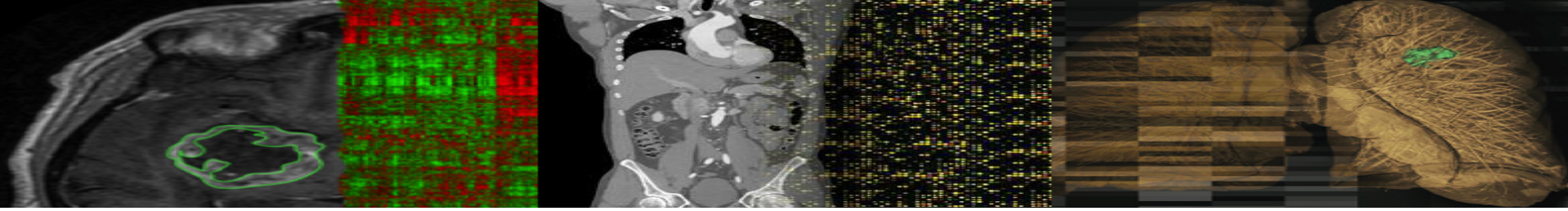


BIOMAT 2017

INTEGRATING IMAGING AND OMICS DATA: A REVIEW

Laura Antonelli, Mario R. Guarracino, Lucia Maddalena, Mara Sangiovanni

Moscow, October 30, 2017

ICAR CNR Institute for High-Performance Computing and Networking

Computational and Data Science Laboratory

AN *OMICS IMAGING* RATIONALE

Bridging **imaging** and **omics** worlds, and exploring their connections, has the potential to provide important **new insights** into the **phenotypic characteristics** and **molecular mechanisms** of normal and/or disordered biological structures and functions.

These in turn will impact the **development** of new **diagnostic, prognostic, therapeutic, and preventive approaches**.

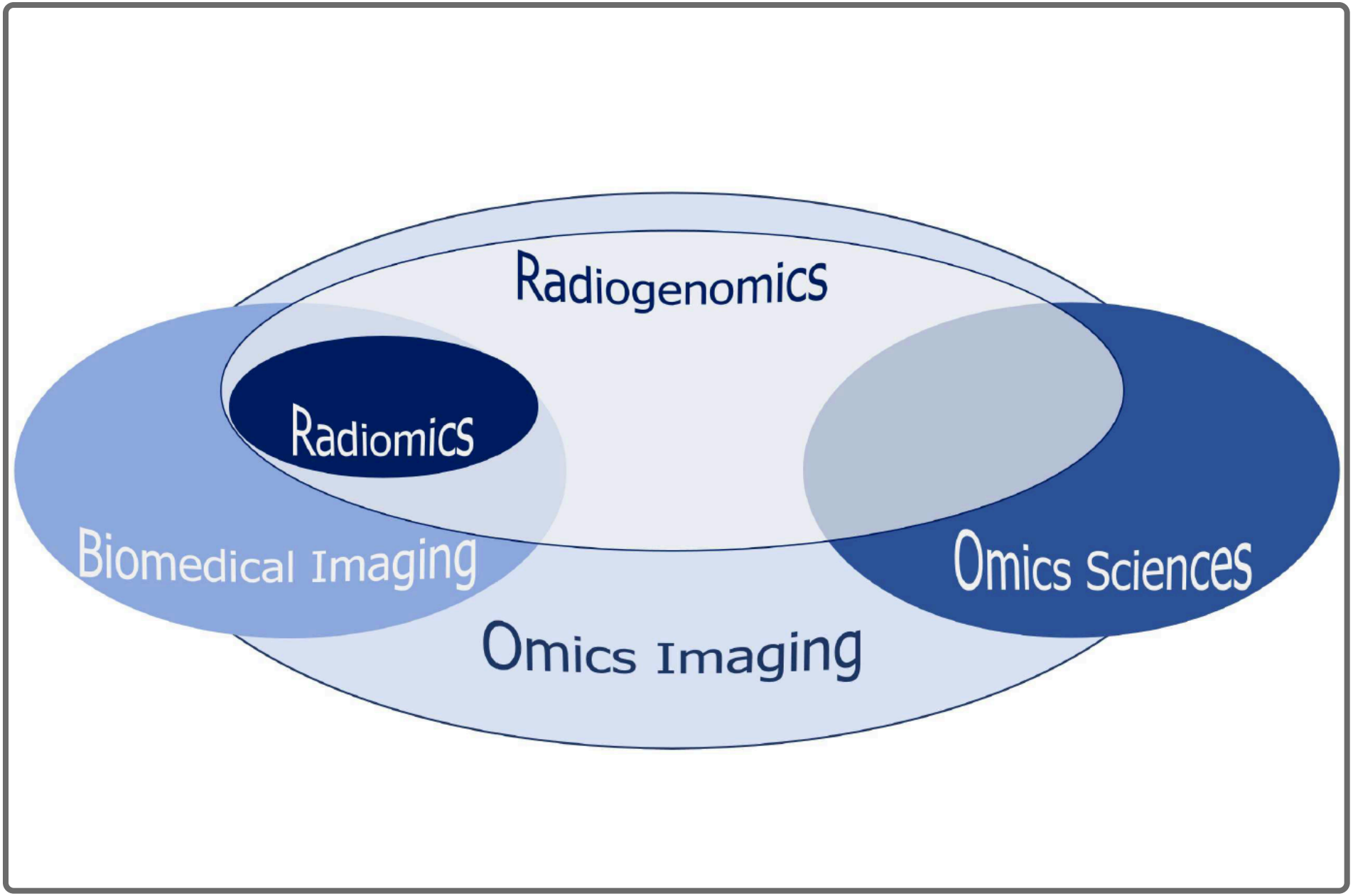
Definition

Omics Imaging investigates algorithms, methods and software tools for analysis and integration of omics (genomics, transcriptomics, proteomics, other omics) and structural, functional, and molecular imaging data.

Also known as *Imaging Genomics*, it includes Radiogenomics, which refers to the integration of radiological imaging features and genomics data.

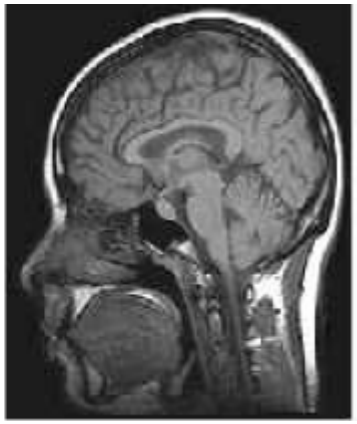
It is not to be confused with the study of genetic variations associated with response to radiation (Radiation Genomics).

FROM RADIOGENOMICS TO *OMICS IMAGING*



EXAMPLES OF IMAGING DATA

CT

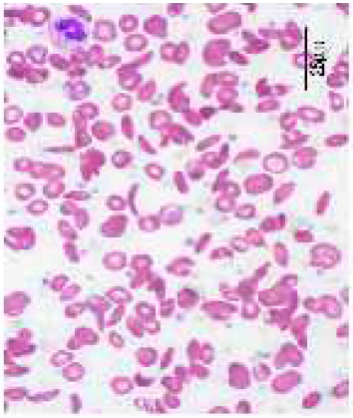


MRI

SPECT

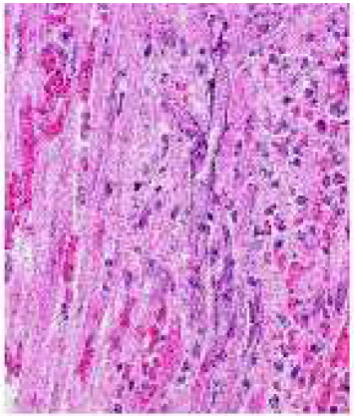
fMRI

DERMOSCOPIC



BRIGHT FIELD MICROSCOPY

PHASE-CONTRAST MICROSCOPY



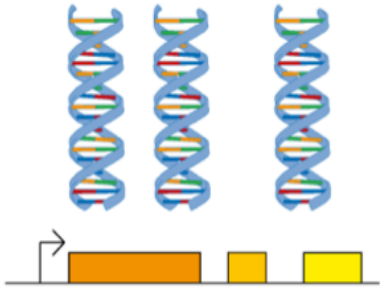
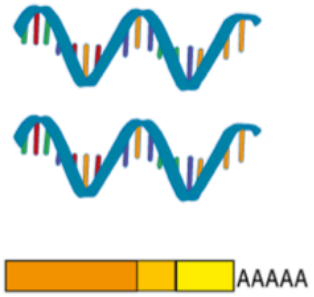
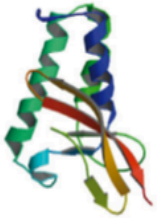
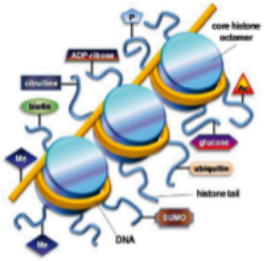
HYSTOLOGICAL

IMAGING DATA

IMAGING DATA SOURCE	ACRONYM	TYPE OF DATA
Computerized Tomography	CT	STRUCTURAL
Magnetic Resonance Imaging	MRI	STRUCTURAL
Diffusion MRI	dMRI	STRUCTURAL
Functional MRI	fMRI	FUNCTIONAL
Dynamic Contrast-Enhanced MRI	DCE-MRI	FUNCTIONAL
Positron Emission Tomography	PET	FUNCTIONAL
Single Photon Emission Computed Tomography	SPECT	FUNCTIONAL
Combined PET and CT	PET-CT	STRUCTURAL **and** FUNCTIONAL
LOw-Resolution brain Electromagnetic Tomography	LORETA	STRUCTURAL **in TIME**
Microscopy Imaging	MI	ANATOMICAL

IMAGING DATA USED IN *OMICS IMAGING* PUBLISHED WORKS

EXAMPLES OF OMICS DATA

GENOMICS	TRANSCRIPTOMICS	PROTEOMICS	EPIGENOMICS
		 > Replicase polyprotein 1ab, P19811 MATFSATGFG GSFVRDWSLD LPDACE- HGAG LCCEVDGSTL CAECFRGCEG M	
- genomic DNA Sequencing - annotations	-microarray hybridization -RNA-seq	- liquid chromatography + - Mass Spectrometry	- ChIP-Seq - bisulfite seqencing
- SNPs - Rare variants - Structural variations	- RNA expression - Alternative splicing	- Protein expression - Protein structure	- DNA methylation - Histone modifications

OMICS DATA

OMICS CATEGORY	TYPE OF DATA
GENOMICS	Single Nucleotide Polymorphisms (SNPs)
	Rare variants
	Copy Number Variations (CNVs)
EPIGENOMICS	DNA Methylation
TRANSCRIPTOMICS	messenger RNA (mRNA) expression
	micro RNA (miRNA) expression
	long non-coding RNA (lnRNA) expression
PROTEOMICS	Protein expression
METABOLOMICS	Metabolite abundances

OMICS DATA USED IN *OMICS IMAGING* PUBLISHED WORKS

OMICS IMAGING APPLICATIONS

The scale and complexity of multi-dimensional imaging and omics data provide **unprecedented opportunities** in enhancing mechanistic **understanding** of **complex traits and disorders**, which in turn can facilitate the diagnostic, prognostic, and therapeutic progress.

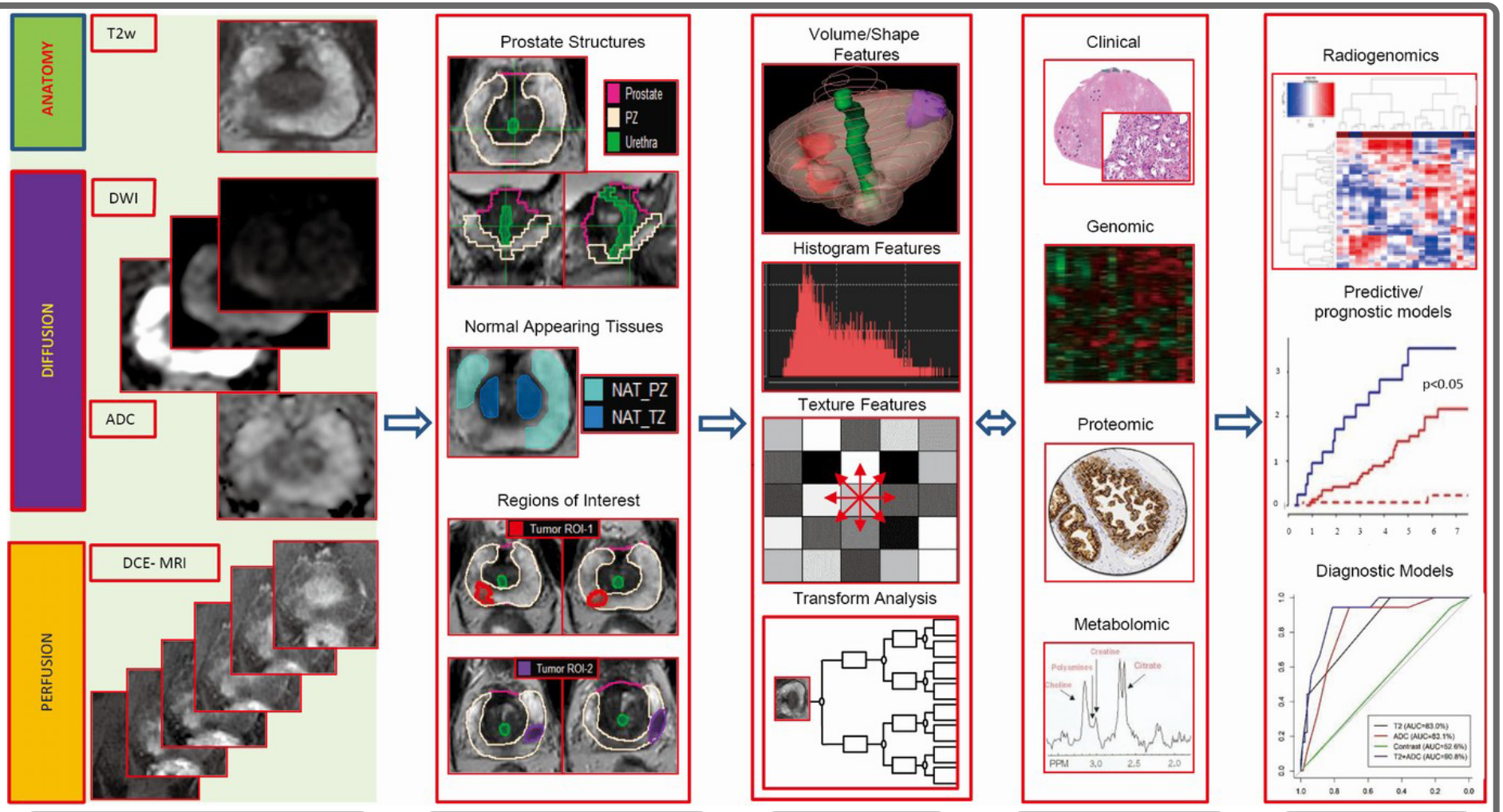
OMICS IMAGING APPLICATIONS

CATEGORY	TYPE
CANCER	Breast Cancer
	GlioBlastoma Multiforme (GBM)
	HepatoCellular Carcinoma (HCC)
	LUng ADenocarcinoma (LUAD)
	Non-Small Cell Lung Cancer (NSCLC)
HEALTHY CONTROLS	Genetic variations associated with human brain volume
NEUROLOGICAL DISEASES	Alzheimer Disease (AD)
	Parkinson Disease (PD)
	Schizophrenia (SCZ)
PSYCHIATRIC DISORDERS	Anxiety and Stress Response
	Attention Deficit Hyperactivity Disorder (ADHD)
	Specific Learning Disorders (SLD)

OMICS IMAGING RESEARCH

- METHODS
  - Imaging data processing
  - Omics data processing
  - Data integration and analysis
- DATA
  - Availability
  - Reliability and consistency
  - Integrated resources

OVERVIEW OF THE PROCESS: AN EXAMPLE FROM LITERATURE



- MULTIPARAMETRIC MRI: STRUCTURE + FUNCTION

- IMAGE REGISTRATION AND PREPROCESSING
- ROI EXTRACTION

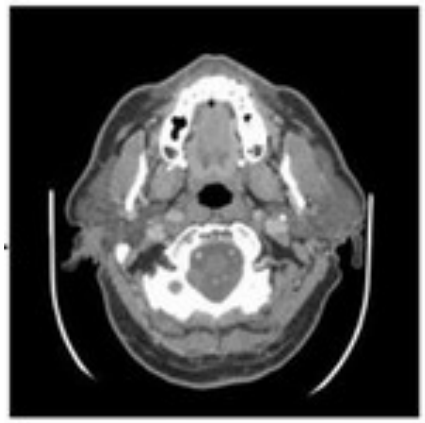
- IMAGING FEATURES EXTRACTION

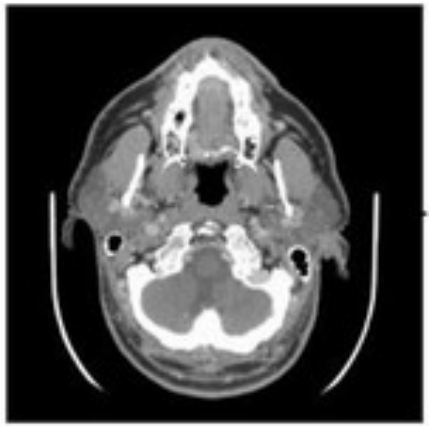
- OMICS FEATURES EXTRACTION

- INTEGRATION AND ANALYSIS

Stoyanova et al. ***Prostate cancer radiomics and the promise of radiogenomics.*** **(2016)**

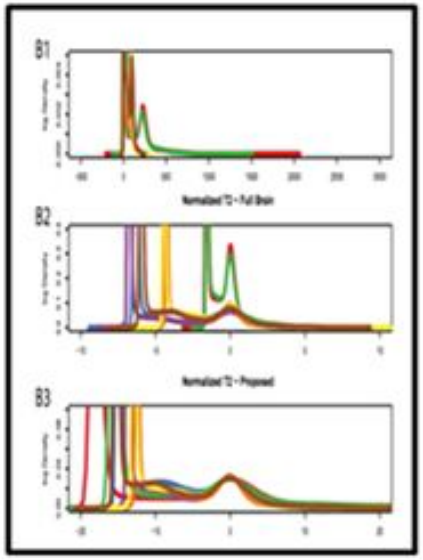
IMAGING DATA PROCESSING AND ROI EXTRACTION



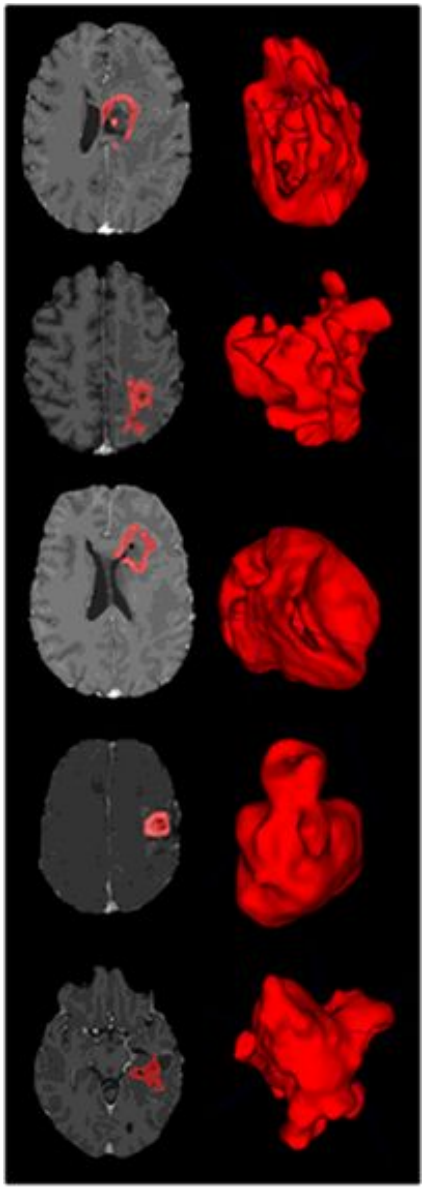


- Image registration across the extracted slices

- Basic image pre-processing
  - Histogram equalization
  - Denoising
  - …
- Imaging-dependent pre-processing



ROI EXTRACTION BY SEGMENTATION



- Image segmentation across the extracted slices: extraction of the Region Of Interest (ROI)
  - Manual
  - Semiautomated detection (Computer-Aided Detection)

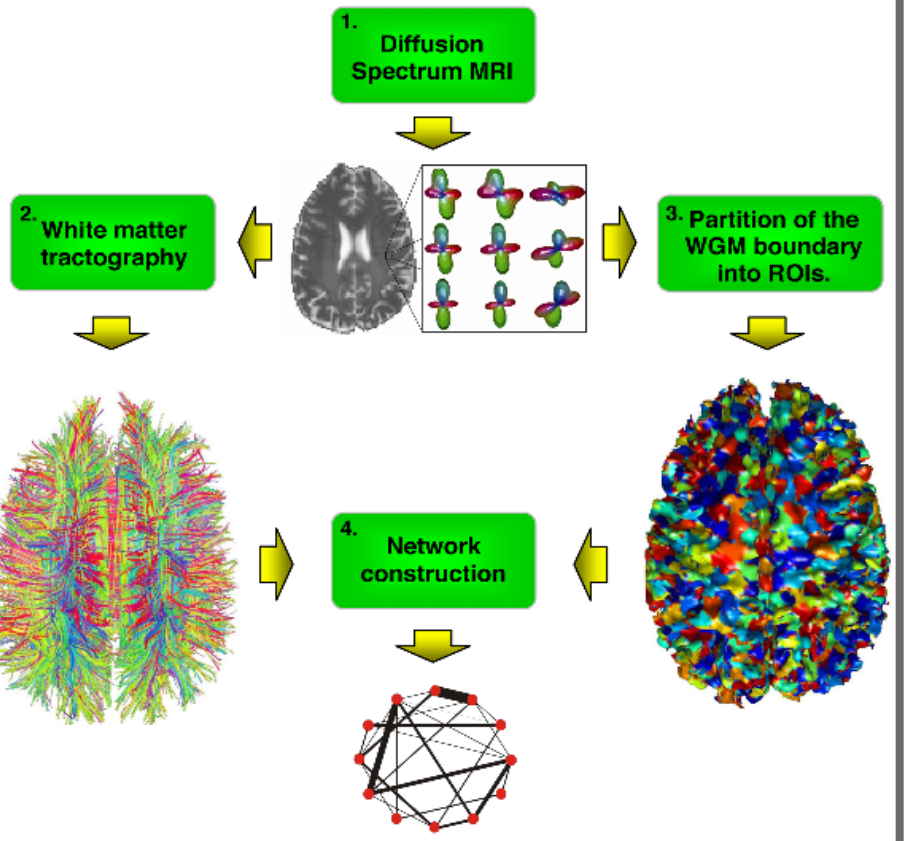
SUITABLE FOR TUMORS OR LESIONS

OTHER ROIs: EXAMPLE FROM BRAIN IMAGING

Designed for Brain Connectivity Analysis

- Diffusion maps
- Tractography
- Activity networks

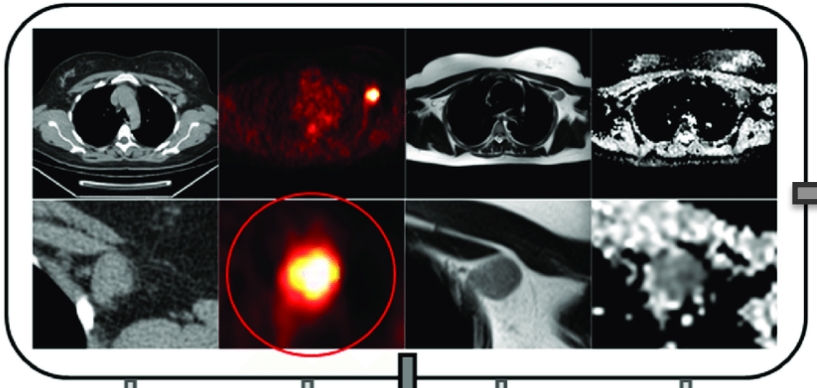
- **Parcelation** of activated regions to extract ROIs
  - Activation Likelihood Estimation (ALE)
  - Automated Anatomical Labeling of Activations (AAL)

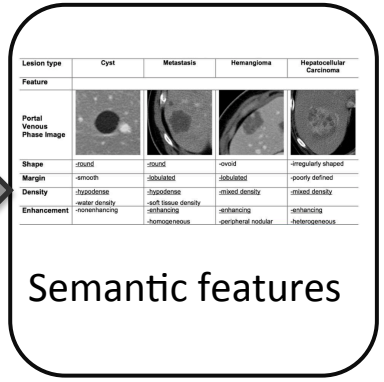


Adapted from: Hagmann et al. *Mapping Human Whole-Brain Structural Networks with Diffusion MRI.* (2007)

IMAGING FEATURE EXTRACTION

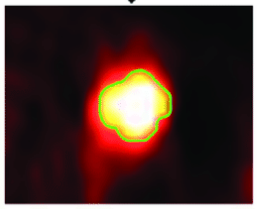
MULTIMODAL IMAGE ACQUISITION





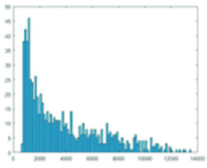
Semantic features

SEGMENTATION



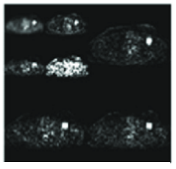
ESTIMATION OF DESCRIPTIVE FEATURES

Shape features



Intensity features

Texture features

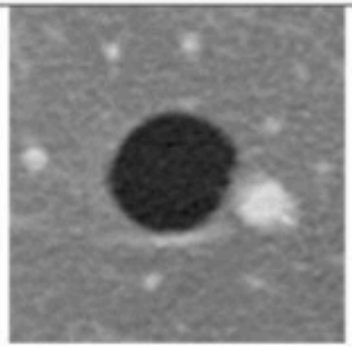
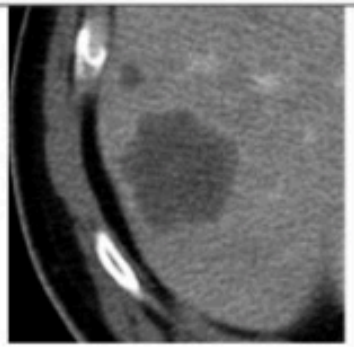
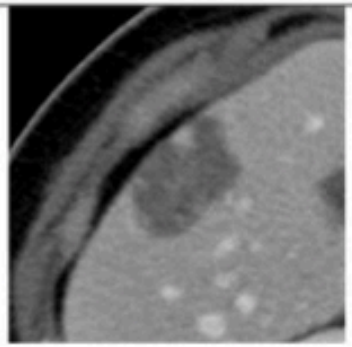
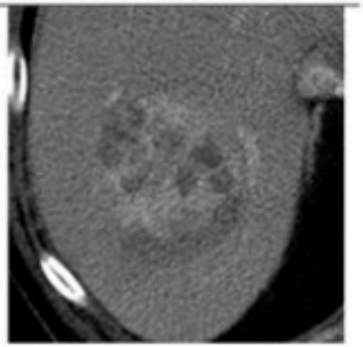


Higher-order statistics

- Semantic Features
- Computational imaging features

Adapted from : Incoronato et al. *Radiogenomic Analysis of Oncological Data: A Technical Survey.* (2017)

SEMANTIC FEATURES: EXAMPLES FROM LIVER LESIONS

Lesion type	Cyst	Metastasis	Hemangioma	Hepatocellular Carcinoma
Feature				
Portal Venous Phase Image				
Shape	-round	-round	-ovoid	-irregularly shaped
Margin	-smooth	-lobulated	-lobulated	-poorly defined
Density	-hypodense -water density	-hypodense -soft tissue density	-mixed density	-mixed density
Enhancement	-nonenhancing	-enhancing -homogeneous	-enhancing -peripheral nodular	-enhancing -heterogeneous

- Compiled by **experts of the field.**
- Based on **application-specific**, standardized lexicons.
- Standard Ontologies for **data comparison**: Imaging Reporting And Data System (RADS). Available for Breast, Prostate, Liver cancer.

COMPUTATIONAL IMAGING FEATURES: AN OVERVIEW

CATEGORY	AIMED AT	MEASURE OR METHOD
SHAPE FEATURES	CAPTURING NUMERICAL INFORMATION REGARDING **GEOMETRIC PROPERTIES** OF THE ROI	**For 2D ROIs**: perimeter, area, circularity, major-minor axis ratio, compactness, solidity, smoothness, eccentricity.
		For 3D ROIs: surface area, volume, surface-to-volume ratio, sphericity.
INTENSITY FEATURES (FIRST-ORDER STATISTICAL DESCRIPTORS)	MEASURING **DISTRIBUTION OF INTENSITY AND/OR COLOR** VALUES IN THE ROI, IGNORING THE SPATIAL RELATIONSHIPS (**HISTOGRAM-BASED** MEASURES)	**Central tendency**: mean, median, and mode.
		Variation: standard deviation, variance, minimum, and maximum.
		Randomness: entropy, uniformity, and energy.
		Asymmetry: skewness, geometric and color asymmetry.
		Sharpness: kurtosis.

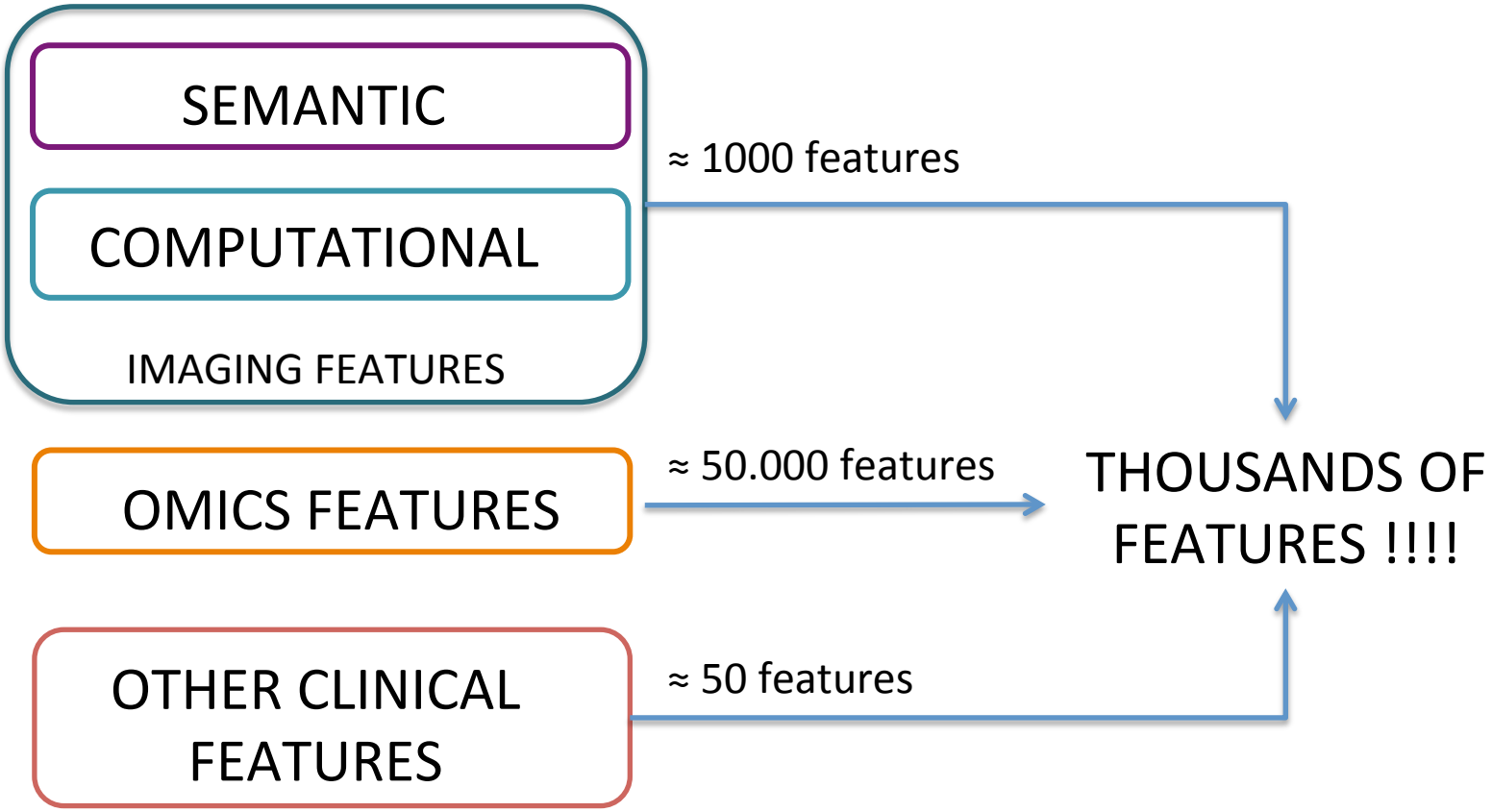
COMPUTATIONAL IMAGING FEATURES: AN OVERVIEW

CATEGORY	AIMED AT	MEASURE OR METHOD
TEXTURE FEATURES (SECOND-ORDER STATISTICAL DESCRIPTORS)	EXTRACTING **STATISTICAL RELATIONSHIP**S AMONG ROI ELEMENTS BASED ON SIMILAR/DIFFERENT **CONTRAST VALUES**.	**Gray-Level Co-occurrence Matrix (GLCM) based measures**: entropy, contrast, correlation
		Gray-Level Run-Length Matrix (GRLRM) based measures: short and long run emphasis: run length non-uniformity, low and high gray-level run emphasis, run percentage
HIGHER-ORDER STATISTICS	DISCOVERING **REPETITIVE PATTERNS**	Fractal dimensions
		Wavelets
		Gabor Filters
		Laplacian of Gaussian bandpass filter

OMICS FEATURE: AN OVERVIEW

OMICS DATA TYPE	EXTRACTED FEATURES		TYPE AND RANGE OF ASSOCIATED NUMERICAL FEATURE
Single Nucleotide Polymorphisms	Position on the chromosome		INTEGER [0,1,2]
Rare Variants			
Copy Number Variations	Read counts		INTEGER > 0
	Z-Score and Normalized Read Counts		REAL [0,1]
DNA Methylation	Peak counts		INTEGER > 0
	Normalized differential expression of peaks		REAL [0,1]
messenger RNA	Microarray	Absolute	REAL > 0
long non-coding RNA		Differential (log-transformed)	REAL [-inf,+inf]
	RNA-seq	Read Counts	INTEGER > 0
micro RNA		Normalized Read Counts	REAL [-inf,+inf]

FEATURE INTEGRATION AND ANALYSIS



FEATURE INTEGRATION AND ANALYSIS

CLASS	METHODOLOGY
SUPERVISED CLASSIFICATION	Sparse features multiple kernel learning (MKL)
REGRESSION	Linear regression
	Second level regression
	Generalized linear model
	Ordinary Least Squares (OLS) Analysis
	Canonical Correlation Analysis (CCA)
	Outcome-regularized sparse canonical correlation analysis
	Multiple regression and likelihood ratio assessment
	Lasso regression
	Cox regression
	Logistic regression and likelihood ratio assessment
UNSUPERVISED LEARNING	Hierarchical Cluster Analysis
NETWORK ANALYSIS	Gene Interaction Network Anlysis

OUR METHODOLOGY

IMAGING DATA:

- PET
- MRI
- DIFFUSION MRI

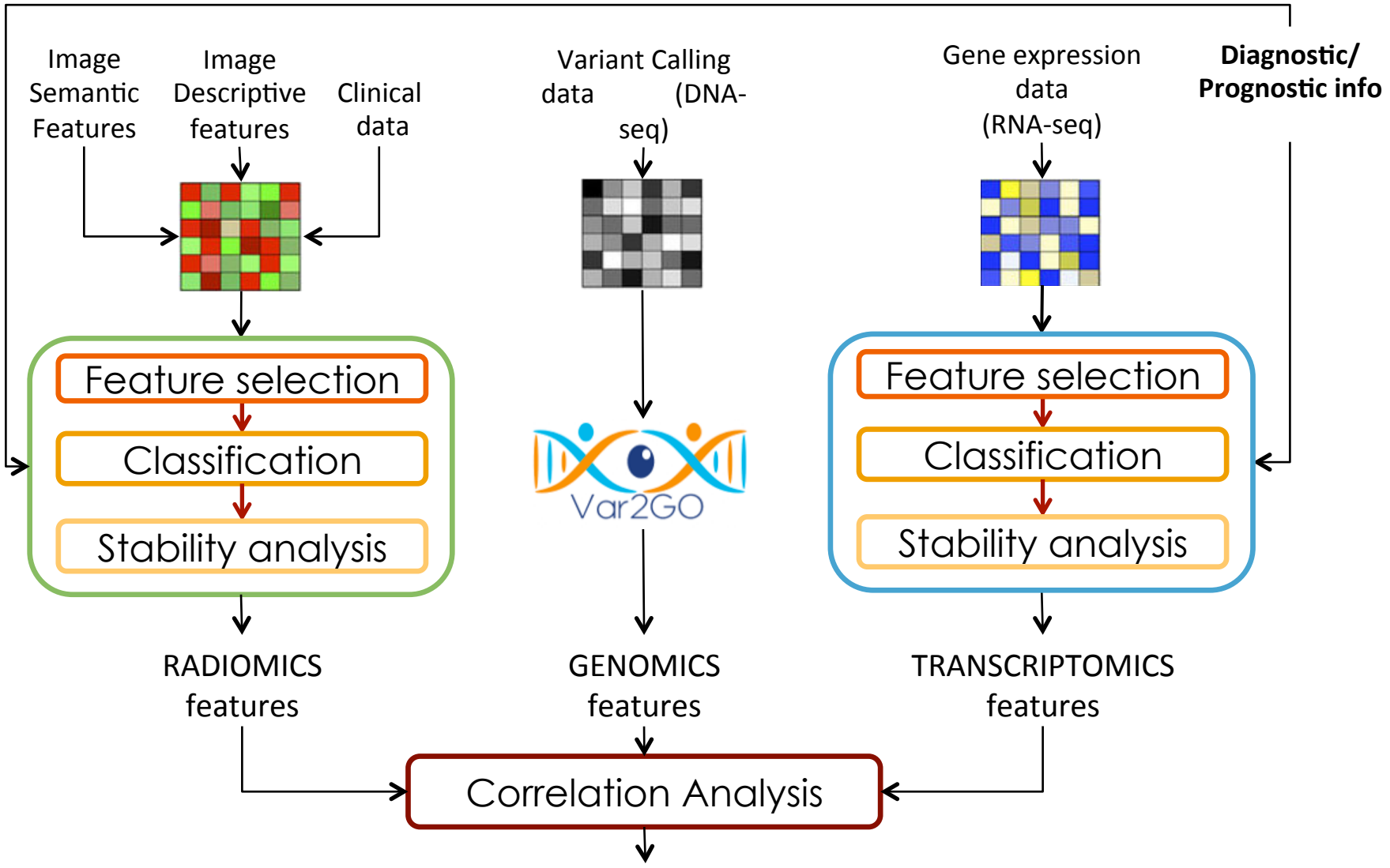
OMICS DATA:

- VARIANT CALLING
- GENE EXPRESSION

INTEGRATION AND ANALYSIS:

- FEATURE SELECTION
- SUPERVISED CLASSIFICATION
- CORRELATION ANALYSIS

FEATURE SELECTION AND CLASSIFICATION



Var2GO: a web-based tool for gene variants selection. I. Granata et al. BMC Bioinformatics, 2016.

DATA AVAILABILITY

Integrated, curated and anonymised data are **essential** to design and test **new** *omics imaging* methodologies.

DATA AVAILABILITY

Integrated, curated and anonymised data are **essential** to design and test **new** *omics imaging* methodologies.

HOWEVER

Annotated *omics imaging* data are **usually not freely available** to the scientific community!

Main reasons:

- the **privacy** of the patients/study partecipants
- the **costs** of collecting and storing data

THE ENIGMA CONSORTIUM APPROACH

The **ENIGMA** consortium (*Enhancing NeuroImaging Genetics through Meta-Analysis*) relies on the following approach:

- ENIGMA enforces the usage of **common, harmonized data analysis and meta-analysis protocols** that are **freely** available to the collaborators.
- **Each partner** processes his own data, providing only the **resulting meta-analyses** that are thoroughly **comparable** since have been obtained using the **same protocols.**

RESULT INTEGRATION

OK IF: YOU HAVE YOUR OWN MULTIMODAL DATA, AND ADDRESS A DISEASE/ TRAIT IN **COMMON** WITH OTHERS IN THE CONSORTIUM

DATA AVAILABILITY: MOTIVATED REQUEST

Several consortia (such as COGNOMICS) disclose their data only after the presentation of a well documented research project, usually evaluated by an internal committee.

MOTIVATED REQUEST

OK IF: YOU HAVE **ALREADY A WELL-ESTABILISHED RESEARCH** DIRECTION AND EXPERIENCE IN THE OMICS IMAGING ANALYSES

DATA AVAILABILITY: FREE ACCESS

Free access to multi-modal data is provided by some institutions. Sometimes, as in the case of UK Biobank, a CV attesting the *bona fide* of the research is required.
These data are often only collected without any check of their reliability.

FREE ACCESS

OK IF: YOU WANT TO **START TESTING YOUR METHODOLOGIES ON MULTIMODAL DATA**, WITHOUT CHOICE ABOUT THE ADDRESSED PROBLEM, AND THE CONSISTENCY IN TERM OF ACQUISITION AND PREPROCESSNING PROTOCOLS OF THE DATA.

EXAMPLES OF *OMICS IMAGING* INTEGRATED RESOURCES

IMAGING REPOSITORY	ACRONYM	ACCESSIBILITY
Allen Human Brain Atlas	AHBA	**FREE**
Alzheimer's Disease Neuroimaging Initiative	ADNI	UPON MOTIVATED REQUEST
Dementias Platform United Kingdom	DPUK	UPON MOTIVATED REQUEST
Mind Clinical Imaging Consortium	MCIC	UPON MOTIVATED REQUEST
Parkinson's Progression Markers Initiative	PPMI	UPON MOTIVATED REQUEST
Pediatric Imaging, Neurocognition, and Genetics study	PING	UPON MOTIVATED REQUEST
The Cancer Genome Atlas **AND** The Cancer Imaging Archive	TCGA TCIA	**FREE**
The IMAGEN project		UPON MOTIVATED REQUEST
The NeuroIMAGE project		UPON MOTIVATED REQUEST
UKBioBank		**FREE UPON CV**

CHALLENGES AND FUTURE PERSPECTIVES

THE CURSE OF DIMENSIONALITY

- When dealing with omics imaging data, usually the number of samples is at most around the thousands. Combined features may reach hundreds of thousands!!

DEVISE EFFICIENT METHODS FOR FEATURE SELECTION, CLASSIFICATION AND REGRESSION

CHALLENGES AND FUTURE PERSPECTIVES

GOING BEYOND GWAS

- A large amount of studies combines results coming from Genome-Wide Association Studies with imaging data. However GWAS have been demonstrated to explain only a few amount of variation in the population.

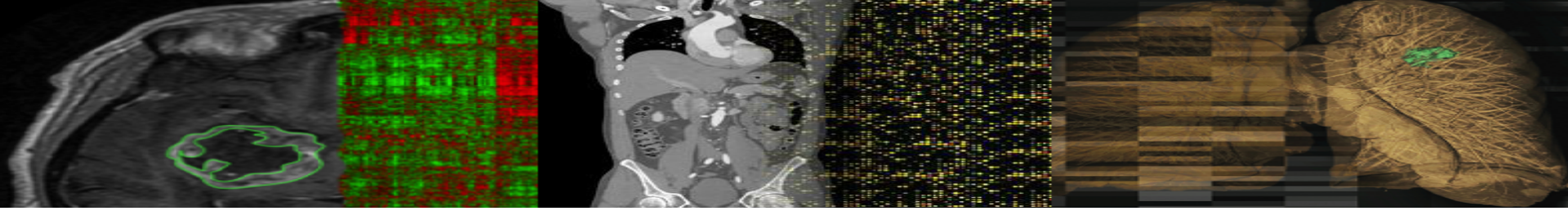
ADDRESS THE INTEGRATION OF IMAGING DATA WITH MULTIMODAL DATA COMING NOT ONLY INCLUDING DIFFERENT GENOMICS DATA (e.g. Rare Variants), BUT ALSO TRANSCRIPTOMICS AND PROTEOMICS.

CHALLENGES AND FUTURE PERSPECTIVES

- **Standardization**: involves the acquisition, analysis, pre- and post-processing of both imaging and omics data, as well as coping with the variability induced by new acquisition tools.

- **Validation**: every result achieved using a set of imaging and omics data should be validated against an independent dataset. This requires the availability of publicly available datasets representative of the problem, covering the variability of the examined data.

- **Reproducibility**: Common standards for statistical testing and reporting results, as well benchmarks for *omics imaging* studies could help to significantly enhance reproducibility.



THANKS

- Panos Pardalos, Mara Sangiovanni

ACKNOWLEDGMENTS

- Biomat Consortium, RSF, PON-MISE

mario.guarracino@cnr.it