

## CONCEPT SPACE

Note : This map presents a preliminary understanding of cancer and its treatment, data & data analyses technics as well as all the ideas related to the use of data for cancer understanding.

If you see any inconsistencies, you are aware of some interesting projects on cancer treatment or you have new ideas on how to go further, don't hesitate to contact us at [sk.chmapi@epidemium.cc](mailto:sk.chmapi@epidemium.cc)

We will complete the map thanks to your inputs!

## USING CANCER DATA

Cancer is a disease

may be fostered by numerous risk factors

that are interconnected in a complex way

related to environment (exposure to radiations UV, exposure to infectious agents, radon, down, asbestos, pesticides...)

related to the profile (age, sex)

related to genetic predispositions

related to lifestyle behaviors (diet/activities, tobacco...)

can be solid (carcinomas (90%) (skin, mucosa, glands) or sarcomas)

can be liquid (leukemias (8%) (blood, bone marrow) or lymphomas)

is characterized by the organ/tissue where it first develops

and consists of normal cells that have mutated

and thus gain new properties ("super powers") :

- Sustaining proliferative signaling
- Evading growth suppressors
- Resisting cell death
- Enabling replicative immortality
- Inducing angiogenesis
- Activating invasion and metastasis
- Reprogramming Energy metabolism
- Evading immune destruction

starting from a single cell that mutates and gains some of these 8 properties (monoclonal cancer)

starting from several cells that mutate on different ways and complement to gather these 8 properties (polyclonal cancer)

By having at least 5 mutations on a cell

Which have not been repaired by automatic control quality from

that first develops into a benign (cluster of cancer cells)

growing through cell division

amplified by the capture of growth signals intended to close

Coupled with non-physical regulation

first at a regular rate

a huge acceleration happens at one point (dysplasia)

when the membrane of epidermal cell is damaged

evolving into a malignant metastase (malignant transformation)

characterized by the TMM (Tumor Nodes Metastasis) process

Cancer is studied

to slow cancer propagation

to block cancer propagation

to "heal" cancer

to prevent cancer

to detect cancer at early stage

to eradicate cancer

by curative therapies

by preventive therapies

for palliative care

Cancer is detected

through screening tests performed by medical staff

through self-screening techniques

through screening techniques performed by third parties

when first symptoms appear

on a regular basis (organised screening)

with a further and more frequent follow-up for people at risk (risk factors, genome analysis, age, sex) (individual screening)

in continuous

That evolves according to the stage of illness

through invasive techniques

through non-invasive techniques

chosen according to the symptoms (absent, well recognizable, visual, non visual, weak, non-specific or vice versa)

the techniques sensibility and specificity

the type of cancer to detect (stage, localisation, features)

the practices of each country

That may have side effects at detection & treatment (over-diagnosis)

that are completely safe for the patients

that slow cancer propagation

with a low ability to anticipate further evolution of benign tumors

with a high ability to anticipate further evolution of the tumor at different stages

Cancer is treated

through invasive techniques

through non-invasive techniques

By targeting the tumor (surgery with tumor removal, radiotherapies). Or by targeting cells with cancer type cells characteristics (chemotherapies)

By strengthening naturally occurring defense systems :  
- biotherapies including immunotherapies  
- hormonal therapies

By acting on the psychological relationship between the patient and his illness (Education, psychological well-being and stress reduction, meditation)

By adopting a lifestyle that undermines physical risk factors (nutrition, physical health, geographical zone...)

By studying cancer mechanisms through alternative medicines (herbs, Chinese medicine, homeopathy, phytotherapy)

By adapting the treatment according to care path

chosen according to the SQR ("Standard, Options, Recommendations")

Adopted according to patient beliefs and the information transmitted

Influenced by external structures

Dependent on the practices of each country (e.g. reimbursement rates impact, norms)

With identical practices in several countries (e.g. EU)

and a treatment efficiency based on

4 parameters :

- the development stage of the tumor
- the type of cancer
- the localisation of the cancer
- the patient (age, medical and surgical histories...)

and may be reduced if cancer cells have mutated into more resistant evolutions ("relapse")

Administered once the disease is detected

preventively when high risk is detected (risk factors)

preventively in all cases (like vaccination)

used

by medical players :  
- research centers / pharmaceutical companies  
- hospitals  
- doctors

by non-experts :  
- patients  
- anyone  
- by associations / foundations

who collected data by himself

who know their existence and ask for them

who can access them through an open platform

who can access them through an interface with data viz

where data is used to estimate unknown interdependencies in a system for a given dataset

to estimate unknown interdependencies in a system between different heterogeneous data sets

to predict new outputs in the given set by using known interdependencies from the other set

driven by risk-benefits balance

focusing on risks only

focusing on benefits only

For research purposes

For patient care purposes

For information purpose & preventing behavior

For management & policy purpose

For funding purpose

For patient advocacy

For shared decision making

**Cancer based studies**

- to discover new types of cancer
- to find genetic or abnormal correlations ;
- to understand correlations between gene expression profiles to disease states or different developmental stages of a cell to model the progression and treatment of cancerous conditions
- to detect genetic predispositions for each patient
- to predict a particular type of cancer or cancer susceptibility recurrence survival
- for each organ understand which type of cancer can occur
- to build right instruments to study cancer

**Global Cancer comprehension, environmental & behavioral data :**

- to better understand cancer risk factors, genetic predispositions, propagation mechanisms, detection/screening
- to study cancer epidemiology (distribution of cancer in population) at the different scales (country, hospitals...)
- to find correlation between heterogeneous data sets (ecological & patient data)
- to better understand the dependencies between cancer and air pollution
- Predict cancer incidence/mortality/survival using risk factors
- to build simulation models
- to determine weak signals across different data sets - accumulation of weak signals or a new type of interdependency
- to estimate unknown factors that influence interdependencies - to get information about cancer on its own (patients) (risk factors, bad qualification, education) & to create an awareness
- to create a common dialogue between different disciplines (sociology, patients)
- to reduce information inequalities

**Cancer treatment & diagnosis & prevention**

- to improve/develop new treatments or screening techniques for cancer
- to reveal how cancer treatment is combined with other treatments & other factors
- to rethink the cancer treatment by taking in account new factors (beyond traditional medical treatment, naturopathy)
- to study the impact of a particular type of treatment for different organs (e.g., chemiotherapy)
- to identify cases similar to a patient case to find the better treatment / care service according to its profile (genomic analogies or scientific literature, clusters)
- to anticipate the efficiency of treatment and side effects according to the patient profile
- to assess more precisely the survival rate of patient and adapt the treatment according to cancer profile
- to assess treatment efficiency or failure ex post including risk & environmental data
- to analyze the secondary effects of treatments, complications, mortality rate

**Advancing general knowledge & understanding of cancer :**

- to develop a continuous comprehension consolidated data sharing on cancer related data from various sources
- to create a shared vision on cancer research (database, datasets)
- to develop dynamic visualization for data cancer
- to display evolution of different types of cancer and variation of different socio-environmental factors like tobacco expenses, pollution
- to develop patient literacy related to cancer ("to decanize" cancer)

**Patient & family care**

- to improve/unify patient care processes of hospitals
- to get information about hospital care services (quality assessment, associations, centers, ambulatory)
- to access to the list of support/care services (built from collected data)
- to improve the distribution of teams and patients in treatment centers (ESPC and turn over)
- to automatically assign patients to different departments based on a type of cancer, socio, treatments
- to improve life quality of patients & their family
- to propose additional services, therapies to improve life quality
- to improve patient - oncologist - general doctor collaboration

**Advance research in other fields (Research reuse) :**

- to get inspired by cancer cells mechanisms for other purposes (than cancer treatment)
- to help research on treatments of other diseases
- to improve patient care services - for economic reasons

**For improved resources & funding allocation :**

- to allocate research fundings according to potential discoveries
- to improve the effectiveness of research paths (balance benefit / cost)
- to improve patient care services - for economic reasons

**Cancer based studies**

- Quick concept : quantitative imaging in cancer : connecting cellular process with therapy tools for drug developers
- Online Gaming : Phyllo Human based computing for bioinformatics to pinpoint mutations in DNA : <http://www.dw.cam.ac.uk/research/research-to-fighting-cancer/ja-17630275>

**Cancer treatment & diagnosis & prevention**

- Childhood cancer -> treatment
- IBM world community grid initiative
- Mooshah initiative to make more therapies available : <https://www.cancer.gov/research-initiatives/mooshah-cancer-initiative>

**Advancing general knowledge & understanding of cancer**

- Cancer epidemiology cohorts AP - HP Compare Project : longitudinal observations of healthy individuals to better understand cancer causes
- Cancer Base project at Stanford

**Patient care**

- Patient advocacy (territories inequality, health issues)
- Start ups like Ma Bulle
- Ethical aspects : Patient & data

**Genomics England (Project whose goal is to sequence 100 000 genomes in the UK)**

**Data**

related to the patient health status (symptoms, health data) and treatments efficiency (and non efficiency)

related to the behavior (nutrition, activity, work)

related to the environment or to other external factors that could affect a person (sociological, demographical factors)

related to cancer genome and mutations (epigenomic data)

related to cancer services

related to the country's economy (PIB, Growth...)

of a patient

of a patient family

of everyone

collected in a reliable (scientifically) way

not collected (scientific desert)

created (through clinical studies/experimentations, from simulation models or from existing research papers)

collected in an unknown way

to answer a specific research question formulated in advance (focus on some variables)

without focus/screening tests, patient treatment...)

by other organisations/instances to answer another research question (data reuse)

through regular medical processes

through clinical studies/lab researches

through "data collection" surveys

through new collaborative techniques/tools (email, doctors)

through patient sharing data (bottom up)

through everyone sharing the data

through social networks

to obtain qualitative data (subject to interpretation)

to obtain quantitative data (through measures/analyses)

in one shot

over time

by the collecting or organisation

by individuals

from one source

from multiple sources

known and controlled

known but uncontrolled

unknown and uncontrolled

In a particular form (texts, pictures, mixed data - clinical and genomic...)

In a standardized form

which is structured

which is unstructured

built and sent

made available

at a cost

freely available

or crowdsourced

which takes time

quickly

instantaneous

by hospitals

by research centers

by patients associations/foundations

by the patient itself (qualified staff on a continuous basis)

by patients family and surroundings

to medical staff / of medical institutions

to everyone

to patients

to be performed

on a targeted sample of patients

on anyone

on a targeted sample of people (healthy, sub-populations)

who give information on a free, voluntary and anonymous basis

who have no choice

who are paid to give information

who have control on the information shared

Data is transformed (aggregated)

Made available as it is

Not transformed because well collected initially (quality and type of data)

to improve their reliability

to extract relevant data  
Dimensionality reduction  
- feature selection and feature extraction

to consolidate data from heterogeneous sources (types of data and contexts like environment or patients profiles)

to convert them into a standard, unified format, exploitable by other studies

to be shared by all actors of cancer ecosystem (create a "common dialog")

to isolate abnormal cases

By applying traditional statistical methods on a small data set : exploration of data based on existing models (epidemiology, molecular bioscience to the performance of clinical trials)

Using methods of supervised learning on larger data set such as task of classification (data mapped in a set of finite classes, e.g. benign or malignant)  
task of clustering (a learning function maps the data into a set of real value variables)  
task of regression  
with tools from machine learning (gradient boosting, KNN, Decision Trees, Random Forest, SVMs,...)

By exploring the data through unsupervised learning methods (eg deep learning) with very large set of data (at least 10^10 points) such as :  
- clustering (find the categories or clusters in order to describe the data items)

Controlling the quality of the model using a defined metric & testing the developed model on a new similar set of data (test set)

By private models of analysis

through Open Science and collaboration

Minimizing the overfitting

By adjusting the new model discovered with pre-existing models

Common Routine

## KNOWLEDGE SPACE

- Ideas explored generated during workshops
- Dominant design for cancer related research
- Potentially interesting ideas to explore

## Epidemium 2016 : Challenge4Cancer projects

- Viz4Cancer
- CancerViz
- Baseline
- Approches prédictives et risque de cancer
- Oncobase
- BD4Cancer
- ELSE
- Venn

Cancer is asymptomatic initially but is highly painful in advanced stages (e.g. bone cancer). Some cancers are incurable

Side effects of radiology and ionisation

No radiation with MRI

self-detection of melanoma through visual observations

IBM Watson Virtual doctor for detection

Complementary therapies - cumulative positive impact with classic approaches :  
- Cancer and Therapeutic Pluralism : <http://euthropiaepidemiareviews.org/359>  
- Cancer and parallel medicine : <http://scd-theses.u->

Patients association advocacy for external informations, collaborative decision making  
art. L1111-4 Right to refuse treatment (in France)

Reimbursement rates varies across different countries

Cancer vaccines

Genomics England (Project whose goal is to sequence 100 000 genomes in the UK)

Data on Facebook and google, use of unexpected data collected from unexpected way

Project data sphere, sharing of cancer patient level data

Ethics, laws about personal data

Data collection standards for clinical research

Deep learning requires large data sets  
Check medical image deep learning, Cornell