DBCLS
Database Center for Life Science

2019/11/11
HwaZhong Argiculrural University

PubCaseFinder : a case-report-based, phenotype-driven differential-diagnosis system for rare diseases

Jin-Dong Kim
Research Organization of Information and systems (ROIS)
Database Center for Life Science (DBCLS)

AJHG

Log in

ARTICLE | VOLUME 103, ISSUE 3, P389-399, SEPTEMBER 06, 2018

PubCaseFinder: A Case-Report-Based, Phenotype-Driven Differential-Diagnosis System for Rare Diseases

Toyofumi Fujiwara • Yasunori Yamamoto • Jin-Dong Kim • Orion Buske • Toshihisa Takagi

Open Archive • Published: August 30, 2018 • DOI: https://doi.org/10.1016/j.ajhg.2018.08.003 •

NORD®
National Organization for Rare Disorders

RareDiseaseFacts™

What is a RARE DISEASE ?

Any disease, disorder, illness or condition affecting **fewer than 200,000** people in the United States is considered RARE.[1]

1 in 10 Americans has a RARE DISEASE

30 million people have a serious, lifelong condition.

More than half are children[1]

Holding hands, they would circle the globe about 1.5 times

7,000 RARE DISEASES exist, with less than 500 FDA-approved treatments[2]

ONLY **5%** of RARE DISEASES have treatments.[2]

Patients with RARE DISEASES are frequently **misdiagnosed or undiagnosed.**

80% of RARE DISEASES ARE GENETICALLY BASED.[2]

Many RARE DISEASES result in premature death of infants & young children or are fatal in early adulthood.[2]

Families & private foundations provide about 3% of ALL medical research funding in the U.S.[6]

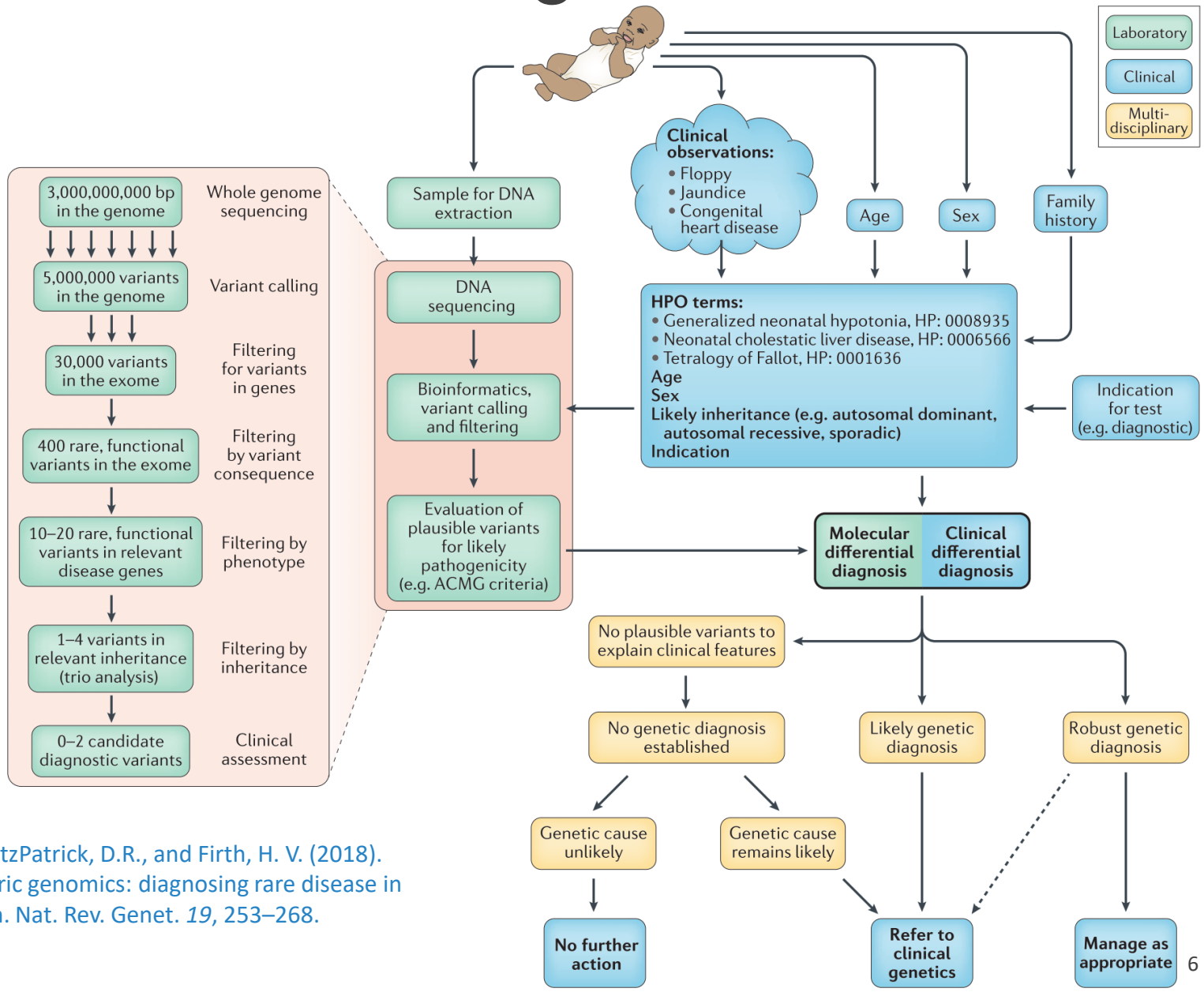
Rare diseases

- For patients of rare diseases, getting diagnosis at an early stage is important
- E.g., Phenylketonuria (PKU)
  - ✓ An inborn error of metabolism
  - ✓ Untreated, it can lead to intellectual disability, seizures, behavioral problems, and mental disorders.
  - ✓ If it is diagnosed early enough, an affected newborn can grow up with normal development through diet, or a combination of diet and medication.

NGS-based diagnosis

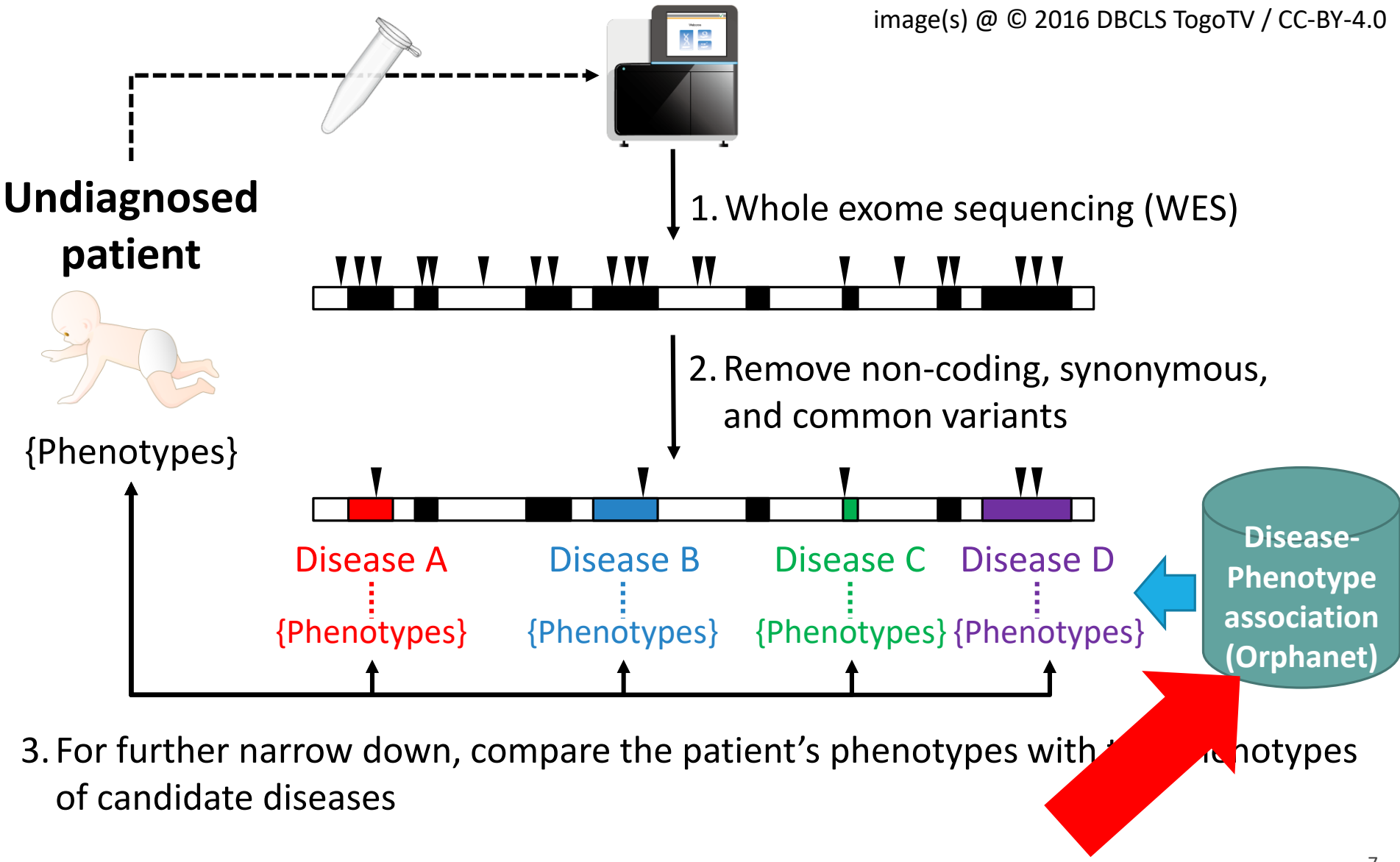
- Sawyer et al., 2016
  - ✓ “Recent advances in sequencing, in particular whole-exome sequencing (WES), are identifying the genetic basis of disease for 25–40% of patients.”

NGS-based diagnosis



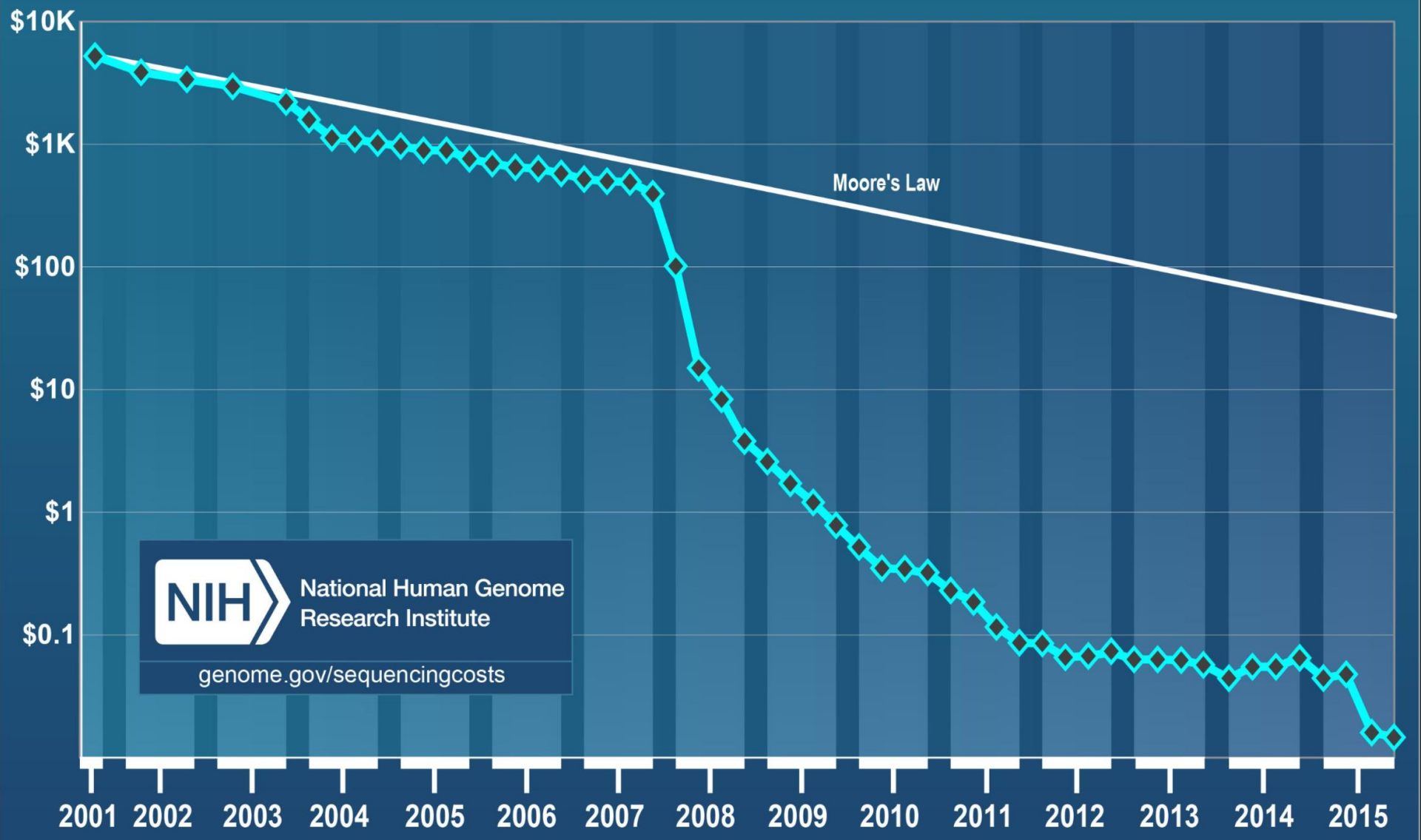
Wright, C.F., FitzPatrick, D.R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. Nat. Rev. Genet. *19*, 253–268.

For undiagnosed patients



3. For further narrow down, compare the patient's phenotypes with ... notypes of candidate diseases

Cost per Raw Megabase of DNA Sequence



IRUD

- Initiative on Rare and Undiagnosed Diseases
- A national system to find diagnoses for undiagnosed patients across Japan, by applying NGS-based diagnostic tests
- Applied to >6,000 cases, so far
- Achieved diagnosis rate of 33%

Initiative on Rare and Undiagnosed Diseases (IRUD)
: Integrating Knowledge for Diagnoses

IRUD Analysis Centers

Analyze patient samples:

- Administer genetic tests, often accompanied by exome or whole-genome sequencing
- Consider other tests, assays, and functional studies

Further analysis

IRUD Diagnosis Committees

DIscuss complex cases:

- Hold conferences with clinical geneticists and medical specialists
- Tie up with local communities

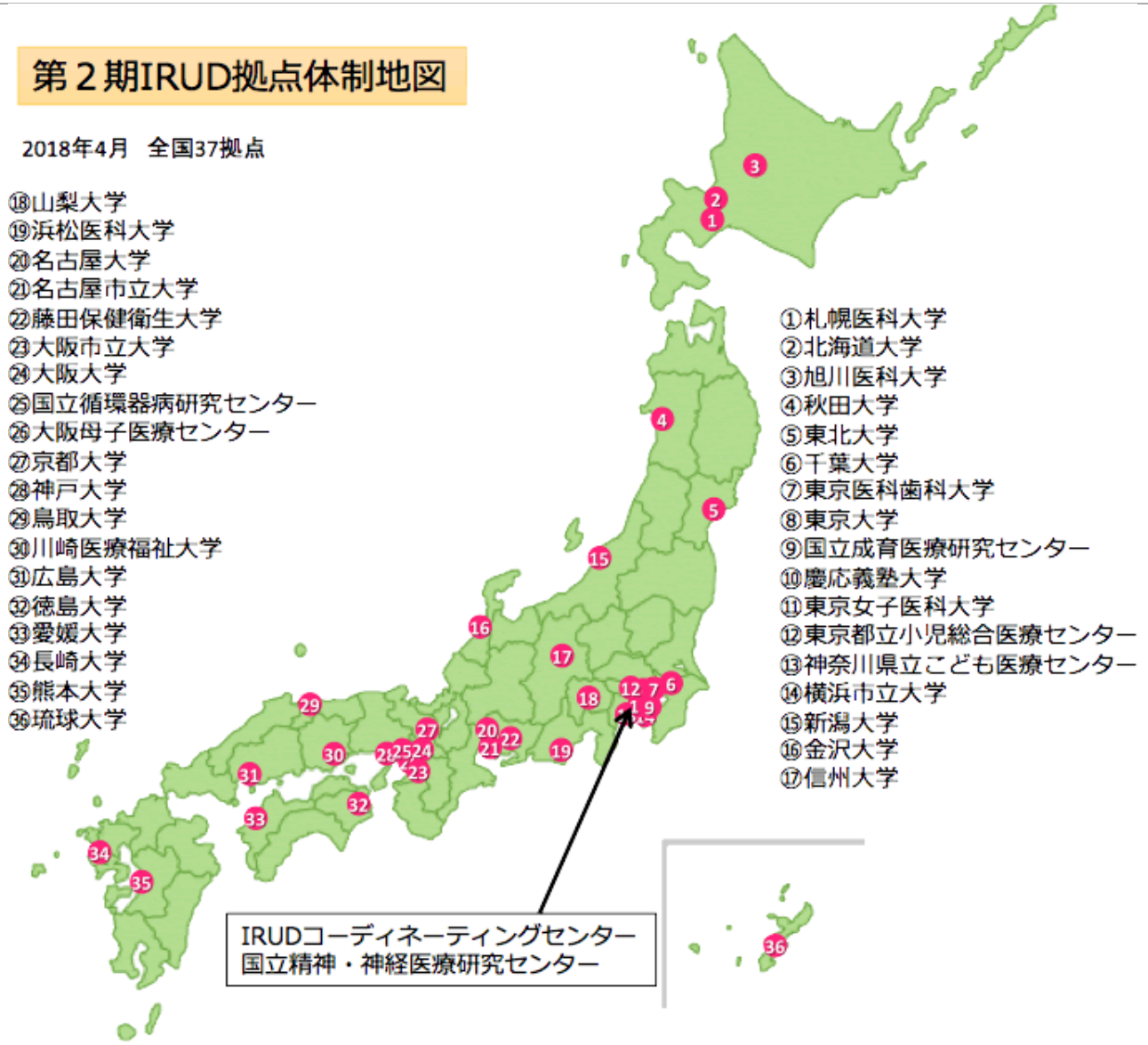
Data

IRUD Data Center

Develop and Manage databases:

- Streamline clinical and genetic data collection
- Enable global data-sharing

IRUD Clinical Centers

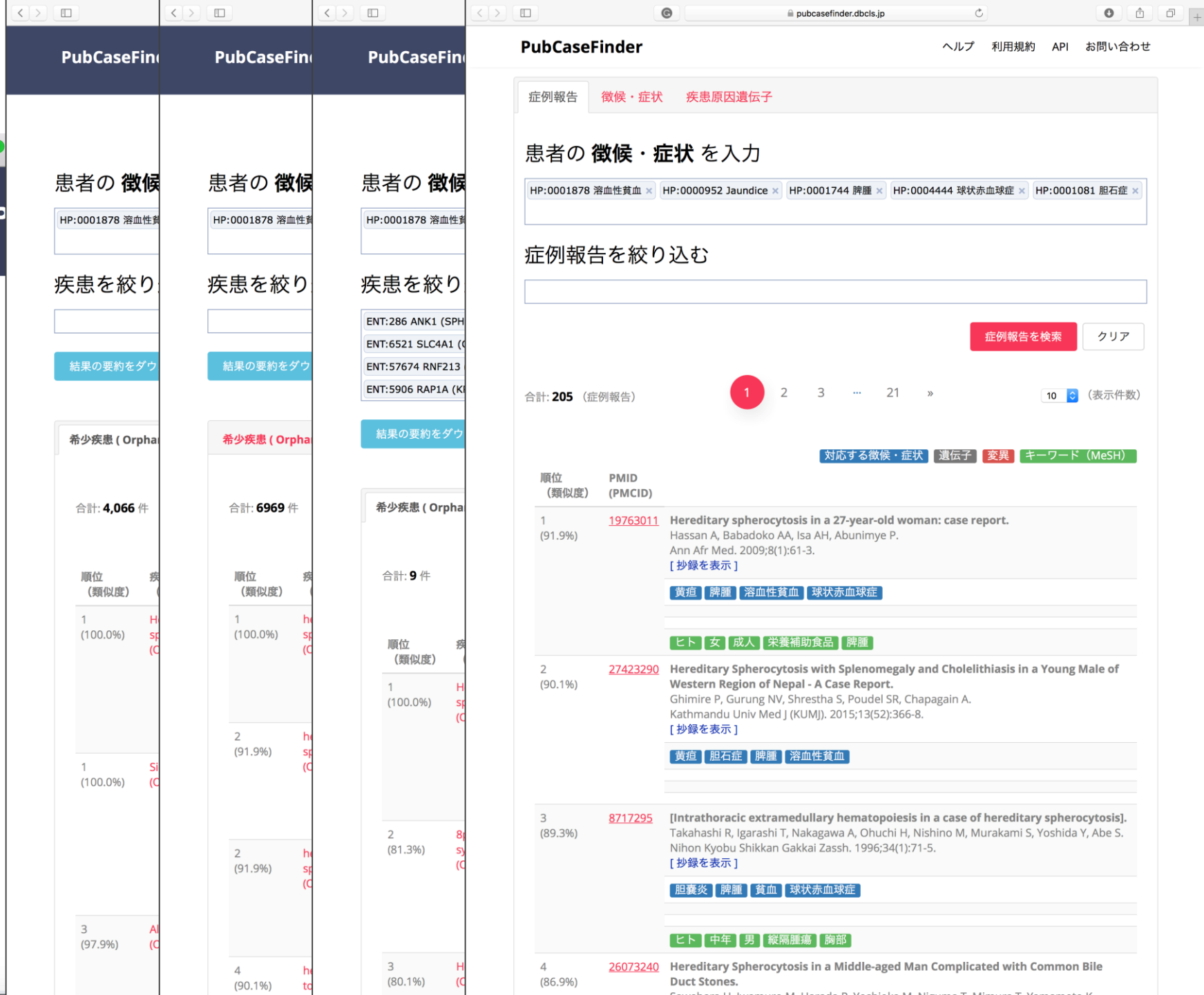


Needs for access to relevant resources

Orphanet
7K rare diseases

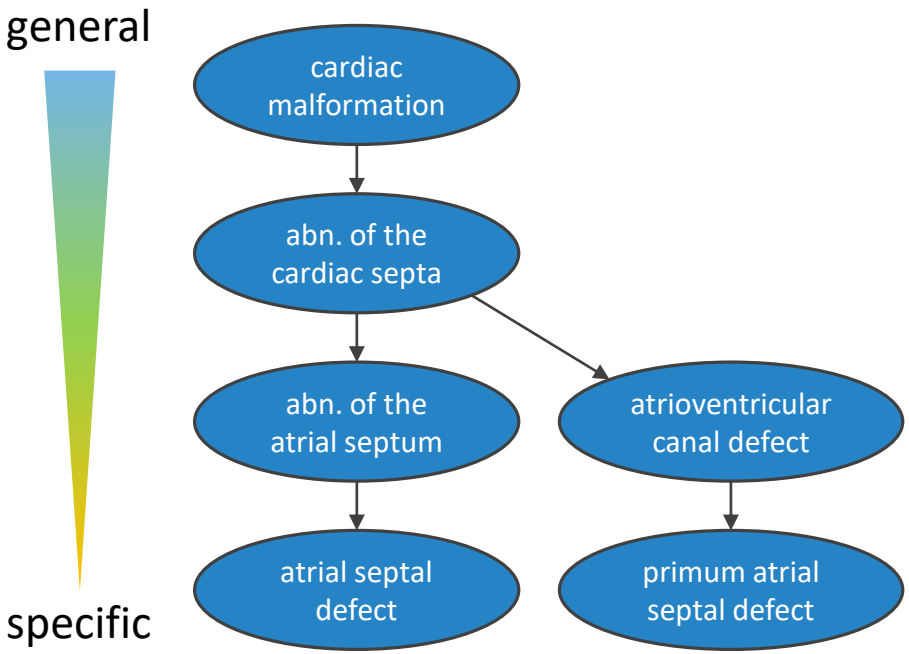
OMIM
8K genetic diseases

PubMed
1M case reports


pubcasefinder.dbcls.jp
PubCaseFinder
ヘルプ
利用規約
API
お問い合わせ
症例報告
徴候・症状
疾患原因遺伝子
患者の 徴候・症状 を入力
HP:0001878 溶血性貧血
HP:0000952 Jaundice
HP:0001744 脾腫
HP:0004444 球状赤血球症
HP:0001081 胆石症
症例報告を絞り込む
症例報告を検索
クリア
合計: 205 （症例報告）
1
2
3
...
21
»
10
（表示件数）
対応する徴候・症状
遺伝子
変異
キーワード（MeSH）
順位（類似度）
PMID (PMCID)
1 (91.9%)
19763011
Hereditary spherocytosis in a 27-year-old woman: case report.
Hassan A, Babadoko AA, Isa AH, Abunimye P.
Ann Afr Med. 2009;8(1):61-3.
[抄録を表示]
黄疸
脾腫
溶血性貧血
球状赤血球症
ヒト
女
成人
栄養補助食品
脾腫
2 (90.1%)
27423290
Hereditary Spherocytosis with Splenomegaly and Cholelithiasis in a Young Male of Western Region of Nepal - A Case Report.
Ghimire P, Gurung NV, Shrestha S, Poudel SR, Chapagain A.
Kathmandu Univ Med J (KUMJ). 2015;13(52):366-8.
[抄録を表示]
黄疸
胆石症
脾腫
溶血性貧血
3 (89.3%)
8717295
[Intrathoracic extramedullary hematopoiesis in a case of hereditary spherocytosis].
Takahashi R, Igarashi T, Nakagawa A, Ohuchi H, Nishino M, Murakami S, Yoshida Y, Abe S.
Nihon Kyobu Shikkan Gakkai Zassh. 1996;34(1):71-5.
[抄録を表示]
胆嚢炎
脾腫
貧血
球状赤血球症
ヒト
中年
男
縦隔腫瘍
胸部
4 (86.9%)
26073240
Hereditary Spherocytosis in a Middle-aged Man Complicated with Common Bile Duct Stones.

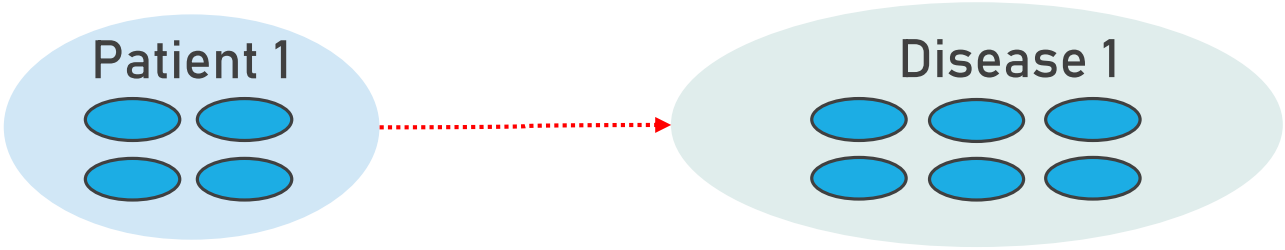
Human Phenotype Ontology (HPO)

- The Human Phenotype Ontology in 2017, Köhler et al., NAR, 2017
  - ✓ HPO is used by many data sets
    - Orphanet, OMIM, ClinVar, MedGen, GARD, …



Similarity Measures

- Similarity between two sets of phenotypes can be computed in many ways



Measure	Equation	Variations	Reference
Resnik(a,b)	$\max_{t \in g^a \cap g^b} IC(t)$	Avg, Max	Rensik, 1995
Lin(a,b)	$\frac{2 * Resnik(a,b)}{IC(a) + IC(b)}$	Avg, Max	Lin, 1998
Jiang-Conrath(a,b)	$\frac{1}{IC(a) + IC(b) - 2 * Resnik(a,b) + 1}$	Avg, Max	Jiang, 1997
simGIC(P,Q)	$\frac{\sum_{t \in g^P \cap g^Q} IC(t)}{\sum_{t \in g^P \cup g^Q} IC(t)}$		Pesquita, 2007
GeneYenta(P,Q)			Gottlieb, 2015

Gene Yenta

- Information Content (IC)

$$P(t) = \frac{|annot_t|}{|annot_{all}|}$$

$$IC_t = -\log P(t)$$

- Similarity between two terms

$$sim_{terms}\left(t, t^{'}\right) = max_{a_t \in A_t \cap A_{t'}}\ IC_{a_t}$$

- Similarity between a case and a disease

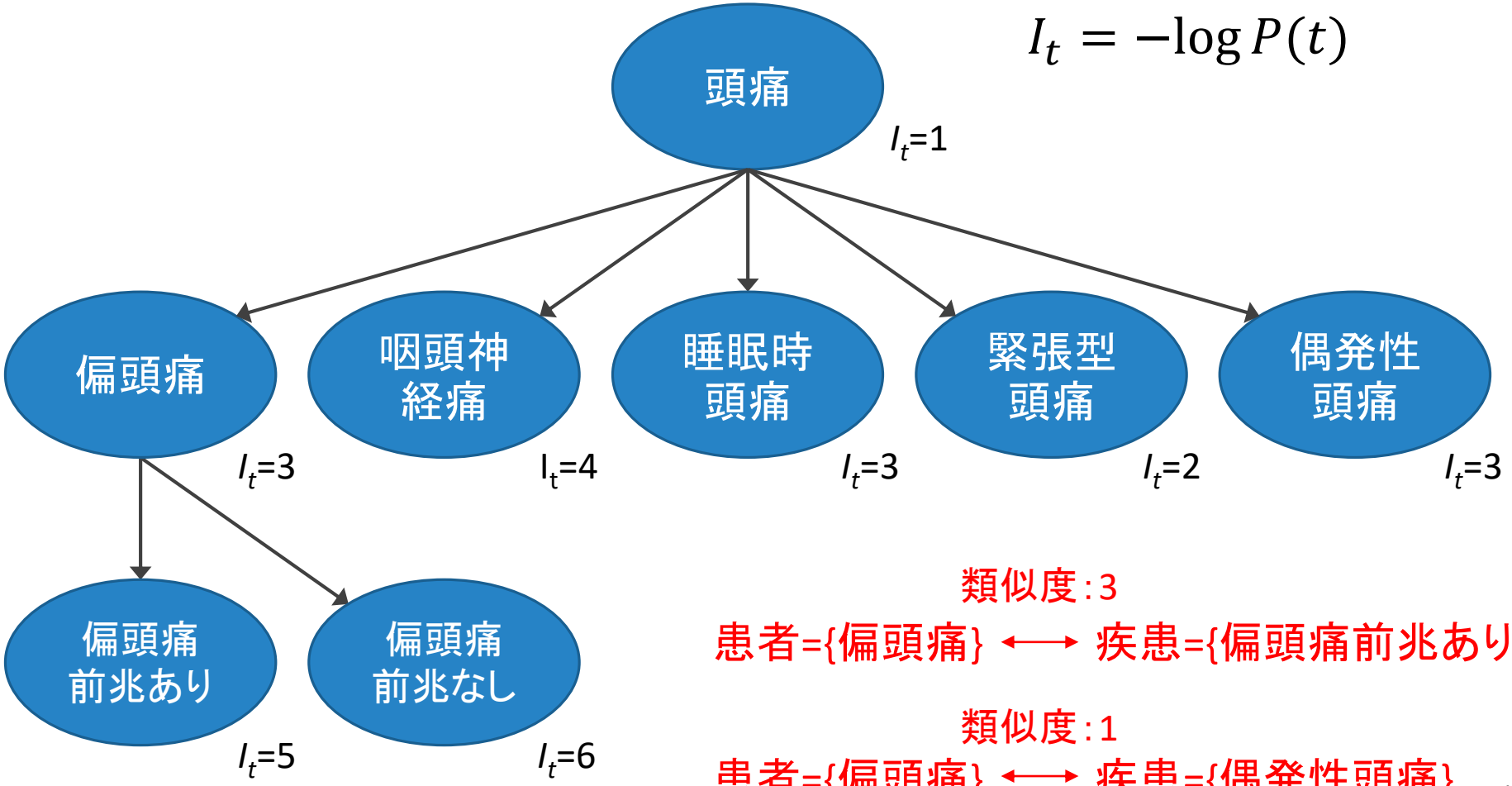
$$sim_{case\_disease}\left(c, d\right) = \frac{\sum_{t \in T_c} R_t \times max_{t^{'} \in T_d} sim_{terms}\left(t, t^{'}\right)}{\sum_{t \in T_c} R_t \times IC_t}$$

症状セットの類似度計算手法

- 症状間の類似度

$$P(t) = \frac{|annot_t|}{|annot_{all}|}$$

$$I_t = -\log P(t)$$



類似度：3

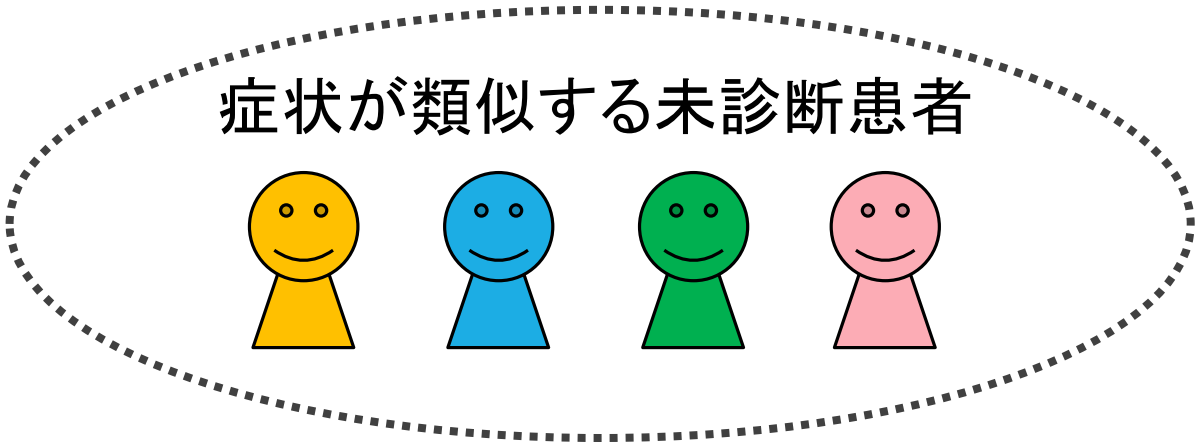
患者={偏頭痛} ⟷ 疾患={偏頭痛前兆あり}

類似度：1

患者={偏頭痛} ⟷ 疾患={偶発性頭痛}

HPOと類似度計算手法の活用事例

- 希少疾患分野では、HPOと類似度計算手法を用いて、症例データの共有が盛んに行われている
  - 症状が類似する複数の未診断患者が集まることで
    - 新規疾患の定義
    - 疾患原因遺伝子の同定



各未診断患者の症状はHPOで表現

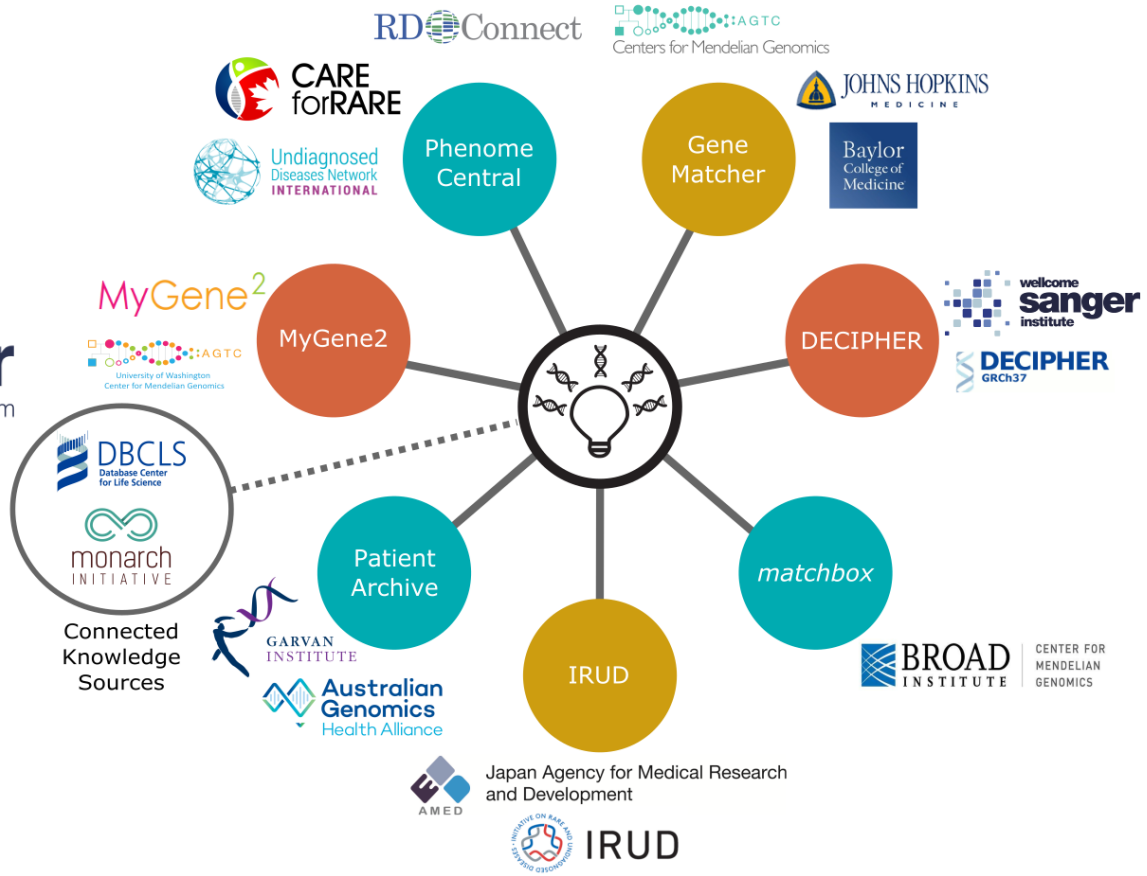
Databases which use HPO

Name	URL
PhenomeCentral	phenomecentral.org
DDD (Deciphering Developmental Disorders)	www.ddduk.org
DECIPHER (DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources)	decipher.sanger.ac.uk
ECARUCA (European Cytogeneticists Association Register of Unbalanced Chromosome Aberrations)	http://umcecaruca01.extern.umcn.nl:8080/ecaruca/ecaruca.jsp
The 100 000 Genomes Project	https://www.genomicsengland.co.uk/
Geno2MP (Exome sequencing data linked to phenotypic information from a wide variety of Mendelian gene discovery projects)	http://geno2mp.gs.washington.edu
NIH UDP (Undiagnosed Diseases Program)	available via phenomecentral.org
NIH UDN (Undiagnosed Diseases Network)	available via phenomecentral.org
HDG (Human Disease Gene Website series)	www.humandiseasegenes.com
Phenopolis (An open platform for harmonization and analysis of sequencing and phenotype data)	https://phenopolis.github.io
GenomeConnect (Patient portal developed by ClinGen (67)	www.genomeconnect.org
FORGE Canada & Care4Rare Consortium	available via phenomecentral.org
RD-Connect	platform.rd-connect.eu
Genesis	thegenesisprojectfoundation.org

MatchMaker Exchange Program

- An official program of GA4GH

PubCaseFinder
Phenotype-Driven Differential-Diagnosis System



Needs for access to relevant resources

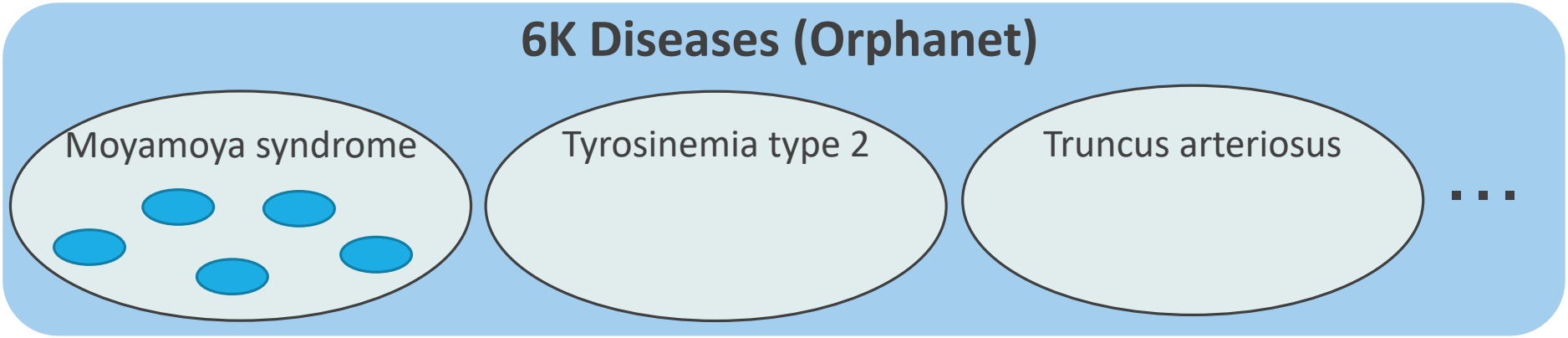
Orphanet
7K rare diseases

OMIM
8K genetic diseases

PubMed
1M case reports

Limited coverated of DBs

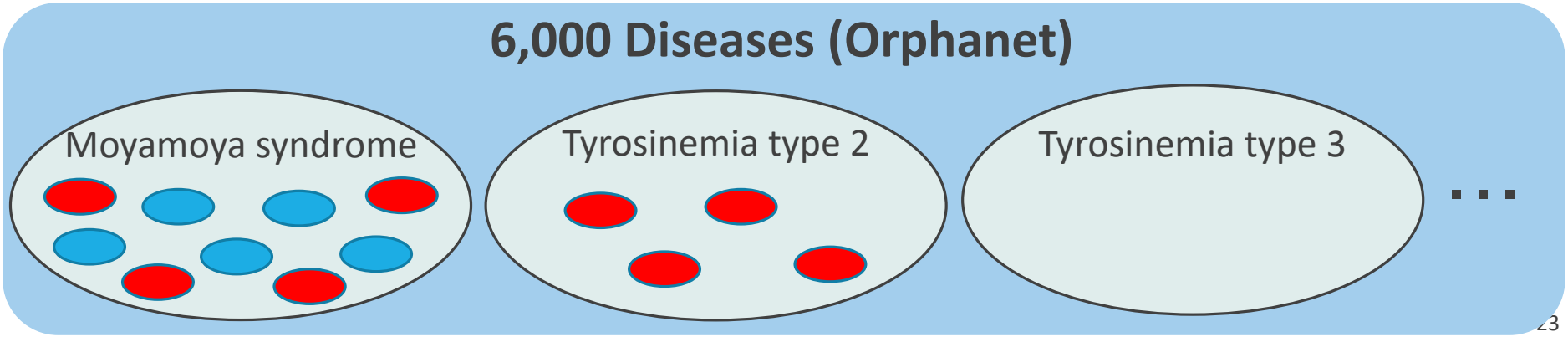
- Manually curated databases like Orphanet inherently have a limited coverage
  - ✓ For example, only 1/3 of rare diseases in Orphanet are associated with phenotypes



Text Mining

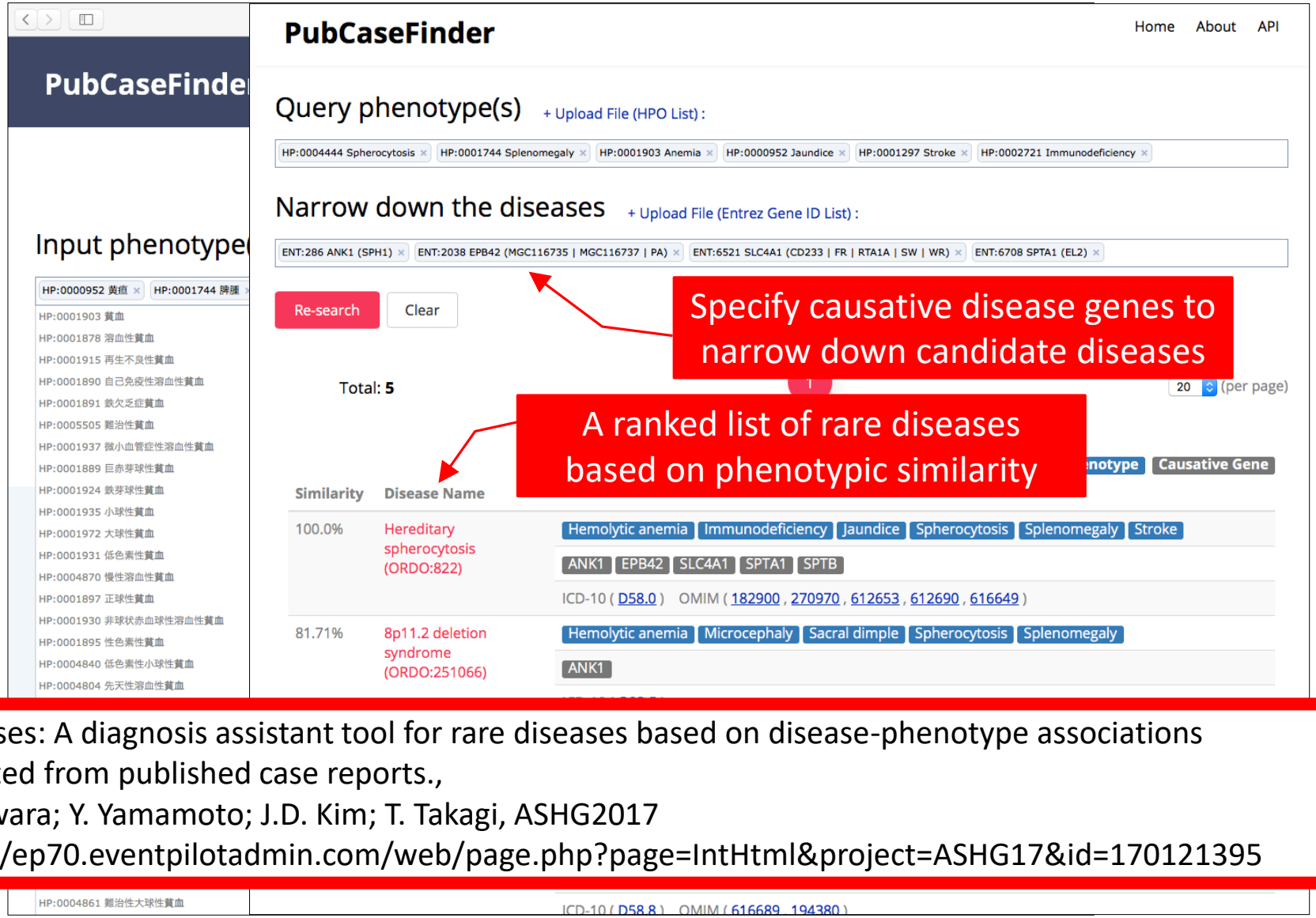
- Text mining-based approach is effective to improve the coverage of disease-phenotype associations
  - To automatically extract the associations from more than one million **case reports** in PubMed

As a result, **2/3** of rare diseases in Orphanet could be associated with phenotypes



PubCaseFinder

- A diagnosis support system for rare diseases



PubCases: A diagnosis assistant tool for rare diseases based on disease-phenotype associations extracted from published case reports.,
T. Fujiwara; Y. Yamamoto; J.D. Kim; T. Takagi, ASHG2017
https://ep70.eventpilotadmin.com/web/page.php?page=IntHtml&project=ASHG17&id=170121395

文献から疾患に関連する症状を抽出

- 課題:「疾患ー症状」関連データの不足が大きな課題
  - 約100万件の症例報告から、テキストマイニング技術で自動取得

希少疾患

例） 遺伝性球状赤血球症（Orphanet:822）

- 溶血性貧血（HP:0001878）
- 黄疸（HP:0000952）
- 脾腫（HP:0001744）

症例報告　PMID: 12355853

青:疾患名　赤:症状

Hereditary spherocytosis is a genetic, frequently familial hemolytic blood disease characterized by varying degrees of hemolytic anemia, splenomegaly, and jaundice. ・・・

症例報告

Publication of Case Reports

文献から疾患に関連する症状を抽出

- オントロジーを用いたアノテーションツール
  - ConceptMapper（Tanenblatt, 2010）、MetaMap（Aronson, 2010）、NCBO Annotator（Jonquet, 2009）
- CRAFT Corpus を利用した、パフォーマンス比較（Christopher, 2014）
  - 8つのオントロジーにおいてF-measureを比較結果、7つのオントロジーでConceptMapperのF-measureが最も高かった
- HPO gold standard（Tudor, 2015）を用いたツール評価（藤原，JSAI2017）

System	F-measure	Precision	Recall
NCBO Annotator	0.51	0.54	0.47
MetaMap	0.56	0.51	0.61
ConceptMapper	0.52	0.52	0.51

System	Processing time (sec)
NCBO Annotator	206.0
MetaMap	351.0
ConceptMapper	4.3

100万件の症例報告の処理に要する時間

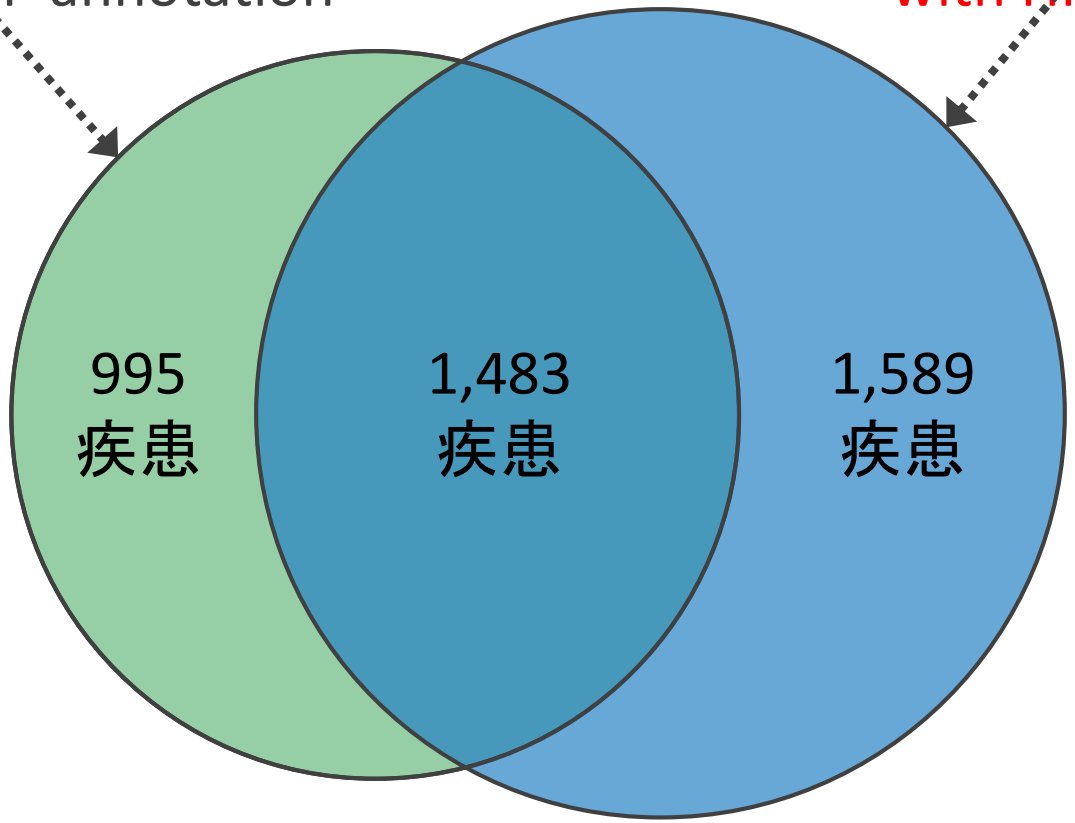
⇒ 17.7 (day)

⇒ 5.2 (hour)

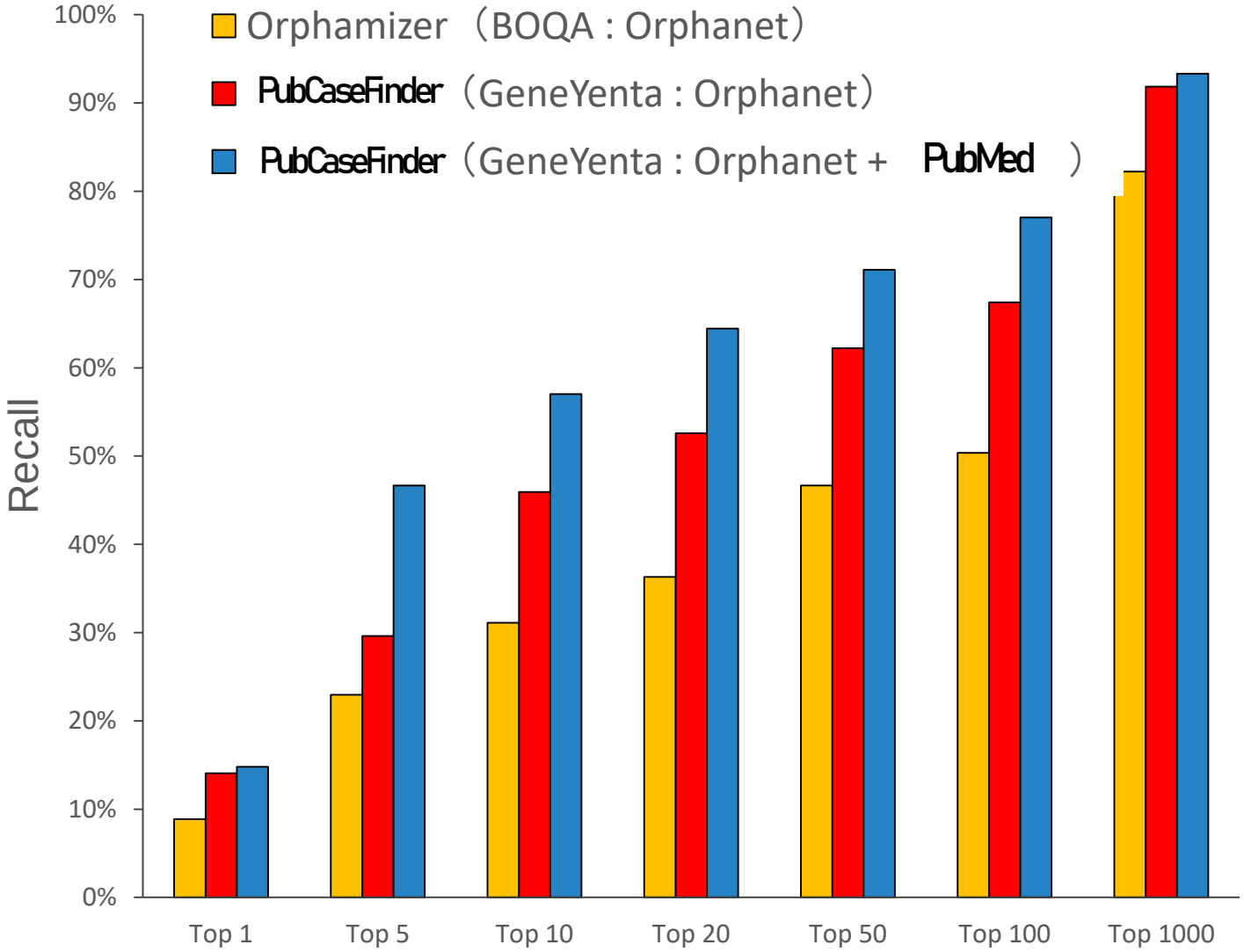
Result of IE from case reports

【Orphanet】
2,478 diseases
with HP annotation

【Case Reports】
3,072 diseases
with HP annotation



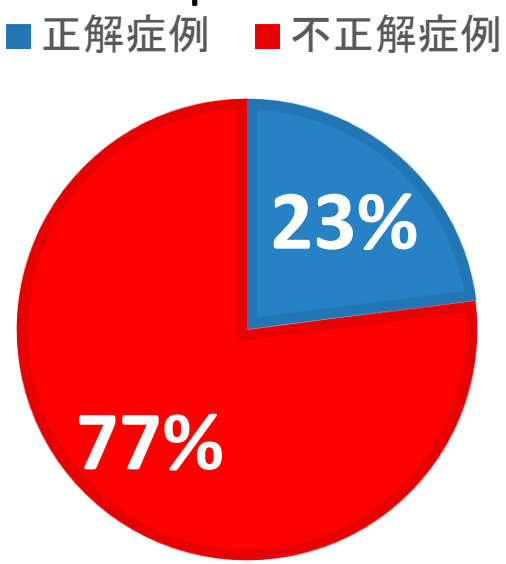
Performance Evaluation



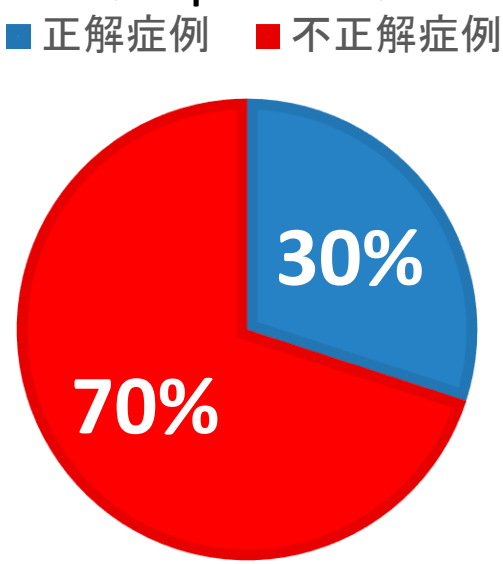
Performance Evaluation

- For 135 test cases
  - ✓ collected from PhenomeCentral
    - maintained by Care4Rare Canada Consortium
  - ✓ Recall @ top5 was calculated

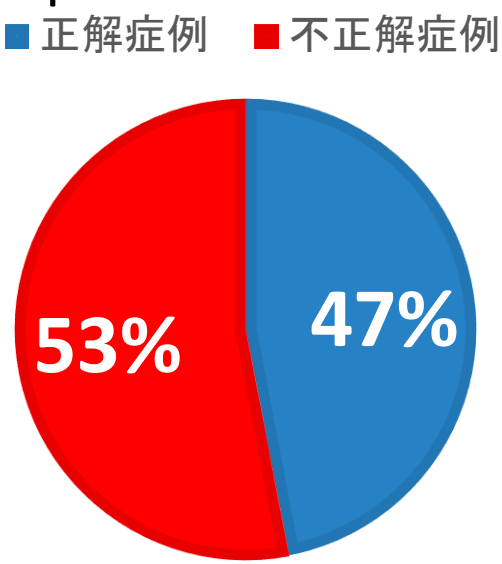
Orphamizer（Orphanet）



PubCaseFidner（Orphanet）

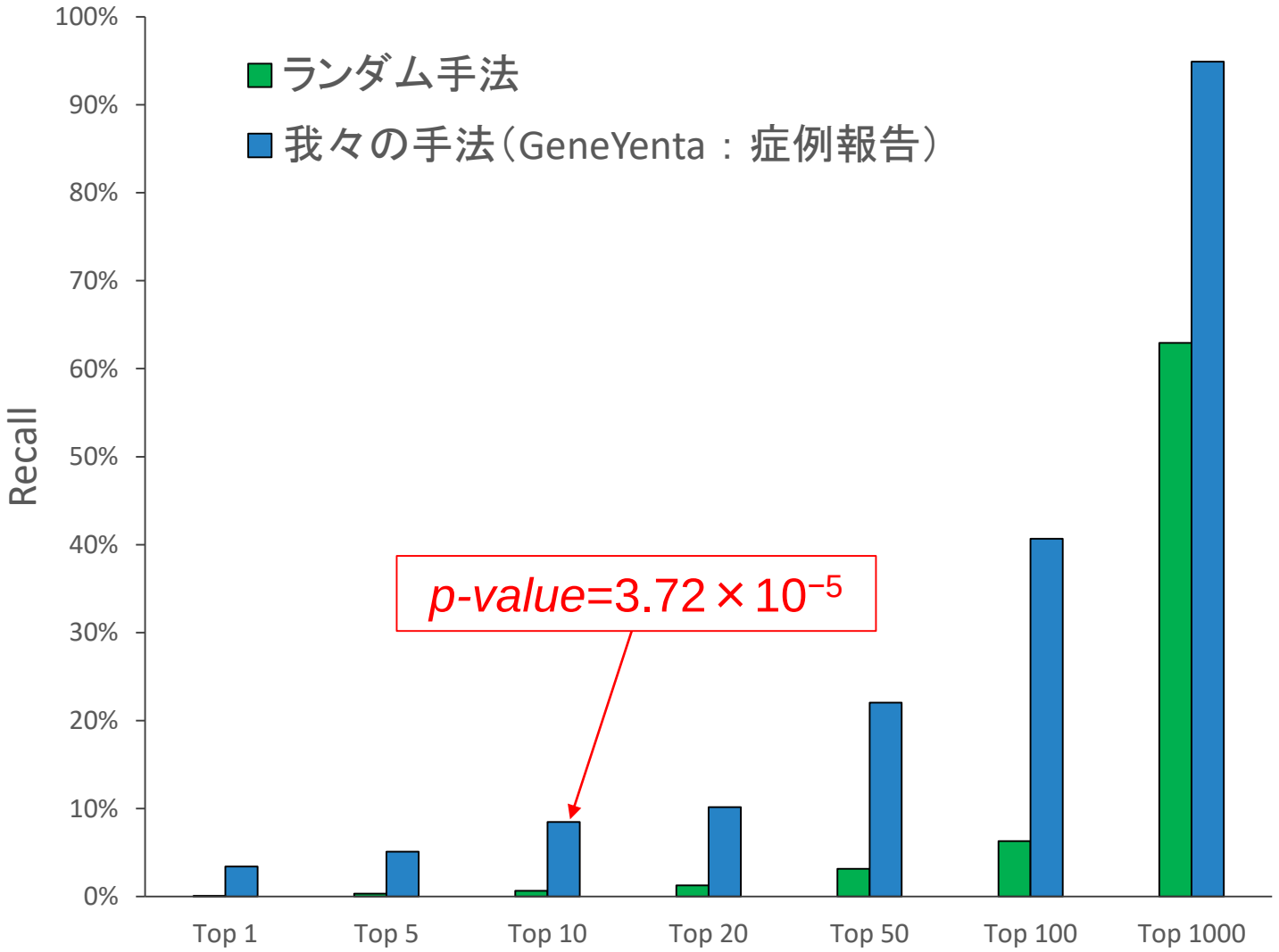


PubCaseFidner（Orphanet + PubMed）

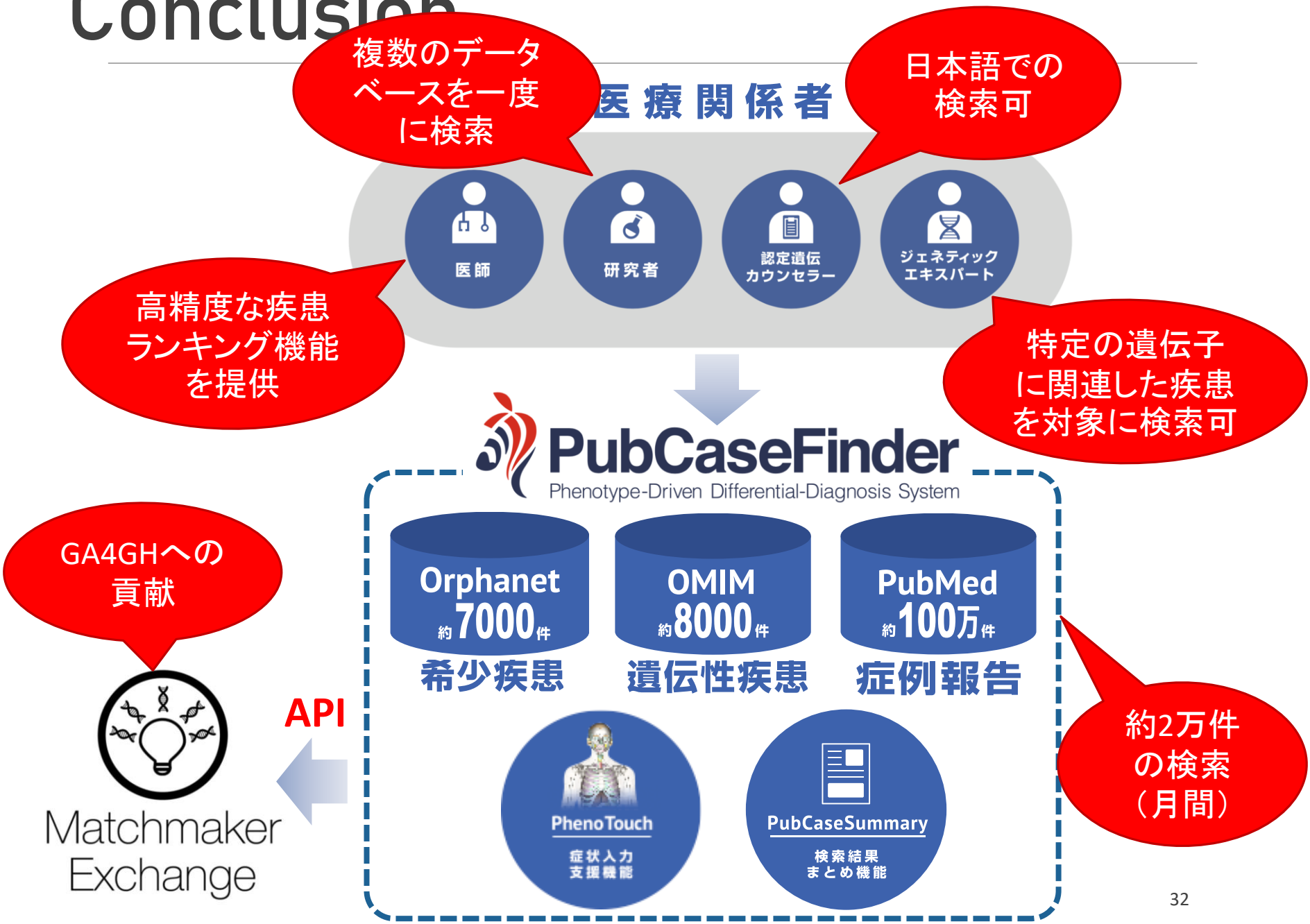


疾患ランキング精度の評価

- 評価用の59症例を利用して、Recall1,5,10,20,50,100,1000を比較



Conclusion


複数のデータベースを一度に検索
医療関係者
日本語での検索可
医師
研究者
認定遺伝カウンセラー
ジェネティックエキスパート
高精度な疾患ランキング機能を提供
特定の遺伝子に関連した疾患を対象に検索可
PubCaseFinder
Phenotype-Driven Differential-Diagnosis System
GA4GHへの貢献
Orphanet 約7000件
希少疾患
OMIM 約8000件
遺伝性疾患
PubMed 約100万件
症例報告
API
Matchmaker Exchange
PhenoTouch
症状入力支援機能
PubCaseSummary
検索結果まとめ機能
約2万件の検索（月間）