

ARTICLE

Open Access

Identification of genomic aberrations associated with lymph node metastasis in diffuse-type gastric cancer

Ji-Hye Choi^{1,2}, Young-Bae Kim³, Ji Mi Ahn³, Min Jae Kim^{1,2}, Won Jung Bae³, Sang-Uk Han⁴, Hyun Goo Woo^{1,2} and Dakeun Lee³

Abstract

Diffuse-type gastric cancer (DGC) is a GC subtype with heterogeneous clinical outcomes. Lymph node metastasis of DGC heralds a dismal progression, which hampers the curative treatment of patients. However, the genomic heterogeneity of DGC remains unknown. To identify genomic variations associated with lymph node metastasis in DGC, we performed whole exome sequencing on 23 cases of DGC and paired non-tumor tissues and compared the mutation profiles according to the presence (N3, $n = 13$) or absence (N0, $n = 10$) of regional lymph node metastasis. Overall, we identified 185 recurrently mutated genes in DGC, which included a significant novel mutation at *CMTM2*, as well as previously known mutations at *CDH1*, *RHOA*, and *TP53*. Noticeably, *CMTM2* expression could predict the prognostic outcomes of DGC but not intestinal-type GC (IGC), indicating pivotal roles of *CMTM2* in DGC progression. In addition, we identified a recurrent loss of heterozygosity (LOH) of DNA copy numbers at the 3p12-pcen locus in DGC. A comparison of N0 and N3 tumors showed that N3 tumors exhibited more frequent DNA copy number aberrations, including copy-neutral LOH and mutations of CpTpT trinucleotides, than N0 tumors ($P = 0.2 \times 10^{-3}$). In conclusion, DGCs have distinct profiles of somatic mutations and DNA copy numbers according to the status of lymph node metastasis, and this might be helpful in delineating the pathobiology of DGC.

Introduction

Gastric cancer (GC) is the fifth most common cancer and the third leading cause of cancer-related mortality worldwide^{1,2}. Histologically, GCs are classified as gland-forming intestinal-type GC (IGC) and scattered-cell diffuse-type GC (DGC) by the Lauren classification³. The incidence of GC has been appreciably declining for several decades, but the incidence of DGC is constantly increasing in Asia and Western countries^{4–6}. Pathologically,

DGC has a massive infiltrative growth pattern with prominent stromal reaction and a propensity for direct serosal invasion and peritoneal carcinomatosis⁷, which are generally considered an aggressive GC phenotype. However, clinical outcomes of DGC patients are highly variable. Some DGCs show widespread lymph node and distant metastases, while others do not, even in very advanced T stages. Lymph node metastasis is a significant prognostic factor for DGC^{8,9} and may contribute to its heterogeneous progression. Delineating the underlying mechanisms of DGC with lymph node metastasis is necessary to develop new and more effective clinical strategies for DGC management.

Advances in high-throughput sequencing technologies have identified genome-wide mutations in GC^{10–14}. The Cancer Genome Atlas Project (TCGA) proposed a

Correspondence: Hyun Goo Woo (hg@ajou.ac.kr) or Dakeun Lee (dakeun@gmail.com)

¹Department of Physiology, Ajou University School of Medicine, Suwon, Republic of Korea

²Graduate School of Biomedical Science, Ajou University School of Medicine, Suwon, Republic of Korea

Full list of author information is available at the end of the article
These authors contributed equally: Ji-Hye Choi, Young-Bae Kim.

© The Author(s) 2018



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

molecular classification dividing GC into four subtypes, including Epstein-Barr virus (EBV)-infected tumors, tumors with microsatellite instability (MSI), genomically stable tumors, and chromosomally unstable tumors¹⁵. The genomically stable phenotype is DGC-enriched and includes mutations of RHOA or fusions involving RHO-family GTPase-activating proteins. DGC displays a lower genomic alteration frequency than other subtypes. However, frequent mutations at *TP53*, *ARID1A*, *CDH1*, and *RHOA* have been noted. Mutations associated with the status of lymph node metastasis in DGC have not been studied thoroughly.

In the present study, by performing whole exome sequencing (WES) analysis of 23 cases of DGC tumors, we compared their mutation profiles and DNA copy number aberrations according to the lymph node metastasis status of N0 ($n = 10$) and N3 ($n = 13$). We were able to identify unique aberrations of DNA copy numbers and mutations between N0 and N3 tumors, which might play pivotal roles in DGC progression with lymph node metastasis.

Materials and methods

Patients and samples

Tumor samples, paired adjacent non-tumor tissues, and clinical data for 23 DGC patients undergoing gastrectomy as a primary treatment option were obtained from Ajou University Hospital, South Korea. Histologically, these tumors were all pure, poorly cohesive carcinomas, showing infiltrative growth and having only a small amount of intra-tumor fibrosis. All samples were collected after obtaining written informed consent from the patients and were frozen at -80°C until use at the Ajou Human Bio-Resource Bank. The Institutional Review Board of Ajou University Hospital approved the study protocol.

No patients had undergone previous chemotherapy or radiation therapy. All tissues required pathology evaluation, and the minimum acceptable percentage of tumor cells was 60%. The presence of EBV or MSI was examined. EBV-encoded small ribonucleic acid in situ hybridization was performed for formalin-fixed paraffin-embedded (FFPE) tissue sections (Ventana Medical Systems, Tucson, AZ, USA). MSI analysis was performed using five microsatellite markers of the Bethesda consensus panel (D2S123, S17S250, D5S346, BAT25, and BAT26)¹⁶. All tumor samples showing the absence of EBV and MSI were included in this study.

WES and variant calling

DNA was extracted from the tumors and matched normal gastric tissues using the QIAamp DNA Mini kit according to the manufacturer's protocol (QIAGEN, Valencia, CA). For targeted exome capture, an Illumina TruSeq exome enrichment kit was used with the given

protocol. The captured samples were sequenced as 101-bp paired-end reads using Illumina GAIIx, and an average of 126 million raw reads per sample was obtained. Low-quality bases with Phred scores of <20 were masked using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). The masked reads were mapped to the hg19 human reference genome using the Burrows-Wheeler Alignment tool (BWA) with default parameters. PCR duplicates were identified and removed using Picard (<http://sourceforge.net/apps/mediawiki/picard/>). Average mapping of the sequence reads to the reference genome ranged from 80.27 to 92.82, and the average coverage of the reads was $85.44 \times (61.19-106.46)$ (Supplementary Table S1). Local realignments of indels were performed using the GATK local realignment walker¹⁷, and the read quality was normalized using the GATK recalibration walker. All variants were annotated using ANNOVAR software¹⁸.

Expression of the mutation signatures was estimated using the 'deconstructSigs' package¹⁹ and the published signatures from the Sanger COSMIC database (<http://cancer.sanger.ac.uk/cosmic>).

Copy number analysis

Estimation of the copy number and ploidy of the tumor samples from the WES data was performed on all the tumor/non-tumor pairs using Sequenza with the default parameters²⁰. DNA copy number gains and losses were determined using a cutoff of ± 0.2 compared to that of the non-tumors. LOH was determined with a B allele frequency cutoff of less than 0.4. LOH without copy number gain or loss was defined as copy-neutral LOH.

Identification of significantly deleterious mutations (SDMs)

We investigated significantly deleterious mutations (SDMs) using four different algorithms for predicting protein functions with default parameters: SIFT (<http://sift.jcvi.org/>), Polyphen2_HDIV (<http://genetics.bwh.harvard.edu/pph2/>), LRT (http://www.genetics.wustl.edu/jflab/lrt_query.html), and MutationTaster (<http://www.mutationtaster.org/>). SDMs were determined as mutations predicted to have a deleterious mutation by at least three of the four algorithms.

PCR and Sanger sequencing validation

The identified sequence variants were validated using PCR reactions and subsequent Sanger sequencing analysis. The specific primers used in the PCR and sequencing reactions are summarized in Supplementary Table S2.

Functional enrichment analysis of gene sets

Enriched functions in the mutated gene sets were analyzed using the Gene Ontology (GO) categories and the

KEGG database as implemented in DAVID 6.8 software (<https://david.ncifcrf.gov>).

Data availability

The data supporting the main findings of this study are available from the corresponding authors upon request.

Cell lines

Human gastric cancer cell lines Kato III, MKN28, MKN1, MKN45, MKN74, SNU216, SNU601, SNU668, AGS, and NCI-N87 and the prostate cancer cell lines PC3 and LNCap were used. Cell lines were purchased from the Korean Cell Line Bank (KCLB) and were maintained in RPMI 1640 (HyClone, South Lagan, UT, USA) supplemented with 10% fetal bovine serum (FBS), 1% penicillin, and streptomycin at 37 °C under a humidified atmosphere containing 5% CO₂. PC3 and LNCap were used as negative and positive controls for CMTM2 expression, respectively.

RNA extraction and reverse transcription PCR (RT-PCR)

Total RNA was isolated using TRIzol reagent (Invitrogen, Carlsbad, CA, USA). cDNA was synthesized from 2 µg of total RNA using the PrimeScript 1st strand cDNA synthesis kit (Takara bio, Shiga, Japan) according to the manufacturer's protocol. For the nested PCR, CMTM2 external primers were used for the primary amplification, and internal primers were used for the following secondary amplification. The primary PCR product (diluted 1: 50,000) was used as the template in a secondary PCR. The final PCR products were electrophoresed on a 1% agarose gel and were identified by SYBR Safe DNA gel staining (Invitrogen, Carlsbad, CA, USA). β-actin was used as an internal control. The primer sequences are described in the Supplementary Methods.

Lentivirus transduction

Detailed procedures are described in the Supplementary Methods.

Cell viability/proliferation assay

The cell viability/proliferation assay was performed with the EZ-Cytox assay kit (Daeil Lab Service, Seoul, Korea). The absorbance of the samples was measured using a microplate reader at 450 nm. The experiments were performed in triplicate.

Transwell invasion/migration assay

Detailed procedures are described in the Supplementary Methods.

Statistical analysis

Differences in the experimental data were compared using the independent Student's *t*-test as implemented in the R program (V. 3.2.2.).

Results

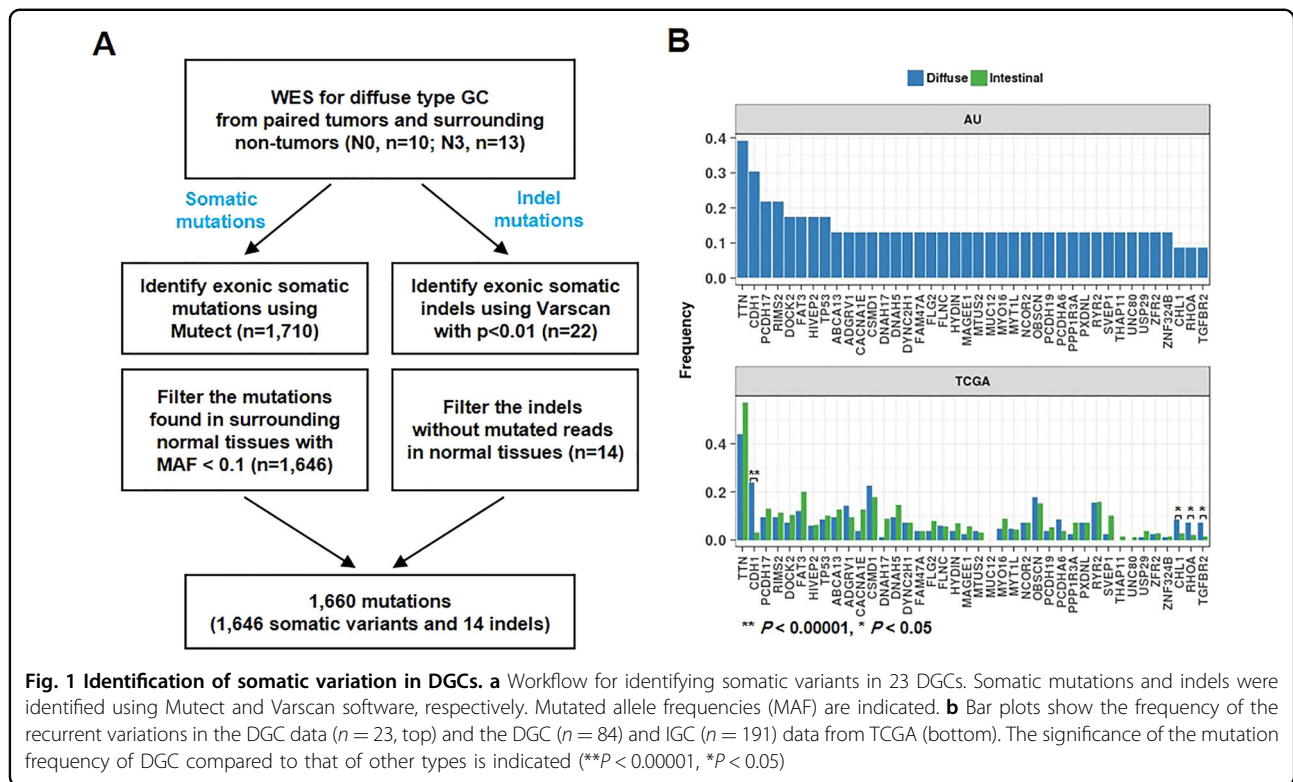
Whole exome sequencing identifies mutation profiles of DGC

WES analysis was performed on 23 cases of DGC tumors and paired adjacent non-tumor tissues obtained from Ajou University Hospital. To delineate mutations associated with lymph nodal metastasis, DGC samples with the presence (N3, *n* = 13) or absence (N0, *n* = 10) of lymph nodal metastasis were examined. The clinical and pathological characteristics of the patients are summarized in Table 1. N3 tumors were more prevalent in males and showed larger tumor size (9.43 ± 4.6 cm) when compared to N0 tumors.

Somatic mutations and indel variants were identified by applying MuTect²¹ and VarScan 2²² (*P* < 0.01), respectively. To remove possible germline mutations, the variants with mutated allele frequencies of greater than 10% in the paired non-tumor tissues were filtered out. Finally, we determined 1646 somatic mutations and 14 indel mutations (Fig. 1a). These included 185 recurrently mutated genes. *TTN* (39.13%) and *CDHI* (30.43%) exhibited the most frequent mutations in DGC tumors (Fig. 1b, top). When we compared the mutation frequencies with the publicly available DGC data from TCGA (*n* = 84), similar mutation rates of the recurrent mutations were evident (Fig. 1b, bottom). In addition, we compared the mutation frequencies between the DGC (*n* = 84) and IGC (*n* = 191) subgroups. Consistent with previous studies, we found that the recurrent mutations in DGC, such as those in the *CDHI*, *CHLI*, *RHOA*, and *TGFBR2* genes, were more frequently mutated in DGC than in IGC tumors (Fisher's exact test for DGC vs. IGC, *P* < 0.05). These results support the suggestion that DGC has a distinct mutation profile compared to other subtypes of GC.

Table 1 Clinical and pathological features of DGC samples

	N0 DGC	N3 DGC
Sex (male/female)	5/5	9/4
Age (mean ± S.D.)	51.8 ± 13.8	53.30 ± 12.4
Gross (Borrmann) type (2/3/4)	0/8/2	0/6/7
Tumor Size (cm) (mean ± S.D.)	4.97 ± 2.1	9.43 ± 4.6
Tumor location (cardia/body/antrum)	2/4/4	0/6/7
T Stage (T2/T3/T4)	1/6/3	0/4/9
N stage		
(No. of metastatic lymph nodes (mean ± S.D.))	N0 (0 ± 0)	N3 (22.30 ± 14.52)
M Stage (M0/M1)	10/0	9/4

**Table 2 List of the significantly mutated genes in DGC**

Gene	CytoBand	Chr	Position	Ref	Alt	AA change	SIFT_score*	P-value**
CDH1	16q22.1	16	68846057	T	A	L343Q	0.02 (D)	3.30×10^{-6}
			68842401	A	T	R154S	0 (D)	
			68842731	G	C	E223Q	0.005 (D)	
			68844180	T	A	N256K	0.001 (D)	
			68847277	A	T	D400V	0 (D)	
RHOA	3p21.31	3	49412922	T	A	Y34F	0 (D)	4.34×10^{-3}
			49412946	C	T	S26N	0.001 (D)	
TP53	17p13.1	17	7578406	C	T	R175H	0 (D)	5.43×10^{-3}
			7578479	G	A	P151S	0.005 (D)	
			7577114	C	T	C275Y	0 (D)	
CMTM2	16q21	16	66613601	G	A	G31S	0.783 (T)	6.05×10^{-3}
			66613531	G	T	K7N	0 (D)	

*D Deleterious (score ≤ 0.05), T tolerated (score > 0.05)

**P-value indicates statistical significance for significantly mutated genes identified by MutSigCV

To pinpoint the most significant mutations, we employed a MutSigCV method and identified *CDH1*, *RHOA*, *CMTM2*, and *TP53* as recurrently mutated genes in DGC (Table 2). These mutated genes, except *CMTM2*, have been noted in DGC^{10,12,23}. The recurrent mutations

of *CMTM2* were newly identified and were validated by Sanger sequencing analysis (Supplementary Fig. S1). To address the functional significance of the *CMTM2* gene, we first screened *CMTM2* mRNA levels in GC cell lines using nested RT-PCR and found that the *CMTM2*

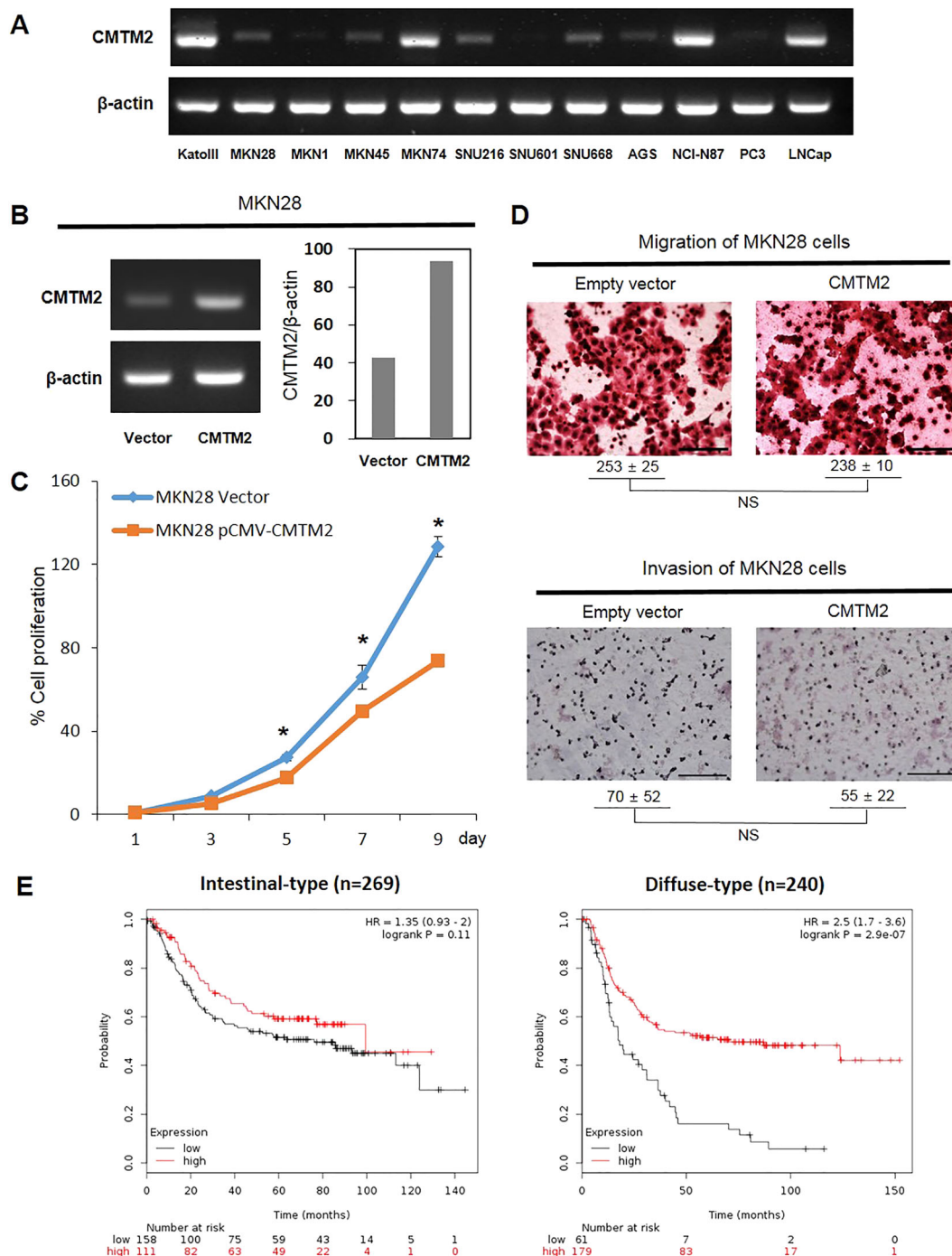


Fig. 2 Functional and clinical significance of *CMTM2* expression. **a** Expression status of *CMTM2* in gastric cancer (GC) cell lines, as detected by nested qRT-PCR. LNCap and PC3 prostate cancer cells were used as positive and negative controls, respectively. **b** Increased *CMTM2* expression in *CMTM2*-induced MKN28 cells. **c** Reduced proliferation of *CMTM2*-overexpressing MKN28 cells. **d** Migration and invasion of *CMTM2*-induced and empty vector-induced MKN28 cells. Scale bar, 100 μm. **e** Kaplan-Meier survival curves compare the overall survival between the *CMTM2*-low group and the *CMTM2*-high group for intestinal-type GC ($n = 269$, left) and diffuse-type GC ($n = 240$, right) cohorts. NS, not significant. * $P < 0.01$

mRNA levels were quite different among cells (Fig. 2a). We selected MKN28 (low *CMTM2* expression) and transduced a *CMTM2*-expressing lentivirus (Fig. 2b).

CMTM2-overexpressing MKN28 cells showed significantly reduced proliferation compared to the control ($P < 0.01$; Fig. 2c). However, *CMTM2* expression did not

affect the migration and invasion of cancer cells (Fig. 2d). To address the clinical significance of *CMTM2*, we analyzed a publicly available database (KM plotter), which contains survival data for 240 cases of DGC and 269 cases of IGC²⁴. When we stratified the DGC patients into two groups based on the expression levels of *CMTM2*, the group with lower expression levels of *CMTM2* (CMCM2-low, $n = 61$) exhibited a worse prognostic outcome of overall survival (OS, Hazards ratio (HR) = 2.5, $P = 2.9 \times 10^{-7}$) than the group with higher *CMTM2* expression (CMTM2-high, $n = 179$). In contrast, IGC patients did not show a similar prognostic difference according to the expression status of *CMTM2* (HR = 1.35, $P = 0.11$). This result suggests that the *CMTM2* variation is pivotal in the progression of DGC but not in that of IGC (Fig. 2e).

Comparison of the mutation patterns between N0 and N3 tumors

To delineate distinct mutation profiles across DGC subgroups, we compared the mutation profiles of N0 and N3 DGCs. The N0 tumors had 635 mutated genes, while the N3 tumors had 863 mutated genes. Only 101 mutations (7.22%) were common to both N0 and N3 tumors, indicating distinct mutation profiles between the subtypes (Fig. 3a). The N0 and N3 tumors exhibited similar average mutation rates of 69.7 for N0 tumors and 74.31 for N3 tumors (Fig. 3b, Supplementary Fig. S2). Moreover, both N0 and N3 DGCs exhibited similar ratios of nonsynonymous vs. synonymous mutations (Fig. 3b, left) and frequencies of C > T/G > A mutations (N0, 67.58%; N3, 61.59%; Fig. 3b, right).

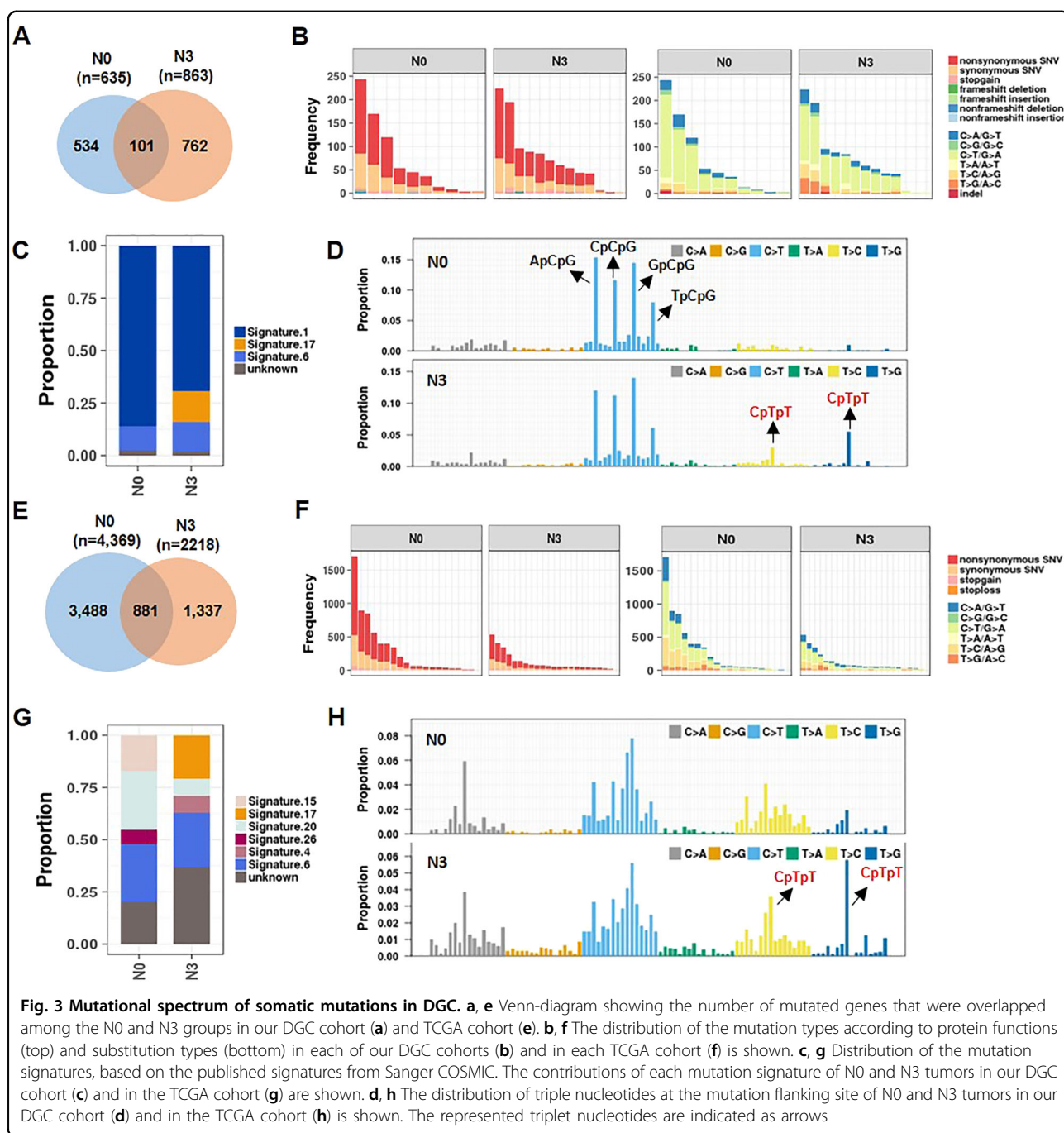
Functional enrichment analysis of the mutated genes demonstrated that distinct functions were enriched in the N0 and N3 tumors. The N0 tumors were enriched with cell adhesion (enrichment scores, ES = 5.02) and transport (ES = 2.11)-related genes, while the N3 tumors exhibited enriched mutations in cell development (ES = 4.14; Supplementary Table S3). This result also supports the view that the different mutation profiles between N0 and N3 tumors might play functional roles in the heterogeneous progression of DGC.

To further address the distinct mutation profiles in DGC, we performed a mutation signature analysis as described in the Materials and Methods. Mutation signature 1 was prevalently enriched in both N0 and N3 tumors. This signature is highly prevalent in various cancer types²⁵. We also found that mutation signature 17 was present in N3 tumors (14.82%) but not in N0 tumors (Fig. 3c), which may imply that mutation signature 17 is associated with tumor progression to lymph node metastasis. Mutation signature 17 has been found in various cancer types including gastric cancer¹⁹, but its etiology remains unknown²⁵.

Next, we performed mutation spectrum analyses for triple nucleotides to determine the characteristics of the mutation profiles. All the DGCs had frequent mutations of C > T at the NpCpG site (where N = A, T, G, or C; Fig. 3d), which is a congruent observation with the prevalent presentation of mutation signature 1. We also found that the N3 tumors had frequent T > C and T > G mutations at CpTpT sites, but the N0 tumors did not. This may support our result that N3 tumors, but not N0 tumors, exhibit mutation signature 17. To validate our findings, we also evaluated DGC data from TCGA. We obtained 43 cases of DGC with information on lymph nodal metastasis. N0 DGCs ($n = 20$) had 4369 mutated genes, N3 DGCs ($n = 23$) had 2218 mutated genes, and 881 mutations were commonly found in both subgroups (Fig. 3e). Differing from our data, DGC in the TCGA data exhibited frequent T < C/A > G mutations (23.82%; Fig. 3f). Preferential presentation of mutation signature 1 was not observed in the TCGA data (Fig. 3g). However, consistent with our data, mutation signature 17 including frequent T > C and T > G mutations at the CpTpT site was preferentially presented in the N3 tumors but not in the N0 tumors (Fig. 3g, h). Taken together, these findings suggest that mutation signature 17 is indicative of lymph nodal metastasis of DGC.

Identification of N3 tumor-specific recurrent mutations

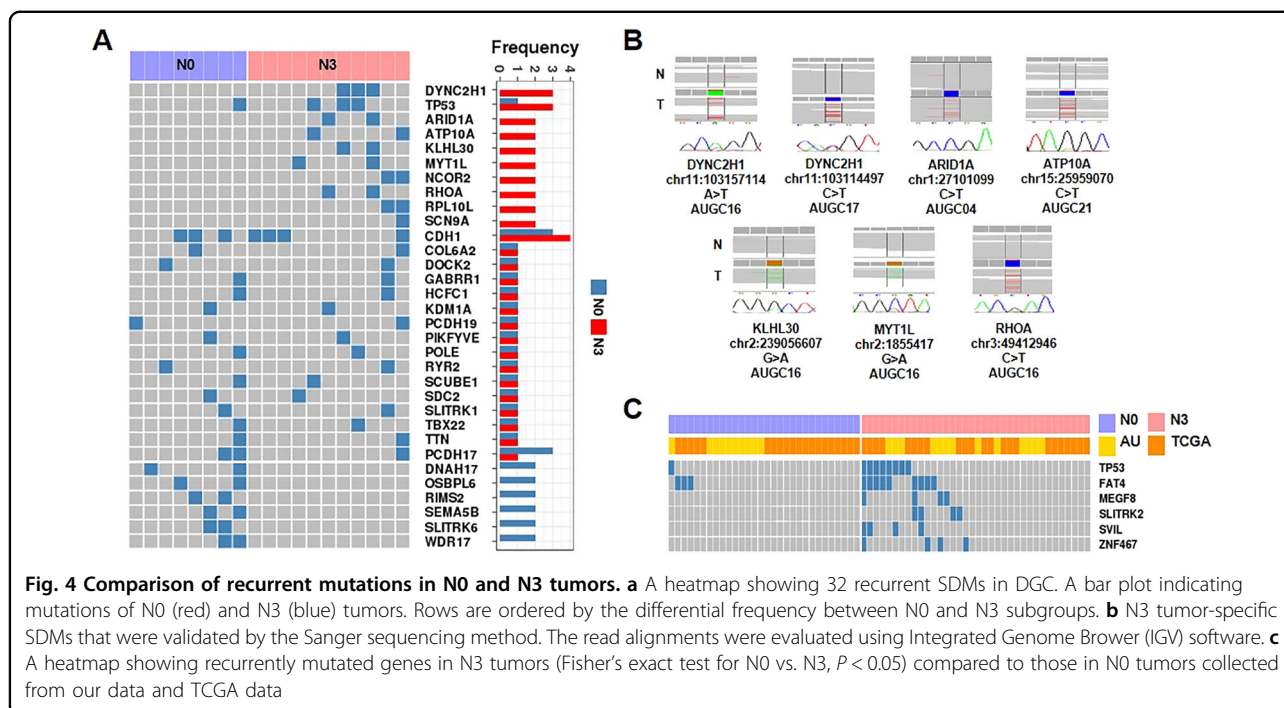
To pinpoint driver mutations for metastatic DGC, we identified significantly deleterious mutations (SDMs) using prediction algorithms for mutation effects on protein functions, as described in the Materials and Methods. This revealed 32 genes with recurrent SDMs, including previously known recurrent mutations, such as *CDH1*, *TP53*, and *RHOA* (Fig. 4a)^{10,15,23,26,27}. Of these, 6 mutations were N0-specific and 10 were N3-specific. Interestingly, previously known cancer genes including *TP53*, *ARID1A*, *NCOR2*, and *RHOA* were recurrently mutated, particularly in N3 tumors. We validated 7 out of 10 N3 tumor-specific mutations (i.e., *DYNC2H1*, *ARID1A*, *ATP10A*, *KLHL30*, *MYT1L* and *RHOA*) by Sanger sequencing analysis (Fig. 4b). In addition, combining our AU and TCGA data allowed the additional identification of recurrent mutations in genes in N3 tumors including *TP53*, *FAT4*, *MEGF8*, *SLITRK2*, *SVIL*, and *ZNF467*, in which mutations were not frequent in N0 tumors (Fisher's exact test for N0 vs. N3, $P < 0.05$; Fig. 4c). In particular, higher mutation rates of *TP53* (22.22%, $P = 0.006$) and *FAT4* (25%, $P = 0.025$) were observed in N3 tumors. Supporting our finding, *FAT4* has a tumor suppressive function and is associated with lymph node metastasis and poor prognostic outcomes in gastric cancer^{28,29}. We suggest that these N3-specific mutations may play critical roles in DGC progression to lymph nodal metastasis.



Aberrant DNA copy number alterations in DGC

DNA copy number aberration was estimated from WES data of the paired tumor and non-tumor tissues. Previously, DGC is known as a genomically stable (GS) subtype in TCGA¹⁵. Indeed, we identified only 17 CNAs from 6 samples (26.08%) (Fig. 5a). Recurrent copy number gains were found at 1q21.2–1q23.2 in N3 tumors of AUGC23 and AUGC01. This region encoded some oncogenes, such as *ARNT*, *MLLT11*, *MUC1*, and *NTRK1*, which may contribute to the progression of DGC.

Recurrent copy number losses were found at 7q31.31–36.3 in N0 (AUGC11) and N3 (AUGC17) tumors. Copy number gains were also found at 5p, 7p, 7q, 8p, 8q, and 18p, while copy number losses were found at 5q, 7q, 9p, 18p, and 18q. When we compared the CNVs between N0 and N3 tumors, N3 tumors exhibited more frequent aberrations of DNA copy number (5 of 13, 38%) than did N0 tumors (1 of 10, 10%). We also estimated loss of heterogeneity (LOH) by calculating allelic copy numbers from WES data. N3 tumors exhibited more frequent



copy-neutral LOH (7.62 per sample) than did N0 tumors (4.1 per sample, Fig. 5b). This result suggests that the N3 tumors had more frequent genomic aberrations than N0 tumors. Notwithstanding this difference between N0 and N3 tumors, we observed recurrent copy-neutral LOHs at 3p12.3-11.1 in both N0 and N3 tumors (Fig. 5c). This chromosomal locus included many tumor-suppressor genes, such as *ROBO1*, *ROBO2*, *CADM2*, *VGLL3*, and *EPHA3* (Fig. 5d). Thus, we suggest that the copy-neutral LOH at 3p12.3-11.1 might play pivotal roles in tumor progression, as described elsewhere^{30,31}.

Discussion

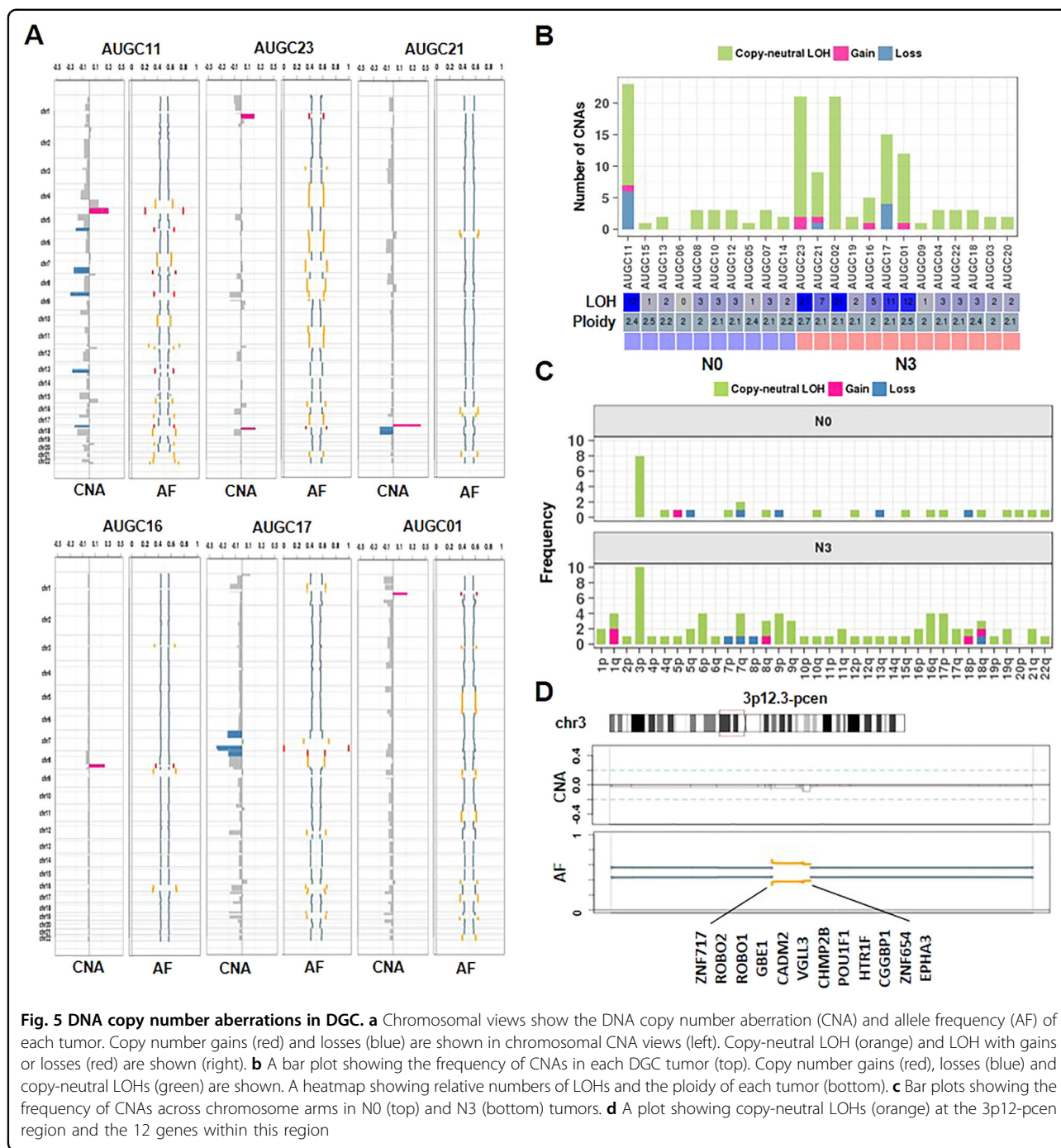
In the present study, by performing WES in DGCs, we identified 185 recurrently mutated genes, including *CDH1*, *RHOA*, *TP53*, and *CMTM2*. *CDH1*, *RHOA*, and *TP53* were frequently mutated in DGCs, and a mutation of *CDH1* (E-cadherin) in DGC has been shown previously. Dysregulation of E-cadherin contributes to tumor invasion and progression³². Somatic mutations of *RHOA* are highly specific to DGC and are absent in IGC^{10,14}. Indeed, *RHOA* mutations have been found in up to 25.3% of DGCs in Japan¹⁰, 15% in DGCs from TCGA data¹⁵, and 14.3% in DGCs in a Hong Kong population¹⁴. However, the frequency of *RHOA* mutations in DGC appears to be quite low in Koreans. No *RHOA* mutation was found in the 14 cases of DGCs²³, and we found only two *RHOA* mutations in DGC (8.6%). Thus, the frequencies of *RHOA* mutations differ significantly across studies from Eastern and Western

countries, reflecting the heterogeneity of DGCs with diverse genetic backgrounds.

In addition to the previously known mutations, we found a novel mutation at *CMTM2* (Table 2), and for the first time, we demonstrated its effect on cellular proliferation. *CMTM2* is a member of a chemokine-like factor superfamily that regulates vesicular transport or membrane apposition events in the endoplasmic reticulum³³. The DGC patients with lower *CMTM2* expression had a shorter overall survival (Fig. 2). Indeed, other *CMTM* family genes (e.g., *CMTM3*, *CMTM4*, *CMTM7*, and *CMTM8*) have tumor suppressive functions^{34–36}. For example, knockdown of *CMTM3* can promote gastric cancer metastasis through the activation of STAT3/ Twist1/EMT³⁷. Thus, we suggest that *CMTM2* mutations may play a pivotal role, especially in DGC progression.

Profiling of DNA copy number aberrations in DGC might identify recurrent copy-neutral LOH at the 3p12-pcen region, which harbors many tumor-suppressor genes. For example, *ROBO1* and *ROBO2* at 3p12.3 are regulators of axon guidance, and the expression levels of these genes are frequently suppressed by methylation and/or LOH in many cancer types³⁸. *CADM2*³⁹ and *VGLL3*⁴⁰ at 3p12.1 also have tumor suppressive functions. Collectively, these results imply that the tumor-suppressor genes at 3p12-pcen are impaired by copy-neutral LOH in DGC.

Comparison of the N0 and N3 tumors revealed distinct aberrations of DNA copy numbers and somatic mutational patterns between the subtypes. The N3 tumors



exhibited more frequent DNA copy number aberrations than did the N0 tumors (see Fig. 5). GC patients with lymph node metastasis are reported to have more frequent DNA copy number gains and losses than patients without metastasis⁴¹. In addition, integrating the analysis of CNA and mutation profiles allowed us to identify N3 tumors with recurrent copy-neutral LOHs at 6p (*PBX2*, *AIF1*, *ZFP57*, and *PGBD1*), 9p (*ACER2*, *DENND4C*, *CCDC171*, *PTPRD*, and *GLIS3*), and 17p (*TP53*, *RAI1*, and

SEN3), although this analysis did not allow us to identify N0 tumors with these features.

Mutational signature analysis also revealed unique features. N3 tumors preferentially had mutation signature 17 with frequent T > C and T > G mutations at the CpTpT site, but the N0 tumors did not. Recurrent T > G mutations at TpT dinucleotides converting CpTpT to CpGpT have been found in esophageal adenocarcinomas⁴² and GCs¹¹. We also identified significantly deleterious

mutations (SDMs) including some in previously known cancer genes (e.g., *TP53*, *ARID1A*, *NCOR2*, and *RHOA*). Frequent mutations of *TP53* have been shown in DGCs. *ARID1A*, encoding a member of the SWI-SNF chromatin remodeling family, has been suggested to act as a tumor-suppressor gene by regulating the cell cycle or promoting apoptosis in various cancers⁴³. Recently, inactivating mutations or protein deficiency of *ARID1A* have frequently been observed in gastric cancers^{29,44}. In addition, a mutation of *DYNC2H1* was newly identified as an N3-specific SDM. Mutations of *DYNC2H1* have been associated with the progression of fibromuscular dysplasia⁴⁵ and short-rib polydactyly syndrome;⁴⁶ however, its cancer-related functions have not been elaborated thoroughly. In addition, an integrated analysis of our data and the TCGA data identified additional N3-specific recurrent mutations (i.e., *TP53* and *FAT4*). *FAT4*, a cadherin-related protein, has been proposed to have a tumor-suppressor role, and its mutation has been noticed in GC. Silencing of *FAT4* promoted cell proliferation, invasion, and migration in GC cells^{28,47,48}. Moreover, the downregulation of *FAT4* expression has been associated with lymph node metastasis and poor survival in GC patients²⁸. We suggest that these differential mutations between N0 and N3 tumors may play crucial roles in DGC progression; thus, they warrant further elucidation.

In conclusion, in this study, we demonstrated somatic mutations and DNA copy number aberrations of DGCs that differ according to the presence or absence of lymph node metastasis. Our analysis can provide an enhanced roadmap for delineating the heterogeneous genomic and clinical progression of DGCs, which might be helpful in developing precision diagnostics and targeted therapeutics for DGC.

Acknowledgements

The biological specimens for this study were provided by the Ajou University Human Bio-Resource Bank (AHBB), a member of Korea Biobank Network, which is supported by the Ministry of Health and Welfare. All samples derived from the National Biobank of Korea were obtained with informed consent under institutional review board-approved protocols. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea, which is funded by the Ministry of Science and ICT (2014R1A1A1007905 and 2017R1C1B2003970) and by the Korea Health Industry Development Institute (KHIDI), which is funded by the Ministry of Health & Welfare of the Republic of Korea (HI15C1551).

Author details

¹Department of Physiology, Ajou University School of Medicine, Suwon, Republic of Korea. ²Graduate School of Biomedical Science, Ajou University School of Medicine, Suwon, Republic of Korea. ³Department of Pathology, Ajou University School of Medicine, Suwon, Republic of Korea. ⁴Department of Surgery, Ajou University School of Medicine, Suwon, Republic of Korea

Authors' contributions

H.G.W. and D.L. designed the study. H.G.W. supervised the exome sequencing. J.-H.C. and H.G.W. performed computational analyses. D.L. and Y.B.K. carried out the pathology review. M.J.K. performed experiments for validating the mutations. J.M.A. and W.J.B. performed functional experiments. S.-U.H. coordinated sample acquisition and provided clinical information. J.-H.C.,

H.G.W., and D.L. wrote the manuscript. J.-H.C., Y.-B.K., H.G.W., and D.L. were involved in critical review and discussion.

Conflict of interest

The authors declare that they have no conflict of interest.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Supplementary information accompanies this paper at <https://doi.org/10.1038/s12276-017-0009-6>.

Received: 24 July 2017 Revised: 29 October 2017 Accepted: 8 November 2017.

Published online: 6 April 2018

References

1. Ferlay, J. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* **136**, E359–E386 (2015).
2. Peleteiro, B., Severo, M., La Vecchia, C. & Lunet, N. Model-based patterns in stomach cancer mortality worldwide. *Eur. J. Cancer Prev.* **23**, 524–531 (2014).
3. Lauren, P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol. Microbiol. Scand.* **64**, 31–49 (1965).
4. Henson, D. E., Dittus, C., Younes, M., Nguyen, H. & Albores-Saavedra, J. Differential trends in the intestinal and diffuse types of gastric carcinoma in the United States, 1973–2000: increase in the signet ring cell type. *Arch. Pathol. Lab. Med.* **128**, 765–770 (2004).
5. Novotny, A. R. et al. Predicting individual survival after gastric cancer resection: validation of a U.S.-derived nomogram at a single high-volume center in Europe. *Ann. Surg.* **243**, 74–81 (2006).
6. Cunningham, S. C. et al. Survival after gastric adenocarcinoma resection: eighteen-year experience at a single institution. *J. Gastrointest. Surg.* **9**, 718–725 (2005).
7. Esaki, Y., Hirayama, R. & Hirokawa, K. A comparison of patterns of metastasis in gastric cancer by histologic type and age. *Cancer* **65**, 2086–2090 (1990).
8. Takagane, A. et al. Evaluation of the ratio of lymph node metastasis as a prognostic factor in patients with gastric cancer. *Gastric Cancer* **2**, 122–128 (1999).
9. Ding, Y. B. et al. Association of VCAM-1 overexpression with oncogenesis, tumor angiogenesis and metastasis of gastric carcinoma. *World J. Gastroenterol.* **9**, 1409–1414 (2003).
10. Kakiuchi, M. et al. Recurrent gain-of-function mutations of RHOA in diffuse-type gastric carcinoma. *Nat. Genet.* **46**, 583–587 (2014).
11. Wang, K. et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* **46**, 573–582 (2014).
12. Cristescu, R. et al. Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. *Nat. Med.* **21**, 449–456 (2015).
13. Holbrook, J. D. et al. Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine. *J. Transl. Med.* **9**, 119 (2011).
14. Wong, S. S. et al. Genomic landscape and genetic heterogeneity in gastric adenocarcinoma revealed by whole-genome sequencing. *Nat. Commun.* **5**, 5477 (2014).
15. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
16. Sargent, D. J. et al. Defective mismatch repair as a predictive marker for lack of efficacy of fluorouracil-based adjuvant therapy in colon cancer. *J. Clin. Oncol.* **28**, 3219–3226 (2010).
17. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
18. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
19. Rosenthal, R., McGranahan, N., Herrero, J., Taylor, B. S. & Swanton, C. DeconstructSigs: delineating mutational processes in single tumors distinguishes

- DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol.* **17**, 31 (2016).
20. Favero, F. et al. Sequenza: allele-specific copy number and mutation profiles from tumor sequencing data. *Ann. Oncol.* **26**, 64–70 (2015).
 21. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
 22. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
 23. Lee, Y. S. et al. Genomic profile analysis of diffuse-type gastric cancers. *Genome Biol.* **15**, R55 (2014).
 24. Szasz, A. M. et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. *Oncotarget* **7**, 49322–49333 (2016).
 25. Alexandrov, L. B. et al. Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
 26. Corso, G. et al. Somatic mutations and deletions of the E-cadherin gene predict poor survival of patients with gastric cancer. *J. Clin. Oncol.* **31**, 868–875 (2013).
 27. Becker, K. F. et al. E-cadherin gene mutations provide clues to diffuse type gastric carcinomas. *Cancer Res.* **54**, 3845–3852 (1994).
 28. Cai, J. et al. FAT4 functions as a tumour suppressor in gastric cancer by modulating Wnt/beta-catenin signalling. *Br. J. Cancer* **113**, 1720–1729 (2015).
 29. Zang, Z. J. et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat. Genet.* **44**, 570–574 (2012).
 30. Cody, N. A. et al. Characterization of the 3p12.3-pcen region associated with tumor suppression in a novel ovarian cancer cell line model genetically modified by chromosome 3 fragment transfer. *Mol. Carcinog.* **48**, 1077–1092 (2009).
 31. Lee, D. J. et al. Multiple tumor-suppressor genes on chromosome 3p contribute to head and neck squamous cell carcinoma tumorigenesis. *Cancer Biol. Ther.* **10**, 689–693 (2010).
 32. Vleminckx, K., Vakaet, L. Jr., Mareel, M., Fiers, W. & van Roy, F. Genetic manipulation of E-cadherin expression by epithelial tumor cells reveals an invasion suppressor role. *Cell* **66**, 107–119 (1991).
 33. Liu, G. et al. Expression and localization of CKLF5F2 in human spermatogenesis. *Asian J. Androl.* **9**, 189–198 (2007).
 34. Plate, M. et al. Identification and characterization of CMTM4, a novel gene with inhibitory effects on HeLa cell growth through inducing G2/M phase accumulation. *Mol. Cells* **29**, 355–361 (2010).
 35. Zhang, W. et al. Down-regulation of CMTM8 induces epithelial-to-mesenchymal transition-like changes via c-MET/extracellular signal-regulated kinase (ERK) signaling. *J. Biol. Chem.* **287**, 11850–11858 (2012).
 36. Li, H. et al. A novel 3p22.3 gene CMTM7 represses oncogenic EGFR signaling and inhibits cancer cell growth. *Oncogene* **33**, 3109–3118 (2014).
 37. Yuan, W. et al. Knockdown of CMTM3 promotes metastasis of gastric cancer via the STAT3/Twist1/EMT signaling pathway. *Oncotarget* **7**, 29507–29519 (2016).
 38. Dallol, A. et al. Tumour specific promoter region methylation of the human homologue of the Drosophila Roundabout gene DUTT1 (ROBO1) in human cancers. *Oncogene* **21**, 3020–3028 (2002).
 39. He, W. et al. Aberrant methylation and loss of CADM2 tumor suppressor expression is associated with human renal cell carcinoma tumor progression. *Biochem. Biophys. Res. Commun.* **435**, 526–532 (2013).
 40. Hong, S. J. et al. The gene-reduction effect of chromosomal losses detected in gastric cancers. *BMC Gastroenterol.* **10**, 138 (2010).
 41. Liang, L., Fang, J. Y. & Xu, J. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. *Oncogene* **35**, 1475–1482 (2016).
 42. Dulak, A. M. et al. Exome and whole-genome sequencing of esophageal adenocarcinoma identifies recurrent driver events and mutational complexity. *Nat. Genet.* **45**, 478–486 (2013).
 43. Wu, R. C., Wang, T. L. & Shih, I. M. The emerging roles of ARID1A in tumor suppression. *Cancer Biol. Ther.* **15**, 655–664 (2014).
 44. Wang, K. et al. Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat. Genet.* **43**, 1219–1223 (2011).
 45. Kiando, S. R. et al. Exome sequencing in seven families and gene-based association studies indicate genetic heterogeneity and suggest possible candidates for fibromuscular dysplasia. *J. Hypertens.* **33**, 1802–1810 (2015).
 46. Merrill, A. E. et al. Ciliary abnormalities due to defects in the retrograde transport protein DYNC2H1 in short-rib polydactyly syndrome. *Am. J. Hum. Genet.* **84**, 542–549 (2009).
 47. Ma, L. et al. Fat4 suppression induces Yap translocation accounting for the promoted proliferation and migration of gastric cancer cells. *Cancer Biol. Ther.* **17**, 36–47 (2016).
 48. Jung, H. Y. et al. Loss of FAT atypical cadherin 4 expression is associated with high pathologic T stage in radically resected gastric cancer. *J. Gastric Cancer* **15**, 39–45 (2015).