저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer
Profiling cancer-associated genetic alterations and molecular classification of cancer in Korean gastric cancer patients

Yoonjung Kim
Department of Medicine
The Graduate School, Yonsei University
Profiling cancer-associated genetic alterations and molecular classification of cancer in Korean gastric cancer patients

Directed by Professor Kyung-A Lee

The Doctoral Dissertation submitted to the Department of Medicine, the Graduate School of Yonsei University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Yoonjung Kim

December 2016
This certifies that the Doctoral Dissertation of Yoonjung Kim is approved.

-----------------------------
Thesis Supervisor: Kyung-A Lee

-----------------------------
Thesis Committee Member#1: Jae Yong Cho

-----------------------------
Thesis Committee Member#2: Woochang Lee

-----------------------------
Thesis Committee Member#3: Jie-Hyun Kim

-----------------------------
Thesis Committee Member#4: Jong Rak Choi

The Graduate School
Yonsei University
ACKNOWLEDGEMENTS

I would first like to express my very profound gratitude to my thesis advisor Prof. Kyung-A Lee for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. I would also like to acknowledge my colleagues from Department of Laboratory Medicine, Yonsei University College of Medicine who provided insight and expertise that greatly assisted the research. I am gratefully indebted to Prof. Jae Yong Cho, Woochang Lee, Jie-Hyun Kim and Jong Rak Choi for their valuable comments that greatly improved this thesis. I am also immensely grateful to Saeam Shin and Aerin Kwon for their inspiration and encouragement. I would also like to thank Prof. Mee-Yon Cho and Sung Nam Kim for providing materials for this study. Finally, I would like to thank all of my family for their unconditional love and support. This accomplishment would not have been possible without them.

Thank you.
<TABLE OF CONTENTS>

ABSTRACT ................................................................. 1

I. INTRODUCTION ........................................................... 4

II. MATERIALS AND METHODS ............................................ 5
   1. Subject selection ...................................................... 5
   2. DNA preparation ..................................................... 6
   3. Detection of EBV and *H. pylori* infection ................... 7
   4. Configuration of a gastric cancer-related target gene panel for
       Korean gastric cancer patients .................................. 7
   5. Targeted sequencing and data analysis ......................... 7
   6. Microsatellite Instability (MSI) Assay ......................... 9
   7. Statistical Analysis ............................................... 10

III. RESULTS ................................................................. 10
   1. Clinical and pathological findings of 107 Korean GC patients ··· 10
   2. Germline variation analysis of hereditary cancer-predisposing
       Syndrome ............................................................. 13
   3. Somatic mutation analysis of gastric cancer with a 43 gene cancer
       panel ................................................................. 17
   4. Copy number analysis in 107 gastric cancer tissues and
       identification of clinical relevant CNVs ...................... 30
   5. Molecular subtype classification and clinical phenotype ........ 33
   6. Prognosis analysis in 107 gastric cancer patients ............. 40

IV. DISCUSSION ............................................................. 42

V. CONCLUSION ........................................................... 47

REFERENCES .............................................................. 49

APPENDICES ................................................................. 56
   1. List of 43 gastric cancer related target gene panel ............. 56
2. Overview of all somatic SNVs and indels identified through NGS in GC ........................................................... 57

3. Somatic variants of the 43 cancer panel genes in 107 Korean gastric cancers and TCGA data .................................. 78

4. Performance of 43 gene pannel through NGS in GC ............... 80

5. Coverage of 43 gene panel through NGS in GC .................... 81

ABSTRACT (IN KOREAN) ........................................ 84

PUBLICATION LIST .............................................. 86
LIST OF FIGURES

Figure 1. Lollipop bar graph representing distribution of mutations across the 39 genes ........................................ 18
Figure 2. Summary of all significant CNAs across 107 gastric tumors ................................................................. 31
Figure 3. Summary of somatic mutations in 107 gastric cancer samples according to molecular subtype ......................... 39
Figure 4. Kaplan-Meier (a) relapse-free survival (RFS) and (b) gastric cancer-specific survival (GCSS) curves were stratified by molecular subtype of gastric cancer (EBV, MSI, CIN, and GS). Kaplan-Meier (c) RFS and (d) GCSS curves were analyzed by AJCC stage ................................................................. 41
Figure 5. Therapeutic implications of somatic genomic alterations in 107 clinical gastric cancer cases ..................... 46

LIST OF TABLES

Table 1. Clinicopathological characteristics of patients with gastric cancer ............................................................... 11
Table 2. Germline SNPs in TP53, STK11, ALK, APC, MSH2, MLH1, and CDH1 genes of 107 Korean gastric cancer patients ................................................................................................................................ 14
Table 3. The location-specific recurrence of somatic variants in
the 43 cancer panel genes .............................................. 25
Table 4. Copy number analysis in 107 gastric cancer tissues ···· 30
Table 5. Somatic mutations in each subtype ······················ 34
Table 6. Patient characteristics according to molecular subtype
.................................................................................... 36
ABSTRACT

Profiling cancer-associated genetic alterations and molecular classification of cancer in Korean gastric cancer patients

Yoonjung Kim

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Kyung-A Lee)

Recently, the Cancer Genome Atlas (TCGA) Research Network and Asian Cancer Research Group (ACRG) provided a new classification of gastric cancer to aid the development of biomarkers for targeted therapy and predict prognosis. We studied associations between genetically aberrant profiles of cancer-related genes, environmental circumstances, and histopathological features in 107 paired tumor-normal tissue gastric cancer (GC) samples. We simplified the molecular subtypes of gastric cancer according to the TCGA system. We classified 6.5% of our GC cases as the EBV subtype, 17.7% as the MSI subtype, 13.1% as the CIN subtype, and 62.6% as the GS subtype. The characteristics of each group were comparable with those of TCGA molecular subtypes. The MSI subtype showed a hyper-mutated status and the best prognosis, which was an identical finding with both the TCGA and ACRG classifications. The P619fs*43 in ZBTB20 (P619fs*43, n=5) was detected approximately 20% of MSI group and was limited to the MSI group. However,
the global molecular portrait of several genetic alterations on PIK3CA, JAK2, CD274, and PDCD1LG2 in EBV-infected GC were not consistent with EBV-infected GC from our study. And the proportion of the GS subtype (62.6%) was relatively larger than that in the TCGA cohort (20%). Some of these differences may be related to the ethnic origins of the patients. Especially, targeted sequencing is not sufficient to detect clusters of CNVs on the whole-exome scale and we focused on CNVs correspond possible homozygous deletion or high-level gain.

We did not observe a substantial difference in overall survival ($p = 0.2828$) or relapse-free survival (RFS, $p=0.3329$) among the four GC subtypes, in contrast to the groups classified according to AJCC stage (RFS, $p=0.004$; GCSS, $p=0.009$). \textit{H. pylori} infection (HR: 0.3, $p = 0.0322$) was associated with a favorable factor in Korean patients with gastric cancer. MSI subtype showed the most favorable prognosis in ACRG, TCGA in our data. Therefore, we thought that the predicting prognosis in GC patients might be performed more simply and effectively using both the MSI subtype and AJCC stage.

Genetic alterations of the RTK/RAS/MAPK and PI3K/PTEN/AKT pathways as well as the \textit{MET} gene were detected in 35.5% of GC cases ($n=38/107$). These mutations may facilitate enrollment of GC patients into clinical trials evaluating targeted therapies and provide the basis for developing solid therapeutic approaches in Korean GC patients.

We analyzed the germline genetic alterations to identify the inherited component of GC in the Korean population and found two pathogenic variants (NM_004360.4: c. 2494 G>A, \textit{p.V832M}) in the \textit{CDH1} gene. Our two cases (1.9%) with V832M were 66-year-old and 75-year-old patients

Here, we classified molecular subtypes using a modality with simplified steps and provide a critical starting point for the design of more appropriate clinical
trials based on a comprehensive analysis of genetic alterations in Korean GC patients.

Key words: Gastric cancer, molecular subtyping, germline mutation, somatic mutation
Profiling cancer-associated genetic alterations and molecular classification of cancer in Korean gastric cancer patients

Yoonjung Kim

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Kyung-A Lee)

I. INTRODUCTION

Gastric cancer is ranked fifth for cancer incidence and second for cancer deaths, and one in 36 men and 1 in 84 women develop stomach cancer before age 79 \(^1\). The histologic classification of gastric carcinoma has been based on the Lauren \(^2\) and 2010 WHO classification systems, which recognize four histological subtypes \(^3\). Neither the Lauren nor the WHO system is particularly clinically useful, as their prognostic and predictive capabilities cannot adequately guide patient management. New classifications are needed for gastric cancer to provide insights into pathogenesis and the identification of new biomarkers and novel treatment targets \(^4\). Recently, advances in technology and high-throughput analysis have improved our understanding of the genetic basis of gastric cancer. To provide a roadmap for patient stratification and trials of targeted therapies, the Cancer Genome Atlas (TCGA) Research Network has characterized 295 primary gastric adenocarcinomas and proposed the new classification of four different tumor subtypes of Epstein-Barr virus (EBV)-positive, microsatellite instability (MSI), genomically stable (GS), and chromosomal instability (CIN) subtypes. \(^5\). The Asian Cancer Research Group
(ACRG) also provided a new classification for gastric cancer, identifying four subtypes: MSI, MSS/EMT, MSS/TP53 (+), and MSS/TP53 (–). One of the most important aspects of the ACRG classification is that it correlates the molecular subtypes with clinical prognosis.

EBV is a human oncogenic gamma-herpes virus that is ubiquitously distributed; more than 90% of the world population is infected with EBV. EBV is now regarded as a GC-causing infectious agent. The vast majority of gastric cancers arise sporadically, and an inherited component contributes to <3% of gastric cancers. We investigated germline mutations in CDH1, MSH2, MLH1, TP53, APC, and STK11, which are associated with hereditary diffuse gastric cancer, hereditary nonpolyposis colon cancer, Li–Fraumeni syndrome, familial adenomatous polyposis, and Peutz–Jeghers syndrome, respectively. Gastric cancer is a heterogeneous disease characterized by epidemiologic and histopathological differences across countries. Here, we studied associations between the genetic aberrant profiles of cancer-related genes, environmental circumstances (EBV and H. pylori), and histopathological features in Korean gastric cancer patients. We also simplified the molecular subtypes of gastric cancer to help identify novel therapeutic strategies for patients with gastric cancer.

II. MATERIALS AND METHODS

1. Subject selection

We obtained a total of 107 gastric tumor and matched normal tissue samples from Yonsei University Wonju Medical Center Biobank (n=138, 69 paired samples) and Samkwang Medical Laboratory Biobank (n= 76, 38 paired samples). Tumor samples were obtained from patients who had not received
prior chemotherapy or radiotherapy. The gastric cancer tissues consisted of 69 fresh-frozen (FF) paired tumor and non-tumor tissue samples and 38 formalin-fixed paraffin-embedded (FFPE) paired tumor and non-tumor tissue samples. Clinical data, including age, sex, clinical follow-up data, and pathologic reports, were provided from the tissue source institutions. The histologic classification of gastric carcinoma has previously been based on Lauren’s criteria \(^2\) and the 2010 WHO classification system \(^3\). Tumor TNM stage assignment was evaluated for consistency with the 7th Edition of the TNM classification by the American Joint Committee on Cancer (AJCC) \(^21\). Pathologic findings were reviewed by experienced gastrointestinal pathologists (S.K. and M.J.). The study was approved by the Institutional Review Boards of Samkwang Medical Laboratories and Yonsei University Wonju College of Medicine.

2. DNA preparation

DNA was extracted from FFPE tumor and adjacent non-tumor gastric tissues using a QIAamp DNA extraction kit (Qiagen, Hilden, Germany) according to the manufacturer’s protocol. H&E-stained sections from FFPE blocks were reviewed by a board-certified pathologist, and representative sections with tumor content or benign tissue were identified. A G-DEX genomic DNA extraction kit (Intron Biotechnology, Korea) was used for FF tumor and matched normal FF tissues according to the manufacturer’s protocol. The quality and concentration of genomic DNA (gDNA) was evaluated by Nanodrop (ND-1000; Thermo Scientific, DE, USA) and the Agilent 2200 Tape Station system (Agilent Technologies, CA, USA) with Genomic DNA Screen Tape according to the manufacturer’s instructions. The DNA Integrity Number (DIN) for determining the integrity of gDNA was calculated from the electrophoretic trace on the 2200 Tape Station system according to the manufacturer’s instructions. The average value (range) of total DNA concentration in FFPE tissue and FF tissue was 322.7 (12.6 ~ 322.9) ng/μL and
952.8 (76.0 ~ 3756.0) ng/uL, respectively. The average value (range) of DIN in FFPE tissue and FF tissue was 2.9 (1.5 ~ 6.4) and 5.6 (1.4 ~ 8.4), respectively.

3. Detection of EBV and H. pylori infection

EBV infection was detected using the Real-Q EBV quantification kit (Biosewoom, Seoul, Korea) and CFX96 real-time PCR system (Bio-Rad, USA) following the manufacturer’s recommendations.

H. pylori infection was detected using Giemsa stain or PCR amplification and sequencing. Primers for PCR and sequencing were derived from a known sequence of the 23S rRNA gene (GenBank Accession No. U27270), as previously described (sense, 5'-CGT AAC TAT AACGGT CCT AAG-3', positions 2365 to 2385; antisense, 5'-TTA GCT AAC AGA AAC ATC AAG-3', positions 2635 to 2653) 22.

4. Configuration of a gastric cancer-related target gene panel for Korean gastric cancer patients

The cancer panel consisted of genes based on the Mutation Analysis (MutSig 2CV v3.1) results of the Cancer Genome Atlas (TCGA) project (http://gdac.broadinstitute.org/runs/analyses_latest/reports/cancer/STAD-TP/index.html, accessed at 2015.03.30) 5,23 and significantly mutated genes in the Asian Cancer Research Group (ACRG) cohort and SMC-2 cohort in primary gastric cancer tissues 6. Genes associated with new targeted therapy of GC (EGFR, ERBB2, FGFR2, and VEGFR2 (KDR)) and hereditary cancer syndromes (CDH1, MSH2, MLH1, STK11, and TP53) were also included in the cancer panel 13.

5. Targeted sequencing and data analysis
DNA fragments of matched tumor and non-tumor tissues were enriched by solution-based hybridization capture, followed by sequencing with the Illumina HiSeq2500 platform (Illumina, San Diego, CA, USA) with the 2 × 125 bp paired-end read module. gDNA was sheared using an Adaptive Focused Acoustics (AFA)™ with the Covaris Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA). The quality and quantity of sheared DNA were assessed using the Agilent 2200 Tape Station system with Agilent D1000 ScreenTape (Agilent Technologies, USA) according to the manufacturer’s instructions. Capture probes for the coding exons of 43 genes (Supplementary 1) were generated by Celemics (Seoul, Korea). Purification and clean-up of samples were performed using a DynaMag™-50 Magnet (Thermo Fisher Scientific Inc., Waltham, MA, USA) with Agencourt® AMPure® XP Kit (Beckman Coulter, Brea, CA, USA). NGS library amplification was performed using a KAPA Library Amplification Kit (Kapa Biosystems, Inc., Wilmington, MA, USA) according to the manufacturer’s instructions. Library preparation, hybridization, capture procedure, and sequencing on the Illumina HiSeq2500 genome analyzer were performed by Celemics according to the protocols recommended by the Celemics User Manual Ver 2.1 (http://www.celemics.com/home/).

The generated reads were trimmed and filtered by Trimmomatic 24 and then mapped against the UCSC hg19 Genome Reference Consortium Human Reference 37 (GRCh37) (http://genome.ucsc.edu/) using the Burrows-Wheeler Aligner (BWA) 25. Picard (http://broadinstitute.github.io/picard/), SAMTools 26, and Genome Analysis Toolkit (GATK,https://www.broadinstitute.org/gatk/) 27 were used for post-processing alignments, base quality score recalibration, and short insertion/deletion (indel) realignment. After variant calling, variants were added to the annotation using ANNOVAR (http://www.openbioinformatics.org/annovar/) 28 and Variant Effect Predictor.
We used VarScan 2 (http://varscan.sourceforge.net) for the detection of somatic single-nucleotide variants (SNVs) and indels. We modified the algorithms for assessing copy number aberrations (CNAs), which were developed for NGS protocols based on hybridization-capture methods in the setting of whole-exome sequencing. We replaced the average coverage of exon pull-down regions with average read counts of regions of interest (ROIs). The ratios of the coding exon segment mean depths from tumor samples and matched non-tumor samples were normalized via log₂-transformation. Performance was measured for CNAs (6–7 copies) and at lower sample purities (20%–30%), with an overall sensitivity >80%. We focused primarily on high amplifications and losses in this study (log₂ copy number ratio: >1.5 or <-1.5).

All acquired candidate variations went through post filters recommended by the authors of these tools. Germline variations were extracted with downstream analysis based on the variant allele frequency of <0.001 in the 1000 Genome Project (http://www.1000genomes.org), ESP6500 (http://evs.gs.washington.edu/EVS/), and Exome Aggregation Consortium (ExAC, http://exac.broadinstitute.org/). We extracted somatic mutations with VarScan2 and post-filtered with downstream analysis for altered allele frequency in tumors > 5%, > 50 x coverage, exonic variants, and population frequency 0.005 less than in the 1000 Genome Project, ESP6500, and ExAC databases. We excluded somatic variants detected >2 times in non-tumor tissue. Visual inspection of filtered calls was performed using Integrated Genomics Viewer 2.3 software (IGV; Broad Institute, Cambridge, MA, USA).

6. Microsatellite Instability (MSI) Assay

Microsatellite status was assessed by the mononucleotid repeat markers BAT-25, BAT-26, NR-21, NR-24, and NR-27 in tumor and
corresponding normal tissues. The five markers were co-amplified in multiplex PCRs performed with Solg2X multiplex PCR Smart mix following the manufacturer’s recommendations. The amplified PCR products were analyzed using the ABI 3500Dx system (Applied Biosystems, Foster City, CA, USA) and GeneMarker software (SoftGenetics, PA, USA). Tumors with two or more of the five markers showing instability were judged as high-frequency MSI (MSI-H), and tumors showing instability in only one locus were classified as low-frequency MSI (MSI-L).

7. Statistical Analysis

Fisher’s exact and Chi-squared tests were performed to evaluate differences in the respective proportion of several factors between subgroups. Patient follow-up periods were calculated as time between date of surgery and date of last follow-up (months). Relapse-free survival (RFS) was assessed based on the absence of loco-regional recurrence, distant metastasis, and death from any cause. GC-specific survival (GCSS) was calculated only for patients who died from any GC-related cause. Kaplan-Meier survival curves with log-rank tests were performed to compare RFS and GCSS according to AJCC stage and molecular subtype. Cox proportional hazard models were performed to assess the influence of prognostic factors on RFS. Univariate analyses for age at surgery, AJCC stage, histologic type, lymphatic invasion, venous invasion, perineural invasion, H. pylori infection, and genomic profile were performed. Any factors from the univariate analysis with a p value less than 0.10 were included in the multivariate analysis. All statistical analyses were performed using SPSS 20.0 (SPSS, Chicago, IL, USA) and MedCalc Software (https://www.medcalc.org/). Except for the univariate analysis, a p value less than 0.05 was regarded as significant.
III. RESULTS

Clinical and pathological findings of 107 Korean GC patients

The male to female ratio of gastric cancer patients was 6:3, and the median patient age was 70 years (range, 32-90). Lauren intestinal type, diffuse type, and mixed type accounted for 53.3%, 26.2%, and 19.6% of cases, respectively. In addition, for gastric cancer classified according to the 2010 WHO classification, the tubular type (64.5%) was observed with the highest frequency, while the poorly cohesive type and mixed type accounted for 13.1% and 15.0% of cases, respectively. More than half of the tumors were located in the antrum or antrum body, and about 7.5% were located in the cardia and gastroesophageal junction. Fifty-two cases (48.6%) were Stages III-IV, and 66 (61.7%) were *H. pylori*-positive. The clinicopathological findings of the 107 gastric cancer patients are summarized in Table 1.

Table 1. Clinicopathological characteristics of patients with gastric cancer (n=107)

| Characteristics       | Number         |
|-----------------------|----------------|
| Age (year)            |                |
| Median (range)        | 70 (32 ~ 90)   |
| Sex                   |                |
| Female                | 36 (33.6%)     |
| Male                  | 71 (66.4%)     |
| Lauren class          |                |
| Diffuse               | 28 (26.2%)     |
| Intestinal            | 57 (53.3%)     |
| Mixed                 | 21 (19.6%)     |
| WHO class             |                |
| Mucinous              | 3 (2.8%)       |
Tubular 69 (64.5%)
Poorly cohesive 14 (13.1%)
Mixed (Tubular_Poorly cohesive) 16 (15.0%)
Uncommon histologic variants 5 (4.7%)

**pT stage**

| pT Stage | Count |
|----------|-------|
| T1a/T1b  | 9 (8.4%)/15 (14.0%) |
| T2       | 11 (10.3%) |
| T3       | 38 (35.5%) |
| T4a/T4b  | 33 (30.8%)/1 (0.9%) |

**pN stage**

| pN Stage | Count |
|----------|-------|
| N0/N1/N2/N3 | 40 (37.4%)/16 (15.0%)/20 (18.7%)/31 (29.0%) |

**M stage**

| M Stage | Count |
|---------|-------|
| M0/M1   | 102 (95.3%)/5 (4.7%) |

**AJCC stage**

| AJCC Stage | Count |
|------------|-------|
| Stages IA/IB | 18 (16.8%)/11 (10.3%) |
| Stages IIA/IIB | 13 (12.1%)/13 (12.1%) |
| Stages IIIA/IIIB/IIIC | 17 (15.9%)/11 (10.3%)/19 (17.8%) |
| Stage IV    | 5 (4.7%) |

**Anatomical regions**

| Anatomical Region | Count |
|-------------------|-------|
| GEJ_Cardia        | 8 (7.5%) |
| Fundus_Body       | 36 (34.3%) |
| Antrum            | 54 (50.5%) |
| Antrum_Body       | 5 (4.7%) |
| Pylorus           | 3 (2.8%) |
| Diffuse           | 1 (0.9%) |

**Epstein-Barr virus infection**

| Epstein-Barr Virus Infection | Count |
|------------------------------|-------|
| Negative                     | 100 (93.5%) |
| Positive                     | 7 (6.5%) |

**Microsatellite instability (MSI)**

| Microsatellite Instability | Count |
|----------------------------|-------|
| MSS                        | 88 (82.2%) |
| MSI-I                      | 4 (3.7%) |
| MSI-H                      | 15 (14.0%) |
**H. pylori** infection

|              |       |
|--------------|-------|
| Negative     | 41 (38.3%) |
| Positive     | 66 (61.7%) |

Abbreviations: GEJ, gastroesophageal junction; pT stage, pathological assessment of the primary tumor (pT); pN stage, pathological assessment of the regional lymph nodes (pN)

**Germline variation analysis of hereditary cancer-predisposing syndrome**

To identify germline mutations, variants were extracted from targeted sequencing using a 43 gene cancer panel on 107 GC tissues. Calls that met all the standards noted herein were retained for downstream analysis, including altered allele frequency > 30%, > 50x coverage, and population frequency less than 0.001 in the 1000 Genome Project, ESP6500, and ExAC databases. These variants were present in both gastric cancer and matched normal tissue. A total of 31 germline variants were observed in **TP53, STK11, ALK, APC, MSH2, MLH1, and CDH1** (Table 2.). Among these, 17 variants, 12 variants, and 2 variants were classified as 'Benign or Likely Benign,' 'VUS,' and 'Likely pathogenic,' respectively. Two likely pathogenic variants (p.V832M) were detected on the **CDH1** gene.
Table 2. Germline SNPs in *TP53*, *STK11*, *ALK*, *APC*, *MSH2*, *MLH1*, and *CDH1* genes of 107 Korean gastric cancer patients

| Sample  | Chr | START   | END     | Ref | Alt | Gene   | Transcript ID | Amino acid change | Total depth | VAF(%) (1) | VAF(%) (2) | Clinical significance |
|---------|-----|----------|---------|-----|-----|--------|---------------|-------------------|-------------|-----------|-----------|----------------------|
| YMC 53  | 3   | 37053562 | 37053562| C   | T   | MLH1   | NM_000249.3   | p.R217C         | 1397        | 46.96%    | 50.40%    | VUS                  |
| YMC 54  | 2   | 29449820 | 29449820| G   | A   | ALK    | NM_004304.4   | p.T1012M        | 637         | 48.51%    | 60.36%    | Benign               |
| YMC 55  | 3   | 37042521 | 37042521| T   | G   | MLH1   | NM_000249.3   | p.S95A          | 233         | 39.06%    | 30.52%    | VUS                  |
| YMC 6   | 3   | 37067240 | 37067240| T   | A   | MLH1   | NM_000249.3   | p.V384D         | 578         | 48.79%    | 44.85%    | Benign               |
| YMC 7   | 2   | 29449820 | 29449820| G   | A   | ALK    | NM_004304.4   | p.T1012M        | 208         | 55.77%    | 43.83%    | Benign               |
| YMC 66  | 2   | 47656972 | 47656972| C   | T   | MSH2   | NM_000251.2   | p.L390F         | 197         | 56.85%    | 52.78%    | Benign               |
| YMC 70  | 2   | 29449820 | 29449820| G   | A   | ALK    | NM_004304.4   | p.T1012M        | 755         | 52.58%    | 57.85%    | Benign               |
| YMC 70  | 3   | 37067240 | 37067240| T   | A   | MLH1   | NM_000249.3   | p.V384D         | 547         | 51.55%    | 47.99%    | Benign               |
| YMC 13  | 5   | 11217777 | 11217777| A   | C   | APC    | NM_001127511.2| p.K2145Q        | 151         | 52.32%    | 33.74%    | Benign               |
| YMC 14  | 3   | 37053562 | 37053562| C   | T   | MLH1   | NM_000249.3   | p.R217C         | 1409        | 45.71%    | 51.65%    | VUS                  |
| Sample | ID  | Chromosome | Allele | Gene   | Transcript | Mutation | Reference Allele | p.Amino Acid Change | Probability | Interpretation  \
|--------|-----|------------|--------|--------|------------|----------|------------------|---------------------|-------------|---------------- \
| YMC 15 | 2   | 47656972   | C      | MSH2   | NM_000251.2| p.L390F  | 1281             | 46.68%              | 79.62%     | Benign         \
| YMC 22 | 3   | 37067240   | T      | MLH1   | NM_000249.3| p.V384D  | 862              | 99.54%              | 100%       | Benign         \
| YMC 22 | 3   | 37067240   | T      | MLH1   | NM_000249.3| p.V384D  | 862              | 99.54%              | 100%       | Benign         \
| YMC 24 | 5   | 11217886   | G      | APC    | NM_001127511.2| p.R2507H | 1142             | 47.90%              | 49.52%     | VUS            \
| YMC 3  | 3   | 37090506   | C      | MLH1   | NM_000249.3| p.Q701K  | 369              | 51.49%              | 41.29%     | Benign         \
| YMC 28 | 2   | 47630344   | C      | MSH2   | NM_000251.2| p.P5Q   | 231              | 53.25%              | 86.79%     | VUS            \
| YMC 29 | 16  | 68867247   | G      | CDH1   | NM_004360.4| p.V832M | 1337             | 52.21%              | 49.75%     | Likely pathogenic \
| YMC 4  | 3   | 37089022   | C      | MLH1   | NM_000249.3| p.L582V | 364              | 48.63%              | 57.47%     | VUS            \
| YMC 37 | 3   | 37053562   | C      | MLH1   | NM_000249 | p.R217C  | 1315             | 49.20%              | 49.76%     | VUS            \
| YMC 37 | 16  | 68867247   | G      | CDH1   | NM_004360.4| p.V832M | 1498             | 50.73%              | 46.41%     | Likely pathogenic \
| YMC 5  | 2   | 47637371   | A      | MSH2   | NM_000251.2| p.I169V  | 711              | 51.34%              | 45.12%     | Likely benign  \
| YMC 48 | 2   | 47656972   | C      | MSH2   | NM_000251.2| p.L390F  | 692              | 51.30%              | 53.31%     | Benign         \
| YMC 48 | 3   | 37067240   | T      | MLH1   | NM_000249.3| p.V384D  | 899              | 43.38%              | 44.61%     | Benign         \
| SKW 40 | 17  | 7578209    | G      | TP53   | NM_000546.5| p.H214Y  | 168              | 39.29%              | 59.58%     | VUS            \
| SKW 31 | 3   | 37053562   | C      | MLH1   | NM_000249.3| p.R217C  | 294              | 55.78%              | 39.71%     | VUS            \

15
| Sample | Chromosome | Start | End | Gene | Reference | Mutation | Codon | Precision | Sensitivity | Risk | Conclusion |
|--------|-----------|-------|------|------|-----------|----------|-------|-----------|-------------|-------|------------|
| SKW 41 | 2         | 47703564 | 47703564 | G  A | MSH2     | NM_000251.2 | p.M688I | 993 | 52.57%     | 50.83%  | VUS        |
| SKW 38 | 5         | 11217654 | 11217654 | G  C | APC      | NM_001127511.2 | p.A1735P | 628 | 45.86%     | 47.30%  | VUS        |
| SKW 21 | 5         | 11217389 | 11217389 | A  C | APC      | NM_001127511.2 | p.E850D | 147 | 55.78%     | 50.93%  | Benign     |
| SKW 18 | 2         | 29519923 | 29519923 | G  A | ALK      | NM_004304.4 | p.L550F | 155 | 38.06%     | 55.39%  | VUS        |
| SKW 16 | 16        | 68856080 | 68856080 | C  G | CDH1     | NM_004360.4 | p.L630V | 928 | 43.74%     | 40.61%  | Benign     |
| SKW 15 | 16        | 68856080 | 68856080 | C  G | CDH1     | NM_004360.4 | p.L630V | 827 | 49.33%     | 71.46%  | Benign     |

Abbreviation: VUS, a variant of unknown significance
Somatic mutation analysis of gastric cancer with a 43 gene cancer panel

To identify driver genes causally linked to tumorigenesis in Korea GCs, variants were extracted from targeted sequencing with a 43 gene cancer panel applied to 107 tumor and matched non-tumor tissues. After mutation calling and stringent filtrations, we identified 317 SNVs on coding sequences that included missense variations (n=164, 54.5%), trunc (nonsense and frameshift) variations (n= 110, 36.5%), in-frame variations (n= 42, 14.0%), and splicing site variations (n=1, 0.3%) (Supplemental Table 2). Somatic variants detected on each gene in each sample were summarized in supplemental table 3. Among the 43 genes, we discovered 39 genes that were mutated in one or more individual samples. Among the 107 samples, TP53 (38.3%), ARID1A (36.4%), CRI (14.0%), APC (11.2%), BCOR (11.2%), CDH1 (10.3%), CIC (9.3%), PIK3CA (9.3%), RHOA (8.4%), ERBB3 (8.4%), ERBB2 (7.5%), CCND1 (6.5%), FBXW7 (6.5%), ALK (5.6%), KRAS (5.6%), and MTOR (5.6%) were identified (Supplemental Table 3). Somatic variants were detected in relatively low frequencies (less than 5%) on the CTNNB1, EGFR, HLA-B, MSH2, PGM5, ZBTB20, IRF2, KDR, LARP4B, MVK, BRAF, PTEN, ACVR1B, CBWD1, FGFR2, JAK2, MEDAG, MLH1, SMAD4, STK11, CD274, MDM2, and MYC genes in this study. We did not detect any somatic variants on the C16orf74, CCNE1, PDCD1LG2, or MET genes (Supplemental Table 3).

The location specific incidences of somatic variants are depicted in Figure. According to the results of lollipop bar chart analysis, p.Gln1334del (n=23/52) in ARID1A, p.Gln1174ThrfsTer8 (n=3/12) in BCOR, p.Glu280del (n=7/7) in CCND1, p.Leu15del (n=5/12) in CDH1, p.Arg2194Ter (n=10/16) in CRI, p.Arg678Gln (n=4/10) in ERBB2, p.Arg385His (n=3/9) in FBXW7, p.Gly13Asp (n=5/6) in KRAS, p.Ile98Val (n=3/6) in PGM5, p.His1047Arg/Tyr (n=5/14) in PIK3CA, p. Arg5Gln/Trp (n=3/9) in RHOA, and p.Pro619LeufsTer43 (n=5/5) in ZBTB20 were recurrently detected in more
than three individual cases. Among 317 somatic variants, 106 (33.4%) were recurrently detected in this study (Table 3 & Figure 1).
Figure 1. Lollipop bar graph representing distribution of mutations across the 39 genes ACVR1B, ALK, APC, ARID1A, BCOR, BRAF, CBWD1, CCND1, CD274, CDH1, CIC, CR1, CTNNB1, EGFR, ERBB2, ERBB3, FBXW7, FGFR2, HLA-B, IRF2, JAK2, KDR, KRAS, LARP4B, MDM2, MEDAG, MLH1, MSH2, MTOR, MVK, MYC, PGM5, PIK3CA, PTEN, RHOA, SMAD4, STK11, TP53, and ZBTB20.
Table 3. The location-specific recurrence of somatic variants in the 43 cancer panel genes

| Gene    | Variants<sup>a</sup> | Mutated samples | % of mutated samples | COMIC counts | COMIC ID                                      |
|---------|----------------------|-----------------|----------------------|--------------|-----------------------------------------------|
| APC     | *p.Glu1464ValfsTer8* | 5               | 4.70%                | 45           | COSM1432412; COSM19694; COSM41622; COSM41622; COSM41622; COSM5030795 |
| ARID1A  | *p.Asp1850ThrfsTer33* | 2               | 1.90%                | 27           | COSM1341426; COSM133001; COSM1666860; COSM1341408; COSM298325; COSM1578346; COSM133030; COSM51218; COSM1238047 |
| ARID1A  | *p.Gln1334del/dup*  | 23              | 21.50%               | 23           | -                                            |
| ARID1A  | *p.Gly87del*         | 2               | 1.90%                | -            | (-)                                           |
| BCOR    | *p.Gln1174ThrfsTer8* | 3               | 2.80%                | 5            | COSM1683572; COSM3732385                      |
| Gene   | Variant          | Count | Frequency | ID(s)                 |
|--------|------------------|-------|-----------|-----------------------|
| BRAF   | p.Pro403LeufsTer8 | 2     | 1.90%     | COSM1448632; COSM5347158; COSM5347157 |
| CCND1  | p.Glu280del      | 7     | 6.50%     | COSM931394            |
| CDH1   | p.Leu15del       | 5     | 4.70%     | (-)                  |
| CR1    | p.Arg2194Ter     | 10    | 9.30%     | COSM301989            |
| ERBB2  | p.Arg678Gln      | 4     | 3.70%     | COSM978678; COSM436498 |
| ERBB2  | p.Ser310Phe      | 2     | 1.90%     | COSM48358; COSM1666868 |
| ERBB2  | p.Val842Ile      | 2     | 1.90%     | COSM14065; COSM1666633 |
| ERBB3  | p.Val104Met      | 2     | 1.90%     | COSM20710             |
| FBXW7  | p.Arg385His      | 3     | 2.80%     | COSM117308            |
| Gene   | Mutation                      | Count | Frequency | ID(s)          |
|--------|-------------------------------|-------|-----------|----------------|
| HLA-B  | p.Glu69Val                    | 2     | 1.90%     | COSM4598273    |
| KRAS   | p.Gly13Asp                    | 5     | 4.70%     | COSM532        |
| LARP4B | p.Thr163HisfsTer47            | 2     | 1.90%     | COSM1638669; COSM4968611 |
| MVK    | p.Ala141ArgfsTer18            | 2     | 1.90%     | COSM1241457    |
| PIK3CA | p.His1047Arg                  | 4     | 3.70%     | COSM775        |
| PGM5   | p.Ile98Val                    | 3     | 2.80%     | COSM1109610    |
| RHOA   | p.Arg5Gln/Trp                 | 3     | 2.80%     | COSM190569; COSM446704; COSM4770224; COSM4770223 |
| RHOA   | p.Thr37Ala/Ile                | 2     | 1.90%     | COSM5064959; COSM1223700 |
| RHOA   | p.Tyr42Cys                    | 2     | 1.90%     | COSM2849892; COSM4770225 |
| Gene   | Mutation     | Count | Frequency | Sample Size |
|--------|--------------|-------|-----------|-------------|
| TP53   | p.Arg248Trp  | 2     | 1.90%     | 56          |
| TP53   | p.Arg342Ter  | 2     | 1.90%     | 151         |
| TP53   | p.Cys275Tyr  | 2     | 1.90%     | 62          |
| TP53   | p.Val173Leu  | 2     | 1.90%     | 65          |
| ZBTB20 | p.Pro619LeufsTer43 | 5 | 4.70% | 26 |

*Recurrent mutations observed in at least two samples.*
Copy number analysis in 107 gastric cancer tissues and identification of clinical relevant CNVs

Twenty-nine CNVs were identified in the genes *BRAF, C16orf74, CCND1, CIC, ERBB2, FBXW7, FGFR2, HLA-B, KRAS, MET, MYC, PDCD1LG2, PTEN, and TP53* (Table 3 & Figure 2). CNVs in the RTK/RAS/MAPK/ERK signaling pathway were harbored in 10 cases (9.4%). Five samples (4.7%) contained an *ERBB2* amplification, and one sample (0.93%) had an *FGFR2* amplification. CNVs in *KRAS, MET,* and *BRAF* were detected in 2 cases (1.9%), 1 case (0.93%), and 1 case (0.93%), respectively. CNVs in *EGFR* and *VEGFR2* were not detected. CNVs in the cell cycle-related genes *CCND1* and *CCNE1* were detected in 5 cases (4.7%). CNVs in *C16orf74* (4.8%) were the most frequent detected in this study. Only one CNV was detected in each of *PDCD1LG2, HLA-B, CIC, FBXW7, PTEN,* and *TP53,* and two samples harbored *MYC* amplifications (Table 4).
Table 4. Copy number analysis in 107 gastric cancer tissues

| Gene       | AMP (Log$_2$ CN ratio) | DEL (Log$_2$ CN ratio) | No of CNVs | Case (%) |
|------------|------------------------|------------------------|------------|----------|
| C16orf74   | 3 (1.6 ~2.9)           | 2 (-1.8 ~ -2.0)        | 5          | 4.7      |
| CCND1      | 3 (1.5 ~2.1)           |                        | 3          | 2.8      |
| CCNE1      | 2 (1.9 ~2.0)           |                        | 2          | 1.9      |
| CIC        | 2 (1.8 ~2.2)           |                        | 2          | 1.9      |
| ERBB2      | 5 (1.9~4.3)            |                        | 5          | 4.7      |
| FBXW7      |                        | 1 (-2.0)               | 1          | 0.9      |
| FGFR2      | 1 (3.3)                |                        | 1          | 0.9      |
| HLA-B      | 1 (1.5)                |                        | 1          | 0.9      |
| KRAS       | 2 (2.2)                |                        | 2          | 1.9      |
| MET        | 1 (2.0)                |                        | 1          | 0.9      |
| MYC        | 2 (2.0 ~2.3)           |                        | 2          | 1.9      |
| PDCD1L G2  | 1 (2.5)                |                        | 1          | 0.9      |
| PTEN       |                        | 1 (-2.0)               | 1          | 0.9      |
| TP53       |                        | 1 (-1.6)               | 1          | 0.9      |
| BRAF       |                        | 1(-1.6)                | 1          | 0.9      |
| **Total**  | 23                     | 6                      | 29         | **27.31%** |

Abbreviations: AMP, amplification; DEL, deletion; CNV, copy number variations. A total of 17 cases harbored amplifications/deletion of target genes, including 5 cases with more than 2 CNVs.
Figure 2. Summary of all significant CNAs across 107 gastric tumors.

(a) Gains and losses are indicated by red squares and green squares, respectively, in the heatmap, b) Significant ($p < 0.05$) gains and losses are indicated by yellow squares and blue squares, respectively, whereas gray squares indicate genes without significant gain or loss.
**Molecular subtype classification and clinical phenotype**

We classified molecular subtypes using genomic data according to subtypes derived by TCGA and correlated clinical covariates of 107 gastric cancer patients with those molecular subtypes (Table 4). The EBV subtype (6.5% of GC) was significantly enriched in EBV burden and characterized as diffuse type and uncommon histologic variants histological subtype (Tables 5 & 6). In the EBV subtype, no samples with a TP53 mutation were detected, but mutations of ARID1A (4 cases, 57.1%), CDH1 (3 cases, 42.9%), and PIK3CA (2 cases, 28.6%) were present with a relatively high frequency. Genetic alterations of the JAK2 and PDL2 genes were not detected in the EBV subgroup. Only one case harbored the mutant CD274 (Figure 3 & Tables 5).

The MSI subtype (17.7% of GC) showed instability in one more locus in the MSI assay. The MSI subtype presented with an elevated mutation rate (6.9 per case) and was characterized by alterations of genes involved in mismatch repair. Mutant MSH2 (n=4/5) and 100% of MLH1 mutations (2 cases) were observed in this subtype. Mutations of BCOR (14.3%) ERBB2 (26.3%), KRAS (26.3%), PIK3CA (36.8%), MTOR (15.8%), ERBB3 (26.3%), EGFR (21.1%), MSH2 (21.1%), MLH1 (10.5%), ZBTB20 (26.3%), LARP4B (15.8%), and MVK (21.1%) were significantly present (Table 5). Interestingly, we observed that mutations of ZBTB20 were limited to the MSI group. Somatic mutations in HLA-B as major histocompatibility complex class I genes were detected in 3 cases (15.8%) of the MSI subtype.

The CIN subtype (13.1% of GC) was characterized by gene amplifications and deletions (13.1% of GC) (Figure 3.) and showed intestinal and tubular histological subtypes (Tables 5&6). The CIN subtype was also characterized by relatively a low-somatic mutation rate (1.9 per case) and a
high frequency of TP53 mutations. In the CIN subtype, we observed amplifications of ERBB2 (35.7%), FGFR2 (7.1%), KRAS (7.1%), and MET (7.1%) and somatic mutation of ERBB3 (7.1%). The genetic alterations of ERBB2, FGFR2, KRAS, MET, and ERBB3 were mutually exclusive to the CIN subtype. Therefore, 50.7% of the CIN subtype harbored amplification of genes belonging to the RTK/RAS/MAPK signaling pathway (Figure 3 & Table 5).

The GS subtype (62.6% of GC) was characterized by a lack of EBV infection, MSI, and somatic CNAs. The GS subtype showed the lowest mutation rate (1.6 per case) and the highest frequency of TP53 mutations (44.8%). In the GS subtype, limited mutations of RHOA (9.0%, 6 cases) were detected in the diffuse or mixed histologic type of Lauren's criteria (Figure 3 & Table 5).
Table 5. Somatic mutations in each subtype

| Somatic mutations | EBV (n=7) | MSI (n=19) | CIN (n=14) | GS (n=67) | \( p \) value |
|-------------------|----------|------------|------------|-----------|--------------|
|                   | n (%)    | n (%)      | n (%)      | n (%)     |              |
| **TP53**          | 0 0.0%   | 5 26.3%    | 6 42.9%    | 30 44.8%  | 0.690        |
| **ARID1A**        | 4 57.1%  | 14 73.7%   | 3 21.4%    | 18 26.9%  | 0.001        |
| **CR1**           | 0 0.0%   | 5 26.3%    | 2 14.3%    | 8 11.9%   | 0.334        |
| **APC**           | 2 28.6%  | 1 5.3%     | 3 21.4%    | 6 9.0%    | 0.148        |
| **BCOR**          | 1 14.3%  | 9 47.4%    | 0 0.0%     | 2 3.0%    | 0.000        |
| **CDH1**          | 3 42.9%  | 1 5.3%     | 0 0.0%     | 7 10.4%   | 0.034        |
| **CIC**           | 0 0.0%   | 5 26.3%    | 0 0.0%     | 5 7.5%    | 0.061        |
| **PIK3CA**        | 2 28.6%  | 7 36.8%    | 1 7.1%     | 0 0.0%    | 0.000        |
| **ERBB3**         | 0 0.0%   | 5 26.3%    | 1 7.1%     | 3 4.5%    | 0.037        |
| **RHOA**          | 2 28.6%  | 1 5.3%     | 0 0.0%     | 6 9.0%    | 0.191        |
| **ERBB2**         | 0 0.0%   | 5 26.3%    | 0 0.0%     | 3 4.5%    | 0.022        |
| **CCND1**         | 0 0.0%   | 3 15.8%    | 2 14.3%    | 2 3.0%    | 0.076        |
| **FBXW7**         | 0 0.0%   | 3 15.8%    | 2 14.3%    | 2 3.0%    | 0.076        |
| **ALK**           | 0 0.0%   | 3 15.8%    | 0 0.0%     | 3 4.5%    | 0.216        |
| **KRAS**          | 0 0.0%   | 5 26.3%    | 0 0.0%     | 1 1.5%    | 0.004        |
| **MTOR**          | 0 0.0%   | 3 15.8%    | 2 14.3%    | 1 1.5%    | 0.028        |
| **CTNNB1**        | 1 14.3%  | 2 10.5%    | 1 7.1%     | 1 1.5%    | 0.096        |
| **EGFR**          | 0 0.0%   | 4 21.1%    | 0 0.0%     | 1 1.5%    | 0.016        |
| **HLA-B**         | 0 0.0%   | 3 15.8%    | 0 0.0%     | 2 3.0%    | 0.129        |
| **MSH2**          | 0 0.0%   | 4 21.1%    | 0 0.0%     | 1 1.5%    | 0.016        |
| **PGM5**          | 1 14.3%  | 4 21.1%    | 0 0.0%     | 0 0.0%    | 0.001        |
| Gene      | Cases | % | Cases | % | Cases | % | Cases | % | Cases | % | Cases | % | Cases | % | Cases | % | Cases | % |
|-----------|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|-------|---|
| ZBTB20*   | 0     | 0.0% | 5 | 26.3% | 0 | 0.0% | 0 | 0.0% | 0.000 | IRF2 | 0 | 0.0% | 2 | 10.5% | 0 | 0.0% | 2 | 3.0% | 0.431 |
| KDR       | 0     | 0.0% | 3 | 15.8% | 0 | 0.0% | 1 | 1.5% | 0.062 | LARP4B* | 0 | 0.0% | 3 | 15.8% | 1 | 7.1% | 0 | 0.0% | 0.012 |
| MVK*      | 0     | 0.0% | 4 | 21.1% | 0 | 0.0% | 0 | 0.0% | 0.030 | BRAF | 0 | 0.0% | 3 | 15.8% | 0 | 0.0% | 0 | 0.0% | 0.140 |
| PTEN      | 0     | 0.0% | 2 | 10.5% | 0 | 0.0% | 1 | 1.5% | 0.223 | ACVR1B | 0 | 0.0% | 1 | 5.3% | 1 | 7.1% | 0 | 0.0% | 0.138 |
| CBWD1     | 0     | 0.0% | 0 | 0.0% | 0 | 0.0% | 2 | 3.0% | 1.000 | FGFR2 | 0 | 0.0% | 2 | 10.5% | 0 | 0.0% | 0 | 0.0% | 0.091 |
| JAK2      | 0     | 0.0% | 1 | 5.3% | 0 | 0.0% | 1 | 1.5% | 0.610 | MEDAG | 0 | 0.0% | 2 | 10.5% | 0 | 0.0% | 0 | 0.0% | 0.091 |
| MLH1      | 0     | 0.0% | 2 | 10.5% | 0 | 0.0% | 0 | 0.0% | 0.091 | SMAD4 | 0 | 0.0% | 1 | 5.3% | 0 | 0.0% | 1 | 1.5% | 0.610 |
| STK11     | 0     | 0.0% | 1 | 5.3% | 1 | 7.1% | 0 | 0.0% | 0.138 | CD274 | 1 | 14.3% | 0 | 0.0% | 0 | 0.0% | 0 | 0.0% | 0.065 |
| MDM2      | 0     | 0.0% | 1 | 5.3% | 0 | 0.0% | 0 | 0.0% | 0.374 | MYC   | 1 | 14.3% | 0 | 0.0% | 0 | 0.0% | 0 | 0.0% | 0.065 |

*\(^a\)p value less than 0.05, n; number of case

Fisher’s exact test was performed to evaluate differences in the respective proportions of somatic mutations between subgroups.
Table 6. Patient characteristics according to molecular subtype

|                   | EBV (n=7)   | MSI (n=19)  | CIN (n=14)  | GS (n=67)   | P value |
|-------------------|-------------|-------------|-------------|-------------|---------|
| **Age**           | Cases (%)   | Cases (%)   | Cases (%)   | Cases (%)   |         |
| ≤50               | 0 (0.0%)    | 1 (5.3%)    | 1 (7.1%)    | 8 (11.9%)   |         |
| 51 ~60            | 2 (28.6%)   | 1 (5.3%)    | 2 (14.3%)   | 13 (19.4%)  |         |
| 61 ~70            | 1 (14.3%)   | 3 (15.8%)   | 4 (28.6%)   | 17 (25.4%)  | 0.189   |
| 71 ~80            | 3 (42.9%)   | 12 (63.2%)  | 3 (21.4%)   | 26 (38.8%)  |         |
| >80               | 1 (14.3%)   | 2 (10.5%)   | 4 (28.6%)   | 3 (4.5%)    |         |
| **Sex**           | Cases (%)   | Cases (%)   | Cases (%)   | Cases (%)   |         |
| Female            | 1 (14.3%)   | 8 (42.1%)   | 5 (35.7%)   | 22 (32.8%)  | 0.608   |
| Male              | 6 (85.7%)   | 11 (57.9%)  | 9 (64.3%)   | 45 (67.2%)  |         |
| **Lauren Class**  |             |             |             |             |         |
| Diffuse           | 3 (42.9%)   | 2 (10.5%)   | 1 (7.1%)    | 22 (32.8%)  |         |
| Intestinal        | 2 (28.6%)   | 15 (78.9%)  | 11 (78.6%)  | 30 (44.8%)  | 0.047   |
| Mixed             | 2 (28.6%)   | 2 (10.5%)   | 2 (14.3%)   | 15 (22.4%)  |         |
| **WHO Class**     |             |             |             |             |         |
| Tubular           | 2 (28.6%)   | 17 (89.5%)  | 13 (92.9%)  | 37 (55.2%)  |         |
| Mucinous          | 0 (0.0%)    | 0 (0.0%)    | 0 (0.0%)    | 3 (4.5%)    |         |
| Poorly cohesive   | 0 (0.0%)    | 0 (0.0%)    | 1 (7.1%)    | 13 (19.4%)  | 0.000   |
| Mixed             | 1 (14.3%)   | 1 (5.3%)    | 0 (0.0%)    | 14 (20.9%)  |         |
| Uncommon          | 4 (57.1%)   | 1 (5.3%)    | 0 (0.0%)    | 0 (0.0%)    |         |
| **pT stages**     |             |             |             |             |         |
| T1a/ T1b          | 2 (28.6%)   | 2 (10.5%)   | 2 (14.3%)   | 18 (26.9%)  |         |
| T2                | 1 (14.3%)   | 3 (15.8%)   | 1 (7.1%)    | 6 (9.0%)    | 0.765   |
|       | 1 | 2 | 3 | 4 | 5 | 6 |
|-------|---|---|---|---|---|---|
| **T3** | 14.3 | 9 | 47.4 | 6 | 42.9 | 22 |
| **T4a/ T4b** | 42.9 | 5 | 26.3 | 5 | 35.7 | 21 |
| **p N stages** |   |   |   |   |   |   |
| **N0** | 42.9 | 10 | 52.6 | 3 | 21.4 | 24 |
| **N1** | 14.3 | 3 | 15.8 | 2 | 14.3 | 10 |
| **N2** | 14.3 | 4 | 21.1 | 5 | 35.7 | 10 |
| **N3** | 28.6 | 2 | 10.5 | 4 | 28.6 | 23 |
| **M stages** |   |   |   |   |   |   |
| **M0** | 85.7 | 19 | 100.0 | 13 | 92.9 | 64 |
| **M1** | 14.3 | 0 | 0.0% | 1 | 7.1% | 3 |
| **AJCC Stages** |   |   |   |   |   |   |
| **Stages IA/IB** | 42.9 | 4 | 21.1 | 2 | 14.3 | 20 |
| **Stages IIA/IIB** | 14.3 | 9 | 47.4 | 3 | 21.4 | 13 |
| **Stages IIIA/IIIB/IIIC** | 28.6 | 6 | 31.6 | 8 | 57.1 | 31 |
| **Stage IV** | 14.3 | 0 | 0.0% | 1 | 7.1% | 3 |
| **Anatomical regions** |   |   |   |   |   |   |
| **Antrum** | 14.3 | 15 | 78.9 | 7 | 50.0 | 31 |
| **Antrum_Body** | 14.3 | 0 | 0.0% | 1 | 7.1% | 3 |
| **Fundus_Body** | 57.1 | 4 | 21.1 | 2 | 14.3 | 26 |
| **GEJ_Cardia** | 14.3 | 0 | 0.0% | 3 | 21.4 | 4 |
| **Diffuse** | 0 | 0.0% | 0 | 0.0% | 0 | 0.0% |
| **Pylorus** | 0 | 0.0% | 0 | 0.0% | 1 | 7.1% |
| **Lymphatic invasion** |   |   |   |   |   |   |
| **Positive** | 57.1 | 17 | 89.5 | 11 | 78.6 | 44 |
| **Negative** | 42.9 | 2 | 10.5 | 3 | 21.4 | 22 |
|                | Missing | 0.0% | 0.0% | 0.0% | 0.0% | 1.0% | 1.5% |
|----------------|---------|------|------|------|------|------|------|
| **Venous invasion** |         |      |      |      |      |      |      |
| Positive       | 2       | 28.6%| 1    | 5.3% | 2    | 14.3%| 1    | 14.9%|
| Negative       | 5       | 71.4%| 18   | 94.7%| 12   | 85.7%| 56   | 83.6%| 0.477|
| Missing        | 0       | 0.0% | 0    | 0.0% | 0    | 0.0% | 1    | 1.5% |
| **Perineural invasion** |         |      |      |      |      |      |      |
| Positive       | 4       | 57.1%| 6    | 31.6%| 6    | 42.9%| 36   | 53.7%|
| Negative       | 3       | 42.9%| 13   | 68.4%| 8    | 57.1%| 30   | 44.8%| 0.319|
| Missing        | 0       | 0.0% | 0    | 0.0% | 0    | 0.0% | 1    | 1.5% |
| **H. pylori infection** |         |      |      |      |      |      |      |
| Negative       | 3       | 42.9%| 8    | 42.1%| 5    | 35.7%| 25   | 37.3%| 0.970|
| Positive       | 4       | 57.1%| 11   | 57.9%| 9    | 64.3%| 42   | 62.7%|      |

Abbreviations: EBV, Epstein-Barr virus; GEJ, gastroesophageal junction; pT stage, pathological assessment of the primary tumor (pT); pN stage, pathological assessment of the regional lymph nodes (pN); MSI, microsatellite instability; Mixed, mixed type with tubular and poorly cohesive; Uncommon, uncommon histologic variants.
Figure 3. Summary of somatic mutations in 107 gastric cancer samples according to molecular subtype.
Prognosis analysis in 107 gastric cancer patients

Among the 107 gastric cancer patients, the date of last follow-up (months), loco-regional recurrence, distant metastasis, and cause of death were obtained from 72 patients. The median follow-up period was 459.5 days, and there were 19 (26.4%) and 12 (16.7%) cases of gastric cancer relapse and gastric cancer-related death, respectively. We conducted a survival analysis but did not observe a substantial difference in overall survival ($p=0.2828$) or relapse-free survival (RFS, $p=0.3329$) among the four GC subtypes (EBV (n=4), MSI (n=7), CIN (n=10), and GS (n=51) (Figures 5(a) & (b)). However, higher AJCC stage was associated with worse prognosis (RFS, $p=0.004$; GCSS, $p=0.009$) than lower AJCC stages I&II (Figures 4 (c) & (d)). In the multivariate analysis, mutant $MTOR$ gene (hazard ratio (HR): 9.9, $p=0.0017$), AJCC stages III&IV (HR: 9.3, $p=0.0395$), mutant $CCND1$ gene (HR: 14.3, $p=0.0265$), and M1 stage (HR: 7.15, $p=0.0306$) were associated with an unfavorable prognosis. $H. pylori$ infection (HR: 0.3, $p=0.0322$) was predicted to be favorable in Korean patients with gastric cancer. Indeed, among 19 cases with gastric cancer relapse or gastric cancer-related death, 11 cases were negative and 8 cases were positive for $H. pylori$ infection. M stage (HR: 5.2, $p=0.0431$) was an independent prognostic indicator of gastric cancer-specific survival in this study.
Figure 4. Kaplan-Meier (a) relapse-free survival (RFS) and (b) gastric cancer-specific survival (GCSS) curves were stratified by molecular subtype of gastric cancer (EBV, MSI, CIN, and GS). Kaplan-Meier (c) RFS and (d) GCSS curves were analyzed by AJCC stage.
IV. DISCUSSION

We analyzed germline mutations with paired normal and tumor GC samples in 107 Korean patients. Two likely pathogenic variants (NM_004360.4: c. 2494 G>A, p.V832M) in the CDH1 gene and 29 non-pathogenic variants in TP53, STK11, ALK, APC, MSH2, and MLH1 were detected. A V832M mutation has been identified in a hereditary diffuse gastric cancer (HDGC) Japanese family and functionally characterized as a pathogenic mutation. It has also been detected in familial lobular breast cancer patients with the wild type BRCA1/2 gene. Recently, a disease-causing mutation of CDH1 was detected in a 44-year-old Korean patient without evidence of GC via genetic screening owing to his family history of GC. The average age of GC patients at diagnosis was 69 years in this study, and our two cases (1.9%) with V832M were diagnosed at age 66 and 75, respectively. Accordingly, genetic screening for germline mutations of CDH1 with a family history of GC could be considered to identify presymptomatic cancer patients for risk-reduction management in high-risk geographical areas.

The location-specific incidence of somatic variants is depicted in Figure 2. According to the results of the lollipop bar chart analysis, the Q1334del/dup (n=23/52) in ARID1A and L15del (n=6/13) in CDH1 were detected at a frequency of 5 ~ 33% of altered alleles in tumor tissue (Table 2 & Supplementary 2). The in-frame indel (Q1334del/dup), which increases the amount of the ARID1A protein in the nucleus and restores its tumor suppressor functions, has also been reported in gastric cancer samples. This SNP were also occasionally reported in COSMIC database (COSMIC v78) and pancreatic cancers. A three-nucleotide deletion c.44_46del TGC (L15del) in exon 1 of CDH1, which is in the signal peptide region of the E-cadherin protein, was identified in Chinese GC patients and not detected in 240 controls and it was also identified endometrial carcinomas. RHOA
belongs to the Rho family, which functions in the regulation of the actin cytoskeleton, and functional evidence indicates that mutant RHOA works in a gain-of-function manner in this gene. An RHOA mutation was observed in 8.4% of GC cases (n=9/107), with mutations in the Arg5, Gly17, Thr37, Tyr42 and Glu64 residues (Table 2 & Supplementary 2). Among these mutations, the Arg5, Gly17, and Tyr42 residues are recurrently detected in GC.

EBV-infected GC comprised 5-10% of all GC cases and the Cancer Genome Atlas project demonstrated that EBV-infected GC is one of four molecular subtypes, and we found that EBV-infected GC grouped as a molecular subtype. As in the EBV-subtype, ARID1A mutations (4 cases, 57.1%) were prevalent, and none of samples with a TP53 were detected. These findings were comparable with TCGA data. Inhibitors of PIK3/Akt/mTOR pathway, JAK2 pathway and PD-1/PD-L1, PD-L2 pathway are considered as potentially applicable targeted therapies in EBV-infected GC. Notably, 80% of EBV-infected GC harbored PIK3CA mutations and amplifications of JAK2, CD274, and PDCD1LG2. However, only 28.6% of EBV-infected GC harbored PIK3CA mutations (n=2/7), and amplifications of JAK2, CD274, and PDCD1LG2 were not detected in this study. In addition, given that genetic alterations of ERBB2 were not detected in the EBV subtype, anti-ERBB2 (HER2) therapy may not be effective for EBV-infected GC. The global molecular portrait of several genetic abnormalities in EBV-infected GC were not consistent with EBV-infected GC our study. Immunomodulatory agents such as lenalidomide, thalidomide, and pomalidomide (LTP) or combination therapies of LTP with alkylating agents and anti-viral agents such as ganciclovir could be worth reviewing for alternative targeted therapy in EBV-infected GC in Korea. Therefore, these genetic abnormalities should be further validated with large scale EBV-infected GC to provide applicable therapeutic options and the basis for clinical trials.
We observed that the MSI subtype was associated with hyper-mutations in genes and was characterized by a more favorable prognosis than the other molecular subtypes. Both TCGA and ACRG classifications also characterized the MSI subtype by the high mutation frequency and best prognosis \(^5,6\). In intestinal-type GC patients, patients with a good prognosis were characterized by a high mutation rate and microsatellite instability. Further, mutations of \(PIK3CA\) (29.4%) and \(KRAS\) (26.5%) were represented in good prognosis subgroup \(^5,6\). In our study, mutations of \(KRAS\) (26.3%) and \(PIK3CA\) (36.8%) were significantly present in the MSI subtype, and we recurrently observed \(KRAS\) G13D (4 cases) and \(PIK3CA\) H1047R mutations (3 cases) (Table 2). \(PIK3CA\) H1047R mutations were also frequently detected in the MSI subtype in a previous study \(^6\). The genetic alteration of \(ZBTB20\) (P619fs*43, n=5) was limited to the MSI group. This SNP (P619fs*43; rs758277701; COSM267785) was also limited to the MSI group in TCGA cohort, and approximately 20% of MSI group harbored P619fs*43 \(^5\). The clinical significance of this variation should be evaluated through further studies.

The proportion of cases in each molecular group of the TCGA cohort and our groups were different, especially in the CIN (13.1% versus 50%) and GS (62.6% versus 20%) molecular subtypes \(^5\). The vast majority of these differences may be related to the ethnic origin of the patients in each cohort, the limitation of technological platforms, and the strategy of classification into each group. Especially, targeted sequencing is not sufficient to detect clusters of CNVs on the whole-exome scale. Therefore, we focused on CNVs corresponding to possible homozygous deletion or high-level gain rather than low confidence events (shallow loss and a low-level gain) which could often be an artifact.

Genetic alterations of the RTK/RAS/MAPK pathway and
PI3K/PTEN/AKT pathway were detected in 35.5% of GC cases (n=38/107) (Figure 5). Receptor tyrosine kinase (RTK) genomic alterations including *ERBB2, EGFR, FGFR, KDR* and *MET* were detected in 21.5% of GC (23 cases). Thirteen samples (12.2% of GC) harbored ERBB2 alterations, 8 contained somatic base substitutions and 5 harbored amplifications, with these events being mutually exclusive. S310F (two cases) and V842I substitutions (two cases) in *ERBB2* were recurrently detected in this study and have been functionally characterized as activating and sensitive to lapatinib in *ERBB2*-negative breast cancers, while the functions of R678Q which was also recurrently detected in this study, related to anti-ERBB2 (HER2)-targeted therapy have not been tested. Four *KDR* substitutions and 5 *EGFR* substitutions were detected, but amplification of *KDR* and *EGFR* were not detected (Figure 5).

Genetic alterations of *PTEN, PIK3CA, KRAS* and *BRAF* were detected in 18.7% of GC cases (n=20/107) (Figure 5). Agents targeting PI3K/AKT pathway are currently in either preclinical or clinical stages. NVP-BKM120 (pan-class I PI3K inhibitor) has showed the increased sensitivity in tumors harboring PIK3CA mutations and the phase I dose-escalation/expansion studies have been performed with solid tumors. Ten cases (9.4%) harbored mutated *PIK3CA*, and *KRAS* G13D co-existed in 4 cases (Figure 5). Effects of co-existence of genetic alterations of *PIK3CA* and *KRAS* on response to therapy are yet to be evaluated. The dual PI3K and STAT3 blockade using NVP-BKM120 and AG490 (STAT3 inhibitor) showed a synergy effect in GC cells harboring mutated *KRAS* by inducing apoptosis.

These biomarkers may facilitate enrollment of GC patients into clinical trials evaluating targeted therapies and provide the basis for developing solid therapeutic approaches in Korean GC patients.
Figure 5. Therapeutic implications of somatic genomic alterations in 107 clinical gastric cancer cases
We did not observe a substantial difference in overall survival ($p=0.2828$) or relapse-free survival (RFS, $p=0.3329$) among our four GC subtypes, which differed from the groups classified according to AJCC stage (RFS, $p=0.004$; GCSS, $p=0.009$). *H. pylori* infection (HR: 0.3, $p=0.0322$) was a favorable factor in Korean patients with gastric cancer. According to a Chinese prospective cohort with 261 gastric cancer patients, *H. pylori* was an independent prognostic factor of cancer-specific survival (HR: 0.485; 95%, confidence interval (CI): 0.265 to 0.889; $p=0.019$) \(^{59}\). The Asian Cancer Research Group (ACRG) subtypes of MSI, MSS/EMT, MSS/TP53 (+), and MSS/TP53 (−) and these molecular subtypes are correlated with clinical prognosis \(^6\). MSI subtype showed the best favorable prognosis in ACRG, TCGA and our data \(^5,6\). However, there was no significant difference between the TP53 (-) group and TP53 (+) group in our study, which were excluded the MSI subtype (data not shown). Therefore, we thought that the predicting prognosis in GC patients might be performed more simply and effectively using both the MSI subtype and AJCC stage \(^{21}\).

V. CONCLUSION

We classified molecular subtypes of gastric cancer according to the TCGA system using a modality with simpler steps than previous studies, and studied associations between the genetic aberrant profiles of cancer-related genes, environmental circumstances (*EBV* and *H. pylori*), and histopathological features in Korean gastric cancer patients. The 43 gene cancer panel consisted of significantly mutated genes from the TCGA and ACRG cohort \(^{5,6,23}\), genes associated with new targeted therapy of GC (*EGFR*, *ERBB2*, *FGFR2*, and *VEGFR2 (KDR)*) and hereditary cancer syndromes (*CDH1*, *MSH2*, *MLH1*, *STK11*, and *TP53*) \(^{13}\). We observed recurrent mutations that could potentially act as driver mutations, and clinically relevant genomic alterations could be used in routine clinical practice to select
therapies and predict prognosis in Korean GCs.

In survival analysis, MSI subtype showed the most favorable prognosis in ACRG, TCGA and our data, however, we did not observe a substantial difference in overall survival \( (p=0.2828) \) or relapse-free survival (RFS, \( p=0.3329 \)) among the four GC subtypes. Therefore, we thought that the predicting prognosis in GC patients might be performed more simply and effectively using both the MSI subtype and AJCC stage. And, \( H.\ pylori \) infection (HR: 0.3, \( p=0.0322 \)) was associated with a favorable factor in Korean patients with gastric cancer.

We classified four molecular subtypes using a modality with simplified steps and provide a critical starting point for the design of more appropriate clinical trials based on a comprehensive analysis of genetic alterations in Korean GC patients.
REFERENCES

1. Fitzmaurice C, Dicker D, Pain A, Hamavid H, Moradi-Lakeh M, MacIntyre MF, et al. The Global Burden of Cancer 2013. JAMA Oncol 2015;1:505-27.

2. Lauren P. THE TWO HISTOLOGICAL MAIN TYPES OF GASTRIC CARCINOMA: DIFFUSE AND SO-CALLED INTESTINAL-TYPE CARCINOMA. AN ATTEMPT AT A HISTOCLINICAL CLASSIFICATION. Acta Pathol Microbiol Scand 1965;64:31-49.

3. Lauwers GY CF, Graham DY. Gastric carcinoma. In: Bowman FT, Carneiro F, Hruban RH, eds. Classification of Tumours of the Digestive System. Lyon:IARC;2010. In press.

4. Corso S, Giordano S. How Can Gastric Cancer Molecular Profiling Guide Future Therapies? Trends Mol Med 2016;22:534-44.

5. Comprehensive molecular characterization of gastric adenocarcinoma. Nature 2014;513:202-9.

6. Wang J, Yu K, Ye XS, Do IG, Liu S, Gong L, et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat Med; doi:10.1038/nm.3850.

7. Fukayama M, Ushiku T. Epstein-Barr virus-associated gastric carcinoma. Pathol Res Pract 2011;207:529-37.

8. Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, et al. Helicobacter pylori infection and the risk of gastric carcinoma. N Engl J Med 1991;325:1127-31.

9. Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, et al. Helicobacter pylori infection and the development of gastric cancer. N Engl J Med 2001;345:784-9.

10. Fukayama M, Hino R, Uozaki H. Epstein-Barr virus and gastric
carcinoma: virus-host interactions leading to carcinoma. Cancer Sci 2008;99:1726-33.

11. Matsunou H, Konishi F, Hori H, Ikeda T, Sasaki K, Hirose Y, et al. Characteristics of Epstein-Barr virus-associated gastric carcinoma with lymphoid stroma in Japan. Cancer 1996;77:1998-2004.

12. Fitzgerald RC, Hardwick R, Huntsman D, Carneiro F, Guilford P, Blair V, et al. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. J Med Genet 2010;47:436-44.

13. McLean MH, El-Omar EM. Genetics of gastric cancer. Nat Rev Gastroenterol Hepatol 2014;11:664-74.

14. Masciari S, Dewanwala A, Stoffel EM, Lauwers GY, Zheng H, Achatz MI, et al. Gastric cancer in individuals with Li-Fraumeni syndrome. Genet Med 2011;13:651-7.

15. van Lier MG, Wagner A, Mathus-Vliegen EM, Kuipers EJ, Steyerberg EW, van Leerdam ME. High cancer risk in Peutz-Jeghers syndrome: a systematic review and surveillance recommendations. Am J Gastroenterol 2010;105:1258-64; author reply 65.

16. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJ, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. Clin Cancer Res 2006;12:3209-15.

17. Watson P, Vasen HF, Mecklin JP, Bernstein I, Aarnio M, Jarvinen HJ, et al. The risk of extra-colonic, extra-endometrial cancer in the Lynch syndrome. Int J Cancer 2008;123:444-9.

18. Hirota WK, Zuckerman MJ, Adler DG, Davila RE, Egan J, Leighton JA, et al. ASGE guideline: the role of endoscopy in the surveillance of premalignant conditions of the upper GI tract. Gastrointest Endosc 2006;63:570-80.

19. Bray F, Ferlay J, Laversanne M, Brewster DH, Gombe Mbalawa C,
Kohler B, et al. Cancer Incidence in Five Continents: Inclusion criteria, highlights from Volume X and the global status of cancer registration. Int J Cancer 2015;137:2060-71.

20. Davis PA, Sano T. The difference in gastric cancer between Japan, USA and Europe: what are the facts? what are the suggestions? Crit Rev Oncol Hematol 2001;40:77-94.

21. Washington K. 7th edition of the AJCC cancer staging manual: stomach. Ann Surg Oncol 2010;17:3077-9.

22. Kim JM, Kim JS, Kim N, Kim YJ, Kim IY, Chee YJ, et al. Gene mutations of 23S rRNA associated with clarithromycin resistance in Helicobacter pylori strains isolated from Korean patients. J Microbiol Biotechnol 2008;18:1584-9.

23. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214-8.

24. Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M, et al. RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res 2012;40:W622-7.

25. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.

26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

27. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-303.

28. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids
29. Koboldt DC, Zhang Q, Larson DE, Shen D, McLellan MD, Lin L, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 2012;22:568-76.

30. Lonigro RJ, Grasso CS, Robinson DR, Jing X, Wu YM, Cao X, et al. Detection of somatic copy number alterations in cancer using targeted exome capture sequencing. Neoplasia 2011;13:1019-25.

31. Ali SM, Sanford EM, Klempner SJ, Rubinson DA, Wang K, Palma NA, et al. Prospective comprehensive genomic profiling of advanced gastric carcinoma cases reveals frequent clinically relevant genomic alterations and new routes for targeted therapies. Oncologist 2015;20:499-507.

32. Buhard O, Cattaneo F, Wong YF, Yim SF, Friedman E, Flejou JF, et al. Multipopulation analysis of polymorphisms in five mononucleotide repeats used to determine the microsatellite instability status of human tumors. J Clin Oncol 2006;24:241-51.

33. Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, et al. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res 1998;58:5248-57.

34. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science 2013;339:1546-58.

35. Yabuta T, Shinmura K, Tani M, Yamaguchi S, Yoshimura K, Katai H, et al. E-cadherin gene variants in gastric cancer families whose probands are diagnosed with diffuse gastric cancer. Int J Cancer 2002;101:434-41.

36. Suriano G, Mulholland D, de Wever O, Ferreira P, Mateus AR,
Bruyneel E, et al. The intracellular E-cadherin germline mutation V832 M lacks the ability to mediate cell-cell adhesion and to suppress invasion. Oncogene 2003;22:5716-9.

37. Schrader KA, Masciari S, Boyd N, Salamanca C, Senz J, Saunders DN, et al. Germline mutations in CDH1 are infrequent in women with early-onset or familial lobular breast cancers. J Med Genet 2011;48:64-8.

38. Choi HJ, Ki CS, Suh SP, Kim JW. Presymptomatic identification of CDH1 germline mutation in a healthy korean individual with family history of gastric cancer. Ann Lab Med 2014;34:386-9.

39. Guan B, Gao M, Wu CH, Wang TL, Shih Ie M. Functional analysis of in-frame indel ARID1A mutations reveals new regulatory mechanisms of its tumor suppressor functions. Neoplasia 2012;14:986-93.

40. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47-52.

41. Chen QH, Deng W, Li XW, Liu XF, Wang JM, Wang LF, et al. Novel CDH1 germline mutations identified in Chinese gastric cancer patients. World J Gastroenterol 2013;19:909-16.

42. Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, Shen H, et al. Integrated genomic characterization of endometrial carcinoma. Nature 2013;497:67-73.

43. Kakiuchi M, Nishizawa T, Ueda H, Gotoh K, Tanaka A, Hayashi A, et al. Recurrent gain-of-function mutations of RHOC in diffuse-type gastric carcinoma. Nat Genet 2014;46:583-7.

44. Wang K, Yuen ST, Xu J, Lee SP, Yan HH, Shi ST, et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. 2014;46:573-82.
45. Jacome AA, Lima EM, Kazzi AI, Chaves GF, Mendonca DC, Maciel MM, et al. Epstein-Barr virus-positive gastric cancer: a distinct molecular subtype of the disease? Rev Soc Bras Med Trop 2016;49:150-7.

46. Jagannath S. Introduction: addressing challenges in multiple myeloma management in an era of new therapeutics. J Natl Compr Canc Netw 2010;8 Suppl 1:S1-3.

47. Jones RJ, Iempridee T, Wang X, Lee HC, Mertz JE, Kenney SC, et al. Lenalidomide, Thalidomide, and Pomalidomide Reactivate the Epstein-Barr Virus Lytic Cycle through Phosphoinositide 3-Kinase Signaling and Ikaros Expression. Clin Cancer Res 2016; doi:10.1158/1078-0432.ccr-15-2242.

48. Ji Jung E, Mie Lee Y, Lan Lee B, Soo Chang M, Ho Kim W. Ganciclovir augments the lytic induction and apoptosis induced by chemotherapeutic agents in an Epstein-Barr virus-infected gastric carcinoma cell line. Anticancer Drugs 2007;18:79-85.

49. Jung EJ, Lee YM, Lee BL, Chang MS, Kim WH. Lytic induction and apoptosis of Epstein-Barr virus-associated gastric cancer cell line with epigenetic modifiers and ganciclovir. Cancer Lett 2007;247:77-83.

50. Bria E, Pilotto S, Simbolo M, Fassan M, de Manzoni G, Carbognin L, et al. Comprehensive molecular portrait using next generation sequencing of resected intestinal-type gastric cancer patients dichotomized according to prognosis. Sci Rep 2016;6:22982.

51. Bose R, Kavuri SM, Searleman AC, Shen W, Shen D, Koboldt DC, et al. Activating HER2 mutations in HER2 gene amplification negative breast cancer. Cancer Discov 2013;3:224-37.

52. Singh SS, Yap WN, Arfuso F, Kar S, Wang C, Cai W, et al. Targeting the PI3K/Akt signaling pathway in gastric carcinoma: A reality for personalized medicine? World J Gastroenterol 2015;21:12261-73.
53. Maira SM, Pecchi S, Huang A, Burger M, Knapp M, Sterker D, et al. Identification and characterization of NVP-BKM120, an orally available pan-class I PI3-kinase inhibitor. Mol Cancer Ther 2012;11:317-28.

54. Wu YL, Zhang LI, Trandafir L, Dong T, Duval V, Hazell K, et al. Phase I Study of the Pan-PI3K Inhibitor Buparlisib in Adult Chinese Patients with Advanced Solid Tumors. Anticancer Res 2016;36:6185-94.

55. Ando Y, Inada-Inoue M, Mitsuma A, Yoshino T, Ohtsu A, Suenaga N, et al. Phase I dose-escalation study of buparlisib (BKM120), an oral pan-class I PI3K inhibitor, in Japanese patients with advanced solid tumors. Cancer Sci 2014;105:347-53.

56. Park E, Park J, Han SW, Im SA, Kim TY, Oh DY, et al. NVP-BKM120, a novel PI3K inhibitor, shows synergism with a STAT3 inhibitor in human gastric cancer cells harboring KRAS mutations. Int J Oncol 2012;40:1259-66.

57. Liu YJ, Shen D, Yin X, Gavine P, Zhang T, Su X, et al. HER2, MET and FGFR2 oncogenic driver alterations define distinct molecular segments for targeted therapies in gastric carcinoma. Br J Cancer 2014;110:1169-78.

58. Xie L, Su X, Zhang L, Yin X, Tang L, Zhang X, et al. FGFR2 gene amplification in gastric cancer predicts sensitivity to the selective FGFR inhibitor AZD4547. Clin Cancer Res 2013;19:2572-83.

59. Wang F, Sun GP, Zou YF, Zhong F, Ma T, Li XQ, et al. Helicobacter pylori infection predicts favorable outcome in patients with gastric cancer. Curr Oncol 2013;20:e388-95.
APPENDICES

Supplementary Table 1. List of 43 gastric cancer related target gene panel.

| ACVR1B | ALK | APC | ARID1A | BCOR | BRAF | C13orf33 (MEDAG) |
|--------|-----|-----|--------|------|------|------------------|
| C16orf74 | CBWD1 | CCND1 | CCNE1 | CDH1 | CIC | CR1 |
| CTNNB1 | EGFR | ERBB2 | ERBB3 | FBXW7 | FGFR2 | HLA-B |
| IRF2 | JAK2 | KRAS | LARP4B | MDM2 | MET | MLH1 |
| MSH2 | MTOR | MVK | MYC | PD-L1 (CD274) | PD-L2 (PDCD1LG2) | PGM5 |
| PIK3CA | PTEN | RHOA | SMAD4 | STK11 | TP53 | VEGFR-2 (KDR) |
| ZBTB20 |      |     |        |      |      |      |
## Supplementary Table 2: Overview of all somatic SNVs and indels identified through NGS in GC

| Group | Sample | chr | pos   | Ref | Alt | Coverage | Alter_Alt_freq (%) | gene_name | Effect            | hgvs_transcript                      | hgvs_protein                        |
|-------|--------|-----|-------|-----|-----|----------|---------------------|-----------|--------------------|-------------------------------------|-------------------------------------|
| GS    | SKW2   | 17  | 7577124 | C   | T   | 685      | 9%                  | TP53      | missense_variant    | ENST00000269305.4:c.814G>T         | ENSP00000269305.4: p.Val272Leu       |
| GS    | SKW5   | 9   | 121567  | T   | A   | 359      | 20%                 | CBW_D1    | missense_variant    | ENST00000314367.10:c.980A>T         | ENSP00000323433.10: p.Asn327Ile      |
| GS    | SKW5   | 19  | 42794626 | G   | A   | 77       | 5%                  | CIC       | missense_variant    | ENST00000160740.3:c.1706G>A         | ENSP00000160740.3: p.Gly569Asp       |
| GS    | SKW5   | 1   | 20778775 | C   | T   | 354      | 13%                 | CR1       | stop_gained         | ENST00000367049.4:c.6580C>T         | ENSP00000356016.4: p.Arg2194Ter      |
| GS    | SKW6   | 17  | 7578406  | C   | T   | 294      | 44%                 | TP53      | missense_variant    | ENST00000269305.4:c.524G>A          | ENSP00000269305.4: p.Arg175His       |
| GS    | SKW13  | 1   | 27101441 | C   | T   | 80       | 5%                  | ARID1A    | missense_variant    | ENST00000324856.7:c.4723C>T         | ENSP00000320485.7: p.Pro1575Ser      |
| GS    | SKW13  | 1   | 27105930 | TG  | T   | 179      | 5%                  | ARID1A    | frameshift_variant  | ENST00000324856.7:c.5548delG        | ENSP00000320485.7: p.Asp1850Thr553Ter33 |
| GS    | SKW13  | 9   | 122002  | C   | A   | 85       | 6%                  | CBW_D1    | missense_variant    | ENST00000314367.10:c.932G>T         | ENSP00000323433.10: p.Trp311Leu      |
| GS    | SKW13  | 19  | 42792038 | C   | A   | 96       | 6%                  | CIC       | missense_variant    | ENST00000160740.3:c.842C>A          | ENSP00000160740.3: p.Ala281Asp       |
| GS    | SKW13  | 4   | 18534068 | G   | T   | 130      | 5%                  | IRF2      | missense_variant    | ENST00000393593.3:c.130C>A          | ENSP00000377218.3: p.His44Asn        |
| GS    | SKW13  | 4   | 18534068 | C   | A   | 126      | 6%                  | IRF2      | missense_variant    | ENST00000393593.3:c.128G>T          | ENSP00000377218.3: p.Arg43Ile        |
| GS    | SKW13  | 2   | 47635593 | G   | T   | 149      | 7%                  | MSH2      | missense_variant    | ENST00000233146.2:c.265G>T          | ENSP00000233146.2: p.Val205His       |
| CI    | SKW15  | 17  | 7578236  | A   | G   | 973      | 41%                 | TP53      | missense_variant    | ENST00000269305.4:c.613T>C          | ENSP00000269305.4: p.Tyr205His       |
| Sample | Pool | Genotype | Chromosome | Start Position | End Position | Allele | Variant Type | Transcript | Protein | Description |
|--------|------|-----------|-------------|----------------|--------------|--------|--------------|------------|---------|-------------|
| GS SKW16 1 20778775 C T 854 7% CR1 stop_gained | ENST0000367049.4:c.6580C>T | CR1 stop_gained | p.Arg2194Ter |
| EB SKW12 1 27100181 CG C 520 5% ARID1A In_Frame_variant | ENSP0000324856.7:c.3999_4001delGCA | ARID1A In_Frame_variant | p.Gln1334del |
| EB SKW12 16 68771347 CG C 61 8% CDH1 In_Frame_variant | ENSP0000261769.5:c.44_46del1TG | CDH1 In_Frame_variant | p.Leu15del |
| EB SKW12 8 12875063 C T 618 9% MYC missense_variant | ENSP00000259523.6:c.131C>T | MYC missense_variant | p.Ala44Val |
| CI SKW17 1 20778775 C T 259 12% CR1 stop_gained | ENSP0000367049.4:c.6580C>T | CR1 stop_gained | p.Arg2194Ter |
| CI SKW17 10 876899 T A 99 6% LARP4B In_Frame_variant | ENSP0000361445.4:c.5551A>G | LARP4B In_Frame_variant | p.Ala1851Asp |
| CI SKW17 1 11190648 T C 188 8% MTO1 missense_variant | ENSP00000326873.7:c.476A>G | MTO1 missense_variant | p.Glu159Val |
| CI SKW17 19 1220383 A G 73 5% STK11 stop_gained | ENSP00003628935.4:c.916C>T | STK11 stop_gained | p.Arg306Ter |
| MS I SKW18 1 27057936 GC G 750 9% ARID1A frameshift_variant | ENSP000003269305.4:c.916C>T | ARID1A frameshift_variant | p.Phe306Val |
| MS I SKW18 17 7577022 G A 847 11% TP53 stop_gained | ENSP00000324856.7:c.1650delC | TP53 stop_gained | p.Glu283Val |
| MS I SKW18 12 25398281 C T 780 10% KRA S missense_variant | ENSP000003256078.4:c.38G>A | KRA S missense_variant | p.Glu126Asp |
| MS I SKW18 13 31495945 GC G 373 6% MEDAG frameshift_variant | ENSP00000380482.4:c.750delC | MEDAG frameshift_variant | p.Glu250Val |
| MS I SKW18 3 37061839 AC A 477 11% MLH1 missense_variant | ENSP00000231790.2:c.927delC | MLH1 missense_variant | p.Glu309Val |
| MS I SKW18 3 17895208 A G 615 9% PIK3CA frameshift_variant | ENSP00000263967.3:c.3140A>G | PIK3CA frameshift_variant | p.Glu1047Arg |
| MS I SKW20 1 27087467 T G 219 8% ARID1A missense_variant | ENSP00000324856.7:c.2041T>G | ARID1A missense_variant | p.Phe681Val |
| MS I SKW20 1 20778775 C T 289 15% CR1 stop_gained | ENSP00000367049.4:c.6580C>T | CR1 stop_gained | p.Arg2194Ter |
| MS I SKW20 1 20778775 C T 289 15% CR1 stop_gained | ENSP00000356016.4:c.6580C>T | CR1 stop_gained | p.Arg2194Ter |
| MS I  | SKW20  | 6    | 31323094 | C  | T  | 145 | 8% | HLA-B missense_variant ENST00000412585.2:c.895G>A ENST00000361445.4:c.5497G>A |
|-------|--------|------|----------|----|----|-----|----|-------------------------------|
| MS I  | SKW20  | 1    | 11190702 | T  | C  | 207 | 5% | MTO missense_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| CLIN  | SKW25  | 4    | 153332725 | C  | A  | 62  | 6% | FBXW7 missense_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| CLIN  | SKW25  | 4    | 153332588 | TC | A  | 64  | 6% | FBXW7 frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| MS I  | SKW21  | X    | 39913252 | TG | T  | 79  | 5% | BCO missense_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| MS I  | SKW21  | 6    | 31324602 | T  | A  | 64  | 8% | HLA-B stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW23  | 1    | 27106732 | C  | T  | 729 | 19%| ARID1A missense_variant  |
| GS    | SKW23  | 17   | 7577547  | C  | A  | 295 | 12%| TP53 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| CLIN  | SKW1   | 17   | 757655   | G  | A  | 678 | 42%| TP53 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| CLIN  | SKW1   | 11   | 69465987 | AG | A  | 201 | 5% | CCN1 frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 2    | 29451783 | AC | A  | 64  | 6% | ALK frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 17   | 7578527  | A  | G  | 157 | 29%| TP53 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 1    | 207787753 | C  | T  | 233 | 16%| CR1 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 3    | 49412913 | G  | A  | 177 | 23%| RHOA stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27057658 | C  | T  | 586 | 14%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27105897 | G  | T  | 879 | 18%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |

| GS    | SKW38  | 2    | 29451783 | AC | A  | 64  | 6% | ALK frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 1    | 207787753 | C  | T  | 233 | 16%| CR1 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 3    | 49412913 | G  | A  | 177 | 23%| RHOA stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27057658 | C  | T  | 586 | 14%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27105897 | G  | T  | 879 | 18%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |

| GS    | SKW38  | 2    | 29451783 | AC | A  | 64  | 6% | ALK frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 1    | 207787753 | C  | T  | 233 | 16%| CR1 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 3    | 49412913 | G  | A  | 177 | 23%| RHOA stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27057658 | C  | T  | 586 | 14%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27105897 | G  | T  | 879 | 18%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |

| GS    | SKW38  | 2    | 29451783 | AC | A  | 64  | 6% | ALK frameshift_variant ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 1    | 207787753 | C  | T  | 233 | 16%| CR1 stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW38  | 3    | 49412913 | G  | A  | 177 | 23%| RHOA stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27057658 | C  | T  | 586 | 14%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| GS    | SKW37  | 1    | 27105897 | G  | T  | 879 | 18%| ARID1A stop_gained ENST00000361445.4:c.5497G>A ENST00000361445.4:c.5497G>A |
| MS   | SKW26  | 30142909 | G | A | 352 | 10% | ALK | missense_variant | ENST0000389048.3:c.617C>T | ENSP0000373700.3: p.Ala206Val |
|------|--------|----------|---|---|-----|-----|-----|------------------|---------------------------|-------------------------------|
| SKW26  | 1   | 27023140 | TG | GC | T | 55 | 7% | ARID1A | ln_Frame_variant | ENST0000324856.7:c.258_260delCGG | ENSP0000320485.7: p.Gly87del |
| SKW26  | 1   | 27101116 | AC | A | 977 | 14% | ARID1A | frameshift_variant | ENST0000324856.7:c.4403delC | ENSP0000320485.7: p.Prol648fsTer13 |
| SKW26  | X  | 39933399 | GC | G | 648 | 32% | BCO2R | frameshift_variant | ENST0000342274.4:c.1199delG | ENSP0000345923.4: p.Gly400AlafsTer42 |
| SKW26  | 7   | 55231495 | G | A | 745 | 19% | EGRF | missense_variant | ENST0000275493.2:c.1701G>A | ENSP0000275493.2: p.Met567Ile |
| SKW26  | 12  | 56486562 | C | A | 1305 | 14% | ERBB3 | missense_variant | ENST0000267101.3:c.1141C>T | ENSP0000267101.3: p.Pro381Thr |
| SKW26  | 12  | 56489493 | T | A | 2582 | 17% | ERBB3 | missense_variant | ENST0000267101.3:c.1958T>A | ENSP0000267101.3: p.Val653Glu |
| SKW26  | 10  | 123263394 | C | T | 1176 | 16% | FGFR2 | frameshift_variant | ENST0000336553.6:c.1076G>A | ENSP0000337665.6: p.Arg359His |
| SKW26  | 12  | 25398281 | C | T | 2457 | 6% | MSH2 | frameshift_variant | ENST0000256078.4:c.1462G>A | ENSP0000256078.4: p.Gly13Asp |
| SKW26  | 2   | 47639587 | GA | G | 1023 | 14% | MVK | frameshift_variant | ENST0000233146.2:c.687delA | ENSP0000233146.2: p.Ala230LeufsTer16 |
| SKW26  | 12  | 11003431  | G | A | 239 | 15% | STK11 | frameshift_variant | ENST0000228510.3:c.1120G>A | ENSP0000228510.3: p.Ala374Thr |
| SKW26  | 19  | 1221313 | GC | G | 468 | 15% | ZBTB20 | frameshift_variant | ENST0000357258.3:c.1856delC | ENSP0000357258.3: p.Arg619LeufsTer43 |
| SKW26  | 3   | 114058002  | AG | A | 1296 | 16% | ARID1A | frameshift_variant | ENST0000324856.7:c.4689delC | ENSP0000320485.7: p.Met1564Ter |
| SKW27  | 1   | 27101401 | GC | G | 183 | 15% | ARID1A | frameshift_variant | ENST0000324856.7:c.4824delC | ENSP0000324856.7: p.Arg715Ter |
| SKW27  | 16  | 68772281 | C | T | 54 | 7% | CDH1 | frameshift_variant | ENST0000261769.5:c.130C>T | ENSP0000261769.5: p.Arg44Cys |
| SKW27  | 9   | 70972209 | G | A | 71 | 6% | PGM5 | frameshift_variant | ENST0000396392.1:c.166G>A | ENSP0000337665.6: p.Arg359His |
| Gene   | Missense Variants | Ensembl | Description |
|--------|-------------------|----------|-------------|
| CIC    | missense_variant  | ENST0000167040.3 | c.193C>T |
| ERBB2  | missense_variant  | ENST0000167041.5 | c.929C>T |
| KRAF1  | missense_variant  | ENST0000167042.4 | c.1636C>A |
| PIK3CA | missense_variant  | ENST0000167043.4 | c.385A>T |
| TP53   | missense_variant  | ENST0000342988.3 | c.358A>T |
| TP53   | missense_variant  | ENST0000167045.4 | c.742C>T |
| PRKCA  | missense_variant  | ENST0000167046.4 | c.585G>A |
| CR1    | missense_variant  | ENST0000167047.4 | c.44_46delTG |
| EGFR   | missense_variant  | ENST0000167048.4 | c.2305G>A |
| LAR    | frameshift_variant| ENST0000342274.4 | c.1534_1536delCA |
| ARID1A  | frameshift_variant| ENST0000342275.7 | c.839_841delAGG |
| CCND1  | frameshift_variant| ENST0000342276.2 | c.2305G>A |
| EGFR   | frameshift_variant| ENST0000342277.3 | c.487delA |
| MSH2   | frameshift_variant| ENST0000342278.2 | c.746delA |
| ARID1A  | frameshift_variant| ENST0000342279.2 | c.3281del |
| ARID1A  | frameshift_variant| ENST0000342280.2 | c.3977dupC |

**Gene Symbols:**
- CIC
- ERBB2
- KRAF1
- PIK3CA
- TP53
- PRKCA
- CR1
- EGFR
- LAR
- ARID1A
- CCND1
- EGFR
- MSH2
- ARID1A
- ARID1A

**Ensembl IDs:**
- ENST0000167040.3
- ENST0000167041.5
- ENST0000167042.4
- ENST0000167043.4
- ENST0000342988.3
- ENST0000167045.4
- ENST0000167046.4
- ENST0000167047.4
- ENST0000167048.4
- ENST0000342274.4
- ENST0000342275.7
- ENST0000342276.2
- ENST0000342277.3
- ENST0000342278.2
- ENST0000342279.2
- ENST0000342280.2

**Missense Variants:**
- c.193C>T
- c.929C>T
- c.1636C>A
- c.385A>T
- c.358A>T
- c.44_46delTG
- c.2305G>A
- c.1534_1536delCA
- c.839_841delAGG
- c.2305G>A
- c.487delA
- c.746delA
- c.3281del
- c.3977dupC
| MS I | SKW41 | X | 39914723 | G | A | 2516 | 34% | BCO R | stop_gained | ENST00000342274.4:c.4537C>T | T
| MS I | SKW41 | 19 | 42794440 | GC | G | 821 | 27% | CIC | frameshift_variant | ENST00000160740.3:c.1526delC | C
| MS I | SKW41 | 19 | 42797375 | GC | G | 648 | 36% | CIC | frameshift_variant | ENST00000160740.3:c.3737delC | C
| MS I | SKW41 | 1 | 20778775 | C | T | 4083 | 17% | CR1 | stop_gained | ENST00000367049.4:c.6580C>T | T
| MS I | SKW41 | 1 | 20778783 | G | T | 3853 | 7% | CR1 | stop_gained | ENST00000367049.4:c.6580C>T | T
| MS I | SKW41 | 17 | 37884008 | C | T | 551 | 34% | ERB2 | missense_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 4 | 15324938 | C | T | 1769 | 28% | FBXW7 | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 6 | 31324207 | AG | A | 193 | 6% | HLA-B | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 6 | 31324508 | C | CT | 600 | 5% | HLA-B | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 12 | 25398281 | C | T | 3172 | 18% | KRA S | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 12 | 69230470 | T | C | 3254 | 47% | MD2 | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 9 | 70993121 | C | T | 3716 | 6% | PGM5 | stop_gained | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 9 | 70993145 | A | G | 3607 | 12% | PGM5 | missense_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 3 | 17895208 | A | G | 1997 | 21% | PIK3CA | missense_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW41 | 3 | 11405800 | AG | A | 1623 | 6% | ZBTB20 | frameshift_variant | ENST00000269571.5:c.3557C>T | T
| MS I | SKW31 | 2 | 30143407 | G | A | 92 | 5% | ALK | frameshift_variant | ENST00000342274.4:c.4537C>T | T

62
| MS    | SKW31  | 27105930 | TG  | T   | 1752 | 20% | ARID | frameshift_variant | ENST00000324856.7:c.5548delG | ENSP00000320485.7: p.Asp1850ThrfsTer33 |
|-------|--------|----------|-----|-----|------|-----|------|-------------------|-------------------------------|-------------------------------------|
| MS    | SKW31  | 42791228 | GC  | G   | 270  | 6%  | CIC  | frameshift_variant | ENST00000160740.3:c.293delC | ENSP00000160740.3: p.Pro98LeufsTer107 |
| MS    | SKW31  | 37879658 | G   | A   | 1353 | 15% | ERB  | missense_variant  | ENST00000269571.5:c.2033G>A | ENSP00000269571.4: p.Arg678Gln  |
| MS    | SKW31  | 56493477 | C   | T   | 1215 | 9%  | ERB  | missense_variant  | ENST00000267101.3:c.2885C>T | ENSP00000267101.3: p.Ala962Val  |
| MS    | SKW31  | 89717769 | T   | TA  | 1885 | 26% | CIC  | frameshift_variant | ENST00000324856.7:c.5548delG | ENSP00000320485.7: p.Asp1850ThrfsTer33 |
| MS    | SKW33  | 7578535  | T   | C   | 720  | 11% | TP53 | missense_variant  | ENST00000269305.4:c.395A>G | ENSP00000269305.4: p.Lys132Arg  |
| MS    | SKW33  | 69465987 | AG  | G   | 111  | 5%  | CCN  | In_Frame_variant  | ENST00000227507.2:c.839_841delAGG | ENSP00000227507.2: p.Glu280del  |
| MS    | SKW33  | 56494920 | G   | 358 | 6%  | ERB  | In_Frame_variant  | ENST00000267101.3:c.3286_3294delTCATCAGAG | ENSP00000267101.3: p.Ser1096_Glu1098del |
| MS    | SKW45  | 27105930 | TG  | T   | 625  | 8%  | ARID | frameshift_variant | ENST00000324856.7:c.5548delG | ENSP00000320485.7: p.Asp1850ThrfsTer33 |
| MS    | SKW45  | 7577548  | C   | T   | 282  | 13% | TP53 | missense_variant  | ENST00000269305.4:c.733G>A | ENSP00000269305.4: p.Gly245Ser  |
| MS    | SKW45  | 14048292 | AG  | A   | 770  | 8%  | BRAF | frameshift_variant | ENST00000288602.6:c.1208delC | ENSP00000288602.6: p.Pro403LeufsTer18 |
| MS    | SKW45  | 69465987 | AG  | A   | 73   | 8%  | CCN  | missense_variant  | ENST00000227507.2:c.839_841delAGG | ENSP00000227507.2: p.Glu280del  |
| MS    | SKW45  | 68845649 | G   | A   | 667  | 9%  | CDH1 | In_Frame_variant  | ENST00000261769.5:c.895G>A | ENSP00000261769.4: p.Ala299Thr  |
| MS    | SKW45  | 68771347 | CG  | CT  | 76   | 8%  | CDH1 | missense_variant  | ENST00000261769.5:c.44_46delITGC | ENSP00000261769.4: p.Leu15del  |
| MS    | SKW45  | 42791265 | G   | A   | 168  | 11% | CIC  | missense_variant  | ENST00000160740.3:c.325G>A | ENSP00000160740.3: |
| Gene   | Sample | Chromosome | Position | Allele 1 | Allele 2 | Length | Type          | Description                                                                 | Reference  |
|--------|--------|------------|----------|----------|----------|--------|---------------|---------------------------------------------------------------------------|------------|
| CIC    | SKW45  | 19         | 42799097 | G        | C        | 205    | frameshift_v  | variant                     | ENST0000160740.3:c.4580delC  | p.Gly109Arg |
| CTNB1  | SKW45  | 3          | 41278163 | G        | C        | 746    | frameshift_v  | variant                     | ENST0000349496.5:c.2046_2047delCT |             |
| EGF    | SKW45  | 7          | 55229242 | G        | T        | 383    | missense_var  | iant                        | ENST0000275493.2:c.1549G>T   |             |
| ERBB2  | SKW45  | 17         | 37881332 | G        | A        | 286    | missense_var  | iant                        | ENST0000269571.5:c.2524G>A   |             |
| ERBB3  | SKW45  | 12         | 56478854 | G        | A        | 638    | missense_var  | iant                        | ENST0000269710.3:c.310G>A    |             |
| KDR    | SKW45  | 4          | 55976174 | G        | T        | 387    | frameshift_v  | variant                     | ENST0000263923.4:c.1111C>A   |             |
| MSH2   | SKW45  | 2          | 47705561 | T        | TA       | 589    | frameshift_v  | variant                     | ENST0000233146.2:c.2362dupA  |             |
| MSH2   | SKW45  | 2          | 47707887 | TA       | T        | 725    | frameshift_v  | variant                     | ENST0000233146.2:c.2513delA  |             |
| MTRR   | SKW45  | 1          | 11182095 | G        | A        | 360    | missense_var  | iant                        | ENST0000361445.4:c.6751C>T   |             |
| RHOA   | SKW45  | 3          | 49412914 | T        | C        | 617    | missense_var  | iant                        | ENST0000418115.1:c.109A>G    |             |
| CR1    | SKW44  | 1          | 20778775 | G        | C        | 193    | stop_gained   |                            | ENST0000367049.4:c.6580C>T   |             |
| IRF2   | SKW44  | 4          | 18534068 | G        | C        | 147    | missense_var  | iant                        | ENST0000393593.3:c.122C>G    |             |
| APC    | YMC69  | 5          | 11217900 | C        | G        | 238    | stop_gained   |                            | ENST0000257430.4:c.7709C>G   |             |
| PIK3CA | YMC69  | 3          | 17893609 | G        | A        | 201    | missense_var  | iant                        | ENST0000263967.3:c.1633G>A   |             |
| TP53   | YMC69  | 17         | 7577507  | T        | G        | 198    | missense_var  | iant                        | ENST0000269305.4:c.774A>C    |             |

GS
| Gene   | Sample | Chromosome | Position | Allele 1 | Allele 2 | Length | Type          | Description                                                                 | Reference  |
|--------|--------|------------|----------|----------|----------|--------|---------------|---------------------------------------------------------------------------|------------|
| CIC    | SKW45  | 19         | 42799097 | G        | C        | 205    | frameshift_v  | variant                     | ENST0000160740.3:c.4580delC  | p.Gly109Arg |
| CTNB1  | SKW45  | 3          | 41278163 | G        | C        | 746    | frameshift_v  | variant                     | ENST0000349496.5:c.2046_2047delCT |             |
| EGF    | SKW45  | 7          | 55229242 | G        | T        | 383    | missense_var  | iant                        | ENST0000275493.2:c.1549G>T   |             |
| ERBB2  | SKW45  | 17         | 37881332 | G        | A        | 286    | missense_var  | iant                        | ENST0000269571.5:c.2524G>A   |             |
| ERBB3  | SKW45  | 12         | 56478854 | G        | A        | 638    | missense_var  | iant                        | ENST0000269710.3:c.310G>A    |             |
| KDR    | SKW45  | 4          | 55976174 | G        | T        | 387    | frameshift_v  | variant                     | ENST0000263923.4:c.1111C>A   |             |
| MSH2   | SKW45  | 2          | 47705561 | T        | TA       | 589    | frameshift_v  | variant                     | ENST0000233146.2:c.2362dupA  |             |
| MSH2   | SKW45  | 2          | 47707887 | TA       | T        | 725    | frameshift_v  | variant                     | ENST0000233146.2:c.2513delA  |             |
| MTRR   | SKW45  | 1          | 11182095 | G        | A        | 360    | missense_var  | iant                        | ENST0000361445.4:c.6751C>T   |             |
| RHOA   | SKW45  | 3          | 49412914 | T        | C        | 617    | missense_var  | iant                        | ENST0000418115.1:c.109A>G    |             |
| CR1    | SKW44  | 1          | 20778775 | G        | C        | 193    | stop_gained   |                            | ENST0000367049.4:c.6580C>T   |             |
| IRF2   | SKW44  | 4          | 18534068 | G        | C        | 147    | missense_var  | iant                        | ENST0000393593.3:c.122C>G    |             |
| APC    | YMC69  | 5          | 11217900 | C        | G        | 238    | stop_gained   |                            | ENST0000257430.4:c.7709C>G   |             |
| PIK3CA | YMC69  | 3          | 17893609 | G        | A        | 201    | missense_var  | iant                        | ENST0000263967.3:c.1633G>A   |             |
| TP53   | YMC69  | 17         | 7577507  | T        | G        | 198    | missense_var  | iant                        | ENST0000269305.4:c.774A>C    |             |
| Sample | YMC69  | 17  | 7577529 | A   | G   | 217 | 46% | TP53 | missense_variant | ENST00000269305.4:c.752T>C | ENSP00000269305.4: p.Ile251Thr |
|--------|--------|-----|---------|-----|-----|-----|-----|------|-----------------|-----------------------------|--------------------------------|
| EB     | YMC63  | 5   | 11217956| C   | T   | 706 | 17% | APC  | missense_variant | ENST00000257430.4:c.8275C>T | ENSP00000257430.4: p.Arg2759Cys |
| EB     | YMC63  | 9   | 5463099 | T   | A   | 527 | 26% | CD27 | missense_variant | ENST00000381573.4:c.318T>A  | ENSP00000381573.4: p.Thr106Gln |
| EB     | YMC63  | 16  | 68849539| AT  | A   | 665 | 5%  | CDH1 | frameshift_variant | ENST0000261769.5:c.1443delT | ENSP0000261769.5: p.Asn481LysfsTer41 |
| EB     | YMC63  | 3   | 41266113| C   | G   | 498 | 35% | CTN  | missense_variant | ENST0000349496.5:c.110C>G   | ENSP0000349496.5: p.Ser37Cys  |
| GS     | YMC62  | 1   | 27100181| CG  | CA  | 357 | 6%  | ARID1A | missense_variant | ENST0000324856.7:c.3999_4001delGCA | ENSP0000324856.7: p.Gln1334del |
| MSI    | YMC68  | 1   | 27100181| CG  | CA  | 298 | 7%  | ARID1A | In_Frame_variant | ENST0000342274.4:c.3519dupA | ENSP0000342274.4: p.Gln1334del |
| MSI    | YMC68  | X   | 39923086| G   | GT  | 248 | 8%  | BCO2R | frameshift_variant | ENST0000367049.4:c.510delC | ENSP0000367049.4: p.Gln1334del |
| MSI    | YMC68  | 1   | 20769697| AC  | A   | 482 | 7%  | CR1  | missense_variant | ENST00002669571.5:c.2033G>A | ENSP00002669571.5: p.Arg678Gln |
| MSI    | YMC68  | 17  | 37879658| G   | A   | 753 | 5%  | ERB2 | missense_variant | ENST00002669571.5:c.2524G>A | ENSP00002669571.5: p.Arg678Gln |
| MSI    | YMC68  | 17  | 37881332| G   | A   | 344 | 6%  | ERB2 | missense_variant | ENST00002669571.5:c.2033G>A | ENSP00002669571.5: p.Arg678Gln |
| MSI    | YMC68  | 10  | 890938  | GT  | G   | 235 | 6%  | LAR4P4B | missense_variant | ENST0000316157.3:c.487delA | ENSP0000316157.3: p.Gln1334del |
| GS     | YMC61  | 1   | 27100181| CG  | CA  | 466 | 6%  | ARID1A | missense_variant | ENST0000324856.7:c.3999_4001delGCA | ENSP0000324856.7: p.Gln1334del |
| GS     | YMC61  | 17  | 7579432 | AG  | A   | 215 | 7%  | TP53 | missense_variant | ENST0000269305.4:c.254delC | ENSP0000269305.4: p.Gln1334del |
| GS     | YMC59  | 12  | 56481922| G   | A   | 622 | 16% | TP53 | stop_gained      | ENST0000267101.3:c.850G>A  | ENSP0000267101.3: p.Gly284Arg |
| GS     | YMC59  | 17  | 7574003 | G   | A   | 254 | 13% | TP53 | stop_gained      | ENST0000269305.4:c.1024C>T | ENSP0000269305.4: p.Arg342Ter |
| Sample | ID     | Chromosome | Start | End | Gene | Mutation Type | Transcript ID | Protein Effect |
|--------|--------|------------|-------|-----|-------|---------------|---------------|----------------|
| GS YMC58 | 16 | 68842751 | TA T | 382 | 22% | CDH1 splicing donor site variant | ENST0000261769.5:c.687+4_687+7delAGTA | |
| GS YMC58 | 3 | 41274899 | G A | 224 | 5% | CTNNB1 stop gained | ENST0000349496.5:c.1149G>A | |
| GS YMC58 | 7 | 55248986 | G A | 127 | 15% | EGF missense variant | ENST0000275493.2:c.2284G>A | |
| GS YMC58 | 3 | 49413010 | G A | 306 | 7% | RHO missense variant | ENST0000418115.1:c.13C>T | |
| GS YMC58 | 17 | 7578212 | G A | 684 | 7% | TP53 stop gained | ENST0000269305.4:c.637C>T | |
| GS YMC58 | 1 | 27092833 | G T | 837 | 9% | ARID1A stop gained | ENST0000324856.7:c.2854G>A | |
| GS YMC58 | 11 | 69465987 | AGAG | A | 76 | 5% | CCND1 In_Frame variant | ENST0000227507.2:c.839_841delAGG | |
| GS YMC58 | 17 | 7577508 | T C | 177 | 6% | TP53 stop gained | ENST0000269305.4:c.773A>G | |
| GS YMC58 | 5 | 11217602 | G T | 411 | 47% | APC stop gained | ENST0000257430.4:c.4729G>T | |
| GS YMC58 | 17 | 7578413 | C T | 255 | 48% | TP53 missense variant | ENST0000257430.4:c.517G>T | |
| GS YMC58 | 1 | 27023140 | TG T | 53 | 6% | ARID1A In_Frame variant | ENST0000324856.7:c.258_260delCGG | |
| GS YMC58 | 17 | 7578550 | G T | 498 | 54% | TP53 missense variant | ENST0000269305.4:c.380C>T | |
| GS YMC58 | 17 | 7578203 | C A | 969 | 13% | TP53 missense variant | ENST0000269305.4:c.646G>T | |
| MS YMC51 | 1 | 27100181 | CG C | 406 | 5% | ARID1A In_Frame variant | ENST0000324856.7:c.3899_40delGCA | |

66
| MS   | I    | CA       | T/A/G | C/T/G | 1A         | riant      | 01del GCA              | p.Gln1334del               |
|------|------|----------|-------|-------|------------|------------|------------------------|----------------------------|
| YMC51| 1    | 27105930 | TG     | T     | 722        | 32%        | ARID1A frameshift variant | ENST0000324856.7:c.5548del G |
| YMC51| 3    | 41277260 | C      | A     | 355        | 20%        | ARID1A missense variant   | ENST0000349496.5:c.1729C>A  |
| YMC51| 4    | 15327363 | A      | AT    | 582        | 33%        | CTN1B frameshift variant   | ENST0000263981.5:c.247dup A  |
| YMC51| 4    | 55955595 | C      | G     | 524        | 27%        | FBXW7 frameshift variant    | ENST0000263923.4:c.3350G>C   |
| YMC51| 10   | 909737   | C      | T     | 667        | 35%        | KDR frameshift variant      | ENST0000316157.3:c.376G>A    |
| YMC51| 3    | 37070348 | AC     | C     | 819        | 32%        | LARP4B frameshift variant   | ENST0000231790.2:c.1489del C  |
| YMC51| 17   | 7577093  | C      | T     | 655        | 23%        | MLH frameshift variant      | ENST0000269305.4:c.845G>A    |
| YMC51| 17   | 7572962  | GT     | G     | 338        | 19%        | TP53 frameshift variant     | ENST0000269305.4:c.1146del C  |
| YMC51| 3    | 11405800 | AG     | A     | 315        | 30%        | ZBTB20 frameshift variant   | ENST0000357258.3:c.1856del C  |
| YMC50| 1    | 27023715 | TG     | T     | 172        | 13%        | ARID1A frameshift variant   | ENST0000324856.7:c.827delG    |
| YMC50| 1    | 27097621 | CA     | C     | 737        | 22%        | ARID1A frameshift variant   | ENST0000324856.7:c.3216del C  |
| YMC50| 1    | 27105930 | T      | TG    | 661        | 23%        | ARID1A frameshift variant   | ENST0000324856.7:c.5548del G  |
| YMC50| X    | 39923086 | G      | GT    | 673        | 20%        | ARID1A frameshift variant   | ENST0000324856.7:c.3216del C  |
| YMC50| 1    | 20778271 | T      | C     | 592        | 21%        | BCR frameshift variant      | ENST0000342274.4:c.3519dup A  |
| YMC50| 7    | 55240761 | C      | T     | 186        | 19%        | CR1 frameshift variant      | ENST0000367049.4:c.597T>C    |
| YMC50| 3    | 11405800 | AG     | A     | 315        | 30%        | EGF frameshift variant      | ENST0000275493.2:c.2005C>T   |

67
| Protein | Accession | Sample | Gene | Chromosome | Coding Region | Description |
|---------|------------|--------|------|-------------|---------------|-------------|
| ERBB3   | ENST00000267101.3 | YMC50 | ERBB3 | c.310G>A | missense variant | p.Val104Met |
| ERBB3   | ENST00000267101.3 | YMC50 | ERBB3 | c.2047C>T | missense variant | p.Arg683Trp |
| FBXW7   | ENST00000281708.4 | YMC50 | FBXW7 | c.2047C>T | missense variant | p.Arg683Trp |
| MVK     | ENST00000228510.3 | YMC50 | MVK | c.2047C>T | missense variant | p.Arg683Trp |
| PGM5    | ENST00000396396.1 | YMC50 | PGM5 | c.2047C>T | missense variant | p.Arg683Trp |
| PIK3CA  | ENST00000263967.3 | YMC50 | PIK3CA | c.2047C>T | missense variant | p.Arg683Trp |
| PIK3CA  | ENST00000263967.3 | YMC50 | PIK3CA | c.2047C>T | missense variant | p.Arg683Trp |

| Protein | Accession | Sample | Gene | Chromosome | Coding Region | Description |
|---------|------------|--------|------|-------------|---------------|-------------|
| APC     | ENST00000257430.4 | GS YMC67 | APC | c.904C>T | stop gained | p.Arg302Ter |
| APC     | ENST00000257430.4 | GS YMC67 | APC | c.466dupA | frameshift variant | p.Arg302Ter |
| CR1     | ENST0000367049.4 | GS YMC67 | CR1 | c.640A>G | missense variant | p.Asp214Gly |
| ERBB2   | ENST00000269571.5 | GS YMC67 | ERBB2 | c.929C>T | missense variant | p.Ser310Phe |
| ERBB2   | ENST00000269571.5 | GS YMC67 | ERBB2 | c.2033G>A | missense variant | p.Arg678Gln |
| HLA-B2   | ENST00000412585.2 | GS YMC67 | HLA-B2 | c.206A>T | missense variant | p.Glu69Val |
| MTO     | ENST0000361445.4 | GS YMC67 | MTO | c.770G>A | missense variant | p.Glu257Val |
| TP53    | ENST00000324856.7 | GS YMC47 | TP53 | c.3999_401delGCA | missense variant | p.Glu1334del |
| Gene | Project | Chromosome | Genomic Location | Mutation Type          | Ensembl Gene ID | Ensembl Transcript ID | Reference | Remark |
|------|---------|-------------|------------------|------------------------|----------------|----------------------|-----------|--------|
| APC  | YMC45   | 5           | 11217567:5       | frameshift Variant     | ENST0000257430.4 | ENSP0000257430.4   | p.Glu1464ValfsTer8 |
| TP53 | YMC45   | 17          | 757836           | In-Frame Variant       | ENST0000269305.4 | ENSP0000269305.4   | p.Pro177_Arg181del |
| ARID1A| YMC43   | 1           | 27100181         | In-Frame Variant       | ENST0000324856.7 | ENSP0000320485.7   | p.Gln1334del   |
| RHO A| YMC42   | 3           | 49405947         | frameshift Variant     | ENST00000418151.1 | ENSP00000400175.1 | p.Glu64Gly   |
| TP53 | YMC42   | 17          | 7578413          | In-Frame Variant       | ENST0000269305.4 | ENSP0000269305.4   | p.Val173Leu   |
| CDH1 | YMC66   | 16          | 68771347         | In-Frame Variant       | ENST0000261769.5  | ENSP0000261769.4   | p.Leu15del   |
| CDH1 | YMC66   | 17          | 7579437          | In-Frame Variant       | ENST0000261769.5  | ENSP0000261769.4   | p.Leu15del   |
| BCO2R| YMC41   | X           | 39921408         | In-Frame Variant       | ENST0000324856.7  | ENSP0000320485.7   | p.Ala84CysTer65 |
| CDH1 | YMC41   | 16          | 68771347         | In-Frame Variant       | ENST0000261769.5  | ENSP0000261769.4   | p.Leu15del   |
| TP53 | YMC41   | 17          | 7577106          | In-Frame Variant       | ENST0000261769.5  | ENSP0000261769.4   | p.Pro278Ser   |
| APC  | YMC40   | 5           | 11217544:4       | frameshift Variant     | ENST0000257430.4  | ENSP0000257430.4   | p.Ser138Gly   |
| APC  | YMC40   | 5           | 11217603         | frameshift Variant     | ENST0000257430.4  | ENSP0000257430.4   | p.Ser138Gly   |
|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|
| 9 | GS | YMC40 | 17 | 7577114 | C | T | 659 | 47% | TP53 | missense_variant | ENST00000269305.4:c.824G>A | ENST00000324856.7:c.3999_40 GCA |
| 9 | GS | YMC39 | 1 | 27100181 | CG | GA | C | 306 | 6% | ARID1A | In_Frame_variant | stop_gained | ENST00000324856.7:c.3999_40 GCA |
| 9 | MS | YMC38 | 2 | 30143174 | C | T | 71 | 6% | ALK | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 1 | 27101174 | C | T | 219 | 29% | ARID1A | In_Frame_variant | stop_gained | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 1 | 27100181 | CG | CA | C | 151 | 5% | ARID1A | In_Frame_variant | stop_gained | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 3 | 18534065 | CT | G | 99 | 17% | IRF2 | frameshift_variant | stop_gained | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 1 | 3029792 | C | A | 122 | 25% | JAK2 | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 4 | 55968082 | C | G | 226 | 31% | KDR | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 12 | 25392821 | C | T | 264 | 28% | KARS | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 1 | 11181147 | T | C | 387 | 16% | MTO | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 9 | 70993145 | A | G | 371 | 6% | PGM5 | missense_variant | ENST00000324856.7:c.4456C>T |
| 9 | MS | YMC38 | 3 | 1789471 | T | C | 615 | 42% | PIK3CA | missense_variant | ENST00000324856.7:c.4456C>T |
| MS | YMC38 | 10 | 89720811 | CA | C | 142 | 49% | PTE | frameshift_variant | ENST0000371953.3:c.968delA | ENSP00000361021.3: p.Asn323MetfsTer21 |
|----|-------|----|-----------|----|----|------|-----|-----|-------------------|---------------------------------|----------------------------------|
| GS | YMC37 | 4  | 55961110  | G  | A | 788  | 9%  | KDR | stop_gained       | ENST00000263923.4:c.2830C>T    | ENSP00000263923.4: p.Arg944Ter   |
| GS | YMC35 | 5  | 11217079  | T  | C | 455  | 22% | APC  | missense_variant  | ENST00000257430.4:c.1892T>C    | ENSP00000257430.4: p.Ile631Thr  |
| GS | YMC35 | 18 | 48604707  | G  | A | 218  | 31% | SMA  | missense_variant  | ENST00000342988.3:c.1529G>A    | ENSP00000341551.3: p.Gly510Glu  |
| GS | YMC35 | 17 | 7578554   | A  | G | 154  | 30% | PTE  | frameshift_variant | ENST00000269305.4:c.376T>C     | ENSP00000269305.4: p.Tyr126His  |
| GS | YMC34 | 17 | 7578190   | T  | C | 827  | 28% | TP53 | missense_variant  | ENST00000269305.4:c.659A>G     | ENSP00000269305.4: p.Tyr220Cys  |
| GS | YMC33 | 16 | 68845617  | A  | T | 938  | 20% | CDH1 | missense_variant  | ENST00000261769.5:c.863A>T     | ENSP00000261769.4: p.Asp288Val  |
| GS | YMC33 | 4  | 15324938  | C  | T | 674  | 13% | TP53 | missense_variant  | ENST00000263981.5:c.1154G>A    | ENSP00000263981.4: p.Arg385His  |
| GS | YMC33 | 10 | 89711875  | G  | A | 428  | 13% | PTE  | frameshift_variant | ENST00000371953.3:c.493G>A     | ENSP00000361021.3: p.Gly165Arg  |
| GS | YMC33 | 17 | 7577557   | A  | G | 245  | 16% | TP53 | missense_variant  | ENST00000269305.4:c.724T>C     | ENSP00000269305.4: p.Cys242Arg  |
| EB | YMC32 | 5  | 11217768  | G  | A | 471  | 15% | APC  | missense_variant  | ENST00000257430.4:c.6397A>G    | ENSP00000257430.4: p.Asp2133Asn |
| EB | YMC32 | 1  | 27100181  | CG | C | 373  | 6%  | ARID1A | In_Frame_variant | ENST00000324856.7:c.3999_4001delGCA | ENSP00000320485.7: p.Gln134del |
| EB | YMC32 | X  | 39922966  | G  | A | 437  | 29% | BCO2 | stop_gained       | ENST00000324857.4:c.3640C>T    | ENSP00000345923.4: p.Gln214Ter  |
| EB | YMC32 | 3  | 17892154  | G  | A | 335  | 16% | PIK3CA | missense_variant | ENST00000263967.3:c.1030G>A    | ENSP00000263967.3: p.Val344Met  |
| EB | YMC32 | 3  | 17893608  | G  | A | 285  | 14% | PIK3CA | missense_variant | ENST00000263967.3:c.1624G>A    | ENSP00000263967.3: p.Glu542Lys  |
| EB | YMC32 | 3  | 49413009  | C  | T | 418  | 18% | RHO  | missense_variant  | ENST00000418115.1:c.14G>A      | ENSP00000400175.1: p.Arg5Gln    |
| Genome | YMC31 | 17 | 757035 | TG | T | 804 | 40% | TP53 | frameshift_variant  |
|--------|-------|----|--------|----|---|------|-----|------|-------------------|
| Genome | YMC65 | X  | 39931937 | T  | G | 105 | 16% | BCO R | missense_variant   |
| Genome | YMC65 | 19 | 42795360 | G  | T | 58  | 5%  | CIC  | missense_variant   |
| Genome | YMC66 | 4  | 15324938 | C  | T | 575 | 7%  | FBX W7 | missense_variant  |
| Genome | YMC67 | 3  | 49412973 | C  | T | 414 | 7%  | RHO A | missense_variant   |
| Genome | YMC68 | 17 | 7577120 | C  | T | 458 | 17% | TP53 | stop_gained       |
| Genome | YMC69 | 5  | 11217459 | C  | A | 171 | 5%  | APC  | missense_variant   |
| Genome | YMC64 | 12 | 56495327 | C  | T | 55  | 5%  | ERB B3 | missense_variant   |
| Genome | YMC30 | 1  | 27023484 | G  | C | 179 | 16% | ARID 1A | In_Frame_variant   |
| Genome | YMC30 | 1  | 27100181 | CG | CA| 268 | 6%  | ARID 1A | In_Frame_variant   |
| Genome | YMC30 | 17 | 7574003 | G  | A | 109 | 44% | TP53 | stop_gained       |
| Genome | YMC21 | 1  | 27092839 | A  | G | 613 | 17% | ARID 1A | In_Frame_variant   |
| Genome | YMC29 | 1  | 27105675 | GG | AA | 169 | 8%  | ARID 1A | In_Frame_variant   |
| Genome | YMC29 | X  | 39932171 | G  | A | 446 | 51% | BCO R | stop_gained       |
| Genome | YMC29 | 4  | 18531016 | T  | C | 472 | 23% | IRF2  | missense_variant   |
| Genome | YMC29 | 3  | 17891663 | G  | A | 461 | 18% | PIK3 CA | missense_variant   |
| Sample | Case ID | Chromosome | Position | Gene | Change Type                  | HgNV | Ensembl Accession | Other Information |
|--------|---------|-------------|----------|------|-----------------------------|------|------------------|-------------------|
| MS I   | YMC29   | 3           | 17895208 | A    | 731                        | 6%   | ENST0000263967.3:c.3140A>G | PIK3CA p.His1047Arg |
| MS I   | YMC29   | 3           | 11405800 | A    | 296                        | 17%  | ENST00000357258.3:c.1856delC | ZBTB20 p.Pro619LeufsTer43 |
| MS I   | YMC28   | 12          | 52370254 | G    | 59                         | 5%   | ENST0000257963.4:c.475C>A | ACVR1B p.Arg159Ser |
| MS I   | YMC28   | 14          | 15324429 | A    | 79                         | 5%   | ENST0000263981.4:c.1626G>T | FBXW7 p.Ala542Asn |
| MS I   | YMC28   | 14          | 15333261 | T    | 57                         | 5%   | ENST00000281708.4:c.341A>G | ENSP00000281708.3:p.Glu114Gly |
| GS     | YMC27   | 17          | 7577580  | T    | 230                        | 53%  | ENST00000269305.4:c.701A>G | TP53 p.Asp121Asn |
| GS     | YMC26   | 1           | 27100181 | C    | 963                        | 33%  | ENST0000324856.7:c.3999_401dupGCA | ARID1A In_Frame_variant |
| GS     | YMC26   | 16          | 68844172 | G    | 1289                       | 5%   | ENST000021769.5:c.760G>C | CDH1A Frameshift_variant |
| GS     | YMC25   | 1           | 27023193 | T    | 58                         | 5%   | ENST000002324856.7:c.300delG | ARID1A Frameshift_variant |
| GS     | YMC25   | 16          | 68771347 | C    | 57                         | 5%   | ENST00000261769.4:p.Gln1334del | CDH1A Frameshift_variant |
| GS     | YMC23   | 1           | 27100181 | C    | 268                        | 7%   | ENST0000324856.7:c.3999_401dupGCA | ARID1A In_Frame_variant |
| MS I   | YMC22   | 2           | 30142964 | A    | 248                        | 35%  | ENST00000263967.3:c.3140A>G | ALK p.His1047Arg |
| MS I   | YMC22   | 1           | 27088769 | T    | 110                        | 39%  | ENST00000263967.3:c.3140A>G | ARID1A Frameshift_variant |
| MS I   | YMC22   | 1           | 27100181 | G    | 215                        | 6%   | ENST00000263967.3:c.3140A>G | ARID1A Frameshift_variant |
| MS I   | YMC22   | X           | 39934068 | G    | 335                        | 38%  | ENST0000160740.3:c.3344delC | ENSP000002342724.4:c.529_530delAG |
| MS I   | YMC22   | 19          | 42796882 | G    | 169                        | 37%  | ENST0000160740.3:c.3344delC | ENSP000002342724.4:c.529_530delAG |
|        |         |             |          |      |                             |      | ENSP000002342724.4:c.529_530delAG | Prol115GlufsTer44 |
| MS I YMC22 | 1 | 20773724  |
|-------------|----|-----------|
| MS I YMC22 | 12 | 56477655  |
| MS I YMC22 | 10 | 12331086 |
| MS I YMC22 | 2  | 47657068  |
| MS I YMC22 | 12 | 11001923  |
| MS I YMC22 | 9  | 70993145  |
| MS I YMC22 | 3  | 17892156 |
| MS I YMC22 | 3  | 17895208  |
| MS I YMC22 | 3  | 11405800  |
| MS I YMC22 | 9  | 11001923  |
| MS I YMC22 | 5  | 11217391  |
| MS I YMC22 | 5  | 11217565  |
| CI YMC19  | 1  | 27100181  |
| CI YMC19  | 1  | 11300361  |
| CI YMC19  | 17 | 7577114   |
| CI YMC18  | 1  | 27100181  |

| Gene | Chr | Ref | Alt | ANNO | Description |
|------|-----|-----|-----|------|-------------|
| CR1  | 1   | G   | A   | 157  | 38%         |
| ERB  | B3  | C   | T   | 228  | 43%         |
| FGF  | R2  | G   | A   | 383  | 42%         |
| MSH  | 2   | G   | A   | 235  | 12%         |
| MVK  |     | GC  | G   | 173  | 43%         |
| PGM  | 5   | A   | G   | 700  | 13%         |
| PIK3 | CA  | A   | G   | 247  | 39%         |
| PIK3 | CA  | A   | G   | 247  | 39%         |
| ZBT  |     | C   | T   | 384  | 40%         |
| ERB  | B2  | AG  | A   | 88   | 22%         |
| APC  |     | C   | T   | 620  | 10%         |
| APC  |     | T   | TA  | 394  | 6%          |
| ARID1A | 32  | CG  | CA  | 160  | 5%          |
| MTO  | R   | T   | A   | 74   | 5%          |
| TP53 |     | C   | T   | 346  | 73%         |
| ARID1A | 32  | CG  | CA  | 281  | 5%          |

| Gene | Chr | Ref | Alt | ANNO | Description |
|------|-----|-----|-----|------|-------------|
| CR1  | 1   | G   | A   | 157  | 38%         |
| ERB  | B3  | C   | T   | 228  | 43%         |
| FGF  | R2  | G   | A   | 383  | 42%         |
| MSH  | 2   | G   | A   | 235  | 12%         |
| MVK  |     | GC  | G   | 173  | 43%         |
| PGM  | 5   | A   | G   | 700  | 13%         |
| PIK3 | CA  | A   | G   | 247  | 39%         |
| PIK3 | CA  | A   | G   | 247  | 39%         |
| ZBT  |     | C   | T   | 384  | 40%         |
| ERB  | B2  | AG  | A   | 88   | 22%         |
| APC  |     | C   | T   | 620  | 10%         |
| APC  |     | T   | TA  | 394  | 6%          |
| ARID1A | 32  | CG  | CA  | 160  | 5%          |
| MTO  | R   | T   | A   | 74   | 5%          |
| TP53 |     | C   | T   | 346  | 73%         |
| ARID1A | 32  | CG  | CA  | 281  | 5%          |

74
| Sample | RefGene ID | Genomic Position | Chromosome | Source | Exon | cDNA Change | protein Change | Gene | Type | Description | Reference |
|--------|------------|------------------|------------|--------|------|-------------|---------------|------|------|-------------|-----------|
| GS YMC17 | 2 | 29448360 | C T | 88 | 8% | ALK | missense variant | ENST00000389048.3:c.3139G>A |
| GS YMC17 | 17 | 7578263 | G A | 396 | 56% | TP53 | stop_gained | ENST00000269305.4:c.586C>T |
| GS YMC16 | 3 | 49412898 | T C | 1858 | 11% | RHO A | missense variant | ENST00000418115.1:c.125A>G |
| MS I | YMC15 | 12 | 52369259 | G GC | 290 | 79% | ACV R1B | frameshift variant | ENST00000257963.4:c.303dupC |
| MS I | YMC15 | 5 | 11217567 | AA GAG | A | 130 | 58% | APC | frameshift variant | ENST00000257430.4:c.4391_4394delAGAG |
| MS I | YMC15 | 1 | 27100181 | CG CA C | 207 | 5% | ARID1A | In_Frame variant | ENST00000324856.7:c.3999_4001delGCA |
| MS I | YMC15 | 7 | 14048292 | AG A A | 214 | 18% | BRAF | frameshift variant | ENST00000288602.6:c.1208delC |
| MS I | YMC15 | 13 | 31480888 | G A | 60 | 7% | MED AG AG AG | frameshift variant | ENST00000380482.4:c.236G>A |
| MS I | YMC15 | 13 | 31480851 | CG C A | 60 | 7% | MED AG AG | frameshift variant | ENST00000380482.4:c.206delG |
| MS I | YMC15 | 12 | 11001923 | GC G | 190 | 26% | MKV | frameshift variant | ENST00000228510.3:c.417delC |
| MS I | YMC15 | 17 | 7579546 | CG C | 390 | 81% | TP53 | frameshift variant | ENST00000269305.4:c.140delC |
| MS I | YMC14 | 11 | 69465027 | AG | 75 | 5% | CCN D1 | In_Frame variant | ENST00000227507.2:c.839_841delAGG |
| MS I | YMC14 | 1 | 20778775 | C T | 601 | 5% | CR1 | stop_gained | ENST00000367049.4:c.6580C>T |
| MS I | YMC14 | 3 | 41268766 | A T | 640 | 6% | CTN NB1 | missense variant | ENST00000349496.5:c.1004A>T |
| MS I | YMC14 | 17 | 7574029 | C CG | 313 | 14% | TP53 | frameshift variant | ENST00000269305.4:c.997dupC |
| GS YMC13 | 5 | 11215500 | AG A | 520 | 31% | APC | frameshift variant | ENST00000257430.4:c.1273del |
| Gene | YMC | Chromosome | Position | Variant | Description | Ensembl | Description | G | Description |
|------|-----|------------|----------|---------|-------------|----------|-------------|---|-------------|
| ARID1A | YMC13 | 1 | 27100181 | In_Frame_variant | ENST0000324856.7:c.3999_401delGCA | ENSP00000320485.7:p.Gln1334del |
| CIC | YMC13 | 19 | 42799299 | frameshift_variant | ENST0000160740.3:c.4784delC | ENSP00000320485.7:p.Glu425LysfsTer29 |
| HLA-B | YMC12 | 6 | 31322976 | missense_variant | ENST00000412585.2:c.920C>G | ENSP00000399168.2:p.Pro307Arg |
| JAK2 | YMC12 | 9 | 5022119 | stop_gained | ENST00000381652.3:c.132C>A | ENSP00000371067.3:p.Tyr44Ter |
| TP53 | YMC12 | 17 | 7578203 | missense_variant | ENST00000269305.4:c.646G>T | ENSP00000269305.4:p.Glu224Val |
| TP53 | YMC11 | 17 | 7578493 | stop_gained | ENST00000269305.4:c.437G>A | ENSP00000269305.4:p.Tyr147His |
| TP53 | YMC10 | 17 | 37879658 | missense_variant | ENST00000269571.5:c.2033G>A | ENSP00000269305.4:p.Glu91Val |
| RHOA | YMC10 | 12 | 56482341 | missense_variant | ENST00000269710.1:c.889C>G | ENSP00000269305.4:p.Pro297Thr |
| CCND1 | YMC10 | 3 | 49412898 | missense_variant | ENST00000418115.1:c.125A>G | ENSP00000400175.1:p.Tyr31His |
| TP53 | YMC9 | 11 | 69465987 | In_Frame_variant | ENST00000275707.2:c.839_841delAGG | ENSP0000027507.2:p.Ile282Val |
| TP53 | YMC9 | 17 | 7577094 | missense_variant | ENST00000269305.4:c.844C>T | ENSP00000269305.4:p.Thr282Ile |
| TP53 | YMC8 | 17 | 7577108 | missense_variant | ENST00000269305.4:c.830G>A | ENSP00000269305.4:p.Arg277Trp |
| ARID1A | YMC6 | 1 | 27023908 | In_Frame_variant | ENST0000324856.7:c.1029_1032delAGCTGCCGCGGC | ENSP0000324856.7:p.Ala345_Ala349del |
| Gene | Sample | Chr | Start | End | VUS Type | Ensembl Transcript | Ensembl Protein | VUS Description |
|------|--------|-----|-------|-----|-----------|-------------------|----------------|-----------------|
| ARID1A | YMC5 | 17 | 7577124 | 7577124 | In_Frame Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | In_Frame variant |
| ARID1A | YMC4 | 17 | 17895208 | 17895208 | In_Frame Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | In_Frame variant |
| ARID1A | YMC4 | 17 | 49413010 | 49413010 | In_Frame Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | In_Frame variant |
| CIC | YMC3 | 19 | 42799177 | 42799177 | Missense Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | Missense variant |
| CR1 | YMC2 | 1 | 42799177 | 42799177 | Missense Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | Missense variant |
| KRA S | YMC1 | 2 | 7577124 | 7577124 | Missense Variant | ENST00000324856.7:c.3999_4001del GCA | ENSP00000324856.7: p.Gln1334del | Missense variant |
Supplementary Table 3. Somatic variants of the 43 cancer panel genes in 107 Korean gastric cancers and TCGA data

| Gene     | In this study (n=107) | TCGA, Nature 2014 (n=289) |
|----------|-----------------------|----------------------------|
|          | No of case | Case (%) | No of case | Case (%) |
| TP53     | 41         | 38.3%     | 138        | 47.8%     |
| ARID1A   | 39         | 36.4%     | 90         | 31.1%     |
| CR1      | 16         | 15.0%     | 21         | 7.3%      |
| APC      | 12         | 11.2%     | 42         | 14.5%     |
| BCOR     | 12         | 11.2%     | 21         | 7.3%      |
| CDH1     | 11         | 10.3%     | 29         | 10.0%     |
| CIC      | 10         | 9.3%      | 26         | 9.0%      |
| PIK3CA   | 10         | 9.3%      | 57         | 19.7%     |
| ERBB3    | 9          | 8.4%      | 31         | 10.7%     |
| RHOA     | 9          | 8.4%      | 16         | 5.5%      |
| ERBB2    | 8          | 7.5%      | 14         | 4.8%      |
| CCND1    | 7          | 6.5%      | 1          | 0.3%      |
| FBXW7    | 7          | 6.5%      | 27         | 9.3%      |
| ALK      | 6          | 5.6%      | 12         | 4.2%      |
| KRAS     | 6          | 5.6%      | 28         | 9.7%      |
| MTOR     | 6          | 5.6%      | 23         | 8.0%      |
| CTNNB1   | 5          | 4.7%      | 19         | 6.6%      |
| EGFR     | 5          | 4.7%      | 15         | 5.2%      |
| HLA-B    | 5          | 4.7%      | 20         | 6.9%      |
| MSH2     | 5          | 4.7%      | 5          | 1.7%      |
| PGM5     | 5          | 4.7%      | 25         | 8.7%      |
| ZBTB20   | 5          | 4.7%      | 28         | 9.7%      |
| IRF2     | 4          | 3.7%      | 15         | 5.2%      |
| KDR      | 4          | 3.7%      | 13         | 4.5%      |
| LARP4B   | 4          | 3.7%      | 27         | 9.3%      |
| MKV      | 4          | 3.7%      | 13         | 4.5%      |
| BRAF     | 3          | 2.8%      | 16         | 5.5%      |
| Gene      | n  | Percentage | n  | Percentage |
|-----------|----|------------|----|------------|
| PTEN      | 3  | 2.8%       | 23 | 8.0%       |
| ACVR1B    | 2  | 1.9%       | 11 | 3.8%       |
| CBWD1     | 2  | 1.9%       | 1  | 0.3%       |
| FGFR2     | 2  | 1.9%       | 12 | 4.2%       |
| JAK2      | 2  | 1.9%       | 12 | 4.2%       |
| MEDAG     | 2  | 1.9%       | 7  | 2.4%       |
| MLH1      | 2  | 1.9%       | 6  | 2.1%       |
| SMAD4     | 2  | 1.9%       | 24 | 8.3%       |
| STK11     | 2  | 1.9%       | 3  | 1.0%       |
| CD274     | 1  | 0.9%       | 3  | 1.0%       |
| MDM2      | 1  | 0.9%       | 15 | 5.2%       |
| MYC       | 1  | 0.9%       | 5  | 1.7%       |
| C16orf74  | 0  | 0.0%       | 2  | 0.7%       |
| CCNE1     | 0  | 0.0%       | 5  | 1.7%       |
| MET       | 0  | 0.0%       | 6  | 2.1%       |
| PDCD1LG2  | 0  | 0.0%       | 1  | 0.3%       |
Supplementary Table 4: Performance of 43 genes panel through NGS in GC

| Sample Name   | Target Size (bp) | Target base | Mean depth over target region | Uncovered over target | 1x~ over target ratio | 10x~ over target ratio | 20x~ over target ratio | 50x~ over target ratio | 100x~ over target ratio |
|---------------|------------------|-------------|-------------------------------|-----------------------|-----------------------|------------------------|------------------------|-----------------------|-------------------------|
| YMC 65-tumor  | 124              | 7672        | 618.05                        | 1.73                  | 98.2                  | 97.8                   | 97.5                   | 96.1                  | 94.4                    |
| YMC 68-non-tumor | 132            | 0030        | 1993                          | %                     | 7%                    | 1%                     | 1%                     | 0%                    | 4%                      |
| YMC 28-tumor  | 124              | 7435        | 599.02                        | 1.51                  | 98.4                  | 97.8                   | 97.3                   | 96.3                  | 94.7                    |
| YMC 38-tumor  | 132              | 8070        | 41839                         | %                     | 9%                    | 4%                     | 0%                     | 8%                    | 6%                      |
| 150601-LKA-1_S10a | 124          | 9187        | 740.13                        | 1.40                  | 98.6                  | 98.1                   | 97.6                   | 96.8                  | 95.4                    |
| 150601-LKA-2_S11a | 132         | 4230        | 33258                         | %                     | 0%                    | 2%                     | 2%                     | 5%                    | 9%                      |
| 150601-LKA-3_S12a | 124          | 8944        | 720.57                        | 1.27                  | 98.7                  | 98.1                   | 97.6                   | 96.8                  | 95.5                    |
| Control_gD    | 132              | 6774        | 78848                         | %                     | 3%                    | 4%                     | 3%                     | 5%                    | 0%                      |
| NA            | 132              | 1755        | 71314                         | %                     | 1%                    | 7%                     | 7%                     | 4%                    | 1%                      |

*a* genomic DNA from normal control, *b* gDNA from NA12878
Supplementary Table 5: Coverage of 43 genes panel through NGS in GC

| Gene   | Gene size (bp) | Uncovered bases in target gene | Uncovered bases in total bases (124,132 bp) | Uncovered (%) | 1X   | 10X  | 20X  | 50X  | 100X |
|--------|----------------|-------------------------------|---------------------------------------------|----------------|------|------|------|------|------|
| CR1    | 7470           | 1666                          | 1.34%                                       | 22.30%         | 77.70%  | 73.16%  | 71.07%  | 68.07%  | 65.33%  |
| FGFR2  | 3207           | 71                            | 0.06%                                       | 2.21%          | 97.79%  | 96.69%  | 96.20%  | 94.48%  | 92.98%  |
| ARID1A | 6858           | 1                             | 0.00%                                       | 0.01%          | 99.99%  | 96.84%  | 94.52%  | 91.51%  | 88.63%  |
| MEDAG  | 912            | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 98.68%  | 88.16%  | 69.52%  |
| CCND1  | 888            | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 81.42%  |
| HLA-B  | 1089           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 94.21%  | 85.58%  |
| PTEN   | 1732           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 98.15%  | 87.64%  |
| CBWD1  | 1234           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 96.35%  | 96.35%  | 91.90%  |
| STK11  | 1302           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 92.17%  |
| CDH1   | 2649           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 98.19%  | 93.85%  |
| BRAF   | 2301           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 98.31%  | 94.00%  | 94.00%  |
| ACVR1B | 1641           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 99.76%  | 94.45%  | 94.45%  | 94.45%  | 94.45%  |
| ALK    | 4863           | 0                             | 0.00%                                       | 0.00%          | 100.00% | 100.00% | 100.00% | 100.00% | 100.00% | 96.94%  |
| Gene  | Count | Mismatch | Repeat | Concordance | Quality | VAF       | CI   | VAF      | CI   |
|-------|-------|----------|--------|-------------|---------|-----------|------|----------|------|
| CIC   | 7621  | 0        | 0.00%  | 100.00%     | 100.00% | 99.20%    | 97.35%|
| EGFR  | 3889  | 0        | 0.00%  | 100.00%     | 98.51%  | 97.74%    | 97.74%|
| ERBB2 | 4418  | 0        | 0.00%  | 100.00%     | 98.80%  | 98.35%    | 97.89%|
| CCNE1 | 1233  | 0        | 0.00%  | 100.00%     | 99.92%  | 98.13%    | 98.13%|
| MSH2  | 2805  | 0        | 0.00%  | 100.00%     | 98.86%  |           |      |
| PGM5  | 1704  | 0        | 0.00%  | 100.00%     | 100.00% | 100.00%   |      |
| APC   | 8697  | 0        | 0.00%  | 100.00%     | 99.53%  |           |      |
| BCOR  | 5268  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| C16orf74 | 231  | 0    | 0.00%  | 100.00%     | 100.00% | 100.00%   |      |
| CD274 | 3578  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| CTNNB1| 2346  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| ERBB3 | 4160  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| FBXW7 | 2631  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| IRF2  | 1050  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| JAK2  | 3399  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| KDR   | 4071  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| KRAS  | 687   | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| LARP4B| 2217  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| MDM2  | 1494  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| MET   | 4227  | 0        | 0.00%  | 100.00%     | 100.00% |           |      |
| Gene    | Count | Low | Low Percent | 50.0% | 50.0% | 50.0% | 50.0% | 50.0% |
|---------|-------|-----|-------------|-------|-------|-------|-------|-------|
| MLH1    | 2271  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| MTOR    | 7650  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| MVK     | 1191  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| MYC     | 1365  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| PDCD1LG2| 822   | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| PIK3CA  | 3207  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| RHOA    | 582   | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| SMAD4   | 1659  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| TP53    | 1263  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
| ZBTB20  | 2250  | 0   | 0.00%       | 100.0%| 100.0%| 100.0%| 100.0%| 100.0%|
ABSTRACT (IN KOREAN)

한국인 위암에서의 암 관련 유전자 분석 및 위암의 분자 생물학적 분류

〈지도교수 이경아〉

연세대학교 대학원 의학과

김윤정

최근 암 게놈 아틀라스 (TCGA) 연구 및 Asian Cancer Research Group (ACRG) cohort에서에서 새로운 위암의 분자 생물학적 분류를 제시하였다. 본 연구에서는 기존 TCGA 분류체계에 사용된 방법에 비해 간소화 방법으로 한국인 위암을 TCGA 분류체계에 따라 분류하고 한국인 위암에서 유의하게 관찰되는 유전자 변이, 향후 표적 치료와 환자 예후 예측에 연관된 유전 변이 및 Hereditary cancer syndromes 관련된 유전자 변이를 관찰하고자 하였다. 이를 위해서 TCGA 및 ACRG 연구의 결과를 토대로 위암에서 유의하게 유전적 변이가 관찰되는 유전자, 표적 치료제의 표적 유전자, 및 Hereditary cancer syndromes에 관련된 유전자를 선택하여 43 cancer panel을 제작하여 차세대 염기서열 검사를 시행하였다.

43 유전자 cancer panel을 이용한 107명의 위암 환자의 genetic alteration 결과와 더불어 환경적 요소 (EBV 감염, H. pylori 감염) 및 MSI 결과를 이용하여 위암의 분자 생물학적 분류하였다. TCGA 시스템에 따라 107명의 위암 환자를 분류하면 전체의 6.5%는 EBV subtype, 17.7%는 MSI subtype, 13.1%는 CIN
subtype 그리고 62.6%가 GS subtype으로 분류되었다. MSI subtype은 TCGA와 ACRG에서 보고된 것과 같이 본 연구에서도 분자생물학적 분류 그룹 중 가장 좋은 예후를 보였다. 약 20%의 MSI subtype에서 ZBTB20 유전자에서 P619fs*43 변이가 관찰되었다. P619fs*43 변이의 경우는 MSI subtype에 한정되어 관찰되는 소견을 보였다. 이 소견으로 TCGA data에서 동일하게 관찰되었다. 기존 연구에서 EBV subtype에서 관찰되는 특정적인 유전자 (PIK3CA, JAK2, CD274 및 PDCD1LG2)의 체세포 변이의 양상이 다소 다르게 관찰되었다. CIN subtype는 전체 위암의 13.1%이고 GS subtype은 62.6%로 기존의 TCGA 분류에서 보고된 CIN subtype보다 상대적으로 작은 수의 환자가 CIN subtype으로 분류되었다.

한국인 위암 환자에서 앞으로의 표적 치료와 환자 예후 등에 관련된 임상적으로 의미 있는 유전적 변이를 관찰하였다. 107 위암 조직 중 38개의 위암 조직에서 RTK/RAS/MAPK, PI3K/PTEN/AKT pathways 및 MET 유전자 유전자 변이가 관찰되었다. 생존율 분석에서는 TCGA 분류체계에 따른 subtype 간의 유의한 차이를 보이지 않았다. 단, H. pylori 감염 여부는 TCGA 체계로 위암 분류 시 영향을 미치지 않았지만 H. pylori 감염된 위암 환자가 감염되지 않은 군에 비해서 좋은 예후를 보였다.

본 연구에서는 기존 TCGA 분류체계에 사용된 방법에 비해서 간소화 방법으로 한국인 위암을 TCGA 분류체계에 따라 분류하였고 TCGA 결과와 상응하는 결과를 보였다.

핵심되는 말: 위암, 분자생물학적 분류, 체세포 변이, 썰매포변이
PUBLICATION LIST