

症例テキストのアノテーションに用いるオントロジーの課題

土肥栄祐¹⁾、高月照江³⁾、建石由佳²⁾、山本泰智³⁾、藤原豊史³⁾

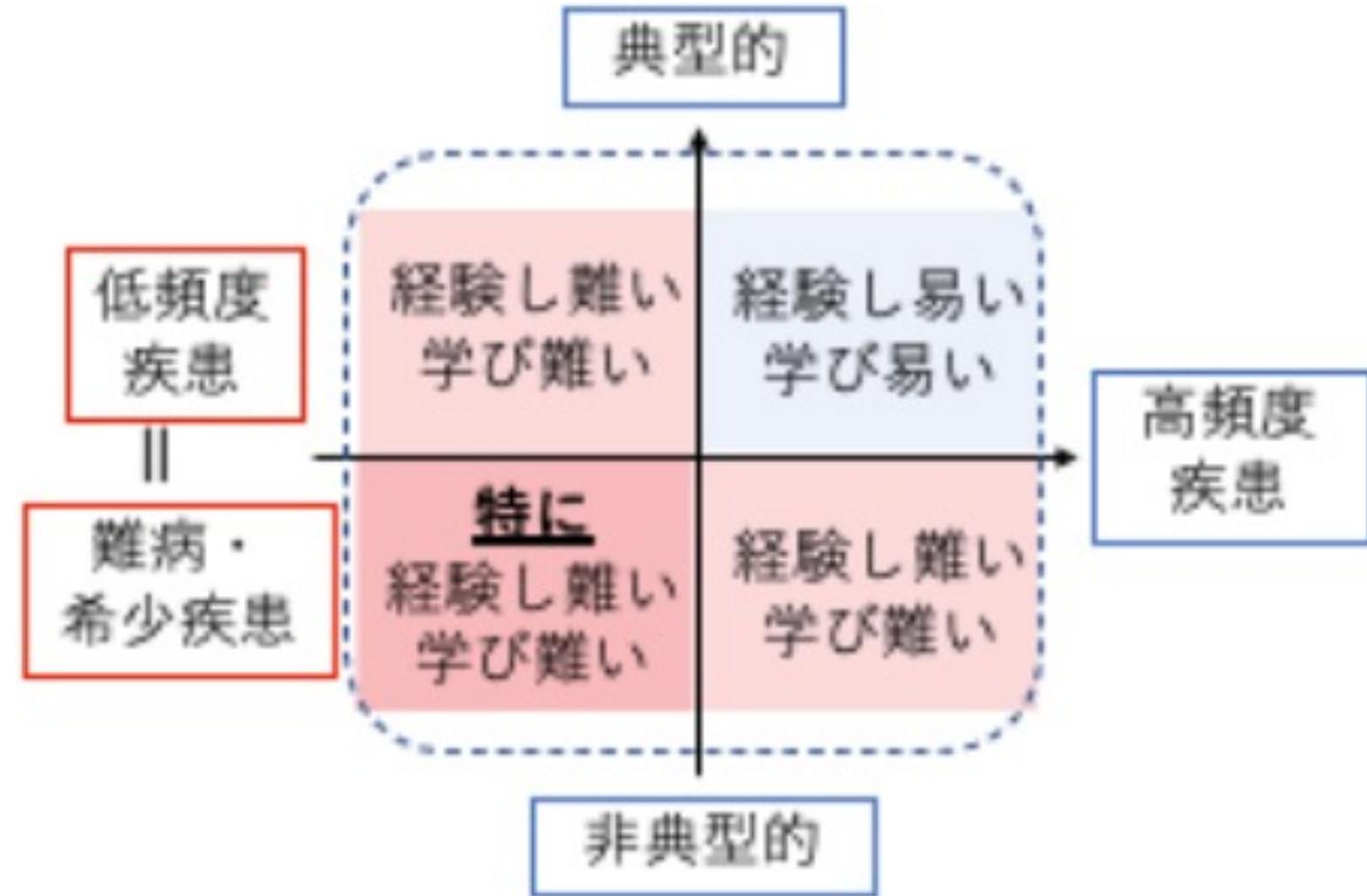
1) 国立精神神経医療研究センター 神経研究所 疾病研究第3部

2) 国立研究開発法人科学技術進行機構 NBDC 事業推進部

3) 情報・システム研究機構 ライフサイエンス統合データベースセンター

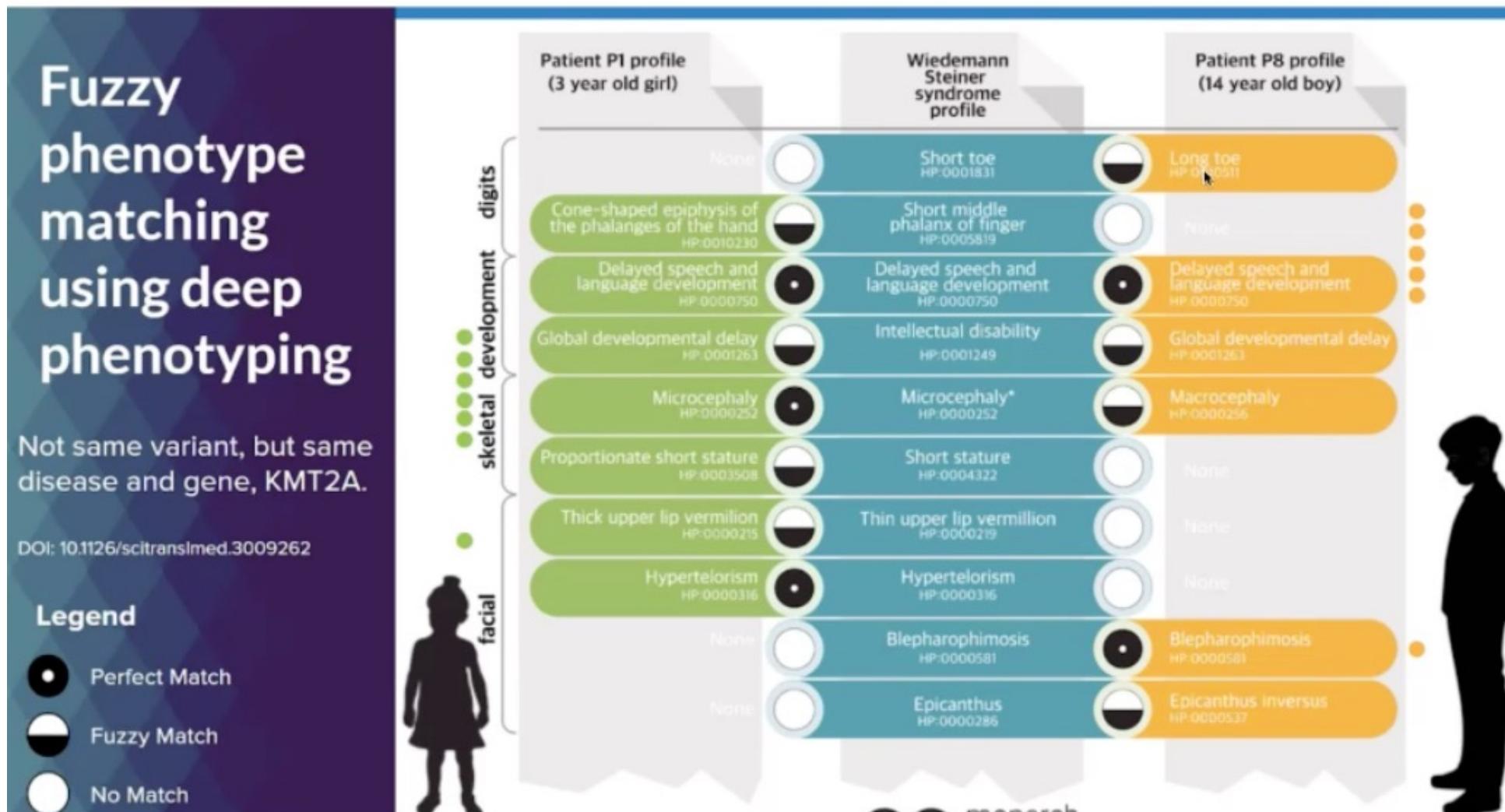
診断がつかず困る時

- 非常に初期、または非典型的
- 難病・希少疾患：10000疾患
- 人口の約5%
- 診断まで平均7-8年



なんとか手掛かりが欲しい

同じ遺伝子異常でも、症状や重症度が異なる



患者の兆候または症状を入力

ex) 小顎、尺骨低形成、短い指、上口唇裂、眼瞼裂

DBCLS

PubCaseFinder

English

Filter(3) Virtual Gene Panel Display options

Filter to narrow your search results Examples: Example1 Example2 Example3

Query box

GENEID:115399 LRRC56 OR GENEID:3265 HRAS OR GENEID:1756 DMD

Genetic Disease Rare Disease Gene Case

10 results

Costello syndrome (73.9%)

Wide anterior fontanel, Macroglossia, Full cheeks, Posteriorly rotated ears, Short neck, Epicanthus, Depressed nasal bridge, Hypertrichosis, Anteverted nares, Micrognathia, Thick lower lip vermillion, Hyperextensibility of the finger joints, Talipes equinovarus, Obstructive sleep apnea, Intellectual disability, Ventricular septal defect, Pulmonic stenosis, Achilles tendon contracture, Autosomal dominant inheritance, HRAS

Costello syndrome (CS) is a rare multisystemic disorder characterized by failure to thrive, short stature, developmental delay or intellectual disability, joint laxity, soft skin, and distinctive facial features. Cardiac and neurological involvement is common and there is an increased lifetime risk of certain tumors. >> Translate(Google)

OMIM:218040 Find images(Google) MONDO:0009026 KEGG:H01747 GeneReviews:NBK1507 GTR:C0587248

Clinical features Show(72) Case reports(PubMed) Show(137)

large congenital melanocytic nevus (61.7%)

Prominence of the premaxilla, Broad forehead, Full cheeks, Open mouth, Broad nasal tip, Hypertelorism, Everted lower lip vermillion, Narrow nasal ridge, Somatic mutation, HRAS, NRAS

Matched Phenotype Mode of Inheritance Causative Gene

100万例以上の症例報告抄録 (本文では無い) 疾患－症状の関係を利用

症状を入力し可能性のある疾患のランキングで提示

遺伝子でフィルターも可

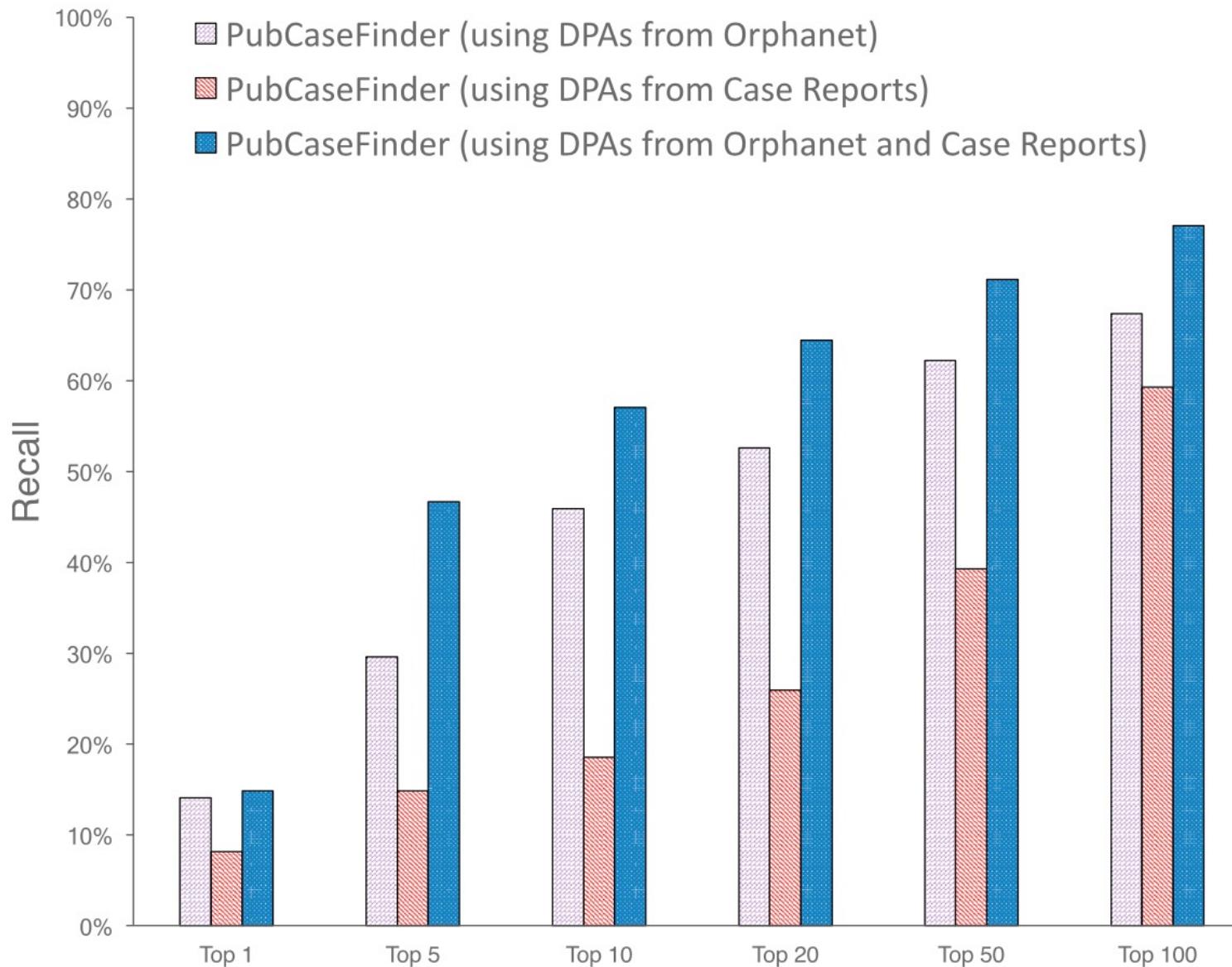
PubCaseFinder uses these sets of disease-gene associations

- MedGen: https://ftp.ncbi.nlm.nih.gov/gene/DATA/mim2gene_medgen
- Orphanet: http://www.orphadata.org/data/xml/en_product6.xml
- GenCC: <https://search.thegencc.org/download>

Fujiwara T et al., 2018 Am J Hum Genet
Fujiwara T et al., 2022 Hum Mut



CC BY 4.0 DEED



性能は
背景のデータに依存

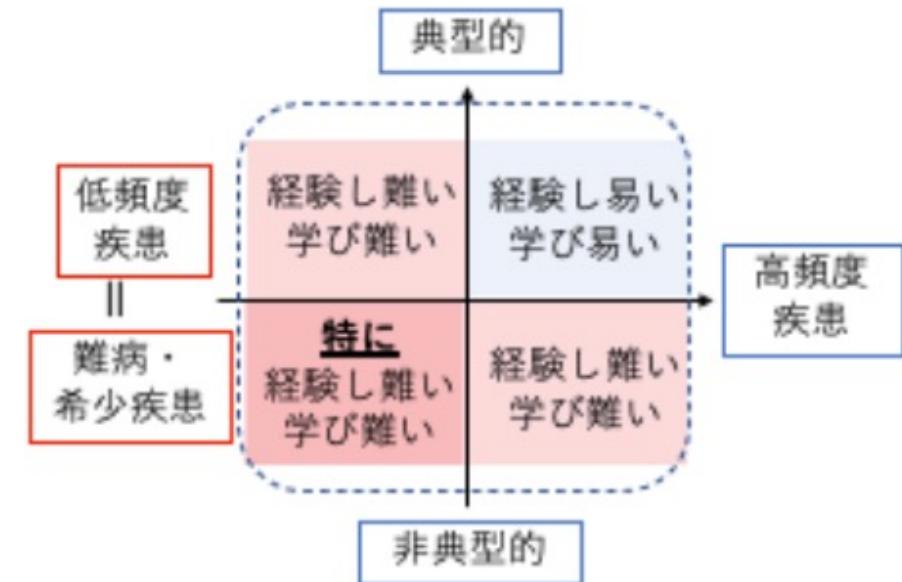
より詳細なデータを得るには？

患者さんや医療者にとって
わかりやすい形のためには？

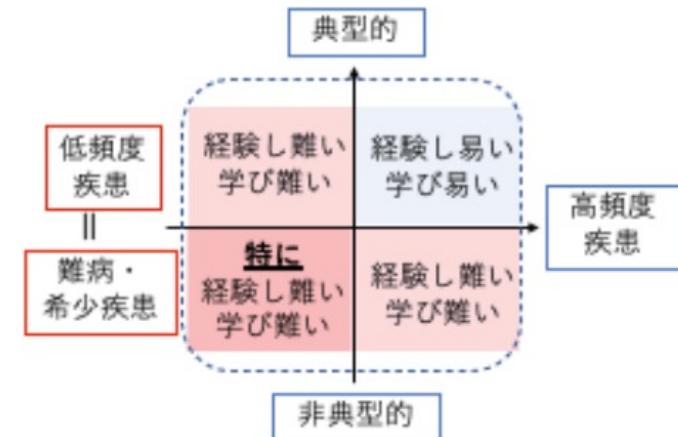
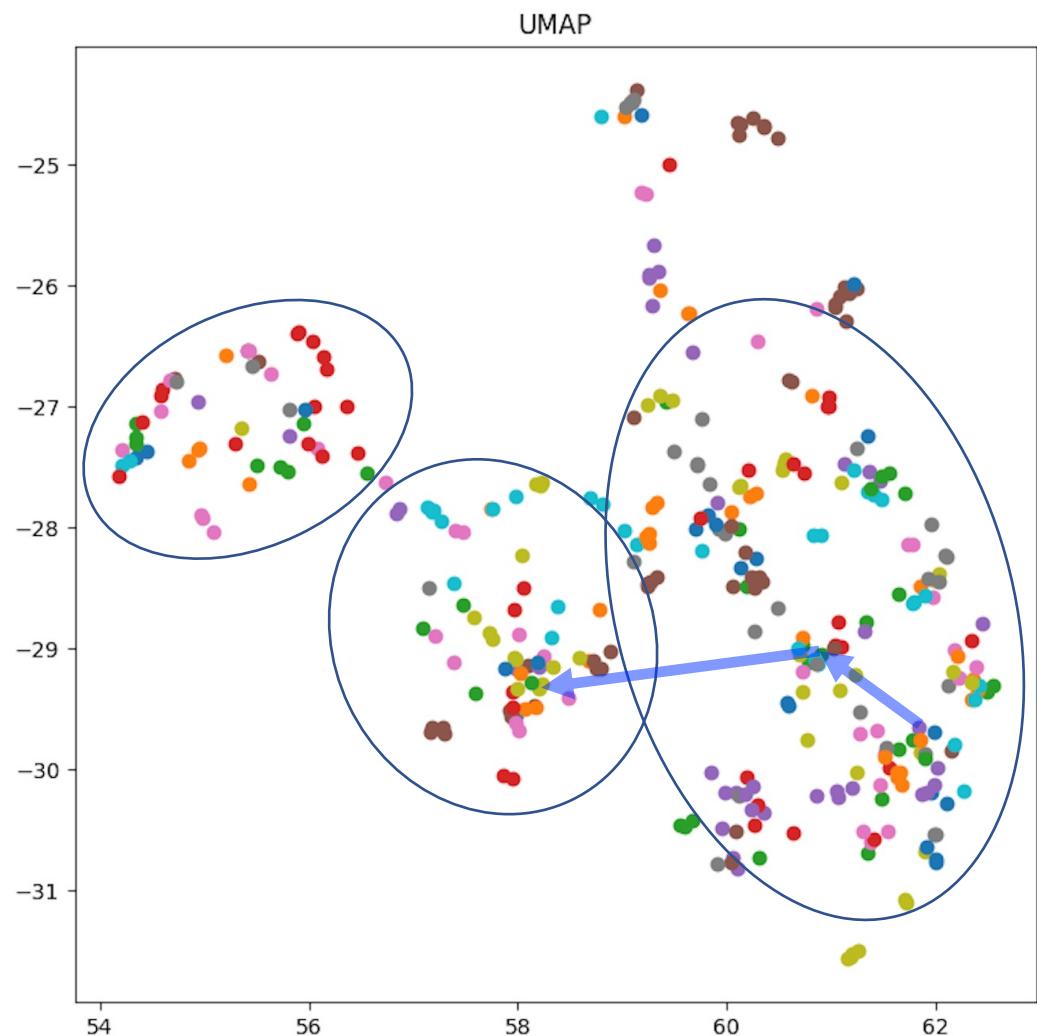
Fujiwara T et al., 2022 Hum Mut

一つの疾患の全体像から非典型例を含めて可視化したい！

- ・アレキサンダー病 英文症例報告200例
- ・症例データを構造化
- ・症状・所見を抽出
- ・アノテーション
- ・個々の症例をベクトル化
- ・目の前の患者さんが
“どれだけ稀（非典型的）なのか？”

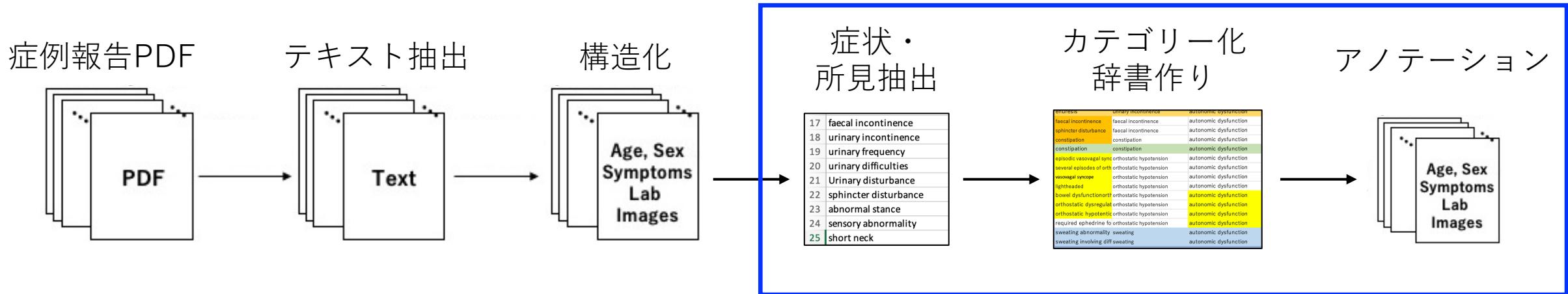


綺麗なアノテーションができれば

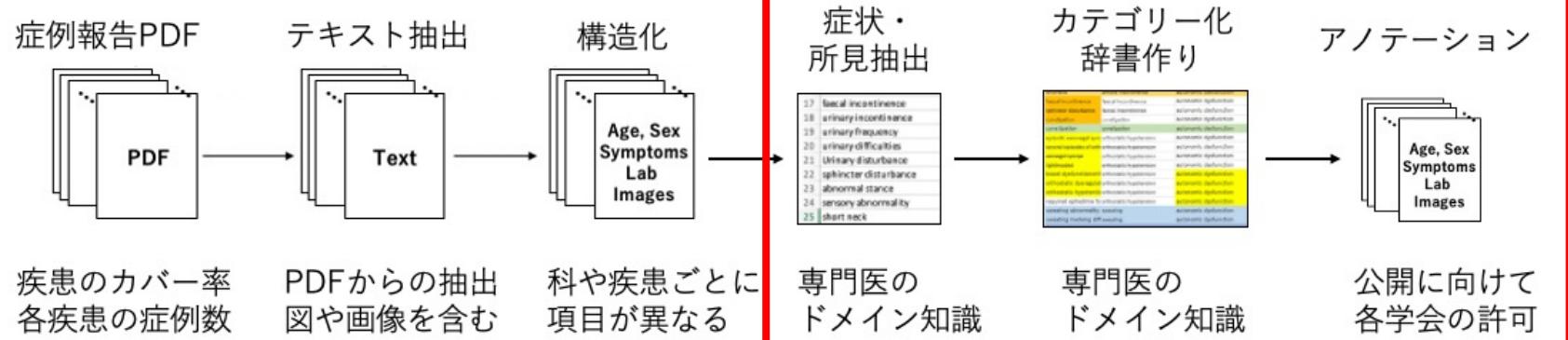


- ・アレキサンダー病 340例
(時系列データ含む)
- ・構造化データ
- ・統制されたカテゴリー辞書
- ・自由な可視化
- ・時系列の可視化
- ・**非典型例の推定**

日本語症例報告からどれだけ情報が集められるか？



CC BY 4.0 DEED



■ 表現型情報(HPO)の入力支援機能

PubCaseFinder

⑦ Guides ⑧ API ⑨ History ⑩ Terms ⑪ Datasets ⑫ RDF ⑬ Contact

抽出結果（アノテーション結果）

【全体像(1)】コントロール不良の高血圧(Hypertension)・糖尿病(Diabetes mellitus)で定期通院中の82歳女性の尿蛋白・全身浮腫。ネフローゼ症候群を中心とした腎疾患を疑い、腎生検を含めた精査目的での入院。

【主訴】全身倦怠感、全身浮腫

【現病歴(2)】入院1か月前より倦怠感が出現し徐々に増悪、2週間前に足のむくみに気付き、1週間前にはいつもの靴が履けなくなり前医受診。ネフローゼ症候群疑いで○月○日当院紹介入院となつた。週単位で増悪傾向の浮腫あり。圧痕性浮腫で、下肢に目立つが顔・上肢にも見られ全身性。靴下の跡がつくがすぐに消える。皮膚の熱感や疼痛の自覚なし。心不全症状なし。肝不全症状なし。先行感染症状なし。

【既往歴(3)】高血圧症(20年前に指摘され内服治療中、普段は150/80 mmHg前後)、糖尿病(15年前に指摘され内服治療中、普段はHbA1c 8%前後)、急性腎盂腎炎(5年前)、急性虫垂炎(65年前に虫垂切除術)。輸血歴なし、アレルギー歴なし、妊娠歴:3妊2産1流。

合計 20件(症状あり20件、症状なし0件)

- 高血压・高血压(HP:0000822)
- Hypertension・高血压(HP:0000822)
- 糖尿病・糖尿病(HP:0000819)
- Diabetes mellitus・糖尿病(HP:0000819)
- 尿蛋白・尿蛋白(HP:0000093)
- 全身浮腫・全身浮腫(HP:0012050)
- ネフローゼ症候群・ネフローゼ症候群(HP:0000100)
- 腎疾患・肾疾患(HP:0012622)
- 全身浮腫・全身浮腫(HP:0012050)
- むくみ・むくみ(HP:0000969)

アノテーションを保存

症状や所見のIDである、HPO

自動抽出が可能

しかし完璧では無い

臨床医の感覚にマッチしない。。。



CC BY 4.0 DEED

LLMを用いた症状抽出

DO

You

下記の文章から症状を抜き出してください

現病歴1歳7ヶ月頃までは順調に発育していたが、1歳9ヶ月頃より特異な手の動きが出てきた。はじめは手を開いて胸の前でうちあわせる動きであったが次第に手指を揃え、軽く第2指間関節で曲げた形で右手掌に左手をのせる様にたたく。そして時々左指を口へ持ってゆくという動きが出てきた。この頃より今までできていたものをつかむ、つまむ、指差し等しなくなり、いたずらも減り、言葉も出なくなり、意志表示が不明瞭になってきた。母親との視線も合わなくなり、周囲への関心も減り、表情が乏しくなり、笑わなくなってきた。種々病院を受診し相談するも、脳波に異常所見(左優位に棘波頻発)が認められた以外、血液・尿初見は全て正常範囲であった。3歳4ヶ月紹介され当クリニックを受診する。

症状の抽出は格段に向上了

HPDのID付けは不十分



ID付けが上手くいかないのか?
辞書 (HPD) に無いのか?

ChatGPTs4



HPO Visualizer

この文章に基づき、以下の症状が抽出されます：

1. 特異な手の動き - 1歳9ヶ月頃から現れる。
2. 手を開いて胸の前で打ち合わせる動き。
3. 手指を揃え、軽く第2指間関節で曲げる動き。
4. 右手掌に左手をのせるようにたたく動き。
5. 左指を口へ持っていく動き。
6. ものをつかむ、つまむ、指差し等の能力の喪失。
7. いたずらが減少。
8. 言葉の喪失。
9. 意志表示が不明瞭になる。
10. 母親との視線が合わなくなる。
11. 周囲への関心の減少。
12. 表情が乏しくなる。
13. 笑わなくなる。
14. 脳波異常 (棘波頻発、左優位)。

Llama2 (7b)

- 意志表示が不明瞭になった
- いたずらが減った
- 言葉が出なくなった
- 視線が合わなくなった
- 周囲への関心が減った
- 表情が乏しくなった
- 笑わなくなった

これはあくまで一意見でしかありません。

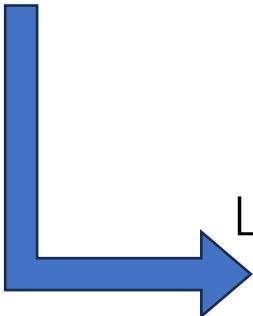


CC BY 4.0 DEED

Clinical Text



Exact match
ambiguous match



LLM (ChatGPT)

症状・所見の抽出：⇒ **非常に良い!**

HPOとのアノテーション… ChatGPT⇒ 非常に難しい…

PubCaseFinder

Guides API History Terms Datasets RDF Contact

抽出結果（アノテーション結果）

【全体像(1)】コントロール不良の高血圧(Hypertension)・糖尿病(Diabetes mellitus)で定期通院中の82歳女性の尿蛋白・全身浮腫。ネフローゼ症候群を中心とした腎疾患を疑い、腎生検を始めた精査目的で入院。
【主訴】全身倦怠感、全身浮腫
【現病歴(2)】入院1か月前より倦怠感が出現し徐々に増強。2週間前に足のむくみに気付き、1週間前にはいつもの靴が履けなくなり前医受診。ネフローゼ症候群疑いで5月5日当院紹介入院となった。週単位で増悪傾向の浮腫あり。圧痕性浮腫で、下肢に目立つが面・上肢にも見られ全身性。靴下の跡がつくがすぐに消える。皮膚の熱感や疼痛の自覚なし。心不全症状なし。肝不全症状なし。先行感染症なし。
【既往歴(3)】高血圧症(20年前に指摘され内服治療中、普段は150/80 mmHg前後)、糖尿病(15年前に指摘され内服治療中、普段はHbA1c 8%前後)、急性腎盂腎炎(5年前)、急性虫垂炎(65年前に虫垂切除術)。輸血歴なし、アレルギー歴なし、妊娠歴:3姉2産1流。

合計 20件(症状あり20件、症状なし0件)

- 高血圧・高血圧(HP:0000822)
- Hypertension・高血圧(HP:0000822)
- 糖尿病・糖尿病(HP:0000819)
- Diabetes mellitus・糖尿病(HP:0000819)
- 尿蛋白・尿蛋白(HP:0000093)
- 全身浮腫・全身浮腫(HP:0012050)
- ネフローゼ症候群・ネフローゼ症候群(HP:0000100)
- 腎疾患・腎疾患(HP:0012622)
- 全身浮腫・全身浮腫(HP:0012050)
- むくみ・むくみ(HP:0000069)

アノテーションを保存



You

下記の文章から症状を抜き出してください

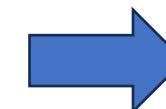
現病歴1歳7ヶ月頃までは順調に発育していたが、1歳9ヶ月頃より特異な手の動きが出てきた。はじめは手を開いて胸の前でうちあわせる動きであったが次第に手指を揃え、軽く第2指間関節で曲げた形で右手掌に左手をのせる様にたたく。そして時々左指を口へ持ってゆくという動きが出てきた。この頃より今までできていた、ものをつかむ、つまむ、指差し等しなくなり、いたずらも減り、言葉も出なくなり、意志表示が不明瞭になってきた。母親との視線も合わなくなり、周囲への関心も減り、表情が乏しくなり、笑わなくなってきた。種々病院を受診し相談するも、脳波に異常所見(左優位に棘波頻発)が認められた以外、血液・尿初見は全て正常範囲であった。3歳4ヶ月紹介され当クリニックを受診する。



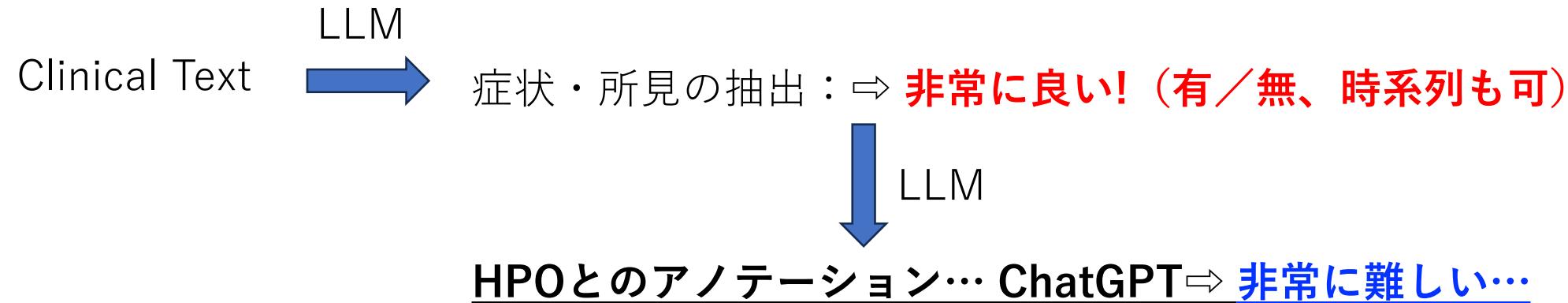
HPO Visualizer

この文章に基づき、以下の症状が抽出されます：

- 特異な手の動き - 1歳9ヶ月頃から現れる。
- 手を開いて胸の前で打ち合わせる動き。
- 手指を揃え、軽く第2指間関節で曲げる動き。
- 右手掌に左手をのせるようにたたく動き。
- 左指を口へ持っていく動き。
- ものをつかむ、つまむ、指差し等の能力の喪失。
- いたずらが減少。
- 言葉の喪失。
- 意志表示が不明瞭になる。
- 母親との視線が合わなくなる。
- 周囲への関心の減少。
- 表情が乏しくなる。
- 笑わなくなる。
- 脳波異常 (棘波頻発、左優位)。



CC BY 4.0 DEED

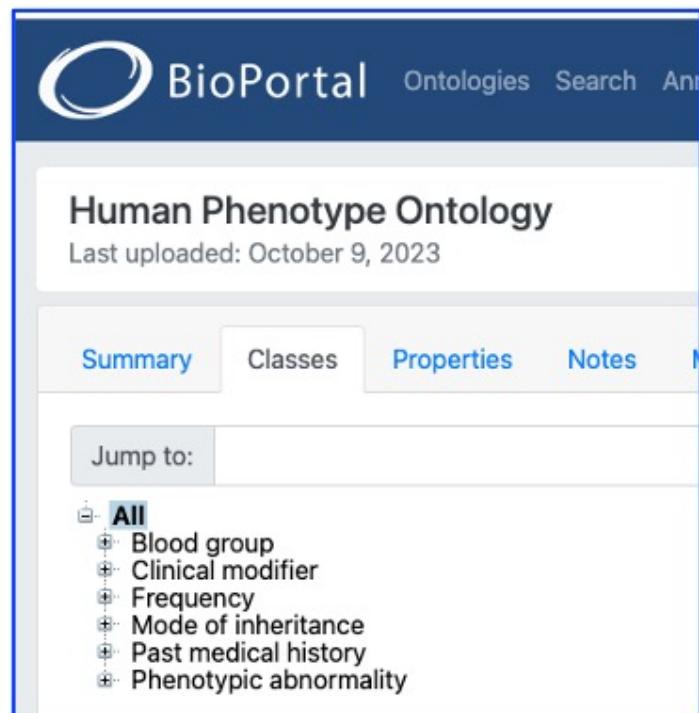


おそらく2つの理由

- 1: アノテーション機能の不足
- 2: そもそも、HPOに対応する語彙が無い？

HPOを検索・探索しようとするが・・・

視認性が非常に悪い
解剖や生理、システムなどに
全く準拠していない



BioPortal Ontologies Search Analytics

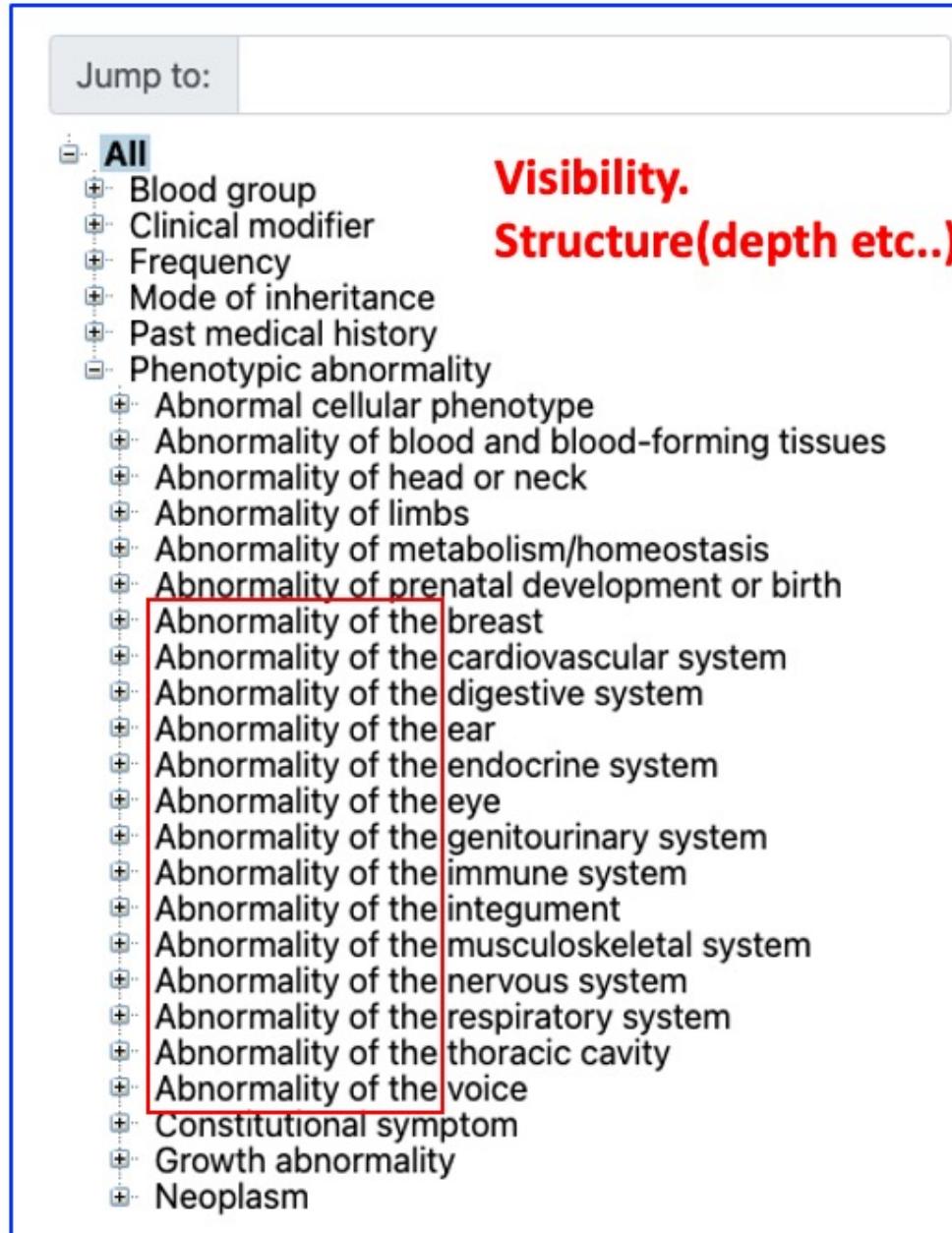
Human Phenotype Ontology

Last uploaded: October 9, 2023

Summary Classes Properties Notes

Jump to:

- All
- Blood group
- Clinical modifier
- Frequency
- Mode of inheritance
- Past medical history
- Phenotypic abnormality



Jump to:

- All
- Blood group
- Clinical modifier
- Frequency
- Mode of inheritance
- Past medical history
- Phenotypic abnormality
- Abnormal cellular phenotype
- Abnormality of blood and blood-forming tissues
- Abnormality of head or neck
- Abnormality of limbs
- Abnormality of metabolism/homeostasis
- Abnormality of prenatal development or birth
- Abnormality of the breast
- Abnormality of the cardiovascular system
- Abnormality of the digestive system
- Abnormality of the ear
- Abnormality of the endocrine system
- Abnormality of the eye
- Abnormality of the genitourinary system
- Abnormality of the immune system
- Abnormality of the integument
- Abnormality of the musculoskeletal system
- Abnormality of the nervous system
- Abnormality of the respiratory system
- Abnormality of the thoracic cavity
- Abnormality of the voice
- Constitutional symptom
- Growth abnormality
- Neoplasm

Visibility.
Structure(depth etc..)

BioPortal

HPO

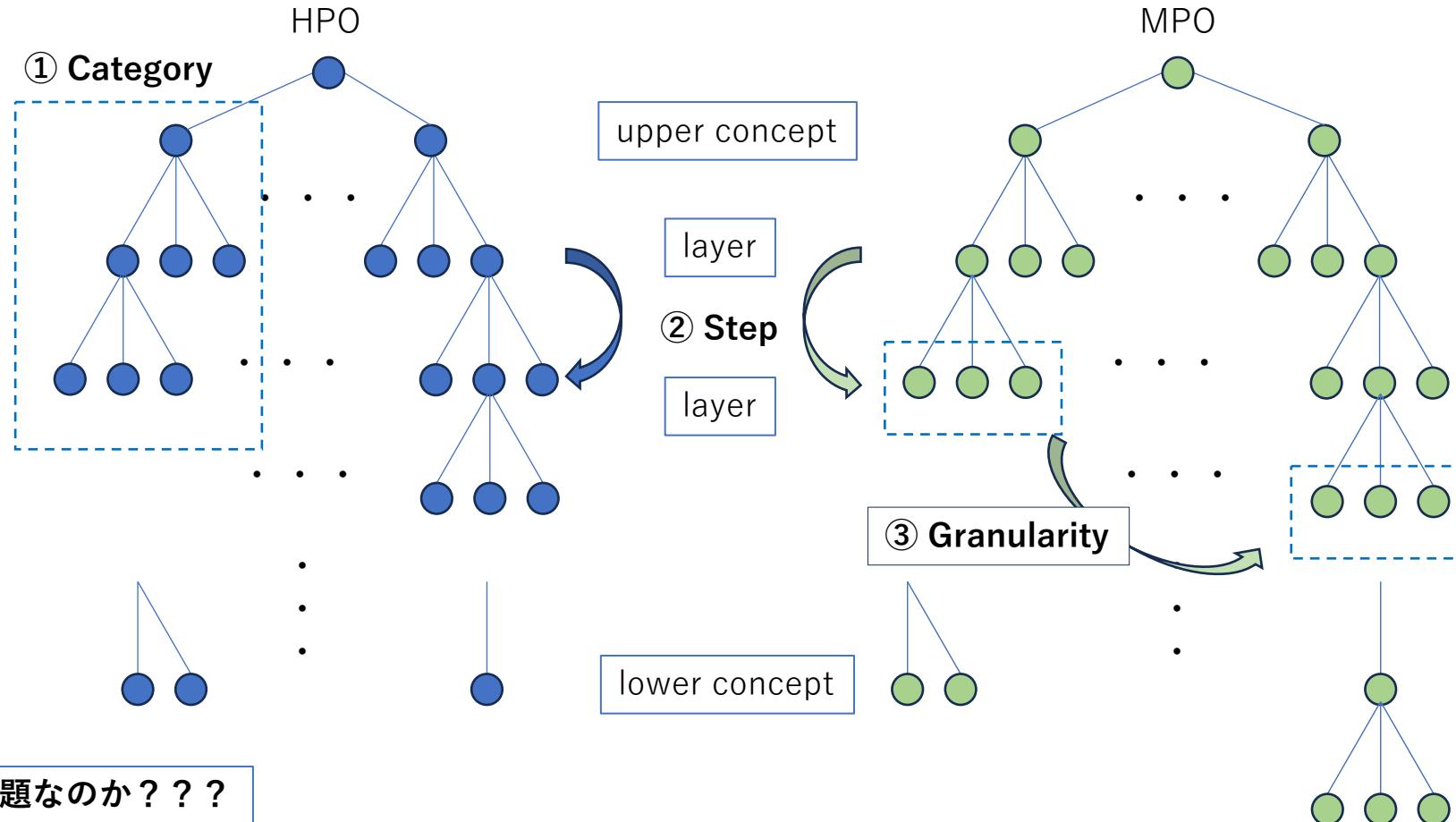
⇒Phenotypic abnormality

並びが臨床医や研究者に
優しくない (😅)



CC BY 4.0 DEED

臨床症状・所見の辞書である、HPO自体の構造化も工夫が必要



何が問題なのか???

- ① **カテゴリー**: 解剖（臓器）、システム（免疫、代謝etc）、時点など、… etc
- ② **階層**: カテゴリーの対象とするものの複雑性に依存して深さが異なる。
- ③ **粒度**: エンドフェノタイプを示すものが、別の階層に存在する。
⇒ これが、俯瞰的な見方や全体像を掴むことを難しくしている

- End-phenotype matching.
- What happened in each step

- ⇒ Checking with experts
- ⇒ To capture the concept of each term

N of Layer in HPO and MPO

	HPO	MPO
1	23	29
2	153	97
3	777	843
4	2340	2750
5	4410	4008
6	5303	4392
7	5467	3583
8	4877	2394
9	3659	1453
10	2604	828
11	1530	531
12	893	257
13	485	98
14	70	30
15	30	9
16	1	

Dohi E et al. *BioHachrXiv* 2024

HPOの整理や構造化（GTPsも活用）

Abnormality of the genitourinary system	1039
Abnormal cellular phenotype	191
Abnormality of blood and blood	570
Abnormality of head or neck	961
Abnormality of limbs	1688
Abnormality of metabolism/homeostasis	1219
Abnormality of prenatal development or birth	192
Abnormality of the breast	24
Abnormality of the cardiovascular system	957
Abnormality of the digestive system	496
Abnormality of the ear	213
Abnormality of the endocrine system	273
Abnormality of the eye	777
Abnormality of the immune system	871
Abnormality of the integument	713
Abnormality of the musculoskeletal system	2743
Abnormality of the nervous system	1586
Abnormality of the respiratory system	445
Abnormality of the thoracic cavity	4
Abnormality of the voice	21
Constitutional symptom	81
Growth abnormality	68
Neoplasm	435

HPOのPhenotypic abnormality
26カテゴリーの各リストを作成

解剖
病理所見（細胞、組織、etc）
疫学的情報

診察異常
症状
生化学・生理検査データ
画像検査データ

それぞれの語句が使用さ
れうる意味の整理
(構造化、タグ付など)

- LLM
- 複数の臨床医・専門家

にて行っている

MPO(mammalian phenotype ontology)とHPOとの取り組み

MPOの日本語訳作成の効率化

LLMのみでは、専門用語の翻訳は難しい・・・

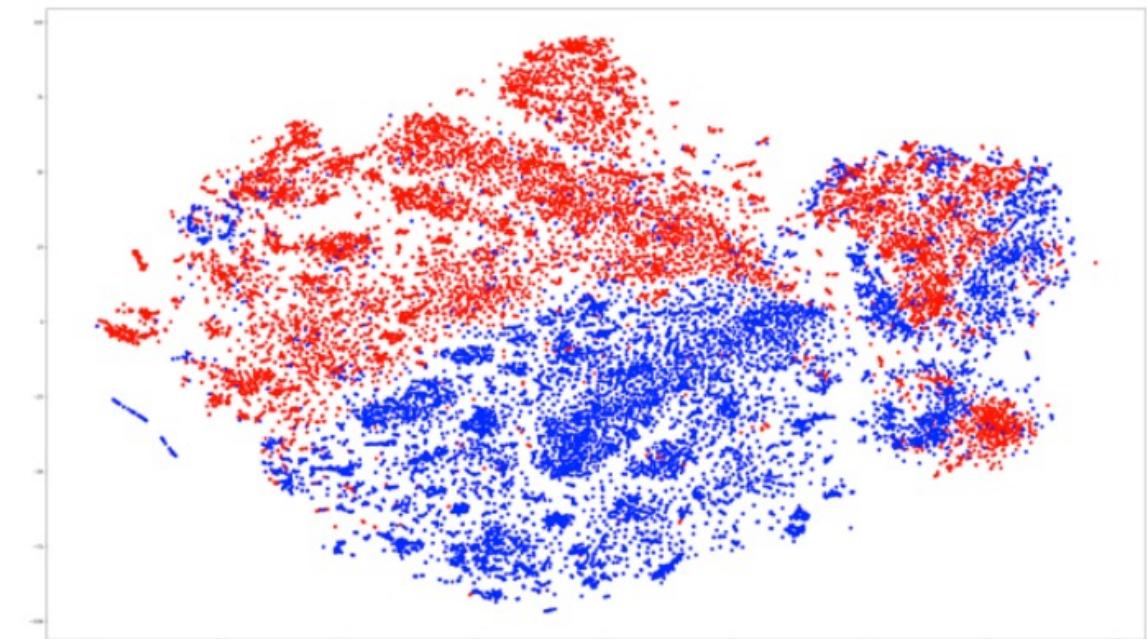
Custom GPTs

これまでの翻訳で作成した対応表を**RAG**として与えて、プロンプトを工夫すると

Custom ChatGPT “MP Japanese”

HPOとMPOのマッピングにおける課題

言葉の定義まで遡ってもマッピングは10%未満
それぞれのオントロジーの構造の整理が必要



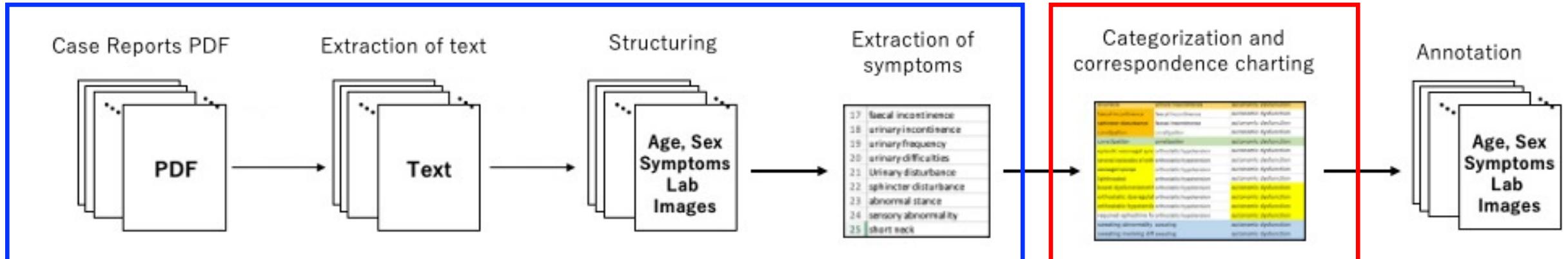
Takatsuki T et al. *Genomics and Informatics* (under revision)

https://github.com/dbcls/MP_Japanse/tree/main

Dohi F et al. *BioRxiv* 2024



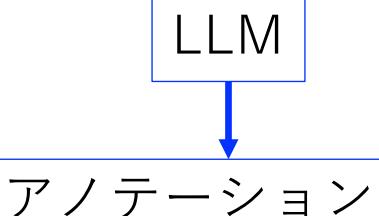
適切なアノテーションをHPOで行うためには・・・



自動化が進行中

Corresponding chart

Provided database (RAG: Retrieval augmented generation)



A 13-year-old boy exhibited cervical kyphosis ["muscle-skeletal"](#) and deteriorating cognition ["dementia"](#). He was born normally and achieved developmental milestones without delay. The family history was noncontributory. At the age of 2, he had 2 simple febrile seizure ["seizure"](#)s. At the age of 8, he had a generalized febrile convulsion that lasted about 10 min. Subsequent electroencephalography (EEG) revealed bilateral frontal spikes during sleep. Valproate was started under the diagnosis of epilepsy. Since then, his school performance slowly began to deteriorate ["psychomotor regression"](#).

適切な知識を与えれば、
全アノテーションが可能に

しかし、語句の対応だけでは、
臨床医が得る印象・内容とは異なる
⇒

工夫したオントロジー構造も提案？



CC BY 4.0 DEED

Acknowledgement

Research Organization of Information and system

Database Center for Life Science (DBCLS)

Toyofumi Fujiwara, Jae-Moon Shin, Yasunori Yamamoto

Terue Takatsuki, Jaemoon Shin, Thomas Liener

RIKEN Bio Resource Research Center

Hiroshi Masuya, Tatsuya Kushida, Yuki Yamagata

Japan Science and Technology Agency

Yuka Tateishi

King Abdullah University of Science and Technology

Robert Hoehndorf

Patient Salon, Patient Net

Nobuyuki Suzuki, Tomomi Yoshida

Tokyo Medical and Dental University

Katsuyuki Takeuchi, Takeshi Hase, Yoritsugu Iida



BioHackathon 2023



Parvez MD SOLWER ALAM

National center of neurology and psychiatry

Department of Mental Disorder

Takeshi Iwatsubo, Takahiro Kochi, Parvez MD SOLWE ALAM

Translational Medical Center

Kenji Hatano, Kotaro Hattori, Ryo Matsumura

Department of Sleep-Wake Disorders

Shingo Kitamura



1st Undiagnosed Hackathon 2023



ISEV IOC 2024



CC BY 4.0 DEED