BLAH4, 10 Jan 2018

Identifying disease-phenotype associations using PubAnnotation and PubDictionaries for supporting diagnosis of intractable diseases

Toyofumi Fujiwara

Database Center for Life Science

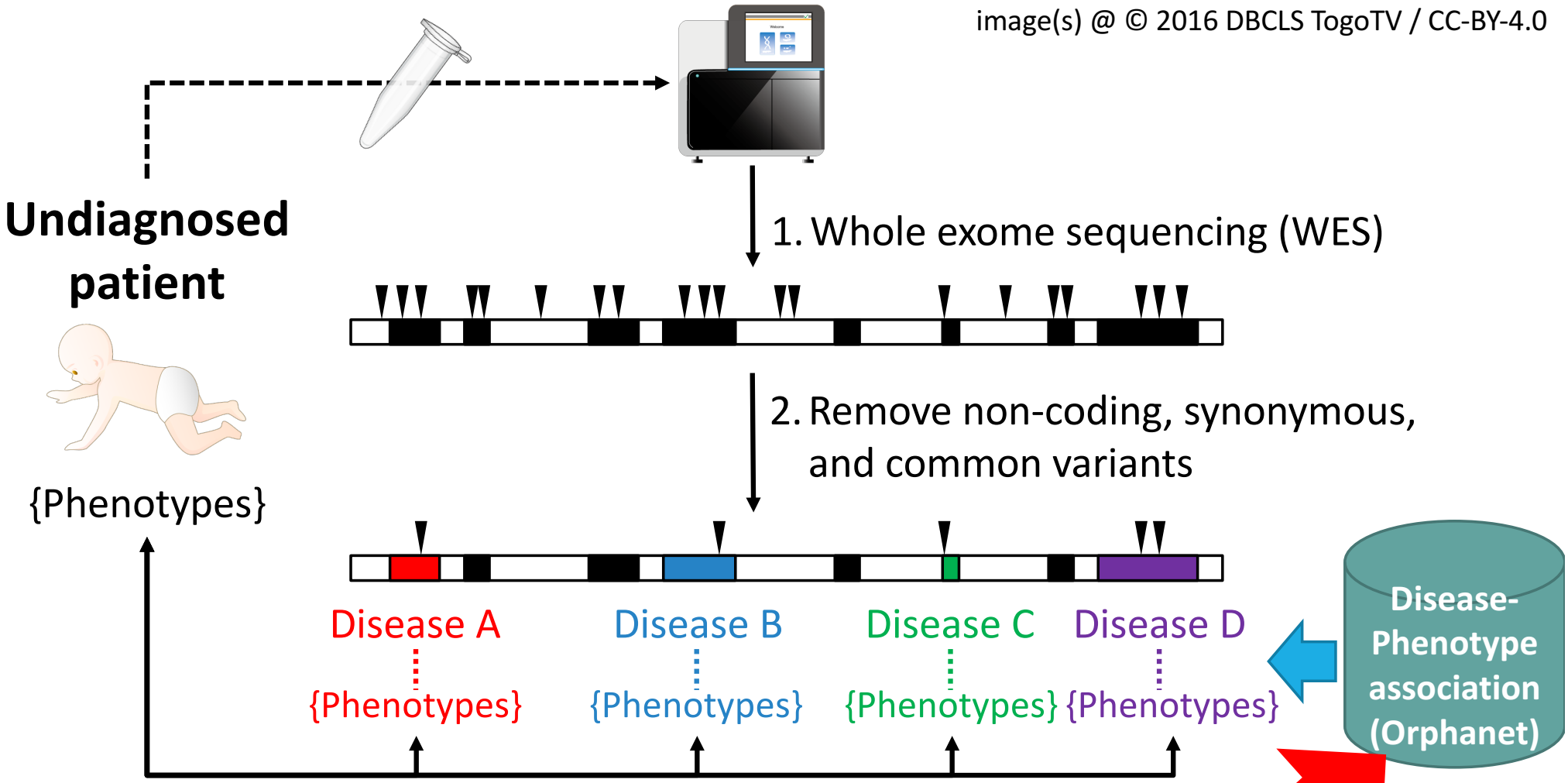
Background

- Japan has a long history of research for diagnosis of “rare and intractable diseases” (called “Nan-Byo”)
  - The Ministry of Health, Labour and Welfare of Japan (MHLW) designated 330 diseases as Nan-Byo
  - MHLW made MHLW Guidelines (330 documents) for diagnosis of Nan-Byo
  - MHLW supports Nan-Byo patients with medical expenses

There are many undiagnosed patients

Source : Adachi, Takeya, et al. "Japan’s initiative on rare and undiagnosed diseases (IRUD): towards an end to the diagnostic odyssey." European Journal of Human Genetics 25.9 (2017): 1025.

For undiagnosed patients



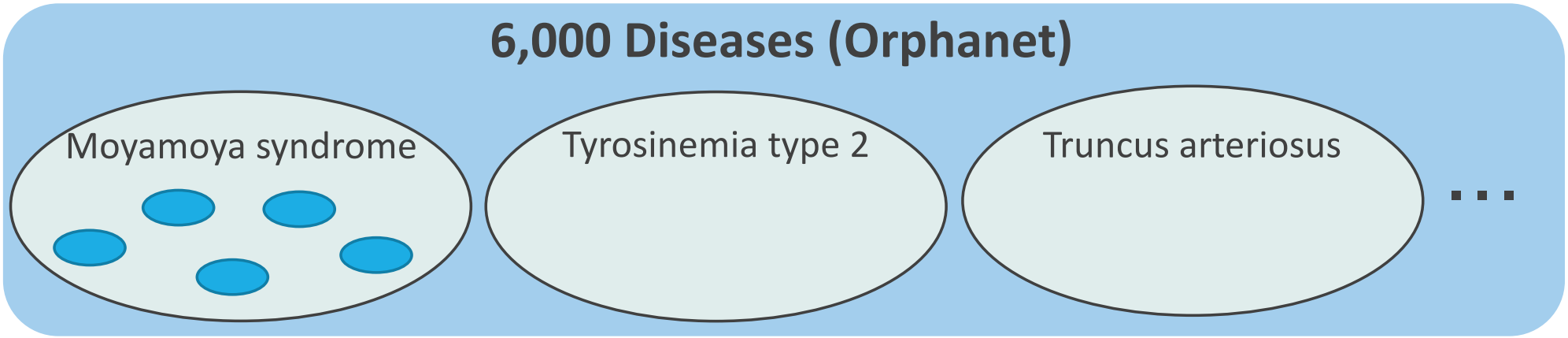
3. For further narrow down, compare the patient's phenotypes with t notypes of candidate diseases

Disease-phenotype associations are important for this process

Challenge

- But…
  - Manually curated databases like Orphanet inherently have a limited coverage

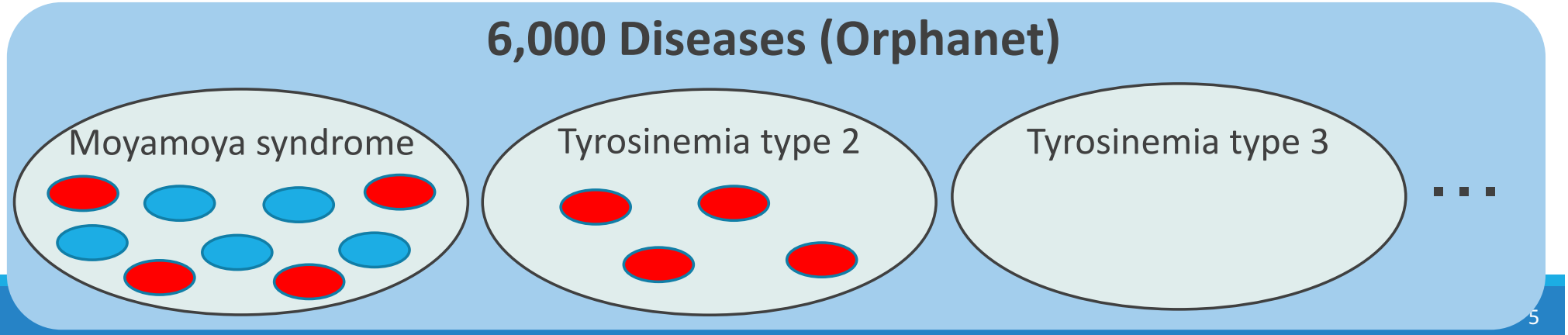
For example, **only 1/3** of rare diseases in Orphanet are associated with phenotypes



Prospective solution

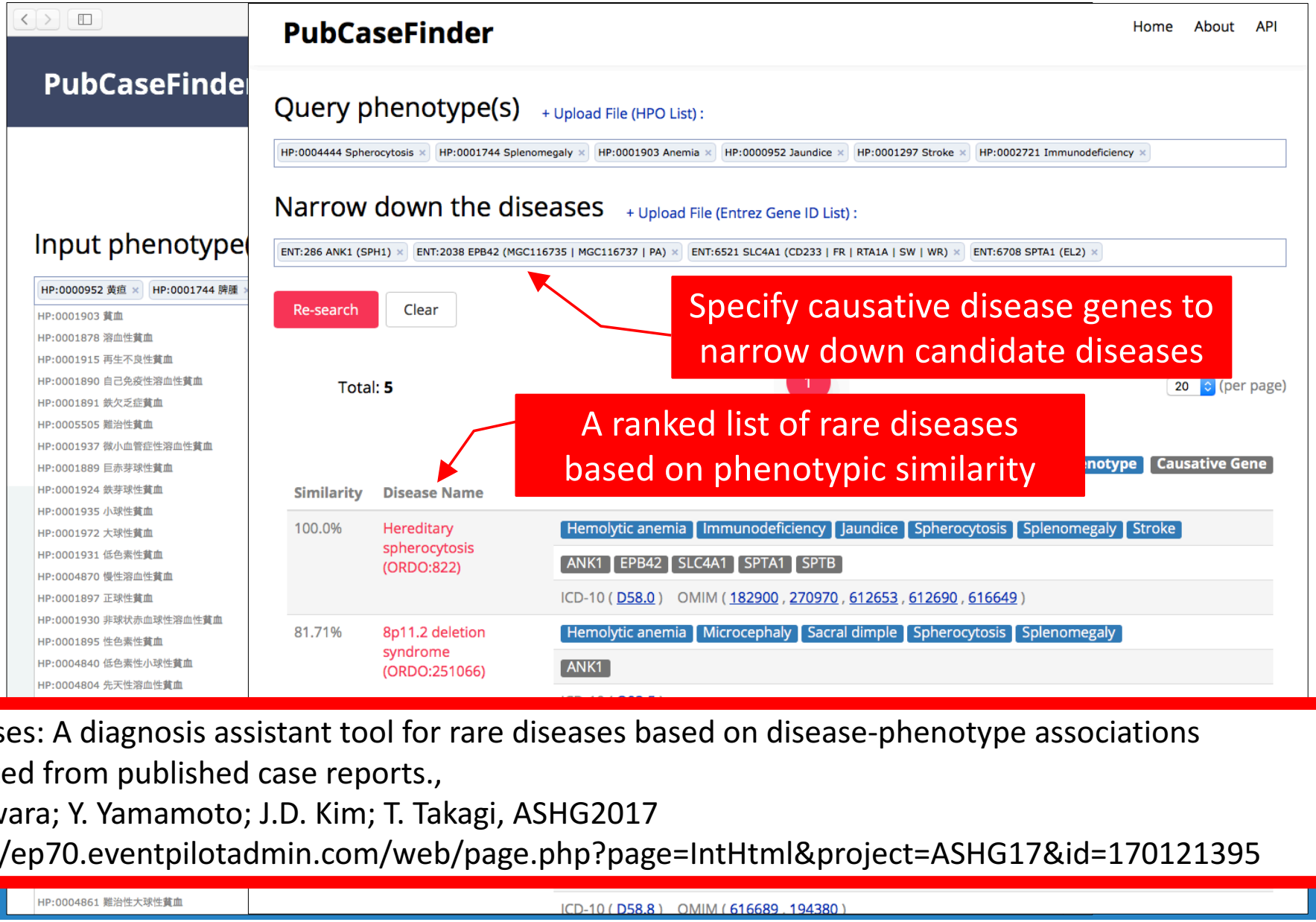
- Text mining-based approach is effective to improve the coverage of disease-phenotype associations
  - To automatically extract the associations from more than one million **case reports** in PubMed

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



PubCaseFinder

- A diagnosis support system for rare diseases

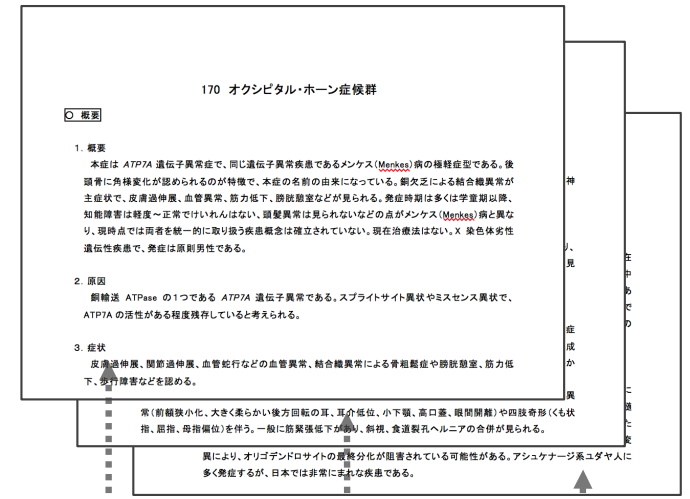


PubCases: A diagnosis assistant tool for rare diseases based on disease-phenotype associations extracted from published case reports.,
T. Fujiwara; Y. Yamamoto; J.D. Kim; T. Takagi, ASHG2017
https://ep70.eventpilotadmin.com/web/page.php?page=IntHtml&project=ASHG17&id=170121395

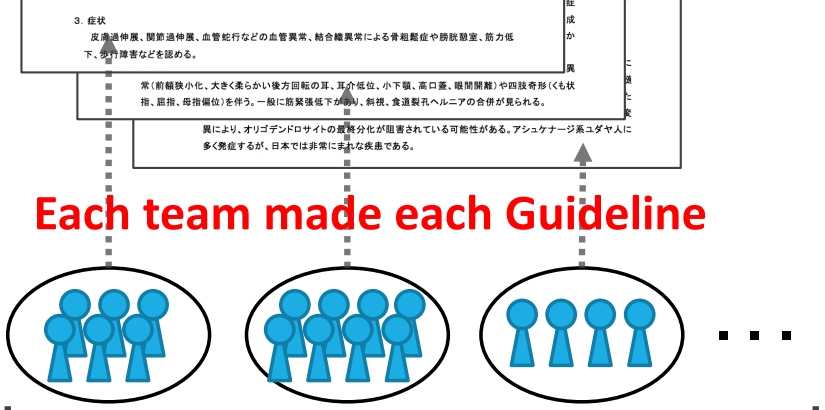
Goal (in this hackathon)

- For getting more accurate disease-phenotype associations
  - Extracting disease-phenotype associations from MHLW Guidelines that include diagnostic criteria for Nan-Byo

330 MHLW Guidelines



Each team made each Guideline



330 research teams (composed of physicians)

HPO Japanese translations

disease-phenotype associations (HPO)

Nan-Byo 1
- HP:0001203,…

Nan-Byo 2
- HP:0023091,…

Nan-Byo 3
- HP:0009808,…

⋮

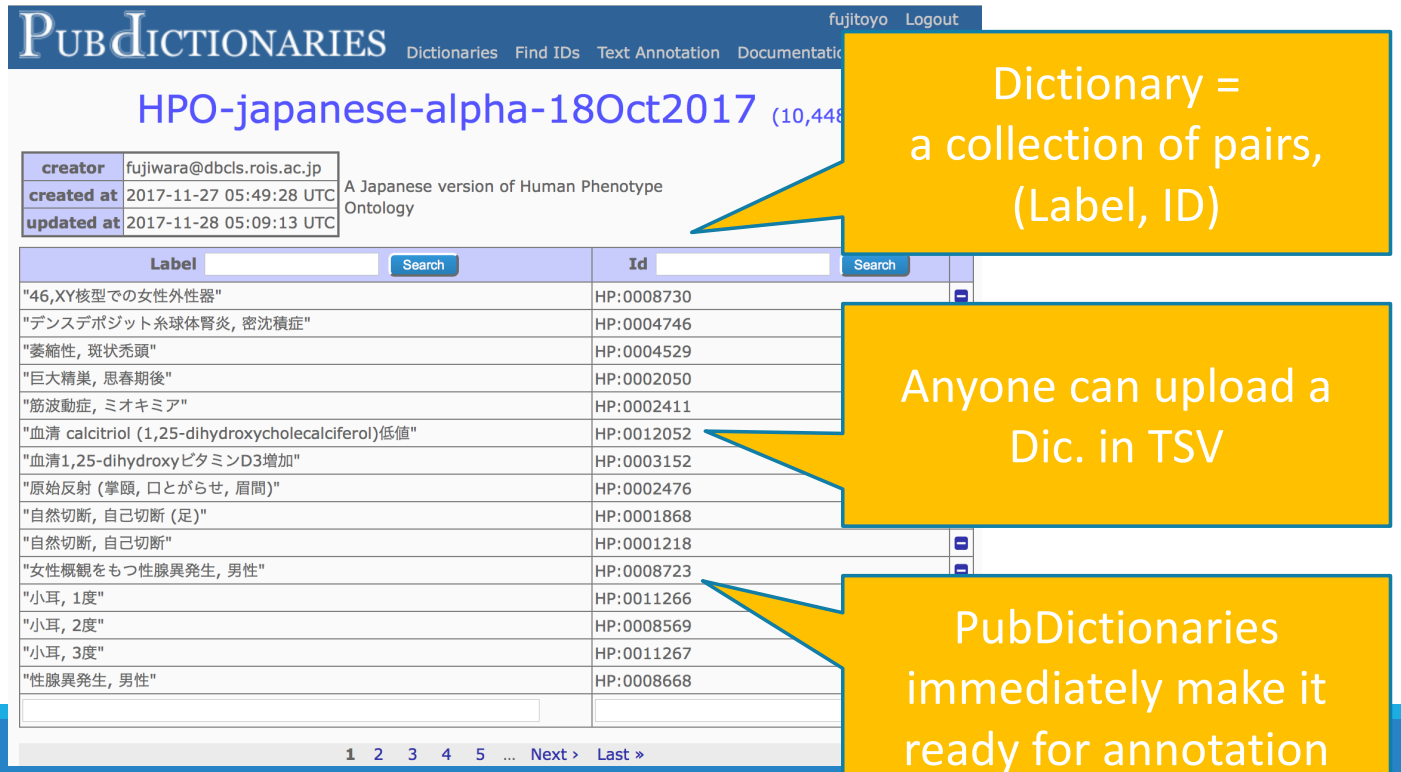
Nan-Byo 330
- HP:0012632,…

Preliminary study

- To extract disease-phenotype associations from MHLW Guidelines for 10 Nan-Byo using PubDictionaries and PubAnnotation

1. To upload HPO Japanese terms to PubDictionaries

HPO Japanese translations



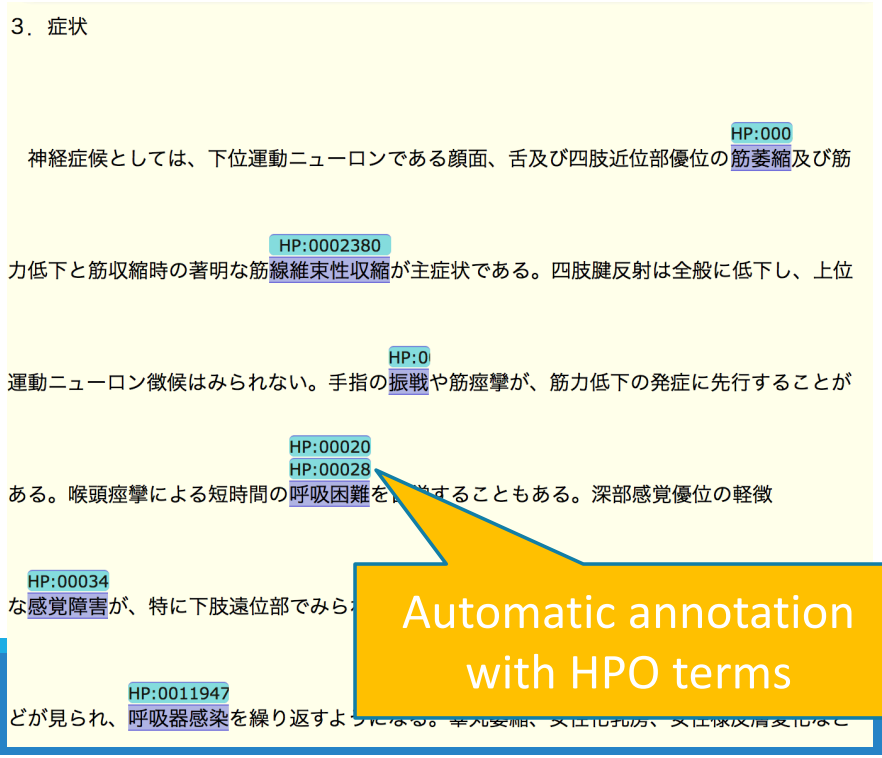
Preliminary study

2. To upload 330 documents to PubAnnotation
3. To get dictionary-based text annotation using a REST service of PubDictionaries

MHLW guideline documents

User can upload any documents

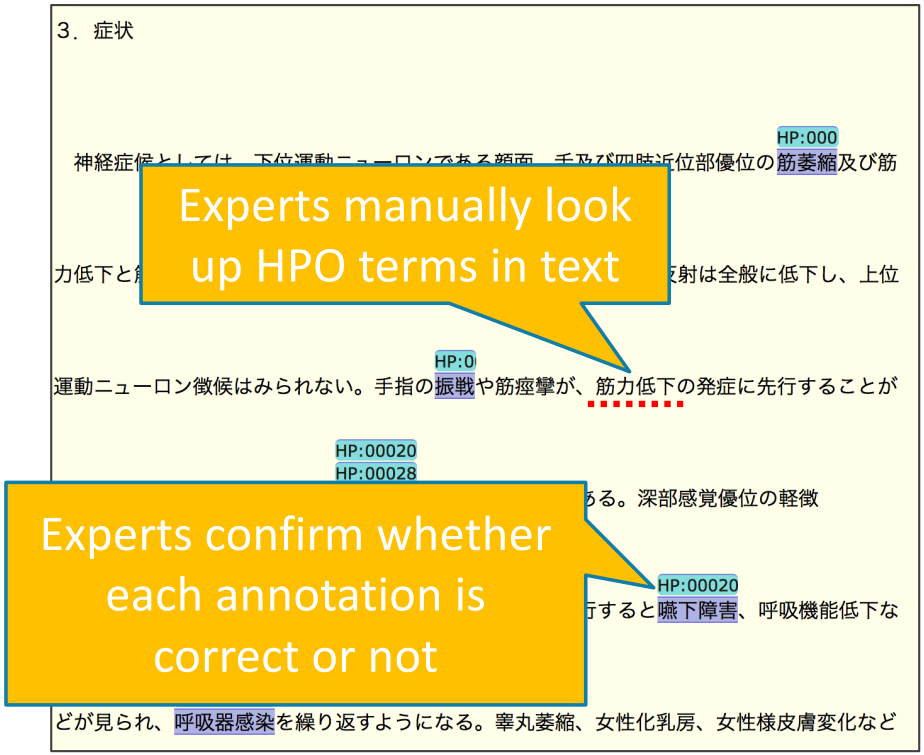
Annotation can be obtained from a REST service of PubDictionaries



Automatic annotation with HPO terms

Preliminary study

4. To manually correct dictionary-based text annotation (using TextAE, http://textae.pubannotation.org)
5. To collect a set of phenotype-associations for a Nan-Byo



Original text	Automatic annotation	Corresponding HPO ID	Corresponding HPO Label (Japanese)	Corresponding HPO Label (English)
筋力低下	×	HP:0001324	筋虚弱	Muscle weakness
筋萎縮	○	HP:0003202	筋萎縮	Skeletal muscle atrophy
振戦	○	HP:0001337	振戦	Tremor
嚥下障害	○	HP:0002015	嚥下障害	Dysphagia

Preliminary result

- For example;

  “moyamoya disease”-phenotype associations

  - Orphanet (by manual curation)

    → 5 HPO terms — Highly reliable

  - Published case reports (by automatic annotation)

    → 129 HPO terms — Less reliable

  - MHLW Guideline (by automatic annotation and manual correction)

    → 21 HPO terms — Highly reliable

Summary & Plan in this hackathon

Summary

- Extracting disease-phenotype associations from MHLW Guidelines that include diagnostic criteria for Nan-Byo
  - PubAnnotation and PubDictionaries facilitated getting automatic annotation and manual correction by experts

Plan

- The performance of automatic annotation can be improved by adding the newly identified terms to PubDictionaries

Acknowledgements

DBCLS

- Yasunori Yamamoto
- Kim Jin-Dong

University of Tokyo

- Toshihisa Takagi

Experts

- Chiharu Ohta
- Haruka Yoshii
- Hiromi Yamamura
- Noriko Nakabayashi

Funding Source

MEXT KAKENHI
(No. 221S0002)

DBCLS
Database Center
for Life Science