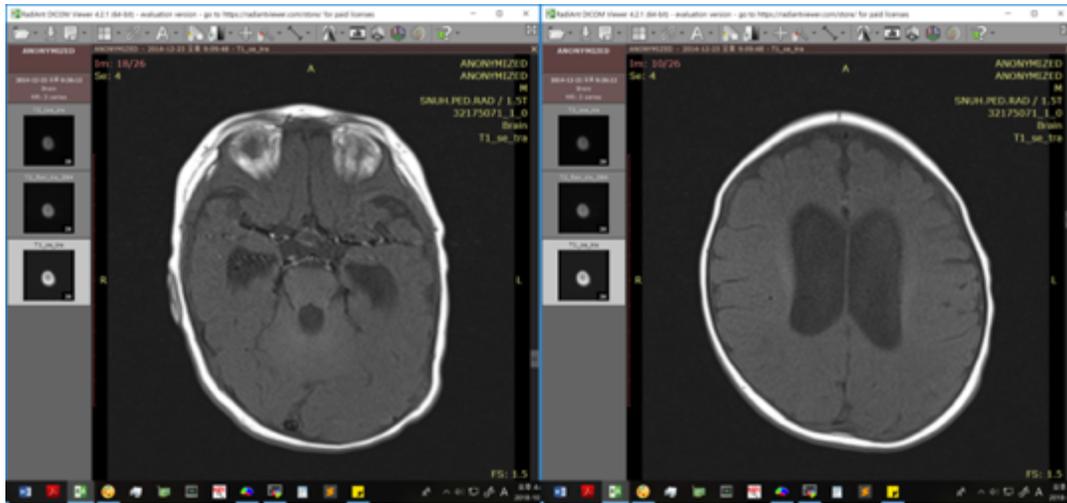
성장이 완료된 성인 뇌의 표준 atlas 는 존재

성장중인 소아, 영유아의 뇌에 대해서는 연령별로 표준화된 atlas 가 필요하나 존재하지 않음

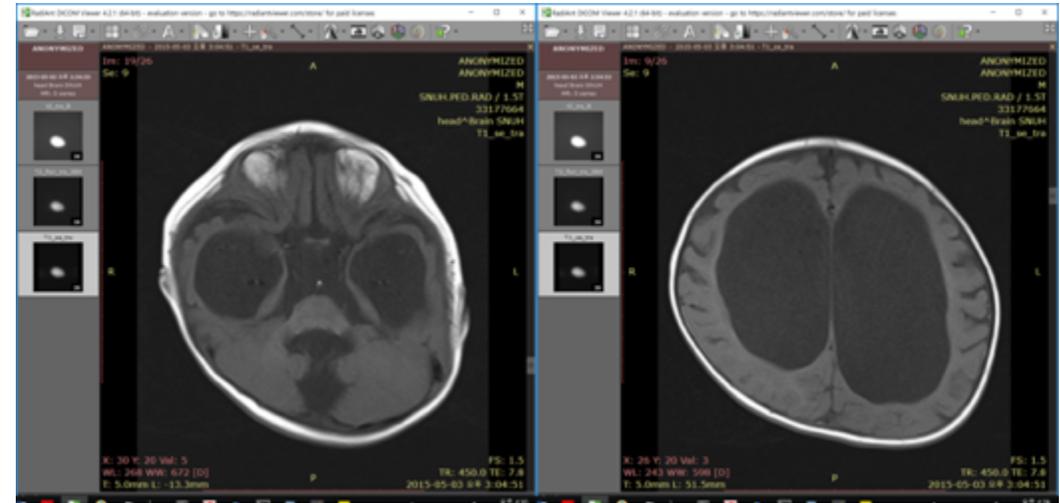
성인의 표준 atlas 를 사용하나 한계가 있음

MRI 영상 활용의 현실적 제약

- ◆ 표준 atlas 에 정합이 어려워 해부학적 영역의 식별이 힘든 경우
 - 뇌의 형태적 기형이 심한 질환군
 - 환자 움직임 등에 의해 잡음이 높은 영상



연령 4개월 소아 환자



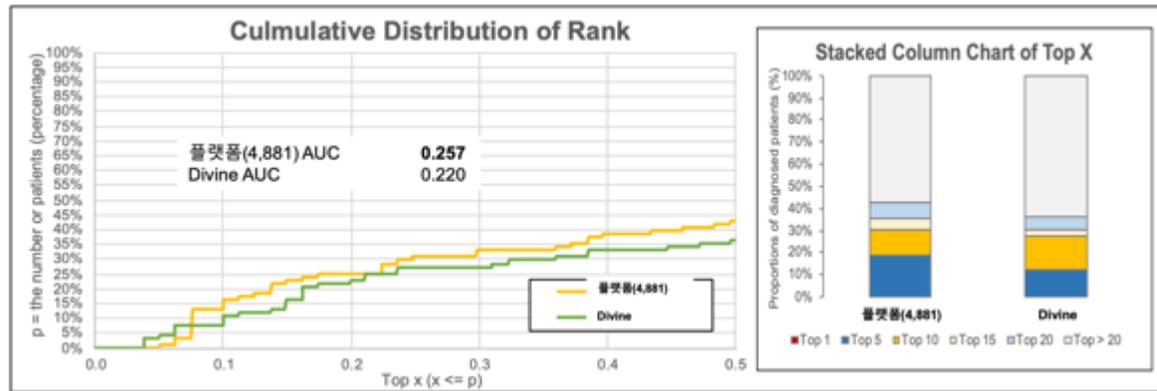
연령 9개월 소아 환자

뇌 해부학적 영역의 수치적 feature 를 사용하지 않고, 신경망 등을 기반으로 영상의 abnormality 존재 여부만 식별하여 feature 로 사용하는 방향으로 접근 방법 변경

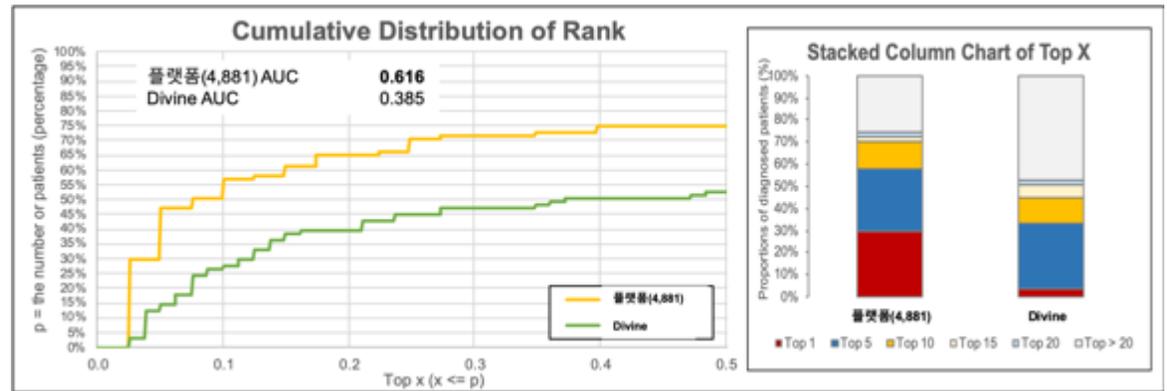
질환 예측의 성능 비교 평가



환자의 임상 증상만으로 예측한 경우



환자의 유전체 정보만으로 예측한 경우



임상 증상 + 유전체 정보로 예측한 경우

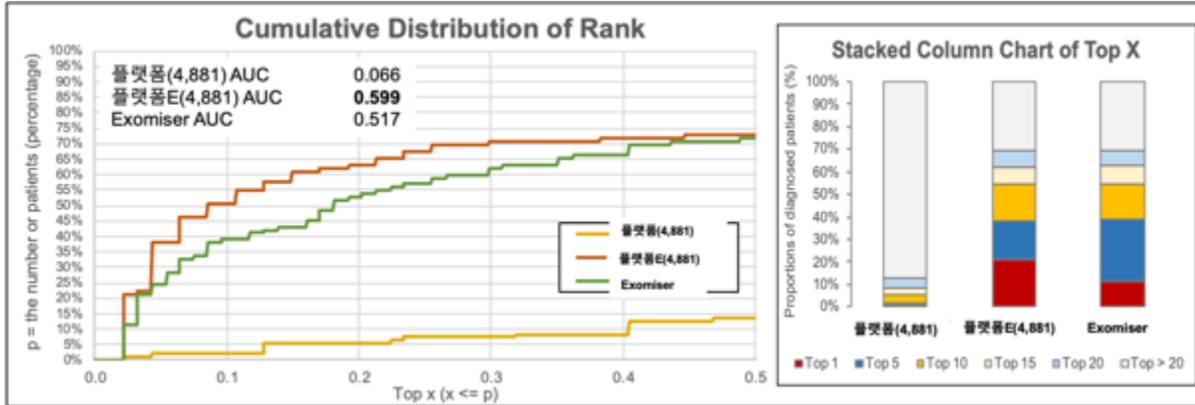
Before publication

비교 대상 1: 미국 클리블랜드 클리닉 개발 Divine

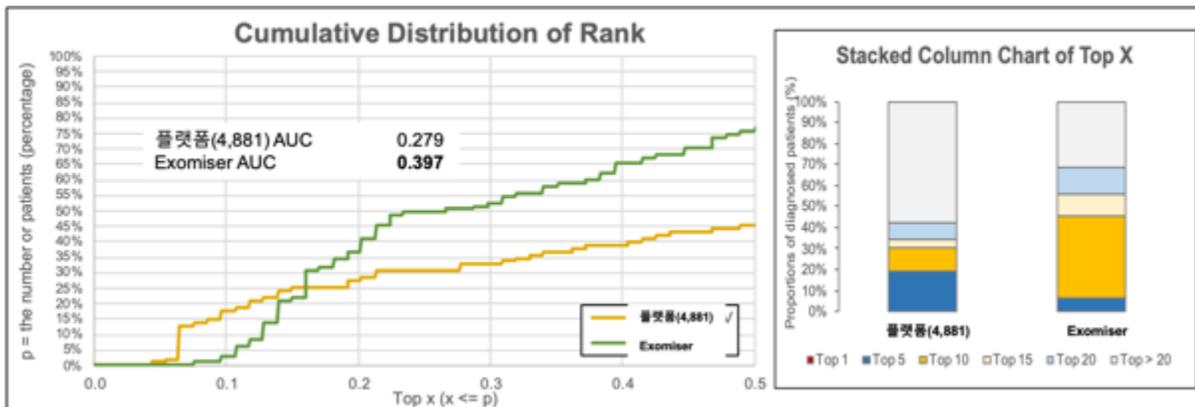
테스트 데이터: 서울대병원에서 진단이 확정된 희귀질환 환자 150명

평가: 상위 X위 이내로 정답 질환을 예측한 케이스 비율

질환 예측의 성능 비교 평가



환자의 임상 증상만으로 예측한 경우



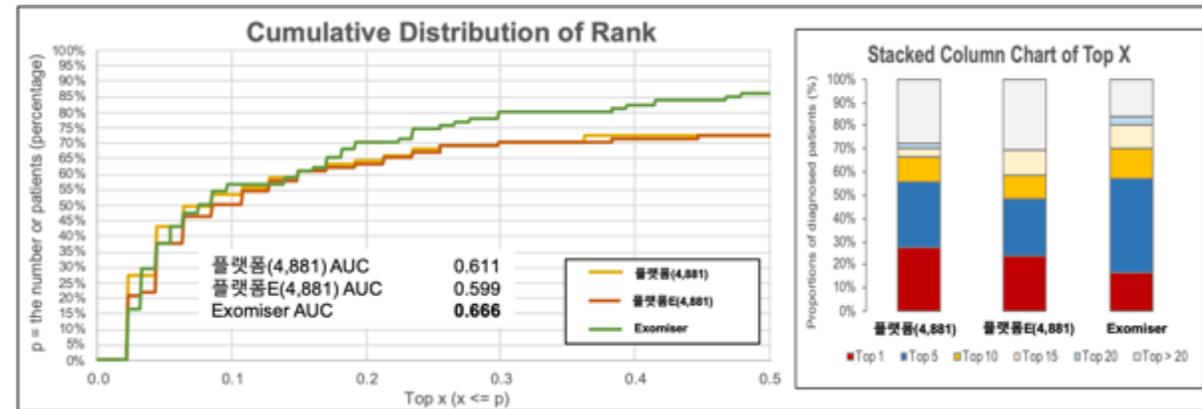
환자의 유전체 정보만으로 예측한 경우

Before publication

비교 대상 2: 영국 Sanger Institute 개발 Exomiser

테스트 데이터: 서울대병원에서 진단이 확정된 희귀질환 환자 150명

평가: 상위 X위 이내로 정답 질환을 예측한 케이스 비율



임상 증상 + 유전체 정보로 예측한 경우