Summary

Data Challenges for Health

WELCOME TO THE DATA CHALLENGES FOR HEALTH EVENT !

2:15 – 6:00

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Data Challenges for Health - Introduction

Data Challenges for Health

David Sainati

eHealth projects director at the Digital Health Ministerial Delegation

David Sainati has a PhD in pharmacy and a graduate of HEC Paris. He founded and managed a digital health start-up for 8 years. With this entrepreneurial experience, he joined the Ministerial Delegation for eHealth (DNS) in 2019 as a project manager and he is in charge of national projects related to entrepreneurship and compliance.

Pr Hossein Khonsari

Medical director of the Health Data Hub

Prof. Roman Hossein Khonsari is Medical Director of the Health Data Hub. Professor of Plastic and Maxillofacial Surgery at the University of Paris and Maxillofacial Surgeon at the Necker - Enfants Malades Hospital, trained in Nantes and Paris, he also holds a master's degree in biology from the Ecole Normale Supérieure de Paris and a PhD in craniofacial development from King's College London.

2:20 – 2:30

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Data Challenges for Health - Introduction

Thoracic Surgery Challenge
Classification of lung tumours from cellular and clinical imaging data

Dermatopathology Challenge
Melanoma relapse risk prediction from pathology data

Allergology Challenge
Machine learning on biological data from allergology

Lung Transplant Challenge
Development of an algorithm capable of detecting graft rejection following transplantation from cellular imaging data

Rare Dental Diseases Challenge
Detection of rare oral and dental diseases from photographs of teeth

Nuclear Medicine Challenge
Diagnosis of Parkinson's syndromes from brain imaging data

Machine Learning and Data Challenges

Data Challenges for Health

Aymeric Floyrac

Data Scientist at Health Data Hub

INSEE administrator, Aymeric Floyrac has been working as a data scientist at Health Data Hub for nearly two years. He works with HDH pilot projects on machine learning pipelines and on a variety of topics including synthetic data generation and data pseudonymization.
Aymeric Floyrac graduated from Ecole polytechnique in applied mathematics in 2020.

Maximilian Oppelt

Research Scientist at the Fraunhofer Institute for Integrated Circuits

Maximilian P. Oppelt is a Research Scientist, at the Fraunhofer Institute for Integrated Circuits and a PhD Candidate at the Machine Learning and Data Analytics Lab of the Friedrich Alexander University. After working for medical device manufacturers, he joined the academic world again and is now working on novel end-to-end deep learning methods to extract stable representations of biosignals. He participated in past data challenges himself and is now supervising students and young researchers in upcoming competitions.

2:30 – 2:45

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

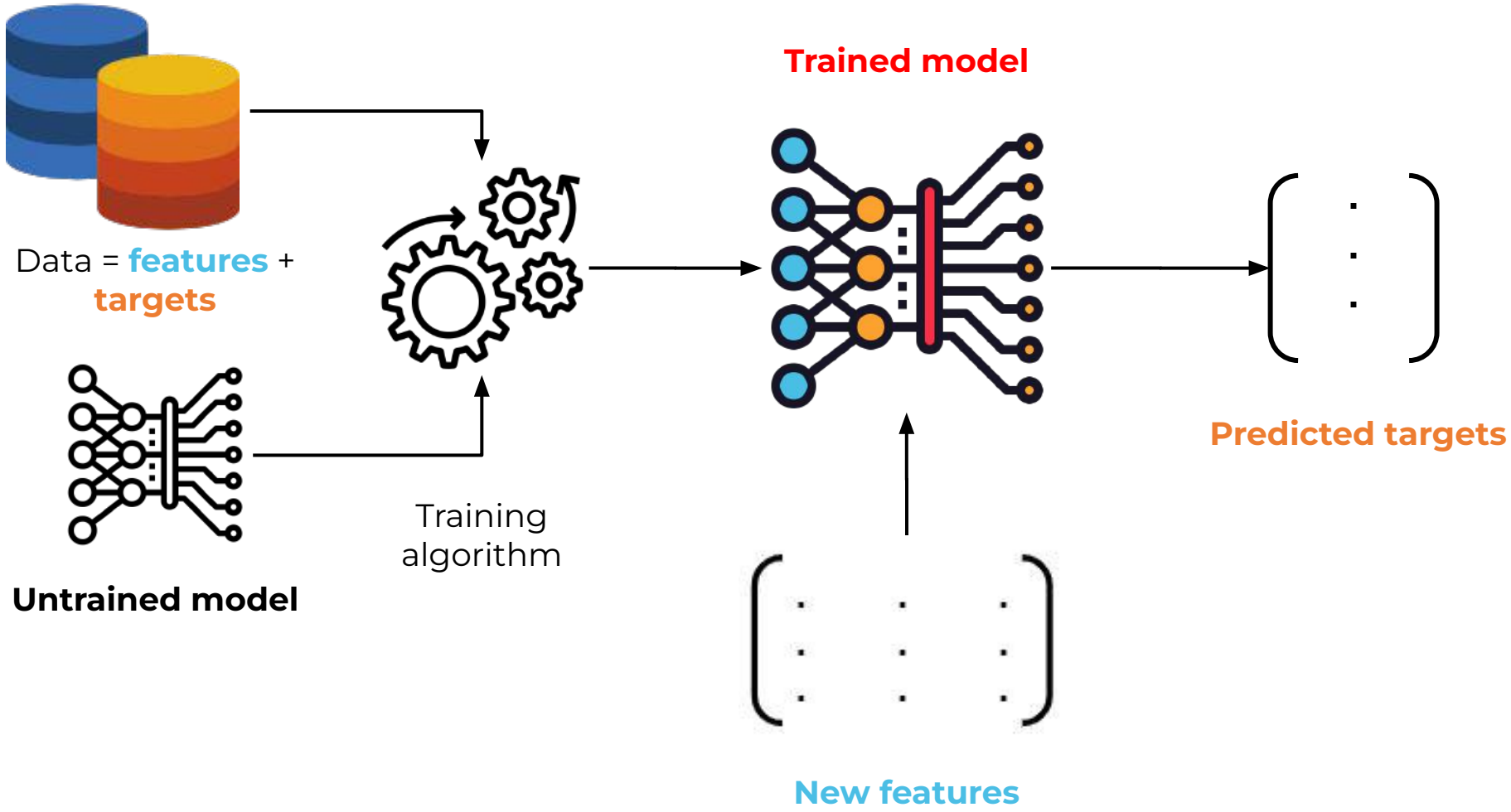
FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

HEALTH DATA HUB

Machine learning
Health Data Hub

24th May 2022

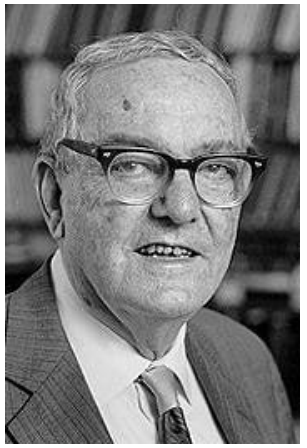
Machine learning: what and what for?



A little bit of history

“Learning is any process by which a system improves from experience.”

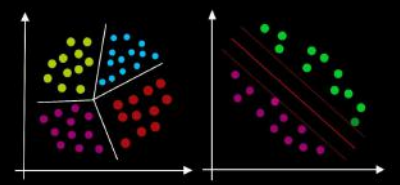
Herbert Simon



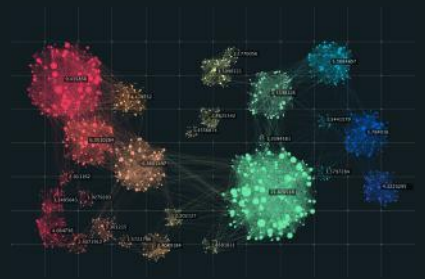
1950s Simple algorithms → **1960s** Bayesian methods → **1970s** AI winter → **1980s** Resurgence in ML research → **1990s** Data driven approach → **2000s** SVM, ensembling, boosting → **2010s** Deep learning supremacy

A little bit of theory

Supervised learning



Unsupervised learning



Reinforcement learning

Some popular applications

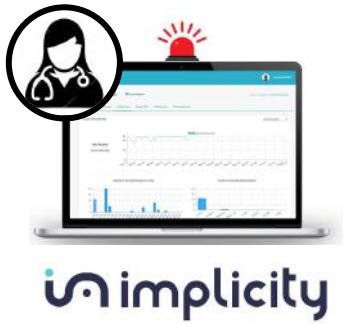
- Natural Language Processing
- Computer Vision
- Speech Recognition
- Recommendation systems
- Generative models
- Trends prediction

Machine learning in action

Context

Problem

A chronic disease: heart failure
A downward spiral of hospital stays
Most hospital stays could be prevented



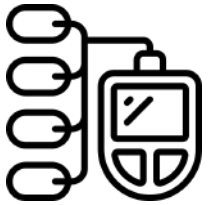
Solution

Use remote monitored pacemakers
Deploy an alerting system for cardiologists

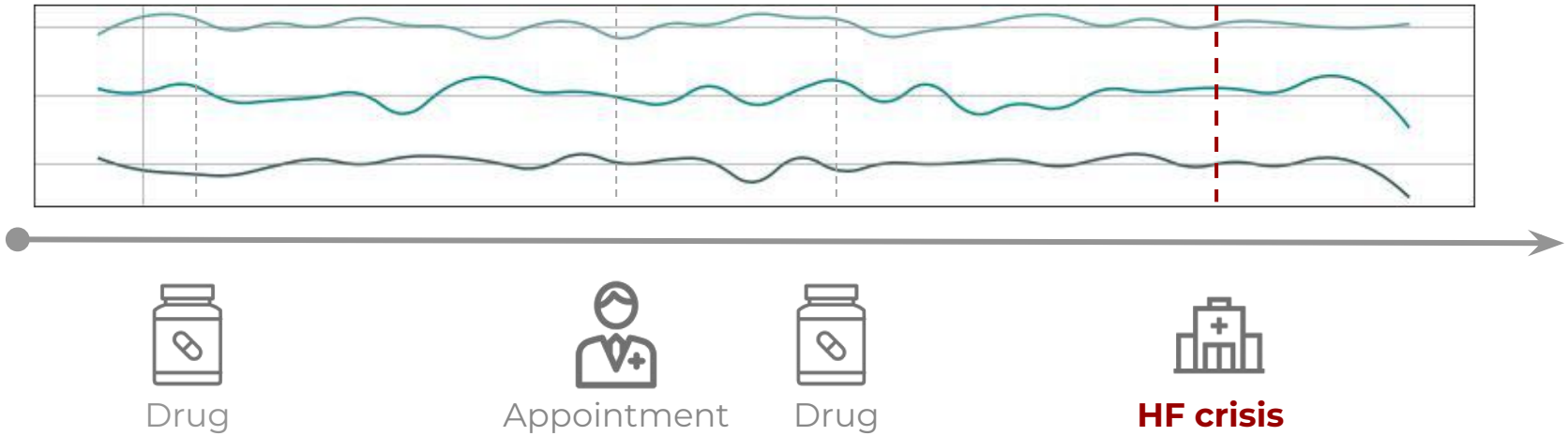
Machine learning in action

Data and model training

Connected medical device

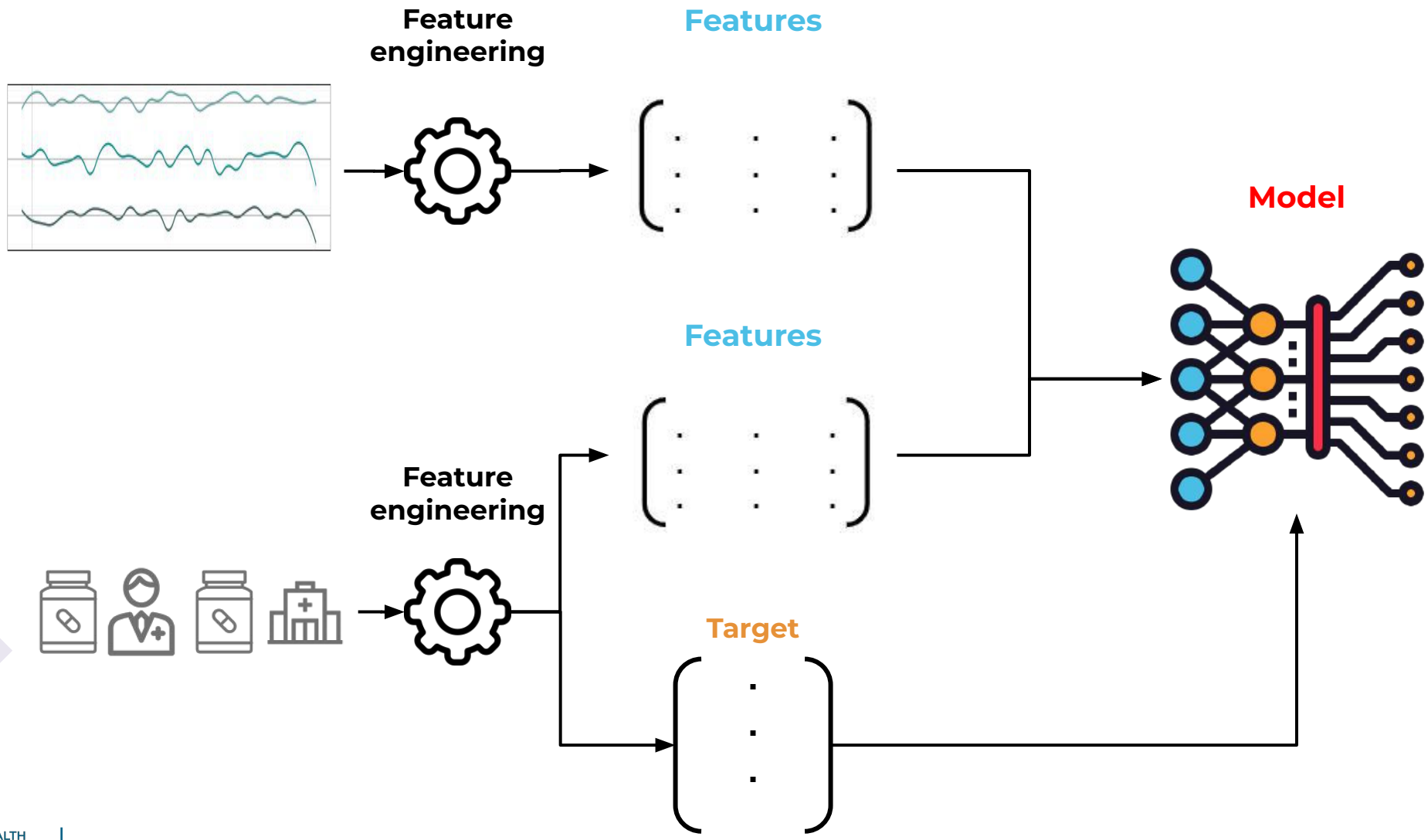


Claims (national health insurance)



Machine learning in action

Data and model training



Thank you

Maximilian P. Oppelt

Data Challenges for Health
A Data Scientist Perspective

Fraunhofer IIS

Fraunhofer Institute for Integrated Circuits IIS

FAU Friedrich-Alexander-Universität Erlangen-Nürnberg

MACHINE LEARNING & DATA ANALYTICS

The Data Scientist Perspective

Data Challenges Can Be Life Changing For Engineers

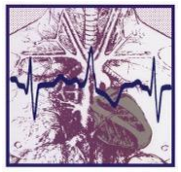
Data Challenge: My Journey

B.Eng Mechatronics

M.Sc. Medical Engineering

Image Processing Software Engineer
Image Quality Engineer

Research Scientist
Physionet 2020 Computers in Cardiology Challenge on ECG Arrhythmia Detection



PhD Candidate

Institute for Integrated Circuits & MaD Lab

- Biomedical Signal Analysis
- Software Engineering and Embedded Design
- Clinical Studies
- Machine and Deep Learning

Motivation for Participation in Data Challenges

Key Messages

1. Employ Novel Methods To Real Problems
   Explore New Domains, Utilize Novel Theoretical Algorithms

2. Compete With Others
   Leaderboards provide instant Feedback, Extra Motivation to Proceed

3. Work Collaborative and Exchange With Other Researchers
   Compare Solutions, Exchange Ideas, Meet AI-Residence and Help Beginners

Combining Scatter Transform and Deep Neural Networks for Multilabel Electrocardiogram Signal Classification

Maximilian P Oppelt[1], Maximilian Riehl[1], Felix P Kemeth[2], Jan Steffan[1]
[1]Department of Image Processing and Medical Engineering, Fraunhofer IIS
[2]Department of Chemical and Biomolecular Engineering, Whiting School of Engineering, Johns Hopkins University

Recipes For Designing a Great Data Challenge

Insights from A Challenger

Building a portfolio of projects! Win price money! Learn new techniques and explore new domains! Benchmark your skill.

1. Participants can help refine the goal and discuss the metric. This increases identification with the challenge and motivate competitors on the long run and encourage data scientists from diverse backgrounds.

2. Build A Community: Google Groups, Discord, Live Events or Forums. This helps competitors too keep motivated in phases of slow progress.

3. Choose your metric wisely: Classic machine learning metrics such as accuracy might not lead to most desirable outcome and choose your validation/test-set wisely!

4. Keep the pedagogical motivation of competitors in mind: Young researchers will compete to learn!

Participating in Future Challenges
Skills and Tools and Requirements

Tips for Participants
- Read the rules and report flaws.
- Make sure the challenge topic is a good fit.
- Network and communicate with other teams.
- Start simple and submit continuously.
- Try to gather domain knowledge and work with the data.
- Learn from similar challenges and publications.
- Track you experiments.
- Have Fun!

Participation in Future Competitions
- Compete with students and new team members for training and faster onboarding.


Fraunhofer
IIS
Fraunhofer Institute for Integrated Circuits IIS
Have Fun!

Data Challenge SFP-SFD : Visio-Mel - Search for a digital signature evaluating the risk of metastatic evolution of localized melanomas

Data Challenges for Health

2:45 – 2:55

Dr Frédéric Staroz

Surgical pathologist at OuestPathologie, Quimper, France

Frédéric Staroz is a dermatopathologist in his practice. He is president of the national professional council in pathology, involved in medical training and innovation within his specialty. He is also a member of the French Society of Pathology and as such coordinates the VisioMel DataChallenge supported by the HDH and the BPI. VisioMel aims to develop a digital fingerprint predictive of recurrence in melanoma.

GOUVERNEMENT Liberté Égalité Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22 PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Search for a digital signature evaluating the risk of metastatic evolution of localized melanomas

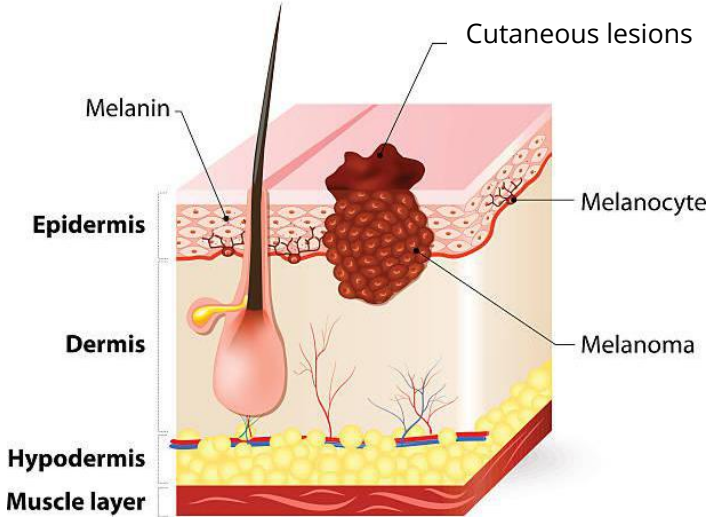
- Dr F. Staroz -

Pathologist at Ouest Pathologie, Quimper site.

Member of the board of directors of the **French Society of Pathology**, and of the **VisioMel** association.

President of the **National Professional Council of Pathologists** (CNPath).

- Context -



Melanoma:

- Multifactorial disease of **melanocytes**
- Current **public health concern**

Origin:

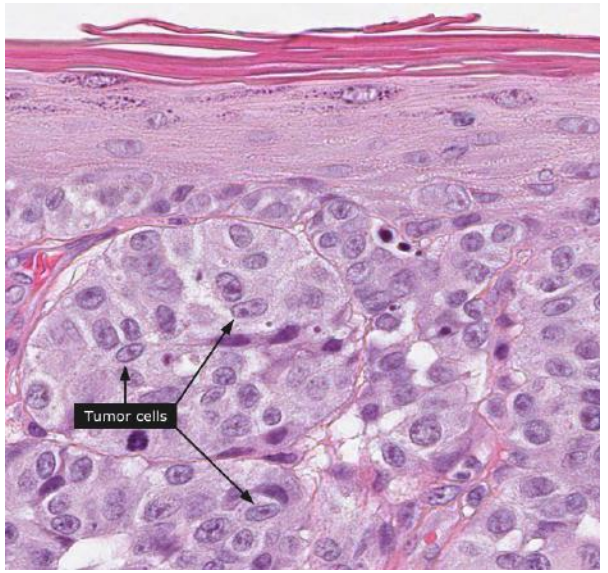
- Mainly depends on the **interaction** between **UV exposure** (period and intensity), **host factors** (presence of atypical nevi, high number of nevi, skin phototype) and **genetic factors**

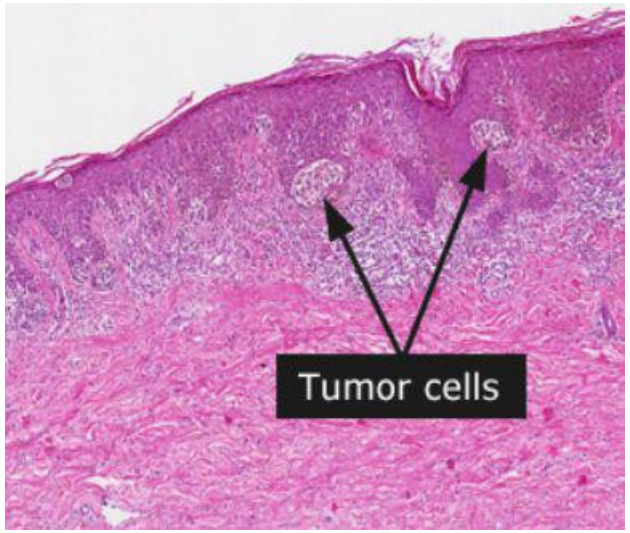
Statistics:

- 3rd most frequent skin cancer but the one of **most serious**
- **Increasing incidence** in white population

Treatment:

- **Surgical resection** of primary melanomas
- **Targeted therapy**, **immunotherapy** : **improve the outcome** of patients with **advanced melanoma** but **severe adverse events** can occur





Human Protein Atlas : https://v15.proteinatlas.org/

Importance to target patients with increased risk of metastatic evolution

- So far -

The risk of **metastastic evolution** is currently **estimated** by anatomopathologists through **histological analysis** of **microscopic lesions** and caracterisation of **cancer stage**

BUT

Main purpose:

There is **a high variability** in the **evolution** of **early stages** melanomas (from 0 to IIC)

 Lack of reliability in the prognosis

Artificial intelligence could **predict which patient is going to relapse**

Secundary purpose:

Targeted therapy is only active in patients with BRAF mutated melanoma

Artificial intelligence could **predict if BRAF is mutated**

- Aim of this data challenge -

In order to **complete the clinical and pathological prognosis**, challengers aim to build an algorithm able to:

1. Identify a **digital fingerprint** that predict the **metastatic evolution** of patients within **5 years following initial diagnosis**

2. Predict the **mutational status** of the tumor for targeted therapy

- Expected available anonymised data: ≈ 3000 patients with primary melanomas -

- **Clinical data**: from **RIC-Mel** database**:**

Network of 49 French centers:
- 43 dermatological centers
- 6 cancer centers

Since the creation in 2012,
≈ 40 000 patients have been included

- **Age, sex**, medical **history**, melanoma **clinical** and **pathological characteristics**, 5 years **outcome**

- Expected available anonymized data: ≈ 3000 patients with primary melanoma -

- **Clinical data**:

- **Age, sex**, medical **history**, melanoma **clinical** and **pathological characteristics**, 5 years **outcome**

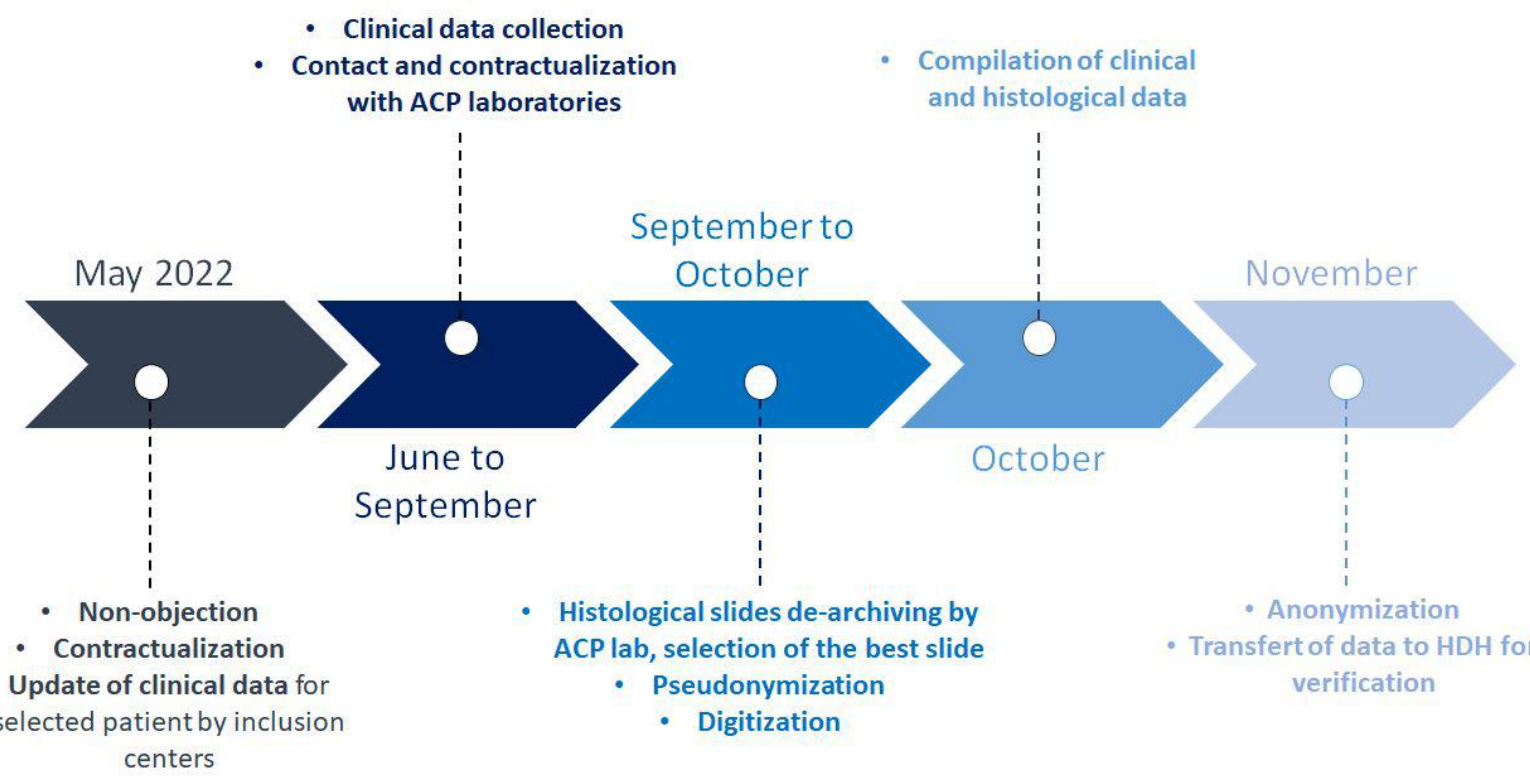
- **ACP data**:

- Cancer **stage** (AJCC, from stage 0 to stage IIC), virtual **histological slides.**

- **Molecular data:**

- Presence/absence of **BRAF mutation**

- Time line -


May 2022
June to September
September to October
October
November
• Non-objection
• Contractualization
• Update of clinical data for selected patient by inclusion centers
• Clinical data collection
• Contact and contractualization with ACP laboratories
• Histological slides de-archiving by ACP lab, selection of the best slide
• Pseudonymization
• Digitization
• Compilation of clinical and histological data
• Anonymization
• Transfert of data to HDH for verification


December
January
2023
February
March
DATA CHALLENGE
April
May
End of May
• Transfert of data to Driven Data platform
• Pre-challenge
• Post-challenge
• Selection of the best algorithm
• Prize giving 25 000€
• Open data in HDH platform

- The data challenge -

Training: ≈1200 slides
+ clinical data
+ molecular data

Test: ≈600 slides

Validation: ≈1200 slides

- The data will be available on the driven data plateform that is hosting this challenge, for a seven weeks period
- The best challengers will be rewarded for their contributions (25 000€)
- Data will be exported to the Health Data Hub and will remain available in open source

Data Challenge Hôpital Foch : DigiLut - Detection of graft rejection following lung transplantation

Data Challenges for Health

2:55 – 3:05

Dr Antoine Roux

(MD, PhD) MCU-PH UMR 0892 VIM, INRAE/U. of Versailles St-Quentin-en-Yvelines

Antoine Roux is a pneumologist and has a PhD in immunology, he is interested in the humoral response to lung transplantation as a post-doctoral fellow at the University of California (UCLA). His current research focuses on the clinical, histological and immunological aspects of rejection in order to develop targeted therapeutic strategies.

GOUVERNEMENT Liberté Égalité Fraternité · Open Government Partnership · Conseil de l'Innovation · bpifrance · Health Data Hub · France22 Présidence française du Conseil de l'Union européenne

HEALTH DATA HUB

DigiLUT

Dr Antoine Roux

MCU PH

Hôpital Foch

UVSQ- Paris Saclay

HOPITAL FOCH

INRAE

UVSQ

université PARIS-SACLAY

Chaire Universitaire de Transplantation

LEADING THE WAY IN ORGAN TRANSPLANTATION

ESOT

BANFF FOUNDATION FOR ALLOGRAFT PATHOLOGY

THE INTERNATIONAL SOCIETY FOR HEART AND LUNG TRANSPLANTATION

A Society that Includes Basic Science, the Failing Heart and Advanced Lung Disease.

LUNG Transplantation

Only curative treatment for end stage Respiratory disease in selected patient

- Remove 2 unfunctional Lung
- Implant 2 functional Lung

4000 LTx/year worldwide

400 LTx/year in France

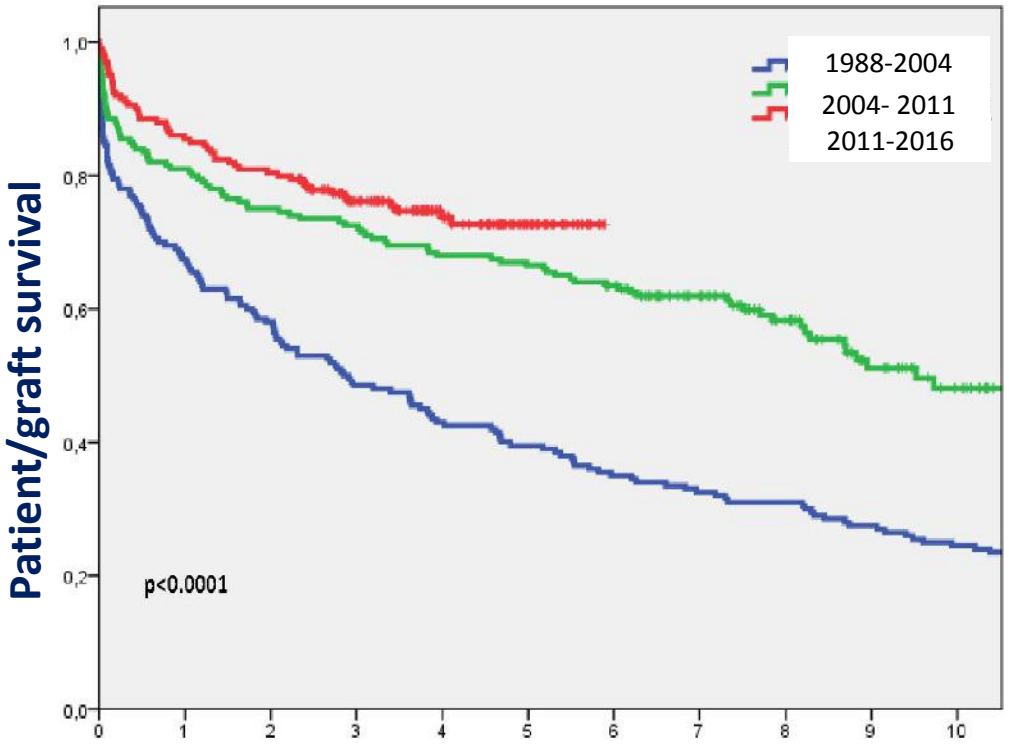
Main indications (>80%)

- COPD
- Pulmonary Fibrosis
- Cystic Fibrosis

High mortality rate without Transplantation

LUNG Transplantation Challenge

Main determinant for survival= REJECTION



Post transplant follow up (years)

Rejection

- Heterogenous process
- Cellular Rejection= A lesion
- Graded 0 to 4

TREATMENT for saving Graft

Histological definition

POOR Reproducibility

Table 5 Interobserver Reliability

Variable	κ Values Median (range)
Biopsy adequacy	0.28 (−0.03 to 0.62)
Acute cellular rejection	0.40 (0.24–0.62)
Airway inflammation	0.23 (0.11–0.56)
Obliterative bronchiolitis	0.18 (0.04–0.58)
Acute lung injury DAD	0.20 (0.03–0.63)
Endotheliitis	0.22 (−0.04 to 0.47)
Alveolar hemosiderosis	0.40 (0.19–0.62)
Capillary inflammation	0.17 (0.03–0.31)
Suspicion for aspiration	0.14 (−0.02 to 0.66)
C4d	0.40 (0.24–0.78)

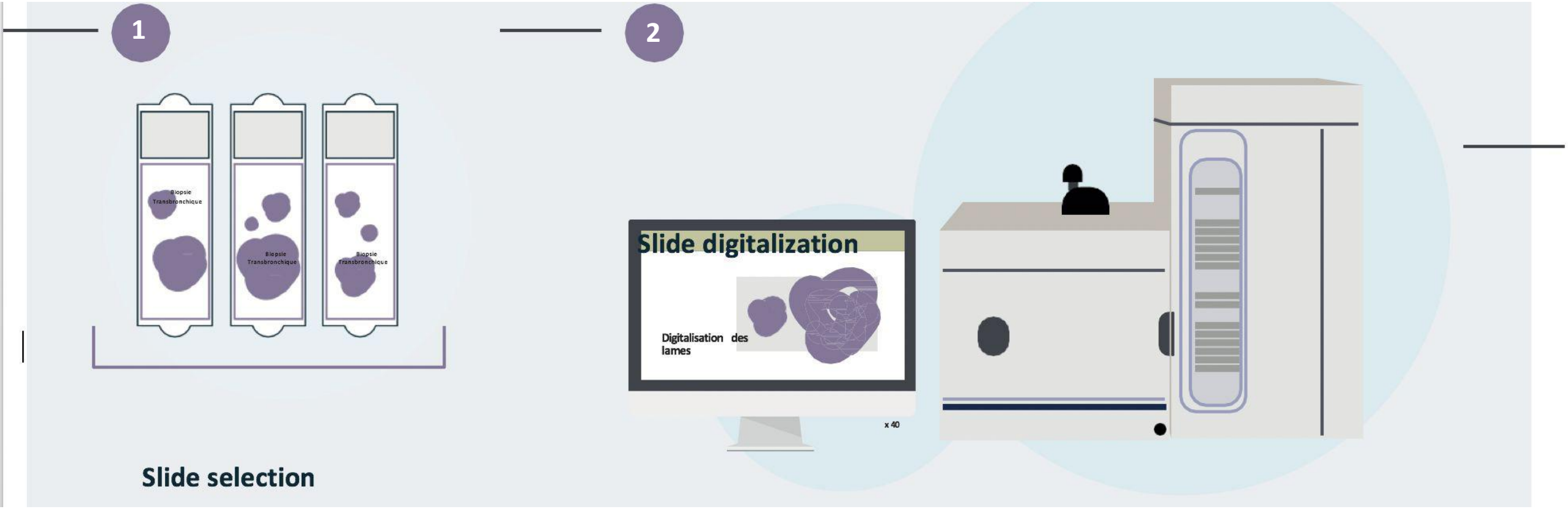
C4d, complement 4d; DAD, diffuse alveolar damage.

Banff study of pathologic changes in lung allograft biopsy specimens with donor-specific antibodies

William Dean Wallace, MD,[a] Ning Li, PhD,[b] Claus B. Andersen, DMSc,[c] A. Valeria Arrossi, MD,[d] Medhat Askar, MD, PhD,[d] Gerry J. Berry, MD,[e] Matthew M. DeNicola, MD,[a] Desley A. Neil, MBBS, PhD, FRCPath,[f] Elizabeth N. Pavlisko, MD,[g] Elaine F. Reed, PhD,[a] Myriam Remmelink, MD,[h] S. Sam Weigt, MD,[i] Birgit Weynand, MD,[j] Jennifer Q. Zhang, PhD,[a] Marie M. Budev, DO,[k] and Carol F. Farver, MD[d]

IMPROVE Acute rejection diagnosis

DigiLUT



IMPROVE Acute rejection diagnosis

DigiLUT

3

Annotation by an International Panel

2000 annotated biopsies

4

Data Challenge :

Datascientist

Reward : 25k€

IMPROVE Acute rejection diagnosis

DigiLUT

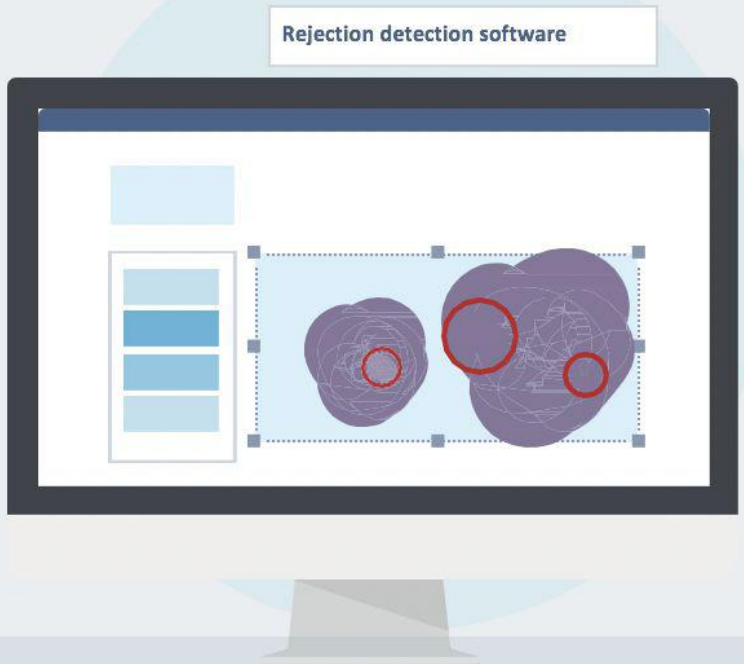
Algorithm :

5

Improve

Acute rejection

Detection



GOAL :

6

Improve LTx treatment

DigiLUT

Challenge for the data scientist community:

Help us to improve Rejection detection and Graft outcome!!!

@drAntoineRoux

HEALTH DATA HUB

Data Challenge SFCTCV : Benefits of AI in thoracic oncology research: about the C PLu NET challenge

Data Challenges for Health

3:05 – 3:15

Romain Vergé

Resident in thoracic surgery at the Toulouse University Hospital, member of the French Society of Thoracic and Cardiovascular Surgery

Romain Vergé is a resident in Thoracic Surgery with a special interest in oncology. He is particularly involved in lung neuroendocrine tumors and is currently conducting a national study with the French Society of Thoracic and Cardiovascular Surgery using the French national database Epithor and classical statistical approaches.
The use of AI appears to him as an additional way to increase knowledge about these rare tumors.

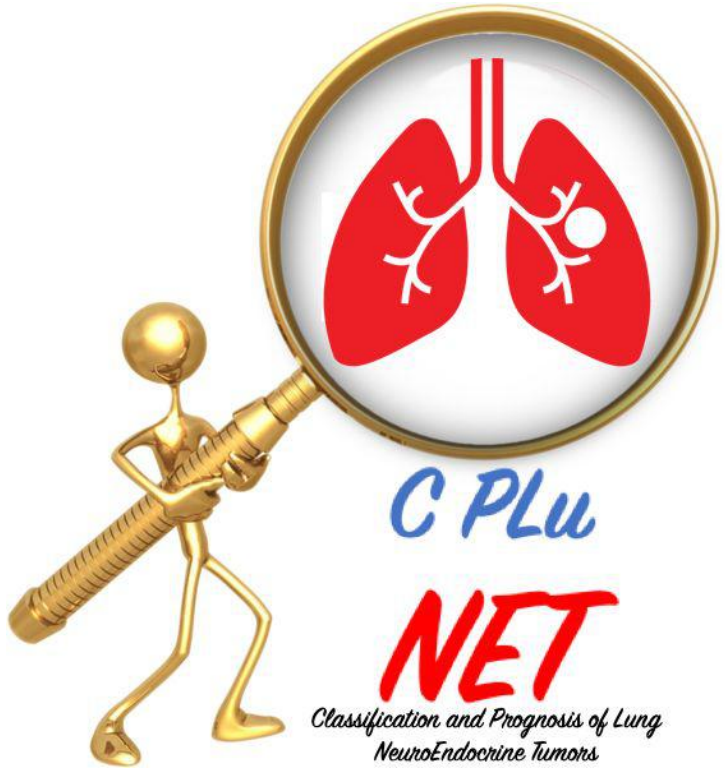
GOUVERNEMENT Liberté Égalité Fraternité · Open Government Partnership · Conseil de l'Innovation · bpifrance · HDH Health Data Hub · France22 Présidence française du Conseil de l'Union européenne

Benefits of AI in thoracic oncology research: about the C PLu NET challenge :

Classification and Prognosis of LUng NeuroEndocrine Tumors

SFCTCV
SOCIÉTÉ FRANÇAISE DE CHIRURGIE THORACIQUE ET CARDIO-VASCULAIRE

CNP·CTCV
Conseil National Professionnel
Chirurgie Thoracique et Cardio-Vasculaire



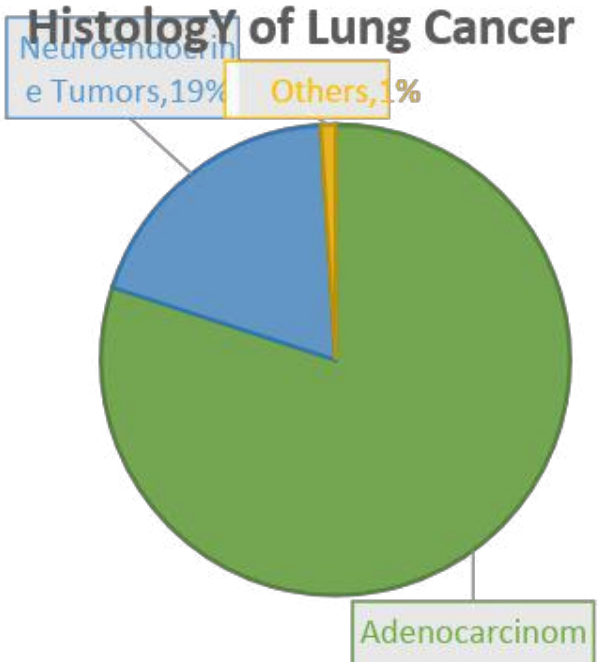
bpifrance

HEALTH DATA HUB

Introduction
Lung Cancer or Lung Cancers ?

- Leading cause of cancer deaths worldwide

- Heterogeneity of lung cancer



Lung Neuroendocrine Tumors :

Agressivity ↓

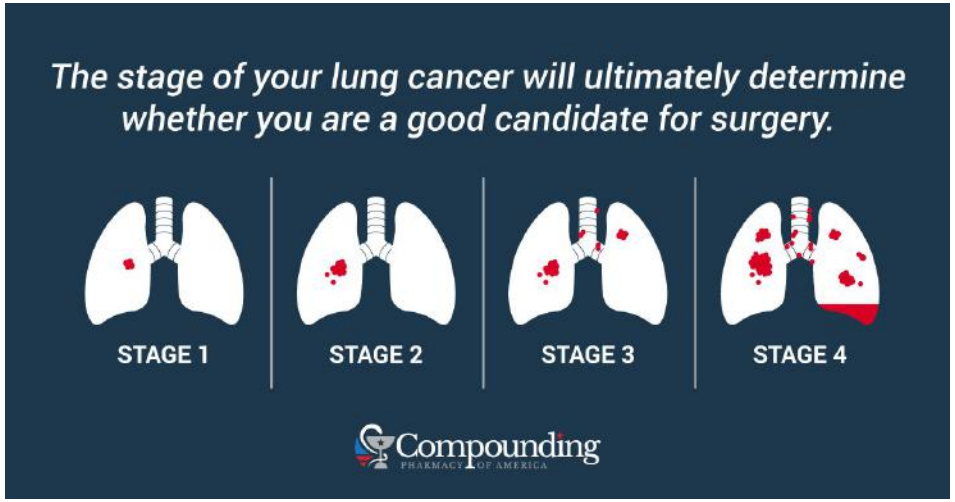
- **Typical Carcinoid (2%)**
- **Atypical Carcinoid (1%)**
- **LGNEC (1%)**
- **SCLC (15%)**

↑ Survival

Scientific Problem

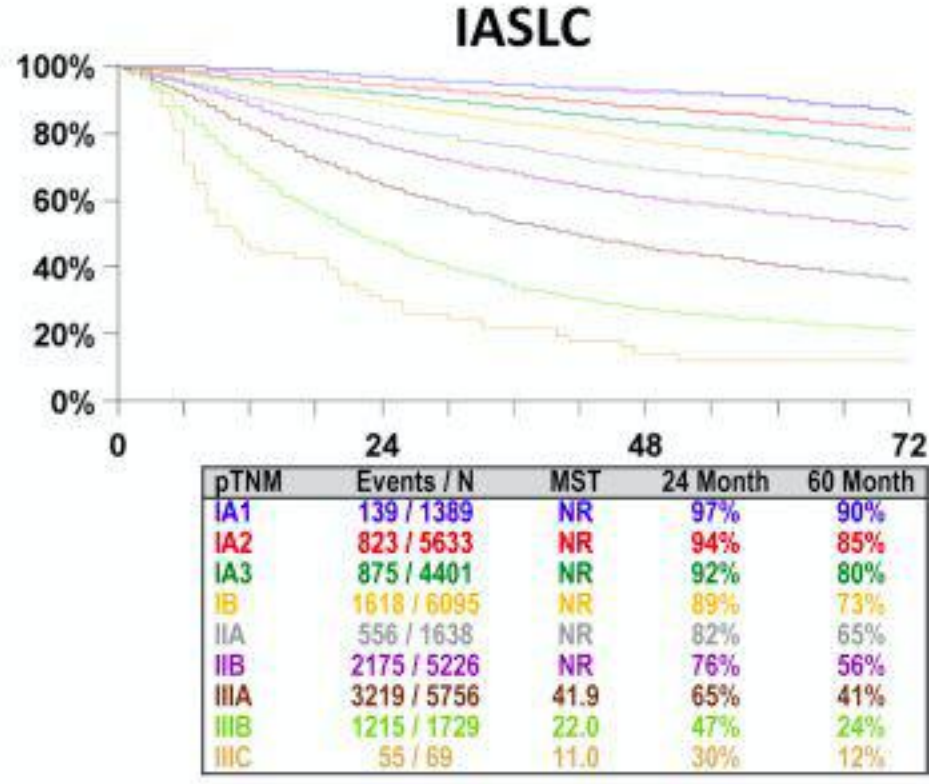
- Classification = Predictive of survival + Necessary for management

Descriptor	Definition
T descriptor	
TX	Primary tumor cannot be assessed or tumor proven by the presence of malignant cells in sputum or bronchial washings but not visualized with imaging or bronchoscopy
T0	No evidence of primary tumor
Tis	Carcinoma in situ
T1	Tumor ≤ 3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus
T1a	Tumor ≤ 1 cm in greatest dimension
T1b	Tumor > 1 cm but ≤ 2 cm in greatest dimension
T1c	Tumor > 2 cm but ≤ 3 cm in greatest dimension
T2 descriptor	
T2	Tumor > 3 cm but ≤ 5 cm or tumor with any of the following features: involvement of a main bronchus regardless of the distance from the carina; invasion of the visceral pleura; associated with partial or complete lung atelectasis or pneumonitis
T2a	Tumor > 3 cm but ≤ 4 cm in greatest dimension
T2b	Tumor > 4 cm but ≤ 5 cm in greatest dimension
T3	Tumor > 5 cm but ≤ 7 cm in greatest dimension or one that directly invades any of the following structures: parietal pleura, chest wall (including superior sulcus tumors), phrenic nerve, parietal pericardium; or separate tumor nodule or nodules in the same lobe
T4	Tumor measuring >7 cm in greatest dimension that invades any of the following structures: mediastinum, diaphragm, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina; or separate tumor nodule or nodules in a different lobe of the same lung
N descriptor	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph nodes
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph nodes
M descriptor	
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumor nodule or nodules in contralateral lung; malignant pleural effusion or pleural thickening or nodules or masses; malignant pericardial effusion or pericardial thickening or nodules or masses
M1b	Single distant (extrathoracic) metastasis in a single organ
M1c	Multiple distant (extrathoracic) metastases in a single organ or multiple organs



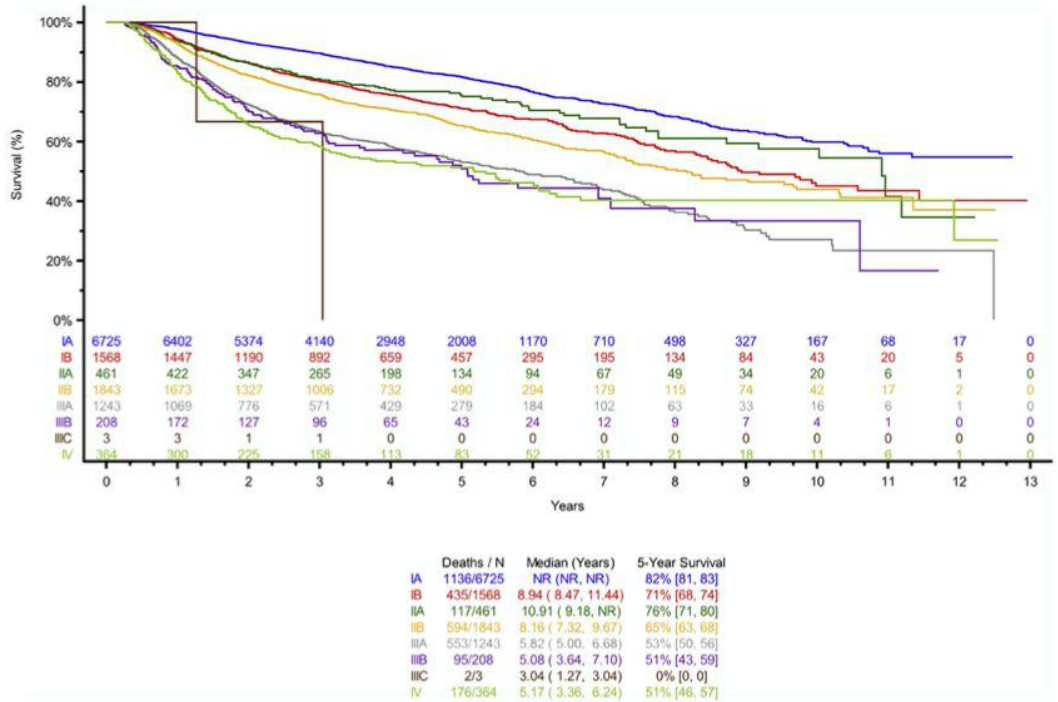
Scientific Problem

- Classification = Predictive of survival + Necessary for management
- Classical Classification = Useless for Lung NET



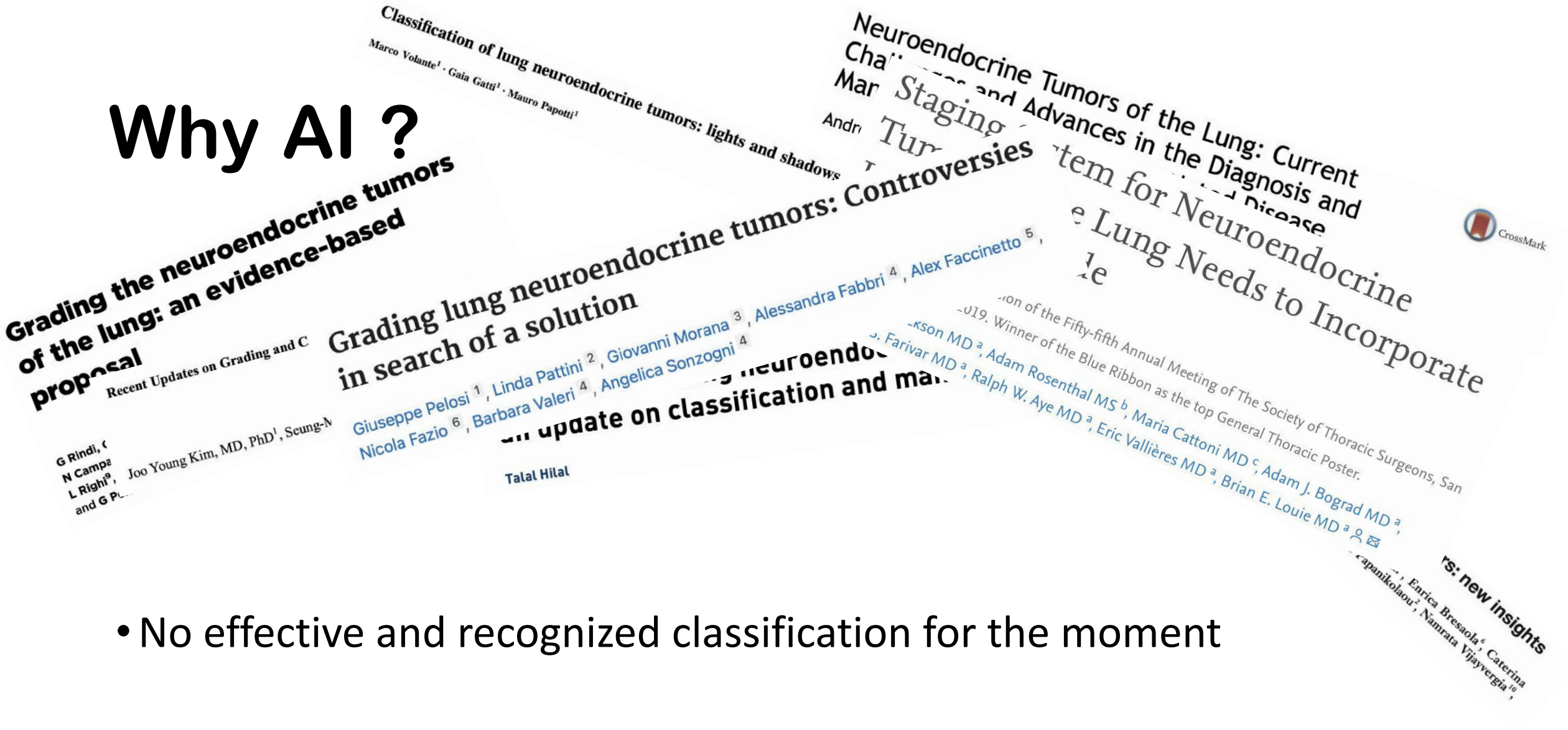
pTNM	Events / N	MST	24 Month	60 Month
IA1	139 / 1389	NR	97%	90%
IA2	823 / 5633	NR	94%	85%
IA3	875 / 4401	NR	92%	80%
IB	1618 / 6095	NR	89%	73%
IIA	556 / 1638	NR	82%	65%
IIB	2175 / 5226	NR	76%	56%
IIIA	3219 / 5756	41.9	65%	41%
IIIB	1215 / 1729	22.0	47%	24%
IIIC	55 / 69	11.0	30%	12%

TNM 8th Edition



Lung NET

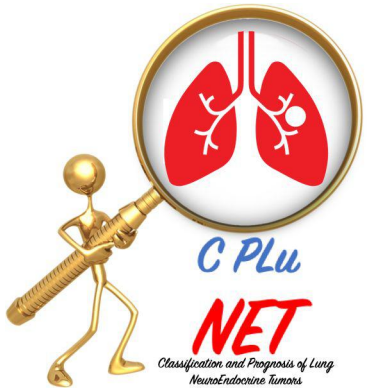
Why AI ?



- No effective and recognized classification for the moment

- Ongoing work of SFCTCV using classical statistical methods

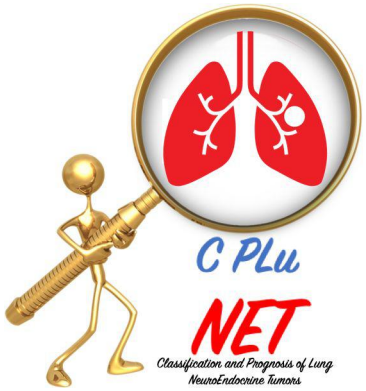
Challenge Objective



- Create a histopronostic classification to predict survival of Lung NET and decide on standardized management

- Obj : Develop an algorithm to predict patient survival using patient clinical data and tumour-specific data

Data Challenge



- 2 types of datasets

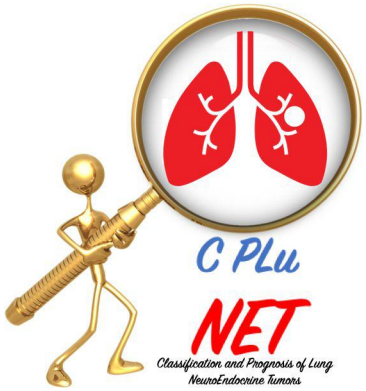
  - Patient clinical data from EPITHOR

  - Tumour-specific data from pathologist reports and pathology slides +++

- For 1500-2000 Patients

- Creation of an unique bank of digital and imaging data

EPITHOR

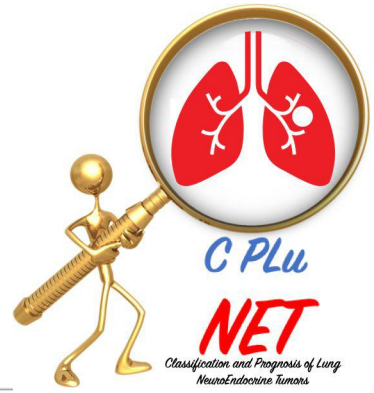


- *French National Database of General Thoracic Surgery*

- Created 2002
- Prospective +++
- 111 centers = most surgical procedures performed in French thoracic surgical departments
- 52 variables per patient, covering information about the patients' characteristics, associated illness, pulmonary function, surgical procedures, cancer staging, and postoperative outcome

- Source of many scientific publications +++

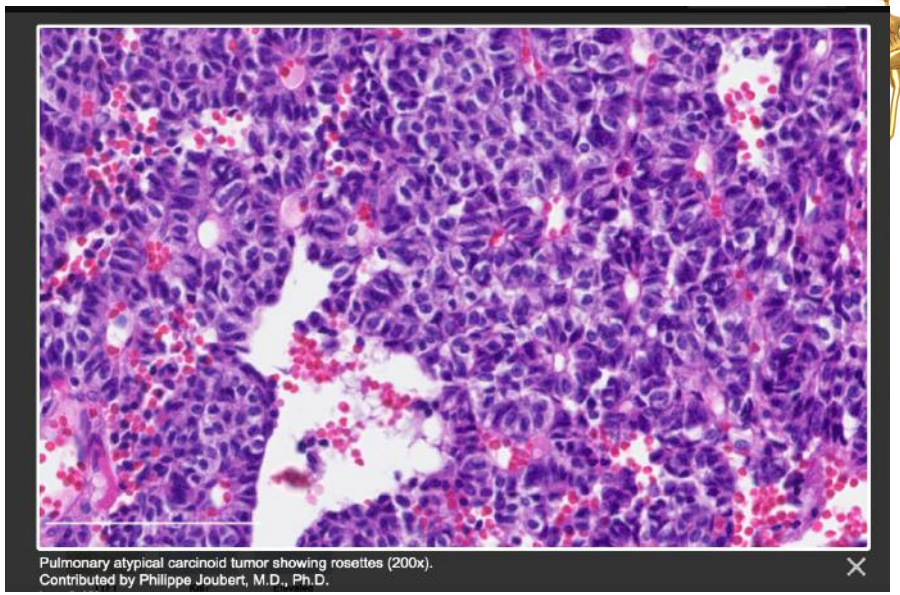
EPITHOR

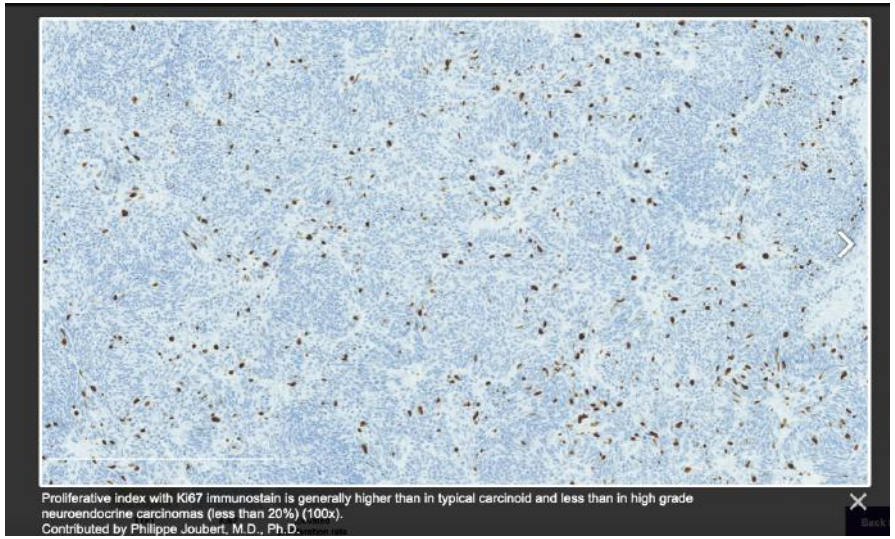


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	PATIENT	Sexe	DDNais	SEJOUR_D/	ENTREE	AGE	POIDS	TAILLE	BMI	OMS	FUME	VEMS	VEMS_	TIFFEN	DLCO
2	640000162_01-00000062	M	#########	2016	07/04/2016	79	75	173	25,1	1	15	83	56		
3	630000479_01-CF_05893	F	#########	2014	07/01/2014	42	80	150	35,6	0					
4	630000479_01-CF_04140	M	#########	2011	09/11/2011	67	90	170	31,1	0		87	87		
5	630000479_01-CF_03087	F	#########	2010	16/02/2010	74	68	171	23,3	1		93	93		
6	630000479_01-CF_03088	F	#########	2010	17/02/2010	37	61	150	27,1	0					
7	630000479_01-CF_05758	M	#########	2013	21/10/2013	38	79	177	25,2	0		108	108		
8	630000479_01-CF_05753	M	#########	2013	03/11/2013	73	98	171	33,5	1		121	121		
9	630000479_01-CF_07469	F	#########	2015	20/10/2015	45	74	170	25,6	0		69	69		
10	630000479_01-CF_04048	M	#########	2010	16/12/2010	76	60	174	19,8	1					
11	630000479_01-CF_04041	F	#########	2011	13/09/2011	45	50	170	17,3	1		76	76		
12	630000479_01-CF_04081	M	#########	2011	07/10/2011	64	88	175	28,7	1		110	110		
13	630000479_01-CF_07393	F	#########	2015	01/10/2015	56	70	158	28	0		65	65		
14	370004467_01-00000237	F	#########	2016	17/04/2016	76	53	168	18,8	0	10	109	86		
15	370004467_01-00000238	M	#########	2016	18/05/2016	65	68	168	24,1	0	0	94	74		
16	630000479_01-CF_04534	M	#########	2012	19/04/2012	74	78	180	24,1	0		52	52		
17	630000479_01-CF_07120	M	#########	2015	22/06/2015	57	103	193	27,7	0		88	88		
18	630000479_01-CF_07051	F	#########	2015	02/06/2015	65	59	150	26,2	0		86	86		
19	630000479_01-CF_07063	F	#########	2015	07/06/2015	66	53	163	19,9	0		99	99		
20	630000479_01-CF_07240	F	#########	2015	30/08/2015	30	72	173	24,1	0					
21	630000479_01-CF_05985	F	#########	2014	11/02/2014	70	57	159	22,5	1		74	74		
22	630000479_01-CF_05989	F	#########	2014	20/01/2014	45	74	168	26,2	1		83	83		
23	630000479_01-CF_04288	M	#########	2011	22/06/2011	66	70	165	25,7	2					

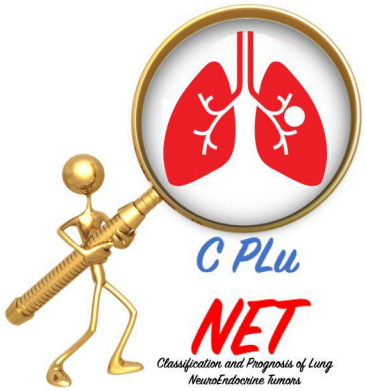
Pathology Slides

- Identify tumour-specific characteristics -> Prognosis
  - Cell « design »
  - Mitoses
  - ….
- From nearly 15 centers
- Scanning +++ (1Go for each slide)
- 2 for each patient :
  -> Classical : H&E
  -> Ki67





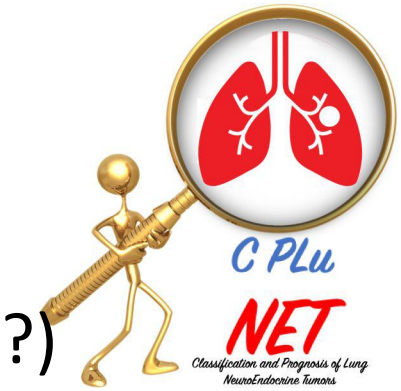
Pathology Slides



- Review of the pathology slides by an expert committee
  - Validate the inclusion of neuroendocrine tumors
  - Underline the areas of interest on which the machines will work

- Creation of a unique virtual pathology slide bank for these rares tumors

Data Challenge



- 2008-2020 : Typical Carcinoid, Atypical Carcinoid, LGCNEC(SCLC?)
- 1500-2000 patients /15 centers in France
  - 20 variables for each patient extracted from Epithor
  - 2 pathology slides per patient = 3-4000 pathology slides = 4To of data

  + Macroscopic characteristics from pathology reports
  - Survival data from INSEE Database
- No missing data

 Primary outcome : Overall survival

• Challenge objective : Test the ability of the algorithm to predict patient survival with both types of data

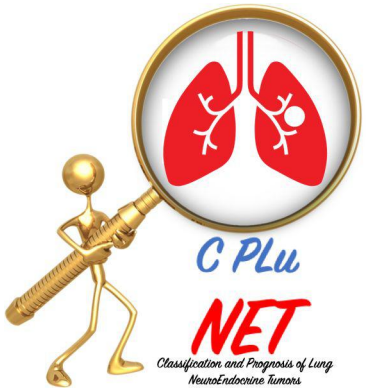
You're a Data Scientist ?

- First time that clinical data and pathology slides of exceptional quality are combined to create a classification using AI

- Unique way to participate in an exciting challenge for data scientists and clinicians

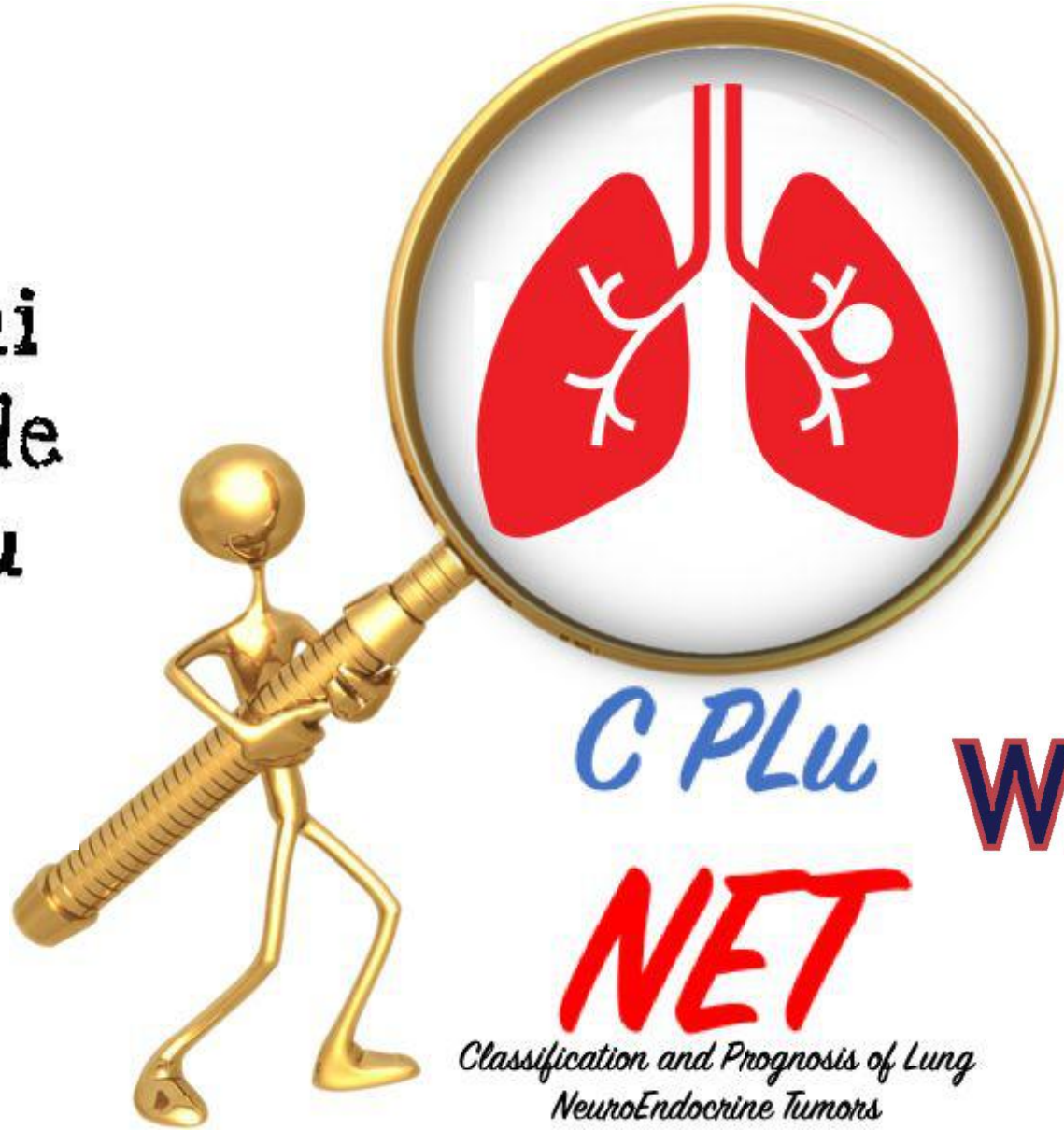
- Last but not least: 25000€

Dates to remember



- Creation of datasets: May-December 2022
- Review by expert committee : November 2022-March 2023
- Challenge : March-May 2023
- Prize-giving ceremony : June 2023

oui
nide
iou



C PLu

NET

Classification and Prognosis of Lung NeuroEndocrine Tumors

WE WANT YOU!

Data Challenges for Health

BREAK TIME !

3:15 – 4:05

STAY INFORMED !

Scan the QR code to get the latest informations on our Data Challenges !

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Sage Bionetworks - Organizing DREAM Challenges for Engaged Health Research

Data Challenges for Health

4:05 – 4:15

Thomas Schaffter, PhD

PhD, Lead of the Research & Benchmarking Technology Workstream Sage Bionetworks

Thomas Schaffter is the Lead of the Research & Benchmarking Technology Work Stream at Sage Bionetworks where he develops cloud infrastructure that enables the organization of biomedical scientific challenges.

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Organizing DREAM Challenges for Engaged Health Research

Thomas Schaffter, PhD
Sage Bionetworks

PFUE Data Challenges for Health event
May 24, 2022

DREAM CHALLENGES
powered by Sage Bionetworks

SageBionetworks

Thomas Schaffter

PhD in Biotechnology & Bioengineering, EPFL, Lausanne, Switzerland.

Post-Doctoral Researcher @ IBM, New York, USA.

Senior Solutions Architect @ Sage Bionetworks, Seattle, USA.

Mission: accelerate biomedical discoveries by improving methods for scientific collaboration and communication

Better Science Together

https://www.linkedin.com/in/tschaffter/

https://www.github.com/tschaffter

Sage Bionetworks Office, Seattle.

SageBionetworks & DREAM CHALLENGES

Our mission is ...

- to contribute to the solution of important **biomedical** problems
- to foster **collaboration** between research groups
- to **democratize access** to data
- to **accelerate research**
- to **objectively assess** algorithms and their performance

SYNAPSE

People and Teams

- Group permissions
- Integrated auth
- ORCID links
- ARs, ACLs, and the ACT

+

Data and Results

- Versioning for data, code
- Provenance to link results
- Distributed data stores
- Continuous benchmarking

+

Research Process

- Connecting pieces of the broader research cycle
- Real-time collaboration
- Sharing across groups, teams, departments

Synapse Platform and Integrated Tools

Modular platform supports many different pieces of the research “puzzle”, and all features can be leveraged in the web client, the analytic (R/Python/CLI) clients, or directly via the API enabling custom integrations/functionality

Recent challenge achievements

First modern model-to-data challenge using Docker (2017)

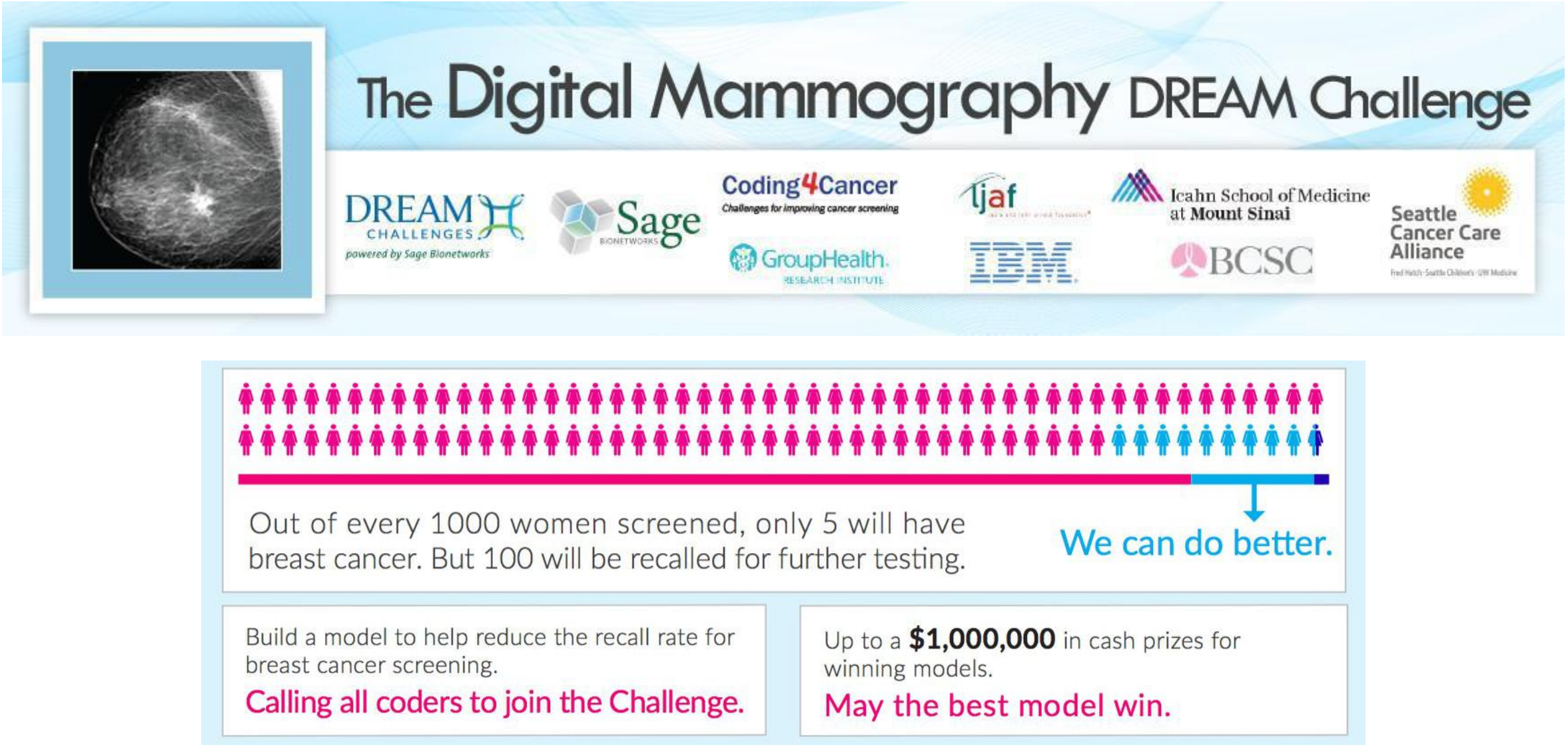
First DREAM challenge on sensitive EHR data (2019)

First continuous challenge - EHR data (2020)

NLPSANDBOX.IO

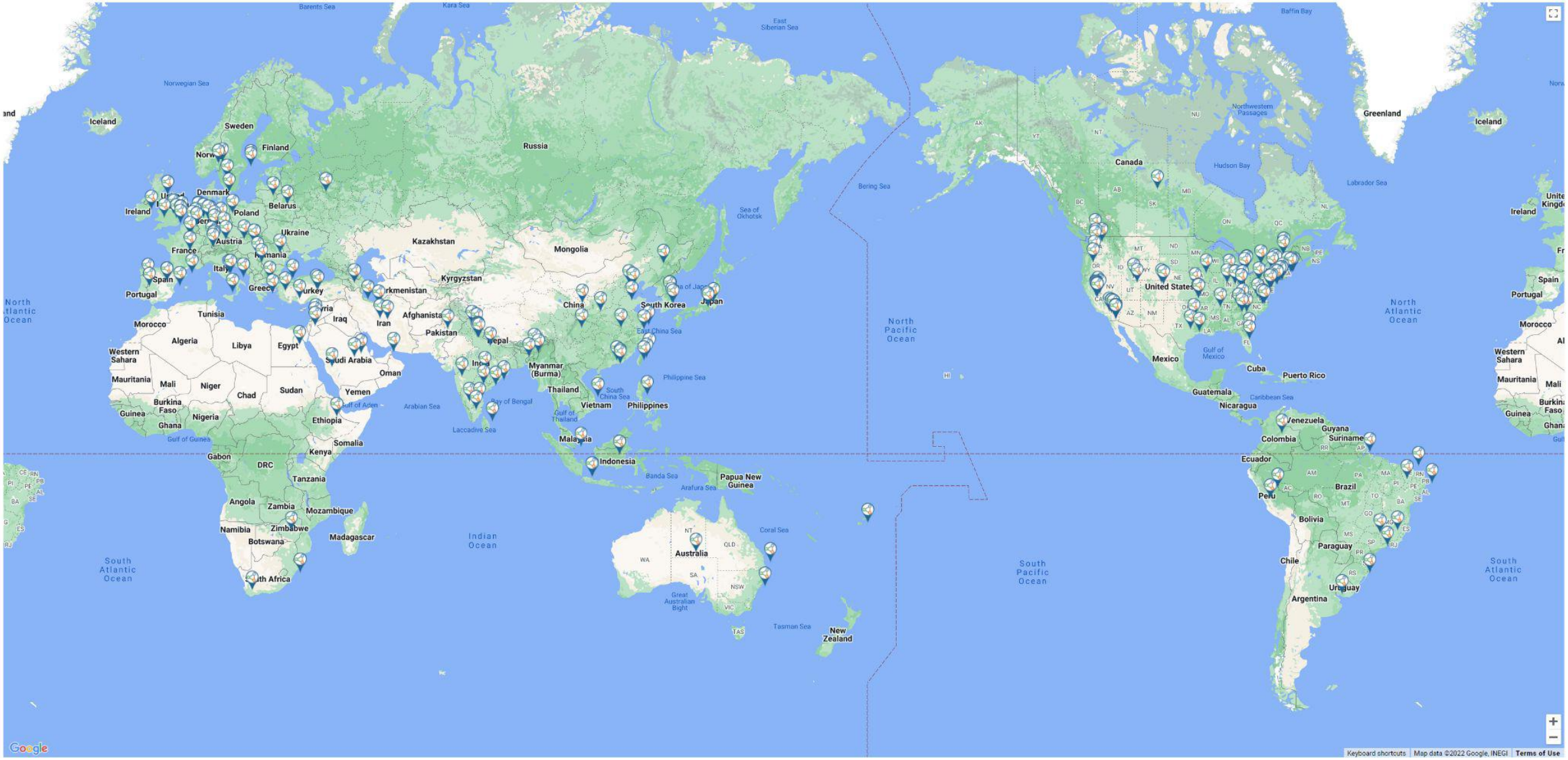
First multi-site evaluation - EHR data (2021)

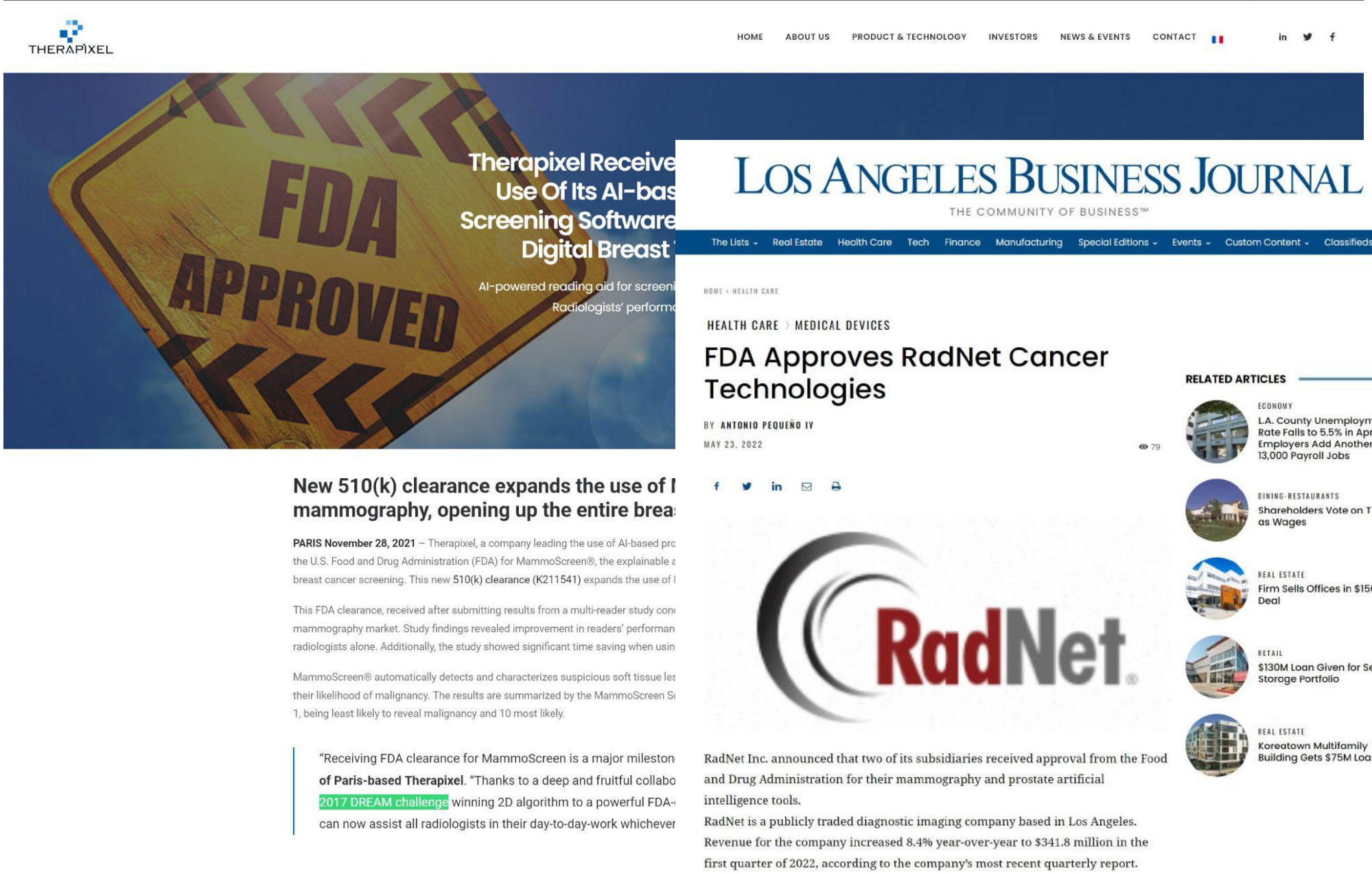
https://dreamchallenges.org



https://www.synapse.org/Digital_Mammography_DREAM_Challenge

DM DREAM Challenge Participants





Schaffter, T., Buist, D.S., Lee, C.I., Nikulin, Y., Ribli, D., Guan, Y., Lotter, W., Jie, Z., Du, H., Wang, S. and Feng, J., 2020. Evaluation of combined artificial intelligence and radiologist assessment to interpret screening mammograms. *JAMA network open*, *3*(3), pp.e200265-e200265.

Thanks!

Data Challenge SFMN : Dat-Hub - Parkinson and dopamine transporter imaging: a data challenge from the Health Data Hub and the French Society of Nuclear Medicine

Data Challenges for Health

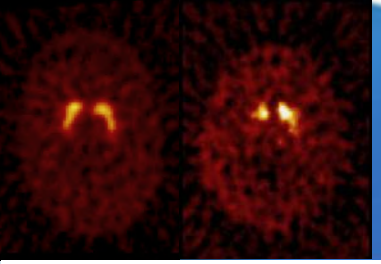
4:15 – 4:25

Pr Eric Guedj

Aix-Marseille University, Timone hospital, SFMN leader of Brain Work Group

Eric Guedj is Professor of Biophysics and Nuclear Medicine at Aix-Marseille University, and head of the Nuclear Medicine Department at Timone hospital. He is an expert in brain molecular imaging and theranostics, and the current SFMN leader of the Brain Work Group. Eric Guedj scientifically coordinates the “DaT-HUB” Data Challenge, with the objective to improve the diagnosis of Parkinson’s disease by brain SPECT imaging of dopamine transporter.

GOUVERNEMENT Liberté Égalité Fraternité · Open Government Partnership · Conseil de l’innovation · bpifrance · Health Data Hub · France22 Présidence française du Conseil de l’Union européenne



SFMN

Société Française de Médecine Nucléaire et Imagerie Moléculaire

DaT-HUB project

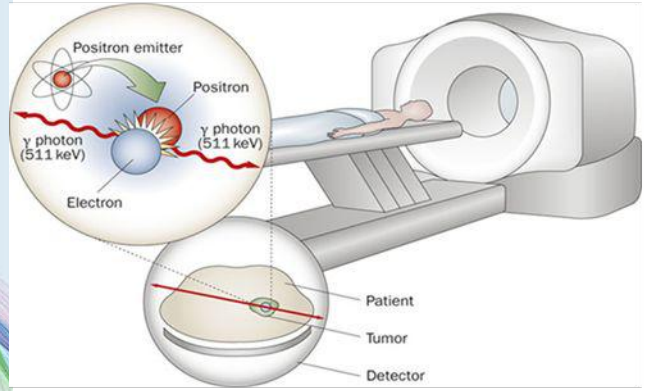
HEALTH DATA HUB

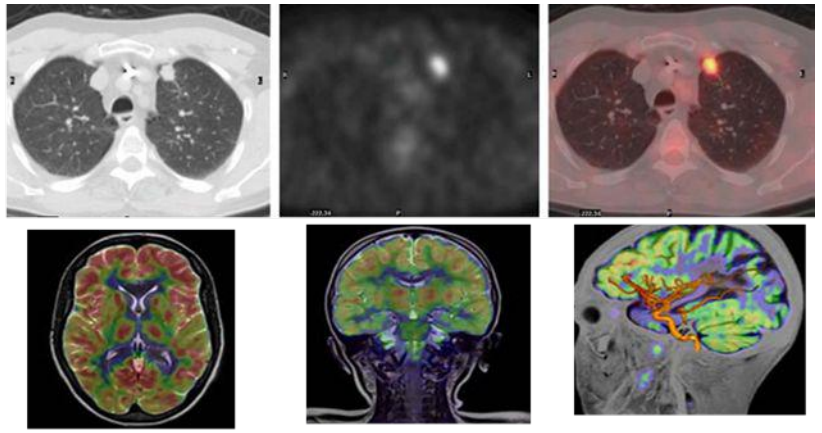
SFMN - HDH Partnership Data Challenge

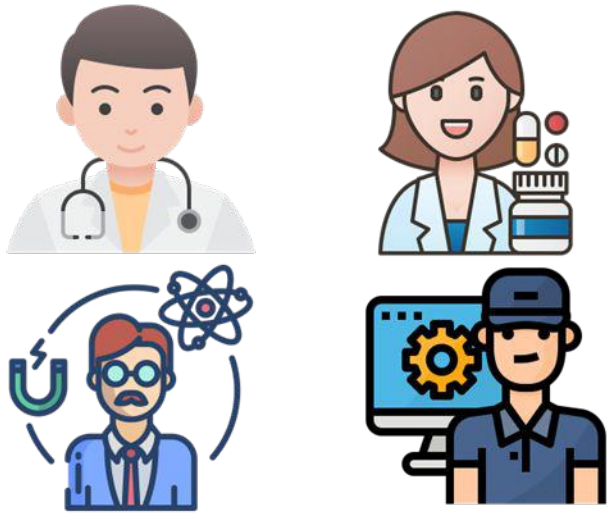
Prof. Eric GUEDJ

Nuclear Medicine in France

« The activity of nuclear medicine consists of the use, in diagnosis or therapy, of a radio-pharmaceutical or an active implantable medical device, in unsealed sources, emitting ionizing radiation, administered to the patient, including the use a single photon emission computed tomography [SPECT] or positron emission tomography [PET] camera and integrated, if applicable, other imaging systems », Art. R. 6123-134, Public Health Code







- 620 cameras in France, and 1.679.388 imaging exams in 2019 (+4.6% per year since 2008)
- 927 physicians, 172 pharmacists, 192 physicists, 1,518 technologists

The medical motives

- Need of **precise** and **early** characterization to:
  - diagnose and guide treatment of diseases otherwise incurable at later stages
  - evaluate efficiency of (expensive) therapeutics to not delay more effective treatment, and not cause unnecessary side effects

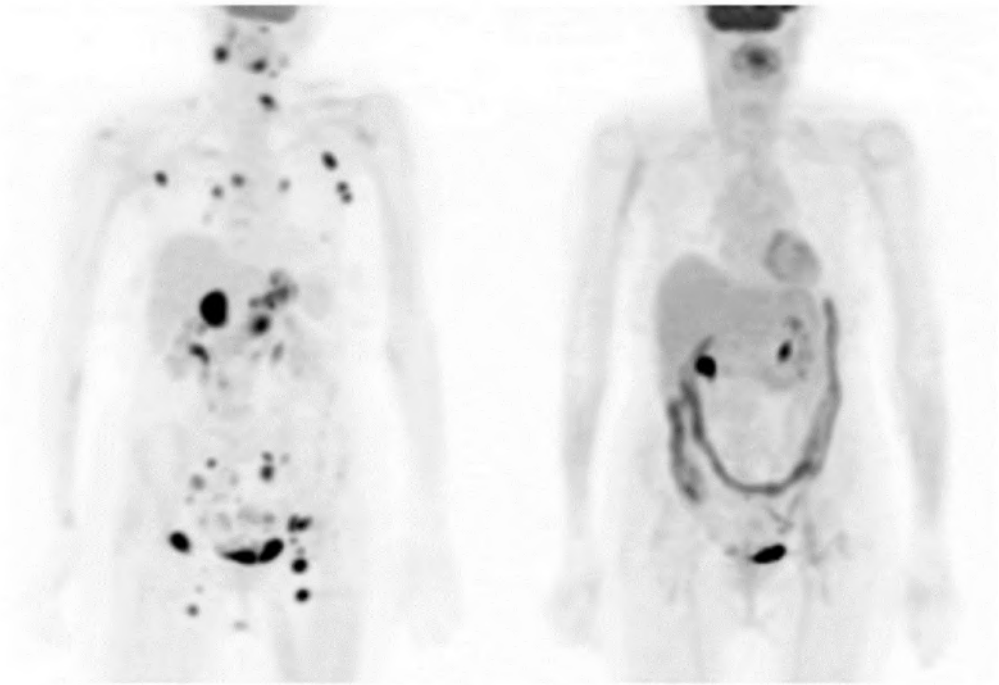
- Evolution towards **personalized medicine**

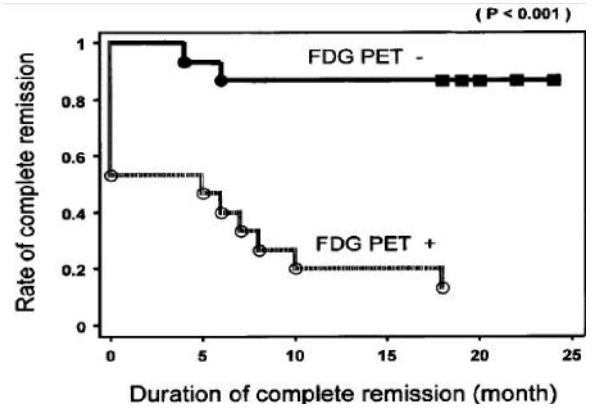
  groups of patients → the individual patient

- Diagnostic and therapeutic advances mainly driven by non-invasive bio-imaging, particularly at individual level, for **molecular signature**
  - molecular changes precede morphological ones (***more sensitive for early diagnosis and evaluation***)

PET Predicts Prognosis After 1 Cycle of Chemotherapy in Aggressive Lymphoma and Hodgkin's Disease

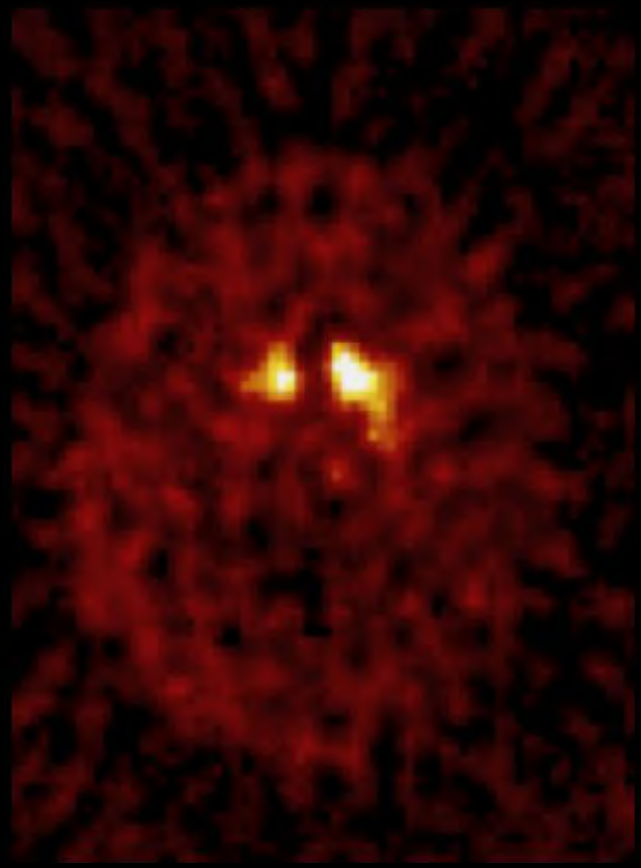
Lale Kostakoglu, MD[1]; Morton Coleman, MD[2]; John P. Leonard, MD[2]; Ichiei Kuji, MD[1]; Holly Zoe[1]; and Stanley J. Goldsmith, MD[1]





THE JOURNAL OF NUCLEAR MEDICINE • Vol. 43 • No. 8 • August 2002

50% of brain dopaminergic loss before first symptoms of Parkinson



Healthy subject

Patient with Parkinson's disease

Parkinson's disease: 2nd most frequent neurodegenerative disease after AD, 1% of subjects after 60yrs, ≈ 200,000 patients in France, 30,000 new patients per year; 30% of diagnostic error (autopsy) with positive impact of early treatments

The medical motives

- Need of **precise** and **early** non-invasive characterization to:
  - diagnose and guide treatment of diseases otherwise incurable at later stages
  - evaluate efficiency of (expensive) therapeutics to not delay more effective treatment, and not cause unnecessary side effects

- Evolution towards **personalized medicine**

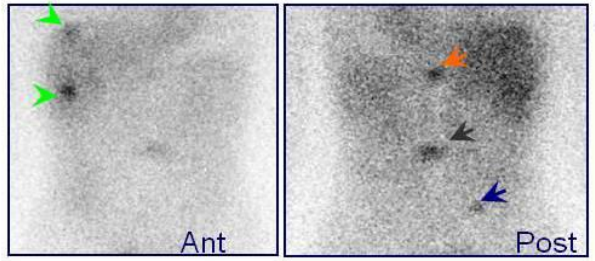
  groups of patients → the individual patient

- Diagnostic and therapeutic advances mainly driven by non-invasive bio-imaging, particularly at individual level, for **molecular signature**
  - molecular changes precede morphological ones (***more sensitive for early diagnosis and evaluation***)
  - molecular **biomarkers** better evaluate complexity of each disease at patient- and lesion- level (***more specific for precise diagnosis and evaluation***)

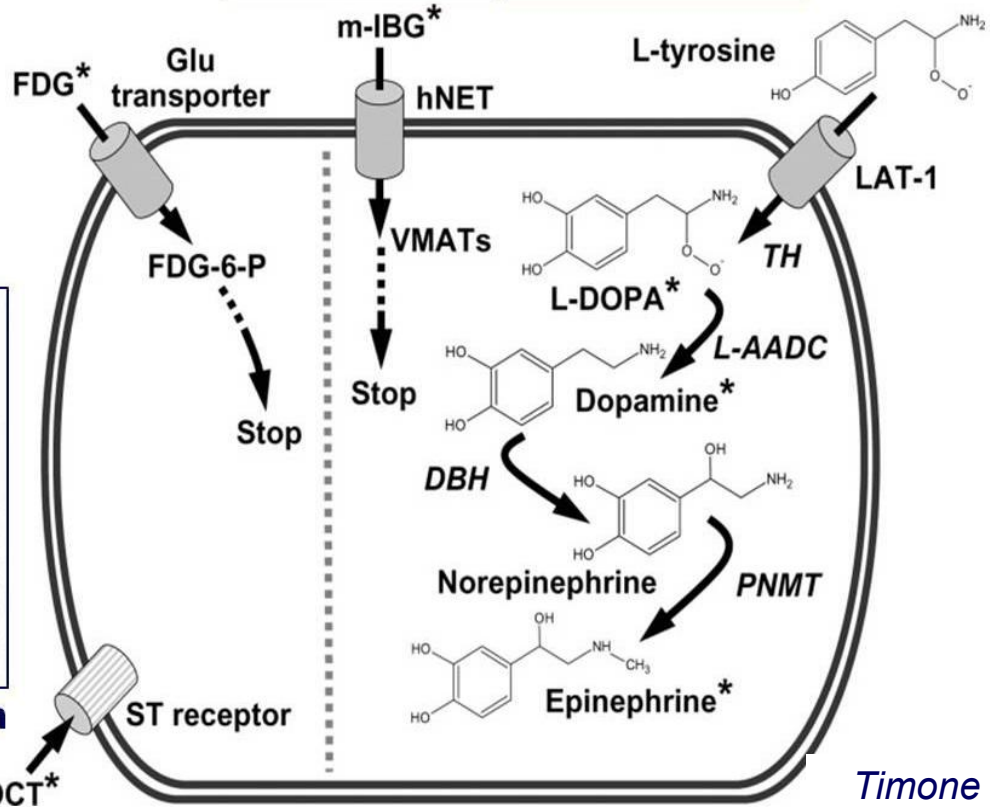
Molecular signature of neuro-endocrine tumors

Assistance Publique Hôpitaux de Marseille

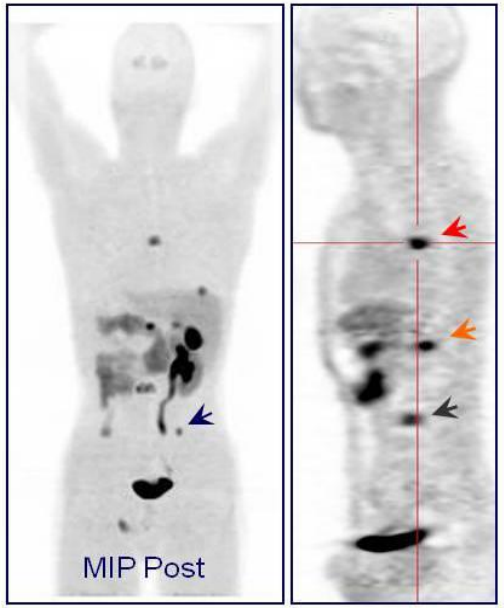
TEP ^{18}FDG



^{131}I-MIBG



OctréoScan SRS



TEP ^{18}FDOPA

Timone Nuclear Medicine Department

The medical motives

- Need of **precise** and **early** non-invasive characterization to:
  - diagnose and guide treatment of diseases otherwise incurable at later stages
  - evaluate efficiency of (expensive) therapeutics to not delay more effective treatment, and not cause unnecessary side effects

- Evolution towards **personalized medicine**

  groups of patients → the individual patient

- Diagnostic and therapeutic advances mainly driven by non-invasive bio-imaging, particularly at individual level, for **molecular signature**
  - molecular changes precede morphological ones (***more sensitive for early diagnosis and evaluation***)
  - molecular **biomarkers** better evaluate complexity of each disease at patient- and lesion- level (***more specific for precise diagnosis and evaluation***)
    - ✔ this molecular complexity is linked to prognosis, and therapy (***companion drugs***)

^{11}C-PiB PET assessment of change in fibrillar amyloid-β load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study

Juha O Rinne, David J Brooks, Martin N Rossor, Nick C Fox, Roger Bullock, William E Klunk, Chester A Mathis, Kaj Blennow, Jerome Barakos, Aren A Okello, Sofia Rodriguez Martinez de Liano, Enchi Liu, Martin Koller, Keith M Gregg, Dale Schenk, Ronald Black, Michael Grundman

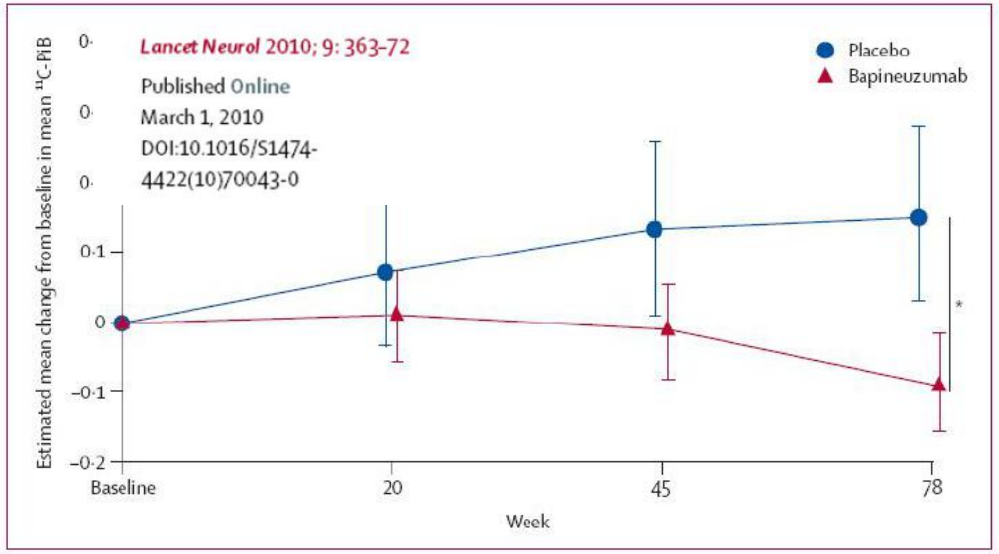
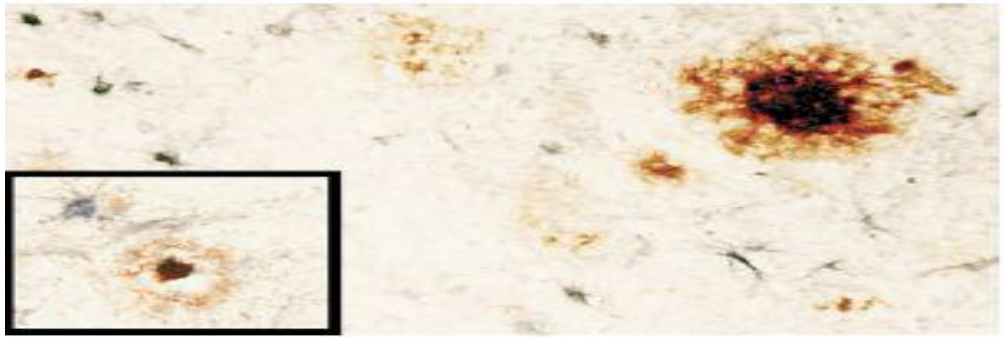


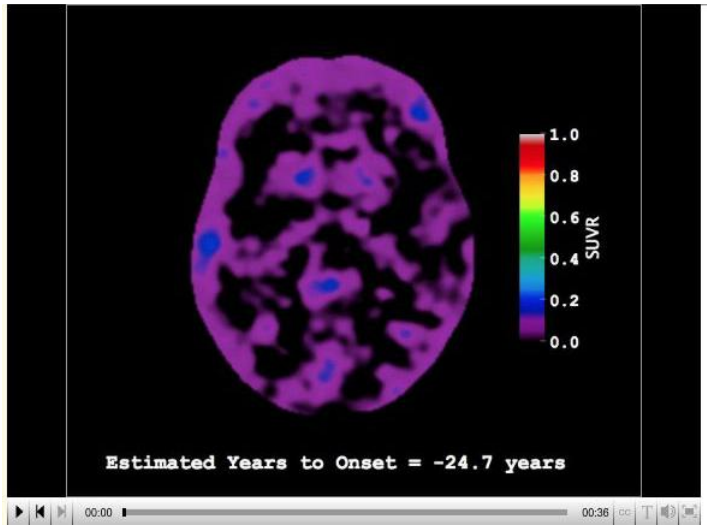
***Figure 2*: Estimated change from baseline over time in mean ^{11}C-PiB PET**
Data are least squares means and 95% CIs. *Difference between patients in the placebo group and those in the bapineuzumab group at week 78=–0·24 (p=0·003). PiB=Pittsburgh compound B.



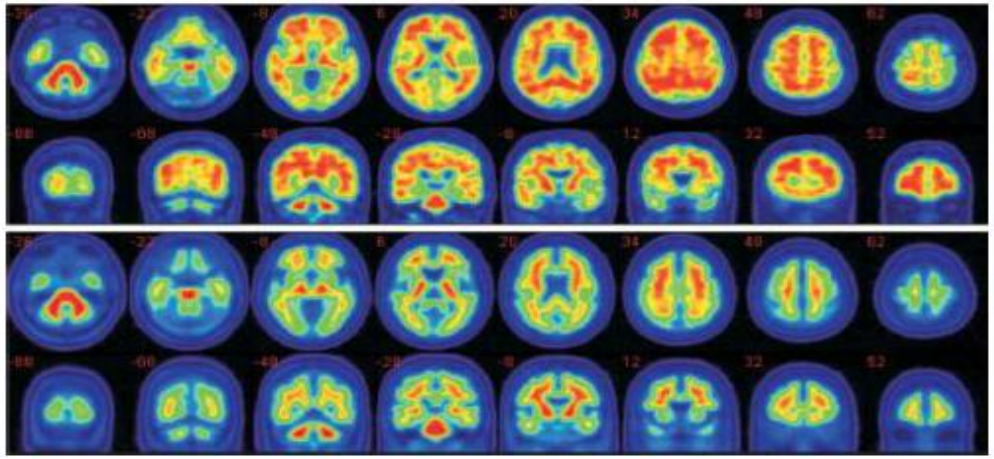
The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 AUGUST 30, 2012 VOL. 367 NO. 9

Clinical and Biomarker Changes in Dominantly Inherited Alzheimer's Disease

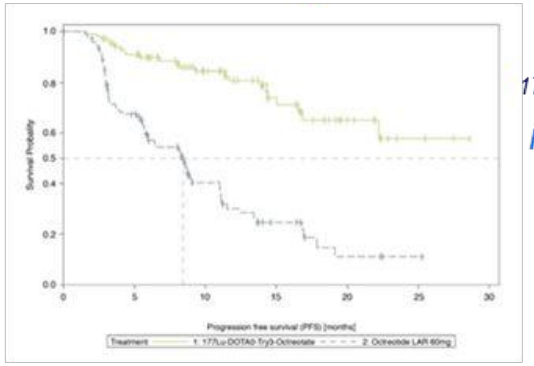


- ^{18}F-PIB Flutemetamol (GE Healthcare)



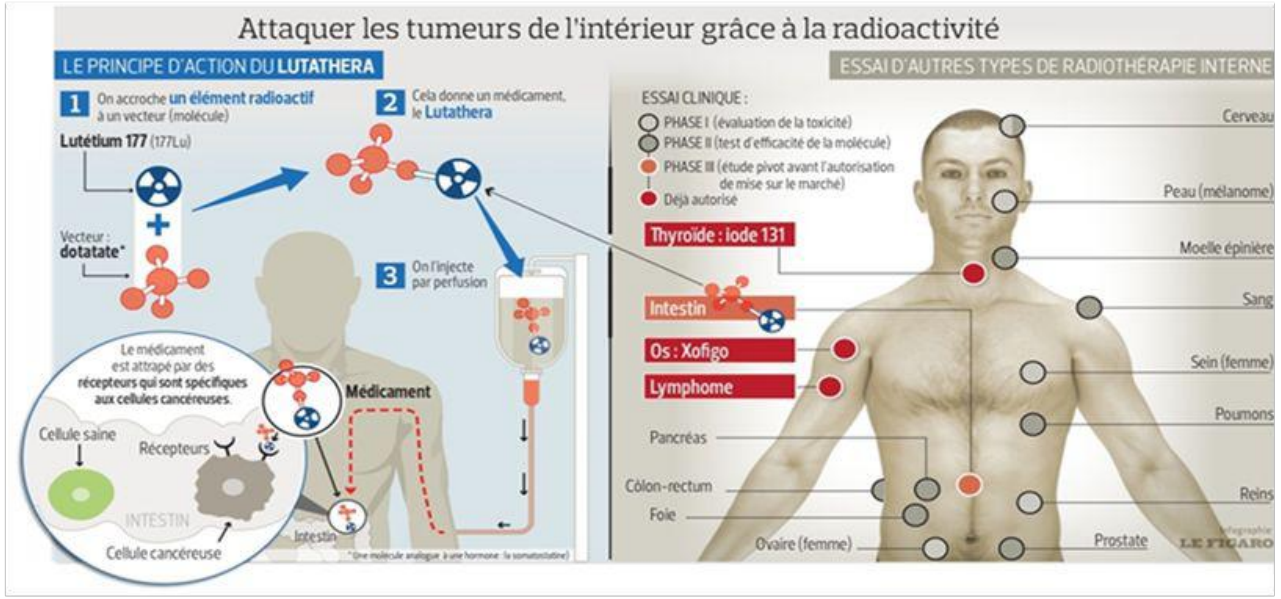
Molecular Imaging & theranostics

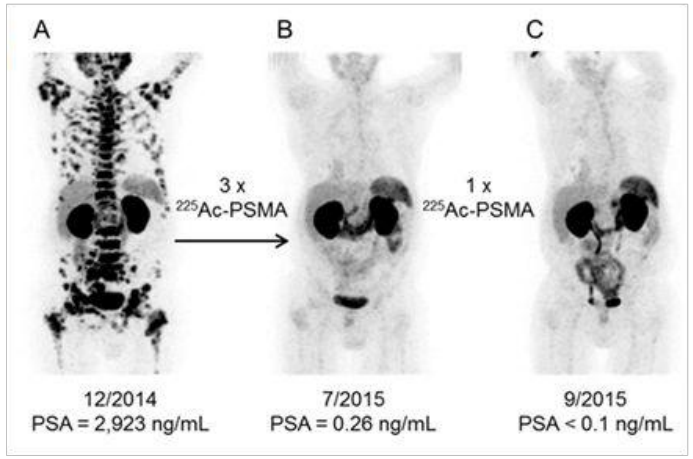
- Molecular process targetted by a tracer
- Labeled by a radio-isotope: **diagnostic & therapeutic (companion pair)**

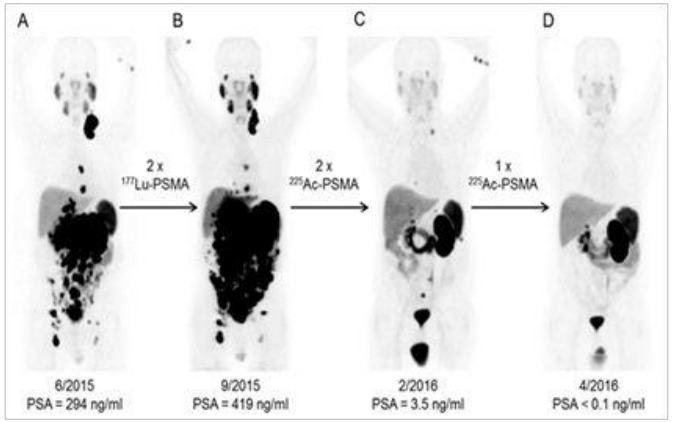


*^{177}Lu-Dotatate Improves Progression-Free Survival in Patients with Midgut Neuroendocrine Tumours: Results of the Phase III **NETTER-1** Trial*

N Engl J Med. 2017 Jan 12;376(2):125-135

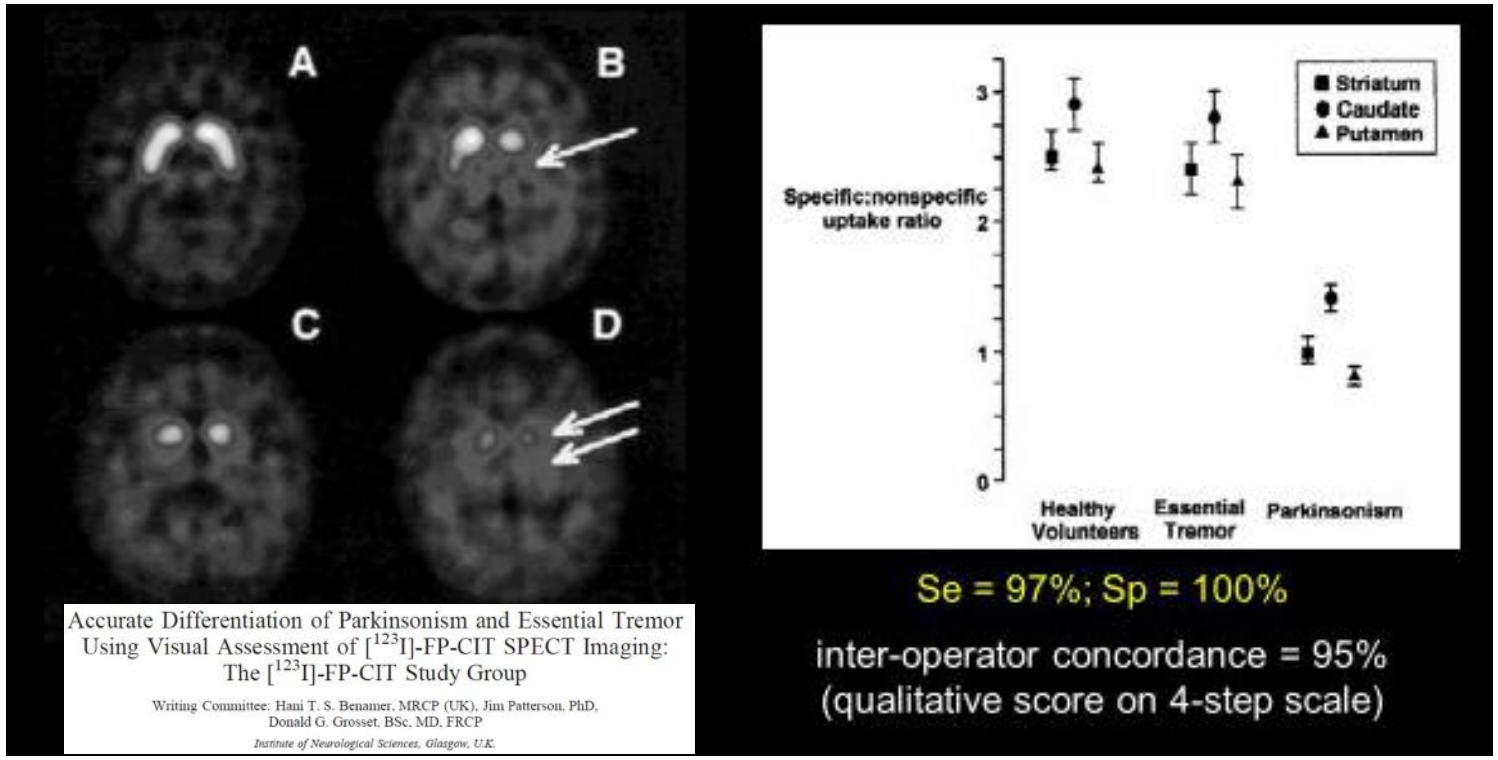


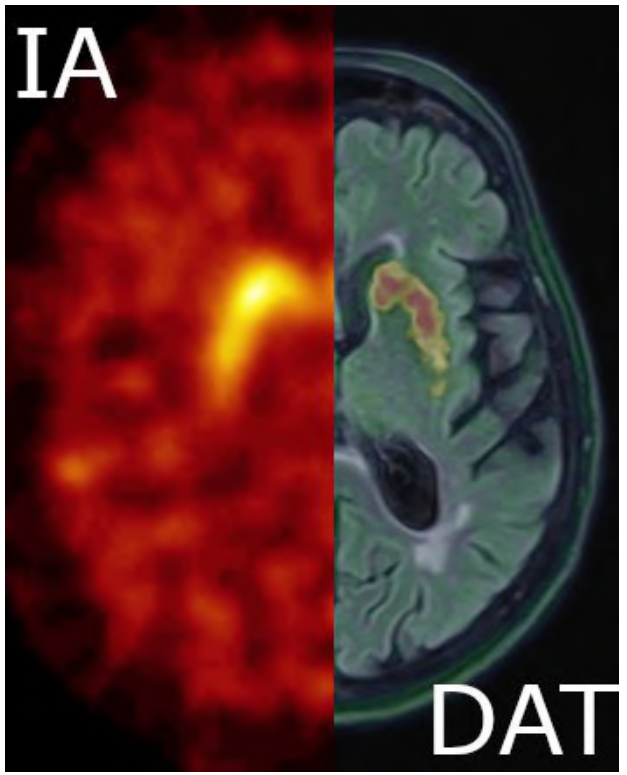




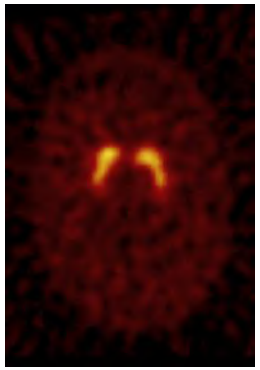
J Nucl Med. 2016 Jul;57(7):1006-13

DaT SPECT Imaging for Parkinsonism

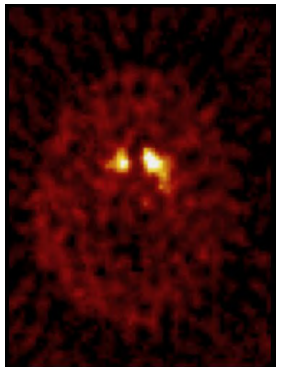




- Brain SPECT imaging of dopamine transporter activity (DaTSCAN©):
  - a binary biomarker of degenerative parkinsonism:
    - ✔ No dopaminergic loss in non-degenerative causes (≈ 1/3 of exams)
    - ✔ Dopaminergic loss in degenerative cause (≈ 2/3 of exams)
  - 50% of dopaminergic loss at diagnosis in case of degeneration
  - 17,250 exams per year in France, in ≈ 200 centres
  - Se/Sp: 97%/100%, inter-operator concordance: 95%
  - 20% of uncertain classification in non-expert interpretation in practice
  - Visual interpretation based on morphological/intensity criteria

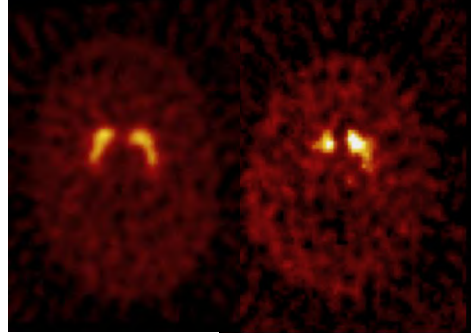


Normal



Pathological

The DaT-HUB data challenge



- A multicentric French national DaTSCAN© database to further develop in 2023 an automated diagnostic aid algorithm identifying normal vs pathological exams on a data challenge

- ≈ 5,000 exams to be collected (3D images, < 10 Mo per scan) with consensual binary interpretation among 3 national experts, and additional data (age, sex, acquisition parameters, confidence index, asymmetry index, severity index)

- The winner will be the algorithm with the performance of maximal concordance with the expert interpretation (25,000 € prize)

- Further scientific objectives on possible interaction with French healthcare database (SNDS):
  - Correlation between the algorithm score and the quantitative visual score, according to the clinical severity
  - Identification of possible iatrogenic or neuroprotective drugs
  - Impact of certain conditions on dopamine activity (such as long COVID)
  - …

Data Challenge HUS : D-IA-GNO-DENT - HDH Data Challenge - Rare oral diseases / Artificial intelligence / Diagnostic assistance

Data Challenges for Health

4:25 – 4:35

Pr Agnès Bloch Zupan

(PU-PH) Hôpitaux Universitaires de Strasbourg,Centre de référence coordonnateur des maladies rares orales et dentaires CRMR O-Rares Université de Strasbourg, Faculty of Dental Surgery and Institute of Genetics and Molecular and Cellular Biology, CNRS- UMR7104, INSERM U1258

Agnès Bloch-Zupan is Professor in Oral Biology at the Faculty of Dentistry, Université de Strasbourg, France. She coordinates the Reference Center for Rare Oral and Dental Diseases, CRMR O-RARES of the University Hospitals of Strasbourg and affiliated French constitutive and competence centers. She supervises clinical research projects and is the scientific director of the D(4)/phenodent database (www.phenodent.org) as well as the genetic diagnosis NGS tool GenoDENT.

GOUVERNEMENT Liberté Égalité Fraternité · Open Government Partnership · Conseil de l'Innovation · bpifrance · HDH Health Data Hub · France22 Présidence française du Conseil de l'Union européenne

D_IA_GNO_DENT : Artificial intelligence helping diagnosis of rare diseases with oral expressions

Pr Agnès Bloch-Zupan

PU-PH professor, clinician, rare disease expert, researcher at IGBMC
UMR Unistra 7104, Inserm U1258

Coordinator CRMR O-Rares Rare diseases Reference Center HUS and network, French Filière TETE COU, ERN CRANIO

Guillaume Icre
Innovation and Valorization HUS
Christine Lecomte E-Health HUS

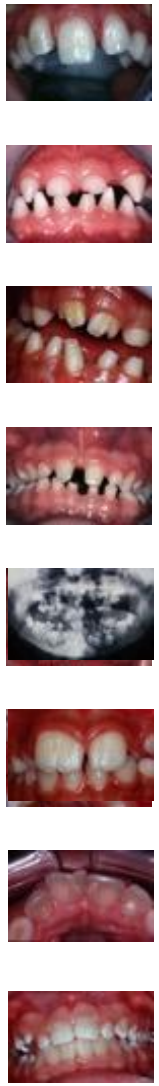
Rare oral and dental Diseases (O-RARES) Network

FILIÈRE SANTÉ MALADIES RARES TeteCou
MALFORMATIONS RARES TÊTE COU DENTS

Fondation Maladies Rares
ATELIER
de co-design en e-santé

2019 Winner

Context and challenges



Developmental abnormalities of the oral cavity and teeth

900 rare diseases/syndromes with **dental anomalies** and **other associated symptoms**

Dental development is regulated by numerous genes

Mineralization fixes in time the anomalies

Dental defects exist in possible combination with **disorders of other systems** : retina, skeleton, kidney, skin, epilepsy...

Currently there is **an important diagnostic wandering :** Patients are not diagnosed with “rare disease" and do not benefit from specific personalized care.

The idea

A virtual expert for diagnosis help. Transdisciplinary analysis **for improved patient care.**

How

An **AI**-based solution analyzing images of the oral cavity, **for automatic recognition of diagnostic signatures.**

Why

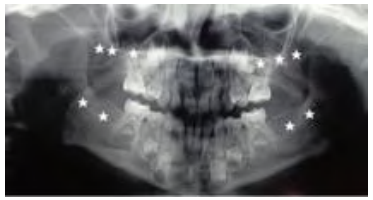
Offer the expertise of the O-Rares network to health professionals.

Move from diagnostic wandering to a personalized diagnostic itinerary!

Current work

1 Rare diseases with oral expression

- Patients, **data D4/phenodent database**, genetic testing NGS **GenoDENT** 567 genes, ARS
- Life and Health research expertise
- An AI-based solution analyzing panoramic radiographs **i-DENT**
  AMI Grand-Est 2019
- For **automatic recognition of radiographic diagnostic signatures**



PAX9 – Absence of molars
AD / Heterozygous deletion
PAX9 (NM_006194.3):
c.(?_-115)_(*62_?)del.

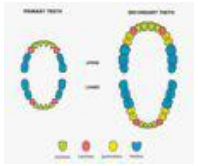
DIGITAL ECONOMY

2 A.I.

Use of a combination of 2 neural networks

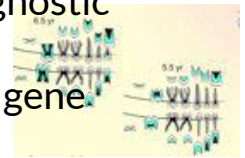
X-Rays:

- Identification and segmentation of teeth types (I,Ca,PreM,Mo)



Coming soon :

- Identification of teeth according to the FDI nomenclature
- Dental age
- Correspondance to a diagnostic signature pattern
- From the signature to the gene



3 HDH : to go further…

- **Combining radiographic and clinical signs (photos)** to identify rare diseases.
- Other parameters may come into play:

size, shape, color, texture,

Patient history…

Data/Cohorts

Amelogenesis imperfecta 200 + 20 - Dentine anomalies 50 - Environmental anomalies 50 - Control 100

Intraoral colour photographs (4 to 6)
Colour, Surface texture, Shape/size

Diagnosing Dental Defects Database

phenodent.org

Key-words D4/phenodent – HPO : AI hypoplastic, hypomineralized, hypomature; Medical information; Gene - GenoDENT; Diagnosis

Developped solution

For whom?

Designed for rare disease patients

- Decrease diagnostic wandering
- Accelarate care in the « right » care pathway

By whom?

Used by doctors and medical research

- Access to enhanced knowledge of rare diseases
- Acces to rapid expertise

How ?

Input and output data

- Inputs : images, reports, diagnostic signatures…
- Output : a (pre-)diagnosis of suspected rare disease with oral manifestations

What ?

Intelligent virtual assistant

- An expert tool to help with pre-diagnosis, or even diagnosis when faced with a suspected rare disease with oral expressions.
- Thanks to a collection of descriptive data, managed by an artificial intelligence engine, it will guide the professional in a personalized way.

Experts

Les Hôpitaux Universitaires de STRASBOURG — igbmc

Expertise in Health Research and Patient Data

+

Platform

Secure and adapted data transfer/development

+

Challengers

A.I. expertise

FONDATION Maladies Rares

Heath Professionals and Patients

→ **Creation and testing of an A.I. based teleexpertise software** ← ORares

Timeline

Step 0

PREPARE HDH/BPI

HDH MOU
BPI application
BPI official notification

BPI signature /HUS

MAY 2022

Step1

BUILD Data Challenge

Scientific committee
Executive committee
Contracts
Data protection, security
RGPD
Data preparation
Communication/challengers

JUNE 2022-FEBRUARY 2023

Step2

CHALLENGE DATA

PLATFORM
COMPETITION

MARCH-MAY 2023

Step3

POST Data Challenge

Algorithms tests
REWARD

Administrative closure
Publications
Open data
Communication

JUNE-OCTOBER 2023

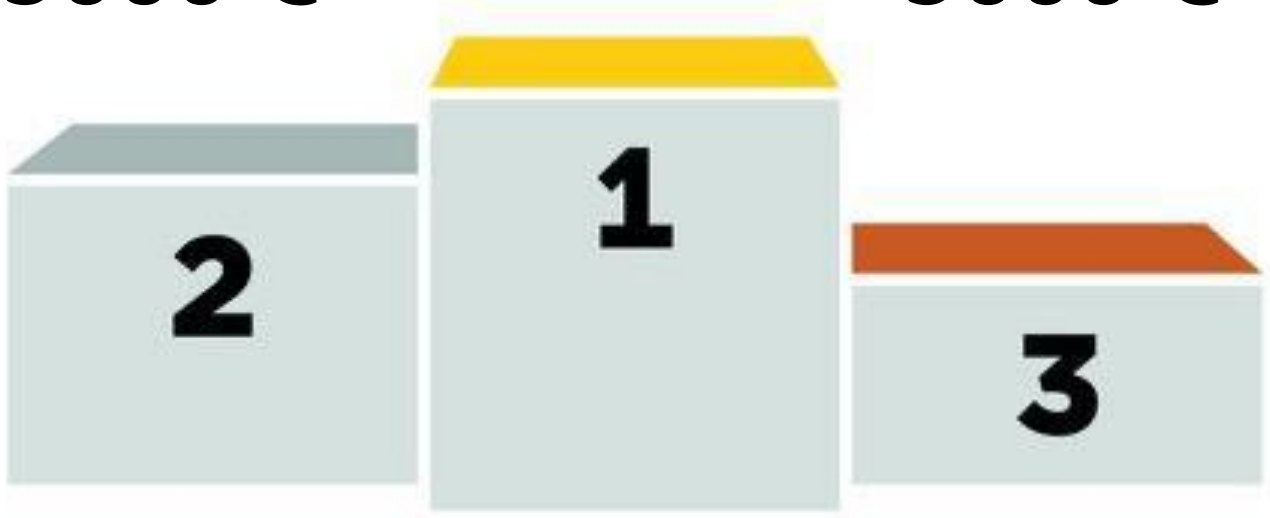
Challenge and reward

<u>European challengers : 30 000 euros</u>

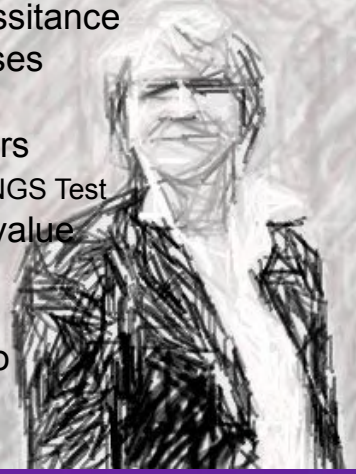
20 000 €

5000 €

5000 €



Virtual expert
Diagnostic Assitance
Rares Diseases
Signatures
Expert Centers
GenoDENT NGS Test
RMT Added value
Upper Rhine
From wandering to **Diagnostic Itinerary**



Les Hôpitaux Universitaires de STRASBOURG

igbmc institut de génétique et de biologie moléculaire et cellulaire

FONDATION Maladies Rares

Université de Strasbourg

PLATEFORME DES DONNÉES DE SANTÉ

ORares MALADIES RARES ORALES ET DENTAIRES

Data Challenge SFA : Allergen Chip Challenge: artificial intelligence at the service of the allergic patient

Data Challenges for Health

4:35 – 4:45

Pr Joana Vitte

MD, PhD, Associate Professor - Hospital Practitioner in Immunology, Aix-Marseille University and Montpellier University, President of the Scientific Committee of the French Society of Allergology.

Joana Vitte is a specialist in immunological mechanisms and tools in allergology, combining hands-on hospital work, innovative diagnostic, translational and epidemiological research and undergraduate and postgraduate training. She currently acts as the coordinator of the Allergen Chip Challenge project in collaboration with Health Data Hub and funded by the Banque Publique d'Investissement, aiming at further developments in the field of open data and artificial intelligence tools for allergists and immunologists.

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

HEALTH DATA HUB

Data Challenges for Health

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

Principal investigator: Société Française d'Allergologie

Les Allergies.fr
SFA
Société Française d'Allergologie

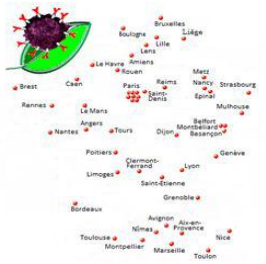
The Allergen Chip Challenge

Speaker: Dr Joana Vitte, MD, PhD

Associate Professor, Aix-Marseille University, MEPHI

Associate Researcher, Montpellier University, Desbrest Institute of Epidemiology and Public Health

President of the Scientific Council of the French Society of Allergology



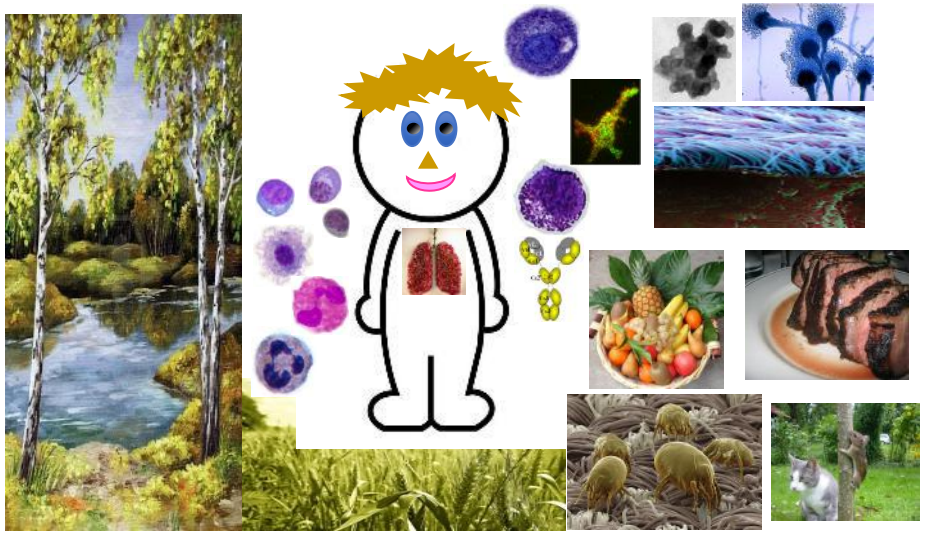
AllergoBioNet
Le Réseau d'ImmunoAllergologie Biologique Hospitalière

IDESP

Inserm

UNIVERSITÉ MONTPELLIER

Context: Exposure to potential allergens is universal, but not everyone develops IgE antibodies and allergy



The exposome concept and the immune response

Allergy = antigen-specific loss of tolerance

Allergen = an antigen that causes an allergic disease by inducing an adaptive immune response

Allergen-specific immunoglobulin (Ig)E = the hallmark of allergic immune response

Allergy diagnosis = clinical history + demonstration of immunoglobulin (Ig) E to culprit allergen

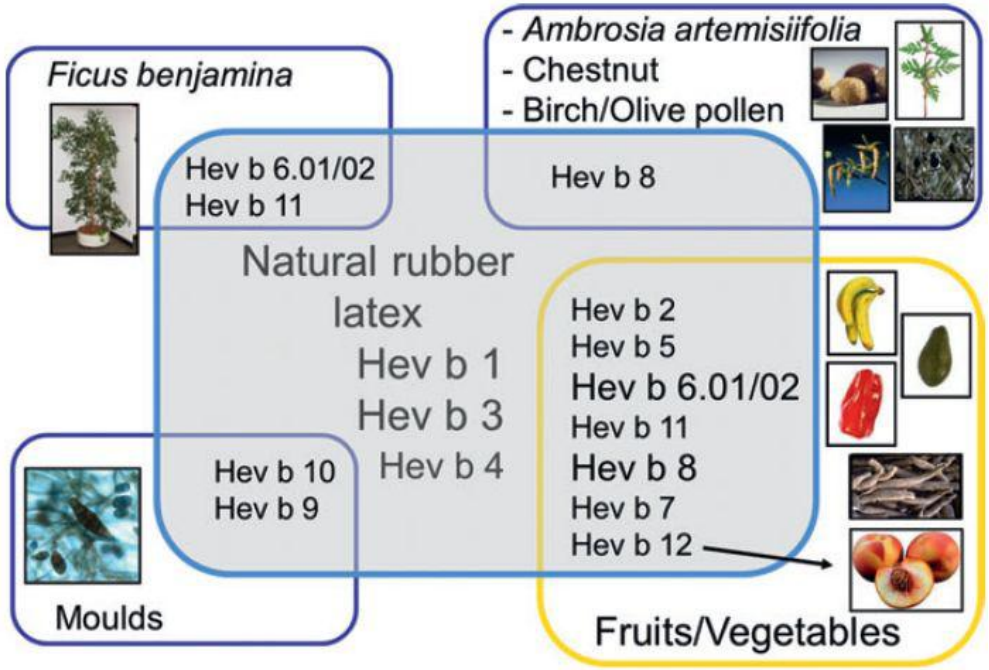
Allergens come in a virtually unlimited variety, and so do IgE

ALLERGEN NOMENCLATURE
WHO/IUIS Allergen Nomenclature Sub-Committee
Financial contribution from IUIS, EAACI, and AAAAI organizations

http://allergen.org/

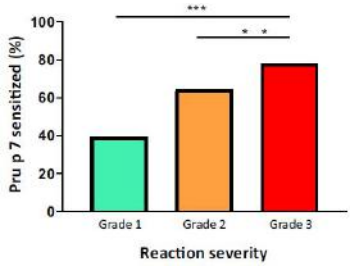
Bacher P & Scheffold A, *J Allergy Clin Immunol* 2018
Chruszcz M *et al*, *Allergy* 2021

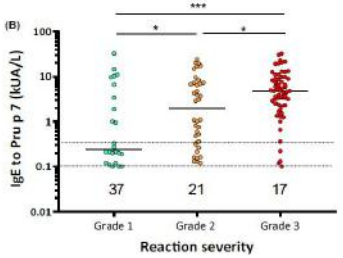
Allergens are species-specific or cross-reactive diagnosis



= cross-recognition by IgE and induction of multiple symptoms

IgE presence, levels, family and clusters are associated to severity stratification

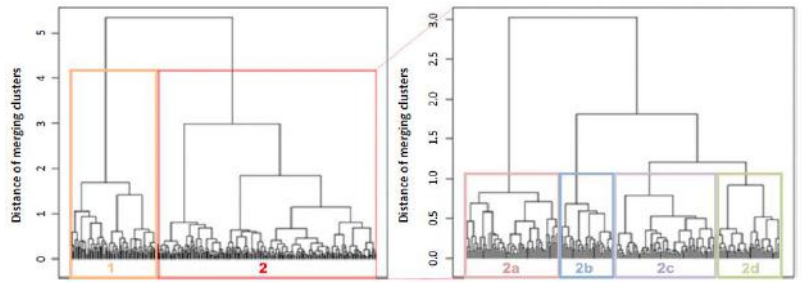




Severity of peach allergy as a function of the prevalence and levels of IgE to allergen Pru p 7

Clinical expression as a function of the allergen family and cluster

IgE to panallergens	Multi-panallergen	Profilin	nsLTP	No-panallergen	PR-10
Number	85	63	36	66	44
Geographic area	North	Center	South	Center	Center
Pollens	birch grass	grass plane olive plantain	pellitory	grass	grass
Triggering foods	all	peach kiwi banana Cucurbitaceae	Rosaceae banana peanut hazelnut	kiwi	apple peach kiwi
Comorbidities	asthma UAH AD	asthma	UAH AD	-	-



Matricardi PM *et al*, *Pediatr Allergy Immunol* 2016;
Mastrorilli C *et al, Allergy* 2016;
Klingebiel C *et al, Clin Exp Allergy* 2019

Visit card of an allergic patients: IgE landscape + clinical history

2. IgE results sorted by protein group

Mainly species-specific food components				
Egg white	nGal d 1	Ovomucoid	<0.3	ISU-E
	nGal d 2	Ovalbumin	<0.3	ISU-E
	nGal d 3	Conalbumin/Ovotransferrin	<0.3	ISU-E
Egg yolk/chicken meat	nGal d 5	Livetin/Serum albumin	2,8	ISU-E
Cow's milk	nBos d 4	Alpha-lactalbumin	<0.3	ISU-E
	nBos d 5	Beta-lactoglobulin	<0.3	ISU-E
	nBos d 8	Casein	<0.3	ISU-E
	nBos d lactoferrin	Transferrin	<0.3	ISU-E
Cod	rGad c 1	Parvalbumin	<0.3	ISU-E
Shrimp	nPen m 2	Arginine kinase	<0.3	ISU-E
	nPen m 4	Sarcoplasmic calcium binding protein	<0.3	ISU-E
Cashew nut	rAna o 2	Storage protein, 11S globulin	1	ISU-E
Brazil nut	rBer e 1	Storage protein, 2S albumin	<0.3	ISU-E
Hazelnut	nCor a 9	Storage protein, 11S globulin	0,5	ISU-E
Walnut	rJug r 1	Storage protein, 2S albumin	<0.3	ISU-E
	nJug r 2	Storage protein, 7S globulin	<0.3	ISU-E
Sesame seed	nSes i 1	Storage protein, 2S albumin	<0.3	ISU-E
Peanut	rAra h 1	Storage protein, 7S globulin	57	ISU-E
	rAra h 2	Storage protein, Conglutin	>100	ISU-E
	rAra h 3	Storage protein, 11S globulin	>100	ISU-E
	nAra h 6	Storage protein, 2S albumin	>100	ISU-E
Soybean	nGly m 5	Storage protein, Beta-conglycinin	<0.3	ISU-E
	nGly m 6	Storage protein, Glycinin	60	ISU-E
Buckwheat	nFag e 2	Storage protein, 2S albumin	<0.3	ISU-E
Wheat	rTri a 19.0101	Omega-5 gliadin	<0.3	ISU-E
	nTri a aA_TI	Alpha-amylase / Trypsin inhibitor	<0.3	ISU-E
Kiwi	nAct d 1	Cysteine protease	<0.3	ISU-E
	nAct d 5	Kiwellin	<0.3	ISU-E

Mainly species-specific aeroallergen components				
Grass pollen				
Bermuda grass	nCyn d 1	Grass group 1	<0.3	ISU-E
Timothy grass	rPhl p 1	Grass group 1	<0.3	ISU-E
	rPhl p 2	Grass group 2	0,6	ISU-E
	nPhl p 4	Berberine bridge enzyme	<0.3	ISU-E
	rPhl p 5	Grass group 5	<0.3	ISU-E
	rPhl p 6	Grass group 6	<0.3	ISU-E
	rPhl p 11	Ole e 1-related protein	<0.3	ISU-E
Tree pollen				
Birch	rBet v 1	PR-10 protein	<0.3	ISU-E
Japanese cedar	nCry j 1	Pectate lyase	0,5	ISU-E
Cypress	nCup a 1	Pectate lyase	0,5	ISU-E
Olive pollen	rOle e 1	Common olive group 5	19	ISU-E
	rOle e 9	Beta-1,3-glucanase	<0.3	ISU-E
Plane tree	rPla a 1	Putative invertase inhibitor	13	ISU-E
	nPla a 2	Polygalacturonase	<0.3	ISU-E

Weed pollen				
Ragweed	nAmb a 1	Pectate lyase	<0.3	ISU-E
Mugwort	nArt v 1	Defensin	<0.3	ISU-E
Goosefoot	rChe a 1	Ole e 1-related protein	<0.3	ISU-E
Wall pelitory	rPar j 2	Lipid transfer protein (nsLTP)	<0.3	ISU-E
Plantain	rPla l 1	Ole e 1-related protein	<0.3	ISU-E
Saltwort	nSal k 1	Pectin methylesterase	<0.3	ISU-E
Animal				
Dog	rCan f 1	Lipocalin	<0.3	ISU-E
	rCan f 2	Lipocalin	2,3	ISU-E
	rCan f 5	Arginine Esterase	<0.3	ISU-E
Horse	rEqu c 1	Lipocalin	<0.3	ISU-E
Cat	rFel d 1	Uteroglobin	1,2	ISU-E
	rFel d 4	Lipocalin	<0.3	ISU-E
Mouse	nMus m 1	Lipocalin	<0.3	ISU-E
Mold				
Alternaria	rAlt a 1	Acidic glycoprotein	<0.3	ISU-E
	rAlt a 6	Enolase	<0.3	ISU-E
Aspergillus	rAsp f 1	Mitogillin family	<0.3	ISU-E
	rAsp f 3	Peroxysomal protein	<0.3	ISU-E
	rAsp f 6	Mn superoxide dismutase	<0.3	ISU-E
Cladosporium	rCla h 8	Mannitol dehydrogenase	<0.3	ISU-E
Mite				
B. tropicalis (HDM)	rBlo t 5	Mite group 5	<0.3	ISU-E
D. farinae (HDM)	nDer f 1	Cysteine protease	<0.3	ISU-E
	rDer f 2	NPC2 family	<0.3	ISU-E
D. pteronyssinus (HDM)	nDer p 1	Cysteine protease	<0.3	ISU-E
	rDer p 2	NPC2 family	<0.3	ISU-E
L. destructor (storage mite)	rLep d 2	NPC2 family	<0.3	ISU-E
Cockroach				
Cockroach	rBla g 1	Cockroach group 1	<0.3	ISU-E
	rBla g 2	Aspartic protease	<0.3	ISU-E
	rBla g 5	Glutathione S-transferase	<0.3	ISU-E

Cross-reactive components				
PR-10 protein				
Peach	rPru p 1	PR-10 protein	<0.3	ISU-E
Soybean	rGly m 4	PR-10 protein	<0.3	ISU-E
Peanut	rAra h 8	PR-10 protein	1,7	ISU-E
Kiwi	rAct d 8	PR-10 protein	<0.3	ISU-E
Celery	rApi g 1	PR-10 protein	<0.3	ISU-E

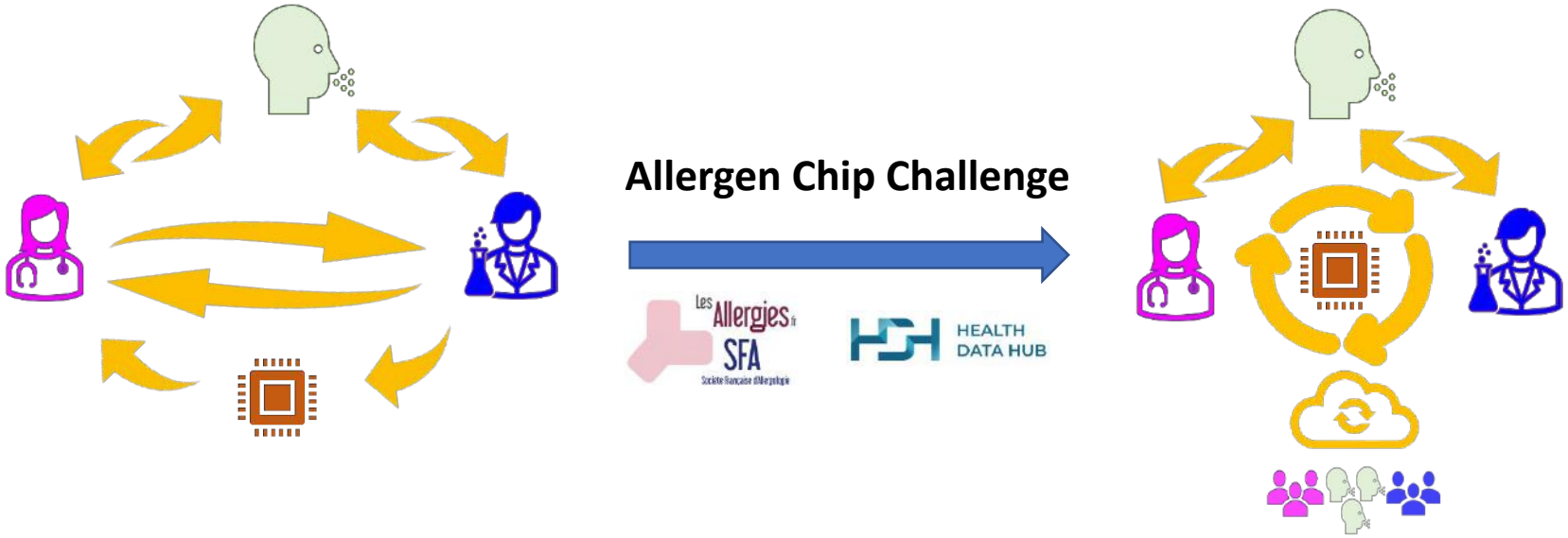
… and counting… to >100 and up to 300 allergen-specific IgE

Female, 12 years
Life-threatening allergic reaction to peanut and soybean
Previously allergic to egg, now tolerates it
Olive pollen rhinitis and asthma

Questions: should this patient carry an epinephrine autoinjector? be allowed to eat cashew nuts? is this patient going to outgrow the peanut/soybean allergy or rather react to further related plant foods? Should the family dog be banned from the sofa in the living room?

How to make the most of this set of data for the patient's management?

Allergen Chip Challenge: the idea behind the project



Data flow for allergy diagnosis using allergen multiplex investigation: current situation and projection of Allergen Chip Challenge outcome.

a, Current situation: multiple pathways and significant disconnection for *in vitro* and clinical data, hampering integrated interpretation.

b, Allergen Chip Challenge objective: data collection, database implementation, data challenge for an AI solution efficiently bridging clinical and *in vitro* data for diagnosis, prognosis, therapeutic choice and follow-up of allergic diseases using allergen multiplex investigation.

Allergen Chip Challenge

Artificial Intelligence project in Molecular Allergology
Partnership: French Society of Allergology and Health Data Hub

Timeline

- May 2021: application
- September 2021: accepted
- January 2022: financial and regulatory application
- Notification of acceptance: May 2022
- Expected completion: December 2023

Background:

- **allergen multiplex investigation** → hundreds of numerical data for each patient → **complex issue** for allergists
- **Health Data Hub** has initiated a **« Grand Challenge program on Artificial Intelligence for improving medical diagnosis »** and will partner with medical societies to implement applied research projects

Objectives:

- **Data collection**: allergen multiplex and clinical data
- **Data challenge**
- **Development of an AI solution** for the clinical interpretation of IgE sensitization data obtained through allergen multiplex investigation
- **Open science**.

Scientific partners:

- **French Society of Allergology** = principal investigator
- **Desbrest Institute of Epidemiology and Public Health** (IDESP, INSERM UMR UA 11, Montpellier, France)
- **AllergoBioNet** (French association of allergy lab specialists from university and public hospitals)

Data:

3,000 allergen multiplex (ISAC, ALEX, FABER) and corresponding clinical data (retrospective collection) → > 500 000 IgE data with corresponding clinical information.

Challenge:

Database hosted by a data challenge platform, open to international competitors. Selection of the algorithm with the best ability to predict clinical data based on IgE sensitization

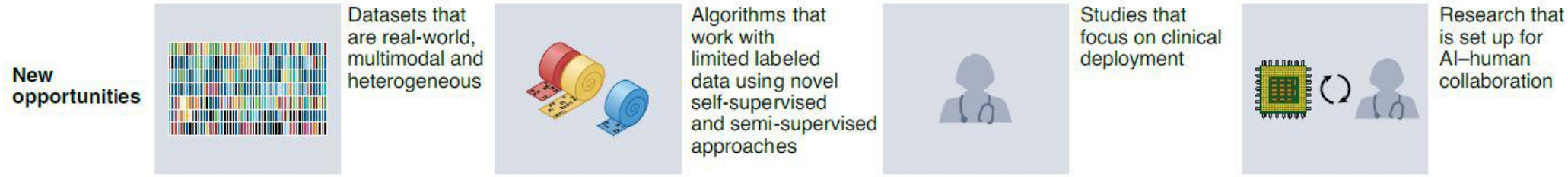
Expected outcome:

- **Open source of AI solution or partnership with the French Society of Allergology**

Key messages

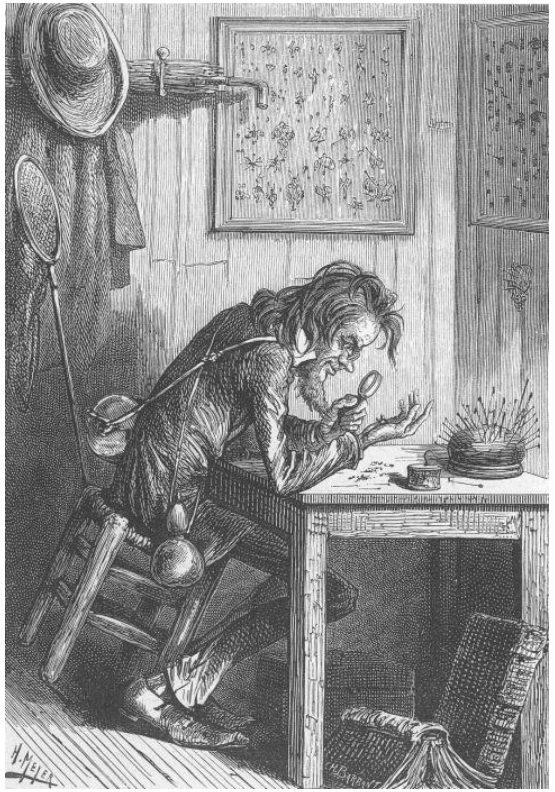
- **Hot topic for Allergology**
- To get hold of the relationship exposome – IgE – clinical expression of allergy
- To efficiently diagnose and stratify allergic patients
- To better understand the IgE clusters as a function of age, symptoms, and climate area

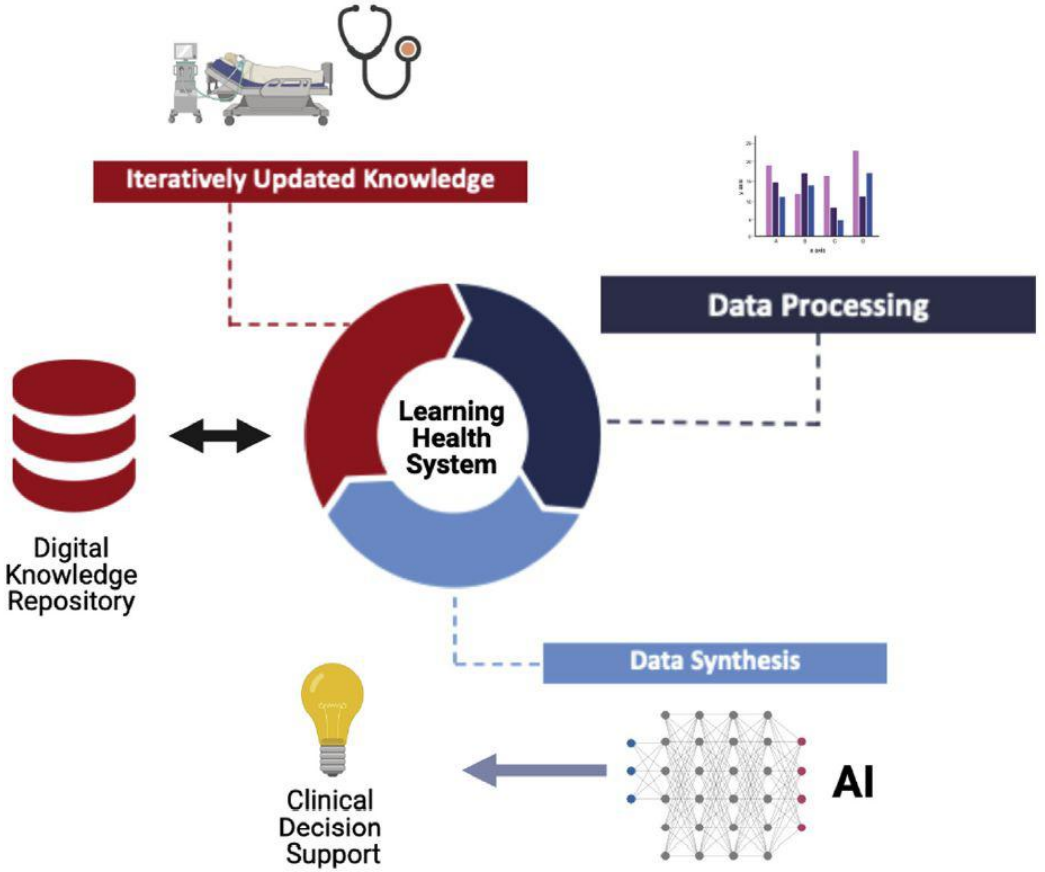
- **Hot topic for data challenge competitors**
- Dataset Excel sheet
- Tackling immune response complexity
- Open data and open science approach
- Data challenge winner: prize: 25 k€



Rajpurkar P *et al*, *Nature Med* 2022

Thank you for your attention





Cousin Bénédict the entomologist
(Dick Sand, *A Captain at Fifteen*, Jules Verne, 1878)

Khoury P *et al*, *J Allergy Clin Immunol Pract* 2022

Open science: sharing and reusing research data

Data Challenges for Health

Dr Lorien Benda

Open science project manager at Health Data Hub

Lorien Benda has a PhD in chemistry and is a project manager at the Health Data Hub. She is in charge of the open science, open source strategy and interoperability strategy aiming at decompartmentalising knowledge around health databases in order to facilitate their reuse.

Isabelle Blanc

National Director of Data, Algorithms and Source Codes for Research and Innovation of the Ministry of Higher Education, Research and Innovation in France.

Isabelle Blanc, originally a researcher in the field of human nutrition, has a leading role in the public policies of Open Science, especially on opendata and opensource. She drives all actions aiming at structuring, sharing and opening up research data, as well as opening up and promoting software source code produced by research.

4:45 – 5:00

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE

The Health Data Hub's open science approach

The Health Data Hub and its **partnerships** are **developing, making available and updating a series of tools designed to facilitate the manipulation of health data**. Because understanding data - their organisation, but also their biases, potentials and limits - represents a considerable effort for each project leader, the Health Data Hub proposes to **share knowledge: this is the open science approach.**

Open science approch

A collaborative work

Data Challenge — Data Challenges for Health

French national medico-administrative database (SNDS)

Open source library of validated algorithms (BOAS project) — boas

Other health database

Contributors

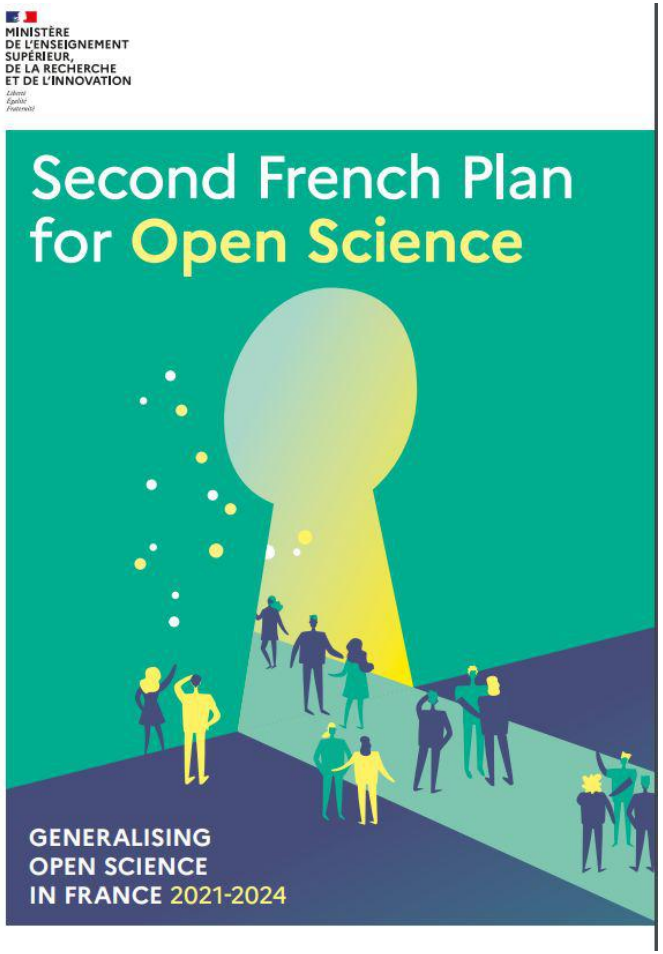
HEALTH DATA HUB

FRANCE22
FRENCH PRESIDENCY OF THE COUNCIL OF THE EUROPEAN UNION

MINISTÈRE
DE L'ENSEIGNEMENT
SUPÉRIEUR,
DE LA RECHERCHE
ET DE L'INNOVATION
Liberté
Égalité
Fraternité

French open science policy : a dedicated chapiter to research data

Structuring, sharing and opening up research data



« We will encourage practices that favor **research data reuse.** »

Give recognition to and boost the reuse of research data

→ Encourage best practices in **data citation**.

→ Award an annual **research data prize** to highlight the work of exemplary projects and teams in preparing to reuse or reusing research data.

→ **Track the changing dynamics in opening up** data sets associated with publications that are stored in a selection of repositories, through the Open Science Barometer.

→ Launch a **Europe-wide call for proposals on reusing research data** as part of the European Research Area Network.

Isabelle Blanc
National Director of Data, Algorithms and Source Codes
Ministry of Higher Education, Research and Innovation

www.ouvrirlascience.fr/home/

@AMDAC_MESRI

Launch of a new call for projects!

Data Challenges for Health

NEW CALL FOR PROJECTS 2022

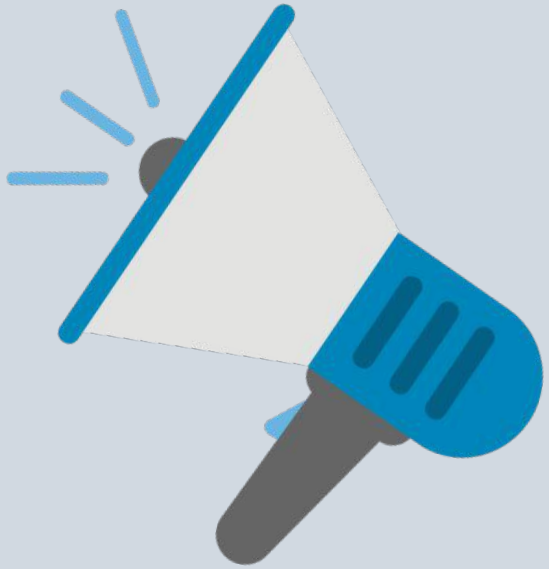
HEALTH DATA CHALLENGE

The Ministerial Delegation for eHealth (DNS) and the Health Data Hub are pleased to announce the launch of a new call for projects in 2022!

For more information, contact us on :
data.challenge@health-data-hub.fr

Would you like to be a **project leader** and propose a new **Data Challenge**?

Apply to be part of the upcoming edition of **Health Data Challenges** !



GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

Data Challenges for Health

THANK YOU !

STAY INFORMED !

Scan the QR code to get the latest informations on our Data Challenges !

GOUVERNEMENT
Liberté
Égalité
Fraternité

Open Government Partnership

CONSEIL DE L'INNOVATION

bpifrance

HEALTH DATA HUB

FRANCE22
PRÉSIDENCE FRANÇAISE DU CONSEIL DE L'UNION EUROPÉENNE