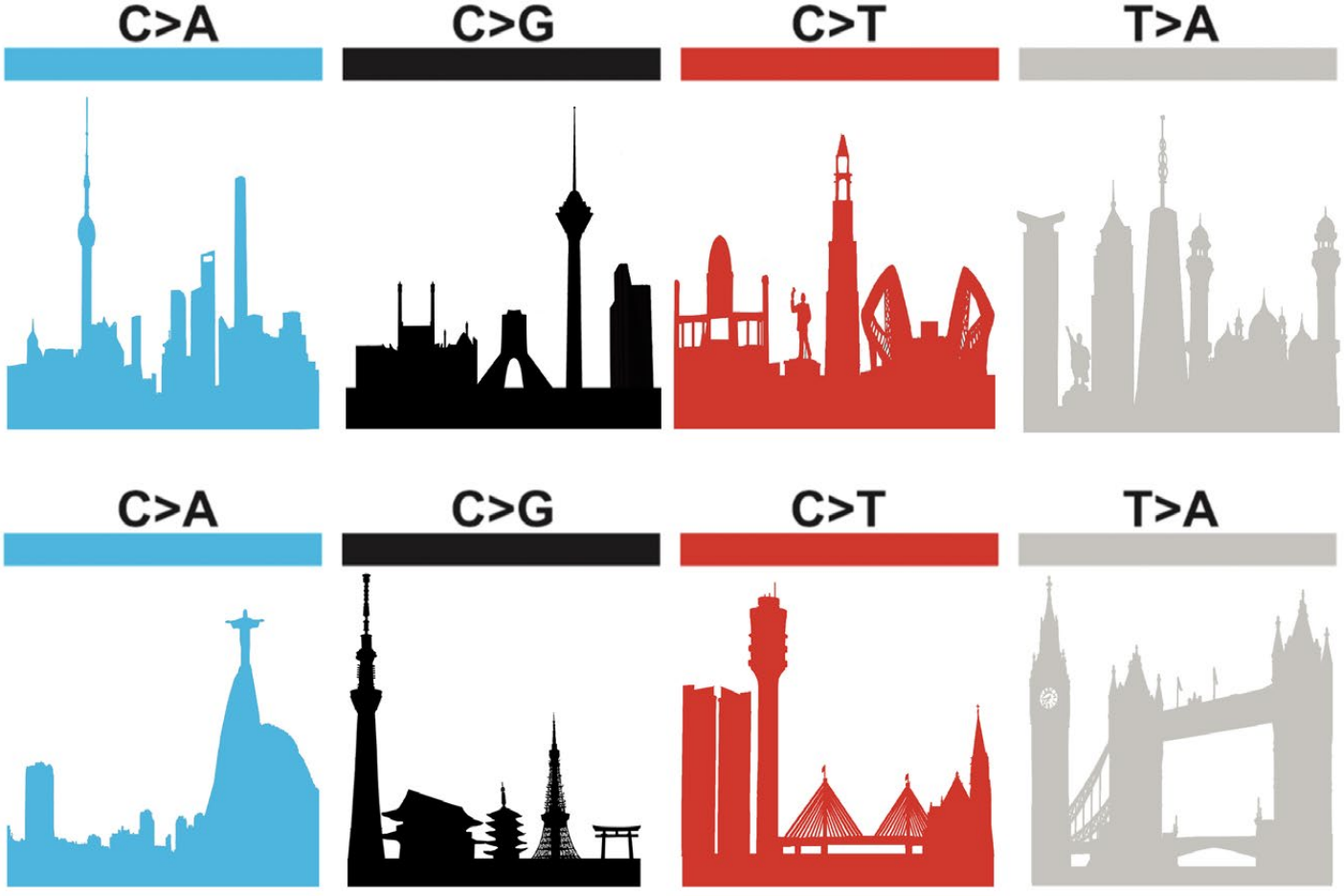
Volume 100 August 2021

nature genetics



Mutational Signatures in esophageal squamous cell carcinoma from eight countries

Courtesy of Joshua Atkins. Reproduced from Moody et al., Mutational signatures in esophageal squamous cell carcinoma from eight countries with varying incidence, *Nature Genetics* volume 53, issue 11, 2021, Springer Nature

Genomic Epidemiology Branch (GEM)

Branch head
Dr Paul Brennan

Deputy branch head
Dr James McKay

Scientists
Dr Behnoush Abedi-Ardekani
Dr Nicolas Alcala
Dr Shaymaa AlWaheidi
Dr Ievgeniia Chicherova
Dr Ana Carolina De Carvalho Peters
Dr Lynnette Fernandez-Cuesta
Dr Aida Ferreiro-Iglesias
Dr Matthieu Foll
Dr Mattias Johansson
Dr Florence Le Calvez-Kelm
Dr Sandra Perdomo Velasquez
Dr Hilary Robbins
Dr Mehrnaz Shamalnasab
Dr Mahdi Sheikh
Dr Shama Virani

Secretariat
Ms Juliette Prazak
Ms Isabelle Rondy
Ms Andreea Spanu-Bermond

Research assistants
Ms Karine Alcala
Mr Thomas Cattiaux
Ms Valérie Gaborieau
Ms Hélène Renard (until March 2023)
Dr Sergey Senkin
Dr Catherine Voegele

Project assistants
Ms Natalia Alves de Oliveira Vaz
Ms Laurène Bouvard (until November 2022)
Ms Sandra Moreno Ayala

Laboratory technicians
Ms Amélie Chabrier
Ms Priscilia Chopard
Ms Nathalie Forey

Postdoctoral fellows
Dr Joshua Atkins (until July 2022)
Dr Ricardo Cortez Cardoso Penha
Dr Claudia Coscia-Requena (until April 2023)
Dr Allison Domingues
Dr Wellington Dos Santos
Dr Rafii Fadoua (until January 2022)
Dr Xiaoshuang Feng
Dr Ryan Langdon
Dr Daniela Mariosa (until September 2023)
Dr Emilie Mathian
Dr Michael Olanipekun
Dr Justina Onwuka
Dr Han La Park
Dr Apiwat Sangphukieo
Dr Sergey Senkin (until March 2022)
Dr Alexandra Sexton-Oates
Dr Laura Torrens Fontanals

Students
Mr Sébastien Calvet (until July 2022)
Dr Gabrielle Drevet
Ms Elmira Ebrahimi
Ms Lipika Lipika (until November 2023)
Ms Laurane Mangé
Ms Fannie Martin (until August 2023)
Mr Simon Niceron (until August 2022)
Mr Eric Rucogoza (until August 2023)
Ms Hana Zahed

Visiting scientists
Dr Anvari Seyd Omid
Dr Patrice Avogbe
Dr Giovanni Centonze (until July 2022)
Dr Eleonora Lauricella (until October 2023)
Dr Delfin Lovelina Francis (until November 2023)
Dr Maike Morrison (until April 2023)
Dr Dariush Nasrollahzadeh Nesheli (until October 2022)
Dr Saeed Nemati
Dr Arash Nikmanesh (until October 2023)
Dr Simon Pahnke
Ms Fatemeh Shafighian (until October 2023)
Dr Jifang Zhou

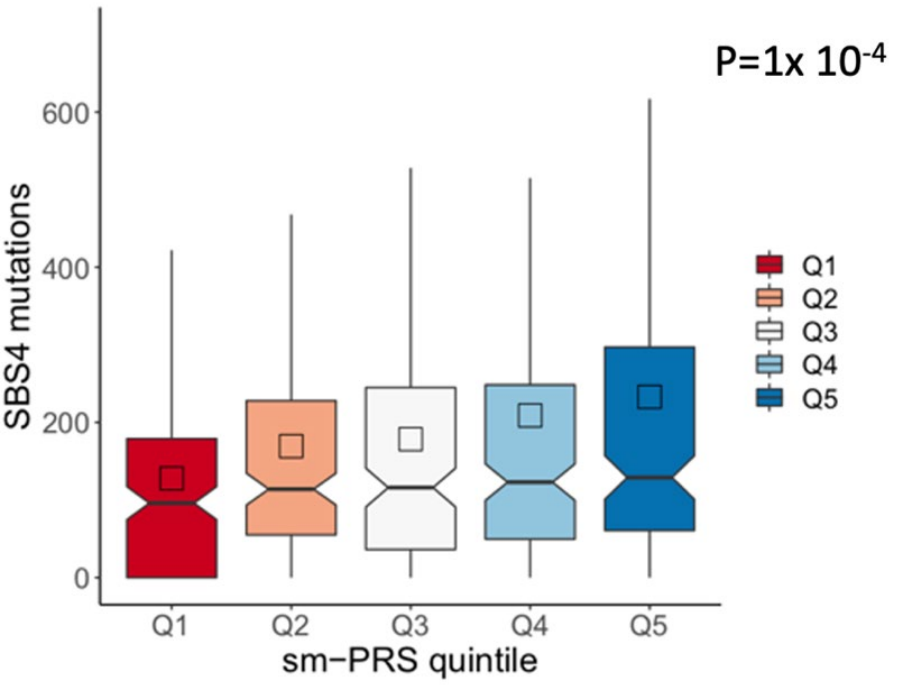
The over-arching goals of the Genomic Epidemiology Branch (GEM) are to further the understanding of cancer prevention and early detection using a combination of genomic and traditional epidemiology methods. This is done by bringing together six broad areas of work, as described here.

Area 1: Understanding the genetic susceptibility to cancer

GEM has continued to explore how genetic variation influences cancer susceptibility, with a focus on large international consortia to assemble expanded genetic data sets of lung cancer, head and neck cancer, renal cancer, and lymphomas. GEM's genetic studies are now on the order of 70 000 for lung cancer, close to 60 000 for lymphomas, 15 000 for head and neck cancer, and 30 000 for kidney cancer. GEM is now working with genotyping laboratories to undertake the genotyping and quality control analysis.

Through this continued expansion, GEM has identified novel susceptibility loci, implicating genetic loci containing genes such as *CHRNA4, CHRNB2, DBH, POLI, CHEK1, ERCC2, CYP1A1,* and *HLA*, and further implicated genes related to additive behaviour, DNA repair, telomere length, metabolic processes, and immune response in the carcinogenic process. GEM has continued to explore how germline genetic variants influence cancer susceptibility. For example, GEM has combined its germline analysis with its genomic analysis of the somatic material to demonstrate that the genetic variants related to aspects of nicotine addiction are also associated with the presence of mutational signatures associated with tobacco exposure in the patient's tumour (Figure 1). This appears consistent with the notion that genetic variants influence individuals' smoking behaviour, which in turn influences the degree of carcinogenic exposure in their lung tissue and, consequently, their somatic mutation burden (Gabriel et al., 2022).

Figure 1. GEM has combined its germline analysis with its genomic analysis of the somatic material to demonstrate that the genetic variants related to aspects of nicotine addiction are also associated with the presence of mutational signatures associated with tobacco exposure in the patient's tumour. sm-PRS, polygenic risk score for smoking; Q, quintile. Reproduced from Gabriel et al. (2022). © Gabriel et al., 2022. Published by Oxford University Press.



GEM used a similar approach to explore the influence of telomere length on susceptibility and lung adenocarcinoma tumour expression profiles (Cortez Cardoso Penha et al., 2023) and to implicate mutations in *BRCA2* in susceptibility to oesophageal squamous cell carcinomas. GEM also made important contributions to facilitate the efforts of these consortia, hosting an online conference during the travel restrictions imposed during the COVID-19 pandemic, and more recently hosting in-person meetings at IARC that also include online participation through hybrid meeting formats using the up-to-date facilities in the new IARC building. GEM continues to embrace technological advances, by developing consortia frameworks that enable data sharing of the consortia resources in a safe and efficient manner while supporting researchers from around the world (Figure 2).

Figure 2. The annual meeting of the International Lymphoma Epidemiology Consortium (InterLymph), at IARC in France in June 2023. © IARC.

Area 2: Identifying novel causes of cancer through genomics studies

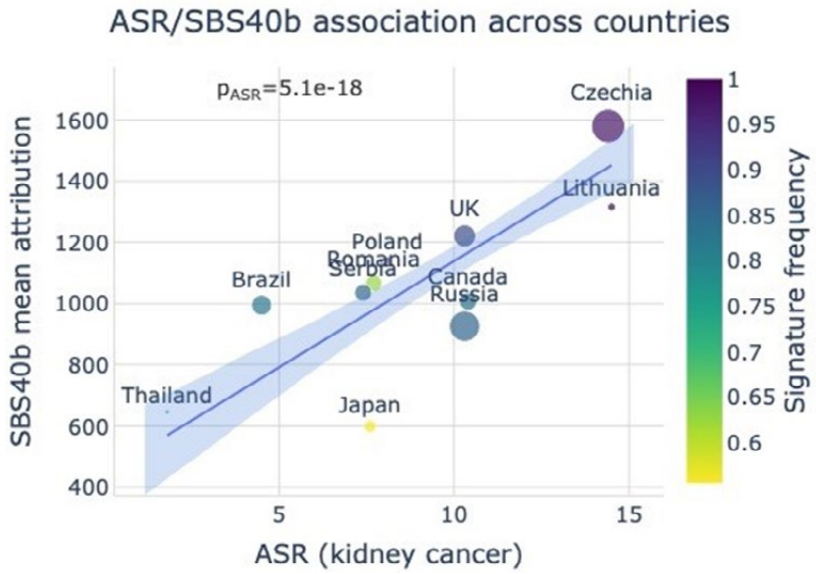
Mutographs

Mutographs is a Cancer Grand Challenges project that aims to understand the causes of five different cancer types across five continents by generating mutational signature profiles. The initial recruitment of about 6000 cases has been completed, and samples from 4000 cases have been successfully processed at IARC and sent to the Wellcome Sanger Institute (United Kingdom) for whole-genome sequencing. Genomic, exposure, and clinical data will be publicly available through the International Cancer Genome Consortium Accelerating Research in Genomic Oncology (ICGC ARGO) platform. The analysis of 552 cases of oesophageal cancer from eight countries with varying incidence was reported in 2021 and illustrated the importance of non-mutagenic causes of oesophageal cancer in high-incidence regions. Analysis of about 1000 kidney cancers across 11 countries has been completed, and the results are shedding light on the contribution of environmental causes to the high risk of kidney cancer in central Europe. In particular, the results highlighted a novel signature (SBS40b) that correlates strongly with incidence of kidney cancer (Figure 3). Understanding the cause of this signature could help to understand why incidence of kidney cancer is particularly high in central Europe.

Other results included the presence of a signature (SBS22) in south-eastern Europe that is linked to the mutagen aristolochic acid, and a separate signature (SBS12) that was present in Japan. These results raise the possibility that many millions of people in these regions are exposed to common mutagens.

The emerging results from the Mutographs project are changing the thinking about how environmental agents cause common cancers, and they have led to two large additional projects: (i) PROMINENT (see the text box) and (ii) DISCERN.

Figure 3. The results from the Mutographs project highlighted a novel signature (SBS40b) that correlates strongly with incidence of kidney cancer. ASR, age-standardized incidence rate. Reproduced from Senkin et al. (2023). Geographic variation of mutagenic exposures in kidney cancer genomes. medRxiv, 2023.06.20.23291538.



DISCERN

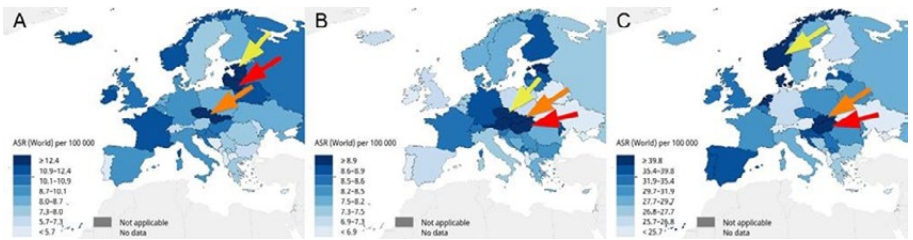
The Discovering the Causes of Three Poorly Understood Cancers in Europe (DISCERN) project was started in 2023 and is funded as part of the European Commission Cancer Mission initiative. The overall goal of DISCERN is to understand the causes of three poorly understood cancers in Europe – renal cancer, pancreatic cancer, and colorectal cancer – and to help explain the geographical distribution of these cancer types, including their high incidence in central and eastern Europe (Figure 4).

This will be achieved by combining large-scale European biorepositories comprising population-based cohorts and tumour case series with state-of-the-art exposomics and proteomics, as well as genomics technologies that analyse both normal and tumour tissue. DISCERN will provide the critical evidence base required to develop new prevention strategies for these cancer types in Europe. DISCERN builds on ongoing pan-European initiatives, including the European Human Exposome Network (EHEN), the Partnership for the Assessment of Risks from Chemicals (PARC), the Exposome-Powered Tools for Healthy Living in Urban Settings (EXPANSE) project, and the Mutographs project.

Area 3: Early cancer detection to reduce mortality and morbidity

Over the past few years, GEM has invested substantially in research aiming

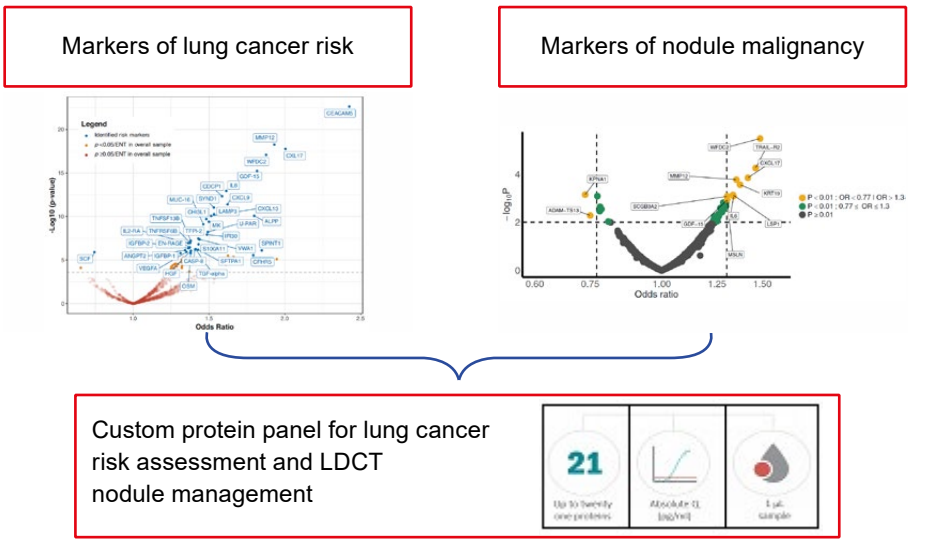
Figure 4. The overall goal of DISCERN is to understand the causes of three poorly understood cancers in Europe – (A) renal cancer, (B) pancreatic cancer, and (C) colorectal cancer – and to help explain the geographical distribution of these cancer types, including their high incidence in central and eastern Europe. From Ferlay et al. (2020). Global Cancer Observatory: Cancer Today. Lyon, France: IARC. Available from: https://gco.iarc.who.int/today.



to improve early detection of cancer. GEM's approach has focused on three research domains: (i) developing and evaluating risk models to inform the identification of individuals who may benefit from screening, (ii) identifying novel risk biomarkers that may improve existing risk models for use in determining screening eligibility, and (iii) developing minimally invasive early cancer biomarkers that may indicate an early undetected cancer.

In early detection of lung cancer, GEM research has been carried out in the context of screening with low-dose computed tomography (LDCT). LDCT screening has been shown to reduce lung cancer mortality in individuals who are at high risk, which is defined as having a history of heavy tobacco exposure. The current screening criteria only consider former and current smokers and typically involve a pack-year threshold (e.g. ≥ 20 or 30 pack-years of smoking exposure), time since quitting (e.g. 15 years), and an age range (e.g. 50–75 years). The use of risk models that provide absolute risk estimates based on individual risk profile data is also being evaluated in different settings. The choice of screening eligibility criteria will have a different impact depending on the setting, and GEM recently carried out an analysis that compared different strategies in Brazil. A major issue is that all commonly used screening eligibility criteria will leave many individuals who are destined to develop lung cancer ineligible for screening, and GEM has carried out extensive research aiming to develop biomarkers that can improve existing risk models. This research has been carried out using resources from the Lung Cancer Cohort Consortium (LC3), a major initiative coordinated by GEM since 2011 that involves 24 population cohorts from around the world with almost 3 million research participants followed up over time (Robbins et al., 2023). Based on these resources, GEM recently identified 36 robust protein biomarkers of lung cancer risk (Lung Cancer Cohort Consortium (LC3), 2023) that were able to substantially improve on traditional risk models (Feng et al., 2023a). Together with collaborators (Khodayari Moez et al., 2023), GEM is now developing a protein-based tool that can inform both the eligibility criteria for lung cancer screening and the management of nodules detected on LDCT screening (Figure 5).

Figure 5. A custom protein panel for assessment of lung cancer risk and management of nodules detected on low-dose computed tomography (LDCT) screening. (left) Reproduced from Lung Cancer Cohort Consortium (LC3) (2023). © Springer Nature. (right) Reproduced from Khodayari Moez et al. (2023), by permission of Oxford University Press.



For several years, GEM has organized research aiming to develop early biomarkers of human papillomavirus (HPV)-associated cancers. A seminal observation was a study in 2013 that observed frequent blood positivity for antibodies against the HPV16 E6 oncoprotein, several years before diagnosis of oropharyngeal cancer. The study also determined that this biomarker is rarely seen in healthy controls, opening up the possibility of using it as an early detection tool for HPV-associated cancer. This work stimulated the initiation of the HPV Cancer Cohort Consortium (HPVC3), which involves many population cohorts from around the world. One important question was to quantify the risk of oropharyngeal cancer that an individual would have after a positive HPV16 E6 blood test; a GEM study based on HPVC3 estimated the 10-year risk of oropharyngeal cancer in an HPV16 E6-seropositive individual at age 60 years to be 27.1% in men and 5.5% in women (Robbins et al., 2022a). This high level of risk may warrant periodic, minimally invasive surveillance after a positive HPV16 E6 serology test, particularly for men in high-incidence regions. However, an appropriate clinical protocol for surveillance remains to be established.

Bladder cancer is the 10th most common cancer type worldwide, and no urinary test has demonstrated sufficient performance to be useful for early detection purposes. Somatic mutations in the promoter of the telomerase reverse transcriptase (*TERT*) gene are common in urothelial cancer, and previous GEM research has demonstrated that it is possible to detect such mutations (TERTpm) in urine. Therefore, GEM scientists have developed a sensitive assay (uTERTpm) based on droplet digital PCR (ddPCR) with the view to use it as a non-invasive biomarker for early detection and monitoring of bladder cancer. The protocol for this ddPCR assay was recently published with step-by-step instructions for use in *TERT* mutation screening, including recommendations for sample preparation (Zvereva et al., 2023). GEM recently evaluated the ddPCR-based uTERTpm assay in a high-risk population in Kerman Province in the Islamic Republic of Iran, where bladder cancer is the most common cancer type in men (Pakmanesh et al., 2022). The uTERTpm assay detected 100% of primary bladder cancers, with a low false-positivity rate (12%) based on control subjects. The test was less sensitive (50%) for recurrent bladder cancer. Overall, this study shows promise for using the ddPCR uTERTpm assay as a non-invasive urinary marker of bladder cancer.

Area 4: Building global capacity for cancer science

GEM has made significant strides in promoting international collaboration in cancer research by addressing key challenges posed by data sharing and protection laws, such as the General Data Protection Regulation (GDPR). With the prime objective of ensuring enhanced access to harmonized genetic and epidemiological data for cancer studies, GEM supported the successful launch of the IARC Scientific IT platform in close collaboration with the Information Technology Services (ITS) team and the DAF Office. This centralized platform securely manages and stores data and enables remote data access without the need to transfer individual-level data. This approach not only streamlines data sharing but also aligns with the stringent data protection standards set by international laws.

Crucially, the project established an efficient administrative framework to manage data access requests. It leveraged existing consortium protocols and introduced a Data Use Agreement, thereby simplifying data sharing processes. Critical consortium studies have been integrated into the platform, including the European Prospective Investigation into Cancer and Nutrition (EPIC), the Lung Cancer Cohort Consortium (LC3), and the International Lymphoma Epidemiology Consortium (InterLymph). The success of this endeavour will undoubtedly fast-track global cancer research collaborations in the future and has also demonstrated a potential model for other scientific research networks.

The importance of data integration and sharing is exacerbated for the study of rare cancers. The ground-breaking efforts of the Rare Cancers Genomics Team (RCG) with the MESOMICS project underscore this principle. By meticulously integrating the largest whole-genome sequencing data set for malignant pleural mesothelioma with previous multi-omics studies (Mangiante et al., 2023), RCG has created an unparalleled resource. This endeavour does not stop at data sharing; by making the data accessible via the TumorMap web portal, RCG ensures interactive visualization and hypothesis generation without the need for intricate computational expertise (Di Genova et al., 2022). By democratizing access to high-quality data sets, ensuring reproducibility through shared bioinformatics pipelines (https://github.com/IARCbioinfo), and providing intuitive visualization tools, RCG is setting new benchmarks in advancing the collective knowledge of rare cancers.

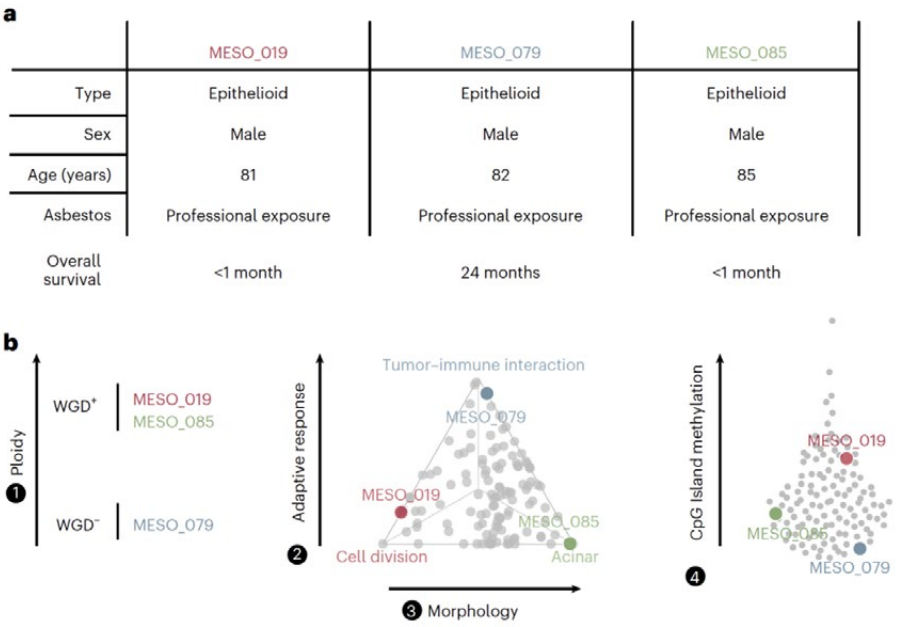
Large-scale biorepositories and databases are essential to generate equitable, effective, and sustainable advances in cancer prevention, early detection, and surveillance. The Mutographs project has created a large genomic data set and biorepository of more than 7800 cancer cases from 30 countries across five continents with extensive demographic, lifestyle, environmental, and clinical information. This collection has resulted in more than 85 000 biological samples currently stored in the IARC Biobank. Whole-genome sequencing data are being generated for nearly 4400 cancer cases.

Area 5: Somatic cancer genomics

The Rare Cancers Genomics Team (RCG) aims at the molecular characterization of rare cancers (https://rarecancersgenomics.com/; https://www.iarc.who.int/teams-rcg/), including malignant pleural mesothelioma (MESOMICS) and lung neuroendocrine neoplasms (lungNENomics). For MESOMICS, RCG has lifted the curtain on molecular differences between malignant pleural mesotheliomas (Figure 6) through the identification of molecular axes and specialized tumour profiles driving the intertumour heterogeneity (Mangiante et al., 2023). RCG has also generated a molecular phenotypic map of this disease (Di Genova et al., 2022). For lungNENomics, RCG has worked on developing a new anomaly-detection deep-learning algorithm, HaloAE, to identify patterns in images that could help to discriminate regions for tumour proliferation or aggressiveness.

RCG has also contributed to review the current biological and clinical knowledge on lung neuroendocrine neoplasms (Fernandez-Cuesta et al., 2023), to unveil that changes in *OTP* gene DNA methylation are responsible for its differential expression in lung neuroendocrine tumours (Moonen et al., 2022), and to better understand the mechanisms

Figure 6. The utility of a four-criteria classification of mesothelioma. (a) Three patients with mesothelioma (identified as MESO_019, MESO_079, and MESO_085) had similar clinical characteristics yet different outcomes. (b) The three patients have vastly different tumour profiles based on the four-criteria classification. Arrows are directed from low to high values for each criterion (e.g. from ploidy of 1 to ploidy of 4), and grey dots represent mesothelioma tumours. WGD, whole-genome doubling (ploidy > 2). Reproduced from Mangiante et al. (2023). © Mangiante et al., 2023. Published by Springer Nature.



behind the transformation of epidermal growth factor receptor *(EGFR)*-mutant lung adenocarcinomas into small-cell lung cancers (Mc Leer et al., 2022). RCG's contribution has also recently expanded to developing mathematical models for cancer evolution (Alcala and Rosenberg, 2022; Morrison et al., 2022).

RCG's projects have a strong computational biology component, particularly for the analysis and integration of –omics data (whole-genome and transcriptome sequencing, methylation arrays, single-cell and spatial transcriptomics data), the interpretation of histopathological images with deep-learning algorithms, and the modelling of cancer evolutionary processes. RCG actively shares these tools as open-source packages (https://github.com/IARCbioinfo), ultimately building capacity for cancer genomics (https://rarecancersgenomics.com/datasets/) and strongly contributing to research in Area 4.

Area 6: Understanding variations in cancer incidence and survival in diverse populations

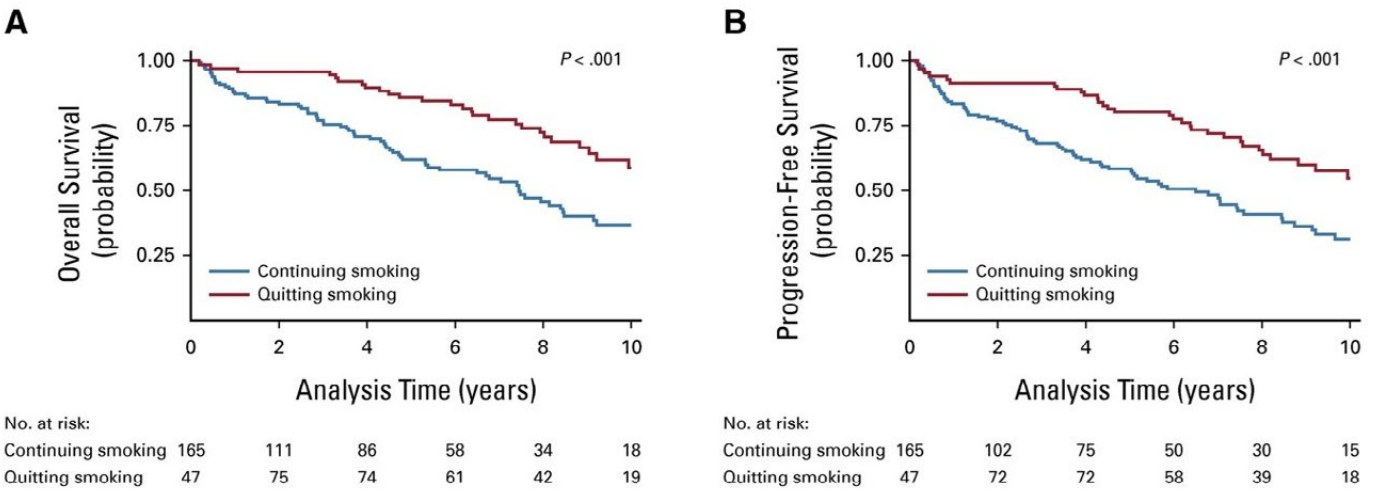
Two GEM initiatives are assessing variations in cancer incidence. The Opioid Cohort Consortium (OPICO) study is generating a single database with detailed individual-level information on opioid use, cancer incidence, and confounders, as well as exploring Mendelian randomization methods, to evaluate the association between regular opioid use and cancer incidence and mortality (Sheikh et al., 2023a). The Latin American Study of Hereditary Breast and Ovarian Cancer (LACAM) has started to describe how both germline pathogenic variants and modifiable lifestyle risk factors influence the risk of breast cancer and ovarian cancer (Díaz-Velásquez et al., 2023) in high-risk individuals in six countries in Latin America.

Additional studies between GEM and external collaborators show the effect of modifiable risk factors on cancer survival. GEM and collaborators from nine centres in the Russian Federation, Poland, Serbia, Czechia, and Romania concluded a survival analysis of 2052 patients with stage I–IIIA non-small cell lung cancer diagnosed and followed up in 2007–2016 (Sheikh et al., 2023b). The results revealed an overall 5-year survival rate of 50%. In patients from central and eastern Europe, higher risk of death and disease progression was observed in individuals with higher-stage tumours (hazard ratio [HR] for stage IIIA vs stage I, 5.54; 95% confidence interval [CI], 4.10–7.48), those who were current smokers (HR, 1.30; 95% CI, 1.04–1.62), and those who were alcohol drinkers (HR, 1.22; 95% CI, 1.03–1.44).

Another study, in collaboration with the N.N. Blokhin National Medical Research Center of Oncology (Russian Federation), recruited 212 patients with primary renal cell carcinoma in 2007–2016 and showed that quitting smoking after diagnosis of renal cell carcinoma may significantly improve survival (HR, 0.51; 95% CI, 0.31–0.85) and reduce the risk of disease progression (HR, 0.45; 95% CI, 0.29–0.71) and of cancer mortality (HR, 0.54; 95% CI, 0.31–0.93) in patients who smoke (Sheikh et al., 2023c) (Figure 7).

Two other projects are exploring variations in survival in diverse populations. The Translational Studies of Head and Neck Cancer in South America and Europe (HEADSpAcE) study is investigating the causes of late-stage diagnosis and the effect on survival of head and neck cancer cases in Europe and South America. The uTERTpm study is evaluating the use of non-invasive biomarkers for monitoring of bladder cancer recurrence.

Figure 7. Extended Kaplan–Meier curves illustrating the probability of (A) overall survival and (B) progression-free survival among smoker patients with renal cell carcinoma during the quitting smoking versus continuing smoking periods. Quitting smoking after diagnosis of renal cell carcinoma may significantly improve survival (hazard ratio [HR], 0.51; 95% confidence interval [CI], 0.31–0.85) and reduce the risk of disease progression (HR, 0.45; 95% CI, 0.29–0.71) and of cancer mortality (HR, 0.54; 95% CI, 0.31–0.93) in patients who smoke. Reproduced from Sheikh et al. (2023c). © 2023 by the American Society of Clinical Oncology.



PROMINENT: Discovering the molecular signatures of cancer PROMotion to INform prevENTion

In 2022, GEM received a Cancer Grand Challenges award from Cancer Research UK and the United States National Cancer Institute for the PROMINENT project. This project is co-led by GEM in collaboration with the Nutrition and Metabolism Branch (NME), along with 10 other partners in five countries (France, the United Kingdom, the USA, Spain, and Sweden). PROMINENT brings together a diverse team of experts who will use advanced high-throughput genomic, proteomic, and functional methods to uncover the main factors and processes that drive the transformation of normal cells into cancer cells.

This project builds on a unique collection of several thousand human samples – both normal and matched tumour samples – collected from more than 20 countries and stored in the IARC Biobank. These samples come from regions with varying levels of cancer risk, and detailed exposure information is available. Analysis of these samples, together with intervention studies in human populations, mouse models, and human organoids, will enable the development of a roadmap of tumour promotion, from individual normal cells with driver mutations all the way to full malignant progression.

PROMINENT team leads and members at the announcement of the awarded Cancer Grand Challenges teams at the 2022 Cancer Grand Challenges Summit in Washington DC, USA. © IARC.