Example of project: AI-based classification of cancers of unknown primary

Julien VIBERT

MD Medical Oncology - PhD Bioinformatics

DITEP - Gustave Roussy

U830 Institut Curie

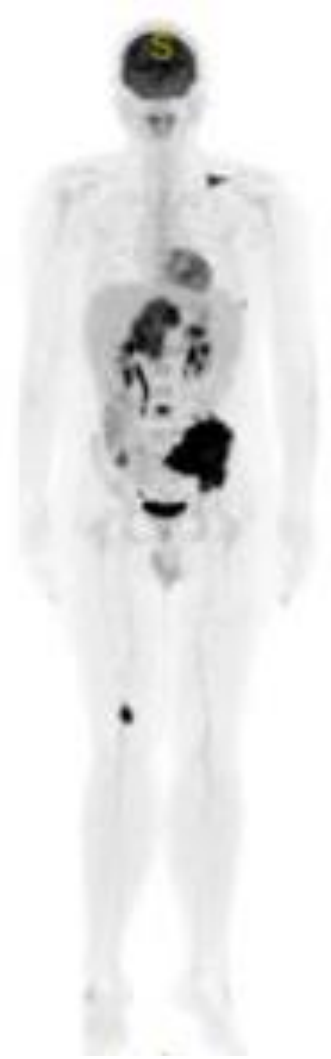
IFSBM Module 12

Gustave Roussy

20 January 2022

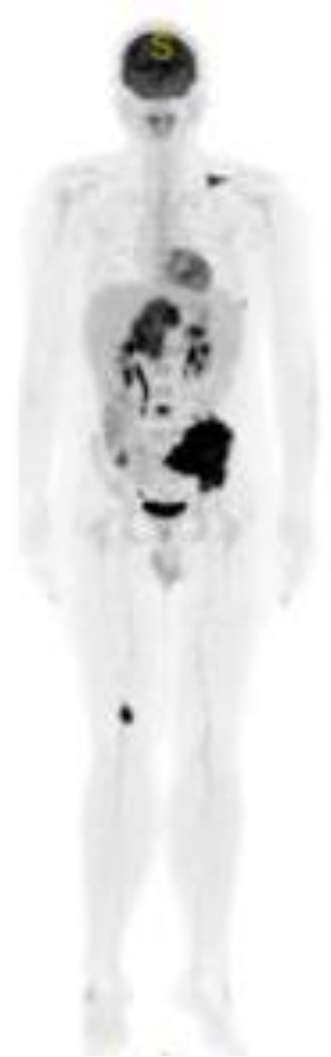
Clinical case: patient n°1

- 30-year old male patient
- Rapid degradation of health status
- Abdominal mass, diffuse bone lesions
- Addressed to Institut Curie for suspicion of bone sarcoma
- **Unclassified carcinoma** according to pathologists
- Proposed treatment: unspecific cytotoxic chemotherapy



Clinical case: patient n°1

- 30-year old male patient
- Rapid degradation of health status
- Abdominal mass, diffuse bone lesions
- Addressed to Institut Curie for suspicion of bone sarcoma
- **Unclassified carcinoma** according to pathologists
- Proposed treatment: unspecific cytotoxic chemotherapy



RNA-seq available in Curie for diagnosis of sarcomas

Cancers of unknown primary

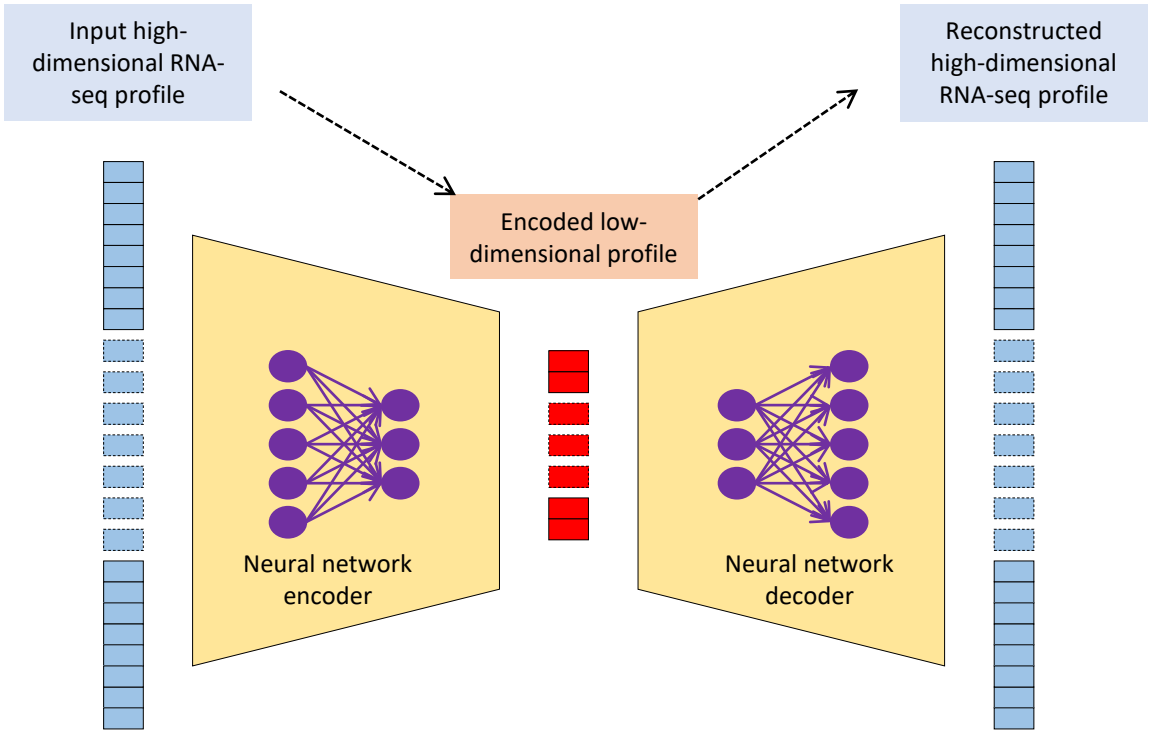
- **Cancers of unknown primary (CUP)** : 2-3 % of metastatic cancers
- The primitive tumor has not been found despite extensive explorations (radiology, pathology…)

Cancers of unknown primary

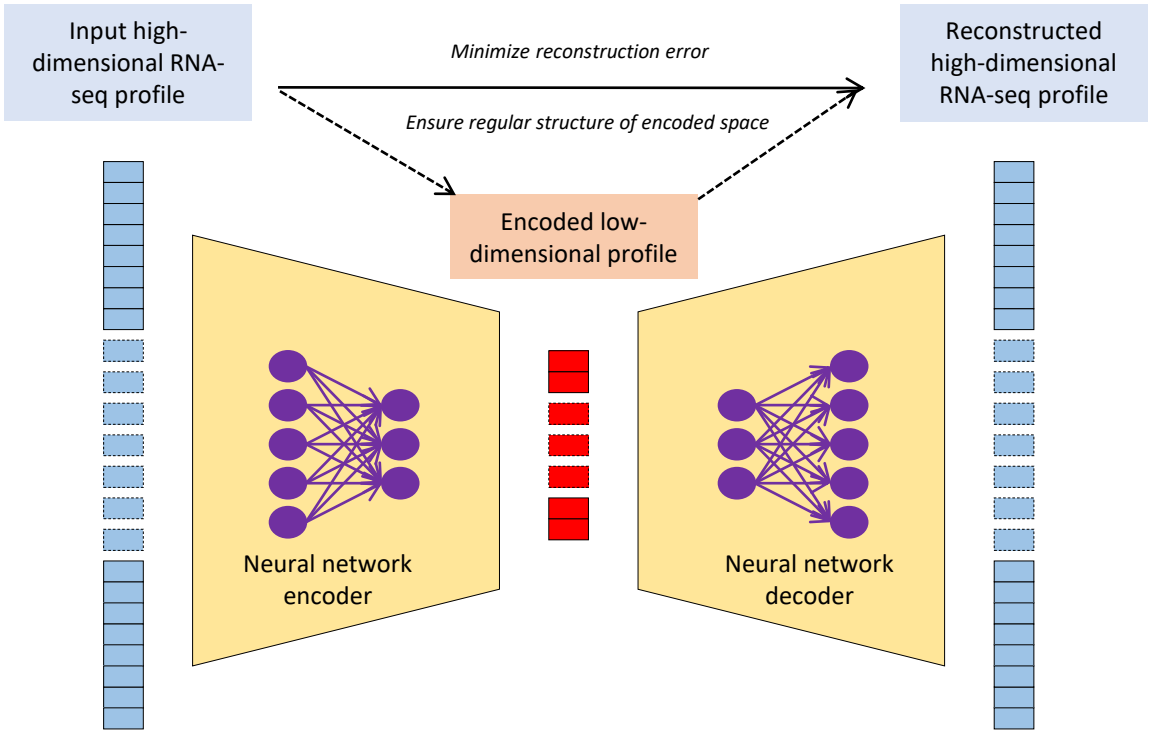
- **Cancers of unknown primary (CUP)** : 2-3 % of metastatic cancers
- The primitive tumor has not been found despite extensive explorations (radiology, pathology…)

-> *Could we identify the tissue of origin based on the transcriptomic profile (> 50,000 features)?*

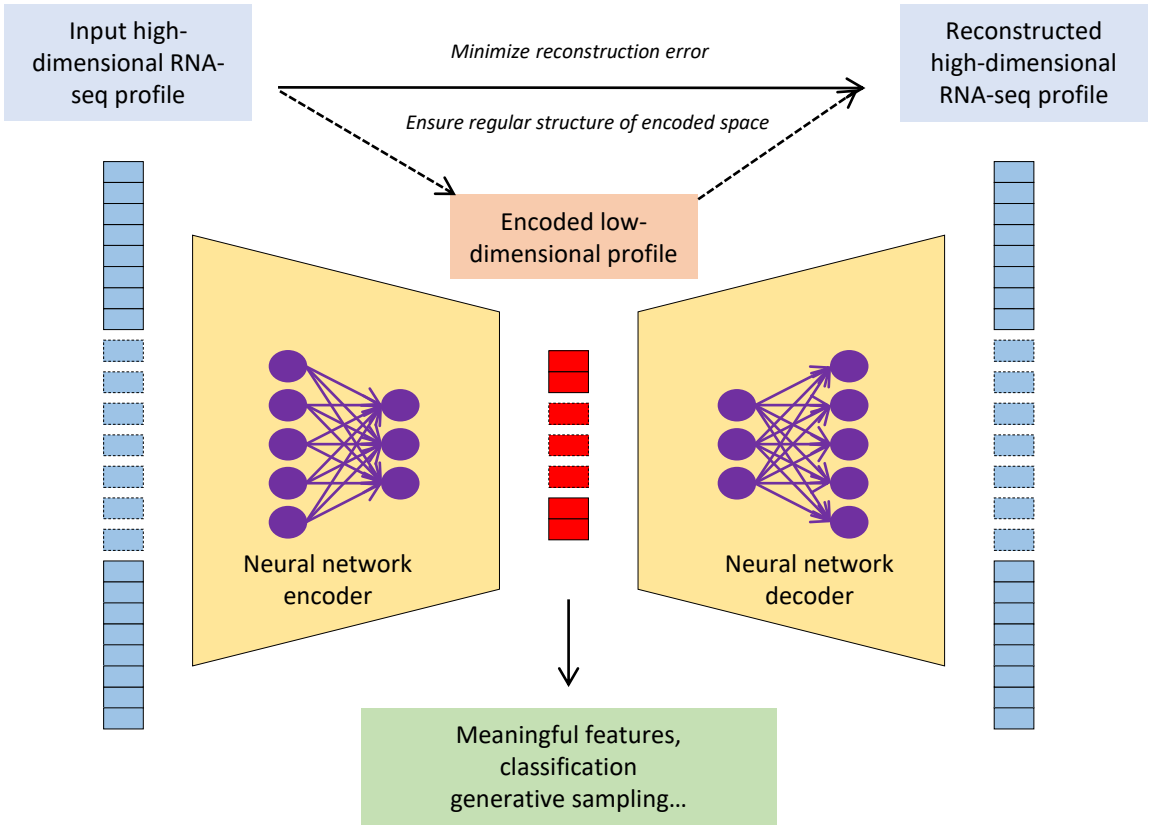
A deep learning algorithm: the variational autoencoder (VAE)



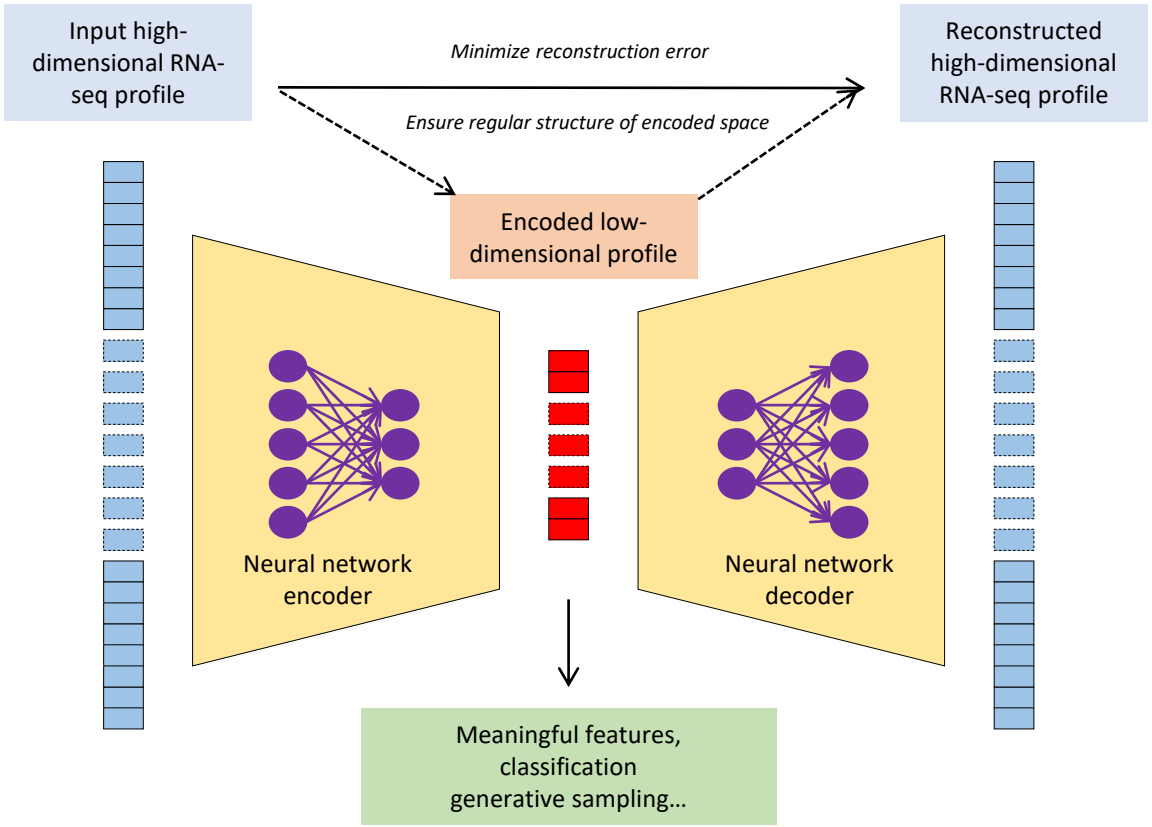
A deep learning algorithm: the variational autoencoder (VAE)



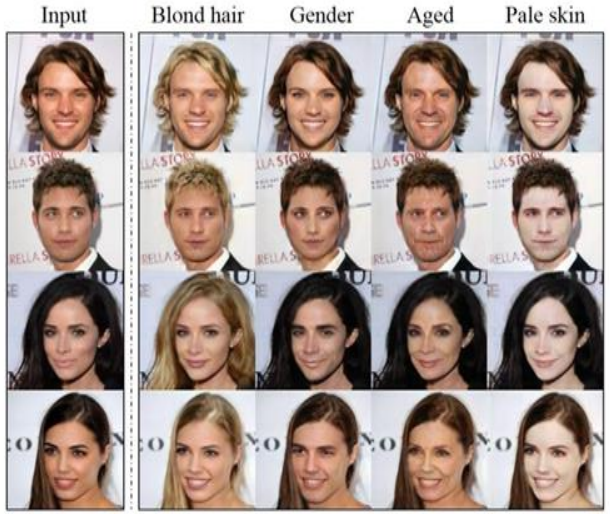
A deep learning algorithm: the variational autoencoder (VAE)



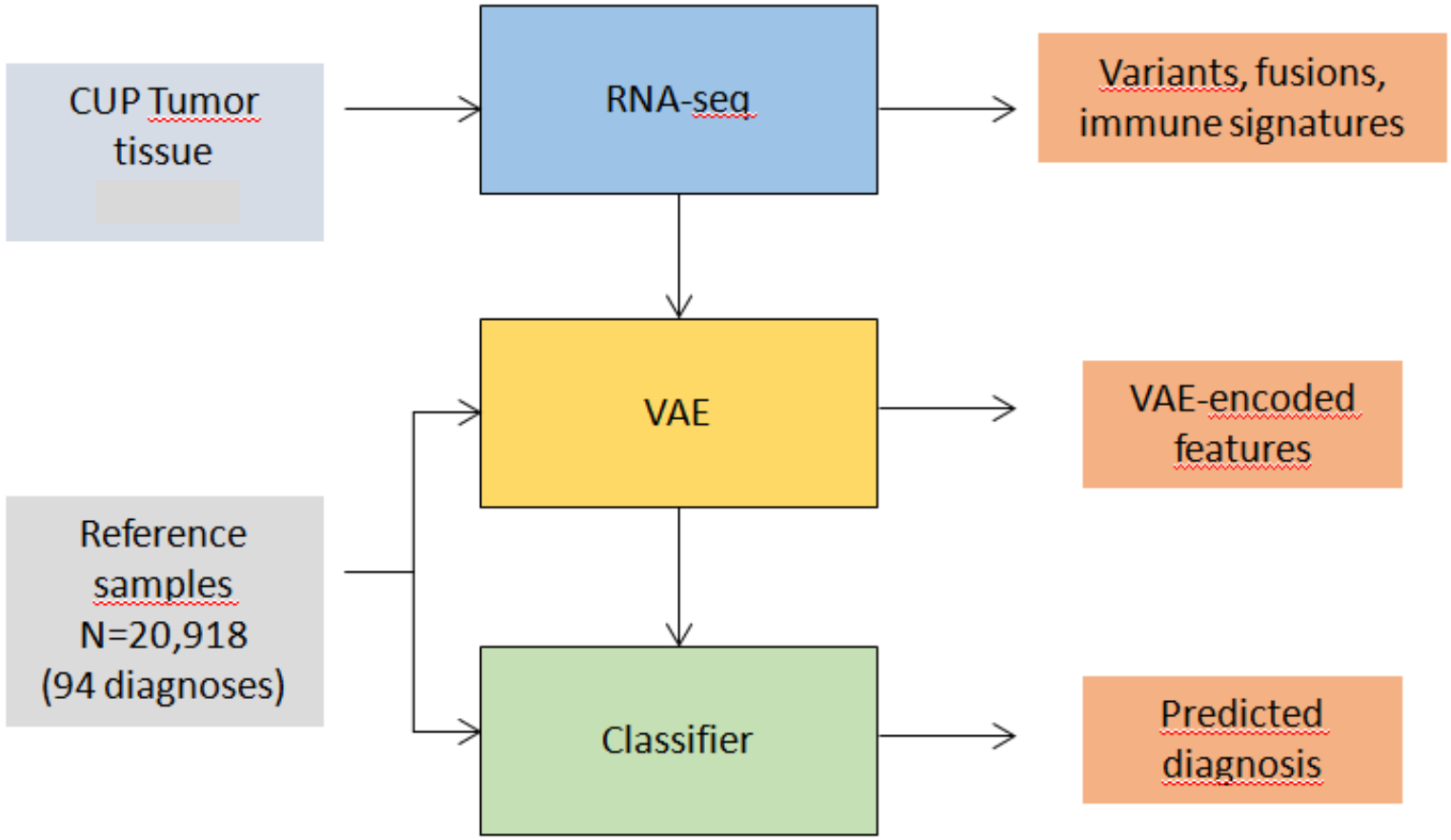
A deep learning algorithm: the variational autoencoder (VAE)



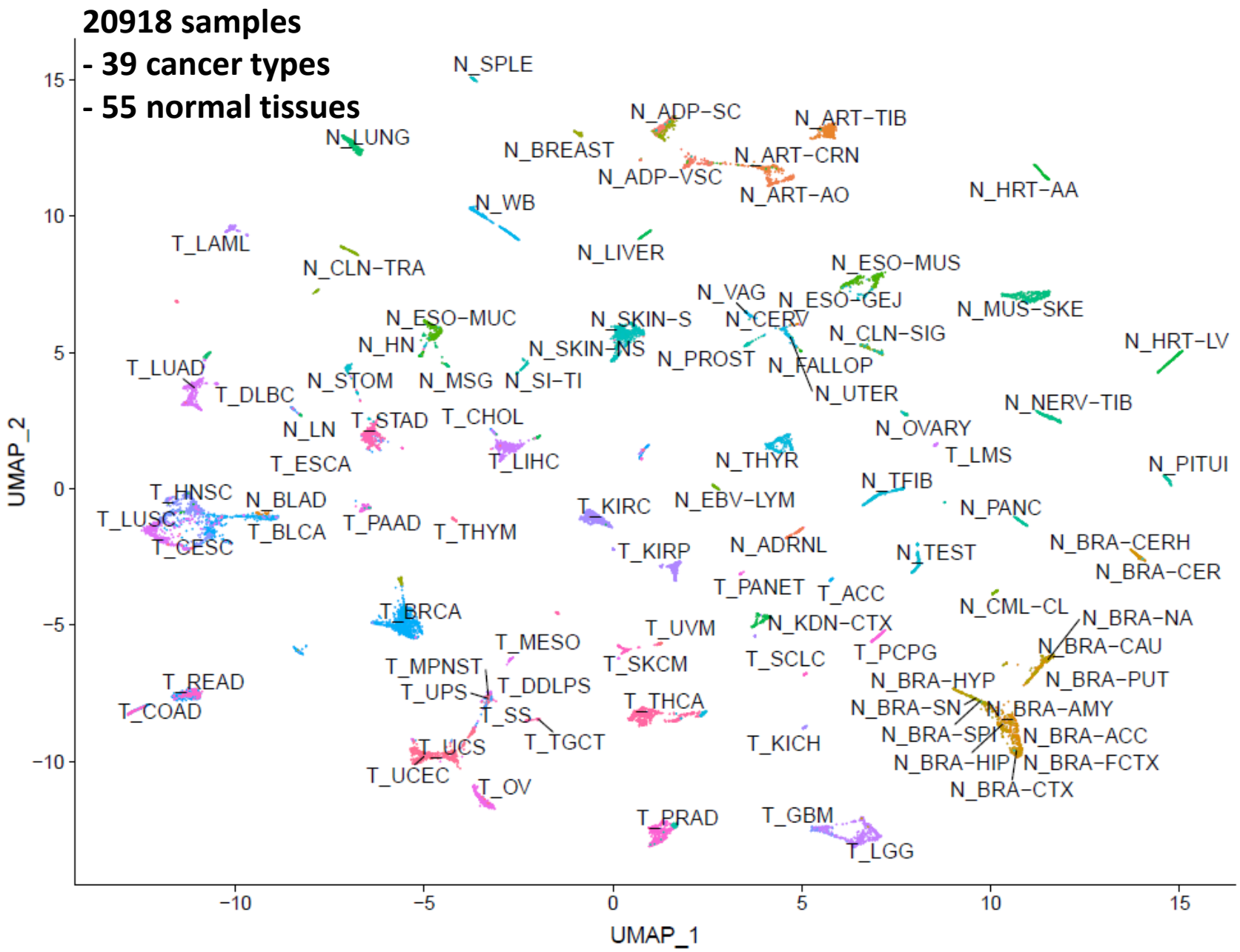
For images



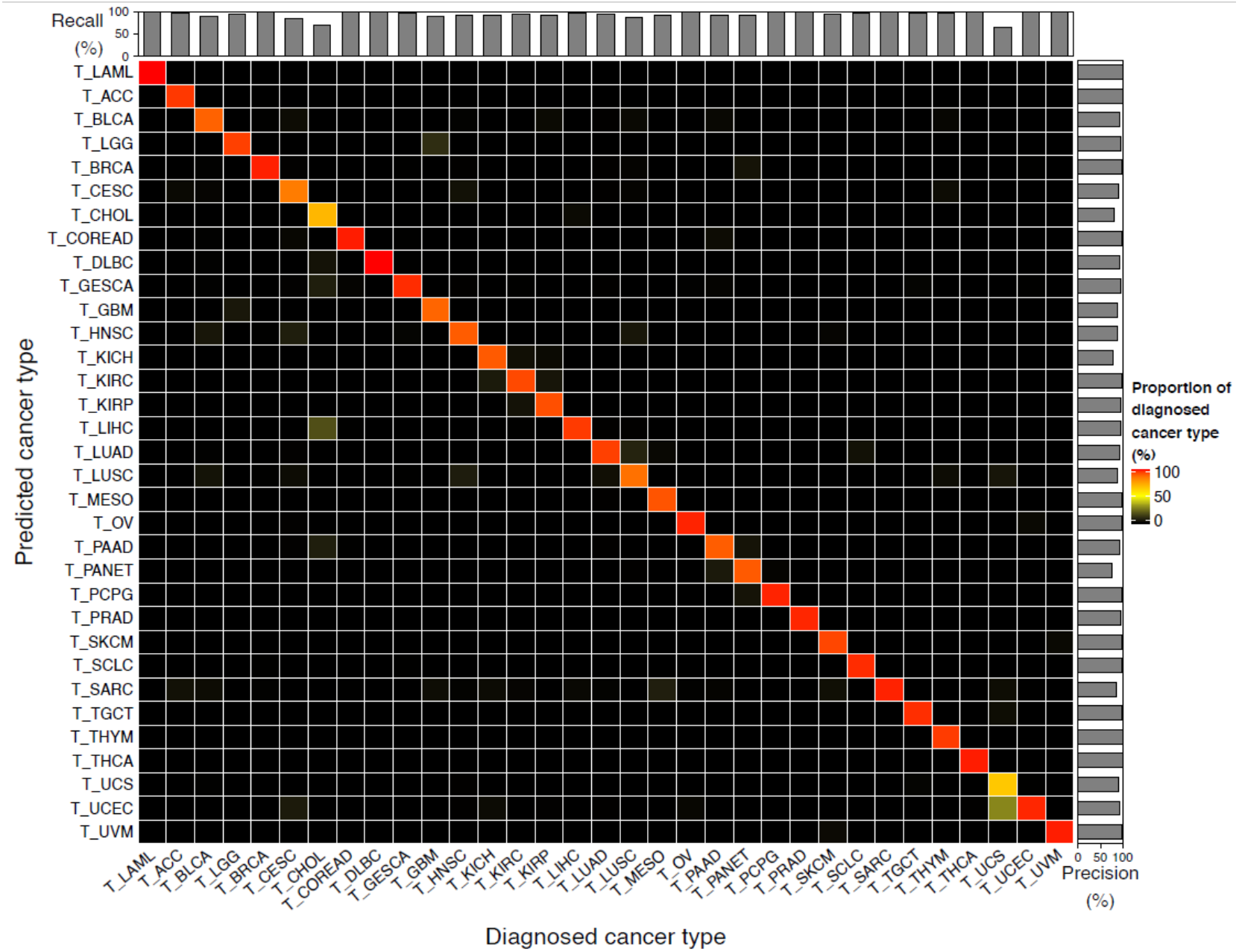
Workflow for classification of CUPs using RNA-seq



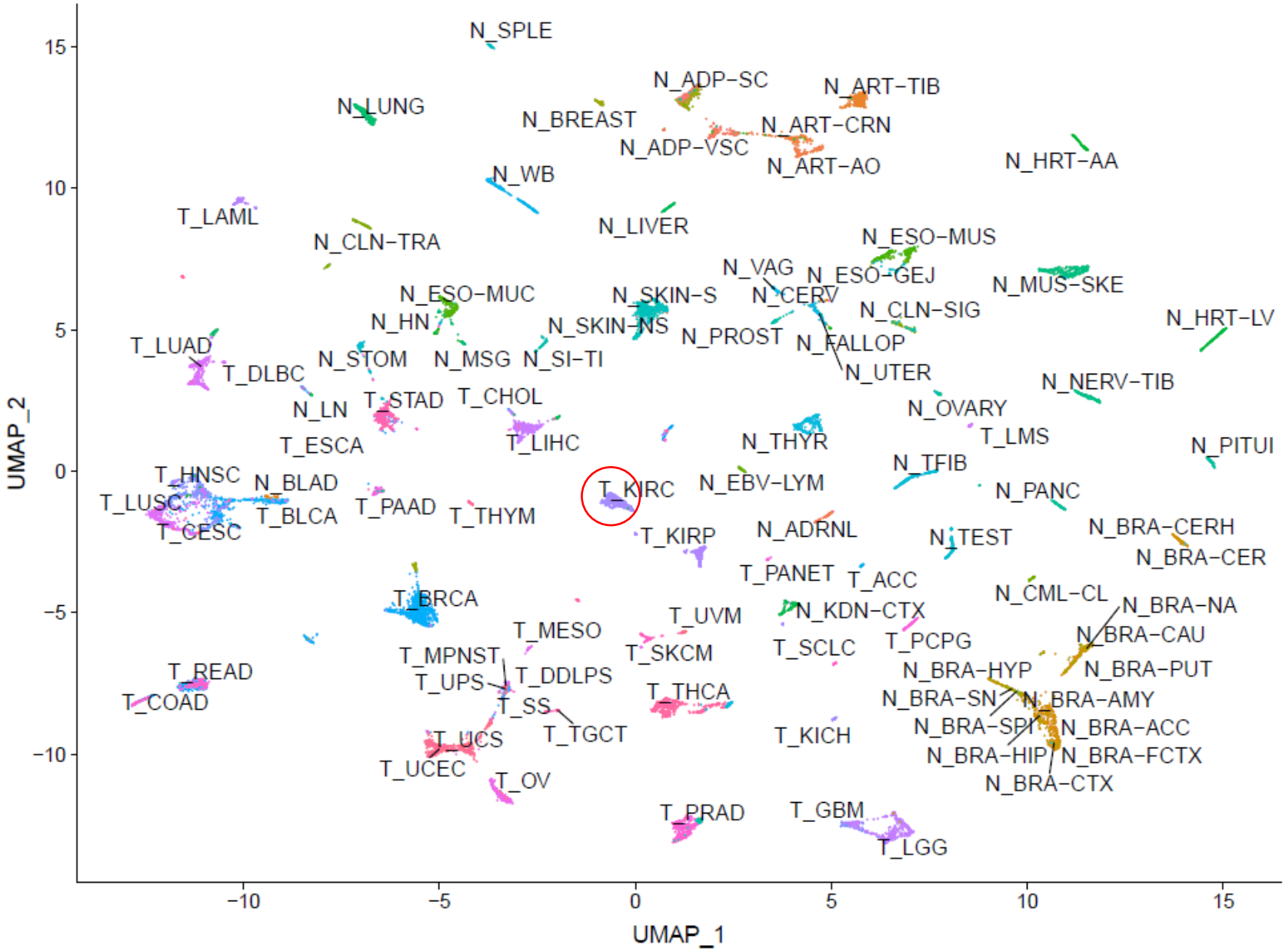
UMAP (2D representation) of all reference samples



Cross-validation confusion matrix for reference samples



Patient n°1


N_SPLE
N_ADP-SC
N_ART-TIB
N_LUNG
N_BREAST
N_ART-CRN
N_ADP-VSC
N_ART-AO
N_HRT-AA
N_WB
T_LAML
N_LIVER
N_CLN-TRA
N_ESO-MUS
N_VAG
N_ESO-GEJ
N_MUS-SKE
N_ESO-MUC
N_SKIN-S
N_CERV
N_CLN-SIG
N_HRT-LV
N_HN
N_SKIN-NS
N_PROST
N_FALLOP
T_LUAD
T_DLBC
N_STOM
N_MSG
N_SI-TI
N_UTER
N_NERV-TIB
N_LN
T_STAD
T_CHOL
N_OVARY
T_LMS
T_ESCA
T_LIHC
N_THYR
N_PITUI
N_TFIB
T_HNSC
N_BLAD
T_KIRC
N_EBV-LYM
T_LUSC
T_BLCA
T_PAAD
T_THYM
N_PANC
T_CESC
T_KIRP
N_ADRNL
N_TEST
N_BRA-CERH
N_BRA-CER
T_PANET
T_ACC
T_BRCA
N_CML-CL
N_BRA-NA
T_UVM
N_KDN-CTX
T_MESO
N_BRA-CAU
T_SCLC
T_PCPG
T_MPNST
T_SKCM
N_BRA-HYP
N_BRA-PUT
T_READ
T_UPS
T_DDLPS
T_THCA
N_BRA-SN
N_BRA-AMY
T_COAD
T_SS
N_BRA-SPI
N_BRA-ACC
T_KICH
T_UCS
T_TGCT
N_BRA-HIP
N_BRA-FCTX
T_UCEC
T_OV
N_BRA-CTX
T_PRAD
T_GBM
T_LGG
UMAP_2
UMAP_1
15
10
5
0
−5
−10
−10
−5
0
5
10
15

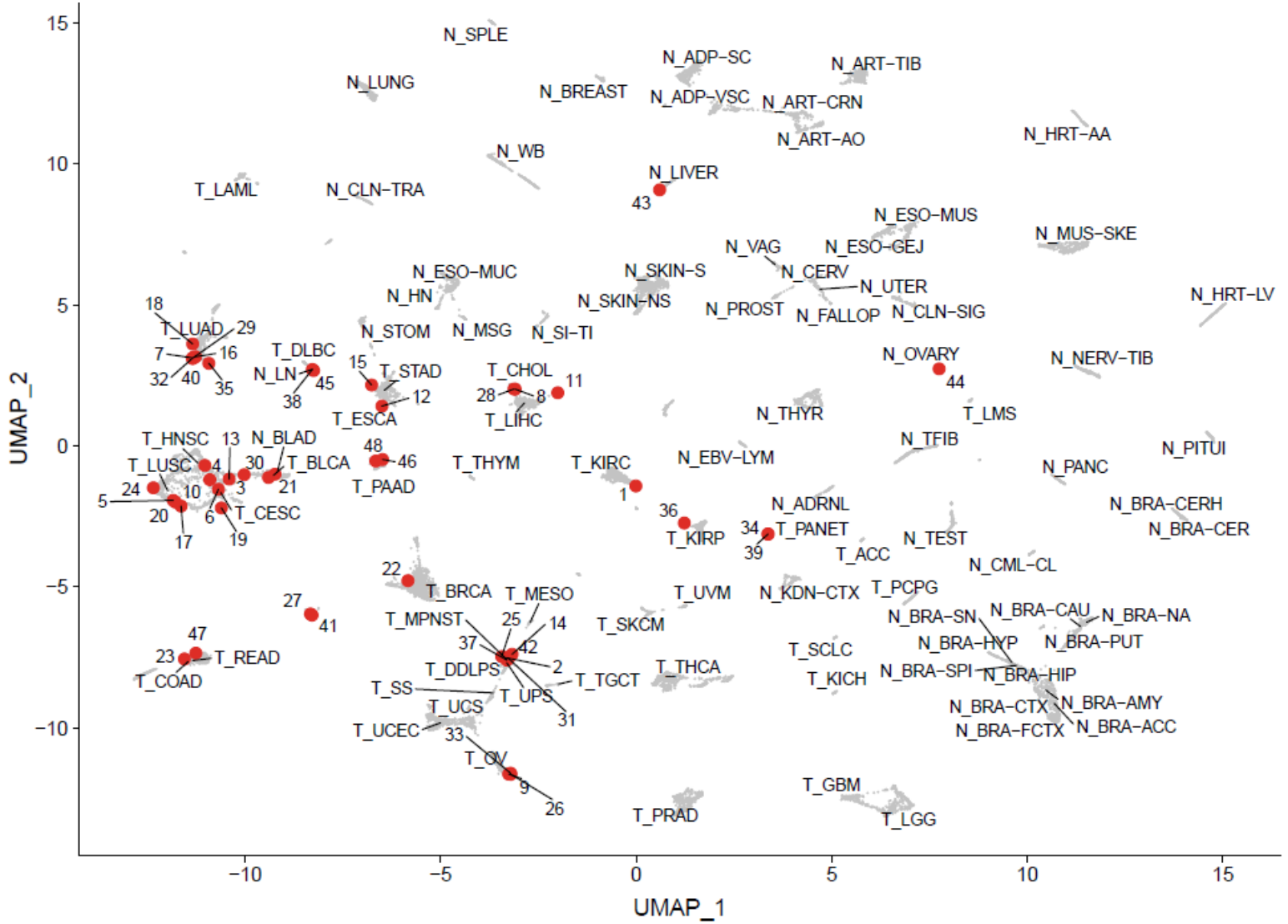
Patient n°1

- **Extrarenal renal cell carcinoma**
- Patient treated accordingly with anti-angiogenic therapy + immunotherapy

Patient n°1

- **Extrarenal renal cell carcinoma**
- Patient treated accordingly with anti-angiogenic therapy + immunotherapy
- -> *Complete response (disappearance of tumor)*

Diagnostic prediction using RNA-seq on a series of 48 CUPs



The Journal of Molecular Diagnostics, Vol. 23, No. 10, October 2021

the Journal of Molecular Diagnostics
jmdjournal.org

Identification of Tissue of Origin and Guided Therapeutic Applications in Cancers of Unknown Primary Using Deep Learning and RNA Sequencing (TransCUPtomics)

Julien Vibert,[*] Gaëlle Pierron,[†] Camille Benoist,[‡] Nadège Gruel,[*§] Delphine Guillemot,[†] Anne Vincent-Salomon,[¶] Christophe Le Tourneau,[‖] Alain Livartowski,[**] Odette Mariani,[¶] Sylvain Baulande,[††] François-Clément Bidard,[**‡‡] Olivier Delattre,[*†] Joshua J. Waterfall,[§§] and Sarah Watson[*,**]

From the INSERM U830, Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, the Department of Translational Research,§ PSL Research University, the Institut Curie Genomics of Excellence (ICGex) Platform,†† PSL Research University, and the INSERM U830,§§ PSL Research University, Institut Curie Research Center, Paris; the Somatic Genetics Unit,† Department of Genetics, the Clinical Bioinformatic Unit,‡ Department of Diagnostic and Theranostic Medecine, the Department of Diagnostic and Theranostic Medecine,¶ and the Department of Medical Oncology,** Institut Curie Hospital, Paris; the Department of Drug Development and Innovation,‖ INSERM U900, Paris-Saclay University, Institut Curie Hospital and Research Center, Paris and Saint-Cloud; and the INSERM CIC-BT 1428,‡‡ UVSQ, Paris-Saclay University, Saint-Cloud, France*

Accepted for publication July 14, 2021.

Address correspondence to Sarah Watson, M.D., Ph.D., INSERM U830, Équipe Labellisée Ligue Nationale Contre le Cancer, Diversity and Plasticity of Childhood Tumors Lab, PSL Research University, Institut Curie Research Center, Paris, France, Department of Medical Oncology, Institut Curie Hospital, Paris, France, 26 rue d'Ulm 75005 Paris, France. E-mail: sarah.watson@curie.fr.

Cancers of unknown primary (CUP) are metastatic cancers for which the primary tumor is not found despite thorough diagnostic investigations. Multiple molecular assays have been proposed to identify the tissue of origin (TOO) and inform clinical care; however, none has been able to combine accuracy, interpretability, and easy access for routine use. We developed a classifier tool based on the training of a variational autoencoder to predict tissue of origin based on RNA-sequencing data. We used as training data 20,918 samples corresponding to 94 different categories, including 39 cancer types and 55 normal tissues. The TransCUPtomics classifier was applied to a retrospective cohort of 37 CUP patients and 11 prospective patients. TransCUPtomics exhibited an overall accuracy of 96% on reference data for TOO prediction. The TOO could be identified in 38 (79%) of 48 CUP patients. Eight of 11 prospective CUP patients (73%) could receive first-line therapy guided by TransCUPtomics prediction, with responses observed in most patients. The variational autoencoder added further utility by enabling prediction interpretability, and diagnostic predictions could be matched to detection of gene fusions and expressed variants. TransCUPtomics confidently predicted TOO for CUP and enabled tailored treatments leading to significant clinical responses. The interpretability of our approach is a powerful addition to improve the management of CUP patients. *(J Mol Diagn 2021, 23: 1380–1392; https://doi.org/10.1016/j.jmoldx.2021.07.009)*

- Tool currently in use at the Institut Curie
- To be deployed nationally with the French program of genomic profiling of CUPs

Other AI approaches for CUP classification

Deep learning on mutational data (DNA-based)

ARTICLE

https://doi.org/10.1038/s41467-019-13825-8 OPEN

A deep learning system accurately classifies primary and metastatic cancers using passenger mutation patterns

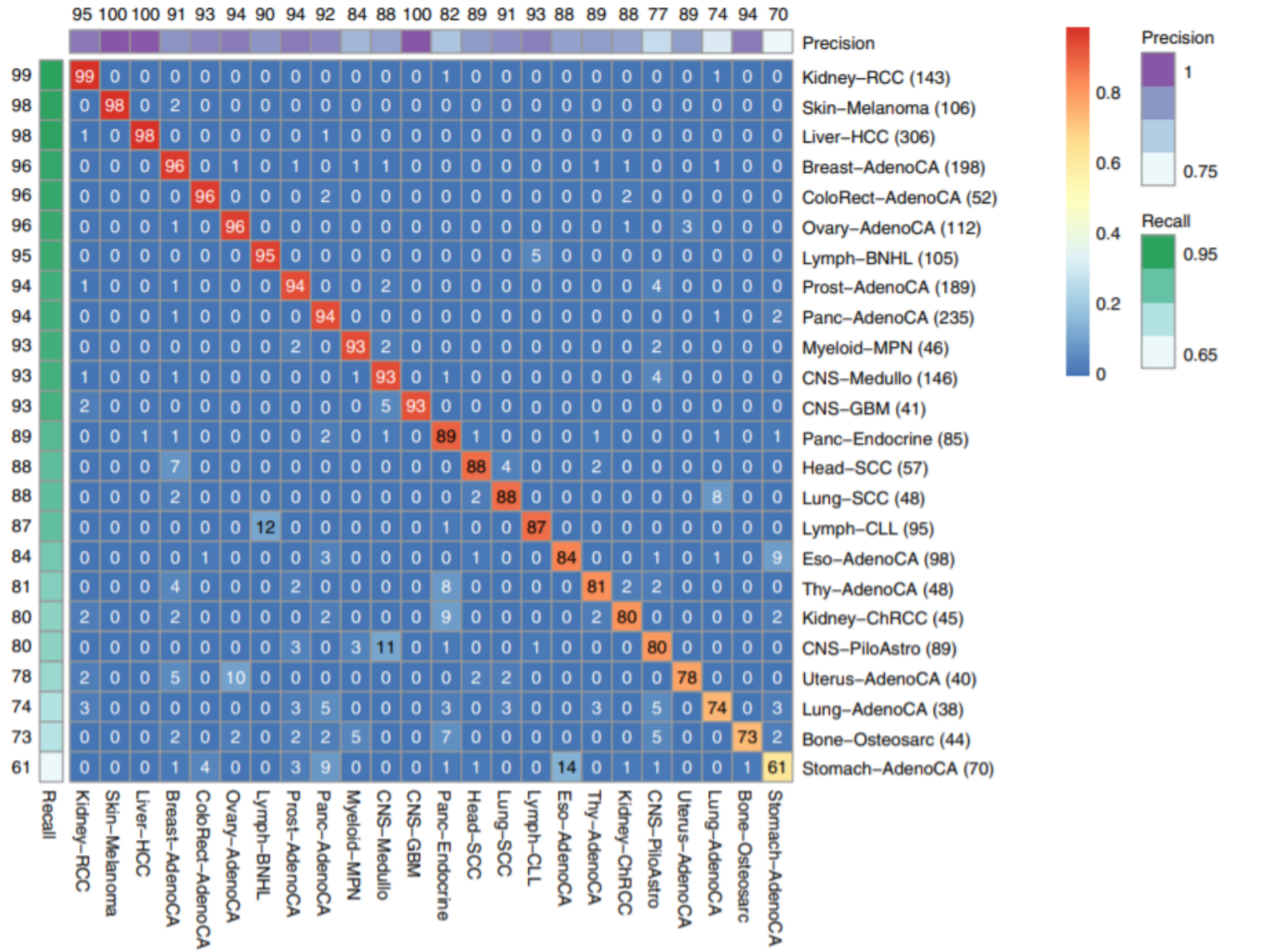
Wei Jiao[1,63], Gurnit Atwal[1,2,3,63], Paz Polak[4,63], Rosa Karlic[5], Edwin Cuppen[6,7], PCAWG Tumor Subtypes and Clinical Translation Working Group, Alexandra Danyi[8], Jeroen de Ridder[8], Carla van Herpen[9], Martijn P. Lolkema[10], Neeltje Steeghs[11], Gad Getz[12], Quaid Morris[1,2,3,13,14,64], Lincoln D. Stein[1,2,64*] & PCAWG Consortium

Methods

- Deep learning classifier
- Trained on whole-genome sequencing (WGS) data
- Uses patterns of mutations to predict tissue of origin

Table 2 WGS feature types used in classifiers.

Feature category	Feature type	Feature count	Description
Mutation distribution	SNV-BIN	2897	Number of SNVs per 1-Mbp bin, and per chromosome, normalised against the total number of SNVs per sample
	CNA-BIN	2826	Number of CNAs per 1-Mbp bin
	SV-BIN	2929	Number of SVs per 1-Mbp bin, and per chromosome, normalised against the total number of SV per sample
	INDEL-BIN	2757	Number of SNVs per 1-Mbp bin, and per chromosome, normalised against the total number of INDEL per sample
Mutation type	MUT-WGS	150	Type of single-nucleotide substitution, double- and triple-nucleotide substitution (plus its adjacent nucleotide neighbours)
Driver gene/pathway	GEN	554	Presence of an impactful mutation in a suspected driver gene
	MOD	1865	Presence of an impactful mutation in a gene belonging to a suspected driver pathway


95 100 100 91 93 94 90 94 92 84 88 100 82 89 91 93 88 89 88 77 89 74 94 70
Precision
99 Kidney–RCC (143)
98 Skin–Melanoma (106)
98 Liver–HCC (306)
96 Breast–AdenoCA (198)
96 ColoRect–AdenoCA (52)
96 Ovary–AdenoCA (112)
95 Lymph–BNHL (105)
94 Prost–AdenoCA (189)
94 Panc–AdenoCA (235)
93 Myeloid–MPN (46)
93 CNS–Medullo (146)
93 CNS–GBM (41)
89 Panc–Endocrine (85)
88 Head–SCC (57)
88 Lung–SCC (48)
87 Lymph–CLL (95)
84 Eso–AdenoCA (98)
81 Thy–AdenoCA (48)
80 Kidney–ChRCC (45)
80 CNS–PiloAstro (89)
78 Uterus–AdenoCA (40)
74 Lung–AdenoCA (38)
73 Bone–Osteosarc (44)
61 Stomach–AdenoCA (70)
Recall
Kidney–RCC
Skin–Melanoma
Liver–HCC
Breast–AdenoCA
ColoRect–AdenoCA
Ovary–AdenoCA
Lymph–BNHL
Prost–AdenoCA
Panc–AdenoCA
Myeloid–MPN
CNS–Medullo
CNS–GBM
Panc–Endocrine
Head–SCC
Lung–SCC
Lymph–CLL
Eso–AdenoCA
Thy–AdenoCA
Kidney–ChRCC
CNS–PiloAstro
Uterus–AdenoCA
Lung–AdenoCA
Bone–Osteosarc
Stomach–AdenoCA
0
0.2
0.4
0.6
0.8
Precision
1
0.75
Recall
0.95
0.65

Deep learning on pathology slides

Article

AI-based pathology predicts origins for cancers of unknown primary

https://doi.org/10.1038/s41586-021-03512-4

Received: 27 June 2020

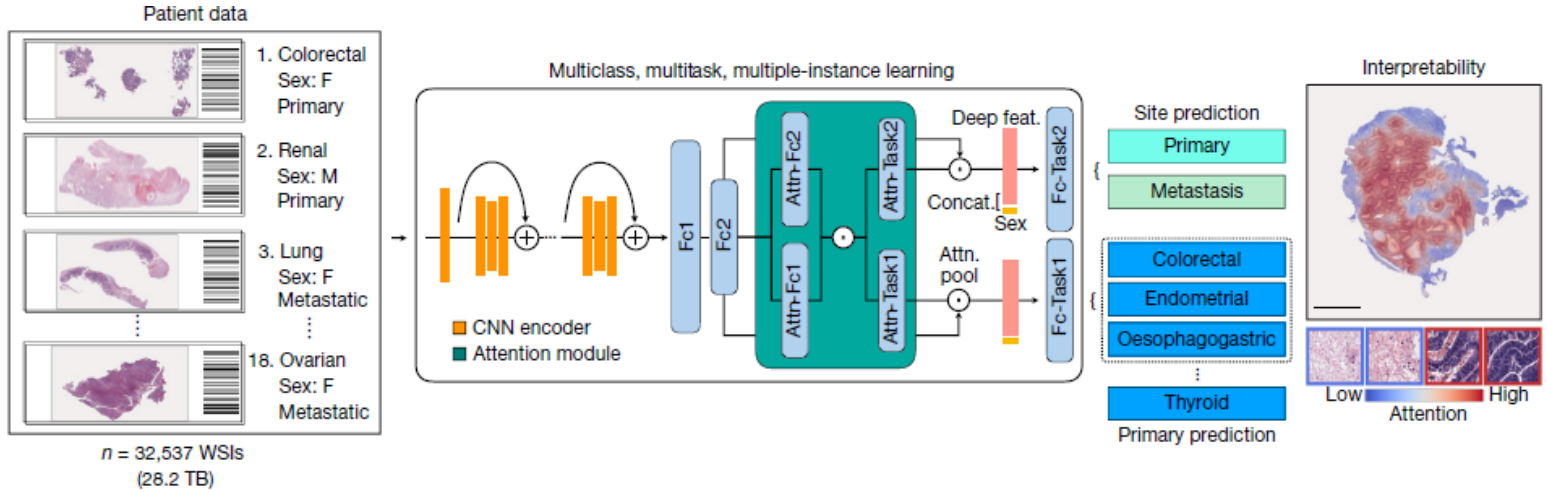
Accepted: 1 April 2021

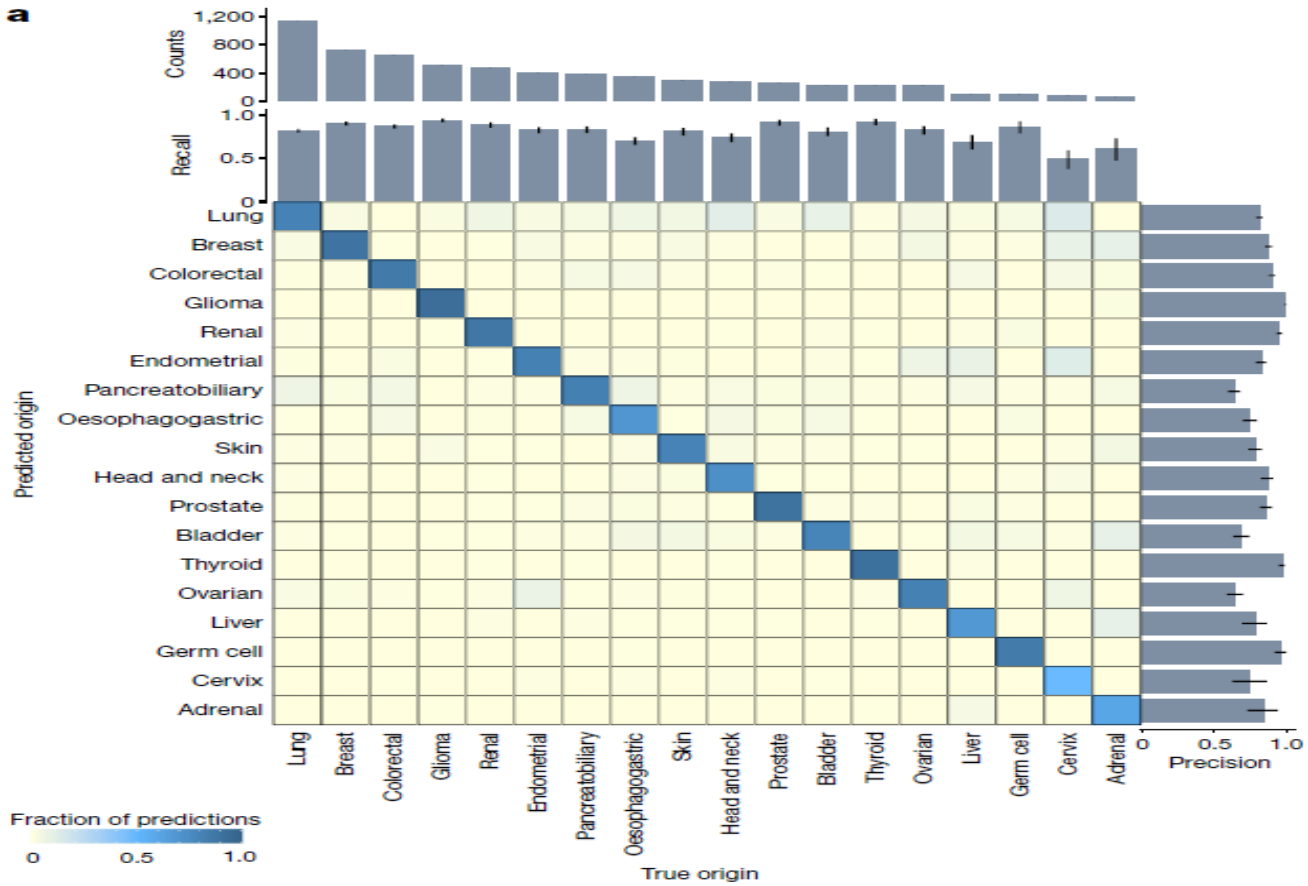
Published online: 5 May 2021

Check for updates

Ming Y. Lu[1,2,3], Tiffany Y. Chen[1,2,5], Drew F. K. Williamson[1,2,5], Melissa Zhao[1], Maha Shady[1,2,3,4], Jana Lipkova[1,2,3] & Faisal Mahmood[1,2,3]✉

Cancer of unknown primary (CUP) origin is an enigmatic group of diagnoses in which the primary anatomical site of tumour origin cannot be determined[1,2]. This poses a considerable challenge, as modern therapeutics are predominantly specific to the primary tumour[3]. Recent research has focused on using genomics and transcriptomics to identify the origin of a tumour[4–9]. However, genomic testing is not always performed and lacks clinical penetration in low-resource settings. Here, to overcome these challenges, we present a deep-learning-based algorithm—Tumour Origin Assessment via Deep Learning (TOAD)—that can provide a differential diagnosis for the origin of the primary tumour using routinely acquired histology slides. We used whole-slide images of tumours with known primary origins to train a model that simultaneously identifies the tumour as primary or metastatic and predicts its site of origin. On our held-out test set of tumours with known primary origins, the model achieved a top-1 accuracy of 0.83 and a top-3 accuracy of 0.96, whereas on our external test set it achieved top-1 and top-3 accuracies of 0.80 and 0.93, respectively. We further curated a dataset of 317 cases of CUP for which a differential diagnosis was assigned. Our model predictions resulted in concordance for 61% of cases and a top-3 agreement of 82%. TOAD can be used as an assistive tool to assign a differential diagnosis to complicated cases of metastatic tumours and CUPs and could be used in conjunction with or in lieu of ancillary tests and extensive diagnostic work-ups to reduce the occurrence of CUP.


Patient data
1. Colorectal
Sex: F
Primary
2. Renal
Sex: M
Primary
3. Lung
Sex: F
Metastatic
18. Ovarian
Sex: F
Metastatic
n = 32,537 WSIs
(28.2 TB)
Multiclass, multitask, multiple-instance learning
CNN encoder
Attention module
Fc1
Fc2
Attn-Fc2
Attn-Fc1
Attn-Task2
Attn-Task1
Deep feat.
Concat.
Sex
Attn. pool
Fc-Task2
Fc-Task1
Site prediction
Primary
Metastasis
Colorectal
Endometrial
Oesophagogastric
Thyroid
Primary prediction
Interpretability
Low
High
Attention


a
Counts
1,200
800
400
0
Recall
1.0
0.5
0
Lung
Breast
Colorectal
Glioma
Renal
Endometrial
Pancreatobiliary
Oesophagogastric
Skin
Head and neck
Prostate
Bladder
Thyroid
Ovarian
Liver
Germ cell
Cervix
Adrenal
Predicted origin
True origin
Precision
0
0.5
1.0
Fraction of predictions
0
0.5
1.0

Acknowledgments

- Sarah Watson
- My thesis directors:
  Olivier Delattre and Joshua Waterfall
- U830 Institut Curie
- Unité de Génétique Somatique (UGS)
- Patients and families
- Doctors and other healthcare workers

- **THANK YOU FOR**
- **YOUR ATTENTION!**