

Whole-Exome Sequencing Reveals Molecular Characterization of Early-Onset and Peritoneal Metastasis in Gastric Cancers

Zhi-wei Zhou

Sun Yat-sen University Cancer Center

Hai-bo Qiu

Sun Yat-sen University Cancer Center

Xu-jun Wang

Shanghai Jiaotong University: Shanghai Jiao Tong University

Wei Wang

Sun Yat-sen University Cancer Center

Ting Wu

Sun Yat-sen University Cancer Center

Ying-bo Chen

Sun Yat-sen University Cancer Center

Xiao-wei Sun

Sun Yat-sen University Cancer Center

Run-cong Nie

Sun Yat-sen University Cancer Center

Yuan-Fang Li

Sun Yat-sen University Cancer Center

Shu-qiang Yuan

Sun Yat-sen University Cancer Center

Jiping Wang (✉ jwang39@bwh.harvard.edu)

Brigham and Women's Hospital <https://orcid.org/0000-0002-7466-098X>

Research

Keywords: GASTRIC CANCER, ONCOGENES, MOLECULAR MECHANISMS, CANCER GENETICS, METASTASIS

Posted Date: September 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-880086/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Increasing numbers of Gastric cancer(GC) patients were diagnosed at younger age with aggressive behavior including early disease recurrence, more frequent peritoneal metastasis and poor overall survival. To investigate the driver genes and mechanisms of the early-onset and more aggressive nature of the GC.

Methods: Gastric adenocarcinoma and matched germline samples were obtained from patients in Sun Yat-sen University Cancer Center between 2007 and 2013. Exome sequencing were performed for 198 pairs of primary gastric adenocarcinoma fresh tissue and blood samples from young or elder patients. Besides, matched tumor / blood exome sequencing was performed in 80 patients with peritoneal seeding and 51 patients without. Then bioinformatics analysis was performed to find genomic variants and their clinical meanings.

Results: Early-onset gastric cancers (EOGCs) have fewer somatic mutations but some deleterious germline variates in FAT genes, patients carried none, one, two, three of the 4 Single nucleotide polymorphisms (SNPs), the mean ages of diagnosis are 60, 50, 40 and 35 respectively. Somatic mutations in *CDH1*, *TGFBR1* and *CTNNB1* are related to early-onset cancers. These variants are all linked to WNT pathways. Somatic mutations in genes involved in cancer aggressiveness like *MAP2K7*, *CDH1* and *RhoA* are related to cancer progression and metastasis. Patients carrying *RhoA*, *ITGAV*, *TGFBR1*, *CDH1*, *CTNNB1*, *MYO9B*, *VAV1*, *SALL1*, *CDX4* somatic mutations or simultaneously carrying three *FAT* germline SNPs have been diagnosed at younger age than those have only *TP53* mutations.

Conclusions: The molecular characterization gives us a novel insight into the carcinogenesis and tumor progression mechanisms and may provide a guide to developing novel targeted therapy in GC.

Background

Although GC is the 5th most common cancer type and the 3rd leading cause of cancer death worldwide¹. Diagnoses at advanced stages and lack of effective chemotherapy and/or target therapy for metastatic cancers are two of the main causes for the high disease specific mortality. EOGC is usually associated with diffuse infiltration type, late stage at the time of diagnosis, and frequently associated with peritoneal metastasis, which leads to poor outcome²⁻⁴.

With the rapid development of massively parallel sequencing and high throughput molecular profiling techniques, many studies identified commonly mutated genes in the GC genomes, including *TP53*, *PTEN*, *PIK3CA*, *CTNNA1*, *CTNNB1*, *ARID1A*, *CDH1* and *RhoA*⁵⁻⁸. Data from The Cancer Genome Atlas (TCGA) project further classified GC into four groups: EBV positive, MSI, genomic stable, and chromosome instability⁹. However, the tumorigenesis mechanism for the early-onset gastric and/or peritoneal metastasis cancer patients and the genetic alternations responsible for its aggressive phenotype have been poorly understood. It has been observed that younger GC patients tend to be diagnosed at late stage

with aggressive behavior as comparing to older patients^{10, 11}. Germline *CHD1* mutation has been implicated in EOGC, although it only appears in < 2% of EOGC¹². In addition, *RhoA* mutation was reported to be associated with poor prognosis^{13, 14}. The goal of current study is to identify candidate new cancer driver genes responsible for GC tumorigenesis and progression.

In this study, we analyzed the somatic and germline variants to characterize genetic features of early-onset and metastasis GC. From the analysis, we found some aggressive makers stimulating cancer progression correlated with EOGC and metastasis GC. These results supported a cancer progression-accelerating hierarchical model that EOGC result from the aberration of highly aggressive tumors with metastatic preference and WNT over-activation that progresses rapidly instead of keeping latent for years. Our mutation group analysis also support this hypothesis.

Methods

Sample preparation

Gastric adenocarcinoma and matched germline (adjacent normal tissue or blood) samples were obtained from 198 patients under Institutional Review Boards (IRB) approved protocol. Eleven patients' tumor primary sites were in the fundus, 67 patients in the body and 110 patients in the antrum or pylorus of the stomach. The tumors located in the esophagus-gastric junction were excluded. Fresh-frozen tumors were obtained from surgical resection performed between 2007 and 2013 in Sun Yat-sen University Cancer Center, China. All tumor samples were collected from patients who were not previously treated with chemotherapy or radiation therapy. Tumors were staged according to the 8th edition of the American Joint Committee on Cancer. The study was approved by the Institutional Review Board on Ethics of Sun Yat-sen University Cancer Center (YB2015-002-01) and all patients signed Institutional Review Board approved consent forms.

Sample collection and processing were performed in terms of the protocol of a central Biospecimen Core Resource of TCGA. After a gross examination, the non-necrotic parts were picked up and excised from the tumor specimens by pathologists, and adjacent normal tissues were collected at least 2 cm from the tumor border in the luminal side of the specimen. Light microscopic examinations were performed on top slides by pathologists, to analyze the tumor-rich area before isolating genomic Deoxyribonucleic acid (DNA) and Ribonucleic acid (RNA) from frozen tumor tissue samples. Then, genomic DNA was extracted from macro-dissected frozen tumor tissues with maxwell RSC Instrument according to the manufacturer's instruction of the Maxwell® RSC Tissue DNA Kit and Maxwell® RSC Blood DNA Kit (Promega, Madison, WI, USA) for whole-exome sequencing (WES). Genomic DNA was also extracted from buffy coats using the same kit.

Our study aimed to identify the genetic alterations driving the development EOGC and those driving peritoneal seeding/recurrence. To better understand EOGC, we performed exome sequencing for 198 pairs of primary gastric adenocarcinoma fresh tissue (200X sequence depth) and blood samples (30X

sequence depth) from young (before 40 years of age, 53 patients), medium (40~70 years old, 100 patients) or older (>70 years old, 45 patients) patients.

To define genomic drivers in GC peritoneal seeding process, we also performed matched tumor / blood exome sequence on 80 patients with peritoneal seeding at the time of surgery (T4aNxM1) or experiencing abdominal recurrence within 3 year follow-up after surgery, and 51 patients without .Exome capture was conducted using (company's capture sequencing kit), and the captured DNA was sequenced on Illumina HiSeq at 200X coverage for the tumor samples and 30X for the blood samples, respectively. By comparing the exome sequence data between young and older patients, and between with or without peritoneal seeding patients, we aimed to characterize the specific genomic alterations in EOGC and peritoneal metastasis process.

Library preparation and sequencing.

Genomic DNA was isolated from peripheral blood, and sequencing libraries were prepared according to manufacturer's instructions for the Illumina sequencing platform. Briefly, ~3–5 µg genomic DNA was sheared into 200–800-base pair (bp) fragments, and the ends were repaired and ligated to Illumina standard paired-end adaptors. These ligated fragments were size selected on an agarose gel and then amplified by Polymerase chain reaction (PCR). Paired-end sequencing of the processed libraries was performed on the Illumina HiSeq 2000 platform. The read length was 90 bp, and the average sequencing coverage for each individual was ~20-fold.

Whole-exome sequencing (WES) analysis and variants calling

The whole exomes of tumor samples and matched blood samples were sequenced for each patient. DNA was fragmented and hybridized to the SureSelect Human All Exome Kit V5 (Agilent Technologies, Santa Clara, CA, USA), containing exon sequences from 27,000 genes. Exome shotgun libraries were sequenced on the Illumina Xten platform, generating paired-end reads of 150 bp at each end. Image analysis and base calling were performed with CAVSAVR (Illumina, San Diego, CA, USA) using default parameters. Sequencing adaptors and low-quality reads were removed to obtain high-quality reads. These were aligned to the NCBI human reference genome hg19 using the Burrows-Wheeler Aligner alignment algorithm.

BWA (v0.6.2-r126) was used to map reads against the human reference genome (GRCh 37/ hg19), allowing ≤ 3 mismatches across a single read. Each FASTQ read file was mapped using the "mem" function of BWA to align the sequence.

Genome Analysis Toolkit (GATK, version 3.5) was used to pre-process the reads. Localized insertion-deletion (indel) realignments were performed by GATK. GATK Realigner Target Creator was used to identify regions for realignment. For single-nucleotide variant (SNV) calling, candidate somatic SNVs was identified with the MuTect2¹⁵ algorithm in tumors by comparison with the matched control blood sample from each patient. For indel detection, tumor and blood samples were also analyzed with

MuTect2 algorithm. The MuTect2 algorithm was implemented by Sentieon software (<https://www.sentieon.com/>). All parameters were set to their defaults. SNV and indel annotation was performed by SnpEff. SIFT software was used to predict if a SNP has a functional effect on protein structure.

Somatic mutation signature analysis

Somatic mutation signatures were analyzed based on the SNV called by MuTect2. The analysis was implemented by R package "SomaticSignatures". Hg19 reference genome was used as the reference.

Results

The overall Mutation characteristics

The somatic single nucleotide mutation frequencies of C>A, C>G, C>T, T>A, T>C and T>G appear very similar among the three age groups (Young: <=40, Medium: 40-70 and Elder: >=70) (Figure 1A) and among patients with/without tumor metastasis (Figure 1B). These results suggested that the proposed etiologies are similar among the groups and hence some external factors are unlikely to contribute to EOGC risks or metastasis risks. Patients with tumor metastasis were significantly younger than those without metastasis (average 48 vs. 59 years old, respectively, Figure 1C). The observed somatic mutation frequency for *TP53*, *PTEN*, *PIK3CA* were 37%, 15% and 12% respectively and other genes *CTNNA1*, *CTNNB1*, *ARID1A*, *CDH1* and *RhoA* (Figure 1D). According to the somatic mutation patterns, the samples could be clustered into three groups (Figure 2A). In cluster 1, the patients have higher numbers of somatic mutations and the diagnosis ages are the highest (Figure 2B). In contrast, in cluster 2 and 3, the patients have lower numbers of somatic mutations and diagnosis ages, but the percentage of peritoneal implantation and mixture type of Laurence classification and are much higher (Figure 2C). The observations suggest that different mutation characteristics are associated with diverse clinical features.

Germline variations

Germline variations were compared between early and late-onset GCs. SNPs at loci rs3733415, rs2304024, rs80293525, rs150453320, located in the coding regions of *FAT1*, *FAT2*, *FAT3*, and *FAT3* respectively. In younger patients, both allele frequencies and proportions of patients carrying such SNPs are higher than those in the older groups (Figure 3A,B). Interestingly, number of cooccurrence of the 4 SNPs inversely correlated to the age of GC diagnosis (correlation=-0.5, p=0.0001). For patients carried none, one, two, three of the 4 SNPs, the mean ages of diagnosis are 60, 50, 40 and 35 respectively (figure 3C). Therefore, the *FAT* deleterious SNPs additively contributed to cancer risks.

Since the "concurrent deleteriousness" to *FAT* proteins implicated higher cancer risks, we then asked whether such concurrent effects could be inferred by somatic mutations. To answer this question and further confirm the influence of *FAT* deleterious variants, somatic mutation profiles of *FAT* were specially analyzed. As expected, *FAT* genes were frequently and concurrently mutated in GCs (figure S2). Moreover,

the co-occurrence trend was reproducible in TCGA STAD dataset. Hence, it could be inferred that *FAT1-3* probably worked as tumor suppressors rather than passengers in GC and the co-occurrence of deleterious both germline and somatic variants in *FAT1-3* may synergistically contribute to oncogenesis.

In addition, another SNPs in *ALDH* is negatively correlated with diagnosis ages. the relations between *ALDH* SNPs and esophageal cancer has already been observed from other genomic studies and our data indicated *ALDH2* SNPs were also associated with early carcinogenesis (Figure 3D). On the other hand, *BCLAF1* appears to be an oncogene since *BCLAF1* SNPs were associated with cancer diagnosis at older age (Figure 3D).

Somatic mutation associated with earlier onset of GC

On the basis of germline analysis, some germline SNPs in *FAT* genes are correlated with earlier age of cancer onset. These SNPs probably damage *FAT* genes' functions and then cause WNT pathway aberration. Next, to identify whether there exist somatic mutations triggering tumorigenesis at early ages, the characteristics of somatic mutation profiles in younger, medium, and older groups were further explored. As expected, the mutation counts of the patients are positively correlated with diagnosis ages (Figure 4A). The higher the age is, the fewer counts the patient has. The correlations reveal an accumulation of somatic mutations in the transformation of a normal cell into a tumor cell (Figure 4B).

Linear regression was then conducted to identify the relevance between diagnosis ages and mutations of individual genes. To identify the key drivers in the biological network among the top candidates related to younger GC, we queried STRING protein-protein interactions (PPI) database for core genes and involved biological connections. According to the PPI analysis, top 20 genes showed enriched pairwise protein-protein interactions (Figure 4C, enrichment significance, $p=0.00731$). Such significant interactions indicated tight cooperative or regulatory relationships with each other instead of random or isolated relationship among the top candidates. Other over-represented pathways are related to invasion and metastasis, such as cell adhesion ($fdr= 0.00471$), catenin complex ($fdr= 0.0207$) and plasma membrane part ($fdr=0.0454$) (Figure 4D). Of note, *CTNNB1* was the top-ranked protein frequently mutated in younger group. In addition, it was also the hub of the PPI network, which suggested a dominate role in this network. There were four patients harboring *CTNNB1* mutations in our data. The average diagnosis age is around 34.5 years old (from 25 to 41, whole cohort average is 52.6 years old).

In addition such aggressive genes are associated with diffuse type but the association doesn't exist in the other genes (Figure S3). Taken together, the risks of carcinogenesis at younger ages were linked to aberration of some aggressive pathways that accelerate cancer progression and metastasis through both inherited (germline SNPs) and postnatal ways (somatic mutations).

Somatic mutations associated with metastasis

Driver alterations involved in aggressive pathways were identified to correlate with early-onset of tumorigenesis. However, driver mutations responsible for GC progression and metastases need further

investigating. The patients with peritoneal implants were assigned to the metastasis group. Gene mutation features specific to metastatic tumors were then explored. Top 40 genes whose mutations are enriched in metastasis group were illustrated by STRING protein interaction network (Figure 5A). Individual genes without any interactions with other candidates were eliminated from the figures. *MAP2K7* mutations were observed in six cases and five have metastatic cancers. Among the six mutations, 4 of them were locating within the kinase domain of *MAP2K7* proteins. Similarly, we identified 12 *RhoA* mutations in 11 cases and these were over-represented in metastasis groups (Figure 5C). *RhoA* mutations occurred in three (3/74, 4.05%) versus eight patients (8/99, 8.08%) in non-metastasis and metastasis groups respectively (odds ratio= 2.072, p=0.2). The most common mutants, Y42C, L57V, G17E and D59G were seen in two, one, one and two cases, respectively. A notable and unreported mutant Y42H, lying in the effector-binding region of *RhoA*, was discovered in one metastatic case. Other mutations in *RhoA* were closed to such hotspot sites (Figure S4). In addition, *RhoA* and *CDH1* somatic mutations had a tendency towards Co-occurrence (log odds= 1.93, p=0.006), especially in metastasis group (log odds=2.1 p=0.013). The Co-occurrence in TCGA dataset was highly reproducible¹⁶ (log odds =1.90, p=0.002, Figure 5B). Besides, mutations of *RhoA* and *MAP2K7* also showed a slight tendency towards mutual exclusive both in our GC data and TCGA datasets (log odds= -inf and -0.33, respectively, Figure S4).

A proposed model

Our analysis on SNPs has demonstrated that some germlines SNPs in aggressive pathways associated with WNT signaling are remarkably correlated with EOGC, which is corresponding to our findings on somatic mutations. Indeed, Emergency Medical Technicians(EMT) genes and RhoA pathway can cooperatively induce cancer aggressiveness and metastasis through crosstalk with WNT signaling. According to our results and TCGA molecular classifications⁹, we, therefore, raised the hypothesis that EOGCs are caused by the aberration of highly aggressive tumors with metastatic preference (especially WNT over-activation) which progresses rapidly instead of keeping latent for years. The schematic illustration of GC development we proposed (Figure 6) may explain the heterogeneity in age at onset: 1) cancer mutations usually initiate at young ages. 2) If mutations occur and locate in some aggressive markers related to WNT pathways, they will drive the tumors more aggressive and progress rapidly. Carcinogenesis process of this type is explicit and doesn't last long, which matches to genomically stable(GS) group in TCGA⁹. 3) If mutations only located in TP53 loci, they will cause the genomics instability and accelerate alteration accumulation to reach the advanced level during the carcinogenesis process^{9, 17}, which matches to Chromosomal instability (CIN) group in TCGA. 4) Otherwise, mutations only prefer to occur in microsatellite/simple repeat regions or vulnerable regions without dramatically stimulating or promoting. This process requires longer time and thus diagnosis ages of this Microsatellite instability(MSI) group are oldest in TCGA cohorts⁹. EBV groups could not be characterized in detail due to data limitation.

To validate this hypothesis, we divided the patients into several groups: 1) patients with *TP53* mutations and aggressive mutations (patients carrying *RhoA*, *ITGAV*, *TGFBR1*, *CDH1*, *CTNNB1*, *MYO9B*, *VAV1*,

SALL1, *CDX4* somatic mutations or simultaneously carrying three *FAT* germline SNPs), 2) only aggressive mutations without *TP53* mutations; 3) *TP53* mutations alone without aggressive mutations 4) The rest (Figure S6). There were marked differences between these three groups. As the boxplots shown, aggressive markers are the primary factors affect diagnosis age. Patients carrying aggressive mutations (group 1 & group 2) have been diagnosed at younger age than those have only *TP53* mutations. The average ages of group 1 and group 2 are 38.9 and 47.4 while average ages of group 3 are 57.9 (group 2 vs group 3, $p= 0.03789$; group 1 vs group 3, $p= 0.0001159$, figure S6). Furthermore, the coordination between *TP53* mutations and aggressive mutations make the diagnosis ages nearly 9 years earlier than group 2 (group 1 vs group 2, $p= 0.0464$). Therefore, these results enhanced our hypothesis that aggressive mutations primarily shorten the duration of carcinogenesis process while *TP53* mutations accompanying with the aggressive mutations dramatically accelerate the oncogenesis. However, *TP53* alone has fewer effect on diagnosis age than aggressive mutations.

Discussion

GC is a common lethal disease - the 3rd leading cause of death from cancers globally. There is a huge need to identify new cancer driver genes causing carcinogenesis and tumor progression. We hypothesized that specific genomic alterations could induce EOGC and peritoneal metastasis process respectively. Through integration with clinical characteristics and genomic profiles, we investigated the molecular features of EOGC and metastatic GC. Importantly, both EOGC and metastatic GC have a strong preference for molecular alterations particularly involved in tumor aggressiveness and aggravation.

Our data show (Fig. 1) that the commonly mutated known driver genes showed similar mutation frequency to those reported in literature⁹. The metastasis drivers identified in our datasets are the modulators related to RhoA pathways, Epithelial-Mesenchymal Transition process and MAPK signaling, which highly correspond to previous studies. These pathways are known to have close interplays with WNT signaling and the crosstalk between each other can reinforce the cancer aggressiveness and metastasis. Particularly, our findings of *RhoA* and *CDH1* co-occurrence mutations raised the possibility for their contributory role in cancer progression. The metastatic tumors showed preference for mutations of *CDH1*, *RhoA* and *MAP2K7*, three cores in this network and known drivers involved in EMT, RhoA pathway and MAP kinase signaling^{9, 18, 19}.

Our findings (Fig. 5, Figure S4) suggested a complicated crosstalk among RhoA, MAPK and EMT pathways in cancer metastasis. PPI analysis reveals that CTNNB1/WNT related genes are tight neighborhood of these top candidates, though beta-catenin itself was not beyond the cutoff yet. Recent reviews described that RhoA activation could potentiate the transcriptional activating capacity of WNT for targets or might activate WNT through YAP-dependent manner²⁰ and that loss of CDH1/E-cadherin can attenuate the β -catenin sequestration at sites of cell-cell contact^{17, 21}. These literature reviews indicated that WNT could co-operatively work with candidate mutations to stimulate cancer progression and metastasis.

On the other hand, EOGC is also associated with alterations locating in mediators of cell adhesion or WNT related genes. We found *CTNNB1*, *ITGAV*, *TGFBR1* and other regulators are frequently somatic mutated inversely to the age.

CTNNB1, also named beta-catenin, is known as an intracellular signal transducer in the Wnt signaling²², the same pathway highlighted by germline analysis. Other genes connecting to *CTNNB1* in PPI network like *CDH1*²³, *TGFBR1*²⁴, *ITGAV*^{25, 26}, together with *CTNNB1*, are known cancer progression and metastasis regulators. Interestingly, some germline SNPs in these genes were reported to associate with increased cancer risk^{27, 28}. Therefore, our findings are consistent with previous germline analysis, though such germline polymorphisms in our data are not as significant possibly due to population differences or ethnicity. The integration of PPI and literature evidences suggest that such genes could exert aggressive effects on cancer development by coordinating with cancer aggressiveness markers (like WNT signaling.)

Moreover, we found patients carrying SNPs in *FAT1-3* are prone to get GCs at younger ages, all these SNPs were predicted to be “DELETERIOUS” according to the SIFT results²⁹. *FAT* genes belong to the cadherin family, which are responsible for maintaining cell adhesion thus are considered tumor suppressor genes. These genes involved in tumor suppression, planar cell polarity as well as cell adhesion, appears to play an important role in cell since it exists at 3 chromosome locations^{30, 31}. Although the function of SNP in *FAT* regions in cancer progression was not well established, frequent *FAT* somatic mutations are exhibited in variant cancer types, which are reported to promote cancer cell growth and the WNT/beta-catenin activation^{32, 33}. Anastas et al. reported that *FAT1* frequent somatic mutations in multiple cancer types were linked to aberrant Wnt activation, a well-known oncogenic signaling pathway promoting cancer initiation, aggressiveness as well as metastasis^{34, 35}. In addition, several phylogenetic studies on *FAT* family reveals that *FAT1*, *FAT2* and *FAT3* as orthologs of *Drosophila Fat1* are conserved in sequence and functions^{30, 36}. However, another *FAT* family gene *FAT4*, the orthologs of *Drosophila Fat3*³⁰, were divergent from *FAT1-3* in domain architecture³⁷. This difference may explain why *FAT4* SNPs are not associated EOGC. Taken together, our analysis revealed that *FAT* deleterious germline SNPs may confer the EOGC risk through a WNT dependent manner.

Both early-onset related SNPs and somatic mutations were linked to WNT dysregulation. High consistency of results from both germline and somatic variant analysis enhanced our findings that aggressive pathways accelerating cancer progression contribute to the EOGC via some specific inherited or acquired genomic patterns.

Our systematical investigation on mutation profiles of metastasis GC and EOGC inspire us to hold the point of view that the shortened latency periods of tumorigenesis and aggressive phenotypes are more major causes of GC in younger patients. Interestingly, some published researches on GC incidence of immigrants concluded that the cancer rates which are high in countries of origin but occur infrequently in countries of adoption tend to decrease in succeeding immigrants and vice versa. Those studies in one hand demonstrated childhood environmental exposure is important in setting the individual's cancer

destiny. On the other hand, they suggested that the initiation of tumorigenesis is usually determined during childhood and hence ages of onset mainly depend on the length of incubation period. On the basis of results and literature evidences, we then have proposed a hierarchical Model for alterations in contributing to cancer acceleration and progression (Fig. 5). In addition, we also uncovered the commonality of aggressive genomic patterns between metastasis GC and EOGC, which partially explained why younger patients usually have higher proportions of diffuse types and more aggressive tumor behaviors.

Conclusion

Our comprehensive analyses give us a novel insight into the mechanism understanding of tumor genesis and metastasis. The molecular characterization gives us a novel insight into the carcinogenesis and tumor progression mechanisms and may serve as a valuable adjunct to the early diagnosis and prevention. These observations provide a guide to developing novel targeted therapy in GC.

Abbreviations

GC: Gastric Cancer

EOGCs: Early-onset Gastric Cancers

DNA: Deoxyribonucleic acid

RNA: Ribonucleic acid

SNPs: Single Nucleotide polymorphisms (SNPs)

TCGA: The Cancer Genome Atlas

IRB: Institutional Review Boards

WES :Whole-exome sequencing

PCR: Polymerase chain reaction

SNV: Single-nucleotide Variant

PPI: protein-protein interactions

EMT: Emergency Medical Technicians

GS: genomically stable

EBV: Epstein–Barr virus

MSI: Microsatellite instability

CIN: Chromosomal instability

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board on Ethics of Sun Yat-sen University Cancer Center (YB2015-002-01) and all patients signed Institutional Review Board approved consent forms.

Consent for publication

Consent to publish has been obtained from all authors.

Availability of data and material

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors disclose no conflicts.

Funding

This work was supported by the Chinese Society of Clinical Oncology Roche foundation grants: Y-2019Roche-157; Natural Science Foundation of Guangdong Province (grant number: 2018A030313631); Wu Jieping Medical Foundation (grant number: 320.6750.17513) and Young Teacher Training Program of Sun Yat-sen University (grant number: 19kypy172).

Authors' contributions

Zhi-Wei Zhou, Jiping Wang: study concept and design; acquisition of data; analysis and interpretation of data; drafting of the manuscript; critical revision of the manuscript for important intellectual content; statistical analysis; obtained funding; study supervision.

: study concept and design; acquisition of data; analysis and interpretation of data; technical and material support; critical revision of the manuscript for important intellectual content.

Hai-bo Qiu: study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content.

Xu-jun Wang: acquisition of data; technical support, analysis and interpretation of data; material support,

Ting Wu: acquisition of data; analysis and interpretation of data; material support.

Wei Wang: study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content.

Ying-bo Chen, Yuan-Fang Li: acquisition of data; analysis and interpretation of data.

Xiao-wei Sun, Shu-qiang Yuan: study concept and design; analysis and interpretation of data; critical revision of the manuscript for important intellectual content.

Wei Li: acquisition of data; analysis and interpretation of data; material support.

Run-cong Nie: analysis and interpretation of data; material support of the manuscript for important intellectual content.

Acknowledgements

The authors acknowledge all the patients and their families for anticipating in this study.

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. Nov 2018;68(6):394-424. doi:10.3322/caac.21492
2. Takatsu Y, Hiki N, Nunobe S, et al. Clinicopathological features of gastric cancer in young patients. *Gastric Cancer*. Apr 2016;19(2):472-478. doi:10.1007/s10120-015-0484-1
3. Lee J, Lee MA, Kim IH, Roh SY. Clinical characteristics of young-age onset gastric cancer in Korea. *BMC Gastroenterol*. Sep 6 2016;16:110. doi:10.1186/s12876-016-0528-y
4. Milne AN, Offerhaus GJ. Early-onset gastric cancer: Learning lessons from the young. *World J Gastrointest Oncol*. Feb 15 2010;2(2):59-64. doi:10.4251/wjgo.v2.i2.59
5. Nagarajan N, Bertrand D, Hillmer AM, et al. Whole-genome reconstruction and mutational signatures in gastric cancer. *Genome Biol*. Dec 13 2012;13(12):R115. doi:10.1186/gb-2012-13-12-r115
6. Zang ZJ, Cutcutache I, Poon SL, et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet*. May 2012;44(5):570-4. doi:10.1038/ng.2246
7. Wong SS, Kim KM, Ting JC, et al. Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. *Nat Commun*. Nov 19 2014;5:5477. doi:10.1038/ncomms6477
8. Chen K, Yang D, Li X, et al. Mutational landscape of gastric adenocarcinoma in Chinese: implications for prognosis and therapy. *Proc Natl Acad Sci U S A*. Jan 27 2015;112(4):1107-12. doi:10.1073/pnas.1422640112
9. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature*. Sep 11 2014;513(7517):202-9. doi:10.1038/nature13480

10. Rona KA, Schwameis K, Zehetner J, et al. Gastric cancer in the young: An advanced disease with poor prognostic features. *J Surg Oncol*. Mar 2017;115(4):371-375. doi:10.1002/jso.24533
11. Sasao S, Hiyama T, Tanaka S, Yoshihara M, Yasui W, Chayama K. Clinicopathologic and genetic characteristics of gastric cancer in young male and female patients. *Oncol Rep*. Jul 2006;16(1):11-5.
12. Bacani JT, Soares M, Zwingerman R, et al. CDH1/E-cadherin germline mutations in early-onset gastric cancer. *J Med Genet*. Nov 2006;43(11):867-72. doi:10.1136/jmg.2006.043133
13. Kakiuchi M, Nishizawa T, Ueda H, et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat Genet*. Jun 2014;46(6):583-7. doi:10.1038/ng.2984
14. Wang K, Yuen ST, Xu J, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet*. Jun 2014;46(6):573-82. doi:10.1038/ng.2983
15. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. Mar 2013;31(3):213-9. doi:10.1038/nbt.2514
16. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer discovery*. May 2012;2(5):401-4. doi:10.1158/2159-8290.CD-12-0095
17. Jeggo PA, Pearl LH, Carr AM. DNA repair, genome stability and cancer: a historical perspective. *Nat Rev Cancer*. Jan 2016;16(1):35-42. doi:10.1038/nrc.2015.4
18. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. Mar 8 2007;446(7132):153-8. doi:10.1038/nature05610
19. Zhou J, Hayakawa Y, Wang TC, Bass AJ. RhoA mutations identified in diffuse gastric cancer. *Cancer Cell*. Jul 14 2014;26(1):9-11. doi:10.1016/j.ccr.2014.06.022
20. Yu OM, Brown JH. G Protein-Coupled Receptor and RhoA-Stimulated Transcriptional Responses: Links to Inflammation, Differentiation, and Cell Proliferation. *Molecular pharmacology*. Jul 2015;88(1):171-80. doi:10.1124/mol.115.097857
21. Liu X, Chu KM. E-cadherin and gastric cancer: cause, consequence, and applications. *Biomed Res Int*. 2014;2014:637308. doi:10.1155/2014/637308
22. Clevers H, Nusse R. Wnt/beta-catenin signaling and disease. *Cell*. Jun 08 2012;149(6):1192-205. doi:10.1016/j.cell.2012.05.012
23. Schell MJ, Yang M, Missiaglia E, et al. A Composite Gene Expression Signature Optimizes Prediction of Colorectal Cancer Metastasis and Outcome. *Clinical cancer research : an official journal of the American Association for Cancer Research*. Feb 01 2016;22(3):734-45. doi:10.1158/1078-0432.CCR-15-0143
24. Cichon MA, Moruzzi ME, Shqau TA, et al. MYC Is a Crucial Mediator of TGFbeta-Induced Invasion in Basal Breast Cancer. *Cancer research*. Jun 15 2016;76(12):3520-30. doi:10.1158/0008-5472.CAN-15-3465

25. Li X, Wu WK, Xing R, et al. Distinct Subtypes of Gastric Cancer Defined by Molecular Characterization Include Novel Mutational Signatures with Prognostic Capability. *Cancer Res.* Apr 1 2016;76(7):1724-32. doi:10.1158/0008-5472.CAN-15-2443
26. Berghoff AS, Kovanda AK, Melchardt T, et al. alphavbeta3, alphavbeta5 and alphavbeta6 integrins in brain metastases of lung cancer. *Clin Exp Metastasis.* Oct 2014;31(7):841-51. doi:10.1007/s10585-014-9675-0
27. Gao L, Nieters A, Brenner H. Meta-analysis: tumour invasion-related genetic polymorphisms and gastric cancer susceptibility. *Alimentary pharmacology & therapeutics.* Sep 01 2008;28(5):565-73. doi:10.1111/j.1365-2036.2008.03760.x
28. Jiang B, Zhu K, Shao H, Bao C, Ou J, Sun W. Lack of association between the CDH1 polymorphism and gastric cancer susceptibility: a meta-analysis. *Sci Rep.* Jan 20 2015;5:7891. doi:10.1038/srep07891
29. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic acids research.* Jul 01 2003;31(13):3812-4.
30. Katoh M. Function and cancer genomics of FAT family genes (review). *Int J Oncol.* Dec 2012;41(6):1913-8. doi:10.3892/ijo.2012.1669
31. Gee HY, Sadowski CE, Aggarwal PK, et al. FAT1 mutations cause a glomerulotubular nephropathy. *Nature communications.* Feb 24 2016;7:10822. doi:10.1038/ncomms10822
32. Morris LG, Kaufman AM, Gong Y, et al. Recurrent somatic mutation of FAT1 in multiple human cancers leads to aberrant Wnt activation. *Nature genetics.* Mar 2013;45(3):253-61. doi:10.1038/ng.2538
33. Lin DC, Hao JJ, Nagata Y, et al. Genomic and molecular characterization of esophageal squamous cell carcinoma. *Nat Genet.* May 2014;46(5):467-73. doi:10.1038/ng.2935
34. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature.* Aug 22 2013;500(7463):415-21. doi:10.1038/nature12477
35. Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer.* Jan 2013;13(1):11-26. doi:10.1038/nrc3419
36. Zhang X, Liu J, Liang X, et al. History and progression of Fat cadherins in health and disease. *Onco Targets Ther.* 2016;9:7337-7343. doi:10.2147/OTT.S111176
37. Katoh Y, Katoh M. Comparative integromics on FAT1, FAT2, FAT3 and FAT4. *Int J Mol Med.* Sep 2006;18(3):523-8.

Figures

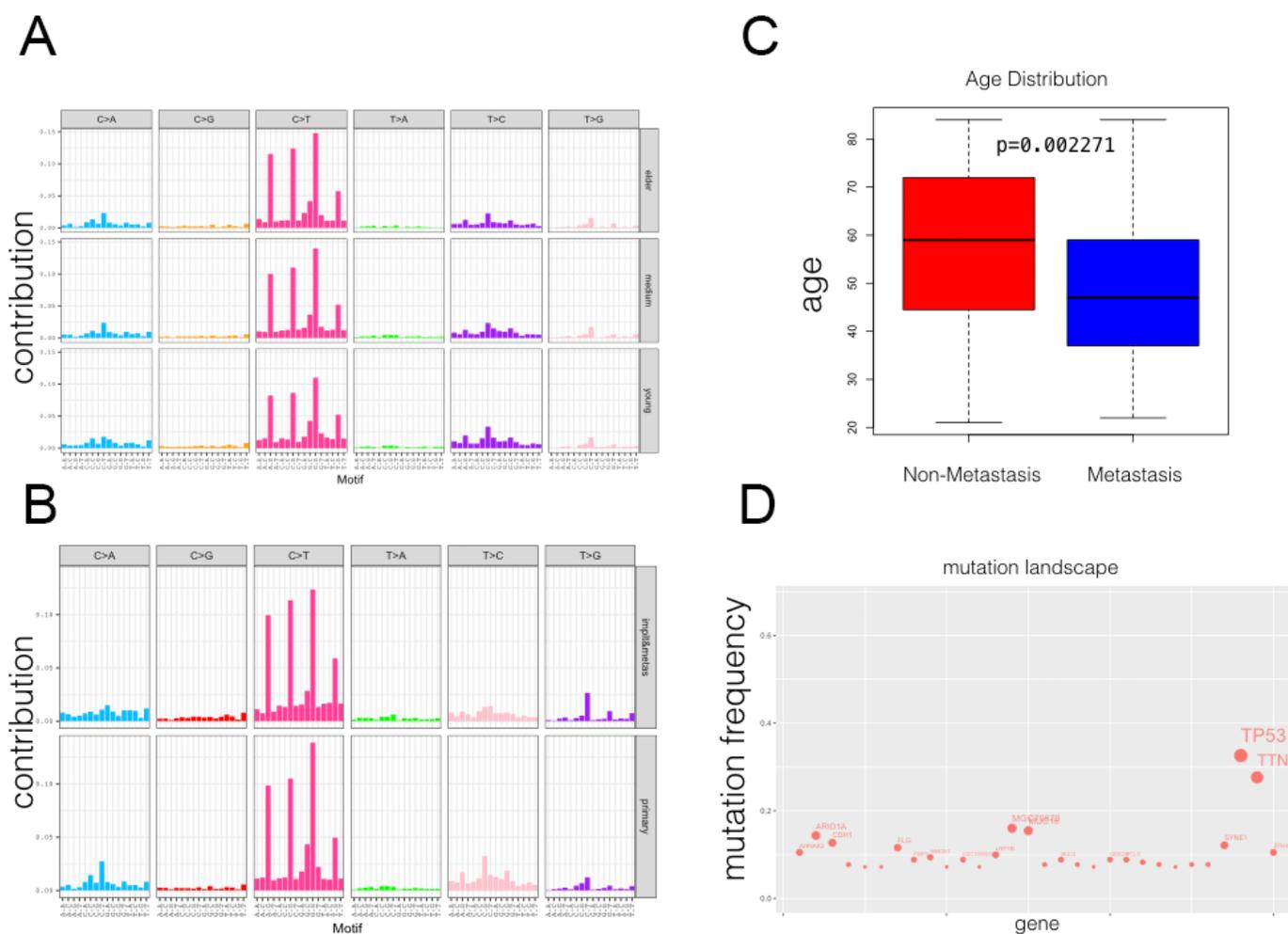


Figure 1

Overview of mutation signatures (A-B) the mutation signatures of different groups: (A) age elder, medium and younger patients. (B) metastasis and non-metastasis patients. The mutation signatures are almost the same across the groups suggesting a similar proposed aetiologies among the groups. (C) the diagnosis age of the patients with metastasis and non-metastasis. (D) the mutation landscape of high frequency genes. The y-axis is the frequency. The genes is ordered by alphabetical order.

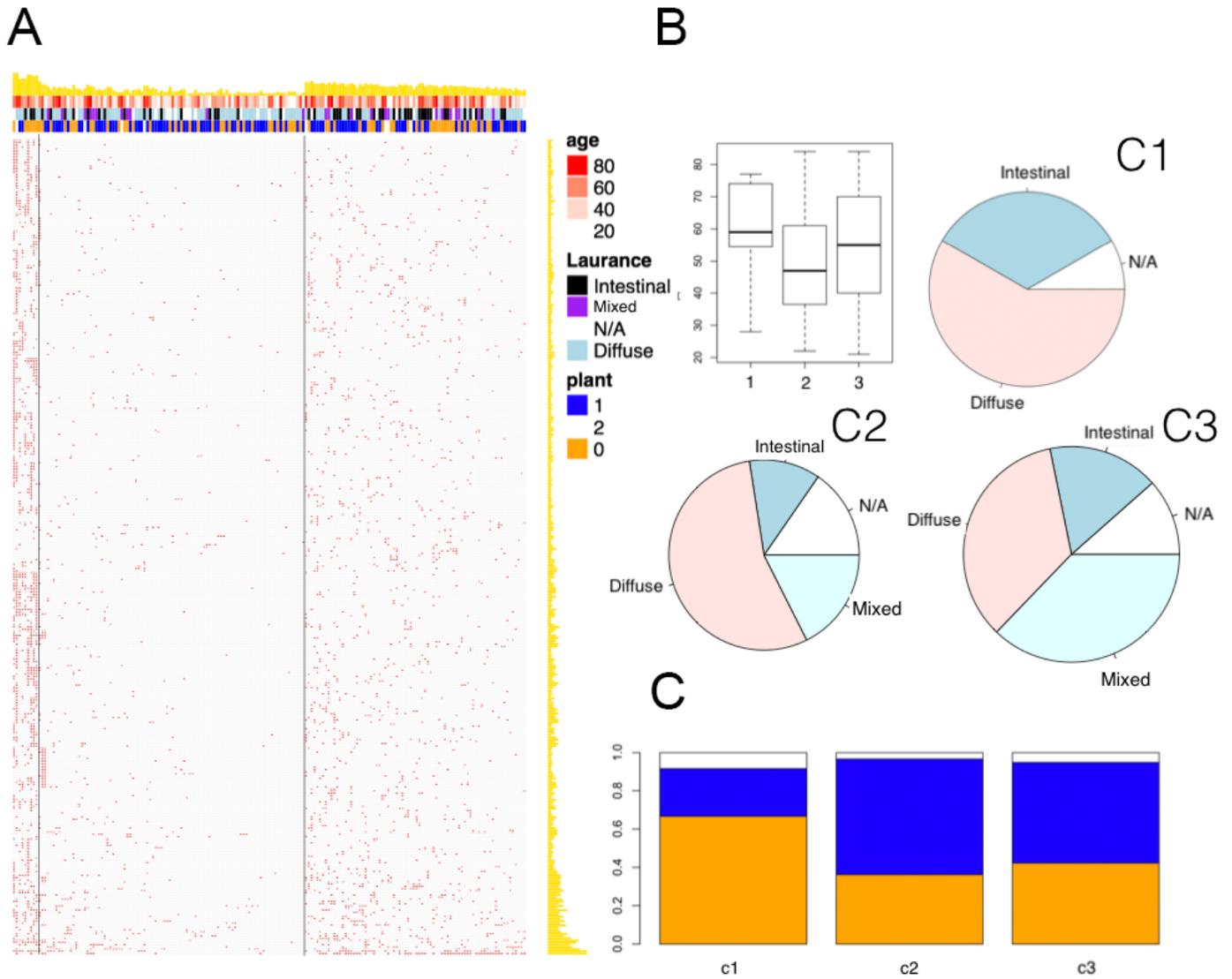


Figure 2

Clusters of patients with mutation profiles and clinical information. (A) heatmap of mutations of GC patients and annotated by the clinical information including Laurance subtypes, metastasis status and ages. Manhattan distance was used to evaluate the distance between samples. Patients were clustered into 3 groups and divided by the black lines. (B) boxplot plots of patients' ages of each cluster derived by heatmap. (C-e) pie charts of percentage of Laurance subtypes for cluster 1 (C1), cluster 2 (C2) and cluster 3 (C3). NA means the subtype information is not available. (C). The percentage of metastasis status of each group. Blue represents the metastasis, orange represents the non-metastasis and white represents "not available".

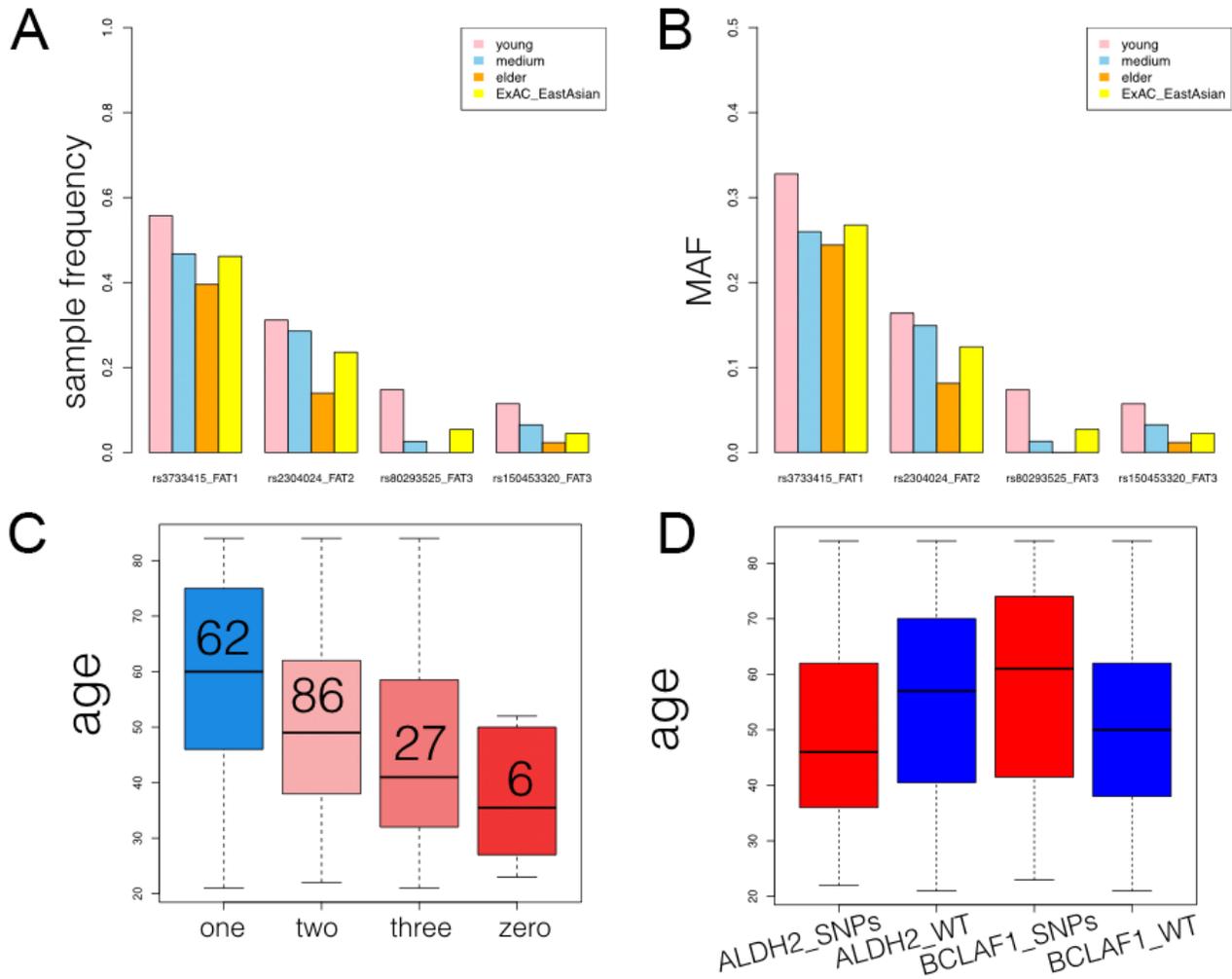


Figure 3

Germline variations related to EOGCs. Germline SNPs located in FATs are associated with early onset gastric cancer. Germline SNPs located in FAT family are associated with EOGCs. Barplots of candidate SNPs, (A) is the frequency of samples that carry candidate SNPs. (B) is the MAF(allele frequency). yellow bar is the frequency information of east asia samples in ExAC database. (C) boxplot for age distribution of the samples without /with one, two or three SNPs simultaneously. (D) the boxplot for age distribution of the samples with /without SNPs in ALDH2 or BCLAF1.

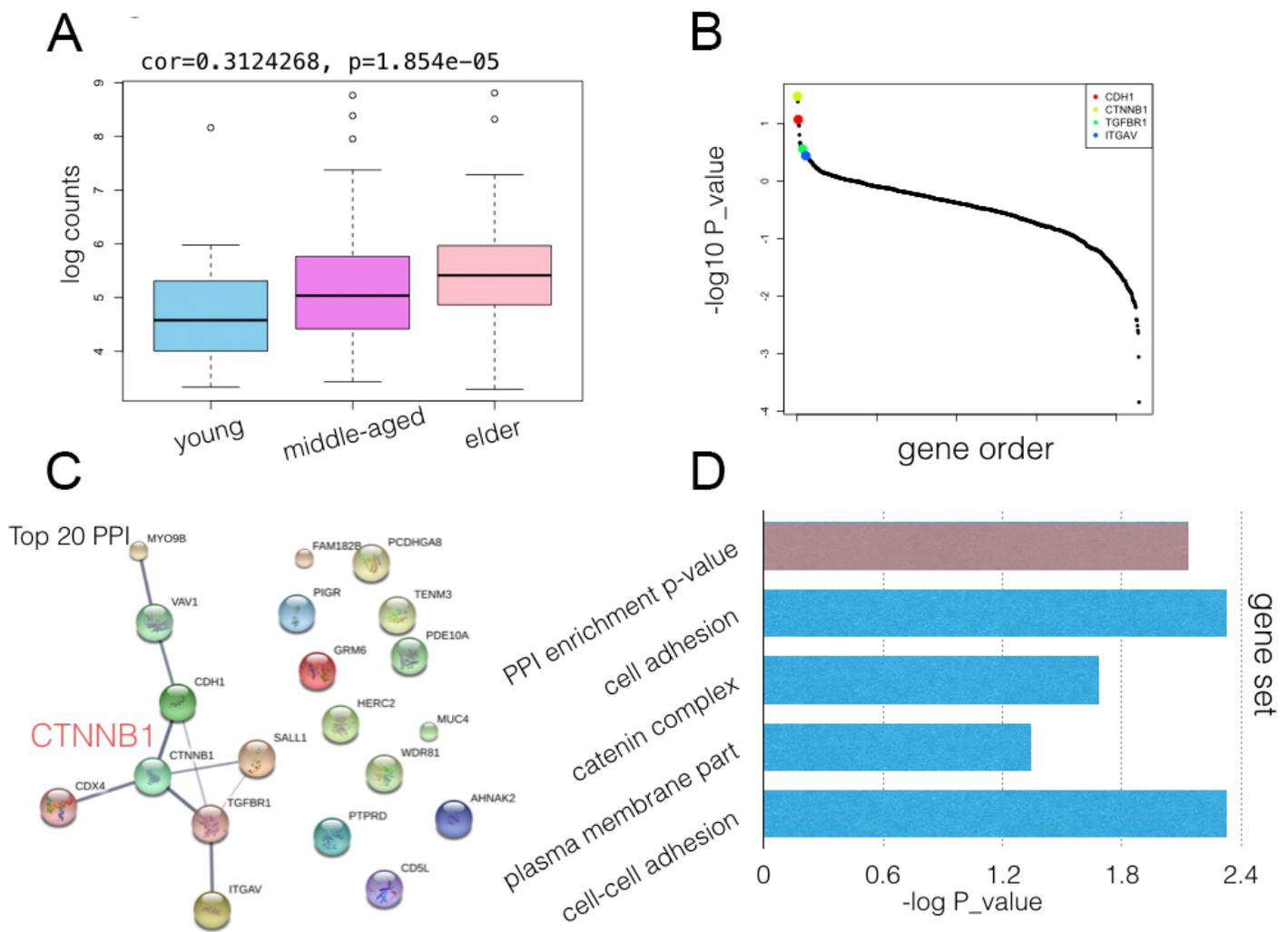


Figure 4

Somatic mutations related to EOGCs. (A) Boxplot for somatic mutation counts of elder, medium and younger groups. Somatic mutation counts are positive correlated with diagnosis age. Mutation counts were calculated into log scale. (B) significance scores of linear regression with direction (negatively correlated >0 , positively correlated <0) of somatic mutations, mutations with frequency $>2\%$ were considered. (C) PPI pattern (STRING database) of top 40 genes that were ranked by significance score derived from linear regression (panel b). (D) enriched pathways of top 40 candidates.

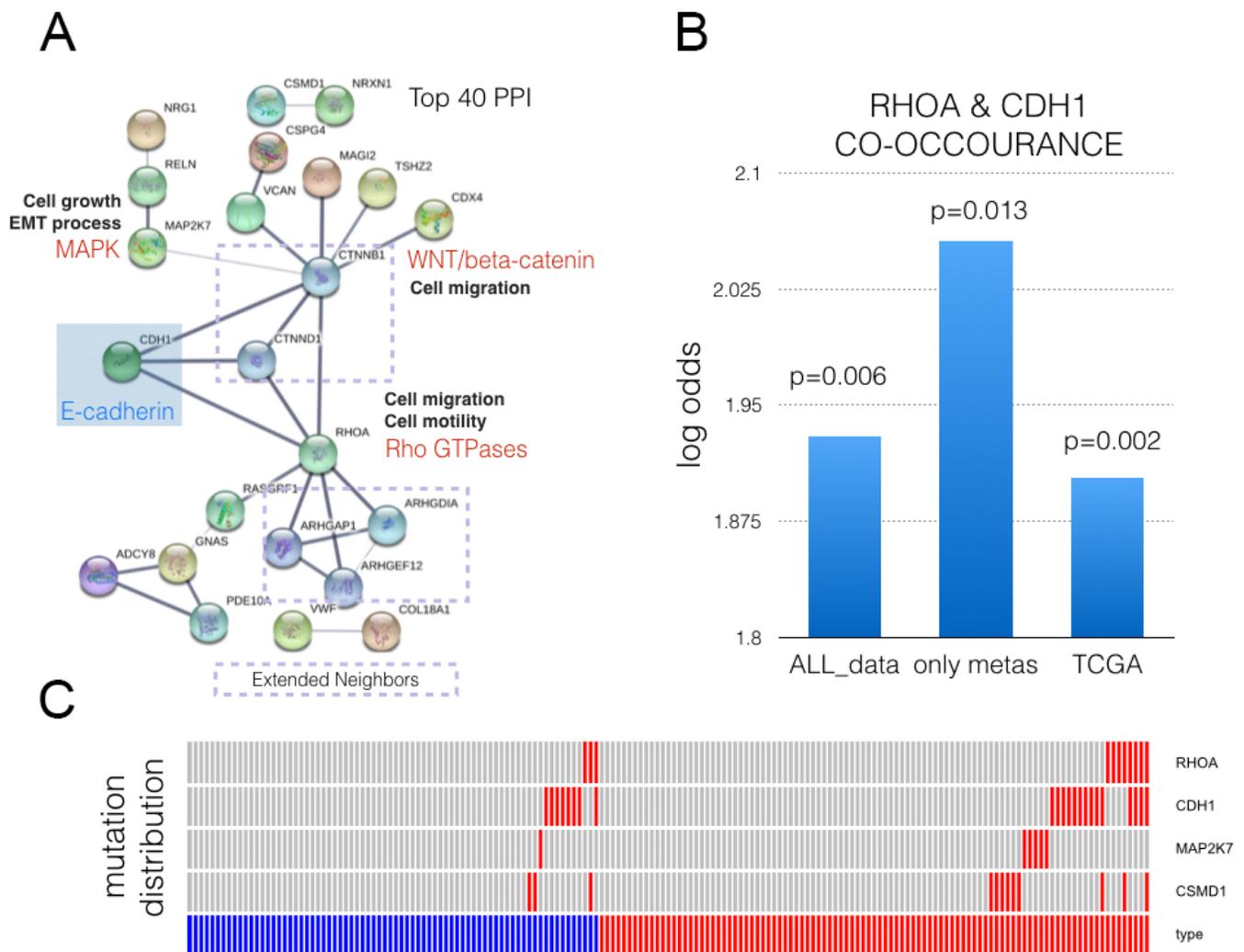


Figure 5

Somatic mutation related to metastasis. (A) Top 40 genes and extended neighbors PPI network. Top genes are frequently mutated in metastasis group, which is ordered by p-value of one-tail fisher test. Genes in the dashed box is not involved in top 40 genes but are the tight neighbors of them. Other genes without any interactions are now shown. (B) barplot of odds of three groups. 1) all gastric cancer, 2) only the metastasis group. 3) TCGA STAD groups. (C) mutation heatmap of RhoA, CDH1, MAP2K7 and CSMD1. Type represents the metastasis status where red means metastasis and blue means non-metastasis.

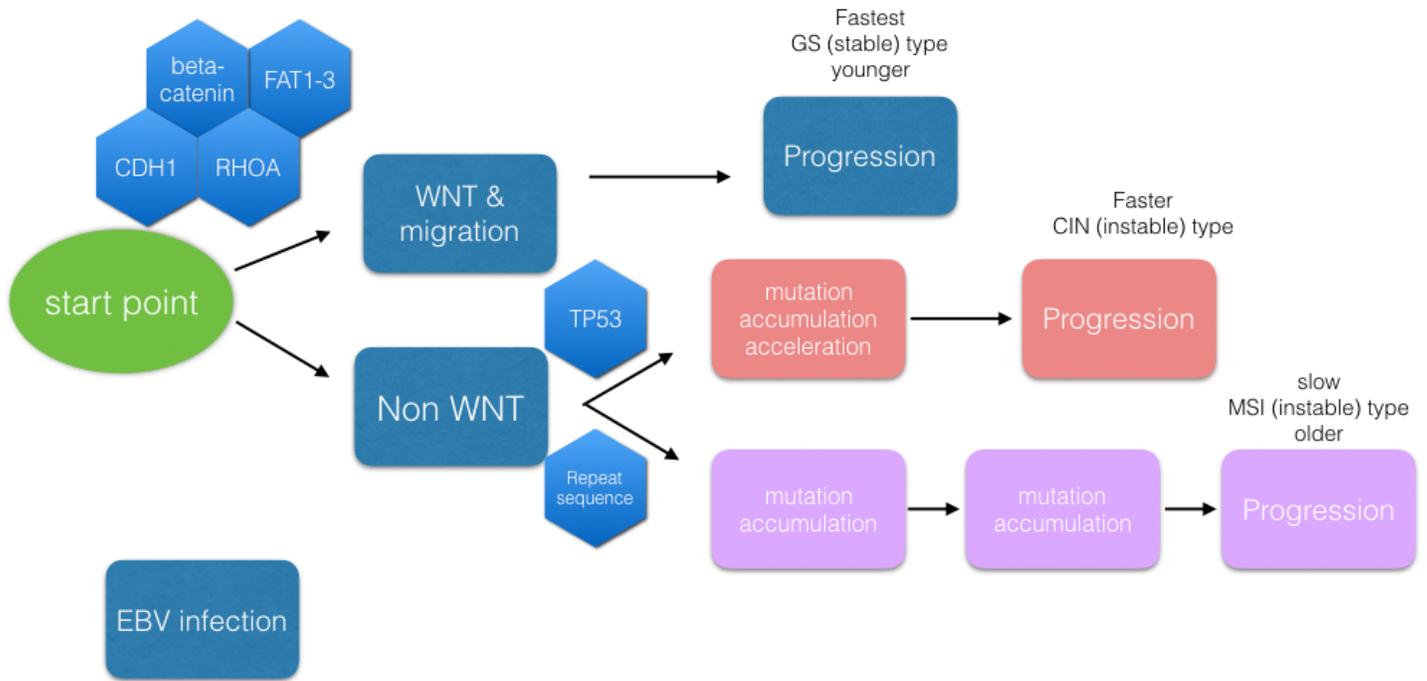


Figure 6

A proposed model for EOGCs and related to TCGA gastric cancer subtypes.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.png](#)
- [FigureS2.png](#)
- [FigureS3.png](#)
- [FigureS4.png](#)
- [FigureS5.png](#)
- [FigureS6.png](#)