

<https://doi.org/10.1038/s41540-024-00455-0>

Multi-bioinformatics revealed potential biomarkers and repurposed drugs for gastric adenocarcinoma-related gastric intestinal metaplasia

Check for updates

Gøran Troseth Andersen^{1,2,3}, Aleksandr Ianevski¹, Mathilde Resell¹, Naris Pojskic⁴, Hanne-Line Rabben¹, Synne Geithus¹, Yosuke Kodama¹, Tomita Hiroyuki⁵, Denis Kainov⁶, Jon Erik Grønbech^{1,2}, Yoku Hayakawa⁶, Timothy C. Wang⁷, Chun-Mei Zhao⁸ & Duan Chen⁸ ✉

Biomarkers associated with the progression from gastric intestinal metaplasia (GIM) to gastric adenocarcinoma (GA), i.e., GA-related GIM, could provide valuable insights into identifying patients with increased risk for GA. The aim of this study was to utilize multi-bioinformatics to reveal potential biomarkers for the GA-related GIM and predict potential drug repurposing for GA prevention in patients. The multi-bioinformatics included gene expression matrix (GEM) by microarray gene expression (MGE), ScType (a fully automated and ultra-fast cell-type identification based solely on a given scRNA-seq data), Ingenuity Pathway Analysis, PageRank centrality, GO and MSigDB enrichments, Cytoscape, Human Protein Atlas and molecular docking analysis in combination with immunohistochemistry. To identify GA-related GIM, paired surgical biopsies were collected from 16 GIM-GA patients who underwent gastrectomy, yielding 64 samples (4 biopsies per stomach x 16 patients) for MGE. Co-analysis was performed by including scRNAseq and immunohistochemistry datasets of endoscopic biopsies of 37 patients. The results of the present study showed potential biomarkers for GA-related GIM, including GEM of individual patients, individual genes (such as RBP2 and CD44), signaling pathways, network of molecules, and network of signaling pathways with key topological nodes. Accordingly, potential treatment targets with repurposed drugs were identified including epidermal growth factor receptor, proto-oncogene tyrosine-protein kinase Src, paxillin, transcription factor Jun, breast cancer type 1 susceptibility protein, cellular tumor antigen p53, mouse double minute 2, and CD44.

The prevalence of gastric intestinal metaplasia (GIM) is approximately 15% in those undergoing routine endoscopy in Europe^{1,2}. GIM can be found in approximately 50% of patients with gastric ulcers and almost 100% of patients with intestinal type gastric adenocarcinoma (GA), the major subtype of gastric cancer³. According to Correa's cascade⁴, the progression rate from GIM to GA varies from 0.25% to 42% during a course of 5 years^{5,6}, and this malignant conversion can occur even after *H. pylori* eradication^{7,8}.

Nevertheless, it remains unclear whether GIM glands can directly transform into GA or share a clonal origin⁹⁻¹¹, despite GIM being susceptible to somatic mutations and copy number aberrations commonly found in GA¹².

Currently, the American Gastroenterological Association (AGA), the European Society of Gastrointestinal Endoscopy and the British Society of Gastroenterology recommend no further intervention against GIM, except for the eradication of *H. pylori*¹³⁻¹⁵. Of note, in patients with GIM, the AGA

¹Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. ²Department of Surgery, St. Olav's Hospital, Trondheim, Norway. ³Department of Surgery, Namsos Hospital, Namsos, Norway. ⁴Laboratory for Bioinformatics and Biostatistics, University of Sarajevo - Institute for Genetic Engineering and Biotechnology, Sarajevo, Bosnia and Herzegovina. ⁵Department of Tumor Pathology, Gifu University Graduate School of Medicine, Gifu, Japan. ⁶Department of Gastroenterology, Tokyo University Hospital, Tokyo, Japan. ⁷Department of Digestive and Liver Diseases and Herbert Irving Comprehensive Cancer Center, Columbia University Medical Center, New York, USA. ✉e-mail: duan.chen@ntnu.no

suggests against routine use of endoscopic surveillance and routine repeat short-interval endoscopy with biopsies for the purpose of risk stratification. Based on H&E staining, histological classification of GIM has been suggested, as incomplete GIM (goblet cells without a brush border) was associated with a greater risk of progression to GA in comparison to complete GIM (with the brush border), but more investigation on the potential benefit of implementing this routine pathological characterization is needed^{13,16–19}. Thus, routine endoscopic surveillance with 3- and 5-year intervals is conditionally recommended in patients diagnosed with incomplete GIM or if there are any additional risk factors, i.e., family history and extensive GIM. On the other hand, in countries with a high risk of gastric cancer such as Japan or South Korea, annual endoscopic surveillance is recommended. Observational studies suggest that advanced GIM and long-term use of acid suppression are associated with higher risk of GA development after *H. pylori* eradication^{7,20–22}.

Biomarkers associated with the progression from GIM to GA could provide valuable insights into identifying patients with increased risk for GA. This information could help in the targeted selection of individuals who would benefit the most from screening and surveillance endoscopy. While biomarkers such as pepsinogen (I and II) levels are commonly used in Asian countries, their studies and applications in the United States and Europe remain limited¹³. CD10, a brush border protein specific to normal small intestinal mucosa and absent in the colon²³, contrasts with Das1, a monoclonal antibody that binds colonic epithelial protein²⁴. CD10 and Das1 have been suggested as potential candidate biomarkers to distinguish complete and incomplete GIM based on gene expression data and immunohistochemistry²⁴. Somatic mutations in GA can be found in adjacent GIM²⁵. Currently, the discovery of effective cancer biomarkers poses a challenge due to considerations of predictive efficacy and clinical application. Rarely can gastric tumorigenesis be attributed solely to an individual molecule, such as a gene or protein, as GA progression is believed to arise from the collective interactions of multiple molecules. Thus, bioinformatics advancements could show promise; for instance, reports on progression gene signatures for lung cancer and glioblastoma have emerged from single-cell RNA sequencing (scRNA-seq) and other transcriptomic analyses²⁶. Single-cell atlases based on scRNA-seq are increasingly accessible for a variety of tissues, organs, and organisms, offering new prospects from data mining to gaining biological insights²⁷. Such single-cell atlases are available for GA and GIM^{28,29}. Nevertheless, various sets of genes specific to GA have been suggested. The differences in GA biomarker proposals in these separate studies could be attributed to several potential reasons, including distinct methods for clinical tissue

sampling, diverse annotation approaches for cell clusters, and disparate data mining techniques^{28,29}.

In the present study, we aimed to utilize the multi-bioinformatics to reveal potential biomarkers for a subset of patients with progression from GIM to GA and to identify the drug targets and accordingly repurposed drugs. Thus, we compared the microarray gene expression (MGE) of GIM, GA and normal tissues within the individual patients, enabling to identify GA-related GIM. We utilized the newly developed computational platform ScType for cell-type identification based on given scRNA-seq data along with a comprehensive cell marker database as background information³⁰. We then created heatmaps of the individual patients and of the pathological diagnosis with sub-classifications by gastric histological activity (inflammation, epithelial defects, mucosal atrophy, hyperplasia, pseudopyloric metaplasia, and dysplasia or neoplasia), locations (antrum, cardia, corpus, ventriculi, major and minor curvatures), and cell types (e.g. goblet cells, metaplastic stem-like cells, chief cells, and cancer cells). Furthermore, we used bioinformatics including Ingenuity Pathway Analysis (IPA), pagerank centrality, GO and MSigDB enrichments, Human Protein Atlas, Cytoscape, ScType and molecular docking analysis. We found a network consisting of 865 RNA markers that might be associated with the progression from GIM to GA. Some of these RNAs encode proteins in the network could serve as potential targets for drug repurposing as well as drug development for the prevention of the progression from GIM to GA.

Results

GIM and GA displayed convoluted gene expression matrixes (GEM)

GEM was created according to histology, gastric histologic activity index (GHAI), and location of lesions (antrum, cardia, corpus, major and minor curvature) in combination of single cell atlas (including various cell types, such as goblet cells, chief cells, pit mucosa cells, stem-like cells and immune cells). It showed convoluted profiles in individual patients and in comparisons between GIM, GA and normal tissues (Fig. 1a, b).

GIM displayed higher levels of gene expression than GA

Co-analysis using ScType visualized nine cell types, including cancer cells and tumor microenvironment-related cells, such as vascular endothelial cells, smooth muscle cells, CD8 + NKT-like cells, B cells, myeloid dendritic cells, mast cells, squamous epithelial cells and neuroendocrine cells (Fig. 2a). Furthermore, it annotated the cell type-related gene expression according to five pathological appearances, including chronic atrophic gastritis (CAG), early gastric cancer (EGA), severe intestinal metaplasia (IMS), mild

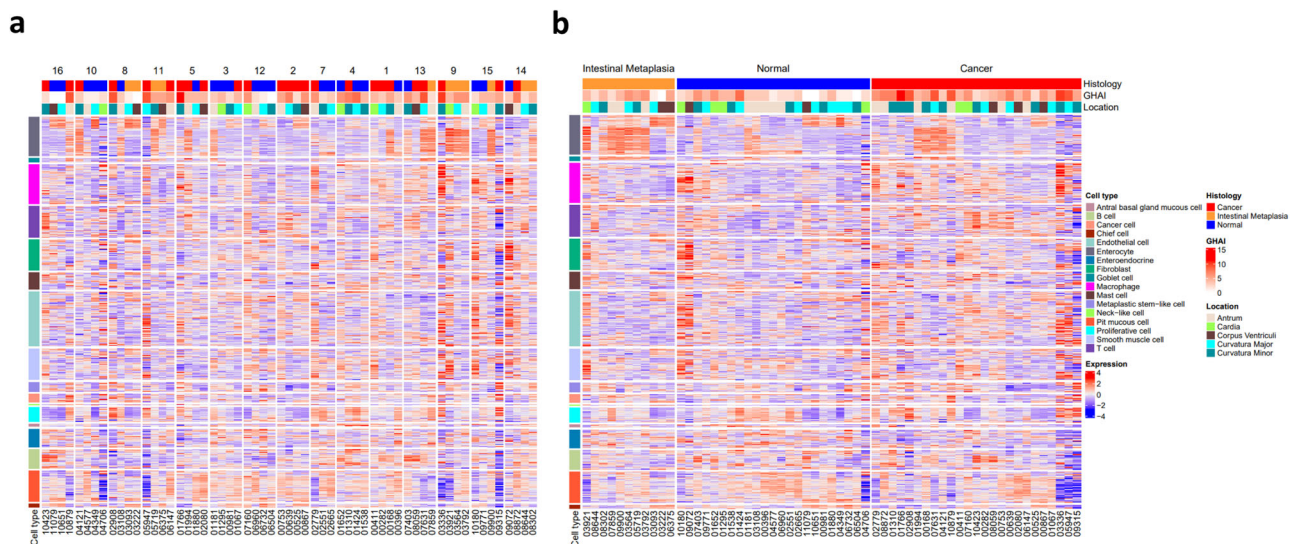


Fig. 1 | GEM. Of note, the gene expression heatmaps annotated according to histology, GHAI and location of lesions as well as single cell atlas in individual patients (a) and in GIM, normal and GA (b). Patient numbers (1-16, except 6) and sample numbers below heatmaps were included.

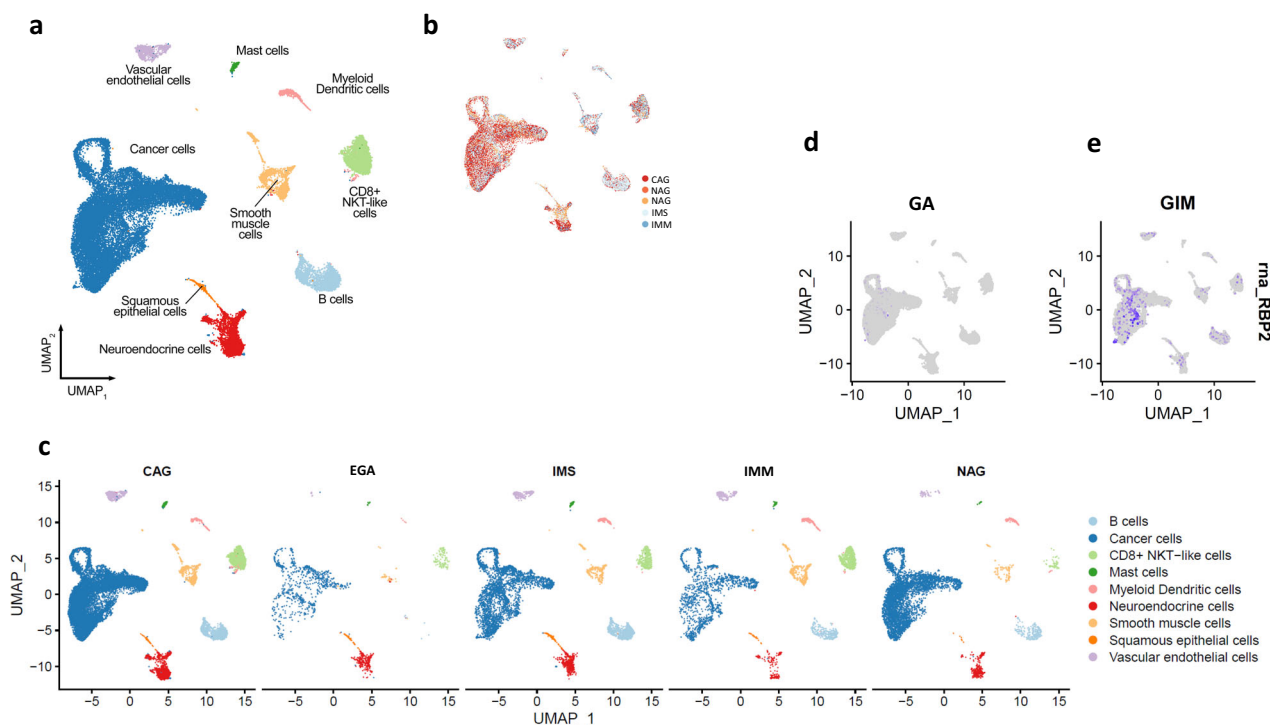


Fig. 2 | UMAP plots. Of note, cell type clusters (a) and annotations according to the pathological appearances (b, c) and low expression of the RBP2 gene in GA (d) but high expression in GIM (e). CAG chronic atrophic gastritis, EGA early gastric

adenocarcinoma, IMS severe intestinal metaplasia, IMM mild intestinal metaplasia, and NAG non-atrophic gastritis.

intestinal metaplasia (IMM), and nonatrophic gastritis (NAG) (Fig. 2b, c). EGA displayed the lowest gene expression in relation to cancer cells, vascular endothelial cells, smooth muscle cells, myeloid dendritic cells, B cells, CD8 + NKT-like cells, and mast cells among the five pathological appearances. GIM, including both IMM and IMS, displayed similar distributions of cell-type gene expression in comparison with NAG or CAG. Differential expression analysis between GIM and GA showed gene profiles and identified genes that were uniquely expressed in cancer cells, particularly RBP2 expression not in GA but in GIM (Fig. 2d, e).

GIM exhibited different signaling pathways than GA and displayed GA-related molecular network

Pathway analysis showed 110 canonical pathways that were either activated or inactivated in GIM or GA (Fig. 3a, b and supplementary Table S1). It should be noticed that there was a negative (albeit weak) correlation between GIM and GA (Fig. 3c). In particular, activated pathways in both GIM and GA were regulation of cellular mechanics by calpain protease, neuregulin signaling, IL-22 signaling, Rac signaling, MIF regulation of innate immunity, Fcγ receptor-mediated phagocytosis in macrophages and monocytes, regulation of eIF4 and p70S6K signaling, salvage pathway of pyrimidine ribonucleotides, HGF signaling, ERK/MAPK signaling, Wnt/β-catenin signaling, p38 mark signaling, role of BRCA1 in DNA damage response, sirtuin signaling pathway, NRF2-mediated oxidative stress response, γ-glutamyl cycle, CDK5 signaling, D-myo-inositol-5-phosphate metabolism, glycogen degradation, ephrin receptor signaling, mTOR signaling, IL-8 signaling, and apelin liver signaling pathway. HIPPO, p53 toll-like receptor signaling, aryl hydrocarbon receptor signaling, LPS/IL-1-mediated inhibition of RXR function, EIF2 signaling, and the NER pathway were inactivated in GIM but activated in GA, whereas fatty acid β-oxidation, nicotine degradation II and III, estrogen biosynthesis, melatonin degradation I, the superpathway of melatonin degradation and serotonin degradation were activated in GIM but inactivated in GA.

Microarray gene expression (MGE) revealed 19,223 out of 20,918 gene IDs and scRNAseq mapped 17,921 out of 22,910 gene IDs. A core analysis

was run on each dataset with the criteria of adjusted p -value < 0.05 and an absolute \log_2 FoldChange of 1. Co-analysis of MGE and scRNAseq was used to filter the genes in scRNAseq dataset and showed that 1,158 genes were expressed in both datasets (excluding the genes that were present in more than one cell cluster). Furthermore, 865 out of 1158 genes had equivalent IDs in IPA (Table S2).

Pathway enrichment analysis using a combination of three databases, i.e., Protein-Protein Interaction Networks Functional Enrichment Analysis (<https://string-db.org/>), Protein, Genetic and Chemical Interactions (<https://thebiogrid.org/>) and IntAct Molecular Interaction Database (<https://www.ebi.ac.uk/intact/home>) was performed. A biomarker network of the 865 genes in connection with the genes/proteins and metabolites was created (Fig. 4). Of note, this analysis reveals not only protein-protein interactions but also genetic and chemical interactions (e.g. elaidic acid and D-glucose). The most influential node was Wnt signaling pathway which has directly connections with HIPPO signaling, hepatocyte growth factor (HGF) signaling, CDK5 (pro-apoptotic signaling), Rac signaling, and EpCAM signaling and indirectly connections with other signaling and metabolic pathways (Fig. 4).

Furthermore, functional enrichment analysis of the 865 genes using Cytoscape (including NetworkAnalyzer) was performed. It revealed a comprehensive network of signaling pathways or terms (Fig. 5, Table S3) and topological coefficient (Table 1 and Table S4). Of note, there are overlaps, to a large extent, between the significant pathway and the gene ontology term. The Cytoscape analysis confirmed (with $p < 0.01$) the involvements of Wnt and Wnt-related signaling (Wnt signaling in cancer, Wnt target genes, Wnt5A-dependent internalization of FZD4, Wnt mediated activation of DVL, Hippo, Yap1, RUNX3 regulates Wnt signaling, catenin Beta 1 (CTNNB1) and binding if TCF/LEF: CTNNB1 to target gene promoters). It was noticed that 'lactose synthesis' was nearly 100% involved (Table S3).

The stability of network in resisting to external interference and attacks depends on key nodes³¹. The network topological analysis revealed potential key nodes. Of note, cytochrome c-mediated apoptosis, COPI-mediated

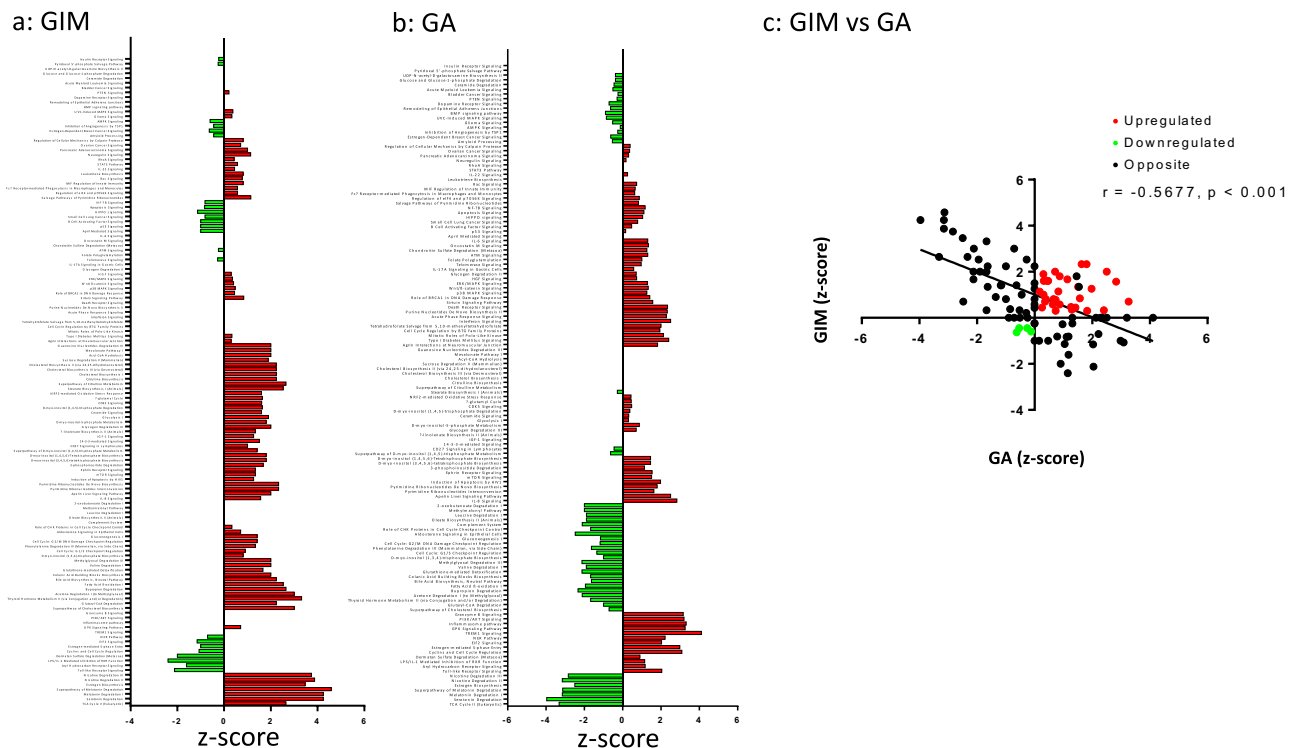


Fig. 3 | Study design. Canonical pathway activation and inhibition in GIM (a) and GA (b). Note: Pathways with obtained z-scores included [z-score > 0: activation (red); z-score < 0: inhibition (green); z-scores from -4 to 6 in (a) and from -6 to 6 in (b)]. Pearson correlation with linear regression between GIM and GA (c). $r = -0.5677$ and $p < 0.001$ in (c).

anterograde transport, membrane trafficking, vesicle-mediated transport, laminin interactions and apoptosis were among the top key nodes. Other key nodes included disorders of nervous system development and axon guidance, Ras mutation and RAF signaling (Table 1). (for a completed list of pathways/terms, see Table S4).

Computational predictions of potential drug targets and drug repurposing

Protein-protein interactions (PPI) analysis using IPA showed 330 out of 865 GIM-related proteins having PPI > 3 were identified, particularly, including epidermal growth factor receptor (EGFR), proto-oncogene tyrosine-protein kinase (SRC), paxillin (PXN), Jun proto-oncogene (JUN), breast cancer type 1 susceptibility protein (BRCA1), tumor protein P53 (TP53 or p53), mouse double minute 2 homolog (MDM2) (also known as E3 ubiquitin-protein ligase MDM2), and CTNNB1 (catenin β 1) and CD44 (Fig. 6, Table 2 and Table S5). It should be pointed out that CD44 was selected because CD44 is an important biomarker of stem cells, particularly in connection with Wnt signaling³²⁻³⁶. It has been known that Wnt/ β -catenin pathway and mucins play important roles in regulating neoplastic transformation and malignant growth, including GA^{33,37,38}. In GIM, CD44 had PPIs with MMP2, MMP7, MMP9, IFN γ , SLC9A and CSPGs (Fig. 6). IPA drug repurposing of approved and investigational drugs/compounds (e.g., existing at www.clinicaltrials.gov) showed the following targeted drugs: dasatinib with or without erlotinib, nilotinib, afatinib on SRC, SM1-71 on SRC, JNJ-26483327 on EGFR, Afatinib on EGFR, and ingenol mebutate on protein kinase C (PKC) family (Fig. 6).

Utilization of the Human Protein Atlas (v23.proteinatlas.org)³⁹ confirmed that the expression of the 8 out 9 proteins (except CD44) in both normal/GIM and GA tissue samples (Fig. 7a). Apparently, quantity of immunostaining was higher in GA (< 75%) than GIM (< 25% or 25-75%). Of note, tissue microarray data at the Human Protein Atlas didn't show CD44 (Fig. 7a, Table S6). Thus, our immunohistochemistry showed that the expression of CD44 was expanded from basal part of mucosa of normal and GIM stomachs to upper part of mucosa of GA stomach (Fig. 7b).

The gene expression of these 9 proteins showed that PXN and CD44 were up-regulated in both GIM and GA ($q < 0.001$, $n = 15$), SRC and CTNNB1 were upregulated in GA but not in GIM ($q < 0.001$ for SRC and $q = 0.05$ for CTNNB1, $n = 15$), and JUN was upregulated in GIM but down-regulated in GA ($q < 0.001$, $n = 15$) in comparison with the normal tissues from the same stomachs (Table S6). Furthermore, scRNAseq showed these proteins (except BRCA1) were expressed in goblet cells in addition to other cell types (Table S6).

The molecular docking studies on FDA-approved compounds against EGFR, SRC, PXN, JUN, BRCA1, p53 (monomer and dimer), MDM2, CTNNB1 and CD44 demonstrated high binding affinities of these compounds to the respective proteins. After initial analysis, 15 compounds were selected for each protein based on their binding affinities followed by a structural analysis of the interaction between the protein of interest and selected components, including the binding affinity and interacting residues (length, number and type of bonds). Accordingly, top five compounds per target protein were chosen based on their binding affinity to further analyze their interactions and the nature of the bonds formed. PyMOL provided excellent visualization of these interactions, revealing a detailed and extensive array of bonds between the selected compounds and the amino acids of the target proteins (Table 3). Out of 1453 compounds, the five highest binding affinities for EGFR were obtained for E155, gliquidone, gossypol, troglitazone, ZINC3830383. In regard to SRC, the highest binding affinity was noticed in cases of ZINC3830342, ZINC3830343, ZINC3830371, ZINC3830384, and ZINC3830386. For PAX, accolate, deferin, troglitazone, ZINC1612996 and ZINC3830342 were found. For the case of Jun the highest binding affinities were obtained for differin, E155, ZINC3830369, rolapitant and ZINC1612996. In regard to BRCA1 the highest binding affinities were noticed for ZINC3830342, ZINC3830343, ZINC3830384, E155 and differin. The highest binding affinity in case of p53 dimer were obtained for E155, E155, ZINC607700, ZINC1612996, ZINC3830342 and ZINC3830384. When it comes to MDM2 the highest binding affinities were acquired for rolapitan, risperdal, accolate, ZINC3830371 and ZINC3830372. The highest binding affinity for

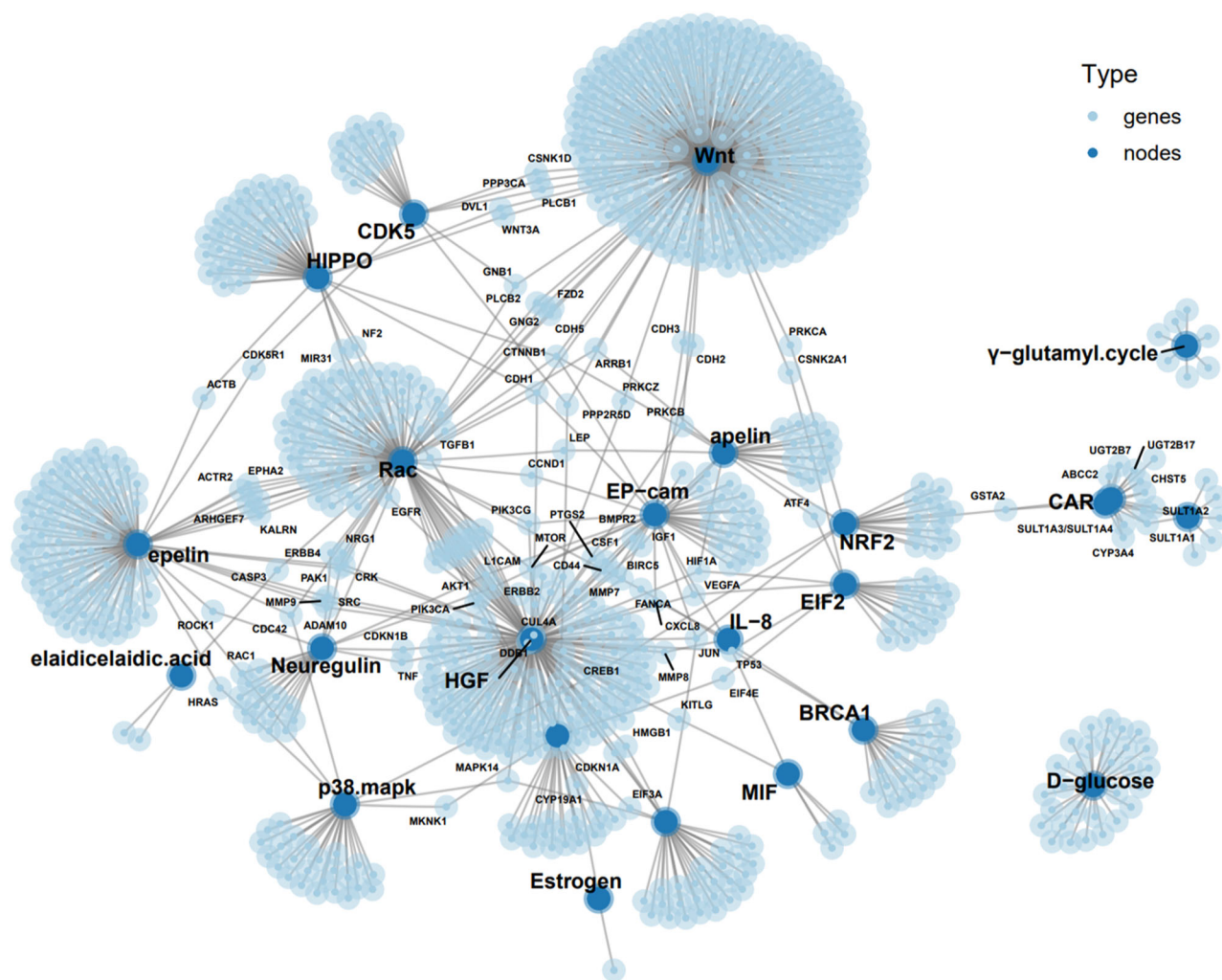


Fig. 4 | Network of molecules highlighting 23 proteins/genes/metabolites characteristic to GA-related GIM, particularly in connection to stem cell-related Wnt-HIPPO genes. Note: sizes of nodes and genes reflect numbers of interactions.

CTNNB1 was obtained for ZINC3830383, E155, ZINC3830369, ZINC3830385 and ZINC3830386. For CD44 the highest binding affinities were acquired in cases of these molecules E155, ZINC3830342, ZINC3830383, and ZINC3830430 (Table 3). It should be noticed that ZINC3830342 and ZINC3830343 are relatively identical compounds with the same molecular characteristics, but with a slight difference in the structural formula, and therefore identical results are obtained (Table 3). Considering that it was a blind docking analysis, the active/binding sites were not taken into account when creating the gridbox. The listed compounds (E155, rolapitant, ZINC1612996 and ZINC3830369) bound and showed the highest binding affinity for a specific region, although interactions with similar residues were achieved, the number and length of hydrogen bonds were different.

Discussion

Biomarkers are needed in risk assessment, screening, differential diagnosis, determination of prognosis, prediction of response to treatment, and monitoring of progression from GIM to GA^{40,41}. Particularly, the biomarkers specifically tailored for GA-related GIM could serve as an ideal tool in establishing personalized endoscopic surveillance programs and treatments.

It has been proposed that the incomplete GIM by pathological evaluation might be useful as a biomarker for GA-related GIM, as there is an association between the incomplete GIM and GA in comparison with the complete GIM. Some facilities (particularly in Japan) have recently

attempted to perform Endoscopic Grading of Gastric Intestinal Metaplasia such as the Operative Link on Gastric Intestinal Metaplasia Assessment (OLGIM)^{42–44}. However, it is practically a challenge due to economic concern, invasiveness to the patients and complex task against clinical daily task flow. The literature also shows high activities in identified potential biomarkers of incomplete GIM^{45,46}. In the present study, our aim was not to identify biomarkers for complete vs incomplete GIM rather than to identify biomarkers of GA-related GIM. Accordingly, the study subjects were GA patients and the study samples of GIM and GA were collected within the same stomachs of GA patients.

By creating GEM based on pathological diagnosis and single cell atlas, we were able to show ‘the gene expression profile’ for the individual patients or subgroups of patients. However, it was difficult to visualize specific clusters of genes for GIM-related GA within the user’s knowledge domain. It would be possible in the future if powerful visual analytics can be developed, e.g. on linking a reordered GEM heatmap and dual 2D projections of its rows and columns, which can be recomputed conditioned by subsets of genes and/or samples selected by the user during the analysis⁴⁷.

We further utilized our computational platform ScType (developed by AI et al.³⁰) to create UMAP plot and found that the gene expression by “cancer cells” and immune cells (such as dendritic cells, B cells, CD8 + NKT-like cells and mast cells) appeared to be lowest in GA than non-dysplastic lesions including chronic atrophic gastritis, GIM and non-atrophic gastritis. Presumably, the inflammation in the tumor micro-environment suppressed the gene expression by the cancer cells, particularly

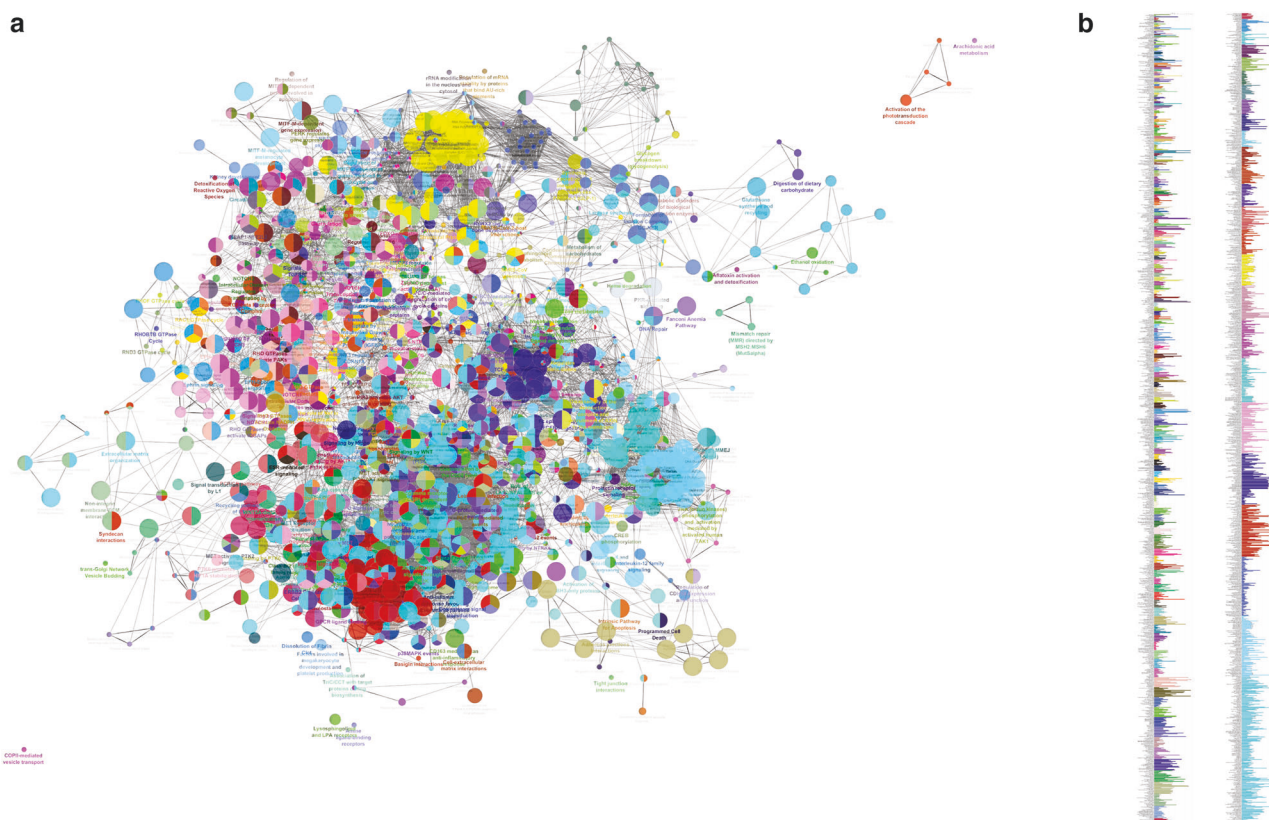


Fig. 5 | Network of signaling pathway characterized as ‘Team’ for GA-related GIM. Note: sizes of nodes reflect percentages of team-related genes per pathway and links presents the connections between pathways (a); percentages of gene/team (b) (for detailed percentages of each team, see Table S3).

Table 1 | Selected key nodes with topological coefficients

Term	T_n	p -Value*	UNIQUER_ID
Cytochrome c-mediated apoptotic response	1.0	0.0307	HSA:111461
COPI-mediated anterograde transport	1.0	0.0323	HSA:6807878
Membrane Trafficking	0.95	0.0171	HSA:199991
Vesicle-mediated transport	0.95	0.0099	HSA:5653656
Laminin interactions	0.93	0.0298	HSA:3000157
Apoptosis	0.92	0.0322	HSA:109581
Programmed Cell Death	0.92	0.0487	HSA:5357801
Disorders of Nervous System Development	0.87	0.0330	HSA:9697154
Asparagine N-linked glycosylation	0.84	0.0281	HSA:446203
Intrinsic Pathway for Apoptosis	0.75	0.0364	HSA:109606
Apoptotic factor-mediated response	0.75	0.0281	HSA:111471
Signaling by RAS mutants	0.52	0.0459	HSA:6802949
Paradoxical activation of RAF signaling by kinase inactive BRAF	0.52	0.0459	HSA:6802955
Signaling downstream of RAS mutants	0.52	0.0459	HSA:9649948
Signaling by RAF1 mutants	0.52	0.0351	HSA:9656223
Axon guidance	0.50	0.0167	HSA:422475

T_n : topological coefficient ($T_n = \text{avg}(J(n,m)) / k_n$).

* p Value: Corrected with Benjamini-Hochberg.

at the early stage. It was unexpected, as it has been well documented that inflammation predisposes to the development of cancer and promotes all stages of tumorigenesis⁴⁸. It was also unexpected, as it is known that chronic inflammation facilitates tumor progression, whereas acute inflammatory

may stimulate the maturation of dendritic cells and antigen presentation, leading to anti-tumor immune responses⁴⁹, and the tumor-promoting inflammation may be regarded as the host’s defense against malignancy⁵⁰.

Furthermore, we found RBP2 expression in GIM but not in GA. Retinoblastoma binding protein 2 (RBP2), also known as cellular RBP2 (CRBP2), belongs to the family of intracellular lipid-binding proteins known as fatty acid-binding proteins (FABPs). Several reports showed that RBP2 stimulated various processes such as HIF-1 α -VEGF-induced angiogenesis in non-small cell lung cancer through the Akt pathway⁵¹, malignant progression of GA through the TGF- β 1-(p-Smad3)-RBP2-E-cadherin-Smad3 feedback circuit⁵², induction of stem-like cancer cells by promoting epithelial-mesenchymal transition in renal cell carcinoma⁵³, and initiation of ER and IGF1R-ErbB signaling in tamoxifen resistance in breast cancer⁵⁴. However, it should be kept in mind that attributing a single gene as a biomarker would be risky, considering that GA is neither a monogenic nor a polygenic malignancy.

By concentrating on pathways as potential biomarkers for GA-related GIM, we used Ingenuity Pathways Analysis (IPA) to create the bar graphs consisting of 110 canonical signaling pathways. We found that while the activated signaling pathways in GIM might be deemed biomarkers for GA, some of these pathways were inactivated in GA when compared with GIM. Thus, more attention should be given to the dynamic changes in the progression from GIM to GA. Furthermore, we found that it was a moderate negative correlation between GIM and GA, which was unexpected as GIM is believed to be the primary precancerous lesion in GA tumorigenesis. As shown in the literature, only some GIM, but not all, is considered to be the precursor of gastric cancer⁵⁵. In the present study, more samples and patients are needed for establishing the topologic association and more importantly the causality between GIM and GA in the future.

It should be kept in mind that association should not be confused with causality; if GIM causes GA, then the two are associated (dependent). However, associations can arise between variables in the presence (i.e., GIM

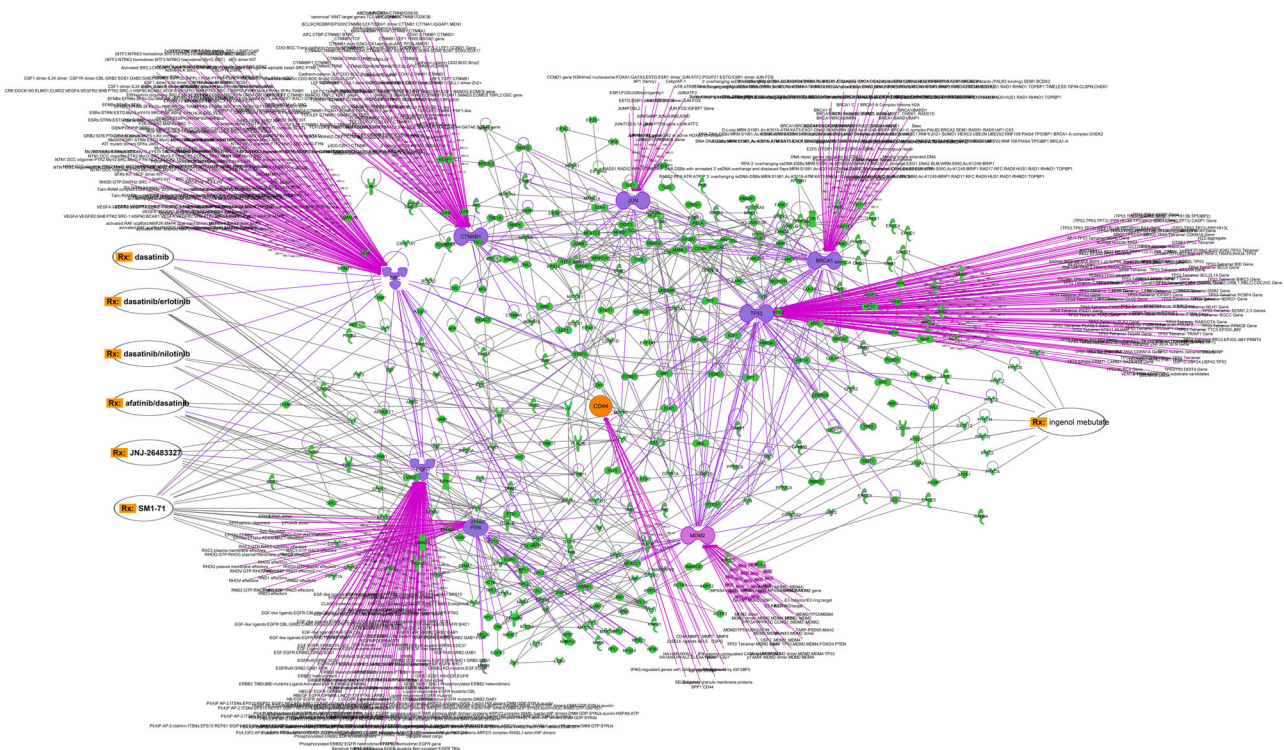


Fig. 6 | Protein-protein interactions. Note: Hub proteins in GIM-GA “niche” (indicated in dashed line) and links with the repurposed drugs (Rx, indicated in orange) revealed by IPA analysis.

Table 2 | Hub proteins with drug-repurposing targets

Protein name	Degree	Drugs
EGFR	92	JNJ-26483327, Dasatinib, dasatinib/erlotinib, afatinib/dasatinib, dasatino/nilotinib,
SRC	124	Dasatinib, dasatinib/erlotinib, dasatino/nilotinib, afatinib/dasatinib, SM1-71, ingenol mebutate (via STAT3)
PXN	48	SM1-71
JUN	30	
BRCA1	66	
P53	119	
MDM2	77	
CTNNB1	83	
CD44	13	
Multi-proteins		Ingenol mebutate

The hub proteins are ranked by their degree of connectivity within the network.

causes GA) and absence (i.e., they have a common cause) of a causal relationship⁵⁶. By performing the functional enrichment analysis which was based on our data in combination with public available databases, we have created the biomarker network and the network of signaling pathways of GA-related GIM. We found Wnt/ β -catenin pathway to be one of the most important biomarkers for GA-related GIM. This was in line with the literature of Wnt signaling in regulating neoplastic transformation and malignant growth, including GA^{33,37,38}. By additional functional annotation and validation analysis (e.g. immunohistochemistry), we found gastric stem cell marker CD44 as GA-related GIM. These findings may support the notion that GIM is the primary precancerous lesion in GA tumorigenesis, as both share the “cause” (i.e., hyperactivity of Wnt signaling and the presence of tissue-resident stem cells)^{32–34}. Wnt signaling is often implicated in stem

cell control as a proliferative and self-renewal signal. Mutations in Wnt genes or Wnt pathway components lead to specific developmental defects, while various diseases, including cancers, are caused by abnormal Wnt signaling⁵⁷. Although molecular targeting of the Wnt signaling system has been proposed and trialed for the treatment of various cancer types without significant success yet^{58,59}, Wnt signaling, particular CD44, should be potential target in the future.

In addition to drug targeting on individual hub/key genes and proteins that are highly expressed in GA-related GIM, we should consider the key nodes that keep the stability of network as potential targets. By topological analysis, we found several potential nodes for GA-related GIM, e.g., apoptosis (such as cytochrome c-mediated apoptosis and laminin interactions and apoptosis) and intracellular trafficking (such as COPI-mediated anterograde transport, membrane trafficking, vesicle-mediated transport). Indeed, dysregulation of apoptosis has been well recognized as a hallmark of cancer cells due to mutations in the extrinsic, intrinsic, p53, and the related signaling pathways. Accordingly, current efforts have been made to develop agents that target apoptotic pathways either directly or indirectly⁶⁰. Targeting membrane trafficking particularly in connection with Wnt signaling has also been proposed as a strategy for cancer treatment⁶¹.

The results of the present study by the topological analysis revealed the nervous system development and axon guidance as potential key node in GA-related GIM, which was in line with our previous studies that demonstrated the role of vagal nerve in gastric tumorigenesis^{33,62}. Another key node was RAS/RAF/MAPK pathway. Indeed, RAF inhibitors (RAFi) combined with MEK blockers have been taken as an FDA-approved therapeutic strategy for numerous RAF-mutant cancers, including melanoma, non-small cell lung carcinoma, and thyroid cancer⁶³. It would be of interest to include GA-related GIM as a prevention strategy.

We utilized IPA analysis of PPI and identified the hub proteins including EGFR, SRC, PXN, Jun, BRCA1, TP53, MDM2 and CTNNB1. CD44 was included not because of its PPI but the biological importance in GIM-related GA as forementioned. The hub proteins were initially identified by deploying the Causal Network Analysis, enabling us to

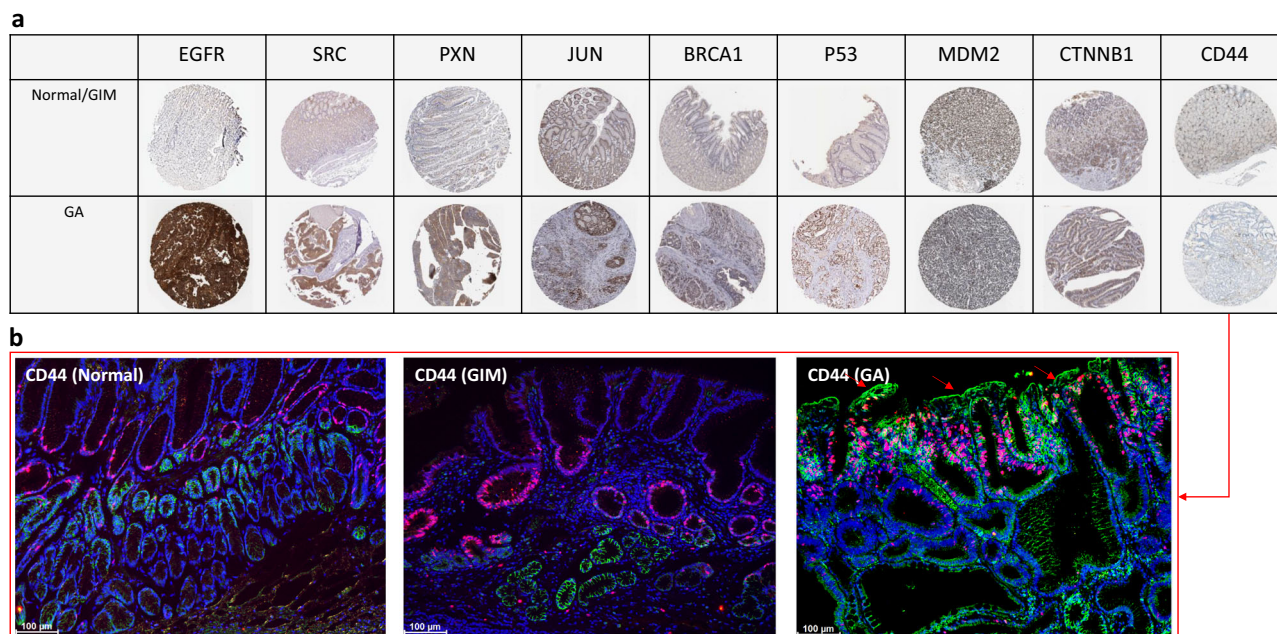


Fig. 7 | Representative immunohistochemistry of 9 hub proteins. Note: tissue microarray at Human Protein Atlas (a) and immunofluorescent staining (b) in which CD44 is indicated in green and Ki67 in red, and CD44 on goblet cells in (arrows). Bar = 100 μ m.

identify master regulators (hubs) that act directly upon other dataset molecules, or through one or more intermediate upstream regulators. The analysis was performed using data from gene expression profiling and then validated by tissue microarray (i.e., immunohistochemistry) that are publicly available at the Human Protein Atlas and immunofluorescence for CD44. Furthermore, by performing drug-hub protein targeting interactions, we identified the following potential repurposed drugs for treatments of GA-related GIM.

Dasatinib is a tyrosine kinase inhibitor (Sprycel; Bristol-Myers Squibb) and has been used for the treatment of chronic myeloid leukemia and Philadelphia-chromosome-positive acute lymphoblastic leukemia⁶⁴. Recent studies suggest an alternative mechanism of action for dasatinib, which involves augmenting the population of functionally active CAR T cells⁶⁵. SM1-71 is a multitargeted kinase inhibitor that includes MKNK2, MAP2K1/2/3/4/6/7, GAK, AAK1, BMP2K, MAP3K7, MAPKAPK5, GSK3A/B, MAPK1/3, SRC, YES1, FGFR1, ZAK (MLTK), MAP3K1, LIMK1, and RSK2⁶⁶. JNJ-26483327 is a multitarget tyrosine kinase inhibitor⁶⁷. Adapalene, also known as CD271 and differin, belongs to the 3rd generation of retinoids that were approved by the Food and Drug Administration (FDA) in 1996 to treat acne vulgaris. A recent study using *in silico*, *in vitro*, and *in vivo* models suggested that adapalene could be used for treatment of multiple myeloma and Leukemia⁶⁸. Irinotecan liposome has been approved by FDA on Feb. 13, 2024 as first-line treatment of metastatic pancreatic adenocarcinoma (<https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-irinotecan-liposome-first-line-treatment-metastatic-pancreatic-adenocarcinoma>). Gliquidone has been widely used for the treatment of type 2 diabetes. There was no report yet for treatment of cancer but possibly for neuroinflammatory disease⁶⁹, which may share some common features with GIM as shown in this study. Gossypol, also called AT-101, has been tested in clinical trials as a single agent or in combination with standard therapy for various cancer types, showing a trend toward increased overall survival and progression-free survival⁷⁰. Troglitazone as a PPAR γ ligand has been reported to exhibit antiproliferative effects and/or enhance cancer immunotherapy^{71,72}. Other drugs, such as rolapitant that is used as an antiemetic in oncology⁷³, accolate for chronic treatment of asthma, and risperidol for mental/mood disorders, and other ligands shown in Table 3 will be of interest in searching for repurposing to treat GIM.

Computational molecular docking has been used as a tool for the discovery of repurposed drugs^{74,75}. By the docking analysis with PyMOL, we identified and visualized five compounds per target protein based on their binding affinity, the nature of the bonds formed and interactions. We presented the repurposed drugs and believed that the information will be useful not only for the identified drugs but also for development of new drugs as potential ‘prototypes’, particularly for CD44 which has the highest binding affinities with the following molecules including E155, ZINC3830342, ZINC3830383, and ZINC3830430.

Taken together, the methods and the results of this study could help in the targeted selection of GIM individuals who would benefit the most from screening and surveillance endoscopy and in the development of treatment for GA-related GIM. The multi-bioinformatics included GEM, ScType on scRNA-seq, IPA, Cytoscape, tissue microarray at HPA, Gene set enrichment analysis (GSEA) (also called functional enrichment analysis or pathway enrichment analysis) and molecular docking analysis, in addition to preliminary analyses with pagerank centrality, GO and MSigDB enrichments (data not shown).

This study has at least the following four limitations. First, after identifying potential biomarkers, the next step is validation. One of possible *in vivo* models would be the INS-GAS mice in which the majority are with high risk of spontaneous progression from GIM to GA⁷⁶. Unfortunately, we didn’t have capacity at the time of validating the repurposed drugs as reported in the present study. Furthermore, we didn’t find the individual specific biomarkers that were correlated with the spatial progression of GIM to GA in terms of express levels and grade of dysplasia, and we were unable to identify the individual specific biomarkers that were upregulated in patients with nondysplastic GIM that later progress to GA. Therefore, we proposed the GEM, signaling pathways, molecular network and network of signaling pathways. Presumably, this study (together with other studies in the literature) might contribute to the future development of AI/ML-supported *in silico* models in prediction of the progression of GIM patients to GA. Second, the repurposed drugs proposed in this study should be assessed in detail in terms of risk-benefit analysis. Many of the drugs that were listed in the present study are fairly toxic and would not be recommended (either alone or in combinations) unless the risk of GA progression is extremely high. However, it would be possible to take these repurposed drugs and other molecular ligands as “lead compounds or prototypes” for the drug

Table 3 | The hub proteins and ligand (repurposed drugs) interacting residues

Protein	Ligand	Binding affinity kcal/mol	Interacting residues
EGFR	E155 (ZINC3830332)	-10.3	Leu718; Gly719; Asp837 (3.7 Å); Arg841 (2.3 Å); Asn842, Ile878 (2.7 Å); Lys879 (2.3 Å) (four hydrogen bonds)
	Gliquinone (ZINC1482077)	-10.6	Gly762 (3.3 Å); Met766; Ala859 (3.5 Å); Leu861 (2.7 Å) (three hydrogen bonds)
	Gossypol (ZINC3775575)	-10.5	Gly721; Ala722 (3.1 Å); Gly724 (3.0 Å); Lys745 (2.0 Å); Met793 (3.4 Å) (four hydrogen bonds)
	Troglitazone (ZINC968276)	-10.5	Met793 (1.9 Å, 3.3 Å); Asp837 (3.3 Å); Arg841 (2.7 Å, 3.2 Å, 3.6 Å); Asn842 (2.0 Å) (forming a total of seven hydrogen bonds)
	ZINC3830383	-10.6	Lys745 (2.7 Å); Asp837 (3.4 Å, 3.8 Å); Arg841 (three hydrogen bonds)
SRC	ZINC3830342	-12.2	Met343 (3.8 Å); Gly346; Ser347 (3.3 Å); Ala392 (two hydrogen bonds)
	ZINC3830343	-12.2	Met343 (3.8 Å); Gly346, Ser347 (3.3 Å), Ala392 (two hydrogen bonds)
	ZINC3830371	-10.6	Arg162; Pro363 (3.5 Å); Asp367 (3.8 Å); Gly397 (3.4 Å); Asp399 (2.4 Å, 3.3 Å, 3.9 Å); Leu400 (2.0 Å), (seven hydrogen bonds)
	ZINC3830384	-11.6	Gln277 (3.4 Å); Cys279; Phe280 (3.0 Å); Lys297 (2.3 Å, 2.8 Å); Thr340; Ser347 (2.3 Å, 3.5 Å); Ala392 (3.7 Å), Asn393; Asp406 (3.3 Å, 3.9 Å) (nine hydrogen bonds)
	ZINC3830386	-11.5	Gln277 (3.4 Å); Gly278; Cys279 (2.0 Å); Lys297 (2.1 Å, 2.9 Å, 3.0 Å); Asp388 (3.4 Å); Arg390 (2.0 Å); Asn393; Asp406 (3.0, 3.9 Å) (nine hydrogen bonds)
PXN	Accolate (ZINC896717)	-5.9	Met616 (3.5 Å); Asp617 (3.4 Å, 4.0 Å); Asp618 (2.6 Å); Leu622 (3.4 Å); Asp625 (four hydrogen bonds)
	Differin (ZINC3784182)	-5.9	Leu622 (4.0 Å); Asp625 (3.3 Å) (two hydrogen bonds)
	Troglitazone (ZINC968276)	-5.9	Asp617 (3.7 Å); Asp618 (1.8 Å); Asp625 (two hydrogen bonds)
	ZINC1612996	-6.0	Leu623 (without forming hydrogen bonds)
	ZINC3830342	-5.9	Asp620 (without hydrogen bonds)
JUN	Differin (ZINC3784182)	-8.2	Arg272 (3.1 Å); Arg276; Glu281 (3.7 Å) (two hydrogen bonds)
	E155 (ZINC3830332)	-8.1	Ser269; Arg272 (2.1 Å, 3.6 Å); Glu275 (3.4 Å); Arg276 (2.6 Å); (four hydrogen bonds)
	Rolapitant (ZINC3816514)	-8.1	Ser269 (2.5 Å); Arg276 (2.6 Å) (two hydrogen bonds)
	ZINC1612996	-8.1	Arg272 (3.1 Å); Arg276; Glu281 (3.7 Å) (two hydrogen bonds)
	ZINC3830369	-8.1	Arg276 (2.2 Å, 3.3 Å); Glu281 (3.2 Å) (three hydrogen bonds)
BRCA1	ZINC3830342	-9.3	Trp1782; Gln1785 (without hydrogen bonds)
	ZINC3830343	-9.3	Trp1782; Gln1785 (without hydrogen bonds)
	ZINC3830384	-8.7	Sep406 (3.2 Å, 3.3 Å, 3.4 Å); Leu1657 (3.4 Å, 3.8 Å); Pro1659 (3.8 Å); Asn1678; Lys1702 (2.0 Å) (seven hydrogen bonds)
	E155 (ZINC3830332)	-9.1	Sep406 (3.3 Å, 3.5 Å); Leu1657 (3.2 Å, 3.7 Å); Thr1658 (3.5 Å), Arg1670; Asn1678 (2.8 Å) (six hydrogen bonds)
	Differin (ZINC3784182)	-8.6	Gln1811 (2.6 Å); Arg1835 (2.2 Å); Glu1836 (3.7 Å) (three hydrogen bonds)
p53 (dimer)	E155	-7.7	Gln100 (3.1 Å); Ala138 (3.4 Å); Thr140 (3.1 Å); Ser166 (3.1 Å, 3.4 Å); Asp186; Arg196 (1.9 Å); Glu198 (3.8 Å); Asn235 (3.7 Å) (eight hydrogen bonds)
	ZINC607700	-10.6	Lys164 (3.9 Å); Asp186 (3.4 Å); Arg196 (2.7 Å) (three hydrogen bonds)
	ZINC1612996	-10.5	Gln100 (2.1 Å); Lys164 (3.8 Å); Ser166 (2.0, 2.8 Å); Val197 (3.8 Å); Gly199 (2.5 Å); Asn200 (1.9 Å); Asn235 (3.8 Å) (eight hydrogen bonds)
	ZINC3830342	-10.6	Thr150; Asp228 (without hydrogen bonds)
	ZINC3830384	-10.4	ZINC 3830384 -10.4 - Leu137 (3.4 Å), Lys139 (2.2 Å), Ser166 (2.3, 2.6, 2.8 Å), Gln167, Asp186 (3.5 Å), Arg196 (2.1, 3.1 Å), Glu198 (3.4 Å) (nine hydrogen bonds)
MDM2	ZINC3830371	-10.2	Gln72 (3.2 Å, 3.4 Å, 3.6 Å); His96 (3.0 Å); Ile99; Tyr100 (four hydrogen bonds)
	ZINC3830372	-10.4	Gln18 (3.3 Å); Ile19 (2.2 Å, 3.6 Å); Gln24 (2.7 Å); Val93 (3.9 Å); His (2.1 Å) (six hydrogen bonds)
	Accolate ZINC896717	-10.3	Gln24, Leu54 (3.5 Å); His96 (one hydrogen bond)
	Risperidal ZINC538312	-10.6	Ile19 (3.5 Å); Pro20; Gln24 (2.4 Å); Leu54 (3.2 Å); Val93 (three hydrogen bonds)
	Rolapitant ZINC3816514	-10.7	Leu54 (3.5 Å) (one hydrogen bond)
CTNNB1	ZINC3830383	-9.2	His223, His260, Asn261, Lys292, Asp295, Asp299, Tyr333, Thr339, Arg342
	E155	-9.1	Asp249, Ser250, Thr289, Asn290
	ZINC3830369	-8.9	Asn290, Asp299, Thr339, Arg342
	ZINC3830385	-8.6	Gln266, Gly268, Ala269, Lys270, Met271, Tyr306, Gly307, Asn308
	ZINC3830386	-8.6	His260, Ala295, Asp299, Lys335, Thr339, Arg342
CD44	E155 (ZINC3830332)	-9.2	Lys42 (2.2 Å, 2.9 Å); Arg45 (1.8 Å, 3.3 Å); Tyr46 (3.1 Å, 3.7 Å); Ser47; Ser117 (2.1 Å); His118 (2.6 Å, 2.7 Å); Arg167 (2.5 Å); Asp172 (3.7 Å) (11 hydrogen bonds)
	Irinotecan (ZINC1612996)	-9.5	Asn29; Glu41 (3.6 Å); Glu79 (3.4 Å); Arg155 (two hydrogen bonds)
	ZINC3830342	-9.9	Glu41; Cys81 (2.4 Å) (one hydrogen bond)
	ZINC3830383	-9.5	Arg33 (2.9 Å); Phe60 (3.8 Å); Asn125 (1.9 Å, 2.4 Å, 3.4 Å, 3.6 Å); Ser127 (2.4 Å, 2.7 Å, 3.5 Å); Thr138 (nine hydrogen bonds)
	ZINC3830430	-9.3	Asn29 (2.9 Å, 3.6 Å); Val30; Thr31; Tyr34 (2.1 Å); His39 (3.0 Å); Glu79 (3.8 Å); Arg94 (1.9 Å, 2.2 Å, 2.6 Å); Asn (2.1 Å); Arg155 (nine hydrogen bonds)

development for treatment of GA-related GIM. Third, based on the results of this study and the literature, we proposed a hypothesis that GIM and GA share the common tissue-resident stem cells, leading to the progression from GIM to GA. One of possible validation methods could be organoid models of GIM vs. GA. Forth, the interpretations of the results reported in the present study (including main and supplementary results) were limited due to consideration of the length of Discussion.

Methods

Patients and study design

This study has been performed in accordance with the Declaration of Helsinki and approved by the Regional Committees for Medical and Health Research Ethics Central Norway (REK 2012-1029). The surgical samples were taken from 16 GA patients at St. Olav's Hospital. The stomach specimens were taken from four pre-determined positions in corpus (major and minor curvature), cardia and antrum per stomach (Fig. 8), including tumors, adjacent non-tumor tissues, and distanced non-tumor tissues, immediately after gastrectomy during 2005 to 2010. All patients were diagnosed histologically as primary gastric adenocarcinoma of stage I-IV. Patients have since been followed-up for five years at St. Olavs Hospital, Trondheim, Norway. TNM status was defined, and samples were classified according to Lauren's classification, (Intestinal, diffuse or mixed/combined type), WHO classification (tubular, papillary, mucinous and poorly cohesive), WHO grading (well, moderately or poorly differentiated). All histological samples without ones from one patient (no. 6) due to a technical error were reviewed according to the Japanese pathological classification. Samples were assigned gastric histopathology scoring including inflammation, epithelial defects, oxyntic atrophy, epithelial hyperplasia and dysplasia and an overall GHAI score. The samples from patient no. 6 did not undergo pathological evaluation and were not analyzed further. Thus, 15 patients were included for gene expression profiling. One or more of the four tissue samples displayed the tumor, and the remaining tissue samples showed a "normal" appearance. The samples were then evaluated histologically and subjected to Illumina microarray gene expression. In addition, the tissue biopsy samples were obtained as routine medical diagnosis by endoscopy from patients with gastritis and/or gastric cancer and were subjected to pathology assessment. Total eight samples from eight patients were used for the immunofluorescent staining of CD44 at Tokyo University Hospital, and the three representative samples were used in this study. Ethics protocol was approved at Tokyo University Hospital and written informed consent was obtained from all patients.

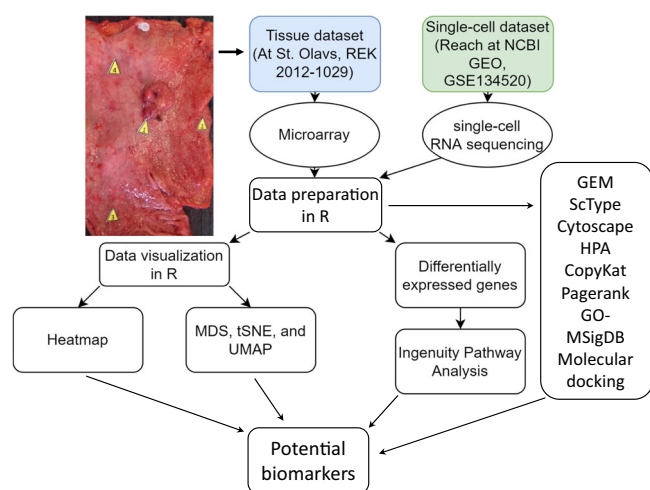


Fig. 8 | Study design showing multi-bioinformatics used for analysis of surgical biopsies (4 samples per stomach = total 60 samples from 16 patients), single-cell RNA sequencing from 13 patients⁹¹, immunohistochemistry from 19 patients.

Multi-bioinformatics

Illumina microarray gene expression data were extracted as an Excel document, while the control data results were extracted. Data wrangling and normalization were performed with R version 4.1.1 and the tidyverse package version 1.3.1. The data frame was filtered using the sample vector list containing the patient samples that were not used for further analysis and was then saved as a text file to be imported into lumi. The control data were similarly imported from a text file to the data frame of control before it was filtered and saved as a text file. Lumi version 2.44.0 was used to filter the unexpressed genes from the data. Genes with detection threshold < 1% in 10 or fewer samples were removed. A limit of 11 was set because the smallest group is GIM, which contains 11 samples. The patient pathology was imported as pdf (pathology dataframe) and filtered to be used as a design matrix. Any flawed samples were removed before the samples were filtered and categorized as GA, GIM or normal. Samples were categorized as cancer if they had a pathological evaluation as GA or a biopsy classification of 2 out of 5, where a biopsy classification of 1 is certain tumor tissue. Samples were categorized as GIM if they were not classified as GA but had a histopathological scoring that indicated intestinal metaplasia. Samples were categorized as normal if they were neither classified as GA nor GIM. After categorization, a new column concerning the sample pathology was made.

A single-cell transcriptomic atlas of premalignant lesions and early GA was modified for data visualization and analysis. All genes that had an adjusted *p* value above 5% were removed. Genes found in more than one cell type were removed when using the single-cell transcriptome biomarkers as cell type biomarkers for the tissue data. In this way, only the RNAs that were not significant in any other cell type would be used when making a heatmap or when utilizing cell type-specific clustering. Genes without a match in the sample data frame were cross-referenced with the UniProt knowledgebase. The genes from the single-cell transcriptomics were then sorted to coincide/align with the microarray data. The data of the single-cell transcriptomic atlas were also used to confirm the findings by using two different datasets in Seurat version 4.0.4. Cells expressing less than 400 genes, more than 7000 genes or cells containing more than 20% genes correlated to the mitochondria were filtered out.

The housekeeping genes *CTBP1*, *CUL1*, *DIMT1L*, *FBXW2*, *GPBP1*, *LUC7L2*, *OAZ1*, *PAPOLA*, *SPG21*, *TRIM27*, *UBQLN1*, *ZNF207*, *AGPAT1*, *B2M*, *CAPN2*, *CYCC (CCNC)*, *PMM1*, *SDHA*, *RPL29*, *RPL29-B2M*, and *B2M-GAPDH* were used for normalization using geNorm, a computational method, from ctrlGene version 1.0.1. It should be noted that *OAZ1* was removed from the normalization due to high intensity output. *SPG21* was removed from the normalization due to a very low expression value. The data were also normalized with DESeq2 version 1.32.0. (It should be mentioned that the effects of different normalization methods on the outcomes were not included in the present study).

Seurat version 4.0.4 was used to perform dimensional reduction. Normalized data from the patients were inputted into Seurat and scaled, clustered and plotted through Seurat functions: CreateSeuratObject, ScaleData, FindVariableFeatures, RunPCA, FindNeighbors, FindClusters, RunUMAP and RunTSNE. The clustering methods performed were UMAP and tSNE. Important parameters included 1) selection.method = vst, 2) nfeatures = 5000, 3) dims = 1:10 and 4) perplexity = 10. Heatmaps were created by ComplexHeatmap version 2.8.0. The differential equations were made with both DESeq2 and limma. The single-cell transcriptomic atlas was tested in DESeq2, whereas the microarray data were tested in both DESeq2 and limma.

The tissue and single-cell transcriptome datasets and their differentially expressed genes (DEGs) were imported into IPA (Qiagen). The microarray data mapped 1 9223 of 2 0918 gene IDs, and the single-cell transcriptome mapped 17921 if 22910 gene IDs to equivalent IDs in IPA. A core analysis was then run on each dataset with the criteria of an adjusted *p* value < 0.05 and an absolute log₂FoldChange of 1. Each dataset was considered for their signaling pathways. A comparison analysis was performed between the scRNA-seq and the background-corrected lima processed dataset.

A co-analysis of all patient samples was performed with ScType to visualize cell populations and the cell type abundance for each tissue type³⁰. To define the subset of cancerous cells, aneuploidy analysis was performed using CopyKat. The underlying idea is that gene expression levels of many adjacent genes can provide depth information to infer genomic copy number in that region, and cells with extensive genome-wide copy number aberrations (aneuploidy) are considered cancer cells. In the case of the raw fastq files or BAM file, real variant calling was performed to validate the subset of cancer cells in consideration of CopyKat, which is well correlated with SNV calling, especially for solid tumors³⁰. Furthermore, to investigate latent disease-related regulatory changes that are invisible based on clustering or differential expression analysis, regulatory networks from single-cell data and quantify gene centralities were inferred. PageRank centrality was calculated to identify the nodes with high Pagerank centrality, indicating “popular” genes involved in multiple regulatory pathways. GO and MSigDB enrichments were performed to visualize overrepresentation of KEGG_CELL_ADHESION_MOLECULES_CAMS⁷⁷. Network graph was built using igraph v1.4.3 and visualized with ggplot2 v3.3.6.

Functional enrichment analysis was performed using Protein-Protein Interaction Networks (<https://string-db.org/>), Protein, Genetic and Chemical Interactions (<https://thebiogrid.org/>) and IntAct Molecular Interaction Database (<https://www.ebi.ac.uk/intact/home>) and Cytoscape 3.10.

The Human Protein Atlas was used, including pathology data with immunohistochemistry using semi-quantitative tissue microarrays based on The Human Protein Atlas version 23.0 and Ensembl version 109 and RNA single cell read count data based on 31 datasets (<https://www.proteinatlas.org/>)^{39,78–84}.

Blind docking analysis was performed using 100 random compounds from the FDA-approved drug list. The three-dimensional crystal structures of target proteins were retrieved from the RCSB Protein Data Bank (PDB) in PDB format⁸⁵. These proteins include the epidermal growth factor receptor (PDB ID: 5U8L), proto-oncogene tyrosine-protein kinase Src (PDB ID: 8XN8), paxillin (PDB ID: 5W93), transcription factor Jun (PDB ID: 5T01), breast cancer type 1 susceptibility protein (PDB ID: 4Y2G), tumor protein p53 (PDB ID: 8A31), E3 ubiquitin-protein ligase MDM2 (PDB ID: 4MDN), and CD44 antigen (PDB ID: 55C3). Initially, these structures were in complex formations with other molecules. To facilitate further analysis, protein preparation was performed using AutoDock Tools 1.5.6⁸⁶, which involved removing water molecules, adding polar hydrogen atoms, and assigning Kollman charges. Following these modifications, the structures of the target proteins were converted from PDB to PDBQT format. Additionally, AutoDock Tools 1.5.6 was utilized to define a grid box for each protein, setting the stage for subsequent docking simulations. The grid box dimensions for EGFR were 48 × 49 × 61 Å and it was centered at 5.549, -6.743, and -25.974. For SRC the dimensions were set to 64 × 62 × 66 Å and it was centered at 23.298, 0.183, and 15.596. For PAX the grid box dimensions were 24 × 24 × 24 Å, centered at 3.757, -7.759 and 32.776. The grid box dimensions for Jun were 68 × 70 × 68 Å and it was centered at -16.564, 21.969 and 24.614. For BRCA 1 the dimensions were 68 × 50 × 40 Å, centered at -26.133, 15.257 and -27.353. For p53 the grid box dimensions were 40 × 48 × 50 Å, centered at 98.225, 80.184 and -29.347. The grid box dimensions for MDM2 were 36 × 34 × 44 Å and it was centered at -14.982, 0.643 and 1.031, while for CD44 the dimensions were set to 36 × 48 × 40 and it was centered at 0.933, -0.201 and 5.715. For each analysis the value of spacing (ångstrom) was set to 1.0. Three-dimensional structures of FDA approved drugs were available in the ZINC database⁸⁷. Compounds were initially downloaded in the appropriate SDF format and then prepared and converted to PDBQT format using OpenBabel 3.1.1 software with default settings. During this preparation phase, various adjustments were applied to the ligands, including the addition of charges, hydrogen atoms, assignment of atom types, conversion of bond types, and establishment of the root. It is important to highlight that the default settings of OpenBabel generate the molecule's 3D structure in a neutral state without considering any ionization states, a feature maintained in the PDBQT format of the ligands employed in this study⁸⁸. Molecular docking simulations were conducted utilizing

AutoDock Vina 1.1.2, employing an energy range of four and an exhaustiveness value of 32⁸⁹. The interactions between the receptors and the chosen compounds were subsequently visualized and analyzed using PyMOL 2.5.4⁹⁰.

Data availability

All data used in this study are available within the manuscript and the raw data will be provided per request.

Received: 5 June 2024; Accepted: 13 October 2024;

Published online: 04 November 2024

References

- Eriksson, N. K., Karkkainen, P. A., Farkkila, M. A. & Arkkila, P. E. Prevalence and distribution of gastric intestinal metaplasia and its subtypes. *Dig. Liver Dis.* **40**, 355–360 (2008).
- Olmez, S., Aslan, M., Erten, R., Sayar, S. & Bayram, I. The Prevalence of Gastric Intestinal Metaplasia and Distribution of Helicobacter pylori Infection, Atrophy, Dysplasia, and Cancer in Its Subtypes. *Gastroenterol. Res Pr.* **2015**, 434039 (2015).
- Eidt, S. & Stolte, M. Prevalence of intestinal metaplasia in Helicobacter pylori gastritis. *Scand. J. Gastroenterol.* **29**, 607–610 (1994).
- Correa, P., Piazuelo, M. B. & Wilson, K. T. Pathology of gastric intestinal metaplasia: clinical implications. *Am. J. Gastroenterol.* **105**, 493–498 (2010).
- de Vries, A. C. et al. Gastric cancer risk in patients with premalignant gastric lesions: a nationwide cohort study in the Netherlands. *Gastroenterology* **134**, 945–952 (2008).
- Rokkas, T., Filipe, M. I. & Sladen, G. E. Detection of an increased incidence of early gastric cancer in patients with intestinal metaplasia type III who are closely followed up. *Gut* **32**, 1110–1113 (1991).
- Shichijo, S. et al. Histologic intestinal metaplasia and endoscopic atrophy are predictors of gastric cancer development after Helicobacter pylori eradication. *Gastrointest. Endosc.* **84**, 618–624 (2016).
- Jencks, D. S. et al. Overview of Current Concepts in Gastric Intestinal Metaplasia and Gastric Cancer. *Gastroenterol. Hepatol. (N. Y)* **14**, 92–101 (2018).
- Gutierrez-Gonzalez, L. et al. The clonal origins of dysplasia from intestinal metaplasia in the human stomach. *Gastroenterology* **140**, 1251–1260 e1251-1256 (2011).
- McDonald, S. A. et al. Mechanisms of field cancerization in the human stomach: the expansion and spread of mutated gastric stem cells. *Gastroenterology* **134**, 500–510 (2008).
- Kinoshita, H., Hayakawa, Y. & Koike, K. Metaplasia in the Stomach-Precursor of Gastric Cancer?. *Int. J. Mol. Sci.* **18**, 2063 (2017).
- Kumagai, K. et al. Expansion of Gastric Intestinal Metaplasia with Copy Number Aberrations Contributes to Field Cancerization. *Cancer Res.* **82**, 1712–1723 (2022).
- Gupta, S. et al. AGA Clinical Practice Guidelines on Management of Gastric Intestinal Metaplasia. *Gastroenterology* **158**, 693–702 (2020).
- Pimentel-Nunes, P. et al. Management of epithelial precancerous conditions and lesions in the stomach (MAPS II): European Society of Gastrointestinal Endoscopy (ESGE), European Helicobacter and Microbiota Study Group (EHMSG), European Society of Pathology (ESP), and Sociedade Portuguesa de Endoscopia Digestiva (SPED) guideline update 2019. *Endoscopy* **51**, 365–388 (2019).
- Banks, M. et al. British Society of Gastroenterology guidelines on the diagnosis and management of patients at risk of gastric adenocarcinoma. *Gut* **68**, 1545–1575 (2019).
- Filipe, M. I. et al. Incomplete sulphomucin-secreting intestinal metaplasia for gastric cancer. Preliminary data from a prospective study from three centres. *Gut* **26**, 1319–1326 (1985).
- Gonzalez, C. A. et al. Incomplete type of intestinal metaplasia has the highest risk to progress to gastric cancer: results of the Spanish follow-up multicenter study. *J. Gastroen Hepatol.* **31**, 953–958 (2016).

18. Shao, L. M. et al. Risk of gastric cancer among patients with gastric intestinal metaplasia. *Int J. Cancer* **143**, 1671–1677 (2018).
19. Shah, S. C., Gawron, A. J., Mustafa, R. A. & Piazuelo, M. B. Histologic subtyping of gastric intestinal metaplasia: overview and considerations for clinical practice. *Gastroenterology* **158**, 745–750 (2020).
20. Arai, J. et al. OLGIM staging and proton pump inhibitor use predict the risk of gastric cancer. *Gut* **71**, 1043–1044 (2022).
21. Arai, J. Machine learning-based personalized prediction of gastric cancer incidence using the endoscopic and histologic findings at the initial endoscopy. *Gastrointest Endosc.* **95**, 864–872 (2022).
22. Arai, J. et al. Letter: predictive model for gastric cancer after eradication of *Helicobacter pylori*—a survival analysis using a deep learning algorithm. *Aliment Pharm. Ther.* **54**, 528–529 (2021).
23. Lloyd, J. M. & Owens, S. R. CD10 immunohistochemistry stains enteric mucosa, but negative staining is unreliable in the setting of active enteritis. *Mod. Pathol.* **24**, 1627–1632 (2011).
24. Koulis, A. et al. CD10 and Das1: a biomarker study using immunohistochemistry to subtype gastric intestinal metaplasia. *BMC Gastroenterol.* **22**, 197 (2022).
25. Businello, G. Molecular Landscapes of Gastric Pre-Neoplastic and Pre-Invasive Lesions. *Int. J. Mol. Sci.* **22**, 9950 (2021).
26. Sheng, K. L. et al. An integrated approach to biomarker discovery reveals gene signatures highly predictive of cancer progression. *Sci. Rep.* **10**, 21246 (2020).
27. Deng, M., Wang, Y. & Yan, Y. Mining cell-cell signaling in single-cell transcriptomics atlases. *Curr. Opin. Cell Biol.* **76**, 102101 (2022).
28. Zhang, P. et al. Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* **27**, 1934–1947 e1935 (2019).
29. Kim, J. et al. Single-cell analysis of gastric pre-cancerous and cancer lesions reveals cell lineage diversity and intratumoral heterogeneity. *NPJ Precis Oncol.* **6**, 9 (2022).
30. Ianevski, A., Giri, A. K. & Aittokallio, T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications* **13** (2022).
31. Doncheva, N. T., Assenov, Y., Domingues, F. S. & Albrecht, M. Topological analysis and interactive visualization of biological networks and protein structures. *Nat. Protoc.* **7**, 670–685 (2012).
32. Kinoshita, H., Hayakawa, Y. & Koike, K. Metaplasia in the stomach—precursor of gastric cancer? *Int. J. Mol. Sci.* **18**, 2063 (2017).
33. Zhao, C.-M. et al. Denervation suppresses gastric tumorigenesis. *Sci. Transl. Med.* **6**, 250ra115–250ra115 (2014).
34. Rabben, H.-L. et al. Computational drug repositioning and experimental validation of ivermectin in treatment of gastric cancer. *Front. Pharmacol.* **12**, 625991 (2021).
35. Takaishi, S. et al. Identification of gastric cancer stem cells using the cell surface marker CD44. *Stem Cells* **27**, 1006–1020 (2009).
36. Hayakawa, Y., Nakagawa, H., Rustgi, A. K., Que, J. & Wang, T. C. Stem cells and origins of cancer in the upper gastrointestinal tract. *Cell Stem Cell* **28**, 1343–1361 (2021).
37. Pai, P., Rachagani, S., Dhawan, P. & Batra, S. K. Mucins and Wnt/ β -catenin signaling in gastrointestinal cancers: an unholy nexus. *Carcinogenesis* **37**, 223–232 (2016).
38. Hayakawa, Y. et al. Nerve growth factor promotes gastric tumorigenesis through aberrant cholinergic signaling. *Cancer Cell* **31**, 21–34 (2017).
39. Uhlén, M. et al. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
40. Henry, N. L. & Hayes, D. F. Cancer biomarkers. *Mol. Oncol.* **6**, 140–146 (2012).
41. Hartwell, L., Mankoff, D., Paulovich, A., Ramsey, S. & Swisher, E. Cancer biomarkers: a systems approach. *Nat. Biotechnol.* **24**, 905–908 (2006).
42. Lim, N. R. & Chung, W. C. *Helicobacter pylori*-associated chronic atrophic gastritis and progression of gastric carcinogenesis. *Korean J. Gastroenterol.* **82**, 171–179 (2023).
43. Kawamura, M., et al. Endoscopic and histological risk stratification for gastric cancer using gastric intestinal metaplasia. *Journal of Gastroenterology and Hepatology* (2024).
44. Kawamura, M. et al. Endoscopic Grading of Gastric Intestinal Metaplasia Using Magnifying and Nonmagnifying Narrow-Band Imaging Endoscopy. *Diagnostics* **12**, 3012 (2022).
45. Gingold-Belfer, R. et al. The transition from gastric intestinal metaplasia to gastric cancer involves POPDC1 and POPDC3 downregulation. *Int. J. Mol. Sci.* **22**, 5359 (2021).
46. Wei, H. et al. OLFM4 promotes the progression of intestinal metaplasia through activation of the MYH9/GSK3 β / β -catenin pathway. *Mol. Cancer* **23**, 1–19 (2024).
47. Díaz, I. et al. Exploratory analysis of the gene expression matrix based on dual conditional dimensionality reduction. *IEEE J. Biomed. Health Inform.* **27**, 3083–3092 (2023).
48. Greten, F. R. & Grivnickov, S. I. Inflammation and cancer: triggers, mechanisms, and consequences. *Immunity* **51**, 27–41 (2019).
49. Zhao, H. et al. Inflammation and tumor progression: signaling pathways and targeted intervention. *Signal Transduct. Target. Ther.* **6**, 263 (2021).
50. Hagemann, T., Balkwill, F. & Lawrence, T. Inflammation and cancer: a double-edged sword. *Cancer Cell* **12**, 300–301 (2007).
51. Qi, L. et al. Retinoblastoma binding protein 2 (RBP2) promotes HIF-1 α -VEGF-induced angiogenesis of non-small cell lung cancer via the Akt pathway. *PLoS One* **9**, e106032 (2014).
52. Liang, X. et al. Histone demethylase RBP2 promotes malignant progression of gastric cancer through TGF- β 1-(p-Smad3)-RBP2-E-cadherin-Smad3 feedback circuit. *Oncotarget* **6**, 17661–17674 (2015).
53. Zhou, D. H. et al. RBP2 induces stem-like cancer cells by promoting EMT and is a prognostic marker for renal cell carcinoma. *Exp Mol Med* **48** (2016).
54. Choi, H. J. et al. Role of RBP2-Induced ER and IGF1R-ErbB Signaling in Tamoxifen Resistance in Breast Cancer. *J Natl Cancer Inst* **110** (2018).
55. Sugano, K., Moss, S. F. & Kuipers, E. J. Gastric intestinal Metaplasia: real culprit or innocent bystander as a precancerous condition for gastric cancer? *Gastroenterology* (2023).
56. Altman, N. & Krzywinski, M. Association, correlation and causation. *Nat. Methods* **12**, 899–900 (2015).
57. Parsons, M. J., Tammela, T. & Dow, L. E. WNT as a Driver and Dependency in Cancer. *Cancer Discov.* **11**, 2413–2429 (2021).
58. Lei, Z.-N. et al. Signaling pathways and therapeutic interventions in gastric cancer. *Signal Transduct. Target. Ther.* **7**, 358 (2022).
59. Chen, Y., Chen, M. & Deng, K. Blocking the Wnt/ β -catenin signaling pathway to treat colorectal cancer: Strategies to improve current therapies (Review). *Int J Oncol* **62** (2023).
60. Tian, X. et al. Targeting apoptotic pathways for cancer therapy. *The Journal of Clinical Investigation* **134** (2024).
61. Tejada-Muñoz, N., Mei, K.-C., Sheladiya, P. & Monka, J. Targeting membrane trafficking as a strategy for cancer treatment. *Vaccines* **10**, 790 (2022).
62. Rabben, H.-L. et al. Neural signaling modulates metabolism of gastric cancer. *iScience* **24** (2021).
63. Bahar, M. E., Kim, H. J. & Kim, D. R. Targeting the RAS/RAF/MAPK pathway for cancer therapy: from mechanism to clinical studies. *Signal Transduct. Target. Ther.* **8**, 455 (2023).
64. Kantarjian, H., Jabbour, E., Grimley, J. & Kirkpatrick, P. Dasatinib. *Nat. Rev. Drug Discov.* **5**, 717–718 (2006).
65. Braun, A. H., Frank, A. M., Ho, N. & Buchholz, C. J. Dasatinib is a potent enhancer for CAR T cell generation by CD3-targeted lentiviral vectors. *Mol. Ther.-Methods Clin. Dev.* **28**, 90–98 (2023).
66. Rao, S. et al. Leveraging compound promiscuity to identify targetable cysteines within the kinome. *Cell Chem. Biol.* **26**, 818–829. e819 (2019).
67. Konings, I. R. H. M. et al. Phase I and pharmacological study of the broad-spectrum tyrosine kinase inhibitor JNJ-26483327 in patients with advanced solid tumours. *Br. J. Cancer* **103**, 987–992 (2010).

68. Boulos, J. C., Chatterjee, M., Shan, L. & Efferth, T. In silico, in vitro, and in vivo investigations on adapalene as repurposed third generation retinoid against multiple myeloma and leukemia. *Cancers* **15**, 4136 (2023).
69. Kim, J. et al. The anti-diabetic drug gliquidone modulates lipopolysaccharide-mediated microglial neuroinflammatory responses by inhibiting the NLRP3 inflammasome. *Front. Aging Neurosci.* **13**, 754123 (2021).
70. Renner, O. et al. Systematic review of Gossypol/AT-101 in cancer clinical trials. *Pharmaceuticals* **15**, 144 (2022).
71. Muller, C. et al. The troglitazone derivative EP13 disrupts energy metabolism through respiratory chain complex I inhibition in breast cancer cells and potentiates the antiproliferative effect of glycolysis inhibitors. *Cancer Cell Int.* **24**, 132 (2024).
72. Jia, X. et al. PPAR γ agonist pioglitazone enhances colorectal cancer immunotherapy by inducing PD-L1 autophagic degradation. *Eur. J. Pharmacol.* **950**, 175749 (2023).
73. Rapoport, B. L. Rolapitant: An NK-1 Receptor Antagonist for the Prevention of Chemotherapy-Induced Nausea and Vomiting. *Rev. Recent Clin. Trials* **12**, 193–201 (2017).
74. Sethi, A., Joshi, K., Sasikala, K. & Alvala, M. Molecular docking in modern drug discovery: Principles and recent applications. *Drug Discov. Dev. - N. Adv.* **2**, 1–21 (2019).
75. Bhagat, R. T. et al. Molecular docking in drug discovery. *J. Pharm. Res. Int.* **33**, 46–58 (2021).
76. Wang, T. C. et al. Synergistic interaction between hypergastrinemia and Helicobacter infection in a mouse model of gastric cancer. *Gastroenterology* **118**, 36–47 (2000).
77. Iacono, G., Massoni-Badosa, R. & Heyn, H. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biol.* **20**, 110 (2019).
78. Thul, P. J. et al. A subcellular map of the human proteome. *Science* **356**, eaal3321 (2017).
79. Karlsson, M. et al. A single-cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, eabh2169 (2021).
80. Uhlen, M. et al. A pathology atlas of the human cancer transcriptome. *Science* **357**, eaan2507 (2017).
81. Uhlen, M. et al. Towards a knowledge-based human protein atlas. *Nat. Biotechnol.* **28**, 1248–1250 (2010).
82. Berglund, L. et al. A gene-centric Human Protein Atlas for expression profiles based on antibodies. *Mol. Cell. Proteom.* **7**, 2019–2027 (2008).
83. Uhlen, M. et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell. Proteom.* **4**, 1920–1932 (2005).
84. Pontén, F., Jirstrom, K. & Uhlen, M. The Human Protein Atlas—a tool for pathology. *J. Pathol.: A J. Pathological Soc. Gt. Br. Irel.* **216**, 387–393 (2008).
85. Berman, H. M. et al. The protein data bank. *Nucleic acids Res.* **28**, 235–242 (2000).
86. Morris, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Computational Chem.* **30**, 2785–2791 (2009).
87. Irwin, J. J. et al. ZINC20—a free ultralarge-scale chemical database for ligand discovery. *J. Chem. Inf. Modeling* **60**, 6065–6073 (2020).
88. O’Boyle, N. M. et al. Open Babel: An open chemical toolbox. *J. Cheminformatics* **3**, 1–14 (2011).
89. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Computational Chem.* **31**, 455–461 (2010).
90. Schrodinger, L. The PyMOL molecular graphics system. *Version* **1**, 8 (2015).
91. Zhang, P. et al. Dissecting the Single-Cell Transcriptome Network Underlying Gastric Premalignant Lesions and Early Gastric Cancer. *Cell Rep.* **30**, 4317 (2020).

Acknowledgements

This study was supported by the Liaison Committee between the Central Norway Regional Health Authority (Helse-Midt Norge RHF) and Norwegian University of Science and Technology (NTNU), Joint Program of the Medical Faculty of NTNU and St. Olavs University Hospital, the Cancer Foundation of St. Olavs Hospital (Kreftfondet ved St. Olavs Hospital). The roles of the funding bodies included the design of the study and material collection.

Author contributions

GTA participated in study development, data collection and management, data analysis, and manuscript writing. AI participated in data analysis and collection, and manuscript writing. MR participated in data analysis, and manuscript writing. NP participated in data analysis, particularly molecular docking analysis, and manuscript writing. HR participated in study development, data collection and analysis, manuscript editing. SG participated in study development, data analysis, and manuscript writing. YK participated in study development, data collection, and manuscript writing. HT participated in study development, data collection, data analysis, and manuscript writing. DK participated in data analysis, and manuscript writing. JEG participated in study development, data collection and management, data analysis, and manuscript writing. TCW participated in study development, data analysis, and manuscript writing. CMZ participated in study development, data collection and management, data analysis, and manuscript writing. DC participated in study development, data collection and management, data analysis, and manuscript writing. All authors have read and approved the final manuscript.

Funding

Open access funding provided by NTNU Norwegian University of Science and Technology (incl St. Olavs Hospital - Trondheim University Hospital).

Competing interests

Parts of the results were included in an MSc thesis by S. Geithus (<https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2993496?show=full>) and a PhD thesis by G.T. Andersen (Doctoral theses at NTNU, 2023:360). Otherwise, the authors declare that they have no conflict of interest.

Additional information

Supplementary information The online version contains supplementary material available at

<https://doi.org/10.1038/s41540-024-00455-0>.

Correspondence and requests for materials should be addressed to Duan Chen.

Reprints and permissions information is available at <http://www.nature.com/reprints>

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024