

○平田誠、坂手龍一、甲斐陽子、木村友則

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

難病の創薬研究においても、様々なデータをつないで新しい知見を得ていくことが重要であり、そのためには疾患名等の整備が必要不可欠である。本研究では指定難病333疾患の同義語等を含む疾患名を整備し、これをもとにPubMedから論文を抽出して薬物名等の有無を調べ、臨床試験情報と比較した。医薬基盤・健康・栄養研究所のDDrare (ddrare.nibiohn.go.jp) の難病英語名をもとにPubMed論文のタイトルと要旨を検索した結果、328疾患について約74万論文が見つかった。疾患ごとに、DDrareの臨床試験数と論文数を比較すると、上位10疾患のうち7疾患が同じであった。その中で、パーキンソン病は、約1,300の臨床試験数に対し、約50倍の6万以上の論文数があり、その他の疾患においても、10倍を超えるオーダーで論文数が臨床試験数より多い。薬物についてはDrugBankの薬物名をもとに検索し、328疾患の

約35万論文から、約5,000薬物が見つかった。薬物数上位の11疾患で1,000薬物を超え、その多くは臨床試験での開発薬物数の5倍を超える結果が得られた。遺伝子についても解析を進めており、今後、文章中での意味合いを明らかにしていくことで、創薬研究に有用な情報を得ることができると考えられる。

# 1. PubMed論文要旨から指定難病の情報を抽出

## ■ 目的・解析手法

PubMedから指定難病の論文を抽出し、そこから薬物と遺伝子の情報を取得するパイロット研究を実施した。臨床試験DBであるDDrareの情報を利用することで、難病創薬の情報源として、論文情報と臨床試験情報との比較を試みた。

まず、論文タイトルと要旨から、DDrareの疾患名（英名）をクエリーとして検索し、ヒットした論文から、さらに薬物名（DrugBank）、遺伝子名（HGNC）を検索して抽出した。それぞれ、同義語等（解析データ参照）や表記揺れ（Parkinson disease、Parkinson's disease等）への対応も図った。

### 解析データ（2020年6月取得）

	ソース	サイズ	備考
論文	PubMed	約3,000万	全論文（タイトルのみも含む）
疾患	DDrare	333	英名（同義語、細分名を含むと1,294）
薬物	DrugBank v5.1.6	13,564	Name, Synonyms
遺伝子	HGNC	42,152	Approved, Previous, Alias symbols/names

※ DrugBankとHGNCのPubMedリンク（順に1,069論文、1,294論文）もマージした。

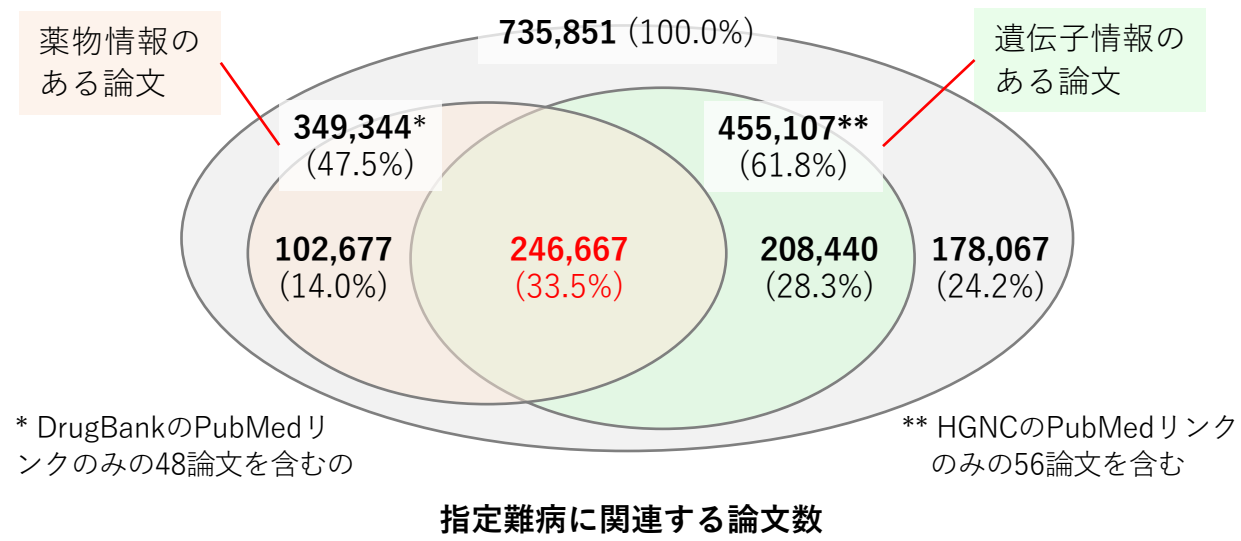
## ■ 結果 1

PubMedの約3,000万論文から、328疾患について約74万論文（1909年-2020年）が抽出できた。そのうち、約35万論文（47.5%）から328疾患についての**4,783薬物**、約46万論文（61.8%）から327疾患についての**12,102遺伝子**の情報が得られた。薬物と遺伝子、両方の情報が得られたのは**約25万論文（33.5%）**であった（**326疾患**についての**4,382薬物**、**9,939遺伝子**）。

### 論文数の多い指定難病トップ10

	論文数	告示番号	疾患名	疾患名（英名）
1	65,848	46	悪性関節リウマチ	Malignant rheumatoid arthritis*
2	60,525	6	パーキンソン病	Parkinson disease
3	49,198	13	多発性硬化症／視神経脊髄炎	Multiple sclerosis, Neuromyelitis optica
4	33,257	49	全身性エリテマトーデス	Systemic lupus erythematosus
5	30,145	299	嚢胞性線維症	Cystic fibrosis
6	28,830	96	クローン病	Crohn disease
7	26,433	65	原発性免疫不全症候群	Primary immunodeficiency
8	23,522	97	潰瘍性大腸炎	Ulcerative colitis
9	17,857	113	筋ジストロフィー	Muscular dystrophy
10	16,989	19	ライソゾーム病	Lysosomal storage disease

\* 同義語にRheumatoid arthritisを含む



\* DrugBankのPubMedリンクのみの48論文を含むの

\*\* HGNCのPubMedリンクのみの56論文を含む

### 論文数の多い薬物の例

Dopamine (15,022)、Histidine (14,787)、Levodopa (9,784)、Methotrexate (8,356)

### 論文数の多い遺伝子の例

TNF (20,073)、SOD1 (12,222)、IL6 (8,978)、INS (insulin) (8,384)

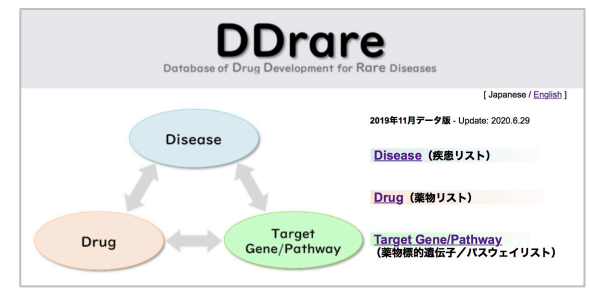


# 2. 臨床試験情報 (DDrare) との比較

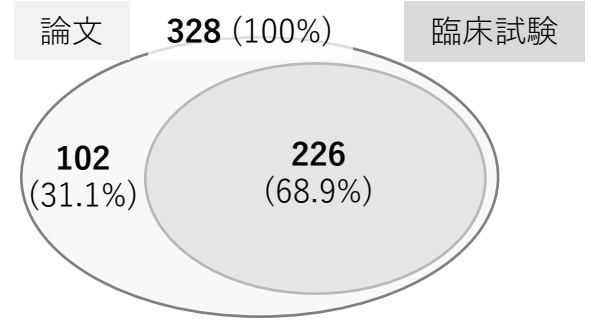
## ■ 結果2

抽出した指定難病に関する論文の薬物 (DrugBank) と遺伝子 (HGNC) の情報を、医薬基盤・健康・栄養研究所の DDrareの臨床試験情報と比較した。

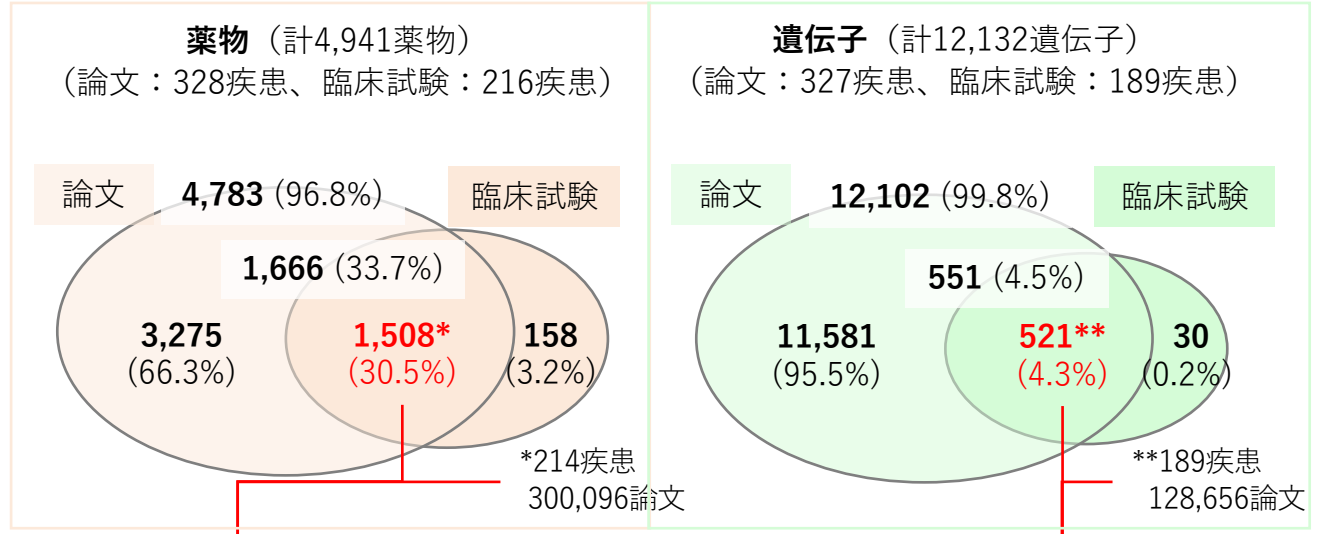
各疾患の論文数は臨床試験数と比較して数10倍のスケールがあり、臨床試験情報のある全疾患が論文情報に含まれていた。一方で、薬物と遺伝子については、臨床試験情報 (薬物標的遺伝子) のみにあるものがあった。



DDrare : 難病臨床試験情報  
ddrare.nibiohn.go.jp



論文と臨床試験 (疾患数) の比較



薬物と遺伝子を臨床試験情報と共有  
70,050論文の188疾患、1,315薬物、521遺伝子

論文と臨床試験 (薬物数、遺伝子数) の比較

### 論文数と臨床試験数の比較 (論文数トップ10)

告示番号	疾患名	論文数(A)	臨床試験数(B)	比率(A/B)
1	46 悪性関節リウマチ	65,848	1,999	32.9
2	6 パーキンソン病	60,525	1,307	46.3
3	13 多発性硬化症/視神経脊髄炎	49,198	1,398	35.2
4	49 全身性エリテマトーデス	33,257	429	77.5
5	299 嚢胞性線維症	30,145	856	35.2
6	96 クローン病	28,830	890	32.4
7	65 原発性免疫不全症候群	26,433	273	96.8
8	97 潰瘍性大腸炎	23,522	820	28.7
9	113 筋ジストロフィー	17,857	289	61.8
10	19 ラインゾーム病	16,989	399	42.6

### ■ まとめ・考察

1. 全論文の2.5%である約74万論文が、指定難病に関する論文として抽出できた。
2. 1/3の約25万論文に薬物と遺伝子、両方の情報があり、難病創薬の情報発掘対象として有用と考えられる。
3. 薬物と遺伝子については、臨床試験の情報が必ずしも論文にないケースがある。
4. 論文と臨床試験とで薬物と遺伝子を共有する約7万論文を抽出した。

論文からの情報抽出は機械学習等により高度化が進行しつつある。より情報抽出の質・量を向上すべく、新規解析技術の取り入れや、キュレーションを行う計画である (「3. 表示システムの開発」参照)。



# 3. 表示システムの開発

PubMed論文タイトルと要旨、そこから抽出した疾患名、薬物名、遺伝子名をハイライト表示するデータベースWebサイトを開発中である。これを利用してキュレーションを行い、解析手法の改良を進めている。

- 疾患名 (DDrare : Disease ID のXXXは指定難病の告示番号)
- 薬物名 (DrugBank)
- 遺伝子名 (HGNC)

DD6 =>PubMed[ [27186120](#) : 2012]

Disease ID	Dis-Name	Anony-names
* DD6	パーキンソン病	Parkinson disease;Parkinson's disease;PD;PD

\*Hugo Gene Symbol(PubMedID)  
 \*Hugo Gene (search word[id])  
 monoamine oxidase B:[6834](#)

\*Drug Bank(g,t,e) , search Drug  
[DB00640](#):Adenosine  
[DB00988](#):Dopamine  
[DB01235](#):Levodopa  
[DB06654](#):Safinamide

\*Hugo Gene (search Word[ids])

TITLE

Investigational agents in the treatment of Parkinson's disease: focus on **safinamide**

Abstract

The authors review management issues in **Parkinson's disease (PD)** and with a specific focus on **safinamide**. Current therapeutic management of **PD** involves **dopamine**, using **levodopa**, **dopamine** agonists, and inhibitors of dopamine metabolism. **Safinamide** as monotherapy and as an adjunct to **dopamine** agonists improved motor symptoms in a randomized, placebo-controlled trial involving 168 patients given a median dose of 50 mg daily. **Safinamide** increased the proportion of responders - defined as patients improving their motor symptoms - from baseline - after 3 months (37.5% for **safinamide** versus 21.4% for placebo). **Safinamide** also improved dyskinesia compared with the placebo in another trial, but dyskinesia severity was similar to that of the placebo. Further Phase III trial data for **safinamide** and other **PD** therapeutics: modified formulations of available compounds and combination-based therapies.

DD3 DD10 =>PubMed[ [17035524](#) : 2006]

Disease ID	Dis-Name	Anony-names
* DD3	脊髄性筋萎縮症	Spinal muscular atrophy;Myelopathic muscular atrophy
* DD10	シャルコー・マリー・トゥース病	Charcot-Marie-Tooth disease

\*Hugo Gene Symbol(PubMedID)  
 \*Hugo Gene (search word[id])  
 CMT2D:[4162](#)  
 GARS:[4162](#)  
 YARS:[12840](#)

\*Drug Bank(g,t,e) , search Drug  
[DB00145](#):Glycine  
[DB11017](#):Yeast

\*Hugo Gene (search Word[ids])

TITLE

Functional analyses of glycyl-tRNA synthetase mutations suggest a key role for tRNA-charging enzymes

Abstract

**Charcot-Marie-Tooth disease** type 2D (**CMT2D**) and distal **spinal muscular atrophy** type V (dSMA-V) are both characterized by a phenotype that is more severe in the upper extremities. We previously implicated mutations in the gene **GARS** as the cause of **CMT2D** and dSMA-V. **GARS** is a member of the family of aminoacyl-tRNA synthetases responsible for charging tRNAs with amino acids; **GARS** ligates **glycine** to tRNA(**Gly**). Here, we present functional analyses of disease-associated mutations in **GARS** expression levels; that the majority of identified mutations impair viability; and that, in most cases, mutant **GARS** protein mislocalizes in neuronal cells. Indeed, for some mutations, we observed **GARS**-associated granules in the neurite projections of cultured neurons and in the peripheral nerve axons, a feature particularly important in light of the recent identification of CMT-associated mutations in another tRNA synthetase gene [Glycyl-tRNA synthetase (GlyRS)]. Together, these findings suggest that tRNA-charging enzymes play a key role in maintaining peripheral nerve function.

DD3 =>PubMed[ [30302786](#) : 2019]

Disease ID	Dis-Name	Anony-names
* DD3	脊髄性筋萎縮症	Spinal muscular atrophy;Myelopathic muscular atrophy;SMA

\*Hugo Gene Symbol(PubMedID)  
 \*Hugo Gene (search word[id])  
 SMN2:[11118](#)  
 SMA:[11117](#)

\*Drug Bank(g,t,e) , search Drug  
[DB01167](#):Itraconazole  
[DB15305](#):Risdiplam

\*Hugo Gene (search Word[ids])

TITLE

A phase 1 healthy male volunteer single escalating dose study of the pharmacokinetics and pharmacodynamics of **risdiplam** (RG7916, RO7034067), a **SMN2** splicing modifier.

Abstract

**Risdiplam** (RG7916, RO7034067) is an orally administered, centrally and peripherally distributed, survival of motor neuron 2 (**SMN2**) mRNA splicing modifier for the treatment of **spinal muscular atrophy (SMA)**. The objectives of this entry-into-human study were to assess the safety, tolerability, pharmacokinetics (PK) and pharmacodynamics of **risdiplam**, and the effect of the strong CYP3A inhibitor **itraconazole** on the PK of **risdiplam** in healthy male volunteers. Part 1 had a randomized, double-blind, adaptive design with 25 subjects receiving single ascending oral doses of **risdiplam** (ranging from 0.6-18.0 mg, n = 18) or placebo (n = 7). A Bayesian framework was applied to estimate **risdiplam**'s effect on **SMN2** mRNA. The effect of multiple doses of **itraconazole** on the PK of **risdiplam** was also assessed using a two-period cross-over design (n = 8). **Risdiplam** in the fasted or fed state was well tolerated. **Risdiplam** exhibited linear PK over the dose range with a multi-phasic decline with a mean terminal half-life of 40-69 h. Food had no relevant effect, and **itraconazole** had only a minor effect on plasma PK indicating a low fraction of **risdiplam** metabolized by CYP3A. The highest tested dose of 18.0 mg **risdiplam** led to approximately 41% (95% confidence interval 27-55%) of the estimated maximum increase in **SMN2** mRNA. **Risdiplam** was well tolerated and proof of mechanism was demonstrated by the intended shift in **SMN2** splicing towards full-length **SMN2** mRNA. Based on these data, Phase 2/3 studies of **risdiplam** in patients with **SMA** are now ongoing.