6-1-2017

Molecular landscape and sub-classification of gastrointestinal cancers: A review of literature

Bita Fakhri
Washington University School of Medicine in St. Louis
Kian-Huat Lim
Washington University School of Medicine in St. Louis

Review Article

Molecular landscape and sub-classification of gastrointestinal cancers: a review of literature

Bita Fakhri, Kian-Huat Lim

Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO 63110, USA

Contributions: (I) Conception and design: All authors; (II) Administrative support: None; (III) Provision of study material or patients: None; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Kian-Huat Lim. Division of Oncology, Department of Internal Medicine, Washington University School of Medicine, Saint Louis, MO 63110, USA. Email: kian-huat.lim@wustl.edu.

Abstract: The historical approach of diagnosing cancer types based entirely on anatomic origin and histologic features, and the "one-size-fit-all" therapeutic approach, are inadequate in modern cancer treatment. From decades of research we now know that cancer is a highly heterogeneous disease driven by complex genetic or epigenetic alterations. The advent of various high throughput molecular tools has now enabled us to view and sub-classify each cancer type based on their distinct molecular features, in addition to histologic classification, with the promise of individualized treatment strategies tailored towards each specific subtype to improve patient outcomes. In this review, we have made an effort to systematically review the most up-to-date, leading literature in molecular analysis and/or subtyping of major gastrointestinal cancers. These include esophageal squamous cell carcinoma (ESCC), gastric cancer (GC) adenocarcinoma, pancreatic ductal adenocarcinoma (PDAC), hepatocellular carcinoma (HCC), gallbladder cancer (GBC), and colorectal cancer (CRC). For each cancer type we summarized the global mutational landscape, subgroup classification based on genomics, epigenetics, gene expression and/or proteomic analysis, and their salient clinicopathological features. We have highlighted the actionable mutations or mutational pathways that could help guide targeted therapies in the future.

Keywords: Gastrointestinal malignancies; molecular subtypes; mutational landscape; next-generation sequencing; genomics; epigenetics

Submitted Jul 12, 2016. Accepted for publication Sep 02, 2016.
doi: 10.21037/jgo.2016.11.01
View this article at: http://dx.doi.org/10.21037/jgo.2016.11.01

Introduction

Accurate classification is a requisite in the diagnosis, treatment and prognostication of cancer. Historically cancers have been classified according to the anatomic site of origin and histologic appearance, with the broad assumption that cancers from the same origin share common pathogenic processes. However, morphologically similar cancers can have widely variable clinical course and response to the same treatment, indicating fundamental differences in the pathogenic processes that drive each cancer. As cancer arises from distinct alterations in genetic and epigenetic events that culminate in transformed cellular behavior, being able to identify the crucial events that drive each cancer should ideally help optimize treatment decisions and improve patient outcomes, which are the principal goals of "personalized oncology". The emergence of high throughput molecular tools in the past two decades including mRNA microarray, next generation DNA and RNA sequencing, methylation array and various proteomic tools have brought us closer towards these goals. We are now able to comprehensively visualize the salient genomic characteristics, gene expression profiles and proteomic information of each cancer case and start categorizing them

into "molecular" subtypes, which theoretically should better reflect the biology and behavior of each cancer. In fact, the past decade has seen an explosion of cancer genomic and transcriptomic analyses, and tremendous efforts to sub-categorize each cancer type based on these molecular features. Though still at its infancy, this new way of cancer classification holds promise to allow more refined clinical trial designs, more optimal patient allocation to targeted therapeutics, and prognostication, which will ultimately help both the patients and treating physicians.

In this article, we have reviewed the most recent landmark literature in molecular analysis and/or subtyping of major gastrointestinal cancers. These include esophageal squamous cell carcinoma (ESCC), gastric adenocarcinoma, pancreatic ductal adenocarcinoma (PDAC), hepatocellular carcinoma (HCC), gallbladder cancer (GBC) and colorectal cancer (CRC). More attention will be paid towards gastric, pancreatic ductal and colorectal adenocarcinoma since these cancers are more common and best studied.

Overview of molecular landscape and classifications of GI cancers

Gastric cancer (GC)

GC, which consists predominantly of adenocarcinoma, is the fifth most common cancer globally, and third leading cause of cancer deaths in 2012 (1). The first and most comprehensive molecular characterization of gastric adenocarcinoma was reported by the Cancer Genome Atlas Research Network (2). In this study, 295 treatment naïve primary gastric adenocarcinoma samples were characterized using six different molecular platforms including array-based somatic copy number analysis, whole-exome sequencing (WES), array-based DNA methylation profiling, messenger RNA sequencing, microRNA (miRNA) sequencing and reverse-phase protein array. Integrated analysis of data from the platforms identified four distinct subtypes. Notably, no survival or racial differences were found among patients from each subgroup.

(I) Epstein-Barr virus-infected (EBV, 9% samples): signified by high EBV burden, extensive DNA promoter hypermethylation (including universal CDKN2A promoter hypermethylation), frequent PIK3CA (80%), ARID1A (55%) and BCOR (23%) mutations, amplification of 9p24.1 locus (15%) containing genes encoding JAK2, PD-L1 and PD-L2. These data suggest potential therapeutic role for PI3K inhibitors, JAK2 inhibitors and immune checkpoint antagonists in this subgroup. EBV-GC were mostly located in the gastric fundus or body (62%) and more frequently found in male patients (81%).

(II) Microsatellite instability (MSI, 22% samples): signified by hypermutated genome and DNA hypermethylation (including MLH1 promoter hypermethylation). Mutations in PIK3CA, EGFR, ERBB2, and ERBB3 were seen. MSI-GC tumors were diagnosed at an older age (median age 72 years), with a slightly higher prevalence in female patients (56%).

(III) Genomically stable (GS, 20% samples): signified by tumors with low somatic copy-number aberrations. GS-GC is enriched for CDH1 mutations (37%), which underlie hereditary diffuse GC syndrome, and either RHOA mutations or CLDN18-ARHGAP rearrangements (30%, mutually exclusive) which may enhance invasiveness and disrupt intercellular cohesion and contribute to the diffuse histology found in 73% of this subtype. GS-GC was diagnosed more frequently in younger patients (median age 59 years).

(IV) Chromosomal instability (CIN, 50% samples): signified by high somatic copy-number aberrations. High frequency of TP53 mutations (73%), genomic amplifications of genes in the receptor tyrosine kinase—Ras pathway including VEGFA, EGFR (10%), ERBB2 (24%), ERBB3 (8%), c-Met (8%), amplification of genes encoding cell cycle mediators, such as CCNE1, CCND1 and CDK6. These findings suggest potential use of many targeted agents towards this subtype. CIN-GC is found more frequently in the gastroesophageal junction/cardia (65%) and exhibits an intestinal histology.

Another major study was from the Asian Cancer Research Group (ACRG) (3,4). In this study, 300 GC were profiled using gene expression, genome-wide copy number microarray and targeted sequencing, yielding four distinct molecular subtypes. Importantly, each molecular subtype is associated with distinct prognosis.

(I) MSI-high (22.7% or 68/300 cases): this subtype occurred frequently in the antrum (75%), with >60% of cases exhibiting the intestinal subtype, and more than half of the cases diagnosed at early stages (I/II). This subtype had the best prognosis, and was associated with the presence of hypermutation, in addition to mutations in genes such as KRAS (23.3%), the PI3K-PTEN-mTOR pathway (42%), ARID1A (44.2%) and ALK (16.3%).

(II) Microsatellite stable/epithelial-mesenchymal transition (MSS/EMT, 15.3% or 46/300 samples): this subgroup was found to occur at a significantly younger age with the majority diagnosed with diffuse-type histology and at advanced stages (III/IV). This subtype had the worst overall prognosis and a higher chance of recurrence compared to the MSI subgroup.

(III) Microsatellite stable/epithelial/TP53 intact (MSS/epithelial/TP53$^+$, 26.3% or 79/300 samples): EBV infection occurred predominantly in this subgroup compared to other groups. This subtype had the second-best prognosis followed by MSI.

(IV) Microsatellite stable/epithelial/TP53 loss (MSS/epithelial/TP53$^-$, 35.7% or 107/300 samples): tumors falling under this subgroup had a less favorable prognosis compared to MSI and MSS/epithelial/TP53$^+$. As expected, this subgroup had the highest rate of TP53 mutations (60%).

The ACRG work (3,4) supplemented the TCGA analysis (2) by introducing TP53 activity and EMT in the classification. The two studies had similarities and differences. The MSI tumors were found in both datasets, and there was some overlap between GS, EBV+, and CIN subgroups defined by TCGA and MSS/EMT, MSS/TP53$^+$, and MSS/TP53$^-$ of the ACRG respectively. However, several differences in terms of molecular mechanisms, driver genes and prognosis were observed between the two cohorts.

PDAC

PDAC is the fourth leading cause of cancer death in the United States and is projected to be the second by 2020 (5). The overall 5-year survival of PDAC is 7% (6). The extremely poor prognosis of PDAC highlights the urgent need to understand and target the molecular aberrations that drive this disease. Collisson *et al.* (7) first classified micro-dissected pancreatic adenocarcinoma samples using gene expression analysis into three subtypes:

(I) Classical subtype: this subtype is characterized by high expression of adhesion-associated and epithelial genes, overexpression of GATA-binding protein 6 (GATA6), higher KRAS mRNA level and dependence on KRAS, more sensitivity to erlotinib and best survival outcomes.

(II) Quasimesenchymal (QM) subtype: this subtype is characterized by high expression of mesenchyme-associated genes, more sensitivity to gemcitabine and has the worst survival outcomes of all three subtypes.

(III) Exocrine-like subtype: as the name implies, this subtype has the highest presence of tumor cell-derived digestive enzyme genes.

In a more elaborate study, Waddell *et al.* performed whole-genome sequencing (WGS) and copy-number variation (CNV) analysis of 100 PDACs. Based on structural variations, four subtypes were proposed (8).

(I) Stable subtype (subtype 1): accounts for 20% of samples, contains <50% structural variation events. Point mutations of *KRAS*, *SMAD4*, and telomere length in this subtype was not different compared to other subtypes.

(II) Locally rearranged subtype (subtype 2): accounts for 30% of all samples. About 1/3 of these samples showed focal regions of gain/amplifications, leading to copy number gain of genes including *KRAS*, *SOX9*, *GATA6*, and at a lower prevalence (1–2% of patients) of targetable mutations such as *ERBB2*, *MET*, *CDK6*, *PIK3CA and PIK3R3*. The remaining 2/3 of this subtype contained complex genomic events.

(III) Scattered subtype (subtype 3): 36% of samples belonged to this subtype with an intermediate range of non-random chromosomal alterations and <200 structural variation events.

(IV) Unstable subtype (subtype 4): accounts for 14% of samples, characterized by a large number (>200, maximum 558) of structural variations suggestive of defects in DNA stability. This subtype showed strong relationship with mutations in BRCA pathway genes (*BRCA1*, *BRCA2 and PALB2*). Patients under the unstable subtype with a high *BRCA* germline mutational burden were found to be better responders to platinum based therapy.

The same group later performed a larger scale, more comprehensive analysis including whole genome sequencing, copy number variation and gene expression analyses, histopathologic and clinical correlation, on 456 PDAC samples (9). In this study, 32 significantly mutated genes assembled in 10 molecular pathways were identified. These include:

- Activating mutations of *KRAS* in 92%;
- Disruption of G1/S checkpoint mechanism in 78% (*TP53*, *CDKN2A*, *and TP53BP2*);
- TGF-B signaling in 47% (*SMAD3*, *SMAD4*, *TGFBR1*, *TGFBR2*, *ACVR1B and ACVR2A*);
- Histone modification in 24% (*KDM6A*, *SETD2*, ASCOM complex members *MLL2* and *MLL3*);

- ❖ The SW1/SNF complex in 14% (*ARID1A*, *PBRM1* and *SMARCA4*);
- ❖ The BRCA pathway 5% germline and 12% somatic (*BRCA1*, *BRCA2*, *ATM*, *PALB2*);
- ❖ WNT signaling defects through *RNF43* mutation in 5%;
- ❖ RNA processing genes in 16% (*SF3B1*, *U2AF1*, and *RBM10*).

In addition, transcriptome analysis of 96 PDAC tumors with ≥40% epithelial cellularity revealed four different subtypes (9):

(I) Squamous subtype: associated with mutations in *TP53*, *KDM6A*, activated α6β1, α6σ4, EGF signaling, hypermethylation and subsequent downregulation of genes determining pancreatic endodermal cell fate. In general, gene sets involved in inflammation, hypoxia response, metabolic reprogramming, TGF-B signaling, MYC pathway activation, autophagy and upregulated expression of TP63(delta)N characterize this subtype. This subtype is associated with the worst survival outcomes.

(II) Pancreatic progenitor subtype: characterized by activated transcription factors such as PDX-1, that regulate differentiation from endoderm to pancreatic lineage, inactivating mutations of *TGFBR2*, and gene programs involved in fatty acid oxidation, steroid hormone biosynthesis, drug metabolism, and O-linked glycosylation of mucins.

(III) Aberrantly differentiated endocrine exocrine (ADEX) subtype: defined by activated transcription factors that regulate later stages of pancreatic development and differentiation, including *NR5A2*, *MIST1*, *RBPJL*, *INS*, *NEUROD1*, *MAFA*.

(IV) Immunogenic subtype: this subtype shares many of the characteristics of the progenitor class, but is associated with significant immune infiltrate. Immune gene sets with a role in B- and T-cell signaling pathways, and antigen presentation delineate this subtype. Particularly upregulation of CTLA4 and PD1 acquired tumor response pathways in this subtype have therapeutic implications.

Three of the subtypes introduced by Bailey *et al.* directly overlap with Collisson subtypes presented earlier. QM-PDAC subtype in the Collisson study was renamed to squamous to better elucidate the common features of this specific type in PDAC with squamous type seen in other organs such as breast, bladder, lung, and head and neck; classical was renamed to pancreatic progenitor to better explain the presence of gene sets involved in early pancreas development in this subtype; and exocrine like was changed to ADEX to further include the endocrine differentiation in addition to exocrine differentiation.

Given the fact that genomic analysis of PDAC is frequently hampered by the sparse tumor cellularity and the presence of abundant stroma intermixed with normal endocrine and exocrine cells, Moffitt *et al.* performed virtual microdissection of 145 primary and 61 metastatic PDAC samples to overcome this challenge (10). In their study, they identified tumor-specific and stroma-specific subtypes with prognostic and biological relevance. Their study identified two stroma-specific subtypes:

(I) Activated stromal subtype: patients belonging to this subtype had a worse median survival time of 15 months and 1-year survival rate of 60% *vs.* 24 months and 80% when compared to the normal stromal subtype group. This subtype was characterized by a more diverse set of genes associated with macrophages, such as integrin ITGAM and chemokine ligands CCL13 and CCL18, WNT family members (WNT2 and WNT5A), MMP9 and MMP11, and FAP which has been associated with poor outcomes (11).

(II) Normal stromal subtype: this subtype was characterized by high expression of known markers for pancreatic stellate cells (ACTA2, VIM, and DES, encoding for actin, vimentin and desmopressin respectively).

Moffitt *et al.* also identified two separate tumor-specific subtypes, independent of normal and stromal factors:

(I) Classical tumor subtype: there was a strong overlap between genes of this subtype and the classical subtype defined by Collisson *et al.* earlier in this section (7). In concert with Collisson *et al.* results, the classical subtype in this study was also enriched for genes associated with GATA6 overexpression. In addition, SMAD4 expression was found to be consistently higher in this subtype consistent with the observation that loss of SMAD4 confers a more aggressive tumor phenotype.

(II) Basal-like tumor subtype: patients in this tumor subtype had a worse median survival time of 11 months and 1-year survival rate of 44% compared to 19 months and 70% of that of classical tumor subtype. Despite worse prognosis, patients belonging to this subtype revealed a strong trend toward response to adjuvant therapy. Manually curated RNA-seq data showed that KRAS

mutation encoding p.Gly12Asp was significantly overrepresented in this subtype.

Basal like and classical tumors were found in both the normal and activated stromal subtype. As expected, tumors from the classical subtype with normal stromal subtype had the best prognosis, and tumors from the basal-like subtype with activated stromal subtype had the worst prognosis (10).

CRC

CRC, or adenocarcinoma, is a disease with extensive intraclonal heterogeneity resulting in various outcomes and drug responses (12). However, molecular subtyping reported by different groups, based mainly on gene expression analysis, showed low consistencies, thereby posing great challenges in clinical translation (13-20). To reconcile these differences, the international CRC subtyping consortium (CRCSC) was formed to re-analyze 18 published datasets (N=4,151) (21). Four consensus molecular subtypes (CMS) were identified:

(I) CMS1; MSI immune (14%): characterized by hypermutated genome with few somatic copy number aberrations (SCNAs); encompasses the majority of MSI tumors and exhibits proteomic features suggestive of defective DNA mismatch repair; high frequency of *BRAF* mutations; increased expression of genes associated with a diffuse immune infiltrate along with strong activation of immune evasion pathways. These features make this subtype an appropriate target for immune checkpoint inhibition. Clinically, these tumors were dominant in female patients, mostly presenting as right-sided lesions with higher histopathological grade. In terms of clinical outcomes, these tumors proved to have a very poor survival rate with every recurrence.

(II) CMS2; canonical (37%): characterized by high prevalence of SCNAs, and amplifications of transcription factor *HNF4A*. Gene expression profiling showed epithelial differentiation and strong upregulation of *WNT* and *MYC* downstream targets, both of which have been indicated in CRC carcinogenesis. These tumors mostly presented as left-sided lesions.

(III) CMS3 (13%); metabolic: this subtype had distinct genomic and epigenomic features when compared with other chromosomal instability (CIN) subtypes (CMS2-4). In brief these tumors were found to have (i) consistently fewer SCNAs, (ii) hypermutation in 30% of samples, an overlap feature with CMS1 tumors, (iii) a higher prevalence of CpG island methylator phenotype (CIMP)-low clusters with intermediate levels of genome hypermethylation. A high frequency of KRAS mutations were observed in this subtype in addition to other genes involved in metabolism.

(IV) CMS4; mesenchymal (23%): gene expression profiling of this subtype revealed clear upregulation of genes involved in epithelial-to-mesenchymal transition (EMT), activation of transforming growth factor-β (TGF-β), angiogenesis, matrix remodeling pathways and complement associated inflammatory response. Clinically, these tumors were mostly diagnosed at more advanced stages (III/IV). In terms of clinical outcomes, these tumors tended to have worse overall survival and worse relapse-free survival.

ESCC

Esophageal cancer is the 8th most common cancer globally. ESCC is the most common histological type worldwide (1,22).

The mutational landscape of ESCC was investigated by Gao *et al.* through WES of 113 tumor-normal paired samples of treatment-naïve ESCC (23). The median number of mutations was found to be 2.9 non-silent mutations per megabase, far fewer than that of smoking-related bronchogenic carcinomas and UV-mediated melanomas. In addition, SCNA were found in 72% of cases. Genes responsible for regulating cell cycle, apoptosis, and DNA damage were found to be mutated in 99% of cases [*TP53* (93%), *CCND1* (33%), *CDKN2A* (20%), *NFE2L2* (10%), and *RB1* (9%)]. Histone modifying genes were found to be mutated in 63% of all ESCC cases [*MLL2* (19%), *MLL3* (6%), *KDM6A* (7%), *EP300* (10%), and *CREBBP* (6%)]. The Hippo and NOTCH pathways were also frequently dysregulated through a number of inactivating mutations [*FAT1-4* (27%), *AJUBA* (7%), *NOTCH1-3* (22%) and *FBXW7* (5%)] (23).

ESCC cases with *EP300* mutation had a less favorable prognostic outcome when controlled for other confounders such as TNM staging, age, gender, smoking and drinking history. This mutational analysis found major similarities between ESCC with squamous cell carcinoma of other primaries (head and neck and lung squamous cell carcinomas), but showed considerable differences when compared to that of esophageal adenocarcinomas (23).

HCC

HCC is the third cause of cancer related mortality in the world (22,24). In the past 10 years, considerable efforts have been made to segregate HCC, primarily based on gene expression technologies, into different prognostic groups (25,26). However, several factors underlie the heterogeneity of HCC. Epidemiologically, the incidence of HCC varies widely across different geographic regions, ancestral heritage, and genders. Furthermore, the etiologies of HCC differ markedly according to ancestry: chronic hepatitis B is the predominant cause of HCC in East Asia and Africa, whereas hepatitis C is more common in Japan and Western countries; Aflatoxin B1 exposure is a causative environmental factor in Asia and Africa, as opposed to alcohol consumption in Western countries (27,28). Therefore, it is reasonable to account for these differences in genomic analysis and subtyping of HCC. In a collaborative work between the International Cancer Genome Consortium (ICGC) and The Cancer Genome Atlas (TCGA) project, Totoki *et al.* conducted the first trans-ancestry HCC genome sequencing in 608 cases of liver cancer (503 sequenced by Totoki *et al.*, integrated with 105 from TCGA) (29). This cohort contained samples from patients with diverse ancestral backgrounds: European, US-Asian, African-American and Japanese. Among the cohort, 212 patients were HCV positive, 117 HBV positive, and 150 not infected by hepatitis viruses.

The average mutation rate was found to be 2.8 mutations per megabase. In terms of CNAs, about 29% of the cases showed gross chromosomal loss (average ploidy 3.87, and the average number of CNAs was 11.58). The following oncogenic pathways held significance in this analysis:

- TP53-RB pathway. Inactivation of this pathway is a recurring theme in HCC tumorigenesis. TP53 and RB1 mutations were observed in 31% and 4.4% of tumors respectively. Overall 72% of cases had mutations in component genes of one or both of these pathways such as *CDKN1A* and *CDKN2A*.
- WNT pathway. Activating *CTNNB1* mutations and inactivating mutations of *AXIN1* and *APC* were found to be signatures of this pathway. Overall 66% of HCC cases harbor a mutation in one of the WNT pathway genes.
- Chromatin and transcription modulators. Mutations involved in this pathway include alterations in *NFE2L2*, and nucleosome remodelers *ARID1A*, *ARID2* and *BRD7*.
- mTOR-PIK3CA pathway. Recurrent inactivating mutations in *TSC1-TSC2* and activating mutations of *PIK3CA* are the signatures of this pathway. Other modulators such as *NF1*, *PTEN*, *INPP4B* and *STK11* were also mutated. In general, 45% of cases had some kind of alteration in the mTOR-PIK3CA pathway.
- Interestingly, *TERT* (which encodes telomerase) promoter mutations were detected in total of 54% of cases, with the highest frequency among HCV positive patients (121/188; 64%), and lower frequencies in non-viral and HBV positive cases, 59% and 37% respectively. *TERT* promoter mutations significantly co-occurred with WNT pathway gene mutations in HCV and non-viral cases suggesting a permissive oncogenic activity between the two in these group of patients. These findings suggest that *TERT* is a driving mutation with potential therapeutic targetability. Alterations of *ATRX*, which allows telomerase-independent telomere maintenance, have also been reported. More than 68% of patients had either mutations in *TERT* or *ATRX* as the most frequent molecular event.

In addition to the abovementioned findings, this study also found ancestry dependent diversity in HCC mutation signatures regardless of hepatitis virus status or gender (29). One signature featured by CTG>CAG mutations dominated the cases of US-Asian male and female patients, whereas another signature characterized by AT>AC mutations was frequently seen in Japanese male cases. These findings indicate more complicated intra- and inter-ancestry variations and/or environmental exposures.

GBC

GBC is a rare aggressive tumor with median survival time of less than 1 year (30). To further understand the somatic mutation spectrum in GBC, Li *et al.* performed a combined WES and ultra-deep sequencing of 57 tumor-normal pairs of pathologically confirmed cases of GBC (31). Their sequencing efforts revealed *TP53* and *KRAS* as being recurrently mutated, with mutation rates of 47.1% and 7.8% respectively. Interestingly, this study identified recurrent mutations in ErbB signaling pathway: *ERBB1* (also known as EGFR) 3.9%, *ERBB2* 9.8%, *ERBB3* 11.8%, and *ERBB4* 3.9%. This study showed that overexpression of each *ERBB2* and *ERBB3* mutations result in a significant increase in proliferation in at least one cell line when compared to the wild type, highlighting the role of ErbB family receptors

in GBC development and progression.

In addition, Li *et al.* found that cases harboring one of the ErbB family signaling pathway mutations have worse prognostic outcomes in terms of overall survival when compared to their control counterparts (median of 8 *vs.* 13 months; P=0.012) (31). These results suggest that patients with one of ErbB family mutations will potentially benefit from targeted therapies against these specific mutations.

Conclusions

In this article, we reviewed different subclassification systems proposed for each major gastrointestinal cancer type based on various molecular tools including next generation sequencing, gene expression analysis, and application of sophisticated analytic algorithms. Although these subgroups have deepened our understanding of the global mutational landscape and signaling aberrations of each cancer types, they have also posed major challenges and elicited important questions that need to be further investigated. For instance:

- How can different sub-classification systems be reconciled between study groups?
- How does one interpret the significant discrepancies in results derived from genomic, transcriptomic and proteomic tools to appropriately guide therapeutic decision making?
- What is the standardized way of preparing and analyzing tumor samples in order to allow accurate sub-classification?
- Virtually all solid cancers are clonally heterogeneous and constantly evolving especially under the selection pressure of therapeutics (32). Do cancer subtypes change throughout the disease course?
- Within each subclass, how can driver and passenger mutations be further distinguished to inform treatment decision?
- How could these sophisticated analyses be distilled down to simple, valid and reproducible assays that can be applied in patient care in a timely fashion?
- Do tumor specimens derived from different sites (primary *vs.* metastatic) share the same molecular subtype?

Numerous basic and translational research are in progress to improve our current understanding and fill existing knowledge gaps in cancer biology. The tremendous quest in molecular subtyping of gastrointestinal malignancies, as is being done in cancers from other anatomic sites, is only meaningful if it can truly impact patient care. Much work remains to be done to standardize and streamline sample processing, data analysis and establishment of algorithms, to create reproducible and reliable diagnostic tools that can be used in daily practice. Of equal significance, development of effective therapeutic agents must follow the rapid pace of development in molecular diagnostics. Lastly, all these sophisticated and costly level of work that already are the basis of personalized oncology must be cost-effective and affordable to patients. Regardless of all these challenges, cancer genomics has irreversibly changed our mindset on cancer classification as many diseases rather than one disease, and has revolutionized our approach towards individualizing treatment.

Acknowledgements

Funding: This work was supported in part by AACR Pancreatic Cancer Action Network Career Development Award 16-20-25-LIM (KH Lim) and Siteman Investment Program Pre-R01 Award (KH Lim).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-86.
2. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202-9.
3. Wong SS, Kim KM, Ting JC, et al. Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. Nat Commun 2014;5:5477.
4. Cristescu R, Lee J, Nebozhyn M, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med 2015;21:449-56.
5. Rahib L, Smith BD, Aizenberg R, et al. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74:2913-21.

6. American Cancer Society. Cancer Facts & Figures 2016. Atlanta: American Cancer Society; 2016.
7. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500-3.
8. Waddell N, Pajic M, Patch AM, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518:495-501.
9. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47-52.
10. Moffitt RA, Marayati R, Flate EL, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015;47:1168-78.
11. Cohen SJ, Alpaugh RK, Palazzo I, et al. Fibroblast activation protein and its relationship to clinical outcome in pancreatic adenocarcinoma. Pancreas 2008;37:154-8.
12. Sottoriva A, Kang H, Ma Z, et al. A Big Bang model of human colorectal tumor growth. Nat Genet 2015;47:209-16.
13. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 2012;487:330-7.
14. Budinska E, Popovici V, Tejpar S, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. J Pathol 2013;231:63-76.
15. De Sousa E Melo F, Wang X, Jansen M, et al. Poor-prognosis colon cancer is defined by a molecularly distinct subtype and develops from serrated precursor lesions. Nat Med 2013;19:614-8.
16. Marisa L, de Reyniès A, Duval A, et al. Gene expression classification of colon cancer into molecular subtypes: characterization, validation, and prognostic value. PLoS Med 2013;10:e1001453.
17. Perez-Villamil B, Romera-Lopez A, Hernandez-Prieto S, et al. Colon cancer molecular subtypes identified by expression profiling and associated to stroma, mucinous type and different clinical behavior. BMC Cancer 2012;12:260.
18. Roepman P, Schlicker A, Tabernero J, et al. Colorectal cancer intrinsic subtypes predict chemotherapy benefit, deficient mismatch repair and epithelial-to-mesenchymal transition. Int J Cancer 2014;134:552-62.
19. Sadanandam A, Lyssiotis CA, Homicsko K, et al. A colorectal cancer classification system that associates cellular phenotype and responses to therapy. Nat Med 2013;19:619-25.
20. Schlicker A, Beran G, Chresta CM, et al. Subtypes of primary colorectal tumors correlate with response to targeted treatment in colorectal cell lines. BMC Med Genomics 2012;5:66.
21. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350-6.
22. Jemal A, Bray F, Center MM, et al. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
23. Gao YB, Chen ZL, Li JG, et al. Genetic landscape of esophageal squamous cell carcinoma. Nat Genet 2014;46:1097-102.
24. Forner A, Llovet JM, Bruix J. Hepatocellular carcinoma. Lancet 2012;379:1245-55.
25. Hoshida Y, Villanueva A, Kobayashi M, et al. Gene expression in fixed tissues and outcome in hepatocellular carcinoma. N Engl J Med 2008;359:1995-2004.
26. Lee JS, Chu IS, Heo J, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology 2004;40:667-76.
27. Augustine MM, Fong Y. Epidemiology and risk factors of biliary tract and primary liver tumors. Surg Oncol Clin N Am 2014;23:171-88.
28. El-Serag HB. Epidemiology of viral hepatitis and hepatocellular carcinoma. Gastroenterology 2012;142:1264-73.e1.
29. Totoki Y, Tatsuno K, Covington KR, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. Nat Genet 2014;46:1267-73.
30. Boutros C, Gary M, Baldwin K, et al. Gallbladder cancer: past, present and an uncertain future. Surg Oncol 2012;21:e183-91.
31. Li M, Zhang Z, Li X, et al. Whole-exome and targeted gene sequencing of gallbladder carcinoma identifies recurrent mutations in the ErbB pathway. Nat Genet 2014;46:872-6.
32. Schmitt MW, Prindle MJ, Loeb LA. Implications of genetic heterogeneity in cancer. Ann N Y Acad Sci 2012;1267:110-6.

Cite this article as: Fakhri B, Lim KH. Molecular landscape and sub-classification of gastrointestinal cancers: a review of literature. J Gastrointest Oncol 2017;8(3):379-386. doi: 10.21037/jgo.2016.11.01