

ライフサイエンス統合データベースセンター
申在紋東京医科歯科大学 難治疾患研究所
宮冬樹
ライフサイエンス統合データベースセンター
藤原豊史東京都市大学
山口敦子

Abstract

次世代シーケンサーを用いた遺伝性疾患の診断では、バリエーションのフィルタリング・解釈に専門家の労力を必要とする。特に、患者の疾患原因となる遺伝子を複数の候補遺伝子の中から決定する際、その根拠となる文献やデータを検索するために多くの時間を要する。そこで近年では、同プロセスを支援するために、患者の症状と関連が高い順に疾患原因候補遺伝子をランキングするシステムが開発され、活用されている。これらのシステムは、文献から人手で抽出した遺伝子-症状関連データを用いている。

本研究では、遺伝子-症状関連データを拡張することで、より精度の高いシステムを構築することを目指した。手法としては、PubMedに含まれる全文献から自動抽出した遺伝子-疾患関連データを活用することで、遺伝子-症状関連データを拡張した。さらに同データを用いて患者症状に基づく疾患原因候補遺伝子ランキングシステムを開発した。同システムをPubCaseFinder (<https://pubcasefinder.dbcls.jp>) に組み込み、18件の臨床症例を用いて、PubCaseFinder、AMELIE、Phenolyzer、PhenoMatcherの精度を比較した。その結果、PubCaseFinderがこれらシステムの中でも最も高い精度を示すことを確認した。本発表では、比較結果の詳細と今後について報告したい。

背景

- 次世代シーケンサーを用いた遺伝性疾患の診断では、バリエーションのフィルタリング・解釈に専門家の労力を必要とする。
- 特に、患者の症状を用いてバリエーションをフィルタリングする際に（図1）、根拠となる文献やデータを検索するために多くの時間が必要となる。
- 近年では患者の症状と関連が高い順に疾患原因候補遺伝子をランキングするシステムが開発され、上記のバリエーションのフィルタリングに活用されている。

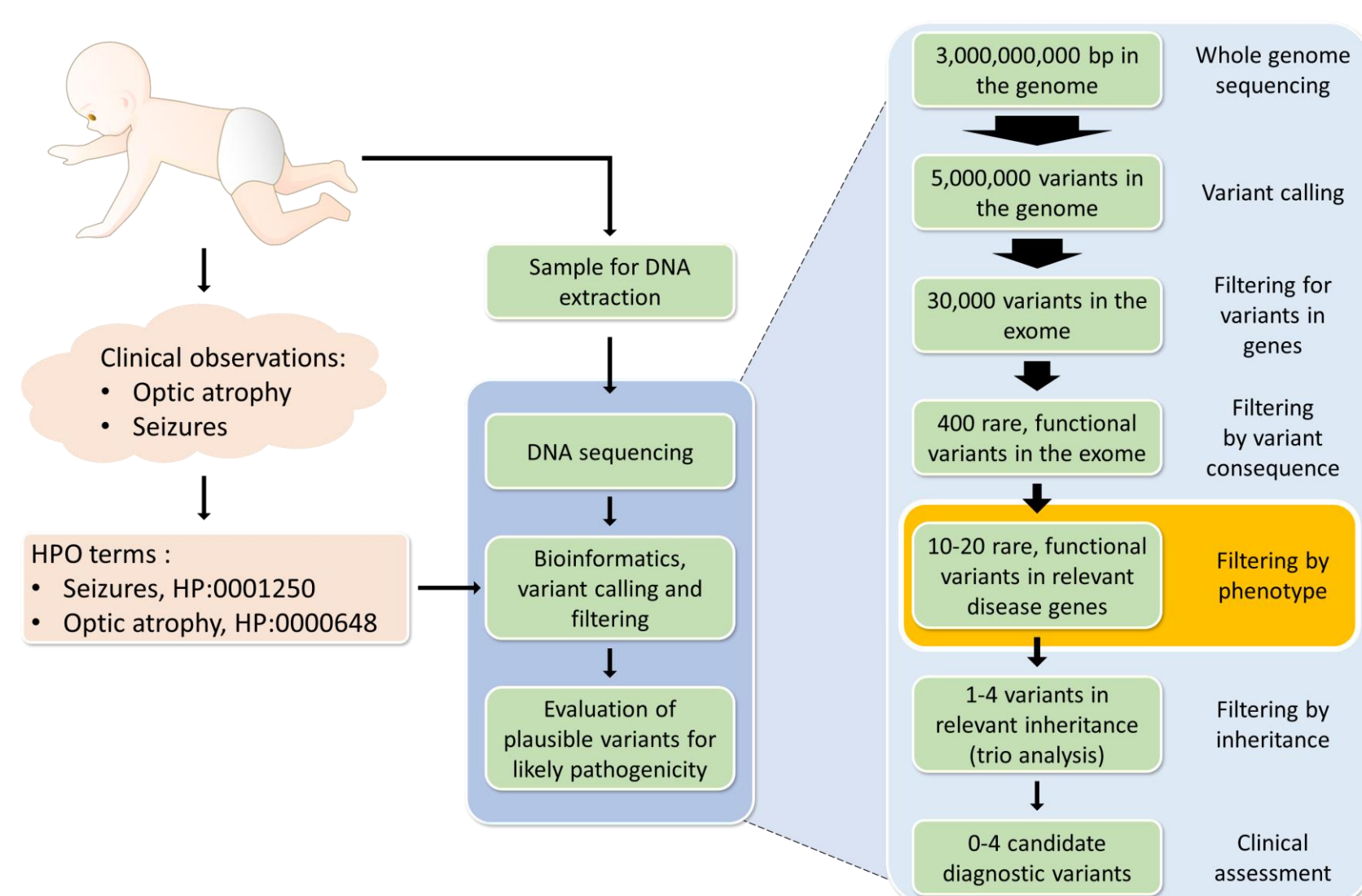


図1. エクソーム解析におけるバリエーションのフィルタリング例

システムの構築

- 患者症状に基づく疾患原因候補遺伝子ランキング機能をPubCaseFinderシステム上に開発した（図2）。
- 上位にランキングされた遺伝子が疾患原因候補遺伝子として可能性が高いと考えられる。

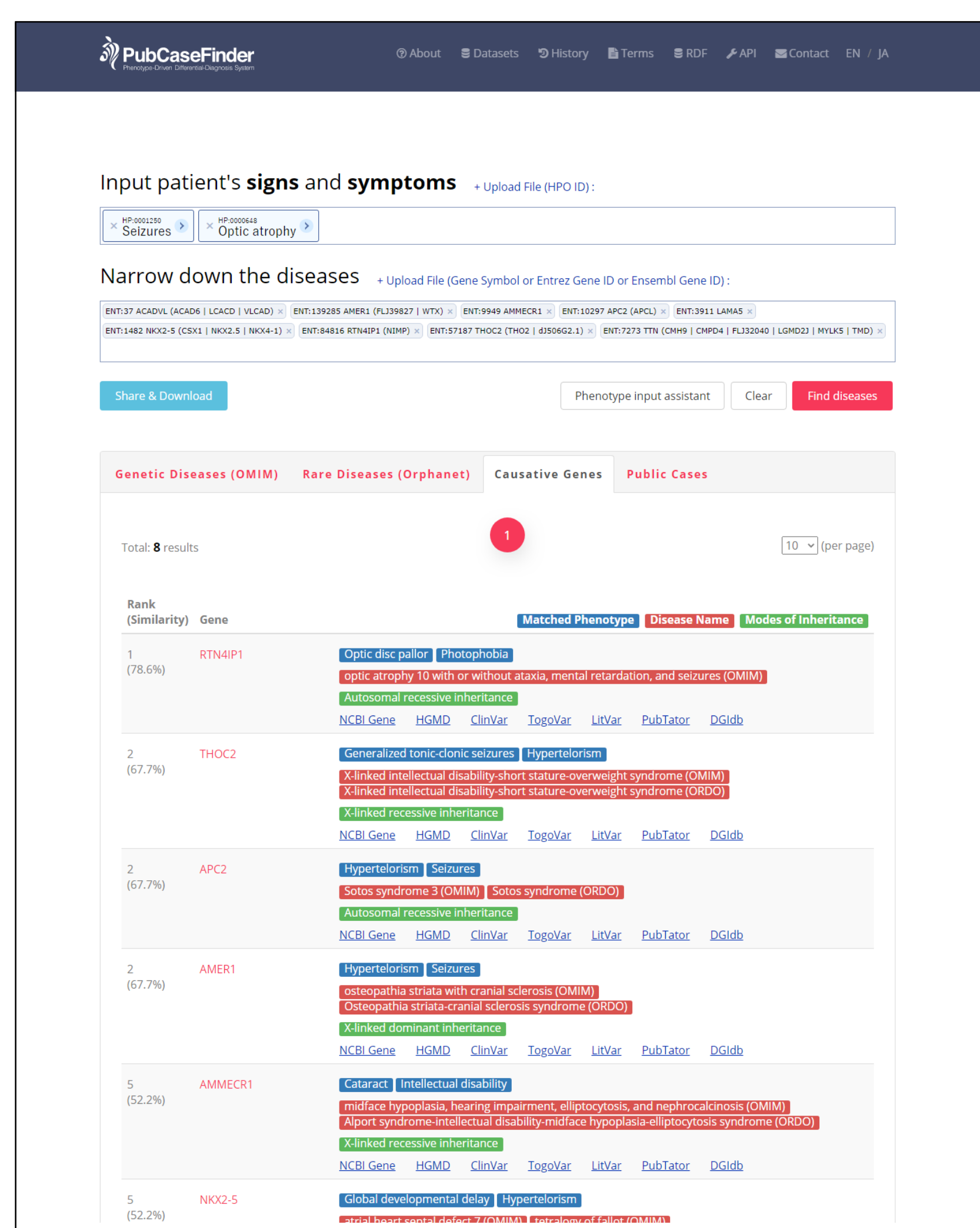


図2. PubCaseFinderの疾患原因候補遺伝子ランキング機能

評価用データセット

性別	年齢	臨床所見	臨床所見 (HPO)	解析方法	解析バリエーション後の候補バリエーション数	論文で報告した疾患原因遺伝子数	報告論文	PubCaseFinder	Phenolyzer	AMELIE	PhenoMatcher
								論文報告遺伝子の順位	論文報告遺伝子の順位	論文報告遺伝子の順位	論文報告遺伝子の順位
男	3歳	小脳低形成	HP:0001321,小脳低形成	エクソーム解析 (Trio)	34	HNRNP2	AHGCに海外の研究者が該当遺伝子の報告。2016	1	11	5	3
男	13歳	てんかん 視神経萎縮	HP:0001250,発作 HP:0000648,視神経萎縮	エクソーム解析 (Trio)	26	RTN4P1	Okamoto et al. J. Hum. Genet. 62, 927-929 (2017).	1	12	1	1
男	10歳	皮膚弛緩症	HP:000973,弛緩性皮膚	エクソーム解析 (Trio)	27	ALDH18A1	Nozaki et al. Brain Dev. 38, 678-684 (2016).	1	1	2	1
男	5歳	白質異常 皮質異常	HP:0002500,大脳白質の異常 HP:0100704,Cortical visual impairment	エクソーム解析 (Trio)	36	KIF1A	Okamoto N. et al. J. Hum. Genet. 59, 639-641 (2014).	2	1	1	1
男	1歳	低身長 精神運動発達障害 骨質異常	HP:0004322,低身長 HP:0002361,精神運動発達障害 HP:0009294,骨格の異常	エクソーム解析 (Trio)	16	FAM20C	Tamai et al. Am. J. Med. Genet. A, 176, 682-686 (2018).	4	7	3	1
女	1歳11か月	macrocephaly cortical dysplasia	HP:0002526,Macrocephaly HP:0002539,Cortical dysplasia	エクソーム解析 (Trio)	36	AKT3	Negishi Y. et al. BMC Med. Genet. 18, 4 (2017).	1	1	2	1
女	2歳	水腫脳症 小脳低形成 脳幹低形成	HP:0002324,水腫脳症 HP:0001321,小脳低形成 HP:0002365,脳幹低形成	エクソーム解析 (Trio)	18	TUBA1A	Yokoi S. et al. Sci. Rep. 5, 15165 (2015).	1	1	1	1
女	15歳	多小脳回 異所性灰白質	HP:0002126,多小脳回 HP:0002281,異所性灰白質	エクソーム解析 (Trio)	22	GNAI2	Hamada N. et al. J. Neurochem. 140, 82-95 (2017).	6	1	1	6
女	2歳	小頭症 脳室周囲性灰白質 小脳球萎縮	HP:0000252,小頭症 HP:0007165,脳室周囲性灰白質 HP:0000568,小脳球萎縮	エクソーム解析 (Trio)	21	PLK4	Tsutsumi M. et al. Eur. J. Hum. Genet. 24, 1702-1706 (2016).	1	3	1	1
男	4歳	半球機能異常 脳梁欠損	HP:0002079,脳梁欠損 HP:0001274,脳梁欠損	エクソーム解析 (Trio)	34	NFIA	Negishi Y. et al. Hum. Genome Var. 2, 15007 (2015).	1	1	1	1
女	NA	てんかん 知的障害	HP:0001250,発作 HP:0001249,知的障害	エクソーム解析 (Trio)	3	EPG5	Hori et al. Sci. Rep. 14, 3552 (2017).	1	1	1	1
女	NA	多小脳回 異所性灰白質	HP:0002126,多小脳回 HP:0002281,異所性灰白質	エクソーム解析 (Trio)	23	NEDD4L	Kato et al. J. Hum. Genet. 62, 861-863 (2017).	1	2	1	1
女	NA	小頭症 全身長 白質異常	HP:0000252,小頭症 HP:0004322,低身長 HP:0000712,白質異常	エクソーム解析 (Trio)	16	ASXL3	Hori I. et al. Am. J. Med. Genet. A, 170, 1863-1867 (2016).	1	9	1	2
女	NA	小頭症 てんかん性脳症	HP:0000252,小頭症 HP:0200134,てんかん性脳症	エクソーム解析 (Trio)	46	STAMBP	Hori et al. J. Hum. Genet. 63, 957-963 (2018).	1	9	5	3
男	NA	巨脳症	HP:0001355,巨脳症	エクソーム解析 (Trio)	34	MYCN	Kato et al. J. Med. Genet. 56, 388-395 (2019).	1	12	2	2
女	NA	小頭症	HP:0000252,小頭症	エクソーム解析 (Trio)	5	ASPM	F. Miya et al. Sci. Rep. 5, 9331 (2015).	1	1	1	1
女	NA	脳梁欠損	HP:0001274,脳梁欠損	エクソーム解析 (Trio)	38	EPG5	Hori et al. Sci. Rep. 14, 3552 (2017).	2	1	1	1
男	NA	水頭症	HP:0000238,水頭症	エクソーム解析 (Trio)	25	TUBA1A	Yokoi S. et al. Sci. Rep. 5, 15165 (2015).	1	4	1	1

表1. 18件の臨床症例(症状+疾患原因候補バリエーション)

評価

- 東京医科歯科大学 宮先生から提供された18件の臨床症例を用いて（表1）、PubCaseFinder、AMELIE、Phenolyzer、PhenoMatcherのランキング精度を比較した。

評価方法

- PubCaseFinder、AMELIE、Phenolyzer、PhenoMatcherのそれぞれに、症状と疾患原因候補遺伝子（フィルタリング後のバリエーションを変換したもの）を入力し、疾患原因候補遺伝子ランキングを取得する（図3）。

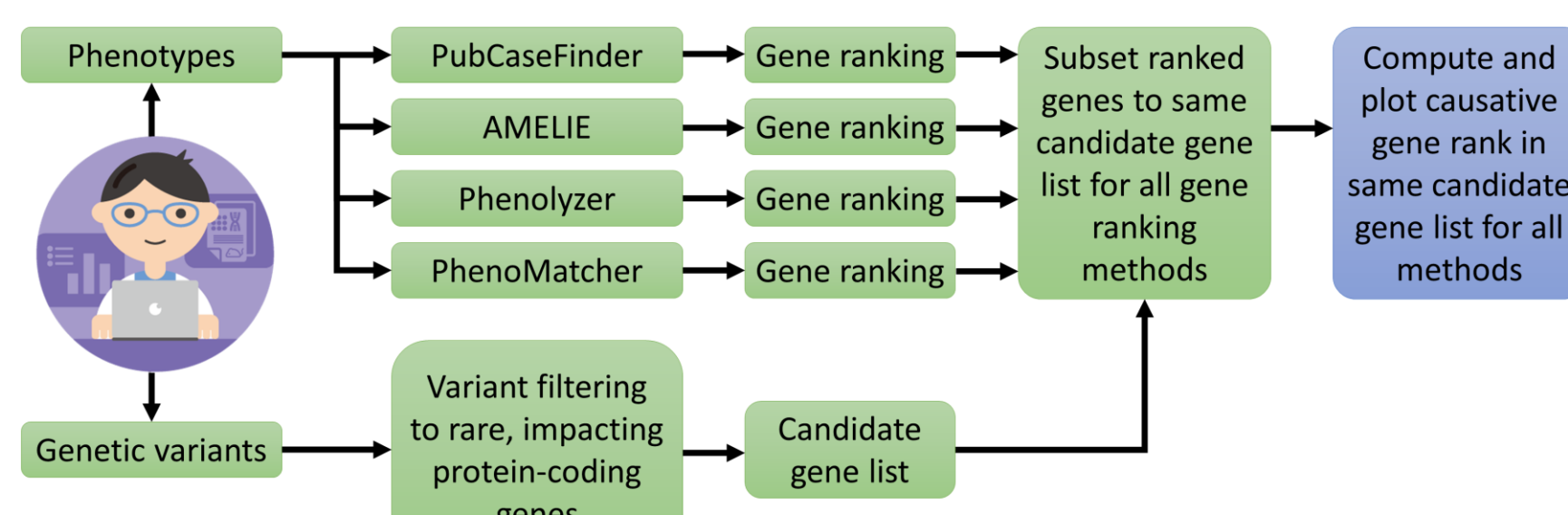


図3. 疾患原因候補遺伝子ランキングシステムの評価フロー

評価結果

- PubCaseFinderがこれらシステムの中で、ランキング上位の精度が最も高いことを示した（図4）。

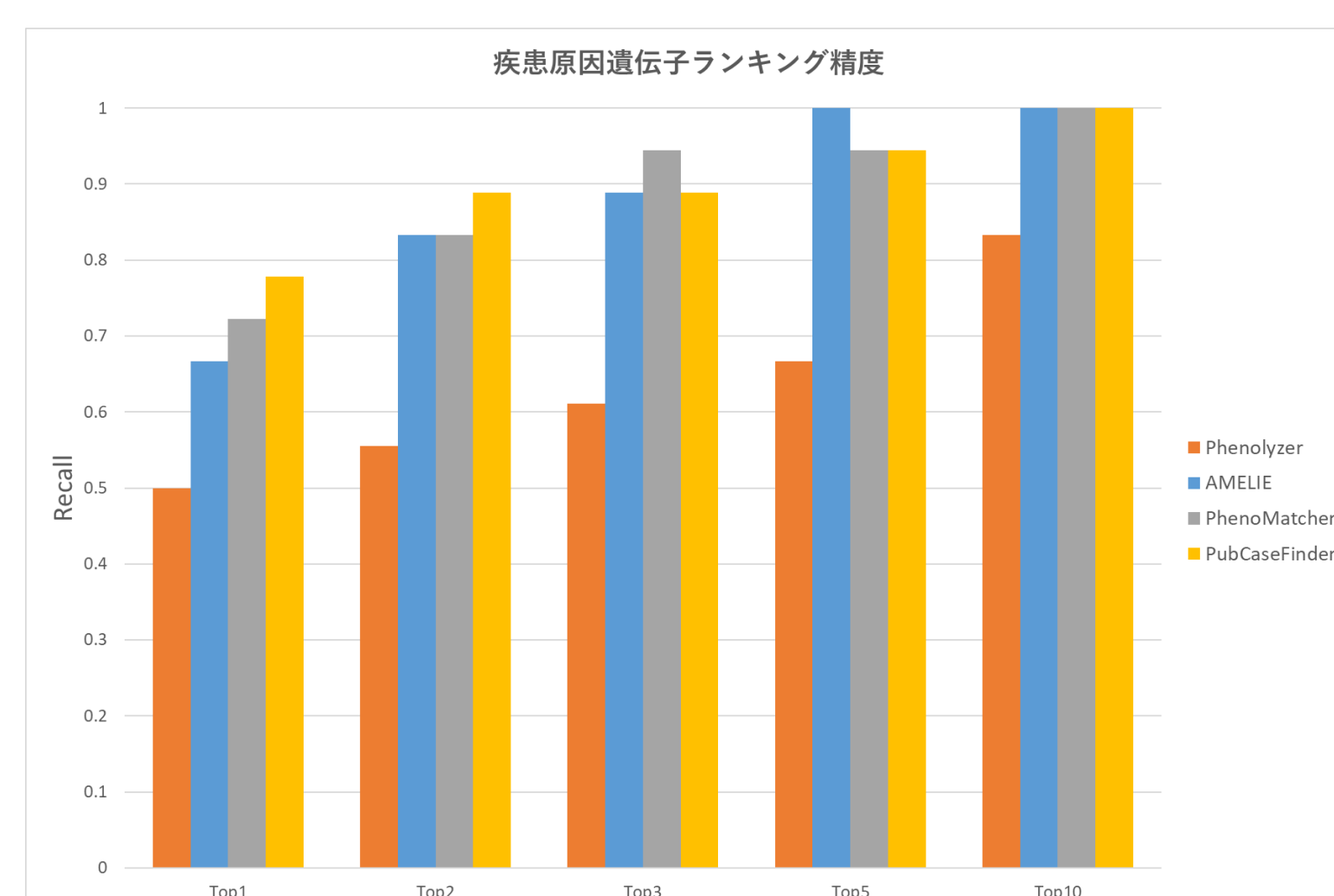


図4. 疾患原因候補遺伝子ランキング比較結果

今後

- 疾患原因候補遺伝子ランキングシステムに必要な遺伝子-症状関連データを拡張することで、より精度の高いシステムを構築することを目指す。



- 先行研究として、PubMedに含まれる全文献から自動抽出した遺伝子-疾患関連データ (DisGeNET) を活用することで、遺伝子-症状関連データを拡張した。

利用データセット

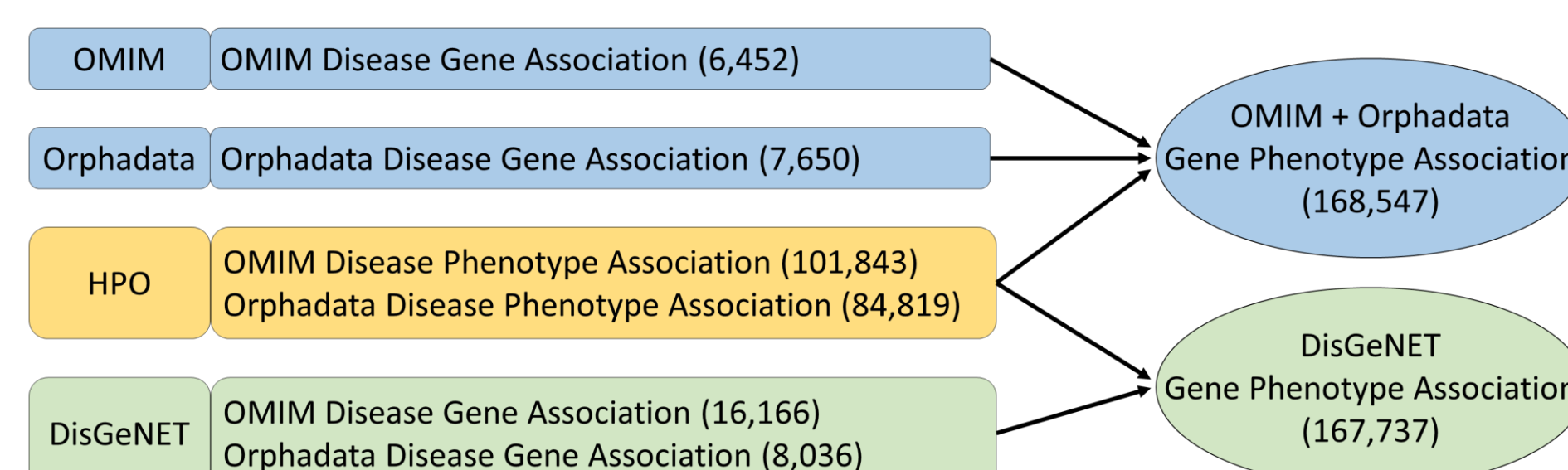


図5. 遺伝子-症状関連データを拡張

OMIM + Orphadata GPA DisGeNET GPA

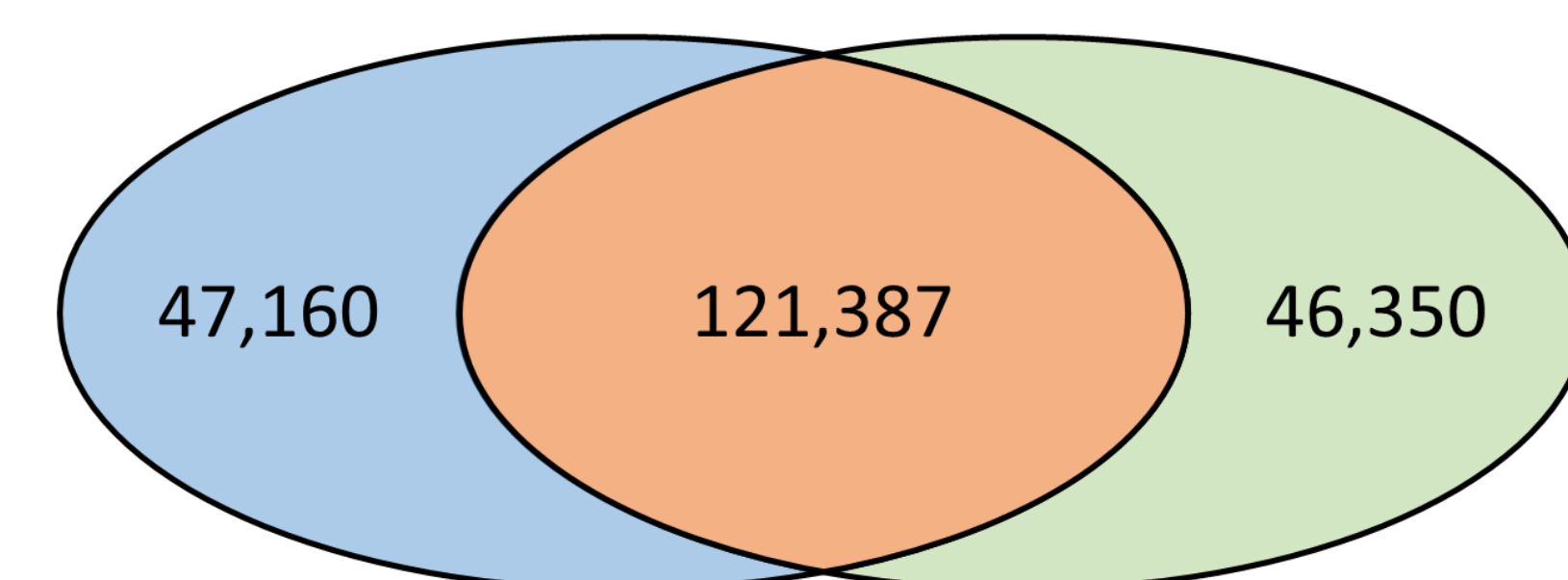


図6. 拡張した遺伝子-症状関連データの比較