UNIVERSITY OF CAMBRIDGE

CANCER RESEARCH UK | CAMBRIDGE CENTRE

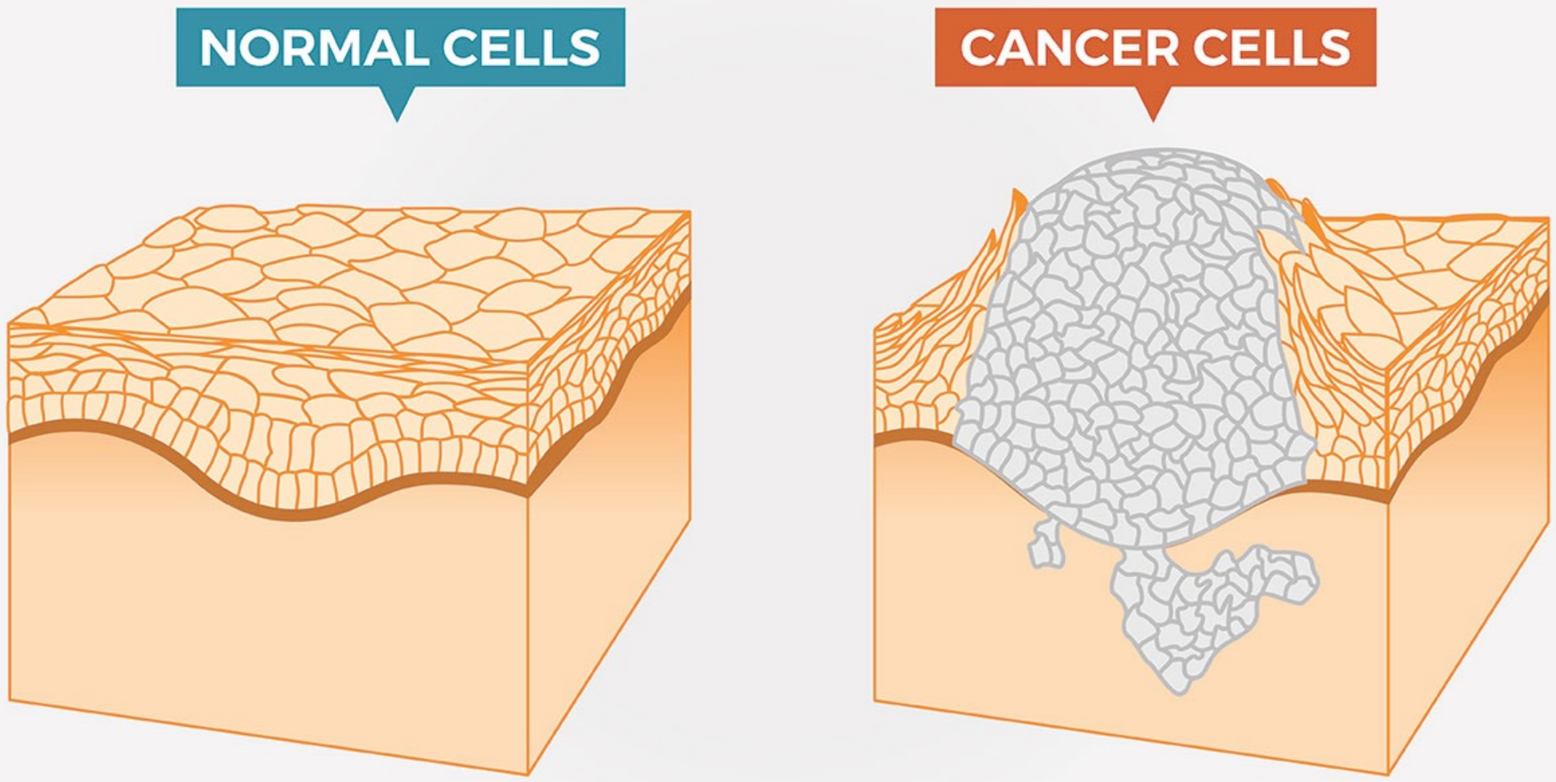
CRISPINLAB
Computational Cancer Medicine

ML, multi-omics and oncology

Mireia Crispin-Ortuzar
Department of Oncology, University of Cambridge
crispinlab.com

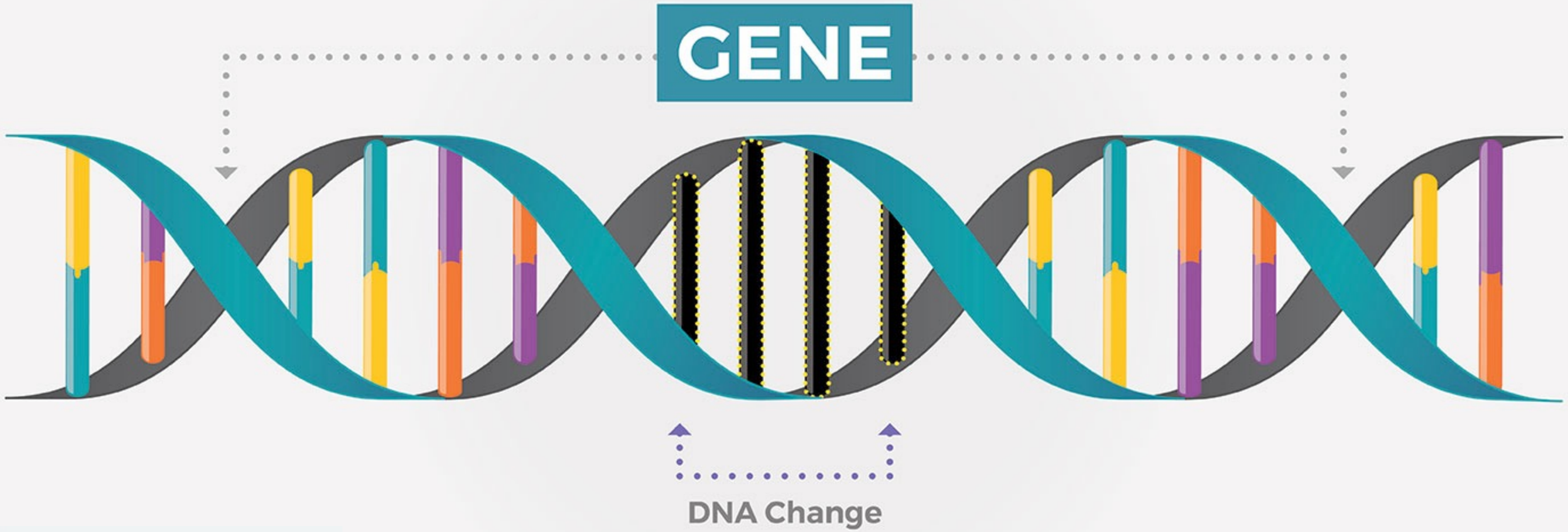
OxML Summer School
10 August 2022

How does cancer form?



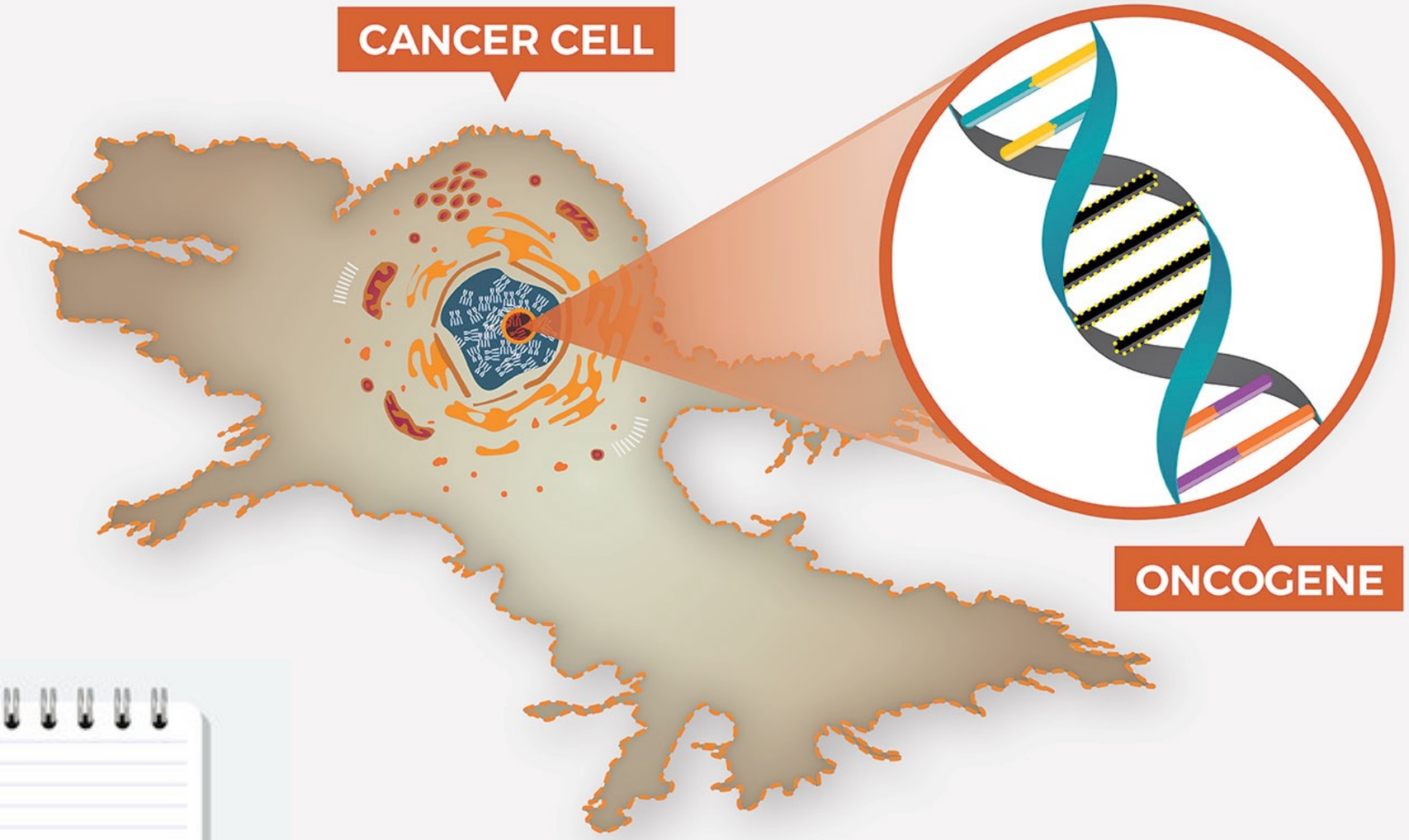
cancer.gov

What causes cancer?



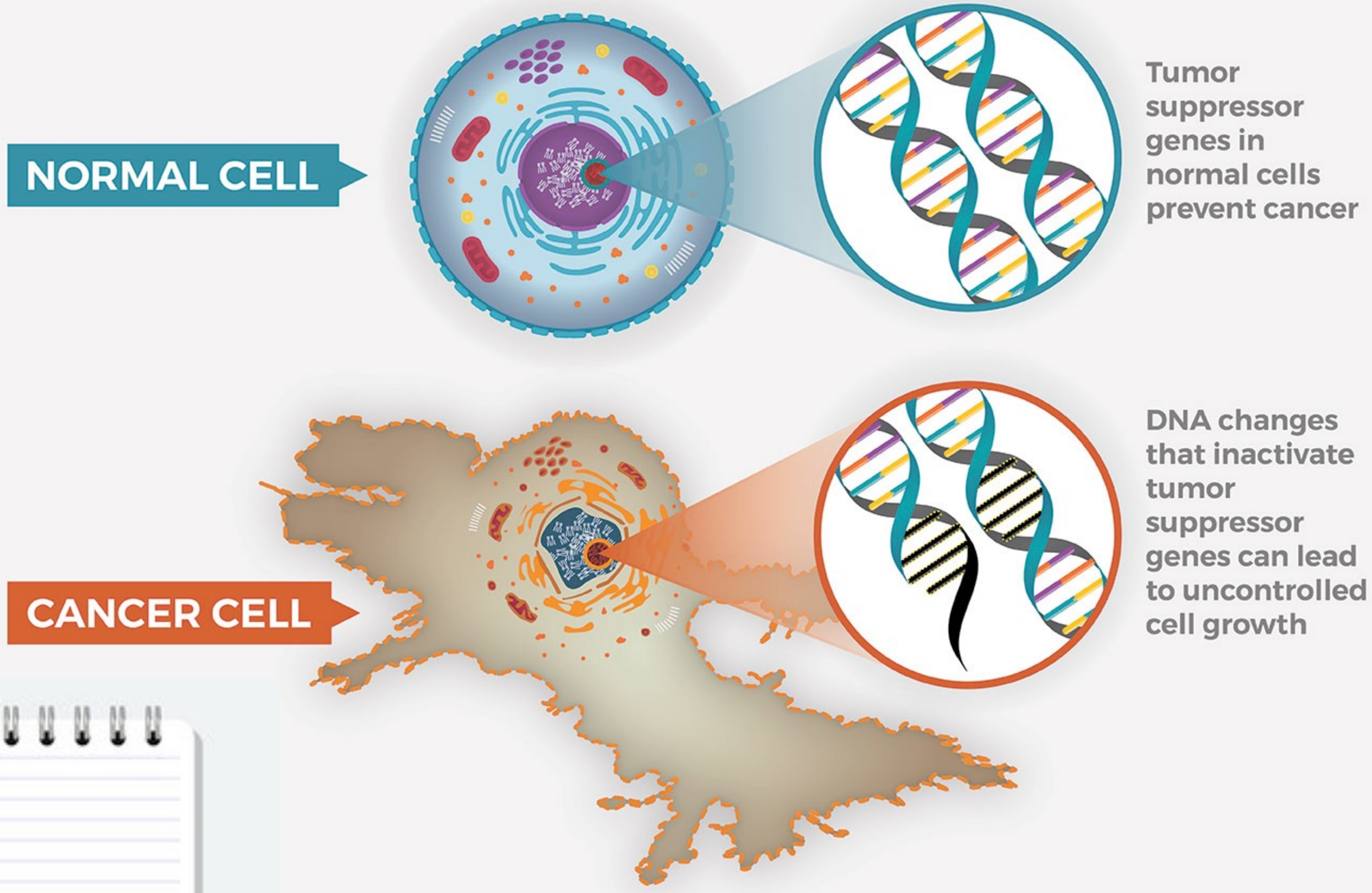
cancer.gov

What are oncogenes?



cancer.gov

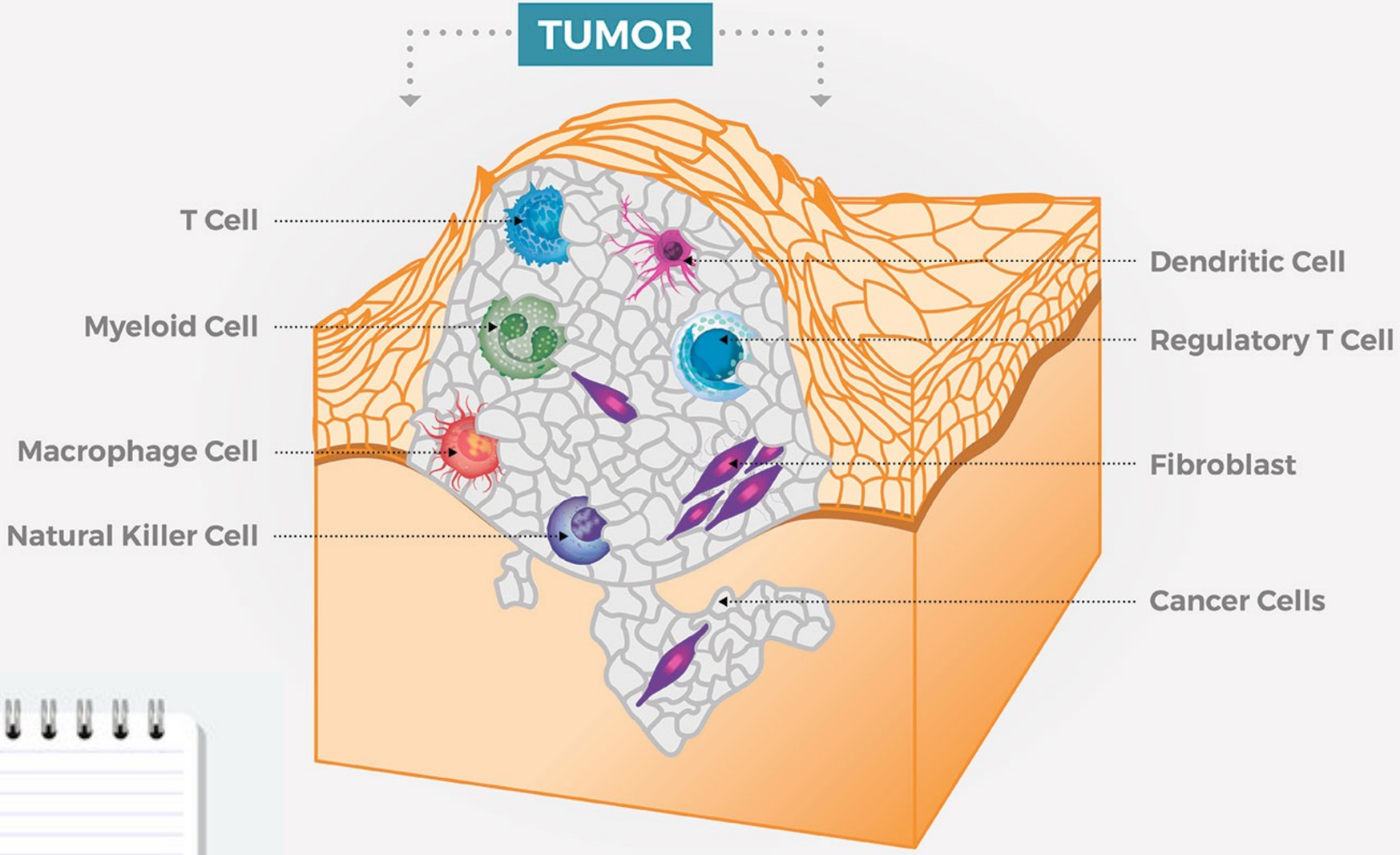
What are tumour suppressor genes?



Genes
Cell morphology

cancer.gov

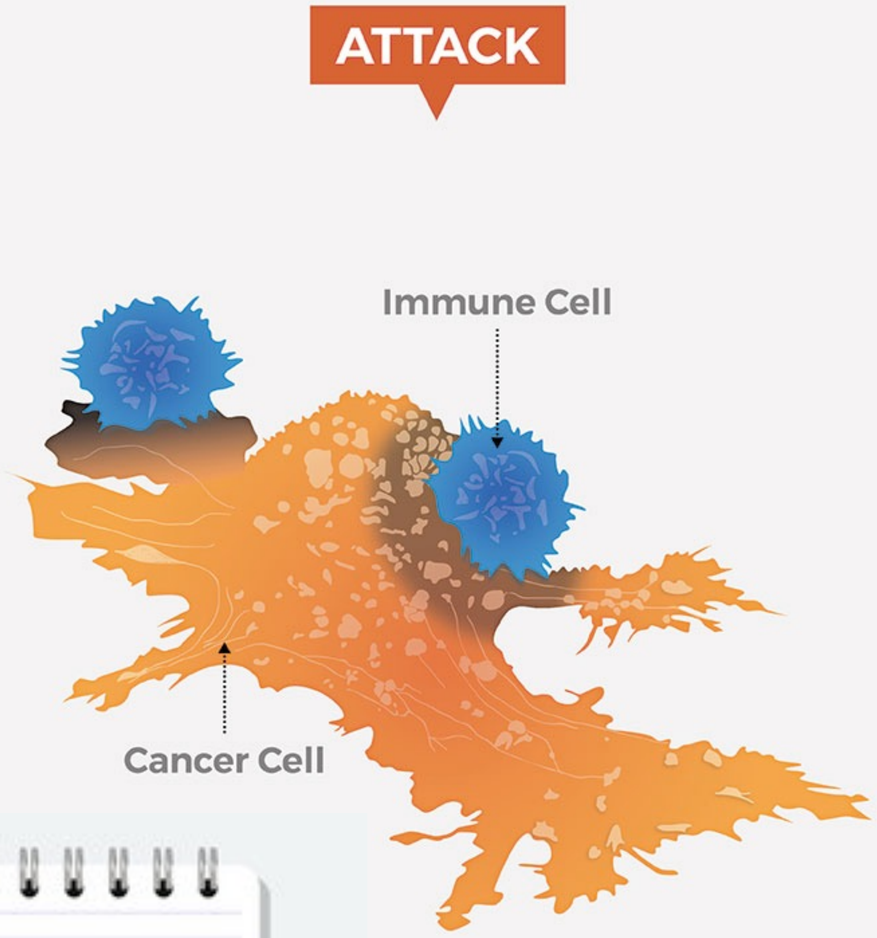
What is the tumour microenvironment?

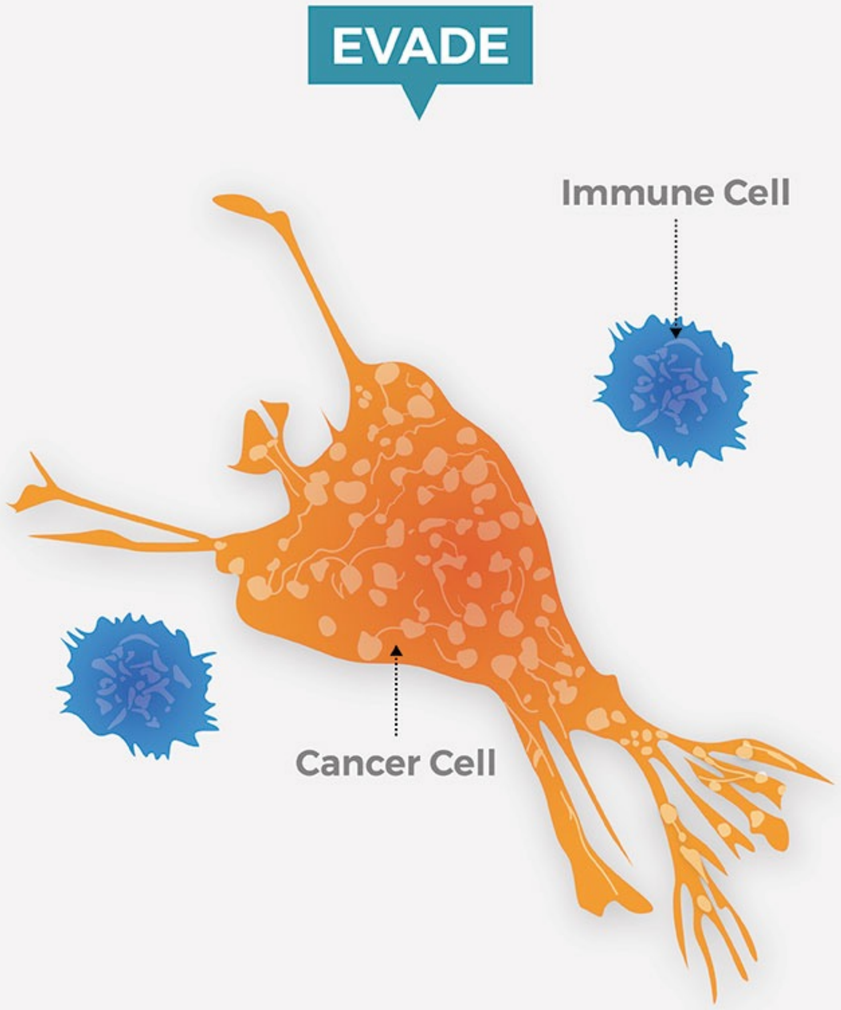


Genes
Cell morphology
Tissues

cancer.gov

How does the immune system interact?

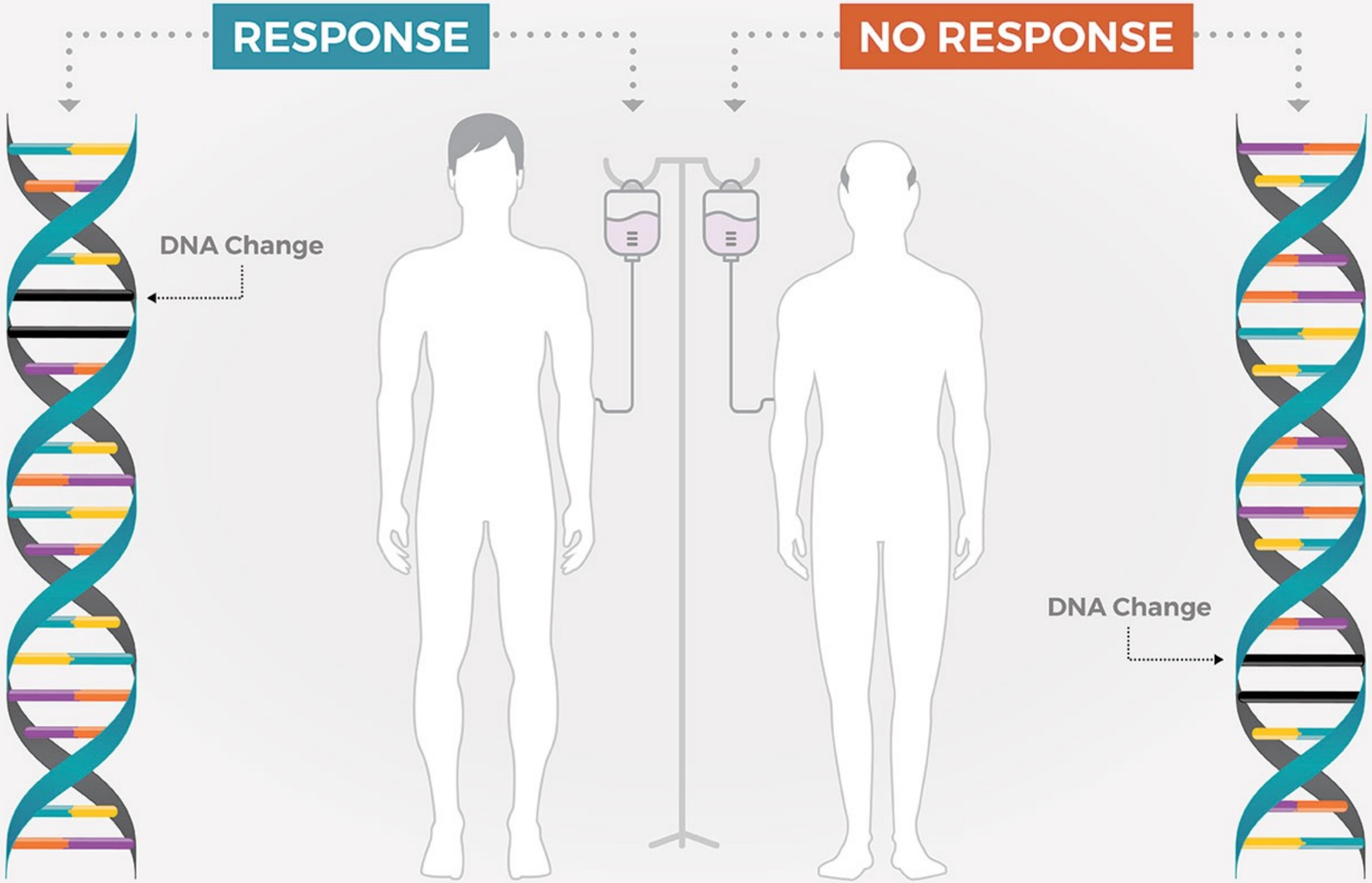




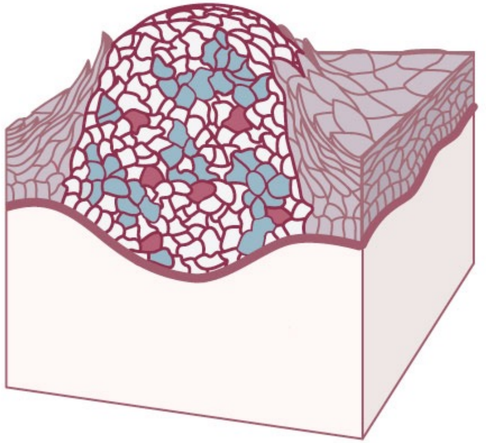
Genes
Cell morphology
Tissues
Cell function

cancer.gov

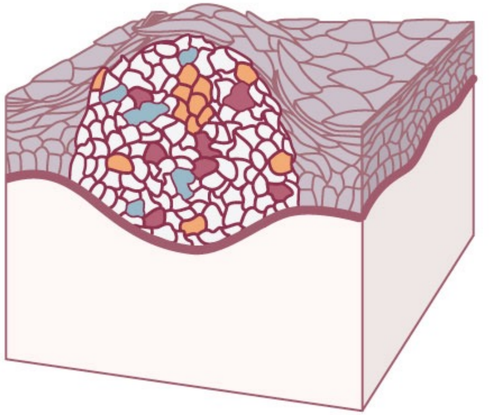
How do genetic changes affect treatment?



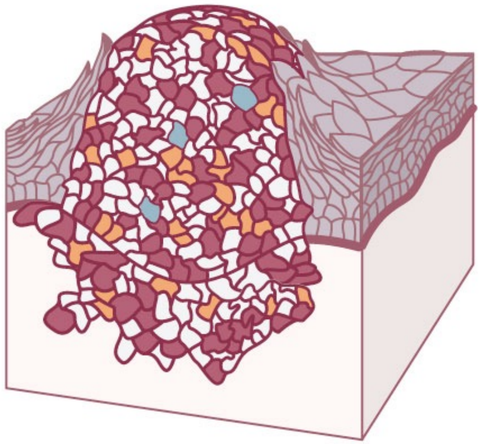
cancer.gov



Before Treatment
Tumors consist of cancer cells with different molecular features, which may make them sensitive or resistant to different types of treatments.

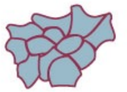


Responding to Treatment
Although, a drug may kill some cancer cells (the sensitive cells), a subset of them almost invariably survives (the resistant cells).



Developing Drug Resistance
The cancer cells that are resistant will multiply, contributing to the re-growth of the tumor.

KEY



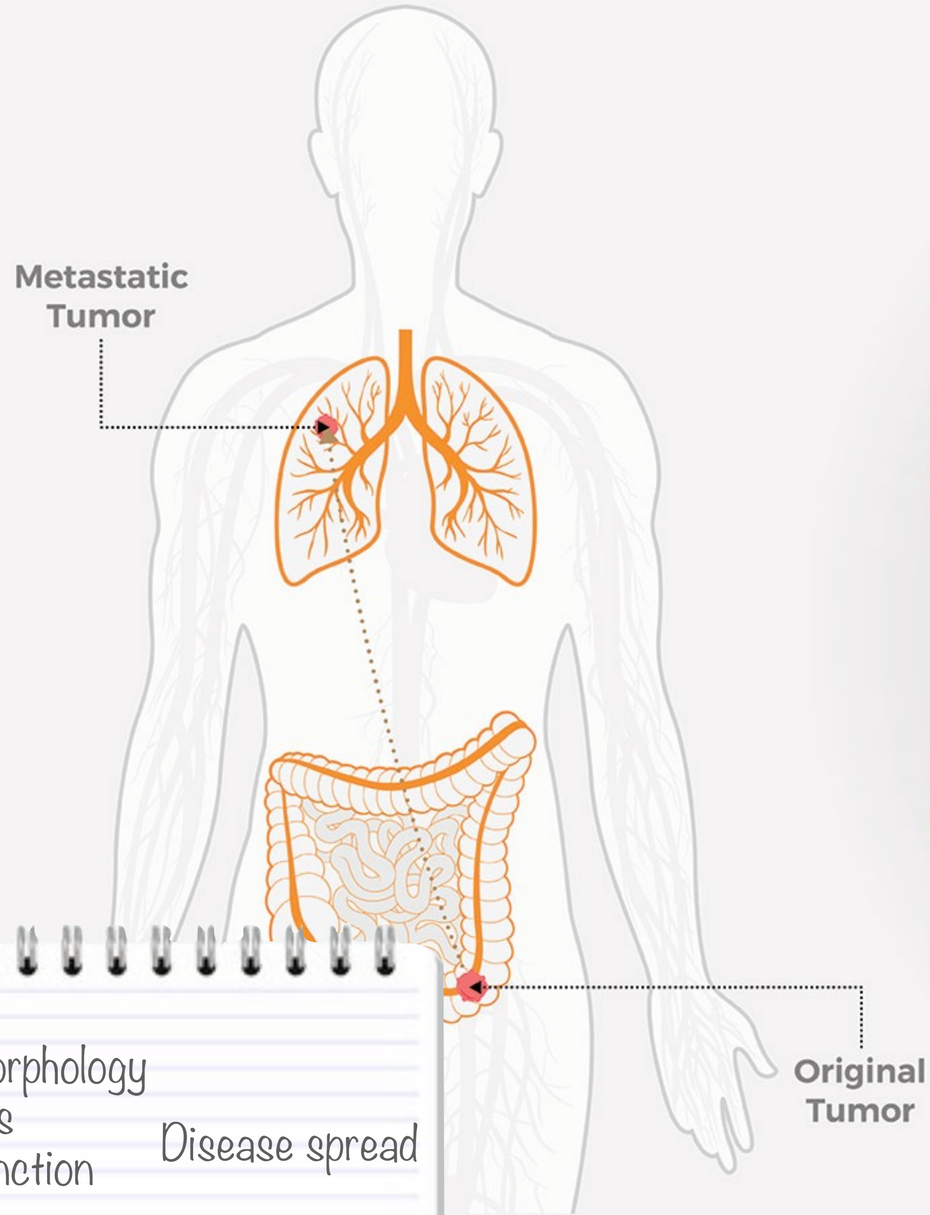
Sensitive cancer cells

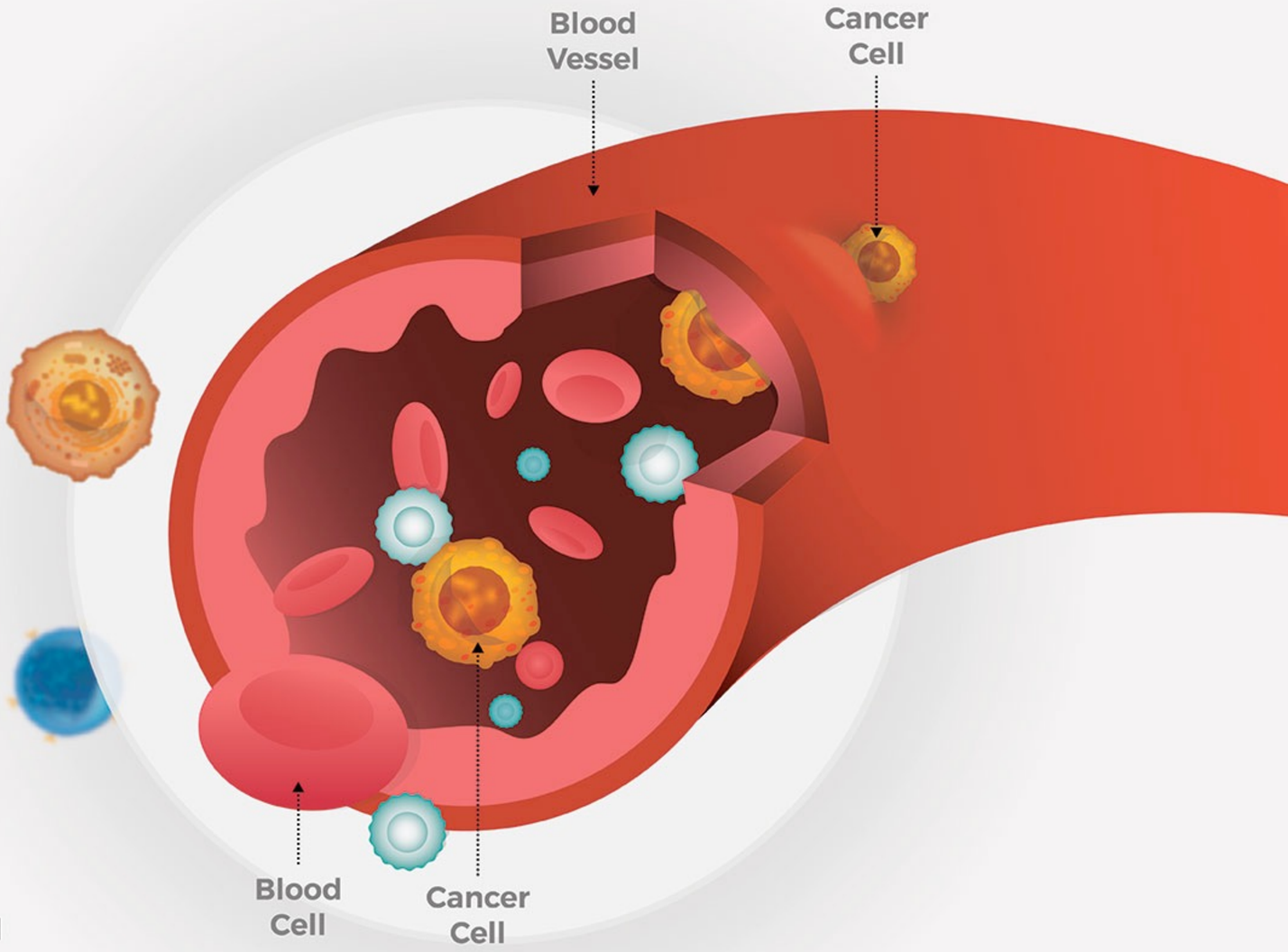
Resistant cancer cells

Therapy-induced resistant cancer cells

cancer.gov

What is metastasis?





Genes
Cell morphology
Tissues
Cell function

Disease spread

cancer.gov

Multi-omics

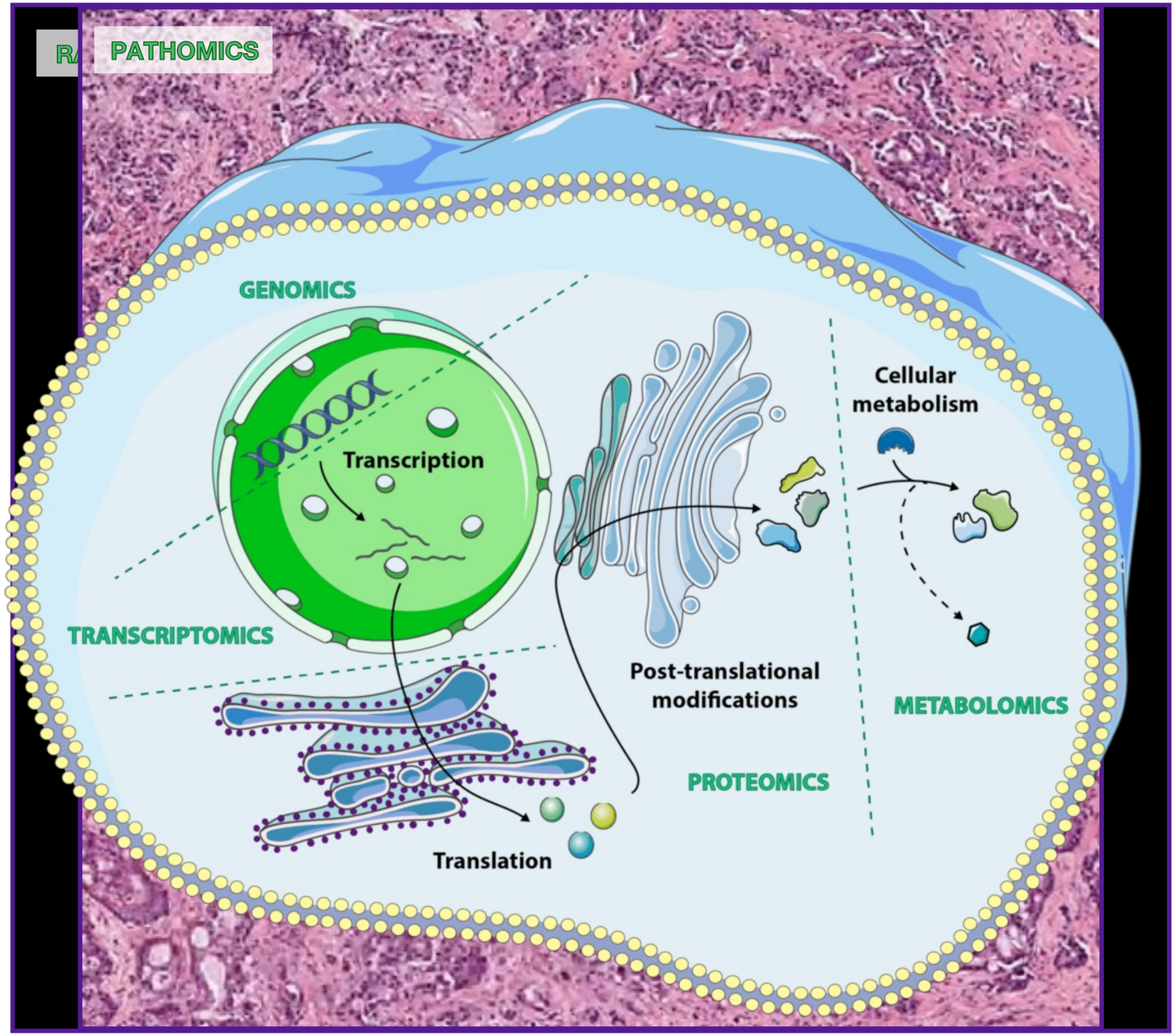
Genes

Cell function

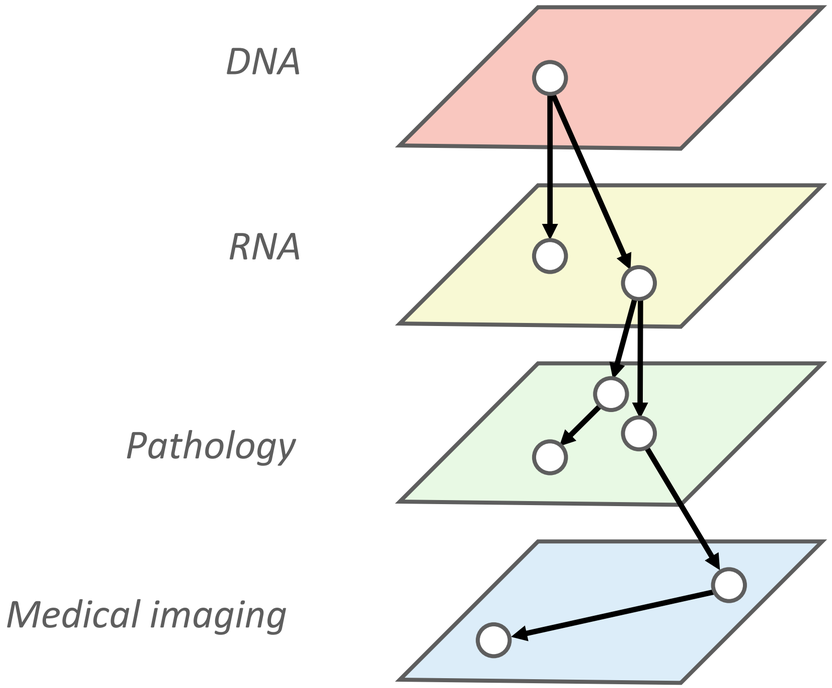
Cell morphology

Tissues

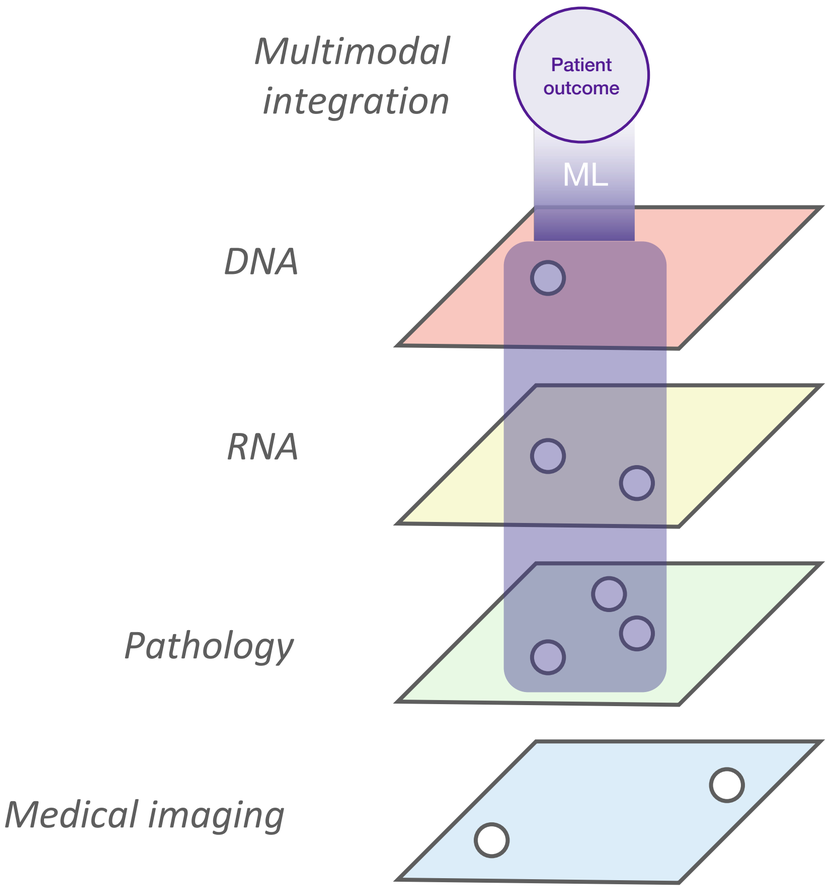
Disease spread



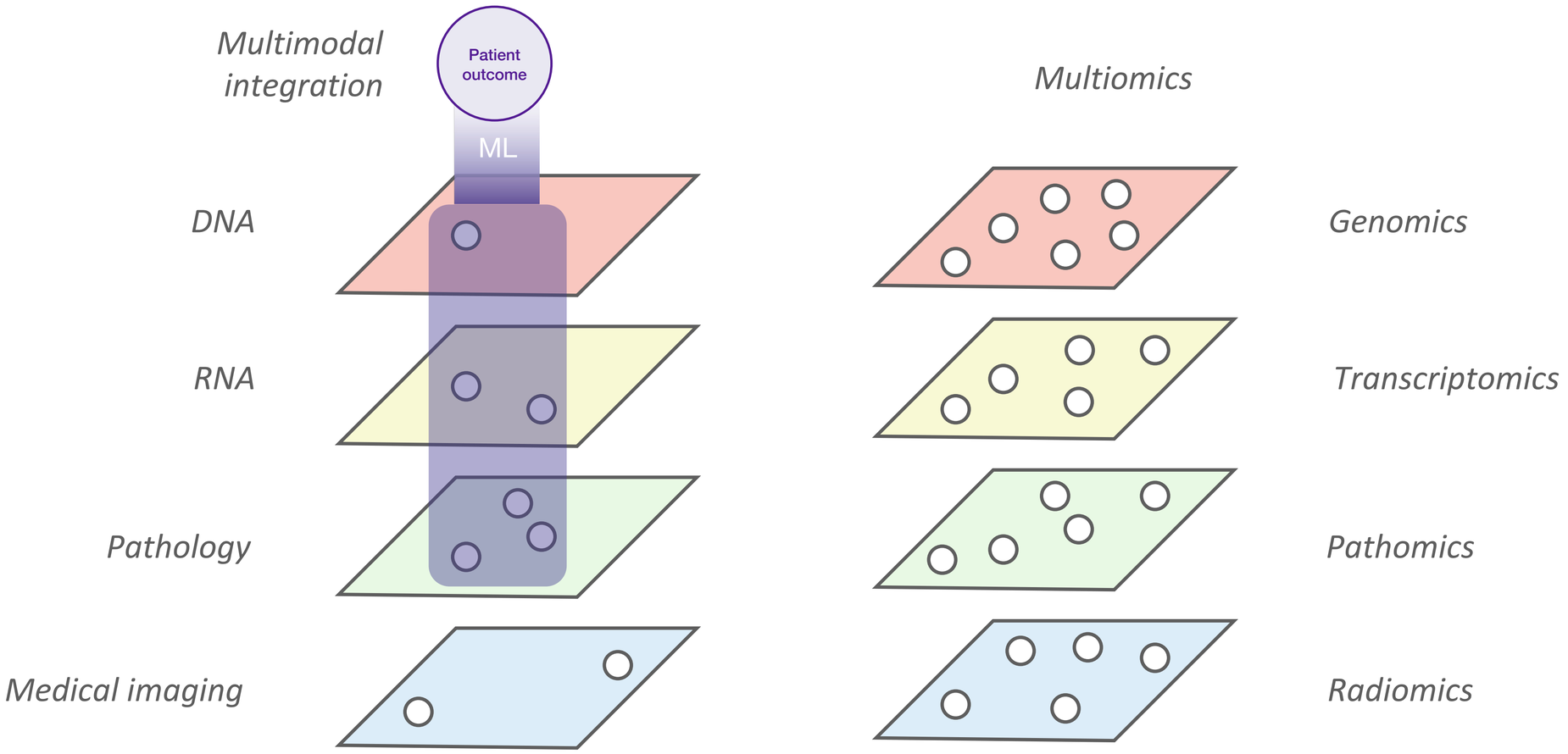
Multi-omics



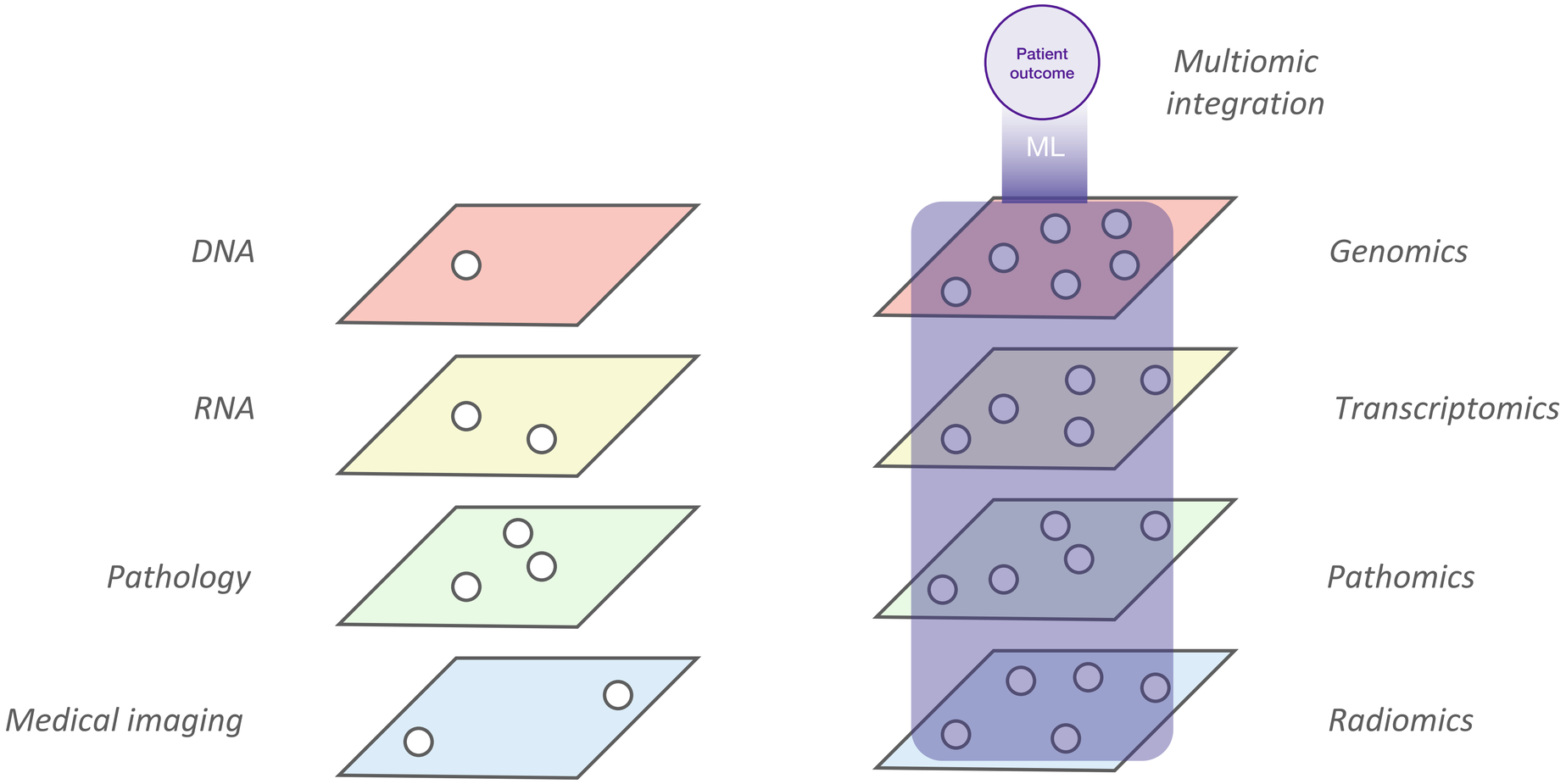
Multi-omics


Multimodal integration
Patient outcome
ML
DNA
RNA
Pathology
Medical imaging

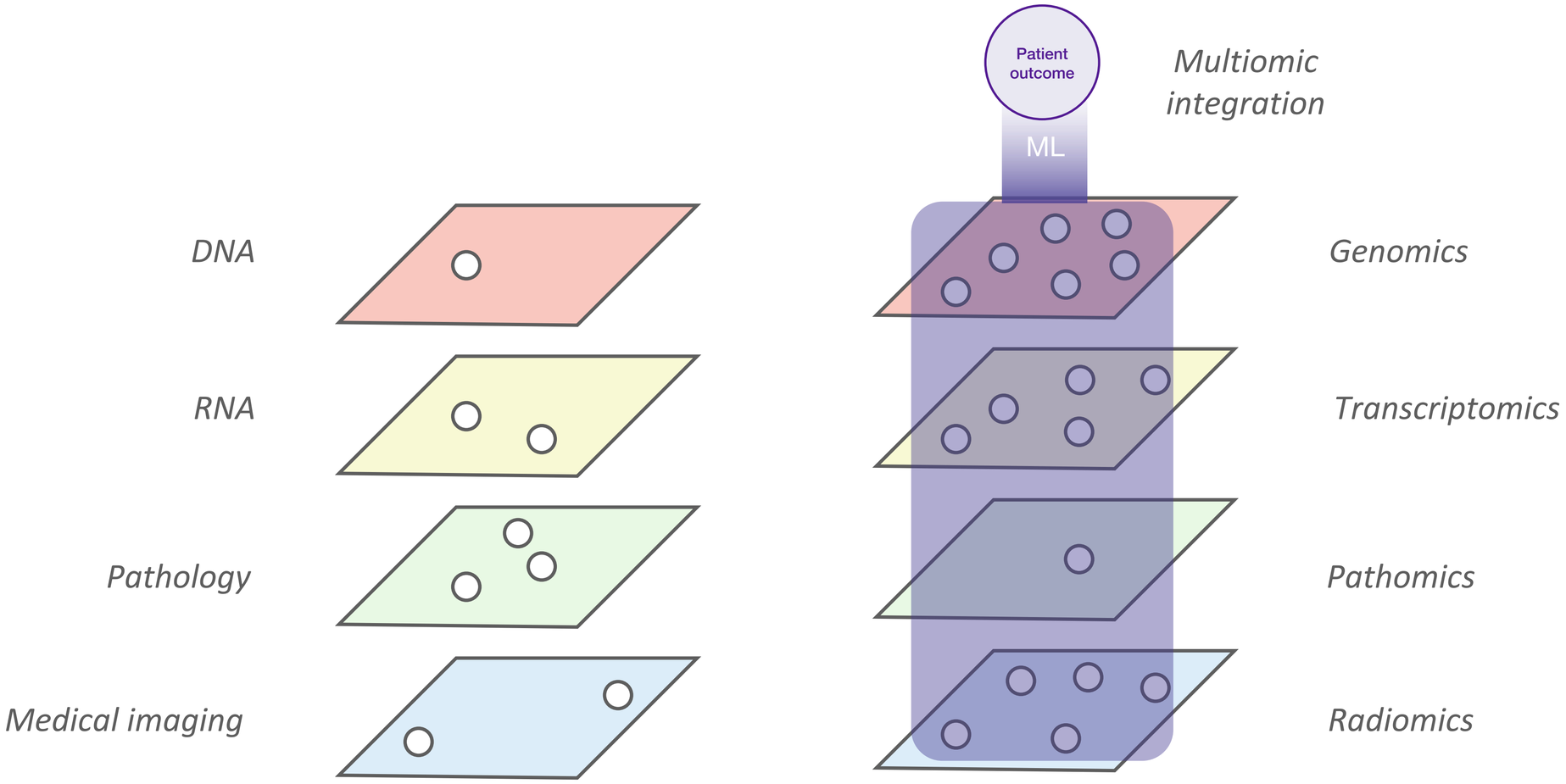
Multi-omics


Multimodal integration
Patient outcome
ML
DNA
RNA
Pathology
Medical imaging
Multiomics
Genomics
Transcriptomics
Pathomics
Radiomics

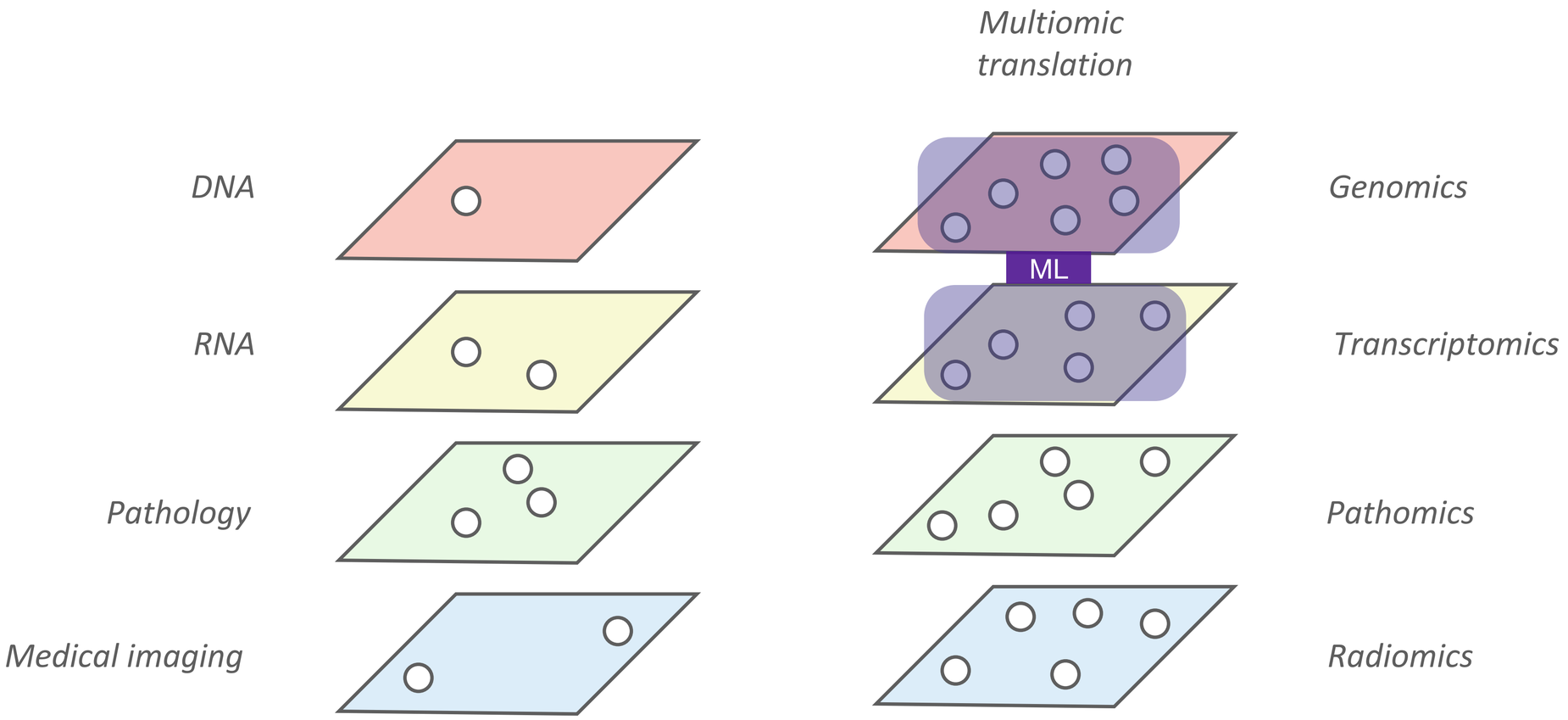
Multi-omics


Patient outcome
Multiomic integration
ML
DNA
RNA
Pathology
Medical imaging
Genomics
Transcriptomics
Pathomics
Radiomics

Multi-omics

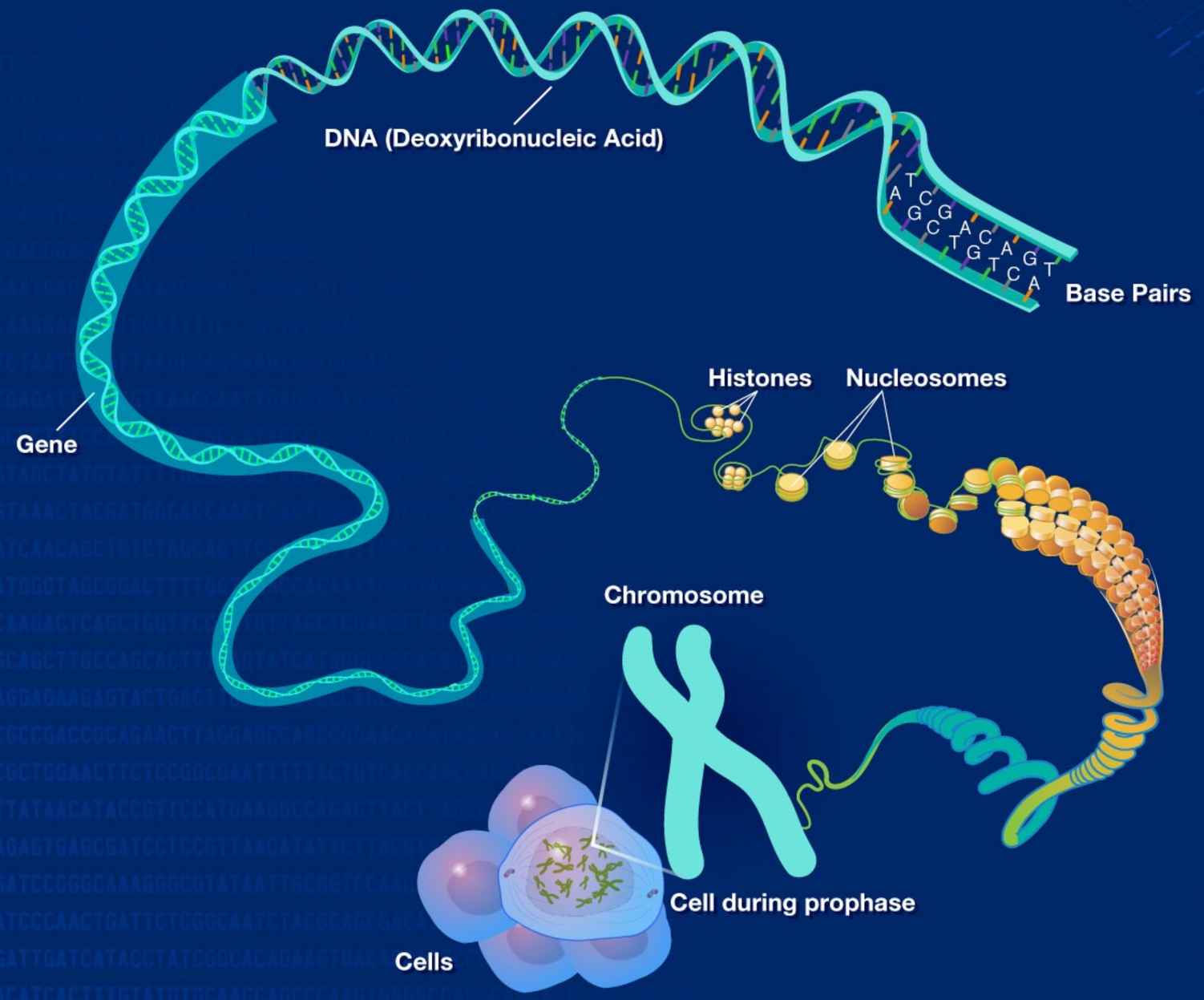

Patient outcome
Multiomic integration
ML
DNA
RNA
Pathology
Medical imaging
Genomics
Transcriptomics
Pathomics
Radiomics

Multi-omics


Multiomic translation
DNA
RNA
Pathology
Medical imaging
ML
Genomics
Transcriptomics
Pathomics
Radiomics

Genome

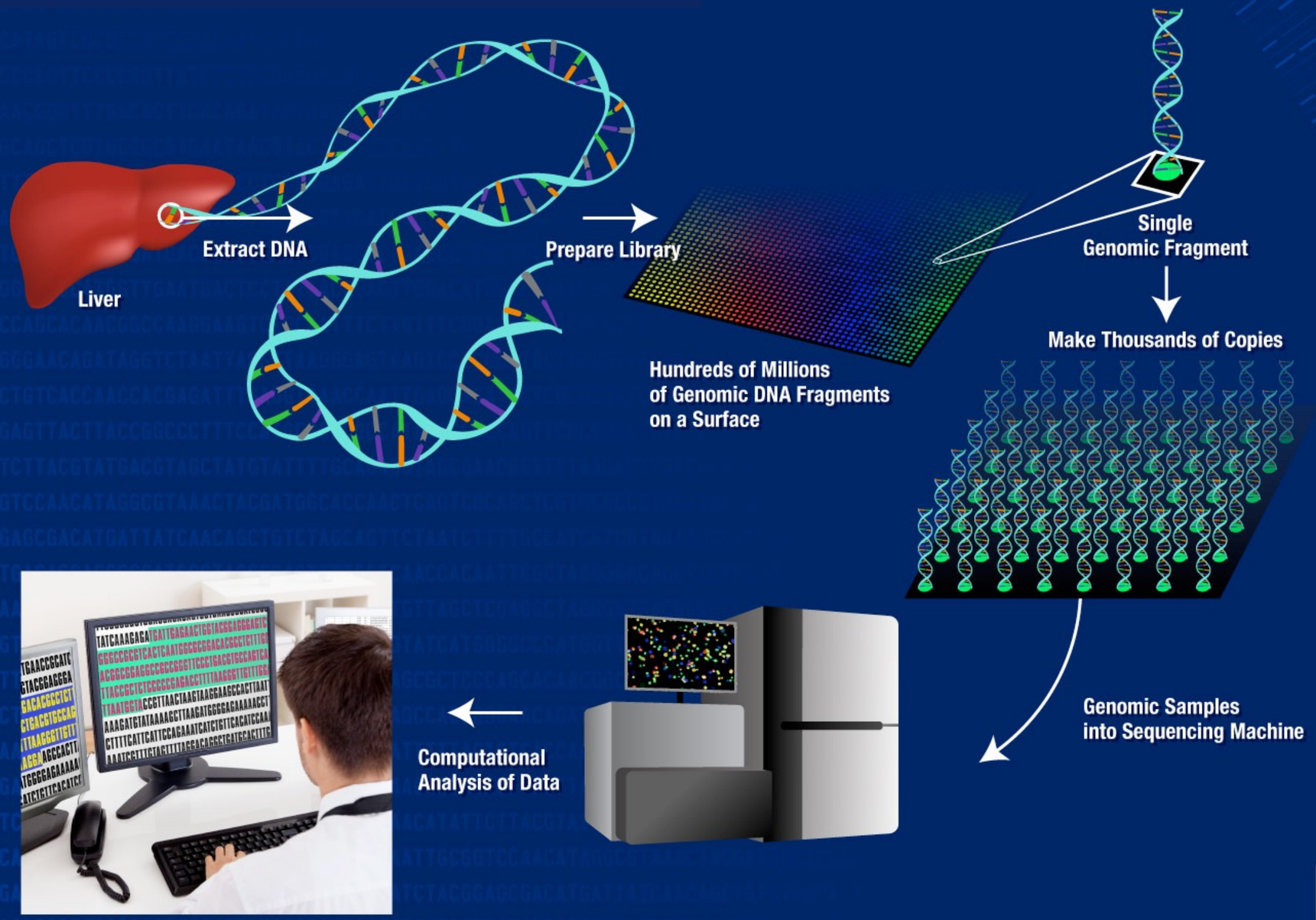
NHGRI FACT SHEETS
genome.gov



National Human Genome Research Institute

Genomic sequencing

NHGRI FACT SHEETS
genome.gov

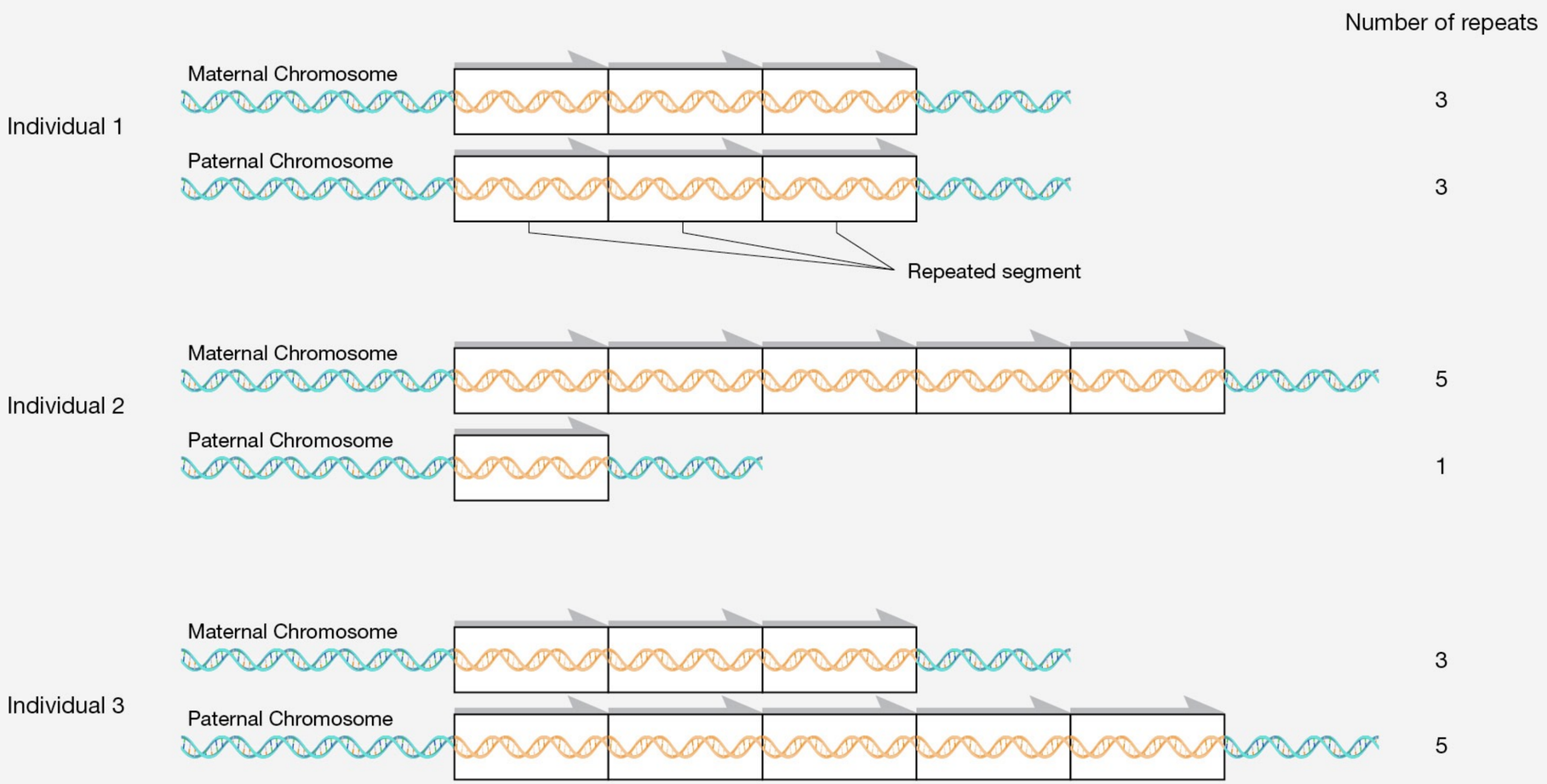


National Human Genome Research Institute

Point mutations

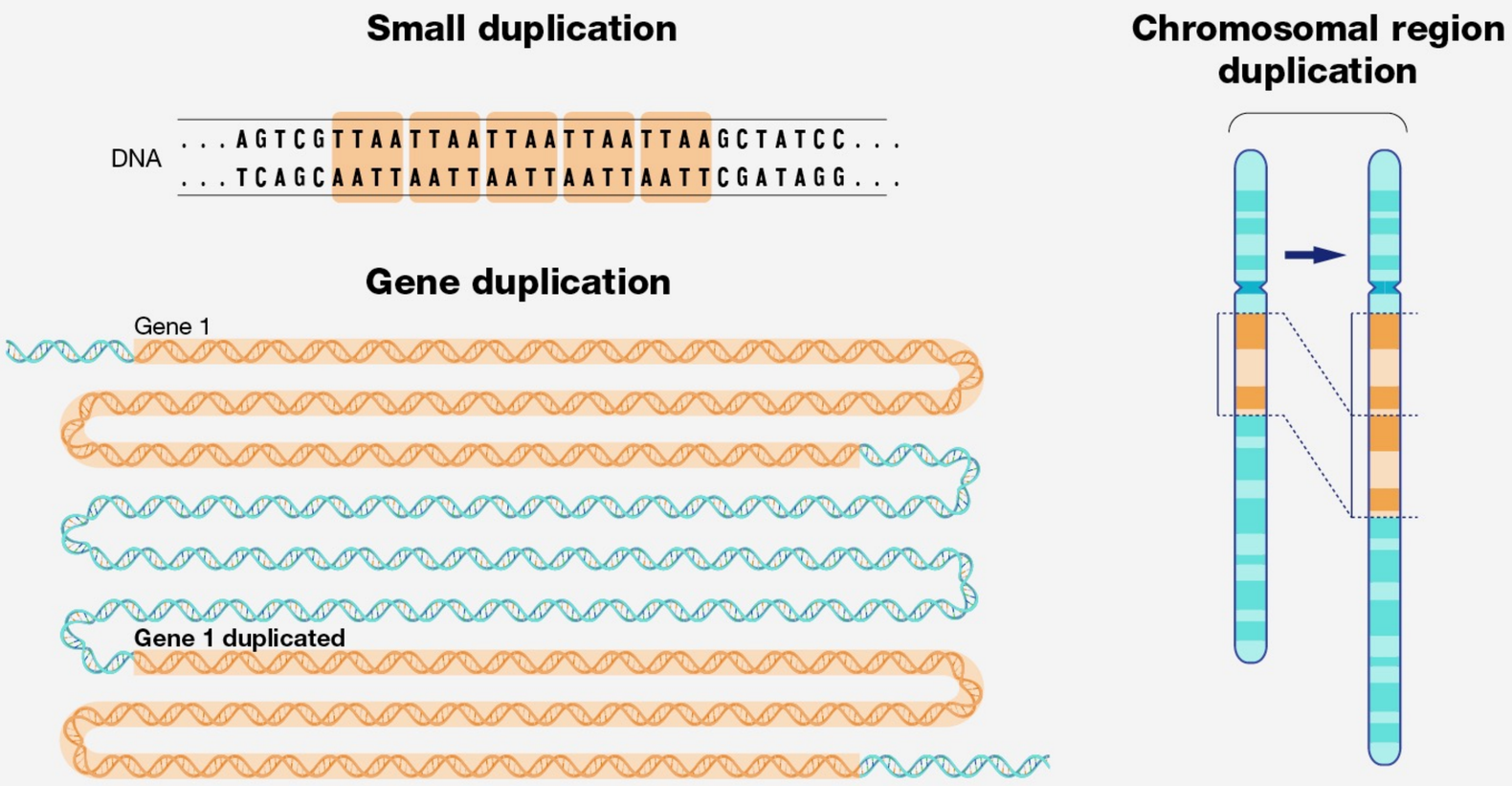
National Human Genome Research Institute

Copy number alterations



National Human Genome Research Institute

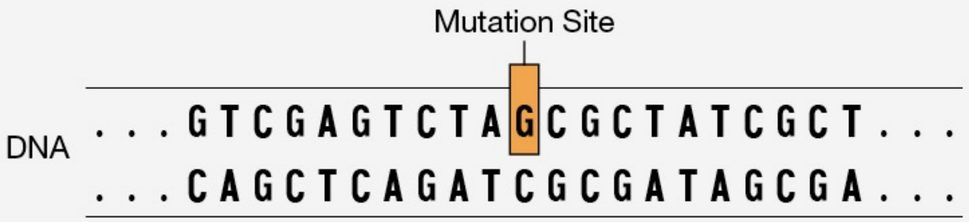
Whole genome duplication



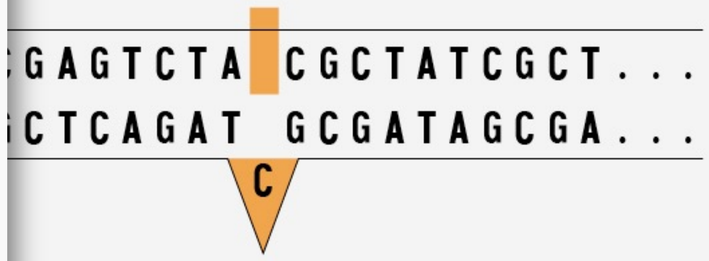
National Human Genome Research Institute

Deletions (& amplifications)

Microdeletion

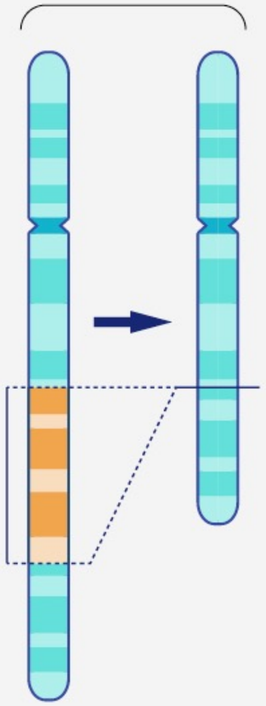


Deletion



	Sample 1	Sample 2	Sample 3
GENE 1	0	0	1
GENE 2	1	0	0
WGD	0	1	0
CNSig1	0	1	0
CNSig2	1	0	0
TMB	0	1	1

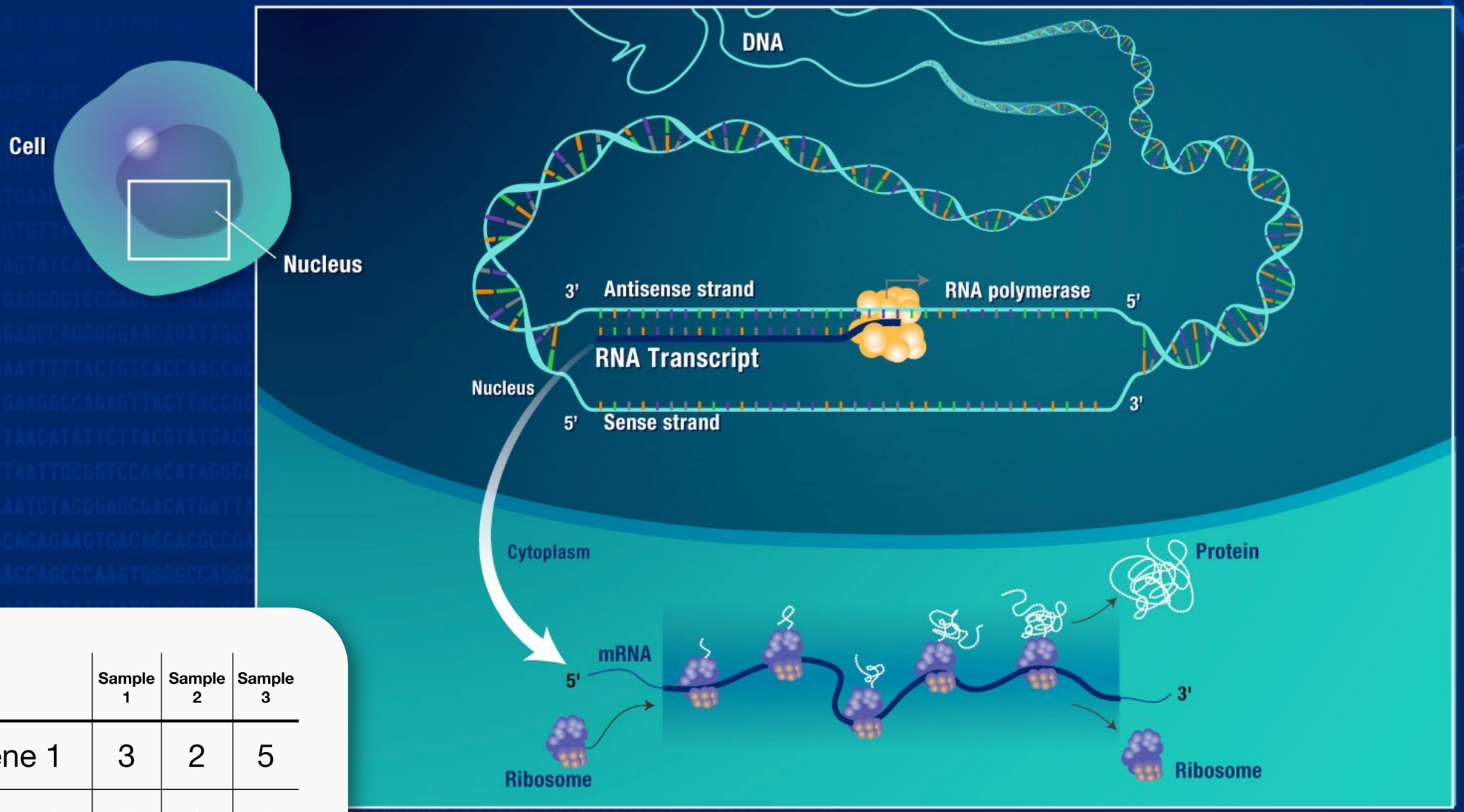
Macrodeletion



National Human Genome Research Institute

Transcriptome

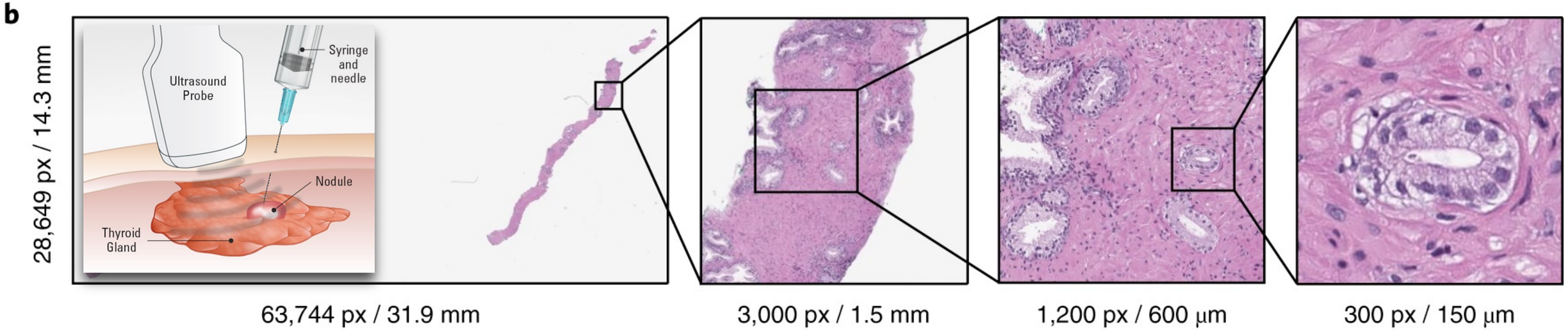
NHGRI FACT SHEETS
genome.gov



	Sample 1	Sample 2	Sample 3
Gene 1	3	2	5
Gene 2	1	9	4
Gene set 1	7	1	6

NIH National Human Genome Research Institute

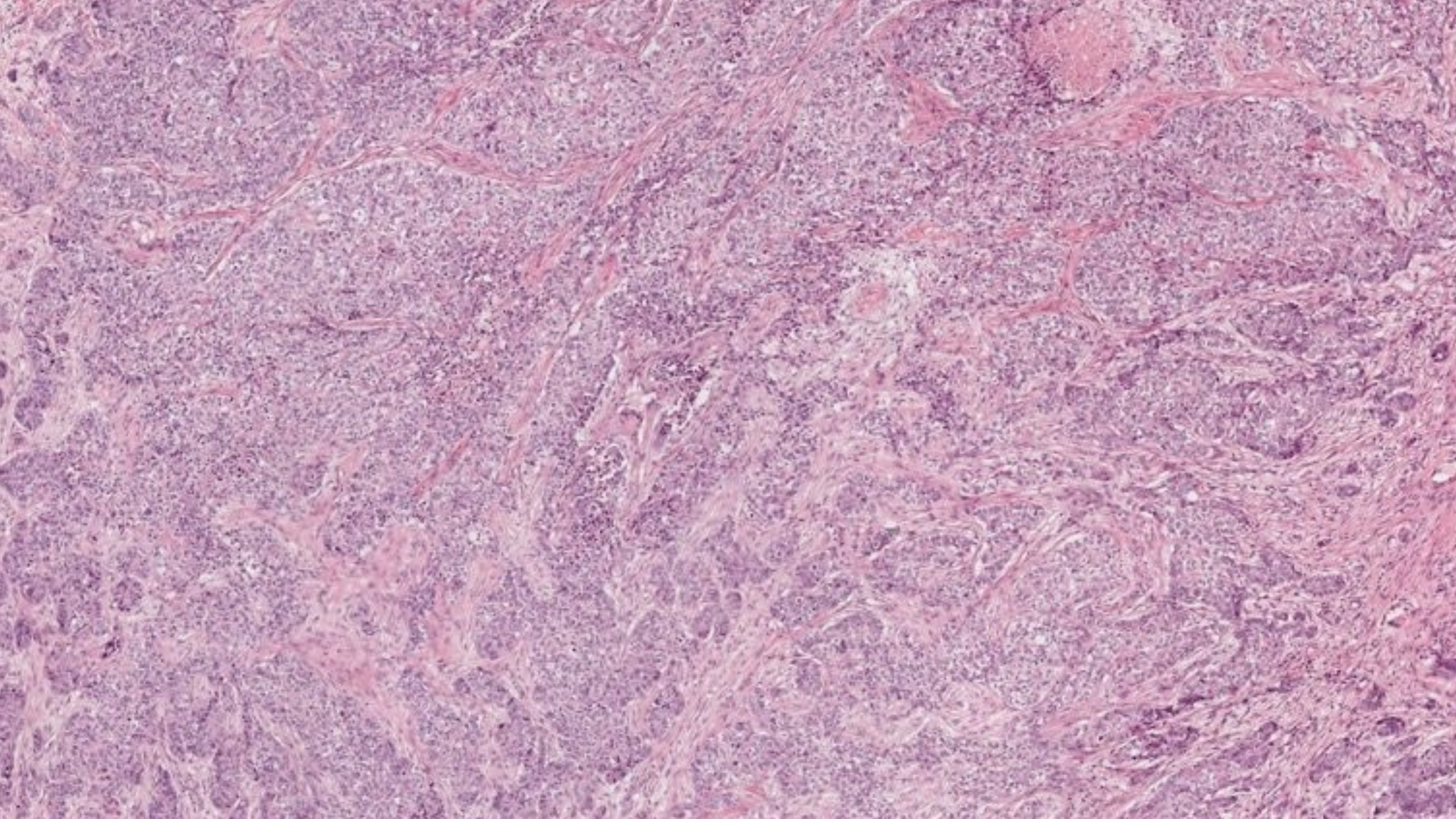
Pathomics



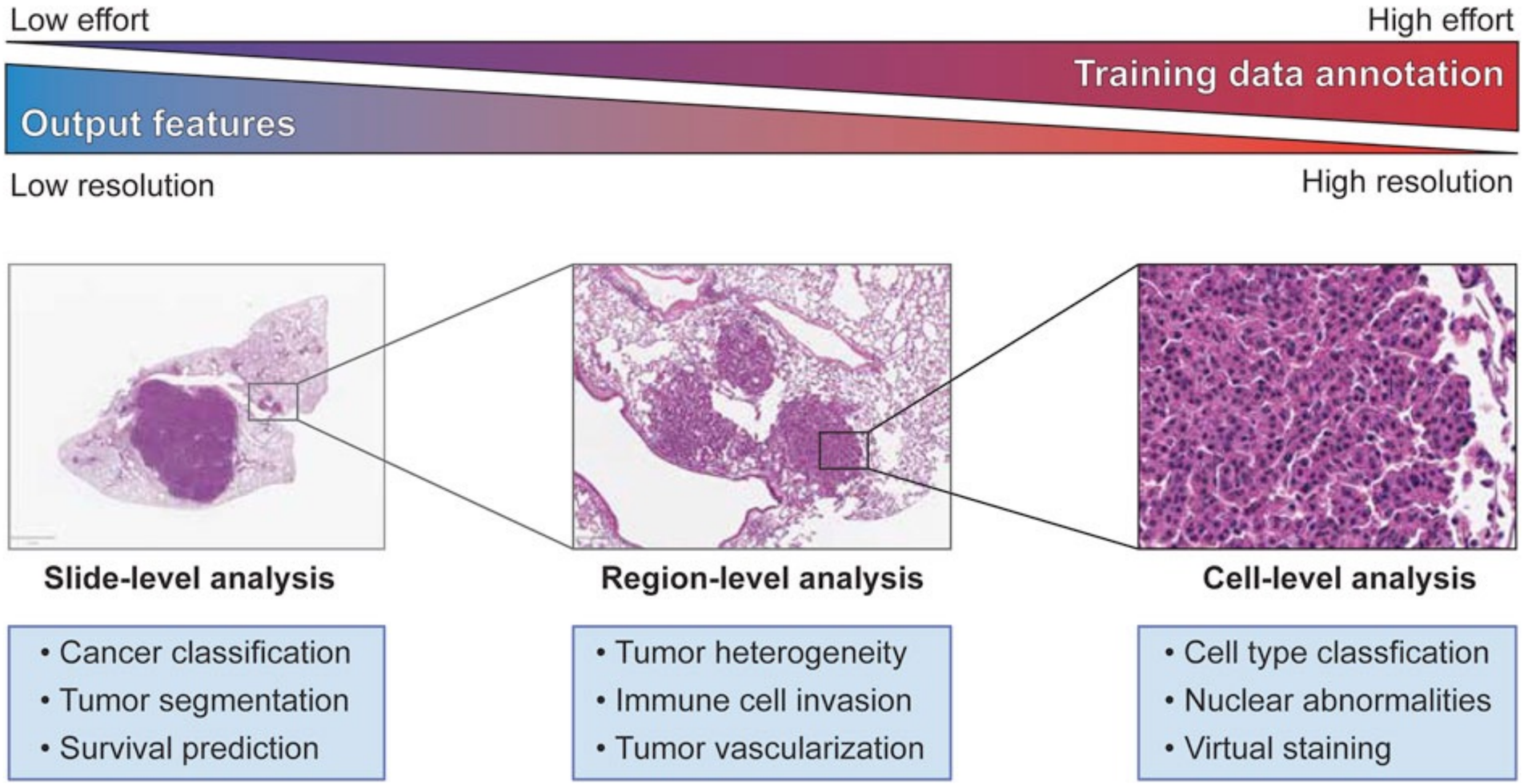
Learn more

G. Campanella et al., Nature Medicine 2019

Image from Thyroid Cancer Canada



Pathomics

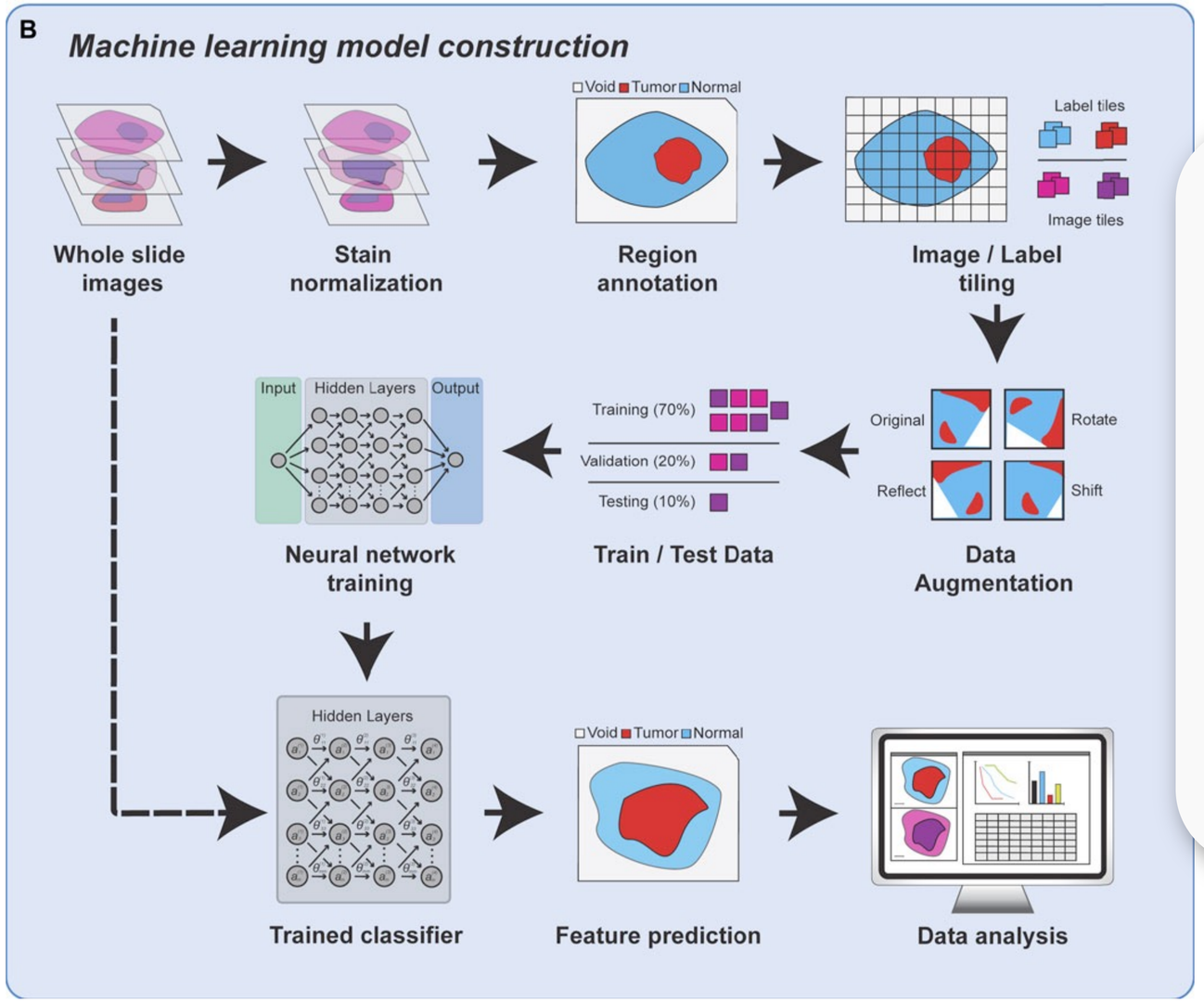


Learn more

Lee et al, Front Artif Intell, 2021

Pathomics

Deep learning



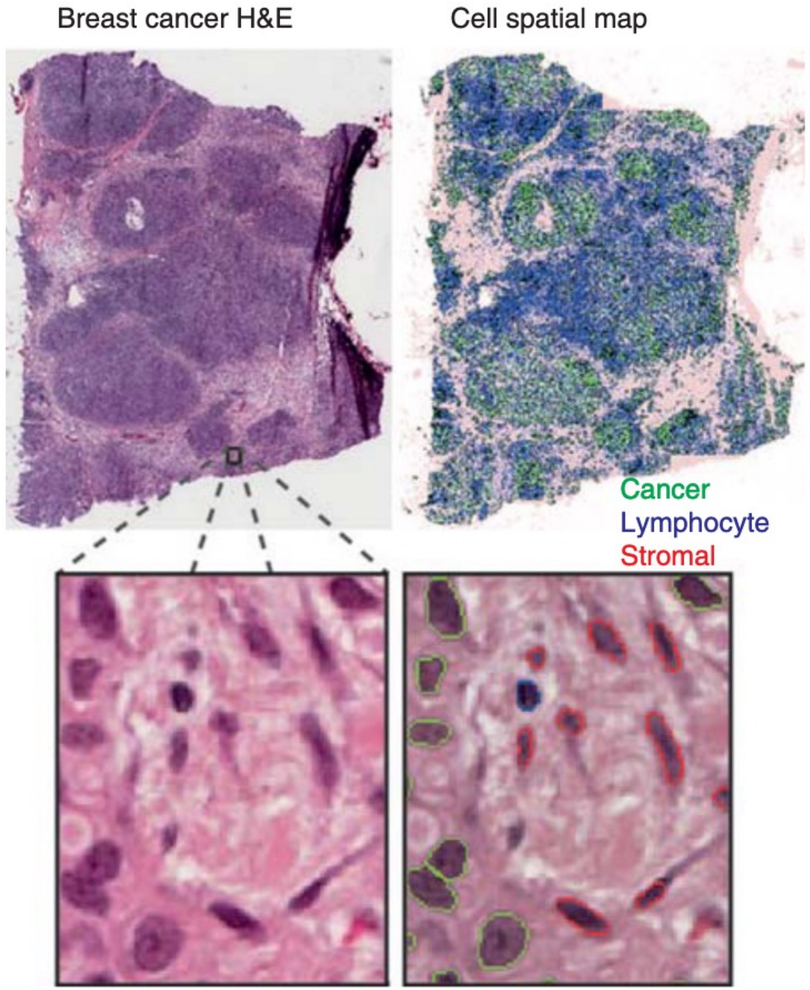
	Tile 1	Tile 2	Tile 3
Latent feature 1	8	1	0
Latent feature 2	5	3	10
Latent feature 3	3	8	4
Latent feature 4	4	7	7
Latent feature 5	12	5	9
Latent feature 6	6	5	4
Latent feature 7	2	6	3

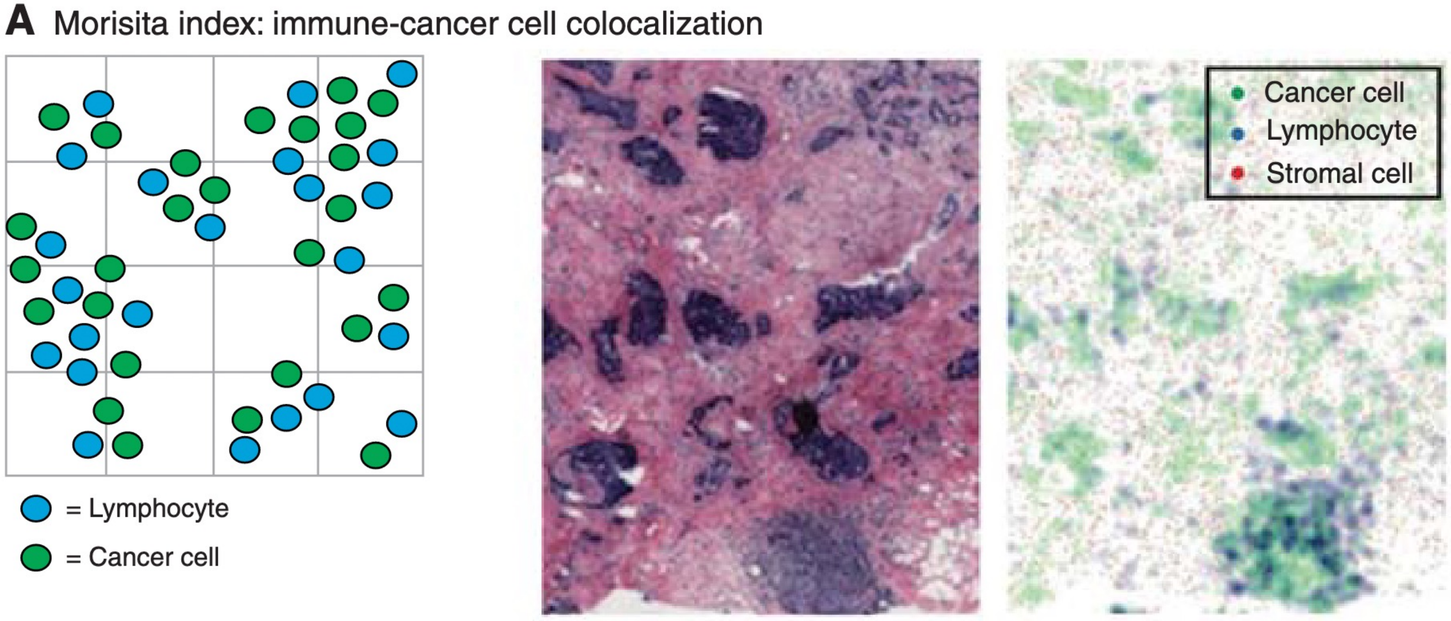
Learn more

Lee et al, Front Artif Intell, 2021

Pathomics

Microenvironment features



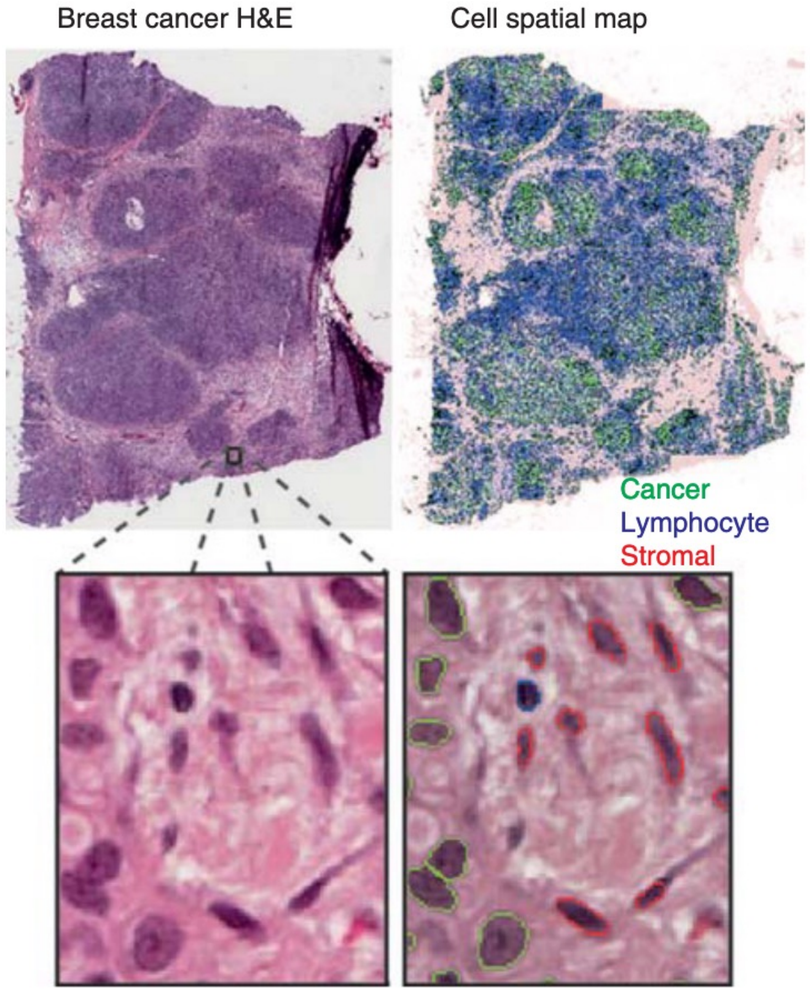


Learn more

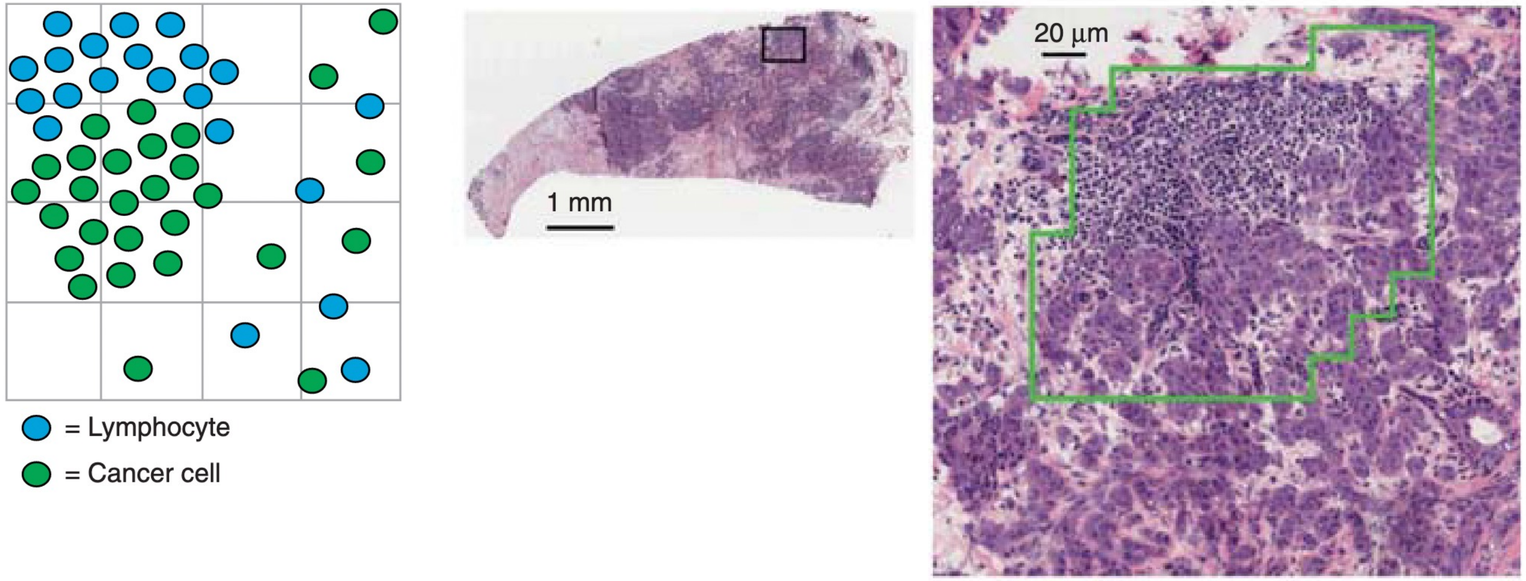
Y. Yuan, CSH Perspectives in Medicine, 2016

Pathomics

Microenvironment features



B Colocalized hotspots: spatial clusters of immune and cancer cells

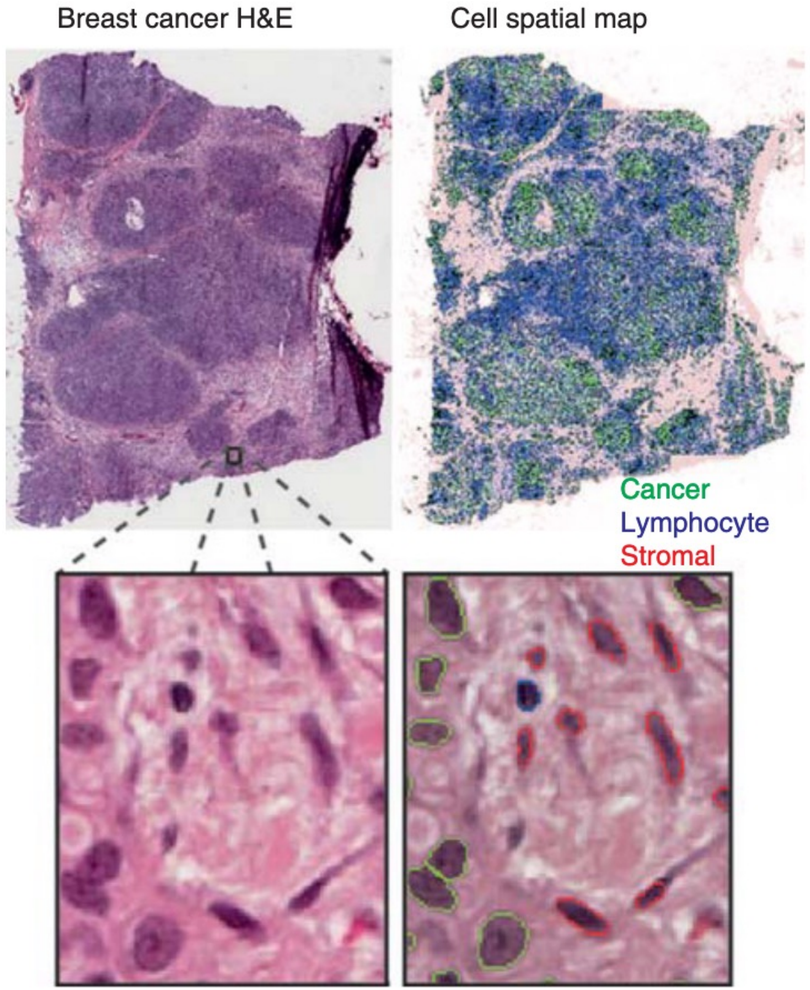


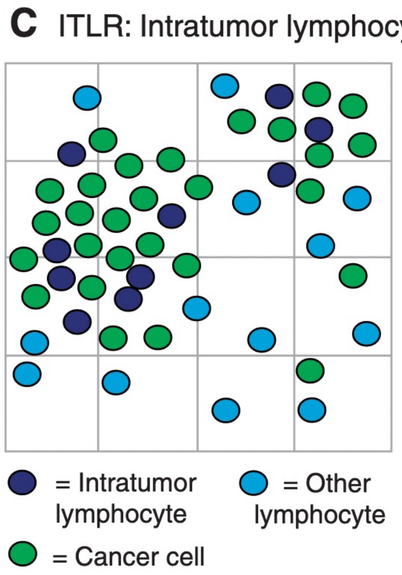
Learn more

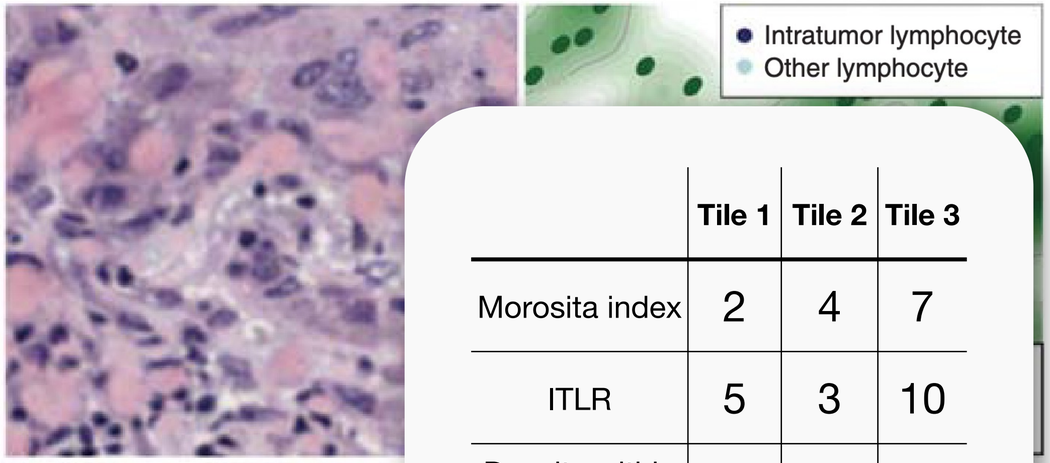
Y. Yuan, CSH Perspectives in Medicine, 2016

Pathomics

Microenvironment features







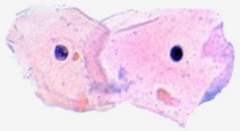
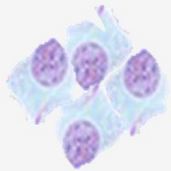
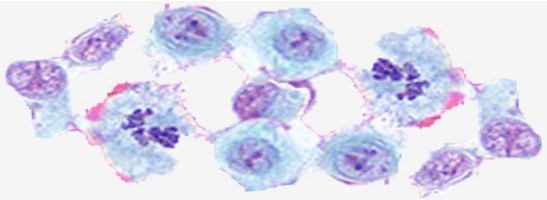
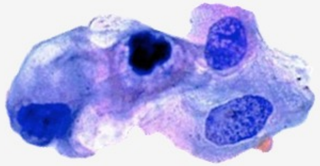
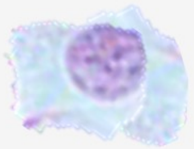
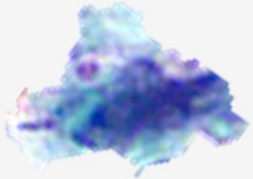
	Tile 1	Tile 2	Tile 3
Morosita index	2	4	7
ITLR	5	3	10
Density within 2 μm radius	3	8	4
Density within 5 μm radius	4	7	7
Distance to closest cell	12	5	9
Shannon entropy	6	5	4
Lymphocyte ratio	2	6	3

Learn more

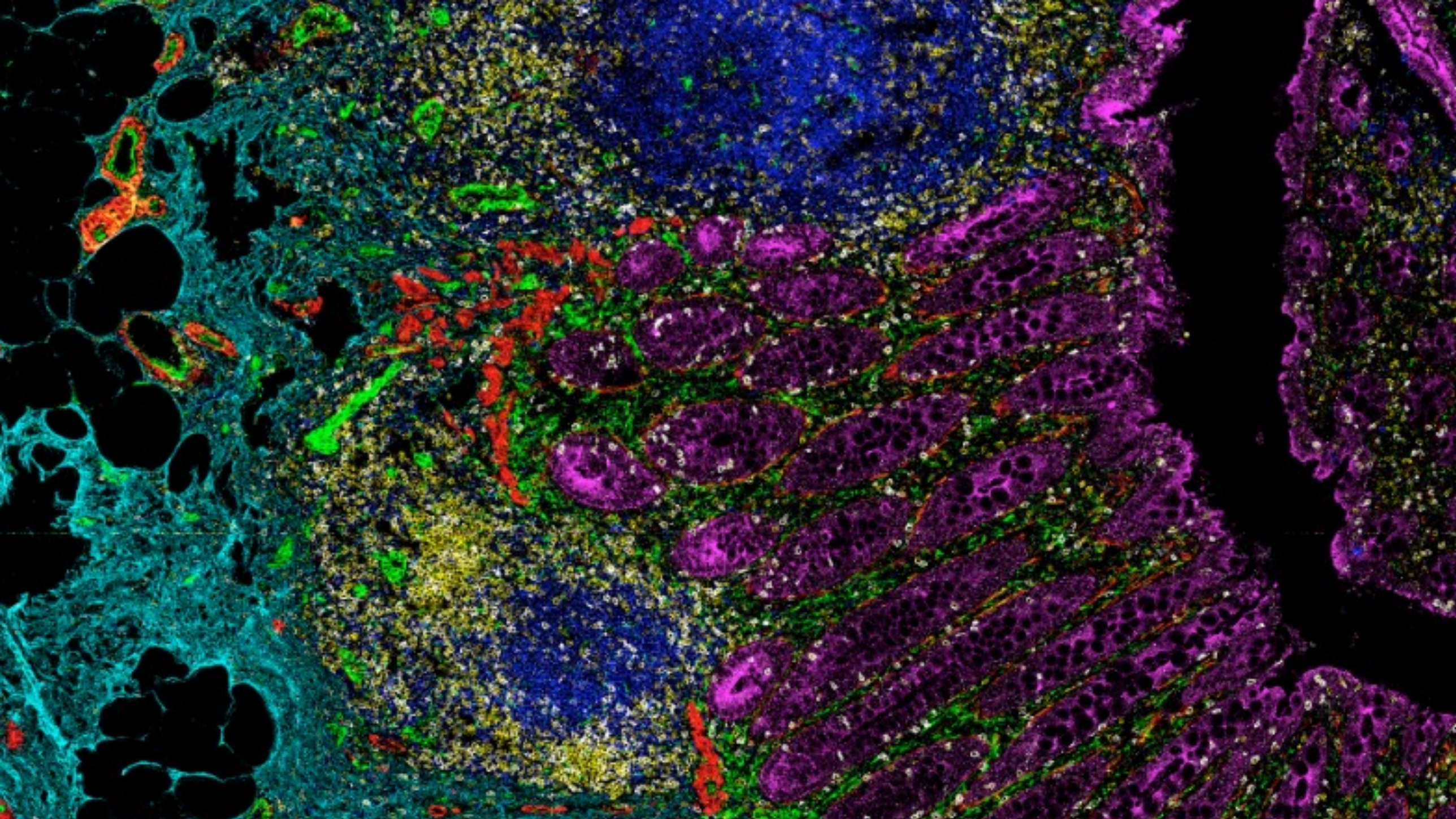
Y. Yuan, CSH Perspectives in Medicine, 2016

Pathomics

Cell morphology

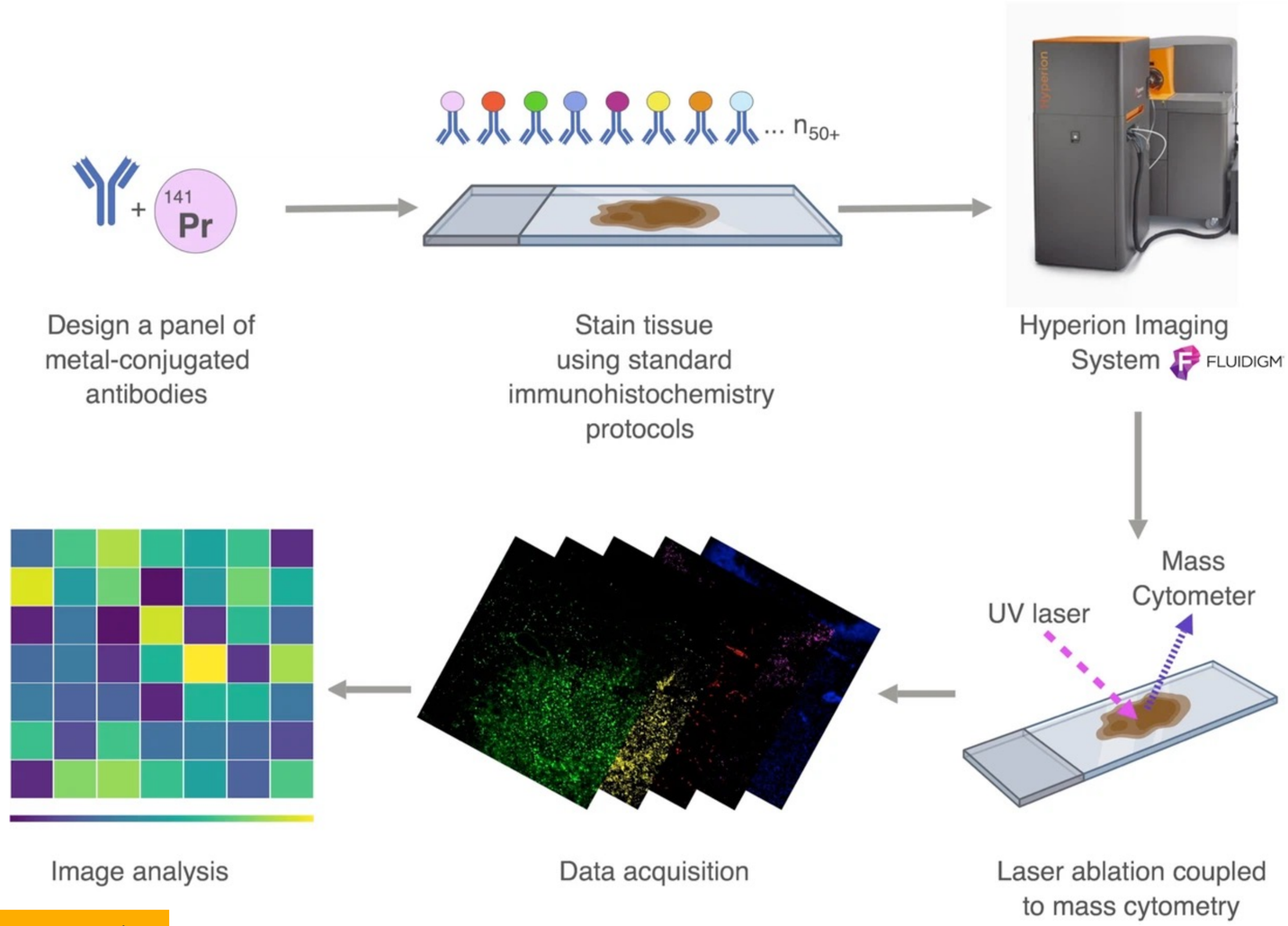
Normal	Cancer	
		Large, variably shaped nuclei
		Many dividing cells; Disorganized arrange
		Variation in size and s
		Loss of normal featur

	Cell 1	Cell 2	Cell 3
Perimeter	2	4	7
Surface	5	3	10
Mean	3	8	4
Median	4	7	7
Entropy	12	5	9
Energy	6	5	4
Run Length	2	6	3



Pathomics

Spatial transcriptomics

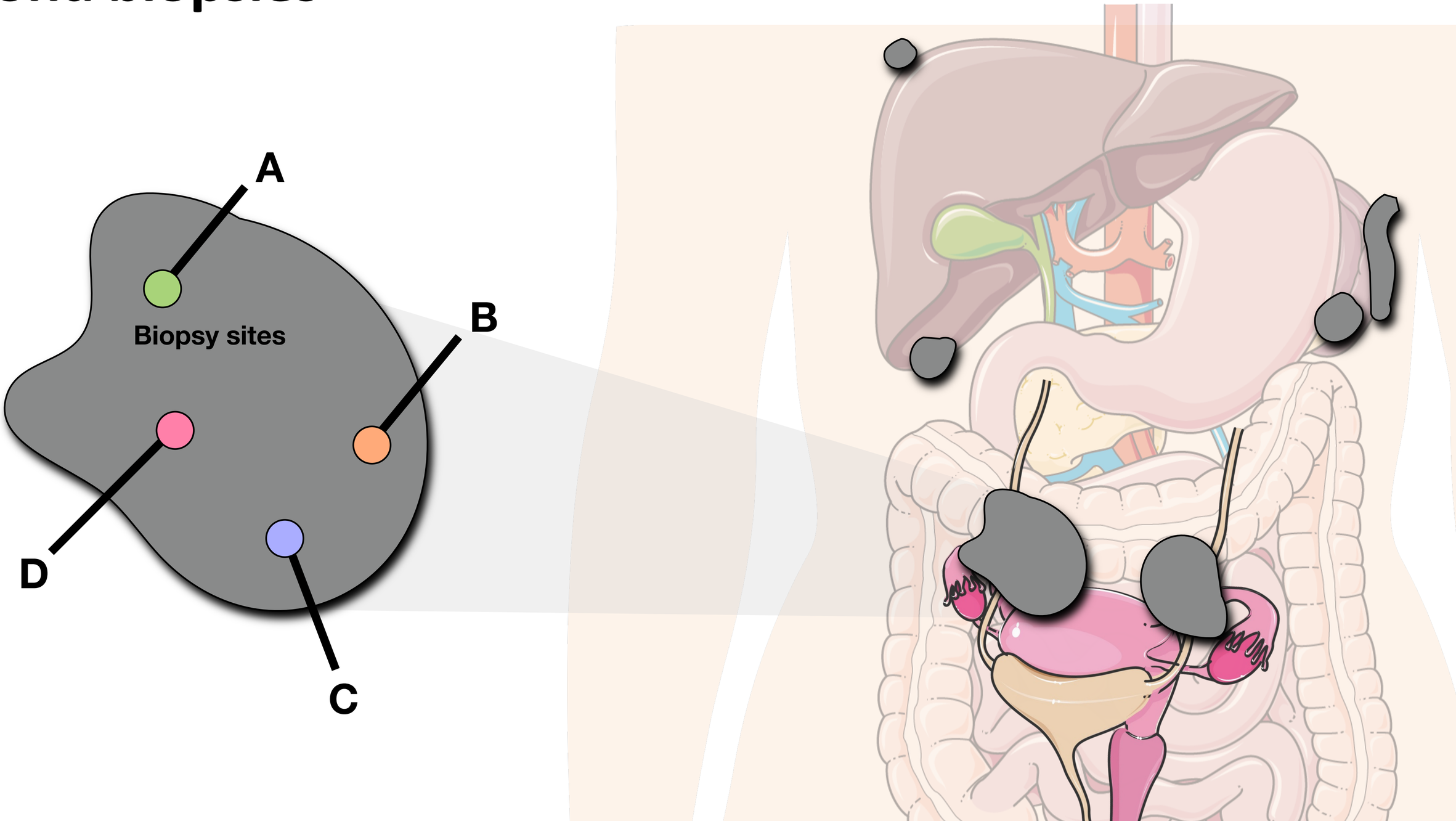


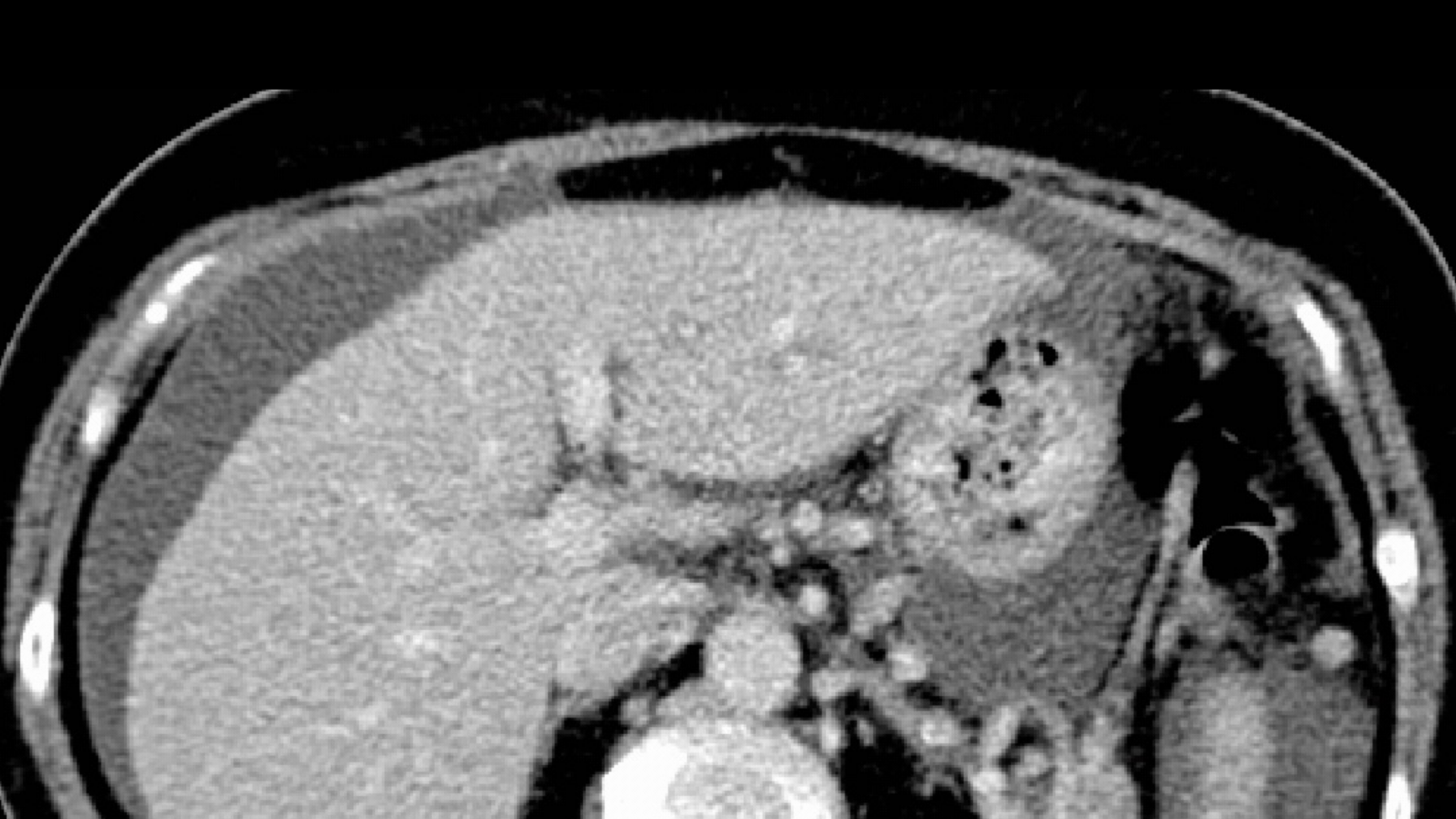
	Cell 1	Cell 2	Cell 3
Protein 1	2	4	7
Protein 2	5	3	10
Protein 3	3	8	4
Protein 4	4	7	7
Protein 5	12	5	9
Protein 6	6	5	4
Protein 7	2	6	3

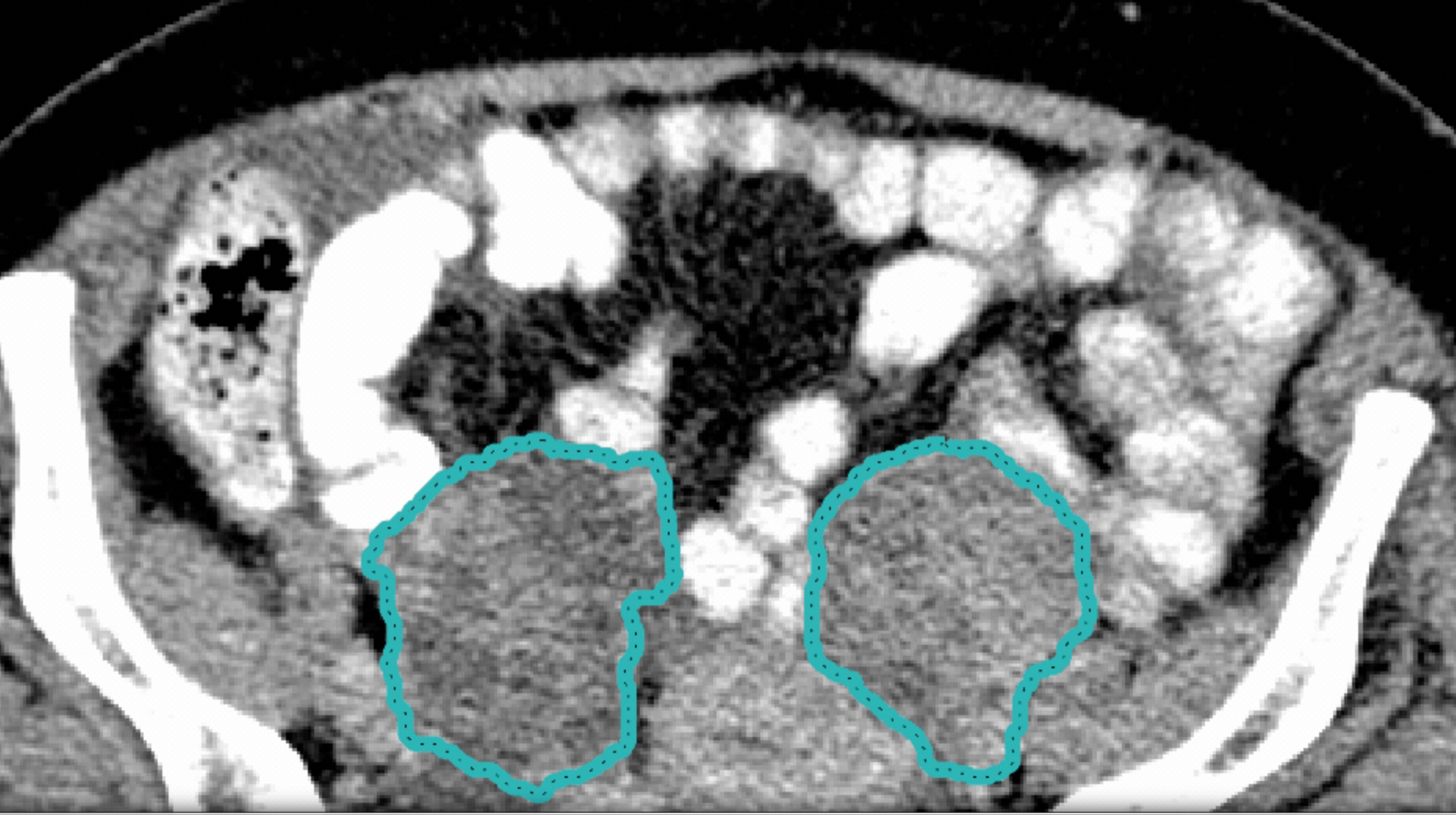
Learn more

Ramaglia et al, eLife 2020

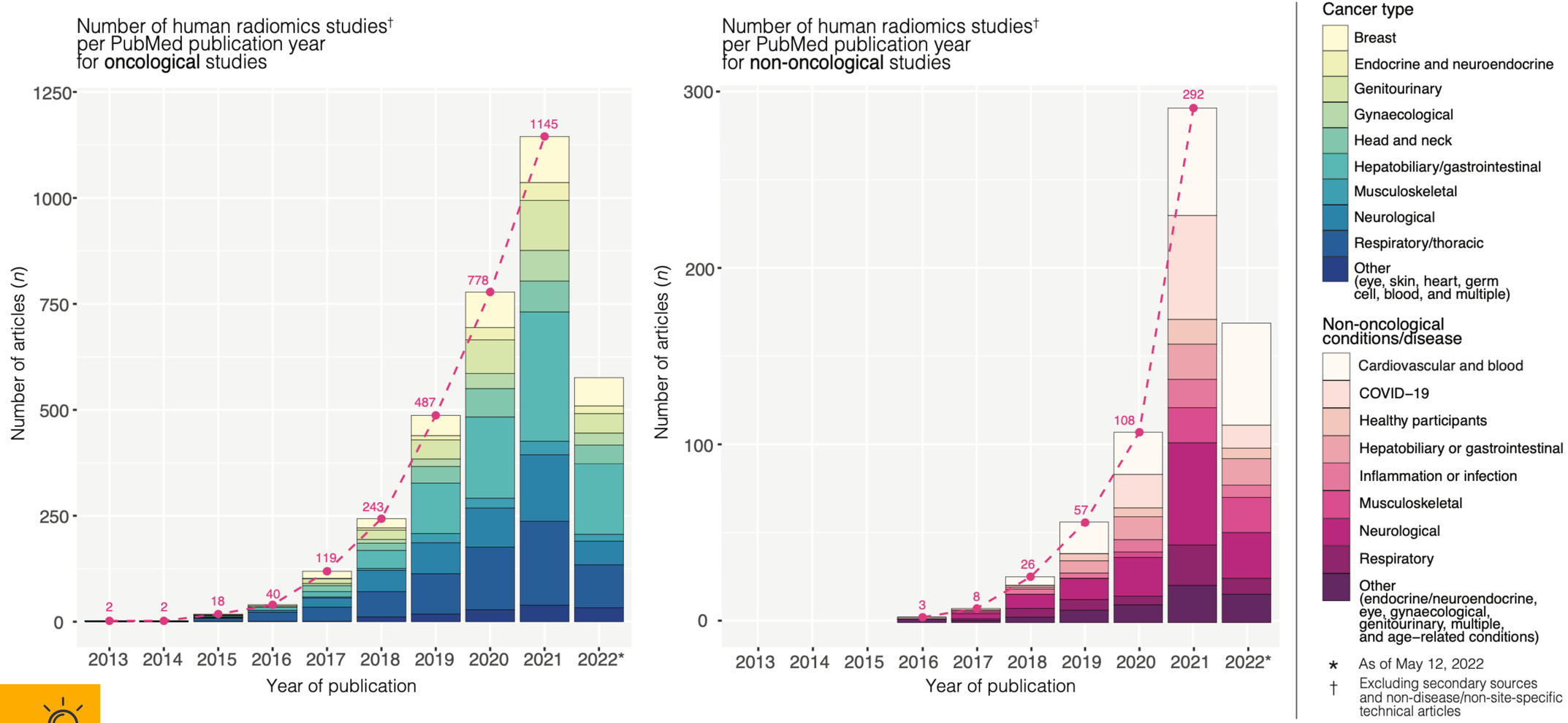
Beyond biopsies







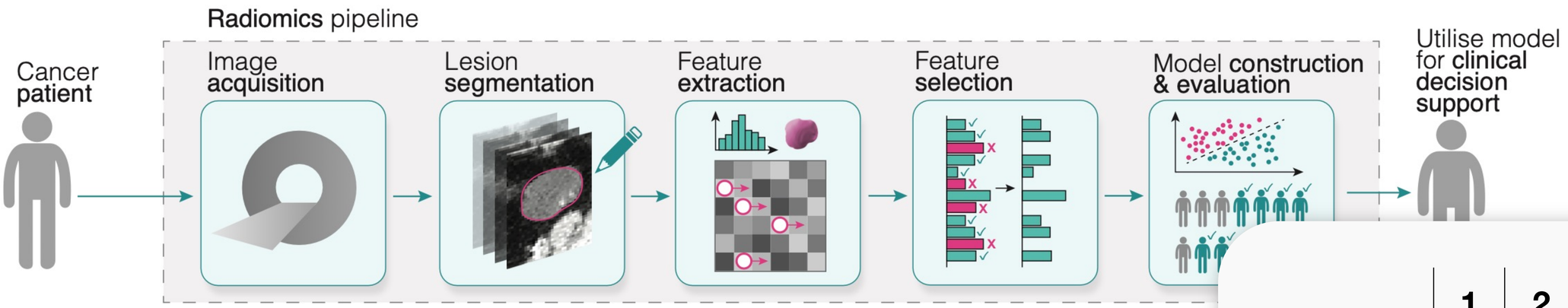
Radiomics



Learn more

C. McCague et al.,
Under review, 2022

Radiomics

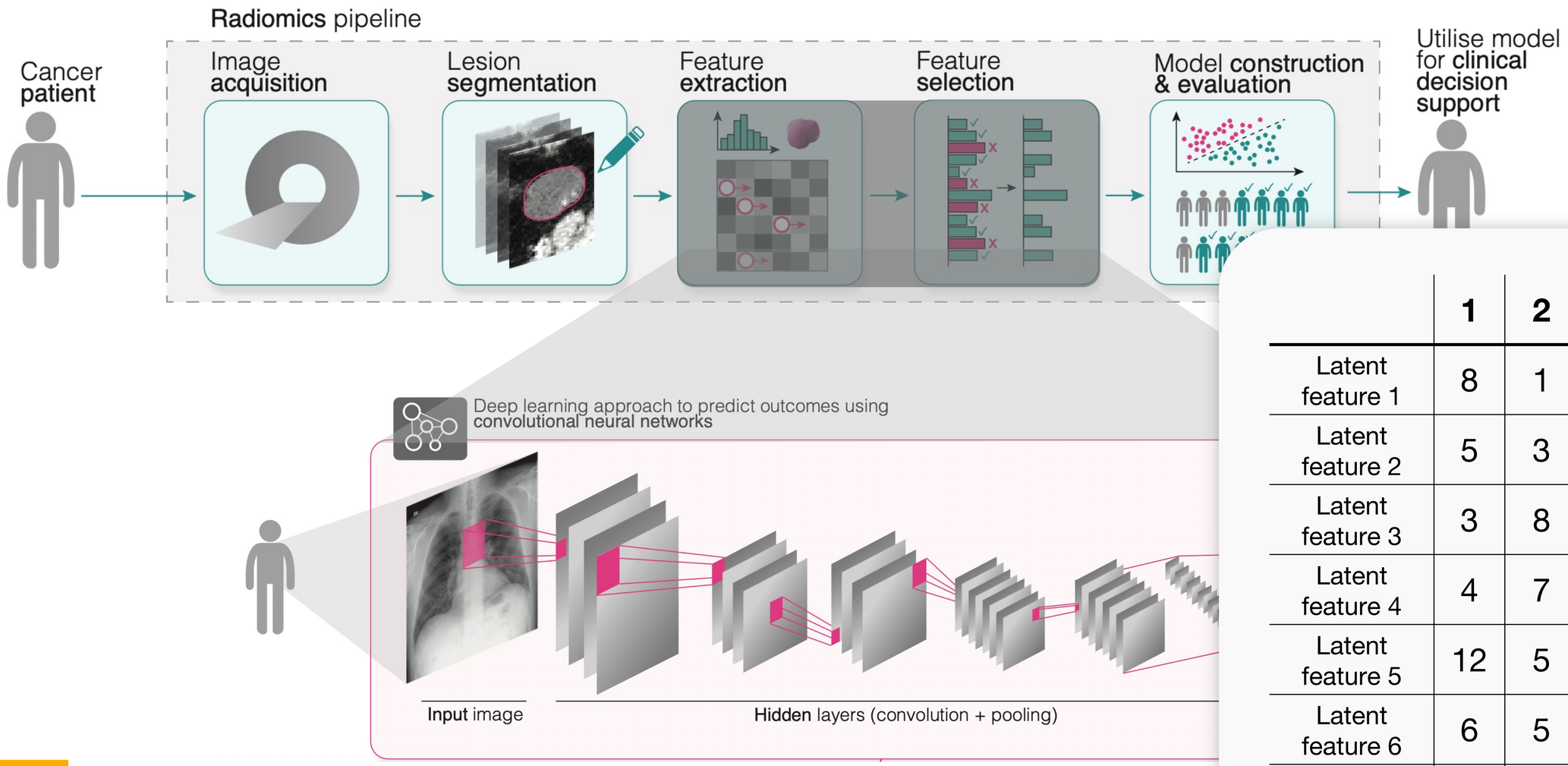


	1	2	3
Volume	32	22	4
Surface	5	3	10
Mean	3	8	4
Median	4	7	7
Entropy	12	5	9
Energy	6	5	4
Run Length	2	6	3

Learn more

C. McCague et al., Under review, 2022

Radiomics



	1	2	3
Latent feature 1	8	1	0
Latent feature 2	5	3	10
Latent feature 3	3	8	4
Latent feature 4	4	7	7
Latent feature 5	12	5	9
Latent feature 6	6	5	4
Latent feature 7	2	6	3

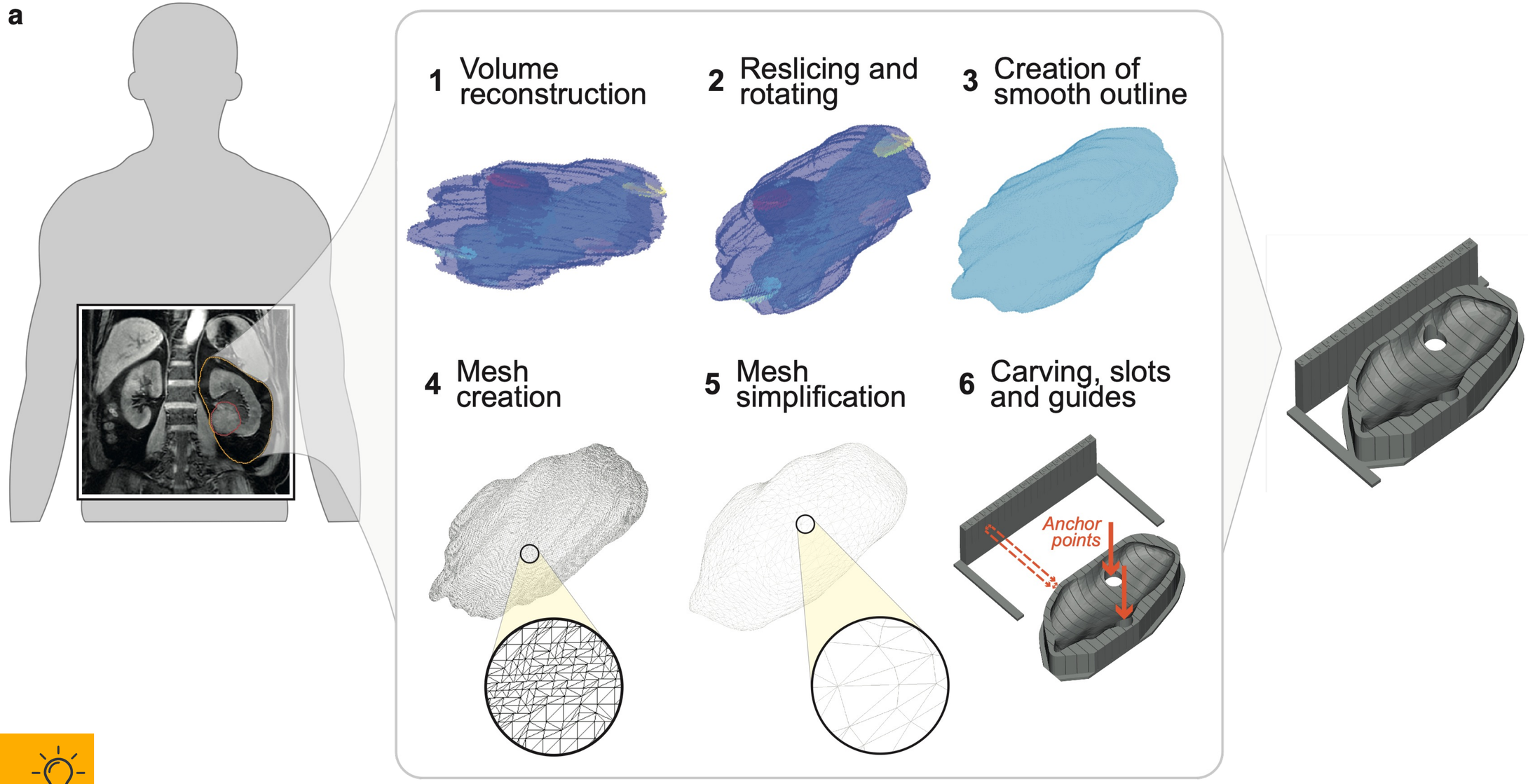
Learn more

C. McCague et al., Under review, 2022

Question

What challenges for integration does the different nature of biopsies vs. medical imaging pose?

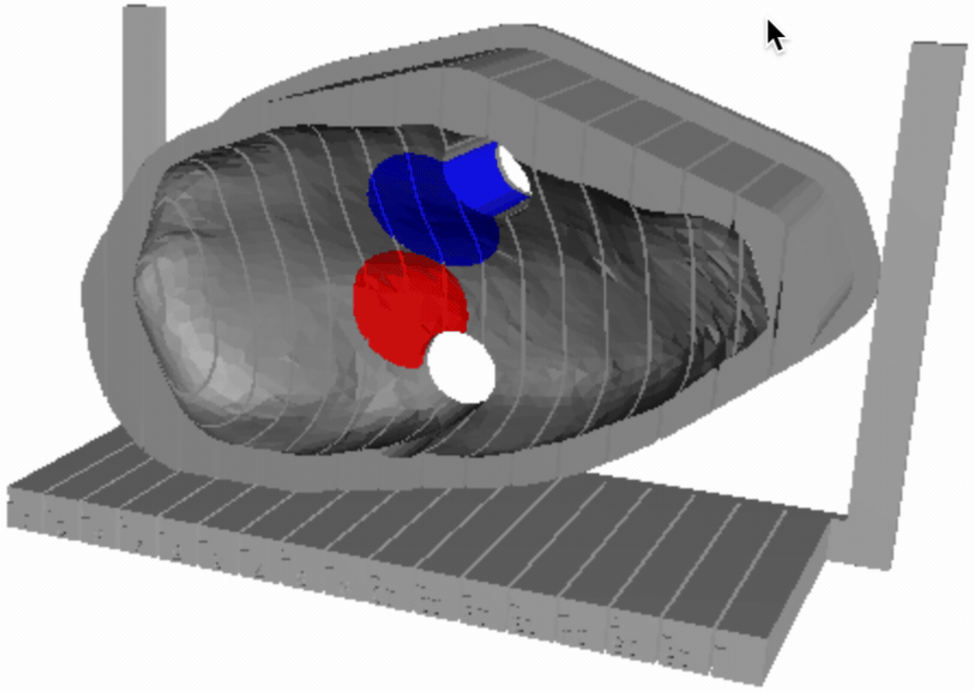
BONUS: 3D printing



Learn more

Crispin et al, JCO CCI 2020

BONUS: 3D printing

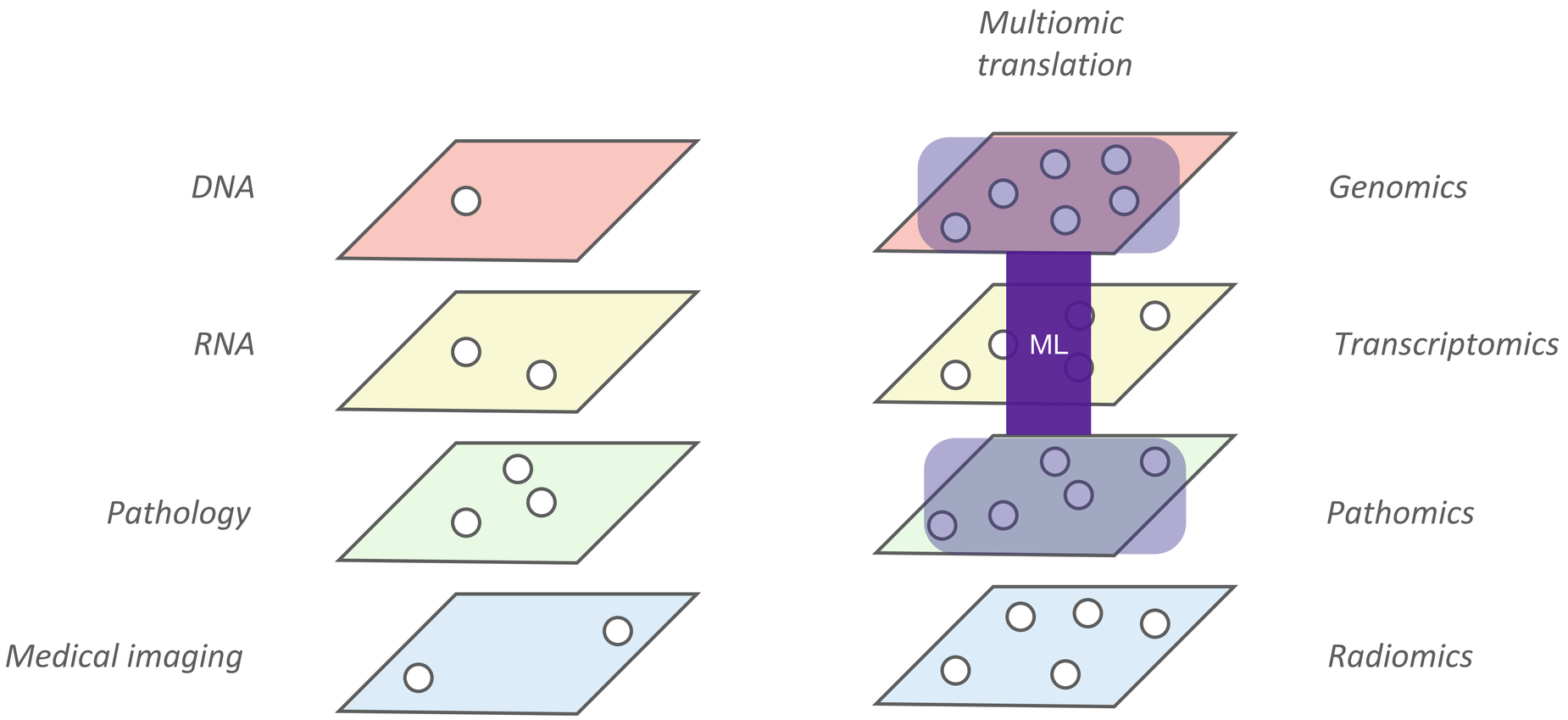


Learn more

Crispin et al, JCO CCI 2020

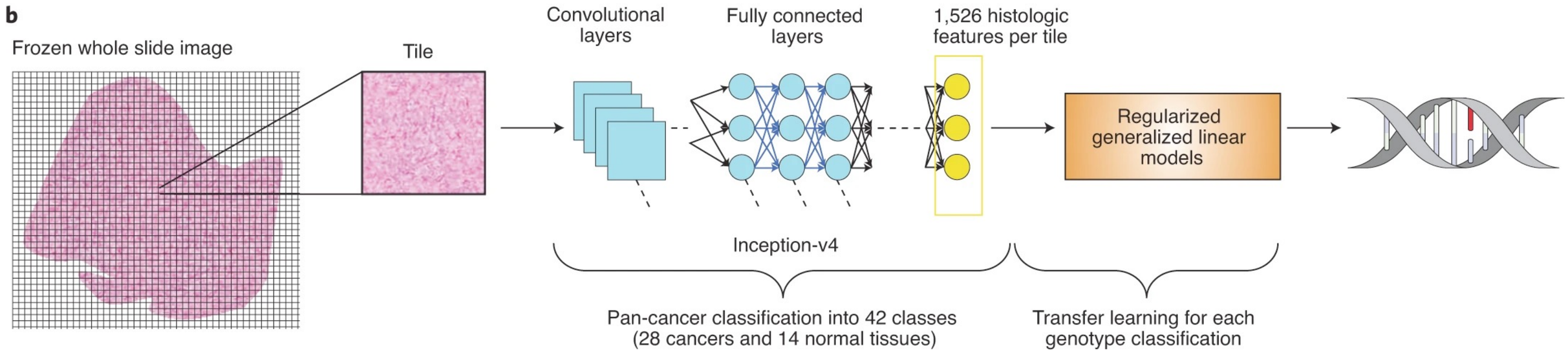
Multiomic translation

Multi-omic translation



Pathomics to predict genomics

Fu et al, 2020

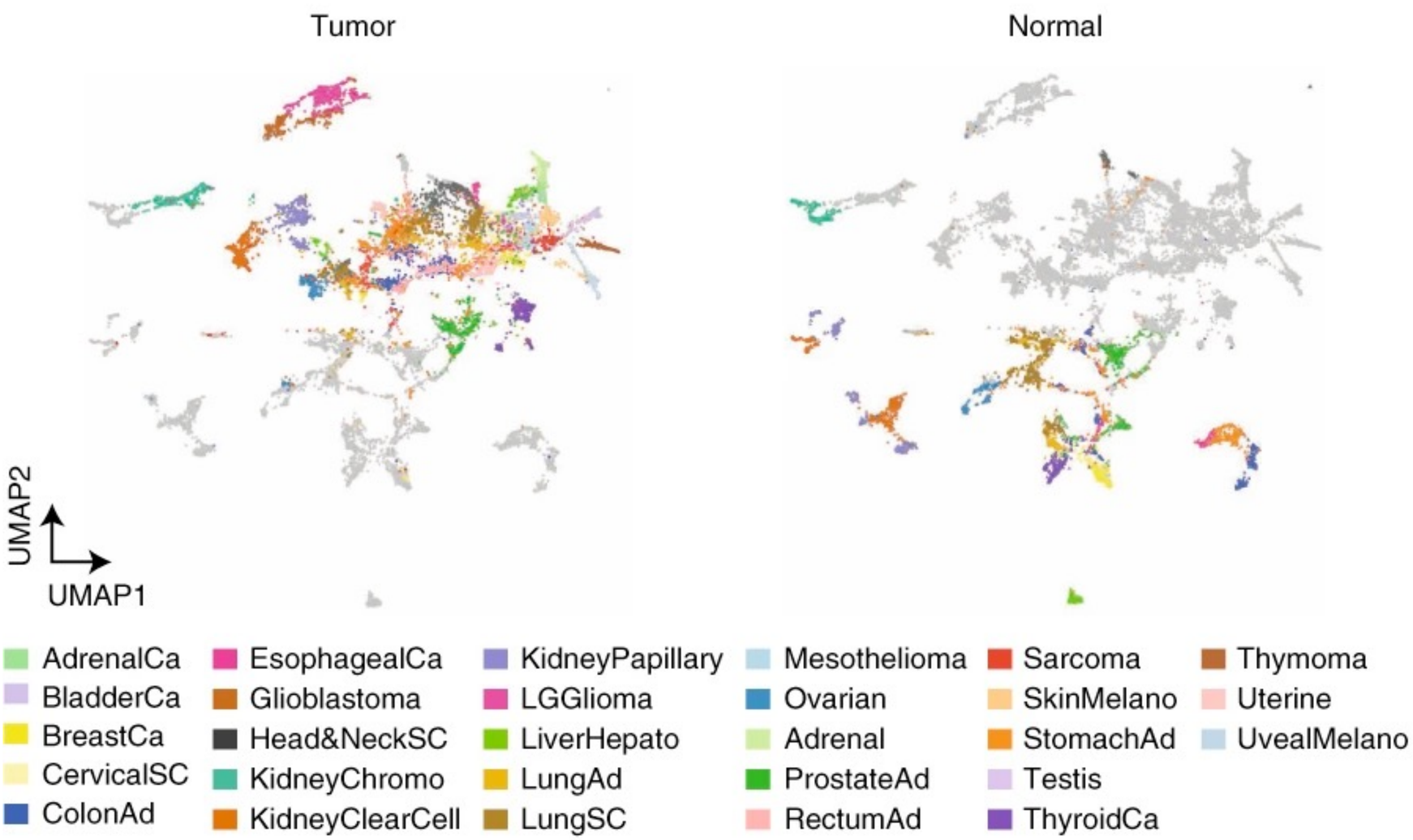


Learn more

N. Coudray and A. Tsirigos, Nature Cancer 2020

Pathomics to predict genomics

Fu et al, 2020

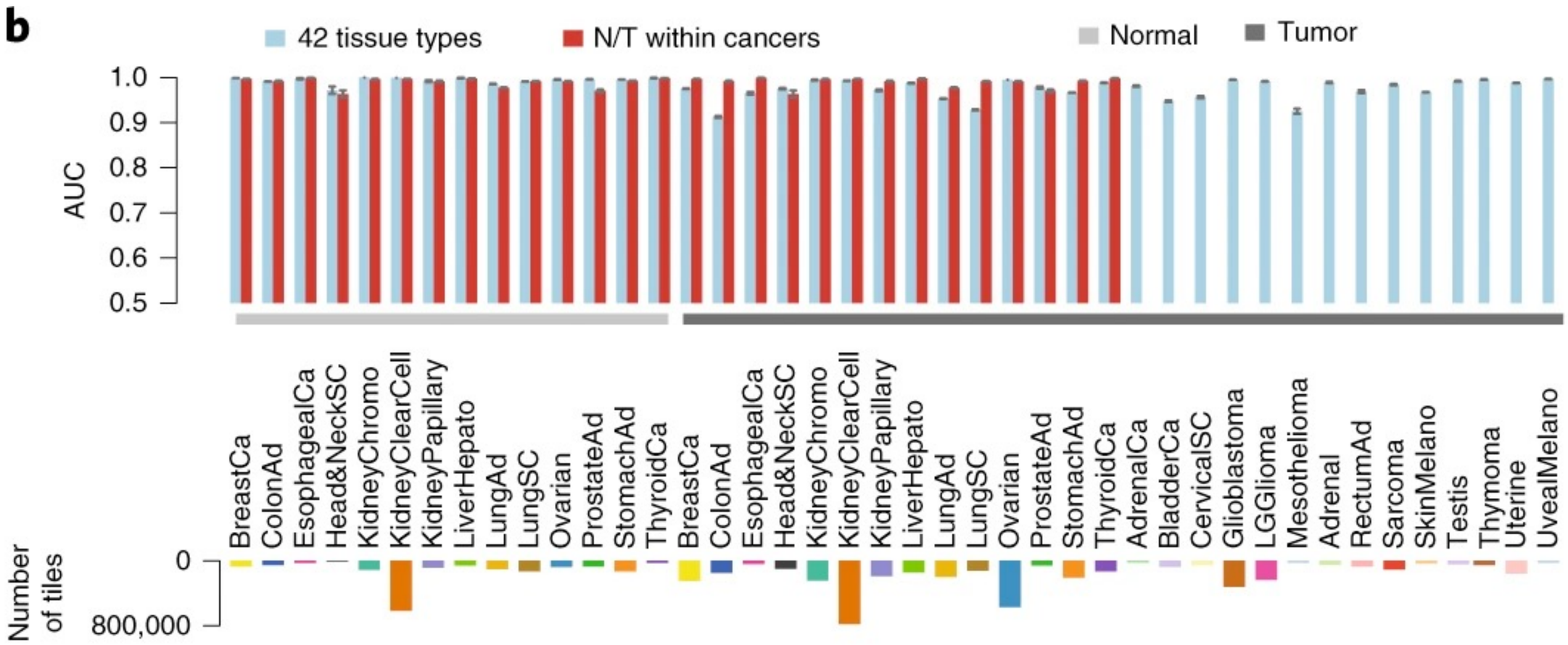


Learn more

Fu et al., Nature Cancer 2020

Pathomics to predict genomics

Fu et al, 2020

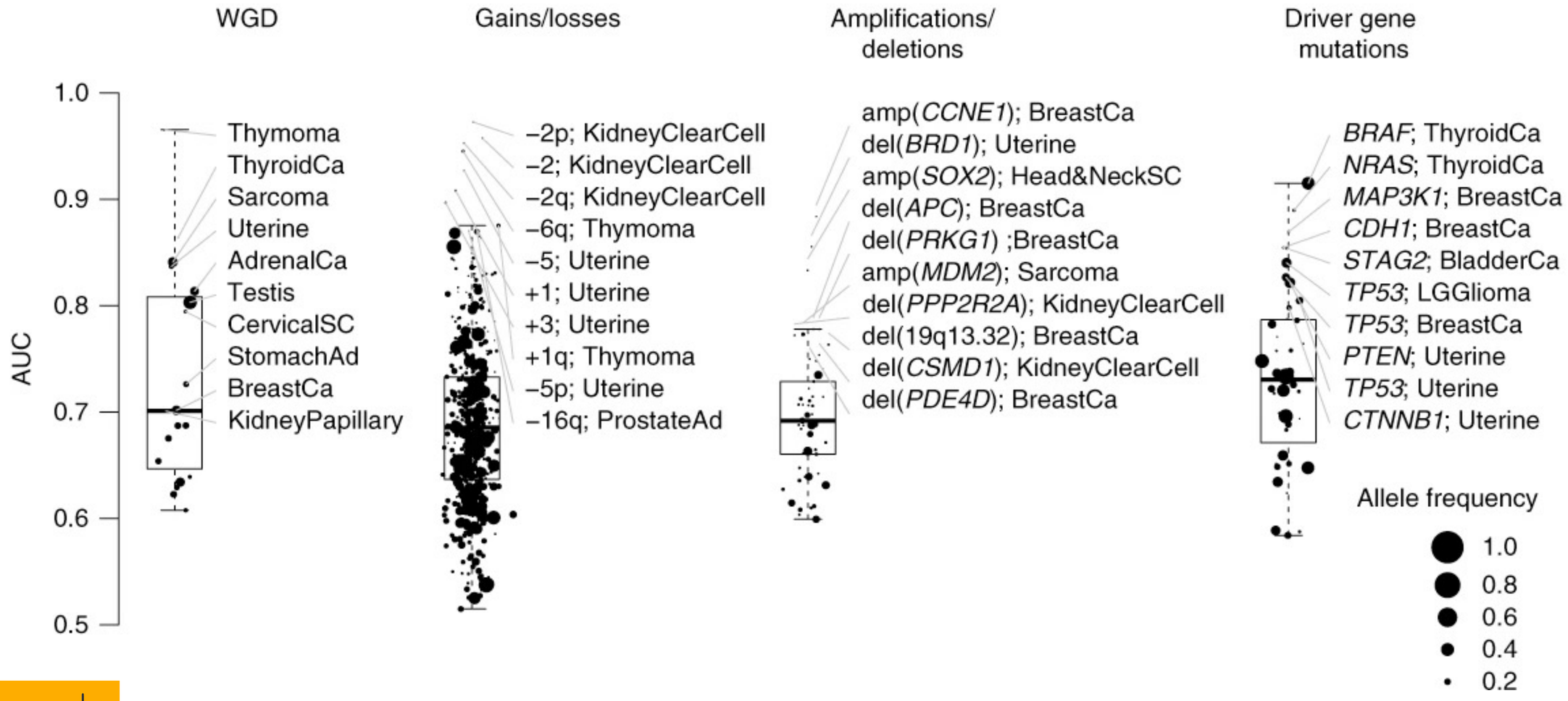


Learn more

Fu et al., Nature Cancer 2020

Pathomics to predict genomics

Fu et al, 2020

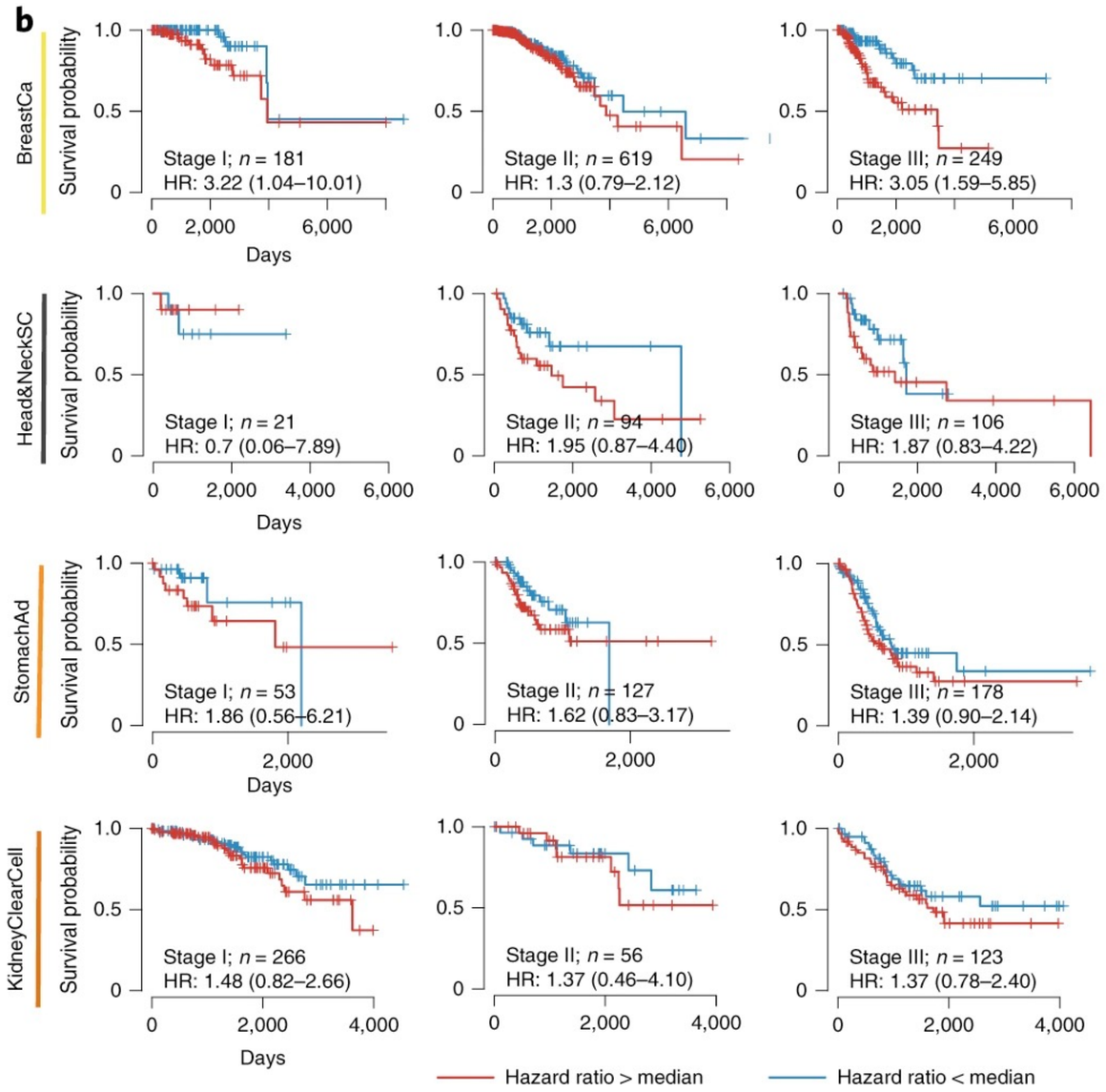


Learn more

Fu et al., Nature Cancer 2020

Pathomics to predict genomics

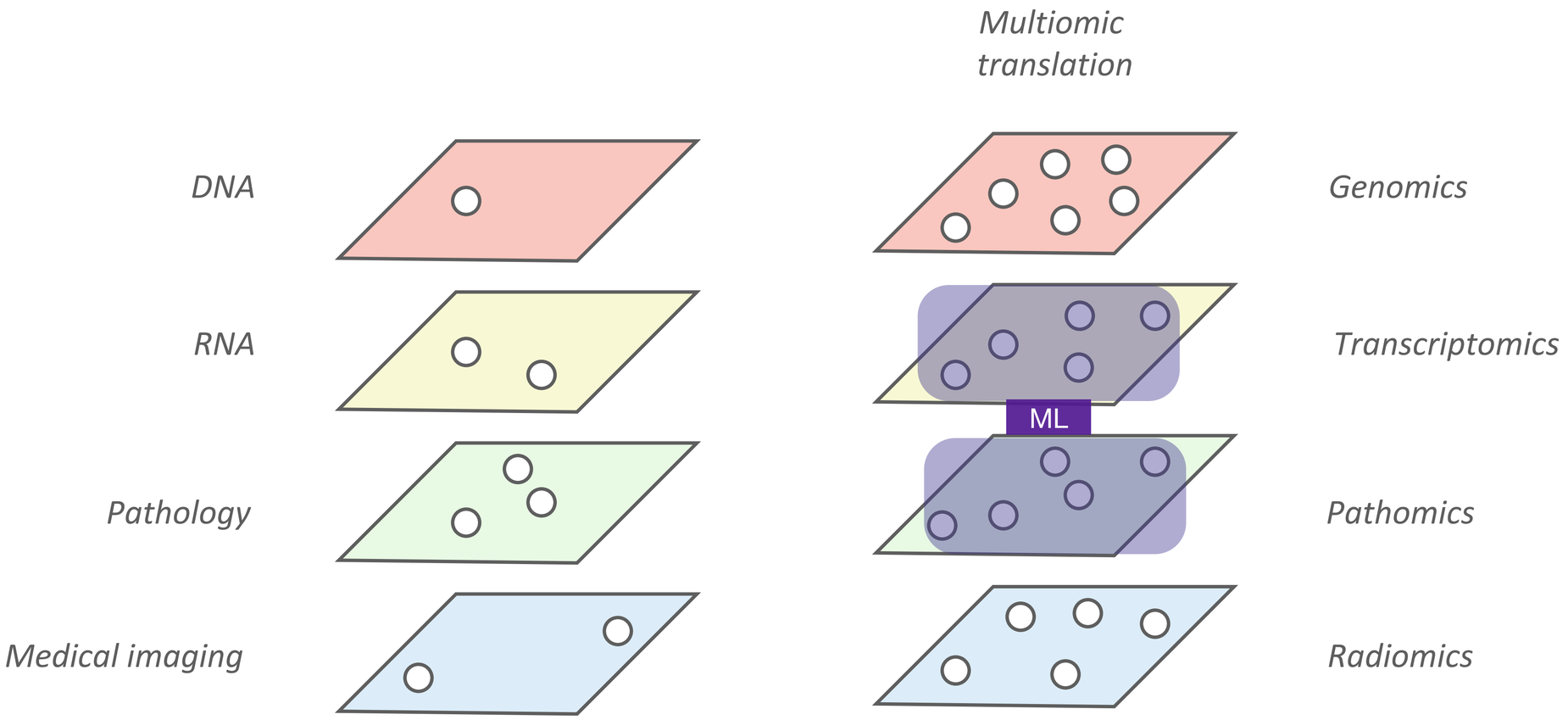
Fu et al, 2020



Learn more

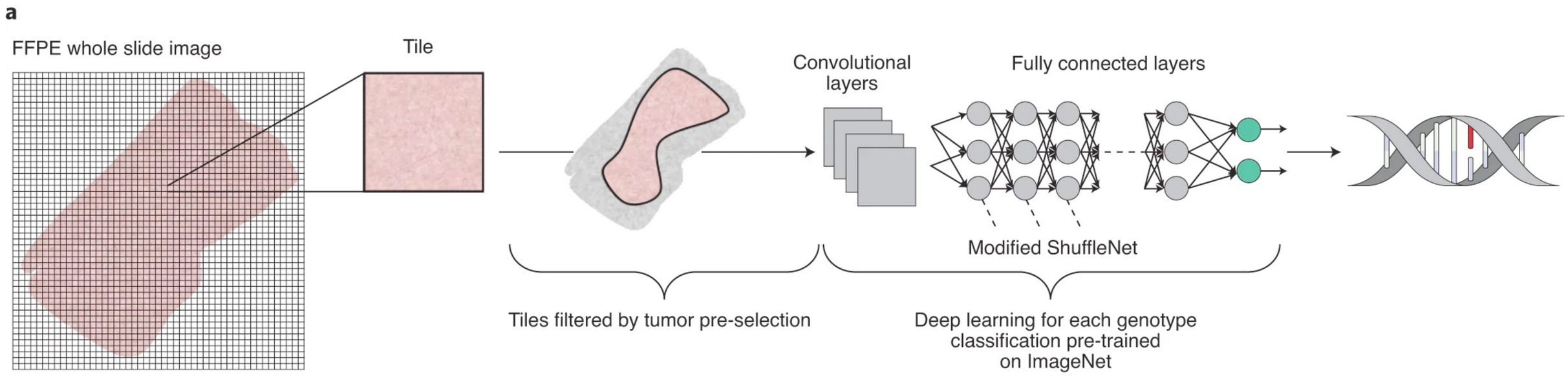
Fu et al., Nature Cancer 2020

Multi-omic translation



Pathomics to predict transcriptomics

Kather et al, 2020

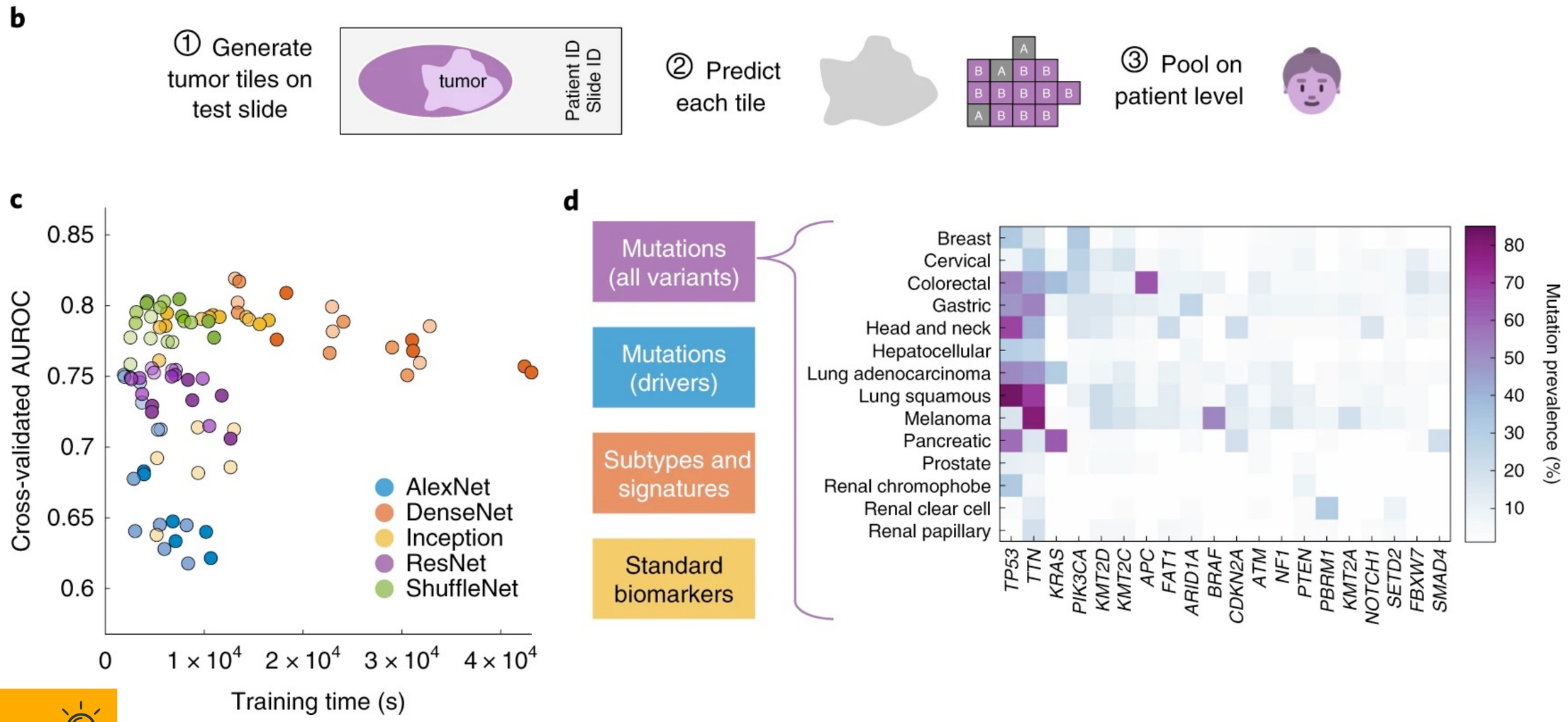


Learn more

N. Coudray and A. Tsirigos, Nature Cancer 2020

Pathomics to predict transcriptomics

Kather et al, 2020

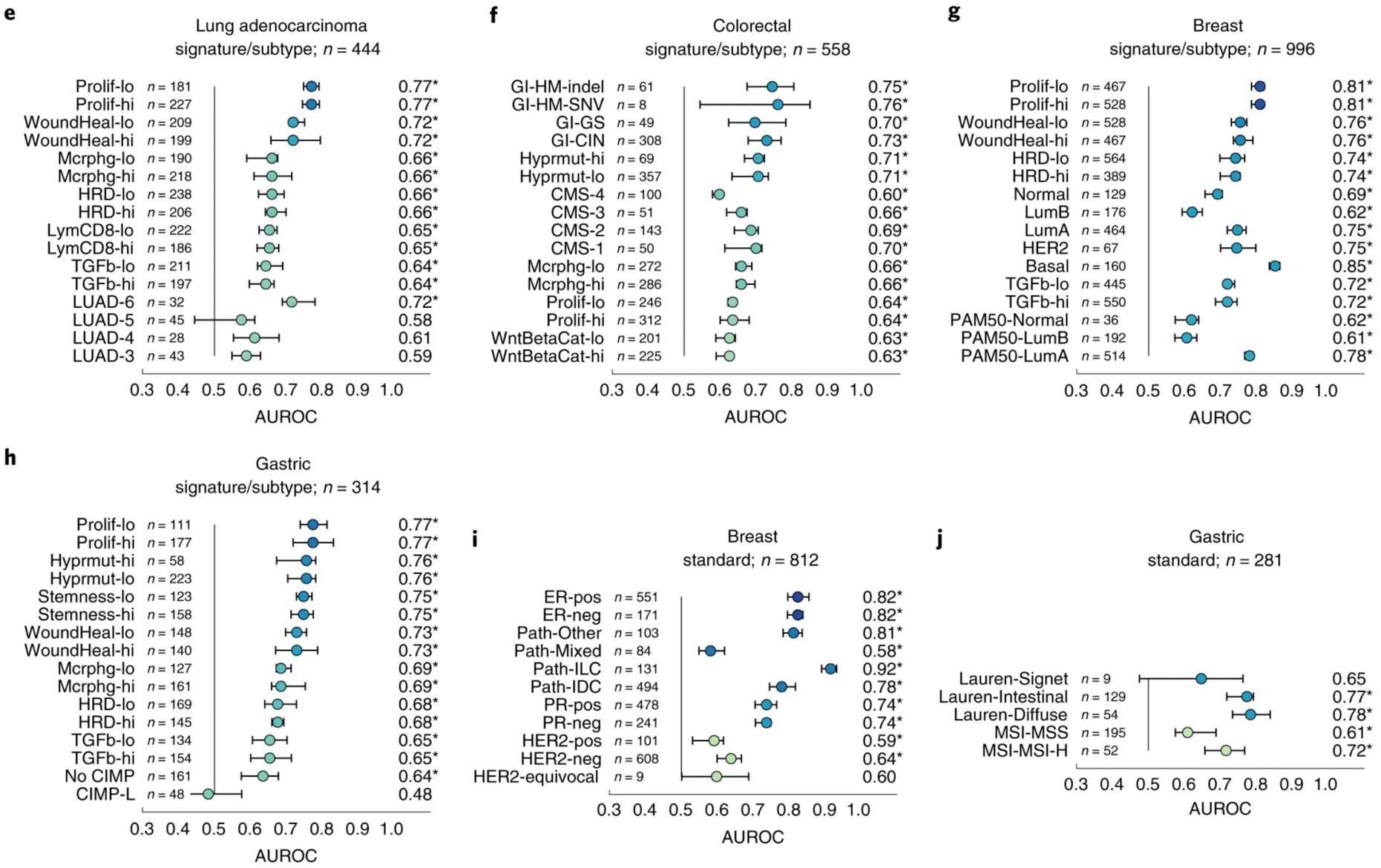


Learn more

J. N. Kather et al., Nature Cancer 2020

Pathomics to predict transcriptomics

Kather et al, 2020



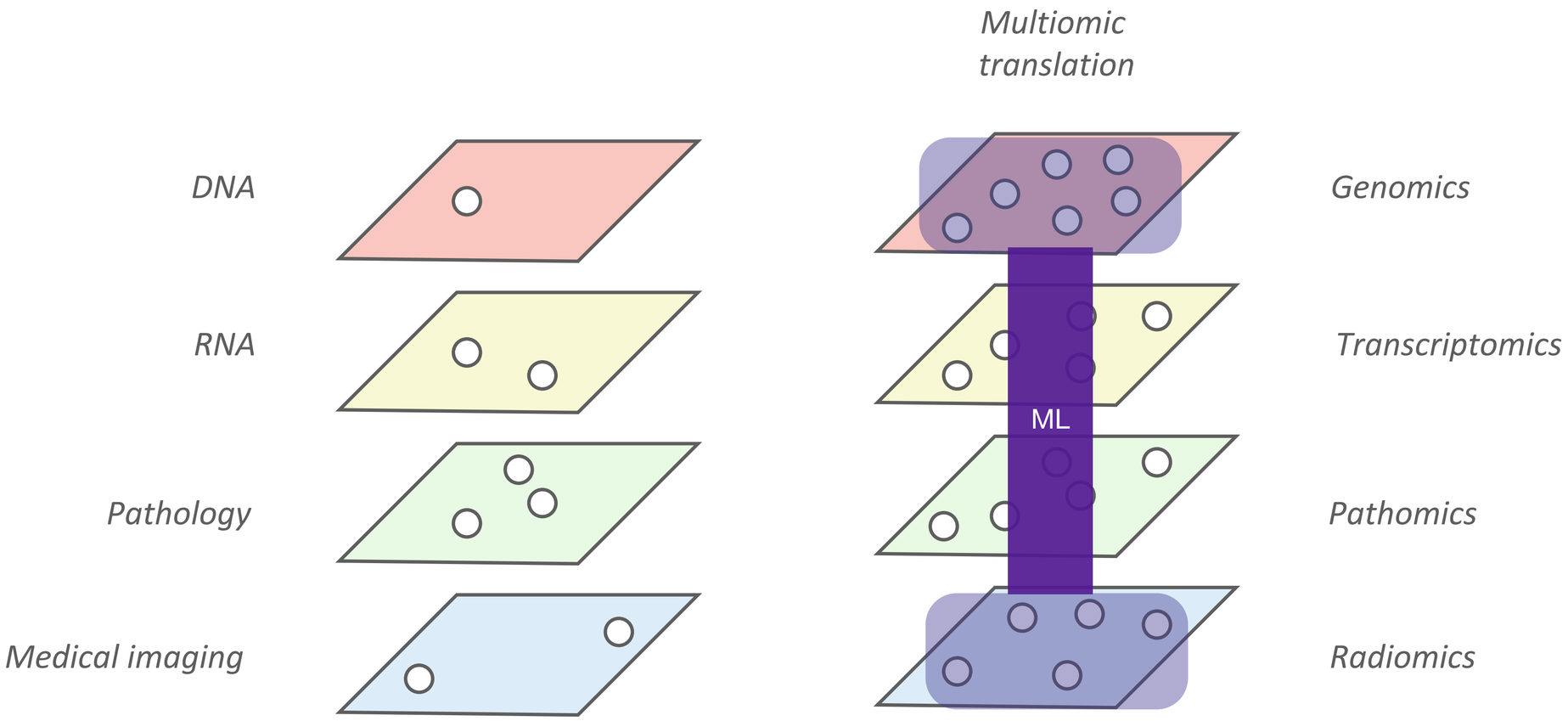
Learn more

J. N. Kather et al., Nature Cancer 2020

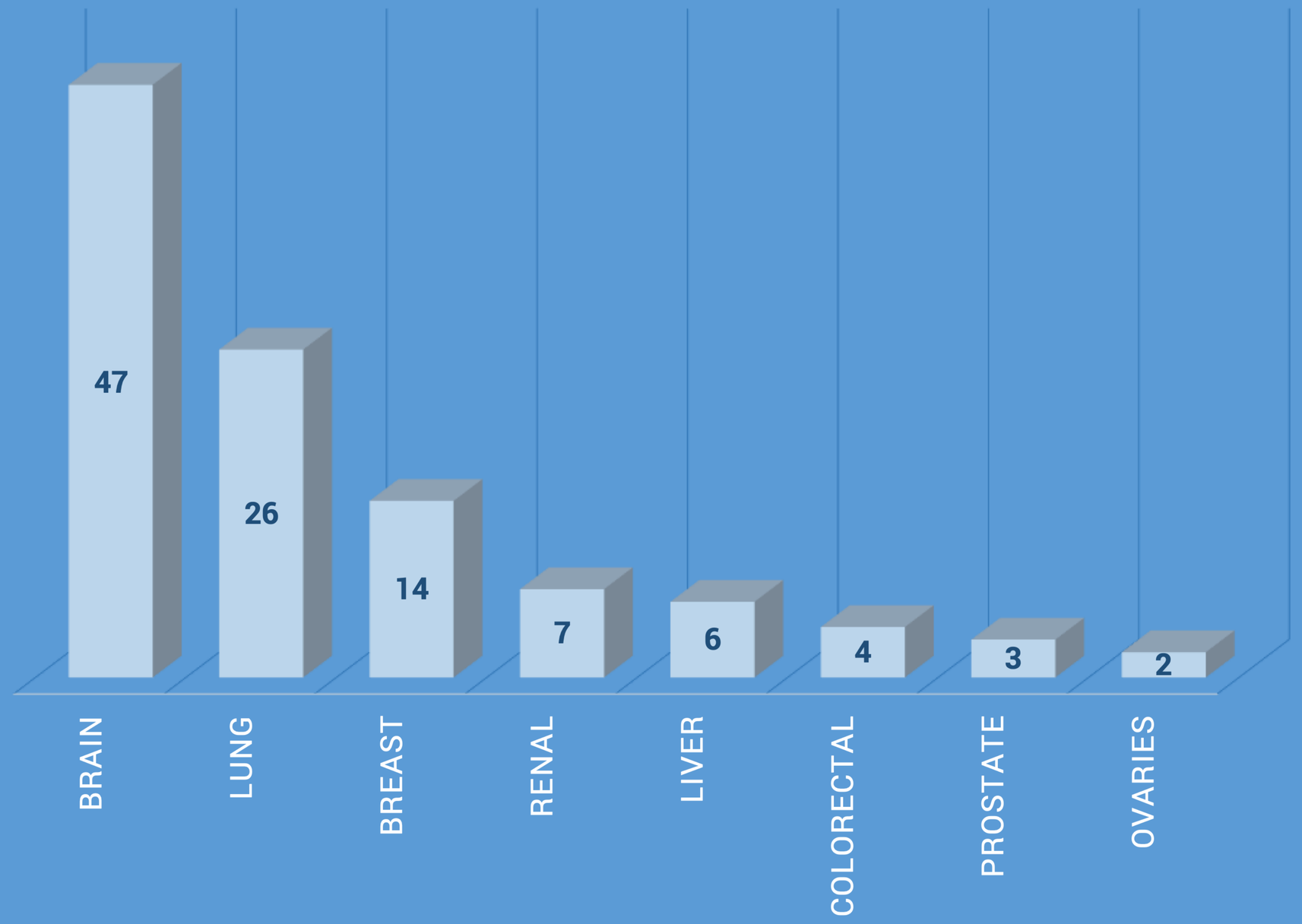
Question

What are the strengths and limitations of these two approaches?

Multi-omic translation



Radiogenomics



Learn more

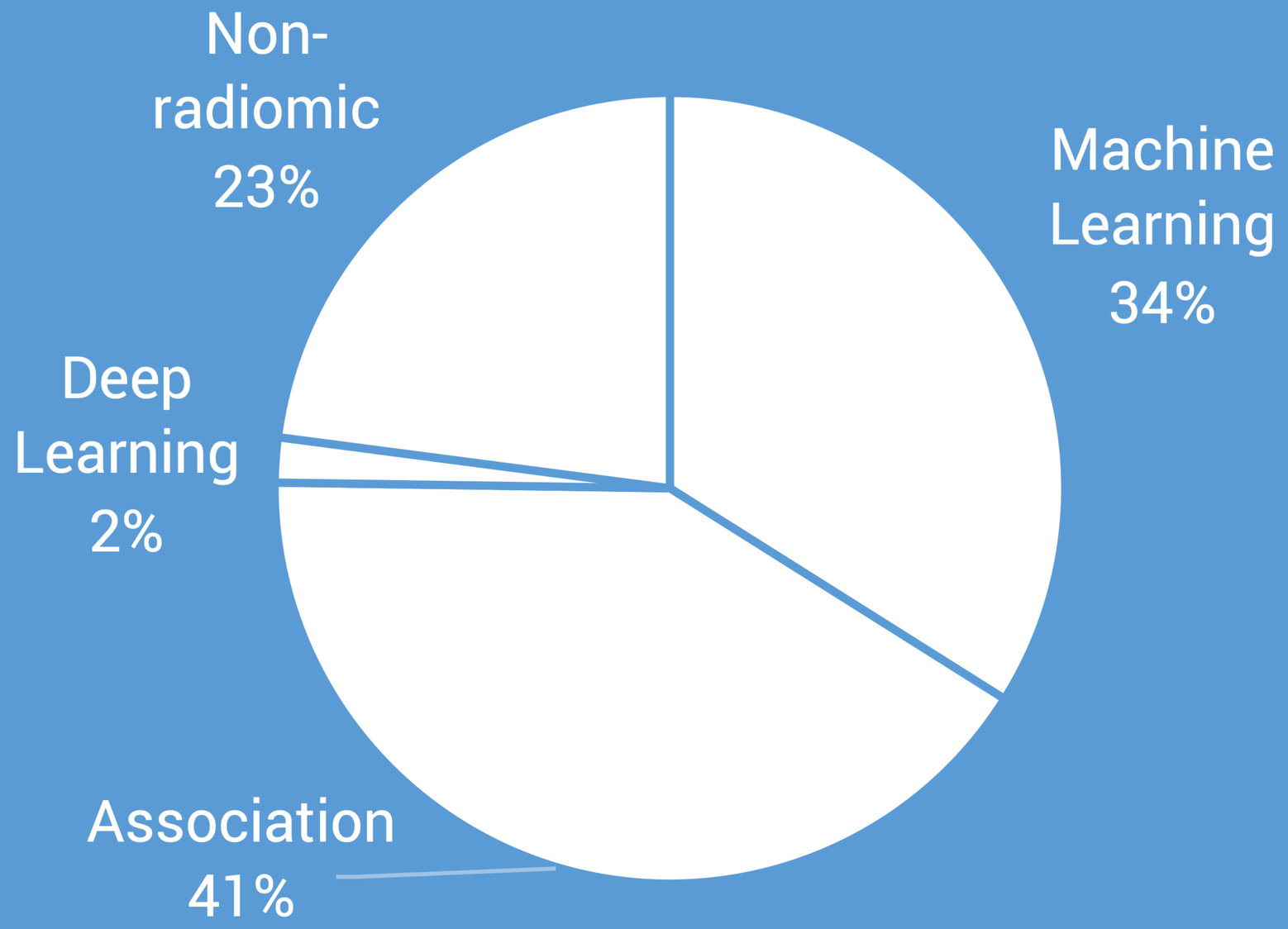
Bodalal et al, Abdominal Radiology 2019

An even split

Learn more

Bodalal et al,
Abdominal Radiology
2019

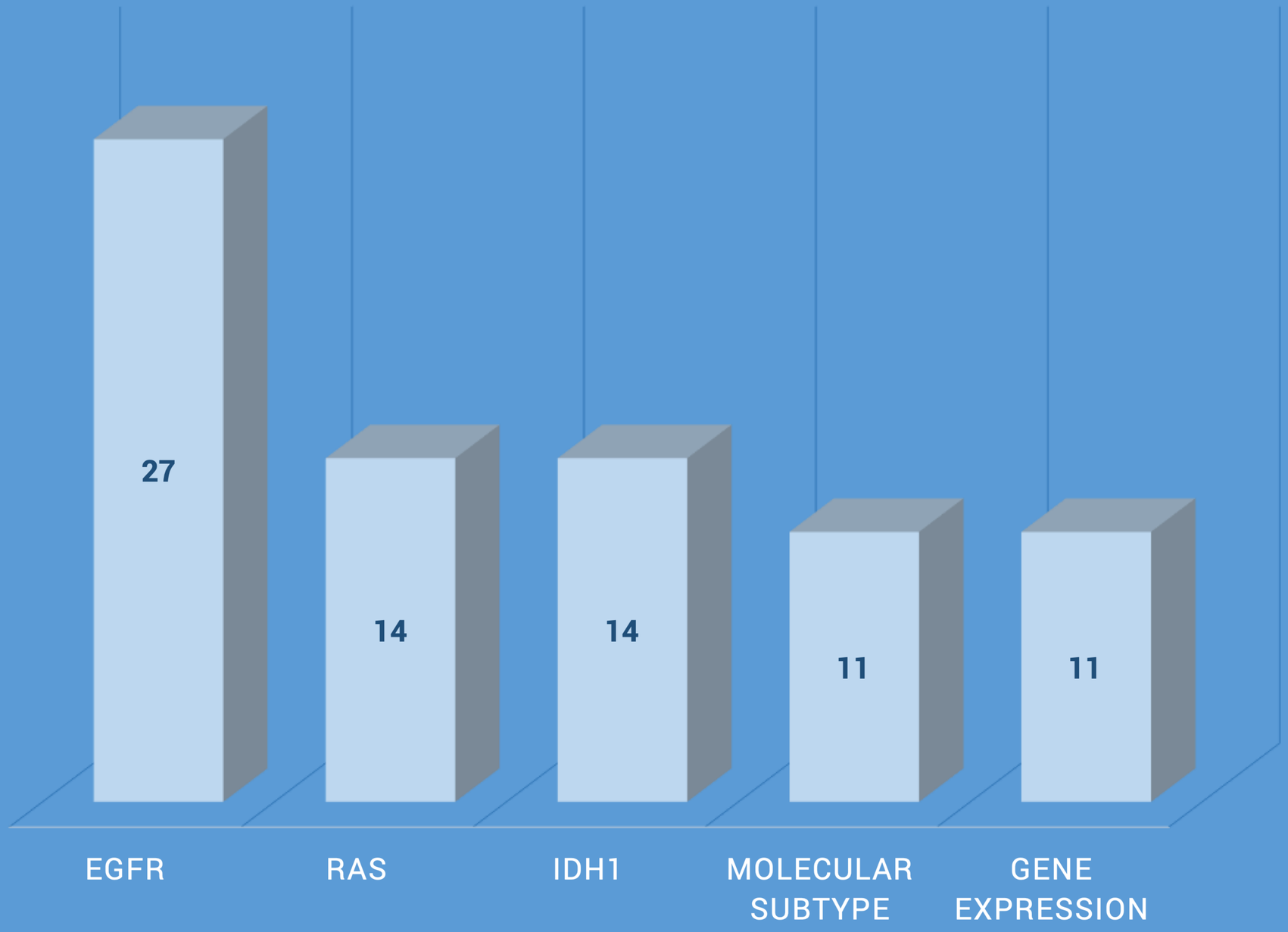
Machine learning is still in the minority



Learn more

Bodalal et al, Abdominal Radiology 2019

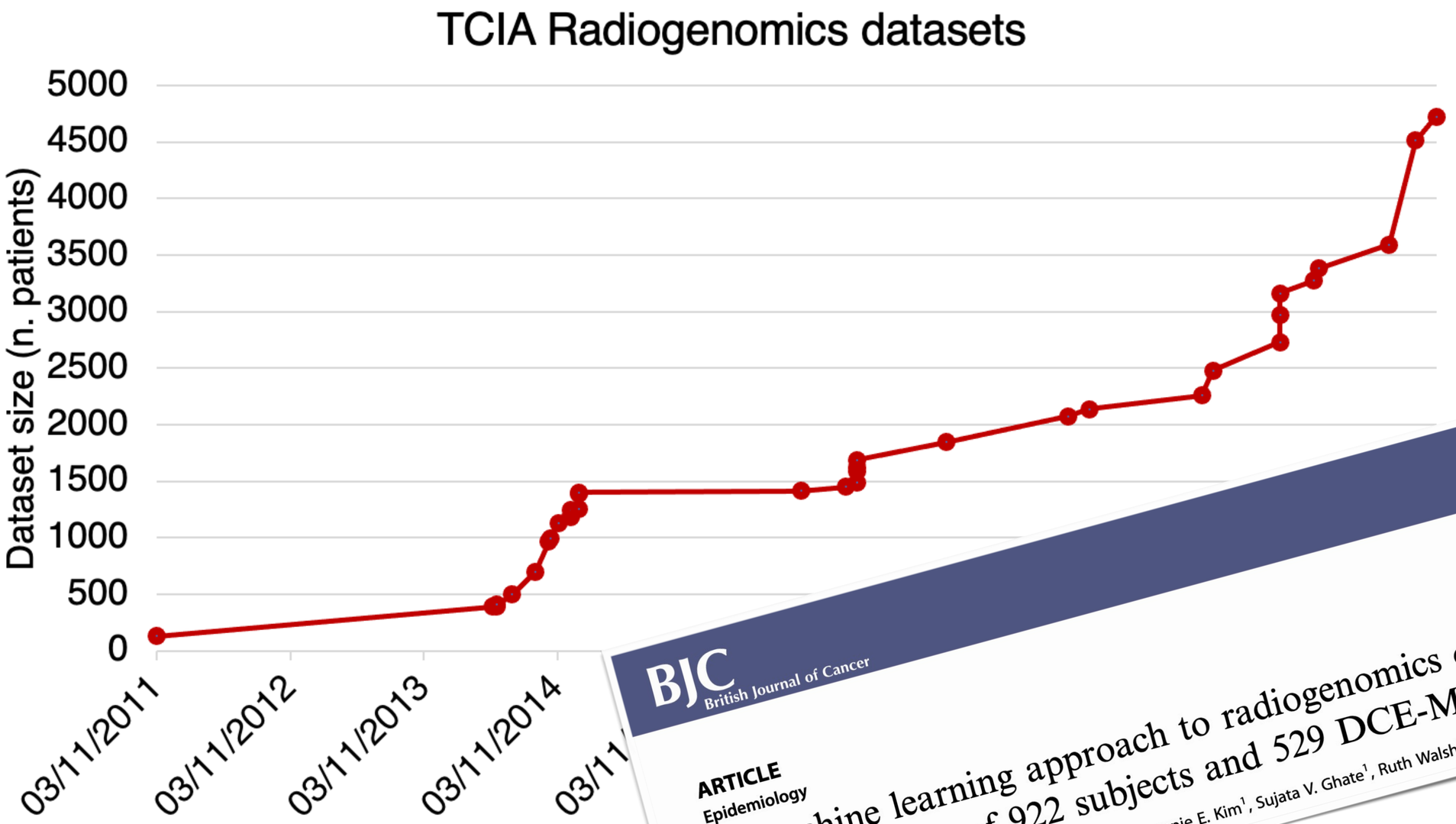
What types of genomic features?

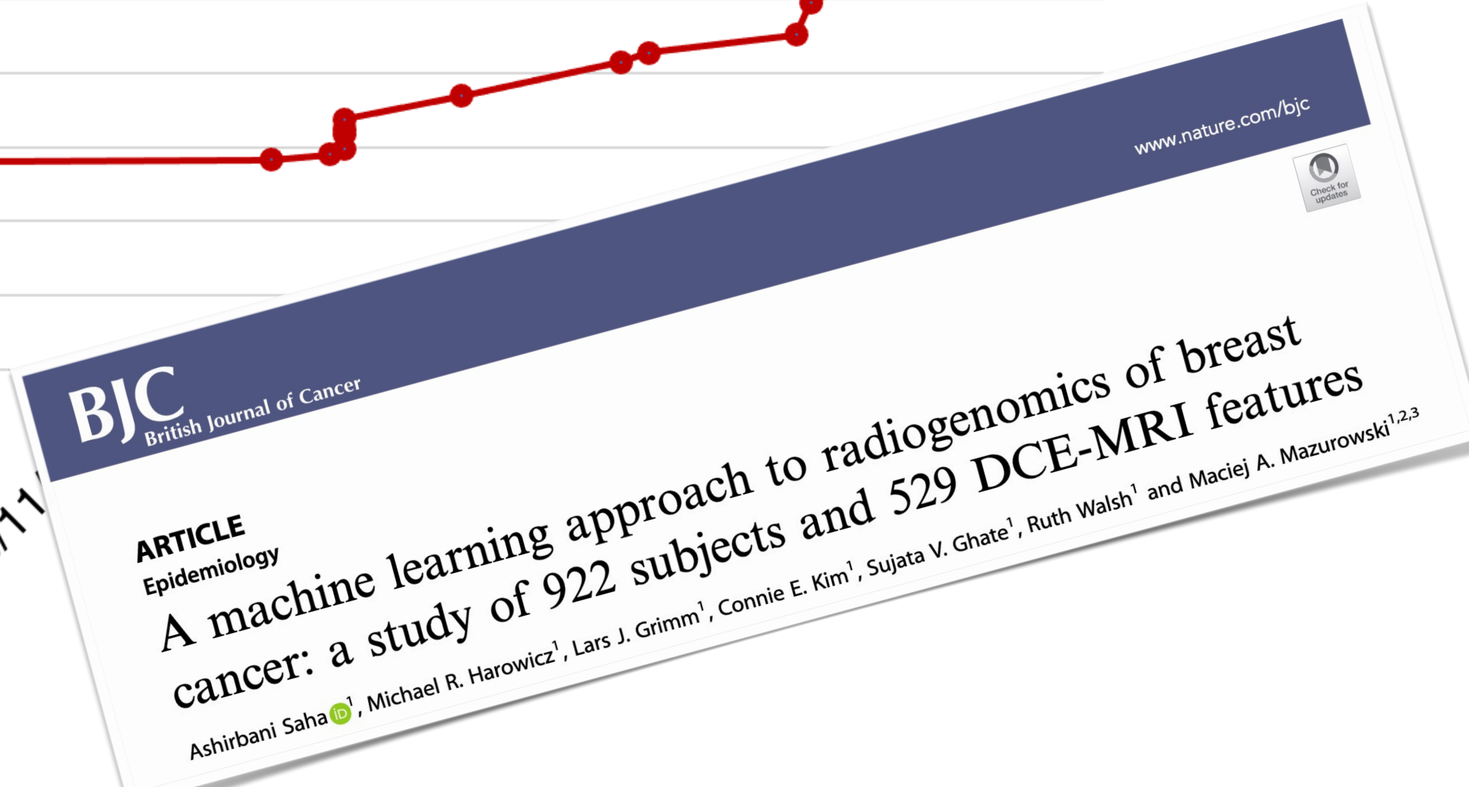


Learn more

Bodalal et al, Abdominal Radiology 2019

Radiogenomic datasets





Radiogenomics of breast cancer

Table 1. Prior studies reporting association of breast MR imaging features and genomic characteristics

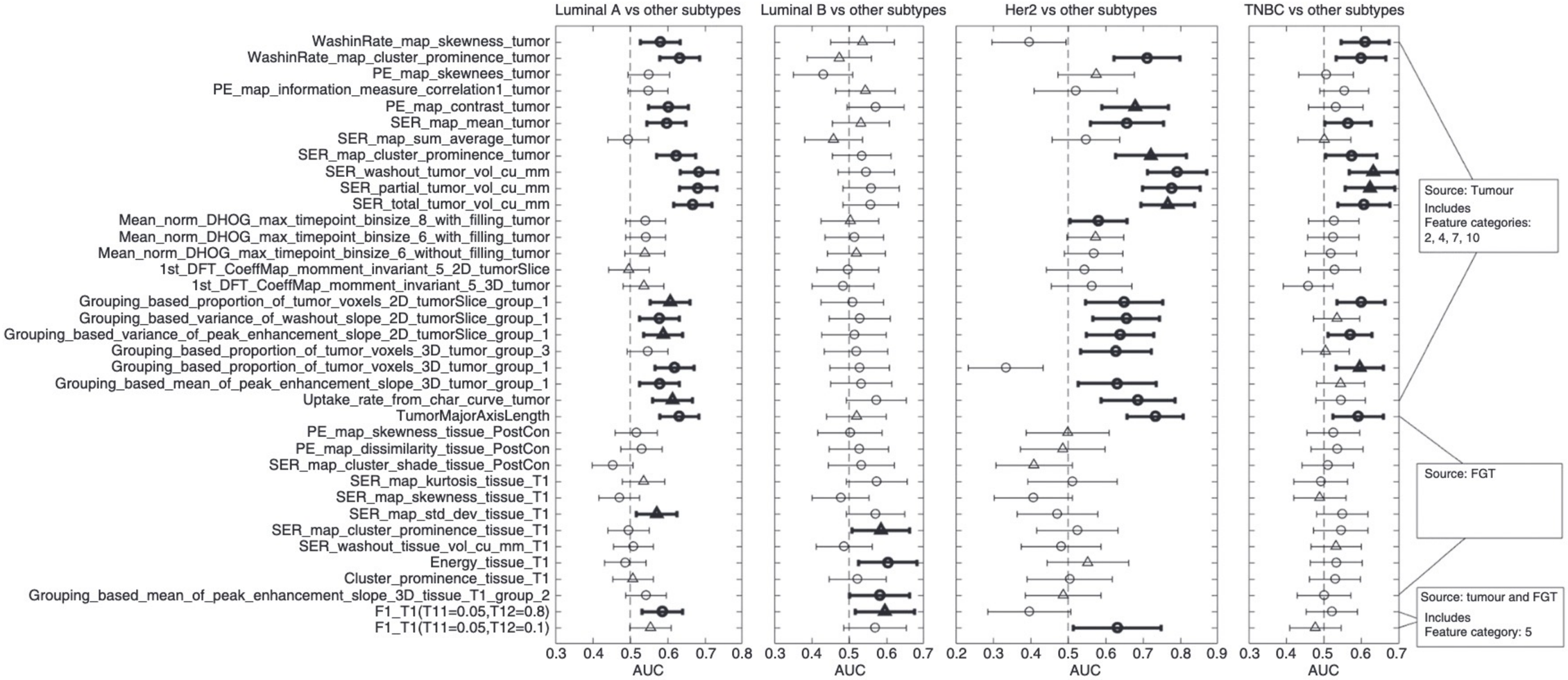
First author, year and reference	Number of imaging features from MR	Number of patients (dataset information: S for single and M—for multiple institutions)	Breast cancer related principal research question
Uematsu et al.[9]	9	176 (S)	Correlation of Imaging features and pathologic findings in TNBC and non-TNBC
Costantini et al.[14]	14	225 (S)	Comparison of imaging features of TNBC and non-TNBC
Yamamoto et al.[7]	26	10 (M)	Association of imaging features and interferon breast cancer subtype
Sung et al.[8]	7	321 (S)	Comparison of imaging features of TNBC and non-TNBC
Agner et al.[13]	120	65 (S)	Imaging features for predicting TNBC and other cancers
Mazurowski et al.[15]	23	48 (M)	Association of imaging features and molecular subtypes
Blaschke et al.[10]	6	112 (S)	Association of imaging features and molecular subtypes
Grimm et al.[6]	56	275 (S)	Association of imaging features and molecular subtypes
Guo et al.[12]	38	91 (M)	Integrated radiomics and genomics data to predict clinical phenotypes
Wang et al.[11]	85	84 (M)	Association of imaging features of background parenchymal enhancement and TNBC
Yamaguchi et al.[16]	5	186 (S)	Association of imaging features and molecular subtypes
Li et al.[17]	37	91 (M)	Association of imaging features and molecular subtypes
Fan et al.[3]	88	96 (S)	Association of imaging features and molecular subtypes
Wu et al.[18]	35	210 (M)	Association of imaging features and molecular subtypes

TNBC triple negative breast cancer

Learn more

Saha et al., BJC 2018

Radiogenomics of breast cancer



Learn more

Saha et al., BJC 2018

Radiogenomics of breast cancer

Table 4. AUC, CI, and *p*-values obtained for the trained multivariate models in the test set

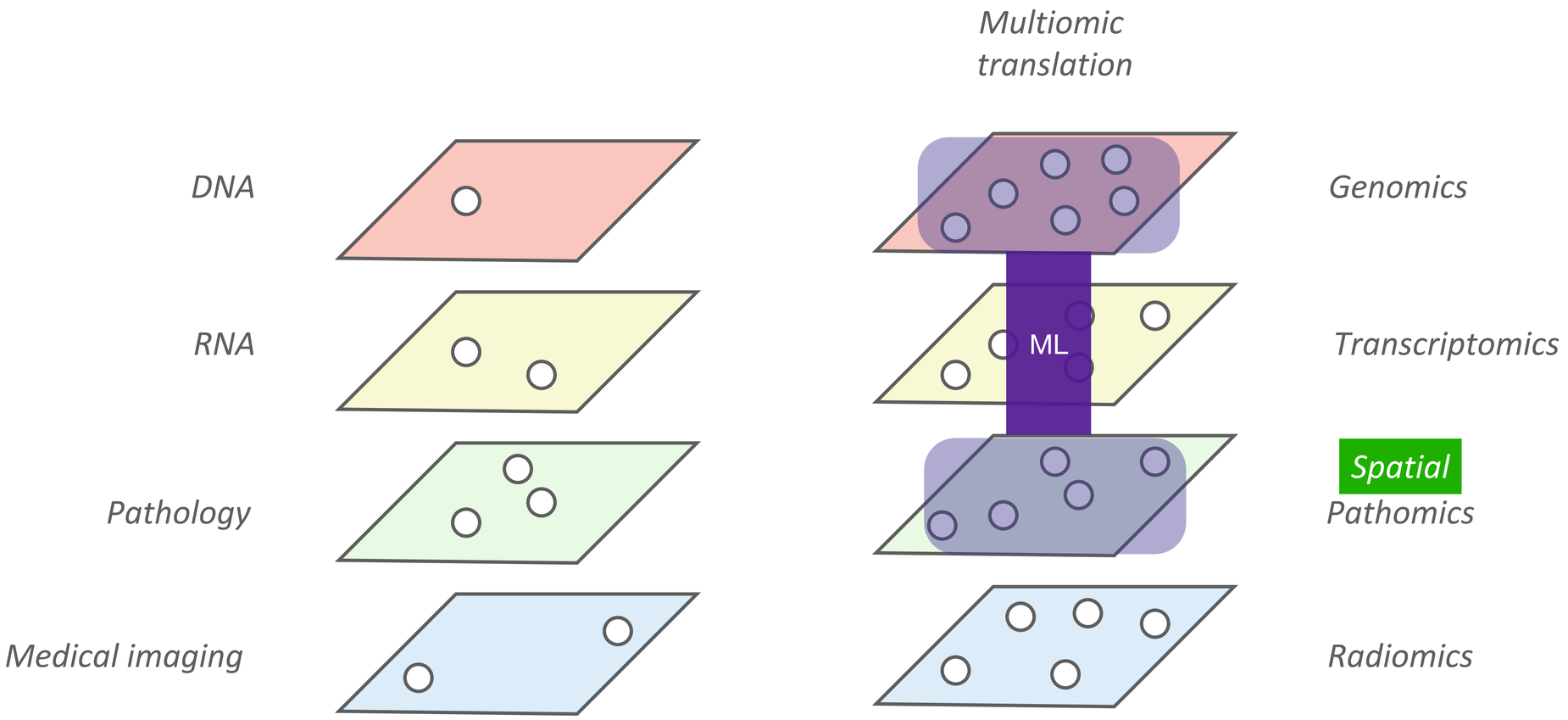
Name of the Task	AUC in test set with 95% CI	*p*-Value for the model in test set
Luminal A vs other subtypes	0.697 (0.647–0.746)	<0.0001*
Luminal B vs other subtypes	0.566 (0.494–0.638)	0.13
HER2 vs other subtypes	0.633 (0.539–0.727)	0.03
TNBC vs other subtypes	0.654 (0.589–0.720)	<0.0001*
ER positivity vs ER negativity	0.649 (0.591–0.705)	<0.0001*
PR positivity vs PR negativity	0.622 (0.569–0.674)	<0.001*
HER2 positivity vs HER2 negativity	0.500 (0.433–0.567)	0.81
High Ki-67 vs low Ki-67	0.624 (0.531–0.718)	0.01

AUC area under the curve, CI confidence interval, ER oestrogen receptor, HER2 human epidermal growth factor, PR progesterone receptor. *Statistically significant $p < 0.00625$

Learn more

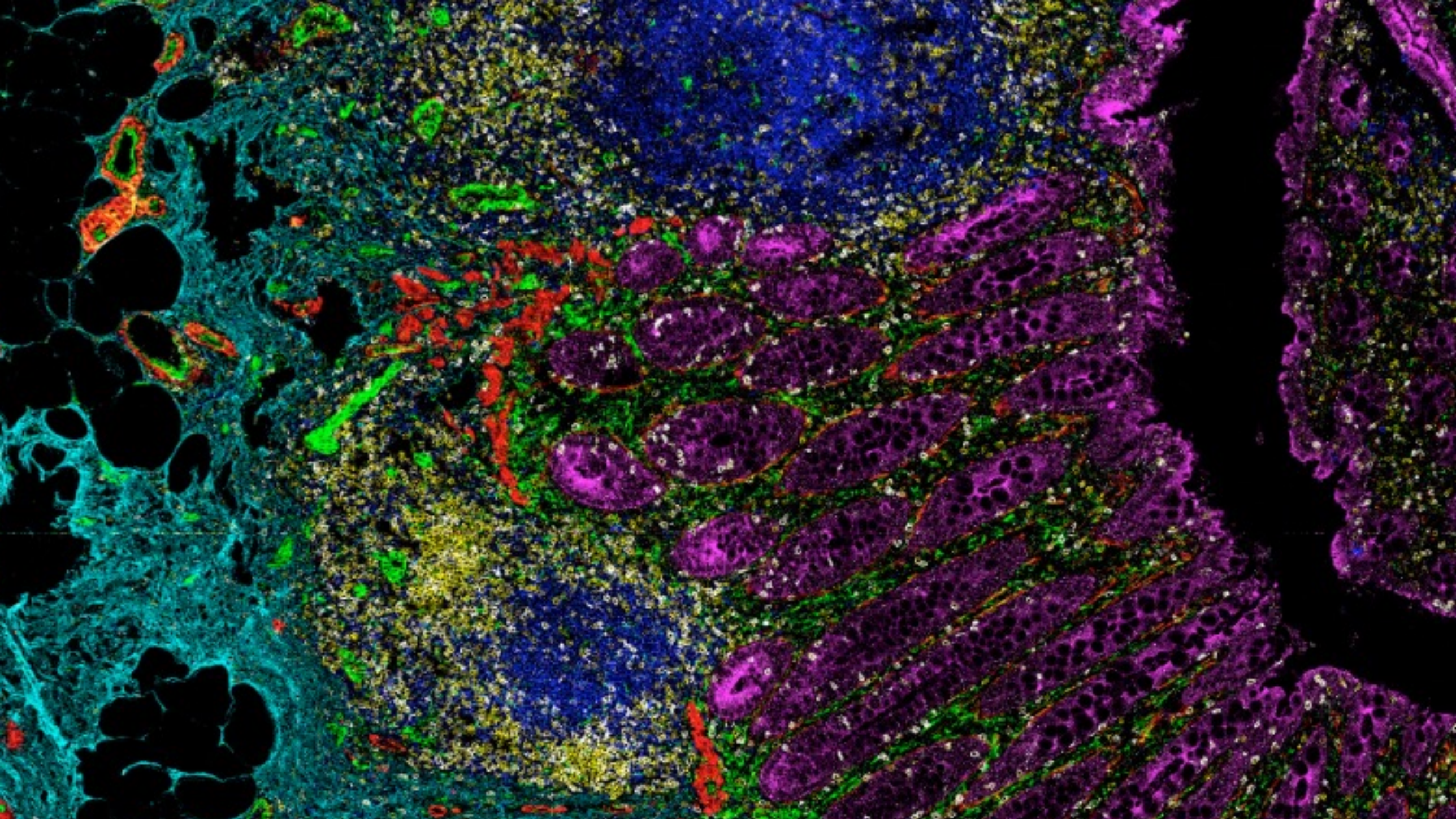
Saha et al., BJC 2018

Multi-omic translation

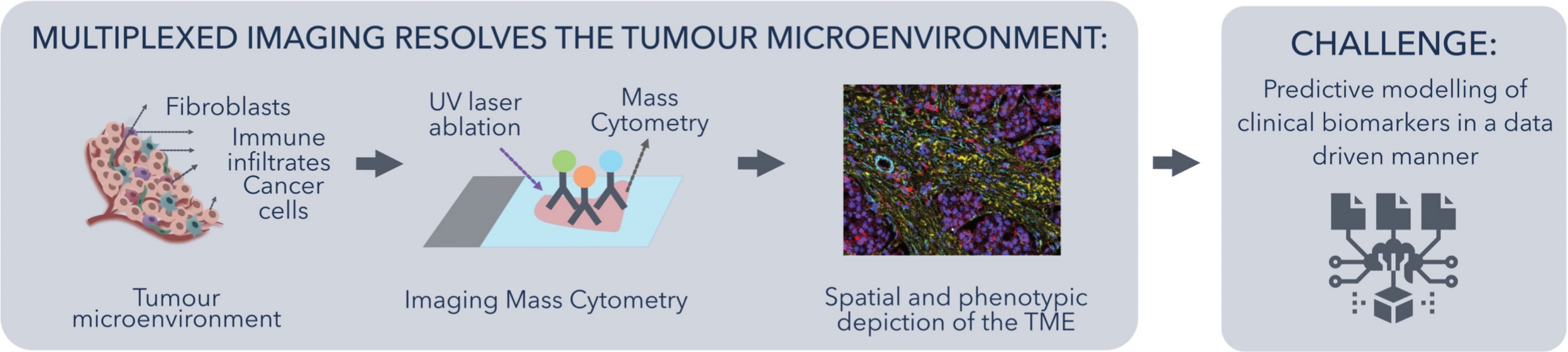


Learn more

Martin-Gonzalez et al, IMMIC @ MICCAI 2021



Spatial transcriptomics

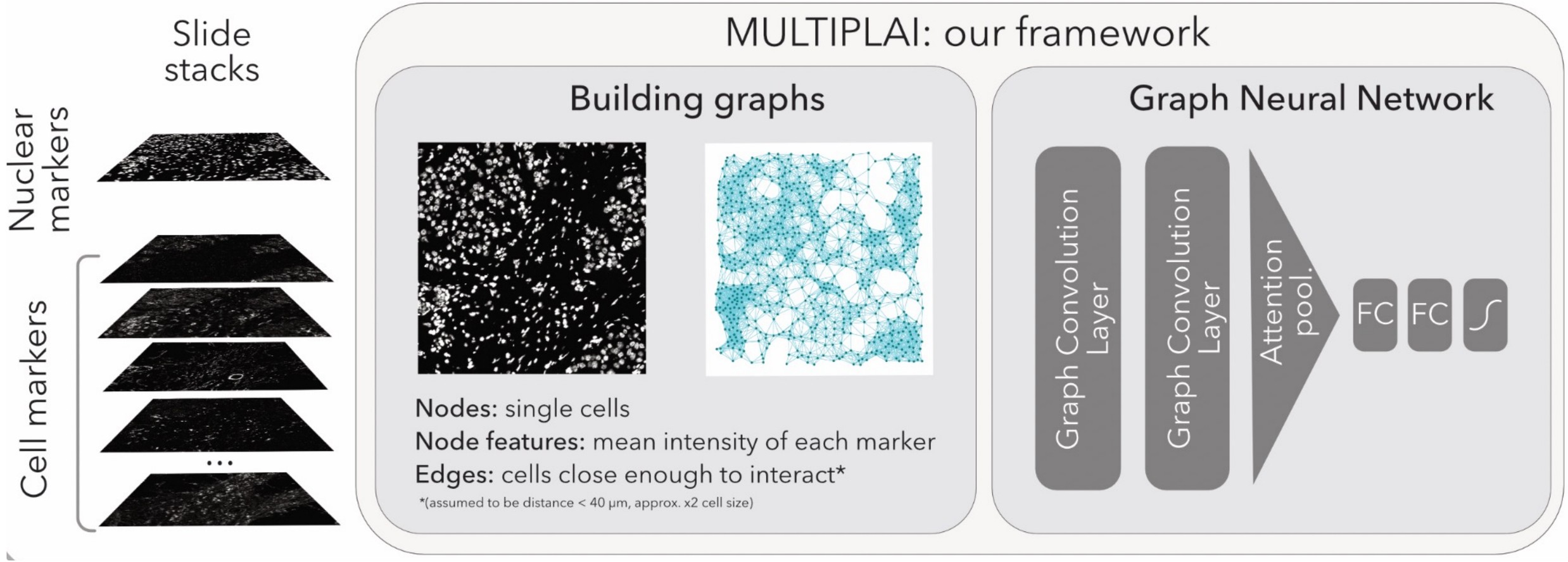


Learn more

Martin-Gonzalez et al, IMMIC @ MICCAI 2021

Spatial transcriptomics

MULTIPLAI (Multiplexed Pathology AI)



Learn more

Martin-Gonzalez et al, IMMIC @ MICCAI 2021

Spatial transcriptomics

MULTIPLAI (Multiplexed Pathology AI)

Model	Architecture		Best model	
	Number of layers	Pooling	BCE Loss CB	Accuracy
PA Graph Kernel	N.A.		7.306	66.18
GNN	1	Mean	0.174	78.4
		Sum	0.353	85.72
		Attention	0.199	86.86
	2	Mean	0.162	86.00
		Sum	0.278	81.16
		Attention	**0.159**	**90.00**

Learn more

Martin-Gonzalez et al, IMMIC @ MICCAI 2021

Spatial transcriptomics

MULTIPLAI (Multiplexed Pathology AI)

b.

METABRIC set
University of Cambridge

n = 456 cores
One core per patient
Cores of 0.6 mm of diameter
43 proteins studied

Train (n= 240)	Development (n= 104)	Internal test (n= 112)

External test set (i): Basel
Basel Hospital

n = 281 cores
One core per patient
Cores of 0.8 mm diameter
38 proteins studied

External test set (ii): Zurich
Zurich Hospital

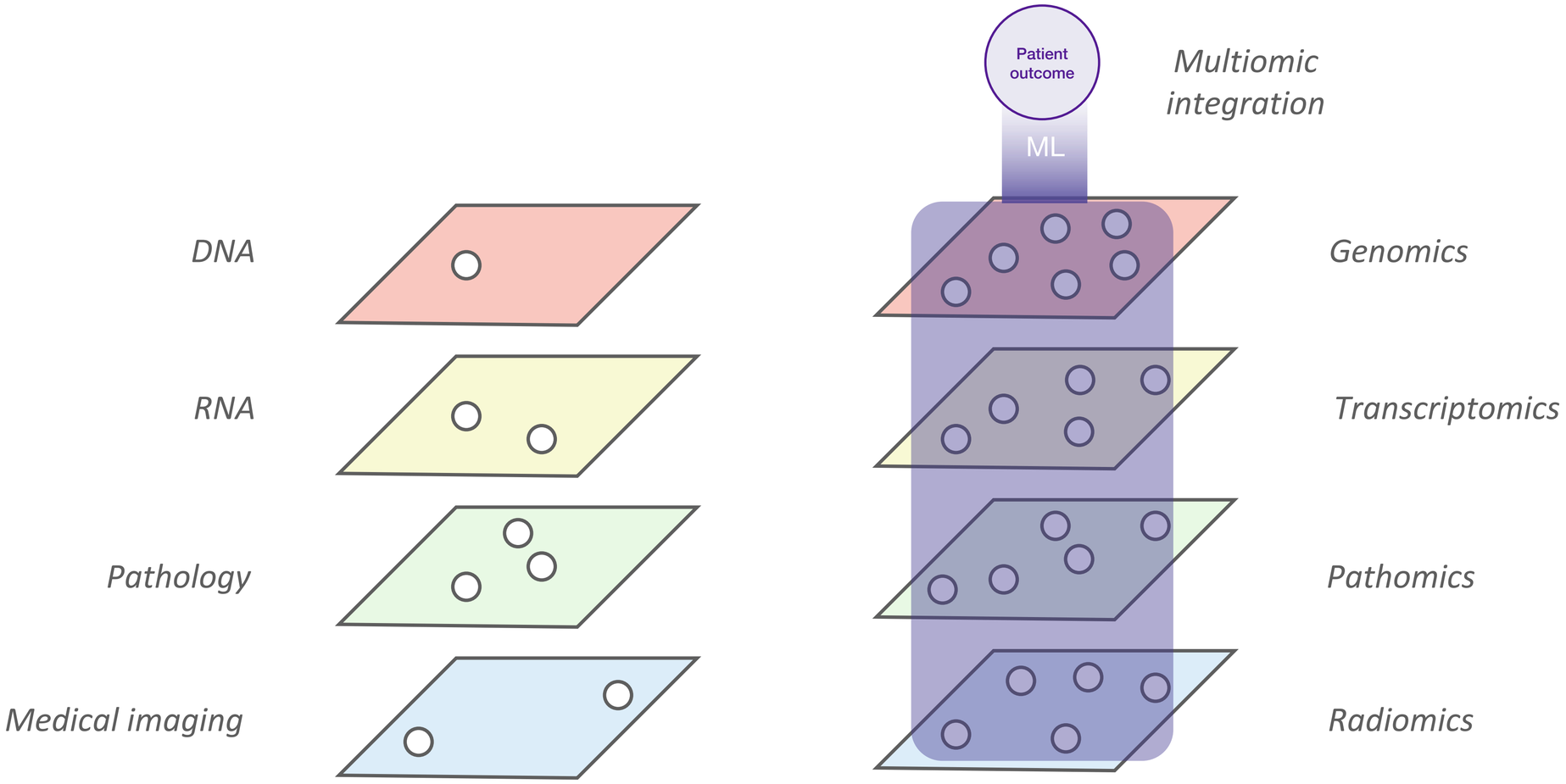
n = 350 cores
70 patients, multiple cores
Cores of 0.6 mm diameter
38 proteins studied

Class Balanced Accuracy		
	Strategy 1	Strategy 2
Internal test (Metabric)	78.96	**78.96**
External test (Basel)	60.16	**65.71**
External test (Zurich)	68.19	**75.10**

Learn more
Martin-Gonzalez et al, IMMIC @ MICCAI 2021

Multi-omics Integration

Multi-omics


Patient outcome
Multiomic integration
ML
DNA
RNA
Pathology
Medical imaging
Genomics
Transcriptomics
Pathomics
Radiomics

Types of prediction

Survival

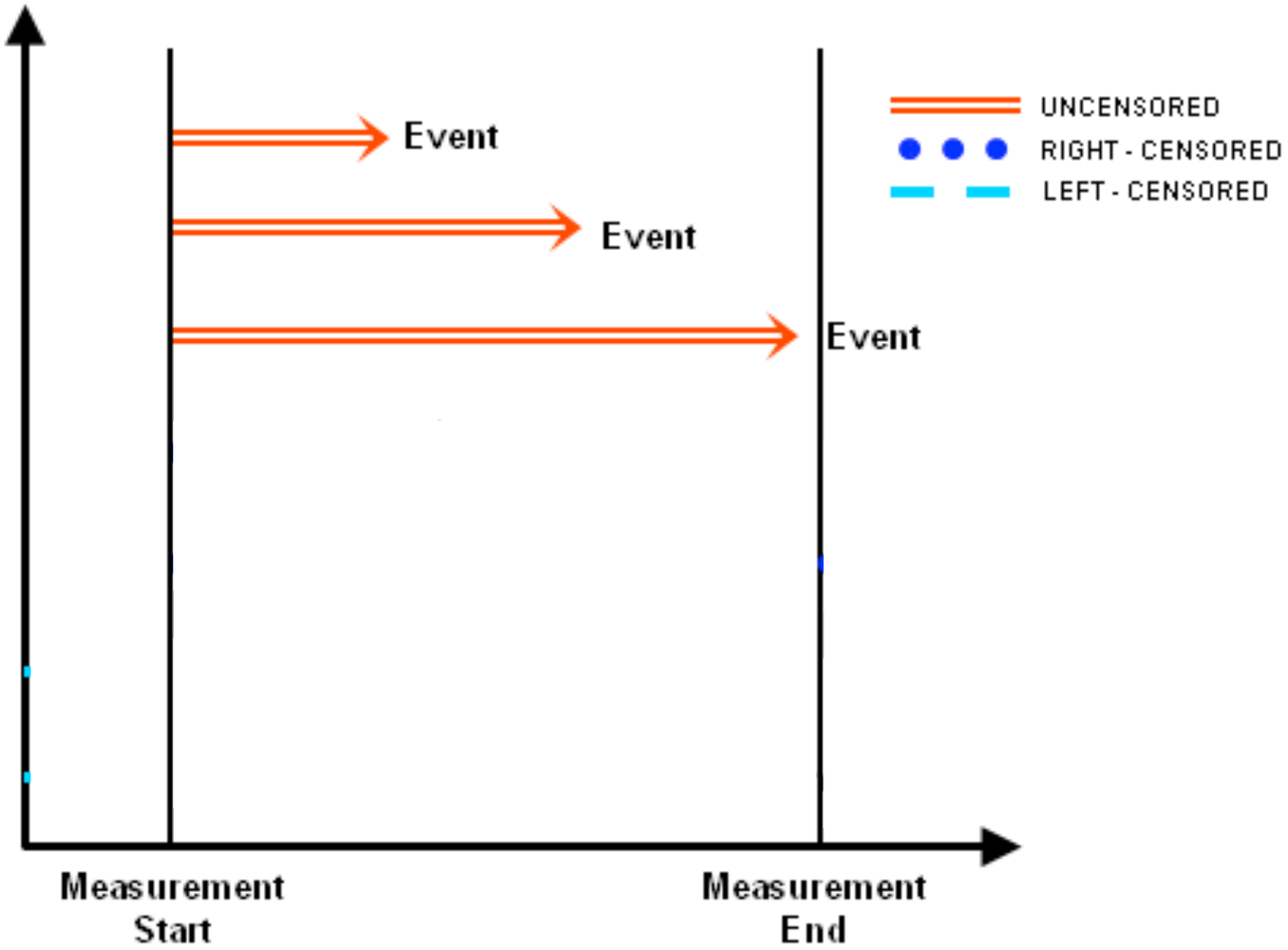


Image from WebFOCUS Stat

Types of prediction

Survival

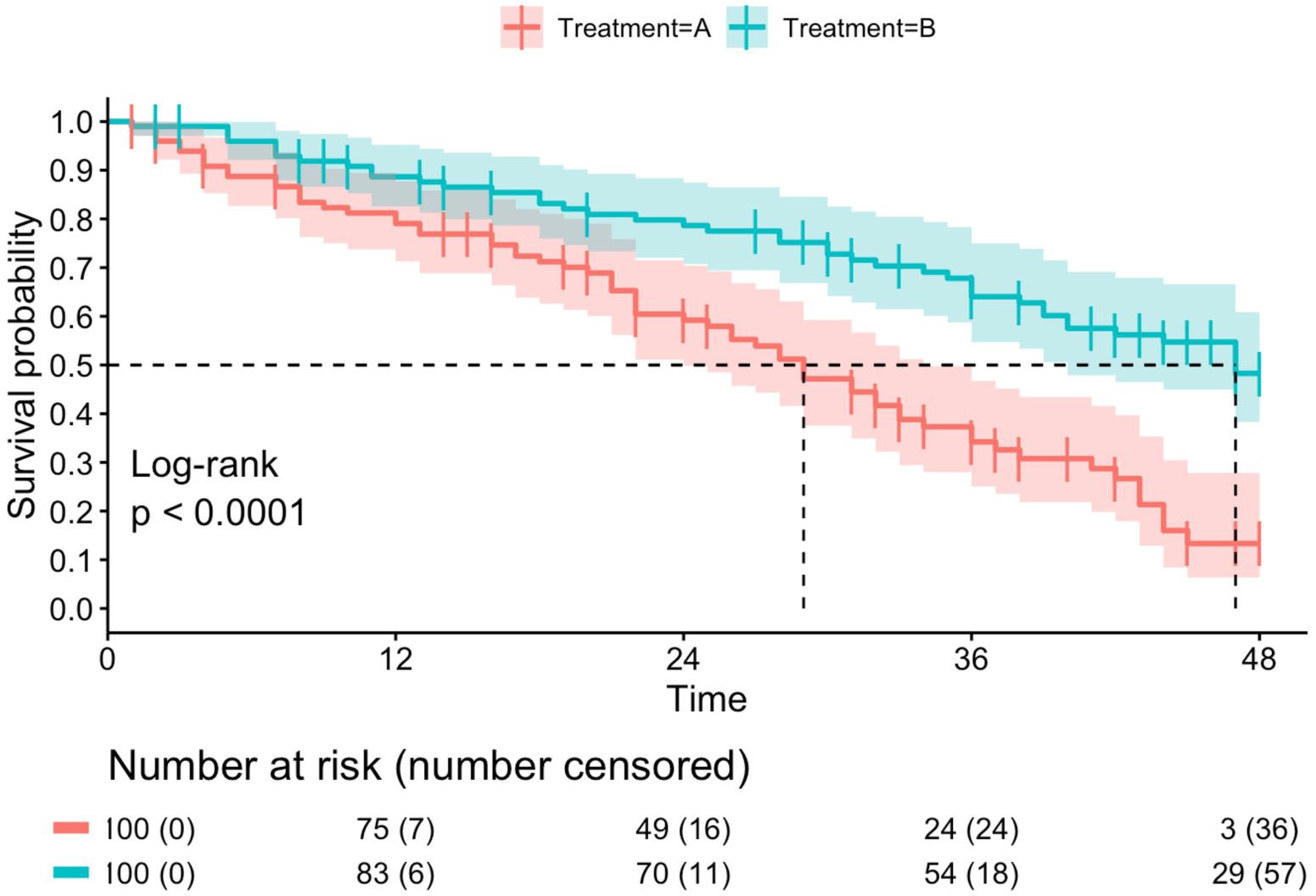
- The survival probability, also known as the survivor function S(t), is the probability that an individual survives from the time origin (e.g. diagnosis of cancer) to a specified future time t.
- The hazard, denoted by h(t), is the probability that an individual who is under observation at a time t has an event at that time.

$$h(t) = h_0(t) \times exp(b_1 x_1 + b_2 x_2 + \ldots + b_p x_p)$$

Cox proportional hazards

Types of prediction

Survival

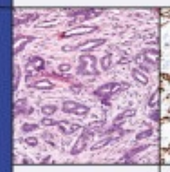
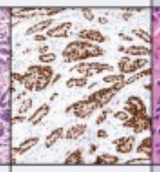
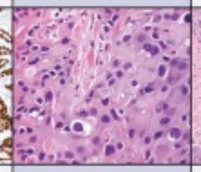
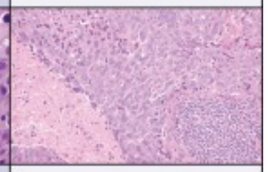
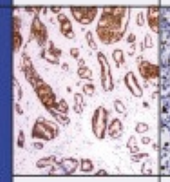
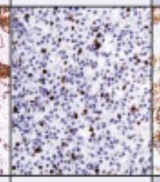
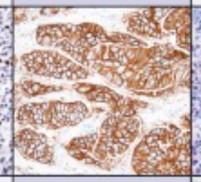
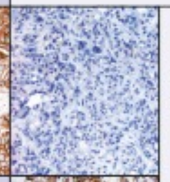
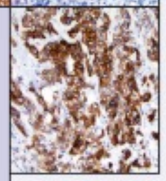


Kaplan-Meier curve

Image from Ruben van Paemel

Types of prediction

Cancer subtype

Molecular subtype	Luminal (A and B)		HER2	Basal
Genetic profile	↑Luminal CKs and ER-related genes (A>B) B↑ in proliferation-related genes		↑HER2-related genes	↑Basal CKs
Histologic correlates				
	A Lower-grade ER+	B Higher-grade ER+	High-grade, ± apocrine features	High-grade, sheet-like, necrosis inflammation
Surrogate markers				 ER/PR- HER2-
	A Strong ER+, PR±, HER2-, low Ki67	B Weaker ER+, PR±, HER2±, ↑Ki67	HER2+, ± ER/PR	 CK5/6± EGFR±
Prognosis	Good	Intermediate	Worse	Worse
Response to chemotherapy	Lower	Intermediate	Higher	Higher
Targeted therapies	Hormone therapies		HER2-targeted therapies	Currently investigational

Example: breast cancer

Learn more

K. H. Allison, Diag Mol Path, 2017

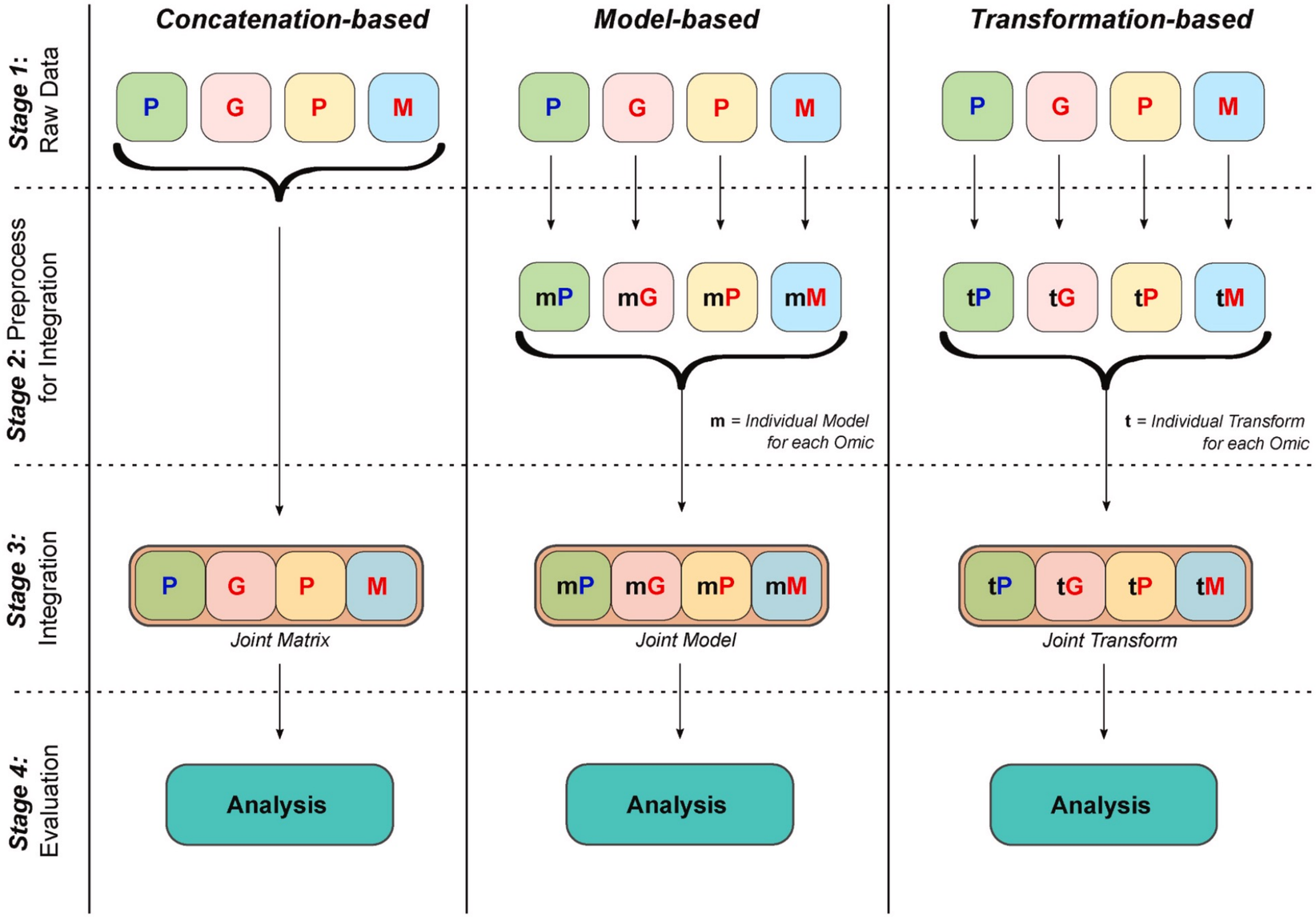
Types of prediction

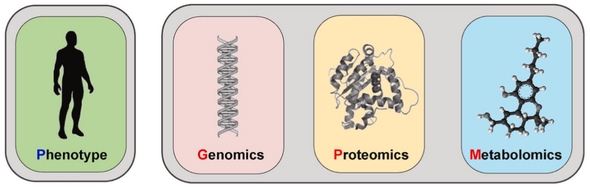
Response status

	RECIST 1.1
Complete Response	Disappearance of all target lesions or lymph nodes <1 mm in the short axis
Partial Response	>30% decrease in tumour size or >=15% decrease in tumour attenuation at CT, no new lesions
Progressive Disease	>20% increase of SPD of target lesions with an absolute increase of >=5mm, new lesions
Stable Disease	None of the above

Other options: pathologic complete response / residual tumour burden; chemotherapy response score; etc.

Integration approaches



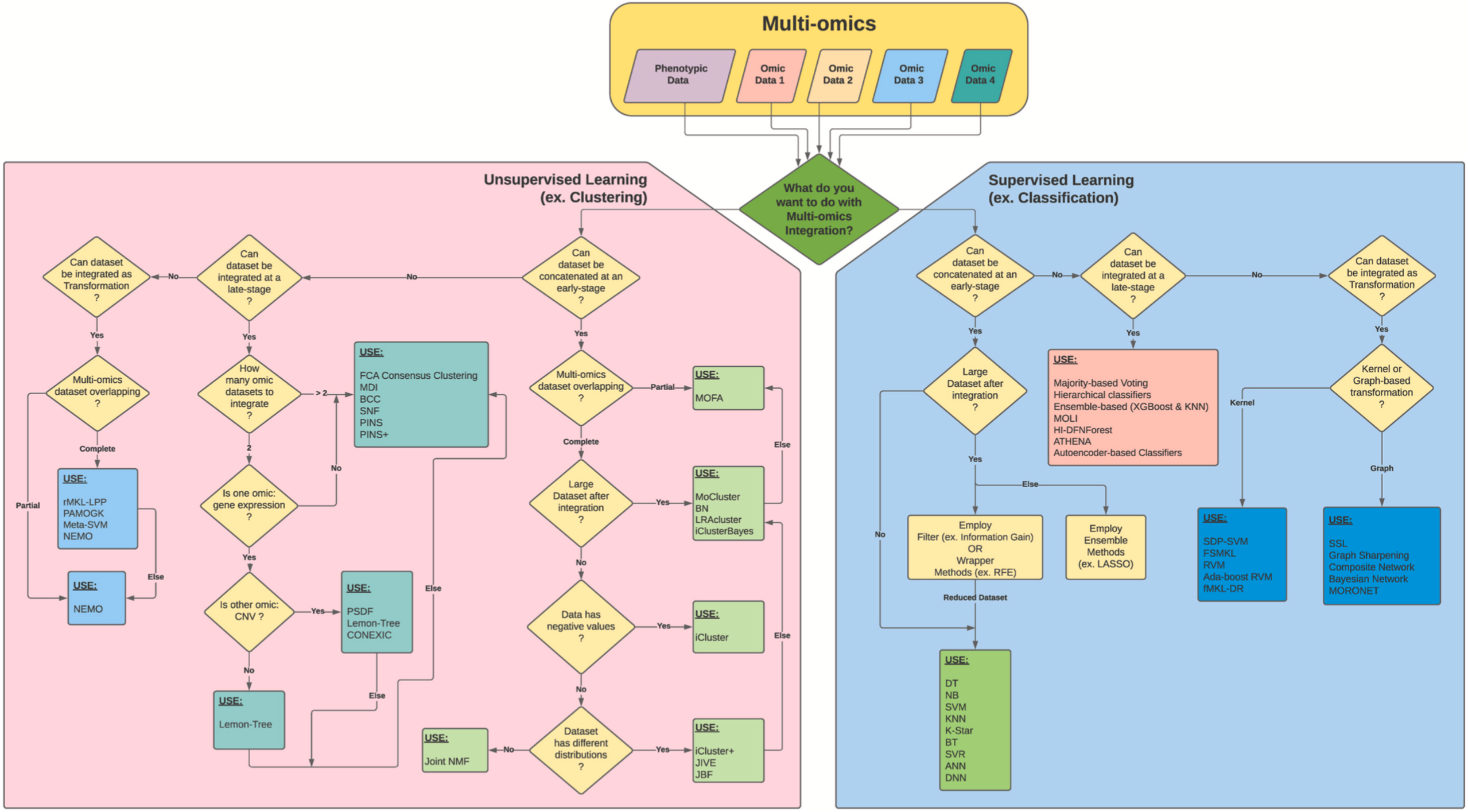


Learn more

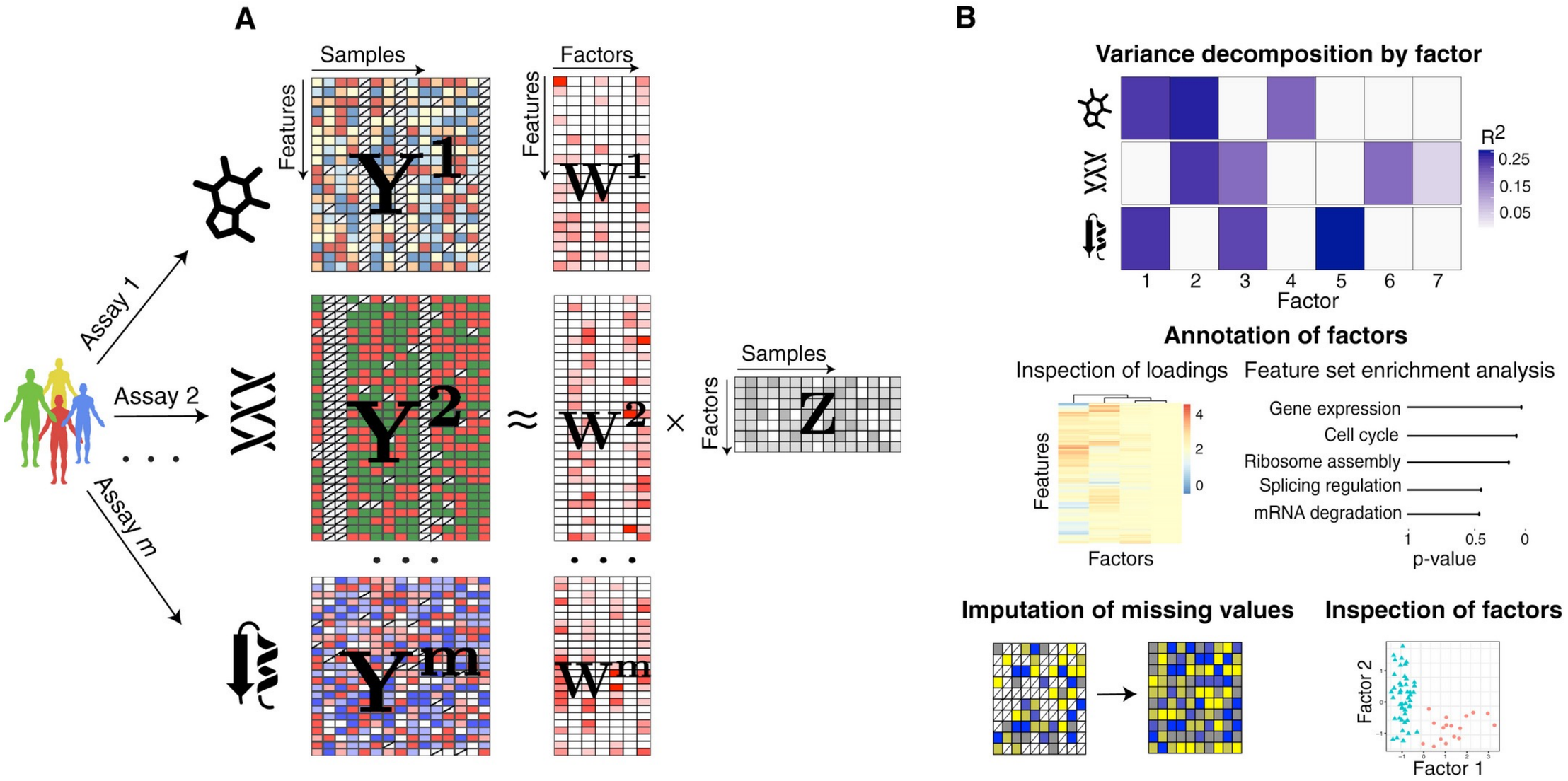
Reel et al, Biotech Adv 2020

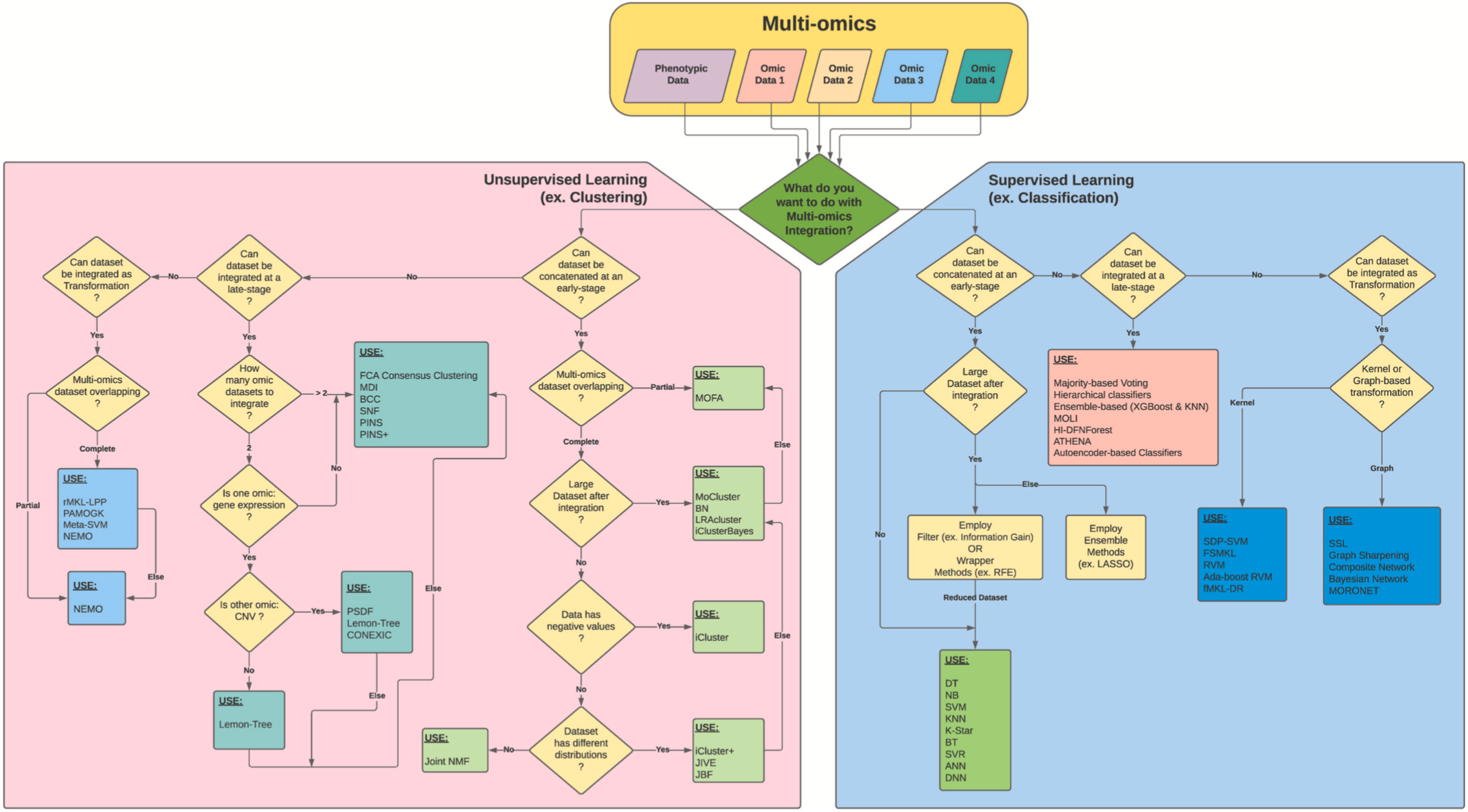
Integration algorithms

		Multi-omics Integration Methods		
Learning Type	Supervised	Concatenation-based • Classical ML (DT (Quinlan, 1993), NB (Domingos and Pazzani, 1997), ANN (Bishop, 1995), SVM (Vapnik, 1995), KNN (Altman, 1992), K-Star (Cleary and Trigg, 1995) • LASSO (Zou, 2006; Nicolai and Peter, 2010; Mankoo et al., 2011) • BT (Elith et al., 2008) • SVR (Awad and Khanna, 2015) • DNN (Tang et al., 2019)	Model-based • Majority-based voting (Drăghici and Potter, 2003) • Hierarchical Classifiers (Bavafaye Haghighi et al., 2019) • Ensemble-based classifiers (XGBoost (Ma et al., 2020a) and KNN (Shen and Chou, 2006)) • MOLI (Sharifi-Noghabi et al., 2019) • HI-DFNForest (Xu et al., 2019a) • ATHENA (Chung and Kang, 2019; Holzinger et al., 2014)	Transformation-based • SDP-SVM (Lanckriet et al., 2004) • FSMKL (Seoane et al., 2014) • RVM (Bowd et al., 2005; Tipping, 2001) • Ada-boost RVM (Wu et al., 2010) • fMKL-DR (Giang et al., 2020) • SSL (Tsuda et al., 2005; Culp and Michailidis, 2008; Kim et al., 2015; Yue et al., 2017; Bhardwaj and Van Steen, 2020), • Graph sharpening (Shin et al., 2010, Shin et al., 2007) • Composite network (Mostafavi and Morris, 2010) • BN (Rhodes et al., 2005) • MORONET (Wang et al., 2020)
	Unsupervised	• Joint NMF (Zhang et al., 2012) • iCluster (Shen et al., 2009) • iCluster+ (Mo et al., 2013) • JIVE (Lock et al., 2013) • JBF (Ray et al., 2014) • BN (Fridley et al., 2012; Zhu et al., 2012) • MoCluster (Meng et al., 2016a) • iClusterBayes (Mo et al., 2018) • MOFA (Argelaguet et al., 2018)	• PSDF (Yuan et al., 2011) • FCA consensus clustering (Hristoskova et al., 2014) • MDI (Kirk et al., 2012) • BCC (Lock and Dunson, 2013) • Lemon-Tree (Bonnet et al., 2015) • SNF (Wang et al., 2014)	• rMKL-LPP (Speicher and Pfeifer, 2015) • PAMOGK (Tepeli et al., 2019) • Meta-SVM (Kim et al., 2017) • NEMO (Rappoport and Shamir, 2019)


Multi-omics
Phenotypic Data
Omic Data 1
Omic Data 2
Omic Data 3
Omic Data 4
What do you want to do with Multi-omics Integration?
Unsupervised Learning (ex. Clustering)
Can dataset be integrated as Transformation ?
No
Yes
Multi-omics dataset overlapping ?
Complete
Partial
USE:
rMKL-LPP
PAMOGK
Meta-SVM
NEMO
Else
USE:
NEMO
Can dataset be integrated at a late-stage ?
No
Yes
How many omic datasets to integrate ?
> 2
2
USE:
FCA Consensus Clustering
MDI
BCC
SNF
PINS
PINS+
No
Is one omic: gene expression ?
Yes
Is other omic: CNV ?
Yes
USE:
PSDF
Lemon-Tree
CONEXIC
No
USE:
Lemon-Tree
Else
Else
Can dataset be concatenated at an early-stage ?
Yes
Multi-omics dataset overlapping ?
Partial
USE:
MOFA
Complete
Large Dataset after integration ?
Yes
USE:
MoCluster
BN
LRAcluster
iClusterBayes
No
Data has negative values ?
Yes
USE:
iCluster
No
Dataset has different distributions ?
Yes
USE:
iCluster+
JIVE
JBF
No
USE:
Joint NMF
Else
Else
Supervised Learning (ex. Classification)
Can dataset be concatenated at an early-stage ?
No
Yes
Large Dataset after integration ?
Yes
Employ Filter (ex. Information Gain) OR Wrapper Methods (ex. RFE)
Else
Employ Ensemble Methods (ex. LASSO)
Reduced Dataset
No
USE:
DT
NB
SVM
KNN
K-Star
BT
SVR
ANN
DNN
Can dataset be integrated at a late-stage ?
No
Yes
USE:
Majority-based Voting
Hierarchical classifiers
Ensemble-based (XGBoost & KNN)
MOLI
HI-DFNForest
ATHENA
Autoencoder-based Classifiers
Can dataset be integrated as Transformation ?
Yes
Kernel or Graph-based transformation ?
Kernel
USE:
SDP-SVM
FSMKL
RVM
Ada-boost RVM
fMKL-DR
Graph
USE:
SSL
Graph Sharpening
Composite Network
Bayesian Network
MORONET

MOFA

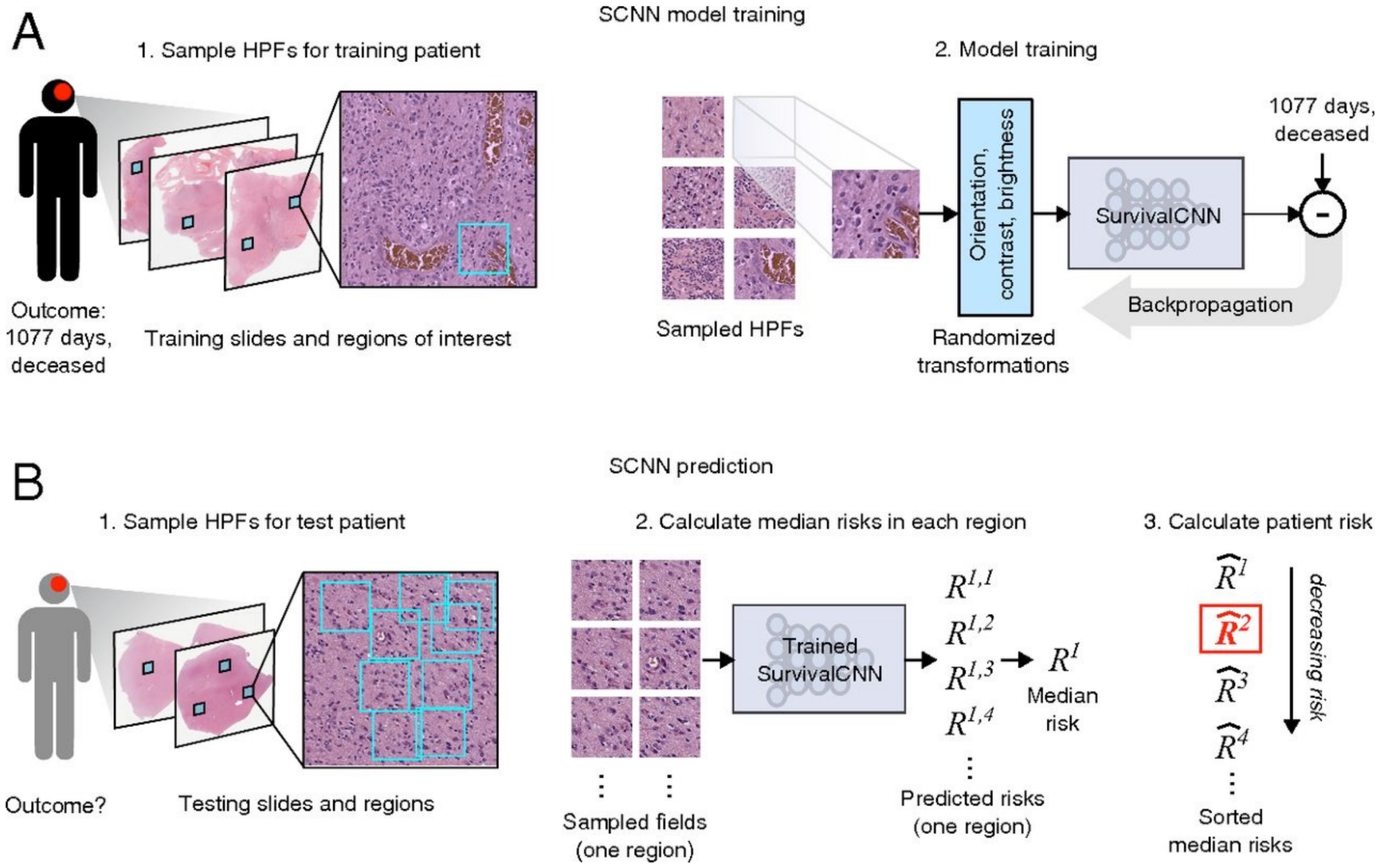

A
Samples
Features
Factors
Assay 1
Assay 2
Assay m
Y1
W1
Y2
W2
Ym
Wm
Z
B
Variance decomposition by factor
R2
0.25
0.15
0.05
1 2 3 4 5 6 7
Factor
Annotation of factors
Inspection of loadings
Feature set enrichment analysis
Features
Factors
Gene expression
Cell cycle
Ribosome assembly
Splicing regulation
mRNA degradation
p-value
Imputation of missing values
Inspection of factors
Factor 2
Factor 1


Multi-omics
Phenotypic Data
Omic Data 1
Omic Data 2
Omic Data 3
Omic Data 4
What do you want to do with Multi-omics Integration?
Unsupervised Learning (ex. Clustering)
Can dataset be integrated as Transformation ?
No
Yes
Multi-omics dataset overlapping ?
Complete
Partial
USE:
rMKL-LPP
PAMOGK
Meta-SVM
NEMO
Else
USE:
NEMO
Can dataset be integrated at a late-stage ?
No
Yes
How many omic datasets to integrate ?
> 2
2
USE:
FCA Consensus Clustering
MDI
BCC
SNF
PINS
PINS+
No
Is one omic: gene expression ?
Yes
Is other omic: CNV ?
Yes
USE:
PSDF
Lemon-Tree
CONEXIC
No
USE:
Lemon-Tree
Else
Else
Can dataset be concatenated at an early-stage ?
Yes
Multi-omics dataset overlapping ?
Partial
USE:
MOFA
Complete
Large Dataset after integration ?
Yes
USE:
MoCluster
BN
LRAcluster
iClusterBayes
No
Data has negative values ?
Yes
USE:
iCluster
No
Dataset has different distributions ?
Yes
USE:
iCluster+
JIVE
JBF
No
USE:
Joint NMF
Else
Else
Supervised Learning (ex. Classification)
Can dataset be concatenated at an early-stage ?
No
Yes
Large Dataset after integration ?
Yes
Employ Filter (ex. Information Gain) OR Wrapper Methods (ex. RFE)
Else
Employ Ensemble Methods (ex. LASSO)
Reduced Dataset
No
USE:
DT
NB
SVM
KNN
K-Star
BT
SVR
ANN
DNN
Can dataset be integrated at a late-stage ?
No
Yes
USE:
Majority-based Voting
Hierarchical classifiers
Ensemble-based (XGBoost & KNN)
MOLI
HI-DFNForest
ATHENA
Autoencoder-based Classifiers
Can dataset be integrated as Transformation ?
Yes
Kernel or Graph-based transformation ?
Kernel
USE:
SDP-SVM
FSMKL
RVM
Ada-boost RVM
fMKL-DR
Graph
USE:
SSL
Graph Sharpening
Composite Network
Bayesian Network
MORONET

Pathomics + Genomics

Patho-genomics to predict survival

Mobadersany et al, 2018

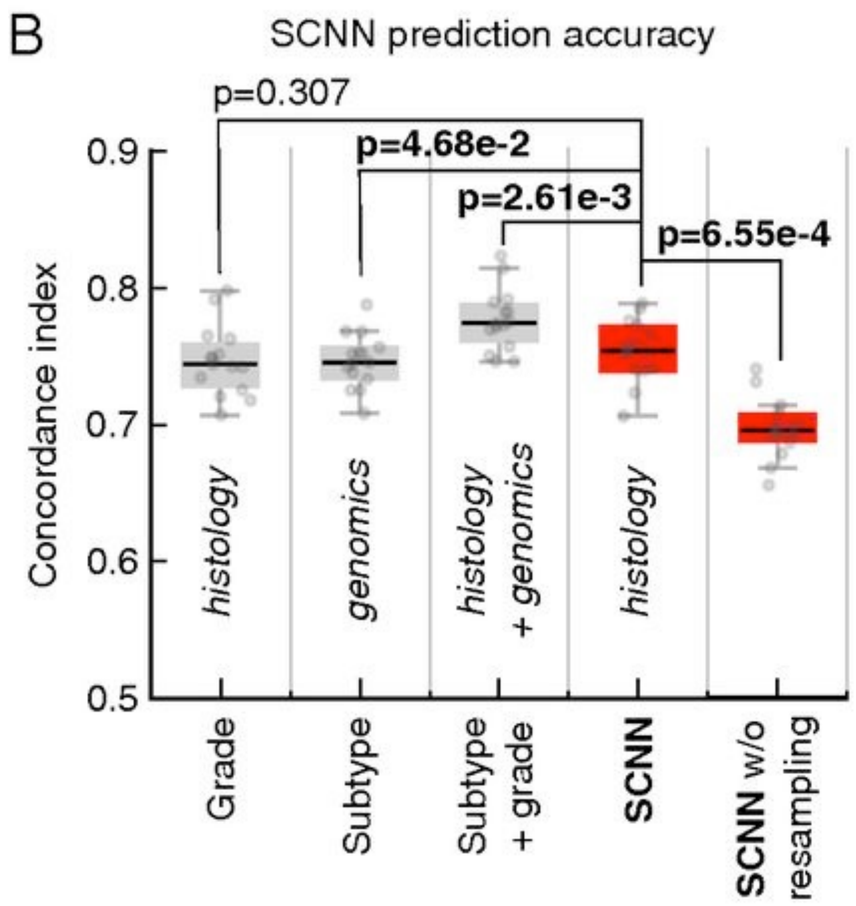


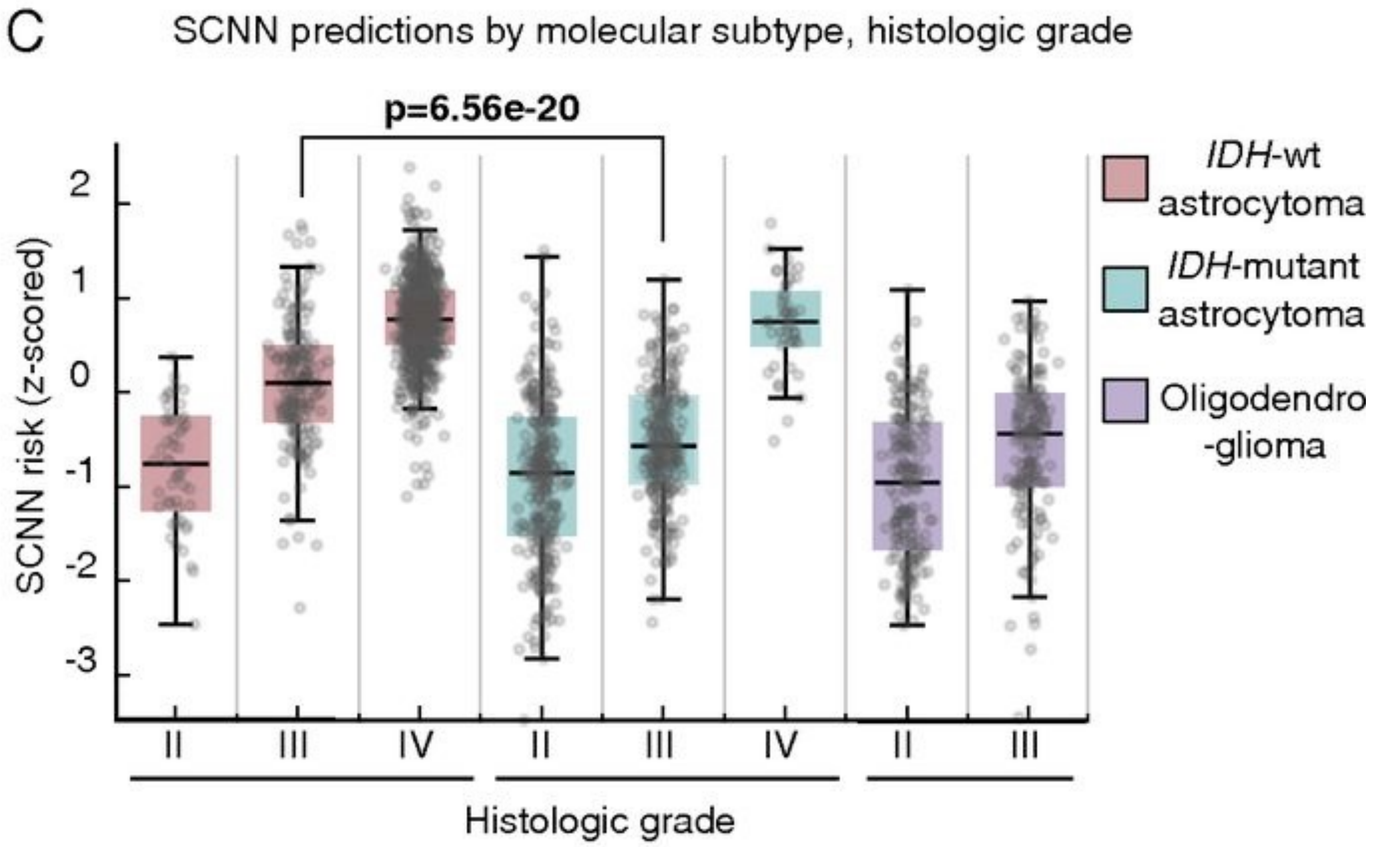
Learn more

P. Mobadersany et al., PNAS 2018

Patho-genomics to predict survival

Mobadersany et al, 2018



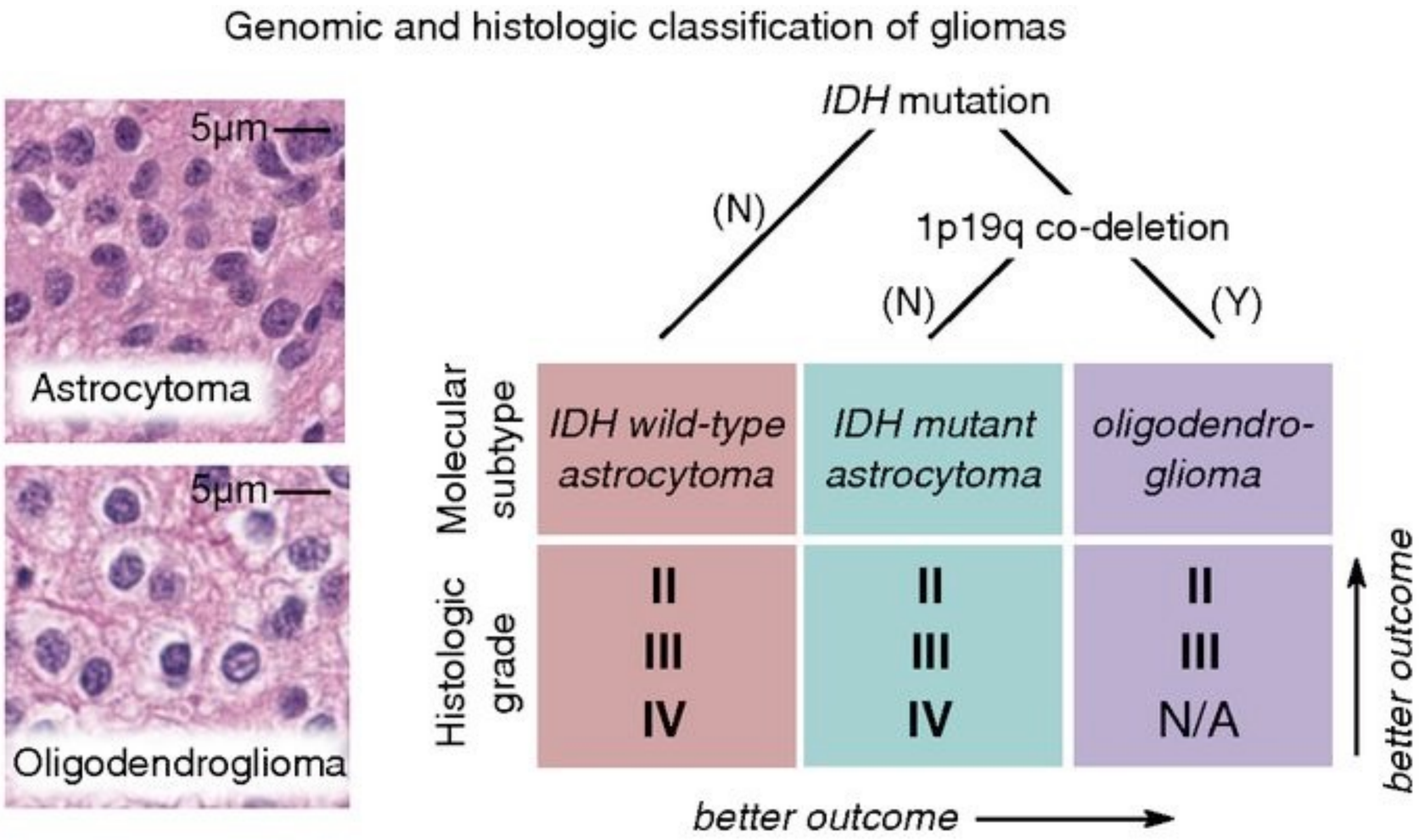


Learn more

P. Mobadersany et al., PNAS 2018

Patho-genomics to predict survival

Mobadersany et al, 2018

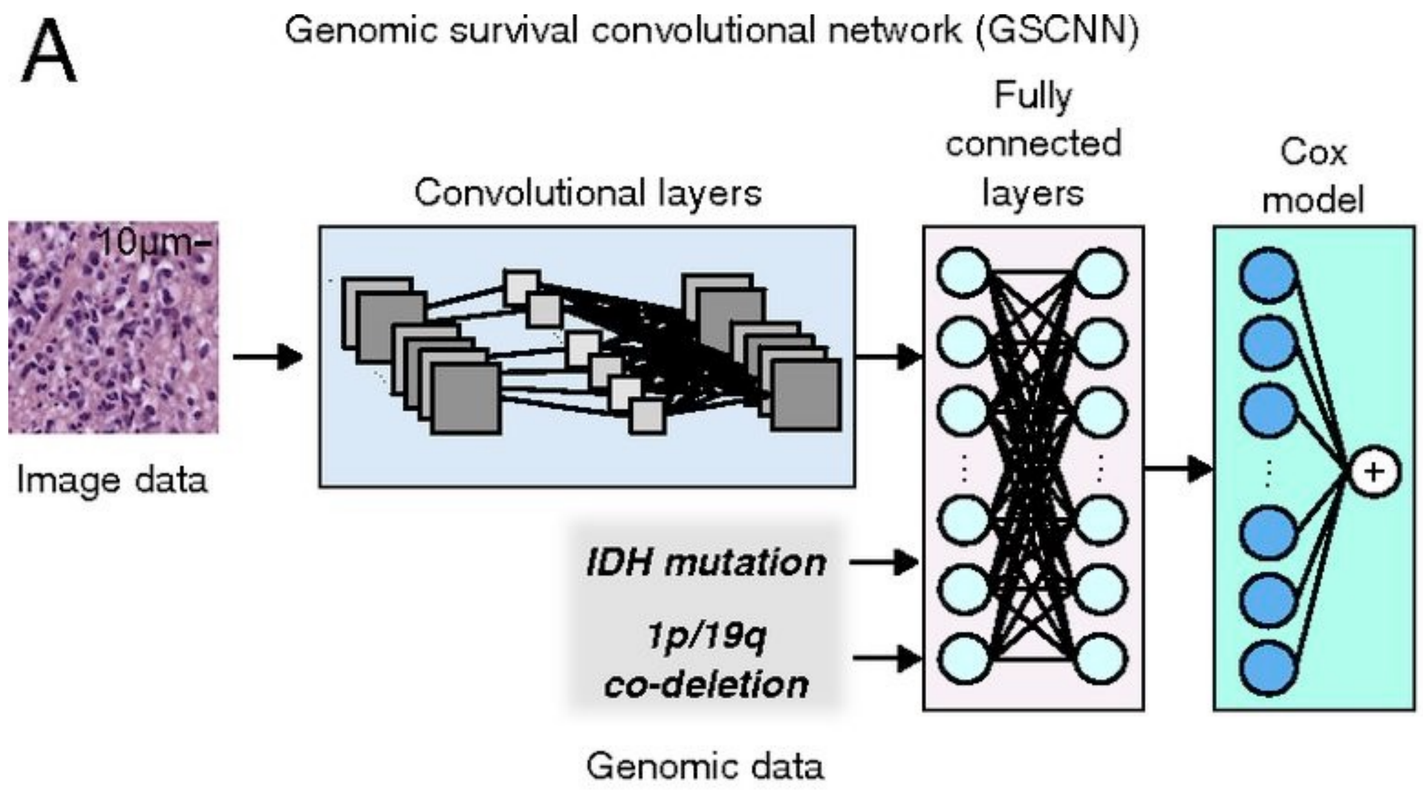


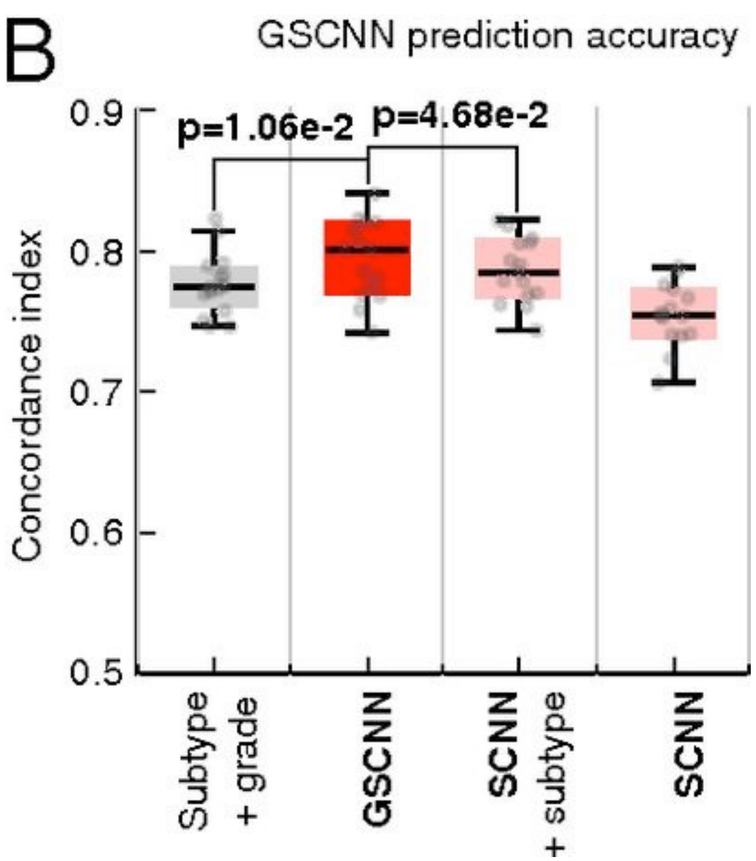
Learn more

P. Mobadersany et al., PNAS 2018

Patho-genomics to predict survival

Mobadersany et al, 2018



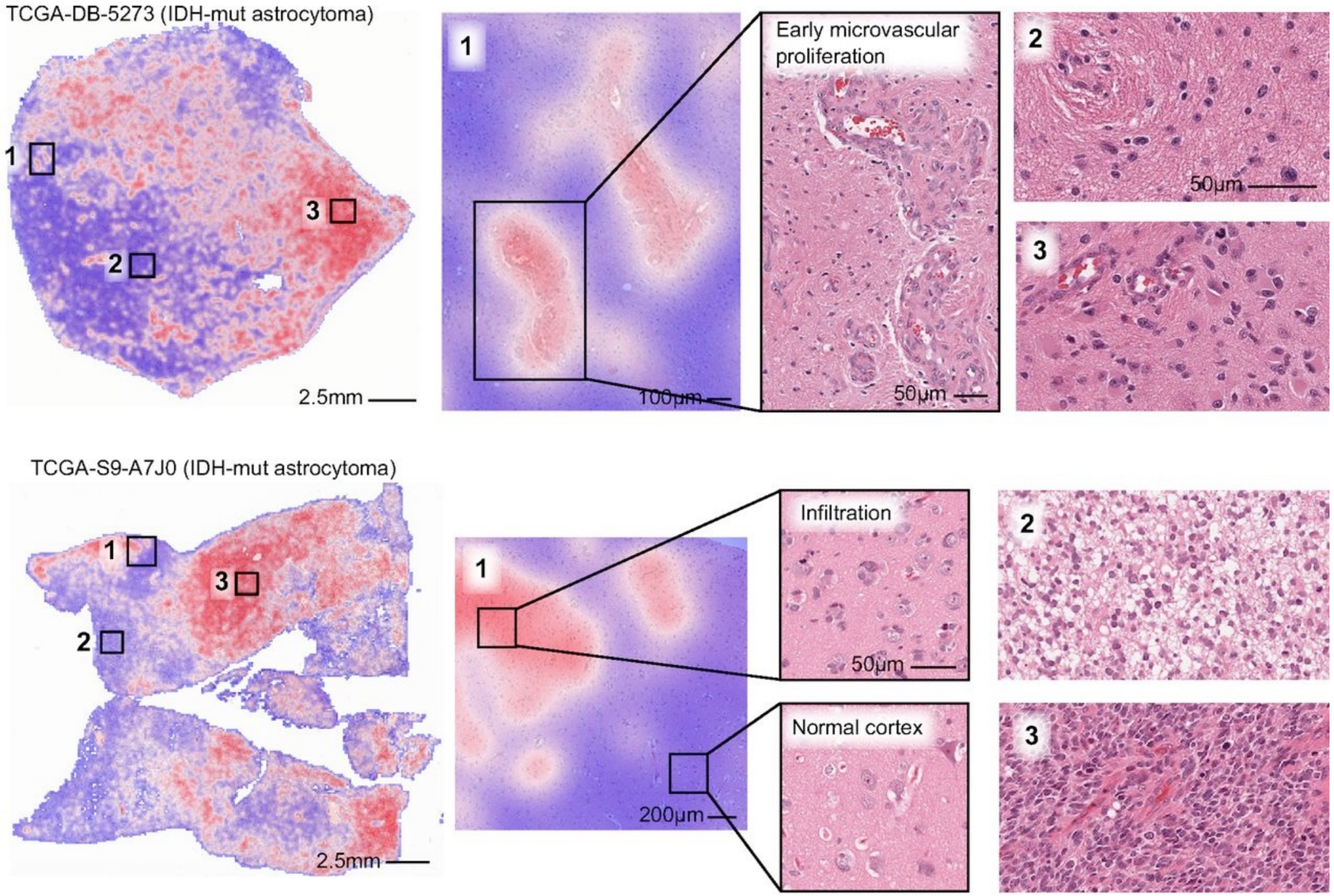


Learn more

P. Mobadersany et al., PNAS 2018

Patho-genomics to predict survival

Mobadersany et al, 2018

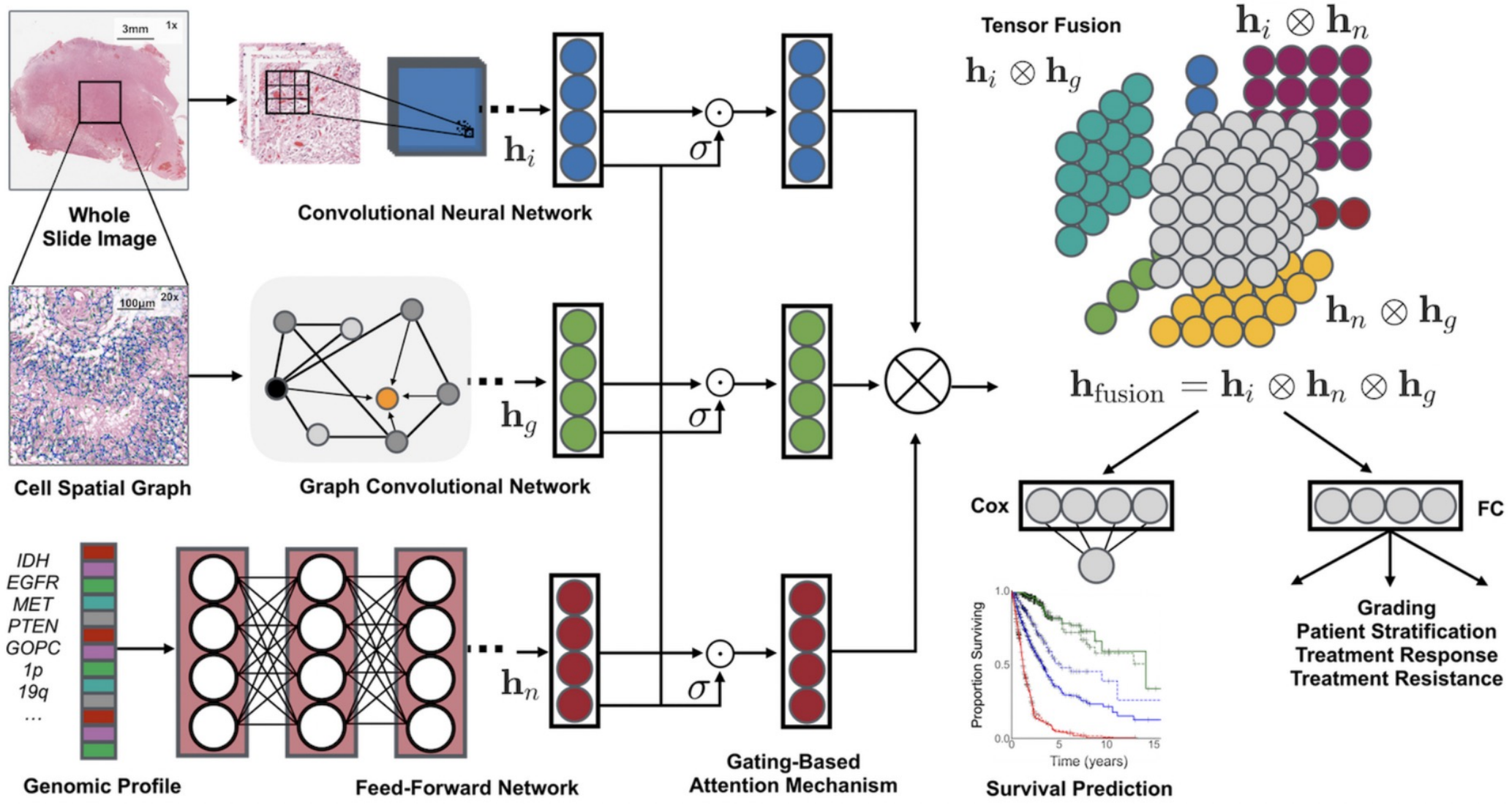


Learn more

P. Mobadersany et al., PNAS 2018

Patho-genomics to predict survival

Chen et al, 2022

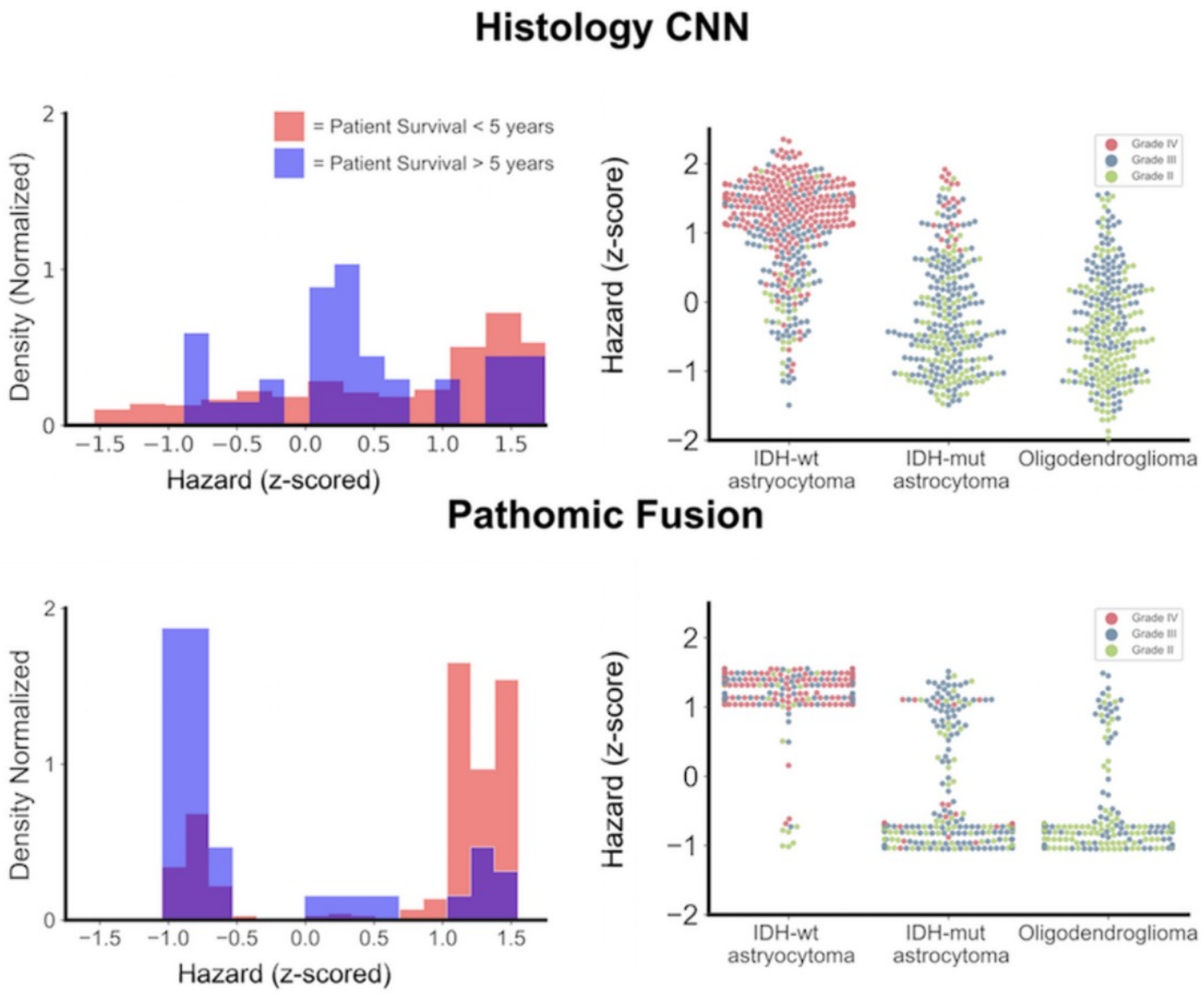


Learn more

Chen et al., IEEE Trans Med Imag, 2022

Patho-genomics to predict survival

Chen et al, 2022

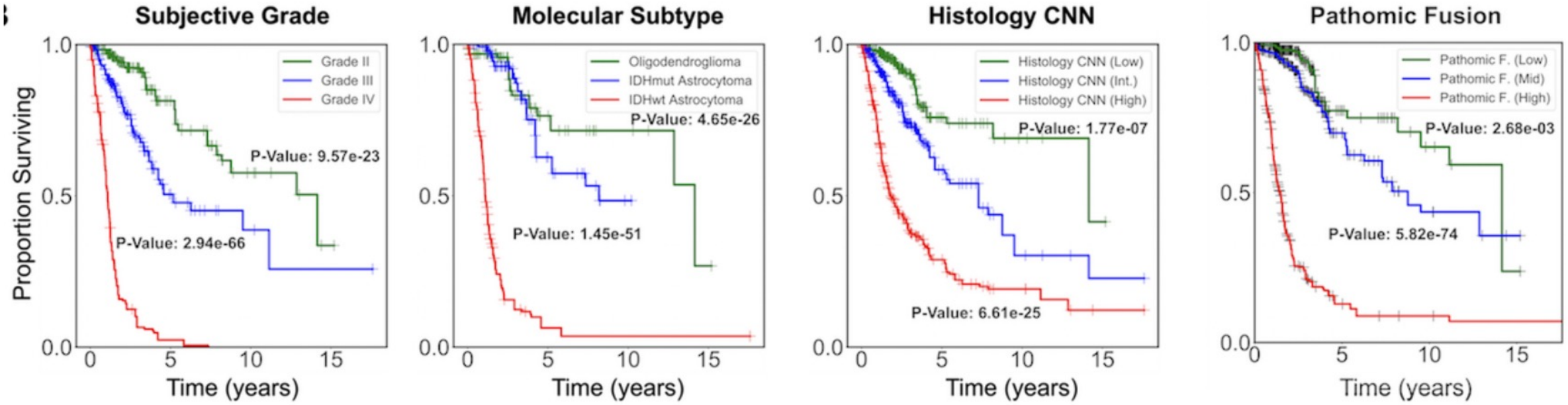


Learn more

Chen et al., IEEE Trans Med Imag, 2022

Patho-genomics to predict survival

Chen et al, 2022



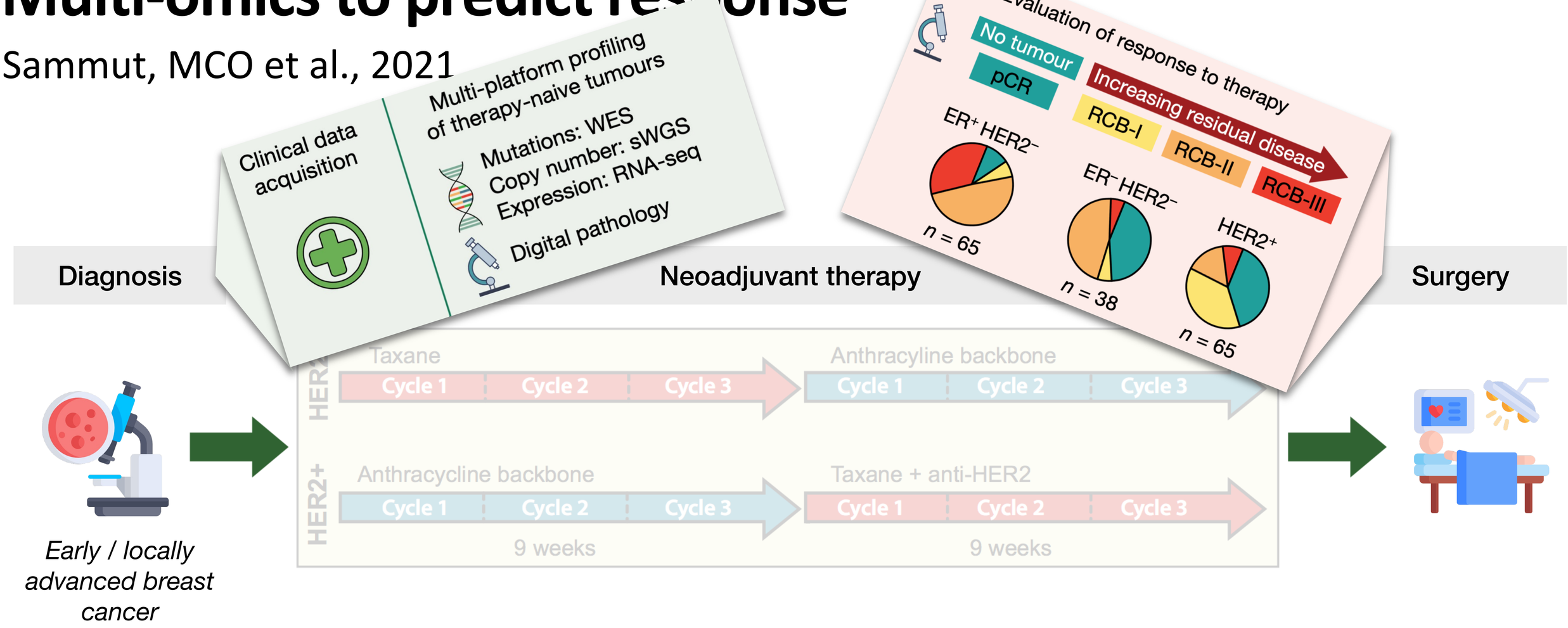
Learn more

Chen et al., IEEE Trans Med Imag, 2022

All -omics

Multi-omics to predict response

Sammut, MCO et al., 2021

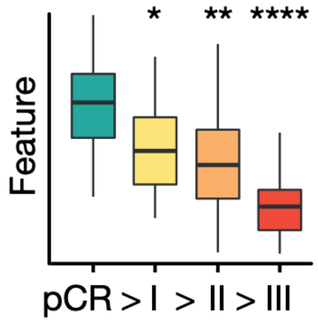


Learn more

Sammut et al., Nature 2021

Multi-omics to predict response

Step 1: Identification of individual features associated with response



Univariate tests for association

- Clinical
- DNA
- RNA
- Digital pathology

Tumour features (n = 19)

Including:
- ER, HER2 expression
- Mutational landscape
- Chromosomal instability

Immune activation (n = 4)

Including:
- Lymphocyte density
- Cytolytic activity
- Neoantigen expression

Immune evasion (n = 4)

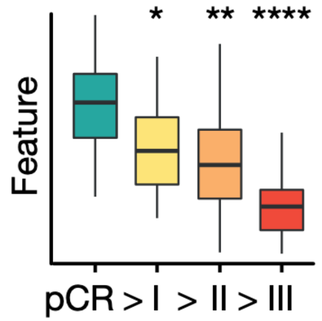
Including:
- HLA LOH
- T cell dysfunction
- T cell exclusion

Learn more

Sammut et al., Nature 2021

Multi-omics to predict response

Step 1: identification of individual features associated with response



Univariate tests for association
- Clinical
- DNA
- RNA
- Digital pathology

Tumour features (n = 19)

Including:
- ER, HER2 expression
- Mutational landscape
- Chromosomal instability

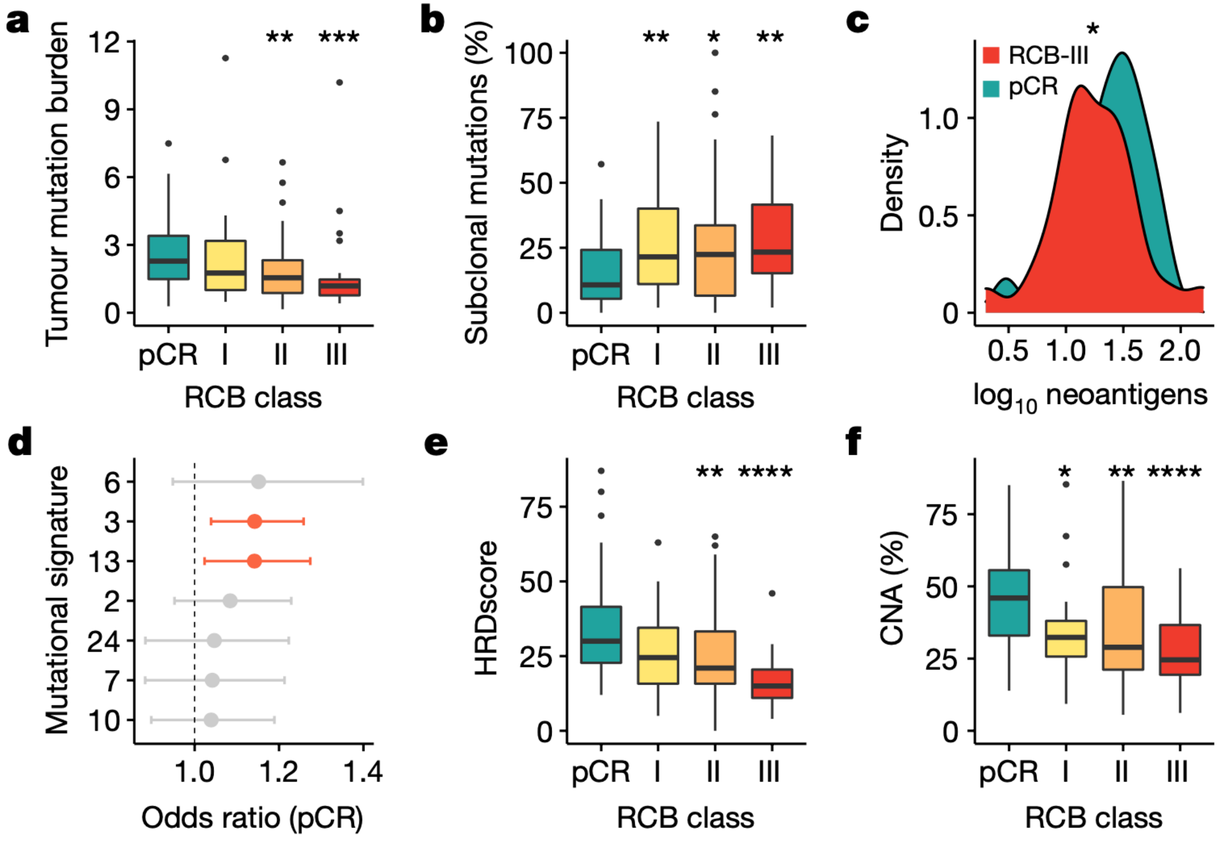
Immune activation (n = 4)

Including:
- Lymphocyte density
- Cytolytic activity
- Neoantigen expression

Immune evasion (n = 4)

Including:
- HLA LOH
- T cell dysfunction
- T cell exclusion

Genomic features associate with response

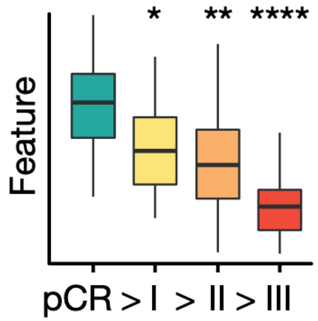


Learn more

Sammut et al., Nature 2021

Multi-omics to predict response

Step 1: identification of individual features associated with response



Univariate tests for association

- Clinical
- DNA
- RNA
- Digital pathology

Tumour features (n = 19)

Including:
- ER, HER2 expression
- Mutational landscape
- Chromosomal instability

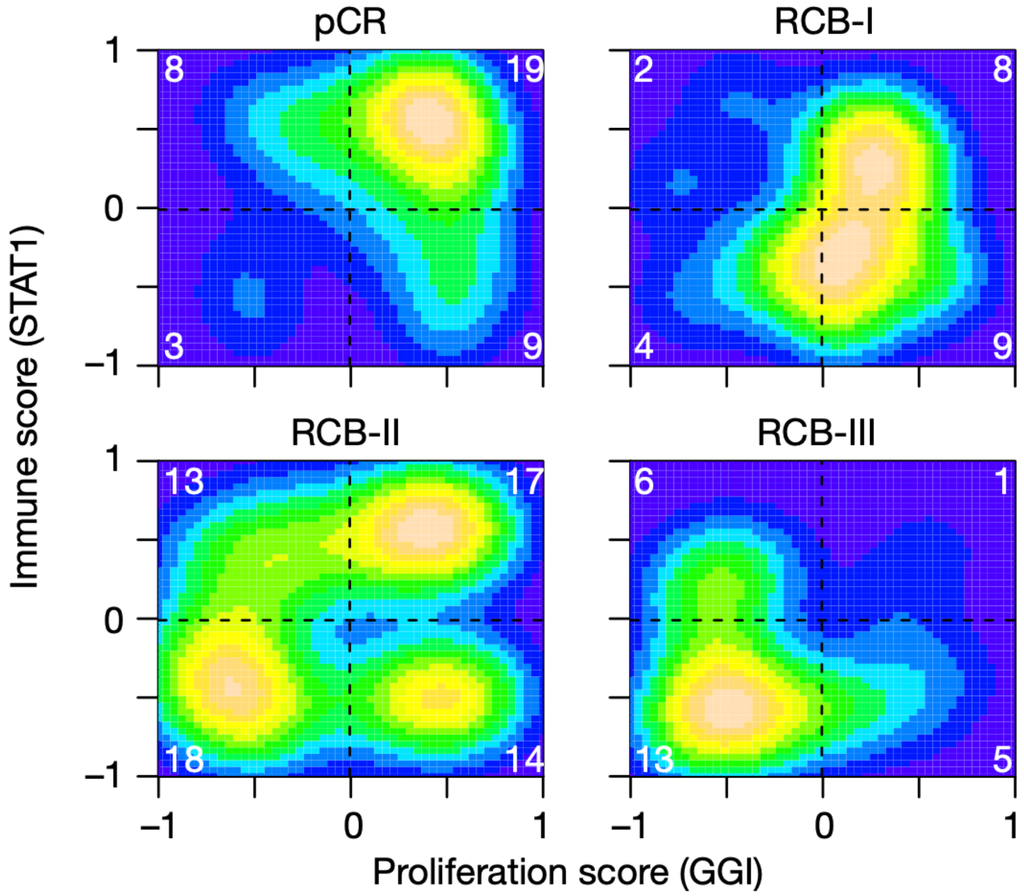
Immune activation (n = 4)

Including:
- Lymphocyte density
- Cytolytic activity
- Neoantigen expression

Immune evasion (n = 4)

Including:
- HLA LOH
- T cell dysfunction
- T cell exclusion

High proliferation and immune activation associate with pCR

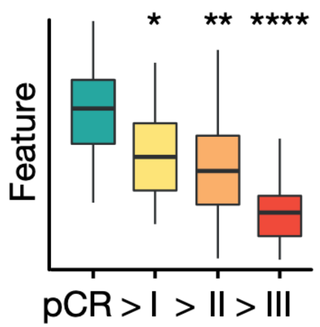


Learn more

Sammut et al., Nature 2021

Multi-omics to predict response

Step 1: identification of individual features associated with response



Univariate tests for association
- Clinical
- DNA
- RNA
- Digital pathology

Tumour features (n = 19)

Including:
- ER, HER2 expression
- Mutational landscape
- Chromosomal instability

Immune activation (n = 4)

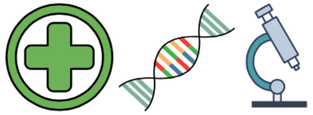
Including:
- Lymphocyte density
- Cytolytic activity
- Neoantigen expression

Immune evasion (n = 4)

Including:
- HLA LOH
- T cell dysfunction
- T cell exclusion

Step 2: data integration of features identified in step 1 using a multi-omic machine learning model

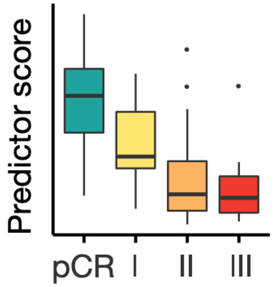
Pre-therapy features



Post-therapy response

pCR or Residual disease

Machine learning predictor of response



Validation in independent dataset (n = 75)

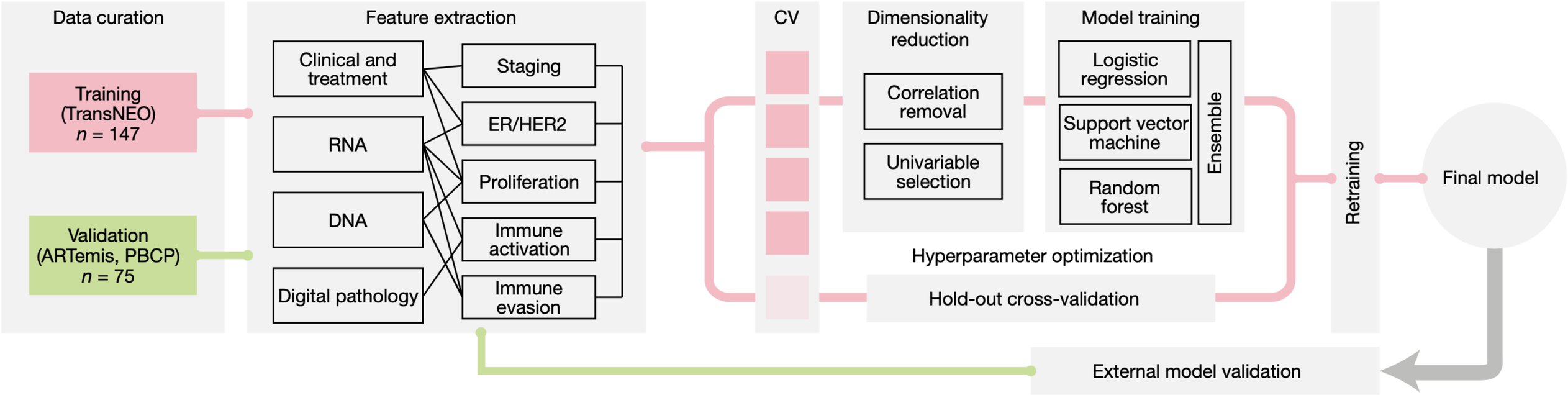
Personalized therapy decisions

Predict response using
- Clinical
- Clinical + DNA or RNA
- Clinical + DNA + RNA
- Clinical + DNA + RNA + Digital pathology

Learn more

Sammut et al., Nature 2021

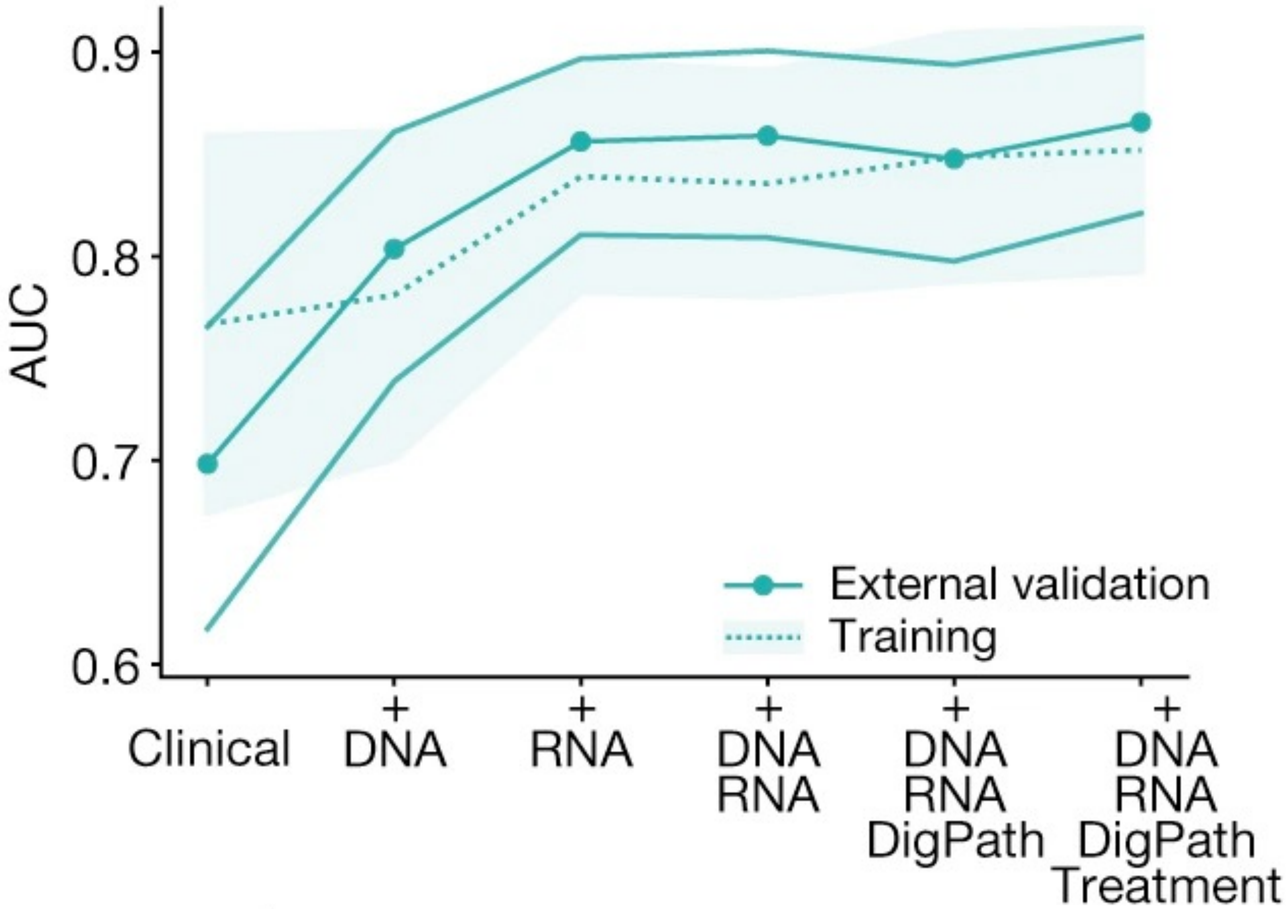
Multi-omics to predict response



Learn more

Sammut et al., Nature 2021

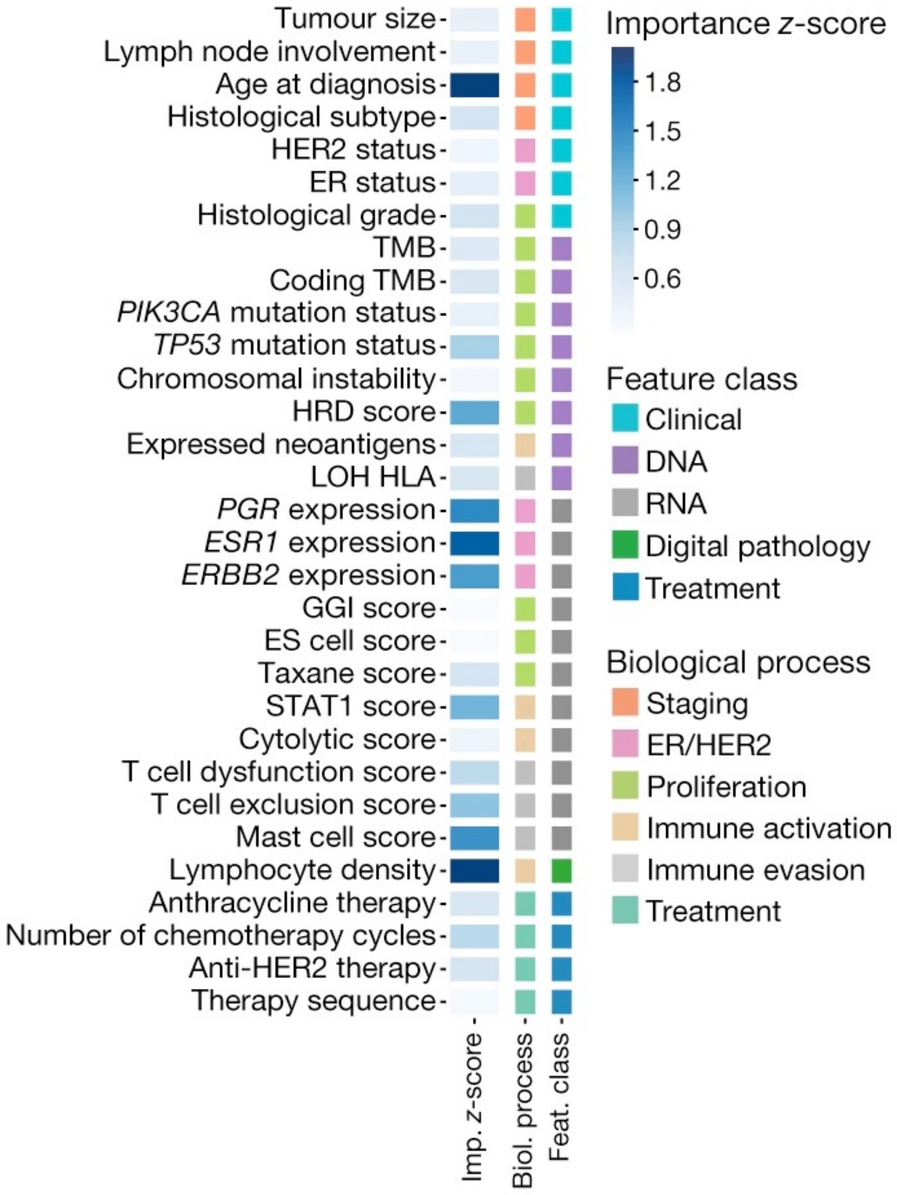
Multi-omics to predict response



Learn more

Sammut et al., Nature 2021

Multi-omics to predict response

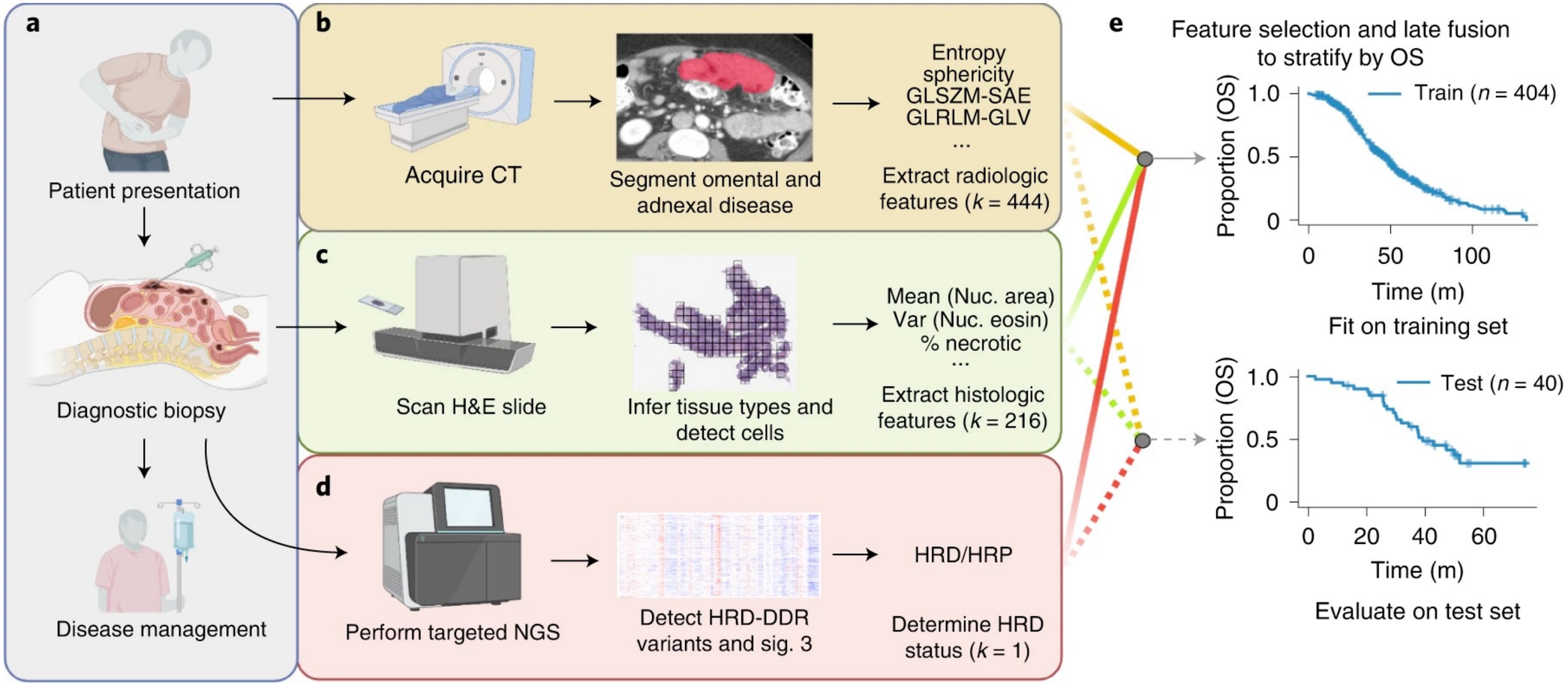


Learn more

Sammut et al., Nature 2021

Multi-omics to predict survival

Boehm et al, 2022

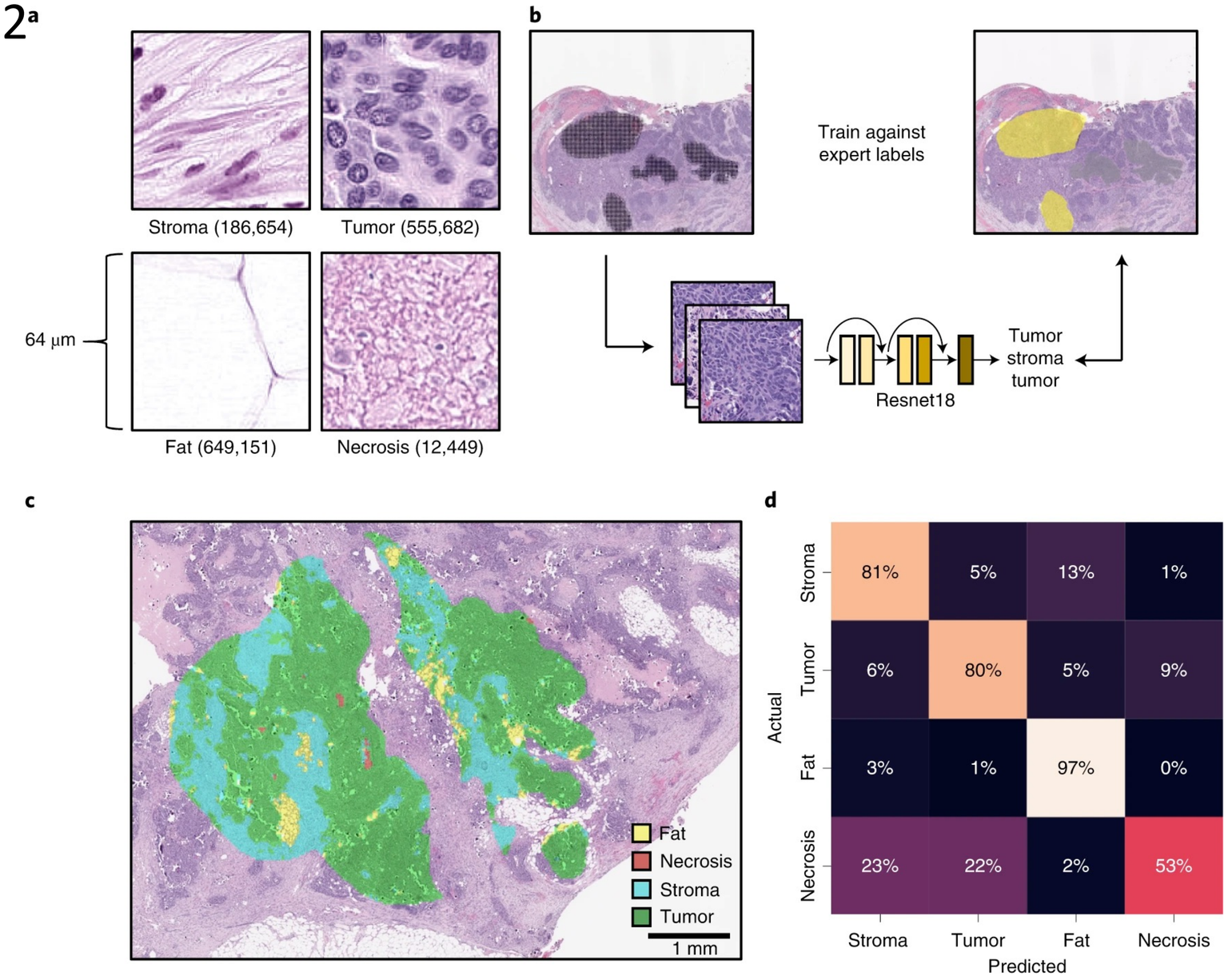


Learn more

K. Boehm et al., Nature Cancer 2022

Multi-omics to predict survival

Boehm et al, 2022[a]

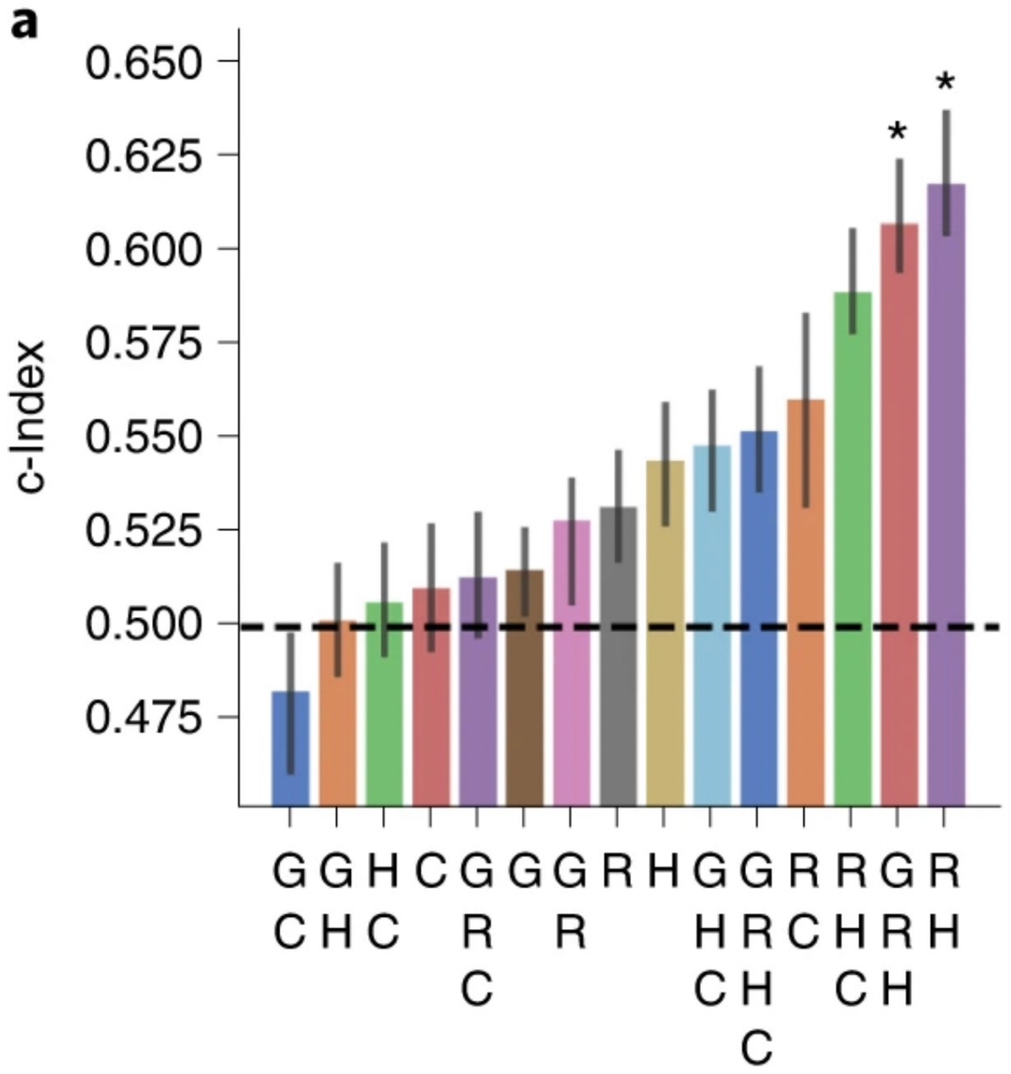


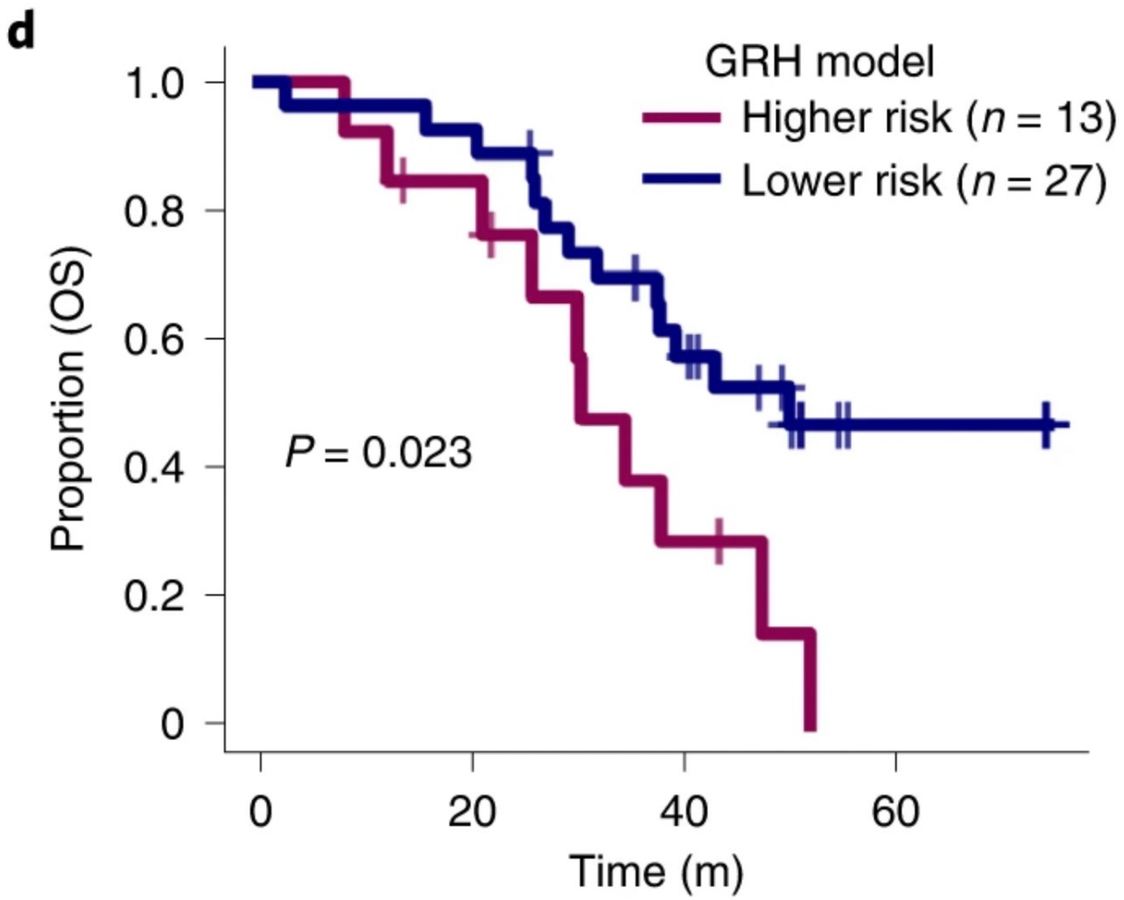
Learn more

K. Boehm et al., Nature Cancer 2022

Multi-omics to predict survival

Boehm et al, 2022



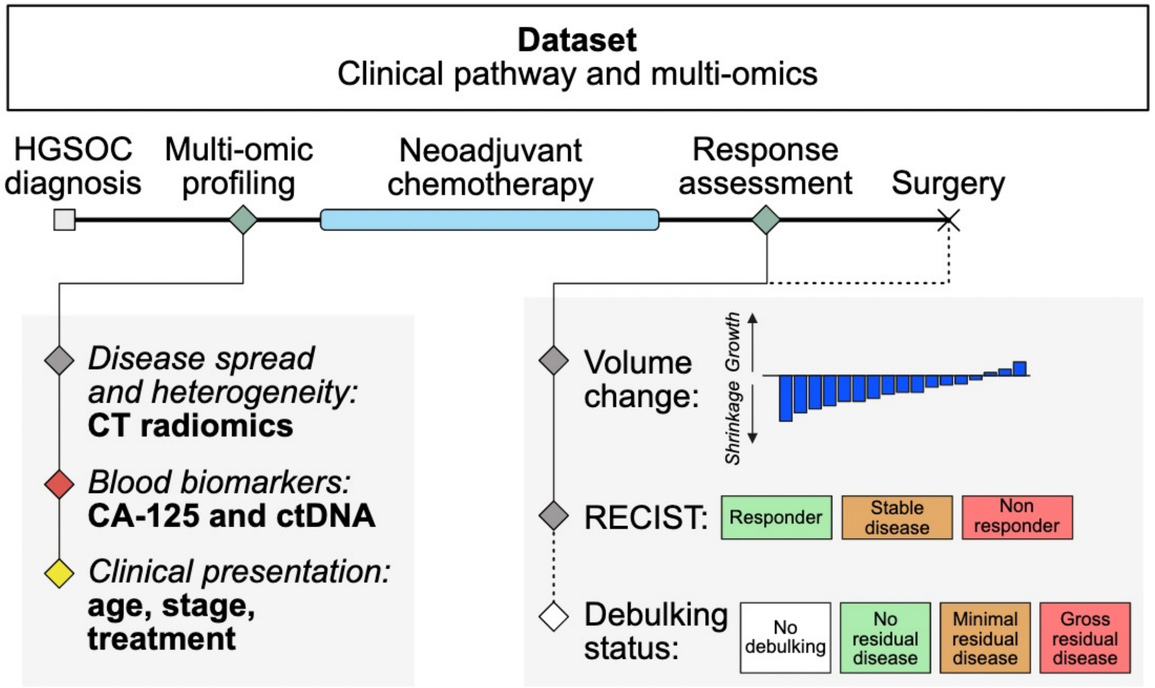


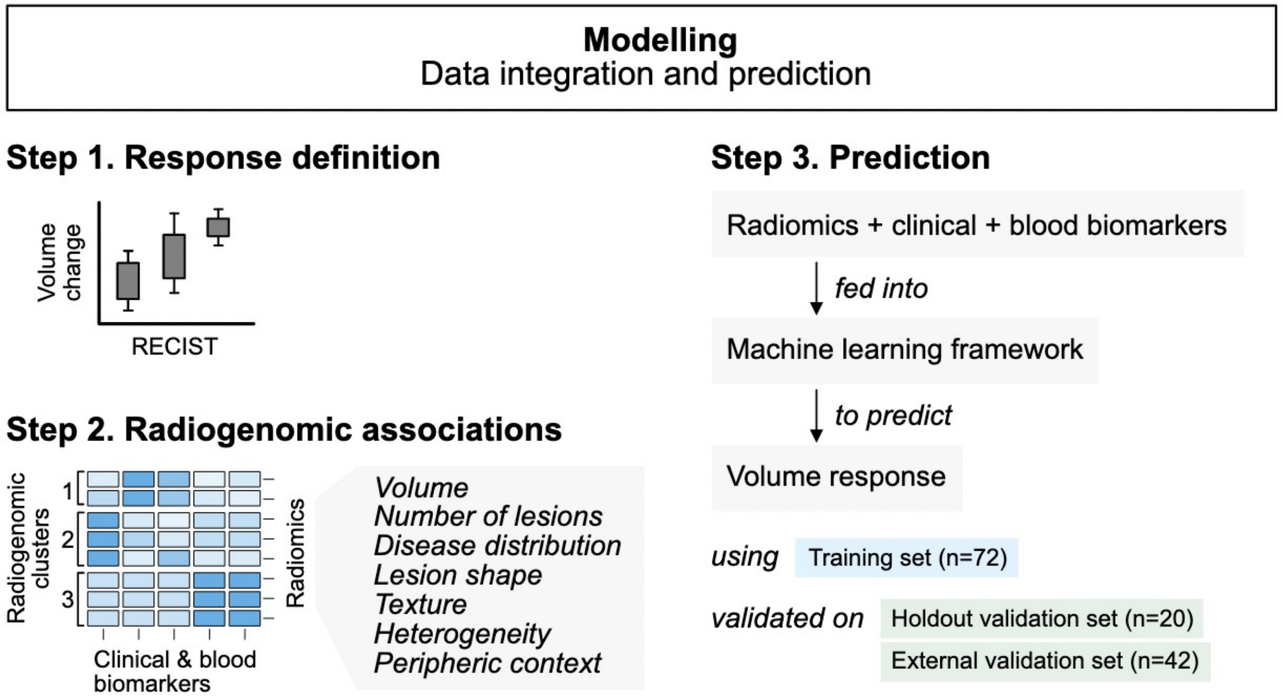
Learn more

K. Boehm et al., Nature Cancer 2022

Multi-omics to predict response

Adding ctDNA



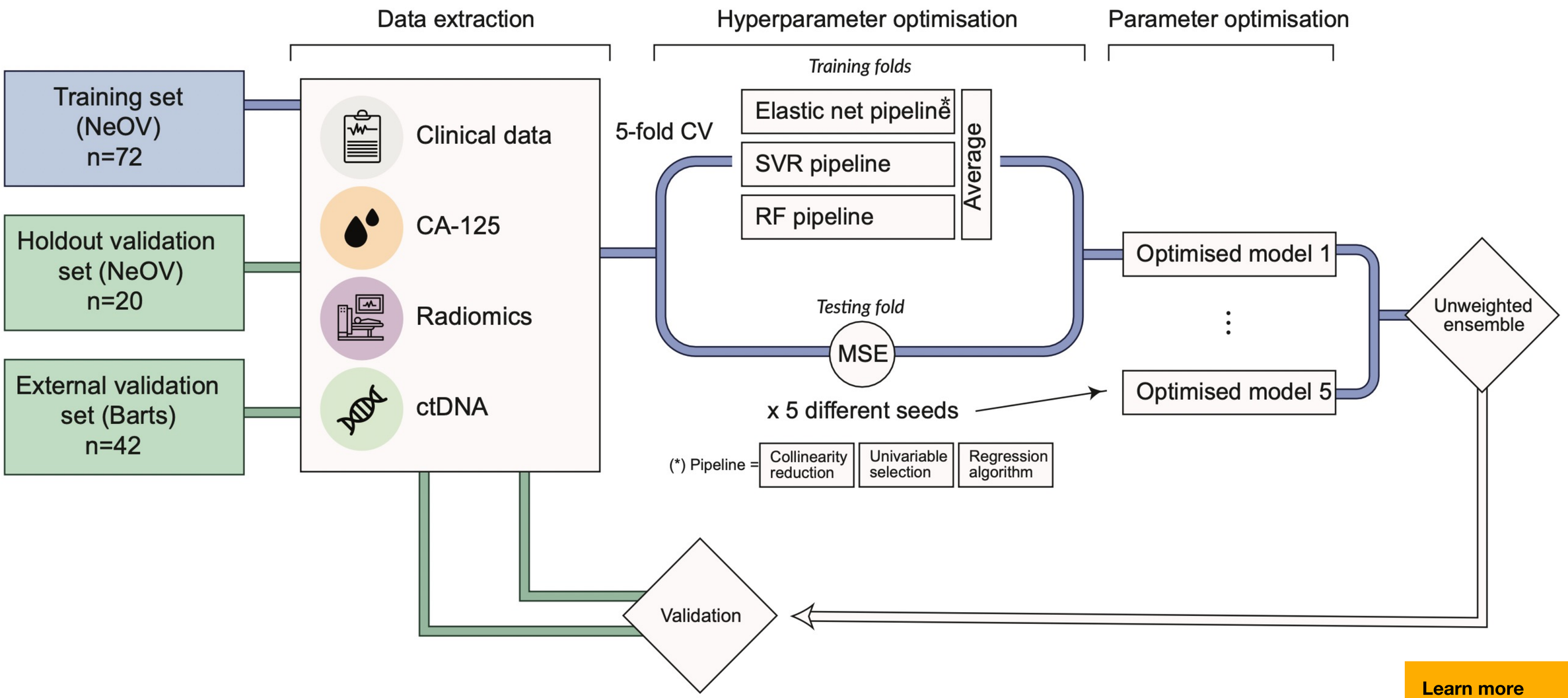


Learn more

Crispin et al, under review 2022

Multi-omics to predict response

Adding ctDNA



Learn more

Crispin et al, under review 2022

Multi-omics to predict response

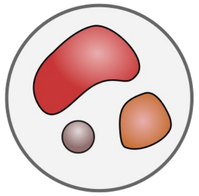
Semantic

Semantic descriptors
Ascites, Pleural effusion

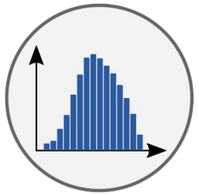
Clinical

Treatment characteristics
Treatment regimen, number of cycles, treatment duration

Radiomics



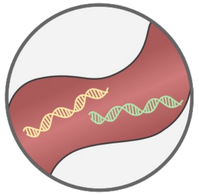
Tumour shape and volume
Compactness, elongation, flatness, major and minor axes, sphericity, surface, volume



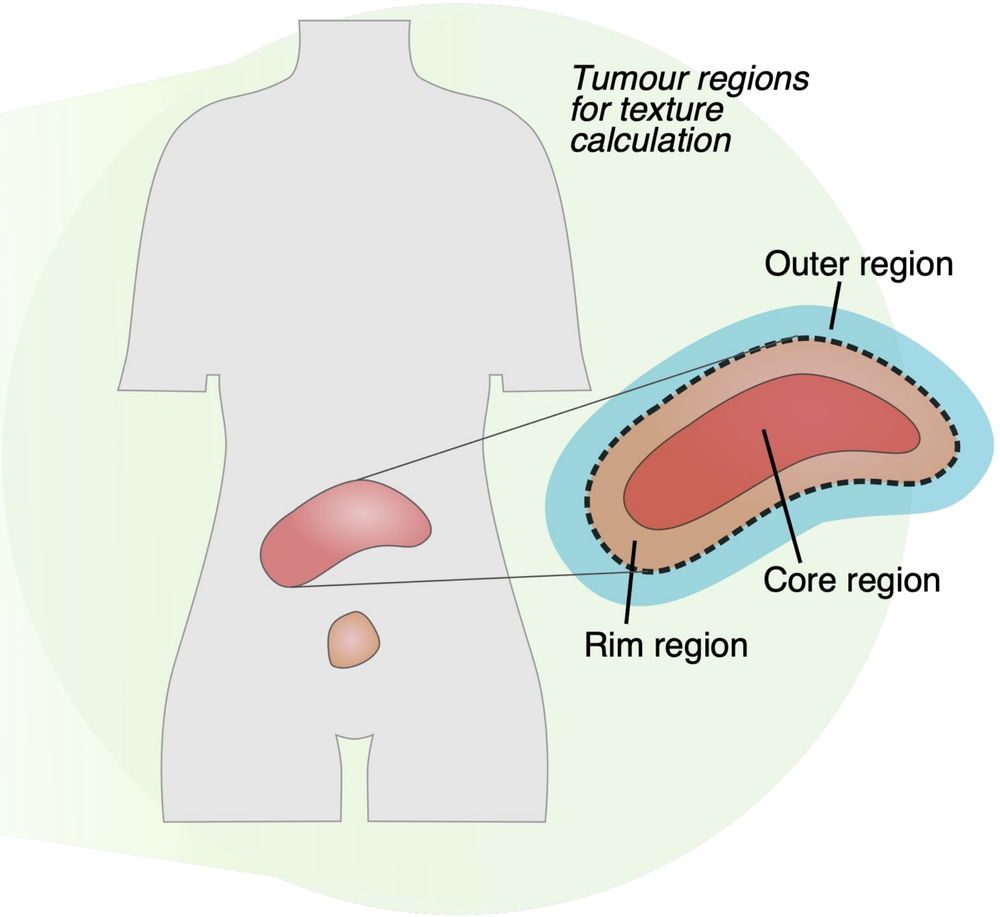
Intensity
Mean, median, skewness, kurtosis, standard deviation, volume for intensity above given HU

Texture features
Haralick energy, entropy, correlation, contrast, cluster prominence, shade...

ctDNA



Circulating Tumour DNA
tMAD, *TP53* MAF



Learn more

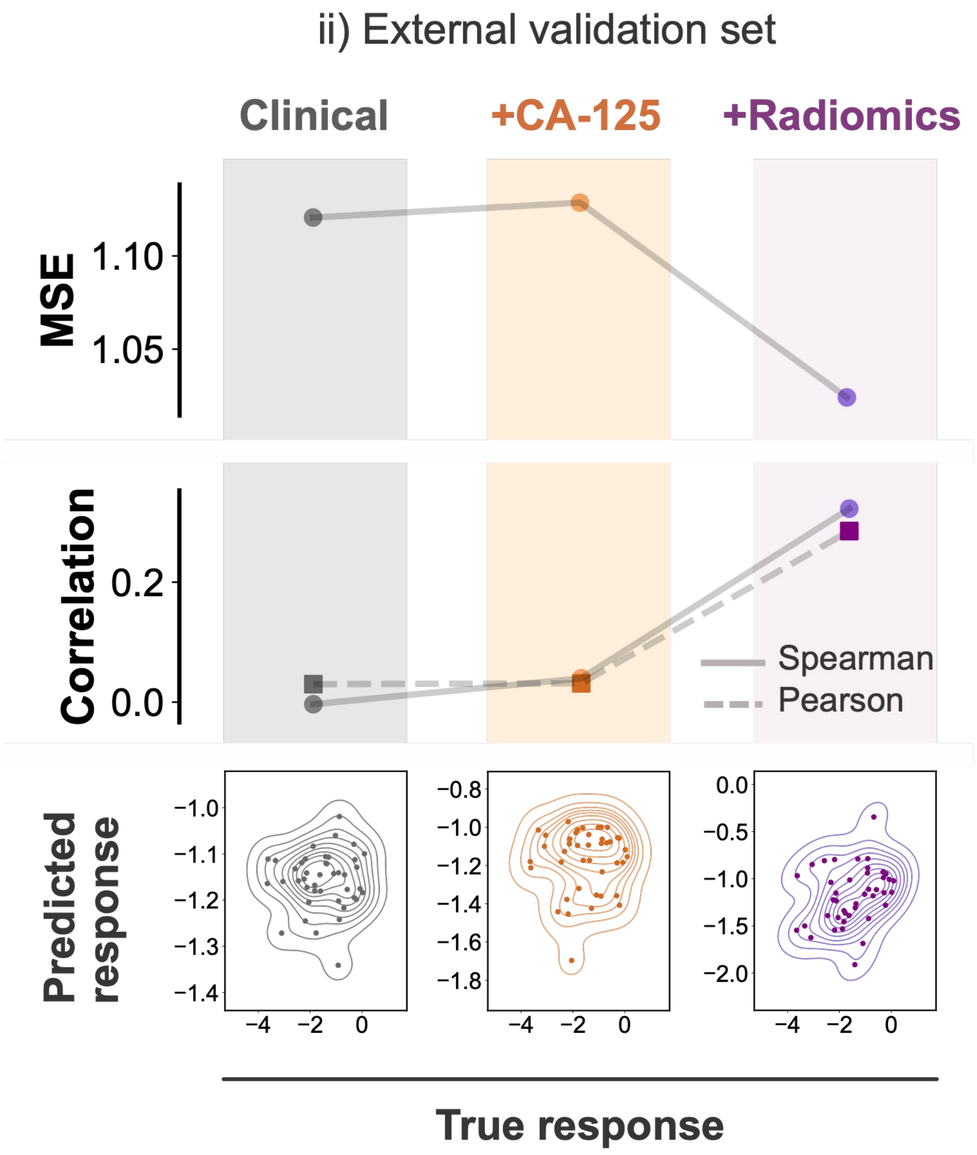
Crispin et al, under review 2022

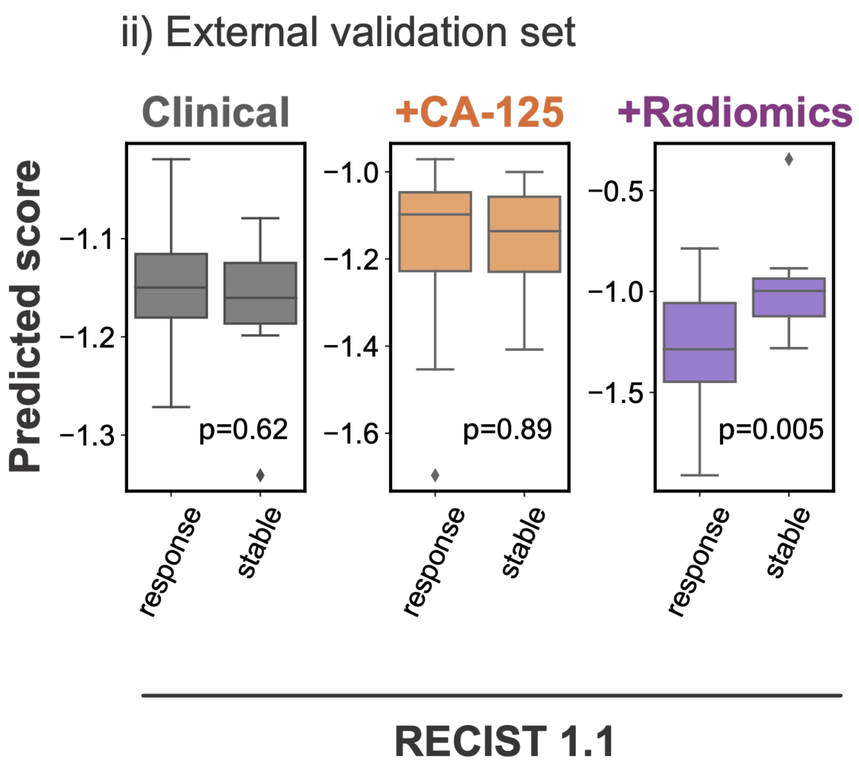
Multi-omics to predict response

Learn more

Crispin et al, under review 2022

Multi-omics to predict response





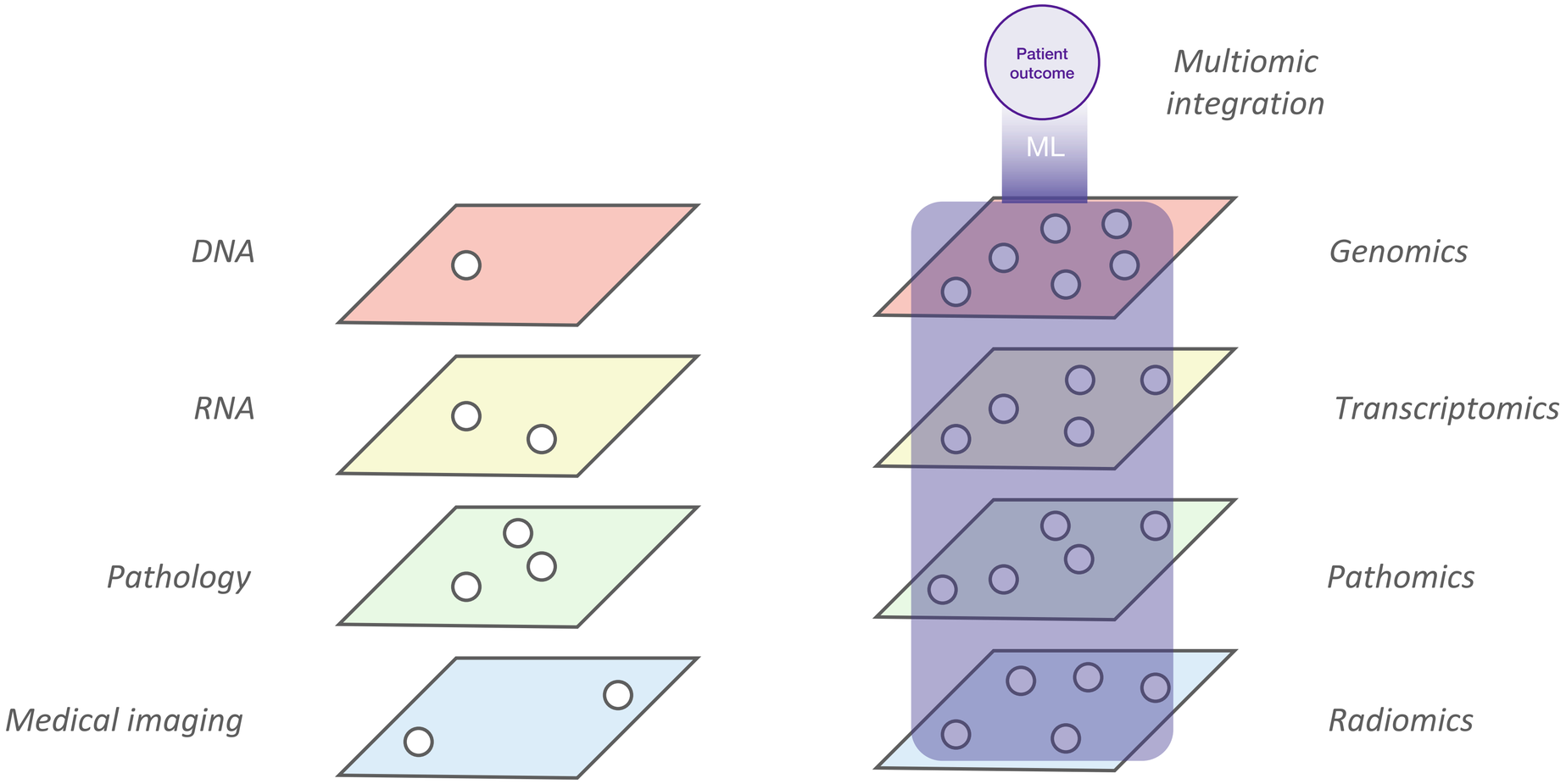
Learn more

Crispin et al, under review 2022

Wrapping up

- Cancer primer
- Multi-omics data: from biopsies to medical imaging
- Approaches to data integration
- Multi-omic translation: examples
- Multi-omic integration: examples

Multi-omics


Patient outcome
Multiomic integration
ML
DNA
RNA
Pathology
Medical imaging
Genomics
Transcriptomics
Pathomics
Radiomics

Thank you!

mc973@cam.ac.uk

BACKUP

Chen et al - GCN

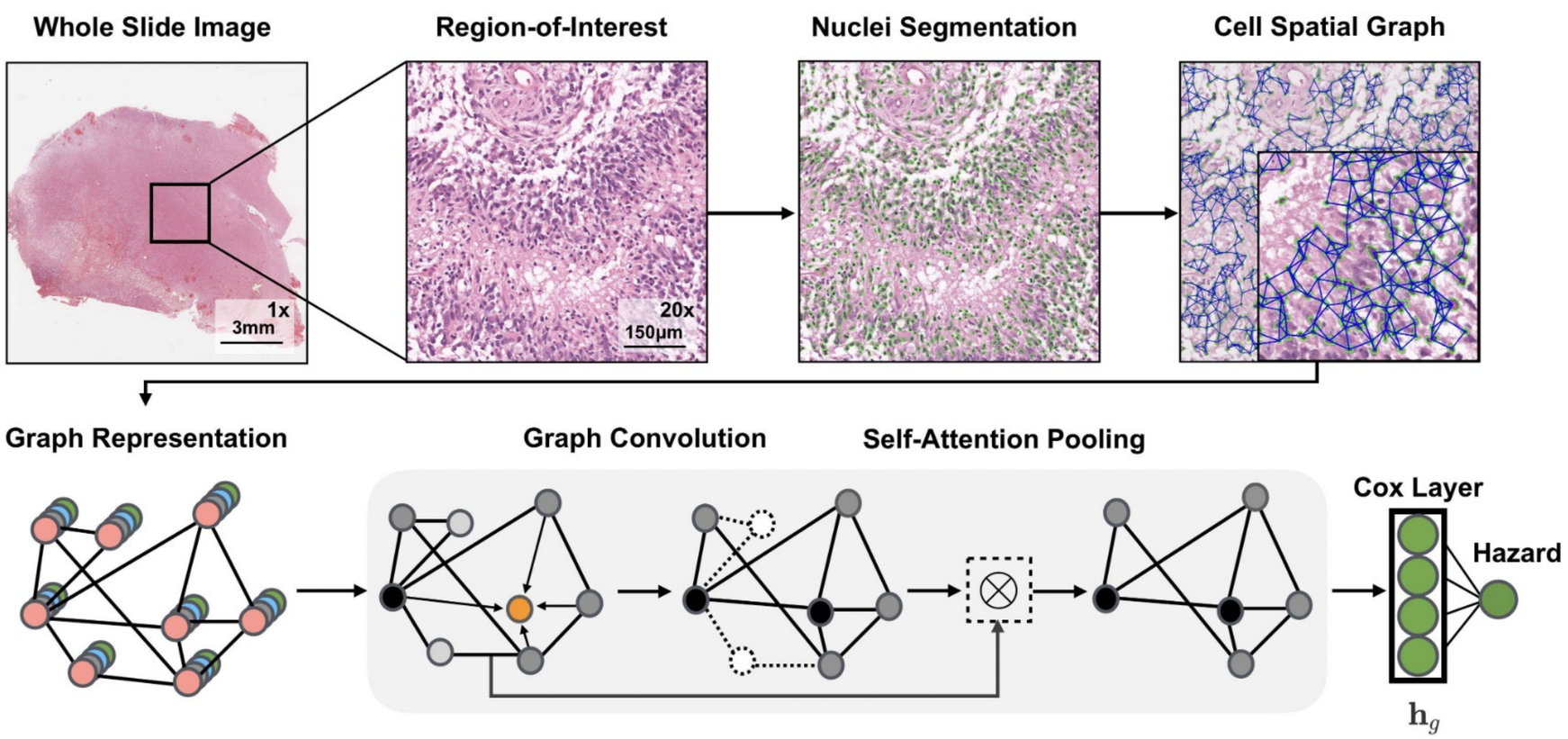


Fig. 2. Graph convolutional network for learning morphometric cell features from histology images. We represent cells in histology tissue as nodes in a graph, where cells are isolated using a deep learning-based nuclei segmentation algorithm and the connections between cells are made using KNN. Features for each cell are initialized using handcrafted features as well as deep features learned using contrastive predictive coding. The aggregate and combine functions are adopted from the GraphSAGE architecture, with the node masking and hierarchical pooling strategy adopted from SAGEPool.