

○藤原豊史¹、山口敦子²

1.情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター、2.東京都市大学 教育開発機構数理 データサイエンス教育センター

- 日本医療研究開発機構（AMED）が主導する未診断疾患イニシアチブ（IRUD）は、日本全国の未診断患者に対してゲノム解析をすることで、診断しようとする試み

- 3,634家系の全エクソン解析を完了(2019年9月末)

- ▶ 1,593家系(43.8%)で診断が確定

出所：医学のあゆみ，未診断疾患イニシアチブ（IRUD）の成果，
273巻7号



50～60%の未診断患者はゲノム解析を用いても診断が見つからない



従来通り，症状の組み合わせなどから診断を試みる必要がある

- しかし，疾患数の多さに比較して各疾患の頻度が低い中で，一人の医療者が十分な症例数の診療経験を積むことは困難であり，症状の組み合わせから診断に至るのは容易ではない

出所：小崎健次郎，ゲノム医療研究開発の稀少疾患領域における国内外の動向，ヒト疾患のデータベースとバイオバンク，羊土社，Vol. 35, No. 17, pp. 12-15(2017)



症状の組み合わせから関連性の高い疾患を検索し診断を支援するシステムPubCaseFinderを開発し2018年からウェブサービスとして公開

出所：Toyofumi Fujiwara, Yasunori Yamamoto, Jin-Dong Kim, Orion Buske, and Toshihisa Takagi., PubCaseFinder: A Case-Report-Based, Phenotype-Driven Differential-Diagnosis System for Rare Diseases, The American Journal of Human Genetics 103.3 (2018): 389-399.

○藤原豊史¹、山口敦子²

1.情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター、2.東京都市大学 教育開発機構数理 データサイエンス教育センター

□ PubCaseFinderで最も重要なのは、ランキング上位に診断候補を正確に提示する能力

□ PubCaseFinderのランキング生成には、類似度計算手法GeneYentaを採用してきたが、入力症状が不正確である場合や誤りがある場合に、その影響を強く受けるという問題があった

出所：Gottlieb, M. M., et al.: GeneYenta: A phenotype based rare disease case matching tool based on online dating algorithms for the acceleration of exome interpretation, Hum. Mutat., Vol. 36, No. 4, pp. 432-438(2015)

出所：Robinson PN, Ravanmehr V, Jacobsen JOB, et al. Interpretable Clinical Genomics with a Likelihood Ratio Paradigm [published online ahead of print, 2020 Jul 29]. Am J Hum Genet. 2020;S0002-9297(20)30230-5. doi:10.1016/j.ajhg.2020.06.021



入力症状に含まれるノイズに対してよりロバストな手法にGeneYentaを改良した

○藤原豊史¹、山口敦子²

1.情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター、2.東京都市大学 教育開発機構数理 データサイエンス教育センター

- 新たな手法は3つのパラメータ $E(>1)$, $K(>1)$, $W(<1)$ を使い、GeneYentaを以下のように変更した。

T_{HPO} : a set of terms in HPO

$D = \{d \mid d \subset T_{HPO}\}$: a set of diseases

INPUT: $c \subset T_{HPO}$

OUTPUT: $SIM(c, d)$ for all d in D

$$P(t) = \frac{|annot_t|}{|annot_{all}|}, IC_t = -\log P(t)$$

$$sim(t, t') = \max_{a_t \in A_t \cap A_{t'}} IC_{a_t}$$

$$C(c, d) = \{t \in c \mid IC_t - \max_{t' \in d} sim(t, t') < E\}$$

$$R_t = \begin{cases} 1 & |C(c, d)| < K \text{ or } IC_t - \max_{t' \in T_d} sim(t, t') < E \\ W & \text{otherwise} \end{cases}$$

$$SIM(c, d) = \frac{\sum_{t \in c} R_t \cdot \max_{t' \in d} sim(t, t')}{\sum_{t \in c} R_t \cdot IC_t}$$

- 以下の方法で作成したデータセットを用いてグリッドサーチを行いパラメータを最適化した

- OMIMの遺伝性疾患の中から、ランダムに500疾患を選択
- それぞれの疾患に対するHPOアノテーションについて、以下の処理を行う
 - ✓ 5~10件のHPOアノテーションをランダムに残す
 - ✓ ノイズとなるHPOアノテーションを1~3件ランダムに追加
 - ✓ ランダムにHPOアノテーションを選択し、1つ上の親ノードに変換
 - ✓ ランダムにHPOアノテーションを選択し、2つ上の親ノードに変換

○藤原豊史¹、山口敦子²

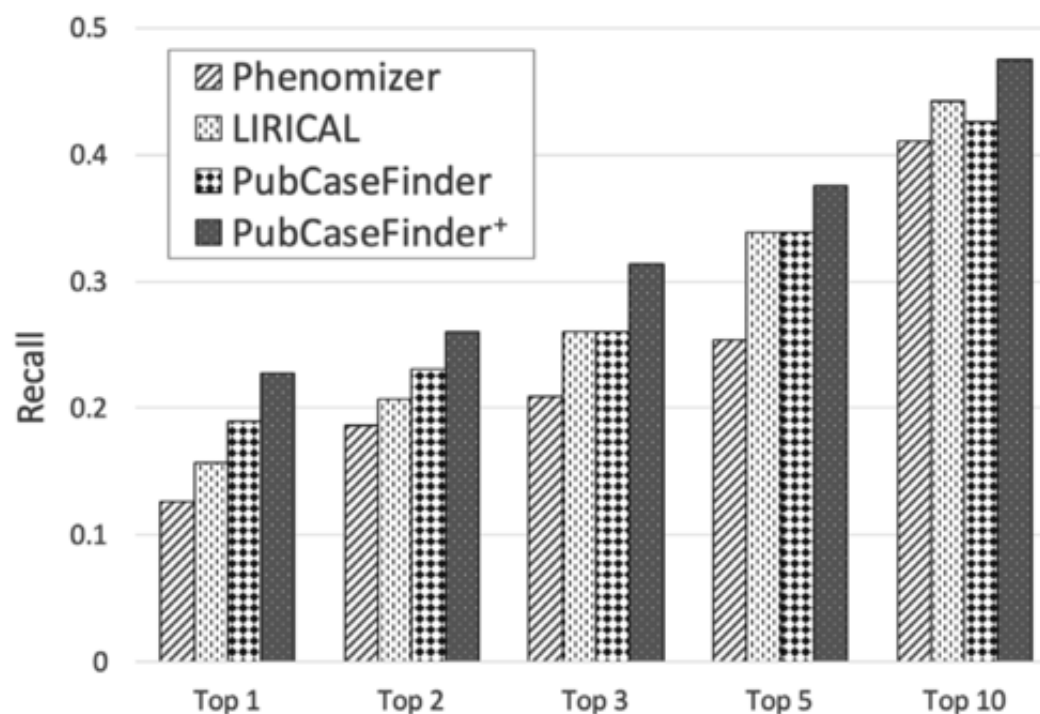
1.情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター、2.東京都市大学 教育開発機構数理 データサイエンス教育センター

□ カナダのCare4Rareから提供された243件の症例データセット（症状+診断結果）を用いて、複数のシステムの疾患ランキング精度を比較

- Phenomizer：HPOリリース（2009年）と同時に発表され、よく利用されているシステム
- LIRICAL：最も精度が高いと2020年に発表されたシステム
- PubCaseFinder：GeneYentaを採用
- PubCaseFinder⁺：GeneYentaの改良手法を採用

Phenomizer：[https://www.cell.com/ajhg/fulltext/S0002-9297\(09\)00399-1](https://www.cell.com/ajhg/fulltext/S0002-9297(09)00399-1)LIRICAL：[https://www.cell.com/ajhg/fulltext/S0002-9297\(20\)30230-5](https://www.cell.com/ajhg/fulltext/S0002-9297(20)30230-5)

PubCaseFinder⁺の精度が最も高いことを確認



○藤原豊史¹、山口敦子²

1.情報・システム研究機構 データサイエンス共同利用基盤施設 ライフサイエンス統合データベースセンター、2.東京都市大学 教育開発機構数理 データサイエンス教育センター

【遺伝子ランキング機能】

患者の徴候・症状を入力 + Upload File (HPO ID):

疾患を絞り込む + Upload File (Gene Symbol or Entrez Gene ID or Ensembl Gene ID):

共有・ダウンロード 症状入力支援 クリア 疾患を検索

遺伝性疾患 (OMIM) 希少疾患 (Orphanet) 疾患原因遺伝子 公開症例

合計: 4312件 (表示件数)

順位 (類似度)	遺伝子	対応する徴候・症状	疾患名	遺伝形式
1 (100.0%)	DHOOD	上口唇裂 小顎 尺骨低形成 眼瞼裂 短い母指	軸性性肢端鼻骨症 (OMIM) 軸性性肢端鼻骨症 (ORDO)	常染色体劣性遺伝
1 (100.0%)	SEMA3E	上口唇裂 小顎 尺骨低形成 眼瞼裂 短い母指	低コチドトロピン性性腺機能低下症74/A 無味覚 (OMIM) CHARGE 症候群 (OMIM) CHARGE 症候群 (ORDO)	常染色体劣性遺伝 常染色体劣性遺伝 孤児性
1 (100.0%)	CHD7	上口唇裂 小顎 尺骨低形成 眼瞼裂 短い母指		

【N-of-1症例検索機能】

患者の徴候・症状を入力 + Upload File (HPO ID):

疾患を絞り込む + Upload File (Gene Symbol or Entrez Gene ID or Ensembl Gene ID):

共有・ダウンロード 症状入力支援 クリア 疾患を検索

遺伝性疾患 (OMIM) 希少疾患 (Orphanet) 疾患原因遺伝子 公開症例

合計: 1893症例 (表示件数)

順位 (類似度)	症例	対応する徴候・症状	対応する遺伝子
1 (80.4%)	DECIPHER:276525	腕骨低形成 第1中手骨無形成/短形成 連続眉毛 長い人中	
2 (80.4%)	MyGene2:831	下顎欠損 小顎 腕尺骨癒合 短い母指 軟口蓋裂	
3 (80.0%)	DECIPHER:275487	幅広い母指 平坦な人中 眼瞼裂斜下 短い母指 短い足	
4 (79.3%)	DECIPHER:304548	下顎突出 内眼角贅皮 短い第5指 薄い上口唇唇紅部	
5 (79.1%)	DECIPHER:283150	下顎後退 口蓋裂 尺骨低形成 軸性性肢端鼻症	

【今後】

- ゲノム解析における症状を用いた未知バリエーション評価のために、新たにリリースした「遺伝子ランキング機能」および「N-of-1症例検索機能」を強化する。