Molecular Oncology

FEBS PRESS science publishing by scientists

REVIEW

Global mapping of cancers: The Cancer Genome Atlas and beyond

Carlo Ganini[1,2], Ivano Amelio[1], Riccardo Bertolo[1,3], Pierluigi Bove[1,3], Oreste Claudio Buonomo[1], Eleonora Candi[1,2], Chiara Cipriani[1,3], Nicola Di Daniele[1], Hartmut Juhl[4], Alessandro Mauriello[1], Carla Marani[1,3], John Marshall[5], Sonia Melino[1], Paolo Marchetti[6], Manuela Montanaro[1], Maria Emanuela Natale[1,3], Flavia Novelli[1], Giampiero Palmieri[1], Mauro Piacentini[1], Erino Angelo Rendina[6], Mario Roselli[1], Giuseppe Sica[1], Manfredi Tesauro[1], Valentina Rovella[1], Giuseppe Tisone[1], Yufang Shi[1,7,8], Ying Wang[7] and Gerry Melino[1]

1 Department of Experimental Medicine, Torvergata Oncoscience Research Centre of Excellence, TOR, University of Rome Tor Vergata, Italy
2 IDI-IRCCS, Rome, Italy
3 San Carlo di Nancy Hospital, Rome, Italy
4 Indivumed GmbH, Hamburg, Germany
5 Medstar Georgetown University Hospital, Georgetown University, Washington, DC, USA
6 Sant'Andrea Hospital, University of Rome Sapienza, Italy
7 CAS Key Laboratory of Tissue Microenvironment and Tumor, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China
8 The First Affiliated Hospital of Soochow University and State Key Laboratory of Radiation Medicine and Protection, Institutes for Translational Medicine, Soochow University, China

Keywords
artificial intelligence; cancer; molecular signature; omics; whole-genome sequencing

Correspondence
I. Amelio or G. Melino, Department of Experimental Medicine, Torvergata Oncoscience Research Centre of Excellence, TOR, University of Rome Tor Vergata, via Montpellier 1, Rome 00133, Italy
Tel: +39 06 72596976
E-mails: ivano.amelio@uniroma2.it (IA) or melino@uniroma2.it (GM)

(Received 1 March 2021, revised 4 May 2021, accepted 9 July 2021, available online 20 July 2021)

doi:10.1002/1878-0261.13056

Cancer genomes have been explored from the early 2000s through massive exome sequencing efforts, leading to the publication of The Cancer Genome Atlas in 2013. Sequencing techniques have been developed alongside this project and have allowed scientists to bypass the limitation of costs for whole-genome sequencing (WGS) of single specimens by developing more accurate and extensive cancer sequencing projects, such as deep sequencing of whole genomes and transcriptomic analysis. The Pan-Cancer Analysis of Whole Genomes recently published WGS data from more than 2600 human cancers together with almost 1200 related transcriptomes. The application of WGS on a large database allowed, for the first time in history, a global analysis of features such as molecular signatures, large structural variations and noncoding regions of the genome, as well as the evaluation of RNA alterations in the absence of underlying DNA mutations. The vast amount of data generated still needs to be thoroughly deciphered, and the advent of machine-learning approaches will be the next step towards the generation of personalized approaches for cancer medicine. The present manuscript wants to give a broad perspective on some of the biological evidence derived from the largest sequencing attempts on human cancers so far, discussing advantages and limitations of this approach and its power in the era of machine learning.

Abbreviations
AI, artificial intelligence; ARGO, Accelerating Research in Genomic Oncology; CNAs, copy number alterations; DNS, double-nucleotide substitutions; ICGC, International Cancer Genome Consortium; InDels, insertions/deletions; PCAWG, Pan-Cancer Analysis of Whole Genomes; SNS, single-nucleotide substitutions; SVs, structural variations; TCGA, The Cancer Genome Atlas; TSS, transcription starting site; WGD, whole-genome duplications; WGS, whole-genome sequencing.

1. Global genomic profiling: from the TCGA to the IGCG-ARGO projects

As a result of the continuous advances in DNA sequencing techniques and a massive reduction of the associated costs, scientists have been able to move from a classic mechanistic approach, in which a single gene or a set of a few genes were studied to elucidate their roles in cancer development, to global observational analyses. This step has led to the evaluation of the genomic alterations in cancers as a global network of molecular events, generating a huge amount of data from single cancer specimens [1]. Moreover, together with advances in genomics, many other '-omics' techniques have emerged and have been made available, allowing the generation of multidimensional datasets (genomes, transcriptomes, proteomes, phosphoproteomes, metabolomes) from individuals [2–10]. This global approach might be regarded as capturing any aspect of the biology of cancer, but also implies a shift in our ability to interpret data and to generalize evidence derived from a single patient to a multitude of individuals with the same disease.

At the early stage of the 'genomic era', the accomplishment of the Human Genome Project [11] in sequencing the entire human genome led to the idea that a similar attempt could be applied to cancer genomes. The first ambitious programme with this goal emerged in 2005 – The Cancer Genome Atlas (TCGA). This international multicentre genome sequencing effort took approximatively 8 years to reach completion (Fig. 1A) [12–14]. TCGA collected exome sequencing data of more than 11 000 cancer samples, characterizing 33 cancer types after an initial exploratory phase on three specific cancer entities (glioblastoma multiforme, lung and ovarian cancer [14–16]). The amount of data generated, in the order of millions of terabytes, clearly pointed out a crucial issue in the technological support required to process and handle this burden of data. Thus, cloud computing became an essential part of the process, together with the development of more sophisticated algorithms for data interpretation [11,17–19].

TCGA analysis of cancer samples mainly focussed on exome sequencing, but complementary approaches such as gene expression profiling and the analysis of copy number alteration (CNAs), single nucleotide polymorphisms, DNA methylation profiling and, to some extent, microRNAs expression were also applied. Only a marginal subset of TCGA cancer samples (< 10%) was used for whole-genome sequencing (WGS).

The huge amount of data generated allowed scientists to paint a broader picture of the mutational status of cancers and helped in both confirming some of the existing data or in generating new biological hypotheses; for example, on cancer-related genes [20–22], apoptotic regulators [23–25], protein stability [26–29] and redox regulators [30–35] or their structural motif [36,37]. TCGA, however, could not inform on large structural variations (SVs) as noncoding regions were not included in the sequencing.

In recent years, noncoding regions of the human genome expanded as a new focus in the scientific community, with the evidence that alteration of regulatory elements plays an important role in proliferative disorders [38–44]. Therefore, a further massive sequencing effort to understand cancer from a genomic perspective was launched by the International Cancer Genome Consortium (ICGC). The Pan-Cancer Analysis of Whole Genomes (PCAWG) started in 2012 with the aim of producing genomic sequences of whole genomes across 38 cancer types, bringing our knowledge on cancer alterations to a more advanced level and allowing the detection of new driver events and the evaluation of large SVs [45] that could not be described by TCGA dataset [46,47]. In February 2020, the PCAWG published a large part of the results obtained by comparing almost 2700 cancer genomes to their existing normal matching controls, together with almost 1200 transcriptomes (Fig. 1B) [48]. This huge effort allowed scientists to explore, for the first time and in a systematic way, noncoding regions of cancer genomes, and to postulate their role in cancer evolution. The development of mathematical modelling of cancer progression and algorithms also allowed the introduction of 'molecular timing' [49] to trace the temporal evolution of a single cancer from a single biopsy.

TCGA and PCWAG still represent the early stage of global mapping of cancer. The two projects still lack a comprehensive collection and analysis of the clinical data of the patients and do not cover proteomics, phosphoproteomics and metabolomics data. The ICGC is therefore now developing the ARGO (Accelerating Research in Genomic Oncology)–IGCG project, aimed at coupling more than 80 000 whole-cancer genomes to more accurate clinical data from patients (Fig. 1C) [50], but many more projects are starting all around the world [51].

2. Mutational signatures from WGS

The massive amount of data generated by the sequencing of whole-cancer genomes might be used for diverse purposes, mainly dependent on the mathematical approaches used to perform their analysis [52].

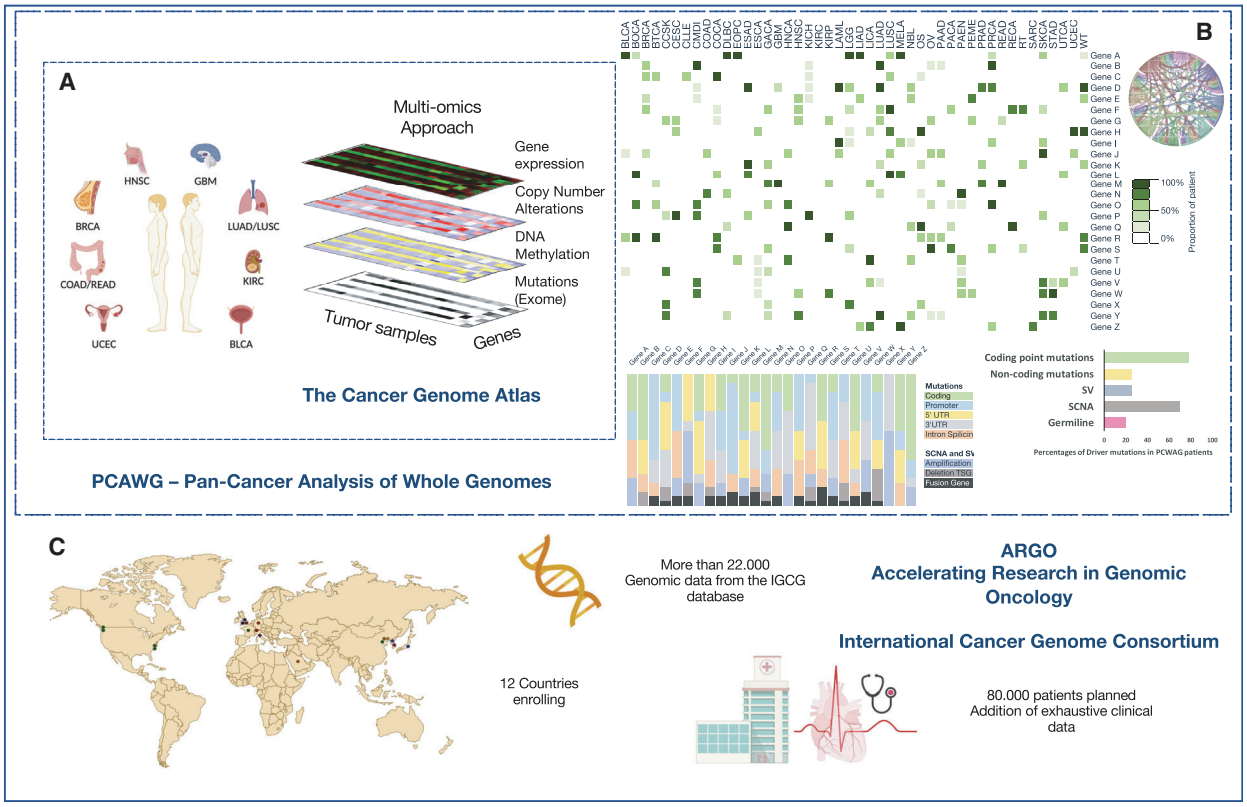

Fig. 1. Global cancer genomics approaches. (A) Multiomics approach in The Cancer Genome Atlas (TCGA), the first international project to catalogue the mutational landscape of human cancers. Data from more than 10 000 patients worldwide have been analysed in terms of gene expression, CNAs, DNA methylation and mutations in the coding regions of the genome, providing the mutational landscape of 12 common cancers. (B) The Pan-Cancer Analysis of Whole Genomes (PGAWG) project analysed more than 2600 whole-cancer genomes from the International Genome Cancer Consortium (IGCG), building upon previous data from TCGA. Cancer type alteration burden has been evaluated regarding mutations (single base substitutions, double base substitutions, small insertion and deletions), CNAs, SVs and RNA expression (heatmap); genomic alterations have been catalogued according to the site of occurrence, coding region of a gene, regulatory regions as promoter, 5′ or 3′ untranslated regions (5′UTR and 3′UTR) or intron splicing variants, or for CNAs and SVs, providing a specific global profile for each gene alteration; among each class of alterations, driver mutations have been recognized and coding point mutations, together with somatic CNAs (SCNAs), represent the highest number of driver events in cancers (bar chart) [80]. (C) ARGO, Accelerating Research in Genomic Oncology, is the ongoing phase of the global-scale omics approach of the ICGC, aimed at collecting omics and clinical data from more than 80 000 patients, with the goal to address key biological and clinical questions for each cancer type. This would allow the development of personalized medicine approaches for each cancer patient.

A common approach is to describe mutational signatures for specific cancer types. Mutational or structural variations are grouped based on their nature [single- or double-nucleotide substitutions (SNS or DNS), small insertions or deletions (InDels)] [53] and then analysed through cancer types to recognize precise patterns of variations among all the mutations occurring in each cancer sample compared to what happens in the general population. This kind of analysis has been adopted in the context of all the global genomic approaches developed so far (from the TCGA to the PCAWG), as well as in more limited analyses [54–60].

The most significant attempt applying this approach used data obtained from almost 24 000 cancer samples from different global genomic cancer databases, comprising 2600 samples from the PCAWG. This resulted in the identification of 67 mutational signatures, among which 49 were considered of biological significance (Fig. 2A,B). Many of those could be associated with a specific mechanism through which the tumours arise or progress, as in the case of signatures associated with defective DNA repair [61,62] or could be linked to clinical aspects and/or therapies, as in the case of signatures associated with platinum-based compounds [63] or ultraviolet exposure [64].

The comparative algorithms developed to detect mutational signatures were associated with different levels: SNS-based signatures associated with tobacco habit correlated in a statistically significant way to the tobacco-associated DNSs or the corresponding InDel signatures, somehow auto-validating the system [54,61].

The constant increase of cancer genomic data has also allowed us to refine already known signature, to better separate overlapping ones or to dissect some of them into sub-signatures [65]. Moreover, mutational signatures could also be associated with endogenous or exogenous exposure to biologically relevant substances linked to cancer development or to known pathogenic processes [66,67].

With notable exceptions (lung and colorectal cancer), the number of DNS signatures correlated with the number of the SNS. The number of signatures attributable to a cancer type also seems to correlate with the age of diagnosis, suggesting a possible temporal trend in the acquisition of these specific patterns of genomic alterations, introducing the concept that a mutational signature might be active already throughout the cell lineage of the tissue from which the cancer arises, from the fertilized egg onwards.

Although the validity of this approach can generate many biological hypotheses that might better unravel unexplored cancer mechanisms, some of the observations might be biased by the mathematical method applied to derive each signature and therefore need external validation through classical molecular biology approaches. The complexity of the cancer signalling during cancerogenesis and progression [68–71], as well as the mutational landscape of each signature, could be reproduced in *in vitro* or *in vivo* systems using CRISPR/Cas9 screening libraries (Fig. 2C) [72].

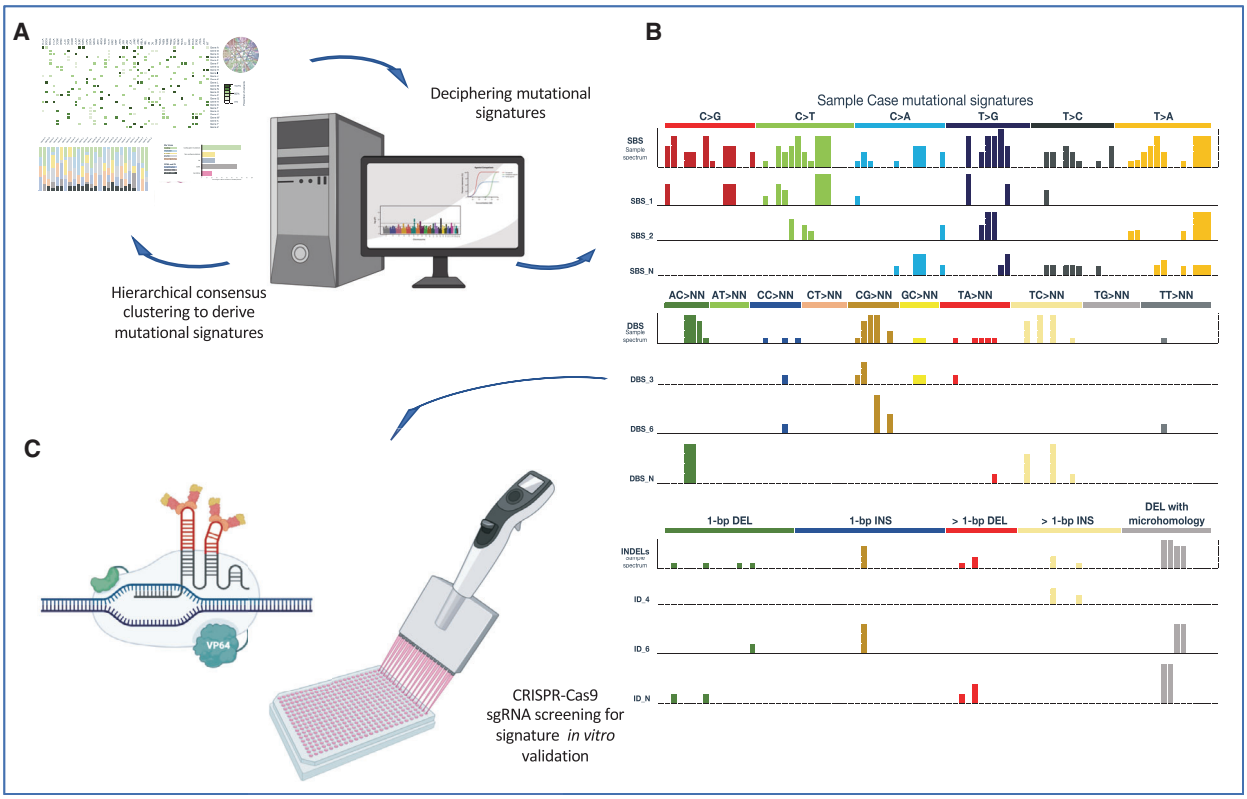

Fig. 2. Mutational signatures of cancers. (A) Global genomics data obtained in the Pan-Cancer Analysis of Whole Genomes (PCAWG) have been processed with fitting algorithm models to recognize mutational signatures in each cancer type. (B) The mutational profile of each cancer type can be dissected in multiple signatures according to the distribution of single base pair (bp) substitutions (SBS), double base substitutions (DBS) or small insertions/deletions (InDels); this approach allows researchers to correlate a specific signature with biological programme alterations in cancers (the APOBEC signature is an example) or with the clinical history of the patient (smoking-associated signatures) [54]. (C) Signatures can be further investigated in their role in cellular or animal models using CRISPR-Cas9 technology and single-guided-RNA screening platforms.

3. Application of WGS on large cancer specimen databases

The PCAWG project has been built upon the line traced by TCGA, with more than 2600 cancer samples collected and analysed for WGS, together with transcriptomics data and annotation of somatic small nucleotide variants, CNAs, small InDels, SVs, germline mutations, some retro-transposition events and mitochondrial DNA defects. Genomic alterations have been ranked based on their recurrence and on their functional consequences, finally developing a clustering methodology to discriminate between potential driver events [46]. The extension of the sequencing to intergenic regions allowed evaluation of the burden of putative driver mutations in noncoding regions: on the pan-cancer database, 13% of all mutations were represented by driver point-mutation events in an intergenic region, with 25% of all PCAWG cancers analysed bearing at least one, one-third of which occurred in the *TERT* promoter, confirming its role in cancer [73–79]. On the counterpart, 91% of all cancers harboured a somatic driver event in a coding region of a gene (Fig. 3).

The global genomic approach from the PCAWG also raised some interesting perspectives that might be explored in further investigations, since some *bone fide* genetic drivers have not been confirmed with this WGS approach. Not all the mutations in a known cancer-associated gene (considered a driver gene) are necessarily drivers if analysed through a 'ranking' approach, where mutations are ranked not only according to their frequencies but also by their putative functional consequence. This evaluation approach also led to the identification of 181 cancers (from the more than 2600 total analysed cases) without probable driver events (mostly in hepatocellular carcinomas, prostate cancer, medulloblastoma, pancreatic neuroendocrine tumours and renal chromophobe cancers). The heterogeneous nature of this group needs to be further analysed to understand whether possible common features are associated with the lack of putative driver events. This issue also pinpointed a provocative question: is no detection of driver mutations a biological phenomenon yet to be explained, or is it just a technical limitation that could be overcome with alternative approaches?

Besides the general overview on genetic aberrations and on their driver potential, specific cancer types, such as chromophobe renal cancer and neuroendocrine pancreatic tumours, surprisingly showed a higher number of driver mutations if compared to any other cancer type [80].

The WGS approach also explored some possible mechanistic links between catastrophic genomic alteration events, such as kataegis [81]. Kataegis is a mutational event in which a single strand of DNA is hypermutated with clusters of nucleotide substitutions. It was found in more than half of the cancer samples analysed by the PCAWG consortium and was more abundant in lung and bladder cancer, melanoma and sarcomas. Among the driver mutations frequently associated with kataegis foci, some genes are well-known cancer drivers, such as *CDKN1B*, *EGFR*, *FOXO1*, *MYC*, *SMAD4* and *TP53*, and frequently associated with the APOBEC gene signature [61,82–85]. Chromothripsis represents a mutational catastrophe in which hundreds of double-strand breaks occur in clusters on a few chromosomes [86–88]. This mutational event was found in almost 600 cases from the PCAWG, frequently in sarcomas, glioblastoma, lung squamous carcinoma melanoma and breast cancer and might explain some of the known pathogenetic features of these diseases [89–91]. Among the most associated driver genes involved in chromothripsis, *TP53* is prominently represented, reaching statistical significance on the pan-cancer analysis [10,85,92,93]. *MDM2* and *TERT* amplifications have been associated with chromothripsis in liposarcoma, while *EGFR*, *MDM2* and loss of *CDKN2A* in glioblastoma. As opposed to kataegis, chromothripsis showed correlation to clinical variables being more abundant in women or in late-onset prostate cancer patients as compared to early onset [80].

4. Large structural variations from WGS

Mutations that have been described in the context of cancer signatures, ranging from point mutations to small InDels, are unlikely to affect the genome structurally [94–96]. Anyway, it is well established that large SVs that endanger the whole structure of chromosomes are at the root of some of the hallmarks of cancers.

Structural variations, represented by large deletions, insertions, translocations or inversions, tend to occur in clusters. Those clusters can be spatially or temporally linked. Most of the time, clusters are both interlinked in space and time, suggesting a possible mechanistic liaison [45].

A WGS approach has the power to detect mutational signatures related to structural variants, and some of the analysis from the PCAWG consortium focussed on this point.

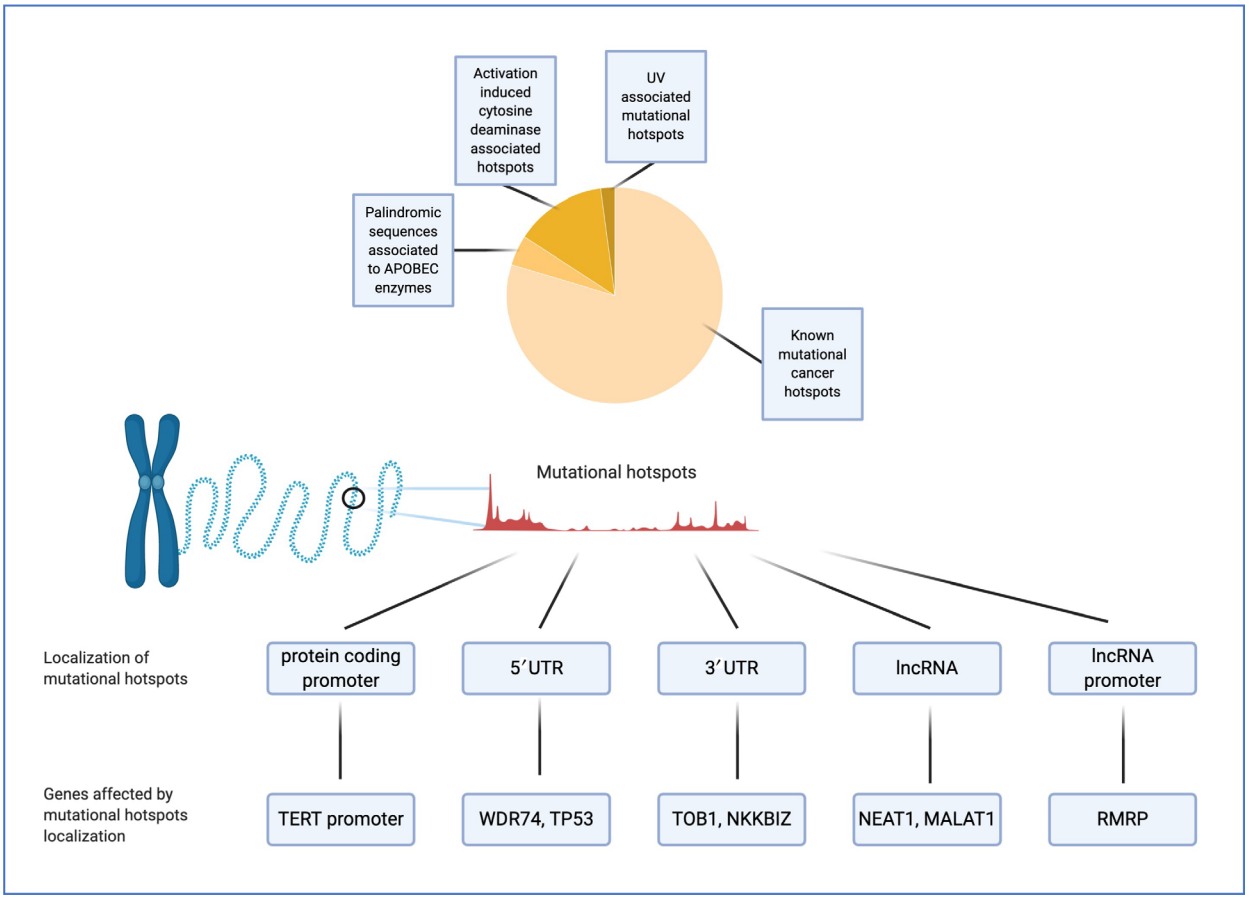

Fig. 3. Mutational hotspots of cancers in noncoding regions of the genome. Mutational hotspots in cancer are frequently localized in known mutated genes and can act as drivers. Their frequency in noncoding regions has been recently evaluated [48]. Apart from known hotspots in coding regions, 25% of cancers show clusters of mutations that are localized at the 5′UTR or 3′UTR of genes, as well as on long noncoding RNAs and on their promoters. These hotspots can also be linked to specific signatures, such as UV, activated induced cytidine deaminases and APOBEC enzymes activity [112].

To expand our knowledge on larger SVs, a new classification needs to consolidate classical categories used so far (translocations, inversions, deletions) with the newly identified ones based on the modality of occurrence (the 'cut and paste' or the 'copy and paste' approach).

A classification of all the possible SVs might be based on the possibility that the inserted segment returns, or not, to the original chromosome. SVs can be classified into chains if they do not return to the original chromosome or into bridges and cycles (leaving a gap or replicating numerous times) if they do return. The way SVs occur can be linked to specific regions of the genome and shed light on the mechanism through which they form: this is the case of the *TERT* regions, where SVs are almost inevitably represented by cycles of templated insertions. This alteration is also associated with SVs affecting tumour suppressor genes, as in the case of the *RB1* gene [45].

Among the classic categories, simple inversions have a relatively low frequency in the PCAWG database and are not associated with copy number gain. A comprehensive evaluation of the complexity of SVs has proved to be challenging; therefore, mathematical approaches like the ones used to highlight mutational signatures have also been used in the context of SVs [97]. At the level of SVs, as in the case of SNS, DNS and InDels signatures, there is a correlation with some described biological processes, as in the case of mutations in the *BRCA2* gene [98], which is associated with small deletions signatures as well as with chromoplexy. This might help in correlating the nature of SVs with a possible mechanism of insurgence and generate more biological hypotheses to explain, from a causative point of view, their insurgence.

5. Molecular timing of cancer evolution

Among the possible approaches to analyse global genomics data, an extremely promising one resides in applying bioinformatics algorithms to determine the timing of the evolution of a single cancer. This approach would grant valuable knowledge and allow the development of more accurate and precise early detection techniques [49].

Evaluating the number of allelic copies of a mutation of any type might help to discriminate between clonal drivers that are early or late in the evolution of a tumour, whereas a relative ratio among duplicated and nonduplicated mutations would allow determination of the 'molecular timing' of a mutation insurgence [99], and the data necessary to develop this kind of approach are obtainable, theoretically, from a single biopsy (Fig. 4A,B) [100], and analysed through a massive sequencing approach, as shown by the data from the PCAWG. Each cancer specimen can be analysed to detect the molecular timing of the mutations found in its genome, and all the information obtained can also be processed through artificial intelligence (AI) approaches to generalize the observations derived from single cases to the comprehensive cancer type to which they belong [99–101].

One of the first results derived from the PCAWG approach confirmed that many of the most common driver mutations are early clonal. The main example is p53 mutations [102–104] which are almost exclusively emerging at an early stage of cancer development. Moreover, more than half of all early clonal mutations occur in just nine genes, and these are rarely mutated in later stages or in the subclonal phase (Fig. 4C). This approach can confirm some of the well-known cancer progression models, as in the case of the *APC–KRAS–TP53* colon model [105,106].

Mutations are not the only features of cancer development that can be analysed thorough a timing approach. Signatures, which can also be detected from global genomic sequencing, can be assigned to a molecular time of development. In the PCAWG analysis, signatures associated with exogenous mutagens are invariably found in early development clonal stages, while other signatures tend to accumulate throughout the whole-cancer evolution, as in the case of the APOBEC signature [49].

The molecular timing analysis considers that the evolution process and the rate of mutations of cancer follow a nonlinear kinetics. This kind of approach allows us to develop a probable timeline of cancer evolution and might represent a good opportunity in cancer prevention, especially in the case of malignancies that are not preceded by a known premalignant lesion. In many cancer types, the early mutations seem to proceed tumour diagnosis by many years. A major limitation is that this analysis is based on point mutations, and therefore, it does not consider any other possible genetic aberrations that might represent *per se* fundamental driver events. Moreover, although the analysis of multiple cancer samples allows the development of this kind of temporal retrospective evolution, many single cases fall out of the prediction, underlining the fact that the nature of cancer is far from deterministic.

6. RNA dysregulation in cancer

While the TCGA database systematically collected cancer genomic data by sequencing platforms technologies [107], transcriptomic data have been less methodically collected and, in many cases, the corresponding genomes of the transcriptomes analysed have not been sequenced [108]. Alterations of transcriptomes have therefore been difficult to attribute either to intra-tumour or inter-tumour heterogeneity or to the underlying altered genomic landscape. RNA alterations can occur independently from DNA mutations, and therefore, an integrated RNA–DNA sequencing approach is mandatory.

One step forward was made in 2017 by the Genotype-Tissue Expression (GTEx) consortium, which analysed transcriptomes from 54 healthy tissues from more than 1000 donors, paired with the corresponding WGS [109–111].

The most recent attempts analysing transcriptomes together with WGS in cancer is again embedded in the PCAWG consortium. The WGS approach allowed researchers to link gene expression variations to transcriptionally inactive genomic sites, such as heterochromatin; more than 2500 cancer specimens were analysed, identifying mutational signatures associated with transcriptional alterations. More than 1100 genes have been associated with these signatures, with an increased number of mutations near the transcription starting site (TSS) of major promoters rather than minor or inactive ones. Mutations in the TSS of major promoters do not seem to significantly alter transcription, since it is more prominent in some cases, such as melanoma, but does not seem to have a role in colorectal cancer [112].

Among the promoters which are highly associated with transcriptional alterations, the *TERT* promoter seems to be the most involved in this kind of deregulation [109,113,114].

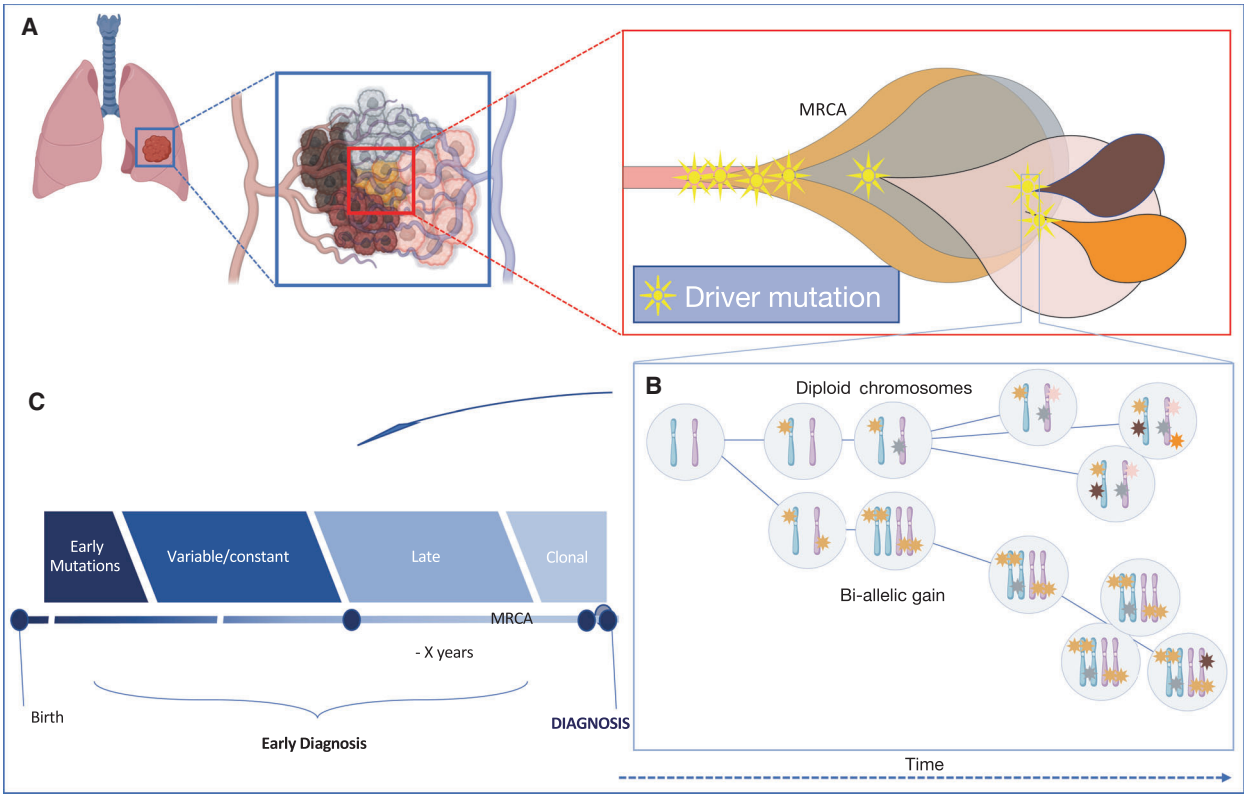

Fig. 4. Evolutionary history of cancers, molecular timing and early detection. (A) The mutational history of each cancer can be evaluated from a single biopsy by considering the evolution of tumour heterogeneity. (B) The clonal allelic status of point mutations can be used as a model to classify mutations as preferentially early, variable, constant, late or subclonal. The first two classes of mutations usually harbour driver mutations among many genes, whereas the late and the subclonal classes usually do not contain driver mutations. (C) The classification of mutations according to their type [driver, CNAs, mutational signatures (Sigs)] and their allelic burden allows the reconstruction of a timeline for the development of each tumour [49], potentially extending the time for an early diagnostic approach. MRCA, most recent common ancestor.

The exact role of RNA alterations has also been evaluated in the context of signalling pathways, more than focussing on a single gene. The NOTCH and the TGF-β pathways are largely affected by transcriptional alterations, more than other signalling pathways [112,115]. Moreover, *KRAS* exhibited more RNA alterations rather than DNA mutations in the context of various cancer types (not all). This finding might also have an impact on the prognostic role of *KRAS* in colorectal cancer and might be considered as a more precise biomarker to determine the fate of this group of patients [112,116].

Although cancer remains a disease governed primarily by DNA alterations, some driver events are directed by perturbations of RNA expression (which can also depend on noncoding RNAs) [117–119], which are not depending on underlying genomic abnormalities and gene expression alterations but have rather been shown to be far more associated with CNAs rather than on gene mutations. Anyhow, the PCAWG analysis showed that these cases are quite rare [112]. Altered expression of genes was found in more than 700 genes. For some of them (e.g. *TP53*), RNA alterations were associated with more abundant DNA alterations, whereas others, such as *GAS7*, behaved oppositely. In total, 87 cancer samples could not show any detectable DNA that could justify the RNA alteration observed.

7. Capturing genomic alterations during cancer evolution

The focus of the most recent global genomic approaches has been on primary cancer samples. A systematic approach involving WGS on metastatic cancer has been attempted by Priestley *et al.* [120]. WGS of more than 2500 metastatic cancers was analysed and matched with the corresponding genetic

background from patient-derived circulating mononucleated cells.

Among the most interesting highlights derived from this study, more than 80% of the samples harboured whole-genome duplications (WGDs). This finding contrasts highly with the setting of primary tumours from the PCAWG, where just 30% of the analysed samples showed this kind of genomic duplication [49]. Moreover, the report does not find a recurrent mutation in metastasis, somehow confirming that the metastatic process does not derive from a single driver event but is rather governed by a more pervasive programme.

This ambitious work misses information derived from a parallel WGS of the matching primary cancers. To overcome this limitation, the PCAWG database has been analysed in parallel and confirmed a high genomic concordance between primary and metastatic lesions, also showing that the most common mutation in primary cancers is the same found in metastasis, although at a higher prevalence [121] This finding, together with the evidence of more common WGDs, suggested that a hallmark of metastatic progression might be represented by genomic instability [122,123]. Conversely to previous results, the tumour heterogeneity of metastatic lesions seems to be less important than the one found in primary cancer; this might further corroborate the idea that a founding cancer cell could colonize a metastatic site and be predominant, but also warns about possible technical limitations derived from biopsy techniques.

8. Advantages and limitations of current global cancer genomics approaches and introduction of the executable cancer models

After almost two decades of efforts aimed at collecting large numbers of cancer specimens, generating terabytes of information, where have we arrived? A step towards personalized cancer medicine has undoubtedly been made, since global genomics, as well as global omics, approaches to cancer patients are available almost everywhere and quite accessible in terms of costs [124].

The generation of data has stressed the necessity to work in large international cooperative networks, including TCGA and PCAWG. The attention of the scientific community towards big data generation in cancer has means that the two cited sequencing programmes are not the only ones available: many others, also developed and run by private companies, are concurrently ongoing [125,126].

A global omics approach also has the prerogative to capture unforeseen cancer alterations that might fit in the context of repurposing and drug rediscovery, which might represent a valuable way in which to exploit pharmacological weapons that we already have but that have not been considered for a given disease.

Alongside the improvement of high-throughput sequencing techniques, AI has also shown impressive steps forward, with the improvement of both the hardware as well as the computational power of machines [127,128]. Together with the generation of big data from the -omics, AI can be used to develop algorithms to detect cancer signatures that might play a role as more accurate and multidimensional cancer biomarkers [2,129–133]. Moreover, multiomics signatures can be associated with specific mechanistic modalities of cancer development (e.g. in the case of signatures associated with exposure to exogenous carcinogens) and can suggest unexplored mechanisms that might cause the oncogenic transformation of cells, therefore impacting on cancer prevention, early detection or therapeutic decisions [121,134–141].

Some of the algorithms that have been developed can also allow the early diagnosis of cancer; for example, they provide the possibility to evaluate the 'molecular timing' of cancer development, allowing us to infer, from a single biopsy, the mutational evolution of cancer, determining a timeline of evolution for each tumour sample. This kind of approach has shown that, in most cancers, the early-stage mutations can precede cancer diagnosis by many years. The perfecting of this approach might have an impact on the detection of cancer at a very early phase, also in the context of those neoplasms that do not show a proper premalignant lesion [142–145].

Many limitations are present in this kind of approach. Undoubtedly, one biological limitation is that cancer heterogeneity is difficult to evaluate from single cancer specimens; moreover, cancer evolution can be inferred from an algorithm, but the dynamic nature of cancer cannot be considered from this single-biopsy-based approach. Therefore, the development of less-invasive procedures to generate -omics data are extremely valuable, and liquid biopsies might represent one of the possible new breakthroughs that could have an impact on cancer treatment [146–153].

An extra level of complexity derives from the evidence that, besides tumour heterogeneity, the tissues employed for omics analysis are mixtures of cancer cells in a complex tumour microenvironment. The complex network between cancer cells and the stroma (immune infiltrates, fibroblasts, endothelial cells) has taken the stage in cancer sciences during the last

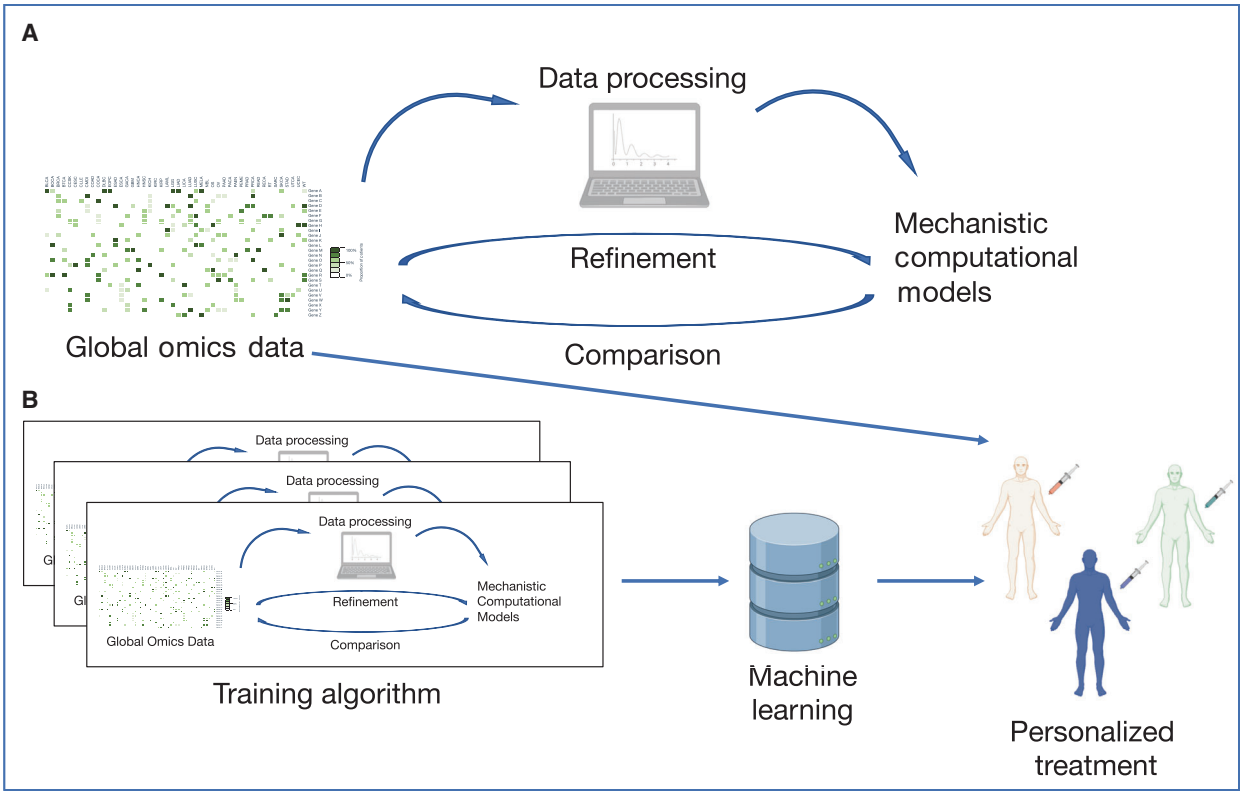

Fig. 5. Executable cancer models. (A) Experimental data from a global omics approach can be used as a matrix source for mechanistic computational models that can be continuously processed and refined using data from different cancer types and patients. This will provide data-based mechanistic hypotheses on each cancer sample. (B) The data obtained through a global omics approach can be further integrated in a machine-learning system, which is able to refine its ability to highlight mechanistic processes at the root of each cancer sample and can be further integrated with patient-derived omics and clinical data to develop more precise information of cancer stage and development, ultimately allowing precise personalized medicine interventions.

decade, especially focussing on the immune system liaisons with cancer, resulting in major therapeutical breakthrough, such as the use of checkpoint inhibitors [154–158]. From one side, prediction algorithms can describe the formation of tumour neoantigens and reconstruct their possible steric presentation by the human leukocyte antigen system of a patient. These *in silico* predicted neoantigens can then be used for the formulation of vaccines that are already tested in clinical trials in some cancer entities [159]. Furthermore, RNA-seq data can now be deconvoluted to understand the cell types present in a cancer specimen, to evaluate their enrichment, but also weigh the heterogeneity of B and T lymphocytes, thanks to the large-scale application of the sequencing techniques, which are now more available than before [160,161].

One other limitation to the current approaches is that the focus is on cancer genomics and transcriptomics. By now, the two largest programmes have only produced a limited amount of data from other -omics; data from cancer proteomics, phosphoproteomics, metabolomics and so on could be of greater value if they could be integrated with WGS.

Finally, among the most important limitations linked to past efforts on global genomics approaches is the poor amount of clinical data collected. This point is crucial to integrate this approach in the context of personalized medicine. The importance of clinical data in this setting is clear to the scientific community, and international programmes are already recruiting patients to further expand the -omics pool of data supported by a more accurate clinical description of the cases, such as in the GENIE (Genomics Evidence Neoplasia Information Exchange) programme [162] or the ICGC–ARGO [50].

9. Executable cancer models: successes and challenges

The large amount of genomic and transcriptomic data available so far contribute to our knowledge on cancer development and progression, but this information is

not yet ready to be translated into the clinic. The possibility to choose the right treatment for the right patient is still a goal to reach. Nowadays, it is surely right around the corner, but the huge amount of data obtained from a single patient need wise tools to be interpreted and embedded into the decision-making process of cancer treatment [163,164].

Artificial intelligence might just help with this issue. The possibility to develop executable cancer models would allow scientists to search among multiple datasets for the discovery of signatures at every level (genomic, transcriptomic, proteomic, clinical data) to detect key features of the biological behaviour of interest [20,69,165–167]. Biological systems can be treated by AI as networks of information that can be programmed, i.e., reconstructed as a matrix of data [168,169].

Once the network model is created, the AI continues perfecting it by embedding new data, constantly correcting the generating algorithm to fit the new data. Moreover, discrepancies between a dataset confronted with the executable model might suggest experiments that can be used to further ameliorate or refute the initial hypothesis sustained by the model.

Due to the extreme plasticity of this kind of approach and to its relative ease of use, AI can overcome one of the limitations that personalized medicine is going through, that is the inability to act according to the evolving patient's response in a timely manner. Nowadays, the maturity of the executable cancer models as such can be easily embedded in a dynamic scenario, such as the one represented by cancer evolution in response to therapies (Fig. 5).

Altogether, the development of AI and the machine-learning approach should be considered one of the most precious tools for the management, analysis and clinical translation of the endless data obtained from the application of high-yield sequencing methodologies and proteomics to cancer science and will represent the next major advance in the field, allowing personalized medicine to become an everyday reality.

Acknowledgements

We apologize for those whose contributions could not be cited due to space constraints. Figure parts, as well as whole figures, have been generated using Biorender.com. This work has been supported by the Associazione Italiana per la Ricerca contro il Cancro (AIRC) to GM (IG#20473; 2018–2022), to IA (AIRC Start-Up ID 23219; 2020–2024), to EC (IG#22206; 2019–2023), Ministry of Health & MAECI Italy-China Science and Technology Cooperation (#PGR00961) to EC, GM and YW, to Ministry of Health and IDI-IRCCS (RF2019.12368888).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

CG, IA and GM wrote the manuscript. CG prepared the figures. All the other indicated authors (RB, PB, OCB, EC, CC, NDD, HJ, AM, CM, JM, SM, PM, MM, MEN, FN, GP, MP, EAR, MR, GS, MT, VR, GT, YS, YW) made substantial contribution to the conception of the manuscript and critically revised it. All the Authors have approved this submitted version.

References

1 Ding L, Bailey MH, Porta-Pardo E, Thorsson V, Colaprico A, Bertrand D, Gibbs DL, Weerasinghe A, Huang K-L, Tokheim C *et al.* (2018) Perspective on oncogenic processes at the end of the beginning of cancer genomics. *Cell* **173**, 305–320.e10.

2 Mantini G, Pham TV, Piersma SR & Jimenez CR (2020) Computational analysis of phosphoproteomics data in multi-omics cancer studies. *Proteomics* **21**, e1900312.

3 Luan M, Song F, Qu S, Meng XI, Ji J, Duan Y, Sun C, Si H & Zhai H (2020) Multi-omics integrative analysis and survival risk model construction of non-small cell lung cancer based on The Cancer Genome Atlas datasets. *Oncol Lett* **20**, 58.

4 Zhang B, Yang L, Wang X & Fu D (2021) Identification of a survival-related signature for sarcoma patients through integrated transcriptomic and proteomic profiling analyses. *Gene* **764**, 145105.

5 Wu L, Yang Y, Guo X, Shu X-O, Cai Q, Shu X, Li B, Tao R, Wu C, Nikas JB *et al.* (2020) An integrative multi-omics analysis to identify candidate DNA methylation biomarkers related to prostate cancer risk. *Nat Commun* **11**, 3905.

6 Li C, Sun Y-D, Yu G-Y, Cui J-R, Lou Z, Zhang H, Huang YA, Bai C-G, Deng L-L, Liu P *et al.* (2020) Integrated omics of metastatic colorectal cancer. *Cancer Cell* **38**, 734–747.e9.

7 Burke L, Guterman I, Palacios Gallego R, Britton RG, Burschowsky D, Tufarelli C & Rufini A (2020) The Janus-like role of proline metabolism in cancer. *Cell Death Discov* **6**, 104.

8 Das S, Chandrasekaran AP, Suresh B, Haq S, Kang J-H, Lee S-J, Kim J, Kim J, Lee S, Kim HH *et al.* (2020) Genome-scale screening of deubiquitinase subfamily identifies USP3 as a stabilizer of Cdc25A

regulating cell cycle in cancer. *Cell Death Differ* **27**, 3004–3020.
9 Krämer M, Plum PS, Velazquez Camacho O, Folz-Donahue K, Thelen M, Garcia-Marquez I, Wölwer C, Büsker S, Wittig J, Franitza M *et al.* (2020) Cell type-specific transcriptomics of esophageal adenocarcinoma as a scalable alternative for single cell transcriptomics. *Mol Oncol* **14**, 1170–1184.
10 Kim SY, Jeong HH, Kim J, Moon JH & Sohn KA (2019) Robust pathway-based multi-omics data integration using directed random walks for survival prediction in multiple cancer studies. *Biol Direct* **14**, 8.
11 Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
12 Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA *et al.* (2013) Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339.
13 Cancer Genome Atlas Research Network (2014) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209.
14 Cancer Genome Atlas Research Network (2008) Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068.
15 Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP *et al.* (2010) Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110.
16 Mehrjardi NZ, Hänggi D & Kahlert UD (2020) Current biomarker-associated procedures of cancer modeling-a reference in the context of IDH1 mutant glioma. *Cell Death Dis* **11**, 998.
17 Tian L, Li Y, Edmonson MN, Zhou X, Newman S, McLeod C, Thrasher A, Liu YU, Tang BO, Rusch MC *et al.* (2020) CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol* **21**, 126.
18 Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, Kovatich AJ, Benz CC, Levine DA, Lee AV *et al.* (2018) An integrated TCGA Pan-cancer clinical data resource to drive high-quality survival outcome analytics. *Cell* **173**, 400–416.e11.
19 Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S *et al.* (2018) Oncogenic signaling pathways in The Cancer Genome Atlas. *Cell* **173**, 321–337.e10.
20 Huang L, Xu C, Yang W & Yu R (2020) A machine learning framework to determine geolocations from metagenomic profiling. *Biol Direct* **15**, 27.
21 Bellomaria A, Barbato G, Melino G, Paci M & Melino S (2012) Recognition mechanism of p63 by the E3 ligase Itch. *Cell Cycle* **11**, 3638–3648.
22 Lisitsyna OM, Kurnaeva MA, Arifulin EA, Shubina MY, Musinova YR, Mironov AA & Sheval EV (2020) Origin of the nuclear proteome on the basis of pre-existing nuclear localization signals in prokaryotic proteins. *Biol Direct* **15**, 9.
23 Schulman JJ, Szczesniak LM, Bunker EN, Nelson HA, Roe MW, Wagner LE, Yule DI & Wojcikiewicz RJH (2019) Bok regulates mitochondrial fusion and morphology. *Cell Death Differ* **26**, 2682–2694.
24 Denton D & Kumar S (2019) Autophagy-dependent cell death. *Cell Death Differ* **26**, 605–616.
25 Nazio F, Bordi M, Cianfanelli V, Locatelli F & Cecconi F (2019) Autophagy and cancer stem cells: molecular mechanisms and therapeutic applications. *Cell Death Differ* **26**, 690–702.
26 Ding X, Jia X, Wang C, Xu J, Gao SJ & Lu C (2019) A DHX9-lncRNA-MDM2 interaction regulates cell invasion and angiogenesis of cervical cancer. *Cell Death Differ* **26**, 1750–1765.
27 Nicklas S, Hillje A-L, Okawa S, Rudolph I-M, Collmann FM, van Wuellen T, del Sol A & Schwamborn JC (2019) A complex of the ubiquitin ligase TRIM32 and the deubiquitinase USP7 balances the level of c-Myc ubiquitination and thereby determines neural stem cell fate specification. *Cell Death Differ* **26**, 728–740.
28 Senichkin VV, Streletskaia AY, Gorbunova AS, Zhivotovsky B & Kopeina GS (2020) Saga of Mcl-1: regulation from transcription to degradation. *Cell Death Differ* **27**, 405–419.
29 Strappazzon F, di Rita A, Peschiaroli A, Leoncini PP, Locatelli F, Melino G & Cecconi F (2020) HUWE1 controls MCL1 stability to unleash AMBRA1-induced mitophagy. *Cell Death Differ* **27**, 1155–1168.
30 Qu Q, Li Y, Fang X, Zhang L, Xue C, Ge X, Wang X & Jiang Y (2019) Differentially expressed tRFs in CD5 positive relapsed & refractory diffuse large B cell lymphoma and the bioinformatic analysis for their potential clinical use. *Biol Direct* **14**, 23.
31 Ciocci M, Iorio E, Carotenuto F, Khashoggi HA, Nanni F & Melino S (2016) H_2S-releasing nanoemulsions: a new formulation to inhibit tumor cells proliferation and improve tissue repair. *Oncotarget* **7**, 84338–84358.
32 Mauretti A, Neri A, Kossover O, Seliktar D, di Nardo P & Melino S (2016) Design of a novel composite H_2S-releasing hydrogel for cardiac tissue repair. *Macromol Biosci* **16**, 847–858.

33 Nepravishta R, Sabelli R, Iorio E, Micheli L, Paci M & Melino S (2012) Oxidative species and S-glutathionyl conjugates in the apoptosis induction by allyl thiosulfate. *FEBS J* **279**, 154–167.
34 Wang L, Luo Y, Zheng Y, Zheng L, Lin W, Chen Z, Wu S, Chen J & Xie Y (2020) Long non-coding RNA LINC00426 contributes to doxorubicin resistance by sponging miR-4319 in osteosarcoma. *Biol Direct* **15**, 11.
35 Pallucca R, Visconti S, Camoni L, Cesareni G, Melino S, Panni S, Torreri P & Aducci P (2014) Specificity of ε and non-ε isoforms of Arabidopsis 14-3-3 proteins towards the H+-ATPase and other targets. *PLoS One* **9**, e90764.
36 Liu Y (2020) On the definition of a self-sustaining chemical reaction system and its role in heredity. *Biol Direct* **15**, 15.
37 Gallo M, Paludi D, Cicero DO, Chiovitti K, Millo E, Salis A, Damonte G, Corsaro A, Thellung S, Schettini G *et al.* (2005) Identification of a conserved N-capping box important for the structural autonomy of the prion α3-helix: the disease associated D202N mutation destabilizes the helical conformation. *Int J Immunopathol Pharmacol* **18**, 95–112.
38 Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV *et al.* (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120.
39 Cheung HH, Lee TL, Davis AJ, Taft DH, Rennert OM & Chan WY (2010) Genome-wide DNA methylation profiling reveals novel epigenetically regulated genes and non-coding RNAs in human testicular cancer. *Br J Cancer* **102**, 419–427.
40 Coetzee GA, Jia L, Frenkel B, Henderson BE, Tanay A, Haiman CA & Freedman ML (2010) A systematic approach to understand the functional consequences of non-protein coding risk regions. *Cell Cycle* **9**, 256–259.
41 Wojcik SE, Rossi S, Shimizu M, Nicoloso MS, Cimmino A, Alder H, Herlea V, Rassenti LZ, Rai KR, Kipps TJ *et al.* (2010) Non-codingRNA sequence variations in human chronic lymphocytic leukemia and colorectal cancer. *Carcinogenesis* **31**, 208–215.
42 Zhou J-D, Zhang T-J, Xu Z-J, Deng Z-Q, Gu YU, Ma J-C, Wen X-M, Leng J-Y, Lin J, Chen S-N *et al.* (2020) Genome-wide methylation sequencing identifies progression-related epigenetic drivers in myelodysplastic syndromes. *Cell Death Dis* **11**, 997.
43 Zheng H, Bi F-R, Yang Y, Hong Y-G, Ni J-S, Ma L, Liu M-H, Hao L-Q, Zhou W-P, Song L-H *et al.* (2019) Downregulation of miR-196-5p induced by hypoxia drives tumorigenesis and metastasis in hepatocellular carcinoma. *Horm Cancer* **10**, 177–189.
44 Ferrari E & Gandellini P (2020) Unveiling the ups and downs of miR-205 in physiology and cancer: transcriptional and post-transcriptional mechanisms. *Cell Death Dis* **11**, 980.
45 Li Y, Roberts ND, Wala JA, Shapira O, Schumacher SE, Kumar K, Khurana E, Waszak S, Korbel JO, Haber JE *et al.* (2020) Patterns of somatic structural variation in human cancer genomes. *Nature* **578**, 112–121.
46 ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020) Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93.
47 Goldman MJ, Zhang J, Fonseca NA, Cortés-Ciriano I, Xiang Q, Craft B, Piñeiro-Yáñez E, O'Connor BD, Bazant W, Barrera E *et al.* (2020) A user guide for the online exploration and visualization of PCAWG data. *Nat Commun* **11**, 3400.
48 Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, Hornshøj H, Hess JM, Juul RI, Lin Z *et al.* (2020) Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* **578**, 102–111.
49 Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, Mitchell TJ, Rubanova Y, Anur P, Yu K *et al.* (2020) The evolutionary history of 2,658 cancers. *Nature* **578**, 122–128.
50 ARGO-ICGC (n.d.) Retrieved September 6, 2020, from https://www.icgc-argo.org/
51 Joos S, Nettelbeck DM, Reil-Held A, Engelmann K, Moosmann A, Eggert A, Hiddemann W, Krause M, Peters C, Schuler M *et al.* (2019) German Cancer Consortium (DKTK) – a national consortium for translational cancer research. *Mol Oncol* **13**, 535–542.
52 Gehring JS, Fischer B, Lawrence M & Huber W (2015) SomaticSignatures: inferring mutational signatures from single-nucleotide variants. *Bioinformatics* **31**, 3673–3675.
53 Youk J, An Y, Park S, Lee J-K & Ju YS (2020) The genome-wide landscape of C:G>T:A polymorphism at the CpG contexts in the human population. *BMC Genom* **21**, 270.
54 Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN *et al.* (2020) The repertoire of mutational signatures in human cancer. *Nature* **578**, 94–101.
55 Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F *et al.* (2015) Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet* **47**, 505–511.
56 Chen Y-L, Zhang Y, Wang J, Chen NA, Fang W, Zhong J, Liu YI, Qin R, Yu X, Sun Z *et al.* (2019) A 17 gene panel for non-small-cell lung cancer prognosis identified through integrative epigenomic-

transcriptomic analyses of hypoxia-induced epithelial-mesenchymal transition. *Mol Oncol* **13**, 1490–1502.
57 Del Puerto-Nevado L, Minguez P, Corton M, Solanes-Casado S, Prieto I, Mas S, Sanz AB, Gonzalez-Alonso P, Villaverde C, Portal-Nuñez S *et al.* (2019) Molecular evidence of field cancerization initiated by diabetes in colon cancer patients. *Mol Oncol* **13**, 857–872.
58 Messmer MN, Snyder AG & Oberst A (2019) Comparing the effects of different cell death programs in tumor progression and immunotherapy. *Cell Death Differ* **26**, 115–129.
59 Fina D, Franzè E, Rovedatti L, Corazza GR, Biancone L, Sileri PP, Sica G, MacDonald TT, Pallone F, Di Sabatino A *et al.* (2011) Interleukin-25 production is differently regulated by TNF-α and TGF-β1 in the human gut. *Mucosal Immunol* **4**, 239–244.
60 Franzè E, Dinallo V, Rizzo A, Di Giovangiulio M, Bevivino G, Stolfi C, Caprioli F, Colantoni A, Ortenzi A, Grazia AD *et al.* (2018) Interleukin-34 sustains pro-tumorigenic signals in colon cancer tissue. *Oncotarget* **9**, 3432–3445.
61 Petljak M & Maciejowski J (2020) Molecular origins of APOBEC-associated mutations in cancer. *DNA Repair* **94**, 102905.
62 Supek F & Lehner B (2017) Clustered mutation signatures reveal that error-prone DNA repair targets mutations to active genes. *Cell* **170**, 534–547.e23.
63 Loveday C, Litchfield K, Proszek PZ, Cornish AJ, Santo F, Levy M, Macintyre G, Holryod A, Broderick P, Dudakia D *et al.* (2020) Genomic landscape of platinum resistant and sensitive testicular cancers. *Nat Commun* **11**, 2189.
64 Formica V, Luccchetti J, Cunningham D, Smyth EC, Ferroni P, Nardecchia A, Tesauro M, Cereda V, Guadagni F & Roselli M (2014) Systemic inflammation, as measured by the neutrophil/lymphocyte ratio, may have differential prognostic impact before and during treatment with fluorouracil, irinotecan and bevacizumab in metastatic colorectal cancer patients. *Med Oncol* **31**, 166.
65 Martínez JRW, Vargas-Salas S, Gamboa SU, Muñoz E, Domínguez JM, León A, Droppelmann N, Solar A, Zafereo M, Holsinger FC *et al.* (2019) The combination of RET, BRAF and demographic data identifies subsets of patients with aggressive papillary thyroid cancer. *Horm Cancer* **10**, 97–106.
66 Toda H, Seki N, Kurozumi S, Shinden Y, Yamada Y, Nohata N, Moriya S, Idichi T, Maemura K, Fujii T *et al.* (2020) RNA-sequence-based microRNA expression signature in breast cancer: tumor-suppressive miR-101-5p regulates molecular pathogenesis. *Mol Oncol* **14**, 426–446.
67 Lena AM, Cipollone R, Amelio I, Catani MV, Ramadan S, Browne G, Melino G & Candi E (2010) Skn-1a/Oct-11 and ΔNp63α exert antagonizing effects on human keratin expression. *Biochem Biophys Res Commun* **401**, 568–573.
68 Vasylenko L, Feldman MW & Livnat A (2020) The power of randomization by sex in multilocus genetic evolution. *Biol Direct* **15**, 26.
69 Harris ZN, Dhungel E, Mosior M & Ahn TH (2019) Massive metagenomic data analysis using abundance-based machine learning. *Biol Direct* **14**, 12.
70 Angelucci S, Sacchetta P, Moio P, Melino S, Petruzzelli R, Gervasi P & di Ilio C (2000) Purification and characterization of glutathione transferases from the sea bass (*Dicentrarchus labrax*) liver. *Arch Biochem Biophys* **373**, 435–441.
71 Ryan FJ (2019) Application of machine learning techniques for creating urban microbial fingerprints. *Biol Direct* **14**, 13.
72 Drost J, van Boxtel R, Blokzijl F, Mizutani T, Sasaki N, Sasselli V, de Ligt J, Behjati S, Grolleman JE, van Wezel T *et al.* (2017) Use of CRISPR-modified human stem cell organoids to study the origin of mutational signatures in cancer. *Science* **358**, 234–238.
73 Zhang Y, Chen F, Fonseca NA, He Y, Fujita M, Nakagawa H, Zhang Z, Brazma A, PCAWG Transcriptome Working Group, PCAWG Structural Variation Working Group *et al.* (2020) High-coverage whole-genome analysis of 1220 cancers reveals hundreds of genes deregulated by rearrangement-mediated cis-regulatory alterations. *Nat Commun* **11**, 736.
74 Tan J, Liu R, Zhu G, Umbricht CB & Xing M (2020) TERT promoter mutation determines apoptotic and therapeutic responses of BRAF-mutant cancers to BRAF and MEK inhibitors: Achilles Heel. *Proc Natl Acad Sci USA* **117**, 15846–15851.
75 Posch A, Hofer-Zeni S, Klieser E, Primavesi F, Naderlinger E, Brandstetter A, Filipits M, Urbas R, Swiercynski S, Jäger T *et al.* (2020) Hot spot TERT promoter mutations are rare in sporadic pancreatic neuroendocrine neoplasms and associated with telomere length and epigenetic expression patterns. *Cancers* **12**, 1625.
76 Dizman N, Lyou Y, Salgia N, Bergerot PG, Hsu JoAnn, Enriquez D, Izatt T, Trent JM, Byron S & Pal S (2020) Correlates of clinical benefit from immunotherapy and targeted therapy in metastatic renal cell carcinoma: comprehensive genomic and transcriptomic analysis. *J Immunother Cancer* **8**, e000953.
77 Schank M, Zhao J, Wang L, Li Z, Cao D, Nguyen LN, Dang X, Khanal S, Nguyen LNT, Thakuri BKC *et al.* (2020) Telomeric injury by KML001 in human T

cells induces mitochondrial dysfunction through the p53-PGC-1α pathway. *Cell Death Dis* **11**, 1030.
78 Siddiqui A, Gollavilli PN, Schwab A, Vazakidou ME, Ersan PG, Ramakrishnan M, Pluim D, Coggins S, Saatci O, Annaratone L *et al.* (2019) Thymidylate synthase maintains the de-differentiated state of triple negative breast cancers. *Cell Death Differ* **26**, 2223–2236.
79 Lamastra FR, De Angelis R, Antonucci A, Salvatori D, Prosposito P, Casalboni M, Congestri R, Melino S & Nanni F (2014) Polymer composite random lasers based on diatom frustules as scatterers. *RSC Adv* **4**, 61809–61816.
80 ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020) Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93.
81 Cortés-Ciriano I, Lee JJ-K, Xi R, Jain D, Jung YL, Yang L, Gordenin D, Klimczak LJ, Zhang C-Z, Pellman DS *et al.* (2020) Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing. *Nat Genet* **52**, 331–341.
82 Zapatka M, Borozan I, Brewer DS, Iskar M, Grundhoff A, Alawi M, Desai N, Sültmann H, Moch H, PCAWG Pathogens *et al.* (2020) The landscape of viral associations in human cancers. *Nat Genet* **52**, 320–330.
83 Amelio I & Melino G (2020) Context is everything: extrinsic signalling and gain-of-function p53 mutants. *Cell Death Discov* **6**, 16.
84 Celardo I, Melino G & Amelio I (2020) Commensal microbes and p53 in cancer progression. *Biol Direct* **15**, 25.
85 Mantovani F, Collavin L & del Sal G (2019) Mutant p53 as a guardian of the cancer cell. *Cell Death Differ* **26**, 199–212.
86 Chen T-W, Lee C-C, Liu H, Wu C-S, Pickering CR, Huang P-J, Wang J, Chang I-F, Yeh Y-M, Chen C-D *et al.* (2017) APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nat Commun* **8**, 465.
87 Li Z, Abraham BJ, Berezovskaya A, Farah N, Liu Y, Leon T, Fielding A, Tan SH, Sanda T, Weintraub AS *et al.* (2017) APOBEC signature mutation generates an oncogenic enhancer that drives LMO1 expression in T-ALL. *Leukemia* **31**, 2057–2064.
88 The Cancer Genome Atlas Research Network (2017) Integrated genomic and molecular characterization of cervical cancer. *Nature* **543**, 378–384.
89 Sathyamoorthy N & Lange CA (2020) Progesterone and breast cancer: an NCI workshop report. *Horm Cancer* **11**, 1–12.
90 Fowler AM, Salem K, DeGrave M, Ong IM, Rassman S, Powers GL, Kumar M, Michel CJ & Mahajan AM (2020) Progesterone receptor gene variants in metastatic estrogen receptor positive breast cancer. *Horm Cancer* **11**, 63–75.
91 Liu L, Wang G, Wang L, Yu C, Li M, Song S, Hao L, Ma L & Zhang Z (2020) Computational identification and characterization of glioma candidate biomarkers through multi-omics integrative profiling. *Biol Direct* **15**, 10.
92 Ham SW, Jeon H-Y, Jin X, Kim E-J, Kim J-K, Shin YJ, Lee Y, Kim SH, Lee SY, Seo S *et al.* (2019) TP53 gain-of-function mutation promotes inflammation in glioblastoma. *Cell Death Differ* **26**, 409–425.
93 Huang S, Li Y, Yuan X, Zhao M, Wang J, Li Y, Li Y, Lin H, Zhang Q, Wang W *et al.* (2019) The UbL-UBA ubiquilin4 protein functions as a tumor suppressor in gastric cancer by p53-dependent and p53-independent regulation of p21. *Cell Death Differ* **26**, 516–530.
94 Yang L (2020) A practical guide for structural variation detection in the human genome. *Curr Protoc Hum Genet* **107**, e103.
95 Jiang L, Huguet G, Schramm C, Ciampi A, Main A, Passo C, Jean-Louis M, Auger M, Schumann G, Porteous D *et al.* (2020) Estimating the effects of copy-number variants on intelligence using hierarchical Bayesian models. *Genet Epidemiol* **44**, 825–840.
96 Akdemir KC, Le VT, Chandran S, Li Y, Verhaak RG, Beroukhim R, Campbell PJ, Chin L, Dixon JR, Futreal PA *et al.* (2020) Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. *Nat Genet* **52**, 294–305.
97 Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V & Korbel JO (2012) DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* **28**, i333–i339.
98 Xia B, Sheng Q, Nakanishi K, Ohashi A, Wu J, Christ N, Liu X, Jasin M, Couch FJ & Livingston DM (2006) Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Mol Cell* **22**, 719–729.
99 Welch J, Ley T, Link D, Miller C, Larson D, Koboldt D, Wartman L, Lamprecht T, Liu F, Xia J *et al.* (2012) The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278.
100 Yates LR, Gerstung M, Knappskog S, Desmedt C, Gundem G, Van Loo P, Aas T, Alexandrov LB, Larsimont D, Davies H *et al.* (2015) Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med* **21**, 751–759.
101 Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P *et al.* (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366**, 883–892.

102 Pitolli C, Wang Y, Mancini M, Shi Y, Melino G & Amelio I (2019) Do mutations turn p53 into an oncogene? *Int J Mol Sci* **20**, 6241.

103 Lonetto G, Koifman G, Silberman A, Attery A, Solomon H, Levin-Zaidman S, Goldfinger N, Porat Z, Erez A & Rotter V (2019) Mutant p53-dependent mitochondrial metabolic alterations in a mesenchymal stem cell-based model of progressive malignancy. *Cell Death Differ* **26**, 1566–1581.

104 Li Y, Cao Y, Xiao J, Shang J, Tan Q, Ping F, Huang W, Wu F, Zhang H & Zhang X (2020) Inhibitor of apoptosis-stimulating protein of p53 inhibits ferroptosis and alleviates intestinal ischemia/reperfusion-induced acute lung injury. *Cell Death Differ* **27**, 2635–2650

105 Doffe F, Carbonnier V, Tissier M, Leroy B, Martins I, Mattsson JSM, Micke P, Pavlova S, Pospisilova S, Smardova J *et al.* (2020) Identification and functional characterization of new missense SNPs in the coding region of the TP53 gene. *Cell Death Differ* **28**, 1477–1492.

106 Sibio S, Di Giorgio A, D'Ugo S, Palmieri G, Cinelli L, Formica V, Sensi B, Bagaglini G, Di Carlo S, Bellato V *et al.* (2019) Histotype influences emergency presentation and prognosis in colon cancer surgery. *Langenbecks Arch Surg* **404**, 841–851.

107 Machnik M, Cylwa R, Kiełczewski K, Biecek P, Liloglou T, Mackiewicz A & Oleksiewicz U (2019) The expression signature of cancer-associated KRAB-ZNF factors identified in TCGA pan-cancer transcriptomic data. *Mol Oncol* **13**, 701–724.

108 Mandal P, Saha SS, Sen S, Bhattacharya A, Bhattacharya NP, Bucha S, Sinha M, Chowdhury RR, Mondal NR, Chakravarty B *et al.* (2019) Cervical cancer subtypes harbouring integrated and/or episomal HPV16 portray distinct molecular phenotypes based on transcriptome profiling of mRNAs and miRNAs. *Cell Death Discov* **5**, 81.

109 Rivas MA, Pirinen M, Conrad DF, Lek M, Tsang EK, Karczewski KJ, Maller JB, Kukurba KR, DeLuca DS, Fromer M *et al.* (2015) Human genomics. Effect of predicted protein-truncating genetic variants on the human transcriptome. *Science* **348**, 666–669.

110 Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, Young TR, Goldmann JM, Pervouchine DD, Sullivan TJ *et al.* (2015) Human genomics. The human transcriptome across tissues and individuals. *Science* **348**, 660–665.

111 GTEx Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.

112 PCAWG Transcriptome Core Group, Calabrese C, Davidson NR, Demircioğlu D, Fonseca NA, He Y, Kahles A, Lehmann KV, Liu F, Shiraishi Y *et al.* (2020) Genomic basis for RNA alterations in cancer. *Nature* **578**, 129–136.

113 Kahraman A, Karakulak T, Szklarczyk D & von Mering C (2020) Pathogenic impact of transcript isoform switching in 1,209 cancer samples covering 27 cancer types using an isoform-specific interaction network. *Sci Rep* **10**, 14453.

114 Nie X, Xiao D, Ge Y, Xie Y, Zhou H, Zheng T, Li X, Liu H, Huang H & Zhao Y (2021) TRF2 recruits nucleolar protein TCOF1 to coordinate telomere transcription and replication. *Cell Death Differ* **28**, 1062–1075.

115 Korkut A, Zaidi S, Kanchi RS, Rao S, Gough NR, Schultz A, Li X, Lorenzi PL, Berger AC, Robertson G *et al.* (2018) A Pan-cancer analysis reveals high-frequency genetic alterations in mediators of signaling by the TGF-β superfamily. *Cell Syst* **7**, 422–437.e7.

116 Ma X, Liu YU, Liu Y, Alexandrov LB, Edmonson MN, Gawad C, Zhou X, Li Y, Rusch MC, Easton J *et al.* (2018) Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature* **555**, 371–376.

117 Agostini M, Ganini C, Candi E & Melino G (2020) The role of noncoding RNAs in epithelial cancer. *Cell Death Discov* **6**, 13.

118 Vishnubalaji R, Shaath H, Elkord E & Alajez NM (2019) Long non-coding RNA (lncRNA) transcriptional landscape in breast cancer identifies LINC01614 as non-favorable prognostic biomarker regulated by TGFβ and focal adhesion kinase (FAK) signaling. *Cell Death Discov* **5**, 109.

119 Wang M, Mao C, Ouyang L, Liu Y, Lai W, Liu NA, Shi Y, Chen L, Xiao D, Yu F *et al.* (2019) Long noncoding RNA LINC00336 inhibits ferroptosis in lung cancer by functioning as a competing endogenous RNA. *Cell Death Differ* **26**, 2329–2343.

120 Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, Duyvesteyn K, Haidari S, van Hoeck A, Onstenk W *et al.* (2019) Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210–216.

121 Qiu S, Nikolaou S, Fiorentino F, Rasheed S, Darzi A, Cunningham D, Tekkis P & Kontovounisios C (2019) Exploratory analysis of plasma neurotensin as a novel biomarker for early detection of colorectal polyp and cancer. *Horm Cancer* **10**, 128–135.

122 Bakhoum SF & Cantley LC (2018) The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell* **174**, 1347–1360.

123 Zheng T, Zhou H, Li X, Peng DI, Yang Y, Zeng Y, Liu H, Ren J & Zhao Y (2020) RBMX is required for activation of ATR on repetitive DNAs to maintain genome stability. *Cell Death Differ* **27**, 3162–3176.

124 Cieslik M & Chinnaiyan AM (2020) News & views Global cancer genomics project comes to fruition. *Nature* **578**, 40.

125 Neumeister VM & Juhl H (2018) Tumor pre-analytics in molecular pathology: impact on protein expression and analysis. *Curr Pathobiol Rep* **6**, 265–274.
126 Samsen A, von der Heyde S, Bokemeyer C, David KA, Flath B, Graap M, Grebenstein B, Heflik L, Hollburg W, Layer P *et al.* (2018) Multi-omic based molecular profiling of advanced cancer identifies treatable targets and improves survival in individual patients. *Oncotarget* **9**, 34794–34809.
127 Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kamińska B, Huelsken J, Omberg L, Gevaert O *et al.* (2018) Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell* **173**, 338–354.e15.
128 Chierici M, Francescatto M, Bussola N, Jurman G & Furlanello C (2020) Predictability of drug-induced liver injury by machine learning. *Biol Direct* **15**, 3.
129 Koch K, Hartmann R, Tsiampali J, Uhlmann C, Nickel A-C, He X, Kamp MA, Sabel M, Barker RA, Steiger H-J *et al.* (2020) A comparative pharmaco-metabolomic study of glutaminase inhibitors in glioma stem-like cells confirms biological effectiveness but reveals differences in target-specificity. *Cell Death Discov* **6**, 20.
130 Doccini S, Morani F, Nesti C, Pezzini F, Calza G, Soliymani R, Signore G, Rocchiccioli S, Kanninen KM, Huuskonen MT *et al.* (2020) Proteomic and functional analyses in disease models reveal CLN5 protein involvement in mitochondrial dysfunction. *Cell Death Discov* **6**, 18.
131 Oktay K, Santaliz-Casiano A, Patel M, Marino N, Storniolo AMV, Torun H, Acar B & Madak Erdogan Z (2020) A computational statistics approach to evaluate blood biomarkers for breast cancer risk stratification. *Horm Cancer* **11**, 17–33.
132 Lin C, Li H, Liu J, Hu Q, Zhang S, Zhang NA, Liu L, Dai Y, Cao D, Li X *et al.* (2020) Arginine hypomethylation-mediated proteasomal degradation of histone H4-an early biomarker of cellular senescence. *Cell Death Differ* **27**, 2697–2709.
133 Han Y, Ye X, Wang C, Liu Y, Zhang S, Feng W, Huang K & Zhang J (2019) Integration of molecular features with clinical information for predicting outcomes for neuroblastoma patients. *Biol Direct* **14**, 16.
134 Yan K, Da T-T, Bian Z-H, He YI, Liu M-C, Liu Q-Z, Long J, Li L, Gao C-Y, Yang S-H *et al.* (2020) Multi-omics analysis identifies FoxO1 as a regulator of macrophage function through metabolic reprogramming. *Cell Death Dis* **11**, 800.
135 Yang S, Liu T & Liang G (2020) The benefits of smoking cessation on survival in cancer patients by integrative analysis of multi-omics data. *Mol Oncol* **14**, 2069–2080.
136 Fiorino C, Guckemberger M, Schwarz M, van der Heide UA & Heijmen B (2020) Technology-driven research for radiotherapy innovation. *Mol Oncol* **14**, 1500–1513.
137 Hoang LT, Domingo-Sabugo C, Starren ES, Willis-Owen SAG, Morris-Rosendahl DJ, Nicholson AG, Cookson WOCM & Moffatt MF (2019) Metabolomic, transcriptomic and genetic integrative analysis reveals important roles of adenosine diphosphate in haemostasis and platelet activation in non-small-cell lung cancer. *Mol Oncol* **13**, 2406–2421.
138 Michaletti A, Mancini M, Smirnov A, Candi E, Melino G & Zolla L (2019) Multi-omics profiling of calcium-induced human keratinocytes differentiation reveals modulation of unfolded protein response signaling pathways. *Cell Cycle* **18**, 2124–2140.
139 Chen P-H, Wu J, Ding C-K, Lin C-C, Pan S, Bossa N, Xu Y, Yang W-H, Mathey-Prevot B & Chi J-T (2020) Kinome screen of ferroptosis reveals a novel role of ATM in regulating iron metabolism. *Cell Death Differ* **27**, 1008–1022.
140 Alsahafi E, Begg K, Amelio I, Raulf N, Lucarelli P, Sauter T & Tavassoli M (2019) Clinical update on head and neck cancer: molecular biology and ongoing challenges. *Cell Death Dis* **10**, 540.
141 Ivano A (2019) How mutant p53 empowers Foxh1 fostering leukaemogenesis? *Cell Death Discov* **5**, 108.
142 Fiala C & Diamandis EP (2018) Utility of circulating tumor DNA in cancer diagnostics with emphasis on early detection. *BMC Med* **16**, 166.
143 Haghjoo N, Moeini A & Masoudi-Nejad A (2020) Introducing a panel for early detection of lung adenocarcinoma by using data integration of genomics, epigenomics, transcriptomics and proteomics. *Exp Mol Pathol* **112**, 104360.
144 Jing Y-Y, Cai F-F, Zhang L, Han J, Yang LU, Tang F, Li Y-B, Chang J-F, Sun F, Yang X-M *et al.* (2020) Epigenetic regulation of the Warburg effect by H2B monoubiquitination. *Cell Death Differ* **27**, 1660–1676.
145 Mihaylov I, Kańduła M, Krachunov M & Vassilev D (2019) A novel framework for horizontal and vertical data integration in cancer studies with application to survival time prediction models. *Biol Direct* **14**, 22.
146 Zubor P, Kubatka P, Kajo K, Dankova Z, Polacek H, Bielik T, Kudela E, Samec M, Liskova A, Vlcakova D *et al.* (2019) Why the gold standard approach by mammography demands extension by multiomics? Application of liquid biopsy miRNA profiles to breast cancer disease management. *Int J Mol Sci* **20**, 2878.
147 Wu X, Li J, Gassa A, Buchner D, Alakus H, Dong Q, Ren N, Liu M, Odenthal M, Stippel D *et al.* (2020) Circulating tumor DNA as an emerging liquid biopsy biomarker for early diagnosis and therapeutic

monitoring in hepatocellular carcinoma. *Int J Biol Sci* **16**, 1551–1562.
148 Ankney JA, Xie L, Wrobel JA, Wang L & Chen X (2019) Novel secretome-to-transcriptome integrated or secreto-transcriptomic approach to reveal liquid biopsy biomarkers for predicting individualized prognosis of breast cancer patients. *BMC Med Genomics* **12**, 78.
149 Amelio I, Bertolo R, Bove P, Buonomo OC, Candi E, Chiocchi M, Cipriani C, Di Daniele N, Ganini C, Juhl H *et al.* (2020) Liquid biopsies and cancer omics. *Cell Death Discov* **6**, 131.
150 Tavassoly I, Hu Y, Zhao S, Mariottini C, Boran A, Chen Y, Li L, Tolentino RE, Jayaraman G, Goldfarb J *et al.* (2019) Genomic signatures defining responsiveness to allopurinol and combination therapy for lung cancer identified by systems therapeutics analyses. *Mol Oncol* **13**, 1725–1743.
151 Amelio I, Bertolo R, Bove P, Candi E, Chiocchi M, Cipriani C, Di Daniele N, Ganini C, Juhl H, Mauriello A *et al.* (2020) Cancer predictive studies. *Biol Direct* **15**, 18.
152 Werner J, Géron A, Kerssemakers J & Matallana-Surget S (2019) mPies: a novel metaproteomics tool for the creation of relevant protein databases and automatized protein annotation. *Biol Direct* **14**, 21.
153 Dobon B, Montanucci L, Peretó J, Bertranpetit J & Laayouni H (2019) Gene connectivity and enzyme evolution in the human metabolic network. *Biol Direct* **14**, 17.
154 Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, Hassel JC, Rutkowski P, McNeil C, Kalinka-Warzocha E *et al.* (2015) Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med* **372**, 320–330.
155 Motzer RJ, Escudier B, McDermott DF, George S, Hammers HJ, Srinivas S, Tykodi SS, Sosman JA, Procopio G, Plimack ER *et al.* (2015) Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* **373**, 1803–1813.
156 Brahmer J, Reckamp KL, Baas P, Crinò L, Eberhardt WEE, Poddubskaya E, Antonia S, Pluzanski A, Vokes EE, Holgado E *et al.* (2015) Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. *N Engl J Med* **373**, 123–135.
157 Weber JS, D'Angelo SP, Minor D, Hodi FS, Gutzmer R, Neyns B, Hoeller C, Khushalani NI, Miller WH, Lao CD *et al.* (2015) Nivolumab versus chemotherapy in patients with advanced melanoma who progressed after anti-CTLA-4 treatment (CheckMate 037): a randomised, controlled, open-label, phase 3 trial. *Lancet Oncol* **16**, 375–384.
158 Rizvi NA, Mazières J, Planchard D, Stinchcombe TE, Dy GK, Antonia SJ, Horn L, Lena H, Minenza E, Mennecier B *et al.* (2015) Activity and safety of nivolumab, an anti-PD-1 immune checkpoint inhibitor, for patients with advanced, refractory squamous non-small-cell lung cancer (CheckMate 063): a phase 2, single-arm trial. *Lancet Oncol* **16**, 257–265.
159 Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L *et al.* (2017) An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221.
160 Finotello F & Eduati F (2018) Multi-omics profiling of the tumor microenvironment: paving the way to precision immuno-oncology. *Front Oncol* **8**, 430.
161 Bolotin DA, Poslavsky S, Davydov AN, Frenkel FE, Fanchi L, Zolotareva OI, Hemmers S, Putintseva EV, Obraztsova AS, Shugay M *et al.* (2017) Antigen receptor repertoire profiling from RNA-seq data. *Nat Biotechnol* **35**, 908–911.
162 GENIE (n.d.) Retrieved February 1, 2021, from https://www.aacr.org/professionals/research/aacr-project-genie/
163 Clarke MA & Fisher J (2020) Executable cancer models: successes and challenges. *Nat Rev Cancer* **20**, 343–354.
164 Sumsion GR, Bradshaw MS, Beales JT, Ford E, Caryotakis GRG, Garrett DJ, LeBaron ED, Nwosu IO & Piccolo SR (2020) Diverse approaches to predicting drug-induced liver injury using gene-expression profiles. *Biol Direct* **15**, 1.
165 Ben-David U, Beroukhim R & Golub TR (2019) Genomic evolution of cancer models: perils and opportunities. *Nat Rev Cancer* **19**, 97–109.
166 Chen JC & Tyler AD (2020) Systematic evaluation of supervised machine learning for sample origin prediction using metagenomic sequencing data. *Biol Direct* **15**, 29.
167 Han Y, Ye X, Cheng J, Zhang S, Feng W, Han Z, Zhang J & Huang K (2019) Integrative analysis based on survival associated co-expression gene modules for predicting Neuroblastoma patients' survival time. *Biol Direct* **14**, 4.
168 Caputo A, Fournier P-E & Raoult D (2019) Genome and pan-genome analysis to classify emerging bacteria. *Biol Direct* **14**, 5.
169 Larmuseau M, Verbeke LPC & Marchal K (2019) Associating expression and genomic data using co-occurrence measures. *Biol Direct* **14**, 10.