Article

Molecular Network Profiling in Intestinal- and Diffuse-Type Gastric Cancer

Shihori Tanabe [1,*], Sabina Quader [2], Ryuichi Ono [3], Horacio Cabral [4], Kazuhiko Aoyagi [5], Akihiko Hirose [1], Hiroshi Yokozaki [6] and Hiroki Sasaki [7]

[1] Division of Risk Assessment, Center for Biological Safety and Research, National Institute of Health Sciences; stanabe@nihs.go.jp
[2] Innovation Centre of NanoMedicine (iCONM), Kawasaki Institute of Industrial Promotion; sabina-q@kawasaki-net.ne.jp
[3] Division of Cellular and Molecular Toxicology, Center for Biological Safety and Research, National Institute of Health Sciences; onoryu@nihs.go.jp
[4] Department of Bioengineering, Graduate School of Engineering, The University of Tokyo; horacio@bmw.t.u-tokyo.ac.jp
[5] Department of Clinical Genomics, National Cancer Center Research Institute; kaaoyagi@ncc.go.jp
[6] Department of Pathology, Kobe University of Graduate School of Medicine; hyoko@med.kobe-u.ac.jp
[7] Department of Translational Oncology, National Cancer Center Research Institute; hksasaki@ncc.go.jp
* Correspondence: stanabe@nihs.go.jp; Tel.: +81-44-270-6686

Simple Summary: Cancer has several phenotypic subtypes where the responsiveness towards drugs or capacity of migration or recurrence are different. The molecular networks are dynamically altered in various phenotypes of the cancer. To reveal the network pathways in epithelial-mesenchymal transition (EMT), we have profiled gene expression in mesenchymal stem cells and diffuse-type gastric cancer (GC), as well as intestinal-type GC. Gene expression signatures revealed the molecular pathway networks altered in intestinal- and diffuse-type GC. The artificial intelligence (AI) recognized the differences in molecular network pictures of intestinal- and diffuse-type GC.

Abstract: Epithelial-mesenchymal transition (EMT) plays an important role in the acquisition of cancer stem cell (CSC) feature and drug resistance, which are the main hallmarks of cancer malignancy. Although previous findings have shown that several signaling pathways are activated in cancer progression, the precise mechanism of signaling pathways in EMT and CSCs are not fully understood. In this study, we focused on the intestinal and diffuse-type gastric cancer (GC), and analyzed the gene expression of public RNAseq data to understand the molecular pathway regulation in different subtypes of gastric cancer. Network pathway analysis was performed by Ingenuity Pathway Analysis (IPA). Total 2815 probe set IDs were significantly different between intestinal- and diffuse-type GC data in cBioPortal Cancer Genomics. The 10 genes including *male-specific lethal 3 homolog (Drosophila) pseudogene 1* (*MSL3P1*), *CDC28 protein kinase regulatory subunit 1B* (*CKS1B*), *DEAD-box helicase 27* (*DDX27*), *golgi to ER traffic protein 4* (*GET4*), *chromosome segregation 1 like* (*CSE1L*), *translocase of outer mitochondrial membrane 34* (*TOMM34*), *YTH N6-methyladenosine RNA binding protein 1* (*YTHDF1*), *ribonucleic acid export 1* (*RAE1*), *par-6 family cell polarity regulator beta* (*PARD6B*), and *MRG domain binding protein* (*MRGBP*) were found to have difference in gene expression in intestinal- and diffuse-type GC. Total 463 direct relationships with 3 molecules (MYC, NTRK1, UBE2M) were found in the biomarker-filtered network generated by network pathway analysis. The networks and features in intestinal- and diffuse-type GC have been investigated and profiled in bioinformatics. Our results revealed the signaling pathways networks in intestinal- and diffuse-type GC, bringing new light for the elucidation of drug resistance mechanisms in CSCs.

Keywords: cancer stem cell; epithelial-mesenchymal transition; molecular network

1. Introduction

Different cell types show a variety of molecular networks. Gastric cancer (GC) has several subtypes, which includes intestinal- and diffuse-type GC [1, 2]. Intestinal-type GC has a trend to be more rigid. In contrast, diffuse-type GC has a tendency to be more loose or sparse, which confers the diffuse-type GC malignant property and the migration capacity to the secondary site of the cancer. It is important to distinguish the subtypes of GC, since the prognosis is different, and the anti-cancer drug resistance may also be involved in diffuse-type GC [3]. Thus, the therapeutic strategy may be different in each subtypes of GC. Although the gene mutations of *CDH1* and *RHOA* distinguished gastric cancer from colorectal and esophageal tumors, and these mutations were specific to diffuse-type GC, it is still challenging to discriminate the intestinal-type and diffuse-type GC in molecular gene expression networks [4]. We have previously revealed that the mRNA ratios of *CDH2* to *CDH1* distinguish the intestinal- and diffuse-type GC [2]. Epithelial-mesenchymal transition (EMT) is associated with malignancy of GC and diffuse-type GC [5]. EMT is one of the important features in cancer stem cells (CSCs), which play an important role in drug resistance and are the therapeutic target [6]. To reveal the network pathways in EMT, we have profiled gene expression and networks in mesenchymal stem cells and diffuse-type GC, as well as intestinal-type GC [2, 7]. To better understand the pathogenesis of GC and treat EMT-like malignant diffuse-type GC, it is essential to know and predict the network pathway difference between intestinal- and disuse-type GC.

The importance and potential to use the molecular network profile to distinguish diffuse- and intestinal-type GC are increasing in digital era. The previous study clearly demonstrated that the gene regulatory network construction identified nuclear transcription factor Y subunit alpha (NFYA) as a prognostic factor in diffuse-type GC [8]. Recent progress in computational analysis and public databases enables multi-disciplinary assessment for big data, including network analysis of the RefSeq data. In this study, the open-sourced RefSeq data of intestinal- and diffuse-type GC were compared, followed by molecular network analysis and gene ontology analysis. In the meantime, the prediction modeling utilizing Artificial Intelligence (AI) for the molecular networks has been established. This research is integrating the gene expression, molecular networks and AI for the future networking.

2. Results

2.1. Genes altered in intestinal- and diffuse-type GC

Genes altered in intestinal- and diffuse-type GC were analyzed in CIN type and GS type samples in TCGA RNAseq data. Table 1 shows top 10 genes altered in intestinal- and diffuse-type GC. The top 10 genes include *male-specific lethal 3 homolog (Drosophila) pseudogene 1* (*MSL3P1*), *CDC28 protein kinase regulatory subunit 1B* (*CKS1B*), *DEAD-box helicase 27* (*DDX27*), *golgi to ER traffic protein 4* (*GET4*), *chromosome segregation 1 like* (*CSE1L*), *translocase of outer mitochondrial membrane 34* (*TOMM34*), *YTH N6-methyladenosine RNA binding protein 1* (*YTHDF1*), *ribonucleic acid export 1* (*RAE1*), *par-6 family cell polarity regulator beta* (*PARD6B*), and *MRG domain binding protein* (*MRGBP*). Gene expression profile of the top 10 genes in intestinal- and diffuse-type GC are shown in Figure 1. Total 2815 IDs were significantly altered in intestinal- and diffuse-type GC (t-test, $p < 0.00001$) (Supplementary Table 1).

Figure 1. Gene expression profile of top 10 genes altered in intestinal- and diffuse-type gastric cancer (GC). The gene expression of top 10 genes which have significant difference between CIN (chromosomal instability; intestinal-type) and GS (genomically stable; diffuse-type) gastric cancer (GC) in TCGA RNAseq data are shown in Tableau visualization.

Table 1. Top 10 genes altered in intestinal- and diffuse-type gastric cancer (GC). The top 10 genes which have significant difference between CIN (chromosomal instability; intestinal-type) and GS (genomically stable; diffuse-type) in TCGA RNAseq data are shown. Total 2815 probe set IDs were significantly different between CIN and GS (Student's t-test, $p < 0.00001$). Gene ontology of the 10 genes are shown from DAVID analysis.

Gene Symbol	Gene Name	GOTERM_BP_DIRECT
MSL3P1	male-specific lethal 3 homolog (Drosophila) pseudogene 1	GO:0006338~chromatin remodeling,GO:0006342~chromatin silencing,GO:0006351~transcription, DNA-templated,GO:0016575~histone deacetylation,GO:0043967~histone H4 acetylation,GO:0043968~histone H2A acetylation,
CKS1B	CDC28 protein kinase regulatory subunit 1B	GO:0007049~cell cycle,GO:0007346~regulation of mitotic cell cycle,GO:0008283~cell proliferation,GO:0044772~mitotic cell cycle phase transition,GO:0045737~positive regulation of cyclin-dependent protein serine/threonine kinase activity,GO:0045893~positive regulation of transcription, DNA-templated,GO:0051301~cell division,
DDX27	DEAD-box helicase 27	GO:0006364~rRNA processing,GO:0010501~RNA secondary structure unwinding,
GET4	golgi to ER traffic protein 4	GO:0006810~transport,GO:0051220~cytoplasmic sequestering of protein,GO:0071816~tail-anchored membrane protein insertion into ER membrane,GO:1904378~maintenance of unfolded protein involved in ERAD pathway,
CSE1L	chromosome segregation 1 like	GO:0006606~protein import into nucleus,GO:0006611~protein export from nucleus,GO:0006915~apoptotic process,GO:0008283~cell proliferation,
TOMM34	translocase of outer mitochondrial membrane 34	GO:0006626~protein targeting to mitochondrion,
YTHDF1	YTH N6-methyladenosine RNA binding protein 1	GO:0045948~positive regulation of translational initiation,
RAE1	ribonucleic acid export 1	GO:0000972~transcription-dependent tethering of RNA polymerase II gene DNA at nuclear periphery,GO:0006406~mRNA export from nucleus,GO:0006409~tRNA export from nucleus,GO:0006606~protein import into nucleus,GO:0007077~mitotic nuclear envelope disassembly,GO:0010827~regulation of glucose transport,GO:0016032~viral process,GO:0016925~protein sumoylation,GO:0019083~viral transcription,GO:0031047~gene silencing by RNA,GO:0071407~cellular response to organic cyclic compound,GO:0075733~intracellular transport of virus,GO:1900034~regulation of cellular response to heat,
PARD6B	par-6 family cell polarity regulator beta	GO:0006461~protein complex assembly,GO:0007043~cell-cell junction assembly,GO:0007049~cell cycle,GO:0007163~establishment or maintenance of cell polarity,GO:0007409~axonogenesis,GO:0030334~regulation of cell migration,GO:0051301~cell division,GO:0070830~bicellular tight junction assembly,
MRGBP	MRG domain binding protein	GO:0006351~transcription, DNA-templated,GO:0006357~regulation of transcription from RNA polymerase II promoter,GO:0016573~histone acetylation,GO:0040008~regulation of growth,

2.2. Networks generated from genes altered in intestinal- and diffuse-type GC

Networks of genes altered in intestinal- and diffuse-type GC were analyzed using IPA. Total 2815 IDs which had significant difference between intestinal- and diffuse-type gastric cancer were analyzed in Ingenuity Pathway Analysis (t-test, $p < 0.00001$). Total 25 networks generated from genes

which have significant difference between intestinal- and diffuse-type GC are shown in Table 2. The Network #1 which is related to cancer, gastrointestinal disease, organismal injury and abnormalities is shown in Figure 2.

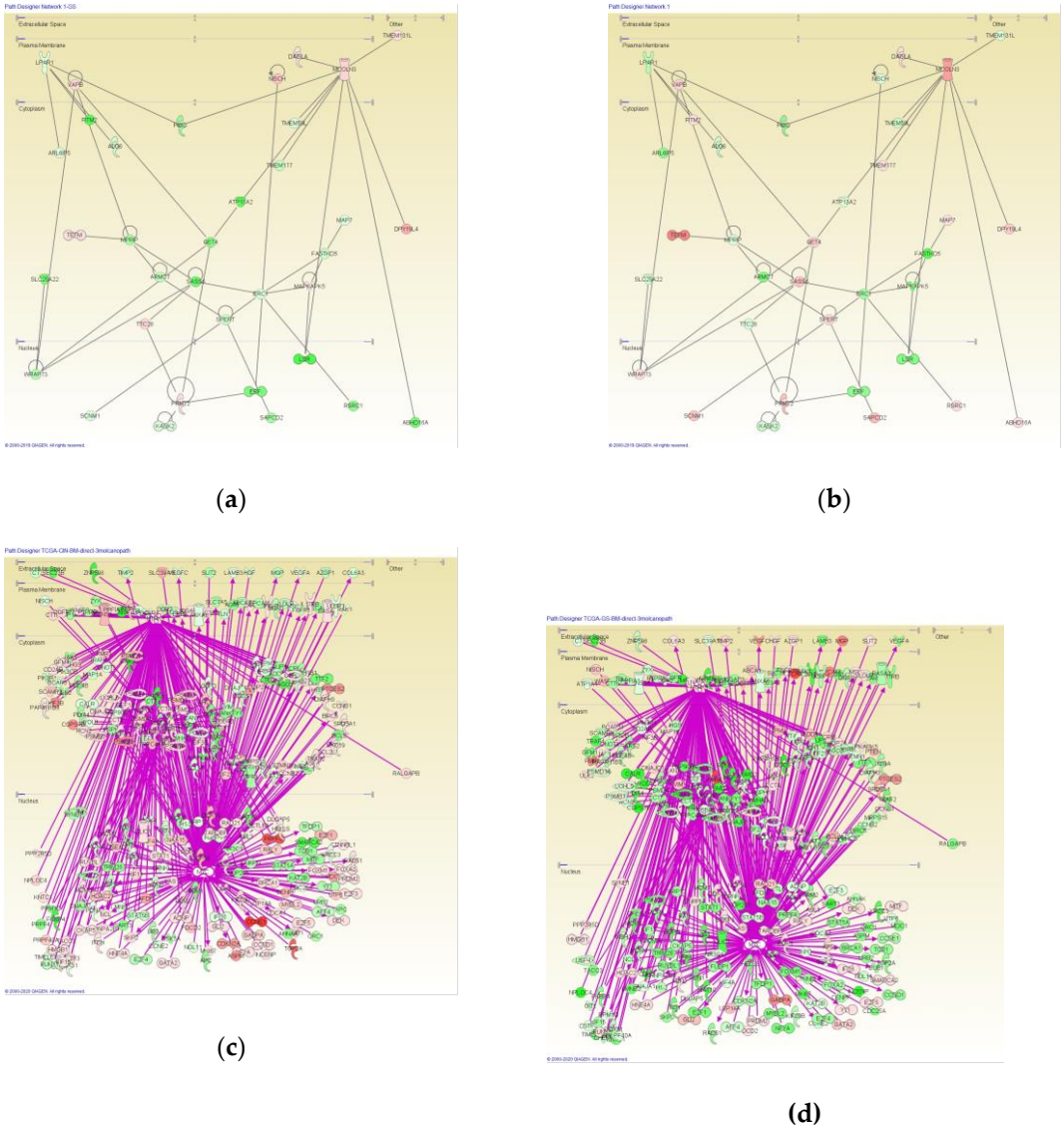


Figure 2. Networks generated from genes altered in intestinal- and diffuse-type gastric cancer (GC). Total 2815 IDs which had significant difference between intestinal- and diffuse-type GC were analyzed in IPA, and Network 1 related to cancer, gastrointestinal disease, organismal injury and abnormalities is shown. (**a**) Network in intestinal-type GC; (**b**) Network in diffuse-type GC. Total 463 direct relationships with 3 molecules (MYC, NTRK1, UBE2M) are shown in the network of biomarker-filtered genes in intestinal-type GC (**c**) and diffuse-type GC (**d**). From 613 genes biomarker-filtered (human, blood, cancer), 285 genes including MYC, NTRK1 and UBE2M are included in the network. All relationships were 609.

Table 2. Networks generated from genes which have significant difference between intestinal- and diffuse-type gastric cancer (GC). The networks were generated from total 2815 probe set IDs differentiated between CIN (intestinal-type) and GS (diffuse-type) gastric cancer (GC) (Student's t-test, $p < 0.00001$).

ID	Focus Molecules	Top Diseases and Functions
1	35	Cancer, Gastrointestinal Disease, Organismal Injury and Abnormalities
2	35	Amino Acid Metabolism, Molecular Transport, Small Molecule Biochemistry
3	34	Cardiovascular Disease, Gene Expression, Protein Synthesis
4	34	Developmental Disorder, Hereditary Disorder, Neurological Disease
5	34	Dental Disease, Dermatological Diseases and Conditions, Post-Translational Modification
6	34	Hereditary Disorder, Infectious Diseases, RNA Post-Transcriptional Modification
7	34	Carbohydrate Metabolism, Lipid Metabolism, Post-Translational Modification
8	34	Connective Tissue Disorders, Developmental Disorder, Hereditary Disorder
9	34	Cell Cycle, Molecular Transport, Protein Trafficking
10	33	Connective Tissue Disorders, Dermatological Diseases and Conditions, Developmental Disorder
11	33	Cell Morphology, Cellular Assembly and Organization, Cellular Function and Maintenance
12	33	Gene Expression, Post-Translational Modification, RNA Damage and Repair
13	33	Cell Cycle, Cellular Growth and Proliferation, Reproductive System Development and Function
14	32	Infectious Diseases, Molecular Transport, Post-Translational Modification
15	32	Cell Cycle, Cellular Assembly and Organization, DNA Replication, Recombination, and Repair
16	32	Developmental Disorder, Hereditary Disorder, Molecular Transport
17	32	Carbohydrate Metabolism, Nucleic Acid Metabolism, Small Molecule Biochemistry
18	31	Cellular Assembly and Organization, Cellular Response to Therapeutics, DNA Replication, Recombination, and Repair
19	31	Developmental Disorder, Lipid Metabolism, Small Molecule Biochemistry
20	31	Cell Morphology, Cellular Assembly and Organization, Skeletal and Muscular System Development and Function
21	31	Cancer, Cellular Assembly and Organization, Skeletal and Muscular Disorders
22	31	Cell Cycle, Cellular Assembly and Organization, Cellular Compromise
23	31	Molecular Transport, RNA Post-Transcriptional Modification, RNA Trafficking
24	31	Nervous System Development and Function, Neurological Disease, Organ Morphology
25	31	Gene Expression, Neurological Disease, Organismal Functions

2.3. Regulator effect networks related to cancer in intestinal- and diffuse-type GC

Regulator effects were analyzed by Ingenuity Pathway Analysis (IPA). The target disease was selected as cancer in the analysis. Type of regulators analyzed include biological drug, canonical pathway, chemical drug (Figure 3). Table 3 shows regulator effect networks related to cancer in intestinal-type GC. Regulator effect networks related to cancer have been generated. Table 4 show regulator effect networks related to cancer in diffuse-type GC.

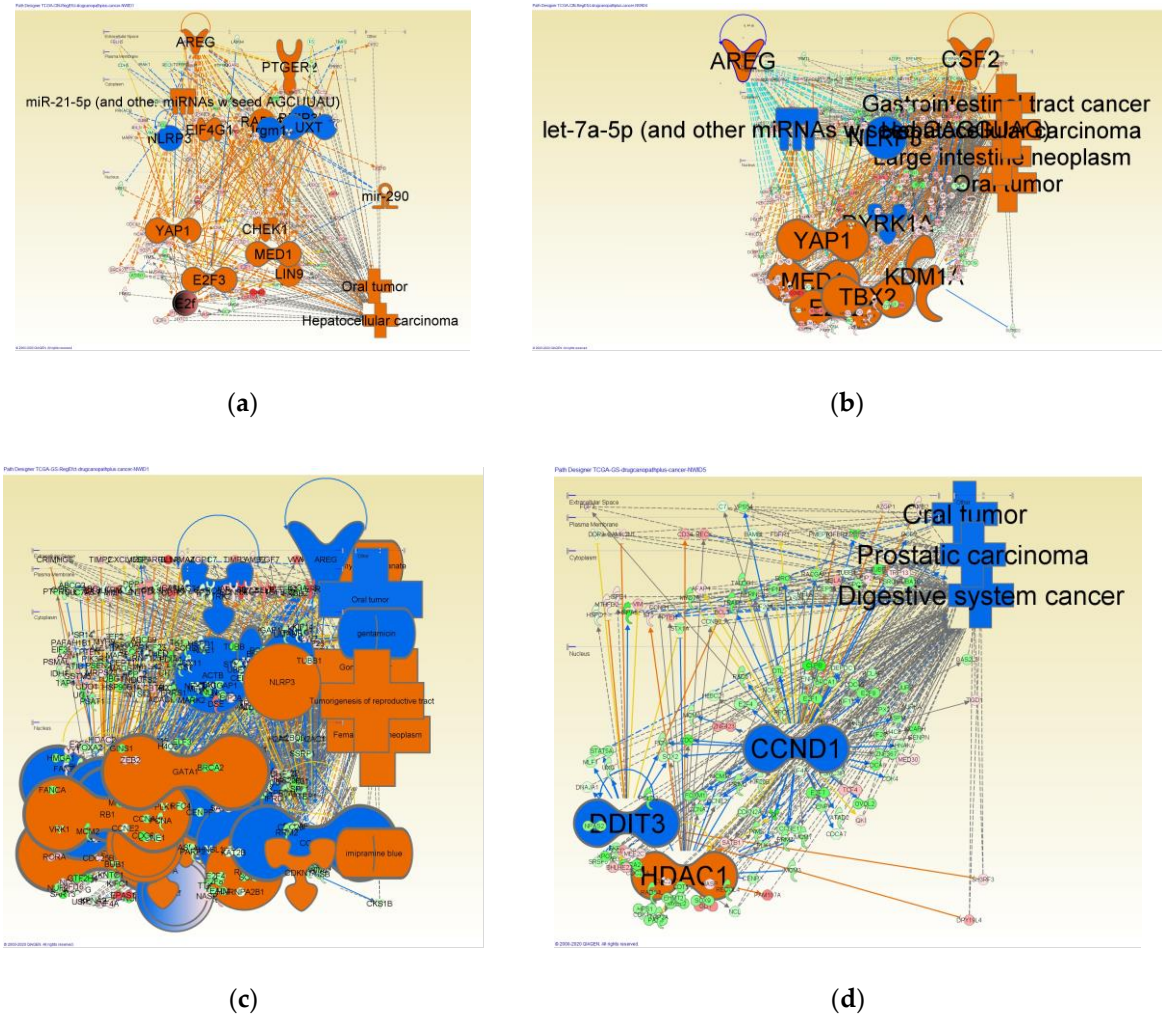


Figure 3. Networks for regulator effects related to cancer in intestinal- and diffuse-type gastric cancer (GC). Regulator effects were analyzed by IPA. The target disease was selected as cancer in the analysis. Type of regulators analyzed include biological drug, canonical pathway, chemical drug. (**a**) Regulator effect network ID1 (Hepatocellular carcinoma, Oral tumor) related to cancer in intestinal-type GC; (**b**) Regulator effect network ID4 (Gastrointestinal tract cancer, Hepatocellular carcinoma, Large intestine neoplasm, Oral tumor) related to cancer in intestinal-type GC; (**c**) Regulator effect network ID1 (Female genital neoplasm, Gonadal tumor, Oral tumor, Tumorigenesis of reproductive tract) related to cancer in diffuse-type GC; (**d**) Regulator effect network ID5 (Digestive system cancer, Oral tumor, Prostatic carcinoma) related to cancer in diffuse-type GC.

Table 3. Regulator effect networks related to cancer in intestinal-type gastric cancer (GC). Regulator effect networks related to cancer have been generated. Type of regulators include biological drug, canonical pathway, chemical drug.

ID	Regulators	Target Total	Diseases & Functions
1	AREG,BNIP3L,CHEK1,E2f,E2F3,EIF4G1,Irgm1,LIN9,MED1,miR-21-5p (and other miRNAs w/seed AGCUUAU),mir-290,NLRP3,PTGER2,RABL6,UXT,YAP1	94	Hepatocellular carcinoma,Oral tumor
2	AREG,ERG,KDM5B,MIR17HG,TFDP1,YAP1	123	Hepatocellular carcinoma,Intestinal cancer,Large intestine neoplasm
3	AREG,KDM5B,miR-21-5p (and other miRNAs w/seed AGCUUAU),mir-290,MIR17HG,PTGER2,SMARCB1,TCF3,UXT,YAP1	70	Hepatocellular carcinoma
4	AREG,CSF2,DYRK1A,E2F2,KDM1A,let-7a-5p (and other miRNAs w/seed GAGGUAG),MED1,NLRP3,TBX2,YAP1	200	Gastrointestinal tract cancer,Hepatocellular carcinoma,Large intestine neoplasm,Oral tumor
5	MYCN	3	Cell death of osteosarcoma cells
6	EGFR,ERBB2,HRAS,miR-205-5p (and other miRNAs w/seed CCUUCAU),tanespimycin,tazemetostat,YAP1	57	Oral tumor
7	calcitriol,medroxyprogesterone acetate	112	Gastrointestinal adenocarcinoma,Intestinal carcinoma
8	TP53	298	Gastrointestinal carcinoma
9	5-fluorouracil	28	Liver tumor
10	TAL1	31	Liver tumor
11	NUPR1	25	Hepatocellular carcinoma
12	MITF	20	Hepatocellular carcinoma
13	26s Proteasome	23	Liver tumor
14	EP400	19	Liver tumor
15	CDKN2A	69	Intestinal cancer,Large intestine neoplasm
16	FOXO1	45	Hepatobiliary system cancer
17	E2F1	47	Hepatocellular carcinoma
18	HGF	35	Hepatocellular carcinoma
19	arsenic trioxide	32	Liver tumor
20	let-7	27	Hepatocellular carcinoma
21	TP73	36	Hepatobiliary system cancer
22	mir-21	13	Oral tumor
23	valproic acid	12	Cell death of osteosarcoma cells

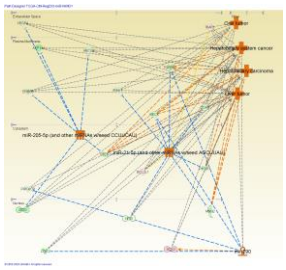
Table 4. Regulator effect networks related to cancer in diffuse-type gastric cancer (GC). Regulator Effect networks related to cancer have been generated. Type of regulators include biological drug, canonical pathway, chemical drug.

ID	Regulators	Target Total	Diseases & Functions
1	ACTB,AREG,BRD4,CCND1,CDKN1A,DYRK1A,E2f,E2F3,EIF4G1,EWSR1,FOXM1,GATA1,gentamicin,imipramine blue,LIN9,MED1,MYCN,NLRP3,NTRK2,phenethyl isothiocyanate,Rb,RB1,RBL2,TCF3,TFDP1	276	Female genital neoplasm,Gonadal tumor,Oral tumor,Tumorigenesis of reproductive tract
2	ATF4,ATF6,BNIP3L,E2f,EIF4G1,epothilone B,ERG,FOXM1,GATA1,gentamicin,imipramine blue,Irgm1,KDM5B,let-7,miR-24-3p (and other miRNAs w/seed GGCUCAG),NLRP3,phenethyl isothiocyanate,RABL6,Rb,RB1,RBL1,RBL2,SMARCB1,ZNF281	231	Cell death of osteosarcoma cells,Female genital neoplasm,Gonadal tumor,Tumorigenesis of reproductive tract
3	alvespimycin,decitabine,EGFR,EWSR1,gentamicin,KAT6A,miR-34a-5p (and other miRNAs w/seed GGCAGUG),phenethyl isothiocyanate,SYVN1,tazemetostat,YAP1	67	Oral tumor
4	alvespimycin,calcitriol,decitabine,E2F2,EGFR,ERBB2,estrogen,EWSR1,mir-181,phenethyl isothiocyanate,tazemetostat,Vegf,YAP1	210	Oral tumor,Prostatic carcinoma
5	CCND1,DDIT3,HDAC1	140	Digestive system cancer,Oral tumor,Prostatic carcinoma
6	ATF4,ATF6,EIF4G1,EP400,FOXM1,gentamicin,Irgm1,MYC,NELFA,NELFCD,NELFE,NLRP3,ZNF281	94	Ovarian tumor
7	KDM4C,UXT	18	Frequency of tumor,Incidence of tumor
8	CSF2,DDIT3,ERG,ESR1,miR-291a-3p (and other miRNAs w/seed AAGUGCU)	131	Prostatic carcinoma
9	TP53	131	Prostatic carcinoma
10	E2F1	87	Tumorigenesis of reproductive tract
11	HGF	67	Female genital neoplasm
12	TBX2	38	Digestive system cancer
13	SMARCB1	39	Abdominal carcinoma
14	TP63	42	Tumorigenesis of reproductive tract
15	PTGER2	40	Abdominal carcinoma
16	MITF	33	Female genital neoplasm
17	mir-21	27	Prostatic carcinoma
18	NFE2L2	23	Gonadal tumor
19	CD3	19	Oral tumor
20	DNMT3B	18	Female genital neoplasm
21	cephaloridine	18	Tumorigenesis of reproductive tract
22	CDKN2A	47	Tumorigenesis of reproductive tract
23	NFE2L1	6	Cell death of

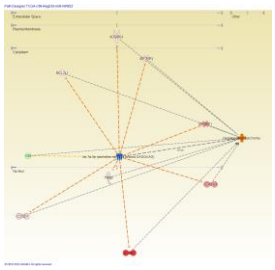
			osteosarcoma cells
24	EIF4E	11	Ovarian tumor
25	5-fluorouracil	5	Cell death of osteosarcoma cells
26	mibolerone	9	Ovarian tumor
27	KDM1A	25	Female genital neoplasm,Tumorigenesis of reproductive tract
28	TRAP1	6	Oral tumor
29	fulvestrant	45	Female genital neoplasm
30	NCOA3	15	Tumorigenesis of reproductive tract
31	MEF2D	9	Female genital neoplasm,Tumorigenesis of reproductive tract

2.4. MicroRNA (miRNA)-related regulator effect networks in intestinal- and diffuse-type GC

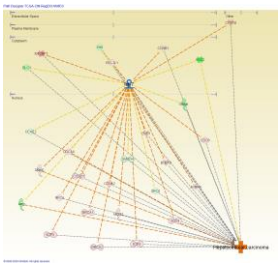
MicroRNA (miRNA)-related regulator effect networks were analyzed in intestinal- and diffuse-type GC (Figure 4). Table 5 shows miRNA-related regulator effect networks in intestinal-type GC, whereas Table 6 shows miRNA-related regulator effect networks in diffuse-type GC.



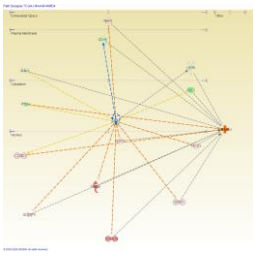
(a)



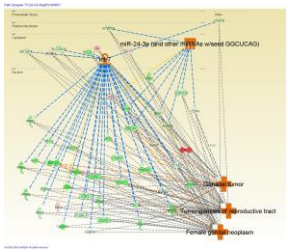
(b)

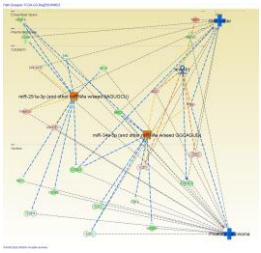


(c)



(d)





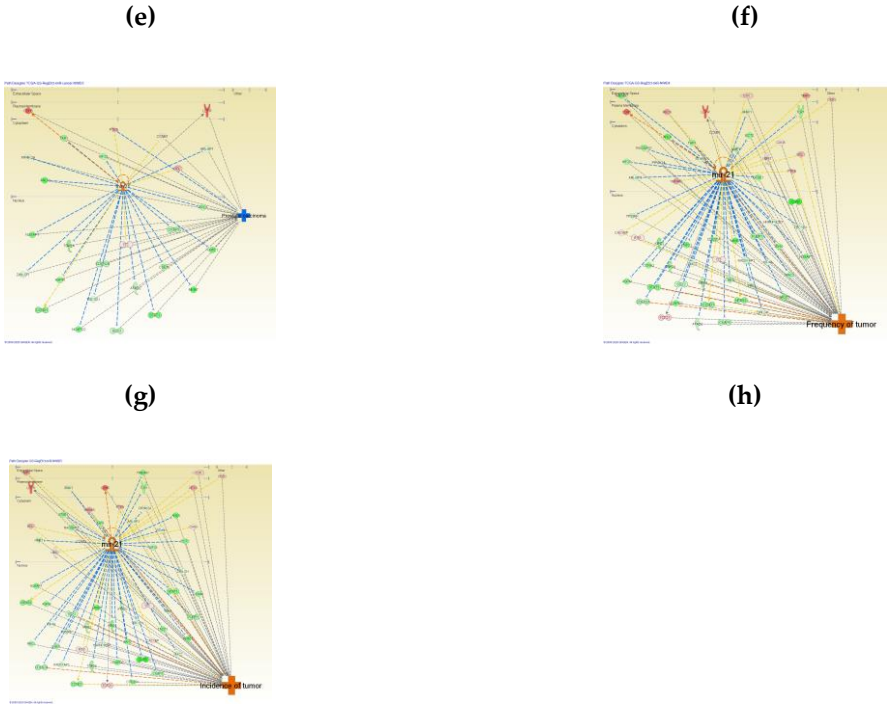


Figure 4. MicroRNA (miRNA)-regulated networks in intestinal- and diffuse-type gastric cancer (GC). Regulators of which type was set as miRNA and mature miRNA were analyzed in data set of intestinal-type (**a-d**) or diffuse-type (**e-i**) GC. Four networks were generated in intestinal-type GC, while 5 networks were generated in diffuse-type GC. (**a**) Network ID#1 regulated by miR-205-5p (and other miRNAs w/seed CCUUCAU), miR-21-5p (and other miRNAs w/seed AGCUUAU), and mir-290 in diffuse-type GC; (**b**) Network ID#2 regulated by let-7a-5p (and other miRNAs w/seed GAGGUAG) in diffuse-type GC; (**c**) Network ID#3 regulated by let-7 in diffuse-type GC; (**d**) Network ID#4 regulated by mir-21 in diffuse-type GC; (**e**) Network ID#1 regulated by let-7, miR-24-3p (and other miRNAs w/seed GGCUCAG) in intestinal-type GC; (**f**) Network ID#2 regulated by mir-181,miR-291a-3p (and other miRNAs w/seed AAGUGCU), miR-34a-5p (and other miRNAs w/seed GGCAGUG) in intestinal-type GC; (**g**) Network ID#3 regulated by mir-21 in intestinal-type GC; (**h**) Network ID#4 regulated by mir-21 in intestinal-type GC; (**i**) Network ID#5 regulated by mir-21 in intestinal-type GC.

26 **Table 5. MicroRNA (miRNA)-related regulator effect networks in intestinal-type gastric cancer**
27 **(GC).**

ID	Regulators	Target Molecules in Dataset	Diseases & Functions	Known Regulator-Disease/Function Relationship
1	miR-205-5p (and other miRNAs w/seed CCUUCAU),miR-21-5p (and other miRNAs w/seed AGCUUAU),mir-290	ABCC2,ATP1A1,BCL2L1,CDH5,CDK2,ERBB3,IRAK1,MSH2,NFIB,PIK3R1,PRKACB,PTEN,RECK,SOX2,TGFBR2,TIMP3,VEGFA,ZEB2	Hepatobiliary carcinoma,Hepatobiliary system cancer,Liver tumor,Oral tumor	42% (5/12)
2	let-7a-5p (and other miRNAs w/seed GAGGUAG)	ADGRG1,BCL2L1,CCND1,CCNE1,CDKN2A,IGF2BP1,IGF2BP3,TYMS,VIM	Hepatocellular carcinoma	100% (1/1)
3	let-7	AGO2,APC,AURKA,BCL2L1,BRCA1,BRCA2,BUB1,BUB1B,CCNA2,CCNB1,CCND1,CCNE2,CDC6,CDCA8,CKS1B,DLC1,E2F5,E2F8,IGF2BP1,MCM2,ORC6,RFC4,RRM1,RRM2,SMAD4,SOX9,VIM	Hepatocellular carcinoma	100% (1/1)
4	mir-21	BCL2,CCND1,CDH5,CDKN2A,DLGAP5,IRAK1,KNTC1,LEPR,PTEN,STAT1,TACC3,TIMP3,TOP2A	Oral tumor	0% (0/1)

28

29

30 **Table 6. MicroRNA (miRNA)-related regulator effect networks in diffuse-type gastric cancer**
31 **(GC).**

ID	Regulators	Target Molecules in Dataset	Diseases & Functions	Known Regulator-Disease/Function Relationship
1	let-7,miR-24-3p (and other miRNAs w/seed GGCUCAG)	ACVR1B,APC,AURKA,AURKB,BCL2L1,BRCA1,BRCA2,BUB1,BUB1B,CCNA2,CCNB1,CCND1,CCNE2,CDC20,CDC25A,CDC6,CDK1,CDK4,CDKN2A,CKS1B,DBF4,DLC1,E2F4,E2F8,FANCD2,FBL,FEN1,HMGA1,IGF2BP1,MCM10,MCM2,MCM7,MCM8,NOLC1,NUF2,PLAGL2,RFC4,RFC5,RRM1,RRM2,SALL4,SLC25A13,SMAD4,SOX9,TARBP2,VIM,XPO5	Female genital neoplasm,Gonadal tumor,Tumorigenesis of reproductive tract	50% (3/6)
2	mir-181,miR-291a-3p (and other miRNAs w/seed AAGUGCU), miR-34a-5p (and other miRNAs w/seed GGCAGUG)	ADCY9,ARHGEF3,BCL2,BIRC5,CCND1,CD46,CDK4,CDKN2A,CENPF,E2F3,E2F5,FAM13B,KIF23,MCM10,NIN,PRC1,PRKACB,PTEN,SOX2,TFAP4,TIMP3,VEGFA,ZEB2	Oral tumor,Prostatic carcinoma	0% (0/6)
3	mir-21	ANLN,ARL6IP1,ASPM,ATAD2,BCL2,CCNB1,CCND1,CDH5,CDKN2A,CKAP5,CSE1L,KIF23,KNTC1,LEPR,MKI67,NCAPD2,NUSAP1,PIP4K2A,PRC1,PTEN,SOX2,STAT1,TAP1,TBC1D1,TOP2A,YY1,ZWILCH	Prostatic carcinoma	0% (0/1)
4	mir-21	ANLN,ARL6IP1,ASPM,ATAD2,BCL2,C1R,CACYBP,CANX,CCNA2,CCNB1,CCND1,CDC25A,CDH5,CDKN2A,CKAP5,CKS2,CLPB,CSE1L,ECT2,FUBP1,GTSE1,HNRNPA2B1,IFI16,IRAK1,KIF23,KIF4A,KIFC1,KNTC1,LEPR,MKI67,MSH2,NCAPD2,NME1,NPAS2,NUSAP1,PIP4K2A,PRC1,PTEN,RACGAP1,RAD51AP1,RECK,SMC2,SOX2,STAT1,STMN1,TACC3,TAP1,TBC1D1,TCF21,TIMP3,TLR1,TMEM97,TOP2A,TP53RK,UBA7,VRK1,YWHAB,YY1,ZW10,ZWILCH	Frequency of tumor	100% (1/1)
5	mir-21	ANLN,ARL6IP1,ASPM,ATAD2,BCL2,C1R,CACYBP,CANX,CCNA2,CCNB1,CCND1,CDC25A,CDH5,CDKN2A,CKAP5,CKS2,CLPB,CSE1L,ECT2,FUBP1,GTSE1,HNRNPA2B1,IFI16,IRAK1,KIF23,KIF4A,KIFC1,KNTC1,LEPR,MKI67,MSH2,NCAPD2,NME1,NPAS2,NUSAP1,PIP4K2A,PRC1,PTEN,RACGAP1,RAD51AP1,RECK,SMC2,SOX2,STAT1,STMN1,TACC3,TAP1,TBC1D1,TCF21,TIMP3,TLR1,TMEM97,TOP2A,TP53RK,UBA7,VRK1,YWHAB,YY1,ZW10,ZWILCH	Incidence of tumor	100% (1/1)

2.5. Upstream regulators in intestinal- and diffuse-type GC

Upstream regulators of genes altered in intestinal- and diffuse-type GC were defined by IPA analysis. Top 25 upstream regulators of the altered genes in intestinal- and diffuse-type GC are shown in Table 7. The top 25 upstream regulators include NUPR1, CSF2, PTGER2, TP53, EGFR, let-7, ERBB2, calcitriol, RABL6, MITF, E2F1, CDKN2A, KDM1A, E2F3, EP400, BNIP3L, YAP1, MYCN, MYC, HGF, E2f, AREG, TBX2 and KDM5B.

Table 7. Upstream regulators in intestinal- and diffuse-type gastric cancer (GC) (Top25 regulators).

Upstream Regulators	TCGA CIN	TCGA GS
NUPR1	-4.457	6.685
CSF2	4.849	-6.057
PTGER2	4.427	-5.06
TP53	-4.044	5.394
EGFR	3.75	-5.207
let-7	-3.031	5.836
ERBB2	2.986	-5.804
calcitriol	-3.349	5.194
RABL6	3.28	-5.154
MITF	2.927	-5.436
E2F1	2.141	-5.933
CDKN2A	-2.944	5
KDM1A	3.328	-4.551
E2F3	2.496	-5.334
EP400	3.183	-4.482
BNIP3L	-3.714	3.571
YAP1	3.103	-4.161
MYCN	4.044	-2.997
MYC	1.087	-5.862
HGF	2.874	-4.014
E2f	2.984	-3.881
AREG	3.525	-3.213
TBX2	2.619	-4.104
KDM5B	-4.075	2.537

2.6. Gene Ontology (Biological Process) of genes regulated in intestinal- and diffuse-type GC

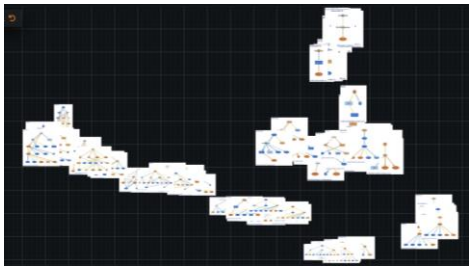
Gene Ontology (GO) was analyzed in genes regulated in intestinal- and diffuse-type GC. Total 2815 IDs were analyzed for enrichment analysis in DAVID database, which resulted in 2762 DAVID gene IDs analyzed in GO Biological Process. Top 21 GOs are shown in Table 8 (modified Fischer Exact p value < 1E-06, $p < 0.005$ in Bonferroni statistics).

Table 8. Gene Ontology (Biological Process) of genes regulated in intestinal- and diffuse-type gastric cancer (GC). The total 2815 probe set IDs were analyzed for enrichment analysis in DAVID, which resulted in 2394 genes analyzed in Biological Process.

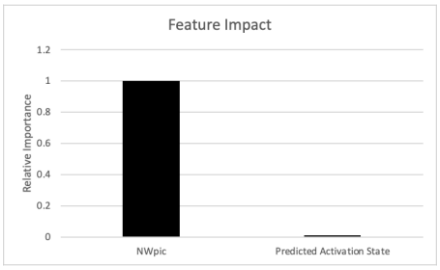
Category	Term	Count
GOTERM_BP_DIRECT	GO:0051301~cell division	121
GOTERM_BP_DIRECT	GO:0007062~sister chromatid cohesion	54
GOTERM_BP_DIRECT	GO:0007067~mitotic nuclear division	91
GOTERM_BP_DIRECT	GO:0006260~DNA replication	67
GOTERM_BP_DIRECT	GO:0031145~anaphase-promoting complex-dependent catabolic process	40
GOTERM_BP_DIRECT	GO:0051436~negative regulation of ubiquitin-protein ligase activity involved in mitotic cell cycle	37
GOTERM_BP_DIRECT	GO:0000082~G1/S transition of mitotic cell cycle	44
GOTERM_BP_DIRECT	GO:0051437~positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	36
GOTERM_BP_DIRECT	GO:0006281~DNA repair	74
GOTERM_BP_DIRECT	GO:0006521~regulation of cellular amino acid metabolic process	27
GOTERM_BP_DIRECT	GO:0006270~DNA replication initiation	20
GOTERM_BP_DIRECT	GO:0043488~regulation of mRNA stability	39
GOTERM_BP_DIRECT	GO:0006364~rRNA processing	62
GOTERM_BP_DIRECT	GO:0007059~chromosome segregation	29
GOTERM_BP_DIRECT	GO:0031047~gene silencing by RNA	39
GOTERM_BP_DIRECT	GO:0038061~NIK/NF-kappaB signaling	27
GOTERM_BP_DIRECT	GO:0060071~Wnt signaling pathway, planar cell polarity pathway	33
GOTERM_BP_DIRECT	GO:0002479~antigen processing and presentation of exogenous peptide antigen via MHC class I, TAP-dependent	26
GOTERM_BP_DIRECT	GO:0007077~mitotic nuclear envelope disassembly	21
GOTERM_BP_DIRECT	GO:0000398~mRNA splicing, via spliceosome	60
GOTERM_BP_DIRECT	GO:0070125~mitochondrial translational elongation	31

2.7. Prediction model for molecular networks of intestinal- and diffuse-type GC

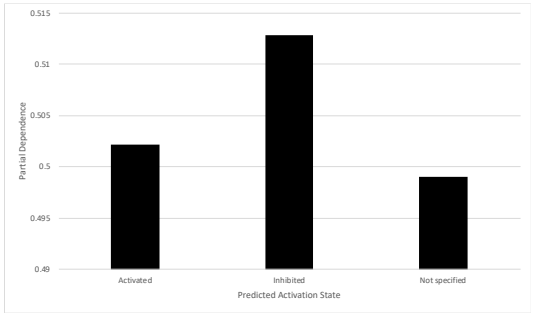
The results of upstream analysis of intestinal- and diffuse-type GC data were analyzed in DataRobot Automated Machine Learning version 6.0 for creating prediction models. The list of upstream regulators was up-loaded linked with network picture data, followed by the target prediction setting as subtype differences in intestinal- and diffuse-type GC (Figure 5). Among various prediction models DataRobot created, Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance) was the highest predictive accuracy model with AUC of 0.7185 in cross-validation score. For this model, the feature impact chart using Permutation Importance showed that the most important features for accurately predicting the subtype of GC ("Analysis" values) were upstream network pictures (NWpic) (Figure5a) and Predicted Activation State (Figure5b). Figure5c shows the Partial Dependence Plot in Predicted Activation State. Figure 5Ddshows the Word Cloud of the target molecules. The size of the molecules indicates the appearance in the dataset, and the color shows coefficient. Figure5e shows the activation maps where the attention of AI is highlighted.



(a)

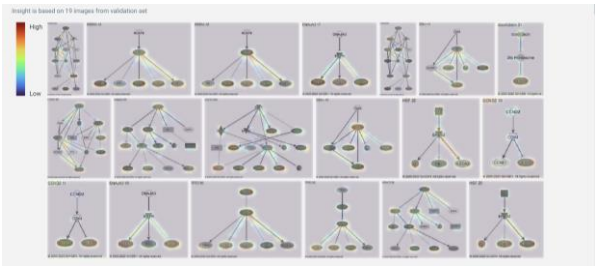


(b)

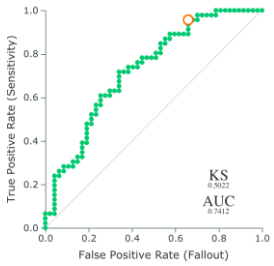


(c)

(d)



(e)



(f)

Figure 5. AI-oriented prediction model in intestinal- and diffuse-type gastric cancer (GC). The results of upstream analysis of intestinal- and diffuse-type GC data in IPA were analyzed in DataRobot Automated Machine Learning version 6.0 (DataRobot) for creating prediction models. The list of upstream regulators was up-loaded linked with the network picture data, followed by the target prediction setting as subtype differences in intestinal- and diffuse-type GC. Among various prediction models DataRobot created, Elastic-Net Classifier (mixing alpha=0.5 / Binomial Deviance) was the highest predictive accuracy model with AUC of 0.7185 in cross-validation score. For this model, the feature impact chart using Permutation Importance showed that the most important features for accurately predicting the subtype of GC ("Analysis" values) were upstream network pictures (NWpic). (**a**) The Image Embedding of 93 images for creating the insight; (**b**) Feature Impact for showing the important features for predicting the subtype of GC; (**c**) The Partial Dependence Plot in Predicted Activation State; (**d**) The Word Cloud of the target molecules. The size of the molecules indicates the appearance in the dataset, and the color shows coefficient; (**e**) The activation maps where the attention of AI is highlighted; (**f**) ROC curve for the model.

2.8. EMT molecular pathway and diffuse-type GC mapping

The canonical pathways for Regulation of the EMT pathway include TGF-beta pathway, Wnt pathway, Notch pathway and Receptor Tyrosine Kinase pathway (Figure 6). In each pathway related to EMT, genes of which expression was altered in diffuse-type GC compared to intestinal-type GC are mapped in pink (up-regulated) or green (down-regulated) color. The activation states of the pathways are predicted with IPA, and shown in orange (activation) or blue (inhibition) color. RNA-RNA interaction analysis identified interacted miRNAs as let-7, mir-10, mir-126, mir-181, mir-26, mir-515, MIR100-LET7A2-MIR125B1, MIR124, MIR99A-LET7C-MIR125B2, and MIRLET7.

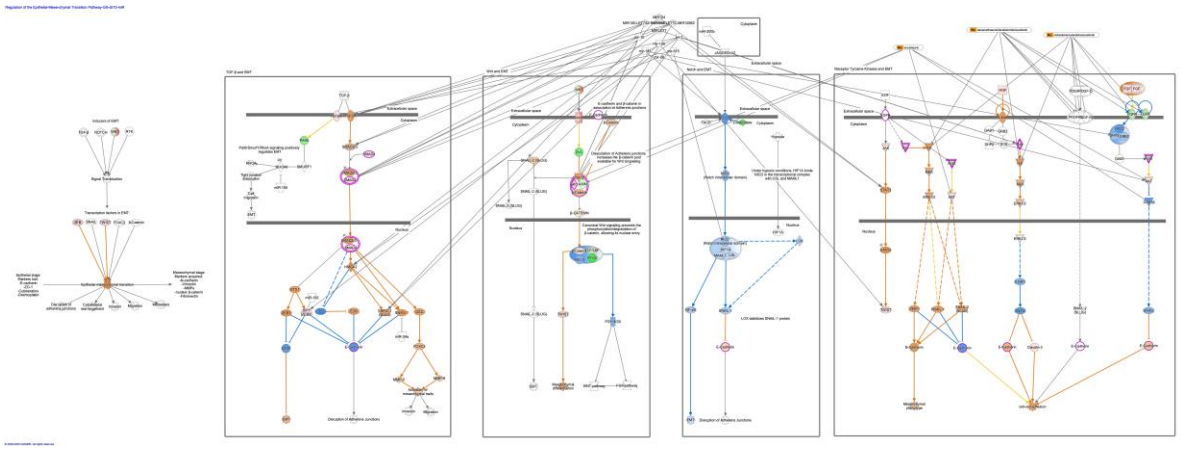


Figure 6. Canonical EMT molecular pathway and diffuse-type GC-related gene mapping. Canonical pathways for Regulation of the EMT pathway are shown. The genes of which expression was altered in diffuse-type GC compared to intestinal-type GC are shown in pink (up-regulated) or green (down-regulated).

3. Discussion

It is important to distinguish the intestinal- and diffuse-type GC for effective therapeutic strategies since the pathogenesis and prognosis are quite different in these subtypes. We previously revealed the gene signature of intestinal- and diffuse-type GC, which is indicated by the ratio of gene expression in *CDH2* to *CDH1* [2]. CDH1 and CDH2 are important factors as the signatures for distinguishing the subtypes of GC. Since our previous reports, the abundant useful open-source data, including RefSeq data for the intestinal- and diffuse-type GC have been available in public [9-12]. Our current study highlights the relevance of using open-source data for human health. In this study, the RefSeq data of intestinal- and diffuse-type GC has been analyzed for exploring the molecular networks and AI modeling application. Top 10 genes of which gene expression was altered in intestinal- and diffuse-type GC RefSeq data included *CKS1B*, *CSE1L*, *DDX27*, *GET4*, *MRGBP*, *MSL3P1*, *PARD6B*, *RAE1*, *TOMM34* and *YTHDF1*. The network analysis of altered genes in intestinal- and diffuse-type GC generated networks related to cancer, gastrointestinal disease, organismal injury and abnormalities, amino acid metabolism, molecular transport, small molecule biochemistry, and so on. Several miRNAs including miR-205-5p, miR-21-5p, let-7a-5p, let-7, miR-24-3p, miR-291a-3p were identified to regulate networks involved in intestinal- and diffuse-type GC. Since previous studies have revealed the involvement of miR-200s in promoting metastatic colonization by inhibiting EMT and promoting mesenchymal-epithelial transition (MET), it may be a very interesting approach to reveal miRNA networks in EMT [13, 14]. The several miRNAs are involved and regulated in EMT and MET, which would be critical for progression and metastasis process [15-17]. DataRobot Automated Machine Learning created prediction models to distinguish intestinal- and diffuse-type GC with results of up-stream analysis and the network picture data. The image recognition of molecular networks by AI would distinguish the intestinal- and diffuse-type GC. It was indicated that Predicted Activation State can anticipate the subtypes of

GC with approximately 0.5 of partial dependence, which showed that the predicted activation state of the molecular networks may distinguish the subtypes of GC.

The intestinal- and diffuse-type GC can be distinguished with the mRNA ratios of *CDH2* to *CDH1* as previously shown [2]. The molecular network profiling is very important to reveal the mechanisms behind the differences between the intestinal- and diffuse-type GC, such as EMT and drug resistance in CSCs. The research exploring the differences between molecular networks in intestinal- and diffuse-type GC would reveal the interesting mechanisms leading to the therapeutic target identification. It is easier to detect miRNAs in the blood than to analyze the tissues. The current study exploring the miRNA regulation in intestinal- and diffuse-type GC might identify the miRNAs involving the EMT in diffuse-type GC, and these miRNAs might be detected in blood. The profile in the molecular networks of RNAs detected in blood would be the next pathways to be reveal in the near future research.

4. Materials and Methods

4.1. Data Collection

The RefSeq data of intestinal- and diffuse-type GC are publicly available in The Cancer Genome Atlas (TCGA) database (http://www.cbioportal.org/) [9-11] in NCI Genomic Data Commons (GDC) (https://portal.gdc.cancer.gov/) [18]. From the data stomach adenocarcinoma (TCGA, PanCancer Atlas), intestinal- and diffuse-type GC data, which are noted as chromosomal instability (CIN) and genomically stable (GS), respectively in TCGA Research Network publication, were compared [11].

4.2. Network Analysis

Data of intestinal- and diffuse-type GC in TCGA cBioPortal Cancer Genomics were uploaded and analyzed through the use of Ingenuity Pathway Analysis (IPA) (QIAGEN Inc., https://www.qiagenbioinformatics.com/products/ingenuitypathway-analysis) [19].

4.3. Gene Ontology Analysis

Gene Ontology was analyzed in the Database for Annotations, Visualization and Integrated Discovery (DAVID) Bioinformatics Resources 6.8 (Laboratory of Human Retrovirology and Immunoinformatics, https://david.ncifcrf.gov/) [20, 21].

4.4. AI Prediction Modeling

To create a prediction model by using multi-modal data including images and text description of molecular networks, an enterprise AI platform (DataRobot Automated Machine Learning version 6.0; DataRobot Inc.) was used. For the modeling, the 116 molecular networks of IPA upstream analysis in intestinal- and diffuse-type GC were collected and input as image data in the DataRobot (58 images in each subtype), that automatically created and tuned prediction models using various machine learning algorithms (e.g. eXtreme gradient-boosted trees, random forest, regularized regression such as Elastic Net, Neural Networks) [22, 23]. Finally, the AI model with the highest predictive accuracy on DataRobot was identified and various insights (such as Permutation Importance or Partial Dependence Plot) obtained from the model were reviewed.

4.5. Data Visualization

The results of gene expression data of RefSeq and network analysis were visualized by Tableau software (https://www.tableau.com/).

4.6. Statistical Analysis

The RefSeq data were analyzed by Student's t-test. Z-score in intestinal- and diffuse-type GC samples were compared, and the difference was considered to be significant in p value < 0.00001. For DAVID Gene Ontology (GO) enrichment analysis, data was analyzed in the default setting. GO

enrichment was considered significant in modified Fischer Exact p value < 1E-06. Bonferroni statistics showed p value < 0.005.

5. Conclusions

The regulatory molecular networks are altered in intestinal- and diffuse-type GC. Networks generated from genes altered in intestinal- and diffuse-type GC included a network related to cancer, gastrointestinal disease, and organismal injury and abnormalities. We demonstrated that several miRNAs regulated the networks in intestinal- and diffuse-type GC. Machine learning of network image data created prediction models to distinguish the subtypes of the GC. Our results support further identification of GC subtypes through visual changes in molecular networks.

Supplementary Materials: The following are available online at www.mdpi.com/xxx/s1, Table S1: List of 2815 gene ID altered in intestinal- and diffuse-type gastric cancer (GC).

Author Contributions: Conceptualization, Shihori Tanabe and Hiroki Sasaki; Data curation, Shihori Tanabe, Kazuhiko Aoyagi and Hiroki Sasaki; Formal analysis, Shihori Tanabe; Funding acquisition, Shihori Tanabe, Sabina Quader, Ryuichi Ono and Akihiko Hirose; Investigation, Shihori Tanabe; Methodology, Shihori Tanabe; Project administration, Shihori Tanabe, Kazuhiko Aoyagi, Hiroshi Yokozaki and Hiroki Sasaki; Resources, Shihori Tanabe; Software, Shihori Tanabe; Supervision, Shihori Tanabe and Akihiko Hirose; Visualization, Shihori Tanabe; Writing – original draft, Shihori Tanabe; Writing – review & editing, Shihori Tanabe, Sabina Quader and Horacio Cabral.

Funding: This research was funded by Japan Agency for Medical Research and Development (AMED), grant number JP20ak0101093 (ST, RO and AH) and JP20mk0101163 (RO), and Strategic International Collaborative Research Program, grant number JP20jm0210059 (ST and SQ), Ministry of Health, Labour, and Welfare (MHLW), grant number H30-KAGAKU-IPPAN-002 (ST and RO), and JSPS KAKENHI grant number 18K19315 (RO).

Acknowledgments: The authors would like to acknowledge Shinpei Ijichi and Kohei Kessoku for assisting the DataRobot application.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Perrot-Applanat, M.; Vacher, S.; Pimpie, C.; Chemlali, W.; Derieux, S.; Pocard, M., Bieche, I. Differential gene expression in growth factors, epithelial mesenchymal transition and chemotaxis in the diffuse type compared with the intestinal type of gastric cancer. *Oncol Lett* **2019**, *18*, 674-686 [PMID: 31289541 DOI: 10.3892/ol.2019.10392]

2. Tanabe, S.; Aoyagi, K.; Yokozaki, H., Sasaki, H. Gene expression signatures for identifying diffuse-type gastric cancer associated with epithelial-mesenchymal transition. *Int J Oncol* **2014**, *44*, 1955-1970 [PMID: 24728500 DOI: 10.3892/ijo.2014.2387]

3. Assumpção, P.P.; Barra, W.F.; Ishak, G.; Coelho, L.G.V.; Coimbra, F.J.F.; Freitas, H.C.; Dias-Neto, E.; Camargo, M.C., Szklo, M. The diffuse-type gastric cancer epidemiology enigma. *BMC Gastroenterol* **2020**, *20*, 223-223 [PMID: 32660428 DOI: 10.1186/s12876-020-01354-4]

4. Hoang, T.; Ganesan, A.K.; Hiyama, D., Dayyani, F. Gene mutations distinguishing gastric from colorectal and esophageal adenocarcinomas. *J Gastrointest Oncol* **2020**, *11*, 45-54 [PMID: 32175104 DOI: 10.21037/jgo.2019.12.06]

5. Sohn, S.H.; Kim, B.; Sul, H.J.; Kim, Y.J.; Kim, H.S.; Kim, H.; Seo, J.B.; Koh, Y., Zang, D.Y. INC280 inhibits Wnt/β-catenin and EMT signaling pathways and its induce apoptosis in diffuse gastric cancer positive for c-MET amplification. *BMC Res Notes* **2019**, *12*, 125 [PMID: 30871613 DOI: 10.1186/s13104-019-4163-x]

6. Tanabe, S.; Quader, S.; Cabral, H., Ono, R. Interplay of EMT and CSC in Cancer and the Potential Therapeutic Strategies. *Frontiers in Pharmacology* **2020**, *11*, [DOI: 10.3389/fphar.2020.00904]

7. Tanabe, S.; Kawabata, T.; Aoyagi, K.; Yokozaki, H., Sasaki, H. Gene expression and pathway analysis of CTNNB1 in cancer and stem cells. *World J Stem Cells* **2016**, *8*, 384-395 [PMID: 27928465 DOI: 10.4252/wjsc.v8.i11.384]

8. Cao, B.; Zhao, Y.; Zhang, Z.; Li, H.; Xing, J.; Guo, S.; Qiu, X.; Zhang, S.; Min, L., Zhu, S. Gene regulatory network construction identified NFYA as a diffuse subtype-specific prognostic factor in gastric cancer. *Int J Oncol* **2018**, *53*, 1857-1868 [DOI: 10.3892/ijo.2018.4519]

9. Cerami, E.; Gao, J.; Dogrusoz, U.; Gross, B.E.; Sumer, S.O.; Aksoy, B.A.; Jacobsen, A.; Byrne, C.J.; Heuer, M.L.; Larsson, E.; Antipin, Y.; Reva, B.; Goldberg, A.P.; Sander, C., Schultz, N. The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discovery* **2012**, *2*, 401 [DOI: 10.1158/2159-8290.CD-12-0095]

10. Gao, J.; Aksoy, B.A.; Dogrusoz, U.; Dresdner, G.; Gross, B.; Sumer, S.O.; Sun, Y.; Jacobsen, A.; Sinha, R.; Larsson, E.; Cerami, E.; Sander, C., Schultz, N. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **2013**, *6*, pl1 [PMID: 23550210 DOI: 10.1126/scisignal.2004088]

11. Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **2014**, *513*, 202-209 [PMID: 25079317 DOI: 10.1038/nature13480]

12. Hoadley, K.A.; Yau, C.; Hinoue, T.; Wolf, D.M.; Lazar, A.J.; Drill, E.; Shen, R.; Taylor, A.M.; Cherniack, A.D.; Thorsson, V.; Akbani, R.; Bowlby, R.; Wong, C.K.; Wiznerowicz, M.; Sanchez-Vega, F.; Robertson, A.G.; Schneider, B.G.; Lawrence, M.S.; Noushmehr, H.; Malta, T.M.; Cancer Genome Atlas, N.; Stuart, J.M.; Benz, C.C., Laird, P.W. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **2018**, *173*, 291-304.e296 [PMID: 29625048 DOI: 10.1016/j.cell.2018.03.022]

13. Korpal, M.; Ell, B.J.; Buffa, F.M.; Ibrahim, T.; Blanco, M.A.; Celià-Terrassa, T.; Mercatali, L.; Khan, Z.; Goodarzi, H.; Hua, Y.; Wei, Y.; Hu, G.; Garcia, B.A.; Ragoussis, J.; Amadori, D.; Harris, A.L., Kang, Y. Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* **2011**, *17*, 1101-1108 [PMID: 21822286 DOI: 10.1038/nm.2401]

14. Sun, Z.; Zhou, S.; Tang, J.; Ye, T.; Li, J.; Liu, D.; Zhou, J.; Wang, J., Rosie Xing, H. Sec23a mediates miR-200c augmented oligometastatic to polymetastatic progression. *EBioMedicine* **2018**, *37*, 47-55 [PMID: 30301603 DOI: 10.1016/j.ebiom.2018.10.002]

15. Ma, L.; Young, J.; Prabhala, H.; Pan, E.; Mestdagh, P.; Muth, D.; Teruya-Feldstein, J.; Reinhardt, F.; Onder, T.T.; Valastyan, S.; Westermann, F.; Speleman, F.; Vandesompele, J., Weinberg, R.A. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol* **2010**, *12*, 247-256 [PMID: 20173740 DOI: 10.1038/ncb2024]

16. Pan, Y.; Li, J.; Zhang, Y.; Wang, N.; Liang, H.; Liu, Y.; Zhang, C.Y.; Zen, K., Gu, H. Slug-upregulated miR-221 promotes breast cancer progression through suppressing E-cadherin expression. *Sci Rep* **2016**, *6*, 25798 [PMID: 27174021 DOI: 10.1038/srep25798]

17. Choi, P.W., Ng, S.W. The Functions of MicroRNA-200 Family in Ovarian Cancer: Beyond Epithelial-Mesenchymal Transition. *Int J Mol Sci* **2017**, *18*, [PMID: 28587302 DOI: 10.3390/ijms18061207]

18. Grossman, R.L.; Heath, A.P.; Ferretti, V.; Varmus, H.E.; Lowy, D.R.; Kibbe, W.A., Staudt, L.M. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* **2016**, *375*, 1109-1112 [PMID: 27653561 DOI: 10.1056/NEJMp1607591]

19. Krämer, A.; Green, J.; Pollard, J., Jr., Tugendreich, S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* **2014**, *30*, 523-530 [PMID: 24336805 DOI: 10.1093/bioinformatics/btt703]

20. Huang da, W.; Sherman, B.T., Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009**, *4*, 44-57 [PMID: 19131956 DOI: 10.1038/nprot.2008.211]

21. Huang da, W.; Sherman, B.T., Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **2009**, *37*, 1-13 [PMID: 19033363 DOI: 10.1093/nar/gkn923]

22. Breiman, L. Random Forests. *Machine Learning* **2001**, *45*, 5-32 [DOI: 10.1023/A:1010933404324]

23. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann Statist* **2001**, *29*, 1189-1232 [DOI: 10.1214/aos/1013203451]