Designing an ontology for Knowledge representation of molecular biomarkers

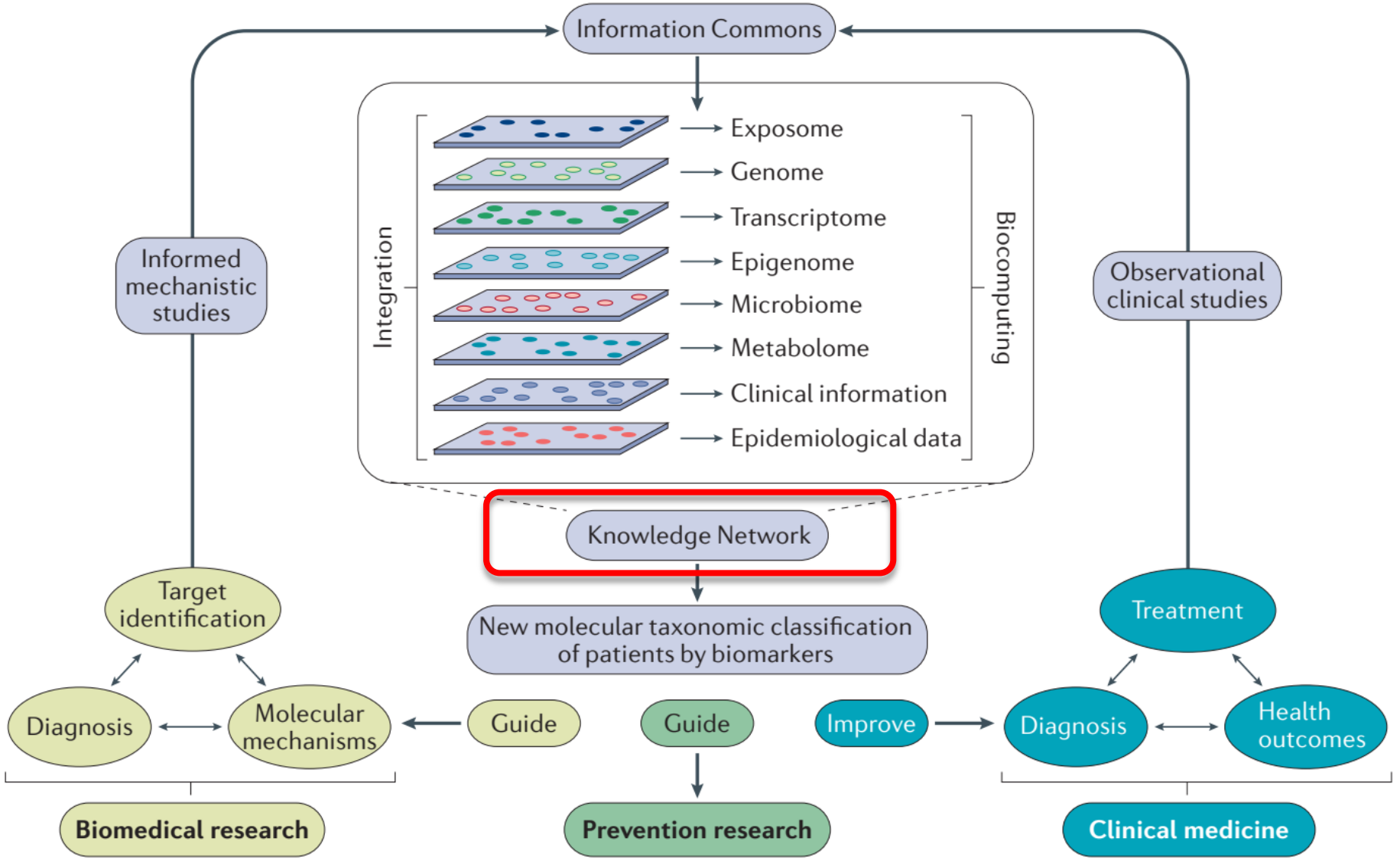
邵晨
SHAO CHEN
国家蛋白质科学中心 (北京)
National Center for Protein Sciences · Beijing
北京蛋白质组研究中心
军事科学院军事医学研究院生命组学研究所

国家蛋白质科学中心（北京）

国家蛋白质科学中心（北京），又称凤凰中心，是国家发展改革委员会批准立项的国家重大科技基础设施之一。凤凰中心的生物信息平台致力于为生命科学、特别是生命组学研究提供一站式的大数据服务。包括：

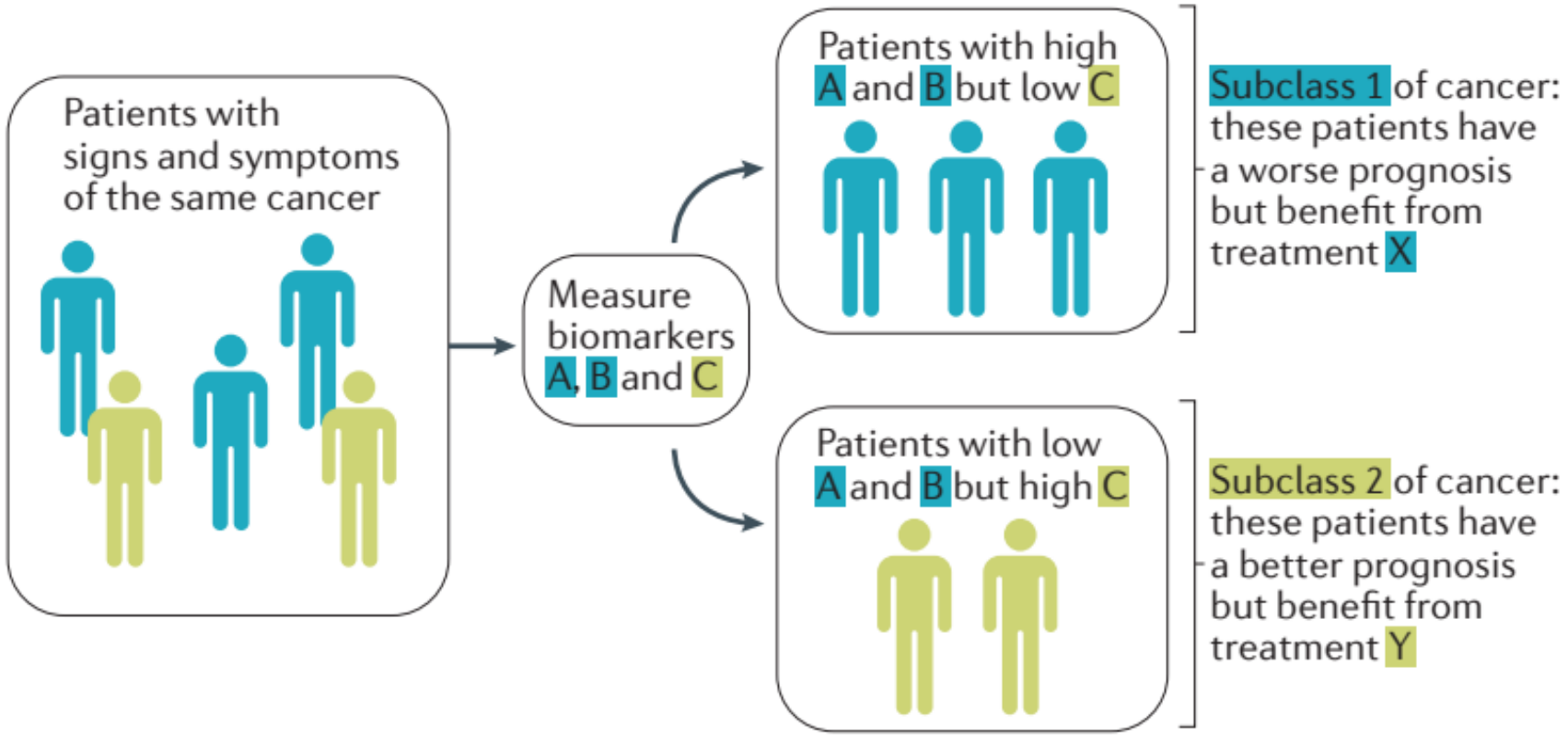
- 数据获取与处理
- 数据标准化、审编与注释
- 数据管理与整合
- 数据的分析、挖掘与展示

Biomarker plays a core role in precision medicine



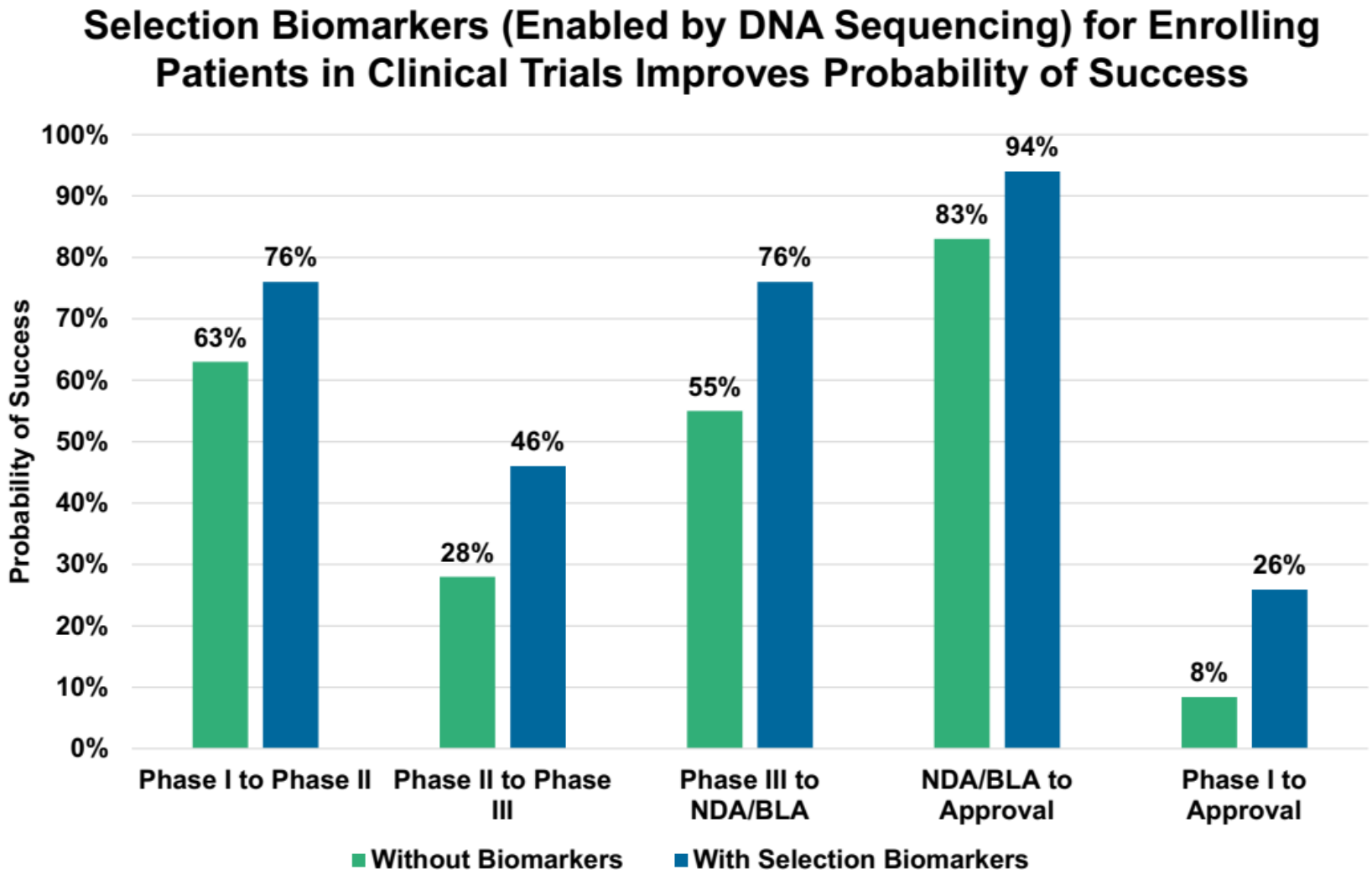
Nat Rev Cancer. 2016 Aug;16(8):525-37

Patients with the same signs and symptoms can be classified into new, specific taxa by molecular biomarkers



Nat Rev Cancer. 2016 Aug;16(8):525-37

…New Data Streams = Enhancing & Perhaps Accelerating Clinical Trials…

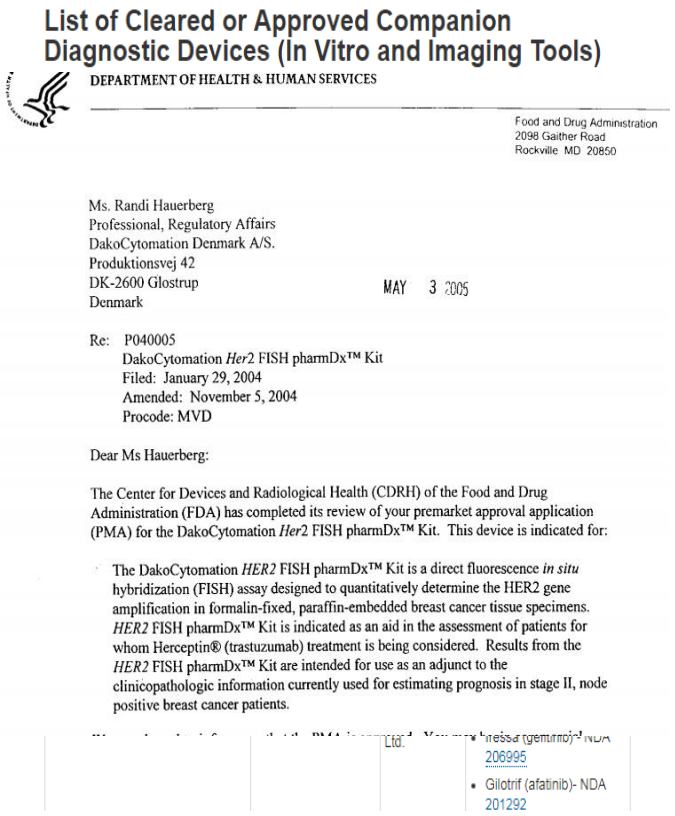


KLEINER PERKINS

Source: Biotechnology Innovation Group, Biomedtracker, Amplion (5/16)
Note: Based on 9,985 phase transitions of trials between 2006 – 2015. 512 phase transitions incorporated selection biomarkers for patient stratification; phase transitions identified by mapping NCT numbers from ClinicalTrials.gov with Amplion's BiomarkerBase and Biomedtracker's transition database.

Non-commercial resources for biomarkers

FDA databases & documents



List of Cleared or Approved Companion Diagnostic Devices (In Vitro and Imaging Tools)

DEPARTMENT OF HEALTH & HUMAN SERVICES

Food and Drug Administration
2098 Gaither Road
Rockville MD 20850

Ms. Randi Hauerberg
Professional, Regulatory Affairs
DakoCytomation Denmark A/S.
Produktionsvej 42
DK-2600 Glostrup
Denmark

MAY 3 2005

Re: P040005
DakoCytomation *Her2* FISH pharmDx™ Kit
Filed: January 29, 2004
Amended: November 5, 2004
Procode: MVD

Dear Ms Hauerberg:

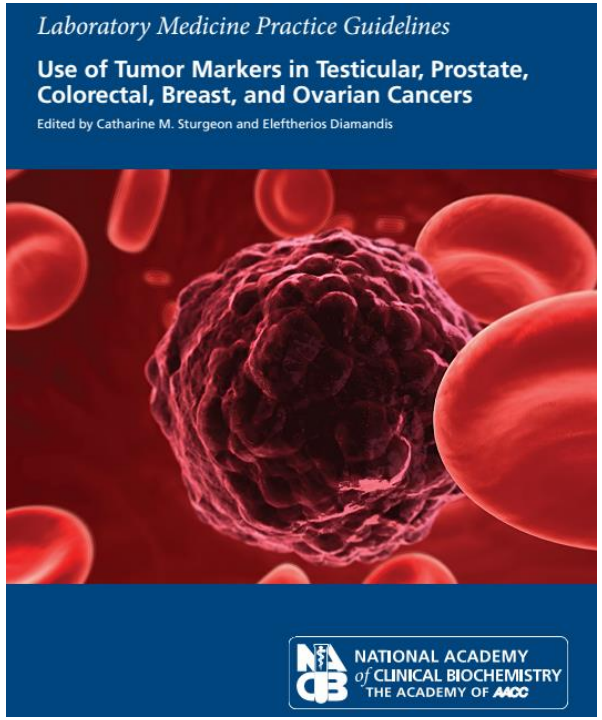
The Center for Devices and Radiological Health (CDRH) of the Food and Drug Administration (FDA) has completed its review of your premarket approval application (PMA) for the DakoCytomation *Her2* FISH pharmDx™ Kit. This device is indicated for:

- The DakoCytomation *HER2* FISH pharmDx™ Kit is a direct fluorescence *in situ* hybridization (FISH) assay designed to quantitatively determine the HER2 gene amplification in formalin-fixed, paraffin-embedded breast cancer tissue specimens. *HER2* FISH pharmDx™ Kit is indicated as an aid in the assessment of patients for whom Herceptin® (trastuzumab) treatment is being considered. Results from the *HER2* FISH pharmDx™ Kit are intended for use as an adjunct to the clinicopathologic information currently used for estimating prognosis in stage II, node positive breast cancer patients.

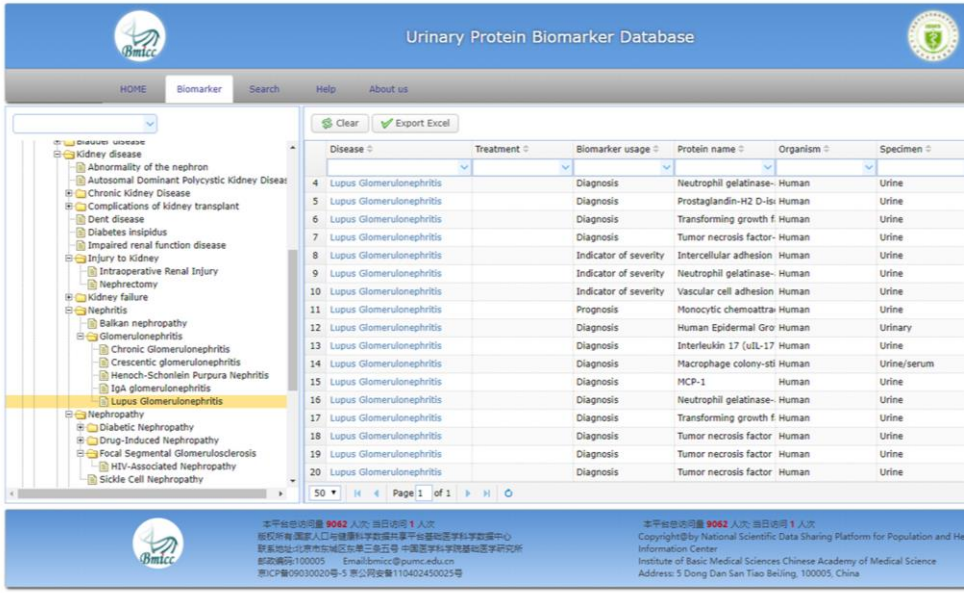
Ltd.

- Gilotrif (afatinib)- NDA 206995
- Gilotrif (afatinib)- NDA 201292

Guidelines



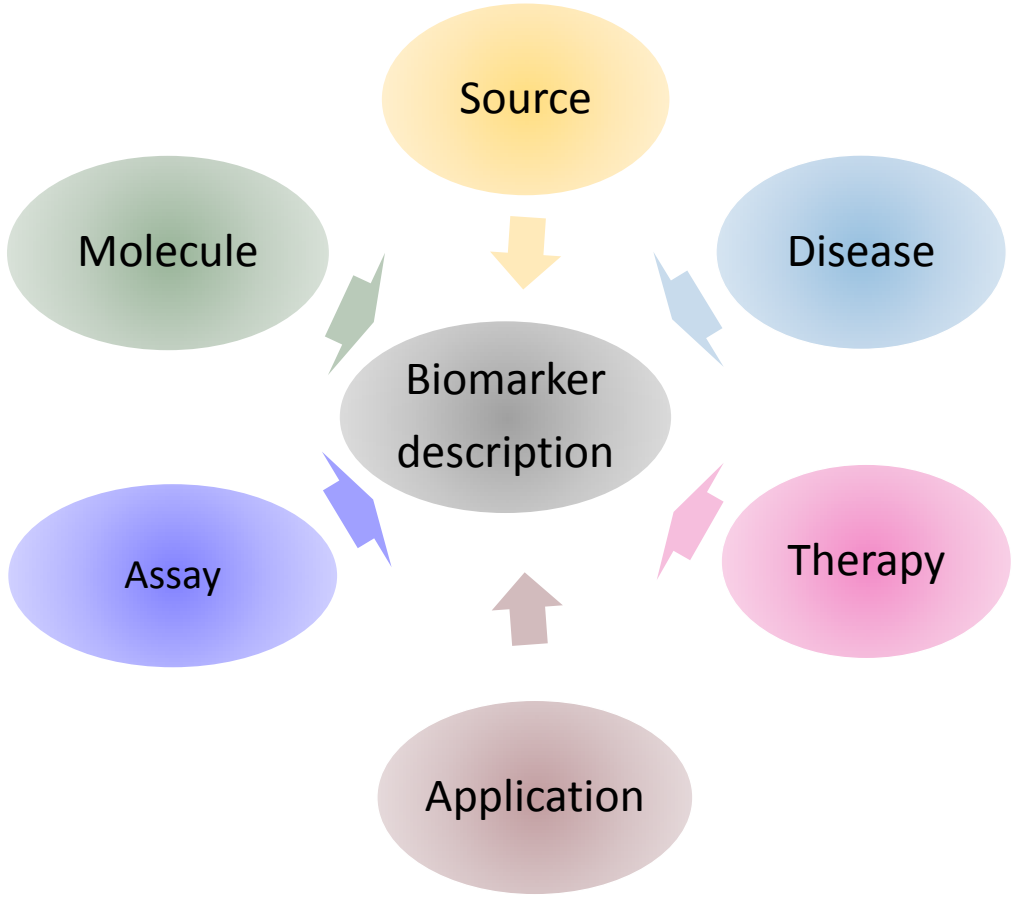
Literature and literature-curated databases



Approval statement of a biomarker test in FDA

Approval for the DAKO Herceptest.

This device is a **semi-quantitative immunohistochemical assay** to determine **HER2 overexpression** in **breast cancer tissues routinely processed for histological evaluation**. HercepTest is indicated as an aid in the **assessment of patients for whom HERCEPTIN(R)(Transtuzumab) treatment is being considered.**



Principles for biomarker knowledgebase construction

- The FAIR data principles
  Findable, Accessible, Interoperable & Reusable

- Common understanding of biomarker between clinicians, biologists, data scientists, and even the computational programs
  - Explicit definition of biomarker-related concepts
  - Relationships between concepts
  - Use machine-interpretable language

- Facilitating reuse of domain knowledge & data

Biomarkers in existing ontologies

Filters

Term type

× Class

class 265

Ontologies

Filter by ontology

NCIT 99

EFO 66

CHEBI 51

ORDO 27

PR 10

HP 3

Search results for *biomarker*

Previous *Showing **1** to **10** of **265** results* Next

Biomarker NCIT:C16342

http://purl.obolibrary.org/obo/NCIT_C16342

A characteristic that can be objectively measured and serves as an indicator for normal biologic processes, pathogenic processes, state of health or disease, the risk for disease development and/or prognosis, or responsiveness to a particular therapeutic intervention.

Ontology: NCI Thesaurus OBO Edition NCIT

biomarker CHEBI:59163

http://purl.obolibrary.org/obo/CHEBI_59163

A substance used as an indicator of a biological state.

Ontology: Chemical Entities of Biological Interest CHEBI

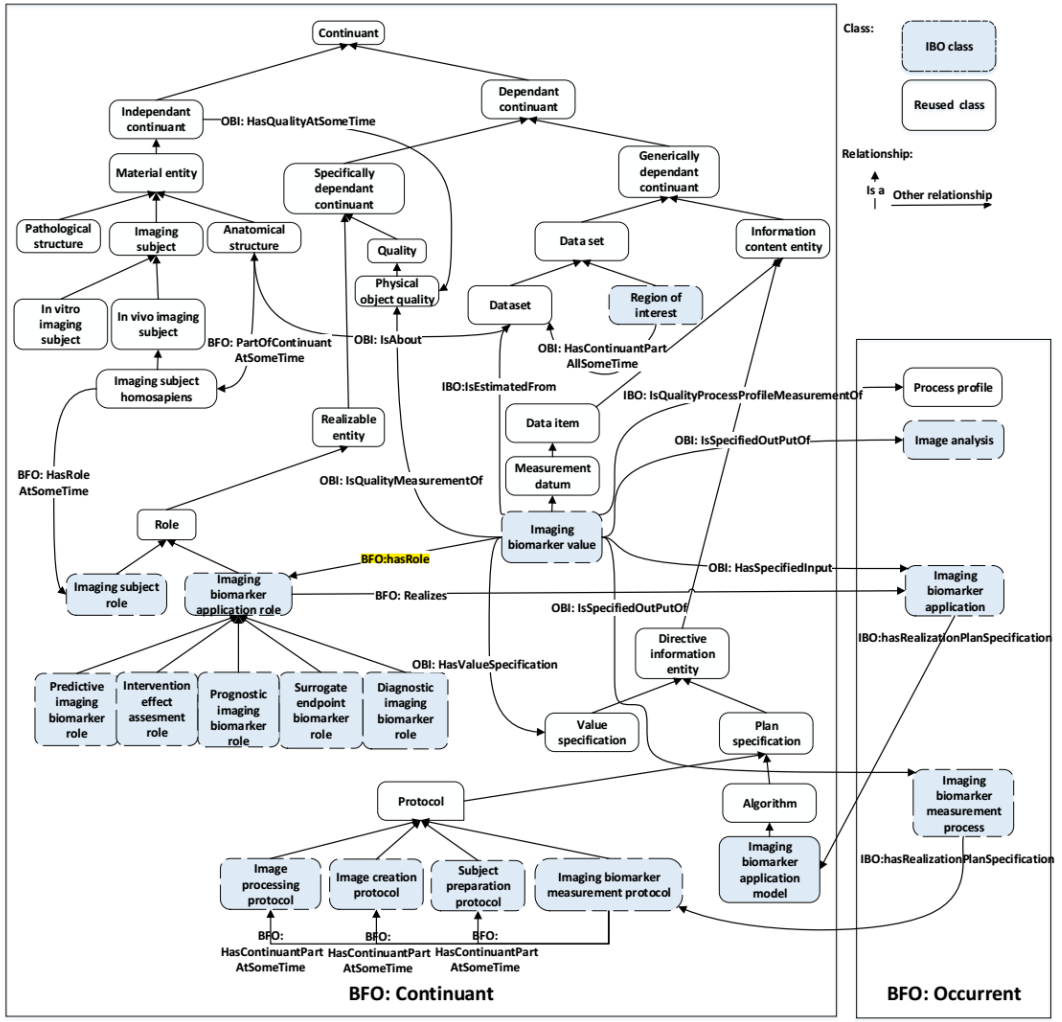
Also in: BAO NBO ENVO MICRO PECO ECOCORE TO ENM EFO

Biomarker Analysis NCIT:C63333

http://purl.obolibrary.org/obo/NCIT_C63333

A method of biological assay that looks for the presence of unique molecules or sequences considered indicative for a condition or state.

IBO: Imaging Biomarker ontology



Steps in MBO development

- Knowledge collection
  - Biomarker information
  - Terminologies that are widely accepted by the community
- Ontology design
  - Convert textual definitions into ontological definitions
    - Reuse classes and relationships in existing domain ontologies
    - Define new classes and relationships
  - Build up hierarchy & design pattern
    - Basic Formal Ontology (BFO) as top level ontology
- Formal representation of individual biomarker knowledge
  - Ontology guided curation
  - Assign MBO ID for each biomarker
- Validation

Unambiguous definition of “biomarker”

A defined characteristic that is measured as an indicator of normal biological processes, pathogenic processes, or responses to an exposure or intervention, including therapeutic interventions.

Widely accepted, but not ontologically correct

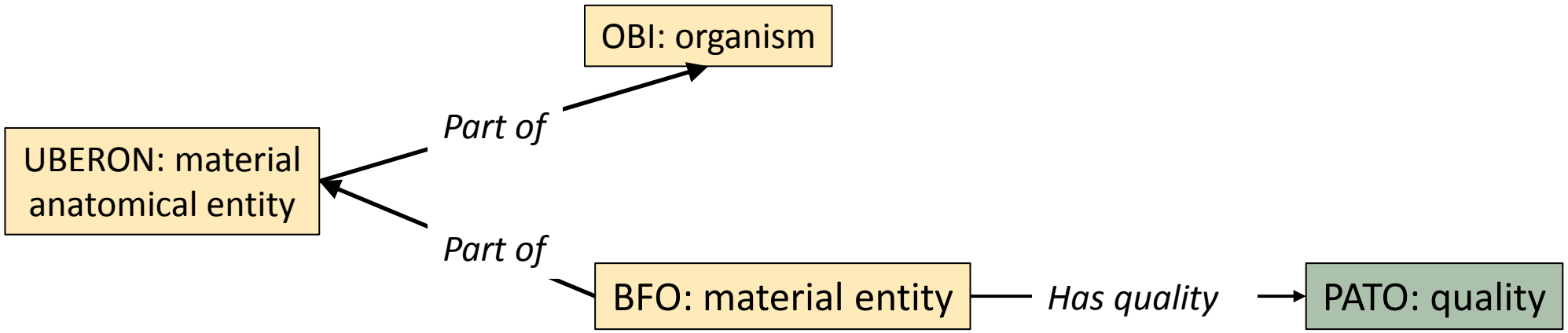
https://www.ncbi.nlm.nih.gov/books/NBK326791/

What is a “characteristic”?

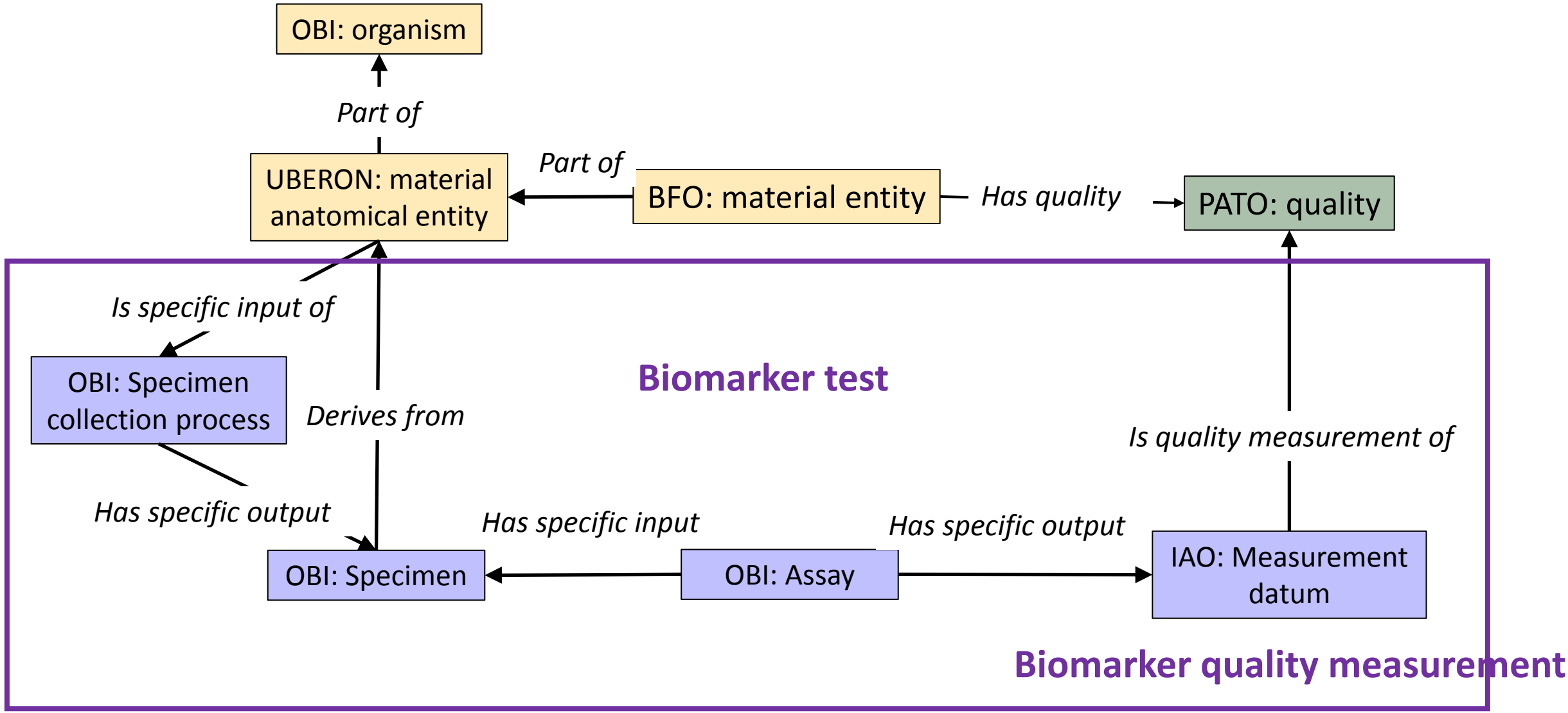
A characteristic = a characteristic of a material entity

For instances,

Blood glucose concentration, Presence of EGFR variants in tumor tissue



Define “measure”



Define “indicator of”

- In ChEBI, they define biomarker as a “role”

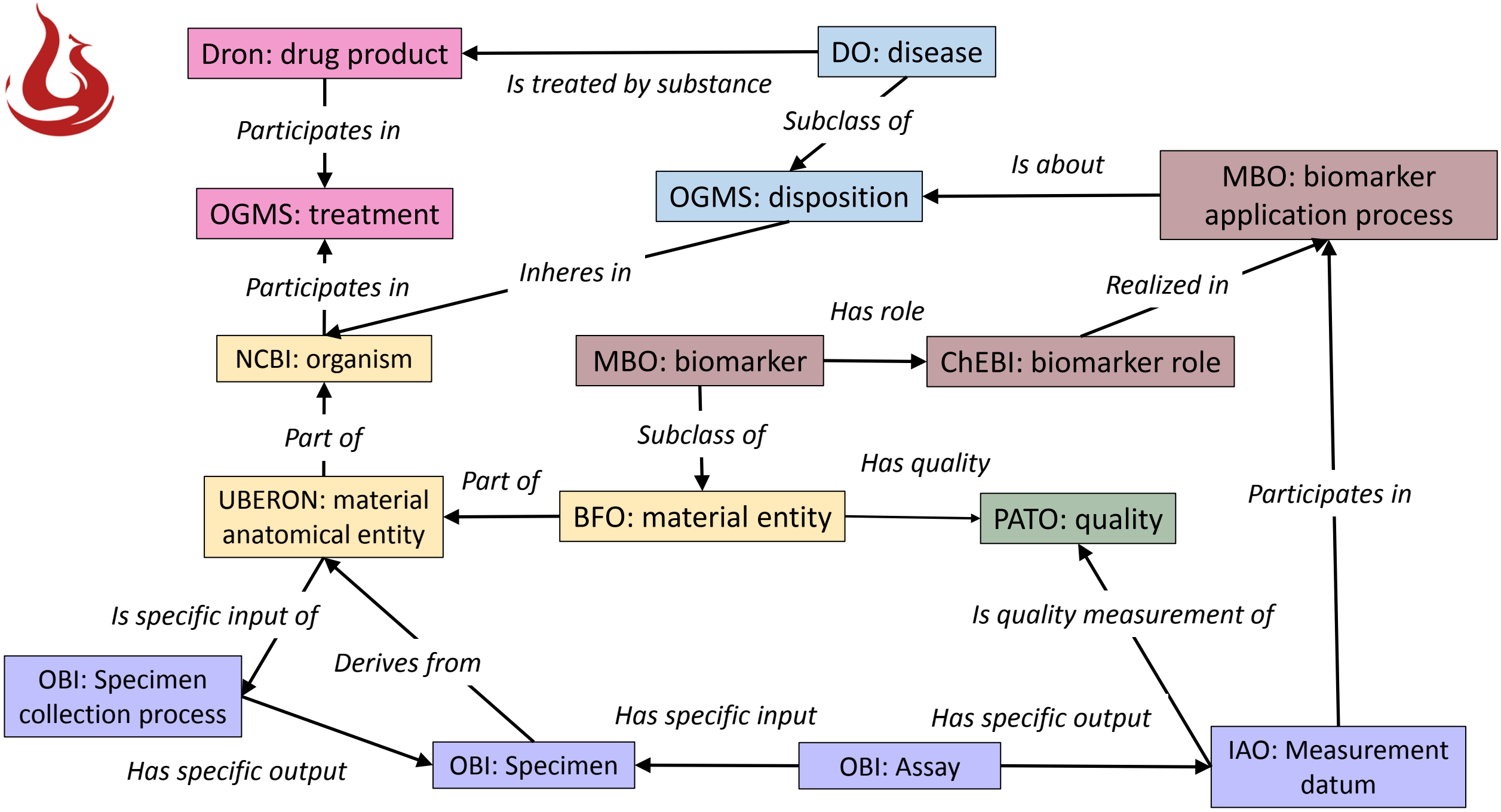
  def. A substance used as an indicator of a biological state.

- In MBO, we use two distinct classes, “biomarker” & “biomarker role”

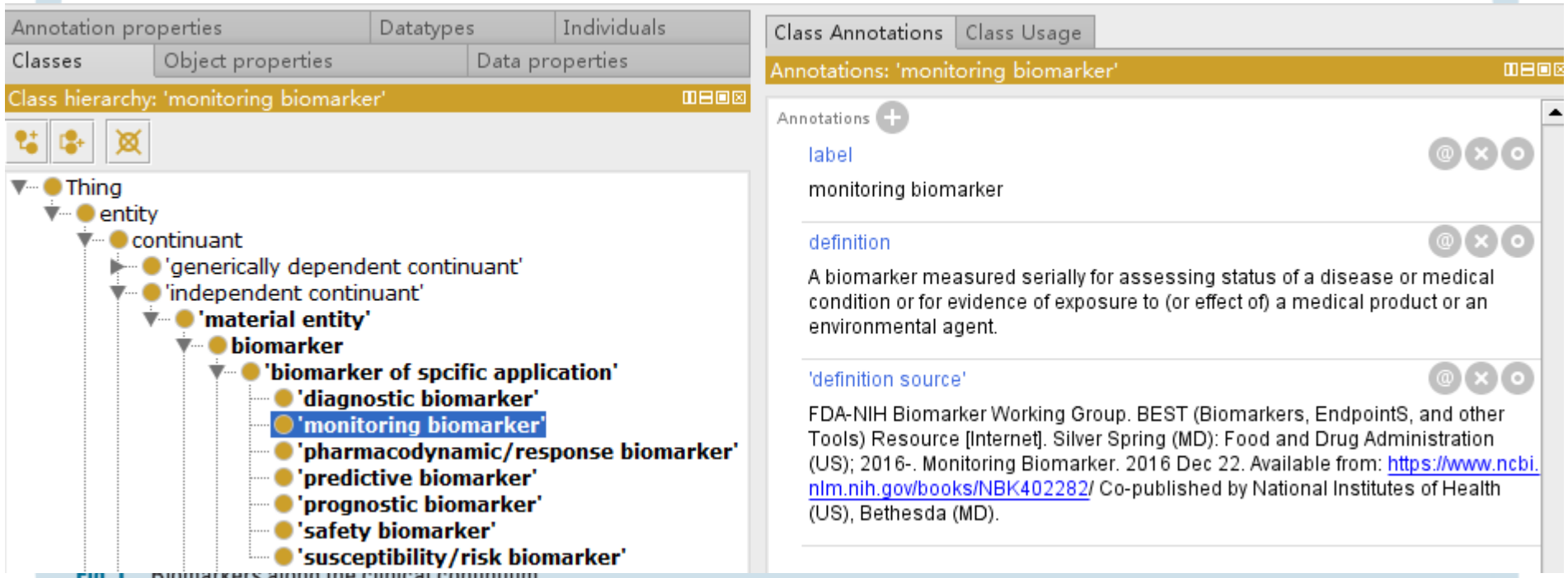
  biomarker = def. “material entity” and (“has role” some “biomarker role”)

A material entity can have different roles under different circumstances.

e.g. EGFR can have both “biomarker role” and “drug target role”.


Dron: drug product
DO: disease
Is treated by substance
Participates in
Subclass of
OGMS: treatment
OGMS: disposition
Is about
MBO: biomarker application process
Participates in
Inheres in
Realized in
Has role
NCBI: organism
MBO: biomarker
ChEBI: biomarker role
Part of
Subclass of
Part of
Has quality
UBERON: material anatomical entity
BFO: material entity
PATO: quality
Participates in
Is specific input of
Derives from
Is quality measurement of
OBI: Specimen collection process
Has specific input
Has specific output
Has specific output
OBI: Specimen
OBI: Assay
IAO: Measurement datum

Classify biomarkers by their applications



Material entity m "*is a*" "*diagnostic biomarker*"

m "*has role*" some "diagnostic biomarker role"

("measurement datum" "*is quality measurement of*" some ("*quality of*" some m)) "*participates in*" some "diagnostic process"

Neuro-Oncology 20(9), 1162–1172, 2017

It’s important to known which categories the biomarker belongs to

新普通体检男性（2847元）

预约此套餐

	序号	项目	内容
化验	1	血常规	全血细胞五分类
	2	肝功能	丙氨酸氨基转移酶ALT、总蛋白TP、白蛋白Alb、白蛋白球蛋白比A/G、总胆红素TBil、直接胆红素DBil、谷氨酰转肽酶GGT、碱性磷酸酶ALP、天门冬胺酸氨基转移酶AST、乳酸脱氢酶LD
	3	肾功能	尿素Urea、肌酐Cr、尿酸UA、葡萄糖GLU、钙Ca、磷P、钾K、钠Na、氯Cl、胱抑素C
	4	血脂	总胆固醇TC、甘油三脂TG、高密度脂蛋白HDL-C、低密度脂蛋白LDL-C、超敏C反应蛋白
	5	肿瘤筛查	甲胎蛋白AFP、癌胚抗原CEA、CA19-9、前列腺PSA、PSA-F、ProGRP、Cyfra211、SccAg、SPE
	6	糖化血红蛋白	糖化血红蛋白
	7	甲状腺功能	游离三碘甲状腺原氨酸FT3、游离甲状腺素FT4、三碘甲状腺原氨酸T3、甲状腺素T4、促甲状腺激素TSH3、甲状腺球蛋白抗体A-Tg、甲状腺过氧化物酶抗体A-TPO
	8	胃蛋白酶原	胃蛋白酶原PGI/II
	9	尿常规	尿10项
	10	便潜血	便OB

Screening biomarker, a subclass of diagnostic biomarker that diagnosis disease before a person has any symptoms.

Translate textual description into logic definition

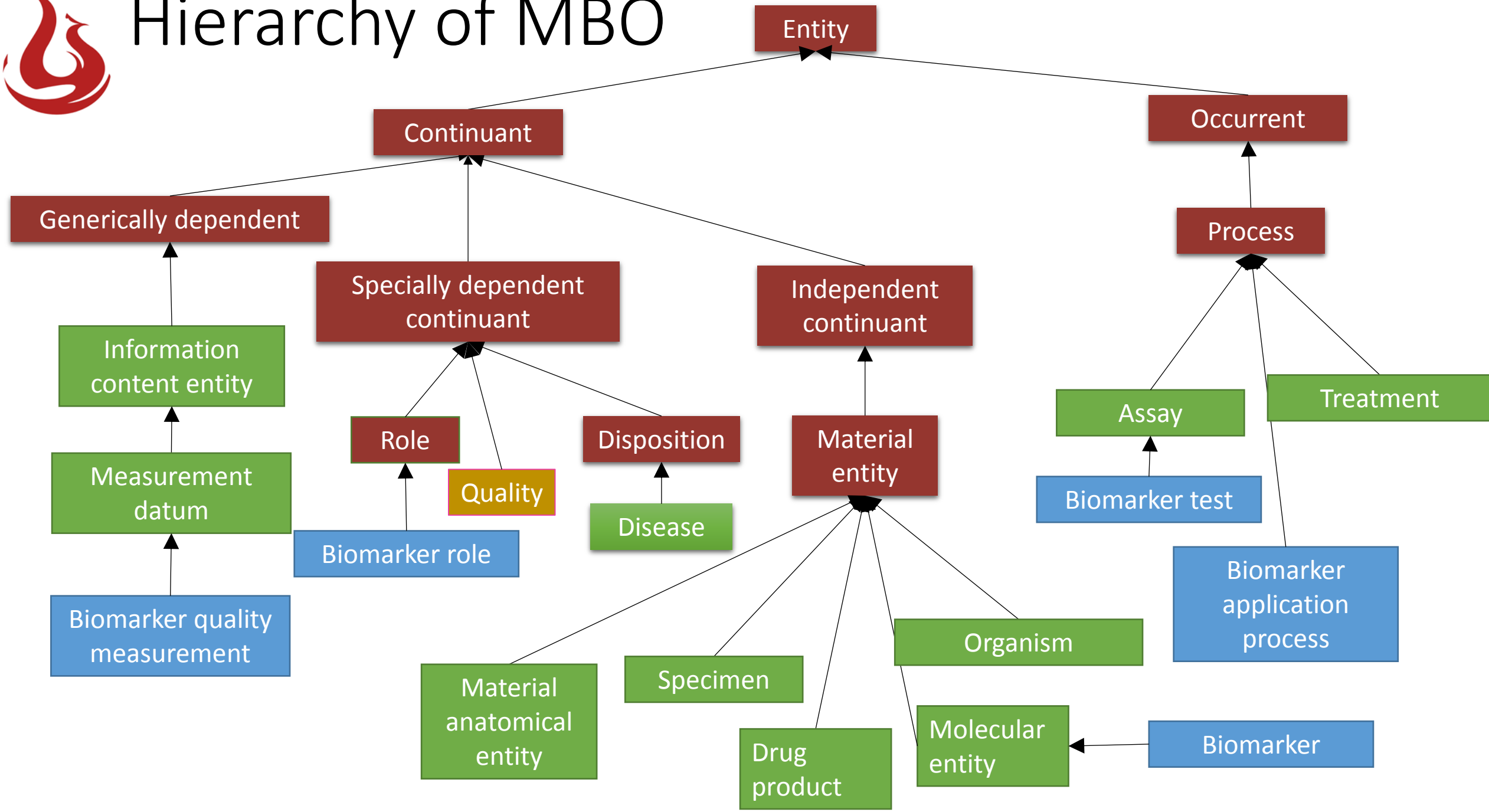
Approval for the DAKO Herceptest.

This device is a **semi-quantitative immunohistochemical assay** to determine **HER2 overexpression** in **breast cancer tissues routinely processed for histological evaluation**. HercepTest is indicated as an aid in the **assessment of patients for whom HERCEPTIN(R)(Transtuzumab) treatment is being considered.**

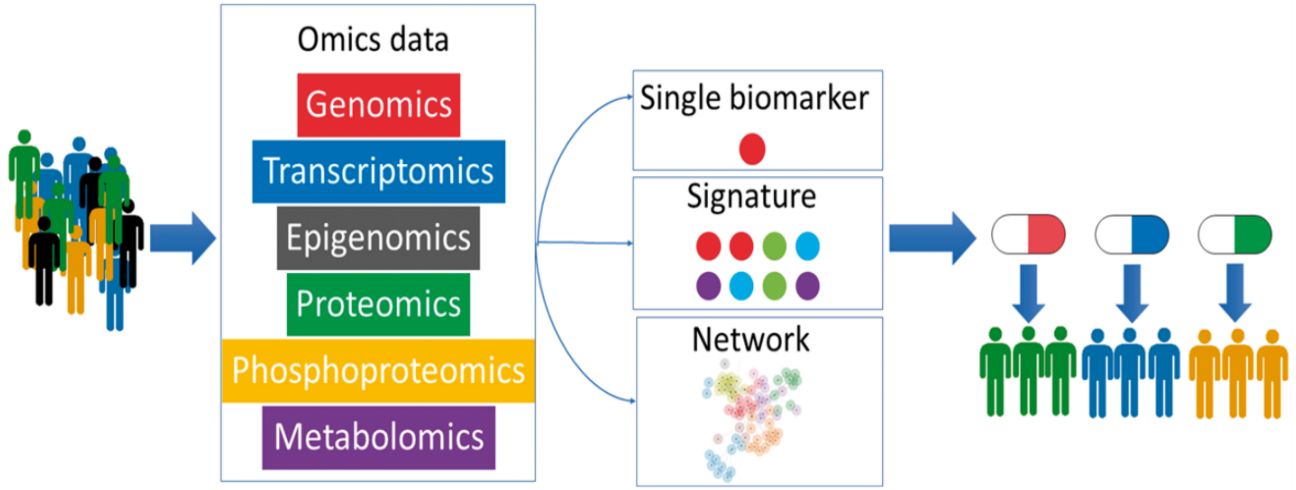
("HER2 gene" and ("*expressed in*" some ("breast caner cell" and "*part of*" some "breast cancer tissue" and "*part of*" some human))) and "*has role*" some ("predictive biomarker role" and "*realized in*" some ("biomarker application of predicting therapy effect" and "*is bout*" some ("breast cancer" and "*is treated by substance*" some Herceptin®)) and ("*has biomarker test*" some "DAKO Herceptest")

DAKO Herceptest = def. ("*subclass of*" some "FDA approved biomarker test") and ("*has specific applicant*" only "Dako A/S") and "*subclass of*" some ("semi-quantitative immunohistochemical assay " and ("*has specific input*" some "processed breast cancer tissue specimen") and ("*has specific output*" some ("expression level" and "*is quality measurement of*" some "HER2 gene")))

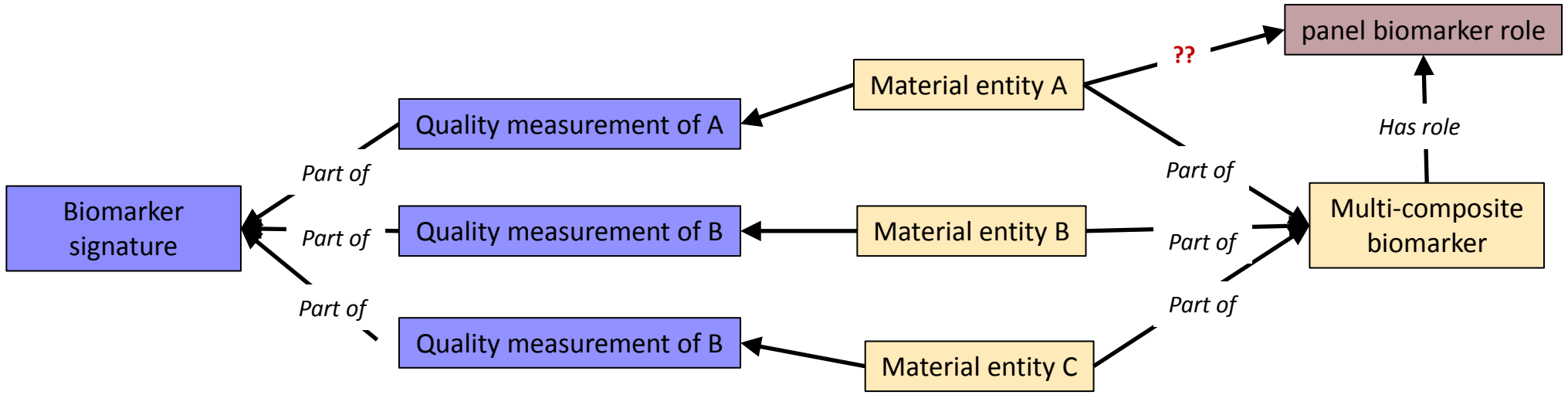
Hierarchy of MBO



MBO meets precision medicine: novel biomarker types



5-Protein signature (OVA1®)
21-Gene signature (Oncotype DX®)
70-Gene signature (Mammaprint®)



Network biomarker

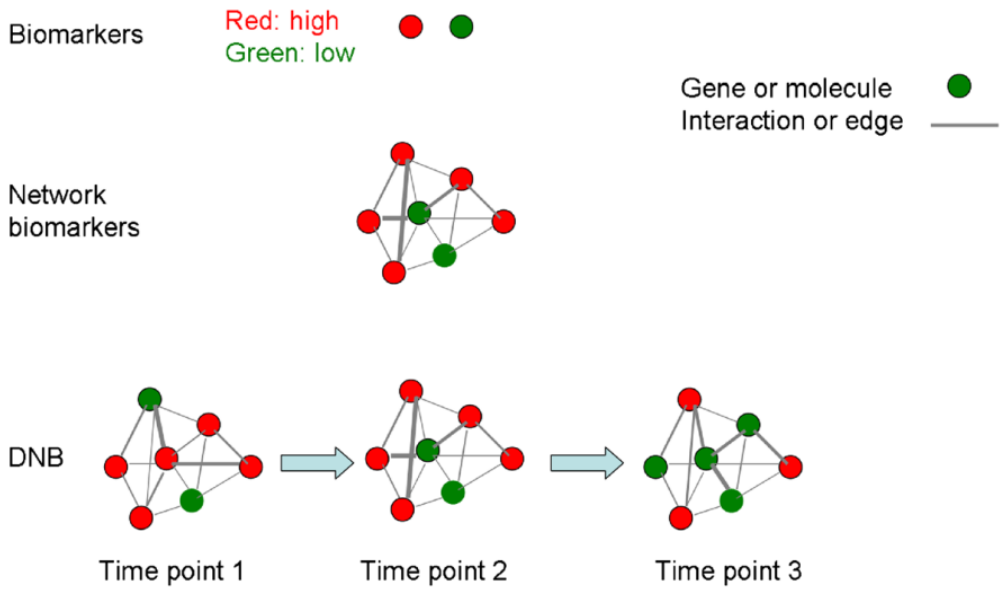
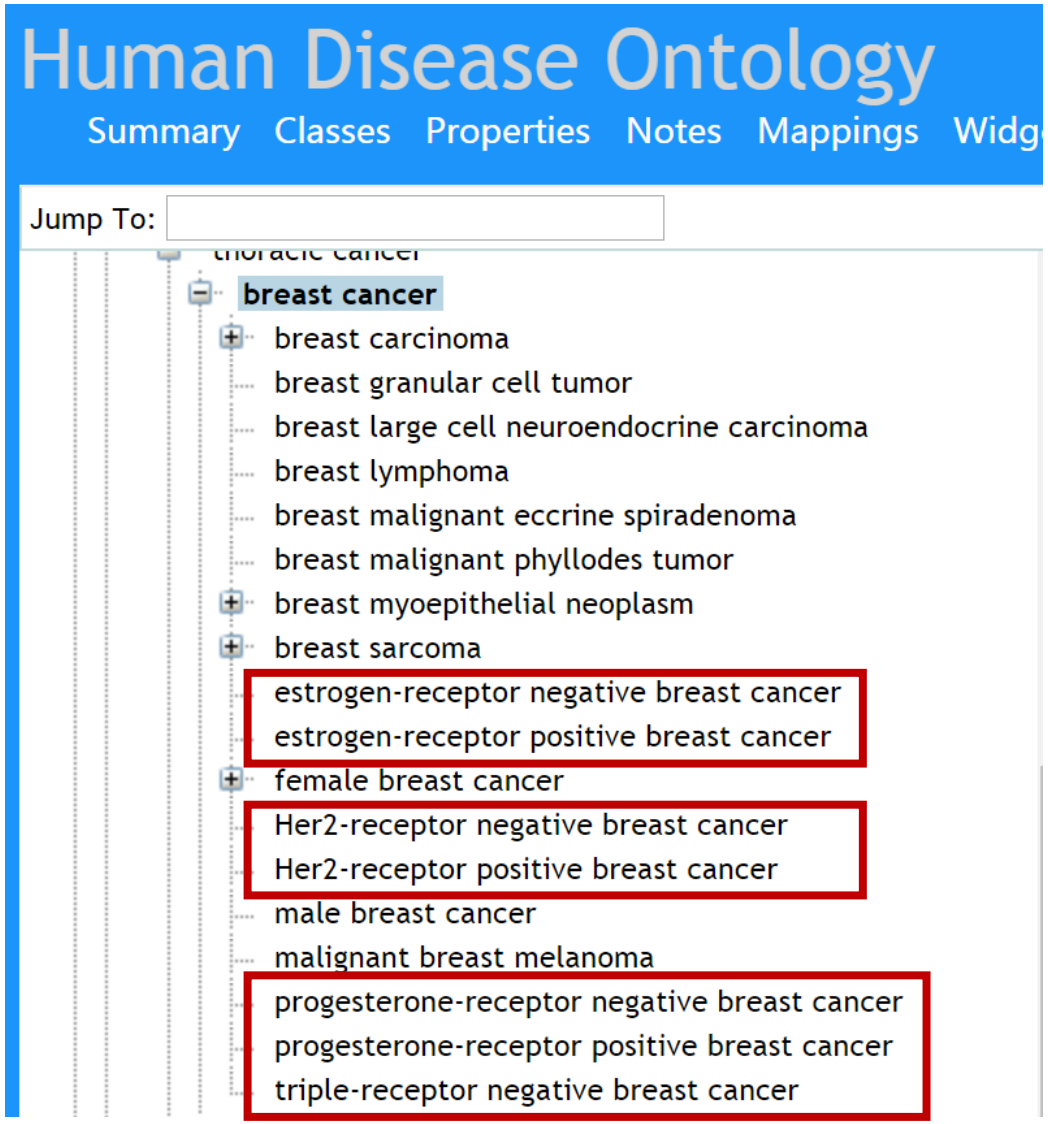


Figure 1 Biomarkers, network biomarkers and dynamical network biomarkers. Biomarkers provide one dimensional information, while network biomarkers provide two dimensional information by adding interactions. DNBs provide a three dimensional image of biomarker-biomarker interactions by showing time-dependent stronger or weaker interactions among biomarkers in the network. DNB: Dynamical network biomarkers.

"*has composite*" only ("*material entity*" and "*participates in*" only ("*subclass of*" some ("network interaction process (INO)" and "*has quality*" some ("dynamic network quality" and ("*has quality measurement*" some ("dynamic network quality measurement" and "*participates in*" some "biomarker application process")))))

Wu et al. Clinical and Translational Medicine 2014, 3:16

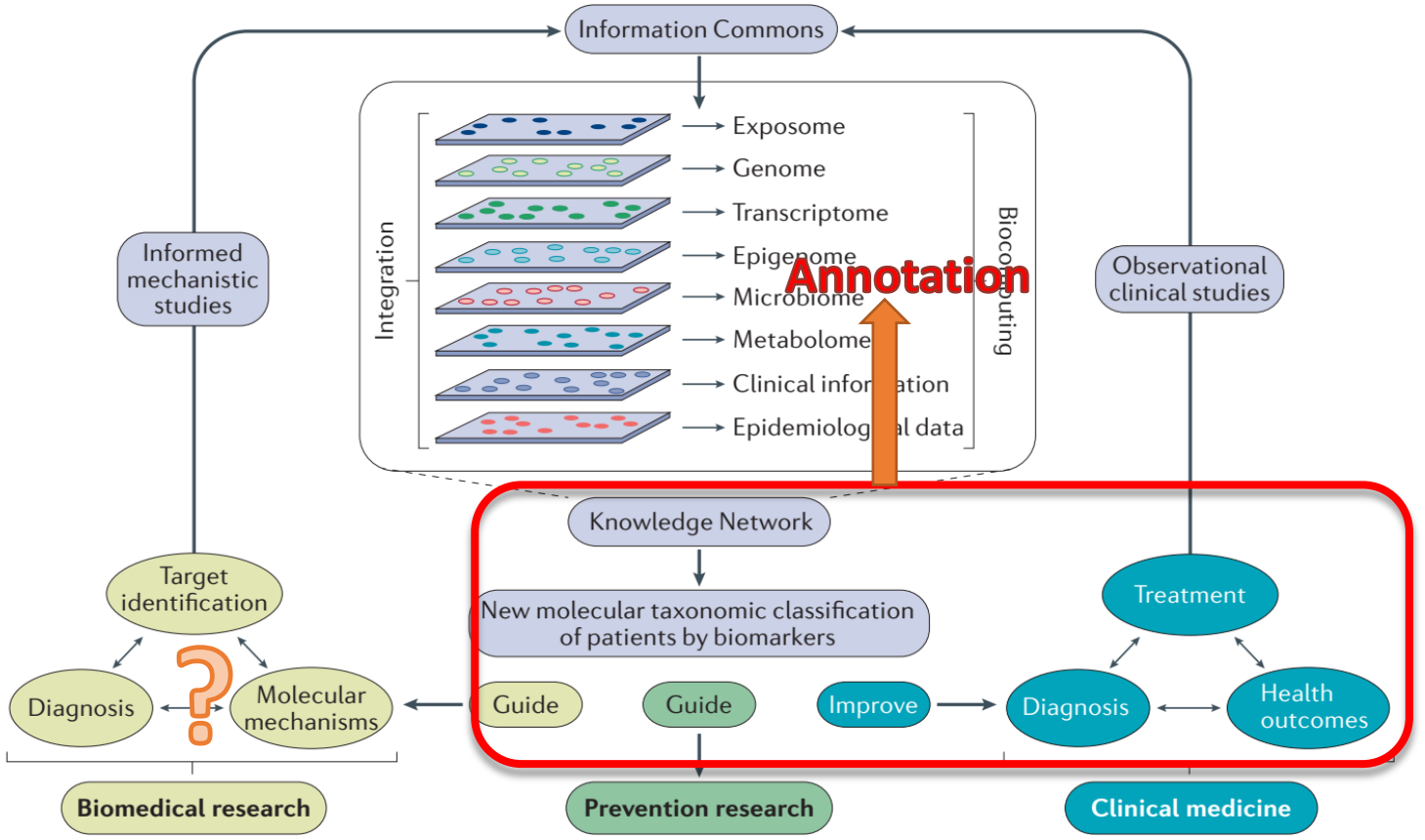
MBO meets precision medicine



Her2 is both “predictive biomarker” & “prognosis biomarker” for breast cancer

“Her2-receptor positive breast cancer” has Her2 as “diagnostic biomarker”.
“Her2-receptor positive breast cancer” has poor prognosis.
“Her2-receptor positive breast cancer” can be treated with anti-Her2 target therapy.

MBO meets precision medicine: interoperation with knowledge and ontologies from other domains



Summary

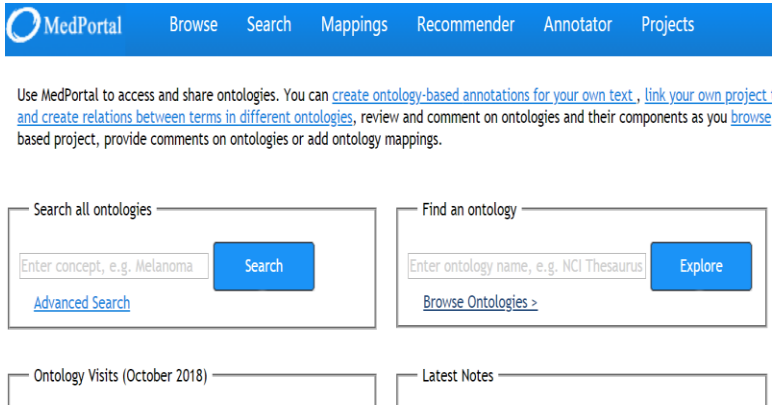
- Achievements
  - A semantic model for biomarker knowledge representation
  - Many terms in MBO can be reused in a universal description of biomarkers

- Ongoing work
  - Adding biomarker instances to MBO, making it a knowledgebase for omics data annotation
  - Refine MBO's design pattern

- Intended use of MBO
  - Database schema design, biocuration and data integration
  - Database query (using SPALQL language)
  - Text mining (with accurate and comprehensive relationship definitions)
  - Reasoning and inference new biomarker knowledge

Acknowledgements

Bmicc OntoChina



http://medportal.bmicc.cn/

http://ontoanimals.bmicc.cn/