XVIII IPOPI GLOBAL PATIENTS' MEETING

an **IPOPI** event

16-19 OCTOBER 2024
MARSEILLE, FRANCE

#GPM2024

SUPPORTING ORGANISATIONS

ESID

INGID

SPONSORS

Biotest
From Nature for Life

Chiesi
global rare diseases

CSL Behring

Takeda

X4 Pharmaceuticals

IPOPI

IPOPI.ORG

XVIII IPOPI GLOBAL PATIENTS' MEETING
an IPOPI event

#GPM2024

SUPPORTING ORGANISATIONS

SPONSORS

IPOPI.ORG

Reducing the diagnosis wandering: the example of Dr Warehouse

Nicolas Garcelon

Head of data science platform

Institut Imagine

institut imagine
GUÉRIR LES MALADIES GÉNÉTIQUES

Disclosure

Co funder of CODOC, Spin Off of Imagine Institute

Scientific advisor for Takeda

Objective: To find patients with rare diseases in hospitals who have not been diagnosed.

Conditions to success:

No additional work for the physicians => Based on raw data collected in hospitals

Ability to screen all the patients of a hospital

⇒ What type of data should we use?

⇒ How to access all the data of all the patients?

⇒ How to compute the data to find the patients?

⇒ How to find the patients? Especially with few patients

Getting access to all data through an hospital data warehouse

Patients data in hospitals:

- Administrative data
- Clinical reports (free text): consultation, hospitalisation, imaging, etc
- Lab results: structured data
- Specific forms

⇒ Most of the phenotypic information are only in free texts (Neuraz et al 2020)

⇒ We built in 2014, **the first data warehouse oriented toward documents**, inspired by Google search engine, to help physicians to easily search for patients: **Dr Warehouse®**.

Neuraz A et al.Natural Language Processing for Rapid Response to Emergent Diseases: Case Study of Calcium Channel Blockers and Hypertension in the COVID-19 Pandemic. J Med Internet Res. 2020 Aug 14;22(8):e20773.

Getting access to all data through an hospital data warehouse

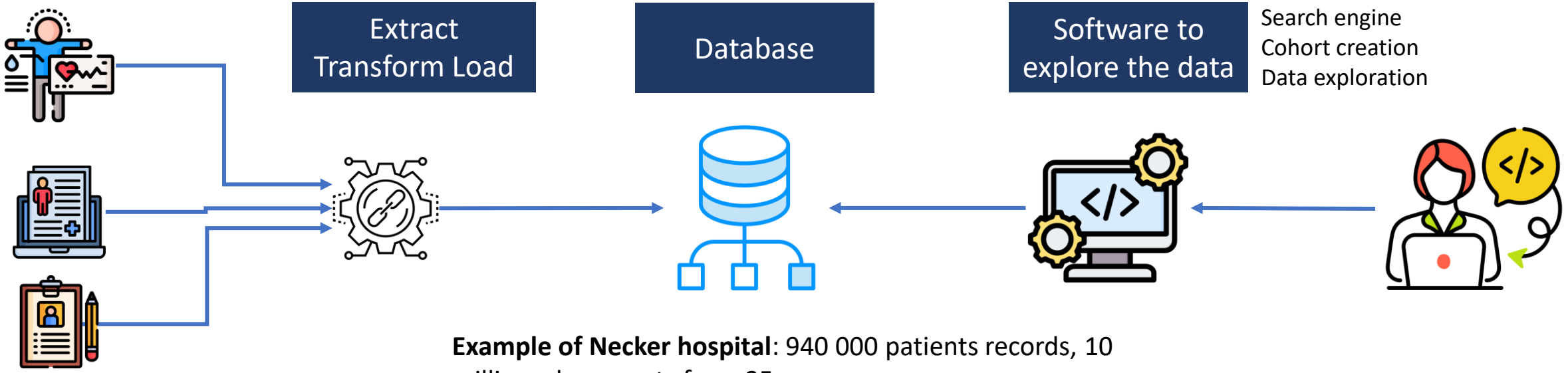
Dr Warehouse®

Extract Transform Load

Database

Software to explore the data

Search engine
Cohort creation
Data exploration



Hospital data

Example of Necker hospital: 940 000 patients records, 10 millions documents from 35 sources

Creation of codoc in 2017, Imagine's spin off to industrialise and deploy Dr Warehouse® in other hospitals

End user

Garcelon N, Neuraz A, Salomon R, Faour H, Benoit V, Delapalme A, Munnich A, Burgun A, Rance B. A clinician friendly data warehouse oriented toward narrative reports: Dr. Warehouse. J Biomed Inform. 2018 Apr;80:52-63.

Primary data search engine

Dr. Care | Accueil | Recherche | Cohortes | Tableau de bord | Admin | Rechercher un patient

Recherche

Population — 4 651 — 4 651 dans l'EDS

A Critère Textuel — 78 — 78 dans l'EDS

"rett syndrome" OR FOXG1

Pour rechercher la phrase complète, ajoutez des guillemets.

Étendre aux synonymes

› Filtres avancés

+ Ajouter un critère

Rechercher

Using AI to improve search (negation detection and family history)

Capture Suite codoc,
Example with fake data

Résultats

78
811

Récapitulatif de la recherche — Ajouter à la bibliothèque

A Documents contenant **"rett syndrome" OR FOXG1**, en excluant les négations.

Dossiers patients | Démographie | Biologie | PMSI

Alimenter une cohorte

GILLETTE Rebecca
F, 39
Née le 10/08/1959 - 63 ans

Brain & development, le 01/11/2008, par Dr Harada Koto...
Hypoplastic hippocampus in atypical Rett syndrome with a...

Proceedings of the National Academy of Sciences of the...
sensitivity in a rat model of Rett syndrome. Children with

Advances in experimental medicine and biology, le...
gentle transactivation of the Foxg1 transcription factor gene,...

› Afficher plus 7

KNOWLES James
H, 40
Né le 29/01/1954 - 69 ans

Stem cell investigation, le 14/04/2012, par Dr Gomathi...
Novel therapeutic approaches: Rett syndrome and human...

Developmental neurorehabilitation, le 15/06/2007, par D...
Early development in Rett syndrome - the benefits and...

Epilepsia, le 30/03/2007, par Dr Lim Zhan - Pubmed
ting with features similar to Rett syndrome. Cardinal features ...

› Afficher plus 7

LEE David
H, 41

Current biology : CB, le 14/04/2012, par Dr Franco Luis M ...
ired inhibition in autism and Rett syndrome, these findings...

Stem cell research, le 14/04/2012, par Dr Hunihan Lisa -...
he mutant MECP2 allele from a Rett Syndrome patient fibrobl...

1 2 3 4 — 78 résultat(s) / 20 par page

Brain & development ×

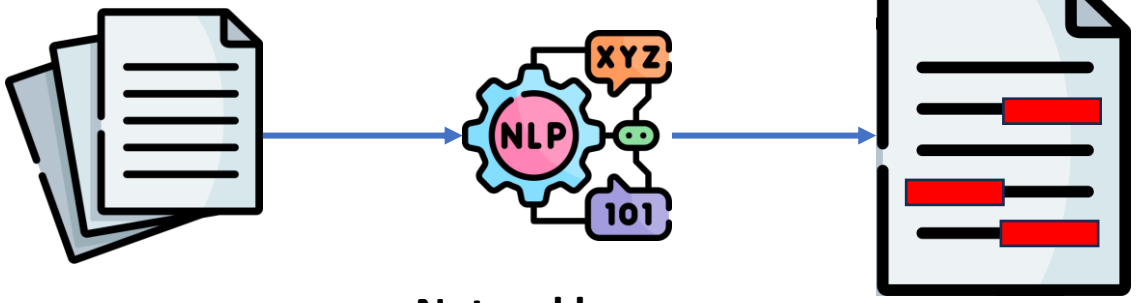
Brain & development
Le 01/11/2008 par Dr Harada Kotoha
GILLETTE Rebecca - Née le 10/08/1959 (63 ans)

Hypoplastic hippocampus in atypical Rett syndrome with a novel FOXG1 mutation.

The forkhead box G1 (FOXG1) gene encodes a brain-specific transcription factor and is associated with a congenital variant of atypical Rett syndrome (RTT); several FOXG1 mutations have been identified. The congenital variant of RTT shows a hypoplastic corpus callosum, delayed myelination, and frontal and temporal atrophy. Although no report has described a hippocampal abnormality in humans, the current study suggests that FOXG1 also regulates neurogenesis in the postnatal hippocampus. In the present case, severe developmental delay was observed in a patient with a congenital variant of RTT from about 4months, in conjunction with acquired microcephaly, hypotonia, limited motor function, absent purposeful hand use, and repetitive jerky movements of the upper limbs. A novel missense mutation was identified in FOXG1 on gene analysis (c. 569T>A, p. Ile190Asn). The patient showed not only the typical cerebral abnormalities of a congenital variant of RTT, but also a hypoplastic hippocampus. This

How to make free text machine-readable (and human-readable) ?

Automatic extraction of clinical signs and diagnosis from hospital reports



Natural language processing

Clinical signs	Polarity	Context
• Fever	Affirmation	Patient
• Anaemia	Negation	Patient
• Immuno-deficiency	Affirmation	Family
• infection	Affirmation	Patient

Automatic phenotyping & normalisation with international thesaurus (HPO)

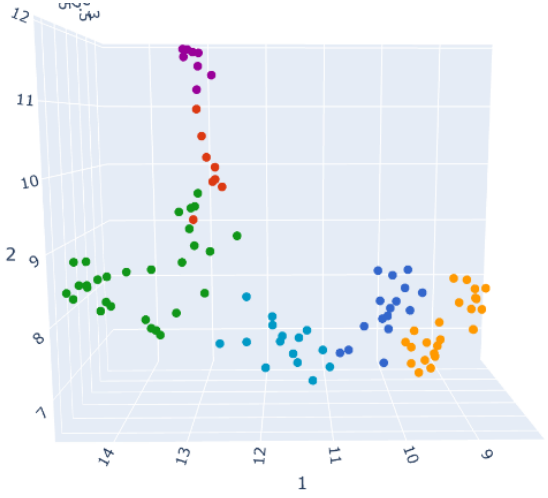
- Each patient is represented by a list of **clinical signs, phenotypes, diagnosis** extracted from their reports
- With **frequencies** of the phenotypes, **age** at first mention, **polarity** and **context**.
- In Necker Hospital we extracted a total of **50 millions phenotypes** from 5 millions hospital reports.

Garcelon et al, Orphanet J Rare Dis, 2018 ; Vincent M et al, MedInfo, 2022

What can we do with computational patient records ?

Automatic description of cohorts based on the aggregation of phenotypes (PheWAS)

Compute **similarity** between patients for **clustering** or **diagnosis**



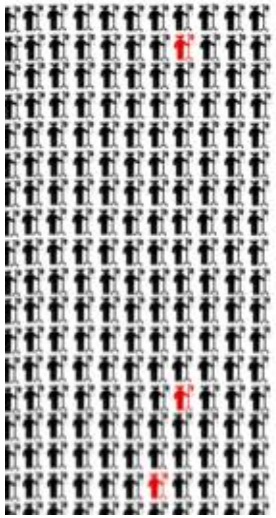
Show 10 entries Search:

Order	Concepts	# patients	See	FreqRes	TF-IDF	PSS	Case-Weighted PSS	Median age
1	apds	49		100	21.44	58.3	23336.94	12.12
2	deficit immunitaire	44		89.8	3.99	.2	27.86	10.25
3	Adenopathie	42		85.7	4.86	.2	25.99	10.95
4	Toux	37		75.5	1.64	.1	8.66	9.63
5	Lymphopenie	36		73.5	2.47	.3	41.98	11.92
6	infections	35		71.4	1.32	.2	24.52	10.31
7	Fievre	33		67.3	.9	0	0	11.25
8	lymphoproliferation	33		67.3	4.82	3	582.41	10.57
9	Otite	33		67.3	1.24	.1	8.76	8.865
10	TCMH	31		63.3	1.07	.1	8.39	13.07

Showing 1 to 10 of 1,195 entries Previous 1 2 3 4 5 … 120 Next

Lo Barco T et al, epilepsia, 2023 ; Yang D et al Orphanet J Rare Dis, 2022 ; Chen et al 2024

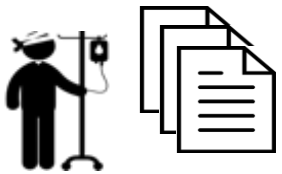
Reducing diagnosis wandering through phenotypic similarity



In Necker Hospital data warehouse

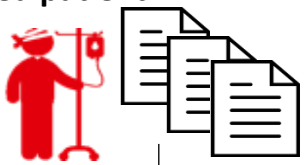
Necker ENFANTS MALADES HÔPITAL UNIVERSITAIRE

940,000 patients



Concepts	Nb	TF-IDF
Diabetis	2	0.02
Cataract	4	0.2
Fever	5	0.001
Proteinuria	3	0.02
Cytopenie	2	0.03
Asthenia	1	0.01

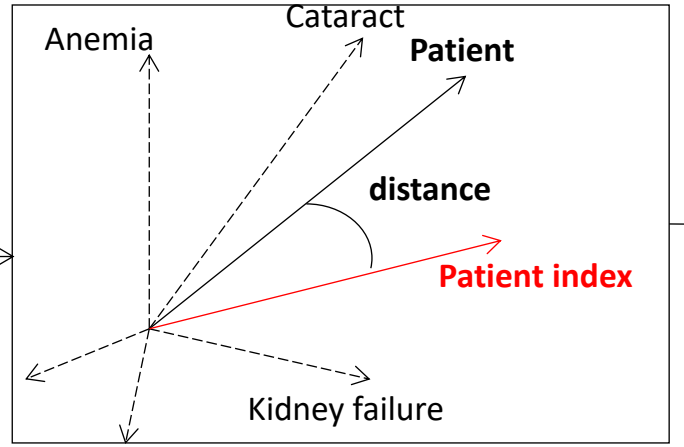
Diagnosed patient



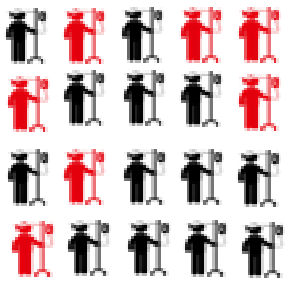
Concepts	Certainty
Lowe syndrome	1
Anemia	-1
Kidney failure	1
Cataract	1

if ≥ k concepts in common

Similarity : cosine



Top30 similar patients



Evaluation on 5 rare diseases: Lowe Syndrome, Dystrophic Epidermolysis Bullosa, Activated PI3K delta Syndrome, Rett Syndrome, Dowling Meara

In average 51% patients with same rare disease in top30

Improvement of the methods on ciliopathies with the C'IL-LICO project (Faviez & Chen)

Garcelon N, Neuraz A, Benoit V, Salomon R, Kracker S, Suarez F, Bahi-Buisson N, Hadj-Rabia S, Fischer A, Munnich A, Burgun A. Finding patients using similarity measures in a rare diseases-oriented clinical data warehouse: Dr. Warehouse and the needle in the needle stack. J Biomed Inform. 2017 Sep;73:51-61. doi: 10.1016/j.jbi.2017.07.016.

Similarity algorithm available for Physicians in the hospital

Genetics inMedicine

www.nature.com/gim

Check for updates

BRIEF COMMUNICATION

Deep phenotyping unstructured data mining in an extensive pediatric database to unravel a common *KCNA2* variant in neurodevelopmental syndromes

Marie Hully[1], Tommaso Lo Barco[1], Anna Kaminska[1,2], Giulia Barcia[1,3], Claude Cances[4], Cyril Mignot[5], Isabelle Desguerre[1], Nicolas Garcelon[6,7], Edor Kabashi[8] and Rima Nabbout[1,8]✉

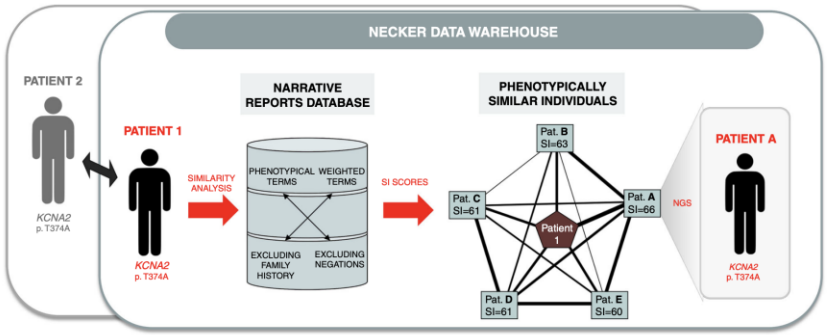


Fig. 1 Display of the two patients (patient 1 from our institution and patient 2 from another institution in our reference center network) sharing the same phenotype and the same *KCN2A* variant. Similarity analysis with all data warehouse narrative reports was performed, yielding a high similarity index (SI) in five patients (patients A–E). Exome sequencing validated that patient A, who had the highest SI, harbored the same *KCNA2* variant. NGS next-generation sequencing.

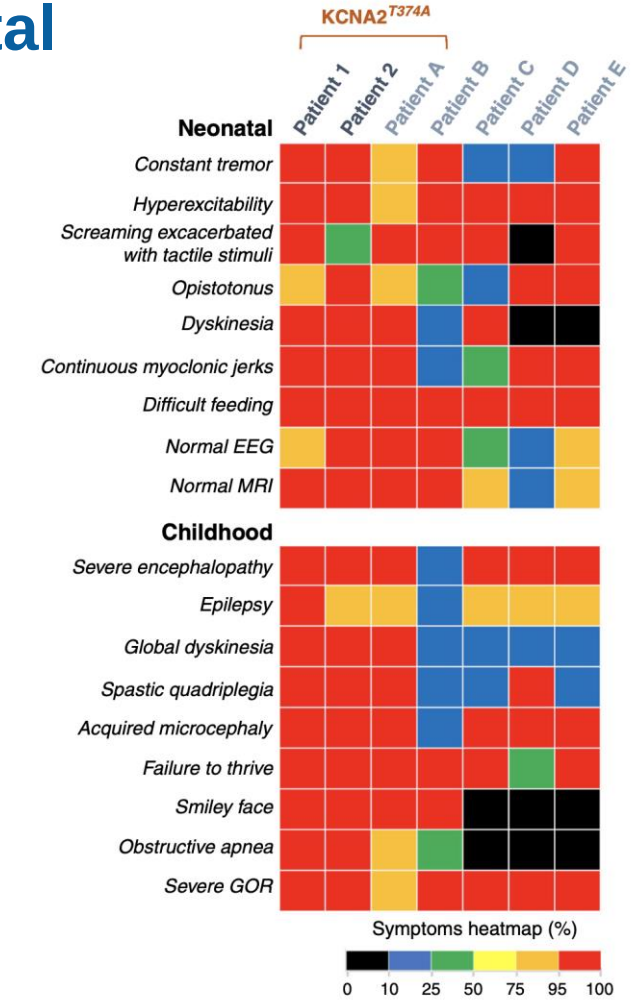


Fig. 2 Clinical heat map describing the detailed characteristics of the patients in this study. Heatmap for patient 1 and 2 with

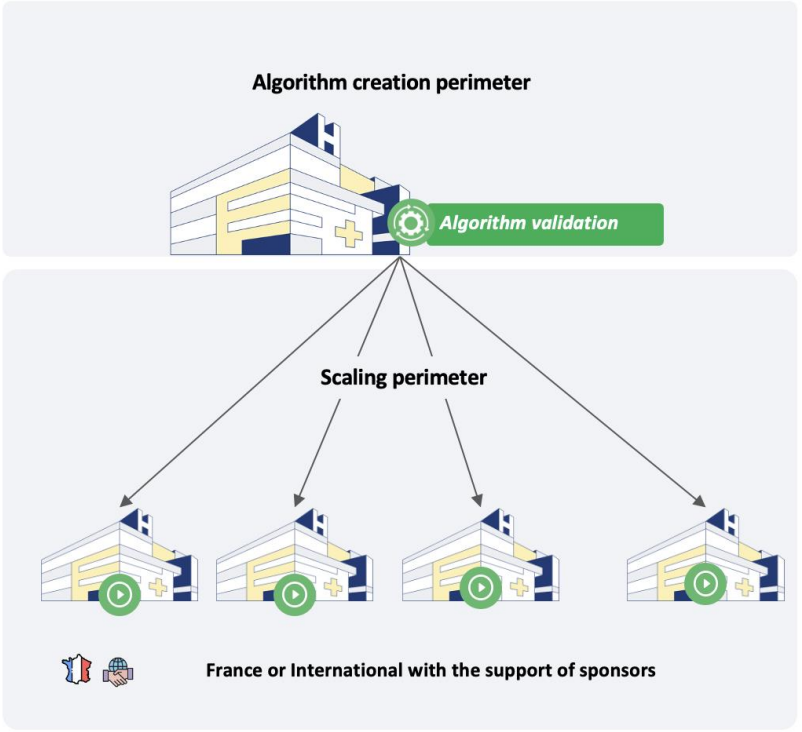
Hully et al Genetics in Medicine 2021

How to export this concept in other hospitals ?

Spin off **codoc** has **industrialised** Dr Warehouse and already **deployed** it in **several hospitals in France and Belgium**.

We want to use this **network to apply the similarity algorithms in these hospitals.**

But we cannot use real patients (confidentiality, ethic).

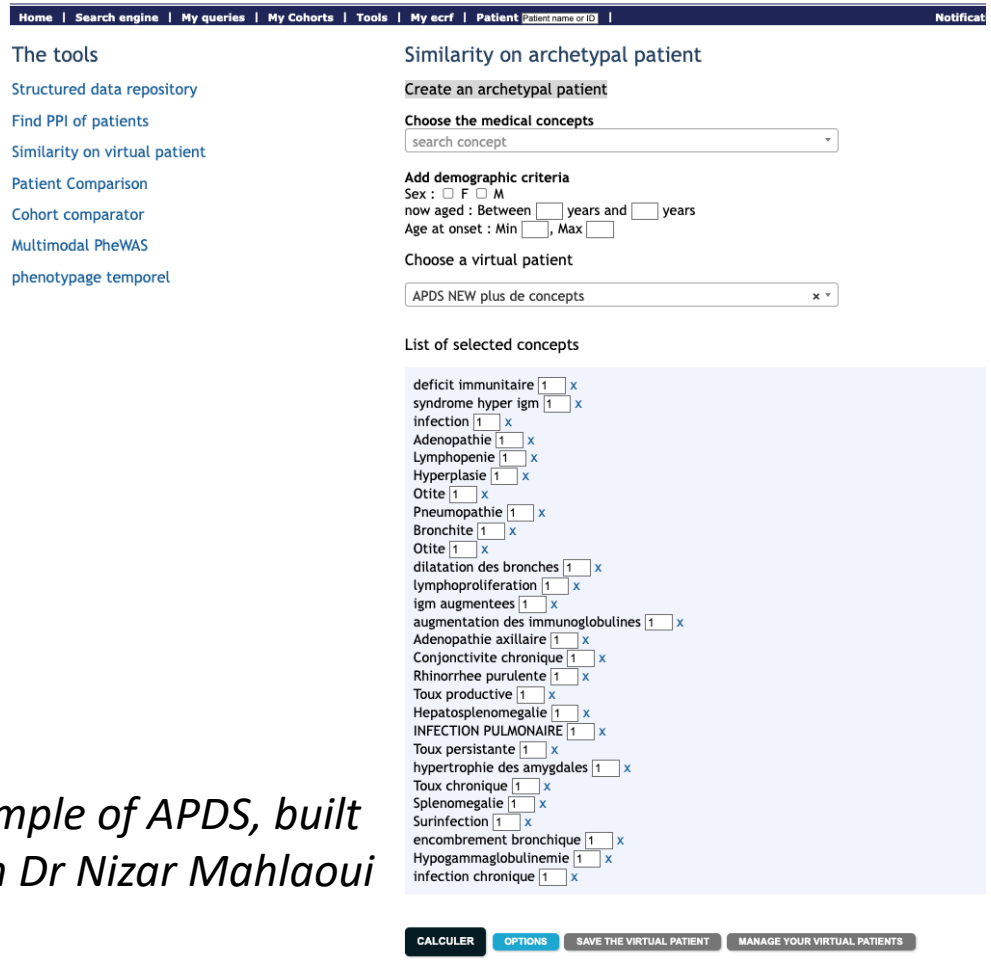


Build an archetypal patient for a specific disease

Instead of computing similarity with an existing patient, creation of **a disease specific archetype**

Based on:

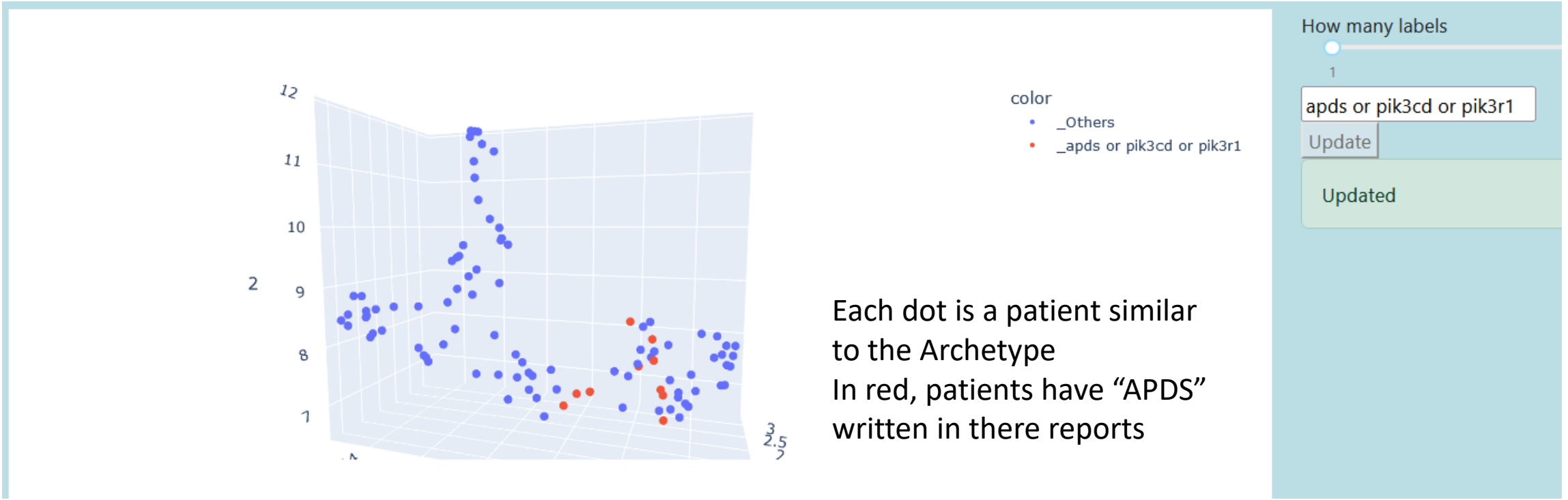
- Experts and physicians
- Knowledge databases (Orphanet, OMIM etc)
- Real world data (health reports)
- Patient knowledge



Example of APDS, built with Dr Nizar Mahlaoui

Screen all patients in a hospital data warehouse

Display the top 100 most similar patients, and cluster on the results



Each dot is a patient similar to the Archetype
In red, patients have "APDS" written in there reports

Conclusion

Aim : to find undiagnosed patient in hospitals

1) Build a data warehouse to gather all the data in several hospitals ✅
2) Transform patient data to make it computable : natural language processing methods to extract clinical signs ✅
3) Create an archetype for specific diseases (e.g. APDS) ✅
4) Develop an algorithm to compute similarity between the archetype and all the patients of a hospital data warehouse ✅ but to adapt and evaluate for each disease
5) Apply the algorithm in hospitals and explore the top N patients maybe undiagnosed ⏭️

#GPM2024

SUPPORTING ORGANISATIONS

SPONSORS

IPOPI.ORG