Computational and Structural Biotechnology Journal 20 (2022) 5065–5075

COMPUTATIONAL ANDSTRUCTURAL BIOTECHNOLOGY JOURNAL

journal homepage: www.elsevier.com/locate/csbj

Mini review

The seen and the unseen: Molecular classification and image based-analysis of gastrointestinal cancers

Corina-Elena Minciuna [a,b], Mihai Tanase [c,d], Teodora Ecaterina Manuc [b,e], Stefan Tudor [a,b], Vlad Herlea [f,g], Mihnea P. Dragomir [h,i,j,*], George A. Calin [k,l,*], Catalin Vasilescu [a,b,*]

[a] Department of Surgery, Fundeni Clinical Institute, Bucharest, Romania
[b] Carol Davila University of Medicine and Pharmacy, Bucharest, Romania
[c] Department of Automatic Control and Computers, Politehnica University of Bucharest, Bucharest, Romania
[d] University of Bucharest, Bucharest, Romania
[e] Department of Gastroenterology, Fundeni Clinical Institute, Bucharest, Romania
[f] Department of Pathology, Fundeni Clinical Institute, 022328 Bucharest, Romania
[g] "Titu Maiorescu" University, Bucharest, Romania
[h] German Cancer Consortium (DKTK), Partner Site Berlin, and German Cancer Research Center (DKFZ), Heidelberg, Germany
[i] Berlin Institute of Health (BIH), Berlin, Germany
[j] Institute of Pathology, Charité – Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[k] Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
[l] Center for RNA Interference and Non-Coding RNAs, The University of Texas MD Anderson Cancer Center, Houston, TX 77054, USA

ARTICLE INFO

Article history:
Received 5 July 2022
Received in revised form 7 September 2022
Accepted 7 September 2022
Available online 12 September 2022

Keywords:
Artificial intelligence
Molecular classification
Image-based classification
Gastric adenocarcinoma

ABSTRACT

Gastrointestinal cancers account for 22.5% of cancer related deaths worldwide and represent circa 20% of all cancers. In the last decades, we have witnessed a shift from histology-based to molecular-based classifications using genomic, epigenomic, and transcriptomic data. The molecular based classification revealed new prognostic markers and may aid the therapy selection. Because of the high-costs to perform a molecular classification, in recent years immunohistochemistry-based surrogate classification were developed which permit the stratification of patients, and in parallel multiple groups developed hematoxylin and eosin whole slide image analysis for sub-classifying these entities. Hence, we are witnessing a return to an image-based classification with the purpose to infer hidden information from routine histology images that would permit to detect the patients that respond to specific therapies and would be able to predict their outcome. In this review paper, we will discuss the current histological, molecular, and immunohistochemical classifications of the most common gastrointestinal cancers, gastric adenocarcinoma, and colorectal adenocarcinoma, and will present key aspects for developing a new artificial intelligence aided image-based classification of these malignancies.

Contents

* Corresponding authors at: Institute of Pathology, Charité - Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin (M.P. Dragomir). Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA (G.A. Calin). Department of Surgery, Fundeni Clinical Institute, Bucharest, Romania (C. Vasilescu).

E-mail addresses: mihnea.dragomir@charite.de (M.P. Dragomir), gcalin@mdaderson.org (G.A. Calin), catvasilescu@gmail.com (C. Vasilescu).

https://doi.org/10.1016/j.csbj.2022.09.010

1. Introduction

Gastrointestinal cancers (GIC) account for 22.5 % of cancer related deaths worldwide and represent circa 20 % of all cancers [1]. In order to keep this review concise, we will focus on two main GIC subtypes, gastric adenocarcinoma (GAC) and colorectal adenocarcinoma (CRC), for which the most interesting **molecular classification** (terms outlined in bold are further presented in the Glossary section) and the first steps for image based-analysis were recently developed.

Specifically, GAC is the fourth most common cancer type [2] and the number three cause of cancer related death worldwide [3] being an important public health issue. Around 35 % of GAC patients are diagnosed with metastatic disease [4]. Prior to chemotherapy, the survival was of only 5 months [5,6]. Although adjuvant chemotherapy increased survival to 8 months [5], the prognosis remains unfavourable, with a median survival of under 1-year [5,7]. Recently, neoadjuvant chemotherapy emerged showing improved overall survival compared with surgery alone, 5-year survival rate of 36 % vs 23 %, and a progression-free survival benefit [8] becoming the preferred treatment option for Stage IB resectable GAC or higher according to ESMO Guidelines [9] and NCCN Guidelines [10]. In what concerns CRC, it is the fourth most deadly cancer, with the highest rates of incidence in developed countries [11]. Novel therapeutic strategies have doubled the overall survival for advanced stage disease to 3 years [11], but giving the fact that it becomes symptomatic late on, it represent approximately 10 % of cancer-related mortality in western countries [12].

The introduction of immunotherapy, in particular immune checkpoint inhibitors (ICIs), was an important advance in oncology. The results of this new therapy in advanced stage melanoma [13] encouraged oncologists around the world to use immunotherapy in the treatment of digestive cancers, including GAC and CRC. Unfortunately, the therapeutic effect recorded in stage IV melanoma has not been recapitulated for GAC: median survival remained around one year in metastatic disease [2] even though targeted immunotherapies have been introduced in the clinical use for advanced GAC [2]. The 5-year survival rate under 5 % in advanced stage disease emphasizes the need for a better biologic understanding of this neoplasia in order to develop novel therapies [14]. Not all GAC patients respond effective to immunotherapy. Hence, new strategies need to be developed to discover the patients that can benefit from this therapy.

Likewise, CRC is a complex heterogenous disease and despite all the advancements made in treating CRC there are many current and potential variables influencing the treatment plan: location (right-side colon cancer is a negative prognostic factor for the overall survival in patients who underwent treatment with curative intent for colon cancer [15] and it steered treatment decision in first line [16]), stage, cancer grade, and genomic biomarkers [17] (CIMP - CpG island methylator phenotype High, MSI - microsatellite instability High, *MLH1* methylation, *BRAF* mutation, CIN - chromosomal instability, CMS - consensus molecular subtypes, *RAS* mutational status, EGFR/HER family, TP53-APC/β -catenin and various **microRNAs** (miRNAs) - miR-31, miR-99a, miR-125b, miR-181a [18,19]).

Currently immune checkpoint therapy is approved for PD-L1 positive GAC (defined as a Combined Positive Score-CPS $\geq$ 1) [10], and it is well known that microsatellite instable high (MSI-H), or PD-L1 CPS $\geq$ 10 tumors show a much better response rate [20]. Other score considered and reported in some trials is tumor proportion score (TPS) [10]. Other monoclonal antibodies approved are trastuzumab for HER2-positive tumors and ramucirumab for VEGFR2 positive [21]. In CRC, drug resistance against chemotherapeutic regimens poses a serious challenge, finally leading to chemotherapy failure [22]. Hopes have come from anti-EGFR agents as cetuximab or panitumumab [23], anti-angiogenesis agents as bevacizumab and anti PD-L1 as pembrolizumab, that are currently approved only for MSI-H metastasized CRC [22,24].

These markers only partially reflect the tumor biology. It became clear that in both cases the tumor heterogeneity is a crucial cause of lack of response to ICIs and it holds profound implications for therapy selection [3].

Great efforts have been made to overcome the barrier of the molecular heterogeneity of GAC and sub-classify these tumors. The so called "molecular revolution" tried to identify optimal gene sets in order to predict the disease course and the response to chemotherapy or ICI.

2. The morphological classifications

The first classifications of gastrointestinal cancers were made based on the tumor morphology using simple hematoxylin and eosin (H&E) slides.

Stomach tumors are mainly epithelial. There are multiple types of epithelial malignant tumors of the stomach, but by far the most common one is GAC. GAC is defined as a neoplasia with glandular differentiation originating from the stomach mucosa. Laurén classified GAC in 1965 into diffuse type and intestinal type [25]. Diffuse type GAC is characterized by infiltrating isolated cells or small bundles of cells that do not form glands. Often these cells have a signet-ring cell morphology. On the other hand, intestinal type GAC is composed of well-structured glands. As expected, diffuse type GAC have a poor prognosis compared to intestinal type GAC. The diagnosis of diffuse type GAC is more challenging, especially if metastasized, the poorly cohesive cells hide in an inflammatory or fibrous background making their detection sometimes difficult.

Currently, the WHO proposes five main morphological subtypes of GAC [26]. The five subtypes are tubular adenocarcinoma (if well/moderately differentiated matches the intestinal type, if poorly it is similar to the diffuse type), papillary adenocarcinoma (matches the intestinal type), poorly cohesive also including signet-ring cell carcinoma (matches the diffuse type), mucinous adenocarcinoma (can be both intestinal or diffuse) and mixed adenocarcinoma (matches a mixed subtype of Laurén containing both intestinal and diffuse type). Additionally, other subtypes are described, but these are rare and are beyond the scope of this review. All these subtypes not only look different but also have a different prognosis and clinical course. For example, the papillary adenocarcinoma (a former intestinal type) frequently is associated with liver metastases and has a poor outcome [27]. Tubular GAC is the most common subtype of GAC and is more frequent in Japanese population, and if solid components are present it associates with MSI [28]. Poorly cohesive GAC is well known to have an unfavorable prognosis and can harbor *RHOA* mutations [29]. Mucinous

GAC is rare, shows MSI, and *TP53* is the most common driver mutation [30]. Moreover, mixed GAC shows multiple phenotypes, and the poorly cohesive component is *E-cadherin* mutated [31]. Therefore, it seems that a lot of information is hidden in the banal morphology of adenocarcinomas, and we ask ourselves how much of it are we missing.

The other very common subtype of gastrointestinal tract cancers is CRC. CRC is the main epithelial neoplasia originating from the colon. Regarding the morphology most CRC are classified as Not Otherwise Specified (NOS) [32]. Despite this, the WHO proposes nine other subtypes of CRC, all having specific clinical features and prognostic impact: mucinous adenocarcinoma, adenoma-like adenocarcinoma, serrated adenocarcinoma, micropapillary adenocarcinoma, signet-ring cell carcinoma, medullary carcinoma, adenosquamous carcinoma, carcinomas with sarcomatoid components and undifferentiated carcinomas [26]. For example, mucinous CRCs have a prognosis similar to CRCs NOS, although being enriched in MSI [33,34]. Additionally, mucinous tumors show more frequently *BRAF*, and *PIK3CA* mutations and alterations of the transforming-growth-factor-beta pathway [34]. Moreover, these tumors have a high number of tumor infiltrating lymphocytes [34]. Signet-ring cell CRCs are rare (∼1%), have an unfavorable prognosis and are also enriched in MSI and *BRAF* mutations being depleted in *KRAS* mutations [35]. Medullary morphology is associated with a good prognosis and is associated with MSI [36]. Serrated CRCs harbor *BRAF* mutations and *MLH1* methylation and have a specific intestinal microbiota [37]. Adenoma-like adenocarcinoma morphology is associated with a favorable prognosis often showing *KRAS* mutations [38]. Micropapillary adenocarcinomas of the colon and rectum show high levels of lymph node metastases, vascular and perineural invasion, being associated with an unfavorable prognosis [39]. Adenosquamous carcinoma of the colon and rectum usually have a higher stage at diagnosis and consequently a shorter survival [40]. Finally, carcinomas with sarcomatoid components show alterations of the SWItch/sucrose nonfermentable (SWI/SNF) chromatin remodeling complex [41].

All these data together clearly show that the morphology is at least partially containing the molecular information of gastrointestinal neoplasia and the hypothesis of developing an image-based classification of GICs is supported by these data. Unfortunately, because of the interobserver variability, the morphology of tumors traditionally plays a secondary role and more reliable tools for quantifying it are highly necessary.

3. Molecular classifications of gastrointestinal cancers

In 2014, The Cancer Genome Atlas (TCGA) research network presented four molecular subtypes based on genomic profiling data of primary GAC [42]. The four subtypes (EBV - Epstein-Barr virus, MSI - microsatellite instability, CIN - chromosomal instability, and GS - genomically stable) have led to a better selection and stratification of patients that may respond to immunotherapy, and other types of chemotherapy [3]. This molecular classification became a roadmap for stratifying patients in order to develop specific targeted therapies. To define these GAC subgroups, advanced molecular techniques are necessary, which are not feasible nor cost effective in current clinical practice [43]. Therefore, we consider the molecular classification of GAC as one step towards personalized cancer therapy and follow-up.

The widespread application of molecular classification was hampered by the fact that it requires advanced technology, time-consuming bioinformatics methods that extend too much the time-sensitive delivery of the results, the lack of robust classifiers that are platform-independent and especially the remarkably high costs.

For CRC, until 2015, there were a lot of inconsistencies among the transcriptomic-based CRC classifications that impeded the clinical translation [44]. A multicentre group tried to harmonize these classification systems by coalescing into four subtypes that are considered to represent distinct groups [44]. They established a consensus molecular subtype classification based on gene expression profiles from bulk tumours: CMS1-14 % - microsatellite instability immune, CMS2-37 % - canonical, CMS3 - 13 %-metabolic, and CMS4-23 % - mesenchymal [44]. The rest of the samples, 13 %, had mixed features attributed to the tumour heterogeneity or a transition phenotype being considered indeterminate CMS subtype [44].

CMS1 is characterized by MSI-H and CIMP high, generally has good prognosis, but the worst prognosis if relapsing, represents 14 % of total CRC tumours, and is predominant in the proximal colon [45], CMS4 has high somatic copy number alterations, has the worst overall survival, and potentially display high response against heat shock proteins like inhibitors [46]. CMS2 also has high level of somatic copy number alterations and shows MYC and WNT activation, best responds to anti-EGFR and HER2 inhibitors [46] and is predominant found in distal colon and rectum [45]. CMS3 has often *KRAS* mutations, has the 5-year survival rate of approximately 75 %, the second-highest of the four subtypes, with approximately 13 % of the patients having this subtype [47]. This classification has been reproduced across multiple studies [48,49] and represents one of the most accepted molecular classification of CRC. An important limitation of this classification is that a high number of samples do not fall in any of the classes. Therefore, more specific classifications are highly needed.

In 2007, Jass [50] anticipated that a possible solution to understanding CRC heterogeneity would be the study of its molecular features. He proposed a classification comprised of 5 subgroups based on combinations of MSI and CIMP status and existence of *BRAF* and *KRAS* mutations as follow: type 1 (CIMP-high/MSI-H/*BRAF* mutation), type 2 (CIMP-high/MSI-L or MSS/*BRAF* mutation), type 3 (CIMP-low/MSS or MSI-L/*KRAS* mutation), type 4 (CIMP-neg/MSS) and type 5 or Lynch syndrome (CIMP-neg/MSI-H). This classification is a predecessor of the purely transcriptomics-based classification. Currais *et al.* [51] pointed out that MSI-high and CIMP-high correspond to CMS1 group, CMS2 and CMS4 are MSS/MSI-low tumors and CMS3 are CIMP-negative tumors.

Recently, in 2022, Joanito *et al.* [52] proposed a new approach to classifying CRC, refining the CMS [44], by combining intrinsic epithelial subtype (I), microsatellite instability (M) and fibrosis (F). The refined IMF classification has 5 subtypes: iCMS2_MSS_NF, iCMS2_MSS_F, iCMS3_MSS_NF, iCMS3_MSS_F and iCMS3_MSI. Regarding the relationship with the CMS, the IMF found that CMS1 and CMS3 tumours are mainly i3, while CMS2 were mainly i2, but CMS4 with an equal proportion can be i2 or i3 (being stratified by the epithelial subtype). In what concerns the correlation with clinic-molecular characteristics, the right sided tumours were mainly i3, 66 %, and the left sided tumours were mainly i2, 68 %. Consistently with the CMS classification that showed poor relapse free survival for CMS4 subgroup [44], this was also a particular feature of the CMS4/iCMS3 subgroup [52], with the same effect on the overall survival, inferior to all others subgroups. After relapse, the survival was worse for i3 patients than for the i2 ones [52].

4. Immunohistochemical classifications of gastrointestinal cancer

Even though clinically used genetic or immunohistochemical techniques already are used in GAC for identification of EBV, and MSI subgroups, characterizing other more complex and heteroge-

nous subgroups are more technically challenging [3]. A more affordable alternative is required. A solution with limited potential is immunohistochemistry combined with i*n situ* hybridization with the following markers proposed for performing the subclassification: Epstein-Barr positive, p53 and MMR proteins [43,53]. This technique is a surrogate for the molecular subtyping, offering an efficient and reasonably accurate alternative, although separating GS from CIN is difficult, this has potential prognostic and therapeutic implications [54]. Furthermore, in 2018, Birkman *et al.* [53] used immunohistochemistry and in situ hybridization to achieve biologically and clinically relevant subgroups of GAC based on the histological Laurén classification of tumors. This is applicable for both clinical diagnostics and research purposes [53]. Since then, other studies [55–59] emerged trying to overcome the difficulties in implementing the molecular classification techniques by deploying immunohistochemistry staining and in situ hybridization.

Other new research directions that uses immunohistochemistry are those that evaluate the immune components found in GAC and those that try to implement some scoring system of those components. Zhang *et al.* [60] identified that tumor infiltrating lymphocytes have prognostic value in GAC, high levels being associated with a positive prognosis. Other more complex scoring systems as ImmunoScore Signature [61] can predict recurrence and survival and may be a tool to assess if the patient may benefit from adjuvant chemotherapy in stage II and III GAC.

Regarding CRC, in 2018, Hoorn *et al.* [62] described an immunohistochemistry Mini Classifier that in combination with MIS testing can classify CRC into the main molecular subtypes. Immunohistochemistry was also used to develop a two-protein classifier derived from stromal gene-profiling that could assess the pathological response to neoadjuvant treatment in rectal cancer [63].

Galon *et al.* [64,65] combines the immunohistochemistry staining of the immune cells and image analysis methods and developed Immunoscore Colon that is a diagnostic test that predicts the risk of relapse in early-stage colon cancer and so it may have a role in deciding the need for adjuvant chemotherapy. It demonstrated the relevance of specific immune signatures in the prognostic of early-stage CRC [49].

Immunohistochemistry is taking a leading role in directing patients to targeted therapy. Currently-three targeted therapies have been approved for GAC treatment: trastuzumab - against erb-b2 receptor tyrosine kinase 2 (ERBB2), ramucirumab - against vascular endothelial growth factor receptor 2 (VEGFR2), pembrolizumab and nivolumab- both against programmed cell death protein 1 (PD-1) [3]. HER2 is overexpressed in 15–25 % of GAC. Its positivity is defined as immunohistochemistry score 3 + or by immunohistochemistry score 2 + and florescence in situ hybridization positivity [66]. For HER2 positive advanced stage GAC trastuzumab plus chemotherapy is the first line therapy [66]. PD-L1 is used as a biomarker to guide treatment in certain conditions with anti-PD-1 antibodies [66]. Its expression in GAC is evaluated by immunohistochemistry. A positive PD-L1 expression is determined by the combined positive score (CPS) $\geq$ 1 and it corelates with the therapeutic effect of ICI [67].

Other applications of molecular analysis and immunohistochemistry in GAC include identification of *FGFR2* amplifications, *EGFR* amplifications or overexpression, *MET* high expression or amplification, VEGFR overexpression and claudin 18.2 overexpression [66]. They identify the patients that may benefit from specific targeted therapies.

For CRC, the use of immunohistochemistry to evaluate the MSI is employed to evaluate the potential response for PD-1 blockade, but also to predict the outcome [68]. Even though in the last decade CRC research produces important results in order to bring CRC treatment in the era of personalized medicine, many of them did not yet rendered clinical utility [69].

5. MiRNAs as morphogens: How does a tumor get its shape [70]?

The **cancer invasion front** (CIF) configuration seems to correlate with the prognosis and survival, but how its shape impacts prognosis remains an intriguing problem. Here, many characteristics of cancer growth and progression are synchronized both at single- and collective-cell levels [70]. Vasilescu *et al.* [70] proposed a model that supports part of the hypothesis that a morphogenetic activator and inhibitor regulate tumor development and its form, underlining that the gradual change in concentration of the inhibitor is crucial. Overall miRNA expression variations and probably miRNA concentration gradients in the tumor are involved in initiation and progression of cancers [71]. Aberrant expression of miRNAs is a characteristic of the neoplasia, being involved in many biological processes through gene repression [72]. One hypothesis is that miRNAs are the potential inhibitors in determining the tumor border of epithelial malignancies and their **diffusion coefficient** is an important factor in CIF configuration [70]. Can these mechanisms be blocked as a novel therapeutic strategy?

For GAC, multiple miRNAs have been linked to its progression and prognosis. Low level of miR-34a can promote progression and reduce survival [73]. Downregulation of miR-193b was significantly associated with invasion, metastasis, and Lauren sub-type-diffuse, while high levels of miR-196-a was linked to poor differentiation [74]. Some serum miRNAs such as miR-21, miR-146a and miR-148a were linked with gastric cancer pN stage, these may be possible biomarkers to predict the presence of tumor cells in lymph nodes giving the fact that there were no differences noticed by pT stage, Lauren's sub-types, sex or age [75]. Whilst other miRNAs such as miR-17-92 cluster miRNAs are associated with the progression and chemoresistance in patients with gastric cancer treated with oxaliplatin/capecitabine (XELOX) [76]. Other non-coding RNA molecule found in GAC patients that express chemoresistance is *MACC1-AS1* [77]. It is overexpressed in FOLFOX-resistant GAC, and it binds miR-145-5p derepressing CPT1 and ACS [77]. Also, CircAKT3 is upregulated in cisplatin-resistant GAC [77].

In CRC, miRNAs regulate main pathways for development, progression, and metastasis, as well as serve as biomarkers for prognosis and diagnosis [78]. MiR-224 expression in primary CRC patients may have prognostic value and promotes CRC metastasis through the regulation of SMAD4, at least in part [79]. MiRNA dysregulation in CRC are a consequence of genetic and epigenetic changes and transcriptional regulations as suggested by recent studies [78]. A major obstacle to current cancer therapy is drug resistance to chemotherapy, but also to molecular targeted therapy [78]. A novel way to overcome it, as evidence indicate, is by targeting miRNAs [78]. The following miRNAs were identified as inducing resistance to chemotherapy: downregulated expression of miR-4802, miR-18*, miR-145, miR-17-5p and upregulation of miR-21, miR-215 and miR-625-3p; and resistance to molecular-targeted therapies: upregulated miR-31, miR-302a, miR-100, miR-125b and miR-199a-5p [78,80].

Considering their role in tumorigenesis and progression of GICs, as well as the important therapeutic implications, miRNAs may be a key step in elucidating the adequate treatment and management of GIC.

6. Back to image-based classification

The returning to routine histopathological images was aided by whole slide image (WSI) scanning and by Artificial Intelligence (AI)-based technologies that seem to dominate the headlines, especially in the field of diagnostics. Even though WSI was first introduced commercially almost 22 years ago [81] and Computer-Aided Diagnosis system began in 1980 [82], only after

the developing of the **deep learning** (DL) technology which can automatically learn the characteristics of the image through **convolutional neural networks** (CNN) and can search for similar features through a lot of training [82], the field of digital pathology changed completely. **Neural networks** (NN) were first proposed in 1944 by Warren McCullough and Walter Pitts as a model of biological neuronal networks [83]. The implications were twofold, one consisting in beginning to understand how neurons process information and the other consisting in applications in computer science. In NN paradigm, a problem is solved by the computer not by a step-by-step deterministic algorithm i.e., rule-based approach, but by learning from examples. The main advantage of using NN to solve a problem is that the programmer does not need to fully understand the intricacies of the problem or the process of finding the solution. The programmer just needs to design an appropriate NN architecture and collect loads of data for training it. The main disadvantage is precisely that the programmer might never understand these intricacies, or, in other words, the solution lacks explainability. Recently some methods were proposed that can potentially **reverse engineer DL models**; this may be a valuable tool to discover relationships and bring insights into a specific problem [84] or even identify targets for the development of new therapies [85]. Therefore, in some sense, the process of finding the solution (or answer) lacks some sort of controllability as well. The limitation of hardware resources of the last century put NN in the shadow, but with the increase of computing power (e.g., **Graphics Processing Units**-GPUs), memory capacity and availability of data in the last two decades (and especially the last decade), NN (and AI in general) became the main focus of modern computer science. Therefore, the researchers in the field are constantly developing cleverly designed NN architectures. Nowadays, NN can perform impressive tasks. In particular, CNN (and their state-of-the-art architectural versions such as **EfficientNet**, **AlexNet**, **ResNet**, **Inception** and so on), can obtain impressively high accuracy in image recognition tasks. As an example, the **CNN HoverNet** was developed for simultaneous **segmentation** and classification of nuclei on WSI [86]. In Fig. 1 we depict the two possible approaches that can link genotype to phenotype and provide new diagnostic, prognostic and predictive tools in medicine.

A far less expansive and probably more accurate solution for the stratification of GIC would be massive high-resolution acquisition of histopathological images followed by large-scale implementation of algorithms based on deep machine learning, an original and innovational technique [87]. Today, we are able by integrating these methods to obtain an exceptionally large volume of information from the usual H&E stained slides and by enabling their use from the archives of cancer institutions no prospective requirement of patients is necessary. The impact of these great advances in data acquisition and image processing have already improved other areas of imaging and histological diagnosis [88,89]. **Machine-learning** (ML) algorithms identify specific elements of the tumor to help with detection and differentiation from normal tissue that are very important in GAC [90]. In addition to morphological features, the molecular characteristics that distinguish normal from tumoral tissue will improve the ML algorithms and will bring these tools closer to the diagnostic and therapeutic seting

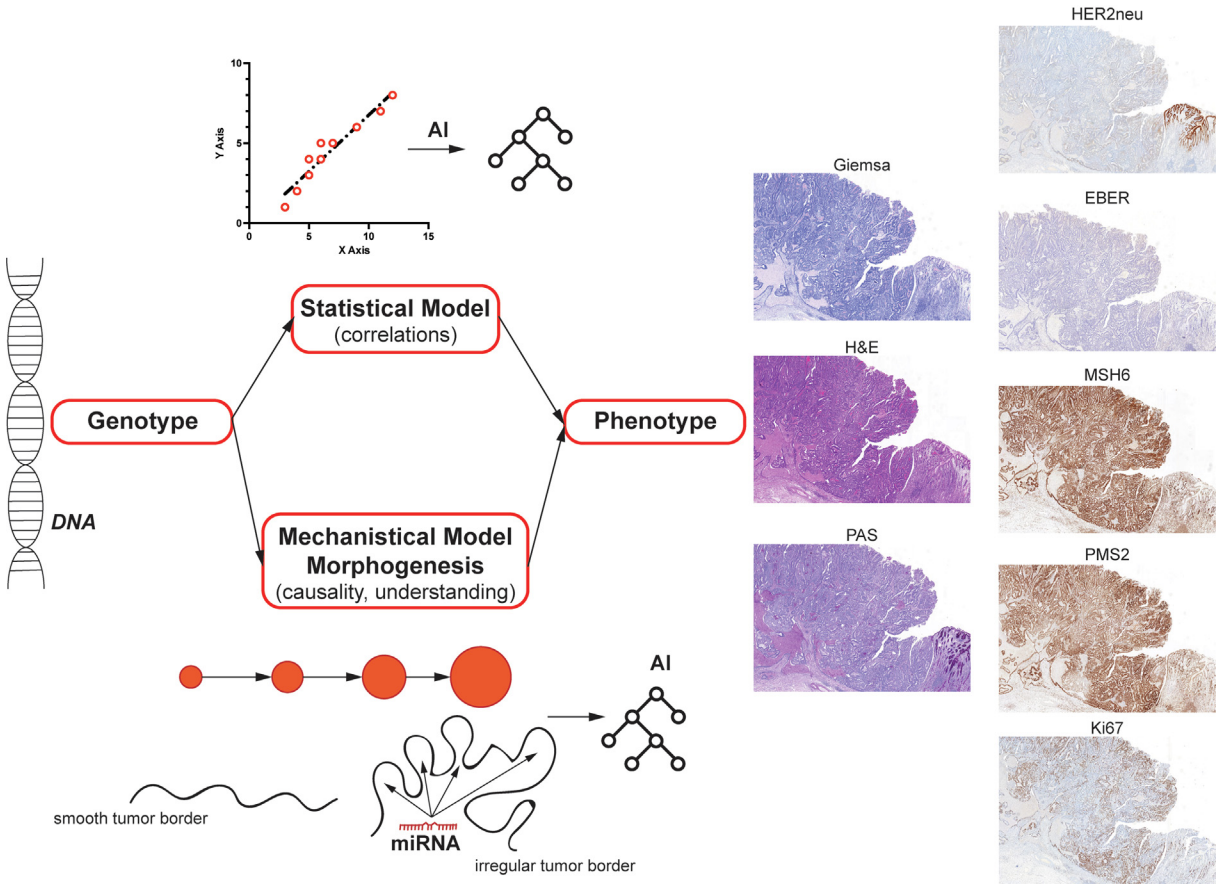

Fig. 1. The two types of approach that may render insights through the AI are the statistical model that is based on correlations and the mechanistical model, that is based on understanding and causality, a type of approach that may give knowledge on morphogenesis. Vasilescu et al. [70] hypothesized that miRNAs are the morphogenic triggers that gives the tumor it's shape. Also considering the information provided by the usually stained H&E that currently the pathologist is not using, the AI seem to extract more than meets the eye.

[90]. DL technologies and quantitative image analysis are enabling researchers to interrogate complex information with applications in immune-oncology [91] and are showing a great potential in subclassifying tumors and predicting patient's outcome [92,93]. Spatial characteristics of tumors may have a key significance in determining the prognosis [94]. They are a game changer in cancer diagnosis and treatment, leading cancer therapy into the era of personalized medicine [95].

Tumor heterogeneity is not limited to inter-patient variations, but also to intra-tumoral variations [96]. Spatial profiling of GAC matched primary and locoregional metastases leads to the discovery of principles of tumor spread, linking regional lymph nodes metastasis to the deeper subregions of the primary tumors [96]. Spatial subregions in GAC revealed clinically significant genomic, transcriptomic and phenotypical heterogeneity [96]. Already some AI image-based classification of GIC variants exists in correspondence to the histopathological subtypes: a DL model for classifying the diffuse-type GAC on WSI [97], differentiating it from other gastric pathologies, as well as glandular structure-guided classification of colorectal WSI [98]. Yoshida *et al.* [99,100] evaluated if automated image analysis software can accurately classify gastric and colorectal biopsy specimens and even if there are some limitations and requirements the results seem promising and may assist the pathologist in near-future. This is utterly important for the field of surgical pathology where it provides time-sensitive information and is a key step for patient management. Also, DL can characterize colorectal polyps and so can reduce the cognitive burden on pathologist [101]. In GIC, AI applications extend to help make the diagnosis, for prognostication, and for genetic and/or molecular testing [102].

This rapidly developing field of digital pathology and computational pathology needed specific reporting guidelines in assessing AI interventions and results. The Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 statement is one of the most widely accepted in presenting results for diagnostic studies, and an AI-specific extension to it is being developed [103]. Luo *et al.* [104] managed to join ML specialists, clinicians and statisticians to develop a set of guidelines: first part referred to a list of reporting items to be incorporated in a paper and second part, a set of steps for developing predictive tools [104].

In regard to the "molecular revolution", the question is if there is enough information on routine stained histopathological images in order to replace or complement the molecular classifications, predictions of overall survival or response to therapy and have a less expensive "image-based surrogate". This hypothesis assumes that genetic alterations in tumor cells induce functional changes that modify the cell/tissue morphology [105]. The reverse of the story is the replacement of the classical image-based pathology by molecular biology methods (i.e. sequencing) that are more specific and sensitive. In this scenario the H&E slide will be used only to define the region for extracting DNA, RNA or proteins.

7. Current image-based classifications of GIC

Cancer is a heterogeneous disease and even though the molecular classifications yields some insights, this analysis of tumor tissue is a demanding task due to the genomic information coming also from the stromal cells, immune cells, and healthy, normal ones [106]. Genetic chaos seem to be a specific feature of solid cancers, the complex pathogenetic mechanism being ruled by disordered genetic and molecular events [107]. In 2012, Yuan *et al.* [106] used the information rendered by the tissue architecture using a computational approach on standard H&E tissue sections and devised a predictor for survival in negative estrogen receptor breast cancers that integrated gene expression and image-based analysis, uncovering insights into breast tumor biology. In the same year, Cooper *et al.* [108] went a step further and hypothesized that quantitative morphometric analysis of WSI may provide mechanistic insides of disease development giving the fact that morphologic variations are often concerted to molecular aberrations.

The heterogeneity of sporadic CRC poses important difficulties in assessing prognostic and treatment response. After performing a molecular classification, Budinska *et al.* [109] evaluated the morphological CRC patterns of the molecular subtypes, subsequently Popovici *et al.* [93] shifted the approach from the molecular classification to a histopathological image-based classifier able to predict CRC molecular subtype, proving significant prognostic value [93]. The underlying molecular traits of the disease seem to be embedded in visual information of histologic specimens; this provide opportunities for integration with genomic analysis using image analysis algorithms to examine the microscopic features of WSI [110,111]. Furthermore, Sirinukunwattana *et al.* [92] provided an image-based classification of CMS of CRC using DL that seems to be a cost effective and reliable tool. This approach of image-based detection of targetable molecular alterations may be the way to understand and quantify genotype-phenotype links in cancer [105]. Deep-learning classification systems exponentially evolve into solving digital pathology problems. Until recently these systems depended on well annotated and large sets of data, but new methods have been developed that perform a fully automated demarcation of any tumor type as long as a specific staining is available for learning [112].

DL-based classifiers are used to detect MSI and EBV status from H&E slides in order to predict the response to immunotherapy in GAC, being an inexpensive biomarker after prospective validation [113]. Another application of DL was classification and mutation prediction of small cell lung cancer [114] and forecast the outcome of CRC patients [115]. Some studies emerged proving an association between microsatellite instability in CRC and routine stained WSI [116–118]. Table 1 presents a compendium of studies using digital pathology in GAC and CRC, representing the current state of knowledge in the field.

Another technique used to detect architectural changes in WSI was developed using methods from algebraic topology, as persistent homology. This method was implemented on prostate cancer WSI and clustered it through a ranked persistence vector, suggesting that it can be a robust quantification method with higher granularity than the existing semi-quantitative measures [119]. The same approach was used to assess the architectural differences between low and high grade prostate cancers with promising results [120]. Another proposed method was based on the association of features extracted by multiscale and multidimensional fractal techniques: **Haralick descriptors** and CNN for pattern recognition [121]. This method was used in CRC, breast cancer and non-Hodgkin lymohomas with promising results [121]. Because of chaos typically results in the appearance of fractals [122], various studies suggested that there is a connection between fractals and cancer [122]. Fractal geometry emerged as a useful tool in describing not only the pathological architecture of tumors, but yield insights into tumor growth mechanisms and angiogenesis that complement the modern molecular methods [123]. Fractal dimension of chromatin was proposed as a potential molecular marker for cancer progression [124–126]. During carcinogenesis and tumor progression it has been shown that the fractal dimension of the stained nuclei increased for intraepithelial lesions of the uterine cervix, anus, adenocarcinomas of the pancreas or oral squamous cell carcinoma [124]. Fractal geometry has been used to unravel the complexity of signaling networks in cancer [127] and even in the detection of colonic cancer images [128]. This method emerges as an useful tool in pathology research [129].

Table 1
A compendium of most relevant studies that used image-based classification for gastrointestinal tract cancer classification.

Year, Author, Journal	Tumor type	Aim of the study	Method	Results/conclusion
2017 Popovici V.[93], Bioinformatics	CRC	Predict the molecular subtypes based on image analysis	Deep CNN	Considerable prognostic value as molecular classification
2017 Awan R. [130], Sci Rep	CRC	Objective grading using computer algorithms	NN	Distinguishes between normal and cancer cells with 97 % accuracy and with 91 % accuracy between normal cells, low- and high-grade cancer.
2018 Bychkov D [115], Sci Rep	CRC	Foresees outcome, without any histopathological classification	Recurrent NN	DL can obtain more prognostic data than an experienced human observer
2019 Geessink OGF [131], Cell Oncol (Dordr)	RC	Computer-aided quantification of intra tumoral stroma in RC WSI	NN	DL-based technology may be a significant aid to pathologists in routine diagnostics
2019 Kather JN [132], PLoS Med	CRC	Extraction of prognostic markers directly from H&E–stained tissue slides	Deep CNN	CNN can predict prognosis directly from histopathological images
2019 Shapcott M [133], Bioeng. Biotechnol	CRC	Identify prognostic features	DL CNN	Tissue morphology relates with a range of clinical features as cell identification algorithm uncovers them
2019 Kather JN [117], Nat. Med.	GIC	Predict MSI from digital tissue slides	Deep residual learning	May identify the subset of patients that benefit from immunotherapy
2020 Kather JN [105], Nature Cancer	CRC, GAC, panc. cancer	Predict molecular alterations from digital tissue slides	NN	DL has the potential to infer mutations, molecular subtypes, gene expression patterns and biomarkers from digital tissue slides
2020 Skrede OJ [134], Lancet	CRC	Develop a prognostic biomarker after primary CRC resection by analyzing digital H&E tissue slides	CNN	Stratification of CRC stage II and III patients into prognostic groups
2020 Fu Y [135], Nature Cancer	CRC, GAC	Predict genomic alterations based on digital tissue slides; Cancer classification	DL	Infer genomic alterations, mutations, immune infiltration and gene expression profiling
2020 Sirinukunwattana K.[92], Gut	CRC	Image-based approach to predict CRC molecular subtypes from standard H&E sections	NN with domain adversarial learning	CRC molecular subtypes can be predicted from digital H&E tissue slides
2020 Echle A [116], Gastroenterology	CRC	Identify mismatch-repair deficiency (dMMR) on H&E slides	Shufflenet DL	96 % accuracy in predicting dMMR
2021 Bilal M [136], Lancet Digit Health	CRC	Assess the status of major molecular pathways and mutations on H&E slides	DL framework involving 3 separate CNN	Identify patients for targeted therapies faster and with lower costs

* CNN– convolutional neural networks, NN– neural networks, DL- deep learning, CRC- colorectal cancer, GAC- gastric adenocarcinoma, GIC- gastrointestinal cancer, MSI- microsatellite instability, panc. cancer – pancreatic cancer.

8. The way to computational pathology

Artificial Intelligence (AI) encompasses the techniques for a machine to replicate or to overrun human intelligence, mainly in the rapid processing capabilities of complex data. In a traditional rule-based approach to AI, the programmer overtly encodes the know-how coming from the pathologist. In contrast, ML uses statistical methods to discover essential patterns from a set of training data without explicit *a priori* instruction from the pathologist. DL is a novel ML tool, inspired from neurobiology networks, used to represent data employing multiple levels of simple but nonlinear modules. While there are numerous applications and research projects employing AI in image analysis in medical field, the vast majority are either relying on DL to solve classification problems, or using image pre-processing applied globally, like thresholding and masking, to extract features, which are analyzed using statistical methodologies to facilitate conclusions [102,137–139]. Both strategies offer good results, but with some disadvantages. DL alone could offer high accuracy predictions, but it lacks explainibility, which is crucial in the scientific research. A CNN will learn how to identify the different elements that constitutes a tissue and then to recognize various pattern that makes the tissue fall into a particular category. In the end, a CNN will be at most capable of labeling the image and highlighting some elements in the image in terms of importance for the decision (i.e. attention maps). As an example, let us consider images of fractals. CNNs will definitely be able to recognize different fractal objects, but our goal is to find a way to extract information about the generating rules of the fractal in question. For instance, in this illustrative scenario, we would be interested in defining and estimating the fractal dimension of a fingerprint. This invariant would be a link between the 'genotype' and the 'phenotype' of the entities in this scenario.

One approach is to attach numerical invariants to the histopathological features in order to identify the type of tissue or a particular class, especially regarding cancer. Ideally, these invariants should belong to a continuous domain, should be as few as possible, and should correlate with the specific class i.e., each class should be associated with a well-defined subdomain of values for the invariants. The fewer the invariants, the easiest is to find a biological interpretation for them.

Considering this, we can emphasize again that the best way a NN to produce values that can be called numerical invariants is by using a bottleneck architecture called autoencoder. For example, the size/shape/color of individual nuclei is indeed important, but there are straight forward procedures dedicated to analyzing these features. Thus, we will ignore these aspects and concentrate on the more interesting problem of their configuration as points in the real plane.

Potential methods:

M1: One method for analyzing the configuration of nuclei called local structure correlation diagram (LSCD) is described by Tanase *et al.* [140].

M2: Second method comes from algebraic topology and is called persistent homology. It consists in calculating homological invariants (**Betti numbers**) that persists across different scales [119].

M3: Third method originates in fractal geometry and consists in computing the **Fourier coefficients** of the histogram associated with the local fractal dimension map [141].

M4: Forth method derives from the perspective of dynamical systems, a version of **Lyapunov exponent(s)** can be computed for an image [142].

By aggregating the data computed by such methods, we aim to obtain a precise model of the shift between classes in terms of configurations of histopathological features and understand the image as a fingerprint of a dynamic process.

Another related research topic developed recently is that of spatial transcriptomics. This technologies capture coding transcripts expression across biological tissue space [143]. There are many technologies available: Spatial Transcriptomics, Slide-seqV2, MERFISH, GeoMx™, DBiT-seq and Stereo-seq, some of them allow, in some cases, even for subcellular detection of RNA [143]. Unfortunately, even these novel methods often rely on defining the tumor cells and stromal cells before performing the analysis, hence image-based segmentation tools are highly necessary.

9. Concluding remarks

There are several challenges that need to be addressed in the field of AI based diagnosis in pathology.

First, is the implementation of AI in daily practice. There are many barriers in what concerns the WSI scanners, and the massive acquisition of images needed to train the CNN and their proper storage. Nowadays there are cloud-based telepathology systems tested in order to evaluate their effectiveness [144], but questions still remain regarding cyber security, access to large data, and ethics. Most probably this challenge will be addressed in the near future, for example it is already widely accepted the advantage of archiving pathology slides in a digital format, in this way they are immediately available for later comparison and analysis (especially in case of recurrences). The digital format also permits the transfer of slides to renowned specialists whose expert-opinion can be consulted for definitive diagnosis. Additionally, the digitalization of slides solves the problem of a physical storage place for the slides.

Second, it is necessary to evaluate their performance of these tools in real-life clinical setting, prospective studies need to be conducted to show the true benefit of AI [87]. The prospective studies need to validate these very new and interesting findings in a real-life clinical setting, where for example core biopsies contain very limited amount of tumor tissue or many cancers are treated with neoadjuvant chemotherapy and the morphology is dominated by regressive changes. Moreover, there needs to be a standardized method of reporting the results. Only in this way the results can be reproduced between groups and methods can be implemented in the clinical setting.

Another issue is related to explainability – can we use a diagnostic tool that we cannot understand? Because of this point many groups perform an explainable pathology, in this way a ready to use tool can be developed. There is also the reverse of this question, we can develop a tool that lacks explainability to obtain a more in depth understanding of cancer biology and by reverse engineering to discover unknown mechanisms.

Finally, there is an important question regarding medical ethics – can we accept to be diagnosed by an algorithm, even an explainable one, instead of a medical doctor? Who will be to blame for the diagnostic errors?

We believe that DL through CNN may be the key into understanding the relationship between genotype and phenotype. If the method is validated through clinical studies, this may open new opportunities for adequate pathological classifications, survival predictions and most important, a tailored treatment. Even though it seems that analyzing the WSI by AI may render enough information on the genotype it is early to draw definitive answers. Nonetheless a *de novo* and unsupervised image-based classification of GIC could be an alternative approach. This could shed new light

on our understanding of these malignancies, but this will have to be backed up by in depth molecular characterization of the researched tumors and it would need an impressive cohort size.

We also consider that by having a better understanding of the molecular mechanisms that shape the morphology of the tumors could rapidly aid the development of digital pathology. As outlined in this review and our previous manuscript [70], miRNAs that could play a role as morphogenic inhibitors, could be the missing link between genotype and phenotype. Hence, a combination between image analyses methods and the molecular methods is the ideal strategy. Furthermore, the AI can be used to additionally integrate miRNA expression patterns in the tumor border obtained by In Situ Hybridizations (ISH). Such integrative analyses will extract more information from tumor slides. Revealing such missing mechanistic elements could be sometimes the much quicker way compared to scanning tens of thousands of slides and not knowing what you are looking for.

Funding

This work was supported by a grant of the Ministry of Research, Innovation and Digitization, CNCS - UEFISCDI, project number PN-III-P4-PCE-2021-1068, within PNCDI III (to C.E.M., M.T., T.E.M., S.T., V.H., M.P.D., C.V.) and by Berlin Institute of Health, Junior Clinician Scientist Program (to M.P.D.), DKTK Berlin Young Investigator Grant 2022 (to M.P.D.), Berliner Krebsgesellschaft (DRFF202204 to M.P.D.), National Cancer Institute at the U.S. National Institutes of Health (1R01CA182905-01 and 1R01CA222007-01A1 to G.A.C.), National Institute of General Medical Sciences at the U.S. National Institutes of Health (1R01GM122775-01 to G.A.C.), and U.S. Department of Defense (W81XWH2110030 to G.A.C.).

CRediT authorship contribution statement

Corina-Elena Minciuna: Conceptualization, Investigation, Data curation, Writing – original draft. **Mihai Tanase:** Resources, Data curation, Writing – review & editing. **Teodora Ecaterina Manuc:** Resources, Writing – review & editing. **Stefan Tudor:** Resources, Writing – review & editing. **Vlad Herlea:** Resources, Writing – review & editing. **Mihnea P. Dragomir:** Conceptualization, Visualization, Writing – original draft, Supervision, Funding acquisition. **George A. Calin:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Catalin Vasilescu:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Kuntz, S., *et al.* Gastrointestinal cancer classification and prognostication from histology using deep learning: Systematic review. *European journal of cancer (Oxford, England : 1990)* 155, 200-215 (2021).
[2] Van Cutsem E, Sagaert X, Topal B, Haustermans K, Prenen H. Gastric cancer. Lancet 2016;388:2654–64.
[3] Ho SWT, Tan P. Dissection of gastric cancer heterogeneity for precision oncology. Cancer Sci 2019;110:3405–14.
[4] Uggeri F et al. Is there a role for treatment-oriented surgery in liver metastases from gastric cancer? World J Clin Oncol 2020;11:477–94.
[5] Glimelius B et al. Randomized comparison between chemotherapy plus best supportive care with best supportive care in advanced gastric cancer. Ann Oncol 1997;8:163–8.
[6] Song A et al. Surgical resection for hepatic metastasis from gastric cancer: a multi- institution study. Oncotarget 2017;8:71147–53.
[7] K. Zhang L. Chen Chinese consensus on the diagnosis and treatment of gastric cancer with liver metastases Therapeutic advances in medical oncology 12 2020 1758835920904803.
[8] Newton AD, Datta J, Loaiza-Bonilla A, Karakousis GC, Roses RE. Neoadjuvant therapy for gastric cancer: current evidence and future directions. J Gastroint Oncol 2015;6:534–43.
[9] Smyth EC et al. Gastric cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up[†]. Ann Oncol 2016;27:v38–49.
[10] Ajani JA et al. Gastric cancer, version 2.2022, NCCN clinical practice guidelines in oncology. J Natl Comprehen Cancer Network 2022;20:167–92.
[11] Dekker E, Tanis PJ, Vleugels JLA, Kasi PM, Wallace MB. Colorectal cancer. Lancet (London, England) 2019;394:1467–80.
[12] Kuipers EJ et al. Colorectal cancer. Nat Rev Dis Primers 2015;1:15065.
[13] Ralli M et al. Immunotherapy in the treatment of metastatic melanoma: current knowledge and future directions. J Immunol Res 2020;2020:9235638.
[14] Lomnicki S, Setia N, Catenacci DVT. Tackling diversity within diversity. Ann Oncol 2020;31:970–2.
[15] Aoyama T et al. Clinical impact of tumor location on the colon cancer survival and recurrence: analyses of pooled data from three large phase III randomized clinical trials. Cancer Med 2017;6:2523–30.
[16] Kafatos G et al. Impact of biomarkers and primary tumor location on the metastatic colorectal cancer first-line treatment landscape in five European countries. Future Oncol (London, England) 2021;17:1495–505.
[17] Kumar R, Harilal S, Carradori S, Mathew B. A comprehensive overview of colon cancer- a grim reaper of the 21st century. Curr Med Chem 2021;28:2657–96.
[18] Zarkavelis G et al. Current and future biomarkers in colorectal cancer. Ann Gastroenterol 2017;30:613–21.
[19] Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. Nat Rev Clin Oncol 2020;17:11–32.
[20] Fuchs CS et al. Pembrolizumab (pembro) vs paclitaxel (PTX) for previously treated advanced gastric or gastroesophageal junction (G/GEJ) cancer: Phase 3 KEYNOTE-061 trial. J Clin Oncol 2018;36:4062.
[21] Weidle UH, Birzele F, Nopora A. microRNAs promoting growth of gastric cancer xenografts and correlation to clinical prognosis. Cancer Genom Proteom 2021;18:1–15.
[22] Hossain MS et al. Colorectal cancer: a review of carcinogenesis, global epidemiology, current challenges, risk factors, preventive and treatment strategies. Cancers 2022;14.
[23] Benson, A.B., *et al.* Colon Cancer, Version 2.2021, NCCN Clinical Practice Guidelines in Oncology. *Journal of the National Comprehensive Cancer Network* 19, 329-359 (2021).
[24] Le DT et al. PD-1 blockade in tumors with mismatch-repair deficiency. N Engl J Med 2015;372:2509–20.
[25] LAURÉN, P. THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. *Acta Pathologica Microbiologica Scandinavica* 64, 31-49 (1965).
[26] Nagtegaal ID et al. The 2019 WHO classification of tumours of the digestive system. Histopathology 2020;76:182–8.
[27] Yasuda K, Adachi Y, Shiraishi N, Maeo S, Kitano S. Papillary adenocarcinoma of the stomach. Gastric Cancer 2000;3:33–8.
[28] Arai T et al. Frequent microsatellite instability in papillary and solid-type, poorly differentiated adenocarcinomas of the stomach. Gastric Cancer 2013;16:505–12.
[29] Kakiuchi M et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. Nat Genet 2014;46:583–7.
[30] Lee JE et al. Clinicopathologic and genomic characteristics of mucinous gastric adenocarcinoma. Gastric Cancer 2022;25:697–711.
[31] Machado JC et al. E-cadherin gene mutations provide a genetic basis for the phenotypic divergence of mixed gastric carcinomas. Laborat Investig J Tech Methods Pathol 1999;79:459–65.
[32] Hugen N, van Beek JJ, de Wilt JH, Nagtegaal ID. Insight into mucinous colorectal carcinoma: clues from etiology. Ann Surg Oncol 2014;21:2963–70.
[33] Graur F et al. Analysis of the MLH1, MLH2, MLH6, PMS2 genes and their correlations with clinical data in rectal mucinous adenocarcinoma. Ann Ital Chir 2022;93:188–94.
[34] Shia J et al. Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology-molecular associations: clinical and biological implications. Mod Pathol 2017;30:599–609.
[35] Liu, X., Huang, L., Liu, M. & Wang, Z. The Molecular Associations of Signet-Ring Cell Carcinoma in Colorectum: Meta-Analysis and System Review. *Medicina (Kaunas, Lithuania)* 58(2022).
[36] Thirunavukarasu P et al. Medullary carcinoma of the large intestine: a population based analysis. Int J Oncol 2010;37:901–7.
[37] Kang X et al. Serrated neoplasia in the colorectum: gut microbiota and molecular pathways. Gut Microbes 2021;13:1–12.
[38] Gonzalez RS et al. Adenoma-like adenocarcinoma: a subtype of colorectal carcinoma with good prognosis, deceptive appearance on biopsy and frequent KRAS mutation. Histopathology 2016;68:183–90.
[39] Lee HJ et al. Colorectal micropapillary carcinomas are associated with poor prognosis and enriched in markers of stem cells. Mod Pathol 2013;26:1123–31.
[40] Khan AH et al. Presentation, treatment, and prognosis of colorectal adenosquamous carcinoma: a contemporary analysis of the surveillance, epidemiology, and end results database. Am J Surg 2022;223:957–62.

[41] Agaimy A et al. SWI/SNF complex-deficient undifferentiated/rhabdoid carcinomas of the gastrointestinal tract: a series of 13 cases highlighting mutually exclusive loss of SMARCA4 and SMARCA2 and frequent co-inactivation of SMARCB1 and SMARCA2. Am J Surg Pathol 2016;40:544–53.
[42] Bass AJ et al. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202–9.
[43] Cisło M et al. Distinct molecular subtypes of gastric cancer: from Laurén to molecular pathology. Oncotarget 2018;9:19427–42.
[44] Guinney J et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350–6.
[45] Valenzuela G et al. Consensus molecular subtypes of colorectal cancer in clinical practice: a translational approach. World J Clin Oncol 2021;12:1000–8.
[46] Singh MP, Rai S, Pandey A, Singh NK, Srivastava S. Molecular subtypes of colorectal cancer: an emerging therapeutic opportunity for personalized medicine. Genes Dis 2021;8:133–45.
[47] Simoneaux R. The four colorectal cancer consensus molecular subtypes. Oncol Times 2018;40:10–1.
[48] Rodriguez-Salas N et al. Clinical relevance of colorectal cancer molecular subtypes. Crit Rev Oncol Hematol 2017;109:9–19.
[49] Dienstmann R et al. Consensus molecular subtypes and the evolution of precision medicine in colorectal cancer. Nat Rev Cancer 2017;17:79–92.
[50] Jass JR. Classification of colorectal cancer based on correlation of clinical, morphological and molecular features. Histopathology 2007;50:113–30.
[51] Currais P, Rosa I, Claro I. Colorectal cancer carcinogenesis: from bench to bedside. World J Gastrointest Oncol 2022;14:654–63.
[52] Joanito I et al. Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. Nat Genet 2022;54:963–75.
[53] Birkman E-M et al. Gastric cancer: immunohistochemical classification of molecular subtypes and their association with clinicopathological characteristics. Virchows Arch 2018;472:369–82.
[54] Gonzalez RS, Messing S, Tu X, McMahon LA, Whitney-Miller CL. Immunohistochemistry as a surrogate for molecular subtyping of gastric adenocarcinoma. Hum Pathol 2016;56:16–21.
[55] Zhao C et al. Protein expression-based classification of gastric cancer by immunohistochemistry of tissue microarray. PLoS ONE 2020;15:e0238836.
[56] Tsai JH et al. An integrative morphomolecular classification system of gastric carcinoma with distinct clinical outcomes. Am J Surg Pathol 2020;44:1017–30.
[57] Díaz Del Arco C et al. Immunohistochemical classification of gastric cancer based on new molecular biomarkers: a potential predictor of survival. Virchows Arch 2018;473:687–95.
[58] Pretzsch E et al. Molecular subtyping of gastric cancer according to ACRG using immunohistochemistry – Correlation with clinical parameters. Pathol Res Pract 2022;231:153797.
[59] Di Pinto F et al. Are immunohistochemical markers useful in phenotypic gastric cancer classification? Oncology 2020;98:566–74.
[60] Zhang D et al. Scoring System for Tumor-Infiltrating Lymphocytes and Its Prognostic Value for Gastric Cancer. Front Immunol 2019;10:71.
[61] Jiang Y et al. Immuno score signature: a prognostic and predictive tool in gastric cancer. Ann Surg 2018;267:504–13.
[62] Ten Hoorn S, Trinh A, de Jong J, Koens L, Vermeulen L. Classification of colorectal cancer in molecular subtypes by immunohistochemistry. Methods Mol Biol (Clifton NJ) 2018;1765:179–91.
[63] Gonçalves-Ribeiro S et al. Prediction of pathological response to neoadjuvant treatment in rectal cancer with a two-protein immunohistochemical score derived from stromal gene-profiling. Ann Oncol 2017;28:2160–8.
[64] Galon J et al. Immunoscore clinical utility to identify good prognostic colon cancer stage II patients with high-risk clinico-pathological features for whom adjuvant treatment may be avoided. J Clin Oncol 2019;37:487.
[65] Galon, J., *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science (New York, N.Y.)* 313, 1960-1964 (2006).
[66] Nakamura Y, Kawazoe A, Lordick F, Janjigian YY, Shitara K. Biomarker-targeted therapies for advanced-stage gastric and gastro-oesophageal junction cancers: an emerging paradigm. Nat Rev Clin Oncol 2021;18:473–87.
[67] Park Y et al. PD-L1 testing in gastric cancer by the combined positive score of the 22C3 PharmDx and SP263 assay with clinically relevant cut-offs. Cancer Res Treat 2020;52:661–70.
[68] Dudley JC, Lin MT, Le DT, Eshleman JR. Microsatellite instability as a biomarker for PD-1 blockade. Clin Cancer Res 2016;22:813–20.
[69] Bărbălan A et al. Immunohistochemistry predictive markers for primary colorectal cancer tumors: where are we and where are we going? Rom J Morphol Embryol 2018;59:29–42.
[70] Vasilescu C et al. How does a tumor get its shape? MicroRNAs act as morphogens at the cancer invasion front. Non-coding RNA 2020;6.
[71] Dragomir MP, Knutsen E, Calin GA. Classical and noncanonical functions of miRNAs in cancers. Trends Genet TIG 2022;38:379–94.
[72] Shin VY, Chu KM. MiRNA as potential biomarkers and therapeutic targets for gastric cancer. World J Gastroenterol 2014;20:10432–9.
[73] Li Z, Liu ZM, Xu BH. A meta-analysis of the effect of microRNA-34a on the progression and prognosis of gastric cancer. Eur Rev Med Pharmacol Sci 2018;22:8281–7.
[74] Mu YP et al. Association of miR-193b down-regulation and miR-196a up-regulation with clinicopathological features and prognosis in gastric cancer. Asian Pacific J Cancer Prevent 2014;15:8893–900.
[75] Kim SY et al. Validation of circulating miRNA biomarkers for predicting lymph node metastasis in gastric cancer. J Mol Diagn 2013;15:661–9.
[76] Fan B et al. miR-17-92 cluster is connected with disease progression and oxaliplatin/capecitabine chemotherapy efficacy in advanced gastric cancer patients: a preliminary study. Medicine 2018;97:e12007.
[77] Chen B et al. Targeting non-coding RNAs to overcome cancer therapy resistance. Sig Trans Targeted Ther 2022;7:121.
[78] Huang X et al. Dissecting miRNA signature in colorectal cancer progression and metastasis. Cancer Lett 2021;501:66–82.
[79] Ling H et al. The clinical and biological significance of MIR-224 expression in colorectal cancer metastasis. Gut 2016;65:977–89.
[80] Dragomir MP, Kopetz S, Ajani JA, Calin GA. Non-coding RNAs in GI cancers: from cancer hallmarks to clinical utility. Gut 2020;69:748–63.
[81] Pantanowitz L et al. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. J Pathol Inform 2018;9:40.
[82] Ai S et al. A state-of-the-art review for gastric histopathology image analysis approaches and future development. Biomed Res Int 2021;2021:6671417.
[83] Benko, A. & Lányi, C.S. History of artificial intelligence. in *Encyclopedia of Information Science and Technology, Second Edition* 1759-1762 (IGI Global, 2009).
[84] Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. J Pathol Inform 2016;7:29.
[85] Echle A et al. Deep learning in cancer pathology: a new generation of clinical biomarkers. Br J Cancer 2021;124:686–96.
[86] Graham S et al. Hover-net: simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal 2019;58:101563.
[87] van der Laak J, Litjens G, Ciompi F. Deep learning in histopathology: the path to the clinic. Nat Med 2021;27:775–84.
[88] Vasilescu C et al. Morphometrical differences between resectable and non-resectable pancreatic cancer: a fractal analysis. Hepatogastroenterology 2012;59:284–8.
[89] Watanabe H et al. Quantification of structural heterogeneity using fractal analysis of contrast-enhanced CT image to predict survival in gastric cancer patients. Dig Dis Sci 2021;66:2069–74.
[90] Toh J et al. Profiling of gastric cancer cell-surface markers to achieve tumour-normal discrimination. BMJ Open Gastroenterol 2020;7.
[91] Shakya R, Nguyen TH, Waterhouse N, Khanna R. Immune contexture analysis in immuno-oncology: applications and challenges of multiplex fluorescent immunohistochemistry. Clin Transl Immunol 2020;9:e1183.
[92] Sirinukunwattana K et al. Image-based consensus molecular subtype (imCMS) classification of colorectal cancer using deep learning. Gut 2021;70:544–54.
[93] Popovici V, Budinská E, Dušek L, Kozubek M, Bosman F. Image-based surrogate biomarkers for molecular subtypes of colorectal cancer. Bioinformatics (Oxford, England) 2017;33:2002–9.
[94] Tsujikawa T et al. Prognostic significance of spatial immune profiles in human solid cancers. Cancer Sci 2020;111:3426–34.
[95] Levy-Jurgenson A, Tekpli X, Kristensen VN, Yakhini Z. Spatial transcriptomics inferred from pathology whole-slide images links tumor heterogeneity to survival in breast and lung cancer. Sci Rep 2020;10:18802.
[96] Sundar, R., *et al.* Spatial profiling of gastric cancer patient-matched primary and locoregional metastases reveals principles of tumour dissemination. *Gut*, gutjnl-2020-320805 (2020).
[97] Kanavati F, Tsuneki M. A deep learning model for gastric diffuse-type adenocarcinoma classification in whole slide images. Sci Rep 2021;11:20486.
[98] Awan R, Al-ma'adeed S, Alsaady R, Bouridane A. Glandular structure-guided classification of microscopic colorectal images using deep learning. Comput Electr Eng 2019;85:106450.
[99] Yoshida H et al. Automated histological classification of whole-slide images of gastric biopsy specimens. Gastric Cancer 2018;21:249–57.
[100] Yoshida H et al. Automated histological classification of whole slide images of colorectal biopsy specimens. Oncotarget 2017;8:90719–29.
[101] Korbar B et al. Deep learning for classification of colorectal polyps on whole-slide images. J Pathol Inform 2017;8:30.
[102] Calderaro J, Kather JN. Artificial intelligence-based pathology for gastrointestinal and hepatobiliary cancers. Gut 2021;70:1183–93.
[103] Sounderajah V et al. Developing specific reporting guidelines for diagnostic accuracy studies assessing AI interventions: the STARD-AI Steering Group. Nat Med 2020;26:807–8.
[104] Luo W et al. Guidelines for developing and reporting machine learning predictive models in biomedical research: a multidisciplinary view. J Med Int Res 2016;18:e323.
[105] Kather JN et al. Pan-cancer image-based detection of clinically actionable genetic alterations. Nat Cancer 2020;1:789–99.
[106] Yuan, Y., *et al.* Quantitative Image Analysis of Cellular Heterogeneity in Breast Tumors Complements Genomic Profiling. *Science Translational Medicine* 4, 157ra143-157ra143 (2012).
[107] Calin GA, Vasilescu C, Negrini M, Barbanti-Brodano G. Genetic chaos and antichaos in human cancers. Med Hypotheses 2003;60:258–62.

[108] Cooper LA et al. Integrated morphologic analysis for the identification and characterization of disease subtypes. J Am Med Inform Assoc 2012;19:317–23.
[109] Budinska E et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. J Pathol 2013;231:63–76.
[110] Cooper, L.A., *et al.* Novel genotype-phenotype associations in human cancers enabled by advanced molecular platforms and computational analysis of whole slide images. *Laboratory investigation; a journal of technical methods and pathology* 95, 366-376 (2015).
[111] Lafarge MW, Koelzer VH. Towards computationally efficient prediction of molecular signatures from routine histology images. Lancet Digital Health 2021;3:e752–3.
[112] Su A et al. A deep learning model for molecular label transfer that enables cancer cell identification from histopathology images. NPJ Precis Oncol 2022;6:14.
[113] Muti HS et al. Development and validation of deep learning classifiers to detect Epstein-Barr virus and microsatellite instability status in gastric cancer: a retrospective multicentre cohort study. Lancet Digital Health 2021;3:e654–64.
[114] Coudray N et al. Classification and mutation prediction from non-small cell lung cancer histopathology images using deep learning. Nat Med 2018;24:1559–67.
[115] Bychkov D et al. Deep learning based tissue analysis predicts outcome in colorectal cancer. Sci Rep 2018;8:3395.
[116] Echle A et al. Clinical-grade detection of microsatellite instability in colorectal tumors by deep learning. Gastroenterology 2020;159:1406–1416.e1411.
[117] Kather JN et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25:1054–6.
[118] Park JH et al. Artificial intelligence for predicting microsatellite instability based on tumor histomorphology: a systematic review. Int J Mol Sci 2022;23.
[119] Lawson P, Sholl AB, Brown JQ, Fasy BT, Wenk C. Persistent homology for the quantitative evaluation of architectural features in prostate cancer histology. Sci Rep 2019;9:1139.
[120] Lawson, P., Schupbach, J., Fasy, B. & Sheppard, J. *Persistent homology for the automatic classification of prostate cancer aggressiveness in histopathology images*, (2019).
[121] Candelero, D., Freire, G., Zanchetta do Nascimento, M., Rozendo, G. & Neves, L. *Selection of CNN, Haralick and Fractal Features Based on Evolutionary Algorithms for Classification of Histological Images*, (2020).
[122] Sokolov I, Dokukin ME. Fractal analysis of cancer cell surface. Methods Mol Biol (Clifton NJ) 2017;1530:229–45.
[123] Baish JW, Jain RK. Fractals and cancer. Cancer Res 2000;60:3683–8.
[124] Metze K. Fractal dimension of chromatin: potential molecular diagnostic applications for cancer prognosis. Expert Rev Mol Diagn 2013;13:719–35.
[125] Metze K. Fractal dimension of chromatin and cancer prognosis. Epigenomics 2010;2:601–4.
[126] Metze K, Adam R, Florindo JB. The fractal dimension of chromatin – a potential molecular marker for carcinogenesis, tumor progression and prognosis. Expert Rev Mol Diagn 2019;19:299–312.
[127] Garland J. Unravelling the complexity of signalling networks in cancer: a review of the increasing role for computational modelling. Crit Rev Oncol Hematol 2017;117:73–113.
[128] Esgiar AN, Naguib RN, Sharif BS, Bennett MK, Murray A. Fractal analysis in the detection of colonic cancer images. IEEE Trans Inform Technol 2002;6:54–8.
[129] Bianciardi G. Differential diagnosis: shape and function, fractal tools in the pathology lab. Nonlinear Dynam Psychol Life Sci 2015;19:437–64.
[130] Awan R et al. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. Sci Rep 2017;7:16852.
[131] Geessink OGF et al. Computer aided quantification of intratumoral stroma yields an independent prognosticator in rectal cancer. Cellular oncology (Dordrecht) 2019;42:331–41.
[132] Kather JN et al. Predicting survival from colorectal cancer histology slides using deep learning: a retrospective multicenter study. PLoS Med 2019;16:e1002730.
[133] Shapcott M, Hewitt KJ, Rajpoot N. Deep learning with sampling in colon cancer histology. Front Bioeng Biotechnol 2019;7:52.
[134] Skrede O-J et al. Deep learning for prediction of colorectal cancer outcome: a discovery and validation study. Lancet 2020;395:350–60.
[135] Fu Y et al. Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. Nat Cancer 2020;1:800–10.
[136] Bilal M et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit Health 2021;3. e763–e772.
[137] Chen ZH et al. Artificial intelligence for assisting cancer diagnosis and treatment in the era of precision medicine. Cancer Commun (Lond Engl) 2021;41:1100–15.
[138] Yoshida H, Kiyuna T. Requirements for implementation of artificial intelligence in the practice of gastrointestinal pathology. World J Gastroenterol 2021;27:2818–33.
[139] Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. J Ambient Intell Hum Comput 2022;1–28.
[140] Tanase M, Waliszewski P. On complexity and homogeneity measures in predicting biological aggressiveness of prostate cancer; implication of the cellular automata model of tumor growth. J Surg Oncol 2015;112:791–801.
[141] Olteanu, M. & Tanase, M. An algorithm for the analysis of fractal-like structures and miscellaneous applications.
[142] Khodadadi H, Sedigh A, Ataei M, Jahedmotlagh M-R. Applying a modified version of Lyapunov exponent for cancer diagnosis in biomedical images: the case of breast mammograms. Multidimension Syst Signal Process 2018;29.
[143] Noel T, Wang QS, Greka A, Marshall JL. Principles of spatial transcriptomics analysis: a practical walk-through in kidney tissue. Front Physiol 2022;12:809346.
[144] He X et al. Effectiveness of a cloud-based telepathology system in china: large-sample observational study. J Med Int Res 2021;23:e23799.

Glossary

Betti numbers: the ranks of the homology groups capturing information about holes of different dimensions in a topological space.
Cancer invasion front (CIF): the border between an infiltrating neoplasia and normal tissue.
Convolutional Neural Networks (CNNs): a class of neural networks that use convolution filters to analyze images.
Convolutional Neural Network Hover-Net: a branched CNN which within a single network can perform segmentation and classification of nuclei.
Deep learning (DL): the algorithm(s) used to modify the parameters of a neural network model based on the response of the network to training input data in order to increase its accuracy (i.e. it is a learning method for neural networks).
Diffusion coefficient: the quantity of a given substance that diffuses across a unit of space in one second under the influence of a gradient of one unit.
EfficientNet, AlexNet, ResNet, Inception: examples of special state-of the-art CNN architectures.
Fourier coefficients: complex numbers that form the frequency spectrum of a signal.
Graphic Processing Units (GPUs): specialized electronic circuit designed for parallel processing of information.
Haralick descriptors: a texture descriptor composed of several statistical features.
Lyapunov exponent(s): a quantity that characterizes the rate of separation of infinitesimally close trajectories that reflects chaotic behavior of a dynamical system.
microRNA (miRNA): small non-coding RNA molecule that inhibits messenger RNA translation at a post-transcriptional level.
Machine learning (ML): a class of algorithms capable of learning from data to improve performance of a given task.
Molecular classification/molecular subtypes in oncology: the use of molecular big data (transcriptomics, genomics, ***epi***-genomics) to sub-classify a cancer type.
Neural networks (NN): a class of algorithms based on model inspired from the study of the brain neurons that is capable of finding patterns in the input data in order to generate a desired output.
Reverse engineer deep learning models: understanding how a deep learning model produces its output for a given input.
Segmentation (in image analysis and processing): the process of dividing an image into disjointed areas so that each area represents a different pattern/structure/object (for example tumor tissue, stroma, cells).