

All models are wrong and yours are useless

Making clinical prediction tools
impactful for patients

<https://doi.org/10.1038/s41698-024-00553-6>

All models are wrong and yours are useless: making clinical prediction models impactful for patients

Florian Markowetz



Check for updates

Most published clinical prediction models are never used in clinical practice and there is a huge gap between academic research and clinical implementation. Here, I propose ways for academic researchers to be proactive partners in improving clinical practice and to design models in ways that ultimately benefit patients.

“All models are wrong, but some are useful” is an aphorism attributed to the statistician George Box. There is humility in claiming your model is

By all academic standards our work was a success, it has been widely noted in the community, and the papers are well-cited. Still, it feels to me like there is something missing. Over the years I have come to see academic papers not as ends in themselves, but as the beginning of the journey to clinical implementation, and I am frustrated with how little of my own work ever had clinical impact. Looking back, here are some lessons I had to learn over the last two decades:

Observation 1: *Success in academia is not the same as success in the*

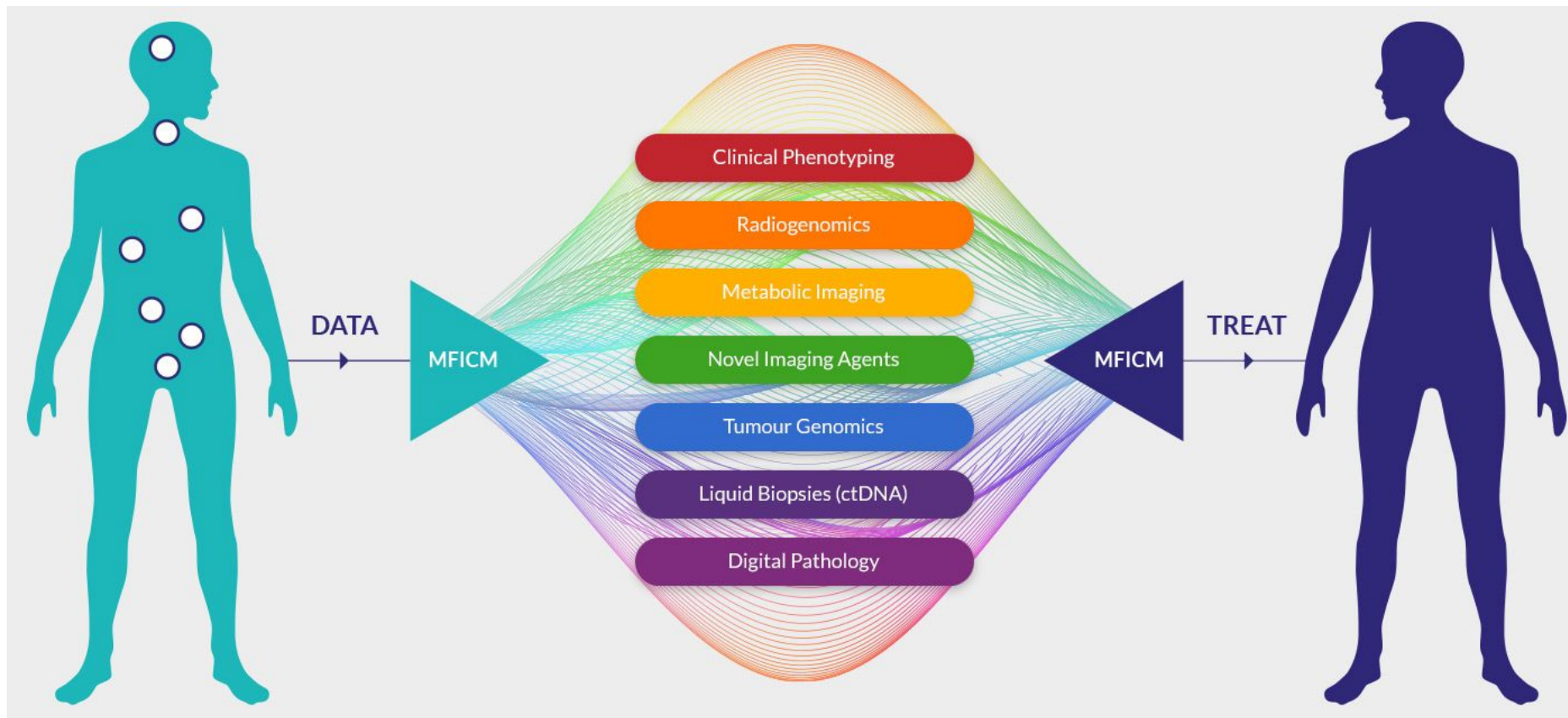
R

RESTRICTED



**UNDER 17 REQUIRES ACCOMPANYING
PARENT OR ADULT GUARDIAN**

Strong Views And Harsh Truths



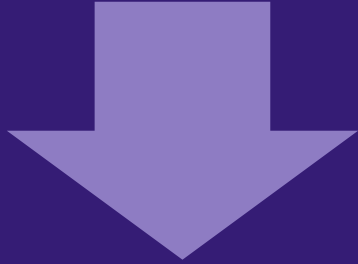
**“This model could
be used in the clinic”**

**“This model could
probably have the
potential to maybe
be used in the clinic”**

**“realising the potential of
pReclsIon MeDlcīNe”**

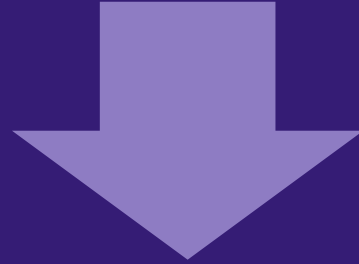
Science

Discovery



Insights

Translation



Actions

A meme featuring a woman with red hair crying. The image is composed of three overlapping rectangular frames. The top frame shows the back of her head and shoulders. The middle frame shows her face from the side, looking down. The bottom frame is the largest and shows her face from the front, crying with her eyes closed. A hand is visible on the right side of her face, touching her cheek. The text "YOUR MAN" is in the bottom left, and "AND NONE OF THEM ARE USED" is at the bottom. A watermark "imgflip.com" is in the bottom left of the bottom frame.

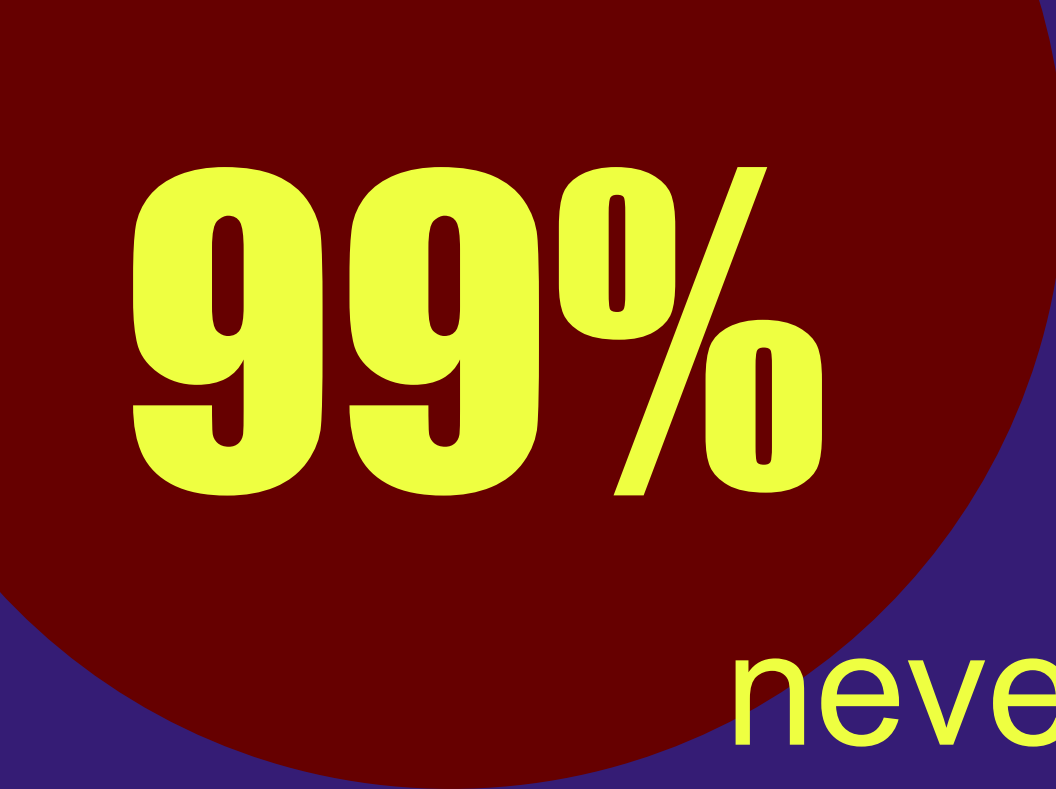
**YOUR
MAN**

AND NONE OF THEM ARE USED

imgflip.com

Landscape of clinical prediction models

- 408 models for COPD prognosis (Bellou, 2019)
- 363 models for cardiovascular disease general population (Damen, 2016)
- 263 prognosis models in obstetrics (Kleinrouweler, 2016)
- 258 models mortality after general trauma (Munter, 2017)
- 232 models related to COVID-19 (Wynants, 2020)
- 160 female-specific models for cardiovascular disease (Baart, 2019)
- 119 models for critical care prognosis in LMIC (Haniffa, 2018)
- 101 models for primary gastric cancer prognosis (Feng, 2019)
- 99 models for neck pain (Wingbermhühle, 2018)
- 81 models for sudden cardiac arrest (Carrick, 2020)
- 74 models for contrast-induced acute kidney injury (Allen, 2017)
- 73 models for 28/30 day hospital readmission (Zhou, 2016)
- 68 models for preeclampsia (De Kat, 2019)
- 67 models for traumatic brain injury prognosis (Dijkland, 2019)
- 64 models for suicide / suicide attempt (Belsher, 2019)
- 61 models for dementia (Hou, 2019)
- 58 models for breast cancer prognosis (Phung, 2019)
- 52 models for pre-eclampsia (Townsend, 2019)
- 52 models for colorectal cancer risk (Usher-Smith, 2016)
- 48 models for incident hypertension (Sun, 2017)
- 46 models for melanoma (Kaiser, 2020)
- 46 models for prognosis after carotid revascularisation (Volkers, 2017)
- 43 models for mortality in critically ill (Keuning, 2019)
- 42 models for kidney failure in chronic kidney disease (Ramspek, 2019)
- 40 models for incident heart failure (Sahle, 2017)
- 37 models for treatment response in pulmonary TB (Peetluk, 2021)
- 35 models for in vitro fertilisation (Ratna, 2020)
- 34 models for stroke in type-2 diabetes (Chowdhury, 2019)
- 34 models for graft failure in kidney transplantation (Kabore, 2017)
- 31 models for length of stay in ICU (Verburg, 2016)
- 30 models for low back pain (Haskins, 2015)
- 27 models for pediatric early warning systems (Trubey, 2019)
- 27 models for malaria prognosis (Njim, 2019)
- 26 models for postoperative outcomes colorectal cancer (Souwer, 2020)
- 26 models for childhood asthma (Kothalawa, 2020)
- 25 models for lung cancer risk (Gray, 2016)
- 25 models for re-admission after admitted for heart failure (Mahajan, 2018)
- 23 models for recovery after ischemic stroke (Jampathong, 2018)
- 23 models for delirium in older adults (Lindroth, 2018)
- 21 models for atrial fibrillation detection in community (Himmelreich, 2020)
- 19 models for survival after resectable pancreatic cancer (Stijker, 2019)
- 18 models for recurrence hep. carcinoma after liver transplantation (Al-Ameri, 2020)
- 18 models for future hypertension in children (Hamoen, 2018)
- 18 models for risk of falls after stroke (Walsh, 2016)
- 18 models for mortality in acute pancreatitis (Di, 2016)
- 17 models for bacterial meningitis (van Zeggeren, 2019)
- 17 models for cardiovascular disease in hypertensive population (Cai, 2020)
- 14 models for ICU delirium risk (Chen, 2020)
- 14 models for diabetic retinopathy progression (Haider, 2019)



99%

of models
never get used
for patient benefit

When is a prediction model successful?



- Novelty of the algorithm
- Number of citations
- Glamour factor of journal

When is a prediction model successful?



If it helps patients!

Real success =

- Makes people's lives better
(impact beyond my CV)
- Routinely used
(not just one-off in academic lab)
- World/Europe/UK-wide
(not just in a single centre of excellence)



ARTICLE

The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups

Christina Curtis^{1,2,*}, Sohrab P. Shah^{3,4,*}, Suet-Feung Chin^{1,2,*}, Gulisa Turashvili^{3,4,*}, Oscar M. Rueda^{1,2}, Mark J. Dunning², Doug Speed^{5,6,*}, Andy G. Lynch^{1,2}, Shamith Samarajiwa^{1,2}, Yinyin Yuan^{1,2}, Stefan Gräf^{1,2}, Gavin Ha¹, Gholamreza Haffari¹, Ali Bashashati¹, Roslin Russell¹, Steven McKinney^{1,2}, METABRIC Group^{1,2}, Anita Langerod⁷, Andrew Green⁸, Elena Provenzano⁹, Gordon Wishart¹⁰, Sarah Pinder¹¹, Peter Watson^{1,12,13}, Florian Markowitz^{1,2}, Leigh Murphy^{1,2}, Ian Ellis¹, Annie Parushotham^{1,13}, Anne-Lise Borresen-Dale^{1,12}, James D. Brenton^{2,11}, Simon Tavaré^{1,2,5,14}, Carlos Caldas^{2,5,14} & Samuel Aparicio^{1,14}

The elucidation of breast cancer subgroups and their molecular drivers requires integrated views of the genome and transcriptome from representative numbers of patients. We present an integrated analysis of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in ~40% of genes, with the landscape dominated by *cis*- and *trans*-acting CNAs. By delineating expression outlier genes driven in *cis* by CNAs, we identified putative cancer genes, including deletions in *PPP2R2A*, *MTAP* and *MAP2K4*. Unsupervised analysis of paired DNA–RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, oestrogen-receptor-positive 1q/13/14 *cis*-acting subgroup and a favourable prognosis subgroup devoid of CNAs. *Trans*-acting aberration hotspots were found to modulate subgroup-specific gene networks, including a TCR deletion-mediated adaptive immune response in the ‘CNA-devoid’ subgroup and a basal-specific chromosome 5 deletion-associated mitotic network. Our results provide a novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome.

Inherited genetic variation and acquired genomic aberrations contribute to breast cancer initiation and progression. Although somatically acquired CNAs are the dominant feature of sporadic breast cancers, the driver events that are selected for during tumorigenesis are difficult to elucidate as they co-occur alongside a much larger landscape of random non-pathogenic passenger alterations^{1,2} and germline copy number variants (CNVs). Attempts to define subtypes of breast cancer and to discern possible somatic drivers are still in their infancy^{3–6}, in part because breast cancer represents multiple diseases, implying that large numbers (many hundreds or thousands) of patients must be studied. Here we describe an integrated genomic/transcriptomic analysis of breast cancers with long-term clinical outcomes composed of a discovery set of 997 primary tumours and a validation set of 995 tumours from METABRIC (Molecular Taxonomy of Breast Cancer International Consortium).

A breast cancer population genomic resource

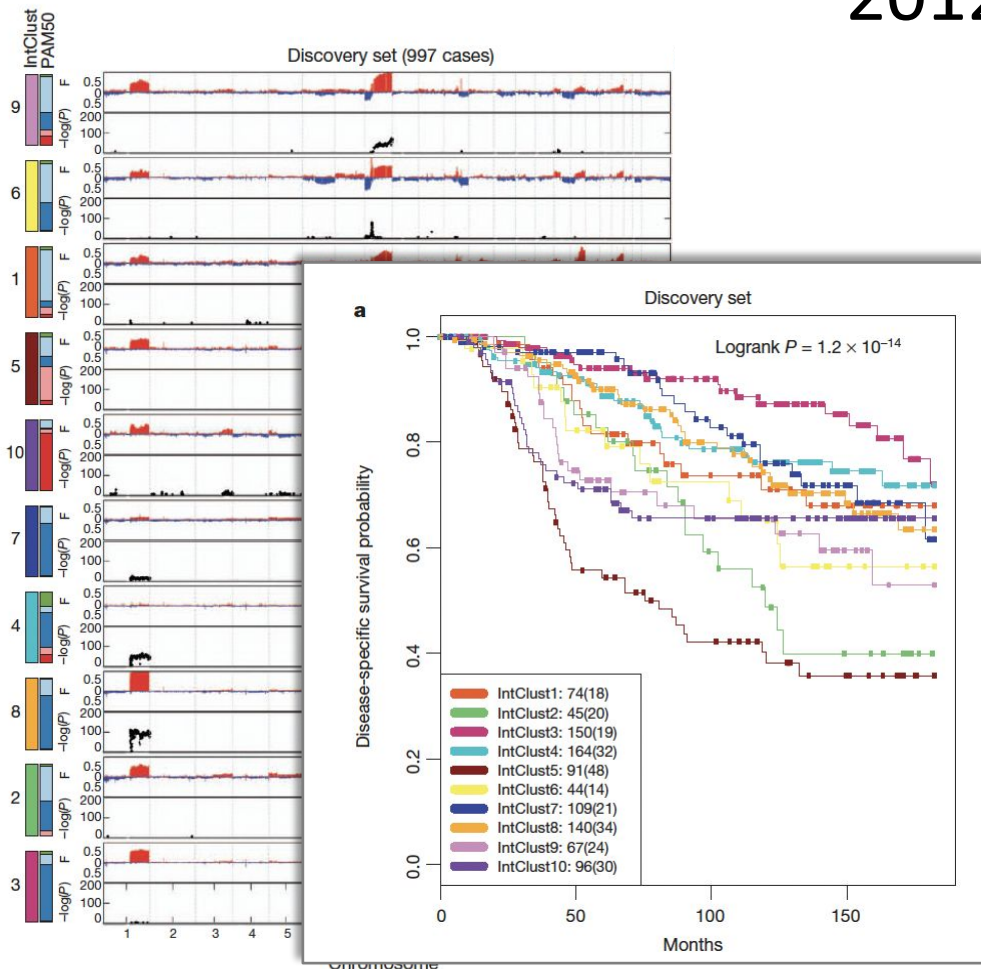
We assembled a collection of over 2,000 clinically annotated primary fresh-frozen breast cancer specimens from tumour banks in the UK

and Canada (Supplementary Tables 1–3). Nearly all oestrogen receptor (ER)-positive and/or lymph node (LN)-negative patients did not receive chemotherapy, whereas ER-negative and LN-positive patients did. Additionally, none of the HER2⁺ patients received trastuzumab. As such, the treatments were homogeneous with respect to clinically relevant groupings. An initial set of 997 tumours was analysed as a discovery cohort and a further set of 995 tumours, for which complete data later became available, was used to test the reproducibility of the integrative clusters (described below). An overview of the main analytical approaches is provided in Supplementary Fig. 1. Details concerning expression and copy number profiling, including sample assignment to the PAM50 intrinsic subtypes^{4,7} (Supplementary Fig. 2), copy number analysis (Supplementary Tables 4–8) and validation (Supplementary Figs 3 and 4 and Supplementary Tables 9–11), and TP53 mutational profiling (Supplementary Fig. 5) are described in the Supplementary Information.

Genome variation affects tumour expression architecture

Genomic variants are considered to act in *cis* when a variant at a locus has an impact on its own expression, or in *trans* when it is associated

doi:10.1038/nature10983



¹Department of Oncology, University of Cambridge, Hills Road, Cambridge CB2 2ZL, UK. ²Cancer Research UK, Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ³Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia V6T 2B5, Canada. ⁴Molecular Oncology, British Columbia Cancer Research Centre, Vancouver, British Columbia V6T 1Z5, Canada. ⁵Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Centre for Mathematical Sciences, Cambridge CB3 0WA, UK. ⁶Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Montebello, 0310 Oslo, Norway. ⁷Department of Histopathology, School of Molecular Medical Sciences, University of Nottingham, Nottingham NG5 1PB, UK. ⁸Cambridge Breast Unit, Addenbrooke's Hospital, Cambridge Hospital NHS Foundation Trust and NIHR Cambridge Biomedical Research Centre, Cambridge CB2 0QQ, UK. ⁹King's College London, Breastthrough Breast Cancer Research CB2 0LL, UK. ¹⁰Memorial Institute of Ovarian Biology, University of Manchester, Manchester M13 9PL, UK. ¹¹NIHR Comprehensive Biomedical Research Centre at Guy's and St Thomas' NHS Foundation Trust and King's College London, London WC2R 2LS, UK. ¹²Institute for Clinical Medicine, Faculty of Medicine, University of Oslo, 0316 Oslo, Norway. ¹³Cambridge Experimental Cancer Medicine Centre, Cambridge CB2 0RE, UK. ¹⁴Molecular and Computational Biology Program, University of Southern California, Los Angeles, California 90089, USA. *Present address: ¹⁵Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California 90033, USA (Ch.C.); ¹⁶University College London, Genetics Institute, WC1E 6BT, UK (D.S.).

*These authors contributed equally to this work.

†List of participants and affiliations appear at the end of the paper.

ANALYSIS

nature
medicine

The consensus molecular subtypes of colorectal cancer

Justin Guinney^{1,21}, Rodrigo Dienstmann^{1,2,21}, Xin Wang^{3,4,21}, Aurélien de Reyniès^{5,21}, Andreas Schlicker^{6,21}, Charlotte Sonesson^{7,21}, Laetitia Marisa^{8,21}, Paul Roepman^{8,21}, Gift Niyamundanda^{9,21}, Paolo Angelico⁷, Brian M Bot¹, Jeffrey S Morris¹⁰, Iris M Simon⁸, Sarah Gerster⁷, Evelyn Fessler³, Felipe De Sousa E Melo³, Edouard Missiaglia⁷, Hena Ramay⁷, David Barras⁷, Kristian Homicsko¹¹, Dipen Maru¹⁰, Ganiraju C Manyam¹⁰, Bradley Broom¹⁰, Valerie Boige¹², Beatriz Perez-Villamil¹³, Ted Laderas¹, Ramon Salazar¹⁴, Joe W Gray¹⁵, Douglas Hanahan¹¹, Josep Tabernero⁷, Rene Bernards⁶, Stephen H Friend¹, Louis Laurent-Puig^{16,17,22}, Jan Paul Medema^{3,22}, Anguraj Sadanandam^{9,22}, Lodewyk Wessels^{6,22}, Mauro Delorenzi^{7,18,19,22}, Scott Kopetz^{10,22}, Louis Vermeulen^{3,22} & Sabine Tejpar^{30,22}

Colorectal cancer (CRC) is a frequently lethal disease with heterogeneous outcomes and drug responses. To resolve inconsistencies among the reported gene expression-based CRC classifications and facilitate clinical translation, we formed an international consortium dedicated to large-scale data sharing and analytics across expert groups. We show marked interconnectivity between six independent classification systems coalescing into four consensus molecular subtypes (CMSs) with distinguishing features: CMS1 (microsatellite instability immune, 14%), hypermutated, microsatellite unstable and strong immune activation; CMS2 (canonical, 37%), epithelial, marked WNT and MYC signaling activation; CMS3 (metabolic, 13%), epithelial and evident metabolic dysregulation; and CMS4 (mesenchymal, 23%), prominent transforming growth factor- β activation, stromal invasion and angiogenesis. Samples with mixed features (13%) possibly represent a transition phenotype or intratumoral heterogeneity.

We consider the CMS groups the most robust classification system currently available for CRC—with clear biological interpretability—and the basis for future clinical stratification and subtype-based targeted interventions.

Gene expression-based subtyping is widely accepted as a relevant source of disease stratification¹. Despite the technique's widespread use, its translational and clinical utility is hampered by discrepant results, which are probably related to differences in data processing and algorithms applied to diverse patient cohorts, sample preparation methods and gene expression platforms. In the absence of a clear methodological gold standard² to perform such analyses, a more general framework that integrates and compares multiple strategies is needed to define common disease patterns in a principled, unbiased manner. Here we describe such a framework and its application to elucidate the intrinsic subtypes of CRC.

A full list of affiliations appears at the end of the paper.

Received 6 March; accepted 6 September; published online 12 October 2015; doi:10.1038/nm.3967

Inspection of the published gene expression-based CRC classifications^{3–9} revealed only superficial similarities. For example, all of the groups identified one tumor subtype enriched for microsatellite instability (MSI) and one subtype characterized by high expression of mesenchymal genes, but they failed to achieve full consistency among the other subtypes. We envisioned that a comprehensive cross-comparison of subtype assignments obtained by the various approaches on a common set of samples could resolve inconsistencies in both the number and the interpretation of CRC subtypes. The CRC Subtyping Consortium (CRSC) was formed to assess the presence or absence of core subtype patterns among existing gene expression-based CRC subtyping algorithms. Recognizing that transcriptomics represents the level of high-throughput molecular data that is most intimately linked to cellular or tumor phenotype and clinical behavior, we also wanted to characterize the key biological features of the core subtypes, integrate and confront all other available data sources (mutation, copy number, methylation, microRNA and proteomics) and assess whether the subtype assignment correlated with patient outcome. Furthermore, our aim was to establish an important paradigm for collaborative, community-based cancer subtyping that will facilitate the translation of molecular subtypes into the clinic, not only for CRC but for other malignancies as well.

RESULTS
Comparison of published molecular subtyping platforms

We evaluated the results of six CRC subtyping algorithms^{3–5, 8}, each developed independently using different gene expression data sets and analytical approaches (Supplementary Tables 1 and 2). Figure 1 summarizes the workflow of our analysis. A total of 18 CRC data sets ($n = 4,151$ patients) from both public (GSE42284, GSE33113, GSE35582, GSE35896, GSE13067, GSE13294, GSE14333, GSE17536, GSE20916, GSE2109 and The Cancer Genome Atlas (TCGA)) and proprietary^{3,10} sources (Supplementary Table 3)—which consisted of multiple gene expression platforms (Affymetrix, Agilent and RNA-sequencing), sample types (fresh-frozen samples and formalin-fixed paraffin-embedded (FFPE) samples) and study designs (retrospective and prospective series and one clinical trial¹⁰)—were uniformly preprocessed and normalized from the raw formats to reduce technical variation. The six expert groups applied their

Personalized Medicine and Imaging

Practical and Robust Identification of Molecular Subtypes in Colorectal Cancer by Immunohistochemistry

Anne Trinh^{1,2}, Kari Trumpf¹, Felipe De Sousa E Melo^{4,5}, Xin Wang^{1,6}, Joan H. de Jong⁴, Evelyn Fessler⁷, Peter J.K. Kuppen⁷, Marlies S. Reimers⁷, Marloes Swets⁷, Miriam Koopman³, Iris D. Nagtegaal⁸, Marnix Jansen^{3,9}, Gerrit K.J. Hooijer⁹, George J.A. Offerhaus³, Onno Kranenburg³, Cornelis J. Punt¹¹, Jan Paul Medema⁴, Florian Markowitz⁷, and Louis Vermeulen⁴

Abstract

Purpose: Recent transcriptomic analyses have identified four distinct molecular subtypes of colorectal cancer with evident clinical relevance. However, the requirement for sufficient quantities of bulk tumor and difficulties in obtaining high-quality genome-wide transcriptome data from formalin-fixed paraffin-embedded tissue are obstacles toward widespread adoption of this taxonomy. Here, we develop an immunohistochemistry-based classifier to validate the prognostic and predictive value of molecular colorectal cancer subtyping in a multicenter study.

Experimental Design: Tissue microarrays from 1,076 patients with colorectal cancer from four different cohorts were stained for five markers (CDX2, FRMD6, HTR2B, ZEB1, and KER) by immunohistochemistry and assessed for microsatellite instability. An automated classification system was trained on one cohort using quantitative image analysis or semiquantitative pathologist scoring of the cores as input and applied to three independent clinical cohorts.

Results: This classifier demonstrated 87% concordance with the gold-standard transcriptome-based classification. Application to three validation datasets confirmed the poor prognosis of the mesenchymal-like molecular colorectal cancer subtype. In addition, retrospective analysis demonstrated the benefit of adding cetuximab to bevacizumab and chemotherapy in patients with RAS wild-type metastatic cancers of the canonical epithelial-like subtypes.

Conclusions: This study shows that a practical and robust immunohistochemical assay can be employed to identify molecular colorectal cancer subtypes and uncover subtype-specific therapeutic benefit. Finally, the described tool is available online for rapid classification of colorectal cancer samples, both in the format of an automated image analysis pipeline to score tumor core staining, and as a classifier based on semiquantitative pathology scoring. *Clin Cancer Res* 23(2): 367–38. ©2016 AACR.

Introduction

Colorectal cancer is a heterogeneous disease with an overall 5-year survival of below 60% (1). There is an urgent need to improve selection of early-stage patients who may benefit from adjuvant therapy, or to identify patients with metastasis who may profit from a specific targeted therapy. To facilitate this, stratification methods based on histopathologic characteristics are extensively implemented: For example, only patients with colorectal cancer with high-risk features such as high-grade and poorly differentiated morphology are believed to benefit from adjuvant chemotherapy (2). Although histopathologic classification is difficult to implement uniformly, associations with molecular characteristics have been noted, such as microsatellite instability (MSI) in serrated tumors (3). This provides a more robust/objective means of determining the suitability of a patient for a given therapy. For example, mutation in the KRAS/BRAF axis is a well-characterized determinant of resistance to anti-EGFR therapy in metastatic disease (4, 5). However, current mutational profiling provides only limited biomolecular understanding of the disease, particularly in chromosomal instable disease where the large heterogeneity in patient response to

¹Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, United Kingdom; ²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts; ³Cancer Center UMC Utrecht, University Medical Center Utrecht, Utrecht, the Netherlands; ⁴Laboratory for Experimental Oncology and Radiobiology (LEXOR), Center for Experimental Molecular Medicine (CEMM), Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; ⁵Department of Molecular Oncology, Genentech Inc., South San Francisco, California; ⁶Department of Biomedical Sciences, City University of Hong Kong, Kowloon Tong, Hong Kong; ⁷Department of Surgery, Leiden University Medical Center, Leiden, the Netherlands; ⁸Department of Pathology, Radboud University Medical Center, Nijmegen, the Netherlands; ⁹Department of Pathology, Lymphoma and Hematopathology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands; ¹⁰Centre for Tumour Biology, Barts Cancer Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, London, United Kingdom; ¹¹Department of Medical Oncology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands.

Note: Supplementary data for this article are available at *Clinical Cancer Research* Online. <http://dx.doi.org/10.1158/1078-0432.CCR-16-0680>

Corresponding Author: Louis Vermeulen, Academic Medical Center, University of Amsterdam, Meibergdreef 3, 1105AZ Amsterdam, the Netherlands. Phone: 31-20-5664777; Fax: 31-20-5677792; E-mail: l.vermeulen@amc.uva.nl

doi:10.1158/1078-0432.CCR-16-0680

©2016 American Association for Cancer Research.

www.aacrjournals.org

2013-2016

Clinical
Cancer
Research

Downloaded from <http://aacrjournals.org/doi/pdf/10.1158/1078-0432.CCR-16-0680> by University of Cambridge user on 06 October 2022

AACR 387

Article

Multi-omic machine learning predictor of breast cancer therapy response

<https://doi.org/10.1038/s41586-021-04278-5>

Received: 30 July 2021

Accepted: 23 November 2021

Published online: 7 December 2021

Open access

Check for updates

Stephan-John Sammut^{1,2}, Mireia Crispin-Ortuza^{1,3}, Suet-Feung Chin^{1,5}, Elena Provenzano³, Helen A. Bardwell¹, Wenxin Ma⁴, Wei Cope⁶, Ali Darilush^{1,5}, Sarah-Jane Dawson^{6,7}, Jean E. Abraham^{3,7}, Janet Dunn⁸, Louise Hillier⁹, Jeremy Thomas^{8,10}, David A. Cameron¹, John M. S. Bartlett^{11,12}, Larry Hayward³, Paul D. Pharoah^{11,13}, Florian Markowetz¹, Oscar M. Rueda^{1,4}, Helena M. Earl^{2,3} & Carlos Caldas^{1,3,14}

Breast cancers are complex ecosystems of malignant cells and the tumour microenvironment¹. The composition of these tumour ecosystems and interactions within them contribute to responses to cytotoxic therapy². Efforts to build response predictors have not incorporated this knowledge. We collected clinical, digital pathology, genomic and transcriptomic profiles of pre-treatment biopsies of breast tumours from 168 patients treated with chemotherapy with or without HER2 (encoded by *ERBB2*)-targeted therapy before surgery. Pathology end points (complete response or residual disease) at surgery³ were then correlated with multi-omic features in these diagnostic biopsies. Here we show that response to treatment is modulated by the pre-treated tumour ecosystem, and its multi-omics landscape can be integrated in predictive models using machine learning. The degree of residual disease following therapy is monotonically associated with pre-therapy features, including tumour mutational and copy number landscapes, tumour proliferation, immune infiltration and T cell dysfunction and exclusion. Combining these features into a multi-omic machine learning model predicted a pathological complete response in an external validation cohort (75 patients) with an area under the curve of 0.87. In conclusion, response to therapy is determined by the baseline characteristics of the totality of the tumour ecosystem captured through data integration and machine learning. This approach could be used to develop predictors for other cancers.

Neoadjuvant treatment, that is, systemic therapy (chemotherapy with or without targeted therapy) administered before surgery, is increasingly used in the management of breast cancer to improve rates of breast-conserving surgery and increase survival⁴. However, many patients do not have a good response^{5,6}. Features associated with response to neoadjuvant therapy have been derived from clinical⁷, molecular^{8,9} and digital pathology analysis^{10,11}. However, these studies have been frequently small, combined data from patients receiving different treatments and used single platform profiling that fails to capture the complexity of the tumour ecosystem. Unsurprisingly, physicians continue to select patients for neoadjuvant therapies using empirical clinical risk stratification¹².

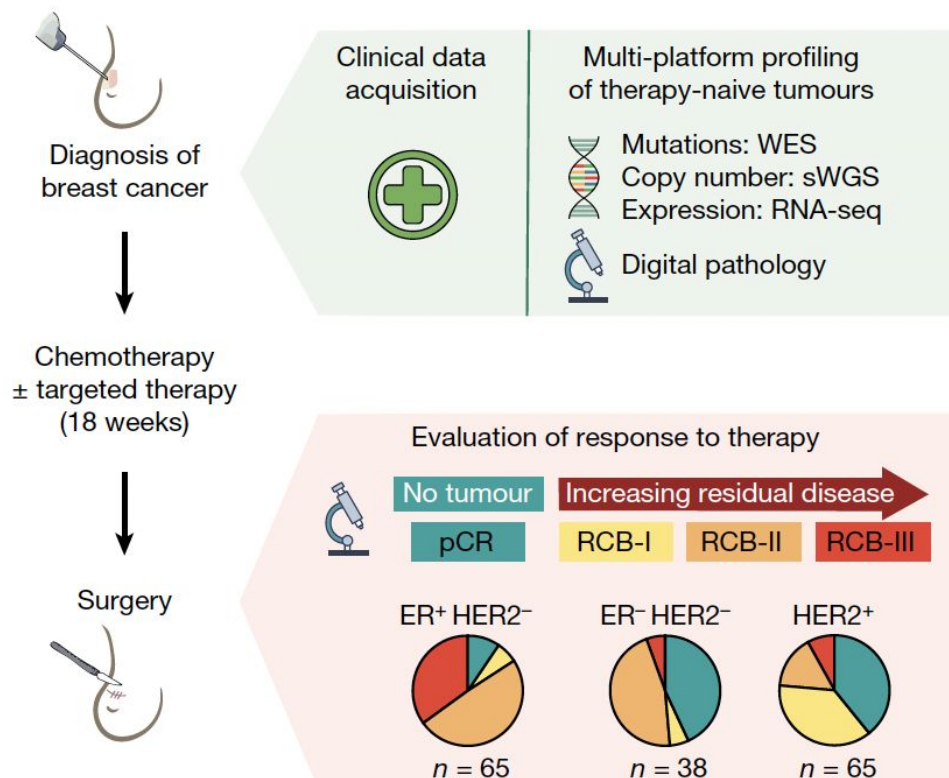
Tumour ecosystems are increasingly recognized as major determinants of treatment response² and we hypothesized that improved prediction models need to account for tumours as complex ecosystems, comprising communities of malignant clones within a microenvironment of stromal, vascular and immune cell types that are perturbed by therapy.

Here we characterized biological parameters extracted from a prospective neoadjuvant study that collected detailed pre-therapy tumour multi-omic data and associated these with eventual response. We found that malignant cell, immune activation and evasion features were associated with treatment response. These features, derived from clinicopathological variables, digital pathology and DNA and RNA sequencing, were used as input into an ensemble machine learning approach to generate predictive models. We validated the accuracy of the predictive models in independent, external cohorts and demonstrated that the best performers integrated clinicopathological and molecular data. The overall approach is widely applicable to other cancers and can be customized to include both fewer and newer features.

Multi-platform profiling of tumour biopsies

We prospectively enrolled 180 women with early and locally advanced breast cancer undergoing neoadjuvant treatment into a molecular

Clinical work flow and data acquisition



¹Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, UK. ²Department of Oncology, University of Cambridge, Cambridge, UK. ³CRUK Cambridge Centre, Cambridge Experimental Cancer Medicine Centre (ECMC) and NIHR Cambridge Biomedical Research Centre, University of Cambridge and Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. ⁴School of Clinical Medicine, University of Cambridge, Cambridge, UK. ⁵Institute of Astronomy, University of Cambridge, Cambridge, UK. ⁶Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia. ⁷Centre of Cancer Research and Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia. ⁸Warwick Clinical Trials Unit, University of Warwick, Coventry, UK. ⁹Edinburgh Cancer Research Centre, Western General Hospital, Edinburgh, UK. ¹⁰Q2 Laboratory Solutions, Livingston, UK. ¹¹Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ¹²Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. ¹³Strangeways Research Laboratory, University of Cambridge, Cambridge, UK. ¹⁴MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. *These authors contributed equally: Mireia Crispin-Ortuza, Suet-Feung Chin. ✉e-mail: carlos.caldas@cruc.ac.uk

So what ?!?



Triage-driven diagnosis of Barrett's esophagus for early detection of esophageal adenocarcinoma using deep learning

Marcel Gehrung^{1,2}, Mireia Crispin-Ortuzar¹, Adam G. Berman¹, Maria O'Donovan^{3,4},
Rebecca C. Fitzgerald^{3,5} and Florian Markowetz^{1,5}

Deep learning methods have been shown to achieve excellent performance on diagnostic tasks, but how to optimally combine them with expert knowledge and existing clinical decision pathways is still an open challenge. This question is particularly important for the early detection of cancer, where high-volume workflows may benefit from (semi-)automated analysis. Here we present a deep learning framework to analyze samples of the Cytosponge-TFF3 test, a minimally invasive alternative to endoscopy, for detecting Barrett's esophagus, which is the main precursor of esophageal adenocarcinoma. We trained and independently validated the framework on data from clinical trials, analyzing a combined total of 4,662 pathology slides from 2,331 patients. Our approach exploits decision patterns of gastrointestinal pathologists to define eight triage classes of varying priority for manual expert review. By substituting manual review with automated review in low-priority classes, we can reduce pathologist workload by 57% while matching the diagnostic performance of experienced pathologists.

Early detection of cancer often leads to better survival¹, because pre-malignant lesions and early-stage tumors can be more effectively treated². Most pre-malignant lesions amenable to early detection rely on targeted sampling and show only minor tissue changes on pathology assessment^{3,4}. In addition, pathology procedures often involve laborious and time-consuming steps that can lead to errors and adversely affect patient care⁵. Recent developments in artificial intelligence (AI) have achieved excellent performance on diagnostic tasks^{6–9}. However, understanding how these techniques can be integrated into clinical workflows most efficiently and assessing the actual benefits they bring remain a challenge. The design of a clinical decision support system needs to balance its performance against workload reduction and potential economic effect. Replacing pathologists entirely could lead to substantial workload reduction, but such an approach would be viable only if performance remains similar to that of human experts. Between a fully automated approach and the status quo of fully manual review lies a semi-automated approach that uses computational methods to triage patients and presents pathologists only with equivocal cases. A semi-automated approach will not reduce workload as much as a fully automated approach, but its performance benefits from existing expert knowledge and heuristics. Here we present such a semi-automated triage system using deep learning for the detection of Barrett's esophagus (BE), a precursor of esophageal adenocarcinoma (EAC).

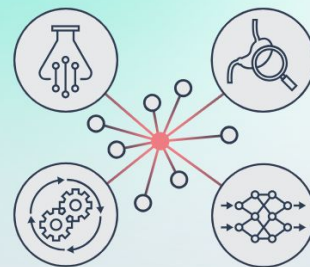
Esophageal cancer is the sixth most common cause of cancer-related deaths¹⁰. Patients usually present at an advanced stage with dysphagia and weight loss, and the 5-year overall survival of EAC—one of two pathological subtypes—is 13%¹¹. EAC can arise from a precursor lesion called BE¹², providing an effective starting point for early detection. BE occurs in patients with gastroesophageal reflux disease (GERD), a digestive disorder where acid and bile

from the stomach return into the esophagus, often leading to heartburn symptoms. In Western countries, 10–15% of the adult population are affected by GERD¹³ and, therefore, are at an increased risk of having BE. The pathognomonic feature of BE is intestinal metaplasia (IM), a process whereby the stratified squamous epithelial lining localized in the lower esophagus is replaced with columnar epithelium containing goblet cells^{14,15}. The conventional diagnosis of BE requires an invasive endoscopic procedure of the upper gastrointestinal tract. However, there is no routine endoscopic screening of the GERD population and, thus, the vast majority of patients with BE are undiagnosed¹⁶.

Cytosponge-TFF3 is a non-endoscopic, minimally invasive diagnostic test for BE^{17–19}. It is a cell collection device consisting of a compressed sponge on a string inside a soluble capsule. The capsule is swallowed by the patient and dissolves in the stomach, releasing the sponge. The expanded sponge is withdrawn by the attached string, sampling superficial epithelial cells from the top of the stomach, the esophagus and the oropharynx (Fig. 1a). Therefore, the cellular composition of the sample is dominated by squamous cells, gastric columnar epithelium and respiratory epithelium as well as any IM cells, if present. After removal, the device is placed in a container with preservative solution, and the sampled cells are processed, embedded in paraffin and stained with hematoxylin & eosin (H&E) as well as immunohistochemically stained with TFF3 (trefoil factor 3)²⁰. H&E stains allow the identification and quantification of cellular phenotypes, which is critical for quality control. TFF3 is over-expressed in mucin-producing goblet cells, which are a key feature of BE. TFF3 also functions as a protector of the mucosa from insults, stabilizes the mucus layer and promotes healing of the epithelium²¹. TFF3 stains allow the identification and quantification of goblet cells, which are indicative of IM. Therefore, TFF3 is the key diagnostic biomarker for BE²².

The Data-Driven Diagnostics Company

Our unique approach combines expertise in gastrointestinal oncology, digital pathology and AI-driven biomarker discovery.

[Learn more](#)[Latest news](#)

¹Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge, UK. ²The Alan Turing Institute, London, UK. ³MRC Cancer Unit, University of Cambridge, Cambridge, UK. ⁴Department of Pathology, Cambridge University Hospitals NHS Trust, Cambridge, UK. ⁵These authors contributed equally: Rebecca C. Fitzgerald, Florian Markowetz. ✉e-mail: RCF29@MRC-CU.cam.ac.uk; florian.markowetz@cruc.cam.ac.uk

The international journal of science / 30 June 2022

nature

ORDER OUT OF CHAOS

Compendium of
chromosomal instability
offers insight into genomic
changes in cancer

Coronavirus

Can antibody therapies
help to quell the next
pandemic?

Supply chains

The challenges
of corporate due
diligence

Radiation zone

A repeating fast radio
burst localized in a
dwarf galaxy

Photo: Mark W. 2022
Nature Publishing Group

taylor bio

COI: I am co-founder and director

PAN-CANCER PRECISION THERAPEUTICS

CONTACT

What is Predict?

Predict is an online tool that helps patients and clinicians see how different treatments for early invasive breast cancer might improve survival rates after surgery.

It is endorsed by the American Joint Committee on Cancer (AJCC).

➤ Start Predict

Change Language ▾

Did you mean to visit [Predict Prostate?](#)



Key decision: adjuvant chemotherapy – yes or no?

Reset

Predict is not designed to be used in all cases. [Click here for more details.](#)
If you are unsure of any inputs or outputs, click on the  buttons for more information.



DCIS or LCIS only?



Yes No

Age at diagnosis



- +

Age must be between 25 and 85

Post Menopausal?



Yes No Unknown

ER status



Positive Negative

HER2 status



Positive Negative Unknown

Ki-67 status



Positive Negative Unknown

Positive means more than 10%

Invasive tumour size (mm)

Tumour grade

Detected by

Positive nodes

Micrometastases only

Predict Breast

- No papers in glamour journals
- No major grants
- Philanthropy
- Hard uphill battle

Enabled when positive nodes is 1.



Treatment options and results will appear here when you have filled in all the information needed above.



“ It takes time! ”



“For 20 years, I thought my job was, as a **basic scientist**, to publish papers and **throw them over the wall** for someone else to apply.

I now realize there’s **no one** on the other side of the wall, just a huge pile of papers we’ve all thrown over.”

Attributed to
Duncan “Six Degrees” Watts





**“ I just follow the
standards in my field ”**

**“ This time
will be
different! ”**



Every translational paper
needs an

Implementation plan

Implementation plan

1. What decision is your model addressing?
2. How will the data be routinely collected?
3. What evidence will convince clinicians?
4. How will you navigate medical device regulations?

Implementation plan

When the paper is out,
what are your concrete
next steps?

Currently:
No incentives for
implementation

**Systemic bias against
implementation
harms patients**

Journals
publish

Universities
promote

Hospitals
innovate

Researchers
implement

Funders
fund

Patients
“So what?!”

Regulators
streamline



HOW DO YOU DO, FELLOW KIDS?



Journals
publish

Universities
promote

Hospitals
innovate

Researchers
implement

Funders
fund

Patients
“So what?!”

Regulators
streamline

Thank you!

Read all about it!

All models are wrong and yours are useless: making clinical prediction tools impactful for patients.

NPJ Precision Oncology 2024

Constructive feedback:

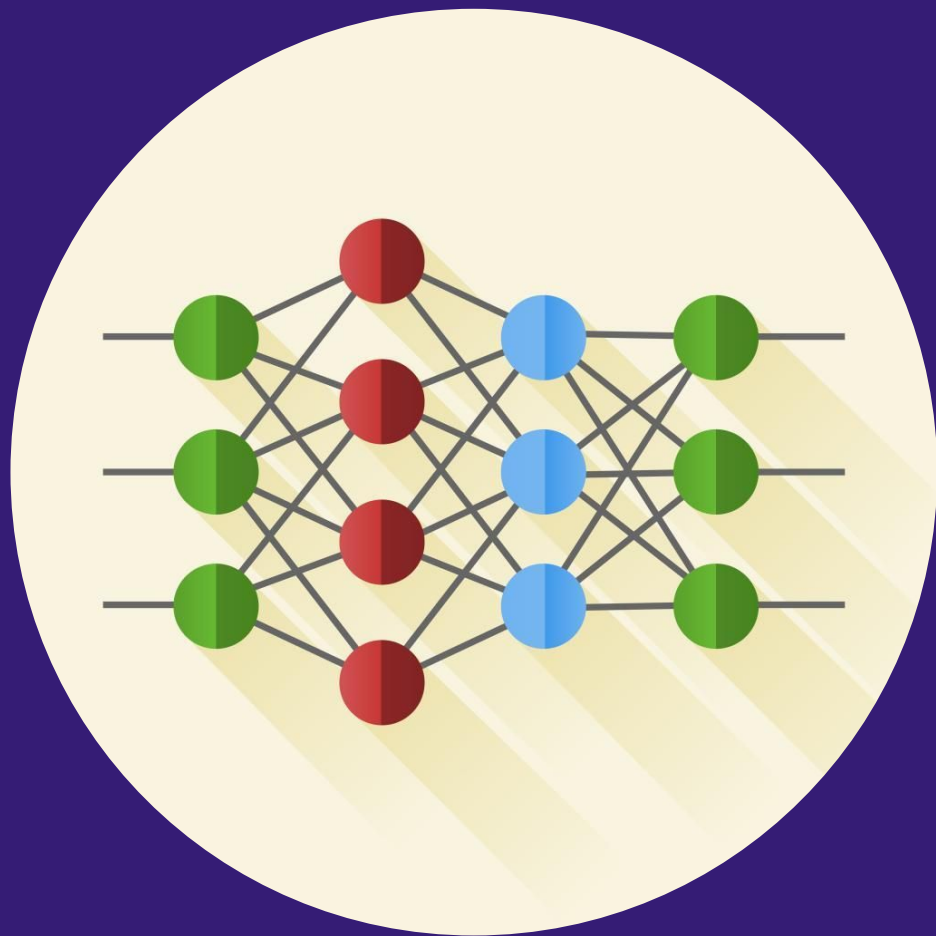
Paul Pharoah, Sarah Burge,
Louis Vermeulen, James Brenton,
Oscar Rueda, Mireia Crispin
Ortuzar, Jason Yip, Geoff Macintyre,
Stephen John Sammut, Raza Ali,
William Prew, Greta Markert



“Amateurs talk
strategy.”

Professionals talk
logistics.”

Military aphorism, source unknown



“Amateurs talk
algorithms.”

Professionals talk
implementation.”