Original article

Rapidly and slowly growing lineages in chromosomal instability-type gland-forming gastric carcinomas as revealed by multisampling analysis of DNA copy-number profile

Tu Thanh Duong[1,2], Diem Thi-Ngoc Vo[2], Takahisa Nakayama[1], Ken-ichi Mukaisho[1], Masamichi Bamba[3], Trung Sao Nguyen [2], Hiroyuki Sugihara[1]

1. Department of Pathology, Division of Molecular and Diagnostic Pathology, Shiga University of Medical Science
2. Department of Pathology, University of Medicine and Pharmacy at Ho Chi Minh City
3. Department of Pathology, Saiseikai Shiga Hospital, Imperial Gift Foundation Inc.

Correspondence: Professor Hiroyuki Sugihara

E-mail: sugihara@belle.shiga-med.ac.jp

Phone: +81-77-548-2168

Fax: +81-77-543-9880

Short running head: Loss-rich and gain-rich lineages in GC

Abstract

Background: To examine whether gastric carcinoma (GC) with chromosomal instability (CIN-type GC), the largest category in the Cancer Genome Atlas classification, consists of a single genetic lineage, we conducted a multisampling analysis of genomic DNA copy-number profile. ***Methods*****:** We performed array-based comparative genomic hybridization using formalin-fixed, paraffin embedded tissues from 54 gland-forming GCs containing a total of 106 DNA samples from mucosal, extra-mucosal invasive, and lymph node lesions. Microarray data were analyzed by unsupervised hierarchical clustering and penetrance plots. Epstein-Barr virus infection status and mismatch repair (MMR) enzyme silencing/p53/mucin expression were examined with *in situ* hybridization and immunohistochemistry, respectively. ***Results*****:** The samples examined were divided into gain-rich clusters A and loss-rich cluster B, which were different in tumor locus and patient age. The T1/T2–4 ratio, the frequency of small cancers (diameter ≤2–4 cm), and intestinal mucin expression were higher in cluster B than in cluster A, whereas there were no significant differences in the frequencies of MMR silencing, mutant p53 pattern, and lymph node metastasis between the two clusters. ***Conclusions*****:** We demonstrated that the CIN-type GC could be categorized into two genetic lineages which were different in rapidity of local extension but similar in nodal metastasis risk.

Keywords

Stomach, adenocarcinoma, chromosomal instability, copy-number alterations, array-based comparative genomic hybridization, metastasis risk

Introduction

Gastric carcinoma (GC) is the fifth most common malignant neoplasm and the third leading cause of cancer-related deaths in both sexes worldwide [1]. Recently, the Cancer Genome Atlas (TCGA) classification of GC includes four subtypes: Epstein-Barr virus (EBV)-positive GCs, microsatellite-unstable GCs, genomically stable GCs, and GCs with chromosomal instability (CIN) [2]. It remains unclear how the risk of progression from early to advanced stage or metastasis risk is assessable and whether the most common CIN-type GCs are genetically homogeneous.

Early GC is defined as a tumor that is limited to the mucosal and submucosal layers, if present, irrespective of the lymph node status. In Japan, early GCs account for approximately 40%–60% of all GC cases [3-5], most of which are detected by endoscopy. Endoscopic resection (ER) of early GCs, which has become the standard treatment in Japan, is gaining greater acceptance worldwide [6]. However, indications for ER of early GC are limited to

those with a very low risk of lymph node metastasis; gland-forming tumors ≤ 2 cm at clinical pT1a stage [7]. When the ESD specimen is pathologically assessed as non-curative, additional treatment follows to prevent recurrence. However, the subsequently removed gastric tissue frequently do not show any evidence of tumor spread or metastasis.

The extent to which early detection and treatment contribute to the reduction of mortality in these patients depends on the lineage continuity between the endoscopically resectable lesions and advanced cancer. In our previous studies, we applied genomic DNA copy-number alteration (CNA) profiling to precursor lesions as well as early and advanced GC specimens, and classified GC samples using unsupervised hierarchical cluster analyses. Using this approach, we have demonstrated that virtually all undifferentiated (diffuse) early GC cases were considered to become advanced [8], whereas around 20% of non-invasive (gland-forming) neoplasms were considered to eventually become invasive [9]. In the present study, we utilized this approach to multiple samples from mucosal, extra-mucosal invasive, and metastatic lymph node lesions of individual tumor specimens to confirm the consistency of individual-specific changes and to examine changes associated with tumor progression.

Patients and Methods

Materials

This study used formalin-fixed paraffin embedded tissue specimens from 57 invasive gastric adenocarcinomas including 22 intramucosal, 6 submucosal (including 1 collision cancer, counted as 2 distinct tumors), and 29 advanced cancers (including 2 double cancers, each counted as 2 tumors) (online suppl. Table 1). Of these tumors, 25 lymph node-positive (N+) and 22 node-negative (N0) tumors were surgically resected from 44 patients in the period between 1998 and 2014, and 10 mucosal GCs (cases M27 to M36; online suppl. Table 1) were removed by ESD during the 2012–2014 period. For diagnosis of N0 tumors, 10 or more lymph nodes examined had to be free from metastasis [10]. We performed multisampling, including mucosal, invasive, and if present, lymph node metastatic samples in all 27 advanced GC cases included in this study. The third edition of the Japanese Classification of Gastric Carcinoma and pTNM staging were used to determine histological characteristics and tumor stages, respectively [11].

Immunohistochemistry and EBV *in situ* hybridization

Immunohistochemical staining of 4-µm-thick paraffin sections was performed using an automated Ventana Discovery XT system (Tucson, AZ, USA) with heat pretreatment, an

amplification kit (Ventana, 760-080) and a DAB detection kit (Ventana, 760-124). The following two monoclonal antibodies against two mismatch repair (MMR) proteins were used to assess enzyme silencing, which is closely related to microsatellite instability (MSI) [12]: MSH6 (clone 44, Ventana) and PMS2 (clone EPR3947, Cell Marque, Rocklin, CA, USA). A monoclonal antibody to p53 protein (DO-7, 1:100; Dako, Glostrup, Denmark) was used to assess p53 expression pattern. Mucin phenotype was analyzed immunohistochemically using monoclonal antibodies against MUC2 (clone MRQ-18, Cell Marque, Rocklin, CA, USA), MUC5AC (clone MRQ-19, Cell Marque), MUC6 (clone MRQ-20, Cell Marque) and CD10 (clone 56C6, Dako, Glostrup, Denmark). The stains were scored according to the percentage of stained neoplastic cells and categorized into gastric (G), intestinal (I), null (N), G>I, and I>G phenotypes based on the previous study [9]. A polyclonal antibody to SEMA3E protein (1:500; Atlas Antibodies, Bromma, Sweden) to evaluate the expression of *SEMA3E* gene.

Loss of MMR enzyme expression (LOM) was defined as complete absence of tumor nuclear staining in specimens with retention of MMR enzyme expression (ROM) in nuclei of normal glands and lymphocytes (internal control) [12]. p53 staining pattern was classified as diffuse, regional, sporadic, and null. The former two and the last patterns were considered to reflect mutation, whereas sporadic pattern was considered as a wild-type (WT) pattern.

EBV status was determined by EBV-encoded small RNA (EBER) *in situ* hybridization [13] using the INFORM EBER Probe (Ventana, 800-2842: Mannheim, Germany).

Genomic DNA extraction

Genomic DNA was extracted from 5-µm-thick tumor and normal gland (reference) sections using laser microdissection (LMD6000; Leica Microsystems, Wetzlar, Germany). Samples were obtained from an area of ≥6 mm^2 in which ≥90% of the cells were neoplastic. A proteinase K solution (200 µg/ml) was used for digestion of the dissected tumor and reference samples for 70 ± 2 h at 37°C, followed by phenol/chloroform DNA extraction. DNA quality assessment was based on a cut-off A260/A280 ratio of >1.5, a cut-off A260/A230 ratio of >1.0, and concentration of double-stranded DNA.

Whole genome amplification

We used the GenomePlex whole genome amplification kit (WGA2 Kit; Sigma, St. Louis, MO, USA) for DNA amplification according to the manufacturer's protocol [14].

Array-based comparative genomic hybridization

For array-based comparative genomic hybridization **(**aCGH), 60-mer length oligonucleotide probes were used according to the manufacturer's guidelines [15]. From each tumor specimen, a set of genomic DNA samples, one from nonneoplastic glands, as a reference,

tumor and the others from tumor parts, were labeled using Cyanine 5 and Cyanine 3, respectively, prior to competitive microarray hybridization (SurePrint G3 CGH Microarray 8x60K, GPL10152 62,976 probes). Intensity of all hybridized probes were captured and qualified by a DNA microarray scanner (Feature Extraction software 10.7.3.1) followed by calculation of the ratio of tumor and reference fluorescence intensities. Next, chromosomal patterns within the microarray profiles were visualized, detected, and analyzed by the Agilent CGH analytic software using the UCSC Genome Browser according to the latest resource content: hg19 assembly-Design ID 021429 (GRCh Build 37). Definition of genomic copy number gain and loss, and amplification were based on base 2 logarithm of the tumor-to-reference (T/R) ratios which were >0.3219, <−0.3219, and >1, respectively. The microarray data were registered in the Gene Expression Omnibus (GEO) data base (Accession number: GSE108507).

Clustering algorithm

Before performing the cluster analysis, average T/R ratio of the probes within each gene was calculated to intensify the signal-to-noise ratio in hybridization analysis. Given that samples from the same tumor might share random generation of gene CNAs and tissue environmental selection, different parts of a tumor might have similar CNA profiles in the presence of a common carcinogenetic process [21, 22]. CNA profiles that have a low noise level could

satisfy this internal standard. Gene size, which is defined as the number of probes in each gene, impacts noise-canceling function of the average of T/R ratio, whereas clustering reproducibility depends on the number of genes. For smaller gene numbers, the reproducibility of clustering analysis becomes lower, whereas gene size is larger. Thus, for determining optimal gene size and number, the clustering analysis was repeatedly performed using the Clustering 3.0 (version 1.52) software and TreeView (version 1.1.6r2) [29, 30]. The unsupervised clustering analysis was based on genomic copy number profile resemblance, and complete linkage and uncentered correlation distance were applied.

Genes showing significantly different CNAs between clusters

Welch's *t* test was used between average $\log_2$(T/R) values of total A cluster and total B cluster samples for 14,977 protein-coding genes. Bonferroni correction was used to correct for multiple comparisons.

Statistical analysis

We used Microsoft Office Excel 2013 and BellCurve for Excel (Social Survey Research Information Co., Ltd., Tokyo, Japan). Correlations among variables were examined using Fisher's exact test and Welch's *t* test. Statistical significance was determined at a *p* value of less than 0.05.

Results

Unsupervised clustering analysis of the CNA profile

To improve the CNA signal-to-noise ratios, probe T/R ratios within a specified gene were averaged (T/R ratios of 55,142 probes in 30,471 gene regions). To classify the samples solely by the similarity in their CNA profile, unsupervised cluster analysis was performed. To determine the optimal clustering condition, we repeated the clustering nine times using varying gene sizes, from 9,487 genes with ≥2 probes to 370 genes with ≥10 probes, [9] and correlated the neighboring or splitting sample distribution in the dendrogram with the presence of large chromosomal changes common to all samples for each case, which we defined as stemline changes. For this purpose, we prepared penetrance plots of individual samples (online suppl. Fig. 1). Of these nine conditions, we chose the condition including genes with ≥5 probes because the neighboring sample distribution in the dendrogram was best correlated with the presence of stemline changes. In the chosen condition, out of the 27 cases included in the multisampling, 21 cases showed a neighboring pattern and 6 cases showed a splitting pattern in the sample distribution (Fig. 1). Of the 21 neighboring pattern cases, stemline chromosomal changes were detected in 17 cases with ROM and not in 4 cases with LOM (online suppl. Fig. 1). Of the six splitting pattern cases included two double

cancers (case# A13 and A17), three cancers with almost no common chromosomal changes in the samples (case# S5, A15, and A29), and one cancer with stemline changes but splitting between the samples with and without 1p/q+ (case# A7).

Clustering results and their relationship to MMR enzyme expression, p53 expression, EBV infection, and mucin phenotype

Consequently, a total of 106 samples were classified into two major clusters: A and B (Fig. 1). The heat map indicated that there was a copy-number gain area common to both clusters. Clusters A and B were characterized by the presence of gain-rich and loss-rich areas, respectively. LOM samples characteristically showed dark in the heat map. There were no differences in the frequencies of LOM and mutant p53 pattern between the two clusters. The LOM samples were not distributed in the clusters. Additionally, we confirmed that MMR enzyme silencing and mutant p53 pattern were mutually exclusive (Fig. 1). The tumors we examined in this study contain only 1 case with EBV infection. The staining results of mucin phenotype are shown in online suppl.Table 1. We found that the relative frequency of gastric predominant phenotype (G, G≥I)/null type (N) was significantly lower in invasive/metastatic parts than in mucosal parts, whereas intestinal predominant phenotype (I, I>G) was almost constant between these parts (online suppl. Fig. 2). Therefore, we can tentatively regard the intestinal expression as a progression-independent lineage marker. This marker expression

was more common in cluster B than in cluster A.

Chromosomal CNAs

The penetrance plots of the tumors with LOM demonstrated infrequent gains (such as 12q+, which was rare in the ROM samples) and scarce losses, whereas gains and losses were conspicuous in the ROM tumors (Fig. 2a, b), which were divided into two patterns by clustering (Fig. 2c, d). The changes common to both clusters were 8q+, 13q+, 20q+, and 5q−. Clusters A and B were characterized by gain-rich (7p/q+, Xp/q+, etc.) and loss-rich (4p/q−, etc.), respectively (Fig. 2c, d). Statistical analyses are shown in Table 1a. The mucosal, invasive, and metastatic ROM samples were compared in clusters A and B (Fig. 3). The changes common to clusters A and B (8q+, 13q+, 20q+, and 5q−) were already present in the mucosal samples, whereas cluster-specific changes (such as 4p/q− and 7p/q+) accumulated during the progression to invasion and metastasis. These were confirmed by statistical analysis of the frequency of each chromosomal CNA, which were determined by counting using the penetrance plot of each sample (online suppl. Fig. 1) and compared between the mucosal and extra-mucosal (invasive and metastatic) samples (Table 1b). In contrast, such differences were not detected between the invasive and metastatic samples. Table 1c shows metastasis-related chromosomal changes. The chromosomal changes common to clusters A and B were not or weakly related to metastasis, whereas the cluster specific changes, such as

7p/q+ in cluster A and 3p−, 4p−, 5q−, and 8p− in cluster B were significantly related to lymph node metastasis. The chromosomal changes detected are summarized as an evolutionary tree in Fig. 4.

Clinicopathological and molecular characteristics of the clusters

Clinicopathological characteristics and molecular changes were compared between the tumors of clusters A and B (Table 2). The tumor locus and patient age were different between clusters A and B. In this tumor-based analysis, representative phenotype in each submucosal or advanced cancers was defined to that of the invasive sample, and we could also confirm that the intestinal expression was more frequent in cluster B than in cluster A.

The T1/T2–4 ratio and the frequency of small cancers (diameter ≤ 2–4 cm) were higher in cluster B than in cluster A, suggesting that the GCs in cluster A is more rapidly growing than in cluster B. Nine of the 10 ESD specimens examined were classified into cluster B. There were however no significant differences in the frequencies of lymph node metastasis.

Genes showing significantly different CNAs between the clusters

We identified 32 genes that significantly contributed to the difference in clusters A and B. Of these genes, 12 were related to growth, as shown in online suppl. Fig. 2. When the gene function and the direction of CNAs were consistent, we regard the CNAs as putative driver

and when inconsistent, we regard the CNAs as putative passenger. There were five putative driver tumor suppressor genes (*CACNA2D3*, *PTPRG*, and *LRIG1* at chromosome 3p, *SLIT2* at chromosome 4p, and *FSTL5* at chromosome 4q) and one possible driver protooncogene (semaphorin 3E [*SEMA3E*]) at 7q. However, immunohistochemically, there was no difference in expression level of SEMA3E protein between 10 tumors that showed greatest copy-number gains and those with greatest copy-number losses.

To compare the growth activity between cluster A and cluster B tumors, we examined the frequency of amplification of growth-related genes, which is a characteristic of CIN-type GC. Of the 55 receptor tyrosine kinases (RTKs) examined, copy-number gains of 12 RTKs, including EGFR, ERBB3, FGFR4 and EPHB3 and amplification of LMTK2, EPHB3, EPHB4 showed significantly different frequency between clusters A and B, and all of them were more frequently detected in cluster A than in cluster B (online suppl. Table 2).

Discussion

Using the multisampling method, we could demonstrate not only the differences among samples but also the reproducibility of changes common to all samples in individual GC patients. To assess the reproducibility of this approach as well as to classify samples, we applied unsupervised, hierarchical clustering analysis to the multisampling aCGH data of 57

early and advanced gland-forming GC specimens. We selected the clustering condition that showed the highest concordance between the neighboring pattern in the clustering dendrogram and the presence of stemline changes. Even in the absence of stemline changes among the samples from individual tumors such as the LOM tumors, the mucosal and invasive/metastatic samples of individual tumors often showed a neighboring position in the dendrogram, likely reflecting a similarity among gene-level CNAs. CNA profiles are considered as individual tumor-specific and progression-independent lineage markers.

During development of individual tumors, random changes as well as essential genetic and epigenetic changes accumulate in a time-dependent manner based on genetic instability, and these changes undergo natural selection by the tissue environment. In this evolutionary process, genomic alteration profiles might become unique to each individual tumor, which might converge in several types by natural selection [16]. We attempted to reveal such converged genotypes using unsupervised hierarchical clustering.

The tumors included in the present analyses were mostly CIN-type and partly MSI-type of TCGA classification, whereas EBV-related GCs were infrequent. A cancer with LOM, which largely overlaps with the MSI phenotype, is characterized by a low frequency of chromosomal changes and the presence of the unique 12q+ (Fig. 2a, b), and may have branched off from the main trunk of gland-forming GCs at earlier stages of GC development

(Fig. 4). Accordingly, it was reported that GCs of MSI phenotype showed the frequency of chromosomal CNAs lower than in GCs without MSI and were associated with high DNA methylation status even at intramucosal stages [17].

We classified ROM GC samples (GCs excluding LOM) into two clusters, A (ROM) and B (ROM) (Fig. 2c, d). Based on the resemblance of gene-level CNA profiles, our clustering approach have successfully disclosed two distinct genetic lineages (gain-rich and loss-rich lineages) that have not been clearly recognized thus far. In the present study, we analyzed sequential accumulation of chromosomal CNAs during progression from the mucosal to the invasive/metastatic growth in each lineage (Fig. 3). Both lineages might have derived from the common trunk with 8q+, 13q+, 20p/q+, and 5q−,which were commonly detected in both clusters A and B. These findings partly confirm the report of Uchida *et al.*, who reported 8p+ and 5q− as early changes of GC [18]. Then, cluster-specific later changes, such as 7pq+ in cluster A and 4pq− in cluster B (Fig. 4).

Next, we classified all samples into two clusters, A and B. There appeared little relationship between this cluster classification and TCGA GC subtypes (Table 2).The tumor locus and patient age were different between clusters A and B. The age difference might be explained by the difference in the frequency of early GCs, whereas the locus difference suggests that this clustering might reflect distinct genetic lineages. This was also supported

by mucin phenotyping; progression-independent intestinal expression was significantly different between clusters A and B. Comparison of other clinicopathological factors between clusters A and B indicated that the tumors in cluster A was more rapidly growing (larger, deeper tumors) than those in cluster B. ESD specimens were mostly included in the latter. These finding implied that early detection and treatment were more difficult in cluster A than in cluster B tumors (Fig. 5).

A rapid growth in the gain-rich lineage (cluster A) might be partly due to the copy-number gains/amplifications of the RTKs mentioned above and growth-related genes in chromosome 7, including *RalA* [19], *EGFR* [20], *MAFK* [21], *GLI3* [22], *CUL1* [23], and *SEMA3E* [24, 25] (online suppl. Table 2). Among these, only *SEMA3E* exhibited significant copy-number differences after Bonferroni correction between clusters A and B. However, its expression is reportedly enhanced [24] or suppressed [25] in GCs. Our immunohistochemical studies for *SEMA3E* failed to demonstrate significant association between protein expression level and gene copy number, and we cannot regard this gene as functional driver or suppressor. Rather, detection of immunoreactivity for amplified growth-related genes may be more promising as cluster A markers. In the loss-rich lineage (cluster B), the above-mentioned tumor suppressor genes on 3p and 4p/q showed significant differences between clusters A and B. In particular,

silencing of *CACNA2D3* [26] and *PTPRG* [27] were reported to play a role in gastric tumorigenesis.

There was no significant differences in the lymph node metastasis risk between the rapidly and slowly growing lineages, whereas there were lineage-specific, invasion/metastasis-related chromosomal changes. These findings suggest that metastasis risk is determined not at an earlier but at a later stage during tumor development as shown in Fig. 4.

These conclusions have several clinical implications for early detection and treatment strategies. This study was conducted with a null hypothesis that there are indolent cancers treated as cancers, which are associated with very low mortality risk. We previously demonstrated that high-grade and low-grade gastric adenomas, of which the former is treated as carcinoma in Japan, were such tumors. Only around 20% of these tumors were inferred to progress to overt GC [9] (Fig. 5). The present study showed that there was no inherently indolent CIN-type early GC cluster and that the two lineages demonstrated in the present study had equal potential for metastasis. Our findings also demonstrated that GCs detected at an early stage were biased to a slowly growing loss-rich lineage. The prevalent use of early detection by endoscopy and subsequent treatment contribute to the reductions in the incidence of advanced GC and age-adjusted mortality rates; however, these reduction sizes were smaller than expected compared with the reduction in the GC morbidity rate [28]. One

major cause might be the increase in the percentage of older patients. Another factor might be the failure of early detection of GCs with a rapidly growing gain-rich lineage, which remains a challenge that should be addressed in future studies.

Acknowledgements

The authors thank Associate Professor Suzuko Moritani and Professor Ryoji Kushima, Department of Pathology, Shiga University of Medical Science Hospital, for kind supports for EBV in situ hybridization.

This study was supported in part by JSPS KAKENHI Grant Numbers JP25460454 and JP16K08689.

Statement of Ethics

All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. The Institution Review Board on Medical Ethics at Shiga University of Medical Science granted permission for conducting this study (Permission number: 30-021). A substitute of written informed consent was obtained from all patients included in the study.

Conflict of Interests

The authors declare that they have no conflict of interest.

References

1. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C RM, Parkin DM, Forman D, Bray F. F.GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11 [Internet] Lyon, France: International Agency for Research on Cancer;2013 [cited 2016 21 Nov]. Available from: http://globocan.iarc.fr.
2. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513:202-9.
3. Everett SM, Axon AT. Early gastric cancer in Europe. Gut. 1997;41:142-50.
4. Shimizu S, Tada M, Kawai K. Early gastric cancer: its surveillance and natural course. Endoscopy. 1995;27:27-31.
5. Tanaka M, Ono H, Hasuike N, Takizawa K. Endoscopic submucosal dissection of early gastric cancer. Digestion. 2008;77(1 Suppl):23-8.
6. Gotoda T. Endoscopic resection of early gastric cancer. Gastric cancer. 2007;10:1-11.
7. Japanese gastric cancer treatment guidelines 2010 (ver. 3). Gastric cancer. 2011;14:113-23.

8. Sonoda A, Mukaisho K, Nakayama T, Diem VT, Hattori T, Andoh A, et al. Genetic lineages of undifferentiated-type gastric carcinomas analysed by unsupervised clustering of genomic DNA microarray data. BMC Med Genomics. 2013;6:25.

9. Vo DT, Nakayama T, Yamamoto H, Mukaisho K, Hattori T, Sugihara H. Progression risk assessments of individual non-invasive gastric neoplasms by genomic copy-number profile and mucin phenotype. BMC Med Genomics. 2015;8:6.

10. Borie F, Plaisant N, Millat B, Hay JM, Fagniez PL. Appropriate gastric resection with lymph node dissection for early gastric cancer. Ann Surg Oncol. 2004;11:512-7.

11. Japanese Gastric Cancer A. Japanese classification of gastric carcinoma: 3rd English edition. Gastric cancer. 2011;14:101-12.

12. Mojtahed A, Schrijver I, Ford JM, Longacre TA, Pai RK. A two-antibody mismatch repair protein immunohistochemistry screening approach for colorectal carcinomas, skin sebaceous tumors, and gynecologic tract carcinomas. Mod Pathol. 2011;24:1004-14.

13. Okamoto A, Yanada M, Miura H, et al. Prognostic significance of Epstein–Barr virus DNA detection in pretreatment serum in diffuse large B‐cell lymphoma. Cancer Sci. 2015;106:1576-1581.

14. Little SE, Vuononvirta R, Reis-Filho JS, Natrajan R, Iravani M, Fenwick K, et al. Array CGH using whole genome amplification of fresh-frozen and formalin-fixed, paraffin-embedded tumor DNA. Genomics. 2006;87:298-306.

15. Barrett MT, Scheffer A, Ben-Dor A, Sampas N, Lipson D, Kincaid R, et al. Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. Proc Natl Acad Sci USA. 2004;101:17765-70.

16. Gleaves M. Cancer: The Evolutionary Legacy. Oxford University Press; 2001.

17. Sugai, T, Eizuka M, Arakawa N, Osakabe M, Habano W, Fujita Y, et al. Molecular profiling and comprehensive genome-wide analysis of somatic copy number alterations in gastric intramucosal neoplasias based on microsatellite status. Gastric Cancer 2018;21:765-775.

18. Uchida M, Tsukamoto Y, Uchida T, Ishikawa Y, Nagai T, Hijiya N, et al. Genomic profiling of gastric carcinoma in situ and adenomas by array‐based comparative genomic hybridization. J Pathol. 2010;221:96-105.

19. Martin TD, Samuel JC, Routh ED, Der CJ, Yeh JJ. Activation and Involvement of Ral GTPases in Colorectal Cancer. Cancer Res. 2011;71:206-15.

20. Oh HS, Eom D-W, Kang GH, Ahn YC, Lee SJ, Kim J-H, et al. Prognostic implications of EGFR and HER-2 alteration assessed by immunohistochemistry and silver in situ

hybridization in gastric cancer patients following curative resection. Gastric cancer. 2014;17:402-11.

21. Okita Y, Kimura M, Xie R, Chen C, Shen LT-W, Kojima Y, et al. The transcription factor MAFK induces EMT and malignant progression of triple-negative breast cancer cells through its target GPNMB. Sci Signal. 2017;10:474.

22. Iwasaki H, Nakano K, Shinkai K, Kunisawa Y, Hirahashi M, Oda Y, Onishi H, Katano M. Hedgehog Gli3 activator signal augments tumorigenicity of colorectal cancer via upregulation of adherence-related genes. Cancer Sci. 2013;104:328-36.

23. Bai J, Zhou Y, Chen G, Zeng J, Ding J, Tan Y, et al. Overexpression of Cullin1 is associated with poor prognosis of patients with gastric cancer. Hum Pathol. 2011;42:375-83.

24. Maejima R, Tamai K, Shiroki T, Yokoyama M, Shibuya R, Nakamura M, Satoh K. Enhanced expression of semaphorin 3E is involved in the gastric cancer development. Int J Oncol. 2016;49:887-894.

25. Chen H , Xie GH , Wang WW , Yuan XL , Xing WM , Liu HJ, et al. Epigenetically downregulated Semaphorin 3E contributes to gastric cancer. Oncotarget. 2015;6:20449-65.

26. Wanajo A, Sasaki A, Nagasaki H, Shimada S, Otsubo T, Owaki S, et al. Methylation of the calcium channel-related gene, CACNA2D3, is frequent and a poor prognostic factor in gastric cancer. Gastroenterology. 2008;135:580-90.

27. Wu CW, Kao HL, Li AF, Chi CW, Lin WC. Protein tyrosine-phosphatase expression profiling in gastric cancer tissues. Cancer Lett. 2006;242:95-103.

28. Tani M ST, Nakanish Y, Ochiai A, Taniguchi H, Sasako M, et al. Chronological trends of gastric carcinoma in Japanese from the pathological view point. Stomach and Intestine (Tokyo). 2005;40:27-36.

Figure legends

Fig. 1: Unsupervised clustering analysis of 106 samples using the genes/markers of ≥5-probe size. Genomic copy-number gains and losses of genes/markers are indicated by red and green squares, respectively in the clustering heat map. The clustering results are related to tumor size (gray squares: ≤2 cm), presence of lymph node metastasis (marked with green background of sample names), Mutant pattern (M, purple) of p53 immunohistochemistry (WT; Wild type), loss of mismatch repair enzyme expression (LOM, pink), and EBV infection status, and mucin phenotype (I/G/N: intestinal/gastric/null). Intestinal predominant (I, I > G) and gastric predominant (G, G ≥ I) phenotypes are marked with yellow and brown background, respectively.

Fig. 2: Frequency of copy-number alterations at the chromosome level (penetrance plots) compared between the tumors with and those without mismatch repair enzyme silencing and between clusters A and B. Statistically significant differences are marked with blue frames. **a**, **b**. LOM: loss of mismatch repair enzyme expression; ROM: retention of mismatch repair enzyme expression. The sample numbers of LOM and ROM are 14 and 92, respectively. **c**, **d**. Clusters A and B consist of 41 and 51 samples, respectively.

Fig. 3: Penetrance plots of ROM tumors in clusters A and B. Cluster-specific, progression-related significant changes are marked with blue frames. **a**, **d.** Mucosal samples; **b**, **e.** extra-mucosal invasive samples; **c**, **f.** lymph node metastasis samples. The sample numbers of **a** to **f** are 18, 15, 8, 31, 12, and 8, respectively.

Fig. 4: Evolutionary tree of gland-forming gastric carcinoma reconstructed from the data of Table 1, and Figs 2 and 3. LOM/ROM: loss/retention of mismatch repair enzyme expression. Chromosomal gains and losses are shown as red and green characters. Bold/normal letters indicate frequent/infrequent significant changes. Chromosomal parts in parentheses indicate insignificant but characteristic changes.

Fig. 5: Progression model of gland-forming gastric neoplasm from intramucosal neoplasm to invasive carcinoma.

Online suppl. Table 1: Clinicopathological features of individual tumors (age, gender, locus, size of mucosal lesion, pT, pN, and excision type) and immunohistochemical and in situ hybridization results (p53, MMR enzymes, EBV, mucin markers) of individual samples.

(xlsx28KB)

Online suppl. Table 2: Frequencies of copy-number gains and amplifications of receptor tyrosine kinases and representative growth-related genes on chromosome 7 in clusters A and B. (xlsx15KB)

Online suppl. Fig. 1: Penetrance plots of individual samples. The sample orders in clusters A and B and LOM corresponds to the sample order in Fig. 1. LOM: loss of mismatch repair enzyme expression; ROM: retention of mismatch repair enzyme expression. (pdf282KB)

Online suppl. Fig. 2: Comparison of mucin phenotype composition between mucosal (M) and invasive/metastatic (I + LN) samples in the cluster A and the cluster B tumors. Based on the online suppl. Table 1, mucin phenotype was classified into gastric predominant phenotype (G, G≥I), intestinal dominant phenotype (I, I>G), and unclassifiable phenotype (N). **a.** mucin phenotype composition of M and I + LN parts in clusters A and B. **b, c**. Statistical analyses of the differences in frequency of gastric predominant expression (b) or intestinal predominant expression (c) between M and I + LN parts and between clusters A and B. (pdf389KB)

Online suppl. Fig. 3: Outline of 32 genes that showed significantly different copy-numbers

between clusters A and B. The mean copy number alterations (CNAs) are expresses as Tumor/Reference (T/R) signal intensity ratio. In the gene function column, tumor suppressor genes and proto-oncogenes are marked with green and pink background, respectively. In the mean T/R ratio columns, copy-number gains and losses are marked with pink and green background, respectively. Concordant pairs of CNA and gene function are marked with black frame. (xlsx15KB)

Table 1: Differences in frequencies of chromosomal copy-number changes between clusters A and B (a), between mucosal and invasive/metastatic samples (b), and between N0 and N1-3 (c) in ROM tumors.

	a			b			c		
	Number of samples		*p value*	Number of samples		*p value*	Number of samples		*p value*
	Cluster A (*n = 41*)	Cluster B (*n = 51*)		M (*n = 53*)	I+LN (*n = 39*)	M vs I+LN	N1-3 (*n = 51*)	N0 (*n = 41*)	N1-3 vs N0
3p−	7	14	0.3192	6	15	**0.0028**	17	4	**0.0114**
3q+	10	6	0.1660	6	10	0.0967	11	5	0.2792
4p−	4	12	0.1020	5	11	**0.0260**	14	2	**0.0097**
4q−	2	14	**0.0051**	8	8	0.5820	13	3	**0.0275**
5p+	12	8	0.1339	8	12	0.0805	14	6	0.2035
5q−	8	16	0.2374	11	13	0.2306	19	5	**0.0083**
7p+	23	7	**<0.0001**	9	21	**0.0003**	24	6	**0.0015**
7q+	12	2	**0.001**	3	11	**0.0064**	13	1	**0.0025**
8p−	6	16	0.0853	7	15	**0.0067**	19	3	**0.0011**
8q+	19	23	1.0000	23	19	0.6746	25	17	0.5309
9p−	7	17	0.0969	9	15	**0.0300**	18	6	**0.0320**
10p+	12	7	0.0763	6	13	**0.0175**	16	3	**0.0047**
13q+	15	18	1.0000	15	18	0.0846	23	10	0.0501
14q−	0	6	**0.0316**	2	4	0.3955	5	1	0.2202
15q+	11	6	0.1033	6	11	0.0565	12	5	0.1877
16p+	11	3	**0.0077**	6	8	0.2525	11	3	0.0806
17p−	2	10	0.0592	6	6	0.7554	9	3	0.2144
18q−	7	10	0.7938	6	11	0.0565	12	5	0.1877
19p−	2	12	**0.0181**	5	9	0.0851	10	4	0.2486
19q−	1	4	0.3764	1	4	0.1587	2	3	0.6528
20q+	27	29	0.3996	28	28	0.0847	39	17	**0.0011**
22q-	0	11	**0.0009**	4	7	0.1934	8	3	0.3343
Xp+	18	5	**0.0002**	8	15	**0.0147**	19	4	**0.0032**
Xq-	1	5	0.2202	1	5	0.0796	6	0	**0.0316**
Xq+	19	8	**0.0024**	10	17	**0.0121**	20	7	**0.0231**

ROM: retention of mismatch repair enzyme expression. *p values* <0.05 are shown in bold.

Table 2: Differences in clinicopathological factors between clusters A and B.

57 tumors	Cluster A	Cluster B	Total	*p value*
Gender*				
- Male	12	22	34	0.5451
- Female	8	9	17	
Age* (mean ± SD)	75.35 ± 8.65	68.55 ± 7.24		**0.006**
Tumor locus				
- Upper (U)	3	9	12	**0.0038**
- Middle (M)	7	19	26	(U+M vs L)
- Lower (L)	13	6	19	
Tumor size				
≤ 2cm	1	14	15	**0.0019**
>2cm	22	20	42	
≤ 3cm	5	19	24	**0.0143**
>3cm	18	15	33	
≤ 4cm	9	26	35	**0.0062**
>4cm	14	8	22	
≤ 5cm	14	27	41	0.1454
>5cm	9	7	16	
Local extension				
T1	5	23	28	**0.0011**
T2-4	18	11	29	
Nodal lymph node status				
N0	12	23	35	0.2769
N1-3	11	11	22	
Mismatch repair enzyme expression				
LOM	5	2	7	0.1057
ROM	18	32	50	
EBV infection				
+	0	1	1	1.0000
−	23	33	56	
p53 expression				
Diffuse/ regional/null	13	26	39	0.1497
Sporadic	10	8	18	
Phenotype				
Intestinal predominant	5	17	23	**0.0264**
Non-intestinal	17	15	32	

*Tumor number is 51 after excluding 2 double cancers and 1 collision cancer (6 tumors). LOM/ROM: loss/retention of mismatch repair enzyme expression.
p values <0.05 are shown in bold.

Fig. 1

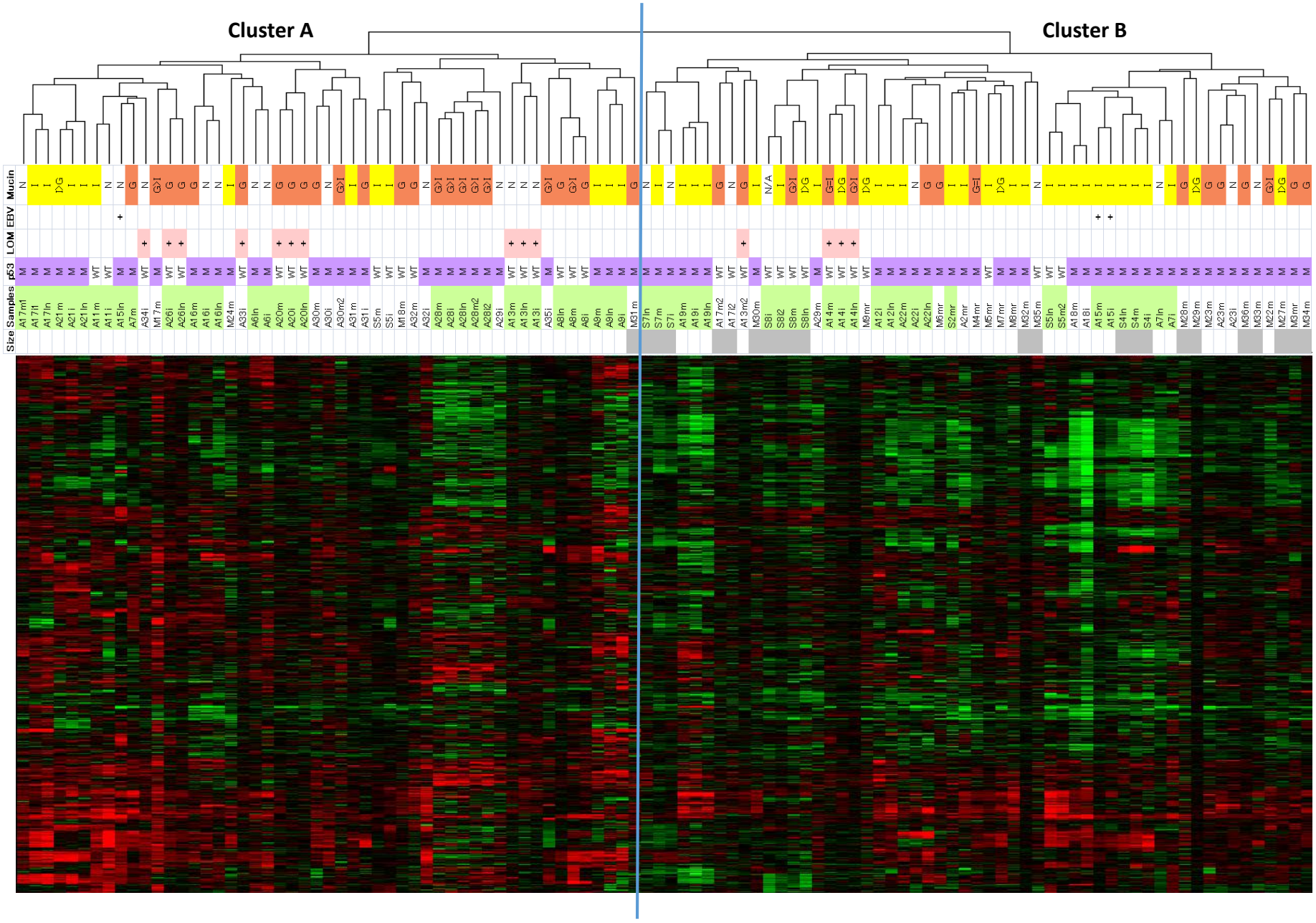


Fig. 2

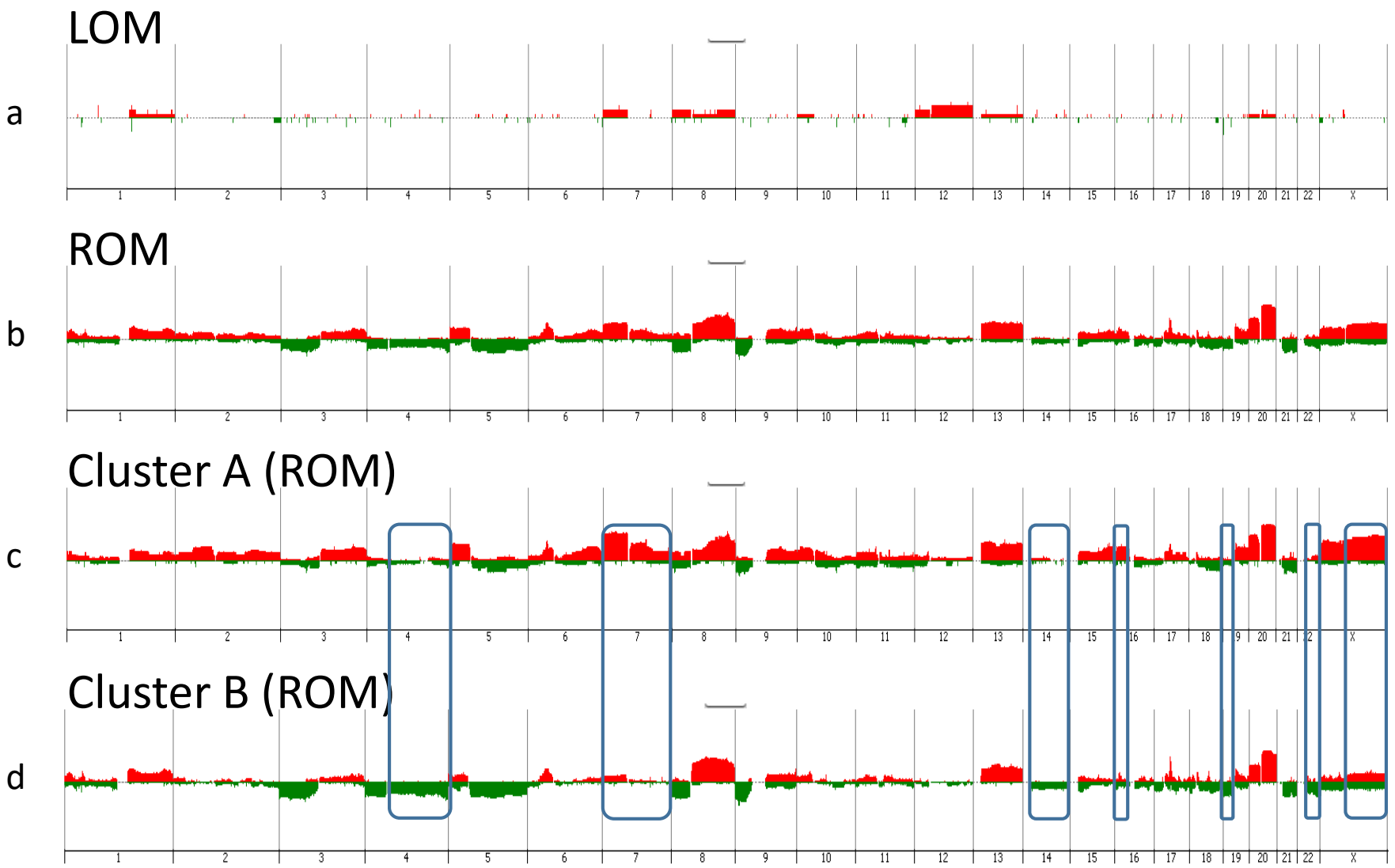

LOM
a
ROM
b
Cluster A (ROM)
c
Cluster B (ROM)
d
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X

Fig. 3

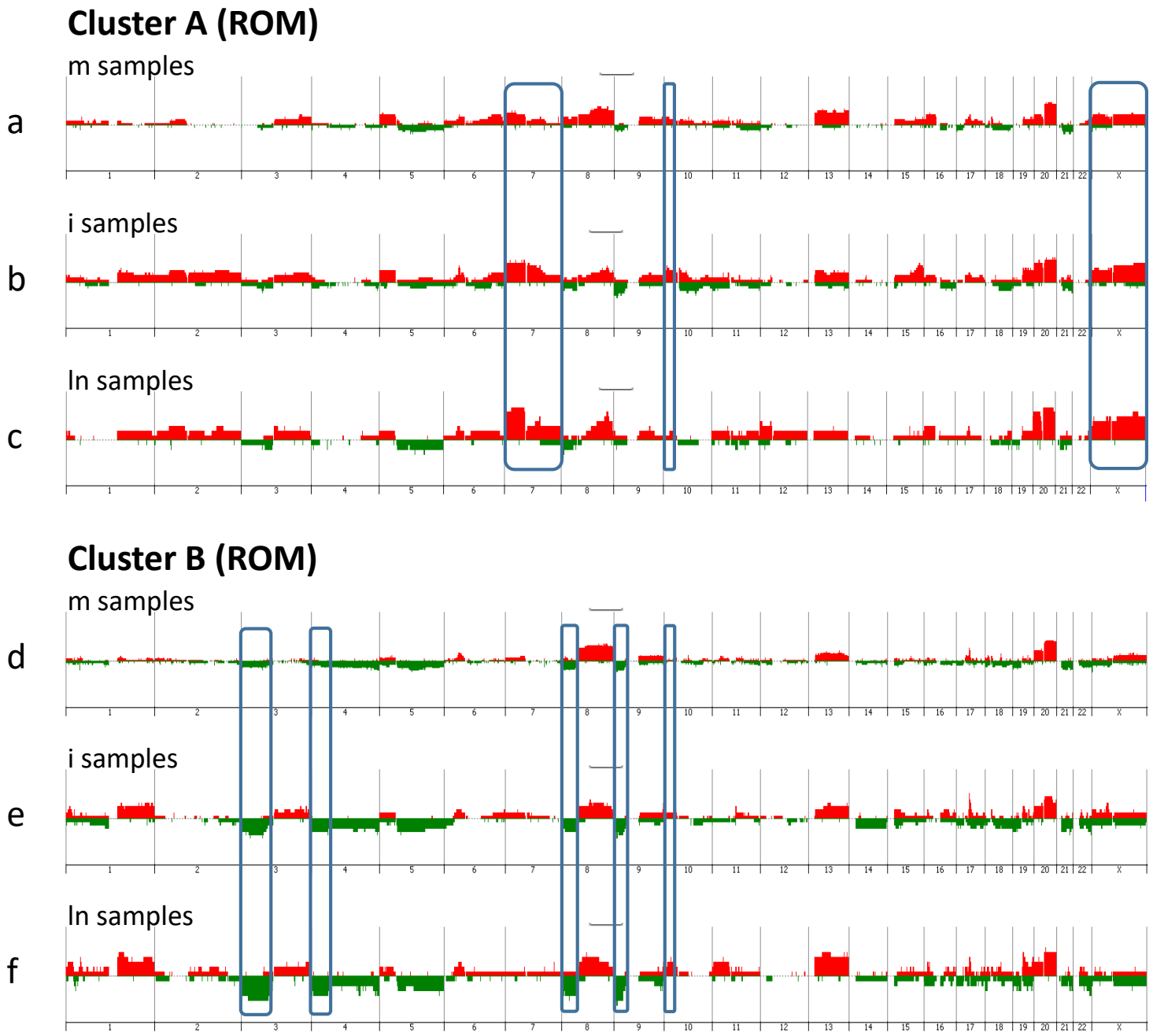

Cluster A (ROM)
m samples
a
i samples
b
ln samples
c
Cluster B (ROM)
m samples
d
i samples
e
ln samples
f

Fig. 4

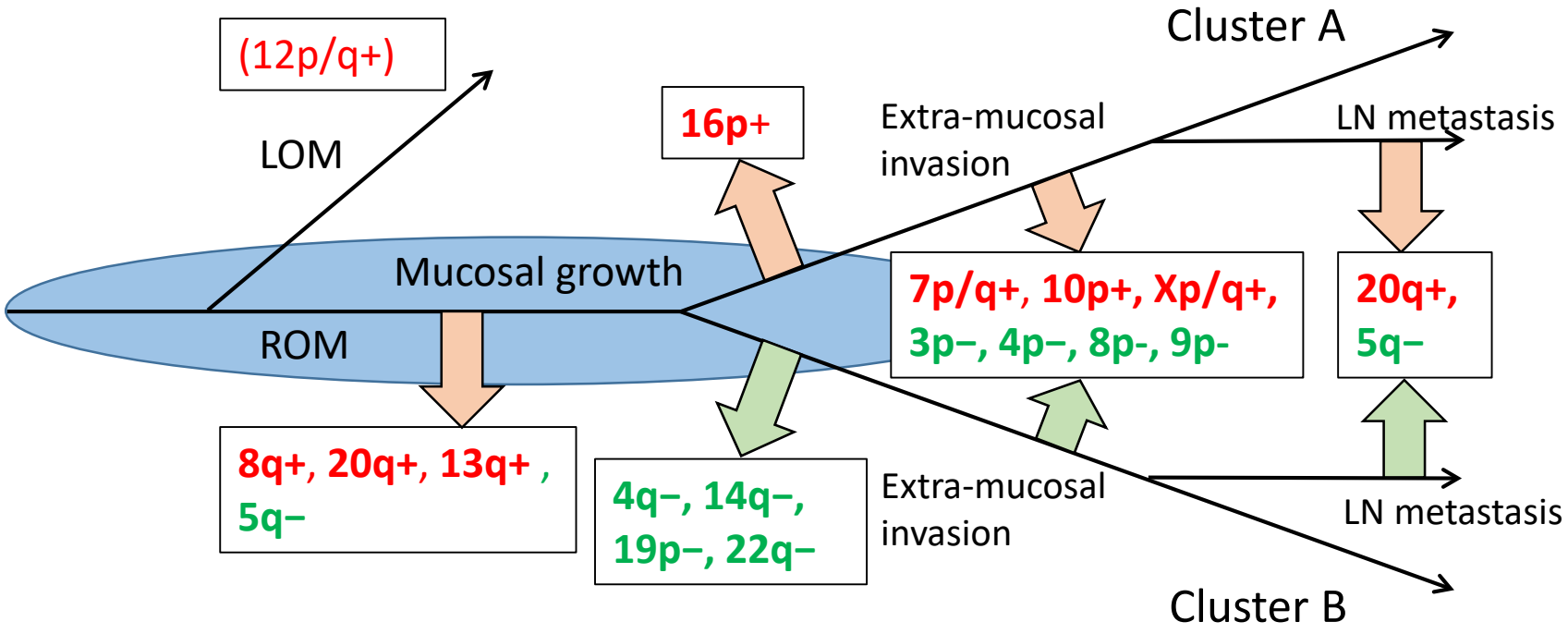
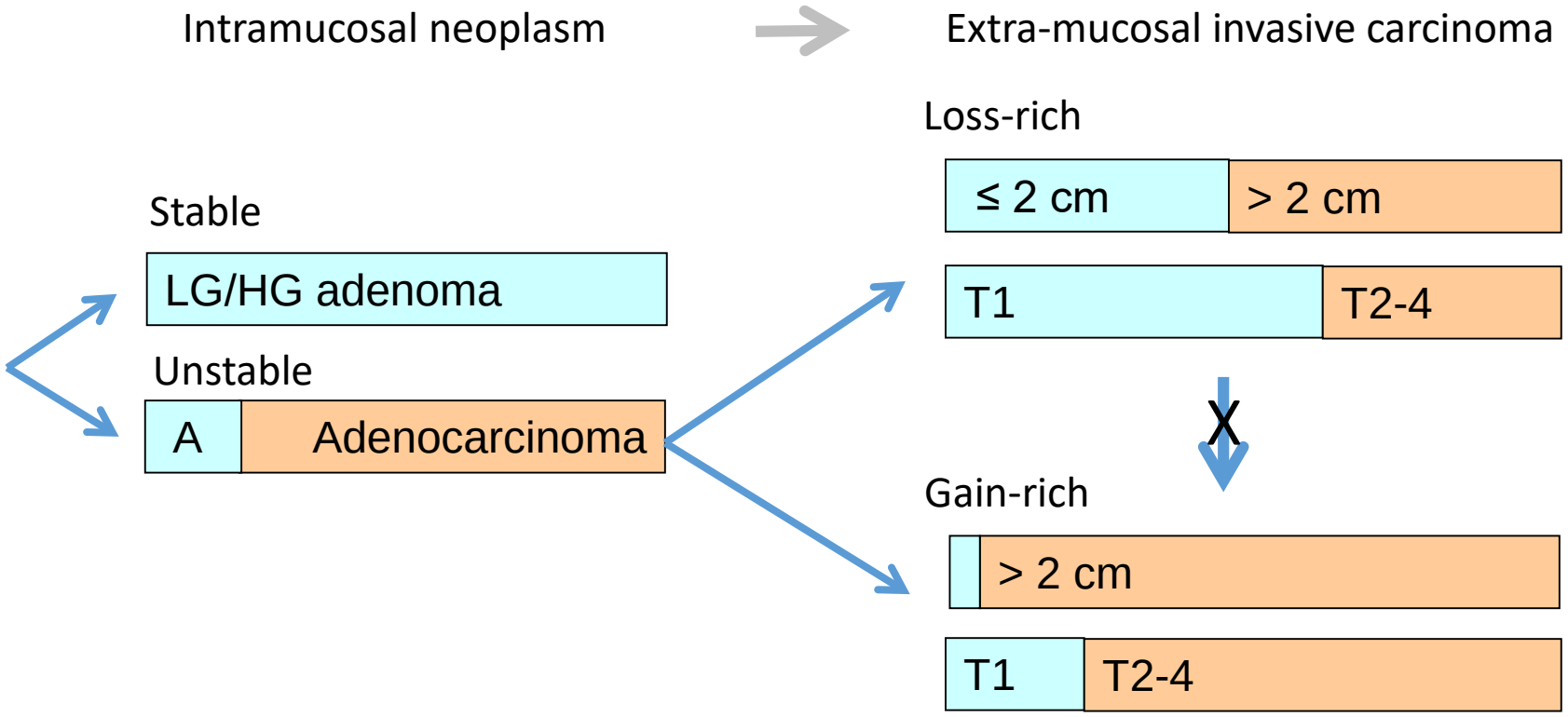

(12p/q+)
LOM
16p+
Cluster A
Extra-mucosal invasion
LN metastasis
Mucosal growth
ROM
7p/q+, 10p+, Xp/q+,
3p−, 4p−, 8p-, 9p-
20q+,
5q−
8q+, 20q+, 13q+ ,
5q−
4q−, 14q−,
19p−, 22q−
Extra-mucosal invasion
LN metastasis
Cluster B

Fig. 5


Intramucosal neoplasm
Extra-mucosal invasive carcinoma
Stable
LG/HG adenoma
Unstable
A
Adenocarcinoma
Loss-rich
≤ 2 cm
> 2 cm
T1
T2-4
X
Gain-rich
> 2 cm
T1
T2-4

Online suppl. Table 1: Clinicopathological features of individual tumors (age, gender, locus, size of mucosal lesion, pT, pN, and excision type). and immunohistochemical and in situ hybridization results (p53, MMR enzymes, EBV, mucin markers) of individual samples.

Tumor N.	Samples	Age	Gender	Locus*	Size	T	N	Excision	p53	MSH6	PMS2	LOM†	EBV	MUC5AC	MUC6	MUC2	CD10	Phenotype‡
#1	M3mr	67	F	M	13	1a	0	Surgical	Null	+	+			2+	—	—	—	G
#2	M4mr	67	M	M	30	1a	0	Surgical	Diffuse	+	+			1+	—	1+	—	G=I
#3	M5mr	74	M	M	60	1a	0	Surgical	Sporadic	+	+			—	—	1+	1+	I
#4	M6mr	65	M	M	25	1a	0	Surgical	Diffuse	+	+			1+	—	—	—	G
#5	M7mr	67	F	M	25	1a	0	Surgical	Diffuse	+	+			—	1+	1+	1+	I>G
#6	M8mr	70	F	L	35	1a	0	Surgical	Null	+	+			—	—	—	1+	I
#7	M9mr	51	M	L	40	1a	0	Surgical	Sporadic	+	+			1+	—	2+	-	I>G
#8	M17m	59	M	L	25	1a	0	Surgical	Diffuse	+	+			2+	2+	1+	2+	G>I
#9	M18m	80	F	L	25	1a	0	Surgical	Sporadic	+	+			3+	1+	—	—	G
#10	M22m	76	M	M	22	1a	0	Surgical	Null	+	+			1+	1+	1+	—	G>I
#11	M23m	65	M	M	35	1a	0	Surgical	Null	+	+			3+	2+	—	—	G
#12	M24m	68	M	M	25	1a	0	Surgical	Diffuse	+	+			—	—	1+	3+	I
#13	M27m	75	M	M	10	1a	0	ESD**	Diffuse	+	+			—	1+	1+	1+	I>G
#14	M28m	58	M	M	7	1a	0	ESD	Diffuse	+	+			1+	1+	—	—	G
#15	M29m	62	F	M	7	1a	0	ESD	Diffuse	+	+			1+	—	—	2+	I>G
#16	M31m	81	F	L	9	1b	0	ESD	Diffuse	+	+			2+	—	—	—	G
#17	M30m	71	M	U	10	1a	0	ESD	Null	+	+			—	—	—	1+	I
#18	M32m	61	F	U	9	1a	0	ESD	Null	+	+			—	—	1+	2+	I
#19	M33m	60	M	M	8	1a	0	ESD	Diffuse	+	+			—	—	—	—	N
#20	M34m	61	M	M	7	1a	0	ESD	Regional	+	+			1+	2+	—	—	G
#21	M35m	70	M	U	5	1a	0	ESD	Sporadic	+	+			—	—	—	—	N
#22	M36m	67	M	M	9	1a	0	ESD	Diffuse	+	+			2+	1+	—	—	G
#23	S2mr	81	F	L	35	1b	1	Surgical	Diffuse	+	+			—	—	—	1+	I
#24	S4m	63	M	L	20	1b	1	Surgical	Diffuse	+	+			—	—	—	1+	I
#24	S4i								Diffuse	+	+			—	—	—	2+	I
#24	S4ln								Diffuse	+	+			—	—	—	2+	I
#25	S5m	82	M	U	35	1b	0	Surgical	Sporadic	+	+			—	—	—	3+	I
#25	S5i								Sporadic	+	+			—	—	—	3+	I
#26	S5m2								Sporadic	+	+			—	—	—	3+	I
#26	S5ln								Sporadic	+	+			—	—	—	3+	I
#27	S7m	75	F	M	20	1b	1	Surgical	Diffuse	+	+			—	—	+	—	I
#27	S7i								Diffuse	+	+			—	—	—	—	N
#27	S7ln								Diffuse	+	+			—	—	—	—	N
#28	S8m	64	F	M	15	1b	1	Surgical	Sporadic	+	+			2+	1+	1+	—	G>I
#28	S8i								Sporadic	+	+			N/A¶	N/A	N/A	N/A	N/A
#28	S8i2								Sporadic	+	+			—	—	—	3+	I
#28	S8ln								Sporadic	+	+			—	2+	1+	3+	I>G
#29	A2mr	71	M	M	100	3	0	Surgical	Diffuse	+	+			—	—	—	1+	I
#30	A6i	81	F	M	35	3	2	Surgical	Diffuse	+	+			—	—	—	—	N
#30	A6ln								Diffuse	+	+			—	—	—	—	N
#31	A7m	79	M	U	70	4a	3	Surgical	Diffuse	+	+			1+	2+	—	—	G
#31	A7i								Diffuse	+	+			—	—	—	1+	I
#31	A7ln								Diffuse	+	+			—	—	—	—	N
#32	A8m	69	F	L	45	3	3	Surgical	Sporadic	+	+			2+	—	1+	—	G>I
#32	A8i								Sporadic	+	+			2+	—	—	—	G
#32	A8ln								Sporadic	+	+			2+	—	—	—	G
#33	A9m	67	F	L	50	4a	2	Surgical	Diffuse	+	+			—	—	3+	2+	I
#33	A9i								Diffuse	+	+			1+	—	2+	1+	I
#33	A9ln								Diffuse	+	+			—	—	2+	—	I
#34	A11m	70	F	M	25	2	1	Surgical	Sporadic	+	+			—	—	—	3+	I
#34	A11i								Sporadic	+	+			—	—	—	—	N
#35	A12i	65	F	L	55	3	1	Surgical	Diffuse	+	+			—	—	—	3+	I
#35	A12ln								Diffuse	+	+			—	—	—	3+	I
#36	A13m	75	F	M	43	2	1	Surgical	Sporadic	+	—	LOM		—	—	—	—	N
#36	A13i								Sporadic	+	—	LOM		—	—	—	—	N
#36	A13ln								Sporadic	+	—	LOM		—	—	—	—	N
#37	A13m2	75	F	U	45	2	0	Surgical	Sporadic	+	—	LOM		1+	—	—	—	G
#38	A14m	77	M	L	70	4a	3	Surgical	Sporadic	+	—	LOM		—	3+	3+	—	G=I
#38	A14i								Sporadic	+	—	LOM		—	1+	2+	—	I>G
#38	A14ln								Sporadic	+	—	LOM		3+	1+	2+	—	G>I
#39	A15m	71	M	U	95	4a	3	Surgical	Null	+	+		+	—	—	2+	1+	I
#39	A15i								Null	+	+		+	—	—	1+	—	I
#39	A15ln								Null	+	+		+	—	—	—	—	N
#40	A16m	80	M	L	80	3	3	Surgical	Diffuse	+	+			1+	1+	—	—	G
#40	A16i								Diffuse	+	+			—	—	—	—	N
#40	A16ln								Diffuse	+	+			—	—	—	—	N
#41	A17m1	82	M	L	40	3	1	Surgical	Diffuse	+	+			—	—	—	—	N
#41	A17i1								Diffuse	+	+			—	—	—	1+	I
#41	A17ln								Diffuse	+	+			—	—	—	3+	I
#42	A17m2	82	M	M	15	2	0	Surgical	Sporadic	+	+			1+	1+	—	—	G
#42	A17i2								Sporadic	+	+			—	—	—		N
#43	A18m	72	M	U	40	3	0	Surgical	Diffuse	+	+			—	—	—	3+	I
#43	A18i								Diffuse	+	+			—	—	—	3+	I
#44	A19m	65	M	U	40	3	3	Surgical	Diffuse	+	+			—	—	+	—	I
#44	A19i								Diffuse	+	+			—	—	+	—	I
#44	A19ln								Diffuse	+	+			—	—	1+	1+	I
#45	A20m	86	M	L	100	3	1	Surgical	Sporadic	+	—	LOM		3+	1+	—	—	G
#45	A20i								Sporadic	+	—	LOM		3+	1+	—	—	G
#45	A20ln								Sporadic	+	—	LOM		3+	3+	—	—	G
#46	A21m	79	M	M	45	4a	2	Surgical	Regional	+	+			—	1+	—	3+	I>G
#46	A21i								Regional	+	+			—	—	—	3+	I
#46	A21ln								Regional	+	+			—	—	—	3+	I
#47	A22m	85	M	M	60	3	3a	Surgical	Diffuse	+	+			—	—	1+	—	I
#47	A22i								Diffuse	+	+			—	—	—	—	N
#47	A22ln								Diffuse	+	+			1+	—	—	—	G
#48	A23m	70	M	M	30	2	0	Surgical	Diffuse	+	+			1+	—	—	—	G
#48	A23i								Diffuse	+	+			—	—	—	—	N
#49	A26i	86	F	L	85	3	2	Surgical	Sporadic	+	—	LOM		2+	—	—	—	G
#49	A26ln								Sporadic	+	—	LOM		2+	1+	—	—	G
#50	A28m	85	M	M	95	3	>1	Surgical	Null	+	+			2+	1+	1+	—	G>I
#50	A28i								Null	+	+			3+	1+	2+	—	G>I
#50	A28ln								Null	+	+			3+	—	1+	—	G>I
#50	A28m2								Null	+	+			2+	1+	1+	—	G>I
#50	A28i2								Null	+	+			2+	2+	2+	—	G>I
#51	A29m	86	M	M	31	3	0	Surgical	Diffuse	+	+			—	—	—	2+	I
#51	A29i								Diffuse	+	+			—	—	—	—	N
#52	A30m	78	F	L	48	3	0	Surgical	Diffuse	+	+			1+	—	—	—	G
#52	A30i								Diffuse	+	+			—	—	—	—	N
#52	A30m2								Diffuse	+	+			—	2+	1+	—	G>I
#53	A31m	80	M	U	56	3	0	Surgical	Regional	+	+			—	—	—	3+	I
#53	A31i								Regional	+	+			2+	—	—	—	G
#54	A32m	71	M	L	60	4b	0	Surgical	Sporadic	+	+			3+	—	—	—	G
#54	A32i								Regional	+	+			—	—	—	—	N
#55	A33i	72	M	L	120	3	0	Surgical	Sporadic	+	—	LOM		1+	2+	—	—	G
#56	A34i	57	M	L	60	3	0	Surgical	Sporadic	+	—	LOM		—	—	—	—	N
#57	A35i	72	M	U	60	2	0	Surgical	Diffuse	+	+			3+	3+	—	1+	G>I

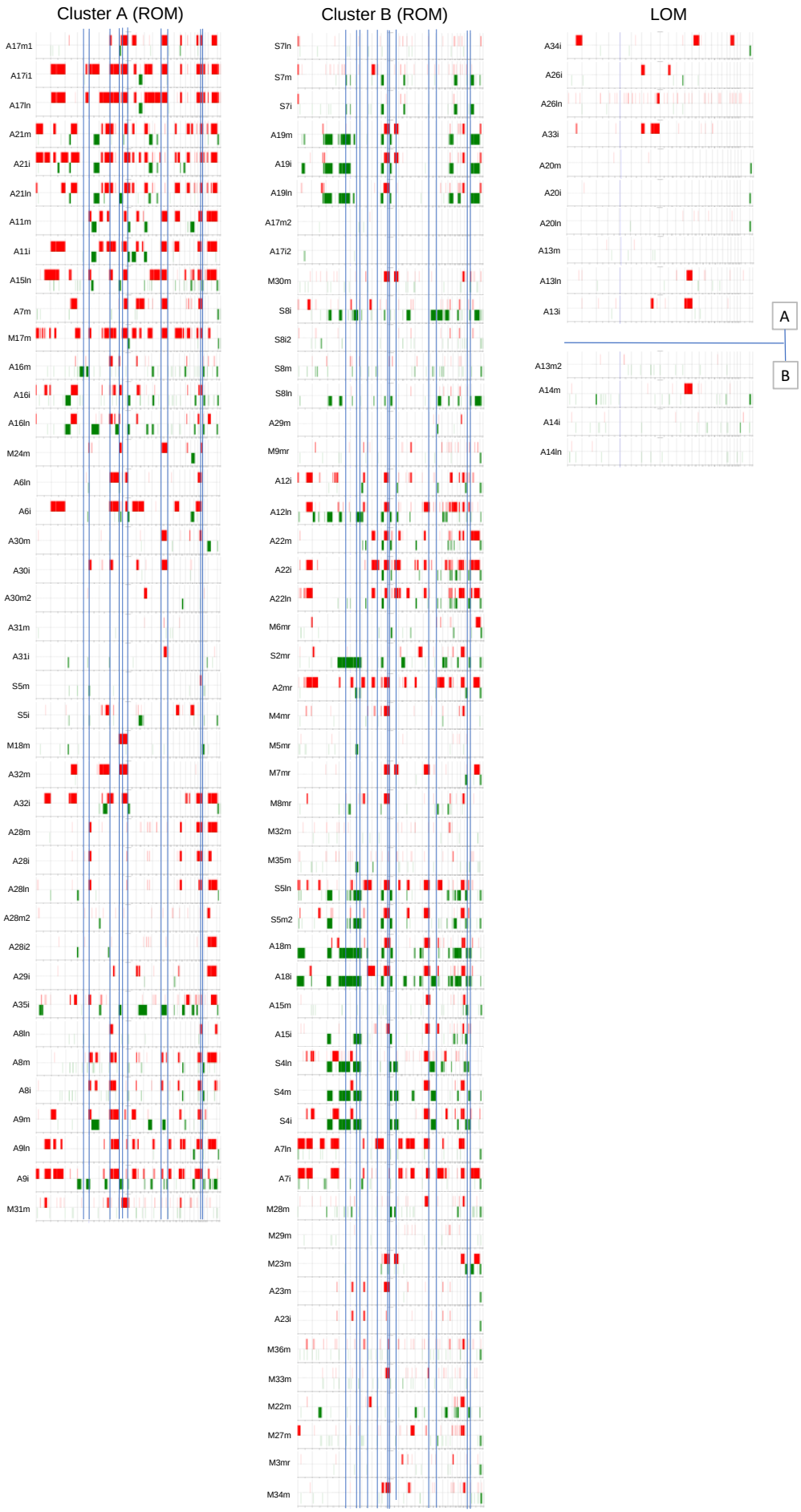
*Locus: "U"/"M"/"L", upper/middle/lower one third of the stomach. **ESD: endoscopic submucosal dissection. †LOM: Loss of mismatch repair enzyme expression. ‡ Phenotype: "G"/"I"/"N", gastric/intestinal/null. ¶N/A: not assessable.

Online suppl. Table 2: Frequencies of copy-number gains and amplifications of receptor tyrosine kinases and representative growth-related genes on chromosome 7 in clusters A and B.

	Name of Gene	Copy-number gain			Amplification		
		Cluster A (*n = 51*)	Cluster B (*n = 5* 5)	*p value*	Cluster A (*n = 51*)	Cluster B (*n = 5* 5)	*p value*
Receptor tyrosine kinases (RTKs)	AATK	8	13	0.3388	1	0	0.4811
	ALK	6	4	0.516	0	0	1
	AXL	15	3	**0.0014**	1	0	0.4811
	CSF1R	10	12	0.8148	0	1	1
	DDR1	20	19	0.6887	3	3	1
	DDR2	16	19	0.8368	2	0	0.2291
	FGFR1	13	12	0.8193	2	1	0.6074
	FGFR2	5	12	0.1155	0	1	1
	FGFR3	21	17	0.314	8	4	0.2249
	FGFR4	17	8	**0.0381**	3	1	0.3495
	FLT1(VEGFR1)	25	19	0.168	5	1	0.1033
	FLT3	16	23	0.3158	3	0	0.1079
	FLT4 (VEGFR3)	19	14	0.2131	1	0	0.4811
	IGF1R	16	6	**0.0154**	2	0	0.2291
	INSR	3	2	0.6699	0	0	1
	INSRR	N/A	N/A	N/A	N/A	N/A	N/A
	KDR (VEGFR2)	4	2	0.4248	1	0	0,.811
	KIT	6	3	0.3075	1	0	0.4811
	LMTK2	29	15	**0.003**	7	0	**0.0048**
	LMTK3	N/A	N/A	N/A	N/A	N/A	N/A
	LTK	N/A	N/A	N/A	N/A	N/A	N/A
	MERTK	14	12	0.6519	0	1	1
	MET	18	11	0.0861	0	1	1
	MST1R	17	7	**0.0191**	1	0	0.4811
	MUSK	13	9	0.3382	0	0	1
	NTRK1	13	17	0.6667	1	1	1
	NTRK2	9	7	0.5901	0	0	1
	NTRK3	14	4	**0.0086**	0	0	1
	PDGFRA	4	3	0.7086	0	0	1
	PDGFRB	7	8	1	0	0	1
	PTK7	22	21	0.6932	13	10	0.48
	RET	19	19	0.8405	2	7	0.1636
	ROR1	5	6	1	0	0	1
	ROR2	7	3	0.1905	0	0	1
	ROS1	10	2	**0.0129**	0	0	1
	RYK	9	3	0.0661	0	0	1
	STYK1	29	20	0.0508	8	3	0.114
	TEK	3	3	1	0	0	1
	TIE1	13	22	0.1484	5	2	0.2575
	TYRO3	17	15	0.5313	1	1	1
	EPHA1	6	7	1	0	0	1
	EPHA2	16	14	0.5248	0	0	1
	EPHA3	14	4	**0.0086**	0	0	1
	EPHA4	9	5	0.2544	0	0	1
	EPHA5	3	8	0.2052	0	0	1
	EPHA6	6	5	0.755	0	0	1
	EPHA7	3	2	0.6699	0	0	1
	EPHA8	6	12	0.2018	2	0	0.2291
	EPHA10	26	29	1	11	14	0.6554
	EPHB1	9	5	0.2544	0	0	1
	EPHB2	21	24	0.8457	2	1	0.6074
	EPHB3	24	15	**0.0444**	8	2	**0.0464**
	EPHB4	28	20	0.0786	14	5	**0.0212**
	EPHB6	32	27	0.1755	14	7	0.0867
	EGFR	20	7	**0.0033**	6	1	0.0537
	ERBB2	28	32	0.8449	7	13	0.2219
	ERBB3	17	8	**0,.0381**	7	3	0.1905
	ERBB4	6	1	0.0537	0	0	1
Growth-related genes on chromosome 7 except RTKs	RALA	18	9	**0.0283**	2	0	0.2291
	PTPN12	16	9	0.1079	4	0	0.0503
	MAFK	34	25	**0.0328**	23	15	0.0693
	ARHGEF5	17	15	0.5313	5	3	0.4772
	BRAF	11	9	0.6206	1	0	0.4811
	CAV1	12	8	0.3213	1	0	0.4811
	GLI3	25	11	**0.0021**	1	0	0.4811
	HOXA1	29	32	1	4	3	0.7086
	RABL5	21	13	0.063	4	3	0.7086
	RBM28	22	21	0.6932	8	7	0.7826
	CUL1	31	20	**0.0192**	11	3	**0.0203**
	IGF2BP3	25	19	0.168	3	2	0.6699
	RELB	25	17	0.0741	3	0	0.1079
	SEMA3E	19	7	**0.006**	1	0	0.4811

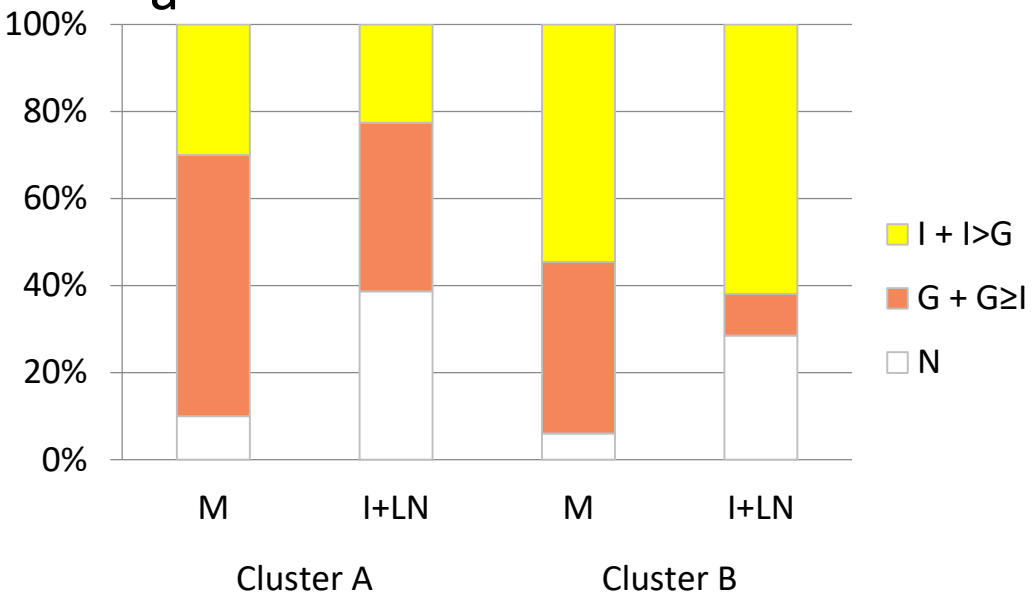
p values <0.05 are shown in bold. N/A: not assessable.

Online suppl. Fig. 1



Online suppl. Fig. 2

a



b

		G + G≥I	N	*p* value
A	M	12	2	**0.0392**
	I+LN	12	12	
B	M	13	2	**0.0062**
	I+LN	2	6	
A+B	M	25	4	**0.0011**
	I+LN	14	18	
A		24	14	1.0000
B		15	8	

c

		I + I>G	Non-I	*p* value
A	M	6	14	0.7432
	I+LN	7	24	
B	M	18	15	0.7784
	I+LN	13	8	
A+B	M	24	29	0.5545
	I+LN	20	32	
A		13	38	**0.0014**
B		31	23	

Online suppl. Fig. 3

No	Gene symbol	Gene name	Gene function	Location	Probe number	Mean T/R*		Difference (A vs B)	
						Cluster A	Cluster B	*p* value	*p* value after BC¶
1	**VRK2**	vaccinia related kinase 2	**apoptosis and growth**	2p16.1	4	0.2216578	-0.091428	1.78882E-06	0.026791191
2	**XIRP2**	xin actin binding repeat containing 2	**actin binding**	2q24.3	8	0.2525296	0.0062581	1.27706E-06	0.019126541
3	**CNTN6**	contactin 6	**cell adhesion**	3p26.3	6	0.00099	-0.322622	8.6214E-07	0.012912274
4	**CNTN4**	contactin 4	**axon connections**	3p26.3-p26.2	19	0.1137656	-0.189603	2.02011E-09	3.02551E-05
5	**RBMS3**	RNA binding motif single stranded interacting protein 3	**c-myc gene binding**	3p24.1	16	0.0883685	-0.259709	1.30358E-12	1.95238E-08
6	**ARPP-21**	cAMP regulated phosphoprotein 21	**nerve function**	3p22.3	5	0.1407465	-0.211016	3.69533E-07	0.005534503
7	**STAC**	SH3 and cysteine rich domain	**SH3 and cysteine rich**	3p22.3-p22.2	4	0.0944739	-0.24017	5.63687E-07	0.008442343
8	**MYRIP**	myosin VIIA and Rab interacting protein	**myosin interacting**	3p22.1	10	0.1020163	-0.178061	1.24906E-06	0.018707211
9	**CACNA2D3**	calcium voltage-gated channel auxiliary subunit alpha2delta 3	**tumor suppressor gene**	3p21.1-p14.3	20	0.1946772	-0.121071	3.18111E-08	0.000476436
10	**PTPRG**	protein tyrosine phosphatase, receptor type G	**tumor suppressor gene**	3p14.2	14	0.0442553	-0.238805	3.32493E-06	0.049797535
11	**LRIG1**	leucine rich repeats and immunoglobulin like domains 1	**tumor suppressor gene**	3p14.1	3	0.2577084	-0.234233	1.39033E-06	0.020822901
12	**SLIT2**	slit guidance ligand 2	**tumor suppressor gene**	4p15.31	11	0.0410187	-0.214485	1.32879E-06	0.019901247
13	**SEL1L3**	SEL1L family member 3	**lymph node/stomach expression**	4p15.2	3	0.1200379	-0.272848	5.63456E-07	0.00843888
14	**FSTL5**	follistatin like 5	**tumor suppressor gene**	4q32.2	16	0.0087773	-0.252876	1.05674E-06	0.015826866
15	**GPM6A**	glycoprotein M6A	**oncogenic potential gene**	4q34.2	9	0.1142247	-0.225603	2.75025E-07	0.004119045
16	**MCCC2**	methylcrotonoyl-CoA carboxylase 2	**carboxylase**	5q13.2	3	0.1370447	-0.323169	2.31579E-07	0.003468354
17	**NXPH1**	neurexophilin 1	**nerve function**	7p21.3	6	0.4400862	0.0622393	6.47209E-07	0.009693249
18	**H2AFV**	H2A histone family member V	**histones nucleosome**	7p13	2	0.6194794	0.1092316	8.29559E-07	0.012424304
19	**SUN3**	Sad1 and UNC84 domain containing 3	**testis expression**	7p12.3	2	0.3635659	-0.120918	8.55676E-07	0.012815461
20	**COBL**	cordon-bleu WH2 repeat protein	**actin regulator**	7p12.1	8	0.4966139	0.111728	4.98461E-08	0.000746545
21	**SEMA3E**	semaphorin 3E	**proto-oncogene?**	7q21.11	7	0.2570146	-0.078017	1.71778E-06	0.025727224
22	**MUC12**	mucin 12, cell surface associated	**tumor suppressor gene**	7q22.1	2	0.7944379	0.2106728	8.58551E-08	0.001285852
23	**ORAI2**	ORAI calcium release-activated calcium modulator 2	**calcium modulator**	7q22.1	2	0.370068	-0.099438	3.29493E-07	0.004934813
24	**PRKD1**	protein kinase D1	**target for oncogenic KRas signaling**	14q12	9	0.0830602	-0.212465	1.15426E-06	0.017287292
25	**SLC39A9**	solute carrier family 39 member 9	**proto-oncogene**	14q24.1	3	0.2246334	-0.265776	2.34539E-08	0.000351268
26	**NRXN3**	neurexin 3	**nerve function**	14q24.3-q31.1	29	-0.009645	-0.199779	1.05375E-06	0.015782062
27	**CHST14**	carbohydrate sulfotransferase 14	**sulfotransferases**	15q15.1	1	0.5904759	-0.107594	2.16714E-06	0.032457219
28	**HS3ST3A1**	heparan sulfate-glucosamine 3-sulfotransferase 3A1	**tumor growth factor related gene**	17p12	5	0.1581726	-0.186387	2.87768E-06	0.043099021
29	**MAP2K7**	mitogen-activated protein kinase kinase 7	**tumor growth factor related gene**	19p13.2	2	0.3651435	-0.067787	3.70778E-07	0.005553136
30	**SCUBE1**	signal peptide, CUB domain and EGF like domain containing 1	**EGF (epidermal growth factor)-like**	22q13.2	4	0.0642863	-0.256986	5.19618E-07	0.007782325
31	**MPPED1**	metallophosphoesterase domain containing 1	**brain/liver expression**	22q13.2	3	0.482682	-0.047041	1.24649E-07	0.001866868
32	**CELSR1**	cadherin EGF LAG seven-pass G-type receptor 1	**non-classic-type cadherin**	22q13.31	4	0.0170639	-0.325309	2.06694E-06	0.030956486

*T/R: Tumor/Reference fluorescence intensity ratio. ¶BC: Bonferroni corretion.