

○平田誠、坂手龍一

国立研究開発法人 医薬基盤・健康・栄養研究所 難治性疾患研究開発・支援センター

厚生労働省の指定難病の創薬については、他疾患で開発された薬剤の適応拡大であったり、逆方向の適応拡大であったりすることがある。そうした適応拡大については、それら疾患間の関連性についても情報が得られるものと考えられる。そこで我々は、**PubMed全論文のタイトルと要旨をテキストマイニングにより解析し、疾患（指定難病、他疾患）と薬物とが共起する論文を抽出**して、研究動向を分析した。指定難病はDDrare¹⁾、他疾患はMalaCards²⁾から、薬物はDrugBank³⁾からリストを作成し、共起関係における頻出語の影響は、IDF (Inverse Document Frequency) 値を算出して除去した。その結果、PubMed約3,400万論文のうち、**約9,100薬物について約1,100万論文**が得られた。そのうち**指定難病332疾患について約36万論文、他疾患約10,000疾患について約517万論文**が得られた。薬物についてみると、

指定難病と他疾患とで4,800薬物が共有されていた。1,000薬物について時系列での論文数の変化をみると、**92.6%の薬物で他疾患の方が指定難病より先に論文報告**され、逆は約1.5%のみであった。疾患間での論文数の時系列変化、及び疾患ごとの薬物の論文数の時系列変化について解析した結果を報告する。

- 1) DDrare : <https://ddrare.nibiohn.go.jp>
- 2) MalaCards : <https://www.malacards.org>
- 3) DrugBank : <https://go.drugbank.com>

1. PubMedのテキストマイニング

表1の解析データを用いて、PubMed全論文を対象に、疾患（指定難病、他疾患）と薬物とが共起する論文を抽出して、研究動向を分析した。共起関係における頻出語の影響を除去するため、IDF値を利用した方法（Modified IDF）により、頻出薬物を抽出した。

その結果、PubMed約3,400万論文のうち、約9,100薬物について約1,100万論文が得られた。そのうち指定難病332疾患について約36万論文、他疾患約10,000疾患について約517万論文が得られた。薬物についてみると、指定難病と他疾患とで4,800薬物が共有されていることがわかった。

Modified IDF Modified Inverse Document Frequency

出現回数が多いが、出現文書数が少ない語句が重要として薬物を抽出

$$mIDF_m = \log \left(\frac{\text{全薬物の論文数}}{\text{薬物}_m\text{の論文数} + 1} + 1 \right) \\ \times \log \left(\text{薬物}_m\text{の論文数} \right)$$

表1. 解析データ（2022年7月取得）

	データソース	サイズ	備考
論文	PubMed	約3,400万	タイトル、要旨
疾患	DDrare (指定難病)	338	英語疾患名（同義語、細分名を含む1,871のうち原則4文字以上のもの）
	MalaCards	約22,000	ここから上記難病を除いた疾患
薬物	DrugBank v5.1.9	約15,000	Name, Synonyms

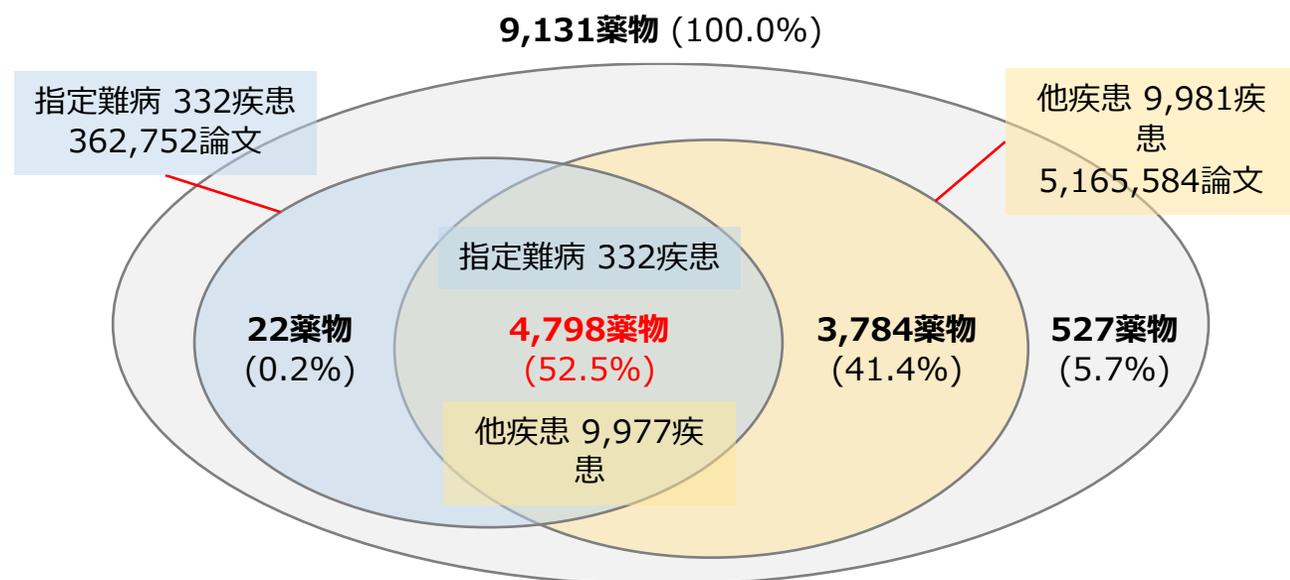


図1. テキストマイニングにより抽出された薬物数



2. 指定難病と他疾患の研究動向の比較

指定難病と他疾患の両方と共起した4,798薬物のうち、mIDFスコアの大きいトップ1,000薬物を解析用に抽出した（トップ20薬物を表2に示す。）。トップ1,000薬物のそれぞれについて、指定難病と他疾患のどちらで先に論文が現れたかを比較した。その結果、92.6%の薬物で他疾患の方が指定難病より先に論文報告されたことがわかった。

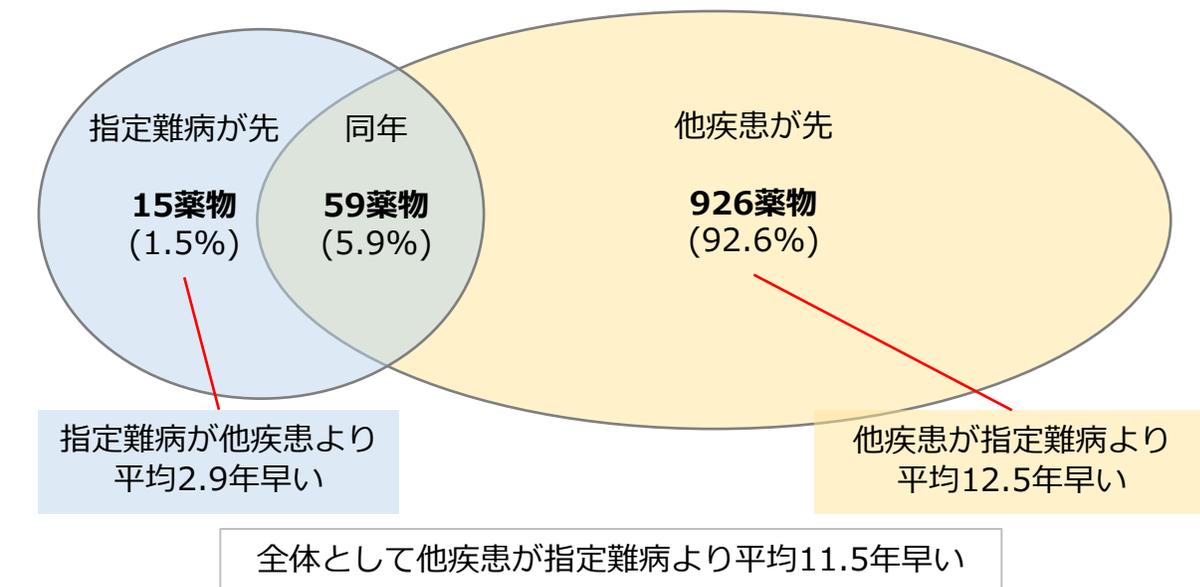


図2. 指定難病と他疾患の両方と共起する薬物トップ1,000における最初の論文出現年の比較

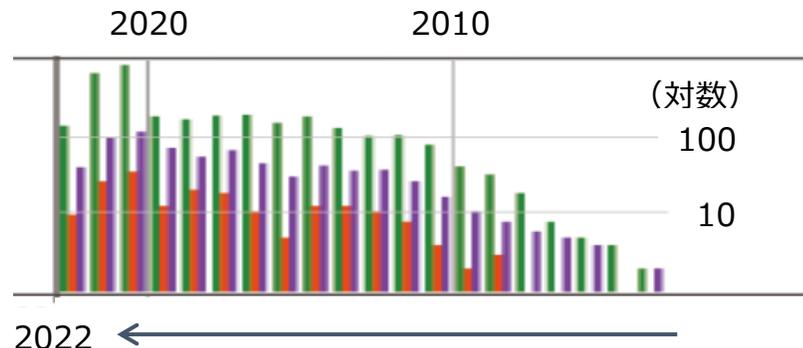
表2. 指定難病と他疾患の両方と共起する薬物トップ20

	mIDFスコア	薬物 (DrugBank)
1	75.25	DB10583 Clostridium tetani toxoid antigen (formaldehyde inactivated)
2	75.23	DB00761 Potassium chloride
3	74.89	DB09421 Protirelin
4	74.81	DB00480 Lenalidomide
5	74.80	DB06273 Tocilizumab
6	74.78	DB09035 Nivolumab
7	74.73	DB00853 Temozolomide
8	74.71	DB00005 Etanercept
9	74.69	DB15274 Pertussis vaccine
10	74.68	DB01073 Fludarabine
11	74.64	DB01101 Capecitabine
12	74.59	DB00317 Gefitinib
13	74.57	DB00530 Erlotinib
14	74.56	DB14954 PIK3CA
15	74.52	DB09037 Pembrolizumab
16	74.44	DB00398 Sorafenib
17	74.43	DB00642 Pemetrexed
18	74.41	DB06186 Ipilimumab
19	74.39	DB01268 Sunitinib
20	74.39	DB00188 Bortezomib

3. 薬物ごとの論文数の時系列変化

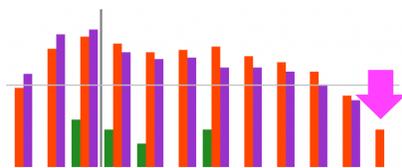
(凡例)

- 指定難病のみの論文
- 他疾患のみの論文
- 指定難病と他疾患の両方の論文



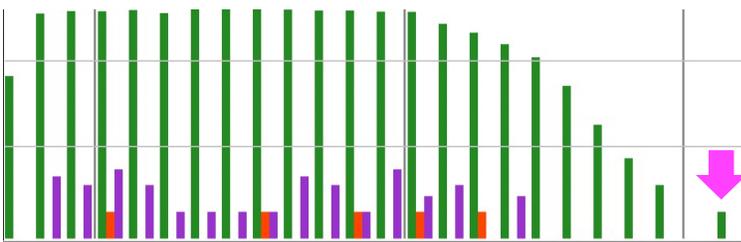
パターン1：指定難病が早い

DB08820 Ivacaftor
全531論文
指定難病：5疾患
他疾患：103疾患



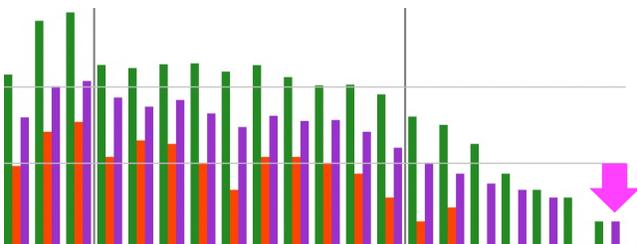
パターン2：他疾患が早い

DB00002 Cetuximab
全5,700論文
指定難病：25疾患
他疾患：591疾患



パターン3：同年

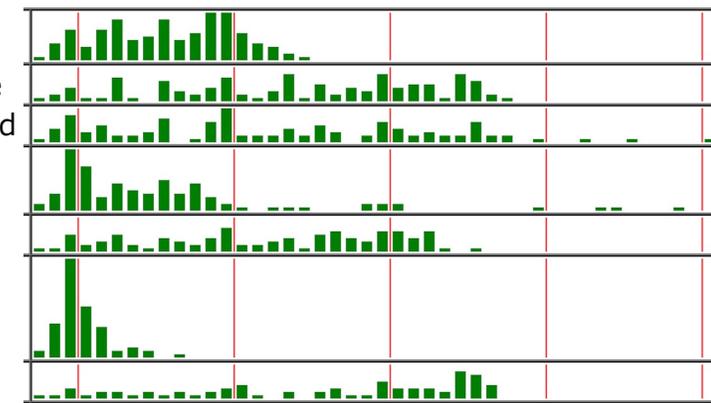
DB06273 Tocilizumab
全4,185論文
指定難病：76疾患
他疾患：793疾患



疾患ごとの薬物の論文数の時系列変化

144. レノックス・ガストー症候群：196薬物

- DB06201 Rufinamide
- DB00555 Lamotrigine
- DB00313 Valproic Acid
- DB00349 Clobazam
- DB00273 Topiramate
- DB09061 Cannabidiol
- DB00949 Felbamate



Cannabidiolは、レノックス・ガストー症候群とドラベ症候群で、2018年にFDAで承認されたが、その時期に論文が多い。

まとめ

本研究では指定難病と他疾患についての研究論文を調査して、共通の約4,800薬物を抽出した。そのうち90%以上の薬物が、他疾患で先に論文報告されている傾向であった。一方で、指定難病で先に論文報告されていたり、論文数が多い薬物も見つかった。また、疾患ごとの薬物のトレンドを明らかにすることができた。これらのデータは、解析を進めることにより、創薬標的探索の基礎データとなることが期待される。

