의학박사 학위논문

Comprehensive analysis of molecular characteristics and tumor immune microenvironment in stage II and III gastric carcinoma

2 기와 3 기 위암의 분자유전학적 특성과 종양 면역 미세 환경에 대한 통합적 분석

2018 년 2 월

서울대학교 대학원

의학과 병리학 전공

고 지 원

Abstract

Tumor microenvironment immune type (TMIT) is the novel classification scheme based on both the expression of PD-L1 and density of CD8-positive tumor infiltrating lymphocytes. We aimed to apply this classification in stage II and III gastric cancer (GC) patients and assess the prognostic and molecular genetic implications of this classification.

A total of 392 Stage II and III GC patients who were treated by curative surgical resection followed by 5-fluorouracil based adjuvant chemotherapy in Seoul National University Bundang Hospital were included in this study. Tissue microarrays were constructed from the formalin fixed paraffin embedded tissue samples, and the clinical information were collected retrospectively.

Based on the immunohistochemistry (IHC) results of PD-L1 and CD8, TMIT classification of GC was performed as follows: type I (PD-L1^{+}/CD8^{High}), type II (PD-L1^{-}/CD8^{Low}), type III (PD-L1^{+}/CD8^{Low}), type IV (PD-L1^{-}/CD8^{High}). The clinicopathologic features including overall survival according to these four types were analyzed for the evaluation of prognostic performance of TMIT.

For the comprehensive assessment of molecular characteristics of GC in immuno-oncology related perspective, IHC for tumor infiltrating immune cell markers (CD8, Foxp3), markers for epithelial-mesenchymal transition (E-cadherin, vimentin), markers representing cancer stem cells (CD44, Sox2, CD133, OCT3/4), as well as EBV in situ hybridization and microsatellite instability testings were performed.

To elucidate the possible relationship between mutational profiles of GC and immune microenvironment, we analyzed gene expression data and clinical information from two publicly available transcriptome database. In addition, we performed deep targeted sequencing on 80 selected cases from all four TMITs, using the targeted sequencing panel of 170 recurrently mutated genes in various types of solid tumors.

I have found that EBV^{+} and MSI-H GCs are distinct subtypes that are tightly associated with TMIT I ($PD\text{-}L1^{+}/CD8^{High}$), and OS within the $CD8^{High}$ group differs according to PD-L1 expression. Therefore, I conclude that co-assessment of PD-L1 and $CD8^{+}$ TILs is clinically relevant, has a possible prognostic role, and warrants further investigation as a predictive marker for immune checkpoint blockade.

Moreover, I have found an inverse association between EMT phenotype and PD-L1 expression, and close association between EMT features and TMIT II in GCs, which are the opposite results compared to other types of solid tumors. Additional TMIT-associated tumor characteristics include cancer stemess: I have found a tight association between CD44 positivity, a cancer stem cell marker, and TMIT I phenotype, which is consistent with recent findings that $CD44^{+}$ tumor cells play important roles on cancer progression by expressing PD-L1.

Finally, by performing deep targeted sequencing on selected GC tissue samples, I have found that TMIT I tumors have more numbers of somatic mutations compared to other groups and are enriched with somatic mutations of major cancer related genes including *PIK3CA*. TMIT II tumors were enriched with mutations of *RUNX1* gene, and *NTRK3* mutations were relatively

specific to TMIT IV. TMIT III had unique somatic mutational profile, harbouring mutations of genes such as *APC*, *TSC1*, *JAK1*, *MET*, *HRAS* and *RHEB*. Clustering analysis based on somatic mutational profiles have identified two groups, one with higher mutational burden (cluster 1) and the other with lower (cluster 2); cluster 1 had significant association with MSI-H GCs and showed the slight tendency of shorter overall survival.

Recent advances of immunotherapy in solid tumors have facilitated the search for valuable predictive factor for favorable treatment outcome. TMIT was developed for better understanding of immune microenvironment and more effective immune treatment strategy. Based on the findings from this study, we conclude that application of TMIT classification in GC would be helpful for selecting the patients who would have favorable response to immunotherapy, and that this classification could be utilized as the significant prognostic indicator in stage II and III GC.

By clarifying the relationship between molecular profile and microenvironment of GC, we expect to have clues for deeper understanding of the pathogenesis of GC as well as the oncogenesis and progression of other types of solid tumor.

……..

Keywords: gastric cancer, tumor microenvironment, PD-L1, Epstein-Barr virus, microsatellite instability, epithelial-mesenchymal transition, cancer stem cell, prognosis, next generation sequencing

Student Number: 2014-22995

Contents

List of Tables

List of Figures

Chapter 1. Introduction

1.1 Disease burden of gastric cancer

Gastric cancer (GC) is the fifth most common cancer worldwide (Jemal *et al*, 2011), the third most common cancer in South Korea (Jung *et al*, 2016), and one of the leading causes of cancer-related death worldwide (Ferlay *et al*, 2015). Though 5-year survival rate of early GC is over 95%, metastatic GC shows less than one year of median survival, and locally advanced GCs, which are categorized into stage II and III GCs, have less than 40% of 5-year survival (Jung *et al*, 2013).

In addition, the treatment strategy in stage II and III GCs are very limited: current standard therapy includes radical gastrectomy followed by fluoropyrimidine (FP)-based adjuvant chemotherapy. The only targeted therapy in GCs is trastuzumab targeting HER2 protein, however, the HER2 positivity rates in South Korean patients are reported to be around 9% (Kim *et al*, 2012); therefore, the innovative treatment options for the majority of patients are desperately needed.

1.2 Gastric cancer as a candidate for immunotherapy

The close relationship between GC carcinogenesis and chronic inflammation caused by *Helicobacter pylori* and Epstein-Barr virus (EBV) infection has been investigated (van Beek, 2004; Suzuki *et al*, 2009), and this unique immune environment is expected to be an effective target of therapy (Das *et al*, 2006).

Clinical trials of immune checkpoint inhibitors have shown favorable outcomes in some solid tumors, including GC (Hodi *et al*, 2010; Herbst *et al*, 2014; Ansell *et al*, 2015). Currently, cell surface expression of PD-L1, as assessed by immunohistochemistry (IHC), is a predictive factor for the favorable response to immune checkpoint inhibitors; however, not all patients benefit from this therapy (Muro *et al*, 2016). Therefore, recent studies have focused on how to predict which patients would clinically benefit from cancer immunotherapy and what lies beyond the mechanism of immune escape.

1.3 Emergence of novel classification: Tumor microenvironment immune type (TMIT)

The scheme of the tumor microenvironment immune type (TMIT) was developed for better understanding of immune microenvironment. The classification is based on the expression of PD-L1 and tumor-infiltrating lymphocytes (TILs) and consists of four types as follows: type I (PD-$L1^{+}$/TIL^{High}, adaptive immune resistance), type II (PD-$L1^{-}$/TIL^{Low}, immune ignorance type), type III (PD-$L1^{+}$/TIL^{Low}, intrinsic induction of PD-L1 in the absence of TILs), and type IV (PD-$L1^{-}$/TIL^{High}, components other than PD-L1 suppressing the action of TILs) (Taube *et al*, 2012).

In detail, type I (PD-$L1^{+}$/TIL^{High}) is the condition representing adaptive immune escape, which is, though there are many TILs in surrounding microenvironment, tumor cells express PD-L1 so as to evade the anti-tumor effects by TILs. Tumors with this type of microenvironment are expected to have the greatest clinical benefit by immune checkpoint inhibitors. Type II (PD-

L1^{-}/TIL^{Low}) is the status of immunologic ignorance or dormancy, therefore, it is thought that this type of tumors would not have much clinical response by immunotherapy, unless some other measures to potentiate immune response are co-implemented. Type III (PD-L1^{+}/TIL^{Low}) tumors express PD-L1 by intrinsic induction mechanism without infiltration of TILs nearby. Though they compose a minor proportion, they are expected to provide important clues for understanding the expression mechanism of PD-L1. Type IV (PD-L1^{-}/TIL^{High}) tumors are thought to be using various immune-suppressive strategies other than PD-L1 in the midst of high TIL infiltration, and they are important target for studying the dynamic interactions between tumor cells and immune microenvironment.

Though this stratification was criticized for being too simplistic (Teng *et al*, 2015), a comprehensive analysis of The Cancer Genome Atlas (TCGA) dataset for various solid tumors, which used *CD8A* expression as a surrogate marker for TILs, revealed significant association between TMIT I (*PD-L1*^{High}/*CD8A*^{High}) and features like high mutational burden and oncogenic viral infection, suggesting the clinical relevance of this classification (Ock *et al*, 2016b).

Recent studies suggest that the type of TILs, especially CD8-positive (CD8^{+}) cytotoxic T cells, is important for the action of immune checkpoint inhibitors (Tumeh *et al*, 2014). In GC, EBV-positive (EBV^{+}) GCs and MSI-high (MSI-H) GCs are frequently accompanied by heavy infiltration of TILs (Rooney *et al*, 2015; Choi *et al*, 2016), which may be associated with a favorable response to immune checkpoint blockades. However, the rest of GCs

are heterogeneous. Recent studies have proposed that additional characteristics, including epithelial-mesenchymal transition (EMT) features and *TP53* mutations, could be used for further molecular classification (Cristescu *et al*, 2015; Setia *et al*, 2016), although little is known about these categories from a tumor microenvironment-related perspective.

Relating various clinicopathologic features with tumor microenvironmental profiles has become one of the major goals of recent cancer research. EMT phenomenon for instance, it has been proposed that close association exists between EMT signature, as determined by mRNA expression data, and PD-L1 expression in various types of solid tumors, specifically lung adenocarcinoma (Mak *et al*, 2016). Since EMT serves the role of mediating tumor progression and metastases, this close association between EMT and PD-L1 expression is considered to have significant clinical and therapeutic implications. Cancer stem cell (CSC) feature is also one of the key characteristics associated with tumor initiation and progression. Stemness of gastric cancer and its influence on patient prognosis is previously well studied (Ryu *et al*, 2012). Recent report suggests a tight association between stemness markers and immune-evading mechanism: tumor cells with CD44 expression, one of the tumor initiating cell (TIC) marker, constitutively express PD-L1 via STAT3 signaling pathway, thereby evading host anti-tumor immunity (Lee *et al*, 2016b). In addition, with recent advances on genetic research methods including next-generation sequencing (NGS), attempts to use somatic mutational status of cancer to predict the response to immune checkpoint inhibitors have been investigated (Rizvi *et al*, 2015; Dong *et al*, 2017), implying

the importance of linkage between cancer genetics and immuno-oncologic features.

Considering the importance of both PD-L1 expression and $CD8^+$ TILs in defining the tumor immune microenvironment (Taube *et al*, 2012; Teng *et al*, 2015; Ock *et al*, 2016b), I co-assessed PD-L1 expression by immunohistochemistry and the density of $CD8^+$ TILs in stage II and III GC cohort tissue samples and applied the scheme of TMIT classification on GC, based on PD-L1 expression/CD8 status. The major goal of this study was to determine the association between TMIT and various clinicopathologic features of GCs, specifically (i) prognostic significance, (ii) molecular subtypes of GCs including EBV and MSI status, (iii) major tumor-propagation associated features including EMT and cancer stemness. In parallel with this study flow, I attempted to apply TMIT scheme using the publicly available gene expression data of GCs, and studied key features listed above according to TMIT to see if similar patterns of association are observed. Additionally, to determine whether somatic mutational profiles of GCs vary among TMIT classes, I planned to perform NGS on selected cases from stage II and III GC cohort, to study the somatic mutational landscape of key cancer-related genes.

Chapter 2. Materials and methods

2.1 Patients and samples

A total of consecutive 406 patients with stage II or III GC who were treated in Seoul National University Bundang Hospital (Seongnam-si, Republic of Korea) from 2006 to 2013 were screened for inclusion. Among them, the tumor tissue samples of 14 patients were found inadequate for immunohistochemistry, thus excluded. All 392 patients who were included in final analysis underwent curative surgical resection (R0 resection) with D2 lymph node dissection followed by FP-based adjuvant chemotherapy (5-fluorouracil (5-FU), capecitabine, or S-1 with cisplatin, if clinically indicated). Clinicopathologic characteristics, including overall survival (OS) were obtained retrospectively from medical records and pathology reports. OS was defined as the time from surgery to the date of death by any cause or censoring.

Surgically resected GC specimens from patients were formalin-fixed and paraffin-embedded (FFPE). In all cases, one representative 2-mm core was selected from the invasive margin of the tumor, and tissue microarrays (TMA) were constructed as described previously (Superbiochips Laboratories, Seoul, Republic of Korea) (Lee *et al*, 2016a).

All human FFPE tissue samples were obtained from the archive of the Department of Pathology, Seoul National University Bundang Hospital. This study was approved by the institutional review board (IRB) of Seoul National University Bundang Hospital (IRB number: B-1606/349-308). Written patient consent and the consent process were waived by the IRB.

2.2 Immunohistochemistry

IHC for CD8, Foxp3, p53, PD-L1, E-cadherin, vimentin, CD44, Sox2, CD133, and OCT3/4 were performed with an automatic immunostainer (BenchMark XT; Ventana Medical Systems, Tucson, AZ, USA), according to the manufacturer's instructions. The IHC antibodies used in this study were as follows: CD8 (C8/114B, Dako, Carpinteria, CA, USA); Foxp3 (236A/E7, Abcam, Cambridge, UK); p53 (DO7, Dako); and PD-L1 (E1L3N, Cell Signaling Technology, Danvers, MA, USA); E-cadherin (clone 36, BD Biosciences, Franklin Lakes, NJ, USA); vimentin (V9, Thermo Fischer Scientific, Waltham, MA, USA); CD44 (DF1485, Novocastra, Newcastle upon Tyne, UK); Sox2 (6F1.2, Milipore Corp., Billerica, MA, USA); CD133 (PAB12663, Abnova, Taipei City, Taiwan); OCT3/4 (sc-5279, Santa Cruz Biotechnology, Dallas, TX, USA).

To interpret the CD8 and Foxp3 staining, immunostained TMA slides were scanned (Aperio ScanScope CS instrument; Aperio Technologies, Vista, CA, USA), and the average $CD8^+$ and $Foxp3^+$ cell densities (positive cell counts per mm^2) in each core of TMA were counted by an Aperio image analysis system (Aperio Technologies). The $CD8^{High}$ and $CD8^{Low}$ groups were defined using the 25^{th} percentile as the cut-off value, and median value was used as the cut-off for Foxp3.

All other immunostainings were interpreted by light microscope while blinded to patient characteristics at the time of interpretation. Membrane staining of PD-L1 on more than 5% of tumor cells was interpreted as positive (Derks *et al*, 2016; Thompson *et al*, 2016). For E-cadherin, complete loss of

membrane staining or aberrant cytoplasmic staining was regarded as altered expression, while complete membrane staining as strong as that in the non-neoplastic gastric epithelium was considered normal expression (Yi Kim *et al*, 2007). For p53, strong nuclear staining in more than 30% of tumor cells was interpreted as p53 overexpression/positive, and cases with less than 30% positive cells including those showing scattered positive or patchy positive cells were considered negative (Chang *et al*, 2000). For vimentin, either membranous or cytoplasmic staining in more than 10% of tumor cells with any intensity was regarded as positive, and interpretations of CD44 (membranous staining), Sox2 (nuclear staining), CD133 (apical membranous staining), and OCT3/4 (nuclear staining) were performed likewise (Wakamatsu *et al*, 2012; Li *et al*, 2014; Nam *et al*, 2017).

2.3 In situ hybridization

EBV in situ hybridization (ISH) was performed with the INFORM EBV-encoded RNA (EBER) probe (Ventana Medical Systems). To detect *PD-L1* mRNA by ISH on the tissue microarray slides, the *PD-L1* RNAscope 2-plex detection kit (Advanced Cell Diagnostics, Hayward, CA, USA) was used according to the manufacturer's guidelines. The results were interpreted according to the instructions in the RNAscope FFPE Assay Kit and were scored as described previously (Kim *et al*, 2013): 0, no staining; 1, staining in <10% of tumor cells, difficult to identify at 40×; 2, staining in ≥10% of tumor cells, difficult to identify at 20× but easy at 40×; 3, staining in ≥10% of tumor cells,

difficult to identify at 10× but easy at 20×; 4, staining in ≥10% of tumor cells, easy to identify at 10×. A score of 4 was considered *PD-L1* overtranscription.

2.4 Microsatellite instability testing

MSI status was assessed by comparing the allele profiles of five markers (BAT-26, BAT-25, D5S346, D17S250, and S2S123) in tumor cells to those in matched normal samples. Hematoxylin-eosin stain slides were reviewed to select appropriate areas with sufficient tumor cellularity and adequaute non-neoplastic gastric mucosa for macrodissection, and DNA extraction was performed. The polymerase chain reaction (PCR) of DNA were performed with a DNA autosequencer (ABI 3731 Genetic Analyzer; Applied Biosystems, Foster City, CA, USA). According to the Revised Bethesda Guidelines, tumors with additional alleles in two or more markers were classified as MSI-H, tumors with novel bands in one allele were defined as MSI-low (MSI-L), and those with identical bands in all five markers were classified as microsatellite stable (MSS) (Umar *et al*, 2004).

2.5 Processing and analysis of publicly available gene expression data

I used the publicly available level 3 data from TCGA downloaded from the UCSC Cancer Browser (http://genome-cancer.ucsc.edu) on June 3, 2015, which included clinical information and mRNA expression data obtained by RNAseq (Illumina HiSeq V2 platform) of TCGA samples. The mRNA

expression data were presented as reads per kilobase per million (RPKM) and were transformed into log 2 values for the analysis. MSI status was available for 414 stomach adenocarcinoma (STAD) samples, and EBV status was referenced from TCGA clinical data.

In addition, I obtained clinical and mRNA expression data from a SMC cohort (Samsung Medical Center, Seoul, Republic of Korea) shared by Cristescu and colleagues (Cristescu *et al*, 2015) (Gene Expression Omnibus, GSE62254) on April 17, 2015. The mRNA expression data were processed by the Affymetrix Human Genome U133plus 2.0 Array (Santa Clara, CA, USA).

For application of the TMIT classification using the genomic data, after merging the log 2-transformed RPKM values of *PD-L1* and *CD8A*, I divided TCGA and SMC cohort samples into four groups using the aforementioned cut-off values (the median for *PD-L1* and lower 25th percentile for *CD8A*).

2.6 Deep targeted sequencing using cancer-related gene panel

Deep targeted DNA sequencing was performed using cancer-related gene panel, which consisted of 170 widely known cancer driver genes, including *TP53*, *PIK3CA*, *BRCA1*, *KRAS*, *CDH1*, *CDKN2A*, and *ERBB2* (**Table 1**). From the stage II/III GC cohort, I selected 80 eligible cases for sequencing, with sufficient tumor cellularity and relatively short cold ischemic time. After 3 μg of genomic DNAs were extracted from FFPE samples, DNA libraries preparation and target enrichment by hybrid capture method were performed according to Illumina's standard protocol using Agilent SureSelect^{XT} Target

Enrichment Kit (Agilent Technologies, Santa Clara, CA, USA). A total of 961,253 bp target region bases were sequenced for each sample on Hiseq 2500 system (Illumina, San Diego, CA, USA), achieving mean coverage depth ranging from 394x to 2,404x reads (Macrogen Inc., Seoul, Republic of Korea).

The adapter sequences found in raw sequencing reads were removed by cuadapt (Martin, 2011). Trimmed reads were aligned to the reference genome (GRCh37/hg19) using Burrows-Wheeler Aligner-MEM (BWA-MEM) (Li, 2013). Poorly mapped reads that have mapping quality (MAPQ) below 20 were removed using Samtools version 1.3.1 (Li *et al*, 2009). Somatic mutations including short nucleotide variants (SNV), small insertions and deletions (INDELs) were detected by MuTect2 algorithm (Cibulskis *et al*, 2013). All the variants were annotated using SnpEff & SnpSift v4.3i (Cingolani *et al*, 2012) with dbNSFP v2.9.3 (Liu *et al*, 2016).

Following criteria were used to filter out less significant variants and narrow down to clinically relevant variants: (i) variants other than those with allele frequency (AF) between 2% and 20% were excluded, (ii) variants with an allele frequency more than 0.1% in Exome Aggregation Consortium (ExAC) East Asian database were excluded (Lek *et al*, 2016), (iii) all synonymous, intronic, 3`- and 5` untranslated region (UTR) variants were excluded, and (iv) variants which were previously reported as benign or likely benign according to ClinVar (2017-06 release) archive (Landrum *et al*, 2016) and (v) benign variants predicted by PolyPhen-2 HDIV in dbNSFP were filtered out (Adzhubei *et al*, 2010).

2.7 Statistical analysis

The associations between clinicopathological characteristics and TMITs were analysed by Chi-square, linear-by-linear, Kruskal-Wallis, and Wilcoxson/Mann-Whitney tests, as appropriate. Spearman rank correlation was used for the correlation analysis between PD-L1 IHC and *PD-L1* mRNA ISH. Kaplan-Meier analysis of OS according to TMIT and molecular classification was used for survival analysis, and the significance of survival differences was determined by the log-rank test. For comparing mRNA expression levels according to each TMIT groups, Tukey's honest significant difference tests were performed. *P*-value less than 0.05 was considered statistically significant.

When analysing the results of targeted sequencing, fuzzy clustering analysis was performed to organize sequencing data into groups harboring similar somatic mutational profile. Fisher's exact tests were used to assess significant differences in the distribution of a certain somatic mutation among TMIT classes.

Most of the statistical analyses were performed using SPSS statistics 22.0 (IBM, Armonk, NY, USA), and the genomic analysis with data presentation were performed using the R statistical package 3.4.2 (http://www.r-project.org).

Table 1. List of 170 cancer-related gene panel

ABL1	*BCL2*	*CDKN1B*	*ERBB3*	*FLCN*	*JAK3*	*MEN1*	*NOTCH3*	*PPARG*	*SMAD4*
ABL2	*BRAF*	*CDKN2A*	*ERBB4*	*FLT1*	*KDR*	*MET*	*NOTCH4*	*PTCH1*	*SMARCA4*
AKT1	*BRCA1*	*CDKN2B*	*ERCC2*	*FLT3*	*KIT*	*MITF*	*NPM1*	*PTEN*	*SMARCB1*
AKT2	*BRCA2*	*CDKN2C*	*ERG*	*FLT4*	*KMT2A*	*MLH1*	*NRAS*	*RAB35*	*SMO*
AKT3	*BRD2*	*CEBPA*	*ERRFI1*	*FOXL2*	*KRAS*	*MPL*	*NTRK1*	*RAD50*	*SRC*
ALK	*BRD3*	*CHEK2*	*ESR1*	*GNA11*	*MAP2K1*	*MSH2*	*NTRK2*	*RAF1*	*STK11*
APC	*BRD4*	*CREBBP*	*ETV1*	*GNAQ*	*MAP2K2*	*MSH6*	*NTRK3*	*RARA*	*SYK*
AR	*CBFB*	*CRKL*	*ETV4*	*GNAS*	*MAP2K4*	*MTOR*	*NUTM1*	*RB1*	*TET2*
ARAF	*CCND1*	*CSF1R*	*ETV5*	*HDAC9*	*MAP3K1*	*MYC*	*PDGFB*	*RET*	*TMPRSS2*
ASXL1	*CCND2*	*CTNNB1*	*ETV6*	*HGF*	*MAP3K4*	*MYCN*	*PDGFRA*	*RHEB*	*TOP2A*
ATM	*CCND3*	*DDR1*	*EWSR1*	*HRAS*	*MAPK1*	*MYD88*	*PDGFRB*	*RICTOR*	*TP53*
ATR	*CCNE1*	*DDR2*	*EZH2*	*IDH1*	*MAPK3*	*NF1*	*PIK3CA*	*RNF43*	*TSC1*
AURKA	*CDH1*	*DNMT3A*	*FBXW7*	*IDH2*	*MAPK8*	*NF2*	*PIK3CB*	*ROS1*	*TSC2*
AURKB	*CDK12*	*DOT1L*	*FGFR1*	*IGF1R*	*MCL1*	*NFKBIA*	*PIK3CD*	*RSPO1*	*VHL*
AURKC	*CDK4*	*EGFR*	*FGFR2*	*IGF2*	*MDM2*	*NKX2-1*	*PIK3R1*	*RSPO2*	*WT1*
AXL	*CDK6*	*EPHA3*	*FGFR3*	*JAK1*	*MDM4*	*NOTCH1*	*PIK3R2*	*RUNX1*	*XPO1*
BAP1	*CDKN1A*	*ERBB2*	*FGFR4*	*JAK2*	*MED12*	*NOTCH2*	*POLE*	*SMAD2*	*ZNRF3*

Chapter 3. Results

3.1 Clinicopathologic characteristics

The baseline clinicopathologic characteristics of the study population are shown in **Table 2**. The median age was 59 years (range, 20 – 87 years). Of the 392 patients, 182 (46.4%) were AJCC 7th TNM stage II, and 210 (53.6%) were stage III. FP-based regimen was applied as adjuvant chemotherapy; 336 patients (85.7%) were treated with FP only, and 56 patients (14.3%) were treated with FP and cisplatin. The number of $CD8^+$ TILs ranged from 6.90 cells/mm^2 to 1374.94 cells/mm^2 with the median value of 195.23 cells/mm^2. The number of $Foxp3^+$ TILs ranged from 1.22 cells/mm^2 to 785.88 cells/mm^2 with the median value of 60.12 cells/mm^2.

PD-L1 IHC was positive in 98 samples (25.0%), and *PD-L1* mRNA overtranscription (a *PD-L1* mRNA ISH score of 4+) was detected in 14 samples (3.6%). When PD-L1 IHC and mRNA ISH were compared, all cases with mRNA ISH score of 4+ were PD-L1 IHC positive, and the correlation coefficient between the 2 tests was 0.467, which was statistically significant at the 0.01 level (**Table 3**).

Representative figures of immunostainings are shown in **Figure 1**. Altered expression of E-cadherin was detected in 61 of 392 samples (15.6%), vimentin IHC was positive in more than 10% of tumor cells in 93 samples (23.9%), and overexpression of p53 was detected in 108 of 392 samples (27.6%). Among four stemness markers studied, CD44 showed the highest

positivity rate of 65.4% (244 / 373), followed by OCT3/4 (55.4%; 209 / 377), Sox2 (52.0%; 194 / 373), and CD133 (42.0%; 158 / 376).

Table 2. Clinicopathologic characteristics of stage II and III gastric cancer cohort

	Tumor microenvironment immune type				Total	P
	I PD-L1^{+} CD8^{High}	II PD-L1^{-} CD8^{Low}	III PD-L1^{+} CD8^{Low}	IV PD-L1^{-} CD8^{High}		
Age	60 (31 – 82)	57 (30 – 87)	68 (43 – 77)	59 (20 – 85)	59 (20 – 87)	0.159
Sex						0.021
Male	70 (27.7%)	51 (20.2%)	9 (3.6%)	123 (48.6%)	253 (64.5%)	
Female	19 (13.7%)	38 (27.3%)	0 (0.0%)	82 (59.0%)	139 (35.5%)	
Lauren classification						0.929
Intestinal	37 (25.3%)	30 (20.8%)	6 (4.2%)	71 (49.3%)	144 (36.7%)	
Diffuse	38 (17.8%)	57 (26.4%)	2 (0.9%)	119 (55.1%)	216 (55.1%)	
Mixed	13 (43.3%)	2 (6.7%)	0 (0.0%)	15 (50.0%)	30 (7.7%)	
Indeterminate	1 (50.0%)	0 (0.0%)	1 (50.0%)	0 (0.0%)	2 (0.5%)	
Lymphatic invasion						0.698
Absent	23 (19.7%)	31 (26.5%)	0 (0.0%)	142 (53.8%)	117 (29.8%)	
Present	66 (24.0%)	58 (21.1%)	9 (3.3%)	174 (51.6%)	275 (70.2%)	
Vascular invasion						0.855
Absent	77 (23.5%)	70 (21.4%)	6 (1.8%)	174 (53.2%)	327 (83.4%)	
Present	12 (18.5%)	19 (29.2%)	3 (4.6%)	31 (47.7%)	65 (16.6%)	
Perineural invasion						0.266
Absent	40 (30.5%)	21 (16.0%)	3 (2.3%)	67 (51.1%)	131 (33.4%)	
Present	49 (18.8%)	68 (26.1%)	6 (2.3%)	138 (52.9%)	261 (66.6%)	
TNM stage						0.072
II	45 (24.7%)	28 (15.4%)	3 (1.6%)	106 (58.2%)	182 (46.4%)	
III	44 (21.0%)	61 (29.0%)	6 (2.9%)	99 (47.1%)	210 (53.6%)	
Chemotherapy regimen						0.177
FP only	80 (23.9%)	64 (19.1%)	9 (2.7%)	182 (54.3%)	335 (85.7%)	
FP + cisplatin	9 (16.1%)	25 (44.6%)	0 (0.0%)	22 (39.3%)	56 (14.3%)	
Foxp3 IHC						< 0.001
High	79 (40.3%)	11 (5.6%)	5 (2.6%)	101 (51.5%)	196 (50.0%)	
Low	10 (5.1%)	78 (39.8%)	4 (2.0%)	104 (53.1%)	196 (50.0%)	
E-cadherin IHC						0.131
N / C	6 (9.8%)	18 (29.5%)	0 (0.0%)	37 (60.7%)	61 (15.6%)	
M	83 (25.1%)	71 (21.5%)	9 (2.7%)	168 (50.7%)	331 (84.4%)	
Vimentin IHC						0.639
Negative	76 (25.7%)	56 (18.9%)	9 (3.0%)	155 (50.8%)	296 (76.1%)	
Positive	13 (14.0%)	32 (34.4%)	0 (0.0%)	48 (51.6%)	93 (23.9%)	
p53 IHC						0.039
Negative	71 (24.3%)	71 (24.3%)	6 (2.1%)	144 (49.3%)	292 (74.5%)	
Positive	18 (18.0%)	18 (18.0%)	3 (3.0%)	61 (61.0%)	100 (25.5%)	
CD44 IHC						0.002
Negative	9 (7.0%)	41 (31.8%)	3 (2.3%)	76 (58.9%)	129 (34.6%)	
Positive	80 (32.8%)	37 (15.2%)	6 (2.5%)	121 (49.6%)	244 (65.4%)	
Sox2 IHC						0.758
Negative	33 (18.4%)	49 (27.4%)	6 (3.4%)	91 (50.8%)	179 (48.0%)	
Positive	55 (28.4%)	30 (15.5%)	3 (1.5%)	106 (54.6%)	194 (52.0%)	
CD133 IHC						0.201
Negative	47 (21.6%)	47 (21.6%)	2 (0.9%)	122 (56.0%)	218 (58.0%)	
Positive	42 (26.6%)	33 (20.9%)	7 (4.4%)	76 (48.1%)	158 (42.0%)	
OCT3/4 IHC						0.207
Negative	42 (25.0%)	39 (23.2%)	6 (3.6%)	81 (48.2%)	168 (44.6%)	
Positive	47 (22.5%)	41 (19.6%)	3 (1.4%)	118 (56.5%)	209 (55.4%)	
Total	89 (22.7%)	89 (22.7%)	9 (2.3%)	205 (52.3%)	392 (100.0%)	

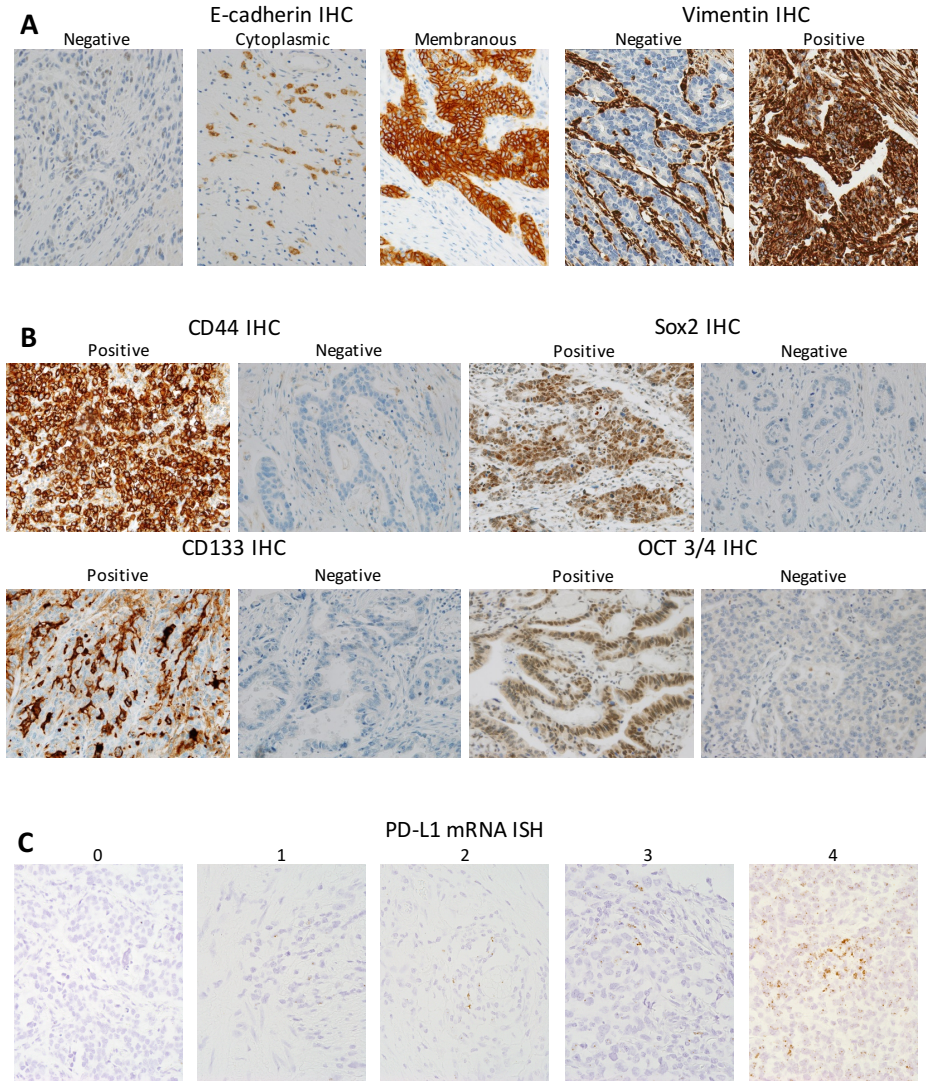
Abbreviations: FP, fluoropyrimidine; IHC, immunohistochemistry; N / C, altered expression (negative or cytoplasmic); M, membranous staining; P, p-value

Table 3. Comparison between two methods of PD-L1 assessment

		PD-L1 IHC			Correlation coefficient
		Negative	**Positive**	**Total**	
PD-L1 mRNA ISH	**0**	280 (94.9%)	57 (58.2%)	337 (85.8%)	0.467
	1+	12 (4.1%)	15 (15.3%)	27 (6.9%)	
	2+	2 (0.7%)	7 (7.1%)	9 (2.3%)	
	3+	1 (0.3%)	5 (5.1%)	6 (1.5%)	
	4+	0 (0.0%)	14 (14.3%)	14 (3.6%)	
	Total	294 (75.0%)	98 (25.0%)	392 (100.0%)	

Abbreviation: IHC, immunohistochemistry; ISH, in situ hybridization

Figure 1. Representative figures of immunohistochemistry and PD-L1 mRNA in situ hybridization



Total loss or altered cytoplasmic expression of E-cadherin and membranous positivity of vimentin immunostainings were considered to be surrogate features of EMT phenomenon (**A**). Cancer stemness was studied by immunostainings of four markers: membranous staining of CD44, nuclear staining of Sox2 and OCT3/4, and apical membranous staining of CD133 (**B**). The interpretation of *PD-L1* mRNA ISH was performed by reading dots on tumor nuclei, and score 4 was defined as *PD-L1* mRNA overtranscription (**C**).

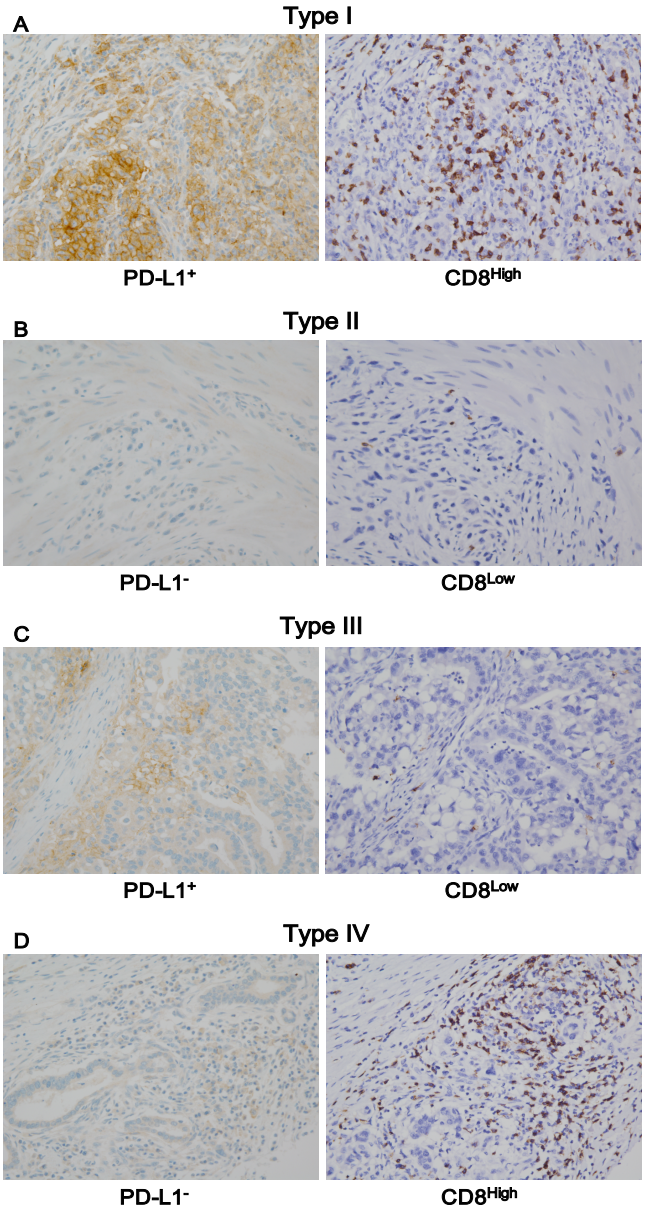
3.2 TMIT in stage II and III GC cohort

I categorized the study population into TMITs I – IV based on the results of PD-L1 IHC and $CD8^+$ TIL density (**Figure 2**). The number and proportion of each type were as follows: type I ($PD\text{-}L1^+/CD8^{High}$), 89 (22.7%); type II ($PD\text{-}L1^-/CD8^{Low}$), 89 (22.7%); type III ($PD\text{-}L1^+/CD8^{Low}$), 9 (2.3%); and type IV ($PD\text{-}L1^-/CD8^{High}$), 205 (52.3%). Type I showed more male predominance than the other types ($P = 0.021$). Type I was associated with $Foxp3^{High}$ status, type II was associated with $Foxp3^{Low}$ status ($P < 0.001$), and p53 IHC positivity showed slight predilection toward TMIT IV ($P = 0.039$).

Striking associations between TMIT I and EBV / MSI status were observed. Twenty-three of the 25 (92%) EBV^+ GCs were type I ($PD\text{-}L1^+/CD8^{High}$); none of the EBV^+ GCs were $CD8^{Low}$, and only two (8.0%) EBV^+ GCs were $PD\text{-}L1^-$. Similarly, MSI-H GCs also had a distinct relationship with TMIT I; 26 of 36 (72.3%) MSI-H cases were $PD\text{-}L1^+$, and 24 cases (66.7%) were classified as TMIT I (**Figure 3A** and **3B**).

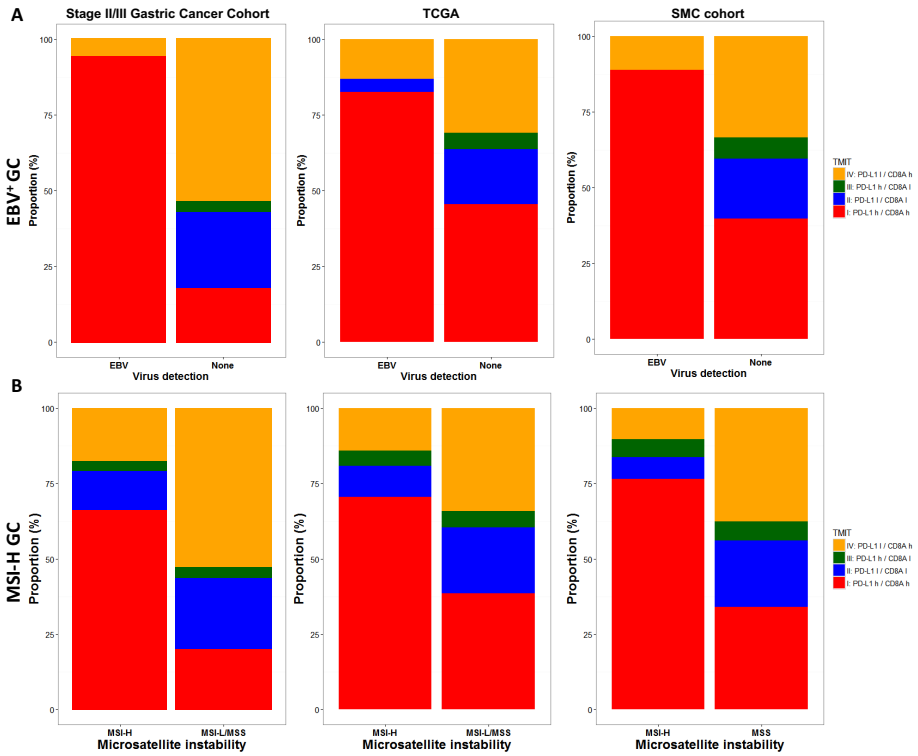
To validate this association between TMIT I and EBV^+ or MSI-H GCs, I performed analysis of the mRNA expression dataset from TCGA and SMC cohort. As shown in **Figure 3A**, the majority of EBV^+ stomach adenocarcinomas in both datasets were classified as TMIT I (81.1% in TCGA and 88.9% in SMC). Genomic analysis according to MSI status showed that, in accordance with the findings from our tissue samples, most of the MSI-H cases were TMIT I (70.5% in TCGA and 76.5% in SMC), followed by type IV, II, and III (**Figure 3B**).

Figure 2. Representative cases in each tumor microenvironment immune types



The TMIT classification is as follows: (**A**) type I (PD-L1^{+}/CD8^{High}), (**B**) type II (PD-L1^{-}/CD8^{Low}), (**C**) type III (PD-L1^{+}/CD8^{Low}), and (**D**) type IV (PD-L1^{-}/CD8^{High}). PD-L1^{+} was defined as PD-L1 membrane staining in more than 5% of tumor cells (**A**, left; **C**, left), and CD8^{High} was defined as a density of CD8^{+} tumor infiltrating lymphocytes (TILs) exceeding the 25th percentile (**A**, right; **D**, right).

Figure 3. Association between TMIT classification and Epstein-Barr virus (EBV) / microsatellite instability (MSI) status



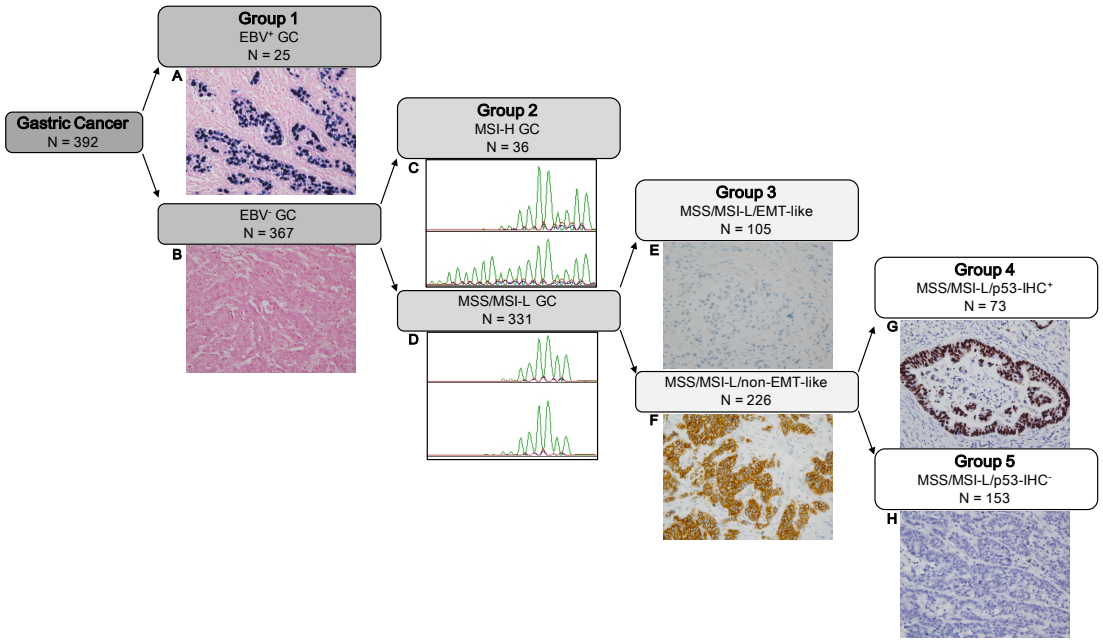
Vast majority of EBV^+ GCs (92%) in stage II and III GC cohort were classified into TMIT I, and concordantly, more than 75% of the cases in both TCGA and SMC datasets were TMIT I (**A**). Similarly, MSI-H GCs were mostly (66.7%) in TMIT I in stage II and III GC cohort. By genomic analysis, MSI-H cases were associated with higher *PD-L1/CD8A* expression, and were thus TMIT I (**B**).

3.3 IHC based molecular classification and TMIT

After observing tight association between TMIT and EBV/MSI status, I modified and adapted previously described molecular classification models for GC (Cristescu *et al*, 2015; Setia *et al*, 2016) in our study population, to further assess the relationship between GCs other than EBV^+/MSI-H GC and TMIT classification. The GC cohort was classified into 5 molecular groups according to the IHC based process described in **Figure 4**: EBV^+ (group 1), MSI-H (group 2), MSS/MSI-L/EMT-like (group 3), MSS/MSI-L/p53-IHC^+ (group 4), and MSS/MSI-L/p53-IHC^- (group 5). EMT-like feature was defined as tumors that histologically resemble mesenchymal cells or show altered E-cadherin expression by IHC.

As a result, of the 392 patients, 25 were in group 1 (6.4%), and 36 were group 2 (9.2%); none of the EBV^+ GCs showed an MSI-H phenotype, and vice versa. The number of patients in groups 3, 4, and 5 were 105 (26.8%), 73 (18.6%), and 153 (39.0%), respectively. To determine the implications of the molecular classification from an immune microenvironment perspective, I compared TMIT and molecular classification. The relationship between the two classifications is shown in **Table 4**. The predilections toward TMIT I in group 1 and 2 were described previously. Within group 3, only 4 of 105 (3.8%) cases were TMIT I, and the proportion of TMIT II cases was relatively high (35/105; 33.3%). In groups 4 and 5, the proportion of each TMIT was similar to that from the whole study population.

Figure 4. Adaptation of immunohistochemistry based molecular classification of gastric cancer



After sorting out the EBV$^+$ GCs (group 1; **A**; EBV ISH), we sorted EBV$^-$ GCs (**B**; EBV ISH) by MSI status. MSI-H cases were categorized as group 2 (**C**), and MSS/MSI-L cases (**D**) were further classified as MSS/MSI-L/EMT-like cases (group 3; **E**; E-cadherin IHC) or MSS/MSI-L/non-EMT-like cases (**F**; E-cadherin IHC). Finally, the MSS/MSI-L/non-EMT-like cases were subclassified according to p53 IHC results as MSS/MSI-L/p53-IHC$^+$ (group 4; **G**; p53 IHC) or MSS/MSI-L/p53-IHC$^-$ (group 5; **H**; p53 IHC).

Table 4. Comparison between molecular classification of gastric cancer and tumor microenvironment immune type

	Tumor microenvironment immune type				Total	P
	I PD-L1^{+} CD8^{High}	II PD-L1^{-} CD8^{Low}	III PD-L1^{+} CD8^{Low}	IV PD-L1^{-} CD8^{High}		
Molecular classification						< 0.001
Group 1 **EBV^{+}**	23 (92.0%)	0 (0.0%)	0 (0.0%)	2 (8.0%)	25 (6.4%)	
Group 2 **MSI-H**	24 (66.7%)	5 (13.9%)	2 (5.6%)	5 (13.9%)	36 (9.2%)	
Group 3 **MSS/MSI-L/EMT-like**	4 (3.8%)	35 (33.3%)	0 (0.0%)	66 (62.9%)	105 (26.8%)	
Group 4 **MSS/MSI-L/p53-IHC^{+}**	12 (17.9%)	10 (14.9%)	3 (4.5%)	42 (62.7%)	67 (17.1%)	
Group 5 **MSS/MSI-L/p53-IHC^{-}**	26 (16.4%)	39 (24.5%)	4 (2.5%)	90 (56.6%)	159 (40.6%)	
Total	89 (22.7%)	89 (22.7%)	9 (2.3%)	205 (52.3%)	392 (100.0%)	

Abbreviations: EBV, Ebstein-Barr virus; MSI-H, microsatellite instability high; MSI-L, microsatellite instability low; MSS, microsatellite stable; EMT, epithelial mesenchymal transition; IHC, immunohistochemistry; P, p-value

3.4 Analysis of prognostic significance

Kaplan-Meier survival analyses according to various measures were performed, and the results showed that patients in the $CD8^{High}$ group had significantly better overall survival (OS) than the $CD8^{Low}$ group ($P < 0.001$; **Figure 5A**) in stage II and III GC patients with standard treatment. PD-L1 IHC positivity itself was not significantly associated with survival ($P = 0.579$; **Figure 5B**). There was no significant survival difference between EBV^{+} and EBV^{-} GCs ($P = 0.486$; **Figure 5C**). Analysis according to MSI status showed that MSI-L patients had worse OS when compared to MSI-H and MSS patients, with borderline statistical significance ($P = 0.063$; **Figure 5D**).

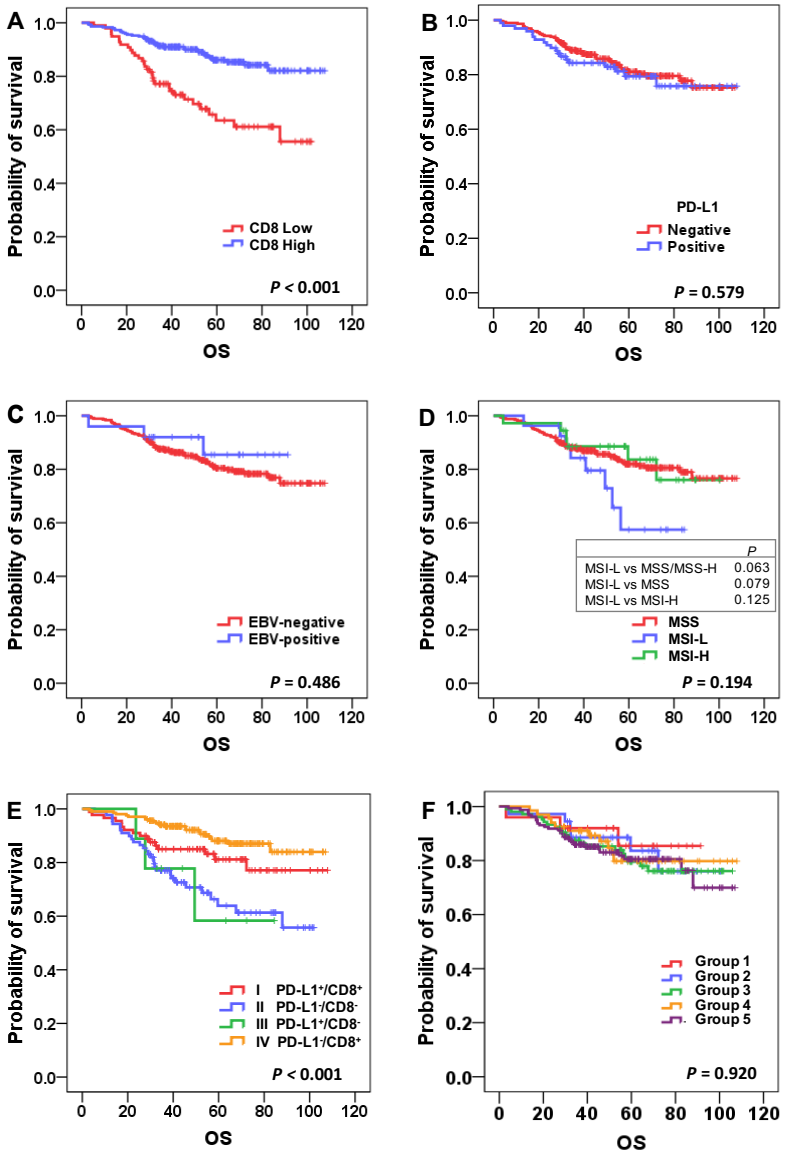
I also performed Kaplan-Meier survival analysis according to TMIT and molecular classification. Of the four TMITs, type IV ($PD\text{-}L1^{-}/CD8^{High}$) had the best OS, and type II ($PD\text{-}L1^{-}/CD8^{Low}$) had the worst OS ($P < 0.001$; **Figure 5E**). Interestingly, when TMITs I and IV (the $CD8^{High}$ groups) were compared, type IV ($PD\text{-}L1^{-}/CD8^{High}$) had better OS, with marginal statistical significance ($P = 0.070$). However, according to the molecular classification, no significant survival differences were detected among the 5 groups ($P = 0.791$; **Figure 5F**).

Subgroup survival analyses stratified by TNM stage were performed to see if the prognostic significance of TMIT classification is still valid. In stage II and III GC cohort where classification was performed by IHC using FFPE tissue samples, similar survival trends were observed with retained statistical significance. However, from the TCGA and SMC cohort datasets, where mRNA expression levels were used for classificaiton, no significant survival

discrimination within stage II, stage III, and stage II/III combined population was observed (**Figure 6**).

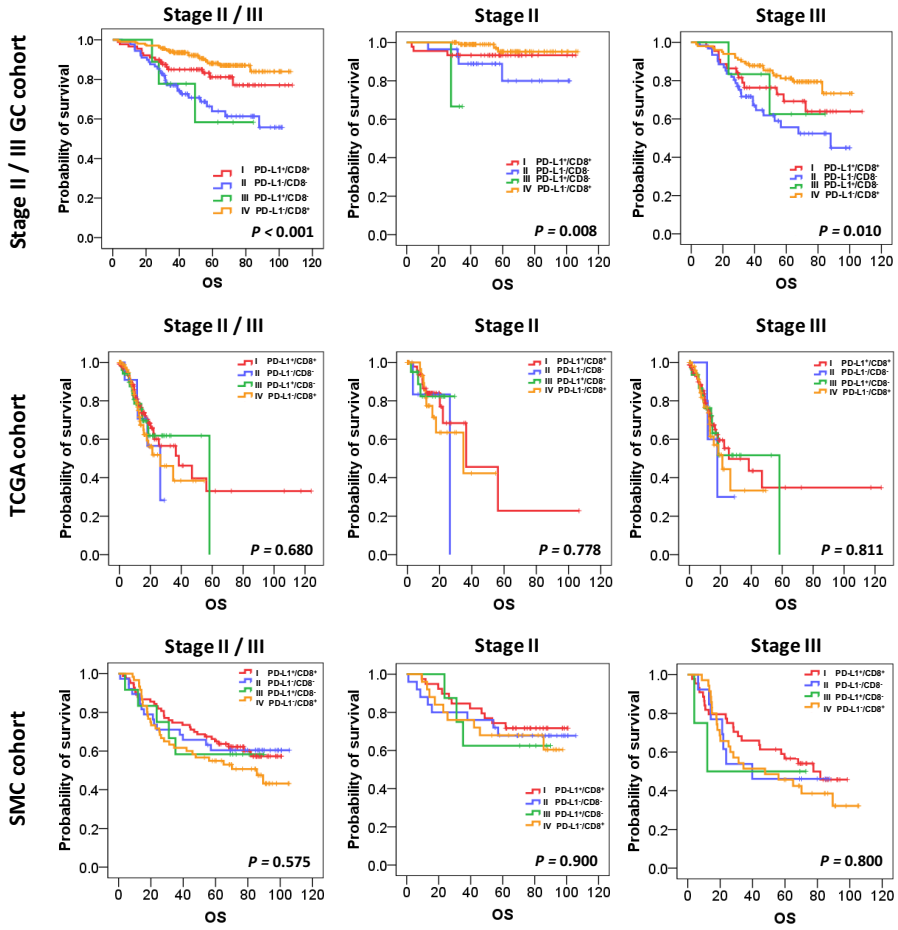
Univariate analysis of OS by Cox proportional hazard model showed that age, vascular invasion, perineural invasion, chemotherapy regimen, TNM stage, $CD8^+$ TILs, $Foxp3^+$ TILs, and TMIT IV are the key clinicopathologic features that are significantly associated with OS (**Table 5**). By multivariate analysis, older age, the presence of vascular invasion, addition of cisplatin to FP-based chemotherapy, higher TNM stage, and $CD8^{High}$ status were significantly correlated with OS. Furthermore, when compared to the type I, II and III, TMIT IV was an independent prognostic factor for OS, with statistical significance (hazard ratios, 2.11, 2.55 and 3.50; 95% confidence intervals, 1.30 – 4.34, 1.41 – 4.62 and 1.03 – 11.88; $P = 0.042$, 0.002 and 0.045 respectively).

Figure 5. Kaplan-Meier survival analysis of overall survival according to major clinicopathologic features



Higher densities of $CD8^+$ cells were associated with better overall survival (**A**; $P < 0.00$), whereas PD-L1 positivity and EBV status were not significant prognostic factors (**B** and **C**; $P = 0.579$ and 0.486, respectively). MSI-L cases showed poor prognosis compared to others (**D**). There were significant survival differences among the four TMITs (**E**; $P < 0.001$), whereas there were no discernible differences according to IHC based molecular classification (**F**).

Figure 6. Subgroup survival analyses according to tumor microenvironment immune types stratified by stage



Compared to stage II and III GC cohort where significant OS differences according to TMIT were observed in all subgroup analyses, Kaplan-Meier study using TCGA and SMC cohort mRNA expression datasets failed to discriminate significant survival differences.

Table 5. Univariate and multivariate analysis of overall survival by Cox proportional hazards model

Variable		Univariate			Multivariate (TMIT)			Multivariate (CD8$^+$ TILs)		
		HR	95% CI	P	HR	95% CI	P	HR	95% CI	P
Age		1.03	1.01 – 1.05	0.002	1.04	1.02 – 1.06	0.001	1.04	1.02 – 1.06	0.001
Sex	Female vs male	1.03	0.63 – 1.67	0.920						
Lymphatic invasion	Present vs absent	1.72	0.98 – 3.09	0.070						
Vascular invasion	Present vs absent	3.70	2.29 – 6.00	< 0.001	2.15	1.29 – 3.58	0.003	2.13	1.28 – 3.54	0.003
Perineural invasion	Present vs absent	3.08	1.58 – 6.01	0.001	1.62	0.78 – 3.38	0.197	1.57	0.76 – 3.27	0.224
Chemotherapy	FP only vs FP+C	4.69	2.81 – 7.82	< 0.001	3.85	2.19 – 6.77	< 0.001	3.82	2.19 – 6.65	< 0.001
TNM stage	III vs II	4.91	2.58 – 9.35	< 0.001	2.36	1.15 – 4.81	0.019	2.44	1.20 – 4.98	0.014
PD-L1 IHC	P vs N	1.16	0.69 – 1.97	0.579						
CD8$^+$ TILs	High vs Low	0.34	0.21 – 0.55	< 0.001				0.44	0.26 – 0.75	0.003
Foxp3$^+$ TILs	High vs Low	0.52	0.32 – 0.85	0.009	0.82	0.43 – 1.57	0.553	1.08	0.61 – 1.91	0.802
EBV status	P vs N	0.67	0.21 – 2.11	0.489						
MSI status	MSI-L vs MSS	1.92	0.92 – 4.03	0.085						
	MSI-H vs MSS	0.90	0.39 – 2.09	0.808						
E-cadherin IHC	M vs N/C	1.07	0.56 – 2.04	0.838						
Vimentin IHC	P vs N	1.37	0.83 – 2.28	0.230						
p53 IHC	P vs N	1.19	0.71 – 2.00	0.510						
CD44 IHC	P vs N	0.70	0.43 – 1.14	0.147						
Sox2 IHC	P vs N	0.86	0.53 – 1.39	0.527						
CD133 IHC	P vs N	0.84	0.51 – 1.38	0.492						
OCT3/4 IHC	P vs N	0.70	0.43 – 1.14	0.154						
TMIT	I vs IV	1.80	0.95 – 3.44	0.073	2.11	1.03 – 4.34	0.042			
	II vs IV	3.61	2.07 – 6.29	< 0.001	2.55	1.41 – 4.62	0.002			
	III vs IV	3.74	1.12 – 12.50	0.032	3.50	1.03 – 11.88	0.045			
Molecular classification	Group 2 vs 1	1.27	0.32 – 5.01	0.737						
	Group 3 vs 1	1.57	0.47 – 5.27	0.464						
	Group 4 vs 1	1.35	0.38 – 4.85	0.643						
	Group 5 vs 1	1.59	0.49 – 5.23	0.444						

Abbreviations: HR, hazard ratio; CI, confidence interval; FP, fluoropyrimidine; C, cisplatin IHC, immunohistochemistry; P, positive; N, negative; TIL, tumor infiltrating lymphocytes; MSI, microsatellite instability; MSI-L, MSI-low; MSI-H, MSI-high; MSS, microsatellite stable; M, membranous statining; N/C, altered expression (negative or cytoplasmic); TMIT, tumor microenvironment immune types; P, p-value

3.5 Analysis of EMT and cancer stem cell markers by IHC

The results of EMT and cancer stemness studied by IHC methods are depicted in **Figure 7**. When TMIT I and II were compared regarding the EMT markers, among 61 cases showing altered E-cadherin expression, only 6 cases (9.8%) were TMIT I and up to 18 cases (29.5%) were in TMIT II ($P = 0.008$). Vimentin positivity was observed in 93 cases: 13 (7.0%) and 32 (34.4%) cases were in TMIT I and II respectively ($P = 0.001$), implying that tumors with EMT phenotype are more likely to be in TMIT II rather than TMIT I.

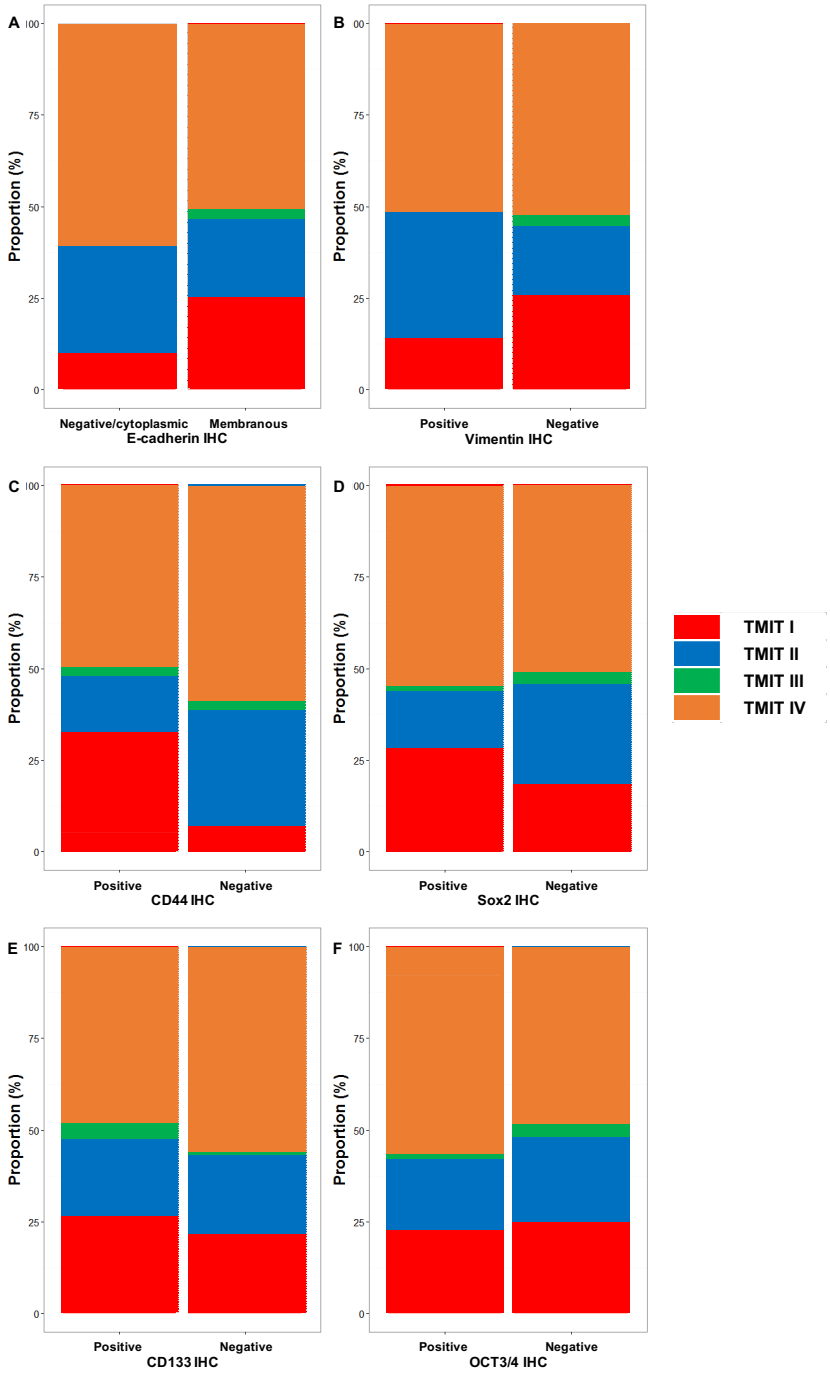
With regards to cancer stem cell markers, CD44 IHC showed marked predilection toward TMIT I; among 244 $CD44^+$ cases, up to 80 (32.8%) were in TMIT I, and within TMIT I group, 80 cases (89.9%) were $CD44^+$, leaving only 9 cases (10.1%) showing no expression of CD44. When positivity rates of CD44 in TMIT I and II were compared, statistically significant differences were observed ($P < 0.001$). Similar pattern was observed by Sox2 IHC: 28.4% of $Sox2^+$ cases were in TMIT I, and Sox2 positivity rate in TMIT I (61.8%) was significantly higher than that in TMIT II (37.2%) ($P = 0.002$). Meanwhile, CD133 and OCT3/4 IHC results did not show different positivity rate among TMITs.

To see if similar patterns of differential gene expression levels are observed according to TMIT, I assessed the mean mRNA expression levels of *CDH1* and *VIM* in each TMIT using TCGA and SMC datasets (**Figure 8**). In TCGA dataset, TMIT IV showed the lowest *CDH1* expression, and only the difference between type IV and II showed statistical significance. In contrast,

analysis of the SMC dataset showed that *CDH1* expression levels did not differ among the 4 TMITs. *VIM* expression in TMITs I and IV of TCGA cohort was significantly higher than in TMITs II and III. In SMC cohort, mean *VIM* expression in TMIT II was the lowest of all with statistically significant difference compared to TMITs III and IV.

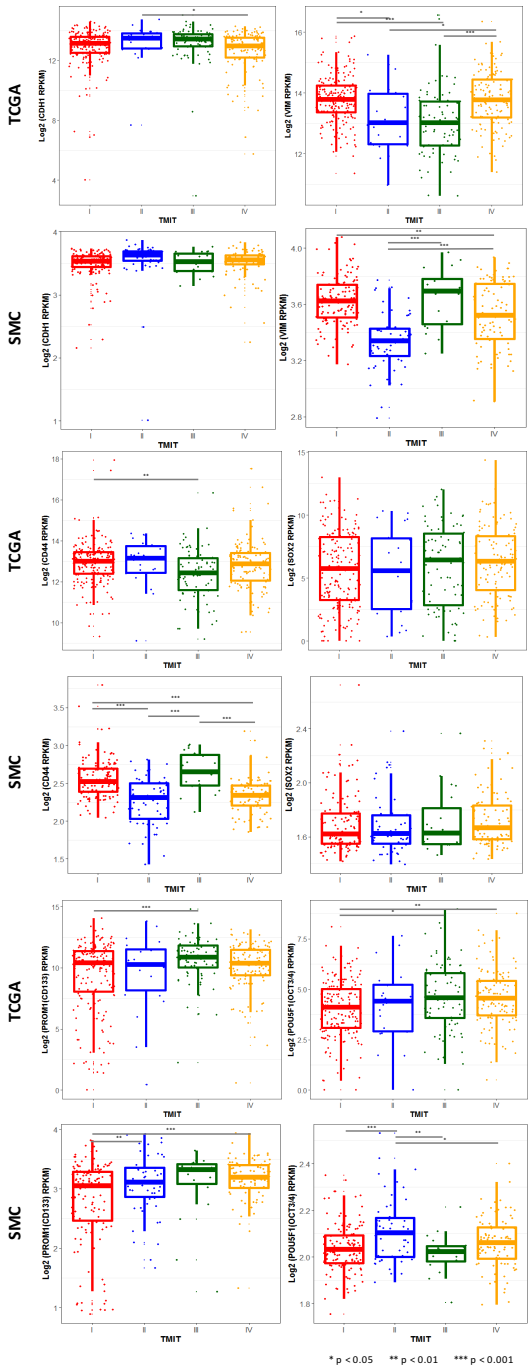
Stemness related genes were shown to be differentially expressed between certain TMITs in TCGA database: higher expression of *CD44* in TMIT I compared to III, and lower expression of *POU5F1* (encoding OCT3/4) in TMIT I compared to II and IV. However, SMC cohort analysis did not show any similar pattern or reproducible data: *CD44* level was significantly higher in TMIT I compared to II and IV, while differences in other genes were inconsistent with IHC or TCGA gene expression analysis results with their clinical significance remain unclear.

Figure 7. Immunohistochemistry results of mesenchymal and stemness markers according to tumor microenvironment immune types



Tumors with altered E-cadherin expression and vimentin positive cases were more frequently observed within TMIT II compared to TMIT I. Among the stem cell markers, tight association between $CD44^{+}$ GCs and TMIT I is notable.

Figure 8. mRNA expression levels of epithelial-mesenchymal transition and cancer stemness associated genes



The mRNA expression levels according to four TMITs from two publicly available datasets are plotted, with statistical analysis by Tukey's honest significant difference tests.

3.6 Targeted sequencing of cancer-related genes in stage II and III GC

3.6.1 Somatic mutational profile of stage II and III GC

A total of 686 somatic mutations in 145 cancer-related genes from 80 patients were found (**Figure 9**). 546 mutations were single nucleotide variations (SNV) and 140 were small insertions and deletions (indels). Each of the 80 cases harbored mutations of 8 genes on average.

Most frequently mutated genes included *GNAQ* (40%), *PIK3CA* (28%), *TP53* (23%), *MAP3K1* (18%), *KMT2A* (16%), *ATR* (16%), *GNAS* (15%), *APC* (15%), *CDH1* (14%), *RUNX1* (14%), *ATM* (13%), *MAP3K4* (13%), *NOTCH1* (13%), *NOTCH3* (13%), *TOP2A* (13%), *TSC2* (13%), *MED12* (11%), *NF1* (11%), *PTEN* (11%), *RAD50* (11%), *RICTOR* (11%), *FLT4* (10%), *MSH6* (10%), *PIK3CB* (10%), *BRAF* (9%), *BRD4* (9%), *MSH2* (9%), *MTOR* (9%), *NOTCH2* (9%), *POLE* (9%), *PTCH1* (9%), *RNF43* (9%), *SMO* (9%) and *XPO1* (9%).

Among the rest, previously reported significant somatic alteration in GC includes *SMAD2* (8%), *SMAD4* (6%), *CTNNB1* (5%), *KRAS* (4%), *FGFR2* (4%), *ERBB3* (4%), *JAK2* (4%), *ERBB2* (3%), *EGFR* (1%) and *CDKN2A* (1%).

Figure 9. Somatic mutational landscape of stage II and III gastric cancer cohort

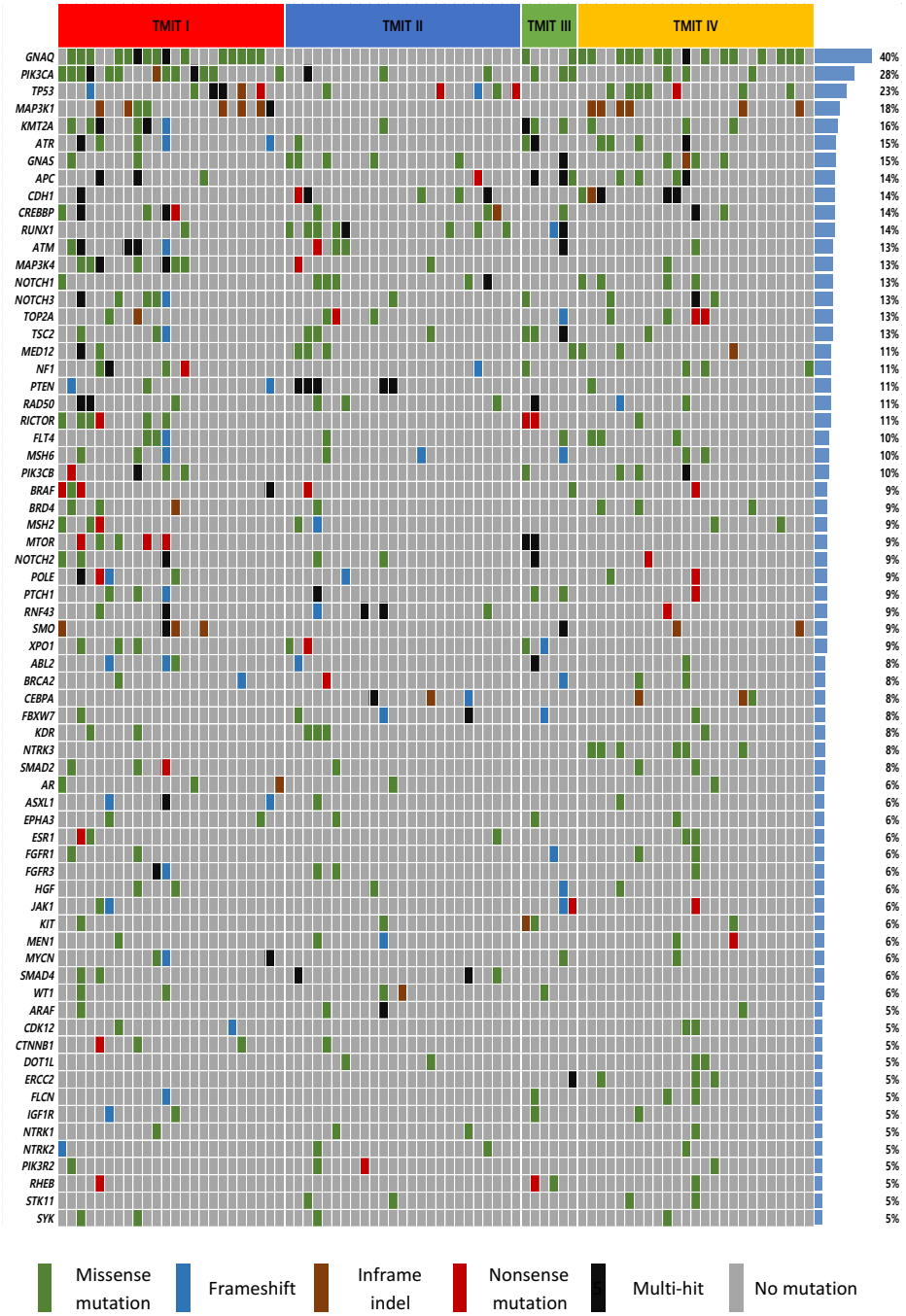
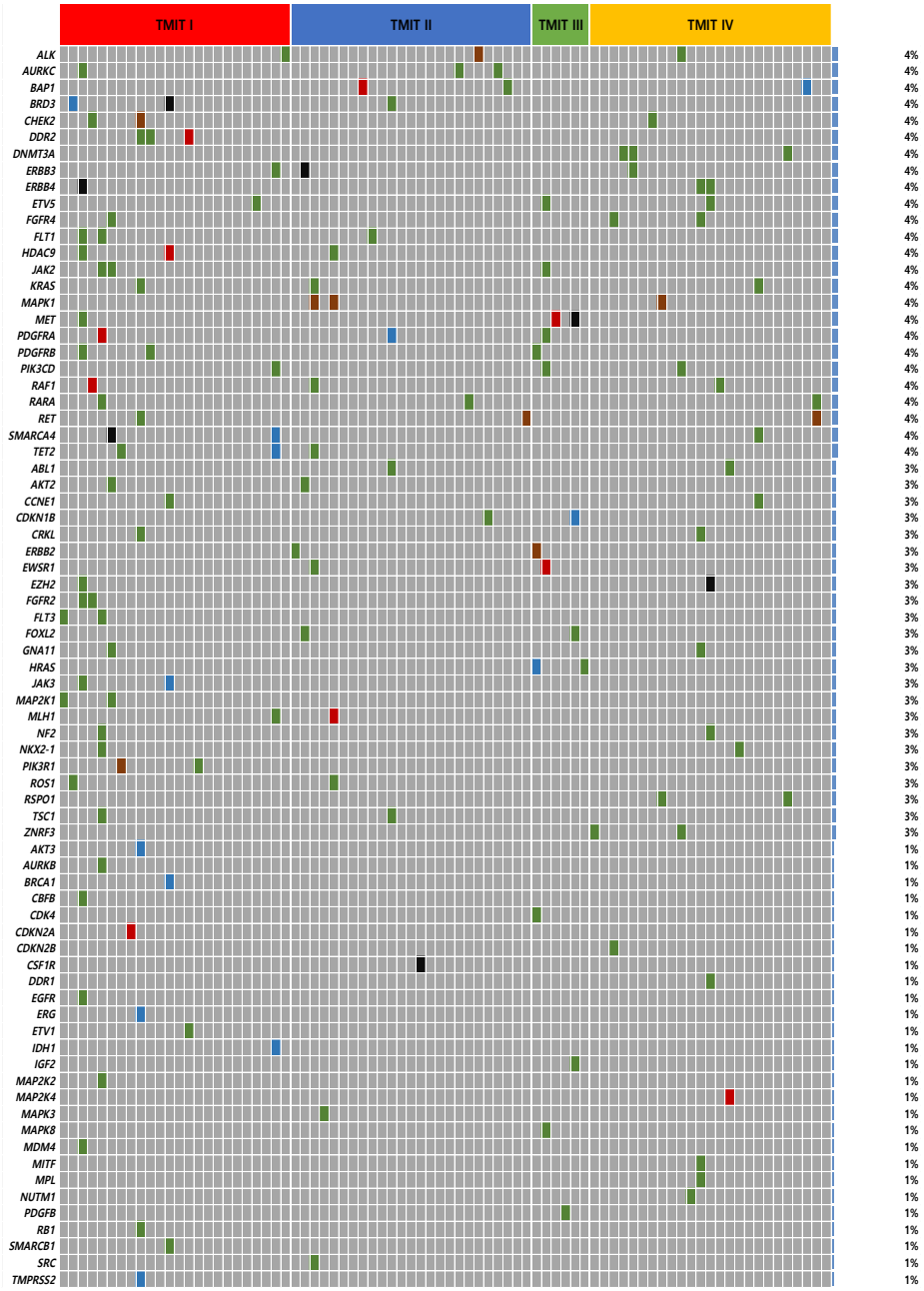


Figure 9. Somatic mutational landscape of stage II and III gastric cancer cohort (cont.)



Among 170 cancer-related genes studies, SNVs and indels in 140 genes were found. Previously well studied genes including *TP53* and *PIK3CA* are noted, as well as genes such as *CREBBP*, *MED12*, *RUNX1*, *FLT4*, and *BRD4* which were less previously reported as recurrently mutated genes in GCs.

3.6.2 Differences in mutational profiles according to tumor microenvironment immune types

Total numbers of somatic mutations in each TMIT varied, with TMIT I carrying the most, 286 mutations of 116 genes, followed by 168 mutations of 94 genes in TMIT IV, 168 mutations of 81 genes in TMIT II, and 97 mutations of 57 genes in TMIT III. TMIT I and IV shared 16 common mutations, including *TP53* R248W, *TP53* Y220C, *PIK3CA* R349Q and *PTEN* H93R, while TMIT II and III had no genetic alteration in common.

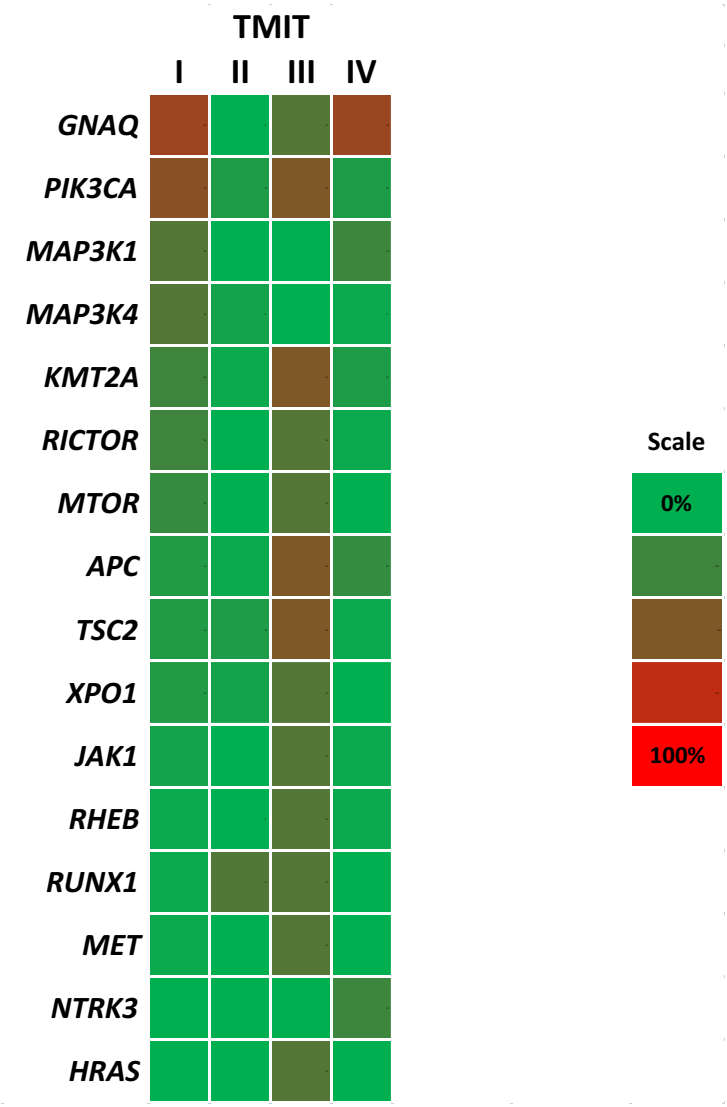
Fisher's exact tests were performed in an attempt to specify which genes are significantly more frequently mutated in certain TMIT group, and the results are plotted in **Figure 10**. *GNAQ* mutations were significantly more frequent in TMIT I (62%) and IV (60%), while none was found in TMIT II (0%) ($P < 0.001$). Mutations in *PIK3CA*, a well-known recurrently mutated gene in GC, were observed more frequently in TMIT I (54%) compared to IV (12%) and II (12%) (I vs II, $P = 0.002$; I vs IV, $P = 0.002$). Other cancer-related genes with enriched mutational profile in TMIT I includes, *MAP3K4* (33%) (I vs IV, $P = 0.01$), *MAP3K1* (33%) (I vs II, $P = 0.001$), *KMT2A* (25%) (I vs II, $P = 0.048$) and *MTOR* (21%) (I vs II, $P = 0.02$; I vs IV, 0.02).

RUNX1, a tumor suppressor gene, was the only gene showing significantly more frequent mutations within TMIT II patients (32%) (II vs I, $P = 0.02$; II vs IV, $P = 0.004$), and *NTRK3* mutations were observed only in TMIT IV (24%) (IV vs I and IV vs II, $P = 0.02$).

TMIT III, when compared with TMIT 1, no significantly different mutational profiles were observed. Compared to TMIT II and IV, TMIT III

showed more distinct somatic mutational profile, harboring significantly more mutations in following genes: *APC* (50%) (III vs II, $P = 0.02$), *TSC2* (50%) (III vs IV, $P = 0.02$), *KMT2A* (50%) (III vs II, $P = 0.02$), *JAK1* (33%) (III vs II, $P = 0.03$), *MET* (33%) (III vs II and III vs IV, $P = 0.03$), *HRAS* (33%) (III vs II and III vs IV, $P = 0.03$), *MTOR* (33%) (III vs II and III vs IV, $P = 0.03$), *RHEB* (33%) (III vs II, $P = 0.03$), *RUNX1* (33%) (III vs IV, $P = 0.03$), and *XPO1* (33%) (III vs IV, $P = 0.03$).

Figure 10. Differentially mutated genes according to four tumor microenvironment immune types



Heatmap shows frequency of mutations observed in each TMITs. *PIK3CA* mutations were enriched in TMIT I and III, while *RUNX1* mutations were more frequently observed in TMIT II. *NTRK3* mutations were found to be the highest in TMIT IV.

3.6.3 Differences in mutational profiles according to Epstein-Barr virus gastric cancer microsatellite instability status

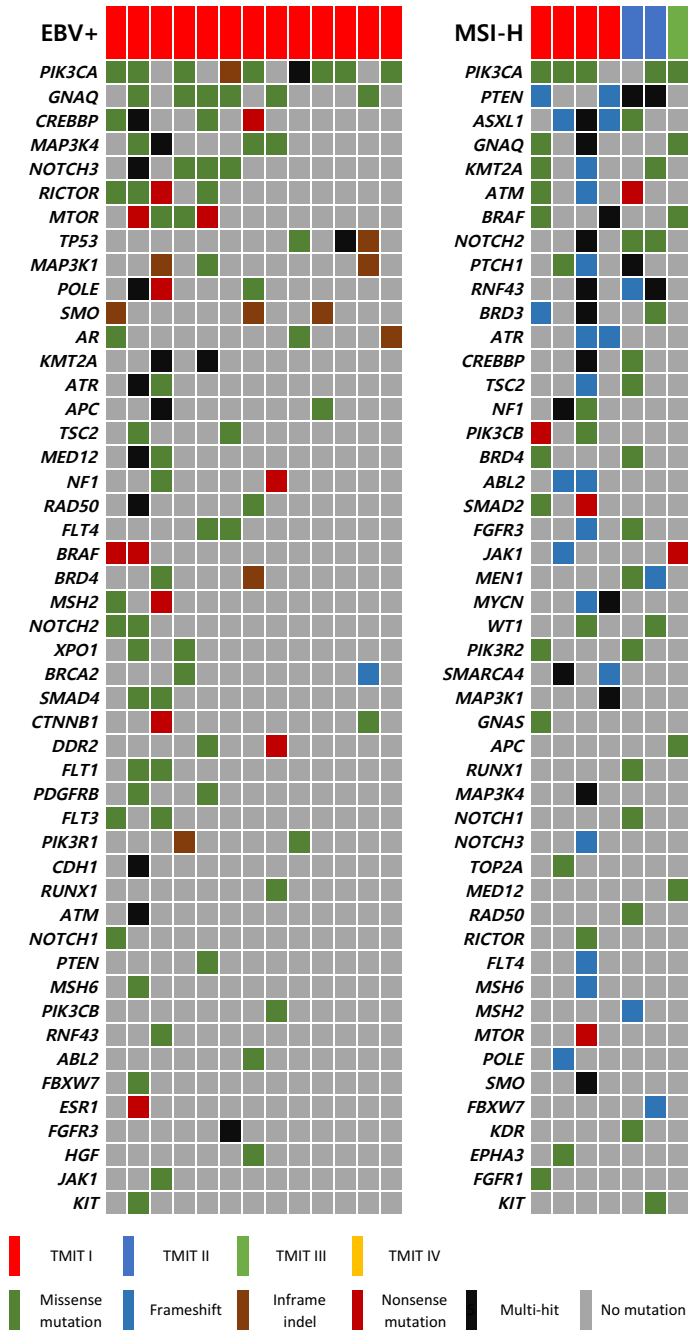
Among 80 stage II and III GC patients who were eligible for deep targeted sequencing, 13 were EBV^+ GCs and they were all in TMIT I. Seven MSI-H GC samples were also sequenced, and four of them were TMIT I, with two TMIT II patients and one in TMIT III. Each EBV^+ GC had mutations in 11 genes on average, and MSI-H GC had mutations in around 17 genes, ranging from 8 to 32 genes, implying higher mutational burden of cancer-related genes in MSI-H cases (**Figure 11**).

PIK3CA was most commonly mutated gene in both EBV^+ GCs (69%) and MSI-H GCs (71%). In addition, *CREBBP* (31% in EBV^+ GCs and 29% in MSI-H GCs), *MAP3K4* (31% in EBV^+ GCs and 14% in MSI-H GCs), *NOTCH3* (31% in EBV^+ GCs and 14% in MSI-H GCs) and *KMT2A* mutations (15% in EBV^+ GCs and 43% in MSI-H GCs) were among the frequently observed mutations in both groups.

TP53 mutation is one of the most commonly observed genetic alteration in GCs, however, none of the MSI-H GCs from this cohort had *TP53* mutation, while three of the 13 EBV^+ GCs (23%) had *TP53* mutations, Y220C, T256I, R248W, and in-frame deletion (PHHERC177del), though this difference was not statistically significant finding. Three of the seven MSI-H GCs had *BRAF* mutations (P = 0.03); however, none of them were V600E missense mutation. Four of the MSH-H GCs had *PTEN* mutations (57%), while only one EBV^+ GC (1%) had mutations in *PTEN* (P = 0.03).

Other genes which had enriched somatic mutation in MSI-H GCs compared to EBV^+ GCs included *ASXL1* (57% vs 0%, $P < 0.001$), *PTCH1* (43% vs 0%, $P = 0.003$), and *BRD3* (43% vs 0%, $P = 0.003$). Of seven cases with *RNF43* mutations among the 80 stage II / III GCs, three of them were MSI-H ($P = 0.013$). *POLE* mutations were found in seven out of this present cohort, and three and one of them were EBV^+ and MSI-H GCs, respectively.

Figure 11. Somatic mutational landscape in Epstein-Barr virus associated gastric cancer and microsatellite instability-high gastric cancer



Heatmap shows the distribution of SNV and indels among EBV^{+} GCs and MSI-H GCs. MSI-H cases had higher mean number of mutated genes (22; range 8 – 32), compared to EBV^{+} GCs (11; 2 – 37).

3.6.4 Clustering analysis based on somatic mutational profile

Fuzzy clustering analysis was performed to suggest a novel classification of stage II and III GCs based on somatic mutational profiles (**Figure 12**). Two distinct clusters were identified: 25 cases were classified into cluster 1 (31.3%) and 55 in cluster 2 (68.7%). Most notably, cluster 1 was composed of GCs with higher number of genetic alterations; while GCs in cluster 1 had 17 mutated genes per cases on average, cluster 2 GCs had four mutated genes on average ($P < 0.001$). When clinicopathologic features were compared, none of the features showed significant differences between two clusters (**Table 6**).

Cluster 1 was enriched with mutations of cancer-related genes including *PIK3CA* (60% vs 13%, $P < 0.001$), *KMT2A* (44% vs 4%, $P < 0.001$), *ATR* (44% vs 2%, $P < 0.001$), *RICTOR* (36% vs 0%, $P < 0.001$), *MAP3K* (32% vs 4%, $P = 0.001$), TSC2 (32% vs 4%, $P = 0.001$), *GNAS* (28% vs 9%, $P = 0.04$), *PTEN* (28% vs 4%, $P = 0.003$), *ATM* (28% vs 5%, $P = 0.003$), *PTCH1* (28% vs 0%, $P < 0.001$), *RAD50* (24% vs 5%, $P = 0.02$), *PIK3CB* (28% vs 4%, $P = 0.01$), *BRAF* (24% vs 2%, $P = 0.003$), *BRD4* (24% vs 2%, $P = 0.003$), *MTOR* (24% vs 2%, $P = 0.003$), *NOTCH2* (24% vs 2%, $P = 0.003$), *RNF43* (24% vs 2%, $P = 0.003$) and *ABL2* (24% vs 0%, $P < 0.001$). *TP53* mutations, however, were significantly more common in cluster 2 compared cluster 1 (29% vs 8%, $P = 0.045$).

Next, I compared this cluster model with previously introduced classification schemes of GCs, TMIT and molecular classification (**Figure 13**). No discernable or significant association between TMIT classification and

somatic mutational cluster model was found ($P = 0.075$). When compared with molecular classification of GC, all MSI-H GCs (group 2) were classified into cluster 1, and group 3, which represents GCs showing EMT-like features, showed predilection toward cluster 2 ($P = 0.017$).

To assess the prognostic significance of the cluster model, Kaplan-Meier survival analysis was performed and cluster 2 showed slightly worse OS compared to cluster 1 ($P = 0.106$) (**Figure 14**).

Figure 12. Clustering analysis based on somatic mutational profile

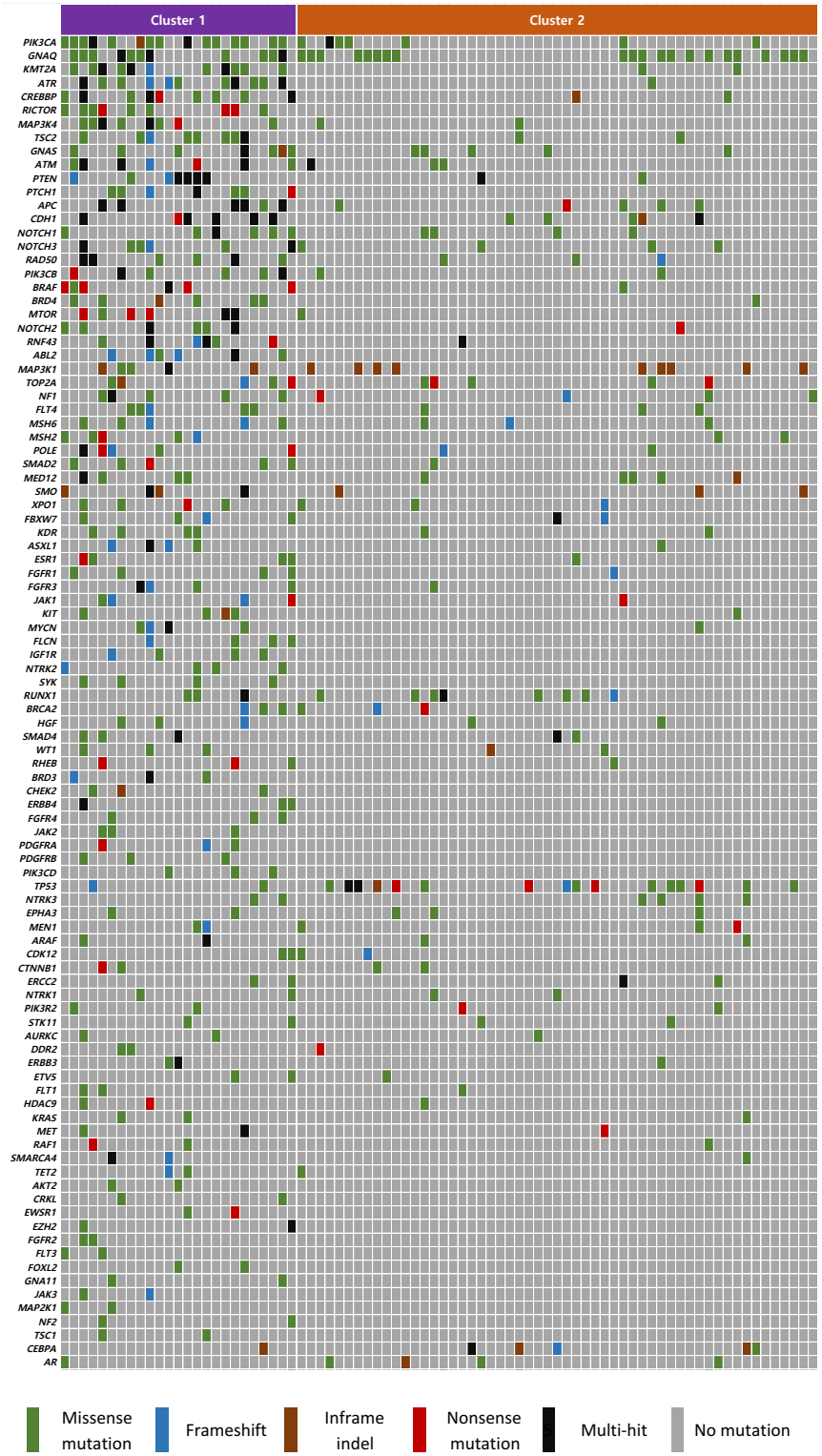
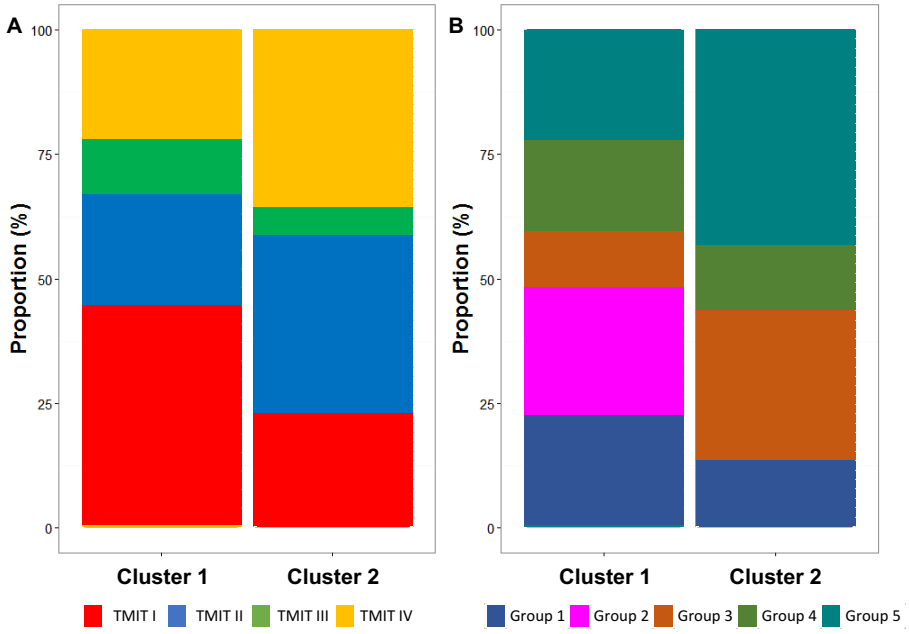


Figure 13. Clustering analysis based on somatic mutational profile (cont.)
Fuzzy clustering method was adapted to classify stage II and III GCs solely based on somatic mutational profile. As a result, two clusters were discriminated: cluster 1 shows markedly larger number of somatic mutations compared to cluster 2, except for *TP53*, the mutations of which gene is more frequently observed in cluster 2.

Table 6. Clinicopathologic characteristics according to cluster groups based on somatic mutational profile

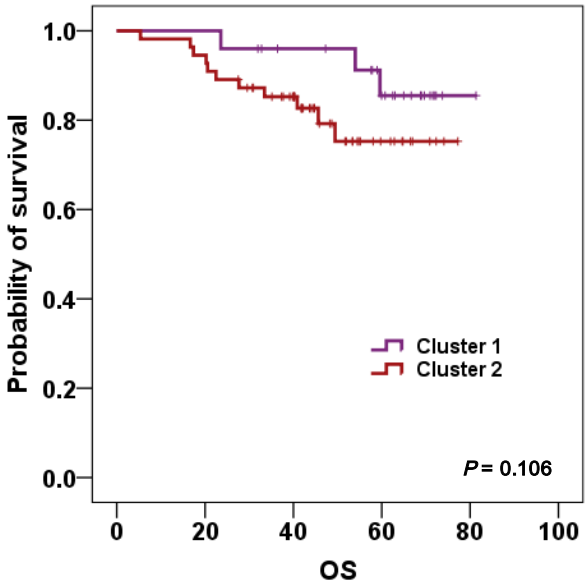
	Cluster 1	Cluster 2	Total	*P* - value
Age				1.000
< 65	13 (32.5%)	27 (67.5%)	40 (50.0%)	
≥ 65	12 (30.0%)	28 (70.0%)	40 (50.0%)	
Sex				0.406
Male	17 (28.3%)	43 (71.7%)	60 (75.0%)	
Female	8 (40.0%)	12 (60.0%)	20 (25.0%)	
Lauren				0.247
Intestinal	9 (27.3%)	24 (72.7%)	33 (41.3%)	
Diffuse	11 (29.7%)	26 (70.3%)	37 (46.2%)	
Mixed	4 (50.0%)	4 (50.0%)	8 (10.0%)	
Indeterminate	1 (50.0%)	1 (50.0%)	2 (2.5%)	
Lymphatic				0.397
Absent	4 (21.1%)	15 (78.9%)	19 (23.7%)	
Present	21 (34.4%)	40 (65.6%)	61 (76.3%)	
Vascular				0.755
Absent	20 (30.3%)	46 (69.7%)	66 (82.5%)	
Present	5 (35.7%)	9 (64.3%)	14 (17.5%)	
Perineural				0.302
Absent	10 (40.0%)	15 (60.0%)	25 (31.3%)	
Present	15 (27.3%)	40 (72.7%)	55 (68.7%)	
pT stage				0.717
T1/T2	4 (40.0%)	6 (60.0%)	10 (12.5%)	
T3/T4	21 (30.0%)	49 (70.0%)	70 (87.5%)	
pN stage				0.755
N0	5 (33.3%)	9 (64.3%)	14 (17.5%)	
N1	20 (30.3%)	46 (69.7%)	66 (82.5%)	
pTNM stage				0.810
II	12 (33.3%)	24 (66.7%)	36 (45.0%)	
III	13 (29.5%)	31 (70.5%)	44 (55.0%)	
Total	25 (31.3%)	55 (68.7%)	80 (100.0%)	

Figure 13. Comparison of three types of gastric cancer classification methods



Three classification models of stage II and III are plotted and compared. Close association of clustering model and molecular classification is observed: Group 2 GCs are only in cluster 1 and group 3 GCs are more commonly classified as cluster 2.

Figure 14. Survival analysis according to two clusters



OS according to cluster model was analysed and plotted, and cluster 2 showed relatively shorter survival, though lacking statistical significance.

Chapter 4. Discussion

4.1 Molecular biologic and clinical significance of TMIT

4.1.1 Molecular biologic significance

In this study, I classified a large cohort of stage II and III GC patients who were managed with standard treatment into one of four TMITs, using immunohistochemical assessment of PD-L1 expression and $CD8^+$ TIL infiltration as the surrogate markers of the tumor microenvironment (TME). I found that TMIT I ($PD\text{-}L1^+/CD8^{High}$) is closely correlated with EBV infection and MSI-H phenotype than TMIT IV ($PD\text{-}L1^-/CD8^{High}$). Additionally, to validate our results, I analysed datasets from TCGA (Cancer Genome Atlas Research Network, 2014) and the SMC cohort, the latter of which is a mostly Asian population (Cristescu *et al*, 2015). The results also showed that the EBV^+ and MSI-H cases in the both datasets were likely to be type I ($PD\text{-}L1^{High}/CD8A^{High}$).

Numerous studies have shown that PD-L1 expression is increased in both EBV^+ and MSI-H GCs (Kim *et al*, 2015; Derks *et al*, 2016; Kim *et al*, 2016a). Likewise, it is well known that EBV^+ GCs and MSI-H GCs are associated with heavy lymphocytic infiltration (Kim *et al*, 2014; Li *et al*, 2016). However, classification of the TME by co-assessment of PD-L1 and TILs had not yet been reported, and a study of a small Western population showed that $CD8^+$ T cell-infiltrated GCs are associated with PD-L1 expression (Thompson *et al*, 2016). Here, I demonstrated, for the first time, the close association of TMIT I ($PD\text{-}L1^+/CD8^{High}$) with EBV^+ and MSI-H, compared to type IV (PD-

L1^{-}/CD8^{High}), using both tissue samples and gene expression data. TMIT I status (PD-L1^{+}/CD8^{High}) implies the adaptive immune escape responses, and based on many previous studies, there is a good chance that GCs with this signature can be reversed by immune checkpoint blockade (Taube *et al*, 2012; Thompson *et al*, 2016). Therefore, I suggest that the type I (PD-L1^{+}/CD8^{High}) TMIT could serve as a biomarker for a good response to immune checkpoint inhibitors, and that PD-L1 and CD8 TIL status should be evaluated in patients with EBV^{+} or MSI-H GC.

4.1.2 Clinical and prognostic significance

In addition, I also found that the TMIT has prognostic value. TMIT II, which implies the immune ignorant state of tumor microenvironment, shows worse survival outcome compared to highly inflamed status (types I and IV), and this finding is consistent with previous studies from diverse tumor types including GC (Kim *et al*, 2014; 2016a). Even more important finding from our survival analysis is that OS within the CD8^{High} group differs according to the differential expression of PD-L1; type I (PD-L1^{+}/CD8^{High}) showed significantly poorer OS than type IV (PD-L1^{-}/CD8^{High}) by multivariate analysis. From this I could infer that although heavy immune cell infiltration might play the favorable anti-tumor effect in gastric cancer, effective immune evading occurs by expression of PD-L1, possibly resulting in decreased OS. Since PD-L1 expression alone failed to discriminate survival in the total study population, the significant survival difference elucidated by differential PD-L1 expression in the CD8^{High} group strongly suggests that the clinical implication of PD-L1

expression could become more meaningful when interpreted in combination with other components of the TME. Therefore, I suggest co-assessment of both PD-L1 and $CD8^+$ TILs as a useful way of defining the TME, which also has a significant prognostic role in stage II and III GC.

Regarding the results of survival analysis using the transcriptome data from TCGA and SMC cohort, significant survival differences according to four TMIT groups were not observed. Part of the reason for this result could be explained by the technical limitation of RNAseq data: tumor-stroma mixture. RNAseq data of TCGA and SMC cohort are derived from the mixture of cancer and surrounding stromal tissue, therefore the *PD-L1* mRNA levels represent the both component of tumor microenvironment, while the analysis using TMA of stage II and III GCs only assessed the PD-L1 expression on tumor cells. More important factor to consider is the fact that treatment strategies of patients in TCGA and SMC cohort varied, while the stage II and III GC cohort patients were all treated with curative surgical resection followed by standard adjuvant chemotherapy.

One step further, I found that previous studies on the prognostic role of PD-L1 expression in GC showed conflicting results. For example, the most recent study of a large Caucasian cohort of GC showed that PD-L1 expression in tumor and stromal immune cells was associated with better tumor-specific and overall survival (Böger *et al*, 2016), while previous studies of an Asian population showed the poor prognostic role of PD-L1 expression (Eto *et al*, 2015; Zhang *et al*, 2015). Some authors attributed these discrepant results to differences in the gene signatures between the Asian and Caucasian populations

(Shen *et al*, 2013; Böger *et al*, 2016). Apart from ethnicity, I suggest other explanations for the conflicting results. Previous survival analyses of GC according to PD-L1 expression were not performed within the context of the immune microenvironment, as discussed earlier. Furthermore, most studies were performed on heterogeneous populations; that is, patients with cancers of various stages with different clinical settings and treatment strategies. In contrast, our study population was relatively homogenous. In Korea, the 5-year survival rate of the localized gastric cancer patients exceeds 92% (Jung *et al*, 2013), therefore, when performing prognostic analysis within the localized gastric cancer group, the chance that the survival outcome of this group may not be directly related to disease itself must be taken into account. In cases of metastatic gastric cancer, the therapeutic approach including chemotherapy regimen widely varies (Lee *et al*, 2014), and this heterogeneity may result in possible confounder of the survival analysis. For these reasons, I have restricted the study population into patients with stage II and III GC who were treated by curative surgical resection followed by FP-based adjuvant chemotherapy, expecting that there would be less bias affecting survival analysis. Therefore, I suggest that the prognostic difference found in the present study of stage II and III GCs is notable and very reliable.

4.1.3 Additional tumor-associated features and immune-oncologic significance

Since the introduction of molecular subtypes of GC in TCGA study, EBV^+ GCs and MSI-H GCs have been consistently regarded as distinct

subtypes (Cancer Genome Atlas Research Network, 2014). Yet, debates regarding the proper classification of the remaining GCs continue, and little is known about these GCs from an immuno-oncologic perspective. Recently, Setia *et al*. suggested a practical molecular classification model mainly based on IHC analysis of E-cadherin and p53 (Setia *et al*, 2016), which I adapted in this study. Based on the previous findings for other types of solid tumors, group 3 (MSS/MSI-L/EMT-like) was expected to be positively associated with PD-L1 expression (Ock *et al*, 2016a; Kim *et al*, 2016b). However, only 3.1% of group 3 cases (4/105) were PD-L1^{+}. This may be due to differences in the biology of GC compared to that of the other cancers for which strong associations were observed.

For this reason, I have come to a hypothesis that the association between EMT and immune escape mechanism via PD-L1 expression would be different in GCs compared to other types of solid tumor. Therefore I have studied vimentin, another marker representing mesenchymal phenotype in addition to E-cadherin. Furthermore, I co-assessed the stem cell markers widely studied in GCs previously, which is also a key tumor-associated feature playing crucial step in cancer progression.

Altered E-cadherin expression and vimentin positivity, representing EMT-like feature, were more frequently observed in TMIT II rather than TMIT I, which is the opposite finding compared to previous studies on pan-cancer RNAseq study, and IHC based studies on lung adenocarcinoma and head and neck squamous cell carcinoma (HNSCC) (Mak *et al*, 2016; Ock *et al*, 2016a; Kim *et al*, 2016b). A clue to explain this finding was found in a subtype of

breast cancer, invasive lobular carcinoma (ILC), which is also well-characterized by altered E-cadherin expression. Most of the studies of PD-L1 expression on breast cancer had been focused on ductal carcinomas, and reports on ILCs are recently introduced, which states that PD-L1 expression on ILCs are relatively rare (Dill *et al*, 2017). Moreover, comprehensive genomic analysis on ILCs has identified two distinct subtypes within lobular carcinoma: one with immune related signature characterized by PD-L1 expression and GATA3 mutation, and the other with hormone related signature associated with EMT with low PD-L1 expression (Michaut *et al*, 2016). Considering the resemblance of ILC cells and GC cells showing altered E-cadherin and vimentin expression, the association between TMIT II and EMT feature in this study may share the similar biological nature with the low PD-L1 expression on ILCs. Further studies to clarify the underlying biological mechanism that can explain these phenomena should be warranted.

Regarding the stem cell features of stage II and III GC cohort, the most striking feature was the close association between CD44 and PD-L1 expression on tumor cells. Compared to EMT phenomenon, the association between cancer stemness and immune evading mechanism is not widely studied yet. Recent study on HNSCC suggested that $CD44^+$ cancer cells constitutively express PD-L1 to evade host immunity via constitutive phosphorylation of STAT3 (Lee *et al*, 2016b). Though temporal association between CD44 expression and PD-L1 expression was not studied in this study, the strong correlation between two markers suggest that CD44 expression on GC cells have immune-oncologic implication. From the mRNA expression analysis using TCGA and SMC

cohort, *CD44* level in TMIT I did not appear to be significantly high, however, I concluded that *CD44* signature from stromal cells may have hindered the association between TMIT I and *CD44* level.

4.1.4 Further consideration

This study has the limitation of being a retrospective study at a single institution. However, compared to other studies, our study population is a large, relatively homogeneous cohort with restricted confounding factors. The cut-off value for PD-L1^{+} is still a matter of debate; applying different cut-off level for PD-L1 IHC results would inevitably result in different proportions among the TMIT subtypes. However, since there is no general consensus in this topic till nowadays, I have done thorough review of previous studies in pursuit of identifying an ideal cut-off criteria for PD-L1 IHC, and chose our criteria referenced from the most recent studies of GC (Derks *et al*, 2016; Thompson *et al*, 2016). In addition, this study was based on the immunostainings on TMA blocks, which enabled us to assess PD-L1 expression in a large cohort of 392 patients. Despite, it is reported that spatial heterogeneity of PD-L1 IHC exists in various types of tumor including non-small cell lung cancer and malignant melanoma (Rehman *et al*, 2017). Therefore, even though I have applied 5% positivity as the cut-off for PD-L1 IHC, the possibility of false-negativity should be considered.

4.2 Somatic mutational profiles of stage II and III gastric cancer

4.2.1 Mutational landscape of GC

In overall, similar somatic mutational landscape was found in stage II and III GC cohort compared to TCGA report (Cancer Genome Atlas Research Network, 2014). *PIK3CA* and *TP53* genes were the most commonly mutated genes, and mutations of genes in Wnt signaling pathway (*GNAQ*, *CDH1*, *APC, CTNNB1, CREBBP, RNF43*), TGF-β pathway (*SMAD4* and *SMAD2*) were observed in the present cohort with similar mutational frequency compared to TCGA.

We also observed that EBV^+ GCs were significantly enriched with *PIK3CA* mutations, as well reported by TCGA group. In addition, among the three *CTNNB1* mutations in the present cohort, two of them were found in EBV^+ GCs, consistent with previous reports (Lee *et al*, 2012). *POLE* gene encodes DNA polymerase epsilon catalytic subunit, and its mutations cause defective DNA proofreading function, resulting in higher mutational burden and therefore enhanced immune response (van Gool *et al*, 2015); three of the seven *POLE* mutated cases were EBV^+ GCs, which could be expected from the fact that virus-associated cancers harbor more numbers of mutations and cause intense immune reactivity.

MSI-H GCs were also enriched with *PIK3CA* mutations, however, none of seven patients harbored *TP53* mutations. *BRAF* mutations are alleged to be very rare in GCs (van Grieken *et al*, 2013), however, I observed seven cases with *BRAF* mutations and three of them were MSI-H GCs; in contrast to MSI-H colorectal cancer, none of them were V600E mutations. Recent studies in colorectal adenocarcinoma and endometrial adenocarcinoma suggested that

RNF43 plays a role in carcinogenesis of MSI-H cancers (Giannakis *et al*, 2014). Three out of seven *RNF43* mutated cases were MSI-H GCs, implying the association between *RNF43* and MSI status ($P = 0.013$).

4.2.2 Consideration of both somatic mutations of tumor cells and TME

Somatic mutational signatures of tumors according to TMIT classification were determined for the purpose of understaning tumor genetics within the context of tumor microenvironment as well. TMIT I tumors were most notably enriched with mutations of *PIK3CA*, which is a highly expectable finding considering the close association of EBV^+/MSI-H GCs and TMIT I. More interesting finding was that TMIT II, which is alleged to be immunologically silent group, were significantly enriched with *RUNX1* mutations. *RUNX1* (runt related transcription factor 1) is a tumor suppressor gene, previously studied mostly in hematolymphoid diseases. Among gastrointestinal malignancies, it was reported that 15% of esophageal tumors have deletions in *RUNX1* (Dulak *et al*, 2012). More recently, it was reported that microRNA-216a-3p (miR-216a-3p) downregulates *RUNX1* in GCs and cause activation of NF-κB signalling pathway (Wu *et al*, 2017), implying the potential role of *RUNX1* gene in GC carcinogenesis.

NTRK3 gene is well known for its fusion with *ETV6* in newly developed entity in salivary gland, the secretory carcinoma (Skálová *et al*, 2010). Its missense mutations, however, are studied only recently in subsets of colorectal adenocarcinoma (Deihimi *et al*, 2017), and somatic mutations of

NTRK3 genes are not well reported in GCs. Considering the role of *NTRK3* as the tyrosine kinase domain, the enriched mutations of *NTRK3* gene in TMIT IV indicates the possible therapeutic target among this subgroup of GCs.

4.2.3 Limitations of gene cluster model

Based on the novel findings derived from the targeted sequencing data, I performed fuzzy clustering analysis for the purpose of developing a novel classification of GCs according to somatic mutational characteristics. This gene clustering model consisted of two groups, one with higher mutational burden (cluster 1) and the other with relatively low genomic alterations (cluster 2). Though I have observed the tendency of cluster 1 having better OS compared to cluster 2, it lacked statistical significance. Moreover, none of the tumor related clinicopathologic characteristics of 80 GCs correlated with this gene cluster model, except for the association between cluster 1 and MSI-H GCs. For a classification scheme of a disease to be clinically meaningful, it is crucial that the classification method could provide prognostic information as well as clinicopathologic associations. Therefore, TMIT is the classification method which is much easier to adapt compared to targeted sequencing, providing more relevant information and better prognostic performance.

4.3 Conclusive remarks

I have found that EBV^{+} and MSI-H GCs are distinct subtypes that are tightly associated with TMIT I ($PD\text{-}L1^{+}/CD8^{High}$), OS within the $CD8^{High}$ group

differs according to PD-L1 expression, and I have proved that co-assessment of PD-L1 and $CD8^+$ TILs is clinically relevant, with a possible prognostic role.

The associations between TMIT classification and major cancer-propagating characteristics – EMT and cancer stemness – were observed. I have found an inverse association between EMT phenotype and PD-L1 expression, and close association between EMT features and TMIT II in GCs. In addition, I have found a tight association between CD44 positivity, a cancer stem cell marker, and TMIT I phenotype.

Finally, by performing deep targeted sequencing on selected GC tissue samples, I have found that TMIT I tumors have more numbers of somatic mutations compared to other groups and are enriched with somatic mutations of major cancer related genes including *PIK3CA*. TMIT II tumors were enriched with mutations of *RUNX1* gene, and *NTRK3* mutations were relatively specific to TMIT IV. TMIT III had unique somatic mutational profile, harbouring mutations of genes such as *APC*, *TSC1*, *JAK1*, *MET*, *HRAS* and *RHEB*. Clustering analysis based on somatic mutational profiles have identified two groups, one with higher mutational burden (cluster 1) and the other with lower (cluster 2); cluster 1 had significant association with MSI-H GCs and showed the tendency of shorter overall survival.

Overall, this study indicates that TMIT classification, which co-assesses both PD-L1 expression on tumor cells and surrounding $CD8^+$ TILs, has clinicopathologic, molecular genetic and clinical implications, and I expect that findings from this study may help to provide additional clues for deeper understanding of the biology of GCs.

Bibliography

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. *Nat Methods* **7**: 248–249, doi:10.1038/nmeth0410-248.

Ansell SM, Lesokhin AM, Borrello I, Halwani A, Scott EC, Gutierrez M, Schuster SJ, Millenson MM, Cattry D, Freeman GJ, Rodig SJ, Chapuy B, Ligon AH, Zhu L, Grosso JF, Kim SY, Timmerman JM, Shipp MA, Armand P (2015) PD-1 Blockade with Nivolumab in Relapsed or Refractory Hodgkin's Lymphoma. *N Engl J Med* **372**: 311–319, doi:10.1056/NEJMoa1411087.

Böger C, Behrens H-M, Mathiak M, Krüger S, Kalthoff H, Röcken C (2016) PD-L1 is an independent prognostic predictor in gastric cancer of Western patients. *Oncotarget* **7**: 24269–24283, doi:10.18632/oncotarget.8169.

Cancer Genome Atlas Research Network (2014a) Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**: 202–209, doi:10.1038/nature13480.

Chang MS, Lee JH, Kim JP, Kim HS, Lee HS, Kim CW, Kim YI, Kim WH (2000) Microsatellite instability and Epstein-Barr virus infection in gastric remnant cancers. *Pathology International* **50**: 486–492.

Choi YY, Noh SH, Cheong J-H (2016) Molecular Dimensions of Gastric Cancer: Translational and Clinical Perspectives. *J Pathol Transl Med* **50**: 1–9, doi:10.4132/jptm.2015.09.10.

Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, Gabriel S, Meyerson M, Lander ES, Getz G (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**: 213–219, doi:10.1038/nbt.2514.

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X (2012) Using Drosophila melanogaster as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front Genet* **3**: 35, doi:10.3389/fgene.2012.00035.

Cristescu R, Lee J, Nebozhyn M, Kim K-M, Ting JC, Wong SS, Liu J, Yue YG, Wang J, Yu K, Ye XS, Do I-G, Liu S, Gong L, Fu J, Jin JG, Choi MG, Sohn TS, Lee JH, Bae JM, Kim ST, Park SH, Sohn I, Jung S-H, Tan P, Chen R, Hardwick J, Kang WK, Ayers M, Hongyue D, Reinhard C, Loboda A, Kim S, Aggarwal A (2015a) Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat Med* **21**: 449–456, doi:10.1038/nm.3850.

Das S, Suarez G, Beswick EJ, Sierra JC, Graham DY, Reyes VE (2006) Expression of B7-H1 on Gastric Epithelial Cells: Its Potential Role in Regulating T Cells during Helicobacter pylori Infection. *The Journal of Immunology* **176**: 3000–3009, doi:10.4049/jimmunol.176.5.3000.

Deihimi S, Lev A, Slifker M, Shagisultanova E, Xu Q, Jung K, Vijayvergia N, Ross EA, Xiu J, Swensen J, Gatalica Z, Andrake M, Dunbrack RL, El-Deiry WS (2017) BRCA2, EGFR, and NTRK mutations in mismatch repair-deficient colorectal cancers with MSH2 or MLH1 mutations. *Oncotarget* **8**: 39945–39962, doi:10.18632/oncotarget.18098.

Derks S, Liao X, Chiaravalli AM, Xu X, Camargo MC, Solcia E, Sessa F, Fleitas T, Freeman GJ, Rodig SJ, Rabkin CS, Bass AJ (2016) Abundant PD-L1 expression in Epstein-Barr Virus-infected gastric cancers. *Oncotarget* **7**: 32925–32932, doi:10.18632/oncotarget.9076.

Dill EA, Gru AA, Atkins KA, Friedman LA, Moore ME, Bullock TN, Cross JV, Dillon PM, Mills AM (2017) PD-L1 Expression and Intratumoral Heterogeneity Across Breast Cancer Subtypes and Stages: An Assessment of 245 Primary and 40 Metastatic Tumors. *Am J Surg Pathol* **41**: 334–342, doi:10.1097/PAS.0000000000000780.

Dong Z-Y, Zhong W-Z, Zhang X-C, Su J, Xie Z, Liu S-Y, Tu H-Y, Chen H-J, Sun Y-L, Zhou Q, Yang J-J, Yang X-N, Lin J-X, Yan H-H, Zhai H-R, Yan L-X, Liao R-Q, Wu S-P, Wu Y-L (2017) Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. *Clinical Cancer Research* **23**: 3012–3024, doi:10.1158/1078-0432.CCR-16-2554.

Dulak AM, Schumacher SE, van Lieshout J, Imamura Y, Fox C, Shim B, Ramos AH, Saksena G, Baca SC, Baselga J, Tabernero J, Barretina J, Enzinger PC, Corso G, Roviello F, Lin L, Bandla S, Luketich JD, Pennathur A, Meyerson M, Ogino S, Shivdasani RA, Beer DG, Godfrey TE, Beroukhim R, Bass AJ (2012) Gastrointestinal adenocarcinomas of the esophagus, stomach, and colon exhibit distinct patterns of genome instability and oncogenesis. *Cancer Res* **72**: 4383–4393, doi:10.1158/0008-5472.CAN-11-3893.

Eto S, Yoshikawa K, Nishi M, Higashijima J, Tokunaga T, Nakao T, Kashihara H, Takasu C, Iwata T, Shimada M (2015) Programmed cell death protein 1 expression is an independent prognostic factor in gastric cancer after curative resection. *Gastric Cancer* **19**: 466–471, doi:10.1007/s10120-015-0519-7.

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer* **136**: E359–E386, doi:10.1002/ijc.29210.

Giannakis M, Hodis E, Jasmine Mu X, Yamauchi M, Rosenbluh J, Cibulskis K, Saksena G, Lawrence MS, Qian ZR, Nishihara R, Van Allen EM, Hahn WC, Gabriel SB, Lander ES, Getz G, Ogino S, Fuchs CS, Garraway LA (2014) RNF43 is frequently mutated in colorectal and endometrial cancers. *Nat Genet* **46**: 1264–1266, doi:10.1038/ng.3127.

Herbst RS, Soria J-C, Kowanetz M, Fine GD, Hamid O, Gordon MS, Sosman JA, McDermott DF, Powderly JD, Gettinger SN, Kohrt HEK, Horn L, Lawrence DP, Rost S, Leabman M, Xiao Y, Mokatrin A, Koeppen H, Hegde PS, Mellman I, Chen DS, Hodi FS (2014) Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**: 563–567, doi:10.1038/nature14011.

Hodi FS, O'Day SJ, McDermott DF, Weber RW, Sosman JA, Haanen JB, Gonzalez R, Robert C, Schadendorf D, Hassel JC, Akerley W, van den Eertwegh AJM, Lutzky J, Lorigan P, Vaubel JM, Linette GP, Hogg D, Ottensmeier CH, Lebbé C, Peschel C, Quirt I, Clark JI, Wolchok JD, Weber JS, Tian J, Yellin MJ, Nichol GM, Hoos A, Urba WJ (2010) Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* **363**: 711–723, doi:10.1056/NEJMoa1003466.

Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D (2011) Global cancer statistics. *CA: A Cancer Journal for Clinicians* **61**: 69–90, doi:10.3322/caac.20107.

Jung K-W, Won Y-J, Kong H-J, Oh C-M, Shin A, Lee J-S (2013) Survival of korean adult cancer patients by stage at diagnosis, 2006-2010: national cancer registry study. *Cancer Res Treat* **45**: 162–171, doi:10.4143/crt.2013.45.3.162.

Jung K-W, Won Y-J, Oh C-M, Kong H-J, Cho H, Lee J-K, Lee DH, Lee KH (2016) Prediction of Cancer Incidence and Mortality in Korea, 2016. *Cancer Res Treat* **48**: 451–457, doi:10.4143/crt.2016.092.

Kim JH, Park HE, Cho NY, Lee HS, Kang GH (2016a) Characterisation of PD-L1-positive subsets of microsatellite-unstable colorectal cancers. *Br J Cancer* **115**: 490–496, doi:10.1038/bjc.2016.211.

Kim JW, Im S-A, Kim M, Cha Y, Lee K-H, Keam B, Kim MA, Han S-W, Oh D-Y, Kim T-Y, Kim WH, Bang Y-J (2012) The prognostic significance of HER2 positivity for advanced gastric cancer patients undergoing first-line modified FOLFOX-6 regimen. *Anticancer Res* **32**: 1547–1553.

Kim K-J, Lee KS, Cho HJ, Kim YH, Yang H-K, Kim WH, Kang GH (2014) Prognostic implications of tumor-infiltrating FoxP3+ regulatory T cells and CD8+ cytotoxic T cells in microsatellite-unstable gastric cancers. *Hum Pathol* **45**: 285–293, doi:10.1016/j.humpath.2013.09.004.

Kim MA, Jung JE, Lee HE, Yang H-K, Kim WH (2013) In situ analysis of HER2 mRNA in gastric carcinoma: comparison with fluorescence in situ hybridization, dual-color silver in situ hybridization, and immunohistochemistry. *Hum Pathol* **44**: 487–494, doi:10.1016/j.humpath.2012.06.022.

Kim S, Koh J, Kim M-Y, Kwon D, Go H, Kim YA, Jeon YK, Chung DH (2016b) PD-L1 expression is associated with epithelial-to-mesenchymal transition in adenocarcinoma of the lung. *Hum Pathol* **58**: 7–14, doi:10.1016/j.humpath.2016.07.007.

Kim SY, Park C, Kim H-J, Park J, Hwang J, Kim J-I, Choi MG, Kim S, Kim K-M, Kang M-S (2015) Deregulation of Immune Response Genes in Patients With Epstein-Barr Virus-Associated Gastric Cancer and Outcomes. *Gastroenterology* **148**: 137–147.e139, doi:10.1053/j.gastro.2014.09.020.

Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Hoover J, Jang W, Katz K, Ovetsky M, Riley G, Sethi A, Tully R, Villamarin-Salomon R, Rubinstein W, Maglott DR (2016) ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res* **44**: D862–D868, doi:10.1093/nar/gkv1222.

Lee J, Van Hummelen P, Go C, Palescandolo E, Jang J, Park HY, Kang SY, Park JO, Kang WK, MacConaill L, Kim K-M (2012) High-throughput mutation profiling identifies frequent somatic mutations in advanced gastric adenocarcinoma. *PLoS ONE* **7**: e38892, doi:10.1371/journal.pone.0038892.

Lee JH, Kim JG, Jung H-K, Kim JH, Jeong WK, Jeon TJ, Kim JM, Kim YI, Ryu KW, Kong S-H, Kim H-I, Jung H-Y, Kim YS, Zang DY, Cho JY, Park JO, Lim DH, Jung ES, Ahn HS, Kim H-J (2014) Clinical practice guidelines

for gastric cancer in Korea: an evidence-based approach. *J Gastric Cancer* **14**: 87–104, doi:10.5230/jgc.2014.14.2.87.

Lee KS, Nam SK, Koh J, Kim D-W, Kang S-B, Choe G, Kim WH, Lee HS (2016a) Stromal Expression of MicroRNA-21 in Advanced Colorectal Cancer Patients with Distant Metastases. *J Pathol Transl Med* **50**: 270–277, doi:10.4132/jptm.2016.03.19.

Lee Y, Shin JH, Longmire M, Wang H, Kohrt HE, Chang HY, Sunwoo JB (2016b) CD44+ Cells in Head and Neck Squamous Cell Carcinoma Suppress T-Cell-Mediated Immunity by Selective Constitutive and Inducible Expression of PD-L1. *Clinical Cancer Research* **22**: 3571–3581, doi:10.1158/1078-0432.CCR-15-2665.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'Donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, DePristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won H-H, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, McCarroll S, McCarthy MI, McGovern D, McPherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, MacArthur DG, Exome Aggregation Consortium (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**: 285–291, doi:10.1038/nature19057.

Li (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078–2079, doi:10.1093/bioinformatics/btp352.

Li N, Wang W, Xu B, Gong H (2014) OCT3/4 expression is correlated with the invasion of gastric carcinoma. *Oncol Lett* **8**: 12–16, doi:10.3892/ol.2014.2112.

Li Z, Lai Y, Sun L, Zhang X, Liu R, Feng G, Zhou L, Jia L, Huang X, Kang Q, Lin D, Gao J, Shen L (2016) PD-L1 expression is associated with massive lymphocyte infiltration and histology in gastric cancer. *Hum Pathol* **55**: 182–189, doi:10.1016/j.humpath.2016.05.012.

Liu X, Wu C, Li C, Boerwinkle E (2016) dbNSFP v3.0: A One-Stop Database of Functional Predictions and Annotations for Human Nonsynonymous and Splice-Site SNVs. *Hum Mutat* **37**: 235–241, doi:10.1002/humu.22932.

Mak MP, Tong P, Diao L, Cardnell RJ, Gibbons DL, William WN, Skoulidis F, Parra ER, Rodriguez-Canales J, Wistuba II, Heymach JV, Weinstein JN, Coombes KR, Wang J, Byers LA (2016) A Patient-Derived, Pan-Cancer EMT Signature Identifies Global Molecular Alterations and Immune Target Enrichment Following Epithelial-to-Mesenchymal Transition. *Clinical Cancer Research* **22**: 609–620, doi:10.1158/1078-0432.CCR-15-0876.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal; Vol 17, No 1* doi:http://dx.doi.org/10.14806/ej.17.1.200.

Michaut M, Chin S-F, Majewski I, Severson TM, Bismeijer T, de Koning L, Peeters JK, Schouten PC, Rueda OM, Bosma AJ, Tarrant F, Fan Y, He B, Xue Z, Mittempergher L, Kluin RJC, Heijmans J, Snel M, Pereira B, Schlicker A, Provenzano E, Ali HR, Gaber A, O'Hurley G, Lehn S, Muris JJF, Wesseling J, Kay E, Sammut SJ, Bardwell HA, Barbet AS, Bard F, Lecerf C, O'Connor DP, Vis DJ, Benes CH, McDermott U, Garnett MJ,

Simon IM, Jirström K, Dubois T, Linn SC, Gallagher WM, Wessels LFA, Caldas C, Bernards R (2016) Integration of genomic, transcriptomic and proteomic data identifies two biologically distinct subtypes of invasive lobular breast cancer. *Sci Rep* **6**: 18517, doi:10.1038/srep18517.

Muro K, Chung HC, Shankaran V, Geva R, Catenacci D, Gupta S, Eder JP, Golan T, Le DT, Burtness B, McRee AJ, Lin C-C, Pathiraja K, Lunceford J, Emancipator K, Juco J, Koshiji M, Bang Y-J (2016) Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncology* **17**: 717–726, doi:10.1016/s1470-2045(16)00175-3.

Nam KH, Yoon H, Lee K, Park DJ, Kim H-H, Lee HS, Shin E (2017) Predictive value for lymph node metastasis of epithelial-mesenchymal transition and cancer stem cell marker expression in early gastric cancer. *Pathol Res Pract* **213**: 1221–1226, doi:10.1016/j.prp.2017.03.010.

Ock C-Y, Kim S, Keam B, Kim M, Kim TM, Kim J-H, Jeon YK, Lee J-S, Kwon SK, Hah JH, Kwon T-K, Kim D-W, Wu H-G, Sung M-W, Heo DS (2016a) PD-L1 expression is associated with epithelial-mesenchymal transition in head and neck squamous cell carcinoma. *Oncotarget* **7**: 15901–15914, doi:10.18632/oncotarget.7431.

Ock CY, Keam B, Kim S, Lee JS, Kim M, Kim TM, Jeon YK, Kim DW, Chung DH, Heo DS (2016b) Pan-Cancer Immunogenomic Perspective on the Tumor Microenvironment Based on PD-L1 and CD8 T-Cell Infiltration. *Clinical Cancer Research* **22**: 2261–2270, doi:10.1158/1078-0432.CCR-15-2834.

Rehman JA, Han G, Carvajal-Hausdorf DE, Wasserman BE, Pelekanou V, Mani NL, McLaughlin J, Schalper KA, Rimm DL (2017) Quantitative and pathologist-read comparison of the heterogeneity of programmed death-ligand 1 (PD-L1) expression in non-small cell lung cancer. *Mod Pathol* **30**: 340–349, doi:10.1038/modpathol.2016.186.

Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, Lee W, Yuan J, Wong P, Ho TS, Miller ML, Rekhtman N, Moreira AL, Ibrahim F, Bruggeman C, Gasmi B, Zappasodi R, Maeda Y, Sander C, Garon EB, Merghoub T, Wolchok JD, Schumacher TN, Chan TA (2015) Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**: 124–128, doi:10.1126/science.aaa1348.

Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N (2015) Molecular and Genetic Properties of Tumors Associated with Local Immune Cytolytic Activity. *Cell* **160**: 48–61, doi:10.1016/j.cell.2014.12.033.

Ryu HS, Park DJ, Kim H-H, Kim WH, Lee HS (2012) Combination of epithelial-mesenchymal transition and cancer stem cell-like phenotypes has independent prognostic value in gastric cancer. *Hum Pathol* **43**: 520–528, doi:10.1016/j.humpath.2011.07.003.

Setia N, Agoston AT, Han HS, Mullen JT, Duda DG, Clark JW, Deshpande V, Mino-Kenudson M, Srivastava A, Lennerz JK, Hong TS, Kwak EL, Lauwers GY (2016) A protein and mRNA expression-based classification of gastric cancer. *Mod Pathol* **29**: 772–784, doi:10.1038/modpathol.2016.55.

Shen L, Shan Y-S, Hu H-M, Price TJ, Sirohi B, Yeh K-H, Yang Y-H, Sano T, Yang H-K, Zhang X, Park SR, Fujii M, Kang Y-K, Chen L-T (2013) Management of gastric cancer in Asia: resource-stratified guidelines. *Lancet Oncology* **14**: e535–e547, doi:10.1016/S1470-2045(13)70436-4.

Skálová A, Vanecek T, Sima R, Laco J, Weinreb I, Perez-Ordonez B, Starek I, Geierova M, Simpson RHW, Passador-Santos F, Ryska A, Leivo I, Kinkor Z, Michal M (2010) Mammary analogue secretory carcinoma of salivary glands, containing the ETV6-NTRK3 fusion gene: a hitherto undescribed salivary gland tumor entity. *Am J Surg Pathol* **34**: 599–608, doi:10.1097/PAS.0b013e3181d9efcc.

Suzuki H, Iwasaki E, Hibi T (2009) Helicobacter pylori and gastric cancer. *Gastric Cancer* **12**: 79–87, doi:10.1007/s10120-009-0507-x.

Taube JM, Anders RA, Young GD, Xu H, Sharma R, McMiller TL, Chen S, Klein AP, Pardoll DM, Topalian SL, Chen L (2012) Colocalization of Inflammatory Response with B7-H1 Expression in Human Melanocytic Lesions Supports an Adaptive Resistance Mechanism of Immune Escape. *Science Translational Medicine* **4**: 127ra37–127ra37, doi:10.1126/scitranslmed.3003689.

Teng MWL, Ngiow SF, Ribas A, Smyth MJ (2015) Classifying Cancers Based on T-cell Infiltration and PD-L1. *Cancer Res* **75**: 2139–2145, doi:10.1158/0008-5472.CAN-15-0255.

Thompson ED, Zahurak M, Murphy A, Cornish T, Cuka N, Abdelfatah E, Yang S, Duncan M, Ahuja N, Taube JM, Anders RA, Kelly RJ (2016) Patterns of PD-L1 expression and CD8 T cell infiltration in gastric adenocarcinomas and associated immune stroma. *Gut* doi:10.1136/gutjnl-2015-310839.

Tumeh PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJM, Robert L, Chmielowski B, Spasic M, Henry G, Ciobanu V, West AN, Carmona M, Kivork C, Seja E, Cherry G, Gutierrez AJ, Grogan TR, Mateus C, Tomasic G, Glaspy JA, Emerson RO, Robins H, Pierce RH, Elashoff DA, Robert C, Ribas A (2014) PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515**: 568–571, doi:10.1038/nature13954.

Umar A, Boland CR, Terdiman JP, Syngal S, la Chapelle de A, Rüschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HFA, Hawk ET, Barrett JC, Freedman AN, Srivastava S (2004) Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *Journal of the National Cancer Institute* **96**: 261–268.

van Beek J (2004) EBV-Positive Gastric Adenocarcinomas: A Distinct Clinicopathologic Entity With a Low Frequency of Lymph Node

Involvement. *Journal of Clinical Oncology* **22**: 664–670, doi:10.1200/JCO.2004.08.061.

van Gool IC, Eggink FA, Freeman-Mills L, Stelloo E, Marchi E, de Bruyn M, Palles C, Nout RA, de Kroon CD, Osse EM, Klenerman P, Creutzberg CL, Tomlinson IP, Smit VT, Nijman HW, Bosse T, Church DN (2015) POLE Proofreading Mutations Elicit an Antitumor Immune Response in Endometrial Cancer. *Clinical Cancer Research* **21**: 3347–3355, doi:10.1158/1078-0432.CCR-15-0057.

van Grieken NCT, Aoyama T, Aoyma T, Chambers PA, Bottomley D, Ward LC, Inam I, Buffart TE, Das K, Lim T, Pang B, Zhang SL, Tan IB, Carvalho B, Heideman DAM, Miyagi Y, Kameda Y, Arai T, Meijer GA, Tsuburaya A, Tan P, Yoshikawa T, Grabsch HI (2013) KRAS and BRAF mutations are rare and related to DNA mismatch repair deficiency in gastric cancer from the East and the West: results from a large international multicentre study. *Br J Cancer* **108**: 1495–1501, doi:10.1038/bjc.2013.109.

Wakamatsu Y, Sakamoto N, Oo HZ, Naito Y, Uraoka N, Anami K, Sentani K, Oue N, Yasui W (2012) Expression of cancer stem cell markers ALDH1, CD44 and CD133 in primary tumor and lymph node metastasis of gastric cancer. *Pathology International* **62**: 112–119, doi:10.1111/j.1440-1827.2011.02760.x.

Wu Y, Zhang J, Zheng Y, Ma C, Liu X-E, Sun X (2017) MiR-216a-3p Inhibits the Proliferation, Migration, and Invasion of Human Gastric Cancer Cells via Targeting RUNX1 and Activating the NF-κB Signaling Pathway. *Oncol Res* doi:10.3727/096504017X15031557924150.

Yi Kim D, Kyoon Joo J, Kyu Park Y, Yeob Ryu S, Soo Kim H, Kyun Noh B, Hwa Lee K, Hyuk Lee J (2007) E-cadherin expression in early gastric carcinoma and correlation with lymph node metastasis. *J Surg Oncol* **96**: 429–435, doi:10.1002/jso.20732.

Zhang L, Qiu M, Jin Y, Ji J, Li B, Wang X, Yan S, Xu R, Yang D (2015) Programmed cell death ligand 1 (PD-L1) expression on gastric cancer and its relationship with clinicopathologic factors. *Int J Clin Exp Pathol* **8**: 11084–11091.

국문초록

2 기와 3 기 위암의 분자유전학적 특성과 종양 면역 미세 환경에 대한 통합적 분석

PD-L1 발현과 CD8 양성 종양 침윤 림프구의 밀도를 기준으로 종양 미세 환경을 네 가지 아형으로 분류하는 면역 미세 환경 분류(tumor microenvironment immune type, TMIT)를 2 기 및 3 기 위암 조직에 적용하여, 이 분류법의 임상적 유용성을 증명하고, 위암의 분자유전학적 요인과의 관련성을 규명하고자 하였다.

2006년부터 2013년까지 분당서울대학교병원에서 근치적 위절제술 및 5-fluorouracil 기반 보조항암요법으로 치료 받은 2기 및 3기 위암 환자들의 포르말린 고정 파라핀 조직으로부터 조직배열(tissue microarray) 블록을 제작하여 연구에 사용하였고, 환자들의 임상정보를 수집하였다.

PD-L1 과 CD8 면역조직화학염색을 시행하였고 이를 기준으로 면역 미세 환경 분류(TMIT)를 다음과 같이 적용하였다: I형(PD-$L1^{+}$/$CD8^{High}$), II형(PD-$L1^{-}$/$CD8^{Low}$), III형(PD-$L1^{+}$/$CD8^{Low}$), IV형(PD-$L1^{-}$/$CD8^{High}$). 이를 토대로 전체 생존기간을 포함한 임상정보에 대한 분석을 시행하여 TMIT 분류의 예후적 가치에 대해 평가하였다.

위암의 여러 분자유전학적 특성에 대한 통합적 평가를 위해 종양 침윤 면역세포(CD3, CD4, Foxp3), 상피-간질 이행 관련 표지자(E-cadherin, vimentin), 암 줄기세포 표지자(CD44, Sox2, CD133, OCT3/4)에 대한 면역조직화학검사와 EBV 동소교잡반응검사 및 현미부수체 불안정성 검사를 시행하였고 종양 면역학적 관점에서 어떠한 의미를 갖는지 고찰하였다.

또한 위암의 유전자 수준에서의 특성과 면역 미세 환경 사이의 관련성 평가를 위해 두 개의 공개 유전체 데이터세트로부터 전사 유전체 및 임상 정보를 얻어 이에 대한 통계적 분석을 시행하였다. 그리고 각각의 TMIT 아형별로 적합한 증례를 선정하여 170 개 유전자에 대한 차세대 염기서열 분석을 통해 위암의 유전자 변이와 면역 미세 환경의 관계에 대해 평가하였다.

PD-L1 과 CD8 양성 종양 침윤 림프구를 기준으로 각각 생존 분석을 시행하였을 경우, 이들에 따른 유의한 생존 기간의 차이는 보이지 않았으나, PD-L1 양성도와 CD8 양성 림프구를 기준으로 한 면역 미세 환경 분류을 적용 시, CD8 양성 림프구가 많은 아형인 I 형과 IV 형 내에서, PD-L1 의 양성도에 따라 유의하게 다른 생존 기간을 보임을 밝혔고, 이는 TMIT 분류법의 임상적 유용성을 시사한다. 또한 EBV 양성 위암과 현미부수체 불안정 위암의 경우 제 I 형 미세 환경과 밀접한 관련성을 보여, TMIT 분류법이 위암의 대표적인 분자적 특성과도 강한 연관성이 있음을 밝혔다. 또한 E-cadherin 의 이상 발현과 vimentin 의 양성 발현을 보이는 위암의

경우 PD-L1 의 발현이 낮은 현상을 관찰하여, 기존의 다른 고형암에서 보인 상피-간엽 이행 현상과 PD-L1 의 발현 사이의 연관성과는 대비되는 현상이 위암에 존재함을 밝혔고, CD44 발현으로 대표되는 암 줄기세포적 특성과 PD-L1 의 발현 사이의 강한 연관성을 최초로 보고하였다.

면역 미세 환경 분류법(TMIT)은 효과적인 면역 치료 전략 수립을 위해 고안된 분류 체계로, 위암의 발생과 진행에 중요한 여러 분자유전학적 특성과 종양 미세 환경 사이의 관계에 대해 본 연구를 통해 규명함으로써, 위암의 병태생리에 대한 이해를 높일 수 있었고, 더 나아가 다른 종류의 고형암의 발생 및 진행을 이해하는 데에도 유용한 단서를 제공할 수 있을 것으로 기대된다.