

A view from Italy
Learning from the experience of Alleanza Contro il Cancro's
Health Big-Data project

Pier Giuseppe Pelicci

Connect to Win (Paris, Nov. 3rd-5th)



UNIVERSITÀ
DEGLI STUDI
DI MILANO

ACC

Vision:

Cancer Research drives Prevention and Cure

Mission:

- Guarantee access to innovation for all citizens and patients
- Provide Personalized Plans of Prevention, Cure and Follow-up
 - Potentiate the culture of patients as partners

Instruments:

- Potentiate Network Activities (Research and Care)
- Consolidate the national and international leadership of ACC in Oncology

ACC main activity:

Implementation of a National Program of Personalized/Precision Oncology (started 2017):

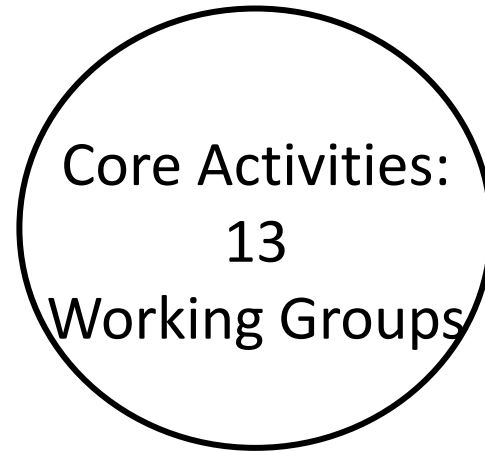
Immediate Goals (first 2 years)

- **Guarantee that all patients of the network are adequately screened (genomic screens)**
- **Guarantee that Genomic-data are properly interpreted and shared among all centers**
- Set-up of NGS-facilities at each center
- Training of a new generation of genomics technologists and clinical bioinformaticians
- Set-up and dissemination of proprietary high-density gene panels (Lung Cancer; Pan-cancer; Leukemia)
- Set-up of a centralized and shared IT infrastructure (ACC data portal)
- Storage and analytical pipelines
- National database of mutations and risk variants
- First draft of a *Prescription Tool*

Current priorities and lines of Activities (2019-today)

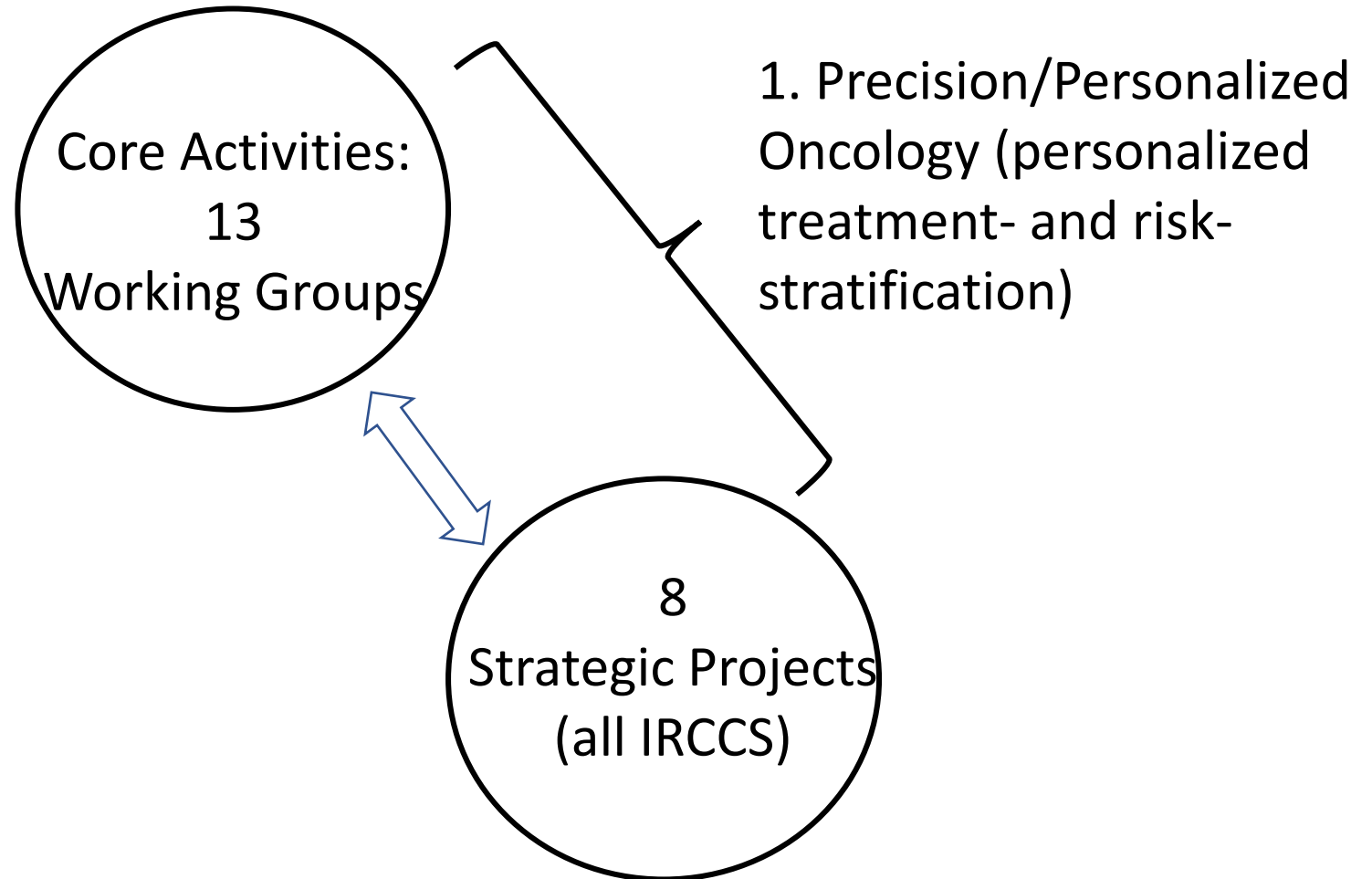
1. Implementation of Personalized/Precision Oncology in the Institutes of the network and in the country (personalized treatment- and risk-stratification)
2. Start Pilot Projects on other areas of Personalized/Precision Oncology:
 - Prevention (screenings and early diagnosis)
 - Health Research (organizational models, health governance; appropriateness, sustainability, outcomes of tests/drugs)
 - Quality of Life (patients, survivors)
3. Potentiate Translational Research

Organization of ACC Activities

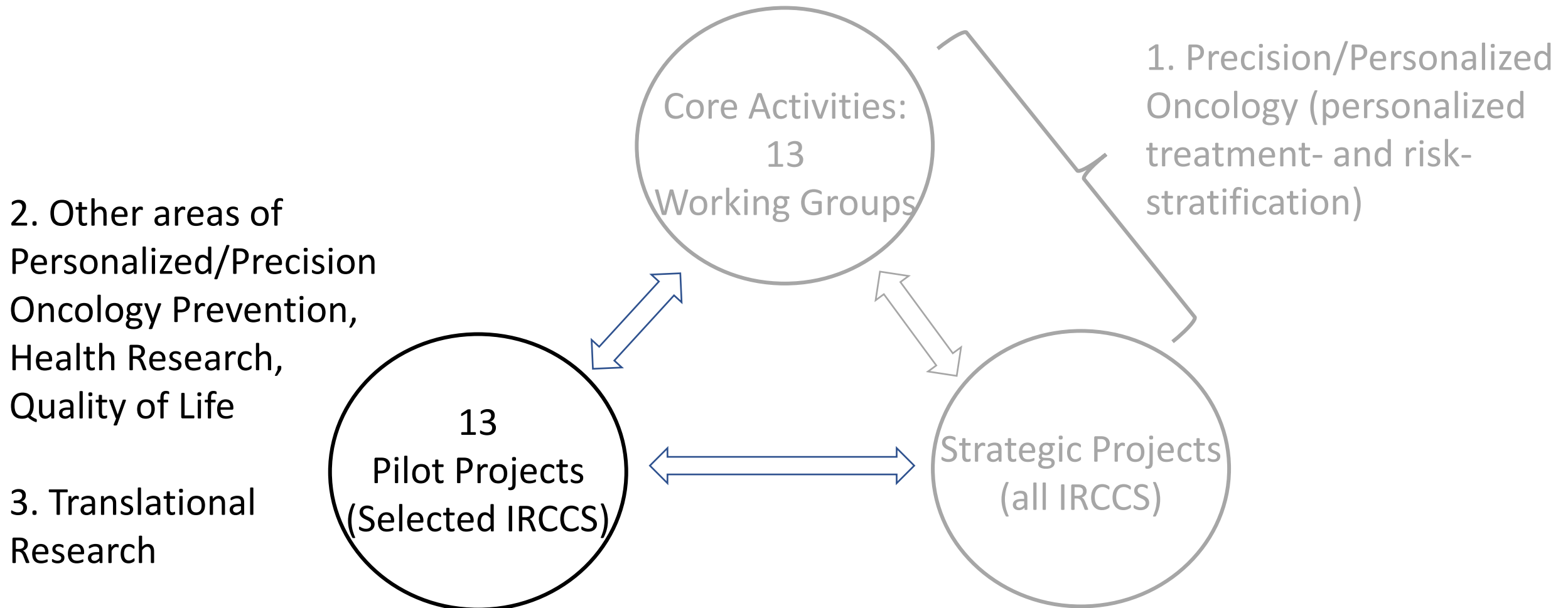


- **8 disease-oriented** (Colon, Breast, Lung, Brain, Melanoma, Hemato-oncology, Sarcoma, Musculoskeletal tumors)
- **5 across diseases** (Genomics and Bioinformatics, Pathology and Biobanks, Immunotherapy, Radiomics, Survivorship care & Nutritional Support)

Organization of ACC Activities



Organization of ACC Activities



Priority 1. Implementation of Precision/Personalized Oncology

Goals:

- I. Consolidate the ability of each IRCCS to perform **genomic screening** (including training of NGS technologists and clinical bioinformaticians)

Special Scientific Projects:

- 1. First National Genomic Screening on Lung Cancer** (Somatic mutations; proprietary Gene Panel): Maeci. Italy-US: 0,7 M E
- 2. First National Genomic Screening on Breast/Ovary/Colon Cancer** (Germline and Somatic Mutations; proprietary pancancer Gene Panel): Ministry of Health 6,5M E
- 3. Dissemination of standard lung-cancer genomic markers nationwide:** Ministry of Health; 5M
- 4. OncoNGS Liquid Biopsy:** H2020: Innovative Procurement: 11 M E

Priority 1. Implementation of Precision/Personalized Oncology

Goals:

II. **Garantee access** of all patients (our IRCCS, other nation hospitals) to innovatite **treatments** (targeted drugs and immunotherapies)

- a) Create a network of Molecular Tumor Board (MTB) for each Institute (somatic mutations and germline variants)
- a) Ensure the availability of off-label drugs
- a) Enhance the generation and collection of Real World Data and the generation of Real World Evidence

Special Scientific Projects:

5. **First National Personalized-Oncology Clinical Trial** (Mutational Oncology Project); ongoing

- Consensus document on the development and organization of Mutational Oncology in Italy” (ACC, AIOM, SIAPEC, RES, Cittadinanza Attiva, many others)
- MTBs, genetic tests, access to new drugs, Data collection and Genomic IT Platform, follow-up and RWD collection; Privacy and GDPR compliance, patient involvement

Priority 1. Implementation of Precision/Personalized Oncology

Goals:

III. Build enabling **Infrastructures**

- a) Create a network infrastructure for multicentric clinical studies
- b) Generate common IT tools for the extraction, interpretation and analysis of interoperable omic, clinical and RWD data (Potentiation of local IT systems; Centralized Databases; Prescription and Analytical Computational Tools)
- c) Create an IT infrastructure for the management of information within of individual MTBs, their coordination and information exchange

Special Scientific Projects:

6. **Forza Project** (Infrastructure for Multicentric Clinical Trials)
7. **Health Big Data Project** (Creation of a centralized structure for sharing/analyses of clinical scientific data of the 51 IRCCS; PoliMI, INFN): Ministry of Finance and Ministry of Health: 55M
8. **Digicore Project** (European Consortium for the generation / sharing / analysis of RWD in oncology)

Priority 2. Significantly extend ACC activities to include:

Pilot Projects:

I. Prevention (screenings and early diagnosis)

1. TP53 in the Pap test (TOWARDS) (***Mario Negri, Gemelli, INT-Pascale, IEO, OSR, INT-MI, CRO, ICH***)
2. Early diagnosis in Head&Neck tumors (***GPII, INT-Pascale, IEO, IRE, OSM***)
3. Neuregulins/HER in the invasive lung adenocarcinoma (***CSS, IRST, OSR, AUSL-Reggio Emilia, Gemelli***)
4. Locally advanced NSCLC (***OSR, IRE, INT-MI, OSM, INT-Pascale, CSS, ICH, Gemelli, San Matteo, IEO, FPO, CRO, IOV, IRST***)

Priority 2. Significantly extend ACC activities to include:

II. Health Research

- organizational models
- health governance
- appropriateness, sustainability, outcomes of tests/drugs

III. Quality of Life

(patients, survivors)

Pilot Projects:

5. **RARITY:** Italian Tumor Registry of Rare Tumors (*INT-MI, IEO, IOV, FPO, ICH, OSR, OSM, IRE*)
6. **Overdiagnosis, Overtreatment and Outcome** in cancer patients (**CRO**, *INT-MI, INT-Pascale, IRST, AUSL-Reggio Emilia, GP11*)
7. **PPRER:** Re-evaluation of prevention, screening and early diagnosis in high-impact cancers (**CRO**, *IRST, IOV*)
8. **BioPerspec:** Novel bio-markers for targeted-drugs: from governance to RWD (**GP11**, *INT-Pascale, IRE, FPO, OSC Don Calabria, IRST*)
9. **Nutritional support** (**Gemelli**, *San Matteo, INT-Pascale, IOV, FPO*)

Priority 3. Potentiate Translational Research

Goals:

- Mechanisms of Metastasis
- Mechanisms of Resistance to targeted drugs and immunotherapies
- Novel biomarkers (risk and treatment stratification)
- New therapeutic approaches

3 Pilot Projects:

1. **M ϕ -score:** Macrophage Immune-score (*ICH, IRE, OSR, San Matteo, AUSL-Reggio Emilia*)
2. **RADECISION:** Precision/Personalized Radiotherapy (*INT-MI, IEO, GPII, CRO, FPO, Gemelli, ICH, OSC Don Calabria, IRE*)
3. **Liquid Biopsy** in Ovary Cancer (*ICH*)

1 Special Project:

4. **CAR-T project:** Ministry of Health: 7 M E

A few examples:

- **The ACC-proprietary Gene-Panels**
- The ACC Lung-Cancer Clinical Study
- The Health Big-Dta Project

ACC-Gene Panels: From tumor-type specific to Pan-Cancer + Germline

ACC-Lung-GenePanel

Large GenePanel (182 genes; 0,8MB) with all the known:

- *Actionable Genes* (161)
- *Lung-cancer Drivers* (33)
- *Lung cancer translocations* (89)
- *Drug-toxicity variants* (196)

Validation/Dissemination:

- *NGS Technology*
- *Bioinformatic pipeline*

Low cost (<500 E)

ACC-Gersom

Large GenePanel (467 genes; 1,6MB) with all the known:

- *Actionable Genes* (135)
- *Cancer Drivers* (299; 2018 TCGA release)
- *Tranlocations* (89)
- *Cancer-risk genes* (150)
- *Drug-toxicity variants* (196)

110 genes optimized for CNV

Low cost (<700 E)

Challenges

- Tumor Mutation Burden
- HLA
- CNV
- Specific signatures (HRD as stratification marker)
- Annotation of risk variants

Goal: patient-drug stratification

ACC-Gene Panels: From tumor-type specific to Pan-Cancer + Germline

ACC-Lung-GenePanel

Large GenePanel (182 genes; 0,8MB) with all the known:

- Actionable Genes (161)
- Lung-cancer Drivers (33)
- Lung cancer translocations (89)
- Drug-toxicity variants (196)

Validation/Dissemination:

- NGS Technology
- Bioinformatic pipeline

Low cost (<500 E)

ACC-Gersom

Large GenePanel (467 genes; 1,6MB) with all the known:

- Actionable Genes (135)
- Cancer Drivers (299; 2018 TCGA release)
- Tranlocations (89)
- Cancer-risk genes (150)
- Drug-toxicity variants (196)

110 genes optimized for CNV

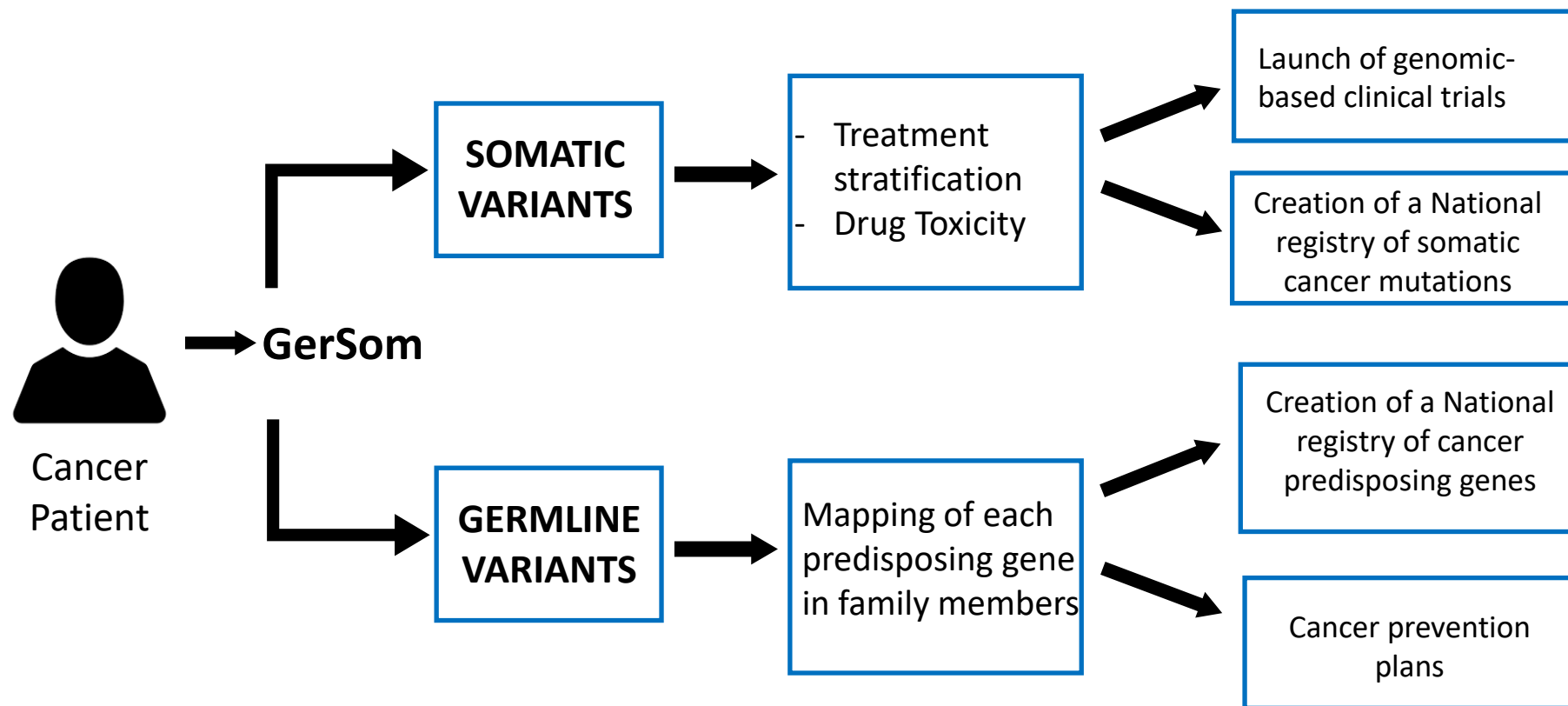
Low cost (<700 E)

Challenges

- Tumor Mutation Burden
- HLA
- CNV
- Specific signatures (HRD as stratification marker)
- Annotation of risk variants

Goal: patient-drug stratification and risk assesment

The GerSom Clinical Strategy



Goal: analyze simultaneously germline and cancer-predisposition variants

ACC-Gene Panels: From tumor-type specific to Pan-Cancer + Germline

ACC-Lung-GenePanel

Large GenePanel (182 genes; 0,8MB) with all the known:

- Actionable Genes (161)
- Lung-cancer Drivers (33)
- Lung cancer translocations (89)
- Drug-toxicity variants (196)

Validation/Dissemination:

- NGS Technology
- Bioinformatic pipeline

Low cost (<500 E)

ACC-Gersom

Large GenePanel (467 genes; 1,6MB) with all the known:

- Actionable Genes (135)
- Cancer Drivers (299; 2018 TCGA release)
- Tranlocations (89)
- Cancer-risk genes (150)
- Drug-toxicity variants (196)

110 genes optimized for CNV

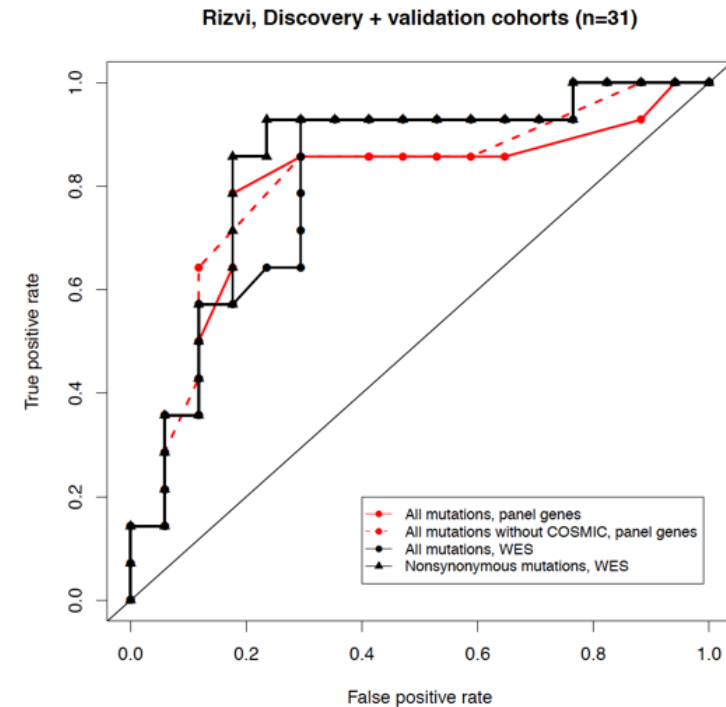
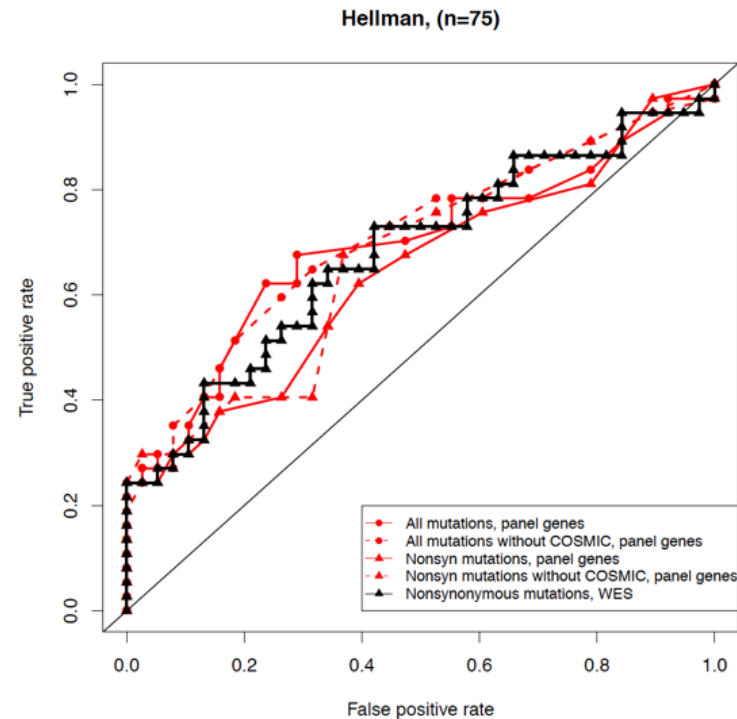
Low cost (<700 E)

Challenges

- Tumor Mutation Burden
- HLA
- CNV
- Specific signatures (HRD as stratification marker)
- Annotation of risk variants

GerSom Pipeline validation and improvement: Tumor Molecular Burden (TMB)

- SNV
 - CNA
 - Fusions
 - TMB
- Simulated analysis based on published WES datasets
 - Good correlation
 - Good ROC



Hemathology gene panels

- Risk of therapy-related myeloid neoplasms in cancer survivors;
- Analysis of clonal hematopoiesis and predisposition to t-MN

MyeloPanel

Large GenePanel (255 genes; 1,39MB)
with all the known:

- Actionable Genes (113)
- AML cancer Drivers (37)
- AML translocations (250)
- Hematological diseases risk-genes (79)
- Common to more than 1 list (26)

Sensitivity: Conventional NGS (5-10%)

Purpose:

- patient stratification
- germline risk

Low cost (<500 E)

CHIP-UMI Panel

High-sensitivity GenePanel (80 genes;
400 KB)

80 most frequently mutated genes in
clonal hematopoiesis:

- Tier1: 43 genes in >1 study and recurrent in >2
- Tier2: 25 genes in >1 study but recurrent in >1
- Tier3: 12 genes in >1 study but never recurrent

Sensitivity: UMI-NGS (0,1-1%)

Purpose: Clonal Hematopoiesis

Low cost (<500 E)

Duplex-Panel

Ultra-high-sensitivity GenePanel
(27 genes; 70 KB)

27 most frequently mutated genes in
clonal hematopoiesis:

- 13 hotspots
- 14 all coding regions

Sensitivity: Duplex sequencing
(>0,001%)

Purpose: Clonal Hematopoiesis

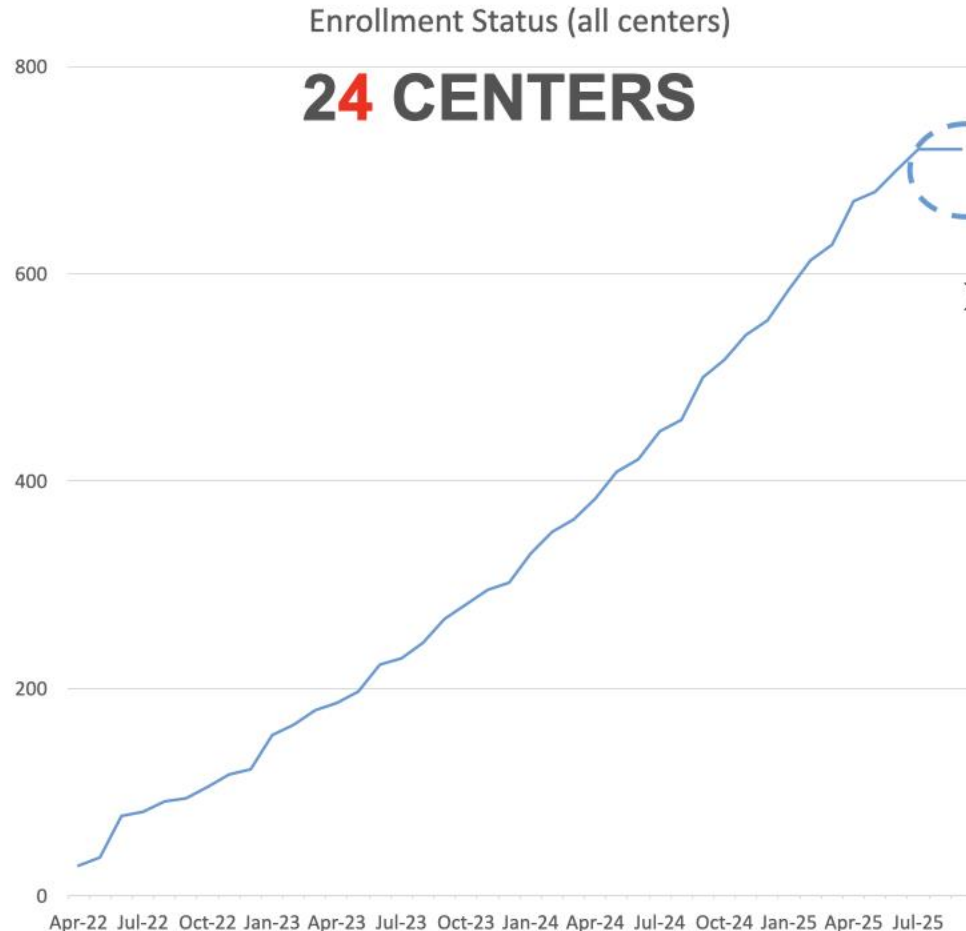
Cost (\approx 800 E)

A few examples:

- The ACC-proprietary Gene-Panels
- **The ACC Lung-Cancer Clinical Study**
- The Health Big-Data Project



The ACC Lung-Cancer Clinical Study



24 CENTERS

720_{/1000} **patients**
(September 2021)

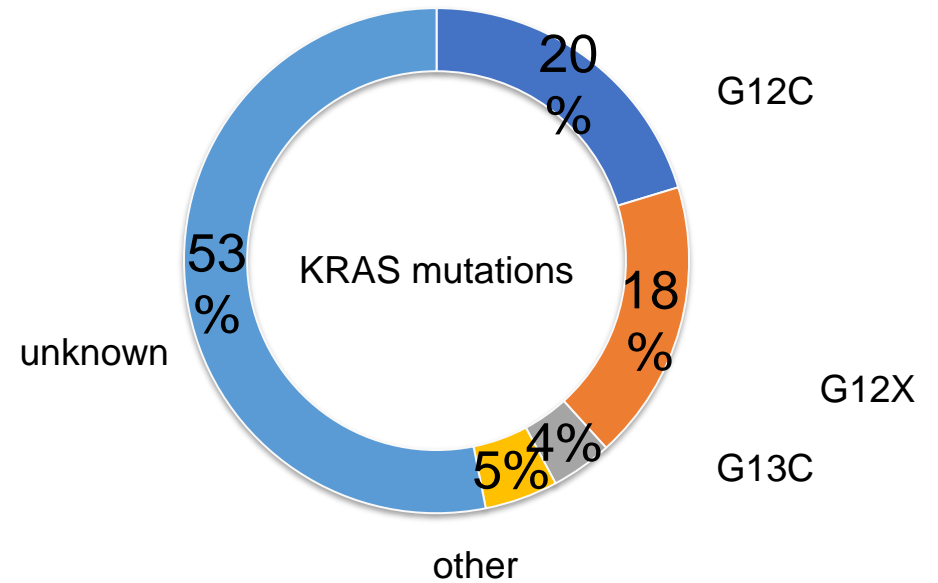
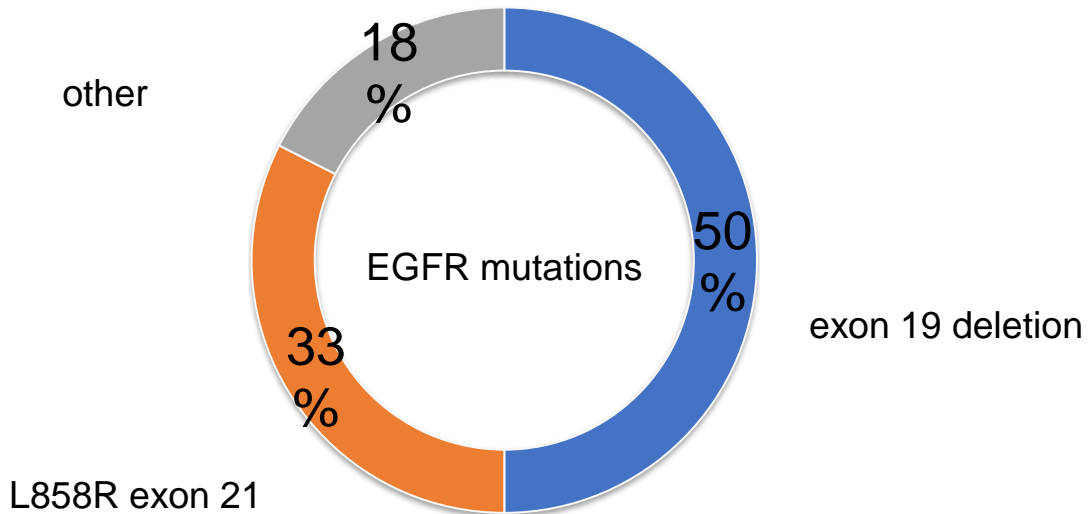
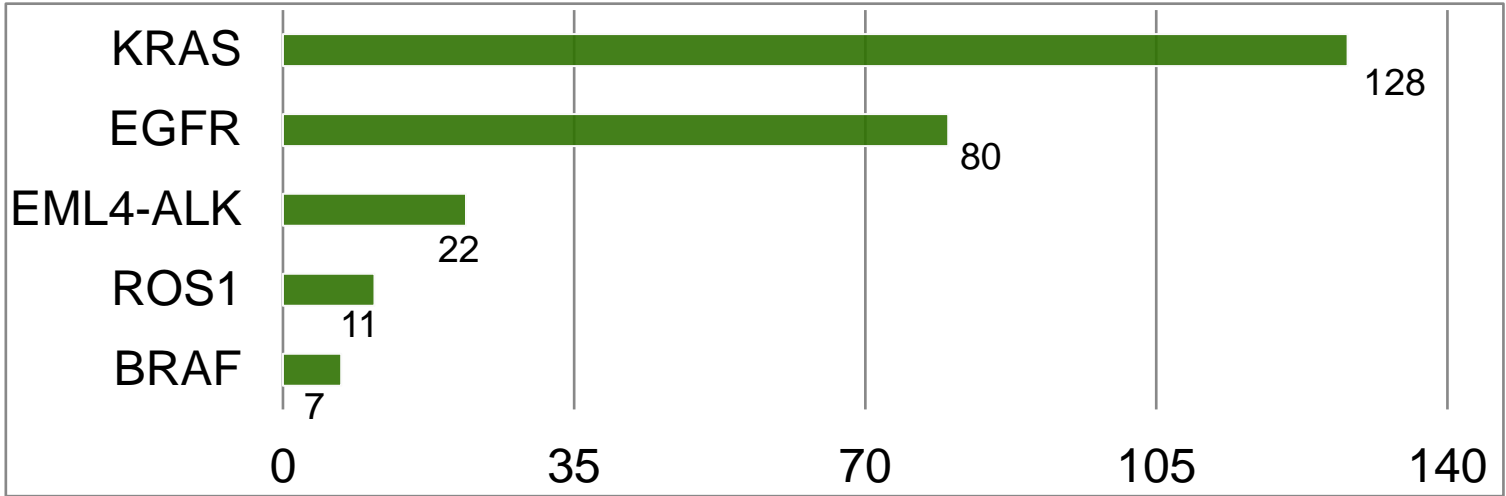
➤ **Possibly reaching 1000 patients by the end of 2022 (according to CRO enrollment data)**

Primary Objective

To validate
the ACC-Lung Panel as a tool for
molecular screening in patients
with NSCLC



Results





The ACC Lung-Cancer Clinical Study

➤ *Satellite studies*

LIQUID
BIOPSY

IMMUNOSCORE

RADIOMICS

Radiomics validation of the ACC Lung Panel in NSCL Cancer (RATIONALE study)

Satellite study to ACC Lung

18 centers authorized - 10 centers active: 800 patients expected

Primary aim

Development of radiogenomics models for :

- prediction of EGR, EML4-ALK, cMET, ROS1, RET, BRAF/KRAS or no mutational status
- treatment response prediction and overall survival

Secondary aims

- Delta radiomics applications for the available dataset
- Development of clinical decision support systems (DSS)
- Development of an automatic segmentation algorithm for lung volumes (CNN)



RECORD study state-of-the-art

Istituto	Lims	OneData
IEO	35	35
INT Milano	35	35
CRO di Aviano	36	36
FP Candiolo	35	35
IRCCS Istituto Clinico Humanitas	51	51
IRST Meldola	-	-
INT Pascale	56	56
IRCCS Ospedale San Raffaele	55	55
Fondazione Policlinico Universitario "A. Gemelli" IRCCS	41	41
Azienda USL di Reggio Emilia IRCCS	9	9
Totale	353	353

- Technical platform ready
- 35% total cases uploaded
- Optimization 50% done

Received: 22 January 2021 | Revised: 13 April 2021 | Accepted: 23 April 2021

DOI: 10.1002/jbc.29110

ONCOLOGY: RESEARCH ARTICLE

Radiogenomics prediction for MYCN amplification in neuroblastoma: A hypothesis generating study

Angela Di Giannatale¹ | Pier Luigi Di Paolo² | Davide Curione² |
Jacopo Lenkiewicz³ | Antonio Napolitano⁴ | Aurelio Secinaro² | Paolo Tomà² |
Franco Locatelli^{1,5} | Aurora Castellano¹ | Luca Boldrini³

Pediatric Blood & Cancer
aspho
The American Society of
Human Oncology
WILEY

cancers

MDPI

Article

A Multicentre Evaluation of Dosiomics Features Reproducibility, Stability and Sensitivity

Lorenzo Placidi^{1,*} | Eliana Gioscio^{2,†} | Cristina Garibaldi³ | Tiziana Rancati² | Annarita Fanizzi⁴ |
Davide Maestri⁵ | Raffaella Massafra⁴ | Enrico Menghi⁶ | Alfredo Mirandola⁵ | Giacomo Reggiori⁷ |
Roberto Sghedoni⁸ | Pasquale Tamborra⁴ | Stefania Comi⁹ | Jacopo Lenkiewicz¹ | Luca Boldrini¹ |
and Michele Avanzo¹⁰



- Data upload
- Quality check
- RT progress
- Procedures
- Pipelines
- Reagent
- Global stats

- Knowledge database
- Data retrieval

ALLIANCE AGAINST CANCER DATA PORTAL.

A data portal to guide researchers and physisists toward the resources generated by the **Alliance Against Cancer organization**.

[FIND OUT MORE](#)

A few examples:

- The ACC-proprietary Gene-Panels
- The ACC Lung-Cancer Clinical Study
- **The Health Big-Data Project**

Health Big-Data Project (start March 2020)

Main goal:

Creation of technological platforms enabling the generation/extraction, collection, sharing and analyses of scientific and clinical data of patients of each of the 50 Research Hospitals of the project.

(Not only ACC, also the Cardiology and Neuroscience Research-Hospital Networks)

Specific goals:

1. Upgrade of IT infrastructures of each IRCCS, aiming at the extraction, integration, and interoperability of clinical and scientific data
2. Potentialiation of the three Network IT-infrastructures
3. Creation of a centralized, integrated and federated IT platform (starting from the 3 network infrastructures)

Health Big-Data Project

Data types :

- omic data (genomic, transcriptomic, proteomic, metabolomic)
- clinical data (electronic medical folder, patient follow-up data, real world data)
- imaging and radiomic data
- data provided by the patient
- (In the mid term) data from biosensors, social and financial data

Key element: extraction of unstructured data (metadata, such as concepts, keywords, relation between concepts etc, in natural language or other free formats). Collaborations are planned with experts of natural language processing.

Health Big-Data Project

The IT platforms will have to guarantee:

1. *Extraction of structured and unstructured data* derived from the different activities of each IRCCS
2. *Connectivity* among IT platforms in each IRCCS of the project and with other Italian and non-italian research institutes, with databases of the national health system and the public international databases (TCGA, COSMIC, CGHub, ClinicalTrials.gov, etc.).
3. *Data analysis*. The platform will be able to:
 - a. Integrate heterogeneous data (omic and clinical) for each single patient (to provide operative tools to the Molecular Tumor Board and for the creation of prescription databases);
 - b. aggregated analysis on populations of patients for various omic and clinical parameters. A key element of the platform of data analysis will be the incorporation of algorithms of artificial intelligence and machine learning.

Health Big-Data Project

The IT platforms will have to guarantee:

4. That sensitive data of the patients are treated according to privacy rules (GDPR)
5. *Policies* for sharing and access to the data generated by each IRCCS, namely:
 - data of the network (generated through multicentric Network activities and owned by the Network)
 - data of each IRCCS (generated by their institutional activity and owned by each IRCCS). The choice of project and levels of access/data sharing will be established through a governance system involving all IRCCS
4. Storage of common (of the networks) and proprietary (of the IRCCS) data

Health Big-Data Project

Parallel activities:

- Implementation of privacy and patient safety data rules (General Data Protection Regulation, GDPR)
- Pilot project network for the generation and collection of omic data
- Generation of real world data, quality and data standardization
- Potentiation of culture of patient as partner
- Generation of policies for the sharing and access of the data generated by the IRCCS

Health Big-Data Project

Project organization:

- **7 Working Groups (WG).** The WGs represent the site of discussion among IRCCS on specific subjects (5-10 participants from each IRCCS):
 - i. Study of text data available in digital format in the different IRCCS (NLP)
 - ii. Extraction of omic data
 - iii. Interaction between electronic clinical folder and programs of clinical data management for research purpose;
 - iv. Platform of data integration
 - v. Systems architecture
 - vi. Data analysis
 - vii. Ethics and privacy

Health Big-Data Project

Project organization:

- a team of >50 IT-scientists devote to the project and assigned to the IRCCS
- a team of >20 IT-scientists and >40 clinicians/scientists from each IRCCS and devoted to the potentiation of the centralized Network IT infrastructures

The project has been launched with the survey of IT infrastructures of each IRCCS

Health Big-Data Project

Collaborators:

INFN

CNAF: “Center for Research and Developm. in Information and Communication Technologies”

- **EPIC Cloud:** cloud service for projects dealing with clinical, biomedical and genomic data (GDPR-compliant, ISO-certified)
- **Leonardo super-computer for Big-data analyses:** one of the eight pre-exascale supercomputers that will form the European high-performance computing network

Politecnico di Milano

Strong community in Data Science

Approaches: crowdsourcing, data extraction, data integration

Technologies: cloud computing, parallel processing for big data, AI, machine learning

Areas: smart cities, business analytics, socials, genomics

The challenge: Systematic collection of high-resolution multi-dimensional data

Disease data

Tumor
Multi-omics

Person Data

Germline
DNA and Microbiome

Medical history, RWD

Environmental and
wearable sensors

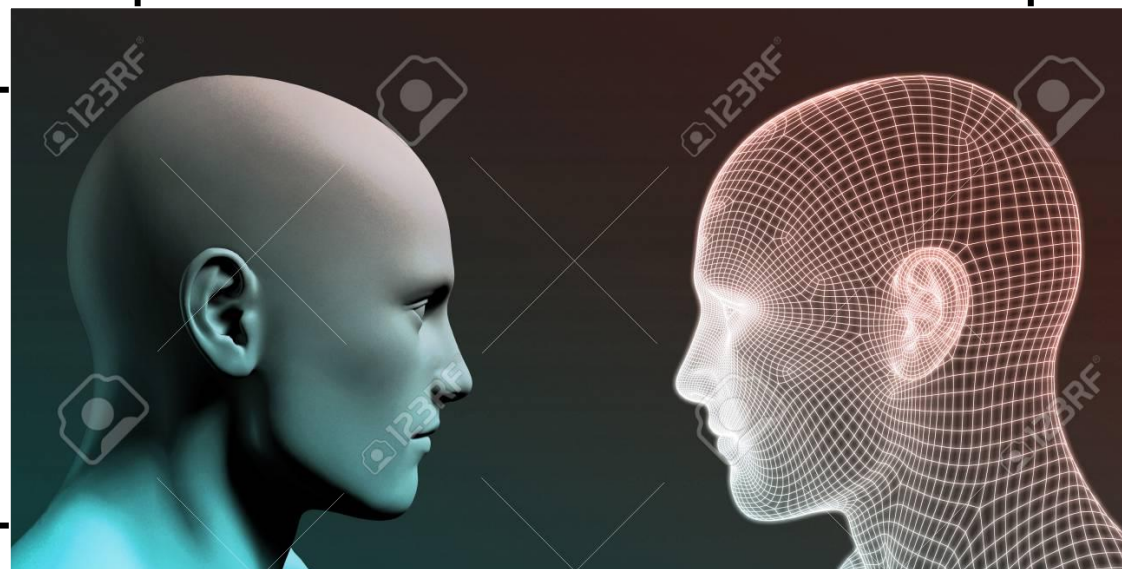
Imaging
(digital imaging, radiomics)

Lyfe-style

Tissue-on-chip

Liquid Biopsies

Generation of Patient/Person Digital Avatars



Generation of a Universal Knowledge Resource to manage health and care

Systematic collection
of high-resolution
data across millions
of patients



Real-time knowledge
generation (artificial
intelligence systems)



Disease re-classification
(not based on human-
defined disease classes)

High-definition Prevention
and Treatment

Single-patient
treatment decisions
(intelligent-decision support)

Real-time data generation and sharing