INSTITUT
JULES BORDET
INSTITUUT

Big Data and oncology: value and limitations

Philippe Aftimos, MD

Clinical Trials Development Leader
Clinical Trials Conduct Unit (CTCU)
Institut Jules Bordet

Disclosure information

- Consulting:
  - Boehringer Ingelheim, Macrogenics, Roche, Novartis, Amcure, Servier, G1 Therapeutics, Radius, Deloitte
- Honoraria:
  - Synthon, Amgen, Novartis, Gilead
- Travel grants:
  - Amgen, MSD, Pfizer, Roche
- Research funding to my institution:
  - Roche

What is Big Data?

Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation.

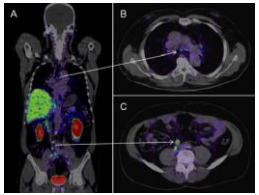
Gartner glossary (circa 2001)

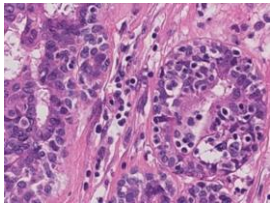
Credit: Ben Chams Fotolia

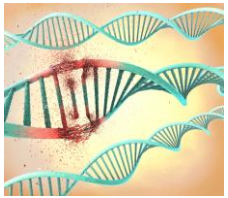
INSTITUT JULES BORDET INSTITUUT

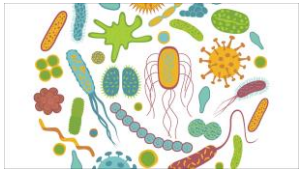
ULB iris

Big data sources in the treatment of cancer (not comprehensive)









INSTITUT
JULES BORDET
INSTITUUT

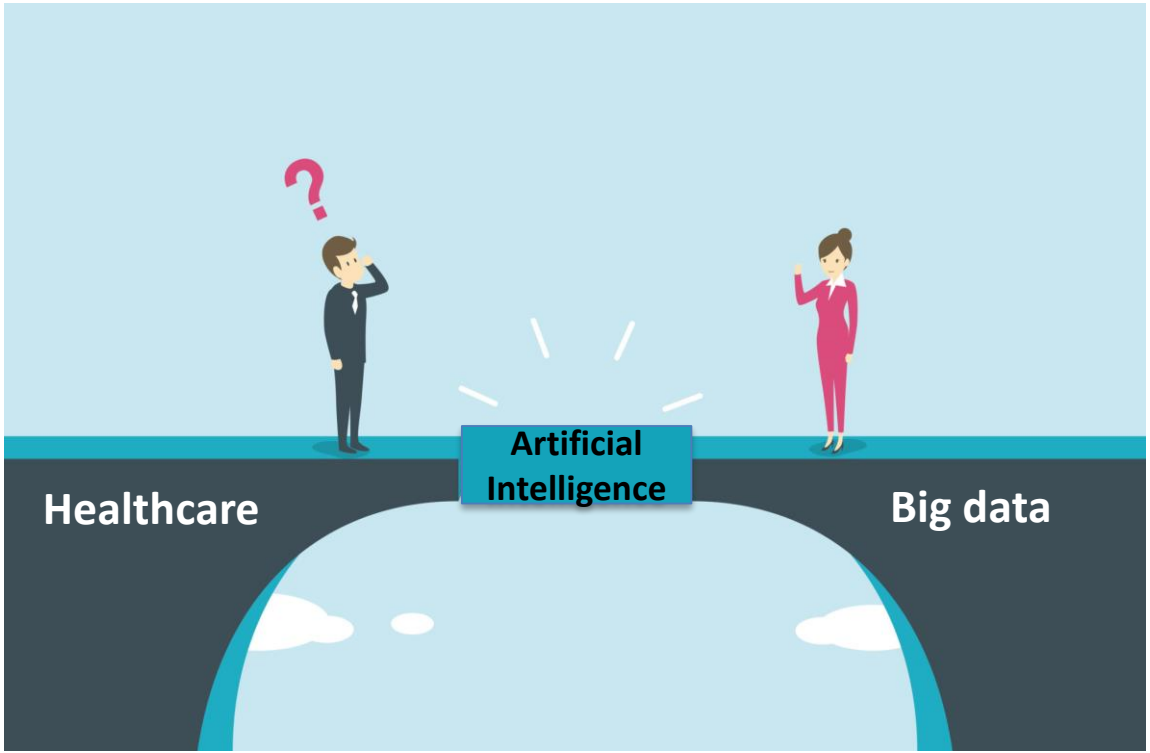
ULB iris

It’s not that simple

Regina Barzilay, an artificial intelligence researcher at MIT who was diagnosed with breast cancer in 2014 at the age of 43, was shocked by the paltry amount of data upon which her doctors based their clinical decision.

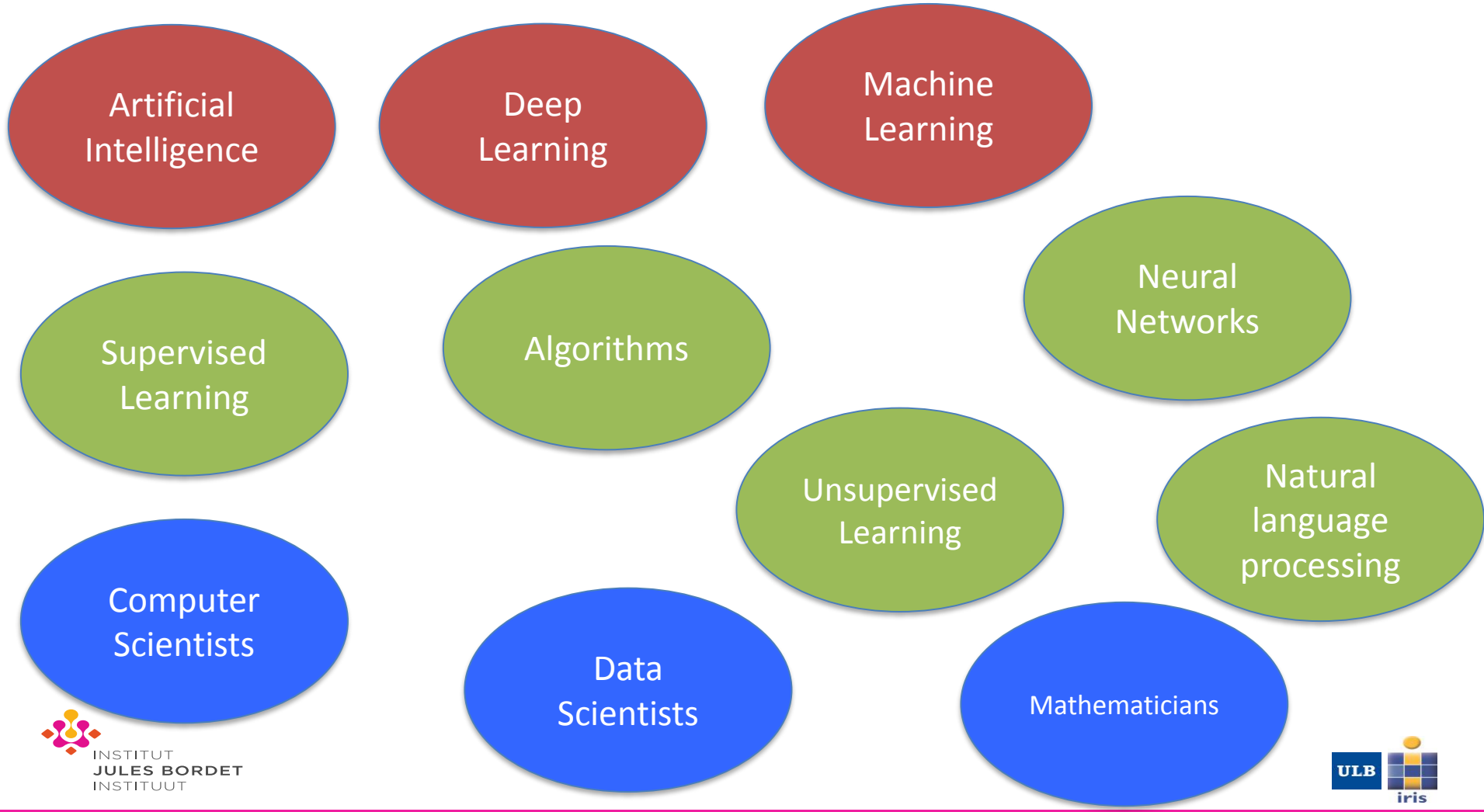
Credit: Sam Ogden (DFCI)

Bridging the gap



Leveraging big data in cancer care

- Early diagnosis and prevention
- Digital pathology and molecular pathology
- Drug discovery
- Treatment decisions
- Matching patients to clinical trials
- Anticipating adverse events



INSTITUT
JULES BORDET
INSTITUUT

ULB
iris

What is AI?

- AI refers to the ability of a machine to perform tasks commonly associated with intelligent human behavior.
- AI can be considered a group of iterative, "self-learning" techniques, which discover relationships within data that can evolve and often be performed faster over time.

Russel SJ et al. 2003.

Machine Learning (ML) vs Deep Learning (DL)

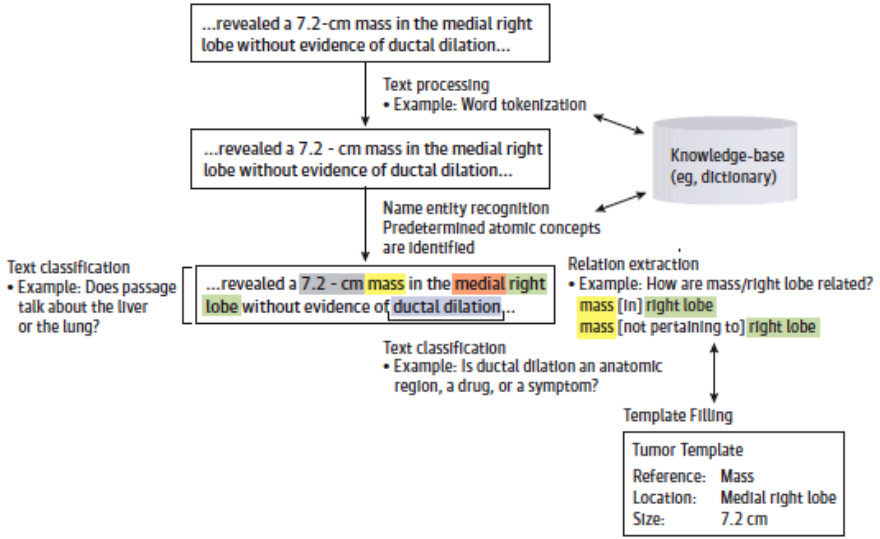
- Machine Learning:
  - ML algorithms, exposed to training data, are able to appreciate hidden patterns within the data which can then be used to perform a task without explicit programming
  - ML tasks are often broadly dichotomized into supervised or unsupervised learning

- Deep Learning:
  - DL is a form of ML that uses layered "artificial neural networks" to develop sophisticated models with the ability to understand data at different levels of abstraction

Samuel AL. IBM J Res Dev. 1959;3:210-29.
LeCun Y, et al. Nature. 2015;521:436-44.

Natural language processing (NLP)

- NLP is any computer-based algorithm that handles, augments, and transforms natural language so that it can be represented for computation
- This technology can harvest important clinical variables trapped in the free-text narratives within electronic medical records



Yim W et al. JAMA Oncology

ASCO CancerLinQ

Since 2015, More than **100** organizations **1,426,015** patients with a primary cancer diagnosis.

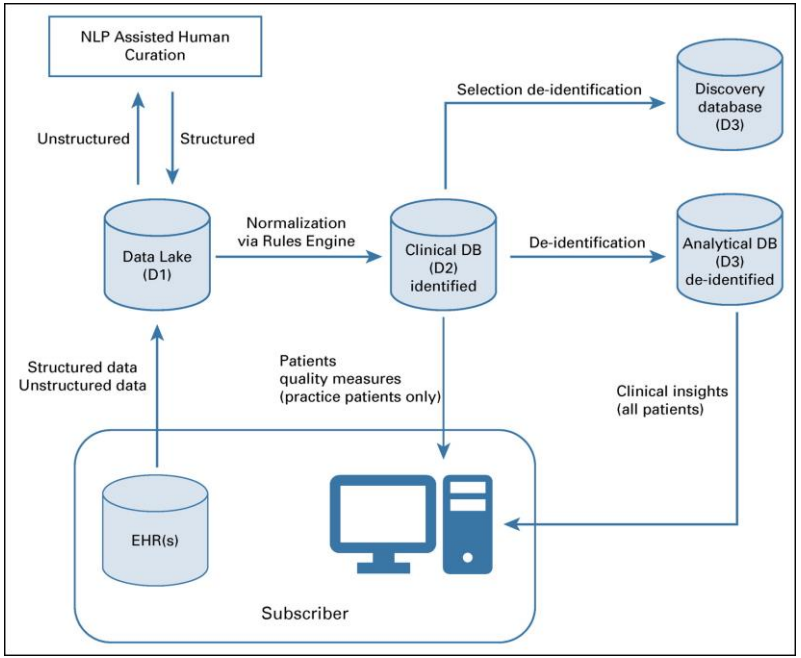


TABLE 1. Overview of CancerLinQ Quality Data Model Data Model Version 1.6.1

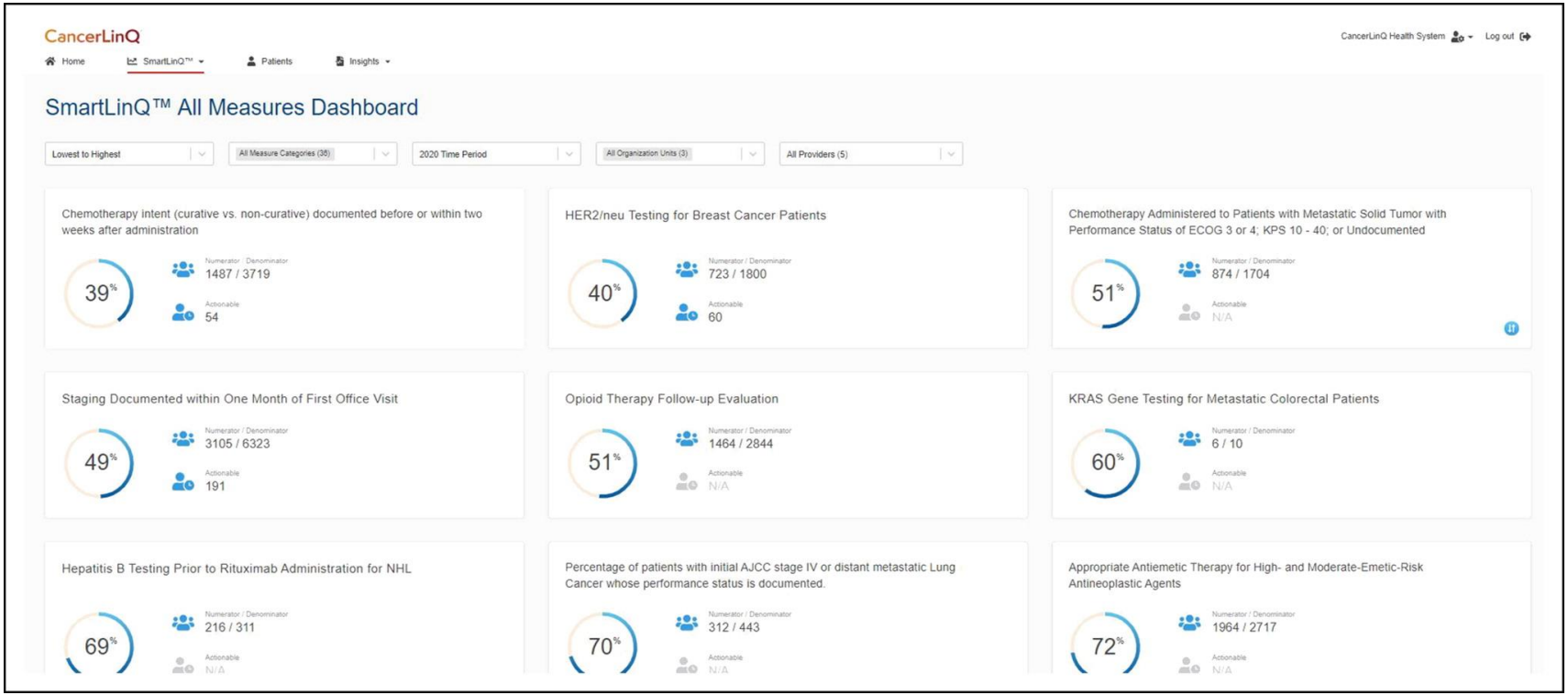
Data Type	Table Name	Contents
Demographics	Patient	Demographics
Diagnoses	diagnosis_active	Diagnoses-all types, problem list, adverse events
Pathology	procedure_performed	Staging, histology, laterality, tumor invasion, surgical margin, tumor size from pathology
Patient assessments	functional_status_performed	Performance status, therapeutic response, tumor progression
Risk assessments	risk_assessment	Mental health, pain, and smoking screening, family history
Physical exam values	physical_exam_performed	Physical examination values for the patient
Laboratory tests	laboratory_test_performed	Laboratory tests, including genetic tests
Care plan	care_plan	Treatment intent
Encounters	encounter_performed	Encounter codes (CPT, SNOMED, HCPCS)
Interventions	intervention_performed	Indication data of smoking cessation counseling, genetic counseling, chemotherapy, hospice, etc
Medications (administered)	medication_administered	Medication administered to the patient
Medications (ordered)	medication_ordered	Medication ordered for the patient
Radiation therapy	radiation_care_plan	Abstracted summary radiation therapy information, including type, location, dose, and total fractions
Radiation therapy	procedure_radiation	Transactional radiotherapy information, where available
Surgery	procedure_surgery	Abstracted summary surgical information
Imaging results	procedure_imaging	Abstracted imaging activities, including size of tumor and progression

Abbreviations: CPT, Current Procedural Terminology; HCPCS, Healthcare Common Procedure Coding System; SNOMED, Systemized Nomenclature of Medicine Clinical Terms.

INSTITUT JULES BORDET INSTITUUT

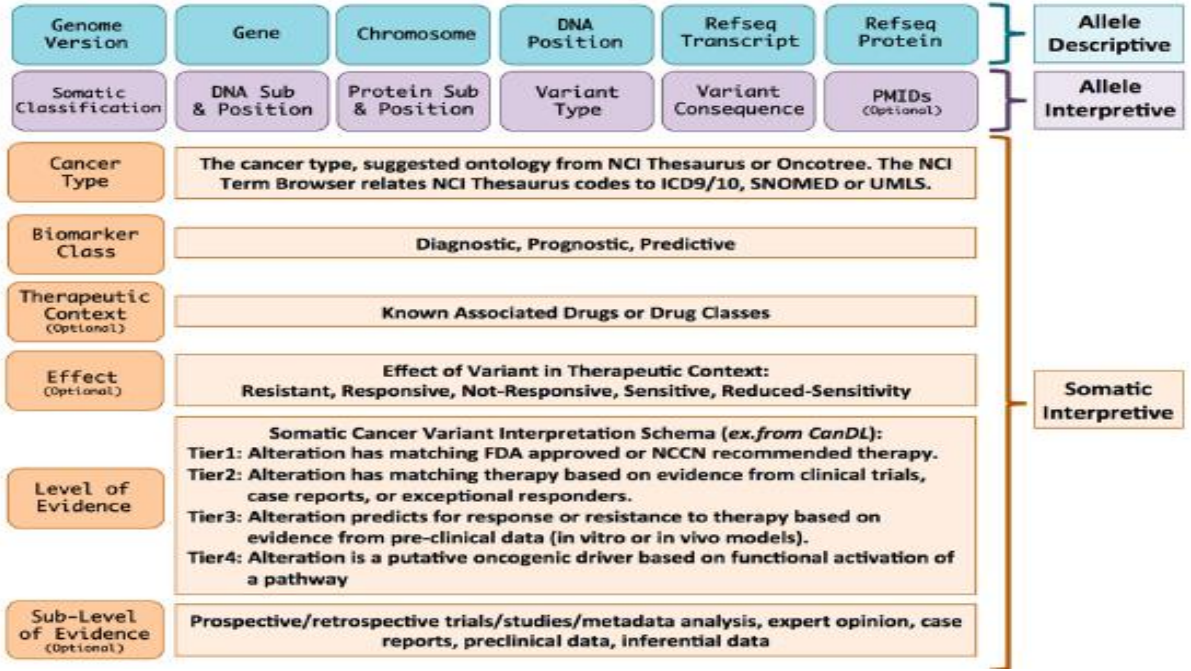
Potter D et al. JCO CCI 2020.

ULB iris



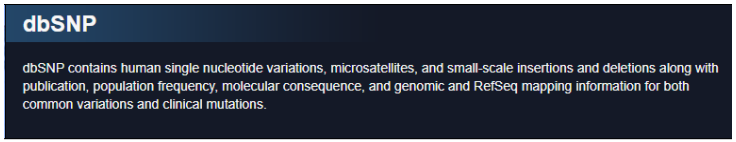
Potter D et al. JCO CCI 2020.

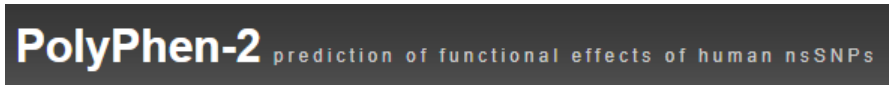
Standardization efforts for genomic data
(Minimum Variant Level Data MVLD)

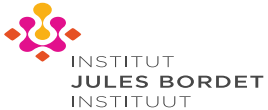


Ritter D et al. Genome Medicine 2016.

Genomics databases and selected tools for tertiary analyses

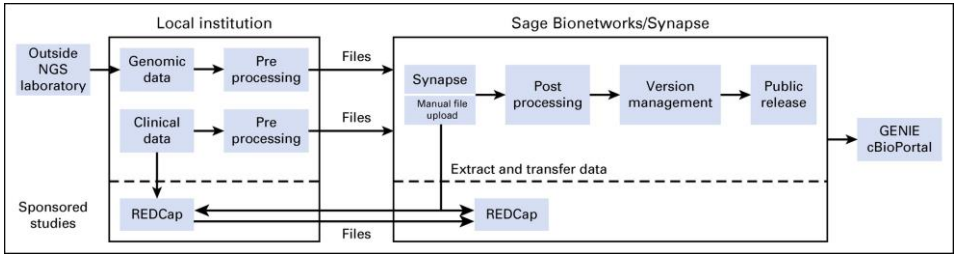


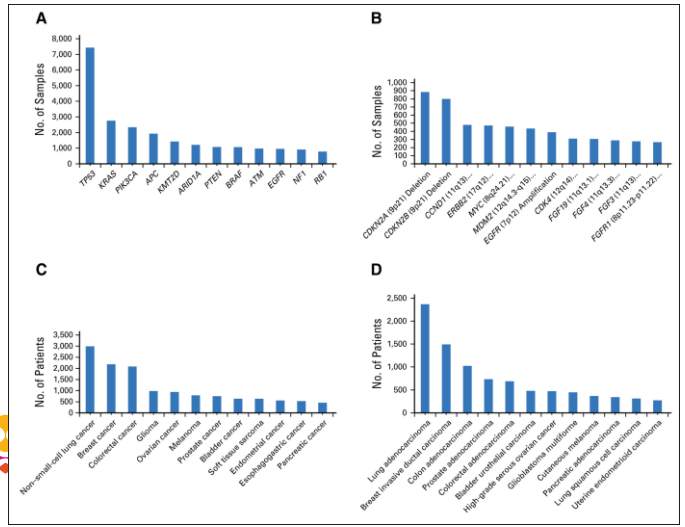




AACR
American Association for Cancer Research
FINDING CURES TOGETHER™

PROJECT GENIE
Genomics Evidence Neoplasia Information Exchange





AACR PROJECT GENIE: PUBLICATIONS

Linked Entity Attribute Pair (LEAP): A Harmonization Framework for Data Pooling
Journal of Clinical Oncology, July 2020

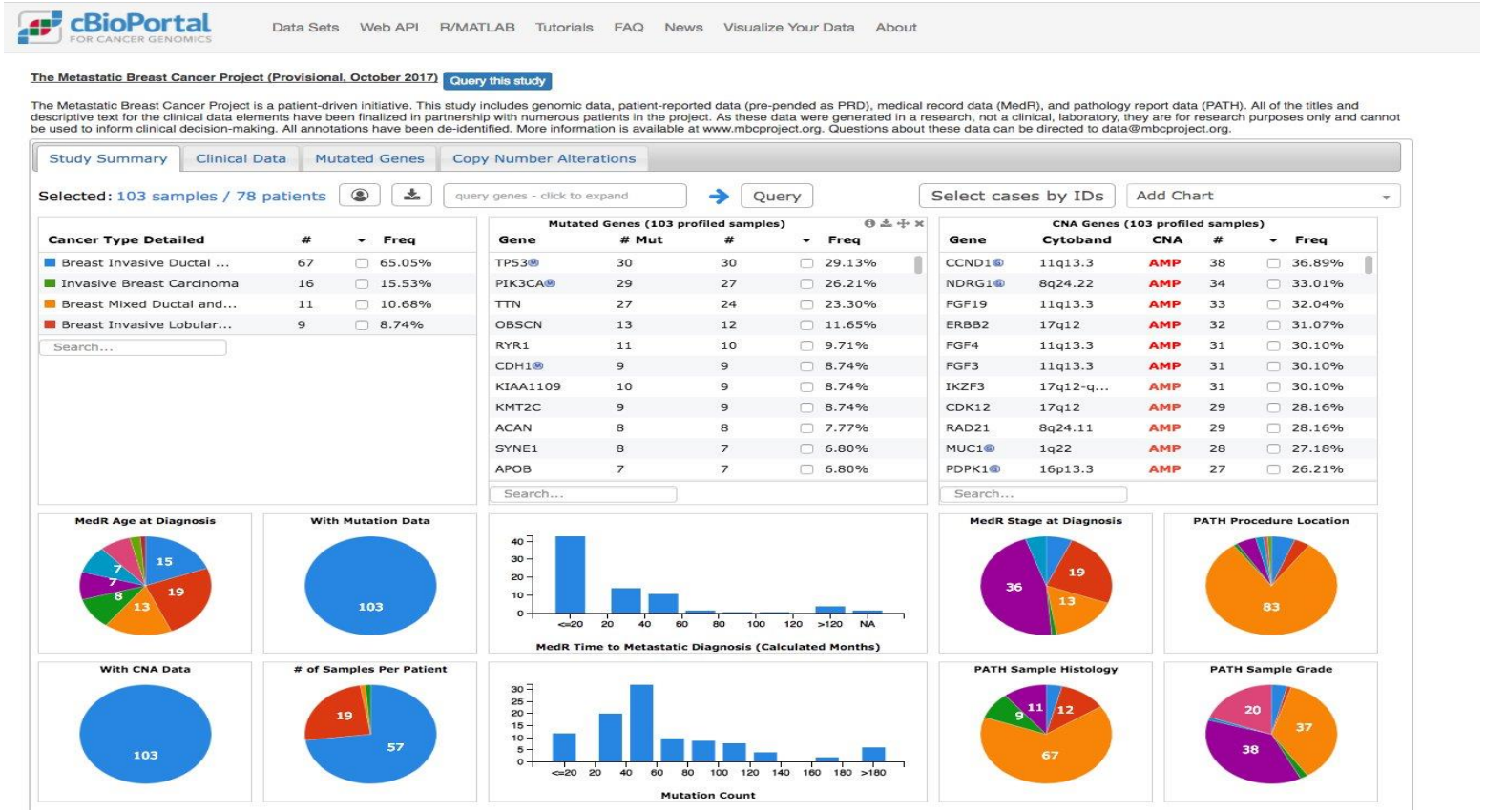
Characteristics and Outcome of AKT1E17K-Mutant Breast Cancer Defined through AACR Project GENIE, a Clinicogenomic Registry
Cancer Discovery, Vol. 10, Issue 4, 2020, April, 2020

American Association for Cancer Research Project Genomics Evidence Neoplasia Information Exchange: From Inception to First Data Release and Beyond—Lessons Learned and Member Institutions' Perspectives
Journal of Clinical Oncology, February 16, 2018

AACR Project GENIE: Powering Precision Medicine through an Internal Consortium
Cancer Discovery, Vol. 7, Issue 8, August 2017

Micheel C et al. JCO CCI 2018.

ULB iris

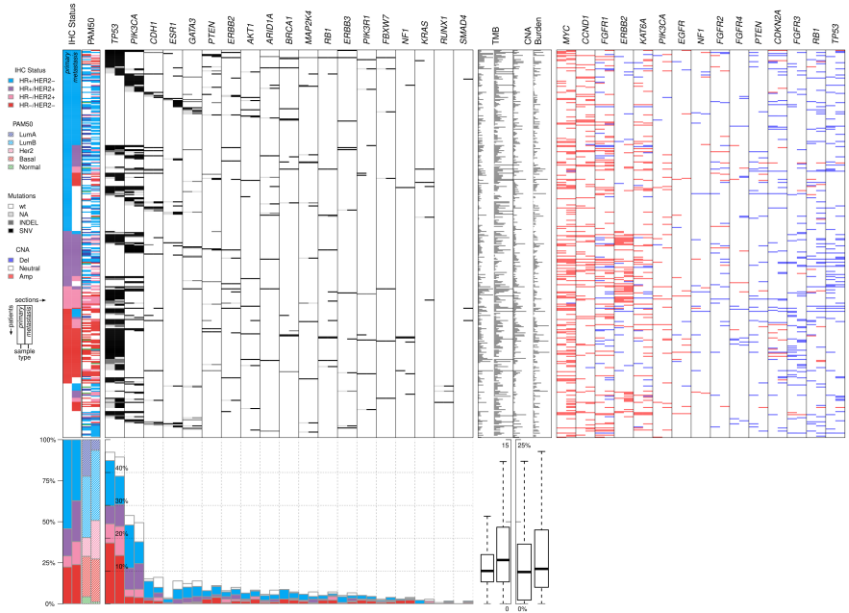


INSTITUT
JULES BORDET
INSTITUUT

ULB
iris

AURORA: Aiming to Understand the Molecular Aberrations in Metastatic Breast Cancer

AURORA
Metastatic Breast Cancer - molecular aberrations

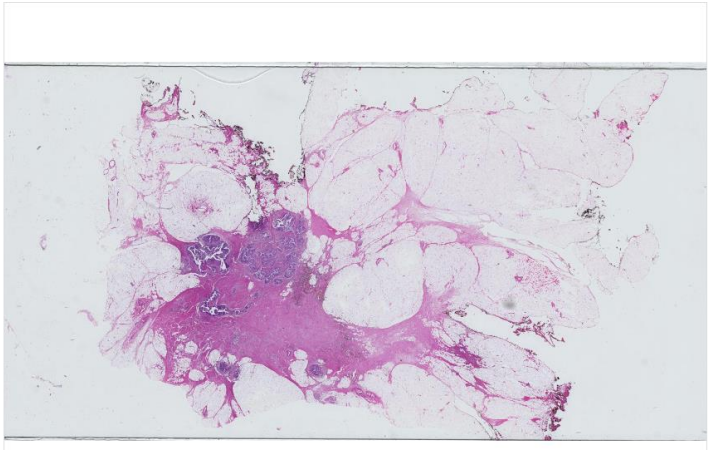


Patient 1345 - pathology images

⚠ A word of caution on internal controls...

- Primary H&E
- Primary ER
- Primary PgR
- Primary HER2
- Primary Ki67

No images found



Aftimos P et al. Manuscript submitted.

INSTITUT JULES BORDET INSTITUUT

ULB iris

INSTITUT
JULES BORDET
INSTITUUT

Applications in the diagnosis and treatment of cancer

ULB iris

Microsoft researchers detect lung-cancer risks in web search logs

Eric Horvitz, technical fellow and managing director of Microsoft's research lab in Redmond, Washington, says search queries may be an early warning of lung cancer. (Photography by Scott Eklund/Red Box Pictures)

New Online Views 1,743 | Citations 0 | Altmetric 83

Brief Report ONLINE FIRST

November 10, 2016

Evaluation of the Feasibility of Screening Patients for Early Signs of Lung Carcinoma in Web Search Logs

Ryen W. White, PhD[1]; Eric Horvitz, MD, PhD[1]

» Author Affiliations

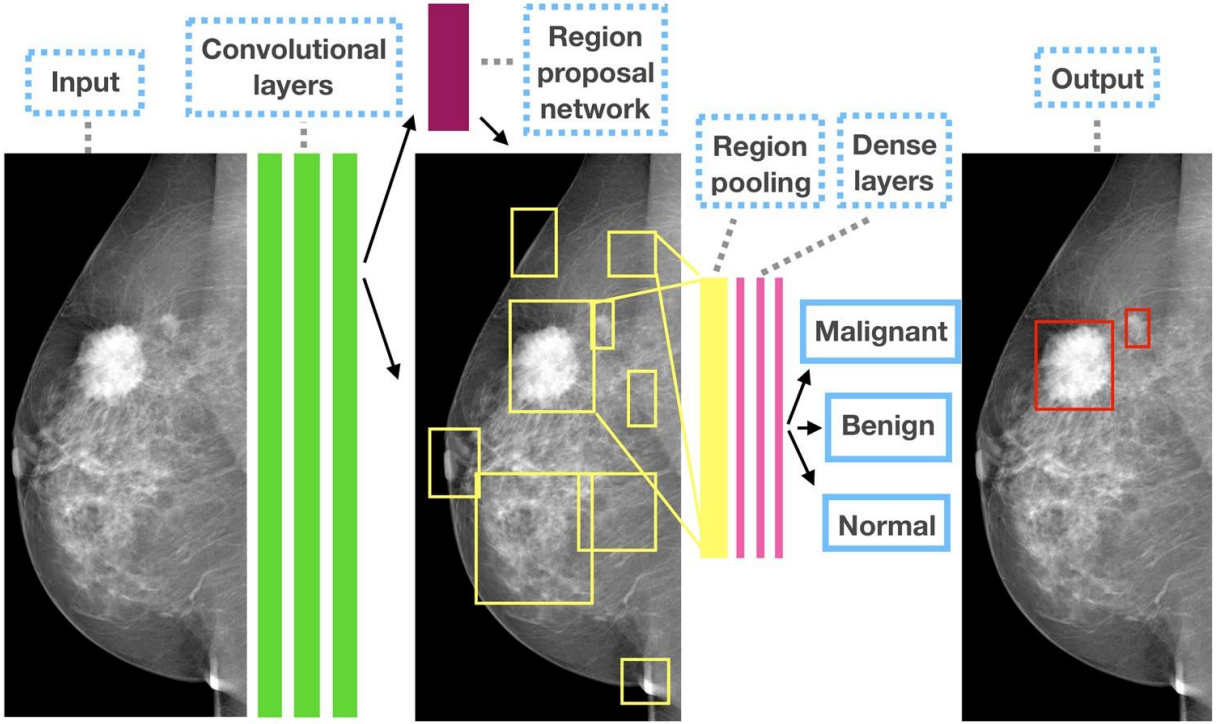
JAMA Oncol. Published online November 10, 2016. doi:10.1001/jamaoncol.2016.4911

JAMA Oncology

INSTITUT JULES BORDET INSTITUUT

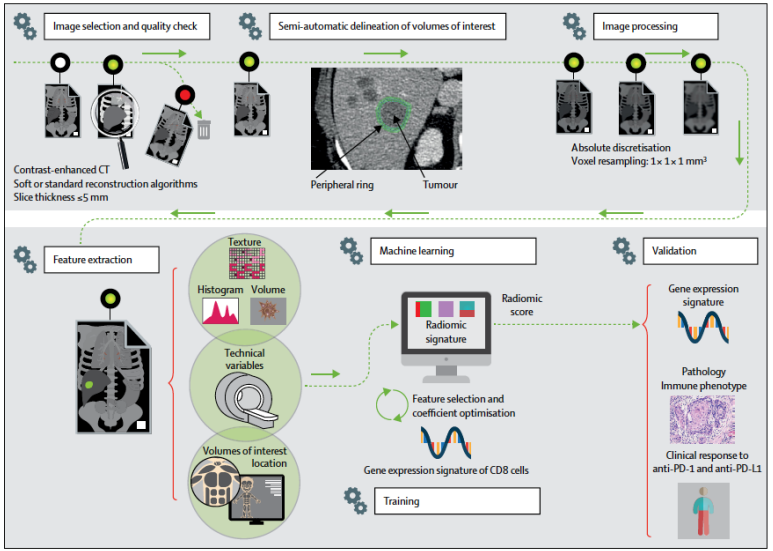
ULB iris

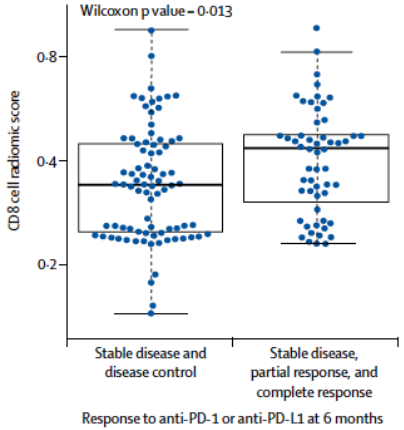
Detecting and classifying breast lesions using a deep convolutional neural network

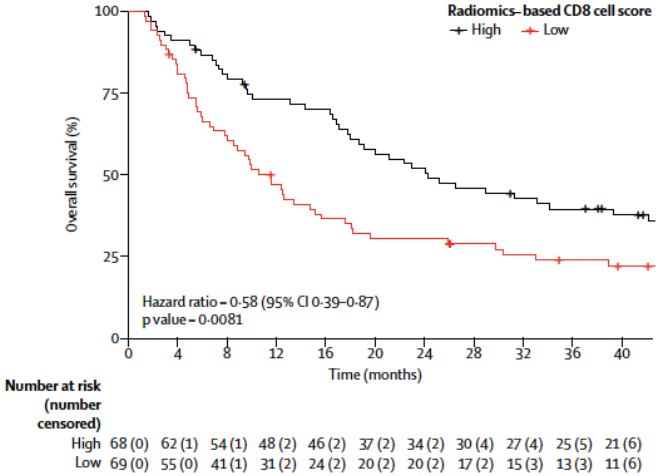


Ribli D et al. Sci Rep 2018.

A radiomics signature predicting CD8+ TILs and response to CPIs







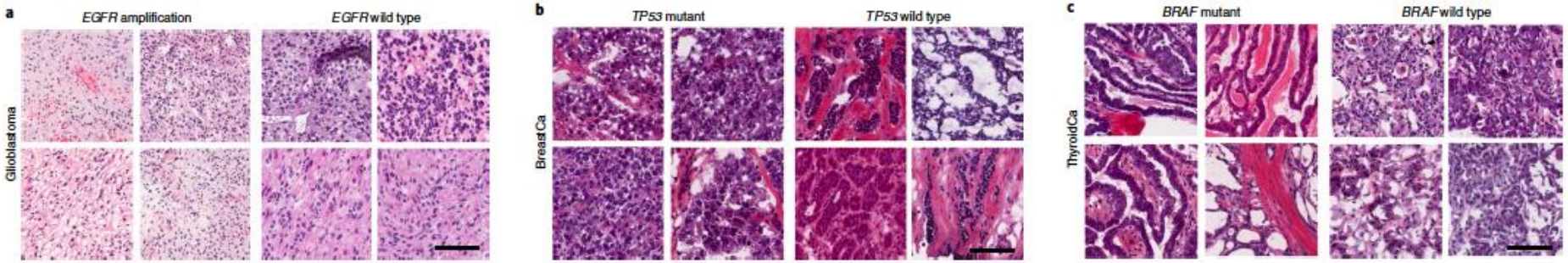
INSTITUT JULES BORDET INSTITUUT

Sun R et al. Lancet Oncol 2018.

ULB iris

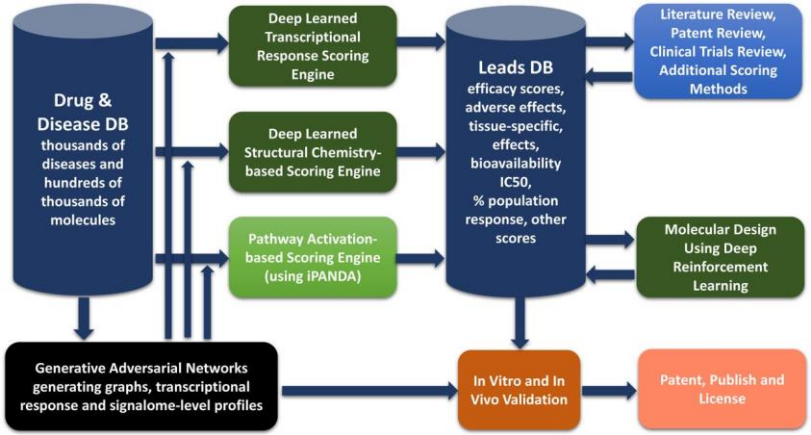
Histopathology reveals driver mutations

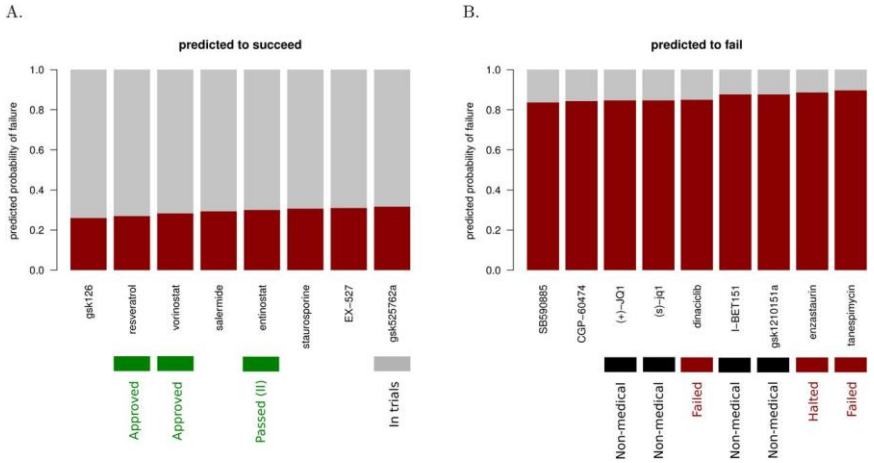
(but also tumor composition, gene expression and prognosis)



Fu Y et al. Nature Cancer 2020.

A deep learning tool to predict outcome of clinical trials





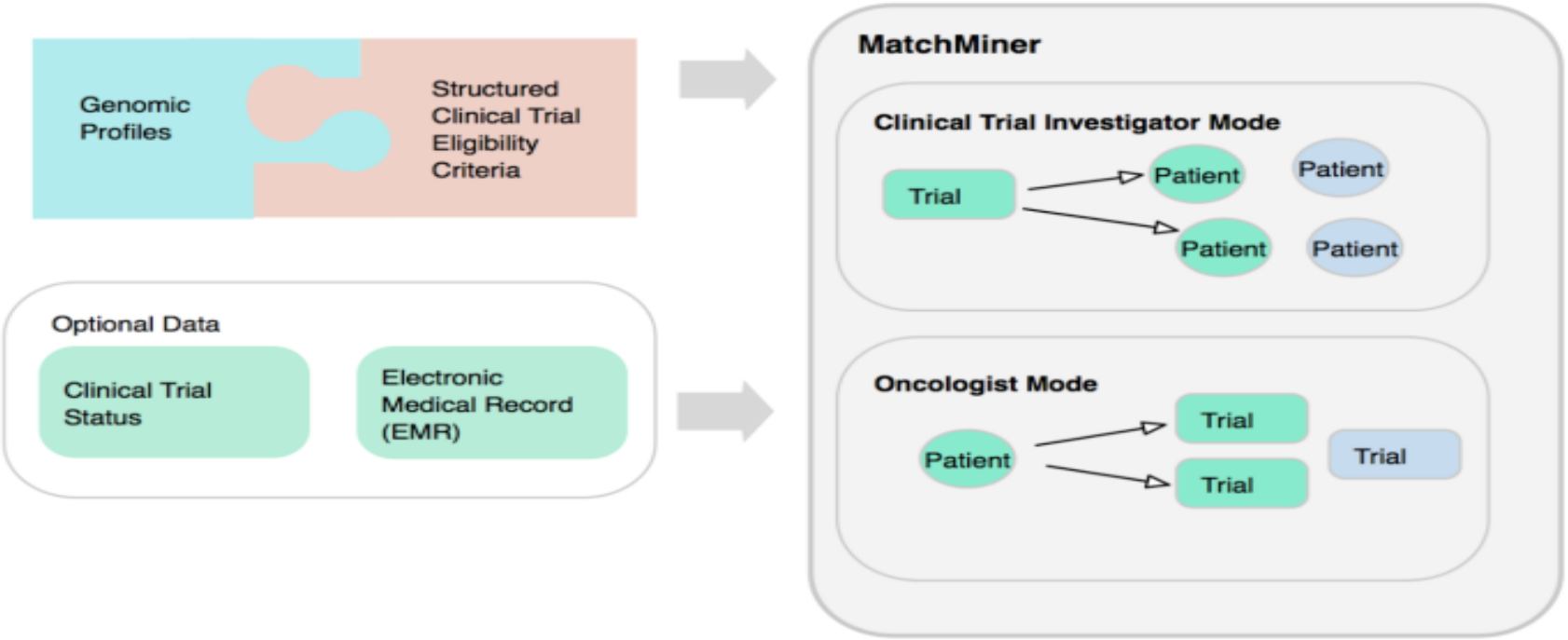
INSTITUT JULES BORDET INSTITUUT

Artemov AV et al. December 29, 2016. https://www.biorxiv.org/content/10.1101/095653v2.

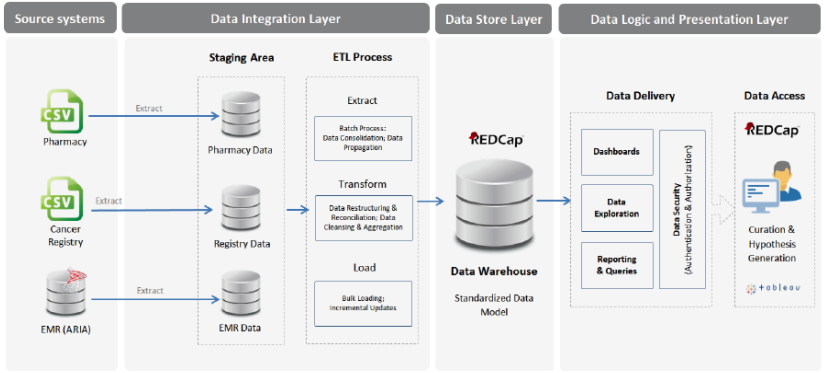
ULB iris

MatchMiner

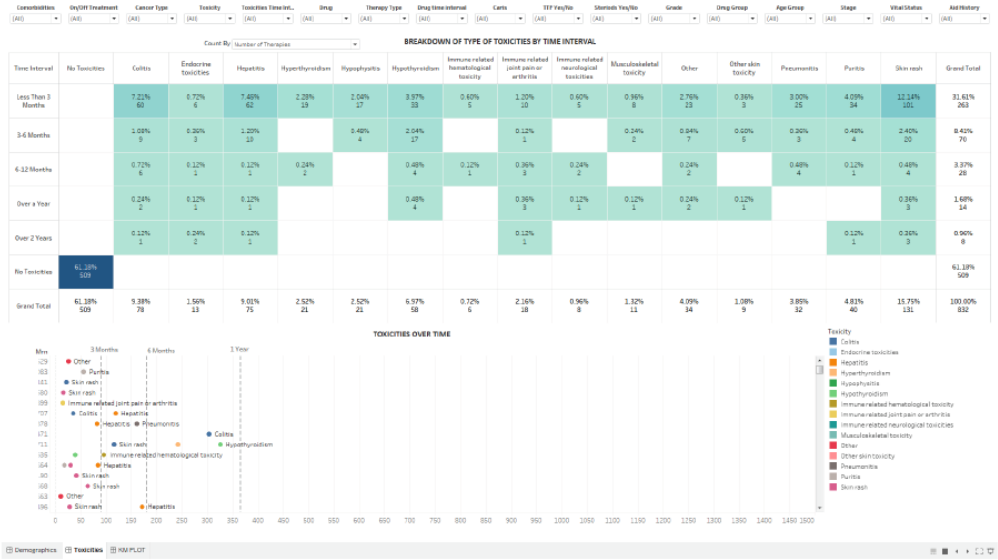
Developed at Dana Farber Cancer Institute and now Open Source



Georgetown Immuno-Oncology Registry



irAE Dashboard

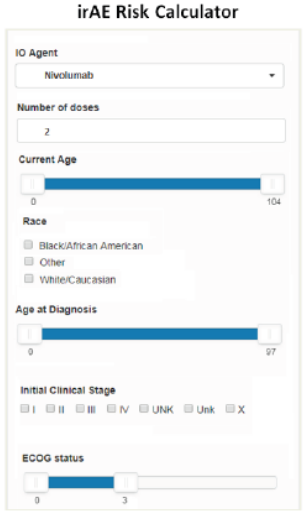


Courtesy Subha Madhavan

Predicting irAE

Can we predict irAE using a machine learning

Can we build a risk calculator for irAE?



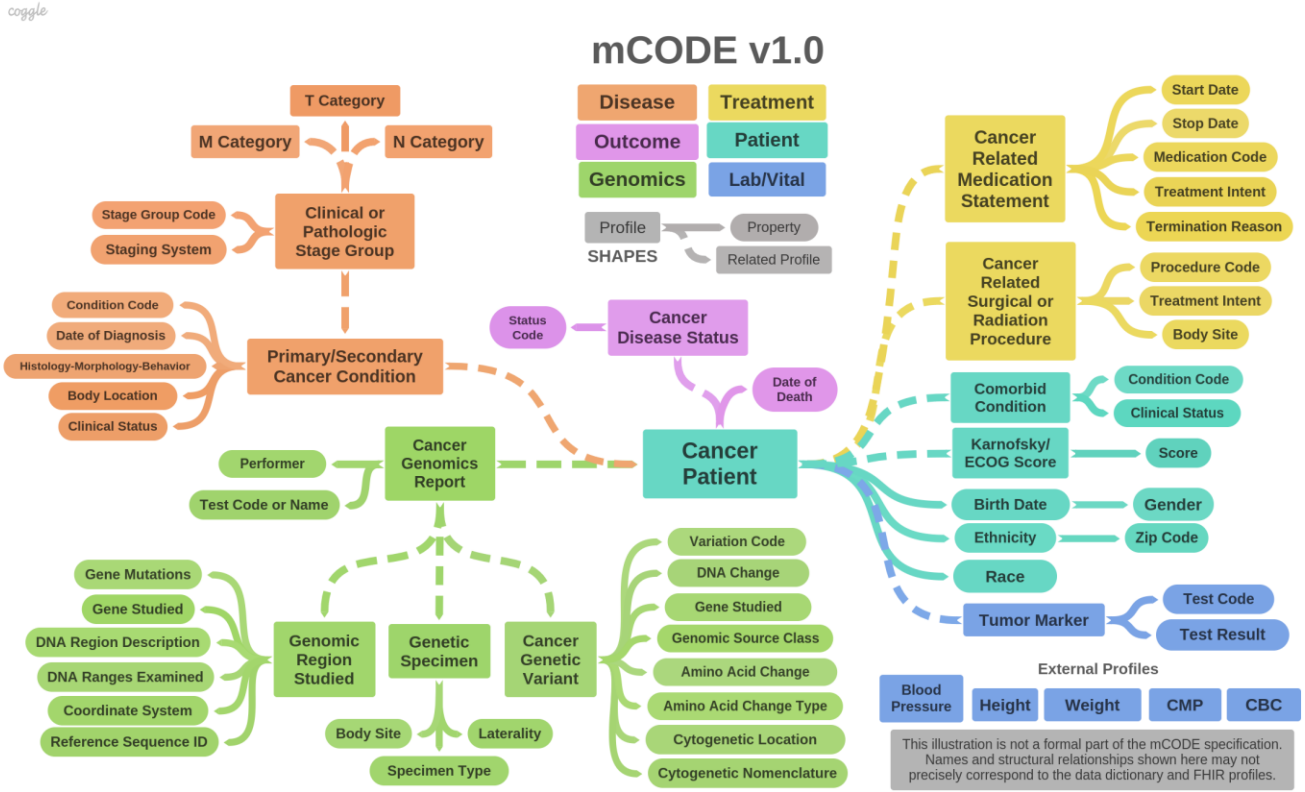
Courtesy Subha Madhavan

Limitations

- Proving generalizability and real-world applications
- Data access and equity
- Interpretability and the black box problem
- Education and expertise

mCODE™: Minimal Common Oncology Data Elements

The Initiative to Create a Core Cancer Model and Foundational EHR Data Elements



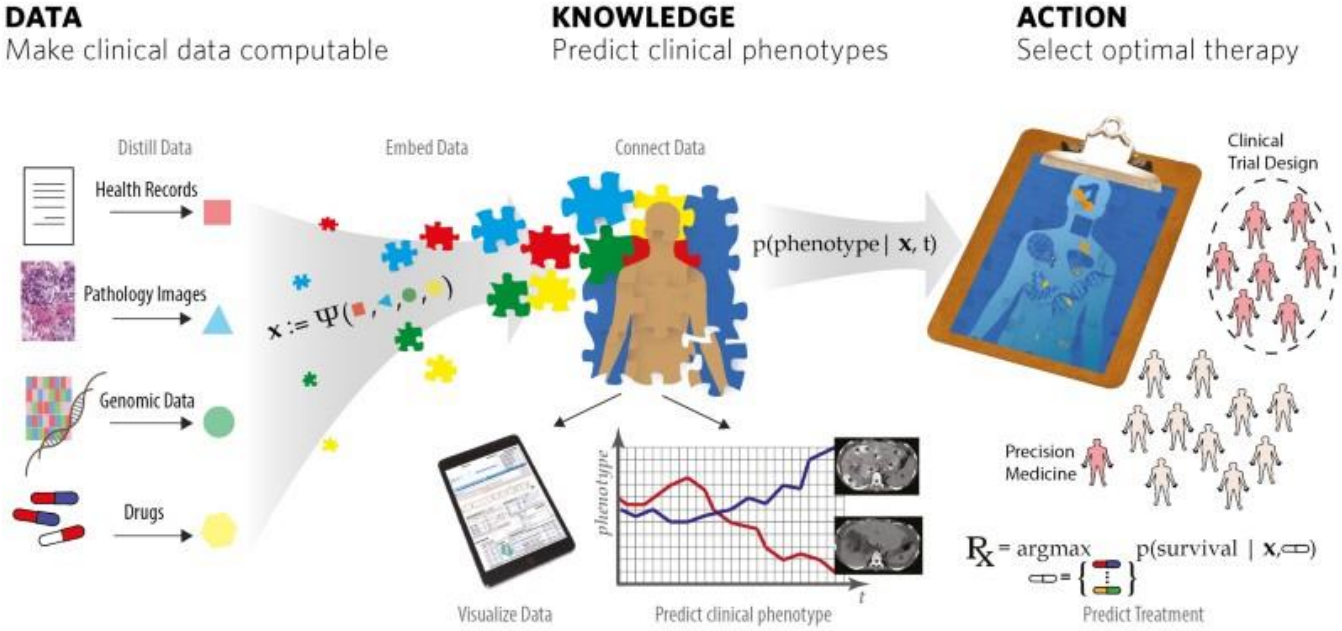
INSTITUT JULES BORDET INSTITUUT

ULB iris

Annals of Oncology 2020. Article in press.

INSTITUT
JULES BORDET
INSTITUUT

ULB
iris



Credit: Andre Kahles, Gunnar Rätsch, Chris Sander

The Institut Jules Bordet Clinical Trials Conduct Unit (CTCU)

Let’s interact…

INSTITUT JULES BORDET INSTITUUT

twitter @aftimosp

ULB iris