Title: Genome-wide analysis of DNA copy number alterations in early and advanced gastric cancers

Running title: Genetic alterations in gastric cancers

Noriyuki Arakawa, MD (1), Tamotsu Sugai, MD (1), Wataru Habano, PhD (2), Makoto Eizuka, MD (1), Ryo Sugimoto, MD (1), Risaburo Akasaka, MD (3), Yosuke Toya, MD (3), Eiichiro Yamamoto, MD (4), Keisuke Koeda, MD (5), Akira Sasaki MD (5), Takayuki Matsumoto, MD (3), and Hiromu Suzuki, MD (4)

(1) Department of Molecular Diagnostic Pathology, School of Medicine, Iwate Medical University, 19-1, Morioka 020-8505, Japan
(2) Department of Pharmacodynamics and Molecular Genetics, School of Pharmacy, Iwate Medical University, 19-1, Morioka 020-8505, Japan
(3) Division of Gastroenterology, Department of Internal Medicine, School of Medicine, Iwate Medical University, 19-1, Morioka 020-8505, Japan
(4) Department of Molecular Biology, Sapporo Medical University School of Medicine, 1-17, Minami-1, Cyuouku, Sapporo city, Japan
(5) Department of Surgery, School of Medicine, Iwate Medical University, 19-1, Morioka 020-8505, Japan

Correspondence:
Tamotsu Sugai, MD

Department of Molecular Diagnostic Pathology

Iwate Medical University, 19-1 Morioka 020-8505, Japan

TEL: +81-19-651-5111; FAX: +81-19-629-1437

E-mail: tsugai@iwate-med.ac.jp

Keywords: copy neutral LOH; copy number alteration; gain; gastric cancer; loss of heterozygosity

Conflict of interest statement: We declare that we have no conflicts of interest.

Abstract

To better understand progressive changes in gastric cancer (GC), early and advanced GCs (EGC and AGC, respectively) were examined for copy number alterations (CNAs). A crypt isolation method was used to isolate DNA from tumors and normal glands in 20 AGCs, and fresh tumor samples were obtained from 45 EGCs. We assessed CNAs for differentiated-type GCs using an Infinium HumanCytoSNP-12v2.1 BeadChip in EGCs and AGCs. The most frequent aberrations in EGC were gains at 8q23.3 (42.2%) and 8q23.2 (40%), and loss of heterozygosity (LOH) at 3p14.2 (24.2%), suggesting that these CNAs were involved in the development of EGC. On the other hand, the highest frequencies of gains in AGC were found at 8q24.21 (65%) and 8q24.3 (60%). The most frequent LOHs in AGC were at 11q24.3-25, 11q23.2-24.1, 11q14.1 and 12p11.21-13.33, whereas that in EGC was at 3p14.2. In addition, regions of copy-neutral LOHs in AGC were detected at 11q21, 11q13.3-14.3, 11q11, 11p13-15.3, 12q21.1, 12q12-13.3 and 5q33.3-35.1. Comparisons of gains in EGC and AGC showed significant differences at 12q22–q23.2, 12q21.33, 11p12, 11p14.1, 12q21.31-32.32, 3p12.3, 3p14.1, 10p15.1, 1q24.2 and 2q12.1. Copy neutral LOHs were significantly higher in AGC than in EGC at 14q32.11-32.33, 14q21.3, 14q11.2, 5q11.2, 5q 13.3, 14q21.1-23.2, 14q13.2-13.3, 5q12.1-12.3, 5q11.1 and 17p13.3. The total lengths of the CNAs were significantly greater in AGC than in EGC. We found that the pattern of CNAs in AGC was quite different from that in EGC. We suggest that increasing numbers of CNAs are associated with disease progression from EGC to AGC.

Introduction

Whereas GC is the fourth most common cancer in the world and the second leading cause of death due to cancer [1], the worldwide incidence of GC has declined rapidly over recent decades [2]. On the other hand, high rates of GC are present in Central and Eastern Europe and South America [2] and the incidence of GC remains high in Japan [1, 2]. Although the diagnosis and therapy of GC have improved over the past decade, better understanding of gastric carcinogenesis is important to provide further advances in the diagnosis and therapy of GC.

Sporadic GC is divided into two histological entities: 'intestinal' and 'diffuse'. They differ in epidemiology, pathogenesis, clinical outcome and genetic profile [3]. At the genetic level, the hallmark of cancer is genomic destabilization. At least two molecular phenotypes have been identified in gastrointestinal cancer, and they have been associated with distinct pathways of genomic instability. Those phenotypes are high-level microsatellite instability (MSI) and chromosomal instability (CIN) [3 – 5]. The MSI phenotype is caused by DNA methylation of the *MLH1* gene [4]. In addition, the CpG island methylation phenotype (CIMP) shows genome-wide methylation patterns in MSI [4]. The CIN phenotype is characterized by large genomic changes at the chromosomal level as represented by copy number alteration (CNA) in cancer cells [6, 7]. CNA refers to a form of genomic structural change that results in variations in the number of copies of one or more sections of DNA [5 – 7]. Mutations in p53 play an important role in genomic instability that is closely associated with the CIN phenotype [5 – 7]. CIN that is characterized by CNAs is a major phenotype in GC [5 – 7].

GC is classified into two categories based on tumor stage: early gastric cancer (EGC) and advanced gastric cancer (AGC) [3]. EGC is defined as an adenocarcinoma that is restricted to the mucosa or submucosa, regardless of lymph node metastasis [8, 9]. On the other hand, AGC is defined as a tumor that invades beyond the muscular layer, regardless of lymph node metastasis [8]. This classification has been accepted worldwide, given that it is based on the patient prognosis of GC. Understanding the differences of genetic alterations between EGC and AGC is important for the development of new approaches to diagnosis and/or therapy in GC.

In the present study, our aim was to identify genetic alterations at the whole chromosome level, that is CNAs, in EGCs and AGCs. In addition, we examined CNAs that were closely associated with tumor progression in GCs.

Materials and Methods

1. Patients

Sporadic GC samples were obtained from 65 patients (45 EGCs and 20 AGCs) who had undergone endoscopic or surgical resection at the Iwate Medical University Hospital between 2012 and 2015. No preoperative neoadjuvant therapy or radiotherapy was given to any of these patients. Patients with a family history of GC were not included in the study. The histological criteria used to diagnose intramucosal adenocarcinoma (IMA) were based on Japanese histological criteria, given that the criteria used by Japanese pathologists are different from those of Western pathologists [9]. Well-differentiated and moderately-differentiated tubular adenocarcinomas and

papillary adenocarcinomas corresponding to the intestinal-type (Lauren's classification) were used to analyze genetic alterations (defined as differentiated-type adenocarcinomas) in AGCs. However, poorly differentiated adenocarcinomas and signet ring cell carcinomas that were classified as diffuse type according to Laurens's classification were not included in the study. Clinicopathological findings are summarized in Table 1.

2. Tumor tissue sampling

2-1. Sampling for early gastric cancer

Tumor tissue was obtained from the resected stomach using biopsy forceps within 30 min of resection. Normal gastric mucosa that was distant from the neoplasm was removed from the submucosa using scissors. Tissue for clinicopathological analysis was obtained from a region of the resected stomach adjacent to the site used for molecular analysis. All of the tissue samples contained cancerous tissue. In tumor samples, neoplastic cells accounted for at least 50% of the tissue. As a control, gastric biopsies from EGC patients with chronic gastritis were included.

2-2. Crypt isolation method (sampling for AGC)

A crypt isolation method was used in AGC to obtain pure tumor and non-neoplastic glands. AGC tumor samples were obtained from the central area of the tumor. Normal gastric mucosa was taken from the most distal portion of the stomach. Crypt isolation from the tumor and normal mucosa was performed in accordance with a

previously reported method [10]. Briefly, fresh mucosa and tumor were cut with a sharp scalpel into minute pieces, then incubated at 37℃ for 30 min in calcium- and magnesium-free Hanks' balanced salt solution (CMF) containing 30 mmol/L EDTA. The tissue was then stirred in CMF for 30-40 min. The isolated glands were immediately fixed in 70% ethanol and stored at 4℃ until used for DNA extraction.

The fixed isolated glands were observed under a dissection microscope (SZ60, Olympus, Tokyo, Japan). The isolated glands were processed routinely to confirm its nature using paraffin-embedded histological sections. Contamination by other materials such as interstitial cells was not seen in the samples that were examined. A representative example of a tumor gland is shown in Figure 4.

3. DNA extraction

DNA was extracted from isolated normal and tumor tissue by sodium dodecyl sulfate (SDS) lysis and proteinase K digestion, followed by a phenol-chloroform procedure as reported previously [10].

4. Copy number alteration analysis

Extracted DNA was adjusted to a concentration of 50 ng/μL. All 65 paired samples were assayed using the Infinium HumanCytoSNP-12v2.1 BeadChip (Illumina, San Diego, CA). It contained 299,140 single nucleotide polymorphism (SNP) loci, according to the Illumina Infinium HD assay protocol. BeadChips were scanned using iScan (Illumina) and analyzed using GenomeStudio software (v.2011.1; Illumina). Log

R ratio (LRR) and B allele frequency (BAF) data from each sample were exported from normalized Illumina data using GenomeStudio. Data analysis was conducted with KaryoStudio 1.4.3 [CNV (copy number variation) Plugin v3.0.7.0; Illumina]. The program was used with default parameters. The copy number alterations were classified by CNA partition algorithms. LRR 0 means a normal diploid region. LRR>0 means copy number gain. LRR<0 means copy number loss-of-heterozygosity (LOH). BAF values range from 0 to 1, homozygous SNPs have BAFs near 0 (A-allele) or 1 (B-allele); heterozygous diploid region SNPs have BAFs near 0.5 (AB genotype). Additionally, LRR and BAF data were used to identify regions of hemizygosity and copy-neutral LOH.

5. Calculation of length of copy number alterations on a genome-wide scale in early and advanced gastric cancers

To quantitate copy number aberrations on a genome-wide scale, we calculated the total lengths of CNAs (losses + gains), total length of CNA gains, total length of CNA LOHs, and total length of CNA-copy neutral LOHs identified by the SNP-array analysis, as previously described (11). We observed a strong correlation between the total number of CNAs in the tested AGC samples and the total lengths of the CNAs. We therefore used the total CNA length as an index representing the degree of chromosomal alteration and assessed the relationship between CNA length (total CNA, CNA gain, CNA LOH and CNA copy neutral LOH) and EGC or AGC.

6. Statistical analysis

Data obtained for copy number variations in EGC and AGC were analyzed using chi-square tests (Stat Mate-III software). If significant differences in the lengths of CNAs were found among the 2 groups, differences between the two groups were analyzed using a Mann–Whitney U test (PRISM6, GraphPad software, La Jolla, CA). Differences with *p* values of less than 0.05 were considered significant.

Results

1. Genomic alterations in EGC

Chromosomal CNAs were observed in all 45 EGCs. The average frequencies of CNAs across the entire genome are shown in Figure 1. Genomic CNAs were detected in more than 30% of EGC cases and are summarized in Table 2. The total number of chromosomal aberrations per patient had a mean value of 145.9 with an average of 113.7 gains, ranging from 0 to 771; the mean number of LOHs was 5.1, ranging from 0 to 58 and copy neutral LOHs averaged 27, ranging from 0 to 367. The most frequent aberrations were gains at 8q23.3, 8q23.2, 8p22-23.1, 9p13.1, 8p11.21-11.23, 8p12-21.3, 8p11.1, 8p23.2-23.3, 8q24.3, 9p23-24.1, 8q24.21 and 17p13.3 (31.1%) (Table 2-a). Although none of the LOH cases occurred in more than 30% of the patients, the LOH at 3p14.2 (24.2%) was the most frequent LOH site in this study (Table 2-a). On the other hand, CN-LOH occurred at a low frequency in EGCs (Table 2-a). Minimal common gain regions were 9p13.1 and 8p23.2-23 (37.8% and 35.9%, respectively). *IGFBPL1* and *CSMD1* were selected in the genes located at 9p13.1 and 8p23.2-23.3, respectively,

from published data. Similarly, we identified only one the region showing frequent loss, The minimal common LOH region was at 3p14.2 where the *FHIT* gene was harboured. A minimal common region was not found in the remaining frequent gains.

2. Genomic alterations in advanced gastric cancer

The genomic CNAs were examined by SNP Array in 20 AGCs. Chromosomal CNAs are summarized in Figure 2. Analysis of the averaged frequencies of copy number gains and losses revealed that CNAs were detected across the entire genome (Figure 1, bottom). The mean total number of chromosomal aberrations per patient was 331.8, with an average of 197.7 gains ranging from 45 to 594; there was an average of 27 LOHs ranging from 0 to 230 and we found 108 copy neutral LOHs ranging from 0 to 313. Regions of gain detected in more than 50% of cases were identified at 8q24.21, 8q24.3, 1q24.2, 8q24.22-23, 8q24.13, 20p12.2-13, 1q41-42.3, 1q31.1-32.3, 1q24.3, 8q24.12, 8q22.3, 10q24.32, 10q23.2-23.33, 11q25, 11q23.1 and 12q21.31-23.2 in decreasing order of frequency (Table 2-b). Regions of LOH detected in more than 30% of the cases were at 11q24.3-25, 11q23.2-24.1, 11q14.1, 12p11.21-13.33 and 3p14.2 in decreasing order of frequency (Table 2-b). On the other hand, regions of copy-neutral LOH detected in more than 50% of cases were at 11q21, 11q13.3-14.3, 11q11, 11p13-15.3, 12q21.1, 12q12-13.3 and 5q33.3-35.1 in decreasing order of frequency (Table 2-b). No minimal common region was detected in the AGCs, given that the alterations were found throughout the entire region. Candidate genes that were

associated with CNAs are listed in Table 3 and the candidate genes related to carcinogenesis are also shown.

3. Comparison of genomic alterations identified in EGCs and AGCs

We compared the patterns of CNAs between the EGCs and the AGCs, as shown in Table 4. Regions of gain detected in more than 40% of cases were selected for comparison of EGCs with AGCs. We found significant differences of gains at 12q22–q23.2, 12q21.33, 11p12, 11p14.1, 12q21.31-32.32, 3p12.3, 3p14.1, 10p15.1, 1q24.2, and 2q12.1 between EGCs and AGCs (Table 4). We also examined regions of loss that were found in more than 30% of cases of EGCs and AGCs, as shown in Table 4. Although statistically significant differences in the frequencies of LOHs between EGCs and AGCs were found at 8p12, 8p21.1, 8p21.2, 8p 21.3, 8p 23.1, 8p23.2 and 8p23.3, these regions constituted less than 30% of the cases (Table 4). However, the number of copy neutral LOHs at 14q32.11-32.33, 14q21.3, 14q11.2, 5q11.2, 5q 13.3, 14q21.1-23.2, 14q13.2-13.3, 5q12.1-12.3, 5q11.1 and 17p13.3 were significantly higher in the AGCs than in the EGCs (Table 4).

We examined common CNA regions in EGCs and AGCs. Gains at 8q24.21 and 17q12 were frequent in both EGCs and AGCs. Gains at 8q24.21, the location of the *c-MYC* gene, were detected in 14 of 45 EGCs (31.1%) and 13 of 20 AGCs (65%). On the other hand, gains at 17q12, where the *ERBB2* gene is harbored, were observed in 9 of 45 EGCs (20%) and 8 of 20 AGCs (40%). *FHIT*, located at 3p14.2, was frequently observed as a LOH locus in 11 of 45 EGCs (24.4%) and 7 of 20 AGCs (35%). Pathological findings and an ideogram of genomic imbalance in EGCs and AGCs are

shown in Figure 3 and 4. Summary data describing CNAs in AGCs and EGCs are shown in the supplementary Tables.

4. Association of length of copy number alterations on a genome-wide scale in early and advanced gastric cancers

The overall total length of CNVs in AGCs was longer than that of EGC ($p<0.001$) (Figure 5). We analyzed genomic losses (LOHs and copy neutral LOHs) and gains separately. The total length of CNV gains in EGCs was significantly greater than that of AGCs, as shown in Figure 5 ($p<0.001$). Significant differences in the total length of losses (LOH and copy neutral LOH) between EGC and AGC were found ($p<0.001$).

Discussion

Using next-generation sequencing and other genomic technologies, we identified detailed molecular alterations in GCs in The Cancer Genome Atlas Project (TCGA) [12]. Those high-throughput technologies now allow a comprehensive analysis of genomic and epigenomic alterations associated with GC and permits the establishment of molecular profiles of GC [12, 13]. Recently, the TCGA network proposed a classification scheme of four GC subtypes based on the underlying molecular alterations [12]. This landmark study has significantly enhanced our understanding of the characteristics of GC at the molecular level [12]. However, the platform used for molecular analysis in the present study differed from that of TCGA. Specifically, the number of probes plotted in the array and the molecular distance between each probe

are different [12]. In addition, an Infinium HumanCytoSNP-12v2.1 BeadChip was used in the present study, whereas Affymetrix SNP 6.0 arrays were used in TCGA [12]. Because the results of chromosomal alterations are platform-dependent, the molecular data derived in the present study cannot be directly compared with that of TCGA. However, we belive that the present data contribute to our understanding of gastric carcinogenesis.

One of the primary goals of analyzing gastric carcinogenesis is to understand the genetic basis that contributes to its development. It was anticipated that genome-wide studies would identify alleles that were closely associated with specific diseases, including various cancers [12, 13]. Indeed, genome-wide studies have identified thousands of variants associated with hundreds of targets [12 - 14]. We attempted to identify specific loci that were associated with tumor development. However, CNAs were found to affect entire regions in the AGCs that we examined [15]. In the present study, we identified 3 minimal common gain regions in EGCs, including gains at 9p13.1 and 8p23.2-23.3 and LOH at 3p14.2. In the present study, candidate genes that might play critical roles in gastric carcinogenesis were found in clusters. We found insulin-like growth factor binding protein-like 1 *(IGFBPL1)* at 9p13.1 and *CUB and* Sushi Multiple Domains 1 *(CSMD1)* at 8p23.2-23.3 as minimal common gain regions [15 – 17]. Furthermore, the *FHIT* gene, located at 3p14.2, was found as a minimal common LOH region [18]. Smith et al. found that the silencing of *IGFBPL1* was strongly associated with aggressive clinical disease, suggesting that *IGFBPL1* acts as tumor suppressor gene, not an oncogene [15]. The *CSMD1* gene is a novel candidate

tumor suppressor located at 8p23.2-23.3. *CSMD1* is frequently found to be deleted, mutated or methylated in many cancers [16, 17]. Shull et al. showed that *CSMD1* expression was frequently lost in breast cancer, indicating that it functions as a tumor suppressor [17]. If both *IGFBPL1* (located at 9p13.1) and *CSMD1* (at 8p23.2-23.3) act as tumor supressors, it is difficult to explain how their gain contributes to tumor development. In the present study, we suggest that the majority of CNAs showing gain may be caused by secondary genetic effects that may not positively contribute to gastric carcinogenesis. On the other hand, the LOH involving 3p14.2 has been reported in gastric and other human cancers [18]. The *FHIT* gene located at 3p14.2 is commonly lost in EGCs and AGCs [18]. In addition, previous studies have shown that the *FHIT* gene is inactivated in the majorty of GCs, thus suggesting that the loss of the *FHIT* gene can play a critical role in early gastric carcinogenesis.

It is well known that the 8q24.21 region harbors the *c-MYC* gene whose amplification is frequently found in various cancers, including gastric, colorectal and ovarian cancers [19 – 23]. In the present study, we showed copy number gains of c-*MYC* at a significantly high prevalence in both AGCs and EGCs. This finding suggests that *c-MYC* plays a critical role in not only advanced gastric carcinogenesis, but also early gastric carcinogenesis. *POU5F1B*, which is located adjacent to *c-MYC* on human chromosome 8q24, is frequently amplified in GC [24]. In a previous study, *POU5F1B* was also found to be amplified at a high level in human GC tissue. Recent study has shown that *POU5F1B* amplification is associated with a poor prognosis in GC patients [24]. Jin et al. showed that the *PUF60*, *BOP1* and *E2F1* genes at 8q24.3 were

significantly over-expressed in tumor tissues with copy number gains [25]. In particular, they observed that copy number gains of *PUF60* show a strong positive correlation with expression in GCs [25]. Gumireddy et al. reported that *PUF60* is an important gene via translational regulatory LncRNA, as it is involved in tumor invasion and metastasis [26]. These observations suggest that copy number gains of *PUF60* may be a major mechanism underlying the over-expression of genes in GC. In addition, a previous study showed that genes located at 8q23-24 might define the development of lymphatic metastases in colorectal cancers [27]. Accordingly, our data support the notion that the gain of this region could play an important role in gastric carcinogenesis and could predict the metastatic potential of primary AGCs or invasion of EGCs.

Here, we found that 17q12 showed a CNA gain, a region that includes the chromosomal locus for the *HER2* gene [28, 29]. *HER2* is a key driver in gastric tumorigenesis, and its gene amplification leads to overexpression in GC [28, 29]. There is increasing evidence that *HER2* amplification and HER overexpression are important biomarkers in GC, and its amplification plays a crucial role in a specific type of gastric tumorigenesis [28, 29]. In a previous study, amplification of *HER2* was strongly associated with poor carcinoma-specific survival, particularly evident in a subgroup of intestinal cancers [30]. In the present study, 17q12 (involving the *HER2* gene) was a common gain in AGCs and EGCs. This finding suggests that gain at 17q12 is a novel candidate biomarker to predict tumor progression of GC.

LOH and copy neutral-LOH provide valuable information for the identification of tumor suppressor genes [31]. Regions affected by LOH and copy neutral-LOH

contain many candidate tumor suppressors, and LOH in tumor cells is thought to play pivotal roles in gastric carcinogenesis [32]. In the present study, however, LOH appeared not to be a driving force for early gastric carcinogenesis because LOH occurred at a low frequency. Our findings showed that the most frequent LOH locus was at 3p14.2, harbouring the *FHIT* gene in EGCs. However, copy-neutral LOH was not detected in the EGCs examined here.The present data suggest that although LOH and copy-neutral LOH were frequent events in AGC, copy-neutral LOH rather than LOH is a major driving force in advanced stages of GC.

The possible association between CNA and expression was examined to confirm the significance of our array data in EGC and AGC. We selected genes from regions with tumor-specific CNAs based on published expression data. The following genes were gained and elevated expression has been noted: *KITLG* [33], *LGR5* [34], *IGF1* [35], *ASCL1* [36], *TRAF6* [37], *FOXP1* [38] and *OCT 1* [39]. However, the loss of *DICER1* [40], *MEG3* [41], *DLK1* [42], *CHGA* [43], *miR-127* [44], *NDRG2* [45], *APE-1* [46] and *PTGER2* [47] was inversly correlated with gene expression. Thus, it is likely that genes that play important roles in gastric carcinogenesis have not yet been reported. Among them, some genes that are responsible for the development of gastrointestinal tumors were selected. Leucine-rich repeat-containing G protein-coupled receptor 5 *(LGR5)* has recently been reported to be a marker of cancer stem cells in colorectal cancer [48]. Many studies have suggested that *LGR5* plays a key role in colorectal carcinogenesis and is associated with the poor outcome of CRC patients [48]. In addition, overexpression of *LGR5* in *in vitro* assays resulted in enhanced proliferation and

resistance to chemotherapy [48]. The KIT ligand gene (*KITLG*), also known as stem cell factor, achieves multiple biological functions during development by triggering its receptor tyrosine kinase protein, c-KIT [31]. *KITLG* has been implicated in the development of several cancers, including colorectal cancer, GC and lung cancer [49]. This finding supports the notion that CRC tissues that over-express KITLG may be subject to tumor growth and invasion. It was reported that octamer transcription factor 1 *(OCT1)* influenced the progression of CRC [50]. Wang et al. showed that *OCT1* expression independently predicted poor patient prognosis in CRC [39]. In addition, *OCT1* is amplified and overexpressed in GC; both are associated with poor survival in patients with GC [50]. In the present study, there was a significant difference in the frequencies of gains at 12q22-23.2 between EGCs and AGCs. We suggest that gain at 12q22-23.2 plays a major role in the progression of cancer cells and that *LGR5, KITLG* and *OCT1* genes are novel candidate markers that drive tumor invasion in GC.

In the genetic model of colorectal carcinogenesis proposed by Vogelstein et al., loss of genetic material is an essential alteration for progression of colorectal cancer [32]. A similar hypothesis has been proposed in gastric carcinogenesis [51]. However, colorectal cancer cells as well as GC cells acquire DNA aneuploidy through gains of genetic materials during tumor progression [13]. In the present study, the alterations of CNAs in AGCs had a total length greater than those of CNAs in EGCs. These findings suggest that alterations of CNAs by gain rather than by loss play an essential role in the progression of gastric tumorigenesis [13].

In conclusion, we comprehensively examined allelotypes in GC based on copy

number changes. We showed diverse chromosomal regions were involved in EGCs and AGCs. These abnormal regions may be associated with tumor progression during development of gastric carcinogenesis. The integrated analysis of gene CNAs pointed to several interesting genes as potential biomarkers for GC although further studies need to be performed. Taken together, these results may be helpful in the understanding of gastric carcinogenesis.

Acknowledgments

We gratefully acknowledge the technical assistance of Ms. E. Sugawara and Mr. T. Kasai. We also thank members of the Department of Molecular Diagnostic Pathology, Iwate Medical University for their support.

References

1. Montori G, Coccolini F, Ceresoli M, et al. The treatment of peritoneal carcinomatosis in advanced gastric cancer: state of the art. Int J Surg Oncol. 2014;2014:912418.
2. Jemal A, Center MM, DeSantis C, Ward EM. Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiol Biomarkers Prev. 2010;19:1893-907.
3. Ferrasi AC, Pinheiro NA, Rabenhorst SH, et al. Helicobacter pylori and EBV in gastric carcinomas: methylation status and microsatellite instability. World J Gastroenterol. 2010;16:312-9.
4. Sugimoto R, Sugai T, Habano W, et al. Clinicopathological and molecular alterations in early gastric cancers with the microsatellite instability-high phenotype. Int J Cancer. 2016;138:1689-97.
5. Lim B, Kim JH, Kim M, Kim SY. Genomic and epigenomic heterogeneity in molecular subtypes of gastric cancer. World J Gastroenterol. 2016 ;22:1190-201.
6. Liang L, Fang JY, Xu J. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. Oncogene. 2016;35:1475-82.
7. Tsukamoto Y, Uchida T, Karnan S, et al. Genome-wide analysis of DNA copy number alterations and gene expression in gastric cancer. J Pathol. 2008;216:471-82.
8. Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma: 3rd English edition. *Gastric Cancer* 2011; **14**: 101–112.

9. Sugai T, Inomata M, Uesugi N, et al. Analysis of mucin, p53 protein and Ki-67 expressions in gastric differentiated-type intramucosal neoplastic lesions obtained from endoscopic mucosal resection samples: A proposal for a new classification of intramucosal neoplastic lesions based on nuclear atypia. Pathol Int 2004; 54: 425-435.
10. Habano W, Sugai T, Nakamura S, Yoshida T. A novel method for gene analysis of colorectal carcinomas using a crypt isolation technique. Lab Invest 1996; 74: 933-940.
11. Sawada T, Yamamoto E, Suzuki H, et al. Association between genomic alterations and metastatic behavior of colorectal cancer identified by array-based comparative genomic hybridization. Genes Chromosomes Cancer. 2013;52:140-9.
12. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202-9.
13. Tan P, Yeoh KG. Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology. 2015;149:1153-1162.
14. Sunakawa Y, Lenz HJ. Molecular classification of gastric adenocarcinoma: translating new insights from the cancer genome atlas research network. Curr Treat Options Oncol. 2015;16:17.
15. Smith P, Nicholson LJ, Syed N, et al. Epigenetic inactivation implies independent functions for insulin-like growth factor binding protein (IGFBP)-related protein 1 and the related IGFBPL1 in inhibiting breast cancer phenotypes. Clin Cancer Res. 2007;13:4061-8.

16. Zhang R, Song C. Loss of CSMD1 or 2 may contribute to the poor prognosis of colorectal cancer patients. Tumour Biol. 2014;35:4419-23.
17. Shull AY, Clendenning ML, Ghoshal-Gupta S, et al. Somatic mutations, allele loss, and DNA methylation of the Cub and Sushi Multiple Domains 1 (CSMD1) gene reveals association with early age of diagnosis in colorectal cancer patients. PLoS One. 2013;8:e58731.
18. Xiao YP, Wu DY, Xu L, Xin Y. Loss of heterozygosity and microsatellite instabilities of fragile histidine triad gene in gastric carcinoma. World J Gastroenterol. 2006;12:3766-9.
19. Kim S, Lee J, Hong ME, et al. High-throughput sequencing and copy number variation detection using formalin fixed embedded tissue in metastatic gastric cancer. PLoS One. 2014;9:e111693.
20. Calcagno DQ, Leal MF, Assumpcao PP, Smith MA, Burbano RR. MYC and gastric adenocarcinoma carcinogenesis. World J Gastroenterol. 2008;14:5962-8.
21. Lee KS, Kwak Y, Nam KH, et al. c-MYC Copy-Number Gain Is an Independent Prognostic Factor in Patients with Colorectal Cancer. PLoS One. 2015;10:e0139727.
22. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012 ;487:330-7.
23. Darcy KM, Brady WE, Blancato JK, et al. Prognostic relevance of c-MYC gene amplification and polysomy for chromosome 8 in suboptimally-resected, advanced stage epithelial ovarian cancers: a Gynecologic Oncology Group study. Gynecol

Oncol. 2009;114:472-9.

24. Hayashi H, Arao T, Togashi Y, et al. The OCT4 pseudogene POU5F1B is amplified and promotes an aggressive phenotype in gastric cancer. Oncogene. 2015;34:199-208.

25. Jin DH, Park SE, Lee J, et al. Copy Number Gains at 8q24 and 20q11-q13 in Gastric Cancer Are More Common in Intestinal-Type than Diffuse-Type. PLoS One. 2015;10:e0137657.

26. Gumireddy K, Li A, Yan J, et al. Identification of a long non-coding RNA-associated RNP complex regulating metastasis at the translational step. EMBO J. 2013;32: 2672–2684.

27. Ghadimi BM, Grade M, Liersch T, et al. Gain of chromosome 8q23-24 is a predictive marker for lymph node positivity in colorectal cancer. Clin Cancer Res. 2003;9:1808-14.

28. Cappellesso R, Fassan M, Hanspeter E, et al. HER2 status in gastroesophageal cancer: a tissue microarray study of 1040 cases. Hum Pathol. 2015;46:665-72.

29. Stahl P, Seeschaaf C, Lebok P, et al. Heterogeneity of amplification of HER2, EGFR, CCND1 and MYC in gastric cancer. BMC Gastroenterol. 2015;15:7.

30. Gravalos C, Jimeno A. HER2 in gastric cancer: a new prognostic factor and a novel therapeutic target. Ann Oncol. 2008;19:1523-9.

31. Chen W, Yuan L, Cai Y, et al. Identification of chromosomal copy number variations and novel candidate loci in hereditary nonpolyposis colorectal cancer with mismatch repair proficiency. Genomics. 2013;102:27-34.

32. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998; 396: 643–649.

33. Yang S, Li WS, Dong F, et al. KITLG is a novel target of miR-34c that is associated with the inhibition of growth and invasion in colorectal cancer cells. J Cell Mol Med. 2014;18:2092-102.

34. Huang T, Qiu X, Xiao J, et al. The prognostic role of Leucine-rich repeat-containing G-protein-coupled receptor 5 in gastric cancer: A systematic review with meta-analysis. Clin Res Hepatol Gastroenterol. 2016;40:246-53.

35. Shiratsuchi I, Akagi Y, Kawahara A, et al. Expression of IGF-1 and IGF-1R and their relation to clinicopathological factors in colorectal cancer. Anticancer Res. 2011;31:2541-5.

36. Lenhart R, Kirov S, Desilva H, et al. Sensitivity of Small Cell Lung Cancer to BET Inhibition Is Mediated by Regulation of ASCL1 Gene Expression. Mol Cancer Ther. 2015;14:2167-74.

37. Han F, Zhang L, Qiu W, Yi X. TRAF6 promotes the invasion and metastasis and predicts a poor prognosis in gastric cancer. Pathol Res Pract. 2016;212:31-7.

38. He M, Gao L, Zhang S, et al. Prognostic significance of miR-34a and its target proteins of FOXP1, p53, and BCL2 in gastric MALT lymphoma and DLBCL. Gastric Cancer. 2014;17:431-41.

39. Wang YP, Song GH, Chen J, et al. Elevated OCT1 participates in colon tumorigenesis and independently predicts poor prognoses of colorectal cancer patients. Tumour Biol. 2016;37:3247-55.

40. Zheng ZH, Sun XJ, Fu WN, et al. Decreased expression of DICER1 in gastric cancer. Chin Med J (Engl). 2007;120:2099-104.
41. Braconi C, Kogure T, Valeri N, et al. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. Oncogene. 2011;30:4750-6.
42. Anwar SL, Krech T, Hasemeier B, et al. Loss of imprinting and allelic switching at the DLK1-MEG3 locus in human hepatocellular carcinoma. PLoS One. 2012;7:e49462.
43. Sciarra A, Monti S, Gentile V, et al. Chromogranin A expression in familial versus sporadic prostate cancer. Urology. 2005;66:1010-4.
44. Zhou J, Lu S, Yang S, et al. MicroRNA-127 post-transcriptionally downregulates Sept7 and suppresses cell growth in hepatocellular carcinoma cells. Cell Physiol Biochem. 2014;33(5):1537-46.
45. Feng L, Xie Y, Zhang H, Wu Y. Down-regulation of NDRG2 gene expression in human colorectal cancer involves promoter methylation and microRNA-650. Biochem Biophys Res Commun. 2011;406:534-8.
46. Futagami S, Hiratsuka T, Shindo T, et al. Expression of apurinic/apyrimidinic endonuclease-1 (APE-1) in H. pylori-associated gastritis, gastric adenoma, and gastric cancer. Helicobacter. 2008;13:209-18.
47. Tian L, Suzuki M, Nakajima T, et al. Thoracic Surgery Group. Clinical significance of aberrant methylation of prostaglandin E receptor 2 (PTGER2) in nonsmall cell lung cancer: association with prognosis, PTGER2 expression, and

epidermal growth factor receptor mutation. Cancer. 2008;113:1396-403.

48. Jiang Y, Li W, He X, Zhang H, Jiang F, Chen Z. Lgr5 expression is a valuable prognostic factor for colorectal cancer: evidence from a meta-analysis. BMC Cancer. 2016;16:12.

49. Salomonsson A, Jönsson M, Isaksson S, et al. Histological specificity of alterations and expression of KIT and KITLG in non-small cell lung carcinoma. Genes Chromosomes Cancer. 2013;52:1088-96.

50. Qian J, Kong X, Deng N, et al. OCT1 is a determinant of synbindin-related ERK signalling with independent prognostic significance in gastric cancer. Gut. 2015;64:37-48.

51. Figueiredo C, Costa S, Karameris A, Machado JC. Pathogenesis of Gastric Cancer. Helicobacter. 2015;20 Suppl 1:30-5.

Figure Legends

Figure 1. Ideogram of genomic imbalance in 45 EGCs. Chromosomes are ordered from 1 to 22. The colored horizontal lines represent the frequencies of gains and LOHs and CNLOHs. Lines on the left indicate losses (red, copy neutral LOH; gray, LOH) and those on the right (green) indicate gain.

Figure 2. Ideogram of genomic imablance in 20 AGCs. Chromosomes are ordered from 1 to 22. The colored horizontal lines represent the frequencies of gains and LOHs and CNLOHs. Lines on the left indicate losses (red, copy neutral LOH; gray, LOH) and those on the right (green) indicate gains.

Figure 3. Representative images of early GC. A. An endoscopic picture shows an elevated lesion on the gastric body. B. An elevated lesion removed from the gastric body. C. Low power view of the lesion. D. High power view of the lesion shows a moderately differentiated adenocarcinoma. E. Ideogram showing copy number alteration. Green, gain; red, LOH; grey, copy neutral LOH.

Figure 4. Representative figure of advanced GC. A. Huge central ulcered lesion on the gastric body. B. Isolated tumor crypt. C. Differentiated adenocarcinoma of the isolated cancer gland. D. High power view of the lesion shows moderately differentiated adenocarcinoma. E. Ideogram showing copy number alteration. Green, gain; red, LOH; grey, copy neutral LOH.

Figure 5. Comparison of total lengths of regions in CNAs between EGCs and AGCs. A. The total length of CNA gains between early and advanced GCs. B. Total length of CNA LOH between early and advanced GCs. C. Total length of CNA copy neutral LOH between early and advanced GCs. D. Total length of overall CNA between early and advanced GCs.

Table 1. Clinicopathological findings of gastric cancer patients

	Total	EGCs	AGCs
Case	65	45	20
Sex (M/W)	47/18	34/11	13/7
Mean age (years)	73.2	73.2	73.2
(Range)	45-91	45-88	48-91
Location			
U	9	4	5
M	24	17	7
L	32	24	8
Macroscopy			
I or IIa	28	28	0
IIc	12	12	0
IIa+IIc or IIc+IIa	5	5	0
type 1	1	0	1
type 2	7	0	7
type 3	12	0	12
Histological type			
tub1	36	33	3
tub2	19	7	12
pap	10	5	5
pStage			
IA	44	44	0
IB	5	1	4
IIA	1	0	1
IIB	6	0	6
IIC	0	0	0
IIIA	2	0	2
IIIB	2	0	2
IIIC	3	0	3
IV	2	0	2

Table 2. Frequent regions of copy number alterations in early and advanced gastric cancers.

a

Chromosomal regions	EGC n=45 (%)
Gain	
8q23.3	19 (42.2)
8q23.2	18 (40)
8p22-23.1	17 (37.8)
9p13.1	17 (37.8)
8p11.21-11.23	16 (35.8)
8p12-21.3	16 (35.8)
8p23.2-23.3	16 (35.8)
8p11.1	15 (33.3)
8q24.3	14 (31.1)
9p23-24.1	14 (31.1)
17p13.3	14 (31.1)
8q24.21	14 (31.1)
LOH	
3p14.2	11 (24.4)
CNLOH	None

b

Chromosomal regions	AGC n=20 (%)	Chromosomal regions	AGC n=20 (%)
Gain		**LOH**	
8q24.21	13 (65)	11q24.3-25	10 (50)
8q24.3	12 (60)	11q23.2-24.1	10 (50)
1q24.2	11 (55)	11q14.1	10 (50)
8q24.22-23	11 (55)	12p11.21-13.33	10 (50)
8q24.13	11 (55)	3p14.2	7 (35)
20p12.2-13	11 (55)		
1q41-42.3	11 (55)	**CNLOH**	
1q31.1-32.3	10 (50)	11q21	10 (50)
1q24.3	10 (50)	11q13.3-14.3	10 (50)
8q24.12	10 (50)	11q11	10 (50)
8q22.3	10 (50)	11p13-15.3	10 (50)
10q24.32	10 (50)	12q21.1	10 (50)
10q23.2-23.33	10 (50)	12q12-13.3	10 (50)
11q25	10 (50)	5q33.3-35.1	10 (50)
11q23.1	10 (50)		
12q21.31-23.2	10 (50)		

Table 3. List of candidate genes that show copy number alteration in early and advanced gastric cancers

EGC

Chromosomal regions	Gene list
Gain	
8q23.3	*EIF3H, PSCA*
8q23.2	*EBAG9*
8p22-23.1	*FGF20, MSR1, CLDN23, TUSC3, TSP1*
9p13.1	*IGFBPL1*
8p11.21-11.23	*SFRP1, FGFR1, TACC1, BAG4, LSM1*
8p12-21.3	*FZD3, TNFRSF10A/B, PEBP4, LZTS1*
8p23.2-23.3	*CSMD1*
8p11.1	*DKK4, CEBPD*
8q24.3	*PSCA, PTP4A3, RECQL4, NDRG1*
9p23-24.1	*PDCD1LG1, TYRP1, PTPRD, KDM4C*
17p13.3	*HIC1, TUSC5, MIR22, YWHAE*
8q24.21	*Myc, POU5F1B, PVT1*
LOH	
3p14.2	*FHIT*
CNLOH	None

AGC

Chromosomal regions	Gene list
Gain	
8q24.21	*Myc, POU5F1B, PVT1*
8q24.3	*PSCA, PTP4A3, RECQL4*
1q24.2	*Oct1*
8q24.22-23	*WISP1*
8q24.13	*ATAD2, MTSS1*
20p12.2-13	*RASSF2, BMP2, CRCS11, CDC25B*
1q41-42.3	*CAPN9, WNT3A, TGFB2*
1q31.1-32.3	*ATF3, NEK2, MAPKAPK2, IL10, CD55*
1q24.3	*FASLG*
8q24.12	*OPG, TNFRSF11B*
8q22.3	*UBR5, FZD6, RRM2B*
10q24.32	*FGF8, BTRC, SUFU, CYP17A1*
10q23.2-23.33	*BMPR1A, PTEN, MINPP1, SNCG*
11q25	*NTM, OPCML*
11q23.1	*miR34B, IL18, COLCA1/2*
12q21.31-23.2	*IGF1, UTP20, ASCL1, BTG1*

Chromosomal regions	Gene list
LOH	
11q24.3-25	*ST14, NTM, OPCML*
11q23.2-24.1	*Synbindin, TTC12, ZBTB16, BLID*
11q14.1	*GAB2*
12p11.21-13.33	*KDM5A, KRAS, CDKN1B, TARBP2, CCND2*
3p14.2	*FHIT*
CNLOH	
11q21	*MMP7, MIR1260B, MRE11A*
11q13.3-14.3	*FGF4, CCND1, ORAOV1 FADD, GAB2*
11q11	*CTNND*
11p13-15.3	*CD44, WT1, TSG101, RRAS2, DKK3*
12q21.1	*RAB21*
12q12-13.3	*NUR77, PPHLN1, ITGA5, TARBP2*
5q33.3-35.1	*FBXW11, USP12*

Table 4. Significant difference in the frequencies of CNAs regions between early and advanced gastric cancers

Chromosomal regions	EGC n=45 (%)	AGC n=20 (%)	P value	Chromosomal regions	EGC n=45 (%)	AGC n=20 (%)	P value
Gain				**CNLOH**			
12q22-23.2	3 (6.7)	10 (50)	< 0.001	14q32.11-32.33	0 (0)	7 (35)	< 0.001
12q21.33	4 (8.9)	10 (50)	< 0.001	14q21.3	0 (0)	7 (35)	< 0.001
11p12	3 (6.7)	9 (45)	< 0.001	14q11.2	0 (0)	7 (35)	< 0.001
11p14.1	3 (6.7)	9 (45)	< 0.001	5q11.2	2 (4.4)	9 (45)	< 0.001
12q21.31-21.32	5 (11.1)	10 (50)	0.001	5q13.3	0 (0)	6 (30)	< 0.001
3p12.3	4 (8.9)	9 (45)	0.002	14q21.1-23.2	0 (0)	6 (30)	< 0.001
3p14.1	4 (8.9)	9 (45)	0.002	14q13.2-13.3	0 (0)	6 (30)	< 0.001
10p15.1	4 (8.9)	9 (45)	0.002	5q12.1-12.3	1 (2.2)	7 (35)	< 0.001
1q24.2	7 (15.6)	11 (55)	0.003	5q11.1	2 (4.4)	8 (40)	< 0.001
2q12.1	3 (6.7)	8 (40)	0.003	17p13.3	2 (4.4)	8 (40)	< 0.001
LOH							
8p12	0 (0)	4 (20)	0.01				
8p21.1	0 (0)	4 (20)	0.01				
8p21.2	0 (0)	4 (20)	0.01				
8p21.3	0 (0)	4 (20)	0.01				
8p23.1	0 (0)	4 (20)	0.01				
8p23.2	0 (0)	4 (20)	0.01				
8p23.3	0 (0)	4 (20)	0.01				

chi-square test

Figure 1

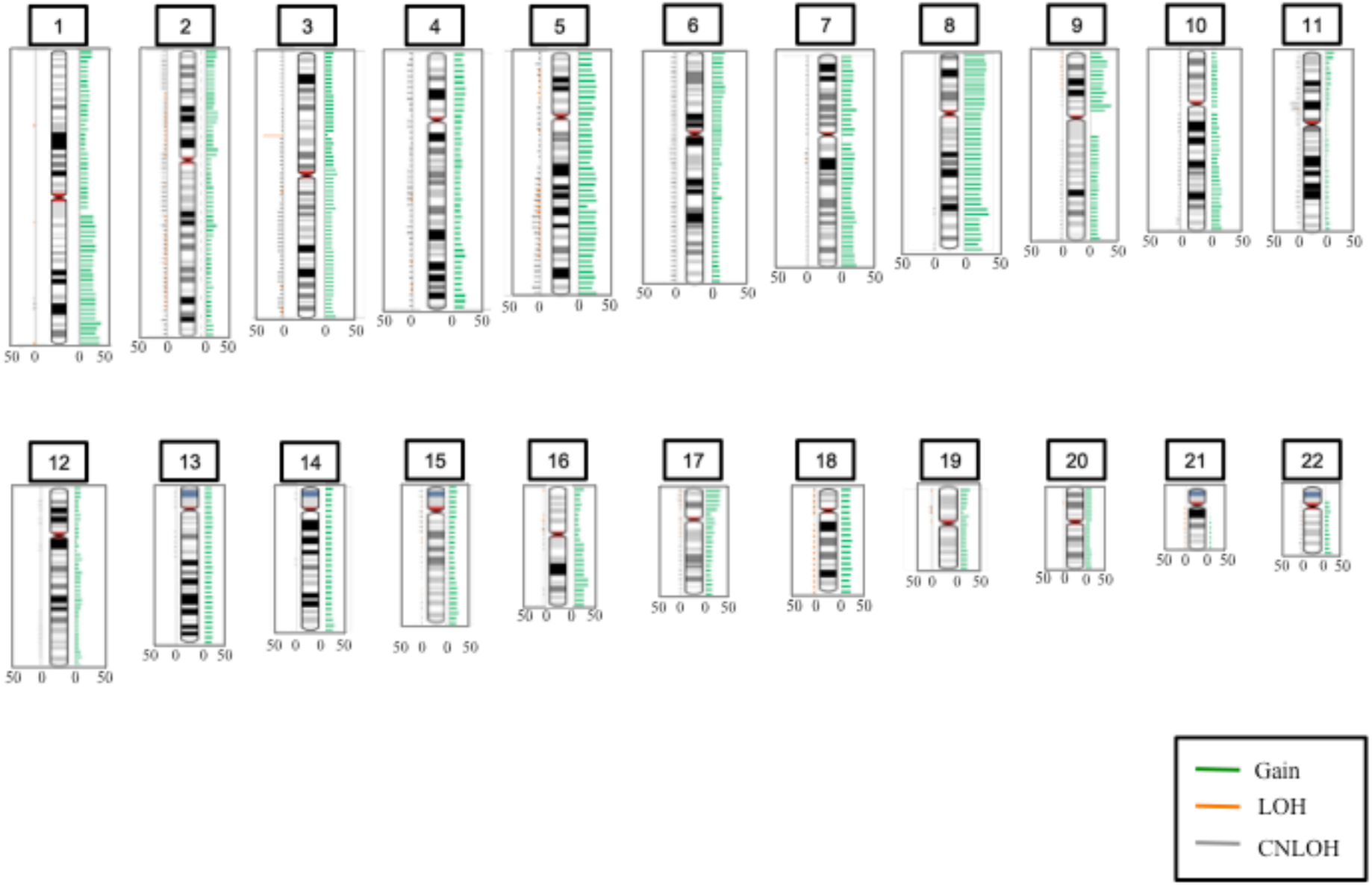

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
50 0 0 50
Gain
LOH
CNLOH

Figure 2

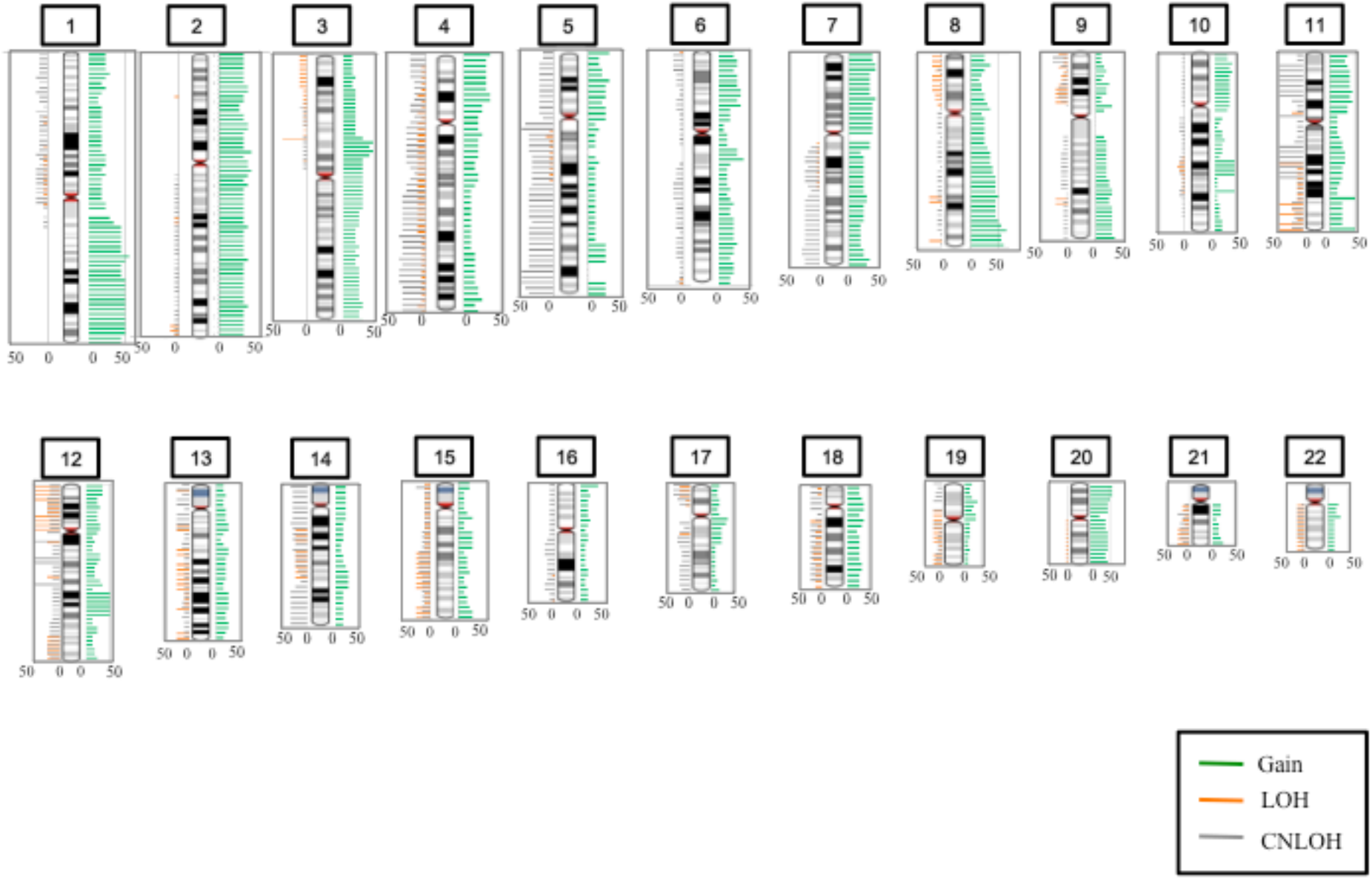
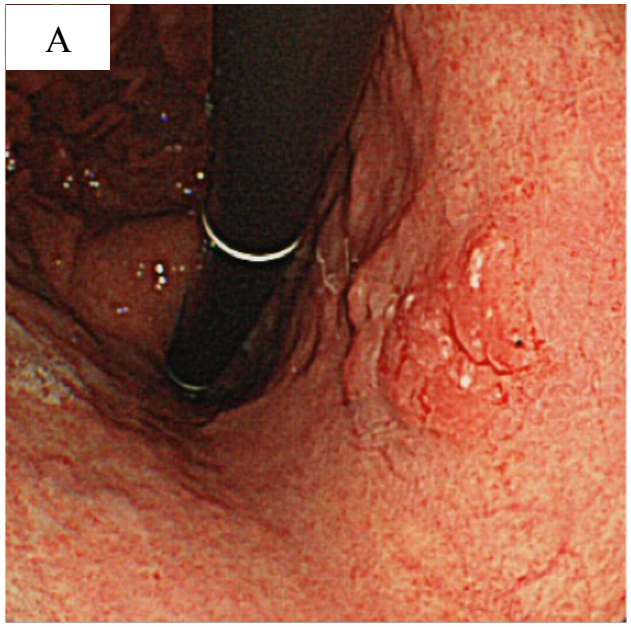
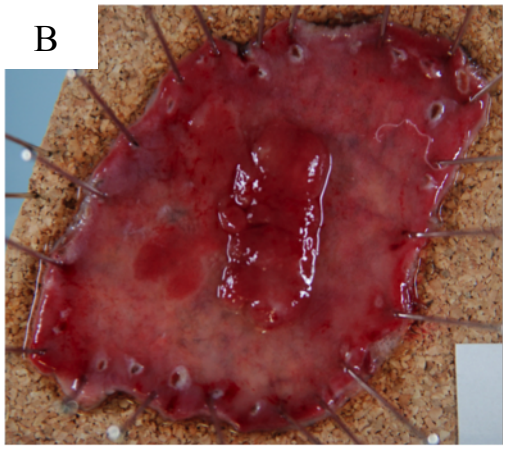
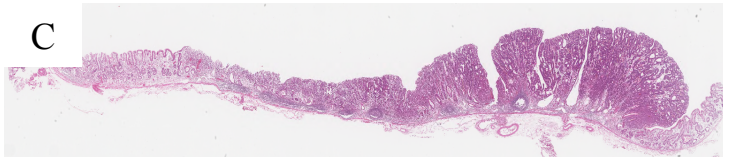
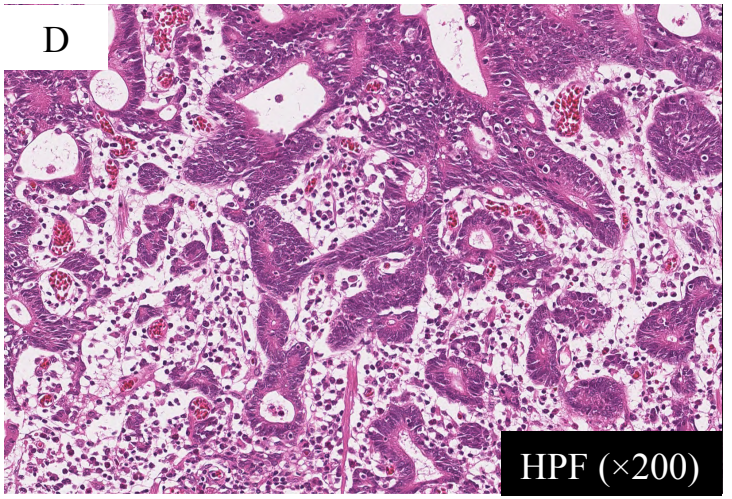

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
50 0 0 50
Gain
LOH
CNLOH

Figure 3


A


B


C


D
HPF (×200)

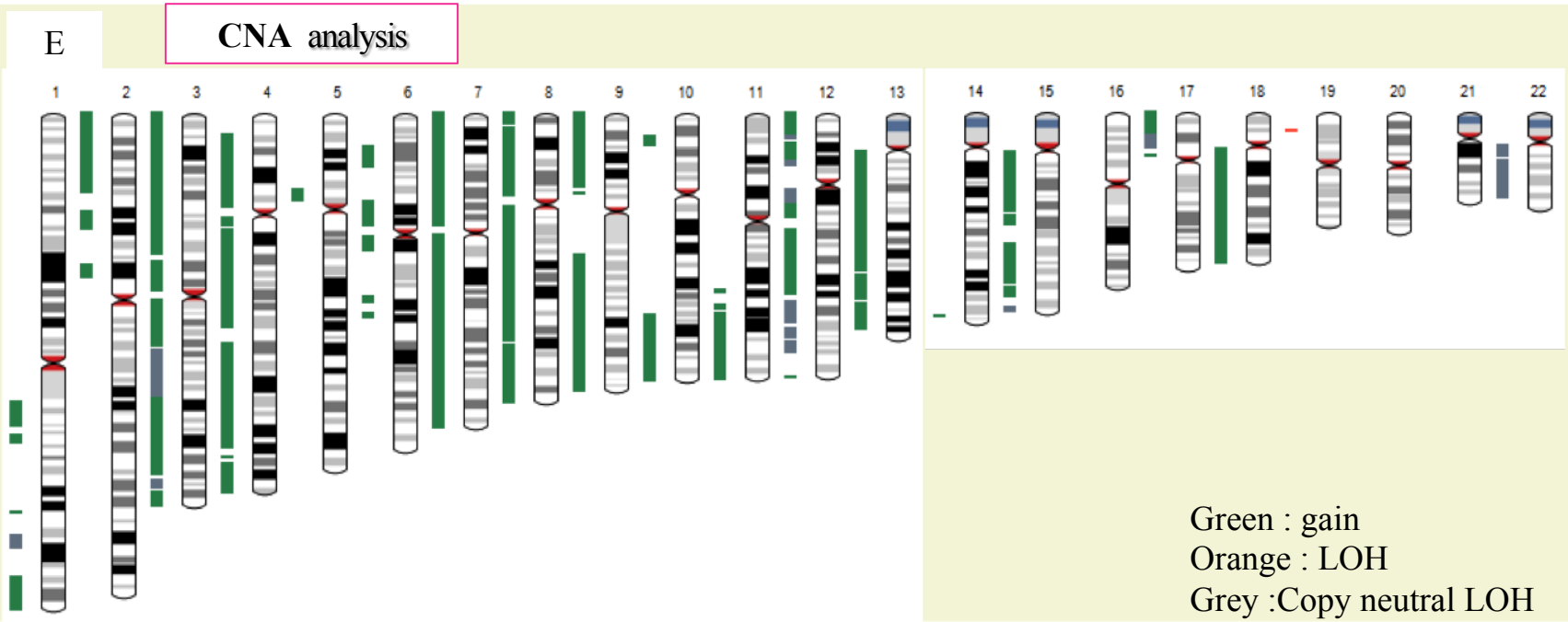
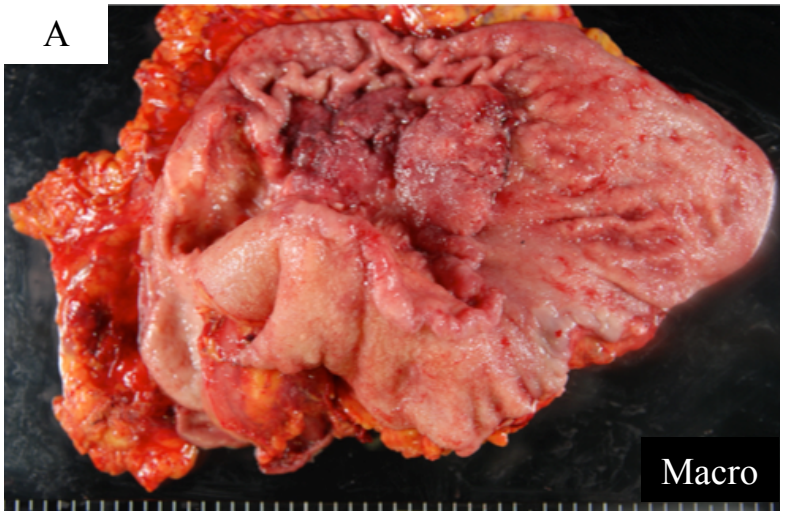
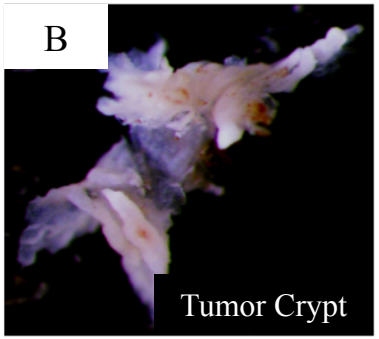
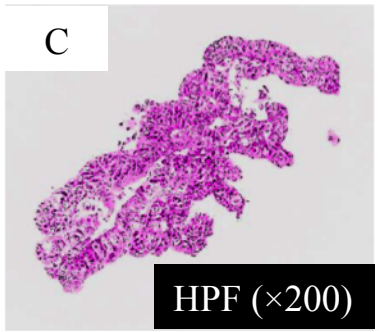
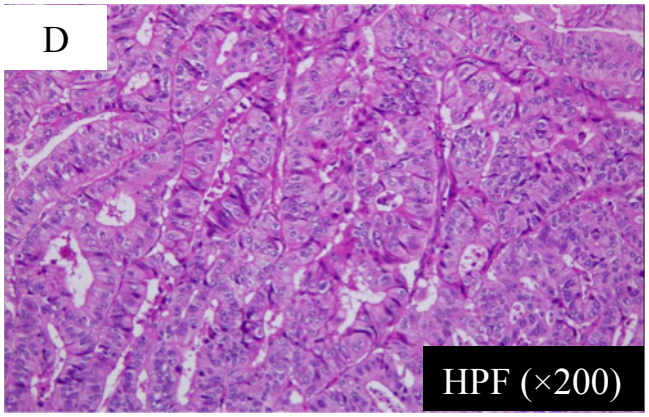

E
CNA analysis
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Green : gain
Orange : LOH
Grey :Copy neutral LOH

Figure 4


A
Macro


B
Tumor Crypt


C
HPF (×200)


D
HPF (×200)

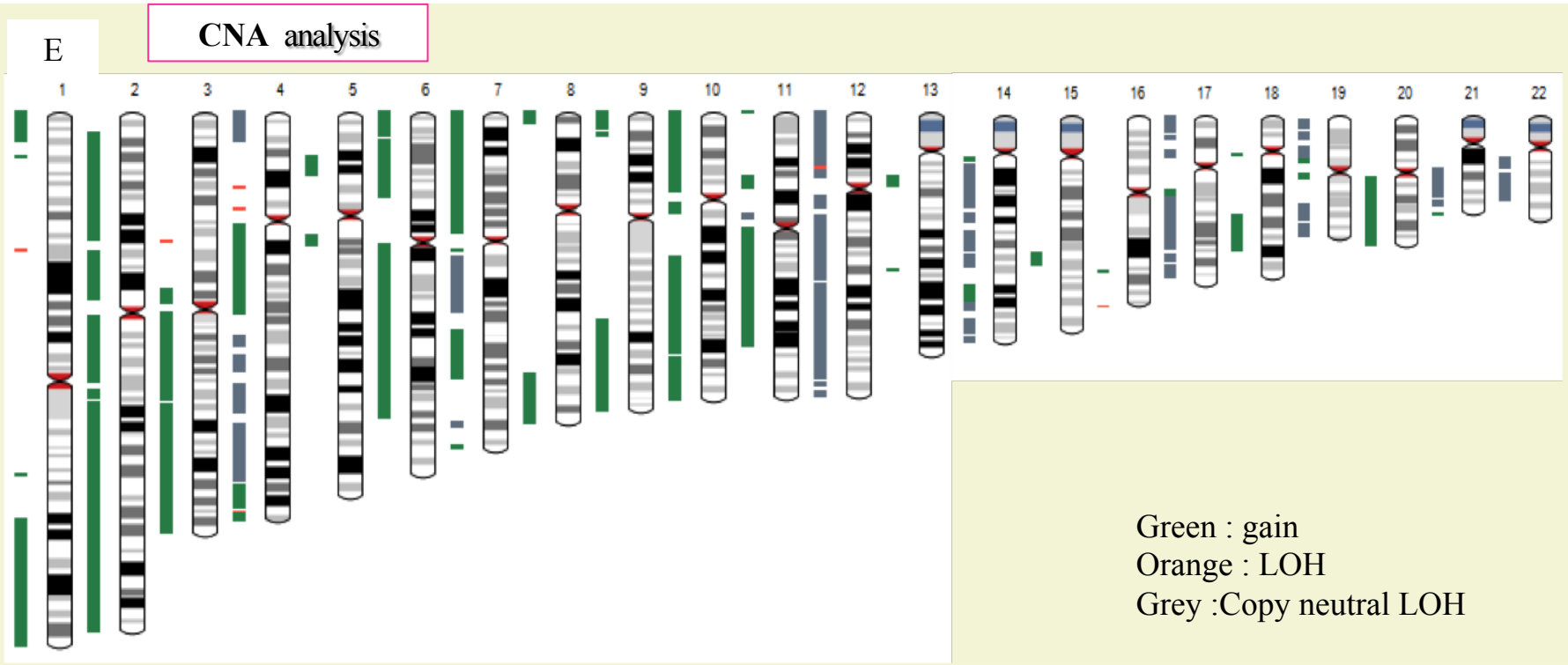

E
CNA analysis
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Green : gain
Orange : LOH
Grey :Copy neutral LOH

Figure 5. Comparison of total length of region in CNAs length between EGCs and AGCs.